Location-based marketing

Location-Based Marketing

Six evidence-based Narratives on the Future of Hyper-Targeting in Cities

Hannah Sophie Schmitt

ISBN: 978-94-6469-224-2

Cover design by Hannah Sophie Schmitt

Printed in the Netherlands by ProefschriftMaken | www.proefschriftenmaken.nl

Location-Based Marketing

Six evidence-based Narratives on the Future of Hyper-Targeting in Cities

DISSERTATION

To obtain the degree of Doctor at Maastricht University,

on the authority of the Rector Magnificus, Prof. Dr. Pamela Habibović,

in accordance with the decision of the Board of Deans,

to be defended in public

on Monday, 27th February 2023, at 13:00 hours

by

Hannah Sophie Schmitt

Supervisor

Prof. Dr. Jos G.A.M. Lemmink

Co-Supervisor

Dr. Benjamin J. Lucas (University of Nottingham)

Assessment Committee

Prof. Dr. Gaby J. Odekerken-Schröder (Chair)

Prof. Dr. Joop de Kraker

Prof. Dr. Martin G.M. Wetzels (EDHEC Business School, Lille)

Dr. Francisco J. Villarroel Ordenes (LUISS Guido Carli University, Rome)

Für meine Familie

Acknowledgements

"Life is like riding a bicycle – to keep the balance, you must keep moving."

This wonderful quote has become one of my favorite ones over the years and it could not be more fitting for this personal introduction to my dissertation. Not only have I had a sign with this quote in my (PhD) student room and later home office room for many years, but it also became a mantra during the PhD journey and a reminder to keep going when things were tough. I would like to modify the quote a little bit by saying "Doing a PhD is like riding a bicycle – to keep the balance, you must keep moving" and then, this metaphor fits even better. First, my thesis topic focuses on location-based marketing and customer mobility in urban areas and bicycles are a core part of urban mobility. Second, Maastricht and the Netherlands, a passionate biking nation, was home to me for 7 years and an academic home for 10 years. And last, keeping the (work-life) balance has been a goal throughout my PhD and I am happy to say that I achieved that. Many people contributed to me keeping the balance and keeping moving in the past six years in various ways and I would like to take this opportunity to thank you all.

Staying within the metaphor of the bicycle, almost exactly 6 years ago, on 5th February 2017 I started my bike ride as a PhD student at Maastricht University and had no idea what was ahead of me. Looking back, I can confidently say that it was a crazy ride with many unexpected turns, various highs and lows, steep learning curves and a whole lot of fun. First and foremost, I would like to thank my supervisors. Jos, my steering wheel and the most wonderful supervisor, thanks for everything you have done for me. You are one of the most genuine and supportive people I have ever met and your enthusiastic and warmhearted personality are so inspiring. Thank you for guiding me through the sometimes-bumpy hills of the PhD journey and for steering me into the right direction when I was lost. Thank you for giving me the freedom to work on what I am most interested in, for familiarizing yourself with new topics to assist me and for your open-mindedness to creative and out-of-the-box research ideas. Most importantly, thanks for always believing in me when I was not able to do so myself – your continuous positivity and trust in my abilities has kept me sane throughout the years. And Ben, the wheels of my bike, you offered me the PhD position and brought me onto this crazy ride. Thanks for helping me find my niche and shaping my research interest into a concrete dissertation topic. Your enthusiasm for everything that is "different" has inspired me and your crazy visions have positively influenced me in many ways. Thanks for introducing me to many research approaches and tools outside of traditional marketing research tools and thus, making this dissertation such a special one. When I was sitting in my marketing class during my bachelor exchange in Australia listening to your tutorial, I definitely did not think that 8 years later you would be sitting next to me during my PhD defense. In German we have this saying "Man sieht sich immer zweimal im Leben", meaning that you always meet a person twice in your life for a good reason, and I am very thankful that this saying holds for us.

Second, I would like to thank the assessment committee Gaby Odekerken-Schroeder, Joop de Kraker, Martin Wetzels and Francisco Villarroel Ordenes. In my metaphor, I guess you would be the Dutch bicycle police that frequently checks whether your bike is in good conditions and fines you if it is not. Gaby, thank you for agreeing to chair this

committee and thanks to all of you for taking the time to read my dissertation, for the valuable feedback and for concluding that my "bike" was in good condition after 6 years.

Next, I would like to thank my fellow cyclists, my inner circle and social home in Maasi Nina, Timna, Dinah and Tim, who were with me on this crazy journey all along. We share the same experience and I knew I could always turn to you all whenever I had issues with my "bike" or anything else. We would try to fix things together, whether that meant long discussions one-on-one, group lunches at SBE and later weekly Zoom calls to stay in touch, board game nights or simply dinner, wine and high-quality German Trash TV. Ich wüsste wirklich nicht, was ich ohne die *Intellektuelle Stimulation* machen würde und wie ich die letzten 6 Jahre überstanden hätte. Nina, meine SBE Sonne, wir kennen uns nun seit über 10 Jahren und haben so viel zusammen erlebt seit der INKOM 2012. Ich bin einfach unendlich froh, dass es dein riesengroßes Herz und deine unermüdliche Unterstützung in meinem Leben gibt und hoffe, dass das noch sehr lange so bleibt. Timna, auch wenn wir zusammen studiert haben, haben wir uns erst als PhDs richtig kennen gelernt und ich bin wahnsinnig froh darüber – erst fellow MSCMers, dann schnell Freundinnen und später sogar kurz Mitbewohnerinnen. Ich danke dir für dein offenes Ohr, deine Unterstützung und Denkanstöße und ganz besonders für unsere daily morning motivation calls, ohne die wir wahrscheinlich beide immer noch nicht fertig wären. Dinah, irgendwie muss ich gerade einfach daran denken, wie du in deinem Office hinter den großen Bildschirmen sitzt und wie wild in die Tastatur hämmerst oder durch die Flure der SBE wirbelst. Danke für deine herzliche Art, deine liebenswerte Ruhrpottschnauze und danke an Lars und dich für eure Gastfreundschaft all die Jahre und als Auffangstation für Pendler. Tim, auch wir sind uns im Studium nie begegnet und haben uns erst kennen gelernt, als du dir das Büro mit Nina geteilt hast. Danke für inspirierende Diskussionen, scharfe Analysen, unverblümte Ehrlichkeit und dass du es echt 6 Jahre mit dem Hühnerhaufen ausgehalten hast und den ein oder anderen Cringe-Moment vor dem Fernseher ertragen hast. Nina and Timna, it is a great pleasure and honor to have you both standing by my side today as my paranymphs and a tremendous thanks for your help in organizing the defense and for making this day so special!

And when I talk about my Maasi inner circle, Inka you are definitely part of it! Thank you for being the best roommate one could ask for, you are one of a kind and one of the most affectionate people I know. Even though we did not know each other before, we clicked the first time I viewed the apartment, bonded over a cup of tea and the rest is history. Your chilled and at the same time hard-working personality amazes me every day and I am incredibly grateful for all the evenings and weekends we spend together on the couch, the balcony barbequing or dog-sitting Dommel.

Further, I would like to thank all my colleagues at BISS 1.0, the bicycle frame that provided me the structure and frame to work and grow in. Thanks for giving me the opportunity to do a PhD with societal relevance in a vibrant, innovative environment close to industry where data-driven approaches to business questions fall on fruitful ground and where an applied research focus is appreciated. Despite all the challenges that come with working in a newly founded research institute, I am thankful for the trust that was put into me and that I had the opportunity to do my first steps in the academic world here. Special thanks goes to Rudolf Müller as the director of BISS and to the project managers Martine, Racheel and Ellen who provided the base of our company co-creation projects. Alex and Wiebke, thank you two so much for the fun commutes to Heerlen, the motivational and insightful talks, for challenging me to continuously reflect, for sharing your experience and for your guidance in my early PhD years. Jeroen and Michelle, my fellow BISS PhDs, thanks for the fun lunches and table football matches NL vs. DE including the office score board and the "BISS, booze & boss" nights at Centrifuge to pay off football debts. Deniz, thanks for the positive energy in the office and Kay, thanks for the long chats about life and your hilarious metaphors for every situation. Also thanks to all other colleagues and project partners for the fruitful and warm atmosphere of support, inspiring talks and projects.

When my first bike BISS 1.0 "broke down" in spring 2019, the wonderful colleagues at the MSCM department offered me a new bike and I could smoothly continue my journey as a PhD. Thank you Gaby and Dominik for guiding the department as department heads during my time as a PhD student. Steffi, we started this PhD journey together in the department in February 2017 and discovered life as a PhD together with all the joys and hurdles. Thanks to the office F2.02, Ruud, Kars, Tim & Steffi, for giving me a home in Maastricht during BISS times and for the fun office days at SBE in the first two years; and of course also office F1.13, Mark S, Mark G and Mahdi, thanks for all the great conversations about work and life and the always supportive atmosphere. Mark S, you positive energy and craziness made days at SBE more enjoyable for sure. Pascalle and Nicole, thanks for your constant and never-ending assistance in all admin topics and most importantly for being the good souls of the department. Last but not least, thanks to my fellow MSCM PhDs: Susan and Martina for a fun time at Frontiers 2019 and for introducing me to the world of academic conferences, Eric and Mathilde for numerous coffee chats about data-driven research and PhD life, and thanks all other PhDs and colleagues whom I met over the years for a great time at SBE.

Bicycles cannot move without a chain and the chain of my PhD bike was the data that allowed me to execute my research ideas. Thanks to Jonas Schorr for a wonderful summer course in Berlin that sparked my interested in urban mobility and for inviting me to a conference the same year where I made the connection to InnoZ for my data collection. Special thanks to Enrico Howe and Lena Damrau from InnoZ Berlin and Modalyzer for their positive energy, the amazing collaboration on the data collection and the data anonymization. Further, thanks to the Gemeente Heerlen and Maastricht Bereikbaar for assisting in recruiting participants for the data collection. Last but definitely not least, thanks to Bram Oosterbroek for sharing your knowledge about spatial data analysis and ArcGIS with me and for patiently answering all my questions over the years.

A welcoming biking destination on the other side of the river was always IDS with its inspiring and data-oriented "e-cyclists". Special thanks to Chang and Nadine for starting YDS together, growing and learning while changing the format multiple times. Chang, it was a true pleasure organizing two WiDS Datathons with you and I am glad we met during our times in Maastricht. Also thanks to all other colleagues at IDS for the fun times and open doors whenever I visited your offices.

Riding a bike is more fun in a larger group. Therefore, thanks to the fellow PhDs I met outside of MSCM and BISS. Nora, Marina, Lidwien, Anouk and Irene, thanks for the fun evenings in the Vijfharingenstraat and for letting me join your Finance girls nights.

Martijn and Mike, thanks for bringing in the necessary humor to SBE days whenever I met you in the hallways. Nina S and Kasia, thanks for great times whenever I was out and about with the OSE crowd and Nina S, thanks for the wonderful days in Hamburg and brainstorming on creative projects. Sean, it's always fun having you around for a BBQ, karaoke session, bar night or party. Alex, thanks for fun board game evenings, handball WM watching and skiing with the crew. Suus, I am glad I got to know you better during our skiing trips and thanks for the nice evenings when you hosted me in the guest room of V8C01.

Last and most importantly, a huge thank you goes out to my family and to my friends outside of the PhD world. In all of this, you are the bike shop that even allowed me to get out on the road and experience this crazy ride and at the same time the repair shop that fixed my bike countless times and the café that offered me a tea and cake when skies were grey and I needed a shoulder to rest on. To my lovely friends, you are best people one can ask for and I consider myself incredibly grateful to have you in my life. Jasmin, seit fast 20 Jahren erleben wir gemeinsam alle Höhen und Tiefen des Lebens und unterstützen uns gegenseitig so gut es geht. Ich könnte mir keine bessere Beste Freundin als dich vorstellen – keine hört so gut zu wie du, findet immer die richtigen Worte und ist immer für mich da, egal wie viele Kilometer uns auch gerade trennen. Du bist ein Geschenk und ich freue mich wahnsinnig, dass es jetzt nur noch zwei Stunden zu dir sind und nicht mehr sechs. Lenni und Flo, wir kennen uns seit der ersten Klasse und es ist einfach verrückt, dass ihr immer noch in meinem Leben seid. Danke für unzählige Abenteuer, die ich jetzt nicht alle aufzählen kann und für euren fantastischen Humor, der mir immer wieder ein Lächeln ins Gesicht zaubert. Caro, du treue Seele seit der Krabbelgruppe, ich bin unseren Eltern sehr dankbar, dass sie sich damals gut verstanden haben, denn ich möchte dich in meinem Leben nicht missen. Du kennst alle Herausforderungen einer Doktorarbeit und ich liebe es, dass auch diese Erfahrung uns verbindet. Ich freue mich schon sehr, wenn wir nächste Saison hoffentlich wieder gemeinsam über die Pisten düsen – das muss zur Tradition werden. Britta, Paula und Michelle, meine Maasi Girls aus dem Studium, danke, dass ich immer auf euch zählen kann, dass ihr immer ein offenes Ohr habt und ihr es mir verziehen habt, wenn ich mich nicht so regelmäßig gemeldet habe bzw. euch dann erst recht gemeldet habt um nachzufragen, ob alles ok ist. Ich glaube ich habe euch viel zu selten gesagt, wie froh ich bin, dass es euch gibt und ich freue mich auf viele weitere gemeinsame Jahre voller wertvoller Freundschaft, Kaffeekränzchen und stundenlanger Telefonate. Danke auch an Joachim, Elke und Alex für eure Unterstützung, euer Interesse und die herzliche Aufnahme in die Familie.

Meine lieben Geschwister, Jannis und Mareike, auch wenn ich es nicht immer leicht mit euch hatte und ihr sicherlich auch nicht mit mir (vor allem als wir jünger waren), bin ich Mama und Papa unendlich dankbar, dass es euch beide gibt. Geschwister sind Freunde fürs Leben, die man geschenkt bekommt und ich könnte nicht stolzer auf uns sein, was für eine tolle Beziehung wir zueinander haben vor allem jetzt, wo wir alle Kinderstreitigkeiten hinter uns gelassen haben. Nur zu gerne denke ich zum Beispiel an den letzten Sommer zurück, wo wir als Geschwister zusammen ein ganzes Wochenende auf einem Festival waren und ich hoffe, dass wir uns jetzt noch häufiger sehen, wo ich endlich wieder zurück bin im Süden. Mama und Papa, ich weiß gar nicht, wo ich anfangen soll. Ich bin euch wahnsinnig dankbar, für alles was ihr für mich getan habt.

Ihr habt mich immer bei allem unterstützt mit eurer bedingungslosen Liebe und mir alles ermöglicht, wovon ich geträumt habe. Vieles davon weiß ich heute mehr zu schätzen denn je und ohne euch wäre ich heute nicht der Mensch, der ich bin. Ich verdanke es euch auch, dass ich heute hier stehen darf, weil ihr immer das Potenzial in mir gesehen und gefördert habt und meine Ausbildung ermöglicht habt. Von ganzem Herzen danke für alles! Und zum guten Schluss Jan, vor etwa 9 Jahren bist du vollkommen unvorhergesehen in mein Leben gestolpert und einfach geblieben. Ich weiß, dass du es nicht immer einfach hattest mit mir und bin umso glücklicher, dass du dich durchgebissen hast und immer noch hier bist. Ich könnte mir niemand Besseren vorstellen, mit dem ich das Abenteuer Leben gemeinsam bestreiten möchte. Du bist mein Fels in der Brandung, meine Schulter zum Ausheulen, du bringst mich auf den Boden, wenn mir alles zu Kopf steigt, unterstützt all meine Ziele und bringst mich immer wieder zum Lachen. Dafür danke ich dir!

Thank you everybody for being part of this crazy bike ride called PhD journey. You made it very special!

Hannah Schmitt,

Maastricht, February 2023

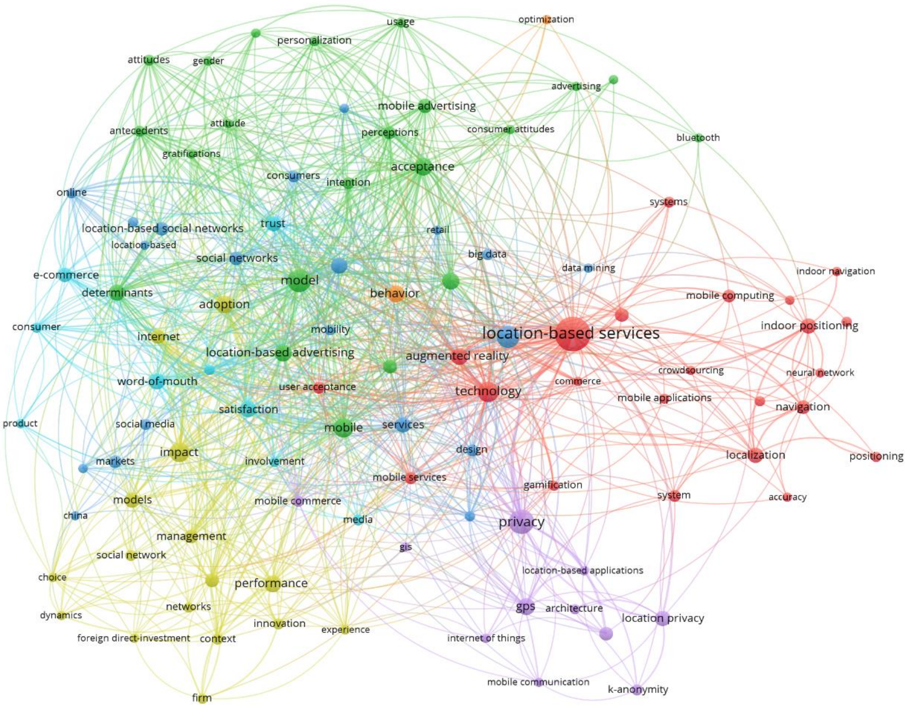
Table of Contents

List of Figures

List of Tables

List of abbreviations

BAG	Basisadministratie Gebouw
CBM	City-based marketing
CBS	Centraal Bureau voor Statistiek
DBSCAN	Density-based spatial clustering of applications with noise
GDPR	General Data Protection Regulation
GIS	Geographic information systems
GPS	Global positioning system
GSM	Global system for mobile communication
KDE	Kernel Density Estimation
LBM	Location-based marketing
MIT	Massachusetts Institute of Technology
OD	Origin-Destination
OSM	OpenStreetMap
POI	Point of interest
UAM	Urban air mobility
UNA	Urban network analysis
VTOL	Vertical takeoff and landing



Introduction

Using City-Based Marketing to Shape the Future of Hyper-Targeting

Location-based marketing, city-based marketing and hyper-targeting

Have you ever pulled into an IKEA parking lot and immediately received a notification from the IKEA app on your phone that welcomes you, invites you to the bistro for a free coffee, and offers personalized coupons? This is a typical example of hyper-targeting through location-based marketing (LBM) meant to enhance the customer experience. LBM refers to a direct marketing strategy that uses location information to deliver marketing content on mobile devices relevant to a particular location. IKEA uses LBM to increase customer satisfaction and trigger unplanned or complementary purchases. However, such an LBM strategy only works for customers who have the company's app installed; the app uses beacons or geo-fencing to detect that the customer is approaching a store location.

To reach more potential customers who do not have the company's app installed, IKEA and like-minded businesses can engage in what we call city-based marketing (CBM). This term refers to a geographically bounded form of LBM that includes more data about the city and urban living—not merely the static customer positioning observed via an app or mobile phone. CBM uses additional data on phenomena such as the following: pollution, the built environment and physical infrastructure, public transport usage, consumer location, consumer interests (derived from social media), and consumer online and offline purchasing behavior. The goal is to generate a more complete picture of the customer's behavior and interests in their preferred environment and to deliver specific advertising for products, services, or events in the city. This practice of placing very specific advertising in an online and offline context is known as hyper-targeting. Previous researchers have overlooked a great deal of information with the narrow definition of LBM, resulting in a noteworthy knowledge gap. We need to fill this gap by gaining an understanding of the dynamics of consumer behavior in cities, which will allow us to provide super-specific, curated marketing offerings. Hardly any previous studies have applied movement data from proxemic technologies to marketing, and little is known about such data's worth and the technicalities of how to include it in marketing research. Therefore, the overall goal of this dissertation is to *provide evidence that proxemic technologies can enhance hyper-targeting and that these technologies are vital to the dynamic perspective in marketing*.

According to Czarny (2018), usage of location data is "one of the most misunderstood areas in marketing," and marketers struggle to find suitable ways to utilize it. Previous researchers have examined different aspects of location data that collectively highlight why geography is relevant in marketing; we can group these aspects into four categories.

First, the manner in which an individual's favorite locations are scattered is indicative of their general mobility behavior. Bettman et al. (1998) find that location choice correlates with product preference. Further, Provost et al. (2015) emphasize that individuals with similar demographics and socioeconomic characteristics tend to visit similar locations and thus construct the so-called geo-social networks of mobile phone users. This finding is supported by McPherson et al. (2001) theory on homophily and the extension that this phenomenon is mainly caused by the limited opportunities an individual has within a certain geographic radius (Kossinets & Watts, 2009).

Consequently, geographic boundaries drive location homophily, which determines the effectiveness of targeted location-based advertising.

Second, there is a significant interaction effect between trip timing and location choice. Luo et al. (2014) find that mobile marketing achieves higher response rates when it matches the mindset of the targeted consumer. Estimations of the current mindset are inferred from factors such as the time of day or the starting and ending time of trips combined with the specific locations (Bargh et al., 2001). The dependency on time and location is further supported in a study by Molitor et al. (2012), which finds that consumers who use the same smartphone app in the same area at the same time have similar product preferences and make similar purchasing decisions (Zubcsek et al., 2016). Consequently, the temporal dimension of a movement matters in marketing, and it is crucial for understanding human decision-making in cities.

Third, the long-standing economic principle of the time value of money (Marshall, 1926) leads to the integration of geography in marketing. This principle states that humans assign costs to their time spent on all activities, including movement and travel. Therefore, the mobility choice is also affected by an individual's "psychological time," which refers to the phenomenon in which travel speed influences the perception of trip duration (Brown, 1995).

Finally, the contextual features around movement influence consumers' perceptions of location-based advertising and their tendencies to make actual purchases (Andrews et al., 2016; Zhang & Katona, 2012). Such contextual features of a trip include the purpose or activity at the end of the movement, whether the movement is a multi- or single-purpose trip, available modes of transport, and externalities such as weather conditions. All the studies cited above advocate for the integration of spatial data into marketing and consumer research.

LBM efforts based on the consumer's static location, such as ad hoc price promotions, cover only a small part of the many opportunities associated with spatial data in marketing. Hyper-targeting presents another opportunity for the use of spatial data. In this practice, marketers use as many attributes about the customer as possible to deliver extremely customized messages and place them where the customer is likeliest to see them. This makes the customer feel special and increases brand loyalty. The data that marketers use to draw clear personas among their target groups can be divided into five categories. First, demographic information about the customer, such as age, gender, job or income level, and family situation, builds the groundwork for the persona. Second, via GPS on the customer's phone or geo-tagged social media posts, marketers can learn about frequently visited locations and geographic behavior such as the origin and destination of a movement, the mode of transport used, and trip timing. Third, via social media accounts, we can gather information on the customer's interests based on their own posts and content they liked. Fourth, a company can use its own website and social media pages, as well as credit card information, to observe the customer's online purchase behavior. Finally, proxemic technologies can be used to track the customer's offline purchase behavior and store visit routines. When combined, these five categories of data allow for the generation of a very concrete picture of the customer, which, in turn, allows for the accurate placement of advertising.

Evolution of the field over time

Now that the importance of including the dynamic spatial perspective in LBM is clear and we have established hyper-targeting as one use case, let us examine the development of the academic field over time and which topics have been addressed so far. In a relatively new research domain such as LBM, the time trend and content of publications can be particularly useful in providing insights into the general understanding of the domain. I decided to take a data-driven approach to creating such an overview with the help of VOSviewer and Web of Science. Searching Web of Science for "location-based marketing" in July 2021 led to 702 results. The platform itself provides a visualization of the time trend in absolute numbers of annual publications (Figure 0.1).

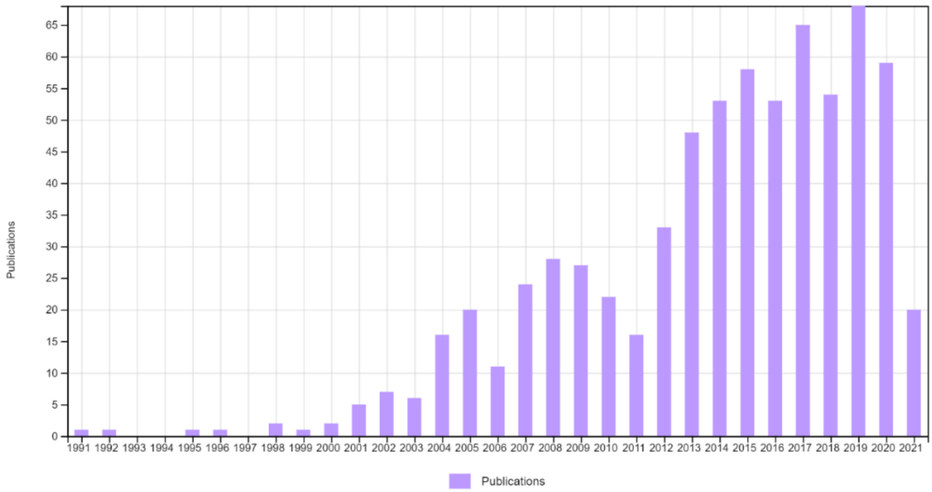


Figure 0.1: Time trend of location-based marketing publications

Prior to 2004, between one and four papers were published each year. From 2004 to 2011, between 12 and 28 papers were published each year, demonstrating the increased interest of academia in the effects of digitalization. In the beginning, LBM was only out-of-home, non-mobile marketing in public places. Over the years, digital marketing has been added to the LBM portfolio, and nowadays, many LBM strategies are purely digital, neglecting out-of-home marketing.

As of 2012, a rapid increase in publications on LBM is observable. This rise in attention makes sense; in 2007, the first smartphone (iPhone) introduced smart sensing and GPS technology into ubiquitous consumer devices. Industry professionals quickly jumped on this train of opportunities and developed not only growth strategies and marketing campaigns but also entire new business ideas around this disruptive innovation. Popular industry examples include the introduction of smart watches and smart home solutions, which, along with numerous apps, make use of smart sensing data. Further, companies started using cookies and GPS data for personalized advertising or took the opportunity to reach and interact with the consumer 24 hours a day. In addition to these industry gains, the introduction of smart sensing allowed academics to collect large amounts of data about individuals, including their current locations and their (movement) behavior. The first proxemic technology data used in LBM research was from Foursquare check-ins and similar check-in data, in which consumers shared their current locations. Today, the sensing data landscape is vastly different; it includes data from many sources that cover a variety of urban living aspects. The most popular devices are wearable consumer devices such as smartphones and smartwatches, as well as stationary devices used to observe pollution or weather conditions and to store geo-fencing data.

The explanation for the five-year delay between the introduction of smartphones and the increase in publications is twofold: (1) it took academics time to understand the opportunities and to develop appropriate methods and tools to make sense of the newly available big data; (2) it took consumers time to adopt smartphones and integrate them into their daily lives. Since 2012, the topics within the LBM field that are of interest to academics have diversified and progressed. Figure 0.2 illustrates the literature development over time, based on the 702 articles on "location-based marketing" from Web of Science.

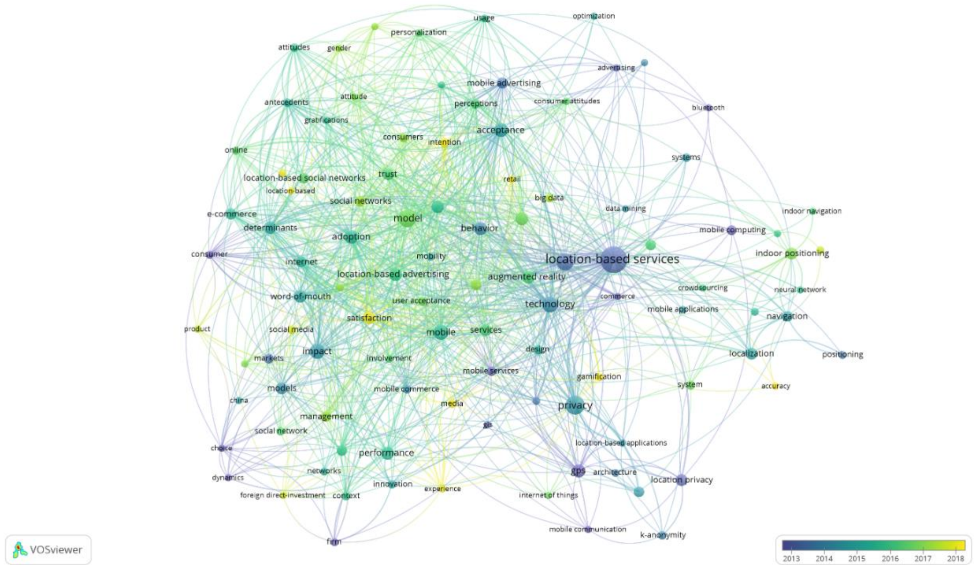


Figure 0.2: Literature development over time

This deeper analysis of the evolution of topics over time was carried out in VOSviewer. The tool draws maps based on network data, bibliographic data, or text data, and it allows various analysis specifications (for more detail on the complete functionalities of VOSviewer, see (van Eck & Waltman, 2010, 2014; Waltman et al., 2010). In this overview, the focus is on a co-occurrence analysis with all the included keywords from bibliographic data. The graph presents all 103 keywords that appeared at least five times, starting in 2012. The publications in 2012 and 2013 centered on location-based services through mobile computing and advertising enabled by emerging technologies such as GPS. In 2014, academia added research on localization and navigation as well as user acceptance and privacy. Publications in 2015 focused more on the impact of the internet, e-commerce, and word-of-mouth marketing, while publications in 2016 connected mobile marketing and competition with LBM. In 2017, publications shifted toward social media and big data, followed by indoor positioning, modeling, and trust-related topics in 2018 and then user acceptance and products in LBM in 2019. Since 2020, publications in the LBM field have focused on satisfaction, experience, intention, media, gamification, and accuracy. The research presented in this dissertation integrates recently included concepts, adding new perspectives and insights.

Current topic clusters within location-based marketing

In addition to the visualization of the development of LBM literature, I would like to provide an interactive network visualization of topic clusters. When using the same dataset of 702 publications (up to July 2021) and the resulting 103 keywords with at least five occurrences, I identified seven clusters: three large, three medium-sized, and one small (Figure 0.3).

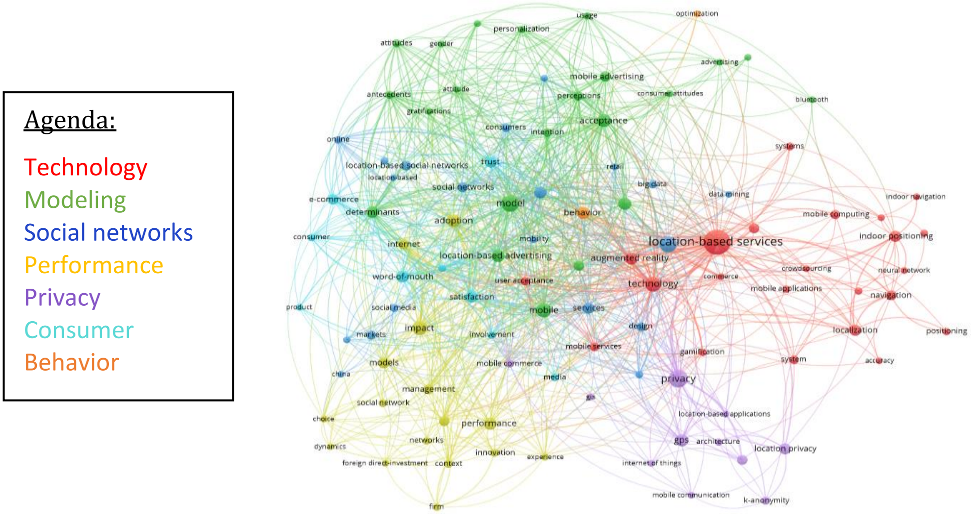


Figure 0.3: Topic clusters in location-based marketing

The largest cluster (red) is themed around technology and includes keywords such as (indoor) positioning, mobile computing, navigation, localization, and accuracy. Unsurprisingly, these technological aspects of location-based services constitute the largest cluster. In the search summary provided by Web of Science, it was already clear that most of the 702 research papers on LBM were published in the engineering/computer science domain and that less than 20% of these papers were published in the business/management domain. The second largest cluster (green) focuses on marketing modeling and mobile advertising, as well as their usage, acceptance, perception and personalization. In the evolution of the field, we have seen that advertising was among the first use cases of LBM; as a rather advanced topic within the domain, advertising forms a larger cluster. The third cluster (dark blue) is built around social networks and includes big data and data mining from social media as well as mobile marketing on such platforms. Big data and social media analytics in marketing are very young but rapidly growing disciplines; therefore, I expected to find them in combination with LBM as well.

The fourth cluster (yellow) presents impact, a topic that came up quite early and has been of stable interest in the LBM field. Within this cluster, we can also find keywords such as performance, competition, innovation, dynamics, and adoption; thus, papers assigned to this cluster are highly relevant for the management of location-based services. In a way, this cluster acknowledges that LBM has a dynamic component that is relevant to the competition between businesses and to the reaction to advertising. However, I did not find any studies that satisfactorily explored and exploited the myriad possibility of the dynamic perspective in marketing. Another medium-sized cluster (purple) deals with privacy and data security, topics that are still underrepresented in many research fields, including LBM, although it is a good sign that the consumer's privacy receives attention at all. Research has addressed privacy continuously since 2014. However, academia currently only examines the privacy concerns of consumers and does not yet focus on how researchers can overcome the data gap caused by privacy regulations. Next, a younger, smaller, and mainly marketing-driven cluster (light blue) contains research efforts that put the consumer in the spotlight. Publications of this sort include keywords such as satisfaction, trust, word-of-mouth, product, e-commerce, and media.

Finally, the smallest cluster (orange), on behavior, is centrally placed in the network and has connections to all other clusters. This centrality of the smallest cluster is uncommon and thus quite remarkable. It indicates that this keyword has tight connections to other, nearby keywords and that it plays a crucial role in all other topic clusters. This cluster is still small because, as we have seen in the timeline, behavior is the most recent research topic. Hence, behavioral aspects combined with LBM will be the next trend and will affect and connect to the other six clusters within the research field. I will address this topic in my dissertation and add the dynamic perspective of individuals, which is behavioral by nature. Moreover, the final chapter will focus solely on citizen behavior while implicitly addressing all the other topics.

Research gaps

In this dissertation, I address three research gaps—in the following labeled as mismatches between common research practice and research goal—currently existent in the LBM field. First, surveys among marketing professionals and marketing agencies found that approximately 45% of companies were dissatisfied with their LBM effectiveness (Bues et al., 2017). Despite using state-of-the-art technologies and practices, these firms do not manage to capitalize on their LBM spending. Desirable outcomes include increased spending of the targeted customers, increased store or brand awareness in the area, increased overall shop revenue, etc. One possible explanation of this dissatisfaction among professionals is the limited research on consumer behavior in terms of LBM. As mentioned above, only ~15% of the articles on LBM accessible via Web of Science originate from the business/management domain. The remaining papers stem mostly from computer science/engineering and, consequently, have a different take on LBM, mainly focusing on the technical feasibility and use cases of GPS position, navigation, localization, etc. (Barbosa et al., 2018; Celik & Dokuz, 2017; Dokuz & Celik, 2017; Pengfei Wang et al., 2017). As a result, even though we are discussing location-based *marketing*, there is limited knowledge regarding consumer perception of and reaction to LBM efforts. This mismatch between the research domain that has published the most about LBM (computer science/engineering) and the domain that implements LBM in practice (business/management) could be one explanation for the current dissatisfaction with LBM effectiveness among marketing professionals.

Second, previous research only focuses on the consumer's location—not on the infrastructure in the area where the consumer currently is. For example, a store collects data about customer visits via platforms such as Foursquare or uses geo-fencing and its own app to contact potential customers via push messages as soon as they enter a predetermined radius around the store's location (Noë et al., 2016; Noulas et al., 2012). Therefore, the focus is solely on the consumer's proximity—not on the environment itself. Examining the surrounding area of a store location is beneficial from two distinct perspectives: (1) retailers looking for customers can utilize complementary stores and facilities to raise awareness about their own store; (2) customers looking for specific retailers or service providers benefit from meaningful conglomerates of stores (D'Silva et al., 2018).

Third, until now, LBM has dealt with static information tied to the specific customer position (Ameen et al., 2021). However, using static information on the customer's proximity to a store represents only the tip of the iceberg of LBM's potential. The massive, submerged part of the iceberg revolves around dynamic information and collects data about customers' trips to the store based on all features, such as start of the trip, time, distance, mode of transport used, and combination with other purposes. This is the third research gap I address in this dissertation. Several studies have established that "humans follow simple reproducible patterns" (Gonzalez et al., 2008) and that their movements are particularly defined by time and space (Eagle and Pentland, 2009). People spend most of their time in a few locations and the rest across a broad variety of places—typically ranging from 5 to 50 places (Gonzalez et al., 2008; Jiang et al., 2012). Song et al. (2010) detect 93% predictability in the mobility patterns of mobile phone users. However, all the aforementioned studies use aggregated, anonymized mobile phone data and lack relevant information about individual trips. Because these telecommunication data do not provide the level of detail needed for individual marketing analyses, we should enable researchers to focus on switching to GPS tracking with clearer origin and destination information, routing, modal choice, and ideally even individual demographic information. Due to data privacy regulations, such data are difficult to collect, but they are necessary for meaningful behavioral analyses. When neglecting this dynamic perspective of LBM whereby we study how an individual gets to the place where they pursue an activity and how frequently they return to this location, many important behavioral insights are lost. By ignoring these additional dimensions of *where from* and *by which mode of transport*, marketers run the risk of missing opportunities to acquire new customers and retain existing ones, making inaccurate assumptions about their customer base, which leads to unfavorable strategic decisions for their businesses. CBM is the natural extension of LBM that aims to include more information and helps overcoming the described mismatches.

Dissertation overview

Based on the automated literature review I presented in this introduction, I asked *where the LBM field is heading and how CBM can shape the future of hyper-targeting*. To fill in the research gap discussed above, it is important that CBM will broaden its horizons by embracing multidisciplinary research. Overall, we can summarize the three mismatches as a gap between the consumer and the environment. On the one hand, the field of urban planning uses spatial data to study cities on an aggregate level and derives policy implications based on such studies. On the other hand, the field of consumer behavior uses surveys and experiments to evaluate individuals' behavior and decision-making. An explicit link between the two fields is missing; therefore, by using human movement data in South Limburg, this dissertation aims to provide this link between consumers and the environments in which they move.

This dissertation contains six distinct chapters in which I provide some evidence that proxemic technologies and the data they generate are imperative to LBM and CBM research and that this data gives marketers the potential to engage in curated marketing via hyper-targeting. Each of the six chapters addresses one of the topic clusters established earlier, except for the technology cluster; I decided not to discuss this topic because it is the most dominant one in the network and most of the computer science/engineering-based literature already boosts this topic. Even though each chapter focuses on one topic cluster, we cannot look at the topics in isolation. As the network visualization has shown, the topics are intertwined; thus, it would be an information loss to solely focus on one topic at a time while ignoring other topics that potentially influence scenarios or research problems. The advancement of LBM and CBM research is only possible when we examine these topics simultaneously. Each chapter provides an example of a research approach that benefits hyper-targeting. Additionally, I want to use this setup to demonstrate that small steps toward multidisciplinary research can deepen our understanding and result in a more holistic picture of the research problem.

All six chapters are structured similarly; they first introduce the phenomenon and the background of the topic cluster, followed by a description of the research method, which allows for the incorporation of the geographic or dynamic perspective. Next, I illustrate the results achieved through adding spatial data, and then I draw a conclusion with respect to the phenomenon and more broadly elaborate on the impact of the research results for the future of hyper-targeting. Between chapters, I enhance the narrative with transition paragraphs that illustrate the line of thought leading to the next chapter. Directly below, I briefly introduce each chapter.

Chapter 1. According to Czarny (2018), usage of location data is "one of the most misunderstood areas in marketing." Until now, LBM has focused on the positioning of a consumer at a certain point in time and used this static location information to place targeted advertising on social media platforms or in a company's app. However, location data are much more than just static location information; they include information on the infrastructure of a city, the buildings and their function, and the consumer's dynamic positioning in the form of human movement data. In this chapter, I list and briefly introduce a selection of methods suitable for spatial data; subsequently, in the following chapters, I demonstrate these methods in practice. Further, Chapter 1 introduces the datasets used for the analysis in the other five chapters. These datasets are good examples of a spatial datasets for marketing research purposes; further I link human movement data from a diverse sample of people to environment data for the entire observation area. The main takeaway from this chapter is that current LBM literature is too limited in scope and that adding dynamic data is both feasible and desirable.

Chapter 2. Building interconnectivity analysis is a useful tool for gaining a clearer picture of a city's structure. The composition of the built environment allows us to view the city as a servicescape, which can help to improve the customer experience in cities. In this chapter, we use the urban network analysis toolbox, developed at the Massachusetts Institute of Technology (MIT), to run a straightness analysis on all buildings in the region and determine which buildings are reachable without having to pass through many intersections and junctions. The goal of this analysis is to sketch the city and prepare for a future with urban air mobility (UAM). Based on this analysis, which is enriched with data on public transport infrastructure and residents' purchasing power, we can provide suggestions on where in the region UAM vertiports should be placed to ensure that they are well embedded in the current public transport services and serve potential customers who can afford such an exclusive service.

Chapter 3. Adding to the previous chapter, we continue to sketch the city, including variables beyond the positioning of buildings. We use supplementary information about the built environment to model the popularity of locations and consider diverse factors. We propose that building function, competition, density of the built environment, and public transport availability predominantly impact the popularity of a location, which is measured in frequency of visits. We run a regression model to test our hypotheses and find several significant direct effects and interaction effects. In particular, we find that density, competition, and cell function positively impact location attractiveness and we find that there is an interaction between public transport availability and cell function as well as between the cell function of a trip's origin and the cell function of a trip's destination. Moreover, in this chapter, we shed light on research from the early 2000s that emphasized the use of spatial econometrics in marketing but diminished after the introduction of ubiquitous consumer devices.

Chapter 4. Individuals with similar socioeconomic and demographic characteristics and those closely connected in an (offline) social network tend to behave similarly. More specifically, they tend to visit the same locations for the same reasons at roughly the same times. However, uncovering such geo-social networks is easier said than done. Therefore, in this chapter, we use data visualization to uncover hidden geo-social networks in our dataset. Data visualization is a powerful approach that can complement traditional computational research approaches and bring previously undetected relationships and interdependencies to the surface. Graphs and visualizations are easier for the human brain to process; therefore, researchers frequently uses data visualization when presenting results to an audience of industry professionals. Through this research, we can gain a deeper understanding of the network structure underlying the movement data, and we can visualize multiple layers of information within one graph using colors and glyphs.

Chapter 5. During the analyses of the previous research problems, it became clear that the insights would be more meaningful if more information were available in the human movement dataset. Specifically, it would be valuable to know which movements in the dataset correspond to the same individual as well as the purpose of those movements (i.e., which activity the individual pursues at their destination). However, due to privacy-preserving regulations in Europe, this information is not available in our data; thus, we use feature engineering and clustering to reverse engineer the trip purpose. The method proves to be successful: we identify 12 distinct clusters, which we present in a heat map table. We put the 12 clusters into four broader categories so that public and private stakeholders can integrate these findings into their future policy and business decisions.

Chapter 6. Each chapter discusses one piece of the puzzle, leading to a complete picture of consumers' behavior in cities. Because mobility decisions are complex, this chapter incorporates transport mode into activity-based travel behavior by hypothesizing that people's mode choices depend on their activities at trip destinations. We conduct a complementary survey among the participants of the movement data collection and observe the respondents' transport-mode choices for specific activities. Based on these data, in combination with additional citizen information, a number of clusters with similar mode-activity patterns emerge. We note distinct differences in transport mode choice between clusters for various activities. This finding demonstrates the relevance of examining mode choice and activities simultaneously; only with this approach are we able to capture a holistic picture of human behavior and decision-making in cities. In the end, we draw implications from this study for businesses and sustainable transport policymakers.

Table 0.1 presents an overview of all dissertation chapters and Figure 0.4 presents the dissertation's structure and logic in a framework.

Chapter	**Marketing contribution**	**Lens**	**Data**	**Method**
1	Call for integration of geographic data into marketing	Proposition to broaden LBM's perspective towards dynamic movements and physical context	Non-empirical essay-style chapter that introduces all datasets of the dissertation	—
2	Introducing a new city service (UAM)	Cities as servicescapes	OD trajectories, built environment from BAG, purchasing power from CBS, public transport availability from OSM	KDE, UNA toolbox straightness, Anselm Moran I, multi-layer visualization
3	Drivers of location popularity	(Static) Urban infrastructure	OD trajectories, built environment from BAG, public transport availability from OSM	OLS regression
4	Using data visualization to uncover complex relationships	Urban dynamics	OD trajectories, built environment from BAG	Data visualization (i.e., flow analysis, diurnal graphs, hot spot analysis, classification)
5	Trip purpose discovery	Privacy-preserving data-driven marketing	OD trajectories, built environment from BAG	DBCSAN clustering
6	Understanding current mobility behavior to trigger changes	Activity-based mode of transport choice	Complementary survey	k-means clustering

Table 0.1: Overview dissertation chapters

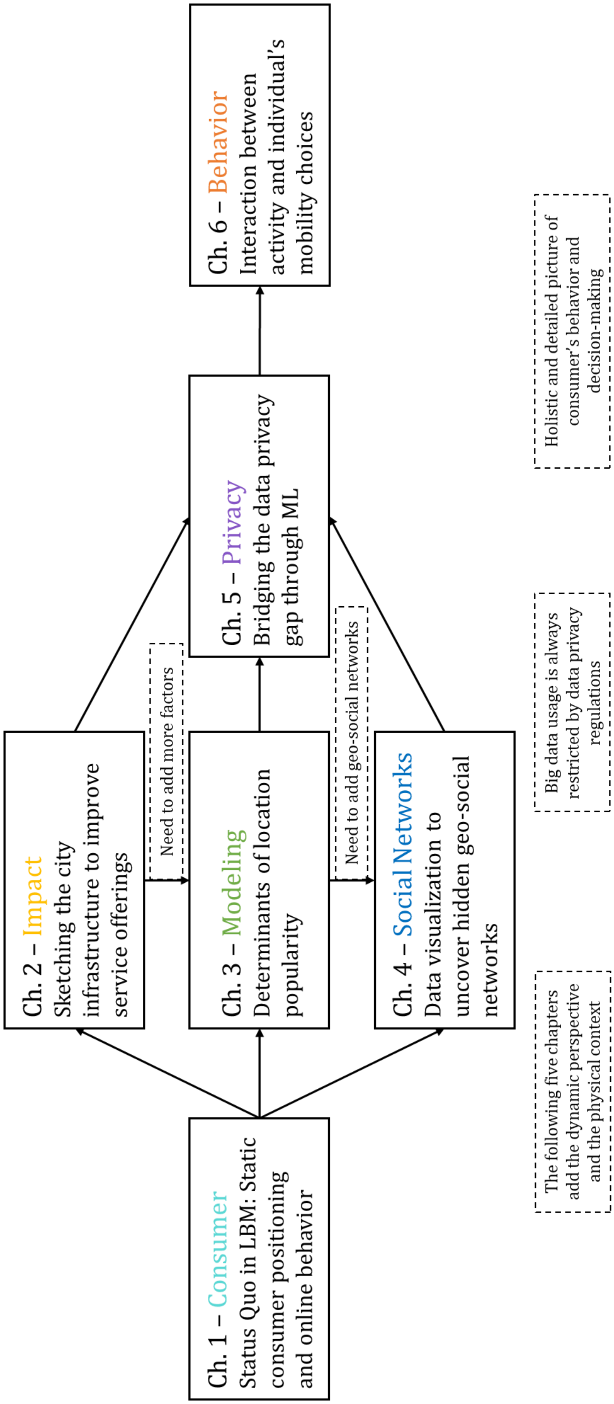
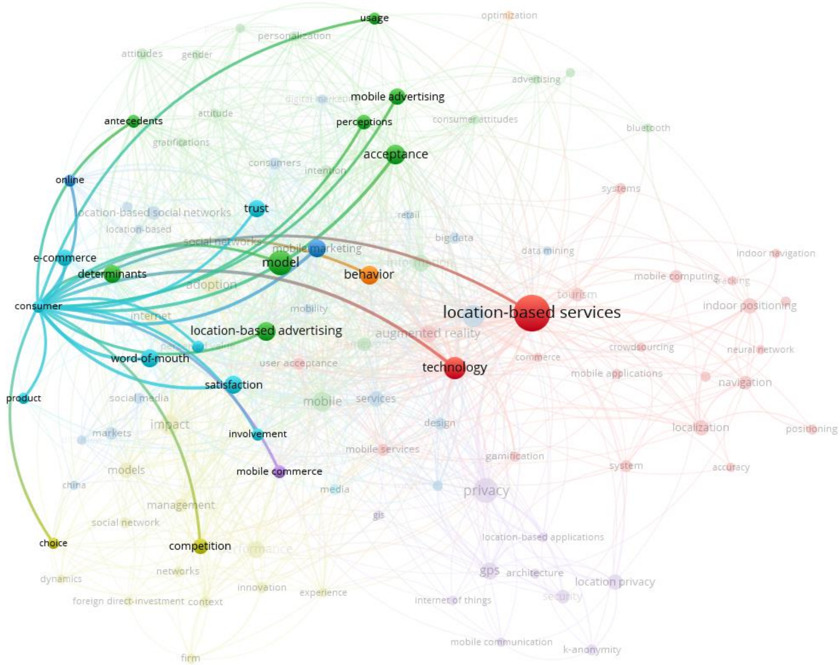


Figure 0.4: Dissertation framework



Chapter 1

Location-Based Marketing as an Iceberg—Jumping into Unknown Waters to Look Beyond the Surface

1.1 Introduction

Have you ever walked down a street in the city center and, when you passed by your favorite fashion retailer, received a push notification about special offers and a personalized coupon from the retailer's app? Have you ever enjoyed the sun in the park and, while scrolling through social media platforms on your phone, received an advertisement for your favorite coffee—and the corresponding coffee shop is just around the corner from the park in which you are sitting? Situations like these are perfect examples of companies' current LBM efforts. Via company-owned apps and social media behavior, businesses collect consumer preference information and, combining this with geo-fencing, place ad hoc advertisements and mobile marketing. Such placements are a core part of LBM; they can be effective in prompting unplanned purchases. However, these advertisements form only the tip of the iceberg; LBM entails a lot more potential when we look beyond the surface. Two aspects of the abovementioned scenarios limit LBM to its current scope: (1) LBM is static and only considers the consumer's current position while ignoring the dynamic consumer movements in the city; (2) LBM applies a purely consumer-focused perspective and works mainly with the online shopping behavior of the target audience. Thus, the field neglects the physical context in which the consumer lives. We can broaden the current scope of LBM and move towards CBM by placing it within the larger field of geography in marketing decision-making. This integration of geographic information requires a jump into unknown waters, but the benefit will be extensive.

The building composition of a neighborhood is good to know, especially for certain businesses (e.g., shops, restaurants, banks, doctors' offices, entertainment facilities, and transport service providers such as shared mobility and last mile delivery). For a restaurant, for example, it is important not only to focus on the quality of the food and use LBM for advertising and special offers via social media coupons but also to understand the geographic area around the building. A restaurant located in a business district should focus on quick lunches and fancy business dinners, establishing deals with companies in the surrounding neighborhood—such as special offers or delivery services to the office—and striving to retain existing customers who need to grab lunch close to their office every day. Conversely, a restaurant next to a theatre in the city center of a larger town probably attracts a lot of theatre visitors or tourists. They can be targeted up front with special offers in combination with a theatre ticket or via advertising printed on the parking tickets of the closest parking garage.

Thus far, the computer science and urban planning literature has noted the importance of the surrounding area and trajectory analysis to business success (D'Silva et al., 2018; Gonzalez et al., 2008; Song et al., 2010). Further, transport journals have mainly been interested in research with geographic information and they usually focuses on policy and sustainability implications. Recent marketing literature has hardly considered spatial data and its usefulness and importance in marketing decision-making (Bradlow et al., 2005; Bronnenberg, 2005; Elhorst, 2017); the literature covers areas such as survey-based behavioral studies in geographically bounded retail, location-based services, and mobile marketing and advertising (Bues et al., 2017; Ketelaar et al., 2017; Ketelaar et al., 2018). However, this view on LBM fails to illuminate the impact that the surrounding area has on consumer behavior, on the attractiveness of the focal location, and consequently, on the success of a business. This is problematic because we study humans, who are extraordinarily complex; they make decisions about when and where to shop or attend an activity by weighing many factors. Therefore, we must broaden the LBM horizon and adopt core concepts, methods, and data sources that are part of geography in marketing decision-making (e.g., the attractiveness of the geographic area where a consumer is). This broadening can include spatial modeling of decisions and behavior, location decision modeling, hotspot detection, etc. Although this angle is relevant to marketing, it is currently only addressed in other domains and contexts.

This chapter aims to kick-start the use of spatial data for marketing decision-making, a new LBM perspective that will deepen insights into urban marketing and behavior in cities. Research about infrastructure and planning ideally builds on static information (i.e., data about the built environment, such as buildings and physical [mobility] infrastructure) and uses dynamic information (i.e., movements of people tracked via GPS) to verify that these locations are indeed frequently visited and not only popular based on the environment. Conversely, research can use movement data as a starting point to observe behavior and consider infrastructure as influential variables.

The remainder of this chapter is structured as follows: first, I reflect on the current literature on geography in marketing, particularly LBM and related components; second, I introduce a methodology suitable for working with spatial data for marketing purposes; third, I introduce spatial data that allow researchers to answer marketing research questions; finally, I propose an agenda for new research directions to broaden the scope of LBM.

1.2 Background and research question

According to Bernritter et al. (2021), marketers struggle to find suitable ways to utilize geographic data in marketing. At present, when we talk about geography in marketing, we usually associate it with LBM. However, the LBM field does not consider traditional spatial data, especially not in a larger geographical context such as a city. Instead, it investigates the effectiveness of marketing efforts, such as ad hoc price promotions and personalized coupons, or store awareness/attention campaigns without financial incentives. If spatial data are used for analysis, they are usually GPS movements on a small scale (e.g., within a shopping mall; (Ghose, Li, et al., 2019) or GPS data via Foursquare (Noulas et al., 2012). Even though LBM covers only part of *geography in marketing decision-making*, it is still worth having a closer look. Because LBM and CBM belong to the social science domain, the consumer as the focal point in the research, and most studies investigate how to use knowledge of an individual's whereabouts to influence their shopping behavior and increase store profits. While LBM used to be a combination of out-of-home marketing and digital marketing, recent trends have shown that the former has decreased and that many companies now focus solely on the latter. Consequently, LBM research mainly focuses on location-based advertising and mobile marketing. In this study, we divide consumer-centered LBM into in-store vs. out-of-store and into mobile vs. non-mobile.

1.2.1 Status quo: Location-based marketing

In-store LBM examines how the consumer perceives non-mobile and mobile marketing messages presented inside a store as well as their effects on sales, coupon redemption rates, and consumer retention rates (Bues et al., 2017; Ketelaar et al., 2018). However, Bernritter et al. (2021) question whether in-store mobile LBM is a profitable investment, noting that these consumers are already in the store intending to purchase a product and that the beacon technology required for this type marketing is expensive. Counterintuitive to their proposition, they find that in-store is often more effective than out-of-store mobile marketing, potentially because customers are drawn to products other than those they initially intended to purchase (Bernritter et al., 2021).

A non-mobile in-store advertisement, also known as a point-of-sale advertisement, is displayed on the shelf where the product is placed. Therefore, it is unlikely that the customer will overlook the advertisement. Further, researchers have found that mobile and non-mobile advertisements inside a store are equally effective and that attention toward an advertisement does not depend on the medium of exposure (Ketelaar et al., 2017). Mobile in-store marketing typically uses beacons in combination with a company's smartphone app, sending pop-up advertisements to smartphones (Bues et al., 2017). Thus, it is limited to larger firms that have their own apps and are dependent on customers downloading their apps or using store Wi-Fi. Mobile in-store marketing uses sound and vibration to get the customer's attention when a message is sent (Ketelaar et al., 2017). Approximately 14% of consumers already use mobile retailer apps, and 63% would be willing to use them if their stores had one available; moreover, 43% of the companies with such apps are still dissatisfied with their mobile in-store marketing effectiveness asking for improvement suggestions (Bues et al., 2017). The most promising features of mobile in-store marketing campaigns include the close spatial proximity of the consumer to the product, inspiration, and perceived value (van de Sanden et al., 2019) as well as the clear relevance of the advertisements (van't Riet et al., 2016).

Non-mobile out-of-store LBM has been around for a long time; it includes the road signs and large posters on the sides of highways, which often indicate the fast-food places available at the next exit, as well as the signs and flyers on a main street or in a marketplace, which might draw attention to a store—and its daily promotions—in an alley around the corner. Mobile out-of-store LBM examines how smartphones can help attract consumers to a store near their current location (Dubé et al., 2017; Fang et al., 2015). These campaigns usually work with geo-fencing; they either strategically place advertisements, promotions, and coupons in the consumer's social media and public transport apps or they send coupons via a company's own app if installed on the consumer's phone. Similar to in-store LBM, research has found that spatial proximity has a positive effect on store visits, purchases, and coupon redemption (Danaher et al., 2015; Molitor et al., 2020).

A great deal of information has been gathered about consumers' perceptions of mobile LBM and its effectiveness (Bruner & Kumar, 2007). Companies must walk a thin line to attain good LBM via smartphone; they have to find the right balance between personalization of offers and preservation of consumer privacy in order to create information and value instead of irritation and discomfort (Aguirre et al., 2015; Xu et al., 2009). LBM is said to have longer dynamic effects than behavior-based mobile promotion while also encouraging future sales (Luo et al., 2013). Further, consumers are more receptive to mobile LBM on sunny days than on rainy days (Li et al., 2017), and they are more receptive to it in crowded subways than in empty subways (Andrews et al., 2016; Ghose, Kwon, et al., 2019), and they are more receptive to high-involvement products and utilitarian products (Bart et al., 2014).

Ghose, Li, et al. (2019) place LBM and the consumer in a broader context, acknowledging that LBM should take spatial data into account. More precisely, they argue that the location of a marketing placement, timing of a marketing placement, context of the activity (i.e., purpose, time spent on the activity, and connection to other activities), and movement speed need to be considered to understand individual decisions. Knowing the where, when, and how behind a consumer's decisions and consequently the why behind their current location allows for more targeted LBM efforts. Therefore, it is crucial to understand how consumers make trip decisions.

1.2.2 Consumers' trip decision-making

Dellaert et al. (2008) argue that consumers, when making trip decisions, first develop simplified mental representations of their trips according to their previous experiences (Dellaert et al., 2014; Dellaert et al., 2008). When people become familiar with their surrounding areas, they include habits in their decision-making processes (De Ceunynck et al., 2011). In a second step, they incorporate attributes, benefits, and context-dependent features (Arentze et al., 2014): attributes (instrumental factors) include speed, travel time, distance, type of store, ease of parking, and flexibility; benefits include convenience, efficiency, freedom, and relaxation; and context-dependent features include weather and location. Further, "psychological time," an individual's own perception of time spent, can influence their emotions, their cognition, and consequently, their decision-making (van Rijn, 2014). Moreover, context significantly affects attributes and benefits, such that people's final decisions on modes of transport and locations combine context-dependent features of the available alternatives and personal needs (Dellaert et al., 2014; Innes et al., 1990).

Extant research has identified the four main drivers of transport mode choice: socioeconomics and demographics, instrumental factors, geographic factors, and trip purpose (Carrasco et al., 2008). Socioeconomic and demographic attributes, which are highly relevant, include age, gender, occupation, household size, income, and car ownership (De Jong et al., 2004; Garikapati et al., 2016; Hunecke et al., 2010; Kurniawan et al., 2018). Instrumental factors include values such as safety, health and well-being, the meaning of emerging transport technologies, and the meaning of owned versus shared vehicles (Kurniawan et al., 2018). Geographic factors in mode choice equally involve trip origin and trip destination, with rural locations possibly limiting options (Hunecke et al., 2010; Manaugh et al., 2010; Noulas et al., 2012). Finally, trip purpose incorporates factors such as commuting behavior, job availability in urban areas, flexible leisure activities, and the interplay between trip purpose and resource accessibility (Manaugh et al., 2010).

Trip timing is another widely studied component of trip decision-making. Initially, researchers studied commuter trip timing; interest in non-work trip timing arose much later (Bhat & Steed, 2002). The crucial difference between the two trip types is the temporal flexibility of non-work trips. Consumers connect leisure activities to work schedules as well as family and household schedules; they shop at the most convenient times of the day or week, and their trips often feature chains of multiple stops (Bhat &

Steed, 2002). Travel cost plays only a minor role in leisure trip decisions, and few individuals consider the environmental aspects of different modes of transport (De Ceunynck et al., 2011). In addition, researchers have found similarities in preferences among people with identical smartphone usage behavior. More precisely, individuals using the same app at roughly the same time and place, tend to share preferences (Molitor et al., 2012; Zubcsek et al., 2016). Hence, it is critical to understand the interplay between time and location to effectively make use of the colocation effect in LBM.

Ghose, Kwon, et al. (2019) claim that commuters are three times more receptive than non-commuters to LBM efforts on public transport apps. They show that consumers' contexts, as well as their movements and destinations, affect their responsiveness to advertising and can help nudge them toward changing their shopping behavior in the future (Ghose, Li, et al., 2019; Zhang & Katona, 2012). Assuming that such relationships exist for other modes of transport and marketing placements, studying the travel mode, travel timing, and potential areas to pursue activities is important for a holistic LBM strategy. My study builds on the work of Ghose, Li, et al. (2019) and focuses on LBM more broadly. Previous LBM research has studied the current locations of consumers and has used this information to place advertisements. I suggest additional research on the movement that gets consumers to a location, on the favorable infrastructure features of an area, and on ways to use spatial data in an LBM strategy. We can derive our initial insights into location attractiveness from the econometrics literature on facility location selection.

1.2.3 Potential of using store location in marketing

Researchers have investigated store location selection elaborately; studies of this sort usually consider, among other factors, the geographic surrounding area of a potential store location. Most of this research is quantitative and tests various econometric approaches (e.g., Canel et al. (2001); Klose (1999); Melkote and Daskin (2001) as well as machine learning approaches (e.g., Kahraman et al. (2003); Karamshuk et al. (2013) to solve the facility location problem. However, once a store opens, these ideal location features are not used further for purposes such as marketing; marketers are not econometricians or geographic information systems (GIS) experts, and academic research is often not multidisciplinary enough to identify these strong connections and the potential for new insights.

From the facility location literature, we can still derive four attributes that make a location attractive to businesses. First, depending on whether a business is market- or resource-oriented, a suitable location provides access to either markets or skilled labor respectively (Andersson et al., 2014; Larsson, 2014). Second, accessibility to public transportation within walking distance and the presence of other transport infrastructure features, such as vehicle sharing (stations) or bike paths, tend to positively influence the visitation rate to a location (Mejia-Dorantes et al., 2012). Third, the physical infrastructure of the area, distance to the city center, popularity of the area, and socioeconomic structure of the nearby neighborhoods collectively play an important role in the suitability of an area for a specific business (Dong et al., 2018; Karamshuk et al., 2013). The last aspect is especially important; there must be a good match between the business category and the demand in the immediate surroundings, particularly for small businesses (Larsson & Öner, 2014). Finally, competition and co-location are crucial to the ideal placement of a business and, therefore, must be considered. While stores with profound interaction (e.g., clothing stores and shoe stores, doctors' offices and pharmacies) should co-locate, those selling fast-moving consumer goods and perishables (e.g., supermarkets, bakeries, and butcher shops) or durables (e.g., hardware stores) should spread out (Larsson & Öner, 2014). Using these attractiveness attributes beyond location selection and considering them in the (location-based) marketing strategy of stores can increase effectiveness and thus increase revenue. Business owners can estimate long-term success probability by examining where their store can attract a certain number of clients, their nearby rivals, and their complementary business partners.

1.2.4 Research gap

Until now, only LBM, which covers the effects of mobile advertising on purchases, has addressed geography in marketing. This is the consumer-based perspective typical in marketing. LBM needs to be broadened with the addition of movement and spatial data. Therefore, this chapter aims to show how researchers can use spatial information to enrich marketing strategy. We list a variety of methods and tools for different analysis options and research goals, while introducing a dataset that fits these methods. This dataset will then be used in the remaining five chapters to address different aspects of LBM in urban areas.

1.3 Research methods for spatial data in location-based marketing

Traditionally, movement data were collected proactively using paper-and-pencil interviews; followed by telephone-aided interviews and computer-aided self-interviews (Chen et al., 2019; Nguyen et al., 2020; Zhao et al., 2020). However, information from these sources is regarded to be of poor quality (Chen et al., 2019; Nguyen et al., 2020). The self-administered nature implies that the data are subject to inaccuracies and missing information caused by the under- and over reporting of relevant trip information (Nguyen et al., 2020). Moreover, the data usually covers only a relatively short duration and small geographic space due to the high costs and heavy respondent burden (Chen et al., 2019). Other limitations include the high non-response rate and long gaps between periodic travel surveys (Nguyen et al., 2020).

Technological advancements in information and communication technology, driven by the development and wide adoption of GPS, have revolutionized the collection methods and focus of human mobility surveys (Barbosa et al., 2018; Gong et al., 2016; Gong et al., 2014; Nguyen et al., 2020; Shen & Stopher, 2014). Nowadays, people use various devices—on a daily basis—that collect data about their movements, such as smartphones, smartwatches, or fitness trackers. This real-time information about individuals' actual behavior is the basis for urban analytics. As mentioned above, spatial analysis allows for more diversity in the field of marketing and adds a new perspective on behavior and decision-making in cities. Ultimately, there is no one-size-fits-all solution for integrating geographic information into marketing research. The most appropriate method is dependent on the available data and its format as well as on the specific research question and goal. Therefore, this section briefly introduces some methods that are applied to spatial data in the following chapters.

In general, spatial data are best processed in GIS software solutions, such as ArcGIS. However, it is also possible to load and analyze them in less specialized programs, such as R. Depending on the nature of the information, spatial data are usually stored in csv files and shapefiles. Shapefiles contain geospatial vector data such as points, lines, or polygons with geographic coordinates. Thus, they can be viewed on a map in GIS software, but they can also be read and visualized in programs such as R. ArcGIS provides numerous tools for geoprocessing and calculations with geospatial data layers. One of the most popular tools is the kernel density estimation (KDE), which estimates the probability density function based on a random variable. KDE calculates the distance between a point on a map and its reference point, sums the calculations together for all surfaces for that particular reference location, and iterates this step for all points until a so-called kernel is placed over each observation. Adding all the kernels together gives a density estimation for that area and, ultimately, a smooth and continuous inference for the region based on the sample (Anderson, 2009).

Another popular research method for the spatial analysis of movement data is the Hot Spot Analysis Getis-Ord Gi*. In this method, an algorithm uses Getis-Ord Gi* statistics to identify significant hot and cold spots on a map, based on a (weighted) feature set. When the aim is to get a visual overview of the city, input feature sets can be the movement trajectories or building densities of certain categories, as in Chapter 4 of this dissertation. Hotspot analysis can also be applied to infection rates (to determine the epicenters of a spreading virus), addresses of crime scenes (to determine safe and unsafe neighborhoods), and accident information (to identify unsafe intersections that require interventions to improve road safety).

Next, origin–destination (OD) flow analysis, also referred to as human trajectory analysis, is a core tool in urban dynamics. The goal of this method is to discover urban activity patterns based on the OD dyads or exact routing information of the population. Based on historical or real-time data, researchers can measure phenomena such as service levels in a city or traffic flow in an area. By enriching the raw trajectories with additional attributes of the movement, the algorithm aims to infer activity patterns and relationships between urban hotspots.

A few years ago, researchers from the MIT Senseable City Lab developed a toolbox that can be loaded as an external extension to the ArcGIS standard toolboxes. This urban network analysis (UNA) toolbox computes graph analysis measures of spatial networks based on map information and the physical infrastructure features of a city. The centrality tool of the toolbox focuses on the buildings, calculating reach, gravity, betweenness, straightness, and closeness. Complementarily, the redundancy tool uses street and junction data, calculating a redundancy index, redundant paths, or a wayfinding index for the city (Sevtsuk & Mekonnen, 2012). Chapter 2 highlights the straightness tool.

Aside from these specific GIS tools, spatial data—especially OD trajectories—can be analyzed using commonly known machine learning techniques. This usually requires thorough data pre-processing and feature engineering but often leads to meaningful insights. In Chapter 5, we assess using the popular k-means clustering algorithm in R to reverse engineer trip purpose based on spatial data. In Chapter 6, we apply the k-means algorithm to transport mode choice patterns to distinguish behavioral groups in our sample. This algorithm is an unsupervised technique that partitions observations into clusters, with minimum variance within one cluster but maximum variance between clusters.

Finally, researchers do not necessarily have to use machine learning to handle large datasets; it is also possible to run a regression on human trajectories. The modeling of spatial data for marketing caught the interest of researchers in the early 2000s, but this interest did not last. Chapter 3 brings it back, using spatial data of human movements and city infrastructure in simple regression models.

1.4 Description of spatial datasets applicable in location-based marketing

The data used in the following five chapters exemplify spatial data with marketing applicability. We work with an OD tracking dataset of movements within South Limburg—a region with a surface area of approximately 660 km^2 that is located in the southern part of the Dutch province of Limburg—and we add a built environment dataset retrieved from the Basisadministratie Gebouw (BAG) for more information about the buildings in the region. Additional data include census data from the Dutch Centraal Bureau voor Statistiek (CBS) and transport infrastructure data for the region from OpenStreetMap (OSM). In 2018, South Limburg had a population of 599,025 inhabitants and an average population density of 921 people per km^2, making it one of the most densely populated regions in the Netherlands (CBS, 2021). At that time, the region comprised 18 administrative regions and the municipality of Maastricht, which is the largest district and contains the province's capital. A topographic map of the area reveals that most of the urbanized, industrial, and commercial zones are located in the western and northeastern parts whereas the more rural areas and smaller districts are situated in the center and the south. Furthermore, the area maintains good, well-developed infrastructural elements such as walking and bicycle lanes, recreational routes, roads, highways, a railway system, and an airport. The average commuting distance of people living in South Limburg is 30.6 km, and 70% of the working population works outside their residential municipality (ESZL, 2018).

1.4.1 Origin–destination tracking dataset

The OD tracking dataset originates from a comprehensive data collection on human mobility, which I executed in cooperation with the InnoZ research institute in Berlin, using their Modalyzer smartphone app. This app was particularly designed for mobility behavior research on individuals that conforms to the General Data Protection Regulation (GDPR). To participate in the data collection, people had to be live or work in Maastricht or frequently commute to the city for any purpose. To ensure a diverse sample, people were recruited via the university's social media channel, internal course pages for students, and the researchers collaborated with Maastricht Bereikbaar and through them contacted a panel of citizens who frequently participate in surveys aimed at improving urban mobility in the region. Participants were asked to sign up in the app and to activate the tracking function themselves via a slide button. Once activated, the recorded movements were only visible to the app user and they had to indicate in the projects tab of the app that the collected data is a contribution to a research project. The app's core algorithm automatically recorded the trips of each participant using a combination of GPS, GSM, smartphone motion sensors, maps, and public transport schedule information. Moreover, the algorithm automatically detected nine different modes of transport, and the participants could manually select six additional modes. To ensure excellent data quality, the participants had to review, correct, and confirm the logged trips inside the app. Only after the participant had verified them were the trips included in the data collection.

The resulting raw tracking dataset consists of 6,355 trips by 98 participants between November 18 and December 10, 2018. Because not all participants tracked their movements for the full three weeks, the raw dataset comprises 1,032 days and 53,231 km tracked collectively by all the participants—a daily average of 147 trips and 2,314 km. Each observation consists of six variables related to the trip: mode of transport, length in kilometers, starting date and time, finishing date and time, origin, and destination. For privacy reasons, the specific GPS coordinates of the origin and destination are anonymized by reducing the spatial granularity to a 300-meter hexagonal cell in a grid that encompasses the entire region. A shapefile linking the origin and destination cells' ID numbers to the polygon-based geometries of the hexagonal grid completes this data.

We know some basic demographics about the sample and can confirm the balanced diversity of the sample. Specifically, 43.9% of the 98 participants are female, and 56.1% are male. The participants are 18–25 (9.2%), 26–40 (24.5%), 41–59 (52%), and 60+ (14.3%) years old. Ninety-five participants possess a driver's license, ensuring full flexibility in transport mode choice, and the home addresses of all participants are evenly distributed across all neighborhoods within Maastricht as well as the smaller towns in the region. Finally, only five travel modes remain after pre-processing the data. Driving (36.50%), walking (35.31%), and cycling (21.72%) are the most dominant modes, while traveling by public transport, such as by train (4.61%) and bus (1.86%), are less common. Moreover, most trips are less than 5 km and completed within 10 minutes.

1.4.2 Built environment dataset

The built environment dataset contains information about South Limburg's physical infrastructure composition. These data are available on two levels of aggregation, individual buildings and grid cells. One dataset contains all 339,233 buildings in the region. Each observation in the dataset represents one building and includes variables on the latitude and longitude of the building, the surface of the building in m^2, and the main function of the building as text and as a binary in 11 columns of possible functions. The categories are residential, social gathering/restaurant, telecommunication, health, industry, office, hotel/accommodation, education, sport, shops, and other. The other dataset has all 3,256 grid cells covering the area and corresponding to the start and end points of the trips. In this dataset, moreover, each observation consists of the unique cell ID and 11 variables—one for each building function category that counts the number of buildings within the borders of the respective cells that have the particular function category (Table 1.1). These two datasets also come with matching shapefiles to visualize and process the observations and all their attributes in GIS software.

N	Variable	Type	Mean	Mean (or "Example")
1	Cell ID	Numeric	Number corresponding to the grid	"249432"
2	Residential use	Numeric	From 0 to 1,653 buildings	91.6 buildings
3	Social gathering/restaurant	Numeric	From 0 to 83 buildings	1.1 buildings
4	Telecommunication	Numeric	From 0 to 1 buildings	0.001 buildings
5	Health	Numeric	From 0 to 85 buildings	0.5 buildings
6	Industry	Numeric	From 0 to 70 buildings	1.7 buildings
7	Office	Numeric	From 0 to 134 buildings	1.4 buildings
8	Hotel/accommodation	Numeric	From 0 to 200 buildings	0.8 buildings
9	Education	Numeric	From 0 to 9 buildings	0.2 buildings
10	Sport	Numeric	From 0 to 5 buildings	0.1 buildings
11	Shops	Numeric	From 0 to 317 buildings	2.2 buildings
12	Other	Numeric	From 0 to 250 buildings	6.2 buildings

Table 1.1: Description of variables in the built environment dataset

1.4.3 Data pre-processing and feature engineering

The raw OD tracking dataset contains 6,355 trip observations and six variables, while the raw built environment dataset includes 3,256 cell observations and 12 variables. Some of the analyses in the following chapters focus on one of the two datasets and use the other for validation, while other analyses benefit from the combination of the two and feature engineering made possible this way. The combined cleaned dataset consists of 4,296 trip observations and 39 variables. Trips are excluded for multiple reasons, such as the origin or destination being outside the observation area, modes of transport being unavailable in the area (i.e., boat or airplane), and measurement errors by the app, such as unrealistically low or high travel speed or very short travel distances (i.e., less than 0.1 km). Table 1.2 provides an overview of all the features in the combined dataset.

N	Variable	Type	Range	Mean (or "Example")
1	Mode	Nominal	Five travel mode categories	"Bicycle," "Bus," "Car," "Train," and "Walk"
2	Started at	Datetime	2018-11-19 07:50:33 to 2018-12-10 08:31:34	"2018-11-22 10:25:18"
3	Started at day of week	Nominal	Seven weekday categories	"Monday" until "Sunday"
4	Started on the weekend?	Binary	1 = True, 0 = False	"1" or "0"
5	Started at hour	Numeric	0 to 23 hours	14 hours
6	Started at daypart	Nominal	Four daypart categories	"Morning peak," "Day off-peak," "Evening peak," "Evening/night off-peak"
7	Finished at	Datetime	2018-11-19 08:02:16 to 2018-12-10 08:44:05	"2018-11-22 11:09:56"
8	Finished at day of week	Nominal	Seven weekday categories	"Monday" until "Sunday"
9	Finished on the weekend?	Binary	1 = True, 0 = False	"1" or "0"
10	Finished at hour	Numeric	0 to 23 hours	14 hours
11	Finished at daypart	Nominal	Four daypart categories	"Morning peak," "Day off-peak," "Evening peak," "Evening/night off-peak"
12	Length	Numeric	From 0.08 to 72.48 km	5.15 km
13	Duration	Numeric	From 2.00 to 88.05 min	11.71 min
14	Speed	Numeric	From 2.01 to 125.25 km/h	22.19 km/h
15	First	Numeric	Number corresponding to the 200m grid	"229320"
16	Last	Numeric	Number corresponding to the 200m grid	"242299"
17	Same cell?	Binary	1 = True, 0 = False	"1" or "0"
18	Total buildings first	Numeric	From 0 to 1,934 buildings	612 buildings
19	Use residential first	Numeric	From 0 to 100%	80.63%
20	Use social gathering/restaurant first	Numeric	From 0 to 100%	2.06%
21	Use health first	Numeric	From 0 to 44.3%	0.74%
22	Use industry first	Numeric	From 0 to 100%	2.53%

23	Use office first	Numeric	From 0 to 100%	4.32%
24	Use hotel/accommodation first	Numeric	From 0 to 96.6%	0.31%
25	Use education first	Numeric	From 0 to 33.3%	0.94%
26	Use sport first	Numeric	From 0 to 100%	0.29%
27	Use shop first	Numeric	From 0 to 100%	3.57%
28	Use other first	Numeric	From 0 to 100%	3.99%
29	Total buildings last	Numeric	From 0 to 1,934 buildings	612 buildings
30	Use residential last	Numeric	From 0 to 100%	80.70%
31	Use social gathering/restaurant last	Numeric	From 0 to 44.3%	2.01%
32	Use health last	Numeric	From 0 to 100%	0.75%
33	Use industry last	Numeric	From 0 to 100 %	2.52%
34	Use office last	Numeric	From 0 to 100 %	4.33%
35	Use hotel/accommodation last	Numeric	From 0 to 97.0%	0.32%
36	Use education last	Numeric	From 0 to 33.3%	0.97%
37	Use sport last	Numeric	From 0 to 100%	0.28%
38	Use shop last	Numeric	From 0 to 82.4%	3.63%
39	Use other last	Numeric	From 0 to 46.7%	4.00%

Table 1.2: Overview of variables of the final combined dataset

1.4.4 Data exploration

Figure 1 shows the spatial distribution of trips in South Limburg and the trip distributions across building use categories. There is hardly any discrepancy between the geographic distribution of the trip origins and destinations; most trips start and end in cells located in and around South Limburg's villages and cities (Figure 1.1a). Moreover, the trip distributions per building use behave similarly for the origins and destinations (Figure 1.1b). Unsurprisingly, residential buildings dominate, constituting more than 85% of all buildings in South Limburg—a predominantly residential region with few large industry clusters.

Each residential building serves only a few people, whereas a special-purpose building, such as a school, restaurant, gym, office building, or hospital, can serve hundreds or thousands of people. Consequently, residential buildings are the most common type of building. For research goals where residential buildings are less important than the composition of the remaining categories, researchers can assign weights to categories to decrease the dominance of one category.

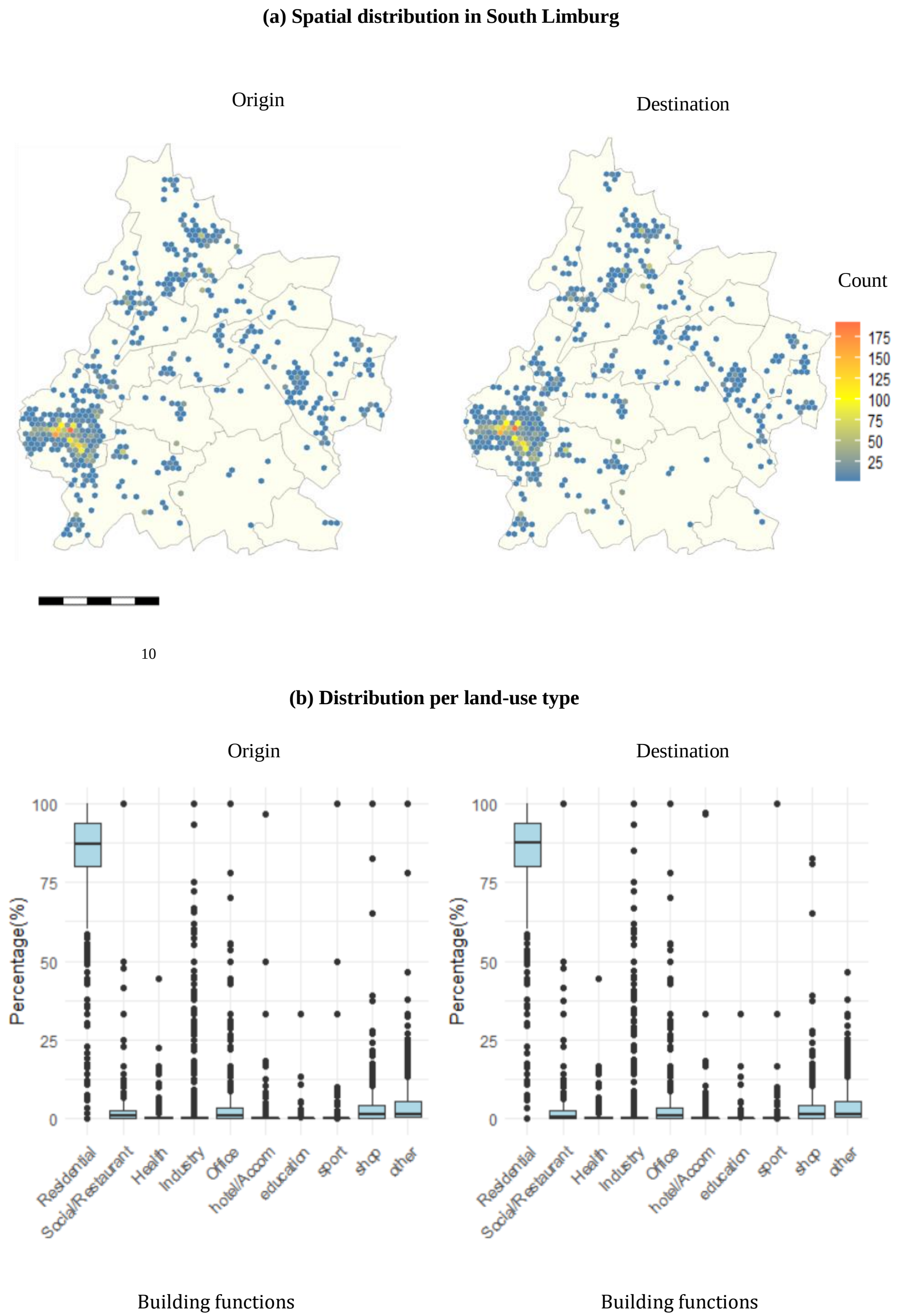


Figure 1.1: Spatial distribution of South Limburg as per building use category

The majority of trips (80.45%) occur during the week, while the remaining trips (19.55%) take place on the weekend, which is a slight skewness toward weekdays. Additionally, the largest proportion of trips are recorded on Friday (18.25%), whereas the smallest proportion (7.43%) are logged on Sunday.

Figure 1.2 shows the distribution of trip starting times, expressed in hours of the day and during the week vs. the weekend. The left histogram for the distribution on weekdays shows two peaks: one peak between 7AM and 9AM and another larger peak between 4PM and 7PM. These are in line with the morning and evening peak traffic hours of the study area. In contrast, the right histogram shows a bell-shaped curve with the peak in the middle of the day. Also, more trips start in the late afternoon and at night on the weekend than on weekdays.

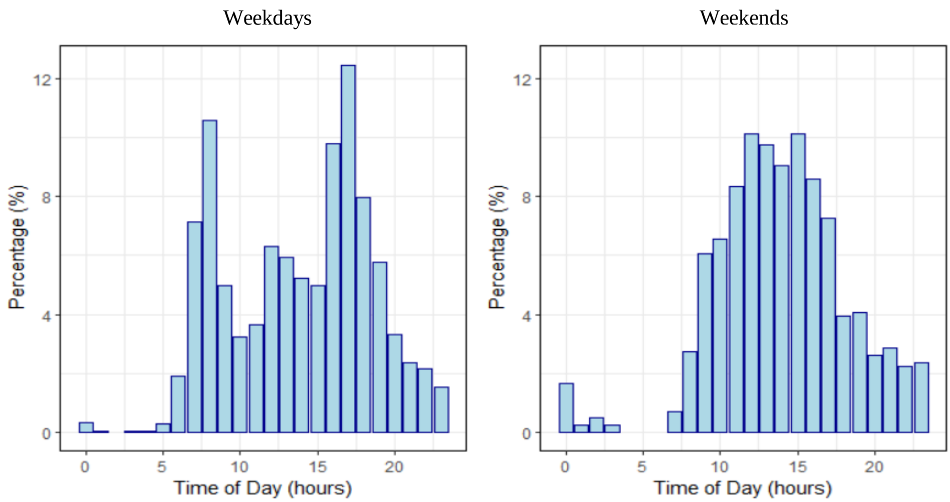


Figure 1.2: Temporal distributions of the starting of trips—weekdays vs. weekends

1.5 Conclusion

In this chapter, I have shown that there are many different branches within the LBM literature, such as location-based advertising, mobile marketing, and location-based services, whereby researchers explore the effects that these marketing initiatives have on store visitation rates, purchase intention, actual purchase rate, and brand awareness. Research also compares the effectiveness of in-store vs. out-of-store LBM or mobile marketing in contrast to non-mobile marketing. However, all these research streams only focus on one aspect of LBM; they base marketing efforts on the temporary positioning of a consumer and use that information to place advertisements on marketing channels such as social platforms or company-owned apps. In doing so, they ignore the core concept of LBM: the location itself—consumers' geographic surroundings and their movements in space. Using the building (composition) at a consumer's destination, we can identify the activity they are likeliest to be pursuing, gather information about their daily schedule or behavioral routines, or infer their current mood. This information generates an advantage for targeted marketing and hyper-targeting. Even though geographic information in the form of spatial data has been neglected in marketing research so far, the integration of such information in future research is desirable because it will open up new perspectives and improve hyper-targeting.

Spatial data comes in different formats and therefore requires the appropriate software and methods for data processing. Often, geographic information is stored in vector layers of points, lines, or polygons with geographic coordinates. To work with such shape files and other datasets in more commonly known formats, the ArcGIS program in combination with R provides a suitable software setup for marketing researchers who want to include spatial data in their research. Within these software packages, there is a diversity of tools available that not only addresses a variety of research goals and fits different data formats but also caters to all levels of experience in working with spatial data.

Additionally, ArcGIS is designed to make another crucial exercise easy when working with spatial data. Spatial analysis frequently requires combining multiple data sources (e.g., maps, building data, grids, trajectories, infrastructure features, and land features). Researchers engage in intense feature engineering to extract as much information as possible from the data sources and to achieve insightful analyses. To make this more tangible for marketers just getting started with using spatial data, I introduced two datasets from South Limburg—one with building usage information and one with human movement trajectories—that are typical examples of spatial data suitable for marketing decision-making and research. In the following five chapters, I show how to implement some suitable methods that were briefly touched upon earlier in this chapter.

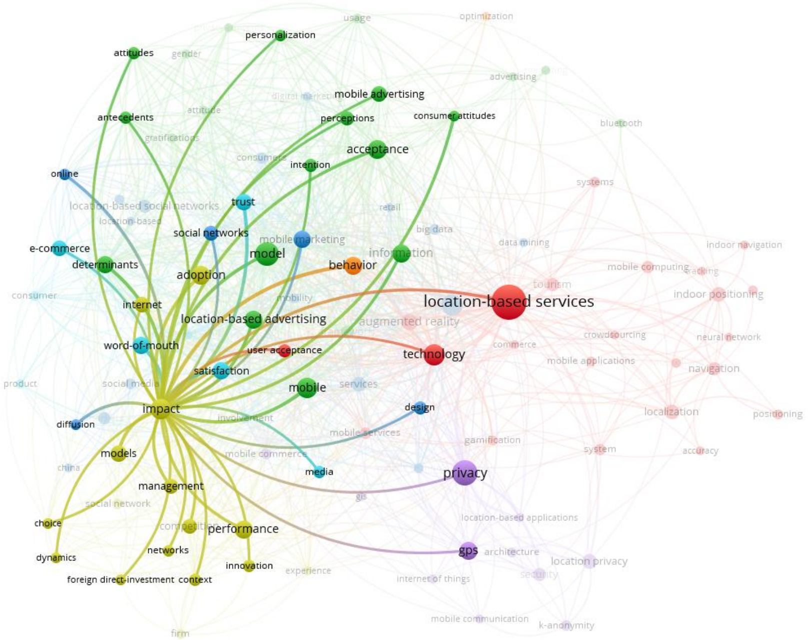
1.6 Impact on hyper-targeting

In the introduction to this dissertation, I mentioned that hyper-targeting can draw information from five sources: demographics, location, interests, online behavior, and offline behavior. In this chapter, it has become clear that LBM practices and literature currently focus only on the demographics and online behavior of consumers, even though the other sources are equally important. Observing (dynamic) geographic information about a consumer and information about the city is relatively easy nowadays. Proxemic technologies, such as wearable user devices, observe dynamic consumer movements in cities via GPS. Combining other offline behavior measures, such as travel card information or credit card data, with online behavior data provides powerful insights that allow for the creation of precise customer personas. Further, digital advertisement needs to be city-specific because consumers value customized mass advertisement more than non-customized promotions. More information about the physical infrastructure enriches the database and paves the way for improved hyper-targeting. Consequently, we can conclude that location information—both in the form of dynamic consumer movements and physical infrastructure information of a city—is crucial to the ability to provide the super-specific marketing offers of hyper-targeting.

Transition 1: Adding physical context

CBM research is currently missing the dynamic perspective. Therefore, the following five chapters apply different viewpoints and approaches to discuss why this perspective is fruitful for research. Each chapter provides an example of how we can implement these approaches with different methods that fit each research goal and data source. More importantly, CBM lacks the physical context; its horizons should be broadened to encompass the use of geography in marketing. Chapter 1 concluded that CBM only considers demographics and online behavior. Therefore, Chapters 2 to 5 include location data and offline behavior information into CBM studies, and Chapter 6 combines interests, demographics, and offline behavior.

Chapter 2 focuses on buildings, infrastructure, and neighborhood characteristics to deepen our understanding of how well certain places within a city are connected to each other and what this means for the use of mobility services and customer experience in cities. I decided on this analysis as the next step because I believe that a clear understanding of the city is important before I start adding the citizens to the research picture. The addition of layers, features, factors, and influences increases the complexity of the analysis and makes the overall topic more difficult to grasp, so I decided to look at the static infrastructure of the geographic area before adding the dynamic movements of individuals. In the upcoming chapter, I use the example of a UAM service, which will most likely be introduced in the next decade, to critically assess the usefulness of building interconnectivity analysis. Because we are focusing this analysis on infrastructure and not on human movements, the implications are also interesting for infrastructure providers and city authorities. Municipalities can use this method and the results to evaluate the performance of the current infrastructure and make necessary improvements.



Chapter 2

Sketching the City Space I—A Multi-Layer Approach to Preparing for Future Urban Air Mobility Services

Under review at the *Journal of Services Marketing* for the special issue on *The role of physical context and the customer experience*.

Schmitt, H.S., and Lemmink, J.G.A.M. (2022), “Sketching the City Space – A Multi-Layer Approach to Preparing for Future Urban Air Mobility (UAM) Services.", *Journal of Services Marketing*.

The authors thank Duy Thai Dao for carrying out part of the literature review and data analysis.

2.1 Introduction

How can we make a city more accessible? How do customers perceive the current transport offerings? Where are the transport service gaps? What could infrastructure improvements look like? To understand customer experience in transport service settings, we must recognize that people travel through the city to fulfill needs in various locations. So far, service research has focused on experience criteria along the customer journey such as access to information and visibility prior to the service encounter as well as timing, service frequency, pricing, and capacity during the service encounter (Ameen et al., 2021; Lemon & Verhoef, 2016). However, the impact of the built environment on city services has not yet been addressed, even though it can be considered a core criterion. Buildings are the places where people interact; thus, evaluation of city services of any kind needs to look at the positioning of buildings in relation to each other and their accessibility by various modes of transport. For example, dense and labyrinthine city centers with narrow alleys, one-way streets, and no parking spots are inconvenient for cars; therefore, they need good public transport service. Whenever the physical context of a service encounter plays a substantial role, building interconnectivity introduces a new perspective for identifying service gaps or suitable service spots. Consequently, we must integrate geographic information into marketing and service research via appropriate approaches and methods.

To date, we have hardly seen the use of spatial data in service research (Bernritter et al., 2021; Ghose, Li, et al., 2019). A potential explanation could be that once we consider customer experience in a physical context, the research problem becomes inherently multidisciplinary, requiring unfamiliar research methods and a solid database that covers all input features. Datasets must be large enough for robust results without violating personal privacy, and they must be high-resolution enough to obtain meaningful, concrete conclusions. Nevertheless, it is important to work on answering the aforementioned questions and to find new ways of approaching complex problems. Sketching the city space involves studying a city's infrastructure based on chosen criteria and mapping it out visually or numerically. This activity relies on the combination of multiple data sources and state-of-the-art analysis techniques to generate meaningful insights regarding the city. Among other things, spatial data can assist marketers in continuously monitoring and evaluating emerging transport innovations. In this chapter, we focus on the fictive scenario of determining suitable locations for urban air mobility (UAM) access points before the service is ready to be launched.

UAM, a new service concept, makes use of drones to transport passengers who are able to afford it or who rely on fast transportation through the sky (Straubinger et al., 2021). Cities with a lot of congestion could benefit from UAM, which would reduce overall travel times (Rimjha et al., 2021). However, the concept of UAM is relatively new; approximately 200 companies worldwide are still working on being the first provider to offer these services to the public (Straubinger et al., 2020). Meanwhile, many questions still need to be answered to offer safe and useful travels with UAM. Aside from the technical and legal challenges pertaining to UAM, one economic challenge is determining suitable locations for these services to appeal to the needs of the public and build trust within society. Finding spots for UAM landing sites is the first step in its integration into the existing mobility service environment.

This chapter aims to find these spots in Maastricht by combining multiple data sources to pinpoint optimal locations that would not be identifiable with only one of these datasets. In the end, we determine optimal locations based on a broad set of factors, including but not limited to safety, noise pollution, built environment, existing transport infrastructure, topography of the region, characteristics of the neighborhoods, citizens' behavior and preferences, and privacy. In this investigation, we limit ourselves to behavioral data from GPS tracking, census data, building data, and transportation data. By analyzing individuals' trips, we establish a better understanding of the mobility demands in a certain geographical area. Integrating spatial data from the surrounding buildings and census data of the region helps to identify suitable locations for UAM landing sites, also known as vertical takeoff and landing (VTOL) stations (Rimjha et al., 2021; UBER, 2016). When mobility services are placed where they are needed the most, the value for citizens increases because distances to the closest service spot shorten, and people can use this service to get to their destination faster—rather than being stuck in traffic or waiting for public transport.

2.2 Background and research gap

2.2.1 Customer experience in transport

Service scholars have called for research on the role of the physical context in customer experience. This niche topic in service marketing is backed by two well-established marketing concepts. First, customer experience describes an individual customer's perception of a service that is subjective and highly contextual (Akaka et al., 2015; Edvardsson et al., 2005; Lemon & Verhoef, 2016; Still et al., 2018). Lipkin and Heinonen (2022) expand the perspective from *customer experience between firms and customers* to *customer experience in ecosystems of multiple actors*. Findings show that actors in the ecosystem influence a customer's experience throughout the customer journey (Lipkin & Heinonen, 2022) and that the physical environment impacts customer experience in shopping malls (Ameen et al., 2021). Hence, we propose broadening the perspective even further and inspecting the customer experience within city limits.

Consequently, the second concept frames the physical context of a city as a servicescape, which means that the entire city represents the environment in which the service can take place and that service encounters can occur in multiple locations within the servicescape (Bitner, 1992). Shortly after the introduction of this framework, researchers recognized its importance for customer satisfaction in different settings (Wakefield & Blodgett, 1994), and its importance has only increased with the diversification of the service landscape through technology and digitalization (Wakefield & Blodgett, 2016). Sheng et al. (2016) view cities as servicescapes to examine servicescape attributes that affect consumer well-being, and they identify public transportation as an attribute that leads to satisfaction with urban life. On-time performance, travel speed, service frequency, network size, and information availability are the main drivers of customer experience with public transport services (Edvardsson, 1998; Hutchinson, 2009; Mouwen, 2015) and the starting point for the identification of transport service gaps. Therefore, the next section elaborates on UAM as a new service concept that can fill transportation gaps in cities.

2.2.2 Urban Air Mobility

Whether the benefits of a car outweigh the disadvantages depends on the user's perspective and the situation. The door-to-door solution of driving directly from the home to any other location is convenient. However, the emissions produced by fossil-fueled cars have a negative impact on the environment and human health. Thus, driven by strict government regulations and policies, car manufacturers have announced a shift toward electric air vehicles instead of combustion engine cars. The South Korean government and Hyundai plan to have the first UAM routes operating by 2025 (Garrett-Glaser, 2020). The UAM flight vehicle with an electric propulsion battery is expected to transport 2–5 people to and from VTOL sites that are well integrated into the existing infrastructure (UBER, 2016). The UAM community differentiates between air metro and air taxis for passenger transportation: air metro resembles current public transport options, such as buses and subways; air taxis, considered a niche service, offer door-to-door flights similar to ride-hailing options (NASA, 2018). To avoid confusion, this chapter will focus on the air metro concept when referring to UAM.

Existing public transportation modes can accommodate only a fraction of the urban mobility demand within a city because many citizens revert to car usage (Tachet et al., 2017). UAM seeks to supplement traditional public transport in scenarios in which the latter cannot serve an individual's mobility needs. However, only in combination can public and commercial travel modes yield a viable alternative to using private vehicles (ERTICO, 2019). For UAM to attract society's masses, it must be placed in locations that maximize utility. The ground infrastructure for UAM landing sites consists of VTOL sites for multiple or single aircrafts that also function as charging stations. They are either based in areas that are frequently visited but time-consuming to reach or at transport hubs; moreover, they are built on top of buildings, existing helipads, or parking garages (UBER, 2016). Finding these mobility hotspots is challenging because accurate mobility data of a specific region is not always readily available. Furthermore, due to privacy rights, the data depends on travelers voluntarily providing the relevant information (Jevinger & Persson, 2019).

2.2.3 Placement of vertical take-off and landing sites

To find possible locations for VTOL sites, a study by UBER (2016) applied the k-means clustering algorithm to trip data. With this method, they reduce the number of potential locations for VTOL sites to a smaller, manageable set. These spots represent locations where trips started or ended—locations where people could opt for air vehicles instead of cars. However, the study is based on many assumptions and only examines the UAM on-demand concept, as opposed to a scheduled air metro network concept. Rajendran and Srinivas (2020) explore suitable locations for air taxis by clustering travel data from actual taxi trips in New York City, and they use the k-means clustering method to visualize 300 million individual taxi trips. The results show that the main mobility locations are next to airports, large parks, and sightseeing spots. Willey and Salmon (2021) consider the constraints of the vehicle (e.g., travel speed and battery range) when conducting their research; based on existing public transportation routes, their study further analyzes potential designs for a UAM network. They reason that, in the beginning, networks of distributed VTOL sites are more useful than dispersed networks with single start- and end-point travel routes. Stations can be placed in popular locations to significantly reduce travel times for inhabitants and visitors. Rimjha et al. (2021) also indicate that UAM can only succeed if it is offered at low fares. Additionally, its reliability must be comparable to that of automobile transportation.

Sevtsuk and Mekonnen (2012) show that buildings along straight thoroughfares are assigned higher straightness values than buildings on side streets. However, the former are not particularly relevant to our research, because they are located along a main road and are easy to access. Furthermore, modes of public transport are already in place on heavily trafficked roads, making the usage of UAM in these areas questionable; UAM is most beneficial in places where congestion and overcrowded infrastructure prolong travel times. Therefore, people living in or visiting districts with low-straightness buildings would benefit from UAM because the air vehicle is able to move through the sky in a straight line. As detours are prevented with a direct flight, transportation as a whole becomes more efficient and less time intensive. Hence, placing VTOL sites in these areas would make traveling more efficient.

The VTOL network needs to match the demand. Thus, two additional types of VTOL sites will likely be necessary: (1) inter-regional transport hubs where people arrive by train or plane and want to use UAM for their final trip segment to a low-straightness destination are reasonable (Fadhil, 2018); (2) there are wealthy individuals who would use UAM to leave their home in the morning to avoid rush hour traffic on the roads and thus would appreciate VTOL sites in their neighborhoods (Straubinger et al., 2021).

2.2.4 Research gap

Fadhil (2018) uses a GIS tool to determine suitable locations for UAM in Los Angeles and Munich. However, his analysis focuses on criteria such as the minimum physical and technical requirements for VTOL sites without considering the spatial relationship between buildings. In particular, the study does not incorporate the built environment of the city and its impact on VTOL placement, which includes examining the distance between buildings as well as their accessibility, the shortest direct walkways between buildings, and the natural barriers between them (e.g., intersections or traffic lights). Except for Cho and Yoon (2021), no researchers consider the spatial relationship of the buildings when choosing suitable locations for UAM infrastructure. This is, perhaps, because the UAM field is relatively new. However, in this chapter, we view the integration of building data into the analysis as crucial because buildings are the places where people meet and interact with each other; therefore, buildings are important elements in the determination of VTOL sites (Lipkin & Heinonen, 2022; Sheng et al., 2016). Based on the literature presented above, this chapter establishes four criteria that locations should fulfill to qualify as suitable for VTOL sites:

- High mobility activity
- Weak direct connection to other buildings in the city
- Above-average household income
- Good match with existing modes of public transport

2.3 Research approach, software, and data

In this chapter, we use a combination of techniques to reach the goal of finding suitable UAM sites based on the four criteria presented above. This section addresses the research flow, highlights the steps we take, and introduces the analysis software. The chapter uses both datasets introduced in Chapter 1 of this dissertation (i.e., OD trajectory data and built environment data on individual building level) and adds openly accessible purchasing power data from CBS and public transport availability data from OSM.

2.3.1 Research flow

There are five steps in our analysis. First, we run a KDE algorithm on the movement data to determine where most mobility activity occurs. Second, we apply the straightness measure of the UNA toolbox, which captures the connectivity of buildings and calculates the travel distance difference between the shortest street path connecting two buildings and the direct flight distance between the same buildings; as a result, each building in the region receives a straightness score. Third, we filter for buildings with low straightness (i.e., buildings hard to reach from other buildings because many turns must be taken and many intersections must be crossed). Because UAM circumvents these turns and can use the straight line from one building to another, low-straightness buildings are the ones most attractive for UAM service flights and thus most relevant to this analysis. For filtering, we use the *Anselin Moran I* statistic for spatial association. Because the number of resulting clusters is too large for meaningful interpretation, we apply KDE to the low-straightness clusters for clearer differentiation between high- and low-straightness areas. Fourth, we combine the data layers of common trip starting points and low-straightness buildings to determine whether there is an overlap. Such locations would be appropriate for VTOL because the many people who depart from these places could reach their destination more quickly via air travel. Finally, we add the purchasing power of the neighborhoods and public transport availability as additional layers to our map to include more features that influence the adoption and success of UAM in an area. Based on this combined map, we are able to suggest suitable locations for UAM landing ports. It is worth mentioning here that Maastricht is unlikely to be the first city to implement UAM; the required investment is large, and UAM is most suitable for large metropolitan areas. Nevertheless, we use data from Maastricht and use the city as an example to demonstrate the method. Figure 2.1 summarizes the research approach.

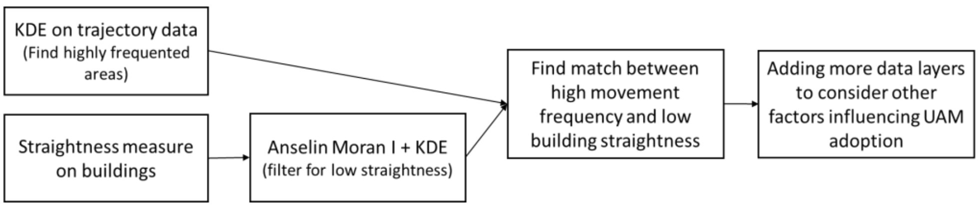


Figure 2.1: Research approach

2.3.2 ArcGIS as an analysis software

GIS is commonly used in spatial analysis (e.g., for identifying traffic accident hotspots or busy retail streets; (Anderson, 2009; Porta et al., 2009). Most urban activity locations have an underlying network structure that intertwines with the built environment (Natapov et al., 2018). Consequently, using the ArcGIS Desktop software and the UNA toolbox to determine suitable locations for VTOL sites seems to be a viable option. The toolbox is an extension to the standard ArcGIS environment and provides "powerful methods for analyzing spatial accessibility, pedestrian or bicycle flows, facility patronage along spatial networks and/or the effects of pedestrian or bike infrastructure on non-motorized access and flow" (MIT, 2022). Sevtsuk and Mekonnen (2012) in association with the City Form Research Group at MIT, developed this tool for ArcGIS Desktop to calculate the spatial relationship of buildings with other buildings in an area. ArcGIS Desktop is a GIS software that can work with large amounts of data and can, among other things, analyze spatial relationships and map desirable locations for new businesses. The UNA tool incorporates the street nodes, edges, and buildings of a spatial network into its analysis to answer social, economic, and transportation questions (Sevtsuk, 2017).

2.4 Empirical analysis

2.4.1 Kernel density estimation on trajectory data

Every movement is represented by cell ID numbers corresponding to the cells in which they start and end. Figure 2.2 depicts the distribution of the starting and ending cells for all movements in the dataset. Here, if multiple trips start or end in a specific cell, they are bundled for this visualization, and the cell is only shown once. We can see that most movements start or end in Maastricht and that the rest of the trips originate or terminate mainly in the Heerlen and Sittard areas. Because a participation criterion for the data collection was that people live or work in Maastricht or frequently visit the city for other reasons, the skewness toward origins and destinations being in Maastricht is natural. However, because the analysis in this chapter focuses on the city—not on the entire region—this skewness is not an issue.

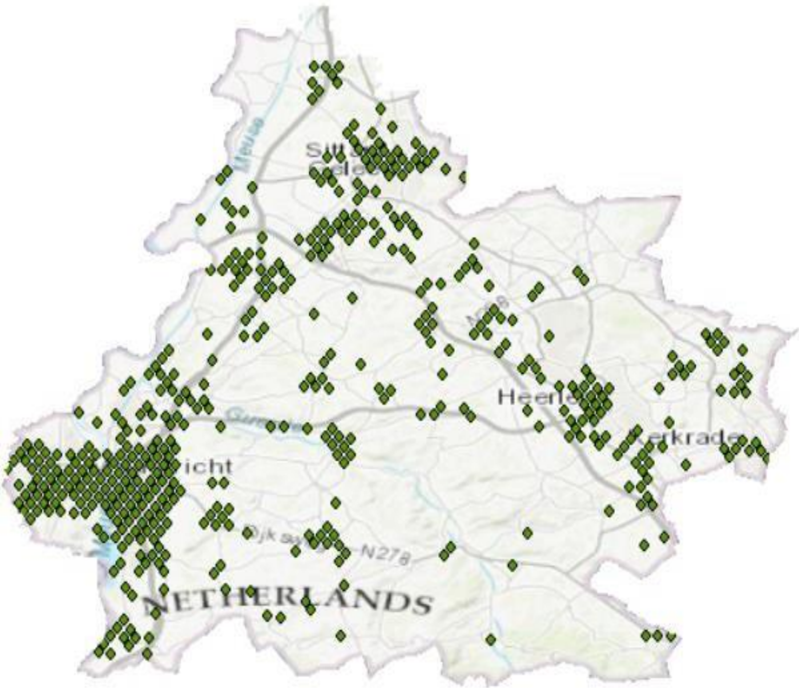


Figure 2.2: Visualization of all cells that recorded at least one trip origin or destination

This first glimpse of the geographic distribution of trips provides us with an understanding of where movements with a potential demand for UAM services occur. To reveal highly frequented areas and differentiate them from rarely frequented areas, we apply the KDE method to all recorded travel activities; this method divides the entire study area into predefined cells of the same size and calculates the distance from the point of interest to a reference location (Anderson, 2009). The calculations are then summed for all surfaces in that particular reference location. By iterating this step for all points, a circular area—or "kernel"—is placed over each observation. Adding up all the kernels gives a density estimation for that area, and as a result, a smooth and continuous surface is created. By inserting the travel activity for all cells into the KDE algorithm, we create a map that shows the mobility activity density within the region (Figure 2.3).

Depending on the mobility activity, the areas that show more activities will be denser and therefore darker. The algorithm provides density values ranging from 0 to 4.29, which are divided into nine different categories. We use human trajectory data rather than open-source population density data because population density does not necessarily reflect highly frequented areas. For example, industrial areas with large factories and office buildings are highly frequented during the day by people going to work, but industrial areas usually have almost no residential buildings and thus have a low population density. To avoid this bias in the analysis, we opt for trajectory data. Further, we argue that current high mobility activity presents a good proxy for future mobility demand and thus potential locations for UAM. It would not make sense to place VTOL sites in locations where little activity is taking place; therefore, we consider the upper six KDE categories and disregard the lower three categories. To be more precise, only areas with a kernel density above 1.43 are included in this analysis.

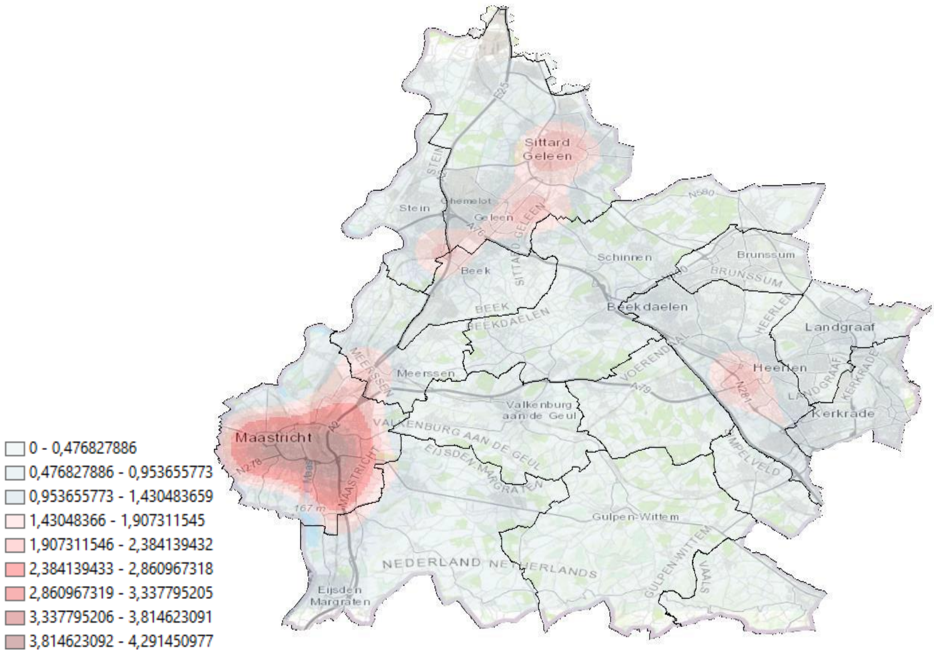


Figure 2.3: Kernel density estimation of the movements in Limburg

2.4.2 Straightness

In theory, vertiports or vertistops can be placed everywhere within the area identified by the KDE method. However, to narrow down the location options, we continue the analysis by calculating the straightness values for every building in Maastricht. Two core components are required when conducting building interconnectivity analysis in ArcGIS. First, we need the latitude and longitude information of each building in the region for the exact positioning of each data point. Second, we need to create a network dataset of the streets in the region; for this, we use the open-source data from OSM.

A network dataset models the loose collection of lines into a network by capturing the junctions and turns of the lines. Moreover, by retaining features such as street length or allowed driving directions, a network of streets can be used for various analyses. In this study, the network dataset serves as a basis for calculating the straightness measure of every building in the region. Using the UNA toolbox, we create an OD matrix in which we calculate every possible path to each building in a particular area to determine the associated straightness value. Straightness is defined as follows:

$$C_s^r[i] = \sum_{j \in G-[i], d[i,j] \le r} \frac{\partial[i,j]}{d[i,j]} \qquad \text{(Equation 2.1)}$$

Where

$C_s^r[i]$ is the straightness of building i within the search radius of r,

$\partial[i,j]$ is the Euclidean distance between buildings i and j,

$d[i,j]$ is the shortest network distance between buildings i and j.

The idea behind this formula is to find the shortest street path between two buildings in a given urban network and compare it to a direct flight route (Equation 2.1). In other words, we take the straight-line Euclidean distance that connects building ***i*** with another building ***j*** and compare that distance to the walking/ driving distance—taking into account turns and crossings along the way. The fewer deviations the walking/driving route has in comparison to the virtual straight line connecting two buildings, the higher the straightness value for that particular building. Hence, Straightness value measures the extent to which a building can be reached directly in straight line from all other buildings in a city (Porta, 2009; Sevtsuk, 2012).

Buildings with low straightness values are not well connected to other buildings in the urban environment. Therefore, only buildings with low straightness values are considered for the analysis. In Maastricht, we observe a wide range of straightness values because the area is abyrinthine and densely built, and shorter distances between pairs of buildings increase the difference between the as-the-crow-flies and path distances. Another reason that the straightness measure is qualified to determine suitable locations for UAM is that it innovatively includes the built environment and interrelated building positioning in the calculations. As mentioned above, the first VTOL sites will most likely be on top of buildings, existing helipads, or parking garages. Moreover, buildings are where people meet and interact; hence, examining how buildings are interconnected will lead to a more informed decision regarding which building roofs VTOL sites should be placed on.

By running the UNA tool for Maastricht, we create a map that shows buildings with high and low straightness values (Figure 2.4). Each point on the map represents one building, and the color indicates the degree of straightness calculated for this building. Buildings with high straightness scores (i.e., buildings reachable from many other buildings without passing infrastructural obstacles such as traffic lights or junctions) are marked in red. Buildings with low straightness scores (i.e., buildings difficult to reach from most other buildings because of infrastructural obstacles) are marked in dark green. The colors orange, yellow, and light green fill the spectrum in between. We can observe that the city center on both sides of the river Maas, as well as some neighborhoods further outside, shows low straightness. This makes intuitive sense, especially for the city center; Maastricht is an old city that expanded its city boundaries naturally and not through planned neighborhood developments, and the city center is densely built with many small streets and junctions.

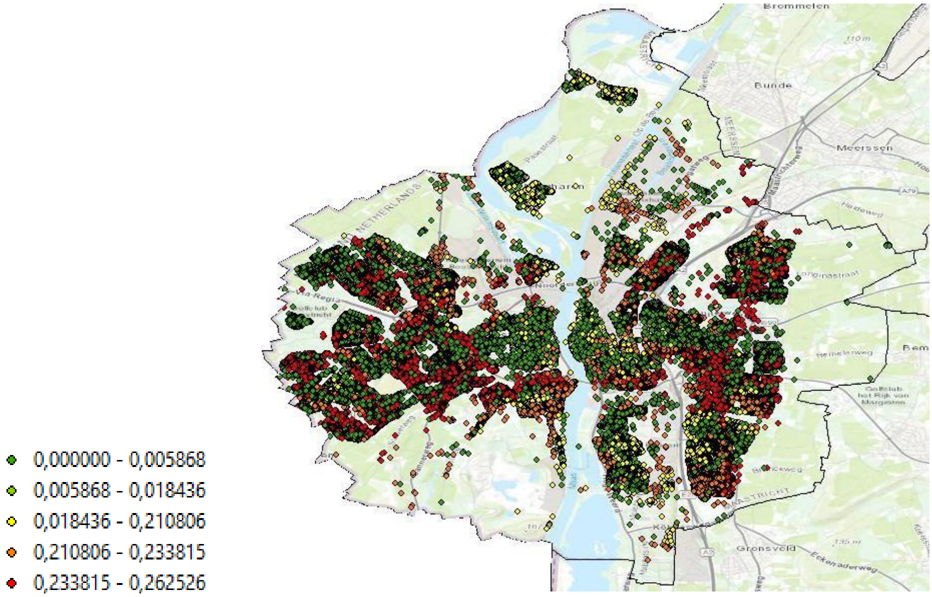


Figure 2.4: Straightness values for all buildings in the Maastricht municipality

2.4.3 Anselin Local Moran's I statistic combined with kernel density estimation

As mentioned above, buildings with high straightness values are of no interest for this research because they are already well connected. To filter these out, the Anselin Local Moran's I statistic of spatial association is applied. It is designed to identify statistically significant spatial clusters of high and low values, as well as outliers in the data, by calculating the Anselin Local Moran's I value, a z-score (representing standard deviations), and a pseudo p-value (representing the significance level; (Anselin, 1995). In general, the output, which is given as the local Moran's index I, indicates whether the building feature (straightness value) is part of a cluster or an outlier. If the local Moran's index has a positive value, the building feature has neighboring features with similar high or low attribute values and is part of a cluster. If the index has a negative value, the building feature is labeled as an outlier. Whether the output is statistically significant is determined by the z-score and pseudo p-value results associated with each building feature. In addition, the results of the Anselin Local Moran's I statistic are automatically corrected using the false discovery rate for multiple testing and spatial dependency; this correction accounts for weaknesses in the analytical approach to return results that are more accurate.

Inserting the straightness values of all the buildings into this algorithm gives an output of clusters with high straightness values and clusters with low straightness values. We remove all clusters with high straightness values, and those with low straightness values remain, scattered across the map. Every depicted building cluster represents a potential location for a VTOL site. Because the large amount of clusters is difficult to interpret, we apply the KDE method to all the remaining buildings with low straightness values. Applying the same logic as before, we consider only the upper six KDE categories. KDE returns values ranging from 0 to 13; thus, setting the threshold value at 4.337 allows for a better picture of where most of the buildings with low straightness values are concentrated in the region. Figure 2.5 shows the outcome.

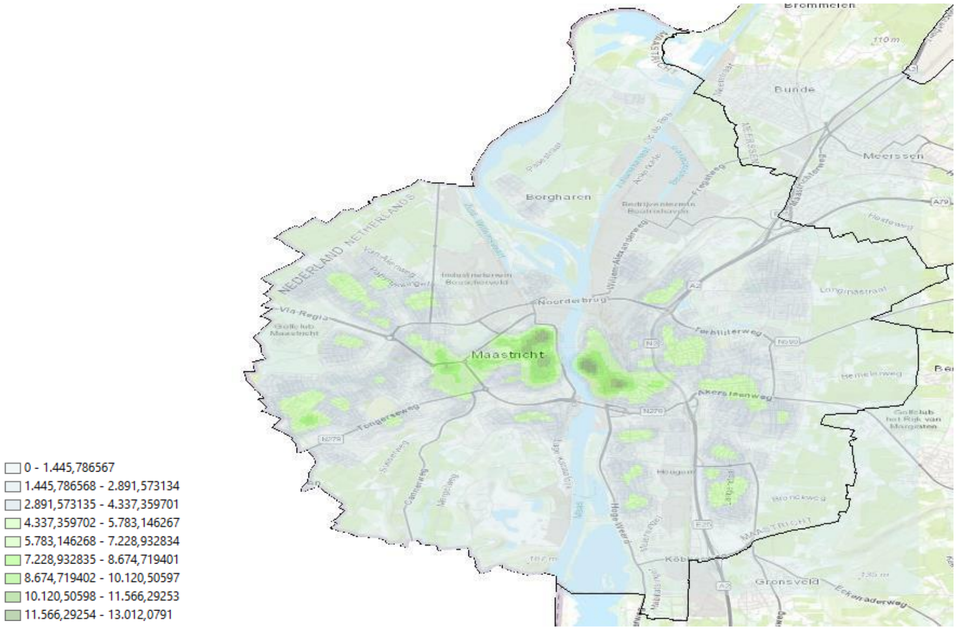


Figure 2.5: Kernel density estimation for low straightness buildings in Maastricht

Now that we have calculated the kernel density for the two core determinants of UAM landing ports, we overlay the highly frequented origin and destination kernels with the kernels of the low straightness buildings. We find the highest density of low straightness buildings in the city center of Maastricht. In general, buildings with low straightness values are mainly clustered around the city center and specific suburban areas for all municipalities (Figure 2.6). Further, we can observe a significant overlap between highly frequented areas and low-straightness building areas. This is the first promising sign toward identifying suitable UAM locations in the region.

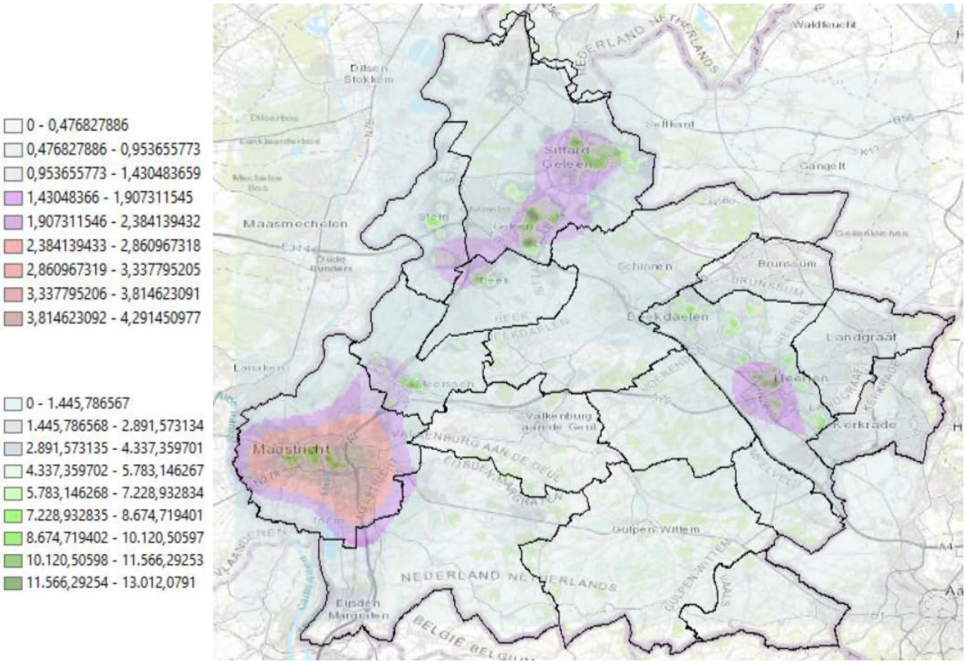


Figure 2.6: KDE of starting frequency (high trip density in red, medium trip density in purple) and KDE of low-straightness buildings (low building straightness in dark green, medium building straightness in light green) in South Limburg

2.4.4 Adding supplementary data layers

As we noted in the introduction of this chapter, sketching the city space and improving infrastructure are complex tasks because of the many perspectives and influential factors involved. Therefore, additional data sources such as neighborhood income level—because initially the service will be expensive and thus used mainly by wealthy people—should be used to fine-tune the results. Therefore, we incorporate an existing data layer about the purchasing power of the population in Maastricht in 2019.

Adding the purchasing power per capita data layer on neighborhood level with an average value ranging from 18,500 to 23,900 EUR, we see that predominantly low-straightness building areas overlap with only a few neighborhoods with purchasing power above the average. For example, one of the densest areas with low-straightness buildings in Maastricht lies in the Wyck neighborhood, which has a purchasing power per capita between 23,900 and 29,000 EUR. The remaining dense areas with low-straightness buildings, including city centers and most suburban areas of other municipalities, are mainly located in neighborhoods with average or below-average purchasing power per capita (Figure 2.7). Consequently, they are not ideal for the placement of UAM VTOL sites.

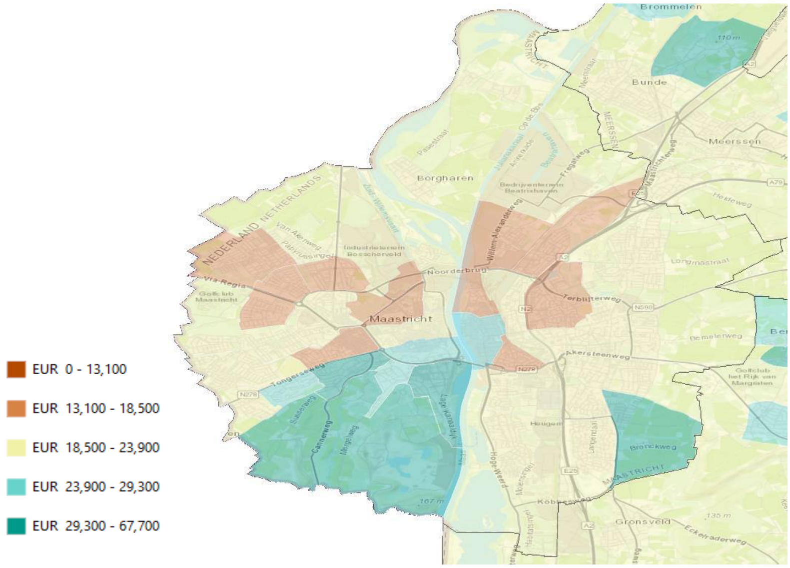


Figure 2.7: Purchasing power per capita in Maastricht

In addition to purchasing power, existing transportation infrastructure is of interest to UAM service providers. UAM needs to fit in with the existing public infrastructure to supplement instead of cannibalize. For example, placing VTOL sites at locations where users can easily transfer to long-distance modes of transport will most likely increase the usage of UAM services in the future. Thus, we add public parking spots and public transportation stops to the map (Figure 2.8).

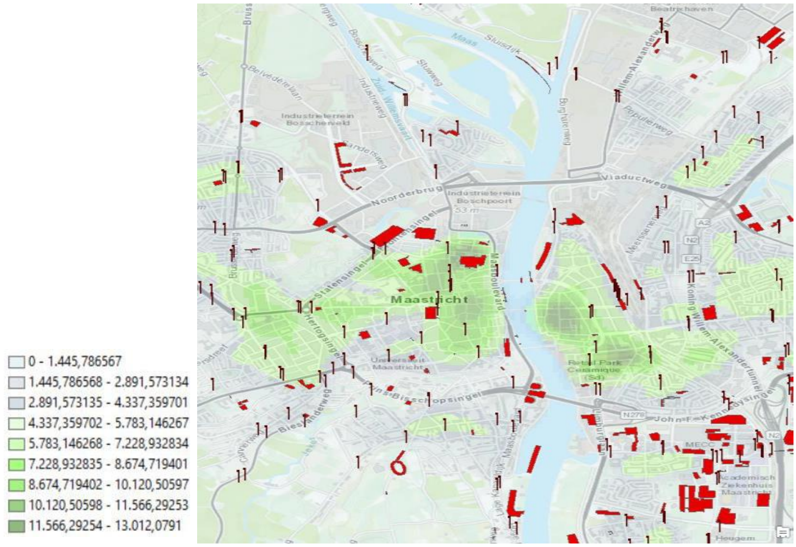


Figure 2.8: KDE of low-straightness buildings and specific public spots in the Maastricht city center (areas are parking lots; vertical lines are public transport stops)

2.4.5 Overlaying all data layers

The final step is to overlay all the layers and extract the most suitable locations for vertiports and vertistops in Maastricht. By adjusting several transparency measures for each layer, we create a proper visualization of suitable spots for UAM service locations (Figure 2.9). The first proposed VTOL site is in the Wyck neighborhood (second circle from the right). The high purchasing power per capita, strong public transport connections, high movement frequency, and low building straightness make this neighborhood a potentially valuable location for UAM providers. Here, we find Maastricht's central station and many bus stops; thus, this is a transport hub for multimodal traveling. Consequently, a VTOL site in this area would perfectly integrate into the existing public transport network. A typical UAM trip via this VTOL site would be taken by a wealthy Wyck resident who walks from their home to the station and uses UAM for a regional flight within Limburg, quickly getting to Heerlen, Sittard, or Venlo during rush hour, when trains are delayed, or when the highway is crowded. Alternatively, the VTOL site in Wyck can serve a visitor who arrives by train at the central station; the air vehicle, which is not dependent on bridges to cross the river, can take the visitor to the western part of the city faster than a bus or car could.

The second proposed location (second circle from the left) has similar infrastructure features but a slightly different reasoning for being used as a VTOL site. The city center of Maastricht, with its small and labyrinthine alleys, has low interconnectivity among buildings and low straightness values, even though the built environment is dense and the number of buildings in proximity is high. Additionally, a large pedestrian area makes most buildings difficult or impossible to reach by car, at least during the daytime. The purchasing power of residents is rather low, but the area is predominantly a city center—not a residential area—with shops, restaurants, hotels, and leisure activities; therefore, non-residents frequently visit this area to pursue a variety of activities. For example, for those on a tight schedule who just attended a meeting in the city center, a VTOL site in this location would be a convenient option to quickly get to the airport or even to Hasselt or Aachen for international travel, because cross-border connections by bus and train are poorly serviced and tiring.

Third, we propose a location close to the Belgian border (left circle). This area is a residential neighborhood with sufficient public transport connections to nearby neighborhoods and the city center, but it also has a relatively high purchasing power per capita—a crucial layer to consider when determining where to place VTOL sites in residential areas. Most trips from this site will be residents who want to get out of the city quickly, and they will need to be able to afford this new, expensive service. A typical UAM trip starting from this VTOL site would be taken by a wealthy resident who prefers door-to-door mobility solutions over public transport and wants to circumvent rush-hour traffic. Thus, they would use UAM to get to their office building somewhere in Limburg, where the air vehicle could land on the building roof. Notably, the neighborhoods close to the Belgian border have a higher elevation level than the city center or station; thus, UAM will bypass natural barriers for customers in this case as well.

Finally, we suggest a fourth location in the southeast of the city (right circle), an area that is also quite low in building interconnectivity. The highway cuts this area off from the majority of the city, which might generate additional demand for UAM, making this VTOL site prone to residents' urgent intra-city flights to the city center and neighborhoods close to the Belgian border. Alternatively, people coming by car to Maastricht could use this VTOL site similarly to the Park+Ride concept. Even after the completed construction of the highway tunnel, people traveling into the city face daily congestion due to the river and the fact that only two bridges are for motorized vehicles. Park+Fly via this eastern VTOL site could be an attractive way to circumvent traffic and shorten travel times.

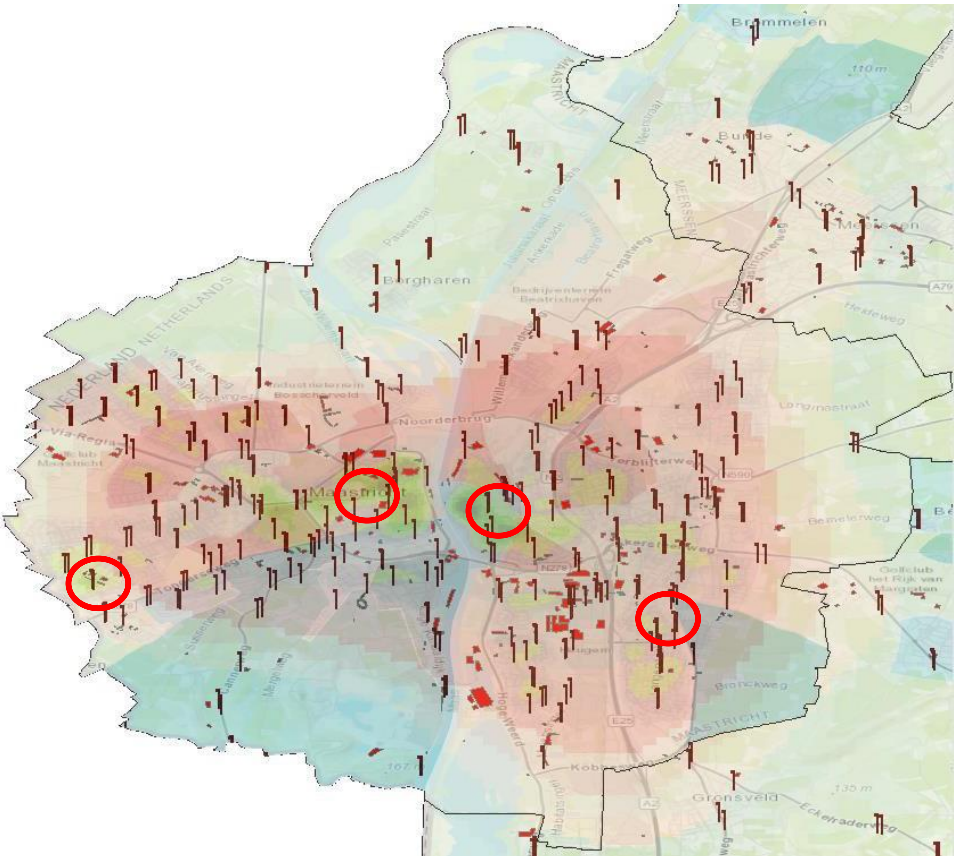


Figure 2.9: Map of Maastricht with all data layers

2.5 Conclusion

UAM remains a fictive scenario, but companies and researchers around the world are working hard on the vehicles to be able to launch the service soon. Many questions need to be answered before the launch, one of which is about the placement of UAM VTOL sites. To determine suitable locations, we assumed that such spots are situated in areas with high mobility activity, a weak connection to other buildings in the city, and above-average household incomes; moreover, these areas had to match well with the existing public transportation system, meaning the determined location is either well connected for multimodal trips or poorly connected and thus needs improvement in transport services. The following section discusses the advantages and disadvantages of UAM as a new service and of the analytical approach before broadening the scope and suggesting new contexts for the applied approach.

2.5.1 Reflection on urban air mobility as a new mobility service

Short-distance air travel in small electric vehicles for two to five people moving within a city or metropolitan area has advantages and disadvantages. On the one hand, UAM allows passengers to circumvent being stuck in traffic during rush hour or when there has been an accident on the road. Further, the movement above ground enables the flight vehicles to easily cross natural barriers, such as rivers, lakes, and mountains. Therefore, UAM is applicable not only in large metropolitan cities in Asia trying to fight congestion but also in cities built on uneven ground (e.g., Andean cities such as Medellin, where many people live in neighborhoods in the hills far away from the city center and business district). In addition, traveling through the air is more efficient, and introducing UAM to a city diversifies the urban mobility landscape. Finally, once it is affordable for the majority of the population, UAM can substitute or complement other modes of transport. With stations being conveniently located, UAM can substitute car travel with almost-door-to-door transport. Conversely, some VTOL sites can serve as concepts like Park+Fly to prevent overcrowding in highly urbanized districts.

On the other hand, the introduction of UAM to cities also brings some disadvantages to people and the environment. For example, when UAM is used as a substitute for cycling, it can have a negative effect on people's health. Further, noise pollution caused by air vehicles—due to their rotor blades—is probably higher than that caused by electric cars, and noise pollution has proven to have negative effects on (mental) health. Moreover, in the beginning, the service will be expensive and especially appealing to wealthy citizens. This naturally excludes some societal groups from access to the service, which is not desirable for a new service concept. In addition, it will be difficult to provide a network of VTOL sites without blind spots. Even with continuous evaluation and improvement, there will always be areas with service gaps. In this context, another aspect to consider is the sustainability of the service. While electric air vehicles produce less emissions and air pollution than cars or buses with combustion engines, the batteries of electric vehicles are difficult to recycle, and charging can only be considered sustainable when renewable energy is used (Jang et al., 2021; Smith & Hensher, 2020).

2.5.2 Reflection on analysis

This chapter used a combination of multiple methods and four datasets to answer the research question, and it is important to critically reflect on the methods and data used in the analysis. To the best of our knowledge, no previous study has used the same approach; the built environment and building interconnectivity have never been considered as factors in UAM location selection. This is surprising because buildings are an essential factor in urban life and activity—buildings are people's main meeting points for various activities.

The analysis simultaneously considers four factors for location selection and uses four distinct datasets to cover these factors. Examining multiple criteria within one analysis allows for a more complete picture of the situation, but a study is unlikely to ever capture all criteria at once. In this chapter, we had to limit ourselves to four factors, but we are aware that more need to be tested and included in future analyses. For example, we did not consider the degree of congestion on a route, which is a relevant factor because people generally want to avoid being stuck in traffic and are thus likely to opt for UAM on these routes. Low building straightness is usually caused by labyrinthine road infrastructures with many traffic lights, which result in an increased volume of traffic, but this factor might only hint at congestion without representing actual congestion volumes. Further, we did not filter our observed trips for mode of transport or trip length, and we included all movements. To generate more robust results, a replication of this analysis should consider only trips that are long enough to be applicable for UAM use or filter for trips taken in motorized vehicles.

Since a requirement for participation in the data collection was to live in, work in, or frequently visit Maastricht, the analysis only includes the city, even though many trips start or end elsewhere in the South Limburg region. This is mainly done to avoid bias in the results. Despite these restrictions on data use, we were able to achieve satisfying results for VTOL site locations in Maastricht using KDE and building interconnectivity analysis from the UNA toolbox.

The combination of methods and datasets allowed us to produce multiple data layers that we overlaid in ArcGIS for interpretation. We interpreted the results and inspected the individual layers and findings numerically. When answering a complex research question that requires diverse input features, an exceptional and unusual approach is required. We developed a research flow that works well with the four datasets and is also flexible enough to include other data layers if desired.

Researchers should be aware that the interpretation of the results requires knowledge about the geographic and cultural circumstances of the study area. For example, the final step in the analysis showed that not all factors are equally important for potential VTOL sites. City centers and transport hubs are frequently visited by citizens with a broad variety of backgrounds; thus, the residents' purchasing power is less relevant in these areas than in purely residential neighborhoods. Additionally, the Netherlands is known as a cyclist country, and UAM will not be attractive to individuals who will always choose to cycle regardless of distance and weather conditions. Nonetheless, the analysis conducted in this chapter—with its unique research approach and combination of multiple data sources—tackled UAM VTOL site placement, uncovering what had previously been left unexamined.

2.5.3 Using the method in a different context

The research flow developed in this chapter can also be applied to other scenarios; it is especially suitable for use cases with a geographic component and a dependency on the local built environment. With its versatility and multiple algorithms, this research flow can be adjusted not only by incorporating different data layers in the analysis but also by selecting between types of building interconnectivity analysis. For example, the method proposed in this study can also be used to find suitable locations for other new mobility services in a city (Echeverri & Salomonson, 2019); examples include public charging station placement for electric vehicles or the optimal placement of bike-sharing stations when a station-based concept is used (Bao et al., 2017; Xing et al., 2020). Furthermore, the proposed research flow could be used to determine where delivery drones can make a valuable contribution. In particular, when building interconnectivity is low in a neighborhood, a drone can deliver a parcel much more quickly than a mail truck can—the latter needs to stop at every traffic light and make many turns to get to the next house (Frachtenberg, 2019). Additionally, when researchers weigh the buildings in their datasets and focus more on one or more categories of buildings, building interconnectivity analysis can be included in a pre-entry market analysis. This helps in determining how many competing businesses are within a certain radius of a location of interest and can also be used to identify the number of complementary businesses nearby.

Hence, new areas of research that consider how buildings are connected to each other can be explored. Laying several data layers on top of each other in ArcGIS gives the researcher the freedom to test different datasets and make a data-driven decision regarding the research question. In general, such studies can contribute to the large research body on urban planning, not to mention new business development and transport operation optimization, all of which are dependent on the spatial structure of the city. Understanding infrastructure can lead to a better understanding of individuals' mobility behavior (Gonzalez et al., 2008).

2.6 Impact on hyper-targeting and future research

In this chapter, we proposed a new research framework that combines multiple methods to sketch the city infrastructure for improved service offerings. We combined spatial analysis techniques and multiple data layers to answer one concrete mobility-related service research question, thereby assessing the importance of considering the built environment and the usefulness of building interconnectivity analysis for different scenarios in city servicescapes and customer experience in urban areas. The flow of people through the streets is what keeps a city alive, just like blood running through a body. Therefore, a city's infrastructure should serve its citizens, matching their ideas of a livable city. The built environment plays a central role in this setup because buildings are interaction hubs. Consequently, the factors pertaining to the area around a building, such as the functionality of the nearby buildings and the transport infrastructure must match the purpose of the (majority of the) building(s) and the characteristics of the neighborhood to ensure a pleasant customer experience in the city. UAM will extend transport service offerings and thus add to customer satisfaction in the context of cities as servicescapes. However, cities considering the implementation of UAM should also be aware of its disadvantages, such as its exclusion of some societal groups and its debatable contribution to urban sustainability.

Cities should continuously evaluate new and innovative services, examining their applicability to the respective region. We demonstrated that with the right datasets and sufficient computing power, our research framework is easy to use for this purpose, and basic computer science and/or geography knowledge is sufficient to successfully apply this method. Rather, researchers need to be knowledgeable about their study area and the characteristics of the local population. Marketers can apply this method, combine multiple spatial datasets into powerful insights, and dive deep into the current situation in their city. Working with such information and non-traditional methods for marketing broadens the researcher's perspective regarding customer experience in urban areas.

With this chapter, we add to related marketing research streams (e.g., store location selection; (D'Silva et al., 2018; Duan & Mela, 2009; Hunneman et al., 2021); and service distribution across cities; (Mittal et al., 2004; Yang & Allenby, 2003).

This method and UAM can take hyper-targeting for transport services to the next level via new techniques, data streams, and combinations of datasets. The introduced method can cover all five data sources—depending on the available data sources—relevant to hyper-targeting in one study. Additionally, the introduction of UAM will broaden the mobility landscape. This allows but also requires more diversified mobility offers for customers. Transport needs vary greatly between individuals, so hyper-targeting is particularly applicable in this industry. Moreover, hyper-targeting in this industry can assist in making mobility more sustainable.

Finally, we would like to address the limitations of the study described in this chapter. First, the geographical scope is limited; Maastricht is a relatively small city, especially for an exclusive, investment-intensive service such as UAM. Generally, Maastricht does not (yet) have many areas that fulfill all the requirements necessary for ideal VTOL sites. Though every city center of each municipality where high mobility activity occurs has dense areas with low-straightness buildings and relatively good connections to existing public transport, many do not show above-average household incomes. This leads to the theoretical conclusion that the introduction of UAM in Maastricht is currently not viable. Nevertheless, by assuming that the cost of UAM will decrease in the future and that the purchasing power per capita will increase, some suitable locations can be determined for VTOL sites. Moreover, we suggest validating the introduced research flow and methods in a large metropolitan city that is likely to be a first mover in the adoption of UAM.

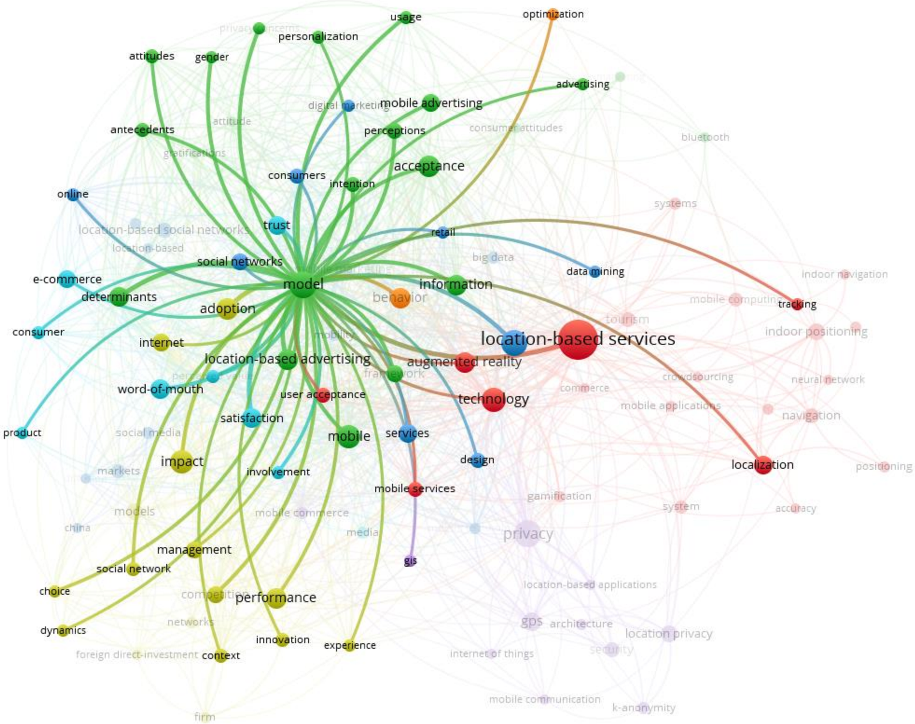
Second, approximately 6,350 movements across three weeks is an acceptable dataset size. However, these movements do not capture the behavior of all citizens; thus, some societal groups might be excluded through selection bias. In particular, the dataset excludes people who do not own a smartphone. Tourists and other visitors to the city are also excluded from this research. Future studies could use telecommunication data to observe the movements of more individuals, regardless of whether these individuals are residents. However, these data are aggregated, and ideal UAM positioning would be less precise than in our trajectory dataset.

Finally, more information on the region and additional data layers will further refine the analysis. In our transport data layer, we did not include the airport and taxi stands, congestion on traveling routes in the city, etc. Additional data layers about behavior, infrastructure, and neighborhood characteristics will further improve result accuracy. After all, the multi-layer approach introduced in this chapter has proven to be useful for explaining customer city experiences, which is an important factor in the context of cities as servicescapes.

Transition 2: Beyond building interconnectivity

Two basic factors are crucial to the success of a store or service offered in urban areas: (1) enough demand for the offered product or service, which can be identified via neighborhood characteristics—as we did in the previous chapter—, through pass-by rates or other options; (2) a healthy balance between competition and cooperation. However, an attractive location is determined by more than just demand and competition. City dynamics are a complex construct, and many different aspects can influence whether a certain location is popular. Detecting these aspects and their strengths is easier said than done, but more information leads to a deeper analysis, to a clearer picture of the situation, and finally, to better and more robust results. Therefore, Chapter 3 extends Chapter 2's building interconnectivity analysis, examining variables beyond building location and interrelated positioning. We build and evaluate a model with multiple factors that potentially influence the popularity of a location.

We derive and compute the input variables from the different datasets introduced in Chapter 1 and use traditional statistics in the form of a regression. Additionally, since we use spatial data as an input into a regression with a marketing target, we aim to use the next chapter to lobby for the renaissance of spatial modeling in marketing. Let's see how well traditional methods work with spatial data and how well such a model captures influences on location popularity. Identifying places where competition, complementation, demand, and other factors are in balance is another aspect of sketching the city.



Chapter 3

Sketching the City Space II—Using Spatial Data to Model Infrastructure Influences on Location Popularity

3.1 Introduction

Streets, squares, and other physical places in cities are popular for different reasons. The popularity of a location is interesting to consider for various stakeholders in a city, both public and private. Around the world, people struggle to create livable, eco-efficient, equitable cities with access to and connections between important destinations and services. At the same time, innovation and technological advancement cause continuous changes in urban lifestyles and citizen behavior (e.g., the competition between offline and online shopping). Maastricht and Heerlen, two cities in South Limburg, exemplify how this challenge differs even between nearby cities of similar size. Maastricht is currently struggling with how to balance the demands of its residents, while coping with the ever-growing number of visitors to the city center. Twenty kilometers away, Heerlen is struggling to keep its "dying" city center alive. In each case, a clear picture of the current physical configuration and the offering of products and services could support urban decision-makers in solving the challenge. Understanding why people perceive a place as attractive would allow Heerlen to revitalize its dying city center and would give Maastricht guidelines on how to increase the attractiveness of some non-central neighborhoods to relieve the pressure on the city center.

Some shops or services are reliant on walk-in customers and thus have different location requirements than specialty stores, which customers approach with a clear purpose. Additionally, some businesses require a particular transport infrastructure; for example, hardware stores that sell bulky and heavy items need large parking lots, while hospitals and health-care services require excellent public transport connections so that everyone has a fair chance of reaching the facilities. Consequently, to determine vital areas in cities, we must go beyond building interconnectivity analysis. Sketching the city is more than just looking at buildings and their interrelated positioning. Popularity analysis should use human movement data to determine the inflow and outflow rates of an area while including other variables and factors derived from infrastructure and movements to estimate why a particular area is more or less popular than another.

In recent years, technologies and modern sensing devices that collect human movement data have developed rapidly, and GIS researchers have explored many suitable techniques to analyze the data. Despite this disruptive change in urban data collection that has the potential to broaden and extend this research field, we have yet to see the comeback of spatial modeling in marketing. Most literature on the topic stems from before the ubiquitous computing era and thus does not work with big data and real-time observations of urban behavior. In the existing work, some factors influencing popularity have been studied in isolation (e.g., geographic proximity to the urban center, density of an area, retail accessibility, and demographic/socioeconomic characteristics of neighborhoods; (Andersson et al., 2014; Hunneman et al., 2022; Larsson, 2014; Öner, 2017; Yang & Allenby, 2003). However, the holistic picture that simultaneously considers multiple factors is still missing. To revitalize cities that currently lack popularity, it is vital that we attain this complete picture of what attracts customers to a place. Nowadays, we can identify these crucial factors by combining multiple sources of big data and using the most suitable methods.

This chapter focuses on the influence of infrastructure features on the popularity of a location determined via sensing data and spatial modeling. We study previous research, examining spatial modeling in marketing as well as location popularity research in related fields. Next, we model location popularity based on infrastructure features from a spatial dataset and then present our results, discussing the contribution of our findings. Finally, we broaden the narrative again and consider the impact of this chapter on the LBM field. Through this research, which takes advantage of opportunities presented by new forms of data, we want to support the few researchers (Del Gatto & Mastinu, 2018; Hunneman et al., 2022; Öner, 2017)—who have recently addressed this topic—in bringing back spatial modeling in marketing.

3.2 Background

3.2.1 Spatial modeling in marketing

One core assumption of traditional marketing research is that one individual behaves independently from another individual (Bradlow et al., 2005). However, when we deal with people's behavior in space, this assumption usually does not hold because choices are dependent on the surrounding area in which a decision-maker is located. In fact, the opposite is the case, and we assume a correlation between the choices and behavior of individuals who live close to each other or are positioned near one another at a particular time (Bradlow et al., 2005; Hunneman et al., 2022; Kossinets & Watts, 2009; Noë et al., 2016). Geographic circumstances, climate, lifestyle, demographics, and other components can influence a consumer's taste and decision-making process (Jank & Kannan, 2005; Mittal et al., 2004). Therefore, researchers started to consider spatial econometric models for marketing modeling tasks (Bradlow et al., 2005; Bronnenberg, 2005; Elhorst, 2017; Hartmann et al., 2008). Early applications of such models dealt with developing a deeper understanding of new product adoption and growth. Further, they have been used as a method for examining the physical boundaries of geographic markets or for inferring information about certain phenomena when data sources are sparse (Bronnenberg, 2005). However, marketing scholars stretch the term *space* beyond its original meaning, including not only geographic maps but also socioeconomic and demographic data layers that allow for the identification of social networks and special-purpose areas.

Choosing the right spatial model is important because such maps of a non-geographical nature can uncover latent information and relationships (Bradlow et al., 2005). For example, retailers use spatial models to determine new pricing strategies, taking into account other retailers in the neighborhood (Bronnenberg & Mahajan, 2001). Hunneman et al. (2022) extend this finding by using a spatial model to better predict store sales at the store level and at the zip code level. For such research problems, spatial models are powerful; they allow for the weighing of certain influences based on how strong their presence is in a particular research area (Mittal et al., 2004). Additionally, Del Gatto and Mastinu (2018) apply the Huff model to retail data from Italy and are able to explain store visits from certain neighborhoods to retail areas, but they cannot model how a change in the retail landscape influences consumer spending. Other use cases of spatial modeling in marketing include the ideal placement of ATMs to serve as many customers as possible within the target radius and the modeling of product market entries across regions to optimize advertising placement (Bronnenberg et al., 2005). Finally, spatial models can be used to study the competitive pricing of products based on local market conditions, production facility locations, and manufacturers' delivery costs (Anderson & De Palma, 1988; Bronnenberg, 2005).

Despite the thorough examination of the use of spatial models in marketing, they have been mostly ignored in the location popularity literature. Hunneman et al. (2022) and Del Gatto and Mastinu (2018) have published the only two recent papers on spatial modeling in marketing. In the early 2000s, spatial modeling seemed to be of interest, but it was soon neglected and has not experienced a revival, even though one would think that GPS data and all the associated opportunities would spark new interest. Granted, most marketing science literature uses spatial models to understand consumer choices (Bell & Song, 2004; Yang & Allenby, 2003), but little attention is paid to leveraging geographic knowledge to identify the influences of the environment on consumers (Wood & Reynolds, 2012). To fill in this research gap, spatial models in marketing must include infrastructure features as a supplement to demographic and socioeconomic information when modeling attractiveness in urban settings.

3.2.2 Urban hotspots

In urban planning, we often call popular locations "hotspots." Urban hotspots are areas that have higher volumes of people inflows and flourishing businesses than neighboring areas during a certain time period (Li et al., 2021). Because people inflow is used as the determinant for popularity, the physical infrastructure consisting of roads, railways, and other public transport as well as buildings and their purposes play a crucial role in urban hotspot detection. For example, places where various modes of transport intersect and where changing modes within an acceptable timeframe is possible are natural hotspots. Similarly, buildings such as supermarkets and train stations are popular, frequently visited places in urban areas (Yang et al., 2017). Hotspot analysis can play an important role in the commercial layout of a city. While public authorities can learn what attracts consumers to vital areas and use that information to revitalize dying city centers, private stakeholders can examine their surrounding rivals, examine their complementary business partners, and identify locations where their stores will attract a certain number of clients. Alternatively, researchers in other fields can examine other types of hotspots, such as crime (Kalinic & Krisp, 2018), environmental (Visner et al., 2021), and epidemic hotspots (Shariati et al., 2020).

In GIS research, hotspots have been studied from a methodology perspective, including the identification of urban hotspots supported by mobile phone and GPS data (Cai et al., 2018; Gonzalez et al., 2008; Kang et al., 2010; Sagl et al., 2012). Several advanced techniques are used for hotspot analysis, such as kernel density, line density, and hotspot similarity (Cai et al., 2018). A conventional method to study urban traveling hotspots is spatial autocorrelations, such as Moran's I and Getis-Ord Gi* analysis, which identify hot or cold urban areas based on quantitative statistical tests (Visner et al., 2021). Cai et al. (2018) investigate the spatiotemporal distribution of urban hotspots and their similarity by using kernel density methods. The geographical space that exhibits the largest distribution of trajectory points represents the space with the highest kernel density value. The analytical results present the six main hotspot categories in a color-encoded city map. Comparably, Li et al. (2012) examine the spatiotemporal dynamics of urban hotspots by using a two-phase clustering method. The results show that the overall spatial distribution of the hotspots is relatively similar in different periods, and most of the hotspots are located in the outer ring of the urban city. The particularity of these results varies based on region and can be generalized whenever there are similar geographic and infrastructure characteristics (Li et al., 2021). Research has found similarities and symmetries of human mobility in test areas, indicating that different urban areas share similar mobility patterns (Sagl et al., 2012).

Now we know from the urban planning and geography literature how we can detect and measure hotspots and where they would be located in some studied areas. However, we do not yet know why certain spots are hotspots or what characterizes a hotspot. Therefore, the following section identifies the features that contribute to location attractiveness.

3.2.3 Determinants of location popularity

The general transportation infrastructure, the public transport system, economic viability, and sociodemographic characteristics cannot be examined individually because they largely impact location popularity in a collective manner (LeSage & Pace, 2009). Economic viability includes the diversity of shops, facilities, and other amenities available in an area, and contributes to the popularity and growth of cities (Clark, 2003; Clark et al., 2002). Due to the many influential factors involved, producing a precise description of location attractiveness is a complex task. In the existing literature in different fields, we identified six factors that influence location popularity. Most studies stem from economic geography and related fields; hence, they measure popularity through house prices in a neighborhood. In general, it has been established that people are willing to pay a premium to live in certain locations; consequently, the main question is which neighborhood characteristics generate such a premium (Öner, 2017). The following six factors have been identified and are described in the next paragraphs: landscape, density, access to public transport, geographic proximity to the city center, access to retail, and socioeconomic structures.

First, people value landscape features such as parks, greenery, rivers, lakes, (urban) forests, farmland, and other open spaces (Anderson & West, 2006; Gibbons et al., 2014; Tyrväinen & Miettinen, 2000). Depending on how rural or urban a particular neighborhood is, these natural amenities are present to a greater or lesser degree. Further, individuals have their own preferences regarding urban and rural living; some will pay a high price for a city apartment, while others will pay a similar price for a rural home. Research has found that access to a supermarket within the same neighborhood is important to people in cities, whereas such access does not influence place attractiveness in rural areas (Öner, 2017).

Second, a place's density in terms of building count and population affects location attractiveness. People associate density with two benefits: (1) access to a variety of essential and non-essential product and service providers; (2) better job market opportunities with higher wages and increased chances of finding jobs that match their profiles. This combination of benefits makes density a crucial factor for location popularity (Andersson et al., 2014; Larsson, 2014).

Third, related to density, accessibility to public transportation within walking distance and the presence of other transport infrastructure features, such as vehicle sharing (stations) or bike paths (Mejia-Dorantes et al., 2012) is attractive to most people. Naturally, denser areas have larger, more highly frequented public transport networks. However, residents also care about transportation costs and compare public transport with individual motorized transport when making their transport mode choice.

Fourth, in line with density and transport availability, location popularity is influenced by geographic proximity to urban centers and to the relevant amenities of everyday life. Proximity is closely tied not only to travel distance and time but also to accessibility when an individual does not have all common transportation options available to them (Öner, 2017).

Fifth, access to retail and the total number of consumption possibilities are determining factors of place attractiveness. In this case, we talk about retail and consumption in the broadest sense, including such things as cultural experiences, leisure activities, public and private services, gastronomy, tourism activities, and shops. A diverse set of shops in an area is considered a pull factor for consumers—it increases the area's popularity (Hunneman et al., 2022; Öner, 2017). Competition and co-location are also crucial from a business perspective. While stores with profound interaction (e.g., clothing stores and shoe stores, doctors' offices and pharmacies) should co-locate, shops selling fast moving consumer goods (FMCG) and perishables (e.g., supermarkets, bakeries, butcher shops) or durables (e.g., hardware stores) should spread out (Larsson & Öner, 2014; Mao et al., 2019).

Finally, the socioeconomic structure of the neighborhoods and the current geographic positioning of consumers play an important roles in predicting location popularity and consumer choices (Yang & Allenby, 2003). People surround themselves with like-minded people who have similar preferences and views on popular locations (Noë et al., 2016). For businesses, particularly small ones, this is important; there must be a good match between the business category and the demand in the surrounding area (Dong et al., 2018; Karamshuk et al., 2013; Larsson & Öner, 2014).

Based on these studies from multiple fields, which examine spatial models and place attractiveness from different perspectives, this chapter explores how we can use spatial modeling to determine the extent to which infrastructure features influence location popularity and what roles other factors play. To date, research has measured location attractiveness through house prices, but we innovatively use GPS data to measure location attractiveness through the inflow and outflow of people. In light of the revitalization of urban centers, this seems to be a more suitable and promising proxy for popularity. Studies that examine a single infrastructure feature are outdated. Therefore, we use this chapter to examine infrastructure influence as a whole and contribute to closing the knowledge gap on how geographic information can be leveraged for marketing. Finally, spatial modeling has been dormant since before the dawn of ubiquitous devices and the urban sensing era. Thus, we aim to assist fellow researchers in bringing about the renaissance of spatial modeling in marketing.

3.3 Methodology

3.3.1 Conceptual model and description of variables

Based on the previous research discussed, we hypothesize that the infrastructural circumstances at a destination influence its popularity and that this relationship is moderated by the circumstances at the origin of a trip. In particular, we suggest that the following aspects of a destination impact whether it is attractive to consumers and citizens: density, dominant function, competition between businesses of the same category, and availability of public transport. Additionally, we assume that the same set of circumstances at the origin of an individual's movement moderates this relationship. Figure 3.1 below summarizes the hypothesized relationships.

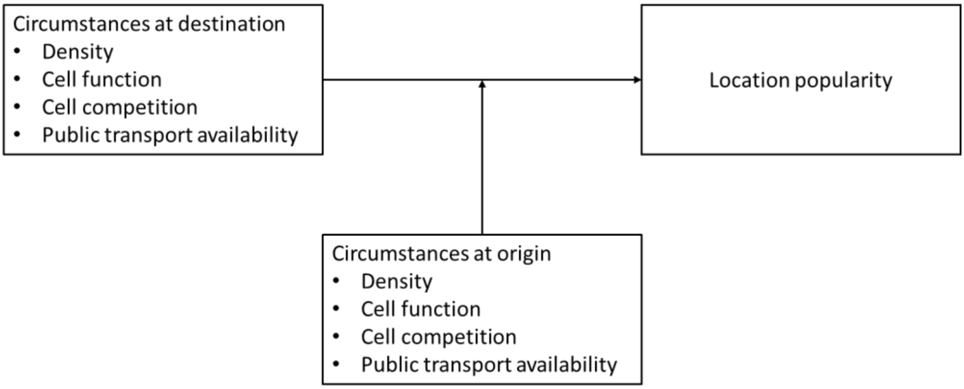


Figure 3.1: Conceptual framework of this chapter

We use the two datasets on OD trajectories and built environment introduced in Chapter 1 as the foundation of this analysis, enriching them with additional open source-data about public transportation from OSM. In line with the logic of the research question and framework, the dataset is set up to represent the 4,296 recorded movements for which the start and end points are anonymized on a hexagonal grid with cell IDs. Therefore, each observation contains the ID number of the start and end point, popularity, density, cell function, cell competition as well as public transport availability at origin and destination. The following paragraphs expand on these variables, explaining the reasons for their inclusion in the model as well as the calculation of variables not included in the raw dataset.

Cell popularity—This study aims to determine the impact of infrastructure features on location popularity. In the urban planning literature, the dependent variable location attractiveness is typically measured through house prices in an area (Öner, 2017). However, in the context of this research, we decided on a more suitable indicator of popularity. We use the inflow numbers in a grid cell divided by the total number of movements across the observation area during the entire observation period of three weeks. We then multiply the resulting fraction by 100 for better interpretability of the results.

Density—Dense urban areas are associated with immediate access to all the utilities of everyday life; thus, we hypothesize that densely built areas are more popular than sparsely built areas (Andersson et al., 2014; Larsson, 2014). In our study, we use the absolute count of buildings in a respective grid cell as the density variable. To this end, we use the built environment dataset, which lists the count of buildings per category, and take the sum across all categories as the indicator of building density.

Cell function—Every human movement has a purpose; thus, the functionality of the built environment can provide researchers with an indication of what a person is likeliest to be doing at their destination. According to Chen et al. (2020), the function of the region greatly affects its attractiveness; therefore, it makes sense to consider the functionality of buildings in a destination cell. For the variable of cell function, we compare all nine categories of the built environment dataset; whichever category indicates the highest count is established as the function of this cell.

Cell competition—As mentioned above, having competitors nearby can be simultaneously advantageous and disadvantageous; this is partially dependent on the nature of the business (Larsson & Öner, 2014; Mao et al., 2019). Consequently, we want to consider this factor in our model. To compute the variable, we exclude the residential function because it is irrelevant to this variable. Among the remaining functions, we search for the most dominant function in a grid cell—the one that presents a larger count than all other functions—and we establish this function as the competition in the respective cell.

Public transport availability—Access to public transport within walking distance and the presence of other transport infrastructure features such as vehicle sharing (stations) or bike paths positively influence location attractiveness (Mejia-Dorantes et al., 2012). Therefore, we used OSM to download information on where public transport stations (i.e., bus stops and train stations) are located in our study area. Via the latitude and longitude of each point, we are able to determine how many public transport options are available in each grid cell, and we integrate these figures as a variable in our model.

3.3.2 Research method

The dataset we compiled for this analysis is based on the 4,296 trips we recorded in South Limburg. Because every movement has an origin and a destination, the dataset is dyadic—the origin and destination are not independent from one another. Further, from the data collection procedure we know the data is technically hierarchical, however it does not include an identifier that allows connecting multiple trips to the same individual. Initially, we used a mixed effects model to test the hypothesized relationships because it seemed to be a suitable technique. Mixed effects models are used for data that allow for an outcome to be measured more than once on the same person. In our case, each tracked person recorded movements whose features we assess in our analysis and since movements are nested within individuals, it is possible to observe the same movement done by the same person twice. However, the results of this mixed effects model showed a variance for the random term of zero, meaning that the model did not pick up any random effect in the data. Consequently, research suggests switching to a simple linear regression model.

The linear model analyzes the popularity of the destination location; thus, the predictors in the model represent the circumstances at the destination (i.e. density, cell function, cell composition, and public transport availability). However, it is important to not neglect the dyadic nature of the data structure or ignore the origin circumstances. It has been shown that there is an interdependency between the infrastructure features at the origin and destination of a trip. This becomes especially evident in the example of transport availability. An individual can only opt for this solution if public transport is available at the origin and destination of a movement. Similarly, if we consider cell function as a proxy for trip purpose, it is unlikely that many movements will take place between cells of the same function: Why would a person go somewhere else to fulfill a need that is also available at their current location? Consequently, we examine the extent to which the circumstances at the origin moderate the effect that the circumstances at the destination have on location attractiveness.

3.4 Empirical results

Using the dataset and variables described above, we run a linear regression model that is significant overall (F-statistic = 70.04, $p < 0.000$) and has an adjusted R-squared value of 0.6165, which means that the model explains 61.65% of the variance in the dependent variable. As depicted in the model graph above (Figure 3.1), our model includes direct independent variables and interaction variables. The next section elaborates on the direct effects found in the analysis, followed by the significant interaction effects.

3.4.1 Direct effects

In line with our expectations, we find that denser grid cells are more popular than those with less densely built environments. The coefficient for the density of the destination is 0.001142 ($p < 0.000$), which means that with every additional building constructed in a cell, the popularity of that cell increases by 0.001142. This makes sense because a greater number of buildings either increases the variety of available building categories or increases the options to choose from within one building category; thus, denser areas provide more possibilities to consumers, which leads to a higher location attractiveness score.

The functionality and competition in a cell are measured in the categories of the most dominant function; thus, the impact of the variables is given in relation to a baseline function. We choose residential buildings as a baseline for the cell function for three reasons: most buildings fall into this category, every individual returns to their home at least once a day, and most homes are located in cells categorized as residential. In our analysis of the direct effect of cell function on location popularity, we find that office and education cells are more popular than residential cells. When a cell is categorized as office or education, the location attractiveness score is 0.8708 ($p = 0.003$) or 1.836 ($p < 0.000$) units higher, respectively, than when it is residential. A possible explanation for this could be that offices and educational facilities serve many people, while residential buildings serve only one family or—in the case of an apartment building—a few residents. Further, people tend to come from their homes to offices and educational institutions, and people's homes are spread across all the neighborhoods in the region. Thus, because residents from the entire region come together in offices and educational institutions, these places attract more people than neighborhoods do.

Aside from these positive effects, we find that industry and shopping cells are less popular than residential cells. When a grid cell is categorized with the residential cell function, the popularity score is 1.082 ($p < 0.000$) or 1.658 ($p < 0.000$) units higher than when it is categorized as industry or shopping, respectively. Regarding the negative coefficient between shopping and residential, we argue that people go home at the end of each day, whereas shopping is an infrequent activity so these grid cells are visited less often.

To examine the direct effect of cell competition on location popularity, we chose social cells as the baseline because it is one of the functions in which businesses tend to locate themselves close to the competition to increase choice for consumers in a single location. Our regression analysis indicates that when competition is high among industry buildings and shops, a cell is more popular than when competition is high among social activities and restaurants. In particular, location popularity increases by 0.7075 ($p < 0.000$) units when industry buildings compete, and it increases by 1.101 ($p < 0.000$) units when shopping facilities compete for customers. Industry and shopping facilities are both functions that tend to build conglomerates of similar businesses. Most towns and cities have industry parks on the outskirts of the city where many larger companies and production plants are located next to each other. Likewise, most city centers feature myriad shops so that customers can visit multiple brands and stores in the same area, compare and evaluate choices, and enjoy a nice shopping experience.

Finally, contrary to our expectations, the direct effect of public transport availability decreases the location popularity score. More precisely, every additional public transport station decreases location attractiveness by 0.2388 ($p < 0.000$) units. We do not have an explanation for this counterintuitive finding. Nevertheless, the interaction effects of cell function and public transport availability show the opposite, in line with what we hypothesized. Therefore, we will elaborate on these findings instead of the direct effect.

3.4.2 Interaction effects

Given the data at hand, we find two main interaction effects: one between cell function and public transport availability and one between origin cell function and destination cell function. For the interaction effect of destination cell function and public transport availability, we find a positive and significant effect for almost all cell functions compared to the baseline residential cell function. In particular, the coefficients indicate that the location popularity score increases by 0.5007 ($p < 0.000$) units with every additional public transport stop added to the grid cells when the cell function is social compared to residential. It increases by 0.1179 ($p = 0.003$) units for health, 0.3141 ($p < 0.000$) units for industry, 0.4962 ($p < 0.000$) units for office, and 0.275 ($p < 0.000$) units for shopping (Figure 3.2). We expected to find this relationship because previous research has established that denser areas have more public transport stops. Additionally, social facilities such as restaurants and leisure activities as well as shops are usually located in dense areas, and most municipalities and regional governments are trying to make commuting to work by public transport convenient and attractive. Consequently, it is reassuring for governments that we find a positive interaction term for industry and offices, because that means their efforts are worthwhile. A more extensive public transport network within these cells makes the area more attractive for employees and potentially for employers as well.

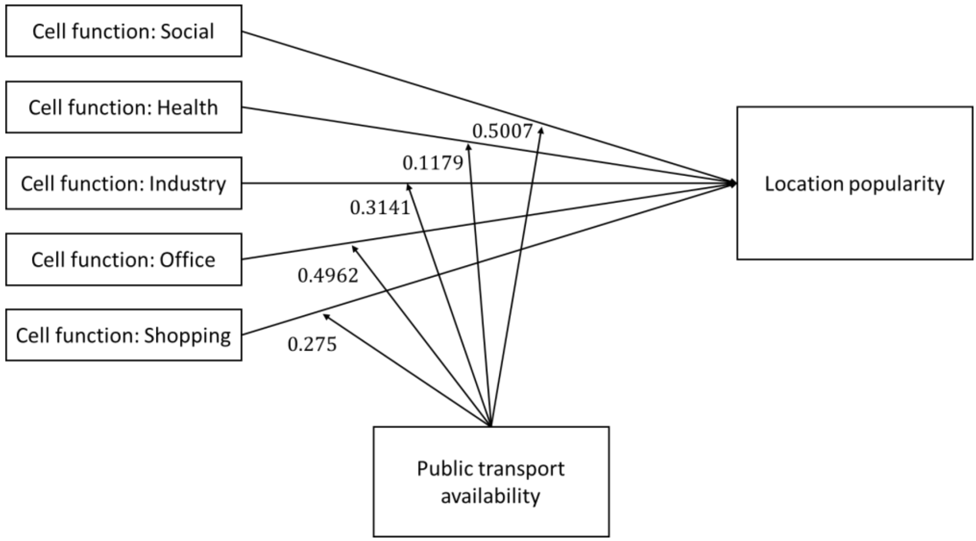


Figure 3.2: Significant interaction effects of public transport availability (baseline cell function: residential)

The second group of significant interactive effects refers to how the cell function of a trip's origin moderates the impact of a trip's destination cell function on location popularity. In the results of our regression, we find a variety of significant OD cell function combinations, and we will elaborate on some of these structures. A complete overview of all significant interaction effects including their effect size can be found in Figure 3.3 below. For a trip ending in a social cell, we find that the effect on the popularity is different with respect to residential if the trip starts in a social cell, an office cell, a sports cell, or a shopping cell. More precisely, the slopes of the regression lines between popularity and destination functions are steeper for social ($\beta = 0.5447$, $p = 0.014$), sports ($\beta = 1.891$, $p < 0.000$), and shopping starting cells ($\beta = 0.9391$, $p < 0.000$); the slope is flatter for office starting cells ($\beta = -0.5898$, $p = 0.008$) compared to residential cells. In other words, the effect of a social destination on popularity is 1.891 units higher compared to a residential destination whenever the trip started in a sports cell. In practice, this means that people are more likely to go to a bar or restaurant with their training partner after a workout instead of going home immediately. The same logic applies for shopping, meaning that after shopping people tend to visit a restaurant or another leisure activity before going home. As an additional example, we find that the effect on the popularity of a trip ending in an office cell is different compared to residential if a trip starts in a social cell ($\beta = -0.4668$, $p = 0.033$) or health cell ($\beta = -1.219$, $p = 0.002$). In practice, we can conclude that if people were in a social or health cell before, they are less likely to go to an office cell next instead of a residential cell. In other words, if you are sick and just saw a doctor, you are going home and not to work. Similarly, after dinner and drinks with friends, people also go home and not to work. In fact, the negative coefficient here implies that people go from their office to social cells.

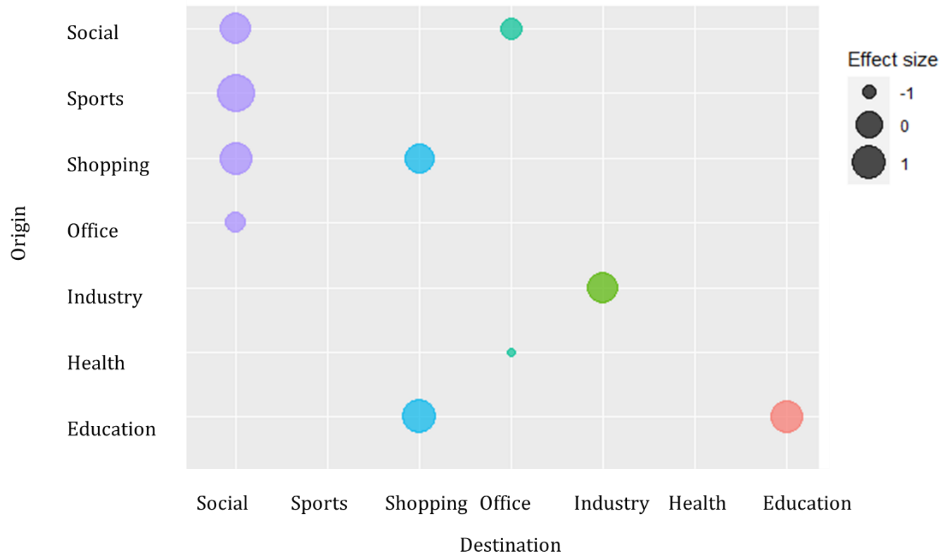


Figure 3.3: Interaction effects between origin cell and destination cell functions

Overall, we find support for all four direct effects, three of which are in line with our expectations. We examine the interaction effect of the four factors and determine the extent to which the respective circumstances at the trip origin influence the effect of the four circumstances on location popularity. We identify a moderating effect on the relationship between the destination cell function and popularity based on the origin cell as well as a moderating effect between public transport availability and cell function.

3.5 Conclusion

Location attractiveness depends on many factors. In this study, we aimed to examine the direct and moderating effects of four of these factors. This section highlights the most important implications of our findings and elaborates in greater depth where the significant interaction effects potentially stem from and how we can use them for marketing strategies.

One important takeaway from this study is that public transport access is important for location attractiveness and that the importance varies with cell function. This means that for some business categories, the availability of public transport does not influence people inflow and location attractiveness, whereas other business categories can increase the popularity of their businesses and the area they have settled in by providing access to more transport solutions. For example, a health cluster (and all the businesses in the area) benefits from additional public transport stations because they give more people access to their services and increase potential customer reach. Similarly, industry and office cells naturally are areas with large people inflows because employees travel to their workplaces from across the region. A smooth and well-functioning transport system is crucial to preventing traffic, and public transport can be one contributor to smooth traffic. Hence, municipalities are working hard to increase public transport availability in industry and office areas. Our results confirm that their efforts are worthwhile and that additional bus stops and train stations increase attractiveness. This increase in attractiveness holds for the citizens working in these places as well as for companies that own land and facilities in these areas, or those planning to settle there and want to give their employees as many travel options as possible.

Moreover, for the interpretability of the interaction effect between the functions of origin and destination cells, we consider the destination cell function as a proxy for the trip purpose. Every movement has a trip purpose, or the activity that the individual pursues once they arrive at their destination. If we look at origin and destination cell functions from this perspective, the interpretations and implications become more concrete and easier to grasp. Among the significant effects reported in the previous section, we can identify three broader categories of cell function interaction effects; these three categories allow marketing implications to be derived. First, trips starting in social, sports, or shopping cells and ending in social cells as well as trips starting in education or shopping cells and ending in shopping cells are most likely so-called chain trips. Here, consumers attend multiple activities, fulfill multiple needs, or combine multiple purposes into one route with several stops instead of going home between each activity. For example, after going to the gym or after a long shopping trip with friends, people go to a social area to have drinks or dinner together. Likewise, after school or university, students go to the central shopping area of the city with their friends to spend their free time.

We observe significant effects for trips between two shopping cells and two social cells. These movements also make intuitive sense because a person out on a shopping trip is most likely visiting multiple shops that span an area larger than one grid cell. The same goes for combining social activities with dinner or drinks in the city. With this phenomenon of chain trips in mind, various stakeholders can derive implications for themselves. New businesses looking for suitable store locations should ensure that the right business categories are located in close proximity. Established businesses should make sure their shops are attractive to chain-trip customers by ensuring that there is a large parking lot or a bus stop nearby, working with neighboring shops to enable customers to combine stops and decrease total trip time, etc. Chain trips are crucial to consider for transport providers when planning infrastructure or public transport routes to ensure that the service matches customer needs optimally. For urban planners and city authorities, similar implications apply, but they are responsible for the overall city infrastructure and for ensuring that the offered services as a whole sufficiently fulfill citizens' needs.

The second category of cell function interaction effects describes trips starting and ending in industry cells and trips starting and ending in education cells. We assume that these trips are rather short and stay within the same cell or end in a neighboring cell of the same dominant function category. For businesses, this is fascinating to observe, because that means employees at their workplaces or students and staff of educational institutions leave during the day to fulfill certain needs nearby. This could include going to a restaurant or takeout place to buy lunch, buying cigarettes or food at a supermarket, or going to the hairdresser or physical therapist during lunch break, for example.

Third, we find negative interaction terms for going from social or health cells to the office. This is reasonable, because one would assume that the movements take place the other way around. People leave their homes (reminder: residential is the baseline here) in the morning and go straight to work. On their way back, they might immediately go to dinner with friends or a doctor's appointment but not the other way around. For leisure and health services, this means that it is worth considering placing a business so that it is convenient to reach from office areas where the majority of the people flow originates.

More broadly, we can conclude from this research that we can use people inflow to a grid cell as an indicator of location attractiveness, especially in a marketing and behavioral context. We consider people flow to be a better proxy than house prices for behavioral research because it directly represents how many people actually visit a specific location or neighborhood and, depending on the data richness, how long they stay and by which mode of transport they got there. Considering that this is an early attempt to model the influence of infrastructure on location popularity with spatial data and with people inflow as a popularity indicator, the factors we included in the model explain a satisfying amount of variation.

However, the results indicated that roughly 40% of the variation is still unexplained, and more research needs to be done. Additional infrastructure variables that could be included are greenery such as parks and urban forests or transport infrastructure options such as vehicle sharing stations or bike paths. The constellation of buildings and their functionality are interesting to study further, and other related context variables of individuals' movements should be examined for additional insights. For example, the purpose of the trip and the time available for a certain activity impact location attractiveness in a given situation. Whenever people are out for leisure shopping and have multiple hours to go from shop to shop to find the ideal product, they prefer areas with many shops and are less sensitive to difficult parking situations and high parking fees, which we usually find in a city center. In contrast, when people are on a tight schedule and need to buy one specific item for which they already have a suitable shop in mind, they might prefer to go to a place less crowded than a city center because that will save time.

Further, there are certain activities and duties for which people simply do not have a choice of where to go, but it should still be the goal of urban planners to make the areas as attractive as possible. More concretely, a person's work, school, or university location is a fixed place that they usually cannot choose, but employers and employees still appreciate an attractive environment. Likewise, most towns and medium-sized cities have only one hospital and a few specialized doctors per discipline. This does not mean that the hospital surroundings do not need to be attractive to citizens.

As well as additional variables, a larger dataset with more movements would lead to more accurate popularity measurements. We have established in this study that people inflow works well as a location popularity indicator, but the more movements and inflows into grid cells we have available in the dataset, the more robust the popularity index will be. In addition, methods for collecting new data should find ways to observe true trip purposes, such as using an anonymized personal identifier to connect several movements to one person or determining their home and work location grid cells. This would allow for deeper insights and the confirmation of derived implications, such as the chain trip when movements between certain cell functions take place.

3.6 Impact on hyper-targeting

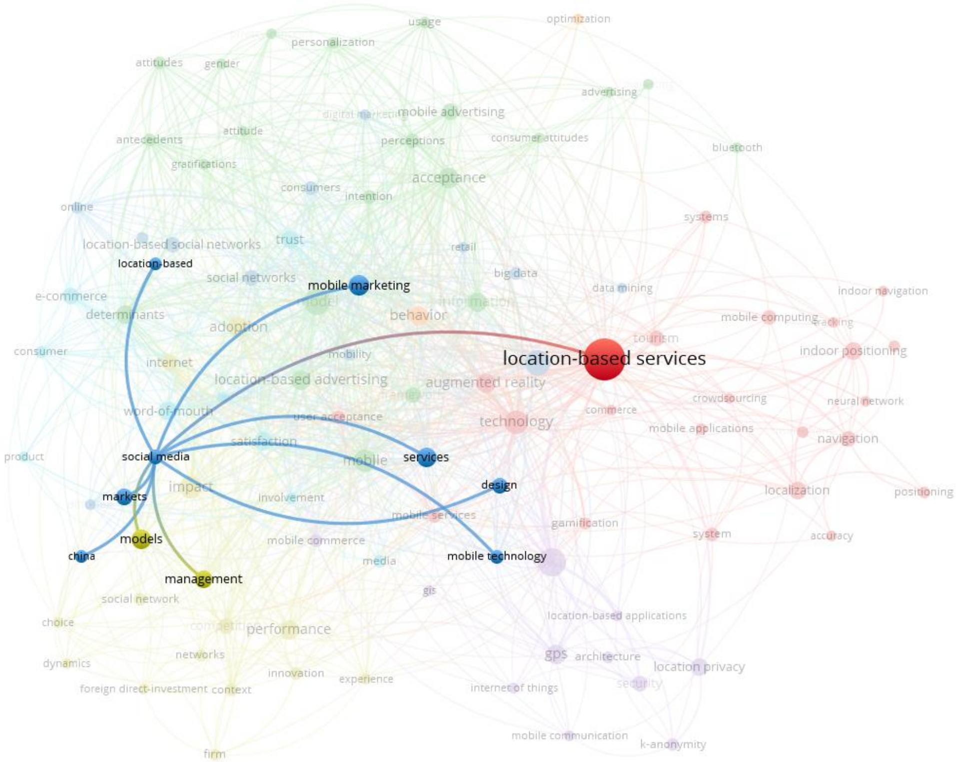
As in the previous chapters, the last section reflects on the impact of this study on hyper-targeting. The factors that drive location popularity indicate a business' choice of whom to partner with. Competition and cooperation between businesses are omnipresent, but once aware of the competing and complementary businesses nearby, shop owners can turn their presence into an advantage. In particular, a shop or restaurant can build partnerships with complementary neighboring shops to increase revenues and attract each other's customers. Marketers can learn from this chapter that accessibility and transport infrastructure are vital to location attractiveness and consequently to business success. Companies can take initiative and approach their municipalities in case they see a need to improve the transport landscape in their areas. Alternatively, they can negotiate special mobility deals with local public transport providers or parking garage operators. Knowing which locations are popular with your target audience allows marketers to work with geo-fencing and to place mobile advertisements. Finally, for super-specific, curated marketing offers, more is more when it comes to information about the target audience. The more we can explain consumer behavior (e.g., by adding spatial data, offline location information, or mobility preferences and attractiveness), the better and more successful the hyper-targeting strategy.

Aside from the content impact on CBM, this chapter extends the building interconnectivity analysis from Chapter 2 with more variables than just building location and positioning. We discuss that the mere location and positioning of buildings to each other is more insightful when we add their functionality to the equation. Furthermore, we emphasize that static and dynamic spatial data are equally important and complementary to spatial analysis and have many use cases in addition to marketing and behavioral research. With the realization that the dynamics of consumer movements are quite insightful for marketers, the following chapter will focus more closely on the observed movements.

Transition 3: It's time to add the consumer … again, but this time dynamically

So far, we have mainly examined building networks or infrastructure features to sketch the city and to understand the impact of the built environment on citizen behavior, consumer transport experience, and location attractiveness. Further, we considered the impact of the factors impacted by the built environment —citizen behavior, consumer transport experience, and location attractiveness— on the future of hyper-targeting. In these analyses, we used the movements of citizens only as a side component, and focused on static spatial information about the city. Now, it is time to move the citizens into focus and to examine the observed human movements in more detail. Aside from the origin and destination cells, a time stamp and the mode of transport that the individual used for a particular trip describe the movements. With the knowledge about the structure of the built environment from Chapters 2 and 3, I now explore how citizens use their city and how they deal with the given physical infrastructure. Because humans are social beings and their decisions and behavior are not independent from the behavior of their peers, we use the following chapter and add geo-social networks as a new component that influences LBM and CBM. Geo-social networks are online or offline networks between people who live or work close to each other or frequently visit the same locations with similar time patterns. They readily allow for the capture of the dynamics in a city and can uncover new indicators for location popularity, business success, and service acceptance and usage.

Such networks and dynamics are complex, and are therefore hard to find and even harder to extract from a datasheet in an understandable format. For that reason, I decided to take a data visualization approach to geo-social networks in cities. I critically assess the power of this technique for questions with complex datasets and multiple perspectives.



Chapter 4

Let the Pictures Talk—A Data Visualization Approach for Revealing Dynamics in Complex Urban Datasets and Movement Structures

I would like to thank Madalina Vrancean for carrying out part of the analysis.

4.1 Introduction

Urban dynamics are complex, and urban planners need to take many different factors and perspectives into account when making decisions. Therefore, decision makers in urban planning and city policy often turn to data analysts when dealing with data-driven evidence for a scenario and ask for advice. It is the analysts' job to help decision makers find their way through the data jungle and make informed decisions. Depending on the issue or question at hand, data analysts combine various data sources and use a large portfolio of analysis techniques. Often, data analysts present the results to practical questions in enormous tables of numbers, complex spreadsheets, or very shallow managerial reports. Both extremes are inconvenient for thoughtful decision-making, and in recent years, a new approach has attracted attention. Data visualization aims to make complex datasets and scenarios more understandable and accessible to broader audiences through the careful selection of graphics.

Human mobility and urban dynamics are largely interested in how individuals move within geographic space. Thus far, a consumer's location is tracked and analyzed only via various social media platforms, such as Foursquare check-ins or location tagging on Facebook and Instagram. This information is collected over a long period of time to draw pictures of an individual's movement patterns, of the network structures of frequently visited locations, and of people who interact with each other or follow similar patterns (Bernritter et al., 2021; Ghose, Li, et al., 2019; Gonzalez et al., 2008). Further, these platforms use GPS for on-the-spot mobile marketing. The aim of this chapter is to understand offline geo-social network spaces through a clever combination of datasets and visualizations of many different features.

The theoretical understanding of offline social networks is built around the understanding that similar people form ties with the same people and tend to like similar things (Kossinets & Watts, 2009; McPherson et al., 2001). Various studies observe that people who meet frequently have closer social ties than people who meet less frequently (Leng et al., 2018). Likewise, people with similar personalities and interests visit the same locations (Celik & Dokuz, 2017; Dokuz & Celik, 2017; Noë et al., 2016) and enjoy the same activities (Carrasco et al., 2008). These findings support and deepen this theory. However, "homophily limits people's social worlds" (McPherson et al., 2001); people who, for example, live in different neighborhoods but work in the same area can act as social bridges between groups or communities (Dong et al., 2018).

Brown et al. (2014) extend this notion and study offline social networks in combination with visited places. They detect that groups visit known places, while individuals explore new places only with one other person and tend to choose new places that are familiar to others in their social networks. Further, Coscia and Hausmann (2015) study networks of people based on calls between individuals and compare the resulting social networks with the mobility networks of individuals captured via the GSM of their respective mobile phones. This comparison allows us to conclude that call-based social networks and movement networks are isomorphic, meaning that people who interact frequently also show similar movement patterns within their cities.

An essential role for strategic planners is to understand how a city's characteristics connect to its inhabitants' movement patterns and, consequently, to understand how individuals plan trips on regional and city levels. Conventional urban analysis can embed complex analytical tasks that require strong expertise and in-depth knowledge. Hence, it is essential to explore the data in a flexible, creative manner and include visual human perceptions for meaningful interpretations. Visual analytics is the technique that translates data and information into interactive visual representations and assists decision-making and knowledge building for different application domains (Thomas & Cook, 2006).

Although the importance of data visualization is recognized, the difficulty in discovering these urban patterns lies in their variability and dispersion. It challenges the strategic planner to understand the urban mobility phenomenon comprehensively. Developing an integrated visual system that leverages individuals' trajectories, service distributions, and density patterns would allow analysts to explore more facets of the same phenomenon. To address this issue, this research takes a holistic visualization approach to analyze geo-social patterns and assist public authorities in making better strategic decisions.

4.2 Background: The data visualization landscape for urban mobility

Urban dynamics, also known as movement analysis, is defined by variations in individuals' movements between locations and across time periods (Sagl et al., 2012). It is one of the most researched and challenging topics in data visualization (Andrienko & Andrienko, 2013). Such research uses geospatial data (expressed in geographic coordinates); attribute information about the object, movement, or location; and temporal information about the movement. On a more general level, we can further classify this type of data as static or dynamic. Static data represent stationary data points without time components, such as the built environment or physical infrastructure of an area (roads, railways, stations, bus stops, etc.). Dynamic data embed data points that change with respect to location and time. We classify trajectory data as dynamic because their main components are the origin and destination locations of an individual's movement at different points in time. Depending on the richness of the data, trajectory data can incorporate additional variables and points of research such as mode of transport, recurrence of trip patterns, or socio-demographic attributes (Stock & Guesgen, 2016). The collection and availability of large amounts of data recorded via mobile phone records, GPS, or social media allow researchers to study more people across a larger geographical space and in greater depth, but with limited investment in data collection efforts. Methods to visualize such data are diverse, and this paper chooses trajectory analysis, hotspot analysis, and mobility in context.

4.2.1 Trajectory analysis

Trajectories of moving objects are considered some of the most complex data for evaluation (Andrienko et al., 2017). Trajectory analysis, often referred to as OD flow analysis, encompasses various techniques and visualization methods (Demissie et al., 2013; Guo et al., 2011; Liu & Ratti, 2009; Ma et al., 2015; Pérez-Messina & Graells-Garrido, 2019; Von Landesberger et al., 2015). The most common type of movement visualization is a linear map, where linear symbols represent object trajectories. Further, we classify trajectory patterns into quasi-continuous data and episodic data. Quasi-continuous data contain features such as speed, acceleration, or direction

(Andrienko et al., 2017). Episodic data include trajectories that are separated by large time gaps or intermediate locations such that they cannot be accurately constructed as chain trips or multimodal trips. Trajectories are often aggregated into matrices or flows, such that each matrix cell or flow reflects all travels from a given starting point to a given destination. Aggregation techniques are used to ensure the data privacy of the tracked objects. Additionally, matrix views capture only direct OD connections rather than the routing on a map; the same holds for episodic data.

Due to the large gaps between points, each trajectory segment is seen as an OD move, which facilitates the formation of visual clutters. To reduce the clutter, the literature stresses the use of clustering techniques or spatial simplification by grouping origin and destination locations into larger regions and aggregating them into the same flow. One such technique is used in the work of Graser et al. (2019) and is known as edge bundling. This technique is designed to visually bundle similar edges and reduce visual clutter within a graph. This novel approach determines the bundle strength to produce faster and more accurate node grouping (Graser et al., 2019). Similarly, Von Landesberger et al. (2015) spatially simplify mobility flow movements by clustering time steps based on their spatial similarity. Guo et al. (2011) analyze traffic trajectory data by using arrow-shaped glyphs to represent specific directional movement. Glyphs are visual design objects that can be used independently to show certain data attributes or comprise a set of data records (Borgo et al., 2013). Demissie et al. (2013) use both flow clustering maps and glyphs to explore the mobility patterns of Lisbon, Portugal. Glyphs are commonly applied in information visualization to analyze OD flows (Demissie et al., 2013; Guo et al., 2011; Liu & Ratti, 2009; Ma et al., 2015; Pérez-Messina & Graells-Garrido, 2019).

4.2.2 Urban hotspots

Trajectory data allow for more than flow analysis and routing problems. A wide area of research has focused on identifying urban hotspots with the support of mobile phone and GPS data (Cai et al., 2018; Gonzalez et al., 2008; Kang et al., 2010; Sagl et al., 2012). Urban hotspots are areas with large volumes of people inflow and flourishing businesses during a certain period (Li et al., 2021). Hotspot analysis can provide scientific bases and references for location-based service expansion using real-time data. Moreover, this type of analysis can benefit multiple stakeholders from different areas of expertise and facilitate the decision-making process. For instance, it can serve as a tool for urban planners in allocating and developing sustainable city infrastructure. Furthermore, hotspot analysis can play an important role in the commercial layout of a city, in which business owners can examine where their stores can reach a certain number of clients. Alternatively, researchers in other fields can examine other types of hotspots, such as crime (Kalinic & Krisp, 2018), environmental (Visner et al., 2021), and epidemic hotspots (Shariati et al., 2020).

A conventional method for studying urban traveling hotspots is spatial autocorrelations, such as Moran's I and Getis-Ord Gi* statistics, which identify hot or cold urban areas based on quantitative statistical tests (Anselin, 1995; Visner et al., 2021). Mueller et al. (2015) use Getis-Ord Gi* statistics to evaluate free-floating electric vehicle services and to identify where the hotspots for starting a rental are located in Berlin and Munich, whereas Malleson and Andresen (2015) advocate that the Getis-Ord

Gi* method is sensitive to using a fitting data source and demonstrate this based on the example of crime rate calculations. Sagl et al. (2012) provide a comparison of collective human mobility hotspots between different urban environments by employing the line density method and find that different urban areas share similar mobility patterns. A similar line of validation was found in the work of Gonzalez et al. (2008), which used the probability density function to explore human mobility patterns. The research concludes that trajectories of mobile phone users depict a high degree of temporal and spatial regularity. Thus, regardless of the diversity of urban areas and movement history, individuals follow analogous, reproducible travel patterns.

Cai et al. (2018) investigated the spatiotemporal distribution of urban hotspots and their similarities using kernel density methods. The geographical space that exhibits the largest distribution of trajectory points represents the space with the highest kernel density value. The analytical results report the formation of six main hotspot categories, visualized on a color-encoded city map. They detect surprising findings for the hotspot similarities. Two main hotspot clusters were formed with the same predominant size both for the weekends and for weekdays, which depict a high inflow of people regardless of the period. Li et al. (2021) examined the spatiotemporal dynamics of urban hotspots using a two-phase clustering method. The results showed that the overall spatial distribution of the hotspots is relatively similar during different periods, and most of the hotspots were located in the outer ring of the urban city. The particularity of these results varies from region to region and can be generalized whenever there are similar geographic and infrastructure characteristics. Although spatial clustering has received considerable attention for urban hotspot detection, the lack of spatial context constrains its ability to explain the causes and effects of hotspot development. Hence, the following section aims to connect hotspot analysis with the built environment and time attributes for a more holistic picture of urban mobility.

4.2.3 Mobility in context

Urban mobility literature considers cities as complex systems that incorporate various processes and elements. The rapid advancement of technology not only opened a new spectrum in creating digital traces but also provides information regarding urban buildings (Cecchini et al., 2019). Developing a system for real-time monitoring of urban mobility patterns can measure individuals' daily mobility and evaluate their relationships to socioeconomic and land-use patterns. In addition to the research conducted in hotspot identification, spatial division and morphology are valuable extensions for studying urban mobility patterns. The work of Sevtsuk and Ratti (2010) accounts for differences in mobility patterns by using built environment indicators at varying times. The results suggest that cells with more retailers have lower shares of human inflow during weekdays.

Alternatively, Kang et al. (2012) examine the impact of morphological characteristics of buildings, such as compactness and size, on human mobility inside cities. The general patterns indicate that individuals living in larger cities commonly have to commit to longer travel distances on a daily basis. The opposite applies to smaller cities with compact built environments. Chen et al. (2015) also explore mobility patterns from social media and the corresponding relationships with local land-use characteristics at different periods. The research applies a clustering-based method to identify the temporal distribution of patterns and assigns modes of transport afterward. Li et al. (2018) explore urban dynamics by combining multiple spatiotemporal and network analysis units. The research also considers built environment evaluation to investigate the social relationships between individuals, groups, and places. The study refers to social relationships as the connections among urban objects and as the direct link to human trajectories; similarly, trips carried out to work, friends, or shopping create social connections between individuals and places (McPherson et al., 2001).

In addition to the advantageous contributions of building usage and sociodemographic information, the mode of transport can further increase the predictive power of travel behavior models. Ton et al. (2020) use the latent clustering technique to classify different transport mode patterns. The research concludes that the majority of travelers engage in multimodal transport behavior. A similar modeling technique was applied by Kroesen (2014) to explore the qualitative differences in travel behavior patterns; he also finds predominantly multimodal travel behavior.

4.3 Research method: Visual analytics toolbox in detail

Visual analytics is "the science of analytical reasoning facilitated by interactive visual interfaces" (Thomas & Cook, 2006), which assists decision-making and knowledge building for different application domains. Specifically, computers generate, aggregate, and process data, and humans make associations and generate insights based on these data (Andrienko et al., 2008). Traditional visualization approaches include a direct depiction of each record in a dataset; the analyst extracts and interprets patterns by looking at and interacting with the visualization. However, a considerable volume of visual clutter characterizes large and complex datasets, and as a result, users encounter difficulties in assimilating multiple visual elements that change simultaneously. To reduce the complexity of large datasets, designers and engineers must use visual and statistical techniques to summarize the data (Thomas & Cook, 2006).

Data summarization may involve numerous forms of abstraction, such as aggregation, generalization, and sampling (Andrienko et al., 2008). These types of transformations should be consistent and appropriate for the sample data being investigated. For instance, elementary transformations such as aggregations may be applied to examine a particular set of spatial units or specific temporal unit, or to classify certain attributes of the unit (Andrienko et al., 2017). Instead of directly analyzing the original data, summarization techniques serve as a simplified approach to extract and interpret data patterns. For instance, dot symbols or polygons distributed on the map represent spatial units in geospatial data. Points may embed certain attributes that are visually differentiated through size, color, or symbols. The spatial distribution of moving objects can be represented on a map by symbols or diagrams positioned at various locations across a territory with sizes proportional to the counts of objects. Other characteristics, such as modes of transport, can be color-coded (Andrienko et al., 2017). Researchers also apply distinct visual representations for dynamic data. Trajectory maps, for instance, are commonly illustrated using arrows to depict the directional linkage between OD locations. To account for the potential visual clutter formed by multiple overlapping lines, the counts of the objects that moved, the quantities of commodities carried, or other numerical features of the flows are integrated into the widths of the lines, which are proportional to the magnitude of flows (Andrienko et al., 2017; Stephen & Jenny, 2019).

Density maps are effectively employed for visualizing traffic channels or inflow concentrations in certain urban regions. Contrary to the OD flow maps, density maps do not show the direction of movement. The variety of colors or shades in such a map conveys variations in presence or movement throughout an area. Kernel density estimations are a special function from density map analysis that uses a unique weighting methodology to illustrate a spatial situation. A density field is generated for each time interval, which reflects the distribution of spatial events or motions during that interval. As a result of this aggregation, a time series of spatial circumstances is generated (Andrienko et al., 2017).

Several studies have developed a visual analytics interface that enables a better understanding of human mobility dynamics and integrates multiple visualization techniques for exploring crowd flow volumes, urban trajectories, and spatiotemporal patterns (Chen et al., 2015; Guo et al., 2011; Ma et al., 2015). Chen et al. (2015) provide a good example of a visual system interface for understanding urban mobility using geotagged social media data. The study's main visualization is the aggregated movements of individuals between places. The research engages in highly dynamic filtering and aggregation mechanisms, allowing the user to analyze the different aspects of the movement patterns. The user can extract information such as the number of visits or the average time interval of the movement by clicking on the inner layers of the nodes. Moreover, the capacity to filter, group, or animate certain graphical representations reduces visual clutter, which fosters a more accurate judgment. In this chapter, we apply a selection of the introduced methods and best practices to movement data in the region of South Limburg in the Netherlands (for a detailed description of the dataset, see Chapter 1). The goal is to explore and present a complex dataset in a tangible and understandable manner and to draw conclusions from the data that local authorities can use as an input for future developments and policies.

4.4 Empirical analysis

4.4.1 Trajectory analysis

This study employs force-directed edge bundling to control visual clutter and improve graph legibility. This method consists of aggregating the edges with the same origin and destination ID into one line. The links have varying widths, depending on the number of movements represented by the line (Stephen & Jenny, 2019). Figure 4.1 shows this first visualization of the trajectories. The current visual element has the purpose of helping the user understand which parts of the geographical space are highly frequented. The results indicate a high concentration of inflows and outflows in the Maastricht area, which is the largest city in the region and offers the most activities and special-purpose facilities. We also see a large number of movements from suburban areas south and southeast of Maastricht, which are most likely commuter movements to work or trips for purchases in the city center. Finally, we also detect movements between Maastricht and the other two cities in the region, Heerlen and Sittard, but we do not observe direct movements between these two places.

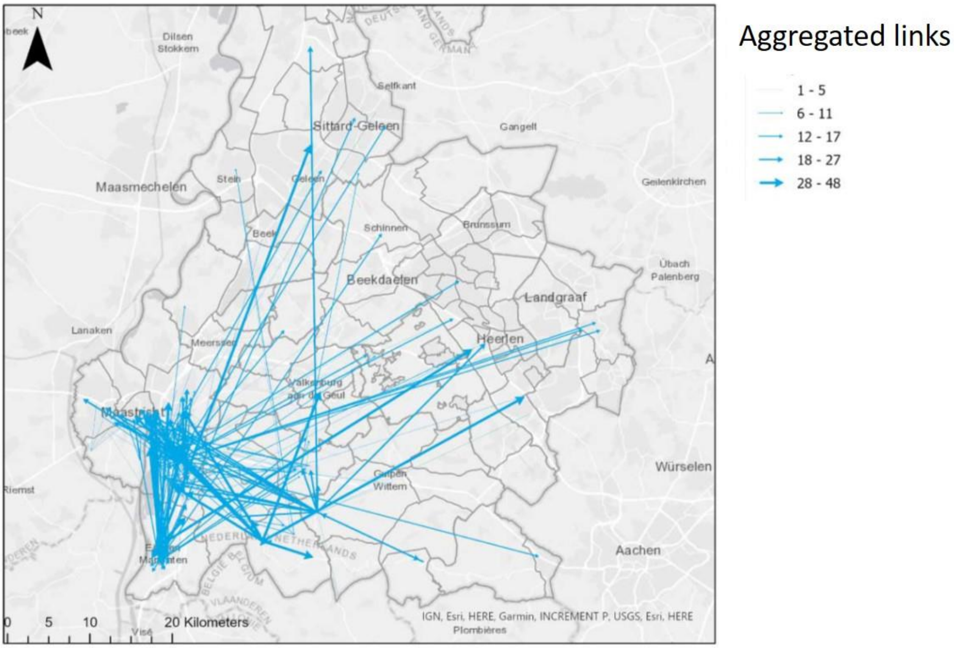


Figure 4.1: Aggregated origin–destination link map of Limburg. Links with the same origin–destination combinations are aggregated into one line.

Figure 4.2 displays the movement count per hour every weekday, which can be seen as a diurnal "curve" heat map. The darker color indicates a large number of movements during that specific hour and day. For all weekdays, we can clearly see peaks during the morning rush hour from 8AM to 9AM and during the afternoon rush hour between 5PM and 6PM, with movement numbers rising prior to the peak and slowly diminishing in the hour after. Further, we observe that people seem to finish work earlier on Fridays, because significantly more movement is observed around 2PM compared to other weekdays. Especially in the middle of the week, we can see activity during lunchtime, which may indicate that people leave their workplaces to get lunch. The diurnal curves for the weekend look significantly different from those for the weekdays. We can see that people start their days later; there is very little activity before 9AM on Saturday or before 11AM on Sunday. People also end their days earlier on weekends, and most nighttime activity is recorded for Friday evenings.

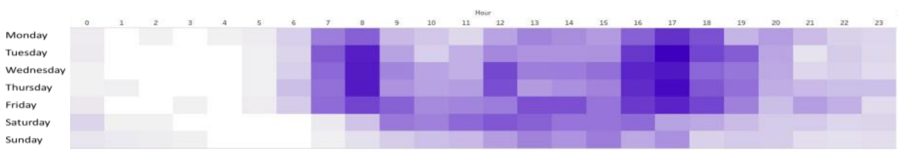


Figure 4.2: Temporal variations of movement. The dark purple indicates a larger number of movements during a specific time of day. The lighter the color, the lower the movement frequency.

To add more context to the flow analysis, this research includes building usage as an additional visualization layer. We choose to include this information because it can potentially explain some of the movements in the region. To extract this potential connection between a movement and its purpose, we paired morning trips with industries and offices as they depict the common morning working destinations of individuals. The larger the circle, the greater the number of buildings within this category that are within the 300 m radius of a given point. Figure 4.3a shows the high density of industrial buildings around the Sittard area, which can explain the home–work OD links directed to this location. Additionally, relatively high densities of offices are located in the central and eastern parts of Maastricht. Midday links are associated with secondary activities such as lunch, leisure, or shopping activities; thus, we pair such buildings with movements occurring during the day (Figure 4.3b).

In rural locations, the choice of activity settings within convenient travel distances is frequently limited (Millward & Spinney, 2011), which can explain the variability of movement from rural toward urban locations. Alternatively, half-day work trips or school–home journeys can also deliver part of the evidence. Figure 4.3c shows an increased link magnitude on the outskirts of Maastricht and another group of outflow links originating from the southeast part toward the central parts of the urban area, where multiple educational buildings are located.

Finally, the evening OD links are characterized by an increased link magnitude advancing to Heerlen. The urban region is known for its relatively cheaper real estate market compared with the other cities in Limburg. Hence, multiple residential buildings are located in the area, which explains the work–home trips during the evening hours (Figure 4.3d). We see that using the building functions of destination areas as a proxy for pursued activities leads to a deeper understanding of why people move to particular places at different times of the day. Cities can use these insights for sustainable travel campaigns or when working on solutions to spread traffic more equally across the day. Municipalities should work with transport providers and employers to achieve their goals for the future of urban mobility.

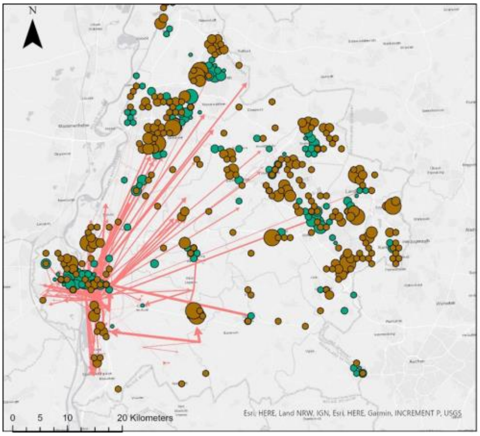


Figure 4.3a: Morning OD links with building frequencies of industry (brown) and offices (green).

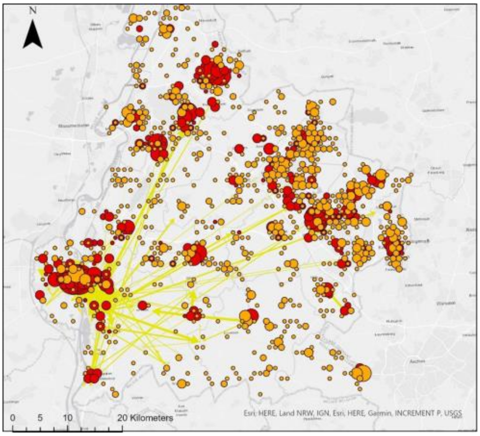


Figure 4.3b: Afternoon OD links with building frequencies of shops (red) and restaurants (yellow).

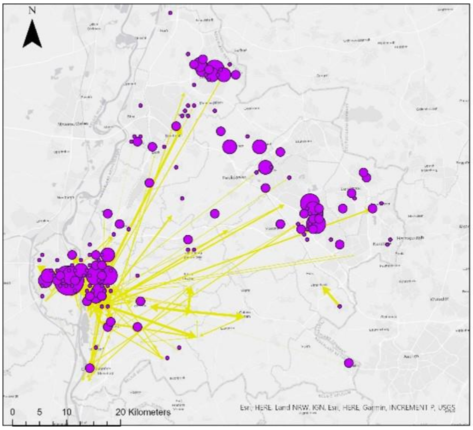


Figure 4.3c: Afternoon OD links with building frequency of educational institutions (purple).

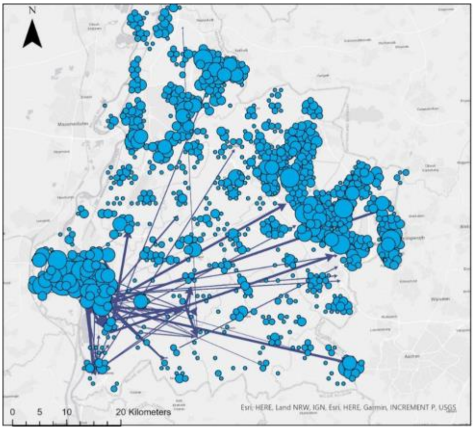


Figure 4.3d: Evening OD links with building frequency of residential buildings (blue).

4.4.2 Hotspot analysis

We can find hotspot analysis, also known as density analysis, in a variety of fields (Anselin, 1995; Cai et al., 2018; Visner et al., 2021). According to Hu et al. (2014), hotspots are regions with a large amount of people inflow or where events occur more often than in other areas. More generally, O'Sullivan and Unwin (2003) define density analysis as the capacity to compare and characterize intensities of processes across space as well as assess whether there is a substantial difference between them. Density analysis collects quantities of a specified variable and distributes them across geographical space based on what is being measured at each location. Consequently, density functions produce a continuous surface built on the concentration of feature points along a geographical area.

For a more flexible visualization exploration, ArcGIS Pro offers a collection of symbolization techniques that include more visualization methods beyond KDE (for details on the functionality of KDE, see Chapter 2). Heat mapping is one such option, and we use it in this analysis to draw the density points of the tracking data as a continuous color gradient. The relative density of points is displayed as a dynamic raster visualization that uses color-coding to represent density values. The color scheme consists of a smoothly changing collection of colors ranging from cold (blue, low point density) to hot (red, high point density).

Figure 4.4 displays Limburg's sparse and dense areas concerning trip starts, with an additional closer overview of the three main cities in the province (Maastricht, Sittard, and Heerlen). The more detailed the visualization gets, the easier it is to distinguish between sparse and dense urban regions. In line with the previous flow analysis, we see two groups of starting points. There is a substantial number of trips starting within the city boundaries, which can be assigned to people commuting between the three main cities or moving within one of them. We detect heat islands in the triangle between and especially along the north-south highway. A detailed look at the three cities allows us to see that Maastricht is much more frequented than the other two cities, where starting locations are more scattered and not equally or densely distributed across the entire city.

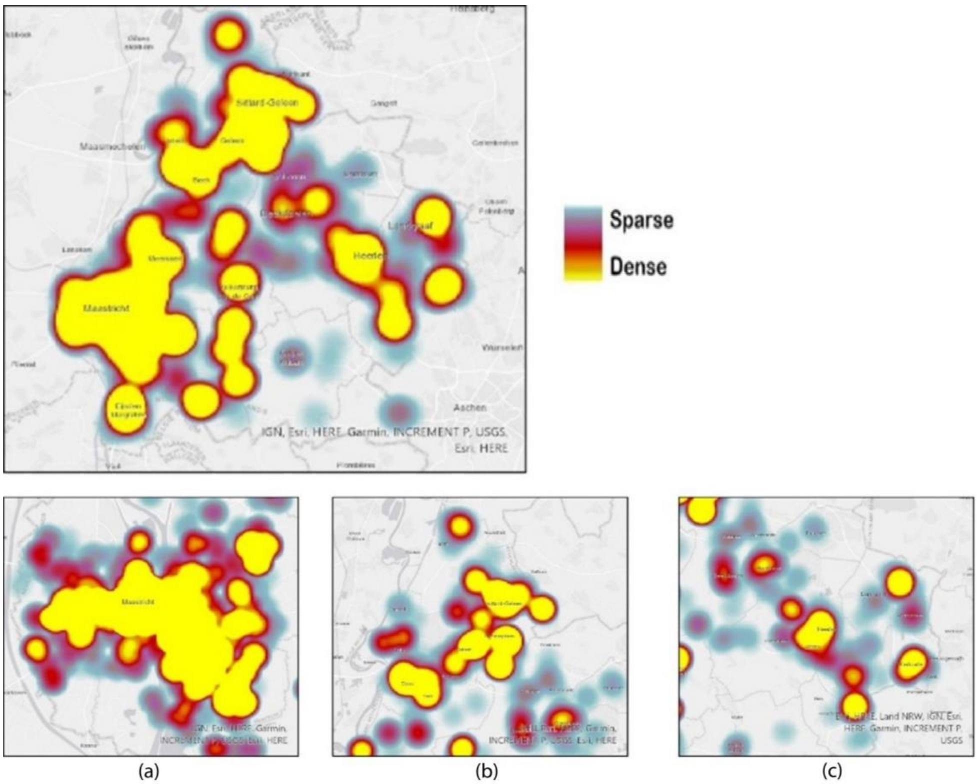


Figure 4.4: KDE heat map visual output. (a) Maastricht region, (b) Sittard region, (c) Heerlen region

To account for the temporal variations in mobility density, we divide the dataset by weekdays. Figure 4.5 illustrates the fluctuation of mobility patterns during weekdays and weekends. Three urban circles with radii of 6 km emphasize the urbanized regions in South Limburg. The green points within each circular boundary represent the mean centers of each of the three regions. The temporal graphical representation illustrates the regularity of density values throughout the week, with the most movement recorded on Wednesday, Thursday, Friday, and Saturday and the least movement on Sunday (Figure 4.5). In line with previous research, weekdays are characterized by greater consistency in travel behavior, with home–work, work–home, and home–school trips being the predominant behavioral patterns in most city areas. Conversely, weekends show relatively low and more diversified movement densities. Comparing all days, movements are spread across the area proportionally equal to each other.

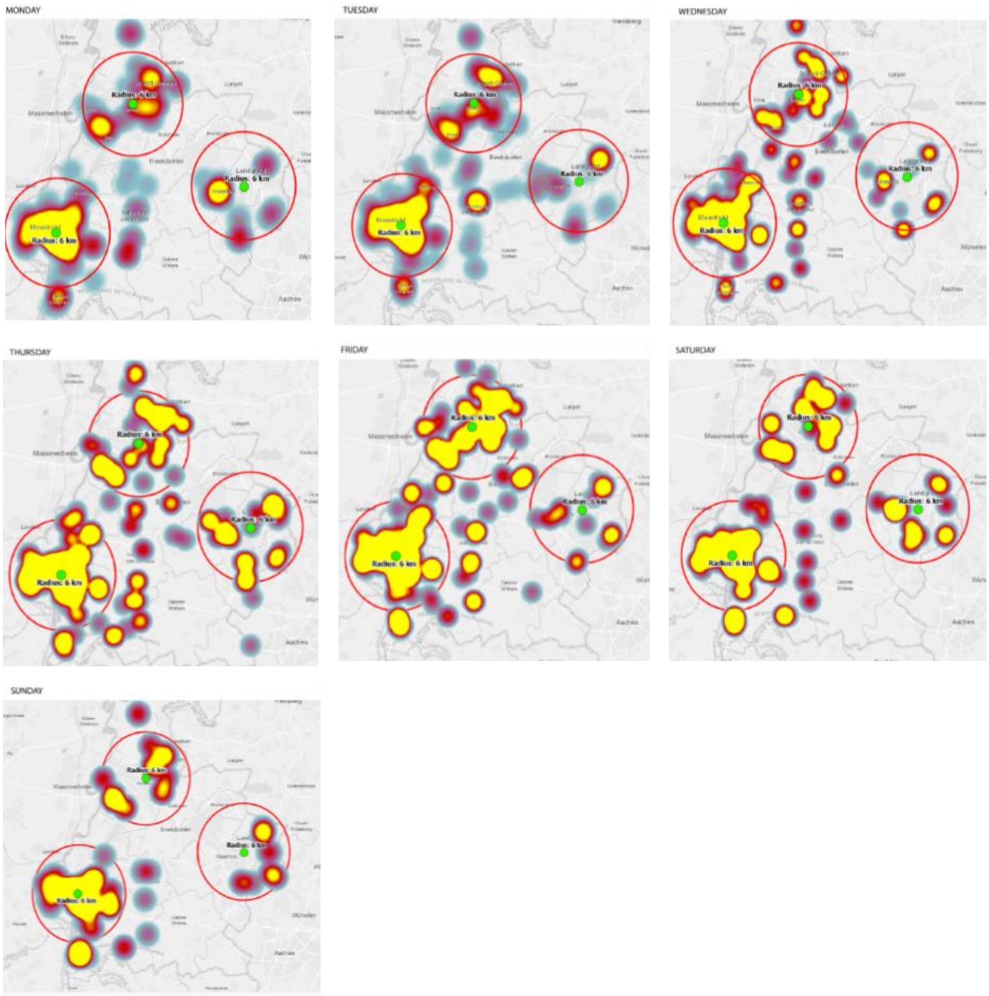


Figure 4.5: Temporal fluctuation KDE heat map during weekdays. Three main high-density islands (Maastricht, Sittard, Heerlen) are highlighted with red circles, indicating the densest locations throughout the week

KDEs, such as those presented in Figure 4.4 and Figure 4.5, can show where clusters are located, but the results do not incorporate the statistical significance of these clusters. Clusters are potentially formed randomly or because of the spatial structure or sample bias of the tracking data. An alternative method that assesses patterns and their significance is the Hot Spot Analysis Getis-Ord Gi*. This analysis labels hotspots as statistically significant high-value clusters and cold spots as statistically significant low-value clusters. The greater the z-score, the denser the clustering of high values (hot spot); the smaller the z-score, the denser the clustering of low values (cold spot). Figure 4.6 displays the visual output for the Getis-Ord Gi* analysis on the movement data. The results illustrate that the Maastricht area records a significant and highly dense hotspot area, whereas the central-east part of the province reports a significant cold spot value with negative z-scores. Further, the northern and eastern parts of Limburg record non-significant outputs.

The urban-rural divide can serve as a potential explanation; more densely populated areas naturally record more movements, whereas sparsely populated areas have a small absolute number of movements. However, popular residential areas are not the only hotspots for trip origins and destinations, as workplaces are also common trip origins. Based on the optimized hotspot analysis, we can assume that Maastricht provides residential areas, workplaces, and leisure activities, whereas the cold spot area does not attract any movement traffic because it lacks these features. However, another explanation for this distribution of movements is the sample population used in this research, which shows origin and destination records skewed toward originating in and leading to the Maastricht urban area.

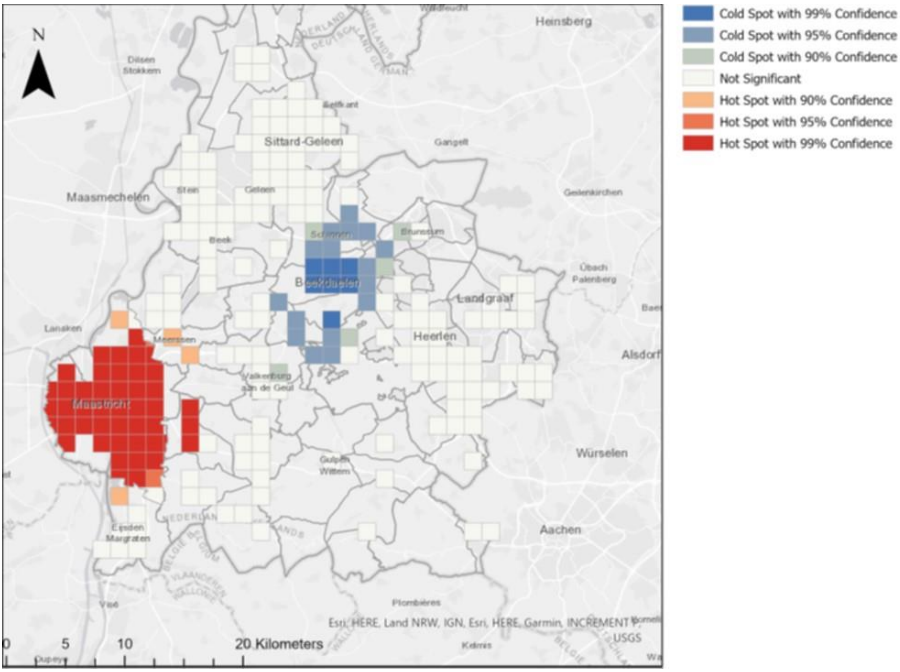


Figure 4.6: Optimized hotspot analysis. Red indicates statistically significant hotspots, blue indicates statistically significant cold spots, and white indicates non-significant results

In Figure 4.6, we can see that the movement data show one large hotspot in the Maastricht area, whereas Sittard and Heerlen are insignificant. Consequently, we wonder whether this is a sample-specific finding or whether the built environment, its density, and available activities in this area can explain it. Therefore, we apply hotspot analysis in the same area on the sum of non-residential buildings per cell. Figure 4.7 displays the results, where Maastricht, Sittard, and Heerlen present highly significant hotspot locations. Again, Figure 4.7 confirms the formation of three main hotspot islands in Limburg. It is notable that especially in the Heerlen area (eastern hotspot), there are considerable numbers of buildings with specific purposes that could and should attract people. However, in the hotspot analysis, density map, and flow analysis, Heerlen is less frequented than one would assume based on this graph. The explanation could again be the dataset that might be biased toward Maastricht or that Heerlen is unattractive for other reasons. If the latter is a realistic option, regional governments should take such findings seriously, start a second data collection to confirm this trend, and, if confirmed, identify the reasons so that they may act upon them.

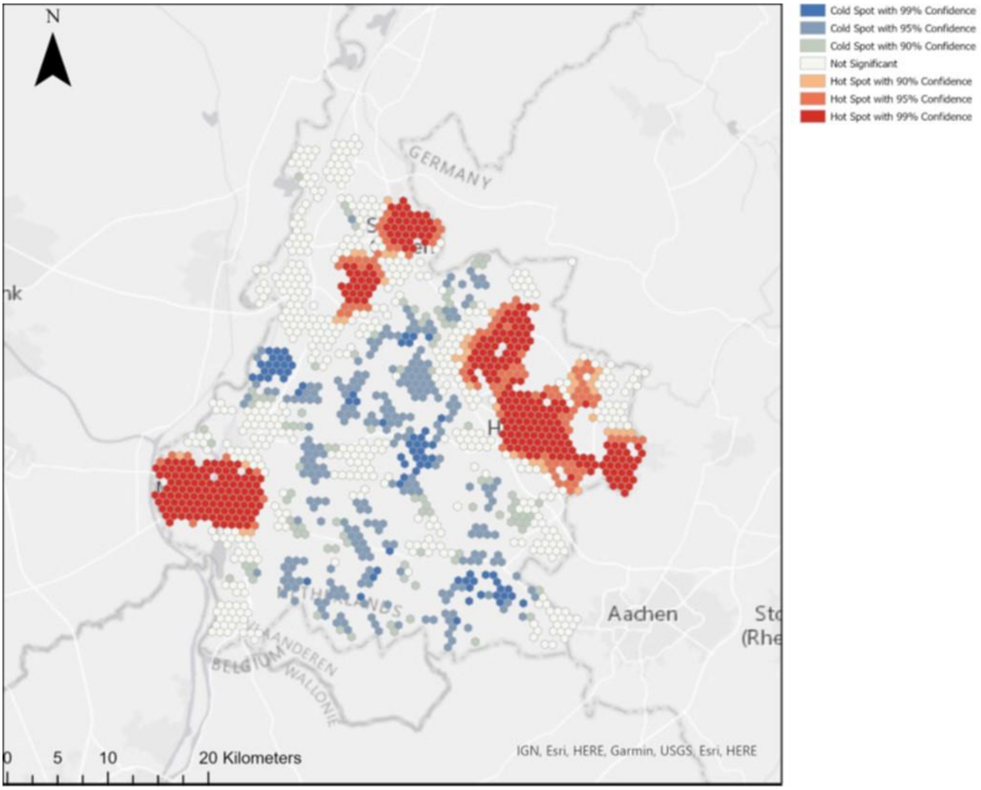


Figure 4.7: Optimized hotspot analysis of building usage. Red indicates statistically significant hot spots, blue indicates statistically significant cold spots, and white indicates non-significant results

4.4.3 Mobility in context

When we talk about human mobility, the built environment and the movements within this environment are the two core components to evaluate. However, examining building locations and flows alone allows for only a limited understanding of the situation. Researchers in this field need to add more attributes to their analyses to put their findings into context. For example, mode of transport plays an essential role in predicting and understanding urban behavior and permits differentiation of the flow analysis for deeper insights. According to Pan et al. (2009), neighborhood characteristics including easy access to local facilities, diverse land-use, and effective public transportation are associated with reduced private car usage and hence lower emissions. Although a pedestrian- and cyclist-friendly urban design may not be able to stop the general income-driven trend of growing motorization, it does assist in moderating the rate of expansion and lessens the need for high levels of car ownership. Pedestrian- and cyclist-friendly urban design makes non-motorized modes viable alternatives, which is critical for reducing vehicle dependency. The entire province of Limburg is characterized by a systemized public transport system, with bus stop locations situated on each main street and in every neighborhood.

Figure 4.8 shows this infrastructure on a map, both for the region of South Limburg and the city center of Maastricht. The thick blue lines show cycle paths, the green lines indicate small alleys and paths that are only accessible for pedestrians, the brown lines represent streets that are pedestrian-only areas during the daytime and shopping hours, and the red lines highlight all streets that provide sidewalks to pedestrians. Based on the map, we can point out two noteworthy insights. First, cycle paths cover not only the cities but the entire region, and they connect cities with suburban areas. They provide safe options for non-motorized transport between communities and provide the infrastructure required for commuters and leisure cyclists. Second, within cities and villages, where people usually cover shorter distances that do not necessarily require car use, citizens can rely on an extensive network of pedestrian-friendly infrastructure.

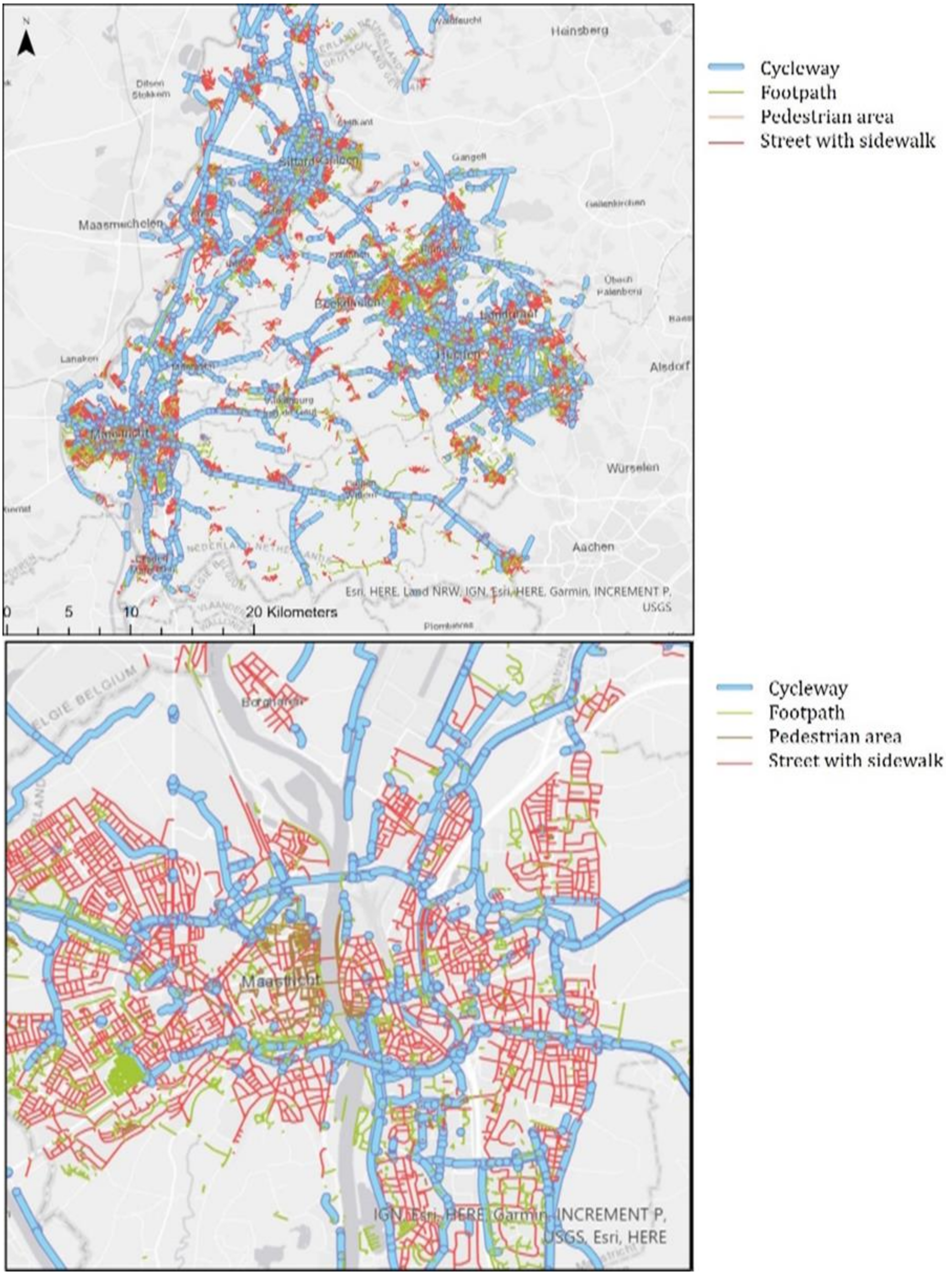


Figure 4.8: Limburg road infrastructure

Knowing that the required safe infrastructure is in place, we would like to explore the extent to which it is currently being used. To illustrate the use of different types of transport, we apply unique values symbolization. This tool groups categorical variables into classes and displays them in different urban locations. Figure 4.9 presents the classification map for the mode of transport in South Limburg and colors each cell according to the most dominant mode in absolute numbers. The encoded colors for the transportation type allow the reader to easily discern the regions that use non-motorized modes of transport for movement. Maastricht appears to record the highest number of bicycle trips, followed by Sittard and Heerlen. Walking trips are primarily visible in the northern part of the province, and on a general level, the use of cars predominates the entire location. Combining this graph with the previous presentation of the cycling and walking infrastructure, we draw a twofold conclusion. People seem to make use of the pedestrian infrastructure within all three cities as well as in surrounding villages to an extent that we consider satisfactory, but there is room for improvement in bicycle usage. Few movements to or from less urbanized areas are made by bike, and most people still opt for cars, even though the cycling infrastructure is in place. Potential reasons for this finding are numerous; municipalities should explore them so that cycling takes place both within cities and between communities.

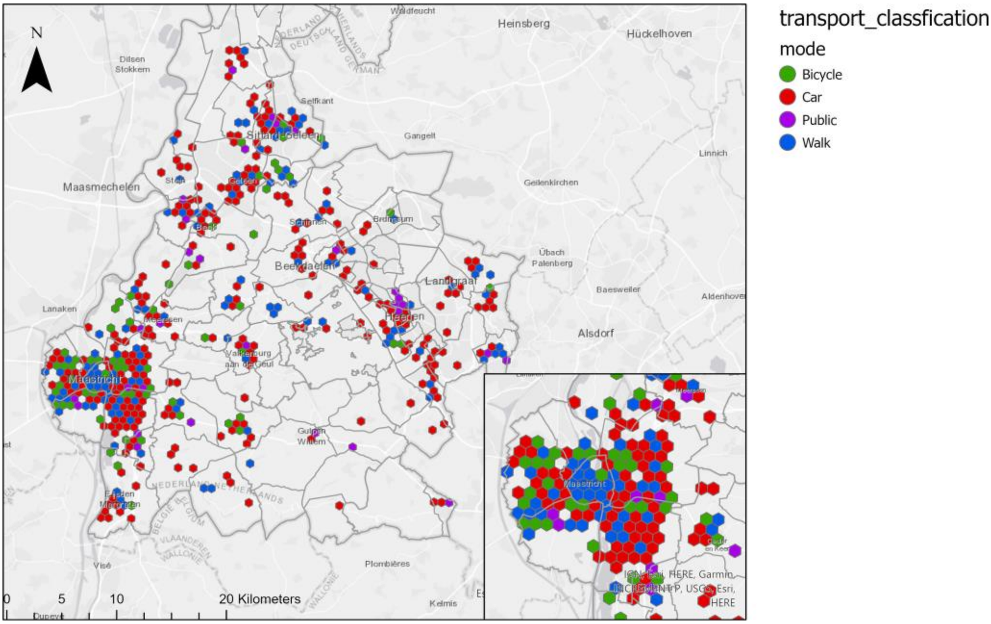


Figure 4.9: Classification of transport mode. This graph groups transport types into classes and assigns a different color to each transport mode

Although the unique value symbology is a tool that tries to prevent the visual clutter and distinguishes between various categorical variables, Figure 4.9 is limited and is not sufficiently representative to evaluate the full transport mobility picture. Hence, we decided to incorporate the OD link proximity tool once more and to identify which type of transport is predominant in a specific urban area. The results confirm what we assumed based on the previous classification. Cars are the preferred mode of transport for longer trip distances, trips between cities, and trips to the suburban areas and villages in between. Figure 4.10 illustrates the collection of links split into the four categories (car usage, public transport, walking, and cycling). There seems to be a minority of people who commute between Maastricht and Heerlen or Maastricht and Sittard by regional train as well as a few bus trips within Maastricht. Walking is naturally bound to short distances within cities. Cycling mainly takes place within city boundaries, with few movements to the closest northern and southern suburbs of Maastricht.

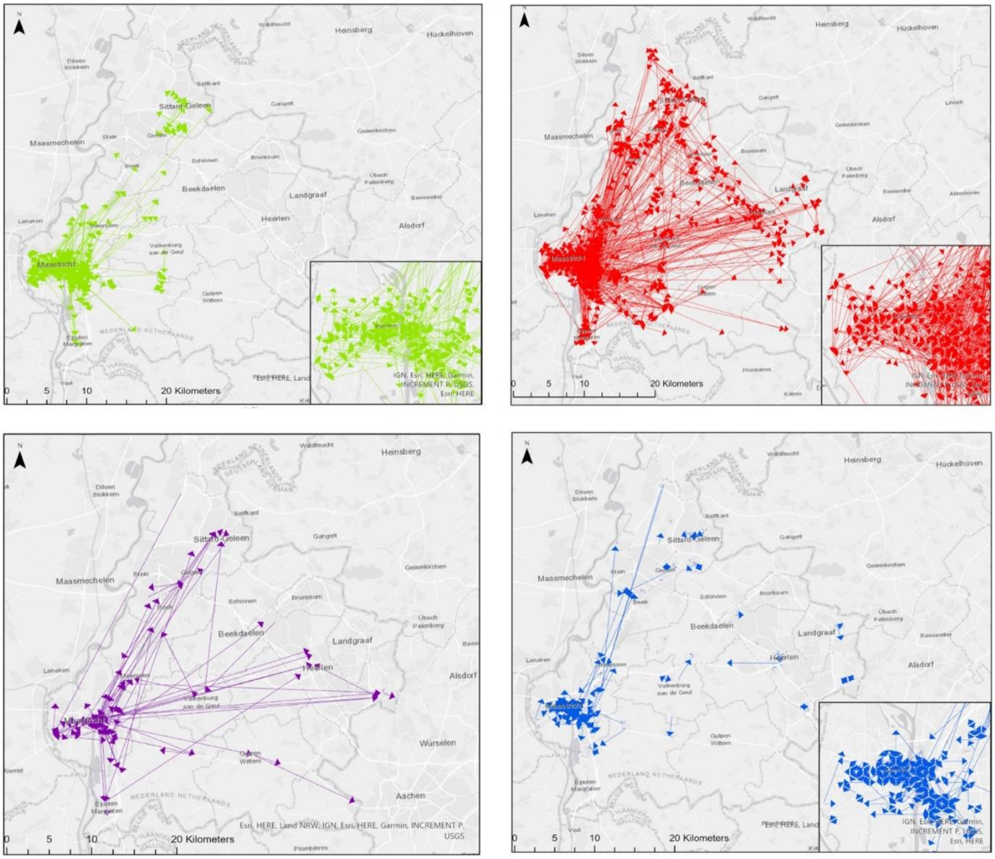


Figure 4.10: OD links by mode of transport: bicycle (green), car (red), public transport (purple), and walking (blue)

4.5 Conclusion

Data visualization is an efficient approach for analyzing urban mobility dynamics and movement structures. The combination of different techniques is a powerful tool to draw a holistic picture and to get a good overview of the situation for several reasons. First, geospatial data usually have a location component (XY-coordinates), a time component, and often a dynamic location component such as OD specifics or routing information plus additional features that add to the complexity of the problem. Data visualization can handle the multidimensional nature of geospatial data by using different colors and symbols to capture as many details about a phenomenon as possible (Demissie et al., 2013; Guo et al., 2011). Additional techniques not presented in this paper, such as 3D visualizations, interactive maps, or animated maps, further increase the number of layers to be visualized.

Second, and in line with the previous thought, data visualization allows researchers to look at multiple datasets simultaneously. For example, we look at trajectories (dataset 1) on a map (dataset 2) in combination with the purposes of the buildings in destination areas (dataset 3). In a computational setting or a results table, the dependencies of these three datasets would be quite difficult to present in an understandable manner.

Third, visualizations are frequently used in the exploratory stage of a research project because graphic representations of data make it easier to discover relationships and the human brain can process visuals much better than numeric representations of linkages because they reduce complexity (Thomas & Cook, 2006). Thus, data visualization can give a first glimpse of which relationships are to be expected, and where it is worth diving deeper to confirm a relationship computationally.

Finally, the integration of complementary contextual variables gives researchers the opportunity to emphasize specific aspects of urban living that might have been neglected before or that would have been overlooked if not for the graphic representation. We presented an example of this in Figure 4.3, in which we divided movements based on trip timing and combined them with a subset of relevant buildings for the corresponding time window. Likewise, the differentiation between different modes of transport in Figure 4.9 and Figure 4.10 sheds light on where and for which routes certain modes of transport are used in the region.

We were able to derive concrete results for the region that the municipality of Maastricht can incorporate in their future policy-making and urban planning. Because human movements are always subject to the geographical circumstances of the observation region, generalization beyond the Maastricht area is difficult and should be verified through a second data collection and analysis. In fact, it is already hard to draw conclusions for the municipalities of Heerlen and Sittard because most of the tracked people lived in and around Maastricht; thus, the data at hand are skewed toward this city and not evenly distributed across all three places. Our analysis even backs this skewness, because the optimized hotspot analysis of the buildings (Figure 4.7) matches the flow analysis and density map much better than the optimized hotspot analysis of the movements (Figure 4.6). Consequently, we can conclude that a large and appropriate data source is a crucial factor for achieving meaningful results for the region of interest.

In general, data visualization can be used as a standalone exploratory approach such as the one presented in this chapter. However, it is often seen as complementary to traditional statistics. It can be used either at the beginning of the research process for initial exploration and oversight or at the end to support the interpretations of the mathematical concepts. When researchers use data visualizations in the data exploration phase of complex problems, they select suitable techniques to sketch all dimensions involved, to disentangle complexity layers, and to get a first glimpse on the relation between data layers, before they verify these assumptions with computational techniques. Alternative, researchers might prefer to approach a research problem numerically but they would still like to make their findings more tangible and thus, decide to provide data visualizations after the calculations. Especially when data visualization is used toward the end to support and summarize research results, interactive dashboards can be convenient. Such dashboards comprise multiple visualizations and allow the user to filter for different components. This way, managers do not need to read a lengthy report about all analyses, but can instead look at concrete scenarios to make informed decisions and explore trends (Chen et al., 2015). In this dissertation, the data visualization serves as both verification and exploration. While the hotspot analysis supports the findings of the city sketching (Chapters 2 and 3) and highlights once more where popular places can be found and what defines them, the trajectory analysis combined with further context variables kicks off the inclusion of movements as the focus of the analysis. All previous chapters have used the trajectory data only as a supporting data layer but not as a core part of the analysis or research question.

Finally, this chapter showed that human trajectory patterns form a network of high- and low-frequented locations connected through movements with different modes of transport. Coming back to the geo-social networks addressed in the introduction, selected data visualization techniques show inflows of people from specific neighborhoods to certain cells or even individual stores while other neighborhoods do not register any movements to that specific location. One explanation implies that offline social networks and communities yield isomorphic mobility networks (Brown et al., 2014; Coscia & Hausmann, 2015). With regard to CBM, this is important to realize, because offline social networks are a core aspect in word-of-mouth marketing, which can generate additional store traffic if executed effectively. Attracting a few new customers who are well connected in their communities may have a big impact on future store traffic. Related marketing assets should be placed strategically and target those people centrally located in the network or, even better, who can be classified as "social bridges" (Dong et al., 2018) from the existing customer base to a new neighborhood or community.

4.6 Impact on hyper-targeting

Well-executed data visualization of urban dynamics supports municipalities and businesses in detecting service gaps. Through a clever combination of multiple datasets, decision-makers can identify heavily traveled roads to increase road safety, install additional bike paths, or adjust public transport timetables if demand does not match the current supply. Data visualization dashboards can be used to simulate different scenarios and give early warnings about eventualities that might have been neglected otherwise. Based on these scenarios, city authorities can prioritize their investment projects and start where improvement is needed the most.

In addition, trajectory analysis adds a dynamic perspective to CBM. Currently, CBM focuses solely on where a consumer is at a particular moment and places mobile advertising on this basis. However, dynamic trajectory information answers questions about where the consumer came from, what they may have been doing previously, and which mode of transport they used to get to their current location. The inclusion of physical infrastructure puts to-be-targeted consumers into a geographical context and uncovers more behavior insights and daily routines.

Information about the mode of transport, in combination with trip timing, opens up many additional opportunities for improved hyper-targeting. Businesses and transport providers should consider collaborating to increase store traffic for the businesses while making mobility more sustainable but still suitable for the consumer. For example, some people live in rural areas and are dependent on cars because public transport availability is limited and using it is time-consuming. In other cases, a person's trajectory patterns may indicate that they have a part-time job or work shifts and are free every afternoon. These people should receive incentives such as cheaper parking for weekday afternoons to make it more attractive for them to go to the city during the week rather than on the weekend. Alternatively, collaborations between businesses and transport providers could include discounts at shops for using a sustainable mode of transport and similar offers.

Businesses should try to monetize the dynamic information about where the customer comes from. For example, one knows that most customers combine their store visits with other store visits or in a chain trip on their way home from work. This is valuable information for LBM that we can use to target customers at the right locations and with appropriate offers. Further, if a shop or restaurant owner knows that the majority of customers come from a given neighborhood and that other neighborhoods produce fewer customers, advertising in the respective neighborhoods can be increased or decreased according to the marketing strategy. Complementary neighborhood statistics are usually openly available from a country's national statistics agency.

Generating trajectory data, including the residential locations of tracked citizens required for such an analysis, will be complicated because working with big data always comes with restrictions. Nonetheless, data visualization is a very flexible method; one can add a variety of data layers to illustrate interests, demographics, shopping behavior, and similar topics. As always, the more data layers included, the more precise the personas and the more successful the hyper-targeting.

In a perfect world, researchers would be able to use more detailed and complete information about human mobility behavior for an even better representation of reality. However, due to personal privacy restrictions, most information needs to be stored and processed on an aggregated level. Therefore, the next chapter introduces an attempt to use big data and machine learning to bridge this gap within the legal boundaries.

Transition 4: What hinders deeper behavioral insights?

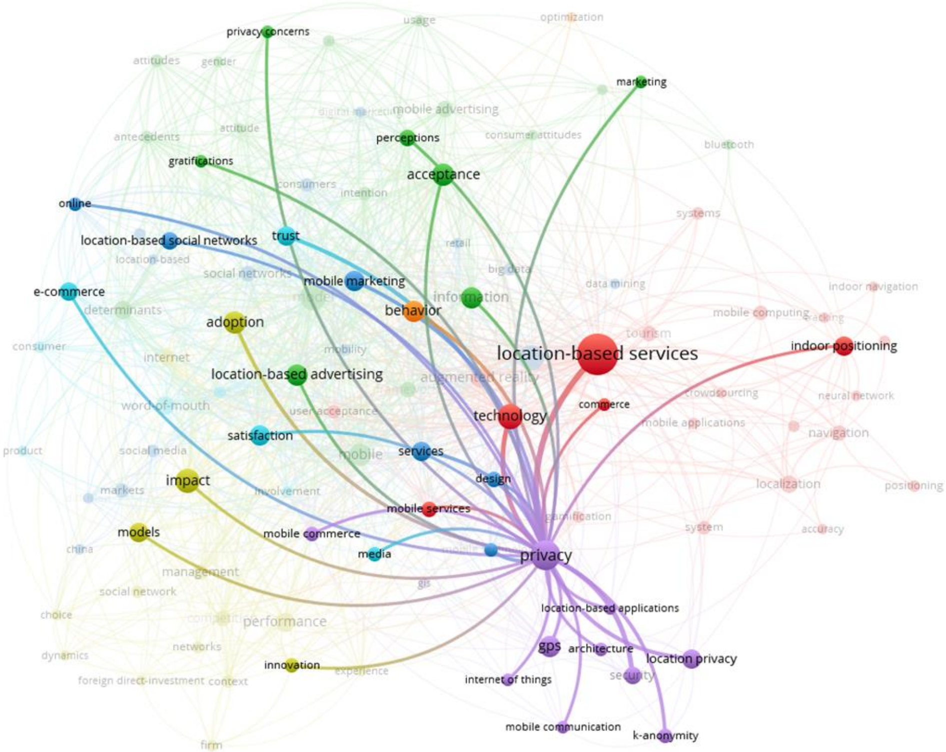
In Chapter 4, we were not able to dive deeper than the regional or city level for two reasons. First, the study is limited by the small sample size. This makes it impossible to divide the sample on neighborhood level, as the subsets would be too small to reach any robust, meaningful, or significant results. Second, to preserve the privacy of the tracked individuals, data privacy regulations in Europe do not allow very detailed examinations of the collected human movement data.

In particular, these data privacy regulations impose restrictions on researchers. First, trajectory datasets cannot carry personal identifiers. Such identifiers would be desirable from a research perspective because they would allow the recurring movement patterns of an individual to be extracted. This would deepen the understanding of routines and ensure the continuity of movement patterns over a longer period of time.

Second, data privacy regulations limit the granularity of the origin and destination points on a map. Researchers are not allowed to store the exact coordinates of a trip's start and end points, and they must anonymize this information on a less granular grid. We were able to achieve sufficient data privacy with a 300 m grid, which is a satisfyingly high granularity compared to other studies.

Third, data privacy restricts the avenues for collecting additional demographic and personal information about the tracked individuals. In practice, this means that it is not possible for researchers to contact the study participants at all. If they do have contact details to gather additional information, the regulations make it impossible to connect individual survey responses to movement data. Being able to match a survey response with demographic and personal data to all observed movements of an individual would be ideal and would allow for fascinating behavioral insights about human mobility. However, we do not live in a perfect research world, and researcher are limited in their ability to dive deeper and uncover meaningful behavioral patterns.

Chapter 5 introduces an attempt to use unsupervised machine learning to bridge this data gap within legal boundaries with the goal of reverse engineering trip purposes from movements.



Chapter 5

Bridging the Data Privacy Gap—An Attempt to Reverse Engineer Trip Purpose from Smartphone-Based GPS Trajectories

I would like to thank Niels Schoeters for carrying out part of the literature search and the analysis.

5.1 Introduction

Human mobility is an interdisciplinary field that employs techniques and methods from social science, physics, computer science, transportation science, and human geography to create models that capture patterns about the ways individuals and human populations move through time and space (Barbosa et al., 2018; Wang et al., 2019). Studying human mobility is important because the daily movement of human populations imposes a significant strain on the environment and society itself (Barbosa et al., 2018). Within the field of human mobility, the ability to infer and predict the purposes of trips from travel survey data has been a long-standing research topic (Chen et al., 2019). According to an activity-based view on understanding the movement dynamics of a population, the demand for travel is derived from people's need to participate in activities that are distributed in space and time (Gong et al., 2016; Zhao et al., 2020). A trip is hereby defined as the journey from an origin point toward a destination point with the intent of performing a particular activity (Zhao et al., 2020). Consequently, the purpose of a trip is highly entangled with human mobility and critical for analyzing, modeling, and predicting human travel behavior (Gong et al., 2016).

Scholars have access to growing volumes of spatiotemporal information, which are collected passively from smartphones, public transit smart cards, taxi probes, e-bike platforms, and geo-tagged social media data (Gong et al., 2016). These GPS-enabled sources overcome many of the limitations of conventional datasets and allow researchers to examine data on a larger scale, track participants over a longer period, decrease the burden on participants, and keep costs low (Chen et al., 2019; Nguyen et al., 2020). In addition, these sources make it possible to capture trajectory data that accurately reveal when and where subjects traveled at a high spatiotemporal resolution and with fine granularity (Nguyen et al., 2020).

However, the collected GPS data lacks information such as the actual trip purpose, spending behavior, or demographic and socioeconomic data including a home or work address (Chen et al., 2019; Gong et al., 2016; Zhao et al., 2020). Even if a user has some of this information stored in their smartphone, the tech company providing the operating system of the device is not allowed to retrieve this information, connect it to other data collected via the same smartphone, or pass this database to third parties. All of this is secured under the European GDPR agreement introduced in 2018 to preserve consumers' privacy (EU, 2018). Consequently, studies aim to develop innovative techniques that can derive the missing trip purpose information from these "trajectory rich" but "semantically poor" datasets to complement or substitute for conventional data collection methods (Chen et al., 2019; Nguyen et al., 2020; Zhao et al., 2020).

Most of the existing literature has formulated such tasks as classification problems, which involves the use of supervised learning methods for which the ground truth needs to be included in the dataset to train and validate the models. For GPS-enabled spatiotemporal datasets, the trip purpose information is unavailable, and the true values are typically collected through a complementary prompted recall survey (Chen et al., 2019). It requires that researchers predefine a list of purpose categories to ask about on that recall survey (Zhao et al., 2020). An alternative approach to bridge the data gap caused by the GDPR is an unsupervised clustering task. It also implies focusing on trip purpose discovery to find and profile naturally occurring representative trip purpose categories rather than correctly inferring predefined purpose categories (Zhao et al., 2020).

The primary goal of the present study is therefore to investigate the following: How can we use unsupervised clustering techniques to reverse engineer trip purpose from GPS-enabled smartphone spatiotemporal tracking data? As sub-objectives, we review which conventional clustering technique performs best.

Given its popularity, simplicity, and wide applicability, we focus on density-based spatial clustering of applications with noise (DBSCAN). Once the trips have been clustered into naturally occurring groups, we examine the spatiotemporal patterns and basic trip attributes to profile and label any inherent structures for trip purpose categories. In particular, we focus on the temporal distributions as trips with the same purpose tend to have similar temporal regularities. Additionally, we examine the spatial distribution of trips and their distribution per building category since regions with similar building configurations likely support related trip purposes. Hence, profiling these spatiotemporal patterns while also considering basic trip characteristics commonly used in the inference literature, such as length, duration, and transport mode, may be useful in labeling the corresponding clusters along distinct trip purpose categories.

5.2 Background and research gap

5.2.1 Data privacy in data collection

The introduction of the GDPR in Europe in spring 2018 has had a major influence on public and private entities working with big data. The aim of this regulatory framework is to reunite data privacy and innovation, but that is easier said than done (Maldoff, 2016). The GDPR allows researchers to collect only the minimum amount of data that is directly required to answer a particular research problem. The data subject (i.e., the observed individual) needs to be informed about the data collection, and it should be easy to find out who is collecting, storing, and processing the data and for what purpose (Kotsios et al., 2019). In the era of big data, this is an extremely difficult task and hinders many researchers who wish to use the data most suitable to their research question.

Although there are some exemptions in place for academia, commonly used resources such as social network data, GPS data from smartphones, and other indirect data collection methods are harder to use now, especially in combination with other sources (Kotsios et al., 2019; Maldoff, 2016). The GDPR requires pseudonymization or anonymization, which means that the data cannot be traced back to their subjects. Thus, it is no longer possible to combine automated data collection with traditional survey data or similar combinations of online big data and individual information (Maldoff, 2016). Additionally, sharing collected data has become more complex and restricted under the GDPR, which is problematic for companies sharing their data with researchers for joint projects (Greene et al., 2019).

Switching perspectives to the data subject, Libaque-Sáenz et al. (2021) as well Paul et al. (2020) found that people have fewer privacy concerns now that the GDPR is in place compared to before 2018. This is a very positive development; while the current situation might be less than ideal for researchers, the data privacy of individuals should be treasured, and no entity should take advantage of people's decreasing privacy concerns. However, some researchers argue that the GDPR is too vague with respect to data protection because it controls the data collection stage but has few mechanisms to monitor how companies analyze and process data (Wachter, 2018, 2019). Thus, consumers might lull themselves into a false sense of security, and that is just as dangerous.

In summary, the introduction of the GDPR has caused both intended and unintended effects on innovation, research, efficiency, competition, and data science developments (Gal & Aviv, 2020). They are necessary consequences of preserving data privacy. Academics are faced with the new challenge of finding ways to deal with the limitations and creatively developing new methods to bridge the information gap.

5.2.2 Supervised GPS-enabled trip purpose inference

Most of the extant research over the last two decades consists of supervised works (Nguyen et al., 2020). Given the recent ambitions for completely substituting the traditional household travel survey with the cost-effective GPS-only study, exploring machine learning methods for analysis was a necessary step (Shen & Stopher, 2014). This research stream is dedicated to developing and improving techniques that can automatically predict the correct trip purpose. To train a classification algorithm, the actual trip purpose information needs to be included in the dataset and is used to validate the performance of these models in the test phase. This trip purpose information is mostly collected using complementary prompted recall surveys (Chen et al., 2019). The researchers predefine the list of trip purposes from which participants can choose. Once these two datasets are merged into one larger dataset, classification of the pre-processed GPS data and a wide variety of input variables begins.

The need for the actual trip purpose is the major limitation of trip purpose inference combined with supervised machine learning approaches. With the data collection tools we currently have and the data privacy regulations in place, it is not possible to collect this information automatically on a large scale (Kotsios et al., 2019). GPS data from portable devices are not always connected to users' personal information; even if researchers wanted to put the burden on people by asking them to indicate their trip purposes manually, it would not be technically feasible (Barbosa et al., 2018).

Another limitation stems from the use of predefined trip purpose categories that might not be representative of the real trip purposes. Additionally, studies that decide to ask participants to indicate trip purposes manually often lack sufficient explanations for each category (Shen & Stopher, 2014). There is no consensus among researchers on the number of categories, or which ones to use. To illustrate, in the highlighted studies, the number of trip purpose categories ranges from 3 to 11, with only some overlap in the definitions (Chen et al., 2019; Nguyen et al., 2020; Shen & Stopher, 2014). As a result, caution is advised when interpreting the reported accuracies to compare and contrast existing research. An alternative method to infer supervised trip purposes is needed. We will examine existing literature on unsupervised methods that focus on trip purpose discovery, consider the true nature of human travel data, and are able to uncover and profile inherent patterns.

5.2.3 Unsupervised GPS-enabled trip purpose discovery

Unsupervised learning methods have been used previously to analyze mobility data, albeit for different research objectives. Clustering methods based on similarity measures such as k-means, partition around medoids, hierarchical agglomerative clustering, and DBSCAN clustering are commonly used to uncover mobility behavior patterns in spatiotemporal datasets (Moreau et al., 2020; Shi & Pun-Cheng, 2019). Subsequently, dimension reduction techniques including principal component analysis or uniform manifold approximation and projection are used to extract only the primary dimensions and facilitate the visualization of the outcome (Moreau et al., 2020). For instance, a study by Jiang, Ferreira and Gonzalez (2012) uses k-means clustering via principal component analysis on an activity-based travel survey of Chicago containing household, personal, social demographic, trip detail, and location data to partition the population into eight groups on weekdays, and seven on weekends. The authors provide significant evidence for repeating patterns and spatial and temporal correlations for visited locations. This underlines that human mobility patterns are highly structured and that activities follow similar daily and weekly routines (Barbosa et al., 2018; Gonzalez et al., 2008; Jiang, Ferreira, & Gonzalez, 2012). Considering that trips are taken to perform particular activities at specific locations and times, trips that share similar purposes can thus be expected to have similar spatial and temporal characteristics.

So far, only a few studies focus directly on discovering trip purpose categories inherently present in the data. The existing research mostly analyzes highly detailed trajectory information from e-bike sharing systems, smart card data, or taxi-probe data (Bao et al., 2017; Chen et al., 2019; Pengfei Wang et al., 2017; Xing et al., 2020). All focus on highly urbanized environments, and three of the four identified studies use New York City data. In addition, these works are characterized by the use of complex techniques such as modified topic models and deep learning to uncover hidden patterns (Pengfei Wang et al., 2017). Afterward, the output is typically combined with conventional clustering algorithms such as k-means to aggregate the patterns into different groups of trip purposes. Ultimately, the strong spatial and temporal regularity of human mobility is exploited to describe and profile the uncovered trip purpose categories.

Modified topic models, adapted from text mining, are suitable when high-quality point of interest (POI) information is available for the study area (Bao et al., 2017). Topic models are beneficial since they can extract hidden travel patterns from the datasets as topics. As output, the models indicate here probabilistic values of each POI belonging to those topics. The POIs with the highest probabilities can then be used to semantically increase the understanding of the travel pattern. Following this, clustering algorithms can be applied to aggregate these topics into trip purpose profiles (Chen et al., 2019). Notably, this process can also be conducted in reverse.

For instance, Pengfei Wang et al. (2017) combine a mixture of Hawkes processes with a hierarchical topic model to analyze taxi trajectories in New York City, US. Based on the POI configuration and temporal features, the authors discover ten hidden types of trip patterns. K-means clustering is utilized to further classify the patterns into three broad purpose categories: shopping and outdoors, dining, and homing. More importantly, this study observes the human mobility synchronization phenomenon, which implies that for a specific trip purpose, the temporal arrival rate is similar if two or more spatial areas share the same function (Pengfei Wang et al., 2017). This provides further evidence that human travel patterns are highly routinized and structured in terms of both spatial and temporal regularity.

Likewise, Bao et al. (2017) investigate trip patterns associated with the docked bike sharing system in New York by combining smart card data with POI information. In this study, k-means clustering is first used to group the bike sharing stations into five types based on POI configurations in the nearby area. Subsequently, they employ latent Dirichlet allocation to discover hidden types of trip purposes in the data. Ultimately, they extract fifty trip patterns, and the topics of the ten most important are investigated further to profile the six types of discovered purposes: eating, shopping, transferring to other transit modes, studying, homing, and other.

Chen et al. (2019) develop a novel clustering model called Trip2Vec to examine a dataset containing taxi trajectories in New York. First, time, origin and destination, and human activity popularity augment the context of the trips. The temporal context is extracted using the trip start and travel times. Second, considering the information of nearby POI configurations adds to the origin and destination context and the social media check-in data of an area during a particular period complement the human popularity context. Third, the authors convert the trip data to the latent space using a deep learning auto-coder to reveal hidden patterns. Fourth, they aggregate and partition these latent patterns into five purpose categories using k-means clustering: dining, recreation, work, homing, and other. Finally, Chen et al. (2019) profile the patterns of the discovered purpose types on a city-wide level based on time features such as hour of day, workday or non-workday, and whether the trip occurred during the day or night in the context of the study area.

Xing et al. (2020) explore dockless bike sharing rides in Shanghai, China, by augmenting trip origin and destination positions with the surrounding POI configurations. The authors apply a modified k-means clustering algorithm, k-means ++, to partition the trip patterns into six distinct types: dining, transferring to other public transport modes, shopping, work, integrated locations, and homing. This modified k-means model is essentially the same as the regular algorithm with the exception of optimizing the selection of the initial clustering center, reducing computing time. Afterward, the clusters are profiled through an examination of the temporal and spatial distributions. This study demonstrates how unsupervised clustering techniques can also be used to better inform decision-makers. In particular, the authors analyze the spatial distribution to formulate strategies to improve the dockless bike-sharing system in the area.

5.2.4 Research gap

The present study differs on several key grounds from the existing research on trip purpose discovery. First, it analyzes data from a human mobility data collection survey conducted in the area of Maastricht, which is a city located in the Netherlands, an EU country. This implies that the data conform to the GDPR, which governs how personal data are collected, processed, and stored (EU, 2018). Consequently, the available data are confined to a lower spatial granularity in contrast to the data analyzed in studies from non-EU countries. Additionally, the Maastricht area is significantly less dense and urbanized in comparison to the metropolitan environments of New York and Shanghai. The underlying data collection method differs, as smartphone-based GPS positioning from recruited individuals is used. Trips taken using different travel modes are observed at the cost of a substantially lower sample size. This contrasts with the extant research, which only focuses on one mode type, either taxi trajectory or e-bike sharing data, from unaware individuals, enabling considerably larger samples. However, given the nature of the data collected and the fact that the datasets contain attributes that are commonly used in trip purpose inferencing, basic clustering techniques, as used by Xing et al. (2020), allow for meaningful groupings and interpretations.

5.3 Analytical framework

The contemporary literature on trip purpose discovery has provided evidence that trips with similar purposes tend to share regularities with respect to their temporal and spatial patterns (Barbosa et al., 2018; Gonzalez et al., 2008). Consequently, this study approaches the trip inference problem as an unsupervised learning task in which we use clustering to extract spatiotemporal patterns and reveal inherent trip purpose categories. In essence, clustering is the process of dividing data objects into naturally occurring groups, called clusters, based on a particular distance metric. Clustering is successful when the distance between data objects within clusters is minimized but maximized among clusters. This ensures that the characteristics of observations inside the same cluster are almost identical while simultaneously being distinct from the observations in other clusters. Hence, clustering is ideal for exploring complex datasets when little to no information about the distribution is available a priori. Once we have identified meaningful natural group structures in the data, we examine the unique temporal and spatial patterns to label and profile the best clustering outcomes for the goal of trip purpose discovery. Figure 5.1 shows the analytical framework of the clustering analysis followed in this research. For a detailed description of the underlying data sources, please see Subchapter 1.4.

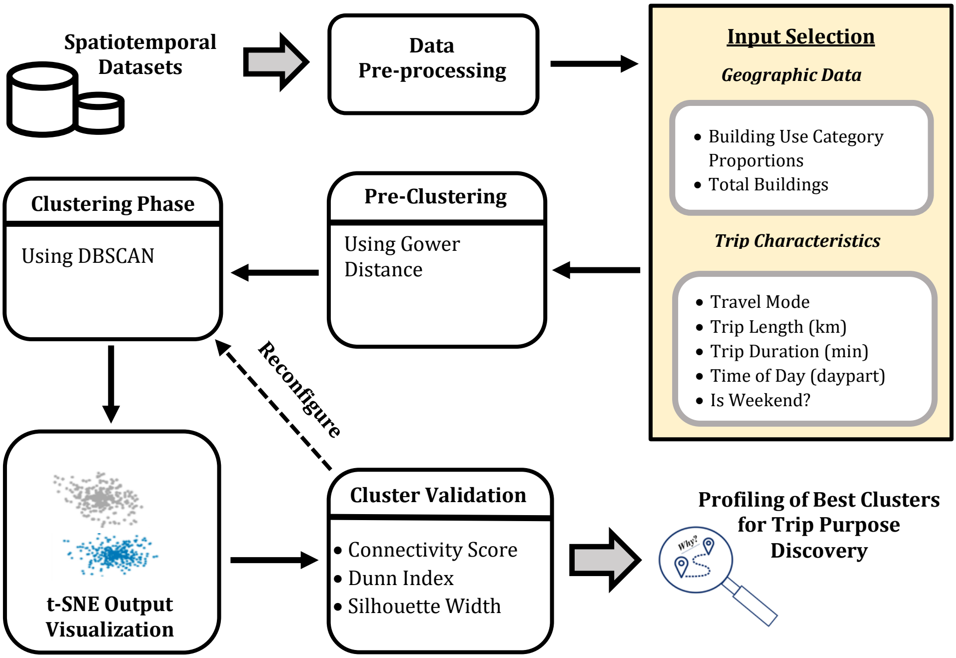


Figure 5.1: Analytical framework of the clustering analysis for trip purpose discovery

5.4 Empirical analysis

5.4.1 Data pre-processing

The analysis is implemented in RStudio, an integrated development environment for the popular R programming language. First, we merge the OD tracking data with the land-use information using the corresponding cell IDs of the origin and destination cells. Trip observations with missing ("NA") values in any of the features are unsuitable for the clustering algorithms and are thus removed. Next, new variables representing the trip characteristics are derived from existing features, such as trip duration, average speed, day of the week, hour, time of day, weekday versus weekend, and peak versus off-peak. Then, heuristic rules based on common sense and regulations are deployed to filter out anomalies and noise caused by GPS errors, Modalyzer app malfunctions, and human negligence during the travel log confirmation process. We delete all trips that are less than 2 minutes or more than 90 minutes. The lower bound is a common rule to account for GPS dwell time while the upper bound deals with extreme outliers discovered in the duration distribution (Nguyen et al., 2020).

Second, we impose a lenient rule that accounts for wrongful activations of the Modalyzer application and eliminate all trips with an average speed of less than 2 km/h. Third, we establish specific rules about the maximum speed according to the transport mode, i.e. bike 45 km/h, bus 100 km/h, car 130 km/h, train 160 km/h and walking 15 km/h.

Thereafter, we adjust the building use variables. First, we create a numeric total building count variable by computing the row-wise sums of the building category counts within the trip origin and destination cells. Fundamentally, this variable is a proxy to capture the relative degree of urbanization between different cells. Furthermore, we convert the building category counts to proportions because the total number of buildings within a cell can vary significantly.

5.4.2 Input selection and pre-clustering

The final cleaned and combined dataset consists of 4,296 trip observations and 39 variables. Table 5.1 provides an overview of all the present features. The row numbers of the variables that we select as input for the clustering algorithms are colored light orange.

N	Variable	Type	Range	Mean (or "Example")
1	Mode	Nominal	5 travel mode categories	"Bicycle," "Bus," "Car," "Train," and "Walk".
2	Started at	Datetime	2018-11-19 07:50:33 to 2018-12-10 08:31:34	"2018-11-22 10:25:18"
3	Started at day of week	Nominal	7 weekday categories	"Monday" until "Sunday"
4	Started at is weekend?	Binary	1 = True, 0 = False	"1" or "0"
5	Started at hour	Numeric	0 to 23 hours	14 hours
6	Started at daypart	Nominal	4 daypart categories	"Morning peak," "Day off-peak," "Evening peak," "Evening/night off-peak"
7	Finished at	Datetime	2018-11-19 08:02:16 to 2018-12-10 08:44:05	"2018-11-22 11:09:56"
8	Finished at day of week	Nominal	7 weekday categories	"Monday" until "Sunday"
9	Finished at is weekend?	Binary	1 = True, 0 = False	"1" or "0"
10	Finished at hour	Numeric	0 to 23 hours	14 hours
11	Finished at daypart	Nominal	4 daypart categories	"Morning peak," "Day off-peak," "Evening peak," "Evening/night off-peak"
12	Length	Numeric	From 0.08 to 72.48 km	5.15 km
13	Duration	Numeric	From 2.00 to 88.05 min	11.71 min
14	Speed	Numeric	From 2.01 to 125.25 km/h	22.19 km/h
15	First	Numeric	Number corresponding to the 300 m grid	"229320"
16	Last	Numeric	Number corresponding to the 300 m grid	"242299"
17	Samecell?	Binary	1 = True, 0 = False	"1" or "0"
18	Total buildings first	Numeric	From 0 to 1934 buildings	612 buildings
19	Use residential first	Numeric	From 0 to 100%	80.63%
20	Use social gathering/restaurant first	Numeric	From 0 to 100%	2.06%

21	Use health first	Numeric	From 0 to 44.3%	0.74%
22	Use industry first	Numeric	From 0 to 100%	2.53%
23	Use office first	Numeric	From 0 to 100%	4.32%
24	Use hotel/accommodation first	Numeric	From 0 to 96.6%	0.31%
25	Use education first	Numeric	From 0 to 33.3%	0.94%
26	Use sport first	Numeric	From 0 to 100%	0.29%
27	Use shop first	Numeric	From 0 to 100%	3.57%
28	Use other first	Numeric	From 0 to 100%	3.99%
29	Total buildings last	Numeric	From 0 to 1934 buildings	612 buildings
30	Use residential last	Numeric	From 0 to 100%	80.70%
31	Use social gathering/restaurant last	Numeric	From 0 to 44.3%	2.01%
32	Use health last	Numeric	From 0 to 100%	0.75%
33	Use industry last	Numeric	From 0 to 100 %	2.52%
34	Use office last	Numeric	From 0 to 100 %	4.33%
35	Use hotel/accommodation last	Numeric	From 0 to 97.0%	0.32%
36	Use education last	Numeric	From 0 to 33.3%	0.97%
37	Use sport last	Numeric	From 0 to 100%	0.28%
38	Use shop last	Numeric	From 0 to 82.4%	3.63%
39	Use other last	Numeric	From 0 to 46.7%	4.00%

Table 5.1: Overview variables final combined dataset

Once the data are pre-processed, we apply the Gower distance to the chosen input variables to determine the pairwise degree of (dis)similarity between the trip observations. In this step, we stipulate a custom user-defined weighting scheme for the input variables to obtain more meaningful high-quality clusters. These weights are specified through trial and error, after it became evident that the default equal weights are predominately driven by the nominal and binary variables. To reduce the importance of these features, the weights of the nominal and binary variables are set to 0.5 and 0.75, respectively. The weights of the diminished land-use type proportions and length and duration variables are set to 1.1 and 1.5 to increase the sensitivity of the clustering formation to those features.

5.4.3 Clustering implementation

We employ the DBSCAN algorithm to cluster the data based on the density of the data regions in the feature space of the Gower dissimilarity matrix. As a reminder, we group objects from high-density regions into clusters and disregard observations in lower-density areas as noise. DBSCAN is executed in R through the *dbscan()* function of the *fpc* package produced by Hennig (2020). Despite having only two parameters, it is critical that they are chosen correctly to achieve good performance and obtain meaningful clusters. Consequently, this study follows the heuristic guidelines examined and validated by Schubert et al. (2017) to properly implement DBSCAN.

First, a value for the function's *MinPts* argument, which determines the minimum number of observations that should be nearest neighbors to start a cluster, needs to be set. A common rule of thumb is to set this parameter equal to at least the number of input variables. However, for high dimensional datasets that may have a lot of noise, it is recommended to set this equal to twice the input dimensionality. The input data consist of 26 features; therefore, the *MinPts* argument is set to 52. Next, the optimal epsilon parameter needs to be established. Researchers usually achieve this heuristically by plotting a sorted k nearest neighbor (kNN) distance graph. Ultimately, we find the optimal epsilon value at 0.125, which results in 12 clusters.

The output in Figure 5.2 shows that the clustering solution of this parameter setup satisfies all the criteria: the largest cluster component, cluster 3, includes only 17.85% (767/4296) of all trip observations, and only 7.64% (328/4296) of the data is interpreted as noise. Admittedly, 12 clusters may be too diversified for the data at hand. However, a larger epsilon value results in cluster partitioning with only 4, 2, or 1 clusters, which clearly violates the largest cluster component criterion. A smaller epsilon value results in partitioning that comprises more than 15 clusters, even less ideal for the profiling analysis, and more noise points than allowed by the noise criterion. Given that a desired number of clusters cannot be pre-specified in this algorithm, the DBSCAN outcome of 12 clusters is considered most appropriate.

```
dbscan Pts=4296 MinPts=52 eps=0.125
         0   1   2   3   4   5   6  7   8   9  10  11 12
border 328  46   9  45  19  10  15 48  17  24  14 249 57
seed     0 443 206 722 525 307 253 15 192  80 227 406 39
total  328 489 215 767 544 317 268 63 209 104 241 655 96
```

Figure 5.2: DBSCAN clustering output

Figure 5.3 displays the t-SNE projection of the DBSCAN clustering results. This visualization also seems to indicate a favorable clustering outcome. Most areas in the 2-dimensional graph that show distinct groups are partitioned into a single cluster. Additionally, dense groups that belong to the same cluster are located in roughly the same region. Only cluster 3 is more spread out over the feature space and interrupted by parts belonging to other clusters. The identification of noise points by DBSCAN also makes intuitive sense because only observations in less-dense regions are correctly classified as noise.

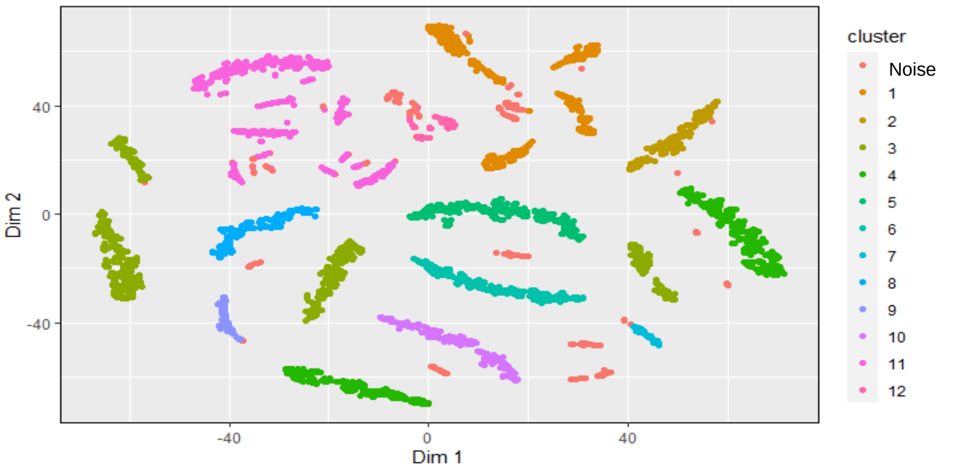


Figure 5.3: t-SNE projection with DBSCAN clusters

5.4.4 Profiling the clusters for trip purpose discovery

We examine the spatiotemporal patterns and characteristics of the travel observations residing in the DBSCAN clustering solution to uncover representative trip purposes inherent to the data. First, we extract key variable statistics of the trips captured by each cluster, the noise points, and the data as a whole and present them in Table 5.2. The table is color-coded to bring the comparative distinctiveness of the clusters to the foreground. Accordingly, we then divide this scale along seven color intervals that highlight the dispersion of a variable statistic value for one cluster relative to the others. A full color-mapping key is available below the table. The scale essentially transforms the table into a heat map that conveniently reveals which features are relatively under- and overrepresented in each of the 12 clusters.

Holistically, the table also provides a greater understanding of how the DBSCAN algorithm partitioned the data. The data are well divided over the clusters as the size component varies between containing 1.47% and 17.85% of all data. Examining the means of the physical trip features shows that the clusters represent trips of different lengths, durations, and speeds. Next, the travel mode variable appears to have had a major impact on the formations since all partitions, aside from clusters 1 and 11, are formed around a single mode. Remarkably, the bus category is only marginally represented by cluster 11. Further, the clusters are completely driven by the weekday. Only clusters 11 and 12 contain trips started during the weekend. Interestingly, the daypart variable is more diversified, as multiple clusters contain trips that started and finished throughout the day. However, each category is fully represented by one cluster at least once. Similar to the weekend variable, the clusters are completely driven by the binary samecell variable: only clusters 1 and 12 include trips that end in the same cell as the starting location. Finally, the mean building count and land-use proportions of the origin and destination cells are uniquely embodied across the clusters.

Variables		Noise	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	All
Percentage of all trips		7.64	11.38	5.00	**17.85**	12.66	7.38	6.24	**1.47**	4.86	2.42	5.61	15.25	2.23	100.00
Mean physical trip features	*Length [km]*	7.75	0.58	2.69	0.76	10.19	9.52	4.54	**17.32**	9.22	2.92	5.05	5.76	**0.39**	5.15
	Duration [min]	15.81	7.78	10.14	8.32	14.37	14.51	14.79	**17.97**	12.60	10.70	16.20	10.82	**5.35**	11.71
	Speed [km/h]	31.56	**4.52**	15.69	5.73	39.20	36.78	17.36	**59.42**	40.75	16.60	18.14	28.33	4.80	22.19
Proportion per travel mode	*Bicycle [%]*	10.06	1.02	**100.00**	/	/	/	**100.00**	/	/	**100.00**	**100.00**	10.23	/	21.72
	Bus [%]	**22.26**	/	/	/	/	/	/	/	/	/	/	1.07	/	1.86
	Car [%]	23.17	0.82	/	/	**100.00**	**100.00**	/	/	**100.00**	/	/	63.82	/	36.50
	Train [%]	31.71	/	/	/	/	/	/	**100.00**	/	/	/	4.73	/	4.61
	Walk [%]	12.80	98.16	/	**100.00**	/	/	/	/	/	/	/	20.15	**100.00**	35.31
Proportion started on the weekend *[%]*		27.14	/	/	/	/	/	/	/	/	/	/	**100.00**	**100.00**	19.55
Proportions started per weekday	*Monday [%]*	16.77	12.68	16.74	17.34	11.40	12.93	19.78	**22.22**	16.75	12.50	15.77	/	/	12.62
	Tuesday [%]	9.45	24.13	11.63	20.21	16.91	17.35	**24.63**	17.46	19.62	15.38	22.41	/	/	15.46
	Wednesday [%]	14.94	21.27	22.79	19.17	**23.90**	23.34	16.42	23.81	17.70	19.23	22.82	/	/	16.85
	Thursday [%]	13.11	20.04	22.79	20.47	22.98	23.34	20.90	19.05	**25.84**	22.12	21.16	/	/	17.27
	Friday [%]	18.60	21.88	26.05	22.82	24.82	23.03	18.28	17.46	20.10	**30.77**	17.84	/	/	18.25
	Saturday [%]	14.94	/	/	/	/	/	/	/	/	/	/	**63.21**	60.42	12.13
	Sunday [%]	12.20	/	/	/	/	/	/	/	/	/	/	36.79	**39.58**	7.43
Proportion started per daypart	*Morning peak (7AM to 10AM) [%]*	30.18	17.79	/	19.43	41.18	/	/	1.59	/	/	**100.00**	8.70	6.25	20.11
	Day-off peak [%]	14.33	36.81	**100.00**	35.46	58.82	/	/	/	/	/	/	59.24	47.92	34.17
	Evening peak [%] (4PM to 7PM)	25.30	21.88	/	29.73	/	**100.00**	**100.00**	98.41	/	/	/	18.93	20.83	28.14
	Evening/night off-peak [%]	30.18	23.52	/	15.38	/	/	/	/	**100.00**	**100.00**	/	13.13	25.00	17.57
Proportion finished per daypart	*Morning peak (7AM to 10AM) [%]*	32.32	18.20	/	19.69	40.26	/	/	1.59	2.87	1.92	**95.85**	7.48	5.21	20.00
	Day-off peak [%]	16.46	35.79	**95.35**	35.20	55.88	/	/	/	/	/	4.15	57.86	47.92	33.59
	Evening peak (4PM to 7PM) [%]	23.78	22.90	4.65	29.47	3.86	**95.27**	**98.13**	95.24	/	/	/	20.92	21.88	28.63
	Evening/night off-peak [%]	27.44	23.11	/	15.65	/	4.73	1.87	3.17	97.13	**98.08**	/	13.74	25.00	17.78

Table 5.2: Variable statistics of best cluster results

Table 5.3: Variable statistics of best cluster results—continued

Variables		Noise	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	All
Proportion starts and ends in same cell *[%]*		34.15	**100.00**	/	/	/	/	/	/	/	/	/	/	**100.00**	16.22
Mean total buildings at origin cell [count]		605	604	663	767	516	487	702	**896**	523	762	**475**	556	651	612
Mean land-use percentages at origin cell	*Residential [%]*	82.43	80.35	81.68	**75.30**	85.54	76.57	77.26	84.78	80.11	83.65	**88.19**	81.88	78.80	80.63
	Social/Restaurant [%]	1.87	2.29	2.54	**2.74**	1.20	1.84	2.59	1.77	1.71	2.46	**1.04**	2.03	2.36	2.06
	Health [%]	0.50	0.96	0.79	0.89	0.77	0.59	**0.97**	0.50	0.87	0.44	**0.43**	0.61	0.46	0.74
	Industry [%]	2.14	1.45	**0.81**	3.79	2.01	**4.01**	1.94	1.20	2.75	1.63	1.77	2.97	3.29	2.53
	Office [%]	4.31	5.51	4.87	5.62	2.63	6.68	**7.00**	3.80	3.67	3.03	**1.17**	2.44	4.66	4.32
	Hotel/Accommodation [%]	0.28	0.24	0.24	0.23	0.24	0.30	**0.20**	0.35	**0.68**	0.29	0.27	0.47	0.60	0.31
	Education [%]	0.44	1.81	1.55	1.49	0.54	0.64	**2.20**	0.24	0.78	0.41	**0.17**	0.26	0.24	0.94
	Sport [%]	0.27	0.37	**0.08**	0.24	0.19	0.64	0.09	0.12	0.22	**1.11**	0.30	0.29	0.21	0.29
	Shop [%]	2.98	2.97	2.89	4.81	2.63	3.25	3.38	4.67	3.84	4.06	**1.51**	4.28	**5.52**	3.57
Mean total buildings at destination cell [count]		632	604	712	**771**	533	483	588	733	**475**	671	568	557	651	612
Mean land-use percentages at destination cell	*Residential [%]*	83.46	80.35	82.46	76.00	78.29	84.41	**88.18**	**66.88**	87.70	87.48	74.91	82.05	78.80	80.70
	Social/Restaurant [%]	1.74	2.29	2.44	2.49	2.00	1.50	1.23	**2.91**	**0.76**	1.32	2.89	1.88	2.36	2.01
	Health [%]	0.72	0.96	0.73	0.82	0.99	0.54	**0.30**	0.37	0.68	0.39	**1.25**	0.56	0.46	0.75
	Industry [%]	1.52	1.45	**1.30**	3.80	3.15	2.33	1.45	**6.46**	1.97	1.87	1.55	2.89	3.29	2.52
	Office [%]	3.67	5.51	3.82	5.54	5.65	2.52	**1.53**	7.78	2.61	2.09	**8.15**	2.54	4.66	4.33
	Hotel/Accommodation [%]	0.23	0.24	0.25	0.21	0.26	0.39	0.26	**0.12**	**0.74**	0.24	0.17	0.54	0.60	0.32
	Education [%]	0.51	1.81	1.08	1.31	1.25	0.39	0.18	**0.10**	0.29	0.27	**3.07**	0.26	0.24	0.97
	Sport [%]	0.24	0.37	0.14	0.24	0.44	0.30	**0.45**	0.44	0.09	0.21	**0.07**	0.29	0.21	0.28
	Shop [%]	3.32	2.97	3.69	4.74	3.25	3.34	2.66	**8.00**	**1.96**	2.78	2.41	4.24	5.52	3.63

Minimum — Far Below — Slightly Below — Average — Slightly Above — Far Above — Maximum

Note: Color mapping key based on the (row-wise) feature range scaling process

Before continuing, it is also worthwhile to examine the noise. It is not indicative of any evocative trip purpose patterns per se because most variables are average. Nonetheless, a manual inspection of the values suggests that the noise is at least partly driven by public transport travel modes. In fact, the train (31.37%) and bus (22.36%) travel modes comprise more than half of the trips (53.73%) specified as noise. This explanation is further supported by the far-above-average mean duration of the trips contained in the noise and the fringe representation of the bus travel mode by the 12 main clusters.

The trip purpose labeling of each cluster is a predominantly subjective process that can be approached in a plethora of distinct ways with varying degrees of complexity, each leading to (slightly) different outcomes. To approach this systematically, this study proposes the creation of a summary table that mainly utilizes the more dominant variable statistics to describe the distinctiveness of each cluster and ultimately extract the purpose. Table 5.4 and Table 5.5 are an implementation of this method with the labeled purposes attached.

With regards to the tables, please note: For the weekday and land-use configuration variables, the weekday or land-use belonging to the highest color intervals are marked in bold. For example, cluster 4 has a residential land-use in the origin that corresponds to the maximum interval scale and is thus always marked in bold. Additionally, it has a health and sport land-use in the destination cell. The health and sport land-use corresponds to the slightly above average color interval while sport corresponds to the "far above" color interval. Accordingly, only sport is marked in bold to indicate a higher level of representation in the cluster.

C	Physical Description		Mode	Weekend?	Weekday	Daypart	Same Cell?	Urbanization Level (Building Count)		Land use (cell function)		Purpose Label
								O	D	O	D	
1	*Length* *Duration* *Speed*	0.58 7.78 4.52	Walk	No	-Tuesday -Wednesday	-Mixed: mainly day off-peak	Yes	Average	Average	-Social/Restaurant **-Health** -Office **-Education**	- Social/Restaurant -Sport	*“Going for lunch during a break from study or work”*
2	*Length* *Duration* *Speed*	2.69 10.14 15.69	Bicycle	No	-Tuesday -Wednesday -Friday	- Day off-peak	No	Average	Far Above	-Social/Restaurant	-Residential - Social/Restaurant	*“Going home after having lunched at a restaurant”*
3	*Length* *Duration* *Speed*	0.76 8.32 5.73	Walk	No	-Tuesday -Wednesday	-Mixed: mainly day-off-peak	No	Slightly above	Max	- **Social/Restaurant** -Health -Industry -Office -Shop	- Social/Restaurant	*“Going for lunch during a midweek day out in the city”*
4	*Length* *Duration* *Speed*	10.19 14.37 39.20	Car	No	- **Wednesday** -Thursday -Friday	-Morning peak -Day off-peak	No	Far below	Far below	**-Residential**	-Health **-Sport**	*“Leaving home to go sporting or to get a health check-up”*
5	*Length* *Duration* *Speed*	9.52 14.51 36.78	Car	No	-Tuesday, -Wednesday	-Evening Peak	No	Far below	Far below	**-Industry** -Office	**-Residential**	*“Going home after the workday ends”*
6	*Length* *Duration* *Speed*	4.54 14.79 17.36	Bicycle	No	-Monday **-Tuesday** -Thursday	-Evening peak	No	Average	Average	-Social/Restaurant **-Health** **-Office** **-Education**	**-Residential** **-Sport**	*“Going home and/or sporting after the workday ends”*

Table 5.4: Summary trip purpose labeling best clusters

C	Physical Description		Mode	Weekend?	Weekday	Daypart	Same Cell?	Urbanization Level (Building Count)		Land use (cell function)		Purpose Label
								O	D	O	D	
7	Length Duration Speed	17.32 17.97 59.42	Train	No	**-Monday** -Wednesday	-Evening peak	No	Highest	Far Above	-Residential -Shopping	**-Social/Restaurant** -Industry **-Office** -Sport **-Shopping**	*"Going for an evening dinner at a restaurant in the city center"*
8	Length Duration Speed	9.22 12.60 40.75	Car	No	**-Thursday**	-Evening/Night off-peak	No	Far Below	Lowest	-Health **-Hotel/ Accommodation**	-Residential **-Hotel/ Accommodation**	*"Returning to the hotel or going home after visiting a tourist destination"*
9	Length Duration Speed	2.92 10.70 16.60	Bicycle	No	-Wednesday -Thursday **-Friday**	-Evening/Night off-peak	No	Average	Average	-Social/restaurant **-Sport**	**-Residential**	*"Going home after having sported or lunched"*
10	Length Duration Speed	5.05 16.20 18.14	Bicycle	No	-Tuesday -Wednesday -Thursday	-Morning peak	No	Lowest	Average	**-Residential**	-Social/restaurant **-Health** **-Office** **-Education**	*"Leaving home in the morning to go to school or work"*
11	Length Duration Speed	5.76 10.82 28.33	Bicycle	Yes	**-Saturday** -Sunday	-Mixed: mainly day off-peak	No	Slightly Below	Slightly Below	-N.A.	-Residential	*"Going home after having done recreational activities in the weekend"*
12	Length Duration Speed	0.39 5.35 4.80	Mixed: mainly Car	Yes	-Saturday **-Sunday**	-Mixed: mainly day off-peak	Yes	Average	Average	-Social/Restaurant -Industry -Hotel/ Accommodation **-Shop**	-Social/Restaurant -Hotel/ Accommodation	*"Going for lunch after shopping during the weekend"*

Table 5.5: Trip purpose labeling best cluster results—continued

The clear representation of all the relevant information ensures that the clusters can be swiftly interpreted and labeled. For instance, cluster 5 represents trips around 10 km by car, midweek, and predominately on Wednesday during the evening peak. The most dominant land-use categories in the origin cell are office and industrial whereas the most representative category at the destination is residential. Hence, we can potentially label this trip *Going home after the workday ends*.

Some cluster descriptions are more ambiguous and require assumptions. For example, cluster 7 contains trips averaging 17 km conducted by train, midweek, mainly on Monday and Wednesday during the evening peak. The origin and destination urbanization levels are relatively high, which implies trips to and from South Limburg's urban places, such as from Heerlen to Maastricht. The most representative land-uses at the origin are residential and shopping whereas at the destination the social/restaurant, office, and shopping land-uses are most dominant, followed by industry and sport. This suggests a mixed cell configuration with many possibilities. However, given that the trips take place close to the evening hours, shopping and work-related trips, although possibilities, are less likely. Accordingly, a heuristic approach that demands a single label would suggest that *Going for an evening dinner at a restaurant in the city center* is more appropriate. Therefore, this cluster already exposes some of the limitations and demonstrates why the discovered labels should be interpreted with caution.

Nevertheless, these cluster labels hint at the existence of larger categories that are present in the trip data. Table 5.6 consolidates the specific purpose labels of each cluster and groups them according to the observed larger trip purpose theme. In the end, the descriptions of the 12 clusters are categorized under four overall trip purpose classes: food-related trips, homing trips, career-related trips, and sport- and health-related trips.

N	Clusters	Specific Purpose Label	Consolidated Trip Purpose Category
1	C1	*Going for lunch during a break from study or work*	Food-related
	C3	*Going for lunch during a midweek day out in the city*	
	C7	*Going for an evening dinner at a restaurant in the city center*	
	C12	*Going for lunch after shopping during the weekend*	
2	C2	*Going home after having lunched at a restaurant*	Homing
	C5	*Going home after the workday ends*	
	C6	*Going home and/or sporting after the workday ends*	
	C8	*Returning to the hotel or going home visiting a tourist destination*	
	C9	*Going home after having sported or lunched*	
	C11	*Going home after having done recreational activities in the weekend*	
3	C10	*Leaving home in the morning to go to school or work*	Career-related
4	C4	*Leaving home to go sporting or to get a health check-up*	Sport and Health-related

Table 5.6: Trip purpose labeling consolidation

In summary, we conduct this analysis to show one way to bridge the privacy gap by applying machine learning. The combination of two rich datasets and intense feature engineering with the variables from these two data sources allow us to reverse engineer trip purposes. We follow multiple steps (i.e., pre-processing, pre-clustering, DBSCAN clustering, t-SNE output visualization, and cluster profiling), and this analysis framework gives insights on an aspect that would not have been visible otherwise.

5.5 Conclusion

GPS-enabled smartphone spatiotemporal tracking data are an important improvement in urban movement data collection compared to traditional data collection methods such as pen-and-paper travel surveys or telephone-aided interviews (Chen et al., 2019; Zhao et al., 2020). Like every innovation, GPS-based data collection has advantages and disadvantages. GPS data can observe the movements of more people and across a larger geographical scope, leading to a dataset that is much larger than one collected through traditional methods. Further, data collection is automated and requires no user interaction after giving consent to being tracked. Through this automated, real-time tracking via portable devices, researchers can achieve more accuracy in their observations and thus increase the quality of their data, models, and resulting predictions.

However, because of the automated nature of the method, there is no opportunity for researchers to ask for additional information about the person being tracked. This stems from privacy regulations that do not allow connections between the GPS-based movement and contact details of the tracked person. The privacy regulations require that starting and ending points of movements be anonymized by aggregating them on a grid with a few hundred meters' radius. Therefore, the challenge brought by the innovation of GPS-based movement data collection is to figure out how we can still generate some of this information that observed individuals previously shared with researchers via surveys.

So far, we find that different machine learning techniques have the potential to bridge the gap caused by the GDPR. Most studies use supervised techniques, but these are limited because they require the ground truth to train the algorithm (Nguyen et al., 2020). Large-scale GPS datasets do not have this ground truth included because the purpose of using GPS data compared to traditional methods is that it is automated and requires zero user interaction to keep the burden on people low. However, machine learning still allows trip details to be added or reverse engineered by combining multiple datasets and through clever feature engineering and smart models. Consequently, unsupervised machine learning models are the only option for trip purpose discovery.

Unsupervised models for trip purpose discovery have hardly been explored. We found only four studies that attempted to uncover trip purpose data (Bao et al., 2017; Chen et al., 2019; Pengfei Wang et al., 2017; Xing et al., 2020). We add to this elite group of articles because our dataset contains movements with five different modes of transport while the other papers include only one (i.e., taxis or shared bike rides). All four studies are set in large metropolitan areas (NYC and Shanghai), and we show in our study that trip purpose discovery is also feasible for smaller cities. Additionally, these four studies each come up with a different number of categories, but the labels for the categories are similar. Our results fit with that scheme. These papers use complex deep learning methods that might be hard to implement for practitioners. We use a less complex methodology and show that this also yields a satisfying result. Unsupervised learning methods for trip purpose discovery have massive potential for future research and simpler, more user-friendly methods to achieve the same results as deep learning algorithms.

The aim of this study is to critically assess an example of a user-friendly method to reverse engineer trip purposes from GPS-enabled smartphone tracking data. We combine static and dynamic spatial data (i.e., land-use information from the region that indicates the function of every building and anonymized GPS tracking data from smartphones on a grid), and use heavy feature engineering to capture as much as possible about the environment. Using the DBSCAN algorithm, we successfully reverse engineered the trip purposes in our dataset. The final analysis leads to 12 distinct clusters, which all have a unique combination of dominant features. Closely analyzing these features, such as mode of transport used, different time features, and characteristics of the starting and ending locations, allows us to determine a unique purpose for all movements assigned to a particular cluster. We group these 12 clusters into four broader categories: food-related trips, homing trips, career-related trips, and sport and health trips. These four labels align with the labels derived by the four other studies on unsupervised trip purpose discovery and confirm the feasibility and suitability of our method. This research shows that with a much simpler unsupervised technique than deep learning, it is possible to get a similar result and discover underlying trip purposes in tracking data equally well. The key to success in trip purpose discovery is data richness and clever feature engineering.

5.6 Impact on hyper-targeting

Data privacy regulations hinder businesses seeking to combine multiple data sources because companies can pass only anonymized raw data to third parties. Thus, integrating multiple sources into one larger data frame is hardly possible. Businesses are very limited in their ability to use various datasets to draw a clearer picture of a customer and their behavior. As a result, hyper-targeting will benefit from any form of gap bridging that researchers can develop. These are typically modern unsupervised machine learning techniques and algorithms, in which researchers use heavy feature engineering to enrich their existing dataset and add as much relevant information as possible. Ideally, the algorithm then uncovers insights that were previously unnoticed.

However, bridging data gaps caused by privacy regulations is often very close to the legal boundaries. By touching and stretching these boundaries, businesses are walking a thin line. Any additional piece of information about the customer base is highly valuable for marketers and especially for hyper-targeting, but at the same time, it is a hazardous game. If legal boundaries are violated, it can cause damages to a firm ranging from reputation loss to bankruptcy. Aside from the legal boundaries, businesses should consider the ethical implications of their practices. Some practices to bridge a data gap may be legal, but that does not automatically make them ethical or morally correct.

Trip purpose discovery advances hyper-targeting in multiple ways. First, being able to predict the trip purpose behind a GPS movement accurately without asking the user is a breakthrough for behavioral research and marketing. Currently, behavioral research relies largely on traditional data collection methods with sample size limits and a lack of observation of actual behavior in the individual's natural environment. GPS data from portable devices overcome these downfalls. Second, once we know what an individual is most likely doing in a particular area, they can be targeted much better with regard to shopping offers and similar mobile advertisements, suggestions or coupons for alternative mobility options, services complementary to their actual trip purpose, and so on.

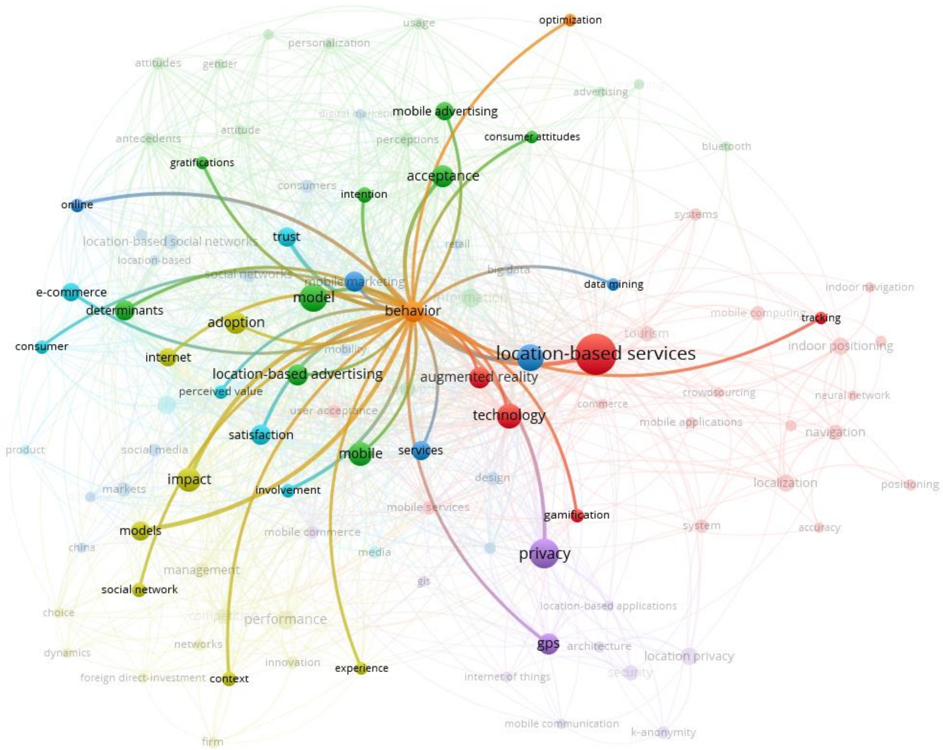
Third, recurrence of movements and movement activity patterns deepens the broader understanding of citizen behavior and allows popular locations and times for activities to be identified. With larger datasets, researchers can identify chain trips (i.e., activities that people commonly combine or pursue after another without returning home in between). Such chain trips are another option that marketers can use as a peg on which to hang their hyper-targeting strategy. Trip purpose is closely connected to behavior because it presents an active conscious choice and can be changed or influenced. This is important to notice; therefore, the next chapter considers trip purpose further but moves the mode of transport into focus and deals with mode choices for different purposes.

Transition 5: Putting the pieces together for behavioral change

In the end, we are doing the previous analyses with the goal of approximating answers to questions about how citizens behave, use their cities, and make mobility-related decisions. The answers to these questions build the foundation on which marketers can estimate the success of their LBM strategies; they can adjust current marketing efforts and identify ways to trigger changes in consumers' behavior. However, the population of a city or metropolitan area is large, and it is nearly impossible to survey a representative example properly. Therefore, we have to use 21st century data collection tools and additional big data that can be a byproduct of city monitoring or databases containing information acquired through standard bureaucratic processes. These data sources allow information from a larger population to be captured with limited intervention and effort.

So far, we have attempted to extract as many insights as possible about citizens' behavior from human movement trajectories, built environment data, and map data. This has led to a clearer understanding of the city and surrounding area, about general human movement dynamics, and about the impact of data privacy regulations on mobility research in urban areas. However, when we want to focus primarily on consumers, it is essential to know as much as possible about them. We need to go beyond anonymously tracked movements because they contain almost no information about the individual behind that movement, which is necessary to target consumers successfully. Fortunately, I was able to contact the participants of my human movement data collection study for a complementary survey, which we will use in the last chapter of this dissertation. Unfortunately, we have to isolate the survey results from the movements because matching is forbidden under the GDPR, but at least we know that we are studying the same group of individuals and that their behavior is represented in all other chapters.

The goal of Chapter 6 is to see which mode of transport people choose for various activities and what characteristics people with similar choice patterns share. In this chapter, the core and guiding thoughts of all previous chapters converge into one holistic picture. Looking at the graph on the cover page of Chapter 6, we can see that all previous chapter topics meet in the center of the network to form the behavior node. Thus, Chapter 6 addresses the fact that the consumer is still at the focal point in behavioral research, although their decisions and choices are complex (Chapter 1). Infrastructure and service performance determine their choices and routines (Chapter 2). Locations can be attractive to consumers for multiple reasons (Chapter 3), and people with similar demographics behave and choose alike (Chapter 4). Finally, data privacy restricts research opportunities (Chapter 5), and we use traditional survey data but an unsupervised clustering technique for our study.



Chapter 6

City-Based Marketing for Behavioral Change—A Holistic Approach to Consumers' Activity-Based Mode-of-Transport Choice Patterns

6.1 Introduction

What if cycling to the cinema instead of driving could finance your Spotify subscription? What if taking the train to work would give you a large discount at your favorite restaurant? This is the business idea of the recently founded European start-ups Liight (2021) and Ciclogreen (2020). Citizens use the start-ups' smartphone apps to track their own sustainable movements and collect virtual currency credits for sustainable behavior. Once citizens have collected enough credits, they can use this virtual currency to purchase subscriptions or get vouchers for participating local businesses. Policymakers see potential in this new way of triggering sustainable transport mode choices; thus, national and international sustainability agencies as well as regional and national governments support these start-ups. Because these apps have only recently been launched, the gamification and reward approach has yet to be proven successful, but their goal of encouraging sustainable transport choices is already widely recognized. Aside from the sustainability goals, marketing can use such apps to attract new customers, retain existing ones, or increase their frequency of return to a shop through vouchers and continuous reminders about offers through the app. In addition, the app provider and participating companies learn about the mobility behavior of their customers and can include it in their marketing strategies.

Mobility is an omnipresent phenomenon in our lives, but its impact is frequently unnoticed and under-researched. However, policymakers and marketers alike need to understand why citizens behave the way they do because that builds the basis for developing strategies to achieve a shift in mobility behavior (Luce, 2012). The wealth of vehicles and mobility services available makes people's trip choices complex, as not every mode is suitable for every activity. When people pursue urban activities, they make multiple decisions including transport mode, destination location, and the timing of their trips (Dellaert et al., 2008). People's personal needs and preferences add to the complexity of trip decision-making (Dellaert et al., 2014; Schoenau & Müller, 2017). Only the combined information of transport mode choice, activity, and personal situation can truly shed light on how citizens currently use transportation. Therefore, it is important to study these three components *simultaneously* to determine who uses which modes, and for which activities.

In this study, we consider mobility decision-making from a behavioral perspective and prioritize mobility behavior as the newest interest field in LBM, which combines all previously studied components to achieve the most complete picture of the target audience and to develop the most suitable strategies to meet the marketing goal. Accordingly, we take a citizen-centric approach to analyzing people's transport mode choices and detecting similarities and patterns among their individual choices. We hypothesize that transport mode choices and activity-based travel patterns emerge from people's personal characteristics, living situations, and attitudes, and these factors help explain their decisions. In a survey, participants indicate which mode of transport they use for various activities, and we leverage this information to (1) identify clusters of people with similar mode choice patterns, (2) analyze these clusters for similarities in demographics and personality, and (3) analyze citizens' trip choices by cluster. The resulting insights into the links between activities and modes of transport can help society move toward more vital, accessible cities and indicate solutions to market-reduced reliance on motorized transport in cities. Our findings also offer implications for businesses and mobility strategists.

6.2 Background

Human mobility is a complex, multidisciplinary topic that has been studied from various perspectives, such as economics (Becker, 2013; Schneider et al., 2013), psychology (Koppelman & Pas, 1980; Luce, 2012), sociology (McPherson et al., 2001), ecology (Goodchild & Janelle, 1984), geography (Hägerstrand, 1989; Harvey & Taylor, 2000; Yu & Shaw, 2008), and urban studies (Axhausen et al., 2002; Ben-Akiva & Bowman, 1998). Extant research identifies four main drivers of mobility behavior and transport mode choice: socioeconomics and demographics, instrumental factors, geographic factors, and trip purpose (Carrasco et al., 2008). We address these factors in the context of shopping and sustainability.

6.2.1 Transport mode choice

Deciding on the mode of transport is a complex but crucial component of trip choice. Basic socioeconomic and demographic attributes, such as age, gender, occupation, household size, income, and car ownership, are highly relevant (De Jong et al., 2004; Garikapati et al., 2016; Hunecke et al., 2010; Kurniawan et al., 2018). Noting debates about age as an indicator, many researchers prefer life cycle stage as a more robust variable to explain mode of transport choice (Hunecke et al., 2010). According to Garikapati et al. (2016), people change their preferences and behavior as they mature and move on to new life cycle stages; when these shifts are accounted for, behavioral differences between generations diminish.

Studies of whether lifestyle, values, and attitudes intrinsically motivate mobility behavior and mode choice have produced contradictory research findings. Whereas Garikapati et al. (2016) find that after accounting for demographic differences, lifestyle and values do not affect mode choice, Kurniawan et al. (2018) assert the opposite; their conceptual model includes values such as safety, health and well-being, the meaning of emerging transport technologies, and the meaning of owned versus shared vehicles.

Transport mode choice depends on geography and infrastructure too, because movements occur in spaces that are defined by infrastructure and land-use (Noulas et al., 2012). Trip origin and destination matter equally; a significant number of trips start or end at home. Therefore, home location affects mode choice and may eliminate mode options (Manaugh et al., 2010). For example, residents of suburban areas choose private cars more frequently than residents of urban areas, possibly because they have limited access to public transport, and this choice is independent of travel distance (Hunecke et al., 2010). Whenever people are familiar with their surrounding areas, they include habits and previous experiences in their decision-making processes (De Ceunynck et al., 2011).

The activity to be pursued determines trip purpose, which is driven by a person's state of mind (Bell, 2012). Factors such as time constraints, stress levels, and hunger lead to reactive movements; flexible leisure activities instead lead to active movements. Job availability in urban areas or close to residential areas is another important aspect. Many people must commute because they do not live close to their workplaces. Limited transit availability in suburban areas and distance often leave individual motorized transport as the only option (Giuffrida et al., 2021). Consequently, pollution and congestion during rush hour are common problems in urban areas (Manaugh et al., 2010). Aside from home/work location discrepancies, the average trip length is proportional to the density of a city, because larger cities have a higher store density and thus require shorter distances (Noulas et al., 2012). Overall, people's movements are active behavioral choices that rely on an interplay between trip purpose and transport resource accessibility.

6.2.2 Activity-based travel patterns

According to behavioral economic theory, humans are rational beings, so their choices are not random. Several studies have established that "humans follow simple reproducible patterns" (Gonzalez et al., 2008); their movements are particularly defined by time and space (Eagle & Pentland, 2009). People spend most of their time in a few locations and the rest across a broad variety of places, ranging from 5 to 50 different locations (Gonzalez et al., 2008; Jiang, Ferreira, & González, 2012). These observations of dominant places and predictable returns to them are quite consistent and especially significant at 24-, 48-, and 72-hour points in time (Gonzalez et al., 2008). Song et al. (2010) detect 93% predictability in the mobility patterns of mobile phone users. However, all these studies are based on aggregated, anonymized mobile phone data, so predictions about trip purpose or who is engaged in the movements are not possible. For this reason, researchers have attempted to understand the connection between activities and mobility patterns by using other data sources. In a study of people's activities and geographic locations over 24 hours, Jiang, Ferreira and González (2012) identify eight clusters, such as regular workers, early-bird workers, students, and overnight adventurers, without accounting for mode of transport.

While there is no complete picture of activity–movement patterns yet, research has examined various components that play a role (Ribeiro et al., 2020). Malokin et al. (2019) study mode perceptions and find that convenience, followed by comfort and benefit/cost, has the greatest impact on mode choice. People also consider whether there are opportunities to be productive while traveling, such as by using their commuting time to complete other tasks. De Ceunynck et al. (2011) find the benefits of efficiency and freedom to be important, and they often relate to the flexibility and travel time attributes. Haustein and Hunecke (2007) classify the attractiveness of a transport mode according to the four dimensions of autonomy, excitement, status, and privacy.

Some studies divide trips into work journeys and leisure journeys. Instrumental features of a transport mode, such as time and cost, are generally important for work journeys, whereas affective features, such as convenience, relaxation, or freedom, are favored for leisure trips (Anable & Gatersleben, 2005). Steg et al. (2001), in their study of the drivers of car usage, show that in addition to costs and driving conditions, excitement and prestige are important. After all, context significantly affects attributes and benefits, such that people's final decisions on modes of transport and locations combine context-dependent features of available alternatives and personal needs (Dellaert et al., 2014; Innes et al., 1990). However, only a few people consider the environmental aspects of different modes of transport (De Ceunynck et al., 2011).

6.2.3 Shopping and mobility

When making shopping trip decisions, consumers first develop simplified mental representations of their trips based on their previous experiences (Dellaert et al., 2014; Dellaert et al., 2008). When people become familiar with their surrounding areas, they include habits and previous experiences in their decision-making processes (De Ceunynck et al., 2011). As a second step, they incorporate attributes, benefits, and context-dependent features (Arentze et al., 2014). Attributes (instrumental factors) include speed, travel time, type of store, or distance; benefits include convenience, efficiency, and relaxation; and context-dependent factors include weather and location. For mode choice, the benefits of efficiency and freedom are important and often relate to the attributes of flexibility and travel time, as well as ease of parking (De Ceunynck et al., 2011). Dellaert et al. (2008) find an explicit link between attributes and benefits, in which a benefit is usually influenced by several attributes. Moreover, context significantly affects attributes and benefits, such that people's final decisions on modes of transport and locations combine context-dependent features of the available alternatives and personal needs (Dellaert et al., 2014; Innes et al., 1990).

Trip timing is another widely studied component of trip decision-making. Researchers first studied commuter trip timing, with interest in non-work trip timing arising much later (Bhat & Steed, 2002). The crucial difference between the two trip types is the temporal flexibility of non-work trips. Consumers connect shopping activities to work schedules, as well as family and household schedules; they shop at the most convenient times of the day or week, and their shopping trips often feature chains of multiple stops (Bhat & Steed, 2002). Travel costs play only a minor role in shopping trip decisions and mode of transport choice in particular; only a few people consider the environmental aspects of different modes of transport (De Ceunynck et al., 2011).

6.2.4 Sustainability in mobility

Sustainable mobility aims to redesign cities to increase quality of life by decreasing car ownership (Banister, 2008). Because each city is different, municipalities require innovative ideas (Banister, 2011). Many governments and municipalities have the long-term goals of providing more sustainable transport systems and generating greater environmental consciousness among their populations. Current initiatives include redesigning existing infrastructure, improving public transport services, and changing education and labor policies to create more flexible work and school hours (De Witte et al., 2008; Kurniawan et al., 2018).

Some researchers examine people's sentiments toward sustainability and various modes of transport, which can be used to create targeted campaigns (Barr & Prillwitz, 2012; Schoenau & Müller, 2017). Prillwitz and Barr (2011) focus on daily travel segmentation and arrive at four clusters: "persistent car users," "frequent car users," "constrained public transport users," and "consistent green travelers." They find that compared with other clusters, the average age of consistent green travelers is lower, and 58% of the members of this cluster live without children, 21% have one child, and 16% live with two children. Millennials' increasing interest in the sharing economy and on-demand mobility services reflects this profile (McDonald, 2015; Polzin et al., 2014). Barr and Prillwitz (2012) segment people according to their attitudes; they identify "addicted car users," "aspiring green travelers," "reluctant public transport users," and “committed green travelers,” with only political views and geographic locations of homes differing among those clusters. Although these studies find that attitudes toward modes of transport and sustainability have crucial effects on daily mobility behavior (Prillwitz & Barr, 2011), the comparison of the results of daily travel segmentation and attitudinal segmentation shows that habit has a larger influence on sustainable behavior than intention (Schoenau & Müller, 2017).

Although all these studies are relevant to the exploration of people’s mode choices in combination with their activities, they mostly focus on attitudes toward only one transport mode, or on sustainability in general, or they examine activity patterns without accounting for mode choice. Therefore, a more holistic understanding of the drivers of mode choice in combination with the underlying trip purpose is needed. This chapter covers precisely this topic by looking at multiple modes of transport simultaneously and including the pursued activity in the analysis. As Figure 6.1 shows, this study focuses on movements and activity equally and observes individuals rather than populations as a whole.

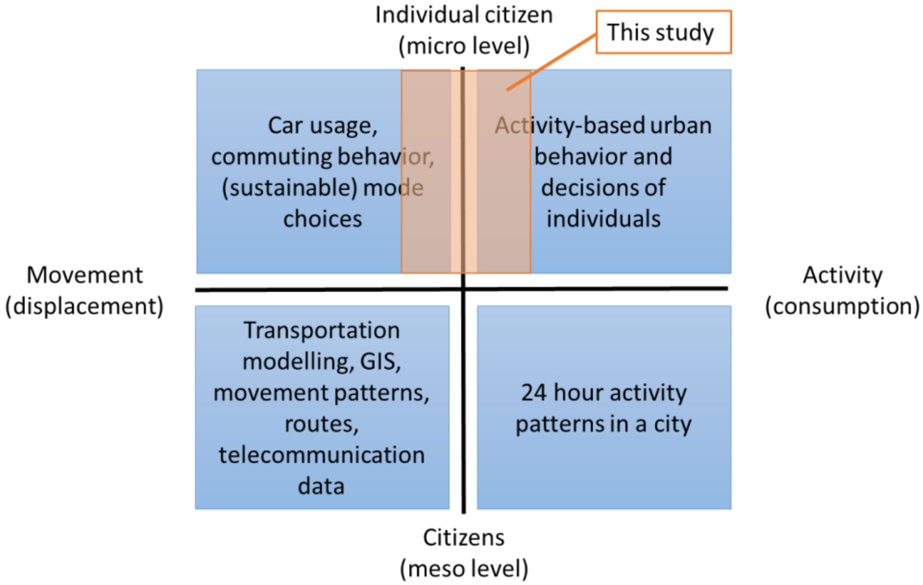


Figure 6.1: Positioning of this research from a holistic view

6.3 Methodology

6.3.1 Data collection

We collected responses to a survey as part of a larger data collection effort in November 2018. Participants voluntarily agreed to be tracked via a smartphone application for three weeks. We invited the tracking exercise participants to fill in a complementary survey at the end of the observation period.

The core part of the survey focused on activity-based transport mode choices. Participants indicated which mode of transport (car, bicycle, walking, train, bus) they choose most frequently for a variety of different activities. According to a framework of activity types used in urban studies (Axhausen et al., 2002; Bowman & Ben-Akiva, 2001), we selected eight activities or trip purposes: work/education, meeting friends, going to sports, having dinner at a restaurant, grocery shopping, shopping for large purchases, partying with friends, or engaging in entertainment such as cinema and theater visits. Participants also answered questions on environmental awareness and willingness to act, expressing opinions, beliefs, and attitudes about the environment (Stone et al., 1995). Environmental consciousness scored $x_{env} = 5.076$ [$s_{env} = 0.885$] out of 7 (where 1 indicates low environmental consciousness and 7 high consciousness). Personality traits related to mobility behavior are sociability (defined as enjoying spending time with others in general and after work), efficiency (defined as planning ahead, sticking to plans, and finishing what was started), and orderliness (defined as doing tasks in order and following schedules and routines). Their corresponding scores were $x_{soc} = 3.768$, $x_{eff} = 5.214$, and $x_{ord} = 4.821$ out of 7, respectively [$s_{soc} = 0.914$; $s_{eff} = 1.012$; $s_{ord} = 0.874$] (Goldberg, 1992). Finally, the survey included socioeconomic and demographic questions.

Concept	Scale Item
Environmental consciousness (Stone et al., 1995)	• Economic growth should take precedence over environmental considerations. • The amount of energy I use does not effect the environment to any significant degree. • There is nothing the average citizen can do to help stop environmental pollution. • I would not carpool unless I was forced to. It is too inconvenient. • My involvement in environmental activities today will help save the environment for future generations.
Sociability (Goldberg, 1992)	• I cannot be without the company of others. • I enjoy silence. • I like to be alone after a stressful day at work. • I enjoy being on the go.
Efficiency (Goldberg, 1992)	• I make plans and stick to them. • I find it difficult to get down to work. • I frequently forget to do things. • I finish what I start.
Orderliness (Goldberg, 1992)	• I do things in a logical order. • I work according to a routine. • I leave my belongings around. • I follow a schedule to complete my tasks.

Table 6.1: List of survey items for personality traits and environmental consciousness

Our dataset consisted of information related to 113 residents of Maastricht and South Limburg. Those who did not live within the city boundaries indicated that they commute into the city frequently for work or leisure purposes. We considered it important to limit the spatial scope of our study because urban activity and mobility are products of personal choices and infrastructural circumstances. The sample included 54% men and 46% women; 7.1% were 18–25 years old, 24.8% were 26–40 years old, 53.1% were 41–59 years old, and 15% were 60 years or older. Among the participants, 110 had driver's licenses. We further grouped postal codes into suburbs of residence.

With regard to the mode distribution of each trip purpose, we observed no clear trends, and each purpose had a unique distribution. In general, cycling and driving were the most attractive modes of transport, whereas participants walked or took public transport less frequently. However, they chose to walk twice as often as they chose to take public transport. Participants in our sample favored bicycles over cars for going to work or school. Distributions for grocery shopping and "regular" shopping were comparable; both indicated a clear preference for doing these activities by car. The distribution of modes for social activities, such as visiting friends, having dinner in restaurants, and engaging in entertainment, was split almost evenly between bicycle and car use, with a recognizable amount of walking and little public transport use.

6.3.2 Data analysis: Patterns and clustering of mobility and activity

New forms of data in transport and mobility have sparked the interest of researchers from other fields, including computer science, in which researchers seek to model and predict how humans move in geographical space (Calabrese et al., 2013; Coscia & Hausmann, 2015; Wang et al., 2011). In general, humans follow simple patterns with regard to the locations, timing, and frequency of their visits (Gonzalez et al., 2008). Such patterns are probably caused by the daily routines observable in activity distribution over 24 hours (Jiang, Ferreira, & González, 2012). In this study, we combine the two factors of activity and mode of transport and perform k-means clustering on the activity/mode choice matrix to identify population segments according to their similar choice patterns.

We use k-means clustering, an unsupervised machine learning method. The algorithm divides a data set into meaningful clusters according to a set of features and then minimizes the distances of data points within one cluster while maximizing the distance between clusters (Likas et al., 2003). To group all cases of similar patterns, we convert the categorical data of the activity/mode choices into a binary matrix. This method, proposed by Ralambondrainy (1995), has been affirmed by Ding and He (2004); Jiang, Ferreira and González (2012); Zha et al. (2002).

6.4 Empirical analysis

Using the binary matrix with eight activities, we identify a completely separated cluster and three clusters that overlap but clearly differ from one another (Figure 6.2). Cluster 4 is larger than the other three clusters; it contains 45 participants (39.8%), whereas clusters 1, 2, and 3 contain 27 (23.9%), 25 (22.1%), and 16 (14.2%) observations, respectively. Age and gender are balanced across all four clusters. Mode of transport is distinct in each cluster, such that we observe one cluster of people who cycle to work but drive when pursuing leisure activities, one cluster of drivers, one cluster of (mainly) walkers, and one cluster of cyclists. In this section, we describe the four clusters that emerged.

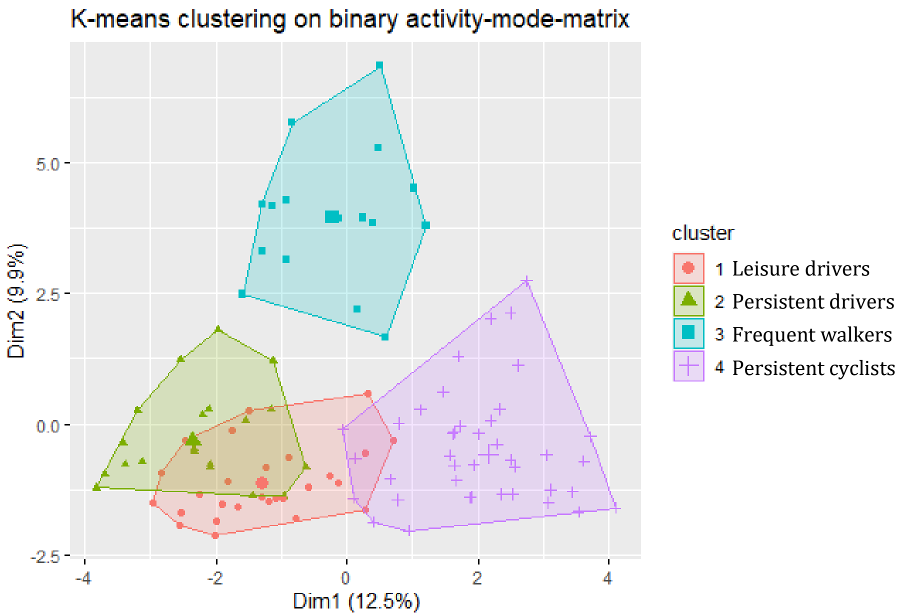


Figure 6.2: K-means clustering results

6.4.1 Cluster of "leisure drivers"

This cluster consists of people who bike to work and use their cars for leisure trips. In this group, 66% are 41 to 59 years of age, slightly higher than in the overall sample, and 59.25% are men. Leisure drivers typically live in suburban neighborhoods or rural regions. We assume they have decided to live closer to their workplaces than the city center, where many leisure activities take place. These locations allow them to cycle to work but require them to drive to social activities such as meeting friends, dining, or going to the cinema (Table 6.2).

However, considering that leisure drivers make their daily commute to work by bicycle, they are not completely averse to cycling. They can be motivated to use their bicycles to get to leisure activities. Yet we find significantly lower scores for sociability and environmental consciousness in this cluster than in other clusters, suggesting that they travel by car because they enjoy having time to themselves. Accordingly, sustainable mobility initiatives should motivate them to cycle to leisure activities, because bicycles are also private modes of transport. It will likely be difficult to motivate people in this cluster to use public transport.

Leisure drivers also tend to shop by car. Travel distance is less important to them, but easy access by car is important. To attract these consumers, businesses can use large parking lots or provide customer-only parking spots when parking is scarce. Smaller shops in dense areas might collaborate with nearby parking garages to offer special deals. Urban developers should notice that these people do all non-work trips by car. It is likely that they combine different purposes in one trip; hence, malls and clusters of different activities are attractive to these people.

6.4.2 Cluster of "persistent drivers"

For this cluster, individual motorized transport by car is clearly the dominant mode of transport. They make 79% of all trips by car; few walk to engage in sports or cycle to leisure activities (Table 6.2). Similar to leisure drivers, persistent drivers score low on sociability, partly explaining why they do not use public transport. Most of them live in the smaller towns of the surrounding area and may rely on their cars because of insufficient access to public transport. In rural areas, people often have to cover longer distances to get to places, making non-motorized transportation modes unsuitable (Figure 6.3a) and leaving individual motorized transport as the only option. Finally, because of their low levels of environmental consciousness, people in this cluster are unlikely to reconsider their modal choices as long as their personal needs are fulfilled. For them, promoting environmental awareness to trigger a change in modal choice will be costly and time-consuming. Promoting the positive features of cycling, such as flexibility, lack of parking issues, and reduced traffic, may be more effective for increasing their sustainable mobility behavior and is an alternative to external forces such as government restrictions or high taxation on car purchases and gas.

Persistent drivers are attracted by areas that offer high parking capacity and easy access to many different facilities and activities in one location. They go to work by car and are likely to combine shopping, restaurant visits, or movie nights with their work commutes (i.e., make chain trips). Therefore, service providers that seek to target this group should locate their businesses next to business districts, industrial areas, highway exits, or similar locations that commuters pass.

6.4.3 Cluster of "frequent walkers"

People in this cluster prefer to walk to all activities except work or school. For work trips of longer distances, frequent walkers choose their cars or the train. Compared with the other clusters, frequent walkers generally embrace the greatest mix of transportation modes and the greatest use of public transport (Table 6.2). This diversity in mode choice may be caused by their high levels of environmental awareness; their dominant modes—walking and public transport—are the most sustainable ways to move around. However, other factors sometimes force people to choose options that are less in line with their attitudes and beliefs but more suited to overall situations. No single category of residential area is dominant among members of this cluster. Consequently, they show that traveling sustainably using non-motorized transportation or public transport is possible, independent of their home locations.

For service providers and retailers, we note two strategies for winning frequent walkers as customers. In urban areas, shops or restaurants should be located within walking distance of their homes. Especially when grocery shopping, these consumers buy smaller quantities more frequently. By making attractive offers to returning consumers using loyalty programs and changing weekly offers, businesses can keep them satisfied. In rural areas, the business location should allow frequent walkers to combine several shops or activities into one trip. Offering home delivery services to surrounding neighborhoods could further attract these consumers.

6.4.4 Cluster of "persistent cyclists"

The largest cluster in our sample contains those who travel to all activities by bicycle. Considering that we collected the data in the Netherlands, known to be extraordinarily bicycle-friendly, this result is not surprising. Even if they cycle to social activities, however, persistent cyclists prefer to drive to grocery stores and for regular shopping tasks (Table 6.2). Most people in this cluster live within the city boundaries and tend to cover shorter distances for activities (Figure 6.3b). The few persistent cyclists who live outside the city do not go to work by car or public transport, so we conclude that some people in this cluster live close to the city center because they like to be close to social activities even if they must commute out of the city for work. However, independent of their residence, the vast majority of people in this cluster cycle to work, which shows that non-motorized transport can be an option for everyone. Their high score on environmental consciousness underlines this mindset.

Overall, we find that persistent cyclists and frequent walkers value efficiency and orderliness. Especially in Dutch cities, where city centers are compact and infrastructure features such as one-way streets, pedestrian areas, and limited parking are barriers to cars, choosing non-motorized modes of transport significantly increases people's flexibility, routing options, and efficiency. Persistent cyclists even shop by bicycle, so stores, restaurants, or entertainment facilities could consider offering "green bonuses" to reward them for using a sustainable mode of transport.

Figure 6.3: Heat map of the area covered by (a) persistent drivers and (b) persistent cyclists

	Leisure drivers (27 participants)	Persistent drivers (25 participants)	Frequent walkers (16 participants)	Persistent cyclists (45 participants)
Age	3.7% 18–25; 18.5% 26–40; 66.7% 41–59; 11.1% 60+	8% 18–25; 32% 26-40; 48% 41–59; 12% 60+	6.25% 18–25; 25% 26–40; 50% 41–59; 18.75% 60+	8.89% 18–25; 24.44% 26–40; 48.89% 41–59; 17.78% 60+
Gender	59.25% men, 40.75% women	48% men, 52% women	56.25% men, 43.75% women	53.33% men, 46.67% women
Residential area	Predominantly outer neighborhoods and surrounding region	Almost all live outside of city, in the smaller towns of the surrounding area	No clear residential area; urban, suburban, rural	Most people live within city boundaries
Activity-mode matrix (split by cluster 1/2/3/4)				

	Work	**Friends**	**Sports**	**Dinner**	**Grocery**	**Shopping**	**Party**	**Cinema**	**Sum**
Bicycle	25/0/3/26	8/0/1/34	18/4/2/24	0/3/0/35	5/5/2/15	3/1/1/23	3/1/2/24	2/0/2/27	64/14/13/208
Walking	0/0/1/1	0/2/7/2	5/7/4/7	3/2/12/3	6/2/5/4	1/2/14/6	3/2/8/3	0/0/11/0	18/17/62/26
Car	0/23/4/7	17/22/5/6	4/7/2/6	22/15/2/1	15/18/7/25	19/18/0/10	15/11/2/2	23/17/2/6	115/113/24/63
Train	1/1/6/4	1/1/1/3	0/0/0/1	1/0/1/1	0/0/0/0	1/0/0/1	1/0/2/3	0/1/0/0	5/3/10/13
Bus	0/0/1/5	0/0/0/0	0/0/0/0	0/0/0/1	0/0/1/0	3/0/0/3	1/0/0/1	0/1/1/6	5/1/3/16
NA	0/0/0/2	0/0/0/0	0/0/0/7	0/0/0/4	0/0/0/1	0/0/0/2	0/0/0/12	0/0/0/6	0/0/0/34

	Leisure drivers (27 participants)	Persistent drivers (25 participants)	Frequent walkers (16 participants)	Persistent cyclists (45 participants)
Sustainable mobility	Use sustainable transport (bicycle) only for work commute; a good start, because it is a daily activity; all other activities by car are non-sustainable. Already commute by bicycle, so should be possible to motivate them to use their bicycles for leisure activities more often	Low environmental consciousness, little sustainable transport use. Use other features, such as flexibility or lack of parking issues, to promote sustainable mobility	High environmental consciousness; walking and public transport already dominant modes; greatest mix of modes	High environmental awareness; sustainable mobility already observable; bicycle clearly dominant mode of transport
Shopping	Use car to shop; attracted by retail areas that are easily accessible and offer good parking options	All activities by car; possibility of chain trips. Possibly attracted to retail areas located along their commuter trip routes	Regularly shop on foot, often grocery shop on foot, sometimes by car; cannot buy numerous or bulky products so visit stores more frequently than other consumers. Keep these consumers satisfied so they will return	Regular shopping by bicycle, grocery shopping mostly by car, likely to be independent trip dedicated to weekly shopping for larger quantities

Table 6.2: Cluster summary

6.4.5 Results summary

In this study, we clustered 113 respondents from the south of the Netherlands according to their detailed activity-based mode choices, and we identified four distinct behavioral groups. Consumers with low levels of sociability choose to use individual motorized transportation, whereas those who emphasize efficiency and orderliness tend to cycle or walk—modes of transport associated with high flexibility for the region subject to our study. Notably, we found that residential locations influence cluster assignments: most persistent cyclists live within city boundaries, whereas most persistent drivers live in rural surroundings. Several factors may explain this phenomenon, including longer distances in rural areas, inconvenience, and limited access to (and high costs of) public transportation.

6.5 Conclusion

The aim of this chapter is to assess the benefits of combining trip purpose and transport mode choice. It also stresses the importance of looking at multiple modes of transport simultaneously. Studying choices between multiple modes of transport instead of each mode in isolation broadens the scope of mobility research. People often have strong opinions about each individual transport mode, but looking at them in comparison and having to make a choice puts things into perspective. Possible misinterpretation caused by a too-narrow view becomes most obvious when looking at leisure drivers and persistent cyclists. A study about mode choice for commuting to work would assign people of both clusters to the same group—the cyclists. However, a closer look at their choice matrix shows that cycling to work is the only choice they have in common and for all other activities, preferences deviate.

Including trip purpose in the multiple transport mode scenario adds an additional perspective to mobility behavior. Independent of the geographic trip origin and travel distance, not every mode of transport is suitable for every activity. Knowing how people get to places where they pursue a particular activity allows businesses and policymakers to include this information in their efforts to create more viable urban areas. Municipalities can discover where to improve the infrastructure for which mode of transport, and this can either support or undermine which modes are favored. Moreover, governments can use this holistic view of mobility to integrate local businesses into the innovative solution to motivate the sustainable movement presented in this chapter's introduction. Citizens record their sustainable transport use via an app and collect virtual currency for their efforts. Later, they can trade their virtual currency for a variety of options, such as free subscriptions to music and movie streaming services or discounts at participating local restaurants and entertainment businesses (Ciclogreen, 2020; Liight, 2021). This innovative app is suitable for all four identified clusters, but for different reasons. "Persistent cyclists" and "frequent walkers" are rewarded for their already sustainable behavior; thus, they are encouraged to continue choosing sustainable transportation. Further, they might share this app with like-minded people. Conversely, this new business model can trigger a change in the behavior of "leisure drivers" and "persistent drivers" because they are more responsive to indirect stimuli such as rewards and cost savings for themselves than to direct environmental awareness campaigns. The app provider can monetize the generated data and provide meaningful marketing analytics reports to participating companies, which, in turn, can use the insights to hyper-target marketing offers.

By asking the question, “How do people move for which purposes?” we provide a novel and insightful perspective that can inform both businesses and sustainability strategists.

6.5.1 Implications for businesses

Urban sensing data on mobility provide insights into the question of who pursues which activities, when, and where. Shopping trip decisions reflect the interplay of transport mode choice, timing, and shopping location (Dellaert et al., 2008). People make different trip choices according to their general mobility behavior; in our sample, most people shop by car, but holistically, their trips differ. Some people prefer chain trips and combine multiple activities in one trip, while others take individual trips for each activity and are thus more flexible in their mode choice.

In particular, we observed four distinct groups of consumers. First, some people do everything by car, possibly because driving is their only option or they love driving, but also because they prefer chain trips that combine leisure activities and shopping with their return from work. Second, there are consumers who do everything by car except for going to work; they are likely to start their trips from home rather than work. Third, the people who do everything by bicycle except shopping for groceries, which they do by car, make trips only for that purpose and are unlikely to combine them with other activities. Fourth, a segment of consumers shop for groceries and other items on foot or by bicycle. These differences are important for retailers or other businesses that target specific customer groups, because they need to reach out to these groups differently. For example, businesses can make special offers in cooperation with other nearby businesses or strategically locate their own business next to a complementary business to attract chain-trip customers.

Business location matters in the effort to target desired consumer groups. When businesses want to open new stores, they should carefully consider their locations to ensure that they match the movement profiles of their target groups. In particular, they should note “hotspots” of urban living, such as intermodal, social, or health hotspots, according to the nature of their businesses. Moreover, they should know what consumers value when making trip decisions, including store type, opening hours, quality and price range of offered services or goods, travel time, accessibility, and ease of parking (De Ceunynck et al., 2011). These attributes and benefits should serve as starting points when developing marketing strategies or new concepts of urban activity.

6.5.2 Implications for urban sustainability

Many cities are putting sustainability, sustainable and flexible transportation, and pollution and noise reduction on their agendas. Therefore, policymakers are interested in human mobility research and urban-sensing data collection (Calabrese et al., 2013; Ferrari et al., 2012). In our study, we show that people’s levels of environmental consciousness influence their mobility behavior and transport mode choices. Citizens with high levels of environmental consciousness choose to cycle, walk, or take public transport, whereas those with low levels of environmental consciousness use their cars.

Neither travel distance nor home location influences this relationship. We also find no single group that uses public transport as their dominant mode. Overall, public transport is underrepresented, compared with cycling, walking, and driving. Therefore, it is of utmost importance for city councils and municipalities to engage people in more sustainable forms of transportation. Our finding that cycling is widely accepted among participants is promising, but substantial room for improvement remains. Knowing what motivates individual mobility choices is a starting point for steering sustainable mobility behavior. Using behavioral insights, urban developers and policymakers can develop campaigns for cities to nudge people into more sustainable transport choices. A few studies have investigated options for bringing about change, including real-time behavioral feedback, social influences, and reputation building related to sustainable alternatives to driving (Gabrielli et al., 2014).

One initiative that builds on behavioral feedback, social influence, and financial rewards was tested a few years ago. The idea was that people who already cycle to work entice their colleagues to switch from car to bicycle. While new cyclists were rewarded 5 EUR per cycling day, the "recruiter" also received a financial reward for each colleague they recruited. A smartphone app registered the cycling trips, but everyone could also see their personal statistics, such as cycled kilometers and saved CO_2 emissions, via a website. On this online dashboard, colleagues could also compare their own performance with that of others, which pushed their own sustainability performance even further. The pilot project was quite successful. "Recruiting colleagues" achieved a conversion rate of 75%, and 80% of them cycled to work multiple times a week. Additionally, 90% enjoyed their new way of commuting so much that they stuck with cycling even after the end of the project, and 42% reported that they now also use their bike more frequently for other trip purposes. Participants indicated that health (41%) and financial rewards (36%) were their main motivators to cycle (Bron & Soeters, 2014). This is an example in which the activity provider (i.e., employer and the municipality) worked together to increase sustainability efforts.

Similar to the commercial strategies, sustainability strategies must match the interests of the target groups. In our sample, two clusters of people score high on environmental consciousness and already behave accordingly. To influence the others, municipalities should focus on aspects of sustainable modes of transport that are more interesting to them, such as costs, efficiency, and flexibility. Alternatively, cities might force sustainable mobility on inhabitants by imposing restrictions and laws that make driving unattractive, including lower speed limits, dedication of road lanes to public transport, increased parking prices and parking controls, and road tolls (Banister, 2008). However, to achieve voluntary switches to sustainable transport, researchers recommend involving people in the change process and giving them a voice in policy development (Banister, 2011). They also recommend informing them about future mobility visions, such as planned changes in infrastructure and the pricing structures of public transport (Fernandez-Heredia & Fernandez-Sanchez, 2020). Municipalities cannot expect people to find such information on their own; they must make the effort to spread the word (Banister, 2008).

6.5.3 Limitations

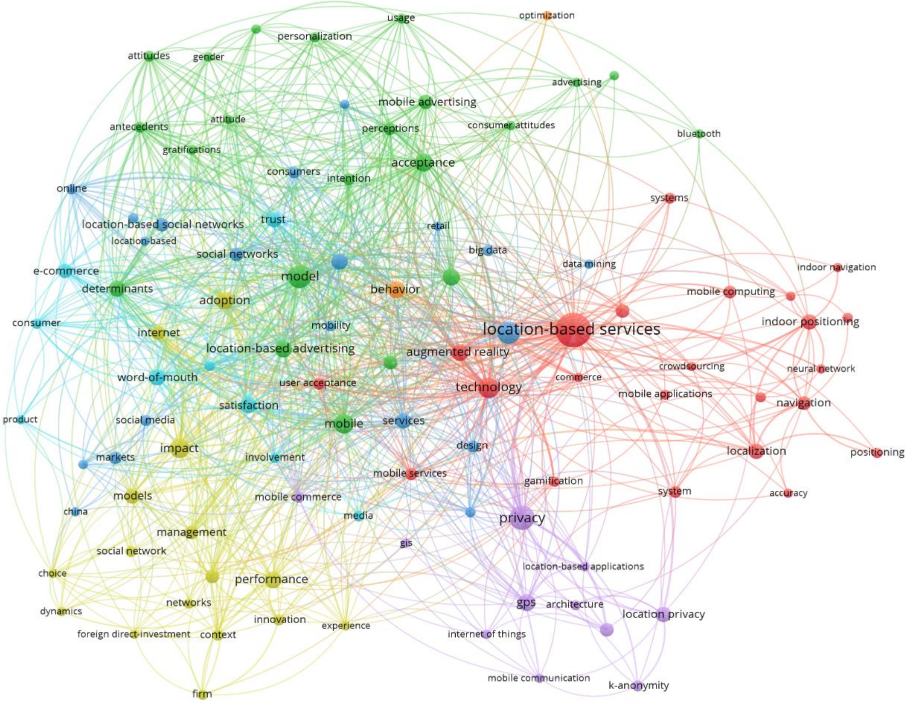
We apply k-means clustering to mobility data and mobility patterns and demonstrate that the method is suitable for these datasets, such that it leads to meaningful results. Our findings underline the importance of treating mobility behavior as a multidisciplinary topic and taking a holistic perspective. We establish activity-based transport mode choice as a distinct research area and elaborate on implications for businesses and sustainability strategists. Notably, mode choice and activity-mode patterns emerge from personal characteristics, attitudes, and living situations. However, other influential factors, beyond those we tested, are also likely, which leaves room for further research. In addition to observing where people live, it would be interesting to record where they work and how far they travel for work and other activities. Although we do not investigate the lifecycle stages or family compositions of our sample, such investigations would likely contribute to a more complete understanding of urban mobility behavior and would provide leads for encouraging sustainable behavior.

As in all studies with geographical components, our results should hold not only for the region in which we collected our data but also for areas with similar mobility cultures and infrastructural circumstances. Grindrod and Lee (2016) provide support for generalizability by showing that (social) structures within cities are similar even for cities that differ in size. Finally, the basis of our study is a mobility behavior survey, which was part of a larger data collection effort that included actual GPS tracking of movements in a city. We believe that connecting our survey results with this tracking information will enrich our data sources and allow for more precise and valid results.

6.6 Impact on hyper-targeting

This final dissertation chapter covers the most recent research topic of them all. The automated literature study in the introduction of this dissertation has shown that there is hardly any LBM research about actual behavior. This chapter adds to this new niche field and demonstrates the importance of looking at deeper behavioral levels to understand people’s choice patterns. We confirm again that more information and details on consumers’ behavior helps to produce curated, hyper-targeted offerings and better LBM strategies. Additionally, the network from the introduction points out that this niche topic of behavioral research in CBM and hyper-targeting connects to all other topics addressed in the previous chapters—or put differently, this up-and-coming domain is the melting pot where it finally all comes together. By collecting the right data from consumers and using appropriate and modern data modeling and analysis techniques, all while preserving individual privacy, we can observe how different social groups behave. On that basis, we can develop hyper-targeted offerings for behavior change toward purchasing, sustainability, and so forth to improve performance of all kinds in the city of the future.

From the study, we can learn that activities pursued in a city and the mode of transport chosen for a trip are an interesting combination with significant interrelation. For hyper-targeting, this means that both transport providers and activity providers should use the other component to make their own services more appealing to consumers. For example, businesses can give out special deals for people coming by bike or walking, or they can ensure that enough parking is available if they know that most of their customers come by car. Depending on how central a location is and how scarce parking is, one might even consider a deal with a nearby parking garage or include reservation of a parking spot during the ticket booking process (e.g., for cinemas, theaters, or restaurants with many daily reservations). For more specific hyper-targeting, companies should look not only at the activity and mode combinations but also at the behavioral clusters we identified. They differentiate the customer base further into smaller clusters based on behavior and demographics. Thus, they can serve as a hook for developing hyper-targeted marketing offers. If a business has a focus quite different from transport, shopping, and sustainability, the company can use the suggested method to develop another set of personas based on different data, which would result in more suitable clusters for the specific scenario. As mentioned before, hyper-targeting benefits most from a combination of as many aspects and sources as possible. Therefore, pushing for greater integration and research on the behavioral topic cluster will advance CBM and consequently hyper-targeting.



Discussion

City-Based Marketing—Quo Vadis?

In the introduction to this dissertation, I stated that the research goal was *to provide evidence that proxemic technologies can enhance hyper-targeting and that these technologies are vital to the dynamic perspective in marketing*. The compilation of chapters has addressed this evidence and shed light on human movement in marketing from different perspectives, assessing the impact of these perspectives on hyper-targeting. In this sequence, the chapters used different sources and forms of spatial data and different appropriate methods for analysis. Further, the implications of the research chapters are relevant to a variety of stakeholders and thus make this dissertation interesting for a broad audience from academia and industry. Each chapter presents a use case in which including spatial data is beneficial to the analysis, and in each example, I discuss how to include geographic information in marketing research. Throughout all the studies, it became clear that geography and marketing can no longer be studied in isolation—especially not in city marketing and marketing of transport services, nor in consumer behavior, citizen behavior, or LBM (including location-based advertising and mobile marketing). LBM should broaden its scope and incorporate diverse data sources and multiple perspectives like those that we see in CBM. Building on this realization, I developed an additional guiding thought for this dissertation and the focal question in this discussion chapter: *Where is the LBM field heading, and how can CBM shape the future of hyper-targeting?* To answer these questions, I will first reflect on the research field using six propositions; the first three suggest general future directions for the LBM research field, while the latter three reflect on the three CBM mismatches I identified in the introduction. In the second step, I will reflect on the future of hyper-targeting in cities, and I will close this discussion with concrete examples of hyper-targeting in different aspects of urban living.

Broadening horizons: Use of location data in marketing

Currently, research with spatial data and research on behavioral marketing are two distinct research fields with clear borders and no overlap. However, during my research projects, it became evident that marketing, consumer behavior, and consumer decision-making on one side and urban planning and human geography on the other are two complementary disciplines. Therefore, I advocate for greater integration of the two domains and more multidisciplinary research when studying LBM, consumer behavior in cities, performance of city services and businesses, urban needs of citizens, and so on. Combining the two disciplines will be mutually beneficial for both research fields. Urban planning researchers will see the effects that their researched and implemented changes to infrastructure and built environment have on citizens' behavior. Currently, this research stream either stops at the policy advice or implementation stage and largely neglects the validation stage and after-implementation observations.

Marketing research will gain a more holistic understanding of external influences on behavior. To date, marketing has been focused on the consumer itself and largely considers only variables that relate to the individual, such as demographics, socioeconomic features, personality, and opinions and preferences, but often neglects external influences and the environment. Further, such an integration of the two disciplines will lead to a clearer understanding of the power dependencies in urban areas because the three main stakeholders—municipalities, businesses, and citizens—

represent different interests and have different mechanisms for making their voices heard and pushing for their visions.

Most importantly, not integrating location data into marketing presents an information loss, and the combination of research fields can fill niches and knowledge gaps that a single discipline cannot fully grasp alone (Bernritter et al., 2021). Interdisciplinary research can uncover additional influences that individual disciplines would not consider. Furthermore, it can generate a more holistic picture of the studied scenario and specifically stress the added value of using spatial data in marketing research. I demonstrate these benefits of interdisciplinary research in all six chapters of this dissertation.

Eyeballing the purpose: Working on concrete scenarios and use cases

One of the determinants of good research is applicability and relevance. Companies or governments invest money in university research projects with the aim of receiving research outcomes with societal relevance and meaningful implications that can be transformed into noticeable improvements of any form. Especially when we study humans and their behavior, research needs to ensure that we address real, relevant, and urgent questions with the right approaches and methods (Barbosa et al., 2018). Depending on the nature of the phenomenon or problem, this can be direct data collection asking questions of a representative sample of people or via a lab experiment. Alternatively, it can be indirect data collection where actual behavior is knowingly (individual level) or unknowingly (aggregate level) observed via digital data crumbs that people generate every day when they pursue the activities and duties of their daily lives.

With the location factor included as a core component of the research, CBM has the advantage that there is automatically a clear geographic application area for the results. CBM research projects are especially suitable for collaboration with municipalities or national statistics centers and companies with an interest in citizen behavior, because these institutions share similar interests. They usually collect data about citizens that are of use to the research field, or they have the tools and resources to assist universities in the collection of applicable datasets on citizen behavior and related aspects. Such datasets contain annual national or regional travel surveys, public transport usage information from smart ticketing solutions, road traffic observations, telecommunication and GSM information from smartphones, other urban sensing data, urban development plans about infrastructure and building composition, etc.

In addition, it is vital that data collection for LBM, CBM, urban behavior, and related fields also involve citizens. It is hard to reach out to a large and representative sample of citizens, especially if one is interested in all travels around an area, including those from visitors, commuters, and tourists. However, especially when we are studying the city as a servicescape and are aiming to improve urban living, it is crucial to involve citizens in the entire research process. Citizens can be an inspiration during the idea generation phase of a research project, such as when their opinions on the current situation in a city are gathered. Further, regionally bound projects can—and need to—ask active members of the community to participate in data collection efforts, test pilots of new service concepts, and give feedback on future developments to be implemented.

Beyond that, municipalities ideally involve citizens in later stages of city projects as well, to ask for continuous evaluations of city services.

Adjusting to the 21st century: Observing actual behavior with sensory devices

From decades of consumer behavior research, we have seen that perception of a situation or intention to do, use, or buy something is a commonly used proxy that works for some scenarios and research questions. However, for other research problems, this proxy is not applicable; we know that intention does not necessarily lead to actual behavior (Ajzen et al., 2009; Warshaw & Davis, 1985). This is especially the case for complex scenarios that depend on many influential factors that tend to change rather frequently. Nonetheless, the academic literature contains numerous intention and perception studies about city behavior and transport choices, mostly based on (online) consumer surveys (De Jong et al., 2004; Garikapati et al., 2016; Hunecke et al., 2010). Even very recent publications such as Schrage et al. (2022) paper still use intention information from a survey and achieve interesting results, but their work would benefit from real-time observations of actual behavior. This was nearly impossible for a long time. Data collection based on travel diaries or pen-and-paper interviews was tedious and often led to incomplete observations, sparse datasets, and long and expensive data collection rounds. Frustrated participants frequently dropped out (Nguyen et al., 2020; Zhao et al., 2020).

Within the last ten to fifteen years, we have experienced technological development and improvement that has given urban behavior data collection a new perspective. The era of ubiquitous computing has integrated numerous sensory devices into the everyday lives of citizens, and these devices allow us to collect behavioral data from individuals in an unstructured format and with limited user interaction during the collection (Barbosa et al., 2018; Gong et al., 2014; Shen & Stopher, 2014). Aside from these advantages and improvements compared to traditional data collection methods, sensory data also have a downside (Chen et al., 2019; Gong et al., 2016; Gong et al., 2014). Leading up to the data collection, the researcher might have to purchase devices in case they cannot ask participants to use an app on their own smartphones or they have to purchase special software for data analysis. Further, research projects often work with longitudinal data, which are collected across a few weeks or months, necessitating that participants carry the observing device with them across this time period. Participants may also be asked to intervene (in)frequently or fill in complementary surveys adjacent to the sensory data. This combination of factors can make it more difficult to recruit participants. Nevertheless, the use of GPS and other sensory data in LBM and urban behavior research in the 21st century is unavoidable.

Comparing apples and oranges: Caveat in an interdisciplinary research field

The introduction of this dissertation highlighted the topic clusters that future CBM research should address, and I touched upon one topic in each of my chapters. There was an overlap of topics; over the course of this dissertation research, I observed that the chapter topics are intertwined and cannot be looked at in isolation. In the introduction, I showed that 85% of published LBM research papers stem from computer science and engineering. However, LBM clearly belongs to the social science domain; by definition, it focuses on targeted offerings of products and services while using the customers' geographic locations and service providers as core marketing strategy components (Bernritter et al., 2021). This mismatch illustrates a severe problem in the field. We cannot draw conclusions about successful marketing methods or citizen behavior from research that focuses on introducing an algorithm instead of producing social science implications. To be clear, the development of new methods and algorithms that help social scientists study citizen behavior is equally important to the social science implications. Nevertheless, they are not comparable because they aim to answer research questions from two different domains. LBM is not the only research field in which the majority of published research comes from a "foreign" domain. In fact, this is a typical pitfall in interdisciplinary research, and stakeholders drawing on the research results must be cautious. They have to pay close attention to what has been studied, why, and how to ensure that the goal, results, and implications are applicable to their own purposes. Oftentimes, marketers are not satisfied with their current LBM efforts, but they cannot find promising alternatives or advice in the academic literature. Therefore, in this dissertation, I present an attempt to overcome this mismatch between the publishing and application fields in Chapters 2 to 6. They all address social science research problems with the help of spatial data, and I derive implications for the LBM field from each of the five studies. However, much more social science research needs to be conducted to close the gap between the research and application domains.

Consumers on the move: Adding the dynamic perspective to decision-making in cities

So far, LBM and marketing research in general have paid little attention to geographic data. As mentioned at the beginning of this dissertation, researchers have considered only static information about consumers' whereabouts. This mainly includes the geographic proximity of a consumer to a company's shop, which is detected via beacons or geo-fencing. Considering only a consumer's current position limits the scope of the CBM research field and prevents deeper insights. Thus, new research projects should include dynamic data about consumer movements in the studied region, which will open up a whole new world. In particular, human trajectory data and other dynamic data on consumer movements allow us to extract where a person comes from and whether a trip is dedicated to a specific purpose or is a stop in a chain of trips. Additionally, dynamic human movements often include a variable listing the mode of transport identified for a movement. These two components—trip origin and chosen mode of transport—are crucial pieces of information in CBM research because the *where from* and *how* of a movement add a new perspective to consumer behavior in cities and urban decision-making.

From previous research, we know that people follow routines and patterns in their daily activities (Song et al., 2010; Wang et al., 2011). Further, most people visit very few locations quite heavily, and the rest of their movements are distributed across a large number of places that they visit rarely (Gonzalez et al., 2008; Jiang, Ferreira, & Gonzalez, 2012). These studies have either tested their hypotheses on a very small geographical scale, such as a university campus, using aggregated, anonymized telecommunication data, or they intended to test an algorithm and only reported these outcomes as a byproduct of their actual research (Barbosa et al., 2018; Eagle & Pentland, 2009). Still, these papers allow us to assume that peoples' preferred locations are a significant indicator for other choices and behavioral patterns.

Building on these findings, in Chapter 1, I discuss the importance of creating awareness of this mismatch between used data and available data. Further, I use this chapter to argue in greater depth why adding this dynamic perspective to LBM is so important and what type of data marketing research needs to include. In Chapters 4 and 5, I use the collected human trajectory data as a core component of my analysis and show that they deliver new insights that had been invisible. Aside from the use cases illustrated in Chapters 4 and 5, additional topics concerning dynamics in urban behavior can be explored. First, GPS data typically not only observe the origin and destination of a trip but also include data points along the way. Thus, researchers can study movement schemata and commonly used trip routings. Second, if data collection allows for an anonymous personal identifier, researchers should pay special attention to recurrence of individuals' movement patterns as well as those of different societal groups. Third, it would be interesting to study the effects of home and work locations to trip decisions, such as trip timing, chosen mode of transport, and destination. Finally, the dynamic perspective of human movements in urban areas could shed additional light on what would motivate a change in trip decisions, if we were to be interested, for example, in motivating more sustainable transport.

Local conditions drive behavior: Including the surrounding infrastructure

In line with their temporary positioning but not considered in research so far, consumers' behavior is largely influenced by the surrounding environment in which they are moving. The product and service offerings in a particular place as well as the offerings of neighboring stores determine the attractiveness of a location and the financial success of a business (Hunneman et al., 2022; Öner, 2017). Additionally, factors such as nearby public transport availability and other infrastructure components are not to be neglected, because consumers' choices are based on the available opportunities in their current geographic positions (Mejia-Dorantes et al., 2012). Consequently, human movements cannot be separated from the local built environment and physical city infrastructure. City infrastructure is easier to observe, and predictions are more stable across time because there are fewer fluctuations in the built environment and store availability compared to human trajectory streams.

In the introduction of this dissertation, I addressed this third mismatch in the focal units of analysis and noted that focusing only on the consumer and disregarding the infrastructure is a critical issue. Therefore, I explain the core problem of the mismatch in more depth in Chapter 1, while in Chapters 2 and 3 I fill this gap. In particular, the second chapter uses infrastructure data to sketch the city, understand how buildings are connected, and determine where transport service gaps that could be filled by the UAM service in the near future are located. The third chapter examines building usage data to understand the impact of the built environment composition on location popularity for citizens.

In addition to these two examples where built environment information added value to the analysis, there are more open research questions for the future that, once tackled, will further diminish this mismatch. First, researchers should dive deeper into understanding the effects of co-location and competition on trip decisions by using spatial data about buildings surrounding focal firms. Second, LBM can build on urban planning literature and examine whether multiple smaller “city centers” compared to one large center would change the trip decisions of citizens, visitors, and commuters. Finally, researchers should build on our analysis in Chapter 2 and simulate the extent to which building connectivity can be further improved for more city convenience.

The future of hyper-targeting in cities

CBM shapes the future of hyper-targeting in two ways. In the beginning of this dissertation, I mentioned that marketers can draw on five types of data sources when engaging in hyper-targeting. These are customer demographic data, customer interests, locations, online behavior, and offline behavior. Currently, research and practice on hyper-targeting uses demographics, online behavior, and sometimes customer interest information, while location and offline behavior are rarely used. However, over the course of this dissertation, I critically assessed that physical context is a crucial component for CBM and hyper-targeting alike. Static as well as dynamic location information about customer whereabouts largely strengthen the quality of the hyper-targeted marketing offers. Location information is a broad term, and at the same time, it is also tied to offline behavior. We usually consider location information such as geo-fencing, GPS data, human trajectories, public and private transportation data, built environment information, and other landscape information. CBM uses these data as well and studies people’s movement behavior in cities, frequent destinations, recurring movement patterns, popular time and location combinations, popular online and offline behavior combinations, and related scenarios. All these study results from CBM are vital inputs for developing hyper-targeted marketing offers.

The second aspect of CBM research that will shape the future of hyper-targeting is the increasing data variety. Cities are continuously implementing more sensors to glean a multitude of information about urban living. While some of these sensors and data might have very different initial purposes, they may also serve hyper-targeting well. Across all chapters of this dissertation, it became evident once more that more data means more knowledge and thus more power. Good CBM enriches the data landscape that marketers can theoretically draw from. Via different identifiers or clever feature engineering, researchers can then use these additional datasets and connect them with the original data they have about their customer bases. The more knowledge a company has about the interests, habits, needs, and options of its customers, the more concrete personas it can develop. In the end, it is the goal of a hyper-targeting strategy to have a clear understanding of customer preferences, either individually or in small groups of personas, and to deliver super-specific, curated marketing offerings that fit these preferences.

However, enriching existing datasets with external information and combining multiple datasets with personal information is risky. There is a thin line between legal and illegal data handling practices, and companies have to be careful about which datasets they use and how they merge them into a larger database. In 2018, the European Union introduced the GDPR as a legal groundwork for data handling and privacy protection of individuals (EU, 2018). While this data protection regulation is a good start, it is far from complete—it mainly addresses privacy regulations for data collection but is vague about data processing, and it does not explicitly mention law enforcement bodies and practices to ensure that companies comply with the GDPR (Wachter, 2018, 2019). Aside from the GDPR and its potential weaknesses, one always has to keep the following in mind: just because an action is legal, this does not make it ethical. Some data processing practices might be technically feasible and not strictly forbidden by law, but their implementation is questionable from an ethical perspective. This is the case when data are anonymized, but reverse engineering the identity of the individual is rather easy because of the dense information available. When the sample size or formed subgroups are too small and patterns of individuals are very explicit, privacy preservation is likely violated, at least morally. Therefore, companies working with large datasets of personal information from multiple sources must be extremely careful. Understandably, from an economic perspective, it is tempting to combine as much information as possible about the customer base to develop and deliver specific, curated marketing offers. However, there is a limit to everything, and personal privacy is one of the greatest goods in data analytics.

Impact of hyper-targeting on various aspects of city life

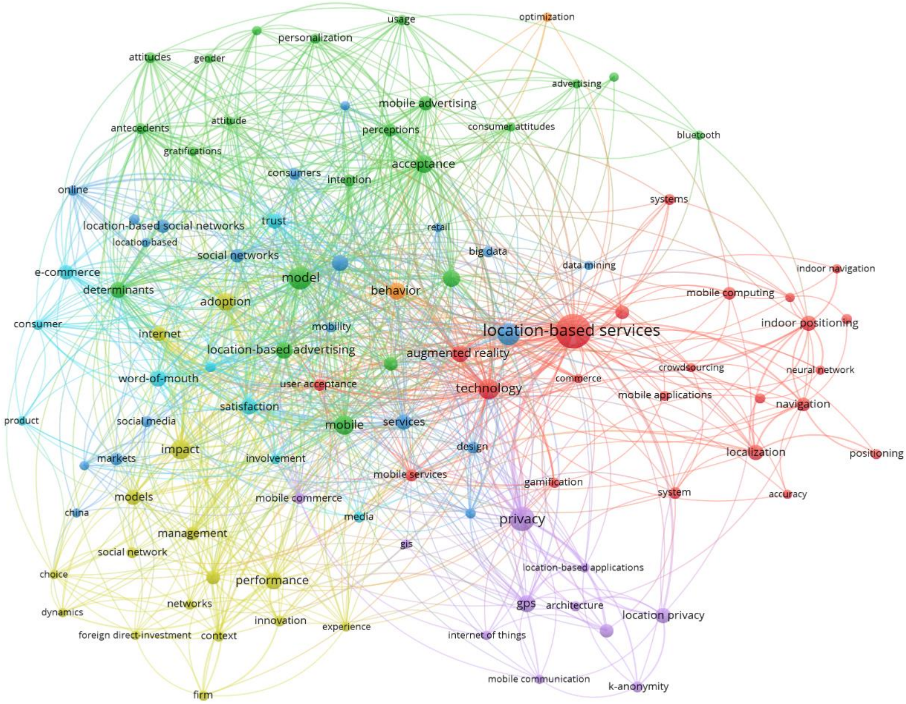
As a last step within this discussion, I would like to address some learnings from my dissertation and implications that the use of hyper-targeting has on different research fields and aspects of city life. First, Chapter 2 has taught us that we can use spatial data to create and evaluate new transport services and that we can monitor and improve the existing transport landscape. By developing an understanding of the current transport situation in a city and the personal preferences of travelers, we can use hyper-targeting to encourage citizens, commuters, and tourists to choose an alternative transport option. Ideally, this suggested alternative suits the individual's preferences better and is more sustainable than their current mode of transport choice. Sustainability is an important aspect of urban living, and its presence in discussions is continuously increasing. Transport emissions are among the biggest polluters in cities; thus, using hyper-targeting for more sustainability in the city, be it transport-related or otherwise, is a concept for the future.

Second, and probably most obviously, hyper-targeting can significantly influence the economic performance of a business and the economic well-being of a city as a whole. As touched upon in Chapter 3, CBM studies investigate, among other topics, the attractiveness of places in cities. They try to filter out why people perceive a certain location as attractive and why not all people rate a specific location as similarly attractive. While doing so, cities also develop an understanding of societal groups' preferences. Combining these two aspects, CBM allows businesses to evaluate their locations in relation to their target groups and their preferences. In a later step, businesses can use hyper-targeting and deliver super-specific offers to their target audiences and consequently increase their economic performance.

Third, the health and well-being of citizens in a city is crucial. In Chapter 4, I demonstrate how we can use CBM to uncover service gaps of different kinds in a city. Service gaps generally have negative effects on citizens' well-being. Having to search for an available service or covering long distances to reach the next service point raises stress levels among the population. Therefore, it is of the utmost importance that municipalities ensure a dense service landscape for the most important city services. Once CBM identifies these gaps and municipalities and businesses fill them, hyper-targeting can assist in increasing awareness of the services. Personal marketing offerings can help customers find the closest service when they need it, or they can generally increase awareness about new service spots that a customer might remember in the future.

Fourth, hyper-targeting affects the moral compass of a municipality and its citizens. As I mentioned previously, more information leads to more personalized offers and thus to higher sales. This is tempting to all businesses, but companies always have to ask themselves whether adding this extra puzzle piece is worth it. One should not close one's eyes to the high risks involved. Aside from legal consequences, these include—but are not limited to—acquiring a bad reputation or facing bankruptcy due to customer boycotts or dissolved partnerships. Therefore, in Chapter 5, I addressed the topic of data privacy preservation and introduced how feature engineering of non-critical datasets can help bridge the data gap.

Finally, hyper-targeting is applicable to many aspects of urban living. In Chapter 6, a combination of demographics, interests, locations, and offline behavior is used to generate clusters of citizens and their behavior. Such clustering exercises work with a variety of input datasets and can consequently address many aspects of urban living. Additionally, knowing about offline behavior gives marketers the advantage of placing advertising messages online and offline where their target audiences will receive them. The more precise and detailed the behavioral clusters are, the more powerful and effective the curated, hyper-targeted offerings will be, and the higher the return on the marketing investment.


optimization
usage
personalization
attitudes
gender
advertising
mobile advertising
antecedents
attitude
perceptions
consumer attitudes
bluetooth
gratifications
consumers
intention
acceptance
online
systems
trust
location-based social networks
retail
location-based
big data
social networks
data mining
indoor navigation
e-commerce
model
determinants
behavior
mobile computing
adoption
indoor positioning
consumer
mobility
location-based services
internet
location-based advertising
augmented reality
crowdsourcing
neural network
word-of-mouth
commerce
user acceptance
technology
mobile applications
satisfaction
navigation
product
social media
mobile
services
impact
design
localization
positioning
markets
involvement
mobile services
gamification
system
accuracy
models
mobile commerce
china
media
privacy
management
gis
social network
location-based applications
choice
performance
networks
gps
architecture
location privacy
dynamics
innovation
foreign direct-investment
context
experience
internet of things
mobile communication
k-anonymity
firm

References

Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2015). Unraveling the Personalization Paradox: The Effect of Information Collection and Trust-Building Strategies on Online Advertisement Effectiveness. *Journal of Retailing, 91*(1), 34-49.

Ajzen, I., Czasch, C., & Flood, M. G. (2009). From Intentions to Behavior: Implementation Intention, Commitment, and Conscientiousness1. *Journal of Applied Social Psychology, 39*(6), 1356-1372.

Akaka, M. A., Vargo, S. L., & Schau, H. J. (2015). The context of experience. *Journal of Service Management*.

Ameen, N., Tarhini, A., Shah, M., & Madichie, N. O. (2021). Going with the flow: smart shopping malls and omnichannel retailing. *Journal of Services Marketing, 35*(3), 325-348.

Anable, J., & Gatersleben, B. (2005). All work and no play? The role of instrumental and affective factors in work and leisure journeys by different travel modes. *Transportation Research Part A: Policy and Practice, 39*(2-3), 163-181.

Anderson, S. P., & De Palma, A. (1988). Spatial price discrimination with heterogeneous products. *The Review of Economic Studies, 55*(4), 573-592.

Anderson, S. T., & West, S. E. (2006). Open space, residential property values, and spatial context. *Regional Science and Urban Economics, 36*(6), 773-789.

Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention, 41*(3), 359-364.

Andersson, M., Klaesson, J., & Larsson, J. P. (2014). The sources of the urban wage premium by worker skills: Spatial sorting or agglomeration economies? *Papers in Regional Science, 93*(4), 727-747.

Andrews, M., Luo, X., Fang, Z., & Ghose, A. (2016). Mobile ad effectiveness: Hyper-contextual targeting with crowdedness. *Marketing Science, 35*(2), 218-233.

Andrienko, G., Andrienko, N., Chen, W., Maciejewski, R., & Zhao, Y. (2017). Visual analytics of mobility and transportation: State of the art and further research directions. *IEEE Transactions on Intelligent Transportation Systems, 18*(8), 2232-2249.

Andrienko, G., Andrienko, N., Dykes, J., Fabrikant, S. I., & Wachowicz, M. (2008). Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research.

Andrienko, N., & Andrienko, G. (2013). Visual analytics of movement: An overview of methods, tools and procedures. *Information visualization, 12*(1), 3-24.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical analysis, 27*(2), 93-115.

Arentze, T. A., Dellaert, B. G., & Chorus, C. G. (2014). Incorporating mental representations in discrete choice models of travel behavior: modeling approach and empirical application. *Transportation Science, 49*(3), 577-590.

Axhausen, K. W., Zimmermann, A., Schönfelder, S., Rindsfüser, G., & Haupt, T. (2002). Observing the rhythms of daily life: A six-week travel diary. *Transportation, 29*(2), 95-124.

Banister, D. (2008). The sustainable mobility paradigm. *Transport Policy, 15*(2), 73-80.

Banister, D. (2011). Cities, mobility and climate change. *Journal of Transport Geography, 19*(6), 1538-1546.

Bao, J., Xu, C., Liu, P., & Wang, W. (2017). Exploring bikesharing travel patterns and trip purposes using smart card data and online point of interests. *Networks and Spatial Economics, 17*(4), 1231-1253.

Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., & Tomasini, M. (2018). Human mobility: Models and applications. *Physics Reports, 734*, 1-74.

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automated will: nonconscious activation and pursuit of behavioral goals. *Journal of personality and social psychology, 81*(6), 1014.

Barr, S., & Prillwitz, J. (2012). Green travellers? Exploring the spatial context of sustainable mobility styles. *Applied Geography, 32*(2), 798-809.

Bart, Y., Stephen, A. T., & Sarvary, M. (2014). Which products are best suited to mobile advertising? A field study of mobile display advertising effects on consumer attitudes and intentions. *Journal of Marketing Research, 51*(3), 270-285.

Becker, G. S. (2013). *The economic approach to human behavior*. University of Chicago press.

Bell, D. R., & Song, S. (2004). Social contagion and trial on the Internet: Evidence from online grocery retailing. *Wharton School Marketing Department Paper*.

Bell, W. J. (2012). *Searching behaviour: the behavioural ecology of finding resources*. Springer Science & Business Media.

Ben-Akiva, M., & Bowman, J. L. (1998). Integration of an activity-based model system and a residential location model. *Urban Studies, 35*(7), 1131-1153.

Bernritter, S. F., Ketelaar, P. E., & Sotgiu, F. (2021). Behaviorally targeted location-based mobile marketing. *Journal of the Academy of Marketing Science, 49*(4), 677-702.

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of consumer research, 25*(3), 187-217.

Bhat, C. R., & Steed, J. L. (2002). A continuous-time model of departure time choice for urban shopping trips. *Transportation Research Part B: Methodological, 36*(3), 207-224.

Bitner, M. J. (1992). Servicescapes: The impact of physical surroundings on customers and employees. *Journal of Marketing, 56*(2), 57-71.

Borgo, R., Kehrer, J., Chung, D. H., Maguire, E., Laramee, R. S., Hauser, H., Ward, M., & Chen, M. (2013). Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. Eurographics (State of the Art Reports),

Bowman, J. L., & Ben-Akiva, M. E. (2001). Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice, 35*(1), 1-28.

Bradlow, E. T., Bronnenberg, B., Russell, G. J., Arora, N., Bell, D. R., Duvvuri, S. D., Hofstede, F. T., Sismeiro, C., Thomadsen, R., & Yang, S. (2005). Spatial Models in Marketing. *Marketing Letters, 16*(3), 267-278.

Bron, J., & Soeters, R. (2014). *Project Fietsactie Regio Twente*. https://docs.google.com/viewerng/viewer?url=https://beterbenutten.nl/document/download/15514/

Bronnenberg, B. J. (2005). Spatial models in marketing research and practice. *Applied Stochastic Models in Business and Industry, 21*(4-5), 335-343.

Bronnenberg, B. J., Dhar, S. K., & Dubé, J.-P. (2005). Market structure and the geographic distribution of brand shares in consumer package goods industries. *Available at SSRN 651321*.

Bronnenberg, B. J., & Mahajan, V. (2001). Unobserved retailer behavior in multimarket data: Joint spatial dependence in market shares and promotion variables. *Marketing Science, 20*(3), 284-299.

Brown, C., Lathia, N., Mascolo, C., Noulas, A., & Blondel, V. (2014). Group colocation behavior in technological social networks. *PloS one, 9*(8), e105816.

Brown, S. W. (1995). Time, change, and motion: The effects of stimulus movement on temporal perception. *Perception & psychophysics, 57*, 105-116.

Bruner, G. C., & Kumar, A. (2007). Attitude toward location-based advertising. *Journal of Interactive Advertising, 7*(2), 3-15.

Bues, M., Steiner, M., Stafflage, M., & Krafft, M. (2017). How Mobile In-Store Advertising Influences Purchase Intention: Value Drivers and Mediating Effects from a Consumer Perspective. *Psychology & Marketing, 34*(2), 157-174.

Cai, L., Jiang, F., Zhou, W., & Li, K. (2018). Design and application of an attractiveness index for urban hotspots based on GPS trajectory data. *IEEE Access, 6*, 55976-55985.

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira Jr, J., & Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies, 26*, 301-313.

Canel, C., Khumawala, B. M., Law, J., & Loh, A. (2001). An algorithm for the capacitated, multi-commodity multi-period facility location problem. *Computers & Operations Research, 28*(5), 411-427.

Carrasco, J. A., Hogan, B., Wellman, B., & Miller, E. J. (2008). Collecting social network data to study social activity-travel behavior: an egocentric approach. *Environment and Planning B: Planning and Design, 35*(6), 961-980.

Cecchini, M., Zambon, I., & Salvati, L. (2019). Housing and the city: a spatial analysis of residential building activity and the socio-demographic background in a Mediterranean city, 1990–2017. *Sustainability, 11*(2), 375.

Celik, M., & Dokuz, A. S. (2017). Discovering socio-spatio-temporal important locations of social media users. *Journal of computational science, 22*, 85-98.

Centraal Bureau voor de Statistiek (CBS). (2021). Regionale kerncijfers Nederland. *Statline*. https://opendata.cbs.nl/#/CBS/nl/dataset/70072ned/table?dl=55749

Chen, C., Liao, C., Xie, X., Wang, Y., & Zhao, J. (2019). Trip2Vec: a deep embedding approach for clustering and profiling taxi trip purposes. *Personal and Ubiquitous Computing, 23*(1), 53-66.

Chen, S., Yuan, X., Wang, Z., Guo, C., Liang, J., Wang, Z., Zhang, X., & Zhang, J. (2015). Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. *IEEE transactions on visualization and computer graphics, 22*(1), 270-279.

Chen, Y., Chen, X., Liu, Z., & Li, X. (2020). Understanding the spatial organization of urban functions based on co-location patterns mining: A comparative analysis for 25 Chinese cities. *Cities, 97*, 102563.

Cho, J., & Yoon, Y. (2021). Extracting the topology of urban airspace through graph abstraction. *Transportation Research Part C: Emerging Technologies, 127*, 103116.

Ciclogreen. (2020). *https://www.ciclogreen.com/*. Introduction

Clark, T. N. (2003). Urban Amenities: Lakes, opera, and juice bars: Do they drive development? In T. Nichols Clark (Ed.), *The City as an Entertainment Machine* (Vol. 9, pp. 103-140). Emerald Group Publishing Limited.

Clark, T. N., Lloyd, R., Wong, K. K., & Jain, P. (2002). Amenities Drive Urban Growth. *Journal of Urban Affairs, 24*(5), 493-515.

Coscia, M., & Hausmann, R. (2015). Evidence that calls-based and mobility networks are isomorphic. *PloS one, 10*(12), e0145091.

Czarny, B. (2018). How Brands Are Getting More Sophisticated at Using Location Data. https://www.adweek.com/performance-marketing/how-brands-are-getting-more-sophisticated-at-using-location-data/

D'Silva, K., Jayarajah, K., Noulas, A., Mascolo, C., & Misra, A. (2018). The Role of Urban Mobility in Retail Business Survival. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2*(3), 100.

D'Silva, K., Noulas, A., Musolesi, M., Mascolo, C., & Sklar, M. (2018). Predicting the temporal activity patterns of new venues. *EPJ Data Science, 7*(1), 13.

Danaher, P. J., Smith, M. S., Ranasinghe, K., & Danaher, T. S. (2015). Where, when, and how long: Factors that influence the redemption of mobile phone coupons. *Journal of Marketing Research, 52*(5), 710-725.

De Ceunynck, T., Kusumastuti, D., Hannes, E., Janssens, D., & Wets, G. (2011). What drives people? Analyzing leisure-shopping trip decision making. Proceedings of the 90th Annual Meeting of the Transportation Research Board. Washington, DC, USA,

De Jong, G., Gunn, H., & Ben-Akiva, M. (2004). A meta-model for passenger and freight transport in Europe. *Transport Policy, 11*(4), 329-344.

De Witte, A., Macharis, C., & Mairesse, O. (2008). How persuasive is 'free'public transport?: a survey among commuters in the Brussels Capital Region. *Transport Policy, 15*(4), 216-224.

Del Gatto, M., & Mastinu, C. S. (2018). A Huff model with firm heterogeneity and selection. Application to the Italian retail sector. *Spatial Economic Analysis, 13*(4), 442-456.

Dellaert, B. G., Arentze, T. A., & Horeni, O. (2014). Tourists' mental representations of complex travel decision problems. *Journal of Travel Research, 53*(1), 3-11.

Dellaert, B. G., Arentze, T. A., & Timmermans, H. J. (2008). Shopping context and consumers' mental representation of complex shopping trip decision problems. *Journal of Retailing, 84*(2), 219-232.

Demissie, M. G., de Almeida Correia, G. H., & Bento, C. (2013). Exploring cellular network handover information for urban mobility analysis. *Journal of Transport Geography, 31*, 164-170.

Ding, C., & He, X. (2004). K-means clustering via principal component analysis. Proceedings of the twenty-first international conference on Machine learning,

Dokuz, A. S., & Celik, M. (2017). Discovering socially important locations of social media users. *Expert Systems with Applications, 86*, 113-124.

Dong, X., Suhara, Y., Bozkaya, B., Singh, V. K., Lepri, B., & Pentland, A. S. (2018). Social bridges in urban purchase behavior. *ACM Transactions on Intelligent Systems and Technology (TIST), 9*(3), 33.

Duan, J. A., & Mela, C. F. (2009). The Role of Spatial Demand on Outlet Location and Pricing. *Journal of Marketing Research, 46*(2), 260-278.

Dubé, J.-P., Fang, Z., Fong, N., & Luo, X. (2017). Competitive Price Targeting with Smartphone Coupons. *Marketing Science, 36*(6), 944-975.

Eagle, N., & Pentland, A. S. (2009). Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology, 63*(7), 1057-1066.

Echeverri, P., & Salomonson, N. (2019). Consumer vulnerability during mobility service interactions: causes, forms and coping. *Journal of Marketing Management, 35*(3-4), 364-389.

Economische Samenwerking Zuid-Limburg (ESZL). (2018). Factsheet Zuid-Limburg. https://www.eszl.nl/wp-content/uploads/2018/10/Factsheet-Zuid-Limburg.pdf

Edvardsson, B. (1998). Causes of customer dissatisfaction - studies of public transport by the critical-incident method. *Managing Service Quality: An International Journal, 8*(3), 189-197.

Edvardsson, B., Enquist, B., & Johnston, R. (2005). Cocreating customer value through hyperreality in the prepurchase service experience. *Journal of service research, 8*(2), 149-161.

Elhorst, J. P. (2017). Spatial Models. In P. S. H. Leeflang, J. E. Wieringa, T. H. A. Bijmolt, & K. H. Pauwels (Eds.), *Advanced Methods for Modeling Markets* (pp. 173-202). Springer International Publishing.

ERTICO. (2019). *Mobility as a Service (MaaS) and Sustainable Urban Mobility Planning*. E. P. o. S. U. M. Plans. https://www.eltis.org/sites/default/files/mobility_as_a_service_maas_and_sustainable_urban_mobility_planning.pdf

EU. (2018). *The GDPR: new opportunities, new obligations: What Every Business Needs to Know about the EU's General Data Protection Regulation*. https://ec.europa.eu/info/sites/default/files/data-protection-factsheet-sme-obligations_en.pdf

Fadhil, D. N. (2018). A GIS-based analysis for selecting ground infrastructure locations for urban air mobility. *inlangen]. Master's Thesis, Technical University of Munich*.

Fang, Z., Gu, B., Luo, X., & Xu, Y. (2015). Contemporaneous and delayed sales impact of location-based mobile promotions. *Information systems research, 26*(3), 552-564.

Fernandez-Heredia, A., & Fernandez-Sanchez, G. (2020). Processes of civic participation in the implementation of sustainable urban mobility systems. *Case Studies on Transport Policy, 8*(2), 471-483.

Ferrari, L., Berlingerio, M., Calabrese, F., & Curtis-Davidson, B. (2012). Measuring public-transport accessibility using pervasive mobility data. *IEEE Pervasive Computing, 12*(1), 26-33.

Frachtenberg, E. (2019). Practical Drone Delivery. *Computer, 52*(12), 53-57.

Gabrielli, S., Forbes, P., Jylhä, A., Wells, S., Sirén, M., Hemminki, S., Nurmi, P., Maimone, R., Masthoff, J., & Jacucci, G. (2014). Design challenges in motivating change for sustainable urban mobility. *Computers in Human Behavior, 41*, 416-423.

Gal, M. S., & Aviv, O. (2020). The Competitive Effects of the GDPR. *Journal of Competition Law & Economics, 16*(3), 349-391.

Garikapati, V. M., Pendyala, R. M., Morris, E. A., Mokhtarian, P. L., & McDonald, N. (2016). Activity patterns, time use, and travel of millennials: a generation in transition? *Transport Reviews, 36*(5), 558-584.

Garrett-Glaser, B. (2020). South Korea Plans to Launch Air Taxi Service by 2025. Will Hyundai Be Ready?

Ghose, A., Kwon, H. E., Lee, D., & Oh, W. (2019). Seizing the commuting moment: Contextual targeting based on mobile transportation apps. *Information systems research, 30*(1), 154-174.

Ghose, A., Li, B., & Liu, S. (2019). Mobile targeting using customer trajectory patterns. *Management Science, 65*(11), 5027-5049.

Gibbons, S., Mourato, S., & Resende, G. M. (2014). The Amenity Value of English Nature: A Hedonic Price Approach. *Environmental and Resource Economics, 57*(2), 175-196.

Giuffrida, N., Le Pira, M., Inturri, G., & Ignaccolo, M. (2021). Addressing the public transport ridership/coverage dilemma in small cities: A spatial approach. *Case Studies on Transport Policy, 9*(1), 12-21.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological assessment, 4*(1), 26.

Gong, L., Liu, X., Wu, L., & Liu, Y. (2016). Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science, 43*(2), 103-114.

Gong, L., Morikawa, T., Yamamoto, T., & Sato, H. (2014). Deriving personal trip data from GPS data: A literature review on the existing methodologies. *Procedia-Social and Behavioral Sciences, 138*, 557-565.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *nature, 453*(7196), 779.

Goodchild, M. F., & Janelle, D. G. (1984). The city around the clock: Space—time patterns of urban ecological structure. *Environment and Planning A, 16*(6), 807-820.

Graser, A., Schmidt, J., Roth, F., & Brändle, N. (2019). Untangling origin-destination flows in geographic information systems. *Information visualization, 18*(1), 153-172.

Greene, T., Shmueli, G., Ray, S., & Fell, J. (2019). Adjusting to the GDPR: The impact on data scientists and behavioral researchers. *Big data, 7*(3), 140-162.

Grindrod, P., & Lee, T. (2016). Comparison of social structures within cities of very different sizes. *Royal Society open science, 3*(2), 150526.

Guo, H., Wang, Z., Yu, B., Zhao, H., & Yuan, X. (2011). Tripvista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. 2011 IEEE Pacific Visualization Symposium,

Hägerstrand, T. (1989). Reflections on "what about people in regional science?". Papers of the Regional Science Association, Hartmann, W. R., Manchanda, P., Nair, H., Bothner, M., Dodds, P., Godes, D., Hosanagar, K., & Tucker, C. (2008). Modeling social interactions: Identification, empirical methods and policy implications. *Marketing Letters, 19*(3), 287-304.

Harvey, A. S., & Taylor, M. E. (2000). Activity settings and travel behaviour: A social contact perspective. *Transportation, 27*(1), 53-73.

Haustein, S., & Hunecke, M. (2007). Reduced use of environmentally friendly modes of transportation caused by perceived mobility necessities: An extension of the theory of planned behavior 1. *Journal of Applied Social Psychology, 37*(8), 1856-1883.

Hennig, C. (2020). *fpc: Flexible Procedures for Clustering*. https://CRAN.R-project.org/package=fpc

Hu, Y., Miller, H. J., & Li, X. (2014). Detecting and analyzing mobility hotspots using surface networks. *Transactions in GIS, 18*(6), 911-935.

Hunecke, M., Haustein, S., Böhler, S., & Grischkat, S. (2010). Attitude-based target groups to reduce the ecological impact of daily mobility behavior. *Environment and behavior, 42*(1), 3-43.

Hunneman, A., Elhorst, J. P., & Bijmolt, T. H. A. (2022). Store sales evaluation and prediction using spatial panel data models of sales components. *Spatial Economic Analysis, 17*(1), 127-150.

Hunneman, A., Verhoef, P. C., & Sloot, L. M. (2021). The impact of hard discounter presence on store satisfaction and store loyalty. *Journal of Retailing and Consumer Services, 59*, 102405.

Hutchinson, T. P. (2009). The customer experience when using public transport: a review. *Proceedings of the Institution of Civil Engineers - Municipal Engineer, 162*(3), 149-157.

Innes, J. D., Ircha, M. C., & Badoe, D. A. (1990). Factors Affecting Automobile Shopping Trip Destinations. *Journal of Urban Planning and Development, 116*(3), 126-136.

Jang, S., Caiati, V., Rasouli, S., Timmermans, H., & Choi, K. (2021). Does MaaS contribute to sustainable transportation? A mode choice perspective. *International Journal of Sustainable Transportation, 15*(5), 351-363.

Jank, W., & Kannan, P. (2005). Understanding geographical markets of online firms using spatial models of customer choice. *Marketing Science, 24*(4), 623-634.

Jevinger, Å., & Persson, J. A. (2019). Exploring the potential of using real-time traveler data in public transport disturbance management. *Public Transport, 11*(2), 413-441.

Jiang, S., Ferreira, J., & Gonzalez, M. C. (2012, 11/01). Clustering daily patterns of human activities in the city. *Data Mining and Knowledge Discovery, 25*. https://doi.org/10.1007/s10618-012-0264-z

Jiang, S., Ferreira, J., & González, M. C. (2012). Clustering daily patterns of human activities in the city. *Data Mining and Knowledge Discovery, 25*(3), 478-510.

Kahraman, C., Ruan, D., & Doğan, I. (2003). Fuzzy group decision-making for facility location selection. *Information sciences, 157*, 135-153.

Kalinic, M., & Krisp, J. M. (2018). Kernel density estimation (KDE) vs. hot-spot analysis–detecting criminal hot spots in the City of San Francisco. Proceeding of the 21?? Conference on Geo-Information Science, Kang, C., Gao, S., Lin, X., Xiao, Y., Yuan, Y., Liu, Y., & Ma, X. (2010). Analyzing and geo-visualizing individual human mobility patterns using mobile call records. 2010 18th international conference on geoinformatics,

Kang, C., Ma, X., Tong, D., & Liu, Y. (2012). Intra-urban human mobility patterns: An urban morphology perspective. *Physica A: Statistical Mechanics and its Applications, 391*(4), 1702-1717.

Karamshuk, D., Noulas, A., Scellato, S., Nicosia, V., & Mascolo, C. (2013). Geo-spotting: mining online location-based services for optimal retail store placement. Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining,

Ketelaar, P. E., Bernritter, S. F., van't Riet, J., Hühn, A. E., van Woudenberg, T. J., Müller, B. C. N., & Janssen, L. (2017). Disentangling location-based advertising: the effects of location congruency and medium type on consumers' ad attention and brand choice. *International Journal of Advertising, 36*(2), 356-367.

Ketelaar, P. E., Bernritter, S. F., van Woudenberg, T. J., Rozendaal, E., Konig, R. P., Hühn, A. E., Van Gisbergen, M. S., & Janssen, L. (2018). "Opening" location-based mobile ads: How openness and location congruency of location-based ads weaken negative effects of intrusiveness on brand choice. *Journal of Business Research, 91*, 277-285.

Klose, A. (1999). An LP-based heuristic for two-stage capacitated facility location problems. *Journal of the Operational Research Society, 50*(2), 157-166.

Koppelman, F. S., & Pas, E. I. (1980). Travel-choice behavior: Models of perceptions, feelings, preference, and choice. *Transportation Research Record*(765).

Kossinets, G., & Watts, D. J. (2009). Origins of homophily in an evolving social network. *American journal of sociology, 115*(2), 405-450.

Kotsios, A., Magnani, M., Rossi, L., Shklovski, I., & Vega D'Aurelio, D. (2019). *An Analysis of the Consequences of the General Data Protection Regulation (GDPR) on Social Network Research*.

Kroesen, M. (2014). Modeling the behavioral determinants of travel behavior: An application of latent transition analysis. *Transportation Research Part A: Policy and Practice, 65*, 56-67.

Kurniawan, J. H., Ong, C., & Cheah, L. (2018). Examining values and influences affecting public expectations of future urban mobility: A Singapore case study. *Transport Policy, 66*, 66-75.

Larsson, J. P. (2014). The neighborhood or the region? Reassessing the density–wage relationship using geocoded data. *The Annals of Regional Science, 52*(2), 367-384.

Larsson, J. P., & Öner, Ö. (2014). Location and co-location in retail: a probabilistic approach using geo-coded data for metropolitan retail markets. *The Annals of Regional Science, 52*(2), 385-408.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing, 80*(6), 69-96.

Leng, Y., Santistevan, D., & Pentland, A. (2018). Familiar Strangers: the Collective Regularity in Human Behaviors. *arXiv e-prints*.

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Li, C., Luo, X., Zhang, C., & Wang, X. (2017). Sunny, rainy, and cloudy with a chance of mobile promotion effectiveness. *Marketing Science, 36*(5), 762-779.

Li, F., Shi, W., & Zhang, H. (2021). A two-phase clustering approach for urban hotspot detection with spatiotemporal and network constraints. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14*, 3695-3705.

Li, M., Ye, X., Zhang, S., Tang, X., & Shen, Z. (2018). A framework of comparative urban trajectory analysis. *Environment and Planning B: Urban Analytics and City Science, 45*(3), 489-507.

Li, X., Pan, G., Wu, Z., Qi, G., Li, S., Zhang, D., Zhang, W., & Wang, Z. (2012). Prediction of urban human mobility using large-scale taxi traces and its applications. *Frontiers of Computer Science, 6*(1), 111-121.

Libaque-Sáenz, C. F., Wong, S. F., Chang, Y., & Bravo, E. R. (2021). The effect of Fair information practices and data collection methods on privacy-related behaviors: A study of Mobile apps. *Information & Management, 58*(1), 103284.

Liight. (2021). *https://liight.es/*.

Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern recognition, 36*(2), 451-461.

Lipkin, M., & Heinonen, K. (2022). Customer ecosystems: exploring how ecosystem actors shape customer experience. *Journal of Services Marketing, 36*(9), 1-17.

Liu, L. B., Assaf, & Ratti, C. (2009). Urban mobility landscape: Real time monitoring of urban mobility patterns. *Proceedings of the 11th international conference on computers in urban planning and urban management*, 1-16.

Luce, R. D. (2012). *Individual choice behavior: A theoretical analysis*. Courier Corporation.

Luo, X., Andrews, M., Fang, Z., & Phang, C. W. (2014). Mobile targeting. *Management Science, 60*(7), 1738-1756.

Luo, X., Gu, B., Fang, Z., & Xu, Y. (2013). Quantifying the Dynamic Sales Impact of Location-Based Mobile Promotion Technologies. ICIS,

Ma, Y., Lin, T., Cao, Z., Li, C., Wang, F., & Chen, W. (2015). Mobility viewer: An Eulerian approach for studying urban crowd flow. *IEEE Transactions on Intelligent Transportation Systems, 17*(9), 2627-2636.

Maldoff, G. (2016). How GDPR changes the rules for research. https://iapp.org/news/a/how-gdpr-changes-the-rules-for-research/

Malleson, N., & Andresen, M. A. (2015). The impact of using social media data in crime rate calculations: shifting hot spots and changing spatial patterns. *Cartography and Geographic Information Science, 42*(2), 112-121.

Malokin, A., Circella, G., & Mokhtarian, P. L. (2019). How do activities conducted while commuting influence mode choice? Using revealed preference models to inform public transportation advantage and autonomous vehicle scenarios. *Transportation Research Part A: Policy and Practice, 124*, 82-114.

Manaugh, K., Miranda-Moreno, L. F., & El-Geneidy, A. M. (2010). The effect of neighbourhood characteristics, accessibility, home–work location, and demographics on commuting distances [journal article]. *Transportation, 37*(4), 627-646.

Mao, X., Zhao, X., Lin, J., & Herrera-Viedma, E. (2019). Utilizing multi-source data in popularity prediction for shop-type recommendation. *Knowledge-Based Systems, 165*, 253-267.

Marshall, A. (1926). Principles of economics 8th ed. *London: McMillan*.

McDonald, N. C. (2015). Are millennials really the “go-nowhere” generation? *Journal of the American Planning Association, 81*(2), 90-103.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology, 27*(1), 415-444.

Mejia-Dorantes, L., Paez, A., & Vassallo, J. M. (2012). Transportation infrastructure impacts on firm location: the effect of a new metro line in the suburbs of Madrid. *Journal of Transport Geography, 22*, 236-250.

Melkote, S., & Daskin, M. S. (2001). Capacitated facility location/network design problems. *European journal of operational research, 129*(3), 481-495.

Millward, H., & Spinney, J. (2011). Time use, travel behavior, and the rural–urban continuum: results from the Halifax STAR project. *Journal of Transport Geography, 19*(1), 51-58.

MIT. (2022). *Urban Network Analysis toolbox for Rhino3D*. City Form Lab / MIT. Retrieved April 21 from https://cityform.mit.edu/projects/una-rhino-toolbox

Mittal, V., Kamakura, W. A., & Govind, R. (2004). Geographic patterns in customer service and satisfaction: An empirical investigation. *Journal of Marketing, 68*(3), 48-62.

Molitor, D., Reichhart, P., & Spann, M. (2012). Location-based advertising: Measuring the impact of context-specific factors on consumers’ choice behavior. *Available at SSRN 2116359*.

Molitor, D., Spann, M., Ghose, A., & Reichhart, P. (2020). Effectiveness of location-based advertising and the impact of interface design. *Journal of Management Information Systems, 37*(2), 431-456.

Moreau, C., Devogele, T., Peralta, V., Etienne, L., & de Runz, C. (2020). Methodology for Mining, Discovering and Analyzing Semantic Human Mobility Behaviors. *arXiv preprint arXiv:2012.04767*.

Mouwen, A. (2015). Drivers of customer satisfaction with public transport services. *Transportation Research Part A: Policy and Practice, 78*, 1-20.

Mueller, J., Schmoeller, S., & Giesel, F. (2015). Identifying users and use of (electric-) free-floating carsharing in Berlin and Munich. 2015 IEEE 18th International Conference on Intelligent Transportation Systems,

NASA. (2018). *Urban Air Mobility (UAM) Market Study*. NASA. https://www.nasa.gov/sites/default/files/atoms/files/uam-market-study-executive-summary-v2.pdf

Natapov, A., Czamanski, D., & Fisher-Gewirtzman, D. (2018). A network approach to link visibility and urban activity location. *Networks and Spatial Economics, 18*(3), 555-575.

Nguyen, M. H., Armoogum, J., Madre, J.-L., & Garcia, C. (2020). Reviewing trip purpose imputation in GPS-based travel surveys. *Journal of Traffic and Transportation Engineering (English Edition)*.

Noë, N., Whitaker, R. M., Chorley, M. J., & Pollet, T. V. (2016). Birds of a feather locate together? Foursquare checkins and personality homophily. *Computers in Human Behavior, 58*, 343-353.

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012). A Tale of Many Cities: Universal Patterns in Human Urban Mobility. *PloS one, 7*(5), e37027.
O'Sullivan, D., & Unwin, D. (2003). *Geographic information analysis*. John Wiley & Sons.
Öner, Ö. (2017). Retail city: the relationship between place attractiveness and accessibility to shops. *Spatial Economic Analysis, 12*(1), 72-91.
Pan, H., Shen, Q., & Zhang, M. (2009). Influence of urban form on travel behaviour in four neighbourhoods of Shanghai. *Urban Studies, 46*(2), 275-294.
Paul, C., Scheibe, K., & Nilakanta, S. (2020). Privacy concerns regarding wearable IoT devices: how it is influenced by GDPR? Proceedings of the 53rd Hawaii International Conference on System Sciences,
Pérez-Messina, I., & Graells-Garrido, E. (2019). Visualizing Transportation Flows with Mode Split using Glyphs.
Polzin, S. E., Chu, X., & Godfrey, J. (2014). The impact of millennials' travel behavior on future personal vehicle travel. *Energy Strategy Reviews, 5*, 59-65.
Porta, S., Strano, E., Iacoviello, V., Messora, R., Latora, V., Cardillo, A., Wang, F., & Scellato, S. (2009). Street centrality and densities of retail and services in Bologna, Italy. *Environment and Planning B: Planning and Design, 36*(3), 450-465.
Prillwitz, J., & Barr, S. (2011). Moving towards sustainability? Mobility styles, attitudes and individual travel behaviour. *Journal of Transport Geography, 19*(6), 1590-1600.
Provost, F., Martens, D., & Murray, A. (2015). Finding similar mobile consumers with a privacy-friendly geosocial design. *Information systems research, 26*(2), 243-265.
Rajendran, S., & Srinivas, S. (2020). Air taxi service for urban mobility: a critical review of recent developments, future challenges, and opportunities. *Transportation research part E: logistics and transportation review, 143*, 102090.
Ralambondrainy, H. (1995). A conceptual version of the K-means algorithm. *Pattern Recognition Letters, 16*(11), 1147-1157.
Ribeiro, P., Fonseca, F., & Meireles, T. (2020). Sustainable mobility patterns to university campuses: Evaluation and constraints. *Case Studies on Transport Policy, 8*(2), 639-647.
Rimjha, M., Hotle, S., Trani, A., & Hinze, N. (2021). Commuter demand estimation and feasibility assessment for Urban Air Mobility in Northern California. *Transportation Research Part A: Policy and Practice, 148*, 506-524.
Sagl, G., Loidl, M., & Beinat, E. (2012). A visual analytics approach for extracting spatio-temporal urban mobility information from mobile network traffic. *ISPRS International Journal of Geo-Information, 1*(3), 256-271.
Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of The Royal Society Interface, 10*(84), 20130246.
Schoenau, M., & Müller, M. (2017). What affects our urban travel behavior? A GPS-based evaluation of internal and external determinants of sustainable mobility in Stuttgart (Germany). *Transportation research part F: Traffic psychology and behaviour, 48*, 61-73.
Schrage, R., Meißner, L., Schütte, R., & Kenning, P. (2022). Acceptance of in-store apps: factors that influence the intention to adopt location-based retail apps–

insights from Germany. *International Journal of Retail & Distribution Management*.

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS), 42*(3), 1-21.

Sevtsuk, A. (2017). Analysis and Planning of Urban Networks. In.

Sevtsuk, A., & Mekonnen, M. (2012). Urban network analysis. A new toolbox for ArcGIS. *Revue internationale de géomatique, 22*, 287-305.

Sevtsuk, A., & Ratti, C. (2010). Does Urban Mobility Have a Daily Routine? Learning from the Aggregate Data of Mobile Networks. *Journal of urban technology, 17*(1), 41-60.

Shariati, M., Mesgari, T., Kasraee, M., & Jahangiri-Rad, M. (2020). Spatiotemporal analysis and hotspots detection of COVID-19 using geographic information system (March and April, 2020). *Journal of Environmental Health Science and Engineering, 18*(2), 1499-1507.

Shen, L., & Stopher, P. R. (2014). Review of GPS travel survey and GPS data-processing methods. *Transport Reviews, 34*(3), 316-334.

Sheng, X., Siguaw, J. A., & Simpson, P. M. (2016). Servicescape attributes and consumer well-being. *Journal of Services Marketing, 30*(7), 676-685.

Shi, Z., & Pun-Cheng, L. S. C. (2019). Spatiotemporal Data Clustering: A Survey of Methods. *ISPRS International Journal of Geo-Information, 8*(3), 112-128. https://www.mdpi.com/2220-9964/8/3/112

Smith, G., & Hensher, D. A. (2020). Towards a framework for Mobility-as-a-Service policies. *Transport Policy, 89*, 54-65.

Song, C., Qu, Z., Blumm, N., & Barabási, A.-L. (2010). Limits of Predictability in Human Mobility. *Science, 327*(5968), 1018.

Steg, L., Vlek, C., & Slotegraaf, G. (2001). Instrumental-reasoned and symbolic-affective motives for using a motor car. *Transportation research part F: Traffic psychology and behaviour, 4*(3), 151-169.

Stephen, D. M., & Jenny, B. (2019). Methods and tools for automated production of origin-destination flow maps. *Abstracts of the ICA, 1*, NA-NA.

Still, J., Komulainen, H., & Nätti, S. (2018). Contextual layers of service experience in professional business services. *Journal of Business & Industrial Marketing*.

Stock, K., & Guesgen, H. (2016). Geospatial reasoning with open data. In *Automating open source intelligence* (pp. 171-204). Elsevier.

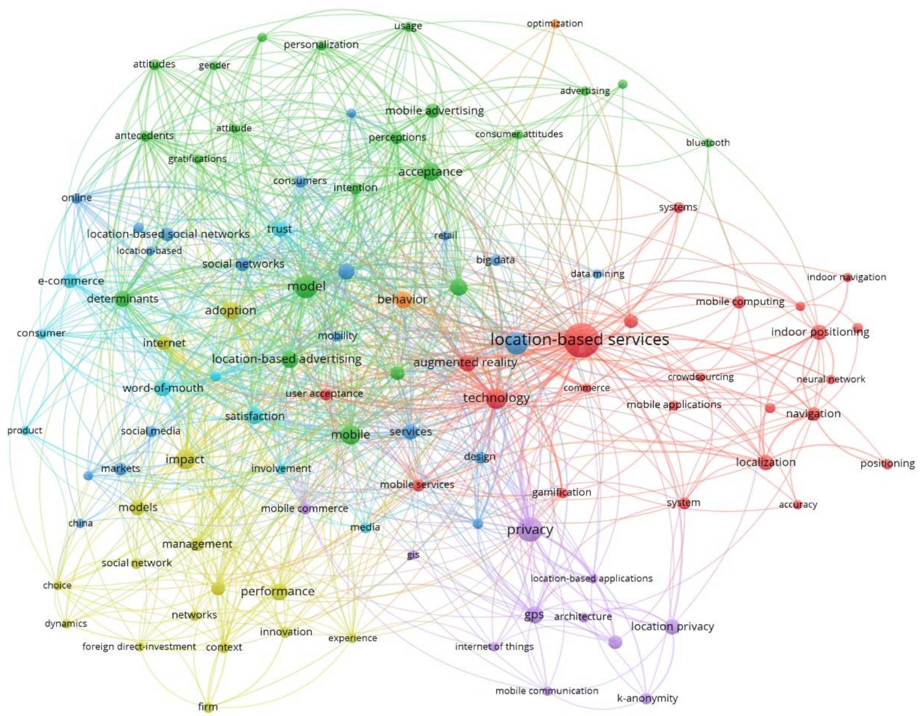
Stone, G., Barnes, J. H., & Montgomery, C. (1995). Ecoscale: a scale for the measurement of environmentally responsible consumers. *Psychology & Marketing, 12*(7), 595-612.

Straubinger, A., Rothfeld, R., Shamiyeh, M., Büchter, K.-D., Kaiser, J., & Plötner, K. O. (2020). An overview of current research and developments in urban air mobility–Setting the scene for UAM introduction. *Journal of Air Transport Management, 87*, 101852.

Straubinger, A., Verhoef, E. T., & de Groot, H. L. (2021). Will urban air mobility fly? The efficiency and distributional impacts of UAM in different urban spatial structures. *Transportation Research Part C: Emerging Technologies, 127*, 103124.

Tachet, R., Sagarra, O., Santi, P., Resta, G., Szell, M., Strogatz, S. H., & Ratti, C. (2017). Scaling law of urban ride sharing. *Scientific Reports, 7*(1), 1-6.
Thomas, J. J., & Cook, K. A. (2006). A visual analytics agenda. *IEEE computer graphics and applications, 26*(1), 10-13.
Ton, D., Zomer, L.-B., Schneider, F., Hoogendoorn-Lanser, S., Duives, D., Cats, O., & Hoogendoorn, S. (2020). Latent classes of daily mobility patterns: the relationship with attitudes towards modes. *Transportation, 47*(4), 1843-1866.
Tyrväinen, L., & Miettinen, A. (2000). Property Prices and Urban Forest Amenities. *Journal of Environmental Economics and Management, 39*(2), 205-223.
UBER. (2016). *Fast-Forwarding to a Future of On-Demand Urban Air Transportation*. U. Elevate. https://evtol.news/__media/PDFs/UberElevateWhitePaperOct2016.pdf
van't Riet, J., Hühn, A., Ketelaar, P., Khan, V.-J., Konig, R., Rozendaal, E., & Markopoulos, P. (2016). Investigating the effects of location-based advertising in the supermarket: Does goal congruence trump location congruence? *Journal of Interactive Advertising, 16*(1), 31-43.
van de Sanden, S., Willems, K., & Brengman, M. (2019). In-store location-based marketing with beacons: from inflated expectations to smart use in retailing. *Journal of Marketing Management, 35*(15-16), 1514-1541.
van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping [journal article]. *Scientometrics, 84*(2), 523-538. https://doi.org/10.1007/s11192-009-0146-3
van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In *Measuring scholarly impact* (pp. 285-320). Springer.
van Rijn, H. (2014). It's time to take the psychology of biological time into account: speed of driving affects a trip's subjective duration. *Frontiers in Psychology, 5*, 1028.
Visner, M., Shirowzhan, S., & Pettit, C. (2021). Spatial Analysis, Interactive Visualisation and GIS-Based Dashboard for Monitoring Spatio-temporal Changes of Hotspots of Bushfires over 100 Years in New South Wales, Australia. *Buildings, 11*(2), 37.
Von Landesberger, T., Brodkorb, F., Roskosch, P., Andrienko, N., Andrienko, G., & Kerren, A. (2015). MobilityGraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. *IEEE transactions on visualization and computer graphics, 22*(1), 11-20.
Wachter, S. (2018). The GDPR and the Internet of Things: a three-step transparency model. *Law, Innovation and Technology, 10*(2), 266-294.
Wachter, S. (2019). Data protection in the age of big data. *Nature Electronics, 2*(1), 6-7.
Wakefield, K. L., & Blodgett, J. (2016). Retrospective: the importance of servicescapes in leisure service settings. *Journal of Services Marketing, 30*(7), 686-691.
Wakefield, K. L., & Blodgett, J. G. (1994). The Importance of Servicescapes in Leisure Service Settings. *Journal of Services Marketing, 8*(3), 66-76.
Waltman, L., van Eck, N. J., & Noyons, E. C. M. (2010, 10//). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics, 4*(4), 629-635. https://doi.org/https://doi.org/10.1016/j.joi.2010.07.002
Wang, D., Pedreschi, D., Song, C., Giannotti, F., & Barabasi, A.-L. (2011). *Human mobility, social ties, and link prediction* Proceedings of the 17th ACM SIGKDD

international conference on Knowledge discovery and data mining, San Diego, California, USA.
Wang, J., Kong, X., Xia, F., & Sun, L. (2019). Urban human mobility: Data-driven modeling and prediction. *Acm Sigkdd Explorations Newsletter, 21*(1), 1-19.
Wang, P., Fu, Y., Liu, G., Hu, W., & Aggarwal, C. (2017). *Human Mobility Synchronization and Trip Purpose Detection with Mixture of Hawkes Processes*. https://doi.org/10.1145/3097983.3098067
Wang, P., Liu, G., Fu, Y., Zhou, Y., & Li, J. (2017). Spotting Trip Purposes from Taxi Trajectories: A General Probabilistic Model. *ACM Trans. Intell. Syst. Technol., 9*(3), 1-26.
Warshaw, P. R., & Davis, F. D. (1985). Disentangling behavioral intention and behavioral expectation. *Journal of Experimental Social Psychology, 21*(3), 213-228.
Willey, L. C., & Salmon, J. L. (2021). A method for urban air mobility network design using hub location and subgraph isomorphism. *Transportation Research Part C: Emerging Technologies, 125*, 102997.
Wood, S., & Reynolds, J. (2012). Leveraging locational insights within retail store development? Assessing the use of location planners' knowledge in retail marketing. *Geoforum, 43*(6), 1076-1087.
Xing, Y., Wang, K., & Lu, J. J. (2020). Exploring travel patterns and trip purposes of dockless bike-sharing by analyzing massive bike-sharing data in Shanghai, China. *Journal of Transport Geography, 87*, 102787.
Xu, H., Oh, L.-B., & Teo, H.-H. (2009). Perceived effectiveness of text vs. multimedia location-based advertising messaging. *International Journal of Mobile Communications, 7*(2), 154-177.
Yang, S., & Allenby, G. M. (2003). Modeling interdependent consumer preferences. *Journal of Marketing Research, 40*(3), 282-294.
Yang, Z., Lian, D., Yuan, N. J., Xie, X., Rui, Y., & Zhou, T. (2017). Indigenization of urban mobility. *Physica A: Statistical Mechanics and its Applications, 469*, 232-243.
Yu, H., & Shaw, S. L. (2008). Exploring potential human activities in physical and virtual spaces: A spatio-temporal GIS approach. *International Journal of Geographical Information Science, 22*(4), 409-430.
Zha, H., He, X., Ding, C., Gu, M., & Simon, H. D. (2002). Spectral relaxation for k-means clustering. Advances in neural information processing systems,
Zhang, K., & Katona, Z. (2012). Contextual advertising. *Marketing Science, 31*(6), 980-994.
Zhao, Z., Koutsopoulos, H. N., & Zhao, J. (2020). Discovering latent activity patterns from transit smart card data: A spatiotemporal topic model. *Transportation Research Part C: Emerging Technologies, 116*, 102627.
Zubcsek, P. P., Phan, T. Q., & Lu, X. (2016). Homophily and influence: pricing to harness word-of-mouth on social networks. *Available at SSRN 2562167*.



Impact paragraph

Despite noticeable progress that has been made in the past decades towards citizen-centric city services and sustainable urban living, research on geography in marketing and consumer behavior in cities is still underrepresented. The era of ubiquitous computing has given researchers many new opportunities to accelerate existing research streams and to rethink the boundaries of research fields. Ubiquitous computing devices are omnipresent in everyday life and the sensory data they collect allow for real-time observation of actual consumer behavior. The opportunities inherent in these datasets are enormous; using carefully selected methods to translate raw urban behavior and mobility datasets into meaningful implications for public and private stakeholders is the big challenge I addressed in this dissertation.

The combination of data analytics and social science with high applicability to practice was a core focus of the BISS Institute, where I spent the first 2 years of my PhD. My supervisor Jos Lemmink, initiator of the BISS Institute, imagined this research institute on the Brightlands Smart Services Campus (BSSC) in Heerlen to be a ”Test Kitchen” where innovative, out-of-the-box, and interdisciplinary research ideas would fall on fruitful ground and where experimenting with the diverse data landscape of the ubiquitous computing era is supported and appreciated. This mindset has accompanied me throughout the past six years. Consequently, in my dissertation, I use urban (sensing) datasets and I examine five ways of integrating these into marketing research for deeper insights on human behavior in urban areas and on the future of location-based marketing. The resulting implications are threefold: conceptual, methodological and societal. The following paragraphs will elaborate on each of these aspects.

Concerning conceptual implications, in the introduction of my dissertation, I define LBM as *a direct marketing strategy that uses location information to deliver marketing content on mobile devices relevant to a particular location*. In Chapter 1, it becomes evident that LBM currently applies a rather narrow definition of location and mainly uses the current consumer positioning to place advertisement for nearby shops. At the end of the first chapter, I argue that researchers should broaden the definition of location and include dynamic consumer movements and static location information of the surrounding area. The following five chapters demonstrate use cases and approaches for LBM research with a broader definition of location. Thereby, these chapters advocate for LBM to approximate the definition of CBM. CBM refers to *a geographically bounded form of LBM that includes more data about the city and urban living*. Depending on the goal, the scope, and the datasets of an LBM research project, the geographical boundaries are not necessary or not feasible for reasons such as privacy. However, they are helpful for deriving clear implications for a target region.

Apart from extending the LBM definition towards the CBM definition and including more data beyond the static customer positioning, I challenge the long-standing proposition that the consumer is the focal unit of analysis in marketing research. In the end, I agree with this proposition because marketing usually aims at selling a product, services, or idea to a consumer in the most appropriate and efficient way. At the same time, I believe that many other topics and units of analysis can the studied to learn more about human behavior and after all, to deliver a marketing message successfully. In the context of this dissertation, we saw that even though the consumer is still at the focal point in behavioral research, their decisions and choices are complex (Chapter 1).

Infrastructure and service performance determine their choices and routines (Chapter 2), locations can be attractive to consumers for multiple reasons (Chapter 3), and people with similar demographics behave and choose alike (Chapter 4). Data privacy restricts research opportunities (Chapter 5) and all these aspects directly or indirectly influence consumers' activity-based mode of transport choice (Chapter 6).

Regarding methodological implications, my dissertation disseminates the potential of spatial data in marketing. Spatial data can include static information about the landscape, built environment, or physical infrastructure of an area as well as dynamic observations such as human trajectory data, traffic volumes, pollution levels, etc. Many of these observed circumstances influence or are influenced by humans and their daily movements in their environment. Consequently, marketing research would benefit from integrating more such datasets into research projects that have a clear connection to the physical context. Throughout my dissertation, I highlight the benefit of spatial data in my research projects and point out that some insights would not have been observed if I had not used these datasets. Precisely, in Chapter 2 and Chapter 4, I demonstrate that both infrastructure and human trajectory data can assist researchers in uncovering service gaps in a city. In a second step, similar datasets and additional neighborhood information from national statistics centers allow developing a deeper understanding of neighborhood dynamics and consequently, developing personalized services and curated marketing offers that fit the need of citizens.

As a second methodologically contribution—and in line with the guiding thought of researching in a "Test Kitchen"—, I tested and critically assessed five different approaches to spatial data in marketing and corresponding methods that fit my use cases. As a result, I work with MIT's UNA toolbox to identify optimal UAM VTOL sites from a demand perspective (Chapter 2), and I employ regression analysis on spatial data to model location attractiveness based on its surrounding infrastructure (Chapter 3). Further, I spend time becoming acquainted with the world of data visualization for geographic data and utilize a variety of visualization options to discover urban dynamics (Chapter 4). I show that careful data pre-processing, clever feature engineering and the DBSCAN clustering algorithm can bridge data gaps from privacy regulations; with this unsupervised method, I reverse engineer the trip purpose of recorded movements, identify common characteristics of each cluster and categorize clusters into broader purpose categories (Chapter 5). Additionally, I apply k-means clustering to activity-based mode of transport choices observed via a complementary survey and describe four distinct behavioral clusters of mobility users (Chapter 6). For generalizability, I ensured that all introduced methods are transferrable to other cities and regions as long as suitable datasets are available.

The last set of implications—societal implications—are the most important ones for applied research from a "Test Kitchen". Universities are government's and society's institutions for the creation of new knowledge; in the past decades and centuries, universities have had a major impact on developing the society we live in today. I strongly believe that research always needs to have a societal and practical impact, because if a project does not help anybody or does not advance practices of any societal group, then what is the point of running the project. To make sure my research would have concrete practical implications, I worked together with the City of Maastricht to recruit participants for my data collection and I used additional information from CBS and BAG to enrich my datasets. This gave me a concrete geographically bounded area of research and knowing the region from living there was an additional benefit.

Applying the aforementioned methods in Maastricht and South Limburg, I find the following practical implications for the region. First, in Chapter 2, I show that there is a transport service gap in some areas of the city and a subset of these is suitable for UAM services. By overlaying multiple data layers, I propose four VTOL sites across the city and argue why they would be frequently used and by whom. More precisely, I identify two locations in the city center, which people would be using for direct regional flights to avoid traffic or to reach the densely built city center fast, and two in the outskirts of the city in wealthy neighborhoods, which residents would use for fast intra- and intercity connections.

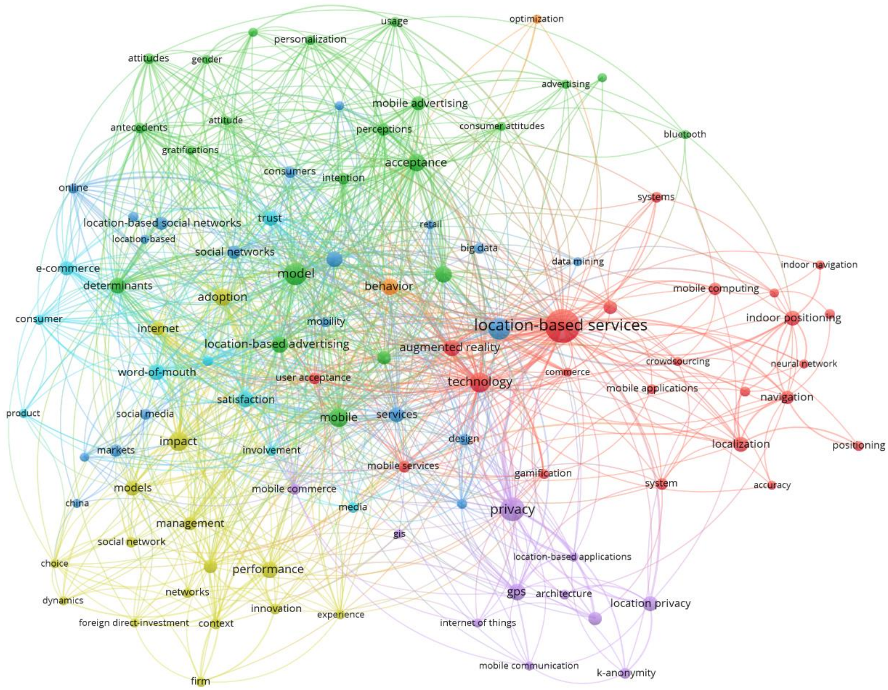
Second, in Chapter 3, I divide the region into 300m grid cells and demonstrate that we can use people inflow to a grid cell as an indicator of location attractiveness, especially in a marketing and behavioral context. When testing the direct effects of destination features on location attractiveness, I find that denser areas are more popular than rural areas and that certain grid cell functions (used as a proxy for trip purpose) are more popular than others are. Further, I find that competition between businesses has an effect on location attractiveness and that public transport access varies in importance by cell function. Additionally, the results indicate that trip origin features moderate the attractiveness of the destination location, meaning that depending on the function of the origin, certain destination functions are more popular than others are and we can clearly differentiate between chain trips, short trips within the direct surroundings and trips to or from home.

Third, in Chapter 4, different data visualization techniques uncover urban dynamics and service gaps previously unrecognized. Here, the combination of multiple graphs and representations is key and data visualization has the unique ability to present multiple dimensions of a dataset in an understandable way through colors and glyphs. For Maastricht and South Limburg, I detect traffic peak hours which differ between weekday and weekend; I find that some cell functions are more dominant in specific areas of the region and that the flow to and from these functions differs during the day. The heat maps of flow hotspots show the main activity in Maastricht, Heerlen and Sittard and I compare these results between weekdays and with a hotspot analysis of buildings. Moreover, I illustrate the flow links by mode of transport; there seems to be a minority of people who commute between Maastricht and Heerlen or Maastricht and Sittard by regional train as well as a few bus trips within Maastricht. Walking is naturally bound to short distances within cities. Cycling mainly takes place within city boundaries, with few movements to the closest northern and southern suburbs of Maastricht.

Fourth, in Chapter 5, I cluster all observed movements and identify 12 categories, which I can label with a certain trip purpose based on the common characteristics of each cluster. Such characteristics include average trip length and duration, mode of transport used, weekday, time of the day, urbanization level, and cell function. I find clusters of trips where people go for lunch during a break from work or study, where people go home after work or after leisure activities or where people leave their home in the morning to go to work or school. Further, I am able to group the 12 clusters into four broader themes: food-related trips, home trips, career-related trips and sports-/health-related trips. Having an understanding of the most likely purpose behind a trip is interesting for public and private stakeholders alike. Amongst other use cases, public stakeholders can provide the necessary transport services and infrastructure to manage the flow volume and private stakeholders can use trip purpose information to market their business.

Last, in Chapter 6, I use activity-based mode of transport choice patterns to allocate citizens to one of four behavioral clusters: Leisure drivers, persistent drivers, frequent walkers, and persistent cyclists. People within one cluster have very similar choice patterns whereas people in other clusters choose quite differently. More precisely, I find that the persistent cyclists build the largest cluster in the region and that most of these people live within the city boundaries of Maastricht. Together with the frequent walkers, they show a high environmental consciousness, whereas persistent drivers care less about sustainability. Thus, in a sustainable mobility campaign for Maastricht, the persistent cyclists and frequent walkers should be rewarded for their non-motorized transport use to keep them motivate to behave similarly in the future, while persistent drivers need to be targeted differently with incentives such as less parking issues, less traffic, and higher flexibility.

In summary, integrating geography research into marketing and behavioral research is still a big challenge and there is a long way left to go. However, we see some first promising steps of interdisciplinary research in this area and continuous attempts to deploy new algorithms and methods in social sciences use cases. Additionally, researchers worldwide increasingly value applied research projects and pay close attention to the societal impact of their work.


optimization
usage
personalization
attitudes
gender
advertising
mobile advertising
antecedents
attitude
consumer attitudes
perceptions
bluetooth
gratifications
acceptance
consumers
intention
online
systems
trust
location-based social networks
retail
location-based
social networks
big data
data mining
indoor navigation
e-commerce
model
determinants
behavior
mobile computing
adoption
consumer
mobility
location-based services
indoor positioning
internet
location-based advertising
augmented reality
crowdsourcing
neural network
word-of-mouth
commerce
user acceptance
technology
mobile applications
satisfaction
navigation
product
social media
mobile
services
impact
design
localization
positioning
markets
involvement
mobile services
gamification
models
system
accuracy
mobile commerce
china
media
privacy
management
gis
social network
location-based applications
choice
performance
networks
gps
architecture
location privacy
dynamics
innovation
experience
foreign direct-investment
context
internet of things
mobile communication
k-anonymity
firm

Summary

To deliver specific advertising for products, services and events in their city to potential customers and to increase customer experience with the brand, companies are starting to engage in hyper-targeting. Developing these super-specific, curated marketing offers works best with much information from multiple data sources. Thereby, location information is an impactful component, especially for local events and geographically bounded deals. Consequently, firms can use location-based marketing to sharpen their hyper-targeting efforts and spatial marketing analytics can shape the future of hyper-targeting research and practice. However, current literature on location-based marketing mainly stems from computer science and engineering, even though social science and management is home to this field. Few marketing studies use spatial data from proxemics technologies and we know little about its worth and the technicalities how to include such data in marketing research. In this dissertation, I address this research gap in six distinct chapters.

In the first chapter, "Location-Based Marketing as an Iceberg – Jumping into Unknown Waters to Look beyond the Surface", I review previously published work on location-based marketing to get an understanding of the topic from the business domain and to filter out why marketers tend to be dissatisfied with their location-based marketing effort. I conclude that the current scope is much too limited, because it focuses on static consumer positioning observed via geo-fencing and online behavior. As a mitigation, I introduce examples of geographic data and elaborate on techniques that allow integrating these datasets into marketing research. Such spatial data and respective methods constitute the foundation of the following five chapters where I add physical context and in particular the dynamic perspective to location-based marketing.

In the second chapter, "Sketching the City Space I – A Multi-Layer Approach to Preparing for Future Urban Air Mobility (UAM) Services", I demonstrate that the built environment is one core location aspect and that it enables us to understand the city as a servicescape better. Based on information about all buildings in the city of Maastricht and the physical infrastructure network of roads, I run a building interconnectivity analysis and I uncover service gaps in the current public transport service landscape of the city. Urban Air Mobility is an uprising concept in the transport industry that is said to soon fill this service gap in large metropolitan areas, but some problems still need to be solved before its implementation. Using a multi-layer approach, I show how cities can figure out ideal placements of vertical take-off and landing sites for air vehicles and exemplary I indicate where these spots would be in Maastricht and why.

In the third chapter, "Sketching the City Space II – Using spatial data to model infrastructure influences on location popularity", I add more factors than buildings and their interconnectivity to determine what makes a location attractive to citizens. I hypothesize that the circumstances at the destination, i.e. density of an area, the dominant function of an area, the competition between businesses of the same category and the availability of public transport solutions influence location popularity and that the circumstance at trip origin moderate this effect. Using a regression model, I provide evidence that (a) all factors have a direct effect on the location popularity, (b) public transport availability increases location popularity of some destination functions and (c) that certain origin-destination function combinations impact popularity.

In the fourth chapter, “Let the Pictures Talk – A data Visualization Approach for Revealing Dynamics in Complex Urban Datasets and Movement Structures”, I move the citizens into the focus of the analysis. Moreover, I step away from looking at the built environment and other static infrastructure features. Instead, I analyze human trajectory data of Maastricht’s citizens to uncover urban dynamics. Working with a variety of data visualization tools, I show that this approach is an excellent complement to traditional computational methods, especially when working with complex datasets. Temporal distribution graphs or heat maps provide a nice overview of the data, while flow analyses or hotspot analysis allow deeper insights. For example, in a flow analysis, I use colors and glyphs on a map to show where traffic during different times of the day originates and terminates and how that is related to building function at destination. Moreover, I am able to differentiate urban dynamics by transport mode and draw valuable conclusions for public authorities.

In the fifth chapter, “Bridging the Data Privacy Gap – An Attempt to Reverse Engineer Trip Purpose from Smartphone-Based GPS Trajectories”, I present an attempt to bridge the gap that researchers face because of data privacy regulations. Personal privacy is one of the highest goods and thus, restrictions are important. Still, they limit research in using the ideal data sources and especially in combining sources. Through a combination of multiple non-critical datasets and clever feature engineering that preserves all personal privacy, I demonstrate that unsupervised machine learning techniques can assist researchers to reverse engineer unobservable aspects about the phenomenon. In particular, I am able to reverse engineer the trip purpose from human trajectory data clustered into twelve distinct purposes, which I group into four broader categories.

In the sixth chapter, “City-Based Marketing for Behavioral Change – A Holistic Approach to Consumers’ Activity-Based Mode-of-Transport Choice Patterns”, I draw on aspects of the previous five chapter to provide a holistic and detailed picture of consumer’s behavior and decision-making in urban contexts. I incorporate the dimension of transport mode into activity-based travel behavior by hypothesizing that people’s mode choices depend on their activities at trip destinations. In a complementary survey conducted together with the tracking data collection, I observe respondents' transport-mode choices for specific activities. I use the resulting data matrix, in combination with additional consumer information, to cluster consumers with similar mode-activity patterns into four distinct groups. Between these identified groups, I find differences in transport mode choice for various activities, but also differences in personality or lifestyle. I also derive implications for business and sustainable transport policy makers and conclude how these insights can improve hyper-targeting in the future.

About the author

Hannah Sophie Schmitt was born on August 7th 1993 in Karlsruhe, Germany. In 2012, she obtained her Abitur from Heimschule Lender in Sasbach; during her high school time, she also spent one year at Sant Bani School in Sanbornton, New Hampshire in the USA. In September 2012, Hannah moved to the Netherlands to pursue her Bachelor studies in International Business at the School of Business and Economics (SBE) of Maastricht University (UM). During her undergraduate studies, she spent an exchange semester at the University of Newcastle (UoN) in NSW, Australia; at UoN, she also met her co-supervisor Ben for the first time when he taught her tutorial in Advertising & Promotion Strategy during his Ph.D. years there. After completing her B.Sc. degree in June 2015, Hannah did two internships at Deutsche Bundesbank and Lufthansa Cargo AG in Frankfurt, Germany. In February 2016, she enrolled in the M.Sc. International Business program with a specialization in Supply Chain Management at SBE and graduated in January 2017 with distinction (cum laude).

In February 2017, Hannah joined the Business Intelligence and Smart Services Institute (BISS 1.0) via the Department of Marketing and Supply Chain Management (MSCM) as a Ph.D. candidate under the supervision of Prof. Dr. Jos Lemmink and Dr. Benjamin Lucas. During the first two years of her Ph.D. trajectory, she participated in several company co-creation projects of BISS and fellow BSSC members. In her role as a project researcher, she worked on challenges such as the Smart Service City Heerlen, parking garage occupation optimization for Q-Park Netherlands, or AI in the classroom at UM. After being integrated back into the department in spring of 2019, Hannah supervised various Master theses and Bachelor MaRBLe projects. Further, she presented her doctoral research at multiple industry symposia and leading international conferences such as Frontiers in Services 2019 in Singapore.

During her Ph.D., Hannah was also actively involved in the Ph.D. community at UM and the data science community initiated by the Institute of Data Science. In the latter, she co-founded the data science community for students, junior researchers and people just getting started with data science together her fellow Ph.D.s Chang Sun and Nadine Rouleaux. Additionally, she was engaged in the local Women in Data Science (WiDS) community at UM, where she organized the WiDS Datathon together with Chang Sun for three consecutive years in 2020, 2021, and 2022, and assisted with the local WiDS Maastricht conference in the same years. At the Ph.D. Academy Maastricht, Hannah was a board member and head of the sports committee for two years and she organized the annual ski trip for all Ph.D. candidates of UM to Davos, Switzerland.

Currently, Hannah works as a Project Lead in the Data-Driven Sales team of Mercedes-Benz Mobility AG in Stuttgart, Germany.