Geo-information Science and Remote Sensing

Thesis Report GIRS-2022-11

Walk this way

Assessing variables of spatiotemporal behavior of pedestrians on a national scale

Julia Sipkema

20-03-2023

WAGENINGEN UNIVERSITY & RESEARCH

Walk this way

Assessing variables of spatiotemporal behavior of pedestrians on a national scale

Julia Sipkema

Registration number 1101692

Supervisors:
Ron van Lammeren (WUR)
Sander van der Drift (Dat.mobility)
Tom Thomas (Dat.mobility)

A thesis submitted in partial fulfillment of the degree of Master of Science
at Wageningen University and Research,
The Netherlands.

20-03-2023
Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2020 -11
Wageningen University and Research
Laboratory of Geo-Information Science and Remote Sensing

Abstract

People travel from one point to the other often multiple times a day, using different modes of transport. Walking as a mode of transport is a large part of these daily movements of people because it is a cheap, nature-friendly, and healthy type of mobility. This is especially relevant regarding current issues like population growth, climate change, and increasing urbanization. Therefore it is important to model pedestrian intensities, showing where and how people walk and how this will be impacted by changes in urban fabric or other types of transportation.

However, walking is also a mode of transport that is often forgotten in research and traffic modeling. This is often due to a lack of data available with pedestrian research historically relying on counts and surveys. Furthermore, when pedestrians are included it is often focused on a small extent, like a neighborhood or city, due to this data availability.

This research shows that it is possible to estimate pedestrian intensities using the extent of a whole country, which can be used in pedestrian modeling. This was done by analyzing GPS panel data looking at both spatiotemporal and geographical patterns and comparing it with existing literature.

Patterns were found in pedestrian intensities that can be seen even on the scale of a province and a country. This includes the fact that people tend to walk more often in dense areas (98 vs 85 trips on average in a month), but they walk shorter distances (1094 meters per trip on average compared to 1180). Furthermore, it was found that people on average undertake 1.07 trips by foot per day with an average of 1.1 kilometers. The research also shows that there is a temporal difference in walking behavior: people walk the most on Saturday, 1.22 trips, compared to an average of 1.04 on other days and people walk longer distances on Sundays (1419 meters per trip on average, compared to a 1088 average on other days).

This research shows that although there are differences in the extent and research methods, the results found are comparable to earlier research and national mobility reports. This makes GPS mobile phone data a trustworthy source of data. It also shows that using a large data set can give insight into regional differences. The results show patterns in people's walking behaviors that can be used to create traffic models as well as useful coefficients to finetune them.

This research can be a starting point for more precise pedestrian traffic models. Furthermore, it can be seen as a basis for more research on mobility using large datasets, for example using mobile phone GPS data. Finally, this research can be used as an impulse for more research on pedestrians, as they are often underrepresented in mobility research, even though it is such an important type of mobility looking at climate change, crowding, and health.

Table of contents

1. Introduction

1.1 Context and background

Between 2020 and 2042 there is an expected rise of people living in cities from 4.2 billion people (around half the world population) to over 6 billion people (Kuddus, Tynan & McBryde, 2020). This high rate of urbanization can have drastic effects on life in cities. Examples of this are a decrease in the number of green areas in cities, crowdedness in the streets, and an increase in the number of people who experience health issues due to pollution and a lack of exercise (Lall & Wahba, 2021; Sokolovska, 2020). According to Rudolph & Mátrai (2018), however, a well-designed infrastructure and transportation system can contribute to cities expanding more sustainably and efficiently, curbing the aforementioned problems. For example, efficient planning can lead to cities being less congested and can leave more space for green. But transport is not only important for cities. It also impacts rural areas. These areas are, for example, less and less accessible by public transport, leaving them car-dependent (Monster, 2022). By having a well-designed infrastructure connection between these rural and urban areas, this car dependency can be decreased. Thus, researching mobility flows in cities, but also outside of cities can provide valuable insight for finding suitable solutions for the future (Barbosa et al., 2018).

A means of transportation that is often forgotten when investigating mobility and transport is walking, despite having an important role in the themes of health, crowding, and congestion described above. Promoting walking instead of using other types of transport can, for example, help with decreasing traffic, since fewer cars and bicycles will be used. It also helps against overcrowding of public transportation if people replace short trips with walking (Rudolph & Mátrai, 2018). Walking also is an important type of mobility health-wise. It not only decreases the chance of infectious diseases by avoiding direct contact with other people (Hunter et al., 2021). It also improves pedestrians' mental and physical well-being (Gómez et al., 2010).
Cities can benefit from good walkability, which can be seen as how easy it is to walk in a certain area and can be calculated using many different factors regarding land use and street design (Lo, 2009). However, the consequences of urbanization make walking in cities less and less attractive. Pedestrians prefer a safe, convenient, friendly, and green place to walk (Hsieh & Li, 2018). However, increasing crowdedness, density, and traffic challenge these factors, therefore decreasing the walkability in cities.

While walking is thus an important type of mobility that is not only threatened by current developments but also could have an important influence on future development, it is not an often-studied type of transportation in the field of mobility research. This is due to the dependence on counts and survey data, making research into pedestrians small-scale and labor-intensive (Lee & Sener, 2020). However, now, with the availability of more and more GPS data sources, more research on pedestrian intensities can and should be done. This research into mobility from pedestrians can add useful insights into creating solutions for issues regarding mobility, urbanization, and health. This can be done by using pedestrian traffic models, where the behavior of pedestrians and the pedestrian volumes in places of interest are simulated, for example by using space syntax (Yamu, Van Nes & Garau, 2021). Using a model can give insight not only into current situations but also into planned scenarios, meaning plans can be analyzed this way. Since these models can be very useful but require a lot of input, this thesis will look into which variables and assumptions are the most important to use when building such a model.

1.2 Theoretical framework

1.2.1 Mobility research

As mentioned in the introduction, mobility is an often discussed and relevant research topic. It is also a very broad topic with many possibilities for new research. And more and more possibilities are opening up with the increase in ways to collect and analyze data, for example using GPS and/or mobile phone data. Because of the many opportunities and possibilities, there is already quite some academic literature available on mobility intensities and the modeling of it. Within this broad range of articles, most of the focus has been on cars and public transport. This is because more types of data have been available for these modalities and since these types of transport have a bigger impact on the environment and livability in cities (De Wit et al, 2021; Lee & Sener, 2020).

Lee & Sener (2020) write that currently, more data sources for non-motorized travel (walking and cycling) are emerging, especially through the use of smartphones. They claim that this, however, mainly leads to more research on cyclists, meaning that research on pedestrian behavior is falling behind.

Nonetheless, pedestrian behavior and intensity can be an important opportunity for research. As mentioned in the introduction, insight into walking can, for example, be important during the urbanization process, since modeling pedestrian behavior can help in city planning by simulating the effects of changes in walking behavior after changes in the built environment (Gaber et al., 2020).

From the research that has been done on walking, they were mostly done using a small extent. For example, Bongiorno et al. (2019) look at the area of Greater Boston, De Wit, Versluis & Leferink (2021) look at Rotterdam and Kang (2017) looks at the city of Seoul. There is also a difference in the data they use in their research; Malleson et al. (2018) and Sevtsuk et al. (2021) use passive smartphone data, while Kang (2017) combines different geographical data sources, for example, information about roads, public transport stops, census tract maps information about walking volumes. Within all these different papers not only the extent and the data sources are different, but on top of that many methods and variables have been used. Here, the most important methods will be highlighted.

1.2.2 Traffic modeling

Questions about pedestrian movements and volumes can often be answered by using pedestrian traffic models (PTM). PTM can be defined as a computer simulation of individual pedestrian movement based on multiple variables, which then shows the flow of pedestrians between different zones (Seyfried et al., 2006).

There are multiple types of traffic models estimating traffic flows and intensities, which have been used in many pieces of research. Most traffic models are based on a four-step model.

As the name suggests this is a model that exists of four steps to visualize and predict traffic flows. These four steps are: 1) trip generation, 2) trip distribution, 3) modal split, and 4) allocation. These steps and a couple of corresponding methods are explained below.

1.2.2.1 Trip generation

The trip generation deals with the choice of movement: what trip is a person going to make? This information is based on the socio-economic attributes of the people living in the area of research and other parameters.

A factor that can be of influence on the volume of pedestrians in areas is temporality, meaning the change over time. Malleson et al. (2018) state that there is a daily temporal pattern. They show that there is a peak in pedestrian trips in the morning, at lunchtime, and in the afternoon. This has to do with the motivation for the trips. These trips are likely to go to work or school and home from them. There are also yearly patterns regarding pedestrian behavior. Bongiorno et al. (2019), for example, show in their research on Boston that people are less likely to walk in the summer when the weather is very warm.

A method to estimate this trip generation is by using a categorial analysis. Here, the main starting point is that different qualitative categories can have a distinctive effect on mobility. Jović & Depolo (2011), for example, explain how category analysis can be used to look at differences in transportation between four types of urban zones: central, other built-up areas, rural/suburban, and new residential. Basarić et al. (2016) take a look at another set of categories, not based on geographic attributes but on user attributes. In their research, they look at the impact of gender and age on travel behavior.

1.2.2.2 Trip distribution & modal split

The trip distribution and the modal split look at where people go and what type of transport they use. Here often a network is used together with a cost matrix. These two variables can be used to calculate the cost of travel between two points. People often tend to take the route with the lowest cost.

First, it is important to estimate where people come from and where they go. The attraction of places, therefore, is an important influence on trip distribution. Basu et al. (2022) explain that destination choice influences the routes people take. This means that when an area has a lot of attractive destinations, for example, schools, shops, and parks, the volume of people walking there is higher than when there are very few destinations. They also state that if there is a high intensity of pedestrians in a certain area or on a route people will change their route.

Second, as mentioned before, street networks influence the routes people can take and what places are easy to visit. Özbil & Peponis (2007) and Kang (2017) both analyze the effect of street networks on walking by modeling pedestrian movements in representations of these networks using GIS in Atlanta and Seoul respectively. While the research in Atlanta doesn't show significant results, the case of Seoul concludes that pedestrians prefer to walk on streets that are easy to navigate and where there are dense destinations. Sevtsuk & Basu (2022) analyze how the number of turns in possible routes determines the choice of routes pedestrians take in San Francisco and Boston by analyzing GPS traces. They state that pedestrians prefer to have the least amount of turns possible unless the route gets significantly longer. De Wit et al. (2021) also show that networks are important. They use space syntax, a set of theories based on the assumption that activity is mainly based on the configuration of places in a network, to estimate the volume of pedestrians in Rotterdam.

1.2.2.3 Trip allocation

Lastly, the allocation puts people on these networks and thus has to do with the route choice of people. There are many variables that can influence this route choice. Sevtsuk et al. (2021) approach this topic by using a large dataset with GPS traces from San Francisco and comparing the taken paths with alternative possibilities, taking multiple street qualities into account. For this, they use traces with generalized begin points and endpoints, to guarantee the privacy of the participants. According to their research, travel distance and travel time are the most important variables for route choice, while the number of turns, the broadness of sidewalks, and the presence of stores near the route also affected the route preferences. Basu et al. (2022), on the other hand, do not model the route choices. Instead, they systematically review factors influencing route choices. They conclude the importance of the factors mentioned by Sevtsuk et al. (2021) but add traffic volume, street crossings, route infrastructure, safety, and land use. In another research, Malleson et al. (2018), use passive smartphone data to understand the asymmetry in route choices in Boston. With asymmetry, they mean the degree to which people decide to walk different routes for their trips to and from places.

Route-based models

For the allocation of the trips on the network, there are also models that can be used, as mentioned before. A couple of them will be described below.

Basu et al. (2022) write in their research about two types of models. First, they mention the utility theory-based model, also known as a discrete choice model. This models the route of pedestrians with the highest value, which is called the utility here. The variables to determine the value of a route can be found using a logistic regression model, the second type of model. With this model, the associations between factors and route choice can be calculated. The most influential factors can then be used in the utility model to calculate the best routes.

In their research, Sevtsuk & Basu (2022) and Sevtsuk et al. (2021) use a path size logit (PSL) model. This model aims to find the correlation between different possible routes by taking path size into account. They use this model to estimate the factors influencing route choice, for example, path length and amount of turns. They state that this model is similar to multinomial logit (MNL), which assumes that all data is case-specific and each independent variable has an individual value for each case, but with a correction for path size to account for overlap between routes.

Bazzan & Klügl (2014) review in their research a different type of model to use for the modeling of traffic and transport: the agent-based model. These models consist of autonomous agents that interact with each other in a simulated environment. These agents can represent many things, for example, people, areas, vehicles, and traffic signals. By simulating how these agents interact with each other, flows of traffic can be predicted. De Wit, Versluis & Leferink (2021) also mention agent-based modeling in their research. They describe it as a micropedestrian model that describes how pedestrians walk and how they react to other pedestrians and objects in a very detailed manner.

These and many more types of models can predict all kinds of flows, including the modeling of pedestrian behavior and intensities. However, according to De Wit et al. (2021), pedestrians are currently not often the main focus of a model.

1.3 Problem definition

As stated before, there are many methods to analyze and simulate pedestrian intensities. However, most of the research done with modeling (both analyzing and simulating) focus on other types of mobility, mainly cars and public transport (De Wit, Versluis & Leferink, 2021). Secondly, when research has been done on pedestrian intensities and the analysis and modeling of it, few of them have done their research on the scale of an entire country. Most studies give an account of the research of pedestrians on the scale of a city. For example, Bongiorno et al (2019) looked at Boston, Kang (2017) focused on Seoul, and Sevtsuk et al (2021) used San Francisco as a study case. Some of these are quite large areas, almost half the size of the Netherlands. However, looking at the scale of an entire country could give insight into all commuting and travel activities in an area. While in the previously mentioned cities there is a big chance of commuting going outside the study area, by looking at the entire Netherlands it is possible to get almost all of the movements. The use of GPS adds to this by including all travel, while in other research based on counts and surveys often smaller trips are forgotten and the motives are less clear (Geurs et al., 2015). Furthermore, the use of a large study area also allows scaling down, enabling the possibility to diversify the research by also looking at provinces and cities. This can give a clearer insight into spatial distribution and intensities, while also being able to show regional differences. Within this research, all of these scientific gaps are addressed by focusing on pedestrians in the entirety of the Netherlands using GPS mobile phone data.

1.4. Overall research aim and research questions

The main objective of this thesis is to analyze the drivers of spatiotemporal intensities of pedestrians on a national scale, using both literature and trip data from the Dutch Mobility Panel (DMP), to be used as a basis for a Pedestrian Traffic Model (PTM). This objective can be split into three research questions:

- RQ1: What attributes and variables can be used to estimate pedestrian mobility?

- RQ2: What patterns of pedestrian intensities regarding selected user attributes can be found in the Dutch Mobility Panel?

- RQ3: What patterns of pedestrian intensities regarding geographical attributes can be found in the Dutch Mobility Panel?

2. RQ1: Literature review

Since the methods of the second and third research questions depend on the variables and attributes found in the literature review of the first research question, this literature is described first. In this overview, first, the methods are explained. Second, an exploration is done by looking at attributes of trips described in the literature, followed by a look into the variables used in earlier research. Then a quick explanation is given of which of these attributes and variables will be used in the analysis of research questions 2 and 3.

2.1 Review methods

To answer research question 1 twelve research papers were used to gather data. This amount of papers was chosen since these sources cover a broad range of topics and are therefore able to shed light on multiple parts of pedestrian mobility, while also showing some overlap making it possible to compare the outcomes. To analyze these documents a literature review was done. The intention of doing a literature review is to gather information on different themes systematically and objectively by splitting the texts into different themes (Bryman, 2016). To do this, the computer program Atlas.ti was used. Within this program, it is possible to upload articles, highlight parts of them and divide them into themes. This makes it more convenient to find parts of articles relating to certain topics and makes it easier to compare the different articles used.

In Atlas.ti all of the twelve documents were uploaded and nine codes were created. These codes are used as different 'themes' to which different parts of the text can be assigned (table 2.1). These codes can also have some overlap. In total 126 quotes were made, with most of them being about explaining variables.

Code	**Explanation**	**Amount of quotations**
Assumptions	The assumptions made about pedestrians before the analysis	15
Destination preferences	The production and attraction points (where do pedestrians leave from and where do they go?)	2
Explaining variables	The variables found to explain pedestrian behavior	71
Functions	The functions used to calculate pedestrian behavior?	8
Keywords	The keywords used by the authors to describe their research	8
Method	The methods used to analyze pedestrian behavior	23
Model	The models used to analyze and predict pedestrian behavior	8
Motivation	The reasons pedestrians are traveling	2
Scale	The scale at which the research is performed	4

Table 2.1: codes used for the literature review and their explanation

These quotes were then used to analyze three different parts of the literature. First, the assumptions regarding pedestrians are described. Then the variables mentioned in the literature and their impact on walking are explained. Lastly, some metrics used in the literature are shown.

2.2 Trip attributes

In research, there are always assumptions made and attributes given regarding the topic of investigation. In pedestrian mobility, there are also these kinds of assumptions and attributes about how pedestrians move and use space (table 2.2). A couple of them will be explored below.

2.2.1 Infrastructure

One of the attributes of pedestrian movement is that while pedestrian movement depends on the topography and geography of a place, pedestrians are not necessarily bound by roads or sidewalks (Barbosa et al., 2018). De Wit, Versluis & Leferink (2021) state that, for example, walking is the only type of transport that allows for a movement of 360 degrees around and even up and down. People can duck below low overpasses and climb over small obstacles if necessary. People also can vary speed during their trip; they can walk fast with a purpose, go for a stroll, sit down to enjoy nature, or do some window shopping. On the other hand, pedestrians are more dependent on other factors, for example, land use, topography, and weather conditions.

2.2.2 Flows

Another attribute regarding pedestrian behavior is that the main flows of movement are between centers of commerce (places with job opportunities) and between places where people live (Barbosa et al., 2018). This does not only mean that pedestrians move from their homes to work or university. There is also a counter-current from work and university to home. This relation, however, does seem to weaken the further the origin and destination are apart, also known as distance decay (Barbosa et al., 2018). On top of this movement, there is not only a big flow of people towards the centers of commerce but there is also a high expected intensity of pedestrians in the area itself. Yamu, Van Nes & Garau (2021), for example, assume that pedestrians prefer to walk in places with a high density of buildings, a high density of roads and links, mixed land use, and a mixture of activities. This means that it is expected that people tend to walk more often in urban areas and city centers than in suburbs and smaller towns.

2.2.3 Distance

It is also known that non-motorized trips (e.g. walking and cycling) are shorter on average than motorized trips (Lee & Sener, 2020). According to Kang (2017), people tend to walk more when destinations and origins are close. This also relates to the previous attributes regarding more walking in city centers; greater closeness implies that there is a higher density of destinations (Kang, 2017). This closeness has the most effect on trips to 1000 meters. Bongiorno et al. (2019) also show similar walking distances. They state that most pedestrian trips are between 300 and 1000 meters. On the other hand, it is assumed that people do not automatically take the shortest walking route. According to Sevtsuk & Basu (2022), people prefer taking routes with fewer turns over the shortest possible route.

Dat.mobility (2022) also state some attributes about pedestrians in their starting points for a pedestrian model. They state that people make around 4 journeys a day and that around 60% of all trips are bound to people's homes. They also state that the average distance a person walks should be around 500 meters. This is quite low compared to the other literature, but this could be due to under-registration since walks for leisure are not included in their dataset.

Although all of the attributes mentioned are often defined before starting the research, all of them are possible to verify using measured trip data, for example by using GPS tracking.

Category	Attribute	Author
Unbound by infrastructure	- pedestrians are not dependent on roads and sidewalks	Barbosa et al., (2018)
	- pedestrians move 360° round and even up and down. - varying speed	De Wit, Versluis & Leferink, (2021)
Flows between centers of commerce and homes	- main flows between centers of commerce (jobs, school) and homes - distance-decay	Barbosa et al., (2018)
	- high intensities in places of commerce	Yamu et al., (2021)
Trip distance	- more walking when destinations are close - up to 1000 meter	Kang (2017)
	- people usually walk between 300 and 1000 meters	Bongiorno et al., (2019)
	- people do not always take the shortest route - least amount of turns	Sevtsuk & Basu (2022)
Trip attributes	- people make 4 journeys a day - 60% of trips are to or from home - average distance is around 500 meters	Dat.mobility (2022)

Table 2.2: an overview of the attributes found in the literature

2.2.4 Attributes explored further

Looking at these attributes found in the literature, there are quite some that can be calculated using the DMP combined with other data sets. This will be done in research questions 2 and 3. However, all research has its pros and cons. After these calculations, it is therefore interesting to compare the findings in the literature with the DMP, which will be done in the discussion. This is especially interesting due to differences in the measurement of the data. Most pedestrian research traditionally has an under-registration of trips since short trips are often forgotten in surveys and trips for leisure are not included.

A couple of the attributes found are a bit more difficult and will not be included in the analysis due to either lack of data or time constraints. These are the attributes regarding infrastructure since infrastructure data is not included in this research. The same goes for the number of turns in routes. The flow between home and work will also not be explored.

However, there are a couple of attributes that will be explored. A more general analysis of trip attributes will be done in the second research question. Attributes included here are the trip distance, amount of trips and journeys, and trips connected to home. These attributes can all be easily calculated from the DMP. In the third research question, the attribute that people walk more in places of commerce will be explored by looking at differences in walking per density of shops and meeting places.

2.3 Variables

All of the attributes explained in the previous paragraph are important starting points for pedestrian research. Keeping them in mind, it is now important to research the variables influencing pedestrian intensities. To predict and simulate pedestrian movements variables explaining pedestrian intensities are necessary. In academic research, many of these variables have been found. In their research, Goossen et al. (2021), for example, indicate 160 factors that influence walking. They divide these variables into the following 9 main indicators: 1) infrastructure, 2) quality of the environment, 3) the directness of facilities, 4) accessibility, 5) liveliness, 6) allure, 7) safety, 8) attractiveness, and 9) signage. For readability, these main indicators of Goossen et al. (2021) are used to structure the exploration of the variables below. Variables from other research are then added within these main indicators with a short overview of the most important variables at the end of this section (table 2.3).

2.3.1 Infrastructure

Regarding infrastructure Goossen et al. (2021) name two main indicators: walkability and separation of mobility types (figure 2.1). Walkability here indicates exclusivity for pedestrians. Walkability has a positive impact on walking, meaning that if people can walk more calmly without having to keep other types of mobility in mind, they will walk more. The separation of mobility types goes hand in hand with this. Here, they look at how many other types of mobilities are next to footpaths, looking for example at bicycle lanes and roads. Kang (2017) also states in his research that a mix of mobility types on one road negatively influences how much people walk there.

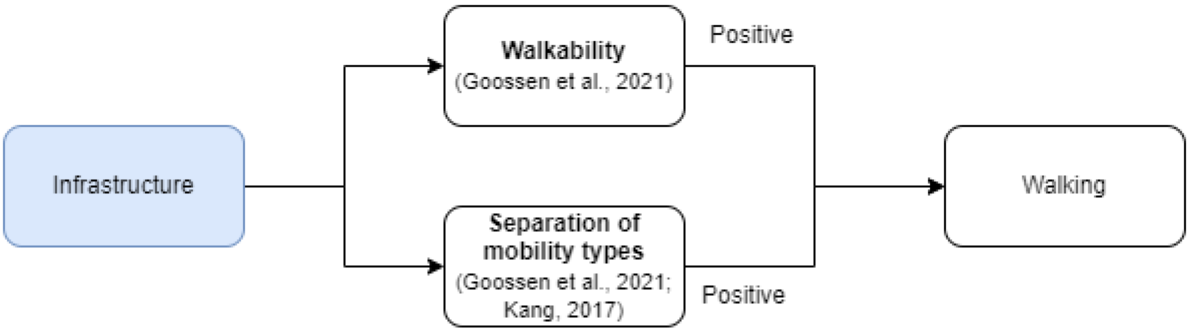


Figure 2.1: Overview of the infrastructure variables and their influence

2.3.2 Quality of the environment

Looking at the environment, Goossen et al. (2021) name five different indicators influencing it. These are noise, quality of air, heat, the chance of flooding and lastly, cooling by nature and water (figure 2.2). The first four indicators have a negative effect on walking and the last one a positive one. Sevtsuk & Basu (2022) state as well that noise has a negative impact on walking. They add to this that weather also has an important impact. If people can walk sheltered from rain, extreme sun, and wind they will prefer to walk there instead of places where they are exposed to these conditions.

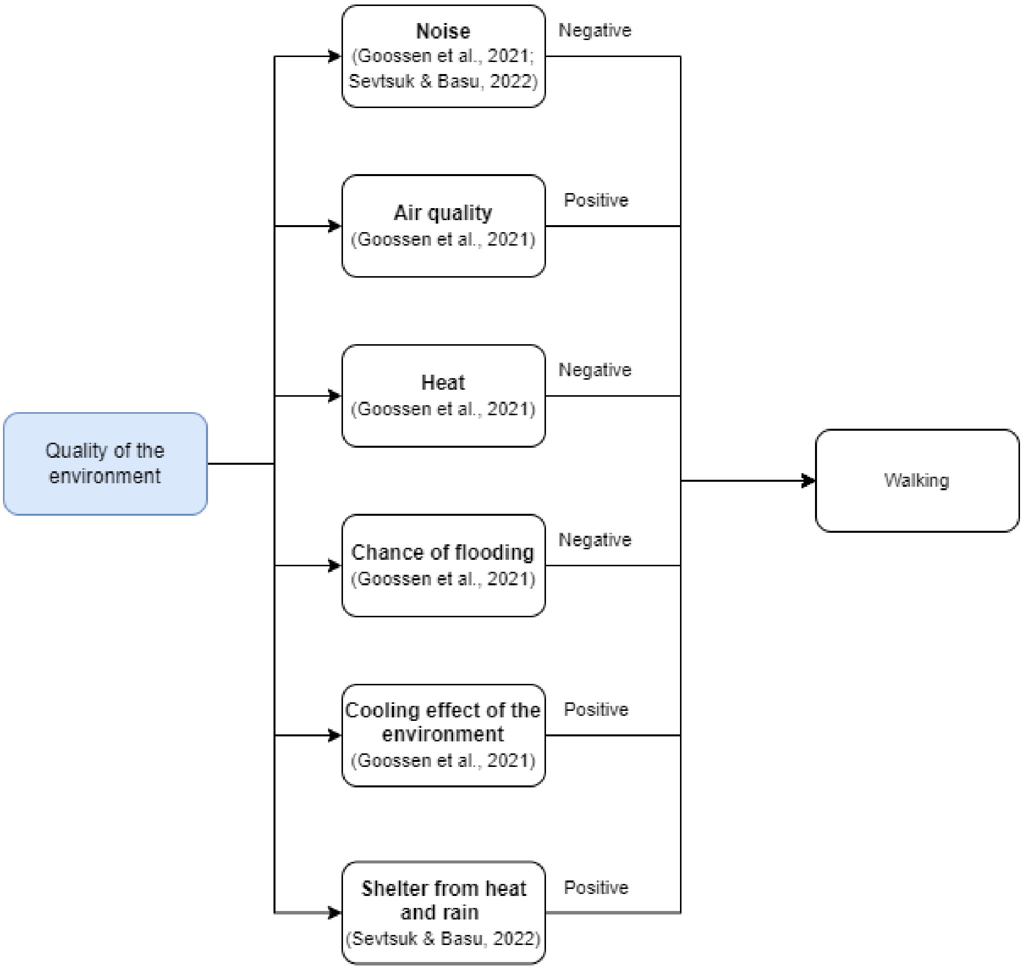


Figure 2.2: Overview of the quality of the environment variables and their influence

2.3.3 Directness of facilities

The indicator directness of facilities, according to Goossen et al. (2021), can be defined as how easy it is to reach a destination. This is measured by looking at variables such as the number of essential services in the area and the amount of public transport stops in the area (figure 2.3). Kang (2017) also mentions the availability of public transport as an important variable. De Wit, Versluis & Leferink (2021) also state that these two variables, public transport, and services, are important and add the density of buildings to it. All of these variables have a positive influence on walking.

Basu et al. (2022) also see directness as an important influence on walking in their research. They state that trip distance is important; the shorter, the better. Kang (2017), Sevtskuk et al. (2021), and Malleson et al. (2018) mention this as well. Adding on to this, according to Basu et al. (2017), shorter walking- and waiting times also impact walkability positively. They also say that pedestrians prefer to avoid hilly walks, as do Sevtsuk et al. (2021). Malleson et al. (2018), Sevtsuk & Basu (2021) and Sevtsuk et al. (2022) also state that the amount of turns influences the directness. People will take longer routes if that means that they have to take fewer turns.

Dat.Mobility (2022) also sees directness as important, but in another way. They state that if places are difficult to reach by car, for example, in city centers, people tend to walk more. They also say that the intensity of stores has an impact on walking, with more people walking if there are more stores.

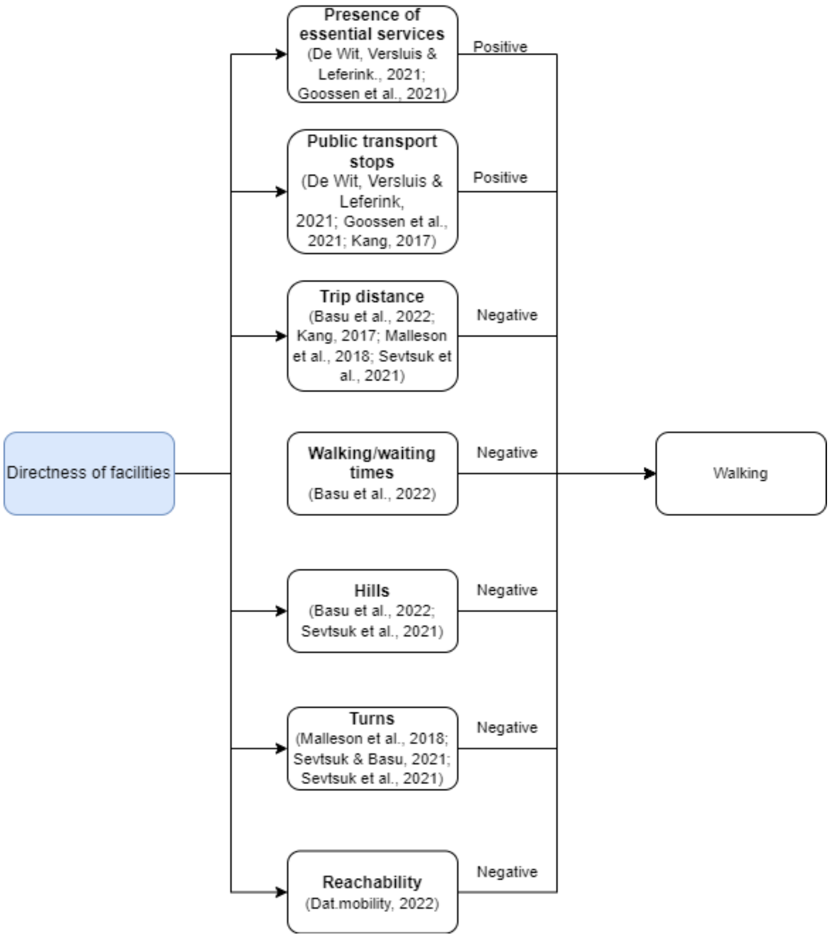


Figure 2.3: Overview of the directness of facilities variables and their influence

2.3.4 Accessibility

Regarding accessibility, meaning how easy it is to walk in an area, many variables influence the amount of walking (figure 2.4). Goossen et al. (2021), for example, name the broadness, length, and hardening of the pavement as important. Basu et al. (2022), De Wit, Versluis & Leferink (2021) and Sevtsuk et al. (2021) also state that pedestrians have a preference for broad and continuous sidewalks. Regarding footpaths, variables such as manhole covers, grates, and guidance lines have a negative influence (Goossen et al., 2021).

Maintenance is also important here. Tree roots or holes in the pavement decrease the accessibility for people with limited mobility, for example, those using a wheelchair or a walking stick. Other obstacles also have an important influence on accessibility. Trees, garbage cans, bicycle parking, containers, poles, and parked cars can all have a negative effect on how easy it is to walk in a certain area (Basu et al., 2022; Goossen et al., 2021).

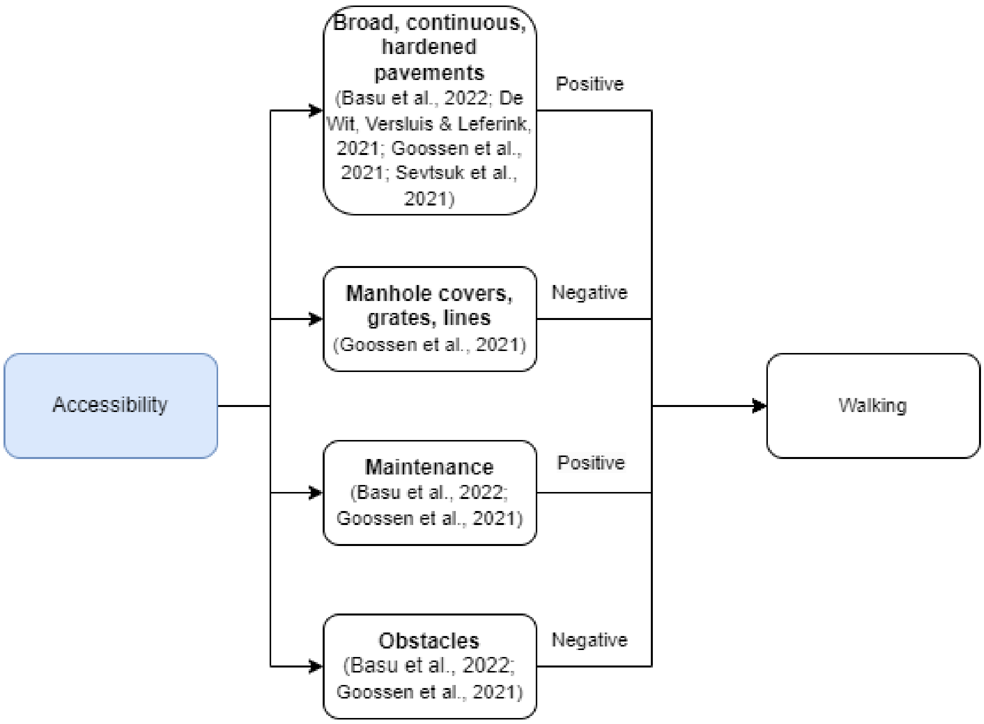


Figure 2.4: Overview of the accessibility variables and their influence

2.3.5 Liveliness

According to Goossen et al. (2021), a lively street increases walkability. This means that when there are other pedestrians or when there are things to do or see in an area, more people will walk there. They state that empty buildings have a negative influence, while mixed working and living, playgrounds, benches, canopies, and stores all have a positive influence (figure 2.5). Kang (2017) also mentions mixed residential and retail areas as positive. He adds to this that a high density of residents is also preferable to pedestrians. Sevtsuk et al. (2021) also mention the presence of shops as a positive influence on walking. On the other hand, Basu et al. (2021) state that too many pedestrians in an area have a negative effect on people's choice to walk. They also say that, interestingly, abandoned buildings do not have a negative effect on walking, but a positive one.

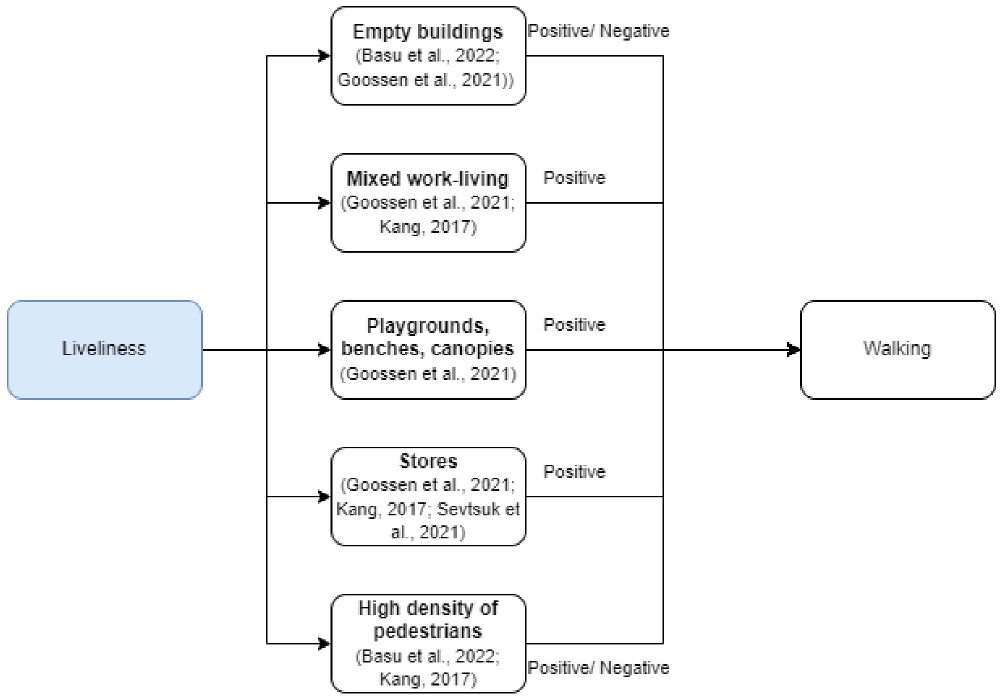


Figure 2.5: Overview of the liveliness variables and their influence

2.3.6 Allure

This indicator, allure, describes how attractive an area is to pedestrians, but in a broad sense. Allure can be seen as an unexplainable, mysterious, and powerful attractiveness of a place. Goossen et al. (2021) use the following indicators here: old buildings, museums, castles, theaters, monuments, boulevards, and the heights of buildings. Here they state that having high buildings can give character to an area, which can be seen, for example, in business districts of large cities. Kang (2017) states as well that mixed-high-rise development increases walking. Basu et al. (2022) on the other hand state that tall buildings have a negative influence on walkability (figure 2.6).

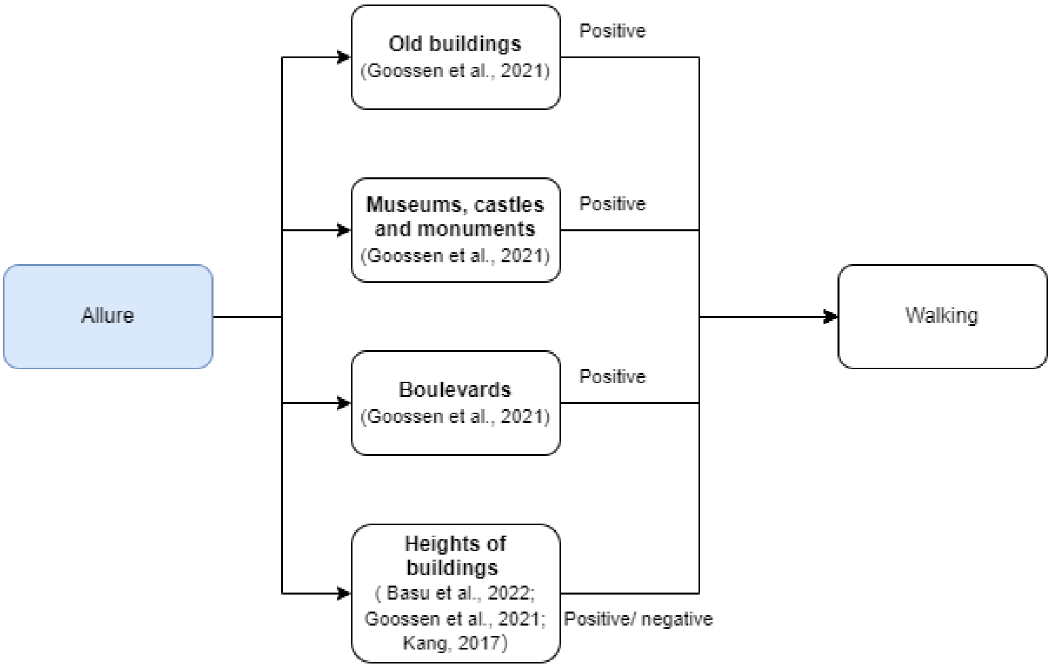


Figure 2.6: Overview of the allure variables and their influence

2.3.7 Safety

Safety is another important factor that influences where people want to walk. The variables that influence this, according to Goossen et al. (2021), are the maximum speed of cars, street lights, camera surveillance, and crossings with traffic lights or a zebra (figure 2.7). Basu et al. (2022) and Kang (2017) also mention safe crossing and street lights as important factors. They add the amount of crime and traffic volume as important safety variables having a negative effect on walking, as do Sevtsuk & Basu (2022) and Sevtsuk et al. (2021).

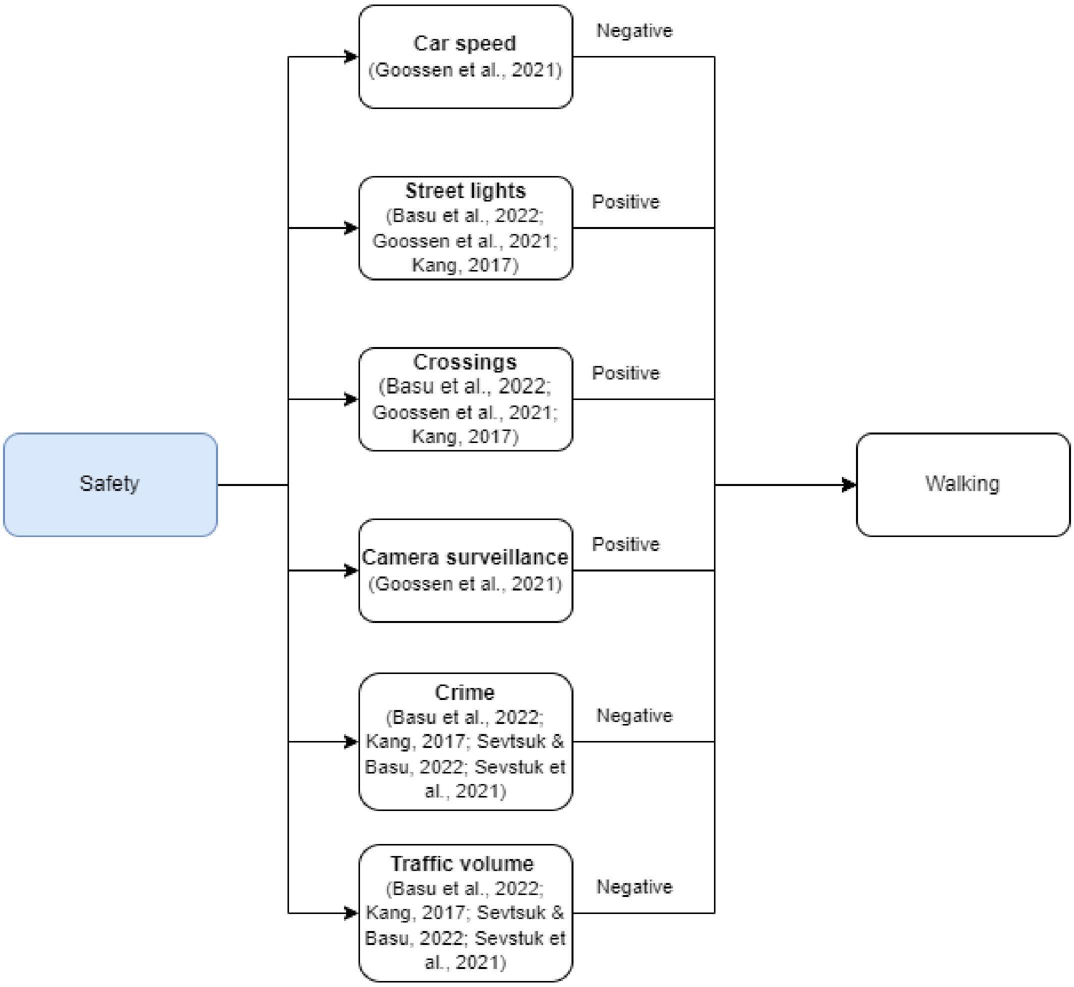


Figure 2.7: Overview of the safety variables and their influence

2.3.8 Attractiveness

According to Goossen et al. (2021), attractiveness is the indicator influencing walkability with the most variables (figure 2.8). Here variables such as nature and water have a positive impact. Art, monuments, fountains, and mixed land use also have a positive influence. According to Basu et al. (2022) green (for example parks) and blue (lakes, rivers) do indeed have a positive influence on how many people walk in an area, as is the presence of stores. De Wit, Versluis & Leferink (2021) also state that the presence of nature invites people to walk more.

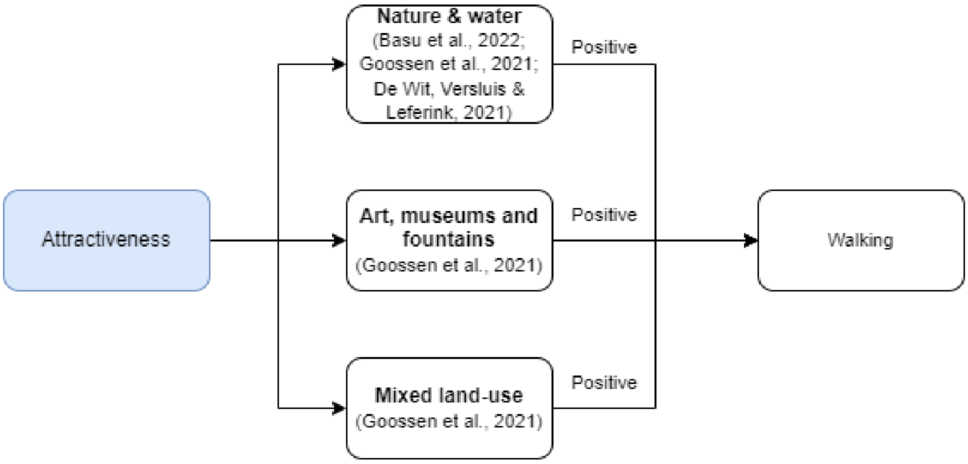


Figure 2.8: Overview of the attractiveness variables and their influence

2.3.9 Signage

The last group of variables according to Goossen et al. (2021) influencing pedestrian behavior is signage. Here, the presence of street names, maps, tourist information, and route signs makes walking easier for pedestrians and thus influences walkability positively (figure 2.9). Basu et al. (2022) also found route signs to have a significant influence on pedestrian behavior.

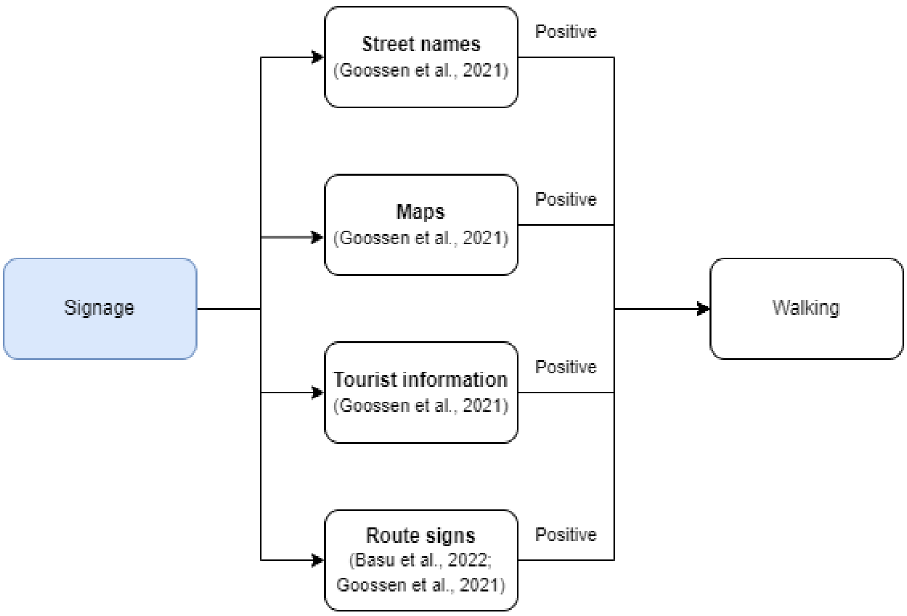


Figure 2.9: Overview of the signage variables and their influence

Category	Variable	Impact
Infrastructure	Separation of mobility types	Positive
	Walkability	Positive
Quality of the environment	Noise	Negative
	Shelter from wind and rain	Positive
Directness of facilities	Essential services in the area	Positive
	Public transport stops	Positive
	Turns	Negative
	Trip distance	Negative
Accessibility	Broadness, length, and continuity of sidewalk	Positive
	Obstacles in sidewalk	Negative
Liveliness	Empty buildings	Positive/negative
	Mixed work/living	Positive
	Presence of other pedestrians	Positive (to a certain amount)
Allure	High buildings	Positive/negative
	Old buildings, museums, monuments	Positive
Safety	Maximum car speed	Negative
	Street lights	Positive
	Pedestrian crossings	Positive
	Crime	Negative
Attractiveness	Green (parks, forests)	Positive
	Water (creeks, lakes, rivers)	Positive
Signage	Route signs	Positive
	Tourist information, maps	Positive

Table 2.3: most important variables per indicator

2.3.10 Relationship between variables

All of these variables do not exist in a vacuum; they do not only influence walking, but also influence each other. The literature used does not show assigned weights or interactions between variables, but the influence of variables on each other can be estimated. For example, mixed work and living have a positive influence on walking, but a mix usually means there are fewer essential services, which in turn has a negative influence on walking. And while the presence of services and buildings has a positive influence, this often means that there is less green and water in the area, which has a negative effect. It is thus important to keep in mind that while one thing can have a positive effect on walking, it can also indirectly influence walking negatively and the other way around.

2.3.11 Assumptions explored further

Although all of these variables described above do have a significant impact on walking, they are difficult to quantify. Looking at the data available only one of them can be easily explored: the distance of trips. A quick analysis of the effect of distance will be done in research question 2.

There are, however, a couple of other variables that relate to each other and can be estimated using the DMP data in combination with information about shops and services. These are the separation of mobility types, essential services, mixed work/living and old buildings, museums, and monuments. All of these variables seem to be connected to (sub)urban zones, meaning that, according to the literature, people should be walking there more. This will be explored in research question 3 by doing a category analysis using the density of shopping and meeting places as an indicator of urbanity.

2.4 Reflection and conclusion

Looking back on the literature, research on pedestrian mobility is broad. There are many different study areas, assumptions, and approaches used to simulate pedestrian intensities and behavior. In this overview, some of the most relevant attributes and variables have been explored. To answer this research question, however, only a subset of the existing literature has been used. What has been chosen depends on the keywords used. This means that there could be other relevant literature with other methods and variables that have not been included in this review.

For the results, some of the attributes and variables were chosen for further research. In research question 2 the following trip attributes will be explored: trip distance, amount of trips and journeys, and percentage of trips connected to home. These will be explored since they are all important for trip generation in a pedestrian model. In research question 3 the impact of the number of shops and meeting places will be explored, indirectly looking at the following variables: the attribute that people walk more in places of commerce, separation of mobility types, essential services, mixed work/living, and old buildings, museums, and monuments. These are interesting variables again for estimating trip generation but also for trip distribution.

3. Data and methods

As explained in the research aim, this thesis project proposes three different research questions each answered by using different methods. The first research question has already been answered in chapter two. The methods of research questions 2 and 3, based on the finding of research question 1, are explained in the paragraphs below.

To give insight into the methods a schematic overview is made (figure 3.1). This schema gives an overview of the data, the methods used, and how it all interacts. It starts at the input and shows per research question (marked red) the steps in the methods and finally the output of each question. For research question 3, a more detailed schema can be found in paragraph 3.3 (figure 3.5)

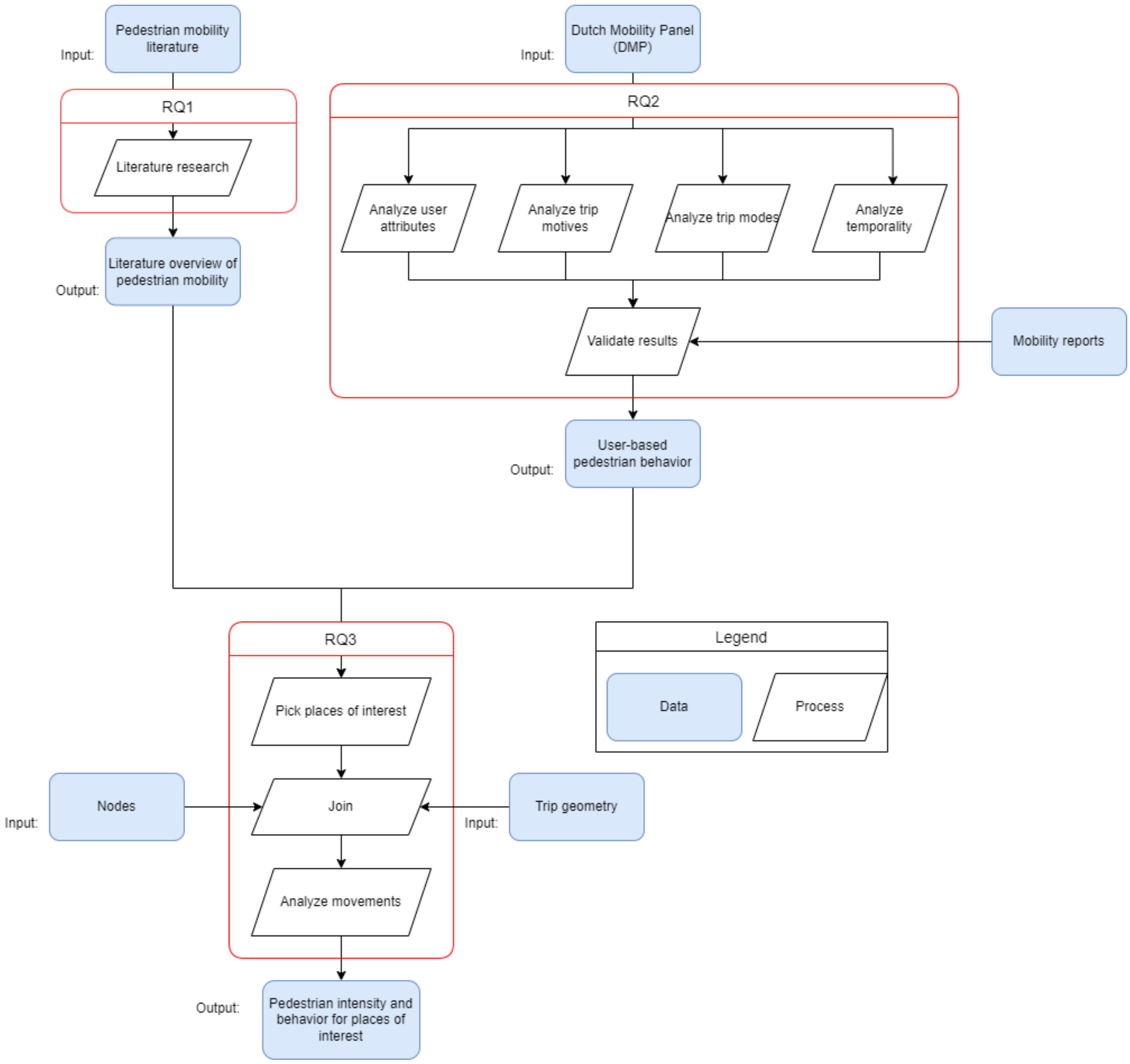


Figure 3.1: Schematic overview of the data processing workflow

3.1 Data and extent

For this study, a couple of different data sources have been used, all using the same study area. First, this study area will be described, followed by a description of the data.

3.1.1 Study area

Since the data of the DMP mainly covers the country of the Netherlands (except for some trips abroad), the main extent of this thesis will be the Netherlands. The Netherlands is a small (41.850 km^2) but densely populated country in the northwest of Europe with 17.5 million inhabitants. The large extent of an entire country has been chosen because often pedestrian models are fit on a very small scale (cities or municipalities at the largest), but analyzing the data on a higher scale may give different insights into generalizable patterns. Since it can also show regional differences (for example between high and low urban densities or between different provinces), an analysis will also be done using provinces as a spatial unit.

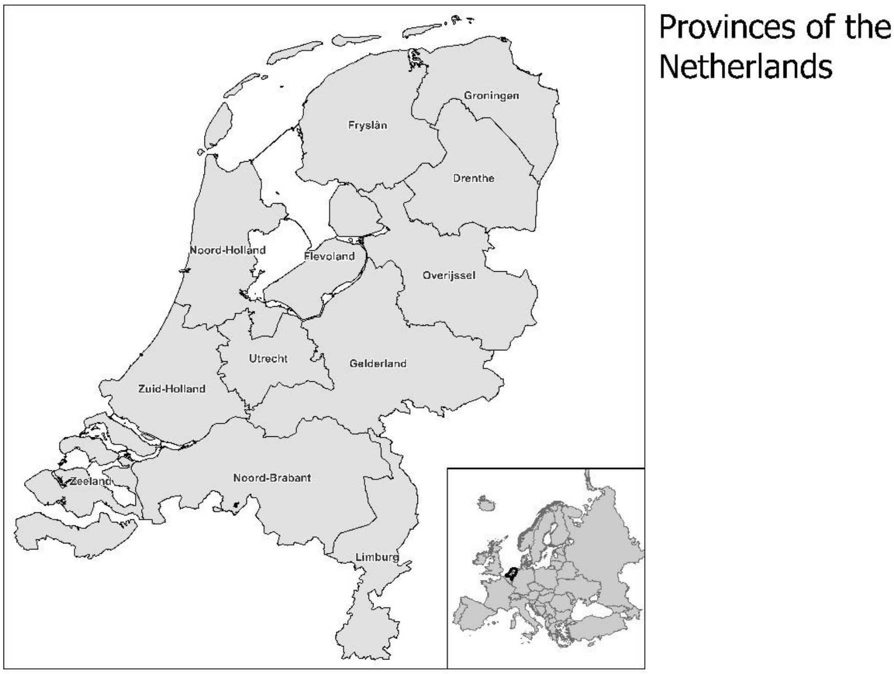


Figure 3.2: Overview of the study area with all the provinces

3.1.2 Dutch Mobility Panel

To answer the second and third research question the main source of data used is the Nederlands Verplaatsings Panel (NVP) or Dutch Mobility Panel (DMP) in English. The DMP collects trip-based movements of the members of a panel continuously, dividing the trips by the types of mobility that were used. It also collects information about the background of the participants and the motives for the trips they make. The panel consists of more than 10.000 people, spread over the Netherlands. This data is collected and stored by Dat.mobility, Kantar, and Mobidot. To conduct this research Dat.mobility has given me access to their data.

3.1.2.1 Data selection and cleaning

From the DMP a selection of the data was made. Firstly, four weeks of data were chosen to not overcomplicate the analysis. Here, dates from September 19 to October 17 2022 were chosen, since these dates are recent and contain no holidays or times of corona restrictions. This selection left records of 395.928 trips, not filtered on walking yet. Secondly, not the whole DMP was used, but a selection of tables which are described below.

From the panel, only the active users were used. Here, people who were tracked for 90 percent or more of a day are considered active. This was calculated for each of the days in the research. The trips were then filtered on only these users.

3.1.2.2 User attributes

Since the NVP consists of a lot of data and variables, it has been split into multiple tables. Not all of them will be used in this thesis. Firstly, there is a table with attributes of the users collecting the DMP data. These attributes consist of the postal code of the participant, gender, age, the urban density of the place they live, education, occupancy, if the person has a driver's license and/or a car, household type, amount of people in the household, and income. This table contains 27.558 records, which are all the users that participated in 2021 and 2022.

3.1.2.3 Trips, journeys, tours, and stays

Secondly, there is a table with all of the sensed records. This table has the following attributes: Sensing id, tracker id (the id of the participant), start and end time, duration (in minutes), tour id, journey id, and the quality of the measurements. As you can see looking at the variables there are multiple types of ids in this table. This is because there are three types of movement measured: trips, journeys, and tours (figure 3.3). These are different terms than those usually used in traffic engineering. There, a trip is called a trip-leg and a journey is a trip. For clarity, the terms stated in the DMP are used. The DMP also collects the places where people stay in one place for a moment: the stays. Each of these has its own tables as well, described below.

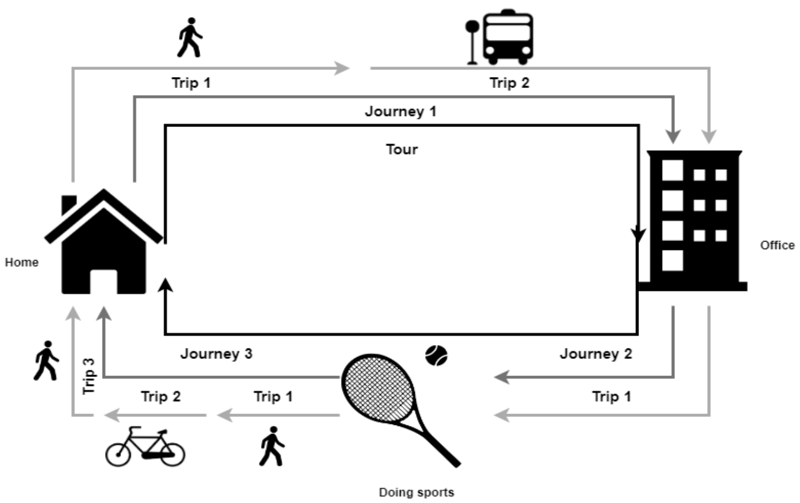


Figure 3.3: Difference between a trip, a journey, and a tour (based on Nitsche et al., 2012)

First, there is a table with sensed trips. Trips are movements using only one type of mobility. If you commute to work using a bike to get to the station, then take the train and finally have a short walk to work from the station, these are three trips. The following attributes are collected: sensing record, origin and destination, distance (in meters), duration, average speed, trip mode (type of mobility), and the purpose of the trip are saved. This table contains 395.592 records for the chosen time frame, from which 100.014 are trips by foot.

As mentioned before, there are not only trips but also journeys in the DMP. Journeys are for example the whole commute to work and thus can include multiple types of transport. The attributes of journeys consist of a journey id, start and end time, duration and distance, dominant transport mode, the purpose of the journey, origin, and destination. This table contains 298.702 records of which 62.331 have walking as the dominant transport mode.

Thirdly, the last type of movement measured is tours. These are similar to journeys but are measured from the start place back to the start place, for example from home back to home. A tour thus usually consists of multiple journeys. Here similar attributes are measured as with journeys. However, adding on to those it is also measured if the tour has an overnight stay or not. There are 138.618 registered tours in this table of which 32.404 are by foot.

Finally, the DMP does not only measure movements but also when people stay in one place, for example when people are at home or work. These records are called stays. The main rule is that if people stay in one place for more than 15 minutes the record will be classified as a stay. However, there are some exceptions to this rule where records can be classified as a stay even when the time of stay is less than 15 minutes. Examples of this are known personal places and stations. Attributes regarding stays are sensing id, start and end time, device id, duration, tour id and journey id, quality of the measurement, and location of stay. Here there are 426.204 records in the table.

DMP table	**Records**	**Records filtered on walking**	**Percentage walking of total**
User attributes	27.558	-	-
Trips	395.592	100.014	25.3
Journeys	298.702	62.331	20.9
Tours	138.618	32.404	23.4
Stays	426.204	-	-

Table 3.1: Overview of the tables used and their records

3.1.2.4 Geometries

Alongside all of these tables, there are also tables containing geographical information. The table used in this thesis contains the traces of the trips. These are the mapped routes corresponding to all of the trips. They include a trip-id and the geometry in the data format of a WKT string. An example of a trace is shown in figure 3.4 with an example of the attribute table. For privacy reasons the begin and end parts of the trace have been removed and multiple sensing records without the full WKT strings are shown.

Trace of a trip with an example of the attributes

	sensing_record_id bigint	sensed_route_wkt text
1	193132432	LINESTRING(5.79...
2	193223824	LINESTRING(4.38...
3	192694553	LINESTRING(4.43...
4	192695294	LINESTRING(4.44...
5	192172348	LINESTRING(5.47...

Legend

Trace

Figure 3.4: example of a trace with a segment of the attribute table

3.1.3 BAG nodes

Adding onto the DMP data, two more data sources were used to get the most out of the analysis. The first one is a node-based density map based on information of the Basis Registration Buildings (BAG) supplied by Dat.mobility. The map consists of coordinates for every network node in the Netherlands with attributes attached to them. These are the number of shops, number of meeting places, and number of shops and meeting places combined (*nshopmeet)* per node*.*

The *nshopmeet* will be used later in the thesis and was calculated the following way:

$$Nshopmeet = \sqrt{(NshopNmeet)}$$

The use of this dataset and the *nshopmeet* variable was chosen based on the findings in the literature review. There it became clear that in city centers and places with mixed shops and living more people walk. Since a high number of shops and meeting places often indicate a center of commerce the *nshopmeet* can be used as an indicator to validate what was found in the literature.

3.1.4 Mobility reports

The other data source are two mobility reports used for validation: one from the Centraal Bureau voor de Statistiek (CBS), Statistics Netherlands in English, and one from the Kennisinstituut voor Mobiliteitsbeleid (KiM), Knowledge Institute for Mobility Policy in English, which is part of the Rijksoverheid, both for 2021. A mobility report can be seen as a quantitative overview of the use of different types of mobility in an area often looking at change. This could be within a year but can also be a comparison of multiple years.

3.2 Research question 2: spatiotemporal patterns of pedestrians

The second research question consists of finding spatiotemporal patterns of pedestrian behavior using the Dutch Mobility Panel (DMP).

From the tables available mainly the trips were used. This is because the trips show all the movements that are done by foot and not only movements that were mainly by foot, as with the journeys and tours. Also, the user attributes were used. These were filtered on the user's tracker id matching with the tracking_ids in the trips. This meant that 6.708 users remained. Using these tables, spatial and statistical analysis of the pedestrians' movements was done using PostGIS and QGIS.

An analysis has been done based on user attributes and trip attributes. In this analysis, multiple variables were calculated based on findings in the literature. First, the number of active participants and their backgrounds (distribution of age and gender) were calculated to see how the sample of the DMP relates to the Dutch population.

Secondly, an analysis was done by looking at the sample as a whole with the entire Netherlands as the extent. Here, the average amount of trips and journeys per day was looked at together with the average amount of trips per transport mode per day and the corresponding average trip distance. Then, the average distance covered per day in general and by foot was explored, to see if this is in accordance with what the literature says.

Thirdly, a more specific geographical analysis was done as the density of shops and buildings seemed to be an important variable in the literature. Here, the differences per density of the living areas of the participants (amount of participants, frequency of trips, trips per participant, average trip distance, and average distance per walking trip) were calculated.

Lastly, a more specific look was given into individual behavior. First, the average amount of walking trips per person per weekday and the corresponding average walking distance were calculated to see if there is a temporal influence on walking. Second, the number of trips per motive and the average distance corresponding to that were looked into, to see if there is a difference between them.

The findings of this research question were then validated using mobility reports by Dutch organizations CBS and KiM. Here, the numbers and patterns found in the DMP data were compared with the reports to estimate how trustworthy the results of the DMP are.

3.3 Research question 3: Categorial analysis of pedestrian intensities

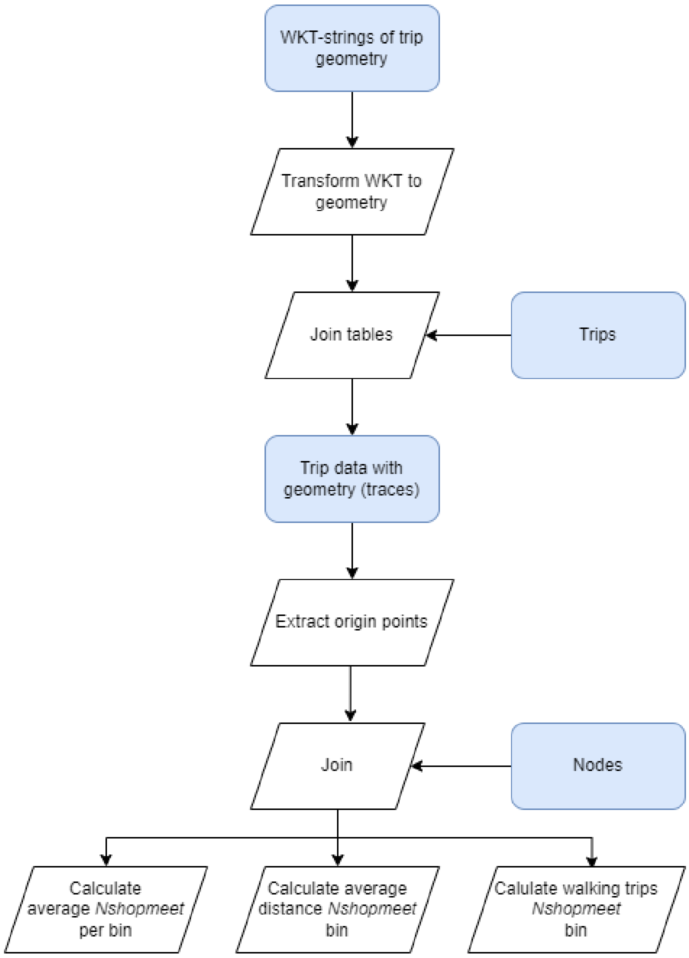


Figure 3.5: Schematic overview of the categorial analysis done in RQ3

In the third research question, a more detailed analysis has been executed regarding places of interest found in the literature (see figure 3.5). This was done by doing a categorial analysis. This means grouping places of interest with similar attributes (function, area, density of amenities) together. Since the density of shops and essential services seemed to be an important influence on walking, these variables were chosen for further analysis (De Wit, Versluis & Leferink, 2021; Goossen et al., 2021; Kang, 2017; Sevstuk et al., 2021).

For the analysis, two spatial units were used: The entire Netherlands and the Dutch provinces. These provinces were then further divided into two groups: The Randstad and the other provinces. The Randstad are the three Dutch provinces with the biggest cities. By comparing these with the Netherlands and with the other provinces the influence of big and dense cities on walking can be derived. For both these spatial units, three different types of data were used: 1) the overall numbers (trip and user attributes), 2) the traces of the trips, and 3) the nodes.

To be able to use the traces, first, the geometry of the journeys had to be made compatible with QGIS, as WKT strings cannot be read directly in the program. This was done by loading the data into Postgres and making a new table, converting the WKT strings to a geometry (see appendix 1). Then, since the geometry table only had the journey ids and the geometry but no other information, the trips table was joined to it. This was done using a full outer join to retain all of the information of all the tables. The column used to match the tables is journey-ids. Because of this join, it is possible to see the geometry of the journeys in QGIS with the corresponding journey and user information.

Then, after loading this data into QGIS the origins and destinations of each of the routes had to be found to analyze what type of places people visit. For this, only the origin points were selected. This was done with the underlying thought that when you visit a place you both get there and you leave, meaning it is both a destination and an origin. By choosing one of those, double points are thus filtered out. To do this the QGIS 'Extract specific vertices' tool was used (processing → toolbox → vector geometry → Extract specific vertices tool). Here, point 0 was chosen, which is the first point and thus the origin of the trip. This was done twice, using both the layer with all the trips and the layer with only the trips by foot to be able to compare the two.

For the final step in the data preparation, each of the origin points was joined with the closest BAG node to get the n*shopmeet* of the origin of each trip. This was done using the join attribute by the nearest tool in QGIS (processing → toolbox → vector general → join attributes by nearest).

Next, the analysis of the joint data, which consists of two parts, has been done. First, the average trip distance per area type was calculated. These area types are defined by their *nshopmeet* value. These values are defined into 5 bins: 0, > 0 and <= 0.1, > 0.1 and <= 0.5, > 0.5 and <= 2 and >2 with 0 having a very low density of shops and meeting places and more than two being very dense, thinking of a city center. The average distance was calculated by creating a flexible field with the mean of the distance and filtering the layer per bin. Secondly, looking at all the trips the percentages of walking trips for each of the five area types were calculated in relation to the total walking trips and to the total trips. The average *nshopmeet* per bin was also calculated. This was done so that the outcome of the analysis could be put into a model at the right density.

The analysis was done for the whole Netherlands and the provinces, grouped into the Randstad and other provinces. This is for two reasons. First, by looking at the entire country general patterns can be found than can be applied to a model with a large extent. Secondly, there could be regional differences in walking behavior, which could be interesting to see by looking at the provinces as well. Looking at the Randstad and the other provinces comparing it with the Netherlands was done since the Randstad both has the highest density of people and a large share of the Dutch population. This made it interesting to compare with the less dense and less inhabited areas of the other provinces to see if there is an effect of larger cities and denser areas on walking intensities.

4. Results

In this chapter, the results per research questions are explained. Before the explanation of the methods, a deeper dive was done into the literature to find the variables and attributes that can be used to estimate pedestrian behavior (RQ1). Here in the results, first, based on the variables found in RQ1, an analysis was done by looking at the trips people take and the attributes corresponding to them using the DMP (RQ2). Second, the impact of a high density of shops and meeting places on walking behavior was done, again using the DMP (RQ3).

4.1 RQ 2: Spatiotemporal patterns of pedestrians

To see if these found assumptions and variables described in research question 1 are similar in the Netherlands and thus can be used in a Dutch pedestrian model, an analysis has been done on data from the Dutch Mobility Panel (DMP). This has been done in two parts. Here, in research question 2, a more general analysis has been done to look at the attributes of the participants of the panel, which will be described below. The SQL queries that have been used can be found in Appendix II.

4.1.1 Participant analysis

First, the background of the participants was explored. As mentioned in the methods, there were a total of 6.708 participants during the time frame of this research. This does not mean that all of them were participating for the whole month. This includes anyone who participated on at least one of the days of the time frame of the study. The average of active participants per day was around 3500. Of these participants, around 46 percent were male and 54 percent female. The range of ages in the study is quite broad, ranging between 16 and 93 (figure 4.1). However, both younger and older people seem to be underrepresented comparing the DMP to the Dutch population (CBS, 2022b). While the age groups 16-20, 21-25, and 26-30 all make up around 7 to 8 percent of the Dutch population, for the first two groups in the DMP this is less than 2 percent, while for the group 26-30, this is around 3.5%. For the group over 80, the DMP also has a lower representation with less than 2 percent compared to around 5 percent in the population. In the age groups from age 36 to 75, however, there is an overrepresentation in the DMP. Here, the age groups represent between 2 and 4 percent more of the DMP sample than in the Dutch population.

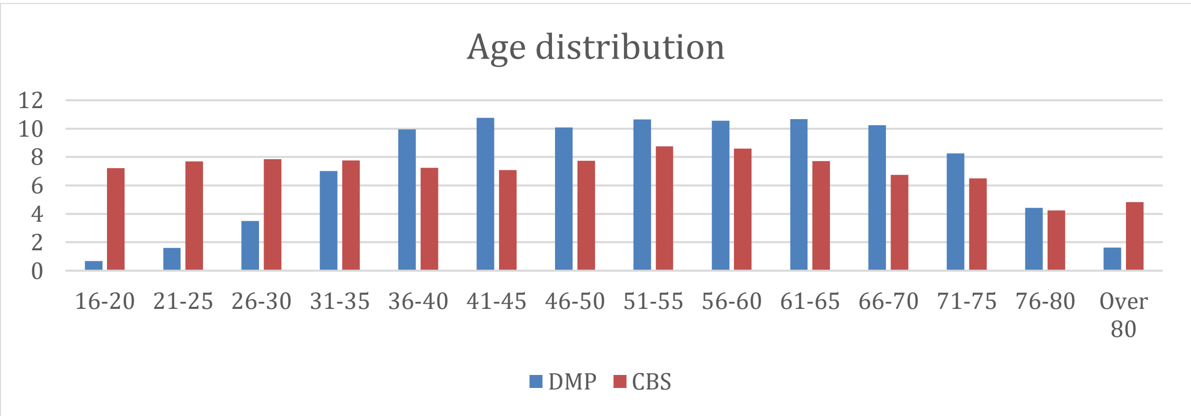
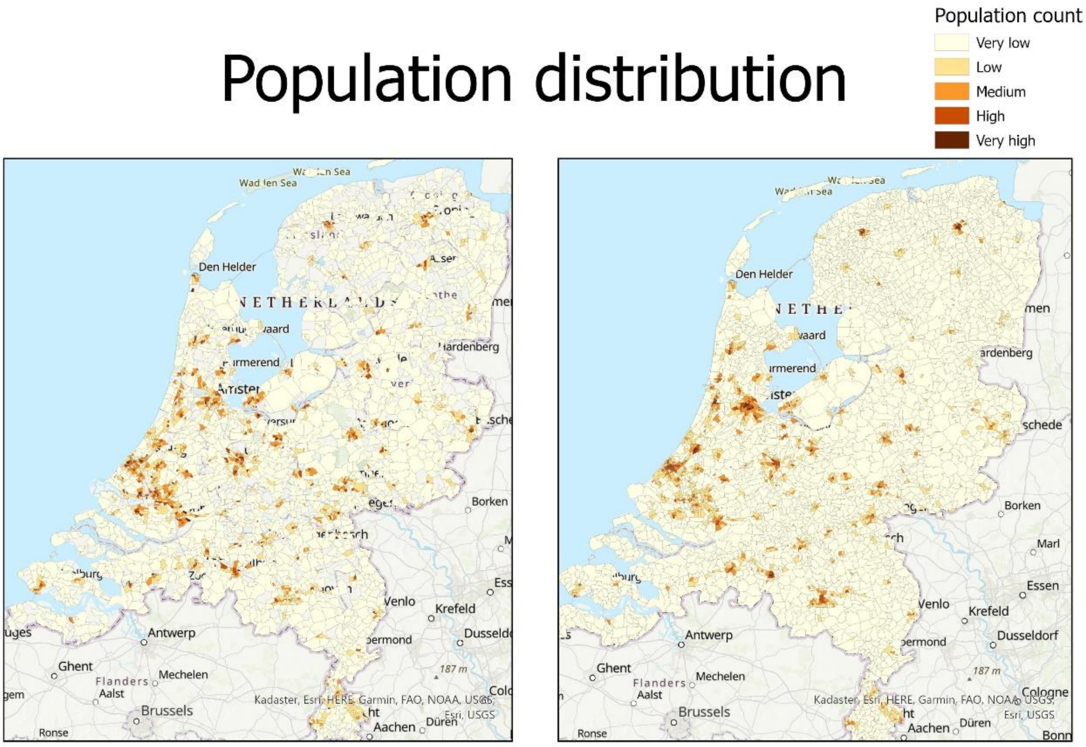


Figure 4.1: Age range of the participants and the general Dutch population

The distribution of the participants per municipality normalized by the area over the Netherlands is also similar to the distribution of the Dutch population (based on numbers of the CBS). The DMP overall shows more places with a high or very high density and also has some areas without participants (the grey areas), but the patterns are similar in both maps (figure 4.2). There is a higher density in the west and south of the country and a lower density in the north and the southwest.



I) DMP

II) CBS

Figure 4.2: The distribution of the population in the DMP and the Dutch population (CBS)

4.1.2 Trip analysis

Secondly, a closer look was taken at the trips people undertake and the underlying patterns. First, the number of trips per day and per mobility will be explored, followed by the average distance per day and per trip. Then, the influence of urban density of the living area of the participants on travelling behavior will be analyzed. After, a look into the temporality of trips will be done. Last, an analysis of the motives of the trips will be explored.

4.1.2.1 Amount of trips and journeys

Calculating the number of journeys, people make on average 3.2 journeys a day. This is a bit lower than in the assumptions for a pedestrian model by Dat.mobility (2022), where the assumption is that people undertake around 4 journeys a day. Looking at the trips, the average per day is a bit more, which makes sense since one journey can consist of more than one trip. The average here is 4.2. This means that each journey on average consists of 1.3125 trips.

Of these 4.2 trips, most of them are by car (2.1), by bike (0.92), and by foot (1.07) (table 4.1). The fact that these types of mobility are used most correlates with the expectations. Looking at the distances, it makes sense that the highest distances are by train. The fact that bicycles and walking have the lowest average distance per trip is also logical because they are non-motorized mobility types.

Mobility type	**Amount of trips per person per day**	**Average distance per trip (km)**
Bike	0.92	3.16
Boat	0.006	4.42
Bus	0.04	9.63
Car	2.1	17.63
Ferry	0.004	4.23
Foot	1.07	1.14
Light Rail	0.002	6.51
Metro	0.01	8.93
Train	0.06	30.31
Tram	0.009	4.75
Other	0.001	1.64
Unknown	0.0008	5.56

Table 4.1: Trips per day and average trip distance per mobility type

4.1.2.2 Distance

On average, people travel 43.6 km a day, of which only 1.065 kilometers are spent walking. Since the average amount of walking journeys is 1.07 a day, the average walking distance per trip is almost the same as the walking distance per day, namely 1140 meters. The minimum distance walked is 0 meters and the highest walking distance in the dataset is 42.5 km per trip.

4.1.2.3 Urban density

Another attribute that is ascribed to the participants is the population density of the area that they live in, with low being rural and high being a city. Here, most of the people live in high-density urban areas (table 4.2). This is in line with the distribution of people in the Netherlands in general. Most trips are made by people living in a high urban density, an area with a density of 1500-2500 addresses per km^2, and the least by people living in a very low urban density, an area with less than 500 addresses per km^2. This is, however, logical looking at the number of participants per density type. What is more interesting is the average trip distance per urban density type. Looking at the table (table 4.2), you can see that the trip distance decreases when the density gets higher. People living in the very low density areas travel around 12.5 kilometers per trip, while people in the very high urban density areas travel on average only three-quarters of that distance, around 9 kilometers per trip. This is most likely since in highly dense areas the amount and density of service, schools, workplaces, and other destinations are higher, meaning people have to travel less far to get to them. The walking distances also decrease, but less than the general trip distances (7.9 percent). The amount of trips per person also gets higher with a higher density. This could again be due to the shorter distances to services, making it less time-consuming and costly to visit them more often.

Urban density	**Frequency of participants**	**Frequency of trips**	**Trips per participant**	**Average distance per trip (m)**	**Average distance per walking trip (m)**
Very low	500	42.430	85	12.448	1180
Low	1.344	114.386	85	11.314	1182
Medium	1.137	105.940	93	10.004	1124
High	2.414	219.835	91	9.929	1126
Very high	1.314	128.899	98	9.042	1094

Table 4.2: Frequency of the participants and trip attributes per urban density type

4.1.2.4 Temporality

There are also some differences in the number of walking trips and average distance per trip per day of the week (figures 4.3 and 4.4). For example, on Monday people tend to make the least amount of walking trips, while on Saturday people tend to walk more often (figure 4.3). This could be since on Mondays many people work at the office, while on other days more people work from home, making it more difficult to put in a walk on Monday. For Saturday the reason is most likely that it is for most people a free day and stores and museums are open, making walking more inviting.

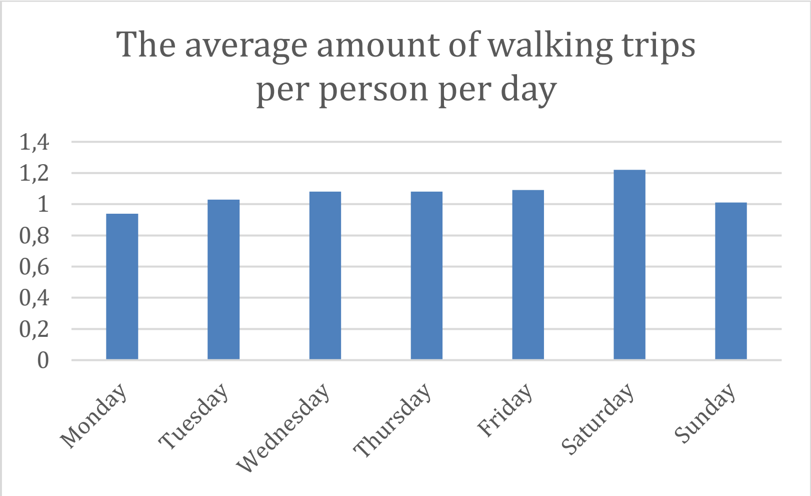


Figure 4.3: Average amount of walking trips per person per day of the week

The average distances per trip are quite similar for all days, except Sunday, where it is a bit longer (figure 4.4). This could be because people tend to walk more for leisure since it is a free day and many places are closed and trips for leisure are often longer than trips with other purposes.

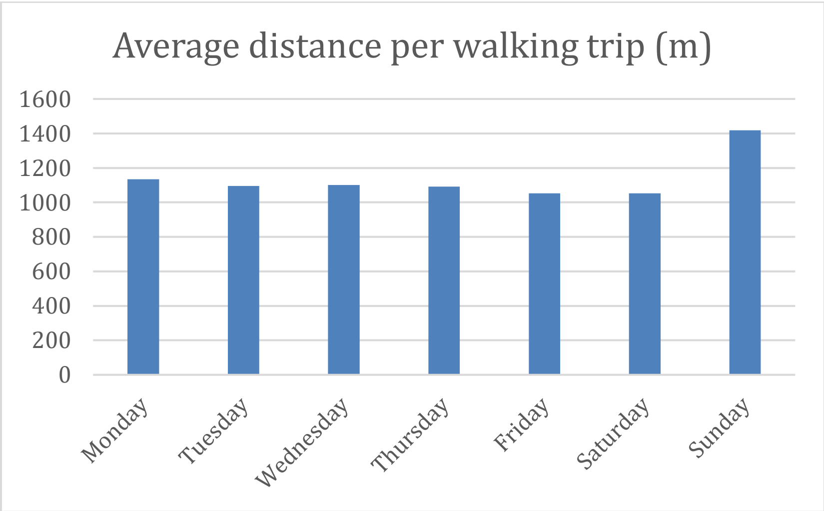


Figure 4.4: The average distance per walking trip per day

4.1.2.5 Motives

In the dataset, five motives for trips are noted, explaining the goal of the trip. Looking at the frequency of these motives and the average distance spent to walk to each of them, there are quite some interesting patterns to be seen (table 4.3). For example, the goal with the highest frequency is 'other'. This means that for most of the walking trips the goal is unknown, making it a bit more difficult to analyze why people are making these trips. The second and third most common goals are recreation/visiting and shopping, meaning people mostly walk for relaxation and socializing or to go to stores. People tend to walk the least often to school or to work, which is most likely because these locations are often too far to walk to.

Looking at the average distance spent to travel for each of the motives, they all seem to be in a similar range between 800 meters and 1.4 kilometers. The longest distances are spent on ‘other’ with around 1.3 kilometers walked on average. This, of course, doesn’t say much, since ‘other’ can contain many different types of trips. The second longest distance walked is for recreation. People walk the least when they go shopping and when they go to school. This could be because with both motives you often have to carry heavy bags, making it more constraining to walk longer distances.

Motive	**Frequency**	**Average distance (m)**
Education	4.503	859
Other	39.545	1.335
Recreation/visiting	28.020	1.099
Work/business	7.606	1.027
Shopping	20.442	899

Table 4.3: Amount of walking trips and average distance per motive

And not only the goals of the trips are noted, but also the origin and destination of the trip. What is interesting to look at is if the trips are to/from home or if they are away from home. Looking at all of the trips combined, 228.363 either have 'home' as the origin or destination and 167.565 do not. This means that almost 60% of all trips connect to the homes of the participants.

4.1.3 Validation

To check if the results found in the DMP are accurate and trustworthy a validation was done using mobility reports. This will be described below.

Looking at the traffic and mobility numbers of CBS (2022a) of 2021 there are a couple of statistics that can be compared with the findings in the DMP. They state, for example, that people travel on average 27.5 kilometers a day from which 1.4 kilometers are spent walking. This is a bit different from the numbers found in research question 2 where the distance per day is stated as 43.6 kilometers and the average walking distance per day is 1.1 kilometers. This could be explained because the CBS does not use a GPS to track the movement of people but use surveys, which could also lead to differences.

The CBS (2022a) also has numbers about the shares of mobility types used, which can also be calculated from the DMP (figure 4.5). CBS shows a lower percentage of car use and a higher percentage of bicycles and 'other' transport modes. However, these percentages found by CBS and the DMP do not differ by much (only a couple of percent), meaning that the DMP data can be seen as quite accurate compared to the CBS data.

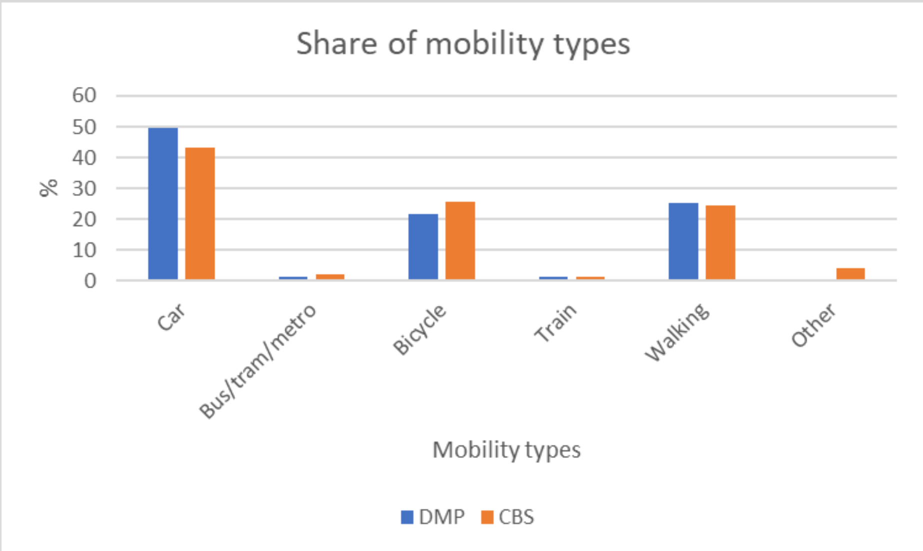


Figure 4.5: Comparison of share of mobility types between the DMP and CBS

The Kennisinstituut voor Mobiliteitsbeleid (KiM) also sets up a mobility report yearly (KiM, 2021). Looking at the version of 2021 there are also some statements made and numbers presented that can be compared with the findings in the DMP. They state, for example, that people traveled an average of 11.000 kilometers in 2019 and 7400 kilometers in 2020. This means around 30 kilometers a day in 2019 and around 20 in 2020. The fact that 2020 has such lower averages is because of the corona pandemic. Taking 2019 as a comparison with the DMP since 2020 was a corona year, the average is again much lower than found in the DMP, same as with CBS. This could have the same reasons as described by CBS. Another reason for the close average between the CBS and KiM is that KiM bases these numbers on a CBS trend model.

For average walking distance, they state that people walked around 300 kilometers in 2019 and 350 in 2020. This is 820 meters and 960 meters a day respectively. This is a bit lower than found in the DMP, which could be again due to how the KiM measures and the rounding off of the totals. Looking at the DMP, the average walking thus seems to be in between the estimation of the KiM and the CBS, making the numbers plausible.

The KiM (2021) also states that there were around 4 billion walking trips in 2020. Dividing this by the total population living in the Netherlands older than 6, as used in the KiM, this leads to 0.66 journeys a day. Since each journey consists of 1.3125 trips in the DMP you would get around 0.87 trips a day in the KiM, compared to 1.07 in the DMP (table 4.4). Why this is so much lower is unclear. It could again be because of differences in data acquisition (KiM is based on surveys, which can lead to an under registration of short trips), measurement (since KiM again based these numbers on a model), or because of the rounding off of the numbers.

	DMP	**CBS**	**KiM**
Average travel distance (km)	43.6	27.5	30
Average walking distance per person per day (km)	1.065	1.4	0.82
Walking trips per person per day	1.07	-	0.87
Walking journeys per person per day	0.83	-	0.66

Table 4.4: Comparison of the DMP, CBS, and the KiM regarding travel distance, walking distance, walking trips and walking journeys

4.1.4 Reflection and conclusion

Looking at all the results described above there are some conclusions that can be drawn. First, looking at the general analysis, there is quite a high number of participants with a wide range of personal attributes, making the data well-suited for large-scale analysis. Looking at where the participants live regarding urban density, it does not seem to have a big effect on walking behavior. While the general trip distance decreases quite a bit with an increase in urban density, the walking distances only decrease a little bit, staying relatively similar.

Looking at the number of journeys (3.2) and trips (4.2) per day, this seems a bit lower than expected looking at the literature. This could be because previous numbers about journeys and trips were often estimations, meaning that they could have been overestimated. The average distance per trip is 1140 meters which is a bit higher than the expectations according to the literature but in line with the mobility reports.

When looking at the validation, the numbers of the DMP are a bit higher than for the validation data of the CBS and KiM. This is most likely due to the fact that the CBS and KiM use surveys for their measurement. This means that people can forget to fill in short trips, while they do get included in the DMP. Taking this into account, the numbers are close enough to state that the DMP can be used as a trustworthy data source.

4.2 RQ 3: Categorial analysis of pedestrian intensities

For the second part of the DMP analysis, a categorial analysis was done. This was done by looking at pedestrian movements in areas with different densities of shops and meeting places (with *nshopmeet* = 0 being very low and *nshopmeet* >= 2 very high), as explained in paragraph 3.3 of the methods. These categories show high densities in city centers and lower densities on the outskirts of the city and rural areas (figure 4.6). The urban areas in the map of the province of Utrecht on the right match the red and orange part of the *nshopmeet* on the left quite well. The outcomes are explained below for the Netherlands, the provinces of the Randstad, and the other provinces. First, the Netherlands will be compared with the Randstad and the other provinces. Second, a more specific insight is given into the individual provinces by looking separately at the Randstad and the other provinces. A more detailed analysis for each of the provinces individually can be found in Appendix III.

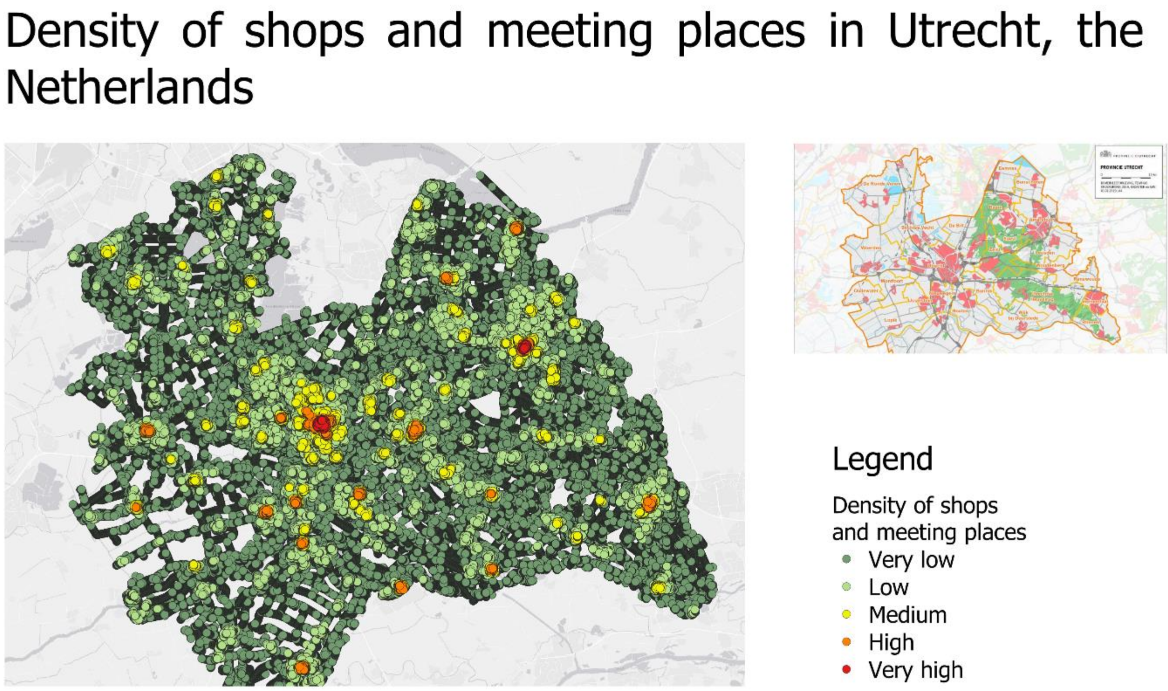


Figure 4.6: the density of shops and meeting places in Utrecht

4.2.1 The Netherlands

4.2.1.1 Share of walking trips

Looking at the share of walking trips compared to other transport modes there is a clear increase when the density of shops and meeting places increases (figure 4.7). For the Netherlands as a whole, the share seems around 20 percent when the fraction of shops and meeting places is 0 and around 60 percent when the fraction is around 3.5. Both the Randstad and the other provinces show a similar pattern.

Some regional differences can also be seen in figure 4.8. In the Randstad (the three provinces with the biggest cities of the Netherlands) the fraction of shopping and meeting places goes up higher than average, with the highest fraction being almost 4, compared to the 3.2 average, while for the other provinces, this is less than 3. The share of walking trips, however, is for all three spatial units the same.

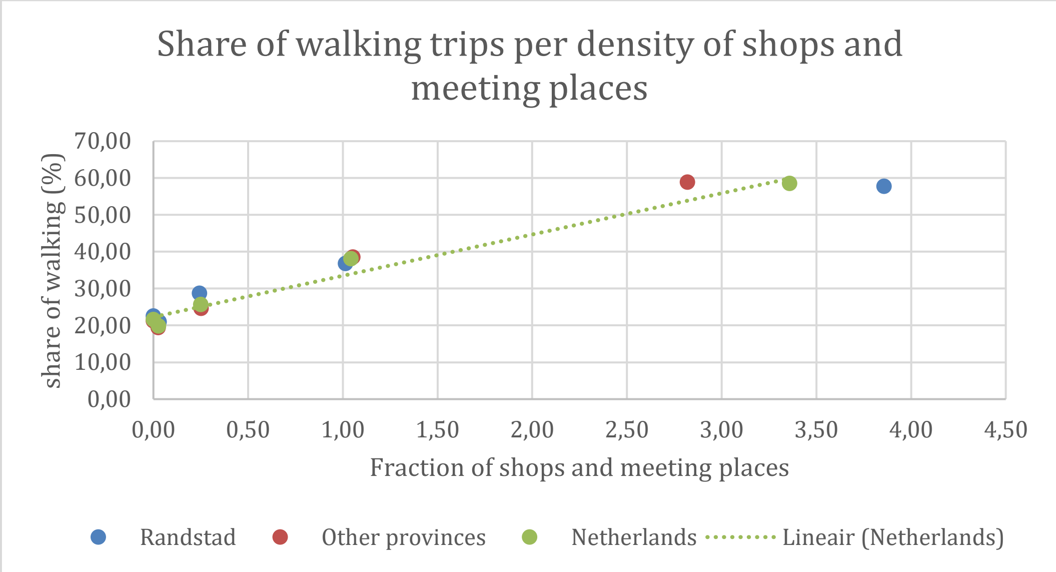


Figure 4.7: share of walking trips per density of shops and meeting places

4.2.1.2 Average walking distance

Looking at the average walking distance for the Netherlands as a whole, the Randstad and the other provinces per density type seem quite similar at a first glance (figure 4.8). The walking distances are the highest in the very low density and then the low density. This could be because there are likely more leisure trips in the less dense areas. For the medium, high, and very high density, there is quite some overlap and there does not seem to be a clear trend. The average distance for the Randstad does seem to be a little bit higher than for the other provinces and the Netherlands as a whole. However, the numbers are close together.

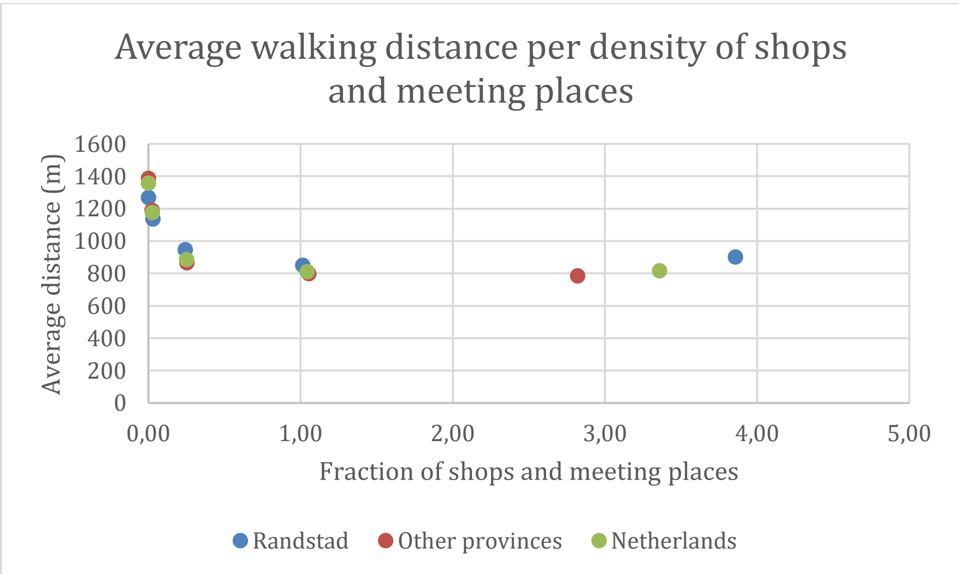


Figure 4.8: average walking distance per density of shops and meeting places

4.2.2 The Randstad

The Randstad, consisting of the three provinces (Utrecht, Noord-Holland & Zuid-Holland) with the biggest cities in the Netherlands, shows some interesting patterns regarding the average walking distance and share of walking. These patterns are explained below.

4.2.2.1 Share of walking

As seen in paragraph 4.2.1.1, the Randstad as a whole has a slightly higher share of walking compared to other modes than the Netherlands as a whole in the lower fractions. Looking at the comparison of the individual provinces with the Netherlands as a whole, the patterns are all similar: a low share at a low fraction and a higher share at a higher fraction (figure 4.9). However, there are still a couple of differences. First, while at a fraction of 0 the share of walking is very similar for the Randstad and the Netherlands, the Randstad does have a higher share than the Netherlands in the other low fractions of shops and meeting places (between 0.1 and 1.1). This difference, however, is only a couple of percent.

In the higher fractions of shops and meeting places (more than 3) the share of walking for the Randstad is lower than for the Netherlands. This could be due to the fact that the larger cities have more travel options in the dense areas (metro, tram) than other cities in the rest of the Netherlands.

There are also some differences between the provinces. For example, Utrecht shows a higher fraction of shops and meeting places than the other provinces. While for Noord-Holland and Zuid-Holland the highest fraction is around 3.2, this fraction measures around 4.9 for Utrecht. This means that Utrecht has a higher density of shops and meeting places in the densest areas. For this density, they also show a higher share of walking (65%) compared to the other provinces (around 50%). For the other densities, the share of walking seems to be similar for all the provinces.

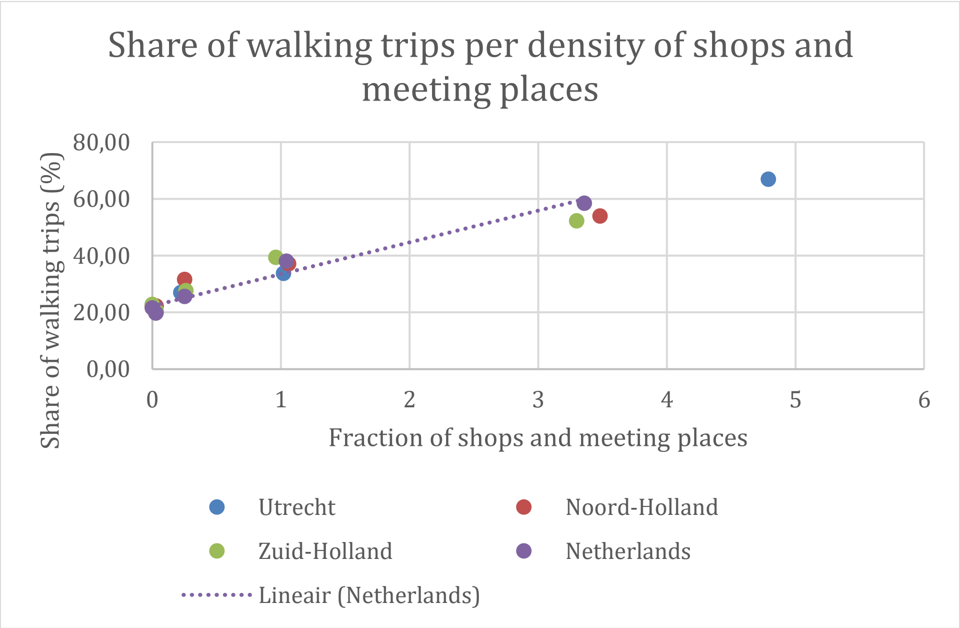


Figure 4.9: share of walking trips per density of shops and meeting places

4.2.2.2 Average walking distance

Looking at the average walking distance per density of shops and meeting places each of the Randstad provinces also show a similar pattern to the Netherlands, namely a higher average at the lowest density and then a constant average for the rest of the density types (figure 4.10).

There are also some differences to note. For example, as mentioned before, the highest fraction of shops and meeting places in Utrecht (4.9) is a lot higher than the Dutch average (3.2). Another difference is the average walking distances in the lower densities. Looking at a density of zero the averages of the provinces are all higher than for the other densities. This could be due to round trips, which are trips for leisure and are often longer than trips with a goal. This could also be due to the longer distances necessary to reach places.

What is interesting in these lower-density areas is that for all of the provinces, the average walking distance is lower than that of the Netherlands as a whole. A reason for this could be a more diverse and constant offer of public transport options compared to other parts of the country. It could also be that since it is a densely populated area people do not like to walk for leisure as much as in other areas.

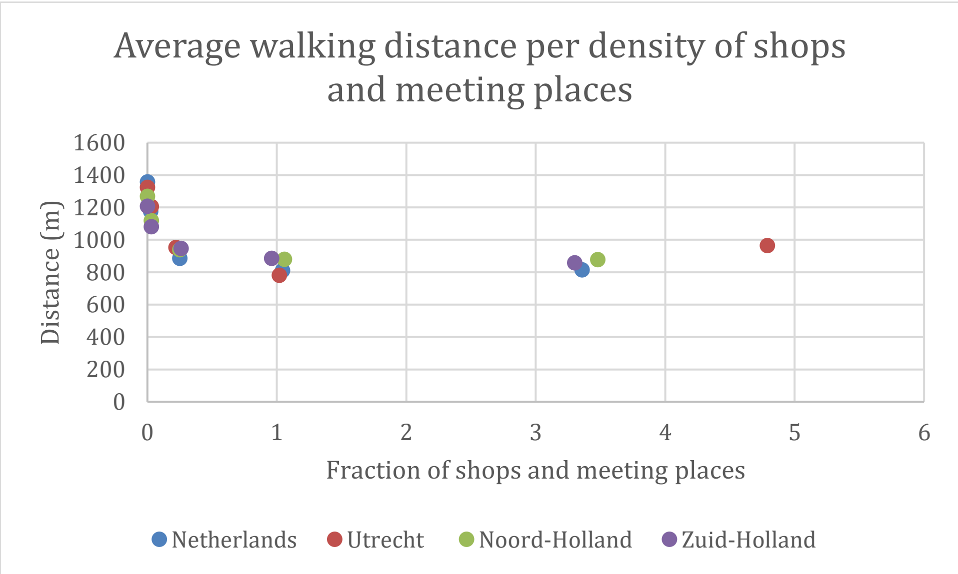


Figure 4.10: average walking distance per density of shops and meeting places

4.2.3 Other provinces

There are nine other Dutch provinces that are not part of the Randstad. These are all provinces without very large cities, most of them having a only a couple of medium size and more smaller cities. The patterns found here will be explored in the paragraphs below.

4.2.3.1 Share of walking

There is quite some diversity in the share of walking between each of the provinces (figure 4.11). At a fraction of shops and meeting places of 0, the provinces show a similar share, ranging between 15 and 25 percent, which is in line with the Netherlands as a whole. However, with an increase in this fraction, the share gets more diverse. For example, with a fraction close to 1 the share of walking in Groningen is low being less than 30 percent, while the other provinces range between 35 and 47 percent with the Dutch average being around 40 percent.

At the higher densities, the shares are even more spread out. Drenthe shows a share of over 65 percent as do Gelderland and Noord-Brabant. There are also a couple of provinces that do not have such a large share of walking even in the higher densities. These are Friesland and Zeeland, with the share of walking in Friesland going up to 50 percent and in Zeeland lower at 40%, which is well below the Dutch average of 60%. Overall, there is quite some diversity between the different provinces.

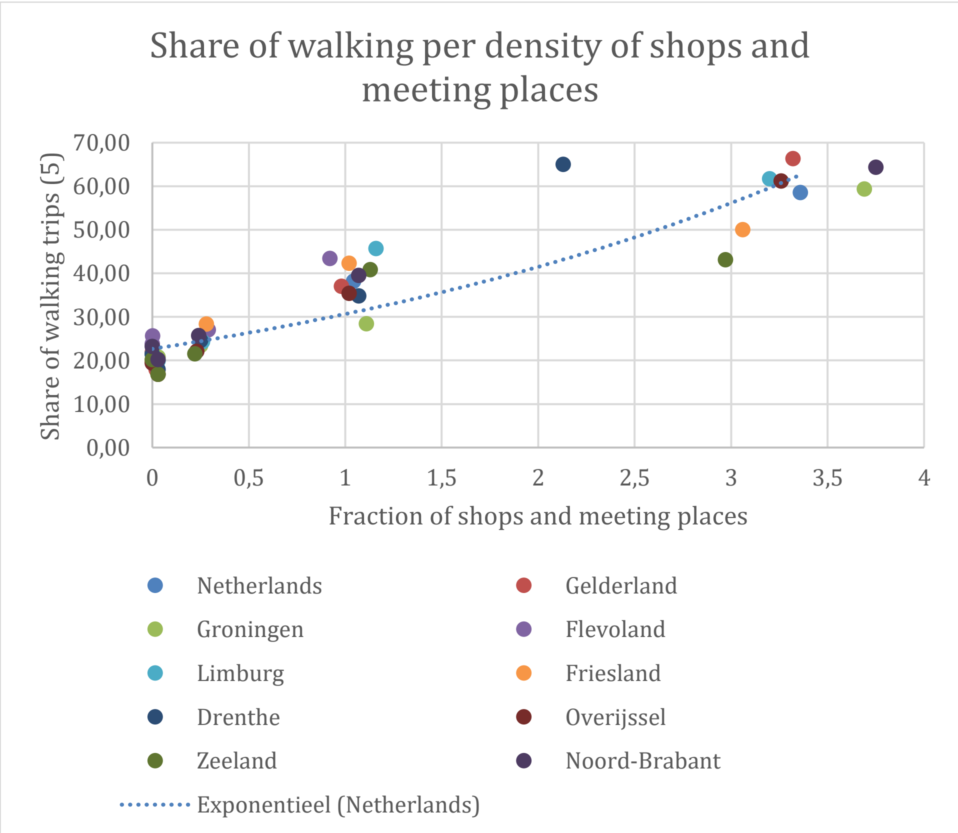


Figure 4.11: share of walking trips per density of shops and meeting places

4.2.3.2 Average walking distance

Looking at the average walking distances for the density of shops and meeting places for the provinces in figure 4.12 there are still some visible differences between the provinces, but less than with the share of walking (figure 4.11). The average walking distance shows the same patterns as the Randstad with higher averages at a fraction of 0 and a constant average at higher fractions. The largest differences are at the lowest fraction with distances ranging from 100 meters (Flevoland) to almost 1600 meters (Zeeland), which is a bit more diverse than in the Randstad (1100-1400 meters). The fact that these other provinces have some higher averages could be due to more leisure walks or because of a less diverse offer of public transport compared to the Randstad.

For the other, higher densities the average distance remains quite constant ranging between 600 and 1000 meters. This is a bit more spread out than for the Randstad where it was between 800 and 1000. Since the numbers are quite constant it seems like the density of shops and meeting places does not have a large influence on walking distances.

The range of the highest density fraction is also diverse. In Drenthe, the highest fraction is only 2.2 while in Noord-Brabant this is around 3.7. The Dutch average here is around 3.4 and only two provinces have a higher ‘maximal’ fraction here: Noord-Brabant as mentioned before and Groningen. The rest are lower meaning that they have less dense shopping areas than on average in the Netherlands.

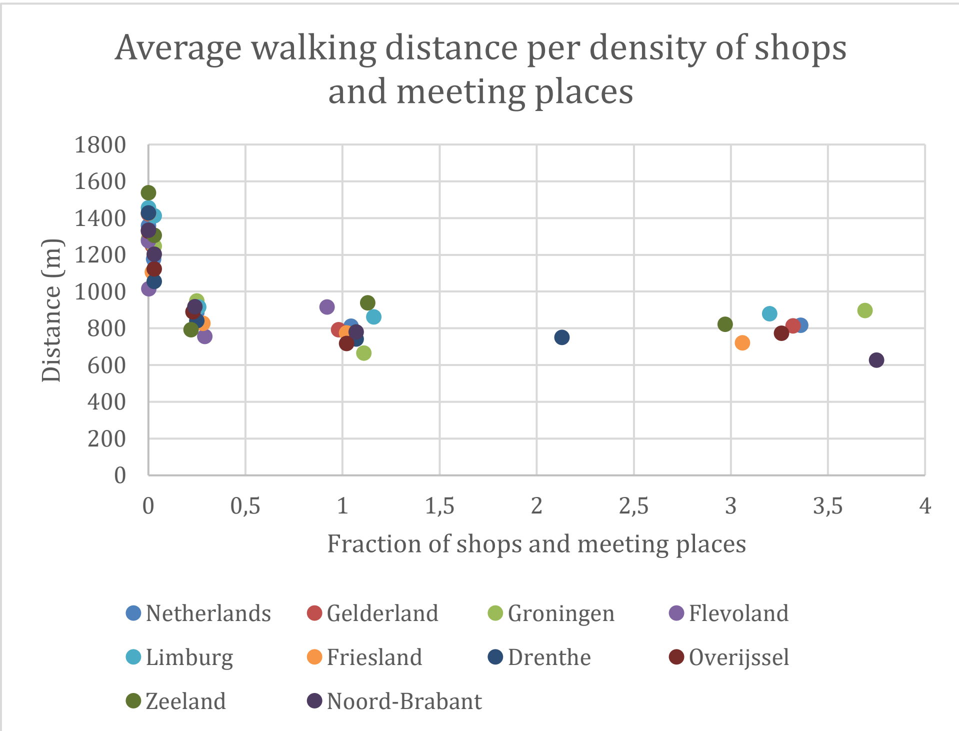


Figure 4.12: average walking distance per density of shops and meeting places

4.3.2 Reflection and conclusion

With the category analysis, there are some interesting patterns to be seen. First, the average walking distance is high when the fraction of shops and meeting places is 0 and then decreases and stays constant at higher fractions. In the Randstad, this constant is on average a bit higher than for the other provinces (800-1000 meters compared to 600-1000 meters), while in the other provinces, the distances are on average a bit higher at a fraction of 0 (1000-1600 meters compared to 1000-1400 meters). The distances, however, are a bit more spread out for the other provinces than for the Randstad. This is likely because quite a number of provinces are included in other provinces (9) while the Randstad only counts 3.

And second, when the density of shops and meeting places increases, the share of walking trips compared to other modes of transport also increases, likely because traveling by car and bike is more difficult and sometimes forbidden in dense areas. The lowest-density areas have on average a share of close to 20 percent, while in the highest density, this is around 60 percent. These shares are on average a bit lower in the Randstad than in the other provinces. The other provinces do, however, show some more diversity in their shares of walking ranging between 40 and 70 percent in the highest densities. But, the numbers of the Randstad and the other provinces are very similar here, suggesting that the presence of large cities does not impact the share of walking.

5. Discussion

The goal of this thesis was to analyze pedestrian intensities on a national scale using both literature and GPS mobile phone data, which can then be used as a basis for a pedestrian traffic model. This was done using three research questions: 1) What attributes and variables can be used to estimate pedestrian mobility?, 2) What patterns of pedestrian intensities regarding selected user attributes can be found in the Dutch Mobility Panel? And, 3) What patterns of pedestrian intensities regarding geographical attributes can be found in the Dutch Mobility Panel?

There are, however, still a couple of things to be discussed regarding the result. This chapter, therefore, starts with a summary of the results given per research question, followed by a comparison with the literature described in research question one. Finally, the limitations of this research will be given.

5.1 Discussion RQ1: existing pedestrian movement knowledge

In the first research question, the attributes and variables concerning pedestrian modeling were collected using a literature review. The expectation was that since it is a broad topic, many different results would be found. Looking at the 12 documents used in the literature review, this seems true. There are many different types of pedestrian mobility research with different extents, types of data collection, and analysis methods. From the literature used in this study, 13 attributes were found about how pedestrians behave regarding the use of infrastructure, flows, trip distance, and trip attributes. A further 41 variables that influence walking were described in 9 different categories. Looking at other literature overviews on pedestrian mobility, Basu et al. (2022) mention 102 factors influencing pedestrian route choice, and Goossen et al. (2021) 103 regarding walkability. This is more than stated in this thesis but matches the fact that there are many components that can be included in pedestrian research.

Overall, it is easy to conclude that pedestrian mobility research is a broad field. Many aspects can be either included or excluded when looking at researching pedestrian intensities depending on the purpose of the research and the data available.

5.2 Discussion RQ2: spatiotemporal patterns of pedestrians

From the Dutch Mobility Panel GPS data, a lot of interesting data could be derived, not only about pedestrians but also for other mobility types and even mobility as a whole. The expectation was that the results found here would match the expectations set in the literature, but could also show regional differences due to the large size of the data set and the extent of an entire country. An example of a result is the average walking distance per trip. In the DMP an average of 1140 meters was found. The literature states that most trips are between 300 and 1000 meters, meaning that the average found is a bit higher than expected. Dat.mobility (2022) also took 650 meters as the average, which is again lower than found here. The fact that the DMP has higher averages than expected could be because the DMP includes all walking trips, while pedestrian research often excludes trips for leisure. The DMP also differs a bit from what was found in the mobility reports of the KiM and the CBS. The CBS showed a higher average walking distance per day (1.4) and KiM a lower one (0.82). This could be due to differences in measurement, with the KiM and CBS using survey data and the DMP using GPS data.

Furthermore, the average journeys and trips per person per day were found. The expected amount was 4 journeys (Dat.Mobility, 2022). However, in the data used a total of 3.2 journeys a day were found. This is lower than expected. This could have multiple reasons. First, the 4 journeys mentioned in the literature were an estimation, it could just be that this was too high. Another reason could be that September was quite a rainy and cold month compared to earlier years (Homan, 2022). This could have impacted how often people were willing to go outside. The KiM on the other hand shows a lower average amount of trips and tours than the DMP. This can be because the KiM is based on surveys, where there often is an under-registration. Smaller trips and journeys are often forgotten when filling in these surveys, leading to a lower average amount of trips and journeys.

Another result was the difference per density of the living places of the participants. Here, the most important patterns were that the denser the living area, the lower the distance per trip (table 4.2). Relating to this is the trip distance found in the literature. Here, variables found in research question 1 state the longer the distances, the fewer people will walk (Kang, 2017). Again looking at the density of where people live, it is likely that in the less dense areas, essential services and other places are further away, thus making trip distances longer. This was also seen in the literature. The table (table 4.2) also shows that the denser the areas get, the more trips per person are taken. This pattern is also one of the first things stated in the assumptions; that centers of commerce are expected to have a high intensity of pedestrians (Barbosa et al., 2018; Yamu et al., 2021).

This also relates to two other variables found in the literature: separation of mobility types and reachability. According to the literature, the first one has a positive impact on walking and the second one has a negative effect. Looking at city centers, where often mobility types are quite separated (especially walking) and therefore more difficult to reach by different types of mobilities, the share of walking is a lot higher than in other areas. The stated influence of the variables in literature can therefore be seen as true.

Lastly, another interesting find is that around 60% of all trips are connected to people's homes. Dat.mobility (2022) also estimated this percentage to be around 60% of all trips. This finding thus matches expectations.

Thus, while there are some differences between the results found in the DMP and the expectations found in the literature, these differences are mostly very low or explainable. This means that the results found in the second research question can be seen as believable. Therefore, the DMP (and other GPS sources) can be seen as a trustworthy data source for pedestrian research.

5.3 Discussion RQ3: drivers of pedestrian movement

Looking at de DMP data combined with the density of shops and meeting places divided into five classes some patterns were found that can be useful when making a pedestrian traffic model. The first is that when the density of shops and meeting places increases, the average walking distance decreases, partially due to leisure trips in the less dense areas. The second one is that the denser, the larger the share of walking trips compared to other modes of transport. This matches the expectations of where people tend to walk and how much they walk.

The average distances per area type found for each of the provinces also are higher than stated in the literature. This could be because walking for leisure is often not included in pedestrian research. The higher averages in the less dense areas could then be explained by people walking more for leisure in these areas. To see if this was the case, a look was taken at round trips, which are trips where people start walking at one place and then return to the same place. This indicates that there was no clear goal to the walk and thus is a walk for leisure. Looking at these round trips, per density types (figure 5.1), people do seem to walk more for fun in areas with fewer shops and meeting places.

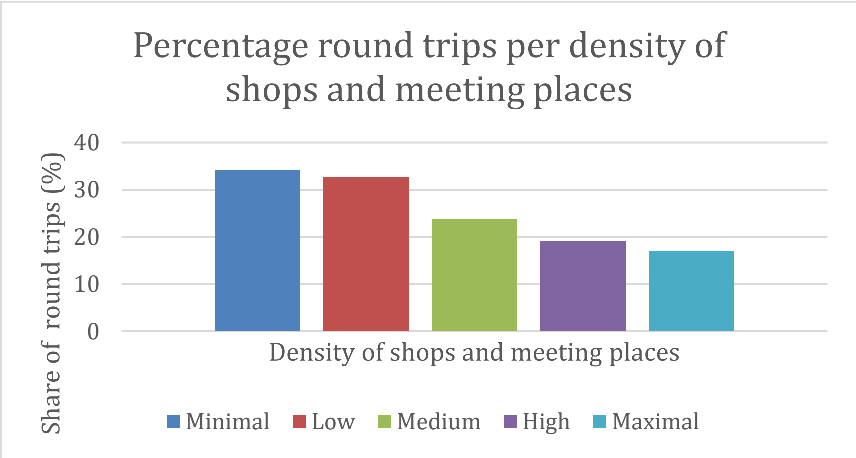


Figure 5.1: Share of tours per density of shops and meeting places

When looking at the average distance per density of shops and meeting places for both the round trips and the non-round trips (figure 5.2), it shows that the average distance is higher for the tours. This line also has a steeper decline when the density goes up, while the average distance for non-tours stays more similar. The average distances found for non-tours are, however, still a bit higher than stated in the literature. Bongiorno et al. (2019) and Kang (2017) state for Boston and Seoul respectively that trips usually are between 300 and 1000 meters, while the DMP shows averages between 800 and 1100 meters. Thus, the presence of tours in the DMP data can partly explain the higher average, but not completely. Both the KiM and the CBS also do include round trips in their reports, therefore this also does not explain the differences found in their average distances and those of the DMP.

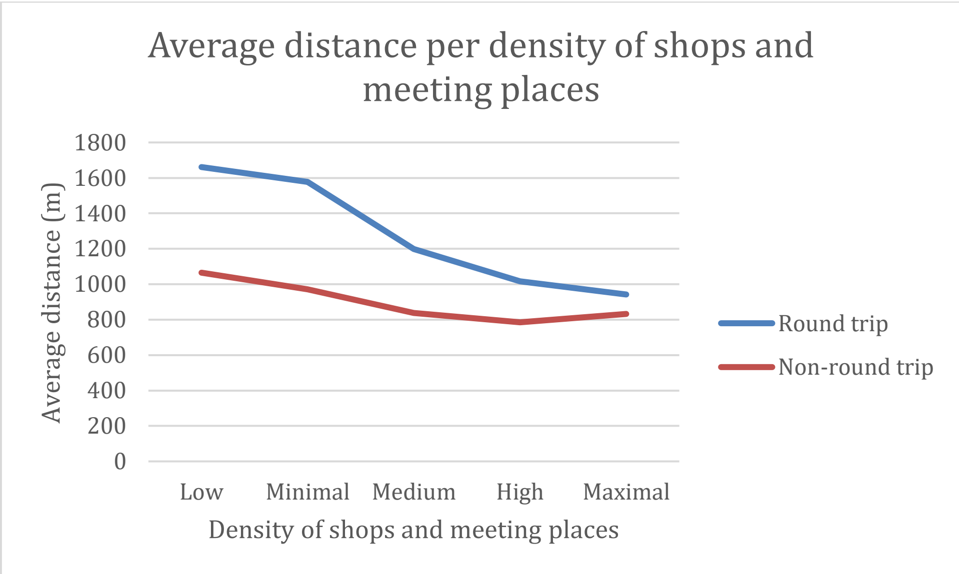


Figure 5.2: Average walking distance per density of shops and meeting places

Lastly, one of the first things stated in the assumptions is that centers of commerce are expected to have a high intensity of pedestrians. Looking at the category analysis this assumption seems to hold up. In all of the provinces, the areas with the highest intensity of shops and meeting places had the highest shares of walking (between 50 and 70 percent) compared to other modes of transport.

5.4 Limitations

While, as seen in the discussion above, the results of this thesis seem very promising, there are also some limitations regarding the research. Two of the most important ones will be described below.

5.4.1 Comparability

First is the comparability of the literature with the panel data. In the first research question literature was used to give an overview of assumptions, variables, and methods used in pedestrian mobility research. This then was used again in the third research question to see how well the literature compared to the DMP dataset used. However, there is a difference in context and methods used between the literature and the DMP. While most literature uses the extent of a neighborhood of a city, the DMP looks at a whole country. And while the DMP uses GPS mobile phone data, other research uses all kinds of methods. This makes it difficult to compare. However, even though these differences exist, the basic principles of pedestrian behavior should be the same no matter the context. This means that it is possible to draw a comparison, the differences in context should, however, be kept in mind when concluding.

5.4.2 Validation

Secondly, there is the validation of the data. In the second research question, the data found in the DMP is validated using mobility reports from the CBS and KiM. These instances collect their data through surveys while the DMP uses mobile phone GPS data. The CBS and KiM also report mobility numbers over a whole year, while in this research only a month of data was used. However, the extent of the studies is the same; the whole Netherlands. Therefore, although there are some differences in methods, the average numbers should be similar, meaning that the mobility reports can be used as a method for validation.

Thus, while there are a couple of things to keep in mind while looking at the data, these limitations do not severely impact the quality of the research. This means that this thesis can add well to the existing pedestrian mobility research, adding new insights on looking at patterns using a large extent and insights on using GPS data.

6. Conclusion and recommendations

The objective of this research was to analyze pedestrian intensities using literature and GPS phone data using a national extent, which can then be used as a basis for a pedestrian traffic model.

6.1 Conclusion

This research shows that although there are many different attributes and variables that can be used to estimate pedestrian intensities and behavior, GPS data can be used to analyze many of these. GPS data can give interesting insights into many aspects of pedestrian behavior. Firstly, it can show interesting patterns regarding the user attributes of the participants, for example, age distribution, gender, and living area. People, for example, travel more often when living in a dense area, but the trips are shorter than for people living in less dense areas. Secondly, it can show insight into why and when people make trips. The data shows that people tend to travel more and longer on weekends and that most walking trips are for leisure. Lastly, when combining GPS with other data sets it can uncover all kinds of interesting patterns. In this thesis, it was shown that when there are more shops and meeting places, people tend to walk shorter distances, but the share of walking trips compared to other modes of transport gets higher.

Looking at a broader perspective, in the literature most research on pedestrian intensities and modeling is based on survey data and counts. The research is also usually done using a small extent, usually a city or a neighborhood. This research shows that mobile phone GPS data with a national extent can be an insightful addition to these researches. GPS can show detailed insights into pedestrian behavior which is not possible using other methods while also being able to show this on higher levels, for example, in provinces of a country. This means that models also can become more accurate and more generalizable.

This research thus shows that research on pedestrian mobility and pedestrian traffic models do not have to be done on the scale of a city or a neighborhood, but that data sets on a national scale can also give important insights. It also shows that GPS data is an interesting type of data to use and can give all kinds of specific insights.

6.2 Recommendations

While this thesis can be used as a good base for pedestrian modeling, there still are some topics regarding pedestrian intensities that would be interesting to explore in future research. A couple of them will be highlighted below.

A first recommendation is to look further into how different user attributes influence walking behavior. In this thesis the participants were looked at as a whole, not looking at different population groups. Only spatial and trip characteristics were taken into account, for example, the density of the living area and the goal of the trip but not personal attributes. However, the Dutch Mobility Panel also registers other more personal user attributes that could be interesting to look at. These include age, gender, education, family type, and income. It could also be interesting to do a combined analysis of multiple of these attributes.

Secondly, the temporal aspect of walking could be further explored. Here only a small analysis was done by looking at the difference in the number of trips and average distances per day. However, it would also be interesting to look at the differences per part of the day or even per hour, since then possible temporal patterns could be exposed which were also mentioned in the literature.

Another interesting follow-up could be to change the bins for the density of the shops and meeting places. For this research, five relatively large bins were created. However, by making more bins or by changing the bounds of the bins the results of the category analysis could differ. Therefore, this could also be useful to look at in other research.

7. References

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

Basarić, V., Vujičić, A., Simić, J. M., Bogdanović, V., & Saulić, N. (2016). Gender and age differences in the travel behavior–a Novi Sad case study. *Transportation research procedia*, *14*, 4324-4333.

Basu, N., Haque, M. M., King, M., Kamruzzaman, M., & Oviedo-Trespalacios, O. (2022). A systematic review of the factors associated with pedestrian route choice. *Transport reviews*, *42*(5), 672-694.

Bazzan, A. L., & Klügl, F. (2014). A review on agent-based technology for traffic and transportation. *The Knowledge Engineering Review*, 29(3), 375-403.

Bongiorno, C., Santucci, D., Kon, F., Santi, P., & Ratti, C. (2019). Comparing bicycling and pedestrian mobility: Patterns of non-motorized human mobility in Greater Boston. *Journal of transport geography*, *80*, 102501.

Bryman, A. (2016). *Social research methods*. Oxford university press.

CBS (2022a). Hoeveel reisden inwoners van Nederland en hoe? *Centraal Bureau voor de Statistiek.* Accessed: 07-02-2023: https://www.cbs.nl/nl-nl/visualisaties/verkeer-en-vervoer/personen/mobiliteit

CBS (2022b). Leeftijdsverdeling Hoe ziet de leeftijdsopbouw van de Nederlandse bevolking eruit? *Centraal Bureau voor de Statistiek.* Accessed: 06-03-2023: https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/leeftijd/bevolking

Dat.Mobility (2022). Uitgangspunten basismodel. *Dat.Mobility.*

De Wit, A., Versluis, L. & Leferink, T. (2021). Van eerste stappen tot sprint: de ontwikkeling van een nieuw soort voetgangersmodel, de Rotterdamse LoopMonitor. *Colloqium Vervoersplanologisch Speurwerk.*

Gaber, N., Ibrahim, A., Rashad, A. B., Wahba, E., El-Sayad, Z., & Bakr, A. F. (2020). Improving pedestrian micro-climate in urban canyons: City Center of Alexandria, Egypt. *Urban Climate*, *34*, 100670.

Geurs, K. T., Thomas, T., Bijlsma, M., & Douhou, S. (2015). Automatic trip and mode detection with move smarter: First results from the dutch mobile mobility panel. *Transportation research procedia*, *11*, 247-262.

Gómez, L. F., Parra, D. C., Buchner, D., Brownson, R. C., Sarmiento, O. L., Pinzón, J. D., ... & Lobelo, F. (2010). Built environment attributes and walking patterns among the elderly population in Bogotá. *American journal of preventive medicine,* 38(6), 592-599.

Goossen, C. M., Rip, F., Staritsky, I., & Thomas, D. (2021). Beloopbaarheid gemeten: pilotstudie naar de mate van beloopbaarheid van ov-knooppunten en hun buurten in de provincie Utrecht.

Homan, C. (2022, October 3). September 2022. *KNMI.* Accessed: 7 February, 2023. https://www.knmi.nl/nederland-nu/klimatologie/maand-en-seizoensoverzichten/2022/september

Hsieh, Y. K., & Li, C. N. (2018). Research on the evaluation indexes of walking-friendly environment in healthy communities from the perspective of mass transit-oriented development-Taipei MRT as an example.

Hunter, R. F., Garcia, L., de Sa, T. H., Zapata-Diomedi, B., Millett, C., Woodcock, J., ... & Moro, E. (2021). Effect of COVID-19 response policies on walking behavior in US cities. *Nature communications,* 12(1), 1-9.

Jović, J., & Depolo, V. (2011). The role of trip generation models in sustainable transportation planning in South-East Europe. *Transport*, *26*(1), 88-95.

Kang, C. D. (2017). Measuring the effects of street network configurations on walking in Seoul, Korea. *Cities*, *71*, 30-40.

KiM (2021, November 18). Mobiliteitsbeeld 2021. *Kennisinstituut voor Mobiliteitsbeleid.* Accessed 08-02-2023): https://open.overheid.nl/repository/ronl-ee43316f-08dc-41a4-80a7-06c4d64af4cf/1/pdf/mobiliteitsbeeld-2021.pdf

Kuddus, M. A., Tynan, E., & McBryde, E. (2020). Urbanization: a problem for the rich and the poor? *Public health reviews*, *41*(1), 1-4.

Lall, S. V., & Wahba, S. N. (2021). Crowded Cities: New Methodology in COVID-19 Risk Assessment. *Sustainability*, *13*(13), 7167.

Lee, K., & Sener, I. N. (2020). Emerging data for pedestrian and bicycle monitoring: Sources and applications*. Transportation research interdisciplinary perspectives*, 4, 100095.

Lo, R. H. (2009). Walkability: what is it? *Journal of Urbanism*, 2(2), 145-166.

Malleson, N., Vanky, A., Hashemian, B., Santi, P., Verma, S. K., Courtney, T. K., & Ratti, C. (2018). The characteristics of asymmetric pedestrian behavior: A preliminary study using passive smartphone location data. *Transactions in GIS*, *22*(2), 616-634.

Monster, J. (2022, November 1). PBL: auto blijft cruciaal voor mobiliteit, vooral in stadsranden en op platteland. Accessed: December 8 2022. *Gebiedsontwikkeling.nu.*

https://www.gebiedsontwikkeling.nu/artikelen/pbl-auto-blijft-cruciaal-voor-mobiliteit-vooral-in-stadsranden-en-op-platteland/

Nitsche, P., Widhalm, P., Breuss, S., & Maurer, P. (2012). A strategy on how to utilize smartphones for automatically reconstructing trips in travel surveys. *Procedia-Social and Behavioral Sciences*, 48, 1033-1046.

Özbil, A., & Peponis, J. (2007). Modeling street connectivity and pedestrian movement according to standard GIS street network representations. In *6th International Space Syntax Symposium.*

Rudolph, F., & Mátrai, T. (2018). Congestion from a multimodal perspective.

Sevtsuk, A., & Basu, R. (2022). The role of turns in pedestrian route choice: a clarification. *Journal of transport geography*, 102, 103392.

Sevtsuk, A., Basu, R., Li, X., & Kalvo, R. (2021). A big data approach to understanding pedestrian route choice preferences: Evidence from San Francisco. *Travel behavior and society*, *25*, 41-51.

Seyfried, A., Steffen, B., & Lippert, T. (2006). Basics of modelling the pedestrian flow*. Physica A: Statistical Mechanics and its Application*s, 368(1), 232-238.

Sokolovska, I. (2020). Current problems and environmental threats due to population migration as a result of urbanization. *Редакційна колегія*, 49.

Yamu, C., Van Nes, A., & Garau, C. (2021). Bill Hillier's legacy: Space syntax—A synopsis of basic concepts, measures, and empirical application. *Sustainability*, 13(6), 3394.

Appendix I – Table of content ZIP

Table of Content of the zip file that accompanies the thesis report.

- ReadMe.txt: Documentation of what is where in the file
- Final-Thesis-JuliaSipkema.docx: a word-version of the final report
- Final-Thesis-JuliaSipkema.pdf: a pdf-version of the final report
- Midterm-JuliaSipkema.pptx: a PowerPoint of the midterm presentation
- FinalPresentation_JuliaSipkema.pptx: a PowerPoint of the final presentation
- Literature: a folder with all the literature used for this thesis

Appendix II – SQL code

An overview of the SQL code used for the analysis of the data

<u>Gender distribution participants</u>

```
select gender, count(distinct tracker_id)
from trips_users
group by gender;
```

<u>Age distribution participants</u>

```
select age, count(distinct(tracker_id))
from trips_users
group by age
```

<u>Participants, trips, and average distance per level of urbanization</u>

```
select urban_density, count(distinct tracker_id), count(distinct(sensing_record_id), avg(distance)
from trips_users
group by urban_density;

select urban_density, avg(infra_distance)
from trips_users
where trip_mode = 'FOOT'
group by urban_density;
```

<u>Getting average journeys per day only using active participants</u>

```
select distinct(date(start_time)), count(distinct(journey_id))
into journeys_date
from journeys
group by distinct(date(start_time));

select *
into journeys_date_active
from journeys_date
inner join active_users
on journeys_date.date = active_users.start_date

select avg(1.0*count_j/users90)
from journeys_date_active;
```

<u>Average trips per day using only active participants</u>

```
select distinct(date(start_time)), count(distinct(sensing_record_id))
into trips_date
from trips
group by distinct(date(start_time));

select *
```

```
into trips_date_active
from trips_date
inner join active_users
on trips.date = active_users.start_date

select avg(1.0*count_t/users90)
from trips_date_active;
```

<u>Trips per day per mode per user</u>

```
select distinct(start_date), trip_mode, infra_distance, count(distinct(sensing_record_id)
into trips_mode
from trips
group by distinct(start_date);

select *
into trips_mode_active
from trips_mode
inner join active_users
on trips_mode.start_date = active_users.start_date;

select trip_mode, avg(1.0*countm/users90)
from trips_mode_active
group by trip_mode;
```

<u>Average distance per mode per trip</u>

```
select trip_mode, avg(infra_distance)
from trips
group by trip_mode;
```

<u>Average distance per person per day</u>

```
select sum(sum2)/avg(users90)/28
from trips_mode_active;
```

<u>Average distance per trip by foot</u>

```
select avg(distance/users90)
from trips_mode_active
where trip_mode = 'FOOT';
```

<u>Goals of walking trips and their count and average distance</u>

```
select rdoel, count(rdoel), avg(infra_distance)
from trips
where trip_mode = 'FOOT'
group by rdoel;
```

Appendix III - extra information RQ3

Utrecht

Looking at walking behavior in the province of Utrecht patterns can be easily spotted in the data (table A1). When analyzing the average walking distance you can see that people tend to walk longer distances per trip when the density of shops and meeting places is low. This could be because places of interest are most likely farther away when the density is lower. The walking distance then decreases when the density increases but goes up again at the highest density which can be found in city centers. This could be because in city centers people often tend to visit multiple stores and walk around for fun more than in non-urban areas.

Looking at the number of trips some intriguing things can be seen. What mainly is interesting is the number of walking trips as a percentage of total trips per area type. Here you can see that the denser the area, the more walking trips there are compared to other transport modes. Especially at the highest density the share of walking trips is high, almost two-thirds of all trips. This could be because distances between stores and meeting places are smaller, making it easier to walk between them. Another reason could be that city centers are often for large parts pedestrian-only, making it impossible to reach with other types of transport. What is interesting is that there is not just a positive correlation between density and the percentage of walking trips. There is a small decrease between the very low and the low density.

The average number of shops and meeting places are added because the bins are quite large. The average makes it easier to see on which end of the bin the distance or share can be placed, which makes it more accurate to put into a model. The low in Utrecht is for example quite on the low end of the bin (n*shopmeet* > 0 and <= 0.1), as are the medium and high (n*shopmeet* > 0.1 and <= 0.5 and n*shopmeet* > 0.5 and <= 2 respectively), while the maximum is relatively high in the bin (*nshopmeet* > 2) meaning that there is a relatively high density of shops in the maximum density area.

Number of shops and meeting places	**The average number of shops & meeting places**	**Average walking distance**	**Percentage walking of total trips per area type**
Very low	0	1324	22.7
Low	0.03	1207	20
Medium	0.22	954	27
High	1.02	781	33.8
Very high	4.79	965	67

Table A1: Walking distance, percentage of walking trips, and percentage of total trips per area type in Utrecht

Gelderland

Looking at the numbers for the province of Gelderland the numbers are quite similar to those of Utrecht (table A2). The average walking distance is a bit higher in less dense areas and a bit lower in dense areas than in Utrecht, but also shows the same decrease with an increase in density and then an increase in the highest density.

The percentages of the walking trips compared to other modes also follow a similar pattern compared to Utrecht with an overall increase when the density increases, but a small decrease between the lowest and second lowest density.

Looking at the average number of shops and meeting places, the numbers are again quite similar to those of Utrecht. The very high is, however, a bit lower than in Utrecht, meaning that the dense areas are less dense than in Utrecht.

Number of shops and meeting places	**The average number of shops and meeting places**	**Average walking distance**	**Percentage walking of total trips per area type**
Very low	0	1422	21.9
Low	0.02	1249	17.9
Medium	0.25	911	25.1
High	0.98	791	37
Very high	3.32	812	66.3

Table A2: Walking distance, percentage of walking trips, and percentage of total trips per area type in Gelderland

Groningen

Looking at the average walking distances, Groningen again shows the same pattern of the average distance decreasing with an increasing density of shops and meeting places and then increasing a little again at the highest density (table A3). The numbers are, however, a bit lower than in the previously mentioned provinces.

The percentages of walking trips compared to trips with another mode of transport show a small difference compared to Utrecht and Gelderland. Instead of showing a small decrease between the very low and low density of shops and meeting places, there is a fully positive correlation with a steady increase when the density increases.

Looking at the average number of shops and meeting places per bin, the numbers are again quite similar to what was shown before. The main differences are in the high, which is a bit higher than in the previous provinces.

Number of shops and meeting places	**The average number of shops and meeting places**	**Average walking distance**	**Percentage walking of total trips per area type**
Very low	0	1284	19.9
Low	0.03	1246	20.8
Medium	0.25	949	23.4
High	1.11	664	28.4
Very high	3.69	897	59.3

Table A3: Walking distance, percentage of walking trips, and percentage of total trips per area type in Groningen

Flevoland

The numbers of Flevoland look a bit different compared to those of the other provinces (table A4). This is because Flevoland is the only province in the Netherlands without areas with the highest density of shops and meeting places (*Nshopmeet > 2)*. Apart from that, the numbers in the four remaining categories are quite similar to those from other provinces. There is again a decline in average walking distance when the density goes up, with a slight increase again at the highest density. However, the average distances are a little lower than those seen earlier.

The walking trips as a percentage of total trips again increase when the density goes up. However, here the share of walking trips in the highest density is lower (43.3%) than in other provinces (more than 60%).

The average number of shops and meeting places also differs a bit from the other provinces, which could be expected since Flevoland does not have areas in the highest density type. But this is not the only difference: the low and the high also have a quite low average distance, which means that in Flevoland there are on average mainly areas with a low density of shops and meeting places.

Number of shops and meeting places	**The average number of shops and meeting places**	**Average walking distance**	**Percentage walking of total trips per area type**
Very low	0	1276	23.8
Low	0.002	1014	25.6
Medium	0.29	756	27
High	0.92	916	43.4
Very high	-	-	-

Table A4: Walking distance, percentage of walking trips, and percentage of total trips per area type in Flevoland

Noord-Holland

Looking at the province of Noord-Holland the numbers again tell a similar story as most of the previous provinces (table A5). The average distance again decreases with an increase in the density of shops and meeting places, but unlike the other provinces described before it does not go up again in the dense city centers. This could be because Amsterdam, the largest city in the province, has trams and subways in the city centers, making it possible to use types of transport other than walking.

The percentage of walking trips compared to other modes again shows an increase with an increase in density, but the percentages are closer together than in other provinces, which could also have to do with more transport options in the bigger cities than in other provinces.

The average number of shops and meeting places does not show a lot of differences compared to other provinces, especially looking at Groningen.

Number of shops and meeting places	The average number of shops and meeting places	Average walking distance	Percentage walking of total trips per area type
Very low	0	1270	22.2
Low	0.03	1120	22.3
Medium	0.25	940	31.6
High	1.06	881	37.2
Very high	3.48	879	54

Table A5: Walking distance, percentage of walking trips, and percentage of total trips per area type in Noord-Holland

Zuid-Holland

Zuid-Holland shows almost the same patterns as Noord-Holland: a continuous decrease in average walking distances and an increase in walking trips as the share of total trips with an increase in density where the numbers are again quite close together (table A6). This similarity could be because, just like Noord-Holland, Zuid-Holland has large cities (The Hague and Rotterdam) with more transport options than other Dutch cities.

The average number of shops and meeting places is quite similar to those of Groningen and Noord-Holland, with only high being a bit lower.

Number of shops and meeting places	The average number of shops and meeting places	Average walking distance	Percentage walking of total trips per area type
Very low	0	1209	22.9
Low	0.03	1081	20.7
Medium	0.26	947	27.7
High	0.96	886	39.4
Very high	3.30	859	52.3

Table A6: Walking distance, percentage of walking trips, and percentage of total trips per area type in Zuid-Holland

Noord-Brabant

When you look at Noord-Brabant (table A7) the average distance again declines when the density of shops and meeting places increases. However, compared to other provinces the average walking distance at the highest density (*nshopmeet* >2) does not increase or is similar to the second highest (*nshopmeet* >0.5 and <= 2), but decreases quite a bit.

The share of walking trips again shows a pattern we have seen in other provinces; a small decline between the lowest and second lowest density. The average number of shops and meeting places also are similar to those of the other provinces.

Number of shops and meeting places	The average number of shops and meeting places	Average walking distance	Percentage walking of total trips per area type
Very low	0	1334	23.2
Low	0.03	1203	20.2
Medium	0.24	917	25.7
High	1.07	779	39.5
Very high	3.75	626	64.3

Table A7: Walking distance, percentage of walking trips, and percentage of total trips per area type in Noord-Brabant

Limburg

Limburg again shows the same kind of pattern as the other provinces looking at the average walking distances (table A8). What is interesting to note is that the province has a quite high average walking distance, especially in the lower densities. While for most provinces it is around 1200 meters, for Limburg it is around 1400.

The share of walking trips compared to other modes of transport also shows a similar pattern as the other provinces: an overall increase in share when the density goes up except for a small decrease in the low-density areas. The shares themselves are also quite similar being around 20 percent in the lower-density areas and over 60 percent at the very high density.

The average number of shops and meeting places again are similar to those in other provinces. Only the maximum is a bit lower than some of the other provinces, meaning that the dense areas are on average less dense than in other parts of the country.

Number of shops and meeting places	The average number of shops and meeting places	Average walking distance	Percentage walking of total trips per area type
Very low	0	1455	21.4
Low	0.03	1411	19.9
Medium	0.26	916	24.3
High	1.16	861	45.7
Very high	3.2	878	61.7

Table A8: Walking distance, percentage of walking trips, and percentage of total trips per area type in Limburg

Friesland

Friesland shows the same patterns as Noord-Brabant with a continuous decline in the average distance. The average distance, however, in Friesland is a bit higher than in Brabant for most densities (table A9).

The percentage of walking trips compared to other trip types also is similar to that of Noord-Brabant with an increase when the density increases, except for the second lowest density type, where the percentage declines a little.

Looking at the number of shops and meeting places per bin the numbers are lower than for most of the provinces, except for the medium density which is close to the average. The low numbers show that in Friesland the numbers of shops and meeting places are lower on average than in other places.

Number of shops and meeting places	**The average number of shops and meeting places**	**Average walking distance**	**Percentage walking of total trips per area type**
Very low	0	1426	19.8
Low	0.02	1105	19.2
Medium	0.28	826	28.3
High	1.02	775	42.3
Very high	3.06	720	50

Table A9: Walking distance, percentage of walking trips, and percentage of total trips per area type in Friesland

Drenthe

When looking at the average walking distance per area type it is again similar to most of the provinces (table A10). Drenthe shows a decrease in walking distances when the density increases, but again shows a slight increase at the highest density. The gap between the lowest density and the second lowest density is larger than in most other provinces.

Looking at the walking trips as a share of the total trips there again is an increase with increasing shop and meeting place density. However, just as with some of the other provinces, there is again a small decline between the lowest (*nshopmeet* = 0) and second lowest (n*shopmeet* < 0 and <= 0.1) density types.

When looking at the average number of shops and meeting places the very low density, low, medium, and high are quite in line with what the other provinces showed before. The very high density, however, is a lot lower. While all of the other provinces have an average of at least 3, the average here is 2.1. This means that the dense shopping and meeting areas are less dense in Drenthe than in other provinces.

Number of shops and meeting places	**The average number of shops and meeting places**	**Average walking distance**	**Percentage walking of total trips per area type**
Very low	0	1427	21.4
Low	0.03	1055	17.9
Medium	0.25	841	24.6
High	1.07	742	34.8
Very high	2.13	750	65

Table A10: Walking distance, percentage of walking trips, and percentage of total trips per area type in Drenthe

Overijssel

Looking at the numbers of the province of Overijssel they again look similar to the numbers of the other provinces (table A11). For the average walking distance, there is again a decline with an increase in the density of shops and meeting places and then an increase at the highest density type.

The percentage of total walking trips compared to other transport modes also is in line with previously mentioned provinces; an overall increase in walking trips when the density increases with a small decrease at the second lowest density type and the highest density almost ⅔ of trips being by foot.

The number of shops and meeting places does not differ in pattern from the other provinces. The average in the medium-density areas is a bit lower but does not differ significantly.

Number of shops and meeting places	**The average number of shops and meeting places**	**Average walking distance**	**Percentage walking of total trips per area type**
Very low	0	1328	19.4
Low	0.03	1123	16.8
Medium	0.23	889	22.1
High	1.02	717	35.4
Very high	3.26	773	61.2

Table A11: Walking distance, percentage of walking trips, and percentage of total trips per area type in Overijssel

Zeeland

Zeeland being the last province it seems unlikely that the patterns differ a lot, but it does show quite some interesting small differences with the rest of the provinces (table A12). For example, the average walking distance at *nshopmeet* = 0 is the highest of all the provinces. And what is also interesting is that the average distance does not only decline or go up at the highest density, but that it goes up already at the second highest density and then decreases again.

The walking trips compared to other transport types again show the same patterns as most provinces with an increase, but a small decrease at the second lowest density type. The overall share of walking trips at the highest *density of shops and meet*ing places is, however, a bit lower than for most other provinces with the share at very high density being 43.1 percent while it is over 60 percent in most provinces.

Looking at the average number of shops and meeting places in Zeeland, they are in line with what we have seen before in the other provinces. The only difference is that in the very high density, the average is a bit lower than in most parts of the Netherlands, meaning less dense city centers.

Number of shops and meeting places	**The average number of shops and meeting places**	**Average walking distance**	**Percentage walking of total trips per area type**
Very low	0	1537	20.2
Low	0.03	1306	16.8
Medium	0.22	792	21.5
High	1.13	938	40.8
Very high	2.97	821	43.1

Table A12: Walking distance, percentage of walking trips, and percentage of total trips per area type in Zeeland