OV-fiets: where to go?

A study on OV-fiets user characteristics and destinations

© NS, 2021

UNIVERSITY OF TWENTE.

OV-fiets: where to go?

A study on OV-fiets user characteristics and destinations

Author:

Bas Pluister
(s1718150)

Master Thesis

Civil Engineering and Management

University of Twente

Version:

Final

September 21, 2022

Internal supervisors

Prof. dr. ing. Karst Geurs
Dr. ir. Baran Ulak

External supervisors

Ir. drs. Jeroen van den Heuvel MBA
Dr. Ir. Danique Ton

Preface

This report marks the final milestone of my study Civil Engineering and Management at the University of Twente in Enschede. After eight months of working on this thesis and living the nomad life between several residences and the office in Utrecht, I am happy to present this report.

The research has been carried out at NS Stations in Utrecht, therefore, I would like to thank NS Stations for the opportunity to conduct my thesis over there and for having a warm welcome in the office with the best view of Utrecht. I want to thank Team Onderzoek: Jeroen, Danique, Rik, Isabel, Do, Rosalie, Mirjam and Jelena for welcoming me to the team. Furthermore for their insights on my work, their experience on various subjects and the “gezelligheid” in the office, especially on “Clubhuisdag”. Next, I want to thank you all for letting me know about your network, work, and insights on working at NS. Also, for the trips, behind the scenes and experience in practice, such as the Zandvoort F1-weekend and the NS treinmodernisering visit. Also, I want to thank my fellow interns at Team Onderzoek: Juliëtte, Florian, Mikael and Olaf, for discussing on the process that is called writing a thesis.

Special thanks to Danique for supervising me in the starting phase of the thesis, helping shape the research, and always encouraging me to seek better insights, data or methods in our “weekly updates”. But especially to make choices in the process and to motivate them. And of course, for organising the “Rondjes buiten”, to never forget the outside experience and why you conduct research in the first place. To Jeroen, for the professional skills and work ethic, that I learned during the process and the ongoing support on my work and doings. For discussing the progress on Fridays after I send the update, although the discussions were for you just saying that I have already answered my own questions. Thanks for your motivation, energy, experience, and advises from you. To Do, for helping and guiding me with your extensive knowledge of surveying and market research, and help with setting up the survey with I&O Research.

Also, I want to thank Jeroen Steenbakkers from Argaleo for providing me with detailed spatial built environment data to conduct the research in the first place. Next, I want to thank all the surveyors of I&O Research who collected the respondents, although the survey hours were before dawn, and to be able to join you and experience the survey collection work myself.

From the University of Twente, I want to thank Baran for his academic view and guidance during the process of the proposal and thesis. Furthermore, for all the input in our meetings on the issues, I stumbled on during the process. And the new insights, directions and methods to think about and together found a solution that seemed unsolvable. I learned a lot from all those insights. Next, I want to thank Karst for connecting me with NS Stations and reminding me to have a critical view of the product and the methods used. And, of course, all my teachers during the master's, I definitely learned a lot during them, and every course was great to participate in, which confirmed me every time on having chosen the correct study.

Finally, I would like to thank my family and friends for my great student life and the support during my thesis. I want to give special thanks to Ruben for reading my thesis and checking for mistakes and discussions on the structure of paragraphs, sections and chapters. The “bierviltjes” will always be with me now. Especially, I want to thank my girlfriend Danielle for always having my back and keep supporting me in the difficult moments, even if that means that the country should be crossed (by train, of course). And last but not least, I want to thank my parents for their support throughout my study.

Bas Pluister,
Enschede, September 2022

Abstract

Not much is known about OV-fiets (a Station-Based Round-Trip bikesharing scheme (SBRT)) destinations and users, although OV-fiets have existed since 2008 with yearly increasing trips. The OV-fiets is operated by NS Stations (a subsidiary of the Nederlandse Spoorwegen (NS)) and allow for the station-to-door part of the trip and improves the door-to-door trip. This accommodates the change of public transport from station-to-station mobility to door-to-door trips. The NS is the Netherlands' national railway operator and a major public transport actor. The NS also changes from a pure railway company to a mobility company. The knowledge gap on OV-fiets makes strategic decisions, such as capacity planning, difficult to make.

The research aims to get insights on OV-fiets user characteristics and OV-fiets user destinations. A revealed preference survey collected 1538 responses with individual characteristics (e.g. socio-economic characteristics, and preferences and habits) and trip characteristics (e.g. destinations, frequencies and purposes). The survey was representative of weekday peak hours, which were around 70% of all weekly trips. With the individual characteristics, a multinomial logit model was estimated to determine the purpose of an OV-fiets user. With the destinations and built environment characteristics, a binomial logistic regression model was estimated to determine the probability of a destination being chosen by an OV-fiets user.

The OV-fiets user in the collected sample was highly educated, young, and regularly travel with OV-fiets to and from work. The median of an OV-fiets trip is 3 kilometres or 10 minutes. High-density areas and facilities with work or business functions, such as offices or meeting locations, increased the probability of a destination being chosen. The distance between 1150 and 3800 meters had the strongest influence on the destination probability.

The results of the models for the example case of the city Eindhoven showed that the main hotspots for destinations were visible in the industrial/office area De Hurk, the offices of Eindhoven Airport, the city centre and the university locations. Contributors of those locations were the high density of buildings and work-focused facilities. In more detail to user characteristics, trips with student travel cards were attracted by the high-tech campus university location more than other card types. Moreover, business card trips were more spatially distributed, while on-balance card trips were more clustered in the area, despite their equal trip amounts.

The research results show the key statistics of OV-fiets trips and the general OV-fiets user for weekday peak hours. Furthermore, it showed the main predictors for the destination of OV-fiets users, which were the distance between 1150 and 3800 metres and a high-density built environment.

Content

Figures

Tables

Abbreviations

Abbreviation	Meaning
ABA	Activity Based Approach
ADASYN	Adaptive Synthetic Sampling Approach for Imbalanced Learning
AIC	Akaike Information Criterion
BAG	Registration addresses and buildings (Dutch: Basisregistratie Adressen en Gebouwen)
BE	Built Environment
BGT	Topographical registration (Dutch: Basisregistratie Grootschalige Topografie)
BSS	Bikesharing system
BTM	Bus, tram, metro
CBD	Central Business District
CBS	Statistics Netherlands (Dutch: Centraal Bureau voor de Statistiek)
DCM	Discrete Choice Models
FSI	Floor Space Index
FSM	Four-step model
GLM	Generalized Linear Models
GIS	Geographic Information System
GSCL	Generalised Spatially Correlated Logit
IIA	Independence of Irrelevant Alternatives
IID	Independent, Identically Distributed
KBM	Ketenmobiliteits Belevings Monitor
KiM	Netherlands Institute for Transport Policy Analysis (Dutch: Kennisinstituut voor Mobiliteitsbeleid)
ML	Mixed Logit
MNL	Multinomial Logit
MPN	Mobility Panel Netherlands (Dutch: Mobiliteitspanel Nederland)
MXI	Mixed land-use Index
NHR	Commercial register (Dutch: Nederlands Handels Register)
NL	Nested Logit
NS	Dutch railways (Dutch: Nederlandse Spoorwegen)
NSSt	NS Stations
OD	Origin - Destination
ODiN	Dutch Mobility Survey (Dutch: Onderzoek Onderweg in Nederland)
OSM	OpenStreetMap
OSR	Open Space Ratio
OTM	OpenTransportMap
OV-fiets	Public Transport-Bicycle (Dutch: Openbaar Vervoer-fiets)
PBL	Netherlands Environmental Assessment Agency (Dutch: Planbureau voor de Leefomgeving
PC4	Postal Code with four digits, resulting in larger areas
PC6	Postal Code with four digits and two letters, resulting in smaller areas
PT	Public Transport
RP	Revealed Preference
RUDIFUN	Spatial Density and Mix use in the Netherlands (Dutch: Ruimtelijke Dichtheden en Functiemenging in Nederland)
SBRT	Station-based round-trip bikesharing
SMOTE	Synthetic Minority Oversampling Technique
SP	Stated Preference
TAZ	Traffic Analysis Zones
WSS	Within Sum of Squares

Extensive summary

Public transport is changing from station-to-station mobility to door-to-door trips. The Nederlandse Spoorwegen (NS) is the national railway operator in the Netherlands and a major actor in the public transport field. The NS changes too, from a pure railway company to a mobility company. This is partly facilitated by the OV-fiets, a Station-Based Round-Trip bikesharing scheme (SBRT) operated by the subsidiary NS Stations, which allows for the station-to-door part of the trip. Although it is possible to make strategic decisions (on matters such as capacity planning) in the short term, there is still a gap in knowledge to make long-term decisions. These decisions can be supported by long-term stable variables relating to the environment of rental locations, such as buildings. After all, buildings are built for the long term and can be described by built environmental characteristics.

Buildings tend to be the destination of OV-fiets users. Additionally, destinations are the reason why people travel in the first place. So in order to understand the OV-fiets system, it is necessary to identify destinations (built environment characteristics) and user characteristics of OV-fiets users. However, not much is known about the characteristics and destinations of OV-fiets users (or those of SBRT systems in general), although OV-fiets has existed since 2008 with yearly increasing trip numbers.

As such, the ability to make long-term decisions could be improved by increased knowledge of the destinations and user characteristics of OV-fiets users. Therefore, this research focuses on bridging this gap, using OV-fiets user characteristics and OV-fiets user destinations. This creates the main research question, as well as five sub-questions:

To what extent can the destinations of OV-fiets users for a specific rental location be predicted based on the individual OV-fiets user's characteristics and the built environment characteristics?

1. What is a suitable topology for OV-fiets rental locations?
2. What are the individual characteristics of OV-fiets users, and what are their destinations?
3. What are the built environment characteristics of OV-fiets users' destinations?
4. What are predicted attractive destination areas for an OV-fiets user group based on the built environment characteristics?
5. What are the limitations of the prediction method?

Station selection

OV-fiets is a nationwide system, and could not be investigated in detail on all locations, therefore, a selection of rental locations was made. OV-fiets rental locations had no classification, so the first sub-question was focused on creating one. The majority of the rental locations are located at train stations, and the OV-fiets user is predominantly a train passenger. Therefore, train station characteristics were used to classify OV-fiets rental locations.

Three clusters of rental locations were defined using a k-means clustering approach based on the silhouette method. The first cluster represented rental locations with longer and more often recreative trips, cluster 2 represented locations with more commuting trips, and cluster 3 represented locations whereby destinations were close to the station or had a short activity period. The rental locations were chosen to broadly represent their cluster. The selected train stations were Groningen, Nijmegen and Maastricht for cluster 1; Arnhem, Eindhoven and Amsterdam Sloterdijk for cluster 2 and Hilversum, Delft and Apeldoorn for cluster 3.

A survey was held to determine OV-fiets user characteristics and destinations at these stations. These results could be used to check the rental location classification. They showed that (compared to clusters 2 and 3) cluster 1 had a different age distribution, more school-going users, a lower trip frequency, more general weekend usage, more social-recreative trip purposes and a higher amount of destinations on the round trip. The satisfaction towards public transport and OV-fiets were similar among all clusters.

Clusters 2 and 3 were similar in the majority of the user characteristics. This means that the clustering approach did not create fully distinguishable clusters. The three clusters differed mainly on trip purpose. Therefore, the best way to separate the OV-fiets rental locations is by trip purposes.

OV-fiets user and travel characteristics

OV-fiets users can be identified by their socio-economic characteristics as well as their travel patterns, habits and their destinations, which answered the second sub-question. Given that around 70% of all OV-fiets trips per week are undertaken on weekdays during peak hours, a survey was also held during this time. As a result, more than 400 responses per cluster and a total of 1538 responses were collected. From those responses, 75% of the OV-fiets users filled in their destinations. The sample size was statistically big enough to have a representative view of the population.

The socio-economic distribution of the sample showed that the OV-fiets user tends to be highly educated, young or middle-aged and predominantly employed. Regarding travel patterns, the sample showed a high weekly usage of OV-fiets. The frequencies matched (semi-)daily commuting patterns. This commuting characteristic was also found with other travel-related factors, like how often travel costs were reimbursed. Finally, the satisfaction towards public transport is high, which is no surprise among daily public transport users.

The main destination tended to be work-related. Other trip characteristics implied the same tendency: users frequently travelled to this destination and mostly travelled to only one destination. The median trip distance to a destination was 3 kilometres or 10 minutes as trip duration. Many OV-fiets users pay for their bicycle themselves, but many others get it reimbursed by their employer, confirming the commuting characteristic. Most respondents stated that they chose the OV-fiets above bus/tram/metro due to convenience, freedom and speed. Finally, the respondents were not influenced in their travelling patterns by Covid-19 measures. During the survey, there were few measures, and respondents indicated that they used the OV-fiets equally (or more) compared to before Covid-19.

Built environment characteristics of destinations

The built environment can be described with the 7D built environment framework. For this research, the following variables were used: function of the area, surrounding facilities, availability of public transport and demographic characteristics of households. Due to time and computational limits, the built environment has only been identified for three cities (in station cluster 2): Amsterdam, Arnhem and Eindhoven. These cities are dominated by residential areas, surrounded by industry and office patches. Furthermore, small clusters of other functions are situated within the residential area, such as schools. Finally, facilities are located most frequently in the city centre and business parks. The destinations that were found in the survey were linked to these built environment characteristics by their location, using the Google Maps' Places and Directions APIs.

Built environment characteristics contributing to OV-fiets user destination choice

The relationship between the socio-economic characteristics plus trip characteristics and chosen destinations were determined using three different models, namely: the trip purpose model, destination model and combination model.

The trip purpose model estimated the OV-fiets user trip purpose, based on their socio-economic and trip characteristics using a multinomial logit model. The model distinguished three trip purpose classes: work, education and leisure. Respondents who chose education over work were significantly less likely to be (self-) employed and less likely to be reimbursed for the OV-fiets by an employer. Being a student, female or not having a daily trip frequency increases the chance of having an educational purpose. On the other hand, there was a higher chance for a leisure purpose when the user had a low trip frequency, when the user themself paid for the OV-fiets, or when the respondent had more destinations on the round trip. Also, older age categories tended to have leisure purposes instead of work purposes.

Next, a binomial logit model was used to make a destination model. This estimated the probability of 1 ha to be chosen as the destination for an OV-fiets user. For this model, the identified destinations of the survey were in the minority compared to the built environment data resolution, creating a highly unbalanced dataset. Therefore, the destinations were oversampled using the ADASYN technique, and the built environment data was undersampled using an unsupervised proportioned selection from a k-means clustering. Those methods were found to be the best performing resampling approach for unbalanced datasets. Based on a forward variable search, a model with 19 variables gave the best trade-off between data-efficiency and model performance.

The results showed that the distance has the highest impact on the probability of choosing a destination, especially in the interval of 1150 to 3800 meters. Furthermore, the probability increased with the presence of a higher density built environment and buildings focused on workers. Buildings and areas focused on recreation were less likely to be chosen.

Next, the destination model and the purpose model were combined. This adds insights into which users go to which destinations, as people choose locations where they can undertake their activity and neglect other locations. This creates the opportunity to connect OV-fiets user characteristics to chosen destinations.

Doing so, the combined model was able to create a spatial probability distribution (which means that each spatial unit has a certain probability of being chosen) for different OV-fiets user groups. Furthermore, a simulation could be performed on which destinations would be chosen by a generated OV-fiets user group to show the model possibilities and results. With this simulation, direct relations between user categories and destination areas could be identified. For example, it showed that destinations of trips done by student travel cards were more clustered, while the destinations of the other card types were more scattered throughout the area. Moreover, OV-fiets users who travel with a business card have more unique locations than OV-fiets users travelling with on-balance OV-chipcards.

Model performance

The trip purpose and destination model were validated to assess their performance. 20% of the survey data had been set aside for this analysis. The trip purpose model reached a 74% accuracy but a macro F1-score of 50%. This model was better at estimating work trips, primarily because commuters were the largest user group during peak hours. The combined trip purpose/destination model was compared to complete spatial randomness. The model performed better than that with 95% confidence.

A comparison was made between actual NS data on egress mode usage and the simulation run with the combination model. This was done to check the external validity and to identify differences between OV-fiets and bicycles. The comparison showed that most often chosen postal code areas by private bicycles were also found in the model for OV-fiets. A difference was that the OV-fiets destinations were more scattered throughout the area and located farther from the station than private bicycles. This diffuse pattern could partially have been caused by the modelling approach in which each grid cell has a (small) probability, which reduces big peaks in the estimation.

Conclusion and application

With the gathered information, the main objective could be answered:

To what extent can the destinations of OV-fiets users for a specific rental location be predicted based on the individual OV-fiets user's characteristics and the built environment characteristics?

Overall, OV-fiets users during peak hours were predominantly young, highly educated and employed. The OV-fiets users have a habit towards public transport and are highly satisfied with the OV-fiets product. Most peak-hour users commute to work or school, use OV-fiets multiple times per week, and get reimbursed by the employer. Next, recreational users use the OV-fiets less often. Furthermore, the median trip distance is around 3 kilometres, and the median trip duration is around 10 minutes. The survey sample was statistically representative for weekdays during peak hours, which is a 70% share for all OV-fiets trips.

The spatial probability distribution of a destination being chosen by an OV-fiets user was predominantly affected by the distance and the function area. High-density areas, facilities and meeting buildings also increased the probability of when a location was chosen. A simulation run with a generated population in the Eindhoven area showed trip destinations for them. Additionally, these showed direct relationships between specific OV-fiets user characteristics and their chosen destinations.

In short, the research gathered key data on OV-fiets distance and duration and provided insights on OV-fiets user characteristics and built environment characteristics of OV-fiets user destinations during weekday peak hours, contributing to scientific research on SBRT systems. The insights could be used for several societal applications. A direct application contributes to the supply capacity planning, forecasting and evaluation of rental locations, which can optimize the systems bicycle availability for OV-fiets users. Other applications of the results can be the integration of OV-fiets in land-use development, or the results might interest local stakeholders for opportunities in transportation planning to initiate a modal shift towards OV-fiets.

1 Introduction

Traditional public transport (PT) changes from station-to-station to a door-to-door trip. This trend is due to emerging digital information technology (mobility apps) and flexible transportation modes (vehicle sharing) (Molander, 2018). These improvements make the journey more convenient or can save travel time. As a result, public transport becomes more attractive to the traveller.

Access and egress are the first-mile and last-mile parts of a multimodal trip. The multimodal trips generally consist of one main mode and the first and last stretch to reach that main mode, respectively, home-end and activity-end, see Figure 1. The transition from a station-to-station trip to a door-to-door trip creates the need for access and egress of public transport to be more integrated with the public transport trip.

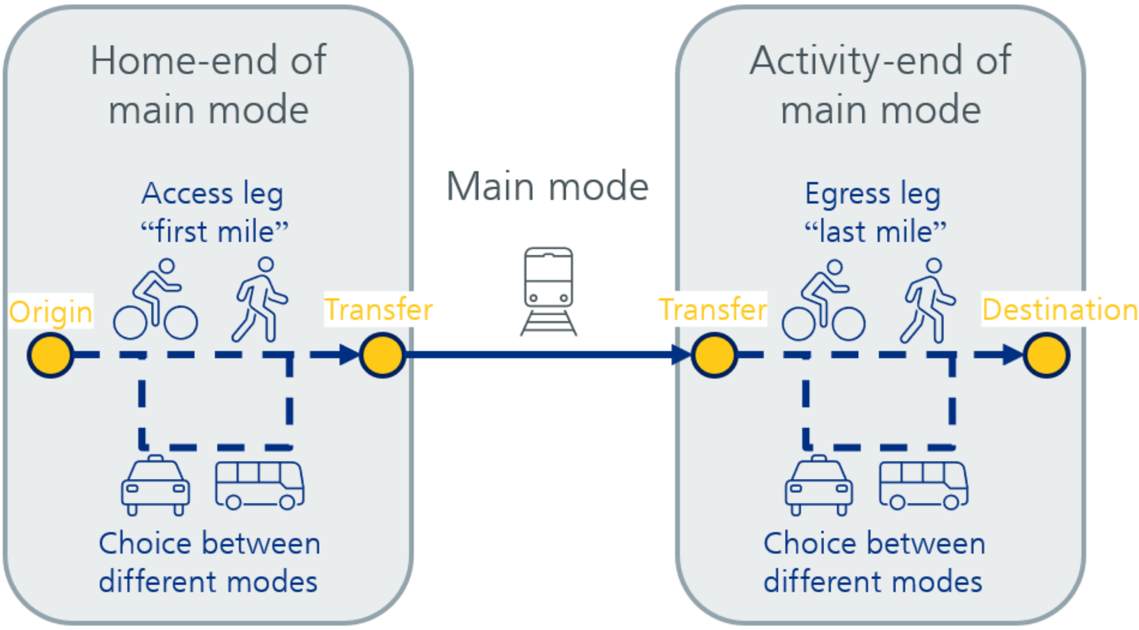


Figure 1: Visualisation of a multimodal trip (Wilkesmann (2022))

A prominent actor in the public transport field of the Netherlands is the Nederlandse Spoorwegen (Dutch Railways (NS)). Besides operating the main railway infrastructure, they own around 400 train stations with their subsidiary NS Stations (NSSt). Nowadays, the NS is focused on bringing travellers from station to station with their trains. To follow the trends within public transport, the NS wants to change to a mobility company to accommodate door-to-door trips (NS, 2020). The Openbaar Vervoer-fiets (Public Transport-Bicycle (OV-fiets)) is a bikesharing system located at train stations and operated by NS Stations. The OV-fiets is one of the ways to help NS with the change towards a door-to-door mobility company, with possibilities of accommodating the station-to-door trip.

1.1 OV-fiets system

The OV-fiets is a bike-sharing system (BSS) with some unique characteristics. The rental locations are solely located at train stations or major bus/metro hubs and can be rented within 3 seconds by OV-chipcard for a fixed price per 24 hours (Ploeger & Oldenziel, 2020). This makes that the system is designed to accommodate flexible station-to-door trips. Over the last years, OV-fiets has experienced an increase in trips with a drop in 2020 due to the Covid-19 pandemic, see Figure 2. The increase in trips emphasises the relevance of the system. Sections 1.1.1 and 1.1.2 describe, in short, the OV-fiets system compared to other systems.

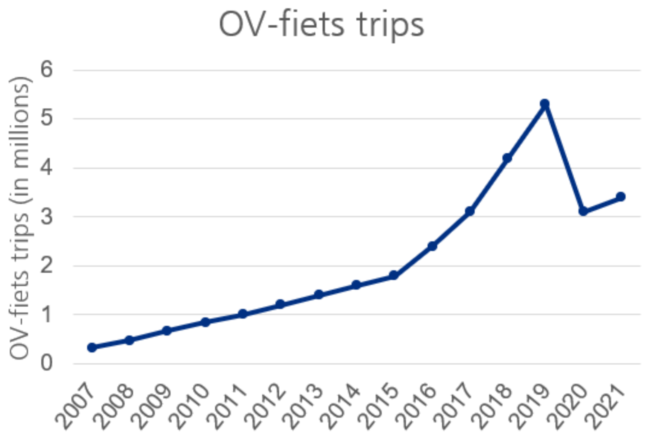


Figure 2: Development of bookings OV-fiets in recent years (NS, 2022a)

1.1.1 OV-fiets bikesharing

The characteristics make the system a so-called station-based round-trip bikesharing (SBRT) scheme (Wilkesmann, 2022). To understand the differences between the OV-fiets system and with more commonly known one-way BSS, van Waes et al. (2018) classified the systems on their use case (one-way versus round-trip) and availability (station-based versus free-floating), see Figure 3.

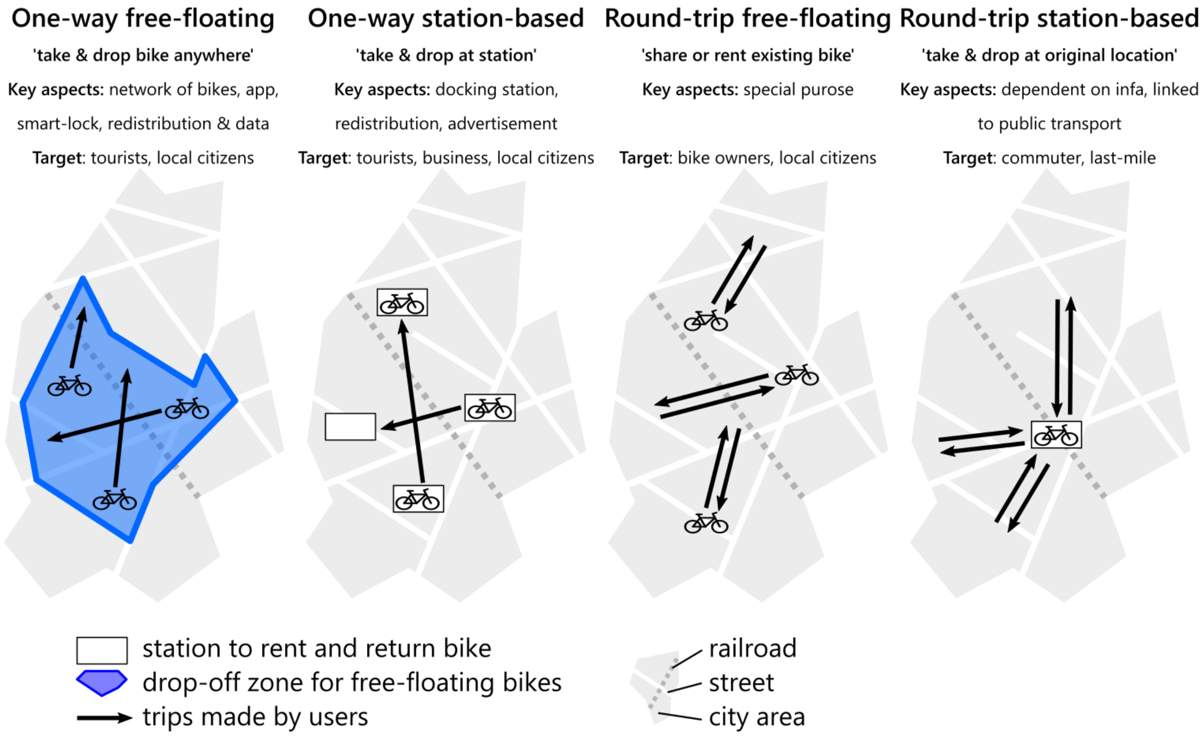


Figure 3: Categorisation of bikesharing systems, based on Waes et al. (2018), Wilkesmann (2022)

One-way systems of bikesharing allow users to pick up and drop off the bicycle anywhere. This can occur for both older and newer systems. One-way station-based systems have docking stations to which the bicycles should be returned, but the rental docking station is not necessarily the returning docking station. Free-floating one-way bikesharing systems do not require the bicycle to be rented or returned to a docking station. These bicycles can be parked anywhere in a particular zone, mainly in a city's built-up area. Those systems are more flexible but require the use of digital systems, such as apps for users and geofencing (DeMaio, 2009; Todd et al., 2021).

On the other hand, round-trip systems require the user to return the bicycle to the location where it was rented. Here too, station-based round-trip systems have docking stations at specific locations, and these locations are mostly dependent on infrastructure and linked to public transport. Furthermore, classic bicycle rentals at hotels or tourist attractions can also be categorised in the category of SBRT systems. The free-floating counterpart is less common worldwide but does exist (DeMaio et al., 2021). These systems are often referred to as peer-to-peer (P2P) bikesharing, in which a resident can rent out their own bicycle and request the lender return the bicycle where it has been picked up. According to van Waes et al. (2018), these systems are considered free-floating round-trip even though technically, each location where the bicycle is available by a person serves as a station due to the round-trip characteristic.

1.1.2 OV-fiets egress mode

Since the rental locations of OV-fiets are situated only at train stations, the OV-fiets user accommodates a seamless transfer for the train passenger (Schakenbos et al., 2016). The OV-fiets cannot be chosen as a mode to the train station or bus stop at the home-end of a trip because the OV-fiets is not available at someone's doorstep. This characteristic means that OV-fiets is only useful as egress mode, which connects the main mode with the destination. On the other hand, travellers do not have all the private modes available on the activity side, which the OV-fiets might fill the gap that the absence of privately owned bicycles leaves (Keijer & Rietveld, 1999).

Access and egress modes determine the availability of PT, whereby their reach determines the catchment area of the main mode (Ortúzar & Willumsen, 2011). When time and distance increase for an access/egress mode, the use of PT decreases (Krygsman et al., 2004). Each egress mode has a different focus area or range, see Figure 4.

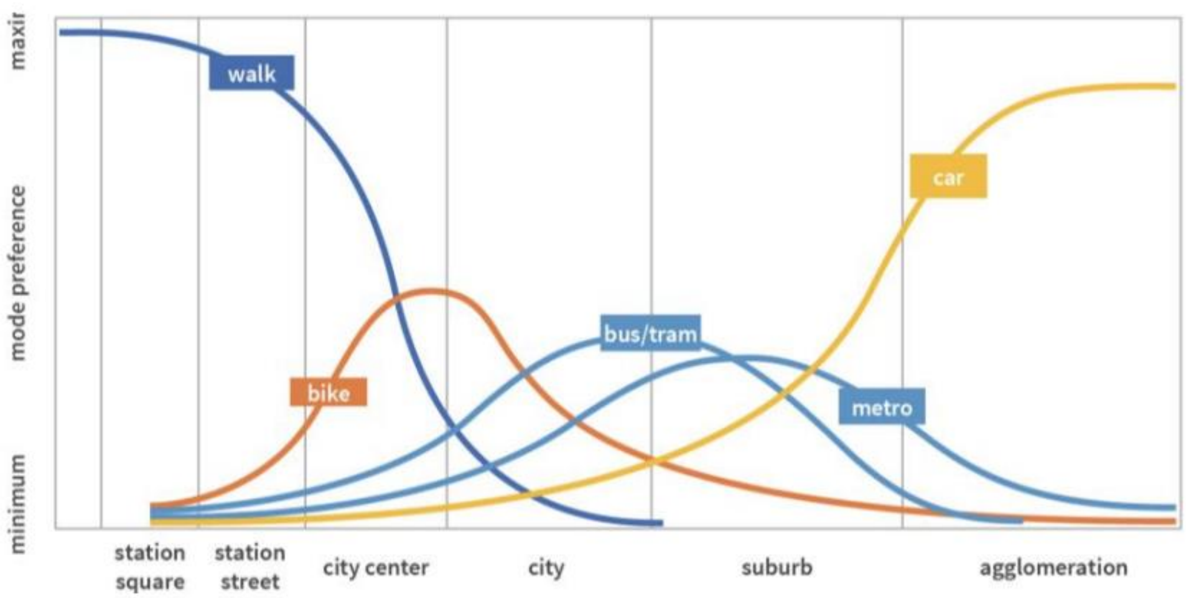


Figure 4: Distribution of mode preference for different egress trip distances (Bureau Spoorbouwmeester, 2012)

It is unknown where the OV-fiets precisely exist in this spectrum, but it is expected to follow a similar distance decay function as the private bicycle. But especially for short distances, it is safe to assume that OV-fiets is not used. For example, people don't pay a fee to cross the station square (Bureau Spoorbouwmeester, 2012).

Another difference for OV-fiets is availability. Due to the characteristics of PT main modes, people arrive at distinct times. The connecting egress mode should be available at those particular times. Walking and cycling are continuous modes, as they can be taken at any time, while BTM is only available at scheduled times (Rietveld, 2000). This availability issue can be translated to reachability too. Walking and cycling can reach every doorstep, while BTM (or car) can only reach stops or parking spots (Jäppinen et al., 2013).

Finally, the OV-fiets have a different payment scheme compared to other paid egress modes. Payment of OV-fiets is a single fee per 24 hours, while with BTM, the payment is per trip. This feature of OV-fiets makes bundling activities in one rental period attractive, like combining grocery shopping with a commute (Krygsman, 2004).

1.2 Problem and objective

The OV-fiets is a different kind of egress mode than the traditional one, but according to the increasing number of trips, the system is becoming more popular with travellers. However, little is known about the OV-fiets system by NS Stations or SBRT systems in general by researchers. Apart from historical patterns from booking information, it is difficult to make well-informed strategic decisions to the system by NS Stations. For instance, it is unknown who is travelling with OV-fiets due to privacy regulations and unknown where those users are going because the bicycles have no GPS-system. Therefore, it is mostly unknown how or why patterns or demand occur, and the system can be seen as a “black box”.

Due to this, it is difficult to estimate demand or plan supply at a location to increase the availability of bikes**.** However, such decisions would help make the system more efficient for door-to-door trips or other applications, such as Mobility as a Service-programmes (Miramontes et al., 2017).

The supply estimation can be divided into two components: the long-term (strategic decisions) and the short-term (operational and tactical decisions) (Tirachini et al., 2013). Recent research by Wilkesmann (2022) focused on seasonal variation of OV-fiets trips, which gave insights for short-term forecasting of demand. Research and insights lack for the long-term or trends. The strategic decisions for the long-term of the OV-fiets system can be improved. This helps the (reactive) decisions on historical patterns to be flipped to (proactive) predictions.

For the long-term, there can be looking at the historical patterns in the long-term (reactive), but it is more interesting what elements attract OV-fiets users and trips (proactive). The OV-fiets user is foremost a train passenger due to the rental locations at train stations. Train station attraction values are known and foremost based or calculated on surrounding destinations in the environment (Cervero et al., 2013; Ewing & Cervero, 2010). For instance, students will go to train stations with universities nearby. The environment is a stable variable because buildings are built to stand the test of time. Due to its constancy, the built environment shows a possibility in helping to identify OV-fiets demand for the long-term.

However, it is unknown how the built environment relates to the OV-fiets because data, information and insights are lacking. Therefore, the ‘black box’ of the OV-fiets should be opened. A first step is to fill the data gap and identify relations between the constant built environment and the OV-fiets destinations and OV-fiets users. This can be done by extracting two issues from the user: the user itself (such as ‘being a student’) and its destination (such as ‘the university’). This leads to the objective of this research, which is twofold: (1) to get more insight into the users of OV-fiets and (2) to get more insight into what destinations those users choose.

1.3 Research questions

The problem and objective of the research show that little information is known about OV-fiets users or SBRT in general. This research aims to fill this knowledge gap by answering the following research question:

To what extent can the destinations of OV-fiets users for a specific rental location be predicted based on the individual OV-fiets user's characteristics and the built environment characteristics?

In order to answer the main research question, several sub-research questions are formulated, and their objectives are briefly mentioned. Each research question focuses on a part of the main aim.

1. What is a suitable topology for OV-fiets rental locations?

Many train stations have OV-fiets rental locations. Due to time and budget limitations, it is impossible to investigate all those rental locations. Therefore, a representative selection of rental locations should be made. Currently, there is no categorisation or classification for OV-fiets rental locations. Therefore, locations that are representative for all (or most) rental locations should be found based on their characteristics. This sub-question will investigate which rental locations are suitable for this representative set to use in further research steps.

2. What are the individual characteristics of OV-fiets users, and what are their destinations?

The second sub-question aims to find the individual characteristics of OV-fiets users and information on the destinations where they are going. Currently, information on the characteristics of OV-fiets users is limited, and users' destinations are unknown. Therefore, this sub-question aims to fill the data gap.

3. What are the built environment characteristics of OV-fiets users' destinations?

The chosen destinations have been determined through sub-question 2. These can not be used yet for this research, as those are only addresses or places. Therefore, this sub-question aims to find the spatial location. Furthermore, the context of these spatial locations should be found too. Therefore, they should be combined with built environment characteristics to enrich the spatial locations.

4. What are predicted attractive destination areas for an OV-fiets user group based on the built environment characteristics?

The previous sub-questions have established user characteristics, chosen destinations and destination characteristics at selected rental locations. Based on these, this sub-question examines the connection between those variables with destination choice models.

5. What are the limitations of the prediction method?

The last sub-question discusses the found relationship and the limitations of the used models. Based on these, it is possible to answer the main research question.

1.4 Scope of the research

This research focuses on filling the research gap for SBRT systems with OV-fiets user characteristic identification. The OV-fiets user characteristic variables that are used in the research are determined in the literature review, described in chapter 2. The method is a survey to gather revealed preference data, which is described in detail in methodology section 3.1.

Next, the research focuses on filling the research gap for destinations chosen by OV-fiets users. It focuses on which built environment characteristics contribute to the destination choice of OV-fiets users. The output of the research would be a probability distribution of destinations chosen by an OV-fiets user. The output can contribute to estimating the number of bicycles per rental location, as described in the research relevance in section 1.3. But this practical relevance is out of the scope of this research. A visualisation of the research scope is displayed in the yellow area in Figure 5.

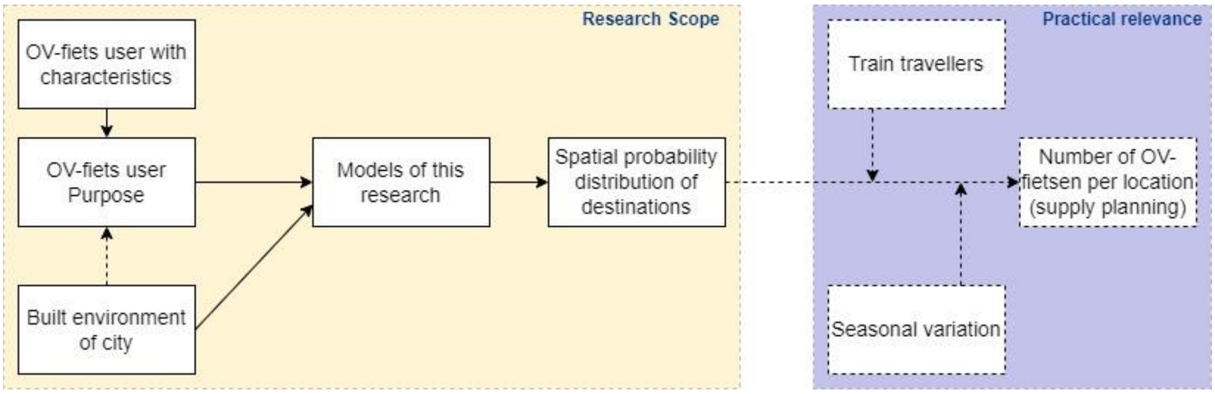


Figure 5: Scope of the research

The goal of this research is to investigate the OV-fiets users and how the built environment contributes to their destination choice. The figure shows the research aspects of the OV-fiets user and, separately, the built environment of the city. There can be noted that the built environment of the city does influence the train traveller. Big cities would attract more users, and universities attract students (Cervero et al., 2013). However, in this research, the relation of the environment to a person is held out of scope. Therefore, the other way around: to what extent the destinations attract an OV-fiets trip is out of scope.

Regarding OV-fiets user patterns, Wilkesmann (2022) showed a clear difference between week and weekend days. Therefore, it is expected that the OV-fiets user characteristics between a week and weekend days are different, and those populations cannot be accumulated. Due to the research's time and budget, it is impossible to analyse both populations.

For both populations hold that there is a knowledge gap. Therefore, this research focuses on the biggest share of OV-fiets trips, which are on weekdays. It is more relevant to focus on the majority of all trips than the minority share.

The study area of the built environment of possible destinations consists of an area around a rental location with a radius of 10 kilometres from that location. This is based on observations from Martens (2007) and Rijsman et al. (2019), who found trip lengths of bicycles around Dutch train stations. Furthermore, the research focus on the built environment characteristics, therefore, land-use is out of scope. This means that the research does not capture parks, nature reserves or other non-built areas.

1.5 Relevance of the research

The results of this research provide an analysis of the characteristics of OV-fiets users and what built environment characteristics contribute to their chosen destinations. This allows for relations between the built environment and OV-fiets destinations. This relationship shows the attractiveness of a destination as a probability of choosing that destination. These results are expected to have societal relevance and scientific relevance.

From a business point of view or societal relevance, the research results will help NS Stations understand their system better. Key statistics on distances travelled by OV-fiets users can help define their system's catchment and effectiveness.

Moreover, the research results can be combined with OV-fiets trip data. The combination can help forecast OV-fiets trips or trends in the long term, allowing for demand identification at rental locations. Together with the seasonal variation of Wilkesmann (2022), it can support supply planning at rental locations.

Furthermore, current supply capacities can be evaluated. The evaluation can then define if there are too many or too few bikes at a rental location. Similarly, future OV-fiets rental locations can be planned with more sophisticated supply capacities.

Improved matching of supply and demand allows for higher availability of the system. Both current OV-fiets users and potential users of the OV-fiets will benefit from this. Knowing that a bike is available will make the transfer to the OV-fiets more convenient and ease the multimodal trip. Furthermore, knowing who is travelling with OV-fiets, dedicated marketing campaigns can be set up. These campaigns can be more focused on the (different) OV-fiets user groups and can increase the effectiveness of the promotion.

The results indicate an attractive built environment for OV-fiets users. When the built environment changes around a rental location, the demand for that location might change too. Therefore, the results can help in responding to those changes in time.

Increasing OV-fiets trips will probably help in increasing public transport trips in general. Both for the public transport operator as well as local stakeholders, this is beneficial. Even the other way around, knowing what attracts OV-fiets users, the built environment can be landscaped in a way to increase OV-fiets trips. This might reduce car usage or shift train egress trips from crowded metro, tram, or bus services towards the bike.

Finally, the insights of this study can contribute to other projects of the company, such as the OV-ebike (an electric OV-fiets equivalent), capacity distribution for multiple OV-fiets rental locations at a single train station, or investigating possibilities for returning the OV-fiets to other rental locations. Here, the knowledge of destination choice for OV-fiets users creates the opportunity to respond to it. Suppose the majority of the attractive destinations are south of the train station with two rental locations. In that case, most of the OV-fiets stock can be located in the southern rental location. Or, if potential attractive areas are outside a convenient trip duration, an e-bike equivalent can create the opportunity to reach them due to the increased speeds (which means increased range).

Besides the business relevance for NS Stations, other potential operators for SBRT systems, such as public transport operators and authorities, also benefit from the research. Insights on SBRT-systems can be used to understand the system and its benefits to the local transport system. Furthermore, the insights can enhance (local) multimodal transportation or integrated land-use and transport planning.

The research has a scientific relevance too. First, this research will fill the knowledge gap on the users of SBRT systems, particularly in the Dutch context. Within the literature, the focus is on systems with one-way bikesharing (for example, Faghih-Imani & Eluru (2015) and Y. Guo & He, (2020)). This is presumably because those systems are more common worldwide and, therefore, are more relevant to study (Médard de Chardon et al., 2017). Therefore, this research gives insights into other types of BSS. In addition, the differences between the one-way and round-trip BSS can be made more explicit.

Although much research on one-way BSS has been conducted, limited research has been done on the relation between bikesharing systems and built environmental variables, both for SBRT and one-way systems. The available articles solely focus on one-way bikesharing systems and are not in the Dutch context (Faghih-Imani et al., 2017; Faghih-Imani & Eluru, 2015; Guo & He, 2020). Despite this, Guo & He (2020) describe that particularly the Dutch context can be interesting due to the cycling culture. Therefore, the results of this research can help understand the characteristics of SBRT-system users in the Netherlands and the built environmental influences on the destination choice of SBRT users.

1.6 Research structure

The remainder of this report is structured as follows. First, a literature review of the subject is described. Following the theory, the methodology is described in chapter 3. This chapter also includes data availability and data gathering. After that, the results of the analysis are described in chapter 4. Next, the conclusions are described in chapter 5. Finally, a discussion is written in chapter 6. Chapters 7 and 8 contain the references and the appendices.

2 Literature review

This chapter provides a literature review on determinants for destination choice and bikesharing usage in section 2.1. After that, the literature on modelling destination choice problems is described in section 2.2. Section 2.3 contains the summary of the literature and the conceptual model, which is used for the methodology.

2.1 Destination choice

The destination choice problem is essential in transportation planning processes. The problem is defined as finding the probability that a person travelling from a given origin will choose a destination among many available alternatives (Bekhor & Prashker, 2008). Those found solutions can then be used in applications such as route prediction, travel intensity calculation or help in making policies.

The probability of a person travelling to a certain destination can be derived from the utilities of the alternatives (Ben-Akiva & Bierlaire, 1999; Ortúzar & Willumsen, 2011). Those utilities can be quantified for each alternative, so they can be used as determinants in predictions. In the literature, many studies are dedicated to researching the topic of destination choice. In the following sections, potential determinants and influences on bikesharing usage and destination choice found in the literature are described, regarding built environment determinants in section 2.1.1, travel mode determinants in section 2.1.3 and individual determinants in section 2.1.5.

2.1.1 Built environment

As described above, destinations can be defined or quantified by physical characteristics or determinants, in other words, by the built environment. The built environment or physical environment is defined as everything in the open space (Brownson et al., 2009). Several variables or characteristics can be subtracted from the environment. However, the built environment captures a large spectrum, and the definition and use of the characteristics vary between studies. Therefore, a framework is described to clarify how the built environment is defined in this study.

Definition

Categories of the built environment have often been named with words beginning with "D". This feature has been conceptualised in the 5D built environment framework: Density, Diversity, Design, Destination, and Distance (Ewing et al., 2009). In more recent studies, the 5D's were not appropriate to capture all variables related to the planning of transportation systems. Two aspects were added to 5D, which were found to be of influence in travel studies, namely: demand management (such as parking costs or bicycle availability related to demand management) and demographics. This resulted in the 7D built environment framework (Ewing & Cervero, 2010):

- **Density** is measured as the variable of interest per unit of area. The variables can be population, dwelling units, employment or floor area.
- **Diversity** refers to the different land uses in a given area and the degree to which they are represented. Entropy measures are widely used in travel studies.
- **Design** refers to street network characteristics within an area. These can vary on the number of intersections or bicycle lane coverage.
- **Destination** is the accessibility measure of ease of access to trip attractions. The destination category also captures particular destinations, such as hospitals or tourist attractions.
- **Distance to transport** refers to the distance between a working or home location to the nearest public transit stop.
- **Demand management** includes variables that influence demand, such as parking supply and costs.

- **Demographics** are variables that cover individual characteristics of the destination, similar to socio-economic variables.

As the 7D framework is used in several studies in urban planning and transportation (see, for example, (Brownson et al., 2009; Cervero et al., 2009; Gao et al., 2021; Ogra & Ndebele, 2014; Sung et al., 2022). This framework will be used to define variables and determinants in this research.

Influences of the built environment

Determinants of the built environment can then be divided by several topics combined with the framework described above. As the research focuses on SBRT OV-fiets, the determinants are described in relation to bikesharing.

Infrastructure

Infrastructure determinants relate to the design category of the built environment. Previous studies found that longer bicycle lanes positively impact bikesharing (Jonkeren et al., 2021; Orvin & Fatmi, 2020). Both studies found that more and better infrastructure for bicycles specifically leads to more use of bikesharing due to a higher safety perception of the user (Eren & Uz, 2020). They also stated that the bicycle infrastructure being a cycling lane or a separate pathway makes no difference. Those studies focused on one-way bikesharing of bike-use in general. Although the impact might be similar for SBRT bikesharing systems, this has not been described in the literature (yet).

Furthermore, better and more infrastructure also leads to better accessibility of jobs (Geurs et al., 2016; Guo & He, 2020). Higher accessibility leads to more travel demand (Geurs & van Wee, 2004), which is also the case for bikesharing systems (Eren & Uz, 2020). Again, studies researched one-way bikesharing, and no distinctive literature has been found for SBRT bikesharing, but the same impact might be true for SBRT systems.

Infrastructure has a direct impact on route choice and accessibility. However, the focus of this research is on destination choice and not on exact route choice. Therefore, the infrastructure is out of the scope of this research. Moreover, the Dutch infrastructure is more advanced than other countries (Fishman, 2016). As cycling infrastructure is generally of constant and high quality, the differences between routes are small. As such, this variable is of minor importance here.

Topography

Topography relates to the terrain. Uneven terrain causes roads and infrastructure to be designed with slopes and gradients. These slopes and gradients are also related to the design category of the built environment framework. Studies into bikesharing concluded that hillier terrain leads to less use of the systems (Todd et al., 2021). It was observed that people predominately use BSS in the downhill slope direction for one-way BSS. This creates rebalancing issues, especially when slopes exceed a 2% gradient (Eren & Uz, 2020).

There has been no literature found on topology related to SBRT systems. With those systems, rebalancing issues are not a problem because people have to return their bikes, creating an obligation to cycle uphill. As this coercion cannot be avoided, hillier terrain might impact the use of such systems. However, in the context of this research, the topography will be left out. OV-fiets only operate in the Netherlands, and the country is mostly flat, which eliminates topography.

Land use

Land use relates to the density and diversity of the built environment framework. Literature states that a higher population density positively impacts usage (Todd et al., 2021) and that a more diverse land use creates more possible destinations and more BSS usage (Guo & He, 2020). This observation is confirmed by Eren & Uz (2020), who found that a higher residential housing density correlates highly with BSS trip generation. On the other end, trip attraction-wise, a high density of office and commercial buildings and a short distance to schools and universities lead to higher attraction during the week, while recreational areas result in higher attraction during the weekend. Furthermore, Faghih-Imani & Eluru (2015) found similar results with the vicinity of parks and central business districts (CBD) to impact BSS use positively.

Looking at SBRT systems, no specific literature has been found on land use. However, it is likely that one-way BSS's impact factors are similar. For example, Faghih-Imani & Eluru (2015) found a daily commuting pattern in their research on one-way trips, indicating that morning and evening peaks exist and that users might be the same. In that case, the trips made by those users are round-trips. They found that nearby CBD or parks have a positive impact on BSS, so those might also have a positive impact on SBRT.

It should be noted that their research covered an American case, where a CBD has a more distinct definition, which can be different in other countries. Urbanity factors can be used to translate those kinds of areas to other places. These factors describe the diversity and density of an area (Harbers et al., 2019). In that regard, these factors can be used for a fair comparison between areas where a CBD can be defined outside the American context.

2.1.2 Trip purposes

A particular element of the built environment framework would be the destination category. This category is heavily related to trip purpose.

Destination

In general, people have somewhere to go. They usually plan to be somewhere to do their activity, such as sports, jobs or recreational matters. Recker & Kostyniuk (1978) found that a destination choice is built by the perception of the destination, accessibility, and the number of opportunities to exercise any particular activity. If people have an activity in mind, they look into that category of destinations where that activity can be fulfilled. This decreases the perception of destinations with other activity categories, eliminating them from the choice options. For example, if someone wants to have a recreational activity or trip, they don't take work facilities into account.

As described earlier, the purpose of a trip is an important factor in destination choice. Further literature confirms this, for instance, Chowdhury et al. (2020) state that the trip's purpose matches that trip's destination. With this in mind, destination prediction can be undertaken. Relevant points of interest could be used in models, and as found by Schimohr & Scheiner (2021), healthcare facilities, kindergartens, sports, tourist attractions, churches, and playgrounds were significant positive predictors in bikesharing use. Like previously discussed literature, those trips were found for one-way bikesharing but are presumably still relevant for SBRT bikesharing. Maybe even more, as SBRT is more used as an egress mode in multimodal trips.

In conclusion, the purpose of a trip can be a good predictor of destination choice. People only consider destinations where their activity can be undertaken and not all locations in the environment. This means that only destinations with the same function are chosen.

2.1.3 Mode

Besides from built environment influences on the destination choice, mode choice is also important in the destination choice. Different modes have different catchment areas and, therefore, different accessibility of activities. This means that mode choice cannot be separated from destination choice. Although, for this research, the mode choice is fixed (OV-fiets), it is still useful to understand mode choice influences. The following sections will describe those influences on bikesharing use and destination choice.

Trip distance

Trip distance indicates the distance between the point of origin and the point of destination. The literature describes that a longer trip distance means less use of one-way bikesharing (Adnan et al., 2019; Faghih-Imani & Eluru, 2015; Salah Mahmoud et al., 2015), which means that far-away destinations are less likely to be chosen. On top of that, users are willing to walk around 400 meters from a public transport access point to their destination (Gu et al., 2019b; X. Zhang et al., 2021). For station-based systems, this metric is useful to space out the docking stations and optimise the system's catchment. Although this creates opportunities for the multimodal public transport system to be more refined, people might experience an extra transfer to their destination due to the obligation to use a bicycle or have to walk a certain distance. This can increase the threshold for making such multimodal trips (Schakenbos et al., 2016).

For SBRT systems, no literature has been found on trip distance, but the same findings might hold. Given that the distance must be travelled twice (round-trip), the willingness to cycle to a certain destination might be shorter. On the other hand, comparing SBRT to it is free-floating sister systems, no docking station has to be found on the activity side. This creates a flexible parking opportunity for the bicycle and minimises the walking distance from the bicycle to the destination.

Availability

A higher density of docking stations or a higher spatial availability of free-floating bicycles throughout an area positively impacts the use of BSS. People prefer having available bicycles as close as possible to them (Médard de Chardon et al., 2017). This is also connected to the walking distance people are willing to cover to access a BSS. For one-way systems, the availability of bicycles is one of the most significant impact factors on usage (Ricci, 2015). For SBRT systems, no specific research has been found. However, the OV-fiets rental locations are primarily located at public transport hubs. Therefore, literature on transportation hubs is relevant, like a described catchment area of those and the available other modes that can be used (Y. Zhang et al., 2017). Moreover, a shorter distance from train platforms to the bikesharing dock creates more demand for bikesharing in general (la Paix et al., 2020). This emphasises the train-bike connection and, therefore, will be described in more detail below.

Weather

Cycling is an outside activity, therefore, bikesharing users have to deal with the weather. Eren & Uz (2020) extensively researched weather impacts on BSS. They found that sunny, low wind, summer days with temperatures between 10-30 degrees Celsius are the best circumstances for bikesharing. Lower temperatures, precipitation and high humidity cause people to avoid bikesharing. Although this research only covered one-way bikesharing, the cycling activity is the same. Therefore, it is expected that for SBRT systems, the same conditions hold.

Wilkesmann (2022) found little difference for morning peaks when there was rain in this research on OV-fiets. However, during rainy weather, the number of trips was slightly lower for the rest of the day, as expected by Eren & Uz (2020). Wilkesmann (2022) concluded that commuters relying on the OV-fiets don’t mind a bit of rain, while recreational or non-commuting-related trips have more flexibility and avoid rain.

Temporality

Temporality relates to times and seasons. As described above, Eren & Uz found that summer has a positive influence on the usage of BSS. Research on weekly patterns was done by several studies covering trip data of BSS. They found clear different patterns between weekdays and weekends (Gu et al., 2019a). This result was confirmed by Todd et al. (2021) in a 300-station-based and free-floating system comparison. Furthermore, they found that morning and evening peaks exist in the system. On SBRT-systems, little research has been found on temporality. As explained before, the research of Faghih-Imani & Eluru (2015) showed that rush hour peaks could be translated to round-trip systems. Additionally, the temporality of the OV-fiets was described by Wilkesmann (2022). He found a clear morning and evening peak and a clear difference between week and weekend days. Although the results were train station-specific, the peak patterns existed in every case.

Alternative modes

The place in the existing (public) transport network is important for bikesharing systems. It will determine if bikesharing is a competitive mode or if people have other modes to choose from. Related to the built environment framework, it is described as the distance to transport. Eren & Uz (2020) described that bikesharing competes with public transport services on shorter trips, while with longer trips, bikesharing serves as an access and egress mode. This is consistent with other research that describes that a higher number of bus stops will lead to more use of that bus and less use of bikesharing (Faghih-Imani & Eluru, 2015; Guo & He, 2020; Orvin & Fatmi, 2020). Furthermore, the availability of escorting services (pick-up by family or friends) will also reduce the use of bikesharing (Adnan et al., 2019). In general, bikesharing is used as an extension of the traditional public transport network to increase catchment and provide access to additional areas (Rijsman et al., 2019). This also means that in areas with less public transport, bikesharing will be the only reasonable alternative for people who rely on public transport (while walking is always available, long distances might cause that it is no reasonable alternative). Therefore, those destinations are likely to be chosen more from a bikesharing perspective. SBRT have no specific research on this topic in literature, but cycling is often researched in combination with a public transport main mode, predominantly train modes.

Bike-train multimodal trip

Recent literature investigated multimodal trip chains focussing on bike-train chains. It was found that integrating bicycle and public transport policies creates more use of both systems (Geurs et al., 2016) and that an integrated transfer creates more use of both systems (Guo & He, 2020; Schimohr & Scheiner, 2021). Moreover, an increase in services of one mode can create higher use of the other mode due to higher passenger amounts or better connections (Cervero et al., 2013; Jonkeren et al., 2021). As explained before, SBRT research lacks specific cases, but research has been done on bikesharing in combination with connecting public transport. The bike-train system can provide the best attractive alternative to car-based interurban mobility (Nello-Deakin & Brömmelstroet, 2021). Therefore, the SBRT will most likely play an important role, especially for multimodal trips.

Eren & Uz (2020) found that it is important to acknowledge different systems in the bikesharing spectrum, as some are increasing public transport catchment, while others are providing an alternative for (overcrowded) public transit lines (Schimohr & Scheiner, 2021). In general, the authors found that bikesharing stations closer to PT hubs have increased usage, especially with systems integrated with the same tariff or card system, as thresholds between systems are low.

2.1.4 Joint choice

Mode choice is not the only choice people make; they also make the destination choice. Those choices are made simultaneously and not after one another (Chowdhury et al., 2020). These joint choices are a trade-off between destination and mode (Ton et al., 2020). Therefore, mode choice cannot be made separately from the destination. For example, if a cinema cannot be reached by public transit, the car might be chosen. But if that person has no access to a car, the trip cannot be made. This can cause the person to choose for another cinema or destination.

In this study, the mode choice is a given (namely OV-fiets), so the influences on mode choice cannot be measured. However, popular destinations might be influenced by the availability of other modes. Therefore, this research will focus on destination variables and available alternative modes

2.1.5 Individual characteristics

The third class of factors influencing destination choice are individual characteristics. These can be divided into objective characteristics (socio-economic characteristics) and subjective characteristics (habits and preferences).

Socio-economic characteristics are individual attributes like age, gender, education level, car ownership income level and the need to pay for public transit. Users of bikesharing are predominantly white, male and highly educated (Schimohr & Scheiner, 2021; Shaheen et al., 2010). One should be aware, however, that BSS are implemented in such neighbourhoods, leading to bias and self-selection and the image of 'wealthy early adopters' (Todd et al., 2021). People who like bikesharing will choose such an area to live or go to where those systems are available.

Despite this, individual characteristics help predict the destination of bikesharing users (Faghih-Imani & Eluru, 2015). However, Ramos et al. (2020) suggest that those personal characteristics predict the trip motive rather than a certain activity. This motive is connected to the trip's purpose, which is described above. In conclusion, structurally, individual characteristics predict the trip purpose or motive, which can be used to predict the destination.

Furthermore, socio-economic characteristics can be used as a kind of built environment characteristics. This has a place in the form of demographics in the framework. These characteristics are a summary of all the individuals who live in that place. It can help in determining and characterising the area. A remark is that it should be residential areas. Otherwise, there is no available data.

On the other hand, habits and preferences relate to people's perceptions. These are subjective reasons to choose a certain travel mode. For example, people with high satisfaction towards public transport are likely to choose public transport more frequently. People who have a strong preference for using public transit will use the connecting bikesharing more often too.

People's preferences for a certain mode can be determined by spatial locations (Chowdhury et al., 2020) or strong driving habits (Aarts et al., 1997). Driving habits act as a strong predictor in choosing the mode (Ramos et al., 2020; Vij et al., 2013). Such preferences can be used as predictors or dividers of particular user groups.

For SBRT systems, no literature has been found. However, NS Stations conduct a customer satisfaction survey of their users. In that survey, almost every respondent rated the OV-fiets higher than 7. This indicates that those users are happy to use the OV-fiets.

As a result, habits and preferences could be relevant determinants for bikesharing use. However, habits and preferences have less impact on destination choice but more on mode choice. In this research, the mode is fixed, which means it might be less useful as a predictor.

2.1.6 Summary

To summarise, built environment characteristics, mode choice attributes and individual attributes (socio-economic characteristics and habits and preferences) can help determine OV-fiets users and their chosen destinations. Individual attributes will help find the purpose of the trip, which is connected to the destination and to the preferences for the OV-fiets as a mode. Furthermore, trip attributes can help determine the system's catchment and reachable destinations. Table 1 shows the summary of the determinants and the influences on bikesharing as far as known.

Table 1: Summary of influences on one-way BSS and SBRT

Determinant	Influence on one-way BSS	Influence on SBRT
Infrastructure	More cycling infrastructure creates more usage	(?) constant for research
Topography	Hilly areas result in lower demand and an imbalance in trips	(?) constant for research
Land-use	Higher density and more diversity lead to higher demand	?
Destination	Specific destinations are a predictor of demand	?
Trip distance	Longer distance is less usage	?
Bicycle availability	Large-scale supply needed to generate demand	Demand is growing as supply grows
Weather	Comfortable cycling weather (sunny) correlates with higher demand	Comfortable cycling weather means higher demand. Although commuters are more flexible
Temporality	Highest demand during peak hours	Highest demand during peak hours. Difference in demand between week and weekend days
Alternative modes	Bikesharing is an extension of public transport and connects low-supply areas	(?) Bike-train relation acts as an extension of train station catchment
Socio-economic characteristics	Tendency to attract 'wealthy early adopters' as a result of providing service mostly in these areas	(?) Higher educated than average
Habits and preferences	Tendency to attract people with satisfaction towards or frequent use of public transport	(?) Satisfaction is high towards the OV-fiets

Based on these influences, a conceptual framework can be formulated showing the relation between the variables. Figure 6 shows the conceptual framework with the most important variables for SBRT systems and their destinations. Some variables are directly related to a destination, while others are related to individual users or to the mode choice. For example, distance to transport variables directly influences mode choice (when a bus stop is nearby, the bus is more likely to be used), reducing the probability of choosing an OV-fiets. Therefore, the variable might be important, but it is not directly influencing the destination choice of OV-fiets users. Individual characteristics are divided into socio-economic factors and individual preferences. The former are objective variables, while the latter are subjective variables of the person. While the first determines the purpose, and therefore the destination, of a person, the second is more likely to determine the mode.

The figure shows an interacting effect between destination and mode choice. From the literature, it was derived that both choices are dependent on each other. In this research, the mode is fixed, namely OV-fiets. Therefore, the mode choice is not a choice anymore. Only the consequences of that choice are possible to investigate. The probability of a destination would not change the mode anymore. Therefore, the relationship is displayed as a dotted line.

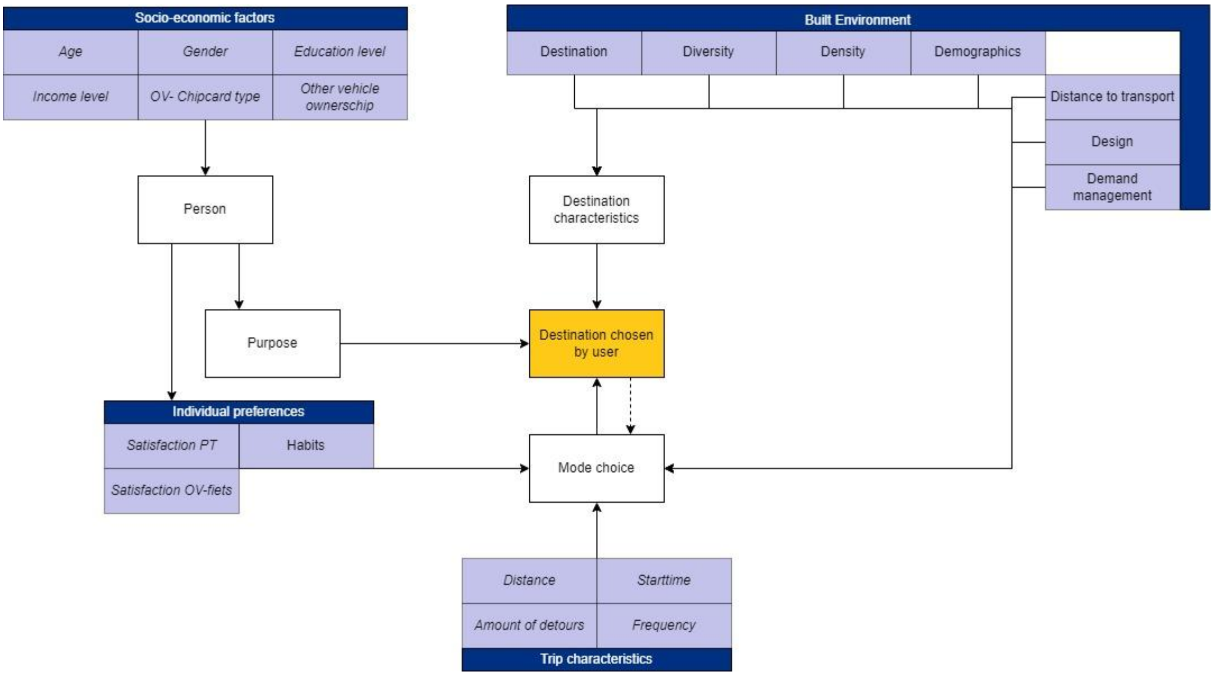


Figure 6: Conceptual framework of influences relating to destination choice

2.2 Destination choice modelling

The following section provides an overview of the literature regarding destination choice. First, the available models in this domain are described. Then, section 2.2.2 describes in detail the most suitable approach, limitations and assumptions.

2.2.1 Available methods

The literature describes two main approaches for destination (choice) modelling: prediction with origin-destination models and choice modelling. The former is based on the gravity model of Newton in combination with the four-step transportation model (FSM) (McNally, 2008) and the latter on the choice of an individual (Train, 2009). In a way, the first is based on attraction and production (focussing on all trips) and the second type is based on behaviour (focussing on an individual).

The gravity (and FSM) model is based on the demand for activity participation. Various traveller characteristics and the land use-activity system are evaluated, calibrated and validated to create attraction and production of zones. The main principles of this estimation are that the larger and closer the origin and destination (OD) are, the more trips will take place between them. This is similar to Newton's gravitation law, hence the gravity model's name. When attraction and production are known, a trip between the travellers' origin and destination will be created onto a transportation network. As a result, an OD-matrix can be compiled, in which destinations can be displayed as the absolute or relative amount of trips from origins. The model is moderately successful in the aggregate; however, it has failed to perform in most relevant policy tests and is better used to analyse a transportation system (Hensher & Button, 2008). Furthermore, gathering data for input and calibrating the models are expensive methods due to extensive surveys and traffic counts.

The gravity model is hardly used in the specific application of destination (choice) modelling. Little applications have been found, but one of them is by using open data to determine the destinations of travellers (Jin et al., 2014). Although the method was effective and low-cost, it was based on an existing OD-matrix. Other applications are found with spatial-interaction models (Fotheringham et al., 2001). In this study, individual trade-offs between spatial convenience and other factors were captured.

Although the method was suitable for these applications, it cannot be used in this research due to the lack of data and many unrealistic assumptions, as addressed by Fotheringham et al. (2001).

The gravity model and FSM models have several limitations, such as ignorance of individual trips and misrepresentation of overall behaviour as an outcome of a true choice process rather than as defined by a range of complex constraints that delimit choice (Hensher & Button, 2008). A shift to an activity-based approach (ABA) was described in the literature. The ABA was set as a "richer, more holistic framework in which travel is analysed as daily or multi-day patterns of behaviour, related to and derived from differences in lifestyles and activity participation among the population" (Jones, 1990). More concrete, travel decisions are activity-based, and an understanding of travel behaviour is secondary to a fundamental understanding of activity behaviour (Hensher & Button, 2008). The method focuses on behaviour and treats every traveller as an individual with choices based on characteristics rather than one big chunk of travel patterns. In this regard of context, destination choices can be better explained because choices are based on behaviour.

Discrete choice models (DCM) are used in many fields to model choice behaviour. Examples are studies on residential location choices in the field of geography, brand choice in marketing or the number of children in sociology (Hensher & Button, 2008). On top of that, DCM is widely used in transportation studies, for example, in mode choice, tourist explanation, and destination choice. Furthermore, it has applications within bikesharing literature (Faghih-Imani & Eluru, 2015).

The DCM has as the underlying mechanic that an individual person chooses between discrete alternatives (Simma et al., 2001). Relationships can be derived based on the characteristics of the decision-maker and the alternative. Estimated models can determine which variables are most important in influencing the decision-makers choice, based on many observations (Ortúzar & Willumsen, 2011).

As described in section 2.1, influences on bikesharing usage or potential destinations are identified. Those are identified for SBRT bikesharing, but the magnitude of the influence is unknown. Moreover, destination choice fits within the spectrum of behavioural choice modelling. Therefore, a more individualistic approach is needed, which can be found in the activity-based approach and discrete choice analysis. Therefore, the discrete choice model is more fitting for this research rather than OD-matrix estimation with FSM or other gravity models. As such, this study will use the DCM and, therefore, will be the focus of the following sections.

2.2.2 Discrete choice modelling

Discrete choice modelling and models are widely described in the literature. Multiple authors have thoroughly described the basics and details (Ben-Akiva & Lerman, 1997; Hensher & Button, 2008; Louviere et al., 2000; Ortúzar & Willumsen, 2011; Train, 2009). In the following sections, the most important concepts are described.

First, for the definitions, the DCM is a collection of procedures that defines the following elements:

1. *Decision-maker*

The unit of a decision can be an individual or a group of persons. It represents the actor who makes the decisions. Individuals face different choice situations and have widely different tastes. As aggregated choices are of interest, the differences between the individuals are important. Those differences can occur between the individual on an objective level (for instance, different age) or upon a more subjective decision level (for instance, different reasons to choose).

2. Alternatives

Each choice is made between a set of alternatives. The decision-makers environment determines the universal set of choices, and which alternatives any single decision-maker considers is the choice set. This choice set consists of every alternative the decision-maker can choose from so that those choices are feasible and known.

3. Attributes

The attractiveness of an alternative is depicted by the attributes. Alternatives can be homogeneous, meaning each alternative has the same attributes, and only the quantities of those attributes are different. On the other hand, alternatives can be heterogeneous, in which attributes are different between the alternatives. An alternative can be simplified to a combination of attributes, also known as the utility.

4. Decision rule

When different alternatives are presented, the choice defines which one is preferred. This is done via a decision rule: the internal mechanisms used by the decision-maker to process the information available and arrive at a unique choice. Four major categories of decision rules are the most commonly used in the literature: Dominance (when one alternative is better for at least one attribute), Satisfaction (each alternative serves a satisfaction criterion), Lexicographic rules (attributes are ranked based on importance to the decision-maker, the most important attribute creates the best alternatives) and Utility (multiple attributes define the alternative, creating an objective function. An utility function can be maximised or minimised to determine the choice). In transportation literature, utility maximisation is predominantly used (Bhat & Guo, 2004) and, therefore will be used as the decision rule in this research as well. The utility of an alternative cannot always be explained by the researcher, and there may be unobserved variables. Therefore, the utility is often displayed with a deterministic component and a random component (Kjaer, 2005).

The most basic and natural discrete choice model is the multinomial logit (MNL) model. This model is mostly used as a reference or to understand the structure of a DCM model. From the MNL model, other random-utility maximising discrete choice models are derived, focussing on relaxing the assumptions of the MNL.

Three basic assumptions underline the MNL model. The first assumption is that the random components of the utilities are independent and identically distributed (IID) with a type I extreme-value (or Gumbel) distribution. It means that no common unobserved factors affect the utilities of the various alternatives. This assumption is often not fulfilled in destination choice (Simma et al., 2001). The second assumption is that the MNL should maintain homogeneity in responsiveness to attributes of alternatives across individuals. This means that different persons react the same in response to attributes of alternatives. For example, some people can be time-conscious and will therefore experience a transfer between transportation modes more tedious than “laid-back” people. If such occurrences are not specified, the assumption is violated. The last assumption of the MNL is that the error variance-covariance structure of the alternatives is identical across individuals. It means that the same competitive structure exists among alternatives for all individuals. In other words, between two alternatives, the unobserved variables should not differ. For example, the level of comfort should not differ between the two route alternatives.

The three assumptions lead to the IIA property: the independence of irrelevant alternatives at the individual level. The property means that the ratio of the probabilities of choosing one alternative over another is unaffected by the presence of any alternative in the choice set. This allows adding or eliminating alternatives without the need for re-estimation of the parameters.

However, the IIA property is often not satisfied in reality, leading to other DCM structures being used. For example, nested logit (NL) models account for scale heterogeneity. NL allows alternatives to share common unobserved components among one another compared with a non-nested alternative. Another example is the mixed logit (ML) model, which allows for random taste variation via a certain probability. It means that the relative importance of attributes is different over the alternatives.

Moreover, spatial issues occur when modelling destination choice (Simma et al., 2001). Spatial issues occur on three levels: spatial dependency, spatial heterogeneity and spatial heteroscedasticity. Spatial dependency describes the presence of unobserved spatial factors, such as a beautiful landscape. Spatial heterogeneity means that the relationship between the dependent variable and independent variable varies across spatial units, which means that there might be no global relationship, but several local ones. The last issue means that the variance of the unobserved influences may be different across spatial units.

Dealing with these spatial issues in DCM is, as Bahamonde-Birke (2021) describes, not a "sexy" problem: its treatment is complicated, it is associated with large computational costs, and it requires the use of advanced econometric techniques (many of which are not easily available from an operational viewpoint), and the results are not particularly appealing (given that the focus of the analysis is mostly not set upon demonstrating the existence or quantifying the spatial correlation, but to establish causal relationships in data that happens to be spatially correlated). Therefore, spatial correlation has become a stumbling block for many modellers, who, for the sake of simplicity, prefer to ignore the issue (Bahamonde-Birke, 2021). To overcome these, Sener et al. (2010) describe that an NL-form model can be used to deal with spatial issues, in which the best structure is described as the generalised spatially correlated logit (GSCL) model, which can accommodate spatial correlation between alternatives. However, this model is difficult to estimate due to its complexity and lack of estimating software. Therefore, this research will simplify the spatial issues and aims to accommodate them as other research addresses those issues.

2.3 Conclusion

As a result of the literature review, potential determinants and influences on the use and destination choice have been identified, which can be used for SBRT systems. The OV-fiets user characteristics will cover socio-economic variables: age, gender, education level, OV-chipcard type, working situation, and individual preferences: travel frequencies, possible alternatives, satisfaction and motivations.

The discrete choice method is a well-established method in literature to determine the destination choice of OV-fiets users. It respects an individual's behavioural choice and uses utility maximisation. This means that relationships can be estimated based on individuals' characteristics and possible destinations, which can be used for destination choice explanation. This is not possible with methods like the gravity model, in which the biggest or closest destination will be chosen. In essence, DCM works on an individual or microscopic level and the gravity model focus on a macroscopic level. The microscopic approach fits better in this research due to the individual characteristics that are captured.

Several concepts of the DCM have been explained, and the assumptions and biases of models were mentioned. To overcome these, other model structures within the DCM family exist. The model used in this research should be determined based on the data and the definition of the individuals, alternatives and choice set. This choice is described in the methodology (chapter 3).

3 Methodology

This chapter describes the methodology which is used to answer the sub-questions. The main aim is to design a combined user characteristics and destination model, as described in section 2.2. The modelling approach for this is displayed in Figure 7. This figure visualizes the connection of the separate sub-questions as well.

This chapter is structured as follows: first, the available sources of data are described to show what is currently available and where gaps exist. Thereafter, a clustering analysis is done to select a set of rental locations for this study in section 3.2 to answer the first sub-question.

Having defined the scope of rental locations to be researched, it is necessary to collect user data through a survey. The design of this survey on user characteristics and destinations is described in section 3.3 to answer the second sub-question.

These survey results can be used to answer the third sub-question, which transforms the destinations to a spatial dimension, as described in section 3.4.1. Also, the built environment data preparation (red box in Figure 7) is described in section 3.4.2. Furthermore, the third sub-question is answered by the combination of the spatial destinations with the built environment, see section 3.4.3.

The results of the second and third sub-questions form the base of the fourth sub-question, in which the three models are formulated: the purpose prediction in section 3.5 (blue box in figure Figure 7), the destination model in section 3.6 (yellow box in figure Figure 7) and the combination model in section 3.7 (green box in figure Figure 7).

The last sub-question describes the implications and limitations of the results. This evaluation is described in the analysis chapter (section 4.6). All the sub-questions support answering the main research question, the yellow block in Figure 7.

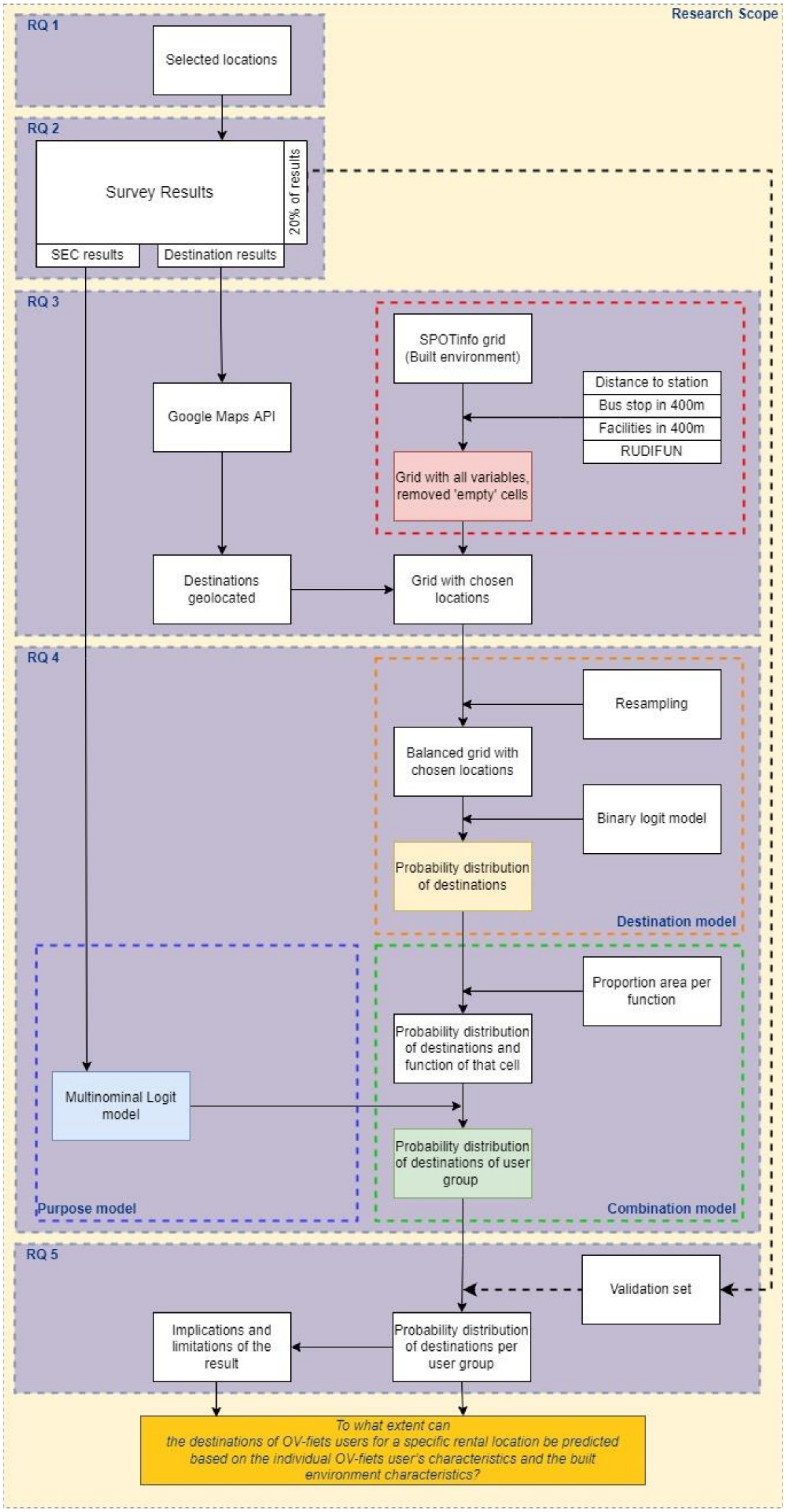


Figure 7: Full modelling workflow within the scope of the research (yellow area)

3.1 Available data

Various data types and sources are required for the models and location selection. The available data should be known to understand the modelling approach and other methodological steps. The available data also shows the exact data gap, which helps specify the survey.

Table 2 shows the source and a general description for each data set. Moreover, the table shows in which research question the data set was used. The sections below display a description per data set with NS Stations related data in section 3.1.1 and the spatial data in section 3.1.2

Table 2: Data sets and sources

Name	RQ use	Description	Source	Spatial unit	Year
Train stations	1	Train stations of the Netherlands	NSSt: KIS10 and Prorail	Point feature	2019
Access/Egress train station	1 & 4	Egress mode information of train stations	NSSt, Natransport	N/A	2019
OV-fiets locations	1	Information on OV-fiets locations	NSSt, Services	N/A	2019
OV-fiets bookings	1	OV-fiets bookings per location, on weekdays in 2019 and 2020	NSSt, Services	N/A	2018, 2019, 2022
OV-fiets user characteristics	2	OV-fiets user characteristics as surveyed once	NSSt, Onderzoek	N/A	2019, 2020
Function floorspace	3	Function of floorspace area building block	RUDIFUN	Building block (m^2)	2022
Grid	3	Grid cells with statistical, spatial and demographic data	Spotinfo/Argaleo	ha	2022
Facilities	3	Point data with BAG definition of company or governmental facility	Spotinfo/Argaleo	Point feature	2022
Bus stops	3	Bus stops with active bus service	Spotinfo/Argaleo; OpenTransportMap	Point feature	2022
Road network	3	Roadlines of the street network	OpenTransportMap	Line feature (m)	2022

3.1.1 NS Stations related data

From NS Stations, multiple datasets were summarized and combined into five sets, two related to train stations and three associated with OV-fiets.

Train station related data

The *train station* dataset consists of characteristics and geolocations of train stations in the Netherlands. In this dataset, the information is twofold. The stations with train services operated by NS Reizigers (Dutch Railways) have extensive data on train passenger numbers, access and egress mode shares and train passenger characteristics (with 288 train stations). This information is based on check-in and check-out data from the OV-chipkaart in combination with user surveys done by NS Stations. Some train services are shared with or operated solely by other train companies than NS (110 train stations). For these stations, the data are incomplete due to competition protections. This train station dataset is combined with the geo-locations and other spatial characteristics of the stations, which are available by ProRail (ProRail, 2022). The dataset is only available for 2019, which means that other datasets that are combined with this one should also be of 2019 for optimal reliability. 2019 is stated as the most reliable year as it is the most recent full year without covid-19 pandemic issues.

Furthermore, the assumption is made that the characteristics used from this dataset are stable over time, while in reality, characteristics might change.

The *Access/Egress train station* dataset consists of more detailed data on access and egress of train stations operated or shared by train services of NS. The dataset is provided by NS Stations (NS Stations, 2022a) and contains several variables. As the variables are not all collected in the same way, they may differ in completeness. For example, the cycling distance from a train station is only known for 28 stations, while the full dataset consists of 246 stations. Therefore, a balance should be found between the available variables and the train stations that are analysed. These two train station-related datasets are the basis for the station selection procedure.

OV-fiets related data

The *OV-fiets locations* dataset consists of all the OV-fiets rental locations and their characteristics, for example: the type of rental location and whether the rental location is connected to a train station. Furthermore, a list of the maximum available bicycles is provided by NS Stations, which can be combined with these rental locations' datasets. This list is only available for 2019, with the same assumption made with the train station dataset, which is that the data are still reliable for the current situation. This dataset is used for the station selection procedure.

The *OV-fiets bookings* dataset contains all individual bookings per OV-fiets rental location. A booking generally concerns one bicycle. However, bookings might contain two bicycles (which is the maximum for a person to rent) or zero bicycles (in the case of an error or maintenance). Per booking, two timestamps are collected, one for the check-out and one for the bicycle check-in. The timestamps are stored per minute, and the dataset only contains bookings for which the origin and destination are the same. This means that bicycles brought back to other locations (which results in a fine for the user) are not considered. The dataset is provided by NS Stations, covers 2019 and is the same as used by Wilkesmann (2022). Therefore, the dataset is already cleaned and reduced, as described in his research.

That research filtered only stations operated by NS, which had more than 10 bicycles available. As a result, 48 out of 313 rental locations were considered, which still represents 75.5% of all the bookings. The variables from this dataset can be used in the station selection procedure.

The *OV-fiets user characteristics* dataset is the last OV-fiets-related dataset. These data are a revealed preference dataset regarding the characteristics of OV-fiets users. The data were gathered via an online survey by the NS-panel, an opt-in customer panel of NS Reizigers. The survey (with a small 2000 respondents) is held every year, with the most recent one in early 2020, in which the Covid-19 restrictions were initiated. The survey covers users who are familiar with OV-fiets, resulting in limited characteristics of users per week, weekend, in general, or per travel purpose. Although this data is the only OV-fiets user characteristics dataset of NS Stations, the dataset might not be representative for all OV-fiets users due to the use of the NS-panel.

Rental location data fusion

The combination of train station datasets is necessary to derive statistical variables from them for further steps in this research. The matching attributes are the names of the train stations and rental locations. The resulting dataset consists of statistical data of each rental location with booking data of the corresponding location. This will make it possible to aggregate booking data per location and generate trip statistics like the average duration of trips per location or the number of trips per location.

3.1.2 Spatial data

Spatial data concerns all the data related to the environment and, most of the time has a geographical component. Examples of spatial data are land use variables, points of interest, accessibility variables and attractiveness. In essence, it covers all the aspects of the 7D-built environment framework as described in section 0. For this research, the spatial data from multiple sources were combined. The sources and sets are described below.

Argaleo/Spotinfo related data

The *mozaiek* (hereafter also described as the *grid*) dataset consists of a 100x100 meter grid cell set with 115 environmental variables (Aalst, 2021). The set is filled with data from several governmental sources like the BAG (basisregistratie adressen en gebouwen; registration addresses and buildings, (Kadaster, 2018)), CBS wijk en buurten (CBS boroughs and neighbourhoods, (Peppel, 2020)) and BGT (basisregistratie grootschalige topografie, large topographical registration, (Kadaster, 2022)). The set is available in the study area from Argaleo and Spotinfo, companies which are specialized in combining governmental spatial and administrative data.

The set consists of administration variables (e.g. postal codes, municipality, province, water authority etc.), physical variables (e.g. soil heights, percentage cell of particular land use), function details (e.g. the percentage of a cell used for sports or events), infrastructure variables (the presence of certain infrastructure structures like roads or railways), building variables (e.g. the number of addresses, amount of buildings with BAG registration on offices, type of house, housing values) and demographic variables (i.e. the percentage of inhabitants of a certain age, household size). The descriptive statistics of this dataset are displayed in Appendix A in section 8.1.1.

The dataset is the base layer unto which other spatial datasets are joined. These variables are used for the analysis. A detailed description of how the datasets are combined can be found in section 4.1.3.

RUDIFUN database

The second spatial dataset is the RUDIFUN database (Harbers et al., 2019). This set consists of areas with information on floor area per function as described in BAG and calculated urbanity variables like mxi (mixed land-use index), fsi (floor space index) and osr (open space ratio). The different function classes can be found in the BAG and are displayed in Table 4. The RUDIFUN dataset is available per province and can be found on building blocks, neighbourhoods, wards and municipalities.

Table 3: Description of the function classes (Kadaster, 2018)

Function class	Description
Residential	Use function for living
Prison	Use function of a prison
Meeting	Use function for the gathering of persons for art, culture, religion, communication, childcare, providing consumptions for on-site use or watching sports
Healthcare	Use function for medical examination, nursing, care, or treatment
Industry	Use function for the commercial processing or storage of materials and goods or for agricultural purposes
Office	Administration usage function
Accommodation	Use function for providing recreational accommodation or temporary accommodation for persons
Education	Use function for teaching
Sport	Use function for practising sports
Shop	Use function for trading materials, goods, or services

For this research, only the RUDIFUN of the provinces of the study area will be used, utilizing only the building block level. The building block level can match the detailed level of the grid. As the RUDIFUN describes the built-up area, it complements the mozaiek, which focuses more on the non-built-up land use area. Furthermore, the mozaiek only describes the number of addresses belonging to a function class, not the areas. Therefore, the function classes of the RUDIFUN are more relevant for this research purpose.

Facilities

The *Facilities* dataset consists of point features of places registered in the NHR (Handelsregister, commercial register, (Kvk, 2022)) and public places like police and hospitals. The dataset contains the address, type and other administrative data of these locations. The dataset is available for the case study areas from Argaleo and Spotinfo.

This research uses the facilities dataset to determine points of interest in an area or grid cell. Due to the addition of facility types, choices can be made on which facilities to consider in the analysis.

Bus stops

The *Bus stops* dataset are point features of bus stops. The dataset is simple in that it only contains the geolocation, the *datakey* and the name of the stop. To enrich the dataset, it is combined with data from the opentransport map (OTM, 2021), in which current regular public transport services are displayed. Again, the bus stops point features are provided by Argaleo and Spotinfo.

The research uses this dataset to determine the available public transport alternatives. Since a bus service is often bi-directional, most stops have two feature points and quays. Therefore, the dataset is reduced based on stop name and geographical location. As a result, a datapoint acts as an entrance or exit of the public transport system.

Road network

The *Road network* dataset are the line features which resemble the roads. The dataset is available per region from Open Transport Map (OTM, 2021). The dataset includes characteristics of roads which are used to distinguish different road types or unidirectional roads.

The dataset is used to construct network distances from the rental locations to a grid cell in this research. The focus is on cycling, which means that highways and other roads on which cycling is not allowed are removed from the dataset. In that way, those routes are not possible anymore and only allowed roads are considered.

3.2 Station classification

The first sub-question focus on rental location selection. This selection is based on a classification, in which the most common statistical method is cluster analysis. Cluster analysis is an explorative method that groups variables to ensure that the observations within a group are similar but different to observations in all other groups (Everitt et al., 2011).

3.2.1 Clustering theory

There are two types of clustering techniques: hierarchical and non-hierarchical (Everitt et al., 2011). In hierarchical clustering, the final number of clusters is not fixed, whereas this number is fixed in non-hierarchical clustering. Around the pre-specified seed points, the clusters are generated. K-means clustering is a common method for non-hierarchical clustering. This method calculates the similarity between the cluster centres (the seed points) and the objects, after which it assigns the object with the highest similarity to the corresponding cluster (Schubert, 2019). Two main assumptions should be taken into account when using k-means clustering:

- The first assumption of the method is that there should be no multicollinearity among the variables (Everitt et al., 2011). Multicollinearity among variables can be problematic as they might capture the same key characteristics. This could lead to an over-weighting of this characteristic and, therefore, to unfavourable solutions.
- Furthermore, there should be no outliers. Outliers can distort the representativeness of the results if they appear as single clusters.

After clustering, the Elbow method, silhouette method or gap-statistic can be used to determine the number of clusters in the k-means technique (Everitt et al., 2011). The basic idea behind the cluster technique is to define clusters such that the total intra-cluster variation is minimised, also called the total within-cluster sum of squares (WSS).

The Elbow method looks for the 'elbow' (or largest curving point) in a WSS-graph. This elbow occurs at the point where the difference in WSS is the largest, and consequently, this is the optimal number of clusters (Everitt et al., 2011).

The silhouette method determines a point's similarity to other points in its cluster (Lletí et al., 2004; Shahapure & Nicholas, 2020). Global measures of the silhouettes are given by averaging them per cluster. With that, the average silhouette width can be calculated. A larger average silhouette width means a better fit of the data and the optimal number of clusters.

Finally, the gap statistic can be used (Tibshirani et al., 2001). The gap statistic is formally a hypothesis test. It constructs a reference distribution by computing a box aligned with the data's principal components and then fills it with uniform random numbers. Then a null model of a single cluster is assumed, and it is rejected in favour of a *k*-clusters model (k>1) if evidence for any such *k* warrants it (Lletí et al., 2004). In the end, the smallest gap-statistic indicates the best amount of cluster to be used.

3.2.2 Station clustering

The OV-fiets rental locations have variables which can be categorised onto three levels:

- Bicycle parking characteristics. These variables relate to the bicycle parking itself. For example, variables like the number of bicycles, amount of OV-fiets trips or bicycle parking type.
- Train station characteristics. These are variables applicable to the train station, for example, the number of passengers or Prorail and NS typology.
- Train station environment characteristics. These are characteristics of train stations for passenger behaviour which are defined by the location of the station. For example, the attraction/production share of a station, duration of OV-fiets trips, travel purpose or catchment of the train station. The variables have a direct influence created by the built environment. For instance, a train station in an urban area has a different attraction/production share than a train station in the city centre. The urban area station is a more production station (more houses), than the city centre station (more facilities, therefore, more attraction).

These variables can be used in the cluster analysis. The selection of these variables is an iterative process as it is not known which variables work best in the cluster analysis. In this iterative process, several conditions or criteria apply:

- Small clusters give little knowledge of the whole station set and should be avoided. Thus, variables that create clear outliers should be excluded.

- High correlating variables should be avoided as those will exaggerate cluster groups.
- Complementing variables should be avoided for the same reason as correlating variables. For example, the variable trip purpose lists three possible answers. These values complement each other and will correlate, which is not always visible in a one-to-one comparison. In short, removing one level of such variables avoids multicollinearity issues.

When the station clustering is finished, a selection should be made on which station to use in the analysis. Stations should be selected to get the broadest view of the station cluster. Although this might reduce the difference between the clusters, it helps to understand each station cluster. The station closest to the station cluster centre (mean) and the two stations furthest apart within the station cluster are chosen for this.

As an additional selection condition, chosen stations should have enough trips in the current situation. The station cluster analysis is done with data from 2019, as 2019 was the last full year before the Covid-19 pandemic. It is assumed that these data represent the research circumstances better than more recent data. However, there might be differences between the situation then and now. Therefore, the selected stations were checked for the number of current trips. When the station does not reach the stated 20.000 trips per year, the station is not chosen, and the next station following the previous criteria will be used.

3.3 Survey design

Since not all data are currently available, the remaining needed data should be gathered. Therefore, the second sub-question was formulated. Data gaps exist for the characteristics of OV-fiets users and the destinations of the OV-fiets users. This data can be gathered with a survey. The results of the survey can be used in the user characteristics and destination models.

Surveys gather choice data, as respondents make their choice among several alternatives per statement. Two types of choice data have emerged as the primary sources of choice response (Hensher et al., 2015). These are revealed preference (RP) and stated preference (SP). RP data refer to situations where the choice is made in real market situations. The data have high reliability and give an overview of the historical choices (Louviere et al., 2000). SP data refer to situations where the choice is made between hypothetical scenarios. This is useful in the case of examining the choice between existing and new alternatives. RP data gathering is needed in this research as only the current alternatives ('the world as it is') will be researched. As stated in the problem in section 1.1, these choice sets do not exist yet.

3.3.1 Revealed preference data gathering methods

The literature in transportation studies describes many ways to gather revealed preference data. Travel data of actual trips of bikesharing systems can be used as an RP dataset. Faghih-Imani & Eluru (2015) were able to use actual travel data as RP data due to the availability of GPS in their analysed system. However, as OV-fiets do not have GPS trackers due to privacy, other ways should be used. These different ways are often via surveys. With survey samples, an RP dataset can be obtained. Surveys can be held online (Adnan et al., 2019; Raux et al., 2017), face-to-face (NS Stations, 2021), or both (Ma et al., 2020). Furthermore, respondents can be asked based on use, or a (representative) panel can give their opinion about the product (McGuirk & O'Neill, 2016). For example, Raux et al. (2017) used travel diaries of users of their system, and Adnan et al. (2019) spread their questionnaires via online platforms.

For this research, a mixed method is used: face-to-face surveys are supplemented by online surveys for users who lack time to fill in the survey directly. The reason for this is that the face-to-face method will result in higher response rates, given that each user can be approached. Second, by directly filling in the survey, the participants have recent knowledge about their visited (or to visit) destination and experiences. However, returning OV-fiets users probably have the main mode (train) left to travel. Therefore, those users might have planned their bicycle trip with a connecting train, leaving little time to fill in the survey. To catch those users, an online variant of the survey is available, which can be accessed with a personalised QR-code on small flyers.

3.3.2 Questions

The questions have several goals to identify the individual and their destination. Therefore, different question groups within the survey can be distinguished:

- **Socio-economic data** – questions about the socio-economic situation of the respondents make it possible to find out who the users of OV-fiets are and their characteristics. Questions included in this cluster are related to gender, age, income, education level, OV-chipcard type (and subscription), but also vehicle ownership to determine alternatives for the respondent.
- **Preferences and habits** – questions about preferences and habits of the respondent towards OV-fiets and PT in general. Questions will cover the respondents' habits, frequency of OV-fiets travelling and satisfaction towards PT and OV-fiets. This information can distinguish users better, although these questions are more subjective to the user than the user's facts.
- **Trip data, destination and motivation** – questions about the last OV-fiets trip of the respondent to provide information about their whereabouts when using an OV-fiets. These characteristics can be used in the analysis to determine the destination. Questions that contribute to finding these characteristics aim towards a specific destination, rental location, travel time, travel distance, amount of detours and purpose of the trip.
- **Coronavirus** – a section of the survey will be about the coronavirus. This survey section aims to find if the respondents now travel differently or have travelled differently with OV-fiets compared to before the pandemic. These questions assess if the research is still representative in a fully non-covid situation. The covid-19 pandemic created a (temporary) change in travel behaviour (de Vos, 2020). Therefore, answers could be different now than in a pre-pandemic situation. Due to this research schedule, the covid aftermath cannot be avoided. Thus, the influences of the pandemic should be described at least.

3.3.3 Required response

The required respondents need to be gathered and approached. Not every OV-fiets user can be reached or can cooperate in the research. Therefore, a sample derived from all the OV-fiets users is needed. This sample should have an appropriate size to represent all OV-fiets users. A sample that is too big risks that the study becomes too complex, although the results are more accurate. However, sample sizes that are too small have a probability that outliers and anomalies are overrepresented in the sample, influencing the results. To determine a correct sample size, a method based on z-scores is often used in literature (Naing, 2003; Qualtrics, 2021). Table 4 shows a few population sizes with appropriate sample sizes as an example:

Table 4: Example of sample sizes at specific population sizes

Population size	Sample size
100	80
500	218
1000	278
5000	357
100000	383

As the research aims to fill the gap in identifying OV-fiets users and their chosen destinations, the target population would be all OV-fiets travellers (with 5.3 million rents), which leads to a sample size of at least 383 users. However, at the OV-fiets rental locations, the number of users and rentals are very different. Some locations have 100 rents per year, while others have more than 100 rents per day. Therefore, it seems that there is heterogeneity within the OV-fiets user group, meaning that the target population might not be the 5.3 million rents.

Due to practicality and budgetary reasons, not every station can be surveyed. Rental locations with 100 rentals per year require a sample size of 80 (Table 4) to be reliable, which is 80% of all the rentals at that location. This means that a surveyor should always be present at the rental location to catch the OV-fiets users. To avoid this, the scope is tightened to focus on rental locations with >20.000 rentals per year. This amount still represents 79% of all OV-fiets trips. But, even for locations with 20.000 rentals a year, the amount of trips averages to 54 rentals per day. In order to reach the required sample size, even with a 100% response rate, it would be required to hold the survey for 8 days per location. Given that this is both unrealistic and impractical, it was decided to use multiple rental locations. To do so, the stations were classified as explained in sub-research question one in section 3.2.

Finally, other constraints of the survey are fieldwork times. To survey effectively, the period with the highest probability of encountering an OV-fiets user was used. Therefore, the survey was held at times with the highest amount of trips. While there is no information yet, it is more useful to focus on the biggest share of users. Additionally, the survey will only be held on weekdays, as defined in the scope.

3.4 Data combination

The third research question aims to locate the found addresses and places of the survey. Furthermore, to combine those dots with the built environment data. The modelling flow is displayed in Figure 8.

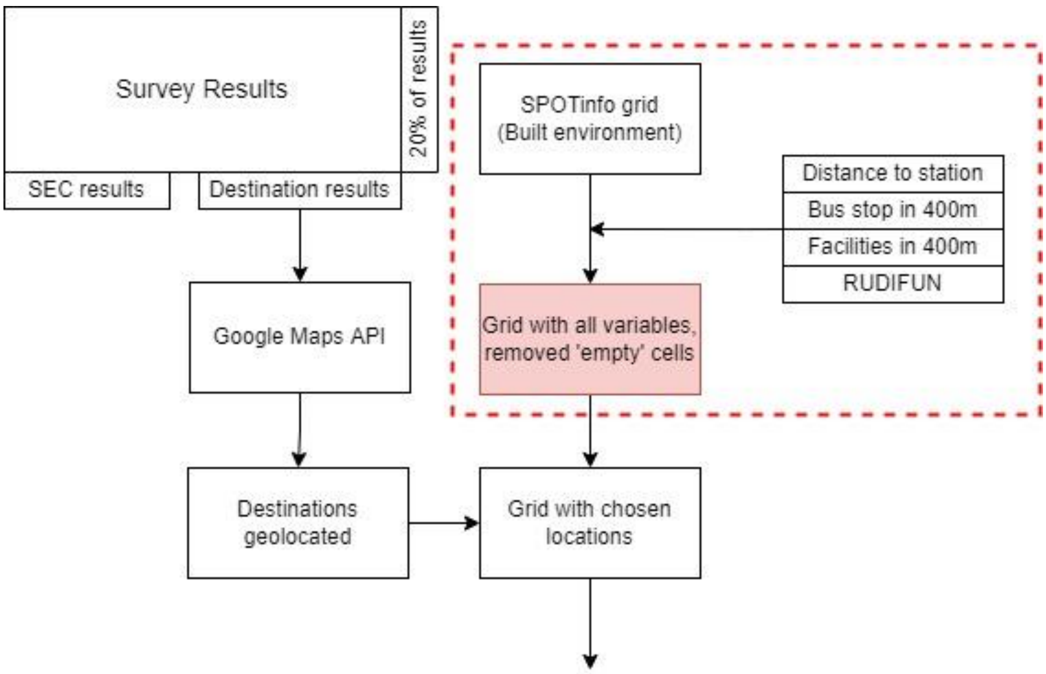


Figure 8: Modelling flow of sub-question 2

3.4.1 Geolocating collected destinations

The destinations are identified in the survey. These destinations are flat text and consist of either postal-code locations, addresses or specific destinations (for example, “Ikea”). The locations should be geocoded to be used in the analysis. To do so, locations can be found using Google Maps. Google maps is an internet tool which can find the geo-locations of a certain text input. When users want to do this several times, an API can be used, known as the Google Maps Places API (Google Maps, 2022b).

The Places API can convert text into a latitude and longitude of a location. A benefit of using the Google engine is that the input can be of all types, so all the types obtained from the survey can be used with the same API. Furthermore, the API can restrict a search area for ambiguous locations. This is a benefit for this research as survey respondents can fill in ambiguous locations like “Ikea”. However, as the sampling location is known, the exact location can still be derived.

A disadvantage of using the Places API is the output. The API can output multiple locations for each call. Then, manually, the best fitting location can be selected. However, it is impossible to compare each output location in an automated way (and reduce manual work). Therefore, the assumption is made that the first result is the right location. The first location corresponds with the first location that Google Maps will show with a manual search. This assumption will be checked with a few samples and compared with the manual output in Google Maps.

3.4.2 Data combination

The data sources should be combined to be able to use them in the research, the red box in Figure 7. The detailed flow is displayed in Figure 9. The following sections describe how this is done.

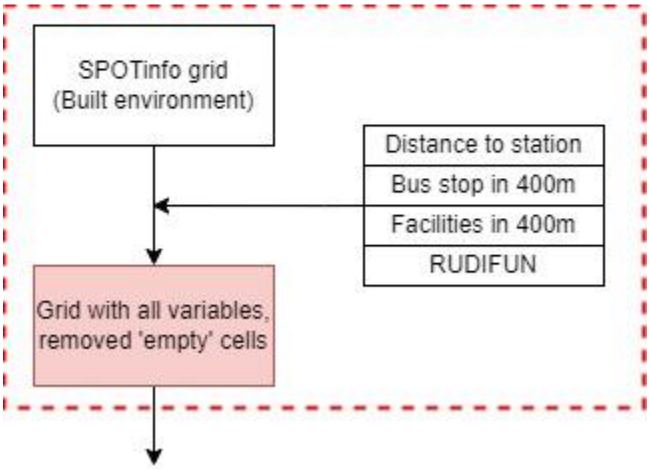


Figure 9: Data combination flow

Spatial data combination

The spatial data are joined on location and are aggregated to the spatial unit of this research. In literature, destination choice models or other spatial analyses are estimated and applied at different units, zones or postal codes. Those units are then used as modelled choice alternatives is usually those units (Bekhor & Prashker, 2008). Traffic analysis zones (TAZs) are used in transport trip and prediction models (Ortúzar & Willumsen, 2011). They are commonly defined as postal code areas or building blocks (Martínez et al., 2009). Other methods of analysis are via uniform formats such as degree-based grid cells (Gao et al., 2021) or single destination points (Faghih-Imani & Eluru, 2015). Krykewycz et al. (2010) used 10-meter by 10-meter grid cells with an aggregated function and, additionally, single destination points for special landmarks or locations.

In this research, the chosen destination can be anywhere theoretically, making a choice set infinite. The built environment does not adhere to administrative units such as postal codes or alternative units such as TAZs. The BE can be stretched out over multiple of these units, or BE characteristics can be very diverse within such units. Therefore, uniform formats such as square grids are more advantageous as these are unbiased. Furthermore, with uniform units, the unit size can be controlled. Due to the practicality of this method, the availability of 100 x 100m grid cells, and the availability of demographic data in this format (vierkantstatistieken (CBS, 2020)), 100 x 100m grid cells are chosen as the spatial unit.

Moreover, this cell size is appropriate to aggregate the functions (e.g. land-use and density) without losing too much detail (Krykewycz et al., 2010). For example, if we have a building block with a certain amount of layers in the RUDIFUN database, then this information might be lost when aggregating to postal code-4 or postal code-6. This is disadvantageous because the number of layers might be an attractor for OV-fiets destinations.

To combine the grid cells with the RUDIFUN-data, a proportioning method, as proposed by Asadi et al. (2022), is used. The function respects the intersected area of a building block with a grid cell:

$$S_i^l = \frac{1}{100^2} \sum_{j=1}^{n} \left(\frac{S_{ij}^l}{B_j} \cdot a_{ij} \right) \tag{13}$$

Where S_i^l is the proportion of the polygon feature S in the cell i; S_{ij}^l is the area of the S_i^l in building block j, and a_{ij} is the area of the building block j that intersects with cell i.

These values are then used to re-calculate the mxi and fsi of the cells, which indicates the ratio of housing function areas to all other functions and the total floor space of buildings in that area, respectively (van den Hoek, 2008). As a result, the grid cells consist of proportionate function floor-space area, fsi and mxi values.

The next step concerning this grid is adding the bus stop and facilities points of interest (poi). The radius of reaching these poi's is 400 meters, as explained in section 2.1.3. The 400-meter range is based on the willingness to walk to a public transport access point and the distribution of BSS docking stations (Gu et al., 2019b; van Soest et al., 2019; X. Zhang et al., 2021). Although this might be different for SBRT, this assumption is made. The grid cells will be buffered to a 400-meter radius and then spatially joined with the poi's. This will be done with built-in tools of the QGIS software (QGIS Development Team, 2009).

At last, the distance to the rental location should be added to the grid cells. This calculation is done with the network analyst of ArcGis (ESRI, 2022). With this tool, the network distances from the rental location to the centroid of a grid cell are calculated via the road network. The tool uses the well-known Dijkstra algorithm to calculate the shortest route (Dijkstra, 1959). The network distance is more accurate for this research than a Euclidian or Manhattan distance, since bicycles will follow the street network. Calculating distances via the network will ensure that barriers, like rivers or railway lines, are taken into account. The road network is available from Open Transport Map (OTM, 2021).

3.4.3 Combination of spatial data and destinations

A spatial join is used to combine the grid cells with the geolocated destinations. This kind of data join is based on the spatial component of a data attribute. Tools to this procedure are available in GIS software such as QGIS, which was used for this research (QGIS Development Team, 2009). As a result, the grid cells contain information on whether the cell is chosen. If multiple respondents answer with the same destination, the cell will is multiplied. In other words, the weight of that cell is doubled.

Furthermore, if a chosen destination is outside the extent of the built environment grid, the respondents' destination is removed from the destination model analysis.

3.5 Purpose modelling approach

The first model is the purpose model, the blue box in Figure 7. This model's goal is to predict a user's trip purpose based on socio-economic characteristics and habits. A user trip purpose can be seen as a non-hierarchical choice between several alternatives. As the theoretical context explains in section 2.2, a discrete choice model serves this purpose. The basic and most flexible type of model form is a multinomial logit model. With this model type, the IIA property must hold. The modelling flow of the purpose model is displayed in Figure 10. The following subsections describe the components and structure of the multinomial logit model in detail.

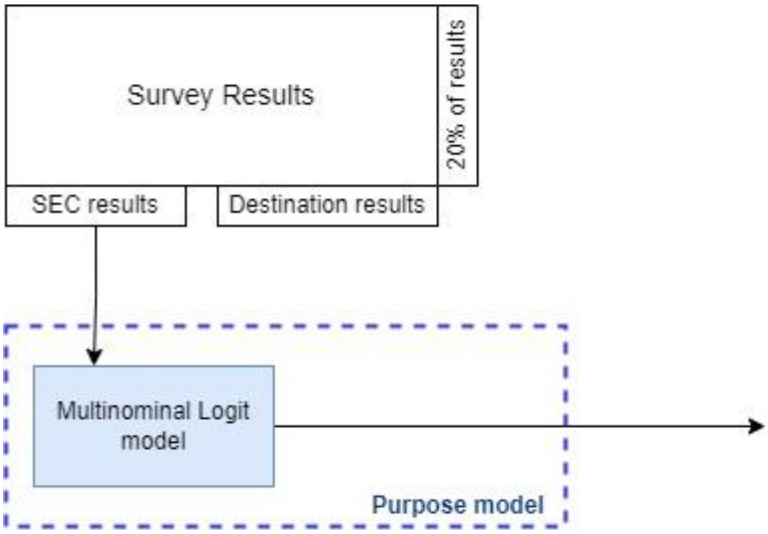


Figure 10: Purpose model flow

3.5.1 Input data

The first component of the MNL model is the input data. The data were gathered with the survey, which is described in section 3.1. A part of the survey consists of questions about the respondent's socio-economic characteristics, preferences, and habits. With this information, the input variables can be made.

3.5.2 Dependent variable

In discrete choice modelling, multiple outcome 'choices' are the alternatives and can be denoted as the dependent variable. The dependent variable of the purpose modelling is the trip purpose. The survey contained a question about trip purpose, so that a relationship between other variables and the trip motive can be modelled.

However, the MNL is burdened with the IIA property. The assumptions from the IIA can hold with trip purpose as the dependent variable. The alternatives are mutually exclusive, as a trip purpose does not overlap with other trip purposes, indicating that respondents can react similar to them. Furthermore, it is assumed that the selected independent variables are the only ones influencing the choice alternative. This assumption can hold because many independent variables have been gathered. Therefore, the unobserved factors will not differ either; for all alternatives, they will be the same.

3.5.3 Independent variables

Independent variables explain the dependent variable. To select the right independent variables for the model, the same selection process is applied as was used for the station cluster analysis (section

3.1). In short, variables that occur with low occurrences in the data can create outliers. Outliers should be removed because they create unreliable outputs. Therefore, those variables should be excluded. Secondly, high correlating variables should be avoided as those exaggerate the prediction. Lastly, complementing variables should be averted to avoid multicollinearity issues.

The independent variables were converted to binary variables. This means that each answer category (or bin) for each question is translated to 1 (true) or 0 (false). The use of binary variables helps in interpreting the model output. An output coefficient is then directly referenced to a characteristic instead of a reference category, which is needed when discrete variables are used. A second benefit is that adding or removing variables in the model is easier. This can improve the model's fitting, as fitting is an iterative process.

The answer categories with less data (the category has no or few responses) are filtered out. This means that certain answer categories (such as 'retired' for the working situation, which probably has a low response rate) cannot be fitted in the model. Although this would be the case for discrete variables anyway, it is a disadvantage for continuous variables. These could be extrapolated when fewer data are available, but this opportunity is lost when translating them to dummy variables.

3.5.4 Model structure

Based on the random utility theory, the utility of an alternative can be described as a probability function (Kjaer, 2005):

$$U_{iq} = Prob_{iq} \tag{1}$$

Where, U_{ig} is the utility of alternative i given by an individual q, and $Prob_{iq}$ is the probability of the alternative i chosen by individual q

Then, this can be configured to a function with a systematic part and a random part (Ortúzar & Willumsen, 2011):

$$U_{iq} = V_{iq} + \varepsilon_{iq} \tag{2}$$

Where, V_{iq} is the systematic component of the utility, reflecting the perception of an average individual q on alternative i, and ε_{iq} is the random factor, reflecting the non-observed behaviour of the individual q

The factor V_{iq} can define multiple attributes, which can therefore be described as the sum of explanatory attributes (Louviere et al., 2000):

$$V_{iq} = \sum_{k=1}^{K} \beta_{ik} x_{ikq} \tag{3}$$

Where, V_{iq} is the level of utility that alternative i provides to individual q, x_{ikq} are the explanatory variables, reflected by the attributes, and β_{ik} is the utility parameter, reflecting the weight of each variable x

In result, the utility function can be written with more simplicity as follows:

$$U_i = V_i + \varepsilon = \sum_n \beta_n * x_{in} + \varepsilon_i \tag{4}$$

Where U_i is the total utility associated with alternative i, V_i is the observed utility associated with alternative i, ε_i is the random error component, β_n is the variable to estimate associated with attribute x_n, and x_{in} is the value of attribute x_n for alternative i

Due to the IIA assumptions of the MNL, the error terms can be eliminated to create a closed-form formula. This allows for short computational times and explains the popularity of the model structure. This results in the probability function of the MNL model, which is described as follows:

$$P_{iq} = \frac{\exp\left(V_{iq}\right)}{\sum_{j}^{J} \exp\left(V_{jq}\right)} \tag{5}$$

Where, P_{iq} is the probability for individual q of alternative i, J is the set of alternatives and V is the utility level of the alternative.

The maximum likelihood estimation (MLE) is the most common way to estimate a MNL model (Louviere et al., 2000; Ortúzar & Willumsen, 2011). The method will maximise the log-likelihood, which is described in equation 6:

$$LL = \sum_{q=1}^{Q} \sum_{j=1}^{J} f_{jq} ln P_{jq} \tag{6}$$

Where, LL is the log likelihood function that should be maximised; f_{jq} is a dummy variable that is either 1 if alternative j is chosen or 0 otherwise; P_{jq} is the probability function of the various alternatives

Maximising the equation with respect to the utility parameters β is an iterative process. For which an initial starting value for the utility parameters can be used. The iterative procedure is often continued until a certain tolerance level is reached and the optimal values for β have been found.

3.5.5 Model performance

The literature describes several manners to evaluate model performance. The most common way to compare models and evaluate the goodness of fit is using the Akaike Information Criterion (AIC). The criterium is beneficial when the number of classes is unknown and can thus be considered using these criteria (Louviere et al., 2000). The AIC can be calculated with the following equation:

$$AIC = 2K - 2\ln\left(LL\right) \tag{7}$$

Where LL is the maximised value of the Log-likelihood function at the estimated parameters and K is the total number of parameters in the model.

The AIC gives a penalty to models which use more parameters for their fit. This encourages a reduction in the number of parameters while fitting the model. A better fitting model results in a lower AIC-value. When comparing models, the one with the lowest AIC should be chosen. The AIC is widely used in literature and is incorporated in model-fitting software. Therefore, the AIC is used in this research.

3.6 Destination modelling approach

The second model, or destination model, aims to find the relative probability that a grid cell is chosen as destination for OV-fiets users. This is the yellow box in Figure 7. A grid cell can be classified as “chosen” or “not chosen”, indicating a binary structure. Because this resembles a ‘choice’, a binary logistic regression model is used for this modelling part (Train, 2009). The modelling flow is displayed in Figure 11. The section is displayed as follows: the input data procedure is illustrated in section 3.6.1. The model and its requirements will be described in sections 0 and 3.6.4. Finally, the model performance method and indicators are discussed.

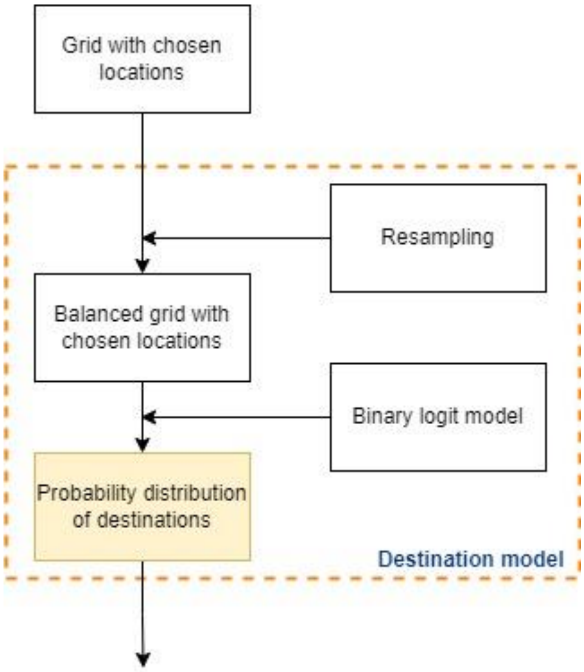


Figure 11: Destination model flow

3.6.1 Input data

The input data of the destination model consists of the grid created from the spatial built environment data and surveyed destination data, which is described in section 3.4.3. This creates the grid with chosen locations.

The input grid consists of many locations around the rental location in which some locations have been chosen by respondents. The survey is targeted to find > 400 samples per cluster to be representative (3 stations). But those 400 locations are way less than the number of grid cells or possible destinations. The grid has a radius of 10 km around the train station and a resolution of 100 x 100 meters, resulting in around 31000 cells. Even if, for instance, 10% of non-built-up cells are removed, it is still around 220 times larger than the sample size. Therefore, it can be stated that the dataset is "unbalanced".

Addressing unbalanced data

A highly unbalanced dataset results in strange or wrong models, which is related to the accuracy paradox (Satpathy, 2020; Uddin, 2019). Suppose that of 100 grid cells, just one is chosen by a respondent. A binary classifier model does not need to be complex to predict 0 for all outcomes (meaning not chosen) and achieves a great accuracy of 99% (which states the model is very good). In such cases, where the class distribution is skewed, the accuracy metric is biased and not preferable. The problem of highly unbalanced datasets occurs in many fields, for example, in fraud detection or traffic safety (Uddin, 2019).

Resampling strategies are often used to solve this problem, for which two methods exist: 1) undersampling, in which the majority class will be reduced and 2) oversampling, in which the minority class will be increased. Most research prefers oversampling, as with undersampling the data instances that may be carrying some important information could be removed. Resampling methods do not have a single 'best' method, some methods perform better, and other methods perform less well, compared with model outputs (Kanellopoulos et al., 2006). Combinations of over- and undersampling can also be used and are found to be a better solution (Morris & Yang, 2021). Therefore, a combination method will be used in this research.

An undersampling method eliminates data instances of the majority class to match the minority class and thereby creates balance. Random undersampling is a non-heuristic method that randomly chooses instances to keep (Kanellopoulos et al., 2006). A heuristic method is k-means clustering in which undersampling can be done more proportionally. According to Morris & Yang (2021), an unsupervised k-means clustering rendered slightly better results than random sampling, but the variance in the data reduced significantly. This ensures that the data representation is much better. Therefore, a k-means clustering method will be used as undersampling method in this research

In the oversampling literature, more methods are available. The two most popular methods are: Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthetic Sampling Approach for Imbalanced Learning (ADASYN). Again, also random sampling can be used as a non-heuristic method, in which random instances of data are duplicated to increase the minority class (Kanellopoulos et al., 2006). SMOTE creates synthetic examples of minority class instances (Chawla et al., 2002). These synthetic instances are based on the distance to n nearest neighbours in the dataset and then multiplied with a random number between 0 and 1. As a result, the algorithm effectively forces the decision region of the minority class to become more general.

On the other hand, Adasyn creates synthetic data based on the n nearest neighbours and a distribution according to the difficulty in learning (He et al., 2008). According to Morris & Yang (2021), the methods are quite similar, where the key difference is the density distribution that regulates the number of samples needed. In their comparison, Adasyn-manipulated data improved predictions with logit models, and outperformed random oversampling and SMOTE oversampling. Therefore, the Adasyn is adopted in this research as oversampling method.

3.6.2 Dependent variable

As explained at the start of this section, the dependent variable of the destination model is the binary variable "chosen". It indicates if a grid cell is chosen or not by an OV-fiets user. After the Adasyn resampling method, around half of the data set is classified as chosen, and the other half is not chosen.

3.6.3 Independent variables

The independent variables are the built environment variables, which were described in section 3.4. The variables from the grid are selected based on the theoretical framework described in section 2.3. The criteria regarding correlation apply to these variables as well. This means that variables with a low amount of data and data outliers should be removed. Additionally, highly correlating variables and multicollinearity among variables should be avoided.

After removing those variables, still many variables remain. A few methods of variable selection can be used to limit the number of variables while keeping as much explanatory power as possible. In literature, backward search, forward search and a combination of the two based on Miller (2002) are often used.

Backward search is the searching method which starts with a model that includes all the considered variables. After this, all possible models are tested wherein one of the independent variables is removed. The model that maintains the best fit based on a defined criteria (e.g. lowest drop in AIC or highest p-value) is selected. That model is chosen, and that independent variable is removed. After this, a new iteration takes place until a predefined threshold is reached or until no improvement of the performance evaluator is reached anymore.

Forward search works the other way around. With this method, the search starts with an empty model, which is filled with a dependent variable which improves the model the most.

Both-direction search combines these two search methods. First, it starts with an empty model, after which every iteration adds the dependent variable that creates the best improvement for the performance evaluator. In contrast to the forward search, for every iteration, it is also possible to remove a variable if that improves the performance evaluator. The algorithm stops when the performance evaluator does not improve anymore.

The searching methods do not consider all combinations of variables, which means there is no guarantee that the best possible combination is found. The searching method can be stuck on a local optimum, which (after a decline in the performance evaluator) might be later improved towards a global optimum. However, these methods have lower computational efforts compared to more extensive searching methods. A common limitation of stepwise searching is that it results in a comparatively unstable variable selection. However, this issue can be ignored when the dataset is sufficiently big, which is the case for this research (Steyerberg et al., 2001).

3.6.4 Model structure

Many independent variables were expected to have explanatory power to the binary dependent variable. Therefore, the model structure of the destination model is a binomial regression (also known as a binary logit model), which is part of the discrete choice model family.

To fit a binomial regression, a generalized linear model can be used. Generalized linear models (GLM) provide a unified approach to many of the most common statistical procedures used in applied statistics (Lindsay, 1997; McCullagh & Nelder, 1989). The approach allows for regression modelling when instances are distributed as one of the members of the exponential family (Myers & Montgomery, 1997). GLMs have proved to be effective in problems of applied statistics and are widely used in observational studies (Khuri et al., 2006). The GLM structure allows for additive terms to the dependent variable, which is helpful in understanding the different variables.

The GLM consists of three components:

1. The elements of a response vector are distributed independently according to a certain probability distribution considered to belong to the exponential family.
2. A linear regression function in the form of

$$\eta(x) = f^T(x)\beta \tag{8}$$

Where η is the linear regression, $f^T(x)$ is a transposed known vector-valued function and β is an unknown parameter vector

3. A link function $g(\mu)$ which relates η to the mean response $\mu(x)$ so that $\eta(x) = g(\mu(x))$

As stated at the start of section 3.6, a binary logit model would be the best fit for this research. Therefore, the model structure would be as follows:

$$g(\mu) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \tag{9}$$

Where $g(\mu)$ is the response variable within a link-function, β_0 is the population y-intercept, β_n is parameter of attribute n, x_n is the independent variable. These parameters are also known as the systematic component.

In contrast to linear models, GLMs have no error component ε. This is because the (random) error term specifies the distribution variance, which is assumed to remain constant due to the link function.

The link function links the systematic and the random components and would be defined as equation 10. The link function provides a translation to the terms in the response variable. This allows for the structure of a logit model to be used with more independent variables.

$$\eta(x) = \log\left(\frac{x}{1-x}\right) \tag{10}$$

The GLM is appropriate for analysing discrete and continuous data, because it generalises both the distributional assumptions made about the data and the systematic component defining the expectations. In the GLM, the systematic component defines a linear combination of predictors. However, this assumption does not hold for the distance variable.

The reason is that people with a very short distance to their destination (like across the station square) will not use the OV-fiets. After a certain threshold, trips will increase steeply to a peak. When distance increase, fewer people will then use a bicycle. The shape of the distance distribution is rather a log-normal distribution.

To find the best fitting relationship for the distance variable, its distribution should be fitted. The distances can be calculated based on acquired destinations from the survey. This can be done using the cycling distance and time of Google Maps. This can also be automated with an API, the Directions API (Google Maps, 2022a).

Then, a distribution of distances can be fitted. Since a fitting process is an extensive iterative process, it is challenging to fit a distribution manually. Therefore, a fitting distribution package was used: fitdistrplus (Delignette-Muller, 2014). In this research, only the log-normal distribution is considered to reduce complexity following the literature and previous observations for cycling distance decay.

With the distance variable distributed as another structure than the link function, the model will be formulated as follows:

$$Y = \beta_0 + \beta_1 x_1 \dots \beta_n x_n + \beta_{distance}\left(\frac{1}{x_{distance}\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln(x_{distance}) - \mu)^2}{2\sigma^2}\right)\right) \tag{11}$$

Where Y is the utility of the alternative which can be used in the link-function, β_n are the estimable parameters associated with attribute x_n, $\beta_{distance}$ and $x_{distance}$ are the parameters regarding the distance attribute, σ is the standard deviation of the log-normal distribution and μ is the mean of the log-normal distribution.

3.6.5 Model performance

As for model evaluators, the AIC can be used again to compare different models. The model with the lowest AIC is the best fitting model and should be used for further research steps.

3.7 Model combination

The third model is the combination model, which, as the name suggests, combines the destination model with the purposes model, the green box in Figure 7. The model aims to predict the probability that a grid cell is chosen by an individual or user group. This can be done by using the destination prediction in combination with the land-use and trip purpose choice. The modelling flow is displayed in Figure 12.

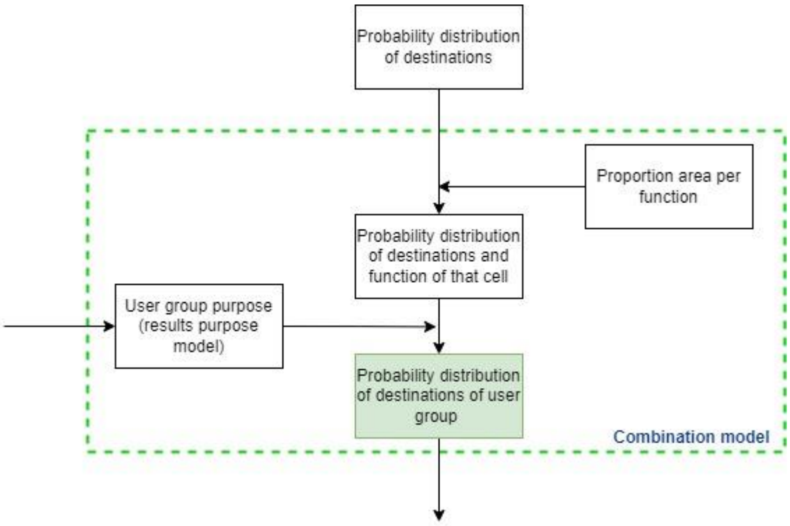


Figure 12: Combination modelling flow

The combination of trip purpose and land use is described in section 2.1.2. In summary, OV-fiets users have an activity to do at the destination they choose. They would not choose a destination where they could not do that activity. Therefore, the building or land-use function is leading in choosing that destination. As a result, the purpose of a trip is connected to the function of the destination.

To translate this connection into a model, input data and mathematical formulas are needed. These are explained in the following sections.

3.7.1 Input data

The destination model creates a probability grid of possible destinations chosen by OV-fiets users. This is used as input data for the combination model. Second, the results of the purpose model are used as input data. The purpose model results in the probability of a certain purpose being chosen by a user or user groups with certain characteristics. Additionally, the building function of a grid cell must be included. This is known from the data preparation, in which the grid was combined with the RUDIFUN database.

3.7.2 Modal structure

The model structure of the combination model consists of relatively simple multiplications:

$$P_{cij} = P_{di} * \sum_{p=1}^{P} (P_{pj} * F_{pi}) \tag{12}$$

Where P_{cij} is the probability of the grid cell i to be chosen by a user (group) j, P_{di} is the probability of cell i to be chosen from the destination model (so based on land-use), P_{pj} is the probability of purpose p of user (group) j, F_{pi} is the function-share of purpose p in cell i.

It means that the probability of choosing a destination (based on the built environment) will be multiplied by the probability of the function-purpose combination for each cell. Within a cell, several functions can exist based on the area. For example, a cell can contain 20% shopping and 80% living area. These shares are combined with the purpose-probability, for example, a user (group) can have a 40% probability of having a shopping-purpose and 60% living-purpose. This means that the total probability of the purpose-function probability is 56% ($(20\% * 40\%) + (80\% * 60\%)$). Combined with a certain destination probability, a total probability of a user choosing that grid cell can be determined.

4 Analysis

The analysis chapter describes the results of the followed methodology. Section 4.1 describes which locations were selected. After that, the survey results from those locations are discussed in section 4.2. The individual characteristics of OV-fiets users are described in section 4.2.4. Section 4.3 covers the estimation of the purpose model, and section 4.4 shows the results of the destination model. After that, section 4.5 illustrate the combination of the models. Finally, section 4.6 covers implications and comparisons of the models.

4.1 Location selection

The rental location selection covers the clustering analysis and selection of the train stations. The selection has two purposes: to define the scope of the survey and to classify OV-fiets rental locations. Based on this, the first research question can be answered. This section is described as follows: section 4.1.1 shows the selected variables, and 4.1.2 shows the clustering method and clusters. Section 4.1.3 goes into detail on the station selection. Appendix C in section 8.3 displays the accompanying figures and tables.

4.1.1 Variable selection

As explained in the methodology, the variables can be distinguished on three levels: bicycle parking, train station and environment. From each level, relevant variables are selected through an iterative approach from the NS-related data sets through the consultation and experience of NS Stations. Many available variables were not relevant to the research.

The considered variables regarding the bicycle parking are: (1) the total number of OV-fiets trips per location and (2) the number of trips per bicycle. The trips per bicycle is the total number of OV-fiets trips divided by the total number of available OV-fiets bicycles at that location. This variable is better than the absolute number of bicycles for the comparison between locations since the variable is relative. However, the absolute number of bicycle trips is taken into account as well, since more trips mean that the OV-fiets is more attractive for train passengers or destinations are more attractive to them.

The train station level has variables such as private bicycle-egress share, station typology, train service levels and number of train passengers. Only the private bicycle-egress share is selected from these because this variable might describe the willingness and attractiveness to cycle from that station. Other variables have strong correlations with each other or are not relevant. The number of train passengers strongly correlates with the number of OV-fiets trips, resulting only in the two busiest stations becoming an isolated cluster. This isolation also happens when stations are classified based on NS-typology and place in the network. After all, these mean that nodes have many travellers as well, due to which they correlate. Thus, the train station level variables are expected to be less useful regarding OV-fiets location clustering.

The last level of variables is the environment level. These variables might distinguish the different environments of the train station. These are (1) attraction of a station (higher attraction results in more egress trips), (2) duration of OV-fiets trip (longer duration might signify other purposes to use the OV-fiets or a more scattered destination density), (3) train trip motives (other environments cause other purposes) and (4) private bicycle distance (cyclable distance to or from a specific station might be a predictor for OV-fiets cyclable distance).

The next step is to check the variables' correlation and outlier risk. Based on the correlation matrix (displayed in Appendix C in section 8.3.2), two variables were removed. The "> 12 h trip duration" was removed due to its almost perfect inverse correlation with the "<10 h trip duration". The ">12 h trip duration” correlated stronger with other variables than the “<10 h trip duration”, hence it was removed. Second, the “Education/Study trip motive” was removed because it created multicollinearity with the other trip motives. The “Education/Study trip motive” was more preferred to be removed because it had a higher correlation with other variables than the other two.

Furthermore, the “distance private bicycle egress” variable was removed due to a lack of data. As this data was only available for 28 stations, it was concluded that it removed too many stations from the analysis. Due to this, the other variables were also scattered, which created less relevant outputs.

Finally, the absolute amount of OV-fiets bookings variable was removed. This variable consistently created a separate cluster for Amsterdam and Utrecht due to the sizes of their bicycle parking. Given the intent to avoid clusters of 2 instances, this variable was removed. The problem was solved to include the relative variable “trips per active bicycle”, which is better for comparing locations.

4.1.2 Clustering

The number of clusters can be defined using the elbow method, silhouette method or gap statistic. The elbow method did not give a clear result. It was not distinguishable which amount of clusters was optimal. The gap-statistic method showed that one cluster was the best for the data. This indicates that no clustering is better than a classification with clusters. However, the gap-statistic for the other cluster amounts was similar, so the result had a risk of being unstable. The silhouette method, however, did show a clear result with 3 clusters as the optimum (the diagram is displayed in Appendix C, section 8.3.3). The cluster means are described in Table 5.

Table 5: Station cluster means, where an underlined value indicates the highest value of the three clusters

Variable	Cluster 1	Cluster 2	Cluster 3	Unit
Trip divided active bicycles (yearly basis)	142,74	205,81	<u>216,79</u>	-
Duration OV-fiets trip < 10 h	62	75	<u>80</u>	%
Private bicycle egress share	20	12	<u>22</u>	%
Attraction of train station	50	<u>61</u>	44	%
Trip purpose train traveller ‘Social/Recreative’	<u>37</u>	29	36	%
Trip purpose train traveller ‘Work/Business’	31	<u>50</u>	43	%

Based on the cluster means, the clusters can be described as follows:

The first cluster has a higher share of OV-fiets trips with long duration, and all three train travel purposes occur roughly equally often. This might indicate that destinations are far away or have a long activity period. The second cluster represents predominately work-related train travel purposes and a low private bicycle egress share. This might indicate that destinations are offices and have a good walk or public transport accessibility. Moreover, the cluster has a high attraction for the train station, which confirms that those stations are more used for activities rather than having residential areas. Cluster 3 represents short OV-fiets trips and high use of OV-fiets in general. This might indicate that destinations are close by or the destinations are more diverse. It can also suggest that bicycles are used for one destination during the daytime and are used again for an evening activity, which results in a higher use per OV-fiets bicycle and a diverse range of destinations. Table 6 shows a short summary of the expected cluster representation.

Table 6: Summarised expected cluster representation

Cluster	Expected representation
Cluster 1	Long, recreational trips
Cluster 2	Commuting
Cluster 3	Mix of destinations close by or short activity period

4.1.3 Station selection

After obtaining different clusters, it is necessary to choose stations from these clusters that represent their whole cluster instead of only a part. This increases the probability of valid survey results. The station closest to the cluster mean and the two stations furthest apart within that cluster are selected to get the broadest view of the cluster. The results of this are displayed in Table 7:

Table 7: Station selection

Cluster	Train station
1	Groningen
1	Nijmegen
1	Maastricht
2	Arnhem Centraal
2	Eindhoven Centraal
2	Amsterdam Sloterdijk
3	Hilversum
3	Delft
3	Apeldoorn*

The station Apeldoorn replaced the station Assen. Assen was the best station to be selected following the procedure. However, as described in the methodology in section 3.2.2, the used data are from 2019 (pre-Covid pandemic), and were compared to the current trip amounts to see if usage is still behind. The number of trips of Assen were too different from the current situation, so Apeldoorn was chosen as the next best option within the selection criteria.

4.2 Survey results

The survey results are described as follows: section 4.2.1 describes the responses, and section 4.2.3 describes the destination geolocation results. After that, section 4.2.4 show the results of the survey in the form of descriptive statistics per variable category. Section 4.2.5 explains the station cluster differences. Finally, section 4.2.6 describe the conclusions of the survey. In Appendix D in section 8.4, the remaining tables and figures are placed.

4.2.1 Survey responses

The final survey questions are displayed in Appendix D in section 8.4. The survey respondents have been collected at each rental location (the train stations) as described before. The times and weather conditions during which the survey was held for each location are displayed in Appendix E in section 8.5.1. The survey times were based on the probability of catching the most users and the availability of the surveyors. These were on weekdays during peak hours. The amount of collected respondents is displayed in Table 8. The weather conditions were predominantly sunny or cloudy, with two exceptions: there were showers once, and another time it was extremely sunny. Those two exceptions did not show irregular patterns for the collected respondents. As such, the weather conditions did not seem to influence the results. Table 8 shows the collected responses per location.

Table 8: Summary of collected responses

	Total	Cluster 1			Cluster 2			Cluster 3		
Location		Groningen	Maastricht	Nijmegen	Amsterdam Sloterdijk	Arnhem Centraal	Eindhoven Centraal	Apeldoorn	Delft	Hilversum
Responses	1045	165	181	104	127	93	130	74	116	55
Responses online	493	62	12	5	45	79	103	92	25	70
Response rate online	27%	20%	12%	56%	23%	32%	45%	34%	19%	22%
Total at location		227	193	109	172	172	233	166	141	125
Total	**1538**	**529**			**577**			**432**		

4.2.2 Survey representativeness

The survey was held by the professional market research institution I&O Research. Each cluster reached the desired number of responses (> 387, see Table 4 in section 3.3.3), indicating that the responses are statistically representative for the population. The commonly used research method of the survey, the professional collection and the number of respondents make the survey representative for the OV-fiets population on weekdays during peak hours. It should be noted that before this research, there was a knowledge gap and information on the population was not known.

The OV-fiets bookings on weekdays have a particular pattern. Figure 13 shows the booking pattern of Eindhoven Zuidzijde in 2022 as an example. The figure shows that most of the bookings and returns are during peak hours, hence these were chosen as the survey moments. This pattern is visible in all surveyed locations. For the three stations in cluster 2, the booking patterns per hour are displayed in Appendix E in section 8.5.2. Additionally, the peak hours account for 68% of all weekday OV-fiets trips in 2022. For 2019 and 2018, the number of trips during peak hours even make up 85% and 88% of all OV-fiets trips on weekdays. For 2019, the share of peak hour OV-fiets trips is 72% of all OV-fiets trips (weekdays and weekend days combined)

As the survey is representative of the OV-fiets population on weekdays during peak hours, and these form the majority of OV-fiets trips on weekdays, the survey represents the majority of weekdays trips.

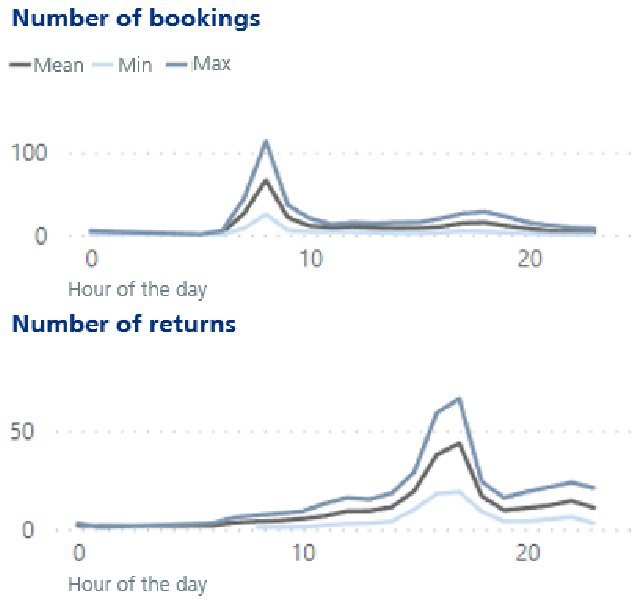


Figure 13: Booking pattern Eindhoven Zuidzijde for weekdays (jan-jun 2022), retrieved from NS Stations OV-fiets dashboard

4.2.3 Geocoding destinations

A component of the survey consisted of filling in the destination of the respondent's OV-fiets trip. It was possible to write the postal code 6 (PC6, for example, "1234 AB"), the street (e.g. "Mauritskade") or the location of interest if the postal code or street name was not known (e.g. "IKEA" (major yellow-blue furniture chain)). The respondents of the survey filled it in as displayed in Figure 14.

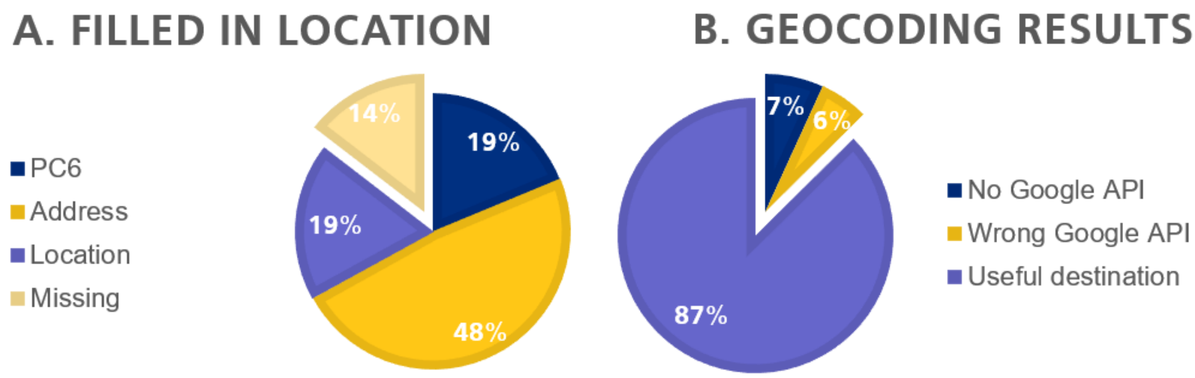


Figure 14: Destination results survey

As can be seen from Figure 14a, about 14% of the respondents did not fill in a destination. In those cases, it was not possible to derive a geolocation. These respondents are not considered with geolocating, although other answers of their response are still used for characterising the OV-fiets user. The destination notations have different accuracies. A PC6 notation is the most accurate because a PC6 area is small, and the middle of that area is not far from the visited building. For Addresses (which required a street name), the accuracy is lower, especially with long streets, as the API geolocates in the middle of the street. Although most of the answers are in this format (56%), some respondents added a house number. This increases the accuracy of geolocating because the API can precisely pinpoint the location. The last type is the Location, which 19% of the respondents used. In this answer format, the accuracy is high regarding specific buildings or facilities. However, it was possible to mention a city (centre) or area (for example, Veluwe). In that case, the geolocation is the middle of such location, which increases the margin of error to the actual location. This was known beforehand and added as an assumption to the research.

Figure 14b shows the results of the Google Places API. The places API could not find 7% of the locations. In these cases, the given location could have been faulty, or it was not found within the radius of the train station, as defined by the scope. Second, 6% of the locations returned with a clear wrong geolocation. Examples were a pizza place in Rome or the postal code “1234 AA”. Such locations were filtered out of the data. The assumption that the first found Google Maps Places solution is deemed correct, was found to be valid. A sample of geolocations was checked and compared with the API result. In all cases, no significant difference was found. In the end, 1162 (76%) of the responses could be geolocated, distributed as 381 (72%), 457 (79%) and 324 (75%), respectively, over the clusters.

4.2.4 Descriptive statistics

The following section shows the descriptive statistics of the collected dataset. Those are divided in socio-economic characteristics, travel patterns and trip characteristics. For each category, the shares of the survey and the shares of each cluster are displayed. The confidence interval to indicate statistically significant differences between the clusters is displayed in Appendix E in 8.5.6. The clusters are compared both to each other and to the sample as a whole. Additional figures of the respondent share are also displayed in Appendix E in section 8.5.3.

Socio-economic characteristics

Table 9: Socio-economic characteristics from the survey

Category		Survey [%]	Cluster 1 [%]	Cluster 2 [%]	Cluster 3 [%]
Gender	Male	*51*	*42*	*56*	*56*
	Female	*44*	*52*	*40*	*41*
Age	< 30	*38*	*50*	*31*	*33*
	31-40	*21*	*16*	*22*	*24*
	41-55	*23*	*18*	*26*	*24*
	56-65	*12*	*9*	*13*	*13*
	66 +	*5*	*5*	*5*	*4*
Education level	Low	*2*	*1*	*2*	*1*
	Middle	*16*	*19*	*18*	*9*
	High	*80*	*77*	*76*	*87*
Working situation	School going	*17*	*30*	*11*	*9*
	Self-employed	*9*	*9*	*9*	*9*
	Employed	*63*	*47*	*70*	*73*
	Retired	*3*	*3*	*3*	*3*

In the category of **gender**, there were more males than females. This indicates that OV-fiets users are more often male. However, a noticeable number of respondents did not fill in the question (3%) or filled in ‘other’ (2%).

Comparing the three clusters, none of the clusters falls in a 90% or 95% confidence interval compared to the whole sample. This indicates that all the clusters are different to the whole sample. Another noticeable difference is that cluster 1 has more females than males. Clusters 2 and 3 are almost the same regarding gender.

Regarding the **age** distribution, the survey collected more younger OV-fiets users, see Figure 15. It shows that OV-fiets users are younger during peak hours. Clusters 2 and 3 are quite similar to the whole sample population. The majority of the age categories are within the 95% confidence interval. Cluster 1, however, differs for all age categories (but 66+). Respondents in cluster 1 are younger in general. This could be explained that a university is located in all three cities of cluster 1, so the OV-fiets users might be dominated by students.

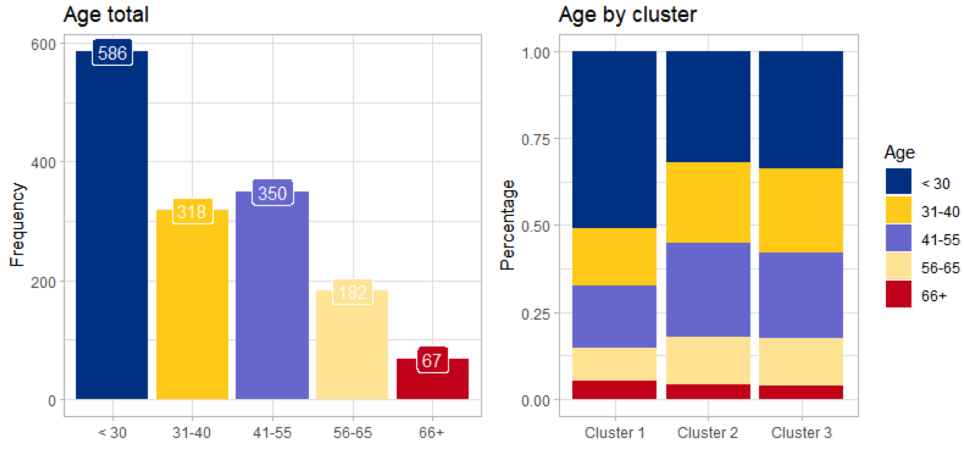


Figure 15: Results of the age distribution of respondents

The **education level** results show that OV-fiets users are predominantly highly educated. The clusters give a similar view as the whole sample population: people are highly educated. Only cluster 3 shows an even larger share of highly educated OV-fiets users.

The **working situation** results show that most OV-fiets users are employed, followed by school-going respondents. This could be explained since retired citizens are commonly not travelling during peak hours and, therefore, were not seen in the survey hours. Comparing the clusters, the self-employed and retired categories have a comparable share of the whole population as for all clusters. The big difference is that cluster 1 has a much higher share of school-going OV-fiets users. This, again, could be explained by the presence of universities in the cities of cluster 1.

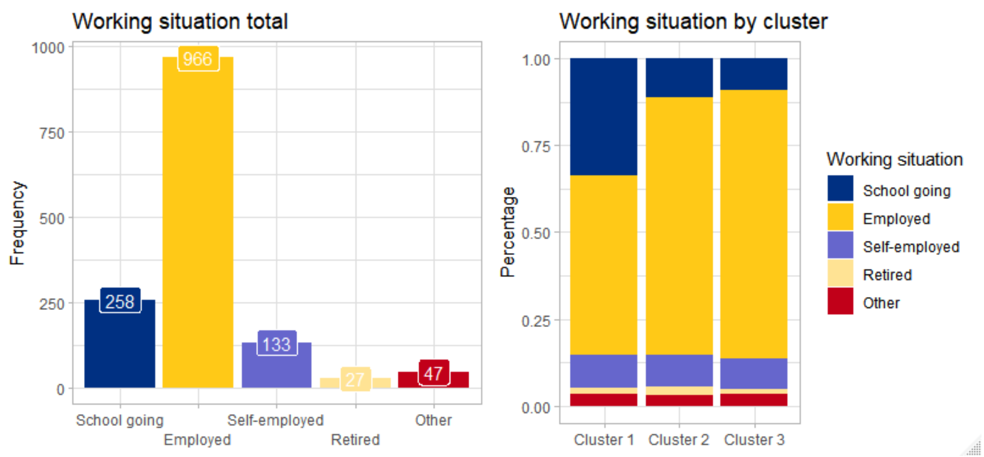


Figure 16: Results of the working situation of respondents

Travel pattern

Travel patterns of OV-fiets users relate to users' general use and habits towards the OV-fiets and public transport. The descriptive statistics of this category are displayed in Table 10.

Table 10: Travel patterns from the survey

Category		Survey [%]	Cluster 1 [%]	Cluster 2 [%]	Cluster 3 [%]
Frequency OV-fiets general	> 1x/week	*41*	*29*	*49*	*44*
	1-3x/month	*23*	*22*	*21*	*25*
	1-11x/year	*21*	*27*	*18*	*19*
	<1/year	*7*	*13*	*3*	*4*
	No general use	*2*	*4*	*2*	*1*
Week/weekend usage	More week	*71*	*63*	*73*	*78*
	More weekend	*12*	*17*	*10*	*8*
Pays for Public Transport	Yes	*30*	*34*	*27*	*28*
	With reduction	*16*	*19*	*13*	*16*
	No (free travel)	*10*	*14*	*7*	*7*
	No (employer pays)	*38*	*24*	*47*	*43*
PT satisfaction	Rating 7 or higher	*88*	*89*	*88*	*89*
	Rating 8 or higher	*58*	*59*	*55*	*59*
OV-fiets satisfaction	Rating 7 or higher	*95*	*94*	*95*	*96*
	Rating 8 or higher	*81*	*81*	*81*	*81*

The **travel frequency of OV-fiets in general** resulted in most respondents using the OV-fiets multiple times per week. This indicates that commuters are using the OV-fiets for their trip to work. Another observation from this category is that relatively few people have a low frequency or make the trip for the first time.

Regarding differences between clusters, cluster 1 shows a noticeable difference in usage. The 'less than one time per year-share is bigger. This might be because those stations are located in the periphery of the Netherlands and so have long train distances. Therefore, train travellers are longer on the train and might be more recreational users who use the OV-fiets less often. Another observation comprises that the multiple times per week usage is more than half of the sample for cluster 2, which emphasises the commuting characteristic of those stations.

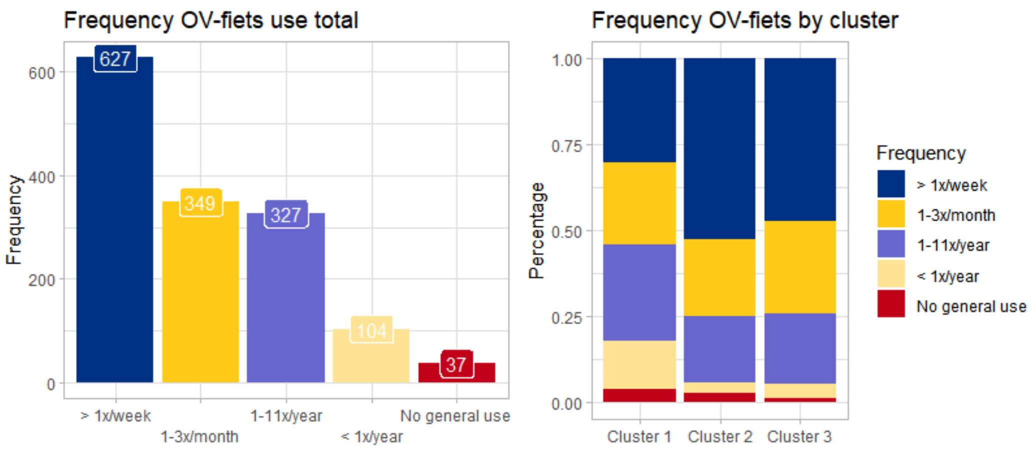


Figure 17: Results of OV-fiets frequency in general

Respondents were also asked about their general usage of OV-fiets on **week or weekend** days. As the survey was held only on weekdays, it was expected that the primary usage of OV-fiets would also be on weekdays. This can also be seen in the survey results, as 86% of the respondents answered that. More remarkable is the difference between clusters. Cluster 1 shows almost a quarter of the responses to use the OV-fiets more at the weekend. This can be the result again of the more recreational purpose focused train stations in cluster 1.

People were also asked about their **usage around the Covid-19 pandemic**. These questions were asked to see if the survey is comparable to the situation before the pandemic and if trip patterns have changed because of Covid-19. The Sankey plot in Figure 18 shows that, in general, the usage of OV-fiets is higher than before covid. Respondents stated that they use the OV-fiets nowadays more than before Covid. This is not surprising as the respondents were actually using the OV-fiets at that moment. However, in the patterns, it can be seen that most of the respondents have increased their frequency. Most of the frequencies increase from 'during covid' to 'after covid'. Also, the frequency groups after covid are bigger than they were before covid, indicating more usage. Therefore, it can be concluded that the survey did not suffer from the Covid regulation aftermath. On the contrary, the usage frequency is even increased.

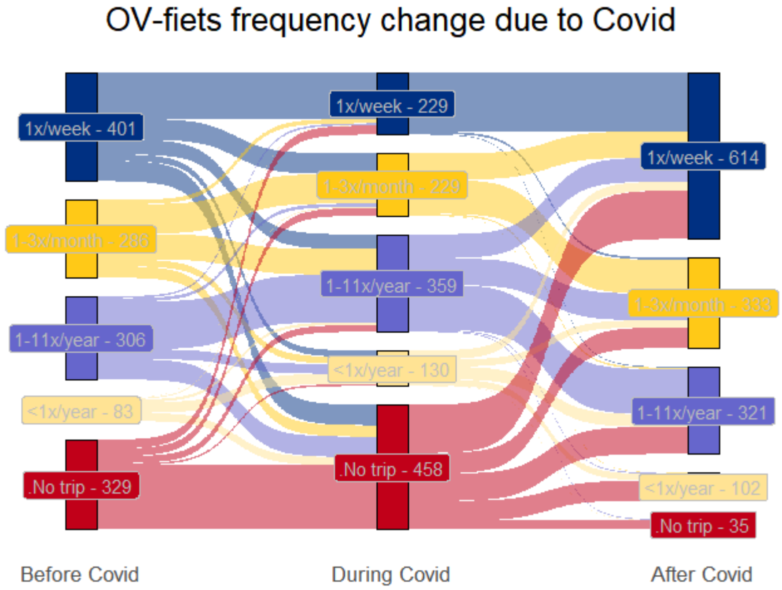


Figure 18: OV-fiets use during Covid

In the **pays for public transport** category, OV-fiets users were asked if they would have had to pay for their public transport at that moment. It could be the case that public transport was free or paid for by the employer. In those cases, the threshold of using public transport is low and might reduce the OV-fiets usage. The results indicate that this is not the case. More than half of the respondents did have reimbursed public transport, but they still used the OV-fiets. It should be noted that the OV-fiets is often reimbursed too, but not always. More on this in the next section.

The difference between clusters predominantly concerns the share of users who get reimbursed by their employer. In clusters 2 and 3, this is the case for almost half of all users, while this holds for only a quarter in cluster 1. Again, this emphasises the commuting characteristics of clusters 2 and 3. Additionally, the share of people paying for public transit is largest in cluster 1. It might be that more students have an OV-chipcard with free weekend travel instead free travel on weekdays. This makes sense as cluster 1 has long train trips, which might cause students to travel only on weekends to their student housing.

The **OV-fiets and public transit satisfaction** identified the preferences of the OV-fiets users. It was expected that OV-fiets users (who are train passengers) are highly satisfied with public transport and OV-fiets. The results of the survey clearly show that. In general, the satisfaction of the OV-fiets (mean = 8.3) is half a point higher than the satisfaction of public transport (mean = 7.6). Moreover, the results confirm the hypothesis that people who use the OV-fiets are also quite satisfied with public transport. The satisfaction marks are comparable with earlier NS research on the OV-fiets product.

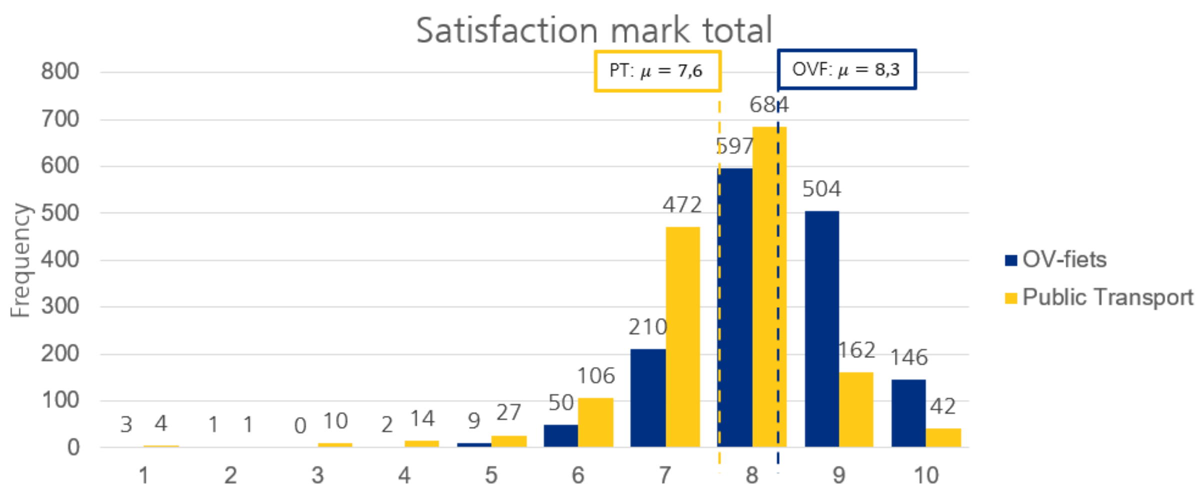


Figure 19: Results of satisfaction rating OV-fiets and public transport

Trip characteristics

This section contains the specified trip characteristics for the OV-fiets trip and the accompanying alternatives to and reasons for the OV-fiets. Respondents could fill in multiple answers for the available alternatives and reasons to use the OV-fiets.

Table 11: Trip characteristics from the survey compared to other research

Category		Survey [%]	Cluster 1 [%]	Cluster 2 [%]	Cluster 3 [%]
Trip purpose	Work	*51*	*30*	*62*	*63*
	Business trip	*8*	*7*	*6*	*12*
	Education	*7*	*12*	*5*	*4*
	Visits	*14*	*19*	*12*	*11*
	Shopping	*1*	*2*	*1*	*0*
	Leisure	*9*	*19*	*5*	*3*
	Sports	*2*	*1*	*3*	*3*
	Other	*6*	*9*	*5*	*4*
Trip frequency	Daily	*9*	*3*	*14*	*8*
	> 1x/week	*34*	*25*	*38*	*41*
	1-3x/month	*20*	*17*	*20*	*22*
	1-11x/year	*11*	*12*	*10*	*12*
	< 1x/year	*2*	*4*	*1*	*2*
	First time	*23*	*37*	*17*	*16*
Available alternatives (mentioned)	Train	*4*	*5*	*3*	*4*
	Bus/Tram/Metro	*50*	*44*	*49*	*56*
	Walking	*26*	*28*	*23*	*25*
	Car	*8*	*5*	*10*	*7*
	Taxi	*2*	*1*	*2*	*2*
	Private bicycle	*11*	*6*	*11*	*16*
	Shared mobility	*11*	*8*	*10*	*12*
	None	*9*	*11*	*9*	*6*
Amount of destinations	1	*70*	*54*	*74*	*83*
	2	*17*	*22*	*17*	*11*
	3	*12*	*20*	*8*	*6*
Pays for OV-fiets	Self	*55*	*68*	*24*	*6*
	Employer	*41*	*52*	*49*	*2*
	Someone else	*3*	*40*	*37*	*1*
Reason (mentioned)	Convenience	*71*	*73*	*66*	*72*
	Freedom	*46*	*38*	*47*	*52*
	Speed	*47*	*38*	*48*	*56*
	Health	*24*	*20*	*26*	*27*
	Environment	*25*	*19*	*27*	*29*
	Comfort	*13*	*13*	*12*	*13*
	Costs	*8*	*9*	*8*	*6*

Regarding **trip purpose**, the work purpose dominates the response. Again, this can be explained by the survey times, which might overrepresent commuters. Comparing the clusters with each other, clusters 2 and 3 are (again) quite similar. Cluster 1, on the other hand, has relatively few work-related trips. In this cluster, trips for visits and leisure occur more often. This might be related to the long train trips and the far-away stations attracting more social-recreative train trips instead of commuter trips.

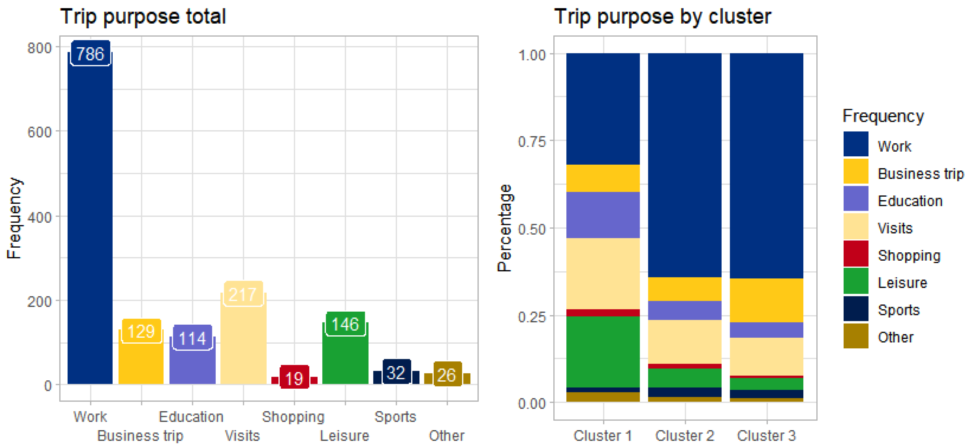


Figure 20: Results of respondents' trip purposes

When analysing and plotting the destinations given by the respondents, **trip travel time and distance** could be calculated. The procedure for this is explained in the methodology, section 3.4.1. The results of the respondent destination in the spatial dimension are displayed in Appendix E in section 8.5.5. The results show that, grosso modo, the single distance of the trips is between 1900 meters and 4200 meters, and the travel time is between 420 seconds (7 minutes) and 900 seconds (15 minutes) based on the interquartile range of the sample. The median is 2800 meters and 630 seconds (10,5 minutes).

Comparing the clusters, it is noticeable that cluster 2 has longer distances and cluster 1 has small distances. This might indicate that cluster 1 has more compact or smaller cities, resulting in smaller distances to the station. It could also mean that the stations of cluster 2 are not in the centre of the city, which causes possible destinations to be further away. Moreover, Figure 21 shows that the clusters' distance and travel time are quite similar. This indicates that there are no major barriers or very slow sections around the stations of a cluster.

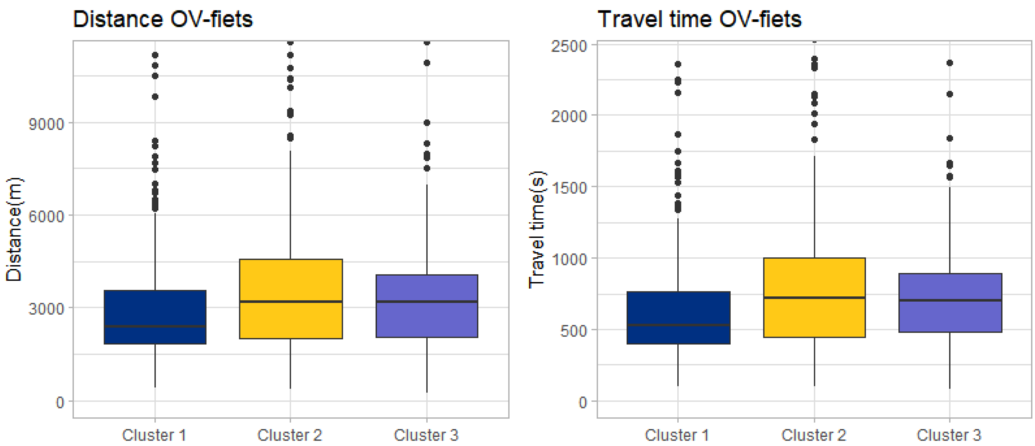


Figure 21: Results of distance and travel time to destinations of OV-fiets users

Another comparison can be made of the distances categorized by trip purpose. The results in Figure 22 show that the leisure trip purpose has longer trips and the largest variation between trip distances. This confirms the expectation that people with a leisure purpose have more time and more diffuse destinations. On the other hand, shopping and sports trip purposes are more concentrated as locations to undertake these activities are concentrated too, namely shopping centres or sports centres.

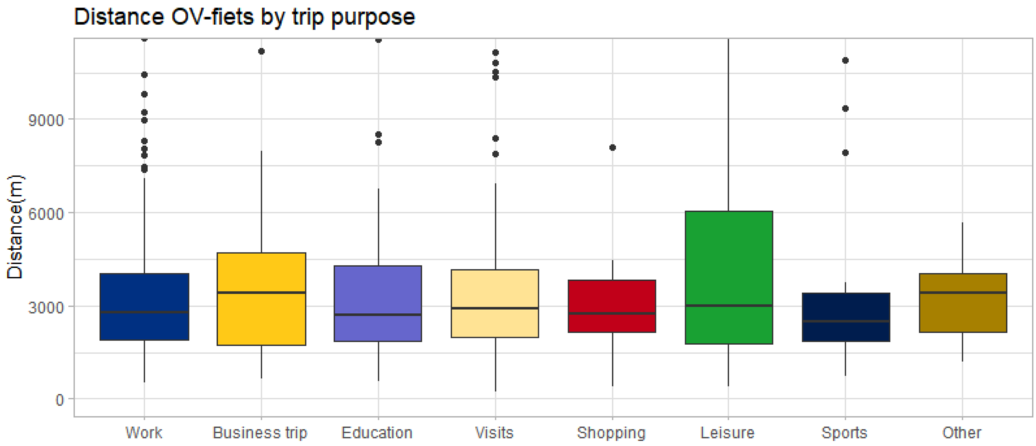


Figure 22: Trip distance categorized per motive (y-axis limited on upper 95% confidence interval of all observations)

The survey respondents were asked to denote their **trip frequency** to the destination they had written down. The largest group of the respondents have a frequency of multiple times per week, confirming the commuting patterns of the peak hours. Another observation is that a quarter (23%) of the respondents answered that this was the first time the trip was made or that there was no regular frequency of this specific trip. Moreover, this first-time frequency is especially visible for cluster 1. An explanation of this can be found with the more prominent recreational aspect in cluster 1.

As OV-fiets can be used for unlimited destinations within the 24-hour rental period, respondents were asked how **many destinations** they visited. Having multiple destinations could be an incentive to choose OV-fiets over other modes. Roughly 30% of the respondents had more than one destination to visit. This indicates that the majority of the OV-fiets users have one clear destination in mind to visit. This also fits in the commuting patterns.

The first cluster differs from clusters 2 and 3 in the multiple destination category. Cluster 3 has a higher share of only one destination as a difference. Therefore, cluster 3 has more single trips and cluster 1 has more multi-destinations trips. This might indicate that the cities of cluster 1 are more diverse as multiple activities are undertaken. Or it might be that the rental duration of cluster 1 is longer than the trips within cluster 3.

In addition to questions on paying for public transport, respondents were also asked if they **paid for OV-fiets**. It could have been the case that the employer reimburses travel costs but not the OV-fiets. Furthermore, users can pay for other group members when people travel in groups. The results of this question show that 55% of the respondents paid the OV-fiets themselves. For the majority of the remaining respondents, the OV-fiets was paid for by their employer. For the peak hours, there is a marginal part of users for whom someone else (other than the employer) pays for the OV-fiets.

Again, there are differences between clusters, especially between cluster 1 and the other two. Clusters 2 and 3 are comparable with the total population. Cluster 1 stands out because more respondents are paying their OV-fiets themselves. Again, this might be explained by the more social-recreative aspect of cluster 1.

The question of **reasons to choose OV-fiets** is different from previous questions. Here, OV-fiets users could mention more than one option (therefore, the percentages in Table 11 did not add up to 100%). The results show that the biggest reasons to choose OV-fiets are convenience (mentioned by 70%), followed by freedom and speed (both almost 50%). This might indicate that the availability of using the bicycle unlimited for 24 hours is a big advantage or that the OV-fiets can get closer to the destination compared to other modes. After that, bicycle-related reasons, such as health and the environment, were mentioned. At last, the reason for costs is mentioned by almost 10%, which indicates that for a small part of the users, the (low) costs of the OV-fiets attract them.

In addition to the reasons for choosing OV-fiets, the **available alternatives** were asked. Again, multiple options could be answered. A side note to this question is that it displayed the *perceived* alternatives. In essence, walking could always be done, but people's willingness to walk might differ. As shown in Figure 23, the most mentioned alternative was the bus/tram/metro. This makes sense, as the user group is foremost train travellers, in which public transport usage is common, and transfers between PT are often easy. Another observation is the low number of people with no other mode available. This indicates that the OV-fiets destinations are reachable by other modes, too (as perceived by users). Thus, the dependence of OV-fiets is only for a relatively small group (around 10%).

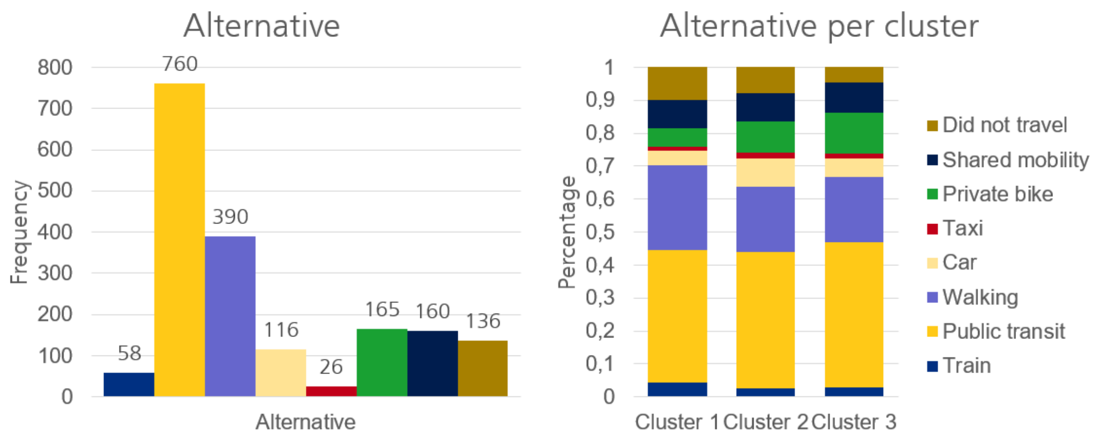


Figure 23: Results of available alternatives for the respondents

4.2.5 Cluster differences

The clusters have some differences between each other and with the total sample (see Table 29 for the 95% confidence interval differences in Appendix E in section 8.5.6). Comparing the total sample with the respondents of each cluster, cluster 2 has the most similarity to the total dataset.

Cluster 1 is the most different from the other two, mostly regarding age, usage of OV-fiets, paying for public transport, trip purposes, trip frequency, and the number of destinations. On the other hand, cluster 2 and cluster 3 were similar to each other, even though the hypothesis was that the station clustering would create different clusters with different OV-fiets users.

Almost every answer option of each category in cluster 3 falls within the 95% confidence interval of cluster 2. This holds for all categories but education level and pays for OV-fiets, which means that the stations of cluster 3 attract different kinds of users in those categories compared to the stations of cluster 2.

4.2.6 Conclusions

It can be concluded that the clustering approach was partly successful, which answers the first sub-research question. The classification successfully defined stations with a different OV-fiets user for cluster 1, but clusters 2 and 3 attract almost the same OV-fiets user type. This means that the used variables and outcome of the cluster analysis were not fully successful in distinguishing different OV-fiets rental locations.

Based on the survey results into account, the second research question can be partially answered because the characteristics of an OV-fiets user have been described. A typical OV-fiets user (for peak hour times) can be characterized as a commuter: young to middle-aged, working people who are highly educated and regularly travel with public transport (and OV-fiets). Also, they have positive satisfaction for both public transport and OV-fiets.

4.3 Purpose model estimation

The purpose model was estimated using a discrete choice model based on the socio-economic characteristics of the whole survey, as described in section 3.4.1. The input data are described in section 4.3.1, the model estimation in section 5.3.2 and validation in section 4.3.3.

The model can estimate the influence of a characteristic on the probability of having a certain trip purpose. The trip purpose outcomes are used in a later stage of this research. 80% of the collected data is used to estimate the model. The other 20% of the data are set aside and are used for the validation steps, see section 4.3.3. Random sampling is used to distinguish the data instances to have a non-biased selection. The 80/20 division is commonly used in modelling and machine learning literature, see, for instance, Chakour & Eluru (2014); Faghih-Imani & Eluru (2015); Washington et al. (2011).

4.3.1 Input data

The dependent variable purpose distinguishes three possibilities: work, education and leisure. The categories are based on the survey taking into account that a category has enough responses to train the model. The dependent variable classes are referenced to the purpose “work”, which was chosen because work is expected to be the most common purpose when travelling during peak hours.

The independent variables are also based on the survey and are converted to binary variables for model estimations. The variables were tested on correlation and covariation with a correlation matrix, which is displayed in Appendix F in section 8.6.1.

Effects on the input data are described in the implications and limitations in section 4.6.

4.3.2 Model estimation

Table 12 shows the estimation of the binomial logistic regression model. Three models are displayed: one model shows a null model as a reference, the second model shows all the possible variables, and the third model shows an estimation without variables that have a low occurrence (which results in a bad prediction due to a small data size). Further iterations were made with other variable combinations, for example, with only the significant variables of the ‘all variables’-estimation. But, those iterations did not perform better than using all the variables (except for the ‘other’-categories).

The model wherein the low-frequency categories were removed had the lowest AIC, which means it fits best. The model performs much better than the reference “null”-model, as the AIC is significantly lower. Therefore, the variables explain a part of the purpose of an OV-fiets trip.

*Table 12: Purpose model estimation. Note: the value between the brackets shows the standard error. Second, the asterisks show the significance level with *p<0.1; **p<0.05, ***p<0.01*

		Reference		Selected variables	
Category	Variable	Education	Leisure	Education	Leisure
Gender (ref = male)	Female			0.941*** (0.299)	0.173 (0.202)
Age (ref = < 30)	31-40			-0.730 (0.481)	0.052 (0.294)
	41-55			-1.202** (0.501)	-0.606** (0.296)
	56-65			-1.354* (0.779)	0.342 (0.355)
	66+			-1.667 (1.101)	0.988** (0.460)
Education level (ref = low)	Middle			0.309 (0.685)	0.448 (0.534)
	High			-0.229 (0.647)	-0.480 (0.502)
Working situation (ref = school going)	Employed			-2.435*** (0.418)	-1.050*** (0.305)
	Self-employed			-2.753*** (0.811)	-1.781*** (0.402)
OV-chipcard type (ref = on balance)	Student			0.805* (0.485)	0.377 (0.437)
	Business card			-0.270 (0.584)	0.726** (0.356)
	NS flex			0.054 (0.410)	-0.191 (0.249)
Frequency OV-fiets (ref = >1x/week)	1-3x/month			1.009** (0.425)	0.212 (0.305)
	1-11x/year			0.249 (0.513)	0.427 (0.307)
	< 1x/year			0.308 (0.649)	0.539 (0.414)
Pays for PT (ref = full tariff)	Discount			-0.853** (0.424)	-0.327 (0.241)
	Employer			0.356 (0.548)	-1.440*** (0.399)
	Free travel			-0.230 (0.483)	-0.171 (0.416)
Rating (continuous)	OV-fiets			0.198 (0.129)	-0.013 (0.083)
	Public transport			-0.019 (0.116)	0.156* (0.088)
Frequency trip (ref = first time)	Daily			-0.264 (0.627)	-17.237*** (0.00000)
	>1x/week			-0.038 (0.459)	-2.200*** (0.325)
	1-3x/month			-0.348 (0.465)	-0.867*** (0.303)
	1-11x/year			0.123 (0.491)	-0.405 (0.296)
Number of Destinations (ref = 1)	2			0.274 (0.353)	0.455* (0.252)
	3			0.647 (0.439)	0.988*** (0.312)
Pays for OV-fiets (ref = self)	Employer			-1.529*** (0.466)	-2.576*** (0.368)
Constant		-2.082*** (0.111)	-0.784*** (0.066)	-1.803 (1.213)	0.759 (0.860)
AIC		**1984.989**		**1197.202**	

Each of the values represents the coefficient β in the deterministic part of the utility function V. A positive coefficient means an increase, and a negative coefficient means a decrease. Having coefficient β is associated with an x change in the log odds of being in the category compared to the base category ("work") while keeping all the other variables constant. For example, being female increases the log-odds by 0.941 for having an educational purpose and 0.173 for having the trip purpose leisure compared to work. In other words, if the OV-fiets user is female, then the probability of having a trip purpose education versus trip purpose work will increase by 0.941. Or when exponentiating the coefficient, being female increases the odds of having an education purpose by 156% compared to a work trip purpose, and increases the odds by 18.9% to have a leisure trip purpose compared to work.

The p-value (displayed as asterisks) shows the significance of the association between the variables. It means that the coefficient is not equal to zero (which means no effect). When the p-value is lower, the evidence of the coefficient having an effect gets stronger. Commonly, in literature, a 0.05 significance level is used. However, a low evidence depicted with a p-value of 0.1 (*) is displayed as well.

The socio-economic variables gender, age, education level, working situation, and OV-chipcard type can be interpreted as follows. Being female positively affects switching to an education or leisure purpose. However, the education purpose has a higher probability, and only this effect is significant. For age, significant effects are visible for the 41 to 55 age group on both education and leisure (negative effect), for the 56 to 65 age group for education (negative) and 66+ age group for leisure (positive). This means that the 41 to 55 age group is less likely to switch to education or leisure purposes compared to work. This group has "work" as the most predominant purpose. The 66+ group is more likely to choose leisure. This makes sense as 66+ age groups are predominantly retired and have time for more leisure activities.

Regarding (self-) employment, the model coefficients (Table 12) shows that there is a strong negative relation of switching to education or leisure purpose compared to work. This makes sense as being employed and travelling during peak hours implies that the user has a work purpose. Finally, the OV-chipcard types suggest that having a student travel card positively affects switching to an education purpose compared to work. This also makes sense, as only students who applied to an educational institution have a student travel card. Second, having a business card positively affects leisure purposes compared to work, with a relatively high odds ratio of 106%. This is surprising as business cards are usually supplied by employers to travel to work. An explanation could be that those users have multiple locations where they have a leisure activity with a work activity.

The travel patterns are indicated by the general frequency of OV-fiets, paying for public transport and satisfaction with PT and OV-fiets. The general usage frequency indicates that OV-fiets usage of 1 to 3 times per month increases the probability of switching to education purposes compared to work. Other frequencies have no significant effects on switching to education and leisure purposes. Regarding the payment for public transport, having a discount on PT had a negative effect on the education purpose compared to the work purpose. Another significant effect is that if the employer pays for PT, a strong negative effect is visible towards the leisure purpose. This is not surprising as leisure trip costs are commonly not reimbursed by the employer. Considering the satisfaction of OV-fiets and PT, the only (small) significant effect is visible on the satisfaction of PT. Suppose the satisfaction of PT increases by 1 point, the odds of having a leisure purpose increase. This indicates that people who have a higher preference for PT are more likely to travel social-recreative.

The last section of variables is the trip variables. These consist of trip frequency to that destination, the number of destinations and how the OV-fiets is paid for. Based on the model, trip frequencies have a significant effect on not switching to the leisure purpose from the work purpose. For trip frequencies of 1 to 3 times per month and higher, there is a strong negative relation for choosing leisure. Especially with a 'daily' frequency, then the log-odds of choosing leisure over work become very small. In essence, only trip frequencies of 1-11x year and first-time users (which is the removed category) indicate that the leisure purpose has a realistic probability of being chosen over the work purpose. On the other hand, the number of destinations positively influences choosing leisure over work. This is not surprising as leisure trips are more flexible, and people have more time to visit activities or locations.

4.3.3 Model performance

The performance or validation of the model can be measured via several indicators. As explained in section 3.5.5, the models themselves are compared on their AIC. Other measures are needed to compare how accurate the results are. Common metrics are accuracy, precision, recall and F1 score (Shmueli, 2019; Sokolova & Lapalme, 2009).

- The accuracy is the percentage of observations that the model correctly predicted. In this case, it means that the predicated purpose from the model is the same as the observed purpose. The accuracy-paradox that occurs with an imbalanced dataset is less of a problem in the purpose model because the purposes are more equally divided than the whole dataset. The higher the accuracy, the higher the correctness of the model.
- The precision explains how reliable the prediction is for a class. It is also known as the positive predicted value and indicates the fraction of relevant instances among the retrieved instances. It shows how dispersed the predictions are. The higher the precision, the more reliable the outcomes of that certain class are.
- The recall states how well the model can predict the purpose. The recall is also known as the sensitivity. A high recall with low precision means that the model performs well within the purpose class but may include other purposes in the results.
- The F1 score is the harmonic mean of the precision and the recall to combine them in a single metric. The F1 score is used predominantly to compare different models with each other in one glance. However, the F1 score comparison does not take domain knowledge into account and comparisons with this metric should be done carefully. In certain domains, precision should be very high, while in others, this is less of an issue (Hand & Christen, 2018).

The precision, recall and F1 score are calculated per class. To calculate them for the whole model, the macro precision, macro recall and macro F1 score can be calculated as the mean of those scores.

For this validation, the set-aside 20% of the data is used. These data are then used as input in the estimated model to create results. The comparison of results can be visualised with a confusion matrix (Ben-Akiva & Lerman, 1997), see Table 13.

Table 13: Confusion matrix purpose model results

		Predicted purpose		
		Work	Education	Leisure
RP purpose	Work	157	8	12
	Education	5	2	15
	Leisure	26	5	46

The table can be used to determine the true positive estimates (the diagonal) and the false positive estimates for each estimated purpose. The statistic values for the validation are displayed in Table 14.

Table 14: Validation statistics purpose model

	Total model with test data	Work	Education	Leisure
Accuracy	74%	-	-	-
Precision	-	83%	13%	63%
Recall	-	88%	9%	59%
F1 Score	-	85%	11%	61%
Macro precision	53%	-	-	-
Macro recall	52%	-	-	-
Macro F1 score	52%	-	-	-

As can be derived from the table, the model is quite good in estimating work purposes, moderately in estimating the leisure purpose and not good in predicting the education purpose. For work and leisure purposes, the accuracy is above 60%, which is evaluated as quite good (Sokolova & Lapalme, 2009). For the education class, the accuracy is not good, and the model estimates more leisure. This might be because trip frequencies and user characteristics are similar to those of leisure users, which makes it hard to predict. Looking at macro evaluators, there can be concluded that the model performs around 50% in contrast to the 74% accuracy. The bad education purpose prediction heavily influences the macro indicators. A macro F1-score between 0,5 and 0,8 is evaluated as "OK", which applies to this model (Allwright, 2022; Yacouby & Axman, 2020).

4.4 Destination model estimation

The next part of the modelling flow is the destination prediction. The destination model was estimated with a binomial logistic regression model based on the built environment and the destinations of the survey, as described in section 3.6. Section 4.4.1 shows the input for the model. Section 4.4.2 covers the model estimation, and section 4.4.3 describes the validation of the model.

The model can define a probability distribution for destinations based on the built environment. This probability is used to define the probability of choosing that grid cell in general. The model's outcomes can be used in further steps of this research. 80% of the collected data from the survey is used to estimate the model. This 80% consists of the same respondents used for the purpose model in section 4.3. The other 20% can be used for validation steps.

The built environment data was available for cluster 2 only. Therefore, in further steps of this research, only cluster 2 is considered. These are the stations Eindhoven Centraal, Amsterdam Sloterdijk and Arnhem Centraal. Cluster 2 was selected for having all land-use functions, which made it possible to investigate all variable categories regarding the built environment. This was helpful in setting up the survey and the research, especially with data fusion. Due to research time, computational time and available resources, the choice was made to analyse cluster 2 only. Later, it became clear that cluster 2 had the most respondents who filled in their destinations, making it even more suitable for fitting the model. In the following sections, Eindhoven Centraal is used in the figures and analysis, while Eindhoven has no big forests or water masses around the station. However, the model was fit for all three stations, which are displayed in Appendix G in section 8.7.

4.4.1 Input data

The model's input data is the built environment grid, which is spatially joined with the destination location part of the responses from the survey (see Appendix E in section 8.5.5 for a visualisation of the geolocations). In this spatial join, 1% of the destinations were excluded because they were outside the 10-kilometre range. Based on this grid, the resampling methods were performed to create balanced input data, as described in the methodology in section 3.6.1.

The undersampling was done with k-means clustering. The silhouette method was used to find the optimal amount of clusters to be used in the undersampling. The other two methods which can be used to find the optimal amount of clusters (gap statistic and elbow method) were unable to find a solution. This was related to the data size. The silhouette method showed that 14 clusters created the best solution (see Appendix G in section 8.7.2). Then, a random sample was drawn from the clusters with respect to the cluster sizes. The oversampling was done with the ADASYN method. The data were oversampled to 8 times the minority class (the chosen grid cells). This amount was selected to have enough data to estimate the model but have appropriate calculation times. Finally, the five nearest neighbours were used to create the synthetic data. This amount is commonly used in literature which uses ADASYN (for example, Morris & Yang (2021)).

As described in the methodology (section 3.5.4), the distance variable was fitted with the fitting package. The results of the fitting are displayed in Figure 24. The blue bins show the distance occurrences, and the yellow line shows the log-normal distribution with $\mu = 7.914803$ and $\sigma = 0.539112$.

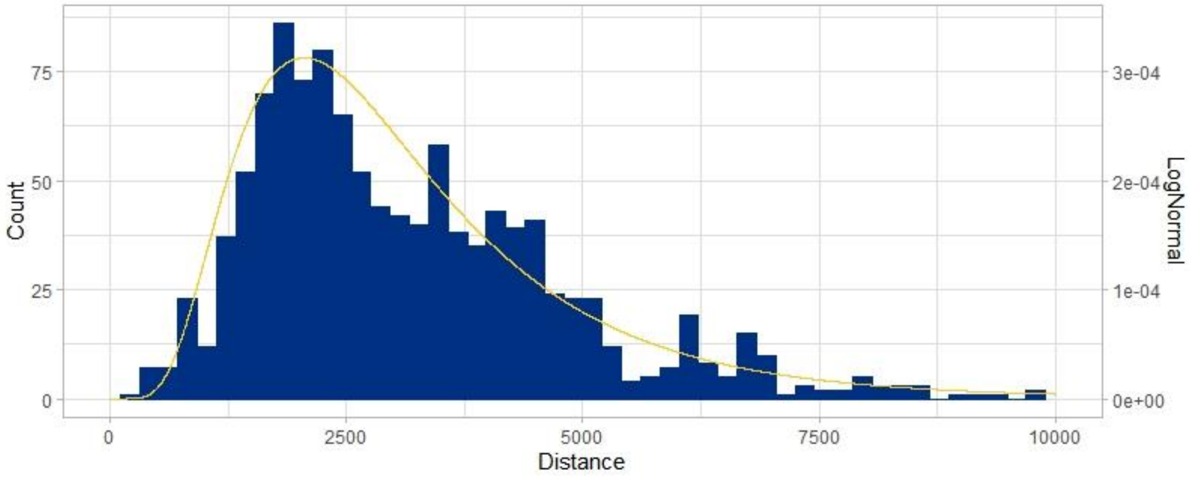


Figure 24: Histogram and transformed distribution of distance

4.4.2 Model estimation

The resampling methods have a random element. A synthetic dataset can sometimes create a much better (or worse) model fit for an iteration than another random run, which can cause incorrect conclusions on the model's coefficients. The model estimation has been run with 1000 iterations to reduce the randomness, each iteration creating a new synthetic dataset and estimating the binomial logistic regression. After that, the means of the coefficients are deemed to be the model coefficients.

The variables used for the estimation are selected following the procedure described in the methodology (section 3.6.3). The correlation matrix of the variables showed that the following variables were highly correlating (bigger than 0.6) with other variables and were therefore removed for the estimation:

- Bouw_adrssn (number of addresses)
- Bouw_mxi (mxi-value of Mozaiek (not recalculated with RUDIFUN))
- Demo_0_15 (percentage of a cell which has residents with age 0 to 15 years)
- Facilities (total amount of facilities in 400m range)
- Geestelijke zorg (amount of facilities with mental health care in 400m range)
- Voedingswinkel (amount of facilities with a food store in 400m range)
- Vuurwerkwinkel (amount of facilities with fireworks store in 400m range)
- FSI (floor space index)

A model with all remaining variables had the lowest AIC, indicating the best model performance. This model is displayed in Appendix G in section 8.7.4. The AIC had much improved compared to the reference model. However, this model consists over 58 variables, which makes it complicated and expensive for practical usage. Therefore, a model with a slightly higher AIC might be better. The model might perform slightly worse but with way fewer variables. Forward, backward and both-direction search was applied to find this trade-off, as explained before in section 3.6.3.

The backward search started with a model with all variables, but as described in the previous paragraph, the full model had the best fit despite the many variables. Hence, the backward search did not show another result.

On the contrary, the forward and both direction search showed a better decrease of variables. The starting variables were defined with the distance (as it was expected to have the highest effect) and the RUDIFUN-areas (as those are easy to gather and expected to have a high effect). Figure 25 shows the results of the forward search with the AIC against the number of parameters and the change in AIC. The results of the both-direction search are displayed in Appendix G in section 8.7.3.

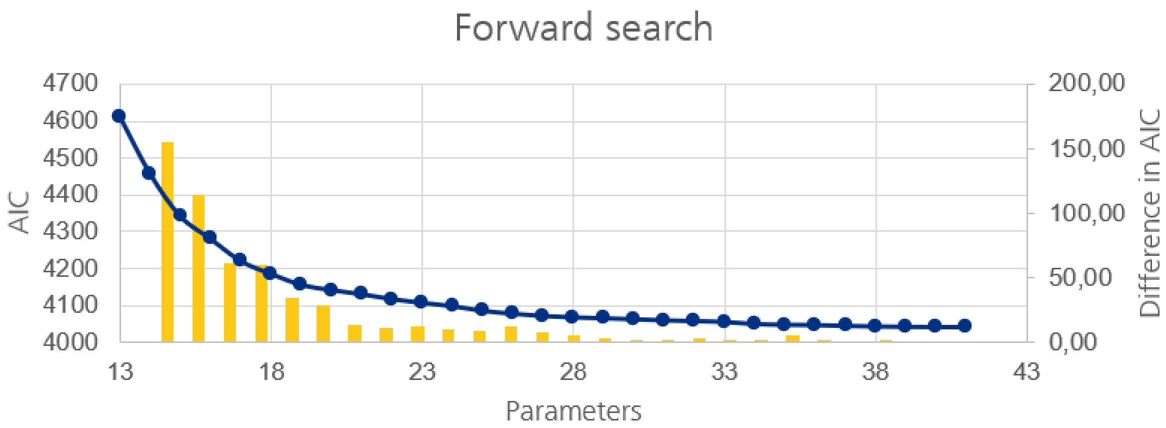


Figure 25: Forward search of parameters

The figure shows that until adding parameter 19 (which was No. meeting buildings) shows the last big drop in AIC value. Therefore, it was chosen to select a model with 19 variables as determined by the forward search. The both-direction search showed similar results, as the same adding order of variables was followed. The both direction search differed from the forward search in later stages when 32 variables were already in the model. As the decrease of AIC was minor, the both-direction had no extra value compared to a forward search. Following the forward search, the model with 19 parameters was considered for further steps.

Table 15 and Table 16 show the estimation for the reference model (with no variables) and the model with the selected 19 variables. The model was able to predict a probability distribution for destinations to be chosen by OV-fiets users.

Table 15: Reference destination binary logit model (or null model)

Variable	Beta	Standard Deviation	t-value	p-value	Sig.
Constant	-0,005	0,017	-3,128	1,998	

Table 16: Destination binary logit model

Variable	Beta	Standard Deviation	t-value	p-value	Sig.
Constant	-2,181 E+00	9,522E-02	-2,290E+02	1,150E-115	***
Area Living function	-6,448 E-05	1,597E-05	-4,038E+01	1,974E-55	***
Area Meeting function	1,882 E-04	1,388E-04	1,356E+01	1,498E-22	***
Area Healthcare function	1,569 E-04	7,047E-05	2,227E+01	2,790E-36	***
Area Industry function	1,431 E-04	3,883E-05	3,685E+01	2,231E-52	***
Area Office function	4,865 E-05	2,902E-05	1,677E+01	4,565E-28	***
Area Accommodation function	6,918 E-05	1,203E-04	5,752E+00	1,511E-07	***
Area Education function	1,938 E-04	1,072E-04	1,808E+01	3,612E-30	***
Area Sports function	3,108 E-04	2,139E-04	1,453E+01	2,887E-24	***
Area Shopping function	-9,875 E-05	1,037E-04	-9,521E+00	7,444E-15	***
Area Other	-2,941 E-04	1,193E-04	-2,465E+01	2,107E-39	***
Area Annexes	1,696 E-04	9,883E-05	1,716E+01	1,064E-28	***
Distance	1,354 E+04	4,030E+02	3,359E+02	3,988E-129	***
No. Households with children	-3,444 E-02	2,920E-03	-1,180E+02	2,113E-92	***
No. KvK registrations	2,051 E-02	1,338E-02	1,532E+01	1,194E-25	***
No. Bus stops (in 400m)	1,820 E-01	2,022E-02	9,000E+01	5,911E-83	***
No. Jewellery stores (in 400m)	2,279 E-01	3,150E-02	7,235E+01	2,288E-75	***
No. Layers	1,315 E-01	2,260E-02	5,816E+01	7,870E-68	***
No. Meeting buildings	1,344 E-01	7,714E-02	1,742E+01	4,009E-29	***

A boxplot of the AIC value of the two models is displayed in Figure 26. The figure shows that the AIC of the estimated model has smaller distribution compared to the reference model. This indicates that the estimation is more consistent than the reference model with each iteration with the resampling, and the reference model is more influenced by the randomness of the synthetic data. Furthermore, the average AIC of the estimated model (4281.94) is much lower than the AIC of the reference model (8074.84).

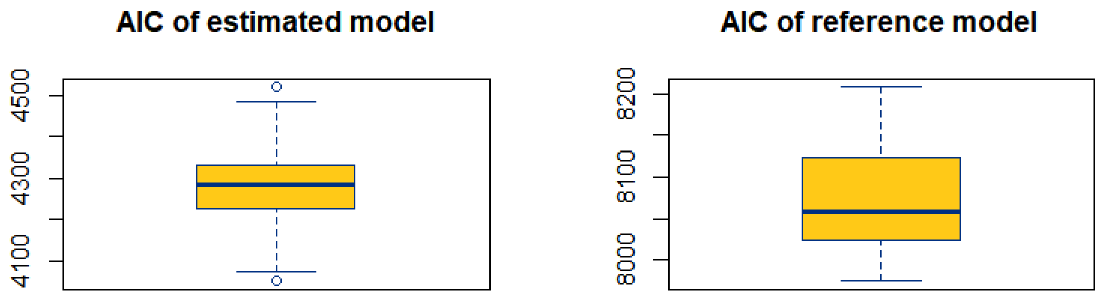


Figure 26: Boxplot of AIC from estimated model and reference model; Remark: Please note that the scale on the y-axis is different

The model structure of being a generalised linear model means that the model is additive and allows for a presence of a phenomenon to add or reduce a certain amount to the total probability. This means that a parameter effect is easily interpretable, as a single parameter has a direct effect on the probability. Therefore, the parameters can be described individually, although the model has many parameters. Here, the model's biggest and most notable parameters are described per group. Positive coefficients add to the odds of a grid cell being chosen, whereas negative coefficients reduce the odds of a grid cell being chosen.

The constant is negative compared to other coefficients. This means that if there is negligible data, the grid cell has a low probability of being chosen. The variables number of bus stops or the number of layers have bigger positive coefficients than most other coefficients. This indicates that variables which are connected to a higher density have positive effects on the probability of a grid cell being chosen. In accordance with that, higher density does create more trips.

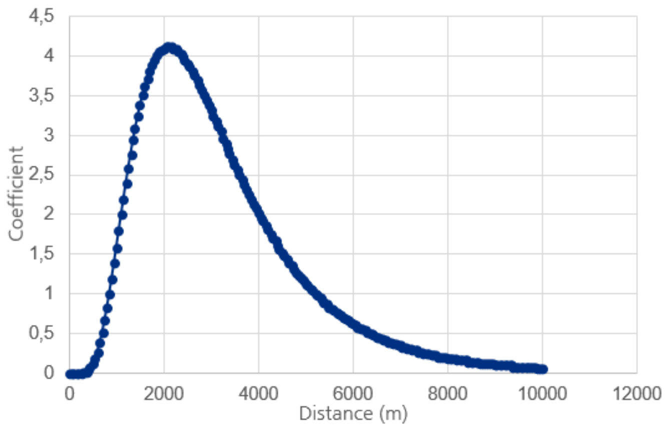
The number of meeting buildings adds to the odds of being chosen for destinations. This is not surprising as the dominant group of respondents in the survey had a working motive and were probably travelling to offices or meeting locations. A more surprising variable is the jewellery stores. This probably happens because jewellery stores are often located in big shopping centres or city centres, which are attractive destinations to choose from due to other factors. It is most likely that jewellery stores are "just a correlation" or by-product, but those can be seen as a summary or indication of a big shopping centre.

The city centres are the most common locations for OV-fiets users in the survey, hence that this effect is visible in the destination model. The demographic variable households with children have a negative effect, which is not surprising, as demographics in the city centres are generally younger and without children. Also, student housing is common in these regions, which is typical for younger people living together and no children living with them.

The proportionated function areas (from the RUDIFUN database) all have a positive effect on the odds of a grid cell being chosen, except for the living area, shopping area and "other" areas. This means that each function area contributes to increasing the probability of choosing that grid cell. This is not surprising as only grid cells with a function area are considered, so each grid cell has a function area. Furthermore, the coefficients are small, meaning that an increase in function area has a small effect on the odds. But the areas are described per square meter, which means that a small coefficient is multiplied by the square meter, leading to a comparable order of magnitude for the effect as to other parameters.

The next positive influence on a destination are the bus stops. The higher the number of bus stops around a grid cell, the higher the probability. This contradicts the literature (see section 2.1.3) as the number of bus stops should negatively affect the destination choice (more bus stops cause more bus usage, which causes less OV-fiets usage). However, the model states otherwise. An explanation for this could be that the bus network in the selected cities is quite dense. Furthermore, city centres and village centres have more bus stops, and the respondents of the survey are mostly going to those places. Additionally, around half of the OV-fiets users in the survey indicated that the bus was their next best option, suggesting that a bus was available (but probably not convenient), and therefore bus stops were present near their destinations.

The last variable is the distance. The distance coefficient is not easy to read as the coefficient is not the standard $y = \beta * x$, but $y = \beta * \left(\frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x_{distance}) - \mu)^2}{2\sigma^2}\right)\right)$ (or the lognormal transformation, see the methodology in section 3.5.4). The beta is denoted as *13536,87* in the estimation, but this connects to a distance on the lognormal function in the range [0;0,0004). Figure 27 shows the translation and, therefore, a more intuitive interpretation of the effect. The figure shows the distribution of the distance and what the model coefficient would do. Interpreting the figure shows that a small distance (< 750 m) has a small effect on the destination probability, while the top (at 2100 m) has the highest effect on the destination probability. The distance effect is larger than the intercept on the interval 1150 to 3800 meters. This means that if all other variables are zero (the data is negligible), the y (and the log-odds) is larger than 0 and therefore, the probability of being chosen is larger than 50%, as a logistic distribution prescribes.



Distance (m)	Coefficient
50	0,000
500	0,118
1000	1,592
2000	4,101
2500	3,912
3000	3,315
3500	2,636
4000	2,023
4500	1,522
5000	1,134

Figure 27: Influence distance on coefficient distance in the destination model

A prediction can be made with the coefficients known from the destination model. Figure 28 shows the area of Eindhoven, and Figure 29 shows the function areas based on the RUDIFUN. The other spatial input variables are displayed in Appendix A in section 8.1.1. The results of the destination model for Eindhoven are displayed in Figure 30. Eindhoven is displayed in the main text as it has no large river, making the prediction clearer at first glance. The results for Arnhem and Amsterdam are displayed in Appendix G in section 8.7.5.

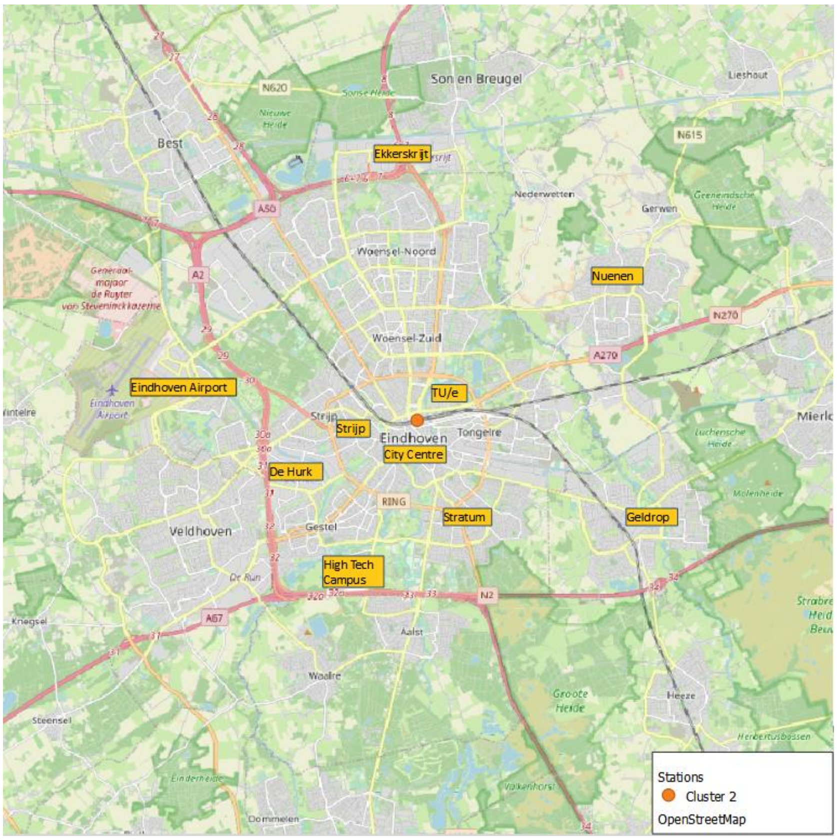


Figure 28: Area of Eindhoven with the most remarkable neighbourhoods

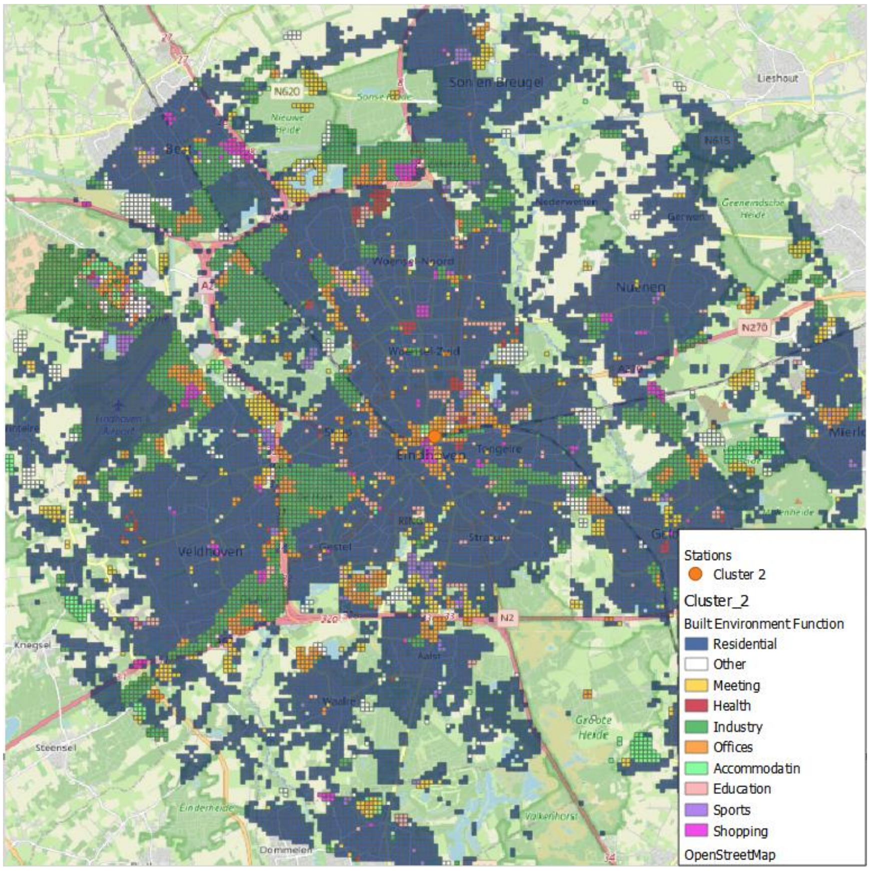


Figure 29: Dominant function type Eindhoven

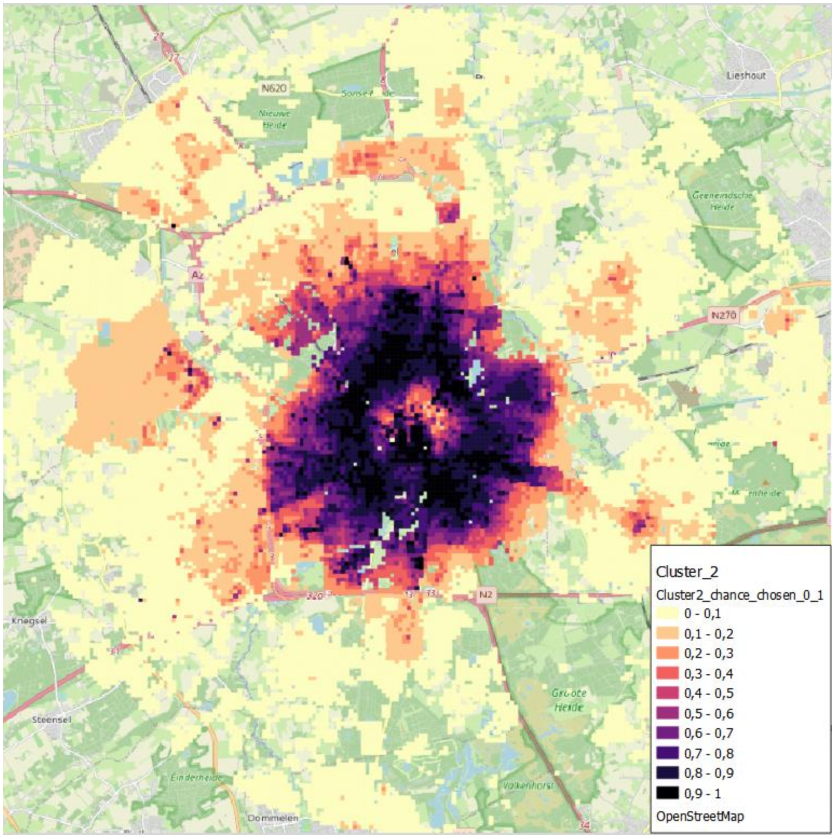


Figure 30: Probability of grid cell being chosen as OV-fiets destination

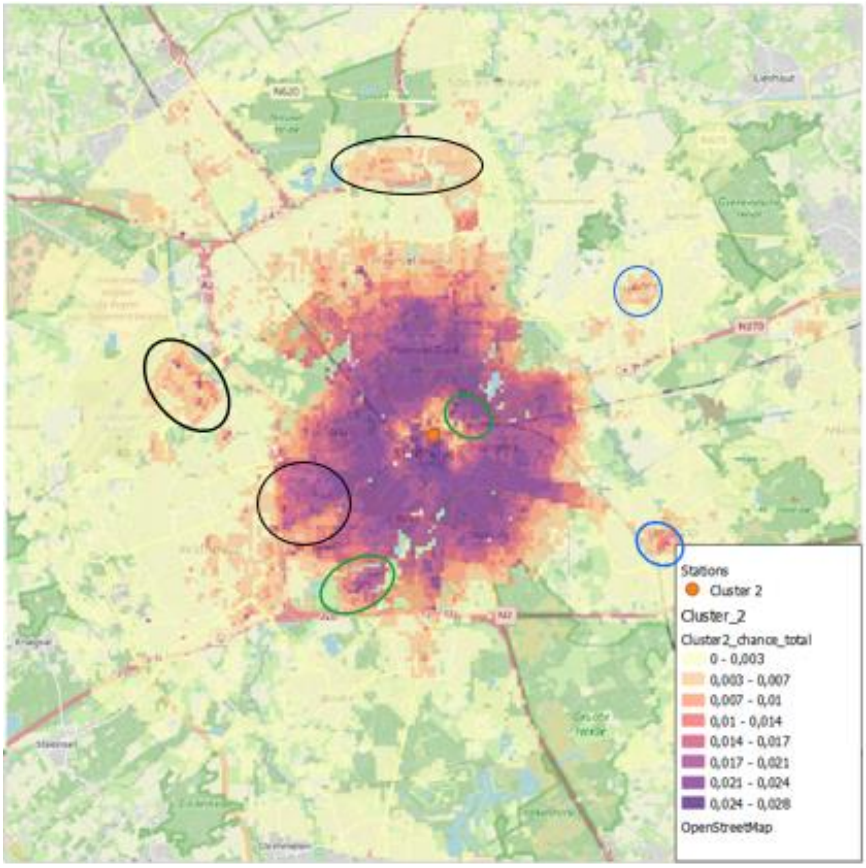


Figure 31: Normalised probability of destination in the Eindhoven area

From Figure 30, some observations can be derived. The first observation is that distance is a major predictor of the probability. Close to the station (which is located in the middle of the city centre), small probabilities are visible, and as the distance increases, so does the probability. The results follow the lognormal distribution of the distance. From Figure 31, the second series of observations are visible. The bigger probabilities occur in the smaller surrounding villages Geldrop and Nuenen (blue circled). As these village centres have many facilities, their attractiveness is higher, and they are visible on the maps. Next, the industry or dense office areas of Ekkerskrijt (north), Strijp/De Hurk (northwest) and Eindhoven Airport (northwest) are also attractive spots (black circles). Together with the city centre, the areas increase the probability of being a destination due to the many facilities in the area. Moreover, the barrier of the highway in the east is visible. The highway creates longer distances in which a cut-off is visible east of De Hurk. Finally, the education areas of the technology campus (south) and the University area near the city centre are often chosen areas (green circles).

4.4.3 Model performance

As described in the methodology in section 3.6.1, the model was based on imbalanced classes. In this way, the metrics used for the purpose model, like accuracy and F1-score, could not be used due to the accuracy paradox. Therefore, this validation used other metrics to validate the model: visual validation and a comparison to complete spatial randomness.

A share of 20% of the survey data was withheld to test the validity, as described for the purpose model in section 5.3.1. This test set can be used to compare the results of the model. Figure 80, Figure 81 and Figure 82 in Appendix G in section 8.7.6 show the test set (blue dots) onto the probability grid of choosing that destination for the rental locations in cluster 2. The destination points lie within the high probability area of the model. This suggests that the model catches the observed locations. Even destinations further away (like the Eindhoven Airport or Amsterdam North) fall in higher probabilities, which would indicate that the model performs well. However, this validation can only be done visually and does not guarantee a valid validation.

Therefore, complete spatial randomness (CSR) is used to quantify the visual validation. CSR is a classical null model for spatial analysis (Delaney et al., 2012). It allows the model to be tested on randomness by comparing a random set of points (Waller & Jacquez, 1995). If the model is better than random, a random set of points should have significant lower predicted values than the validation set. To do so, a set of random points was created with the same geospatial extent as the test set with the survey locations (see Appendix G in section 8.7.6). After that, the test set and the random points were joined by location on the destination model results. This resulted in a predicted probability per location for each point. Thereafter, a t-test could be performed to compare the two sets on the difference.

Table 17 shows the t-test between the datasets. The significance shows that the sets are not similar with 99% confidence. Therefore, it can be concluded that the predicted probabilities of the test set are not the same as random points, and thus, the model does not create a random output.

Table 17: T-test between the 20% test set and random points

	Survey locations 20% sample	Random points
Mean	0.0162	0.0103
Variance	4.508E-05	7.314E-05
Observations (possible joins)	36	30
Degrees of freedom	53	
T-value	2.964	
p-value	0.009	
Significance	p<0,01: ***	

4.5 Model combination

The combination model was estimated with the results of the purpose and destination models (as described in section 3.7.2). Section 4.5.1 shows the input data and, therefore, the relation between the purpose and destination models. Section 4.5.2 show the results, followed by a model performance analysis in section 4.5.3. Finally, the model results are compared with an external source in section 4.6.1.

The model can predict the probability of a single person choosing a single grid cell. Furthermore, the number of trips a grid cell receives can be estimated when running the model for multiple individuals. As the model is a result of the previous models, it was developed with the same 80% of the survey data.

4.5.1 Input data

The input data of the model are the purpose and destination models. An individual could be assigned a purpose based on its characteristics with the purpose model, and attractive destinations were identified with the destination model. The combination of the two probabilities can be done while people go to a place where they can do their activity, as described in section 2.1.2.

Table 18 shows how the purpose and the area are connected. For example, the work purpose is connected to the function area offices, and the education area is connected to the function area education. A more special connection is the leisure connection. Earlier, a decision was made to combine multiple purpose categories into one category to ensure that the class had enough data. This results in many function types corresponding to that purpose. Therefore, many area types are connected to leisure purposes.

Table 18: Translation table for combination model

Purpose survey	Purpose ‘purpose model’	Area function
Work	Work	Area Offices + Area industry
Business trip	Work	Area Offices
Education	Education	Area Education
Visits	Leisure	Area Residential + Area Health
Shopping	Leisure	Area Shop
Leisure	Leisure	Area Meeting + Area Accommodation
Sports	Leisure	Area Sport
Other	Leisure	Area Other

4.5.2 Results

The results of the combination model only allow for certain analyses. Figure 32 shows the probability of a grid cell as a destination for a *young student*. The student has a high probability of having an education or leisure purpose. As a result, the probabilities are relatively high for the residential and educational areas and relatively lower for the office areas. Clear differences against Figure 31 are visible in those office and industrial areas, especially in the north for Ekkerskrijt, Eindhoven Airport, and De Hurk in the west.

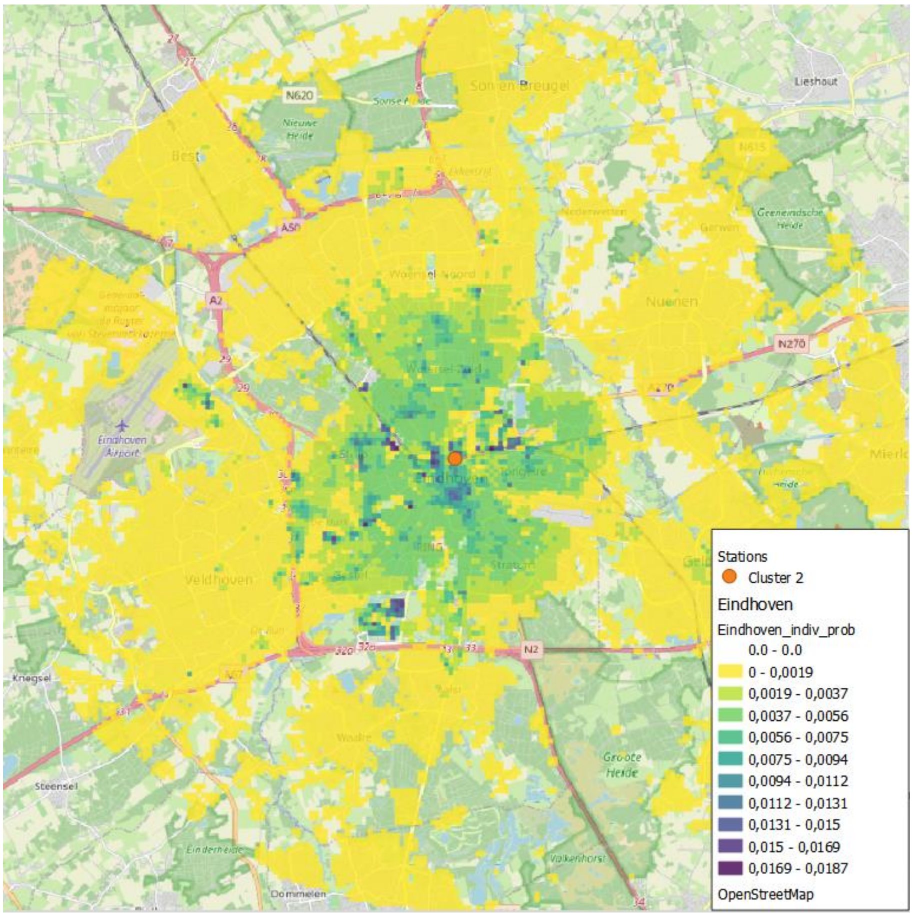


Figure 32: Eindhoven, Probability of grid cell as a destination for the young student

Next, the prediction is made for a set of people. One thousand persons were generated with characteristics based on the survey. Then, for each generated person, the spatial probability distribution was calculated, and one destination was drawn from it. With the 1000-person set, it shows what attractive areas are and what the combination model is capable of doing. The results of this simulation are displayed in Figure 33.

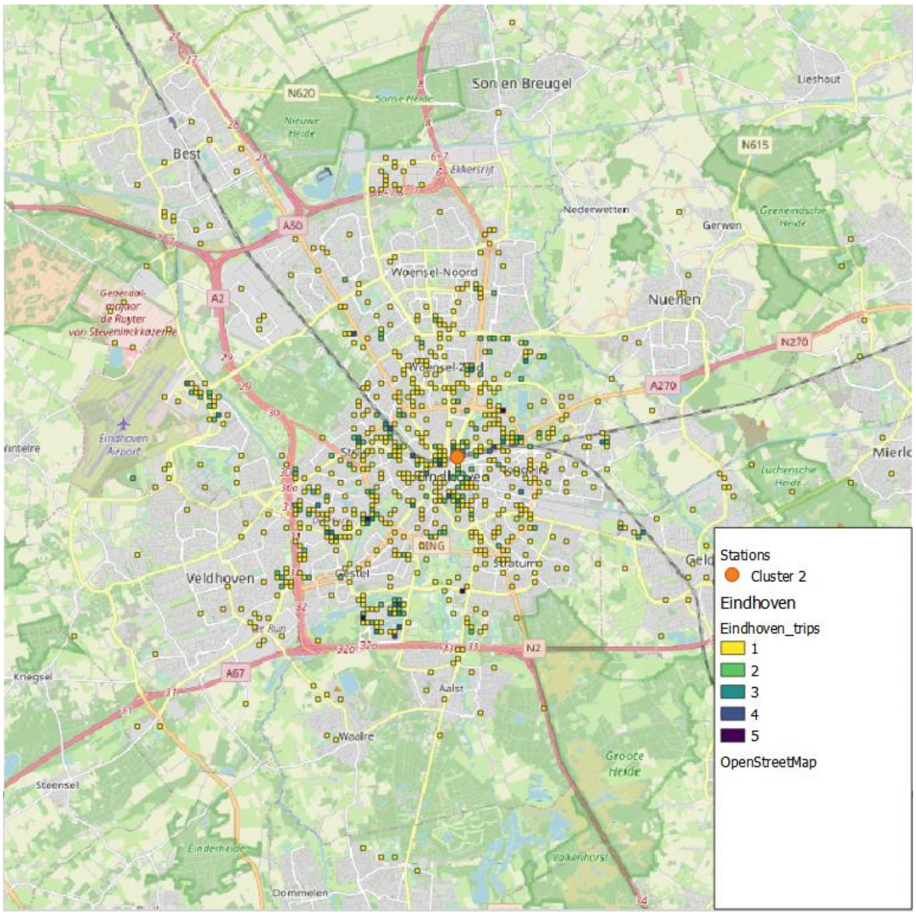


Figure 33: Eindhoven, simulated trips

The figure shows the received trips per grid cell. Appendix H in section 8.8.1 shows the for the other locations in cluster 2. With this, the distribution of the trips in the environment is visible. It shows for the city of Eindhoven that the high-tech campus area, the industrial area De Hurk, the offices area Strijp and the airport act as attraction clusters. Furthermore, in the city centre close to the station, many trips are present. This accentuates the statements on areas with more attraction than others, described above in section 4.4.2.

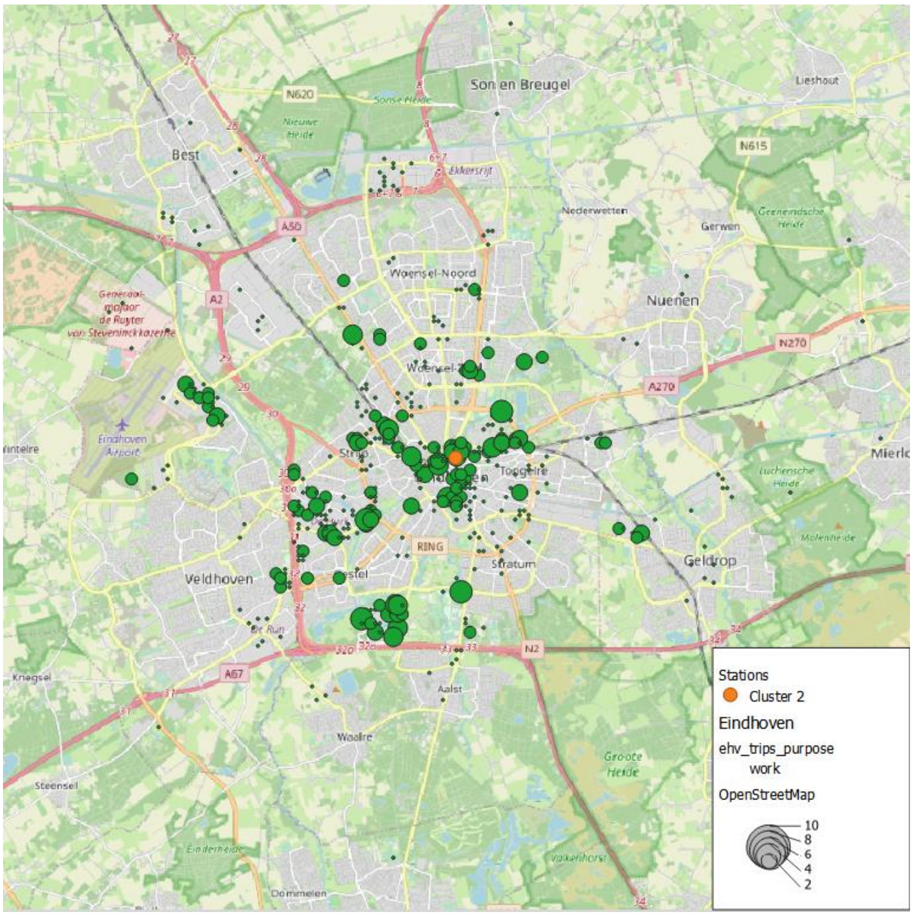


Figure 34: Eindhoven, simulated trips with work purpose

The received trips can be divided into the different purposes, see Figure 34 for work purposed trips. The other trip purposes are displayed in Appendix H in section 8.8.2. This figure shows the direct link of the work purpose to the built environment. Trip clusters are visible around Strijp/De Hurk in the west, Airport Eindhoven (northwest), the city centre (west of the station), the university (TU/e) (northeast of the station) and the high-tech campus (south). These locations are mostly office areas or have many facilities. The number of trips and the direct connection to the office and industry makes those areas more favourable for the simulated trip.

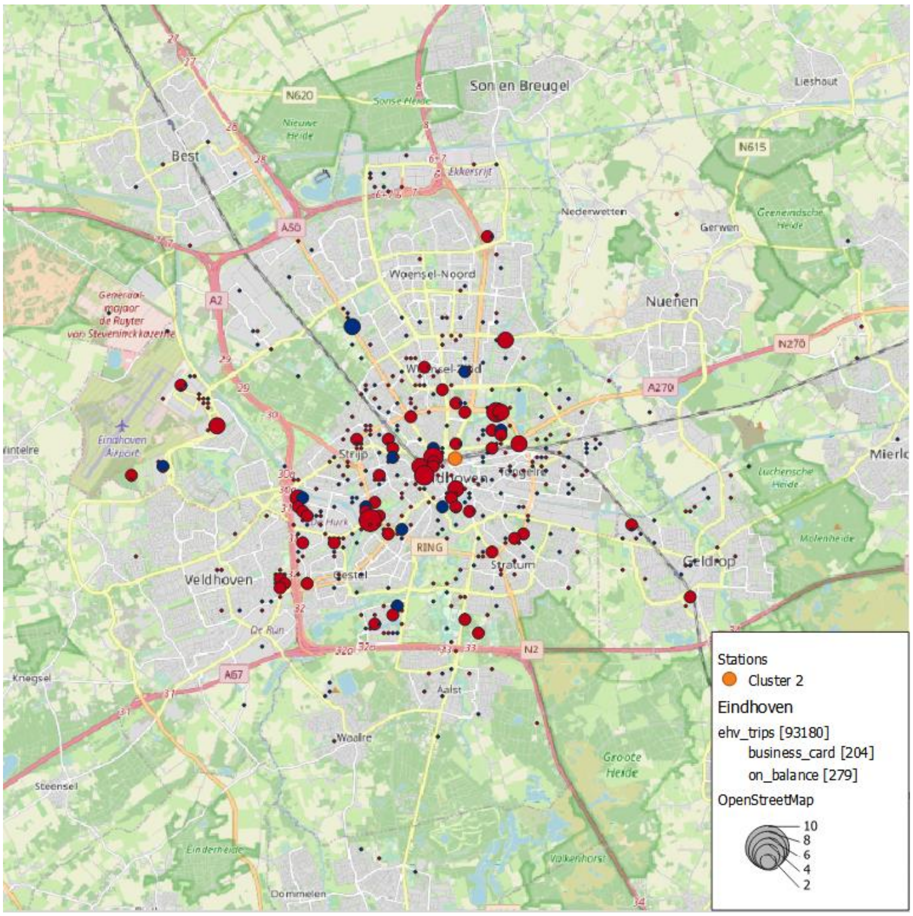


Figure 35: Eindhoven, simulated trips with Business OV-Chipkaart or On Balance OV-Chipkaart

The different OV-chip card types can also be accumulated from the simulated trips. Each simulated trip has a card-type characteristic based on the survey. Although no direct connection is modelled from one card type towards a destination, their patterns might be interesting to analyse. For example, a student travel card is expected to have an educational purpose that leads to those locations. Figure 35 shows the differences in destinations chosen by different card types in the trip simulation (the other card types are displayed in Appendix H in section 8.8.3. It shows that business cards and NS flex destinations are scattered throughout the whole area. A second observation is that the business card OV-fiets users have a larger amount of unique locations than on-balance OV-fiets users, although they had a similar absolute amount of trips in the simulation (both around 30% share).

4.5.3 Model performance

The model can be tested on its performance in a similar way as the destination model. The accuracy metrics cannot be used due to the class imbalance. In contrast to the destination model, the withheld 20% of the data can be modelled. The test set consists of the characteristics and the destinations of respondents. Therefore, this data can be used in the model as input, and the results can be compared with the real answers of the respondents. Furthermore, the results can be compared with random points to assess the complete spatial randomness (the same procedure as used with the destination model).

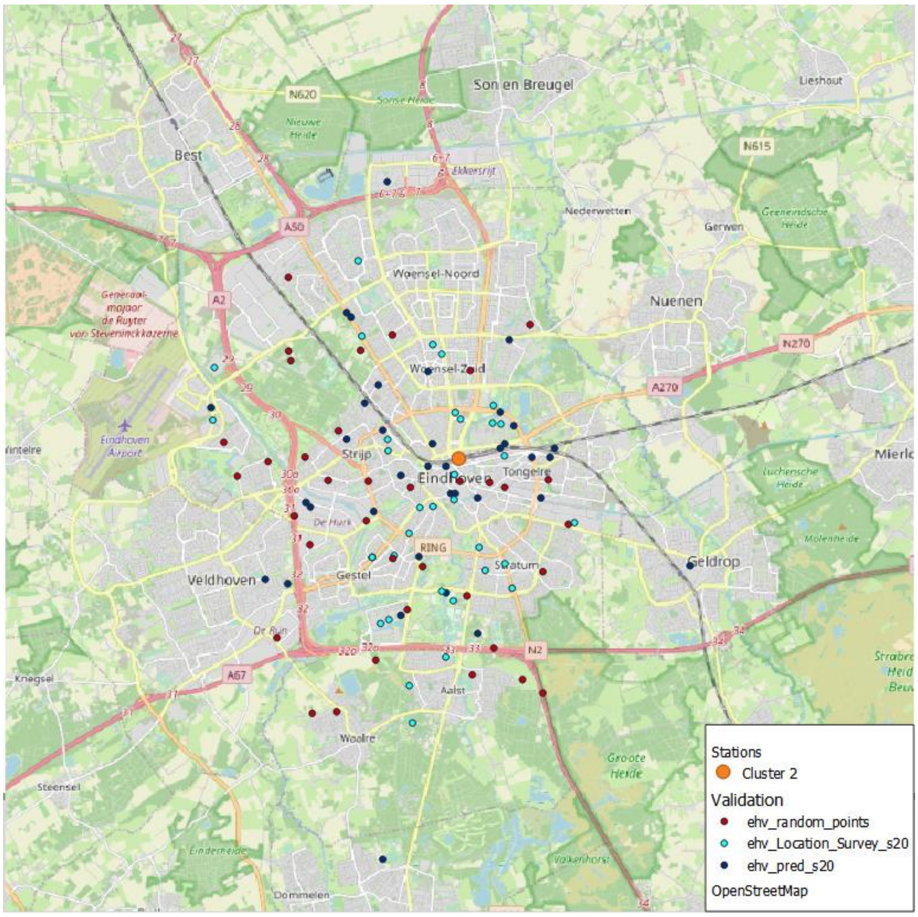


Figure 36: Eindhoven, results of combination model with the test set, 'raw' test set and random points

Figure 36 shows the three point sets. The light blue point set shows the other 20% of the survey locations, the blue points show the prediction of that test set, and the red points are the random points generated to the same extent. The Euclidian distances are calculated from the prediction towards its 'raw' datapoint and to a random point. For example, respondent 1 has a location, and this respondent also has characteristics. These characteristics allow an OV-fiets destination to be calculated and plotted with the combination model. The distances between the answered location (light blue dot) and predicted location (blue dot) is calculated. Next, the distance between the predicted location and a random location (red dot) is calculated. Doing this procedure for all points, a t-test can be performed to determine if the distances are significantly different and assess if the model performs better than a random model.

Table 19 shows the results of the t-test. The p-value lies between 0,01 and 0,05; which indicates that the distances are different with 95% confidence. This shows that the model performs significantly better than a random model or random points would.

Table 19: T-test of distances between the test set to the predicted test set and random points to the predicted test set

	Distance to survey sample	Distance to random points
Mean	3464.246	4462.688
Variance	2.427E+06	4.714E+06
Observations	36	33
Degrees of freedom	58	
T-value	-2.177	
p-value	0.034	
Significance	p<0,05: **	

With the results of the combination model, the fourth sub-research question can be answered. Attractive destinations of OV-fiets users are identified, and it is possible with a fictive individual group to find their destinations. Moreover, the results show that the models can make significantly better predictions compared to randomness.

4.6 Implications and limitations

With estimating models, limitations and implications occur. The following sections describe a small sensitivity analysis for both models. Section 4.6.1 describes the purpose model, section 4.6.2 the destination model, and section 4.6.3 describes the results compared to another data source.

4.6.1 Purpose model sensitivity

The sensitivity of the purpose model can be found by reweighing the input data. The collected data are representative for OV-fiets weekday trips during peak hours. The data shows many work-purposed trips, which is not surprising, as most of the respondents are commuters.

NS has an internal dataset with characteristics of train passengers (KIS10) (NS, 2019). For each station, the purposes of train travellers were identified, so for both week- and weekend days. The KIS10 dataset is the most recent dataset within NS, however, it was before the corona-pandemic. Therefore, the dataset might be outdated, especially with changing traveller patterns as indicated by NS themselves (NS, 2022b). Despite the probably outdated data, it can still be used to check the sensitivity of the trip purpose model.

The KIS10 dataset represented an aggregate of persons. This means that seeing a person has a probability for every purpose. This purpose probability can be different for a specific trip frequency. The probabilities on a trip level are unknown in the NS egress data, hence, the weighing could only be done on a person-level. As a result, the collected survey data were weighted using that dataset to see if the estimated model changes and check the sensitivity to the input data purpose distribution. This means that every instance of the dataset receives a weight factor based on a person's characteristic.

Table 20 shows the respective purpose found in the survey, the share of train travellers with that purpose and the weight factor. The most noticeable matter is that the education purpose gets a high weight factor due to the lower representation in the survey.

Table 20: Weigh factor for sensitivity testing

Purpose	Share of survey	Share of KIS10 (Train travellers)	Weight factor for respondent
Work	50 %	41 %	0.82
Education	6 %	24 %	3.76
Leisure	23 %	35 %	1.53

The reweighted input data was used to re-estimate the purpose model, for which the results are displayed in Appendix F in section 8.6.2. These results show that there are no major changes in coefficients compared to the original model, although many variables increase in significance. However, the AIC is much lower for this model, which indicates a poorer model fit. This might indicate that other variables outside the research scope should have been used for the adjusted input data.

In conclusion, a reweighed input data set led to a poorer fitted model but with more significant variables. On the other hand, the coefficients did not change very much compared to the initial model. Therefore, the collected data of this research were shown to be representative of the weekday peak hours. Additionally, the model coefficients did not change much after reweighing the data, reinforcing that these coefficients were valid. This led to the decision to not reweigh the collected data and work with the representative survey data for the other research steps.

4.6.2 Destination model limitations

The destination model was estimated with 1000 runs to reduce the random impact of the resampling techniques. The mean of the runs for each coefficient was used as a coefficient in the final model. Additionally, the 1000 runs gave the opportunity to analyse how stable the coefficients are by comparing their values for each run. To do so, the 10th up to the 90th percentile was calculated for each variable, see Appendix G in section 8.7.8. If those are very different, the coefficient is less stable or consistent and is vulnerable to changes in the data. On the other hand, if the percentiles are similar, the coefficient is more stable and, therefore, more reliable in the model.

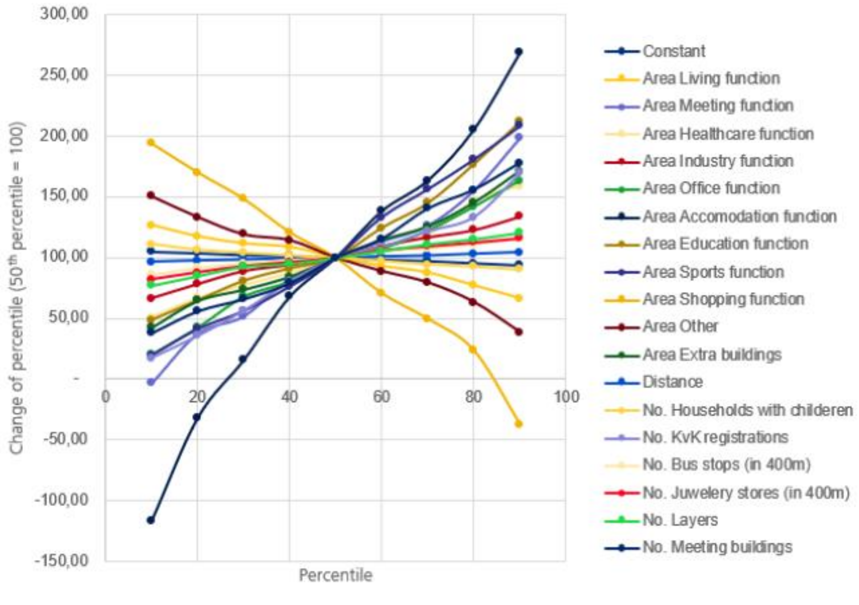


Figure 37: Index number of change in coefficient in 1000 runs per percentile

Figure 37 shows the relative change of a coefficient per percentile. The most noticeable changes occur for the coefficients of Area accommodation function and Area shopping function. Both of these variables also have a smaller effect than the other area function, which might be because these destinations rarely occur. The most consistent or stable coefficient is the distance. This means that in every estimation, the distance has a similar effect on the model.

Another thing to consider in the destination model are destinations with low probabilities. Because the model is additive, different effects add to the total probability of a destination grid cell. So, even a small effect creates a probability, which is visible in the spatial probability distribution maps. Visually, this might be misleading when looking at model results.

4.6.3 Comparison to cycling

The combination model results can also be checked in another way to ascertain if the models make sense. For this, the results are compared to another dataset, which also helps in understanding the correctness of the prediction in a qualitative way. The NS Stations' egress train station data is used to do this analysis (NS Stations, 2022a).

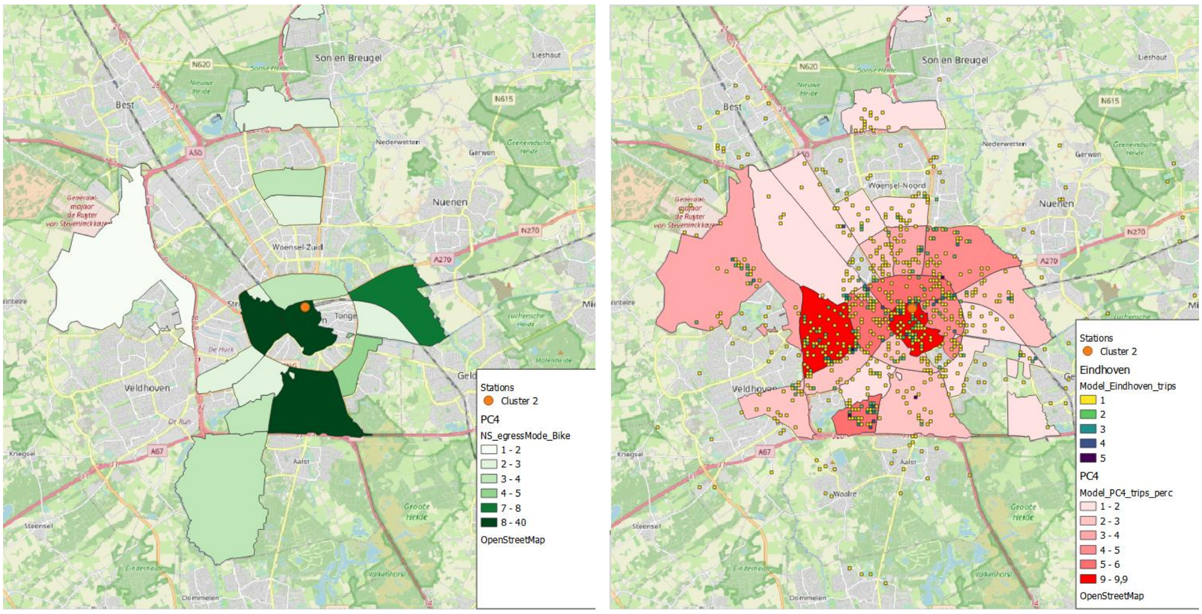


Figure 38a: Bicycle egress modes from Eindhoven CS (>1%), share per PC4-level

Figure 38b: Model prediction OV-fiets trips (>1%), aggregated to PC4-level

Figure 38: Bicycle egress modes and model prediction

Figure 38a shows the bicycle egress share from the NS egress mode data, and Figure 38b shows the model prediction of OV-fiets trips (Appendix H in section 8.8.1 shows a larger figure). The model has a probability for each grid cell, meaning each cell could receive a trip. This results in one trip creating a small percentage and visibility on the map, which is especially the case for larger PC4 areas. Therefore, areas outside city limits or far away from the train station are easily misinterpreted. Thus, the 0 to 1% values have been removed to clarify the analysis for both maps. The bicycle egress share is a combination of private and other bicycles, such as OV-fiets or other BSS. There is no differentiation between those in the NS data, so these are referred to as general bicycles. The legend of the egress data is altered to match the model legend.

The figure shows that the share of bicycle trips is more concentrated in the city centre compared to the model predictions, which are more scattered. This is not surprising, as the model is based on probabilities and a trip simulation. Due to the many low probabilities, a large area size can create a misleading interpretation. Also, people with OV-fiets will not pay for the OV-fiets and cross the station square only, therefore increasing the trip distance.

A second observation is that the egress mode data shows bicycle trip shares to Eindhoven Airport (east) and the High tech Campus (south), which the model favours. The predicted trips are higher for those areas. The Strijp/De Hurk area (west) is a big difference, which receives more trips than the egress data. The trips are more often located to the highway (De Hurk area), whereas the egress data have more trips in the Strijp area.

On the other hand, the data on which the models in this research are based have a higher observation rate than the egress data of NS. This low number of observations is also why PC-4 codes are missing in Figure 38. Therefore, the trips found for the De Hurk area (east) are better represented in this research than in the NS egress data source. This also means that a comparison cannot be made fully.

5 Conclusion

This chapter reports on the conclusions of the research. First, each of the sub-research questions is described. These were formulated in the research approach in section 1.3. After that, the main research question is answered. The main objective of this research was stated as follows:

To what extent can the destinations of OV-fiets users for a specific rental location be predicted based on the individual OV-fiets user's characteristics and the built environment characteristics?

Every sub-question is answered to obtain insights into the objective of this research.

1. What is a suitable topology for OV-fiets rental locations?

The rental locations were clustered with the following characteristics: the first cluster represented trips with a longer duration and a recreative nature, cluster 2 represented commuting, and cluster 3 represented a mixture of destinations that were close to the station or had a short activity period. The rental locations for further research steps were based on their statistical location within the cluster. These were for cluster 1 Groningen, Nijmegen, Maastricht; Arnhem, Eindhoven and Amsterdam Sloterdijk for cluster 2, and Hilversum, Delft and Apeldoorn for cluster 3.

Clusters 2 and 3 were similar in the majority of the user characteristics. This means that the clustering approach did not create fully distinguishable clusters regarding OV-fiets users. The three clusters differed mainly on trip purpose and accompanying characteristics, such as trip frequency. Therefore, the best way to separate the OV-fiets rental locations is for trip purposes.

2. What are the individual characteristics of OV-fiets users, and what are their destinations?

To identify the OV-fiets users, a survey was held at the rental locations during peak hours. Socio-economic characteristics showed that the OV-fiets user is highly educated, young or middle-aged and predominantly employed. Regarding travel patterns, the sample showed a high weekly usage of OV-fiets. The frequencies matched (semi-)daily commuting patterns. The commuting characteristic was also found with other travel-related factors, like reimbursed travel costs. Also, the satisfaction towards public transport is high, which is common among daily public transport users.

The sample showed predominantly work purposes concerning trip characteristics for the main destination. Other trip characteristics were found to have the same tendency: trip frequencies are high to the destination, and mostly one destination is affected. The median distance to a destination is 3 kilometres or 10 minutes trip duration. Many OV-fiets users pay for their bicycle themselves, but many others get it reimbursed by their employer. Finally, most respondents stated that the OV-fiets was chosen above bus/tram/metro due to convenience, freedom and speed.

3. What are the built environment characteristics of OV-fiets users' destinations?

75% of the OV-fiets users collected in the survey indicated their destination. These destinations could be spatially located using Google Maps API's. The locations could be combined with the built environment around the rental locations. The reviewed cities showed a dominance of residential area with industry and office patches around them. Furthermore, small clusters of other functions are situated within the residential area, such as schools. Finally, facilities are most frequent in the city centre and business parks.

4. What are predicted attractive destination areas for an OV-fiets user group based on the built environment characteristics?

The results stated the probability of choosing a destination grid cell increased with the presence of a higher density built environment and, especially, buildings focused on workers. Buildings and areas focused on recreation are less likely to be chosen. The distance was found to be the most dominant effect on whether the destination was chosen or not. Between 1150 to 3800 meters, the distance can solely explain at least 50% of the probability for the destination to be chosen. In other words, there is not much more needed for those destinations to be chosen, such as facilities, while the destination has an attractive distance already.

Destinations of trips done by student travel cards are more clustered, derived from a simulated trip analysis in Eindhoven. The other card types have more scattered destinations throughout the area. Moreover, the business card OV-fiets users have a larger amount of unique locations than on-balance OV-fiets users, although they had a similar absolute amount of trips in the simulation.

5. What are the limitations of the prediction method?

Finally, the OV-fiets destinations are more scattered throughout the area and farther from the station than a private bicycle, as known by NS. This diffuse pattern also had to do with the modelling approach in which each grid cell has a (small) probability. The destination model showed the distance as the most stable predictor, while recreational function areas were found to differ more within the prediction runs.

To what extent can the destinations of OV-fiets users for a specific rental location be predicted based on the individual OV-fiets user's characteristics and the built environment characteristics?

The main objective can be answered with the combined results of the sub-questions. These results show that the OV-fiets destinations for peak hours are influenced at the strongest by the distance to the station. Finally, work-purpose-related buildings influence the destination probability more strongly than recreational destinations. Also, a higher density of the environment creates a more attractive area.

In contrast, accommodation buildings reduce the probability of a destination being chosen by an OV-fiets user. Using the trip purpose-function area connection, it shows that, for example, having a student travel card creates clusters at the University locations, office areas and the city centre.

In short, the research gathered key data on OV-fiets trip distance and trip duration and provided insights on OV-fiets user characteristics and built environment characteristics of OV-fiets user destinations during peak hours.

6 Discussion

This chapter discusses the research output. First, the research limits and possible influences on the outcome are described in section 6.1. Thereafter, potential applications of the results are evaluated in section 6.3. Finally, directions for further research will be described in section 6.4.

6.1 Research limitations

This section describes the potential influences on the results and the limitations of the methodology used in practice. A distinction between three topics is made: the locations and clustering, the survey and the model estimations.

6.1.1 Rental locations

The clustering method to determine the rental locations was able to find clustering means. However, the gap statistic and elbow method could not converge due to computational power limits. Therefore, those methods could not be used to determine the optimal number of clusters. These methods were estimated with fewer stations, but the full analysis could only be done with the silhouette method. Next, it is not certain if the result is applicable to smaller rental locations. Those locations were left out of the analysis due to fewer trips, which creates difficulties in reaching the required sample size for the survey. Wilkesmann (2022) stated that the rental patterns are different for smaller stations, mainly due to the smaller number of bicycles combined with a random error.

6.1.2 Survey

As explained in the methodology (section 3.3), the survey had a bias because only OV-fiets users were filling in the survey. Travellers who might have the same destination but who travel by bus/tram/metro or rental car were not collected by the survey. Therefore, the sample only consists of travellers who have already chosen for the OV-fiets. This should be considered when applying the results in other research.

As explained earlier, the sample showed differences with the limited current data of OV-fiets users known by NS Stations. However, a recent survey was held by NS Stations during this research targeting non-users of OV-fiets. This survey was collected using an independent panel and represented the 18+ population of the Netherlands (NS Stations, 2022b). In that survey, a reference group of OV-fiets users was formulated. The results showed similar results as this research: the OV-fiets users were young, had high trip frequencies, and had work/business trip purposes, while the non-users were older, travelled infrequently, and were more recreational. The report stated that a small proportion of the users make a big share of all trips. A small user group had a daily usage, while the vast majority only used the OV-fiets less than two times per year. These frequencies were similar to the sample found in this research. This means that the collected sample of this research shows correct representatives.

The survey questions had a few limitations as well. First, the questions regarding usage of the OV-fiets for “this trip”, “in general”, and “week or weekend” were confusing for respondents. This was due to the matrix format of those questions. The format was a known risk when creating the survey questions, but due to the survey size, the questions could not be separated into multiple matrices. Furthermore, these questions targeted travel changes due to Covid-19, but many factors could influence travel behaviour. People could have been relocated or changed jobs. All of these influences could have occurred during the pandemic, even though they might not have been caused by it. The causation does not automatically mean the correlation. Therefore, the results on the Covid-implications could be based on many other factors which were not accounted for. However, a general view of travel frequencies is still reliable.

Another question which could be ambiguous was the trip frequency question. The intention of the question was targeted at the frequency of the main destination and trip. This is not an average question for a survey. In general, the total usage of a product is asked. Therefore, people could interpret the question for their general usage due to quick reading and filling in the survey. Although the question was clearly formulated, the words indicating "this destination" might have been more emphasized by using bold or underlined text styles.

Next, the survey question on alternatives did not include the option "walking". This option disappeared somewhere in the process. However, many respondents filled in that walking was an alternative travel mode for them, so it was mentioned in the results. The share of walking as an alternative mode might be underrepresented since other respondents did not see the option, and therefore they are likely not to consider it when filling in the survey.

Finally, the survey location Apeldoorn had automatic locks on the OV-fiets. This reduces the time for renting a bicycle, which is one of the success factors for OV-fiets (Ploeger & Oldenziel, 2020). On the other hand, the lock and its procedure can also be a threshold. New users might experience difficulties with the lock, which was visible during the fieldwork at Apeldoorn. The effects of the automatic lock are unknown and have not been accounted for in the research. Ultimately, Apeldoorn did not show irregular patterns following the survey results. But this is only an indication, while Apeldoorn, on its own, does not have a representative sample (< 400 respondents).

6.1.3 Model

The model estimation also has some limitations and influences due to the built environment data. The model was estimated for cluster 2: Arnhem Centraal, Eindhoven Centraal and Amsterdam Sloterdijk. The station and rental location characteristics are statistically similar. However, the city of Amsterdam might be an outlier in the built environment. Amsterdam is the capital of the Netherlands and is four times larger than Eindhoven and five times larger than Arnhem in terms of residents (CBS, 2021). The function type area is more diverse too. Also, the number of jobs, facilities and tourist attractions are higher. Therefore, the built environment for Amsterdam is an outlier compared to other cities in the Netherlands. This was concluded during the research, while the decision for cluster 2 was already made and could not be changed anymore due to time and budget. However, it is assumed that the results are still valid. In the worst case, the results are underestimated because the built environment values are, on average lower without Amsterdam. An estimation without Amsterdam cause lower density or fewer facilities to predict a higher probability.

Another input variable limitation is geolocating the identified destinations from the survey. Some locations had a clear wrong location located by the Google API. This could be due to dummy answers (like a postal code "1234AB"), an ambiguous location (a street name which occurs in multiple villages) or a wrongly geocoded location. These locations were filtered out manually. The sample size after the selection was still big enough. Therefore, no effort was made to recode the locations manually.

Regarding built environment variables, three issues should be considered. First, the influence of other OV-fiets rental locations was not considered in the analysis. Other OV-fiets rental locations could create another alternative travel mode for the traveller. It is hypothesized that other locations negatively influence destination choice. Especially for Amsterdam, it might be an influence as other locations could be closer to destinations. Second, the bus frequency of a bus stop was not considered. The frequency of a bus line to a train station affects the travel mode decision (as explained in the theoretical framework, section 2.1.3). As the frequency of different bus lines differs over the day, and the data were not fully available, only the bus stop locations were used in the analysis. Third, it is likely that facilities far away from the rental location affect the destination choice differently than those close to the station. Due to research time and model complexity, the interaction effects of the distance with other parameters were not investigated. A more extensive search on these interaction effects might be an interesting direction for further research.

Finally, from the survey respondents, a few destinations were mentioned a significant amount of times. Examples of these are: Rijkskantoor belastingdienst Apeldoorn (Tax office), ASML in Veldhoven (chip factory) and the TU Delft (University). Those destinations probably have a big influence or attraction value on the destination choice by specific rental locations. However, those specific destinations could not be considered in the analysis. This is because these locations have no special built environment characteristics other than being that place. It could be the culture or habits of the personnel which cause those locations to have large OV-fiets use. On the other hand, it is more likely that those big offices/locations have deals with OV-fiets or local transportation, which means that usage of OV-fiets is free for those travellers. As those offices/locations attract extra demand, it is beneficial to do further research to match the supply and demand at those locations. A start in that direction is identifying governmental service offices, like the Belastingdienst offices, DUO office and Provincial offices.

Looking back to the conceptual model presented in section 2.3, especially the density and destination factors of the 7D built environment were major predictors for the destination choice of OV-fiets users. The distance to transport (bus stops) was expected to have a negative effect, although that variable had a positive effect in the analysis. Also, in the final model, not all aspects of the 7D built environment were represented, while not all variables had significant effects or contributions to a significantly lower AIC. Furthermore, the distance was treated as a variable of the destination, not the user trip. This made more sense as trip distances do not easily classify users, but the environment can. Finally, individual preferences were treated as individual characteristics to predict the trip purpose instead of the mode choice. This made sense for this research as the mode choice was fixed but cannot be done when different modes are in play.

6.2 Model results

First, the (trip) purpose model showed a few strange matters, foremost regarding the leisure purpose class. The model outputs showed that having a business card increases the odds for leisure purposes. This is strange, while the employer generally provides a business card for commuting trips. This might be caused by the fact that a business card may also be used for other trips, but those trips might not get reimbursed. To understand this, the other variables should be looked at. The 'pays for OV-fiets' and 'pays for PT' has a big negative effect on the leisure class. This usually occurs when having a business card. However, the correlation displayed in the correlation matrix (section 8.6.1) was not significant to remove one of the variables. An analysis was done with both variables removed. This showed that the business card coefficient did get a negative value, indicating a negative relationship between the business card and leisure. Therefore, when looking at the variable 'business cards' solely, this might be skewed (or card types in general). Who is paying for the PT (employer or the traveller themself) or OV-fiets is more useful.

Next, the results of the purpose model regarding education purpose had a low precision, recall and F1-score. This might be the case due to the relatively low amount of instances and that education is similar to work. A student can go to work or an internship one day and the next day to school. This does not affect their socio-economic characteristics and habits, only the purpose. Therefore, the model did predict many educational trips in the work-purpose class.

Finally, the leisure purpose was difficult to estimate in the combination model. The results were scattered throughout the whole area. This might be because of the relatively low amount of data for leisure trips, which did not create hotspots around recreational areas and facilities. The low amount of leisure trips had to do with the collection times of the survey, which were the peak hours. During these, the number of recreational travellers is low because reduced tariffs are not valid yet, and the commuters dominate the trains.

6.3 Research applications

The research can be used for several applications, which were already described in section 1.5. These were (1) understanding the system better, (2) contributing to trip prognosis models and (3) supply evaluation, (4) land-use planning, (5) transportation planning, (6) (7) SBRT user and system insights and (8) identification contributing factors to destination choice of bikesharing user in a Dutch context. Besides those, a few extra recommendations based on the results can be formulated:

First, the research can be used to reshape the knowledge of the general OV-fiets user for peak hours. This research found another general OV-fiets user than identified in earlier NS data. This might indicate that the peak hour OV-fiets user differs from the standard user or that the earlier data is biased. The recent non-OV-fiets user survey is similar to this research, hence the latter is probably the case (NS Stations, 2022b). This research would help in identifying that bias and re-evaluating the earlier data.

Second, there are models and research on OV-fiets availability prognoses at NS Stations. Currently, these models rely on short-term forecasting based on historical data. The results of this study can help in optimizing these models. Adding attractive destinations and combining them with actual trip numbers can help identify how a determinant or coefficient relates to an OV-fiets trip. Then, it is possible to use the built environment directly to estimate trips, which can be used for availability prognosis and optimization.

Next, the models showed that they could identify big attractive areas. Therefore, they can be used in other locations to identify important destinations. By identifying these destinations, the infrastructure in those areas or the rental location itself can be improved. Also, the division of OV-fiets bicycles between several rental locations at the same train station can be optimized since the models can identify a dominant destination side of the train station.

6.4 Further research directions

Apart from the direct implementation of the research results and models, further research directions are also identified. A first direction would be combining the model with the number of bookings to quantify the connection between the built environment and the user. By doing this, a minimum amount of bicycles can be estimated for a location.

Another research direction is to estimate the models for the other clusters. Especially for cluster 1, this would be interesting since that cluster was significantly different for many user characteristics. The model estimation for other clusters was not researched due to data availability, research time and computational limits.

Furthermore, a research direction could be to estimate the models for just a single trip purpose. This would create the opportunity to compare destinations of different trip motives and the difference in attraction values of different built environment variables. This was not done in this research due to the far lower amount of education and leisure purposes. Let alone the other purposes which respondents could choose.

Another research direction is to investigate the OV-fiets users during other periods, such as the weekend. Wilkesmann (2022) showed that the trip patterns are significantly different for week and weekend days. Trip patterns showed to contribute to having a different trip purpose, which would mean that the destinations are different for the weekend as well. This could be modelled with the same method to compare the results. It would be interesting to see if the destinations are significantly different and if it is useful to have multiple models or if one model can represent all OV-fiets trip destinations.

Finally, a more in-depth identification of destination determinants is a further research direction as well. Interaction effects of variables were ignored in this research, since there were already many variables investigated in the research, interaction effects were deemed to make the research too complicated and set out of scope. However, the research showed that the distance is a major predictor of the destination probability. For example, a certain facility close by to the station has most likely a stronger effect than a certain far away. Therefore, interactions of the distance with facilities is an interesting direction.

7 References

Aalst, J.-W. van. (2021). *Mozaïek Definitiedocument*.

Aarts, H., Verplanken, B., & van Knippenberg, A. (1997). Habit and information use in travel mode choices. *Acta Psychologica*, *96*(1–2), 1–14. https://doi.org/10.1016/S0001-6918(97)00008-5

Adnan, M., Altaf, S., Bellemans, T., Yasar, A. ul H., & Shakshuki, E. M. (2019). Last-mile travel and bicycle sharing system in small/medium sized cities: user's preferences investigation using hybrid choice model. *Journal of Ambient Intelligence and Humanized Computing*, *10*(12), 4721–4731. https://doi.org/10.1007/S12652-018-0849-5/TABLES/5

Allwright, S. (2022, April). *What is a good F1 score? Simply explained*. https://stephenallwright.com/good-f1-score/

Asadi, M., Ulak, M. B., Geurs, K. T., Weijermars, W., & Schepers, P. (2022). A comprehensive analysis of the relationships between the built environment and traffic safety in the Dutch urban areas. *Accident Analysis & Prevention*, *172*, 106683. https://doi.org/https://doi.org/10.1016/j.aap.2022.106683

Bachand-Marleau, J., Lee, B. H. Y., & El-Geneidy, A. M. (2012). Better Understanding of Factors Influencing Likelihood of Using Shared Bicycle Systems and Frequency of Use: *Journal of the Transportation Research Board*, *2314*(2314), 66–71. https://doi.org/10.3141/2314-09

Bahamonde-Birke, F. J. (2021). A brief discussion on the treatment of spatial correlation in multinomial discrete models. *Journal of Transport and Land Use*, *14*(1), 521–535. https://doi.org/10.5198/JTLU.2021.1848

Bekhor, S., & Prashker, J. N. (2008). GEV-based destination choice models that account for unobserved similarities among alternatives. *Transportation Research Part B: Methodological*, *42*(3), 243–262. https://doi.org/10.1016/J.TRB.2007.08.003

Ben-Akiva, M., & Bierlaire, M. (1999). Discrete Choice Methods and their Applications to Short Term Travel Decisions. In *Handbook of Transportation Science* (pp. 5–33). Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-5203-1_2

Ben-Akiva, M., & Lerman, S. R. (1997). *Discrete choice analysis: theory and application to travel demand* (7th ed.). MIT Press series in transportation studies.

Bhat, C. R., & Guo, J. (2004). A mixed spatially correlated logit model: formulation and application to residential choice modeling. *Transportation Research Part B: Methodological*, *38*(2), 147–168. https://doi.org/10.1016/S0191-2615(03)00005-5

Brownson, R. C., Hoehner, C. M., Day, K., Forsyth, A., & Sallis, J. F. (2009). Measuring the Built Environment for Physical Activity State of the Science. *AMEPRE*, *36*, S99-S123.e12. https://doi.org/10.1016/j.amepre.2009.01.005

Bureau Spoorbouwmeester. (2012). *Het Stationsconcept Visie en toepassing*. www.spoorbeeld.nl.

CBS. (2020). *Kaart van 100 meter bij 100 meter met statistieken*. CBS. https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/kaart-van-100-meter-bij-100-meter-met-statistieken

CBS. (2021). *Inwoners per gemeente*. CBS. https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/regionaal/inwoners

Cervero, R., Caldwell, B., & Cuellar, J. (2013). Bike-and-Ride: Build It and They Will Come. *Journal of Public Transportation*, *16*(4), 5. https://doi.org/http://doi.org/10.5038/2375-0901.16.4.5

Cervero, R., Sarmiento, O. L., Jacoby, E., Gomez, L. F., & Neiman, A. (2009). Influences of built environments on walking and cycling: Lessons from Bogotá. *International Journal of Sustainable Transportation*, *3*(4), 203–226. https://doi.org/10.1080/15568310802178314

Chakour, V., & Eluru, N. (2014). Analyzing commuter train user behavior: A decision framework for access mode and station choice. *Transportation*, *41*(1), 211–228. https://doi.org/10.1007/S11116-013-9509-Y/TABLES/5

Chawla, N. v, Bowyer, K., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Article in Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, F., Turon, K., Klos, M., & Slerpinski, G. (2018). Fifth generation of bike-sharing systems: examples from Poland and China. *Scientific Journal of Silesian University of Technology. Series Transport*, *99*, 5–13. https://doi.org/10.20858/sjsutst.2018.99.1

Chowdhury, S., la Paix, L., & Geurs, K. (2020). Inter-and intrapersonal variation in destination choice. *European Journal of Transport and Infrastructure Research*, *20*(4), 194–213. https://doi.org/10.18757/ejtir.2020.20.4.3951

de Vos, J. (2020). The effect of COVID-19 and subsequent social distancing on travel behavior. *Transportation Research Interdisciplinary Perspectives*, *5*, 100121. https://doi.org/10.1016/J.TRIP.2020.100121

Delaney, D. G., Edwards, P. K., & Leung, B. (2012). Predicting regional spread of non-native species using oceanographic models: validation and identification of gaps. *Marine Biology*, *159*, 269–282. https://doi.org/10.1007/s00227-011-1805-5

Delignette-Muller, M. L. (2014). *fitdistrplus: An R Package for Fitting Distributions*. http://cran.r-project.org/package=fitdistrplus.

DeMaio, P. (2009). Bike-sharing: History, Impacts, Models of Provision, and Future. *Journal of Public Transportation*, *12*(4), 3. https://doi.org/http://doi.org/10.5038/2375-0901.12.4.3

DeMaio, P., Yu, C., O'Brien, O., Rabello, R., Chou, S., & Benicchio, T. (2021). *The Meddin Bike-sharing World Map*. https://bikesharingworldmap.com/

Dijkgraaf, J. (2021, September 21). *Deelvervoer in Twente neemt vlucht: 'Binnen 3 minuten is er straks altijd een optie.'* Tubantia.Nl. https://www.tubantia.nl/enschede/deelvervoer-in-twente-neemt-vlucht-binnen-3-minuten-is-er-straks-altijd-een-optie~a3858c0f/

Dijkstra, E. W. (1959). A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, *1*, 269–271.

Donkey Republic. (2021). *Prijzen*. Donkey Republic. https://www.donkey.bike/nl/prijzen/

Duran-Rodas, D., Wright, B., Pereira, F. C., & Wulfhorst, G. (2021). Demand And/oR Equity (DARE) method for planning bike-sharing. *Transportation Research Part D: Transport and Environment*, *97*, 102914. https://doi.org/10.1016/J.TRD.2021.102914

Efthymiou, D., Antoniou, C., & Waddell, P. (2013). Factors affecting the adoption of vehicle sharing systems by young drivers. *Transport Policy*, *29*, 64–73. https://doi.org/10.1016/J.TRANPOL.2013.04.009

Eren, E., & Uz, V. E. (2020). A review on bike-sharing: The factors affecting bike-sharing demand. *Sustainable Cities and Society*, *54*, 101882. https://doi.org/10.1016/J.SCS.2019.101882

ESRI. (2022). *Algorithms used by the ArcGIS Network Analyst extension*. Desktop.Arcgis.Com. https://desktop.arcgis.com/en/arcmap/latest/extensions/network-analyst/algorithms-used-by-network-analyst.htm#GUID-D5832F80-174E-4E56-90F6-0BBADCBA7777

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5th ed.). John Wiley & Sons, Ltd.

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association*, *76*(3), 265–294. https://doi.org/10.1080/01944361003766766

Ewing, R., Greenwald, M. J., Zhang, M., Walters, J., Feldman, M., Cervero, R., & Thomas, J. (2009). Measuring the impact of urban form and transit access on mixed use site trip generation rates—Portland pilot study. *US Environmental Protection Agency*.

Faghih-Imani, A., & Eluru, N. (2015). Analysing bicycle-sharing system user destination choice preferences: Chicago's Divvy system. *Journal of Transport Geography*, *44*, 53–64. https://doi.org/10.1016/J.JTRANGEO.2015.03.005

Faghih-Imani, A., Hampshire, R., Marla, L., & Eluru, N. (2017). An empirical analysis of bike sharing usage and rebalancing: Evidence from Barcelona and Seville. *Transportation Research Part A: Policy and Practice*, *97*. https://doi.org/10.1016/j.tra.2016.12.007

Fishman, E. (2016). Cycling as transport. *Transport Reviews*, *36*(1), 1–8. https://doi.org/10.1080/01441647.2015.1114271

Fotheringham, A. S., Nakaya, T., Yano, K., & Openshaw, S. (2001). Hierarchical destination choice and spatial interaction modelling: a simulation experiment. *Environment and Planning A*, *33*, 901–920. https://doi.org/10.1068/a33136

Gao, K., Yang, Y., Li, A., & Qu, X. (2021). Spatial heterogeneity in distance decay of using bike sharing: An empirical large-scale analysis in Shanghai. *Transportation Research Part D: Transport and Environment*, *94*. https://doi.org/10.1016/j.trd.2021.102814

Garín-Muñoz, T. (2009). Tourism in Galicia: Domestic and Foreign Demand: *Transport Economics*, *15*(4), 753–769. https://doi.org/10.5367/000000009789955107

Geurs, K. T., la Paix, L., & van Weperen, S. (2016). A multi-modal network approach to model public transport accessibility impacts of bicycle-train integration policies. *European Transport Research Review*, *8*(4), 1–15. https://doi.org/10.1007/S12544-016-0212-X/TABLES/6

Geurs, K. T., & van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography*, *12*(2). https://doi.org/10.1016/j.jtrangeo.2003.10.005

Google Maps. (2022a). *Google Maps Platform Documentation | Directions API | Google Developers*. https://developers.google.com/maps/documentation/directions

Google Maps. (2022b). *Google Maps Platform Documentation | Places API | Google Developers*. https://developers.google.com/maps/documentation/places/web-service

Grasso, S. H., Barnes, P., Chavis, C., Grasso, S. H., Barnes, P., & Chavis, C. (2020). Bike Share Equity for Underrepresented Groups: Analyzing Barriers to System Usage in Baltimore, Maryland. *Sustainability*, *12*(18), 1–17. https://EconPapers.repec.org/RePEc:gam:jsusta:v:12:y:2020:i:18:p:7600-:d:413870

Gu, T., Kim, I., & Currie, G. (2019a). To be or not to be dockless: Empirical analysis of dockless bikeshare development in China. *Transportation Research Part A: Policy and Practice*, *119*, 122–147. https://doi.org/10.1016/J.TRA.2018.11.007

Gu, T., Kim, I., & Currie, G. (2019b). Measuring immediate impacts of a new mass transit system on an existing bike-share system in China. *Transportation Research Part A: Policy and Practice*, *124*, 20–39. https://doi.org/10.1016/J.TRA.2019.03.003

Guo, Y., & He, S. Y. (2020). Built environment effects on the integration of dockless bike-sharing and the metro. *Transportation Research Part D: Transport and Environment*, *83*, 102335. https://doi.org/10.1016/J.TRD.2020.102335

Hand, D., & Christen, P. (2018). A note on using the F-measure for evaluating record linkage algorithms. *Statistics and Computing*, *28*(3), 539–547. https://doi.org/10.1007/S11222-017-9746-6

Harbers, A., Spoon, M., Amsterdam, H. van, & Schuit, J. van der. (2019). *Ruimtelijke Dichtheden en Functiemenging in Nederland (RUDIFUN)*.

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. *International Joint Conference on Neural Networks*.

Hensher, D. A., & Button, K. J. (2008). *Handbook of Transport Modelling* (2nd ed.). Emerald, Inc.

Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). *Applied Choice Analysis* (2nd ed.). University printing house.

Jäppinen, S., Toivonen, T., & Salonen, M. (2013). Modelling the potential effect of shared bicycles on public transport travel times in Greater Helsinki: An open data approach. *Applied Geography*, *43*, 13–24. https://doi.org/10.1016/J.APGEOG.2013.05.010

Jin, P. J., Cebelak, M., Yang, F., Zhang, J., Walton, C. M., & Ran, B. (2014). Location-Based Social Networking Data: Exploration into Use of Doubly Constrained Gravity Model for Origin–Destination Estimation. *Journal of the Transportation Research Board*, *2430*(1), 72–82. https://doi.org/10.3141/2430-08

Jones, P. (1990). Developments in dynamic and activity-based approaches to travel analysis. *Oxford Studies in Transport.*, xiii, 465 p.

Jonkeren, O., Kager, R., Harms, L., & te Brömmelstroet, M. (2021). The bicycle-train travellers in the Netherlands: personal profiles and travel choices. *Transportation*, *48*(1), 455–476. https://doi.org/10.1007/S11116-019-10061-3/TABLES/8

Kadaster. (2018). *Gebruikshandleiding BAG*. Kadaster. https://imbag.github.io/praktijkhandleiding/

Kadaster. (2022). *Basisregistratie Grootschalige Topografie (BGT)*. Ministerie van Infrastructuur & Waterstaat. https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/stelsel-van-basisregistraties/10-basisregistraties/bgt/

Kanellopoulos, D., Kotsiantis, S., & Pintelas, P. (2006). Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering*, *30*.

Keijer, M. J. N., & Rietveld, P. (1999). *How do people get to the railway station : a spatial analysis of the first and the last part of multimodal trips*.

Khuri, A. I., Mukherjee, B., Sinha, B. K., & Ghosh, M. (2006). Design issues for generalized linear models: A review. In *Statistical Science* (Vol. 21, Issue 3, pp. 376–399). https://doi.org/10.1214/088342306000000105

Kjaer, T. (2005). *A review of the discrete choice experiment - with emphasis on its application in health care*. Syddansk Universitet.

Krygsman, S. (2004). *Activity and Travel Choice(s) in Multimodal Public Transport Systems*.

Krygsman, S., Dijst, M., & Arentze, T. (2004). Multimodal public transport: an analysis of travel time elements and the interconnectivity ratio. *Transport Policy*, *11*(3), 265–275. https://doi.org/10.1016/J.TRANPOL.2003.12.001

Krykewycz, G., Puchalsky, C., Rocks, J., Bonnette, B., & Jaskiewicz, F. (2010). Defining a primary market and estimating demand for major bicycle-sharing program in Philadelphia, Pennsylvania. *Transportation Research Record*, *2143*, 117–124. https://doi.org/10.3141/2143-15

Kvk. (2022). *Kamer van Koophandel*. Kvk.Nl. https://www.kvk.nl/producten-bestellen/handelsregister-dataservice/

la Paix, L., Cherchi, E., & Geurs, K. (2020). Role of perception of bicycle infrastructure on the choice of the bicycle as a train feeder mode. *International Journal of Sustainable Transportation*, *15*(6), 486–499. https://doi.org/10.1080/15568318.2020.1765223

Lindsay, J. K. (1997). *Applying generalized linear models* (G. Casella, S. Fienberg, & I. Olkin, Eds.; 2nd ed.). Springer-Verlag.

Lletí, R., Ortiz, M. C., Sarabia, L. A., & Sánchez, M. S. (2004). Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. *Analytica Chimica Acta*, *515*(1), 87–100. https://doi.org/10.1016/J.ACA.2003.12.020

Louviere, J. J., Hensher, D. A., Swait, J. D., & Adamowicz, W. (2000). *Stated Choice Methods Analysis and Applications*. Cambridge University Press.

Ma, X., Yuan, Y., van Oort, N., & Hoogendoorn, S. (2020). Bike-sharing systems' impact on modal shift: A case study in Delft, the Netherlands. *Journal of Cleaner Production*, *259*, 120846. https://doi.org/10.1016/J.JCLEPRO.2020.120846

Martens, K. (2007). Promoting bike-and-ride: The Dutch experience. *Transportation Research Part A: Policy and Practice*, *41*(4), 326–338. https://doi.org/10.1016/J.TRA.2006.09.010

Martínez, L. M., Viegas, J. M., & Silva, E. A. (2009). A traffic analysis zone definition: A new methodology and algorithm. *Transportation*, *36*(5), 581–599. https://doi.org/10.1007/S11116-009-9214-Z/TABLES/4

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models* (2nd ed.). Chapman and Hall.

McGuirk, P. M., & O'Neill, P. (2016). Using questionnaires in qualitative human geography. *University of Wollongong - Faculty of Social Sciences - Papers*, *2518*. https://ro.uow.edu.au/sspapershttps://ro.uow.edu.au/sspapers/2518

McNally, A. (2008). *The Four Step Model*. https://escholarship.org/uc/item/0r75311t

Mcneil, N., Dill, J., Macarthur, J., Broach, J., & Howland, S. (2017). *Braking Barriers to Bike Share: Insights from residents of Traditionally Underserved Neighborhoods*.

Médard de Chardon, C., Caruso, G., & Thomas, I. (2017). Bicycle sharing system 'success' determinants. *Transportation Research Part A: Policy and Practice*, *100*, 202–214. https://doi.org/10.1016/J.TRA.2017.04.020

Miller, A. (2002). Subset Selection in Regression. In *Subset Selection in Regression* (Second). Chapman and Hall/CRC. https://doi.org/10.1201/9781420035933

Miramontes, M., Pfertner, M., Rayaprolu, H. S., Schreiner, M., & Wulfhorst, G. (2017). Impacts of a multimodal mobility service on travel behavior and preferences: user insights from Munich's first Mobility Station. *Transportation 2017 44:6*, *44*(6), 1325–1342. https://doi.org/10.1007/S11116-017-9806-Y

Molander, S. (2018). Changing roles and new perspectives: towards market orientation in public transport. *Transportation*, *45*(6), 1811–1825. https://doi.org/10.1007/S11116-017-9781-3/TABLES/5

Morris, C., & Yang, J. J. (2021). Effectiveness of resampling methods in coping with imbalanced crash data: Crash type analysis and predictive modeling. *Accident Analysis & Prevention*, *159*, 106240. https://doi.org/10.1016/J.AAP.2021.106240

Myers, R. H., & Montgomery, D. C. (1997). A Tutorial on Generalized Linear Models. *Journal of Quality Technology*, *29*(3), 274–291. https://doi.org/10.1080/00224065.1997.11979769

Naing, N. N. (2003). Determination of Sample Size. *The Malaysian Journal of Medical Sciences : MJMS*, *10*(2), 84. /pmc/articles/PMC3561892/

Nello-Deakin, S., & Brömmelstroet, M. te. (2021). Scaling up cycling or replacing driving? Triggers and trajectories of bike–train uptake in the Randstad area. *Transportation*, *48*(6), 3239–3267. https://doi.org/10.1007/S11116-021-10165-9/FIGURES/7

NS. (2019). *Keteninformatie Dashboard*. NS. nsmobielkis.nl

NS. (2020). *NS Jaarverslag 2020*. https://www.nsjaarverslag.nl/FbContent.ashx/pub_1002/downloads/v220216221455/NS-Jaarverslag-2020.pdf

NS. (2022a). *NS Jaarverslag 2021*. https://www.nsjaarverslag.nl/FbContent.ashx/pub_1002/downloads/v220225095034/NS-Jaarverslag-2021.pdf

NS. (2022b, August 23). *Halfjaarcijfers NS: Matig herstel in eerste helft 2022*. Ns.Nl. https://nieuws.ns.nl/halfjaarcijfers-ns-matig-herstel-in-eerste-helft-2022

NS Stations. (2021). *KetenBelevingsMonitor OV-fiets*.

NS Stations. (2022a). *NS, Stationsdata.nl*. Stationsdata.Nl. https://stationsdata.ns.nl/portal/home/index.html

NS Stations. (2022b). *Rapportage niet-gebruik OV-fiets*.

Ogra, A., & Ndebele, R. (2014). The Role of 6Ds: Density, Diversity, Design, Destination, Distance, and Demand Management in Transit Oriented Development (TOD). *Neo-International Coference on Habitable Enviornments*, 539–546.

Ortúzar, J., & Willumsen, L. (2011). *Modelling Transport* (4th ed.). Wiley & Sons.

Orvin, M. M., & Fatmi, M. R. (2020). Modeling Destination Choice Behavior of the Dockless Bike Sharing Service Users: *Transportation Research Board*, *2674*(11), 875–887. https://doi.org/10.1177/0361198120950315

OTM. (2021). *Open Transport Map*. Opentransportmap.Info. http://opentransportmap.info/

Peppel, O. van de. (2020, October 8). *Toelichting Wijk- en Buurtkaart 2018, 2019 en 2020*. CBS. https://www.cbs.nl/nl-nl/longread/aanvullende-statistische-diensten/2020/toelichting-wijk-en-buurtkaart-2018-2019-en-2020

Ploeger, J., & Oldenziel, R. (2020). The sociotechnical roots of smart mobility: Bike sharing since 1965: *The Journal of Transport History*, *41*(2), 134–159. https://doi.org/10.1177/0022526620908264

ProRail. (2022). *PDOK, Dataset: Spoorwegen*. Pdok.Nl. https://www.pdok.nl/introductie/-/article/spoorwegen

QGIS Development Team. (2009). *QGIS Geographic Information System*. http://qgis.osgeo.org

Qualtrics. (2021, May 18). *Sample Size*. https://www.qualtrics.com/blog/calculating-sample-size/

Ramos, É. M. S., Bergstad, C. J., & Nässén, J. (2020). Understanding daily car use: Driving habits, motives, attitudes, and norms across trip purposes. *Transportation Research Part F: Traffic Psychology and Behaviour*, *68*, 306–315. https://doi.org/10.1016/J.TRF.2019.11.013

Raux, C., Zoubir, A., & Geyik, M. (2017). Who are bike sharing schemes members and do they travel differently? The case of Lyon's "Velo'v" scheme. *Transportation Research Part A: Policy and Practice*, *106*, 350–363. https://doi.org/10.1016/J.TRA.2017.10.010

Recker, W. W., & Kostyniuk, L. P. (1978). Factors influencing destination choice for the urban grocery shopping trip. *Transportation*, *7*, 19–33.

Ricci, M. (2015). Bike sharing: A review of evidence on impacts and processes of implementation and operation. *Research in Transportation Business & Management*, *15*, 28–38. https://doi.org/10.1016/j.rtbm.2015.03.003

Rietveld, P. (2000). The accessibility of railway stations: the role of the bicycle in The Netherlands. *Transportation Research Part D: Transport and Environment*, *5*(1), 71–75. https://doi.org/10.1016/S1361-9209(99)00019-X

Rijsman, L., van Oort, N., Ton, D., Hoogendoorn, S., Molin, E., & Teijl, T. (2019). Walking and bicycle catchment areas of tram stops: Factors and insights. *MT-ITS 2019 - 6th International Conference on Models and Technologies for Intelligent Transportation Systems*. https://doi.org/10.1109/MTITS.2019.8883361

Salah Mahmoud, M., El-Assi, W., Nurul Habib, K., Khandker Nurul Habib, P., & Khandkernurulhabib, Pe. (2015, January). Effects of Built Environment and Weather on Bike Sharing Demand: Station Level Analysis of Commercial Bike Sharing in Toronto. *94th Annual Meeting of Trans. Res. Board*. https://www.researchgate.net/publication/266968162

Satpathy, S. (2020, September 27). Overcoming Class Imbalance using SMOTE Techniques. *Data Science Blogathon*. https://www.analyticsvidhya.com/blog/2020/10/overcoming-class-imbalance-using-smote-techniques/

Schakenbos, R., Paix, L. la, Nijenstein, S., & Geurs, K. T. (2016). Valuation of a transfer in a multimodal public transport trip. *Transport Policy*, *46*, 72–81. https://doi.org/10.1016/J.TRANPOL.2015.11.008

Schimohr, K., & Scheiner, J. (2021). Spatial and temporal analysis of bike-sharing use in Cologne taking into account a public transit disruption. *Journal of Transport Geography*, *92*, 103017. https://doi.org/10.1016/J.JTRANGEO.2021.103017

Schubert, F. (2019). *Cluster Analysis Technical Lecture*.

Schuberth, F. (2019). *Regression Analysis Technical Lecture*.

Sener, I. N., Pendyala, R. M., & Bhat, C. R. (2010). Accommodating spatial correlation across choice alternatives in discrete choice models: An application to modeling residential location choice behavior. *Journal of Transport Geography*, *18*(5), 624–633. https://doi.org/10.1016/j.jtrangeo.2010.03.013

Shaheen, S., Guzman, S., & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia. *Transportation Research Record*, *2143*, 159–167. https://doi.org/10.3141/2143-20

Shmueli, B. (2019, July 3). *Multi-class metrics made simple, part II: the F1-score*. Towards Data Science. https://towardsdatascience.com/multi-class-metrics-made-simple-part-ii-the-f1-score-ebe8b2c2ca1

Simma, A. ;, Schlich, R. ;, & Axhausen, K. W. (2001). Destination choice modelling for different leisure activities. In *Arbeitsberichte Verkehrs- und Raumplanning* (No. 99; Vol. 99). https://doi.org/10.3929/ethz-a-004347884

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*, 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Steyerberg, E. W., C Eijkemans, M. J., Harrell Jr, F. E., & Habbema, J. F. (2001). Prognostic Modeling with Logistic Regression Analysis: In Search of a Sensible Strategy in Small Data Sets. *Medical Decision Making*, *21*, 45–56.

Sung, H., Lee, S., Cheon, S., & Yoon, J. (2022). Pedestrian Safety in Compact and Mixed-Use Urban Environments: Evaluation of 5D Measures on Pedestrian Crashes. *Sustainability 2022, Vol. 14, Page 646*, *14*(2), 646. https://doi.org/10.3390/SU14020646

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411–423. https://doi.org/10.1111/1467-9868.00293

Tirachini, A., Hensher, D. A., & Rose, J. M. (2013). Crowding in public transport systems: Effects on users, operation and implications for the estimation of demand. *Transportation Research Part A: Policy and Practice*, *53*, 36–52. https://doi.org/10.1016/J.TRA.2013.06.005

Todd, J., O'Brien, O., & Cheshire, J. (2021). A global comparison of bicycle sharing systems. *Journal of Transport Geography*, *94*, 103119. https://doi.org/10.1016/J.JTRANGEO.2021.103119

Ton, D., Shelat, S., Nijënstein, S., Rijsman, L., van Oort, N., & Hoogendoorn, S. (2020). Understanding the Role of Cycling to Urban Transit Stations through a Simultaneous Access Mode and Station Choice Model. *Transportation Research Record*, *2674*(8), 823–835. https://doi.org/10.1177/0361198120925076

Train, K. (2009). *Discrete Choice Methods with Simulation* (Vol. 2). Cambridge University Press.

Uddin, M. F. (2019, November 20). Addressing Accuracy Paradox Using Enhanched Weighted Performance Metric in Machine Learning. *Sixth HCT Information Technology Trends (ITT).*

van den Hoek, J. W. (2008). The MXI (Mixed-use Index) as Tool for Urban Planning and Analysis. *Corporations and Cities: Envsioning Corporate Real Estate in the Urban Future*. www.corporationsandcities.org

van Soest, D., Tight, M. R., & Rogers, C. D. F. (2019). Exploring the distances people walk to access public transport. *Transport Reviews*, *40*(2), 160–182. https://doi.org/10.1080/01441647.2019.1575491

van Waes, A., Farla, J., Frenken, K., de Jong, J. P. J., & Raven, R. (2018). Business model innovation and socio-technical transitions. A new prospective framework with an application to bike sharing. *Journal of Cleaner Production*, *195*, 1300–1312. https://doi.org/10.1016/J.JCLEPRO.2018.05.223

Vij, A., Carrel, A., & Walker, J. L. (2013). Incorporating the influence of latent modal preferences on travel mode choice behavior. *Transportation Research Part A*, *54*, 164–178. https://doi.org/10.1016/j.tra.2013.07.008

Waller, L. A., & Jacquez, G. M. (1995). Disease Models Implicit in Statistical Tests of Disease Clustering. *Epidemiology*, *6*(6), 584–590. https://www.jstor.org/stable/pdf/3703132.pdf?casa_token=xBtCHPHmlaMAAAAA:J1-hQm0NCtJOWGKkVgUP9M9_OdUdpxyq3zngDUADIt6nI4NeGU1d27-niddBAGxrZin9pvaIUVSacitCgnyCDZ_VfE3Qw25Ldi3i7pz_YqqK4Ifk6RY

Washington, S. P., Karlaftis, M. G., & Mannering, F. L. (2011). *Statistical and Econometric Methods for Transportation Data Analysis* (Second). Taylor & Francis Group.

Wilkesmann, F. L. (2022). *Short-Term Forecast of Demand for Train Station-based Round-Trip Bikesharing*. http://repository.tudelft.nl/.

Yacouby, R., & Axman, D. (2020). Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models. *First Workshop on Evaluation and Comparison of NLP Systems*, 79–91.

Yang, Y., Jiang, L., & Zhang, Z. (2021). Tourists on shared bikes: Can bike-sharing boost attraction demand? *Tourism Management*, *86*, 104328. https://doi.org/10.1016/J.TOURMAN.2021.104328

Zhang, X., Shen, Y., & Zhao, J. (2021). The mobility pattern of dockless bike sharing: A four-month study in Singapore. *Transportation Research Part D: Transport and Environment*, *98*, 102961. https://doi.org/10.1016/J.TRD.2021.102961

Zhang, Y., Thomas, T., Brussel, M., & van Maarseveen, M. (2017). Exploring the impact of built environment factors on the use of public bikes at bike stations: Case study in Zhongshan, China. *Journal of Transport Geography*, *58*, 59–70. https://doi.org/10.1016/J.JTRANGEO.2016.11.014

Zuo, T., Wei, H., Chen, N., & Zhang, C. (2020). First-and-last mile solution via bicycling to improving transit accessibility and advancing transportation equity. *Cities*, *99*, 102614. https://doi.org/10.1016/J.CITIES.2020.102614

8 Appendices

8.1 Appendix A – OV-fiets system

8.1.1 Bikesharing Systems

Much research about BSS concerns one-way bikesharing, instead of SBRT systems like OV-fiets. To investigate if that literature is useful for the case of OV-fiets, it should be made clear what the differences and similarities are compared to other BSS. BSS can be distinguished into three generations, as described by Shaheen et al. (2010). The first generation BSS was released in Amsterdam in 1965. The “white bikes” were free to use and placed without locks throughout the inner city. These bikes were often stolen or vandalised and found at dangerous locations, leading to the project's termination. After that, the second generation of bike-sharing started in Copenhagen in 1995. These bicycles had a coin-deposit system, which increased the probability that bicycles returned unharmed. For this purpose, the system had designated docking stations where bicycles are locked. The plan was more reliable, although running costs were higher. Theft of the bicycles was still a problem due to the user’s anonymity.

The third generation BSS is characterised by identifying users via personal cards or other subscription methods. Third-generation BSS programs incorporate information technologies to track bicycles and user information. The main components of third-generation BSS is that they have kiosks or other user interface technology for check-in and check-out, and advanced technology like smartcards is incorporated into the system design. Furthermore, the bicycles are distinguishable (by colour or design), and docking stations are visible. Currently, most active BSS are third-generation systems.

As companies and systems developed, the fourth generation BSS emerged. This generation is characterised by the BSS having smart bicycles, accessed by a mobile app, connected with an integrated traffic management system and real-time information on availability.

The newest generation of BSS is the fifth generation, characterised by free-floating (or dockless), bicycles and possibilities for big data management (Chen et al., 2018). In this generation, technology to lock the bicycle is integrated into the design, and customers use mobile applications to unlock the bicycle. Furthermore, these bicycles can be dropped and picked up everywhere. Shaheen et al. (2010) have made an overview of the different BSS generations and their characteristics. This overview has been extended with the information of Chen et al. (2018), see Table 21.

Table 21: Overview of generations of bikesharing systems

Gen.	Components	Characteristics
1	• Bicycles	• Distinct bicycles (by colour) • Bicycles located haphazardly throughout an area • Bicycles unlocked • No charge for use
2	• Bicycles • Docking stations	• Distinct bicycles (by colour and design) • Bicycles located at specific docking stations • Bicycles with lock
3	• Bicycles • Docking stations • Kiosks or user interface technology	• Distinct bicycles (by colour, design or advertisement) • Bicycles are located at specific docking stations • Bicycles have locks • Smart technology is used for bicycle check-in and check-out • Theft deterrents (i.e. required to provide ID)
4	• Bicycles • Docking stations • Kiosks-user interface • Bicycle distribution system	• Distinct bicycles • Programs may include electric bicycles • Specific docking stations that are more efficient • Improved locking mechanism • Touch screen kiosks-user interface • Bicycle redistribution system • Linked to public transit smartcard
5	• Bicycles with lock-technology • Bicycle distribution system	• Distinct bicycles with GPS • The program may include electric bicycles • Mobile apps to unlock • Improved locking mechanism • Bicycle redistribution system

8.1.2 OV fiets compared to BSS

The OV-fiets has the characteristics of a third-generation BSS and a SBRT system business model. It has distinct bicycles, simple mechanical locks, docking stations and smart technology for check-in and check-out. As stated before, the OV-fiets differ from (international) common BSS based on the business model and scale. OV-fiets is the only system of a round-trip station-based bikesharing system with such a large scale, aside from the smaller Bluebike in Belgium. DeMaio et al. (2021) showed that the system is an exception as they inventoried almost 1900 one-way systems around the world. Therefore, a deeper dive into the characteristics of the OV-fiets system is needed to ascertain the differences with other systems.

The characteristic of round-trips implies that the OV-fiets is targeted at the trip's activity-end of PT (train) trips and cannot be used as home to (train-) station trips or between two main modes of a multimodal trip. This is emphasised with docking stations exclusively located at train or big metro stations. Therefore, most of the OV-fiets users might be train passengers, as opposed to other systems.

A second difference is the payment scheme. The OV-fiets has a fixed fee per 24 hours and is competitive with single bus fares (Martens, 2007). Common BSS use a 'pay as you ride'-payment, in which costs are calculated per minute, mainly with a free starting period (Dijkgraaf, 2021; Donkey Republic, 2021). Literature states that different payment schemes (which can be seen as economic incentives) can attract different users, creating or dissolving equity issues (Bachand-Marleau et al., 2012; Duran-Rodas et al., 2021; Efthymiou et al., 2013; Mcneil et al., 2017). Therefore, the OV-fiets might attract other potential user groups, such as lower-income and immigrant groups (Grasso et al., 2020).

Third, the payment system of OV-fiets is different from a common BSS. The system is designed that a single transaction is below 3 seconds, which is one of the success factors of OV-fiets (Ploeger & Oldenziel, 2020). Incorporating the OV-chipcard means that the OV-fiets system is better integrated with PT. Nowadays, many BSS require an app or other online registration to rent the bicycle (Donkey Republic, 2021). The convenience of the OV-fiets 3-second rent system makes the threshold to use the OV-fiets for regular public transport passengers lower (Schakenbos et al., 2016).

Additionally, the user demographics can differ between the OV-fiets and other BSS, because a personal OV-chipcard is required for the OV-fiets. For instance, foreign tourists cannot use the OV-fiets, as they cannot obtain the necessary OV-chipcard, which requires a Dutch bank account. Conversely, the OV-fiets is catered to business commuters. It works with different OV-chipcard types, such as a business card for employees of companies. The employer can provide this card if the employer reimburses (part of) commuting costs. Therefore, people who use this card travel for free indirectly, which increases the use. This means the OV-fiets is relatively (in)attractive for different users (Garín-Muñoz, 2009; Yang et al., 2021).

To summarise, the main differences between the OV-fiets bike-sharing program and other bike-share programs are displayed in Table 22. These demonstrate the place of the OV-fiets in the bike-sharing spectrum:

Table 22: Main differences between OV-fiets and other BSS

	OV-fiets	Other BSS
Location	Docks situated at train or metro station	Station-based: Docks at specific locations Free-floating: Everywhere
Rental system	Round-trip required	Station-based: Between docks Free-floating: Everywhere
Function	On the activity side of a trip	Both access, egress or single trip
Payment scheme	Rental fee per 24h	Rental fee per minute (sometimes with a free starting period)
Payment system	OV-chipcard with online subscription	Debit card or mobile app (most with online subscription)

8.1.3 OV-fiets as egress mode

The OV-fiets is targeted to a trip's activity side due to its characteristics. Since the rental locations of OV-fiets are situated only at train stations, the OV-fiets user is mostly a train passenger. The OV-fiets cannot be chosen as a mode to the train station or bus stop at the home-end of a trip because the OV-fiets is not available at someone's doorstep. This characteristic means that OV-fiets is only useful as egress mode, which connects the main mode with the destination.

Access and egress modes determine the availability of PT, whereby their reach determine the catchment area of the main mode (Ortúzar & Willumsen, 2011). When time and distance increase for an access/egress mode, the use of PT decreases (Krygsman et al., 2004). Another common rule is that the time spent with an egress mode is relatively constant, but increased speed leads to increased range (Zuo et al., 2020).

Egress modes can be differentiated via their characteristics, mainly by range and availability (Krygsman et al., 2004). For example, walking is characterised with a short-range but is always available (continuous). At the same time, bus/tram/metro (BTM) has a large range but is available at specific times (due to schedules). Krygsman (2004) differentiated seven main access and egress modes: Walking, cycling, car (driver), car (passenger), bus, tram/metro and train.

On the activity side of the trip, travellers do not have all the private modes available. This observation explains the dominant position of walking on the activity side of the multimodal trip (Keijer & Rietveld, 1999). The OV-fiets fills the gap that the absence of privately owned bikes leave. With their higher speed, the OV-fiets can reach more destinations than walking.

Another positive characteristic of (shared) bikes/OV-fiets is availability. Due to the characteristics of PT, people arrive at distinct times. The connecting mode should be available at those particular times. Walking and cycling are continuous modes, as they can be taken at any time, while BTM is only available at scheduled times (Rietveld, 2000). This availability issue can be translated to reachability too. Walking and cycling can reach every doorstep, while BTM (or car) can only reach stops or parking spots (Jäppinen et al., 2013).

Finally, the OV-fiets has a different payment method compared to other paid egress modes. Specifically compared to BTM, the difference is that an online subscription is required. People need to sign up (for free) in advance, while BTM have ticket sales on board, which means that people without an OV-chipcard can still use the mode. The online subscription to use the OV-fiets system can be perceived as an effort, creating a barrier for certain people (Schakenbos et al., 2016). Furthermore, payment of OV-fiets is a single fee per 24 hours, while with BTM the payment is per trip. This feature of OV-fiets makes bundling activities in one rental period attractive, like combining grocery shopping with a commute (Krygsman, 2004).

To summarise, the main characteristics of the OV-fiets compared to other egress modes are shown in the following table:

Table 23: Main differences between OV-fiets and other egress modes

	OV-fiets	Bus, Tram, Metro	Walking	Private vehicles
Availability	Continuous	On schedule	Continuous	Continuous, but limited
Trip costs	Fee per 24h	Fee per trip	No trip costs	No trip costs
Payment system	OV-chipcard and online subscription	OV-chipcard	Not applicable	Not applicable
Destination reach	Medium range, everywhere	Medium to large range, at stops	Short, everywhere	Medium to large, everywhere

8.2 Appendix B – Descriptive statistics datasets

8.2.1 Locations

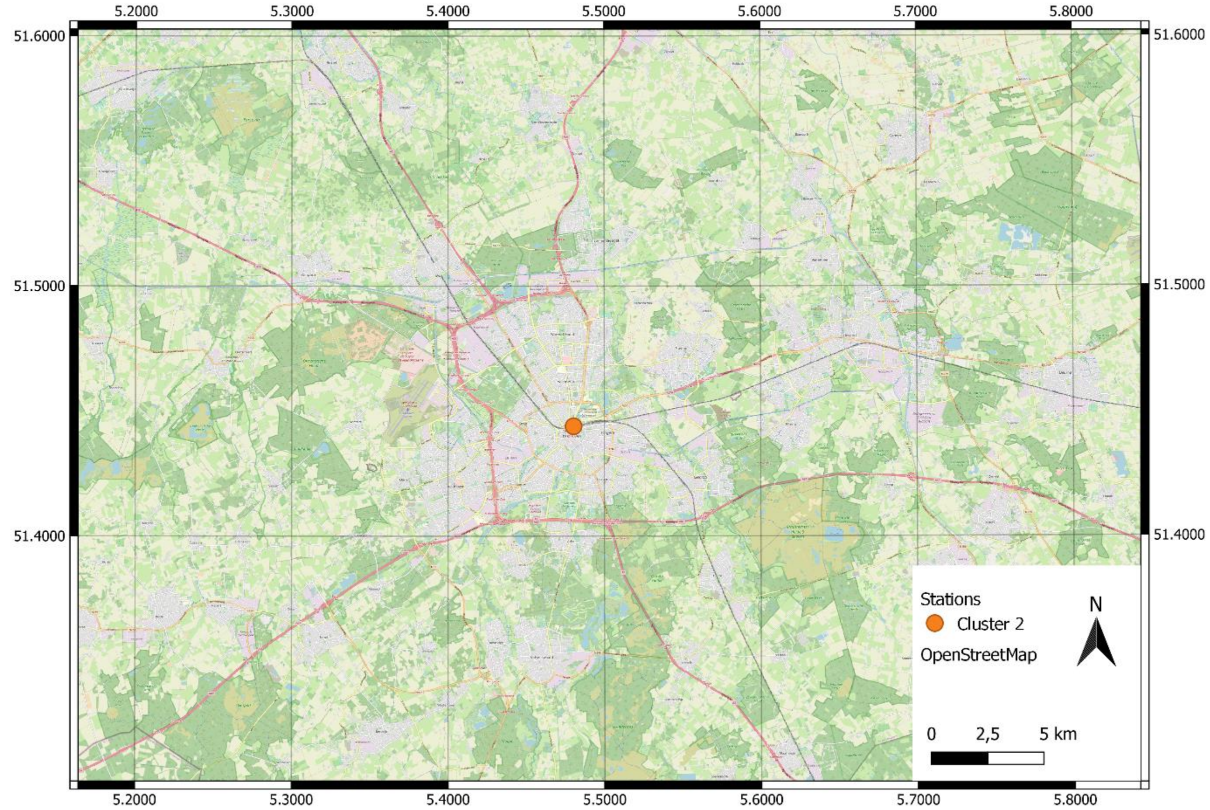


Figure 39: Eindhoven, area

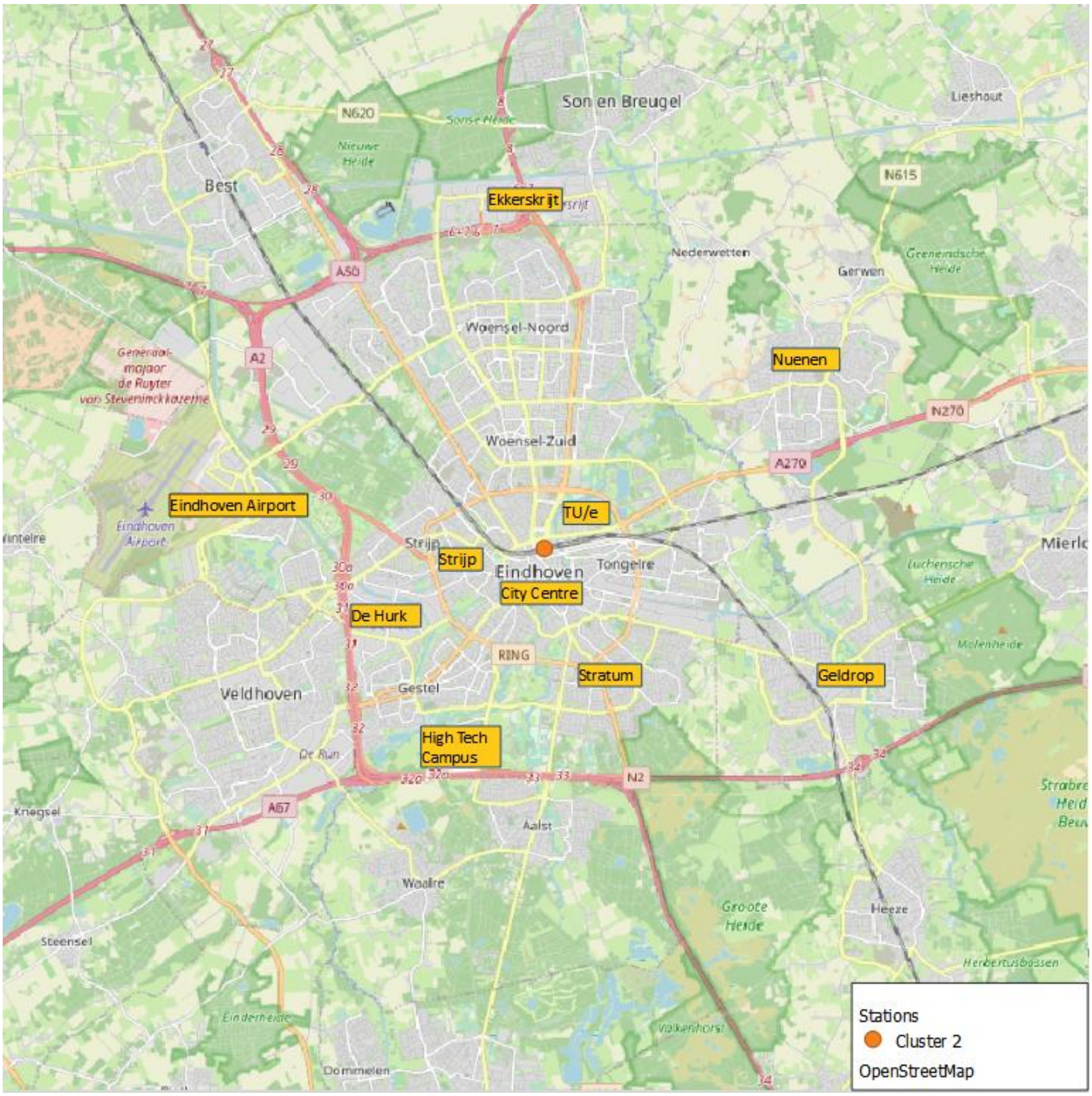


Figure 40: Eindhoven, highlighted areas

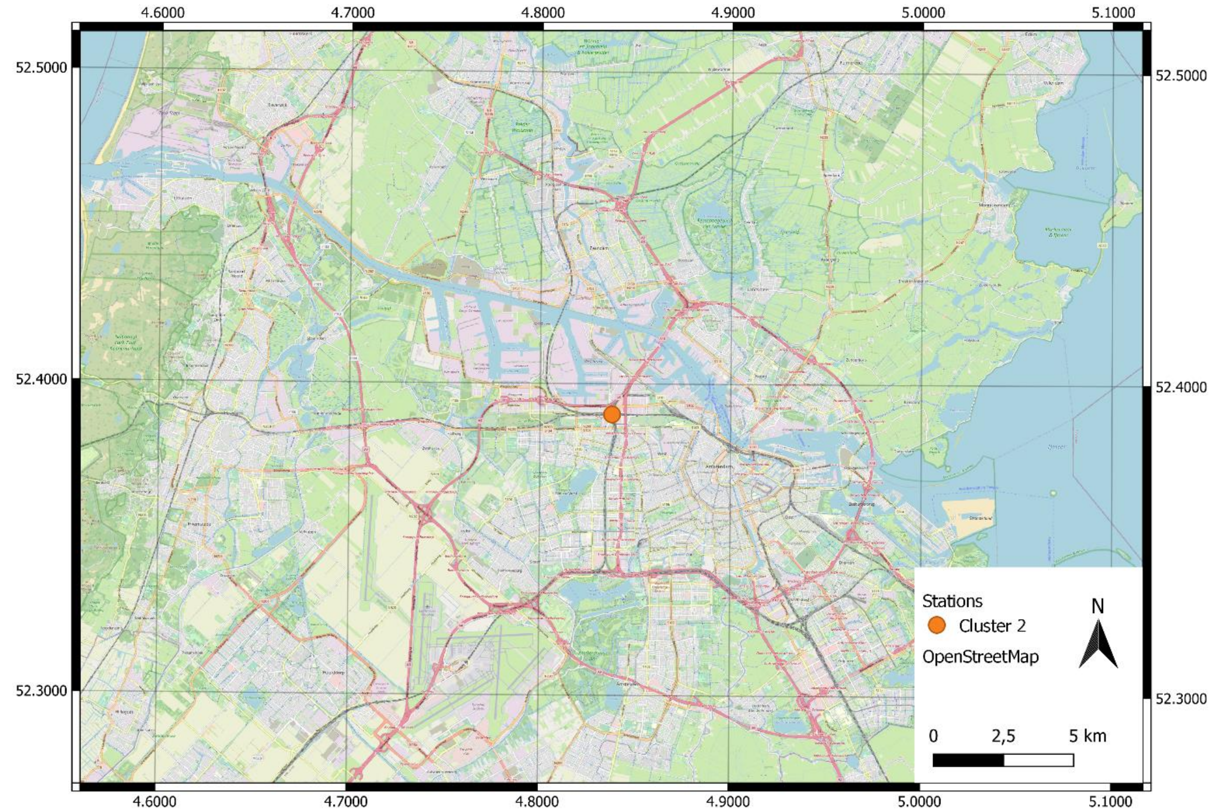


Figure 41: Amsterdam, area

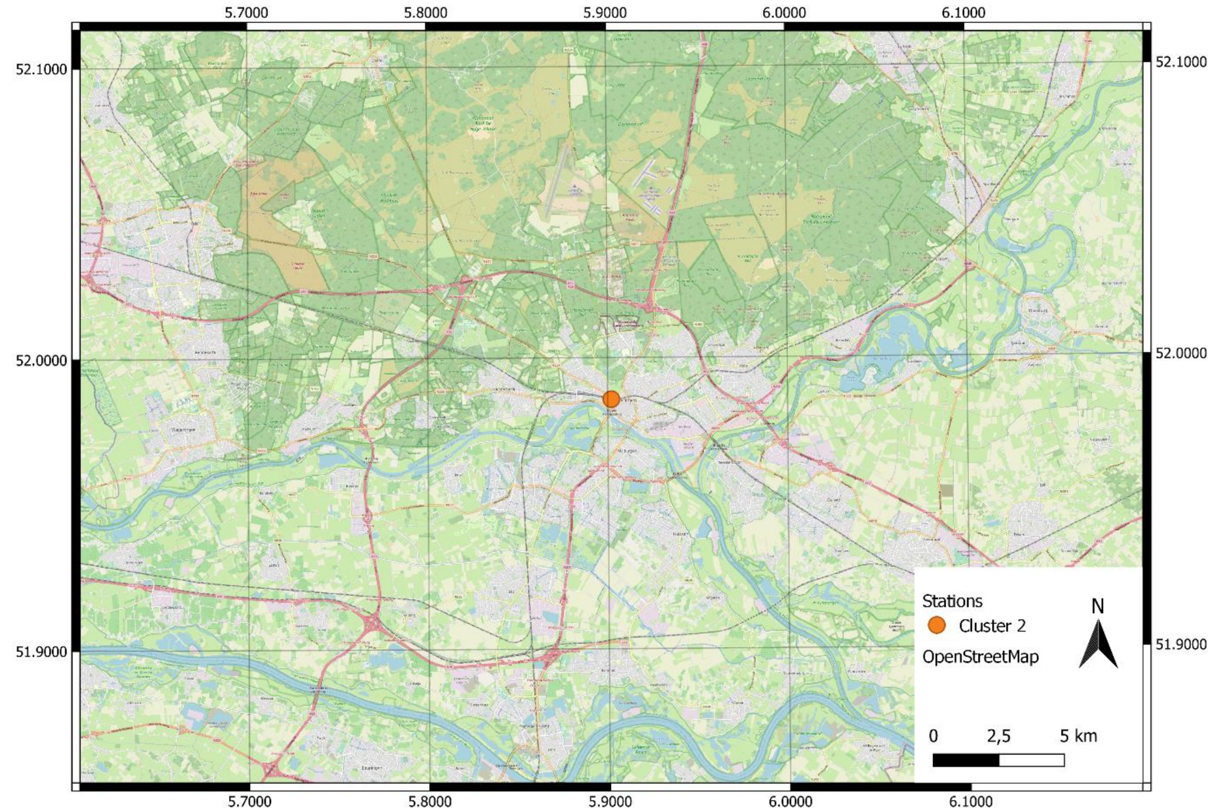


Figure 42: Arnhem, area

8.2.2 Spotinfo grid

Table 24: Descriptive statistics of Mozaiek dataset of Spotinfo/Argaleo

Name	Variable name	class	Min/True	1st Qu/False	Median	Mean	3rd Qu	Max	NA's
Identification number	mid	character							
Postal-4 code	admi_pc4	integer	1011	1505	5651	4623	6816	6994	131
Neighbourhood code	admi_buurtc	character							
Borough code	admi_wijkco	character							
Name of kern	admi_kern	character							
City limits	admi_bebkom	Logical	49825	43355					
Name of city	admi_wpltsc	character							
No. muncipality	admi_gemcd	character							
Name of muncipality	admi_gemnm	character							
No. province	admi_provcd	character							
No. water agency	admi_watsch	character							
No. safety region	admi_vr_nr	character							
No. police	admi_pol_nr	character							
Name of drinking water supplier	admi_watbdr	character							
Name of electricity agency	admi_netbeh	character							
Telephone code	admi_netnmr	character							
Code of BRT	admi_topbld	character							
soil unit code	fysk_bodemv	character							
soil name	fysk_geomrf	character							
Percentage of water	fysk_water	integer	0	0	0	7,336	2	100	
Ground level	fysk_maaivh	integer	-5,5	0,45	13,83	16,19	21,46	105,66	
10th percentile of subsidence	fysk_daling	integer	-863	-190	-110	-95,13	16	149	
Percentage of swampy	fysk_drassg	integer	0	0	0	1,242	0	100	
Flooding probability	fysk_ovrstr	integer	0	0	0	1,064	2	6	
Percentage arable land	fysk_akkerl	integer	0	0	0	7,513	0	100	
Percentage meadows	fysk_grasld	integer	0	0	9	26,54	47	100	
Percentage wasteland	fysk_braakl	integer	0	0	0	0,0113	0	100	
Percentage sand/beaches	fysk_zand	integer	0	0	0	0,1777	0	100	
Percentage dunes	fysk_duin	integer	0	0	0	0	0	0	
Percentage heath	fysk_heide	integer	0	0	0	4,148	0	100	
Percentage cemetery	fysk_bgrfpl	integer	0	0	0	0,1867	0	100	
Percentage forest	fysk_bomen	integer	0	0	0	16,58	18	100	
Percentage fruit farms	fysk_fruitk	integer	0	0	0	0,2495	0	100	
Percentage other land-use	fysk_overig	integer	0	0	2	26,22	56	100	
RIVM nitrogen deposition	fysk_stkstf	integer	1045	1708	1985	2042	2334	4142	
Percentage of business park	func_bedrtr	integer	0	0	0	6,762	0	100	
Percentage of recreation	func_recrea	integer	0	0	0	0,6666	0	200	
Percentage of gold course	func_golftr	integer	0	0	0	0,9827	0	100	
Percentage of extraction	func_winnng	integer	0	0	0	0,1098	0	200	
Marina in cell	func_jachth	Logical	92842	338					
Greenhouse in cell	func_kassen	integer	0	0	0	0,4601	0	100	
Percentage of military property	func_militr	integer	0	0	0	1,155	0	100	
Percentage of national park	func_natprk	integer	0	0	0	3,423	0	100	
Percentage of Natura-2000	func_natura	integer	0	0	0	17,87	0	100	
Percentage of sports amenities	func_sportt	integer	0	0	0	1,249	0	100	
Percentage of aerial activity	func_luchtv	integer	0	0	0	2,081	0	100	
Percentage of solar park	func_zonnep	integer	0	0	0	0,04018	0	100	
Percentage of storage hazard	func_risico	integer	0	0	0	1,953	0	200	
Percentage of event area	func_evenem	integer	0	0	0	0,5127	0	157	
No. monuments	func_monumt	integer	0	0	0	0,09605	0	50	
Railway in cell	infr_spoor	Logical	90363	2817					
Highway in cell	infr_snelwg	Logical	89521	3659					
Provincial (N) road in cell	infr_provwg	Logical	90667	2513					
Nautical fairway in cell	infr_vaarwg	Logical	90095	3085					
High voltage line in cell	infr_hoogsp	Logical	91486	1694					
Subsurface pipeline in cell	infr_buisld	Logical	88697	4483					
No. telecom antennas	infr_telecm	integer	0	0	0	0,05458	0	14	
Major dike in cell	infr_dijkrg	Logical	91828	1352					
No. adresses	bouw_adrssn	integer	0	0	0	10,57	4	670	
No. buildings	bouw_panden	integer	0	0	0	7,765	4	155	

Population	bouw_popula	integer	0	0	0	37,06	24	20934	4839
No. Vulnerable people	bouw_kwetsb	integer	0	0	0	4,603	0	4849	27
No. Child care	bouw_kinder	Logical	92174	1006					
No. Hospitalities	bouw_horeca	Logical	90922	2258					
No. Agricultural adresses	bouw_agrari	Logical	92631	549					
No. Meeting buildings	bouw_bijeen	integer	0	0	0	0,1126	0	53	
No. Religious gatherings	bouw_religi	Logical	92587	593					
No. Art/theater/concert buildings	bouw_kunst	Logical	93035	145					
No. Shops	bouw_winkel	integer	0	0	0	0,2146	0	123	
No. Food stores	bouw_voedng	Logical	91366	1814					
No. Accommodation buildings	bouw_logies	integer	0	0	0	0,03273	0	167	
No. Industry buildings	bouw_indust	integer	0	0	0	0,318	0	65	
No. Health buildings	bouw_gezond	integer	0	0	0	0,04267	0	102	
No. Offices	bouw_kantoo	integer	0	0	0	0,2221	0	30	
No. Education buildings	bouw_onderw	integer	0	0	0	0,01819	0	14	
No. Primary schools	bouw_bssond	Logical	92789	391					
No. Ambulance stations	bouw_ambula	Logical	93177	3					
No. Fire brigades	bouw_brandw	Logical	93130	50					
No. Police offices	bouw_politi	Logical	93163	17					
No. Gas stations	bouw_tankst	Logical	93016	164					
No. Car dealers	bouw_autoga	Logical	92364	816					
No. Hospitals	bouw_zieknh	Logical	93173	7					
No. Kvk businesses	bouw_econac	integer	0	0	0	1,049	0	1190	
Average construction year	bouw_bouwjr	integer	0	0	0	890,6	975	2022	
Average WOZ property worth	bouw_wozwrd	integer	0	0	0	143199	212564	33320000	
No. Buildings >25m	bouw_hoogbw	integer	0	0	0	0,01098	0	20	
No. Demolition buildings	bouw_sloopv	integer	0	0	0	0,006461	0	25	
No. Construction permit	bouw_bouwvg	integer	0	0	0	0,06018	0	94	
No. Large wind turbines	bouw_windtb	integer	0	0	0	0,000408	0	2	
No. Detached houses	bouw_vrijst	integer	0	0	0	0,3664	0	24	
No. Semidetached houses	bouw_2_1kap	integer	0	0	0	0,5153	0	49	
No. Terraced houses	bouw_rijwon	integer	0	0	0	2,293	0	84	
No. Multi-household houses	bouw_appart	integer	0	0	0	4,305	0	631	
Average energy label	bouw_englbl	integer	0	0	0	1,353	3	7	
No. Addresses prone to fraud	bouw_fraude	integer	0	0	0	0,03561	0	80	
Average No. Floors	bouw_etages	integer	0	0	0	0,8683	1	36	
Floor space index	bouw_fsi	integer	0	0	0	19,99	0	32767	
Ground space index	bouw_gsi	integer	0	0	0	0	0	0	
Total floor area	bouw_totbvo	integer	0	0	0	727,8	55	999999	2506
Mixed use index	bouw_mxi	integer	0	0	0	26,48	80	100	
Average liveability score	demo_leefbr	integer	0	0	0	2,55	6	9	5125
Percentage of 0 to 15 years	demo_0_15	integer	0	8	4	2,09	7	36	
Percentage of 15 to 25 years	demo_15_25	integer	0	9	1	0,6	4	80	
Percentage of 25 to 45 years	demo_25_45	integer	0	5	0	0,36	6	76	
Percentage of 45 to 65 years	demo_45_65	integer	0	1	0	4,83	4	59	
Percentage of 65 years and older	demo_65_oud	integer	0	9	6	5,33	2	87	
Percentage of western immigrants	demo_migrwe	integer	0	5	8	9,983	2	80	
Percentage of non-western immigrants	demo_migrnw	integer	0	1	4	9,041	0	95	
Percentage of married	demo_gehuwd	integer	0	2	7	1,12	4	58	
Percentage of divorced	demo_geschd	integer	0	4	6	5,793	8	23	
Percentage of one-person household	demo_1phhdn	integer	0	2	7	2,32	3	98	
Percentage of households with children	demo_mphhdn	integer	0	0	4	0,34	1	78	
Average household-size	demo_gemhgr	integer	0	2	2	2,178	3	3	
No. Residents with AOW	demo_uitaow	integer	0	0	0	2,34	0	1290	
Average income	demo_inkomn	integer	0	0	0	1,652	4	9	2918
No. Residents	demo_inwkrn	integer	0	0	0	51939	158065	865045	
Longitude	lon	integer	4,694	4,902	5,486	5,416	5,845	6,046	
Latitude	lat	integer	51,35	51,48	51,98	51,93	52,35	52,48	

Sources of government for data in mozaiek:

- BAG (Basisregistratie Adressen en Gebouwen)
- BGT (Basisregistratie Grootschalige Topografie)
- NHR (Handelsregister)
- AHN (Actueel Hoogtebestand Nederland)
- LRK (Landelijk Register Kinderopvang)
- BRT (Basisregistratie Topografie)
- BRO (Basisregistratie Ondergrond)
- CBS (Demografische wijk- en buurtdata)
- RRGS (Register Risicosituaties Gevaarlijke Stoffen, uit Risicokaart (PRK)
- WOZ (Waardering Onroerende Zaken)
- NDW (Nationale Databank Wegverkeer)
- RCE (Register Cultureel Erfgoed)
- AR (Antenneregister)
- RHD (RWS Hydrografische data)
- EL (Energielabels)
- LBM (Leefbaarometer)
- EP-O (Energielabels online)
- RIVM (RIVM milieu gerelateerde data)

8.2.3 RUDIFUN

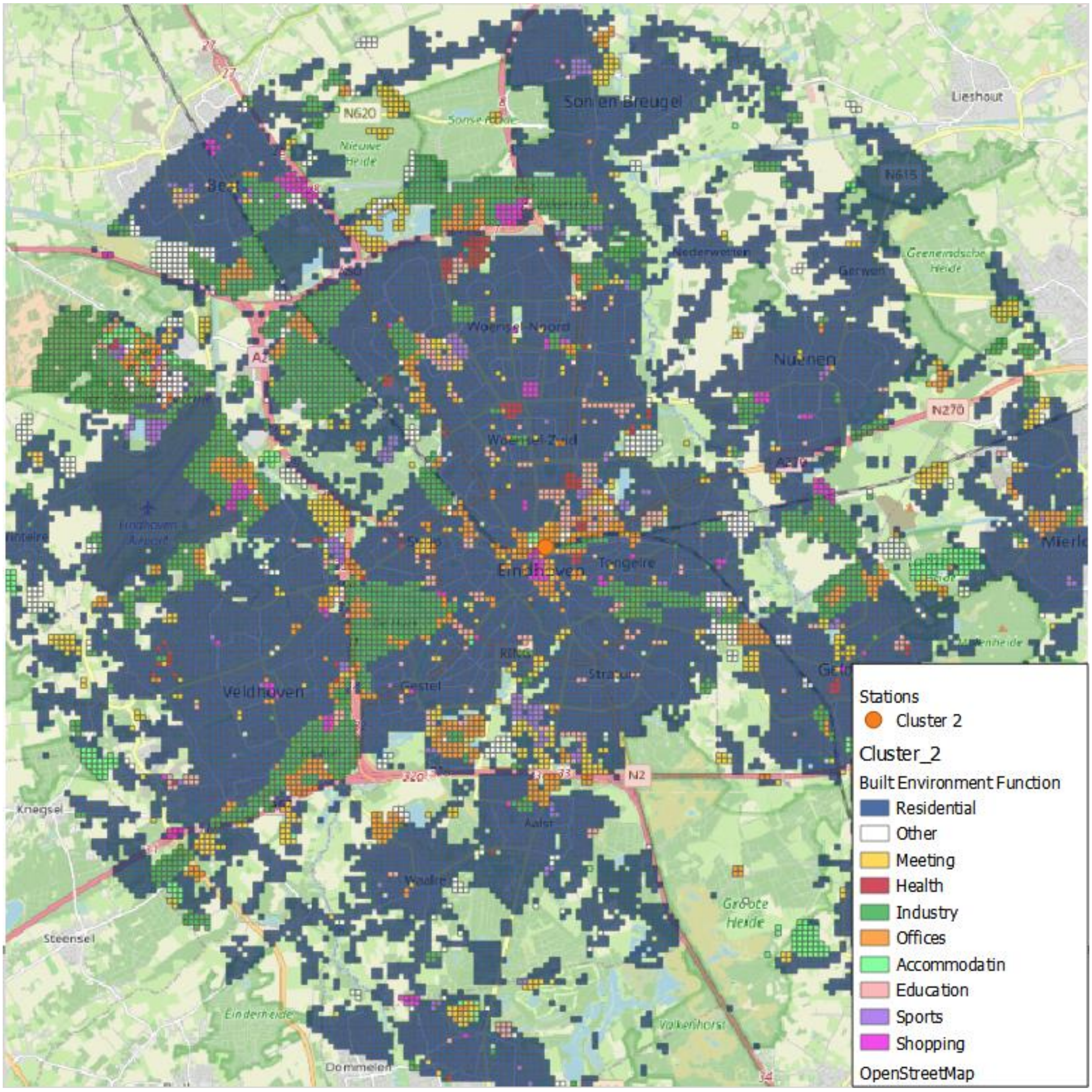


Figure 43: Eindhoven, dominant area function

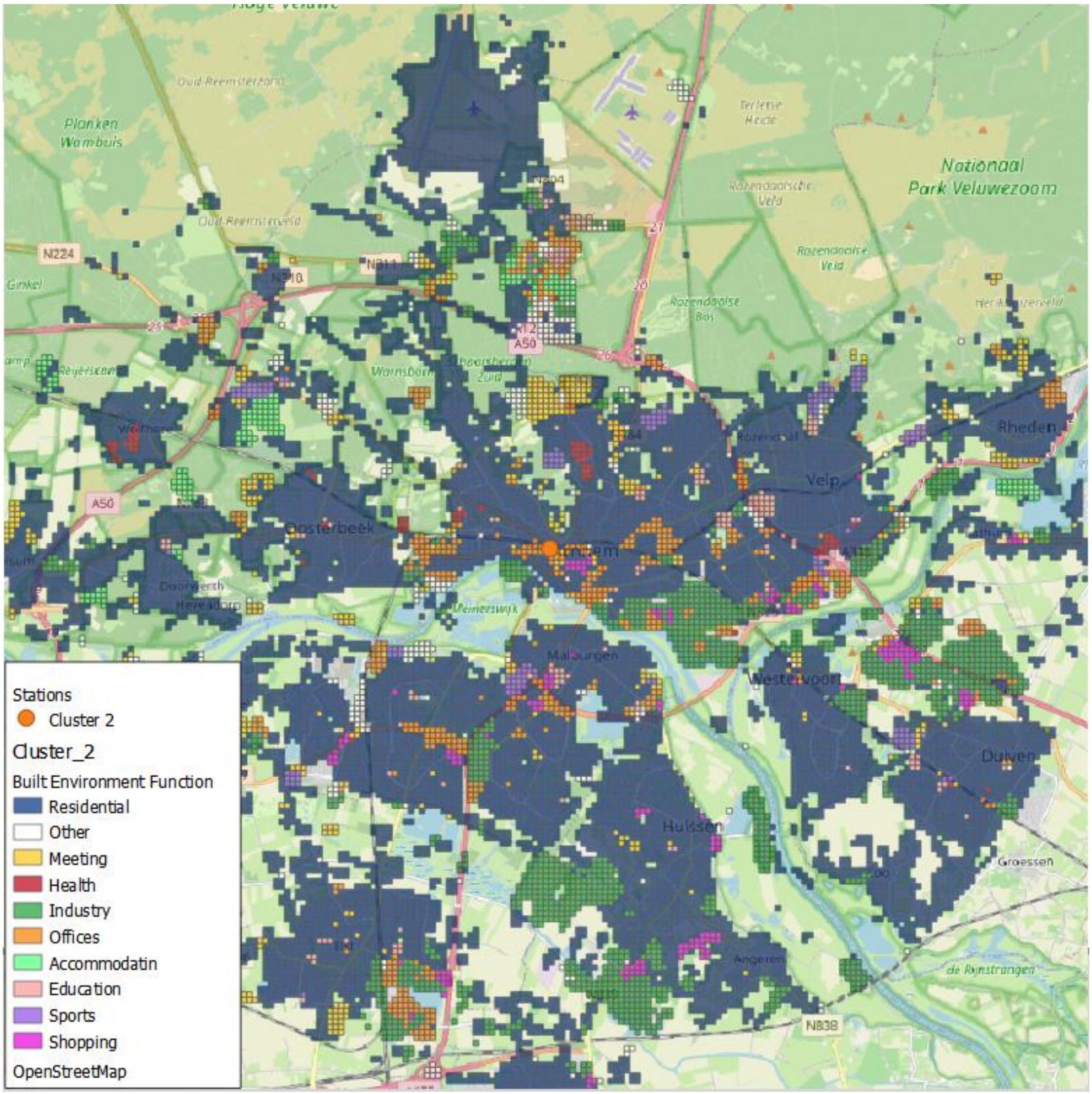


Figure 44: Arnhem, dominant area function

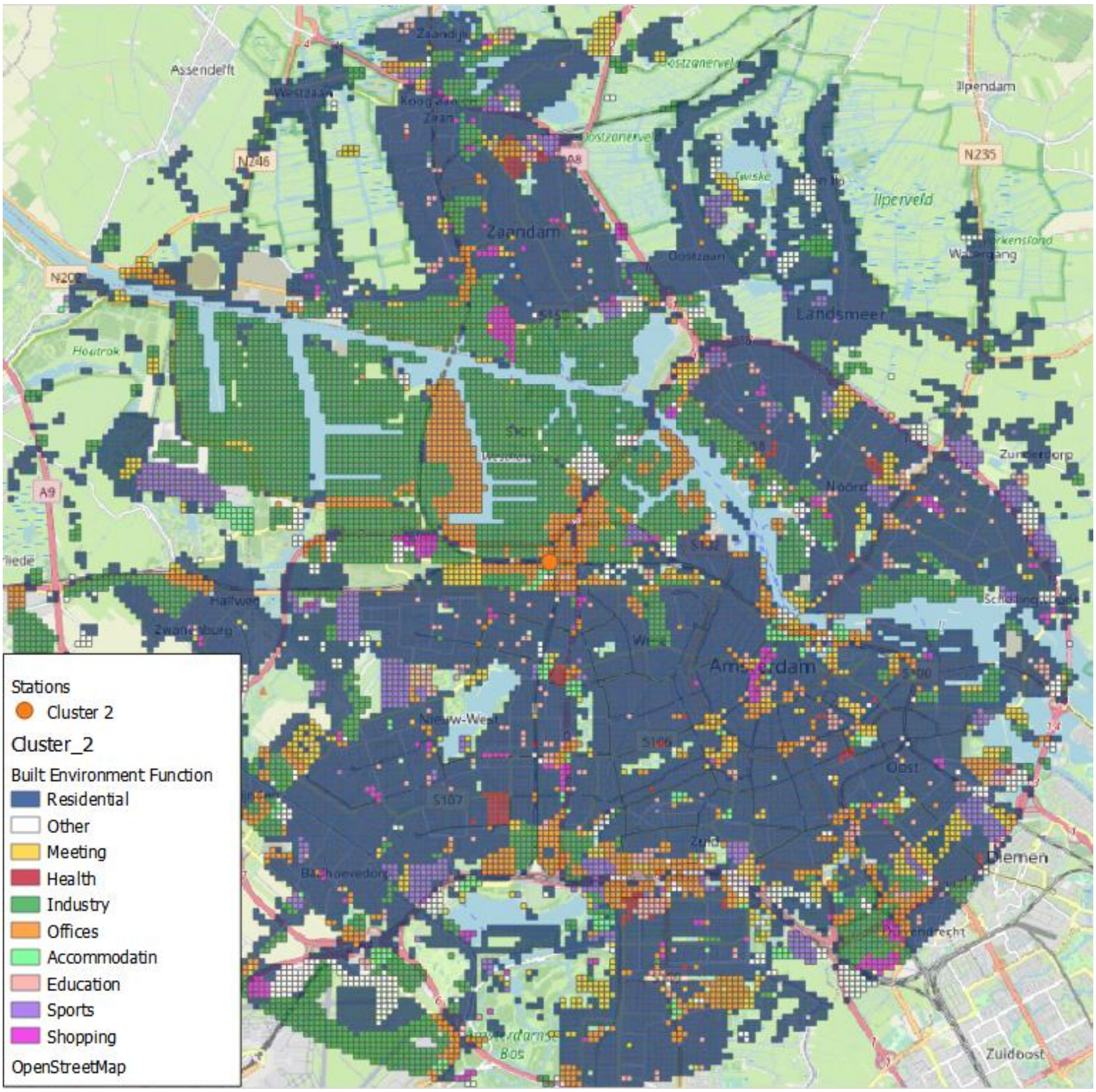


Figure 45: Amsterdam, dominant area function

8.2.4 Point of interest

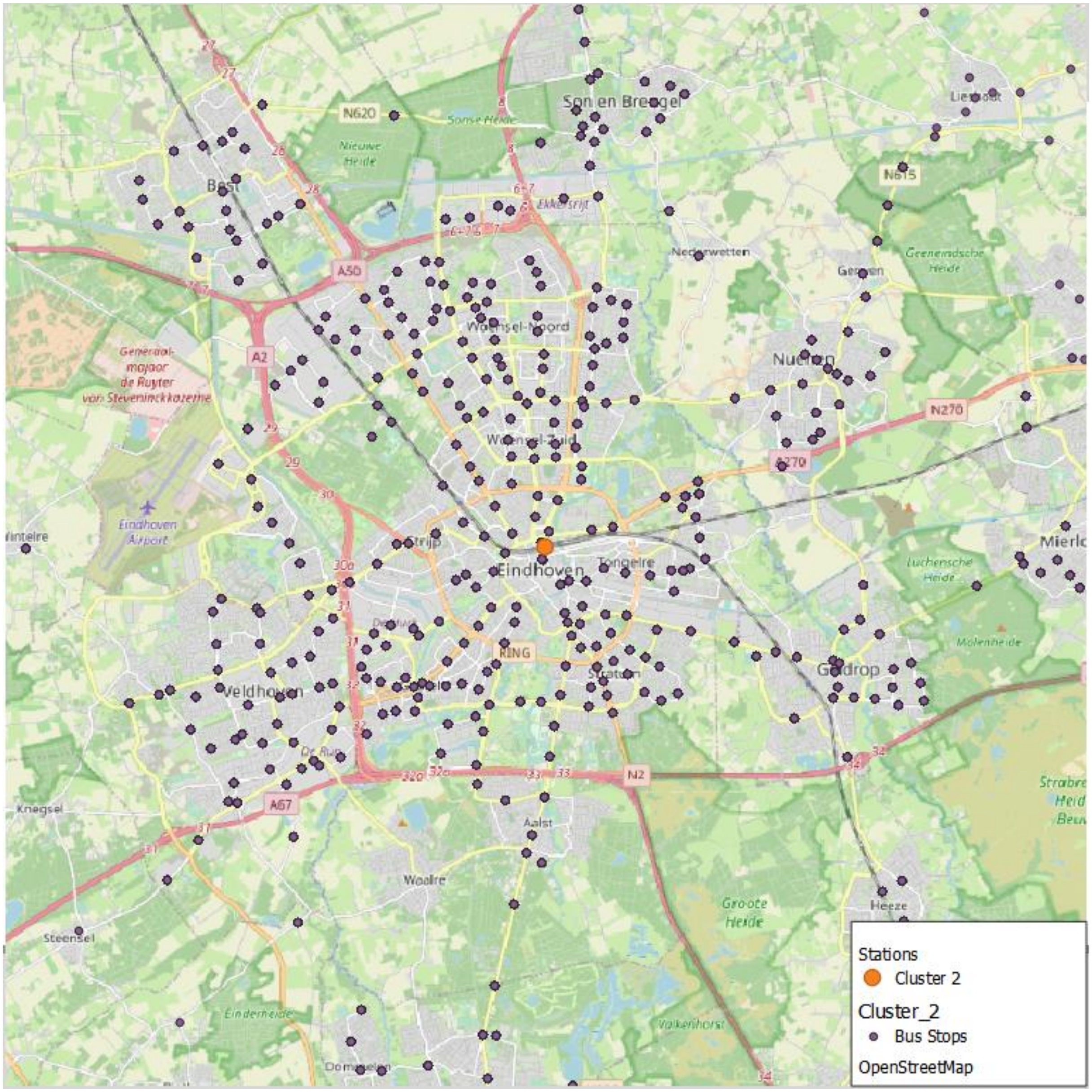


Figure 46: Eindhoven, bus stops

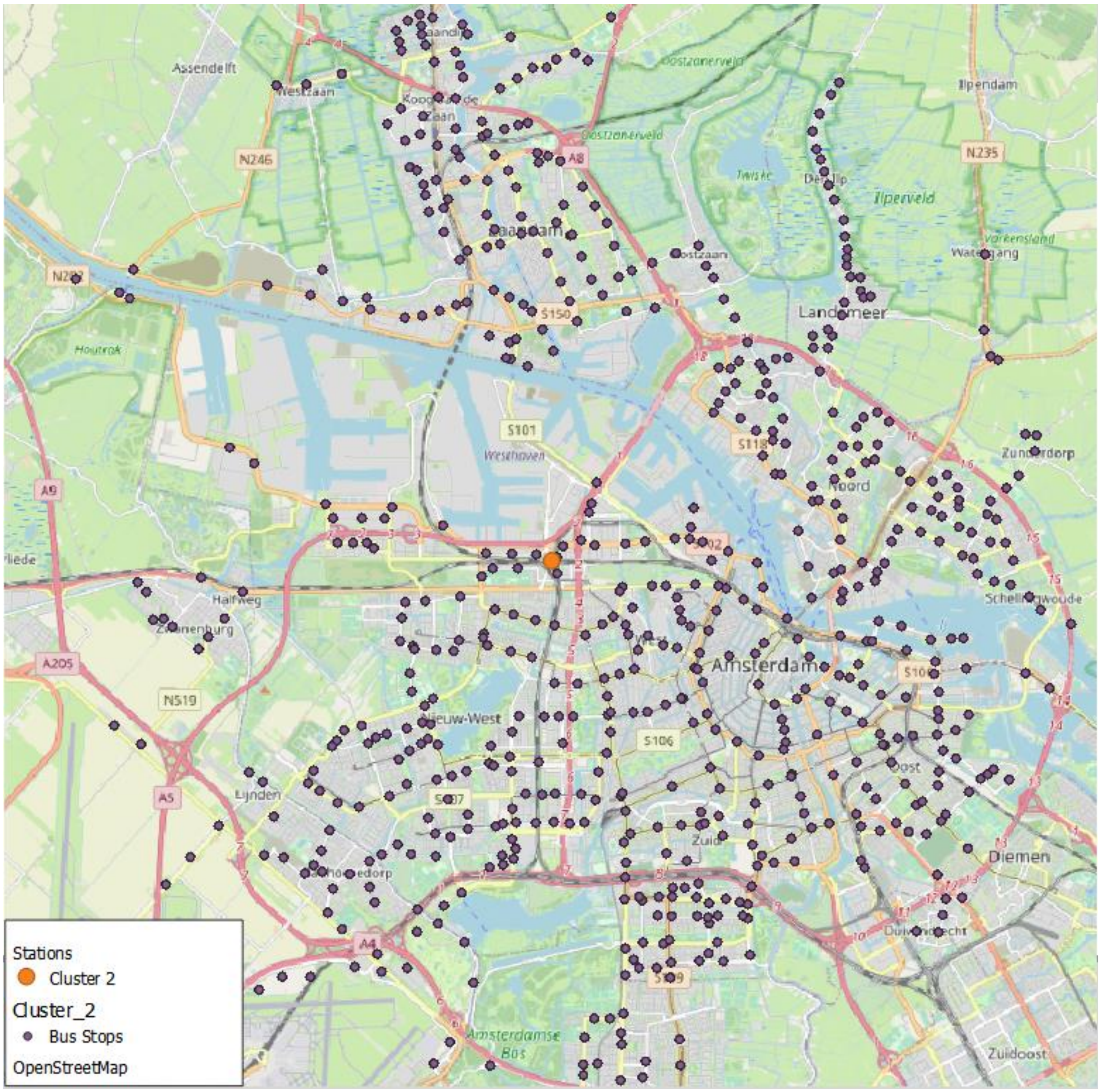


Figure 47: Amsterdam, bus stops

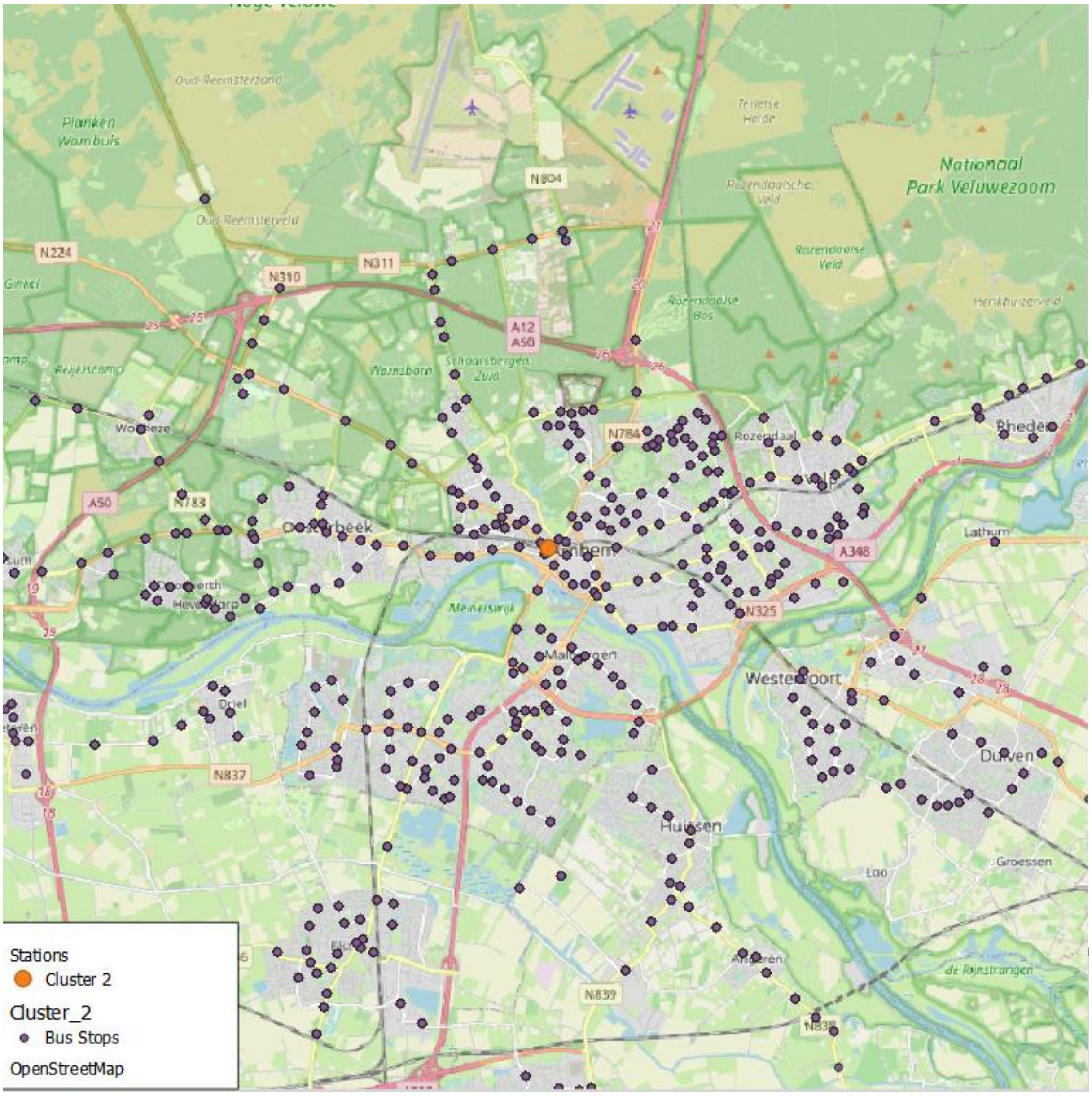


Figure 48: Arnhem, bus stops

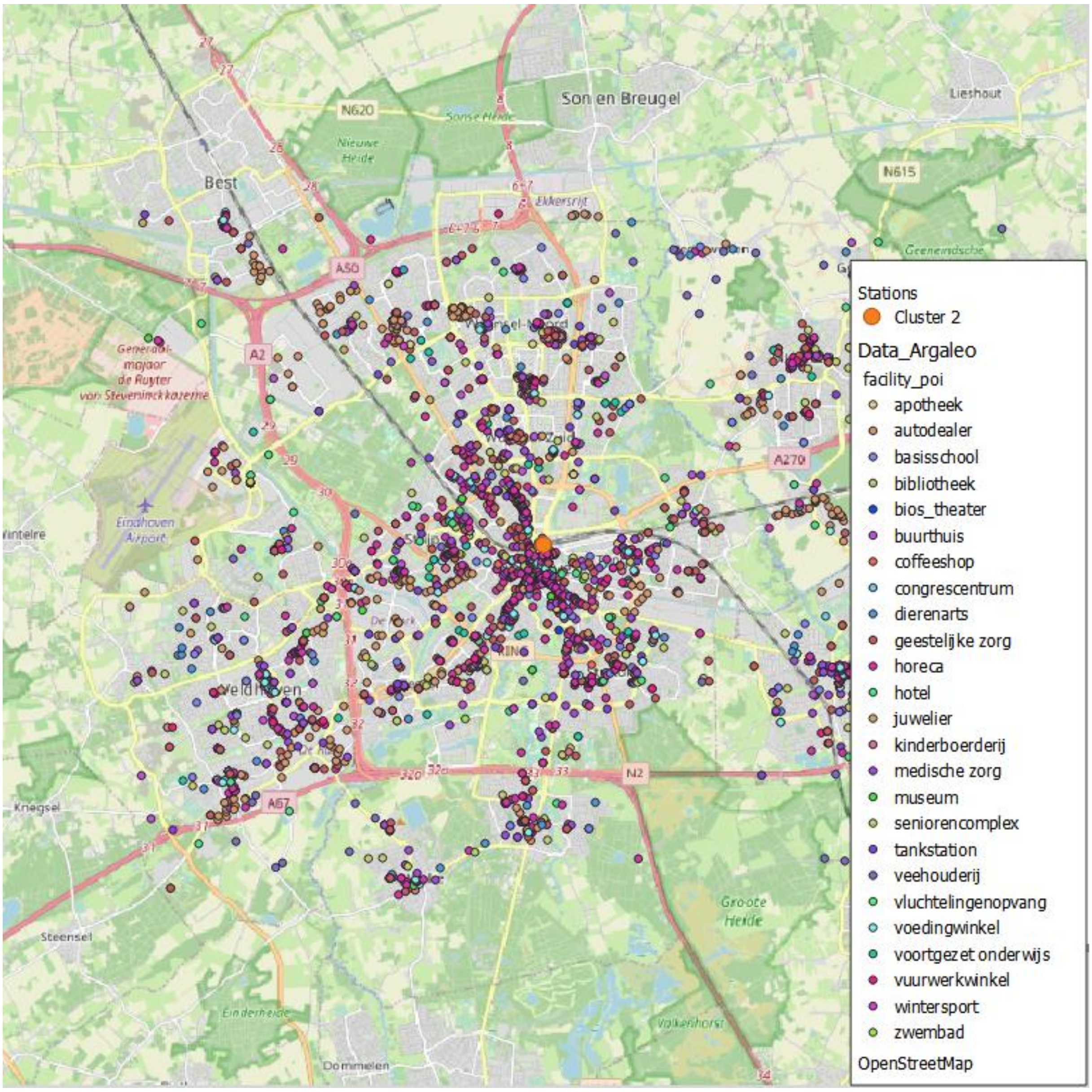


Figure 49: Eindhoven, facilities

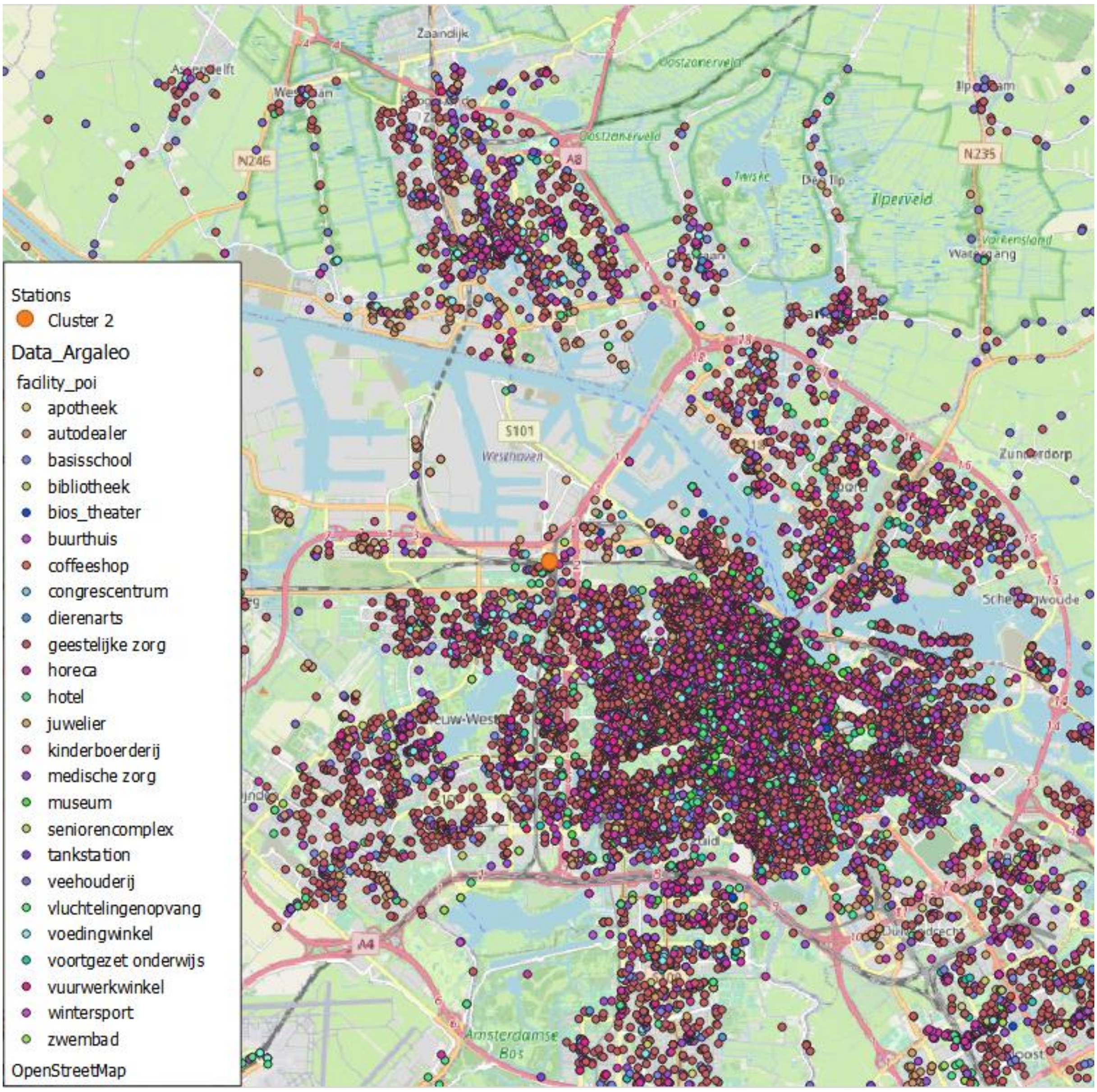


Figure 50: Amsterdam, facilities

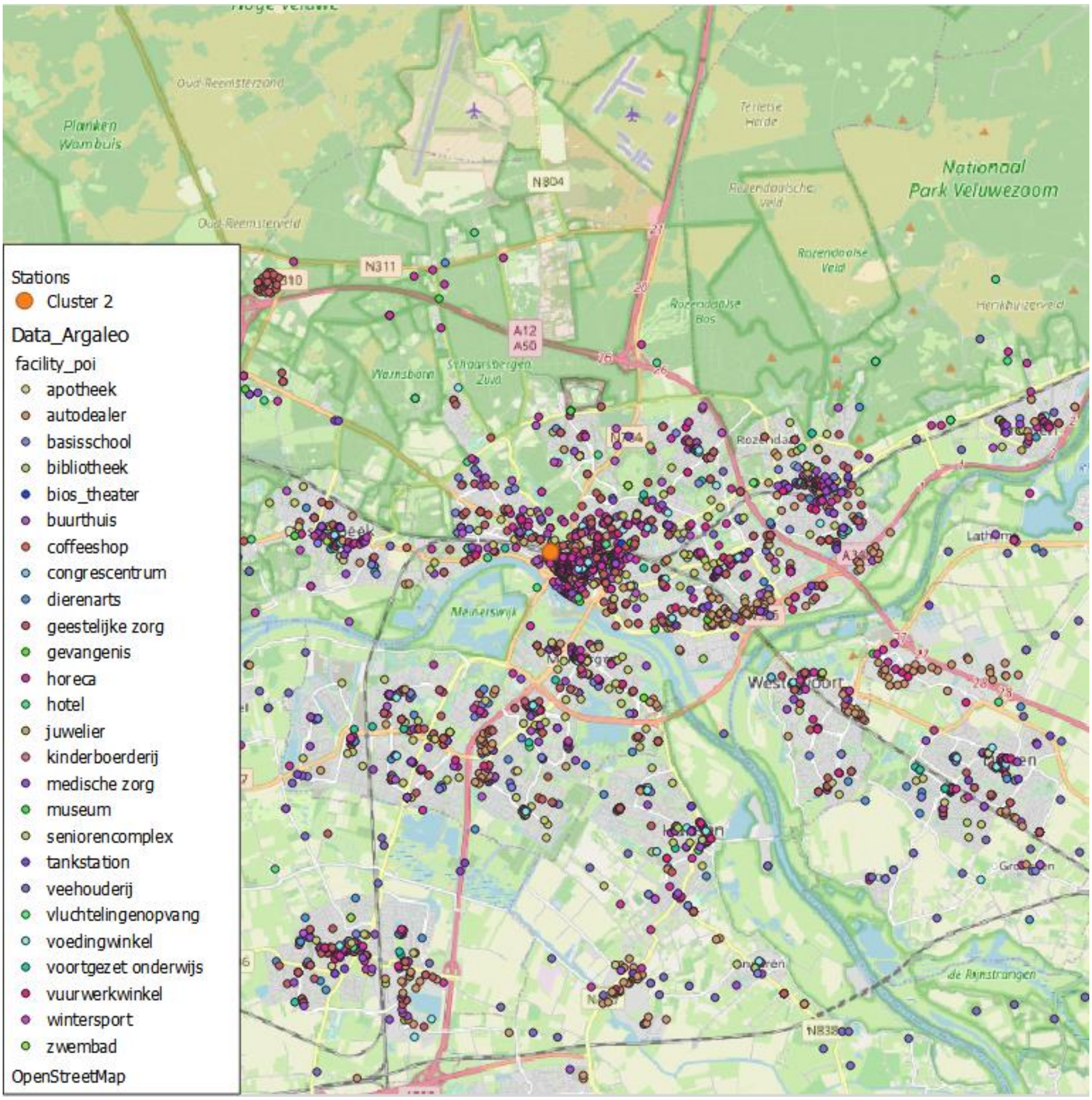


Figure 51: Arnhem, facilities

8.3 Appendix C – Station location selection

8.3.1 Variable selection

Table 25: Available variables for station clustering

Variable	Description	Consideration
Trips divided by active bicycles	Ratio on bicycle usage, therefore how many times a bicycle is used on average per day. Next, it ensures that the absolute amount of trips is divided more fairly comparing locations	More trips per bicycle means that destinations are more close, more dense or have asynchrony functions (working over the day and leisure in the evening)
Private bicycle egress share	Share of the private bicycle as egress mode from train station	More bicycle egress trips might indicate a better cycling environment
Attraction	The attraction rate of a station	A higher attraction rate means more egress trips
Education/Study	Percentage of 1 out of 3 travel motives as indicated by train passengers	Travel motives indicate destinations
Social/recreative	Percentage of 1 out of 3 travel motives as indicated by train passengers	Travel motives indicate destinations
Work/Business	Percentage of 1 out of 3 travel motives as indicated by train passengers	Travel motives indicate destinations
Trips OV-fiets	Amount of OV-fiets trips per year (only weekdays, 2019)	More trips means more destinations or more popular destinations
>12h	Percentage of bookings that are longer than 12 hours, which might indicate recreational usage	Trip duration might correlate and indicate the destination type or travel motive
< 10h	Percentage of bookings that are shorter than 10 hours, which might indicate commuters	Trip duration might correlate and indicate the destination type or travel motive
Distance	The average distance of private bicycle egress mode (this data is only known for 28 stations)	Trip distance indicate the willingness to cycle in the environment.

8.3.2 Correlation matrix

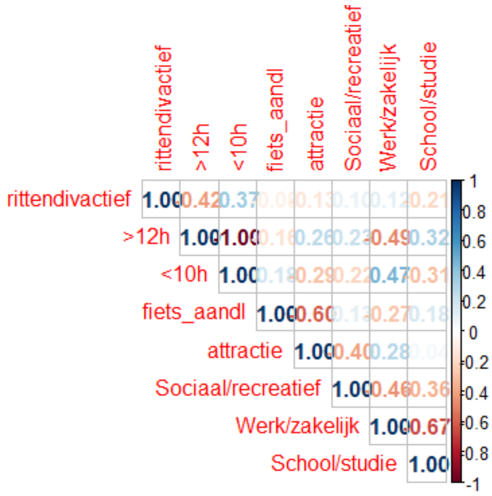


Figure 52: Correlation matrix available variables station clustering

8.3.3 Optimal number of clusters measure

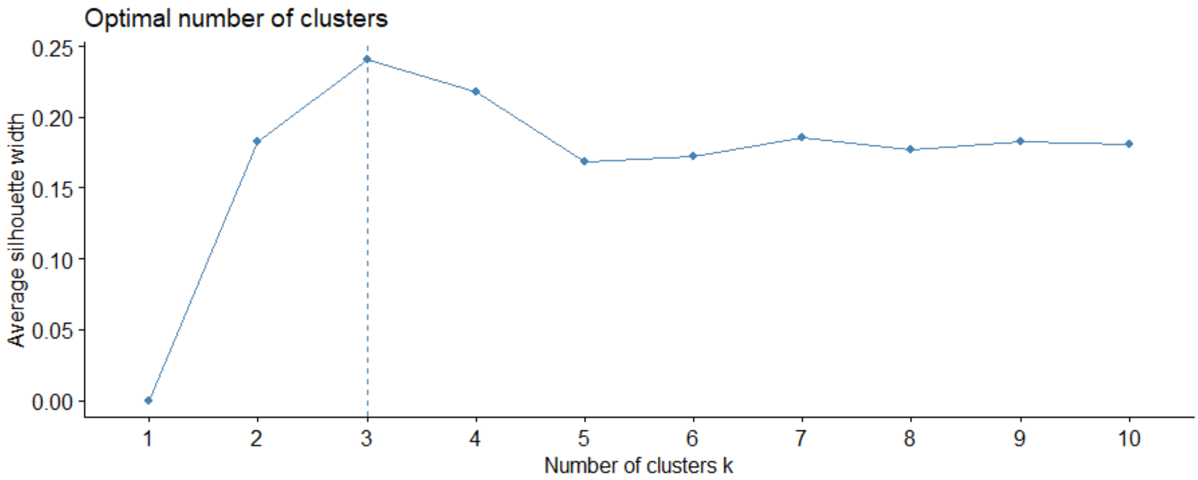


Figure 53: Output of silhouette method

8.3.4 Station cluster list

Table 26: Train stations per cluster

Cluster	Train station
1	Ede-Wageningen
1	Groningen
1	Maastricht
1	Nijmegen
1	Utrecht Vaartsche Rijn
2	Utrecht Centraal
2	Amsterdam Amstel
2	Amsterdam Centraal
2	Amsterdam Sloterdijk
2	Amsterdam Zuid
2	Arnhem Centraal
2	Den Haag Centraal
2	Den Haag Hollands Spoor
2	Eindhoven Centraal
2	Leeuwarden
2	Rotterdam Centraal
2	Utrecht Centraal
2	Zwolle
3	Alkmaar
3	Apeldoorn
3	Assen
3	Breda
3	Delft
3	Deventer
3	Gouda
3	Haarlem
3	Hilversum
3	Leiden Centraal
3	Tilburg

8.4 Appendix D – Survey questions

8.4.1 Original version

NS stations streeft voortdurend naar verbetering van het station en de fietsenstallingen.
Uw mening is hierbij van groot belang. Daarom vragen wij u deze vragenlijst in te vullen.
Het invullen duurt slechts 5 minuten en uw antwoorden worden anoniem verwerkt.

1. Hoeveel bestemmingen gaat u bezoeken met deze OV-fiets huurperiode?

- ☐ 1 bestemming
- ☐ 2 bestemmingen
- ☐ 3 of meer bestemmingen
- ☐ Zeg ik liever niet

2. a. Wat is de straatnaam en postcode van uw <u>belangrijkste</u> bestemming?

- ☐ Straatnaam ________ en/of Postcode ____ __
- ☐ Geen specifieke bestemming (rondrit)
- ☐ Zeg ik liever niet

b. Mocht u de straatnaam en/of postcode niet weten, probeert u dan zo goed mogelijk uw bestemming te omschrijven.

- ☐ Bestemming, bijvoorbeeld "Café Jansen" of "Bioscoop de Kuip" ________

c. Mocht u dit niet weten probeert u dan de volgende vragen te beantwoorden:

- ☐ Wat is de geschatte reistijd in minuten naar uw bestemming van deze OV-fiets rit? ________
- ☐ Wat is de geschatte afstand in kilometers naar uw bestemming van deze OV-fiets rit? ________

3. Wat is voor u de <u>belangrijkste</u> reden om vandaag deze rit met OV-fiets te maken?

- ☐ Van / naar werk
- ☐ Zakenreis / dienstreis
- ☐ Van / naar school, studie, cursus, opleiding
- ☐ Bezoek aan familie / kennissen, zieken(huis)bezoek
- ☐ Winkelen
- ☐ Vakantie / uitstapje / dagje weg
- ☐ Sport / hobby
- ☐ Anders, namelijk ________

4. Wie betaalt er voor deze OV-fiets rit?

- ☐ Uzelf
- ☐ Uw werkgever
- ☐ Iemand anders betaald voor mij
- ☐ Weet ik niet / Zeg ik liever niet

5. Maakt u deze OV-fiets rit vaker, zo ja, hoe vaak?

- ☐ Nee
- ☐ Ja, dagelijks
- ☐ Ja, 1x per week of vaker
- ☐ Ja, 1-3x per maand
- ☐ Ja, 1-11x per jaar
- ☐ Ja, minder dan 1x per jaar

6. Als een OV-fiets niet beschikbaar was geweest, had u deze rit dan ook gemaakt? Zo ja, met welk vervoersmiddel?
U kunt hier meerdere antwoorden geven.

- ☐ Nee
- ☐ Ja, een eigen auto
- ☐ Ja, een eigen (elektrische) fiets
- ☐ Ja, een eigen brommer/scooter
- ☐ Ja, een andere deelfiets
- ☐ Ja, een deelscooter
- ☐ Ja, een deelauto
- ☐ Ja, de bus, tram of metro
- ☐ Ja, de trein
- ☐ Ja, een taxi
- ☐ Anders, namelijk ________

7. Hoe tevreden bent u alles bij elkaar genomen over het openbaar vervoer (bus/tram/metro) in het algemeen, welk rapportcijfer zou u daar aan geven?
(1 = zeer slecht, 10 = zeer goed)

1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

8. Hoe tevreden bent u alles bij elkaar genomen over de OV-fiets in het algemeen, welk rapportcijfer zou u daar aan geven?
(1 = zeer slecht, 10 = zeer goed)

1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Ga verder op de achterzijde >>>

9. Om welke reden maakt u vooral gebruik van OV-fiets in het algemeen?
U kunt hier meerdere antwoorden geven.

- ☐ Snelheid
- ☐ Gemak
- ☐ Vrijheid
- ☐ Comfort
- ☐ Gezondheid
- ☐ Milieu
- ☐ Kosten
- ☐ Anders, namelijk ______

10. Hoe vaak gebruikt(e) u de OV-fiets voor en tijdens de coronaperiode?

		Voor Corona	Tijdens Corona	Nu
Wat was/is de frequentie van het gebruik van OV-fiets in het algemeen?	1x per week of vaker	☐	☐	☐
	1-3x per maand	☐	☐	☐
	1-11x per jaar	☐	☐	☐
	Minder dan 1x per jaar	☐	☐	☐
	Niet	☐	☐	☐
Gebruikte u de OV-fiets	Meer doordeweeks	☐	☐	☐
	Meer in het weekend	☐	☐	☐
Zou u uw huidige rit ook maken in deze situaties?	Nee	☐	☐	☐
	Ja	☐	☐	☐
	Ja, maar niet met OV-fiets	☐	☐	☐

11. Wat is uw geboortejaar? ______

12. Met welk geslacht identificeert u zich? ☐ Man ☐ Vrouw ☐ Anders / Zeg ik liever niet

13. Welk van de volgende kaartsoorten gebruikt u?

- ☐ Reizen op saldo
- ☐ Studentenreisproduct
- ☐ Business card
- ☐ NS flex
- ☐ Anders, namelijk ______

14. Als u naar deze bestemming reist met bus/tram/metro, betaalt u dan voor deze rit?

- ☐ Ja
- ☐ Ja, maar met korting
- ☐ Nee, mijn werkgever betaald
- ☐ Nee, iemand anders betaald
- ☐ Nee, ik heb een vrij reizen abonnement
- ☐ Nee, want deze bestemming kan ik niet bereiken met bus/tram/metro
- ☐ Anders / Zeg ik liever niet

15. Wat is uw hoogst genoten opleiding?

- ☐ Geen of basisonderwijs
- ☐ LBO/VMBO (kader of beroepsgericht)/MBO 1/ VBO
- ☐ MBO 2, 3, 4 of MBO voor 1998
- ☐ Havo of VWO (met diploma) / HBS / MMS
- ☐ HBO of universitair propedeuse
- ☐ HBO of universitair bachelor/kandidaats
- ☐ HBO of universitair master/doctoraal/postdoctoraal
- ☐ Anders / Zeg ik liever niet

16. Wat is uw werksituatie?

- ☐ Schoolgaand / scholier
- ☐ Studerend / student
- ☐ Werkzaam in loondienst
- ☐ Werkzaam bij de overheid
- ☐ Zelfstandig ondernemer
- ☐ Freelancer of ZZP'er
- ☐ Vrijwilliger
- ☐ Werkloos / werkzoekend / bijstand
- ☐ Huisvrouw / huisman
- ☐ Gepensioneerd / VUT
- ☐ Arbeidsongeschikt
- ☐ Anders / Zeg ik liever niet

Hartelijk dank voor uw medewerking. U kunt de enquête weer inleveren bij de enquêteur.

Hieronder niets invullen! ———————————————— *KBM OV-fiets 22*

Datum (DD:MM) | Tijdstip (UU:MM) | Stationscode

Wat is het weer op de startlocatie / plaats van afname?

- ☐ Extreem zonnig en warm
- ☐ Zonnig
- ☐ Bewolkt
- ☐ Buien
- ☐ Regen
- ☐ Onweer
- ☐ Storm
- ☐ Mist

8.4.2 Translated questions

1. How many destinations do you visit this OV-fiets rental period?
    a. 1 destination
    b. 2 destinations
    c. 3 destinatoins
    d. Would rather not answer
2. A. What is the streetname and postal code of your <u>most important</u> destination?
    a. Streetname [line] and/or postal code [four letters, two digits]
    b. No specific destination (recreative tour)
    c. Would rather not answer
2. B. If you don't want to answer the previous question; try to describe your destination.
    a. Destination, for example "Café Jansen" or "Cinema de Kuip" [line]
2. C. If you don't know, try to answer the following questions:
    a. What is the estimated travel time in minutes to your destination? [line]
    b. What is the estimated travel distance in kilometers to your destination? [line]
3. What is your <u>most important</u> purpose to travel with OV-fiets today?
    a. To/From work
    b. Business trip
    c. To/From school, study, course
    d. Visit to family/friends/hospital visit
    e. Shopping
    f. Holiday/day trip
    g. Sport/Hobby
    h. Other, namely [line]
4. Who pays this OV-fiets trip
    a. You
    b. Your employer
    c. Somebody else pays for me
    d. I don't know / would rather not answer
5. Do you make this trip more often, and how often?
    a. No
    b. Yes, daily
    c. Yes, 1 times per week or more
    d. Yes, 1 to 3 times per month
    e. Yes, 1 to 11 times per year
    f. Yes, but less than 1 time per year
6. If the OV-fiets was not available today, did you make this trip anyway? And if yes, with what travel mode (multiple answers possible)?
    a. No
    b. Yes, own car
    c. Yes, own (e-)bicycle
    d. Yes, own scooter
    e. Yes, other shared bicycles
    f. Yes, shared scooter
    g. Yes, shared car
    h. Yes, bus/tram/metro
    i. Yes, train
    j. Yes, taxi

    k. Other, namely [line]
7. How satisfied are you with public transport in general, which grade would you rate that?
    a. 1 till 10
8. How satisfied are you with OV-fiets, which grade would you rate that?
    a. 1 till 10
9. With what reason did you use the OV-fiets in general?
    a. Speed
    b. Convenience
    c. Freedom
    d. Comfort
    e. Health
    f. Environment
    g. Costs
    h. Other, namely [line]
10. How many times did/do you use the OV-fiets during the Corona period?

		Before Corona	During Corona	Now
What was/is the frequency of OV-fiets in general?	1 times per week or more			
	1 to 3 times per month			
	1 to 11 times per year			
	Less than 1 time per year			
	Never			
Did you use the OV-fiets	More on weekdays			
	More on weekend days			
Would you make your current trip also in this situation?	No			
	Yes			
	Yes, but not with OV-fiets			

11. What is your year of birth?
    a. [4 digits]
12. With which gender do you identify?
    a. Male
    b. Female
    c. Other / Would rather not answer
13. Which OV-chipcard do you use?
    a. Travel on balance
    b. Student travel
    c. Business card
    d. NS Flex

    e. Other, namely [line]

14. If you travel to this destination with bus/tram/metro now, would you pay for this trip?
    a. Yes
    b. Yes, but with reduction
    c. No, my employer pays
    d. No, somebody else pays
    e. No, I have free travel
    f. No, I could not reach this destination with bus/tram/metro
    g. Other / Would rather not answer
15. What is your education level?
    a. No, or primary education
    b. LBO/VMBO (kader of beroepsgericht/MBO 1/VBO
    c. MBO 2, 3, 4 or MBO before 1998
    d. Havo or VWO (with graduation) / HBS / MMS
    e. HBO or university propaedeutic
    f. HBO or university bachelor/candidate
    g. HBO or university master/doctorate/post doctorate
    h. Other / Would rather not answer
16. What is your employmet situation?
    a. School going
    b. Student
    c. Salaried employee
    d. Salaried at government
    e. Independent entrepreneur
    f. Freelancer
    g. Job-seeking
    h. Houseman / housewife
    i. Retired / VUT
    j. Incapacitated
    k. Other / Would rather not answer

Questions for the surveyor:

17. What is the date?
    a. [month/day]
18. What is the time
    a. [hour/minute]
19. What is the train station?
    a. [stationcode]
20. What is the weather at the survey location?
    a. Extremely sunny
    b. Sunny
    c. Cloudy
    d. Showers
    e. Rain
    f. Thunderstorm
    g. Windstorm
    h. Fog

8.4.3 QR code survey

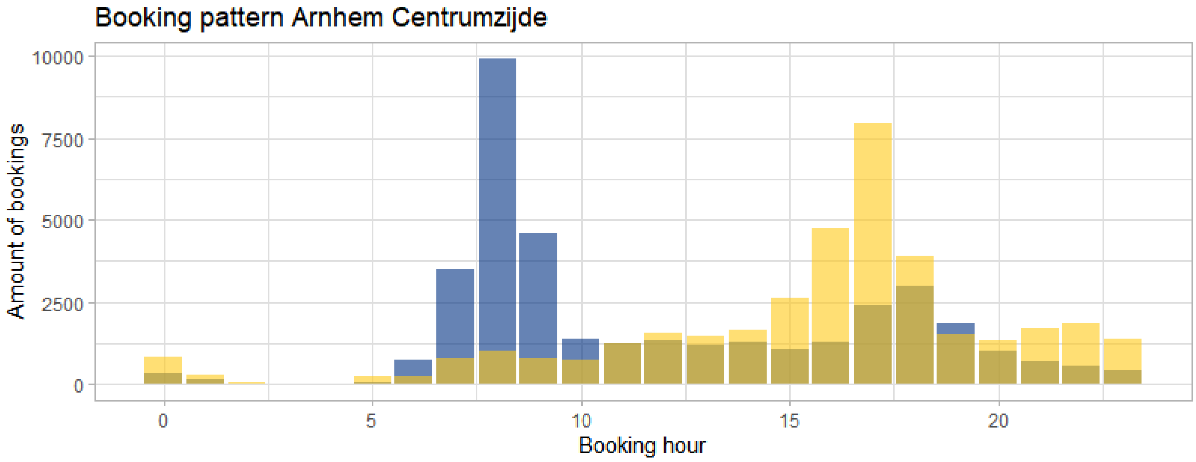
8.5 Appendix E – Survey results

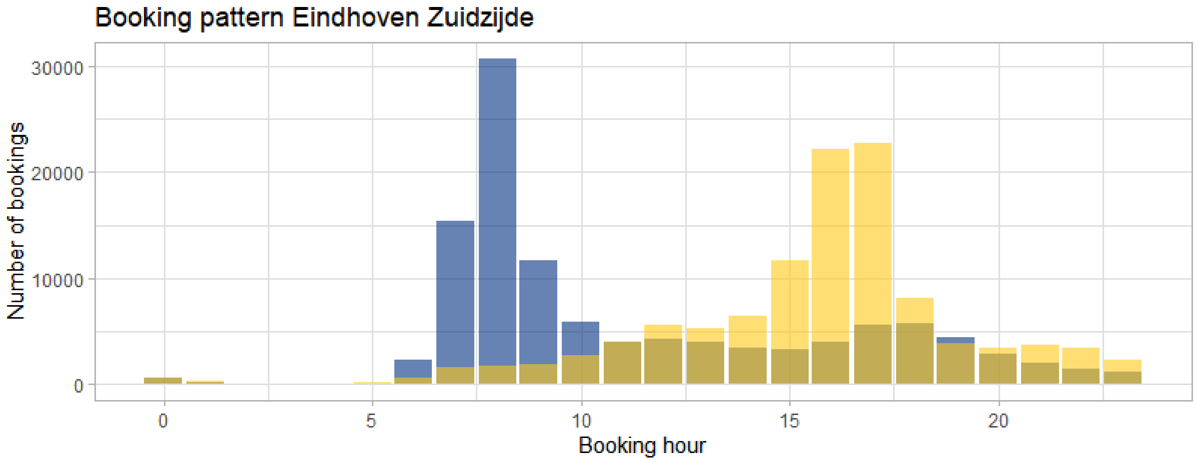
8.5.1 Surveying times

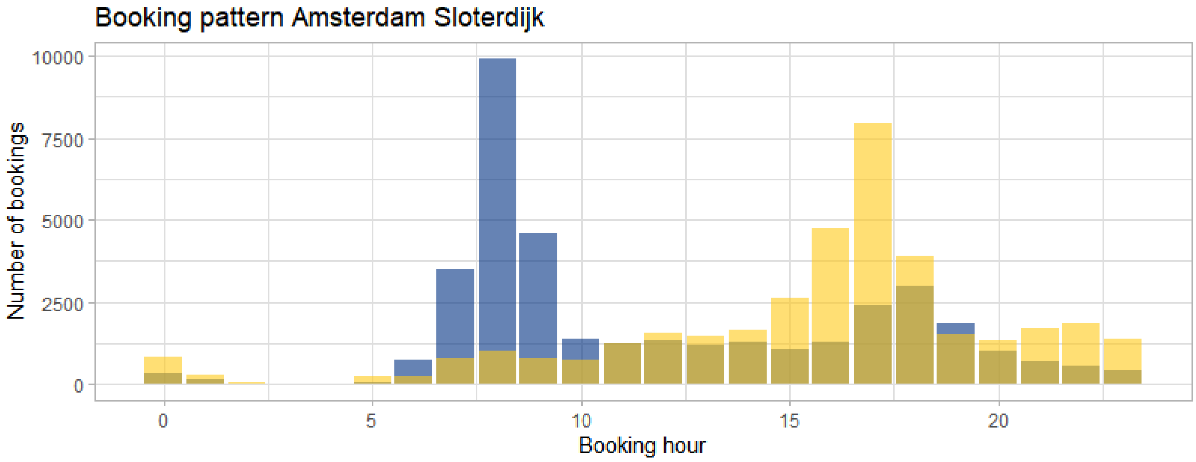
Rental location	Day	Date	Time	Weather condition
Groningen	Friday	22 april	7:00 – 11:00	Sunny
	Monday	9 may	8:30 – 12:30	Sunny
	Monday	9 may	12:30 – 16:30	Sunny
	Monday	25 april	15:00 – 19:00	Sunny/Cloudy
	Monday	2 may	15:00 – 19:00	Sunny/Cloudy
	Wednesday	4 may	15:00 – 19:00	Sunny
Maastricht	Wednesday	20 april	15:00 – 19:00	Sunny
	Thursday	21 april	15:00 – 19:00	Sunny
	Friday	22 april	15:00 – 19:00	Cloudy
	Monday	25 april	15:00 – 19:00	Cloudy
	Tuesday	26 april	15:00 – 19:00	Cloudy
	Thursday	28 april	15:00 – 19:00	Sunny
Nijmegen	Wednesday	11 may	8:00 – 12:00	Sunny
	Wednesday	11 may	15:00 – 19:00	Sunny
Amsterdam Sloterdijk	Monday	16 may	15:00 – 19:00	Sunny/Cloudy
	Tuesday	17 may	15:00 – 19:00	Cloudy
	Wednesday	4 may	15:00 – 19:00	Cloudy
	Monday	9 may	15:00 – 19:00	Sunny
	Wednesday	18 may	15:00 – 19:00	Sunny
Arnhem Centrumzijde	Friday	22 april	07:00 – 11:00	Sunny/Cloudy
	Monday	2 may	07:00 – 11:00	Sunny
	Wednesday	4 may	07:00 – 11:00	Sunny
	Monday	25 april	15:00 – 19:00	Showers
	Friday	29 april	15:00 – 19:00	Sunny/Cloudy
Eindhoven Zuidzijde	Monday	25 april	07:00 – 11:00	Cloudy
	Wednesday	11 may	07:00 – 11:00	Sunny/Cloudy
	Monday	16 may	07:00 – 11:00	Cloudy
	Friday	29 april	15:00 – 19:00	Sunny
	Monday	2 may	15:00 – 19:00	Sunny
	Wednesday	18 may	15:00 – 19:00	(Extremely) sunny
Apeldoorn	Monday	2 may	07:00 – 11:00	Cloudy
	Wednesday	4 may	07:00 – 11:00	Sunny/Cloudy
	Friday	6 may	07:00 – 11:00	Fog/Sunny
	Wednesday	20 april	15:00 – 19:00	Sunny
	Wednesday	11 may	15:00 – 19:00	Sunny
	Tuesday	17 may	15:00 – 19:00	Sunny/Cloudy
Delft CS Stalling 1	Tuesday	3 may	07:00 – 11:00	Cloudy
	Wednesday	18 may	07:00 – 11:00	Sunny
	Thursday	5 may	16:00 – 20:00	Cloudy
	Monday	16 may	16:00 – 20:00	Cloudy
	Friday	20 may	07:00 – 11:00	Cloudy
Hilversum	Friday	22 april	07:00 – 11:00	Sunny
	Monday	2 may	07:00 – 11:00	Sunny

	Wednesday	4 may	07:00 – 11:00	Sunny
	Monday	25 april	16:00 – 20:00	Cloudy/Showers
	Friday	6 may	16:00 – 20:00	N/A
	Tuesday	10 may	16:00 – 20:00	Sunny/Cloudy

8.5.2 Booking patterns







8.5.3 Descriptive result figures

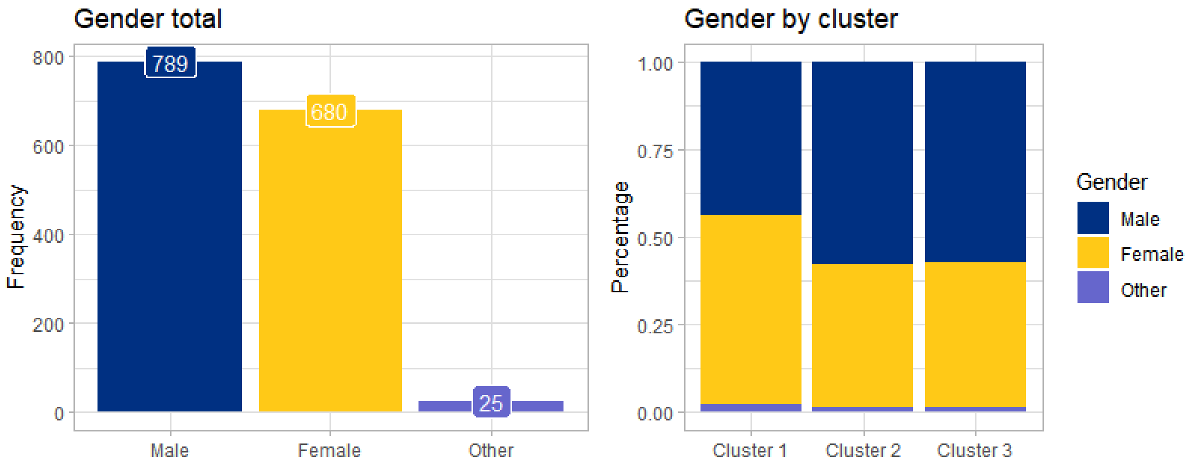


Figure 54: Results of gender of respondents

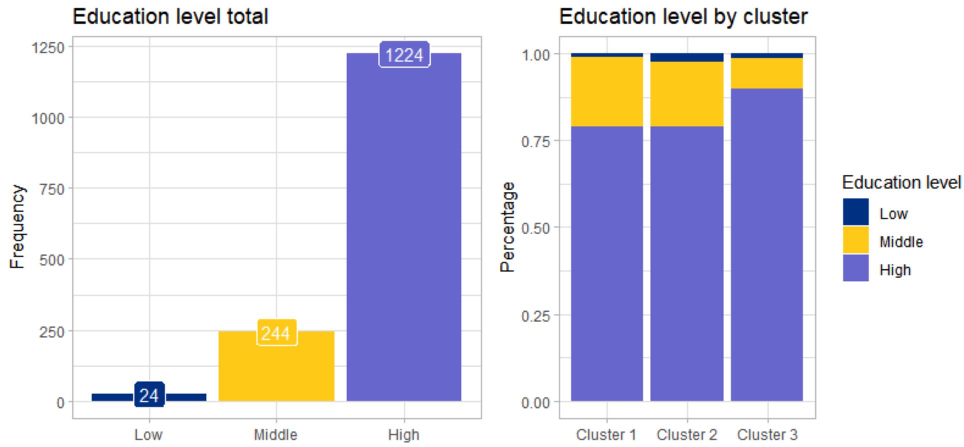


Figure 55: Results of the education level of respondents

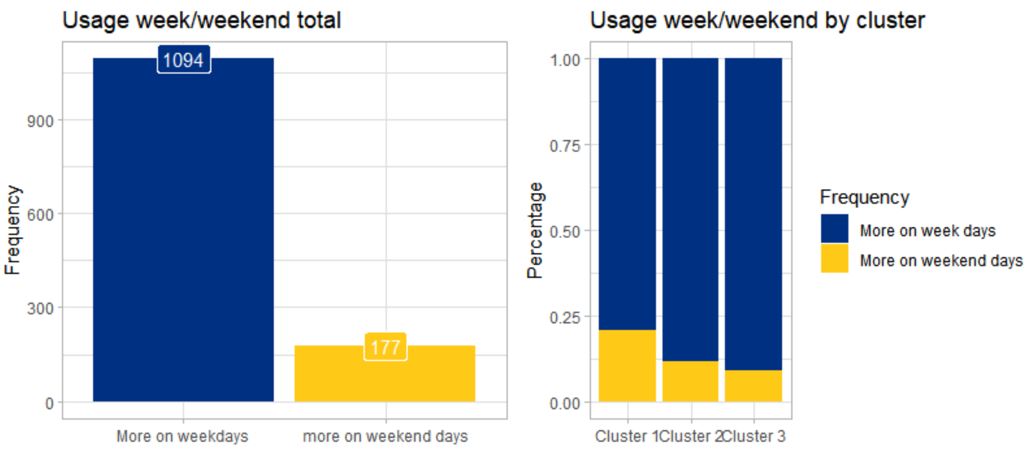


Figure 56: Results of weekday or weekend day usage OV-fiets

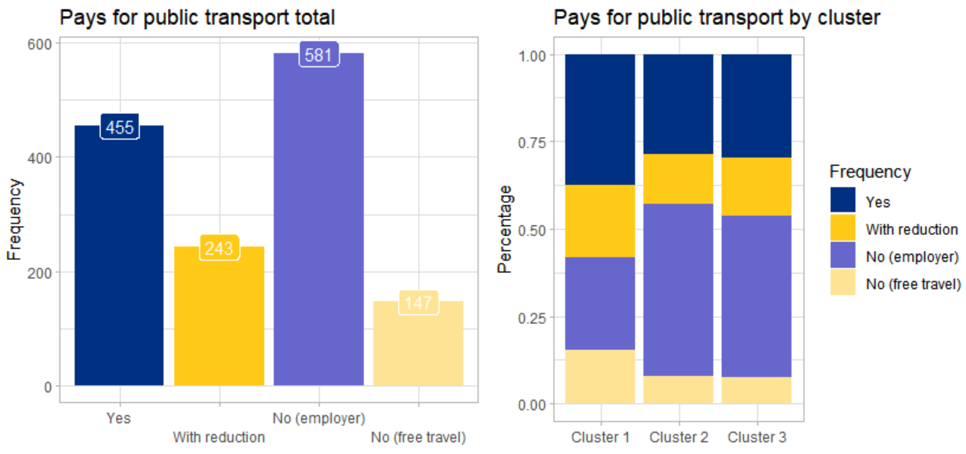


Figure 57: Results of respondents who pay for public transport

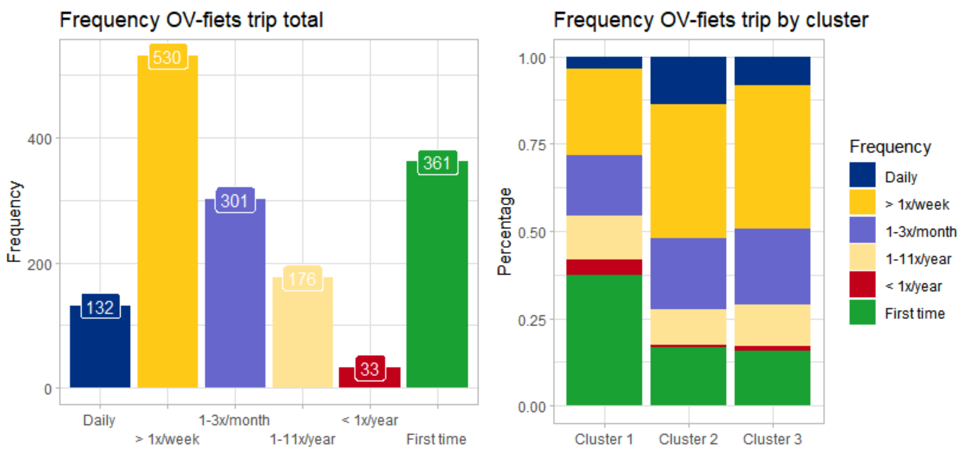


Figure 58: Results of OV-fiets usage frequency

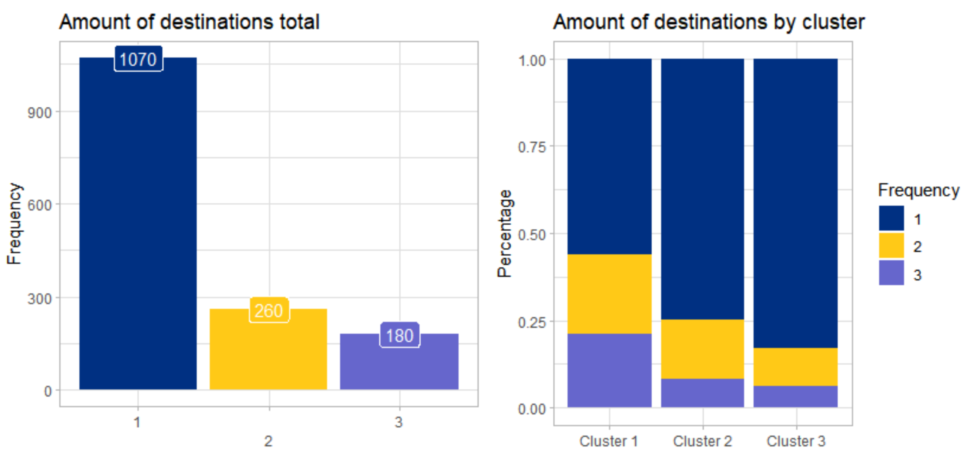


Figure 59: Results of the number of destinations in an OV-fiets round-trip

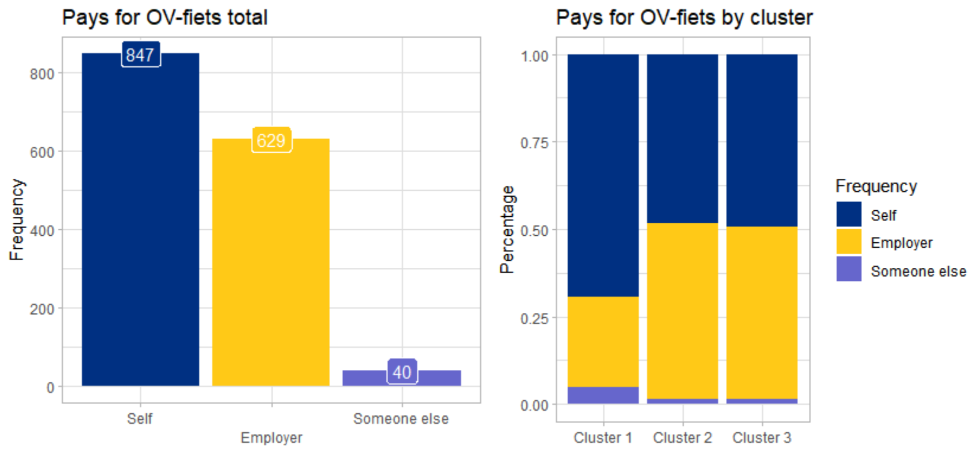


Figure 60: Results of OV-fiets users paying for the OV-fiets

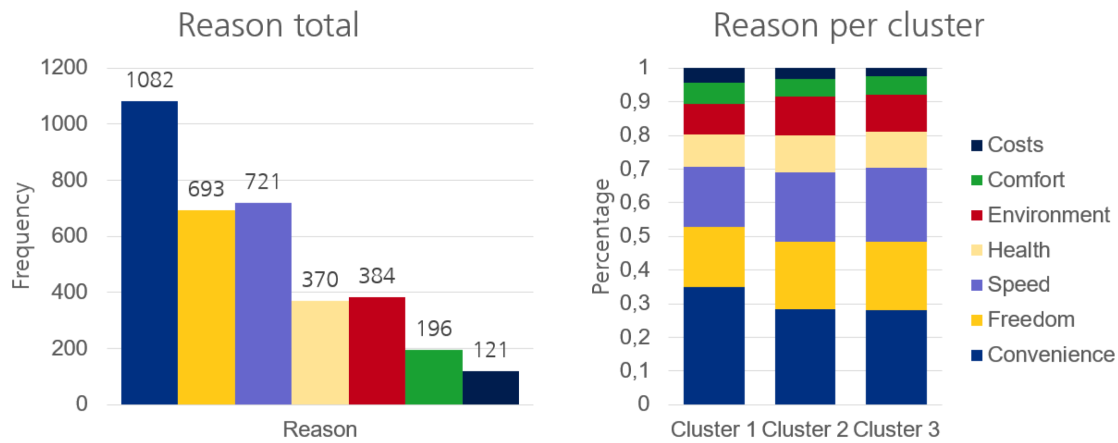


Figure 61: Results of reasons to choose OV-fiets

8.5.4 Destination results

Table 27: Filled in destinations

	Total	Cluster 1	Cluster 2	Cluster 3
N	1538	529	577	432
PC6	289 (19%)	63 (12%)	116 (20%)	110 (25%)
Adress	741 (48%)	225 (43%)	299 (52%)	217 (50%)
Location	285 (19%)	148 (28%)	90 (16%)	47 (11%)

Table 28: Geolocation results

	Total	Cluster 1	Cluster 2	Cluster 3
N	1538	529	577	432
Have wrongly coded location	376 (24%)	148 (28%)	120 (21%)	108 (25%)
No location at all	223 (14% | 59%)	93 (12% | 63%)	72 (12% | 60%)	58 (13% | 54%)
Have location, but no google API	91 (6% | 24%)	19 (4% | 13%)	30 (5% | 25%)	28 (6% | 26%)
Have location, but wrong google API	76 (5% | 20%)	36 (7% | 24%)	18 (3% | 15%)	22 (5% | 20%)
Useful destination N	1162 (76%)	381 (72%)	457 (79%)	324 (75%)

8.5.5 Destination geolocations

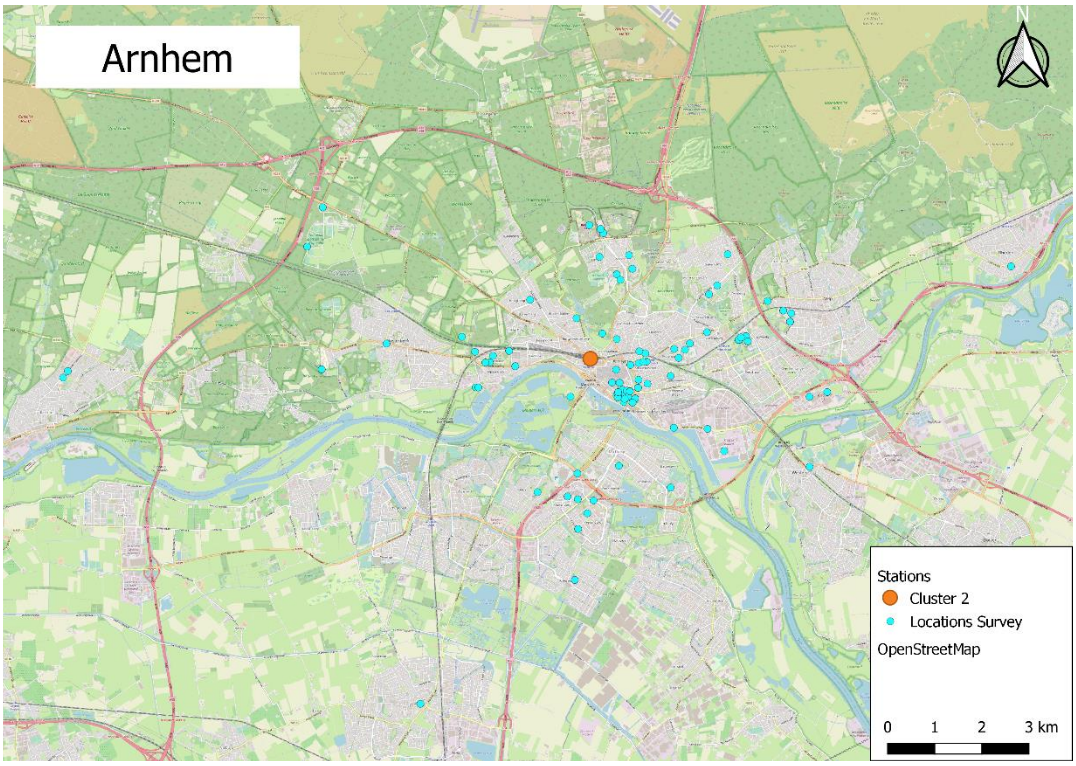


Figure 62: Arnhem, locations survey

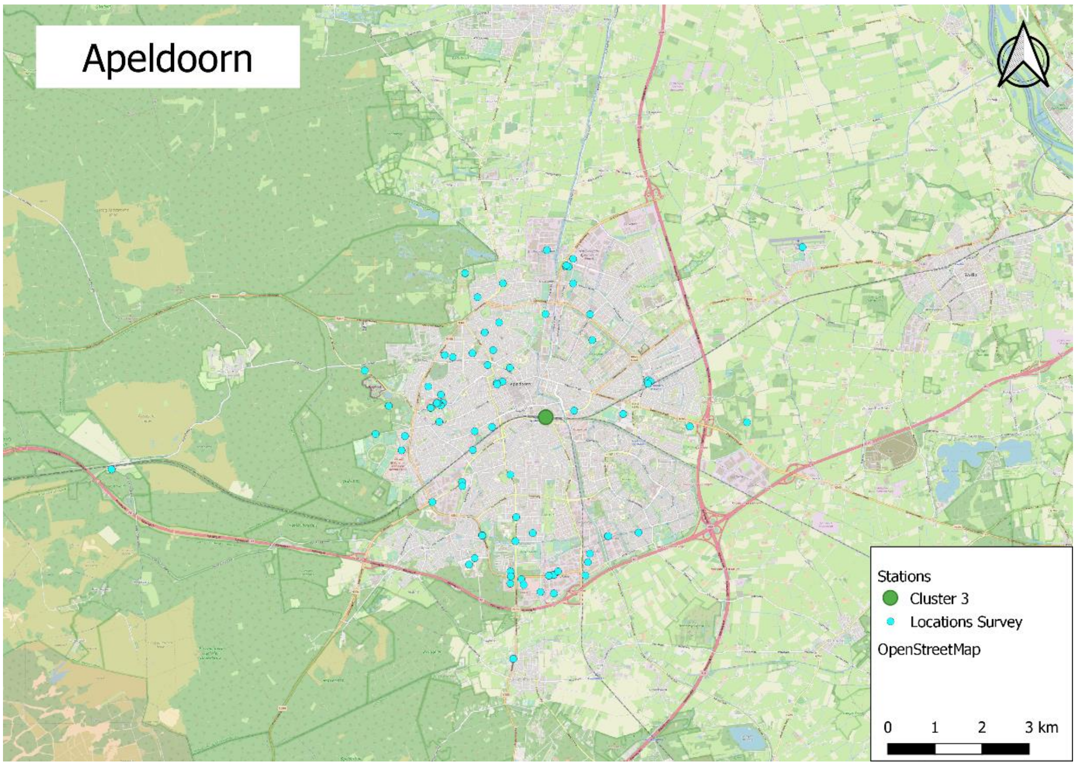


Figure 63: Apeldoorn, locations survey

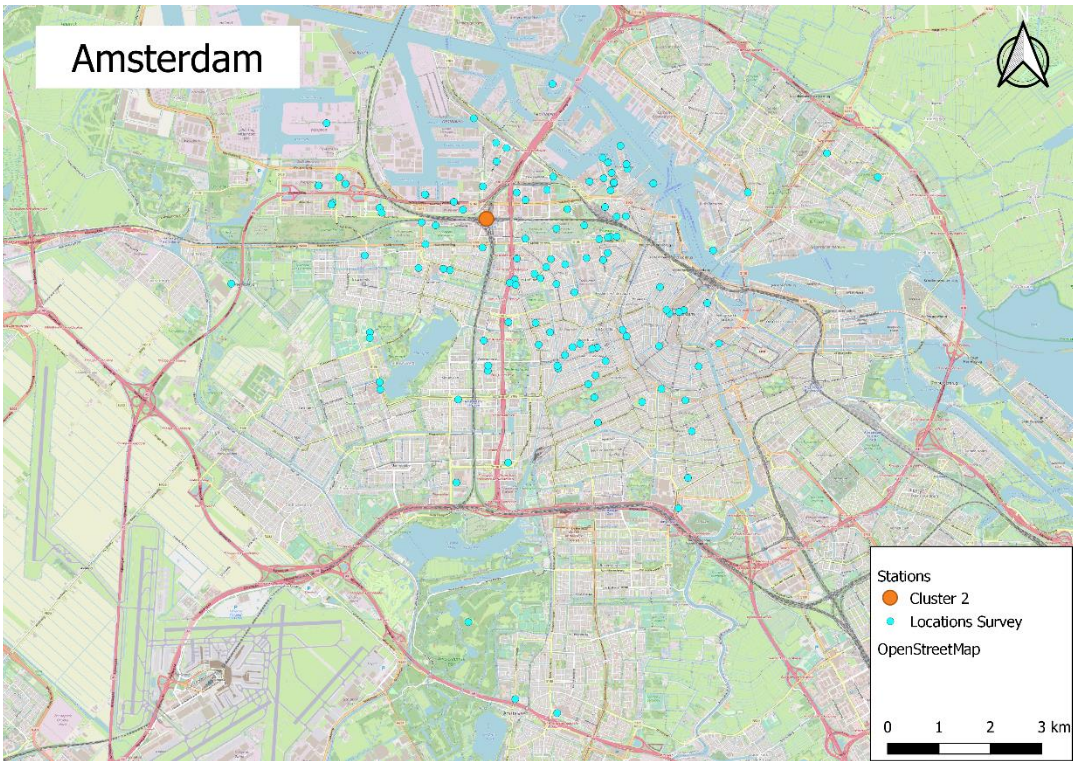


Figure 64: Amsterdam, locations survey

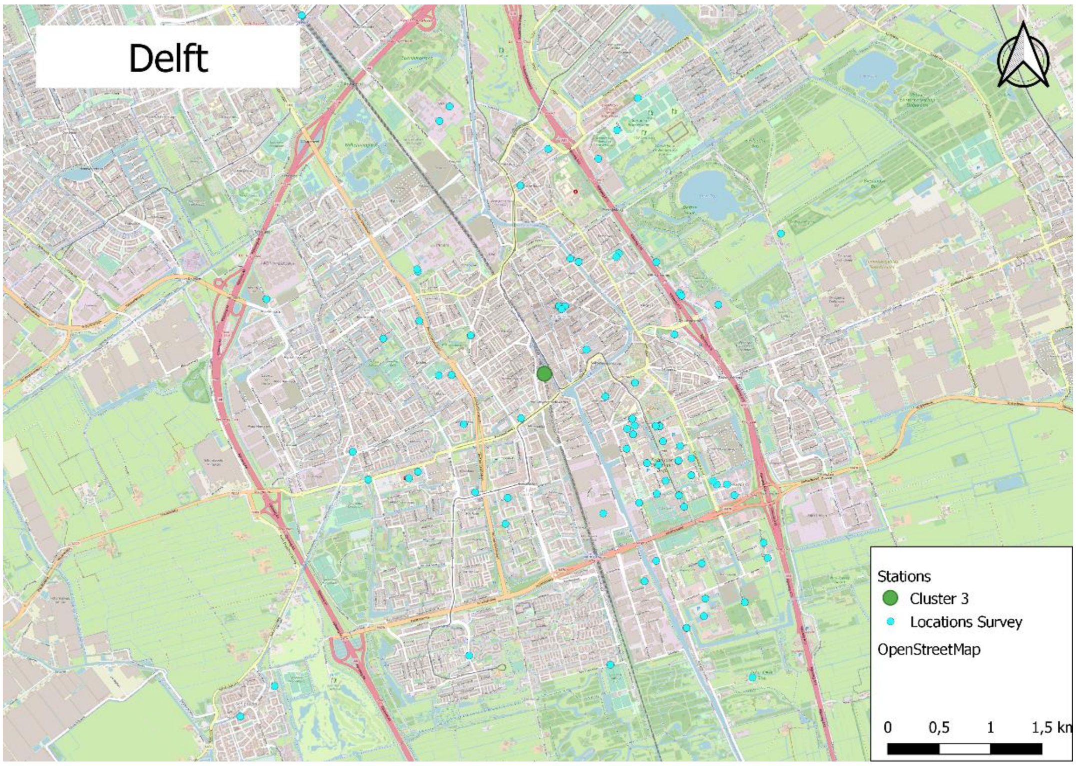


Figure 65: Delft, locations survey

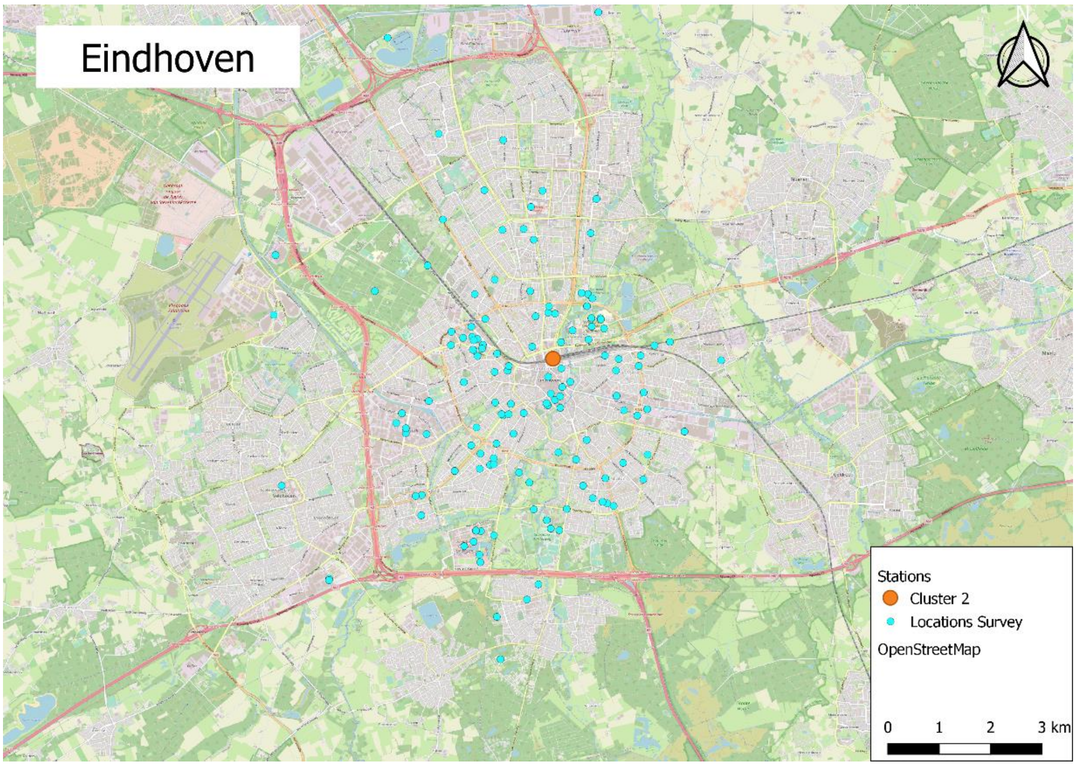


Figure 66: Eindhoven, locations survey

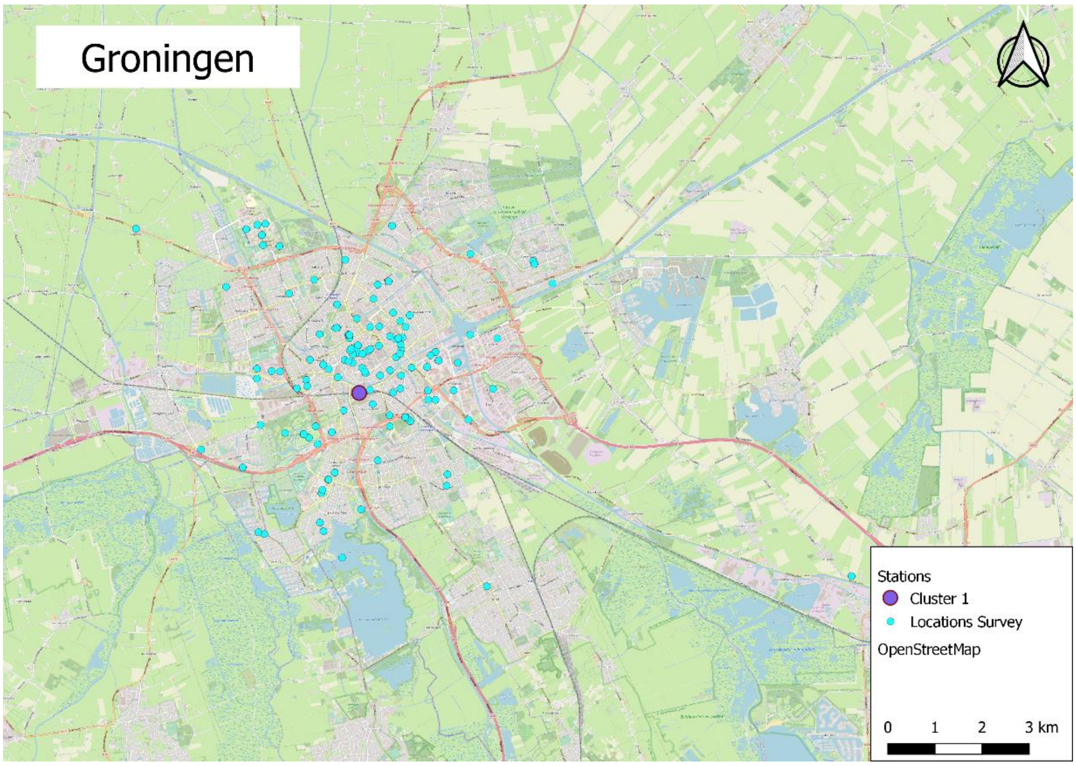


Figure 67: Groningen, locations survey

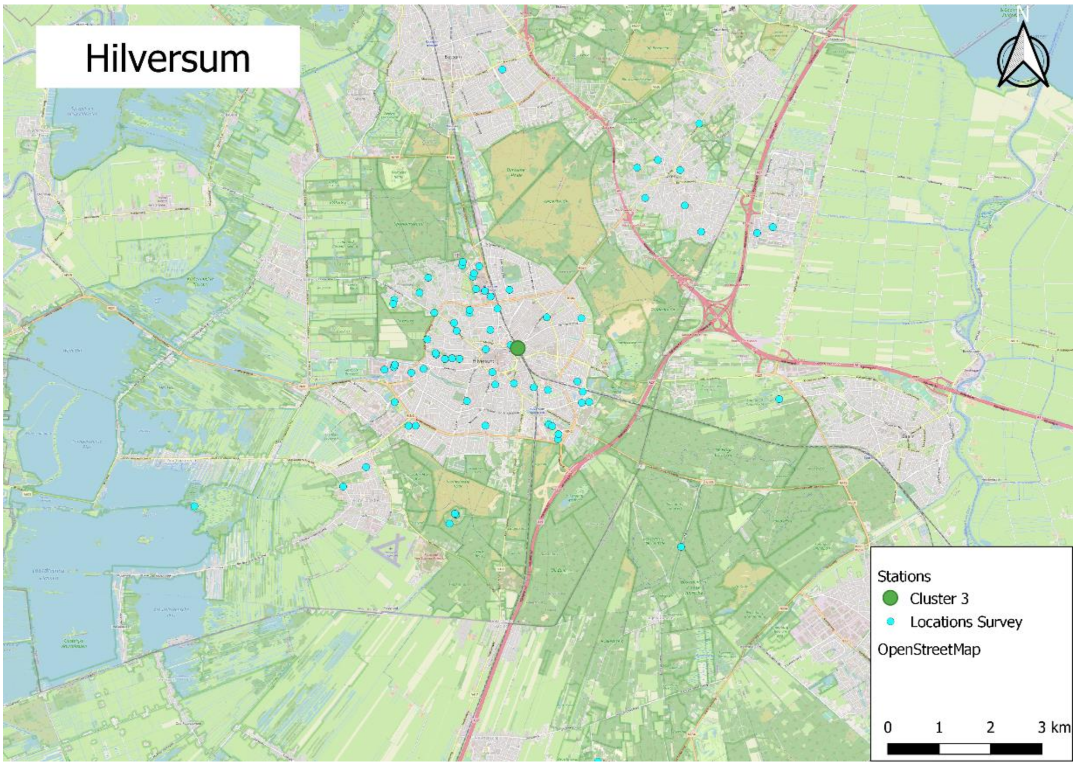


Figure 68: Hilversum, locations survey

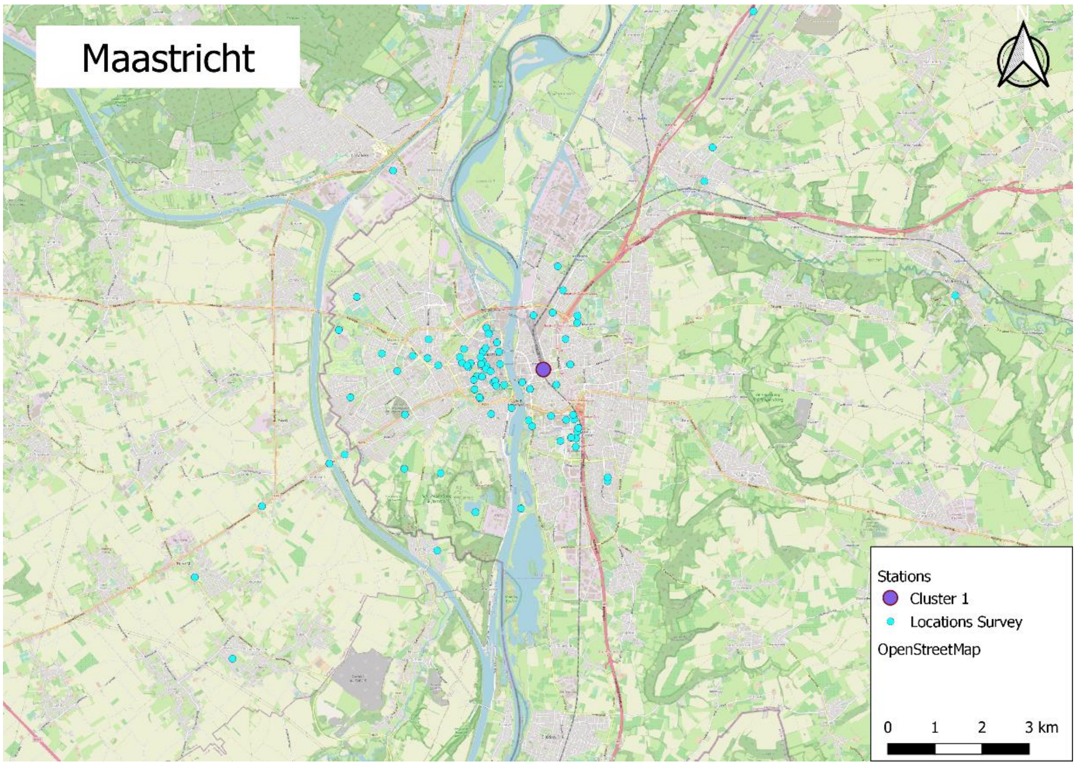


Figure 69: Maastricht, locations survey

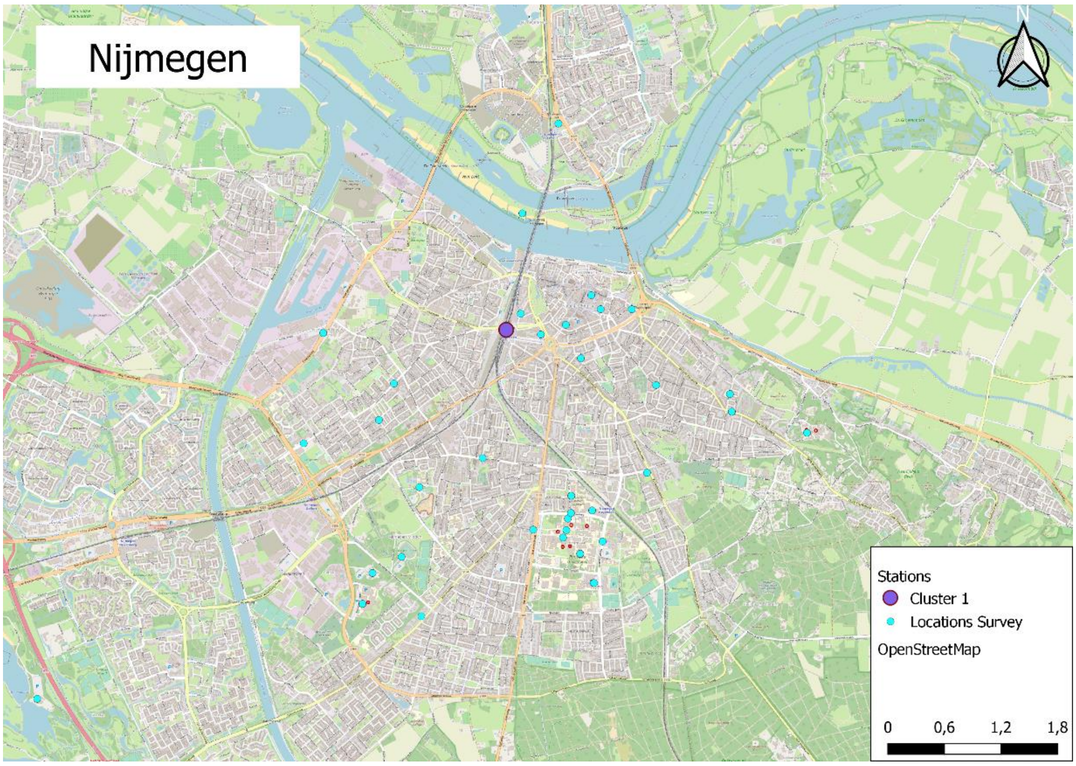


Figure 70: Nijmegen, locations survey

8.5.6 Cluster differences

Table 29: Difference between clusters and total dataset (x means falls <u>within</u> the 95% confidence interval for having <u>no</u> difference between the data)

Category	Base dataset	Total dataset			Cluster 1		Cluster 2
	Comparison	Cluster 1	Cluster 2	Cluster 3	Cluster 2	Cluster 3	Cluster 3
Gender	Male						x
	Female						x
Age	> 30						x
	31-40		x				x
	41-55			x			x
	56-65		x	x			x
	66 +	x	x	x	x	x	x
Education level	Low	x		x		x	
	Middle				x		
	High				x		
Working situation	School going						x
	Self-employed	x	x	x	x	x	x
	Employed						x
	Retired	x	x	x	x	x	x
Frequency OV-fiets general	> 1x/week						
	1-3x/month	x	x		x	x	
	1-11x/year						x
	>1/year						x
	No general use		x		x		x
Week/weekend usage	More weekend		x				
	More week						x
Pays for Public Transport	Yes			x			x
	With reduction			x		x	x
	No (free travel)						x
	No (employer pays)						x
PT satisfaction	Rating 7 or higher	x	x	x	x	x	x
	Rating 8 or higher	x	x	x	x	x	x
OV-fiets satisfaction	Rating 7 or higher	x	x	x	x	x	x
	Rating 8 or higher	x	x	x	x	x	X
Trip purpose	Work						x
	Business trip	x			x		
	Education						x
	Visits						x
	Shopping		x		x		x
	Leisure						
	Sports		x	x			x

	Other		x				x
Trip frequency	Daily			x			
	> 1x/week						x
	1-3x/month		x	x	x		x
	1-11x/year	x	x	x	x	x	x
	< 1x/year			x			x
	First time						x
Amount of destinations	1						x
	2		x				x
	3						x
Pays for OV-fiets	Self						
	Employer				x		
	Someone else				x		

8.6 Appendix F – Purpose model

8.6.1 Correlation matrix

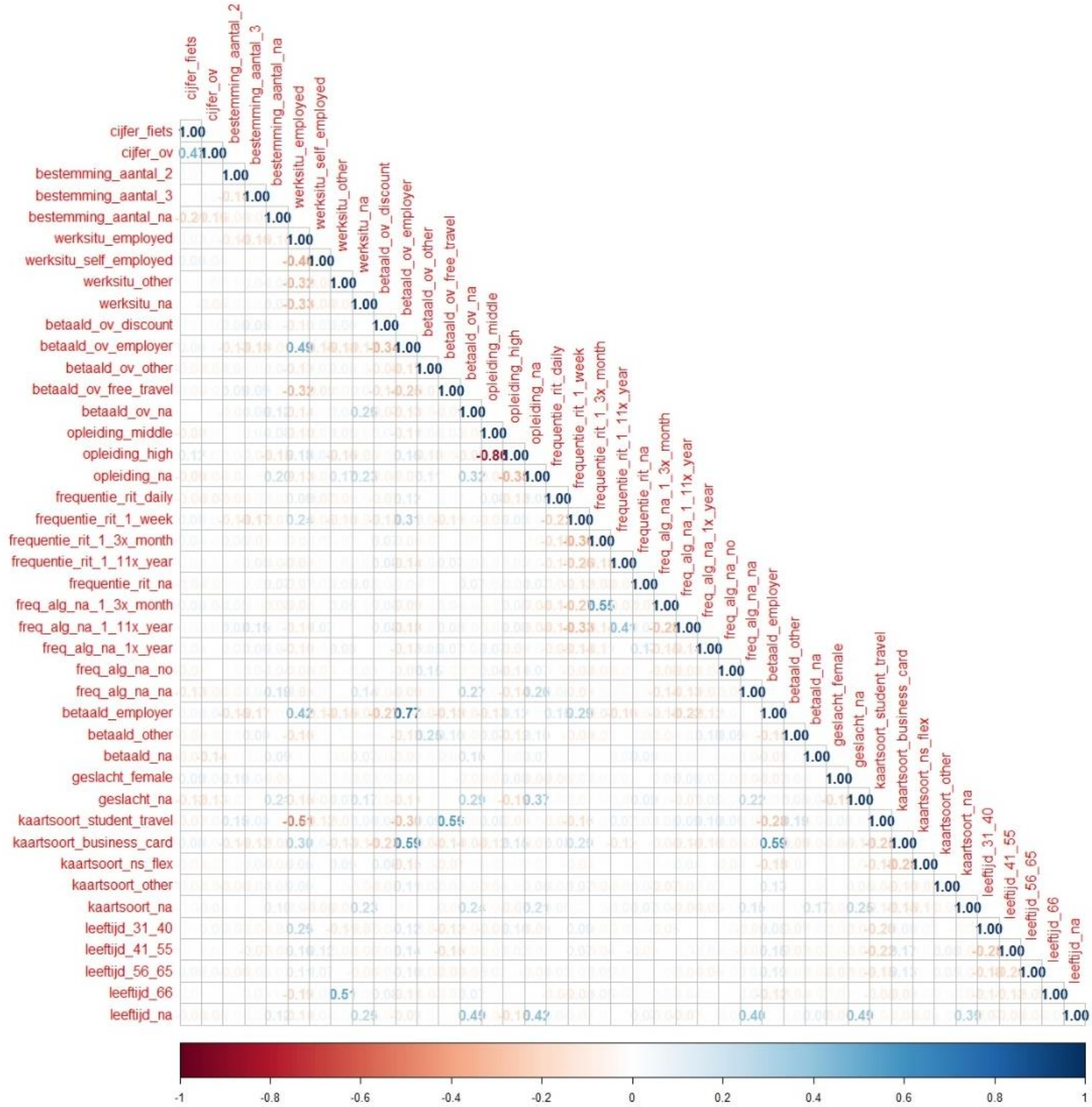


Figure 71: Correlation Matrix input variables purpose model

8.6.2 Sensitivity

		Sensitivity model	
Category	Variable	Education	Leisure
Gender (ref = male)	Female	0.973*** (0.202)	0.189 (0.184)
Age (ref = < 30)	31-40	-0.787*** (0.293)	-0.109 (0.270)
	41-55	-1.165*** (0.304)	-0.540** (0.273)
	56-65	-1.184*** (0.443)	0.451 (0.323)
	66+	-1.731** (0.673)	0.886* (0.454)
Education level (ref = low)	Middle	0.267 (0.490)	0.455 (0.489)
	High	-0.091 (0.456)	-0.360 (0.459)
Working situation (ref = school going)	Employed	-2.372*** (0.294)	-1.172*** (0.295)
	Self-employed	-2.789*** (0.493)	-2.040*** (0.383)
OV-chipcard type (ref = on balance)	Student	0.881** (0.401)	0.503 (0.419)
	Business card	-0.306 (0.334)	0.726** (0.316)
	NS flex	0.072 (0.267)	-0.070 (0.233)
Frequency OV-fiets (ref = >1x/week)	1-3x/month	1.106*** (0.296)	0.173 (0.279)
	1-11x/year	0.403 (0.351)	0.537* (0.291)
	< 1x/year	0.266 (0.468)	0.565 (0.404)
Pays for PT (ref = full tariff)	Discount	-0.695** (0.291)	-0.264 (0.230)
	Employer	0.424 (0.345)	-1.348*** (0.362)
	Free travel	-0.100 (0.398)	-0.213 (0.394)
Rating (continuous)	OV-fiets	0.212** (0.089)	0.030 (0.074)
	Public transport	-0.009 (0.081)	0.170** (0.079)
Frequency trip (ref = first time)	Daily	-0.830* (0.445)	-16.778*** (0.00000)
	>1x/week	-0.330 (0.325)	-2.356*** (0.302)
	1-3x/month	-0.614* (0.337)	-1.004*** (0.292)
	1-11x/year	-0.144 (0.337)	-0.679** (0.282)
Number of Destinations (ref = 1)	2	0.085 (0.256)	0.386 (0.236)
	3	0.457 (0.324)	0.853*** (0.295)
Pays for OV-fiets (ref = self)	Employer	-1.527*** (0.311)	-2.615*** (0.334)
Constant		-0.438 (0.840)	1.009 (0.775)
AIC		**1913.65**	

8.7 Appendix G – Destination model

8.7.1 Correlation matrix

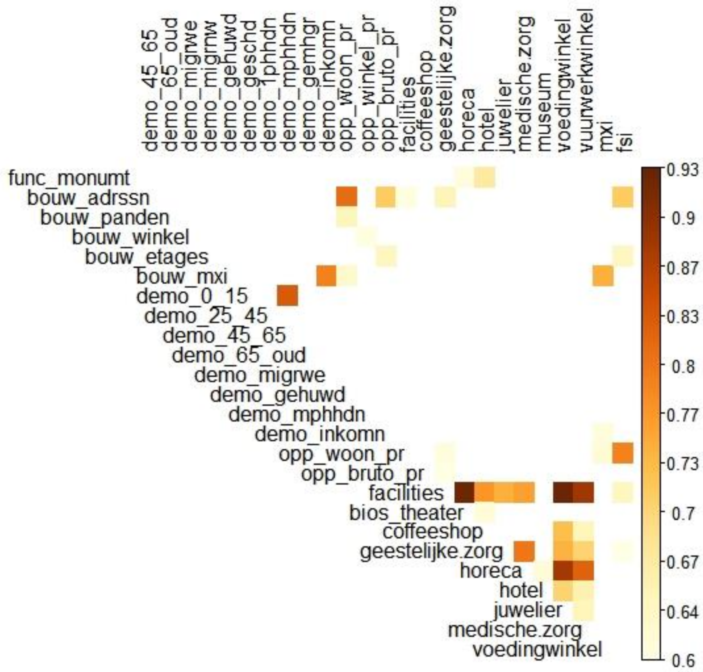


Figure 72: Correlation Matrix Destination model (only variables with r > 0.6 displayed)

8.7.2 Resampling

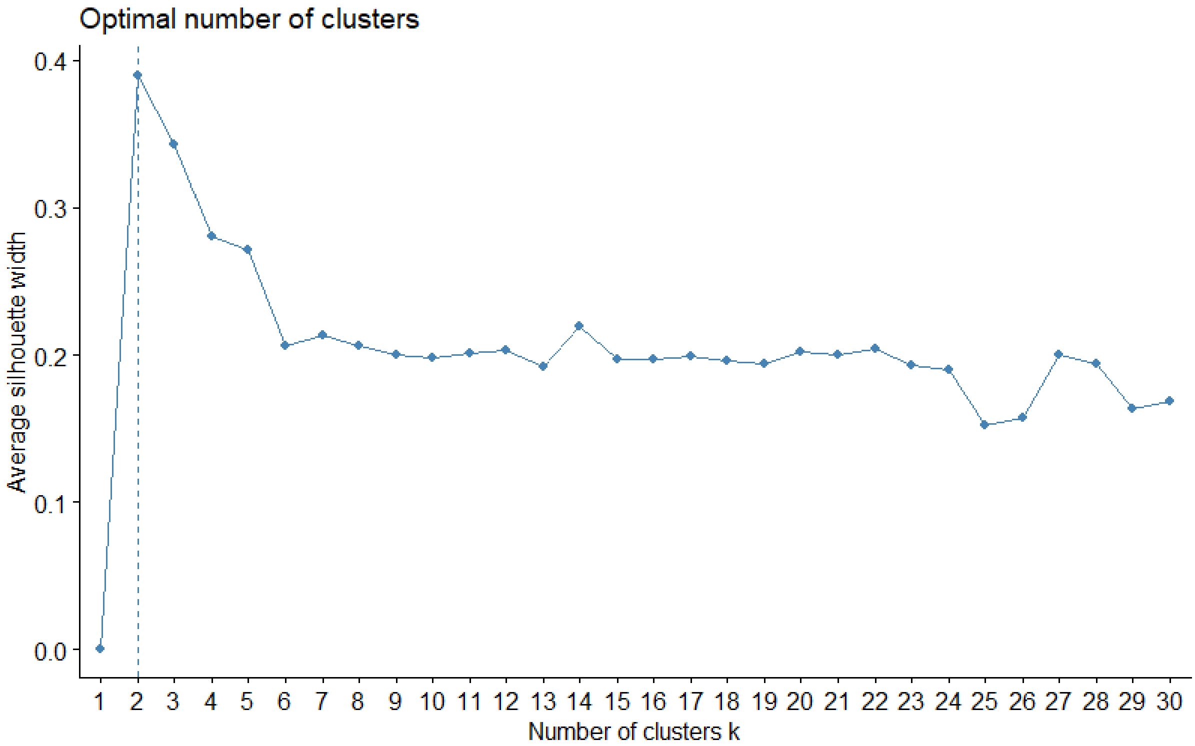


Figure 73: Output of silhouette method for optimal amount of clusters undersampling

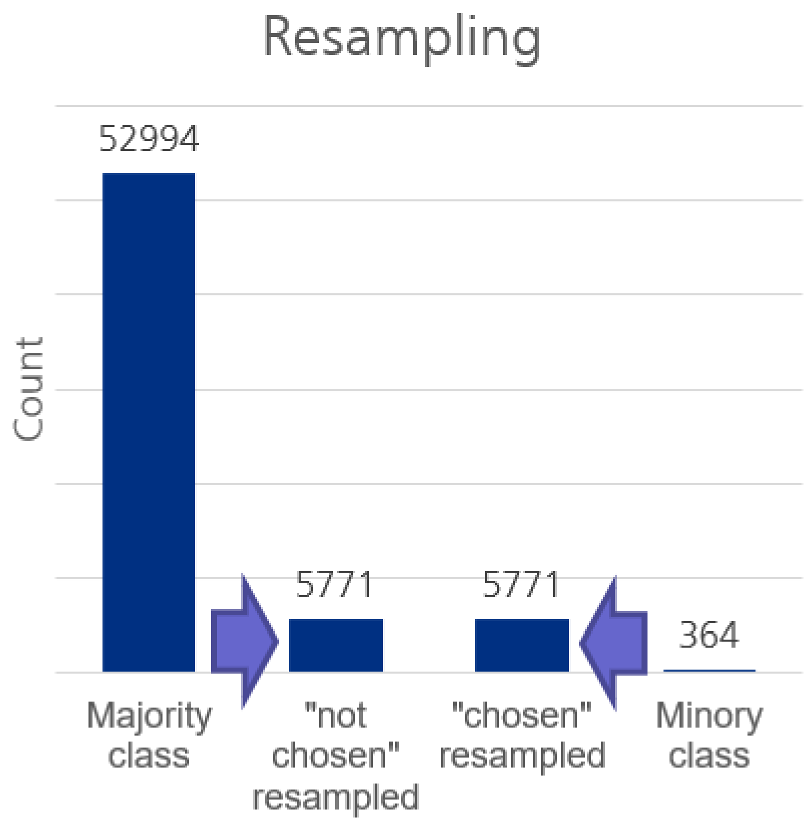


Figure 74: Visualisation of distribution with resampling

8.7.3 Both direction parameter search

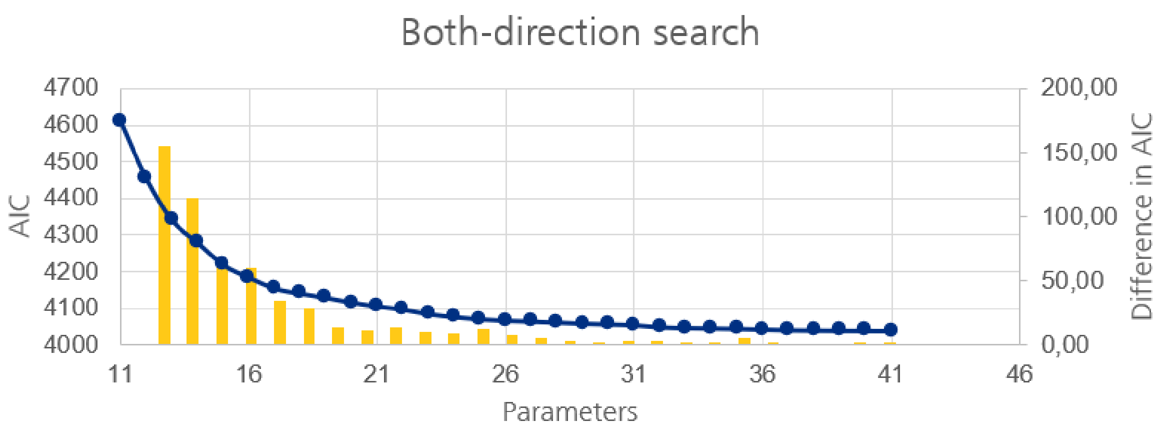


Figure 75: Both direction parameter search

8.7.4 Destination model estimation lowest AIC

Table 30: Destination binary logit model with all variables: AIC = 4161.9

Variable	Beta	Standard Deviation	Lower Margin	Upper Margin	t-value	p-value	Sig.
(Intercept)	-2,321 E+00	0,184	-2,357	-2,285	-126,438	0,000	***
No. Monuments	3,979 E-03	0,050	-0,006	0,014	0,802	0,427	
No. Buildings	1,335 E-02	0,002	0,013	0,014	54,427	0,000	***
Population	-9,235 E-04	0,000	-0,001	-0,001	-26,688	0,000	***
No. Non self-reliance	1,402 E-04	0,001	0,000	0,000	1,604	0,116	
No. Meeting buildings	1,354 E-01	0,089	0,118	0,153	15,249	0,000	***
No. Shopping buildings	6,428 E-02	0,029	0,059	0,070	21,958	0,000	***
No. Accommodations	-4,185 E-01	0,229	-0,463	-0,374	-18,305	0,000	***
No. Industy buildings	2,201 E-02	0,040	0,014	0,030	5,526	0,000	***
No. Health building	-8,992 E-02	0,085	-0,107	-0,073	-10,581	0,000	***
No. Offices	1,267 E-01	0,056	0,116	0,138	22,686	0,000	***
No. Education buildings	2,710 E-02	0,217	-0,015	0,070	1,248	0,219	
No. KvK registrations	1,435 E-02	0,013	0,012	0,017	11,028	0,000	***
Building value	-2,942 E-08	0,000	0,000	0,000	-6,435	0,000	***
No. buildings >25m	-2,642 E-02	0,227	-0,071	0,018	-1,165	0,251	
No. Layers	1,843 E-01	0,027	0,179	0,190	69,181	0,000	***
Perc. of inhabitants age 15-25	1,970 E-02	0,006	0,018	0,021	30,474	0,000	***
Perc. of inhabitants age 25-45	-1,290 E-02	0,004	-0,014	-0,012	-33,121	0,000	***
Perc. of inhabitants age 45-65	-5,492 E-03	0,007	-0,007	-0,004	-7,662	0,000	***
Perc. of inhabitants age 65+	-6,831 E-03	0,005	-0,008	-0,006	-12,692	0,000	***
No. Households with children	-3,541 E-02	0,005	-0,036	-0,034	-70,961	0,000	***
Household size	1,686 E-01	0,084	0,152	0,185	20,072	0,000	***
No. AOW receivers	-2,201 E-03	0,001	-0,002	-0,002	-21,927	0,000	***
Income	7,550 E-02	0,016	0,072	0,079	47,831	0,000	***
Area Living function	-1,146 E-04	0,000	0,000	0,000	-48,266	0,000	***
Area Meeting function	2,239 E-04	0,000	0,000	0,000	14,652	0,000	***
Area Healthcare function	1,415 E-04	0,000	0,000	0,000	15,068	0,000	***
Area Industry function	1,581 E-04	0,000	0,000	0,000	29,773	0,000	***
Area Office function	5,420 E-05	0,000	0,000	0,000	17,029	0,000	***
Area Accommodation function	1,952 E-04	0,000	0,000	0,000	12,793	0,000	***
Area Education function	2,195 E-04	0,000	0,000	0,000	16,892	0,000	***
Area Sports function	2,826 E-04	0,000	0,000	0,000	14,581	0,000	***
Area Shopping function	-1,879 E-04	0,000	0,000	0,000	-15,999	0,000	***
Area Other	-3,210 E-04	0,000	0,000	0,000	-23,721	0,000	***
Area Annexes	1,986 E-05	0,000	0,000	0,000	1,842	0,072	*
No. Bus stops (in 400m)	1,618 E-01	0,026	0,157	0,167	62,632	0,000	***
No. Pharmacies (in 400m)	-8,837 E-02	0,040	-0,096	-0,081	-22,051	0,000	***
No. Car dealers (in 400m)	-6,898 E-02	0,013	-0,071	-0,066	-54,302	0,000	***
No. Lower schools (in 400m)	9,717 E-02	0,046	0,088	0,106	21,106	0,000	***
No. Libraries (in 400m)	1,669 E-01	0,161	0,135	0,198	10,398	0,000	***
No. Cinema's/theaters (in 400m)	-1,417 E-01	0,156	-0,172	-0,111	-9,086	0,000	***
No. Community centres (in 400m)	-8,591 E-03	0,056	-0,020	0,002	-1,521	0,136	
No. Coffee shops (in 400m)	-2,642 E-02	0,093	-0,045	-0,008	-2,828	0,007	***
No. Congress centres (in 400m)	1,365 E-01	0,244	0,089	0,184	5,591	0,000	***
No. Veterinaries (in 400m)	2,223 E-01	0,081	0,207	0,238	27,528	0,000	***
No. Prisons (in 400m)	1,365 E+00	0,681	1,231	1,498	20,045	0,000	***
No. Hospitalities (in 400m)	1,213 E-03	0,004	0,000	0,002	2,739	0,009	***
No. Hotels (in 400m)	-2,177 E-02	0,022	-0,026	-0,017	-9,922	0,000	***
No. Jewellery stores (in 400m)	2,387 E-01	0,060	0,227	0,251	39,644	0,000	***
No. Petting zoos (in 400m)	8,979 E-02	0,149	0,061	0,119	6,035	0,000	***
No. Medical care (in 400m)	7,864 E-03	0,010	0,006	0,010	7,756	0,000	***
No. Retirement homes (in 400m)	2,983 E-02	0,040	0,022	0,038	7,427	0,000	***
No. Middle schools (in 400m)	4,589 E-02	0,058	0,035	0,057	7,964	0,000	***
No. Winter sports stores (in 400m)	-1,067 E-01	0,179	-0,142	-0,072	-5,965	0,000	***
No. Swimming pools (in 400m)	9,354 E-02	0,169	0,060	0,127	5,519	0,000	***
Distance (transformed)	1,436 E+04	512,819	14255,303	14456,324	279,939	0,000	***
Mixed land-use index	-8,000 E-01	0,204	-0,840	-0,760	-39,178	0,000	***

8.7.5 Destination model spatial results

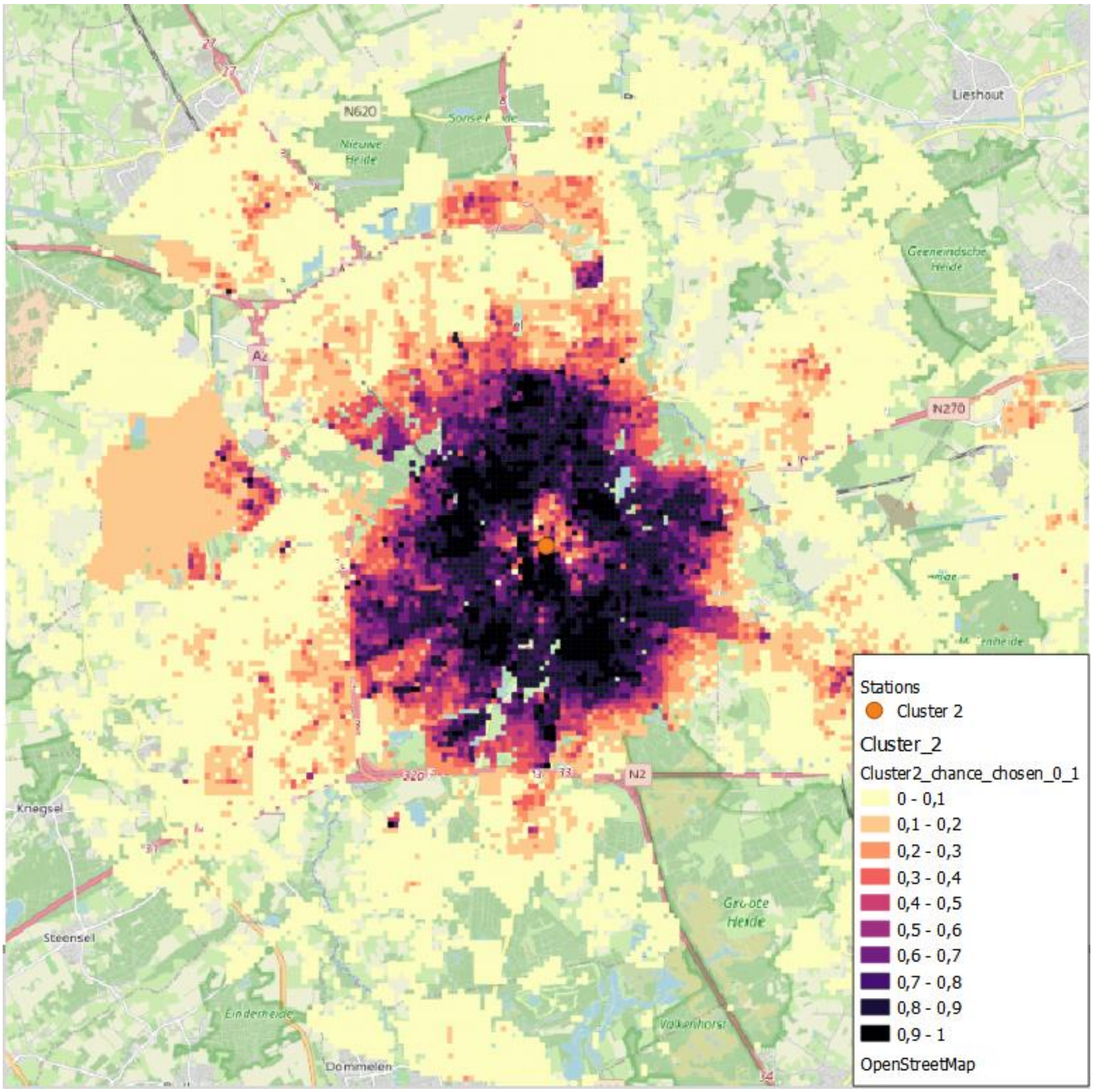


Figure 76: Eindhoven, probability of grid cell being chosen or not

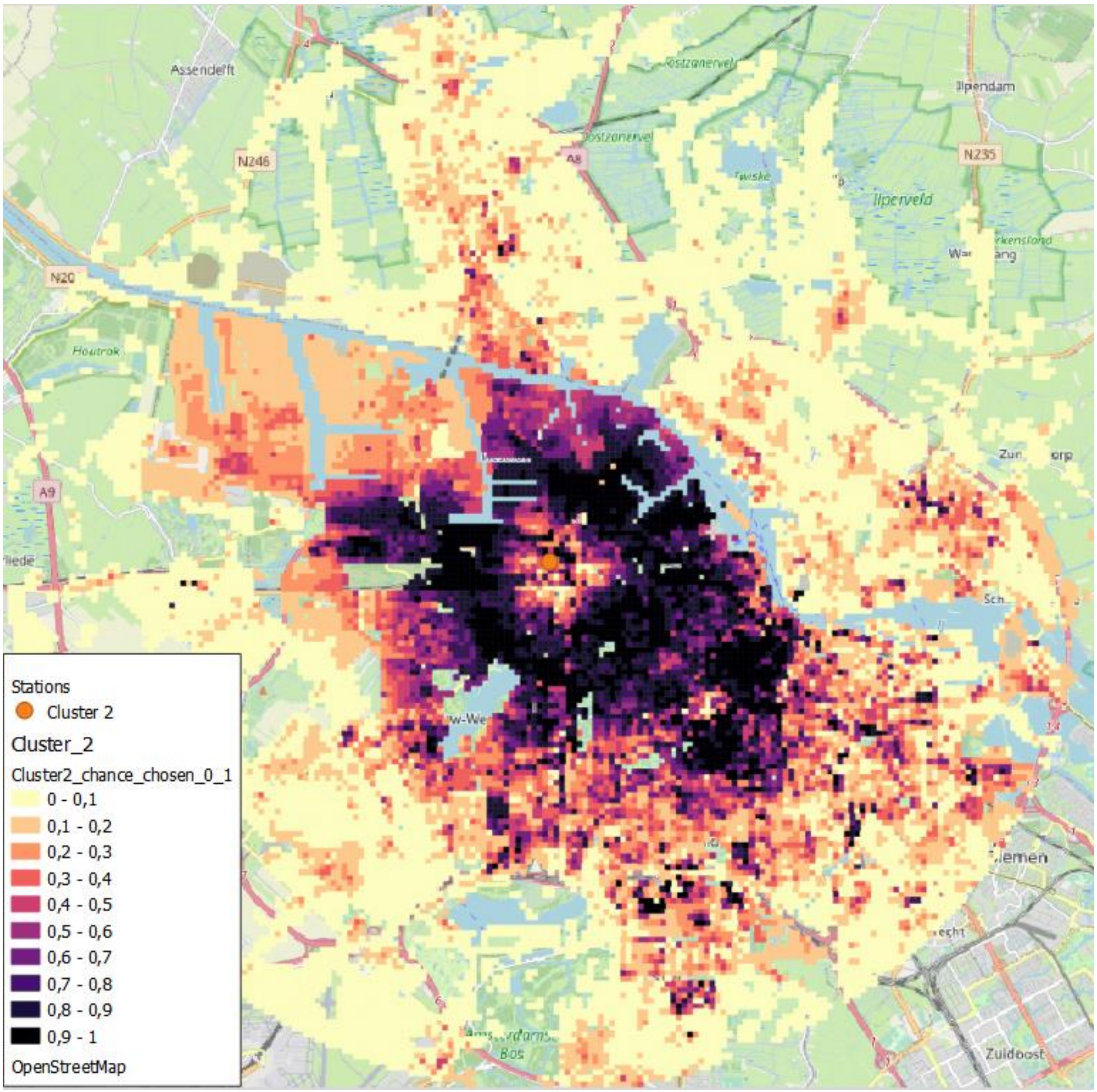


Figure 77: Amsterdam, probability of grid cell being chosen or not

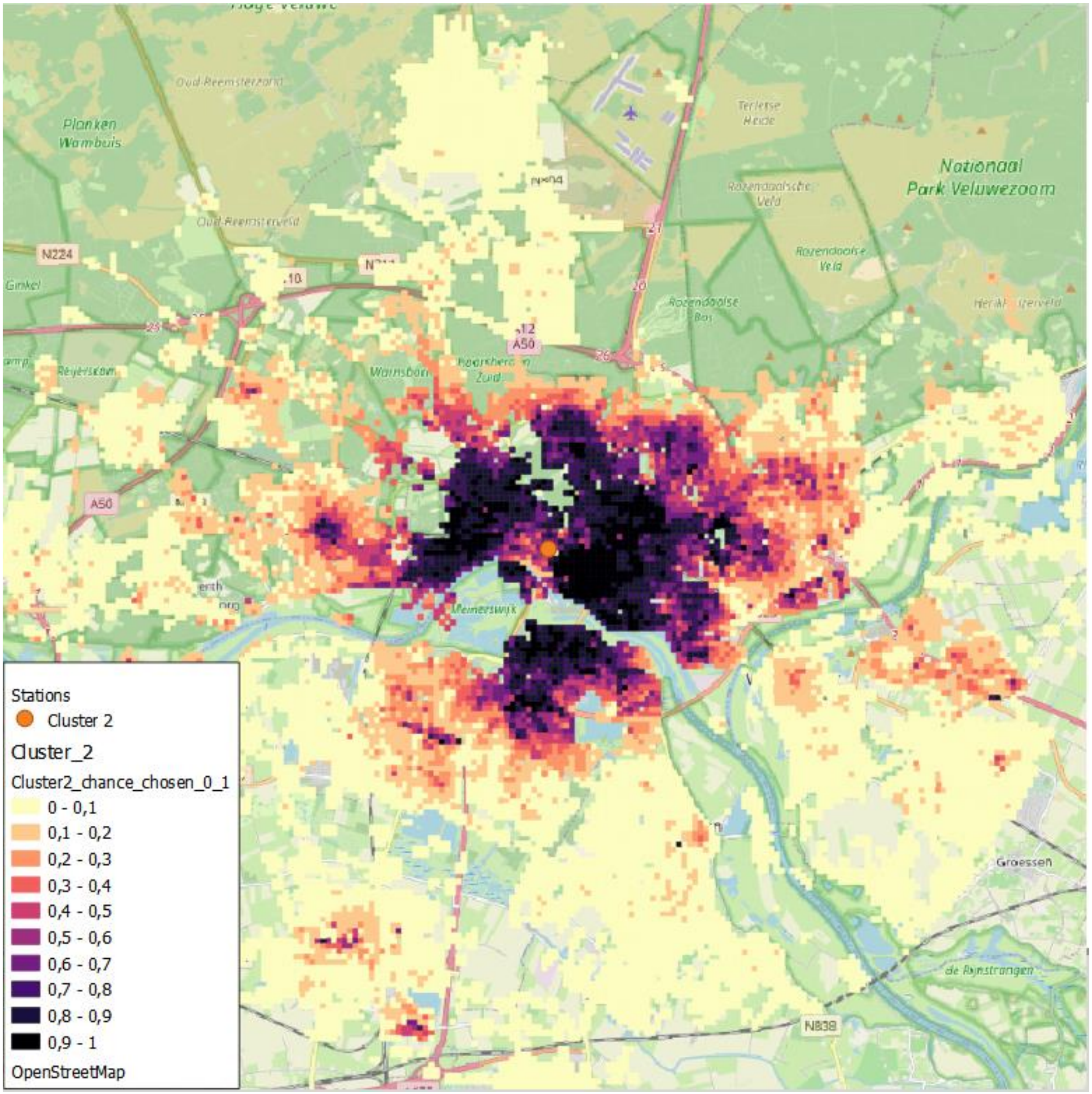


Figure 78: Arnhem, probability of grid cell being chosen or not

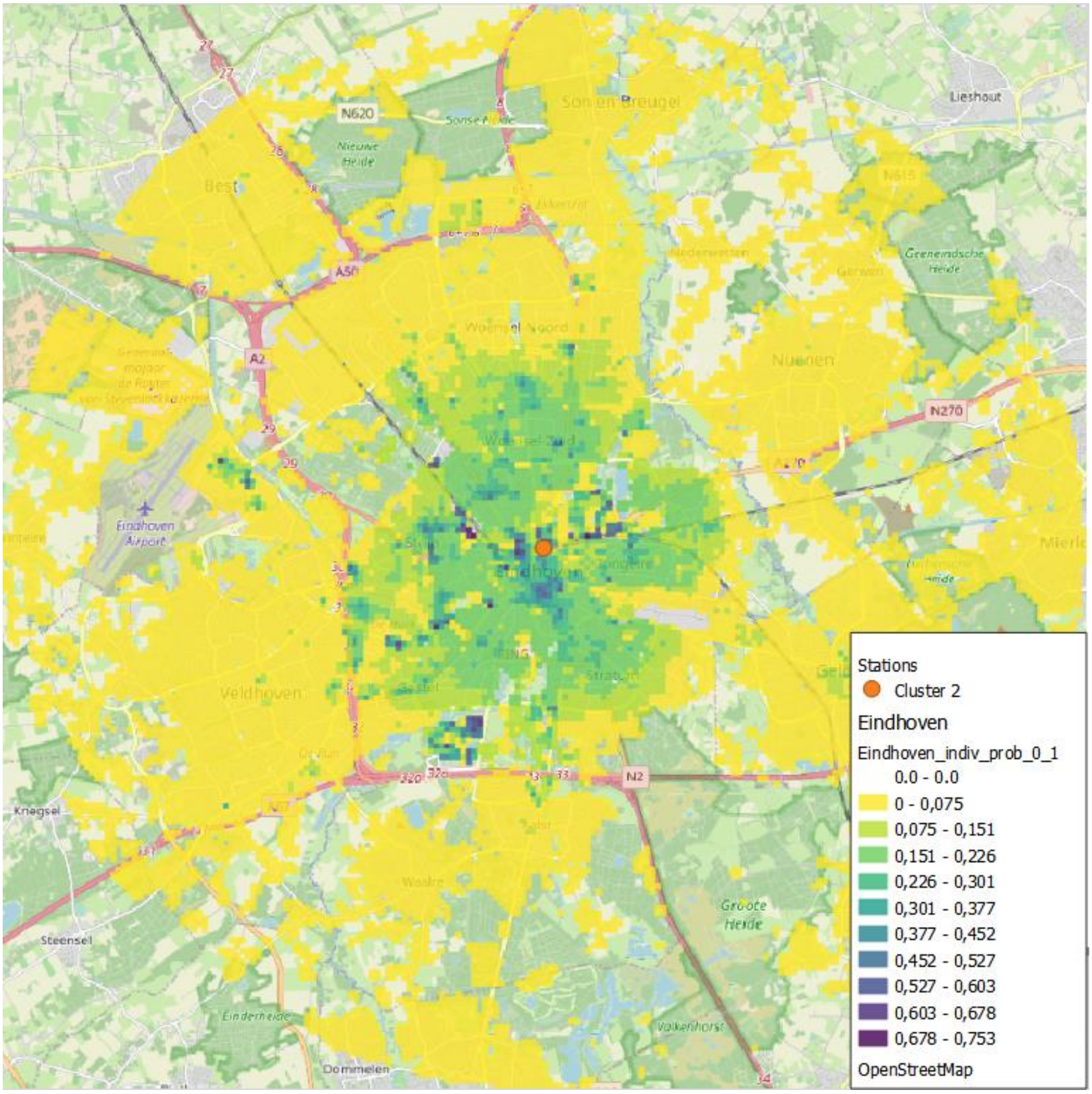


Figure 79: Eindhoven, probability of grid cell being chosen by the young student

8.7.6 Destination model validation

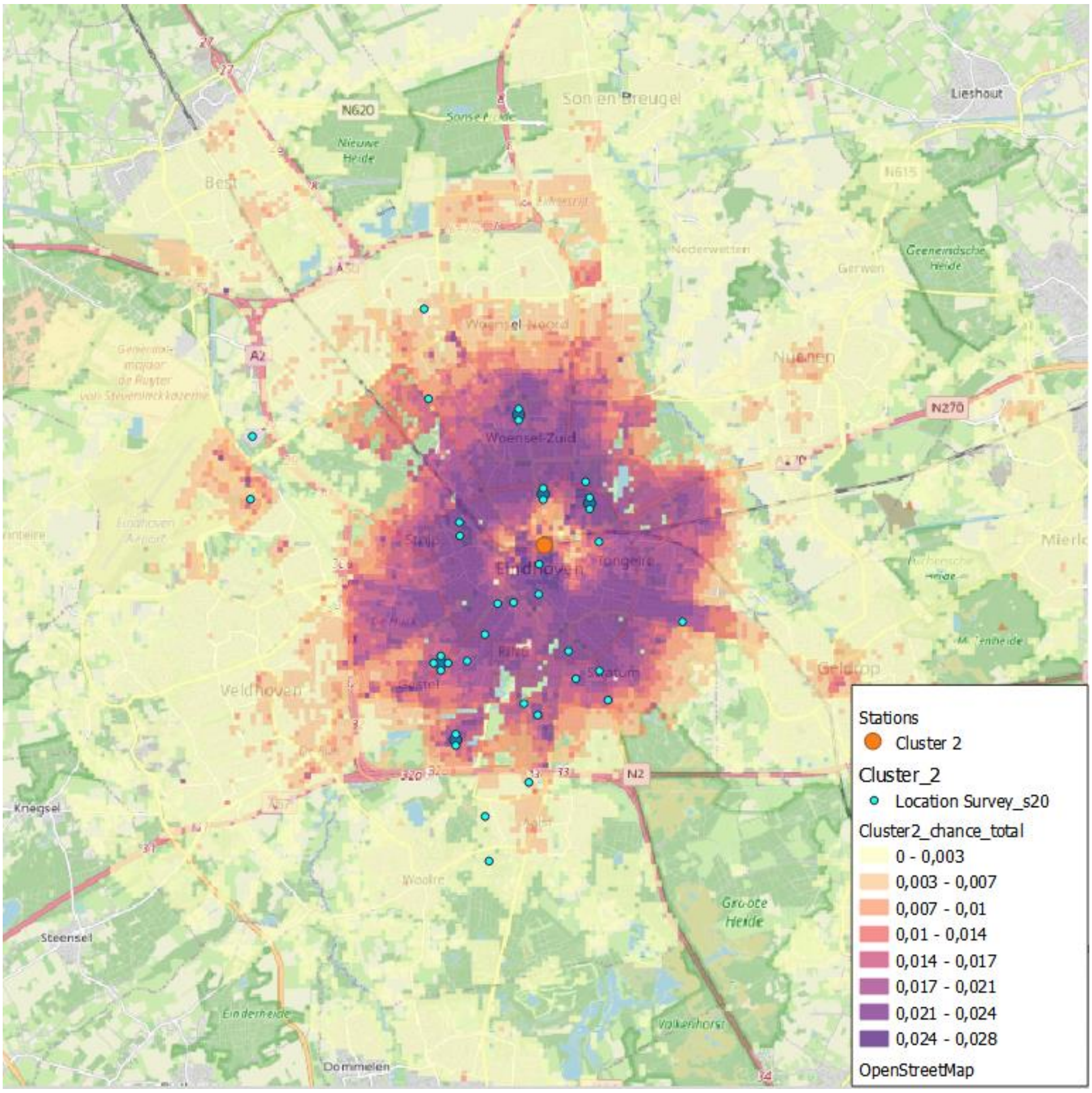


Figure 80: Eindhoven, test set and spatial probability distribution destination

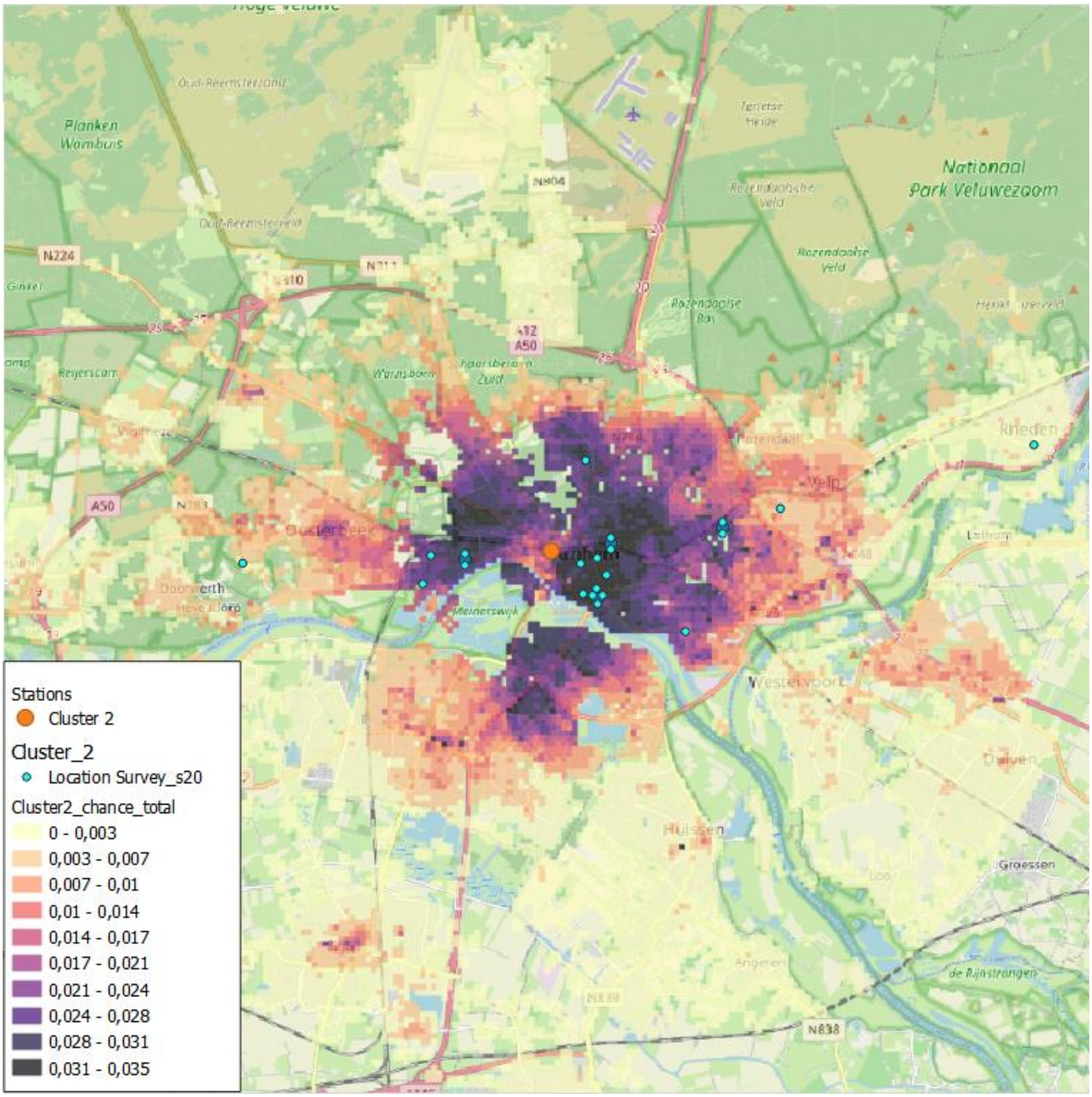


Figure 81: Arnhem, test set and spatial probability distribution destination

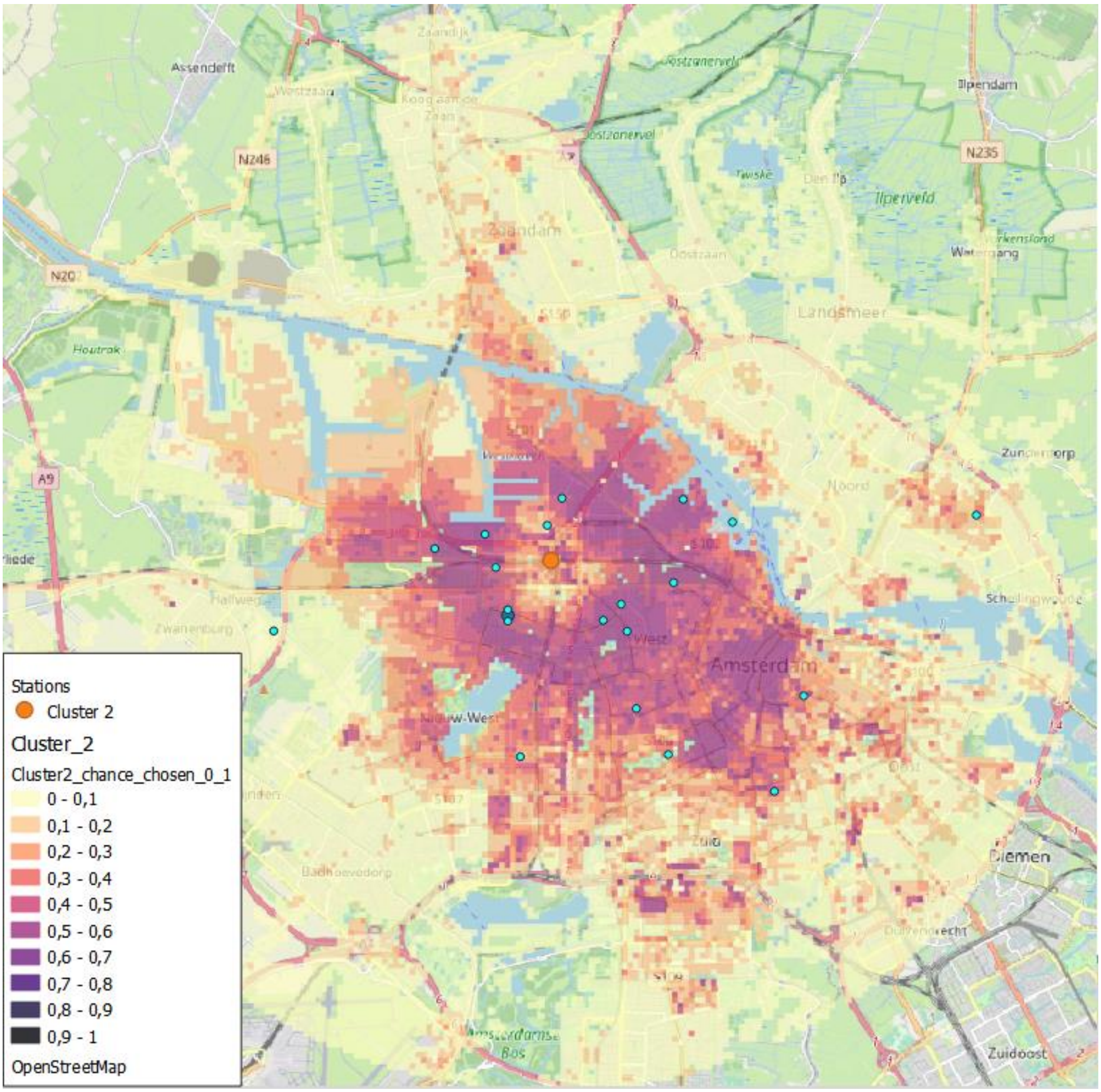


Figure 82: Amsterdam, test set and spatial probability distribution destination

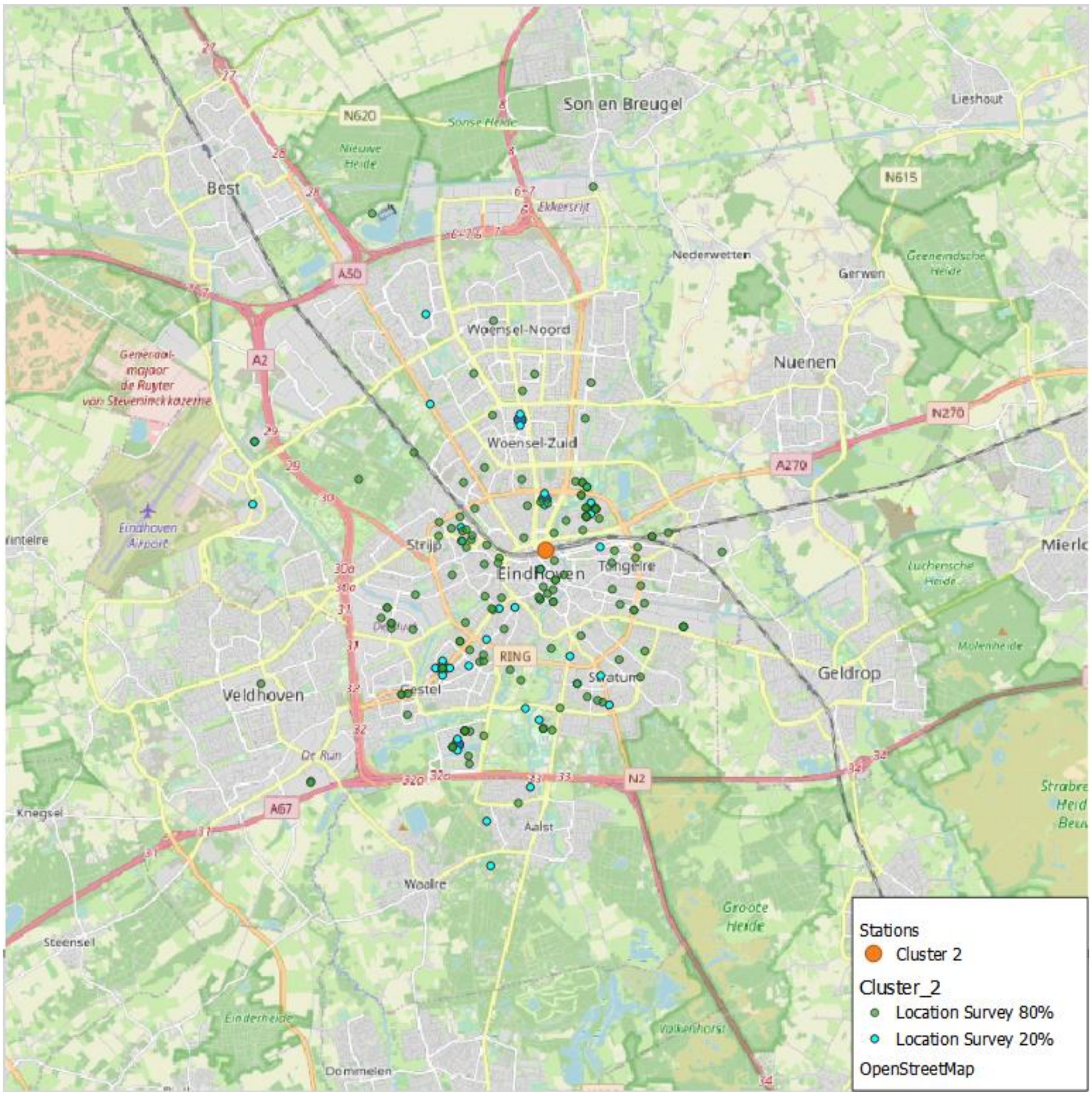


Figure 83: Eindhoven, train and test respondents set of the models

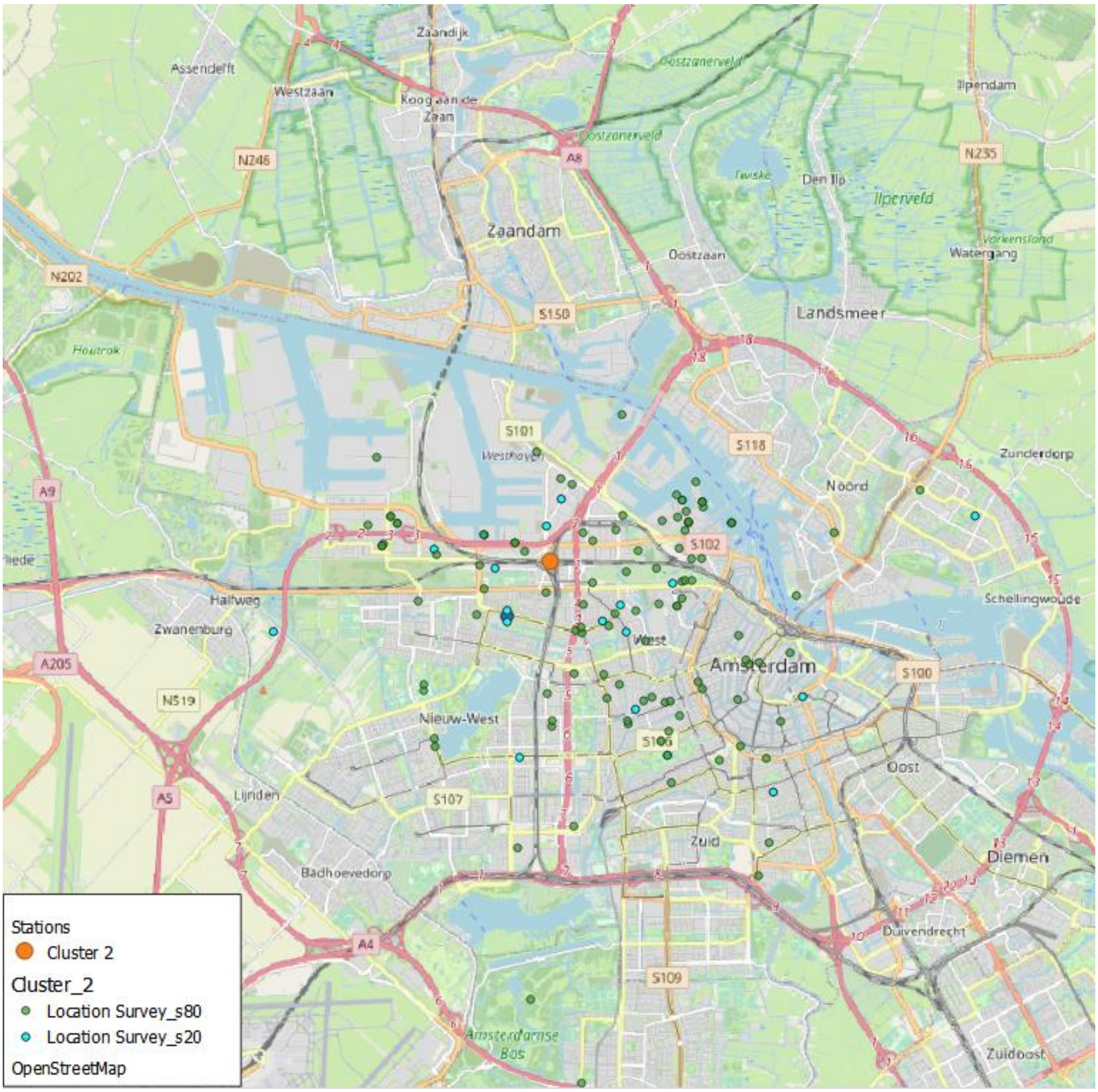


Figure 84: Amsterdam, train and testset of the models

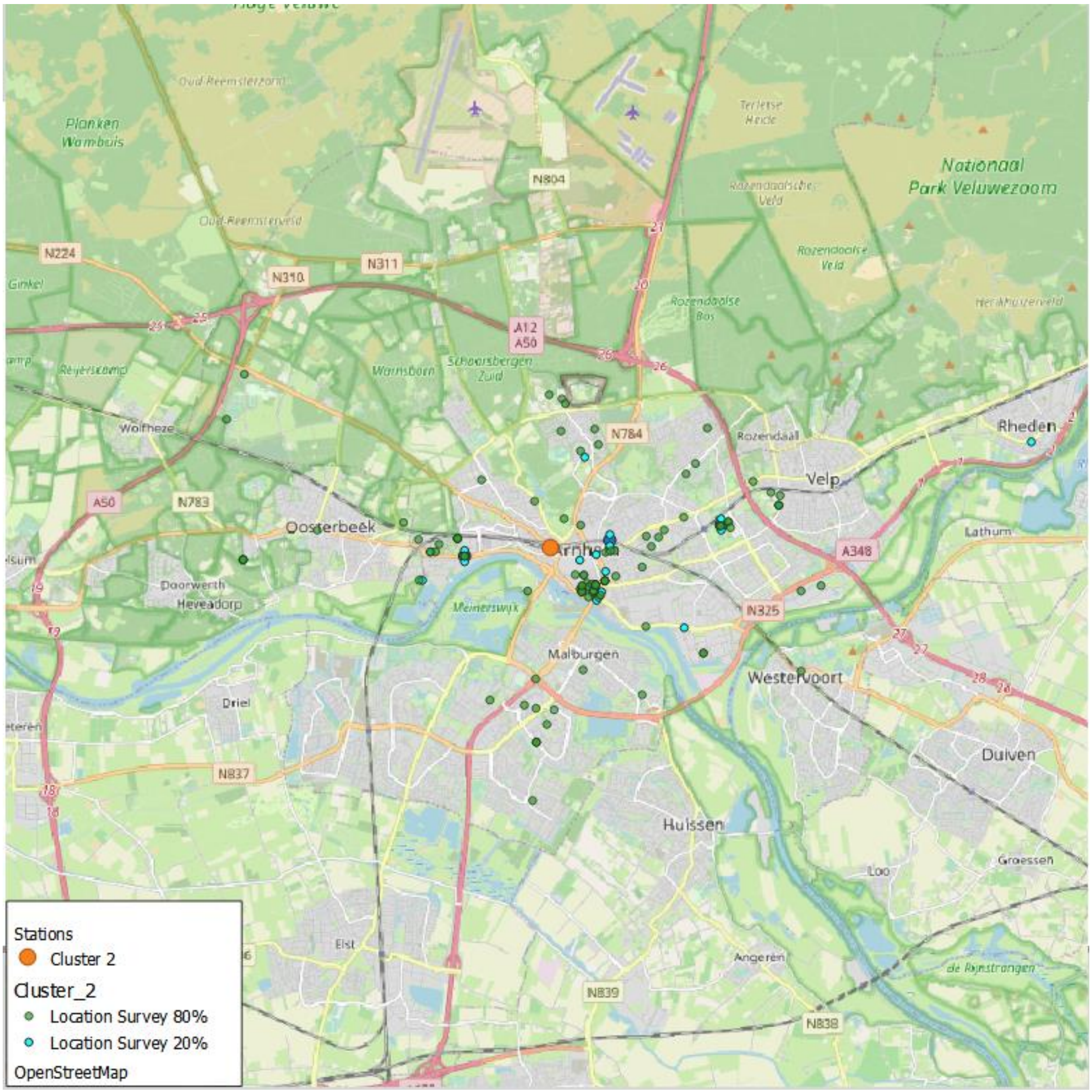


Figure 85: Arnhem, train and test set of the models

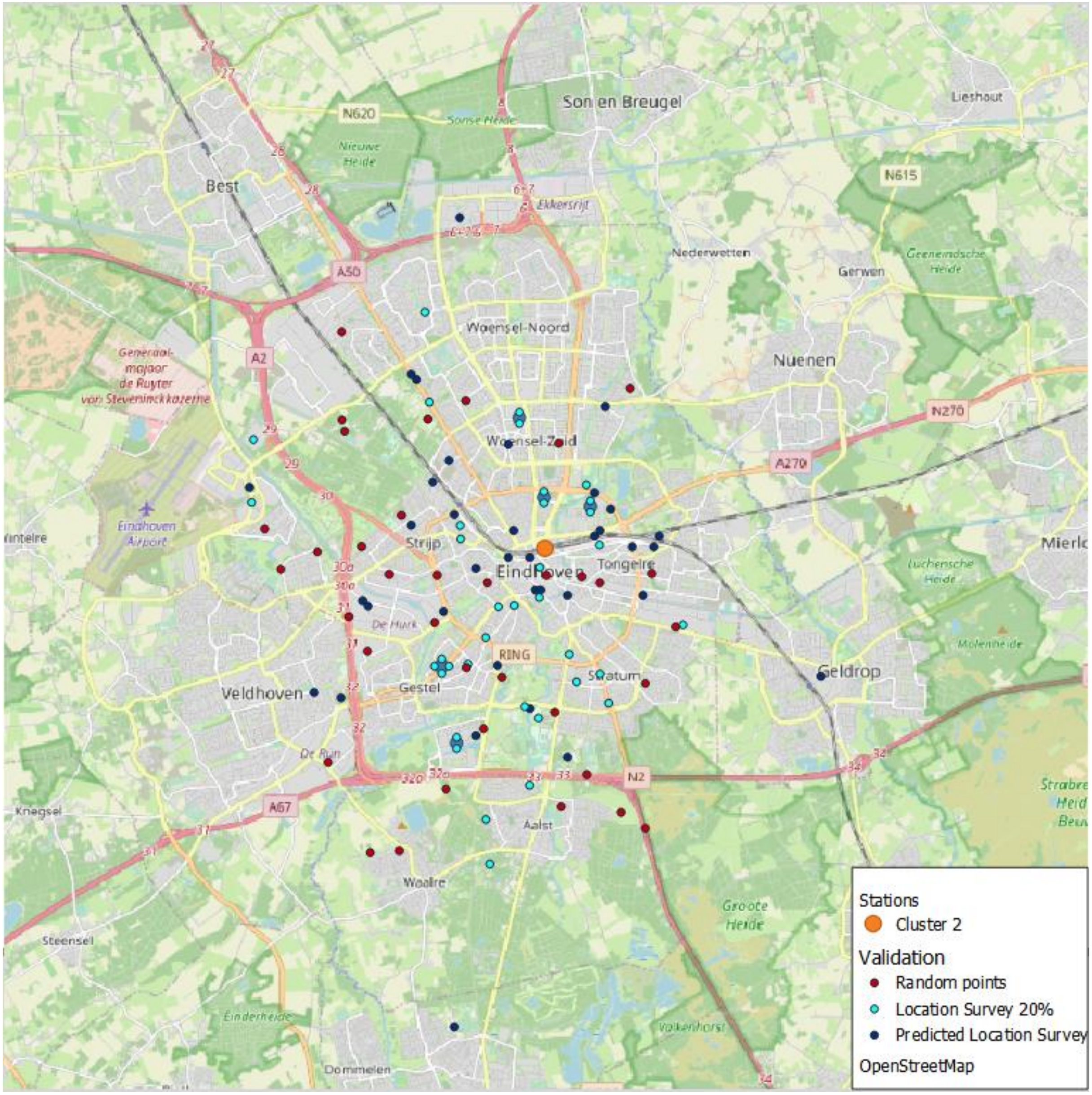


Figure 86: Eindhoven: location survey, location survey predicted by the model and random points

8.7.7 Sensitivity input data

To decide if a reweighed data set would indicate other results. However, for the destination model, the input data are destinations, which are single points in the spatial dimension. These points are not easy to reweigh, while those are binary and can exist in many places. To overcome that, a few analyses have been conducted between the destination area type and the user characteristics. If those are significant, the user characteristics might have a significant effect on destinations. If those were not significant, the user characteristics do not have a significant effect on the destinations, which means that the destinations do not have to be created with reweight data.

The most dominant effect for a destination to be chosen is the activity, in other words, the function type area (Recker & Kostyniuk, 1978). Therefore the destination correlation can be checked on the function-area with the user characteristics. The significance was tested with a single-factor ANOVA analysis (Schuberth, 2019). Both person-based characteristics (e.g. age) and trip-based characteristics (e.g. trip frequency) did not show significant effects on the function type of the area of the destinations. For both analyses, the F-critic was larger than the F-score, which meant that the means were equal, thus keeping the null hypothesis. This means that the means are equal, and there was no significant difference between the groups. The p-value for the analysis defined a significance of 85%.

8.7.8 Bandwidth parameters

variable	mean	10 pctile	20 pctile	30 pctile	40 pctile	50 pctile	60 pctile	70 pctile	80 pctile	90 pctile
Constant	-2,18E+00	-2,30E+00	-2,27E+00	-2,24E+00	-2,23E+00	-2,19E+00	-2,15E+00	-2,13E+00	-2,10E+00	-2,06E+00
Area Living function	-6,45E-05	-8,27E-05	-7,66E-05	-7,28E-05	-7,09E-05	-6,53E-05	-6,06E-05	-5,72E-05	-5,02E-05	-4,34E-05
Area Meeting function	1,88E-04	-5,97E-06	7,22E-05	9,97E-05	1,59E-04	1,92E-04	2,18E-04	2,43E-04	2,98E-04	3,80E-04
Area Healthcare function	1,57E-04	7,65E-05	9,99E-05	1,16E-04	1,31E-04	1,49E-04	1,63E-04	1,78E-04	2,19E-04	2,38E-04
Area Industry function	1,43E-04	9,37E-05	1,10E-04	1,25E-04	1,32E-04	1,41E-04	1,55E-04	1,65E-04	1,73E-04	1,89E-04
Area Office function	4,87E-05	1,06E-05	2,21E-05	3,51E-05	4,15E-05	5,16E-05	5,55E-05	6,31E-05	7,30E-05	8,42E-05
Area Accomodation function	6,92E-05	-9,62E-05	-2,63E-05	1,30E-05	5,62E-05	8,28E-05	1,15E-04	1,35E-04	1,70E-04	2,22E-04
Area Education function	1,94E-04	7,95E-05	1,06E-04	1,32E-04	1,49E-04	1,64E-04	2,04E-04	2,37E-04	2,89E-04	3,46E-04
Area Sports function	3,11E-04	5,55E-05	1,15E-04	1,56E-04	2,12E-04	2,77E-04	3,70E-04	4,34E-04	5,01E-04	5,78E-04
Area Shopping function	-9,87E-05	-2,13E-04	-1,87E-04	-1,63E-04	-1,33E-04	-1,10E-04	-7,75E-05	-5,45E-05	-2,60E-05	4,04E-05
Area Other	-2,94E-04	-4,48E-04	-3,97E-04	-3,54E-04	-3,40E-04	-2,97E-04	-2,64E-04	-2,36E-04	-1,88E-04	-1,15E-04
Area Extra buildings	1,70E-04	6,93E-05	1,05E-04	1,20E-04	1,37E-04	1,63E-04	1,86E-04	2,04E-04	2,36E-04	2,79E-04
Distance	1,35E+04	1,31E+04	1,32E+04	1,33E+04	1,34E+04	1,35E+04	1,36E+04	1,37E+04	1,39E+04	1,41E+04
No. Households with childeren	-3,44E-02	-3,83E-02	-3,67E-02	-3,59E-02	-3,51E-02	-3,45E-02	-3,35E-02	-3,26E-02	-3,20E-02	-3,11E-02
No. KvK registrations	2,05E-02	3,86E-03	7,82E-03	1,24E-02	1,76E-02	2,19E-02	2,39E-02	2,67E-02	2,93E-02	3,71E-02
No. Bus stops (in 400m)	1,82E-01	1,55E-01	1,65E-01	1,71E-01	1,76E-01	1,80E-01	1,89E-01	1,93E-01	2,00E-01	2,08E-01
No. Juwelery stores (in 400m)	2,28E-01	1,86E-01	1,99E-01	2,11E-01	2,18E-01	2,27E-01	2,41E-01	2,48E-01	2,55E-01	2,63E-01
No. Layers	1,31E-01	1,02E-01	1,12E-01	1,22E-01	1,25E-01	1,32E-01	1,39E-01	1,46E-01	1,52E-01	1,59E-01
No. Meeting buildings	1,34E-01	4,83E-02	7,10E-02	8,35E-02	9,99E-02	1,26E-01	1,46E-01	1,77E-01	1,97E-01	2,25E-01

Figure 87: Percentiles of parameters for 1000 estimations destination model

8.8 Appendix H – Combination model

8.8.1 Combination model simulated trips and PC4 share

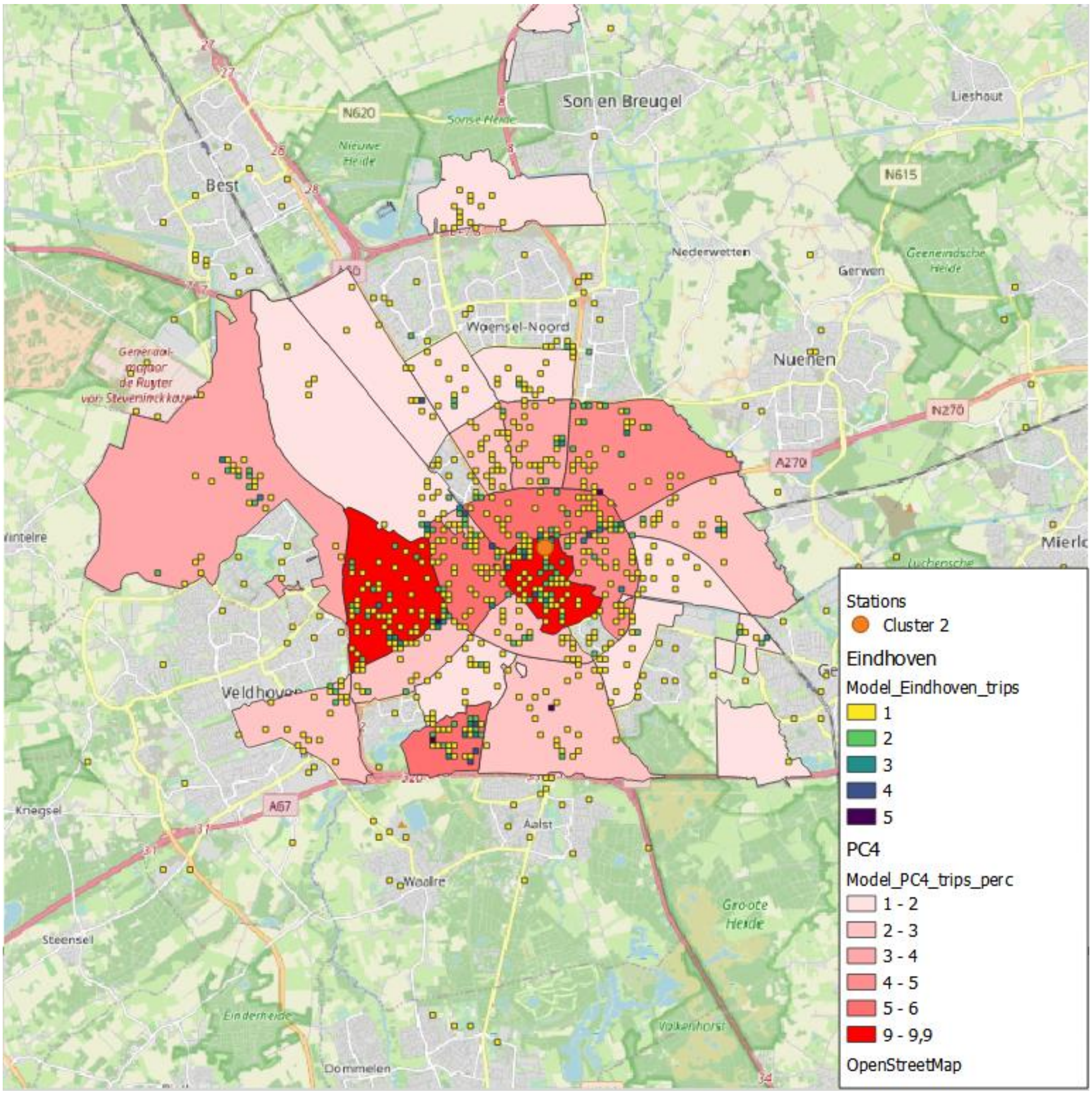


Figure 88: Eindhoven, simulated trips and PC4 share of trips

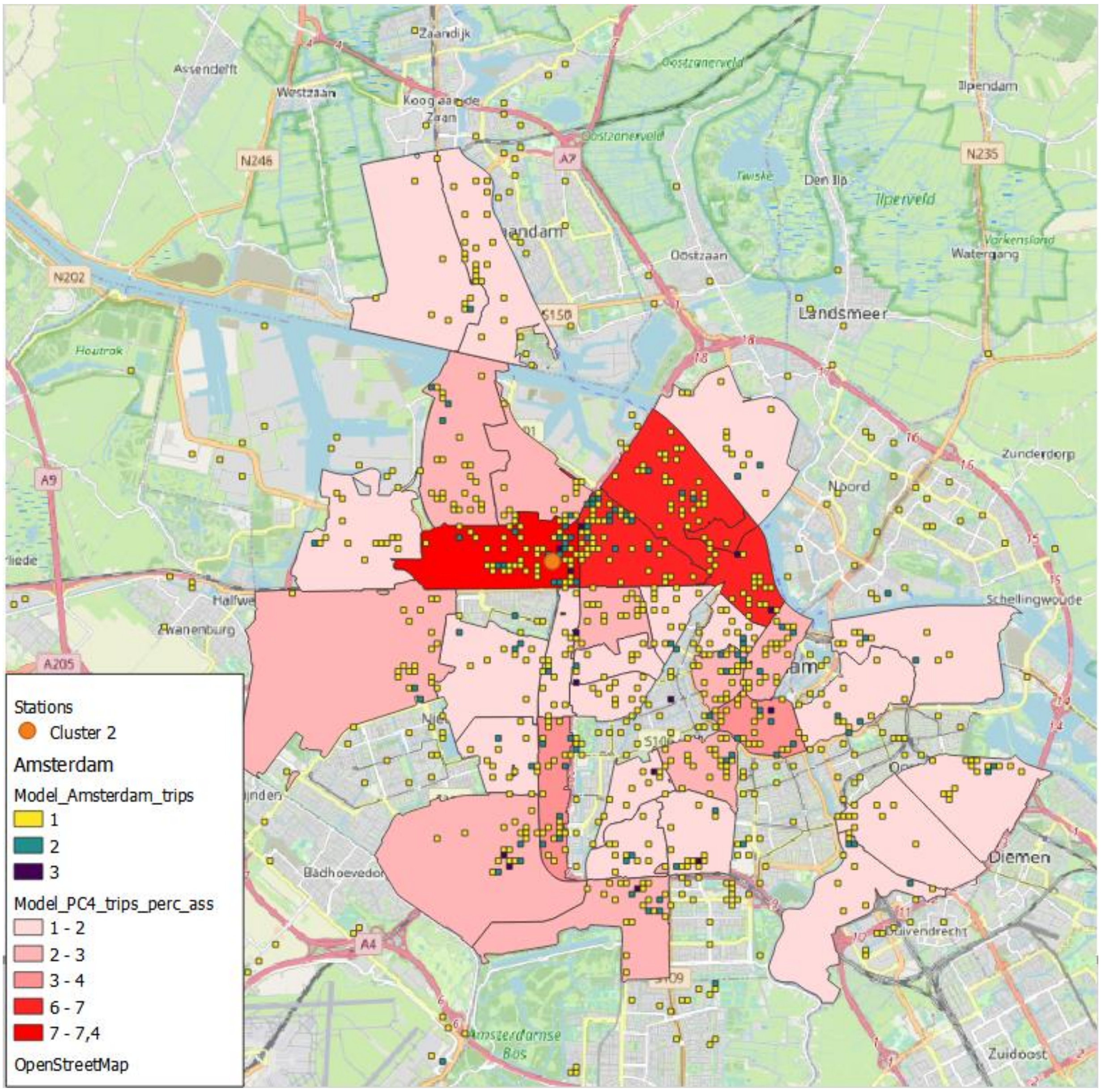


Figure 89: Amsterdam, simulated trips and PC4 share of trips

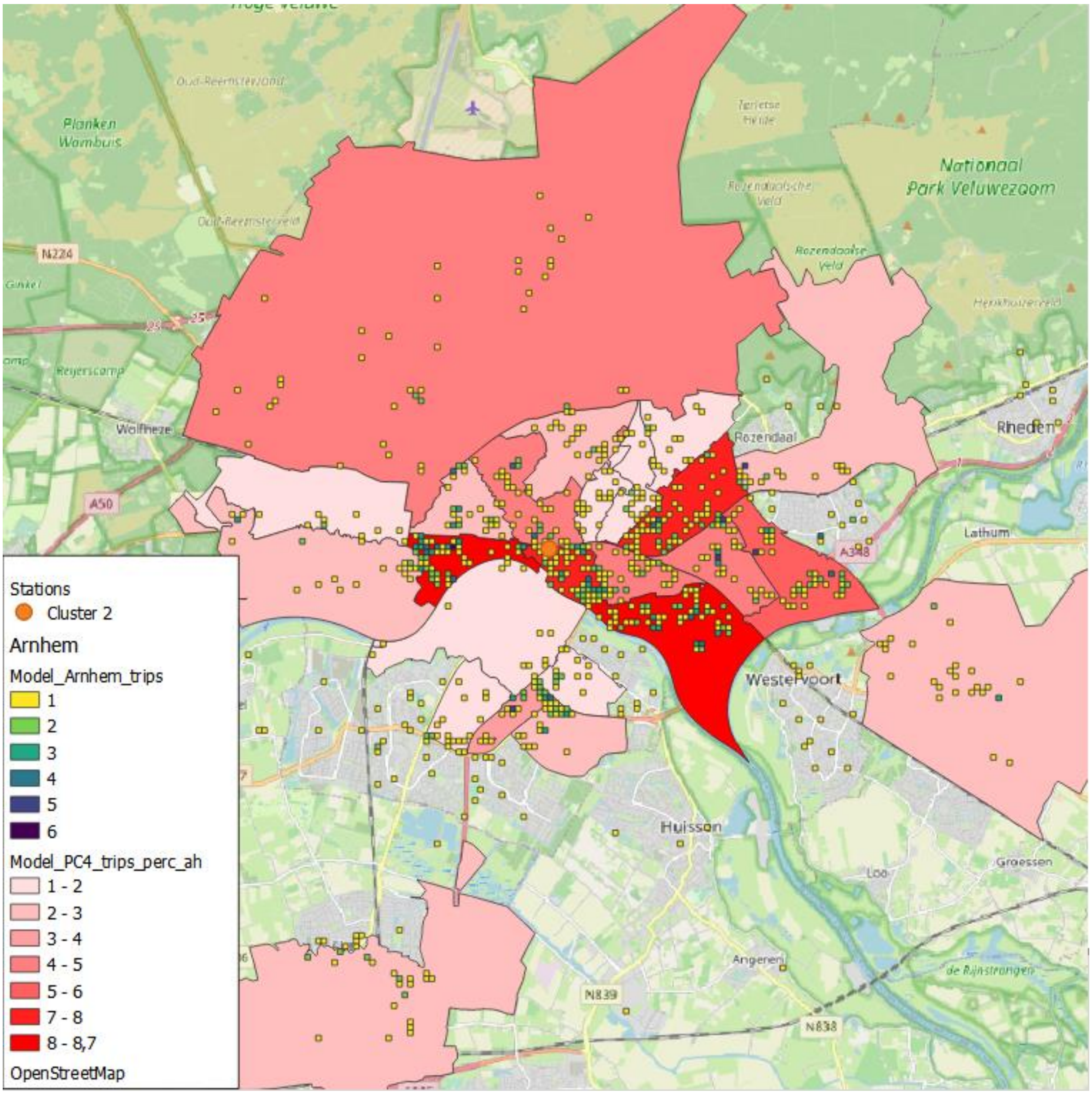


Figure 90: Arnhem, simulated trips and PC4 share of trips

8.8.2 Combination model simulated trips per purpose

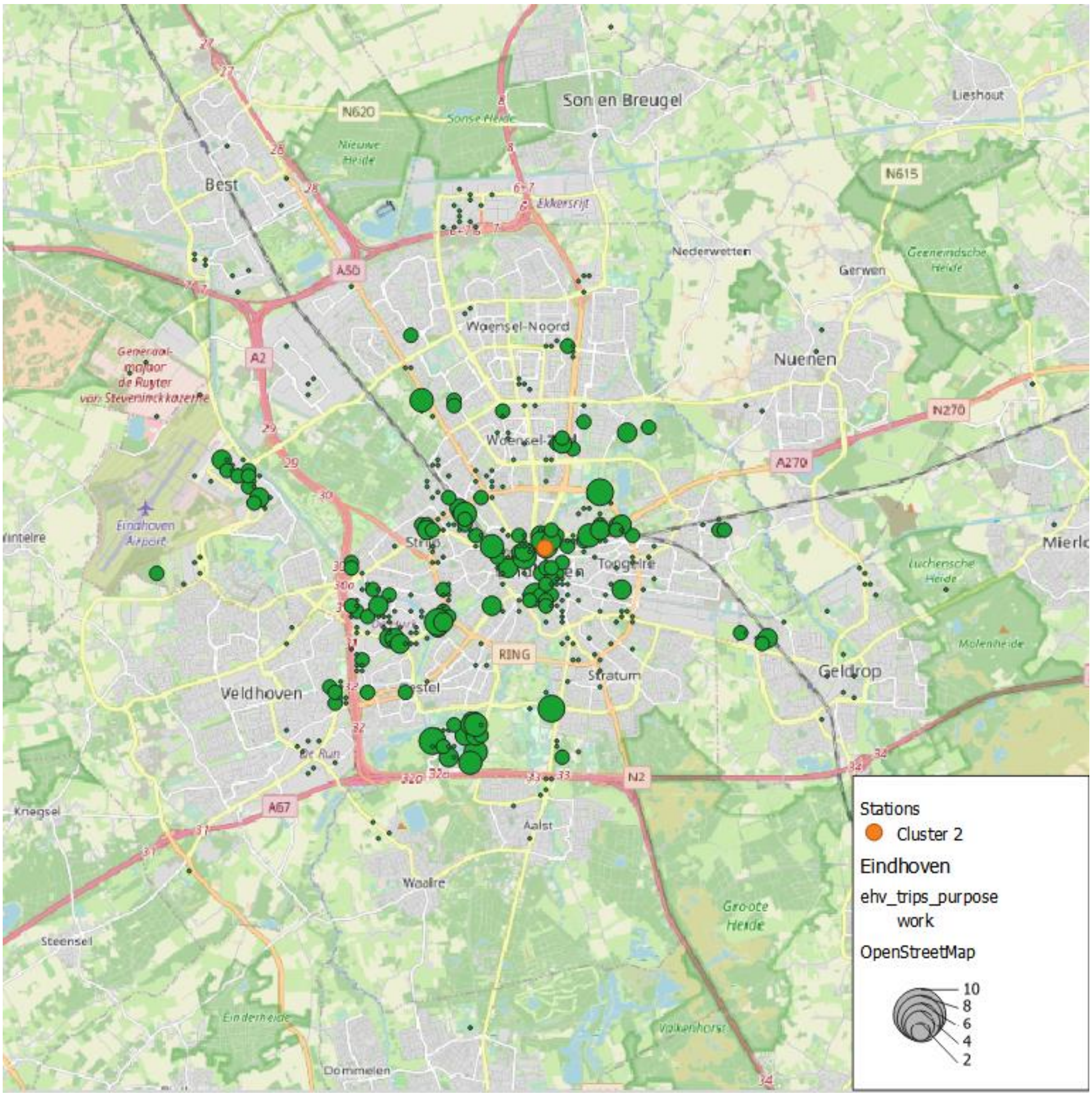


Figure 91: Eindhoven, simulated trips with work purpose (65% of trips)

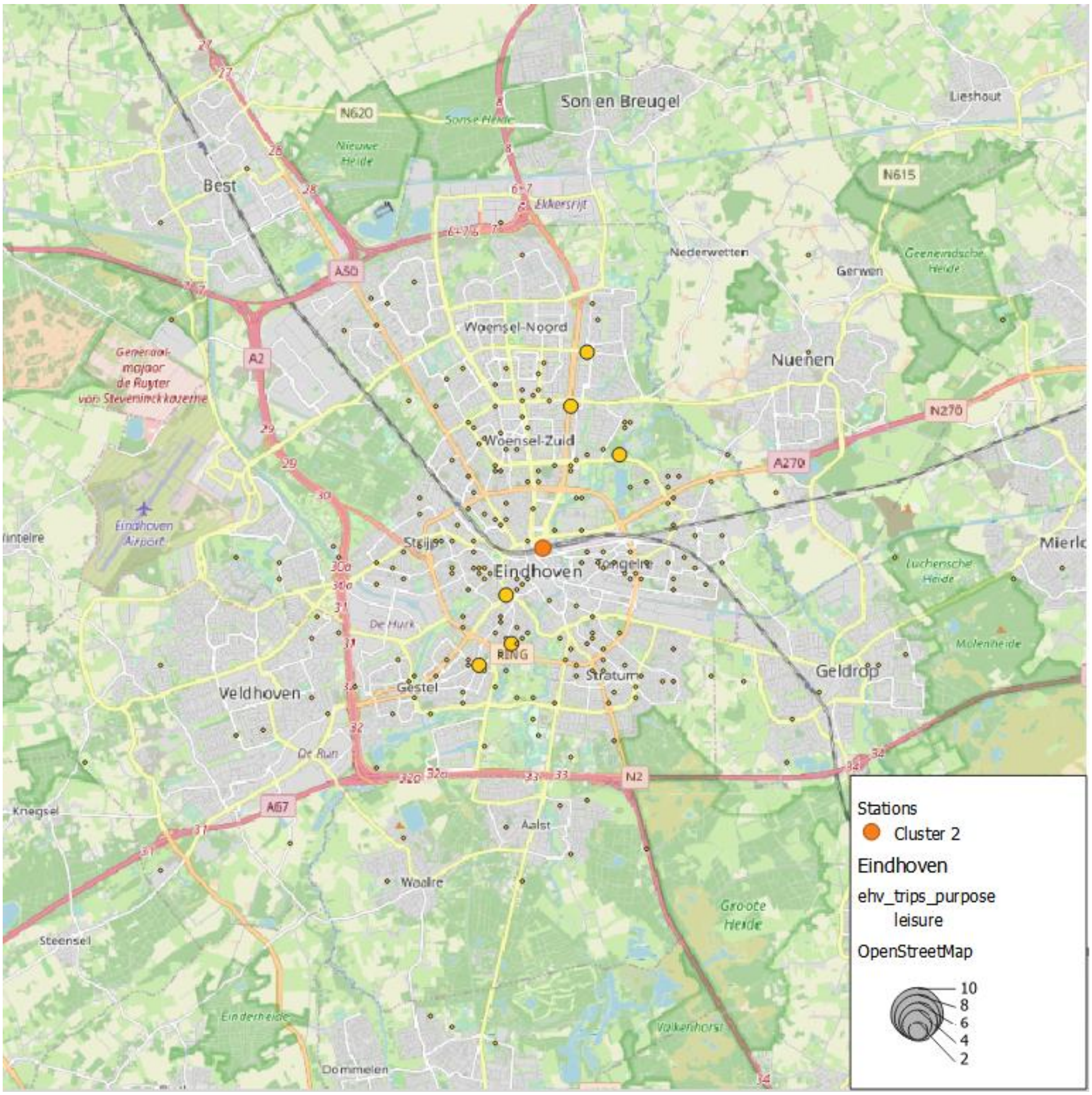


Figure 92: Eindhoven, simulated trips leisure purpose (25% of trips)

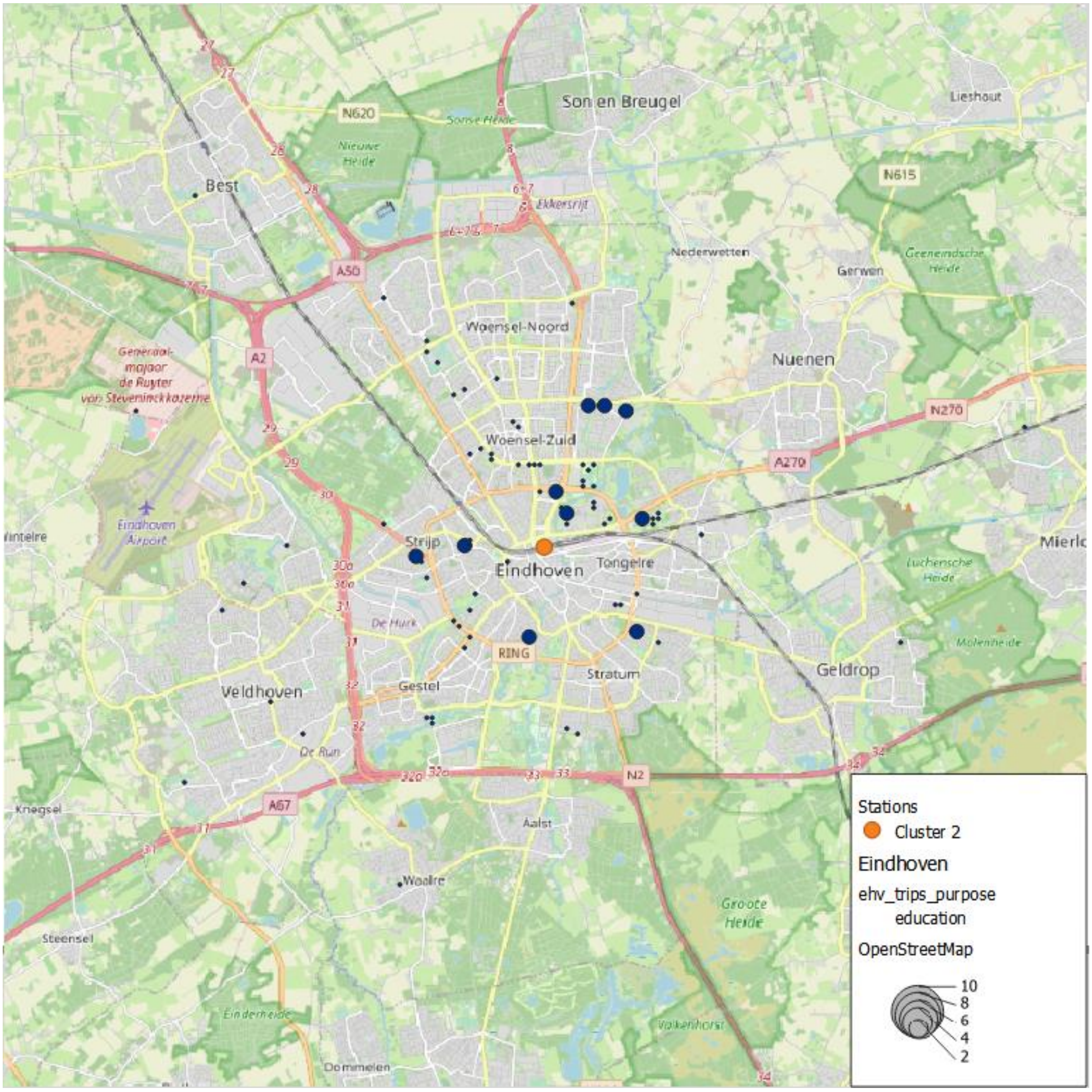


Figure 93: Eindhoven, simulated trips education purpose (9% of trips)

8.8.3 Combination model simulated trips per card type

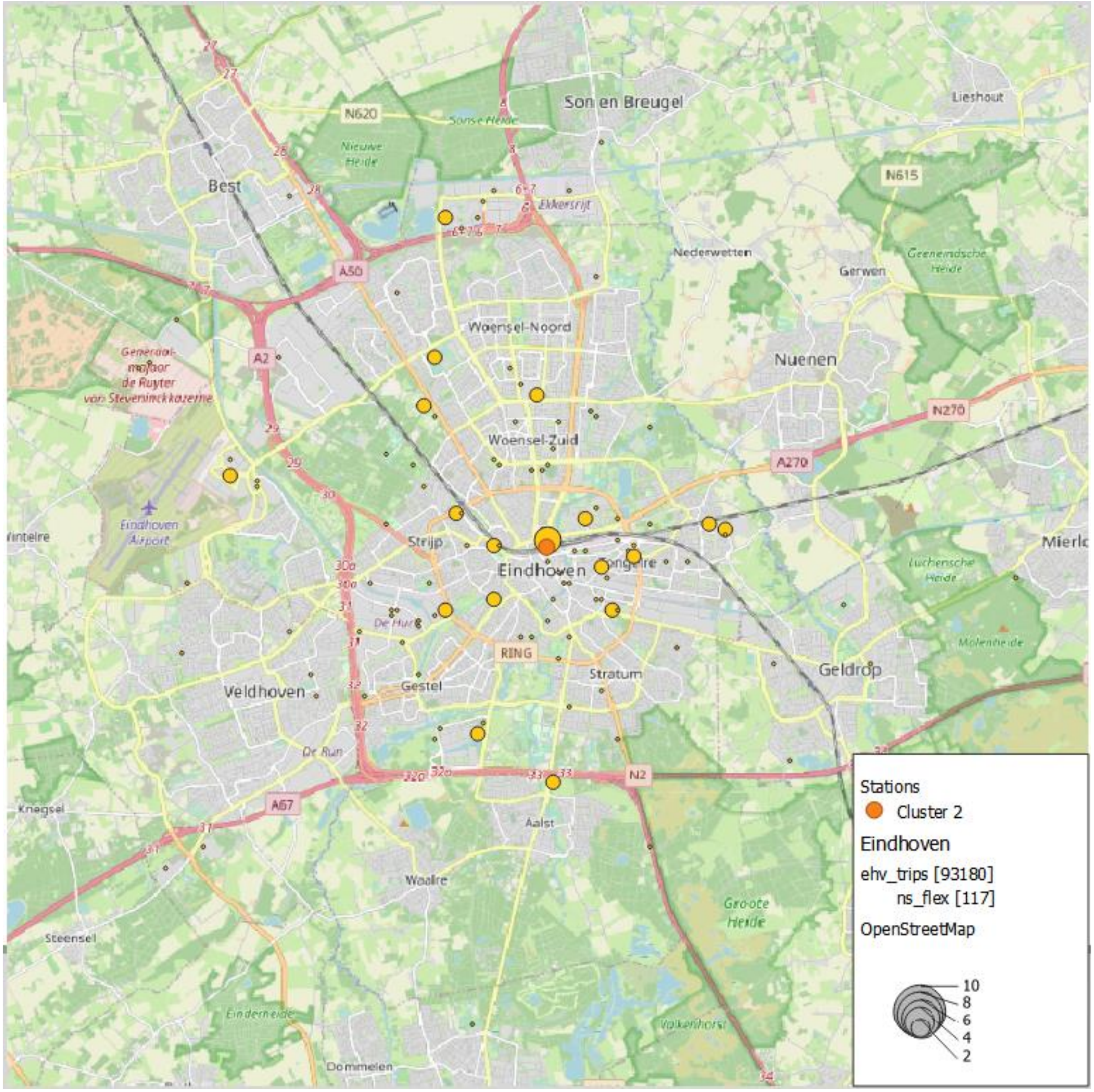


Figure 94: Eindhoven, simulated trips with NS flex OV-Chipkaart (16% of trips)

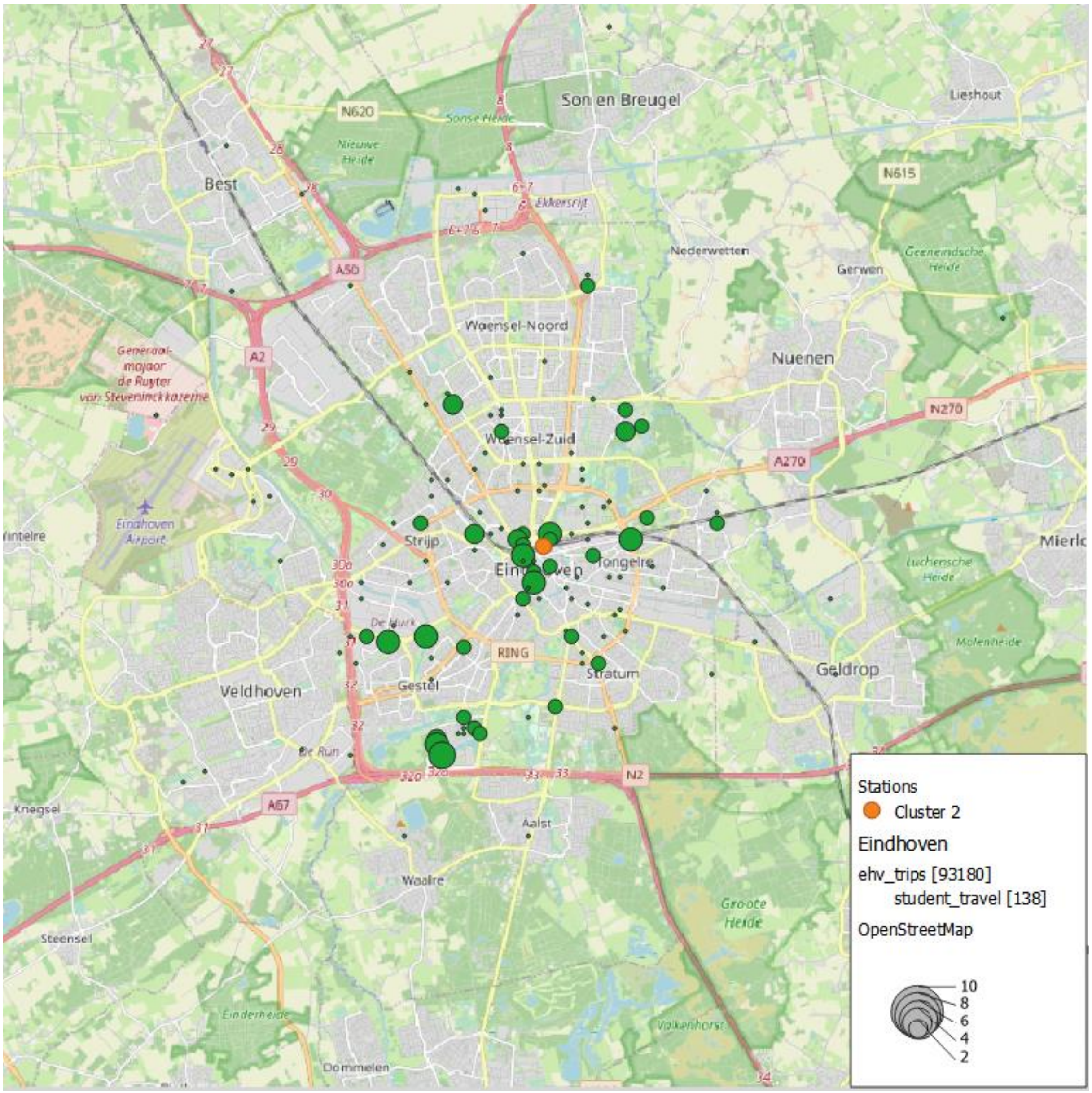


Figure 95: Eindhoven, simulated trips with Student Travel OV-chipkaart (15% of trips)

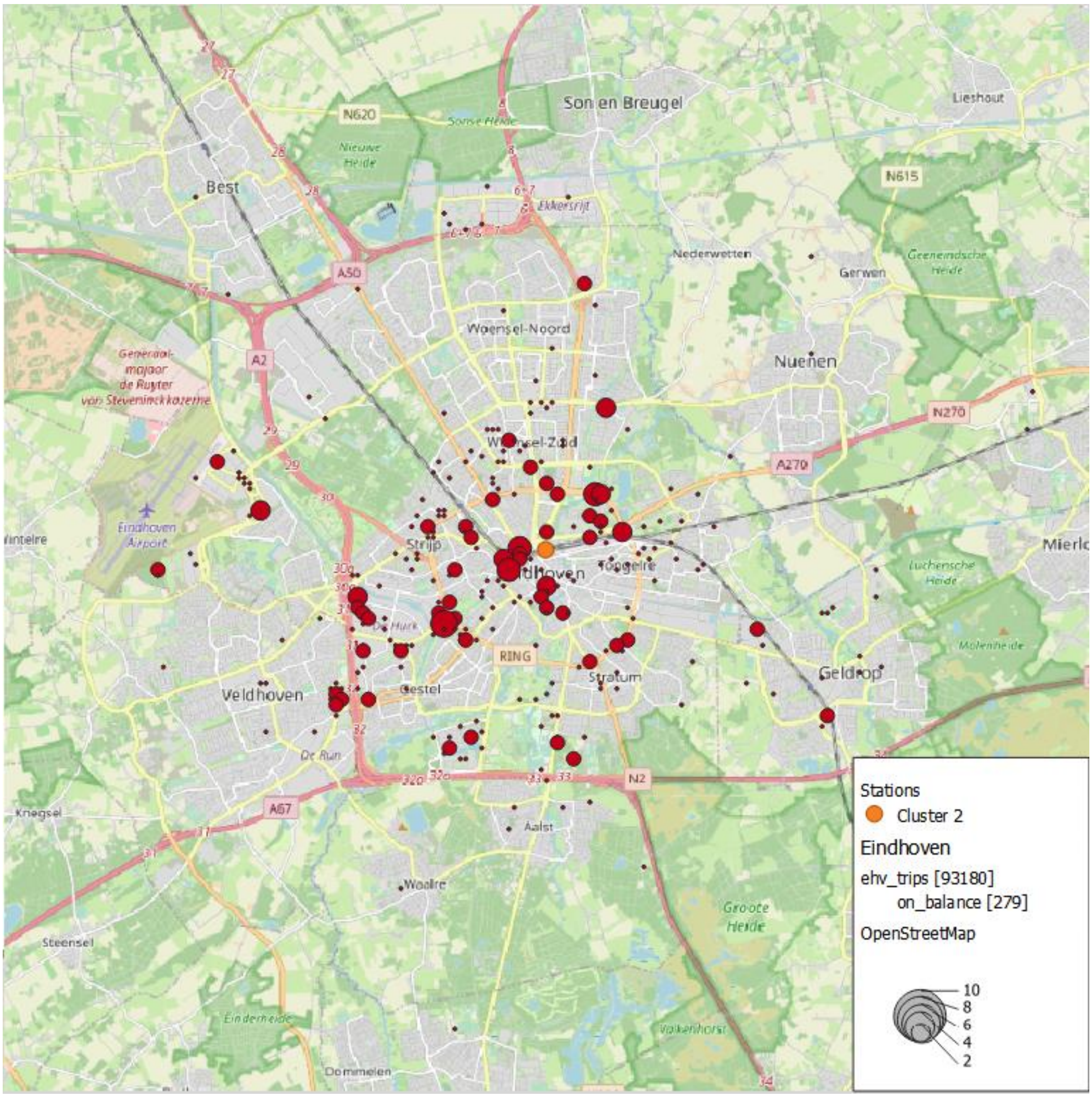


Figure 96: Eindhoven, simulated trips with On Balance OV-Chipkaart (33% of trips)

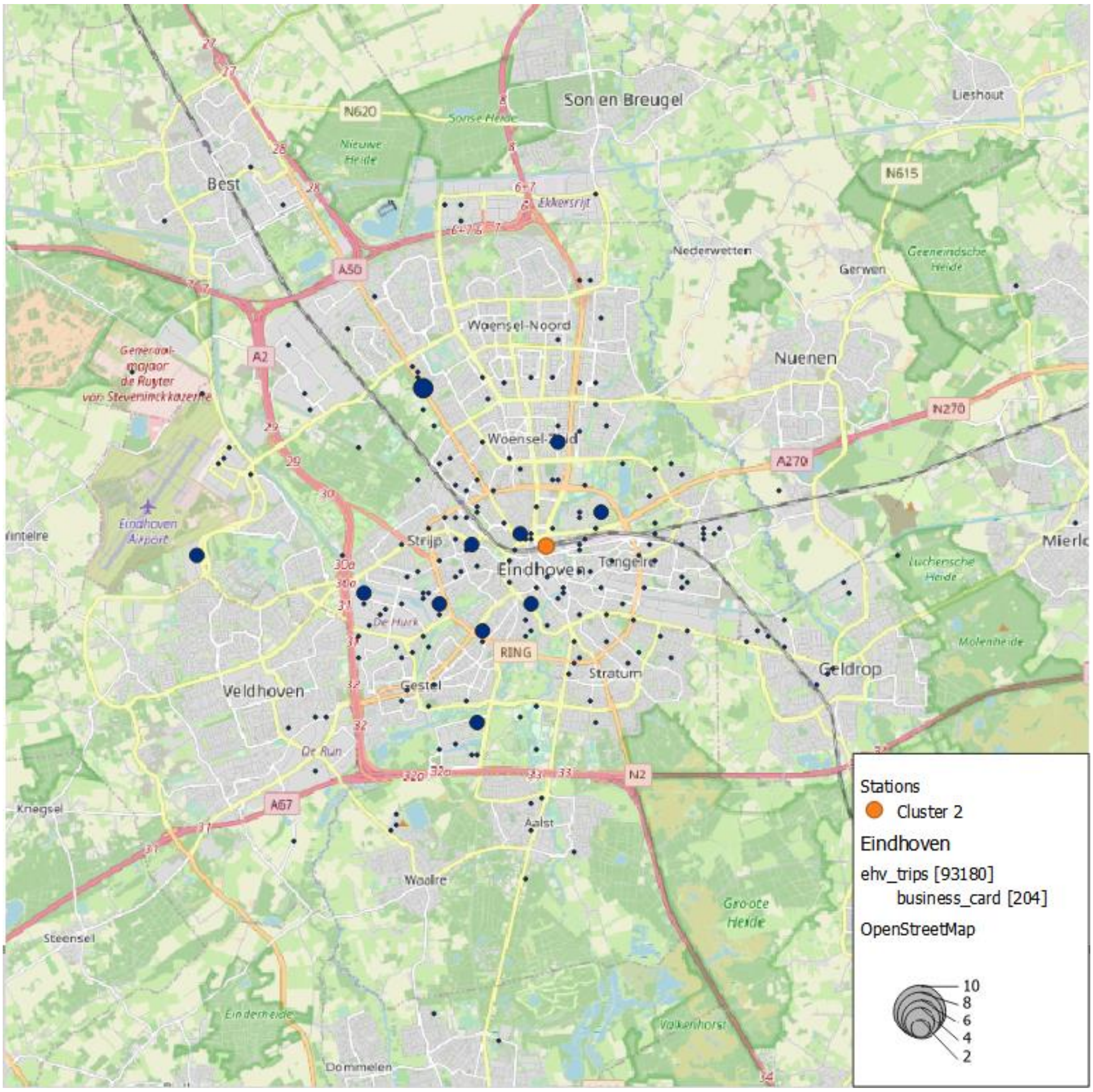


Figure 97: Eindhoven, simulated trips with Business Card OV-Chipkaart (30% of trips)