GEMEENTE ENSCHEDE

THE IMPACT OF HUBS ON THE USE OF SHARED MICRO MOBILITY

HUBS AS AN ALTERNATIVE TO A FREE-FLOATING SYSTEM

ENSCHEDE

The impact of hubs on the use of shared micro mobility

Hubs as an alternative to a free-floating system

Dylan T. van Bezooijen [a]
Author

M. Baran Ulak [a]
Internal Supervisor

Rik Goossens [b]
External Supervisor

July 9, 2023

[a]Department of Civil Engineering, University of Twente, The Netherlands
[b]Gemeente Enschede, The Netherlands

1 Executive summary

Shared mobility has seen an increase in usage the past years. Especially shared micro mobility, such as electric mopeds and bikes, has quickly spread to many cities. This sudden rise to prominence has led to problems, such as vehicles that have been parked obstructively by parking on narrow sidewalks and in front of building entrances. To address these problems and to further integrate shared micro mobility with the public transit network an effort has been made by the municipality of Enschede to pilot the proposed solution to these issues: Shared mobility hubs. The impact of these hubs on the usage of shared mobility, and how land use should be considered when transitioning from a free-floating to a hub system, is little studied and is largely unknown to policy makers. Therefore, this thesis proposes to asses the impact of shared mobility hubs and their relation with points of interest (POIs) on the use of shared mobility in Kennispark and Enschede Noord. In both city regions, a network of hubs has been active since November of 2022 and Juli of 2023 respectively.

To compute results on a hub-by-hub basis a network analysis in a geographic information system (GIS) software was executed in order to identify the service area of each hub. The results in Figure 1 show that the network of hubs effectively covered the entirety of Enschede Noord in such a way that everyone is within 300 metres of a hub, which was the objective of the municipality. This proves that the current approach of the municipality of Enschede is effective in creating a blanket hubs.

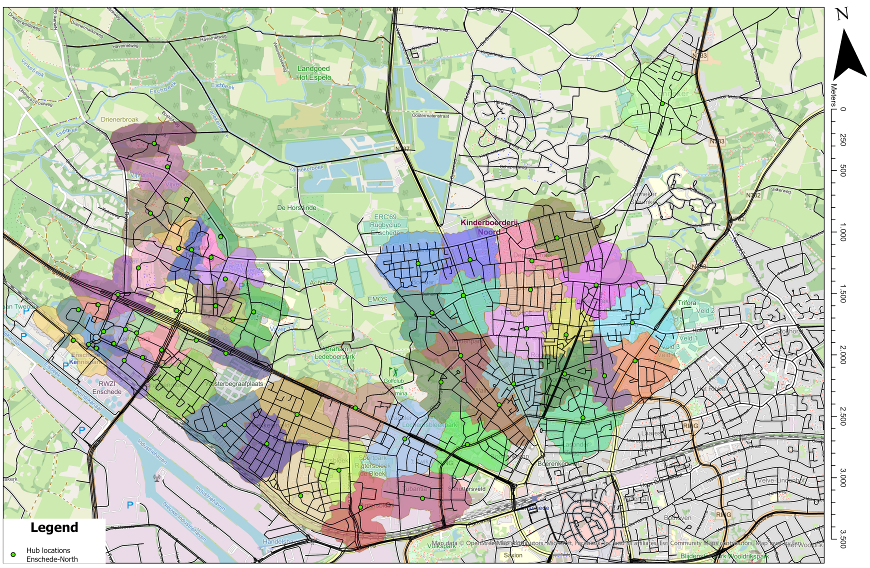

Figure 1: *Service areas for all hubs in Enschede North and Kennispark*

In order to gain insights into the relation between POIs and shared mobility a GIS analysis was done which has as goal to identify what the most visited categories of POIs are. The results seen in Figure 2 count the number of trip end-points, and assign each trip to the closest POI. Figure 2 shows that mobility-related POIs, such as bus stops and train stations, are the most popular destinations for users of shared mobility. This confirms the held notion that shared mobility is often used as a 'last mile solution' and proves there is an inherent synergy between shared mobility and other public transport. Educational facilities are also among the most visited locations, which confirms previous research that young adults are the main demographic group that uses shared mobility. Additionally, the analysis shows that shared mobility is used both as a 'one-off solution' (such as when visiting car or bicycle repair shop to pick up a repaired vehicle) as well as for commuters visiting offices in Kennispark. On the contrary, park and sport-related destinations are visited infrequently by users of shared mobility.

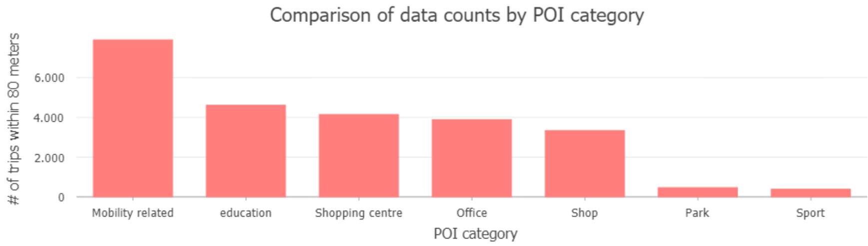

Figure 2: *Count of trip end-points by POI category*

It is not sufficient to analyse the usage of shared mobility before the implementation of hubs to the usage of shared mobility after the implementation of hubs, since the usage is affected by other variables aside from the implementation of hubs. Therefore, to fairly compare the usage of shared mobility before the implementation of hubs to the usage of shared mobility after the hubs, external factors which impact the use of shared mobility need to be accounted for. Several variables were found to be relevant in a negative binomial regression model (n=484 days). The variables that were found to be statistically significant (p¡0.05) are the supply of vehicles available for hire, seasonal effects such as temperature and rain, temporal factors such as holidays and special occasions such as football matches and festivals. These variables were used to train a machine learning algorithm (XGboost) for each individual service area seen in Figure 1 using previous usage data as a training dataset (n=120 days for service areas in Kennispark, n=516 days for service areas in Enschede North). With this model, predictions were then made for all days after the hubs had been implemented (n=211 days for Kennispark, n=22 days for Enschede Noord), this way all variables that influenced the use of shared mobility were accounted for. The predicted number of trips was then compared to the actual number of trips in order to get the residual, which would be the impact of hubs on the usage of shared mobility.

The results of the analysis can be seen in Figures 3 and 4. Hubs have a significant impact on the usage of shared mobility, but outside of their spatial context, it is difficult to say whether or not the use of shared mobility will increase or decrease. In this case study the hubs were shown to positively contribute to the usage of shared mobility in Kennispark by 15%, whilst the hubs contributed negatively to the usage in Enschede Noord by -10%. The data collected in Kennispark is more reliable then the data collected in Enschede Noord due to the limited time frame in which hubs were active in Enschede Noord. Also, because the count data in Enschede Noord only included counts right after the implementation of the hubs, it is possible that the results were inaccurate due to there being a warm-up period in which users have to acclimate to the new system. Therefore, there is reasonable prove in the results that suggests hubs have an overall positive impact on the use of shared mobility.

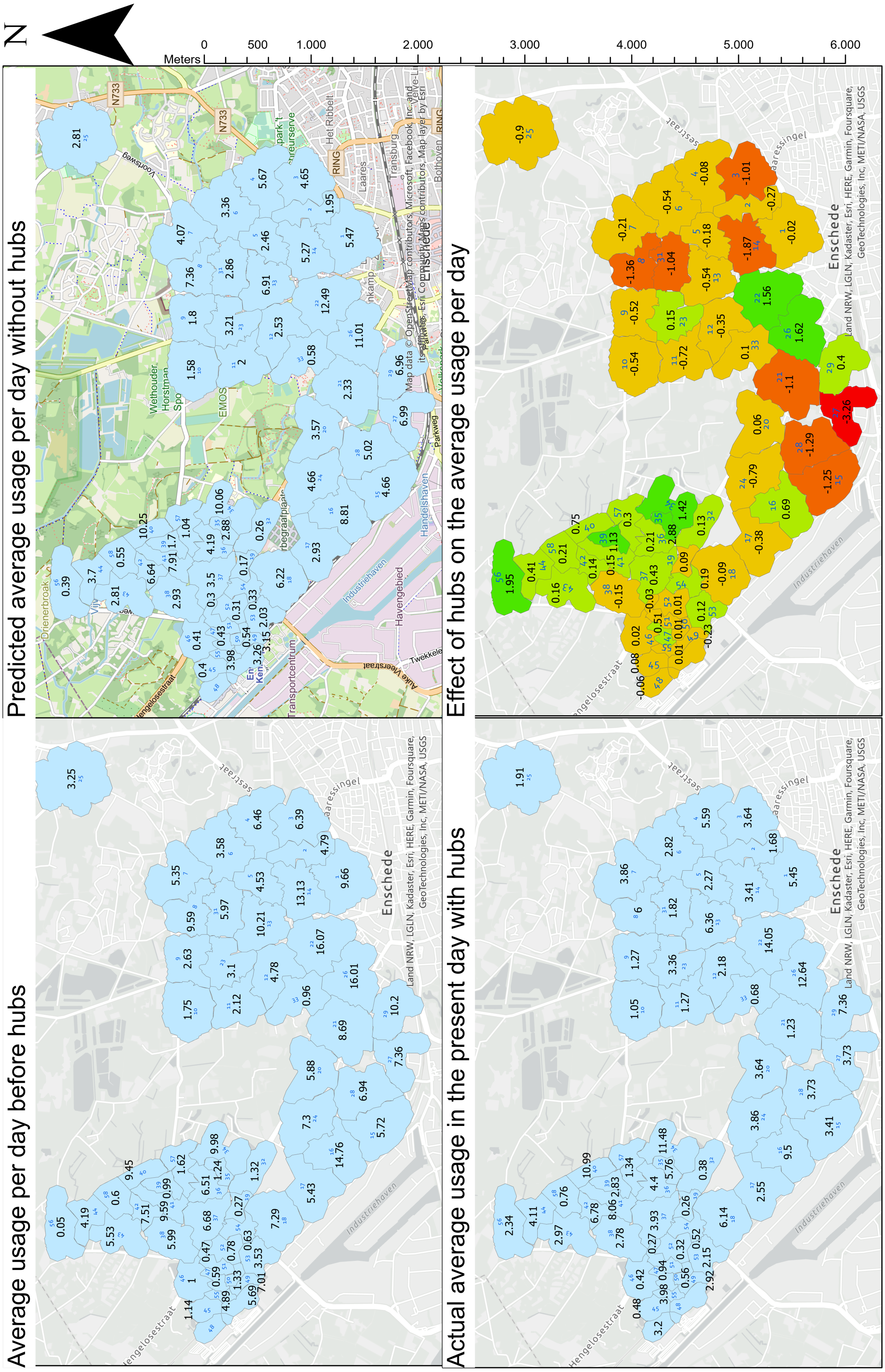

Figure 3: *Impact of hubs on the usage of shared mobility on hub level*

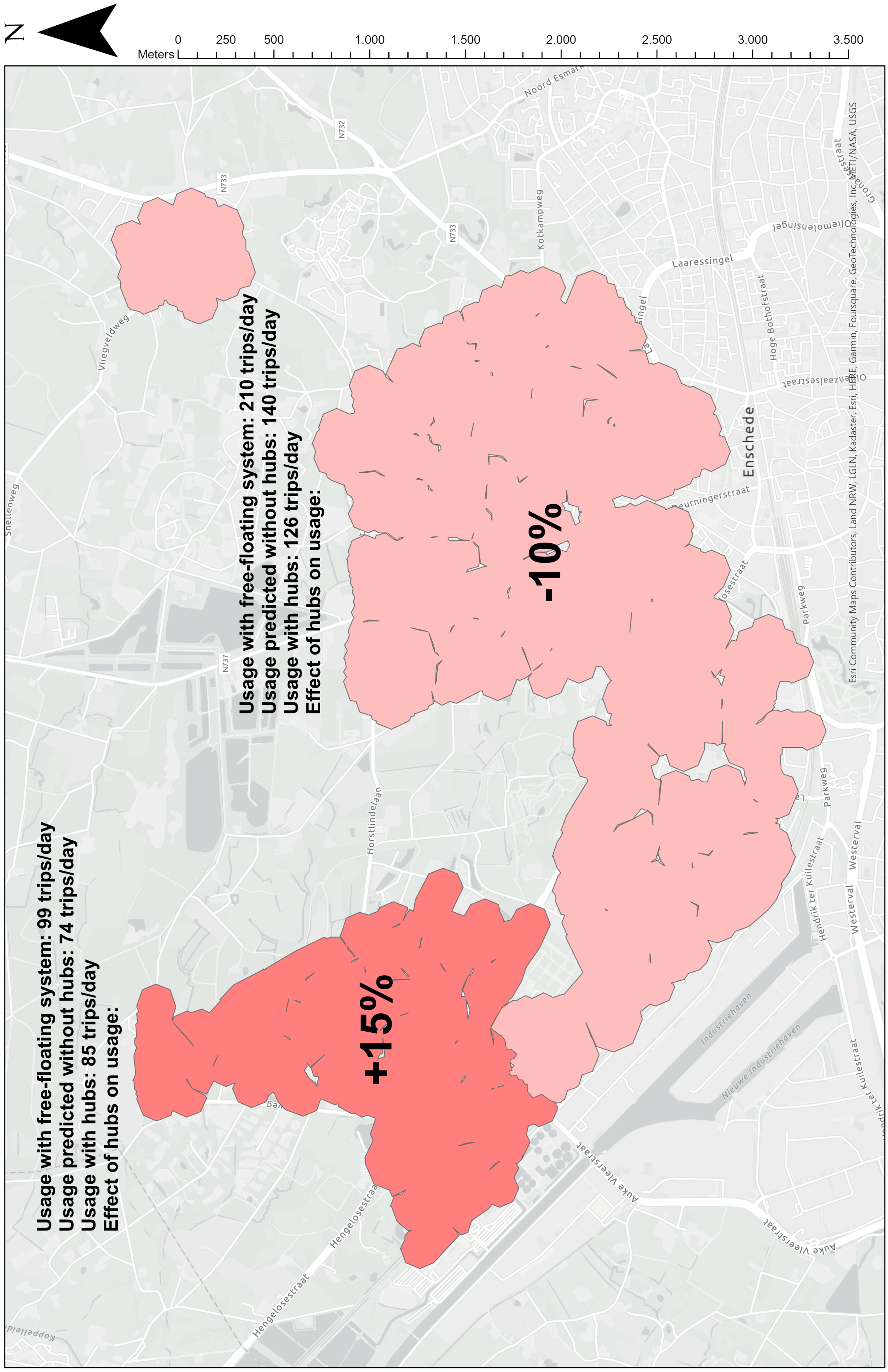

Figure 4: *Impact of hubs on the usage of shared mobility on a city region level*

To gain insight into the land use interactions between the hubs and the POIs, a multi-linear regression model is run to derive any correlation between the number of POIs in the service area of a hub and the impact of the hubs on usage. The results found that the number of mobility related and educational POIs close to the hub have a statistically significant impact. Hubs close to educational POIs are positively impacted by the change from a free-floating system to a hub system; meaning that service areas in which there are a lot of educational POIs, the change in usage caused by hubs was relatively positive. Whilst there seems to be a correlation here nothing can be said about whether they are causally related, this thesis proposes some theories. Young adults visiting educational institutions rely more on shared mobility as a main mode of transport than other demographics, as they often do not have their own motorised vehicle. Furthermore, this target demographic is more willing to accept the increased walking distances and the inconvenience caused by this (Vianen, 2022), which is the main disadvantage of the hub-system compared to the free-floating system. This makes this demographic more susceptible to the positive effect of the hub system, such as recognisability and reliability of finding a vehicle. Mobility related POIs such as bus stops have a negative impact on the hubs when changing from a free-floating system to a hub system; meaning that service areas in which there are more mobility related POIs the change in usage caused by the hubs was relatively negative. A theory is that this is because the free-floating system was the ideal candidate for the last-mile solution. Now that the hub system has reduced the usefulness of shared mobility to fulfill the role of last-mile solution, hubs which previously saw a large percentage of their users travelling from/to public transport services are more heavily impacted by the change.

Previous results indicated that education- and mobility related POIs are important destinations and results now indicate these specific POIs as being sensitive to the change from a free-floating system to a hub system. Thus, it is this thesis' recommendation to prioritize and minimize the walking distances between hubs and these POIs. This is especially of importance if shared mobility is to be a 'last mile solution', since results indicate that public transport users are using shared mobility less due to the change from a free-floating to a hub system.

Whilst this thesis has done its best to draw conclusions from the data and results, some limitations are in place. Firstly, the model which accounted for other variables such as seasonal effects, supply, and other variables is not perfectly accurate, leading to some uncertainty in the model results. Last, the scope of the case study was only limited. Meaning that the conclusions made in the context of Enschede are not certain to hold in other hub networks, cities and cultures. Therefore, it is this thesis' recommendation to replicate similar studies in other hub networks, cities and countries. Also, this thesis suggested that there might be a warm-up period between the transition from a free-floating system to a hub system in which users have to get used to the change. This would be another possible avenue for further research, and can already be done with the current data used in this thesis. Additionally, this thesis would suggest doing an in-depth case study for a hand full of hubs instead of a 'broad' study looking at an entire network. This way more subtle results which data cannot show might come to light. Furthermore, whereas this thesis focused more on the impact of the hub system as a whole, more research could also be done regarding how other aspects of the hubs influence behaviour and usage of shared mobility. Think of this such as size, visibility and recognizability of the hub. Finally, this thesis suggests a research studying the' theories as to why there is a significant relation between the impact of hubs and the number of mobility related POIs and education POIs in the vicinity .

Contents

List of Figures

List of Tables

2 Introduction

In the last years shared mobility has shown to be a popular alternative to more traditional modes of transport, especially in urban areas. This emerging transport mode is radically different from the traditional mobility pattern of owning, using, and maintaining one's own vehicle. Instead, shared mobility providers take care of these services themselves, for which they charge the user. This "mobility as a service" (MaaS) model has several advantages for users that other modes of transport do not have. For example, users are allowed to park their vehicle wherever they want without having to worry further once they have terminated the 'service' . This gives shared mobility the potential to be a last-mile solution[1]. Therefore, more traditional public transport providers, such as the Dutch Railways, have been keen to implement their own version of shared mobility. (Jorritsma et al., 2021). The concept of allowing users to park their vehicle wherever they desire is commonly referred to as a 'free-floating' system (Machado et al., 2018). However, the advantages a free-floating system has for users can be to the detriment of others. (van Steijn et al., 2022). Since the introduction of shared mobility there has been controversy over their parking patterns, which includes examples such as parking on narrow sidewalks or to the entrances of buildings (NOS, 2020). This story has been seen across many large to mid-sized cities in the Netherlands, such as Enschede, who is now at the forefront of implementing shared mobility hubs in order to solve the obstructive parking behaviour.

Figure 5: *A shared mobility hub in Kennispark, Enschede*

In September of 2022 the municipality of Enschede initiated a pilot to introduce shared mobility hubs in Kennispark, a high-tech business park on the north-western edge of Enschede. These mobility hubs only allow users to park shared transport vehicles such as mopeds and bikes in designated zones. At the pilot location these measures have proven to force users to park their vehicle in a 'tidy' manner (van Steijn et al., 2022). Now that the pilot has been proven to solve the issue of obstructive parking, the municipality of Enschede is preparing to roll out the next batch of hubs in the remaining northern parts of Enschede (called Enschede Noord). This batch of 40 new mobility hubs has been implemented in

[1] 'Last-mile solution' is in reference to one of the disadvantages of public transport; it is usually not a door to door transport mode, which makes users responsible for organizing their own transport for the last portion of their trip

July of 2023. Eventually the municipality aims for a network of hubs covering the entirety of Enschede (Goossens, 2022), as there are already plans for hubs in the eastern part of Enschede as well. With the hubs the municipality is giving shared mobility a permanent place in the city, both formalizing the usage as well as providing guarantee to shared mobility providers that there is a definite place for them in the city. Additionally, the municipality has approved new legislation which requires service providers to get a permit before being allowed to operate in the municipality (Enschede, 2023), further formalizing shared mobility. Users have reported to like the concept of the hubs, however they are worried that the hubs will increase walking distances to their end location, which would lead to them using shared mobility less often (van Steijn et al., 2022). This stated problem will be the core of this thesis, and will be elaborated in the next subsection.

2.1 Problem definition

There are two main avenues for research in which the municipality is interested. Firstly, although the hubs have a positive effect on parking behaviour and therefore the municipality is eager to adopt the hubs, little is known by the municipality on the impact of changing from a free-floating system to a hub system (van Steijn et al., 2022). Users suspect that they will use shared mobility less often, since they think walking distances will increase and make shared mobility less attractive compared to other modes such as cycling, public transport and cars. The notion that walking distances increase is based on the fact that in the free-floating system users can decide where to park their vehicle at their own behest, which infers this is as close as they can get to their end destination. The hub system abolishes this privilege and forces users to park at set locations, which means the remaining distance to their destination has increased. This concept is also illustrated in Figure 6. However, this notion remains an assumption. As can be seen in Figure 6 there is uncertainty in the free-floating about the walking distance from the start of a trip to the nearest vehicle, since the nearest vehicle might be anywhere. This distance is fixed with the hub system, possibly reducing the walking distance from the start of a trip to the nearest vehicle. The second avenue for research in which the municipality of Enschede is interested is the interaction between land use and hubs. The assessment framework that is used by the municipality to decide on the location of the hubs includes land use factors such as 'anchor points' or 'Points of Interest' ('POIs') [2](Vianen, 2022). Furthermore, in a study outsourced by the municipality of Enschede in order to asses the hubs in Kennispark, van Steijn et al. (2022) suggest doing a study in a residential area because the researchers suspect the land use has an measurable impact on the usage of shared mobility.

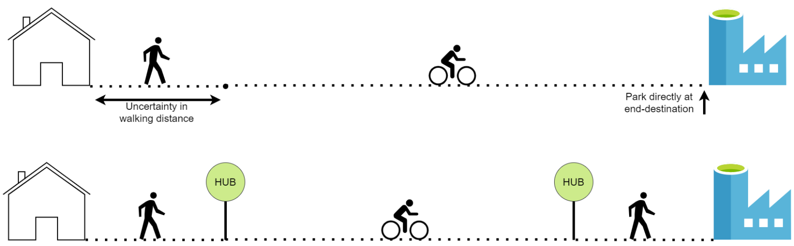

Figure 6: *With the introduction of mobility hubs the free-floating system is abolished, and the distance travelling on the shared vehicle or by foot changes.*

Scientific literature on these two avenues for research is scarce. This lack of knowledge

[2]The notion of anchor points relates to those locations which attract a certain amount of visitors/passenger on a daily basis. The term 'anchor point' is not grounded in scientific literature, therefore the term 'POI' is used instead.

in both the scientific field as well as by policy makers can be translated into scientific gaps. First, there have been few direct studies on the impact of mobility hubs on the use of shared mobility hubs. Second, there is little empirical proof in scientific literature about how factors that are related to land use impact the use of shared mobility. For example, it is agreed upon by experts that so called 'anchor points' are important factors in deciding the performance [3] of a mobility hub (Blad et al., 2022; van Steijn et al., 2022; Vianen, 2022). However there is little actual empirical proof of this available. The crux of the problem can be defined as **a lack of empirical evidence in the practical field as well as the scientific field regarding the impact of changing from a free-floating system to a hub-based system on shared mobility**.

2.2 Reading guide

The remaining part of this thesis is structured in the following manner: First, chapter 3 gives a review of all scientific literature related to our problem. Chapter 4 uses this knowledge in order to create a research objective and a set of research questions which will contribute to the field in a practical sense and in a scientific sense. Chapter 5 is used to elaborate on the method used to answer the research questions. In the chapter a detailed description is given of *how* all sub questions will be answered. That is, the GIS and machine learning tools used will be explained and the necessary data input listed. Chapters 6,7 and 8 will go over the intermediate results used to finally produce the final results in chapter 9. This is done in the manner of several graphs, tables and maps. Chapter 10 discusses the results, after which chapter 11 digests all the results to formulate a conclusions, recommendation to policy makers and paths for further research. The appendices include all scripts used for this thesis.

[3] Measured in the usage of shared mobility.

3 The impact of hubs: literature review

The concept of a mobility hub has been around for some time and is an umbrella term for a large variety of public amenities. Blad et al. (2022), Seker and Aydin (2023) Rongen et al. (2023) and Zhou et al. (2023) all agree that a mobility hub can be defined as a place where multiple modes of transport are linked together. In practice this definition translates to many different scales of hubs; from the relatively small neighborhood mobility hubs which are this thesis' concern, to large district hubs such P+R's and large regional hubs such as train stations. The Ministry of Infrastructure and Water Management's typology is in line with this definition and has made their own definitions for the different scales of mobility hubs, as seen in Figure 7 .

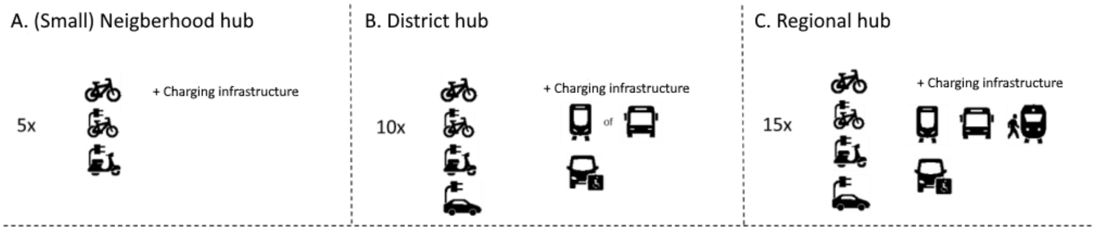

Figure 7: *The typology for hubs as defined by the Ministry of Infrastructure and Water Management*

With the introduction of the hubs, the municipality is transitioning to a back-to-many-station based system (Machado et al., 2018)(for ease of reference this thesis refers to this system as a 'hub system'). As explained in the introduction of this thesis, the hub system changes the walking distances a user will make on their trip, as was conceptualized in Figure 6. This theoretical increase in walking distance is the reason why participants in the study commissioned by the municipality stated that the use of shared mobility has become less attractive to them (van Steijn et al., 2022). This in-situ research suggests that the mobility hubs are a negative factor influencing the usage of shared mobility. Note however that these are not stated-preferences [4]. Rather, they asked existing users what they thought of the hubs. Thus, there is no saying if the hubs might cater towards a whole different group of users to which the hubs might be an added benefit.

The impact of mobility hubs on travel mode choice and modal split was studied with an activity-based-model by Zhou et al. (2023). In the paper, it was found that the introduction of mobility hubs had limited impact on reducing car-ridership. However, combining mobility hubs with sharing services, such as those seen in Enschede, reduced car-ridership by 3.9%. This suggests that mobility hubs are an added benefit to the mobility network when seeking to reduce car-ridership, which in turn suggests that the hubs lead to an increase in usage. However, the research was focused on larger district hubs, in contrast to the smaller neighborhood-level mobility hubs this research is concerned about. The difference in scale can lead to a difference in results.

Most studies into mobility hubs looks into these larger scaled hubs such as district or regional hubs. A relevant master thesis that researches neighborhood-level mobility hubs was done by Vianen (2022). The thesis looked into creating a design approach to determine the most suitable locations for mobility hubs. In it he describes certain 'anchor points',

[4] In a stated preference research a representative sample of the population is asked to state their preference for a modal choice in certain situations.

which are "locations that attract a certain amount of visitors/passenger on a daily basis". Instead of 'anchor points' this thesis refers to these locations as 'POIs', since it is a better term for the aforementioned definition. In the thesis it is described that certain POIs are ideal locations for hubs. This suggests that mobility hubs in proximity to relevant POIs can see a positive influence on the usage of shared mobility. This takeaway is something the municipality also considered in their plans and is why 'POIs' were considered in deciding the locations of the hubs in Kennispark and Enschede Noord.

More general research looking into how spatial factors contribute to travel mode choice also give insights into the impact of mobility hubs on the use of shared mobility: studies by Buehler (2011) and Srinivasan (2005) finds that land-use, land-use balance, and accessibility (as measured in travel distance x utility) were significant factors in determining travel mode choice. With the introduction of the mobility hubs the land-use and land-use balance around the hubs do not change. The only changing variable is accessibility, since walking has a lower utility than cycling due to the decrease in travel speed.[5]. Therefore the introduction of mobility hubs would negatively influence the use of shared mobility. This decrease in utility is also the reason why users in the research by van Steijn et al. (2022) stated that the hubs made shared mobility less attractive; they foresaw walking distances might increase, therefore reducing the utility.

Miramontes et al. (2017) looks at the problem with a more cognitive approach. The paper makes the observation that users of mobility hubs, and in extent the users of shared mobility, to a major extent discover these things by chance when passing by. This suggests that the mobility hubs, if well visible, lead to an increase in awareness of the existence of shared mobility and therefore might increase use. Also, the hubs come with signs which explain how to use the shared vehicles. An example of these signs can be seen in Figure 8. This can have a positive influence in reducing the technological-divide which may be one of the reasons why shared mobility is mainly used primarily by young, highly educated males and less by older generations which are less used to apps and other digital interfaces (Li & Kamargianni, 2018).

Figure 8: *An example of a sign explaining how to use a the shared electric bike at the mobility hub, source: https://www.mobiliteitshubs.nl/nieuws/103-zo-gaan-hubs-er-uitzien*

[5]Utility is the function of all variables which people consider when deciding which transport mode they take. This includes things such as speed of travel, comfort etc.

3.1 Takeaways from the literature review

All in all, the literature seems somewhat divided on the impact of mobility hubs on the use of shared mobility. Initial research (van Steijn et al., 2022) and more traditional research (Buehler, 2011; Srinivasan, 2005, suggest that introduction of mobility hubs might negatively impacts the use of shared mobility. Other research (Zhou et al., 2023,Miramontes et al., 2017) suggests there are reasons to believe the mobility hubs will increase shared mobility usage. In general little research is done directly on the impact of mobility hubs on the use of shared mobility, which gives this thesis an unique opportunity to fill this scientific gap. Furthermore POIs is already used in practice by municipalities (van Steijn et al., 2022; Vianen, 2022), however there is no empirical evidence for POIs close to mobility hubs contributing to the use of shared mobility.

4 Research objective & questions

Based on the problem definition and the literature review the following research objective is appropriate:

To asses the impact of shared mobility hubs and their relation to POIs on the use of shared mobility in Kennispark and Enschede Noord

The 'shared mobility hubs' will be refered to as 'mobility hubs' or simply 'hubs'. 'Shared mobility' refers to only those service providers which will make actual use of the hubs. The forms of shared mobility included are e-bikes (provided by Bolt) and mopeds (provided by Check and Felyx and GoSharing in the past). This research therefore excludes other shared mobility services like Greenwheels (cars), or the OV-fiets (bikes) provided by Dutch Railways. Whilst the cars cannot make use of the hubs for obvious reasons, the OV-fiets is not banned from using the hub and it might very well be the case some people do. However, the OV-fiets is allowed to park outside of the hubs as well and therefore it falls outside the scope of this research.

The main- and sub research questions which serve to complete the research objective are as follows:

Main research question:

1. What is the impact on the use of shared mobility when there is a change from a free-floating system to a hub system?
2. What is the impact of POI type, count and distance to the mobility hub on the use of shared mobility?

The above questions cannot be answered before first answering four sub-questions. These four sub-questions indicate the intermediate results which will be used to answer the main research questions.

Sub-questions:

1. What is the service area of a hub based on walking distances?
2. Which types of POIs are commonly visited by users of shared mobility?
3. How do we fairly compare the use of shared mobility before/ and after the mobility hubs, accounting for external factors which influence the use of shared mobility in Enschede?
4. What are the most important takeaways from the impact analysis for policy makers?

4.1 Scientific relevance

Shared mobility has risen as a new form of transport in the past decade, meaning researchers still have little knowledge about some of it aspects. This thesis further advances the scientific field on the topic two ways: First, by assessing how sensitive shared mobility is to a change in system, in this case from a free-floating system to a hub system. Second, this thesis produces knowledge on land-use and transport interaction in the domain of shared mobility. The interactions between land-use and transportation have been long studied, and the emergence of new types of transportation like shared mobility warrants new research into these interactions. This study hopes to find what features of the build environment (represented in this study by POIs) have an impact on the use of shared mobility, and which do not.

4.2 Practical relevance

The complains about obstructively parked vehicles are part of a larger trend in the municipality of Enschede where there is growing pressure on the use of public space for problems related to mobility, water management, energy-transition, heat-islands and much more. Furthermore, there is no example in the Netherlands of a network of hubs which cover an entire city. Therefore, the municipalities goal of creating a network-covering grid of mobility hubs (Enschede, 2019) is a new concept not previously seen in the Netherlands. It is no surprise then that the ministry of infrastructure and water management and other municipalities are interested in what results the mobility hubs will bring (R. Goossens, personal communication, 12 January, 2023).

5 Method

The research framework in Figure 9 shows the required steps in the research that will be taken in order to reach the research objective. Going from left to right in the Figure: The main study object of the research will be the mobility hubs as well as Enschede Noord, this study object represents all roads, buildings and other build features which may be present. First, a model needs to be made which represents the topological boundaries of a hub. Afterwards, in order to link POIs to service areas, all relevant POIs and their locations need to be determined.

A lot of external factors might influence the use of shared mobility, and therefore skew the comparison between the before-hub and after-hub usage counts. In order to account for these external factors usage data of shared mobility before the implementation of hubs and external factors which might affect this data (such as weather effects and the supply of vehicles), are used to train an extreme gradient boosting algorithm (XGboost) in order to predict the usage of shared mobility. This algorithm can then be used to predict the usage of shared mobility in Enschede Noord without any hubs, given a set of inputs. These predicted values would represent the situation without the hubs. These predicted values can then be compared to actual usage data once the hubs are implemented in order to get data describing the change in use caused entirely by the hubs. This can be done for each service area, giving us a change in use for all the hubs and the corresponding service areas.

Variables related to POIs, such as distance from the hub to POIs and counts of POIs near a hub, can also be linked to the service areas. If this is all fed into a regression model, then the variables can be separated from the impact of the hubs on use of shared mobility. Finally, the results can be digested into an impact assessment for the municipality of Enschede in the form of a discussion and conclusion. All the steps explained in this brief overview are discussed in detail in the following sections.

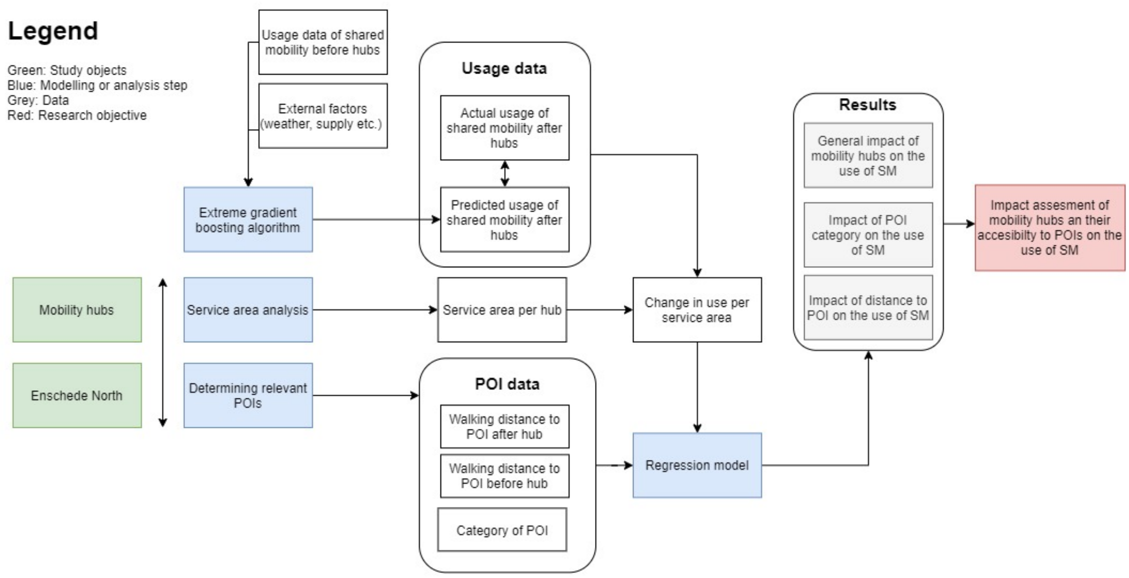

Figure 9: *Overview of the method used in this thesis*

5.1 Service area analysis

The service area analysis is the first step that is performed. Its goal is to generate a set of polygons which represent the service area of each hub. The municipality of Enschede has as an internal goal to create a network of hubs where walking distances to a hub will be no more than 300 meters.

Assumption

- After parking their shared vehicle at a mobility hub, users travel the remaining distance to their (intermediate) destination on foot.

The assumption uses the notion that shared mobility is often a first/last mile solution (Jorritsma et al., 2021), consequently meaning that the last remaining distance to users' front door or intermediate destination is done on foot. This assumption is important to make to determine the service area, since it will be based on walking distances, rather than distances based on other modes of transport. This is also an assumption made by the municipality of Enschede. (Goossens, 2022; van Steijn et al., 2022)

In order to perform a service area analysis, a digital network must be built along which users of shared mobility would travel towards or from their end destination towards the hubs. In this case, a vector network is most suited. in a vector network roads are represented by two-dimensional lines on a planar surface along which users travel. Lines only intersect at designated 'nodes'. Furthermore, it is assumed that by walking users can travel in both directions of a given line segment (i.e. all roads are 'two-way' for a pedestrian) meaning that we do not need to specify this difference as in a directed-network, and that an undirected network is sufficient for the purpose of this analysis. ("The Core of GIScience", 2020). Once the network is built, we can perform a 'service area analysis' in GIS in order to generate the polygons which will represent the service area of each hub. Table 1 summarizes all necessary data that is required for the method.

Table 1: Data overview: Service area analysis

Data description	Use case	Data type	Source
Walking paths	Determine service area	Shape files	PDOK
Location of mobility hubs	Determine source of service area	Point data	Municipality of Enschede
System borders	Outline system borders	Polygon	Municipality of Enschede

5.2 Determining relevant points of interest

This research aims to include the effect of the distance of points of interest to the mobility hub as well as the category of POI on the use of shared mobility. First, it should be determined *Which types of POI are commonly visited by users of shared mobility?*. Three steps in order to answer this several steps have been identified:

1. Map all possible POIs based on existing literature on the topic by Vianen (2022), expert knowledge and wishes from the municipality to look into certain POIs.

2. Induce relevant POIs from usage data using hot spots. Using assumption 1 of Section 5.1 it can be assumed that users park their vehicle at end destinations, if many end destinations are densely parked at the same location, it can be an indication of a possible POI in the vicinity. To visually identify this relationship, POIs must be visualized on the same map.

3. Induce relevant POIs from usage data using parking distances. The assumption is made that users park their vehicles as close as possible to their end-destination. If each trip is assigned to a certain POI based on parking distance, then destinations can be counted to identify relevant POIs. Before this can be done, a maximum parking distance needs to be decided on, within which distance the identified trip destination can be counted with reasonable certainty. For example, in areas where there are little POIs in the vicinity, so that the closest POI is 200 meters away, it would be unrealistic to assign this vehicle to that POI. since the intended destination is probably different. This idea is schematized in Figure 10. However, we must be careful to not choose a maximum parking distance that is too small, since user might not be able to park right next to their destination (in the case of in-door malls, one cannot get their vehicle inside), or a user might be inclined to park at a more accessible sport further away. For example, an existing parking spot for bikes and mopeds. To make the best possible decision for this maximum parking distance a histogram of the parking distances to POIs can help identify what the usual parking distance is from a POI.

   Then, all vehicles which fall outside this maximum parking distance can be assigned to the nearest POI. Finally, the number of trips can be counted for each POI category. In order to validate if the method is accurate, the results can be cross-validated with the method based on heat maps. The results should indicate the same.

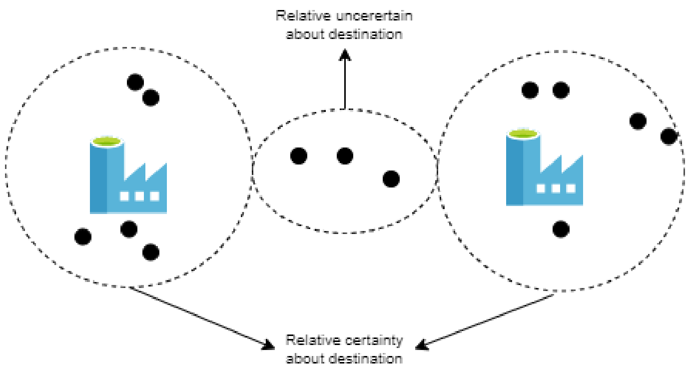

Figure 10: *There is a 'maximum parking distance' after which there is a large uncertainty whether or not the nearest POI is actually the indented POI.*

All the necessary data for these steps has been identified and summarised in Table 2

Table 2: Data overview: Determining POI's

Data description	Use case	Data type	Source
POI's based on expert elicitation	Determine what land-uses or other objects count towards POI's	land-use class or other	Expert knowledge/literature
POIs in Enschede	Label all relevant POI's	Point data	Openstreetmap
Trip end-points of shared vehicles	Methods 2 and 3	Point data	Dashboard deelmobiliteit (CROW)

5.3 Predicting shared mobility usage with XGBoost

In order to predict the usage of shared mobility a type of machine learning algorithm can be used called extreme gradient boosting (XGboost). This algorithm, if we simplify the explanation, makes use of decision trees in order to predict a result (called the dependent variable) based on a number of independent variables[6]. The working of XGboost are quite complicated and are not discussed in this thesis. However, main advantages of XGboost include that it is accurate, fast and is resilient to some problems ordinary regression models struggle with; such as over fitting a model to the data (Morde, 2019). XGboost does have a downside however: it is fairly black-box, meaning that the inner workings of the model are hard to extract. For example, whilst XGBoost does provide a 'gain' factor which indicates the relative contribution of each independent variable in the decision tree, it does not indicate *how* this factor contributes (i.e. in a negative or positive way). Therefore, XGboost is very suited as a way in which to predict the independent variable, but not so much as a model to understand the relation between predictors and dependent variables. This thesis is also expected to show results which give readers insight into what variables influence shared mobility in what way, therefore a regression analysis is also included in the results. Both the XGboost model and the regression model will require data, therefore an overview is given of all the used data.

5.3.1 Dataset

Several independent variables were identified by expert elicitation with M.B Ulak and R. Goossens on the topic, as well as by consulting the service providers Felyx, Check and Bolt. [7]

- Supply of shared vehicles
  The supply of vehicles can have large impact on the use of shared mobility. A large supply directly translates to a higher density of vehicles across the city, which leads to smaller distances to the nearest vehicle, this is an important factor for people deciding whether or not to hire a shared vehicle. (P. Kokos (Felyx regional manager), personal communication, 30/05/2023)

- Seasonal effects
  Seasonal effect plays an important role in modal choice. Seasonal effect considered in this thesis are temperature and rainfall.

- Day of the week
  In Enschede, most shared vehicles are used during weekend days and less during weekdays. This pattern is useful to include in the model since it will improve the ability to predict usage of shared mobility.

- Events
  Events where a large gathering of people are mobilized towards a certain location can generate a lot of traffic, shared mobility is no exception for this. Events which are considered in this thesis include match days of a large football club (FC Twente) as well as public holidays (liberation day, new years etc.) and also the time of introduction for new students at the university (Universitieit of Twente) as well as a school of applied science (Saxion Hogeschool).

[6] A layman's term for independent variable would be 'predictor', which indicate that the variables are co related to the dependent variable and are therefore a good predictor of it.

[7] Service providers often have their own models which they use to determine the fleet size to deploy in a city.

Table 3 summarizes all data for the independent and dependent variables that is required for this part of the method.

Table 3: Data overview: Predicting use of shared mobility

Data description	Use case	Data type	Source
Trip start-points of shared vehicles	Serve as dependent variable	# of trips	Dashboard deelmobiliteit by CROW
Supply of shared vehicles	Control for the supply	# of vehicles available for hire	Dashboard deelmobiliteit by CROW
Seasonal effects	Control for seasonal effects	Weather readings	KNMI weather station 290 (Twenthe)
Days of the weeks	Control for commuters in weekdays and tourists in weekends	Date	N.A
Events	Control for special events such as public holidays or football matches	Date	N.A

The resulting dataset spans from 01/01/2022 to 30/04/2023 (n = 484) for Kennispark and from 01/01/2022 to 31/05/2023 for the rest of Enschede Noord.

5.3.2 Cross-validation technique

To asses the accuracy of the resulting model a k-fold cross-validation technique is used. This validation technique entails training many different XGboost models. All models get to use a random selection of 80% of the data for training, for the remaining 20% the trained model must then make predictions. This is done k-folds resulting in k models. Then, after 1000 folds their are about 200 prediction for each day [8] These predictions are averaged per day and compared to the actual number of trips and a measure of model performance such as RMSE is used. This process is schematized in Figure 11, and the actual process can be coded into R.

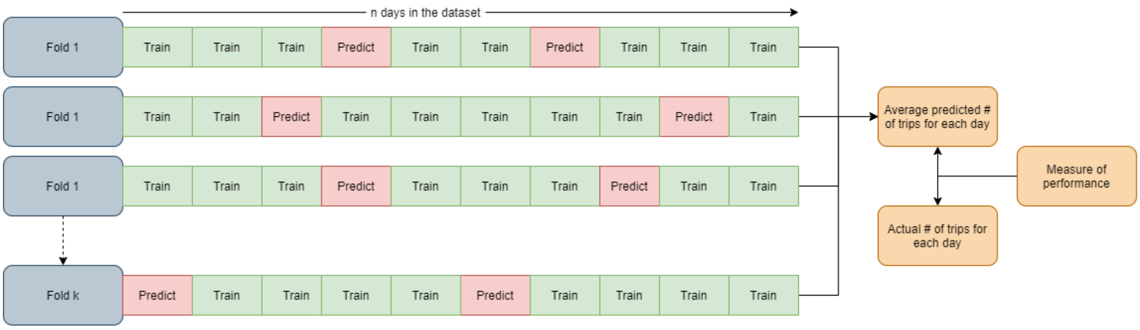

Figure 11: *Schematization of the cross-validation technique used to asses XGboost model performance*

[8]Since 20% of the data is selected for validation there is a 20% chance of selecting a day in a fold which has already been predicted. Then after 1000 folds there are: $1000 * 0.20 = 200$ predictions for each day.

5.4 Computing the impact of hubs on the use

In order to make final conclusions we use all previous intermediate results produced by the previous methods. To do this we first need to add an additional assumption.

Assumption

- Users park their vehicle at the mobility hub which is closest by their end-destination.

This assumption is fundamental to the analysis; To fairly compare the use of shared mobility not only on a broader level in Kennispark and Enschede-North, but also on a mobility-hub level we need to determine the count of users of a mobility hub, before the hub was implemented. This apparent time paradox can be easily solved however by the assumption, since it entails that the users of shared mobility before the mobility hub would have parked their vehicle at the mobility hub that is closest by their end-destination. I.e. a user who parks their vehicle within the service area x of hub y before the hubs were implemented would have theoretically at the mobility hub because this hub is the closest to their end-destination. This notion is schematized in Figure 12.

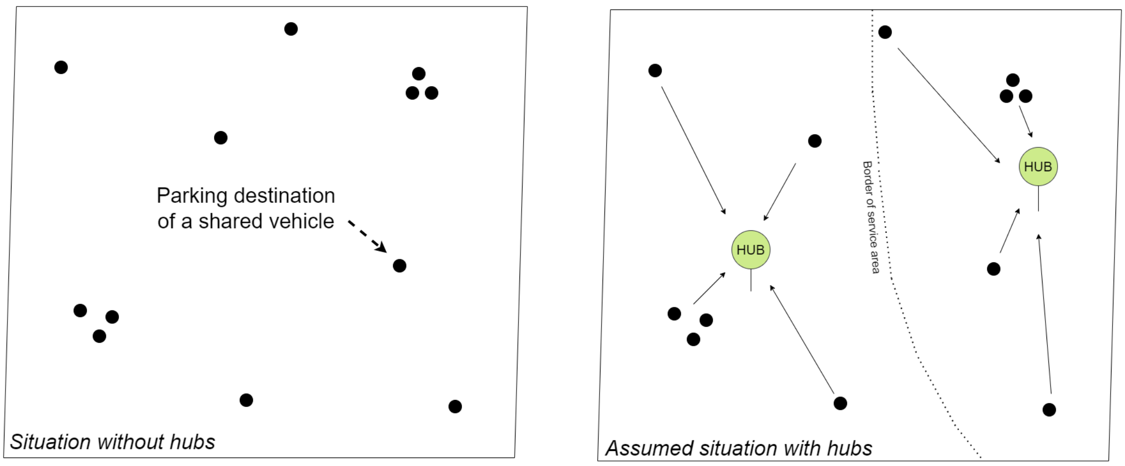

Figure 12: *Users would, with the assumptions, park their vehicle at the hub closest by their end-destination. This method allows us to compare the use of shared mobility before and after the implementation of hubs on a hub-by-hub basis.*

With the assumption the we can count the number of users for a hub before the hub was implemented. Then the independent variables described in the previous subsection can be collected in order to train the XGboost model for each individual hub/service area. This model can then be used to predict the # of trips in the period from the date when the hubs were implemented to the current date. Since the model was trained on data before the hub was implemented, it will make predictions for the use per hub **without** the effect of hubs accounted for. What is accounted for is all other independent variables such as supply, weather and holidays. Subsequently the actual amount of trips **includes** the effect of hubs. Therefore, the difference between the two is the effect of hubs, which will be our key performance indicator for the rest of the analysis (KPI). Usage will be measured in average usage per day , because Kennispark en Enschede North have different moments of hub implementation and absolute values are therefore of no use. For example, it is difficult to compare usage in Kennispark to Enschede North if Kennispark has usage counts for a year, and Enschede North only has counts for a month.

After our KPI has been calculated per hub several other statistics can be computed. First of all, the estimates and importance of the distances to POIs as well as the number of POIs can be derived from a regression model (estimate) as well as from a XGBoosting (importance). Additionally, further variables such as number of inhabitants, gender ration and age ratio are added.

The data necessary for the analysis is summarized in Table 4.

Table 4: Data overview: Comparing use-data

Data description	Use case	Data type	Source
Actual # of trips counted after hubs	Comparison	Count numbers	Dashboard deelmobiliteit
Predicted # of trips counted	Comparison	Counted numbers	XGboost model
POI distances and counts	Derive impact of POI variables	Distance in meters and count	POI results
Demographic variables	Derive impact of demographics on KPI	Postcode 6 data	CBS
Actual # of trips before hubs	Derive impact of number of trips before	Counts	Dashboard deelmobiliteit

The time frame for the trip count data is from 02/11/2022 until 31/05/2023 for Kennispark and from 06/06/2023 until 27/06/2023 for Enschede Noord.

6 Results: Service area analysis

The service area analysis was done with the assumptions and data described in Section 5.1. The analysis was performed in ArcGIS, by performing a service area analysis on a network. This section will serve to showcase the process and the intermediate results.

6.1 Adjustments to the network

An open source data set published every month by the Dutch Ministry of Infrastructure and Water Management, which receives the data directly from road administrators like Rijkswaterstaat ("Dataset: Nationaal Wegen Bestand (NWB)", 2023) is used as a basis for the network. This network is mostly complete, but still some adjustments need to be made. The used data set only contains official roads and pathways that are used by road administrators, which means that only roads that are accessible by car or are dedicated bike paths are registered. Whilst it is assumed that all these roads are also accessible by pedestrians (by means of sidewalks), it cannot be assumed these are the only roads accessible to pedestrians. The data set does not model walking paths in parks, squares, and sidewalks going between buildings. To compensate for this discrepancy between the data set and the actual situation the network was manually edited in order to add all missing paths. Despite carefully exploring satellite imagery to identify possible walking paths, it is likely that some have been missed or that new paths have been created since the date of the satellite imagery. Figure 13 shows the original network as taken from "Dataset: Nationaal Wegen Bestand (NWB)" (2023) together with the additions made.

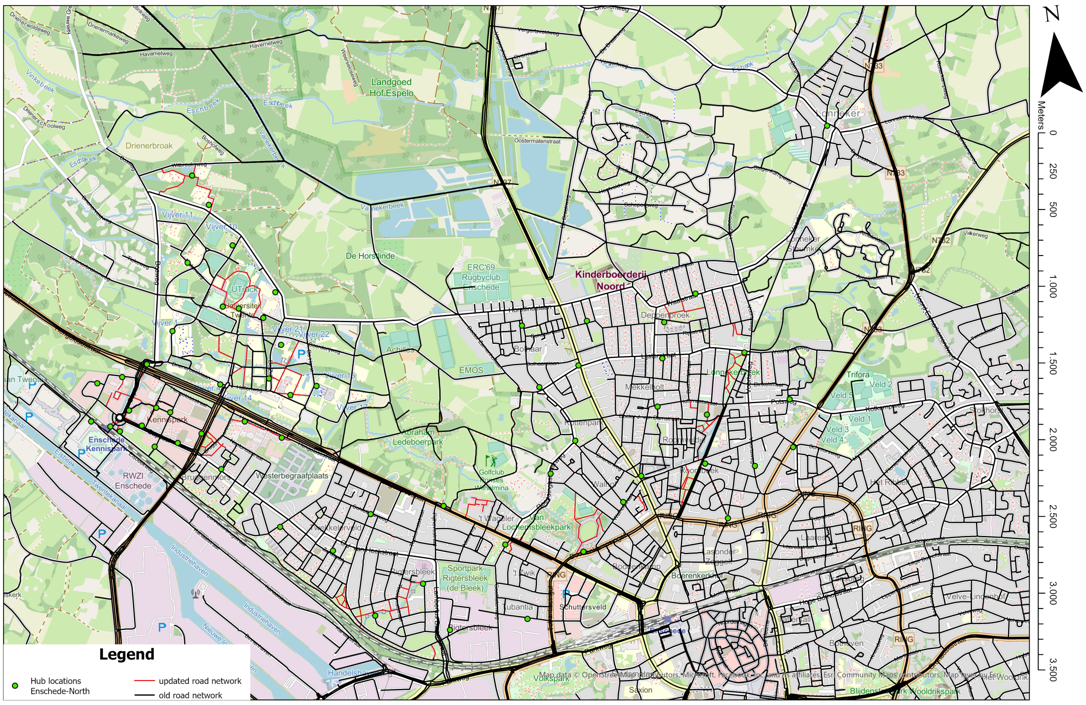

Figure 13: *The undirected network of walking paths used for the network partitioning*

6.2 Service areas

After the network has been validated it can be partitioned. This is done using ArcGIS's own 'service area analysis layer', which partitions the network based on distances from the hub locations. The results of this partitioning can be seen in Figure 14. The municipality's goal is to create a area-covering network with maximum walking distances of 300 meters (Goossens, 2022). This distance is therefore also the edge of the service area, since users falling outside are not expected to use the hubs. As can be seen in the Figure, the municipality has created a network of hubs which cover the entirety of Enschede Noord. With a few minor exceptions here and there such as the western cemetery, which is not expected to be visited often by users of shared mobility.

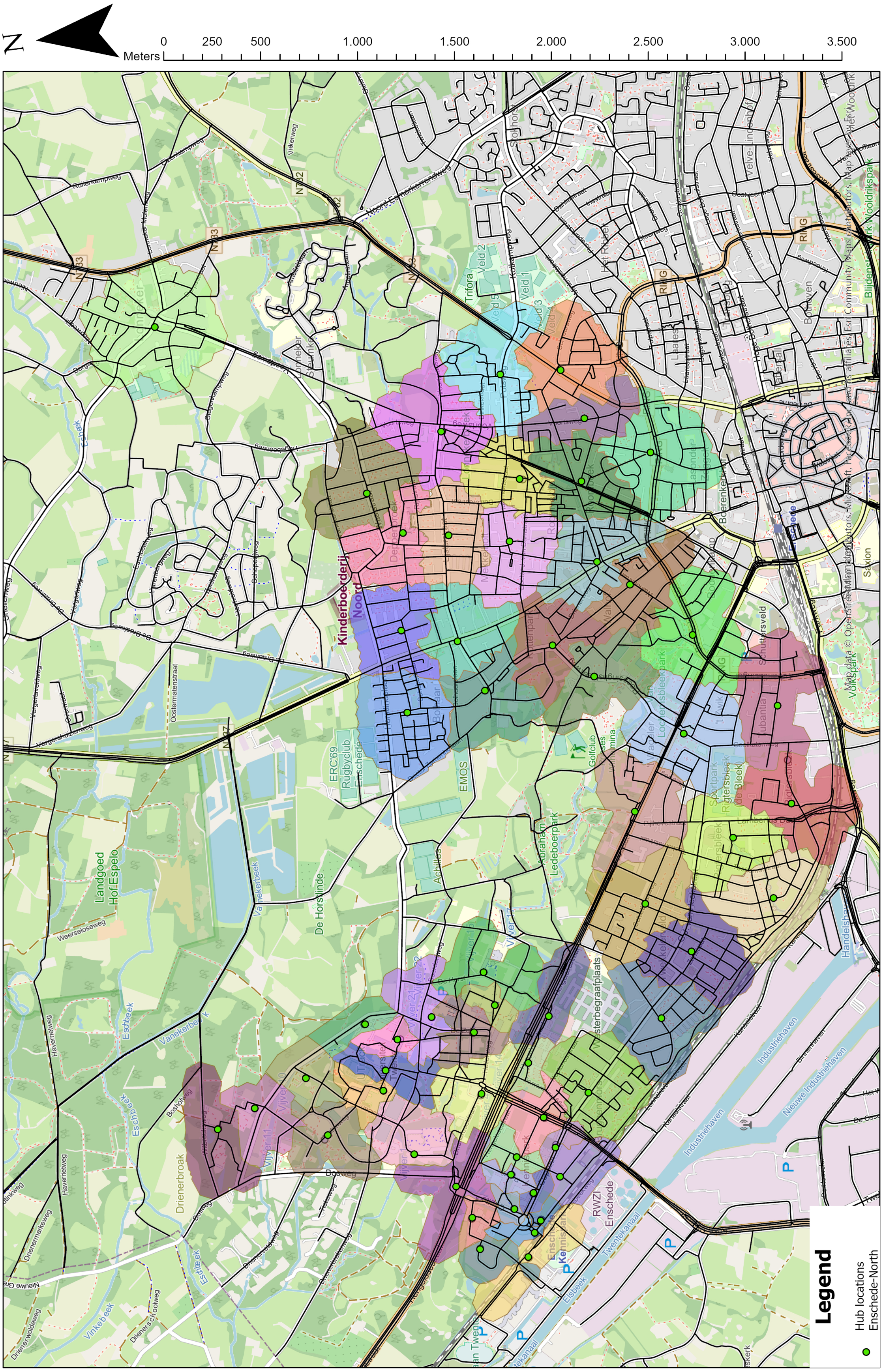

Figure 14: The service area of each mobility hub in Enschede-North.

7 Results: Determining relevant POIs

As described in the method, this results section will show three different ways in which relevant POIs were determined. Besides results, the method requires some intermediate results and decision-making in order to get to the final results. First, the relevant POIs were determined using literature and experts. Afterwards vehicle parking densities are used to visualise hot-spots and finally a quantitative analysis is done on the parking distances to POIs.

7.1 Relevant POIs according to literature & experts

Vianen (2022) describes anchor points as "places which are already embedded in a greater transport network" or places that "attract a certain amount of visitors/passenger on a daily basis." This the study does in the context of selecting appropriate locations for shared mobility hubs, stating that these locations are logical destinations for users of shared mobility to visit. This suggests that the described anchor points are also what this thesis would call 'relevant POIs'. In addition, Blad et al. (2022) in his thesis mentions that hubs should be places which do not only connect mobility related activities, but also connect shared mobility with social and commercial activities. Further expanding the definition of POIs in order to include places with social and commercial functions such as stores and parks.

Consultation with M.B Ulak and R. Goossens proved to lead to the same definition of a POI. Using these definitions a concrete list of POIs was set up and validated with the help of M.B Ulak and R. Goossens. The list of POIs categorized by function can be found in Table 5.

POI category	Example
Mobility related	Bus stop, P+R, train station
Parks	Greenery, playground
Shops	Supermarket, bike repair shop, clothes shop
Sport amenity	Football pitch, tennis court, fitness centre
Educational amenity	Primary-, secondary school, library, museum
Other amenity	Doctors, clinic
Work related	Offices

Table 5: Relevant POIs according to literature and experts

The above list includes most functions within a city aside from residential. Whether all these POIs are as relevant as others is doubtful; therefore, this nuance is further explored in the following subsections.

7.2 Relevant POIs according to hot spots

The data available gives the opportunity to derive relevant POIs based on usage. In addition to trip end points, a large set of POIs was plotted using data from Openstreetmap. The selected POIs are in accordance with Table 5. Figure 15 shows the result.

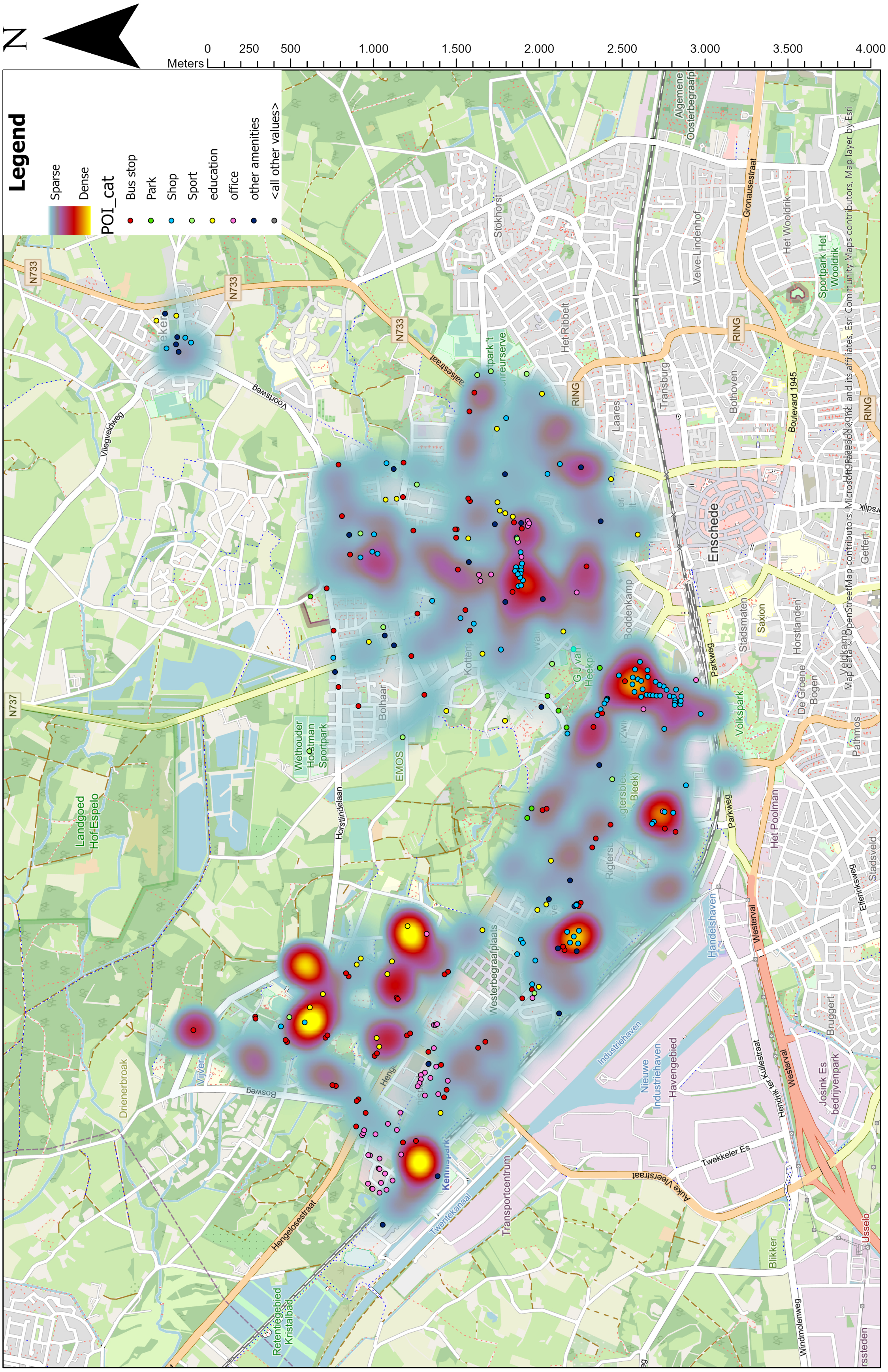

Figure 15: *Heat map of end-of-trip destinations of shared mobility*

All trips made after September 2022 have been deleted from the trip dataset, as this was the date at which the first mobility hubs were implemented. Mobility hubs do not allow users to park their vehicles wherever they want, making it harder to identify possible relevant POIs. What can be seen in the figure is that focal points such as *Winkelcentrum Deppenbroek, Roombeek* and *Woonplein* are clear POIs, with their many shops. Furthermore, areas around bus-stations and near train station *Kennispark* are more often than not popular destinations for trips with shared mobility services. Also among the most visited POIs on the heatmap are educational facilities such as the *University of Twente* (northwestern part of the heatmap). The interest of users to park close to other POIs seems lacking and is hard to discern from the heat map.

7.3 Relevant POIs according to parking distance to POIs

A quantitative analysis is also of value as it does not rely on visual inspection. As explained in the method, first a maximum parking distance must be determined in order to filter any trips which have an uncertain trip destination.

7.3.1 Approximating the maximum parking distance from a POI

Figure 16 shows the distribution of distances from trip end-points to POIs. Most trips end within 0-70 meters from a POI. A clear drop can be seen after the median of 72 meters. This indicates that this may be the cut-off distance at which point the POIs are not a reasonable destination anymore. We can validate this by looking at two samples, illustrated in Figure **??**. What can be seen is that there is indeed a clear cutoff point around 70-80 meters, at which it becomes ambiguous what the end-destination of the trip might be. At *Winkelcentrum Deppenbroek* all trips within 80 meters are clearly centered around the mall, further away it becomes harder to guess what the destination of the trip was. At the *Horst* there is a clear dividing line along the road running to the north-east. This makes sense, since this is also the line where the bike and moped parking for regular (non-shared) vehicles start. Therefore, it seems that 80 meters is about the maximum walking distance at which users park their vehicle. Before the analysis can be finished however, some adjustments to the POIs need to be made, this is explained in the next section.

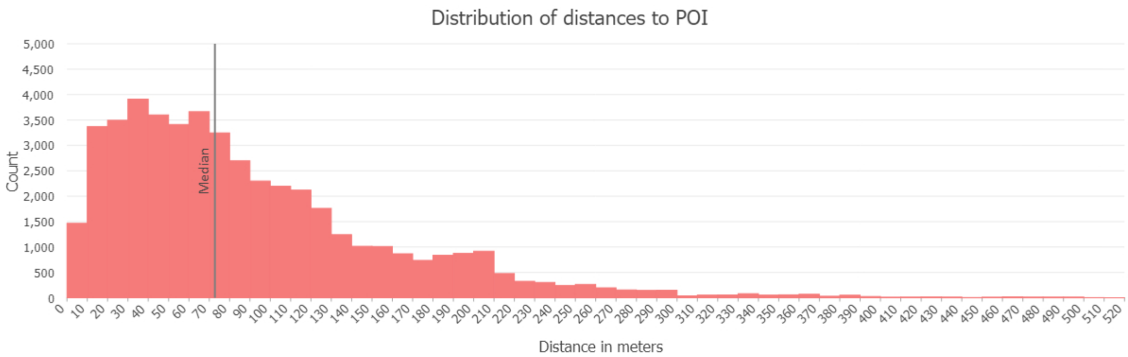

Figure 16: Distribution of parking distances from POIs

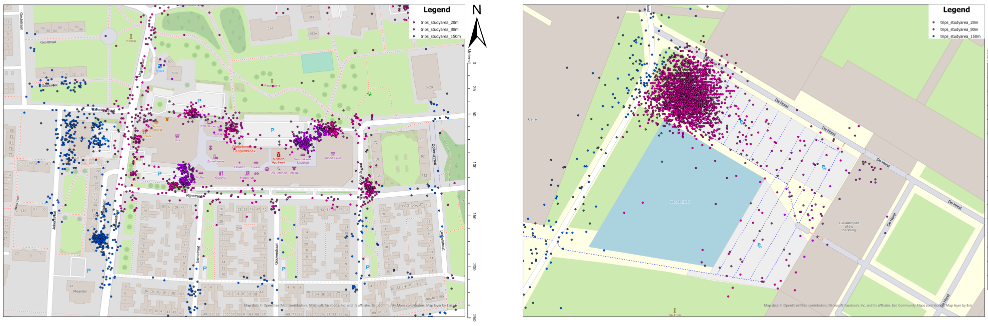

Figure 17: Shared mobility trip end-points at *Winkelcentrum Deppenbroek* (left) and at the *Horst, University of Twente* (right) categorized on distance from nearest POI

7.3.2 Adjustments to POI point data

There are some additional changes that need to be made before the analysis can be run. To show why, first take a look at Figure 18 which shows the counts of trips that ended within 80 meters of a shop, categorized by shop type.

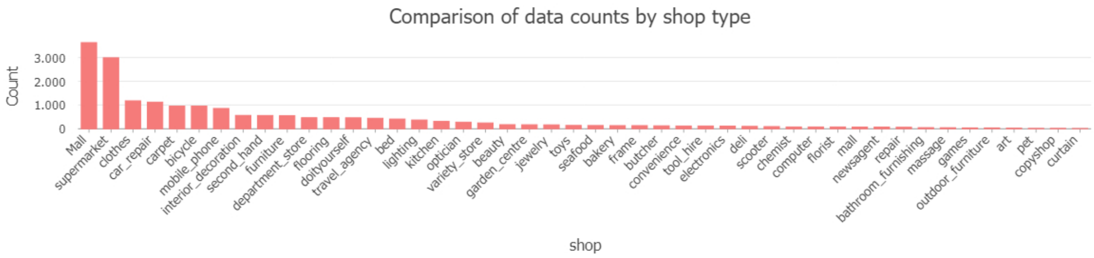

Figure 18: *Count of shop type closest to end-point*

Figure 18 shows that malls are the most visited type of shop, this is expected, since a mall consists out of a big number of shops in a small vicinity of each other, making it a attractive shopping spot. Afterwards come supermarkets, which is also to be expected since it is a shop type visited regularly by almost everyone. Some unexpected shop types are also visited frequently according to the Figure, like carpet and interior decoration shops. This however is not entirely accurate and is a downside of the POI map as it is now; these trips are most likely meant for another POI (in this case a McDonald's in *Woonplein*), but because there is no parking space for vehicles close to the other POI all the vehicles park in front of these POIs. This is illustrated in Figure 19, where it can be clearly seen that the logical place to park shared vehicles is closer to a carpet store and interior decoration shop than to the McDonald's.

Figure 19: *Woonplein streeview (December 2022, Google)*

Therefore, at locations where POIs are in close vicinity to each other the POIs are removed, and replaced by POI markers spread across the area, so that all trips ending in the area are linked to the area. The areas for which this is done are *Woonplein, Winkelcentrum Twekkelerveld, Brouwerijplein* and the shops along the *Hengelosestraat*. Hengelosestraat is not a shopping centre but is nonetheless selected because there are a lot of shops along this particular stretch of road, and there is no parking space directly in front of the shops the due to the narrow sidewalk. These locations are simply labelled as 'Shopping centre' (this is also done for *Winkelcentrum Deppenbroek*). The adjusted POI point data map can be found in Figure 20

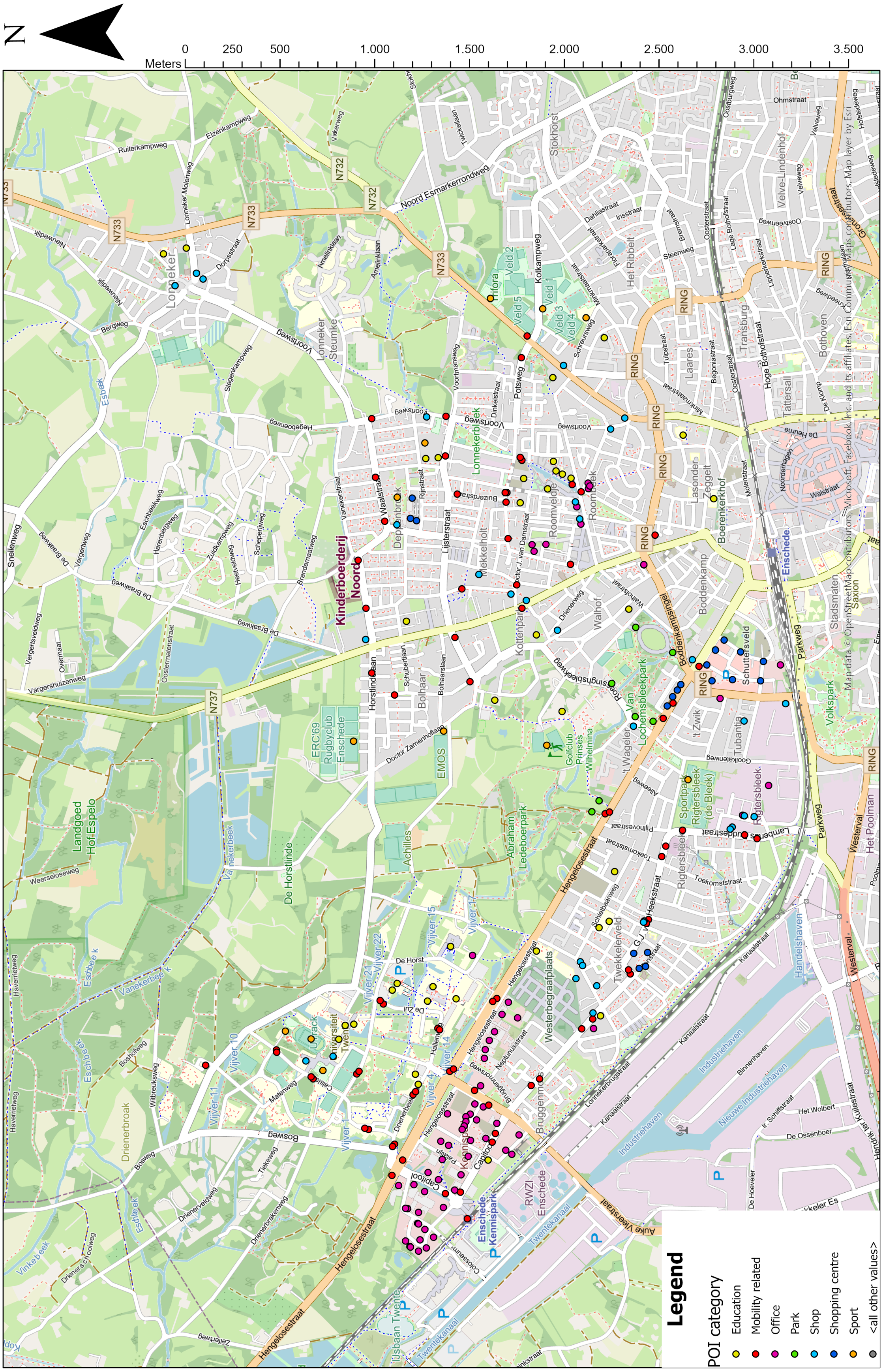

Figure 20: *Final POI map used*

7.3.3 Relevant POIs according to parking distances - Results

Now that everything is adjusted and accounted for the analysis can be executed. First, to ensure everything is in order, the shop type is again counted in 21. The Figure makes sense, as we expect shopping centres to be heavily visited, as well as supermarkets. Also car repair shops and bicycle shops make sense, as these are the type of locations people who are in need of a temporary vehicle travel to in order to retrieve their (repaired) car or bike. This is confirmed by taking a sample as seen in Figure 22. It is clear that these vehicles have parked at the location because the car repair shop is their end-destination, validating that the results in Figure 21 are accurate.

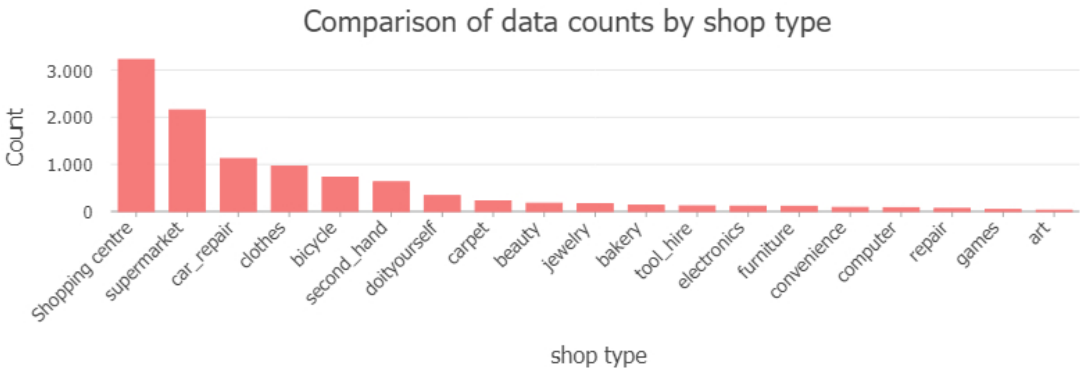

Figure 21: *Count of shop type closest to end-point*

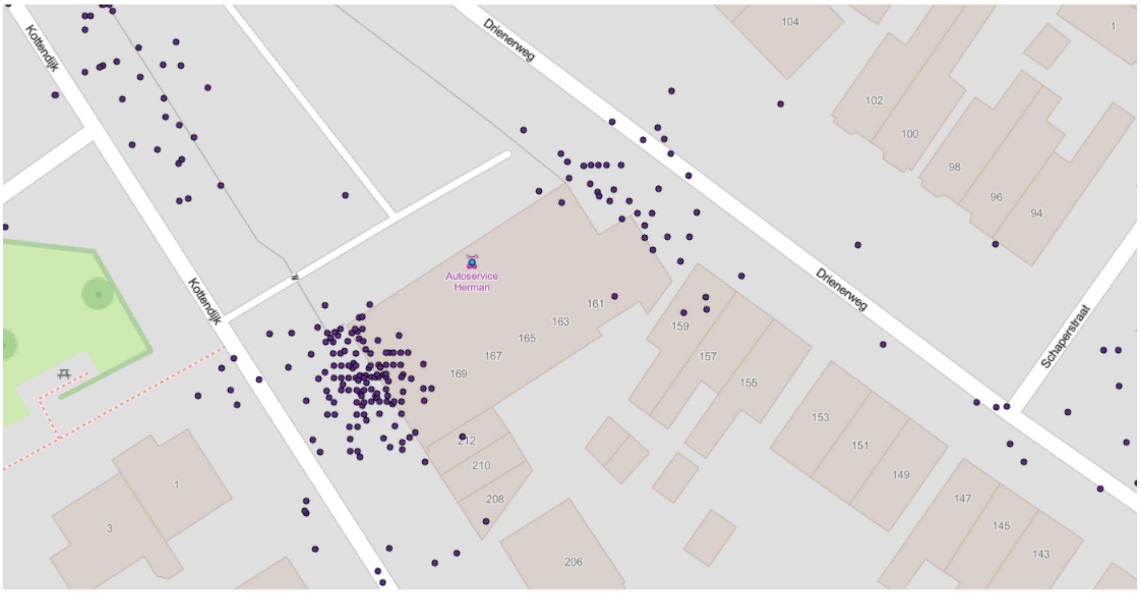

Figure 22: *trip end-points near a car repair shop*

Figure 23 shows the trip end-points within 80 meters of a POI, counted by POI category. Mobility related POIs like bus stops and train stations are the most visited category of POI, followed educational facilities, shopping centres and offices. Other POI categories follow behind. Sports and parks are relatively little visited compared to the other POI categories.

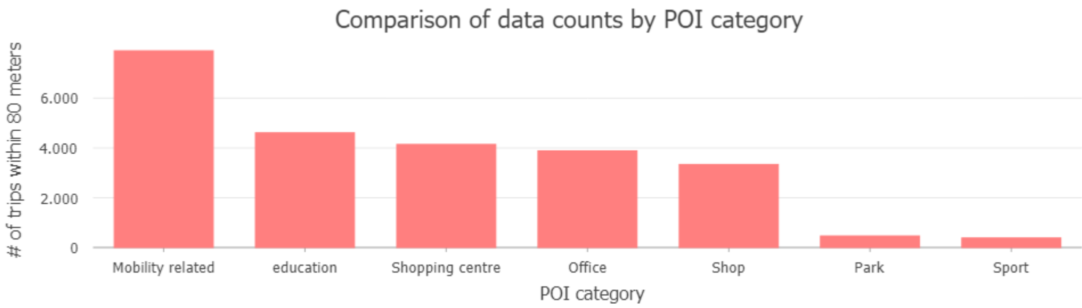

Figure 23: *Count of trip end-points by POI category*

A look at some examples in Figure 24 confirms that mobility related POIs are popular destinations for users of shared mobility. This is not only the case for large mobility nodes like train station Kennispark (top left) but also for smaller nodes such as bus stops.

Figure 24: Shared mobility trip end-points at various mobility related POIs, represented by the red markers.

7.4 Discussing shortcomings in the results

The three methods of identifying relevant POIs, based on experts and literature, hot spots and parking distances have yielded results which are useful for both this thesis' larger objective and as an intermediate product for policy makers. Therefore, this section is tasked with prematurely discussing possible shortcomings and errors in the results.

First, while the list of relevant POIs in 5 gives a concrete list of POIs relevant to shared mobility, it is also very broad in its scope. This results in a dataset which is prone to missing data. For this thesis, the described POI categories in 5 were extracted from Openstreetmap in order to serve as a base for further analysis in the following methods. However, quite a few manual adjustment to this initial data had to be made. For instance, important educational facilities [9] were displayed as polygons rather than as points. This required manual intervention, removing the polygons and placing points at the entrances of buildings. This has been accounted for in this thesis by checking entrance points on street view (using google maps).

[9] More specifically buildings at the *University of Twente*

The second method using the heatmap is a more empirical-based method of determining relevant POIs compared to the first method. For some POIs it is a clear indication that they are relevant, for instance focal points such as shopping centres and the *University of Twente* and to some degree bus stations. Whilst the method is suited for an initial exploration of relevant POIs, it is less suited to give definitive answers on relevancy of all studies POIs.

Finally, the method using parking distances from trip end-points to POIs is a quantitative way of analysing the relevancy of POIs. However, this also makes the method prone to error, since the parking behaviour of users is not always predictable and logical. In the method there has been an attempt to account for this unpredictability by approximating the maximum distance users park from a POI. Nonetheless, this maximum of 80 meters is very rigid, and is not a rule-of-nature. However, it is the best possible approximation that could be made for now. Also, some changes had to be made to the POI dataset generated by Openstreetmap, since the method wrongfully assumes the destination of trips when POIs are clustered close together. As a consequence the analysis is not able to distinguish between shop types in these areas.

7.5 Conclusions from the POI analysis

Based on the results of all three methods, the following conclusions can be made:

- Mobility related destinations such as train station *Kennispark* and bus stops are popular destinations. This suggests that shared mobility used in conjunction with the wider public transport network of Enschede, and confirms the notion also held by Jorritsma et al. (2021) that shared mobility is a viable last-mile-solution.
- Higher education facilities such as universities and colleges are often visited with shared mobility. This validates results by Miramontes et al. (2017) which indicate that a the main user group of (micro)shared mobility are highly educated, young adults. This demographic is also likely to visit offices in Kennispark, which is a big employer of recently graduated students.
- Park and sport destination are infrequently visited by shared mobility. An definitive reason why is uncertain. In parks the main activity undertaken is most often walking, making the option to first go somewhere on a vehicle and then walk perhaps a bit un-intuitive. For sports the same reason could give the answer; why take a (motor driven) vehicle somewhere when ones purpose is to do physical activity.
- The presence of end-destination such as car repair shops and bicycle shops suggest shared mobility are used for one-off trips where a user might only want to use a vehicle one-way (because they have their own repaired vehicle at the shop). However, the presence of other POI categories such as education and office suggest that shared mobility is used also as a commuter option.

8 Results: Predicting shared mobility usage with XGBoost

There is no use in comparing the usage of shared mobility before and after hubs if there is no control for other variables. This relates to sub question 3: *How do we fairly compare the use of shared mobility before and after the implementation of the hubs.* The usage of shared mobility is dependent on several variables which all vary on a daily basis, for example the weather and supply of vehicles. The next section will show the significance of these variables, whilst the section afterwards will use the variables in order to create and test a predictive model.

8.1 Significance of variables

All predictor variables were derived from their respective data source (see the method in section 5.3) and were used as independent variables in a regression model in R. Since we are using count data we can use a Poisson regression. However, a Poisson regression assume that the mean of the data is equal to the variance. The mean trips per day in our dataset is 1100 , whilst our variance is 445685. Thus, mean $\neq$ variance, and mean $<<$ variance. Therefore, a negative binomial regression is more suited because it does not make this assumption. Table 8 shows the list of coefficients for each dependent variable as estimated by the negative binomial regression model (n = 484 days).

Table 6: Overview of the predictor variables

Independent variable	Data type	Estimate	P-value	Significant (yes/no)
Supply	Count	1.7e-03	$<$ 2e-16	Yes
Maximum temperature	Continues	1.9e-02	0.011	Yes
Average temperature	Continues	-5.6e-04	0.95	No
Rain in mm	Continues	-1.2e-02	0.0011	Yes
Hour with rain max	Integer (0-24)	-1.e-03	0.54	No
Matchday	Boolean	1.4e-01	0.027	Yes
Weekday	Integer (0-7)	-1.2e-02	0.065	No
Holiday	Boolean	1.47e-01	0.026	Yes

The most important column is the 'P-value'. The P-value tells us whether an independent variable is statistically significant related to the dependent variable or not. More accurately, it tells us what is the probability of wrongly rejecting our null hypothesis (H_0) [10]. In this case H_0 = The independent variable has no effect on the amount of trips per day. The scientific norm is to accept variables as significant if this value is less than 0.05 [11]. As can be seen, the supply, maximum temperature, the amount of rain, and whether it is a matchday or holiday are all statistically significant variables to indicate the amount of trips. The other variables do not show P-values below our alpha value, which means that the probability of making a type I error (falsely rejecting H_O) is too large to accept

[10] In statistical jargon wrongfully rejecting the null hypothesis is also refered to as a Type I error
[11] The statistical jargon for this value is alpha value

the variable as significant. The 'Estimate' column signifies how the variable impacts the dependent variable [12]. In the case of significant variables, matchday and holidays have the most impact on the use of shared mobility. Afterwards come weather effects such as rain and maximum temperature. Last comes the supply of shared vehicles.

8.2 Cross validation of the model

The next step is to train the model and validate the accuracy using k-fold cross validation technique as described int he method. The results of following this method can be seen in Figure 25 and Figure 25. The first of which plots the actual and predicted number of trips for each day in the dataset. The second figure plots each day in the dataset as a point. The red line in the figure represents the ideal situation in which there is not difference between the predicted value and the actual value.

The RMSE [13] of the model is 278 trips per day, on an average of 1100. However, whilst this seems quite a large uncertainty, the 90% confidence interval showcased later in the results will indicate it is acceptable for the purposes of this model.

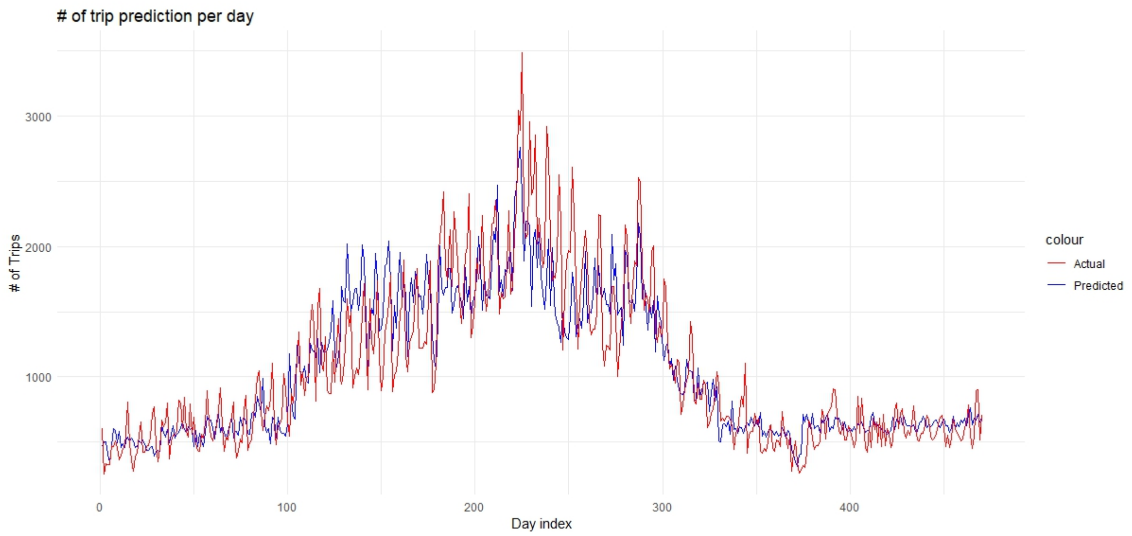

Figure 25: *Actual and predicted # of trips for the training dataset*

[12] A more accurate description of the estimate is that it is the models best estimate for the slope of the variable

[13] The root mean squared error represents the average error for each day. Say that the model predicts 1000 trips with an RMSE of 100, then we can expect the actual number of trips to be 900 or 1100 trips

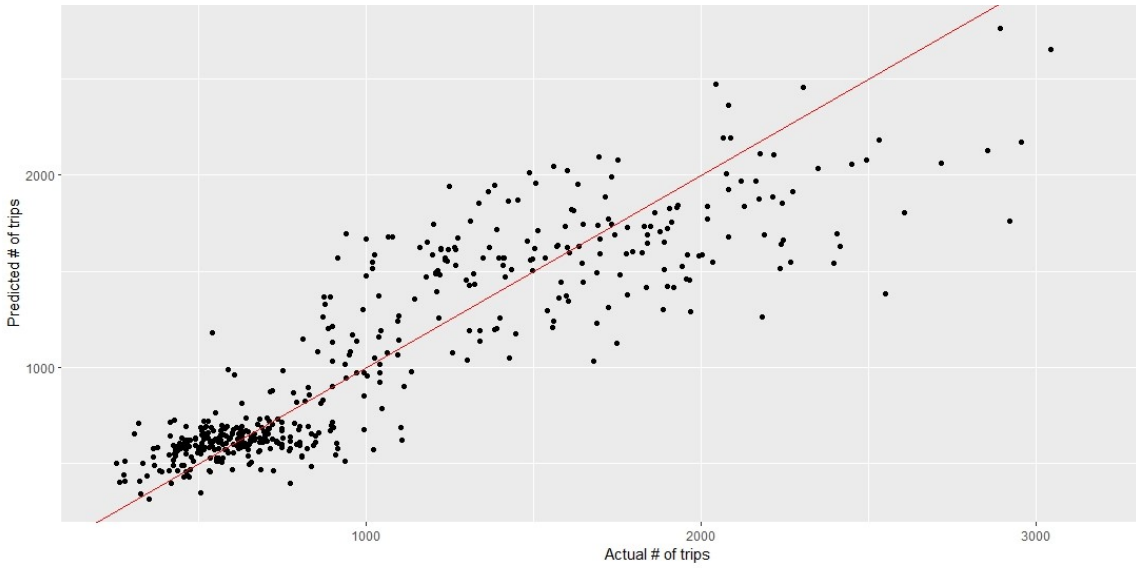

Figure 26: *Actual vs predicted # of trips for the training dataset*

As mentioned in the method, the XGboost model can also showcase the importance of each variable. The importance of every variable in the cross-validation was averaged over the number of folds. The results can be seen in Table 7. The gain (importance) represents how much the independent variable contributed to the decision making in the decision tree.

Table 7: Overview of the predictor variables

Independent variable	Gain (Importance)
Supply	0.73
Maximum temperature	0.089
Average temperature	0.084
Day of the week	0.033
Holiday	0.027
Rainfall in mm	0.018
Hour with rain max	0.012
Matchday	0.0043

9 Results: Impact of hubs on the use of shared mobility

After the predicted amount of trips and the actual amount of trips are compared the impact of all other possible independent variables is controlled for. The remaining change in usage is therefore only possible to be caused by the impact of the hubs. Figure 29 shows the average number of trips per day before the hub (top left), the average predicted number of trips in the present day without the effect of hubs (top right), the actual number of trips in the present day (bottom left), and the difference between the two (bottom right). The last mentioned map (bottom right) contains the data which represents the effect of hubs on the average daily usage of shared mobility. Figure 30 shows the same map, only then the resolution is scaled up to the city regions Kennispark en Enschede Noord, and the effect of hubs is measured in percentages.

The blue numbers in Figure 29 represent the facility ID, and correspond to the facility IDs in Figure 27 and 28. Figure 27 and Figure 28 shows the predicted and actual usage values and includes the 90% confidence interval of the predicted value. Meaning that within this range we are 90% confident the model predicted the number of trips correctly [14]

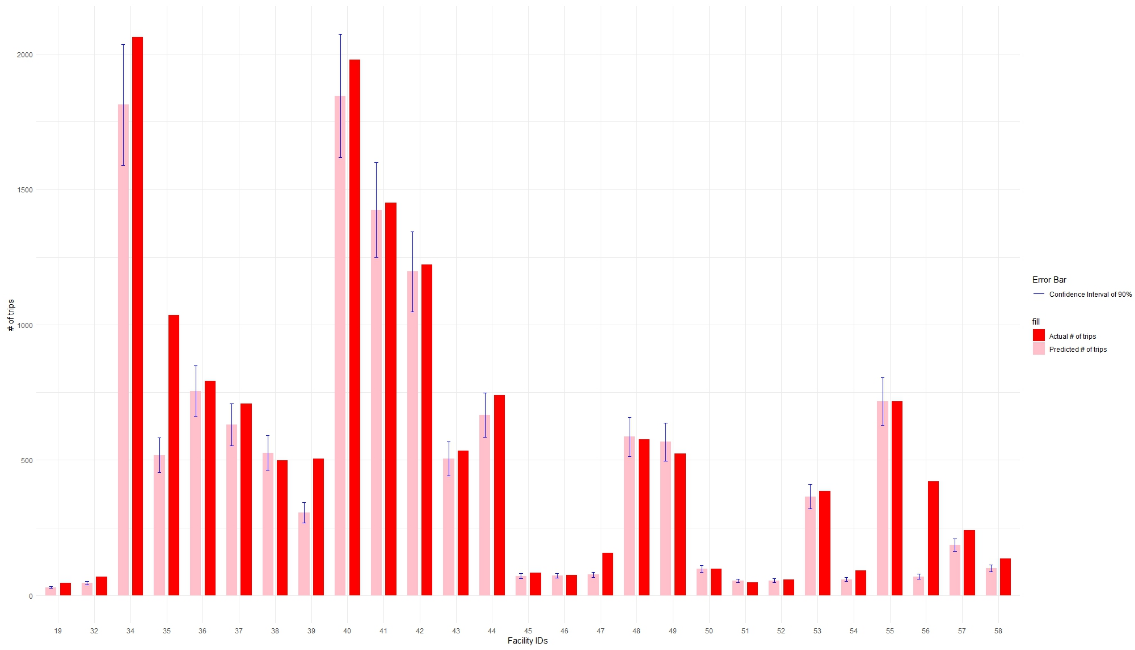

Figure 27: *Actual and predicted values per service area for hubs in Kennispark*

[14]'Correctly' might be a bit confusing for an uninvolved reader, and even for an involved one is rather hard to grasp. It means that, if no hubs had been implemented and the data been collected as usual, and the cross-validation technique used previously in this thesis had been used again, then 'correct' refers to the predicted number of trips being equal to the actual number of trips

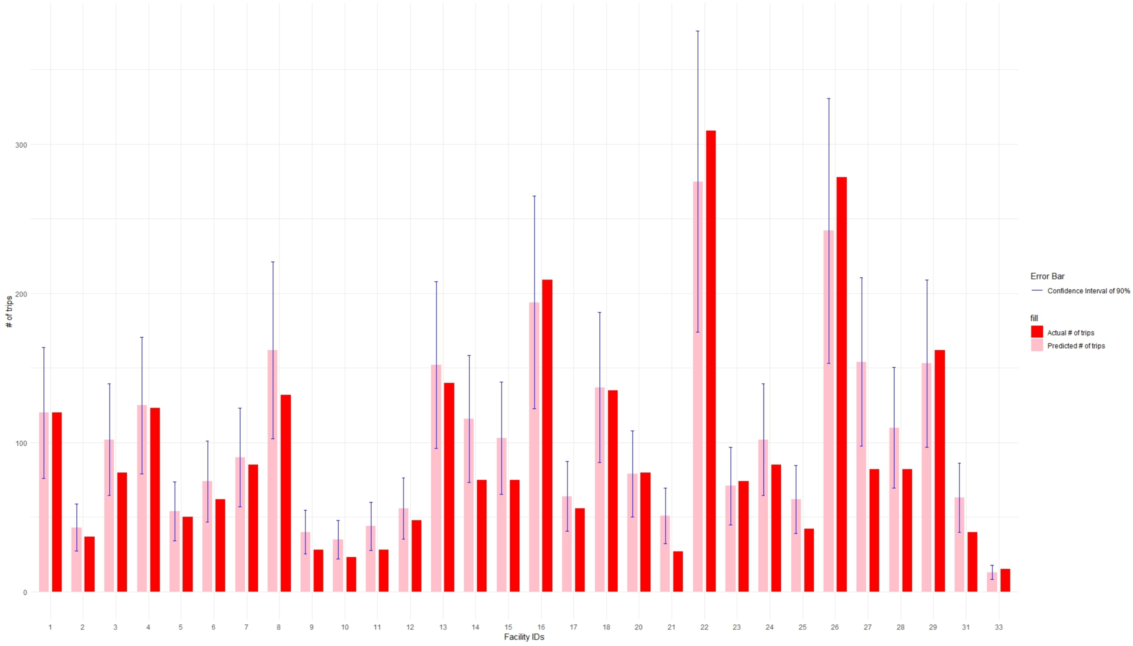

Figure 28: *Actual and predicted values per service area for hubs in Enschede Noord*

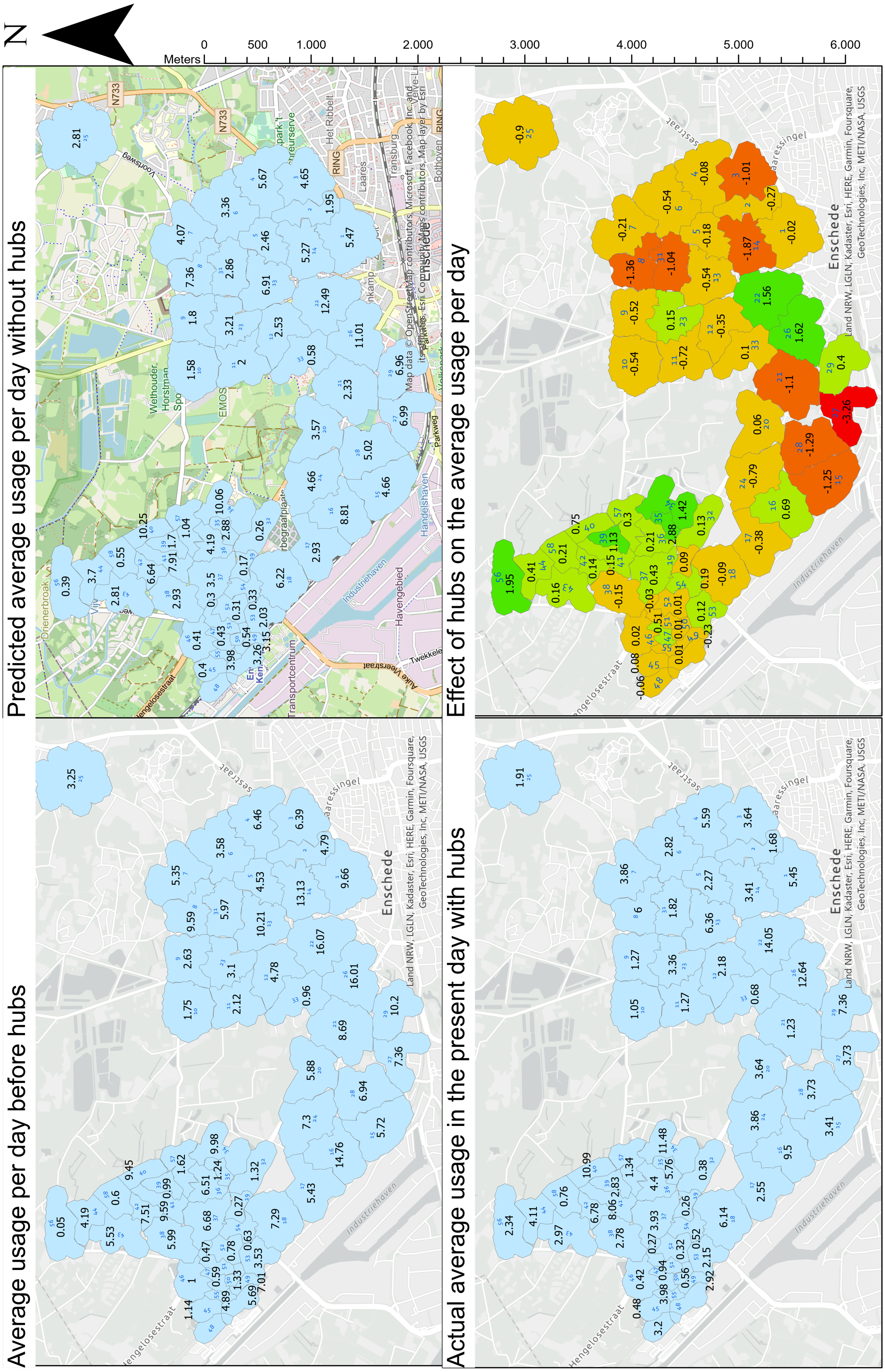

Figure 29: The impact of hubs on the usage of shared mobility on hub scale

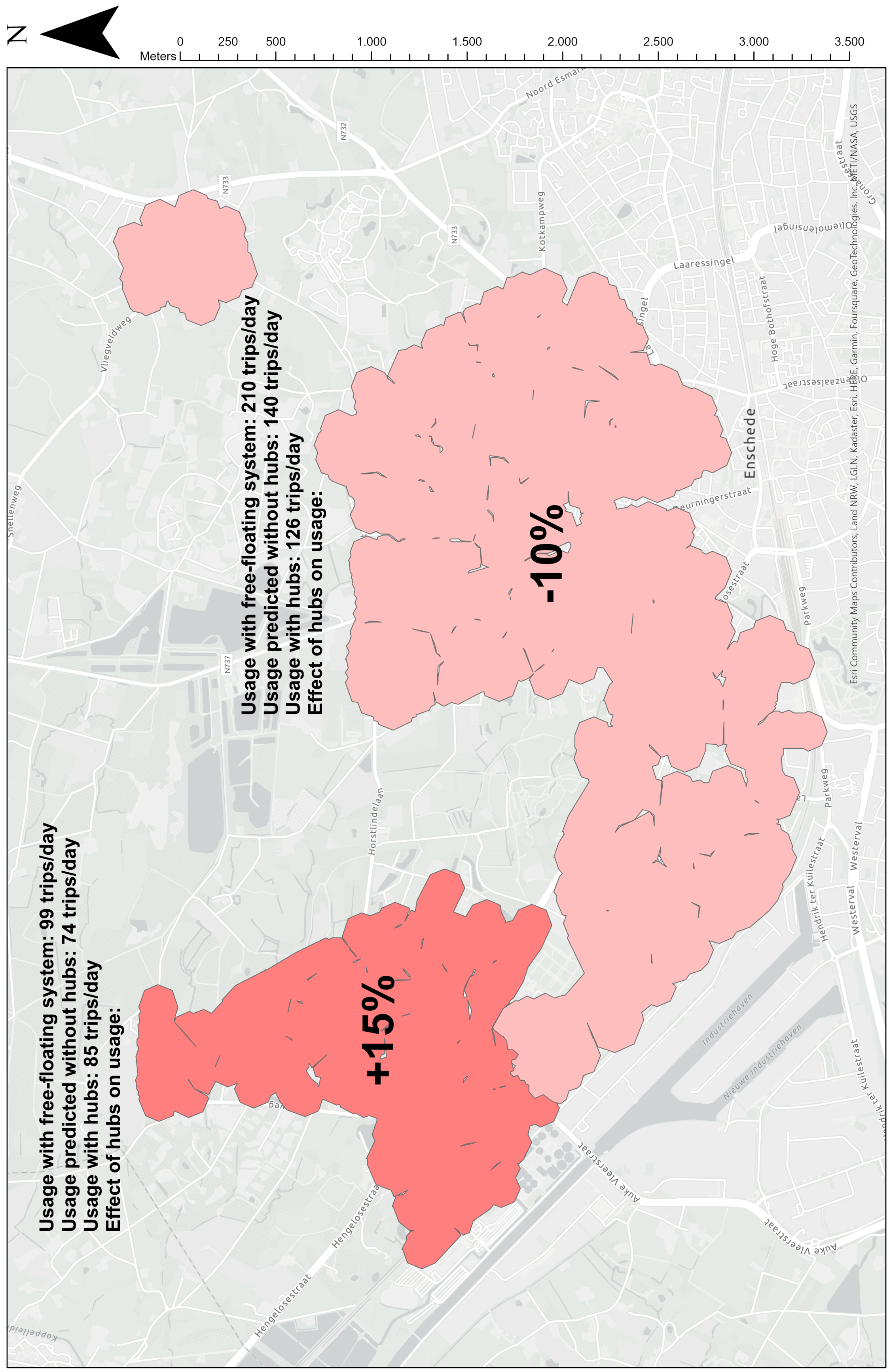

Figure 30: *The impact of hubs on the usage of shared mobility on city region scale*

The next step in the analysis is to get the individual contribution of individual variables on this difference in usage. **The uncertainty in the predictions for Enschede Noord make it irresponsible to do further regression analysis on variables. Therefore, further analysis is done exclusively on the data from Kennispark, which is considerably more reliable**. The reason why predictions for Enschede Noord have such an high uncertainty is due to the small number of days for which measurements could be done (22 days in total. Subsequently, this means that the model has only made predictions for 22 day, leading to outliers in the data having a larger influence on the RMSE, which translates to larger confidence intervals.

A large set of variables were measured for each service area. However, analysis proved that there was high multicollinearity [15] between some variables. In order to asses the degree of multicollinearity the variance inflation factor (VIF) is calculated. This factor measures how much the variance of estimated coefficient of the variable is inflated due to its correlation with other factors (Murray et al., 2012). A value of 1 means there is no multicollinearity, a value of 5 or less is usually considered acceptable. Figure 31 shows the VIF values for all measured variables, with a dotted line at the value of 5. As can be seen, a large group of variables exceed an VIF value of 5. The rational for correlation between some variables can be explained: The total number of POIs is obviously related to the number of POIs for each category. Also the number of POIs is somewhat related to the distance, since a higher number of POIs suggest a higher density of POIs, resulting in decreased distances. Other correlations such as those to the number of inhabitants have no theoretical connection this thesis could identify. After consideration all 'Distance to' variables were removed as well as the total number of POIs and the number of inhabitants. Figure 32 shows the VIF values for the remaining variables. The remaining variables all show some degree of multicollinearity, but since there is not theoretical explanation as to why, and because the value is well under the critical value of 5, the degree of multicollinearity is acceptable.

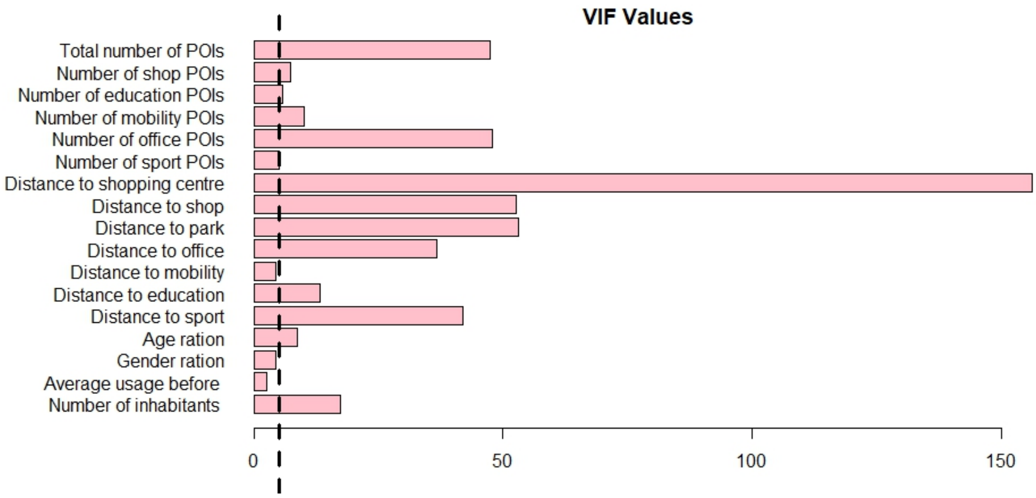

Figure 31: *VIF values for all measured variables*

[15]In simplified terms multicollinearity means that a variable shows a strong statistical relationship with one or several other variables. These variables cannot be included in a regression analysis because of several reasons such as unreliable results and extreme sensitivity of the model

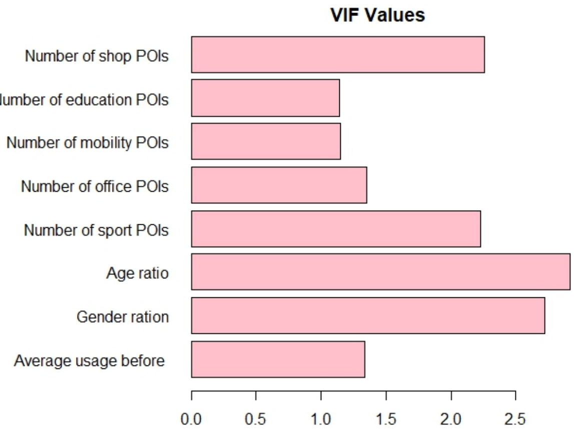

Figure 32: *VIF values for remaining variables*

Table 8 shows the result of a multi linear regression model and an XGboost importance model (n=1000 iterations). In these models the effect of hubs served as the dependent variable (bottom right in Figure 29). The meaning of all the columns has already been explained in section 8.1.

Table 8: Overview of variables in regression analysis and XGboost importance (n=1000 iterations)

Variable	Estimate	Gain (importance)	P-value	Significant yes/no
Average usage before	-0.05117	0.43	0.2577	No
Gender ratio	-0.01363	0.12	0.9630	No
Age ratio	0.15412	0.064	0.4138	No
Number of sport POIs	-0.38485	0.010	0.5148	No
Number of office POIs	-0.04423	0.037	0.3539	No
Number of mobility POIs	-0.22011	0.087	0.0438	Yes
Number of education POIs	0.39333	0.26	0.0216	Yes
Number of shop POIs	0.15818	0.004	0.6163	No

10 Discussion

Figure 29 shows mixed results. Some areas see a positive impact of changing to the hub system on the usage of shared mobility, whilst other see an negative impact. A larger resolution in Figure 30 indicates that there is a difference between Kennispark and Enschede Noord; the hubs have, overall, increased the usage of shared mobility in Kennispark whilst it has decreased in Enschede Noord. Zooming back into the hub resolution in Figure 29, especially hubs on the campus of the University of Twente have seen a significant increase in usage due to the hub system. In the business-park side of the area (south) the impact of the hub is lessened. The increase in usage of 15% caused by the hubs partly contradicts the suspicions by van Steijn et al. (2022), in which users indicated that the hub system decreased their willingness to use shared mobility. Whilst this suspicion has not manifested itself in Kennispark it has in Enschede Noord, in which the hubs have caused shared mobility usage to drop by 10%. There are some uncertainties in the results caused by the model. These uncertainties were communicated through the use of confidence intervals in Figures 27 and 28. These figures indicate that the results about Kennispark are a lot more reliable that the results in Enschede Noord. Therefore, it is hard to draw reliable conclusions from the results for Enschede Noord. The unreliability of the results is higher for Enschede Noord than for Kennispark because of the limited data available for the area. The hubs were implemented in June 2023, giving this thesis only about a months worth of data. This leads to the larger confidence intervals and larger uncertainties. However, there is the possibility that the implementation of hubs requires a certain warm-up period in which time users have to acclimate to the new system, as well as in which time new users need to discover the system (R. Goossens, personal communication 04/07/2023). Therefore, there is a possibility for the trip counts in Enschede Noord to increase in the future.

These results are meant to give some answers to one of the main research questions, *"what is the impact on the use of shared mobility when there is a change from a free-floating system to a hub system?"*, however, an definitive answer remains elusive. Literature review in section 3 indicated the same trend in literature: there is little to no consensus about whether or not the change to a hub system impacts usage positively or negatively. This thesis is in line with the notion that indeed, the answer is not that simple. As varying results per area indicate that assessing hubs outside their spatial context is not enough.

This thesis had tried to anticipate this conclusion by including the hubs relation to POIs. Several variables were included in the initial analysis, but as seen in Figure 31 and 32 not all could be included due to multicollinearity. Nonetheless, some representative variables could still be used. The number of POIs categorized as 'mobility' and 'education' in a service area are the only two significant variables in the regression. This is not surprising and in-line with previous results shown in section 7, in which these two categories of POI were identified as the most visited destinations. The number of mobility related POIs, such as bus stops, in a service area have a negative coefficient, meaning that the difference in usage caused by the implementation of hubs is more severe in areas which include more mobility related POIs such as bus stops. In other words: The more mobility related POIs in an area the more the implementation of hubs will negatively impact usage. The opposite is true for the number of educational POIs: The more educational POIs in an area the more the implementation of hubs will positively impact usage. An important note here is that these coefficient implicate a relative impact on the effect of hubs, they do not say all about the absolute effect of hubs. For example, implementing a hub near an education POI will not necessarily lead to an increase in usage. Rather, the usage might still decrease, but it will decrease to a lesser degree compared to a scenario without an educational POI.

The positive coefficient attached to the number of educational POIs can be hard to explain. The results show that the amount of educational POIs near a hub contribute positively to the use of shared mobility. This suggest that the main user group of shared mobility, young adults Li and Kamargianni (2018), remain loyal to using shared mobility despite other suggestions that the hubs decrease the attractiveness of shared mobility (van Steijn et al., 2022) . However, this does not yet explain the hubs contributing to the average daily usage near educational POIs. A possible theory is that the hubs in this area where placed especially well, and in some cases even decreased walking distances in those cases where the hubs were placed closer to the entrance than the usual parking places where shared vehicles used to park. However, manual investigation of the hubs on the campus showed that this was the case nowhere. Another theory is that the positive coefficient can be attributed to demographic characteristics. For example, young adults are generally known to be more adventurous and open to more physical activity (Vianen, 2022). This may lead to this demographic being impacted less by the increased walking distances whilst still being impacted positively by those things that hubs add, such as recognizability of the hubs in their design and the reliability of finding a vehicle.

The negative coefficient of mobility related POIs can be explained by the fact that the free-floating system was the ideal candidate for a last mile solution (see section 3). With the implementation of hubs this role has been lessened, leading to a decrease in usage for users which intend to use shared mobility in conjunction with other public transport services. This decrease is therefore reflected in the negative coefficient.

The regression analysis on the relation between the impact of hubs and the POI did not include data from Enschede Noord. Therefore, the question remains if the same results that were found in Kennispark would have been found in Enschede Noord. Enschede Noord has few education POIs, the only ones being things such as secondary schools (which are little used by users of shared mobility, as seen in Figure 15), libraries and museums. Therefore, the larger decreases in use caused by the hubs do not contradict the fact that the number of educational POIs near a hub contribute positively to the use of shared mobility. Also, the area has its fair share of mobility related POIs. Thus, the previous conclusion that the number of mobility related POIs contribute negatively to the use of shared mobility is also not contradicted. However, whether these conclusions are confirmed in Enschede North also is not possible to say.

11 Conclusion

Shared mobility hubs as discussed in this thesis are becoming a broadly used concept not only in Enschede, but soon in the entirety of the Netherlands. Little is known on how the hubs impact the usage of shared mobility, and whether their spatial context is of any importance. Therefore it was this thesis' objective to asses the impact of shared mobility hubs and their relation to POIs on the use of shared mobility. This was done by manner of a case-study in Enschede, where hubs were recently implemented in favor of the free-floating system.

Whilst literature both suggested that the implementation of hubs will increase (due to the increase in recognizability and reliability in finding a vehicle) as well as decrease (due to walking distances theoretically getting larger), there was little empirical evidence for either yet. The results in this thesis provided empirical evidence. The hubs have a significant impact on the usage of shared mobility, but outside of their spatial context it is difficult to say whether or not the use of shared mobility will increase or decrease. In this case-study the hubs were shown to positively contribute to the usage of shared mobility in Kennispark by 15%, whilst the hubs contributed negatively to the usage in Enschede Noord by -10%. The data collected in Kennispark is more reliable then the data collected in Enschede Noord due to the limited time frame in which hubs were active in Enschede Noord. Also, because the measurement for Enschede Noord only included measurements right after the implementation of the hubs, it is possible that the results were inaccurate due to there being a warm-up period in which users have to acclimate to the new system. Therefore, there is reasonable prove within the results that hubs have an overall positive impact on the usage of shared mobility.

The spatial context of the hubs was assesed using the number of POIs in the service area of the hub in a multi-linear regression model. The results found that mobility related and educational POIs have a statistically significant impact. Hubs close to educational POIs are positively impacted by the change from a free-floating system to a hub system. Whilst there seems to be a correlation here nothing can be said about whether they are causally related, this thesis proposes some theories. The young-adults who visit educational institutions rely more (this demographic usually has no other motorized vehicle at their disposal) on shared mobility. Furthermore, this target demographic is more willing to accept the increased walking distances, which is the main disadvantage of the hub-system compared to the free-floating system. This makes this demographic more susceptible to the positive effect of the hub system such as recognizability and reliability. Mobility related POIs such as bus stops have a negative impact on the hubs when changing from a free-floating system to a hub system. A theory is that this is because the free-floating system was the ideal candidate for the last-mile solution. Now that the hub system has reduced the use fullness of shared mobility to fulfill the role of last-mile solution, hubs which previously saw a large percentage of their users coming from/to public transport services are more heavily impacted by the change.

Because the results indicate that education- and mobility related POIs are important destinations, in addition to being sensitive to the change from a free-floating system to a hub system it is this thesis' recommendation to prioritize and minimize the walking distances between hubs and POI related to (higher)education and mobility. This is especially of importance if shared mobility is to be a 'last mile solution', since results indicate that public transport users are using shared mobility less due to the change from a free-floating to a hub system.

Whilst this thesis has done its best to draw conclusions from the data and results some limitations are in place. First of all, the model which accounted for other variables such as seasonal effects, supply, and other variables is not 100% accurate, leading to some uncertainty in the model results. Lastly, the scope of the thesis was only limited. Meaning that the conclusions made in the context of Enschede are not certain to hold in other hub networks, cities and cultures. Therefore, it is this thesis' recommendation to replicate similar studies in order to see what the effect is on the use of shared mobility from a free-floating system to a hub system. Also, this thesis suggested that there might be a warm-up period between the transition from a free-floating system to a hub system, in which users have to get used to the change. This would be another possible avenue for further research, and can already be done with the current data used in this thesis. Additionally, this thesis would suggest doing an in-depth case study for a hand full of hubs instead of a 'broad' study looking at an entire network. This way more subtle results which data cannot show might come to light. Furthermore, whereas this thesis focused more on the impact of the hub system as a whole, more research could also be done regarding how other aspects of the hubs influence behaviour and usage of shared mobility. Think of this such size, visibility and recognizability of the hub. Finally, this thesis has stated some possible theories as to why mobility related POIs and education POIs have a significant effect on the impact of changing from a free-floating to a mobility hub system. Further research could therefore be done into accepting or rejecting these theories by means of user interviews.

12 Appendices

12.1 Appendix A

Code in R to train XGboost model, do cross validation, and do negative binomial regression.

```
# Set working directory and load packages -----
setwd(REDACTED)
library(foreign)
library(MASS)
library(ggplot2)
library(dplyr)
library(lubridate)
library(xgboost)
library(dplyr)
library(caret)
library(writexl)

# Import data and initiate for loop ----
dat <- read.csv("Variables.txt")
#set parameters for training different models
training_percentage <- 0.8
train_size <-  round(training_percentage*nrow(dat))
n_iterations <- 40
set.seed(42)
validation_iterations <-  list()
importance_iterations <-  list()

#For loop----
for (i in  1:n_iterations){
  print(i)

  #randomly select a training and validation dataset
  training_indeces <- sample(1:nrow(dat),train_size)
  training <- dat[training_indeces, ]
  validation <-  dat[-training_indeces, ]

  #split into dependent and independent variables
  dependent_var_training <- training$trips
  independent_vars_training <- training %>% select(supply,
     maxtemp, rain,
     matchday,weekday,avgtemp,holiday,rainhour)
  dependent_var_validation <- validation$trips
  independent_vars_validation <- validation %>%
     select(supply, maxtemp, rain,
     matchday,weekday,avgtemp,holiday,rainhour)

  #convert datasets to xgboost matrix
  data_matrix_training <- xgb.DMatrix(data =
     as.matrix(independent_vars_training), label =
     dependent_var_training)
```

```
  data_matrix_validation <- xgb.DMatrix(data =
     as.matrix(independent_vars_validation), label =
     dependent_var_validation)

  #Use xgboost
  params <- list(objective = "count:poisson", max_depth = 3,
     eta = 0.3, gamma = 0.1, subsample = 0.9,
     colsample_bytree = 0.9)  # Define XGBoost parameters
  model <- xgb.train(params, data = data_matrix_training,
     nrounds = 100)  # Train the XGBoost model
  importance_iteration <- xgb.importance(model=model)

  # Make predictions on the remaining 20% data
  validation <- data.frame(date = validation$date,
     trips_predicted = predict(model,
     data_matrix_validation), trips_actual =
     validation$trips)
  validation_iterations[[i]] <- validation

  #store importance of values
  importance_iterations[[i]] <- importance_iteration
}


#Combine results----

#Combine results for trips
validation_combined <- do.call(rbind,validation_iterations)
#group by date and summarize the trips_predicted rename to
   results
results <- validation_combined %>%
  group_by(date) %>%
  summarize(trips_predicted = mean(trips_predicted))
results$trips_actual <- dat$trips #add column for actual
   trips
results$difference <-
   abs(results$trips_actual-results$trips_predicted) #add
   columns for differences


#Combine results for importance
importance_combined <-  do.call(rbind,importance_iterations)
#group by feature and summarize all values
importance <- importance_combined %>%
  group_by(Feature) %>%
  summarize(Gain = mean(Gain),Cover = mean(Cover),Frequency
     = mean(Frequency))


#Make plot for both actual and predicted----
```

```
plot1 <- ggplot(results, aes(x =
   seq_along(trips_predicted))) +
  geom_line(aes(y = trips_predicted, color = "Predicted")) +
  geom_line(aes(y = trips_actual, color = "Actual")) +
  labs(title = "# of trip prediction per day",
       x = "Day index",
       y = "# of Trips") +
  scale_color_manual(values = c("Predicted" = "blue",
     "Actual" = "red")) +
  theme_minimal()
print(plot1)

#make plot for actual vs predicted
Accuracy = RMSE(results$trips_predicted,results$trips_actual)

plot2 <- ggplot(results, aes(x = trips_actual, y =
   trips_predicted)) +
  geom_point() +
  geom_abline(intercept = 0, slope = 1, color = "red") +  #
     Add a reference line for perfect predictions
  labs(x = "Actual # of trips", y = "Predicted # of trips") +
  ggtitle("Actual vs. predicted # of trips")
plot2 <- plot2 +
  annotate("text", x = Inf, y = Inf, label = paste("RMSE =",
     round(Accuracy, 2)),
           hjust = 1, vjust = 1, size = 4)
print(plot2)

#Do negative binomial regression----
model_regression <- glm.nb(trips ~ supply + maxtemp + rain +
   weekday + rainhour + avgtemp + matchday + holiday, data =
   dat)
summary(model_regression)
```

12.2 Appendix B

Code in R to compute predictions per facility

```
# Make predictions for Enschede noord per facility ----

#load data for training
dat_kennispark_before <-
   read.csv('Variables_EnschedeN_before.csv')
dat_kennispark_after <-
   read.csv('Variables_EnschedeN_after.csv')

#set number of facilities (dependent on area which is
   analysed)
n_facilities = 30

#store results per facility
results <- list()
```

```
#Select variables per facility and make predictions
facilityIDs <-  c(1,2,  3,     4,     5,     6,
   7,     8,     9,     10,    11,    12,    13,
   14,    15,    16,    17,    18,    20,    21,
   22,    23,    24,    25,    26,    27,    28,
   29,    31,    33)
for (facility in 1:n_facilities){
  print(facility)

 #select columns of faciltiy in iteration
 column_supply <- paste0("supply_counts_", facility)
 column_trips <- paste0("trips_counts_",facility)

 #select variables and rename them
 vars_facility_before <- dat_kennispark_before %>%
    select(column_trips,column_supply,weather_rain,weather_rain_hour,weathe
 vars_facility_after <- dat_kennispark_after %>%
    select(column_trips,column_supply,weather_rain,weather_rain_hour,weathe
 old_column_names <-
    c(column_trips,column_supply,'weather_rain','weather_rain_hour','weathe
 new_column_names <-
    c('trips','supply','rain','rainhour','avgtemp','maxtemp','weekday','ma
 vars_facility_before <- vars_facility_before %>%
   rename_with(~new_column_names[match(.x,
      old_column_names)], everything())
 vars_facility_after <- vars_facility_after %>%
   rename_with(~new_column_names[match(.x,
      old_column_names)], everything())

 #select indepndent and dependent variables for the facility
 independent_vars_facility_before = vars_facility_before %>%
    select(supply, maxtemp, rain,
    matchday,weekday,avgtemp,holiday,rainhour)
 dependent_var_facility_before = vars_facility_before$trips
 independent_vars_facility_after = vars_facility_after %>%
    select(supply, maxtemp, rain,
    matchday,weekday,avgtemp,holiday,rainhour)
 dependent_var_facility_after = vars_facility_after$trips

 #make data matrix
 data_matrix_facility_before <- xgb.DMatrix(data =
    as.matrix(independent_vars_facility_before), label =
    dependent_var_facility_before)
 data_matrix_facility_after <- xgb.DMatrix(data =
    as.matrix(independent_vars_facility_after), label =
    dependent_var_facility_after)

 #train model on facility
 params <- list(objective = "count:poisson", max_depth = 3,
    eta = 0.3, gamma = 0.1, subsample = 0.9,
```

```
    colsample_bytree = 0.9)  # Define XGBoost parameters
 model_all <- xgb.train(params, data =
    data_matrix_facility_before, nrounds = 100)

 #predict trip values per facility
 trips_prediction <- predict(model_all,
    data_matrix_facility_after)

 #compute results per facility
 trips_after <- sum(dependent_var_facility_after)
 trips_after_average <- mean(dependent_var_facility_after)
 trips_predicted <- sum(trips_prediction)
 trips_predicted_average <- mean(trips_prediction)
 trips_difference <- trips_after-trips_predicted
 trips_difference_average <- trips_after_average -
    trips_predicted_average

 results_facility <-
    data.frame(facilityIDs[facility],trips_after,
    trips_predicted, trips_difference, trips_after_average,
    trips_predicted_average, trips_difference_average)

 #store results per facility
 results[[facility]] = results_facility
}

combined_results <- do.call(rbind,results)
mean(combined_results$trips_difference)
write_xlsx(combined_results,
   "D:/Files/CE/Thesis/EnschedeNoord/ResultsEnschedeN.xlsx")
}
```

12.3 Appendix C

Code in R to compute confidence intervals for facility predictions

```
# Set working directory and load packages -----
directory <- (REDACTED)

#load packages
library(foreign)
library(MASS)
library(ggplot2)
library(dplyr)
library(lubridate)
library(xgboost)
library(dplyr)
library(caret)


#load all csv files ----
```

```
csvFiles <- list.files(directory, pattern = ".csv",
   full.names = TRUE)
#loop over csv files and do cross validation on each
   facility ----
#save RMSE values
RMSE_values <-  list()

for (file in csvFiles){
  print(file)

  #load data
  dat_unfiltered <- read.csv(file)

  #filter data for outliers (>5x mean)
  dat_trips_mean <- mean(dat_unfiltered$trips)
  dat_trips_std <- sd(dat_unfiltered$trips)
  dat <- dat_unfiltered[dat_unfiltered$trips <=
     2*dat_trips_mean,]

  #set parameters for training different models
  training_percentage <- 0.8
  train_size <-  round(training_percentage*nrow(dat))
  n_iterations <- 100
  set.seed(42)
  validation_iterations <-  list()
  importance_iterations <-  list()

  for (i in  1:n_iterations){
    print(i)

    #randomly select a training and validation dataset
    training_indeces <- sample(1:nrow(dat),train_size)
    training <- dat[training_indeces, ]
    validation <-  dat[-training_indeces, ]

    #split into dependent and independent variables
    dependent_var_training <- training$trips
    independent_vars_training <- training %>% select(supply,
       maxtemp, rain,
       matchday,weekday,avgtemp,holiday,rainhour)
    dependent_var_validation <- validation$trips
    independent_vars_validation <- validation %>%
       select(supply, maxtemp, rain,
       matchday,weekday,avgtemp,holiday,rainhour)

    #convert datasets to xgboost matrix
    data_matrix_training <- xgb.DMatrix(data =
       as.matrix(independent_vars_training), label =
       dependent_var_training)
    data_matrix_validation <- xgb.DMatrix(data =
       as.matrix(independent_vars_validation), label =
```

```
        dependent_var_validation)

    #Use xgboost
    params <- list(objective = "count:poisson", max_depth =
        3, eta = 0.3, gamma = 0.1, subsample = 0.9,
        colsample_bytree = 0.9)  # Define XGBoost parameters
    model <- xgb.train(params, data = data_matrix_training,
        nrounds = 100)  # Train the XGBoost model
    importance_iteration <- xgb.importance(model=model)

    # Make predictions on the remaining 20% data
    validation <- data.frame(date = validation$date,
        trips_predicted = predict(model,
        data_matrix_validation), trips_actual =
        validation$trips)
    validation_iterations[[i]] <- validation

    #store importance of values
    importance_iterations[[i]] <- importance_iteration
  }

  #Combine results for trips
  validation_combined <- do.call(rbind,validation_iterations)
  results <- validation_combined %>% #group by date and
     summarize the trips_predicted rename to results
    group_by(date) %>%
    summarize(trips_predicted = mean(trips_predicted))
  results$trips_actual <- dat$trips #add column for actual
     trips
  results$difference <-
     abs(results$trips_actual-results$trips_predicted) #add
     columns for differences
  trip_predicted_sum <- sum(results$trips_predicted) #summed
     trips predicted results

  #store accuracy (RMSE) and calculate confidence levels
  Accuracy =
     RMSE(results$trips_predicted,results$trips_actual)
  RMSE_values[[file]] <- Accuracy
}

#import list of predicted values to create boxplots----
Accuracy_data <- data.frame(actual = c(47,      69,
   2063,    1036,    792,     708,     499,     506,     1979,
   1451,    1221,    534,     739,     84,      76,      157,
   576,     524,     98,      49,      58,      385,     92,
   717,     421,     241,     137))
Accuracy_data$predictions <- c(30,      46,      1812,
   518,     755,     630,     527,     306,     1845,    1424,
   1196,    505,     666, 72,          73,      77,      586,
   567,98,  54,      55,      365      ,59      ,716,    70,
```

```
   187,    100)
facilityIDs <-
   c(19,32,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,5

n <- 180
df <- n -1 #degrees of freedom
confidence_level <- 0.9
for (i in 1:nrow(Accuracy_data)){
  print(i)
  critical_value <- qt((1 - confidence_level) / 2, df)
  Current_RMSE <- Accuracy_data$predictions[i] #select
     current RMSE value
  Accuracy_data$CI_lower[i] <- Current_RMSE -
     (critical_value * (Current_RMSE / sqrt(n)))
  Accuracy_data$CI_upper[i]<- Current_RMSE + (critical_value
     * (Current_RMSE / sqrt(n)))
}

# Convert facilityIDs to a factor with specified levels
facilityIDs <- factor(facilityIDs)

# create a bar chart
# create a bar chart
ggplot(Accuracy_data) +
  geom_bar(aes(x = facilityIDs, y = predictions, fill =
     "Predicted␣#␣of␣trips"), stat = "identity", width =
     0.3, position = position_nudge(x = -0.2)) +
  geom_bar(aes(x = facilityIDs, y = actual, fill = "Actual␣#␣
     of␣trips"), stat = "identity", width = 0.3, position =
     position_nudge(x = 0.2)) +
  geom_errorbar(aes(x = facilityIDs, ymin = CI_lower, ymax =
     CI_upper, color = "Confidence␣Interval␣of␣90%"), width
     = 0.1, position = position_nudge(x = -0.2)) +
  scale_fill_manual(values = c("Actual␣#␣of␣trips" = "red",
     "Predicted␣#␣of␣trips" = "pink")) +
  scale_color_manual(values = "blue", guide =
     guide_legend(title = "Error␣Bar")) +
  xlab("Facility␣IDs") +
  ylab("#␣of␣trips") +
  theme_minimal()
```

12.4 Appendix D

Code in MATLAB to compute and format variables.

```
  %%
clc, clear all
%Author: Dylan van Bezooijen
%Function: Compute variables into usable format for R
%% PREPARE DATA
% Load weather data
```

```
weather =
   readtable('weather_before.txt','ReadRowNames',true);
weather_date(:,1) = table2array(weather(:,1));
weather_temp_max(:,1) = table2array(weather(:,14))./10;
weather_rain(:,1) = table2array(weather(:,22));
weather_rain_hour(:,1) = table2array(weather(:,24));
weather_temp_avg(:,1) = table2array(weather(:,11))./10;
weekday = weekday(weather_date(:,1));

% Load park event data (Supply)
supply = readtable('supply_before.csv','ReadRowNames',true);
DateTime =
   datetime(table2array(supply(:,3)),'InputFormat','yyyy-MM-dd␣
   hh:mm:ss.SSS'); %Convert time format to usable number
   format
supply_date = convertTo(DateTime,'yyyymmdd');
supply_facility = table2array(supply(:,4));

%Load trip event data
trips = readtable('trips_before.csv','ReadRowNames',true);
DateTime =
   datetime(table2array(trips(:,3)),'InputFormat','yyyy-MM-dd␣
   hh:mm:ss.SSS'); %Convert time format to usable number
   format
trips_date = convertTo(DateTime,'yyyymmdd');
trips_facility = table2array(trips(:,4));

% facilities
unique_facilities = unique(supply_facility);

% Initialize a cell array to store the supply and trip counts
supply_counts =
   zeros(length(unique_facilities),length(weather_date));
trips_counts =
   zeros(length(unique_facilities),length(weather_date));

% Iterate over each unique facility
for i = 1:length(unique_facilities)
    % Get the current facility ID
    facility_id = unique_facilities(i);
    % Find the indices of trips made at the current facility
    facility_indices_supply = find(supply_facility ==
       facility_id);
    facility_indices_trips = find(trips_facility ==
       facility_id);
    %Corrospond with date
    facility_dates_supply =
       supply_date(facility_indices_supply);
    facility_dates_trips =
       trips_date(facility_indices_trips);
    % Count based on date for current facility
```

```
    [facility_count_supply, facility_date_supply] =
       groupcounts(facility_dates_supply);
    [facility_count_trips, facility_date_trips] =
       groupcounts(facility_dates_trips);
    %find the indices of at what dates counts were made (not
       a count for
    %each day)
    for k = 1:length(weather_date);
        %check whether date is found in count
        date_indeces_supply = find(facility_date_supply ==
           weather_date(k));
        date_indeces_trips = find(facility_date_trips ==
           weather_date(k));
        %if it is found...
        if date_indeces_supply ~ 0;
            %store value for count
            supply_counts(i,k) =
               facility_count_supply(date_indeces_supply);
        end
        if date_indeces_trips ~ 0;
            trips_counts(i,k) =
               facility_count_trips(date_indeces_trips);
        end
    end
end
supply_counts = supply_counts';
trips_counts = trips_counts';
%load match days
matches = readtable('matches.csv');
DateTime
   =datetime(table2array(matches(:,2)),'InputFormat','yyyy-MM-dd');
   %Convert time format to usable number format
matches_date = convertTo(DateTime,'yyyymmdd');
matches_boolean = zeros(length(weather_date),1);
for t1 = 1:length(weather_date)
    for t2 = 1:length(matches_date)
        if weather_date(t1) == matches_date(t2)
            matches_boolean(t1) = 1;
        end
    end
end

%set 'special days'
holidays_date = [20220505 20230505 20230101 20230407
   20230409 20230410 20230427 ...
    20220101 20220415 20220417 20220418 20220427 20221231
       20220823:20220831];
holiday_boolean = zeros(length(weather_date),1);
for t1 = 1:length(weather_date)
    for t2 = 1:length(holidays_date)
        if weather_date(t1) == holidays_date(t2)
```

```
            holiday_boolean(t1) = 1;
        end
    end
end

variables_table =
   splitvars(table(weather_date,trips_counts,supply_counts,...
    weather_temp_max,weather_rain,weekday,weather_rain_hour,weather_temp_av
    matches_boolean,holiday_boolean));
writetable(variables_table,'Variables_EnschedeN_before.csv','Delimiter',',

%export table for each individual facility
  for f = 1:length(unique_facilities)
      facility_ID = unique_facilities(f);
      %select appropiate columns for newtable
      newTable = variables_table(:,[1 f+1 f+28 56 57 58 59
         60 61 62]);
      %modify variable names
      varNames =
         {'date','trips','supply','maxtemp','rain','weekday',...
          'rainhour','avgtemp','matchday','holiday'};
      newTable.Properties.VariableNames = varNames;

      %generate table name
      tableName = sprintf('Facility_%d',facility_ID)
      writetable(newTable,[tableName '.csv']);
  end
```

References

Blad, K., de Almeida Correia, G. H., van Nes, R., & Annema, J. A. (2022). A methodology to determine suitable locations for regional shared mobility hubs. *Case Studies on Transport Policy*, *10*, 1904–1916. https://doi.org/10.1016/j.cstp.2022.08.005

Buehler, R. (2011). Determinants of transport mode choice: A comparison of germany and the usa. *Journal of Transport Geography*, *19*, 644–657. https://doi.org/10.1016/j.jtrangeo.2010.07.005

The core of giscience. (2020). https://ltb.itc.utwente.nl/page/481/concept/78717

Dataset: Nationaal wegen bestand (nwb). (2023). https://www.pdok.nl/introductie/-/article/nationaal-wegen-bestand-nwb-

Enschede, G. (2019). *Mobiliteitsvisie enschede: Leefbaar, aantrekkelijk en bereikbaar*. https://www.enschede.nl/mobiliteitsvisie-toekomst-van-het-verkeer-in-enschede

Enschede, G. (2023). *Verordening kwaliteit leefomgeving gemeente enschede 2023*. https://enschede.bestuurlijkeinformatie.nl/Agenda/Document/fe79436e-f34d-4d17-a412-6d1769d9d06b?documentId=b9f89999-bae1-4897-8a90-7ac4141d67e81

Goossens, R. (2022). *230111 ienw toelichting pva hubs.*

Jorritsma, P., Witte, J.-J., González, M. J. A., & Hamersma, M. (2021). *Kennisinstituut voor mobiliteitsbeleid — deelauto- en deelfietsmobiliteit in nederland.*

Li, W., & Kamargianni, M. (2018). Providing quantified evidence to policy makers for promoting bike-sharing in heavily air-polluted cities: A mode choice model and policy simulation for taiyuan-china [Factors effecting choosing bikesharing¡br/¿¡br/¿Natural and build environment, includes weather temperature air pollutions¡br/¿¡br/¿Trip attributes: most importantly travel time¡br/¿¡br/¿socio economic characterisitcs: age and gender. Younger highly educated males.]. *Transportation Research Part A: Policy and Practice*, *111*, 277–291. https://doi.org/10.1016/j.tra.2018.01.019

Machado, C. A., de Salles Hue, N. P. M., Berssaneti, F. T., & Quintanilha, J. A. (2018). An overview of shared mobility. *Sustainability (Switzerland)*, *10*. https://doi.org/10.3390/su10124342

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*, *44*, 1325–1342. https://doi.org/10.1007/s11116-017-9806-y

Morde, V. (2019). *Xgboost algorithm: Long may she reign!* https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d

Murray, L., Nguyen, H., Lee, Y.-F., Remmenga, M. D., & Smith, D. W. (2012). Variance inflation factors in regression models with dummy variables. *Conference on Applied Statistics in Agriculture*. https://doi.org/10.4148/2475-7772.1034

NOS. (2020). *Deelscooter is groene belofte, maar vaak ook overlastgever* [author not disclosed]. https://nos.nl/artikel/2359052-deelscooter-is-groene-belofte-maar-vaak-ook-overlastgever

Rongen, T., Lenferink, S., Arts, J., & Tillema, T. (2023). The peripheral mobility hub as a multi-sided platform? applying a fuzzy delphi to identify promising stakeholder interactions. *Research in Transportation Business and Management*. https://doi.org/10.1016/j.rtbm.2023.100953

Seker, S., & Aydin, N. (2023). Fermatean fuzzy based quality function deployment methodology for designing sustainable mobility hub center. *Applied Soft Computing*, *134*, 110001. https://doi.org/10.1016/j.asoc.2023.110001

Srinivasan, S. (2005). Quantifying spatial characteristics of cities. *Urban Studies*, *39*. https://doi.org/10.1080/004209802200001133 van Steijn, J., Kamphuis, R., & Hattink, M. (2022). *Tussenevaluatie deelmobiliteitshubs kennispark enschede rapportage*. www.keypoint.eu

Vianen, J. (2022). *A design approach to determine the locations in combination with amenities for neighbourhood hubs, based on user profiles qualitatively researching the users, amenities, and locations for neighbourhood hubs*. TU Delft. http://resolver.tudelft.nl/uuid:5cc13d7c-04c8-4b35-aa68-a4a58159e91e

Zhou, H., Dorsman, J. L., Mandjes, M., & Snelder, M. (2023). Sustainable mobility strategies and their impact: A case study using a multimodal activity based model. *Case Studies on Transport Policy*, *11*. https://doi.org/10.1016/j.cstp.2022.100945