The influence of infrastructure and land use allocation on the route choice of cyclists

BACHELOR THESIS

IMAGE BY (Fietssnelweg F35, sd)

AUTHOR:	NICK VAN NIJEN
STUDENTNUMBER:	S2147483
SUPERVISOR UT:	BARAN ULAK
SUPERVISOR WITTEVEEN+BOS:	SANDER VEENSTRA
DATE:	24 JUNE 2022

Witteveen + Bos

UNIVERSITY OF TWENTE.

PREFACE

In front of you is the final report of my bachelor thesis with as topic 'The influence of infrastructure and land use allocation on the route choice of cyclists'. I carried out this thesis at the Verkeer en Wegen Breda department of Witteveen+Bos and with this thesis I conclude my bachelor Civil Engineering at the University of Twente. In this thesis, I used a segment approach to evaluate the infrastructural and land use allocation factors of observed routes from 'fietstelweek' data and generated shortest paths. I hope this evaluation contributes to better predictability of bicycle route choices in models and eventually to a better bicycle (built) environment.

This thesis gave me the opportunity to develop skills that I had not learned in the rest of my bachelor study. In the past weeks I learned a lot on how to use Python to make your own model and how to develop a regression model. Moreover, I enjoyed going to office in Deventer to differentiate in work environment and to experience how it is to work at a company.

For that and giving me the opportunity to graduate I want to thank Witteveen+Bos. Besides that, I would like to thank my external supervisor at Witteveen+Bos, Sander Veenstra. He has supported me extensively during my graduation period and was always there to help me with all kinds of (modelling) questions. Furthermore, I want to thank my internal supervisor from the University of Twente, Baran Ulak for his support and helpful feedback. Together with both supervisors I had great discussions on the topic and those discussions improved my thesis even further.

I hope you enjoy reading my thesis.

Nick van Nijen

Enschede, 24 June 2022

SUMMARY

In the Netherlands, the bicycle is strongly embedded in the road infrastructure and many instances are trying to strengthen that position, for example by improving bicycle infrastructure. Witteveen+Bos set up the 'FietsMonitor' (literally translated bicycle monitor) to visualise bicycle flows, which clients can use to substantiate their bicycle policy. One of the major points of improvement for the model is that all routes are assigned on an 'all or nothing' basis to the shortest path, while Witteveen+Bos expects that environmental factors influence the route choice of cyclists substantially. This claim is supported by a substantial number of studies, however these studies used different approaches and obtained different results, indicating the uncertainty that environmental factors have on the route choice of cyclists. Within that gap, the aim of this study is to assess the influence of infrastructural and land use allocation factors on the route choice of cyclists in Enschede by the use of 'fietstelweek' data.

After conducting a literature study, fourteen infrastructural and land use allocation factors emerged as substantially influencing the route choice of cyclists and therefore were included in the analysis, namely the presence of traffic control installations, bicycle lanes, separated bicycle paths, artificial lighting and paved infrastructure, the intensities of motorised vehicles and bicycles, the land use in the area (either residential, commercial, industrial or greenery), land use mix and the floor area ratio.

In this study a segment approach is used, where a comparison is made between the segments that are chosen by cyclists (the observed route) and the segments that are not chosen, but recommended when using the shortest path. Segments of the shortest path overlapping with segments of the corresponding observed route are filtered out. Finally, a logistic regression model is constructed with the aforementioned segments in order to obtain a formula presenting the probability that a cyclist would use a segment based on the characteristics a segment has. This probability is associated with the attractiveness of a segment. Since the dataset was substantially large, a Monte Carlo approach was used to cross-validate the formula.

Moreover, the model was validated for consistency with existing literature and turned out consistent, except the factors '*Artificial lighting', 'Bicycle intensities'* and *'Green land use zone'*. Furthermore, the model was tested on its predictive performance and showed relatively good scores on the four used key performance indicators, namely accuracy, precision, sensitivity and F1-score. In addition, the model of Enschede was projected on the city of Haarlem. However, the model scored substantially less on the key performance indicators for Haarlem, compared to Enschede. Besides that, a model for Haarlem was created. This model showed a substantial difference in the coefficients, especially for the presence of paving stones, the land use mix and the area covered with greenery. Finally, an attractiveness map of Enschede was constructed, which contributed to the validity of the model by evaluating interesting areas.

Finally, the use of the formula is recommended when evaluating a route choice set on their attractiveness, when the trip length of the included routes are comparable. Moreover, the attractiveness map is recommended for indicating substantially less attractive segments in a network and is recommended to policy makers to substantiate their policy on the improvement of the bicycle infrastructure. However, both the model and the attractiveness map are only suitable to use on the study area they are created for and are not generally applicable.

TABLE OF CONTENTS

1. Introduction

1.1. Problem Context

The Netherlands is a country where the bicycle is used often (28% of all trips in 2019 (Centraal Bureau voor de Statistiek, sd)) and therefore is strongly embedded in road infrastructure of the Netherlands. Moreover, municipalities and other governmental institutions are trying to strengthen the position of the bicycle even more by improving bicycle facilities, such as improved bicycle infrastructure (Ensing & Janssen (2020), Gemeente Alkmaar (2020), Gemeente Maastricht (2020) and Gemeente Enschede (2012)). When proposing to improve bicycle infrastructure, it is important for policy makers to understand the current situation. Therefore, Witteveen+Bos set up the 'FietsMonitor' (literally translated bicycle monitor). This model generates and visualises bicycle flows, using the four steps model. The four steps model consists of four steps, namely trip generation, trip distribution, modal split and trip assignment (Gkiotsalitis, 2021). Before the first step is executed, a bicycle network is created. This can be indicated as step zero. In the first step, trips are generated in an origin destination matrix using 'Onderzoek Verplaatsingen in Nederland' (OViN) and 'Onderweg in Nederland' (ODiN) data. In the second step, the trip distribution, the shortest path for all origin destination pairs, is determined using Dijkstra algorithm graph theory. The third step is neglected as the 'FietsMonitor' only models cyclists. Finally, in the fourth step, all trips, from origin to destination, are assigned to the shortest path (Veenstra, Geurs, Thomas, & Van den Hof, 2016).

Witteveen+Bos uses this model to consult clients about the bicycle flows, bottlenecks in the current infrastructure and the influence of proposed measures. The use of the model provides fast and visual insights in bicycle flows and capacity, which are easy to interpret by clients (Witteveen+Bos, sd). However, the model also has some points of improvement as described by Veenstra, Geurs, Thomas & Van den Hof (2016). One of them is the way route choices are modelled. Currently, routes are assigned on an 'all or nothing' basis to the shortest path. This means that all trips from a certain origin to a certain destination follow the same route and this route is always the shortest distance between the origin and the destination. Witteveen+Bos estimates that environmental factors, such as dedicated bicycle infrastructure and green land use, have an influence on the route choice of cyclists and as a result, the model might not resemble reality as much as Witteveen+Bos wants to (Veenstra, 2022). Observed routes from 'fietstelweek' data (Breda University of Applied Sciences, sd) supports this claim, since not all shortest paths overlap fully with the observed route, as visualised in Figure 1. In Enschede only 20.3% of the observed routes fully overlap with the shortest path.

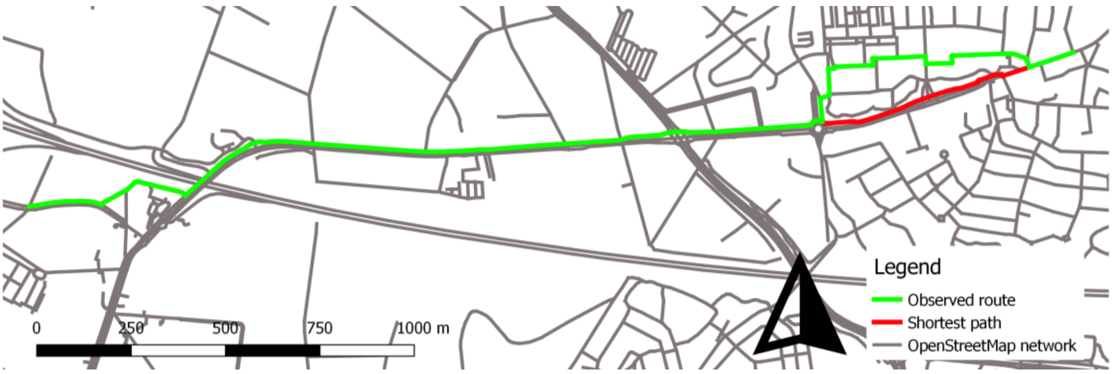

FIGURE 1: OBSERVED ROUTE COMPARED TO THE SHORTEST PATH FOR A TRIP ALONG THE HAAKSBERGERSTRAAT TOWARDS USSELO

Consequently, Witteveen+Bos wants to improve the 'FietsMonitor' by adding environmental factors to the route choice algorithm used in the 'FietsMonitor'. They are planning on adapting the route choice algorithm by generating multiple routes, instead of only the shortest path. Eventually, all trips from a certain origin to a certain destination will be distributed over the potential routes (Veenstra, 2022). In order to do this, it is important to gain more insight in the influences of environmental factors on the route choice of cyclists.

1.2. Research Gap

Research proved that environmental factors may lead to cyclists deviating from the shortest path. Research in this field is executed via multiple ways, including surveys and a number of different (regression) models, as multinomial probit model and mixed logit model. This diversity in approaches shows already that there is no clear way how environmental factors influence the route choice of cyclists and they show that the influence can vary between multiple studies and use cases. This study aims to add to the existing literature by analysing the influence of environmental factors in a Dutch setting by comparing shortest paths with observed routes from 'fietstelweek' data (Breda University of Applied Sciences, sd) and come up with a model that can predict the probability (or attractivity) of segments. 'fietstelweek' data has not been used in literature often (Koch & Dugundji, 2021) and is never used to make a comparison between shortest paths and observed routes.

1.3. Research Aim

This study will focus on the environmental factors that have an influence on the route choice of cyclists and how this can be used to improve the route choice algorithm that is used in the 'FietsMonitor'. This is done by making a comparison between shortest path routes and observed routes from 'fietstelweek' (Breda University of Applied Sciences, sd) data. In this study, environmental factors are specified by infrastructural and land use allocation factors. As a result, the objective of this study is:

"To assess the influence of infrastructural and land use allocation factors on the route choice of cyclists in Enschede by the use of 'fietstelweek' data."

1.4. Research Scope

The research scope defines the boundaries of this study. The scope is identified according to the research aim as discussed in Section 1.3 and elaborates on all elements of that aim.

1.4.1. Method: Quantitative Assessment

In this study, a quantitative approach is chosen where a route set of shortest paths and observed routes will be used to evaluate the influence of infrastructural and land use allocation factors on the route choice of cyclists. This approach is a balance between simplicity, by only considering only two route alternatives, and accuracy, by quantitatively assessing the results. However, there are alternatives to that. For example, other researchers used (stated preference) surveys in order to obtain a qualitative or quantitative result on environmental factors influencing route choice of cyclists (e.g. Mertens et al. (2016) and Band (2022)). Especially in stated preference surveys, it is uncertain whether the outcomes of such a survey correspond with the choice that cyclists would make in real-life. Furthermore, other researchers estimated a route choice model and evaluated the alternative routes (e.g. Koch & Dugundji (2021) and Chen (2016)). Nevertheless, Koch & Dugundji indicated already that the route choice of cyclists come with a high level of stochasticity, which makes simulation of the route choice of cyclists hard.

Moreover, the focus of this study is on the quantification of infrastructural and land use allocation factors. Only these factors will be considered, as these are relatively easy to quantify using existing open-source data. Datasets that already have included related data are 'OpenStreetMap' (Openstreetmap, sd) and 'Basisregistratie Grootschalige Topografie' (PDOK, sd). The advantage of open-source data is that it is easy accessible for the purposes of this study, but also when the data is being implemented in the 'FietsMonitor'.

Furthermore, when using the 'OpenStreetMap' (Openstreetmap, sd) network it is important to note that the network is split into segments. A road is usually split in segments when one or more characteristics change or when one or more other links (roads, bicycle paths etc.) meet or divert from the road. This principle is visualised in Figure 2 with a bicycle path. You can see that the stretch of bicycle path meets other links at both ends of the bicycle path, resulting in the stretch in between.

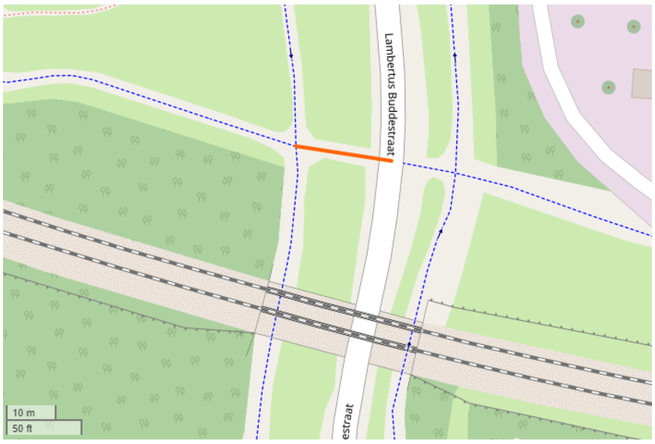

FIGURE 2: VISUALISATION WORKING OF SEGMENTS, A SPECIFIC SEGMENT IS VISUALISED WITH A RED LINE (OPENSTREETMAP, SD)

1.4.2. Infrastructural and Land Use Allocation Factors

Under infrastructural influences fall a large variety of man-made structures that affect bicycle traffic. This includes, but is not limited to the pavement of the bicycle infrastructure, the number and type of intersections that a cyclist may encounter on a certain route, whether the route is lit by artificial lighting and how the bicycle infrastructure is designed. With the design of the bicycle infrastructure is meant what the position of the bicycle path is. There are various possibilities for that including, but not limited to no dedicated bicycle infrastructure, a so called 'fietsstraat' (literally translated 'cycling street', a street in priority with extra facilities regarding cycling), separate bicycle path (one or both directions) and the addition of a bicycle lane to a road dominantly used by cars. For the last type of bicycle infrastructure, the speed of the surrounding cars is important to take into account.

Land use allocation factors include all forms of land use as stated by the land use plan of a certain municipality. This includes, but is not limited to land use classes like residential area, commercial area, industrial area and nature area. Also, mixed land use zones fall under land use allocation.

To set boundaries for this study, not all infrastructural and land use allocation factors are evaluated. The choice which factors to include and which not depends on what conclusions existing literature already draws and the data availability of the factors.

1.4.3. 'Fietstelweek' data

The data of the 'fietstelweek' includes a substantial number of parameters on bicycle behaviour (Breda University of Applied Sciences, sd). The most recent data is from 2016 in which bicycle users were asked to map their cycle behaviour via a smartphone app using GPS during the week of the 19th of September (Koch & Dugundji, 2021). This include parameters that are interesting for this study, for example all routes that a certain cyclist used (Breda University of Applied Sciences, sd). Later in this study, these routes are indicated as the observed routes. The data is open sourced, but also has some limitations. One of them is that when using GPS, the accuracy may not always be sufficient in order to indicate if for example a cyclists travels over a certain stretch of road or a parallel cycle path. This limitation should be taken into account when analysing the results.

The data of the 'fietstelweek' is of great value in this study. However, all the open-source data is anonymised. This is done in two ways that influence this research. Firstly, of all trips between zero and 400 meters were dropped off the beginning and end to mask the true origin and destination (Koch & Dugundji, 2021). Finally, no demographic data on, for example, gender and age are included and therefore also not included in this study. Nevertheless, as other model and questionnaire studies indicate, these aspects have an influence on cycling behaviour and route choice for cyclists (Broach & Dill (2016), Broach, Dill, & Gliebe (2012), Segadilha & Sanches (2014), Fitch & Handy (2020), Winters, Brauer, Setton, & Teschke (2010)). For example, women are more curbed than men from riding with on-street parking, up hills and in high traffic areas (Caulfield, Brick, & McCarthy, 2012). When analysing the results of the study, this aspects should be taken into account. Moreover, the participants in the 'fietstelweek' were mostly experienced cyclist resulting in a relative large share of longer trips. According to the 'fietstelweek' data of 2016, the average distance per trip was 4.44 kilometres. However, the average distance per trip according to Centraal Bureau voor de Statistiek (2017) was 3.56 kilometres.

1.4.4. Study Area

Witteveen+Bos is using the 'FietsMonitor' throughout whole of the Netherlands, so an improvement of the model would be beneficial for the implementation of the model throughout the Netherlands. However, this study focusses on a more specific area, the municipality of Enschede. The location of the municipality of Enschede, in the province of Overijssel is visualised in Figure 3. All observed routes from the 'fietstelweek' (Breda University of Applied Sciences, sd) data that have at least the origin or the destination in the municipality of Enschede are considered. This means that cyclists can travel from other municipalities close to the municipality of Enschede such as Hengelo, Haaksbergen, Losser and Oldenzaal. However, only the part of routes within the municipality borders of Enschede is included, since the data is less trustworthy or not existent outside the boundaries as indicated in Section 3.1.2.6 and Section 3.1.2.7. Nevertheless, this resulted in not fully including trips between two cities resulting in an underestimation of the characteristics of trajectories between two cities. Moreover, Enschede is strategically chosen as the 'fietstelweek' data initiative originates from Enschede. This results in relatively much data from this particular municipality (Fietstelweek, 2016). In total, data of 8,606 observed routes are present in the 'fietstelweek' data for Enschede.

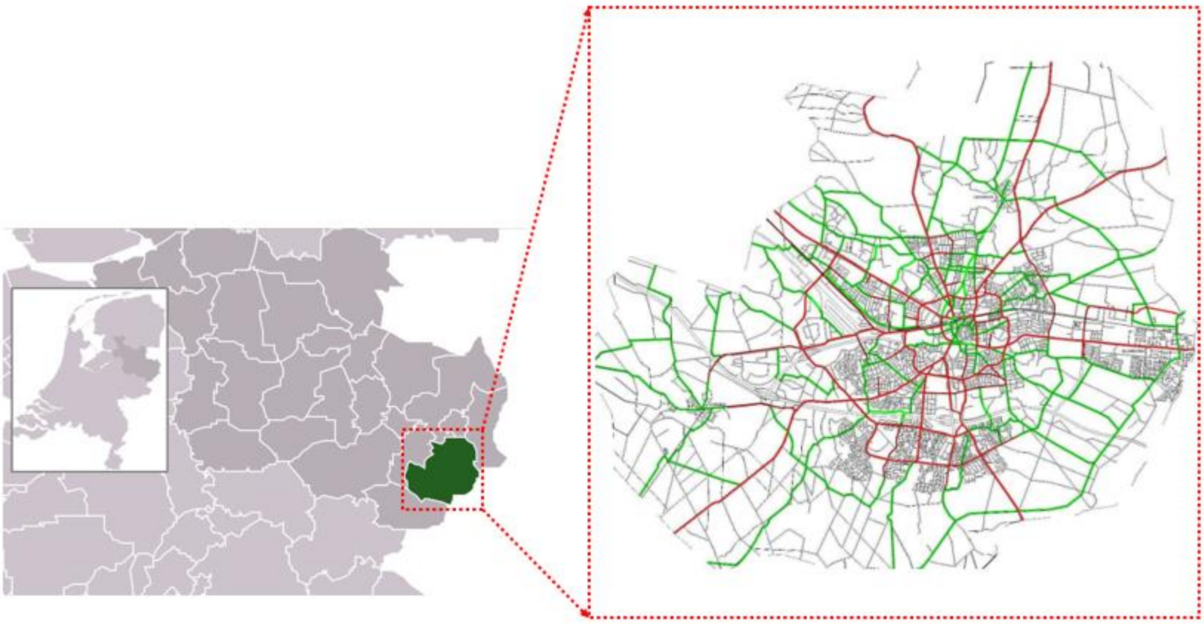

FIGURE 3: A) MUNICIPALITY OF ENSCHEDE AS VISUALISED IN GREEN IN THE PROVINCE OF OVERIJSSEL (CENTRAAL BUREAU VOOR DE STATISTIEK | TOPOGRAFISCHE DIENST KADASTER, 2005), B) MAIN BICYCLE NETWORK OF ENSCHEDE INCLUDING PRIMARY (RED) AND SECONDARY (GREEN) CYCLEWAYS (GEMEENTE ENSCHEDE, 2012)

1.5. Research Questions

The focus of this study is on the quantification of the influence of infrastructural and land use allocation factors on the route choice of cyclists. In order to assess this, a model is created that uses the characteristics of segments to make a comparison between segments of the shortest path and observed route. This can be described as a segment approach. Therefore the main research question is:

MQ ***"What are the influences of infrastructural and land use allocation factors on the route choice of cyclists through a segment approach?"***

This main research question is divided into four sub-questions that have to be answered after each other in order to answer the main research question.

The first sub-question focusses on which infrastructural and land use allocation factors have a substantial influence on the route choice of cyclists. Based on this, certain factors can be either included or excluded from this study. Therefore, the first sub-question is formulated as follows:

Q1 *"What infrastructural and land use allocation factors have a substantial influence on the route choice of cyclists according to literature?"*

With the substantial infrastructural and land use allocation factors known, the model can be constructed and regression modelling, a form of discrete choice modelling, can be performed to obtain a model that determines the probability of a segment to be chosen (or attractivity) by the means of its characteristics. This results in the second sub-question:

Q2 *"What is the quantitative influence of infrastructural and land use allocation factors on the route choice of cyclists through a segment approach and regression modelling?"*

With the model being built, the model can be evaluated for its validity. This is done in three ways, namely a comparison with literature, by testing the model on test data, that is not used in the training of the model and finally projecting the predictions of the model on another study area and verify its performance. This results in the third sub-question:

Q3 *"How valid are the results of the quantitative model when comparing it with existing literature, the use of test data and by projecting the model predictions on another study area?"*

When the model is proven valid, its results can be used in the 'FietsMonitor' as described in Section 1.1. However, the model might be useful for other purposes as well. For example a map can be made that visualises the attractiveness of all segments in a bicycle network. Therefore, the fourth and last sub-question is formulated as follows:

Q4 *"How can the attractiveness of segments in a bicycle network be quantified through the developed model?"*

1.6. Report Outline

This section introduced the problem context, research gap and research aim. Moreover, it bounded the study and stated the research questions, this study tries to answer. In the next section, Section 2, theoretical framework, existing literature on infrastructural and land use allocation factors is introduced. Furthermore, this section elaborates on regression modelling and provides a conceptual model. Thereafter, the methodology is introduced in Section 3. The aim of the methodology is to answer the four sub-questions as stated in Section 1.5. Firstly, the segment approach is introduced and afterwards the reports touches upon the implementation of regression modelling in this study, the validation of the model and the application of the model. In Section 4, the results of the earlier mentioned methodology are presented. Successively, these results are discussed in the discussion in Section 5. Finally, this report concludes with a conclusion in Section 6 and the recommendation for practical implementations and future research in Section 7.

2. Theoretical Framework

In this chapter, the main theories behind the study are analysed. This includes theories that are not only necessary to understand the context and complexity of the study, but also to have a better understanding of the methodology that is described in Section 3.

2.1. Influences on the Route Choice of Cyclists

First of all, the infrastructural and land use allocation factors influencing the route choice of cyclists are introduced. With that, the first sub-question will be answered and acts as input for the model described in Section 3.1.

Firstly, Koch & Dugundji (2021), Broach, Dill & Gliebe (2012), Prato, Halldórsdóttir & Nielsen (2018) and Stinson & Bhat (2003) concluded that the presence of traffic control installations encourages cyclists to take a detour. This can be due to the frequency of stopping, which is not perceived positively by cyclists. Band (2022) supports this claim as he found that cyclists are on average willing to cycle 5.3 minutes extra to get from a bicycle route with five or more stops compared to a routes that have few stops (between zero and two stops). Nevertheless, Broach, Dill & Gliebe (2012) and Khatri, Cherry, Nambisan & Han (2016) found out that traffic control installations are deemed valuable in a lefthand crossing. Especially with high vehicle intensities.

Secondly, many researchers agreed upon the fact that good bicycle facilities improve the attractiveness of a certain route significantly, as it improves bicycle flow and safety. Especially when the bicycle paths are separated from motorised traffic (Koch & Dugundji (2021), Chen (2016), Mertens, et al. (2016), Campos-Sánchez, Valenzuela-Montes, & Abarca-Álvarez, 2019) (2019), Broach, Dill, & Gliebe (2012), Saelens, Sallis, & Frank (2003), Winters, Davidson, Kao, & Teschke (2011), Khatri, Cherry, Nambisan, & Han (2016), Prato, Halldórsdóttir, & Nielsen (2018) and Stinson & Bhat (2003)). Mertens, et al. (2016) even claims that it is the most significant influence on the attractiveness of routes for cyclists. Furthermore, bicycle lanes next to motorised traffic have the same result, but less substantial (Koch & Dugundji (2021), Chen (2016), Mertens, et al. (2016) and Prato, Halldórsdóttir, & Nielsen (2018)). Band (2022) states that cyclists are on average willing to cycle 3.3 minutes to get to a bicycle route with a bicycle lane compared to a route without bicycle facilities. In contrast, Band (2022) found that cyclists want to detour an extra 4.7 and 5.2 minutes if they can use a one-directional bicycle path or a two-directional bicycle path respectively. However, Stinson & Bhat (2003) found that the impact of a separate cycling path is off less significance in the route choice of cyclists compared to a bicycle lane.

Thirdly, the slope of the route has a significant positive effect on the willingness of cyclists to detour (Chen (2016), Winters, Davidson, Kao, & Teschke (2011), Li, Wang, Liu, & Ragland (2012), Stinson & Bhat (2003) and Winters, Brauer, Setton, & Teschke (2010)). According to Broach, Dill & Gliebe (2012), cyclists are already willing to take a significant detour, when the slope of a certain route is 2% or more and Prato, Halldórsdóttir & Nielsen (2018) claim that the perception of distance is about 4.9 times higher when cycling uphill compared to a flat terrain. This could be due to more physical activity needed to cover sloped infrastructure. Although Stinson & Bhat (2003) agree that commuters prefere flat terrain over mountainess terrain, they saw a slight preference for hilly terrain over flat terrain.

Moreover, cyclists are deterrent towards a route, that is not well lit (Winters, Davidson, Kao, & Teschke, 2011). This might be because artificial lighting improves the perceived safety of a route after nightfall. Uttley, Fotios & Lovelace (2020) found that only a minimal change in brightness of the artificial lighting can have a big impact on cycling rates after nightfall. Similarly, an increase in light density may increase cycling rates after nightfall. However, it is advised by Uttley, Fotios & Lovelace (2020) to keep it to a minimum, as the increase in density and brightness does not lead to substantial differences in cycling rates after nightfall when the lighting is already fine.

Next to that, Prato, Halldórsdóttir & Nielsen (2018) and Winters, Davidson, Kao & Teschke (2011) claim that cyclists prefer to use paved infrastructure compared with unpaved infrastructure. This might be due to the fact that cycling over paved infrastructure costs less energy and is safer. Winters, Davidson, Kao & Teschke (2011) noted the importance of surface maintenance as cyclists are deterrent towards a route with potholes, uneven paving and lots of fallen leaves.

In addition, Broach, Dill & Gliebe (2012) claim that the intensities of motorised vehicles have a substantial negative effect on the willingness of non-commuter cyclists to use a certain road stretch. If this road stretch has a motorised vehicle volume of more than 20,000 vehicles per day and has not bicycle path, cyclists are willing to take a detour that are over twice their original route length. Mertens et al. (2016), Band (2022), Li, Wang, Liu, & Ragland (2012) and Stinson & Bhat (2003) support the negative effect of motorised traffic. Moreover, Winters, Brauer, Setton & Teschke (2010) saw an increase in the odds of people taking the bicycle when traffic calming measures were taken. However, Prato, Halldórsdóttir & Nielsen (2018) did not find a significant effect of traffic volumes. This can be causes by the presense of dedicated bicycle facilities at almost all high traffic volume streets in Copenhagen, Denmark, where this study was performed.

Besides that, Band (2022) touched upon the importance of bicycle intensities on the willingness to detour. He found that cyclists are on average willing to cycle an extra 2.2 minutes to avoid a busy bicycle route and use a quiet bicycle route with few cyclists. This can be due to the fact that overtaking and keeping the same speed is harder with higher bicycle intensities. Li, Wang, Liu, & Ragland (2012) support the negative influence of high bicycle volumes on the comfort of cyclists.

Furthermore, land use may have an influence on the attractiveness of a bicycle route. Koch & Dugundji (2021) and Li, Wang, Liu, & Ragland (2012) claim that cyclists try to avoid residential areas. Winters, Brauer, Setton & Teschke (2010) support that, as they claim that the percentage of single family residential land use reduced the odds of taking the bicycle. Nevertheless, Zhao, Ke, Lin & Yu (2020) claim that residential areas have a positive effect on the bicycle frequency. Prato, Halldórsdóttir & Nielsen (2018) split residential areas in high density and low density. They found that cyclists perceive a longer trip length when cycling through high density residential area compared to low density residential area. That finding might be related to a larger amount of potential conflicts caused by, for example bus stops and shopping opportunities. Moreover, Saelens, Sallis & Frank (2003) found no significant influence of residential zoning. The same study concluded no significant influence of commercial zoning. Nevertheless, Koch & Dugundji (2021) and Zhao, Ke, Lin & Yu (2020) saw that cyclists prefer commercial land use. Nonetheless, Winters, Brauer, Setton & Teschke (2010) saw a reduced change of taking the bike when a route travels through a high percentage of commercial land use. Also, Koch & Dugundji (2021) found that cyclists prefer green land use and Zhao, Ke, Lin & Yu (2020) saw a positive effect of green land use on the bicycle frequency. Prato, Halldórsdóttir & Nielsen (2018) support this by stating that green land uses, as forests and parks, have a positive effect on the willingness to detour considering medium to high air temperatures. Nevertheless, Campos-Sánchez, Valenzuela-Montes & Abarca-Álvarez (2019) found that green areas alone do not influence cyclists, but, for example, the proximity to separated cycle path is necessary in order to be more attractive. Moreover, Saelens, Sallis & Frank (2003), Zhao, Ke, Lin & Yu (2020) and Winters, Brauer, Setton & Teschke (2010) touched upon the positive effect of mixed land use on the amount of bicycle trips. This is supported by Maat, van Wee & Stead (2005) as they claim that mixed land use reduces the need to travel by car. This can be due to the fact that more facilities are in closer distance of residential buildings. Furthermore, Prato, Halldórsdóttir & Nielsen (2018) indicated that cyclists do not like industrial areas and are willing to take detours to avoid this land use class. However, Winters, Brauer, Setton & Teschke (2010) found no substantial increase or decrease in the probability of taking the bicycle when the route is through an industrial land use zone.

Finally, Chen (2016) touched upon the positive effects of a low floor area ratio on promoting bicycling. The floor area ratio is the total area a building uses over all floors divided by the gross lot area. A high ratio is often associated with a dense and/or urban area (Metropolitan Council, 2015). Therefore, Chen (2016) indicates that cyclists prefer less densely populated areas. In contrast, Saelens, Sallis & Frank (2003) claim that a high density area positively affects the attractiveness of routes and Winters, Brauer, Setton & Teschke (2010) supports that.

All possible infrastructural and land use allocation factors that could influence route choice of cyclists that are mentioned in the papers described above are summarised in Table 1. Moreover, the in Table 1 indicated influential factors are considered in the analysis of this study. The table also presents the effect of the influential factor on the attractiveness for cyclists to choose a certain route including references. Note that the effect is compared to increasing the presence of the influential factor.

TABLE 1: SUMMARY OF INFRASTRUCTURAL AND LAND USE ALLOCATION FACTORS AND THEIR SOURCES

Influential factor (increasing)	Effect on attractiveness of routes (including references)
Infrastructural	
Traffic control installations	**-** (Koch & Dugundji, 2021), **±** (Broach, Dill, & Gliebe, 2012), **+** (Khatri, Cherry, Nambisan, & Han, 2016), **-** (Prato, Halldórsdóttir, & Nielsen, 2018), **-** (Band, 2022), **-** (Stinson & Bhat, 2003)
Bike lanes	**+** (Koch & Dugundji, 2021), **+** (Chen, 2016), **+** (Mertens, et al., 2016), **+** (Prato, Halldórsdóttir, & Nielsen, 2018), **+** (Band, 2022), **+** (Stinson & Bhat, 2003)
Separate bike path	**+** (Koch & Dugundji, 2021), **+** (Chen, 2016), **+** (Mertens, et al., 2016), **+** (Campos-Sánchez, Valenzuela-Montes, & Abarca-Álvarez, 2019), **+** (Broach, Dill, & Gliebe, 2012), **+** (Saelens, Sallis, & Frank, 2003), **+** (Winters, Davidson, Kao, & Teschke, 2011), **+** (Khatri, Cherry, Nambisan, & Han, 2016), **+** (Prato, Halldórsdóttir, & Nielsen, 2018), **+** (Band, 2022), **+** (Stinson & Bhat, 2003)
Slope	**-** (Chen, 2016), **-** (Broach, Dill, & Gliebe, 2012), **-** (Winters, Davidson, Kao, & Teschke, 2011), **-** (Prato, Halldórsdóttir, & Nielsen, 2018), **-** (Li, Wang, Liu, & Ragland, 2012), **±** (Stinson & Bhat, 2003), **-** (Winters, Brauer, Setton, & Teschke, 2010)
Artificial lighting	**+** (Winters, Davidson, Kao, & Teschke, 2011), **+** (Uttley, Fotios, & Lovelace, 2020)
Paved infrastructure	**+** (Winters, Davidson, Kao, & Teschke, 2011), **+** (Prato, Halldórsdóttir, & Nielsen, 2018), **+** (Stinson & Bhat, 2003)
Motorised vehicle intensities	**-** (Broach, Dill, & Gliebe, 2012), **-** (Mertens, et al., 2016), **-** (Band, 2022), **±** (Prato, Halldórsdóttir, & Nielsen, 2018), **-** (Li, Wang, Liu, & Ragland, 2012), **-** (Stinson & Bhat, 2003), **-** (Winters, Brauer, Setton, & Teschke, 2010)
Bicycle intensities	**-** (Band, 2022), **-** (Li, Wang, Liu, & Ragland, 2012)
Land use allocation	
Residential land use zone	**-** (Koch & Dugundji, 2021), **±** (Saelens, Sallis, & Frank, 2003), **±** (Prato, Halldórsdóttir, & Nielsen, 2018), **-** (Li, Wang, Liu, & Ragland, 2012), **+** (Zhao, Ke, Lin, & Yu, 2020), **-** (Winters, Brauer, Setton, & Teschke, 2010)
Commercial land use zone	**+** (Koch & Dugundji, 2021), **±** (Saelens, Sallis, & Frank, 2003), **+** (Zhao, Ke, Lin, & Yu, 2020), **-** (Winters, Brauer, Setton, & Teschke, 2010)
Greenery land use zone	**+** (Koch & Dugundji, 2021), **±** (Campos-Sánchez, Valenzuela-Montes, & Abarca-Álvarez, 2019), **+** (Prato, Halldórsdóttir, & Nielsen, 2018), **+** (Zhao, Ke, Lin, & Yu, 2020)
Industrial land use zone	**-** (Prato, Halldórsdóttir, & Nielsen, 2018), **±** (Winters, Brauer, Setton, & Teschke, 2010)
Land use mix	**+** (Saelens, Sallis, & Frank, 2003), **+** (Maat, van Wee, & Stead, 2005), **+** (Zhao, Ke, Lin, & Yu, 2020), **+** (Winters, Brauer, Setton, & Teschke, 2010)
Floor area ratio	**-** (Chen, 2016), **+** (Saelens, Sallis, & Frank, 2003), **-** (Prato, Halldórsdóttir, & Nielsen, 2018), **+** (Winters, Brauer, Setton, & Teschke, 2010)

The trip length is considered one of the attributes most often cited in literature as a factor influencing route choice of cyclists (Chen (2016), Broach, Dill, & Gliebe (2012), Heinen, Maat, & Van Wee (2011), Khatri, Cherry, Nambisan, & Han (2016), Prato, Halldórsdóttir & Nielsen (2018), Stinson & Bhat (2003) and Winters, Brauer, Setton, & Teschke (2010)). This factor is outside the scope of this study, since it is difficult to implement this factor using the method of this study, as discussed in Section 3.1.3.

Furthermore, the aforementioned literature indicated other infrastructural factors that may influence the route choice of cyclists, such as the number of righthand or lefthand turns a cyclists encounters along its trip (Broach, Dill, & Gliebe (2012) and Prato, Halldórsdóttir & Nielsen (2018)), number of intersections on a route (Chen (2016) and Prato, Halldórsdóttir & Nielsen (2018)), on-street parking (Li, Wang, Liu, & Ragland, 2012) and the presence of public transport facilities (Li, Wang, Liu, & Ragland (2012) and Koch & Dugundji (2021)). However, the first two factors were not included in this study for the same reason as the trip length. Moreover, the last two factors are indicated by Li, Wang, Liu, & Ragland (2012) as only contributing moderately to the route choice of cyclists and considering computation time these two factors are outside the scope of this study. Koch & Dugundji (2021) supports this by estimating that the positive influence of tram lines is caused by the collinearity between tram lines and the way cyclists find their way in the city centre.

Moreover, slope is not considered in this study, because the Netherlands is relatively flat compared to other countries (Esri, sd). Although quite some research proved the importance of slope, none of these studies were executed in the Netherlands. The studies that were executed with data from the Netherlands all considered the slope when providing background on environmental factors influencing the route choice of cyclists, but none included it in the analysis (Ton, Duives, Cats, & Hoogendoorn (2018), Koch & Dugundji (2021), Band (2022) and Bernardi, La Paix Puello, & Geurs (2018)).

2.2. Regression Modelling

As indicated by Huber et al. (2021), discrete choice models are a valuable tool for analysing the behaviour of individuals when faced with a choice between different alternatives.

In short, discrete choice models postulate the probability that individuals choose a given option. This probability depends on the relative attractiveness of the given options, which is a function of different environmental factors (Gkiotsalitis, 2021). Attractiveness of the alternatives is represented by utility (de Dios Ortúzar & Willumsen, 2011). Utility is defined as what an individual tries to maximise. In this study, attractiveness is defined as the probability a cyclist chooses a certain route. Several families of discrete choice models have been developed and applied, such as multinomial probit regression and logistic regression (de Dios Ortúzar & Willumsen, 2011).

The goal of regression modelling in this study is to find a relationship between the attractiveness of a certain route and the infrastructural and land use allocation characteristics of the routes. Regression modelling is used to determine the effects of one or more independent variables on a dependent variable (Khandelwal (2020) and Washington, Karlaftis, & Mannering (2011)). In the case of this study, the independent variables are the factors influencing bicycle route choice and the dependent variable is the attractiveness of a route.

An example of regression modelling is logistic regression modelling, which is often indicated as the simplest and most popular practical discrete choice model (Domencich & McFadden, 1975). Logistic regression can be used when the dependent variable is binary. This means that for the dependent variable, there are only two options (Washington, Karlaftis, & Mannering, 2011). As the data that used in this study consists of two options, logistic regression is applicable for the study. The observed routes can be associated with a one, when translated to a binary dependent variable, as these are the chosen routes. The shortest paths are associated with a zero, when translated to a binary dependent variable, as these routes are not chosen. However, the outcome of the model is not binary, as the outcome of the model, the dependent variable, is the probability a cyclist chooses a certain route. The equation associated with logistic regression modelling is presented in Equation 1. As presented in the equation, logistic regression models take the natural logarithm of the probability as a regression function of the independent variables (LaValley (2008) and Washington, Karlaftis, & Mannering (2011)).

EQUATION 1

$$\mathrm{Y_i} = \mathrm{log}it(P_i) = LN\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_K x_{K,i}$$

Based on the logistic regression model, the probability that the outcome takes value one, associated with the chosen route, can be determined using Equation 2.

EQUATION 2

$$P_i = \frac{e^{\beta_0+\beta_1 x_{1,i}+\beta_2 x_{2,i}+\cdots+\beta_K x_{K,i}}}{1+e^{\beta_0+\beta_1 x_{1,i}+\beta_2 x_{2,i}+\cdots+\beta_K x_{K,i}}}$$

In Equation 1 and Equation 2, the factors β_i represent the coefficients of a certain infrastructural or land use allocation factor that can be estimated doing regression analysis and $x_{k,i}$ stands for the prediction factor of a certain infrastructural or land use allocation factor (Gkiotsalitis (2021) and Washington, Karlaftis, & Mannering (2011)).

When using logistic regression, it is important to realise that several assumptions have to be made. First of all, the data on which the regression model is based should not contain outliers. Next to that, there should be no correlation between the independent variables (no multicollinearity) (Khandelwal (2020) and Huber, et al. (2021)).

Firstly, outliners will not occur in the data of this study as the dependent variable is binary. Secondly, the correlation between two independent variables can be examined using Pearson r correlation. This coefficient can be calculated by the use of Equation 3.

EQUATION 3

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

The Pearson r correlation coefficient runs from -1 to +1 where a negative value indicates a negative correlation and a positive value indicates a positive correlation. A value of ±0.2 indicate a weak correlation, a value of ±0.5 indicate a moderate correlation, a values of ±0.8 indicate a strong correlation and a value of ±1 indicates a perfect correlation (Ratnasari, Nazir, Toresano, Pawiro, & Soejoko, 2015). Other research indicated that when the Pearson r coefficient indicates a strong correlation, one of two independent variables that are examined should be eliminated (Huber, et al., 2021).

When all independent variables showing multicollinearity are eliminated, regression analysis can be performed to obtain the β_i coefficients of Equation 1. However, not all infrastructural and land use allocation factors influence the regression model in a significant way. In order to obtain an easier formula, some less influential factors could be excluded from the model. For this purpose, the stepwise approach is designed (de Dios Ortúzar & Willumsen, 2011). This approach makes use of the significance of all the factors using the P value. This is calculated using a t-test. If the P value is below alpha, the factor is statistically significant. Alpha is determined using the confidence interval that is chosen for the stepwise approach (Kwok, 2021). Within the stepwise approach multiple iterations are performed, until all included factors are statistically significant. There are multiple ways to perform this stepwise regression step, for example backwards elimination and forward entry (IBM, 2021). Backwards elimination starts with all initial factors being included. Then, in every step, the factor with the highest P value is eliminated, until all factors are statistically significant. In contrast, forward entry adds the factor with the lowest P value to the equation, until no statistically significant factors can be added to the model (Kwok, 2021). Finally, the β_i coefficients of the factors as mentioned in Equation 1 are estimated.

2.3. Conceptual Model

In order to quantify the infrastructural and land use allocation factors that have an influence on the route choice of cyclists, a model is be constructed. To get a better understanding of the steps that should be taken to obtain conclusions for this research, a conceptual model is provided in Figure 4. The steps described in the conceptual model is described in more depth in Section 3. Nevertheless, the model is shortly introduced in the remainder of this section. Furthermore, this section provides an analysis on the data used in this study.

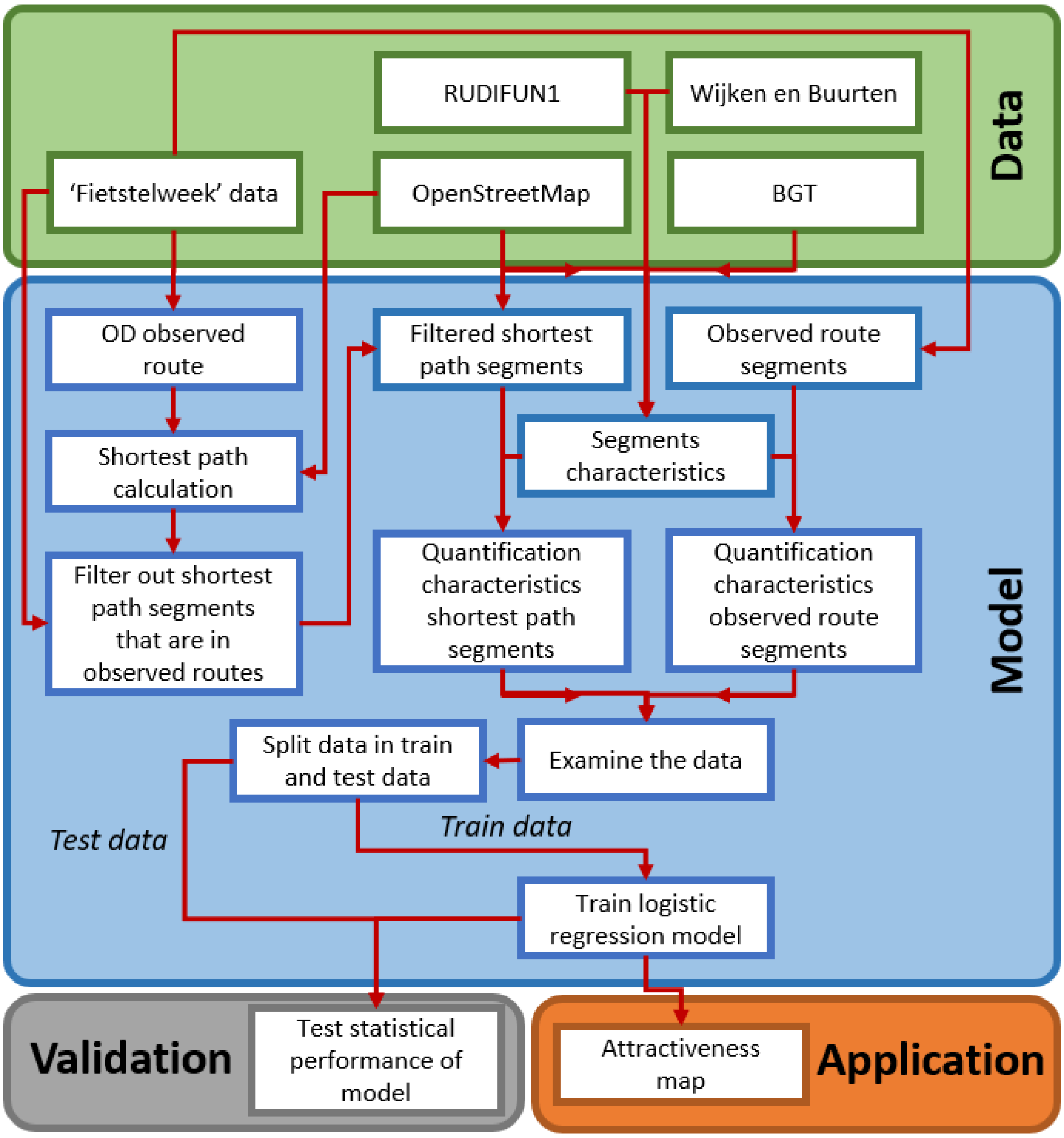

FIGURE 4: CONCEPTUAL MODEL

2.3.1. Model

The model is written by the use of a Python script, which is made from scratch. The segments of the observed route comes from 'fietstelweek' data (Breda University of Applied Sciences, sd) and the shortest path route is calculated by the model itself. Subsequently, the segments of the shortest path that are also in the corresponding observed route are eliminated. After this, the segment method is used to compare segments. More background on this method can be found in Section 3.1.3. Consequently, the data is examined and split in train and test data. Then the train data is used to train the logistic regression model and the test data is used to validate the model on its statistical performance. Moreover, the trained logistic regression model is used to generate an attractiveness map.

2.3.2. DATA

The model uses several data sources to gather data about the observed route of cyclists and the infrastructural and land use allocation characteristics of the observed route and the shortest path. For this study only open-sourced data is used. This makes it easier to acquire the data for this research, but also for future continuation of this study. The data sources are summarised in Table 2. In the remaining of this chapter, these data sources are described in more depth.

TABLE 2: OVERVIEW OF DATA SOURCES

Data	Source	Use case
OpenStreetMap	(Openstreetmap, sd)	Infrastructural (presence of traffic light, separate cycle path, cycle lane, artificial lighting, surface material and road classification) characteristics
Basisregistratie Grootschalige Topografie	(PDOK, sd)	Greenery allocation
Wijken en Buurten	(Centraal Bureau voor de Statistiek, 2021)	Degree of urbanisation
RUDIFUN1	(Planbureau voor de Leefomgeving, sd)	Land use (residential, commercial, industrial and land use mix) allocation
Fietstelweek	(Breda University of Applied Sciences, sd)	Observed routes

Firstly, 'OpenStreetMap' data (Openstreetmap, sd) is an open-source project where anyone can contribute to the network of the data source. It is a data source of geographical information containing infrastructural information with a great level of detail over the globe. 'OpenStreetMap' data is updated continuously and this study uses the most recent version. However, due to the fact that anyone can contribute to the data, the data is not validated and therefore some elements in the network are misidentified. Although this downside need to be considered when implementing the result, the benefit of having details on relatively much information of segments and nodes preponderates.

Moreover, 'Basisregistratie Grootschalige Topografie' data (PDOK, sd) is a data source managed by the Dutch government and Dutch governmental institutions are even obligated to use this data source if they use maps. The data is detailed on 20 centimetres and provides data on all kinds of physical objects in the built environment. Anyone can send any flaws of the data to the government. However, the Dutch government validates all changes, which makes the data highly trustworthy.

Besides that, 'Wijken en Buurten' data (Centraal Bureau voor de Statistiek, 2021) is a data source of the 'Centraal Bureau voor de Statistiek' and obligated by the Dutch law to deliver trustworthy statistical data. 'Wijken en Buurten' data is one of the data they store and consists of data of all neighbourhoods, district and municipalities in the Netherlands and in this study the most recent version from 2021 is used. As the data is government owned and validated, the data is highly trustworthy.

Furthermore, 'RUDIFUN1' data (Planbureau voor de Leefomgeving, sd) is a data source of the 'Planbureau voor de leefomgeving'. This institute is organisational part of the government of the Netherlands and has an independent position. 'RUDIFUN1' data contains data on the built environment and more specific, which land use the buildings fall under and in this study the most recent open-sourced data is used from 2020. As the data is validated, the data is highly trustworthy.

Finally, 'fietstelweek' data (Breda University of Applied Sciences, sd) is already introduced in Section 1.4.3 and therefore not further mentioned in this section.

3. METHODOLOGY

The approach of this study is as follows: first the model is created according to the conceptual model described in Section 2.3. After the model is created, the model is validated via a literature comparison, the use of test data and by projecting the predictive ability of the model on another city. Finally, the model is used for another application, namely a network attractiveness map for the bicycle.

3.1. THE MODEL

The conceptual model described in Section 2.3 is the outline for the creation of the model. The model contains out of four main steps. Firstly, the shortest paths and observed routes are described in the model. Secondly, infrastructural and land use allocation characteristics are connected to the segments of the shortest paths and observed routes. Thirdly, a comparison between shortest paths and observed routes is made along the so-called segment approach. Finally, a regression model is constructed.

3.1.1. SHORTEST PATHS AND OBSERVED ROUTES

Firstly, the observed routes and shortest path need to be described in the model. For this, both data from the 'fietstelweek' (Breda University of Applied Sciences, sd) and 'OpenStreetMap' (Openstreetmap, sd) is used. The observed route could be described using the 'fietstelweek' data. This data provides the segments that are used by cyclists for a certain origin and destination. Next to that, the shortest route is calculated as this is no provided data. For this, the origin and destination combinations of the 'fietstelweek' data are used. Then, using the road network of 'OpenStreetMap', the shortest direct route is calculated by the means of the Dijkstra graph theory algorithm using the length of the segments as weights.

3.1.2. DESCRIBE CHARACTERISTICS

Secondly, all segments are described according to the infrastructural and land use allocation factors that influence the route choice of cyclists as was concluded upon in Section 2.1. This is done by connecting the 'OpenStreetMap' data (Openstreetmap, sd), the 'Basisregistratie Grootschalige Topografie' data (PDOK, sd), the 'Wijken en Buurten' data (Centraal Bureau voor de Statistiek, 2021) and the 'RUDIFUN1' data (Planbureau voor de Leefomgeving, sd) with the segments of the observed routes and shortest path. Consequently, the model is adapted such that it collects the data about the influencing characteristics and connects those characteristics to a segment. The rest of this section is going into depth on how all individual factors are included in the model.

3.1.2.1. TRAFFIC CONTROL INSTALLATIONS

Traffic control installations create extra stopping time for cyclists and this affects the attractiveness of the surrounding segments (e.g. Koch & Dugundji (2021)). 'OpenStreetMap' data (Openstreetmap, sd) provides nodes where a traffic control installation is present. However, traffic control installations do not only affect the directly connected segments, but also segments further from the traffic control installation. This is due to the fact that when bypassing a traffic control installation you have to change more than just one segment of your trip. This principle is visualised in Figure 5, in which it is indicated that another set of segments have to be taken already five segments in advance of the traffic control installation, when travelling from the Northwest to East part of the area.

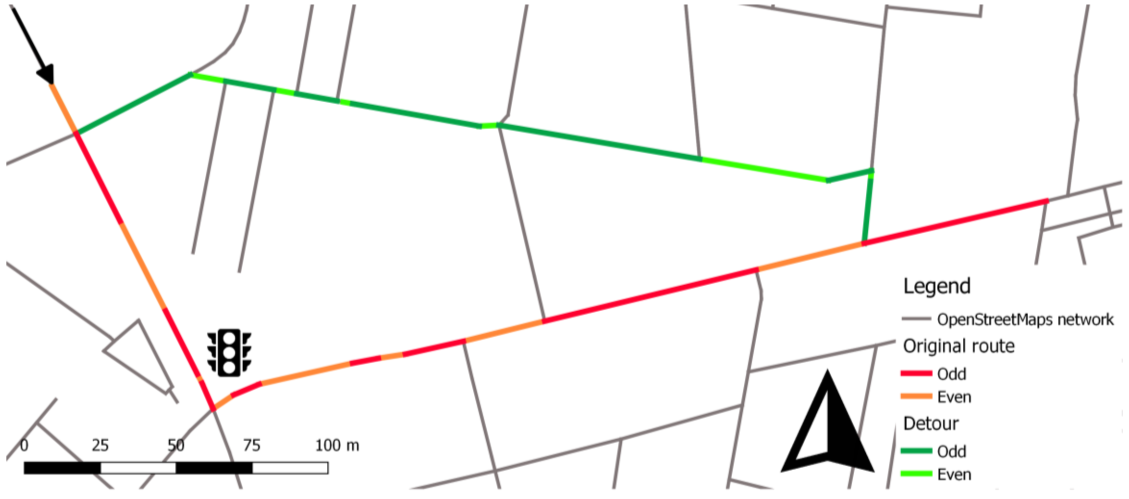

FIGURE 5: EXAMPLE OF USING A DETOUR AROUND A TRAFFIC CONTROL INSTALLATION WHERE EVERY SUBSEQUENT SEGMENT HAS ANOTHER COLOUR

Moreover, traffic control installations are only an obstruction for cyclists when they are traveling towards them. This is visualised in Figure 5 when traveling from the Northwest along the red-orange line, you are in line with the traffic control installation, so there is a high chance you will be obstructed by the traffic control installation. However, if you travel along the dark green-green line, you travel not in line with the traffic control installation and the chance of encountering the traffic control installation is relatively low. Therefore, only if a segment is in line with the closest traffic control installation the distance is considered.

Besides that, considering the regression modelling step, factors cannot be included on a continuous scale (Li S. , 2017). Therefore, the distance to a traffic light is made categorial. For this, the buffer dimensions as described in Strauss & Miranda-Moreno (2013) are used, namely closer than 50 metres, closer than 150 metres, closer than 400 metres, closer than 800 metres and 800 metres and beyond. If a traffic control installation is not in line with a segment, it is also allocated to this last category of 800 metres and beyond.

3.1.2.2. Separate Cycle Paths & Cycle Lanes

Research agrees to the substantiality of the effect of separate cycle paths and cycle lanes on the attractiveness of segments (e.g. Mertens et al. (2016)). Moreover, 'OpenStreetMap' data (Openstreetmap, sd) on the segments include data whether a segment is a separate cycle path or a cycle lane, so this data can be gathered directly from this source.

3.1.2.3. Artificial Lighting

After nightfall, artificial lighting creates a safer environment and can increase the attractiveness of roads (e.g. Winters, Davidson, Kao, & Teschke (2011)). Furthermore, 'OpenStreetMap' (Openstreetmap, sd) data on the segments include data whether a segment is artificially lit, so the data is gathered directly. However, the 'fietstelweek' (Breda University of Applied Sciences, sd) data does not contain data on the exact time a certain trip is cycled, making it impossible to make a distinction between trips after nightfall and trips during the day. Artificial lighting is assumed to have a substantially more influence on the attractiveness of segments after nightfall compared to during daytime. Following this reasoning, it is expected that artificial lighting has no significant contribution in the regression model.

3.1.2.4. *Paved Infrastructure*

The type of surface of infrastructure impacts the safety and the energy that is needed to travel along a segment, which effects the attractiveness of segments (e.g. Prato, Halldórsdóttir & Nielsen (2018)). Besides that, 'OpenStreetMap' (Openstreetmap, sd) data on the segments include different pavement types. In this study two types are included in the model, namely asphalt and paving stones. Both pavement types are visualised in Figure 6.

FIGURE 6: A) STREET WITH A PAVEMENT STONE SURFACE (MAKE BV, CIVIELTECHNISCH ADVIESBUREAU, SD), B) A STREET WITH AN ASPHALT SURFACE (DE BAAN, SD)

3.1.2.5. *Motorised Vehicle Intensities*

Large volumes of motorised vehicles influence the attractiveness of a segment (e.g. Broach, Dill & Gliebe (2012)). Unfortunately, there is no freely accessible data on motorised vehicle intensities available, so an indirect approach is used. Traffic volumes are often depicted by the class of the proxy as stated in the 'OpenStreetMap' (Openstreetmap, sd) data (e.g. primary road, secondary road, residential road) and the maximum speed allowed. Primary roads carry more motorised vehicles compared to secondary roads and are therefore less attractive to use by cyclists. Moreover, a primary road with a speed limit of 50 km/h is deemed more favourable compared to a primary road with a speed limit of 60 km/h, since higher vehicle speeds reduce the perceived safety of cyclists (Rasch, Moll, López, García, & Dozza, 2022). In the end, seven categories were considered including the category zero, when no motorised vehicles are allowed. Although inaccurate, this class approach is suitable for this study as the goal of this factor is to verify if a road stretch has a high motorised vehicle intensity and not to have an exact number for the intensity. Moreover, using a categorial approach, the data can be used in the regression step with ease (Li S. , 2017).

3.1.2.6. *Bicycle Intensities*

Overtaking and maintaining the same speed is harder at segments with high bicycle intensities, indicating that bicycle intensities effect segment attractiveness (e.g. Li, Wang, Liu, & Ragland (2012)). Bicycle intensities are gathered from the 'FietsMonitor' of Witteveen+Bos. Although this model has some drawbacks, as described in Section 1.1, it is deemed a suitable source to provide an indication of the bicycle intensities. Besides that, considering the regression modelling step, factors cannot be included on a continuous scale (Li S. , 2017) and the purpose of the bicycle intensities is to indicate a sense of crowdedness on the infrastructure, which is better represented by a categorial approach. Therefore, the data was made categorial using five categories with intervals as 0, 250, 500, 1000 and 2000 bicycles per day. Using these categories, the bicycle intensities are presented in Figure 7.

FIGURE 7: A BICYCLE INTENSITY MAP OF NORTHWEST ENSCHEDE

It can be imagined that bicycle intensities themselves do influence the route choice of cyclists only to a certain degree. For example the width of the (bicycle) infrastructure can contribute to the route choice, since this influences the capacity of a certain stretch of road. However, this was outside of the scope of this research.

Moreover, the 'FietsMonitor' generates bicycle intensities using the origins and destinations of the municipality that was selected and neighbouring municipalities (Veenstra, 2022). However, this is not done for the neighbouring municipalities, resulting in an underestimation of the bicycle intensities in neighbouring municipalities. This made it only possible to model trips and part of trips that are in the municipality that was chosen, in this case Enschede, as introduced in Section 1.4.4.

3.1.2.7. Land Use

The type of land use may influence the attractiveness of a segment (e.g. Zhao, Ke, Lin & Yu (2020)). In this study, four land use classes are considered, namely residential, commercial, industrial and green land use.

The first three are obtained using 'RUDIFUN1' data (Planbureau voor de Leefomgeving, sd). This data source provides highly detailed polygons on land use classes of the building environment as visualised in Figure 8. The land uses are considered by taking the ratio between the area covered with a certain land use class and the total buffer area. Important to note is that the polygons in Figure 8 represent a two-dimensional area and therefore do not consider the height of the building and the usage of every floor. Moreover, some larger polygons represent multiple land use classes if they are situated in the same building. This was considered in this study by identifying the whole polygon for both land use classes. When analysing the results, both two flaws should be considered.

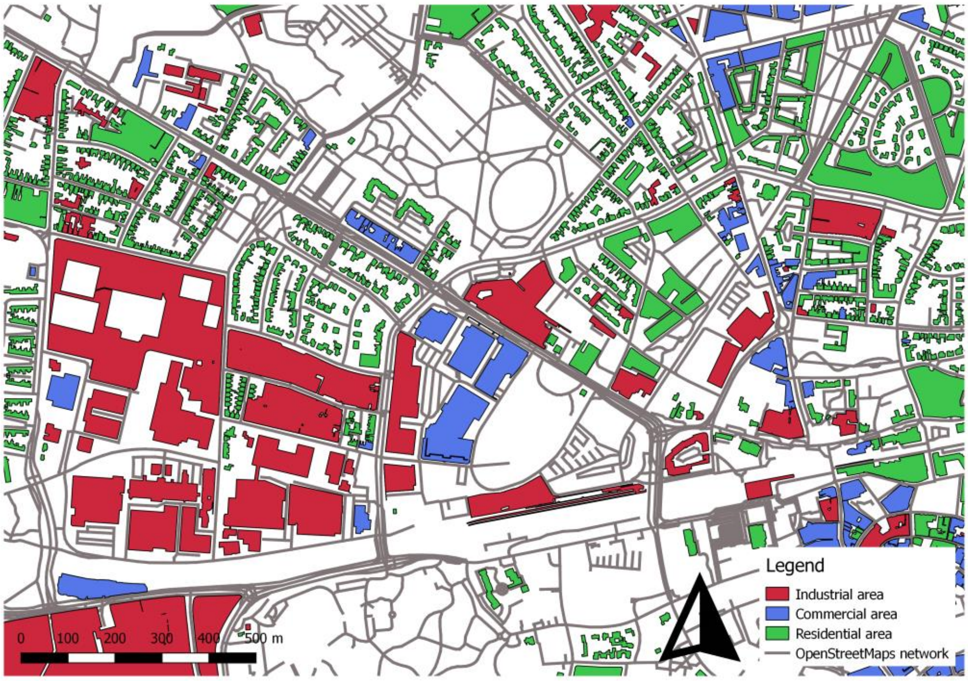

FIGURE 8: A LAND USE MAP OF NORTHWEST ENSCHEDE (NOTE THAT LAND USES MIGHT OVERLAP)

In order to derive the type of building environment a crow-fly buffer of 250 meter is used around the segments. This buffer size is chosen as it is used in two previous studies by Winters, Teschke, Grant, Setton & Brauer (2010) and by Winters, Brauer, Setton & Teschke (2010) and as middle-size buffer it is useful in both dense urban areas as well as more rural areas. The buffer is designed as the area on both sides of a segment. Using this buffer design there is little overlap with other segments.

In the determination of greenery, also this buffer is used. However, for greenery 'Basisregistratie Grootschalige Topografie' data (PDOK, sd) is used. This data source provides highly detailed polygons for all kinds of greenery in Enschede including trees, heathland, greening and more as visualised in Figure 9. Although the data is highly detailed, there is no information far outside the municipality borders of Enschede. Therefore, the municipality of Enschede is used as study area as introduced in Section 1.4.4. For the purpose of this study, the kind of greenery is of less importance as it is being assumed that a cyclist perceives all kinds of greenery equally. Also green area is considered in the model by taking the ratio between the area covered with greenery and the total buffer area.

FIGURE 9: A GREENERY MAP OF THE NORTHWEST OF ENSCHEDE

3.1.2.8. LAND USE MIX

The mixture of land use in a certain area may have an influence on the number of bicycle trips (e.g. Saelens, Sallis & Frank (2003)). The land use mix can be determined using a diversity index, namely the Shannon index. The equation associated with the Shannon index is presented in Equation 4 and is often used by other researchers (Winters, Teschke, Grant, Setton, & Brauer (2010), Zhao, Ke, Lin, & Yu (2020) and Strauss & Miranda-Moreno (2013)).

EQUATION 4

$$S = \frac{-\sum_k p_i \ln p_i}{\ln k}$$

In Equation 4, p_i is the area of a certain land use class and k is the total number of land use classes. The land use mix makes use of the land use classes as described in Section 3.1.2.7 and a value close to 1 indicates an equally divided mixture of land uses.

3.1.2.9. FLOOR AREA RATIO

The floor area ratio, often associated with the density, around a segment could have an influence on the attractiveness of a certain segment (e.g. Chen (2016)). 'Wijken en buurten' data (Centraal Bureau voor de Statistiek, 2021) does provide data on the density of addresses of a certain neighbourhood both continuous and categorical. In this study, the categorial data is more interesting as it is tried to show the difference between an urban environment and a less urban environment. The 'Wijken en buurten' data consists of the categories as presented in Table 3.

TABLE 3: DEGREE OF URBANISATION CATEGORIES IN 'WIJK EN BUURTEN' DATA (CENTRAAL BUREAU VOOR DE STATISTIEK, 2021)

Class number	Degree of urbanisation	Address density [addresses/km^2]
1	Very highly urban	More than 2,500
2	Highly urban	Between 1,500 and 2,500
3	Moderate urban	Between 1,000 and 1,500
4	Little urban	Between 500 and 1,000
5	Not urban	Less than 500

As visualised in Figure 10, primary roads and other substantial roads divide Enschede in multiple neighbourhoods. Nevertheless, it is possible to cycle over most of these roads, so cyclists are influenced by the degree of urbanisation on both sides of the road. Therefore, the same buffer as for the different land use classes is used around the segment is chosen and the neighbourhoods that overlap with this buffer are considered in determining the degree of urbanisation.

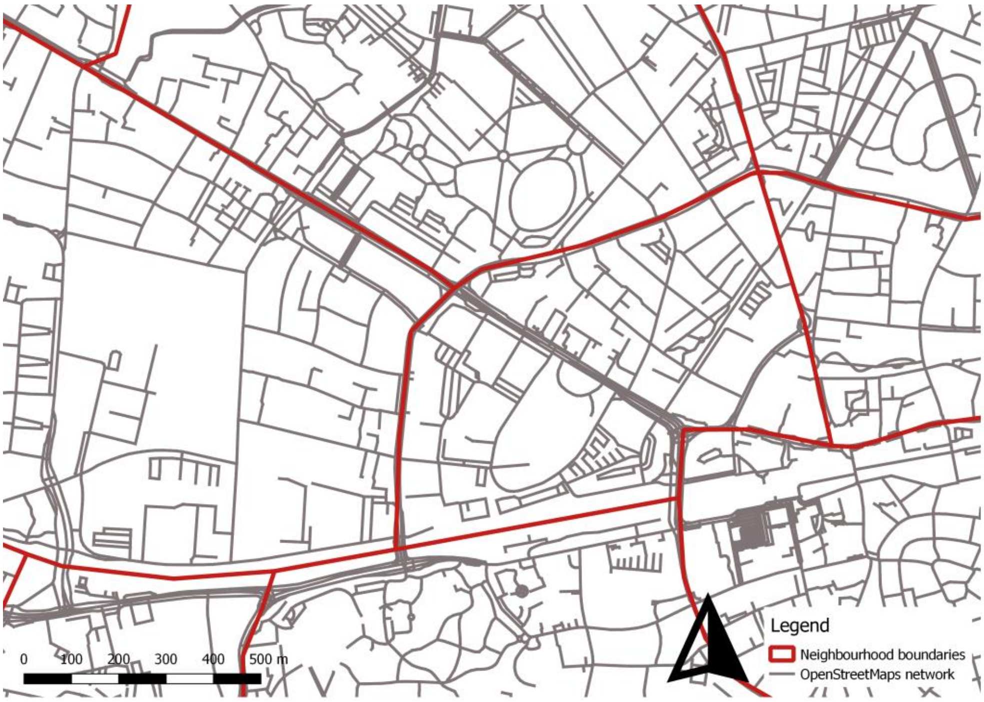

FIGURE 10: MAP OF THE NORTHWEST OF ENSCHEDE REPRESENTING THE NEIGHBOURHOOD BOUNDARIES

3.1.3. Segment Approach

After all segments are characterised by the model, a comparison is made between the segments. This idea is visualised in Figure 11. This method is called the segment method, as this method compares characteristics on a segment level. The comparison is between the segments that are chosen by cyclists (segments of observed routes) and the segments that are not chosen but are recommended when using the shortest path. Segments of the shortest path overlapping with segments of the corresponding observed route are filtered out. In Figure 11, the segments of the observed routes are shown with a dotted line. The solid line represents the shortest path. Segments that are not chosen but are recommended when using the shortest path are visualised in Figure 11 by segments 1 (blue), 2 (blue), 3 (green), 7 (green), 8 (green), 9 (green), 13 (orange) and 14 (orange). In this study, a comparison is made between these segments and the segments of the observed routes.

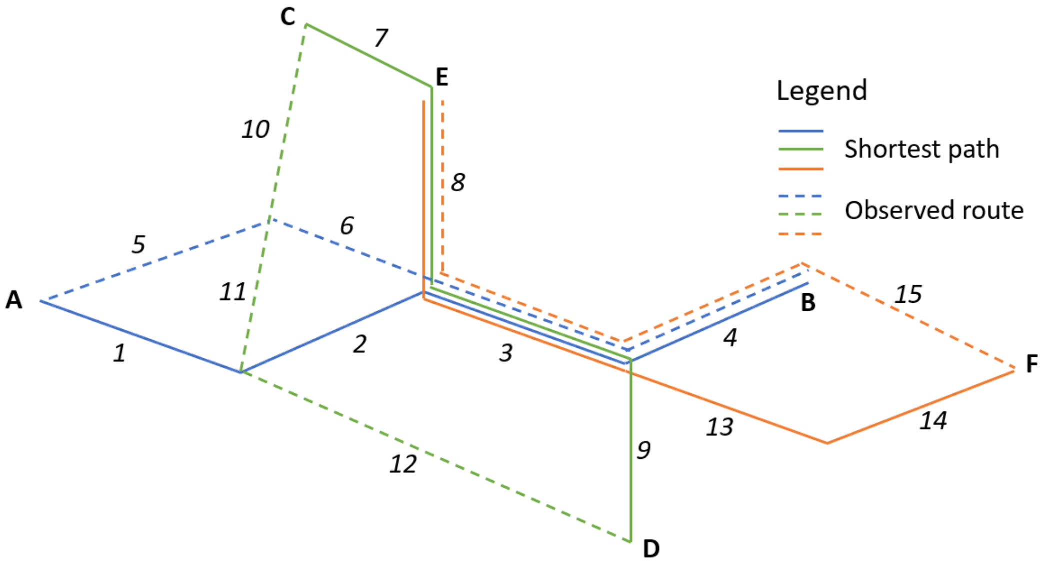

FIGURE 11: VISUALISATION OF THE SEGMENT APPROACH. OBSERVED ROUTES ARE VISUALISED BY A DOTTED LINE AND THE SHORTEST PATH IS VISUALISED BY A SOLID LINE

The segment approach accounts for a problem that occurs when a route approach, an approach where an individual shortest path and observed route are compared, is used, namely when the shortest path and the observed route use the same segments. This is a problem, since no comparison can be made between coinciding routes and therefore cannot be included in the analysis, although this is the most ideal scenario.

However, the segment approach has its downsides, namely route characteristics are hard to include. For example, when one segment is chosen by cyclists for a certain set of characteristics, a cyclists will most likely use the next segments that are in line with the destination. Moreover, trip length cannot be included using a segment approach in a direct way as only segments are compared. However, as mentioned in Section 2.1, trip length is a significant factor influencing the route choice of cyclists as cited often and could be included in a more indirect way using classes of ratios between the length of the observed route and the shortest path in a multinomial regression model. However, that approach is out of the scope of this study. Nevertheless, the aforementioned limitations should be considered when analysing the results.

3.1.4. Regression Modelling

The model is concluded with the regression modelling step. Regression modelling was already introduced in Section 2.2. Before regression modelling can be implemented on the results of the previous step, it must be sure that all conditions described in Section 2.2 are met.

Firstly, possible outliers must be identified. However, since the dependent variable is binary and the independent variables are non-continuous, outliers will not occur as introduced in Section 2.2.

Moreover, the correlation is examined using the Pearson r coefficient. This is done between all variables that are included in the study, which are determined in the step described in Section 2.1. When the Pearson r coefficients exceeds ±0.8, one of the two independent variables should be eliminated before executing regression modelling.

When all conditions for logistic regression modelling are met, the regression analysis can be performed. Using machine learning, the model tries to fit the train data by the use of a solver. In the process, insignificant factors could be excluded from the model. For this a stepwise approach is used, namely the backward elimination, as de Dios Ortúzar & Willumsen (2011) recommend. The confidence interval used in this approach is 95% as this is commonly done in other research where regression modelling in a traffic environment is used (Jakovljevic, Paunovic, & Belojevic (2009), Ang, Chen, & Lee (2017), Washington, Karlaftis, & Mannering (2011), Ton, Duives, Cats, & Hoogendoorn (2018), Winters, Brauer, Setton, & Teschke (2010) and Li, Wang, Liu, & Ragland (2012)). Finally, the β_i coefficients in Equation 1 are estimated and included in the regression modelling and a formula for the attractiveness of segments is constructed.

Although the β_i coefficients of the regression model are estimated, they do not indicate the relative contribution of the factors, since all factors do not use the same scale or measurement unit. Standardised regression coefficients eliminate this problem by using units of standard deviations and give insides in the relative importance of the influential factors. Standardised regression coefficients are calculated using Equation 5 (Siegel & Wagner, 2022).

EQUATION 5

$$\beta_{st,i} = \beta_i \frac{\sigma_{x,i}}{\sigma_y}$$

In Equation 5, β_i is the regression coefficient as mentioned in Equation 6, $\sigma_{x,i}$ is the standard deviation of the input data for the β_i coefficient and σ_y is the standard deviation of the dependent variable.

3.2. Validation of the Model

Validation of the model is important to demonstrate if the model possesses a satisfactory range of accuracy consistent with the intended application of the model (MacLeod, 2022). In this study, the model is validated via three ways. Firstly, if the model is consistent with background knowledge in literature and/or adds something to the existing data. Secondly, if the model gives accurate predictions needed for its uses. Finally, if the model gives accurate predictions on another study area.

As presented in Section 2.1, there is already quite some existing literature in infrastructural and land use allocation factors influencing the route choice of cyclists. The outcomes of the model make more sense if they are in line of multiple existing papers. Therefore, the results are compared with earlier obtained literature.

Secondly, the data for the model is split into training and test data (Polamuri, 2017). The model is made in such a way that it fits the training data as good as possible as mentioned in Section 3.1.4. Both the training and the test data are from the same city, namely Enschede and are split in a 7:3 ratio.

Subsequently, the test data can be used to evaluate the statistical performance of the model. This performance can be visualised using a confusion matrix and by extracting four key performance indicators. The accuracy indicates the percentage of predictions being true. The precision indicates the percentage of positive predictions being true. The sensitivity indicates the percentage of negative predictions being true. Finally, the F1 score is the harmonic mean between the precision and the sensitivity and is an overall metric that incorporates both the precision and the sensitivity (Behesthi, 2022). Scores close to 100 for all key performance indicators contributes to the validity of the model.

Finally, the performance of the model on data from another city is tested. For this, the city of Haarlem is chosen, due to data availability and since it is located at the other side of the Netherlands, in the province of North-Holland (see Figure 12). This substantial difference in location in the Netherlands contributes to the generality of the model, if the model performs well for the city of Haarlem. There are 7,776 observed routes included in 'fietstelweek' data (Breda University of Applied Sciences, sd) that have at least the origin or the destination in the municipality of Haarlem.

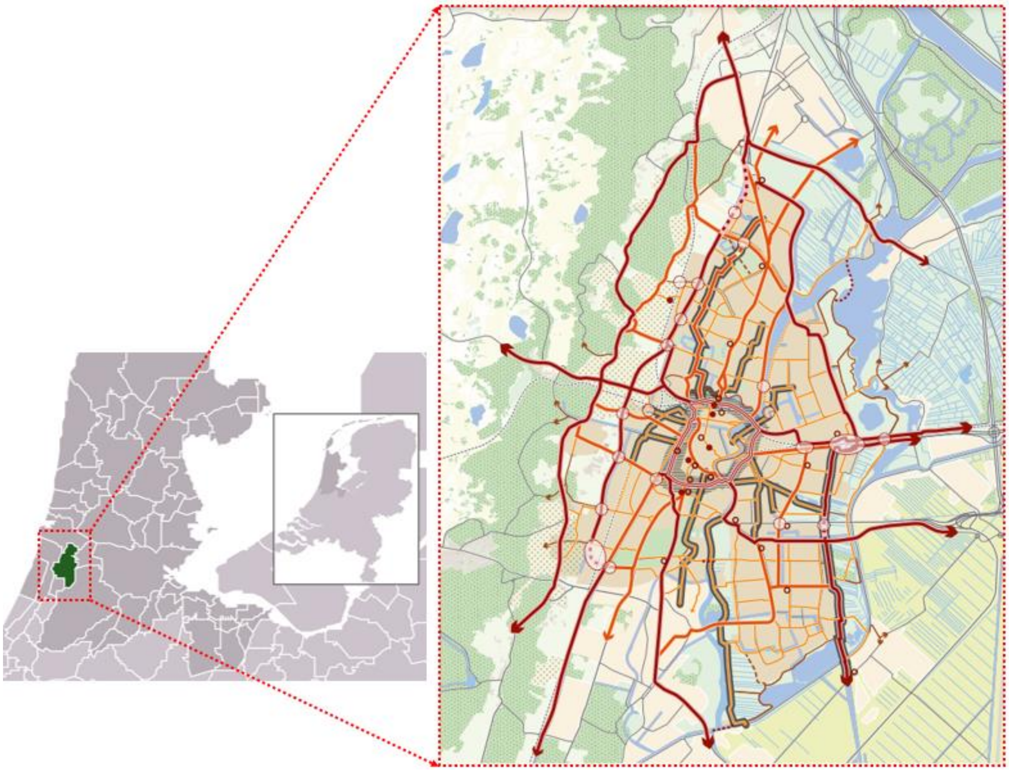

FIGURE 12: A) MUNICIPALITY OF HAARLEM, AS VISUALISED IN GREEN IN THE PROVINCE OF NORTH-HOLLAND (CENTRAAL BUREAU VOOR DE STATISTIEK | TOPOGRAFISCHE DIENST KADASTER, 2005), B) MAIN BICYCLE NETWORK OF HAARLEM INCLUDING REGIONAL (RED) AND MAIN (ORANGE) CYCLEWAYS (GEMEENTE HAARLEM, 2017)

3.3. Application of the Model

Witteveen+Bos is interested in the outcomes of the model in order to improve the route choice algorithm in the 'FietsMonitor', as mentioned in Section 1.1. The outcomes, if proven valid, are of great value for that purpose. Besides that, the model can be used for other purposes. The model provides a formula for the attractiveness of a certain segment. Except from evaluating segments and bicycle routes, it can also evaluate the whole road network using this formula, providing an attractiveness for all segments that can be used for cyclists. This can be helpful for policy makers to identify routes that are not attractive for cyclists according to the factors included in the formula. In order to be easy to interpret the results, this information is visualised in the form of a map.

4. Results

In this section, the results of the executed methodology are presented, starting with the results of the model. Thereafter, the results of the validation of the model and the application of the model are presented successively.

4.1. The Model

As mentioned in Section 3.1, the model consists of four main steps. Although all steps are relevant towards the final regression model, they do not produce results as interesting as the last regression model step. Therefore, this section shows predominantly results of this step.

4.1.1. Verify Multicollinearity

The multicollinearity of the factors is examined using the Pearson's r coefficient as calculated between all factors. The results are visualised using a heatmap and presented in Figure 13. As mentioned in Section 2.2, pairs of factors with a high collinearity (outside ± 0.8) should be examined and one of the factors of the pair should be dropped. However, as presented in Figure 13, this is not the case for the factors included in this study.

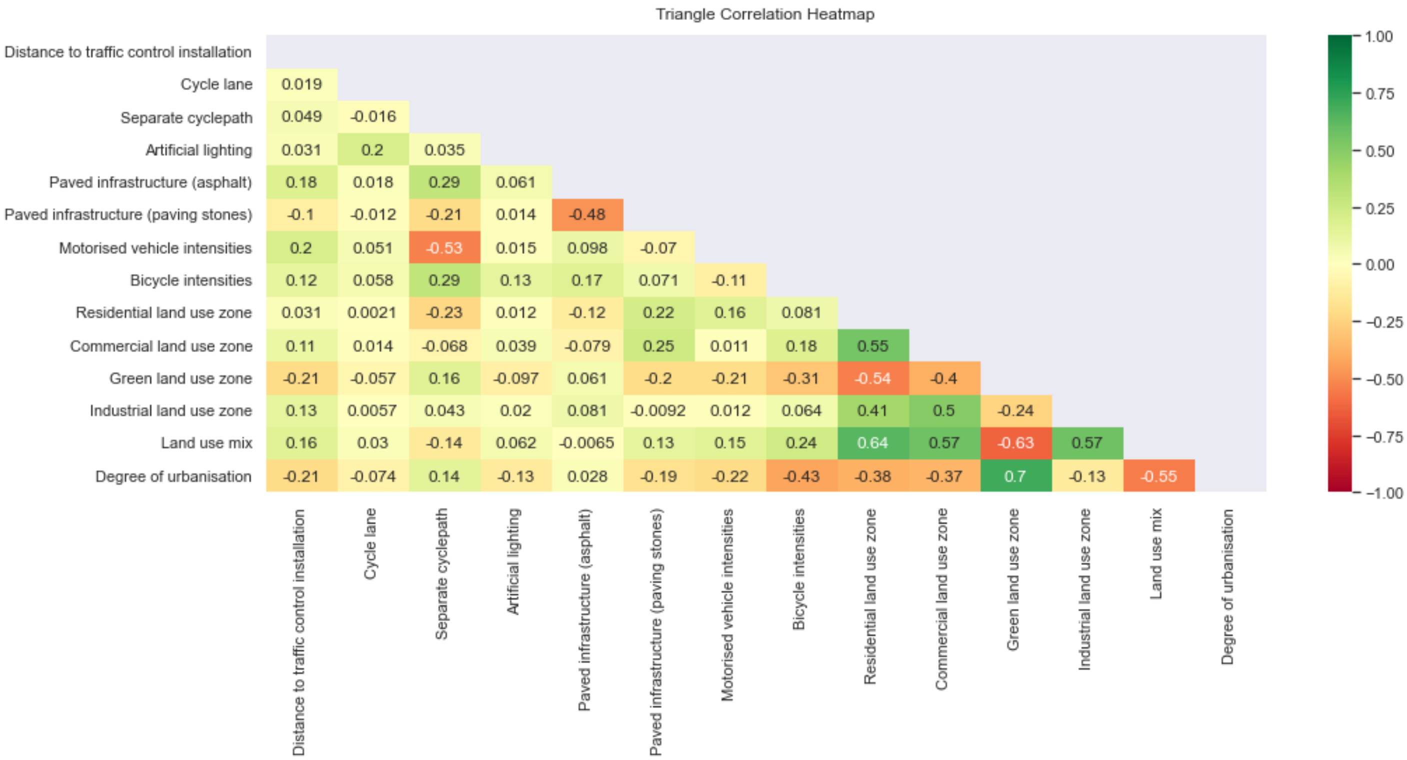

FIGURE 13: MULTICOLLINEARITY HEATMAP

Nevertheless, some factors showed a substantially higher collinearity compared to others. However, these can be certified. Firstly, the factor '*Land use mix*' shows a high collinearity with all land use classes. This is logical, since the land use mix is calculated using the other land use classes as input. In addition, the other land use classes show a high collinearity with the factor *'Degree of urbanisation'* and with each other. This may be caused by the fact that if the degree of urbanisation increases (density decreases), there is less build-up area including residential, commercial and industrial buildings, but there is more greenery. Moreover, the factor *'Motorised vehicle intensities'* shows a high negative collinearity with the factor *'Separate cycle paths'*. This can be explained, since on separated cycle paths, motorised vehicles are not allowed. The same idea holds for the collinearity between both surface materials, asphalt and paving stones, since asphalt and paving stones cannot occur at the same time.

4.1.2. Stepwise Regression

Subsequently, the regression model was constructed using the same factors as included in the multicollinearity analysis. For the solver of the regression model, multiple solvers have been tested, namely the ones indicated by the description of the logistic regression model module that has been used in this study (statsmodels, sd). All solvers resulted in a similar pseudo R-squared value, so could be used. This value can be used to compare models with each other considering both models using the same data and are predicting the same kind of outcome (UCLA, sd). Finally, the Newton-Raphson ('newton') solver is used.

In Table 4, the first iteration of the regression model without the use of stepwise regression is presented. That table indicates a low P value for most factors, except the presence of paving stones and the mixture in land use.

TABLE 4: STATISTICAL SUMMARY OF REGRESSION MODEL

Factor	β	Std. Error	P value
Constant	0.106	0.018	0.000
Infrastructural			
Distance to traffic control installation	-0.037	0.003	0.000
Cycle lane	0.374	0.036	0.000
Separate cycle path	0.866	0.011	0.000
Artificial lighting	-0.339	0.012	0.000
Paved infrastructure (asphalt)	0.591	0.008	0.000
Paved infrastructure (paving stones)	0.022	0.012	0.075
Motorised vehicle intensities	-0.123	0.002	0.000
Bicycle intensities	0.317	0.003	0.000
Land use allocation			
Residential land use zone	1.764	0.032	0.000
Commercial land use zone	-1.483	0.037	0.000
Greenery land use zone	-0.334	0.018	0.000
Industrial land use zone	-0.762	0.036	0.000
Land use mix	0.041	0.003	0.024
Degree of urbanisation	0.092	0.018	0.000

After one iteration of backward elimination, in which the factor *'Paved infrastructure (paving stones)'* was eliminated, all factors were statistically significant. However, the substantially low P value might not only be the cause of the statistical significance of all factors, but also caused by the large sample size that is used as input for the logistic regression (Adedokun, 2008). In that case, P values do not represent the statistical significance of a factor well and therefore cross-validation is added to verify the statistical significance of all factors.

Cross-validation is implemented using the Monte Carlo method. This method is used to account for risks in quantitative analysis and relies on repeated random sampling of the dataset (Palisade (sd) and Raychaudhuri (2008)). In this study, a 10% sample (± 100,000 data points) of the dataset is randomly sampled and a new regression model is generated. This procedure is done 100 times resulting in a mean and 95% confidence interval for all β_x coefficients over the 100 separate models as presented in Table 5, I. Since, 10% samples still consist of a substantially large dataset, the aforementioned procedure is also executed using a 1% sample (± 10,000 data points) and a repetition of 1,000 times and the results are visualised in Table 5, II.

TABLE 5: STATISTICAL SUMMARY OF THE MONTE CARLO METHOD

Factor	β model	I Mean β	I Confidence interval (95%)	II Mean β	II Confidence interval (95%)
Constant	0.099	0.096	[0.085, 0.106]	0.095	[0.084, 0.105]
Infrastructural					
Distance to traffic control installation	-0.036	-0.035	[-0.037, -0.034]	-0.034	[-0.035, -0.032]
Cycle lane	0.376	0.402	[0.382, 0.422]	0.428	[0.404, 0.452]
Separate cycle path	0.868	0.873	[0.867, 0.879]	0.869	[0.862, 0.876]
Artificial lighting	-0.340	-0.331	[-0.338, -0.323]	-0.326	[-0.333, -0.319]
Paved infrastructure (asphalt)	0.596	0.589	[0.585, 0.594]	0.596	[0.591, 0.600]
Motorised vehicle intensities	-0.123	-0.121	[-0.123, -0.120]	-0.124	[-0.125, -0.123]
Bicycle intensities	0.316	0.318	[0.316, 0.320]	0.318	[0.316, 0.320]
Land use allocation					
Residential land use zone	1.762	1.796	[1.776, 1.815]	1.787	[1.767, 1.808]
Commercial land use zone	-1.493	-1.473	[-1.497, -1.451]	-1.468	[-1.490, -1.445]
Green land use zone	-0.331	-0.333	[-0.344, -0.322]	-0.332	[-0.344, -0.320]
Industrial land use zone	-0.758	-0.765	[-0.786, -0.743]	-0.760	[-0.782, -0.739]
Land use mix	0.043	0.027	[0.017, 0.037]	0.037	[0.026, 0.047]
Degree of urbanisation	0.093	0.094	[0.085, 0.106]	0.095	[0.093, 0.097]

From the Monte Carlo approach can be concluded that the factors have relatively small confidence intervals close to the β_x coefficients of the model of the full data set, so the β_x coefficients are considered statistically significant.

Finally, the β_i coefficients are estimated and their statistical significance is tested. These coefficients are substituted in the logit presented in Equation 1, resulting in Equation 6.

EQUATION 6

$$Y_i = 0.099 - 0.036x_{TCI} + 0.376x_{CL} + 0.868x_{CP} - 0.340x_{AL} + 0.596x_{Sa} - 0.123x_{Im} + 0.316x_{Ib} + 1.762x_{res} - 1.493x_{com} - 0.331x_{gre} - 0.758x_{ind} + 0.043x_{LUM} + 0.093x_{DU}$$

In Equation 6 x_{TCI} is the distance to a traffic control installation, x_{CL} is the presence of a cycle lane, x_{CP} is the presence of a separate cycle path, x_{AL} is the presence of artificial lighting, x_{Sa} is the presence of asphalt as surface material, x_{Im} is the motorised vehicle intensity, x_{Ib} is the bicycle intensity, x_{res} is the residential area ratio, x_{com} is the commercial area ratio, x_{gre} is the green area ratio, x_{ind} is the industrial area ratio, x_{LUM} is the land use mix and x_{DU} is the degree of urbanisation.

Subsequently, β_i coefficients do not indicate the relative contribution of the factors and therefore the standardised regression coefficients are calculated using Equation 5. The standardised coefficients are presented in Table 6.

TABLE 6: STANDARDISED REGRESSION COEFFICIENTS

Factor	Standardised β
Constant	0
Infrastructural	
Distance to traffic control installation	-0.124
Cycle lane	0.095
Separate cycle path	1.072
Artificial lighting	-0.246
Paved infrastructure (asphalt)	0.699
Motorised vehicle intensities	-0.675
Bicycle intensities	1.108
Land use allocation	
Residential land use zone	0.689
Commercial land use zone	-0.446
Greenery land use zone	-0.227
Industrial land use zone	-0.232
Land use mix	0.034
Degree of urbanisation	0.365

From Table 6, it can be concluded that the presence of a separate cycle path, paved infrastructure, the motorised vehicle intensities and the area of residential land use have the most substantial importance in the regression model.

4.2. Validation of the Model

The model is validated according to three methods, namely if the model is consistent with background knowledge in literature and/or adds something to the existing literature, if the model gives accurate predictions needed for its uses and if those predictions can be projected on another city, in this case Haarlem.

4.2.1. Literature Consistency

Firstly, the β_i coefficients as presented in Equation 6 are validated for their consistency with literature. As presented in Table 7, not for all factors are consistent with. For the factors '*Artificial lighting*', '*Bicycle intensities*' and '*Green land use zone*' this makes it harder to validate their consistency with literature, but can provide new insides. Although those factors are inconsistent with literature, there might be an explanation why those factors have a different sign.

TABLE 7: LITERATURE CONSISTENCY OF THE MODEL

Factor	Expected sign	Sign in results	Literature validation
Infrastructural			
Distance to traffic control installation	+/±/-	-	e.g. Koch & Dugundji (2021)
Cycle lane	+	+	e.g. Chen (2016)
Separate cycle path	+	+	e.g. Mertens et al. (2016)
Artificial lighting	+	-	-
Paved infrastructure (asphalt)	+	+	e.g. Winters, Davidson, Kao & Teschke (2011)
Motorised vehicle intensities	-	-	e.g. Broach, Dill & Gliebe (2012)
Bicycle intensities	-	+	-
Land use allocation			
Residential land use zone	+/±/-	+	Zhao, Ke, Lin & Yu (2020)
Commercial land use zone	+/±/-	-	Winters, Brauer, Setton & Teschke (2010)
Greenery land use zone	+/±	-	-
Industrial land use zone	±/-	-	Prato, Halldórsdóttir & Nielsen (2018)
Land use mix	+	+	e.g. Maat, van Wee & Stead (2005)
Degree of urbanisation	+/-	+	e.g. Saelens, Sallis & Frank (2003)

Firstly, it was expected that the presence of artificial lighting would have a positive influence on the attractivity of segments, however the model indicated a reversed relationship. Moreover, the relationship is proven to be statistical significant, which was also not expected. This reversed relationship and statistical significancy may be caused by the unavailability of data on the time when the observed routes were used as suggested in Section 3.1.2.3. This might resulted in artificial lighting not being the influential factor, but since the dataset is substantially large, the regression model algorithm found a relationship between the segments equipped with artificial lighting and the attractiveness of roads. In addition, the reversed relationship may also be caused by the allocation of cyclists on the network. In the observed route cyclists are all allocated to the correct position in the network, so if there is a primary road with a separate bicycle path next to it, the cyclist is allocated to this separate bicycle path. However, the shortest path calculation does not use any restriction and cyclists are often allocated on the primary road and primary roads are more often artificially lit resulting in a positive sign.

Secondly, it was expected that high bicycle intensities would have a negative influence on the attractivity of segments, however the model indicated a reversed relationship. This may be caused by the fact that segments along routes with a high bicycle intensities according to the 'FietsMonitor', also occur often in the observed route dataset from 'fietstelweek' (Breda University of Applied Sciences, sd). This indicates that those routes are perceived attractive, which result in a substantial number of observed routes using these segments. This resulted in a positive sign.

Finally, it was expected that area covered in green would have a positive influence on the attractivity of segments, however the model indicated a reversed relationship. This may be caused by the fact that shortest paths are calculated without the use of restrictions. This may result in shortest paths using parks and other paths through greenery, which are prohibited to cycle on and result in a negative sign.

4.2.2. Model Performance

Following the model's literature consistency, the prediction performance of the model is determined. Firstly, the performance of the model excluding trip length is determined. The confusion matrix of this trained model is visualised in Table 8, where positives are associated with the observed route segments and negatives with the shortest path segments. As visualised, the model predicts relatively much true positives. This resulted in an accuracy of 0.82, a precision of 0.82, a sensitivity of 0.99 and a F1-score of 0.90. The model has a relatively high sensitivity, but also tends to predict a route more attractive than it should be. This may be caused by the disbalance in the input data, since 20% of the input data are shortest path segments and 80% of the input data are observed route segments.

TABLE 8: CONFUSION MATRIX OF THE MODEL

	Predicted positive	Predicted negative
Actual positive	242,866	1,127
Actual negative	53,987	1,379

4.2.3. Model Predictions projected on Haarlem

Subsequently, the regression model created with the data of Enschede is projected on Haarlem. The confusion matrix of the model on the data of Haarlem is visualised in Table 9. The predictive performance of the model is lower compared to Enschede, since the accuracy is 0.76, the precision is 0.77, the sensitivity is 1.00 and the F1-score is 0.86. The model projected on Haarlem shows the same behaviour when projected on Enschede, since this model also tends to overestimate the attractivity of segments. However, as Table 9 indicates, this problem is more severe when the model is projected on data of Haarlem

TABLE 9: CONFUSION MATRIX OF THE ENSCHEDE MODEL PROJECTED ON DATA OF HAARLEM

	Predicted positive	Predicted negative
Actual positive	150,418	0
Actual negative	48,018	0

Moreover, a separate regression model for Haarlem is created to indicate the difference in β_i coefficients and also in standardised regression coefficients. The results of that model are presented in Table 10.

TABLE 10: STATISTICAL SUMMARY OF REGRESSION MODEL OF HAARLEM

Factor	β	Std. Error	P value	Standardised β
Constant	-0.866	0.019	0.000	0
Infrastructural				
Distance to traffic control installation	-0.096	0.003	0.000	-0.356
Cycle lane	0.527	0.052	0.000	0.094
Separate cycle path	1.148	0.013	0.000	1.341
Artificial lighting	-0.095	0.009	0.000	-0.099
Paved infrastructure (asphalt)	0.539	0.009	0.000	0.629
Paved infrastructure (paving stones)	0.716	0.012	0.000	0.583
Motorised vehicle intensities	-0.158	0.003	0.000	-0.722
Bicycle intensities	0.358	0.003	0.000	1.218
Land use allocation				
Residential land use zone	0.674	0.025	0.000	0.355
Commercial land use zone	-0.201	0.033	0.000	-0.089
Greenery land use zone	0.446	0.026	0.000	0.194
Industrial land use zone	-0.401	0.030	0.000	-0.199
Degree of urbanisation	0.360	0.006	0.000	0.633

As presented in Table 10, all P values were statistically significant after one iteration of stepwise regression was implemented. During this iteration, the factor *'Land use mix'* was eliminated, since it was statistically significant. Moreover, the standardised β_i coefficients of the model of Haarlem are different compared to the model of

Enschede. The main differences are observed in the factors *'Paved infrastructure (paving stones)', 'Greenery land use zone'*, *'Land use mix'* and *'Degree of urbanisation'*. In Enschede the presence of paving stones was not included since it was not statistically significant. However, the model of Haarlem shows that the presence of paving stones has a positive influence on the attractivity. This may be caused by the fact that Haarlem has a relatively old city centre with a substantial number of segments with paving stones as surface material (12.2% of all segments in the municipality of Haarlem). Since this network of infrastructure with paving stones is so wide spread in the city centre, cyclists have to use it in order to go from one side to the other side of the city centre. For land use mix the same holds, however opposite. The factor *'Land use mix'* was statistical significant in the model of Enschede, however not in the model of Haarlem. In addition, in Enschede greenery had a negative influence on the attractivity of a certain segment. Nonetheless, in Haarlem it had a substantial positive affect. This may be caused by the fact that Haarlem has substantial less area covered in greenery that can act as shortcuts. This may result in the model of Haarlem being less prone to the problem that the shortest path calculation did not use any restrictions as discussed in Section 4.2.1. The other standardised regression coefficients have the same magnitude for both Enschede and Haarlem, except from the factor '*Residential land use zone'* and *'Degree of urbanisation'*. The first factor had more effect on the attractivity in Enschede compared to Haarlem, however the factor *'Degree of urbanisation'* had more effect on the attractivity in Haarlem. The latter might be caused by Haarlem having the third highest density of addresses in the Netherlands, resulting in a higher attractivity of less urban area (Architectenweb, 2017).

Moreover, the performance of this model on the test data of Haarlem is evaluated. The confusion matrix of the model specific for the city of Haarlem is visualised in Table 11. Compared to the model of Enschede projected on Haarlem, this model specific for Haarlem performed better since the accuracy is 0.80, the precision is 0.82, the sensitivity is 0.94 and the F1-score is 0.88. This model is better balanced resulting in less overestimation of the attractivity of segments as the model of Enschede does.

TABLE 11: CONFUSION MATRIX OF THE MODEL SPECIFIC FOR HAARLEM

	Predicted positive	Predicted negative
Actual positive	140,672	9,746
Actual negative	30,352	17,666

4.3. Application of the Model

With the establishment of Equation 6, an attractiveness map is created for the city of Enschede. This map indicates the relative attractivity of a segment based on the characteristics included in this study. This attractiveness map is presented in Figure 14. In the remainder of this section, some interesting areas are highlighted for further validation purposes. Moreover, the city of Haarlem, which was used for validation purposes, is evaluated in this section.

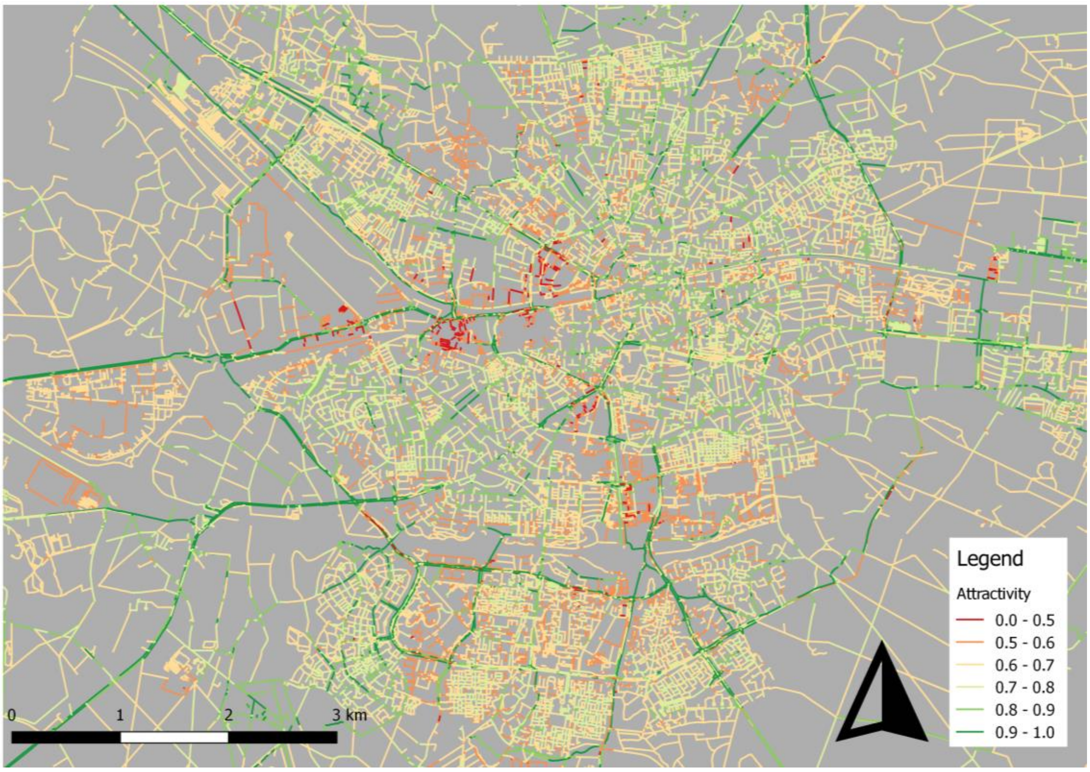

FIGURE 14: ATTRACTIVITY MAP OF ENSCHEDE

4.3.1. ENSCHEDE, BODDENKAMP

Figure 15 presents a close up of the Northwest of Enschede around the 'Boddenkamp' district. This is the same area that has been used for maps presented in Section 3.1.2. The area is interesting, since a lot of commercial and industrial buildings are located there, which should indicate relatively less attractive segments according to Equation 6.

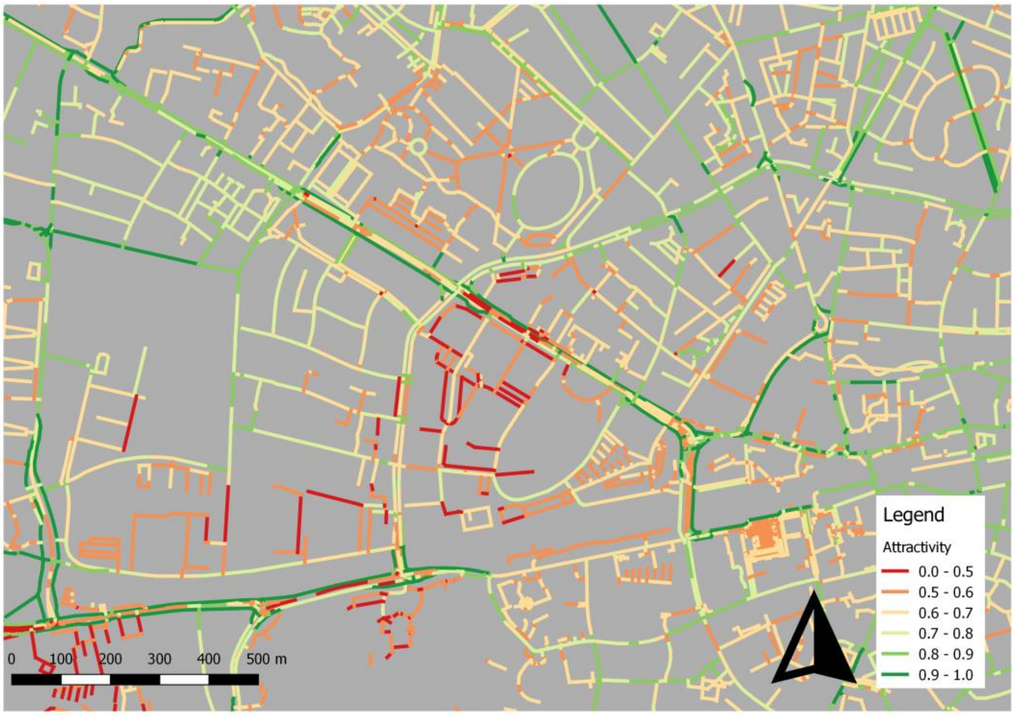

FIGURE 15: ATTRACTIVITY MAP OF ENSCHEDE'S DISTRICT BODDENKAMP

The areas where commercial and industrial buildings are close to the bicycle network resulted in relatively less attractive segments. Furthermore, since industrial areas are often less used by cyclists, no substantial bicycle infrastructure is present. However, these factors have positive β_i coefficients and by not implementing bicycle infrastructure, the attractiveness remains low.

Areas with a similar land use in Enschede like 'Marssteden', the area around the 'Zuiderval' and in Haarlem like 'Waarder- and Veerpolder' show a similar pattern of relatively low attractive segments. Maps visualising this can be found in Appendix A1.

4.3.2. Enschede, Roombeek, F35

Figure 16 A is a close up of the North of Enschede in the 'Roombeek' district. This area is interesting, since the F35, a bicycle highway, runs through this district and since the area is a relatively new-built residential area. Both factors should have a positive effect on the attractiveness according to Equation 6.

FIGURE 16: A) ATTRACTIVITY MAP OF ENSCHEDE, ROOMBEEK, F35 AND B) ATTRACTIVITY MAP OF ENSCHEDE CROMHOFFSBLEEK-KOTMAN

In Figure 16A, the F35 runs through the area as a straight line from the Southwest to the Northeast and has mostly a dark green colour indicating a relatively attractive set of segments. However, some segments are only part of a 'fietsstraat' allowing motorised vehicles. From that point onwards, the F35 is lighter green indicating a relatively less attractive set of segments compared to the other parts of the F35. At other stretches of the F35 like on the Oosterstraat the same difference between segments of designated bicycle paths and shared 'fietsstraten' is observed as presented in Appendix A2. Moreover, in Haarlem along the Amsterdamsevaart the same phenomenon is observed as also presented in Appendix A2. However, in that case the separate cycle path is not part of a bicycle highway.

Moreover, the Roombeek district has been rebuilt quite recently with good land use mix, substantial degree of urbanisation and a well-designed, paved bicycle infrastructure resulting in a relative attractive bicycle network in this particular district. Especially in contrast with relatively older districts like Lasonder-Zeggelt, Boddenkamp and Tubantia-Toekomst as visualised in Figure 15 and Cromhoffsbleek-Kotman as visualised in Figure 16B.

4.3.3. ENSCHEDE, WEST-ENSCHEDE, F35

Figure 17 is a close up of the West of Enschede around the railway track between Hengelo and Enschede. This area is interesting, since it is part of the planned trajectory of the F35 from the West to the station of Enschede (Fietssnelweg F35 (sd) and Valk, et al. (2014)).

FIGURE 17: ATTRACTIVITY MAP OF ENSCHEDE, WEST-ENSCHEDE ALONG THE RAILWAY TRACK BETWEEN HENGELO AND ENSCHEDE

In the West of Figure 17 the F35 can be identified as the attractive set of segments located in the left purple dotted oval. However, at the Lambertus Buddestraat the F35 ends and the attractiveness drops substantially. Moreover, in the Northeast and East of Figure 17 the F35 continues from the station in the direction of both Oldenzaal and Glanerbrug, as located in the right purple dotted oval, and both are also indicated as an attractive set of segments. However, in between the aforementioned trajectories, no direct, attractive set of segments is in place, suggesting that the planned trajectory could be of added value to the existing bicycle network.

4.3.4. HAARLEM

Figure 18 provides an overview of the attractivity of segments within the municipality of Haarlem using the model of Enschede. This area is interesting, since Haarlem is used for validation purposes in Section 4.2.3.

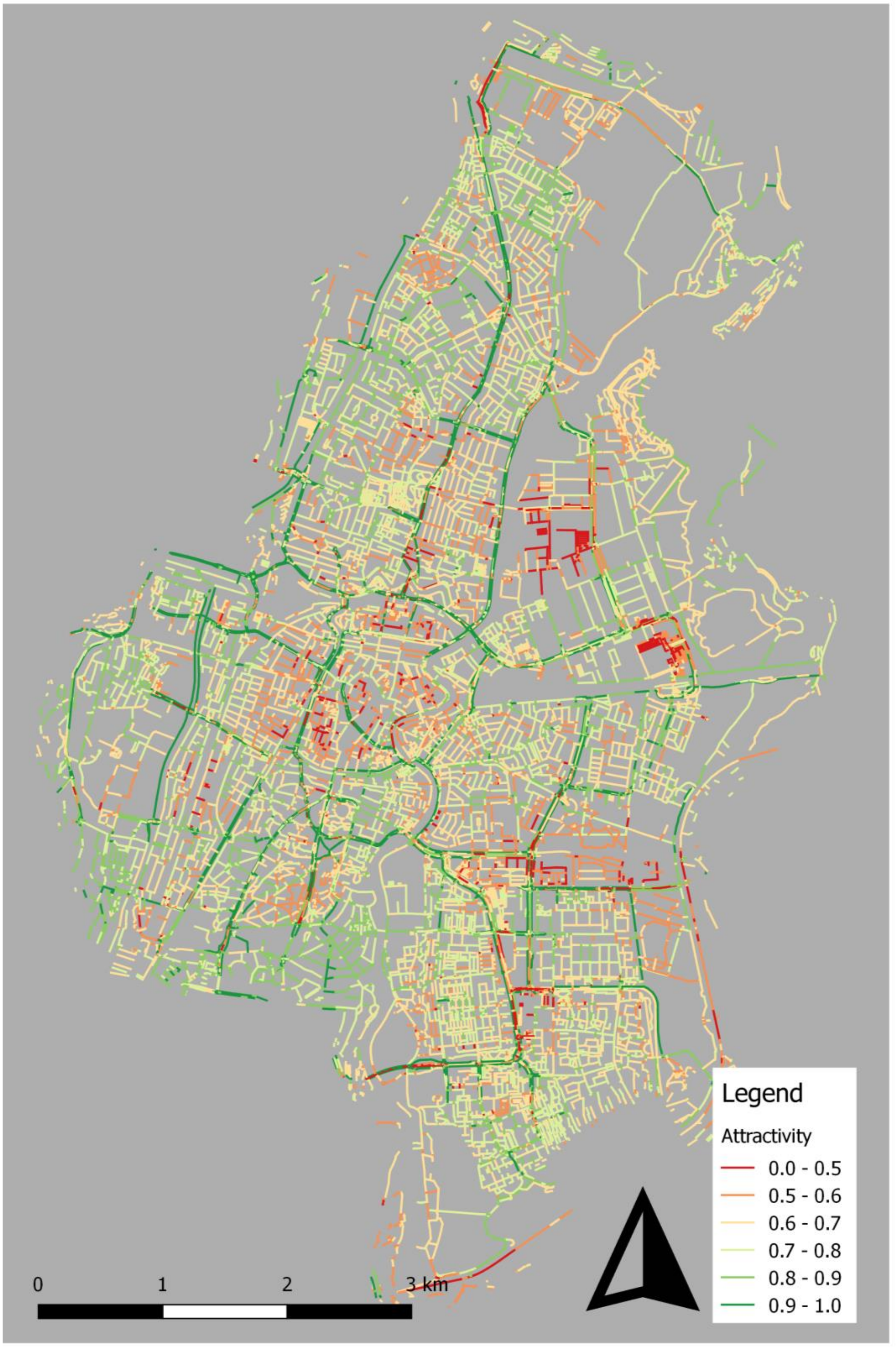

FIGURE 18: ATTRACTIVITY MAP OF HAARLEM

4.3.5. Haarlem, Comparison of Models

Figure 19 provides an overview of the attractivity of segments within the municipality of Haarlem using the model of Enschede (A) and using the model of Haarlem (B). This comparison is interesting, since Section 4.2.3 indicated already a substantial difference in β_i coefficients between both models.

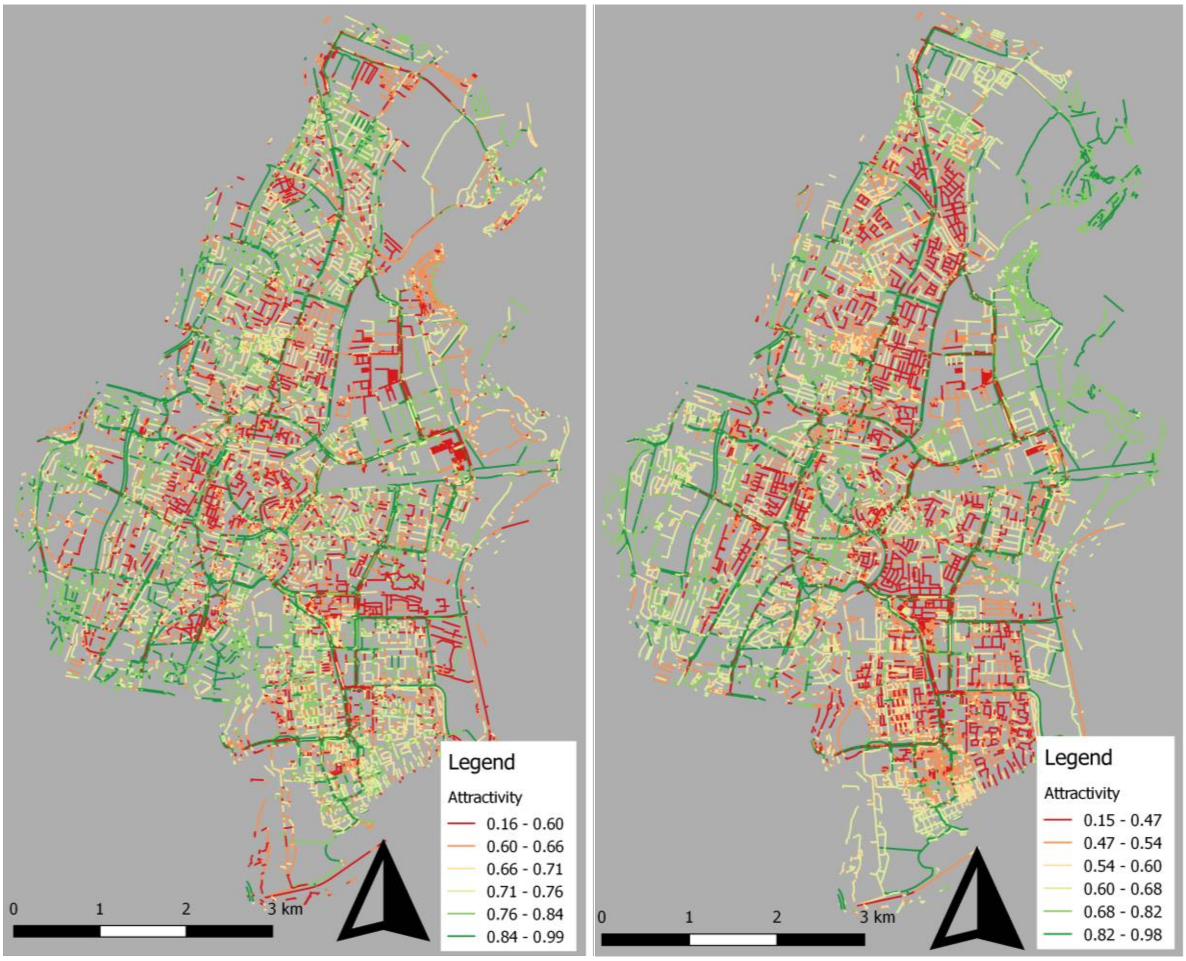

FIGURE 19: ATTRACTIVITY MAP OF HAARLEM USING A) THE MODEL OF ENSCHEDE PROJECTED ON HAARLEM, B) THE MODEL OF HAARLEM

In Figure 19 both maps use a different scale, since the accompanied model of Enschede predicts segments substantially more attractive compared to the model of Haarlem as presented in Appendix B. In order to make a comparison, both maps use quantile classification, which means that every class contains the same number of segments. This made it possible to make a comparison between the maps and clearly visualise, which area differ from each other.

Firstly, both maps clearly illustrate the high attractivity of the main separate cycle path network indicating consistency between both models.

However, Figure 19 B indicates a relatively low attractivity for the residential areas in the North, East and Southeast of Figure 19, compared to the same areas in Figure 19 A. This is consistent with the standardised β_i coefficients for both models, since residential area was substantially more influential in the model of Enschede. Moreover, Figure 19 B indicates a relatively high attractivity in the outskirts of the municipality in the North, East and West of Figure 19, compared to the same areas in Figure 19 A. At these areas more greenery is located, as visualised in Appendix C. This is consistent with the standardised β_i coefficients for both models, since green areas have a positive influence on the attractivity according to the model of Haarlem, however a negative influence according to the model of Enschede.

5. Discussion

The aim of this study was to assess the influence of infrastructural and land use allocation factors on the route choice of cyclists in Enschede by the use of 'fietstelweek' (Breda University of Applied Sciences, sd) data. To do this, the methodology of Section 3 was used. However, this methodology and thus the corresponding results hold assumptions and shortcomings that are considered and discussed in the remainder of this section.

5.1. The Model

First of all, the shortest path has been calculated using the origin destination pairs of the 'fietstelweek' (Breda University of Applied Sciences, sd) data and via Dijkstra graph theory algorithm using the length of the segments as weights. However, there were no restrictions added on certain segments, so in theory cyclists could use all segments included in the 'OpenStreetMap' (Openstreetmap, sd) network, including highways, primary roads and parks. Highways are only faster when traveling a relatively large distance (in a motorised vehicle) especially in Enschede and it is observed that no highway segments are present in the shortest path. However, segments through parks, as well as other segments noted as paths, occur in a substantial number in the shortest path segments. However, such segments, although in a fewer number also occur in the observed route segments and therefore the effect on the results are assumed to be relatively low.

Secondly, the data used in this study was not generated at the same moment in time. This is mostly caused by the 'fietstelweek' (Breda University of Applied Sciences, sd) data being from 2016 and is therefore relatively old compared to the other data sources. This means that renovations on the current bicycle network or extension or subtraction of this network is included in the more recent data sets, but is not present in the network of the 'fietstelweek' data. This could result in different shortest paths and/or different characteristics of certain segments. The effect of this is not expected to be substantial, due to the fact that between 2016 and 2022 not many large scale projects have been developed apart from the elongation of the F35 between the station of Enschede and Roombeek, the station of Enschede towards Glanerbrug in the East and the creation of more bicycle streets (Gemeente Enschede (2017), Valk, et al. (2014) and Fietssnelweg F35 (sd)). However, the latter has minor influence on the results, as bicycle streets do not affect the factors included in this study.

Besides that, the implementation of many different datasets comes with limitations per dataset. These limitations, such as 'OpenStreetMap' (Openstreetmap, sd) being a project where anyone can contribute to the network, are described in Section 2.3.2 and are taken as given in this study. Furthermore, the way these datasets are used had limitations as described in Section 3.1.2 per included factor. For example, the way motorised vehicle intensities are considered in this study is a coarse simplification, since it is mostly based on the proxy of the road as stated in 'OpenStreetMap'. The way datasets are considered in this study are substantiated, however the resulting limitations are taken as given in this study. Section 7.2 introduces recommendations to deal with some of the aforementioned limitations.

Subsequently, the data limitations resulted in only including segments of the observed routes and shortest paths within the municipality borders. This was due to the fact that greenery and bicycle intensity data provided respectively no and less accurate data outside the municipality borders. However, trips between cities may affect the model substantially, since the segments between cities are often through a less urban area with more greenery. In case of Enschede, such trips occur often between, for example Enschede and Hengelo via the F35.

Moreover, the use of the segment approach also comes with shortcomings. The first, on including route characteristics, was already introduced in Section 3.1.3. However, after the analysis another shortcoming arose. This approach compares segments that are not chosen, but are recommended when using the shortest path with the segments of observed routes. However, since the segments of the shortest path that are used in the observed route are filtered out, the classes become imbalanced. In the final dataset of Enschede, 80% of the segments come from observed routes and 20% come from shortest path and therefore show an imbalance (Li S. , 2017). The greater the imbalance, the higher the bias of the model towards the majority class, which is in this case the set of segments of the observed route (Brus, 2021). Relatively many segments were predicted to be a segment of the observed route, but in reality were not (18.0%).

5.2. Validation of the Model

One of the validation steps was to evaluate the key performance indicators, such as the accuracy, of the model. The key performance indicators provide a measure of how well the model can predict whether a segment is observed or not. Although the model performed relatively well on these key performance indicators, these indicators do not tell something about the performance of the model to predict the attractiveness of a segment. This is due to the fact that the attractiveness of a segment ranges between 0 and 1 and is not a binary number.

5.3. Application of the Model

The attractiveness map provides insights in the attractiveness of segments and was validated by evaluating interesting areas. However, the model often produces fluctuating attractivity between neighbouring segments. This behaviour is observed especially at intersections, as visualised in Figure 20.

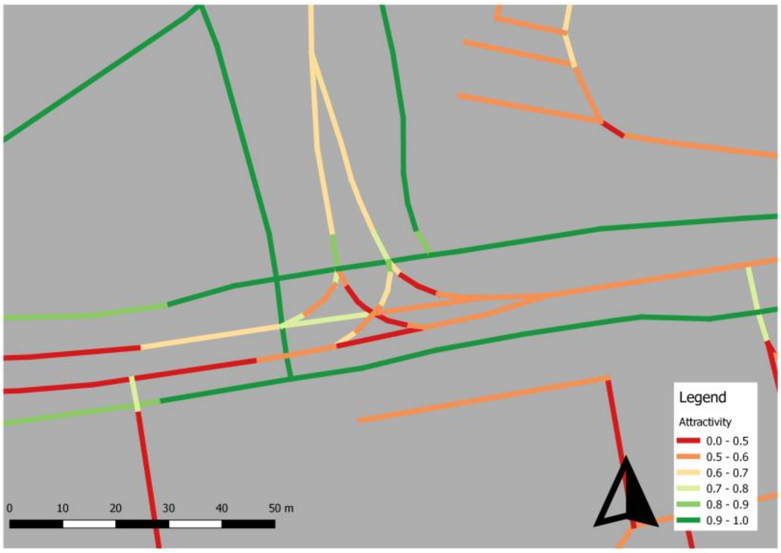

FIGURE 20: ATTRACTIVITY MAP OF ENSCHEDE AT THE INTERSECTION OF THE LAMBERTUS BUDDESTRAAT AND THE PARKWEG

In fact, the attractivity at this intersection ranges between 0.43 and 0.76, despite being in the same 20m by 20m area. Although this infrastructure is not cyclable as there are separate cycle paths, it indicates the sensitivity of the model. This can be the result of the buffer shape around a segment chosen in this study. As mentioned in Section 3.1.2.7, the buffer was flat indicating that the shape of the buffer was a rectangle with only a buffer area on both sides of the segment and not ahead or behind the segment. This theory is strengthened by Figure 20, since an industrial and commercial area is located in the South-West of Figure 20 and a commercial building is located at the Northeast of Figure 20. The least attractive segments are in line with one or both of these areas. The attractive segments are in this case not in line with the industrial and commercial areas, but would have been in case a round buffer was used. Using a flat buffer was deemed favourable in order to neglect the effect of the surroundings of other segments, however as indicated in Figure 20 shows less favourable results when segments are relatively small.

Moreover, the aforementioned limitation in the calculation of the shortest path using no limitations on which part of the network is prohibited to cycle on, also holds for the attractiveness map. At first no limitations were added to the network resulting in, for example highways being noted as moderate or good to cycle on. Therefore, all segments which are prohibited to cycle on are excluded from the attractiveness map However, before the results of the model itself can be used to identify the attractivity of certain routes or other applications, it is important that to add certain restrictions.

6. Conclusion

Cyclists may not always use the shortest path to travel from their origin to their destination as it is expected that they are affected by environmental factors, like infrastructural and land use allocation factors. Research supports this claim, however several different approaches and use cases are used in literature indicating that there is no clear way how environmental factors influence the route choice of cyclists. Therefore, the aim of this study was to assess the influence of infrastructural and land use allocation factors on the route choice of cyclists in Enschede by the use of 'fietstelweek' (Breda University of Applied Sciences, sd) data. In this conclusion, first the four sub-questions are answered. Subsequently, a conclusion is drawn for the main question.

Q1 "What infrastructural and land use allocation factors have a substantial influence on the route choice of cyclists according to literature?"

In this study, fifteen papers on the factors that influence the route choice of cyclists have been evaluated. This reviewed literature indicated fourteen different infrastructural or land use allocation factors that could influence the route choice of cyclists, namely the presence of traffic control installations, bicycle lanes, separated bicycle paths, artificial lighting and paved infrastructure, sloped infrastructure, the intensities of motorised vehicles and bicycles, the land use in the area (either residential, commercial, industrial or greenery), land use mix and the floor area ratio. Moreover, most researchers touched upon the importance of trip length on the route choice of cyclists.

Q2 "What is the quantitative influence of infrastructural and land use allocation factors on the route choice of cyclists through a segment approach and regression modelling?"

The quantitative influence of infrastructural and land use allocation factors on the route choice of cyclists is presented in Equation 6, where all the β_i coefficients represent the quantitative influence of the separate factors. All β_i coefficients were statistically significant when the whole dataset was used due to the large dataset. Therefore, a Monte Carlo method is used to account for the risk in quantitative analysis. This resulted in relatively small confidence intervals close to the β_i coefficients of the model of the full data set, so the β_i coefficients are considered statistically significant.

Q3 "How valid are the results of the quantitative model when comparing it with existing literature and the use of test data?"

The model is evaluated for its consistency with literature, its performance and how it performs on another city, namely Haarlem. The model is consistent with literature, except from the factors '*Artificial lighting*', '*Bicycle intensities*' and '*Green land use zone*'. Moreover, this model has an accuracy of 0.81, a precision of 0.81, a sensitivity of 0.99 and a F1-score of 0.89. Following these key performance indicators, it can be concluded that the model is relatively good at predicting whether a segment is observed or not. Finally, the model performance was evaluated on Haarlem. The model presents an accuracy of 0.76, a precision of 0.76, a sensitivity of 0.99 and a F1-score of 0.86. The key performance indicators were lower for Haarlem than for Enschede and a model made for Haarlem indicated different β_i coefficients for some of the included factors. This both indicate that the model is not generally applicable. It can thus be concluded that the model produces good results for the study area it is designed for but substantially less results for another study area.

Q4 "How can the attractiveness of segments in a bicycle network be quantified through the developed model?"

The model is of great use for generating an attractiveness map and by evaluating interesting areas, it can be concluded that the attractiveness map is valid adding to already existing validity proven at sub-question 3. The attractiveness map can be used for policy makers to identify missing potential attractive links.

MQ ***"What are the influences of infrastructural and land use allocation factors on the route choice of cyclists through a segment approach?"***

Following the answers on the sub-questions, it can be concluded that infrastructural and land use allocation factors influence the route choice of cyclists. According to literature, in total fourteen factors were important. When implementing almost all of these factors in a model using a segment approach, it was concluded that the route choice is influenced by attractivity of segments. The attractivity of an individual segment is positively influenced by the presence of cycle lanes and separate cycle paths, high bicycle intensities, a high ratio of residential area, an equal land use mix and a low degree of urbanisation. In contrast, the attractivity of an individual segment is negatively influenced by the presence of traffic control installation and artificial lighting, high motorised vehicle intensities and a low ratio of commercial area, green area and industrial area. The model was proven to be valid to a large extent using the study area of Enschede. For other study areas, it is recommended to execute further research before implementing it.

7. Recommendations

After the conclusion is drawn, practical recommendations on how the results of this study can be applied in the field of route choice of cyclists and policy making are provided. Moreover, recommendations for future research are presented in this section.

7.1. Practical Recommendations

Firstly, when analysing the route choice of cyclists, the attractiveness of a set of routes for a certain origin destination pair can be determined using the logit function of Equation 6. The β_i coefficients in the function have proven to be valid via three ways. Nonetheless, this study concluded that the model is only valid for a specific study area, so if a new study area is chosen, it is important that a separate model is created. Moreover, the attractiveness is associated with the probability of a cyclist using a certain route. Following this reasoning, the outcomes of the model can also be used to distribute cyclists over a set of routes. For this set of routes, it is important that they have a comparable trip length in order for the model to function well. Currently, the routes in the 'FietsMonitor' are assigned 'all or nothing' to the shortest path. However, the outcomes of this study can help them when their route choice algorithm is improved and is generating multiple routes by evaluating those routes on the probability a cyclist would use it.

Furthermore, the attractiveness map can indicate at which locations within the selected study area the segments are relatively less attractive compared to others and provide policy makers with a substantiation on where bicycle infrastructure should be improved. When policy makers want to improve the attractiveness of the bicycle network, they are advised to investigate the implementation of paved, separated cycle paths through residential areas, which are relatively less urban and result in high bicycle intensities.

7.2. Recommendations for Future Research

First of all, research indicates that trip length is a significant factor influencing the route choice of cyclists. Although including it was out of the scope of this study, it is still interesting to see the influence of distance on the regression model. However, as mentioned in Section 3.1.3, using a segment approach, trip length cannot be implemented directly. A solution could be to use a multinomial regression model using multiple classes for the ratio between the trip length of the observed route and the trip length of the shortest path. This multinomial regression model provides different β_i coefficients for every class with respect to the reference scenario, the shortest path. Then, by comparing the β_i coefficients, the influence of trip length on the other factors can be analysed.

In addition, using the approach as discussed in Section 3.1, all influential factors were immediately included in the model. However, the factors have not been analysed on their individual influence on the attractivity of segments. Although the multicollinearity was proven to be not substantial, certain factors might influence each other when included in the model all together. Therefore, it is recommended to investigate the influence of individual factors on the attractivity of segments.

Furthermore, this study included a substantial number of infrastructural and land use allocation factors that influence the route choice of cyclists. However, there are more factors that could also have an influence on the route choice of cyclists but were not included in this study. Examples are the number of righthand or lefthand turns a cyclists encounters along its trip, number of intersection on a route, car parking facilities and the presence of public transport facilities. Moreover, the slope was excluded in this study, since no studies found substantial evidence on the influence of the slope on the route choice of cyclists in the Netherlands. Even for the Dutch perspective it is interesting to investigate if the influence of the slope is significant.

Besides that, in this study attractivity is defined as the probability a cyclist would use a certain segment based on the factors included in this study. However, in Thüsh, Talens, & Steeneken (2019), which focusses on one of the pillars of CROW (2016), the definition of attractivity includes more. Especially, the relationship between humans and their behaviour plays a prominent role is this document, such as incentives for senses and variety and surprises in the environment. In order to match the definition of attractivity as stated in this document, other, more human-related, factors are recommended to include in future research.

Moreover, in this study only segments within the municipality border are considered in this analysis, as explained in Section 1.4.4. It is recommended to extend the study area when a target municipality is chosen, such that neighbouring municipalities are also included, and study the effects on the β_i coefficients in the regression model. This could be valuable, as the influence of environmental factors on route choice of cyclists could be higher in case the distance of a trip is longer.

Finally, the model can be tested on multiple cities around the country. In this study two opposite cities are evaluated. However, the results indicated that the model for Enschede is not generally applicable on other cities in the Netherlands. Moreover, in this study the regression model is trained with data of one city at the time. Nonetheless, the model could also be trained for multiple cities at once, increasing the generality of the model for other cities and validate the results.

Bibliography

Adedokun, B. O. (2008). P-Value and confidence intervals - facts and farces. *Annals of Ibadan Postgraduate Medicine*, 33-34.

Ang, B. H., Chen, W. S., & Lee, S. W. (2017). Global burden of road traffic accidents in older adults: A systematic review and meta-regression analysis. *Archives of Gerontology and Geriatrics*, 32-38.

Architectenweb. (2017, March 22). *Den Haag dichtstbevolkte stad van Nederland*. Retrieved from Architectenweb: https://architectenweb.nl/nieuws/artikel.aspx?ID=38604#:~:text=Den%20Haag%20is%20de%20dichtstbevolkte,vijfduizend%20mensen%20per%20vierkante%20kilometer.

Band, M. (2022). *A discrete choice model for route preferences of commuters: comparison between conventional and fast bicycles.* Delft.

Behesthi, N. (2022, February 10). *Guide to Confusion Matrices & Classification Performance Metrics*. Retrieved from TowardsDataScience: https://towardsdatascience.com/guide-to-confusion-matrices-classification-performance-metrics-a0ebfc08408e

Bernardi, S., La Paix Puello, L., & Geurs, K. (2018). Modelling route choice of Dutch cyclists using smartphone data. *The Hounral of Transport and Land Use*, 883-900.

Breda University of Applied Sciences. (n.d.). *Download bestanden Nationale Fietstelweek 2015, 2016 en 2017*. Retrieved from Open data cycling intelligence: http://opendata.cyclingintelligence.eu/

Broach, J., & Dill, J. (2016). *Using Predicted Bicyclist and Pedestrian Route Choice to Enhance Mode Choice Models.*

Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, 1730-1740.

Brus, P. (2021, June 1). *Data Imbalance in Regression*. Retrieved from TowardsDataScience: https://towardsdatascience.com/data-imbalance-in-regression-e5c98e20a807

Campos-Sánchez, F. S., Valenzuela-Montes, L. M., & Abarca-Álvarez, F. J. (2019). Evidence of Green Areas, Cycle Infrastructure and Attractive Destinations Working Together in Development on Urban Cycling. *Sustainability*.

Caulfield, B., Brick, E., & McCarthy, O. T. (2012). Determining bicycle infrastructure preferences – A case study of Dublin. *Transportation Research Part D*, 413-417.

Centraal Bureau voor de Statistiek | Topografische Dienst Kadaster. (2005).

Centraal Bureau voor de Statistiek. (2017). *Onderzoek Verplaatsingen in Nederland (OViN) 2016, Plausibiliteitsrapportage.* The Hague: Centraal Bureau voor de Statistiek.

Centraal Bureau voor de Statistiek. (2021, Augustus 6). *Kerncijfers wijken en buurten 2021*. Retrieved from Centraal Bureau voor de Statistiek: https://www.cbs.nl/nl-nl/maatwerk/2021/31/kerncijfers-wijken-en-buurten-2021

Centraal Bureau voor de Statistiek. (n.d.). *Hoeveel fietsen inwoners van Nederland*. Retrieved from Centraal Bureau voor de Statistiek: https://www.cbs.nl/nl-nl/visualisaties/verkeer-en-vervoer/personen/fietsen

Chen, P. (2016). Built environment effects on bicyclists' route preferences: a GPS data analysis. In P. Chen, *Bicycling and the built environment: route choice and road safety* (pp. 19-56). Washington: University of Washington.

CROW. (2016). *Ontwerpwijzer fietsverkeer.*

de Baan, D. (n.d.). *Wegontwerp - basis voor veilige verkeersdeelname*. Retrieved from Verkeer | Verkeersveiligheid | Vorm: https://www.dirkdebaan.nl/vorm.html

de Dios Ortúzar, J., & Willumsen, L. G. (2011). Discrete Choice Models. In J. de Dios Ortúzar, & L. G. Willumsen, *Modelling Transport* (pp. 227-268). Chichester: John Wiley & Sons, Ltd.

Domencich, T., & McFadden, D. (1975). *Urban Travel Demand: A Behavioural Analysis.* Amsterdam.

Ensing, N., & Janssen, K. (2020). *Actieplan Fiets 2020-2022.* Haarlem.

Esri. (n.d.). *AHN4 - Download kaartbladen*. Retrieved from ArcGIS: https://www.arcgis.com/home/webmap/viewer.html?layers=77da2e9eeea8427aab2ac83b79097b1a

Fietssnelweg F35. (n.d.). *Deeltrajecten*. Retrieved from Fietssnelweg F35: https://www.fietssnelwegf35.nl/deeltrajecten/

Fietssnelweg F35. (n.d.). *Met de F35 ben je er zo*. Retrieved from Fietssnelweg F35: https://www.fietssnelwegf35.nl/

Fietstelweek. (2016, November 8). *RESULTATEN FIETS TELWEEK 2016*. Retrieved from Fietstelweek: https://fietstelweek.nl/resultaten-fiets-telweek-bekend/

Fitch, D. T., & Handy, S. L. (2020). Road environments and bicyclist route choice: The cases of Davis and San Francisco, CA. *Journal of Transport Geography*.

Gemeente Enschede. (2012). *Enschede Fietsstad 2020.*

Gemeente Enschede. (2017). *Flyer Fietsstraten.* Enschede.

Gemeente Haarlem. (2017). *Structuurvisie openbare ruimte, Haarlem 2040: Groen en Bereikbaar.* Haarlem.

Gemeente Maastricht. (2020). *Actieplan Fietsen in Maastricht 2020-2025.* Maastricht: Gemeente Maastricht.

Gkiotsalitis, K. (2021). Network Modelling & Forecasting, L3: Mode & Route Choice. Enschede: Center for Transport Studies - University of Twente.

Heinen, E., Maat, K., & Van Wee, B. (2011). The role of attitudes toward characteristics of bicycle commuting on the choice to cycle to work over various distances. *Transportation Research Part D: Transport and Environment*, 102-109.

Huber, S., L. S., Lindemann, P., Muthmann, K., Schnabel, A., & Friedl, J. (2021). Modelling bicycle route choice in German cities using open data, MNL and the bikeSim web-app. *2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 1-6.

IBM. (2021, December 7). *Multinomial logistic regression: Model*. Retrieved from IBM: https://www.ibm.com/docs/en/spss-statistics/beta?topic=regression-multinomial-logistic-model

Jakovljevic, B., Paunovic, K., & Belojevic, G. (2009). Road-traffic noise and factors influencing noise annoyance in an urban population. *Environment International*, 552-556.

Khandelwal, R. (2020, April 7). *Quick and Easy Explanation of Logistic Regression*. Retrieved from TowardsDataScience: https://towardsdatascience.com/quick-and-easy-explanation-of-logistics-regression-709df5cc3f1e

Khatri, R., Cherry, C. R., Nambisan, S. S., & Han, L. D. (2016). Modeling Route Choice of Utilitarian Bikeshare Users with GPS Data. *Transport Research Record*, 141-149.

Koch, T., & Dugundji, E. R. (2021). Taste Variation in Environmental Features of Bicycle Routes. *International Workshop on Computational Transportation Science*.

Kwok, R. (2021, March 9). *Stepwise Regression Tutorial in Python*. Retrieved from TowardsDataScience: https://towardsdatascience.com/stepwise-regression-tutorial-in-python-ebf7c782c922

LaValley, M. P. (2008). Logistic Regression. *Circulation*, 2395-2399.

Li, S. (2017, September 29). *Building A Logistic Regression in Python, Step by Step*. Retrieved from TowardsDataScience: https://towardsdatascience.com/building-a-logistic-regression-in-python-step-by-step-becd4d56c9c8

Li, Z., Wang, W., Liu, P., & Ragland, D. R. (2012). Physical environments influencing bicyclists' perception of comfort on separated and on-street bicycle facilities. *Transport Research*, 256-261.

Maat, K., van Wee, B., & Stead, D. (2005). Land Use and Travel Behaviour: Expected Effects from the Perspective of Utility Theory and Activity-Based Theories. *Environment and Planning B: Planning and Design*, 33-46.

MacLeod, M. (2022, March 10). Modeling: A Critical Approach. Enschede: University of Twente.

Make BV, Civieltechnisch Adviesbureau. (n.d.). *Den Haag – Mecklenburgplein*. Retrieved from Make BV: https://make-bv.nl/den-haag-mecklenburgplein/

Mertens, L., Van Dyck, D., Ghekiere, A., De Bourdeaudhuij, I., Deforche, B., Van de Weghe, N., & Van Cauwenberg, J. (2016). Which environmental factors most strongly influence a street's appeal for bicycle transport among adults? A conjoint study using manipulated photographs. *International Journal of Health Geographics*.

Metropolitan Council. (2015). *CALCULATING FLOOR AREA RATIO.* Saint Paul.

Municipality of Alkmaar. (2020). *Alkmaar Fietst, Actieplan Fiets 2021-2025.* Alkmaar.

Openstreetmap. (n.d.). *Openstreetmap*. Retrieved from Openstreetmap: https://www.openstreetmap.org/#map=19/52.22168/6.86932

Palisade. (n.d.). *Monte Carlo Simulation*. Retrieved from Palisade: https://www.palisade.com/risk/monte_carlo_simulation.asp#:~:text=Monte%20Carlo%20simulation%20performs%20risk,values%20from%20the%20probability%20functions.

PDOK. (n.d.). *Dataset: Basisregistratie Grootschalige Topografie (BGT)*. Retrieved from PDOK: https://www.pdok.nl/introductie/-/article/basisregistratie-grootschalige-topografie-bgt-

Planbureau voor de Leefomgeving. (n.d.). *PBL Dataportaal Downloads*. Retrieved from Planbureau voor de Leefomgeving: https://dataportaal.pbl.nl/downloads/RUDIFUN1/

Polamuri, S. (2017, May 15). *2 WAYS TO IMPLEMENT MULTINOMIAL LOGISTIC REGRESSION IN PYTHON*. Retrieved from DataAspirant: https://dataaspirant.com/implement-multinomial-logistic-regression-python/

Prato, C. G., Halldórsdóttir, K., & Nielsen, O. A. (2018). Evaluation of land-use and transport network effects on cyclists' route choices in the Copenhagen Region in value-of-distance space. *International Journal of Sustainable Transportation*, 770-781.

Rasch, A., Moll, S., López, G., García, A., & Dozza, M. (2022). Drivers' and cyclists' safety perceptions in overtaking maneuvers. *Transportation Research Part F: Psychology and Behaviour*, 165-176.

Ratnasari, D., Nazir, F., Toresano, L. O., Pawiro, S. A., & Soejoko, D. S. (2015). The correlation between effective renal plasma flow (ERPF) and glomerular filtration rate (GFR) with renal scintigraphy 99mTc-DTPA study. *Journal of Physics, Conference Series*.

Raychaudhuri, S. (2008). Introduction to Monte Carlo simulation. *2008 Winter Simulation Conference*, 91-100.

Saelens, B., Sallis, J. F., & Frank, L. D. (2003). Environmental correlates of Walking and Cycling: Findings From the Transportation, Urban Design, and Planning Literatures. *Annals of Behavioral Medicine*, 80-91.

Segadilha, A. B., & Sanches, S. d. (2014). Identification of Factors that Influence Cyclistś Route Choice. *Procedia - Social and Behavioral Sciences*, 372-380.

Siegel, A. F., & Wagner, M. R. (2022). Multiple Regression: Predicting One Variable From Several Others. In A. F. Siegel, & M. R. Wagner, *Practical Business Statistics (Eighth Edition)* (pp. 371-431). Academic Press.

statsmodels. (n.d.). *statsmodels.discrete.discrete_model.Logit.fit*. Retrieved from statsmodels 0.14.0: https://www.statsmodels.org/dev/generated/statsmodels.discrete.discrete_model.Logit.fit.html#statsmodels.discrete.discrete_model.Logit.fit

Stinson, M. A., & Bhat, C. R. (2003). An Analysis of Commuter Bicyclist Route Choice Using a Stated Preference Survey. *Transportation Research*, 107-115.

Strauss, J., & Miranda-Moreno, L. F. (2013). Spatial modeling of bicycle activity at signalized intersections. *The Journal of Transport and Land Use*, 47-58.

Thüsh, M., Talens, H., & Steeneken, P. (2019). *Uitwerking hoofdeisen fietsvriendelijke infrastructuur, Uitwerking hoofdeis 5: Aantrekkelijkheid.* Ede: CROW-Fietsberaad.

Ton, D., Duives, D., Cats, O., & Hoogendoorn, S. (2018). Evaluating a data-driven approach for choice set identification using GPS bicycle route choice data from Amsterdam. *Travel Behaviour and Society*, 105-117.

UCLA. (n.d.). *FAQ: WHAT ARE PSEUDO R-SQUAREDS?* Retrieved from UCLA, Advanced Research Computing, Statistical Methods and Data Analysis: https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/faq-what-are-pseudo-r-squareds/#:~:text=A%20pseudo%20R%2Dsquared%20only,model%20better%20predicts%20the%20outcome.

Uttley, J., Fotios, S., & Lovelace, R. (2020). Road lighting density and brightness linked with increased cycling rates after-dark. *PLoS ONE*.

Valk, R., Bak, G., Wilts, E., Hofman, G., Horsthuis, H., Schilstra, J., . . . ter Avest, R. (2014). *Masterplan Fietssnelweg F35 .* NetzoDruk.

Veenstra, S. (2022). (N. van Nijen, Interviewer)

Veenstra, S., Geurs, K., Thomas, T., & Van den Hof, R. (2016). Alle lichten op groen voor fietsmonitoring in Enschede. *Verkeerskunde*.

Washington, S. P., Karlaftis, M. G., & Mannering, F. L. (2011). Logistic Regression. In S. P. Washington, M. G. Karlaftis, & F. L. Mannering, *Statistical and Econometric Methods for Transportation Data Analysis* (pp. 303-308). Chapman and Hall/CRC.

Winters, M., Brauer, M., Setton, E. M., & Teschke, K. (2010). Built Environment Influences on Healthy Transportation Choices: Bicycling versus Driving. *Journal of Urban Health*, 969-993.

Winters, M., Davidson, G., Kao, D., & Teschke, K. (2011). Motivators and deterrents of bicycling: comparing influences on decisions to ride. *Transportation*, 153-168.

Winters, M., Teschke, K., Grant, M., Setton, E. M., & Brauer, M. (2010). How Far Out of the Way Will We Travel? *Transportation Research Record*, 1-10.

Witteveen+Bos. (n.d.). *FietsMonitor*. Retrieved from Witteveen+Bos: https://digitalsolutions.witteveenbos.com/ruimtelijke-ontwikkeling-wonen-en-mobiliteit/mobiliteit/fietsmonitor

Zhao, Y., Ke, S., Lin, Q., & Yu, Y. (2020). Impact of land use on bicycle usage: A big data-based spatial approach to inform transport planning. *Journal of Transport and Land Use*, 299-316.

APPENDICES

APPENDIX A: ATTRACTIVITY MAPS

APPENDIX A1: OTHER INDUSTRIAL AND COMMERCIAL AREAS

FIGURE 21: ATTRACTIVITY MAP OF ENSCHEDE, MARSSTEDEN

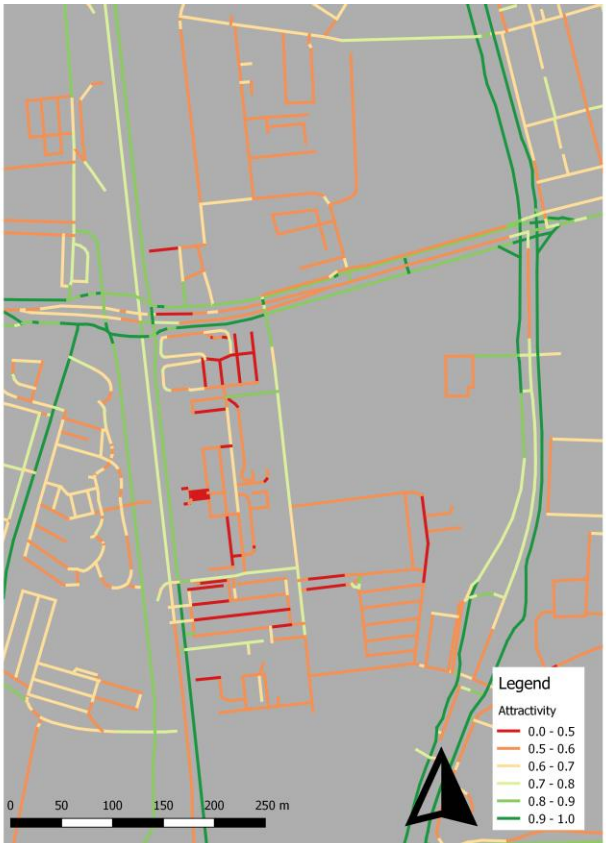

FIGURE 22: ATTRACTIVITY MAP OF ENSCHEDE, ZUIDERVAL

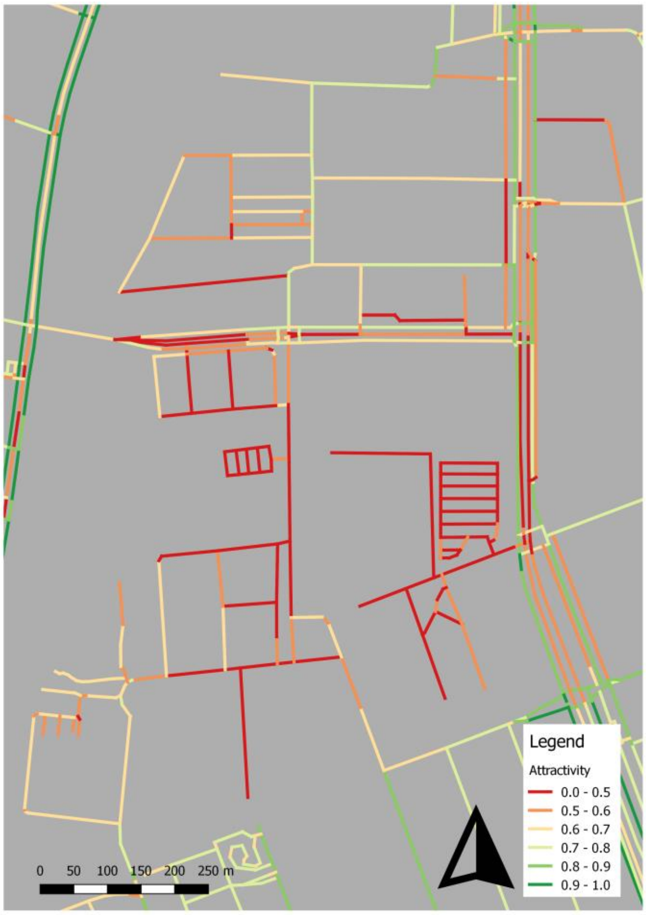

FIGURE 23: ATTRACTIVITY MAP OF HAARLEM, WAARDER- AND VEERPOLDER

APPENDIX A2: OTHER LOCATIONS WITH A TRANSITION FROM SEPARATE CYCLE PATH TO 'FIETSSTRAAT'

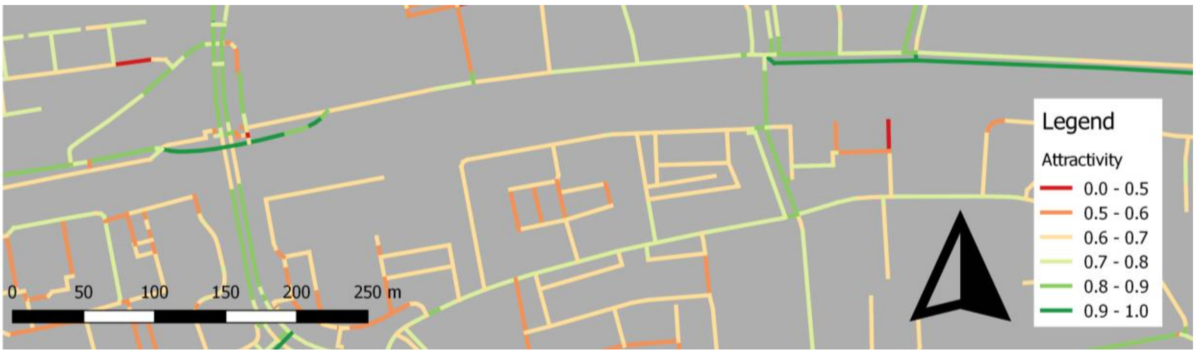

FIGURE 24: ATTRACTIVITY MAP OF ENSCHEDE, OOSTERSTRAAT

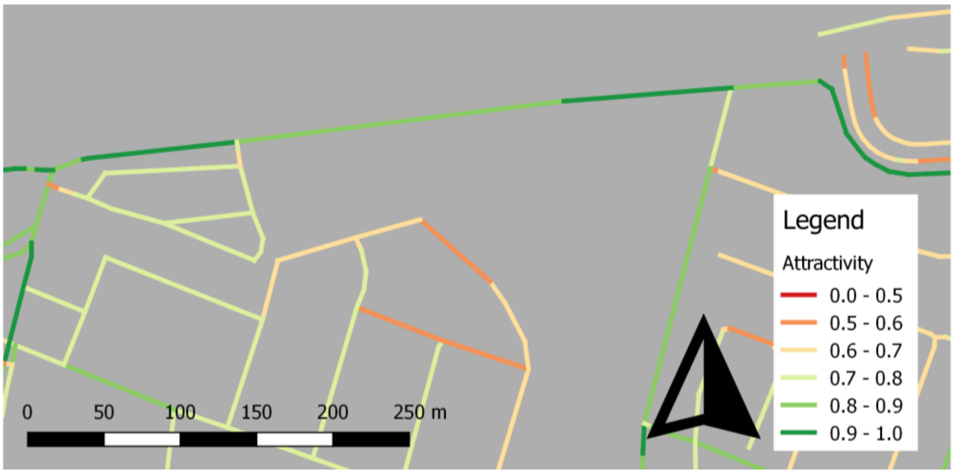

FIGURE 25: ATTRACTIVITY MAP OF HAARLEM, AMSTERDAMSEVAART

APPENDIX B: ATTRACTIVITY MAP USING SCALE OF ENSCHEDE

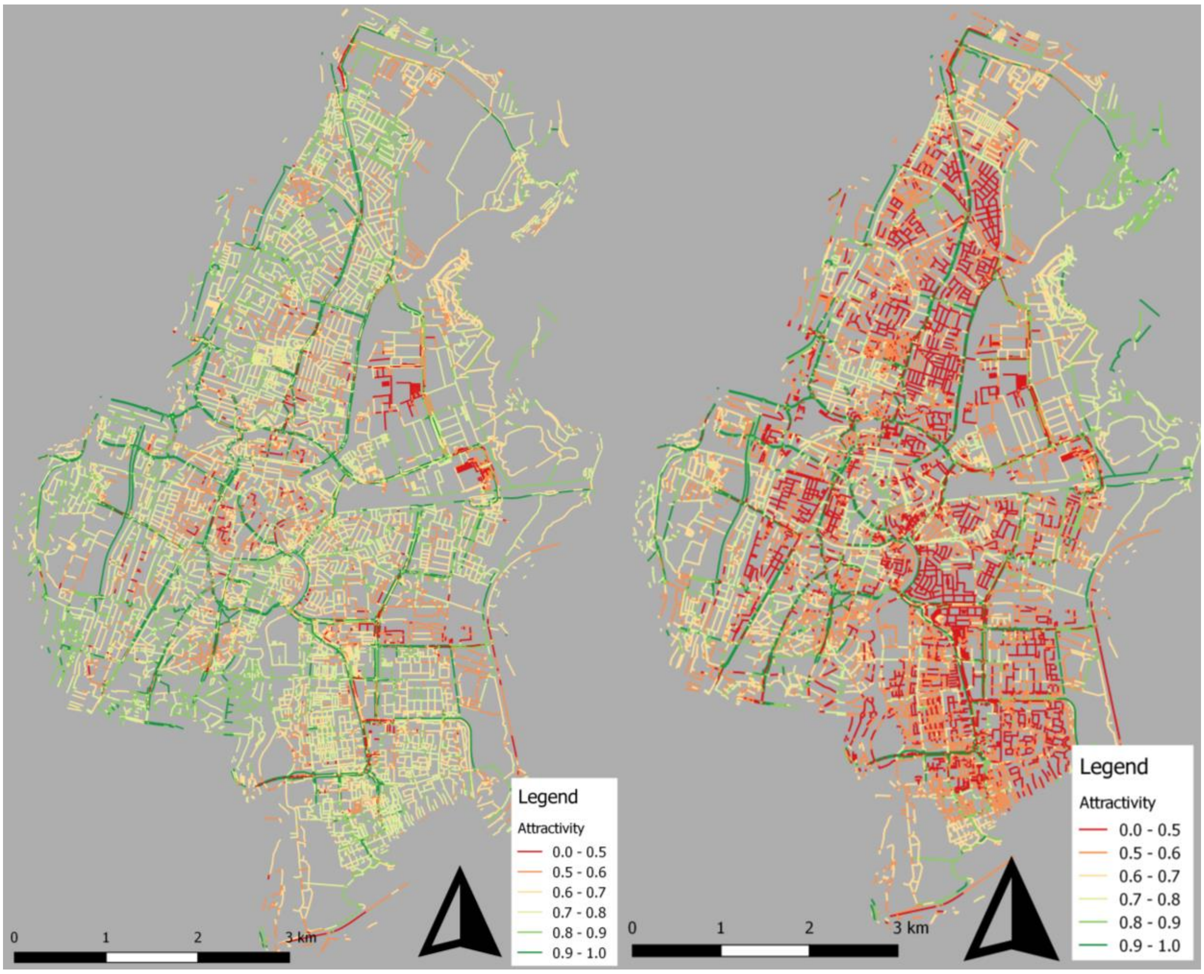

FIGURE 26: ATTRACTIITY MAP OF HAARLEM USING THE SCALE OF THE ENSCHEDE MODEL A) THE MODEL OF ENSCHEDE PROJECTED ON HAARLEM (SAME AS FIGURE 18), B) THE MODEL OF HAARLEM

Appendix C: Attractivity Map including Greenery

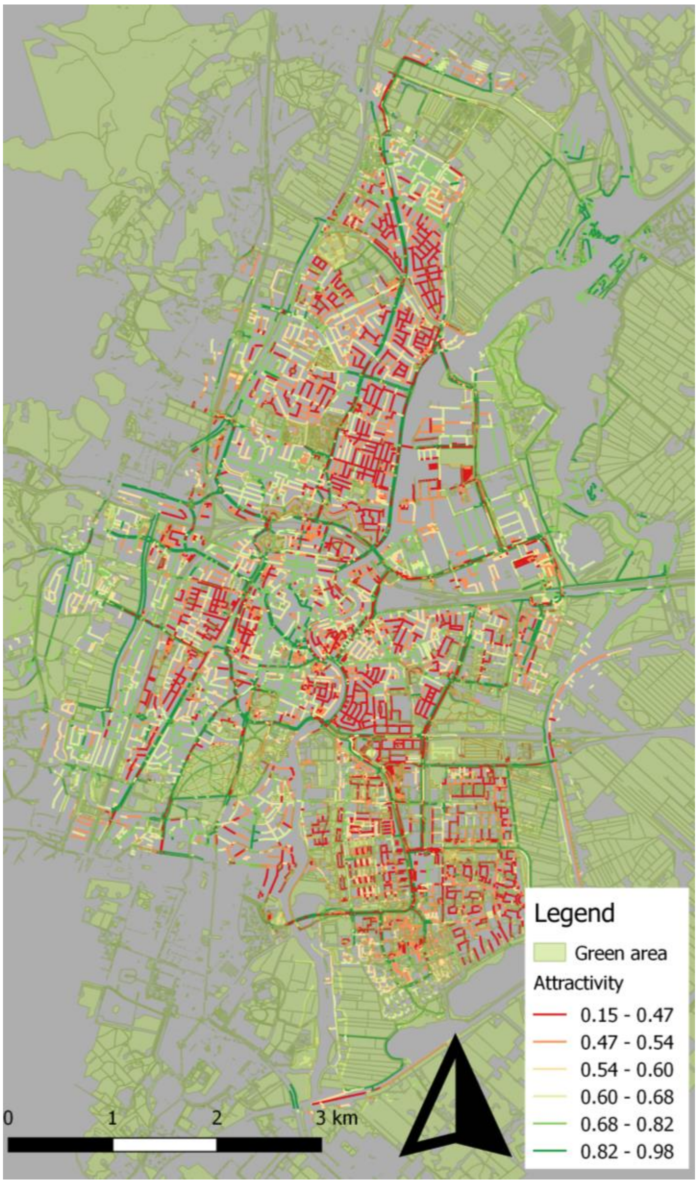

FIGURE 27: ATTRACTIVITY MAP OF HAARLEM INCLUDING GREEN AREA