Geographical Information Management and Applications

Final Thesis Report

Cycling Behaviours: Minimising Travel Distance, Minimising Travel Time and Continuous Cycling

---

Author:
Jimme Smit
(5732352)

Supervisor:
Dr. C. (Kees) Maat
TU Delft

Professor:
Prof.dr.ir. P.J.M. (Peter) van Oosterom
TU Delft

Date:
June 2nd 2021

Preface

Here before you lies my thesis research: *"Cycling Behaviours: Minimising Travel Distance, Minimising Travel Time and Continuous Cycling"* Ever since I was young I have cycled to school, to football and to friends. I have cycled numerous routes to familiar and new destinations. Often I would take the same route, but sometimes I would deviate and change things up. However, I have never actively thought about which route I would take and why. This is why, when I saw the possibility to do my thesis research into cycling behaviour, I was immediately intrigued. What is it that moves me and other people to cycle certain routes. I decided that I preferred to take efficient routes that were either the shortest, saved me time or allowed me to cycle continuously. I wanted to research to what extent other cyclists do this as well.

This research was done as part of my Master's programme: Geographical Information Management and Applications (GIMA). The research took place during the Covid-19 pandemic, which made the data collection process very difficult. This is why I could not have done this without the help of several people, whom I would like to thank.

First of all there is my supervisor Kees Maat, who helped me throughout the duration of my research. He was available when I had questions and would give me a lot of suggestions as well. He also pushed me to continue adding new things which would improve the quality of the research. I also want to thank Danique Ton who helped with the Path Size Logit. Furthermore, I would also like to thank Harmke Vliek and Maaike Kuiper with whom I cooperated on the data gathering process and the methodology of the research. I want to thank Hong Yan who also helped with the parts of the methodology. Lastly, I want to thank my friends and family who participated in the research and supported me throughout the duration of my thesis.

I hope you enjoy reading my thesis and find it insightful.

Jimme Smit

2 June 2021

Abstract

Research into cycling behaviour often concludes that length is one of the most important factors in Route Choice Analysis (RCA). In contrast, RCA studies for cars find that travel time minimisation is a more important motivator. Furthermore, stated preference studies find that cyclists also prefer to cycle continuously. This research aims to find to what extent cyclists minimise travel distance, travel time and cycle continuously. The hypothesis is that cyclists primarily cycle to minimise travel distance, but that they deviate for a shorter travel time or to cycle continuously. To test this hypothesis, cyclists from in different regions in the Netherlands were asked to record their cycling behaviour for a week and to participate in a survey. Between the origins and destinations of the observed routes labelled route alternatives were generated. Spatial and personal characteristics were added whose influence on cycling behaviour is tested through statistical analysis. The characteristics were selected based on what influences these cycling behaviours according to previous studies. The results show that cyclists do prioritise travel distance minimisation, despite not always taking the shortest route. This indicates that cyclists also prioritise other motives. However, from the results it could not be concluded that these motives were travel time minimisation and continuous cycling. This is despite the fact that the participating cyclists did state to have these motives in the survey. However, there are two exceptions. Firstly, men were shown to minimise travel time to some extent. Secondly, when cycling highways are present, people do seem to cycle continuously. This means that under certain circumstances travel time minimisation and continuous cycling were the main alternatives to minimising travel distance, but that generally this is not the case.

Table of Contents

List of Figures and Tables

Figures

Tables

1. Introduction

1.1 Problem and Context

The Netherlands is the most prominent cycling country in the world. Nowhere in the world is the cycling infrastructure as well developed and the cycling culture as ingrained as in the Netherlands (Holligan, 2013). Dutch cities consistently rank high according to the Copenhagenize index (2019). This is due to extensive and safe cycling infrastructure which allows for almost all destinations to be reached by bicycle. The cycling culture also plays a large role. Cycling is learned at a young age. Furthermore, most car drivers tend to be more sympathetic to cyclists, since they're often cyclists as well. These factors allow for a higher cycling participation (Wardlaw, 2014). This is especially the case in the urban environment where the cycling infrastructure in the Netherlands is particularly good (Ton, Cats, Duives & Hoogendoorn, 2017). Around 70% of the trips in major cities like Amsterdam and the Hague are made by bike (Holligan, 2013). Moreover, in the Netherlands cycling is also often used as a means to commute. This is encouraged by the proper infrastructure but also facilities for cyclists by the employer and colleagues that cycle as well (Heinen, Maat & Van Wee, 2013).

Such extensive, safe and advanced cycling network in the Netherlands gives cyclists the freedom to cycle anywhere in the Netherlands. Consequently, this results in many plausible route alternatives from which a cyclists can choose when going from their origin to their destination. Route Choice Analysis (RCA) is the analysis of why people choose a specific route. Although RCA is often applied to car travel such as in Papinksi & Scott (2011), it can also be applied to bicycle travel such as in Ton et al. (2017).

Research into cycling behaviour often concludes that length is one of the most important factors in choosing a route. Cyclists prefer to cycle a short and efficient route to their destination. In RCA studies for cars, travel time minimisation is already an observed phenomenon. The travel time is the focus of the research rather than the route length (Ehrgott, Wang, Raith & Van Houtte (2012). Stated preference studies such as Van Duppen (2012) also see that cyclists in the Netherlands cycle to minimise their travel time and cycle continuously, as they avoid traffic lights and stop signs, in order to not have to stop and waste time. This indicates that travel time minimisation and continuous cycling are also cycling behaviours which should be looked into in order to analyse how cyclists choose their route. The research gap is thus that no empirical research has been done specifically into these two cycling behaviours before. Instead most research focusses on separate influencing characteristics.

1.2 Scientific Relevance

Research outside of the Netherlands shows that a lack of a well-developed cycling infrastructure influences cycling behaviour. Safety concerns cause cyclists to deviate from the shortest route (Menghini, Carrasco, Schüssler & Axhausen, 2010; Yeboah & Alvanides, 2015). Furthermore, research in- and outside of the Netherlands also observes that cyclists often prefer cycling through an attractive environment (Kalter, Coffeng & Groenendijk, 2018). The Dutch context allows for a very free choice of possible route alternatives due to the safe and extensive cycling infrastructure. Cycling in the Netherlands is seen as merely an exercise method or leisure activity but also a serious mode of transportation for utilitarian cycling purposes (Heinen et al., 2013). This context allows for the possibility to research to what extent cyclists cycle in more efficient ways. There are three ways in which a cyclist can be efficient; by minimising their travel distance, by minimising their travel time or by cycling as continuously as possible (Broach, Dill & Gliebe, 2012; Dill & Gliebe, 2008; Ehrgott et al., 2012; Van Duppen, 2012). By primarily focussing on these three cycling behaviours, new insights into route choices by cyclists can be gained.

1.3 Societal Relevance

Dutch cycling paths are becoming busier. Especially in cities, congestion is becoming an increasingly more common sight on Dutch cycling paths (Huiberts, 2020). This increased business is cause for the ANWB (Dutch organisation for traffic and tourism) to advocate for adaptions in Dutch cycling infrastructure to accommodate this increased business (ANWB, 2020). The recent Covid-19 crisis has added to this problem as people are choosing the bicycle over public transportation as a way to avoid being close to other people. Furthermore, cyclists often take informal routes and show anarchistic cycling behaviours such as running red lights, cycling over pavements, taking elephant paths (Van Duppen, 2012, p. 64&65). Te Brömmelstoet (2014) argues that the behaviour of these cyclists is a reflection on the infrastructure. This civil disobedience allows policy makers to learn what makes good cycling infrastructure. A better insight into the motives of Dutch cyclists, and the extent to which cyclists aim to minimise travel distance, minimise travel time and attempt to cycle continuously, can aid in the way in which cycling infrastructure is planned in the Netherlands.

1.4 Research Objective

The goal of this research is to discover the extent to which cyclists cycle according to the three motives that were introduced; minimising travel distance, minimising travel time and continuous cycling. Furthermore this research aims to gain more insight into the influencing characteristics are on these behaviours. The main research question that is central to this research is:

"To what extent do cyclists minimise travel distance, minimise travel time and cycle continuously and which characteristics influence these behaviours?"

In order to answer this main research question, there are six sub questions that need to be answered:

- *"How can minimising travel distance, minimising travel time and continuous cycling be defined and measured?"*
- *"To what extent do cyclists state to minimise travel distance, travel time and cycle continuously?"*
- *"To what extent do cyclists minimise travel distance?"*
- *"To what extent do cyclists minimise travel time?"*
- *"To what extent do cyclists attempt to cycle continuously?"*
- "*Which characteristics influence cycling behaviour and in what way? "*

The first sub question deals with the definition of the three behaviours that are central to this research. These behaviours need to be defined because they have not been researched to the same extent as travel distance and cannot be measured as easily. This is because, for cyclists, there are no speed limits (Broer, 2017). Thus making it harder to calculate the fastest route for cyclists than it is for car travel. In RCA research into cars, comparing the fastest route to shortest route is more common (Papinski & Scott). Cycling speed is wholly dependent on the cyclists themselves and there is also little to no data on cycling congestion (Mobiliteitsplatform, 2020).

The second sub question goes into the extent to which cyclists state that they take routes that are short, fast and continuous. This will allow for comparisons with the stated preference study by Van Duppen (2012) and this allows for a comparison between what is stated and observed. This data is obtained through a survey, which also gathers personal information about the participants.

The next three sub questions look into the extent to which each of these three cycling behaviours are present. This is done by tracking cyclists using GPS trackers and comparing the cycled routes with labelled route alternatives. Labelled alternatives represent certain cycling motives and are not randomly generated (Prato, 2009). Furthermore, the characteristics that define these cycling behaviours are used to find to what extent cyclists show these behaviours.

The last research question answers which characteristics influence cycling behaviour. This data is taken through a spatial analysis, in which data from the environment is attributed to the observed routes and the route alternatives for the statistical analysis. The personal characteristics are also attributed to the routes to determine the influencing characteristics.

1.5 Research Scope

In this thesis, it is thus hypothesised that cyclists primarily cycle to minimise travel distance but that they deviate for a shorter travel time or to cycle continuously. The research goal is limited to researching these 'efficient' cycling behaviours. This means that this thesis will forego further research into other major cycling motives and influencing characteristics that are found in other research, such as cycling safe or attractive routes. Furthermore, this research is limited to the Netherlands. However, within the Netherlands there is no centralised research area, as respondents come from all over the Netherlands. This study is a revealed preference study as cyclists are tracked using GPS trackers in order to determine cycling motives. However, participants are asked about the extent to which they minimise travel distance, minimise travel time and cycle continuously in the survey.

1.6 Thesis Outline

In chapter 2, an overview of the literature and research related to this research topic is provided, in order to answer the first sub question. The chapter concludes with a conceptual model that gives an overview of which relations are researched in this thesis. Chapter 3 describes the process of the data collection and the methodological steps that are taken to obtain the results from the data collection. Chapter 4 presents the results from the survey and the statistical analysis, in order to answer the sub questions. In chapter 5 the results and the methods are reflected upon and interpreted. Chapter 6 concludes with an answer to the sub questions and the main question of this research.

2. Literature Review

In this chapter the literature of other studies into cycling behaviour is discussed. The first sub question: “How can minimising travel distance, minimising travel time and continuous cycling be defined and measured?” is answered in this chapter as well. This chapter starts with defining the three cycling behaviours, followed by an overview of the influencing characteristics on these behaviours that are extracted from the literature. The chapter concludes with the resulting conceptual model.

2.1 Shortest Path and Minimising Travel Distance

Minimising travel distance can be defined as the extent to which the cyclists take the shortest route or prefer shorter route alternatives. Several researchers have found that the route length is one of the most important factors in determining the route choice (Menghini et al., 2010; Casello & Usyukov, 2014; Hood, Sall & Charlton, 2011). This behaviour can be measured by looking at the overlap with the shortest path and assessing to what extent the length influences route choice (Menghini et al., 2010). Broach et al. (2012) found that cyclists have a general preference for shorter routes. Relative deviation from the shortest path matters rather than absolute deviation. Thus, depending on the length of the route, certain levels of deviation can be expected. In Broach et al. (2012) this was found to be a fifth of the total route length. In the study of Ton et al. (2017) in Amsterdam, the importance of distance was found as well, although to a lesser extent than the previously mentioned studies. Still, a third of cyclists took the shortest path. Around 40% chose a route that is only 10% longer than the shortest path.

2.2 Fastest Route and Minimising Travel Time

Although the length of a route plays an important role in route choice, there are other motives which may prevent cyclists from taking the shortest route to their destination. Some examples are that the shortest path is considered to be unsafe or the cyclist has a lack of knowledge on the available routes or the cyclist has a preference for alternative routes (Dill & Gliebe, 2013). Especially the unsafety of the shortest route is a big obstacle that is observed in studies outside of the Netherlands. Cyclists prefer separated cycling paths and routes away from heavy car traffic (Yeboah & Alvanides, 2015). Kalter et al. (2018) found that cyclists prefer their cycling experience over distance minimisation. Cyclists deviate from the shortest path to cycle through an attractive environment.

Another motive for deviating from the shortest route is to minimise travel time. This is the second cycling motive central to this research. Although research into the fastest route and travel time costs is very established in RCA research of cars, this is not researched to the same extent as for cycling behaviour as it is more difficult to calculate (Ehrgott et al., 2012). Broach et al. (2012) found that there is a slight difference between travel distance and travel time. This was confirmed by Hull & O’Holleran (2014) who categorises both the shortest and fastest route under route directness. Although the shortest route and the fastest route mostly overlap, there are various influencing characteristics which affect travel time negatively that cause a deviation from the shortest route. These are discussed in section 2.4.

Minimising travel time can be defined as the extent to which cyclists take the shortest path and avoid characteristics which delay travel time and take routes with characteristics which decrease travel time (Broach et al., 2012). This can be measured by looking at the overlap with the fastest route and assessing to what extent the length and characteristics that cause time delay influence route choices. Characteristics such as traffic lights, stop signs, and gradient can make routes slower. In Menghini et al. (2010) these factors were reasons for cyclists to avoid the shortest route. In contrast, Broach et al. (2012) also found some positive influences from traffic lights as they can increase travel flow.

2.3 The Most Continuous Route and Continuous Cycling

Another reason why cyclists deviate from the shortest path is to cycle continuously. This was discovered by Van Duppen (2012, p. 64), who found that cyclists have a preference for cycling without interruptions. Cyclists want to get to their destinations and prefer not to incorporate intermediate stops within their travels, such as going to a store before going to work.

However, continuous cycling is not limited to destinations and intermediate stops. When going from their origin to their destination, cyclists also want to continue cycling and prevent waiting for traffic lights and stop signs. This results in cyclists running red lights or circumventing red lights in order to continue their flow. This is especially the case when cyclists are familiar with the crossings and thus know whether they can run a red light or during rush hour when the flow of many cyclists continues despite red lights (Van Duppen, 2012; p. 64 & 65). Cycling speed is also related to continuous cycling. Cyclists with higher speeds avoid turns and other barriers that inhibit continuous cycling. Segregated cycling lanes allow for speed and flow to be maintained (Hull & O'Holleran, 2014).

Continuous cycling can thus be defined as the extent to which cyclists avoid barriers which affect cycling flow and prefer route characteristics that allow for continuous cycling. This can be measured by looking at the overlap with the most continuous route and assessing the characteristics that positively and negatively affect cycling flow.

2.4 Influencing Characteristics

There are many influencing factors which have been identified in route choice analysis of cyclists. In this part the influencing factors that affect the motives central to this research are discussed. Several types of influencing characteristics are distinguished; spatial characteristics, personal characteristics, trip purpose and the type of bicycle that is used.

2.4.1 Spatial characteristics

The first type of influencing characteristics pertain to the cycling network and the infrastructure along the route, as well as the physical surroundings of the route.

Route length

The first characteristic that influences cycling behaviour is route length. This was found to have a strong influence in various studies (Broach et al., 2012; Menghini et al., 2010; Ton et al., 2017). Especially for distance minimisation the route length is the most important influencing factor. However, Broach et al. (2012) also found that route length influences travel time minimisation as there is a strong overlap between these two cycling behaviours.

Route obstacles

The presence of route obstacles are one of the primary characteristics to impact the cycling behaviours central to this research. The term route obstacles in Dane et al. (2019) consists of stop signs, traffic lights and turns, especially left turns. Ton et al. (2017) and Menghini et al. (2010) found evidence that route obstacles such as stop signs and traffic lights affect cycling behaviour. Additionally, Van Duppen (2012) found that cyclists often tend to avoid traffic lights in order to continue cycling and minimise travel time. There are two ways in which this is often done. One is by avoiding routes which go past traffic lights. The other method is by running red lights in order to prevent having to stop. Te Brömmelstoet (2014) adds to this that traffic lights cause a pile up of cyclists when waiting for a red light. This prevents other cyclists, that could normally turn right without waiting for the traffic lights, from continuing on their journey. In contrast, Menghini et al. (2010) and Broach et al. (2012) also discovered some positive impact from traffic lights as they can reduce waiting times at busy intersections. Hull & O' Holleran (2014) confirm this as there intersections in the Netherlands with cycle priority traffic lights. Another example of route obstacles are turns at intersections. Cyclists were found to minimise the number of intersections per kilometre travelled (Ton et al., 2017). Broach et al. (2012) further found that the turns on intersections negatively affect travel times of cyclists. Travel time increases for cyclists by going left and straight at intersections because cyclists have to cross other traffic. This especially the case for left turns. However, at lower traffic intensity the difference between left and right turns was found to be less significant.

Gradient
The gradient of a route also impacts cycling behaviour as cyclists tend to avoid routes with higher average or maximum gradients despite it being a shorter route (Menghini et al., 2010). However, the impact of the gradient is not as big as that of the route length, which means cyclists will consider a route with a higher gradient when the alternative rout deviates too much from the shortest path (Menghini et al., 2010). In the Netherlands there is very little relief, so the effects of this characteristic might not be present to the same extent as outside of the Netherlands.

Traffic volume
Traffic volumes also negatively impact cycling behaviour. Especially during commutes it negatively affects the three cycling behaviours central to this research (Li, 2017). Furthermore, traffic volume affects cycling safety and perceived cycling safety, which in turn is shown to negatively influence these three cycling behaviours (Li, 2017; Yeboah & Alvanides, 2015). Traffic volume is not exclusively about car traffic volume but also the traffic volume of bicycles. In the city of Utrecht cyclists have chosen alternative routes to avoid traffic at certain times at rush hour. This is especially the case for those who cannot opt to travel at a different time outside of rush hour. The level of traffic volume on cycle paths is additionally increased due to the corona crisis, since people take the bicycle in order to avoid public transportation (Huiberst, 2020).

Separated cycling paths
The cycling network in the Netherlands is very extensive but not all cycling paths are separated from other traffic. Both Yeboah & Alvanides (2015) and Broach et al. (2012) have found that a lack of separated cycling paths results in a more unsafe cycling experience. This often prevents cyclists from taking the fastest or most direct route. According to Hull & O' Holleran (2014), separated cycling paths further encourage fast and continuous cycling as they are more spacious, allowing for the possibility to overtake other cyclists and maintain cycling speeds. Ton et al. (2017) did not observe a preference for separated cycling paths in the Dutch context in Amsterdam. This may be due to traffic regulations which lower speed limits for cars in areas where traffic is not separated.

Cycling highways
Part of the Dutch cycling network consists of cycling highways or fast cycling routes, which are used to connect offices, schools and main services with the areas where people live. Cycling highways encourage higher cycling speeds and are thus also known as 'fast cycling routes'. The cycling highways are primarily used for commuting, leisure cycling and sports cycling and allow for continuous cycling (Liu et al., 2019; Van Lierop et al., 2020).

Cycling path surface
Cycling path surface influences the physical experience of cycling (Van Duppen, 2012; p. 26). Several researchers also found that the surface quality influences route choice with regards to travel time, safety and comfort (Dill & Gliebe, 2008; Van Lierop et al., 2020). The state of the cycling path surface can impact the cycling experience because smooth surfaces are more beneficial for cycling flow and cycling speed, whereas a rougher surface can negatively impact cycling speeds.

2.4.2 Personal characteristics
The second type of influencing characteristics are personal characteristics of the cyclists, their motives and the bicycle that they use.

Gender
There are several differences between men and women when it comes to cycling behaviours. Both men and women rate minimising travel distance as the most important factor for their route choice. However, women rated this significantly higher than men (Dill & Gliebe, 2008). In contrast, Bernardi et al. (2018) found that especially men tend to take the shortest routes. When it comes to minimising travel time some observations were made as well. Women, on average, cycle at a lower speeds than men. Additionally, women tend to prefer cycling on routes with lower traffic volumes. This is because they generally value safety more than men which could result in them taking slower routes over faster alternatives (Dill & Gliebe, 2008). It should be noted that these observations were made in research outside of the Netherlands. Some of the trends which are observed above can be related back to cycling safety. As the cycling safety in the Netherlands is generally better than abroad this could change the outcomes. Ton et al. (2017) did not incorporate personal characteristics into the results from the research. However, it was recommended to take these into account in future research.

Age
Differences in the age of the cyclists also affects cycling behaviour. According to Menghini et al. (2010) the age of the cyclists determines the extent to which they make minimise travel distance, minimise travel time and cycle continuously. Older people tend to cycle shorter distances compared to younger cyclists (Dill & Gliebe, 2008). Young people also tend to take faster routes compared to older cyclists, who prefer more attractive routes (Kalter et al., 2018). This is confirmed by Stinson & Bhat (2003) who also found that older people are less likely to minimise travel time, which might be because they have more time available. Furthermore, older cyclists also prefer to avoid busier intersections, as they value a safe cycling environment more than younger cyclists.

Route familiarity
The extent to which the cyclist is familiar with the environment in which they cycle and the cycling experience are other important personal characteristics which determine the chances of the shortest or fastest path being taken. This is observed in Van Duppen (2012), where cyclists elaborate on specific cycling behaviours to decrease travel time and allow them to continue cycling as opposed to having to stop for traffic lights. Additionally, this was observed in Van Lierop et al. (2020), where cyclists did not take the shortest path due to lack of knowledge of the cycling network. Furthermore, less experienced cyclists are less likely to take the shortest route (Casello & Usyukov, 2014). However, this inexperience is related to the extent to which cyclists prefer safe route alternatives rather than the extent to which they have motives that are central to this research.

Trip Purpose
Another important influencing factor on cycling behaviour is the purpose of the trip. Cycling purposes are often divided into leisure cycling and utilitarian cycling. Utilitarian cycling is further divided into commuting and non-commuting. Broach et al. (2012) defines non-commuting utilitarian cycling as cycling to visit a store. According to Li (2017), cyclists that commute prefer the shortest path more than leisure cyclists do. Commuter cyclists are also more sensitive to taking the shortest path compared to non-commuting utilitarian cycling trips (Broach et al., 2012). In contrast, Stinson & Bhat (2003) found that cyclists are willing to take a longer route in order to avoid busy intersections and major roads (Stinson & Bhat, 2003). According to Broach et al. (2012), commuter cyclists are willing to avoid these intersections to a lesser extent than cyclists with other cycling purposes. Furthermore, commuting cyclists less sensitive to other infrastructure characteristics and focus mainly on distance minimisation. Additionally, commuting cyclists cycle at higher average cycling speeds than cyclists that cycle for utilitarian non-commuting reasons (Broach et al., 2012).

Type of Bicycle
The last personal characteristic is the type of bicycle that is used. There are several types of bicycles: urban bikes, sports bikes (hybrid), tandems, folding bikes, E-bikes, speed pedelecs, mountain bikes and racing bikes etc. For example, most standard bicycles are used for average cycling behaviour and often offer a more relaxed cycling experience. In contrast, racing bikes are used for cycling at higher speeds (Van Duppen, 2012; p. 27; Spinney 2007). E-bikes decrease the likeliness to take the shortest route but increase the likelihood to take the fastest route (Plazier, Weitkamp & Van den Berg, 2017). Additionally, people with E-bikes want to maximise their speed by taking cycling highways (Van Lierop et al., 2020). Speed pedelecs also have higher cycling speeds compared to normal bikes, but even to a greater extent compared to E-bikes (Steintjes, 2016).

2.5 Conceptual Model
In Figure 2.1 the conceptual model is displayed. The conceptual model gives an overview of the influencing characteristics on the cycling behaviours that are central to this research, based on the literature review. The influencing characteristics are divided into spatial characteristics and personal characteristics.

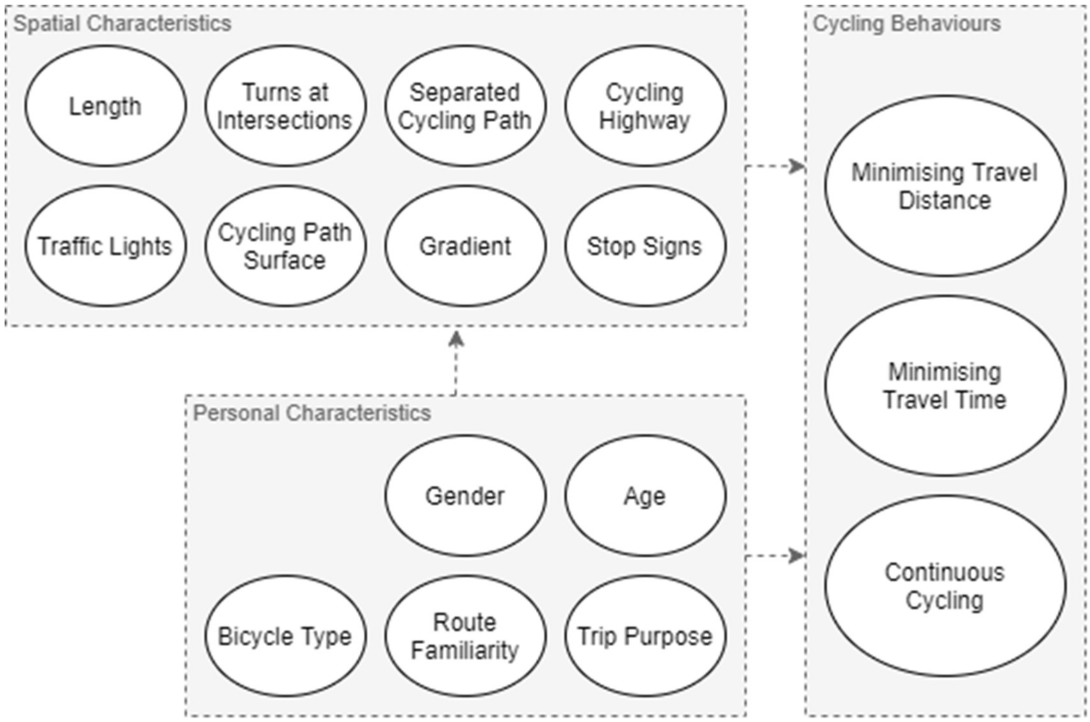


Figure 2.1: Conceptual Model

2.5.1 Spatial Characteristics
Spatial characteristics are aspects of the cycling network, the infrastructure and physical surroundings that affect cycling behaviour. There are spatial characteristics that negatively influence cycling behaviours and characteristics that positively affect cycling behaviour. Length negatively influences minimising travel distance and minimising travel time. Turns at intersections also negatively affect minimising travel time and continuous cycling but in different ways. For minimising travel time turns have a negative influence based on the time costs of the turns. Taking a right turn is assumed to be the least costly because it crosses no other traffic and right turns for cyclists can often go through red lights. Going straight on an intersection is also a time cost because the cyclist crosses other traffic. Left turns are especially costly as traffic from two ways is crossed. Gradient negatively affects travel time. Cycling path surface does the same but also affects continuous cycling. Lastly, stop signs and traffic lights also negatively affect travel time and continuous cycling. However, traffic lights can sometimes

also increase flow and speed. Although this is under specific circumstances. Other positive influences are separated cycling paths and cycling.

2.5.2 Personal characteristics

Personal characteristics also influence cycling behaviours in positive and negative ways. The personal characteristics also interact with the spatial influences on cycling behaviour. Both men and women cycle to minimise travel distance and travel time. Younger people tend to minimise travel time more than older people, while older people cycle shorter routes. People with more route knowledge are also more likely to minimise travel distance and travel time. When it comes to trip purpose, cycling for commuting and other utilitarian purposes positively influences the cycling behaviours, while cycling for recreational purposes negatively affects cycling behaviour. Lastly, faster bicycle types positively influence travel time minimisation and continuous cycling.

2.5.3 Defining minimising travel distance, minimising travel time and continuous cycling

Based on the findings in the literature the first sub question can be answered. The question is "How can minimising travel distance, minimising travel time and continuous cycling be defined and measured?" Minimising travel distance is defined by a negative influence of route length on route choice and overlap with the shortest route alternative. Minimising travel time is defined by a negative influence of length and other time delaying route characteristics and a positive influence of characteristics that allow for faster cycling. Continuous cycling is defined as a negative influence of characteristics that disrupt cycling flow and a positive influence of characteristics that enable cycling flow.

3. Methodology

In this chapter the methodology of this thesis research is discussed. It is important to note that the methodology part of this research is done in cooperation with two other Master's students and a PHD student that research cycling behaviour. This was done in order to gather more data. The chapter starts off with an overview of the methodological steps taken in this research. Then the operationalisation of the influencing characteristics that are taken into account in this research are discussed. This is followed by the data collection is described, followed by a description of the data preparation. The chapter concludes with a description of the methodology of the analysis phase. An overview can be seen in Figure 3.1.

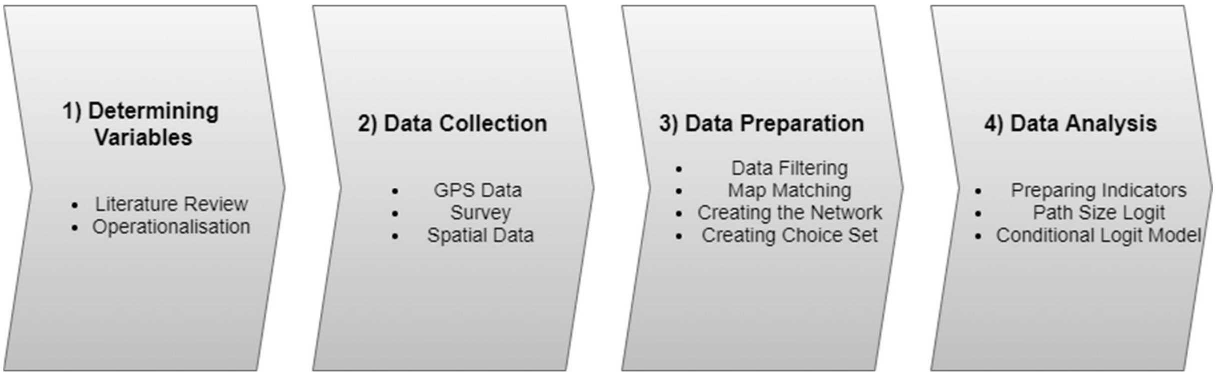


Figure 3.1: The four phases of the research methodology

3.1 Determining Variables

The first phase is determining the relevant variables that influence cycling behaviour through the literature review and operationalisation. The goal of the literature review was to answer the first sub question: *"How can minimising travel distance, minimising travel time and continuous cycling be defined and measured?"* The result of this phase is a definition for each of the three cycling behaviours central to this research and the conceptual model which provides an overview of the important influencing characteristics. The conceptual model displays the summary of the main influences on the three cycling behaviours, found in literature. These are categorised as spatial characteristics and personal characteristics. Not all of the influencing characteristics from the conceptual model can be used in this research. For some there is no (suitable) data available and others cannot be measured. The remaining variables are listed below. The way in which the variables are operationalised is as follows:

- Route Length: The length of the route.
- Turns:
  - Turn costs: The relative time costs or turns. This variable is used for minimising travel time.
  - Number of turns: The number of right and left turns taken in a route. This variable is used for continuous cycling.
- Separated cycling path: Percentage of route that consists of separated cycling paths.
- Cycling highways: Percentage of route that consists of cycling highway.
- Traffic lights: The number of traffic lights on a route.
- Stop signs: The number of stop signs on a route.
- Gender: Gender of the participant.
- Age: Age of the participant.
- Trip purpose: The trip purpose of the route.
- Bicycle type: The bicycle used during the trip.

- Urban/Rural/Nature: The variables urban, rural and nature are added because the research area is in both urban, rural and nature areas. This gives insight into the influence of the environment.

The influencing characteristics that are left out from this research are listed below along with the reason that they are left out.

- Cycling path surface: The reason that this variable is left out is because there was no reliable data source for the cycling path surface. Although this is taken into account in the Fietsersbond cycling network, a lot of data is missing.
- Gradient: Gradient is left out because the Netherlands is a flat country with little height differences, especially in the research area.
- Route familiarity: This variable is different for each of the routes that is cycled by the participants. It is almost impossible to acquire the level of knowledge of the surroundings from all participants for every route.

3.2 Data Collection

The second phase is the data collection. In this phase the trade-off between stated and revealed preference studies is discussed. Furthermore, the way in which the GPS data and other data is collected is discussed.

3.2.1 Stated versus revealed preference

There are two main methods in which route choice analysis for bicycle travel is often carried out. These are stated preference studies and revealed preference studies (Casello & Usyukov, 2014). Through surveys cyclists are asked which route characteristics influence them to take a specific route. The weakness of this methodology is that the real behaviour may differ from the stated behaviours. This is where revealed preference studies come in where cyclists are asked to draw their preferred routes (Casello & Usyukov, 2014). Another method to research revealed preference has been through GPS tracking. This is increasingly becoming the main methodology for route choice analysis of cyclists as technical advancements have made this method more accurate (Montini, Antoniou & Axhausen, 2017; Skov-Petersen, Barkow, Lundhede & Jacobsen, 2018). Cyclists can now be tracked for a longer duration and in larger numbers (Menghini et al., 2010). A disadvantage of revealed preference studies is that they are more difficult and expensive compared to stated preference studies (Dane et al., 2019). In this research the choice is made to do a revealed preference study where the participants are tracked. Their routes are map matched to a network. Information on the routes is gathered through a survey and spatial analysis. This data can then be used in a statistical analysis. Although this research is a revealed preference study, some questions are asked about cycling motives of the participants.

3.2.2 Data

In the operationalisation the influences that are researched in this research are listed. For each of these data is available. In Table 3.1, the data and its sources can be found. Firstly, there is the GPS data which is acquired using GPS trackers, as was determined in 3.2.1. Secondly, the cycling network comes from the Fietsersbond. This is a shapefile which needs to be turned into a network file. The Fietsersbond cycling network also provides data on various network characteristics. Other network characteristics that are not available in the Fietsersbond cycling network are retrieved from OpenStreetMaps (2020). The network data is coupled to the routes using spatial analysis. Personal data and data on trip purpose and bicycle type are retrieved from the survey. Lastly, additional data that is used for creating the route alternatives and determining the route purpose are stated with their sources.

Table 3.1: Data and sources

Data	Source:
GPS cycling data	GPS trackers
Cycling network	Fietsersbond (n.d.)
Separated cycling paths	Fietsersbond (n.d.)
Cycling highways	Fietsersbond (n.d.)
Urban	Fietsersbond (n.d.)
Traffic lights	OpenStreetMaps (2020)
Stop signs	OpenStreetMaps (2020)
Supermarkets and shops	OpenStreetMaps (2020)
Age	Survey
Gender	Survey
Trip purpose	Survey
Bicycle type	Survey
Neighbourhoods and PC6 regions	CBS (2020)

3.2.3 GPS Trackers

The use of GPS trackers in RCA applied to cycling behaviour is not a novel methodology. There are generally two ways of collecting GPS data, using smartphones or separate GPS trackers. Each of these methods has advantages and disadvantages. GPS trackers have a higher accuracy than smartphones. Additionally, the process of tracking drains the battery of smartphones which means it is more invasive for the participant. The downside of GPS trackers is that they, on average, result in a lower number of participants (Pritchard, 2018). In the end the choice was made to use GPS trackers. The GPS trackers that are used are small, allowing participants to carry them with them on their trips when they are cycling. The trackers do not need to be turned on but track continuously. This does mean that non-cycling trips are recorded as well. The GPS trackers have batteries as well. Respondents are instructed to charge the GPS tracker. These instructions can be read in Appendix 1. The GPS trackers sends GPS coordinates of its location every 5 seconds resulting in a table with a location and time. There are only 20 GPS trackers available for this research. In order to gather enough data, these trackers need to be spread between the participants over time. This means that the data collection takes four weeks.

3.2.4 Survey

After the GPS trackers are collected from the participants, they are asked to participate in a survey. This is done in order to determine the attitudes of the participating cyclists towards travel distance minimisation, travel time minimisation and continuous cycling. This validates the research question and allows for a comparison between what is stated and what is observed. The way in which the attitudes towards cycling are asked, is through statements of question 11 in the survey (Appendix 2). Furthermore, data is collected on their personal characteristics and data is collected on cycling purpose and bicycle type. Although it is easier to ask participants afterwards what their cycling goal was for each of their routes, this puts more strain on the participants and could lower participation. This is why participants were asked which bicycle they use for which cycling purpose (question 2, 3 & 4 in Appendix 2). Cycling purpose is later determined in the data analysis. If the origin or destination of a trip is within the postal code area of their employment the route is classified as a commute. This is why participants are asked what the PC6 area of their employment is (question 26 in Appendix 2). This way the bicycle that the participant used for a trip can be determined. The answers of the survey are connected to a GPS tracker through the first question which asks respondents to fill in their id number of their tracker (question 1 in Appendix 2).

3.2.5 Participants and research area

Due to the ongoing corona crisis it was not possible to bring GPS trackers to random participants all over the country. This is why family, friends and acquaintances of the three cooperating Master students were asked. This does have impact on the results in several ways. The respondents are less likely to be representative for the Dutch population as there is a possible bias with regards to personal characteristics such as age, gender and level of education, due to the fact that data is collected through friends and acquaintances of the Master's students. Furthermore, participants can record multiple trips. This means that if certain participants cycle more than others, their preferences might be overrepresented. Another way that this impacts the research is that the research area is very large and scattered. Generally, RCA research is concentrated in one location (e.g. Menghini et al., 2010; Hood et al., 2011, Ton et al., 2017). However, in this research participants are from all over the country. In Figure 3.2 the map of the research area can be seen. In total 73 people were asked to participate in the research. Due to issues with some of the GPS trackers, some participants not filling out the survey and the filtering of some routes, routes of 61 people were taken into account for the final research. This resulted in a total of 421 different routes being used in the statistical analysis of the research.

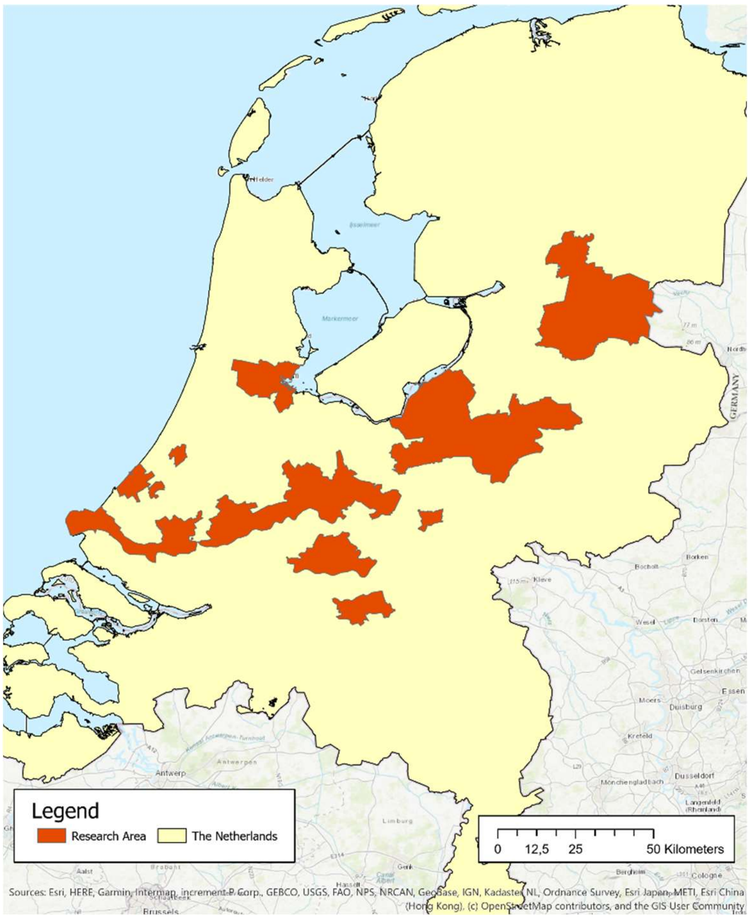


Figure 3.2: Research area

3.3 Data Preparation

The third phase of the methodology is the data preparation. This phase consists of the data filtering, the map matching process, the creation of the cycling network and the creation of choice set.

3.3.1 Data filtering

The first part of the data preparation consists of the filtering of the GPS data. Participants are given a GPS tracker for a period of a week. During this period the GPS tracker is always on. This means that walking trips, car rides and trips with public transportation are recorded as well. Furthermore, because the GPS trackers that are used do not need to be turned on for each trip, the GPS trackers also record at times when the respondent is not on the move. The other modes of transportation and the points that are generated when the tracker does not move need to be filtered out. This is done using a python script. The python script loops through the GPS data and filters them. This script can be found in Appendix 4. The filtering process consists of three steps.

- The first step is filtering out the points where the GPS tracker is not moving. The file contains information on the speed at which the GPS tracker is moving at that moment. All points with a speed of below 5 km/h are filtered out. This removes all the moments when the GPS tracker is not motiving and when the participant is walking. A cut-off of 5km/h is chosen because, on average, people walk 3-5km/h (Huss, Beekhuizen, Kromhout & Vermeulen, 2014).
- The second step is defining what makes a route. In the file all points collected during the week are present. For this research, separate route files are required. By calculating the time difference between each remaining point, a cut-off can be determined. The choice was made to create a cut-off if the time difference between two consecutive points is more than 5 minutes. This is done because if the time is lower, routes can end when people wait at traffic lights or have to look up their route. The gap is not longer than 5 minutes because then the participant can carry out an activity at a destination, meaning the trip should be over.
- The third and final step is to filter the created routes. Once all points of the participant are divided into routes they need to be filtered. This is done by removing all the routes with less than 20 points because they are too short. Then the cycling trips need to be selected from the remaining routes, which is done by looking at the average travelling speeds. Average cycling speeds are 15.8km/h (Fietsersbond, n.d.). Cars and public transportation reach higher speeds of around 50 km/h (Huss et al., 2014). E-bikes and speed pedelecs complicate this matter further as they should be taken into account but have higher average speeds. E-bikes can reach a speed up to 25km/h, while speed pedelecs can reach a speed of up to 45km/h (Plazier et al., 2017). Routes with an average speed between 10 and 30 km/h and a maximum speed of 40 km/h were used. The others were deleted.

3.3.2 Map matching

The second part is the map matching process. After the filtering process, shapefiles of the routes are created. Figure 3.3 shows an example of a point shapefile. The points are not yet placed exactly on the network. Map matching is the process of linking the collected GPS points to the existing cycling network (Schweizer, Bernardi & Rupi, 2016).

Figure 3.3: Point shapefile and network

The process of map matching is a difficult one and comes with certain problems. There are issues with the accuracy from the GPS devices which can make the map matching process more difficult. This is due to interference in the signal. The causes of this interference are different for urban and rural areas. In urban areas (tall) buildings can cause interference. In rural areas trees and other vegetation can cause interference. Another issue of map matching in an urban environment is the fact that cyclists have a tendency to show anarchistic behaviour. This means that they do not use intended infrastructure but cycle over pavements or through parks (Schweizer et al., 2016). This can be seen in Figure 3.4.

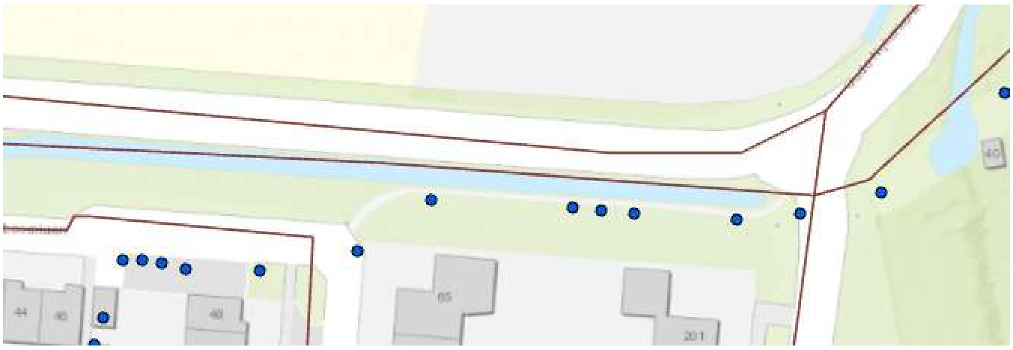


Figure 3.4: Map showing cyclist taking an elephant path

The software that is used is mapmatcher which is created by Simon Scheider. This software takes the GPS points that are gathered and matches them to the network from the Fietsersbond (Schweizer et al. 2016). However, due to various issues with the software not working and the network topology the map matching process was done manually. This resulted in the process taking longer but allowed for the research to continue.

3.3.3 Creating the network

The network that is used in this research is the Fietsersbond cycling network, created by Fietsersbond members. This is similar to the way in which OpenStreetMaps works. Members can draw the network and add information about the segments in the network. The choice for the Fietsersbond cycling network is thus made due to this additional information and the fact that this network is more accurate for cyclists. However, the shapefile that is provided is not ready for use. Outside network data from OpenStreetMaps is added through spatial analysis, fields in the shapefile are reclassified, the network topology is fixed and a network file is created. This consists of 4 major steps:

1. *Spatial joins*

   Additional data that needs to be joined with the network is required for the analysis and the route creation. These are traffic lights, stop signs, supermarkets and shops, bicycle accidents and crime rates per neighbourhood. These are taken from OpensSreetMaps (2020) and joined to the links in the network through spatial joins. The data provided from OpenStreetMaps is in the form of point data which needs to be buffered in order to overlap with the streets. The buffer sizes can be found in Table 3.2. Traffic lights and stop signs have a lower buffer as they are often close to the network. Overlapping buffers are merged because each individual traffic light or stop sign at an intersection is a unique point. Counting them separately would inflate the number of traffic lights and stop signs that are passed. The buffer for supermarkets and shops is 50 meters because people can park their bikes further away from the store which they are shopping at.

Table 3.2: Buffer sizes for outside spatial data

Spatial data	Buffer size
Traffic lights	15 meters
Stop signs	15 meters
Supermarkets and shops	50 meters

2. Reclassification

Data already present within the shapefile needs to be reclassified before it can be used. This data is also required for either the analysis or the route creation. The reason the data is reclassified is twofold. Firstly, almost all the data is in text. In order for it to be used in the analysis or route creation, each field needs to be numerical. Furthermore, a new travel time field is added to display the time it takes to cycle over a segment. This is done using the average cycling speed, which is 15.8km/h or 4.36 m/s (Fietsersbond, n.d.). The calculation can be seen in (1). The reclassifications can be seen in Appendix 3.

$$Travel\ time = \frac{Length}{4.36} \quad (1)$$

3. Fixing network topology

After the outside attribute data is added and the existing attribute data is reclassified, the physical aspects from the network are fixed. The quality of the network topology is not perfect because the network is provided by the Fietsersbond and is generated by its users. Several issues with the network topology can be identified. The main issue is that certain network segments are not connected. Most of these are not clearly visible except when fully zooming in on intersections. This can be seen in Figure 3.5.

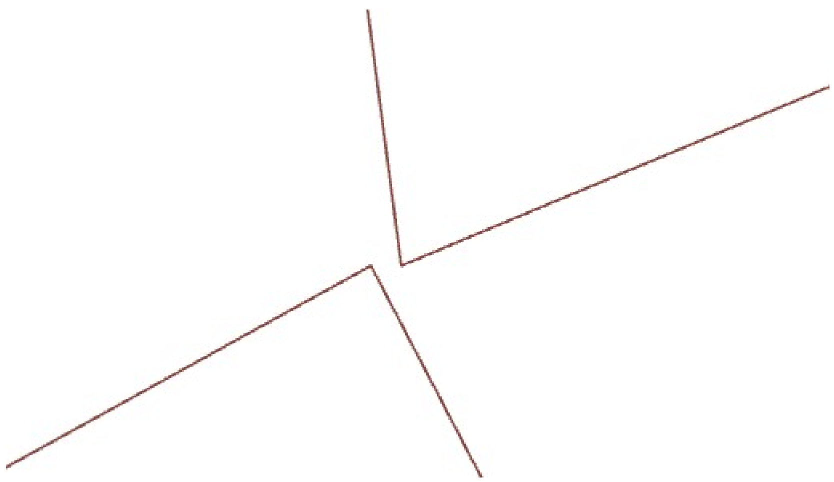


Figure 3.5: Intersection that is not connected

Originally, the goal of fixing the network topology was to allow for the map matching process to be correctly executed. However, even after the topology was fixed the software did not run correctly. Consequently, after the first run of generating the routes, the results proved to be unrealistic. This can be seen in Figure 3.6. The shortest route was a lot longer than the observed route and took a big detour. The cause of these unrealistic routes was that certain issues still remained within the network. Routes are generated by going from point A to point B and taking the segments in between. A segment should be the road segment between the nodes. Each node should be at every intersection or turn where a cyclist can take a turn. However, in reality this was not the case. Some segments continued despite crossing an intersection (Figure 3.7a). Because the network analyst tool in ArcGIS Pro has to take a segment from start to end, this disallowed certain turns to be taken, resulting in unrealistic routes as seen in Figure 3.6. This meant that each segment needed to be split at an intersection (see result in Figure 3.7b). This allowed for better routes to be generated.

Figure 3.6: Unrealistic shortest route before network fix

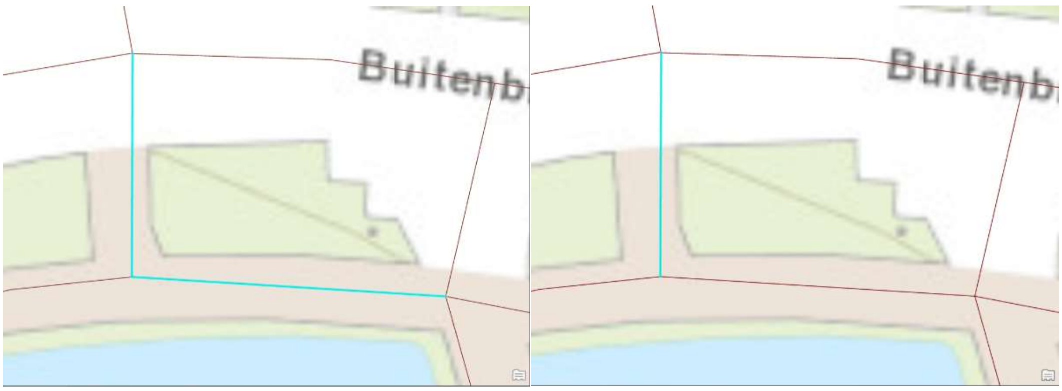


Figure 3.7: (a) Old network: the turn is one segment. (b) New network: the turn is separate segments

4. Create network file

Once the topology is fixed, the shapefile is completed and the network file can be generated. This process is done in ArcCatalog. The creation process consists of several important steps. Global turns need to be able to be modelled. Elevation features are not modelled, because the research area, and the Netherlands in general are mostly flat. Cost and restriction attributes need to be added in order for unique route alternatives to be created. Driving directions were not established in the network dataset due to a lack of data. Once the network dataset is created the network needs to be build. The most important step here are the cost and restriction attributes. Three different cost attributes are added.

1. Length: this cost attribute takes the length field of each segment.
2. Travel time: this cost attribute takes the travel time field of each segment. The cost attribute is a representation of the length expressed in time. The reason this is done is that, in order to add global turn delays, the cost attribute needs to be expressed in a time unit. Turns are determined by looking at the angles between two segments (Figure 3.8). Each turn is then assigned a time penalty.
3. Minimise turns: this cost attribute also utilises the travel time field. However, the global turn delays are different as they punish all turns more in a way to minimise them.

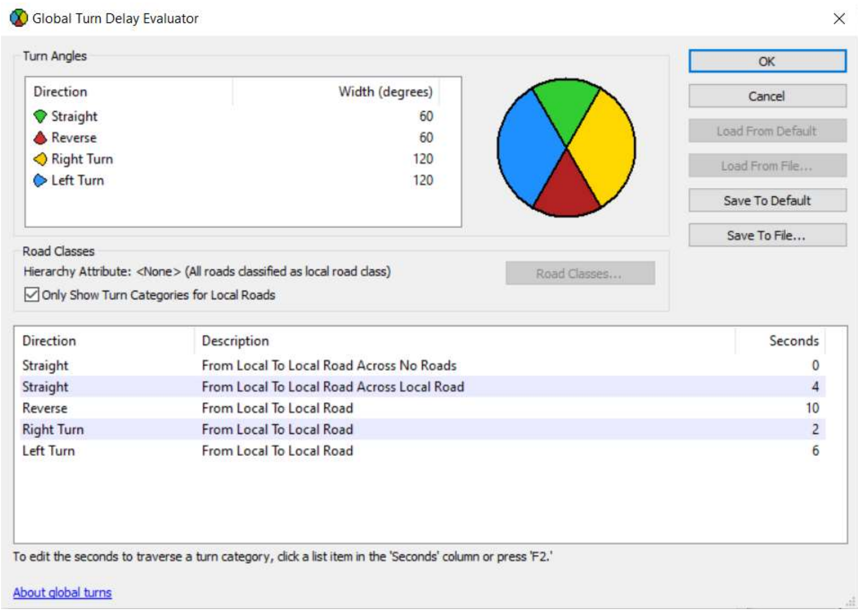


Figure 3.8: *Global turn delay travel time*

Besides the cost attributes there are also restriction attributes. These are attributes which need to be avoided or are preferred for certain route alternatives. These are stop signs, traffic lights, intersections, water, environment (urban, rural, nature), street lighting, crime rates, road type, speed limits and accidents. Which costs and restrictions are used for each alternative is elaborated upon in chapter 3.3.4.

3.3.4 Creating choice set

Creating the choice set is one of the most important steps in the research process. In this step the route alternatives are created. Discrete choice analysis deals with the problem that there are a finite number of possible alternative routes which can be taken by travellers (Ben-Akiva & Lerman, 1985, p.2). There are several elements that influence route choice. The decision maker has personal characteristics and preferences that influence route choice. The environment dictates the different available alternatives. This is called the universal set of alternatives. The alternatives of these universal set that are considered by the decision maker are called the choice set (Ben-Akiva & Lerman, 1985, p.33 & 34). The choice set are the various alternatives that are considered between the origin and destination (OD).

The choice set

According to Bovy (2009), there are several ways to generate a choice set. These are probabilistic methods, constrained enumeration methods and deterministic and stochastic path search-based methods. In this research the choice is made for stochastic path generation as in this method the properties of the route, such as travel distance and travel time determine the probability of the route being taken (Prato, 2009). If random route alternatives are created this is called unlabelled alternatives. However if specific alternative routes are chosen based on the assumption that cyclists have a variety of motives which determine their route choice, then we speak of labelled alternatives (Prato, 2009). In this research, the three motives that are central to this research are taken into account for the choice set. The shortest route is often part of the choice set (Menghini et al., 2010). In Ehrgott et al. (2012) a fastest route alternative is created. For this research a continuous alternative is created as well. In addition, the two motives which cause cyclists to deviate from the shortest path are taken into account as well. These are a green and attractive environment and a safe route (Kalter et al., 2018; Yeboah & Alvanides, 2015). For safe routes the distinction between traffic safety and social safety is made. The resulting six route alternatives are thus:

- The shortest route
- The fastest route
- The most continuous route
- The greenest route
- The most social safe route
- The most traffic safe route

Generating OD points

As stated before, the choice set consists of the considered alternatives between the origin and destination of the observed routes. OD points are generated from the observed routes as input for the Network Analyst tool in ArcGIS Pro. The OD points are generated using the 'generate points along lines' tool. Once the points were generated they are combined into a single file in which the name of the route is present and the sequence of the corresponding points is stated. Only points at the beginning and end are generated for most routes. However, after the first run of the tool, it was discovered that the alternatives to some of the routes were all shorter compared to the observed routes (see Table 3.3). Additionally, the shortest alternative turned out not to be the shortest alternative. Hence new alternatives were generated with new OD points giving better and more accurate results.

Table 3.3: Average old and new route lengths

Alternative	Old average length	New average length
Observed	3162,88	4285,05
Shortest	2982,83	3956,27
Fastest	2742,96	4037,57
Continuous	X	4081,42
Green	2929,25	4661,31
Social Safe	2790,50	4346,21
Traffic Safe	3076,54	4291,76

The reason that the observed routes were far longer, on average, than the alternatives was due to round trips. OD points were generated for the origin and destination, resulting in alternatives that are completely different from the observed routes as can be seen in the round route in Figure 3.9.

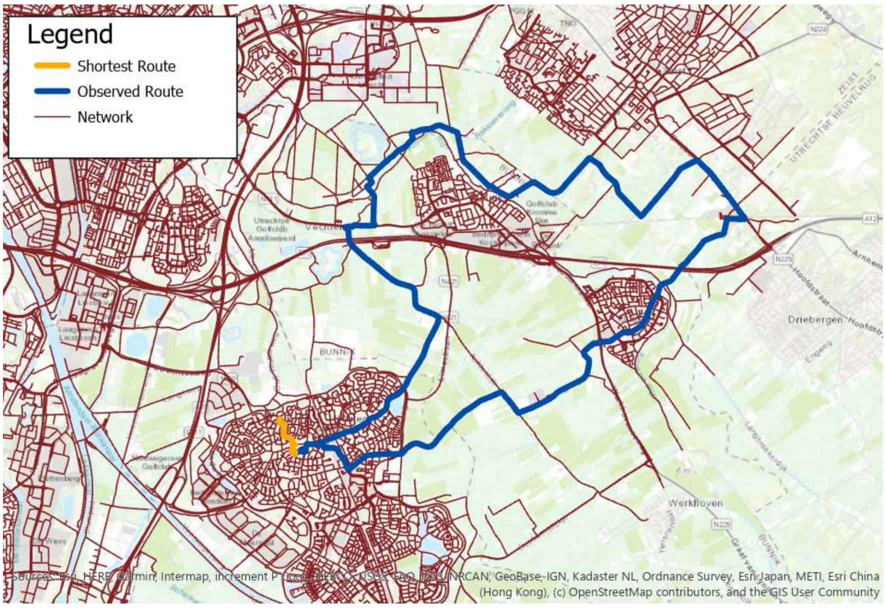


Figure 3.9: The problem of generating alternatives for the round trips

These routes were classified as recreational routes because the origin and destination were the same or because the route between the origin and destination was a very big detour. For the recreational routes, new OD points were generated with additional points which the route alternatives needed to follow. The recreational routes were classified in two groups: round trips and trips that go up and down the same route. The first group was given additional points at 10% intervals along the routes. The alternatives had to reach all these points but could deviate in between. For the second group, one additional point was created at the halfway point of the trip. Between these points the alternatives could still deviate from the observed route.

Route alternatives

As stated before, during the creation of the network file, cost and restriction attributes were added to the network. This was done based on the attributes present in the network after the spatial joins and reclassifications. These attributes are used to generate the six labelled route alternatives. In Table 3.4, an overview of the costs and restrictions that were used for each of the alternatives can be seen.

Table 3.4: Cost and restriction attributes for the alternative routes

Alternative	Cost	Global Turn Delay	Restrictions	Specification
Shortest	Length			
Fastest	Travel time	Straight across no road = 0 Straight across road = 4 Reverse = 10 Right turn = 2 Left turn = 6	Stop signs	Prefer low
			Traffic lights	Prefer low
			Intersections	Prefer low
Continuous	Minimise turns	Straight across no road = 0	Stop signs	Prefer low
			Traffic lights	Prefer low
			Intersections	Prefer low

		Straight across road = 0 Reverse = 30 Right turn = 20 Left turn = 20		
Green	Length		Water	Prefer high
			Environment	Prefer high
			Street lighting	Prefer high
Social safe	Length		Crime rates	Prefer low
			Street lighting	Prefer high
			Environment	Prefer medium
Traffic safe	Length		Road type	Prefer high
			Speed limits	Prefer low
			Street lighting	Prefer high
			Accidents	Prefer low

Each of the labelled alternatives represents a certain common cycling behaviour. The first three are the cycling behaviours central to this research. The other three are other common cycling behaviours according to Yeboah & Alvanides (2015) and Kalter et al. (2018). The goal behind each of these behaviours is explained below.

- *The shortest route*: The goal behind this route is to simulate the behaviour of minimising the travel distance. This is the only thing taken into consideration for the creation of this route. That is why only the length of a route is used as a cost.
- *The fastest route*: The goal behind this alternative is to simulate the behaviour of minimising the travel time. As stated earlier, for cars this is easier to define due to speed limits and traffic models. For bicycles this is not available. This is simulated by taking the travel time as a cost. Additional time penalties are given to all the turns that are taken. Especially turns to the left and crossing other streets results in time penalties as they can delay travel time. Furthermore, restrictions are added to route attributes which negatively affect traffic time such as stop signs, traffic lights and intersections.
- *The most continuous route*: This alternative simulates the behaviour of continuous cycling in which a cyclist attempts to minimise delays and situations where you have to stop. This is done by taking the 'minimise turns cost' attribute which gives time penalties to any turns and the time penalties are higher. Additionally, restrictions are added to attributes which can cause a cyclists to stop such as stop signs, traffic lights and intersections.
- *The greenest route:* This alternative simulates the cycling behaviour where a cyclist prefers to take a more scenic and green route without completely disregarding the length of the trip between the OD points. This is done by taking the length as cost attribute and restriction attributes where the presence of water and street lighting as well as a nice environment along the route are preferred.
- *The most socially safe route:* The goal of this alternative is to simulate the behaviour where cyclists cycle through a safe environment with regards to lighting and crime without disregarding the length of the trip between the OD points. This is done by taking the length as cost attribute and preferring areas with lower crime rates and more street lighting.
- *The most traffic safe route:* The goal of this alternative is to simulate the behaviour where cyclists cycle physically safe routes without disregarding the length of the trip between the OD points. This is done by taking the length as cost attribute and preferring areas with fewer traffic accidents and lower speed limits for cars.

3.4 Data Analysis

The final phase is the data analysis wherein the steps for the statistical analysis of the research are done. This includes preparing indicators, calculating the path size logit and overlap of the routes and finally the conditional logit model.

3.4.1 Preparing data analysis

The final step before the data analysis phase in Stata can begin is the preparation of the data. The observed route shapefiles and the shapefiles from the alternatives are spatially joined with the network on the shared line segments. This way the data from the streets is added to the shapefiles of the routes. However, not all the information in the network is relevant and ready for use. That is why several python scripts are used to condense and transform the data in one excel file that is used as the input in Stata. Because the methodology phase is done in cooperation with other Master students, the script also contains the extraction process of data not used in this research. The script can be found in Appendix 5. The relevant indicators in this research are listed below and can be grouped in the network characteristics, personal characteristics, trip purpose and type of bicycle. The network characteristics are the independent variables in the statistical model. The personal characteristics, trip purpose and type of bicycle are the interaction variables. Below, the indicators are defined and the way in which the data is created is stated.

Independent variables:

- Route length: The length of the routes. The total length is calculated by taking the sum of the segments that make up the route.
- Turn costs: The travel time cost attribute. This variable indicates the relative time cost of turns. A turn to the right is 1. Going straight at an intersection is 2, because other traffic is crossed once. A turn to the left is 3, because other traffic is crossed twice. Joined to the excel file.
- Number of turns: Counts the number of turns to the left and the right on a route. Joined to the excel file.
- Separated cycling path: The percentage of the route that goes over a separated cycling path. If a segment is a separated cycling path, the Boolean is 1 and multiplied with the percentage that segment is of the total route. See 2.
- Cycling highways: The percentage of the route that goes over a cycling highway. If a segment is a cycling highway, the Boolean is 1 and multiplied with the percentage that segment is of the total route. See 2.

$$Total\ percentage = \sum Boolean * Route\ percentage \quad (2)$$

- Traffic lights: The number of traffic lights on the route per km. See 3.
- Stop signs: The number of stop signs on the route per km. See 3.

$$Traffic\ lights = \frac{Number\ of\ traffic\ lights}{Total\ Length} \quad (3)$$

- Path Size Logit: The overlap between the alternatives. See 3.4.2.
- Urban/Rural/Nature: These variables indicate the percentage of the route that goes through an urban, rural or nature environment.

Interaction variables:

- Age: The age of the participant calculated from the birth year stated in the survey.
- Gender: Dichotomous variable indicating the gender of the participant.

- Trip purpose: The trip purpose of the trip is deduced from whether it is a round trip and the location of the OD points. If the trip is considered to be a round trip, the trip purpose is considered to be a recreational trip. If the OD points are located in the PC6 area of their employment the trip purpose is commuting. If the OD points are located within 50 meters from a shop, the trip purpose is utilitarian. If the trip purpose does not meet any of the abovementioned requirements the trip purpose is other. The script can be found in Appendix 7.
- Type of bicycle: The type of bicycle is deduced from the trip purpose and the survey. In the survey participants state which bicycle they use for which purpose. Based on the purpose the bicycle is determined. This can be seen in Appendix 7.

3.4.2 Path Size Logit and overlap

Discrete choice models play an important role in transportation modelling. The goal of the models is to model or represent the complex reality of decision making in transportation based on theory. The most suitable model should carefully be chosen as it needs to fit the application (Bierlaire, 1998). Random utility models are the most commonly used models within transportation research. They simulate the choice of the traveller from among the choice set. However, it is required that the choice set is properly estimated for each individual (Cascetta & Papola, 2001).

When considering only two alternatives for route choice, binary choice models are consulted. These models are often used as they are more simplistic and can thus better illustrate specific situations (Ben-Akiva & Lerman, 1985, p. 59). When more than two alternatives for route choice are considered, multinomial choice models are consulted. These are more complex in nature (Ben-Akiva & Lerman, 1985, p. 100). As stated in earlier, the choice set in this research consists of 6 route alternatives. A property of multinomial choice models is the Independence of Irrelevant Alternatives (IIA), meaning that alternatives are independent from each other (Bierlaire, 1998). For example, in decision making in transportation, taking the train and the bus may seem like different alternatives but actually both represent public transportation. This means that there is overlap between choice alternatives. Standard logit models have difficulty dealing with this. In routes there can be overlap between alternatives in a similar way. When route segments of alternatives overlap for a significant part, this causes the same issue.

Most models are based on the utility theory (4) in which the utility of a route is calculated by a combination of individual characteristics and alternative attributes as well as an error term for unobserved characteristics and attributes (Dane et al., 2019). The Multinomial Logit Model does not account for the aforementioned overlap between the route alternatives. Path Size Logit does take this into account and adds an attribute to account for the overlap between the alternatives. In the second equation (5) you can see the Path Size Logit represented by the PS (Dane et al., 2019).

$$U\_in = V\_in + \in \_in \quad (4)$$

$$U_{in} = V_{in} + \beta_{\rho S} lnPS_{in} + \in_{in} \quad (5)$$

Path Size Logit (6) is calculated by calculating the individual shares of the different links that make up a route. If the links of a route overlap with other route alternatives the share of the entire route that overlaps can be calculated using this equation (6). In Figure 3.10 an example of an observed route can be found. The fastest route has the most overlapping segments with the observed route, followed by the continuous and shortest. Furthermore, the traffic safe has a lot of overlap with the green route. The social safe route completely overlaps with the green route. If the outcome of (6) is 1 then the route does not overlap with other alternatives (Dane et al, 2019). Besides the path size logit, the overlap is also calculated in a percentage for each of the routes. This can be seen in Appendix 6. The goal of this overlap is to give insight into the percentual overlap between the observed route and the alternatives. If an alternative overlaps 100% with another alternative or with the observed route they are combined. The reason for this is that if the overlap is 100%, the alternatives are the same and cannot be counted twice in the model. Once the overlapping routes are removed, the Path Size Logit is calculated as can also be seen in Appendix 6.

$$PS_{in} = \sum_{a \epsilon \Gamma} \frac{\mathrm{L_a}}{\mathrm{L_i}} \frac{1}{\Sigma_{j \epsilon C_n} \delta_{aj}} \quad (6)$$

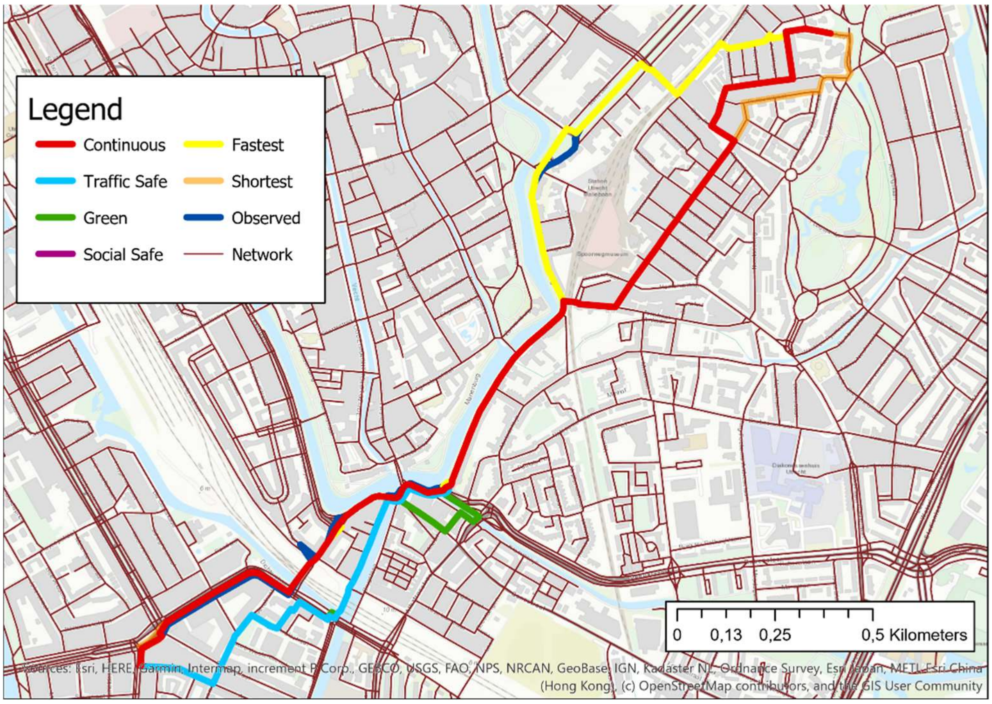


Figure 3.10 Example of Observed route with partially overlapping alternatives

3.4.3 Conditional logit model

In Stata a conditional logit model is used to determine the influence of the independent and interaction variables on the dependent variables. The influence of the selected spatial route characteristics and the interaction with personal characteristics on the route choice is analysed. Variables are taken into account in the model if they do not correlate more than sixty percent. This is why a correlation model is carried out first. The dependent variables are the different route alternatives. An overview can be seen in Figure 3.11. This results in a table which indicates to what extent and in what way the variables influence the route choice. In Stata a mixed logit model is used as a conditional logit model. Random variables are left out The choice was made to use the cmxtmixlogit model. This model is used because participants in this research can cycle multiple routes. This means that the participants make repeated decisions on route choice at different moments in time (Stata, n.d., p. 242). Furthermore, it uses random coefficients to model the correlation of choices across alternatives. These are on alternative-specific variables (Stata, n.d., p. 242). This correlation lessens the IIA assumption that was discussed in 3.4.2. This is because this model models the probability that an alternative is selected at each moment in time rather than calculating the probability of a route alternative being chosen in general.

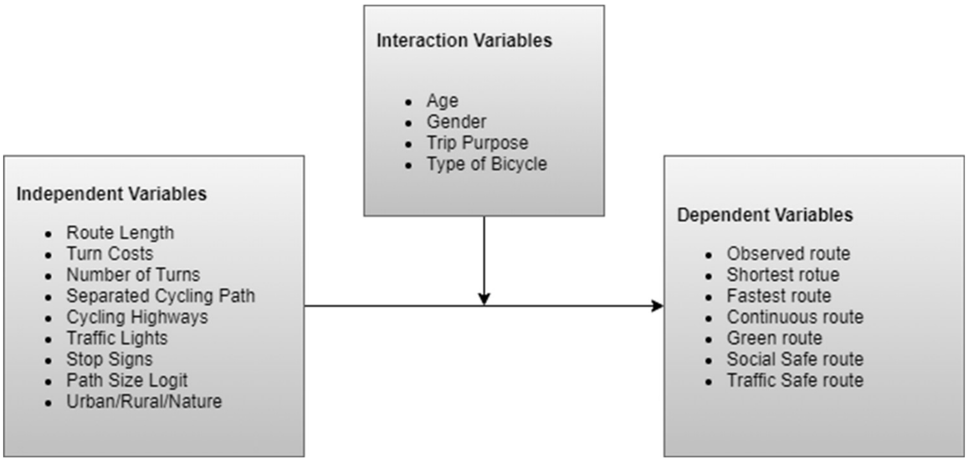


Figure 3.11: Independent variables and interaction variables influence dependent variables

3.4.4 Determining preferred cycling behaviour

The eventual goal of this research is to determine the extent to which cyclists minimise travel distance, minimise travel time and cycle continuously. This was originally measured by creating labelled route alternatives of each of these three cycling behaviours and looking at the percentual overlap between the observed route and these labelled alternatives. However the problem with this methodology is that although the motive of a cyclist can be to cycle continuously, they do not take the most continuous route possible, resulting in little overlap with the most continuous alternative while the route is still continuous. This is why the extent to which cyclists minimise travel distance, minimise travel time and cycle continuously is not measured by the percentual overlap with the shortest, fastest and most continuous alternative. Instead, variables that indicate these behaviours are used to measure the behaviours (Table 3.5). This gives a more accurate representation of whether the cycling behaviours central to this researched are pursued by the participants. This can be seen in Appendix 8. In the next chapter the results from the survey and the statistical analysis are discussed.

Table 3.5: Relevant variables for the cycling behaviours

Cycling Behaviour	Relevant variables
Minimising travel distance	Length
Minimising travel time	Length, turn costs, separated cycling paths, stop signs and traffic lights,
Continuous cycling	Number of turns, cycling highways, stop signs and traffic lights according to the conceptual model

4. Results

This chapter discusses the results of this research in order to answer the main research question: *"To what extent do cyclists minimise travel distance, minimise travel time and cycle continuously and which characteristics influence these behaviours?"* and the sub questions. First, it is established whether the data is representative and the descriptive statistics are given. Subsequently, the results of the survey are discussed. Finally, the results from the statistical analysis are presented.

4.1 Descriptive Statistics

In this part the representativity of the data is discussed. This is done by looking at the personal characteristics of the research population. Additionally, the descriptive statistics of the spatial characteristics are given. This gives an insight into the overall results from the research and how the results can be interpreted.

4.1.1 Personal characteristics

The representativity of the research indicates to what extent the upcoming results are representative for the general population. The representativity is evaluated for the personal characteristics. Gender is further divided into the gender division for the participants and gender division for cycled trips. Age is categorised into four groups in this table. An overview is given in Table 4.1. The first indicator is gender. The gender ratio of the participants is skewed with around two thirds of the participants being female and only a third male. Moreover, this difference becomes even starker when looking at the percentage of routes cycled per gender. Women are responsible for 74.3% of the total routes and men only for 25.7%. The second indicator is age. The age of the participants ranges between 16 and 73. The mean age of the participants is 35.7 years old. Table 4.1, shows that more than half of the participants fall in the youngest two age categories. Only 5% of participants is older than 65. The third indicator is education level. This shows that the research population is predominantly very highly educated at HBO and university level. Only 3.3% falls in the lowest category. The fourth indicator is trip purpose. 41% of the trips that are made by the participants are for a utilitarian purpose. This is followed by 9.3% of trips being commutes and only 6,7% are recreational. The largest group (43%) falls into the 'other' category. The fifth indicator is bicycle type. The most commonly used bicycle for trip is a citybike without gears (36.6%). This is followed by citybikes with gears (35.1%). 10.3% of the trips are made with sports (hybrid) bikes. Only 3.8% of trips are made using E-bikes and less than 1% of trips are made on racing bikes/mountain bikes. For 13,3% of the trips it is unknown which bicycle is used. Overall, the research population is thus predominantly female, young, highly educated, cycling for utilitarian or other purposes and uses citybikes with and without gears.

Table 4.1: Personal Characteristics

Variables	Category	Value	Percentage
Gender ratio (participants)	Female	*	65,6
	Male		34,4
Gender ratio (cycled routes)	Female	0	74,3
	Male	1	25,7
Age	<25	**	38,3
	25 - 44		30,0
	45 - 65		26,7
	>65		5,0
Education	Lbo, mulo, mavo, vmbo or equal	***	3,3
	Havo, vwo, mms, hbs, mbo or equal		20,0
	Hbo or universiteit		76,7
	Other		0
Trip Purpose	Commute	1	9,3
	Recreational	2	6,7
	Utilitarian	3	41,0
	Other	4	43,0
Bicycle type	Electrical Bike	1	3,8
	City bike (no gears)	2	35,1
	City bike (gears)	3	36,6
	Sports (hybrid)	4	10,3
	Racing bike/mountainbike	5	0,9
	Unknown	6	13,3

* Gender ratio in Stata is per cycled routes not participants. Hence, no Value.
** Age is used as a continuous variable in Stata, not as a categories. Hence, no Value.
*** Education is not used as variable in Stata as there was no evidence that education level influences cycling behaviour.
All these variables are in the table in order to determine representativity.

4.1.2 Spatial characteristics

The descriptive statistics of the independent variables give insight into the means and ranges of the variables in Table 4.2. This is done for the observed routes. However, the mean is given of the route alternatives. This shows that on average, the observed routes are shorter than their alternatives. In addition, the range of the routes is very broad, as the shortest route is 230 meters and the longest 89 kilometres. The turn costs of the observed routes, on average, are lower than that of the alternatives. Similarly, the number of turns per route are lower for the observed routes as well, compared to the alternatives, with an average of 14.8 turns. On average, over half of the cycling routes consist of separated cycling paths. Cycling highways are far less common. On average only 0.23% of the route is a cycling highway. Stop signs are also not very common with a maximum of 3 per route and on average only 0.04 per route. Traffic lights on the other hand are more common with an average of 1.95 per route and a max of 19 per route. Most of the routes go through an urban environment. On average 66% of a route goes through an urban environment. This is much more than the 2% for nature and 9% for rural. The alternatives go through rural areas more often, and less through urban areas. The path size ranges from 0.21, which entails complete overlap between the alternatives, and 1 which is no overlap between the alternatives. The average path size is 0.59 which means that there is some overlap between the alternatives but that the alternatives are mostly different from each other with little, but some overlap. Overall, the routes range from short (230 meters) to very long (89 kilometres). On average, the observed routes have less turn costs, fewer turns and encounter fewer stop signs and traffic lights. Most of the routes are in an urban environment.

Table 4.2: Descriptive statistics spatial characteristics

Variables	**Mean Observed**	**Mean Alternatives**	**Min Observed**	**Max Observed**
Length	4.29	5.13	0.23	89.29
Turn Costs	75.1	78.9	0	381
Number of Turns	14.8	15.1	0	93
Separated Cycling Path	56.53	56.23	0	100
Cycling Highway	0.23	0.06	0	18.64
Stop Signs	0.04	0.05	0	3
Traffic Lights	1.95	2.58	0	19
Urban	66.06	63.41	0	100
Nature	2.13	1.06	0	100
Rural	8.94	10.05	0	100
Path Size	0.59	0.47	0.21	1

4.2 Survey Results

Despite being a revealed preference study, this research did ask some questions about cycling preferences with regards to the cycling behaviours central to this research. This was done in the form of statements. The goal of these statements is to answer the sub question: "To what extent do cyclists state to minimise travel distance, travel time and cycle continuously?"

4.2.1 Minimising travel distance

The first statement is in regards to minimising travel distance: "When I cycle I attempt to take the shortest route." Half of the respondents agreed with this statement and twenty percent even strongly agreed (Figure 4.1a). This indicates that most participants state to have a preference for minimising travel distance when cycling. When looking at the difference between male and female, the data shows that women agree more strongly with the statement than men. Additionally, men disagree more often with the statement. When it comes to age, participants between 25 and 44 years old agree the most with this statement. The older age categories see more relative disagreement with the statement. Overall, it can thus be said that the younger age categories state to minimise travel distance more than older participants. This can be seen in Figure 4.1b and c and Table 4.3.

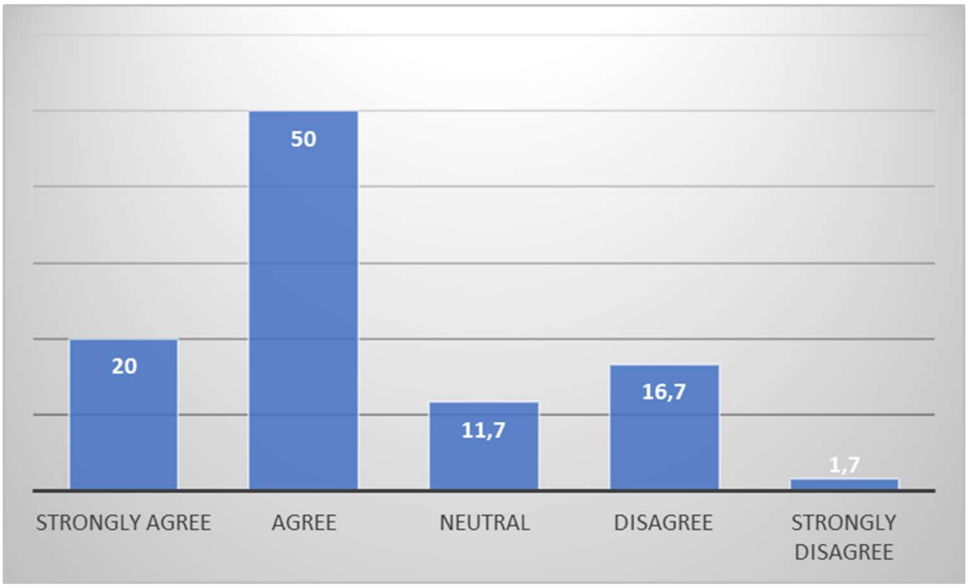
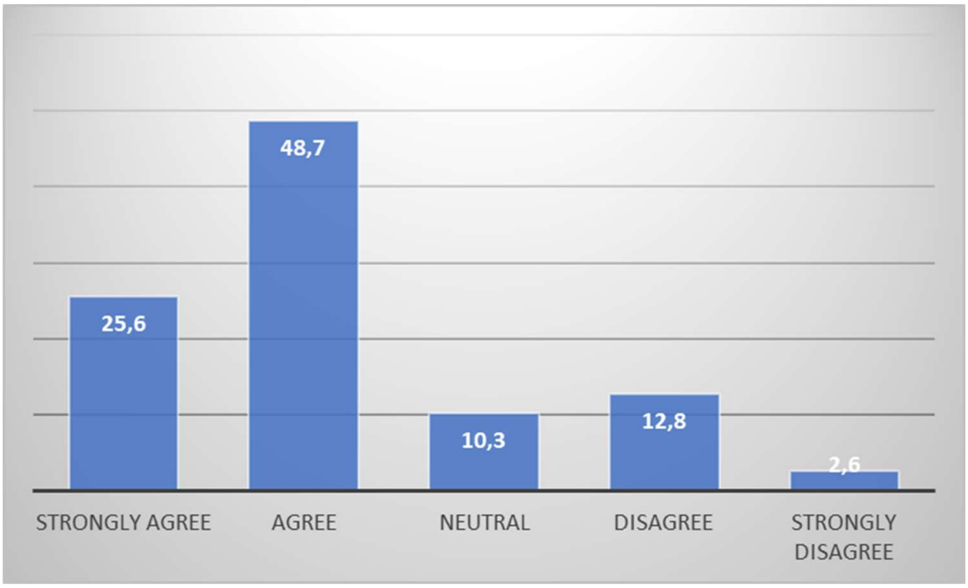
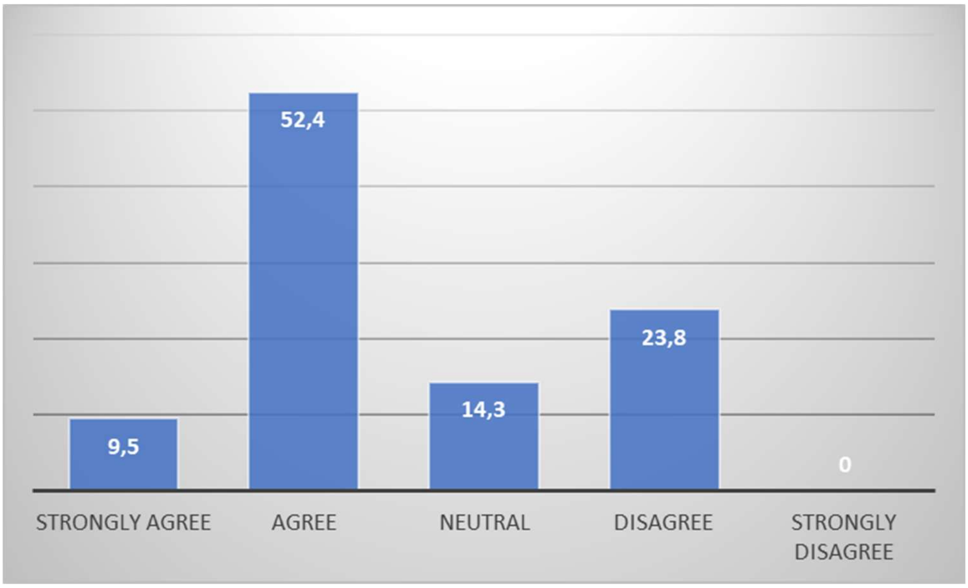


Figure 4.1: When I cycle I attempt to take the shortest route. (in %)
(a) All participants



(b) Female participants



(c) Male participants

Table 4.3 When I cycle I attempt to take the shortest route. (in %) per age category

	Strongly Agree	Agree	Neural	Disagree	Strongly Disagree
<25	30.4	43.5	13	13	0
25 – 44	16.7	61.1	16.7	5.6	0
45 – 65	12.5	50	6.3	31.3	0
>65	0	33.3	0	33.3	33.3

4.2.2 Minimising travel time

The second statement is about minimising travel time: "When I cycle I attempt to minimise travel time". Figure 4.2 shows very similar results to the previous statement. Slightly fewer people strongly agree and slightly more agree with this statement. However, seventy percent agree to some extent. It can thus be said that the participants state that they generally aim to minimise travel time when cycling. Similarly to the previous statement, a higher percentage of women agrees with the statement and more men disagree with the statement. There are also differences in age (table 4.4). The two youngest categories with people up to 44 years old strongly agree with this statement, especially the group younger than 25. People between 45 and 65 still mostly agree, but to a lesser extent. People older than 65 all disagree with this statement to some extent. This shows that both men and women state that they attempt to minimise travel time. Although women state to attempt this to a slightly bigger extent than men. Furthermore, younger people state that they aim to minimise travel time more than older people.

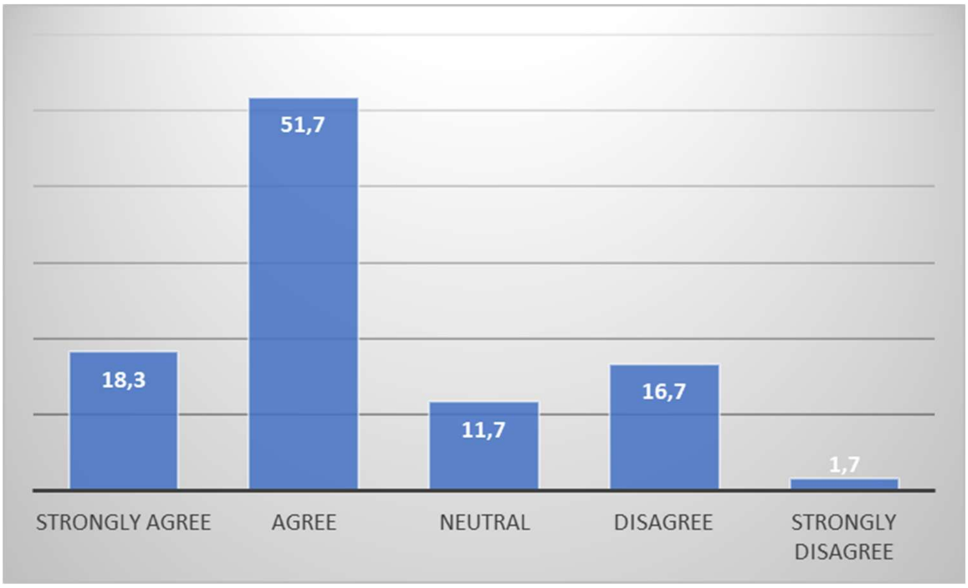
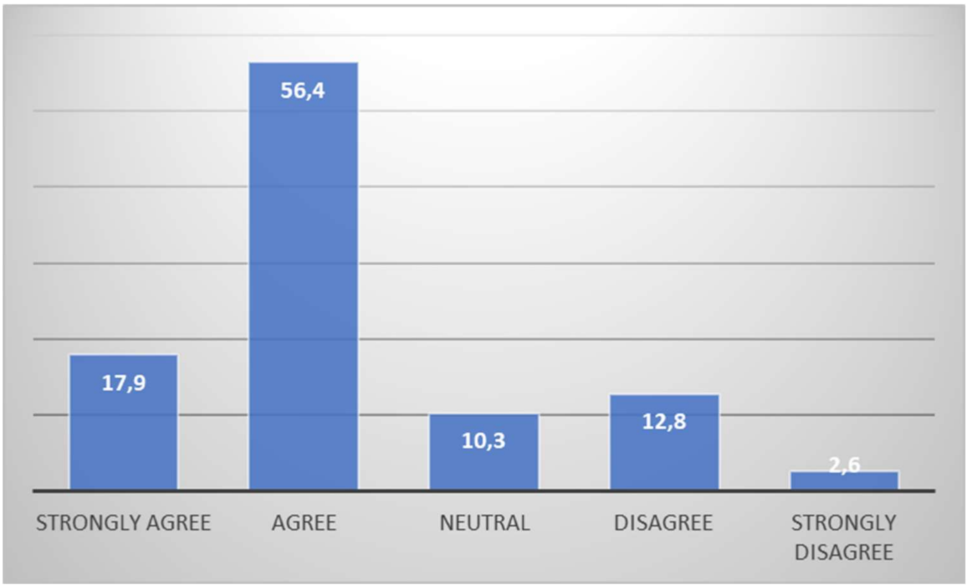
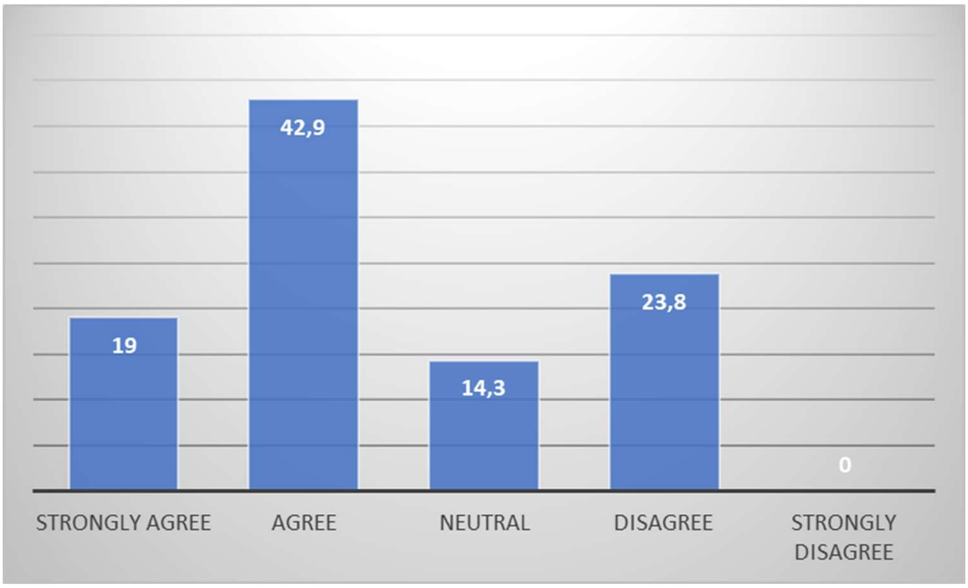


Figure 4.2: When I cycle I attempt to minimise travel time. (in %)
(a) All participants



(b) female participants



(c) male participants

Table 4.4: : When I cycle I attempt to minimise travel time. (in %) per age category

	Strongly Agree	Agree	Neural	Disagree	Strongly Disagree
<25	30.4	56.5	4.3	8.7	0
25 – 44	16.7	61.1	11.1	11.1	0
45 – 65	6.3	43.8	25	25	0
>65	0	0	0	66.7	33.3

4.2.3 Continuous cycling

The third and final statement is about continuous cycling: "When I cycle I attempt to minimise stopping. For example for traffic lights." Figure 4.3 shows that half of the people agree or strongly agree. However, thirty percent of participants are neutral toward this statement. Twenty percent disagree, which is more than with the other statements. Overall, it can still be said that most people state they attempt to cycle continuously. However, not to the same extent as with minimising travel distance and travel time. In contrast to the previous two statements, men agree more with this statement and also agree more strongly. More women disagree with the statement. Around a third of both women and men are neutral towards this statement. When looking at the age of the participants it shows that, similarly to the previous statement, younger people more strongly agree with this statement (Table 4.5). The older categories also disagree more. The 65+ category disagrees more with this statement, although they are divided.

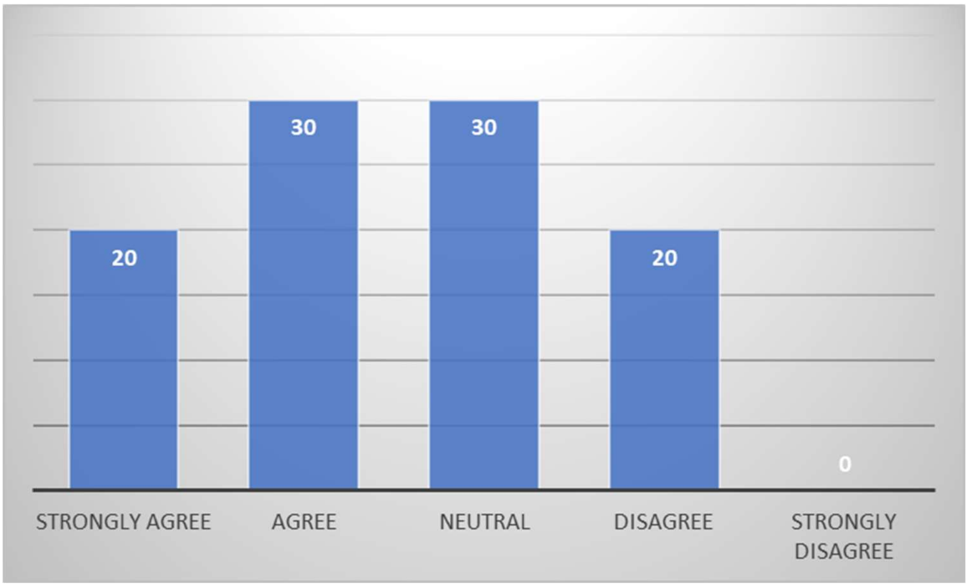
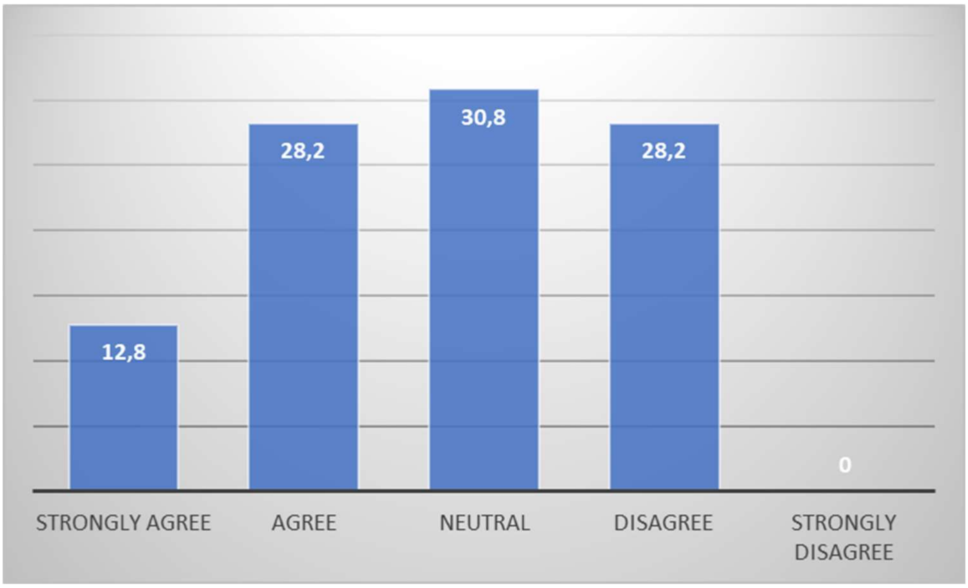
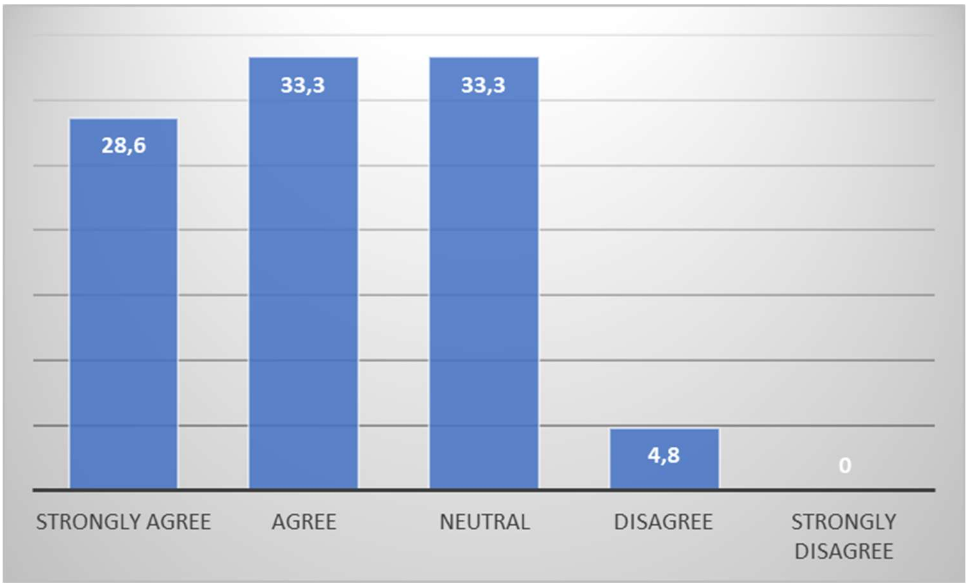


Figure 4.3: When I cycle I attempt to minimise stopping. For example for traffic lights (in %)
(a) All participants



(b) *Female participants*



(c) male participants

Table 4.5: When I cycle I attempt to minimise stopping. For example for traffic lights (in %) per age

	Strongly Agree	Agree	Neural	Disagree	Strongly Disagree
<25	26.1	30.4	30.4	13	0
25 – 44	22.2	27.8	33.3	16.7	0
45 – 65	6.3	31.3	37.5	25	0
>65	0	33.3	0	66.7	0

4.3 Results Statistical Analysis

In this section, the results from the statistical analysis are presented. This is done by estimating a choice model for all the relevant independent variables first. Then, the models for each of the three cycling behaviours central to this research are discussed. Lastly, the influence from the interaction variables that are relevant for this research are discussed.

4.3.1 Correlation

Table 4.6 shows the correlations between the variables. It shows that there are some correlations between the variables, so most are independent from each other. However, there are a few variables that show higher correlation with one another. Various variables have a correlation of higher than 0.3 with the variable ‘length’. These are ‘turn costs’, ‘number of turns’ and ‘stop signs’. This can be explained by the fact that there are naturally more turns, and thus turn costs, when cycling a longer route. Likewise, there are also higher chances of encountering stop signs with longer routes. ‘Cycling highways’ also correlate with ‘length’. Indicating that the percentage of cycling highways is higher on longer routes. ‘Urban’ negatively correlates with ‘length’, indicating that urban routes are shorter. Consequently, ‘rural’ routes are often longer.

Furthermore, the ‘turn cost’ variable and the ‘number of turns’ variable do show a high correlation. This is due to fact that they are both measures for turns. However, because each of these is relevant to one of the three cycling behaviours that is central to this research, they are both be used in the choice model. ‘Traffic lights’ also correlates with ‘turn costs’ Routes with more traffic lights also encounter higher turn costs. Both variables have in common that they negatively affect travel time.

Furthermore, 'traffic lights' negatively correlates with 'separated cycling paths'. This indicates these characteristics are less likely to be found on similar routes. 'Urban' positively correlates with 'separated cycling paths', indicating that these are often found in urban environments. Lastly, 'urban' and 'rural' show a high correlation around the 0.6 threshold. Because these variables both represent the opposite, only urban is used in the choice model.

Table 4.6: Correlation table independent variables

	Length	lnPath Size	Turn Costs	Numb-er of Turns	Separat-ed Cycling Path	Cycling Highwa-ys	Stop Signs	Traffic Lights	Urban	Rural	Nature
Length	1										
lnPathSize	-0.0074	1									
Turn Costs	0.5406	0.0317	1								
Number of Turns	0.5603	-0.0076	0.8909	1							
Separated Cycling Path	-0.0957	-0.0362	-0.0732	0.0156	1						
Cycling Highways	0.3105	0.0125	0.2365	0.2761	-0.0377	1					
Stop Signs	0.4383	0.0235	0.14	0.2134	-0.0203	0.2394	1				
Traffic Lights	0.239	0.0546	0.3727	0.2175	-0.3	0.0512	0.0695	1			
Urban	-0.3474	0.0235	-0.108	-0.0396	0.4055	-0.0795	0.0082	-0.2259	1		
Rural	0.3276	-0.1313	0.0433	0.113	0.0878	0.131	0.0274	-0.1684	-0.5824	1	
Nature	0.2206	0.0423	0.0868	0.0881	-0.0093	0.0205	0.037	-0.0722	-0.1974	0.0142	1

4.3.2 Choice Model

In Table 4.7, the choice model can be found. The model is significant at the 1% level. The Wald indicates that the overall model has a good fit. When looking at the input variables, it can be seen that most are significant at the 1% level as well. However, there are a few exceptions. Stop signs and traffic lights are not significant and the variables 'cycling highways' and 'urban' are significant at a 5% level. Length has an odds ratio of 0.49. This means that for every kilometre that the route length increases, the chance that a route is chosen decreases with 51%.

The path size has an odds ratio of 3.54. The path size needs to be positive to account for overlap between alternatives. The value indicates that there is a correlation between overlapping alternatives (Dane et al., 2019). When the path size increases, the overlap decreases. If the odds ratio is higher, then there is more reason to choose the observed route because the alternatives are different. When there is more overlap, the alternatives do not differ as much from the observed route. This means that there would be less reason to choose that route.

'Turn costs' has an odds ratio of 1.05. This means that if the turn costs of a route increase with one turn cost unit, the chances of the route being chosen increases with 5%. Similarly, the 'number of turns' variable has an odds ratio of 1.095. Each additional turn increases the likelihood that an alternative is chosen by 9.5%. For 'separated cycling paths' the odds ratio is 0.971. This means that for every percentage that the share of separated cycling paths on the route increases, the likelihood of that route being chosen decreases by 2.9%. For cycling highways the odds ratio is 3.02. Every percentage that the share of cycling highways on the route increases, increases the chance of the route being chosen by. 202%. 'Urban' has a odds ratio of 0.987. For every percentage that the urban environment of a route increases, the likelihood of the route being chosen decreases 1.3%. In contrast, the odds ratio of nature is 1.14, meaning that an increase in percentage of nature increases the likelihood of the route being chosen by 14%.

Overall this means that shorter routes, routes with higher turn costs and more turns, routes with more cycling highways and routes with more nature are more likely to be taken by cyclists. Consequently, routes with longer lengths, a higher percentage of separated cycling paths and routes with a higher percentage of urban environment are less likely to be taken by cyclists.

Table 4.7: Choice model all relevant variables

Variable:	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	Conf.
Length	0.4911488	0.0628408	-5.56	0.000	0.3822125	0.631134
lnPathSize	3.541484	0.74665	6.00	0.000	2.342754	5.353575
Turn Costs	1.050361	0.0075611	6.83	0.000	1.035646	1.065286
Number of Turns	1.095202	0.0191618	5.20	0.000	1.058282	1.13341
Separated Cycling Path	0.9713511	0.0064493	-4.38	0.000	0.9587925	0.984074
Cycling Highways	3.02249	1.342686	2.49	0.013	1.265418	7.219309
Stop Signs	1.764736	0.9127883	1.10	0.272	0.6403334	4.86355
Traffic Lights	1.077481	0.0575429	1.40	0.162	0.9704006	1.196377
Urban	0.9870666	0.0053088	-2.42	0.016	0.9767162	0.997527
Nature	1.141429	0.0379767	3.98	0.000	1.069371	1.218342

Wald Chi2(10) = 195.98
Prob > chi2 = 0.0000
Log likelihood = -490.69125
No. persons = 59
No. trips = 405
n = 2160

4.3.2 Cycling behaviours

This research is centred around three types of cycling behaviour. These are minimising travel distance, minimising travel time, and continuous cycling. The significance of the variables related to these cycling behaviours is discussed in this part.

Minimising travel distance

The cycling behaviour of minimising travel distance is defined in the conceptual model by the route length. This is the only variable that is relevant to this cycling behaviour. Based on the choice model in Table 4.7, the length is negative and significant at the 1% level, meaning that shorter routes are more likely to be taken. Travel distance minimisation is thus observed. However, the singular effect of length as a variable proved to be insignificant with a p value of 0.43. In addition to the strong effect seen in the choice model, the average overlap of the observed routes with the shortest alternative is the highest out of all the alternatives (see table 4.8). This is another indication of travel distance minimisation as a cycling behaviour.

Table 4.8: Average overlap of observed route with route alternatives

Shortest	Fastest	Continuous	Green	Social Safe	Traffic Safe
53.6%	49.1%	48.0%	46.0%	49.3%	51.3%

Minimising travel time

In the conceptual model it is assumed that the variables that determine travel time are: length, turn costs, separated cycling paths, cycling highways, stop signs, and traffic lights. In the choice model length, turn costs, separated cycling paths and cycling highways are significant. Several variables indicate that travel time minimisation takes place. A shorter route increases the chance of a route being taken. Furthermore, cycling highways are also preferred by cyclists, further indicating some travel time minimisation. In contrast, higher turn costs increase the chances of a route being taken. Separate cycling paths decrease the chance of a route being taken. These last two variables indicate that travel time is not minimised. Stop signs and traffic lights are not significant even when looking at the singular effect of these variables on route choice. The impact of the other variables does not change either whether they are run separately or not. Additionally, a model was run with only the relevant variables for minimising travel time. This did not change any of the outcomes with regards to significance or impact. The average overlap of the observed routes with the fastest alternatives is 49.1%. This alternative performs around the average compared to the overlap of the other alternatives.

Continuous cycling

Continuous cycling is influenced by the number of turns, separated cycling paths, cycling highways, stop signs, and traffic lights according to the conceptual model. According to the choice model, the number of turns, separated cycling paths and cycling highways turned out to be significant. Counterintuitively, the results indicate that a higher number of turns increases the likelihood of a route being chosen. Furthermore, separate cycling paths decrease the chances of a route being chosen. Although these results indicate that continuous cycling is not present, cycling highways are preferred. This, in contrast, does indicate some preference for continuous cycling. Additionally, rural areas increase the chances of a route being chosen. This could also indicate continuous cycling because there is less traffic allowing for a better cycling flow. As stated earlier, stop signs and traffic lights were not significant. Running the variables individually did not change impact or significance. However, running a separate model with only the relevant variables with regards to continuous cycling does change some outcomes. Firstly, the variable 'cycling highways' is no longer significant. However, traffic lights are significant now. The impact is positive, meaning that routes with more traffic lights have a higher chance of being chosen. Because the presence of traffic lights can, under certain circumstances, also positively influence continuous cycling, this does not eliminate the possibility that continuous cycling is a motive of the research population, although it is unlikely. Lastly, the average overlap of the observed routes with the continuous alternatives is 48.0%. This is the second to last lowest overlap with the observed routes, only above the green alternative.

4.3.3 Interaction variables

In this last part, the influence from personal characteristics is modelled by interaction variables. The interaction variables are gender, age, trip purpose, and bicycle type.

Gender

The first interaction variable is gender. An overview of the influence per gender can be seen in table 4.9. The influence of length is not significant for either gender. Turn costs are relevant for both genders. For both genders the effect is still positive. Although the turn costs are minimised slightly more by women than men. The same can be said for the number of turns. The 'separated cycling paths' variable is only significant for women. However, the odds ratio is only 0.01 higher than in the overall model. Similarly, cycling highways is also only significant for women and the odds ratio is lower than in the general model. Stop signs is not significant for either gender. Interestingly, traffic lights is only significant for men at a 10% level. This was not the case in the general model. The odds ratio is 0.84, meaning that the lower the number of traffic lights along a route, the bigger the chance that the route is taken by men.

Table 4.9: Odds ratio and significance of independent variables per gender

	Length	Turn costs	Number of turns	Separated cycling path	Cycling highway	Stop signs	Traffic lights
Female	0.96 (p = 0.73)	1.04 (p = 0.00)	1.11 (p = 0.00)	0.98 (p = 0.00)	1.31 (p = 0.03)	1.42 (p = 0.39)	1.01 (p = 0.68)
Male	1.08 (p = 0.26)	1.07 (p = 0.00)	1.15 (p = 0.00)	0.99 (p = 0.19)	1.79 (p = 0.25)	0.00 (p = 0.99)	0.84 (p = 0.07)

Age

The second interaction variable is age. An overview of the influence for age can be seen in Table 4.10. Length is significant at a 10% level and has an odds value of 1.00, indicating that there is no difference in age when it comes to choosing routes with a higher length. Cycling highways are significant at a 10% level as well. The odds value is 1.01. Indicating that older people are slightly more likely to take a cycling highway. Turn costs, number of turns, and separated cycling path are all significant at a 1% level. The odds value for all these is 1.00, indicating that there is no difference between old and young people when it comes to these variables. Lastly, stop signs and traffic lights were not significant.

Table 4.10: Odds ratio and significance of independent variables for age

	Length	Turn costs	Number of turns	Separated cycling path	Cycling highway	Stop signs	Traffic lights
Age	1.00 (p = 0.06)	1.00 (p = 0.00)	1.00 (p = 0.00)	1.00 (p = 0.02)	1.01 (p = 0.07)	1.00 (p = 0.85)	1.00 (p = 0.54)

Trip Purpose

The third interaction variable is trip purpose. An overview is given in Table 4.11. The influence of length is not significant for any of the trip purposes. Only the utilitarian purpose is significant but only at a 10% level. Here it shows the opposite effect of the length in the general model. This means that for utilitarian trips the longer the route, the higher the chance the route is chosen. For turn costs all but recreational trips are significant. The effects remain the same compared to the general model. Commutes used routes with higher turn costs the most, followed by utilitarian trips. The same is visible for the number of turns. The variable 'separated cycling paths' is only significant for utilitarian and other trips. However, the influence on route choice is the same compared to the general model. For cycling highway and stop signs, none of the trip purposes are significant. Lastly, for traffic lights, all cycling purposes except for recreational is significant. In the general model, traffic lights is not significant. The effect for all trip purposes is positive, meaning that the more traffic lights, the higher the chance a route gets chosen. This is especially the case for commutes, followed by utilitarian trips, and lastly other trips.

Table 4.11: Odds ratio and significance of independent variables per trip purpose

	Length	Turn costs	Number of turns	Separated cycling path	Cycling highway	Stop signs	Traffic lights
Commute	1.13 (p = 0.64)	1.12 (p = 0.00)	1.34 (p = 0.00)	0.99 (p = 0.61)	1.55 (p = 0.54)	3686172 (p = 0.99)	1.34 (p = 0.00)
Recreational	1.08 (p = 0.40)	1.01 (p = 0.35)	1.01 (p = 0.70)	1.00 (p = 0.91)	1.69 (p = 0.38)	0.00 (p = 0.99)	1.01 (p = 0.70)
Utilitarian	1.33 (p = 0.05)	1.07 (p = 0.00)	1.16 (p = 0.00)	0.98 (p = 0.02)	(omitted)	1.08 (p = 0.91)	1.16 (p = 0.00)

Other	0.92 (p = 0.38)	1.03 (p = 0.00)	1.09 (p = 0.00)	0.98 (p = 0.01)	1.31 (p = 0.90)	1.39 (p = 0.53)	1.09 (p = 0.00)

Bicycle Type
The fourth and final interaction variable is bicycle type. An overview is given in Table 4.12. For the variable length, only city bikes are significant. Similarly to the general model, the effect is negative. For the turn costs, all bicycle types are significant except racing bike/mountainbike. Electrical bike is only significant at 10% level. The effect here is the same as in the general model, but the strongest with E-bikes followed by city bikes with no gears. The pattern for the number of turns is the same as with the turn costs. The difference is that the difference in odds ratio between the bicycle types is bigger. On top of that, the highest odds ratio is that of city bikes with no gears, followed by the 'unknown' category. For the variable 'separated cycling path', only city bikes with gears and without gears are significant. City bikes with gears is only significant at a 10% level. The effect is similar to that in the general model. For cycling highways only the unknown bicycle type is significant and is positive similar to the general model. None of the bicycle types are significant for stop signs. The same is true for traffic lights, except for city bikes without gears. However, this is only significant at a 10% level. The effect is opposite to that in the general model, however. In other words, the higher the number of traffic lights, the lower the chance someone with a citybike without gears chooses the route.

Table 4.12: Odds ratio and significance of independent variables per bicycle type

	Length	**Turn costs**	**Number of turns**	**Separated cycling path**	**Cycling highway**	**Stop signs**	**Traffic lights**
E-bike	1.21 (p = 0.23)	1.08 (p = 0.05)	1.11 (p = 0.07)	1.00 (p = 0.88)	(omitted)	(omitted)	1.09 (p = 0.91)
City bike (no gears)	1.16 (p = 0.43)	1.06 (p = 0.00)	1.24 (p = 0.00)	0.97 (p = 0.00)	(omitted)	0.00 (p = 0.99)	0.85 (p = 0.05)
City bike (gears)	0.69 (p = 0.04)	1.04 (p = 0.00)	1.09 (p = 0.00)	0.99 (p = 0.05)	0.00 (p = 0.99)	1.70 (p = 0.22)	1.03 (p = 0.34)
Hybrid	1.08 (p = 0.30)	1.03 (p = 0.00)	1.06 (p = 0.03)	1.00 (p = 0.87)	1.79 (p = 0.24)	0.00 (p = 1.00)	1.11 (p = 0.39)
Racing bike/ Mountain bike	2.16 (p = 0.74)	1.11 (p = 0.19)	1.23 (p = 0.25)	0.77 (p = 0.30)	(omitted)	(omitted)	176e+07 (p = 0.99)
Unknown	1.05 (p = 0.92)	1.06 (p = 0.00)	1.13 (p = 0.00)	1.00 (p = 0.94)	1.31 (p = 0.04)	(omitted)	0.79 (p = 0.07)

The remaining chapters are the discussion and the conclusion. Here, the results are interpreted and put into context of previous research. The main conclusions are drawn from the results as well in order to answer the main research question and the sub questions.

5. Discussion

In the discussion the outcomes from this research are interpreted and compared to the results from other research. Furthermore, the issues that were encountered during the process with regard to data quality, research methods and Covid-19 will be reflected upon.

5.1 Research Outcomes

This section discusses the outcomes from the research. Firstly, the representativity is reflected upon. Then the results from the survey and the choice model are discussed. Lastly, the influences from the interaction variables are compared to those of previous research.

5.1.1 Representativity of the results

Chapter 4.1 discussed the descriptive statistics. This gave insight into the representativity of the results. Based on the outcomes from the descriptive statistics, the research population is not representative for the Dutch population. The research population is skewed female, young and highly educated. This is the result from using friends and family as participants in the research. Furthermore, the descriptive statistics show that most trips were for a utilitarian purpose. Although no data is available on the usual percentage per trip purpose, it should be noted that due to the lockdown during the Covid-19 pandemic people often worked from home and cycled less for commuting purposes. Lastly, two thirds of the routes went through an urban environment. This should be kept in mind when comparing the results to other studies, which mainly took place in a fully urban environment (e.g. Hood et al., 2011; Ton et al., 2017).

5.1.2 Interpretation of the results

The results from the choice model show that cyclists do attempt to minimise travel distance when cycling. Additionally, participants also state to attempt to minimise travel distance. This matches with the results from various studies that also found route length to be an important deciding factor in route choice (Broach et al., 2012; Menghini et al., 2010; Ton et al., 2017). The result is further confirmed by the fact that the shortest route alternatives had the highest overlap with the observed routes. However, the overlap with the shortest alternative was 53% on average. This indicates that, although people do minimise travel time, they do not often take the shortest route. This could be due to limited knowledge about what the shortest route is, or because other cycling motives are given priority as well.

When it comes to minimising travel time the results are less conclusive. Although cyclists do minimise travel distance and prefer routes with cycling highways, other factors that affect travel time show less promising results. For example, no preference for separated cycling paths was observed, similarly to the research by Ton et al. (2017). Although positive effects from separated cycling paths are mostly observed in research abroad (Yeboah & Alvanides, 2015). Ton et al. (2017) states this difference is due to the safety of Dutch infrastructure. The fact that this variable's effect is negative may thus be due to the Dutch context. Furthermore, the choice model indicates that turn costs are not avoided, indicating that this aspect of travel time is not minimised. This is in contrast to Broach et al. (2012), who states that turns and especially left turns negatively affect travel time. The effect of stop signs and traffic lights on route choice was not significant. This may be because these effects are already cooperated in the other variables or because these variables had both positive and negative effects. Because length also indicates travel distance minimisation and because most variables related to minimising travel time show the opposite effect of what is expected, it seems unlikely that cyclists prioritise travel time minimisation. In contrast, the survey results clearly shows that most participants generally state attempt to minimise travel time. This shows the difference between stated preference and revealed preference by Casello & Usyukov (2014). This disparity is striking and could indicate that people do not cycle as they state they do or that these behaviours are not yet successfully measured. This is further discussed in 5.2.

Similarly to the previous behaviour, some of the major influences on continuous cycling behaviour were not significant in the model. Van Duppen (2012) found that cyclists attempt to minimise their stops by avoiding stop signs and traffic lights. The variables 'stop signs' and 'traffic lights' were not significant in the general model. In the model for the continuous variable 'traffic lights' was significant but had an unexpected positive effect. Cyclists thus prefer routes with traffic lights, which argues against continuous cycling. Other variables that are related to continuous cycling are the avoidance of many turns. However, the effect of the number of turns was negative, which contrasts finding by Van Lierop et al. (2020) that cyclists avoid turns when cycling continuously. This entails that people would rather take routes with more turns, which reinforces that cyclists prefer shorter routes over continuous routes. Additionally, separated cycling paths are also avoided. However, a preference for cycling highways does indicate some continuous cycling behaviour, conform Liu et al. (2019) and Van Lierop et al. (2020). Results from the choice model remain inconclusive as this variable was not significant in the model with only variables related to continuous cycling behaviour, but significant and positive in the general model. However, for the latter it should be taken into account that few routes encountered cycling highways. This means that cycling highways encourage continuous cycling. If cycling highways are present, people do cycle continuously. This is not the case for travel time minimisation because cycling highways have a stronger influence on continuous cycling than fast cycling. As with the previous cycling behaviour, the results from the survey show that people state they do attempt to cycle continuously. Again, this discrepancy could mean that cyclists do not cycle as continuously as they think or the method of measuring continuous cycling is flawed. This is further discussed in 5.2.

5.1.4 Interactions

The interactions between variables have influenced the results in different ways. Firstly there is gender. While both genders prefer routes with more turns and higher turn costs, the odds ratio for both variables is higher for men. This means that men are more likely to take shorter routes which avoid main roads. Consequently, men also prefer to minimise the number of traffic lights. This indicates some travel minimisation behaviour from men. For women this variable was not significant. These outcomes show a difference of how the genders relate to the cycling behaviours. Men seem to more actively look for the best route alternative for them. These findings indicate that men do prefer to minimise travel time to some extent. This would confirm the findings from Broach et al. (2012), that men prefer minimising travel time more than women. This contrasts the findings from Dill & Gliebe (2008), who found the opposite to be true.

When it comes to age, no major differences between the ages were found in the statistical analysis. This contradicts findings from Stinson & Bhat (2003), Kalter et al. (2018) and Menghini et al. (2010), whom all found that older cyclists minimise travel time less and cycle less continuously than younger people. However, in the results from the survey, this pattern can be recognised as higher age ranges stated to prioritise travel distance minimisation, travel time minimisation and continuous cycling to a lesser extent than younger age groups.

None of the categorised cycling behaviours showed any particular preference for minimising travel distance, minimising travel time or continuous cycling. In contrast to the findings from Broach et al. (2012) and Li (2017), participants that cycled for utilitarian purposes were more likely to take longer routes. For the other trip purposes, the length variable was not significant. The other variables showed no indication that the trip purpose affects the extent to which travel time minimisation or continuous cycling occurred differently compared to the general model. The variable 'traffic lights' was significant for most trip purposes. However, the values for all purposes were positive which does not indicate travel time minimisation or continuous cycling. Overall no big influence from any trip purposes were found. This is in contrast to findings by Broach et al. (2012) and Li (2017). The Covid-19 crisis might be the reason that trip purpose does not affect cycling behaviour as expected, as cyclists cycled less for commutes and may have cycled differently for utilitarian purposes. Another reason for these results could come from the fact that the trip purposes were determined based on the OD points, rather than asking participants their trip purpose. This could have resulted in some inaccurate data.

The effect of the bicycle types is hard to determine because for most variables the interaction effect could not be measured as they were not significant. However, almost all bicycle types showed a preference for high turn costs and a higher number of turns, especially E-bikes. This contradicts the findings from Spinney (2007) and Van Lierop et al. (2020) who indicate that continuous cycling and travel time minimisation is more common than with other bicycles.

5.2 Research Limitations

In this section the research limitations will be discussed, as well as suggestions for way in which this research could be improved. This research was carried out during the Covid-19 pandemic, which affected the research. The research has been carried out within the limitations that resulted from this in addition to existing time limits and restrictions on resources.

5.2.1 Data gathering process

As was stated before, this research was carried out during the Covid-19 pandemic. This resulted in various constraints which made the data gathering process more difficult. Firstly, there is the way in which the data is gathered. Originally, the plan was to gather data from randomly selected respondents. However, due to lockdown restrictions this option was deemed irresponsible. Hence, family and friends of the students involved in the project were asked to participate. This choice allowed for the research to be carried out but did result in a less representative research population. Moreover, it resulted in a research area which was spread out over the middle of the Netherlands and both urban and rural municipalities. This resulted in a preference for routes through nature and rural areas. The chosen method of data collection has the advantage that it allows for an additional survey to gather personal data, which is an important influence on cycling behaviour according to literature. However, the data population is smaller compared to existing datasets that do not include personal data. The smaller research population also results in multiple routes being cycled by the same people. This can skew the prominence of certain cycling motives. Future research could opt for random selection of participants to create a more representative and bigger research population.

The process of deciding to approach friends and family resulted in delays in the data gathering process. The data gathering was further delayed by the fact that approval was needed to use the GPS trackers from the TU Delft. Only 20 trackers were available which meant that the data collection phase had to be spread out over time, further delaying the data gathering process. Furthermore, the accuracy of the GPS trackers was not as high as expected, resulting in difficulties during the map matching process. Some of the GPS trackers did not work properly, resulting in a loss of data from some participants.

5.2.2 Data quality

There are also some limitations that were imposed due to the quality of existing datasets that were used. The network dataset that was used was provided by the Fietsersbond. Because the network is maintained by members from the Fietsersbond, the topology of the network is not perfect, resulting in difficulties during the map matching process and the route generation. Because of an innaccurate network topology and a map matching software that did not work, the choice was made to do the map matching process manually. The quality of the network topology still needed to be fixed to generate the route alternatives in the network analyst in ArcGIS Pro. An advantage of the Fietsersbond network is the fact that it contains attributes about the network. This brings with it another limitation, because these attributes are gathered and provided by volunteers. This could lead to inaccuracies in some of the attribute data. In the future another network could be used with better topology and more objective attributes. However, it currently seems that the Fietsersbond network is the most accurate and complete network for modelling cycling behaviour, because of the focus on cycling instead of cars.

5.2.3 Methodological limitations

During the methodology phase of the research various decisions were made that affected the outcomes in this research. Firstly, there is the data filtering process. During this process the cycling trips were filtered based on cycling speeds average and maximum. Routes were cut off after pauses of 5 minutes. By altering the speed range and the cut off point, different routes could have been created from the GPS data. Secondly, there is the process of adding outside spatial data to the network. This data was added through the use of buffers and spatial joins. Different buffers sizes could have resulted in different spatial attributes which can affect the results in the statistical analysis.

Another step in the methodological process that affected the results is the creation of the choice set. During this process the specifications of each of the route alternatives were determined. By altering the cost attributes and the restriction attributes, different route alternatives could have been created. Consequently, different interpretations of what costs and restrictions make for the fastest, most continuous or greenest route alternatives, could impact the resulting routes. For future research different inputs could be explored to create different labelled alternatives. However, the choice could also be made to step away from labelled alternatives and generate random route alternatives instead.

The way in which variables in the statistical analysis are used also affects the outcomes of the research. For example, instead of taking the percentage of the route that goes through an urban environment, a Boolean could be used to indicate whether a route is urban or not. Other variables that were impacted by methodological choices are trip purpose and bicycle type. These variables were determined based on a combination of survey questions and properties of the observed routes. The classifications are made based on certain assumptions. This methodology is not as accurate as asking respondents their trip purpose and bicycle, but it is a feasible way to get insight into these variables. Future research could ask respondents to state their trip purpose and the bicycle they used afterwards. However. Because this process asks more effort from participants, the choice was made not to do this as to not deter potential participants from participating.

Lastly, the way in which this research determines cycling behaviours also affects the results. While travel distance minimisation is only determined by the ‘length’ variable, travel time minimisation and continuous cycling are defined by several variables. Although the chosen variables are based on results from literature and variables for which data was available, it could be argued that some of the used variables do not completely represent the cycling behaviour. Moreover, additional variables that were not used could be missing. This could be an explanation as to why travel time minimisation and continuous cycling were not observed from the choice model, but the participants did state that they cycle according to these cycling motives. Future research could introduce new variables or remove the variables that represent these cycling behaviours. For example, a proper travel time attribute could be introduced for minimising travel time.

6. Conclusion

The question that was central to this research was: *"To what extent do cyclists minimise travel distance, minimise travel time and cycle continuously and which characteristics influence these behaviours?"* The hypothesis is that cyclists primarily cycle to minimise travel distance but that they deviate for a shorter travel time or to cycle continuously. In order to answer the main research question and to establish whether the hypothesis was correct the sub questions are answered first.

6.1 Sub questions

The first sub question is: *"How can minimising travel distance, minimising travel time and continuous cycling be defined and measured?"* This sub question was answered in the literature review and the determination of the variables. Minimising travel distance can be defined as the extent to which cyclists take the shortest route or prefer shorter routes over longer alternatives. Minimising travel time can be defined as the extent to which cyclists take the shortest path, minimise turn costs, avoid stop signs and traffic lights and prefer separated cycling paths and cycling highways. Continuous cycling is defined as the extent to which cyclists minimise the number of turns, avoid stop signs and traffic lights and prefer separated cycling paths and cycling highways.

The second sub question is: *"To what extent do cyclists state to minimise travel distance, travel time and cycle continuously?"* Results from the survey show participants generally state that they attempt to minimise travel distance, minimise travel time and cycle continuously. Participants mostly agreed with statements about minimising travel distance and travel time, while being slightly more neutral towards the statement about continuous cycling. Women were slightly more likely to state to minimise travel distance and travel time, while men were slightly more likely to state to cycle continuously. Younger participants were more likely to agree with the statements for all of all three cycling behaviours compared to older participants.

The third sub question is: *"To what extent do cyclists minimise travel distance?"* In the choice model the variable 'length' was significant and negatively influences route choice. This shows that cyclists do attempt to minimise travel distance. Furthermore, the shortest route alternative had the highest overlap with the observed routes out of all the generated alternatives. However, the average overlap was 53%, indicating that the shortest route is not always taken and that other motives besides travel distance also play a role in route choice.

The fourth sub question is: *"To what extent do cyclists minimise travel time?"* Length was shown to be significant and have a negative influence on route choice. While the preference for shorter routes does indicate some minimisation of travel time, the results from other relevant variables that were significant contradicted the hypothesis. Only the positive influence of cycling highways indicates some travel time minimisation. Because cycling highways are not very common and other variables do not suggest travel time minimisation, it is not likely travel time is minimised. The negative influence of the variable 'length' thus more likely suggest travel distance minimisation. Furthermore, the overlap of the observed routes with the fastest alternative was average among the other alternatives.

The fifth sub question is: *"To what extent do cyclists attempt to cycle continuously?"* Continuous cycling does not seem to be the main cycling motive when deviating from the shortest path. The choice model shows that the number of turns and traffic lights positively affect route choice and that separated cycling paths are a negative influence on route choice. This suggests that continuous cycling is not prioritised. However, cycling highways do have a positive influence on route choice but they are very rare. This indicates that continuous cycling is present when there are cycling highways. Lastly, the overlap of the observed routes with the most continuous route is the second lowest among all alternatives.

The final sub question is: *Which characteristics influence cycling behaviour and in what way? "* Length is the most important influence on route choice and has a negative influence. Cycling highways have a high positive influence but they are not common in the network. Furthermore, the turn costs and the number of turns positively influence route choice, while separated cycling paths negatively influence route choice. Stop signs and traffic lights were not significant in the main model. Lastly, urban areas negatively influence route choice, while nature areas positively influence route choice.

Some personal characteristics also influence cycling behaviour. The interaction of gender on route choice shows that men are more likely than women to have higher turn costs and take more turns. However, men do this to avoid traffic lights, which was significant and had a negative influence on route choice. The age of participants affect route choice. Trip purpose and bicycle type did not show clear influences as most of the interaction effects were not significant.

6.2 Main Research Question

In conclusion it can be said that minimising travel distance is the most important cycling motive due to the strong negative effect of length on route choice and the fact that the shortest route had the highest overlap with the observed route out of all the route alternatives. Cyclists prefer the shortest route, despite increasing the number of turns and turn costs along the way, thus avoiding more continuous routes. However, the average overlap of the shortest routes and the observed routes of 53% indicates that cyclists still deviate from the shortest path.

The hypothesis, that the main motives for deviating from the shortest path are to minimise travel time and cycle continuously, cannot be confirmed. At best, there are opposing effects from influencing characteristics related to these behaviours. The lower overlap of the fastest route alternative and the continuous route alternative with the observed routes further confirm this.

However, there are two exceptions to the outcomes stated above. Firstly, cyclists do choose for continuous routes when cycling highways are present. Cycling highways are not common in the network but have a strong positive influence on route choice. These outcomes contradict the outcomes of the survey, in which the cyclists state they do minimise travel time and cycle continuously. Secondly, men do seem to show some slight travel time minimisation because for men traffic lights have a negative effect on route choice. Moreover, this is done despite the higher turn costs and number of turns men have. This shows that men avoid major roads with traffic lights and prefer winding roads.

Overall, this could mean that travel time minimisation and continuous cycling are the reasons for deviating from the shortest route, but that the way this is measured in this research was not able to properly measure these behaviours. However, it could also mean that other cycling motives play a bigger role in cycling route choice.

6.3 Policy Recommendations and Future Research

In the introduction it is stated that a better insight into cycling motives can provide new perspectives on how cycling infrastructure can be adapted to the needs of the cyclists and the increasing number of cyclists (Huiberts, 2020). Based on the results in this research, alternative routes with fewer turns and separated cycling paths will not incentivise different cycling routes for cyclists. Cyclists still prefer the shortest routes. However, cyclists do show a preference for cycling highways. This means that municipalities in the Netherlands could create more cycling highways, or similar cycling paths, in order to create viable alternative routes for cyclists. If the busy cycling paths are already along more continuous roads, then shorter alternatives with more turns could attract cyclists to deviate from the busy routes.

Because the hypothesis from this research could not be confirmed, future research should look into what other cycling behaviours cyclists prioritise when deviating from the shortest path. Alternatively, future research could also build on the methods in this research to measure travel time minimisation and continuous cycling. There are several ways in which this can be done. Firstly, different cycling data could be used to asses these cycling behaviours. For example, by using a dataset with more participants and cycling data that is concentrated in one location. This would result in a research with a more representative research population in a similar location rather than municipalities all over the Netherlands. Secondly, a different method of generating route alternatives could be used. Future research could use randomly generated route alternatives. Alternatively, other labelled alternatives could be created with different impedance costs and restrictions that better reflect the fastest and most continuous route. Lastly, it could be worthwhile to repeat this research when there are no more Covid-19 restrictions in place. This way a comparison between cycling behaviour under normal circumstances can be compared to cycling behaviour during a lockdown.

References

ANWB (2020, August 19th) Fietspaden te smal en te vol. ANWB website. Consulted on: 01-12-2020. Retrieved from: https://www.anwb.nl/nieuws/fiets/2020/augustus/fietspaden-smal-en-vol

Aultman-Hall, L., Hall, F. L., & Baetz, B. B. (1997). Analysis of bicycle commuter routes using geographic information systems: implications for bicycle planning. *Transportation research record*, *1578*(1), 102 110.

Ben-Akiva, M. E. & Lerman, S. R. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand, MIT Press, Cambridge, Ma

Bernardi, S., La Paix Puello, L., & Geurs, K. (2018). Modelling route choice of Dutch cyclists using smartphone data. *Journal of transport and land use*, *11*(1), 883-900.

Bierlaire, M. (1998). Discrete choice models. In *Operations research and decision aid methodologies in traffic and transportation management* (pp. 203-227). Springer, Berlin, Heidelberg.

Bovy, P. H. (2009). On modelling route choice sets in transportation networks: a synthesis. *Transport reviews*, *29*(1), 43-68.

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*, *46*(10), 1730-1740.

Broer, K. (2017, November 7th) Hoe dwing je een maximum sneleheid of het fietspad af? *Fietsersbond*. Consulted on 5-10-2020. Retrieved from: https://www.fietsersbond.nl/nieuws/hoe-dwing-maximumsnelheid-op-fietspad-af/

Cascetta, E., & Papola, A. (2001). Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. *Transportation Research Part C: Emerging Technologies*, *9*(4), 249-263.

Casello, J. M., & Usyukov, V. (2014). Modeling cyclists' route choice based on GPS data. *Transportation Research Record*, *2430*(1), 155-161.

CBS (2020) Wijk- en buurtkaart 2020. *Centraal Bureau voor de Statistiek [CBS].* Consulted on: 01-03-2020. Retrieved from: https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/wijk-en buurtkaart-2020

Copenhagenize index (2019) *The most bicycle friendly cities of 2019.* Consulted on: 30-09-2020. Retrieved from: https://copenhagenizeindex.eu/.

Dane, G., Feng, T., Luub, F., & Arentze, T. (2019). Route choice decisions of E-bike users: Analysis of GPS tracking data in the Netherlands. In *The Annual International Conference on Geographic Information Science* (pp. 109-124). Springer, Cham.

De Bruijne, R.J. (2016). Revolutie of risico? Een onderzoek naar de verkeersveiligheidsaspecten van de speed pedelec. De Bilt: Grontmij

Dill, J., & Gliebe, J. (2008). Understanding and measuring bicycling behavior: A focus on travel time and route choice.

Ehrgott, M., Wang, J. Y., Raith, A., & Van Houtte, C. (2012). A bi-objective cyclist route choice model. *Transportation research part A: policy and practice*, *46*(4), 652-663.

Fietsersbond (n.d.) Fietsen in Cijfers. Fietsersbond. Consulted on: 6-10-2020. Retrieved from: https://www.fietsersbond.nl/ons-werk/mobiliteit/fietsen-cijfers/

Fietsersbond (n.d.) Cycling network. Retrieved from: https://docplayer.nl/7295641-Metagegevens-database fietsrouteplanner-fietsknooppunten-poi-s.html

Heinen, E., Maat, K., & van Wee, B. (2013). The effect of work-related factors on the bicycle commute mode choice in the Netherlands. *Transportation*, *40*(1), 23-43.

Hood, J., Sall, E., & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation letters*, *3*(1), 63-75.

Huiberts, A. (2020, October 10th) Utrechters mijden steeds vaker de drukke fietspaden. *DUB*. Consulted on: 5 11-2020. Retrieved from: https://www.dub.uu.nl/nl/achtergrond/utrechters-mijden-steeds-vaker-de-drukke-fietspaden?utm_medium=email

Hull, A., & O'Holleran, C. (2014). Bicycle infrastructure: can good design encourage cycling?. *Urban, Planning and Transport Research*, *2*(1), 369-406.

Huss, A., Beekhuizen, J., Kromhout, H., & Vermeulen, R. (2014). Using GPS-derived speed patterns for recognition of transport modes in adults. *International journal of health geographics*, *13*(1), 40.

Kalter, M. J. O., Coffeng, G., & Groenendijk, L. (2018) Aantrekkelijkheid en afwisseling routes meer sturend in keuzegedrag fietsers dan snelheid.

Koh, P. P., & Wong, Y. D. (2013). Influence of infrastructural compatibility factors on walking and cycling route choices. *Journal of Environmental Psychology*, *36*, 202-213.
Li, S. (2017). *Cycling in Toronto: Route Choice Behavior and Implications to Infrastructure Planning* (Master's thesis, University of Waterloo).
Liu, G., Te Brömmelstroet, M., Krishnamurthy, S., & van Wesemael, P. (2019). Practitioners' perspective on user experience and design of cycle highways. *Transportation research interdisciplinary perspectives*, *1*, 100010.
Menghini, G., Carrasco, N., Schüssler, N., & Axhausen, K. W. (2010). Route choice of cyclists in Zurich. *Transportation research part A: policy and practice*, *44*(9), 754-765.
Mobiliteitsplatform (2020, January 10th) Cycle data: nauwkeurige data. Mobiliteitsplatform. Consulted on: 5-10 2020. Retrieved from: https://www.mobiliteitsplatform.nl/artikel/cycle-data-nauwkeurige-data
Montini, L., Antoniou, C., & Axhausen, K. W. (2017). Route and mode choice models using GPS data. In *TRB 96th Annual Meeting Compendium of Papers* (pp. 17-03082). Transportation Research Board.
OpenStreetMap Contributors. (2020). Planet dump retrieved from https://planet.osm.org. Retrieved May 19, 2021, from https://planet.openstreetmap.org/
Papinski, D., & Scott, D. M. (2011). A GIS-based toolkit for route choice analysis. *Journal of Transport Geography*, *19*(3), 434-442.
Plazier, P. A., Weitkamp, G., & van den Berg, A. E. (2017). “Cycling was never so easy!” An analysis of e-bike commuters' motives, travel behaviour and experiences using GPS-tracking and interviews. *Journal of transport geography*, *65*, 25-34.
Prato, C. G. (2009). Route choice modeling: past, present and future research directions. *Journal of choice modelling*, *2*(1), 65-100.
Pritchard, R. (2018). Revealed preference methods for studying bicycle route choice - A systematic review. *International journal of environmental research and public health*, *15*(3), 470.
Schweizer, J., & Rupi, F. (2015). 17 Map matching and cycling infrastructure analyses with SUMO and python. *SUMO2014-Modeling Mobility with Open Data*, 195.
Schweizer, J., Bernardi, S., & Rupi, F. (2016). Map-matching algorithm applied to bicycle global positioning system traces in Bologna. *IET Intelligent Transport Systems*, *10*(4), 244-250.
Skov-Petersen, H., Barkow, B., Lundhede, T., & Jacobsen, J. B. (2018). How do cyclists make their way?-A GPS based revealed preference study in Copenhagen. *International Journal of Geographical Information Science*, *32*(7), 1469-1484.
Spinney, J. (2006). A place of sense: a kinaesthetic ethnography of cyclists on Mt Ventoux. Environment and Planning D; Society & Space 24(5), pp. 709–732
Stata (n.d.) Choice models manual. Stata *Manual*. Retrieved from: https://www.stata.com/manuals/cm.pdf#cmcmtab
Steintjes, S. B. (2016). Comparing and analysing the behaviour of users of conventional bicycles and speed pedelecs: Naturalistic cycling (Master's thesis).
Stinson, M. A., & Bhat, C. R. (2003). Commuter bicyclist route choice: Analysis using a stated preference survey. *Transportation research record*, *1828*(1), 107-115.
Te Brömmelstroet, M. (2014). De choreografie van een kruispunt: Naar een gebruiksgeorienteerde ontwerplogica voor kruispunten. In *Colloquium Vervoersplanologisch Speurwerk* (pp. 1-15).
Holligan, A. (2013, August 8th) Why is cycling so popular in the Netherlands? *BBC*
Ton, D., Cats, O., Duives, D., & Hoogendoorn, S. (2017). How Do People Cycle in Amsterdam, Netherlands?: Estimating Cyclists’ Route Choice Determinants with GPS Data from an Urban Area. *Transportation research record*, *2662*(1), 75-82.
Van Duppen, J. M. A. (2012). Urban Trajectories is een initiatief van Architectuurcentrum Aorta, Casco–Office for Art, Design and Theory en Stichting Vrede van Utrecht in samenwerking met Universiteit Utrecht.
Van Lierop, D., Soemers, J., Hoeke, L., Liu, G., Chen, Z., Ettema, D., & Kruijf, J. (2020). Wayfinding for cycle highways: Assessing e-bike users' experiences with wayfinding along a cycle highway in the Netherlands. *Journal of Transport Geography*, *88*, 102827.
Wardlaw, M. J. (2014). History, risk, infrastructure: perspectives on bicycling in the Netherlands and the UK. *Journal of Transport & Health*, *1*(4), 243-250.
Yeboah, G., & Alvanides, S. (2015). Route choice analysis of urban cycling behaviors using OpenStreetMap: Evidence from a British urban environment. In *OpenStreetMap in GIScience* (pp. 189-210). Springer, Cham.

Appendix

1. **Instruction letter respondents**

Geachte heer of mevrouw,

Allereerst hartelijk dank voor het meedoen aan ons onderzoek.

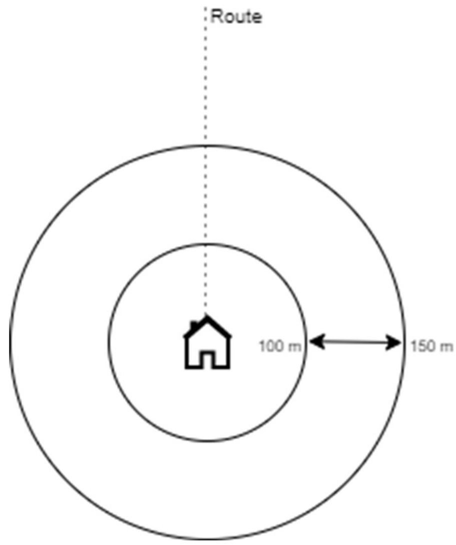
In deze brief vindt u een handleiding voor het gebruik van de GPS-trackers.

Wat we aan u vragen is het volgende:

- De GPS-tracker een week lang bij u te dragen; het maakt niet uit of u hem in uw zak of tas stopt.
- De GPS-tracker iedere dag op te laden; hiervoor ontvangt u een oplaadkabel.
- Achteraf de naar u gestuurde enquête in te vullen.

We benadrukken dat uw privacy goed gewaarborgd wordt:

- Uw data wordt automatisch verstuurd naar een centrale server waartoe alleen onze begeleider toegang tot heeft.
- Onze begeleider anonimiseert de GPS-ritten door de routes op ongeveer 100 tot 150 meter van uw huisadres te laten stoppen. Daardoor is het onmogelijk om uw huisadres te traceren. We weten alleen dat het ergens op 100 tot 150 meter afstand van het eind van de route ligt. Dit is zichtbaar in de afbeelding. Voor ons lijkt het alsof uw huis ergens in de buitenste ring staat terwijl uw huis daadwerkelijk in het midden van de cirkel staat.
- Wij hebben geen toegang tot de originele data; die wordt door onze begeleider verwijderd.
- In lijn met richtlijnen van het rijk, heeft onze begeleider een privacyverklaring getekend. Onderzoekers hebben geen enkel belang bij individuele data, maar kijken alleen naar patronen van de hele groep.
- De data wordt niet doorgegeven aan derden.



Bij vragen kunt u altijd contact opnemen met de persoon van wie u de tracker heeft ontvangen.

- Harmke Vliek:
  g.h.vliek@students.uu.nl
  0611689821
- Jimme Smit:
  j.smit6@students.uu.nl
  0616374277

- Maaike Kuiper:
  m.d.kuiper2@students.uu.nl
  0623407886

Bij voorbaat dank en veel fietsplezier!

2. **Survey**

Enquete Fietsgedrag

Start of Block: Default Question Block

Geachte heer of mevrouw,
Wij zijn studenten aan de Universiteit Utrecht en de TU Delft. Wij doen onderzoek naar fietsroutekeuze. Hiervoor vragen wij u om uw fietsverplaatsingen bij te houden met een zogeheten GPS-tracker. Ook vragen wij u of u deze enquête wilt invullen.
In deze enquête stellen wij u een aantal vragen die ons helpen om uw fietsverplaatsingen te begrijpen. De vragen gaan over de frequentie van uw fietsgedrag en uw voorkeuren met betrekking tot de omgeving waarin u fietst tijdens uw route. Tot slot zullen er nog wat persoonskenmerken gevraagd worden. Uw gegevens worden vertrouwelijk behandeld en volledig anoniem verwerkt, zoals we beschreven hebben in de brief. Het invullen van de enquête vraagt ongeveer 15 minuten van uw tijd.

Alvast hartelijk bedankt voor uw medewerking!

Page Break

Vraag 1 GPS nummer (Deze is te vinden op de achterkant van de tracker):

____

De volgende vragen gaan over uw fietsgewoonten voordat de corona crisis uitbrak.

Vraag 2 Voor welke doelen gebruikt u een fiets (voor de corona crisis)? (meerdere antwoorden mogelijk)

- ☐ Woon-werkrit (1)
- ☐ Zakelijk: ritten tijdens werk (2)
- ☐ School of studie (3)
- ☐ Voorzieningen bezoeken (winkels, supermarkten en dergelijke) (4)
- ☐ Recreatief (5)
- ☐ Sport (6)
- ☐ Vrienden of familie bezoeken (7)
- ☐ Overig (8)

Vraag 3 Wat voor soort fiets(en) heeft u?

- ☐ Stadsfiets zonder versnellingen (1)
- ☐ Stadsfiets met versnellingen (2)
- ☐ Sportieve fiets (hybride) (3)
- ☐ Racefiets (4)
- ☐ Mountainbike (5)
- ☐ Elektrische fiets (max. 25 km/uur) (6)
- ☐ Speed pedelec (max. 45 km/uur) (7)
- ☐ Ligfiets (8)
- ☐ Vouwfiets (9)
- ☐ Tandem (10)
- ☐ OV-fiets (13)
- ☐ Anders, namelijk (11) ________________

Vraag 4 Indien u meerdere fietsen heeft, voor welk doel gebruikt u elke fiets? Selecteer de doelen waarvoor u fietst en selecteer het soort fiets dat u daarvoor gebruikt

	Woon - werkrit (1)	Zakelijk (2)	School of studie (3)	Voorzieningen bezoeken (4)	Recreatief (5)	Sport (6)	Vrienden of familie bezoeken (7)	Overig (8)
Stadsfiets zonder versnellingen (1)								
Stadsfiets met versnellingen (2)								
Sportieve fiets (hybride) (3)								
Racefiets (4)								
Mountainbike (5)								
Elektrische fiets (max. 25km/uur) (6)								
Speed pedelec (max. 45km/uur) (7)								
Ligfiets (9)								
Vouwfiets (10)								
Tandem (11)								
Ov-fiets (12)								
Overig (13)								

Vraag 5 Wat is de onderhoudsstaat van uw meest gebruikte fiets?

- Uitstekend onderhouden (1)
- Goed onderhouden (2)
- Gemiddeld (3)
- Minder goed onderhouden (4)
- Niet onderhouden (5)

Page Break

Vraag 6 Op welke leeftijd bent u zelf gaan fietsen in de openbare ruimte? (Vul alleen het getal in)

____

Vraag 7 Hoe vaak fietst u?

- ○ Dagelijks (1)
- ○ Ongeveer ... dagen per week (2) ____
- ○ Ongeveer ... dagen per maand (3) ____
- ○ Zelden (4)
- ○ Nooit (5)

Vraag 8 Welke afstand (enkele reis) fietst u ongeveer voor de volgende bestemmingen indien van toepassing?

- ▢ Werk: ... km (1) ____
- ▢ School/studie: ... km (2) ____
- ▢ Boodschappen: ... km (3) ____
- ▢ Naar openbaar vervoer: ... km (4) ____

Page Break

Vraag 9
Deze stellingen gaan over het thema omgeving.

Geef uw voorkeuren aan in de volgende uitspraken.
Kruis aan in hoeverre u het eens bent (bedenk dat het anoniem is).

	Helemaal oneens (1)	Oneens (2)	Neutraal (3)	Eens (4)	Helemaal eens (5)
Ik fiets graag langs het *water*, ook als dit niet de kortste route is. (1)	○	○	○	○	○
Ik fiets graag door het *park*, ook als dit niet de kortste route is. (2)	○	○	○	○	○
Ik fiets graag door het *bos*, ook als dit niet de kortste route is. (3)	○	○	○	○	○
Fietsen door een *mooie omgeving* is voor mij belangrijk, ook als dit niet de kortste route is. (4)	○	○	○	○	○
Ik fiets graag via een *levendige route*, ook als dit niet de kortste route is. (5)	○	○	○	○	○
Ik fiets graag langs herkenningspunten zoals kunstwerken of gebouwen door een interessante stedelijke omgeving. (6)	○	○	○	○	○
Met slecht weer zoek ik een beschutte route. (7)	○	○	○	○	○
Ik vermijd het liefst een lawaaiige omgeving, bijvoorbeeld van verkeer. (8)	○	○	○	○	○

Page Break

Vraag 10

De volgende stellingen gaan over het thema veiligheid. Geef uw voorkeuren aan in de volgende uitspraken.

Kruis aan in hoeverre u het eens bent (bedenk dat het anoniem is).

	Helemaal oneens (1)	Oneens (2)	Neutraal (3)	Eens (4)	Helemaal eens (5)
Als het schemert of donker is, neem ik graag een route met goede straatverlichting. (1)	o	o	o	o	o
Ik vermijd verkeerslichten zo veel mogelijk. (2)	o	o	o	o	o
Ik vermijd drukke kruispunten zo veel mogelijk. (3)	o	o	o	o	o
Ik fiets liever op fietspaden die gescheiden zijn van de weg. (4)	o	o	o	o	o
Ik vermijd verkeersdrukte. (5)	o	o	o	o	o
Ik fiets liever op wegen waar de maximale snelheid voor auto's 30 km/u is. (6)	o	o	o	o	o
Ik vermijd onveilige routes, ook als dit niet de kortste route is. (7)	o	o	o	o	o
Ik houd rekening met sociale veiligheid in mijn routekeuze. (8)	o	o	o	o	o
Ik houd rekening met de kans op diefstal als ik mijn fiets parkeer. (10)	o	o	o	o	o
Ik vermijd verlaten gebieden, ook als dit niet de kortste route is. (11)	o	o	o	o	o
Ik vermijd wegwerkzaamheden. (12)	o	o	o	o	o

Page Break

Vraag 11
De volgende stellingen gaan over het thema doorstroming.
Bij deze stellingen gaat het niet over recreatief fietsgedrag. Geef uw voorkeuren aan in de volgende uitspraken.

Kruis aan in hoeverre u het eens bent (bedenk dat het anoniem is).

	Helemaal oneens (1)	Oneens (2)	Neutraal (3)	Eens (4)	Helemaal eens (5)
Als ik fiets dan probeer ik de kortste route te nemen. (1)	○	○	○	○	○
Als ik fiets dan probeer ik de reistijd zo kort mogelijk te houden. (2)	○	○	○	○	○
Als ik fiets dan probeer ik zo veel mogelijk te fietsen zonder te hoeven stoppen (verkeerslichten bijvoorbeeld). (3)	○	○	○	○	○
Als ik fiets dan probeer ik een hoog tempo aan te houden. (4)	○	○	○	○	○
Als ik fiets neem ik wel eens doorsteekjes om een deel van de route af te snijden. (5)	○	○	○	○	○
Ik fiets nooit door een rood stoplicht. (6)	○	○	○	○	○

Page Break

Vraag 12
Geef uw voorkeuren aan in de volgende uitspraken.
Kruis aan in hoeverre u het eens bent (bedenk dat het anoniem is).

	Helemaal oneens (1)	Oneens (2)	Neutraal (3)	Eens (4)	Helemaal eens (5)
Ik vind fietsen ontspannend. (1)	o	o	o	o	o
Ik zie mijzelf als iemand met een goede conditie. (2)	o	o	o	o	o
Ik vind een e-bike duur. (3)	o	o	o	o	o
Ik vind een e-bike het geld waard. (4)	o	o	o	o	o

Page Break

De volgende vragen gaan over de mogelijke verandering van uw fietsgedrag met betrekking tot de corona crisis.

Vraag 13 Werkt u thuis vanwege de corona crisis?

- Ja (1)
- Nee (2)
- Gedeeltelijk (3)
- Niet van toepassing (4)

Vraag 14 Is uw fietsgedrag met betrekking tot werk of school veranderd sinds de coronacrisis?

- Ja, ik fiets meer (1)
- Ja, ik fiets minder (2)
- Nee (3)
- Niet van toepassing (4)

Vraag 15 Is uw recreatieve fietsgedrag veranderd sinds de coronacrisis?

- Ja, ik fiets meer (1)
- Ja, ik fiets minder (2)
- Nee (3)
- Niet van toepassing (4)

Vraag 16 Vermijdt u sinds de coronacrisis drukke fietsroutes?

○ Ja (1)

○ Nee (2)

○ Af en toe (3)

○ Niet van toepassing (4)

Page Break

Ten slotte vragen we u enkele persoonlijke gegevens. Uw gegevens worden anoniem verwerkt.

Vraag 17 Wat is uw geboortejaar?

____

Vraag 18 Wat is uw geslacht?

- ◯ Vrouw (1)
- ◯ Man (2)
- ◯ Anders / wil ik niet zeggen (3)

Vraag 19 Wat is uw hoogste opleiding?

- ◯ Lbo, mulo, mavo, vmbo of gelijkwaardig (1)
- ◯ Havo, vwo, mms, hbs, mbo of gelijkwaardig (2)
- ◯ Hbo of universiteit (3)
- ◯ Anders (4)

Vraag 20 Wat is de samenstelling van uw huishouden?

- ○ Alleenstaand zonder thuiswonende kinderen (1)
- ○ Alleenstaand met thuiswonende kinderen op de basisschoolleeftijd of jonger (2)
- ○ Alleenstaand met thuiswonende kinderen op de middelbare schoolleeftijd of ouder (3)
- ○ Samenwonend/gehuwd zonder thuiswonende kinderen (4)
- ○ Samenwonend/gehuwd met thuiswonende kinderen op de basisschoolleeftijd of jonger (5)
- ○ Samenwonend/gehuwd met thuiswonende kinderen op de middelbare schoolleeftijd of ouder (6)
- ○ Samenwonend met andere volwassenen (zoals studentenhuis, zorgcentrum of woongroep) (7)
- ○ Thuiswonend bij ouder(s) of pleegouder(s) (8)

Vraag 21 Wat is uw (belangrijkste) dagelijkse bezigheid?

- ○ Betaald werk, voltijd (1)
- ○ Betaald werk, deeltijd (minder dan 36 uur/week) (2)
- ○ School/studie (3)
- ○ Geen betaald werk, gepensioneerd, vrijwilligerswerk, overig (4)

Vraag 22 Wat is het netto inkomen van uw huishouden per maand? (als u samenwonend/gehuwd bent, beide inkomens tezamen).

- ◯ Minder dan € 2000 (1)
- ◯ Tussen € 2000 en € 4000 (2)
- ◯ Tussen € 4000 en € 6000 (3)
- ◯ Meer dan € 6000 (4)
- ◯ Dat weet ik niet / dat wil ik niet zeggen (5)

Vraag 23 In welk type woning woont u?

- ◯ Vrijstaand (1)
- ◯ Twee onder een kap (2)
- ◯ Rijwoning (3)
- ◯ Boven- of benedenwoning (4)
- ◯ Portiekwoning, flat, appartement (5)
- ◯ Anders (6)

Vraag 24 Beschikt u over een auto?

- ◯ Ja, altijd (1)
- ◯ Ja, maar in overleg (2)
- ◯ Nee (3)

In de volgende vragen, vragen wij naar postcodes. Hiermee kunnen wij niet uw woon- en werkadres achterhalen.

---

*

Vraag 25 Wat is de postcode van uw woning ? (XXXX AB)

____

---

*

Vraag 26
Wat is de postcode van uw werk of school? (XXXX AB)

____

Page Break

Vraag 27 Heeft u nog opmerkingen of suggesties?

____

End of Block: Default Question Block

3. Reclassifications

WEGNIVEAU2

NoData	0
ONBEKEND	0
Autosnelweg	1
Belangrijke hoofdweg	2
Langs drukke weg	3
Overige weg	4

WEGTYPE2

NoData		0
ONBEKEND		0
Bromfietspad (langs weg)	Scooter Lane	2
Fietspad (langs weg)	Bicycle Lane	3
Fietsstraat	Seperated Cycling Boulevard	4
Normale weg	Mixed Road	1
Solitair bromfietspad	Seperated Scooter path	2
Solitair fietspad	Seperated Cycling path	4
Veerpont	Mixed Road	1
Ventweg	Mixed Road	1
Voetgangersdoorsteekje	Voetgangersdoorsteekje	2
Voetgangersgebied	Mixed Road	1
Weg met fiets(suggestie)strook	Bicycle Lane	3

WEGDEKSRT2

NoData	0
ONBEKEND	0
Asfalt/beton	1
Halfverhard	2
Klinkers	3
Onverhard	4
Overig (hout/kinderkopjes e.d.)	5
Schelpenpad	6
Tegels	7

WEGKWAL2

NoData	0
ONBEKEND	0
Slecht	-1
Redelijk	0
Goed	1

HINDER2

NoData	0
ONBEKEND	0
Zeer weinig	1

Weinig		2
Redelijk		3
Veel		4
Zeer veel		5

VERLICHT2

NoData	0
ONBEKEND	0
Niet verlicht	1
Beperkt verlicht	2
Goed verlicht	3

OMGEVING2

NoData	0	
ONBEKEND	0	
Akkers/weilanden	1	2
Bebouwd (veel groen)	2	1
Bebouwd (weinig of geen groen)	3	1
Bos	4	3
Landelijk of dorps	5	2
Natuur (behalve bos)	6	3

WATER2

NoData	0
Ja	1
Nee	0

SCHOONH2

NoData	0
ONBEKEND	0
Zeer lelijk	-2
Lelijk/saai	-1
Neutraal	0
Mooi	1
Schilderachtig	2

MAXSNELH2

NoData	0		
ONBEKEND	0		
Stapvoets(15)	15	safe	1
30	30	safe	1
50	50	safe	1
60	60	unsafe	2
70	70	unsafe	2
80	80	unsafe	2
100	100	unsafe	2
120	120	unsafe	2

130	130	unsafe	2

KRP_TYPE2

nvt	0
Kruispunt	1
Kruispunt met VRI	2
Rotonde	3

SNELFIETS2

NoData	0
Ja	1
Nee	0

4. Data Filtering Script

```
import csv
import numpy


a = []
description = ["imei,time,lng,lat,angle,speed,altitude"]
route = [description]
name = "resultsTXX_X_01.csv"
inputname = "ResultsT21_4.csv"

data = list(csv.reader(open(inputname)))
name = name[:8] + inputname [8:12] + name[12:]

#Hier worden de punten met een snelheid lager dan 5 eruit gehaald
for i in range(len(data)):
    if i == 0:
        i += 1
    else:
        if int(data[i][5]) <= 5:
            i += 1
        else:
            a.append(data[i])
            i += 1


for j in range(len(a)):
    if j == len(a)-1:
        route.append(a[j])
        totalspeed = 0
        maxspeed = 0
        for k in range(len(route)-1):
            totalspeed += int(route[k+1][5])
            if int(route[k+1][5]) > maxspeed:
                    maxspeed = int(route[k+1][5])
        avgspeed = totalspeed / (len(route)-1)
        #Hier kiezen hoeveel punten een route zijn
        #Hier kiezen welke avg speed de cutoff is
        if len(route) > 20 and avgspeed < 30 and avgspeed > 10 and
maxspeed < 40:
            #with open(name, 'w', newline='', encoding='utf-8') as
myfile:
                #wr = csv.writer(myfile, quoting=csv.QUOTE_MINIMAL)
                #wr.writerows(route)
            with open(name, 'w', newline='', encoding='utf-8') as
csv_file:
                writer = csv.writer(csv_file, delimiter=',')
                writer.writerows(route)

            tempname = int(name[13:15])+1
            if tempname < 10:
                print(name)
                name = name[:14] + str(tempname) + name[15:]

            else:
                name = name[:13] + str(tempname) + name[15:]
        #route = [description]
```

```
        print("hoera")
        #route.clear()
    else:
        #print(a[j][1])
        datetime1 = a[j][1]
        date1 = (datetime1[:10])
        time1 = (datetime1[11:])

        datetime2 = a[j+1][1]
        date2 = (datetime2[:10])
        time2 = (datetime2[11:])
        #print(date1)
        #print(time1)

        Y1= int(date1[:4])
        m1= int(date1[5:7])
        d1= int(date1[8:10])
        H1= int(time1[:2])
        M1= int(time1[3:5])
        S1= int(time1[6:8])

        Y2= int(date2[:4])
        m2= int(date2[5:7])
        d2= int(date2[8:10])
        H2= int(time2[:2])
        M2= int(time2[3:5])
        S2= int(time2[6:8])

        timedifference = (S2+(60*M2)+(3600*H2))-(S1+(60*M1)+(3600*H1))
        #Hier kiezen na hoeveel minuten pauze een nieuwe route begint
        if abs(timedifference) > 300:
            route.append(a[j])
            totalspeed = 0
            maxspeed = 0
            for k in range(len(route)-1):
                totalspeed += int(route[k+1][5])
                if int(route[k+1][5]) > maxspeed:
                    maxspeed = int(route[k+1][5])
            avgspeed = totalspeed / (len(route)-1)
            #Hier kiezen hoeveel punten een route zijn
            #Hier kiezen welke avg speed de cutoff is
            if len(route) > 20 and avgspeed < 30 and avgspeed > 10 and
maxspeed < 40:
                #with open(name, 'w', newline='', encoding='utf-8') as
myfile:
                    #wr = csv.writer(myfile, quoting=csv.QUOTE_MINIMAL)
                    #wr.writerows(route)
                with open(name, 'w', newline='', encoding='utf-8') as
csv_file:
                    writer = csv.writer(csv_file, delimiter=',')
                    writer.writerows(route)

                tempname = int(name[13:15])+1
                if tempname < 10:
                    print(name)
                    name = name[:14] + str(tempname) + name[15:]

                else:
```

```
                name = name[:13] + str(tempname) + name[15:]
            route = [description]

        else:
            route.append(a[j])
```

5. Data Preparation Script

```
#This is an example of the calculation for the observed route. This
script is altered to perform the same calculation for the alternative
routes.

import pandas as pd
import numpy as np
import os
filenames = []
print(os.getcwd()) #path waar je nu in zit
for f in os.listdir():
    if f.endswith(".txt"):
        filenames.append(f)
dataframes = [pd.read_csv(f, delimiter=";") for f in filenames]
tracker_df = []
for df, name in zip(dataframes, filenames):
    df.insert(0, "Route", f"{name.split('.')[0]}")
    df.insert(1, "TrackerID", f"{name[:5]}")
    df.insert(2, "numberoflinks", f"{len(df)}")
    tracker_df.append(df)
final_df = pd.concat(tracker_df).reset_index(drop=True)
final_df.iloc[:, -25:-8] = final_df.iloc[:, -25:-8].convert_dtypes()
final_df.iloc[:, -25:-8] = final_df.iloc[:, -25:-8].apply(lambda x:
x.str.replace(",", "."))
final_df.iloc[:, -25:-8] = final_df.iloc[:, -25:-8].apply(lambda x:
x.astype("float").astype("int"))
final_df.iloc[:, -7:-2] = final_df.iloc[:, -7:-2].convert_dtypes()
final_df.iloc[:, -7:-2] = final_df.iloc[:, -7:-2].apply(lambda x:
x.str.replace(",", "."))
final_df.iloc[:, -7:-2] = final_df.iloc[:, -7:-2].apply(lambda x:
x.astype("float").astype("int"))
final_df["Shape_Le_1"] = final_df["Shape_Le_1"].convert_dtypes()
final_df["Shape_Le_1"] = final_df["Shape_Le_1"].str.replace(",",
".").astype("float").apply(lambda x: round(x, 2))

Length = final_df.loc[:, ["Route", "Shape_Le_1"]]
Length = Length.groupby("Route").sum()
Length["Total_Length"] = Length["Shape_Le_1"]
del Length["Shape_Le_1"]
final_df = pd.merge(final_df, Length, on = "Route")
final_df["Routeperc"] = ((final_df["Shape_Le_1"] /
final_df["Total_Length"]) * 100).apply(lambda x: round(x, 2))

pd.get_dummies(final_df, prefix="", prefix_sep="")
x = pd.get_dummies(final_df.loc[:, ["WEGNIVEAU", "WEGTYPE",
"WEGDEKSRT", "OMGEVING", "VERLICHTIN", "MAXSNELHEI", "KRP_TYPE"]])
x.insert(0, "Route", final_df["Route"])
x.insert(1, "TrackerID", final_df["TrackerID"])
x.insert(2, "numberoflinks", final_df["numberoflinks"])
frame1 = x.groupby(["Route", "TrackerID", "numberoflinks"]).sum()
frame1.reset_index(level = 2, inplace = True)
frame1.reset_index(level = 1, inplace = True)

y = final_df.loc[:, ["Route", "TrackerID", "WEGKWAL2", "HINDER2",
"SCHOONH2", "VERLICHT2"]]
y.convert_dtypes()
y["WEGKWAL2"] = y["WEGKWAL2"].astype(str).astype(int)
```

```
y["HINDER2"] = y["HINDER2"].astype(str).astype(int)
y["SCHOONH2"] = y["SCHOONH2"].astype(str).astype(int)
y["VERLICHT2"] = y["VERLICHT2"].astype(str).astype(int)
y = y.groupby("Route").mean()
frame2 = y.groupby(["Route"]).mean()
frame2.iloc[:, :4] = frame2.iloc[:, :4].apply(lambda x: round(x, 2))

z = final_df.loc[:, ["Route", "WATER2", "SNELFIETS2", "Routeperc"]]
z["Waterperc"] = z["WATER2"] * z["Routeperc"]
z["Snelfperc"] = z["SNELFIETS2"] * z["Routeperc"]
z.loc[:, ["Route", "Waterperc", "Snelfperc"]].groupby("Route").sum()
frame3 = z.loc[:, ["Route", "Waterperc",
"Snelfperc"]].groupby("Route").sum()

q = final_df.loc[:, ["Route", "MAXSNELH2", "Routeperc"]]
q["MAXSNELH3"] = np.where(q["MAXSNELH2"] == 1, 1, 0)
q["Maxsnelperc"] = q["MAXSNELH3"] * q["Routeperc"]
frame4 = q.loc[:, ["Route", "Maxsnelperc"]].groupby("Route").sum()
frame4.insert(0, "Alternative", "Observed")

Wegtype= final_df.loc[:, ["Route", "WEGTYPE", "WEGTYPE2", "Routeperc"]]
Wegtype["Mixed_Road"] = np.where(Wegtype["WEGTYPE2"] == 1, 1 *
Wegtype["Routeperc"], 0)
Wegtype["Non_Cyclingpath"] = np.where(Wegtype["WEGTYPE2"] == 2, 1 *
Wegtype["Routeperc"], 0)
Wegtype["Bicyle_lane"] = np.where(Wegtype["WEGTYPE2"] == 3, 1 *
Wegtype["Routeperc"], 0)
Wegtype["Separated_Bicyclepath"] = np.where(Wegtype["WEGTYPE2"] == 4, 1
* Wegtype["Routeperc"], 0)
Wegtype["Voetgangersdoorsteekje"] = np.where(Wegtype["WEGTYPE"] ==
"voetgangersdoorsteekje", 1 * Wegtype["Routeperc"], 0)
Wegtype2 = Wegtype.loc[:, ["Route", "Mixed_Road", "Non_Cyclingpath",
"Bicyle_lane", "Separated_Bicyclepath",
"Voetgangersdoorsteekje"]].groupby("Route").sum()

Omgeving = final_df.loc[:, ["Route", "OMGEVING2", "Routeperc"]]
Omgeving["Landelijk1"] = np.where(Omgeving["OMGEVING2"] == 1, 1 *
Omgeving["Routeperc"], 0)
Omgeving["Landelijk5"] = np.where(Omgeving["OMGEVING2"] == 5, 1 *
Omgeving["Routeperc"], 0)
Omgeving["Landelijk"] = Omgeving["Landelijk1"] + Omgeving["Landelijk5"]
Omgeving["Bebouwd2"] = np.where(Omgeving["OMGEVING2"] == 2, 1 *
Omgeving["Routeperc"], 0)
Omgeving["Bebouwd3"] =  np.where(Omgeving["OMGEVING2"] == 3, 1 *
Omgeving["Routeperc"], 0)
Omgeving["Bebouwd"] = Omgeving["Bebouwd2"] + Omgeving["Bebouwd3"]
Omgeving["Natuur4"] = np.where(Omgeving["OMGEVING2"] == 4, 1 *
Omgeving["Routeperc"], 0)
Omgeving["Natuur6"] = np.where(Omgeving["OMGEVING2"] == 6, 1 *
Omgeving["Routeperc"], 0)
Omgeving["Natuur"] = Omgeving["Natuur4"] + Omgeving["Natuur6"]
Omgeving2 = Omgeving.loc[:, ["Route", "Landelijk", "Bebouwd",
"Natuur"]].groupby("Route").sum()

NewMerge = pd.merge(frame4, Wegtype2, on = "Route").merge(Omgeving2, on
= "Route")
NewMerge.to_excel("ObservedNewMerge.xlsx")
```

```
result = pd.merge(frame1, frame2, on = "Route").merge(frame3, on =
"Route")

spatjoin = final_df.iloc[:, :1]
spatjoin["Ongevallen"] = final_df.loc[:, "Count_sum"]
spatjoin["VerkeersLichten"] = final_df.loc[:, "Sign_Count"]
spatjoin["StopBorden"] = final_df.loc[:, "Stop_Count"]
spatjoin["Bruggen"] = final_df.loc[:, "Brug_count"]
spatjoinA = spatjoin.groupby("Route").sum()
spatjoinB = final_df.iloc[:, :1]
spatjoinB["Misdaadcijfers"] = final_df.loc[:, "Criminalit"]
spatjoinB["Misdaadcijfers"] =
spatjoinB["Misdaadcijfers"].convert_dtypes()
spatjoinB["Misdaadcijfers"] =
spatjoinB["Misdaadcijfers"].astype("float")#.astype("int")
spatjoinC= spatjoinB.groupby("Route").mean()
spatjoin2 = pd.merge(Length, spatjoinA, on = "Route").merge(spatjoinC,
on = "Route")

spatjoin2["Ongev/len"] = spatjoin2["Ongevallen"] /
spatjoin2["Total_Length"]
spatjoin2["Verkeerslicht/len"] = spatjoin2["VerkeersLichten"] /
spatjoin2["Total_Length"]
spatjoin2["Stop/len"] = spatjoin2["StopBorden"] /
spatjoin2["Total_Length"]
spatjoin2["Brug/len"] = spatjoin2["Bruggen"] /
spatjoin2["Total_Length"]
spatjoin2["Misdaad/len"] = spatjoin2["Misdaadcijfers"] /
spatjoin2["Total_Length"]
spatjoin2["Ongev/len"] = spatjoin2["Ongev/len"].apply(lambda x:
round(x, 2))
spatjoin2["Verkeerslicht/len"] =
spatjoin2["Verkeerslicht/len"].apply(lambda x: round(x, 2))
spatjoin2["Stop/len"] = spatjoin2["Stop/len"].apply(lambda x: round(x,
2))
spatjoin2["Brug/len"] = spatjoin2["Brug/len"].apply(lambda x: round(x,
2))
spatjoin2["Misdaad/len"] = spatjoin2["Misdaad/len"].apply(lambda x:
round(x, 2))
spatjoin2.reset_index(level = 0, inplace = True)
spatjoin3 = spatjoin2.loc[:, ["Route", "Ongev/len", "Stop/len",
"Brug/len", "Misdaad/len"]]

Misdaad = final_df.loc[:, ["Route", "Criminalit", "Routeperc",
"Total_Length"]]
Misdaad["Misdaad"] = Misdaad["Criminalit"] * Misdaad["Routeperc"]
Misdaad["Misdaad"] = Misdaad["Misdaad"].convert_dtypes()
Misdaad["Misdaad"] = Misdaad["Misdaad"].astype("float")
Misdaad
Misdaadtable = Misdaad.groupby("Route").mean()
Misdaadtable.to_csv("Misdaadtable.csv")

Verlichting = final_df.loc[:, ["Route", "VERLICHT2", "Routeperc"]]
Verlichting["Niet_Verlicht"] = np.where(Verlichting["VERLICHT2"] == 1,
1 * Verlichting["Routeperc"], 0)
Verlichting["Verlicht"] = np.where(Verlichting["VERLICHT2"] == 2 | 3, 1
* Verlichting["Routeperc"], 0)
```

```
Verlichting2 = Verlichting.loc[:, ["Route", "Niet_Verlicht",
"Verlicht"]].groupby("Route").sum()
Verlichting2.insert(0, "Alternative", "Observed")
Verlichting2.to_excel("ObservedVerlichting.xlsx")

DonkerlichtWeer = pd.read_csv('SpatJoin/Donkerlicht_Weer.txt', sep=",")

result2 = pd.merge(result, spatjoin3, on =
"Route").merge(DonkerlichtWeer, on = "Route")
result2.insert(3, "Alternative", "Observed")

result2.to_excel("MergeObservedRoutes.xlsx")
```

```
import pandas as pd
import numpy as np
Import os

Fiets = pd.read_csv('Fiets2.csv', delimiter=";")
Fiets.set_index('TrackerID', inplace=True)

OD = pd.read_csv('ODdata.csv', delimiter=",")
del OD["Unnamed: 0"]
OD.set_index('Route', inplace=True)

Round = pd.read_csv('Round.csv', delimiter = ";")
Round.set_index('Route', inplace=True)

WorkPC = pd.read_csv('WorkPC2.csv', delimiter = ";")
WorkPC["TrackerID"] = WorkPC["Tracker_ID"]#.replace({"Tracker_ID":
"TrackerID"})
del WorkPC["Tracker_ID"]
WorkPC.set_index('TrackerID', inplace=True)

df = pd.read_csv('End.csv', delimiter = ";", error_bad_lines=False)
df.set_index(['Route', 'TrackerID'], inplace=True)

final_df = df.join(OD, how="outer")
final_df2 = final_df.join(Round, how="outer")
final_df2.reset_index(level = 0, inplace = True)

newdataframe = pd.merge(final_df2, Fiets, on="TrackerID", how="outer")

final = pd.merge(newdataframe, WorkPC, on ="TrackerID", how="outer")
final.set_index('Route', inplace=True)
final.reset_index(level = 0, inplace = True)

final.to_csv("Goal.csv")

import pandas as pd
import numpy as np
import os
```

```
Final = pd.read_csv('Goal.csv', delimiter=",")
del Final["Unnamed: 0"]

Final["G_Rec"] = np.where(Final["Round"] == 1, 1, 0)
Final["G_Com"] = np.where(Final["Postcode_Work"] ==
Final["origin_PC6"], 1, 0)
Final["G_Com2"] = np.where(Final["Postcode_Work"] ==
Final["destination_PC6"], 1, 0)
Final["G_Uti"] = np.where((Final["Origin_shop"] +
Final["Destination_shop"]) > 0, 1, 0)

Conditions = [
     Final['G_Rec'] == 1,
    Final['G_Com'] == 1,
    Final['G_Com2'] == 1,
    Final['G_Uti']  == 1
    ]

outputs = [
     'Recreational', 'Commute', 'Commute', 'Utilitarian'
]

New = np.select(Conditions, outputs, 'Other')
df = pd.DataFrame (New)
df["Purpose"] = df[0]
del df[0]

Result = Final.join(df, how = "outer")

Conditions2 = [
     Result['Purpose'] == 'Recreational',
    Result['Purpose'] == 'Commute',
    Result['Purpose'] == 'Utilitarian',
    Result['Purpose']  == 'Other'
    ]

outputs2 = [
     Result['Recreatief'], Result['Commute'], Result['Utilitarian'],
Result['Other']
]

New2 = np.select(Conditions2, outputs2)
df2 = pd.DataFrame (New2)
df2["Bicycle"] = df2[0]
del df2[0]

Result2 = Result.join(df2, how = "outer")
Result2.to_excel("Eindfile_real.xlsx")
```

6. Path Size Logit Script

```
#This is an example of the calculation for the observed route. This
script is altered to perform the same calculation for the alternative
routes.

import pandas as pd
import numpy as np

import os
filenames = []
print(os.getcwd()) #path waar je nu in zit
for f in os.listdir():
    if f.endswith(".txt"):
        filenames.append(f)
dataframes = [pd.read_csv(f, delimiter=";") for f in filenames]
tracker_df = []

#create file with required information for PSL: link_id, PS length,
Total length
for df, name in zip(dataframes, filenames):
    df.insert(0, "Route", f"{name.split('.')[0]}")
    df.insert(1, "TrackerID", f"{name[:5]}")
    df.insert(2, "numberoflinks", f"{len(df)}")
    tracker_df.append(df)
final_df = pd.concat(tracker_df).reset_index(drop=True)
final_df["Shape_Le_1"] = final_df["Shape_Le_1"].convert_dtypes()
final_df["Shape_Le_1"] = final_df["Shape_Le_1"].str.replace(",",
".").astype("float").apply(lambda x: round(x, 2))


y = final_df.loc[:, ["Route", "Shape_Le_1", "link_id"]]
y.insert(1, "Alternative", "Observed")

x = y.groupby(["Route"]).sum()
x["Tot_Leng"] = x["Shape_Le_1"]
del x["Shape_Le_1"]

z = pd.merge(y, x, on = "Route")
z["PS_Length"] = z["Shape_Le_1"] / z["Tot_Leng"]
z.to_csv("PS_Observed.csv")
```

```
import pandas as pd
import numpy as np

#import preparation files all alternatives
Check = pd.read_csv('Check.csv', sep=";")
Check.set_index("Route", inplace = True)

Obs = pd.read_csv('PS_Observed.csv', sep=",")
del Observed["Unnamed: 0"]
Continuous = pd.read_csv('Continuous/PS_Continuous.csv', sep=",")
del Continuous["Unnamed: 0"]
```

```
Fastest = pd.read_csv('Fastest/PS_Fastest.csv', sep=",")
del Fastest["Unnamed: 0"]
Green = pd.read_csv("Groen/PS_Green.csv", sep=",")
del Green["Unnamed: 0"]
Shortest = pd.read_csv("Shortest/PS_Shortest.csv", sep=",")
del Shortest["Unnamed: 0"]
SocSafe = pd.read_csv("SocSafe/PS_SocSafe.csv", sep=",")
del SocSafe["Unnamed: 0"]
TrSafe = pd.read_csv("TrSafe/PS_TrSafe.csv", sep=",")
del TrSafe["Unnamed: 0"]

#Create routelinks all alternatives
Obs["routelink"] = Obs["Route"].astype(str) + " " +
Obs["link_id"].astype(str)
Continuous["routelink"] = Continuous["Route"].astype(str) + " " +
Continuous["link_id"].astype(str)
Fastest["routelink"] = Fastest["Route"].astype(str) + " " +
Fastest["link_id"].astype(str)
Green["routelink"] = Green["Route"].astype(str) + " " +
Green["link_id"].astype(str)
Shortest["routelink"] = Shortest["Route"].astype(str) + " " +
Shortest["link_id"].astype(str)
SocSafe["routelink"] = SocSafe["Route"].astype(str) + " " +
SocSafe["link_id"].astype(str)
TrSafe["routelink"] = TrSafe["Route"].astype(str) + " " +
TrSafe["link_id"].astype(str)

#Calculation if Observed segment overlaps with alternative
Obs["Continuous"] =
np.where(Obs["routelink"].isin(Continuous["routelink"]) &
Check["Continuous"] == 1, 1, 0)
Obs["Fastest"] = np.where(Obs["routelink"].isin(Fastest["routelink"]) &
Check["Fastest"] == 1, 1, 0)
Obs["Green"] = np.where(Obs["routelink"].isin(Green["routelink"]) &
Check["Green"] == 1, 1, 0)
Obs["Shortest"] = np.where(Obs["routelink"].isin(Shortest["routelink"])
& Check["Shortest"] == 1, 1, 0)
Obs["SocSafe"] = np.where(Obs["routelink"].isin(SocSafe["routelink"]) &
Check["SocSafe"] == 1, 1, 0)
Obs["TrSafe"] = np.where(Obs["routelink"].isin(TrSafe["routelink"]) &
Check["TrSafe"] == 1, 1, 0)

#Calculation of overlap percentage
Overlap = Obs.groupby("Route").sum()
Overlap = Overlap.drop("T09_1_04")
Overlap = Overlap.drop("T09_1_05")
Overlap = Overlap.drop("T15_2_03")
Overlap = Overlap.drop("T12_4_05")
Overlap = Overlap.drop("T17_4_02")
Overlap.iloc[:, -6:] = Overlap.iloc[:, -6:].apply(lambda x: round(x,
2))
Overlap.to_excel("Overlap.xlsx")

#Path Size Logit Calculation
#Link sum
Obs["LinkSum"] = 1 + Obs["Continuous"] + Obs["Fastest"] + Obs["Green"]
+ Obs["Shortest"] + Obs["SocSafe"] + Obs["TrSafe"]
```

```
#path size for individual segment
Obs["PathSize"] = Obs["PS_Length"] * (1 / Obs["LinkSum"])

#path size for complete route
Pathsize = Obs.groupby("Route").sum()

Pathsize = Pathsize.drop("T09_1_04")
Pathsize = Pathsize.drop("T09_1_05")
Pathsize = Pathsize.drop("T15_2_03")
Pathsize = Pathsize.drop("T12_4_05")
Pathsize = Pathsize.drop("T17_4_02")
Pathsize.insert(0, "Alternative", "Observed")
Pathsize.to_csv("Pathsize_new.csv")

Obs["Continuous_Perc"] = Obs["PS_Length"] * Obs["Continuous"]
Obs["Fastest_Perc"] = Obs["PS_Length"] * Obs["Fastest"]
Obs["Green_Perc"] = Obs["PS_Length"] * Obs["Green"]
Obs["Shortest_Perc"] = Obs["PS_Length"] * Obs["Shortest"]
Obs["SocSafe_Perc"] = Obs["PS_Length"] * Obs["SocSafe"]
Obs["TrSafe_Perc"] = Obs["PS_Length"] * Obs["TrSafe"]
```

7. Determining Trip Purpose and Bicycle Type Script

```
import pandas as pd
import numpy as np

import os
filenames = []
print(os.getcwd()) #path waar je nu in zit
for f in os.listdir():
    if f.endswith(".txt"):
        filenames.append(f)
dataframes = [pd.read_csv(f, delimiter=";") for f in filenames]
tracker_df = []
for df, name in zip(dataframes, filenames):
    df.insert(0, "Route", f"{name[3:-4]}")
    df.insert(1, "ODPair", f"{name.split('.')[0]}")
        tracker_df.append(df)
final_df = pd.concat(tracker_df).reset_index(drop=True)

final_df["BUFF_DIST"] = final_df["BUFF_DIST"].convert_dtypes()
final_df["BUFF_DIST"] = final_df["BUFF_DIST"].str.replace(",",
".").astype("float").astype("int").apply(lambda x: round(x, 2))

selection = final_df.loc[:, ["Route", "ODPair", "FID", "PC6", "shop",
"BUFF_DIST"]]
selection["OorD"] = selection["FID"] + selection["BUFF_DIST"]
selection["Origin_shop"] = np.where((selection["OorD"] == 50), 1, 0)
selection["Destination_shop2"] = selection["OorD"] - 50
selection["Destination_shop"] = np.where(selection["Destination_shop2"]
== 1, 1, 0)

PC = selection.loc[:, ["Route", "Origin_shop",
"Destination_shop"]].groupby("Route").sum()
origindata = pd.read_csv('Selection/origindata.txt', sep=" ")
originPC6 = pd.read_csv('Selection/originPC6.txt', sep=" ")
destinationPC6 = pd.read_csv('Selection/destinationPC6.txt', sep=" ")
origindata["origin_PC6"] = originPC6.loc[:, "origin_PC6"]
origindata["destination_PC6"] = destinationPC6.loc[:,
"destination_PC6"]

ODdata = pd.merge(origindata, PC, on = "Route")
ODdata.to_csv("ODdata.csv")
```

```
import pandas as pd
import numpy as np
import os

Fiets = pd.read_csv('Fiets2.csv', delimiter=";")
Fiets.set_index('TrackerID', inplace=True)

OD = pd.read_csv('ODdata.csv', delimiter=",")
del OD["Unnamed: 0"]
OD.set_index('Route', inplace=True)

Round = pd.read_csv('Round.csv', delimiter = ";")
```

```
Round.set_index('Route', inplace=True)

WorkPC = pd.read_csv('WorkPC2.csv', delimiter = ";")
WorkPC["TrackerID"] = WorkPC["Tracker_ID"]#.replace({"Tracker_ID":
"TrackerID"})
del WorkPC["Tracker_ID"]
WorkPC.set_index('TrackerID', inplace=True)

df = pd.read_csv('End.csv', delimiter = ";", error_bad_lines=False)
df.set_index(['Route', 'TrackerID'], inplace=True)

final_df = df.join(OD, how="outer")
final_df2 = final_df.join(Round, how="outer")
final_df2.reset_index(level = 0, inplace = True)

newdataframe = pd.merge(final_df2, Fiets, on="TrackerID", how="outer")
final = pd.merge(newdataframe, WorkPC, on ="TrackerID", how="outer")
final.set_index('Route', inplace=True)
final.reset_index(level = 0, inplace = True)

final.to_csv("Goal.csv")
```

```
import pandas as pd
import numpy as np
import os

Final = pd.read_csv('Goal.csv', delimiter=",")
del Final["Unnamed: 0"]

Final["G_Rec"] = np.where(Final["Round"] == 1, 1, 0)
Final["G_Com"] = np.where(Final["Postcode_Work"] ==
Final["origin_PC6"], 1, 0)
Final["G_Com2"] = np.where(Final["Postcode_Work"] ==
Final["destination_PC6"], 1, 0)
Final["G_Uti"] = np.where((Final["Origin_shop"] +
Final["Destination_shop"]) > 0, 1, 0)

Conditions = [
     Final['G_Rec'] == 1,
    Final['G_Com'] == 1,
    Final['G_Com2'] == 1,
    Final['G_Uti']  == 1
    ]

outputs = [
     'Recreational', 'Commute', 'Commute', 'Utilitarian'
]

New = np.select(Conditions, outputs, 'Other')
df = pd.DataFrame (New)

df["Purpose"] = df[0]
del df[0]

Result = Final.join(df, how = "outer")
```

```
Conditions2 = [
     Result['Purpose'] == 'Recreational',
    Result['Purpose'] == 'Commute',
    Result['Purpose'] == 'Utilitarian',
    Result['Purpose']  == 'Other'
    ]

outputs2 = [
     Result['Recreatief'], Result['Commute'], Result['Utilitarian'],
Result['Other']
]

New2 = np.select(Conditions2, outputs2)
df2 = pd.DataFrame (New2)

df2["Bicycle"] = df2[0]
del df2[0]

Result2 = Result.join(df2, how = "outer")

Result2.to_excel("Eindfile_real.xlsx")
```

8. Adding Travel Time and Continuousness Script

```
#This is an example of the calculation for the observed route. This
script is altered to perform the same calculation for the alternative
routes.


import csv
import numpy as np
import pandas as pd

O = pd.read_csv('4MeiObserved.txt', sep = ";")
O.insert(0, "Route", O["Name"].str.replace("OD_", ""))

O2 = O.loc[:, ["Route", "Total_Trav", "Total_Mini"]]
O2.set_index('Route', inplace = True)

OR = pd.read_csv('Rest_Observed.txt', sep = ";")
OR["Route"] = OR["Name"]

OR2 = OR.loc[:, ["Route", "Total_Trav", "Total_Mini"]]
OR2.set_index('Route', inplace = True)

Observed = pd.concat([O2, OR2])
Observed[["Total_Trav", "Total_Mini"]] = Observed[["Total_Trav",
"Total_Mini"]].convert_dtypes()
Observed[["Total_Trav", "Total_Mini"]] = Observed[["Total_Trav",
"Total_Mini"]].apply(lambda x: x.str.replace(",", "."))
Observed[["Total_Trav", "Total_Mini"]] = Observed[["Total_Trav",
"Total_Mini"]].apply(lambda x: x.astype("float")).apply(lambda x:
round(x, 2))
Observed.insert(0, "Alternative", "Observed")

frames = [Observed, Continuous, Fastest, Green, Shortest, SocSafe,
TrSafe]
result = pd.concat(frames)
result.reset_index(level=0, inplace=True)
result["Join"] = result["Route"].astype(str) + " " +
result["Alternative"].astype(str)

Final = pd.read_csv('Finaldata_routes.csv', sep = ';')
Final["Join"] = Final["Route"].astype(str) + " " +
Final["Alternative"].astype(str)

Endfile = pd.merge(Final, result, on="Join", how='left')
Endfile.to_excel("Final_Endfile.xlsx")
```