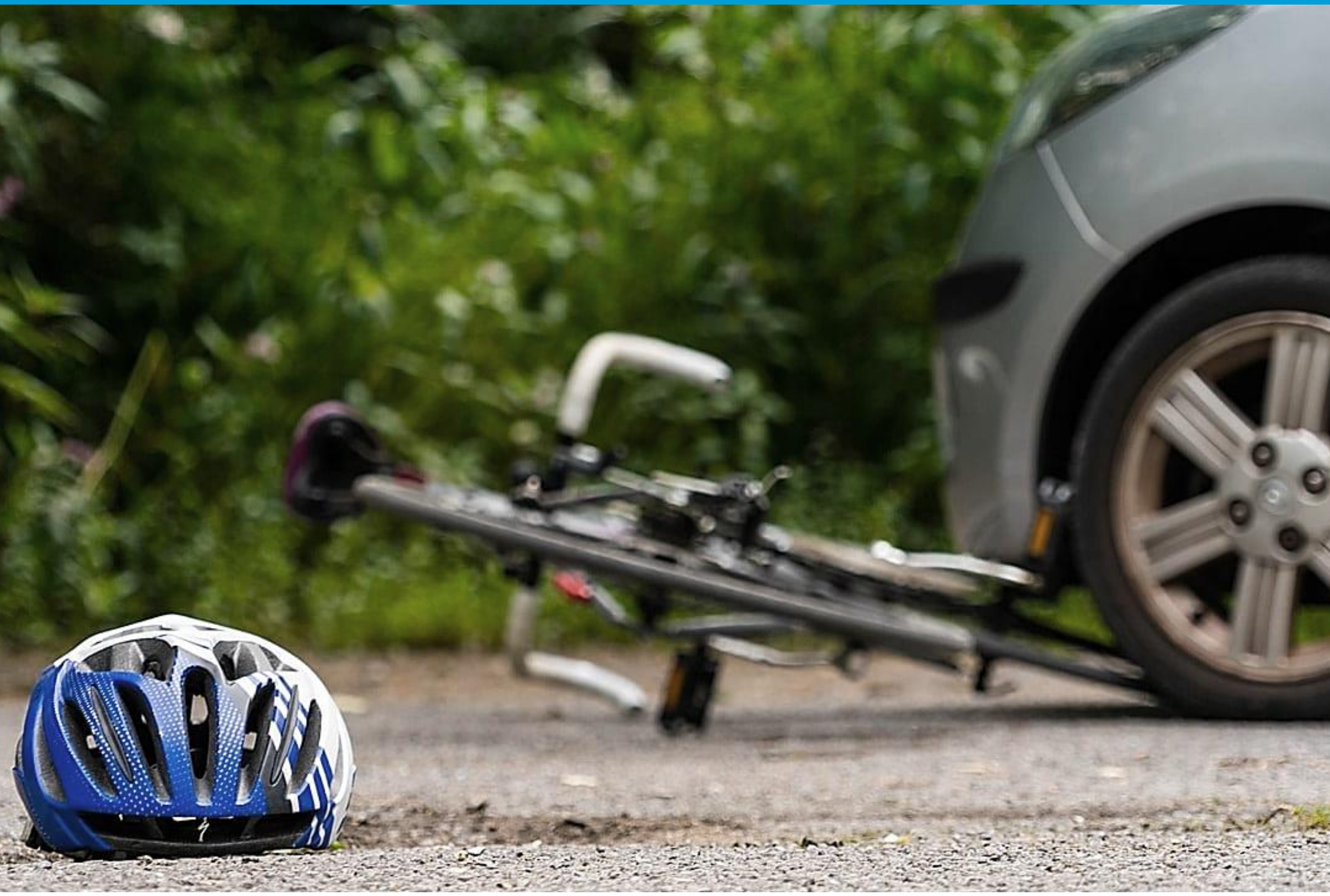
Understanding the Recent Trends of Cyclist Crashes in the Netherlands

Quantitative Analysis of Recent Cycling Crashes in Four Major Dutch Cities and Analysis of Cycling Trends in the Netherlands

Jarno Jorritsma



TU Delft Delft University of Technology

Understanding the Recent Trends of Cyclist Crashes in the Netherlands

Quantitative Analysis of Recent Cycling Crashes in Four Major Dutch Cities and Analysis of Cycling Trends in the Netherlands

By Jarno J.L. Jorritsma

To obtain the degree of Bachelor of Science
at the Delft University of Technology

Faculty of Civil Engineering and Geosciences
Delft University of Technology
BSc Thesis CTB3000-16
Student: 5331072
Delft, June 2024

Examiners:
Dr. Y. Yuan, TU Delft
Dr. S. Sharif Azadeh, TU Delft

Supervisor:
Dr. K. Kavta, TU Delft

External supervisor:
Dr. L. Bisht, TU Delft

Keywords:
Cycling crashes, crash patterns, cycling safety, single-bicycle crashes, cycling crash analysis

Figure on cover: RedactieKW (2022)

TU Delft

Preface

This report presents the results of my Bachelor of Science end thesis project in Civil Engineering, in the domain of Transport and Planning at the Delft University of Technology. The report contains information about how to understand the cycling crashes in recent years in the Netherlands.

This report can be read by anyone interested in Transport and Planning. No prior knowledge is required to understand it.

The work would not have been possible without the support of others. Therefore, special thanks are given to Yufei Yuan, Kuldeep Kavta, Shadi Sharif Azadeh, and Laxman Bisht for providing me information, data, answering my questions, and their guidance during the writing of this report. I also would like to thank my fellow students M. el Aazizi, B. Nauta, C. Polet, and D. van de Brug for reviewing my work weekly.

Delft, June 2024
Jarno Jorritsma

Summary

More than half of all road traffic deaths occur among vulnerable road users, namely pedestrians and (motor)cyclists. Since 2013, cyclist fatalities and serious injuries in the Netherlands have risen more rapidly than those of other road users, with total road deaths doubling from 19% in 1996 to 39% in 2022. This thesis explores cycling crashes in Amsterdam, Rotterdam, Utrecht, and The Hague to gain insights into the recent cycling crashes in the Netherlands. It will create a foundation to understand the crashes. The research question is:

"What patterns, factors, and variables have contributed to the frequency and severity of cycling crashes in Amsterdam, Rotterdam, Utrecht, and The Hague between 2013 and 2023?"

The rise in electric bike usage since 2013 coincides with the increase in bicycle crashes. However, the risk comparison between electric and regular bikes remains inconclusive. Generally, older cyclists often ride electric bicycles, a group that already has relatively serious injuries (on every bicycle type). This means that electric bicycles are not, on average, more dangerous than regular bicycles. In urban areas, overall bicycle usage is rising due to population growth.

A notable pattern is the increase of single-bicycle crashes compared to non-single crashes. Single crashes are more likely to occur at lower speeds (30 km/h) and on straight road segments rather than at junctions compared with non-single-bicycle crashes. Other key factors contributing to single-bicycle crashes include obstacles, lack of road markings, narrow bike paths, and slippery roads. Distractions and cyclist errors are also notable factors.

All four analysed cities have shown an increase in bicycle crashes, with the highest rate of change in The Hague (122 extra crashes per year). Higher maximum speed on the road and leaving the built-up area increases the likelihood of a serious injury or fatal crash. The 4-way junction type contains the most crashes, especially at 50 km/h roads. However, analysis shows no increased likelihood of a serious injury or fatal accident at junctions compared to straight roads when comparing them to material damage crashes.

While weather conditions do not significantly affect the frequency and severity of crashes resulting in serious injuries, road surface conditions do. However, this relationship does not hold for fatal accidents. Adverse weather conditions may reduce the likelihood of fatal accidents compared to material damage crashes.

Key variables and their correlations reveal that most crashes occur at 4-way junctions with a maximum speed of 50 km/h in cities, which is crucial for infrastructure design. To decrease accident severity, straight roads are more dangerous than junctions, especially roads where passenger cars drive, as most bicycle crashes involve passenger cars in recent years, so this is also a crucial lesson when designing infrastructure.

Contents

List of Figures

List of Tables

1 Introduction

Every year, approximately 1.19 million lives are lost globally due to road crashes. Despite population growth, the total number of traffic fatalities has seen a decrease since 2000. This decrease does not hold for every road user. More than half of all road traffic deaths occur among vulnerable road users, namely pedestrians and (motor)cyclists (WHO, 2023). This trend was also visible in the Netherlands. In recent years, namely since 2013, the rise in cyclist fatalities and serious injuries has increased at a higher rate compared to other road users in the Netherlands. Basically road crashes do get less, but not for cyclists, especially in single-bicycle crashes (Aarts, Bijleveld, Bos, & Decae, 2022). According to official statistics from the SWOV report from 2022, the total road deaths doubled from 19% in 1996 to 39% in 2022 for cyclists in the Netherlands. Nowadays, four out of ten road deaths are cyclists.

The EU has an ambitious long-term goal, to move close to zero traffic deaths by 2050. It is called the "Vision Zero"-plan. The first goal of the plan is to halve the number of serious injuries in the EU by 2030 compared to 2020. (EuropeanCommission, 2020). This plan holds for every road user, but as stated before the amount of cycling crashes is rising, so understanding how and why those crashes are happening is important to move towards the goals of the EU to create more safety on the roads and decrease road fatalities.

1.1 Scope and Research Question

This thesis aims to gain insights into the recent cycling crashes in the Netherlands and to understand why the crashes are happening. It is hard to understand all types of cycling crashes in a whole country. Also, crashes might be completely different when comparing crashes in the city and on a farm road. Therefore, this study will focus on the four cities with the highest absolute number of cycling crashes in the Netherlands, namely Amsterdam, Rotterdam, Utrecht, and The Hague, to formulate a more uniform conclusions.(SWOV, 2024)

This research will lay the foundation for understanding cycling crashes in the Netherlands. Understanding these crashes is crucial for creating a safe cycling system and achieving zero fatalities by 2050. Insights that will contribute to creating a safer cycling network will be gained through a literature review of cycling crashes and trends. Followed by a data analysis will be conducted about crashes. The main research question of the thesis is as follows:

"What patterns, factors, and variables have contributed to the frequency and severity of cycling crashes in Amsterdam, Rotterdam, Utrecht, and The Hague between 2013 and 2023?"

To answer this question, the following seven sub-questions are formulated:

1. *Which group is most frequently involved in cycling crashes and has there been a change in recent years in the Netherlands?*

   This sub-question examines which type of group and type of person is more frequently involved in bicycle accidents. Since the dataset used in the Results data analysis of this research does not include information on age, gender, and other personal characteristics, this sub-question is essential for seeking more information within the literature on these statistics to better understand this aspect of bicycle accidents. This sub-question will be answered in Chapter 2, the Literature review.

2. *Which trends can be observed about cyclists and their infrastructure in the Netherlands in the last 10 years?*

   This sub-question will look into the current trends in cycling. It will focus on the last decade, namely between 2013 and 2023. Current trends in the literature can provide insights that cannot be concluded from the used data alone. This will help by understanding what influences cycling crashes. This sub-question will be answered in Chapter 2, the Literature review.

3. *What are the primary factors contributing to the safety of cyclists on the roads in general?*

   By answering this sub-question the most frequent factors contributing to the safety of cyclists on the road will become clear. Current literature and studies about cycling safety will highlight important factors that influence crashes in general, thus not specific to the Dutch context. This sub-question will be answered in Chapter 2, the Literature review.

4. *Which type of research design is used in this study, and why is it suitable for answering the research question?* This sub-question is crucial for this report, as the research design is the guideline for this report. The research design will contain information about the data collection, description, and how to use it. So which tests and analysis will be conducted. The design will be discussed in the Methodology in Chapter 3.

5. *What are the most common causes of crashes for cyclists according to the data in Amsterdam, Rotterdam, Utrecht, and The Hague?*

   This sub-question will create an overview of the most common causes of crashes and the severity of those crashes in the four stated cities according to the used dataset within this research. This sub-question will be answered in Chapter 4, the Results data analysis.

6. *Which patterns in the data about cycling crashes in Amsterdam, Rotterdam, Utrecht, and The Hague can be found?*

   By answering this sub-question an overview of crash patterns over the years between 2013 and 2023 will be created. It will focus on the trend of the types of crashes with their frequency and severity per city. This sub-question will be answered in Chapter 4, the Results data analysis.

7. *What lessons can be drawn from the data analysis and what implementation will contribute to increasing safety in the future?*

   By answering this sub-question, the lessons learned from this study will be noted. These lessons can be recommended for increasing cycling safety. These insights may be important for traffic authorities in the four analysed cities. This will be addressed in Chapter 5, which covers the Conclusions and recommendations.

1.2 Reading guide

To answer these research questions, a literature review will first be conducted to acquire existing knowledge and insights in Chapter 2. The methodology in Chapter 3 involves utilizing the database *BRON*, which contains registered traffic accidents in the Netherlands and it contains a detailed description of the used data analysis methods. The results of the data analysis will be published in Chapter 4. Finally, discussion points related to the research will be presented in Chapter 5, followed by the conclusions, recommendations, lessons, and future research in Chapter 6.

2 Literature review

In this chapter, a closer look into the current literature about cycling crashes will be done. The literature will be about the first three stated research sub-questions in the Introduction. Firstly, a closer look into cycling trends will be conducted. Current trends in cycling safety measures and vehicles will gain insights into current cycling crashes.

Next, the review will explore who is most involved in bicycle accidents. This will involve investigating which groups are most frequently involved in these accidents, whether there has been a shift in these patterns, and the underlying reasons for such changes.

Subsequently, it is crucial to get more knowledge on the primary factors that contribute to the safety of cyclists on the road. The data that will be used in the Results data analysis contains a lot of variables. Those variables will be explained in the Methodology. By investigating this it will become clear which variables have to be taken into account and which not for the data analysis.

2.1 Cycling trends in the Netherlands

This paragraph will focus on the recent trends in the last 10 years within cycling traffic.

Residents of the Netherlands make more than a quarter (28%) of all trips in the Netherlands by bicycle. The majority of these are done with the regular bike (20%) and a smaller portion with the electric bike (8%) (Haas & L, 2023). However, looking at the distance over a quarter of the total distance traveled by bicycle was covered by the electric bike; especially cyclists aged 65 and older opt for an electric bike. The electric bike began its rise in 2013 (Haas & Hamersma, 2020). Figure 1 shows the development of the share of bicycle trips with a regular or electric bike from 2012 to 2022.

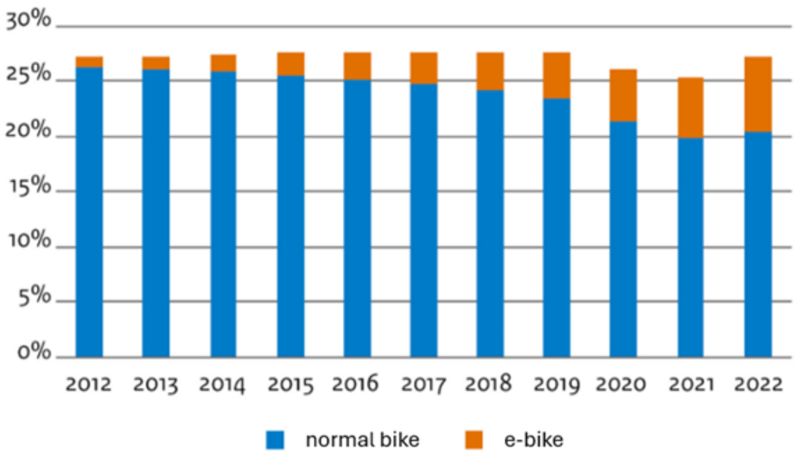


Figure 1: Yearly cyclist for four cities (Boonstra & Brakel, 2023)

As stated before in this paragraph the electric bike has been really popular for people aged 65 and older in the last years. They ride more than half of their cycling kilometers on an electric bike annually. Users of an electric bike covered longer distances on average than users of a regular bike. Currently, the share of cycling kilometers on an electric bike is still growing, especially among cyclists in the age group of 12-50 years. This rise can be seen in Figure 2.

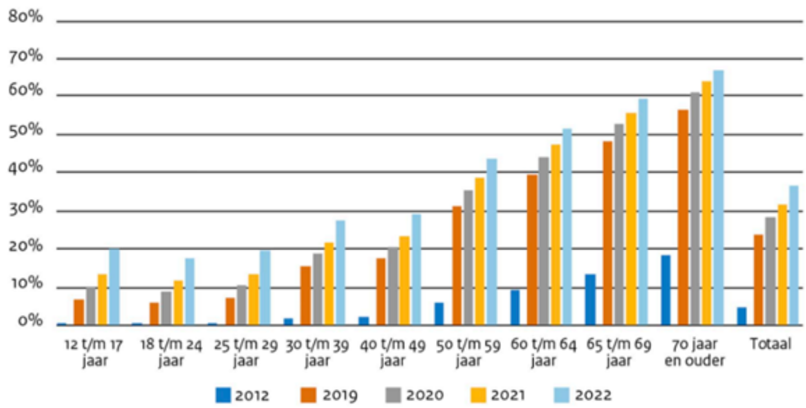


Figure 2: Share of e-bike in total distance covered by bicycle by age (Boonstra & Brakel, 2023).

The share of fatalities riding electric bikes is higher than average for older cycling victims and appears to increase with age (J. Schepers, Weijermars, Boele, Dijkstra, & Bos, 2020). However, there is no consensus within the scientific literature on whether riding an electric bike is riskier than riding a regular bike: some studies report an increased risk for electric bike riders, while others do not (Gogola, 2018a) (Gogola, 2018b) (P. Schepers, Wolt, Helbich, & Fishman, 2020).

So did the Dutch study from VeiligheidNL in 2017 conclude that accidents on electric bicycles cause on average more serious injuries than accidents on a regular bicycle, but corrected for age and gender, this effect disappears. Older cyclists often ride electric bicycles, a group that already has relatively serious injuries (on every bicycle type). This means that the electric bicycle is on average no more dangerous than a normal bicycle (Valkenberg et al., 2017).

Another trend that can be seen in the last ten years is the mode shift within the traveling patterns for citizens. There was an increase in bicycle usage among people living in (very) strongly or moderately urbanized areas. They cycle relatively more than people from non-urbanized areas. This is partly due to the distances to destinations (Haas & L, 2023) Additionally, there has been a significant increase in bicycle usage in all cities in the Netherlands from 2010 to 2019. Between 2012 and 2019, the total distance traveled by bicycles by Dutch citizens increased by almost 8%. An important explanation for this increase is population growth. In Figure 3 the increase of bicycle usage can be seen in some cities in the Netherlands (Haas & Hamersma, 2020). The four cities that will be used in the Results data analysis are put in the front.

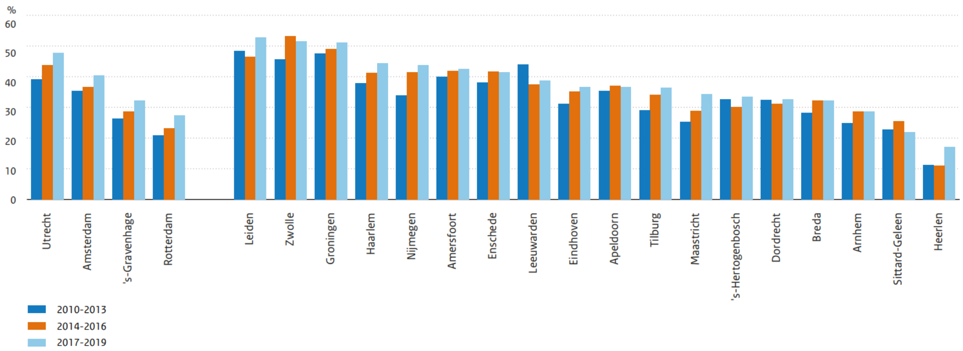


Figure 3: Share of trips made by bicycle in the city (Haas & Hamersma, 2020).

2.2 Cycling crashes in the Netherlands

This paragraph describes statistics of the group of victims of cycling accidents. Risk groups are identified, and a distinction is made between single and multiple accidents. As stated in the Introduction, official statistics from the SWOV report from 2022 the total road deaths doubled from 19% in 1996 to 39% in 2022 for cyclists in the Netherlands. Four out of ten road deaths are cyclists nowadays.

Looking at those crashes via a demographic view it can be observed that the number of people aged 70 and older has increased in recent years. In 2021, 57 percent (119) of cycling fatalities were 70 years or older. In 2010, 49 percent of cycling fatalities were 70 or older. Per 100,000 inhabitants, the number of cycling fatalities among those aged 70 and older was higher in 2021 than in 2010 (4.8 and 4.5, respectively) (CBS, 2022). In the previous paragraph, the rise of the electric bike was highlighted, but studies are not unanimous about the relation between the rise of accidents among people 70 years and older and the rise of electric bikes. The increase in cycling fatalities among people aged 70 and older is mostly explained by the aging population in the Netherlands.

Looking at the gender distribution of bicycle accidents in the Netherlands, the statistics show that men do cycle more than women, but women cover a larger portion of their mobility by bicycle compared to men. Additionally, men are more frequently involved in accidents than women.

2.3 Safety factors of cyclist

Cycling and participating in traffic will always carry a risk. Therefore, the European Union's plan to achieve zero cycling fatalities by 2050 is very ambitious. However, some factors and measures can be taken to closely approach this goal. Bicycle accidents are often the result of a combination of vehicle, road, and behavioral factors. Both road design and road quality play a role in the occurrence of bicycle accidents without the involvement of a motor vehicle, as well as in single-bicycle crashes. In this paragraph, different studies about primary factors that contribute to cycling safety will be analysed.

Obstacles on the road and narrow or slippery bike paths are infrastructural features that can lead to accidents. Cyclist behavior, such as cycling when drunk, ignoring red lights, texting while cycling, and riding without proper lighting, can contribute to an accident. Also, the behavior of other traffic participants plays a role in causing an accident (SWOV, 2023). Ultimately, human errors will always occur, but the behavior that can lead to an error can be influenced. Research indicates that in 2018, two-thirds (68%) of bicycle accidents resulting in emergency hospital visits were single-bicycle incidents (Krul, 2018), meaning no other road users were involved. Research by SWOV in 2023 identifies the following most frequently mentioned factors leading to single-bicycle crashes. Important to mention is that the following four factors are general factors worldwide, so not for the Netherlands only:

1. Obstacles (Krul, Valkenberg, Asscherman, Stam, & Klein Wolt, 2022)
2. Lack of road markings (Algurén & Rizzi, 2022)
3. Narrow bike paths/roads (Boele-Vos et al., 2017)
4. Slippery roads due to weather conditions (Utriainen, 2020)

Other research has found out cycling near an intersection increases the risk of a critical event by four times compared to cycling on a wide-open road on your own. This risk rises to twelve times if the intersection has visual obstructions, such as buildings, hedges, etc. Poor road maintenance increases the risk tenfold. And lastly, the risk of a critical event doubles when a pedestrian or another cyclist crosses the cyclist's path (Dozza & Werneke, 2014). Two studies from the Netherlands confirm the importance of fellow road users. The first studies stated that there are 50-60% fewer accidents on separated bike paths than on non-separate bike lanes (van Petegem, Schepers, & Wijlhuizen, 2021). The second study from 2016 examined the factors of bicycle accidents based on a survey of cyclists who ended up in the emergency department in the Netherlands. 36% of the respondents said their crash was caused by the behaviour of other road users. Other important named causes of the respondent's crash are: Own mistake (44%), state of the road (34%), weather circumstances (20%), and distracted (19%) (Valkenberg et al., 2017).

In Figure 4 the most named causes are visualised in a graph. Those factors are purely based on people who ended up in the hospital. So these factors may be different for people who crashed and only had material damage.

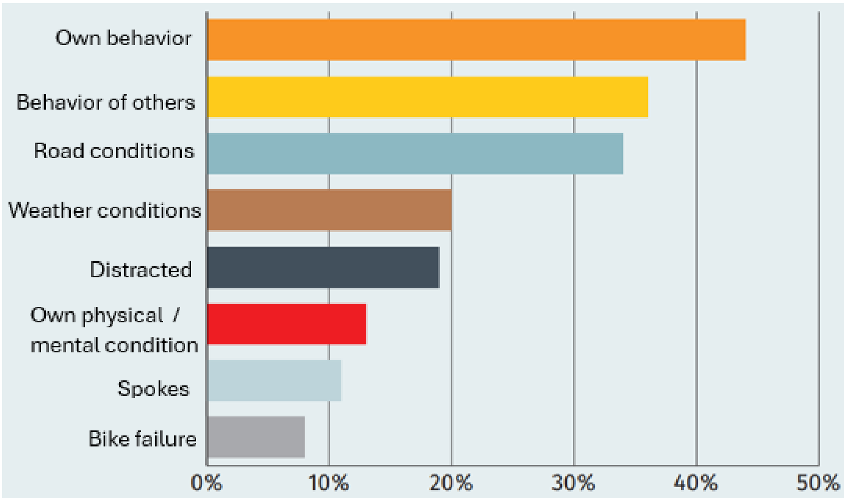


Figure 4: Causes of bicycle accidents in the Netherlands from emergency care in 2016 (Valkenberg et al., 2017)

2.4 Research gaps

In this paragraph, the focus will be on the research and knowledge gaps that currently exist in the literature or on research where limited information is available. These knowledge gaps will be used in the data analysis of this study.

Firstly, there is knowledge about the four most common causes of single-bicycle crashes, but these causes are considered from a global perspective and not specific to the Netherlands. Therefore, it is interesting to examine single-bicycle crashes in the Netherlands to track down the causes of crashes in the Netherlands for single-bicycle crashes.

Additionally, the literature mentions the increased risk of an intersection compared to a road without an intersection. However, there is no knowledge about which types of intersections have the most accidents. Consider junctions with 2, 3, or 4-way splits and roundabouts. There was also no knowledge about the frequency and severity of accidents related to the maximum speed limit on the road.
These research gaps from the literature review will be included in Chapter 4 Results data analysiss next to the other stated points in Section 3.3 the Data analysis method.

3 Methodology

In this chapter, the methodology of the research is explained. It will describe how the (sub-)questions, stated in the Introduction, will be answered. To tackle the first three sub-questions a literature analysis based on the current literature available on cycling crashes was conducted. The sub-question about the research design will become clear in this chapter. The other three sub-questions will be addressed through data analysis in Chapter 4.

3.1 Literature analysis

The first part of the research consisted of a Literature review. With this analysis, the first three sub-questions were answered.

To start with, the research on the current trends within cycling and the infrastructure. This has shown that there was a change in mode shift in the last years and a change of type of bike and their users. This information might help by interpreting the results of the data.

Next, a closer look into which group is most frequently involved in cycling crashes in the Netherlands. Available statistics did highlight which group is mostly involved and it also showed that there is a rate of change within this group in the last years. Since the used data does not contain personal information, c.q. age and gender, etc. This information had to be found in the literature.

Lastly, the third sub-question was answered using the most common causes of crashes according to different studies. It highlighted that there are different causes for single crashes and crashes with more traffic users involved. So a closer look will be conducted in the data analysis to see which factors influence single-bicycle crashes. Furthermore, the Literature review showed the primary factors that contribute to the safety of cyclists.

3.2 Data collection and description

The used data in this research is from *Bestand geRegistreede Ongevallen Nederland* (BRON) (SWOV, 2024). The BRON database includes all traffic accidents in the Netherlands that have been documented by the police or road inspectors from Rijkswaterstaat. This can be done through characteristic reports or incident reports. In reality, many more accidents occur, but the police are not called to all accidents, and lighter accidents may not be recorded as per protocol. BRON annually registers approximately 200.000 accidents for all traffic modes. This data is an open-source provided by the Dutch government (Rijkswaterstaat, 2024). In Overview dataset + all variables A an overview of the database is visualised. On the left side, the different variables can be seen. Under the figure in the appendix, all the variables are noted.

The data is presented in Text files. SWOV, a Dutch institute for scientific research on road safety, has already converted these text files from BRON into Excel spreadsheets. That converted data will be used for this research. The data consists of a lot of information about all the registered road crashes in the Netherlands from 1987 to 2023. In total, there are 5,807,619 registered crashes. It is important to note that not all crashes contain the same level of detailed information, and also the method of recording accidents has evolved over the years. So this makes some data useless. This will be discussed in the paragraph about the Operationalization of variables.

3.3 Data analysis method

In this paragraph, a description of the type of data analysis will be presented and explained. There will be three different tests/analyses, those analyses will be done with the use of Python. First, to start with the general patterns, then a statistical test to look at the correlation between two variables. The chi-square test will be used for this. Lastly, a multinomial logistic regression analysis will be conducted to test the correlation and relation between more than 2 variables. The three types of analysis will be explained in this section.

3.3.1 Cycling crash patterns

First, in the Results data analysis patterns that can be extracted from the data are examined. These patterns between variables are not correlated with each other. The focus is purely on the increase or decrease of a single variable over the analyzed period, namely 10 years. The found results will be presented in tables and graphs. Based on this, conclusions will be drawn about aspects that do occur more often in bicycle accidents nowadays and if there is a rate of change, so a prediction for the future can be made based on the data of the last years. In this way, important variables of the last years will become clear and useful for later tests in the Data analysis.

The following points will be looked at for bicycle crash patterns:

1. **Crashes based on the severity level**
   This point will focus on the severity of the crashes between 2013 and 2022. A closer look will be taken into it to see if there is an increase or decrease within the three possible severities, namely death, injury, or material damage. The needed variables for this point are year, place name, bicycle (involvement) & severity of accident. By creating an overview of the severity of the crashes each year this point can be answered.

2. **City-wise variation**
   This point will look at the variation of cycling crashes over the last 10 years, between 2013-2022. The needed variables for answering this point are year, bicycle (involvement) & place name. The hypothesis is that in each city a rise will be visible because there was a general rise in cycling crashes according to the whole country. This is done to see if a city does show different stats and therefore might need to be investigated more.

3. **Single-bicycle crash pattern**
   This point will look at single-bicycle crashes, so no other traffic road user is involved. The needed variables for this point are year, place name, type of accident, lightning conditions, location (road Segment or intersection), maximum speed (road), road surface condition, road situation, and bicycle (involvement). The type of accident can be filtered on single-bicycle crashes. Then the number of single crashes will be analysed based on lightning conditions, location (road or intersection), maximum speed (road), road surface condition, and road situation. These results from the single crashes will be compared with all types of bicycle crashes with the same variables. Then a conclusion can be drawn if single-bicycle crashes do occur more often with which variable and in which situation.

4. **Crashes vs other modes of vehicles/pedestrians**
   This point will create an overview of which mode of traffic has had the most accidents with a cyclist involved. The needed variables are year, place name, bicycle (involvement), delivery van (involvement), scooter (involvement), bus (involvement), bicycle (involvement), agricultural vehicle (involvement), motorcycle (involvement), passenger car (involvement), pedestrian (involvement), truck (involvement). With the test, a deeper look into crashes with other modes in the four cities will be acquired, and from that conclusions can be drawn.

5. **Crashes for the road section with speed limits**
   This point will first look at which maximum speeds for the roads the most bicycle crashes do occur. Later on, the severity will be taken into account, to see if there is a correlation between severity and maximum speed on a road. The needed variables are year, place name, bicycle (involvement), and maximum speed (road).

6. **Crashes based on the road section type**
   This point will look at the type of road section where the crashes occurred, namely normal road segments or intersections between 2013 and 2022. Also, the severity of the crash will be taken into account. The needed variables are year, place name, bicycle (involvement) & location (road Segment or intersection). As stated in the Literature review cyclists tend to have more chance to crash at an intersection than on a Road Segment. This point will look at whether this also holds for the cyclists in the four cities and what the ratio is.

7. **Crashes based on the junction type**
   This point will focus on crashes on different types of junctions over the years. The previous point highlighted the influence of road section type. Now, a closer look into the junction type will be conducted. The needed variables are year, bicycle (involvement) & road situation. The different analysed road situations are 4-way junction, 3-way junction, corner, straight road, straight road (separated), and roundabout. The literature highlighted the risk of intersections compared to a normal road section but did not mention the type of junction. This point will look at the danger of each junction type. It is expected that more traffic leads to more accidents since the risk of crashes rises when other traffic users are involved as stated in the Literature review. So it is assumed that a 4-way junction or a roundabout will conduct the most crashes since there is coming more traffic from all sides at the junction, so more distraction for the cyclist. By looking at the junction types and their crashes this point can be answered.

3.3.2 Statistical tests

After answering the 7 stated points in the previous section. It will be clear which variables are important. Interesting is to see if some variables show a correlation. Looking at the correlation will give more insights into the relation between the two variables. The statistical test that will be used to test those correlations is the chi-square test. First, the test will be described. Later in this paragraph the correlation between which variables will be described.

The chi-square (χ^2) test is a statistical test used to determine whether there is a significant difference between the observed frequencies and the expected frequencies in a dataset with nominal and or ordinal data. The chi-square test is often used to investigate whether there is an association between two categorical variables. The test works by comparing the observed frequencies to the frequencies that would be expected if there were no associations between the variables (the null hypothesis). If the difference between the observed and expected frequencies is significant enough, the null hypothesis will be rejected, and conclude that there is an association between the variables.

The result of a chi-square test usually includes a χ^2 value and a p-value. The χ^2 value indicates the degree of deviation between the observed and expected frequencies. A higher value suggests a stronger association between the variables. The p-value indicates whether this deviation is significant enough to reject the null hypothesis. The chi-square (χ^2) statistic is calculated as follows:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{1}$$

Where:

- O_i is the observed frequency in category i
- E_i is the expected frequency in category i

The summation runs over all categories into which the data is divided. The expected frequency E_i is calculated as:

$$E_{ij} = \frac{(\sum \text{Observations row } i) \cdot (\sum \text{Observations column } j)}{\sum \text{Observations}} \tag{2}$$

This is calculated for each cell in the contingency table. Next, the χ^2-value is compared to the critical value of the chi-square distribution for a given significance level and degrees of freedom. The formula for degrees of freedom:

$$\text{Degrees of Freedom} = (\text{Number of rows} - 1) \cdot (\text{Number of columns} - 1) \tag{3}$$

When the calculated χ^2 value is greater than the critical value, it indicates a significant relationship between the variables. In general, if the χ^2 value is large, it suggests a large difference between the observed and expected frequencies, indicating a relationship between the variables. If the χ^2 value is small, it suggests a small difference between observed and expected frequencies, indicating no relationship between the variables (Hayes, 2024).

As mentioned in the Introduction, the EU has an ambitious long-term goal to reduce traffic deaths to near zero by 2050. The first objective of this plan is to halve the number of serious injuries in the EU by 2030 compared to 2020. Therefore, it is important to examine the correlation of different variables with the severity of accidents to identify which variables are related to accident severity. Additionally, understanding which variables correlate with the focus on cycling safety is crucial. This knowledge can inform infrastructure design, ensuring a focus on specific factors and variable combinations that enhance safety.

The following variables will be tested on correlation:

1. Severity of accident and road segment
2. Severity of accident and junction type
3. Severity of accident and weather circumstances
4. Severity of accident and maximum speed (road)
5. Severity of accident and lighting conditions
6. Junction type and maximum speed (road)
7. Junction type and lighting conditions
8. Weather circumstances and maximum speed (road)
9. Maximum speed (road) and lighting conditions

3.3.3 Multinomial logistic regression

In this paragraph, a description of a correlation test between more than two variables will be conducted. After finding out all the relevant variables and their inter-correlations that contribute to cycling crashes a multinomial logistic regression analysis will be done. Multinomial logistic regression is a statistical technique used to investigate the relationship between one or more independent variables and a dependent variable with more than two categories. It is often used when the dependent variable has a nominal or ordinal scale with three or more categories. In this case, the dependent variable is on a nominal scale with 3 categories, which makes this a useful method of analysing the relation between the severity of the accident and the junction type. The aim of the analysis is to examine the relationship between the dependent and independent variables. The used variables are listed and explained in Table 1. The dependent variable is the severity of the accident. The independent variable is the junction type. The control variables are inside/outside built-up area, lightning conditions, maximum speed on the road, weather circumstances, and road surface conditions. For an overview of this regression analysis, a conceptual model is created and visualised in Figure 5.

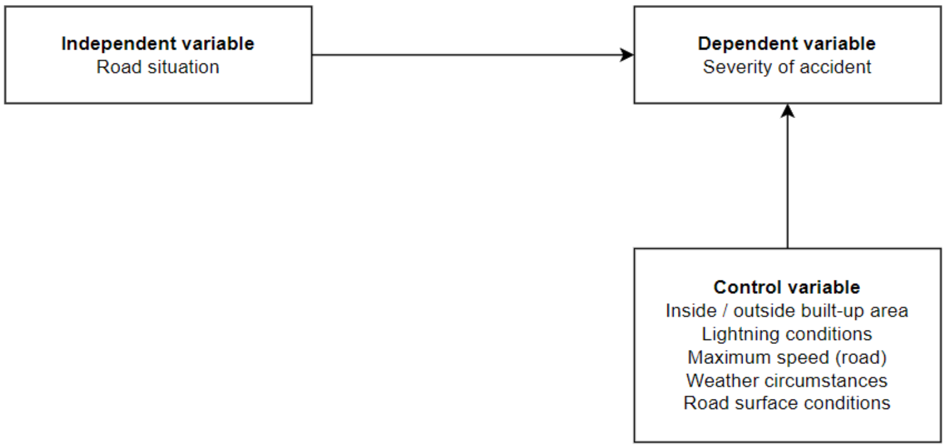


Figure 5: Conceptual Model multinomial logistic regression

3.4 Operationalization of variables

To interpret a model correctly, it is necessary to operationalize the data beforehand so that it can be used in the model. Since it is now clear which types of tests will be conducted, the data can be operationalized so analysis can be done with it. This paragraph will focus on the operationalization of the variables. Some information in the data might be useless since the data also contains information about other crashes without cyclists involved or the level of detail is really low of an accident, which makes it useless for analysis. In total, the dataset consists of 57 variables. Not every variable will be mentioned, only the ones that will be used. In Appendix A all the variables in the dataset are noted.

The variables selected for analyzing the seven points in Section 3.3 Data analysis method are year, severity of accident, place name, type of accident, location (road segment or intersection), maximum speed (road), van (involvement), scooter (involvement), bus (involvement), bicycle (involvement), agricultural vehicle (involvement), motorcycle (involvement), objects (involvement), passenger car (involvement), pedestrian (involvement), truck (involvement). Section 3.3 will specify which variables are needed to answer each point. Some variables will be used for every point, so these variables can already be operationalized.

Firstly, all crashes where a cyclist was not involved will be removed, meaning only incidents where one or more cyclists were involved will be retained. From 1987 to 2023, there are a total of 551,766 registered cycling crashes. However, data from 2023 only includes fatal crashes and lacks records of material damage or injuries. Due to incomplete data for 2023, this year will be excluded from the research.

Firstly, all crashes where a cyclist was not involved will be removed, meaning only incidents where one or more cyclists were involved will be retained. From 1987 to 2023, there are a total of 551,766 registered cycling crashes. However, data from 2023 only includes fatal crashes and lacks records of material damage or injuries. Due to incomplete data for 2023, this year will be excluded from the research. Furthermore, all data before 2013 will be removed because this study focuses on the recent increase in cycling crashes, which has been notable since 2013 as depicted in Figure 6.

It's important to note that data from the years affected by the coronavirus pandemic (2020-2022) will not be deleted. Despite the lockdowns during this period, a significant number of accidents were still recorded in the database. There is no conclusive evidence suggesting that there would have been more or fewer accidents without lockdown measures. Therefore, this data remains valuable for analysis purposes.

As mentioned in the Introduction, understanding all cycling crashes across the Netherlands can be challenging due to significant differences between crashes on rural roads versus those in urban areas. Therefore, this research will specifically focus on the four cities with the highest incidence of cycling crashes. Consequently, all other cities in the dataset will be excluded from the analysis. With this initial operationalization completed, the variables can now be implemented for each point in Section 3.3 Data Analysis Method. Since each point emphasizes different variables, their operationalization and descriptions will be detailed under each respective point in Section 3.3 Data analysis method.

All the operationalized and used variables in the multinomial logistic regression analysis can be seen in Table 1. The reference category of the multinomial logistic regression will be material damage. Furthermore, the speeds 60, 70, 80, and 100 km/h are one category to create more number of cases for the category.

Each category apart had a really low number of cases (30 cases per group) compared to the other groups, namely 15, 30, and 50 km/h (more than 1000 cases per group). In this way, the model will become more accurate and more reliable. The frequencies and number of cases are shown in Table 12.

Table 1: Description of variables

Variable	**Type**	**Description**
Severity of accident	Dependent Nominal	0 = Material damage 1 = Serious injury 2 = Fatal
Road situation	Independent Nominal	0 = Straight road (separated lanes) / corner 1 = Roundabout / 3-way junction / 4-way junction
Inside / outside built-up area	Control Nominal	0 = Inside 1 = Outside
Lightning conditions	Control Nominal	0 = Daylight 1 = Darkness 2 = Twilight
Maximum speed (road)	Control Nominal	0 = 15 km/h 1 = 30 km/h 2 = 50 km/h 3 = 60 / 70 / 80 / 100 km/h
Weather circumstances	Control Nominal	0 = Dry 1 = Raining / fog 2 = Snow
Road surface conditions	Control Nominal	0 = Dry 1 = Wet 2 = Snow / black ice

4 Results data analysis

In this chapter, the data analysis will be conducted as described in the Methodology about the 7 stated points, the statistical tests, and lastly the multinomial logistic regression in Section 3.3 Data analysis method. The results from the data analysis will be published in this chapter in different sections for each of the three tests. All tests and analysis in this research are done with the use of Python.

4.1 Analysis bicycle crash patterns

In this section, the 7 stated points in the Methodology will be conducted. These points will give a first outlook of important variables and their decrease of increase over the last years between 2013 and 2022. All the calculations and Python scripts about the 7 points can be found in Appendix C.

4.1.1 Literature check & crashes based on severity

Firstly, the data from BRON will be rechecked to confirm if the findings align with the Literature review regarding the increase in cycling crashes in recent years in the Netherlands. As depicted in Figure 6, there was a decrease in cycling crashes until around 2012. Since 2013, there has been a notable rise in cycling crashes across the entire Netherlands, validating the information about this trend found in the literature. Therefore, the focus of this research will be on the years 2013 to 2022, as this period reflects the observed increase in cycling crashes.

Figure 6 illustrates the increase in cycling crashes in the Netherlands from 2013 to 2022, with a exception seen between 2019 and 2020, possibly due to reduced traffic during the lockdown period. Therefore, a regression line through the data is made, to see the general increase of the number of bicycle crashes in the Netherlands over the years.

Regarding the severity levels of cycling crashes in Amsterdam, Rotterdam, Utrecht, and The Hague between 2013 and 2022, Figure 7 shows that the majority of crashes are classified as material damage. There are a few deathly crashes, but mostly are serious injury and material damage.

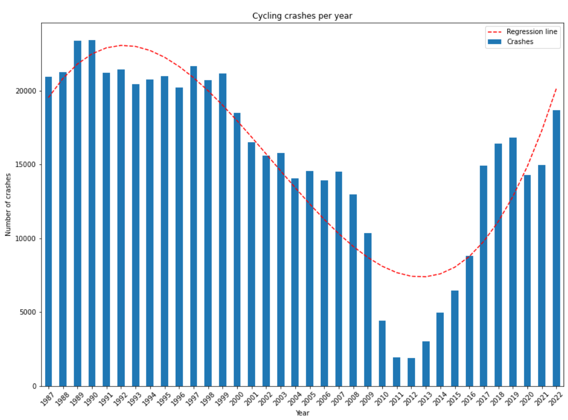


Figure 6: Yearly cyclist crashes from 1987 to 2022.

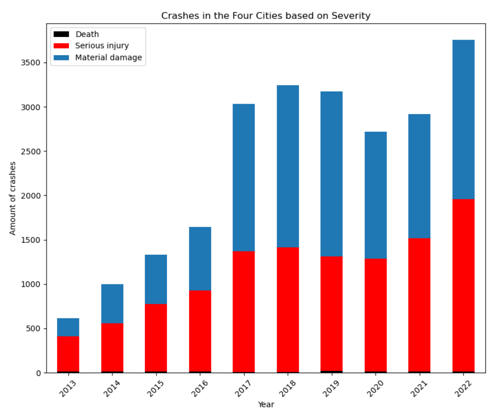


Figure 7: Distribution of Crashes in the Four Cities based on Severity

4.1.2 City wise variation

When looking at the data for the four cities with the most cycling crashes in the Netherlands, namely Amsterdam, Rotterdam, Utrecht, and The Hague a similar rise, as visualised in Figure 6, can be seen. The graph of cycling crashes and the severity of it in the four cities can be seen in Figure 8.

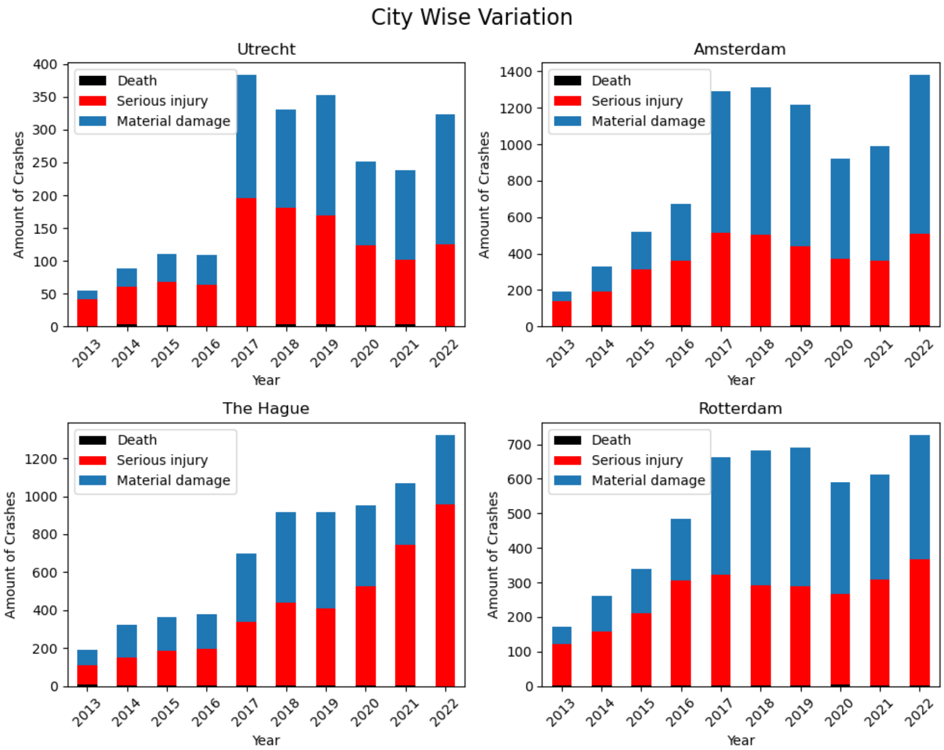


Figure 8: Yearly cyclist crashes from 1987 to 2022.

Looking at the graphs it can be seen that Utrecht and Amsterdam show a similar pattern with a high rise from 2016 to 2017. This rise from 2016 to 2017 is less big for Rotterdam. Furthermore, Utrecht, Amsterdam, and Rotterdam also show a dip from 2019 to 2020, just like in Figure 6. The bicycle crashes in The Hague did rise every year. Looking at a linear regression line the average rate of change between 2013 and 2022 can be observed for each city. Utrecht, Amsterdam, The Hague, and Rotterdam have a respective rate of change of 29.4, 115.2, 56.7, and 122.4 crashes per year. So based on the last 10 years The Hague is the city with the quickest change of rate for bicycle crashes.

4.1.3 Single-bicycle crashes

This section will look at single-bicycle crashes. As stated in the Introduction and Literature review there was an increase of single-bicycle crashes. The most common factors that caused single crashes were: Obstacles, Lack of road markings, Narrow bike paths/roads, and Slippery roads.

In this section, single-bicycle crashes based on Lighting conditions, Location (Road segment or intersection), Maximum speed (road), Road surface condition, and road situation will be analysed.

Firstly, the distribution of single crashes over the last 10 years is visualised in Figure 9. From this graph, it can be obtained that there is a rapid increase in single-bicycle crashes. In 2016 there were 99 single crashes according to the data and in 2022 there were 405 single crashes. This is an increase of 309%.

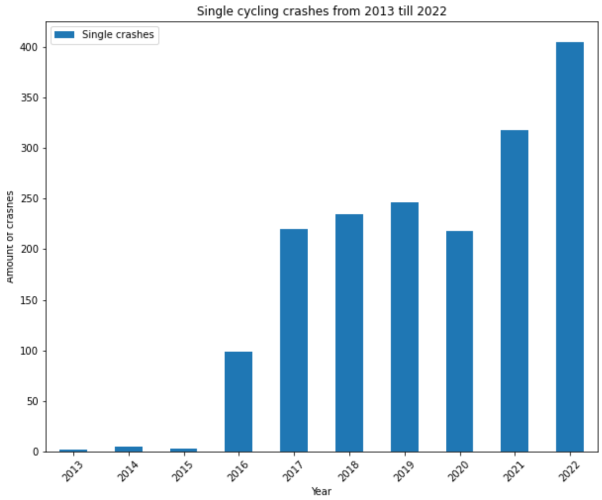


Figure 9: Distribution of Single-Bicycle Crashes

Purely looking at the total crashes without any variables selected, there were 23709 bicycle crashes and there were 1751 single-bicycle crashes between 2013 and 2022 for the four cities. This means that 7.39% of the crashes were single crashes in that decade. But looking at the rate of change it can be stated that every year single-bicycle crashes are rising. In Table 2 the percentage per year is noted. A surprising result is that in 2021 and 2022 more than 10% of the total bicycle crashes were single crashes.

Table 2: Distribution the ratio of Single-Bicycle Crashes

Year	**#Non-single crashes**	**#Single crashes**	**Proportion [%]**
2013	623	2	0.32
2014	1007	5	0.50
2015	1346	3	0.22
2016	1673	99	5.92
2017	3060	220	7.19
2018	3282	235	7.16
2019	3217	246	7.65
2020	2751	218	7.92
2021	2957	318	10.75
2022	3793	405	10.68
Total crashes	**23709**	1751	**7.39**

Now that the rapid increase is stated and transferred into numbers a closer look into the number of crashes for each variable for all crashes will be analysed then the same is done for single-bicycle crashes, so later on conclusions can be drawn from it. In Table 3 the amount of crashes for single-bicycle crashes and normal bicycle crashes are visualised. In total, there were 1235 single-bicycle crashes between 2013 and 2022 in Amsterdam, Rotterdam, Utrecht, and The Hague with information about Light situation, location, max speed, road circumstances, and road situation. Furthermore, from the table, it can be seen that more single-bicycle crashes occur at a straight road segment instead of an junction, which mostly happens for non-single-bicycle crashes.

Table 3: Distribution of Single-Bicycle Crashes

Variable	**Category**	**#Single**	**#Normal**
Lighting Condition	Daylight	876	14927
	Darkness	295	4088
	Dusk	64	1037
Location	Intersection	391	12160
	Road Segment	844	7892
Maximum Speed (Road)	15 km/h	40	238
	30 km/h	385	4541
	50 km/h	802	15164
	60, 70, 80, 100 km/h	8	108
Road surface condition	Dry	884	15193
	Wet	323	4771
	Snow/Black ice	28	88
Road Situation	Corner	53	570
	3-way junction	81	2523
	4-way junction	230	8124
	Straight road	838	7608
	Straight road (separated)	2	47
	Roundabout	31	1171
Total crashes		**1235**	**20052**

4.1.4 Crashes vs other modes

In the previous paragraph, a closer look into single-sided crashes has been done but it is also interesting to look at bicycle crashes with other modes of traffic, so that will be done in this section.

Data analysis has shown that every year between 2013 and 2022 the most crashes did occur with passenger cars looking from the perspective of the cyclist. Between 2013 and 2022 there were respectively 360, 601, 767, 903, 1547, 1684, 1612, 1424, 1404, and 1733 crashes with passenger cars. These numbers are visualised in a graph in Figure 10. This graph looks similar to the total bicycle crashes in the four cities, namely Figure 6. More interesting is to look at the distribution of crashes with all other traffic modes. This distribution is visualised in a pie chart in Figure 11.

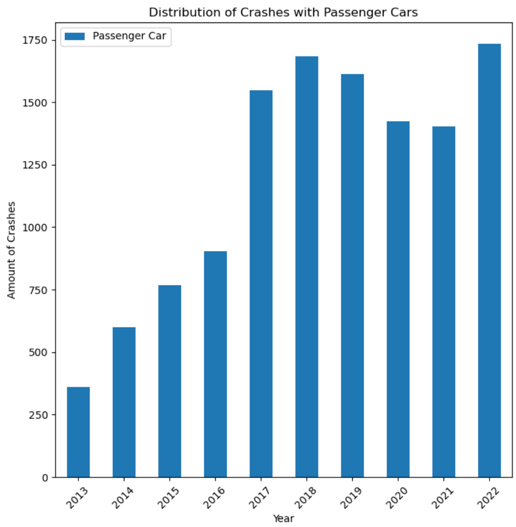


Figure 10: Distribution of Bicycle Crashes based on Road Section Type

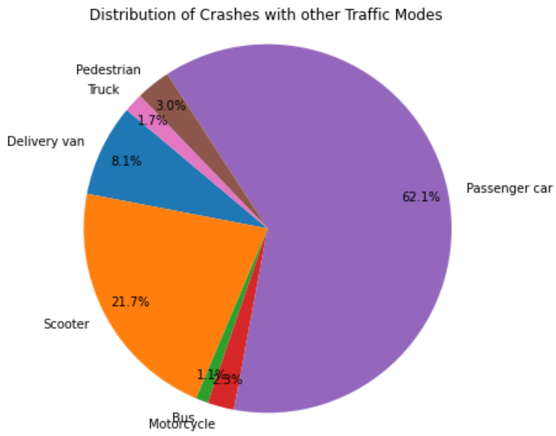


Figure 11: Distribution of Bicycle Crashes with other Traffic Modes

4.1.5 Crashes based on speed limits

In this section, the influence of maximum speed on the roads will be analysed. The maximum speed holds for motorised traffic.

Every single year between 2013 and 2022 the most crashes occurred at a road where the maximum speed for motorised vehicles is 50 km/h. The result of the analysis has shown that the number of crashes each year at a 50 km/h road were respectively 403, 728, 987, 1190, 2209, 2146, 1992, 1816, 1891, and 2257.

To create a perspective all crashes related to the maximum speed are visualised in a pie chart, see Figure 12. From the chart, it can be seen that more than three-quarters of the total crashes occur on roads where the speed limit is 50 km/h. Less than a quarter occurs on roads where the maximum speed is 30 km/h. In the chart other speeds are 15, 60, 70, 70, and 100 km/h. In this case, it is not surprising that this value is low since the focus is on cities and most roads are 30 or 50 km/h.

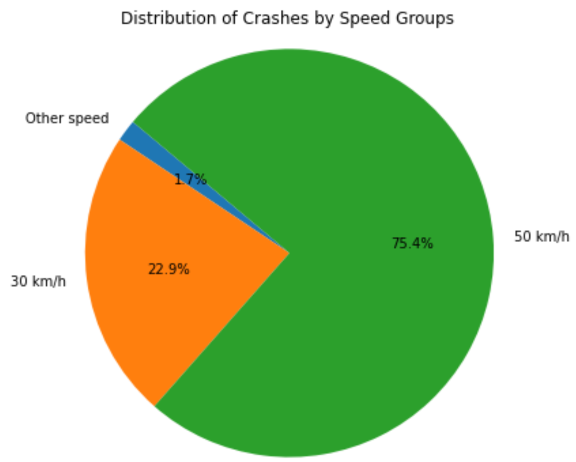


Figure 12: Distribution of Bicycle Crashes by Speed Groups

4.1.6 Crashes based on road section type

In this section, the influence of road section types will be analysed. The two types of road sections are junctions and road segments. With a road segment, a straight road without an junction is meant.

The results of all the crashes based on road section type are visualised in Figure 13. What can be obtained is that the rise and decrease of the two types show a similar rate, but every year there are more crashes on junctions. The slope of the linear regression line for junctions is approximately 176. For road segment crashes the slope of a linear regression line is approximately 151. This means that overall there is an increase in crashes at junctions. To get more knowledge about which type of junctions is the most dangerous, c.q. the most crashes, the next point will look at this, see Section 4.1.7 Crashes based on junction type.

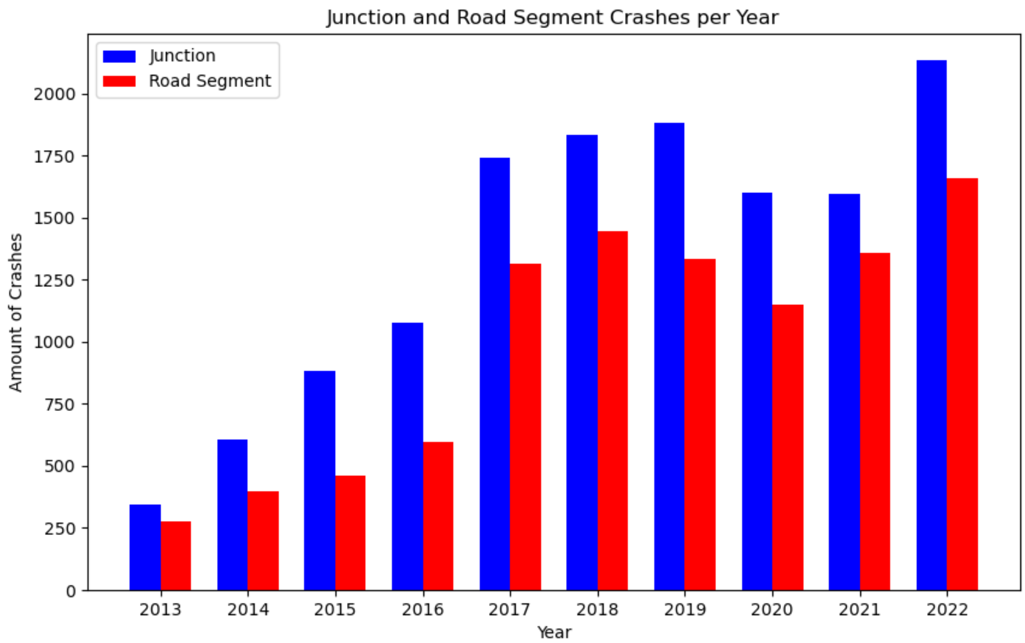


Figure 13: Distribution of Bicycle Crashes based on Road Section Type

4.1.7 Crashes based on junction type

In Section 4.1.6 Crashes based on road section type, it became clear that most crashes do occur on junctions compared to road segments where no other traffic emerges. This point will look closer into the junction types. The different types of junctions are corners, 3-way junctions, 4-way junctions, straight roads, straight road with separated lanes, and roundabouts.

The results of the analysis can be found in Table 4 and Figure 14. Table 4 highlights at which junction type were the most crashes that year. Every year till 2021 the 4-way junction contained the most crashes, but in the last 2 years, most crashes occurred on a straight road. When comparing Figure 13 and Table 4 it can be obtained that in the figure every year the most crashes were at a junction, but in the table the last two years more crashes occurred on a straight road. This is caused by the implementation of more categories within this section, so the level of detail. For example corner, straight road with separated lanes, and straight road are different categories. In Section 4.1.6 Crashes based on road section type, it was a combined category. The higher level of detail gives a more accurate look at this case.

Table 4: Max crashes per year based on Junction Type

Year	Junction Type	Crashes
2013	4-way junction	236
2014	4-way junction	396
2015	4-way junction	627
2016	4-way junction	758
2017	4-way junction	1249
2018	4-way junction	1294
2019	4-way junction	1265
2020	4-way junction	1044
2021	Straight road	1186
2022	Straight road	1517

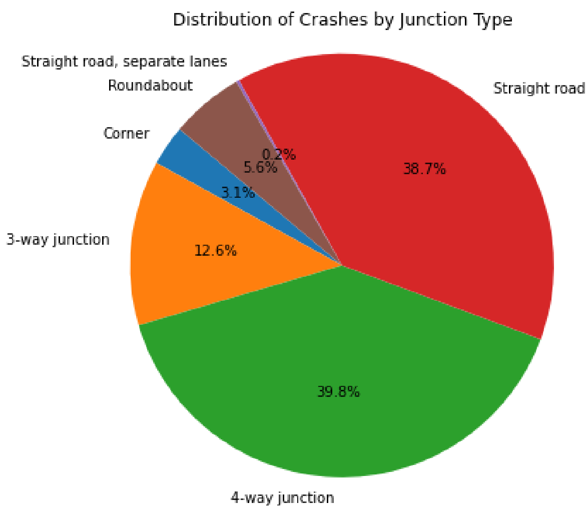


Figure 14: Distribution of Bicycle Crashes based on Junction Type

Now that it is clear that most crashes happened at a 4-way junction between 2013 and 2022, but that in the last two years more crashes occurred at a straight road. It is also interesting to see which junction type is the most dangerous, c.q. the highest risk of getting a serious injury. Therefore, a multinomial logistic regression analysis will be conducted in the next section where the relation between junction type and severity of the accidents will be analysed.

4.2 Statistical correlation test

In this section the correlation tests, as stated in the Methodology, will be conducted. The used statistical test is the chi-square test and it looks at the correlation between two variables. All the calculations and Python scripts can be found in Appendix B.

4.2.1 Correlation severity of accident & road segment

Firstly, the relationship between the severity of accidents and the type of road situation (either a road section or a junction) will be tested. The null hypothesis (H_0) is that there is no association between the severity of an accident and the type of road situation. After performing the test, the following values can be obtained in Table 5.

Table 5: Results of the chi-square test between accident severity and road situation type

Statistic	Value
Chi-Square (χ^2)	62.5
P-Value	$2.56e^{-14}$
Degrees of Freedom (dof)	2

The chi-square value of 62.5 indicates a substantial deviation between the observed and expected frequencies. The p-value of $2.56e^{-14}$ is significantly less than the conventional alpha level of 0.05. Given the extremely low p-value, the null hypothesis is rejected. This indicates that there is a statistically significant association between accident severity and the type of road situation.

4.2.2 Correlation severity of accident & junction type

After finding out the relevance of a junction compared to a road segment on the severity of the accident, it is interesting to look with more detail at the relation of the junction type on the severity of the accident. The data contains six different road situation types, namely corner, straight road, straight road (separated lines), 3-way junction, 4-way junction, and roundabout. For the test between THE severity of an accident and the junction types only the last three named junction types will be taken into account since on that other road situation no other traffic emerges, thus not an junction. The null hypothesis for this test is: There is no association between the severity of the accident and the type of the junction (3 or 4-way, and roundabout). The results of the performed chi-square test can be found in Table 6.

Table 6: Results of the chi-square test between accident severity and junction type

Statistic	Value
Chi-Square (χ^2)	29.0
P-Value	$7.96e^{-6}$
Degrees of Freedom (dof)	4

The chi-square statistic of 29 indicates the magnitude of deviation between the observed and expected values. This high value suggests a stronger association between the variables. The p-value of $7.96e^{-6}$ is extremely small. This means that the likelihood of observing a chi-square statistic as extreme as 29 or greater if the null hypothesis were true, is very low. The p-value is much smaller than 0.05, and therefore the null hypothesis will be rejected, which means that there is a strong relation between the severity of an accident and the junction type.

In Table 7 the values from the data and the expected values from the test are noted. The expected values are calculated based on the fact that there is no correlation between the two variables. In this case, there is a correlation but it shows how big the difference is for example between the observed and expected values for a 4-way junction. There are way more crashes than expected for every severity type. This shows how strong the effect of a junction type is on the amount of crashes.

Table 7: Observed and expected values for severity accident and junction type

	3-way junction		**4-way junction**		**Roundabout**	
	Observed	Expected	Observed	Expected	Observed	Expected
Material	1433	1516	4806	1516	740	672
Injury	1422	1337	4199	1337	530	592
Death	15	17	61	53	2	8

4.2.3 Correlation severity of accident & weather circumstances

The third statistical test will look at the correlation between the severity of an accident and weather circumstances. Both variables are nominal, so the chi-square test will be used again. Weather circumstances are categorised into 5 groups, namely Dry, heavy gusts of wind, fog, rain, and snow/hail. The null hypothesis for this test is: There is no association between the severity of the accident and weather circumstances.

The results of the test can be found in Table 8. The chi-square value is not specifically high and the p-value is above 0.05. So the null hypotheses will not be rejected and therefore it can be said that there is no correlation between the two tested variables, namely the severity of an accident and weather circumstances.

Table 8: Results of the chi-square test between accident severity and weather circumstances

Statistic	**Value**
Chi-Square (χ^2)	6.8
P-Value	0.6
Degrees of Freedom (dof)	8

An assumption for these results is that most crashes happened in dry and rainy circumstances, respectively there were 19122 and 2674 in those conditions. In total, only 142 crashes happened in heavy gusts of wind, fog, and snow/hail combined. After deleting those 142 crashes and only looking at dry and rain circumstances the chi-value got even lower, namely almost 3.

The p-value did get lower but only got reduced to 0.2, so it still did not get rejected. This means that there is no association. A reason could be that cyclist adapt their behavior on the bike when the weather becomes worse.

4.2.4 Correlation severity of accident & maximum speed (road)

The fourth test will look at the association between the severity of an accident and the maximum speed on the road for the vehicles. Since the focus is on 4 cities most of the speeds will be maximal 50 km/h. Therefore for this test, 4 groups are made, namely 15 km/h, 30 km/h, 50 km/h, and 1 last group with 60, 70, 80, 90, and 100 km/h. The null hypothesis for this test is that there is no association between the severity of the accident and the maximum speed on the road for motorised vehicles.

The results of the test can be found in Table 9. The p-value is much smaller than 0.05, and therefore the null hypothesis will be rejected, which means that there is a relation between the severity of an accident and the maximum road speed.

Table 9: Results of the chi-square test between accident severity and weather circumstances

Statistic	**Value**
Chi-Square (χ^2)	42.4
P-Value	$1.52e^{-7}$
Degrees of Freedom (dof)	6

4.2.5 Correlation severity of accident & lighting conditions

Another interesting test is to look at whether a correlation is between the severity of an accident and the lighting conditions at the moment of the crash. The used test is again the chi-square test. Possible outcomes for lighting conditions are daylight, darkness, and twilight. The null hypothesis for this test is that there is no association between the severity of the accident and the lighting conditions.

The results of the test can be found in Table 10. The results show that the null hypothesis will not be rejected, since the p-value is higher than 0.05. So there is no significant correlation between the severity of the accident and the lighting conditions. The severity of accidents and lighting conditions appear to be independent of each other. In other words, changes in lighting conditions do not seem to affect the severity of the accidents significantly.

Table 10: Results of the chi-square test between accident severity and lighting conditions

Statistic	**Value**
Chi-Square (χ^2)	5.82
P-Value	0.21
Degrees of Freedom (dof)	4

4.2.6 Correlation junction type & maximum speed (road)

The null hypothesis for this test is that there is no association between junction type and maximum speed on the road. The result of the chi-square test indicates that the type of junction and/or the speed are significantly associated with the number of accidents. This result is important to keep in mind when infrastructure is designed or adjusted. There is a relation between the amount of crashes for junctions and the maximum speed on the road, so to reduce the number of bicycle crashes it is important to design infrastructure based on this relation. Pay extra attention to 4-way junctions with a speed limit of 50 km/h or just avoid them completely. The statistical values of the outcome of the test are visualised in Table 11.

4.2.7 Correlation junction type & lighting conditions

The null hypothesis for this test is that there is no association between junction type and lighting conditions. The result of the test is that there is a correlation between those variables. When looking at safety measures it is for example important that in dark conditions at junctions, there is enough light created by streetlights, when it is assumed that less visibility due to darkness leads to more crashes. The statistical values of the outcome of the test are visualised in Table 11.

4.2.8 Correlation weather circumstances & maximum speed (road)

The null hypothesis for this test is that there is no association between weather circumstances and maximum speed on the road. The result of the test is that the null hypothesis was accepted. So there is no correlation between weather circumstances and maximum speed on the road when looking at the number of crashes. The statistical values of the outcome of the test are visualised in Table 11.

4.2.9 Correlation maximum speed (road) & lighting conditions

The null hypothesis of the last correlation test is that there is no association between maximum speed on the road and lighting conditions. The result of the test is that this hypothesis was accepted. So there is no association between maximum road speed and lighting conditions on the number of crashes. The statistical values of the outcome of the test are visualised in Table 11.

4.2.10 Summary of all correlation test

In this paragraph, an overview of all the correlation tests will be conducted. The above five did focus on the severity, but also other correlations are interesting for safety measures. In Table 11 all the results of all the correlation tests are noted.

Table 11: Results all correlations tests

Variable 1	**Variable 2**	**Statistic**	**Value**	H_0
Severity of accident	Road segment	Chi-Square (χ^2) P-value Dof	62.5 $2.56e^{-14}$ 2	Rejected
Severity of accident	Junction type	Chi-Square (χ^2) P-value Dof	29.0 $7.96e^{-6}$ 4	Rejected
Severity of accident	Weather circumstances	Chi-Square (χ^2) P-value Dof	6.8 0.6 8	Accepted
Severity of accident	Maximum speed (road)	Chi-Square (χ^2) P-value Dof	42.4 $1.52e^{-7}$ 6	Rejected
Severity of accident	Lighting conditions	Chi-Square (χ^2) P-value Dof	5.82 0.21 4	Accepted
Junction type	Maximum speed (road)	Chi-Square (χ^2) P-value Dof	703.92 $9.07e^{-145}$ 10	Rejected
Junction type	Lighting conditions	Chi-Square (χ^2) P-value Dof	22.24 0.01 10	Rejected
Weather circumstances	Maximum speed (road)	Chi-Square (χ^2) P-value Dof	6.08 0.99 24	Accepted
Maximum speed (road)	Lighting conditions	Chi-Square (χ^2) P-value Dof	22.83 0.06 14	Accepted

4.3 Multinomial logistic regression

In this section, the results of the multinomial logistic regression will be published and described. The focus is on the impact of accident severity on the type of junction in Amsterdam, Rotterdam, Utrecht, and The Hague taking into account the control variables inside/outside built-up area, lighting conditions, maximum speed (road), weather circumstances, and road surface condition. All the calculations and Python scripts can be found in Appendix D.

In Table 12 below, the description is provided for the nominal and ordinal variables along with their frequencies and the number of cases. As stated in Table 1 in Section 3.4 Operationalization of variables the maximum speeds 60, 70, 80, and 100 km/h are categorised as one group because every single speed has a low amount of cases, and therefore may influence the reliability and accuracy of the outcome.

Table 12: Description nominal and ordinal variables

	Frequencies	**Number of Cases**
Severity of Accident		18570
Material damage	9816	
Serious injury	9414	
Death	109	
Inside / outside built-up area		18570
Inside	18274	
Outside	296	
Lightning conditions		18570
Daylight	13833	
Darkness	3775	
Twilight	962	
Maximum speed (road)		18570
15 km/h	224	
30 km/h	4237	
50 km/h	14009	
60, 70, 80, 100 km/h	100	
Weather circumstances		18570
Dry	16079	
Rain / fog	2407	
Snow	50	
Road surface conditions		18570
Dry	14051	
Wet	4438	
Snow / black ice	81	
Road situation		18570
Straight road	7660	
Junction	10910	

The results of the analysis are noted in Table 13. To determine which junction has the highest likelihood of a more severe accident, the category material damage was chosen as the reference category. Material damage is chosen because this is the only category without physical damage, so the least severe outcome in this case. It enables direct comparisons between the more severe outcomes (serious injury and fatal) and material damage, providing the interpretation of the results.

Table 13: Results multinomial logistic regression analysis with reference category material damage

		B-coefficient	Exp(B)	p-value
Serious injury	Constant / Intercept	-0.043	0.958	0.465
	Inside / outside built-up area	0.582	1.790	0.000
	Lightning conditions	-0.043	0.958	0.105
	Maximum speed (road)	0.060	1.062	0.062
	Weather circumstances	-0.073	0.930	0.161
	Road surface conditions	0.111	1.117	0.009
	Road situation	-0.210	0.811	0.000
Fatal accident	Constant / Intercept	-6.164	0.002	0.000
	Inside / outside built-up area	1.553	4.726	0.000
	Lightning conditions	0.123	1.131	0.461
	Maximum speed (road)	0.835	2.305	0.002
	Weather circumstances	-0.939	0.391	0.025
	Road surface conditions	0.322	1.394	0.217
	Road situation	0.120	1.127	0.560

[a] *pseudo* ρ^2 : 0.004111
[b] *n: 18570*

In the table, it can be observed that only the intercept of fatal accidents is significant in the model, whereas that of serious injuries is not. The logit of both intercepts is negative, indicating that under similar conditions (i.e., all other independent variables being 0), the likelihood of the outcome of a bicycle accident being material damage is higher compared to injury and fatal accidents. Regarding (serious) injury accidents, the following variables are significant (p-value < 0.05), namely inside/outside built-up area, maximum speed (road), road conditions, and road situation. About those variables, things can be concluded, not about the non-significant variables. For inside/outside built-up area, maximum speed (road), and road conditions, the Exp(B) value is greater than 1. This implies for inside/outside built-up area that the likelihood is 1.790 times higher outside built-up areas to get a serious injury compared to crashes with material damage. Similarly, for a fatal accident, this likelihood is 4.726 times higher compared to material damage.

Likewise, the likelihood increases by a factor of 1.062 for an accident resulting in injury when the maximum speed increases by one step, so from 0 to 1, or from 1 to 2 compared to crashes with material damage. So the chance of getting a crash with a serious injury increases with a factor of 1.062 when the speed group goes up by 1. For fatal crashes, the likelihood increases by 2.305 times compared to material damage when the speed increases by one step. This is not surprising, considering that the likelihood of injury or a fatal accident already increased when outside built-up areas, where the speed is generally higher than within the built-up areas. Weather conditions are not significant for accidents resulting in (serious) injury. However, road conditions are significant. The likelihood is 1.117 times higher for an accident resulting in injury compared to material damage when road conditions improve by one step, indicating a deterioration of road conditions in this case. When the road situation is increased by one step, from road segment to junction, the likelihood is 0.811 that the accident results in injury compared to an accident with material damage.

Regarding the severity of accidents, inside/outside built-up area and maximum speed on the road have already been discussed, as they were significant. Another significant variable is weather conditions. The result is that the likelihood of a fatal accident is 0.391 compared to an accident with material damage when weather conditions increase by 1 step. In the footnote of the table, the pseudo-R^2 value is provided. It can be observed that both values are very low, indicating high variability. This will be discussed in the Discussion. Where weather circumstances don't play a role in the frequency and severity of a crash resulting in a serious injury compared to material damage, road surface conditions do. This relation does not hold for fatal accidents. Road surface conditions are related to dry, wet, or snowy roads. The likelihood of getting a serious injury from a crash is 1.117 compared to crashes with material damage. So when the road becomes worse, let's say slippery, people find it harder to indicate the grip levels. Where by weather circumstances people can estimate the situation better and adapt to it.

5 Discussion

The results of this research provide insights into the variables and frequencies of cycling crashes in Amsterdam, Rotterdam, Utrecht, and The Hague. Additionally, the research has identified correlations between two variables and developed a model to examine the interrelations of the most significant variables. This chapter addresses some limitations of the research and their impact on the results.

The literature analysis focused on academic papers, public reports and statistics from governmental institutions in the Netherlands. However, some relevant reports or papers may have been missed, potentially leading to incomplete or incorrect information. To mitigate this, multiple sources were used for the literature analysis.

A limitation within the data analysis is the potential exclusion of other relevant variables not available in the dataset. Factors such as traffic density at the time of the crash, accurate time of day, and the speed of the vehicle at the moment of the crash could influence cycling crashes. Including such variables could lead to more accurate research. Additionally, the dataset lacked personal information about the crashes, such as gender, age, and nationality.

Another limitation is the smaller sample size of fatal crashes compared to material damage and injury crashes. While representative, the disparity in the number of cases might have influenced the accuracy of the correlation tests and regression models.

Lastly, the multinomial logistic regression had a low pseudo-R^2 value. This is partly due to the inclusion of many independent variables, which showed low correlations with each other and the dependent variable, making it harder to predict the entire model. As mentioned in the discussion, the low number of fatal crash cases could also have contributed to the low pseudo-R^2 value.

6 Conclusions and recommendations

This chapter aims to form a general conclusion based on the analysis and the discussion given in Chapters 4 and 5. The final goal is to find an answer to the main research question *"What patterns, factors, and variables have contributed to the frequency and severity of cycling crashes in Amsterdam, Rotterdam, Utrecht, and The Hague between 2013 and 2023?"* First, a summary of the findings will highlight all the results, followed by the lessons and recommendations, and lastly, future research will be mentioned.

6.1 Summary of findings

The rise in electric bike usage since 2013 coincides with the increase in bicycle crashes. However, the risk comparison between electric and regular bikes remains inconclusive. Cycling kilometers on electric bikes still continues to grow. In urban areas, the overall bicycle usage is rising. This is related to the population growth in the Netherlands.

All four analysed cities have shown an increase in bicycle crashes. The highest rate of change per year was observed in The Hague. Another notable pattern is the overall rise in single-bicycle crashes compared to non-single crashes. Single crashes are more likely to occur at lower speeds (30 km/h) and on straight road segments rather than at junctions compared with non-single-bicycle crashes. Most accidents were material damage crashes. Important variables affecting the frequency and severity of crashes include the type of road segment (straight road or junction), junction type, and maximum speed limit. Most bicycle crashes involve passenger cars. The likelihood of a serious injury or fatal accident increases as the maximum speed limit increases (e.g., from 15 to 30 km/h, 30 to 50 km/h, or 50 to 60, 70, 80, and 100 km/h), compared to material damage crashes. This increased likelihood is also evident when leaving built-up areas.

Besides the speed on the road, the junction type with a 4-way junction contains the most number of crashes. Especially 4-way junctions with a maximum speed of 50 km/h do lead to more crashes. However, there is no increased likelihood of a serious injury or fatal accident at junctions compared to straight roads when comparing it to material damage crashes.

Weather conditions do not influence the frequency and severity of crashes significantly, as people tend to adjust their behavior based on the weather. No correlation was found between the number of crashes and weather conditions or maximum speed limits. Additionally, there is no correlation between maximum speed and lighting conditions on the frequency of crashes. The analysis also showed no significant correlation between lighting conditions and the severity of accidents. This lack of correlation may be due to well-lit roads in the cities, neutralizing any potential impact of lighting conditions.

While weather conditions do not significantly affect the frequency and severity of crashes resulting in serious injuries, road surface conditions do. However, this relationship does not hold for fatal accidents. In fact, adverse weather conditions may reduce the likelihood of fatal accidents compared to material damage crashes.

6.2 Lessons & recommendations

From the summary of the findings, some important lessons can be drawn regarding cycling infrastructure from a safety perspective. The EU aims to reduce all traffic crashes, including those involving cyclists, so safety measures are essential.

Key variables and their correlations reveal that most crashes occur at 4-way junctions with a maximum speed of 50 km/h in cities. This is a crucial consideration for infrastructure design. Avoiding this combination is advised. Instead, implementing different types of junctions, such as 3-way junctions or roundabouts, or reducing speed limits within cities where possible, is recommended.

To decrease the severity of accidents, the results indicate that straight roads are more dangerous than junctions. Interestingly, more crashes occurred on straight roads in 2021 and 2022 than at junctions. This, combined with the increase in single-bicycle crashes, underscores the importance of designing safer straight roads to meet the EU's safety goals. Key factors include minimizing obstacles, avoiding narrow roads, and ensuring road surfaces are not slippery due to weather conditions.

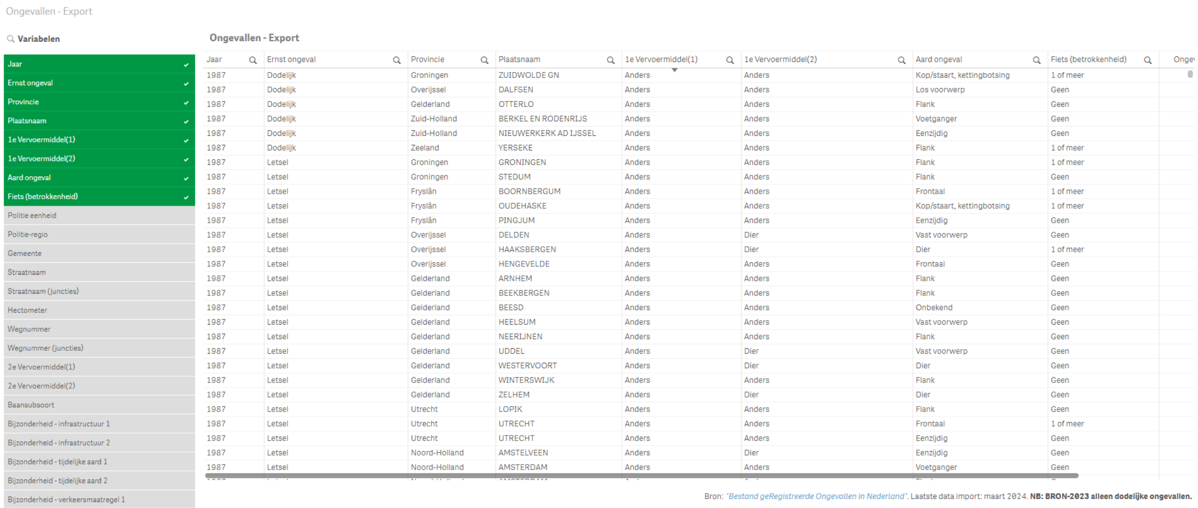
6.3 Future research

This research establishes the foundation for addressing a growing problem in the Netherlands: the rise in bicycle crashes. The focus was on four cities with the highest absolute number of cycling crashes. However, it would be interesting to investigate other cities, as well as areas with less traffic, such as villages or rural areas. On the other side, a rapid increase in The Hague was seen. So looking more closely into The Hague its cycling infrastructure can be really valuable.

Investigating fatal crashes more extensively is another recommendation. Although material damage and serious injury crashes are more common, reflecting real-life conditions, increasing the number of fatal cases through machine learning could create a more accurate model for estimating the relevant variables.

Lastly, a deeper investigation into single-bicycle crashes is crucial. This study highlighted their increase and identified important variables, but focused more on non-single-bicycle crashes. Given the rising number of single-bicycle crashes, it is important to study them more closely.

A Overview dataset + all variables



Jaar	Ernst ongeval	Provincie	Plaatsnaam	1e Vervoermiddel(1)	1e Vervoermiddel(2)	Aard ongeval	Fiets (betrokkenheid)	Onge
1987	Dodelijk	Groningen	ZUIDWOLDE GN	Anders	Anders	Kop/staart, kettingbotsing	1 of meer	
1987	Dodelijk	Overijssel	DALFSEN	Anders	Anders	Los voorwerp	Geen	
1987	Dodelijk	Gelderland	OTTERLO	Anders	Anders	Flank	Geen	
1987	Dodelijk	Zuid-Holland	BERKEL EN RODENRIJS	Anders	Anders	Voetganger	Geen	
1987	Dodelijk	Zuid-Holland	NIEUWERKERK AD IJSSEL	Anders	Anders	Eenzijdig	Geen	
1987	Dodelijk	Zeeland	YERSEKE	Anders	Anders	Flank	1 of meer	
1987	Letsel	Groningen	GRONINGEN	Anders	Anders	Flank	1 of meer	
1987	Letsel	Groningen	STEDUM	Anders	Anders	Flank	Geen	
1987	Letsel	Fryslân	BOORNBERGUM	Anders	Anders	Frontaal	1 of meer	
1987	Letsel	Fryslân	OUDEHASKE	Anders	Anders	Kop/staart, kettingbotsing	1 of meer	
1987	Letsel	Fryslân	PINGJUM	Anders	Anders	Eenzijdig	Geen	
1987	Letsel	Overijssel	DELDEN	Anders	Dier	Vast voorwerp	Geen	
1987	Letsel	Overijssel	HAAKSBERGEN	Anders	Dier	Dier	1 of meer	
1987	Letsel	Overijssel	HENGEVELDE	Anders	Anders	Frontaal	Geen	
1987	Letsel	Gelderland	ARNHEM	Anders	Anders	Flank	Geen	
1987	Letsel	Gelderland	BEEKBERGEN	Anders	Anders	Flank	Geen	
1987	Letsel	Gelderland	BEESD	Anders	Anders	Onbekend	Geen	
1987	Letsel	Gelderland	HEELSUM	Anders	Anders	Vast voorwerp	Geen	
1987	Letsel	Gelderland	NEERIJNEN	Anders	Anders	Flank	Geen	
1987	Letsel	Gelderland	UDDEL	Anders	Dier	Vast voorwerp	Geen	
1987	Letsel	Gelderland	WESTERVOORT	Anders	Dier	Dier	Geen	
1987	Letsel	Gelderland	WINTERSWIJK	Anders	Anders	Flank	Geen	
1987	Letsel	Gelderland	ZELHEM	Anders	Dier	Dier	Geen	
1987	Letsel	Utrecht	LOPIK	Anders	Anders	Flank	Geen	
1987	Letsel	Utrecht	UTRECHT	Anders	Anders	Frontaal	1 of meer	
1987	Letsel	Utrecht	UTRECHT	Anders	Anders	Eenzijdig	Geen	
1987	Letsel	Noord-Holland	AMSTELVEEN	Anders	Dier	Eenzijdig	Geen	
1987	Letsel	Noord-Holland	AMSTERDAM	Anders	Anders	Voetganger	Geen	

Bron: "*Bestand geRegistreerde Ongevallen in Nederland*". Laatste data import: maart 2024. **NB: BRON-2023 alleen dodelijke ongevallen.**

All variables in Dutch: Jaar, Ernst Ongeval, Provincie, Politie Eenheid, Politie-regio, Gemeente, Plaatsnaam, Straatnaam, Straatnaam (Juncties), Hectometer, Wegnummer, Wegnummer (Juncties), 1e Vervoermiddel(1), 1e Vervoermiddel(2), 2e Vervoermiddel(1), 2e Vervoermiddel(2), Aard Ongeval, Baansubsoort, Bijzonderheid-Infrastructuur 1, Bijzonderheid-Infrastructuur 2, Bijzonderheid - Tijdelijke Aard 1, Bijzonderheid - Tijdelijke Aard 2, Bijzonderheid - Verkeersmaatregel 1, Bijzonderheid - Verkeersmaatregel 2, Binnen / Buiten Bebouwde Kom, Lichtgesteldheid, Locatie (Wegvak Of Kruispunt), Maximum Snelheid (Weg), Weer, Wegbeheerder, Wegdek Toestand, Wegdek Toestand - Anders, Wegsituatie, Wegsituatie Anders, Wegsoort-ASW, Wegverharding, Wegverharding - Anders, Wegverlichting, Zichtafstand, Bestelauto (Betrokkenheid), Boom (Betrokkenheid), Brom/Snorfiets (Betrokkenheid), Bus (Betrokkenheid), Fiets (Betrokkenheid), Landbouwvoertuig (Betrokkenheid), Lichtmast (Betrokkenheid), Motor (Betrokkenheid), Objecten (Betrokkenheid), Personenauto (Betrokkenheid), Voetganger (Betrokkenheid), Vrachtauto (Betrokkenheid).

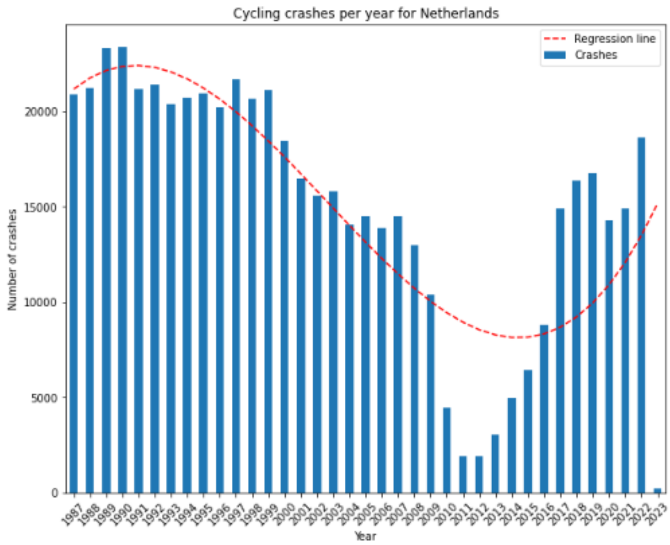
In English: Year, Severity Of Accident, Province, Police Unit, Police Region, Municipality, Place Name, Street Name, Street Name (Junctions), Hectometer, Road Number, Road Number (Junctions), 1st Vehicle(1), 1st Vehicle(2), 2nd Vehicle(1), 2nd Vehicle(2), Type Of Accident, Lane Subtype, Infrastructure Feature 1, Infrastructure Feature 2, Temporary Feature 1, Temporary Feature 2, Traffic Measure 1, Traffic Measure 2, Inside / Outside Built-up Area, Lighting Conditions, Location (Road Segment Or Intersection), Maximum Speed (Road), Weather, Road Authority, Road Surface Condition, Road Surface Condition - Other, Road Situation, Road Situation - Other, Road Type-highway, Road Pavement, Road Pavement - Other, Road Lighting, Visibility Distance, Delivery Van (Involvement), Tree (Involvement), Scooter (Involvement), Bus (Involvement), Bicycle (Involvement), Agricultural Vehicle (Involvement), Light Pole (Involvement), Motorcycle (Involvement), Objects (Involvement), Passenger Car (Involvement), Pedestrian (Involvement), Pedestrian (Involvement)

B Python script variable analysis

Checking the Literature

```
data = pd.read_excel('Jaarlijkse_fietsongelukken.xlsx', index_col='Jaar')
data = data.drop(index= 2023, columns='Fiets_betrokken')
```

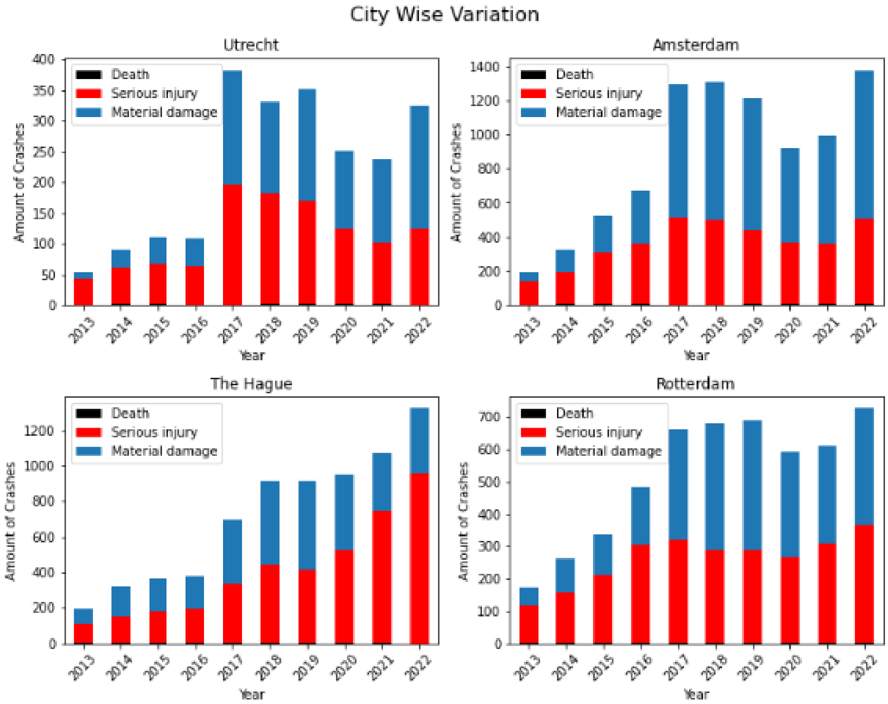
```
plt.figure(figsize=(10,8))

x = np.arange(len(data))
y = data['Ongevallen']
coefficients = np.polyfit(x, y, deg=3)  # Change degree here
poly = np.poly1d(coefficients)
regression_line = poly(x)

data['Ongevallen'].plot(kind='bar', label='Crashes')
plt.plot(x, regression_line, color='red', linestyle='--', label='Regression line')

plt.xlabel('Year')
plt.ylabel('Number of crashes')
plt.title('Cycling crashes per year for Netherlands')
plt.xticks(rotation=45)
plt.legend()
plt.savefig('Yearly_crashes');
```



1. City wise variation

```
city_wise = pd.read_excel('1.City_wise.xlsx', index_col='Jaar')
city_wise = city_wise.drop(columns = ['Fiets_betrokkenheid'])

Utrecht = city_wise[city_wise['Plaatsnaam'] == 'UTRECHT']
Utrecht = Utrecht.drop(columns = ['Plaatsnaam'])
Adam = city_wise[city_wise['Plaatsnaam'] == 'AMSTERDAM']
Adam = Adam.drop(columns = ['Plaatsnaam'])
Rotterdam = city_wise[city_wise['Plaatsnaam'] == 'ROTTERDAM']
Rotterdam = Rotterdam.drop(columns = ['Plaatsnaam'])
Dhaag = city_wise[city_wise['Plaatsnaam'] == 'DEN HAAG']
Dhaag = Dhaag.drop(columns = ['Plaatsnaam'])

Utrecht_dodelijk = Utrecht[Utrecht['Ernst_ongeval'] == 'Dodelijk']
Utrecht_letsel = Utrecht[Utrecht['Ernst_ongeval'] == 'Letsel']
Utrecht_material = Utrecht[Utrecht['Ernst_ongeval'] == 'Uitsluitend materiële schade']

Adam_dodelijk = Adam[Adam['Ernst_ongeval'] == 'Dodelijk']
Adam_letsel = Adam[Adam['Ernst_ongeval'] == 'Letsel']
Adam_material = Adam[Adam['Ernst_ongeval'] == 'Uitsluitend materiële schade']

Rotterdam_dodelijk = Rotterdam[Rotterdam['Ernst_ongeval'] == 'Dodelijk']
Rotterdam_letsel = Rotterdam[Rotterdam['Ernst_ongeval'] == 'Letsel']
Rotterdam_material = Rotterdam[Rotterdam['Ernst_ongeval'] == 'Uitsluitend materiële schade']

Dhaag_dodelijk = Dhaag[Dhaag['Ernst_ongeval'] == 'Dodelijk']
Dhaag_letsel = Dhaag[Dhaag['Ernst_ongeval'] == 'Letsel']
Dhaag_material = Dhaag[Dhaag['Ernst_ongeval'] == 'Uitsluitend materiële schade']
```

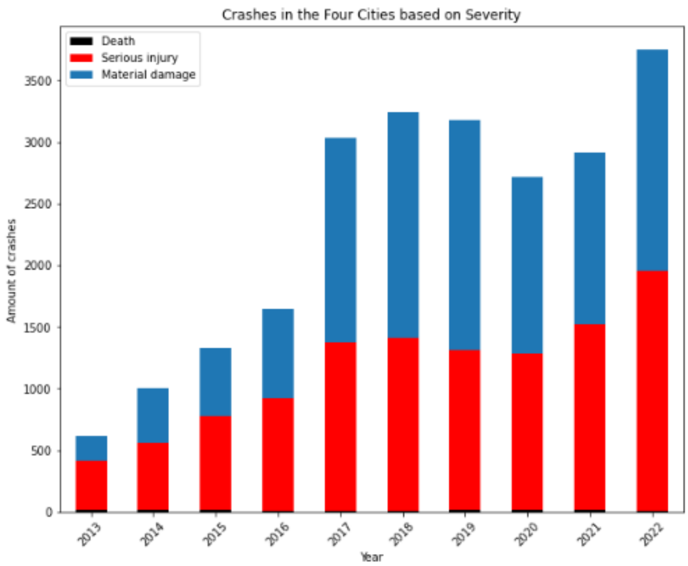
```
fig, axs = plt.subplots(2, 2, figsize=(10, 8))
fig.suptitle('City Wise Variation', fontsize=16)

Utrecht_dodelijk['Ongevallen'].plot(ax=axs[0, 0], kind='bar', label='Death', color='k', legend=True)
Utrecht_letsel['Ongevallen'].plot(ax=axs[0,0], kind='bar', label='Serious injury', bottom=Utrecht_dodelijk['Ongevallen'], color='r', legend=True)
Utrecht_material['Ongevallen'].plot(ax=axs[0,0], kind='bar', label='Material damage', bottom=Utrecht_dodelijk['Ongevallen'] + Utrecht_letsel['Ongevallen'], legend=True)
plt.xticks(rotation=45)
axs[0, 0].set_ylabel('Amount of Crashes')
axs[0, 0].set_xlabel('Year')
axs[0, 0].set_title('Utrecht')
axs[0, 0].set_xticklabels(Utrecht_dodelijk.index, rotation=45)

Adam_dodelijk['Ongevallen'].plot(ax=axs[0, 1], kind='bar', label='Death', color='k', legend=True)
Adam_letsel['Ongevallen'].plot(ax=axs[0, 1], kind='bar', label='Serious injury', bottom=Adam_dodelijk['Ongevallen'], color='r', legend=True)
Adam_material['Ongevallen'].plot(ax=axs[0, 1], kind='bar', label='Material damage', bottom=Adam_dodelijk['Ongevallen'] + Adam_letsel['Ongevallen'], legend=True)
plt.xticks(rotation=45)
axs[0, 1].set_ylabel('Amount of Crashes')
axs[0, 1].set_xlabel('Year')
axs[0, 1].set_title('Amsterdam')
axs[0, 1].set_xticklabels(Utrecht_dodelijk.index, rotation=45)


Rotterdam_dodelijk['Ongevallen'].plot(ax=axs[1, 1], kind='bar', label='Death', color='k', legend=True)
Rotterdam_letsel['Ongevallen'].plot(ax=axs[1, 1], kind='bar', label='Serious injury', bottom=Rotterdam_dodelijk['Ongevallen'], color='r', legend=True)
Rotterdam_material['Ongevallen'].plot(ax=axs[1, 1], kind='bar', label='Material damage', bottom=Rotterdam_dodelijk['Ongevallen'] + Rotterdam_letsel['Ongevallen'], legend=True)
plt.xticks(rotation=45)
axs[1, 1].set_ylabel('Amount of Crashes')
axs[1, 1].set_xlabel('Year')
axs[1, 1].set_title('Rotterdam')
axs[1, 1].set_xticklabels(Utrecht_dodelijk.index, rotation=45)

Dhaag_dodelijk['Ongevallen'].plot(ax=axs[1, 0], kind='bar', label='Death', color='k', legend=True)
Dhaag_letsel['Ongevallen'].plot(ax=axs[1, 0], kind='bar', label='Serious injury', bottom=Dhaag_dodelijk['Ongevallen'], color='r', legend=True)
Dhaag_material['Ongevallen'].plot(ax=axs[1, 0], kind='bar', label='Material damage', bottom=Dhaag_dodelijk['Ongevallen'] + Dhaag_letsel['Ongevallen'], legend=True)
plt.xticks(rotation=45)
axs[1, 0].set_ylabel('Amount of Crashes')
axs[1, 0].set_xlabel('Year')
axs[1, 0].set_title('The Hague')
axs[1, 0].set_xticklabels(Utrecht_dodelijk.index, rotation=45)


plt.tight_layout()
plt.show()
plt.savefig('1City_Wise_Variation');
```




```
<Figure size 432x288 with 0 Axes>
```

2. Crashes based on severity level

```
city_wise = pd.read_excel('1.City_wise.xlsx', index_col='Jaar')
city_wise = city_wise.drop(columns = ['Fiets_betrokkenheid', 'Plaatsnaam'])

City_wise_dodelijk = city_wise[city_wise['Ernst_ongeval'] == 'Dodelijk']
City_wise_letsel = city_wise[city_wise['Ernst_ongeval'] == 'Letsel']
City_wise_material = city_wise[city_wise['Ernst_ongeval'] == 'Uitsluitend materiële schade']

City_wise_dodelijk = City_wise_dodelijk.groupby('Jaar').agg({'Ongevallen': 'sum'}).reset_index()
City_wise_letsel = City_wise_letsel.groupby('Jaar').agg({'Ongevallen': 'sum'}).reset_index()
City_wise_material = City_wise_material.groupby('Jaar').agg({'Ongevallen': 'sum'}).reset_index()

plt.figure(figsize=(10,8))
City_wise_dodelijk['Ongevallen'].plot(kind='bar', label='Death', color='k')
City_wise_letsel['Ongevallen'].plot(kind='bar', label='Serious injury', bottom=City_wise_dodelijk['Ongevallen'], color='r')
City_wise_material['Ongevallen'].plot(kind='bar', label='Material damage', bottom=City_wise_dodelijk['Ongevallen'] + City_wise_letsel['Ongevallen'])
plt.ylabel('Amount of crashes')
plt.xlabel('Year')
plt.xticks(ticks = list(range(10)), labels = np.arange(2013,2023,1), rotation=45)
plt.title('Crashes in the Four Cities based on Severity')
plt.legend()
plt.savefig('2Crashes_based_Severity_Level');
```



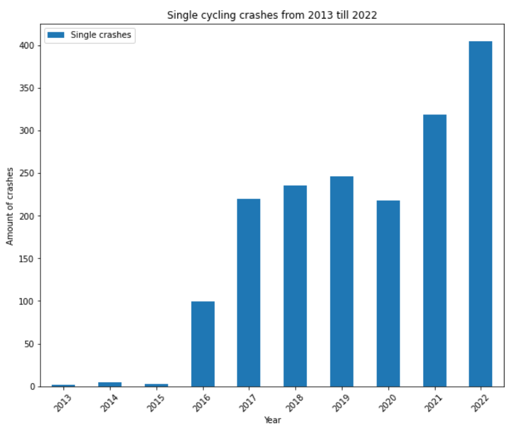
3. Single Bicycle crashes pattern

```
[8]: All_crashes = pd.read_excel('All_crashes.xlsx')
All_crashes = All_crashes.drop(columns = ['Fiets (betrokkenheid)', 'Plaatsnaam'])
All_crashes = All_crashes.groupby('Jaar').sum()

# display(All_crashes)

All_single_crashes = pd.read_excel('All_single_crashes.xlsx')
All_single_crashes = All_single_crashes.drop(columns = ['Fiets (betrokkenheid)', 'Plaatsnaam', 'Aard ongeval'])
All_single_crashes = All_single_crashes.groupby('Jaar').sum()

All_crashes['Single_Crashes'] = All_single_crashes
All_crashes['Proportion'] = round((All_crashes['Single_Crashes'] * 100 / All_crashes['Ongevallen']),2)

# print(All_crashes.Single_Crashes.sum())

display(All_crashes)
```

Jaar	Ongevallen	Single_Crashes	Proportion
2013	623	2	0.32
2014	1007	5	0.50
2015	1346	3	0.22
2016	1673	99	5.92
2017	3060	220	7.19
2018	3282	235	7.16
2019	3217	246	7.65
2020	2751	218	7.92
2021	2957	318	10.75
2022	3793	405	10.68

```
[9]: All_single_crashes = pd.read_excel('All_single_crashes.xlsx',index_col='Jaar')
All_single_crashes = All_single_crashes.drop(columns = ['Fiets (betrokkenheid)', 'Plaatsnaam', 'Aard ongeval'])

print('In total there were', All_single_crashes.Ongevallen.sum(), 'single crashes');
amount_of_sinlge_crashes = All_single_crashes.Ongevallen.sum()
All_single_crashes = All_single_crashes[['Ongevallen']]
All_single_crashes = All_single_crashes.groupby('Jaar')['Ongevallen'].sum()
display(All_single_crashes)

plt.figure(figsize=(10,8))
All_single_crashes.plot(kind='bar', label = 'Single crashes')
plt.title('Single cycling crashes from 2013 till 2022')
plt.ylabel('Amount of crashes')
plt.xlabel('Year')
plt.xticks(rotation=45)
plt.legend()
plt.savefig('3All_Single_Crashes');
```

```
In total there were 1751 single crashes
Jaar
2013      2
2014      5
2015      3
2016     99
2017    220
2018    235
2019    246
2020    218
2021    318
2022    405
Name: Ongevallen, dtype: int64
```



```
[10]: All_crashes = pd.read_excel('3.All_crashes2.0.xlsx', index_col='Jaar')
All_crashes = All_crashes.drop(columns=['Fiets (betrokkenheid)', 'Plaatsnaam'])

print('In total there were', All_crashes.Ongevallen.sum(), 'bicycle crashes with information about Light situation, location, max speed, road curcomstances, road situation')
print('The percentage of single crashes =', amount_of_sinlge_crashes / All_crashes.Ongevallen.sum() * 100,'%') ;

Light = All_crashes.groupby('Lichtgesteldheid')['Ongevallen'].sum().reset_index()
display(Light)

Road_type = All_crashes.groupby('Locatie (wegvak of kruispunt)')['Ongevallen'].sum().reset_index()
display(Road_type)

Max_speed = All_crashes.groupby('Maximum snelheid (weg)')['Ongevallen'].sum().reset_index()
display(Max_speed)

Road_status = All_crashes.groupby('Wegdek toestand')['Ongevallen'].sum().reset_index()
display(Road_status)

Road_situation = All_crashes.groupby('Wegsituatie')['Ongevallen'].sum().reset_index()
display(Road_situation)
```

```
In total there were 20052 bicycle crashes with information about Light situation, location, max speed, road curcomstances, road situation
The percentage of single crashes = 8.732296030321166 %
```

	Lichtgesteldheid	Ongevallen
0	Daglicht	14927
1	Duisternis	4088
2	Schemer	1037

	Locatie (wegvak of kruispunt)	Ongevallen
0	Kruispunt	12160
1	Wegvak	7892

	Maximum snelheid (weg)	Ongevallen
0	100 km/u	9
1	15 km/u	238
2	30 km/u	4541
3	50 km/u	15164
4	60 km/u	29
5	70 km/u	31
6	80 km/u	39
7	Stapv./woonerf	1

	Wegdek toestand	Ongevallen
0	Droog	15193
1	Nat/vochtig	4771
2	Sneeuw/ijzel	88

	Wegsituatie	Ongevallen
0	Bocht	570
1	Kruispunt, 3 takken	2532
2	Kruispunt, 4 takken	8124
3	Rechte weg	7608
4	Rechte weg, gescheiden rijbanen	47
5	Rotonde	1171

Compare with single crashes

```
[11]: single = pd.read_excel('3.single2.0.xlsx', index_col='Jaar')
single = single.drop(columns = ['Fiets (betrokkenheid)', 'Plaatsnaam', 'Objecten (betrokkenheid)', 'Aard ongeval'])
print('In total there were', single.Ongevallen.sum(), 'single bicycle crashes with information about Light situation, location, max speed, road curcomstances, road situation')


Light = single.groupby('Lichtgesteldheid')['Ongevallen'].sum().reset_index()
display(Light)

Road_type = single.groupby('Locatie (wegvak of kruispunt)')['Ongevallen'].sum().reset_index()
display(Road_type)

Max_speed = single.groupby('Maximum snelheid (weg)')['Ongevallen'].sum().reset_index()
display(Max_speed)

Road_status = single.groupby('Wegdek toestand')['Ongevallen'].sum().reset_index()
display(Road_status)

Road_situation = single.groupby('Wegsituatie')['Ongevallen'].sum().reset_index()
display(Road_situation)
```

```
In total there were 1235 single bicycle crashes with information about Light situation, location, max speed, road curcomstances, road situation
```

	Lichtgesteldheid	Ongevallen
0	Daglicht	876
1	Duisternis	295
2	Schemer	64

	Locatie (wegvak of kruispunt)	Ongevallen
0	Kruispunt	391
1	Wegvak	844

	Maximum snelheid (weg)	Ongevallen
0	100 km/u	1
1	15 km/u	40
2	30 km/u	385
3	50 km/u	802
4	60 km/u	1
5	70 km/u	4
6	80 km/u	2

	Wegdek toestand	Ongevallen
0	Droog	884
1	Nat/vochtig	323
2	Sneeuw/ijzel	28

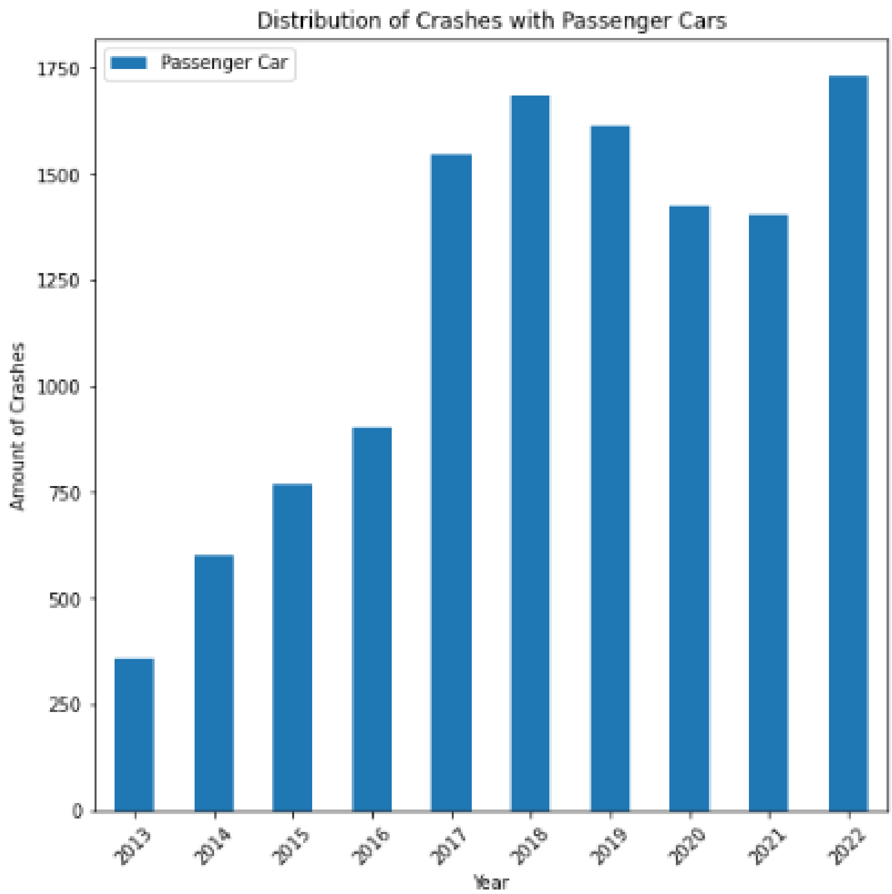
	Wegsituatie	Ongevallen
0	Bocht	53
1	Kruispunt, 3 takken	81
2	Kruispunt, 4 takken	230
3	Rechte weg	838
4	Rechte weg, gescheiden rijbanen	2
5	Rotonde	31

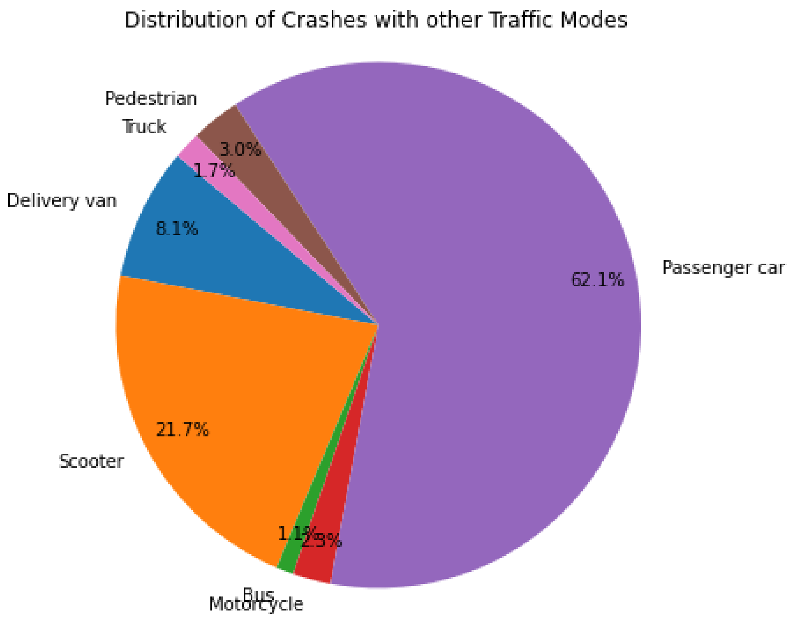
4. Crashes vs other modes of vehicles/pedestrian

```
modes = pd.read_excel('4.other_traffic2.0.xlsx', index_col='Jaar')
modes = modes.drop(columns = ['Fiets (betrokkenheid)', 'Plaatsnaam'])
# display(modes)
print('In total there are', modes.Ongevallen.sum()  ,'crashes with other traffic modes')

modes_grouped = modes.groupby(['Jaar', 'Type']).sum()
modes_per_year = modes_grouped.loc[modes_grouped.groupby('Jaar')['Ongevallen'].idxmax()]
display(modes_per_year)

plt.figure(figsize=(8,8))
modes_per_year.Ongevallen.plot(kind='bar', label='Passenger Car')
plt.title('Distribution of Crashes with Passenger Cars')
plt.ylabel('Amount of Crashes')
plt.xlabel('Year')
plt.xticks(ticks = list(range(10)), labels = np.arange(2013,2023,1), rotation=45)
plt.legend()
plt.savefig('4Graph_PassengerCars_crashes')

modes_sum = modes.groupby('Type').agg({'Ongevallen': 'sum'})
modes_sum = modes_sum[modes_sum.index != 'Landbouwvoertuig']

translation = {
    'Bestelauto': 'Delivery van',
    'Brom_snorfiet': 'Scooter',
    'Bus': 'Bus',
    'Motor': 'Motorcycle',
    'Personenauto' : 'Passenger car',
    'Voetganger' : 'Pedestrian',
    'Vrachtauto' : 'Truck'
}
modes_sum = modes_sum.rename(index=translation)

plt.figure(figsize=(6, 6))
plt.pie(modes_sum['Ongevallen'], labels=modes_sum.index, autopct='%1.1f%%', startangle=140, pctdistance=0.85)
plt.title('Distribution of Crashes with other Traffic Modes')
plt.axis('equal')
plt.savefig('4Crashes_vs_other_mode');
```

```
In total there are 19377 crashes with other traffic modes
```

Jaar	Type	Ongevallen
2013	Personenauto	360
2014	Personenauto	601
2015	Personenauto	767
2016	Personenauto	903
2017	Personenauto	1547
2018	Personenauto	1684
2019	Personenauto	1612
2020	Personenauto	1424
2021	Personenauto	1404
2022	Personenauto	1733




Distribution of Crashes with other Traffic Modes
Pedestrian
Truck
3.0%
1.7%
Delivery van
8.1%
62.1%
Passenger car
21.7%
Scooter
1.1%
2.3%
Bus
Motorcycle

5. Crashes for the road section with speed limits

```
speed =  pd.read_excel('5.Speed2.0.xlsx', index_col='Jaar')
speed = speed.drop(columns = ['Plaatsnaam', 'Fiets (betrokkenheid)'])

print('The database has',speed.Ongevallen.sum(), 'crashes with information about maximum speed limit')

speed_grouped = speed.groupby(['Jaar', 'Maximum snelheid (weg)']).sum()
speed_per_year = speed_grouped.loc[speed_grouped.groupby('Jaar')['Ongevallen'].idxmax()]

display(speed_grouped, speed_per_year);
```

```
The database has 20722 crashes with information about maximum speed limit
```

Jaar	Maximum snelheid (weg)	Ongevallen
2013	15 km/u	2
	30 km/u	60
	50 km/u	403
	60 km/u	2
	70 km/u	3
...	...	...
2022	30 km/u	1099
	50 km/u	2257
	60 km/u	3
	70 km/u	5
	80 km/u	9

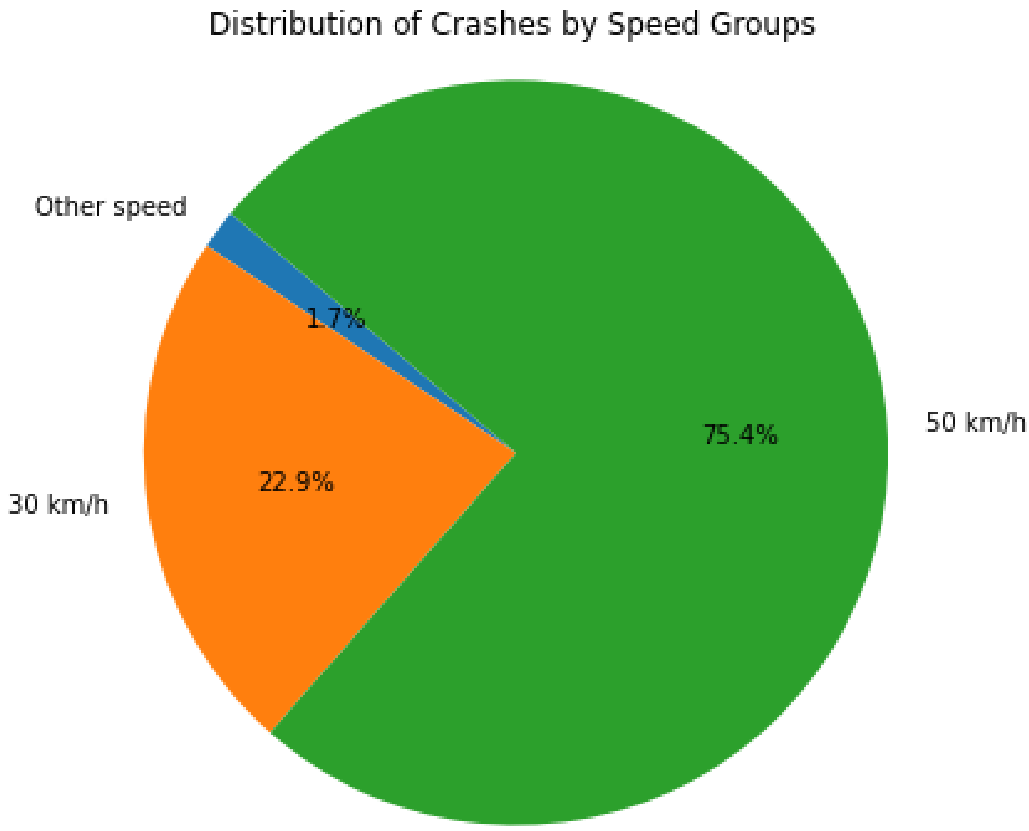
63 rows × 1 columns

Jaar	Maximum snelheid (weg)	Ongevallen
2013	50 km/u	403
2014	50 km/u	728
2015	50 km/u	987
2016	50 km/u	1190
2017	50 km/u	2209
2018	50 km/u	2146
2019	50 km/u	1992
2020	50 km/u	1816
2021	50 km/u	1891
2022	50 km/u	2257

```
grouped_speeds = ['Stapv./woonerf', '15 km/u','60 km/u', '70 km/u', '80 km/u','100 km/u']
separate_speeds = ['30 km/u', '50 km/u']

grouped_df = speed[speed['Maximum snelheid (weg)'].isin(grouped_speeds)].groupby('Jaar').agg({'Ongevallen': 'sum'})
grouped_df['Maximum snelheid (weg)'] = '15, 70, 80 km/u'  # Or any other label you want

separate_df = speed[speed['Maximum snelheid (weg)'].isin(separate_speeds)]
result_df = pd.concat([grouped_df, separate_df])

display(result_df.groupby(['Jaar', 'Maximum snelheid (weg)']).sum())

speed_grouped = result_df.groupby(['Maximum snelheid (weg)']).sum().reset_index()
speed_grouped.iloc[0, 0] = 'Other speed'
display(speed_grouped)

plt.figure(figsize=(6, 6))
plt.pie(speed_grouped['Ongevallen'], labels=speed_grouped['Maximum snelheid (weg)'], autopct='%1.1f%%', startangle=140)
plt.title('Distribution of Crashes by Speed Groups')
plt.axis('equal')
plt.savefig('5Max_Speed_Pie');
```

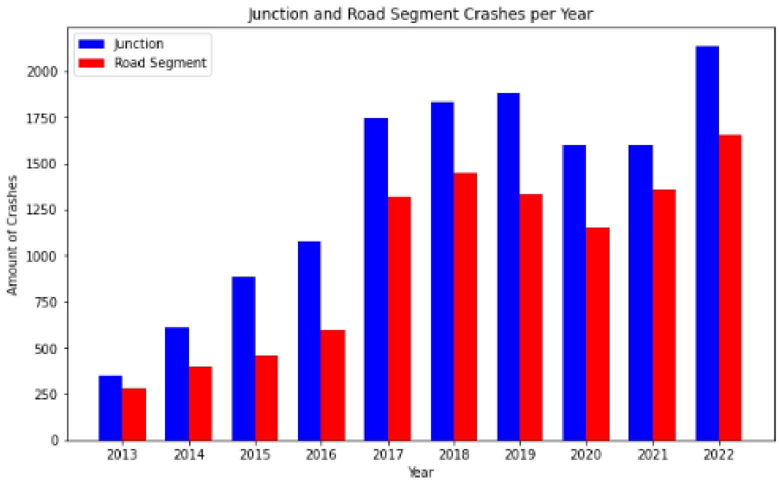
		Ongevallen
Jaar	Maximum snelheid (weg)	
2013	15, 70, 80 km/u	9
	30 km/u	60
	50 km/u	403
2014	15, 70, 80 km/u	8
	30 km/u	125
	50 km/u	728
2015	15, 70, 80 km/u	3
	30 km/u	163
	50 km/u	987
2016	15, 70, 80 km/u	11
	30 km/u	208
	50 km/u	1190
2017	15, 70, 80 km/u	56
	30 km/u	582
	50 km/u	2209
2018	15, 70, 80 km/u	44
	30 km/u	532
	50 km/u	2146
2019	15, 70, 80 km/u	44
	30 km/u	599
	50 km/u	1992
2020	15, 70, 80 km/u	51
	30 km/u	631
	50 km/u	1816
2021	15, 70, 80 km/u	61
	30 km/u	745
	50 km/u	1891
2022	15, 70, 80 km/u	71
	30 km/u	1099
	50 km/u	2257

	Maximum snelheid (weg)	Ongevallen
0	Other speed	358
1	30 km/h	4744
2	50 km/h	15619



6. Crashes based on the road section type

```
section_type = pd.read_excel('6.section_type2.0.xlsx', index_col='Jaar')
section_type = section_type.loc[section_type['Fiets (betrokkenheid)'] == '1 of meer']
section_type = section_type.drop(columns = ['Plaatsnaam', 'Fiets (betrokkenheid)'])
section_type_grouped = section_type.groupby(['Jaar', 'Ernst ongeval', 'Locatie (wegvak of kruispunt)']).sum()

kruispunt = section_type[section_type['Locatie (wegvak of kruispunt)'] == 'Kruispunt']
kruispunt_grouped = kruispunt.groupby(['Jaar']).sum()
wegvak = section_type[section_type['Locatie (wegvak of kruispunt)'] == 'Wegvak']
wegvak_grouped = wegvak.groupby(['Jaar']).sum()

section_type2 = pd.DataFrame({
    'Junction': kruispunt_grouped.Ongevallen,
    'Road_segment': wegvak_grouped.Ongevallen})

fig, ax = plt.subplots(figsize=(10, 6))

bar_width = 0.35
r1 = np.arange(len(section_type2['Junction']))
r2 = [x + bar_width for x in r1]

ax.bar(r1, section_type2['Junction'], color='b', width=bar_width, label='Junction')
ax.bar(r2, section_type2['Road_segment'], color='r', width=bar_width, label='Road Segment')

ax.set_xlabel('Year')
ax.set_ylabel('Amount of Crashes')
ax.set_title('Junction and Road Segment Crashes per Year')
ax.set_xticks([r + bar_width/2 for r in range(len(section_type2['Junction']))])
ax.set_xticklabels(section_type2.index)
ax.legend()
plt.savefig('6Road_Section_type');
```




```
max_crashes_per_year = section_type_grouped.loc[section_type_grouped.groupby('Jaar')['Ongevallen'].idxmax()]
display(max_crashes_per_year)
```

			Ongevallen
Jaar	**Ernst ongeval**	**Locatie (wegvak of kruispunt)**	
2013	**Letsel**	**Kruispunt**	239
2014	**Letsel**	**Kruispunt**	332
2015	**Letsel**	**Kruispunt**	482
2016	**Letsel**	**Kruispunt**	588
2017	**Uitsluitend materiële schade**	**Kruispunt**	963
2018	**Uitsluitend materiële schade**	**Kruispunt**	1079
2019	**Uitsluitend materiële schade**	**Kruispunt**	1124
2020	**Uitsluitend materiële schade**	**Kruispunt**	877
2021	**Uitsluitend materiële schade**	**Kruispunt**	839
2022	**Uitsluitend materiële schade**	**Kruispunt**	1109

7. Crashes based on the junction type

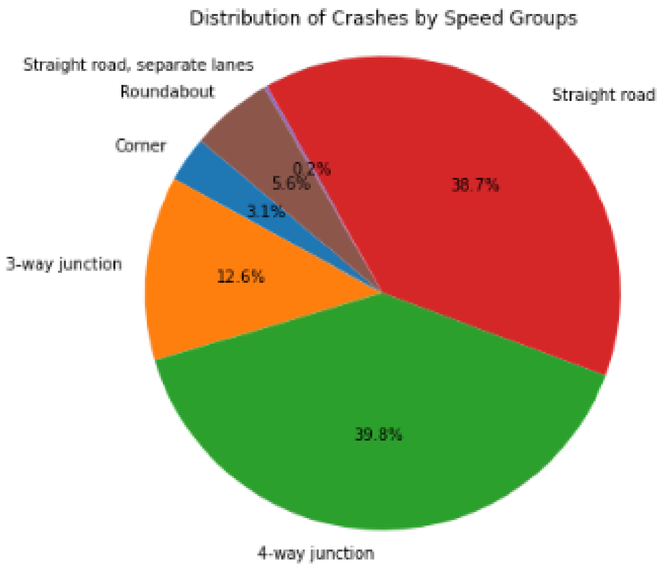
```
junction = pd.read_excel('7.Junction.xlsx', index_col='Jaar')
junction = junction.drop(columns = ['Plaatsnaam', 'Fiets (betrokkenheid)'])

junction_grouped = junction.groupby(['Jaar', 'Wegsituatie']).sum()

max_junction_per_year = junction_grouped.loc[junction_grouped.groupby('Jaar')['Ongevallen'].idxmax()]
junction_sum = junction.groupby('Wegsituatie').agg({'Ongevallen': 'sum'})

display(junction_grouped)
display(max_junction_per_year)
translation = {
    'Bocht': 'Corner',
    'Kruispunt, 3 takken': '3-way junction',
    'Kruispunt, 4 takken': '4-way junction',
    'Rechte weg': 'Straight road',
    'Rechte weg, gescheiden rijbanen' : 'Straight road, separate lanes',
    'Rotonde' : 'Roundabout'
}
junction_sum = junction_sum.rename(index=translation)
display(junction_sum)

plt.figure(figsize=(6, 6))
plt.pie(junction_sum['Ongevallen'], labels=junction_sum.index, autopct='%1.1f%%', startangle=140)
plt.title('Distribution of Crashes by Speed Groups')
plt.axis('equal')
plt.show();
```

		Ongevallen
Jaar	**Wegsituatie**	
2013	**Bocht**	11
	Kruispunt, 3 takken	76
	Kruispunt, 4 takken	236
	Rechte weg	131
	Rechte weg, gescheiden rijbanen	26
	Rotonde	31
2014	**Bocht**	30
	Kruispunt, 3 takken	148
	Kruispunt, 4 takken	396
	Rechte weg	314
	Rechte weg, gescheiden rijbanen	15
	Rotonde	64
2015	**Bocht**	42
	Kruispunt, 3 takken	168
	Kruispunt, 4 takken	627
	Rechte weg	412
	Rotonde	73
2016	**Bocht**	45
	Kruispunt, 3 takken	197
	Kruispunt, 4 takken	758
	Rechte weg	553
	Rotonde	86
2017	**Bocht**	83
	Kruispunt, 3 takken	333
	Kruispunt, 4 takken	1249
	Rechte weg	1163
	Rotonde	156
2018	**Bocht**	100
	Kruispunt, 3 takken	373
	Kruispunt, 4 takken	1294
	Rechte weg	1274
	Rotonde	165

2019	Bocht	110
	Kruispunt, 3 takken	326
	Kruispunt, 4 takken	1265
	Rechte weg	1237
	Rotonde	169
2020	Bocht	80
	Kruispunt, 3 takken	312
	Kruispunt, 4 takken	1044
	Rechte weg	1044
	Rotonde	165
2021	Bocht	102
	Kruispunt, 3 takken	409
	Kruispunt, 4 takken	938
	Rechte weg	1186
	Rechte weg, gescheiden rijbanen	8
	Rotonde	162
2022	Bocht	107
	Kruispunt, 3 takken	528
	Kruispunt, 4 takken	1259
	Rechte weg	1517
	Rechte weg, gescheiden rijbanen	2
	Rotonde	201

		Ongevallen
Jaar	Wegsituatie	
2013	Kruispunt, 4 takken	236
2014	Kruispunt, 4 takken	396
2015	Kruispunt, 4 takken	627
2016	Kruispunt, 4 takken	758
2017	Kruispunt, 4 takken	1249
2018	Kruispunt, 4 takken	1294
2019	Kruispunt, 4 takken	1265
2020	Kruispunt, 4 takken	1044
2021	Rechte weg	1186
2022	Rechte weg	1517

Wegsituatie	Ongevallen
Corner	710
3-way junction	2870
4-way junction	9066
Straight road	8831
Straight road, separate lanes	51
Roundabout	1272



C Python script correlation test

```
[1]: import pandas as pd
     import statsmodels.api as sm
     from statsmodels.formula.api import mnlogit
     import scipy.stats as stats
```

1. Severity vs road segment

```
[2]: chi1 = pd.read_excel('chi1.xlsx')
     chi1 = chi1.drop(columns=['Fiets (betrokkenheid)', 'Plaatsnaam', 'Jaar'])

     aggregated_df = chi1.groupby(['Ernst ongeval', 'Locatie (wegvak of kruispunt)']).agg({'Ongevallen': 'sum'})
     contingency_table = pd.pivot_table(aggregated_df, values='Ongevallen', index='Ernst ongeval', columns='Locatie (wegvak of kruispunt)', aggfunc='sum', fill_value=0)
     display(contingency_table)

     chi2, p, dof, expected = stats.chi2_contingency(contingency_table)
     print("Chi-square:", chi2)
     print("P-value:", p)
     print("Degree of Freedome:", dof)
     print("Expected outcome:\n", expected)
```

Locatie (wegvak of kruispunt)	Kruispunt	Wegvak
Ernst ongeval		
Dodelijk	79	45
Letsel	6397	5183
Uitsluitend materiële schade	7234	4771

```
Chi-square: 62.59077148957819
P-value: 2.562044460444766e-14
Degree of Freedome: 2
Expected outcome:
 [[  71.70441604   52.29558396]
 [6696.26724029 4883.73275971]
 [6942.02834367 5062.97165633]]
```

2. Severity vs Junction type

2. Severity vs Junction type

```
[3]: chi2 = pd.read_excel('chi2.xlsx')
     chi2 = chi2.drop(columns=['Fiets (betrokkenheid)', 'Plaatsnaam', 'Jaar'])

     aggregated_df2 = chi2.groupby(['Ernst ongeval', 'Wegsituatie']).agg({'Ongevallen': 'sum'})

     contingency_table2 = pd.pivot_table(aggregated_df2, values='Ongevallen', index='Ernst ongeval', columns='Wegsituatie', aggfunc='sum', fill_value=0)
     display(contingency_table2)

     chi2, p, dof, expected = stats.chi2_contingency(contingency_table2)
     print("Chi-square:", chi2)
     print("P-value:", p)
     print("Degree of Freedome:", dof)
     print("Expected outcome:\n", expected)
```

Wegsituatie	Kruispunt, 3 takken	Kruispunt, 4 takken	Rotonde
Ernst ongeval			
Dodelijk	15	61	2
Letsel	1422	4199	530
Uitsluitend materiële schade	1433	4806	740

```
Chi-square: 28.965882434338976
P-value: 7.943136258302515e-06
Degree of Freedome: 4
Expected outcome:
 [[  16.9488189    53.53937008    7.51181102]
 [1336.56647486 4222.05981224  592.3737129 ]
 [1516.48470624 4790.40081769  672.11447608]]
```

3. Severity vs weather

```
[4]: chi3 = pd.read_excel('chi3.xlsx')
     chi3 = chi3.drop(columns=['Fiets (betrokkenheid)', 'Plaatsnaam', 'Jaar'])

     aggregated_df3 = chi3.groupby(['Ernst ongeval', 'Weer']).agg({'Ongevallen': 'sum'})

     contingency_table3 = pd.pivot_table(aggregated_df3, values='Ongevallen', index='Ernst ongeval', columns='Weer', aggfunc='sum', fill_value=0)
     display(contingency_table3)

     chi2, p, dof, expected = stats.chi2_contingency(contingency_table3)
     print("Chi-square:", chi2)
     print("P-value:", p)
     print("Degree of Freedome:", dof)
     print("Expected outcome:\n", expected)
```

Weer	Droog	Harde windstoten	Mist	Regen	Sneeuw/hagel
Ernst ongeval					
Dodelijk	107	0	0	8	0
Letsel	9337	23	28	1302	25
Uitsluitend materiële schade	9678	17	20	1364	29

```
Chi-square: 6.829436768712287
P-value: 0.555142178064969
Degree of Freedome: 8
Expected outcome:
 [[1.00238399e+02 2.09681831e-01 2.51618197e-01 1.40172304e+01
  2.83070471e-01]
 [9.33960388e+03 1.95368767e+01 2.34442520e+01 1.30604020e+03
  2.63747835e+01]
 [9.68215772e+03 2.02534415e+01 2.43041298e+01 1.35394257e+03
  2.73421460e+01]]
```

With less data

```
[5]: chi3 = pd.read_excel('chi3.xlsx')
chi3 = chi3.drop(columns=['Fiets (betrokkenheid)', 'Plaatsnaam', 'Jaar'])

chi3 = chi3[~chi3['Weer'].isin(['Harde windstoten', 'Mist', 'Sneeuw/hagel'])]


aggregated_df3 = chi3.groupby(['Ernst ongeval', 'Weer']).agg({'Ongevallen': 'sum'})

contingency_table3 = pd.pivot_table(aggregated_df3, values='Ongevallen', index='Ernst ongeval', columns='Weer', aggfunc='sum', fill_value=0)
display(contingency_table3)

chi2, p, dof, expected = stats.chi2_contingency(contingency_table3)
print("Chi-square:", chi2)
print("P-value:", p)
print("Degree of Freedome:", dof)
print("Expected outcome:\n", expected)
```

Weer	Droog	Regen
Ernst ongeval		
Dodelijk	107	8
Letsel	9337	1302
Uitsluitend materiële schade	9678	1364

```
Chi-square: 3.0970396377527227
P-value: 0.21256237189677407
Degree of Freedome: 2
Expected outcome:
 [[ 100.89144797   14.10855203]
 [9333.77491283 1305.22508717]
 [9687.3336392  1354.6663608 ]]
```

4. Accident severity vs speed

```
[6]: chi4 = pd.read_excel('chi4.xlsx')
chi4 = chi4.drop(columns=['Fiets (betrokkenheid)', 'Plaatsnaam', 'Jaar'])

mapping_speed = {
    '15 km/u': 0,
    '30 km/u': 1,
    '50 km/u': 2,
    '60 km/u': 3,
    '70 km/u': 3,
    '80 km/u': 3,
    '100 km/u': 3
}
chi4['Maximum snelheid (weg)'] = chi4['Maximum snelheid (weg)'].map(mapping_speed)

aggregated_df4 = chi4.groupby(['Ernst ongeval', 'Maximum snelheid (weg)']).agg({'Ongevallen': 'sum'})
contingency_table4 = pd.pivot_table(aggregated_df4, values='Ongevallen', index='Ernst ongeval', columns='Maximum snelheid (weg)', aggfunc='sum', fill_value=0)

display(contingency_table4)

chi2, p, dof, expected = stats.chi2_contingency(contingency_table4)
print("Chi-square:", chi2)
print("P-value:", p)
print("Degree of Freedome:", dof)
print("Expected outcome:\n", expected)
```

Maximum snelheid (weg)	0	1	2	3
Ernst ongeval				
Dodelijk	0	15	95	3
Letsel	131	2277	7640	79
Uitsluitend materiële schade	113	2452	7884	31

```
Chi-square: 42.42153916695183
P-value: 1.5179465981167692e-07
Degree of Freedome: 6
Expected outcome:
 [[1.33069498e+00 2.58722008e+01 8.51808398e+01 6.16264479e-01]
 [1.19256178e+02 2.31865290e+03 7.63386163e+03 5.52292954e+01]
 [1.23413127e+02 2.39947490e+03 7.89995753e+03 5.71544402e+01]]
```

5.Light conditions vs severity accident

```
[7]: chi5 = pd.read_excel('chi5.xlsx')
chi5 = chi5.drop(columns=['Fiets (betrokkenheid)', 'Plaatsnaam', 'Jaar'])

aggregated_df5 = chi5.groupby(['Ernst ongeval', 'Lichtgesteldheid']).agg({'Ongevallen': 'sum'})
contingency_table5 = pd.pivot_table(aggregated_df5, values='Ongevallen', index='Ernst ongeval', columns='Lichtgesteldheid', aggfunc='sum', fill_value=0)
display(contingency_table5)

chi2, p, dof, expected = stats.chi2_contingency(contingency_table5)
print("Chi-square:", chi2)
print("P-value:", p)
print("Degree of Freedome:", dof)
print("Expected outcome:\n", expected)
```

Lichtgesteldheid	Daglicht	Duisternis	Schemer
Ernst ongeval			
Dodelijk	90	27	7
Letsel	8677	2271	631
Uitsluitend materiële schade	8905	2489	611

```
Chi-square: 5.8201703483049965
P-value: 0.21298660198674588
Degree of Freedome: 4
Expected outcome:
 [[9.24298971e+01 2.50374557e+01 6.53264721e+00]
 [8.63101434e+03 2.33797338e+03 6.10012274e+02]
 [8.94855576e+03 2.42398916e+03 6.32455078e+02]]
```

6. Junction type vs Max speed

```
[8]: chi8 = pd.read_excel('Junction_vs_speed.xlsx')
chi8 = chi8.drop(columns=['Fiets (betrokkenheid)', 'Plaatsnaam', 'Jaar'])

values = ['60 km/u', '70 km/u','80 km/u', '100 km/u']
chi8 = chi8[~chi8['Maximum snelheid (weg)'].isin(values)]

aggregated_df8= chi8.groupby(['Maximum snelheid (weg)', 'Wegsituatie']).agg({'Ongevallen': 'sum'})
contingency_table8 = pd.pivot_table(aggregated_df8, values='Ongevallen', index='Maximum snelheid (weg)', columns='Wegsituatie', aggfunc='sum', fill_value=0)
display(contingency_table8)

chi2, p, dof, expected = stats.chi2_contingency(contingency_table8)
print("Chi-square:", chi2)
print("P-value:", p)
print("Degree of Freedome:", dof)
```

Wegsituatie	Bocht	Kruispunt, 3 takken	Kruispunt, 4 takken	Rechte weg	Rechte weg, gescheiden rijbanen	Rotonde
Maximum snelheid (weg)						
15 km/u	11	18	53	156	0	3
30 km/u	208	618	1426	2242	12	64
50 km/u	352	1905	6697	5218	35	1117

```
Chi-square: 703.9150368093844
P-value: 9.067579988323747e-145
Degree of Freedome: 10
```

7. Junction type vs Lighting

```
[9]: chi9 = pd.read_excel('Junction_vs_light.xlsx')
chi9 = chi9.drop(columns=['Fiets (betrokkenheid)', 'Plaatsnaam', 'Jaar'])

aggregated_df9= chi9.groupby(['Lichtgesteldheid', 'Wegsituatie']).agg({'Ongevallen': 'sum'})
contingency_table9= pd.pivot_table(aggregated_df9, values='Ongevallen', index='Lichtgesteldheid', columns='Wegsituatie', aggfunc='sum', fill_value=0)
display(contingency_table9)

chi2, p, dof, expected = stats.chi2_contingency(contingency_table9)
print("Chi-square:", chi2)
print("P-value:", p)
print("Degree of Freedome:", dof)
```

Wegsituatie	Bocht	Kruispunt, 3 takken	Kruispunt, 4 takken	Rechte weg	Rechte weg, gescheiden rijbanen	Rotonde
Lichtgesteldheid						
Daglicht	526	2187	6774	6537	36	936
Duisternis	146	527	1824	1852	8	252
Schemer	38	157	468	446	7	84

```
Chi-square: 22.241173346261547
P-value: 0.013922025092799703
Degree of Freedome: 10
```

8. weather vs Max speed

```
[10]: chi10 = pd.read_excel('Weather_vs_speed.xlsx')
chi10 = chi10.drop(columns=['Fiets (betrokkenheid)', 'Plaatsnaam', 'Jaar'])

# values = ['60 km/u', '70 km/u','80 km/u', '100 km/u']
# chi10 = chi10[~chi10['Maximum snelheid (weg)'].isin(values)]

aggregated_df10= chi10.groupby(['Maximum snelheid (weg)', 'Weer']).agg({'Ongevallen': 'sum'})
contingency_table10 = pd.pivot_table(aggregated_df10, values='Ongevallen', index='Maximum snelheid (weg)', columns='Weer', aggfunc='sum', fill_value=0)
display(contingency_table10)

chi2, p, dof, expected = stats.chi2_contingency(contingency_table10)
print("Chi-square:", chi2)
print("P-value:", p)
print("Degree of Freedome:", dof)
```

Weer	Droog	Harde windstoten	Mist	Regen	Sneeuw/hagel
Maximum snelheid (weg)					
100 km/u	9	0	0	1	0
15 km/u	204	1	0	27	1
30 km/u	3839	10	8	534	11
50 km/u	12813	24	36	1845	38
60 km/u	25	0	0	4	0
70 km/u	29	0	0	2	0
80 km/u	34	0	0	3	0

```
Chi-square: 6.082129260502645
P-value: 0.9999190280609055
Degree of Freedome: 24
```

9. Max speed vs Lighting

```
[11]: chi12 = pd.read_excel('Speed_vs_light.xlsx')
chi12 = chi12.drop(columns=['Fiets (betrokkenheid)', 'Plaatsnaam', 'Jaar'])

# values = ['60 km/u', '70 km/u','80 km/u', '100 km/u', '130 km/u']
# chi12 = chi12[~chi12['Maximum snelheid (weg)'].isin(values)]

aggregated_df12= chi12.groupby(['Maximum snelheid (weg)', 'Lichtgesteldheid']).agg({'Ongevallen': 'sum'})
contingency_table12 = pd.pivot_table(aggregated_df12, values='Ongevallen', index='Maximum snelheid (weg)', columns='Lichtgesteldheid', aggfunc='sum', fill_value=0)

chi2, p, dof, expected = stats.chi2_contingency(contingency_table12)
print("Chi-square:", chi2)
print("P-value:", p)
print("Degree of Freedome:", dof)
```

```
Chi-square: 22.833371835891167
P-value: 0.06303567369534283
Degree of Freedome: 14
```

D Python script multinomial logistic regression

```
[1]: import pandas as pd
     import statsmodels.api as sm
     from statsmodels.formula.api import mnlogit
```

```
[2]: data = pd.read_excel('2013-2022.xlsx')
     data = data.drop(columns=['Fiets (betrokkenheid)', 'Plaatsnaam', 'Jaar'])
     # display(data)
```

```
[3]: duplicated_rows = []

     for index, row in data.iterrows():
         if row['Ongevallen'] > 1:
             for i in range(row['Ongevallen']):
                 duplicated_row = row.copy()
                 duplicated_row['Ongevallen'] = 1
                 duplicated_rows.append(duplicated_row)
         else:
             duplicated_rows.append(row)

     new_df = pd.DataFrame(duplicated_rows)

     new_df.reset_index(drop=True, inplace=True)
     new_df = new_df.drop(columns='Ongevallen')
```

```
[4]: mapping_severity = {
         'Uitsluitend materiële schade': 0,
         'Letsel': 1,
         'Dodelijk': 2
     }

     new_df['Ernst ongeval'] = new_df['Ernst ongeval'].map(mapping_severity)

     mapping_area = {
         'Binnen bebouwde kom' : 0,
         'Buiten bebouwde kom': 1
     }

     new_df['Binnen/buiten bebouwde kom'] = new_df['Binnen/buiten bebouwde kom'].map(mapping_area)

     mapping_light = {
         'Daglicht' : 0,
         'Duisternis' : 1,
         'Schemer' : 2
     }
```

```python
new_df['Lichtgesteldheid'] = new_df['Lichtgesteldheid'].map(mapping_light)

mapping_speed = {
    '15 km/u': 0,
    '30 km/u': 1,
    '50 km/u': 2,
    '60 km/u': 3,
    '70 km/u': 3,
    '80 km/u': 3,
    '100 km/u': 3
}

new_df['Maximum snelheid (weg)'] = new_df['Maximum snelheid (weg)'].map(mapping_speed)

mapping_weer = {
    'Droog': 0,
    'Regen': 1,
    'Mist': 1,
    'Sneeuw/hagel': 2,
    'Harde windstoten' : 3
}

new_df['Weer'] = new_df['Weer'].map(mapping_weer)

['Rechte weg' 'Kruispunt, 4 takken' 'Kruispunt, 3 takken' 'Rotonde'
 'Rechte weg, gescheiden rijbanen' 'Bocht']

mapping_road_conditions = {
    'Droog' : 0,
    'Nat/vochtig' : 1,
    'Sneeuw/ijzel':2
}

new_df['Wegdek toestand'] = new_df['Wegdek toestand'].map(mapping_road_conditions)

mapping_road_situation = {
    'Rechte weg': 0,
    'Rechte weg, gescheiden rijbanen' : 0,
    'Bocht' : 0,
    'Rotonde': 1,
    'Kruispunt, 3 takken': 1,
    'Kruispunt, 4 takken': 1
}

new_df['Wegsituatie'] = new_df['Wegsituatie'].map(mapping_road_situation)
```

```
[5]: display(new_df)
```

	Ernst ongeval	Binnen/buiten bebouwde kom	Lichtgesteldheid	Maximum snelheid (weg)	Weer	Wegdek toestand	Wegsituatie
0	2	0	0	2	0	0	0
1	2	0	2	2	0	0	1
2	2	1	0	3	0	0	1
3	2	0	0	2	0	0	1
4	2	0	0	2	0	0	0
...	...	...	...	...	...	...	...
18565	1	0	0	2	0	0	0
18566	1	0	0	2	0	0	0
18567	1	0	0	2	0	0	0
18568	1	0	0	2	0	0	0
18569	1	0	0	2	0	0	0

18570 rows × 7 columns

```
[6]: Severity_count = new_df['Ernst ongeval'].value_counts()
     print(Severity_count.sum())
     print(Severity_count)

     Built_up = new_df['Binnen/buiten bebouwde kom'].value_counts()
     print(Built_up.sum())
     print(Built_up)

     light = new_df['Lichtgesteldheid'].value_counts()
     print(light.sum())
     print(light)

     Speed_count = new_df['Maximum snelheid (weg)'].value_counts()
     print(Speed_count.sum())
     print(Speed_count)

     Weather_count = new_df['Weer'].value_counts()
     print(Weather_count)
     print(Weather_count.sum())

     Road_conditions = new_df['Wegdek toestand'].value_counts()
     print(Road_conditions)
     print(Road_conditions.sum())

     Road_situation_count = new_df['Wegsituatie'].value_counts()
     print(Road_situation_count)
     print(Road_situation_count.sum())
```

```
18570
0    9446
1    9018
2     106
Name: Ernst ongeval, dtype: int64
18570
0    18274
1      296
Name: Binnen/buiten bebouwde kom, dtype: int64
18570
0    13833
1     3775
2      962
Name: Lichtgesteldheid, dtype: int64
18570
2    14009
1     4237
0      224
3      100
Name: Maximum snelheid (weg), dtype: int64
0    16079
1     2407
2       50
3       34
Name: Weer, dtype: int64
18570
0    14051
1     4438
2       81
Name: Wegdek toestand, dtype: int64
18570
1    10910
0     7660
Name: Wegsituatie, dtype: int64
18570
```

```
7]: Dependent = new_df[['Ernst ongeval']]
    Independent = new_df[['Binnen/buiten bebouwde kom','Lichtgesteldheid','Maximum snelheid (weg)', 'Weer', 'Wegdek toestand', 'Wegsituatie']]

    independent = sm.add_constant(Independent)
    model = sm.MNLogit(Dependent, independent)
    result = model.fit()

    print(result.summary())
```

```
Optimization terminated successfully.
         Current function value: 0.721117
         Iterations 10
                          MNLogit Regression Results
==============================================================================
Dep. Variable:          Ernst ongeval   No. Observations:                18570
Model:                        MNLogit   Df Residuals:                    18556
Method:                           MLE   Df Model:                           12
Date:                Thu, 13 Jun 2024   Pseudo R-squ.:                0.004124
Time:                        14:58:09   Log-Likelihood:                -13391.
converged:                       True   LL-Null:                       -13447.
Covariance Type:            nonrobust   LLR p-value:                 3.964e-18
==============================================================================================
          Ernst ongeval=1       coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------------
const                      -0.0427      0.059     -0.730      0.465      -0.157       0.072
Binnen/buiten bebouwde kom  0.5819      0.124      4.704      0.000       0.339       0.824
Lichtgesteldheid           -0.0432      0.027     -1.614      0.106      -0.096       0.009
Maximum snelheid (weg)      0.0595      0.032      1.869      0.062      -0.003       0.122
Weer                       -0.0725      0.052     -1.401      0.161      -0.174       0.029
Wegdek toestand             0.1200      0.044      2.738      0.006       0.034       0.206
Wegsituatie                -0.2098      0.030     -6.911      0.000      -0.269      -0.150
----------------------------------------------------------------------------------------------
          Ernst ongeval=2       coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------------
const                      -6.1640      0.527    -11.706      0.000      -7.196      -5.132
Binnen/buiten bebouwde kom  1.5518      0.424      3.663      0.000       0.722       2.382
Lichtgesteldheid            0.1220      0.166      0.733      0.464      -0.204       0.448
Maximum snelheid (weg)      0.8354      0.269      3.111      0.002       0.309       1.362
Weer                       -0.9394      0.419     -2.244      0.025      -1.760      -0.119
Wegdek toestand             0.3240      0.262      1.237      0.216      -0.190       0.837
Wegsituatie                 0.1203      0.206      0.583      0.560      -0.284       0.525
==============================================================================================
```

References

Aarts, Bijleveld, Bos, & Decae. (2022). Ernsitg verkeersgewonden 2022. *SWOV*. Retrieved from https://swov.nl/nl/publicatie/ernstig-verkeersgewonden-2022#:~:text=Op%20basis%20van%20deze%20methode,van%20het%20aantal%20in%202021.

Algurén, B., & Rizzi, M. (2022). In-depth understanding of single bicycle crashes in sweden-crash characteristics, injury types and health outcomes differentiated by gender and age-groups. *Journal of Transport & Health*, *24*, 101320.

Boele-Vos, M. J., Duijvenvoorde, K. v., Doumen, M. J. A., Duivenvoorden, C. W. A. E., Louwerse, W. J. R., & Davidse, R. J. (2017, Aug 01). Crashes involving cyclists aged 50 and over: an in-depth study. *Accident Analysis & Prevention*, *105*, 4-10. (15596)

Boonstra, H. J., & Brakel, J. v. d. (2023). Modelling mobility trends - update including 2022 odin data and covid effects. *The Hague : Statistics Netherlands*. Retrieved from https://www.cbs.nl/-/media/_pdf/2023/44/mobility_trends_2023_report_v2.pdf

CBS. (2022). Meer fietsdoden na eenzijdige ongevallen. *Centraal Bureau voor de Statisiek*. Retrieved from https://www.cbs.nl/nl-nl/nieuws/2022/37/meer-fietsdoden-na-eenzijdige-ongevallen

Dozza, M., & Werneke, J. (2014). Introducing naturalistic cycling data: What factors influence bicyclists' safety in the real world? *Transportation research part F: traffic psychology and behaviour*, *24*, 83–91.

EuropeanCommission. (2020). Next steps towards 'vision zero'. *European Commission*. Retrieved from https://op.europa.eu/en/publication-detail/-/publication/d7ee4b58-4bc5-11ea-8aa5-01aa75ed71a1 doi: 10.2832/261629

Gogola, M. (2018a). Are the e-bikes more dangerous than traditional bicycles? In *2018 xi international science-technical conference automotive safety* (p. 1-4). doi: 10.1109/AUTOSAFE.2018.8373344

Gogola, M. (2018b). Are the e-bikes more dangerous than traditional bicycles? In *2018 xi international science-technical conference automotive safety* (pp. 1–4).

Haas, M. d., & Hamersma, M. (2020). Fietsfeiten: nieuwe inzichten. *Kennisinstituut voor Mobiliteitsbeleid | KiM*. Retrieved from https://swov.nl/nl/publicatie/fietsfeiten-0

Haas, M. d., & L, K. (2023). Fietsfeiten 2023. *Kennisinstituut voor Mobiliteitsbeleid | KiM*. Retrieved from https://www.kimnet.nl/publicaties/publicaties/2023/11/28/fietsfeiten-2023

Hayes, A. (2024). Chi-square (2) statistic: What it is, examples, how and when to use the test. *investopedia*. Retrieved from https://www.investopedia.com/terms/c/chi-square-statistic.asp

Krul, I. (2018). Infographic: fietsongevallen 2018 in beeld. *Veiligheid.nl*.

Krul, I., Valkenberg, H., Asscherman, S., Stam, C., & Klein Wolt, K. (2022). Fietsongevallen en snor-/bromfietsongevallen in nederland; seh-bezoeken: inzicht in oorzaken, gevolgen en risicogroepen. *SWOV*. Retrieved from https://www.veiligheid.nl/sites/default/files/2022-06/Rapportage%20%28Snor-%20en%20brom%29fietsongevallen%20in%20Nederland.pdf

RedactieKW. (2022). wielertoerist uit jabbeke in kritieke toestand nadat hij ineenzakt tijdens fietsritje. *Krant van West-Vlaanderen*. Retrieved from https://

kw.be/nieuws/ongevallen/verkeersongeval/wielertoerist-uit-jabbeke-in-kritieke-toestand-nadat-hij-ineenzakt-tijdens-fietsritje/

Rijkswaterstaat. (2024). Verkeersveiligheid en ongevallencijfers. *Rijkswaterstaat, Ministerie van Infrastructuur en Waterstaat*. Retrieved from https://www.rijkswaterstaat.nl/wegen/wegbeheer/onderzoek/verkeersveiligheid-en-ongevallencijfers#:~:text=In%20het%20databestand%20BRON%20worden,middel%20van%20kenmerkenmeldingen%20of%20incidentmeldingen.

Schepers, J., Weijermars, W., Boele, M., Dijkstra, A., & Bos, N. (2020). Oudere fietsers; iongevallen met oudere fietsers en factoren die daarbij een rol spelen. *SWOV*. Retrieved from https://swov.nl/nl/publicatie/oudere-fietsers-0

Schepers, P., Wolt, K. K., Helbich, M., & Fishman, E. (2020). Safety of e-bikes compared to conventional bicycles: What role does cyclists' health condition play? *Journal of Transport & Health*, *19*, 100961.

SWOV. (2023). Swov-factsheet. *SWOV*. Retrieved from https://swov.nl/nl/factsheet/fietsers

SWOV. (2024). Verkeersongevallen. *SWOV*. Retrieved from https://swov.nl/nl/cijfers/verkeersongevallen

Utriainen, R. (2020). Characteristics of commuters' single-bicycle crashes in insurance data. *Safety*, *6*(1). Retrieved from https://www.mdpi.com/2313-576X/6/1/13

Valkenberg, H., Nijman, S., Schepers, P., Panneman, M., & Klein Wolt, K. (2017). Fietsongevallen in nederland seh-behandelingen 2016. *Centraal Bureau voor de Statisiek*. Retrieved from https://www.cbs.nl/nl-nl/nieuws/2022/37/meer-fietsdoden-na-eenzijdige-ongevallen

van Petegem, J. H., Schepers, P., & Wijlhuizen, G. J. (2021, Jul.). The safety of physically separated cycle tracks compared to marked cycle lanes and mixed traffic conditions in amsterdam. *European Journal of Transport and Infrastructure Research*, *21*(3), 19–37. Retrieved from https://journals.open.tudelft.nl/ejtir/article/view/5283 doi: 10.18757/ejtir.2021.21.3.5283

WHO. (2023). Road traffic injuries. *World Health Organization*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries