TIL6022 Research Proposal

Group: 9

Members:
Xyrian Brooijmans   5079519
Ole de Groot   6332188
Minke Groenendijk   5110343
Maud Hendriks   5178061
Renzo van den Brink   6338488

Research questions

The aim of this application project is to investigate the extent to which certain factors influence the number of traffic accidents within the Netherlands. In doing so, a highly urbanized area within the Netherlands will be compared with a non-urban area. The factors being investigated are as follows: weather conditions at the time of the accident, the number of parties involved in the accident and the maximum speed driven at the time of the accident. There will also be a distinction between the different types of accidents, either fatal, injury or merely material damage. Through data processing, an exploratory analysis and visualization we aim to uncover patterns and trends in accident rates within both densely populated and less urbanized regions. This analysis will provide valuable insights to inform future traffic safety measures.

The main research question reads as follows:

What is the influence of environmental factors on the number of accidents per year within an urban area, compared with a rural area in the Netherlands?

This question will be answered using the following sub-questions:

➢ *What is the influence of weather conditions on the number of accidents per year?*

Through this question, the impact of weather on accidents will be examined. Here, the weather conditions are divided into 6 different categories, which are: dry, rain, fog, snow/hail, high winds and unknown.

➢ *What is the influence of the speed limit on the number of accidents per year?*

This will involve examining the impact of the speed limit of the road of the incident.

➢ *To what extent does the number of parties involved per accident differ between urban and rural areas, and how does this relate to the total number of accidents per year?*

This question focuses on examining how often more than one party is involved in an accident in urban versus rural areas.

Data sets

[Verkeersongevallen - Bestand geRegistreerde Ongevallen Nederland | Data overheid](#)
We plan on using the dataset "Verkeersongevallen - Bestand geRegistreerde Ongevallen Nederland 2014_2023" as compiled by the Dutch Rijkswaterstaat. This dataset uses all publicly available data of road accidents spanning from January 1st 2014 up to December 31st 2023, providing us with 10 years of detailed data. This set consists of two subsets: 'ongevallen.txt' gives detailed information about every incident, with distinctions between weather types, number of involved parties and allowed maximum speed on the incident location, to name a few examples.
The dataset "wegvakken.txt" provides information on incident location; street names, municipalities and road types.

[Inwoners per gemeente | CBS](#)
Next to these primary datasets, we will employ a set "Bevolkingsdichtheid, 2024" from the Dutch CBS in order to distinguish between municipalities and their respective urban density. As we will look at certain degrees of density, we will pick a certain number of municipalities from the top and bottom of the list and focus on all road accidents that occurred in these two areas.

Data pipeline

For the data pipeline, the focus will be on processing and integrating data from the two datasets mentioned above. We will retrieve data from the "Verkeersongevallen - Bestand geRegistreerde Ongevallen Nederland 2014_2023" dataset, which contains detailed accident records over a 10-year period, and the "wegvakken.txt" file, which provides geographical context for each incident. Additionally, population density data from "Bevolkingsdichtheid, 2024" will be incorporated to classify municipalities as either urban or rural.

Data merging will be based on incident location and date to create comprehensive accident records. We will then filter the data to focus on accidents within our predefined geographical boundaries. We will also filter for three accident outcomes (fatal, injury, and property damage), and limit the scope to accidents with recorded weather conditions, speed limits, and party involvement. For weather conditions, we will use the categories provided in the dataset: dry, rain, fog, snow/hail, strong wind gusts, and unknown weather.

The accident data will be enriched by linking it to the population density dataset, categorizing each incident as occurring in either an urban or rural area. Finally, after data preparation, we will perform an exploratory quantitative analysis using statistical methods.

For each factor (weather, speed limit, and number of parties involved), we will create two bar charts: one for urban areas and one for rural areas. Each chart will display four columns representing the different types of accidents: fatal accidents, injury accidents, property

damage, and the total number of accidents. This will result in a total of six charts, enabling a clear comparison between urban and rural areas across all accident types and factors.

Geographical Boundary and time scale

The geographical boundary of the study will be based on the categorization from the dataset *"Bevolkingsdichtheid, 2024"*. A comparison will be made between the most and least urban ("Zeer sterk stedelijk vs. "Niet-stedelijk"), where a selection has been made which municipalities will be considered.

For "Zeer stedelijk" this will be the top 3:
- *Den Haag*
- *Leiden*
- *Haarlem*

For "Niet-stedelijk", we will be selecting more municipalities such that the amount of total combined inhabitants will be close to the three municipalities mentioned above. These will be the 100 least urban municipalities.

The time scale that has been chosen is the full time scale of the dataset "Verkeersongevallen - Bestand geRegistreerde Ongevallen Nederland 2014_2023", spanning from January 1st 2014 up to December 31st 2023, providing us with 10 years of detailed data. This time scale has been chosen to ensure:
- There have been at least 100 accidents in all weather conditions considered
- A comparison can be made with a more diverse amount of speed limits, which have changed over the years.

Note that the timescale over which the study is considered is different from the information from the *"Bevolkingsdichtheid, 2024"*, however the order of the least and especially the order of the most dense municipalities have not changed considerably in the last 10 years. On top of this, *"Bevolkingsdichtheid, 2024"* will solely be used to select the studied municipalities, these do not by definition have to always have been the most and least dense municipalities over the full timespan of the study.