MADE Project:
Connection beetween frequency of bicycle accidents and rainy the weather in German cities

Jean-Philipp Schaller
Friedrich-Alexander Universität Erlangen-Nürnberg
January 10, 2024

Contents

1 Introduction

The project objective is to analyse if and how much the impact of rainy weather and slippery roads affects the accident rate. The project will focus on major German cities, because there are more frequent bicycle accident reports and accurate weather station data.
The motivation is to show if more technologies and advances in better slip-resistant roads or visibility with light indicators on road intersections can reduce the amount of bicycle crashes with personal injuries in rainy weather conditions or if investments in road safety should be used differently.

2 Methods

For this project, eight datasets were used. Seven of these datasets contain the crash reports of accidents with personal injury in the years from 2016 until 2022 from `GOVDATA`. Each year is in a new database.
One Dataset is about the weather data from `Deutscher Wetterdienst`. In this study I focus the findings on `Muenchen-Stadt`, but the built database pipeline can handle databases from other cities from the given source if they provide accurate weather data in the time-frame of the crash data reports.

2.1 Dataset GOVDATA

- Metadata URL GER: Metadaten-GER
- Metadata URL ENG: Metadaten-ENG
- Data URL 2016: Unfallorte2016_EPSG25832_CSV.zip
- Data URL 2017: Unfallorte2017_EPSG25832_CSV.zip
- Data URL 2018: Unfallorte2018_EPSG25832_CSV.zip
- Data URL 2019: Unfallorte2019_EPSG25832_CSV.zip
- Data URL 2020: Unfallorte2020_EPSG25832_CSV.zip
- Data URL 2021: Unfallorte2021_EPSG25832_CSV.zip
- Data URL 2022: Unfallorte2022_EPSG25832_CSV.zip
- URI: https://unfallatlas.statistikportal.de/
- License: `dl-de/by-2-0` https://www.govdata.de/dl-de/by-2-0
- Municipality Code: https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/_inhalt.htm

The Datasets from `GOVDATA` were collected from the sources above and used in the data pipeline. There were many attributes that were not needed for this project that could be discarded. After evaluating and cleaning up the data you can see the metadata of remaining attributes with descriptions in table 1. These points represent the location and time of an accident and if it is involving a bicycle or not. Furthermore the data from the category of severity and type of accidents were also included for further and future analysis of the data.
During the data engineering step the project encountered on a problem. At first it was thought that the databases can be connected hourly, because both databases are saving the datapoints with hourly precision. For `GOVDATA` it was the `USTUNDE` and for the `Deutscher Wetterdienst` it was the `HOUR` attribute that was transformed from MESS_DATUM. The encountered problem was that after closer inspection of `GOVDATA` it does not provide hourly precision due to `UWOCHENTAG` only providing what day of the week it was, for example a sunday, and not the day of the month in numerical form, for example the 23 of January.
It is unknown if this is a sort of anonymisation for the reported accidents or an oversight of the data, since it even uses hourly precision afterwards. Therefore the project has to be developed further on a monthly precision instead of an hourly one.

2.2 Dataset Deutscher Wetterdienst

- Metadata and Data URL: hist.zip
- Data Type: Database File as a TXT. The Metadata is also inside the Zip File.
- License: https://opendata.dwd.de/climate_environment/CDC/Nutzungsbedingungen_German.pdf

The data is from `Deutscher Wetterdienst`, which will be transformed to be able to be compared and complement the data of chapter 2.1. This dataset is provided from `CDC OpenData` and is open source.
Slight modification to the data has to be made. The year, month and day are saved in one attribute called MESS_DATUM. The project transforms this value to have separate attributes for the year, month and day for easier comparison. Furthermore, some entries display a uniform `-999` negative value in the `RAINFALL` attribute. Since negative rainfall is impossible, the project assumes a malfunction and deletes the entry. Some attributes are also not interesting for our projects and are tossed. In table 2 the remaining column attributes after the clean-up and transformation can be seen.

2.3 Building the Pipeline

The project uses ETL (Extract, Transform, Load) pipeline approach. After installing the dependencies in `requirements.txt`, the files needed to run the pipeline are located in the `project` folder. The `pipeline.sh` script serves as the entry point for running the python file `pipelinedata.py`, resulting in the extracting and transforming of the eight datasets stored in the folder `data` as SQLite databases. In figure 1 you can see the schematic of the project. After loading the databases the transformation and analysis of the data is done inside the `data-exploration.ipynb` file.

3 Results

3.1 Rainfall

The data of rainfall, in this example in `Muenchen-Stadt` is now grouped by month and summed up. The rainfall can be seen in figure 2.

3.2 Crash data

Afterwards the crash data transformation and analysis part was performed on the data collected from the pipeline. All the data from 2016 until 2022 was used with personal injuries involving a bicycle. The data graph is depicted in figure 3.

3.3 Correlation bicycle accidents and rainfall

Afterwards both of these datasets were merged on a monthly basis, if we have values in the data on both `RAINFALL` in `Muenchen-Stadt` and `TOTALCRASHES` from bicycle accidents in `Oberbayern`. The merged data was used to see if there is a correlation between these two attributes. The result is shown in figure 4. There is a correlation with the value of `0.4885`.

3.4 Correlation all accidents and rainfall

Afterward, the same pipeline was reused to evaluate the correlation between `RAINFALL` in `Muenchen-Stadt` and `TOTALCRASHES` of all accidents in `Oberbayern`. The result is shown in figure 5, showing a correlation with the value of `0.4750`.

3.5 Heat Map

A heat map of the data between `TOTALCRASHES` of bicycle accidents and `RAINFALL` was created in figure 6. This visual representation gives us a good understanding of how rainy weather relates to the frequency of bicycle accidents and how often the data points are represented in our dataset.

4 Conclusions

4.1 Interpretation of the Results

Upon analysing the results from Chapter 3, we can see that the bicycle accidents in the context of rainfall in `Muenchen-Stadt` reveals a notable correlation. Specifically, when focusing on bicycle accidents involving personal injuries from 2016 to 2022, the correlation between rainfall and total bicycle crashes in `Oberbayern` is `0.4885`. This suggests a moderate positive correlation. In other words, as rainfall increases, the number of bicycle accidents tends to rise
Notably, when considering all accidents in `Oberbayern`, not just limited to bicycles, a slightly smaller correlation of `0.4750` is observed. This indicates that the impact of rainfall on overall accident rates is less compared to its influence on bicycle accidents specifically.
The heat map of bicycle incidents in figure 6 also shows that we have a lot of different data points in our project and are not for example saturated with just sunny or rainy months. It also emphasizes the correlation showing less data points in the north-west and south-east part of the heat map.
These results underscore the relationship between rainy weather conditions and accident rates, emphasizing the need for targeted safety measures, especially for cyclists, during rain.

4.2 Limitations

While this study analysed the correlation between rainy weather and traffic accidents, certain limitations need to be acknowledged.

4.2.1 Monthly datapoints

The project relied on monthly weather data, limiting the analysis to monthly patterns like written in chapter 2.1. Hourly data could provide more detailed insights, especially in understanding the immediate impact of weather changes on accident rates. However, due to the absence of being able to tell what day in the month the data was recorded in the `GOVDATA` dataset, the study had to operate on a monthly basis. This can cause counting bicycle accidents as occurring on a rainy month without it being rainy at the specific day.

4.2.2 Geographical variability of weather

Moreover, there can be a great geographical variability of rainy weather. Rainfall patterns can vary significantly even within a small geographical area, and this study focused on point-located data for `Muenchen-Stadt`. This can cause counting bicycle accidents as occurring on a rainy month without it being rainy at the specific location.

4.3 Future scope

Acquiring hourly crash data would improve the accuracy of the project, allowing for a more in-depth examination of its correlation with rainy weather.

5 Tables and Figures

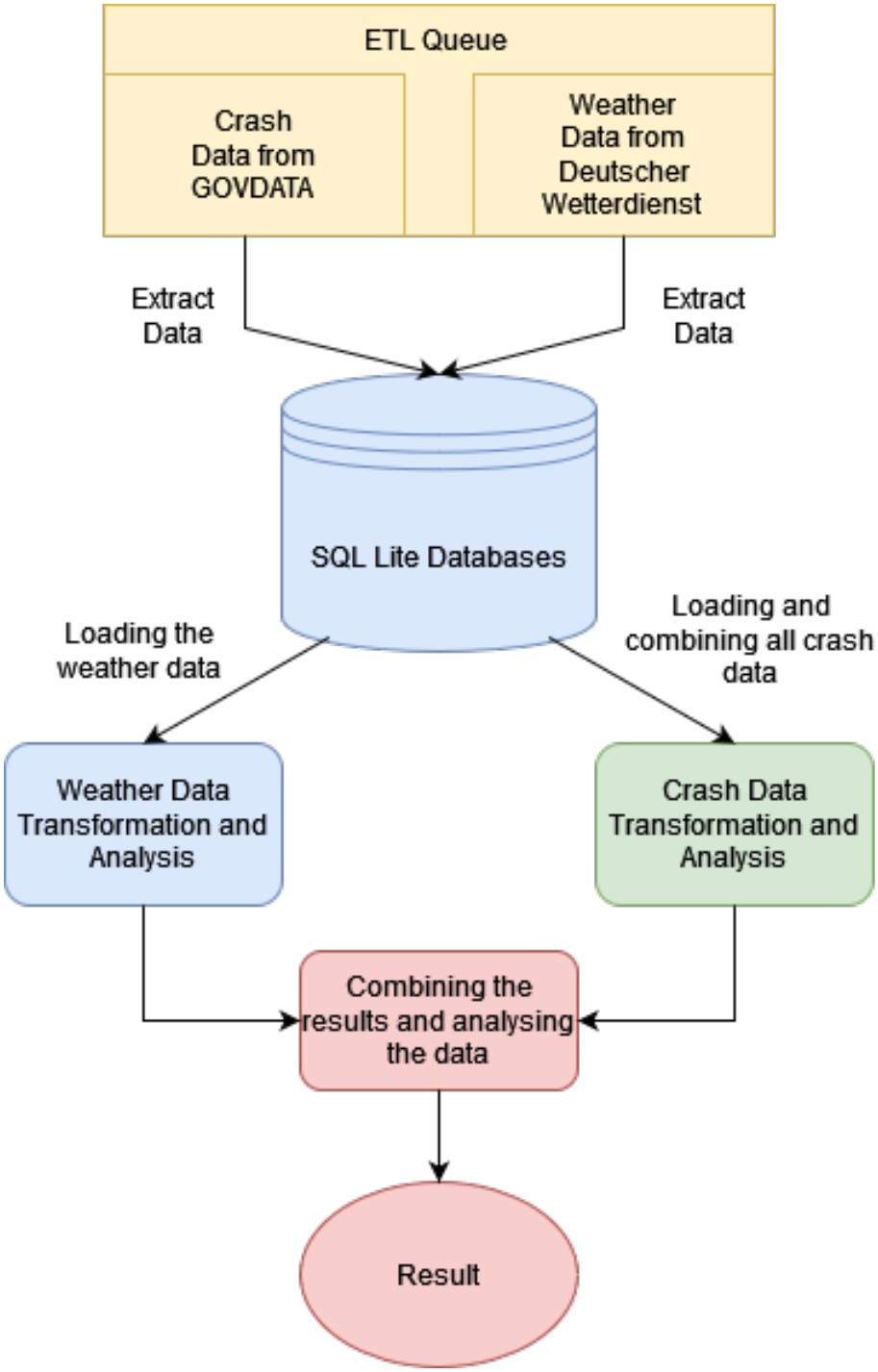

Figure 1: The methodology of the project

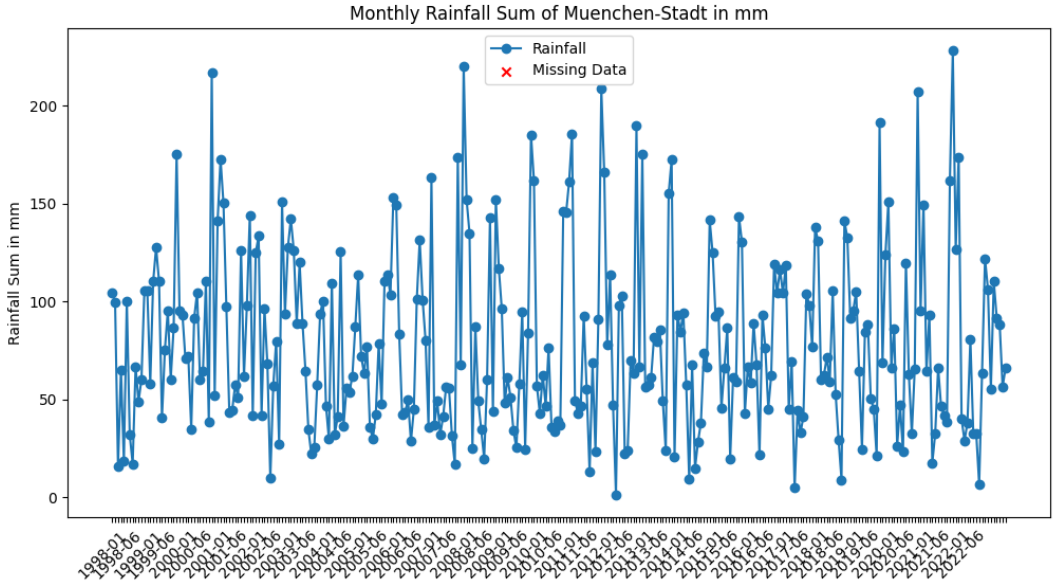

Figure 2: Sum of monthly rainfall in mm from `Deutscher Wetterdienst` in `Muenchen-Stadt`

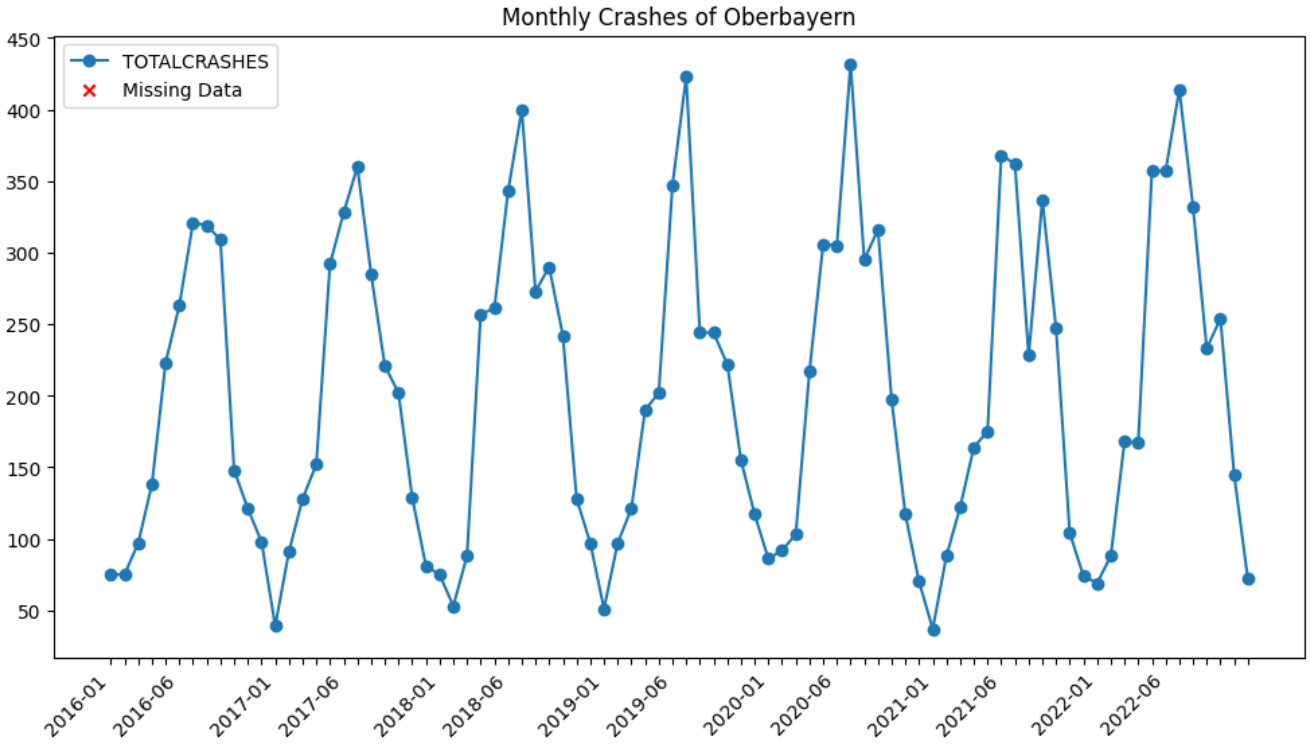

Figure 3: Monthly crashes from `GOVDATA` in `Oberbayern` with personal injuries involving a bicycle

Column_attribute	Description
ObjectID	**Serial number**
ULAND	01 = Schleswig-Holstein 02 = Hamburg 03 = Niedersachsen 04 = Bremen 05 = 'Nordrhein-Westfalen (data as from 2019) 06 = Hessen 07 = Rheinland-Pfalz (data as from 2017) 08 = Baden-Württemberg 09 = Bayern 10 = Saarland (data as from 2017) 11 = Berlin (data as from 2018) 12 = Brandenburg (data as from 2017) 13 = Mecklenburg-Vorpommern (data as from 2020) 14 = Sachsen 15 = Sachsen-Anhalt 16 = Thüringen (data as from 2019)
UREGEZ	**Administrative region**
UKREIS	**Administrative district**
UGEMEINDE	**Form the official Municipality Code. See 2.1 for more details.**
UJAHR	**Year of accident**
UMONAT	**Month of accident**
USTUNDE	**Hour of accident**
UWOCHENTAG	**Day of the week** 1 = sunday 2 = monday 3 = tuesday 4 = wednesday 5 = thursday 6 = friday 7 = saturday
UKATEGORIE	**Road traffic accidents involving personal injury** 1 = Accident with persons killed 2 = Accident with seriously injured 3 = Accident with slightly injured
UART	**Kinds of accidents** 1 = Collision with another vehicle which starts, stops or is stationary 2 = Collision with another vehicle moving ahead or waiting 3 = Collision with another vehicle moving laterally in the same direction 4 = Collision with another oncoming vehicle 5 = Collision with another vehicle which turns into or crosses a road 6 = Collision between vehicle and pedestrian 7 = Collision with an obstacle in the carriageway 8 = Leaving the carriageway to the right 9 = Leaving the carriageway to the left 0 = Accident of another kind
UTYP1	**Type of accidents** 1 = Driving accident 2 = Accident caused by turning off the road 3 = Accident caused by turning into a road or by crossing it 4 = Accident caused by crossing the road 5 = Accident involving stationary 6 = Accident between vehicles moving along in carriageway 7 = Other accident
IstRad	**Accident with bicycle** 0 = Accident without any bicycle involved 1 = accident in which at least one bicycle was involved

Table 1: Metadata of the GOVDATA dataset after cleaning and transformation.

Column_attribute	Description
STATIONS_ID	**The numerical ID of the station**
RAINFALL	**Amount of rain in the recorded hour in mm**
YEAR	**The year of the recorded datapoint. Transformed from the data of MESS_DATUM**
MONTH	**The month of the recorded datapoint. Transformed from the data of MESS_DATUM. Saved as a numerical value between and including 1 to 12**
DAY	**The day of the recorded datapoint. Transformed from the data of MESS_DATUM.**
HOUR	**The hour of the recorded datapoint. Transformed from the data of MESS_DATUM. Uses the 24 hour format**
LOCATION	**Location of the weather station. Provided manually in the data pipeline and has to be extracted from the metadata in 2.2**

Table 2: Metadata of the rain weather dataset after cleaning and transformation. Translated from German to English. For German metadata see 2.2

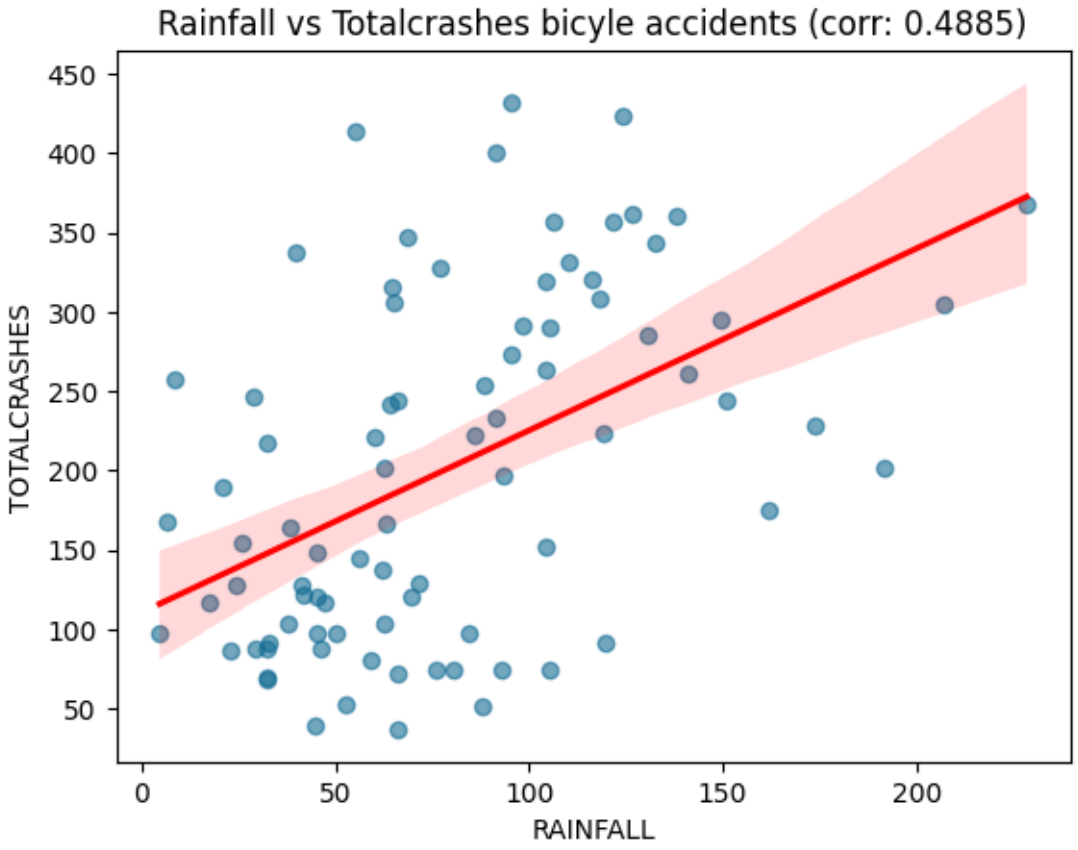

Figure 4: The correlation between rainfall and total crashes involving a bicycle

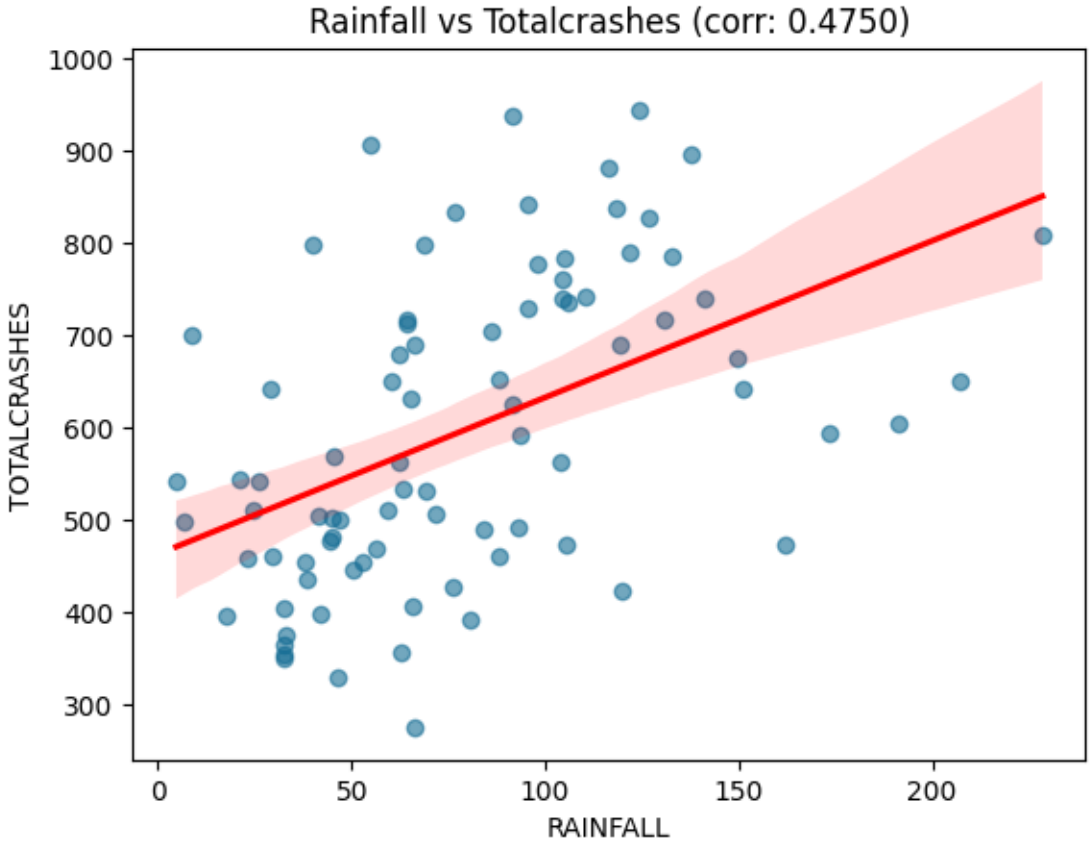

Figure 5: The correlation between rainfall and total crashes involving all accidents

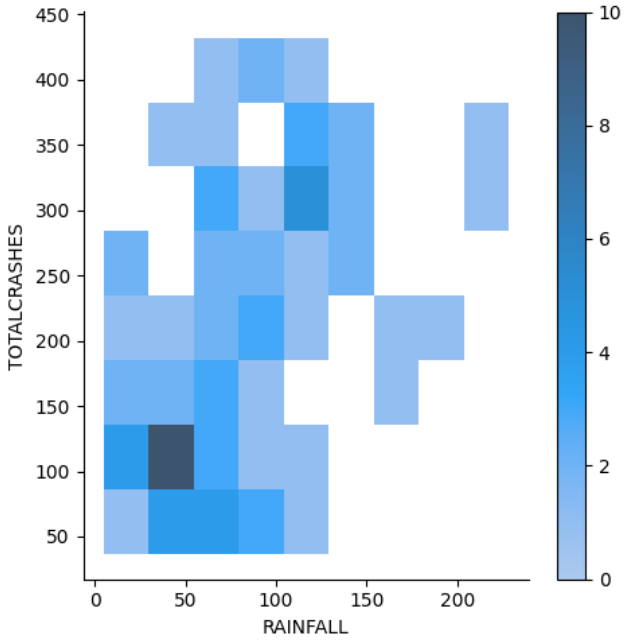

Figure 6: The heatmap between total amount of bicycle crashes in rainy weather and the rainfall