# Data Analysis Project
# P3: Data Analysis and Visualization
# Make Love Not Accidents
# (Group 11)

Edi Zeqiri and Ugur Turhal

University of Basel
Databases (CS244) course
Autumn Semester 2022

## 🌍 Link to map

# Table of Contents

## 1 Introduction:

Speeding is a common and potentially dangerous behavior on public roads and it has been linked to a higher risk of accidents and injuries. In this data analysis project, we aim to examine the relationship between speeding and accident rates on public roads. By analyzing data on the number of speeding cases and the number of accidents occurring on various roads, we hope to gain insights into the prevalence and impact of speeding on public road safety.

To conduct this analysis, we will use data from a range of sources, including traffic and accident records from the local agency of Basel. We will apply statistical techniques to identify patterns and trends in the data and to evaluate the strength of the relationship between speeding and accident rates.

The findings of this data analysis project will have important implications for road safety policy and enforcement, as well as for individual drivers. By understanding the factors that contribute to speeding and accidents on public roads, we can develop strategies to reduce the risks and improve the safety of all road users.

# 2 Datasets

## 2.1 Geschwindigkeitsmonitoring: Einzelmmessung bis 2020

Vehicle speed monitoring dataset of all the years before 2020.

- Source: Dataset 1
- Name: 100097
- Size: 2'559'791'756 bytes
- Format: CSV

## 2.2 Geschwindigkeitsmonitoring: Einzelmmessung ab 2021

Vehicle speed monitoring dataset of all the years after 2021.

- Name: 100200
- Source:Dataset 2
- Size: 3'050'206'951 bytes
- Format: CSV

## 2.3 Strassenverkehrsunfälle

Public road accidents. Mainly parking damage but also other accidents on public roads.

- Name: 100120
- Source:Dataset 3
- Size: 4´2019664´2019895 bytes
- Format: CSV

# 3 Entity Relation Diagram per Data Source

## 3.1 Geschwindigkeitsmonitoring: Einzelmessungen ab 2021

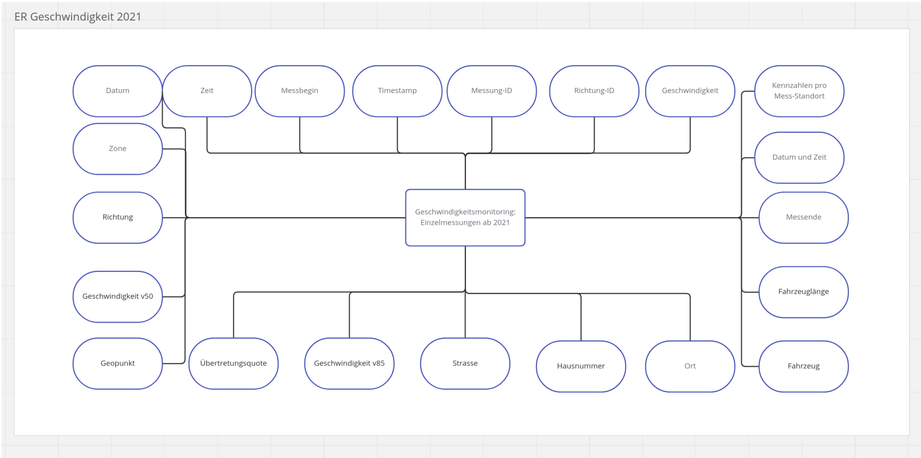


Fig. 1: ER-Geschwindigkeitsmonitoring: Einzelmessungen ab 2021

## 3.2 Geschwindigkeitsmonitoring: Einzelmessungen bis 2020

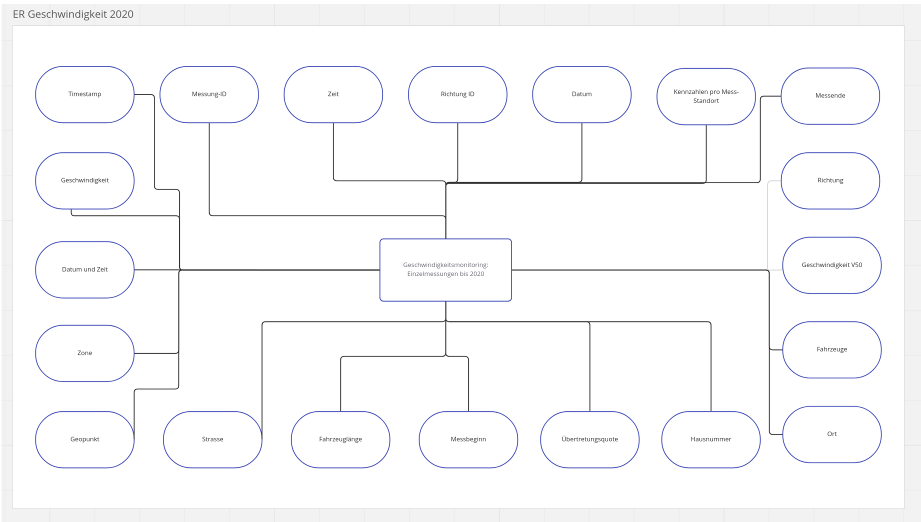


Fig. 2: ER-Geschwindigkeitsmonitoring: Einzelmessungen bis 2020

### 3.3 Strassenverkehrsunfälle

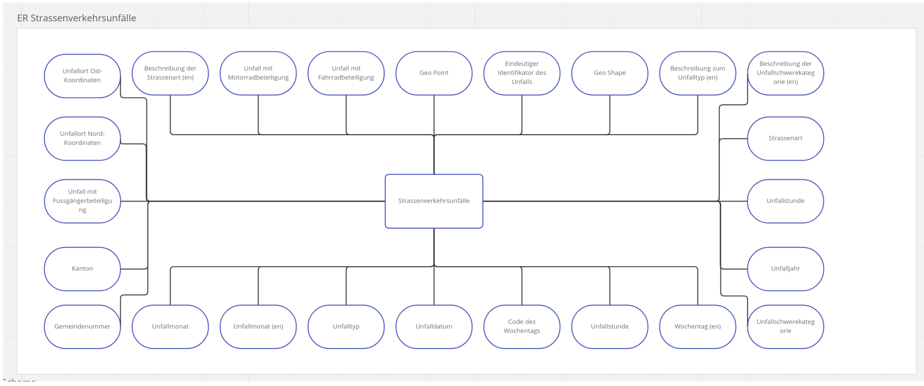


Fig. 3: ER-Strassenverkehrsunfälle

## 4 Integrated ER-Diagram

We have only chosen to display the meaningful attributes which we will consider in our analysis. The other attributes in the respective ER diagrams are not lost. We will include them in the database. Only for the sake of simplicity of this document have we not shown all of them. The same goes for the Relational Schema.

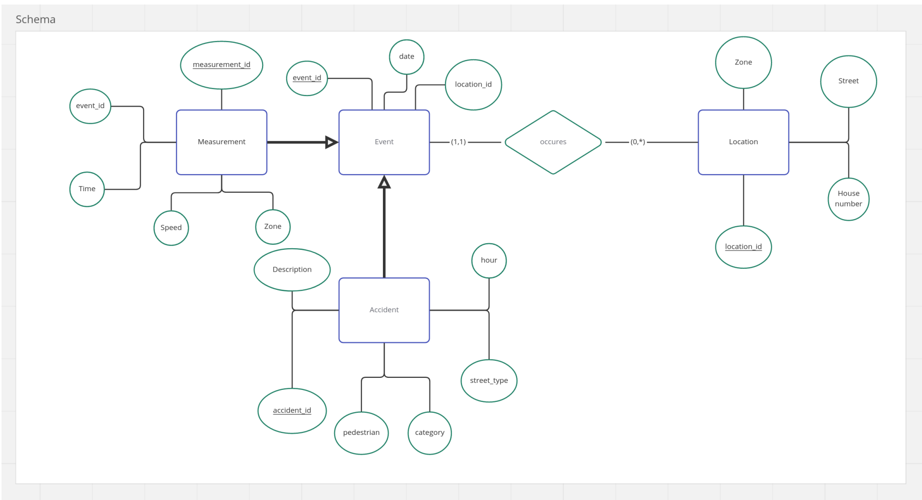


Fig. 4: Integrated ER-Diagram

# 5 Relational Schema

## 5.1 Logical relational schema

Measurement(measurement_id, event_id,Time, Speed, Zone)
Events(event_id, date, location_id)
Location(location_id, Location, Street, House number, Geopoint)
Accidents(accidents_id, event_id, description, category, pedestrain, bicycle, motorcycle, street_type, hour)

## 6 Methods:

This section should describe the methods and techniques used to collect and analyze the data. This might include details about the sample size, data cleaning and preprocessing steps, and statistical or machine-learning techniques applied.

To conduct this data analysis project, we used a range of methods and techniques to collect, integrate, and analyze data on speeding and accident rates on public roads.

First, we sourced data on the number of speeding cases and the number of accidents occurring on various roads from local and national agencies. These data were provided in a range of formats, including CSV and Excel files.

We used MySQL to load and integrate these datasets into a single database, enabling us to analyze the data in a more efficient and standardized way. We also used MySQL to perform basic data cleaning and preprocessing, including the removal of duplicates and formatting of dates and other variables.

Next, we used Python to perform statistical and machine-learning analyses of the data. We used a variety of libraries and packages, including Pandas, NumPy, and scikit- learn, to explore the relationships between speeding and accident rates on public roads. We also used Python to create visualizations of the data, including scatter plots and bar charts, to help illustrate key patterns and trends.

In addition to the data analysis, we also developed a website using Python and the Flask framework to enable users to interactively explore the datasets of speeding and accident cases. The website allows users to view maps and charts showing the distribution of cases by location and time, and to filter the data by various criteria such as road type and weather conditions.

Overall, these methods allowed us to effectively and efficiently analyze and visualize the data on speeding and accident rates on public roads, providing valuable insights into the prevalence and impact of speeding on road safety.

# 7 Results:

## 7.1 Accidents

The Accidents data set consists of these types of accidents:

| **Description** | **Percentage** |
|---|---|
| Accident when parking | 41.61 |
| Accident with skidding or self-accident | 19.90 |
| Accident with rear-end collision | 7.86 |
| Accident involving pedestrian\(s\) | 6.95 |
| Accident when crossing the lane\(s\) | 6.25 |
| Accident when turning-into main road | 5.59 |
| Accident when turning left or right | 4.95 |
| Accident when overtaking or changing lanes | 4.73 |
| Accident with head-on collision | 1.65 |
| Other | 0.39 |

Table 1: Accidents with percentage

Just by the description it might be obvious that not every accident on public roads is related to speeding such as parking accidents. It is not reasonable to assume that parking accidents happen above 10-15 km/h. **Note:** Thus we can disregard 41.61% of the cases in our analysis as they do not involve speeding.

Therefore the other descriptions might involve speeding as a factor of the cause. To verify this claim we need to compare where and when the accidents happened. But first some important key facts need to be analyzed.

| **Category** | **Amount** | **Percent** |
|---|---|---|
| Accident with property damage | 3761 | 64.30 |
| Accident with light injuries | 1495 | 25.55 |
| Accident with severe injuries | 574 | 9.81 |
| Accident with fatalities | 19 | 0.32 |

Table 2: Type of accidents

Obviously, with 41.61% parking accidents the property damage category is high with 64.3%. The more interesting fact is that in the last 10 years, only 19 fatalities have occurred. Looking more closely at the last 10 years a drop in accident rates is visible. From 2016, the peak with 693 accidents, to 2021, with 323 accidents, a drop of 53.39% was measured which is significant.

If we ignore the parking accidents to showcase possible speeding accidents then we receive a similar trend but with less variance and only a drop of 27.6%.


Fig. 5: Amount of accidents visualized in years (left with parking accidents, right without)

Analyzing this further, it is possible to see the categories which have varied the most. With no surprise, in combination with the last figure, property damage has fallen a lot. From our data, we can not state why this descent. The only significant gradient change is the light injuries which first have risen in 2018 and by now have dropped.


Fig. 6: Amount of accidents visualized in years (left with parking accidents, right without)

**Top 5 most accidents pedestrian** In this subsection, we deal with pedestrians who have had accidents. Here we do not distinguish whether an accident with a car, a motorcycle or a bicycle has taken place, but we focus only on the result that a person has been injured. The top street for accidents is Riehenstrasse 20. This is near the Wettsteinbrücke. Also, in this place we have looked at, we had seen that four participants meet each other:

(a) Tram

(b) bicycles

(c) cars

(d) pedestrians

In addition, there are no traffic lights at this intersection in this street. That these four participants meet each other in a blind intersection increases the chance of an accident.

| **Street name** | **House number** | **Pedestrian** |
|---|---|---|
| Riehenstrasse | 20 | 5.0 |
| Feldbergstrasse | 28 | 3.0 |
| Rauracherstrasse | 160 | 3.0 |
| Zürcherstrasse | 111 | 3.0 |
| Aeschenplatz | 6 | 3.0 |

Table 3: Top five of accidents for pedestrians

The frequency of pedestrian accidents in a specific location may depend on a variety of factors, such as the amount of pedestrian traffic, the presence of crosswalks and traffic signals, and the behavior of drivers. Without more information, it is not possible for us to accurately determine why there may be a higher number of pedestrian accidents in a particular location.

**Top 5 most accidents Bicycle**

| **Street name** | **House number** | **Bicycle** |
|---|---|---|
| Reinacherstrasse | 4 | 9.0 |
| Grenzacherweg | 266 | 7.0 |
| Steinentorberg | 12 | 6.0 |
| Mühlenberg | 18 | 6.0 |
| Klingelbergstrasse | 50 | 5.0 |

Table 4: Top 5 bicycle accidents

That the most dangerous street for cyclists is Reinacherstrasse four must be put into perspective. Because there is a big club scene at this place. Therefore it is enormously important for our analysis to know the time of the accident. Since we need to know whether accidents have taken place at night or not since it can be that these accidents have been caused by a Dorgen influence. Therefore, here is the exact time:

| **Date** | **Hour** | **Street** | **House number** |
|---|---|---|---|
| 2011 - 05 | 12 | Reinacherstrasse | 4 |
| 2013 - 03 | 7 | Reinacherstrasse | 4 |
| 2014 - 12 | 9 | Reinacherstrasse | 4 |
| 2015 - 09 | 7 | Reinacherstrasse | 4 |
| 2017 - 09 | 7 | Reinacherstrasse | 4 |
| 2018 - 04 | 18 | Reinacherstrasse | 4 |
| 2020 - 06 | 17 | Reinacherstrasse | 4 |
| 2021 - 03 | 17 | Reinacherstrasse | 4 |
| 2021 - 03 | 17 | Reinacherstrasse | 4 |

Table 5: Bicycle accidents in Reinacherstrasse 4 with times

We can see that the nine accidents happened at this time:

(a) 3 x 7 a.m.
(b) 1 x 9 a.m.
(c) 1 x 12 a.m.
(d) 3 x 5 p.m.
(e) 1 x 6 p.m.

If we analyze these times, then an accident caused by drugs is rather unlikely, but not excluded, because, until the morning hours, the clubs have open. But we can say that these took place at seven o'clock in the morning and nine o'clock, which could be an indicator that the morning stress could have caused the accidents, which is much more likely than riding a bicycle under drugs. Then four accidents in a period from 17:00- 18:59. Which were caused by the rush hour. Therefore, we are inclined to say that this stretch is more likely to be caused by morning stress and evening traffic. **Important:** The frequency of bicycle accidents in a specific location may depend on a variety of factors, such as the amount of bicycle traffic, the condition of the road, and the presence of other hazards. Without more information, it is not possible for us to accurately determine why there may be a higher number of bicycle accidents in a particular location.

### 7.2 Measurements

The measurement datasets consist of approximately 19′872′258 entries. Only 7.89% or 1′585′410 of the measurements are speeding cases. On average in both 50 and 30 zones, vehicles are speeding 10 km/h more than they should drive. To be precise in 30 zones 39.83 km/h and in 50 zones 59.47 km/h

As with the accidents, looking at the trend below some very high peaks are visible. Until the year 2021, the peaks mostly occur in the summer. After that, it is the inverse. It is out of the scope of this analysis to go further into detail but we assume that the Corona measures had something to do with it. We assume that during the lockdown people had more free time and fewer people were on the street which means that the conditions to speed are optimal.

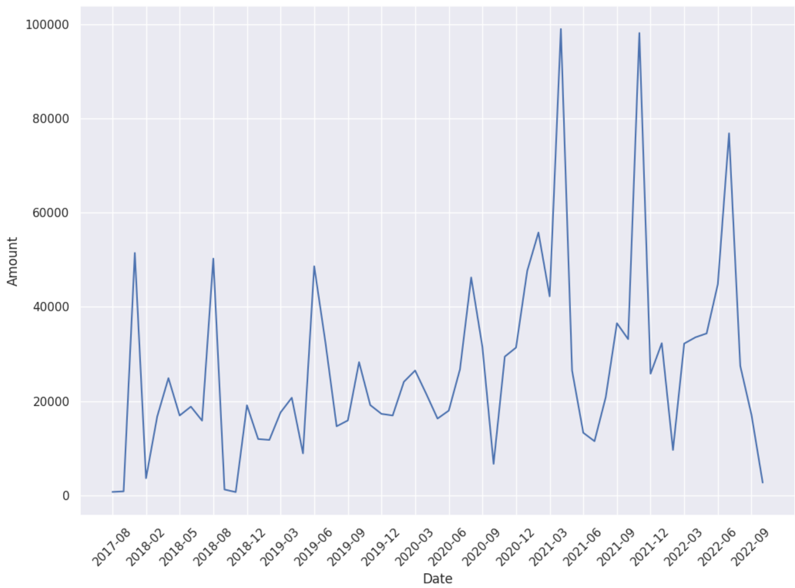


Fig. 7: Speeding in years

**Rasertatbestand Zone 30**[1] In this table, we have the top five speeding exceeding in zone 30. The "worst case" was 159 km/h. We checked the road in our leisure time to evaluate whether 159 km/h can be driven on this road. We estimate as authors that it should be possible. But to be clear, this is dangerous for

[1] no translation for the word

all road users in the street. In addition, all drivers conflicted with the speeding law.

| **Street name** | **House Number** | **Exceeding** | **Amount** | **Date** |
|---|---|---|---|---|
| Unterer Rheinweg | vis 96 | 159 | 1 | 2020 - 11 |
| Allmendstrasse | 124 | 131 | 1 | 2018 - 08 |
| Friedrich Miescher-Strasse | 30 | 124 | 1 | 2020 - 02 |
| Frankfurt-Strasse | Ecke Mailand-Strasse | 113 | 1 | 2021 - 03 |
| Entenweidstrasse | Gasstrasse | 112 | 1 | 2019 - 06 |

Table 6: Top five of most exceeding in zone 30

**Rasertatbestand Zone 50**[2] In this subsection, we look at the top five speeding violations in zone 50. These data are interesting. First, we look at the worst case with 187km/h. Again, we have explored the locality and discovered that there is a highway nearby. As authors, we are convinced that this will be possible to drive 187km/h. However, it should also be mentioned that this is life-threatening for all road users to drive at a speed of 187km/h in a 50 zone. The second interesting fact is that 3 out of five measurements were made in 2021 February. At first nothing unusual, but on closer analysis, this time is exactly the "restrictions time" due to the Corona measures in Switzerland.[2]

| **Street name** | **House Number** | **Exceeding** | **Amount** | **Date** |
|---|---|---|---|---|
| Voltastrasse | 29 | 187 | 1 | 2021 - 02 |
| Voltastrasse | 29 | 156 | 1 | 2021 - 02 |
| Holeestrasse | 119 | 140 | 1 | 2020 - 04 |
| Hiltalingerstrasse | Einfahrt Südquaistr. | 137 | 1 | 2021 - 02 |
| Zürcherstrasse | 180 | 129 | 1 | 2020 - 11 |

Table 7: Top five of most exceeding in zone 30

**Top 5 most exceeding Streets zone 30** If we look closely at where Holeestrasse is located, we see that it borders a main road where 50 can be driven. In addition, we have a document that since 2013 this road must be used as a 30 zone [1]. Which is an indicator that 50 km/h can be driven there. Also important is the fact that this road has a border with a neighboring canton, namely, Binningen, which tells us that this road is undoubtedly more used than others.

[2] no translation for the word

| **Street name** | **House number** | **Total exceeding** |
|---|---|---|
| Holeestrasse | vis 119 | 144017 |
| Steinentorstrasse | 18 | 78016 |
| Elisabethenstrasse | 15 | 66852 |
| Reiterstrasse | 33 | 51411 |
| Spitalstrasse | 12 | 51723 |

Table 8: Top five of 30 zone

**Top 5 most exceeding streets zone 50** For someone driving often in Basel, the Heuwaage-Viadukt is a road that is known. The limit for driving there was once 60 km/h, now only 50, which leads to high speeding, as people are used to a higher limit. This road also is long and straight which increases the likelihood to speed. Schwarzwaldbrücke and Schwarzwaldalle share the same characteristics.

| **Street name** | **House number** | **Total exceeding** |
|---|---|---|
| Heuwaage-Viadukt | Steinentorberg 12 | 114809 |
| Schwarzwaldbrücke | 1 | 92743 |
| Schwarzwaldallee | visa 161 | 59709 |
| Holeestrasse | vis 35 | 57940 |
| Kohlistieg | vis 39 | 47574 |

Table 9: Top five of 50 zone

As we can see, we grouped in figure 8 all the exceeding. We also recognize that Bettingen has the lease exceeding. One reason for this low violation number may be that there was not much measurement made at this location or the people in this place are obeying the speed rules. We also can see that in the suburbs the highest number of violations is in the 30 zone. We can see that the districts: Matthäus, Kleinhünigen, Wettstein, Iselin, Hirzbrunnen and St. Johann have the highest number of violations in the 50 zone. If we look at two neighborhoods, in particular, St. Johann, Iselin and Kleinhüngingen, we have to mention that these residential areas border other countries. Thus, the number of passing vehicles is also greater. Also, a fact and interesting is, that St. Johann and Iselin have two hotspots as streets in each district.

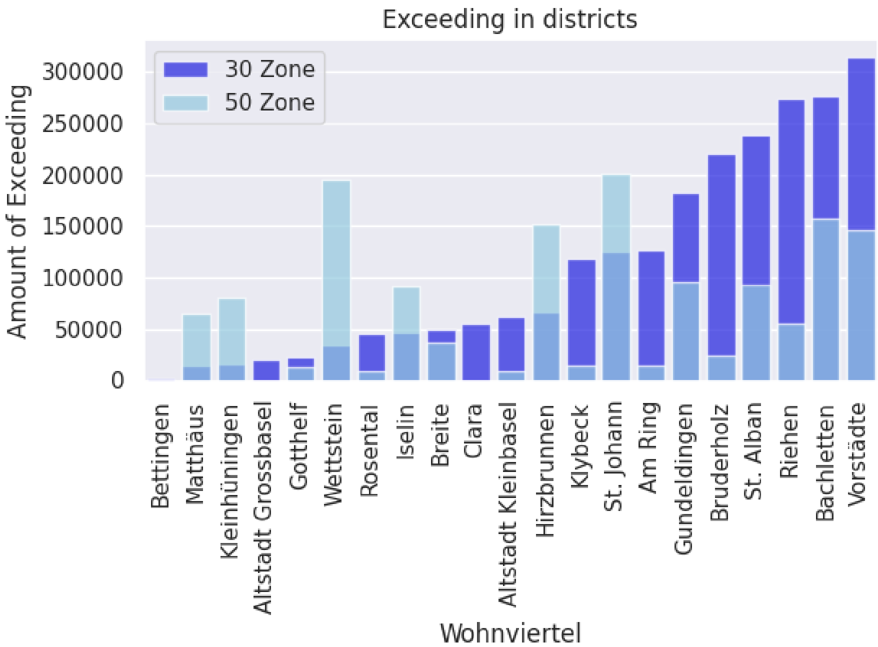


Fig. 8: Amount of exceeding in all districts, stacked

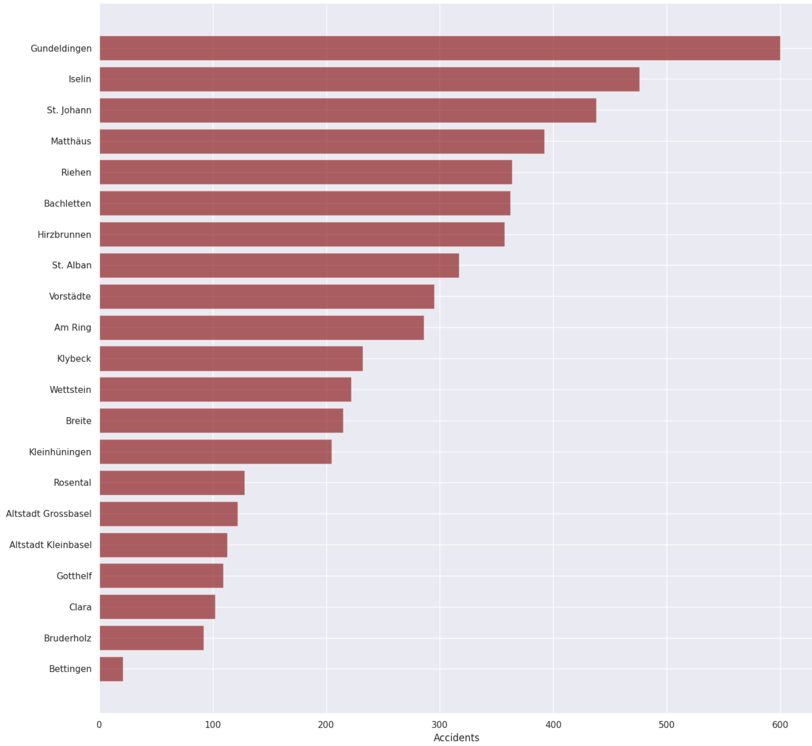


Fig. 9: Accidents per districts

On top of the leaderboard is Gundeldingen with 600 accidents. The district is compact with a lot of trams, many cars, pedestrians, cyclists, and motorcyclists. The Gundeldinger quarter is mainly served by streetcar lines 15, 16 and E11 as well as bus line 36. After several years of work on Güterstrasse (street in district), the completion of the reconstruction of the **main traffic axis** was celebrated at the end of August 2008. If we look at how many streets there are, in the Gundeldingen district, there are 69 streets. If we divide the number of streets by all accidents, then the ratio is: 8.69 accidents per street which is enormous. That this district is a hotspot in the section 7.3 is justified. **Note:** A lot of factors are important in the accidents. Here are a few of them:

1. How much traffic do we have?
2. How many pedestrians do we have?

3. Is the district well-frequented?
4. Is it a main traffic axis?

### 7.3 Hotspots

We define a hot spot in the following way:

- Let the violation $\text{threshold}_1$ be defined as $\mu$
- Let the accident $\text{threshold}_2$ be defined as $\phi$

*A street is a hotspot if the amount of exceeding is higher than $\mu$ and the amount of accident is higher than $\phi$*

| StreetName | Amount of violations in street | Accidents in total |
|---|---|---|
| Burgfelderstrasse | 11400.0 | 54.0 |
| Hegenheimerstrasse | 11241.0 | 70.0 |
| Dornacherstrasse | 22156.0 | 68.0 |
| Klybeckstrasse | 11091.0 | 55.0 |
| Claragraben | 13512.0 | 56.0 |
| Güterstrasse | 14079.0 | 75.0 |
| Grenzacherstrasse | 17951.0 | 60.0 |

Table 10: Hot spot of Basel

As we can see in Table 10 in our project, we have seven streets with $\mu \in \{> 11000\}$ and $\phi \in \{> 50\}$.
In other words our definition is:

**Definition 1.1: Hot spot**

A street is a hot spot if the amount of exceeding is strictly higher than 11'000 and the amount of accidents is strictly higher than 50.

**First approach:**

We also want to know if there is a correlation between "Accidents in total" and "Amount of violations in street", we made a scatter plot, see Fig. 11.
From this plot, we see that the poly-fitted line is linear. We also can see, that we have two out-liners, 70 and 75 accidents. One Accident is very close to the line,

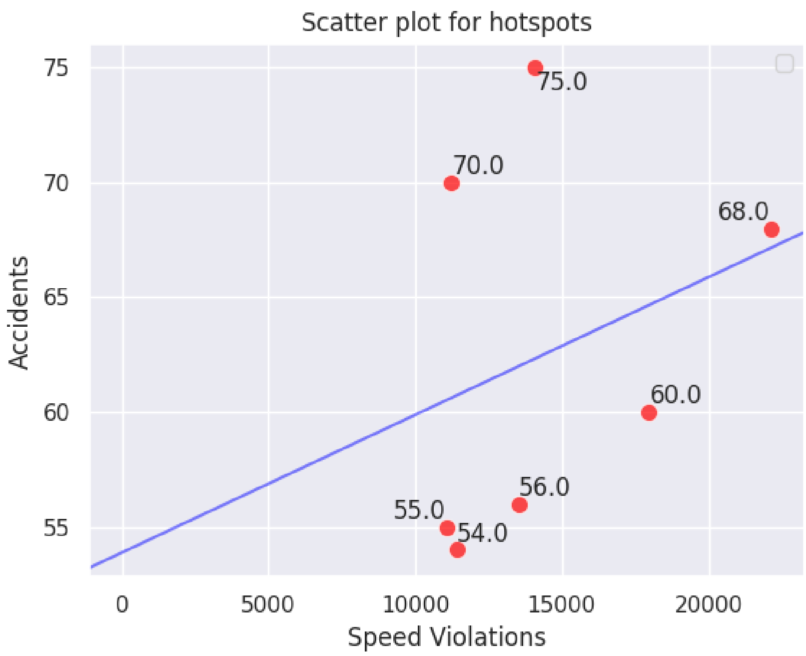


Fig. 10: Violations and Accidents in correlations

but this is in our opinion owed to chance. The reason why we think that must these must be by chance is that we do **not** have any information about the fact, that the accident was caused by speeding.

Since that is the situation, we do not want to make a statement if the accidents were caused by speeding or not. It is reasonable to assume that a correlation is plausible between speeding and accidents intuitively.

The decision to take $\phi > 50$ & $\mu > 11000$ was optimal because if we had a lower threshold, for both, too many streets would be a hot spot and the meaning of a hotspot would lose its identity.

**Second approach:**

To get a better grasp of the correlation scatter plots help. The data on the x-axis is the sum of speeding cases on each street and on the y-axis is the sum of accidents per street. The red line is the least squares polynomial fit.

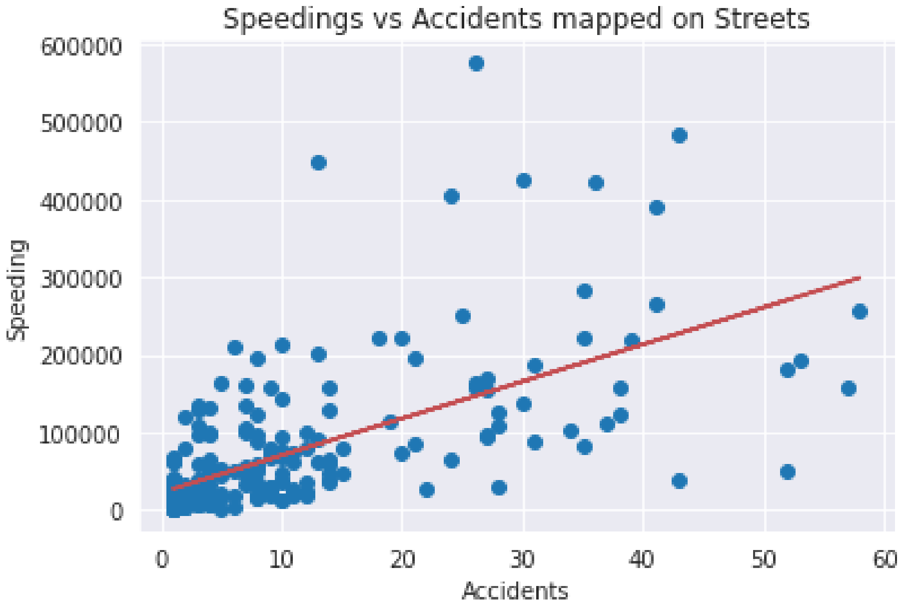


Fig. 11: Correlation of speedings and Accidents

We have calculated the Pearson correlation number and The correlation number is: **0.6318**
The p-value is: **2.27e-26**

# 8 Discussion:

## 8.1 Correlation

The value of 0.63 for the Pearson correlation coefficient and a small p-value of 2.27e-26 indicate that there is a strong positive correlation between the two variables. A correlation coefficient of 0.63 means that as the value of one variable increases, the value of the other variable also tends to increase, but not necessarily at a constant rate. The small p-value of 2.27e-26 indicates that the correlation is statistically significant and is unlikely to be due to chance.

In general, a correlation coefficient of 0.5 or higher is considered to be a strong correlation. A correlation coefficient of 0.63 is considered to be a strong positive correlation.

## 8.2 Analysis

**8.2.1 Our data has its flaws** The first problem which we have not noticed at the beginning was how the data sets documented the time of the events. The accident data set did not have the exact day in date format but only the week code and the weekday like "Monday". House numbers are also not a friend of the accident data set, as we had to reverse engineer the house number through the geocode and a free API. A healthy amount of data cleaning was necessary after the API calls.

**8.2.2 Size matters** Comparing datasets that have a peak of 693 accidents and 510341 speedings in a year is not optimal. Straight conclusions are not simple as we lack data points. Our interpretation could be simply wrong because of the coincidence of the accidents. We hope the accidents dataset gets smaller every year and the trend still goes down, but we did have the feeling that data was missing.

**8.2.3 Politics** Also interesting is that the Green Party in the city of Basel is proposing that speed limits be reduced to 30 km/h in all residential areas of the city. The party claims that such a move would reduce noise and air pollution, and increase traffic safety. The TCS beider Basel, a local transportation organization, does not support the proposal, arguing that traffic flows more smoothly at 50 km/h and that such a measure is unnecessary on cantonal roads.[3]

## 8.3 GeoJson

We used two GeoJson Files, one is to hover over the street names (100189) the other one is to get the shapes from the districts (100042). The two files can be downloaded from here:

1. `https://data.bs.ch/explore/dataset/100042/table/?sort=wov_id` (District)
2. `https://data.bs.ch/explore/dataset/100189/table/` (Street names)

1. 100042.geojson: To use a geoJson file with the exact shape allowed us in this project, to create multiple different choropleth maps in one map. This lets us evaluate e.g. in which districts are the most accidents and mark them with different colors. It allowed us also to extend the GeoJson file with our data.
2. 100189.geojson: To use a geoJson file with the exact street names, allowed us to hover the street names on the map. Which is not necessary, but nice to have. The essential goal of this file was, to add extra data in the geojson file, with the hotspot data and mark them with the additional information of the hotspot.

## 8.4 Map

**8.4.1 Architecture** To understand how the map is created. Let us take a look at how the architecture is. See Figure 12.

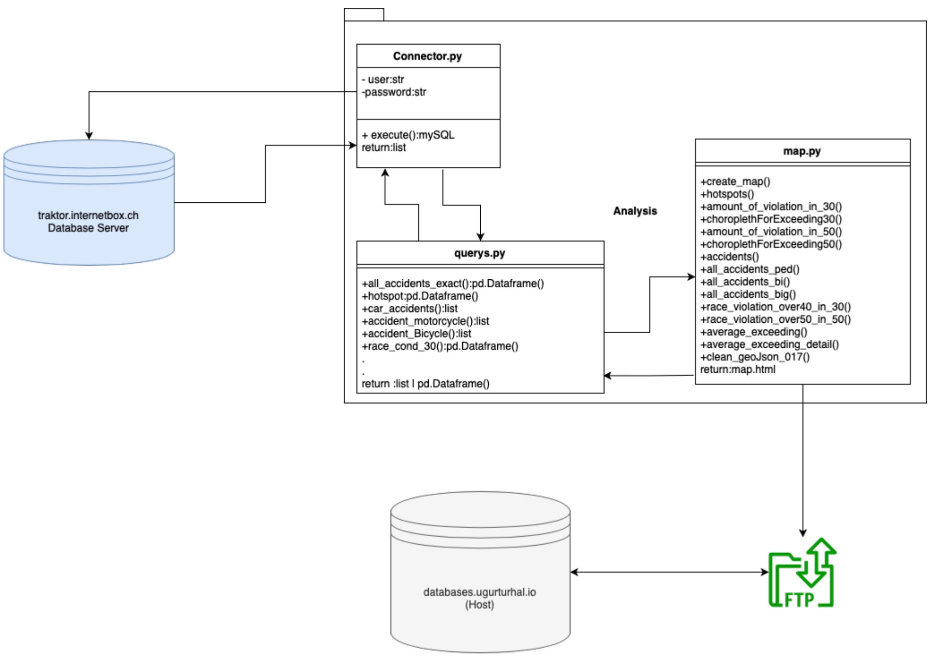


Fig. 12: Architecture of our system

1. `Connector.py:` Our class to connect to the database. Takes as argument only queries and parameters.

2. `querys.py:` Is a helper function with queries for the map.py file.

3. `map.py:` Creates the map. This file contains only functions. The essential module is folium. E.g. if we want all accidents with the exact Coordinates the workflow is like this:
   - map.py contains a function `allaccident()` calls the query which is in querys.py
   - querys.py has an object of the Connector.py, the object takes the queries and if given also parameters. The connection is made.
   - The Database Server responds to the connector. Connection is closed.
   - querys.py gets the response and puts this response in a data frame or list.
   - map.py gets a data frame or a list and adds the coordinates to the map.

**8.4.2 Layers** Since our data set contains about 20 million data and many measurements are made at one point, we were forced to find another way to display the measurements. If we hadn't changed our mind, our map would look like Figure 13 below.

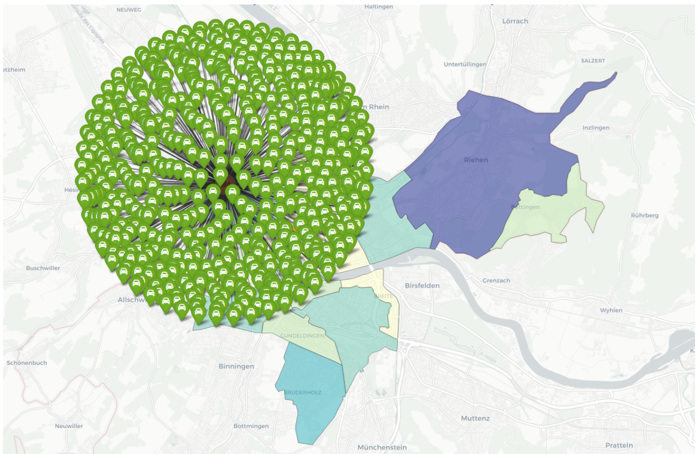


Fig. 13: 10'000 Coordinates at one measurement

We decided to not take the exact points but to count, how many people have driven too fast on that specific street. See Figure 14 for output in the map.

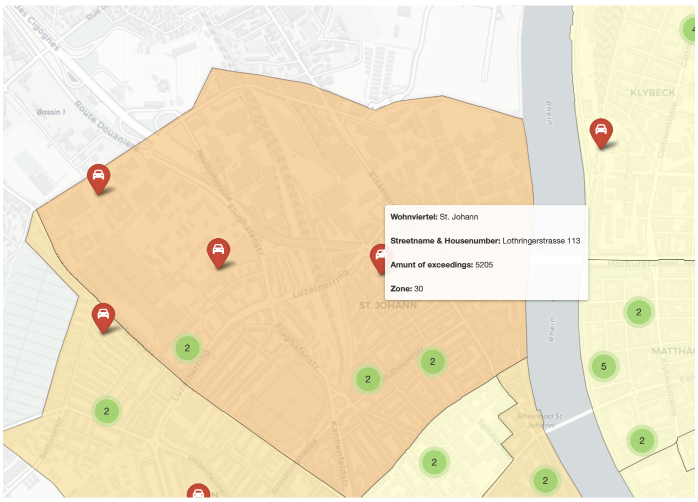


Fig. 14: Better data representation

As we have seen, the map has layers. To be exact, we have 15 Layers. Which can be combined together. Figure 14 has two Layers combined active is:

1. Violation of 30 km/h & amount
2. Amount of Streets in districts (Choropleth).

#### 8.4.3 Layers and Meaning

1. Choose between light and dark mode.
2. Hot-spot
   Shows the streets which have the most accidents with pedestrians, motorcycles, bicycles, cars, and the total violation number in the street.
3. Violation of 30 km/h & amount
   Shows all violations in Zone 30, the amount, and the exact coordinate.
4. Violations in zone 30 (Choropleth) Marks the district with different colors, after the amount in zone 30, and shows the percentage.
5. Violation of 50 km/h & amount
   Shows all violations in Zone 50, the amount, and the exact coordinate.
6. Violations in zone 50 (Choropleth) Marks the district with different colors, after the amount in zone 50, shows the percentage.
7. All Accidents (Choropleth)
   Marks the district with different colors, color is based on the number of total accidents in the district.
8. Accidents involving Pedestrians
   Shows the exact point of an accident with a pedestrian (bicycle or motorcycle).
9. Accidents involving Bicycle
   Shows the exact point where an accident with a bicycle was involved (motorcycle).
10. Accidents involving all participants
    Shows nothing since this combination was not present.
11. Speed violation over 40km/h in Zone 30
    Exact location and amount of exceeding in 30 zone at least over 40km/h
12. Speed violation over 50km/h in Zone 50
    Exact location and amount of exceeding in 50 zone at least over 50km/h
13. Average mean in zone 30 (Choropleth)
    Marks the district, the color is based on the average mean of in zone 30.
14. All exceeding mean
    Shows exact point and the average exceeding speed.
15. Streetnames
    Shows the street name by hovering.

16. Amount of Streets in the district (Choropleth)
    Marks the district, with color, shows the number of streets in the district, color is based on the number.

### 8.5 Searching & Time slider Choropleth

**Searching:** As you can see we can search exactly a street in the speeding zones. If you type a letter it recommends you streets weather in the 30 Zone in the first search tool. And in the second street which were exceeded in the 50 Zone. **Note:** You can only search after one street and not multiple streets, this is default by the `folium module`.

**Time slider Choropleth:** What we miss in our map is: A time slider choropleth which should mark the districts in which the exceedings were in the specific month. We did not achieve this, to combine the map with 3 features:

1. layer for timeslider choropleth
2. layer for choropleth
3. layer for marker

We assume that we can only work with folium choropleth and marker but not with a time slider, in the same map.

### 8.6 Multiprocessing

This section is in the sense of the project, maybe out of topic. But it is necessary to mention multiprocessing, especially in our case. If we closely look, what multiprocessing does, is: Use multiple cores from the CPU to run processes simultaneously (in our case, get the data from the Database Server).

We have 15 layers in the map which are created with the result of the queries from our database. We have an average 10 min waiting time for some queries. Since the CPU has multiple cores, we could force this with a module to use all the cores. This would be possible with the python module: `multiprocessing`. We tried to start multiple processes with a waiting time between the processes with `time.sleep(1)` to generate the Map faster. The problem was: it was not "process safe." That means, while other processes were still executing, the map generation was finished, which led to the map taking the data of the other still running processes.

Our map gets generated by the following functions in this order:

```
if __name__ == '__main__':
    # Example of how to use multiprocessing
    # p0 = multiprocessing.Process(target=create_map)
    create_map()
    hotspots()
    amount_of_violation_in_30()
    choroplethForExceeding30()
    amount_of_violation_in_50()
    choroplethForExceeding50()
    accidents()
    all_accidents_ped()
    all_accidents_bi()
    all_accidents_big()
    race_violation_over40_in_30()
    race_violation_over50_in_50()
    average_exceeding()
    average_exceeding_detail()
    clean_geoJson_017()
```

A better approach would be to start a child process for each query and letting the map creation function wait until every child process is finished. Only then should the map compile with the results of the child processes and their queries in the form of dataframes.

### 8.7 Docker

Since the database was hosted inside a docker container, development had its ups and downs. Deployment of the database was a breeze, as is for many docker applications. Not only for the amount and quality of guides that can be found online but also because of the simplicity of docker.

Some problems arose when restarting the container because of automatic local port selection and port forwarding setup. Docker (Portainer to specific) increments the port on which the container was running. Thus forwarding the port to the router would fail because the port had changed. We had to restart the container twice until we learned how to see the active process list in the database. Metadata locks were the reason for the restarts as we did not understand why some queries ran indefinitely.

The database is reachable from outside the home network. Therefore we could simultaneously work with the database and share our results. This led to more resources on our local computer which were used to create the map.

## 9 Conclusion:

Based on the results of our analysis, we have found a strong positive correlation between the two variables we examined. The Pearson correlation coefficient of 0.59 indicates that as the value of one variable increases, the value of the other variable also tends to increase. The small p-value of 2.49e-25 indicates that this correlation is statistically significant and is unlikely to be due to chance.

These findings suggest that there is a meaningful relationship between the two variables and that changes in one variable are associated with corresponding changes in the other variable. However, it is important to note that this correlation does not necessarily imply causation. Further research is needed to understand the exact nature of the relationship between these variables and to determine the underlying mechanisms at play.

Our results provide valuable insight into the relationship between speeding and accidents on public roads. We found that there were more speeding cases and accidents on certain roads and during certain times of day, but we did not have sufficient data to identify clear causative factors or to make definitive conclusions about the relationship between speeding and accidents.

Overall, our findings suggest that further research and analysis are needed to fully understand the prevalence and impact of speeding on public road safety. While our data analysis provides some insight into the patterns and trends in speeding and accident rates on public roads, it is not sufficient to support a comprehensive understanding of the issue.

To more fully understand the role of speeding in accidents on public roads, it would be necessary to collect and analyze additional data on factors such as the speed at which vehicles were traveling at the time of the accident, road conditions, vehicle types, and driver behavior. By gathering and analyzing this data, we can develop more targeted and effective strategies to reduce the risks and improve the safety of all road users.

## References


[1] *Flyer for the resident, howpublished =* https://www.stsbw.ch/wp-content/uploads/2020/10/Infoflyer_T30_Holeestrasse.pdf*, note = Accessed: 2023-09-01.*

[2] *Offical Website governement of switzerland, howpublished =* https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-82462.html*, note = Accessed: 2023-02-01.*

[3] *Regional TV broadcaster, howpublished =* https://telebasel.ch/2022/02/15/flaechendeckendes-tempo-30-in-basel/?channel=105100*, note = Accessed: 2023-02-01.*