Friedrich-Alexander-Universität
Technische Fakultät

Final Presentation:

Correlation between precipitation and speeding based on data from Cologne, Germany from 2017 to 2022.

Introduction (1)

„The vast majority of most weather-related crashes happen on wet pavement and during rainfall: 70% on wet pavement and 46% during rainfall”

-National Highway Traffic Safety Administration (NHTSA)

Introduction (2)

Therefore:

In order to better understand the driving behavior in wet conditions, the following question was asked in the context of this project:

Is there a significant correlation between precipitation and speeding?

Agenda

01 Data Source Management

02 Data Manipulation

03 Individual Data Review

04 Correlation Review

05 Findings & Summary

Data Source Management

Datasource 1: Geschwindigkeitsueberwachung Koeln (Speed Monitoring Cologne)

- Metadata URL: https://mobilithek.info/offers/8862870771136450928
- Data URL: https://offenedatenkoeln.de/sites/default/files/Geschwindigkeit%C3%BCberwachung_Koeln_Gesamt_2017-2022_0.csv
- Data Type: CSV
- Includes: All speed monitoring data from Cologne between 2017 and 2022
- Imported via: pandas.read_csv

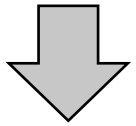
Datasource 2: Wetterrückblick und Klimadaten Koeln (Climate Data Cologne)

- Metadata URL: https://meteostat.net/de/place/de/koln
- Data URL: https://meteostat.net/de/place/de/koln?s=D2968&t=2017-01-01/2022-12-31
- Data Type: CSV
- Includes: All weather and precipitation data from Cologne betweenn 2017 and 2022
- Imported via: meteostat Python module

Data Manipulation

Datasource 1: Speed Monitoring Cologne

- Drop all but the date and the excessive_speed columns
- Format and rename date and excessive_speed
- Export to SQL



- 2 data columns (3550692 entries):
  - **date:** datetime64[ns]
  - **excessive_speed:** float64

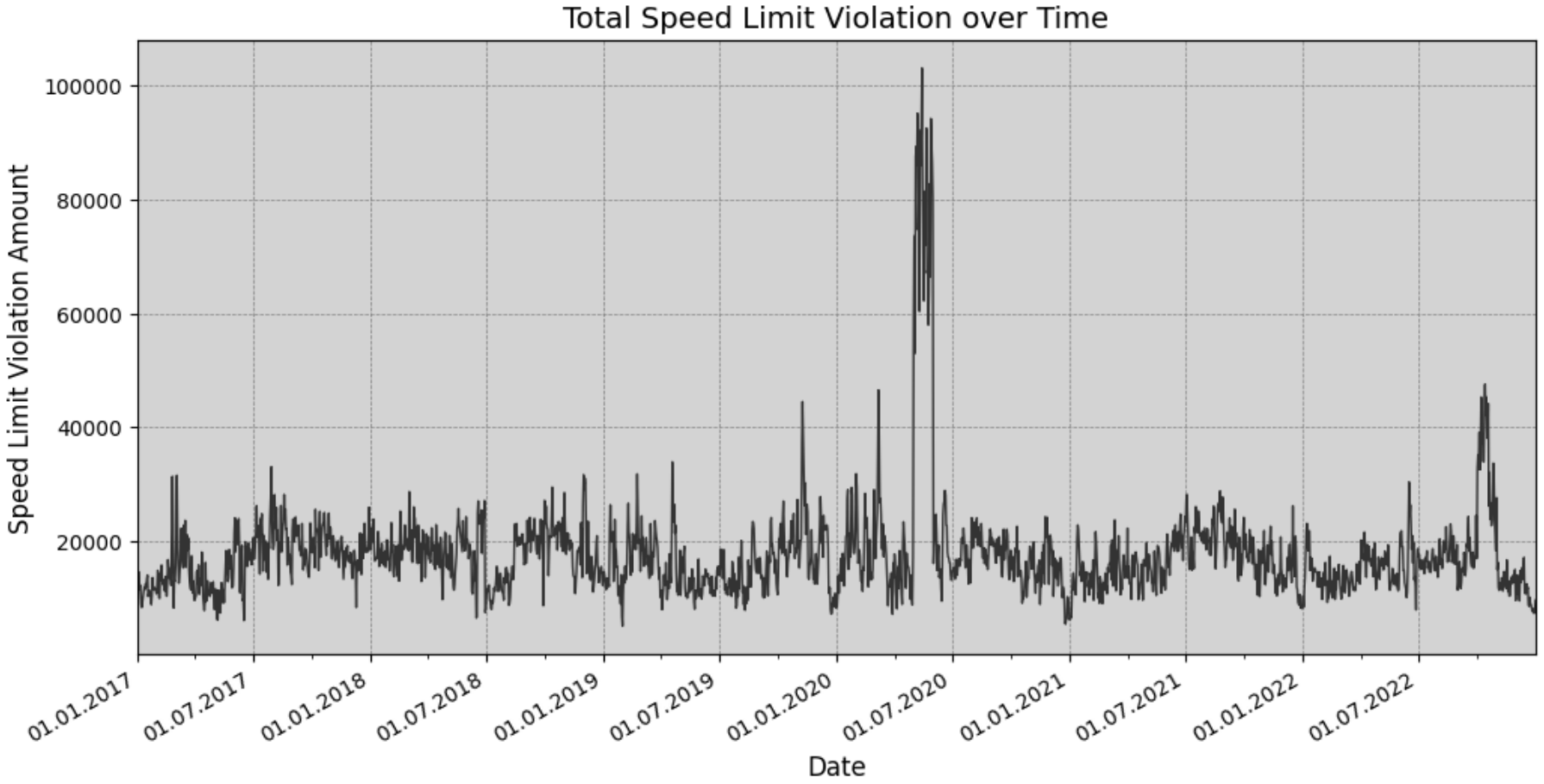
Datasource 2: Climate Data Cologne

- Drop all but the precipitation column
- Add a date column, then format date and precipitation
- Export to SQL

- 2 data columns (2191 entries):
  - **date:** datetime64[ns]
  - **prcp:** float64

Individual Data Review

Speed Monitoring Cologne (1)



In order to visualize the data, the amount of all speeding violations per day is calculated and then illustrated as a line graph.

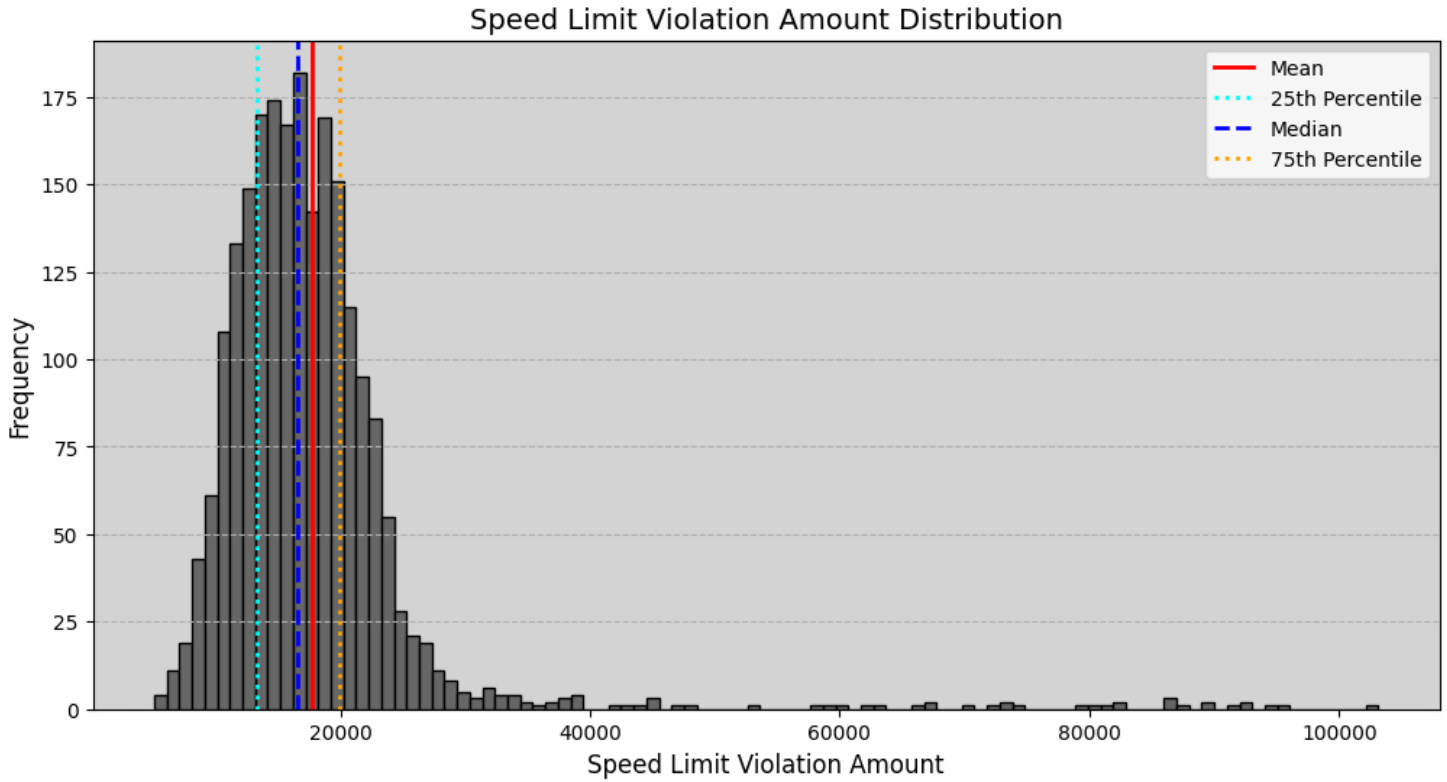
Individual Data Review

Speed Monitoring Cologne (2)

Statistical characteristics:

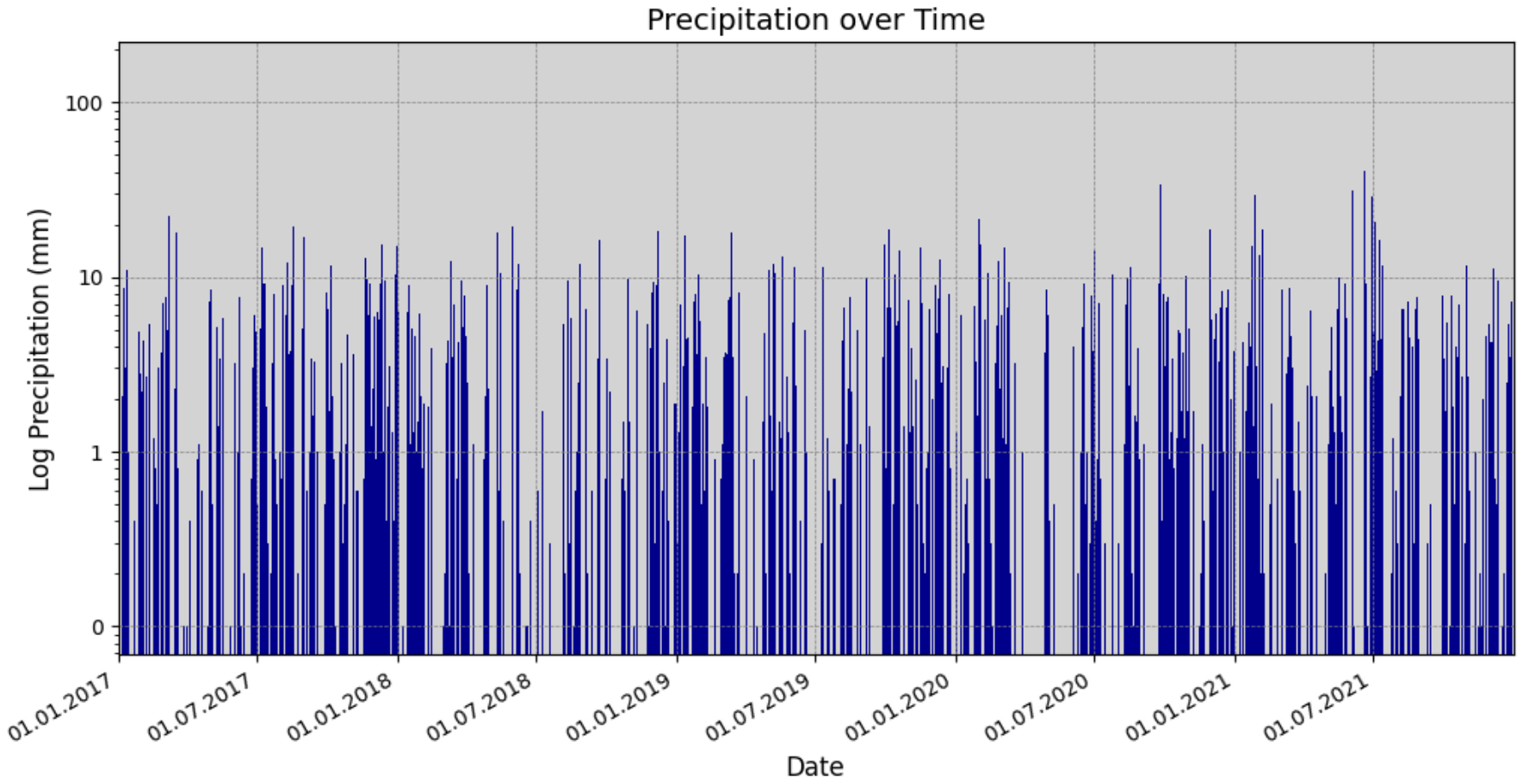
- **Average Speed Limit Violation Amount per Day**: 17731.3
- Range of Speed Limit Violation Amount:
  - Minimum: 5080.0
  - Maximum: 103140.0
- Percentiles:
  - 25th Percentile: 13305.0
  - 50th Percentile (Median): 16532.0
  - 75th Percentile: 19895.5

Histogram:



Individual Data Review

Climate Data Cologne (1)



In order to visualize the data, the daily amount of precipitation is illustrated as a logarithmic bar chart.

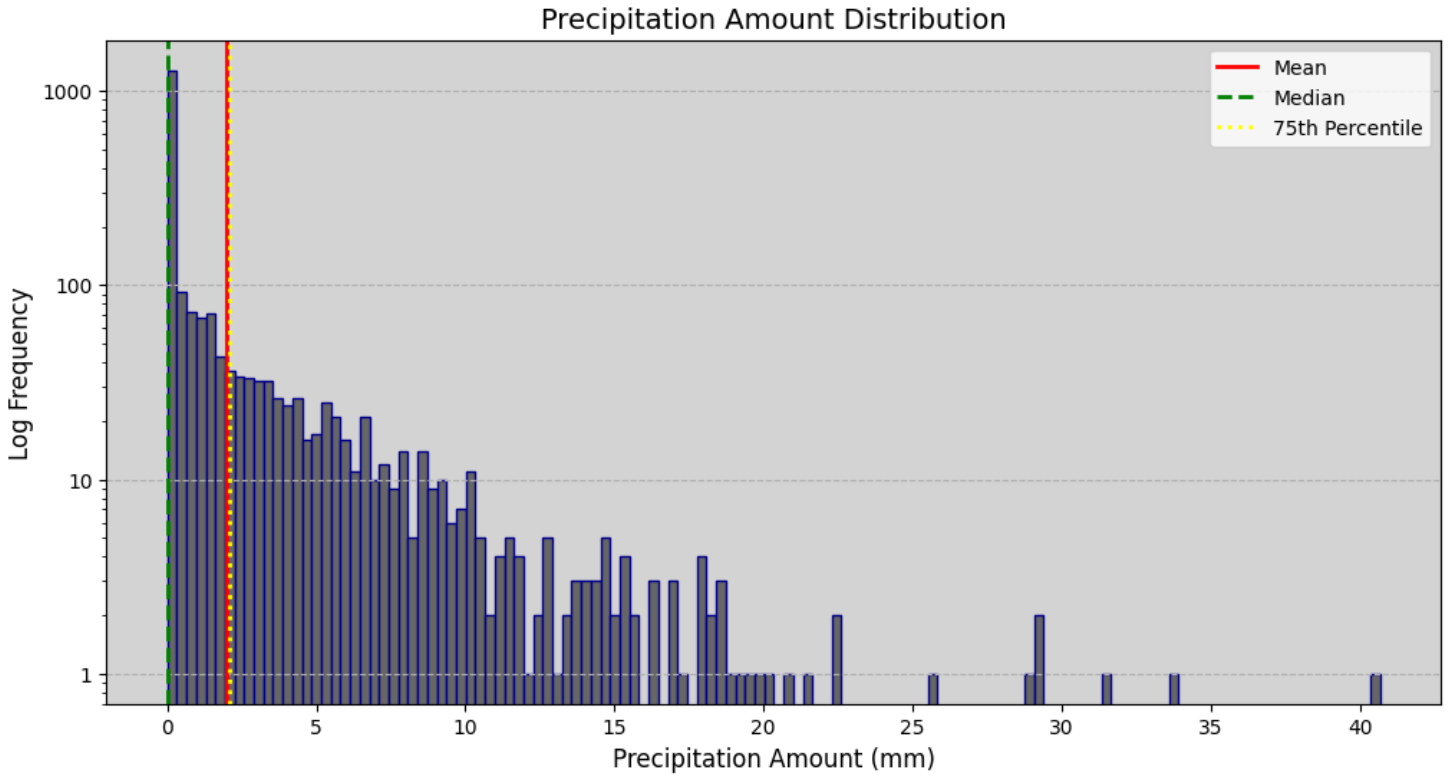
Individual Data Review

Climate Data Cologne (2)

Statistical characteristics:

- **Average Precipitation Amount per Day**: 2.0
- Range of Precipitation Amount:
  - Minimum: 0.0
  - Maximum: 153.5
- Percentiles:
  - 25th Percentile: 0.0
  - 50th Percentile (Median): 0.0
  - 75th Percentile: 2.1

Histogram (log):



Correlation Review

Correlation Coefficient

The Correlation Coefficient is:

```
-0.033438928022421446
```

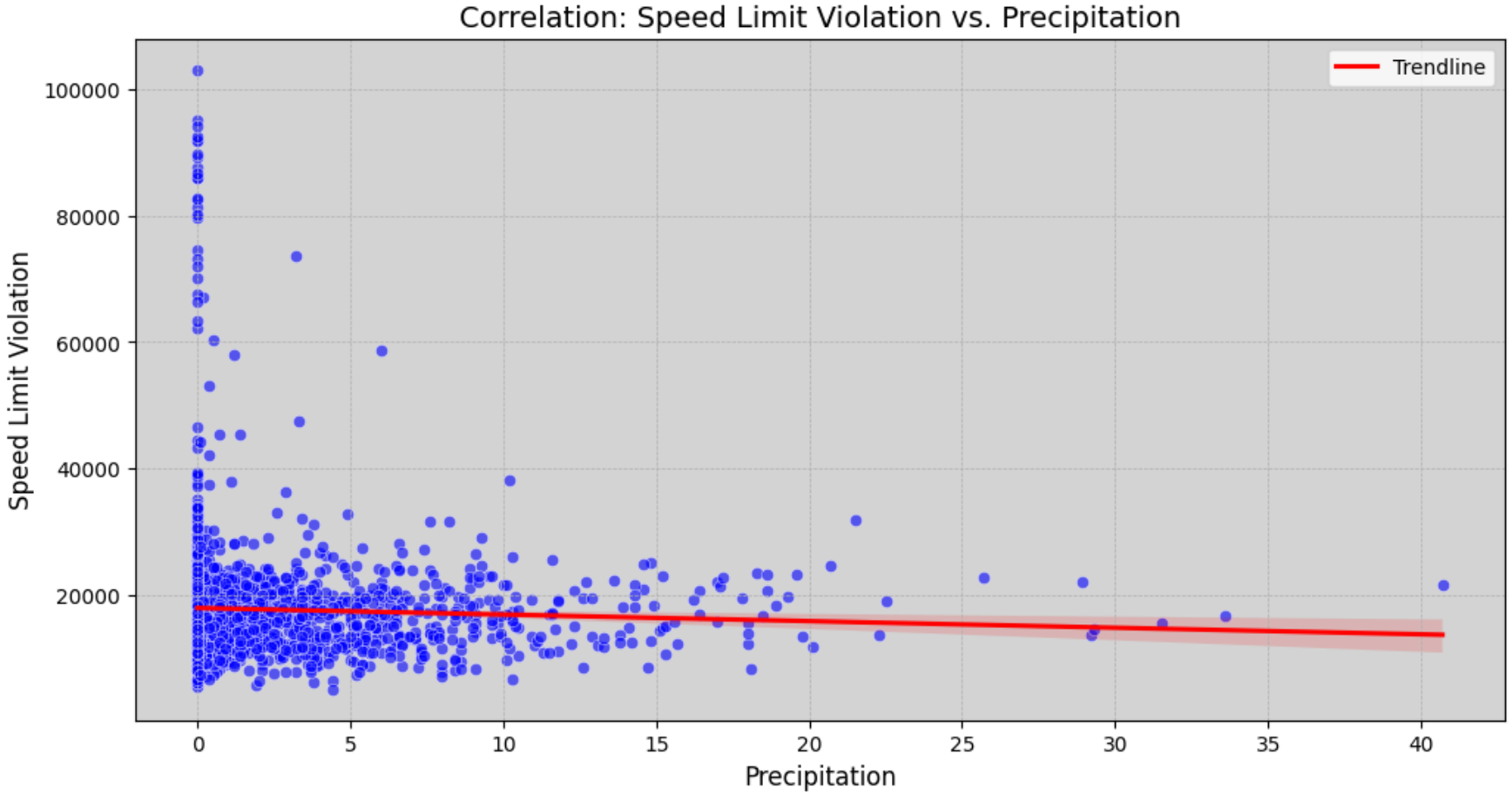
→ Very weak negative correlation between precipitation and the total daily amount of speeding violations

→ Slight tendency for the amount of excessive speed violations to decrease slightly when there is higher precipitation

→ But: **Correlation does not imply causation**

Correlation Review

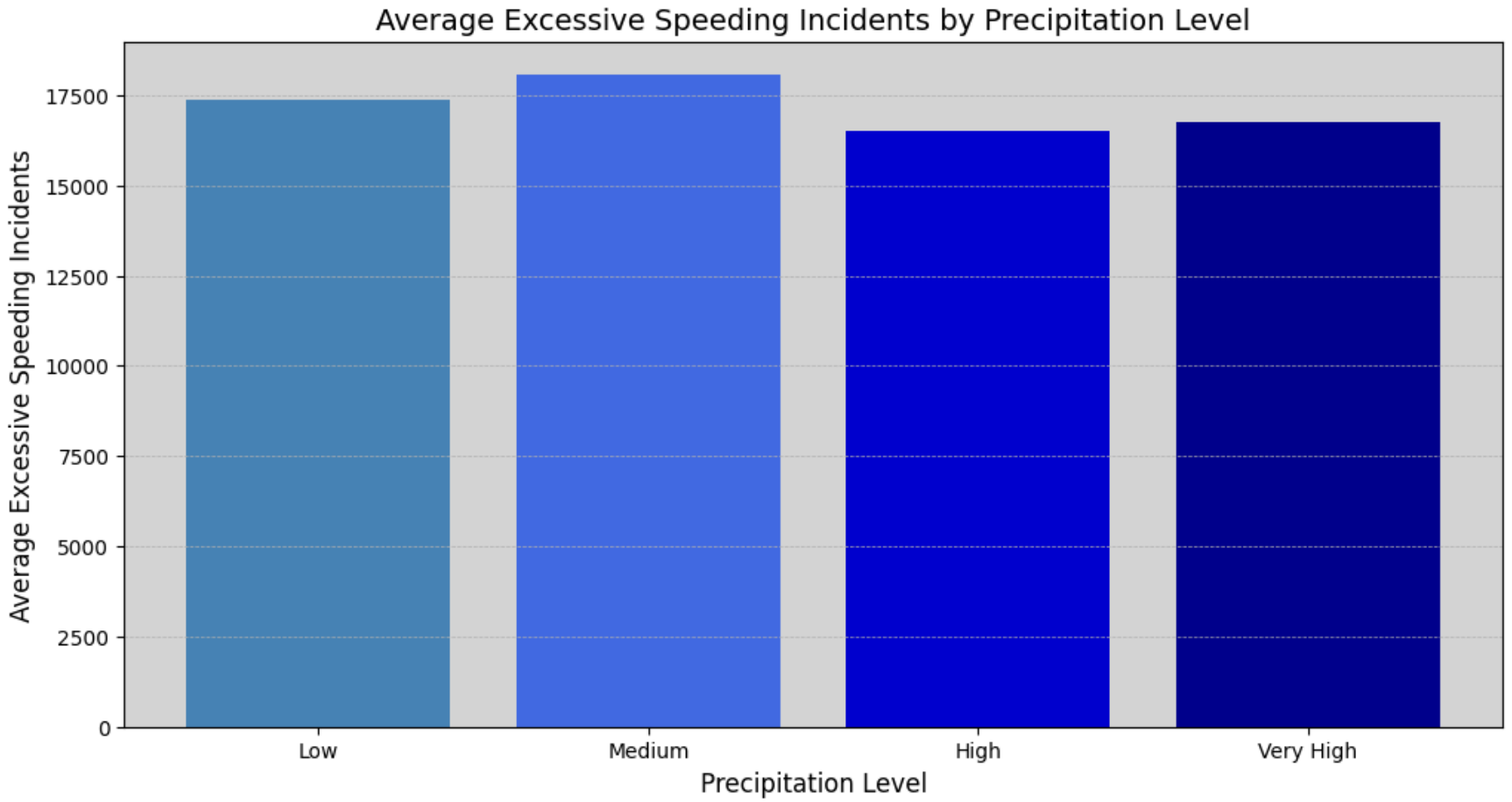
Scatter Plot



A scatter plot using purified data visualizes the slight correlation.

Correlation Review

Bar Chart



In this bar chart, the observation is illustrated in a simpler fashion. A slightly lower speeding amount can actually be observed for high and very high precipitation in comparison to low and medium rainfall.

Findings & Summary

Findings:

- A **correlation coefficient** of approximately **-0.0334** was found between precipitation and speed limit violations

  → This indicates a **very weak negative correlation** between the two variables

- The precipitation levels were **further categorized** into low, medium, high, and very high

  → The number of **speeding incidents was slightly lower for high and very high precipitation** compared to low and medium rainfall

Getting back to the original question:

Is there a significant correlation between precipitation and speeding?

- The correlation coefficient is very close to zero and other factors beyond precipitation could influence driving behavior and speed limit violations
- Without further measures (as in e.g., a regression analysis), the answer is **no**

Thank you
For your attention!