Analysis Report

Mathieu Stenzel

July 3, 2024

Introduction

Do temperature (positively) and rainfall (negatively) correlate with the number of people walking through the pedestrian zone in Erlangen?

Used Data

The data used for the analysis is the output of the data pipeline. When the data is up to date (can be achieved by running `pipeline.sh`) it represents the number of pedestrians together with the average temperature and rainfall per day starting from yesterday to 550 before that day. For that data from the company "HyStreet" and the "Deutscher Wetterdienst" have been combined.

At this point I would like to express my gratitude to the company HyStreet for providing me with historical data of the pedestrian zone in Erlangen.

Analysis

The following steps have been implemented in the file `analysis.py`:

1. Remove the days that are within an event that takes place in Erlangen (e.g. Bergkirchweih, Weihnachtsmarkt, ...)
2. Calculate the average number of pedestrians per weekday.
3. Select the days with more than 10% difference to the median of that weekday.
4. For the remaining days, calculate the correlation between pedestrian counts and temperature and rainfall, separating the data with pedestrian counts above and

below the median.

The following code block contains the output of the Python script executed on `2024-07-01`. The results are based on the data between the dates `2022-12-29` and `2024-06-30`:

```
Statistics for the days where the number of pedestrians has a difference
of more than 10 % to the respective median of that weekday.

Days with more than 10 % above the median:
Average temperature: 15.05 °C
Average rainfall:    0.54 mm

Days with more than 10 % below the median:
Average temperature: 6.97 °C
Average rainfall:    4.13 mm
```

Conclusions

In order to be able to answer the original question, the data was cleaned by filtering out days with events in Erlangen. For the remaining days the median for each weekday was calculated and days with a difference of more than 10 % to the median were selected. The average temperature and rainfall for the days with a pedestrian count above and below the weekday median were calculated.

With these adjustments, the question from the introduction can be answered:
> **Yes, temperature is positively and rainfall negatively correlated with the number of pedestrians in Erlangen.**

While analyzing the data another conclusion was drawn:
> **Social events and Sundays have more influence on the number of pedestrians in Erlangen than rain or temperature**.