Project Proposal
Zürich Air Quality Analysis

Group: Alisha Yasmin Aydin, Luca Franceschetti, Wonkyo Kim, Claude Widmer

General topic and motivation

In this project we investigate the possible correlation between variables describing traffic, weather, and air quality in Zurich. There are multiple variables for each topic. We want to find out how they are correlated in the group itself and how they potentially influence each other.
As one of Switzerland's' biggest cities, Zurich has undergone changes in traffic usage in the last decades. In the same time molecules that influence air quality like Ozone and Carbon monoxide affect the life quality and health in the city. We want to see whether molecules influencing air quality are de- or increasing and their connection to weather and traffic.

Data and source

We are working with data that is freely accessible on the website stadt-zuerich.ch. The data for the traffic, weather and air quality is measured hourly. All of them are measured at the following location in Zurich: Stampfenbachstrasse. The starting year of the measurements varies strongly. The weather data is being collected since 1992, the one surrounding air pollution even since 1983, while the vehicle types are only described since 2007.

- Variables surrounding weather (since 1992), hourly:
https://data.stadt-zuerich.ch/dataset/ugz_meteodaten_stundenmittelwerte
Pressure of air, rain duration, global radiation, temperature, relative humidity, wind direction, vector and scalar velocity
Locations in Zurich: Stampfenbachstrasse, Schimmelstrasse and Rosengartenstrasse

- Variables surrounding air quality (since 1983), hourly:
https://data.stadt-zuerich.ch/dataset/ugz_luftschadstoffmessung_stundenwerte
Ozone (O3), nitrogen oxides (NOx), nitric oxide (NO), nitrogen dioxide (NO2), particulate matter (PM 10 and PM 2.5), carbon monoxide (CO), and sulfur dioxide (SO2)
Locations in Zurich: Stampfenbachstrasse, Schimmelstrasse and Rosengartenstrasse, Heubeeribüel

- Variables surrounding vehicle types (since 2007):
https://data.stadt-zuerich.ch/dataset/ugz_verkehrsdaten_stundenwerte_stampfenbachstrasse
There are three classes that categorize vehicles through length:
0m - 2.7m = 2R (Motorcycles, Bicycles),
2.7m - 7m = PW,
7m - 25m = LKW

Data processing

First, we will tidy the data by removing the missing values, filtering for Stampfenbachstrasse and analyzing if there are outliers and other anomalies in our data. Since our dataset is very large, we will take the data from 2007 to 2021 and reduce/compress our data by adjusting the time scale in order to do an informative EDA. This could be done by transforming the hourly air quality values into the Air Quality Index (AQI) which is on a daily basis. Our data will then be split into different training and validation sets (cross-validation) in order to analyze our models.

Research questions

- How did the variables linked to air quality (ozone, nitrogen oxides etc.) change from 2007 to 2021?
- Which air quality variables are linked with which weather phenomenon?
- How do the different vehicle categories affect the air quality differently?
- Can we predict the variables linked to air quality by regarding the other ones?
- Would reducing or limiting traffic have an effect on the air quality in Zurich?
- Are the measurement locations meaningful and realistic for the rest of Zurich and other cities?
- Can we predict according to weather data how much traffic there is?

Techniques and Algorithms

We try different regression methods and compare them. With cross validation we try to find out which model best describes our data. Because we have so many predicting parameters, interesting methods to consider would be Ridge or Lasso regression to reduce the amounts of variables used.