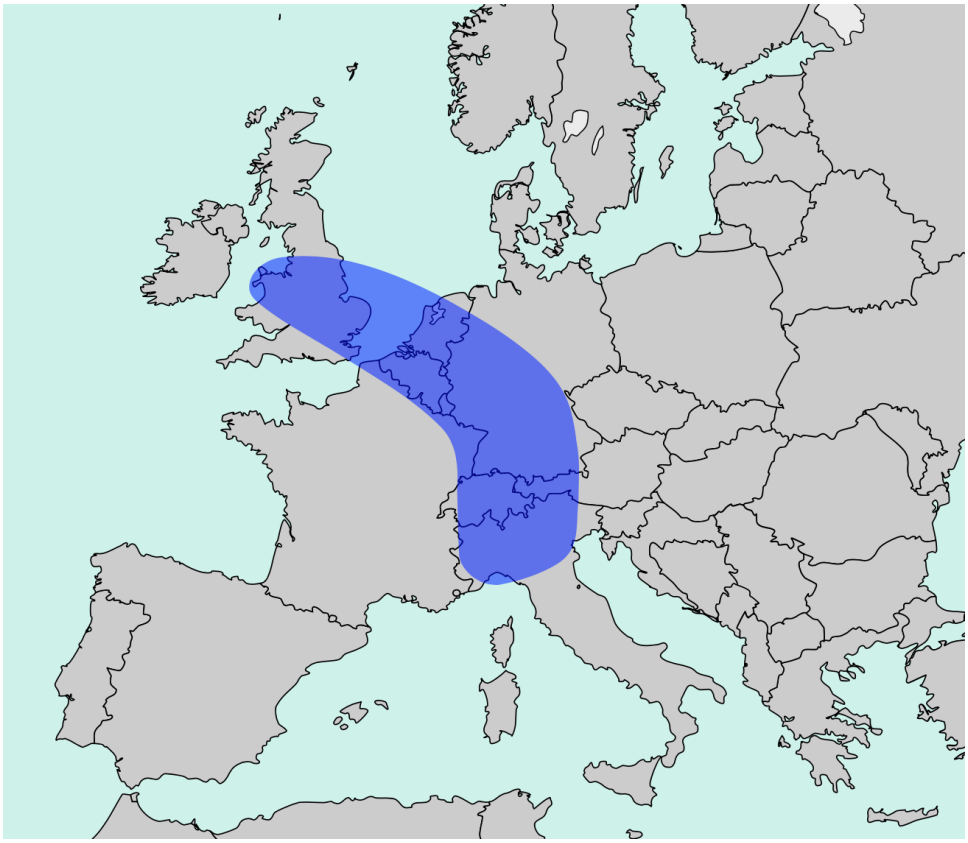
Coursera Project Report – Education Infrastructure in Zurich

1. Introduction

Switzerland is a small country in the heart of Europe, with a population of roughly 8.5 million. However, it belongs to the most densely populated areas of Europe, being part of the



so called "blue banana":

Accordingly, the population growth rate is very high in urban centers of Switzerland as well, most prominently so in Zurich, one of the financial capitals of Europe. With this, there is a higher demand for new social infrastructure, such as hospitals, public transport stations, and schools.

In this project, we want to assist the government in deciding which locations might be best to build new elementary schools (ES). With pupils from ES, it is important to guarantee for them a safe environement in school as well as on their way to school. By far the most

dangerous component of that is traffic, sadly, traffic accidents endanger the lives of many students every year.
Therefore, it is vital to take traffic patterns as well as other geographical data of Zurich city into account when analysing potential locations for education infrastructure. This is a perfect way to apply the data science skills I acquired during the last few weeks to a real-world problem.

2. Data

The Data we use is provided under https://data.stadt-zuerich.ch/dataset/sid_dav_verkehrszaehlung_miv_od2031/resource/fa64fa70-6328-4d47-bcf0-1eff694d7c22 . However, although more recent data is available, I will be making use of the data from Zurich from the year 2019 - pre-Corona, as the Corona pandemic has biased this situation and I would like our analysis to be more applicable to normal, everyday life as we used to know it. This seems to be the most adequate thing to do.
The data we use is very comprehensive: It is documenting the number of cars and motorized vehicles at given counter locations in Zurich – with measurements of hourly traffic over the period of a year, as well as details about the traffic such as direction, et cetera. We want to use the techniques and skills we learned in the IBM course to make statements about the magnitude of traffic in different regions of Zurich, and thus be able to judge which regions might be unfit for a new school to be built.

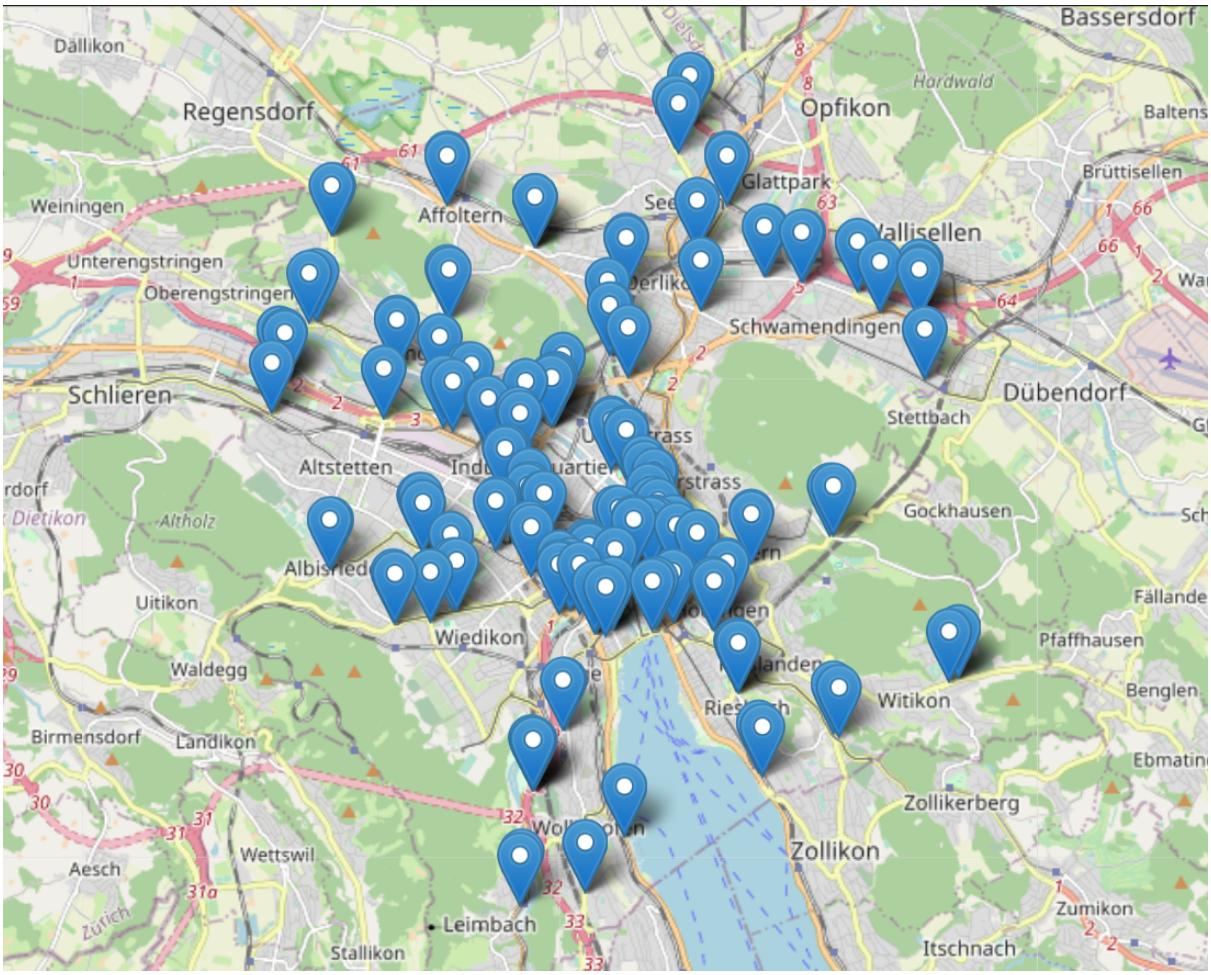
3. Methodology

Essentially, our method is divided in two steps: We first clean the data and translate the information to English, and then try to gain initial insight into the dataset – details about this can be found in the notebook file. We will find a few difficulties: The coordinates are written in a fairly ancient coordinate-type used in Alpine exploration: The “Schweizer Landeskoordinaten”. Translating from this to our normal set of longitudinal and latitudinal

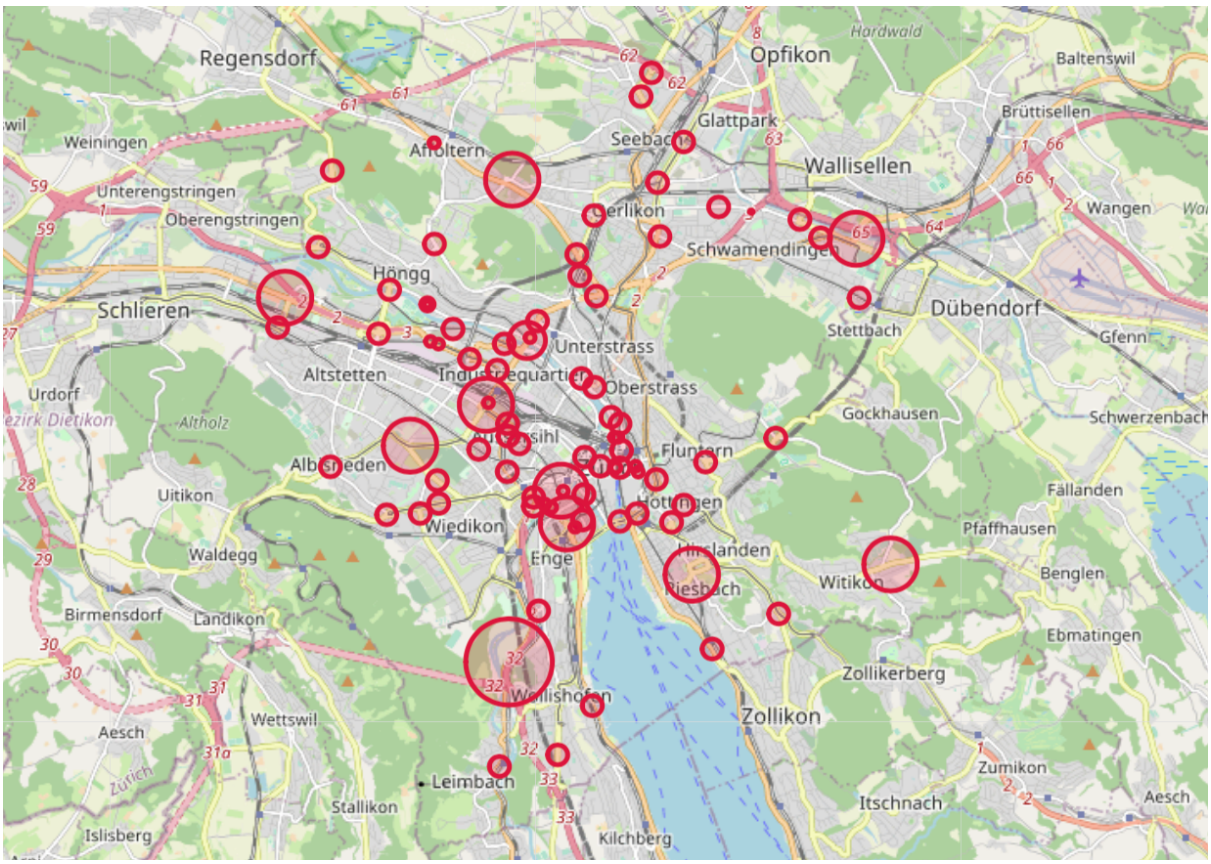
coordinates is tricky indeed. We needed to install additional libraries and packages, but thanks to the skills learnt in the course, we managed to handle the data set accordingly. We then used statistical inference as well as folium plotting to gain insight on the location of the measurements and their numbers.
Here, apart from visualization techniques and geographical imaging with folium, we made use of data grouping, dataframe splitting, offline cluster analysis and bubble diagrams.

4. Results

Using the methods eluded to above, we were able to map the physical locations of the measurement stations correctly:



And then, using what we explained above, we also found the magnitude of the traffic in the different neighborhoods of Zurich:



5. Conclusion

Judging from the data we obtained and the analyses we performed, we were able to determine Oerlikon to be the most suitable region of Zurich for a new school and generally for educational infrastructure for younger pupils. Both the traffic intensity being low, but it still being fairly reachable from all parts of Zurich led us to this conclusion. A close second was the region Altstetten, with similar characteristics.

6. Outlook

As a next step, we could look at the data of different years as well as make seasonal distinctions in the dataset to maybe plan school terms according to the time ranges with the least traffic.