

|  |
| --- |
| **IBM Certification** |
| Finding an apartment to live in Wuppertal, Germany |
| Annis HAJJI |

|  |
| --- |
| April 10, 2019 |

Reference

[I. Introduction 2](#_Toc6486907)

[1. Problem to solve 2](#_Toc6486908)

[2. Interest 2](#_Toc6486909)

[II. Data Acquisition 3](#_Toc6486910)

[2.1. Sources 3](#_Toc6486911)

[2.2. Features Engineering 3](#_Toc6486912)

[2.2.1. Neighborhoods 3](#_Toc6486913)

[2.2.2. Apartments from “ImmobilienScout” 5](#_Toc6486914)

[2.2.3. Station 6](#_Toc6486915)

[III. Mapping the data 8](#_Toc6486916)

[3.1. Venues and Clustering Neighborhood 8](#_Toc6486917)

[3.2. Apartments 10](#_Toc6486918)

[3.3. Final Map 11](#_Toc6486919)

[IV. Results 12](#_Toc6486920)

[4.1. Pre-selection 12](#_Toc6486921)

[4.2. Final selection 12](#_Toc6486922)

[V. Discussions 14](#_Toc6486923)

[VI. Conclusion 15](#_Toc6486924)

# Introduction

## Problem to solve

The challenge to resolve is being able to find a rental apartment unit in Wuppertal (Germany). Therefore, the rental unit is subject to the following conditions:

- Apartment with min 1 bedroom

- Kitchen included

- Cold rent not to exceed 500€/month

## Interest

This is a relevant project for anyone who is moving to a major city due to his job or anything else. The use of FourSquare data and mapping techniques combined with data analysis will help resolve the problem. Lastly, this project is a good practical case toward the development of Data Science skills.

# Data Acquisition

## Sources

In order to make a good choice of a similar apartment in Wuppertal (Germany), the following data is required:

* Apartments for rent in Wuppertal area with descriptions (beds, price, location, address). For this research, We used the website: [http://www.immobilienscout24.de/Suche/S-T/P-%s/%s-%s/%s/%s?pagerReporting=true](http://www.immobilienscout24.de/Suche/S-T/P-%25s/%25s-%25s/%25s/%25s?pagerReporting=true)
* Information on neighborhoods in Wuppertal with their Geodata (latitude and longitude). The data has been found in this url: <https://offenedaten-wuppertal.de/sites/default/files/Stadtbezirke_EPSG4326_JSON.json>
* Venues and amenities in Wuppertal neighborhoods due to FourSquare
* Subway station of Wuppertal which you will find in this url: <https://offenedaten-wuppertal.de/sites/default/files/Schwebebahnhoefe_EPSG4326_JSON.json>

I used the same technique as in the previous class to scrape data from web:

- the apartment from immobilienscout

- the neighborhoods from the website of the city

- the station from the website of the city.

After that, I'm going to do some cleaning and preprocessing in order to response to my problem.

## Features Engineering

### Neighborhoods

The same technique was used to scrape the website of Wuppertal, in order to find the geodata of each neighborhood.



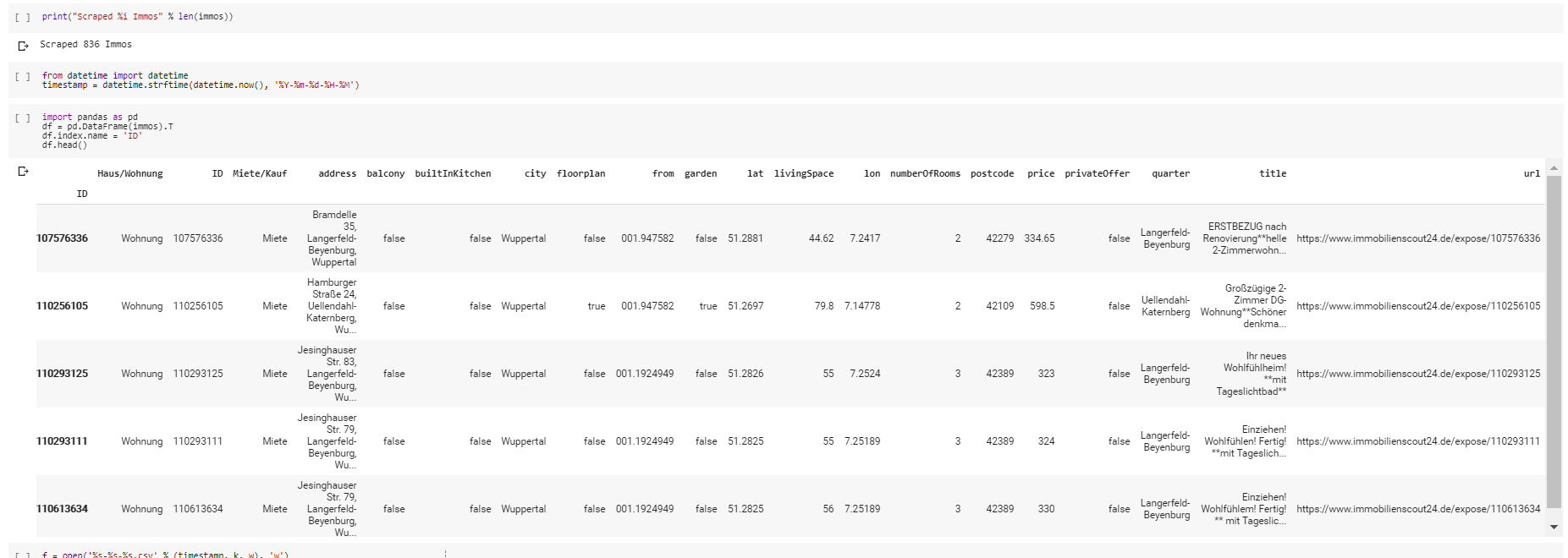
As you can see, we have many coordinates for each Neighborhood. In order to have one coordinate per Neighborhood, we have to clean the data. After that, we have a dataframe as shown below.



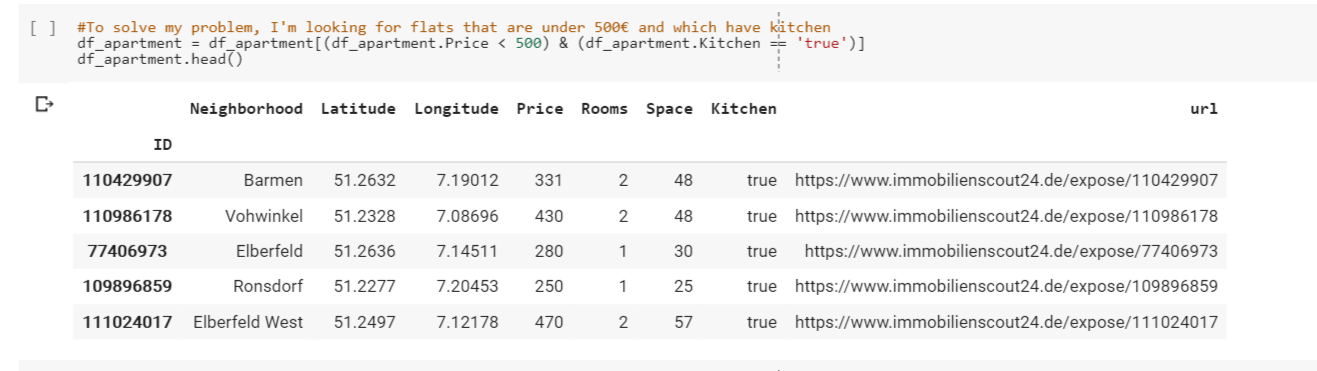
This final dataframe was used for the analysis that we will see in the next section.

### Apartments from “ImmobilienScout”

We scraped the data from immobilienscout, it is a website such as “Craigslist” where you can find a flat to buy or to rend in Germany. The result is the dataframe below.

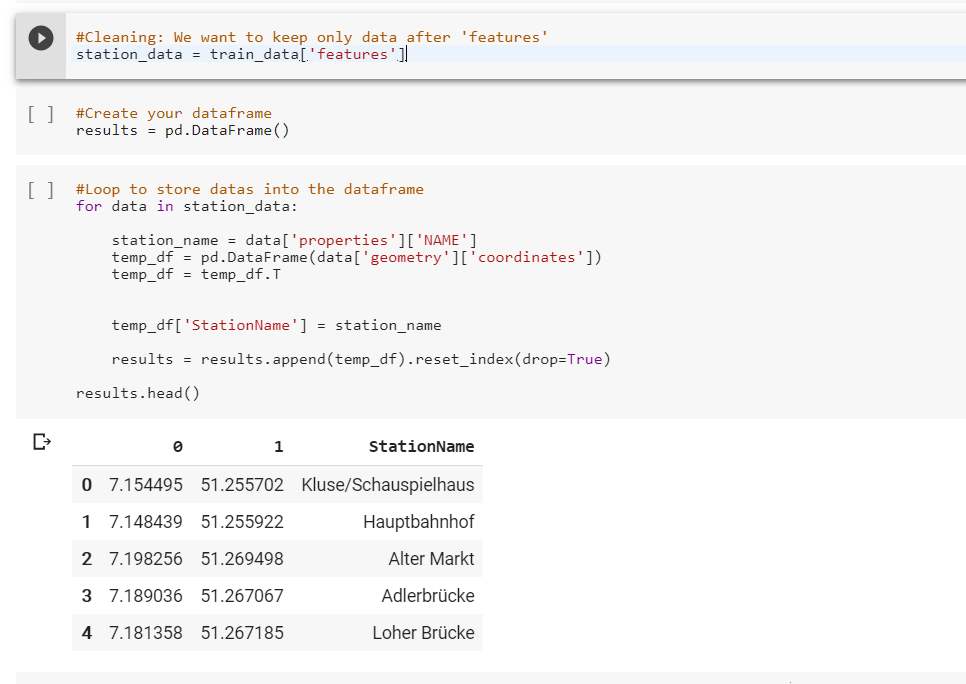


In order to answer to our problem, we had to drop some features and make some filters as you can see in the dataframe below.

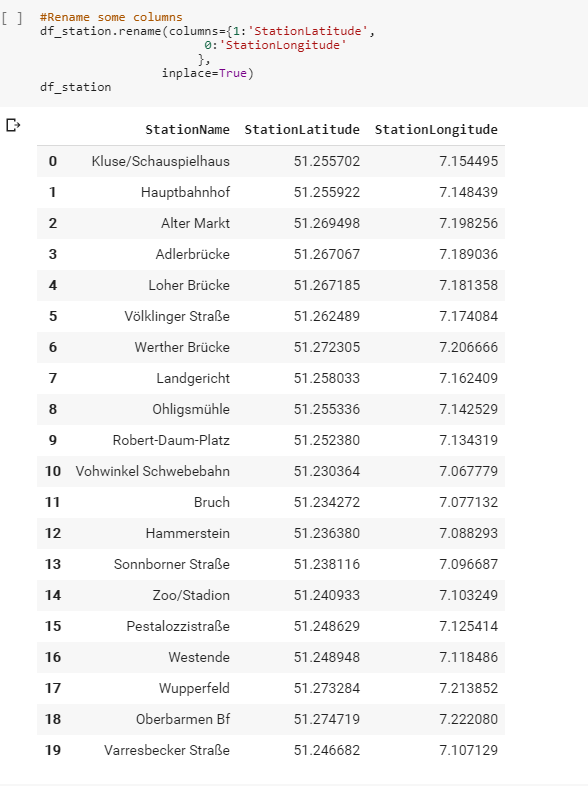


### Station

We scraped the station in order to build our analysis in the following section. In fact, we are looking for a flat that respect some conditions.



After some cleaning, we have the following dataframe.



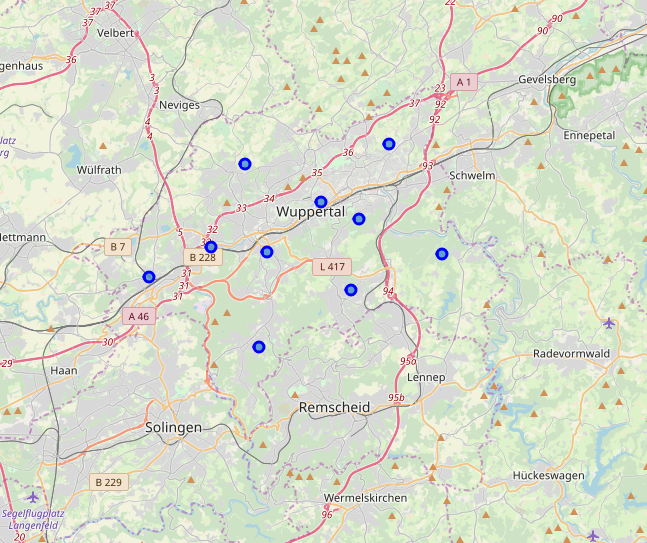
# Mapping the data

In this section, we are going to use the folium to map our dataframe. This will help us to make our decision.

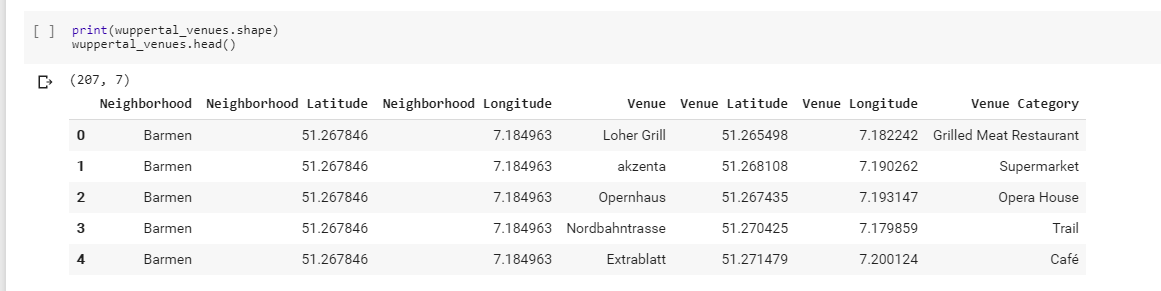
## Venues and Clustering Neighborhood

The first step is to cluster our neighborhoods in order to make a choice of Neighborhood to live.

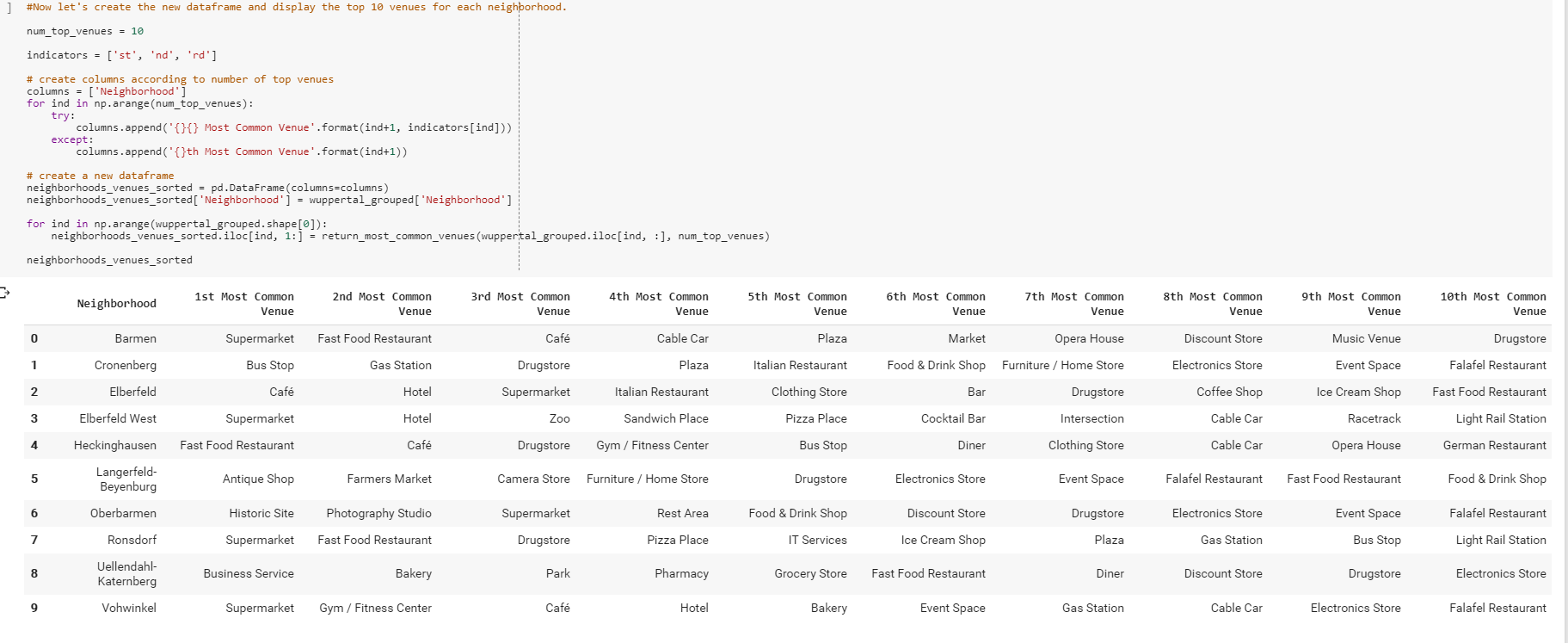
The following map contains the geodata of each neighborhood.



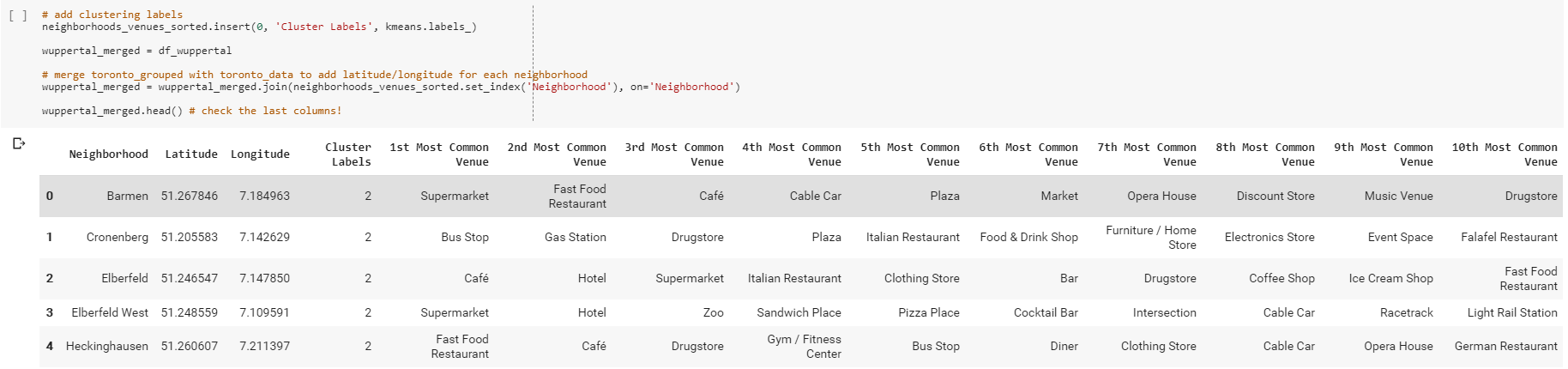
After defining our credential from FourSquare, we get the venues of each neighborhood in the dataframe below.



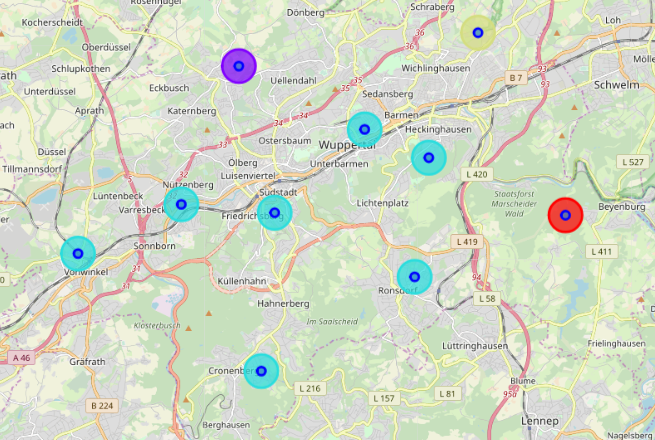
We decided to get the top 10 venues per Neighborhood. This will help us to find a nice neighborhood to live in.



The next step was to attribute a cluster for each Neighborhood.

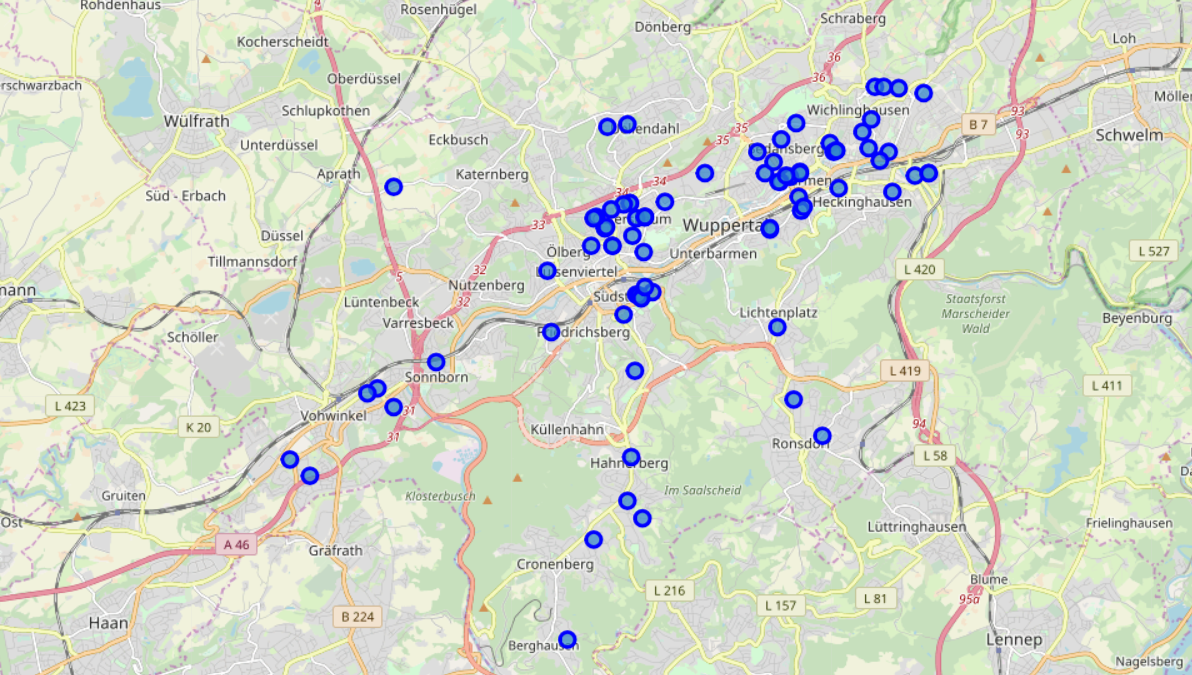


Now we can have the final map with top 10 venues per Neighborhood



## Apartments

In order to visualize the apartment from our dataframe, we used folium to create a map and see each apartment where is located.

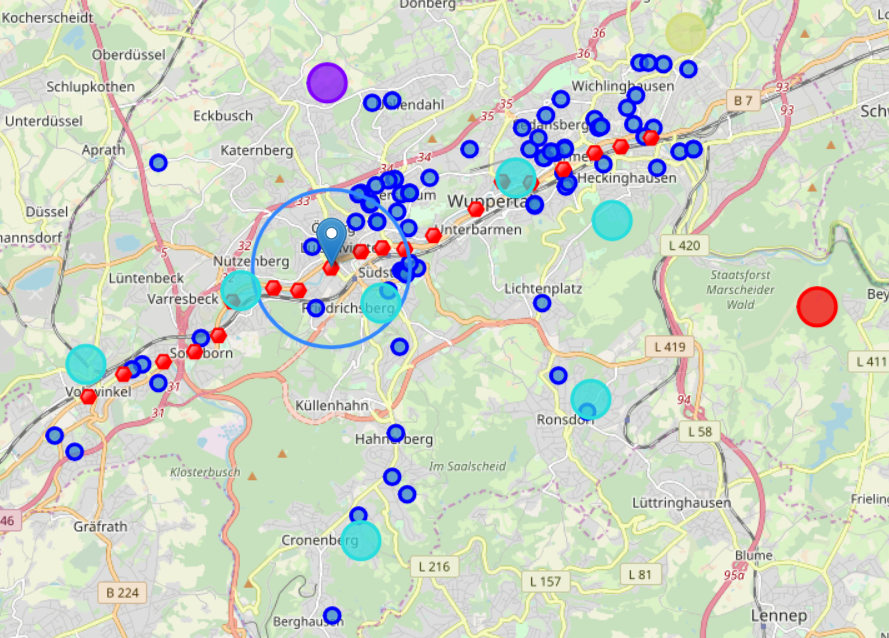


## Final Map

The final map contains in:

* Blue, the apartments
* Red, the stations
* Rainbow colors, the clusters and the venues

However, you can see a marker in one of the stations and a circle of 1500 meters around it. The subway station is my future work place. In fact, I wanted to visualize the apartments close my work place.



# Results

## Pre-selection

After examining the map and the apartments, I decided to focus on these 3 apartments.



## Final selection

In order to make the final decision, I decided to analyze the 3 following apartments.



Based on data analysis, I feel that Cluster 2 type of venues is great place to live. However, the apartment at 420€/month seems to be a better choice since the extra monthly rent is worth the conveniences. In fact, it's close to my work. Furthermore, the apartment is bigger than the others which can be a good solution. In fact, I can rent the 2nd room which can help me to reduce the spends.

# Discussions

This capstone project is a great opportunity to practice and apply all the data science methodology to a real case. In fact, it allows me to find a nice apartment.

However, this project can be presented to my future employers in order to show them my potential as a data scientist.

Furthermore, this project is only the first step to my career and my enthusiasm to work in the data science area. When, you can see the potential of this discipline, you can be only exited.

# Conclusion

This course is a complete way to be able to master all the data science tools. In fact, I am grateful for the time and the efforts spent during these months.

However, I would like to thank all the people that help me online. Especially, “Gnavia007”.

See you in the next episode of the learning step 😊