Berlin Restaurants Analysis

Introduction

- In this project we will try to find an optimal location for an italian restaurant in **Berlin**, Germany.
- Berlin is the capital and the largest city in Germany with lots of touristic and historic sights. This international city is popular with museums, galleries, nightlife, festivals, performing arts and cuisine. We will focus on the **center areas where the young population and the tourists mostly hang out** and enjoy the city. Therefore, we will eliminate some boroughs/neighborhoods at the beginning of our analysis. We are also particularly interested in **the areas with lower crime rates**. We need to mention that Berlin is pretty safe, there are not any drug wars, murderers, dark ghettos or anything like that.
- By using data science powers we will try to find the most promising neighborhoods that fulfill our requirements.

Data

Based on definition of our problem, factors that will influence our decision are:

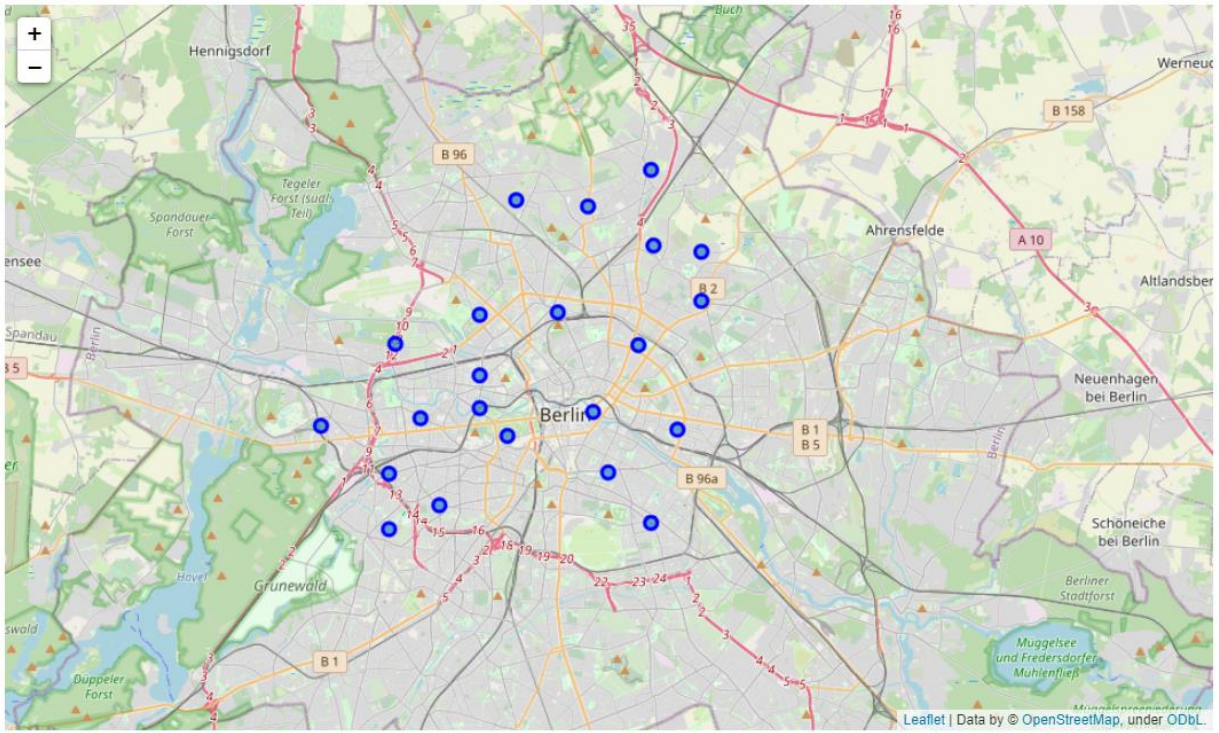
- number of existing Italian restaurants in the neighborhood (any type of restaurant)
- popularity of neighborhood
- crime rates in the neighborhood

Following data sources will be needed:

- name of boroughs and neighborhoods will be obtained from Wikipedia and their coordinates will be generated by using **Google Maps API geocoding**
- number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**
- crime statistics will be obtained from **Kaggle**

Visualizing the neighboorhods on map

After getting the coordinates our preferred neighborhoods, we visualized them on the folium map



	Boroughs	Neighborhoods	Latitude	Longitude
0	Charlottenburg-Wilmersdorf	Charlottenburg	52.515747	13.309683
1	Charlottenburg-Wilmersdorf	Charlottenburg-Nord	52.540525	13.296266
2	Charlottenburg-Wilmersdorf	Halensee	52.497226	13.292999
3	Charlottenburg-Wilmersdorf	Schmargendorf	52.478902	13.292996
4	Charlottenburg-Wilmersdorf	Westend	52.513399	13.255842
5	Charlottenburg-Wilmersdorf	Wilmersdorf	52.487115	13.320330
6	Friedrichshain-Kreuzberg	Friedrichshain	52.512215	13.450290
7	Friedrichshain-Kreuzberg	Kreuzberg	52.497644	13.411914
8	Mitte	Gesundbrunnen	52.550920	13.384846
9	Mitte	Hansaviertel	52.519123	13.341872
10	Mitte	Mitte	52.517885	13.404060
11	Mitte	Moabit	52.530102	13.342542
12	Mitte	Tiergarten	52.509778	13.357260
13	Mitte	Wedding	52.550123	13.341970
14	Neukölln	Neukölln	52.481150	13.435350
15	Pankow	Heinersdorf	52.572825	13.437015
16	Pankow	Niederschönhausen	52.585806	13.401397
17	Pankow	Pankow	52.597917	13.435316
18	Pankow	Prenzlauer Berg	52.539847	13.428565
19	Pankow	Stadtrandsiedlung Malchow	52.571019	13.463285
20	Pankow	Weißensee	52.554619	13.463002
21	Pankow	Wilhelmsruh	52.588012	13.362206

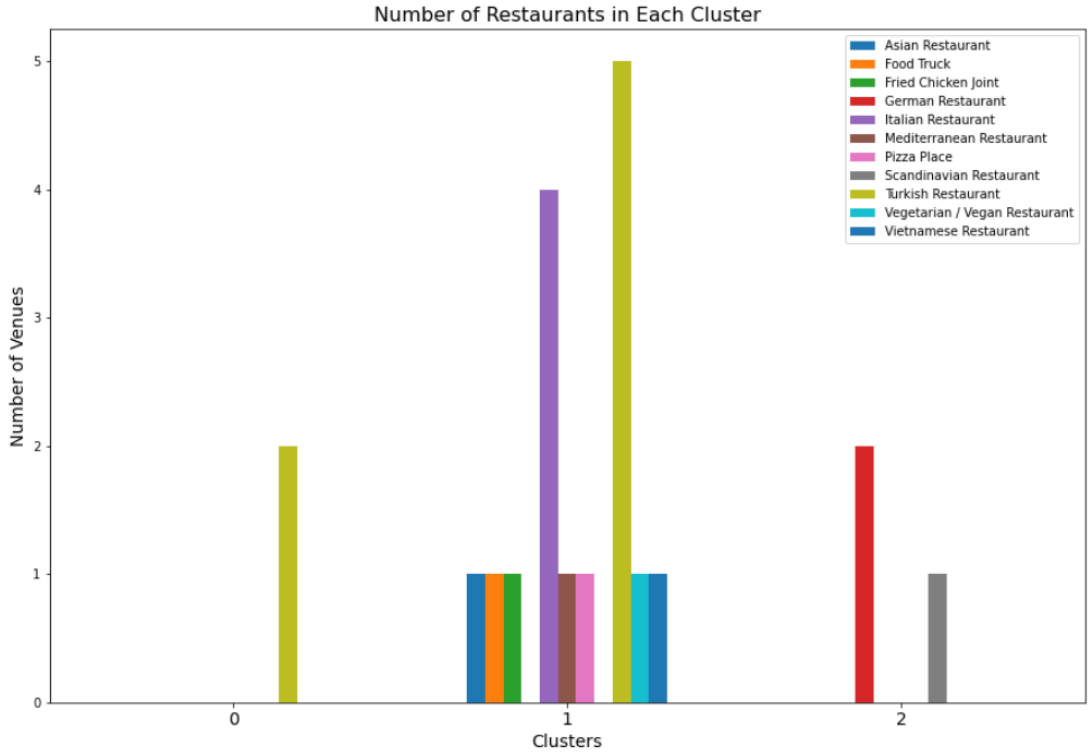
Exploring Venues in Neighborhoods

By using Foursquare API we found 402 existing venues and 62 venue categories in our neighborhoods. Then we analysed them to find out the venue category frequencies in each neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Charlottenburg	Pizza Place	Chinese Restaurant	Italian Restaurant	Vietnamese Restaurant	Burger Joint	Pet Café	Currywurst Joint	German Restaurant	Falafel Restaurant	Fast Food Restaurant
1	Charlottenburg-Nord	Turkish Restaurant	Vietnamese Restaurant	Donut Shop	Halal Restaurant	Greek Restaurant	German Restaurant	Gastropub	Fried Chicken Joint	French Restaurant	Food Truck
2	Friedrichshain	Vegetarian / Vegan Restaurant	Middle Eastern Restaurant	Turkish Restaurant	Bagel Shop	African Restaurant	Burrito Place	Gastropub	Italian Restaurant	Donut Shop	Creperie
3	Gesundbrunnen	Turkish Restaurant	Italian Restaurant	Pizza Place	Chinese Restaurant	Halal Restaurant	Kebab Restaurant	Fast Food Restaurant	Falafel Restaurant	Middle Eastern Restaurant	Donut Shop
4	Halensee	Turkish Restaurant	Italian Restaurant	Japanese Restaurant	Spanish Restaurant	Korean Restaurant	Mediterranean Restaurant	Eastern European Restaurant	Greek Restaurant	Soup Place	German Restaurant

Clustering Neighborhoods

By using K-Means clustering method we clustered the neighborhoods into 3 categories and then we checked the most common venue categories in each neighborhood.

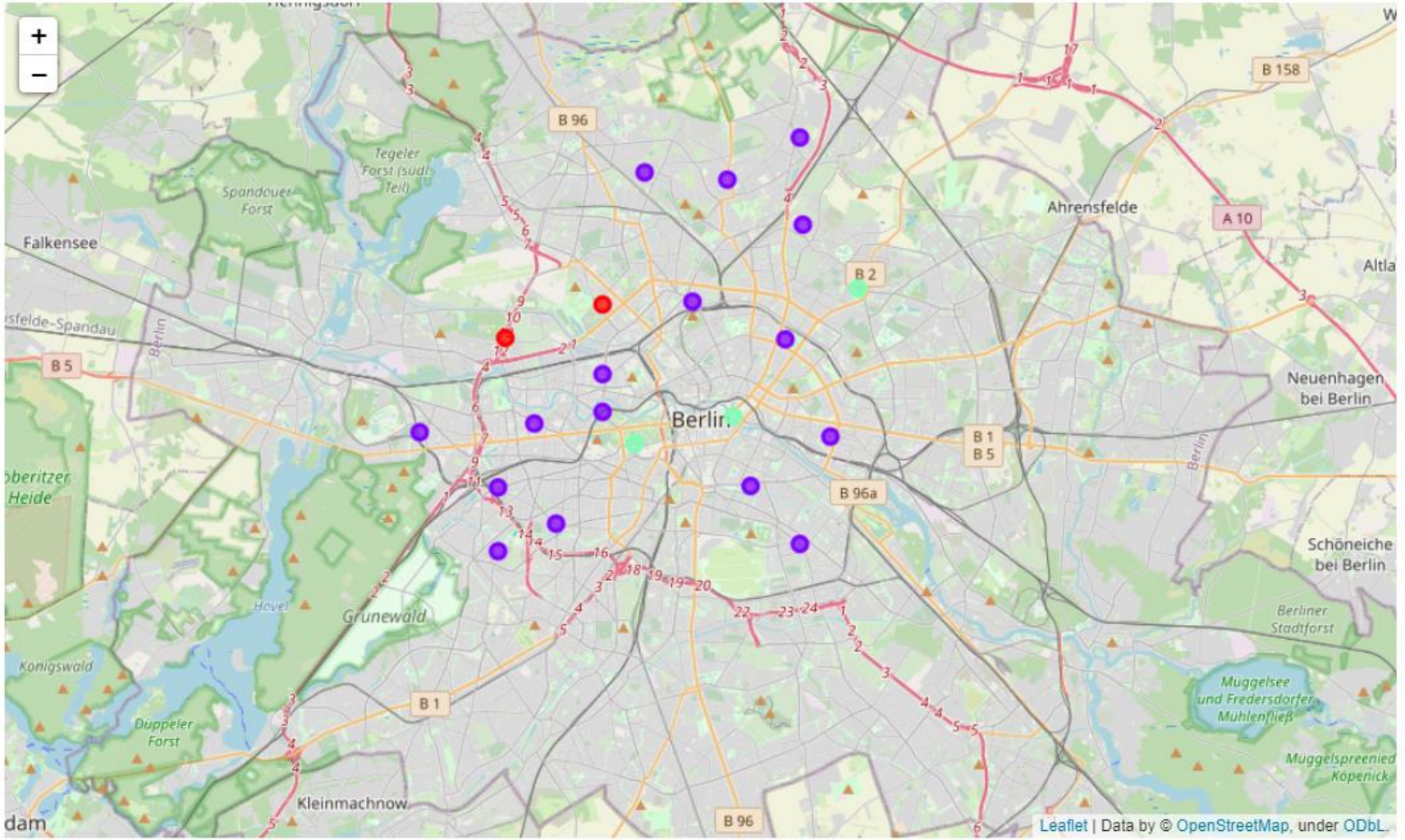


Cluster 0 : Turkish restaurant areas

Cluster 1 : Turkish/Italian restaurant areas

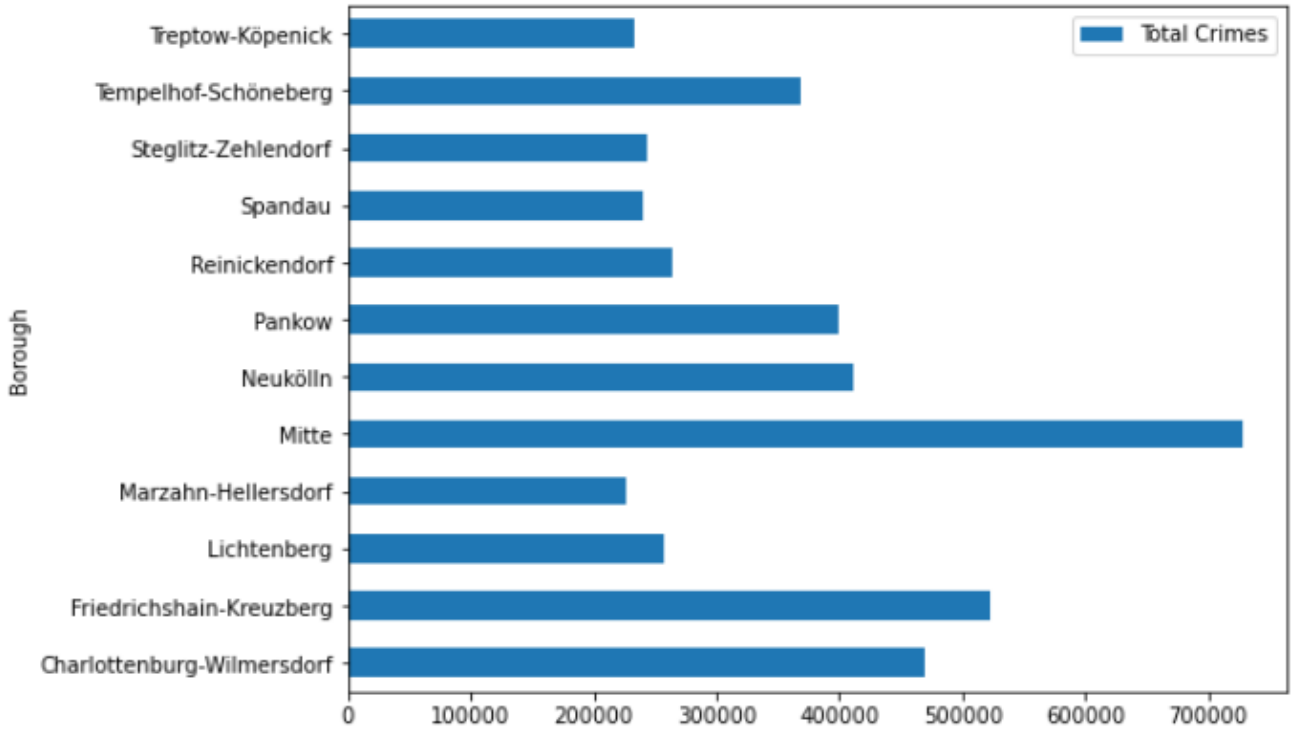
Cluster 2 : German restaurant areas

Visualizing Clusters on map



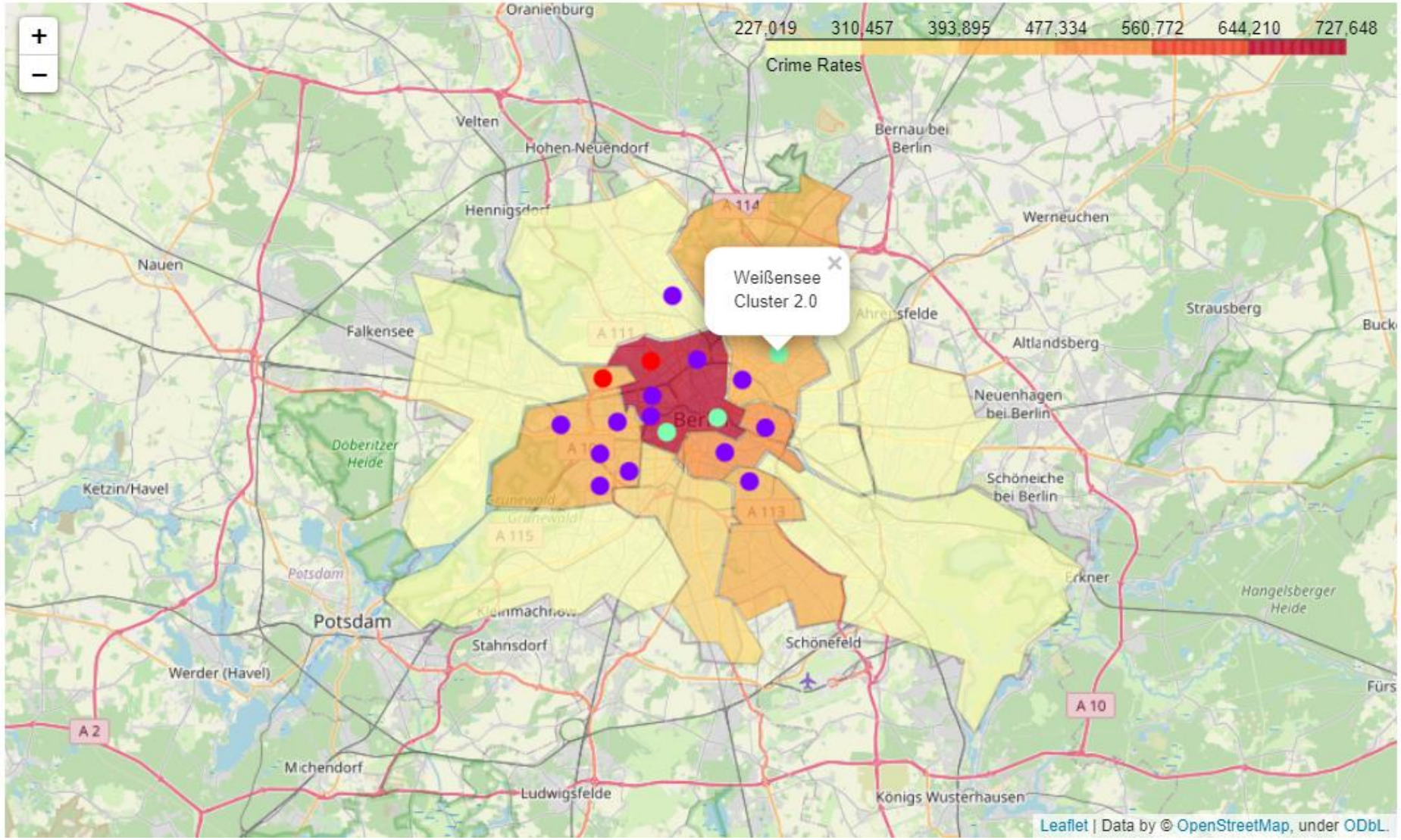
Crime Statistics Analysis

We analyzed the crime statistics from by visaulizing them on a bar-chart.



	Borough	Total Crimes	Latitude	Longitude
0	Charlottenburg-Wilmersdorf	469618	52.507856	13.263952
1	Friedrichshain-Kreuzberg	522186	52.501115	13.444285
2	Lichtenberg	256785	52.532161	13.511893
3	Marzahn-Hellersdorf	227019	52.522523	13.587663
4	Mitte	727648	52.517885	13.404060
5	Neukölln	411786	52.481150	13.435350
6	Pankow	399058	52.597917	13.435316
7	Reinickendorf	263585	52.604763	13.295287
8	Spandau	239582	52.535788	13.197792
9	Steglitz-Zehlendorf	243033	52.429205	13.229974
10	Tempelhof-Schöneberg	368459	52.440603	13.373703
11	Treptow-Köpenick	233173	52.417893	13.600185

Combining cluster and crime statistics



Results and Discussion

- Berlin is the greatest city in Germany with its high energetic young population and very impressive history. Throughout the analysis our focus was on the most popular neighborhoods. Hence, we picked some boroughs like Charlottenburg-Wilmersdorf, Friedrichshain-Kreuzberg, Mitte, Neukölln and Pankow and explored all the restaurant in there. Some of the neighborhoods were eliminated also since they do not have any interesting places.
- After exploring the restaurants in our selected neighborhoods, we clustered them into 3 cluster groups by using an unsupervised learning(K-Means). This process helped us to group the neighborhoods according to frequency of specific restaurant types.
- At the end we combined the crime statistics of the city with our results and visualized them on the map in order to see which boroughs are safer than the others. Moreover, by visualizing the cluster groups on the same map made it easier to make a decision of an optimal location for an Italian restaurant.