Berlin Restaurants Analysis

Yavuz Dogan

May 21, 2021

Table of contents

1. Introduction: Business Problem

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.

2. 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**

2.1. Neighborhood Candidates

As we mentioned in the introduction part, we will eliminate some boroughs that we are not interested in, since they are either too far away from city center or they do not have any touristic sights.

	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

	Boroughs	Neighborhoods	Latitude	Longitude
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

Table 2.1 preferred neighborhoods and coordinates

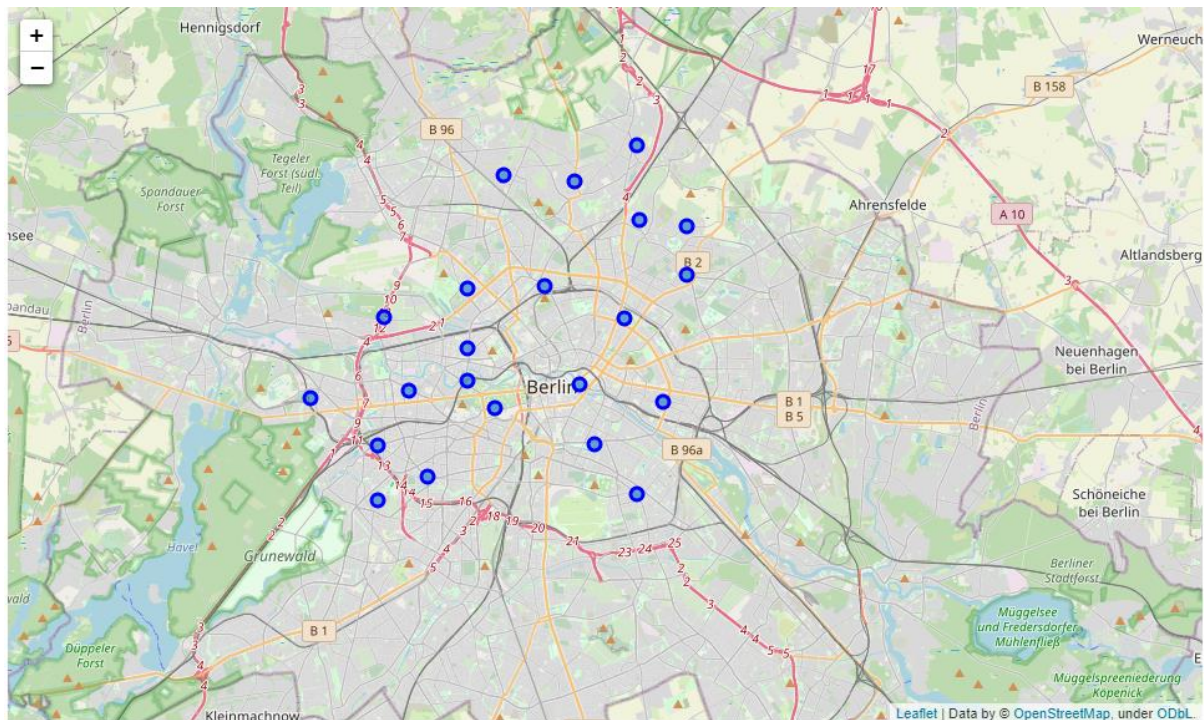Visualizing the neighborhoods on Folium Map:

Fig. 2.1 Neighborhoods on Folium map

2.1. Foursquare

Now that we have our favorite neighborhoods, we need to find out what restaurants are located in these neighborhoods. In order to do that, we use Foursquare API and generate a table consisting of 402 restaurants in those neighborhoods.

Visualizing five of them on the table:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Charlottenburg	52.515747	13.309683	Trattoria Rathaus Piazza	52.516778	13.308748	Trattoria/Osteria
1	Charlottenburg	52.515747	13.309683	Zur Mieze - Katzenmusikcafé	52.515899	13.304765	Pet Café
2	Charlottenburg	52.515747	13.309683	Sole d`Oro CUCINA ITALIANA	52.514660	13.305146	Pizza Place
3	Charlottenburg	52.515747	13.309683	Falafel, Schawarma & Halloumi	52.512515	13.305099	Falafel Restaurant
4	Charlottenburg	52.515747	13.309683	Genki Sushi	52.514541	13.304827	Sushi Restaurant

Table 2.2 Five venues in the neighborhoods

3. Methodology

As a first step we collected the necessary data for our analysis:

- The most touristic and crowded neighborhoods and their coordinates
- The all restaurants that are located in these neighborhoods (Foursquare)

In the next step we will explore the neighborhoods by checking the **restaurant density** in these areas.

In third step we will be clustering the neighborhoods according to the density and type of restaurants by using an unsupervised learning method (**K-Means clustering**).

In fourth and final step we will review **crime statistics of Berlin** and by taking these statistics into the consideration we will try to find the most promising neighborhood candidates for our restaurant.

4. Analysis

4.1. Restaurants Analysis

During our analysis we will try to explore all the restaurants in the neighborhoods by checking their types and locations.

There are 63 unique venue categories totally.

We group the neighborhoods and take the mean value for each venue category. Now we have five top venue category frequencies for each neighborhood.

```
----Kreuzberg----
                venue  freq
0   Turkish Restaurant  0.22
1   Italian Restaurant  0.11
2    Korean Restaurant  0.07
3    German Restaurant  0.07
4   African Restaurant  0.04


----Mitte----
                   venue  freq
0       German Restaurant  0.38
1   Vietnamese Restaurant  0.16
2             Pizza Place  0.06
3      Italian Restaurant  0.06
4     Japanese Restaurant  0.06
```

Table 4.1 Venue category frequencies for Kreuzberg and Mitte

Creating a new dataframe and displaying the top 10 venues for 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

Table 4.2 Dataframe of 10 most common venues in neighborhoods

We use K-Means clustering method, which is an unsupervised machine learning method, to cluster our neighborhoods. We want to divide the neighborhoods into 3 clusters. For the next step we will try to find out what is the most common venue in each cluster.

According to the results:

- Cluster 0 :Turkish restaurants areas

- Cluster 1: Turkish / Italian restaurants areas

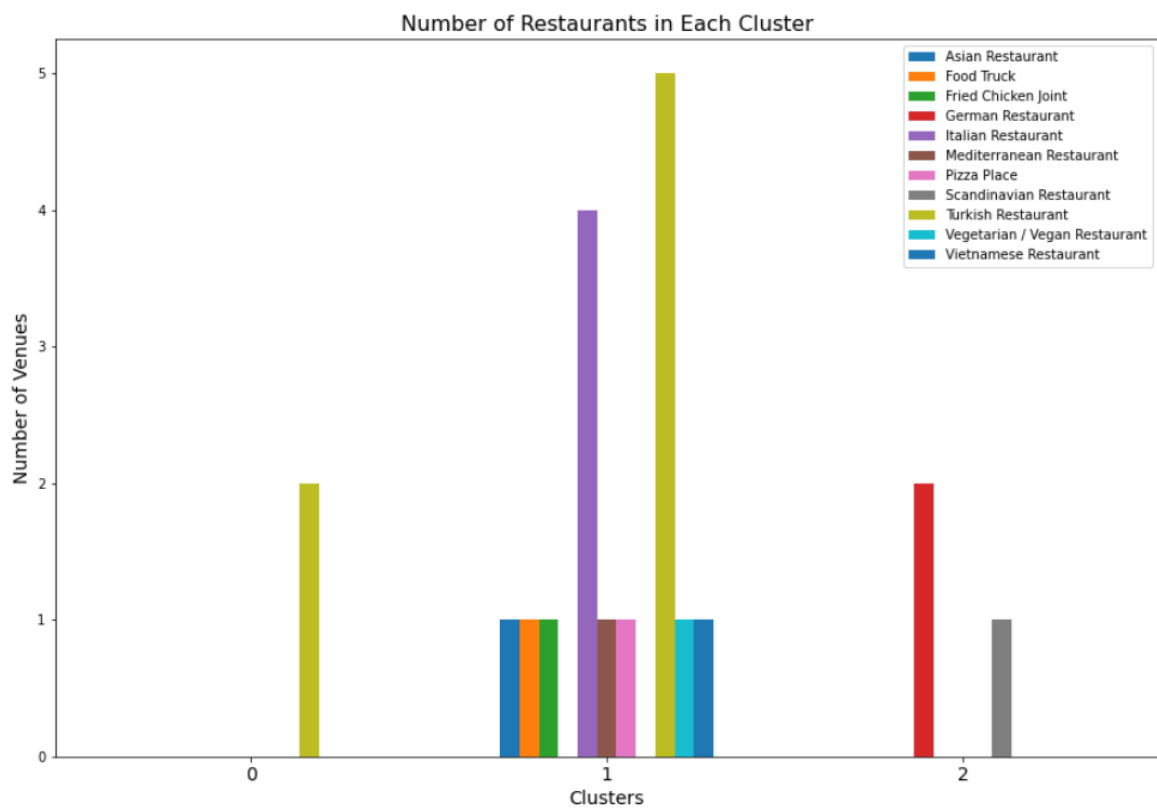- Cluster 2: German restaurants areas
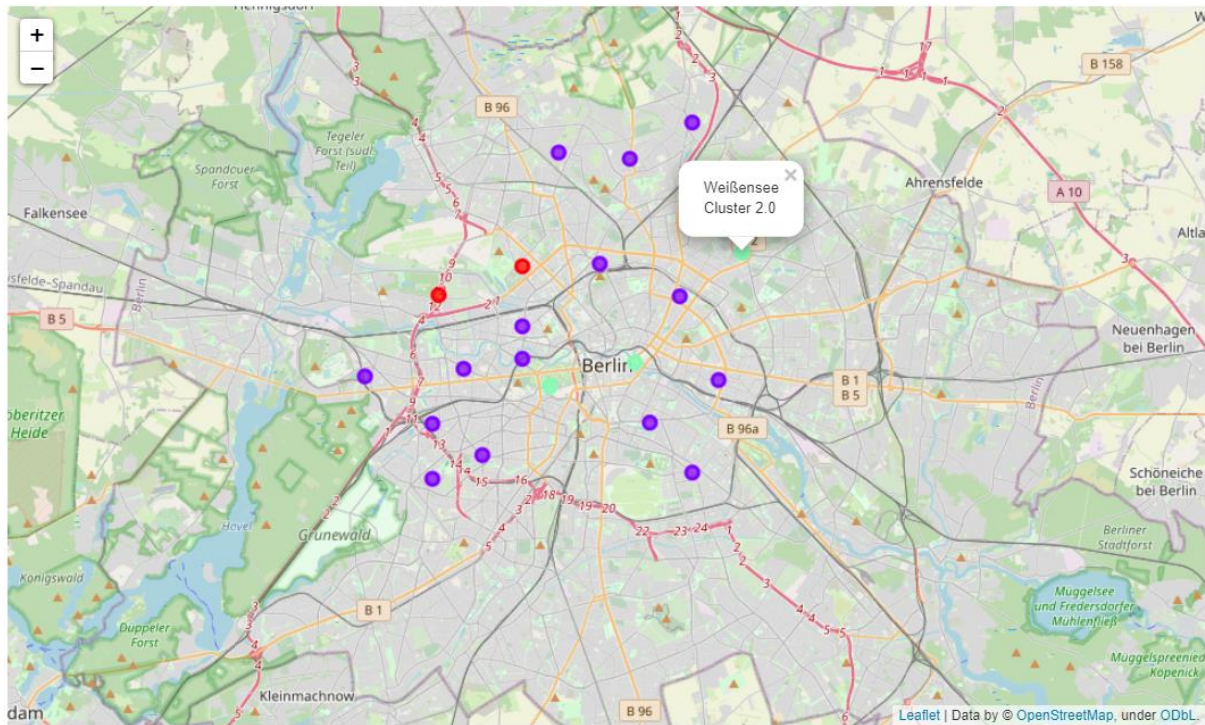

Fig. 4.1 Most common venues in each cluster

Fig. 4.2 Three neighborhood clusters on the map

4.2. Crime Statistics Analysis

In this step we will compare crime statistics (kaggle) of each neighborhood based on the crimes occurred between 2012 and 2019.
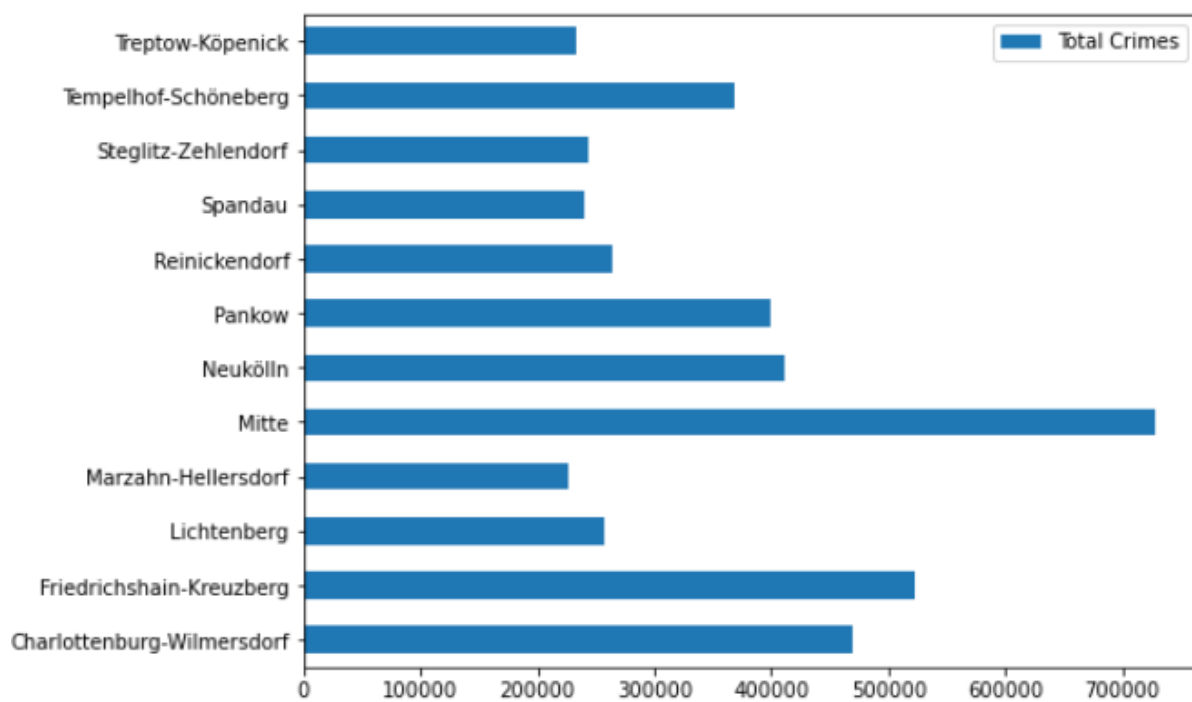

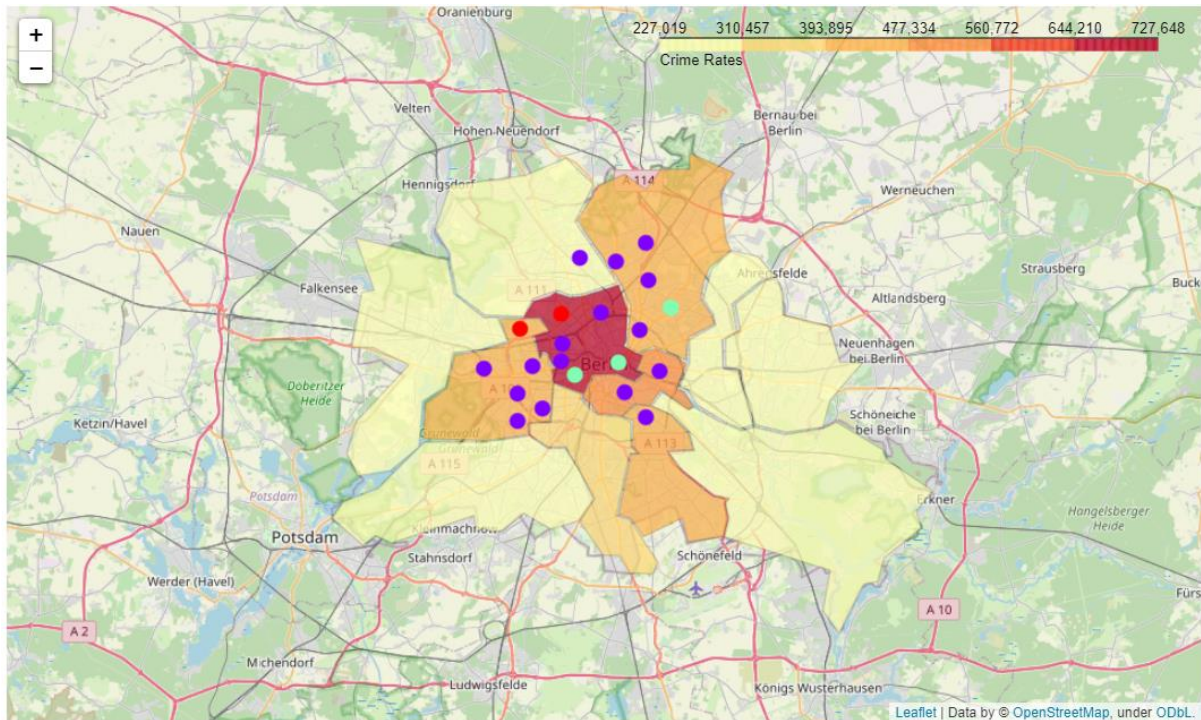Fig. 4.3 Crime rates of neighborhoods

Fig. 4.4 Crime rates and clusters of neighborhoods on the map

5. 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.

6. Conclusion

As we checked the final results on the map, we realize that Mitte has the highest crime rates. As we eliminate this borough, there are only two neighborhoods (Charlettenburg-Nord, Weissensee) remaining as a candidate for our Italian restaurant.

Surely there are many other factors to make a final decision like closeness to parks, water, public transport, levels of noise, rents, real estate availability etc. However, our analysis could give us an idea regarding the current situation of Berlin restaurants and neighborhoods. It still has some potential to be developed further.