The Battle of Neighborhoods

Indian Restaurant in Berlin

Project to obtain Data science professional Certificate

International Business Machines Corporation (IBM)

Nabil Sifo

August 31, 2020

1 Introduction

1.1 Berlin

After the fall of the Berlin wall On 9 November 1989 and the reunification of Berlin those differences between the two parts of the city start to faint with time.Berlin become the hub of cosmopolitan movement in Germany and many people from different nationalities moved to Berlin. Therefore the food market in Berlin turned to be very divers and international.

Berlin is the capital city of Germany. It is known to be a cosmopolitan city with a very high number of foreigners from all over the world and very active tourism. Due to the high diversity of the Berlin population, the food market is divers as well. In Berlin, there are restaurants from almost all cultures in the world, including Indian restaurant. The Indian food is beloved in Germany, and it is probably in third place after the Italian and Turkish food. Therefore, this work will explore the neighborhoods in Berlin to search for a suitable location to open an Indian restaurant.

1.2 Business problem

In this project we will investigate the best place to open an Indian restaurant in Berlin. Although the city center might be a trivial option we will use the data analysis and the machine learning techniques to find all the potential candidates.

1.3 Interest

people who want to open an Indian restaurant. Additionally, people who are interested in opening south Asian restaurant might find this analysis useful.

2 Data acquisition and cleaning

2.1 Data sources

The data will be accuer from deferent resources:

1. The Wikipidia page Verwaltungsgliederung Berlins provides a well ordered list of all the boroughs and there respictive neighboorhoods. This table will be extract using pandas library. 2. The website Das Örtliche provides a full list of all Zip codes in Berlin wich will be used later to get a better accurecy for addresses. the list will be exctract using pandas

3. utilizing Geocoder the geographical coordinates of each zip code will be imported to serve as a center of neighborhood

4. Using Open street maps from Google a geojson file for each Borough in Berlin as well as the whole city will be imported.

5. Utilizing the Foursquare API all th venius will be imported for exploring the neighborhoods in Berlin and later with cluster analysis searching for the neighboors that are good for opining an indian restaurant

2.2 Data cleaning

The data needed to be prepared first merging the two data sets of Berlins addresses and zip codes in one data set. Second, it turned out that both Geocoder and Folium have a problem with dealing with the special characters in the German alphabet. Therefore, the special german alphabet in this data set was replaced with the approximate English litter.

the Geojson for Berlin borough is constructed from merging 12 geojson files of all the boroughs in Berlin.

3 Exploratory Data Analysis

After the collecting the list of boroughs and neighborhoods and merge them together. The coordinate of each address were obtained using Geocoder. In this code the ArcGis decoder is used, but it is not the only option, there are many others like Google Maps, AzureMaps, Bing, etc. Some of them require API keys, while others do not need. The end product data frame look like shown in Figure1.

	PLZ	Borough	Neighborhood	Latitude	Longitude
0	10115	Mitte	Mitte	52.531890	13.378140
1	10115	Mitte	Wedding	52.532142	13.385789
2	10117	Friedrichshain-Kreuzberg	Kreuzberg	52.500610	13.418630
3	10117	Mitte	Mitte	52.521190	13.424140
4	10117	Mitte	Tiergarten	52.516465	13.388040

Figure 1: An overview of the data frame of addresses and locations in Berlin.

The json file that include all the boundaries of borough in Berlin is shown in Figure2. Figure three is the same map after including all the addresses we have

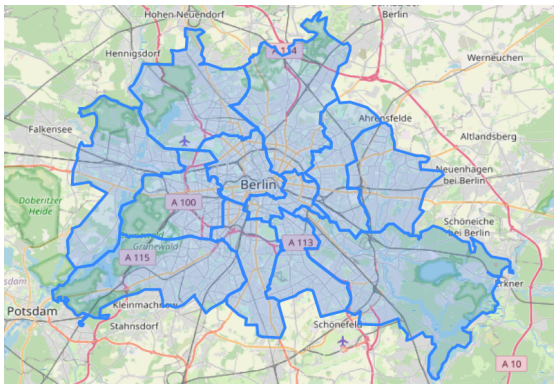

Figure 2: map of the city Berlin with the borough borders.

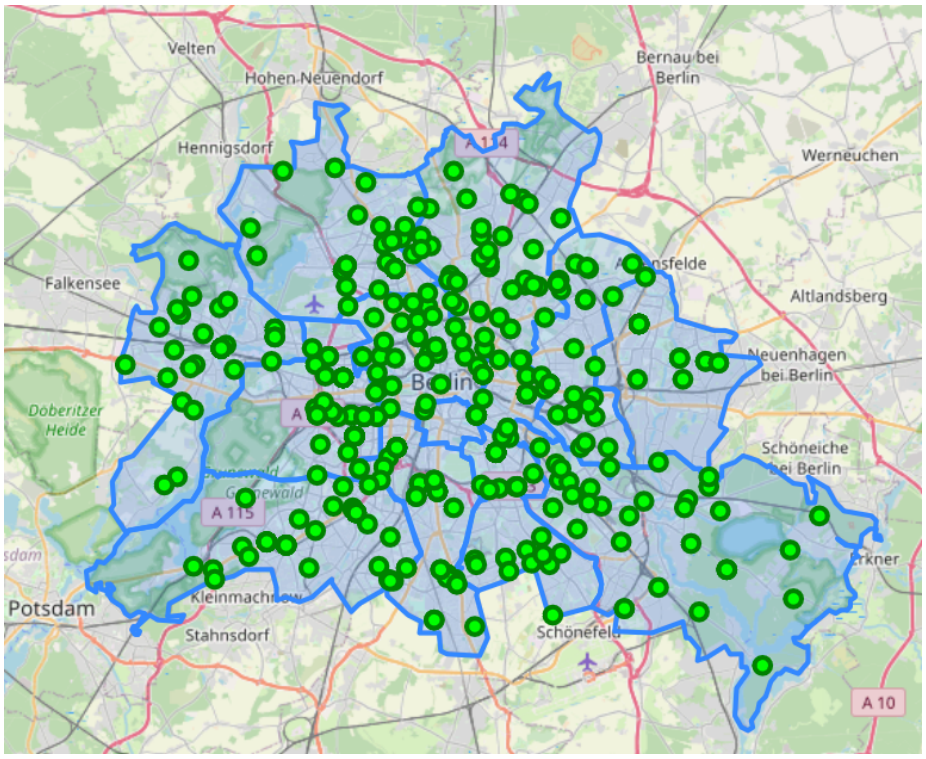

Figure 3: map of the city Berlin with the borough borders and the addresses used in this project as green circles.

After using Foursquer to retrieve 1000 venues in a range of 1000 m from each address we have the number of venues we obtained is shown in Figure4.
Focusing only on the Indian restaurant business and visualized the result as heat map we can see that Marzahn-Hellersdorf have no Indian restaurant.while the highest density of Indian restaurant is in Tempelhof Schoneberg see Figure5.
Finally lets take a look at the 10 most crowded neighborhoods with Indian restaurants. Those 10 neighborhoods will be used as reference to find out which neighborhood clusters are best to open an Indian restaurant.

Borough	ZIP
Charlottenburg Wilmersdorf	2137
Friedrichshain Kreuzberg	1410
Lichtenberg	465
Marzahn Hellersdorf	74
Mitte	1285
Neukolln	1189
Pankow	852
Reinickendorf	218
Spandau	849
Steglitz Zehlendorf	472
Tempelhof Schoneberg	1476
Treptow Kopenick	321

Figure 4: The table shows the number of venues retrieved from Foursquer in each borough in Berlin.

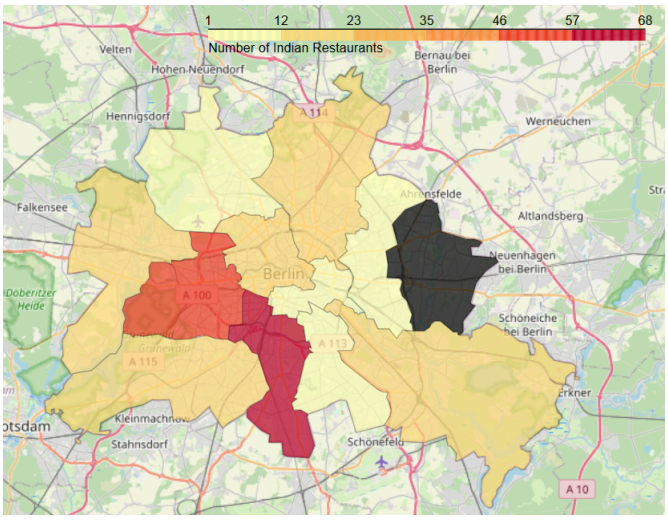

Figure 5: The table shows the number of venues retrieved from Foursquer in each borough in Berlin.

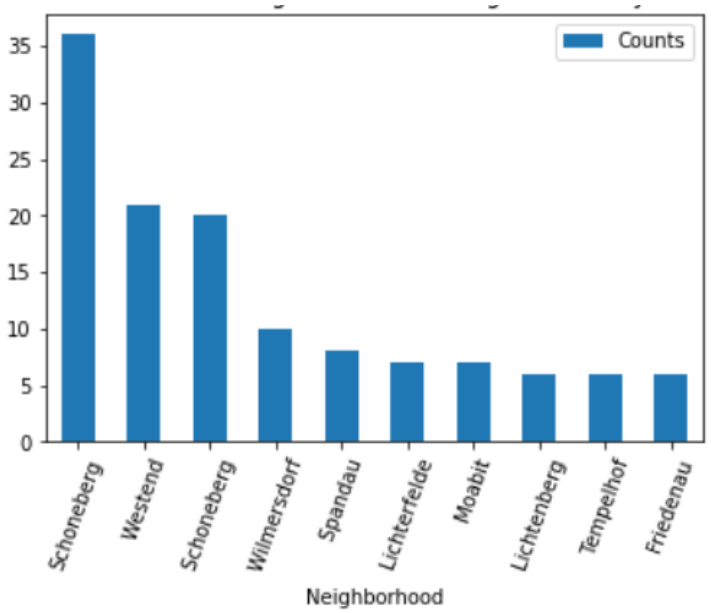

Figure 6: Number of Indian Restaurants in Neighborhood with highest density of Indian Restaurants.

4 Cluster analysis

For the cluster analysis the K-mean methods will be used to find out all the similar groups of neighborhoods in Berlin. The ideal number of the clusters (7) is determined using the elbow methods Figure6.

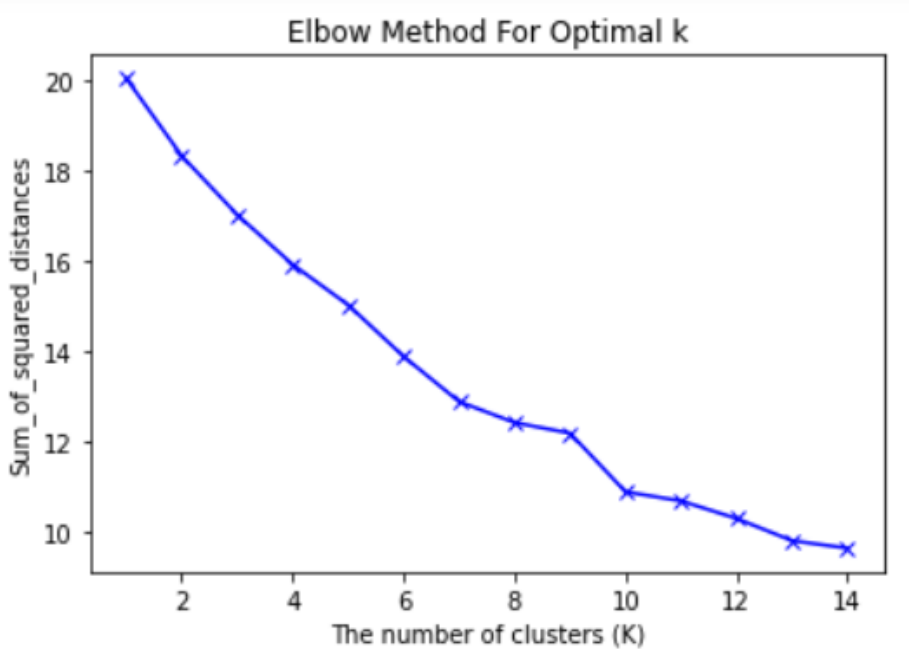

Figure 7: The elbow method to determine the optimal number of clusters K.

In this Project we will search for neighborhoods with the lowest number of Indian restaurants in them, because this means that there is still a chance for a new Indian restaurant to be opened. Neighborhoods with no Indian restaurant will not be considered in these analyses for this might refer to the absence of interest from people who live there in the Indian food.
The analysis will be done by finding the neighborhoods with the highest density of an Indian restaurant. Hence the neighborhoods where the Indian restaurant business is successful, and then with the help of cluster analyses We will we will find the neighborhoods that have a low number of Indian restaurants, but are

similar to the neighborhoods with the highest density of an Indian restaurant. This similarity will be determined by all the venues in the neighborhoods.

The existence of a certain types of venues with a high level of density can tell a lot about the market orientation in this area. This is not only depends on the restaurants and there type and density, but also it depends on all other venues because it tells in an indirect way, what is the general mood of people in those neighborhoods.Therefore, in the cluster analyses all types of venues will be considered to achieve the highest level of accuracy.

Figure7 shows the neighborhoods with highest density of Indian restaurant in red and the recommended neighborhoods to open an Indian restaurant with green.

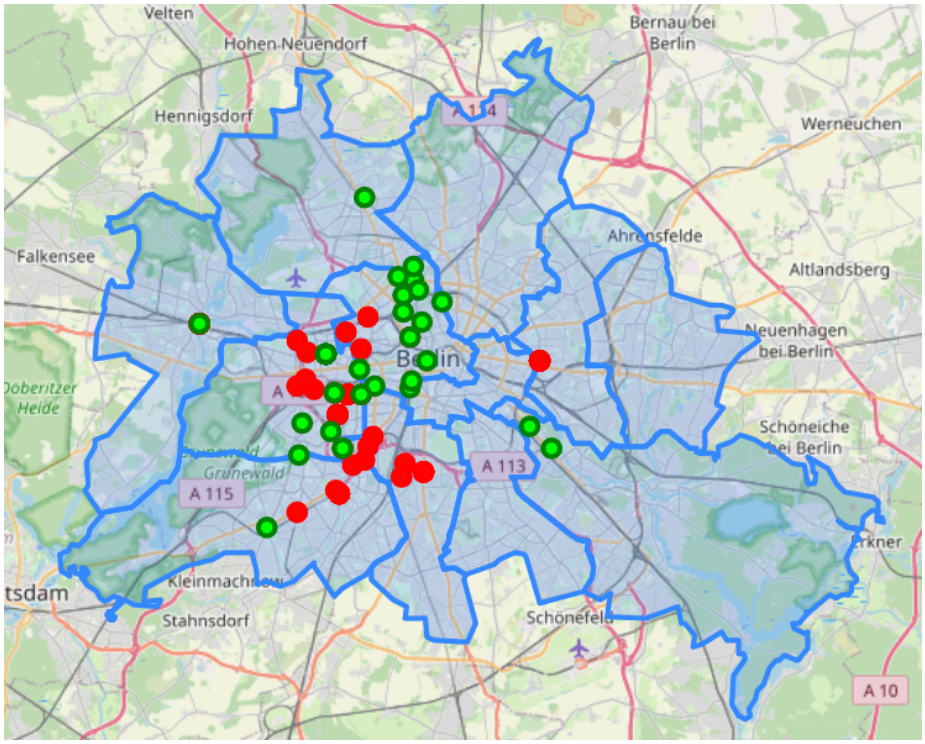

Figure 8: map of the city Berlin with the best Neighborhoods to open an indian restaurant in green and the nieghborhoods with the highest density of indian restaurant in red.

5 Conclusion

5.1 Conclusion

The best Neighborhood to open an Indian Restaurant is mostly the city center with only some exceptions like 14165,Steglitz-Zehlendorf,Zehlendorf and 14109,Spandau,Kladow.
the city center makes a full sense because of the active tourism, and the high density of people in the street. Therefore the city center will be a very suitable place to open the restaurant.
However, opening the restaurant away from the city center might have a lot of risks, but it is also a higher chance to make an extra successful business due to the lake of competition.

5.2 Future research recommendations

There are some neighborhoods which are highly promising of success Indian restaurant business.However, this analysis must be adjusted with better distributed addresses, maybe an artificial grid will produce better results. The addresses based on zip codes follow the houses density and might be not the best core location to do the search of nearby venues.