Capstone Project – The Battle of the Neighborhoods

Author : Palomares Cervantes, Sergio

IBM/Coursera Data Science Professional Certificate

Author Note

This investigation was done as part of the Capstone project as required for the IBM Data Science Professional Certificate offered through the COURSERA platform.

Abstract

Frankfurt am Main is the financial capital of Germany, and with the ongoing “Brexit” process, will soon be considered the financial capital of the European Union, dethroning London. The city is also considered to be an intercultural center with a high percentage of ethnicities and expats from all over the world living there. One of the biggest groups is those of Mexican descent. This has sparked a growth in the demand for authentic Mexican food and commodities. The present project will provide and initial data analysis of the existing Mexican Restaurants in Frankfurt am Main, as well as its relative position to other restaurants and points of high commuter density, in order to evaluate an optimal area for opening a Mexican Restaurant and present it to stakeholders.

Keywords: Mexican Restaurant, Frankfurt am Main, Data Analysis.

Contents

Tables and Figures

Capstone Project – The Battle of the Neighborhoods

INTRODUCTION OF BUSINESS PROPOSAL

This project will be to analyse the possibility of opening a new restaurant that serves authentic Mexican food in Frankfurt. Often considered the city with the most Mexican expats in Germany and having already a fair number of Mexican restaurants but with mostly modified recipes that lose the original taste.

Therefore, the best location will be found that takes into consideration in order of importance the following points:

- Existence of other Mexican restaurants
- Existence of other restaurants
- Proximity to key areas: Close to metro station, plazas, downtown

Acquiring of the data and first cleaning will be performed using Python and its libraries (Pandas, Numpy, Scipy). Subsequently an unsupervised machine learning algorithm will be used to create clusters of neighbourhoods based on the above criteria to then evaluate a recommendation to stakeholders for a location to open the Mexican restaurant.

DATA

Based on the criteria of the task given above, the primary factors will aid in choosing a location are:

- Existing restaurants in each neighbourhood
- Existing Mexican restaurants
- Distance to key establishments: metro station, plazas and downtown.

Neighbourhoods (Stadtteile) will be defined as per description from Wikipedia. And only the main neighbourhoods close to the city centre (Innenstadt, Altstadt) will be considered.

Sources for the data used will be:

- Wikipedia for the neighbourhood denominations
- Restaurant location and type from the Foursquare API
- Folium for providing map visuals

Neighborhoods (Stadtteile)

The different Neighborhoods were taken from the city of Frankfurt am Main Wikipedia page. The corresponding table was scraped using the Wikipedia API and the *geocoders* package the coordinates for the center of each of the neighborhoods was extracted and adjusted to show only the names of the closest neighborhoods to the city center:

Table 1. Neighborhoods (Stadtteile)

INDEX	STADTTEIL	LATITUDE	LONGITUDE
0	Altstadt	50.110442	8.682901
1	Innenstadt	50.112993	8.674341
2	Bahnhofsviertel	50.107741	8.668676
3	Westend-Süd	50.115245	8.662270
4	Westend-Nord	50.126356	8.667921
5	Nordend-West	50.124914	8.677950
6	Nordend-Ost	50.124920	8.692317
7	Ostend	50.115935	8.720546
8	Bornheim	50.133056	8.714932
9	Gutleutviertel	50.097925	8.648964
10	Gallus	50.102309	8.631984
11	Bockenheim	50.122361	8.637532
12	Sachsenhausen-Nord	50.097813	8.673362
13	Sachsenhausen-Süd	50.070725	8.685450
14	Flughafen	50.043030	8.567284
15	Oberrad	50.098791	8.727403
16	Niederrad	50.087053	8.639667
17	Schwanheim	50.082955	8.578917
18	Griesheim	50.101970	8.599960
19	Rödelheim	50.127886	8.604865
20	Hausen	50.134155	8.625835
21	Praunheim	50.147195	8.616688
22	Heddernheim	50.158204	8.641507
23	Niederursel	50.168987	8.622340
24	Ginnheim	50.145069	8.649053
25	Dornbusch	50.147170	8.669729

The Foursquare API was then used to extract the number of venues, their location and type of venue within the vicinity of each neighborhood (below is an excerpt please see attached files to see the full extent of the table):

Table 2. Type and number of Venues

VENUE CATEGORY	NEIGHBORHOOD	NUMBER OF VENUES
AFRICAN RESTAURANT	1	1
AIRPORT LOUNGE	2	2
AIRPORT SERVICE	2	2
AMERICAN RESTAURANT	4	4
APPLE WINE PUB	1	1
ART MUSEUM	8	8
ASIAN RESTAURANT	9	9
ATHLETICS & SPORTS	1	1
AUSTRIAN RESTAURANT	1	1
BBQ JOINT	2	2
BAKERY	12	12
BAR	15	15
BEER BAR	1	1
BEER GARDEN	2	2
BEER STORE	1	1
BISTRO	7	7
BOARD SHOP	1	1
BOOKSTORE	2	2
BOUTIQUE	7	7
BREAKFAST SPOT	4	4
BREWERY	1	1
BUILDING	1	1
BURGER JOINT	10	10
BURRITO PLACE	1	1
BUS STOP	4	4

METHODOLOGY

As a first step the necessary data was gathered, cleaned up and sorted. A preliminary visual analysis using **Folium** to create a labeled map with the location of existing restaurants in each neighborhood was generated.

For the second step, heat maps will be generated to create a more compeling analysis for restaurant density in the different neighborhoods. Particularly compared with the location of our venues of interest, to see where the most transit occurs.

The final step will be to use **K-Means clustering** to identify particular areas with distinct venue characteristics and superimpose them to our heat maps to choose the ideal location with the venues of interest and low restaurant density.

Preliminary mapping

Using Folium, a visual is created to see the locations of Mexican restaurants (red) as well as for all restaurants (green) comparatively to the centers of the neighborhoods (blue).

Figure 1. Mexican restaurant locations

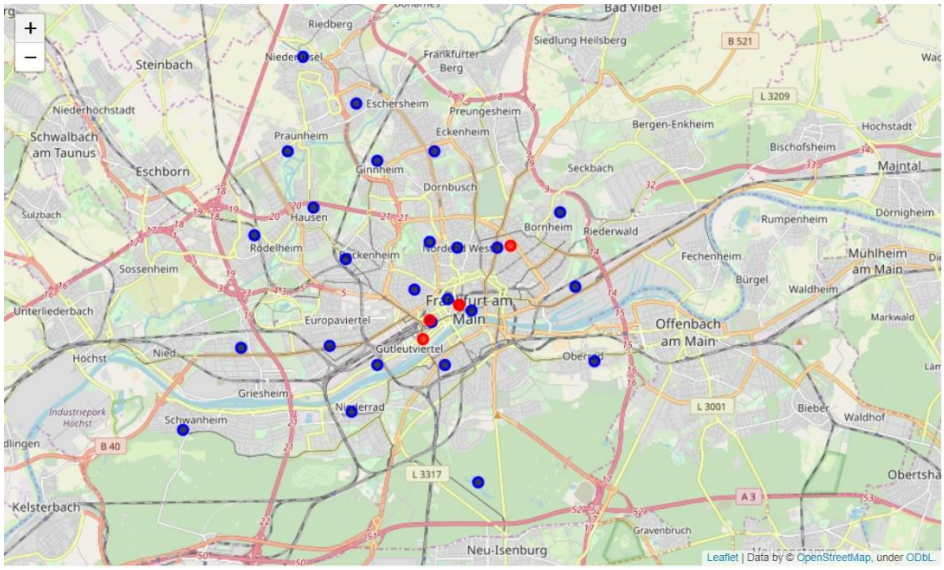
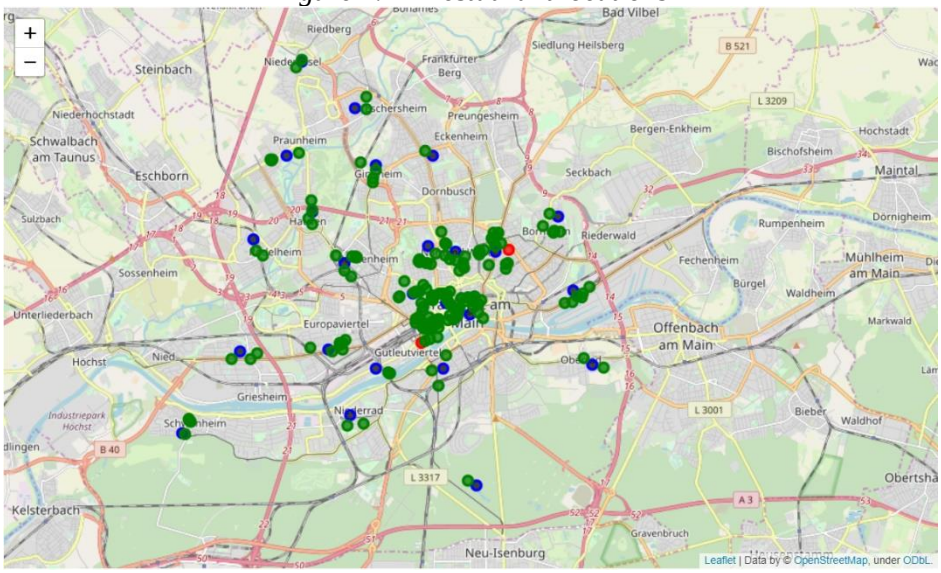


Figure 2. All restaurant locations



Taking the above maps as a reference to where the existing Mexican Restaurants are located, as well as the relative overall restaurant density as compared to the different neighbourhoods chosen, the next step will be to locate places of interest.

The places of interest will be defined by such venues identified by Foursquare as places were there are normally large crowds and/or places with a lot of people transiting such as:

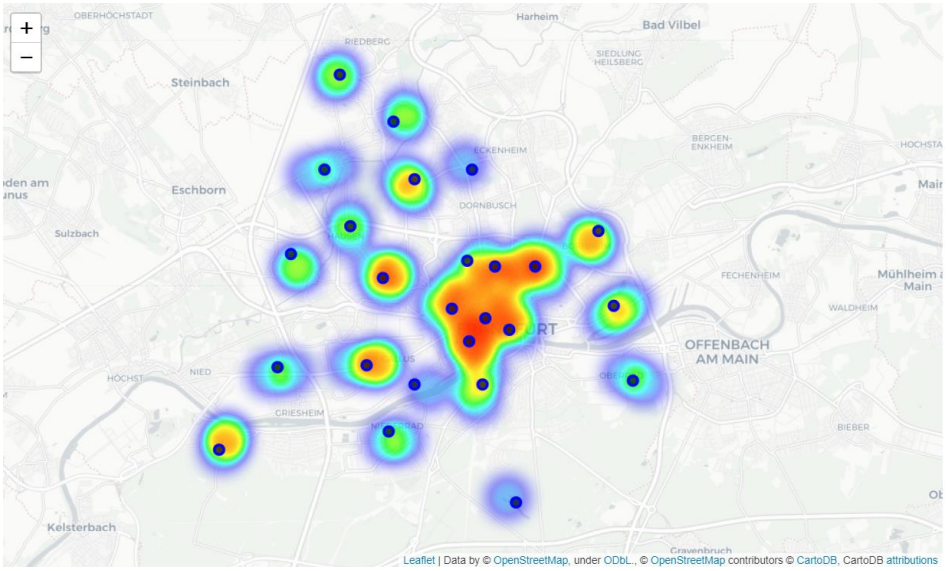
- Bus Stop
- Metro Station
- Hotel
- Park
- Plaza
- Theater

A DataFrame denoting these places as well as the restaurants will be created in order to identify the different clusters using an unsupervised learning algorithm.

RESULTS

The heat map generation function of Folium will be used to get a visual of the restaurant density:

Figure 3. Restaurant Density



Most restaurants seem to be around the **Bahnhofsviertel**. It seems that the existing Mexican restaurants are sparse and only found in four areas of Frankfurt. This looks very promising with respect to location of a new Mexican restaurant.

When coupled with the previous Restaurant heatmap the areas of **Innenstadt** and **Altstadt** both of which are classified as downtown Frankfurt seem to be quite free of restaurant overcrowding.

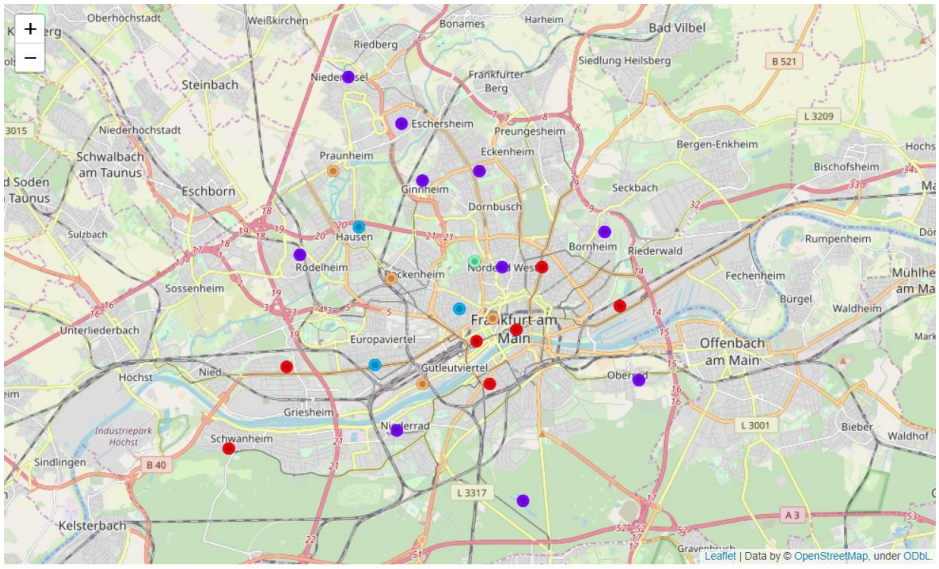
K Means Clustering

Five clusters were chosen given that there are only 26 Neighborhoods considered. Increasing the number of K-clusters would only give us then each neighborhood as its own clusters and lose the effectivity of the model.

For the clustering the information of the venues of interest is included to narrow the options to select an optimum.

A map of the neighborhood centers is rendered with color-coding for their specific cluster labels:

Figure 4. K-Means clustering results

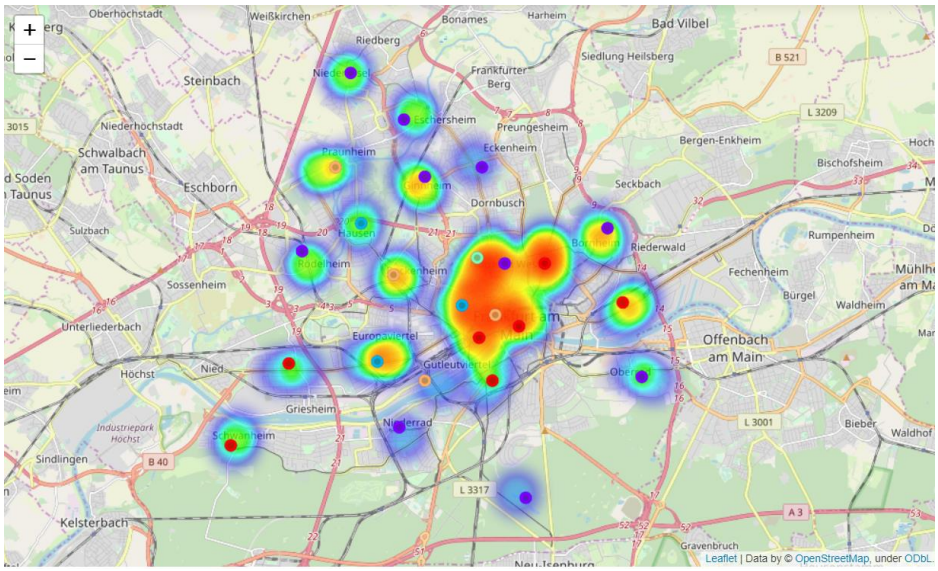


As can be seen from the different clusters, those neighborhoods that are around the downtown area all belong to one cluster giving us an indication of the saturation of venues of interest.

Superimposing the heat map to see how the clustering is affected also by the existence of restaurants one can see the relative saturation of restaurants in the downtown area. However, through further inspections by zooming in to two areas of interest mainly: Innenstadt and Altstadt one can see that they belong to different clusters although a lot of the venues found in the area are very similar. This is due to the proximity of the **Bahnhofviertel** and the existence of a Mexican restaurant close to the Altstadt region, as noted by the orange heat source.

This leaves an option to go more for the Innenstadt, where there are patches and gaps for establishing a Mexican restaurant. Although not belonging to a group where the venues of interest are higher (red label cluster).

Figure 5. Clusters overlaid with restaurant density



DISCUSSION

The analysis shows that there is a relatively low number of Mexican restaurants in Frankfurt. This results in a wide possibility for location for an opening a new Mexican Restaurant in many different neighborhoods. However, under further scrutiny as to the aspects that make several neighborhoods similar/equivalent as per K-Means Clustering method, it seems that there is an optimum close to the city center, Bahnhofviertel, Innenstadt, Altstadt. These three options seem plausible.

When analyzed alongside the density of existing restaurants in general one can see that the Bahnhofviertel is already overcrowded with options, as well as having an existing Mexican restaurant specifically. When looking a bit further, the Altstadt, which belongs to the same cluster, seems also to be an alternative. There is however a Mexican restaurant already, but at the what appears to be the boundary of Altstadt and Innenstadt. Innenstadt, belonging to a different cluster, is however also an alternative.

Other locations belonging to the same cluster are also available with no existence of direct competition as well as lower restaurant density. One drawback, however, is their relative distance to the city center, which could provide less customers, but also a possibility for lower rent cost.

CONCLUSION

The aim of this project was to be able to provide alternatives for stakeholders of the plausibility of opening a Mexican Restaurant in Frankfurt am Main, Germany.

Using K-mean clustering and visual aids (heat maps, map labels) different neighborhoods of Frankfurt can be established based on the venues located there and the existence of restaurants in the vicinity. The lack of Mexican restaurants in the different areas, as well as gaps in different areas with restaurants available make for a promising outlook.

This study, however, only provides a preliminary analysis as further data would be needed for a final decision to be reached. Other factors such as rent costs, permits for venues in the city center, city ordinances, as well as conducted survey for interest in Mexican cuisine, would be needed an placed within an analysis matrix to be able to see if it would be profitable to open a Mexican Restaurant.