IBM Data Science Professional Certificate

Capstone project: "The Battle of Neighbourhoods"

By Yulia Shutko

Table of content:

1. Introduction to business problem.

Frankfurt is a global hub for commerce, culture, education, tourism and transportation, it is the 5th largest city and the financial capital of Germany. This leads to a huge variation of restaurants, bars, night spot, shops, museums and etc. But in this project I would like to consider one part of everyday life of locals: opportunity to make sport.

This project can be interesting for businesspeople who want to open Fitness/Gym in Frankfurt, and they need to know a geographical distribution of sport facilities in different areas of Frankfurt am Main.

2. Data description.

For this project we need following data:

- Geographical data of Frankfurt boroughs with names and coordinates;
- List of sport facilities for each borough.

Following source were used:

- List of areas in Frankfurt am Main: Wikipedia [1];
- Foursquare API [2] with request by specific category;
- Foursquare list of categories [3] to get an id of interested venues;
- GeoJson file with map of Frankfurt am Main [4].

3. Methodology.

In this section following libraries have been used:

- Pandas;
- Requests;
- Bs4;
- Numpy;
- Nominatium;
- Matplotlib .cm and matplotlib.colors;
- Sklearn;
- Folium;
- Geocoder.

As far as data, namely list of areas with districts, from webpage Wikipedia was available, therefore a web scraping technic have been applied to obtain data frame for further processing and analysis. This technic is detailed described in this [5] course. After web scraping follows preparation of data frame for further visualisation. This stage includes saving raw data frame from web page as pandas's, re-naming and re-odering heads and, of course, eliminating of columns with not relevant data.

Next important stage is to associate list of districts with geographical coordinates, such as longitude and latitude. Unfortunately, library Nominatium couldn't work with geographical names of so-called neighbours in this case, therefore coordinates of boroughs were used for this project.

Library Folium was used to mark all boroughs centres of Frankfurt. As an additional layer an GeoJson file [4] with boroughs boundaries was used for better visualisation on the map.

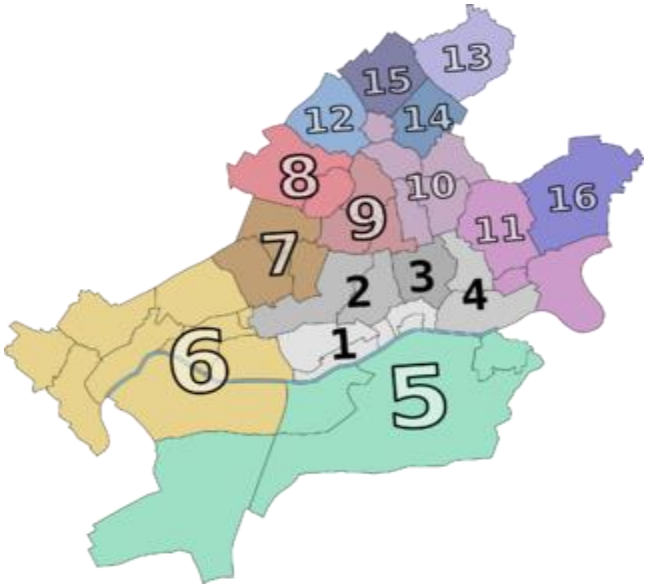
This project focused on distribution analysis of one particular category. Id of any category can be found [3] here. In this project Foursquare API was used to obtain list of venues in category "Gym/Fitness". Obtained data contains not only names of venues with coordinates, but subcategories such as "Gym", "Yoga studio" and etc. Geographical location of all received sport facilities were also visualized on a map with the help of Folium library.

Further steps of data processing include finding top ten common venues for each borough. KMeans from sklearn.cluster library was used for k-mean algorithm. For k-mean algorithm k = 10 was used to obtain satisfied results of clustering. As the final stage of data processing and visualisation, those cluster were marked on the map with the use of Folium library.

4. Results and discussion.

As it was mentioned above, Frankfurt am Main is the big city with a wide variety of venues for each activity. This project is focused only on one aspect, namely gym/fitness facilities and their distribution over the city.

Frankfurt is divided into 16 boroughs:



Picture 1. Boroughs of Frankfurt am Main [6].

Picture 2. Distribution of sport facilities in Frankfurt am Main city.

Nr.	Borough	Neighbourhood
1	Innenstadt I	Gallus, Gutleutviertel, Bahnhofsviertel, Altstadt, Innenstadt
2	Innenstadt II	Bockenheim, Westend-Süd, Westend-Nord
3	Innenstadt III	Nordend-West, Nordend-Ost
4	Bornheim/Ostend	Ostend, Bornheim
5	Süd	Flughafen, Sachsenhausen-Süd, Sachsenhausen-Nord, Oberrad, Niederrad
6	West	Schwanheim, Griesheim, Nied, Sossenheim, Höchst, Unterliederbach, Zeilsheim, Sindlingen
7	Mitte-West	Rödelheim, Praunheim, Hausen, STB 343
8	Nord-West	Niederursel, Heddernheim, STB 426 (Praunheim-Nord)
9	Mitte-Nord	Eschersheim, Ginnheim, Dornbusch
10	Nord-Ost	Eckenheim, Preungesheim, Berkersheim, Frankfurter Berg, Bonames
11	Ost	Seckbach, Riederwald, Fechenheim
12	Kalbach/Riedberg	Kalbach-Riedberg
13	Nieder-Erlenbach	Nieder-Erlenbach
14	Harheim	Harheim
15	Nieder-Eschbach	Nieder-Eschbach
16	Bergen-Enkheim	Bergen-Enkheim

Table 1. List of boroughs and neighbourhood in Frankfurt am Main [1].

[78]:

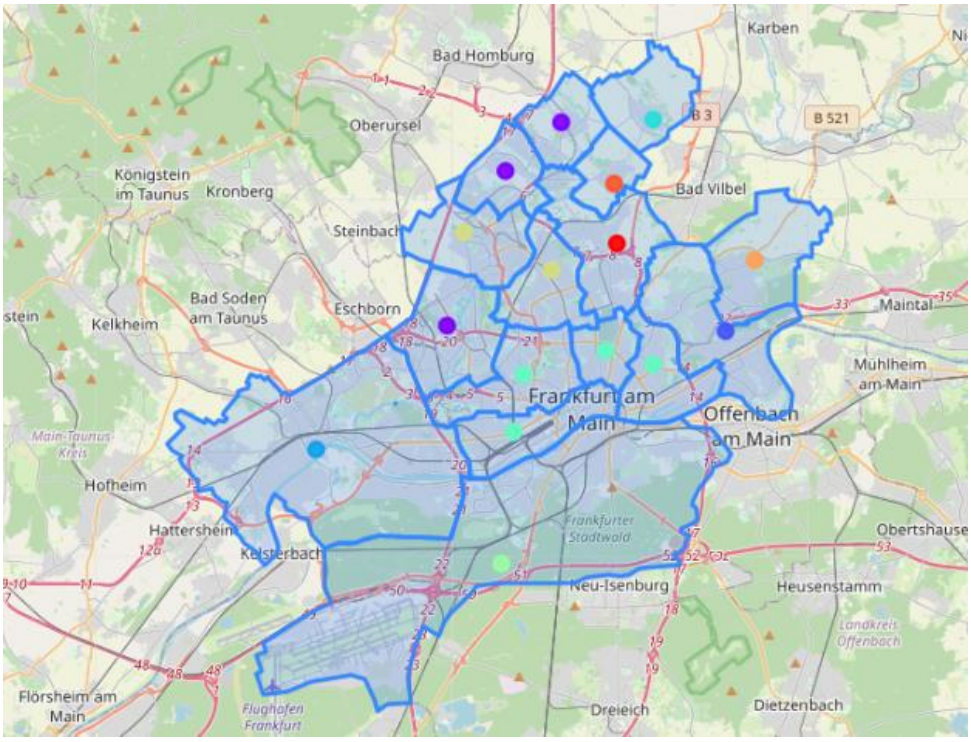
Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bergen-Enkheim	14	14	14	14	14	14
Bornheim/Ostend	47	47	47	47	47	47
Harheim	12	12	12	12	12	12
Innenstadt I	45	45	45	45	45	45
Innenstadt II	45	45	45	45	45	45
Innenstadt III	47	47	47	47	47	47
Kalbach/Riedberg	10	10	10	10	10	10
Mitte-Nord	32	32	32	32	32	32
Mitte-West	35	35	35	35	35	35
Nieder-Erlenbach	6	6	6	6	6	6
Nieder-Eschbach	6	6	6	6	6	6
Nord-Ost	20	20	20	20	20	20
Nord-West	16	16	16	16	16	16
Ost	36	36	36	36	36	36
Süd	21	21	21	21	21	21
West	10	10	10	10	10	10

Table 2. Amount of sport venues per borough.

As it can be seen from Table 2, obviously the highest density of sport facilities is in three main boroughs Innenstadt I-III and Borheim/Ostend. In these areas situated a lot of business centres and touristic places. The second highest density per boroughs belongs to Mitte-Nord, Mitte – West, Ost with 32,35 and 36 sport venues, respectively. The biggest Frankfurt's by population and area borough Süd gas relatively low density of sport venues, only 21.

Top 3 subcategories in almost each borough are: gym/fitness centre, gym, yoga studio. Distribution of clusters are shown in Picture 3.

From this analysis it is obvious that highest density of sport venues (like fitness/yoga studios, gyms) is in the city and business centres. Venues with not so common specialisations located out of these centres.



Picture 3. Cluster distribution in Frankfurt am Main city.

5. Conclusion.

In this project density of indoor sport venues were analysed. Areas with the highest density and top 3 popular categories were determined. Further decisions where and what kind of sport venue to open lays on businesspeople who are interested.

6. Reference.

1. https://de.wikipedia.org/wiki/Liste_der_Ortsbezirke_von_Frankfurt_am_Main
2. https://developer.foursquare.com/
3. https://developer.foursquare.com/docs/build-with-foursquare/categories/
4. https://offenedaten.frankfurt.de/dataset/85b38876-729c-4a78-910c-a52d5c6df8d2/resource/21d455e1-217d-47c5-af3d-ecdef1c50586/download/ffmortsbezirke.geojson
5. https://www.coursera.org/learn/python-project-for-data-science
6. https://de.wikipedia.org/wiki/Liste_der_Ortsbezirke_von_Frankfurt_am_Main#/media/Datei:Ortsbezirke_von_Frankfurt_am_Main.png