

Finding the ideal District in Frankfurt am Main to open a new franchise Café

Andreas Baron

2. November 2020

1. Introduction

Background and Problem

I want to consider the following situation: My friend approached me, since he wants to expand his cafe franchise to Frankfurt am Main/ Germany. Frankfurt is a city with many different kinds of cuisine: local and international and high frequency of business travelers and tourists, due to the closeness of the main airport and industry sector. Since he is not familiar with the city himself, he asked me to help him finding the perfect location. Therefore, I will use clustering algorithms, geodata and population data to find the ideal district to open his new cafe.

An ideal district must fulfill specific requirements:

- It has a high population of a target age group (18 to 64)
- It is accessible via public transportation
- It is central
- It is an area where hotels and restaurants are rather common
- It is an area where cafes are not that common

2. Data acquisition and cleaning

Data Sources

The following Datasets have been used:

a) Street Directory:

<https://offenedaten.frankfurt.de/dataset/strassenverzeichnis-der-stadt-frankfurt-am-main>

This dataset contains 4530 rows and 15 columns, with a lot of unnecessary information. I am only interested in the 46 unique districts of Frankfurt am Main.

b) Demographic Information:

<https://offenedaten.frankfurt.de/dataset/stadtteilprofile-bevoelkerung>

This dataset contains the district-wise distribution of population (total, age groups, foreign, etc.) It contains 46 rows and 146 columns. I am only interested in the total population and population of a certain age for the districts.

c) GeoData:

<https://offenedaten.frankfurt.de/dataset/wahlatlas-2015-geodaten/resource/84dff094-ab75-431f-8c64-39606672f1da>

This JSON file contains all the outlines for every district in Frankfurt am Main. In contrast to the demographical data it only contains 45 districts, since Gutleutviertel and Bahnhofsviertel are joined into one.

d) Foursquare API to find common venues in certain areas

Data Cleaning and Preparation

From the street directory of Frankfurt I have extracted a dataframe with all the unique districts and areas. Combining this table with the demographic information I gained a table with the population of specific age groups in all the districts of the city. To use this table with the GeoData I had to combine two districts (Gutleutvierte and Bahnhofsviertel) into one. I did not have to change the umlauts, since the newer version of folium is able to work with those. In the next step I plotted the population data with the GeoData and gained an overview over all districts and the total population (see Fig. 1).

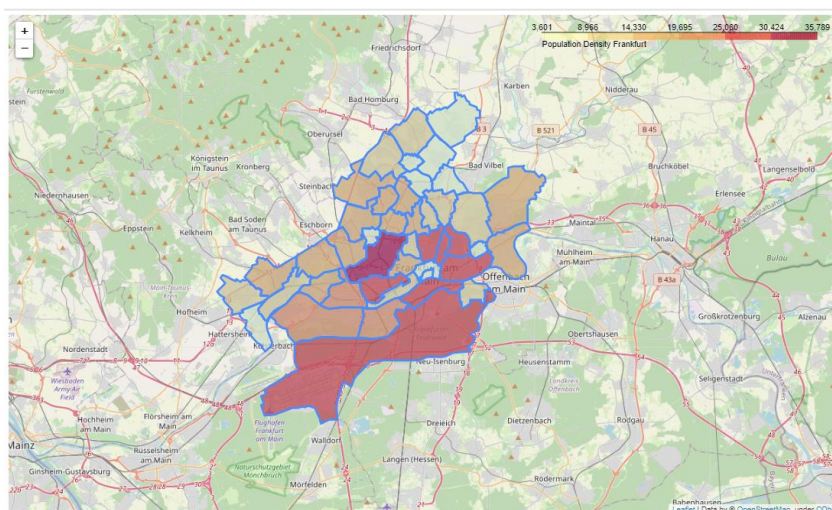


Figure 1: Districts of Frankfurt am Main with total population

To be able to easily gain access to location data in the map, I also included tooltips to the districts. The tooltips tell me the location and population I am currently hovering over. To gain the location data for the districts I used the Nominatim geolocator library, since the GeoData JSON only included the outlines of the districts.

Next I used the Foursquare API to access all available venues for the districts with location data, venue name and category. These were grouped by frequency and then sorted for the ten most common venues (see Fig. 2).

	District	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	Altstadt	Café	Plaza	Restaurant	Art Museum	Burger Joint	German Restaurant	Coffee Shop	Scenic Lookout	Italian Restaurant	Bar
1	Bergen-Enkheim	Plaza	Fountain	Food Court	Food & Drink Shop	Food	Farmers Market	Falafel Restaurant	Exhibit	Event Space	Ethiopian Restaurant
2	Berkersheim	Supermarket	German Restaurant	Trail	Light Rail Station	Zoo	Ethiopian Restaurant	Food Court	Food & Drink Shop	Food	Farmers Market
3	Bockenheim	Café	Italian Restaurant	Asian Restaurant	Supermarket	Japanese Restaurant	Wine Bar	Ice Cream Shop	Spanish Restaurant	Turkish Restaurant	Drugstore
4	Bonames	Metro Station	Café	Italian Restaurant	German Restaurant	Doner Restaurant	Bakery	Ice Cream Shop	Discount Store	Falafel Restaurant	Food Truck

Figure 2: Top 10 most common venues in all districts

To prepare for the Clustering Algorithm the Foursquare data had to be converted “one hot”, so it would only contain 1s and 0s.

3. Clustering using K-Means

The above described data was used to cluster the districts based on the K-Means algorithm. For easier selection of a final district I used a k of 5 and a random state of 0. The result of the clustering was inserted into a table with location data, cluster label and top ten most common venues for easier illustration and visualization.

```
# Cluster the districts into 5 clusters
kcluster = 5

frankfurt_grouped_clustering = frankfurt_grouped.drop('District', 1)

#run k means clustering
kmeans = KMeans(n_clusters = kcluster, random_state=0).fit(frankfurt_grouped_clustering)
```

Figure 3: Code for kmeans clustering of the data

Next, I plotted the clustered districts using a folium map. Each dot represents a district and the color represents the cluster (see Fig. 4). As can be seen immediately: the red cluster (cluster 0) is by far the biggest and most central of the clusters.

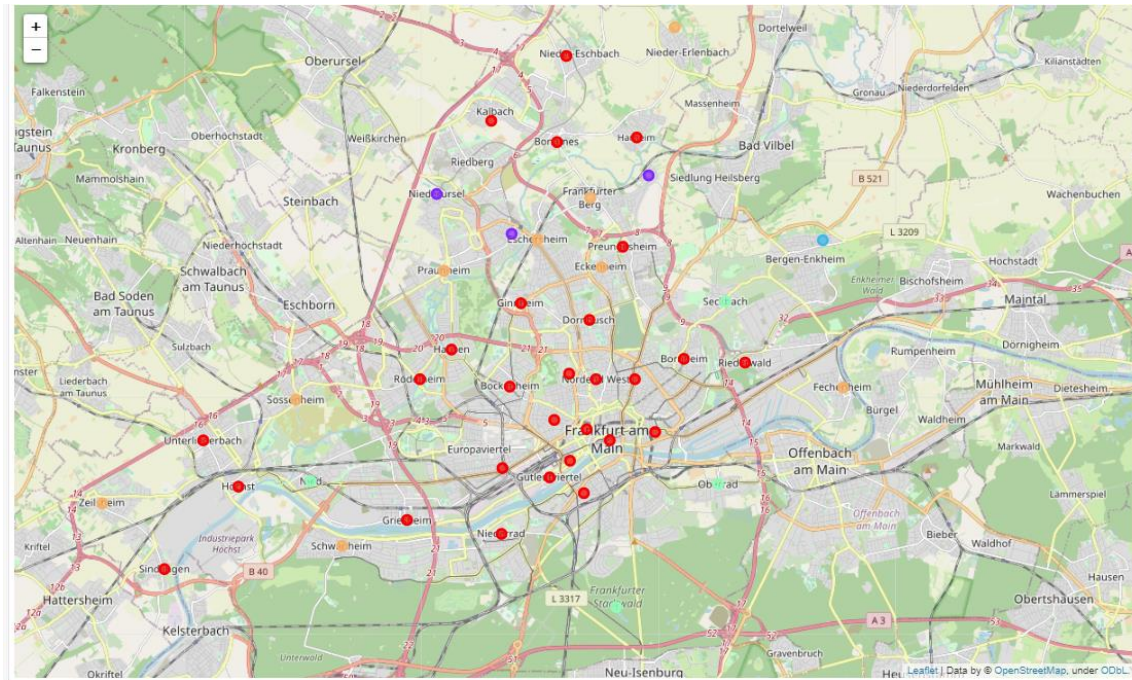


Figure 4: Clustered districts of Frankfurt am Main

In the next step I created tables for each cluster, containing district name and the top then most common venues.

4. Data and Visualization Analysis

In this chapter I interpret the current results and plots. When taking a look at the tables containing the clustered data I observed, that clusters 0 and 4 are the biggest cluster by far. The rest of the clusters seem to be resident areas without many venues that belong to the three groups I am looking for: Hotels, Public Transportation, Restaurants of any kind. The reasoning behind this is as such:

- Hotels: Many hotels means a lot of tourists and business travelers, therefore more customers
- Public Transportation: closeness and availability of public transportation means more fluctuation of tourists or non-residents in general. Generally, customers will have an easier time reaching the café.
- Restaurants of any kind: Having a lot of restaurants in the area means that people generally visit the district to eat, which means there is a higher potential customer base.

Therefore, I exclude all other clusters besides 0 and 4. Furthermore, cluster 4 seems to be less central and most common venues seem to be locations accommodating households in their daily needs (bakeries, supermarkets, etc.). In contrast, cluster 4 is rather central and most common

venues are overwhelmingly restaurants of some kind. Therefore, I will focus on cluster 0 for further analysis. As can be observed in Fig. 5, there are certain areas with a higher total population and therefore more suited for the opening of the café. Before I will draw any conclusion from this observation, I also plotted the same map again, this time with only the population aged 18 to 64 (the target age group) (see Fig. 6). Non surprisingly it results in a rather similar overview. Still, there are some differences and therefor conclusions that can be drawn from this second map. When considering the results there are six districts, that are especially suited to our business endeavour:

- Bockenheim
- Gallus
- Sachsenhausen-Nord
- Ostend
- Bornheim
- Nordend-Ost/ -West

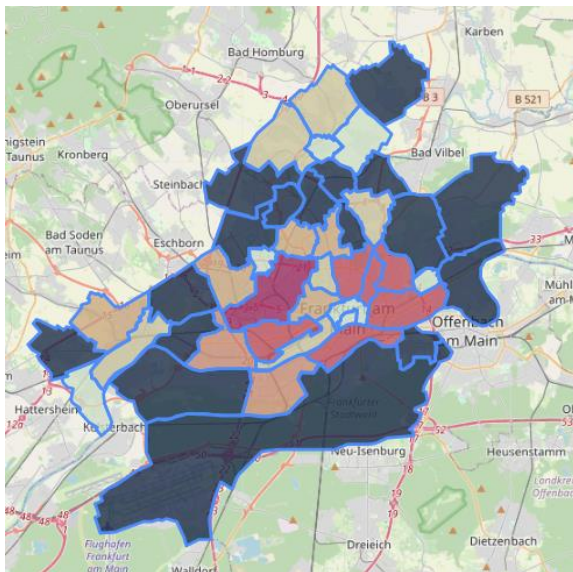


Figure 5: Cluster 0 total population graph

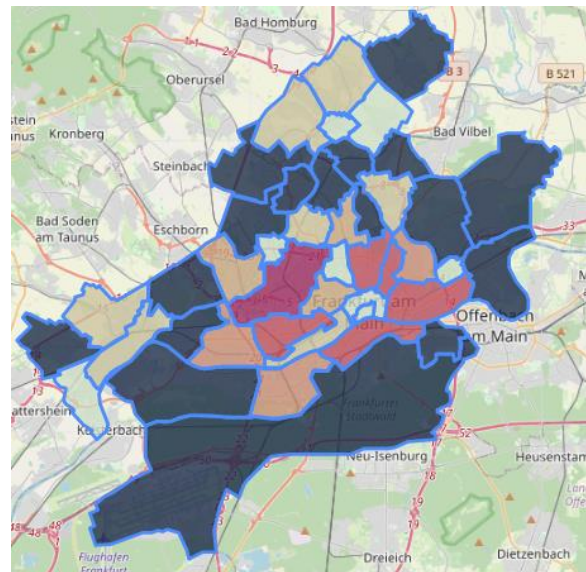


Figure 6: Cluster 0 population aged 18 to 64

In the next step I will use all acquired data to narrow down and select only one district for the café franchise.

5. Results and Discussion

Two districts can be eliminated rather easily: Nordend-Ost and Nordend-West are geographically less central and rather hard to access by public transportation in comparison to the other districts. Also Bornheim is less populated by the age group of 18 to 64, so it can be

excluded as well. Bockenheim has the highest population (total and in target age group), but the most common venues are already cafes. Therefore, I will eliminate Bockenheim as well. For the last three districts I will take a closer look at the three main factors I decided upon in the 4th chapter. To have an objective way of valuing the factors I will use a 10-digit point system.

- a) Closeness to Public Transportation: 1 far and few possibilities, 10 close and many possibilities
- b) Commonness of Cafes: 1 very common, 10 uncommon
- c) Commonness of Hotels: 1 uncommon, 10 very common

Each of the districts will be given points for each of the three factors. The highest overall total will result in the optimal location for the café. I will also give the reasoning behind given points in a specific category.

	District	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	Altstadt	Café	Plaza	Restaurant	Art Museum	Burger Joint	German Restaurant	Coffee Shop	Scenic Lookout	Italian Restaurant	Bar
1	Innenstadt	Coffee Shop	Hotel	Plaza	Boutique	Café	Steakhouse	Gym / Fitness Center	Electronics Store	Japanese Restaurant	Restaurant
2	Gutleut-/Bahnhofsviertel	Hotel	Bar	Indian Restaurant	Restaurant	Asian Restaurant	Seafood Restaurant	Coffee Shop	Vietnamese Restaurant	Park	Modern European Restaurant
3	Westend-Süd	Italian Restaurant	Steakhouse	Bakery	Gym / Fitness Center	French Restaurant	Chinese Restaurant	Restaurant	Ramen Restaurant	Supermarket	Bistro
4	Westend-Nord	Café	Coffee Shop	French Restaurant	Thai Restaurant	Cafeteria	Spanish Restaurant	Garden	Bistro	Ethiopian Restaurant	Supermarket
5	Nordend-West	Italian Restaurant	Café	Thai Restaurant	Theme Restaurant	Brewery	Spanish Restaurant	Road	Falafel Restaurant	Lounge	Beer Garden
6	Nordend-Ost	Café	Italian Restaurant	Bakery	Pub	Pizza Place	Bar	Ice Cream Shop	Plaza	Mexican Restaurant	Middle Eastern Restaurant

Figure 7: Cluster 0 with top 10 most common venues

1. Gallus: 19 points total

- a) 9 points: the main station of Frankfurt is in walking distance with many possibilities like tram, metro, busses, etc.
- b) 10 points: cafes are not in the top ten most common venues list.
- c) 0 points: hotels are not in the top ten most common venues list.

2. Sachsenhausen-Nord: 16 points total

- a) 9 points: many tram stations and access to the eastern train station of Frankfurt
- b) 6 points: 7th most common venues are cafes
- c) 1 point: 10th most common venues are hotels

3. Ostend: 15 points total

- a) 4 points: close to eastern train station of Frankfurt, but not that many other possibilities.
- b) 2 points: 3rd most common venues are cafes.
- c) 9 points: 2nd most common venues are hotels.

Since the district Gallus received the highest overall points, I would recommend building the café in this district. It also fulfills all other requirements I set up in the beginning of the project.

6. Conclusion

The districts of Frankfurt am Main were clustered and displayed on a map with color coding to analyze population and similarity between the districts. The demographics were studied and based on the findings three districts were found to be suitable districts for the new café. In general, the customer could choose from any of those three. I also evaluated the three districts on a 10-digit scale in three main categories to select one top candidate. The district Gallus achieved the overall highest score and would be my top candidate in all regards.