Capstone Project

The Battle of Neighborhoods in Frankfurt am Main, Germany

20 November 2020

1. Introduction

Frankfurt am Main with 763,000 inhabitants is the most important financial centre in Germany. It is also an attractive place to live for young professionals, therefore an interesting place to open a new restaurant. However not all areas have the same potential: boroughs with multi-storey apartment buildings, an older population or families with young children might be less attractive for an investor. On the other side, a high percentage of single-person households and many venues for entertainment and nightlife will be more promising.

Based on these criteria, we will characterize the different boroughs (Stadtteile) of Frankfurt. We will confront the result with the number of restaurants in the different areas: too much competition will not be helpful. We might circumvent the problem, concentrating on a specific category of restaurants, e.g. French restaurants.

2. Data

In order to describe the different boroughs of Frankfurt, we first used the following data:

- The rental price per square meter (in €)
- The population density (population per hectare)
- The average age of the population
- The percentage of the population between 18 and 64 years of age
- The percentage of single-person households

We obtained the rental price per square meter from the site de.statista.com [https://de.statista.com/statistik/daten/studie/262505/umfrage/mietpreise-in-frankfurt-am-main-nach-bezirken/#professional].

The population data came from the official portal of the city of Frankfurt:

- the population density [https://offenedaten.frankfurt.de/dataset/a0feb40c-b5f5-4ba2-a1fc-217229f65a96/resource/8153b993-ee1b-462a-abd8-ed19bc94dcb0/download/bauenwohnen.json],
- the other data [https://offenedaten.frankfurt.de/dataset/3be1af84-12d5-4d91-979a-3a468c77ed4e/resource/d4fc2f98-43cd-4a6c-8511-02ee1d1165a2/download/bevoelkerung.json],
- and the geospatial data [https://offenedaten.frankfurt.de/dataset/85b38876-729c-4a78-910c-a52d5c6df8d2/resource/84dff094-ab75-431f-8c64-39606672f1da/download/ffmstadtteilewahlen.geojson].

The geospatial data were used with geocoding web services of geopy to find the coordinates of each borough.

Data cleaning

The lists of boroughs are not totally identical in the different sources. We had to split or join some rows, remove leading spaces in the names, convert data with decimal commas to numbers with decimal points. A few missing values were replaced by mean values.

Thus, we constructed a dataframe (see figure 1), for 44 boroughs:

	Borough	Population Density	Average Age	Percentage 18-64	Percentage Single Households	Rental Price	B Latitude	B Longitude
0	Altstadt	48.70000	43.4	73.0	66.70	14.75	50.110644	8.682092
1	Bergen-Enkheim	14.00000	44.3	63.6	43.80	11.45	50.139567	8.747393
2	Berkersheim	11.50000	38.9	60.8	36.40	12.20	50.176219	8.697437
3	Bockenheim	66.10000	38.9	75.0	60.40	15.95	50.120524	8.653046
4	Bonames	40.50000	43.1	63.4	44.70	12.20	50.181347	8.663331
5	Bornheim	120.70000	43.2	69.0	62.00	15.05	50.115651	8.701897
6	Dornbusch	84.50000	44.0	63.5	55.10	13.60	50.135764	8.672073
7	Eckenheim	42.21750	41.9	65.9	51.20	12.20	50.145077	8.689725
8	Eschersheim	43.10000	42.4	65.5	53.80	13.60	50.158438	8.655319
9	Fechenheim	22.20000	39.7	64.8	47.80	12.00	50.125715	8.750796

Figure 1 - Data of Frankfurt Boroughs (top 10)

Then, we used the Foursquare API to get the top 100 venues within a radius of 750 m of the center of the borough. We called these areas neighborhoods, as they are not identical to the boroughs (see explanations in the methodology section).

1539 venues were returned; there were 208 unique categories. Figure 2 shows the 5 first entries of this list.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Altstadt	50.110644	8.682092	SCHIRN Kunsthalle	50.110291	8.683542	Art Museum
1	Altstadt	50.110644	8.682092	Römerberg	50.110489	8.682131	Plaza
2	Altstadt	50.110644	8.682092	Weinterasse Rollanderhof	50.112473	8.682164	Wine Bar
3	Altstadt	50.110644	8.682092	Hoppenworth & Ploch	50.110891	8.683701	Café
4	Altstadt	50.110644	8.682092	Kleinmarkthalle	50.112778	8.682958	Market

Figure 2 - Frankfurt Venues (top 5)

3. Methodology

We then had two datasets to work on: the boroughs of Frankfurt with characteristics, and the venues in Frankfurt.

We have to mention two issues:

1. The default Foursquare API limit value of 100 implies that not all venues in a neighborhood are obtained. 36 neighborhoods returned less than 100 venues; 8 reached the limit.
2. The Foursquare API counts venues within a given radius of the borough center, but the area of this circle is not identical to the geographical area of the borough. If the chosen radius is too small, large parts of the boroughs are not covered, if it is too large, there is a lot of overlapping. We tried different radii and finally chose 750 m as a reasonable compromise (see figure 12).

First, we selected information about the venues, which would be relevant to characterise the boroughs. We counted the venues in the categories 'entertainment, 'nightlife', 'food' (bars, pubs) to get a measure of the intensity of the nightlife in a borough. We called it "fun index" and added it to our dataframe of the Frankfurt borrows. Finally the dataframe to be used for clustering the boroughs looked as follows:

	Borough	Population Density	Average Age	Percentage 18-64	Percentage Single Households	Rental Price	Fun Index
0	Altstadt	48.7	43.4	73.0	66.7	14.75	32.0
1	Bergen-Enkheim	14.0	44.3	63.6	43.8	11.45	3.0
2	Berkersheim	11.5	38.9	60.8	36.4	12.20	0.0
3	Bockenheim	66.1	38.9	75.0	60.4	15.95	28.0
4	Bonames	40.5	43.1	63.4	44.7	12.20	3.0

Figure 3 - Clustering Data per Borough (top 5)

Clustering the Data

We wanted to separate the boroughs into groups with similar features. We use KMeans, a machine learning tool, to achieve this. To interpret features with different magnitudes and distributions, we normalized our cluster data with StandardScaler(). We ran the model with different numbers of clusters. We plot the bar chart with the characteristics of the clusters. With 4 clusters, we obtained distinctive clusters with a meaningful interpretation.

Figure 4 shows the bar chart of four clusters, figure 5 the mean values:

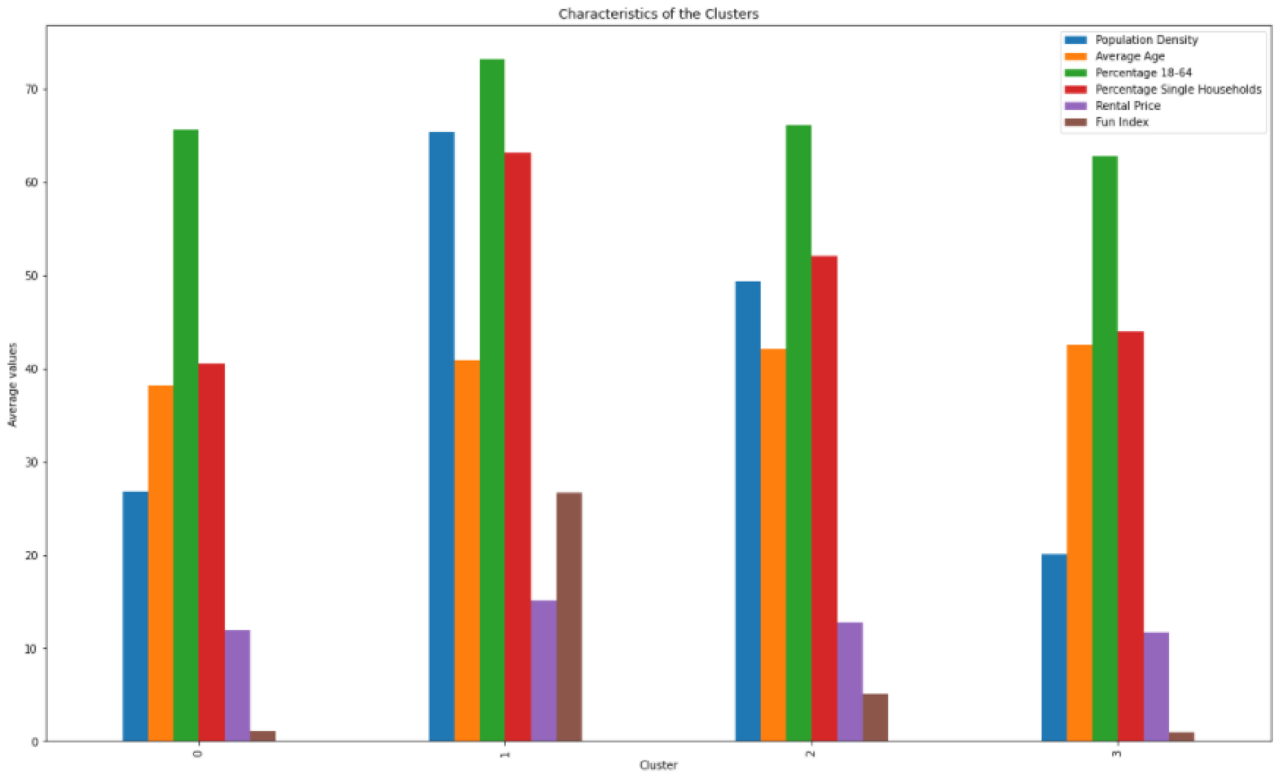

Figure 4 - Bar Chart of the Cluster Features

Labels	Population Density	Average Age	Percentage 18-64	Percentage Single Households	Rental Price	Fun Index
0	26.752917	38.200000	65.616667	40.566667	11.925000	1.166667
1	65.406354	40.891667	73.133333	63.195833	15.079167	26.666667
2	49.410385	42.023077	66.038462	52.030769	12.757692	5.153846
3	20.100000	42.492308	62.800000	43.976923	11.738462	1.000000

Figure 5 - Mean Values of the Cluster Features

In cluster 0 and 3 the features have comparatively low values. These boroughs are of less interest for our purpose. We call them “Low Profile Boroughs”.

In cluster 1, the features have high values. These are the “Nightlife Boroughs”.

Cluster 2 has similar values to cluster 1, but lower, especially a lower “fun index”.

Visualization

To visualize the results, we created a map showing the borough markers in different cluster colors.

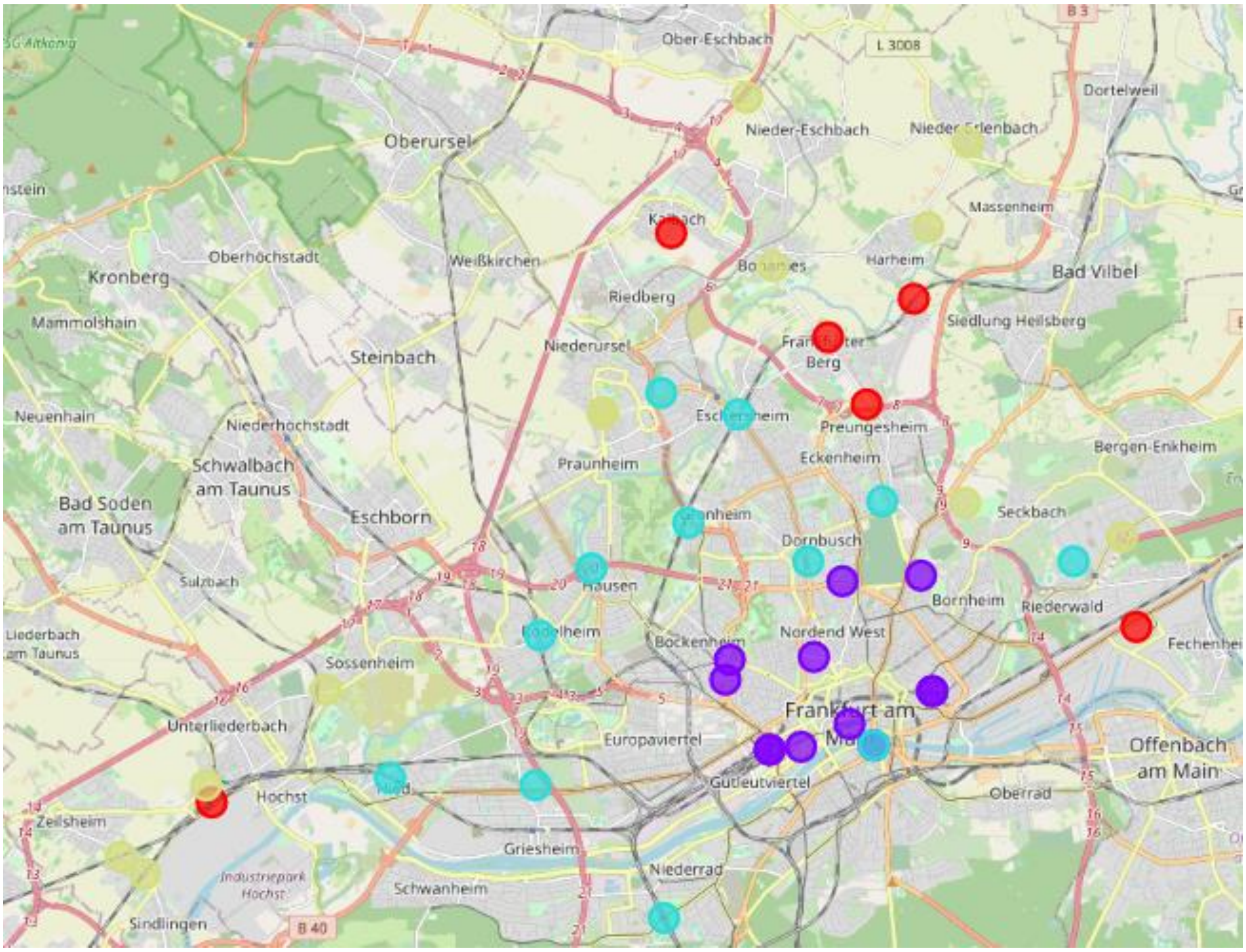

Figure 6 - Map of clustered boroughs

The purple circles belong to cluster 1, the turquois boroughs belong to cluster 2.

We selected all restaurants and built a list of the coordinates of the restaurants (see fig. 7). We superimposed a heatmap with the density of the restaurants to our clusters (fig. 8).

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Altstadt	50.110644	8.682092	Superkato	50.111664	8.679153	Sushi Restaurant
1	Altstadt	50.110644	8.682092	Góc Phố	50.113509	8.681686	Vietnamese Restaurant
2	Altstadt	50.110644	8.682092	Heimat – Essen und Weine	50.111125	8.678286	German Restaurant
3	Altstadt	50.110644	8.682092	Picknickbank	50.111534	8.678509	Moroccan Restaurant
4	Altstadt	50.110644	8.682092	Questione Di Gusto	50.112424	8.682045	Italian Restaurant

Figure 7 - List of restaurants (top 5)

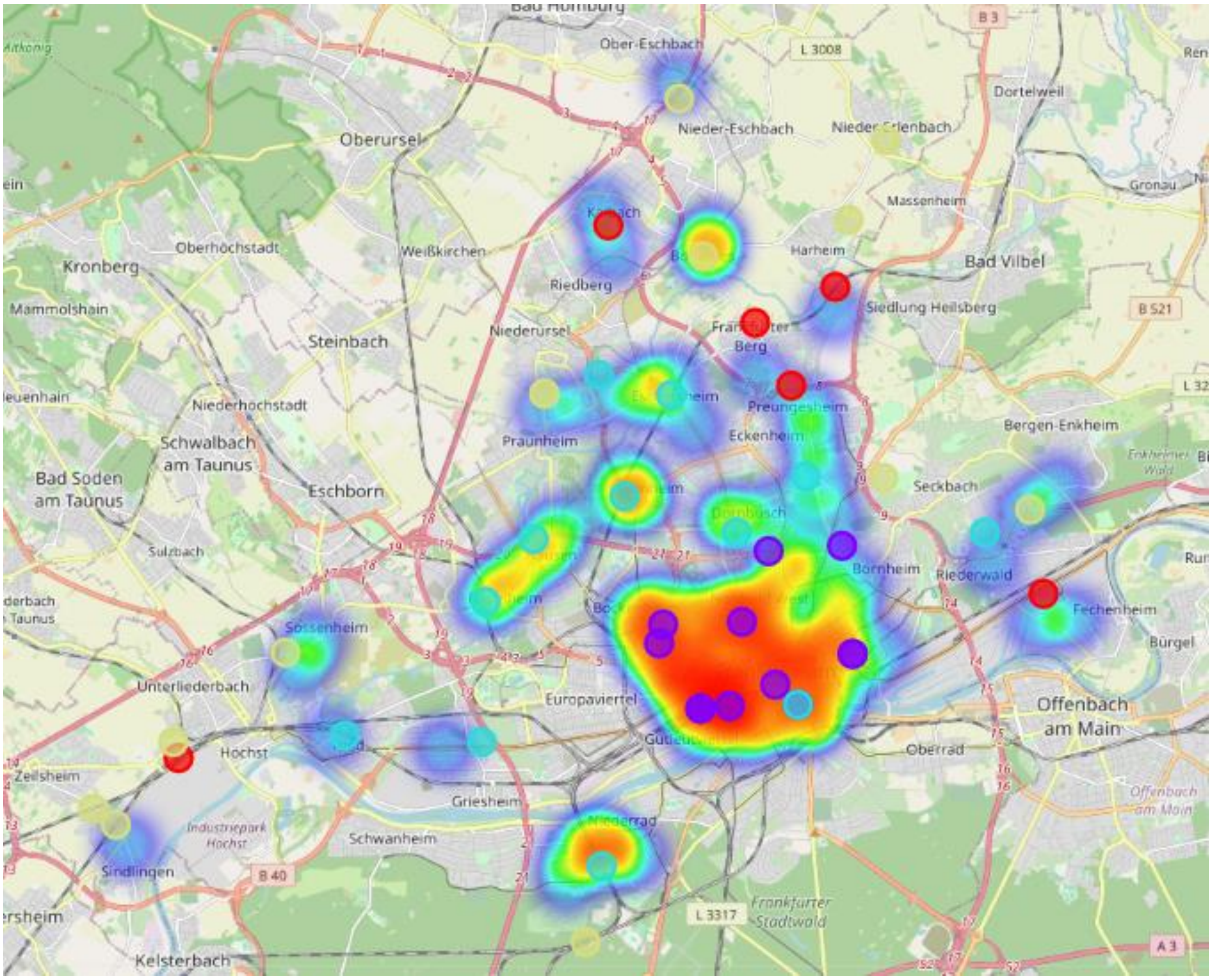

Figure 8 - Clusters with heatmap of restaurants

An interesting area with too many restaurants will yield too much competition. Selecting a category of restaurants of which only a few exist, might be the solution. We tried the list of French restaurants (see fig. 9) and added them to the map (fig. 10).

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Cat
0	Bockenheim	50.120524	8.653046	Lafleur	50.121445	8.655832	French Rest
1	Bockenheim	50.120524	8.653046	Brasserie ici	50.114454	8.651004	French Rest
2	Gutleut-/Bahnhofsviertel	50.107193	8.670254	Holbein's Café-Restaurant	50.102923	8.673752	French Rest
3	Sachsenhausen-Nord	50.107332	8.687672	Lobster	50.105224	8.687848	French Rest
4	Westend-Nord	50.120988	8.673486	Mon Amie Maxi	50.116503	8.667199	French Rest

Figure 9 - List of all French restaurants in Frankfurt

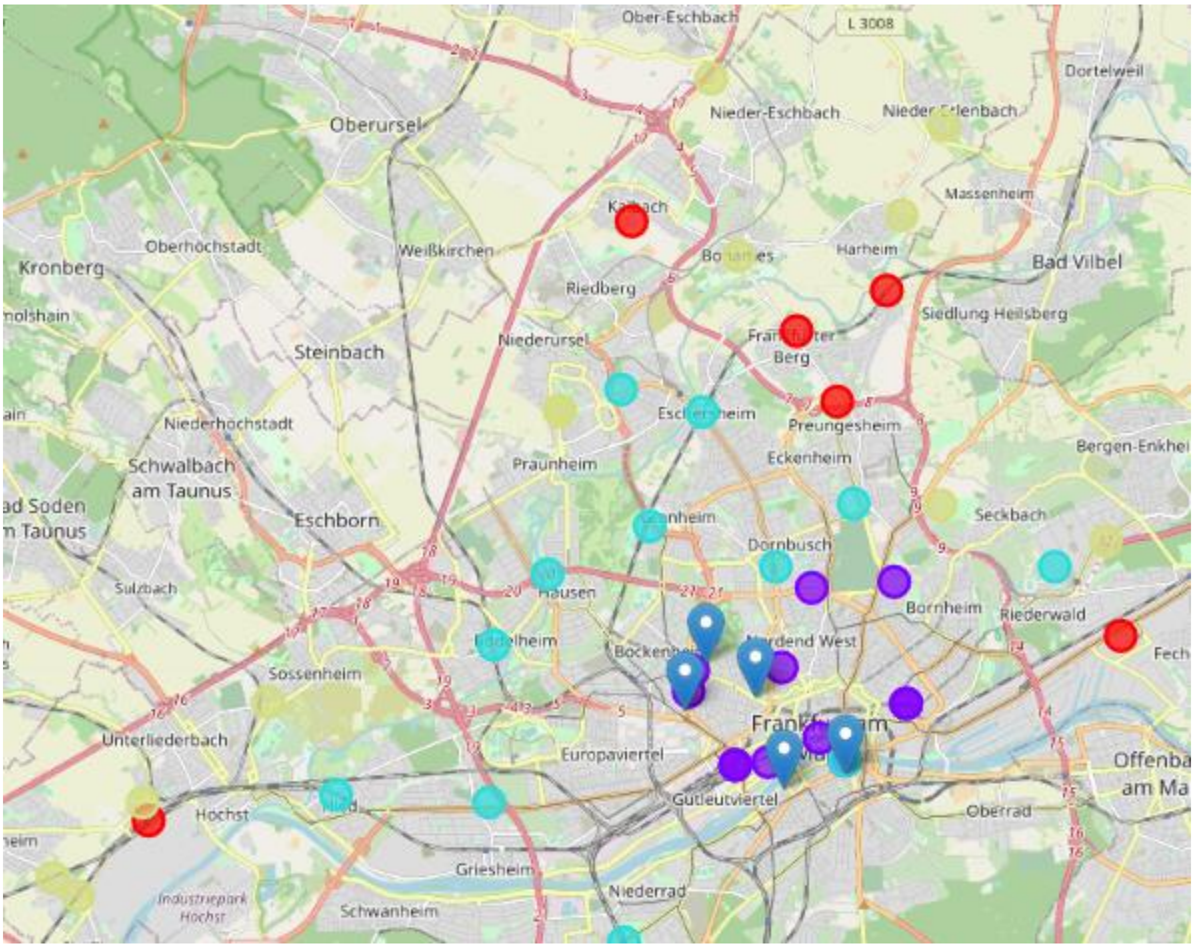

Figure 10 - Map with French restaurants

4. Results

Not surprisingly, the boroughs of cluster 1 are in the center of Frankfurt, whereas the boroughs of cluster 2 with similar features are slightly further away.

The boroughs of clusters 1 (see figure 11) seem the most interesting to open a restaurant, but the superimposed heatmap (fig. 8) reveals that these are also the boroughs with the greatest number of already existing restaurants. However, if we consider a specific category of restaurants, e.g. we look at French restaurants, there are only five in the whole city and they are all in the center (see fig 10). In this case, we should consider to open a French restaurant in a borough with no French restaurant, but a high “fun index”, like Ostend or Innenstadt.

Borough	Population Density	Average Age	Percentage 18-64	Percentage Single Households	Rental Price	B Latitude	B Longitude	Fun Index
Altstadt	48.70000	43.4	73.0	66.70	14.75	50.110644	8.682092	32.0
Bockenheim	66.10000	38.9	75.0	60.40	15.95	50.120524	8.653046	28.0
Bornheim	120.70000	43.2	69.0	62.00	15.05	50.115651	8.701897	34.0
Gallus	31.00000	38.6	73.7	60.00	14.75	50.106654	8.662581	24.0
Gutleut-/Bahnhofsviertel	50.75875	39.3	79.8	68.85	14.75	50.107193	8.670254	32.0
Innenstadt	42.21750	41.6	76.5	71.30	14.75	50.106654	8.662581	24.0
Nordend-Ost	150.60000	40.7	75.2	65.40	15.25	50.133655	8.699082	8.0
Nordend-West	92.00000	41.2	72.5	63.00	15.25	50.132620	8.680232	9.0
Ostend	48.50000	42.5	71.9	62.60	15.05	50.115651	8.701897	34.0
Sachsenhausen-Nord	9.40000	40.6	71.5	60.40	13.50	50.107332	8.687672	39.0
Westend-Nord	54.30000	40.0	69.0	55.30	15.95	50.120988	8.673486	31.0
Westend-Süd	70.60000	40.7	70.5	62.40	15.95	50.117517	8.652180	25.0

Figure 11 - Boroughs of Cluster 1

5. Discussion

We mentioned the issues raised by the discrepancy between boroughs and neighborhoods.

We envisaged to make the radius a variable depending on the area of the borough or to investigate a means to fetch the venues within a geographical area. However, we kept the simple solution, as the limit of 100 venues did not allow us to fetch all the venues anyway. Figure 12 illustrates the problem.

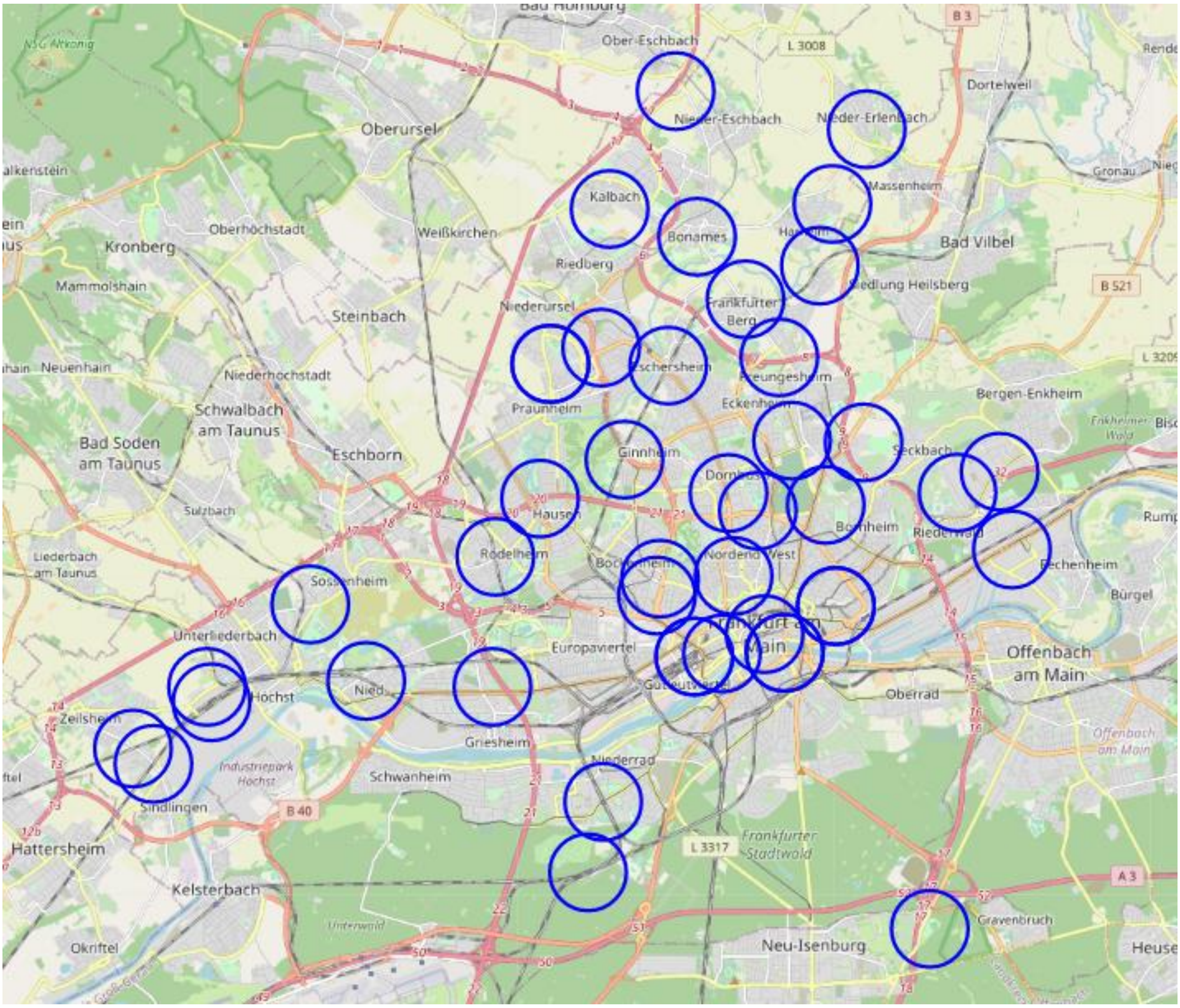

Figure 12 - Frankfurt neighborhoods

6. Conclusion

We investigated boroughs in Frankfurt by machine learning tools to find potential areas, where an investor could open a restaurant. We found that promising areas were also the ones where the most restaurants were. To circumvent that issue, we suggested to select a category of restaurants and choose one of the potential boroughs without that kind of restaurants.

in the discussion section, we also addressed the limitations of the approach due to the Foursquare API venue limit and the circle areas.