IBM Data Science Capstone Project
The Battle of Neighborhoods

Marc Schneider
marcschneider1993@gmail.com

03.02.2020

1 Introduction

The city of Frankfurt am Main is at the center of the Rhine-Main Metropolitan Region in Germany. It covers an area of 248.31 km² and has a population of 753,056 inhabitants. Frankfurt is known to be a global hub for commerce, culture and tourism. For example, the European Central Bank, the Frankfurt Stock Exchange, the Messe Frankfurt and the Frankfurt Airport are some of the busiest and most influential institutions in Europe. As these institutions attract people from all kinds of backgrounds, Frankfurt itself is culturally, ethically and religiously diverse, i.e. a big part of the population has a migration background.

As a city's demography gets more and more diverse, the amount of new venues offering foods from all over the world increases as well. According to an analysis by *Sommer in Hamburg*[1] the five most popular cuisines in Germany are Italian, German, Greek, Chinese and Turkish. It is not too far fetched that a multicultural city like Frankfurt offers an array of venues to try dishes from these cuisines. A useful tool for finding such venues in a city is *Foursquare*[2]. *Foursquare* is an app which lets users search, rate and recommend places to other users.

The goal of this project is to use the *Foursquare* API to explore the culinary landscape of Frankfurt and answer questions like:

- What is the most popular cuisine in Frankfurt?
- Are some cuisines over- or underrepresented?
- Are there correlations between the demography and the venues?

Who can benefit from this analysis? On the one hand, this report might reveal some good places for anyone who wants to try a new venue which serves one of the five mentioned cuisines. On the other hand, this report can give

[1] `https://sommer-in-hamburg.de/kultur/kochen-essen-kulinarik/beliebteste-laender-kuechen/`

[2] `https://foursquare.com/`

recommendations to future founders who want to open a new restaurant but who are uncertain what kind of dishes are popular and/or underrepresented.

2 Data

The data used for this analysis is separated into four parts. In the following sections, they will be introduced briefly.

2.1 GeoData

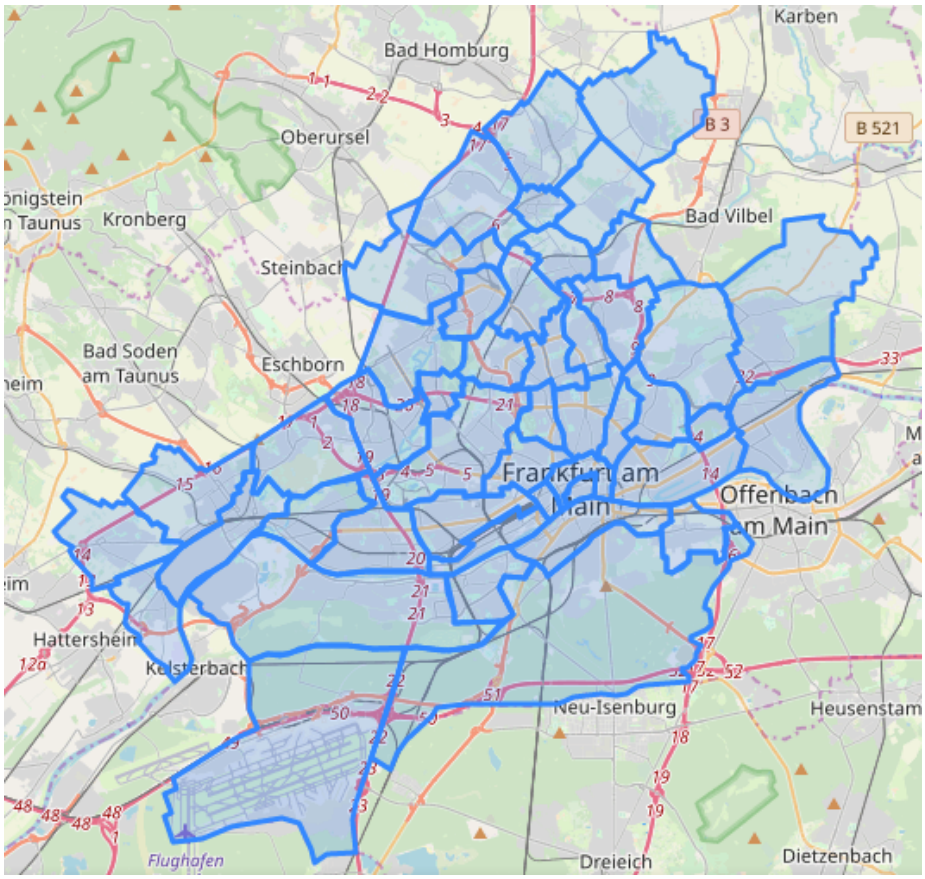

Figure 1: Map of districts

The first part of the data is a GeoJSON file which contains the borders of the 46 districts in Frankfurt. These borders can be seen in figure 1. They will be used

to assign the venues found by the *Foursquare* API to a district. The original file can be found on GitHub[3].

2.2 Census data

	District	Area in square km	Inhabitants	Female	Male	Germans	Foreigners	Foreigners in percent	Inhabitants per square km	Longitude	Latitude
0	Altstadt	0.506	4151	2015	2136	2604	1547	37.3	8204	8.681539	50.110590
1	Innenstadt	1.491	6605	3088	3517	3514	3091	46.8	4430	8.682205	50.112817
2	Bahnhofsviertel	0.542	3561	1378	2183	1638	1923	54.0	6570	8.667658	50.107810
3	Westend-Süd	2.497	18822	9593	9229	13655	5167	27.5	7538	8.661165	50.117950
4	Westend-Nord	1.632	10198	5321	4877	7214	2984	29.3	6249	8.666019	50.129776

Figure 2: Census DataFrame

The second part of the data consists of census data for each district in Frankfurt. The data dates back to 31.12.2018 and contains information like the number of inhabitants, the area or the amount of Germans and foreigners in each district. It does not contain the latitude and longitude of each district but both of these can be derived from the GeoJSON data referenced in section 2.1. After translating and cleaning, the data is stored in a *pandas* DataFrame (DF) as shown in figure 2. It has a shape of $(46, 11)$, one row for each district. The original data in German can be found here[4].

[3] `https://github.com/codeforamerica/click_that_hood/blob/master/public/data/frankfurt-main.geojson`

[4] `https://www.frankfurt.de/sixcms/media.php/738/01_Bev%C3%B6lkerung_Ende2018.pdf`

2.3 Overview of venues

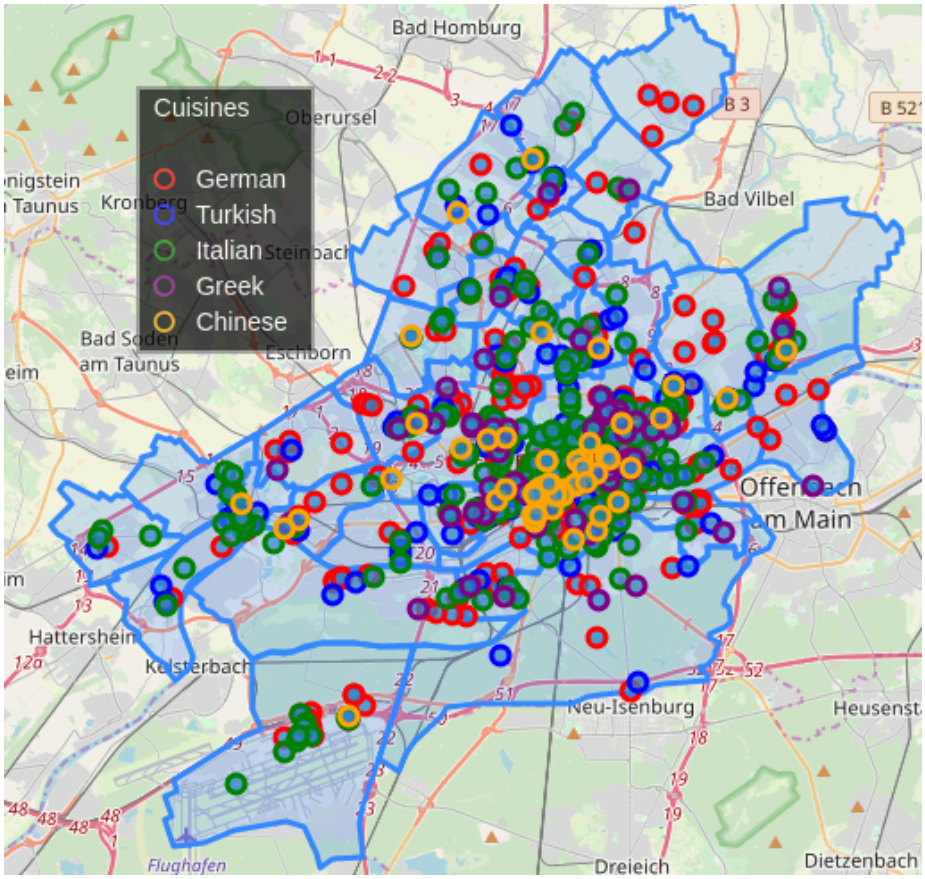

Figure 3: Venues on map

	Name	ID	Address	Lat	Lng	Cuisine	District
0	Zum Standesämtchen	4c3b6d08def90f470beefd2d	Römerberg 23	50.110296	8.682389	DE	Altstadt
1	Wirtshaus am Hühnermarkt	5c180459a4b51b002c093a70	Markt 16-18	50.110777	8.684000	DE	Altstadt
2	Frankfurter Wirtshaus	5b3225a7e65d0c002c94296b	Mainkai 35	50.109050	8.681975	DE	Altstadt
3	Max on One	4e9dc4dff5b95ad93faa8788	Thurn-und-Taxis-Platz 2	50.115110	8.680233	DE	Innenstadt
4	Heimat – Essen und Weine	4cfd2a882c1aa090410e057a	Berliner Str. 70	50.111125	8.678286	DE	Altstadt

Figure 4: Venues DataFrame

The third part of the data is generated by using the *Foursquare* API. The API is used to find all venues for each of the five most popular cuisines in Germany (Italian, German, Greek, Chinese and Turkish). Figure 3 shows where these venues are located on the map. The resulting DF has a shape of $(825, 7)$ and stores the $Name, ID, Address, Latitude, Longitude, Cuisine$ and $District$ of a venue. It is shown in figure 4.

2.4 Venue details

	ID	Likes	Listed	Photos	Price	Rating	Rating Signals	Tips
0	4c3b6d08def90f470beefd2d	25.0	5	86	2.0	6.1	60.0	19
1	5c180459a4b51b002c093a70	2.0	1	4	2.0	NaN	NaN	1
2	5b3225a7e65d0c002c94296b	7.0	1	10	2.0	5.5	21.0	6
3	4e9dc4dff5b95ad93faa8788	16.0	15	44	2.0	8.6	17.0	3
4	4cfd2a882c1aa090410e057a	43.0	136	49	2.0	8.7	59.0	16

Figure 5: Venue details DataFrame

The fourth and final part of the data is also generated with the *Foursquare* API. Each venue contained in the first DF of section 2.3 is queried again to get the venue's details. The API returns information about a venue including ratings, tips, and photos. The results are stored in a separate DF shown in figure 5. It has a shape of $(825, 8)$.

3 Methodology

This section contains a description of the exploratory analysis. It is split into three parts, each having a different emphasis. The first part focuses on districts, the second part on cuisines and the last part on individual venues.

3.1 Analysis of Districts

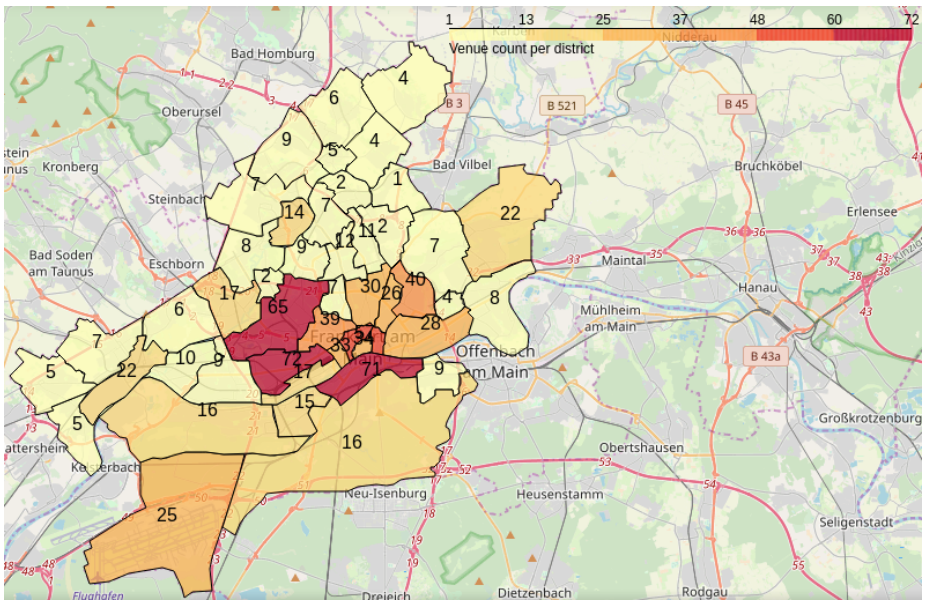

Figure 6: Choropleth map showing venue counts per district

In this part of the analysis the venue information is grouped by **District**. Figure 6 shows a choropleth map overlaying the venue counts per district. The main takeaway is that most of the venue data is concentrated in certain districts around the city center, whereas districts in the north and west have less data. E.g **Gallus** contains 72 venues, **Sachsenhausen-Nord** 71 and **Bockenheim** 65. These 3 district already contain 25% of all venues. In contrast, **Berkersheim** has 1 venue and **Hausen**, **Preungesheim** and **Frankfurter Berg** each contain only 2 venues.

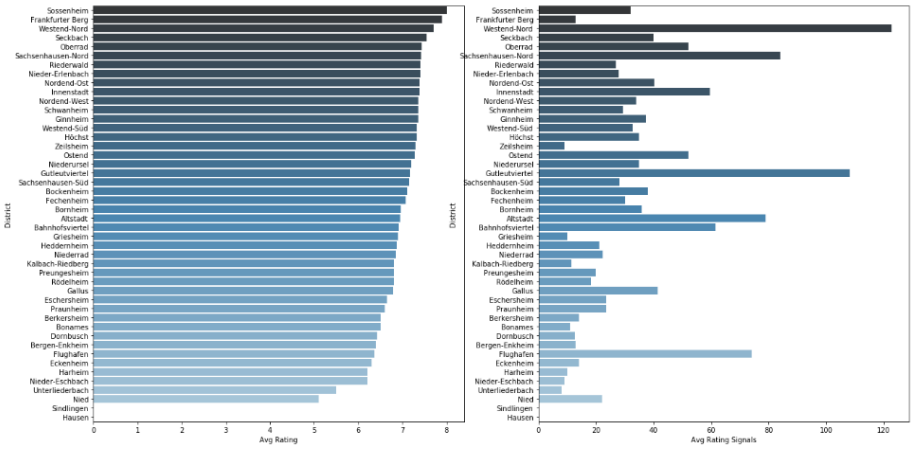

Figure 7: Left: Average venue rating per district,
Right: Average number of ratings per venue in each district

The next point of interest is which districts have better average ratings. Figure 7 shows the average venue rating per district next to the average number of ratings per venue in each district. It is important to look at both information simultaneously as average ratings get more trustworthy when more people rate a venue. It is evident that **Westend-Nord** has the third highest average rating as well as the highest amount of average number of ratings per venue.

3.2 Analysis of Cuisines

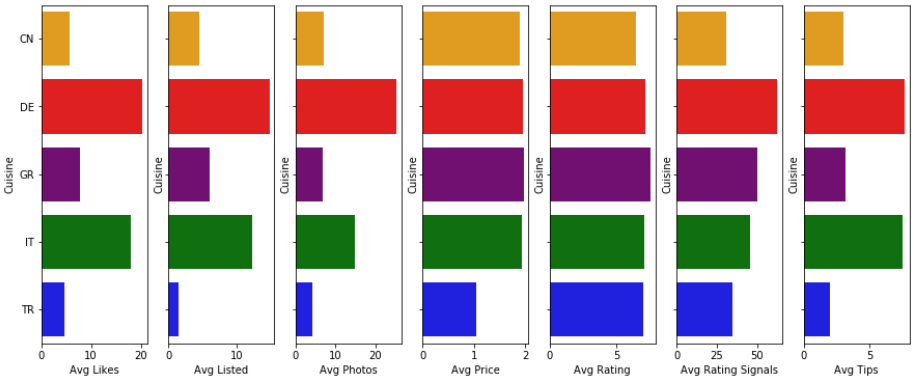

Figure 8: Average statistics for each cuisine

This part of the analysis groups all the venues by their cuisines. Figure 8 shows their average statistics. One can see that *Foursquare* users on average like, list, rate, recommend (Tips) and submit photos for German (**DE**) venues the most. In most categories, Italian venues (**IT**) on average also have many user interactions. Chinese (**CN**), Greek (**GR**) and Turkish (**TR**) venues have less interactions on average.

Another interesting point is that the average price tier for Turkish venues differs significantly from the other cuisines. *Foursquare*'s price tiers go from 1 (least pricey) to 4 (most pricey). Thus, Turkish venues offer cheaper food on average.

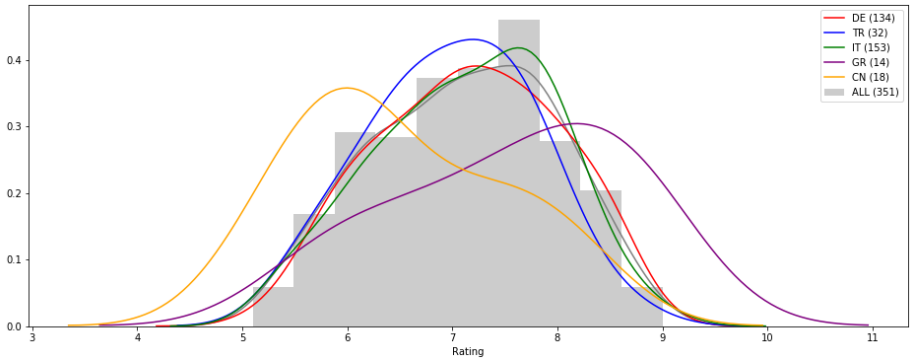

Figure 9: Distribution of ratings grouped by cuisines, the shaded area shows the histogram for ratings of all venues (number of venues in brackets)

The statistic for ratings contains another interesting property. On average, Greek venues are rated the best. Figure 9 shows this statistic in more detail. It depicts the distribution of ratings for the different cuisines. While ratings for German, Turkish and Italian venues are distributed similarly, Chinese venues have a peak shifted to the left and Greek venues have a peak shifted to the right.

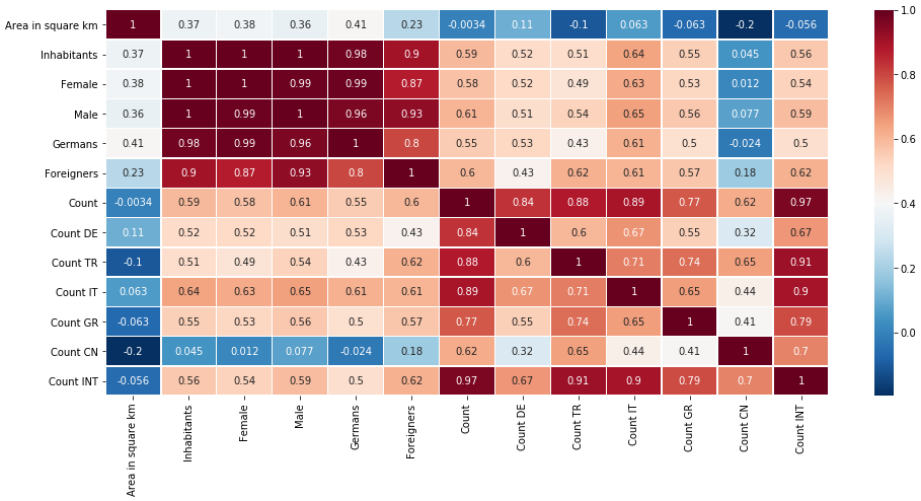

Figure 10: Correlation matrix

	Count	Count DE	Count TR	Count IT	Count GR	Count CN	Count INT
Germans	0.551441	0.525291	0.430631	0.614693	0.499301	-0.024193	0.503028
Foreigners	0.600077	0.425744	0.615729	0.606550	0.568396	0.183584	0.615326
Difference	-0.048636	0.099547	-0.185098	0.008143	-0.069095	-0.207777	-0.112298

Figure 11: Correlation coefficients for Germans/Foreigners and venue counts

Lastly, we want to see whether there are correlations between the census data of section 2.2 and the cuisines. It is especially interesting to find out whether districts with a higher percentage of foreigners have a higher count of non-German venues or vice versa. The heatmap of the correlation matrix in figure 10 shows the correlation values for the merged data. Especially the rows for **Germans** and **Foreigners** and the columns for the venue counts are of interest. In order to make them more visible, they are also shown separately in figure 11. There are slight differences for the column of German venues (0.099547) and international venues (-0.112298).

3.3 Analysis of Venues

	Name	Address	Lat	Lng	Cuisine	District	Likes	Listed	Photos	Price	Rating	Rating Signals	Tips
488	Trattoria i Siciliani	Walter-Kolb-Str. 17	50.104896	8.684428	IT	Sachsenhausen-Nord	168.0	179	73	NaN	9.0	296.0	149
695	Omonia	Eckenheimer Landstr. 126	50.127359	8.685447	GR	Nordend-West	40.0	92	16	2.0	8.9	62.0	24
477	da Cimino	Adalbertstr. 29	50.119496	8.645682	IT	Bockenheim	284.0	442	115	2.0	8.9	410.0	112
723	Parthenon Greek Restaurant	Stresemannallee	50.101037	8.674486	GR	Sachsenhausen-Nord	68.0	39	67	2.0	8.8	100.0	28
4	Heimat – Essen und Weine	Berliner Str. 70	50.111125	8.678286	DE	Altstadt	43.0	136	49	2.0	8.7	59.0	16
724	Ariston Restaurant	Heiligkreuzgasse 29	50.115498	8.689205	GR	Innenstadt	99.0	94	67	2.0	8.7	143.0	40
442	Super Bros	Oeder Weg 55-57	50.121681	8.680445	IT	Nordend-West	39.0	59	19	2.0	8.7	52.0	12
29	Manufactum brot&butter	Bockenheimer Anlage 49-50	50.115958	8.670772	DE	Westend-Süd	94.0	160	46	NaN	8.7	119.0	19
271	Emir'Et	Allerheiligenstr.21	50.113518	8.689303	TR	Innenstadt	14.0	14	17	1.0	8.6	23.0	6
3	Max on One	Thurn-und-Taxis-Platz 2	50.115110	8.680233	DE	Innenstadt	16.0	15	44	2.0	8.6	17.0	3

Figure 12: Top 10 Venues sorted by rating

	Name	Address	Lat	Lng	Cuisine	District	Likes	Listed	Photos	Price	Rating	Rating Signals	Tips
821	Der Goldene Wok	Mainzer Landstraße 793	50.099010	8.562728	CN	Nied	6.0	4	17	2.0	5.1	22.0	7
71	Zum Schwarzen Stern	Römerberg 6	50.110274	8.682466	DE	Altstadt	21.0	5	75	2.0	5.2	50.0	14
804	Ding Ding Sheng	Moselstr. 23	50.107416	8.667175	CN	Bahnhofsviertel	12.0	12	24	2.0	5.3	31.0	11
585	Italissimo	Terminal 1, Gate A22	50.046548	8.567226	IT	Flughafen	18.0	4	43	2.0	5.3	43.0	11
623	Ristorante Il Quadrifoglio	NaN	50.150284	8.622023	IT	Praunheim	3.0	0	3	2.0	5.4	8.0	0
670	Fantasia 5	Berger Str. 191	50.125787	8.707198	IT	Bornheim	12.0	2	15	2.0	5.4	25.0	6
781	China Restaurant New World	Am Hauptbahnhof 8	50.107547	8.665013	CN	Bahnhofsviertel	3.0	0	11	1.0	5.4	8.0	1
580	Vito	Frankfurt Airport	50.051985	8.587016	IT	Flughafen	8.0	1	27	2.0	5.4	22.0	3
138	Ludwigs	NaN	50.052068	8.586865	DE	Flughafen	14.0	0	35	2.0	5.5	31.0	8
677	Ristorante Pizzeria Limoncello	NaN	50.107399	8.543978	IT	Unterliederbach	2.0	1	4	2.0	5.5	8.0	4

Figure 13: Bottom 10 Venues sorted by rating

This part of the analysis focuses on the individual venues, utilizing *Foursquare*'s venue details as seen in figure 5. Figures 12 and 13 show the top 10 and bottom 10 venues sorted by ratings. One thing that is remarkable is the distribution of the cuisines, i.e. there are no Chinese venues in the top 10 and only 1 Italian place. In contrast, there are 5 Italian and 3 Chinese places in the bottom 10.

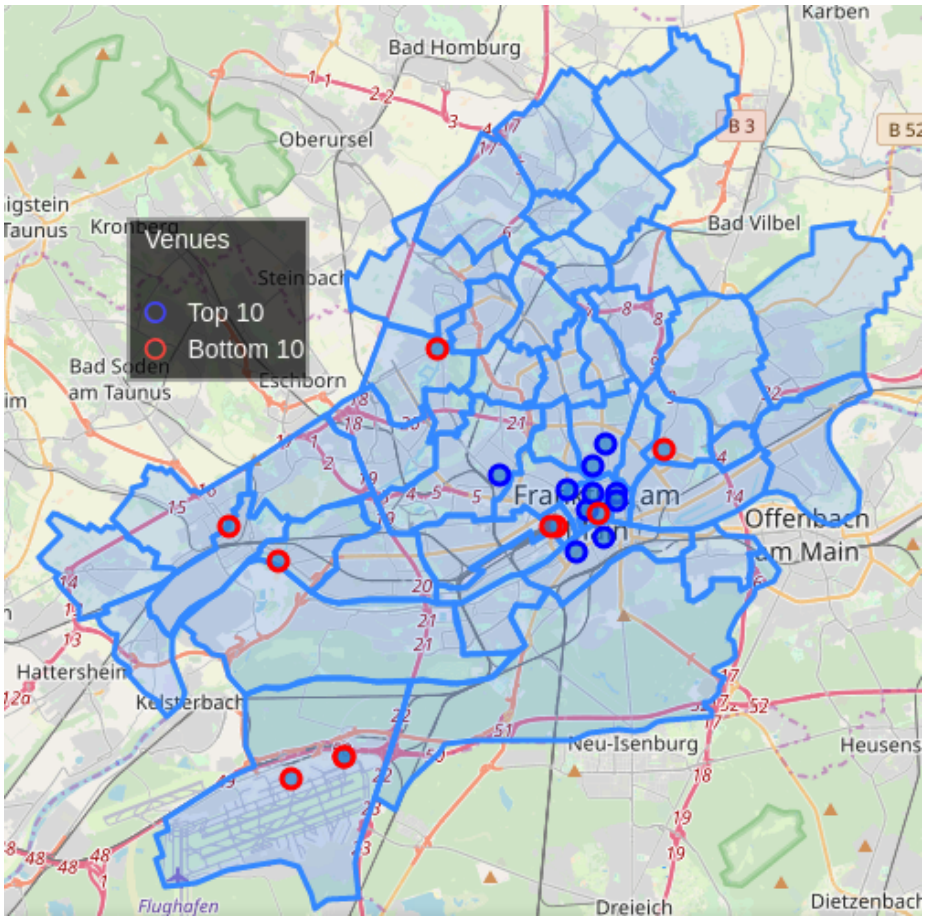

Figure 14: Top 10 and bottom 10 venues on map

Lastly, we want to take a look where these venues are located. Figure 14 shows the locations of the top 10 and bottom 10 venues. It is evident that the top 10 venues are clustered more around the city center, while the bottom 10 venues are more spread out. Notably, 2 of the bottom 10 venues are located at the airport in the bottom left.

4 Results

Based on the available data we can conclude that districts around the city center like **Innenstadt** or **Sachsenhausen-Nord** offer popular and highly rated venues to visit while venues at the airport (**Flughafen**) should be avoided. Regarding the cuisines of Frankfurt, **Greek** venues on average are rated the best

while **Chinese** venues are rated the lowest. Furthermore, **Turkish** venues are the cheapest and **German** and **Italian** venues offer the most variety to choose from.

5 Discussion

My advice for future founders is to open a **Chinese** restaurant in one of the center districts as the average quality of those restaurants is lower and the exposure to people using *Foursquare* is maximized in those area. Alternatively, the airport is another option as the competition there is worse compared to other locations.

As the financial aspect of founding a venue has not been considered in this report, the prices of real estates have to be analysed in a future report.

6 Conclusion

In this report we explored the culinary landscape of Frankfurt. The *Foursquare* API provided the venue data and contained further information about the preferences of its users. Still, *Foursquare* is not very popular in Germany and thus the available data is rather limited.

The analysis grouped the data into three parts, one part grouping the venues by districts, one part grouping the venues by cuisines and the last part looking at the highest and lowest rated venues individually.

Lastly, we made recommendations for visitors and for anyone planning to found a new venue in Frankfurt.