Capstone Project - The Battle of the Neighborhoods

1. Introduction

Stuttgart is the sixth largest city in Germany with a rapidly growing population. It has a huge variety of restaurants for every taste and, thus, to start a restaurant business in this area is not an easy task.

Our stakeholder is willing to open the beer restaurant in the city of Stuttgart with middle-high level prices.

Of course, choosing a location for business is one of the stressful and controversial tasks, since there are a lot of criteria that should be satisfied in order to achieve the highest revenue.

Here are some of them:

- the density of other restaurants

- the density of specifically beer rest

- population density around the locat

- solvency of the population around

- ...

In this project, we will implement the basic analysis and try to find the most optimal Borough to open the beer restaurant according to those criteria. It's obvious, that there are many additional factors, such as distance from parking places or distance from the main streets, but this analysis can be done after choosing the Borough, and thus will not be performed within the scope of this project.

2. Data description

Based on criteria listed above the following data will be utilized in our analysis:

- the number of restaurants within the certain radius of each borough (Foresquare API)
- the net income per person in each borough. Since the restaurant will have middle-high prices, it is important to consider the solvency of population. Source: Socialmonitoring der Landeshauptstadt Stuttgart (https://statistik.stuttgart.de/statistiken/sozialmonitoring/atlas/Stadtbezirke/out/atlas.html)
- the population and the population density of the borough. Source: Statistikatlas Stuttgart (https://statistik.stuttgart.de/statistiken/statistikatlas/atlas/atlas.html?indikator=i0&select=00)
- the population above 18 years age. It is obvious, that potential visitors of our beer restaurant are men and women of full age. Source: Statistikatlas Stuttgart (https://statistik.stuttgart.de/statistiken/statistikatlas/atlas/atlas.html?indikator=i0&select=00)
- the coordinates of the borough. Source: Open street map (https://nominatim.openstreetmap.org/details.php?place_id=17476218)

After cleaning and preparing the data, we defined the following master dataframe:

	Borough	Latitude	Longitude	Population Density	Population	Net income per person	Number of restaurants	Number of beer restaurants
0	Bad Cannstatt	48.807109	9.221557	3798	59685	92.2	24.0	10.0
1	Birkach	48.728574	9.203406	1953	6030	111.9	1.0	0.0
2	Botnang	48.778495	9.129532	5162	11021	103.7	1.0	0.0
3	Degerloch	48.744052	9.180481	1755	14080	114.0	5.0	2.0
4	Feuerbach	48.803635	9.149803	2206	25497	104.1	8.0	3.0
5	Hedelfingen	48.755863	9.219398	1167	8546	96.5	2.0	0.0
6	Möhringen	48.726912	9.151997	1866	28072	103.9	7.0	2.0
7	Mühlhausen	48.845610	9.226179	2361	21528	86.7	1.0	0.0
8	Münster	48.826620	9.202133	2611	5771	88.6	2.0	1.0
9	Obertürkheim	48.772290	9.280816	1324	7228	97.3	2.0	1.0
10	Plieningen	48.711395	9.198991	873	11416	102.5	1.0	1.0
11	Sillenbuch	48.742184	9.222216	2704	20164	109.5	3.0	0.0
12	Stammheim	48.850726	9.154724	2431	10532	97.0	0.0	0.0
13	Stuttgart-Mitte	48.779495	9.179876	5677	21606	103.0	40.0	23.0
14	Stuttgart-Nord	48.796661	9.176252	3459	23578	114.4	8.0	3.0
15	Stuttgart-Ost	48.776972	9.207365	4614	41684	98.0	5.0	2.0
16	Stuttgart-Süd	48.753021	9.132492	4015	38491	105.9	0.0	0.0
17	Stuttgart-West	48.777659	9.151351	6606	45635	107.4	27.0	7.0
18	Untertürkheim	48.783154	9.255246	2339	14164	92.6	4.0	2.0
19	Vaihingen	48.734429	9.088648	1866	38992	107.7	3.0	1.0
20	Wangen	48.773070	9.238802	2260	7742	84.8	3.0	2.0
21	Weilimdorf	48.810704	9.105939	2098	26403	98.4	2.0	0.0
22	Zuffenhausen	48.833891	9.169893	2653	31733	88.8	7.0	1.0

Fig. 2.1 Main dataframe

3. Methodology and Analysis

After cleaning and preparing the data, let us identify the steps, that have to be performed in order to find the most optimal boroughs.

Firstly, we will apply some basic exploratory analysis to our data. For that let's find the location of each borough on the map. Then we can visually inspect some values in our data with the help of bar charts.

Secondly, we have the possibility to reduce the number features in data frame by replacing them with more reasonable data.

Finally, we will perform cluster analysis to find the best cluster of boroughs with meaningful features.

3.1 Exploratory Data Analysis

First, it's is quite useful to visualize the center locations of each borough. For that, the map of Stuttgart was created with boroughs superimposed on top.

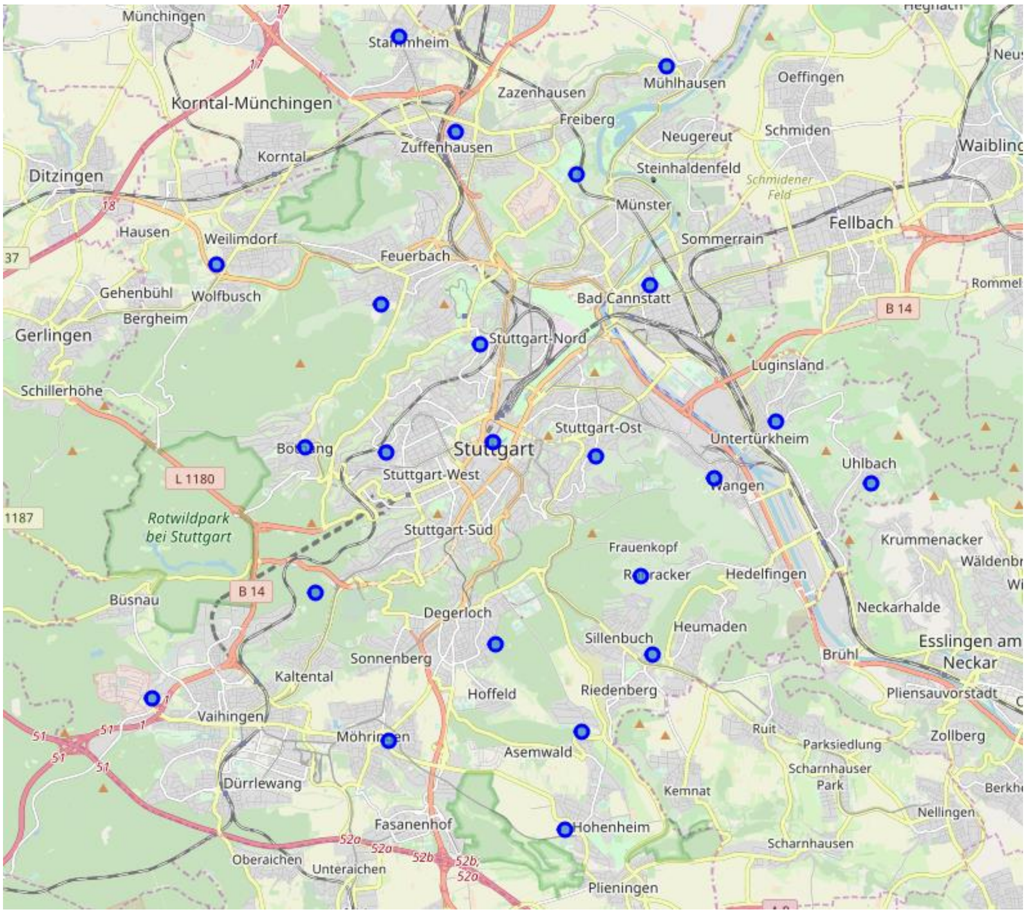


Fig. 3.1 Stuttgart map with centers of the boroughs

Additionally, the several features, that influence our choice, were visualized with respect to the Boroughs.

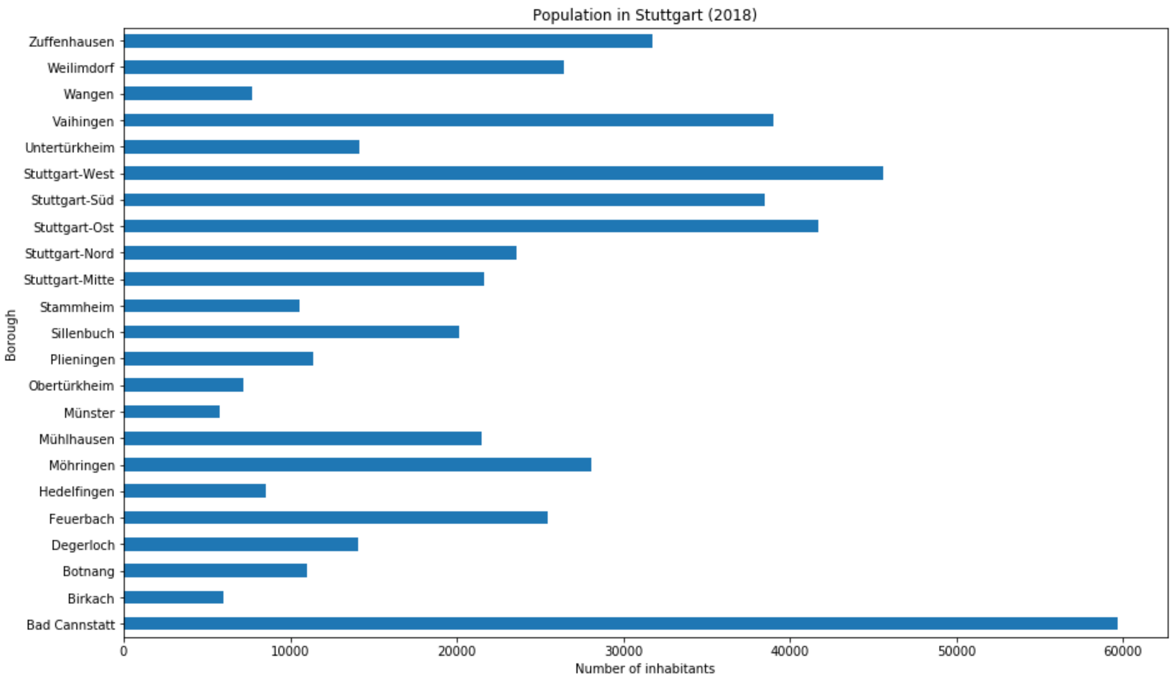


Fig. 3.2 Population in Stuttgart (2018)

Fig. 3.2 shows that Bad Cannstatt is most populous borough in the city. However, since it has one of the largest areas, the population density barely exceeds the average value (Fig. 3.3). From the figure, the most densely populated borough is Stuttgart-West.

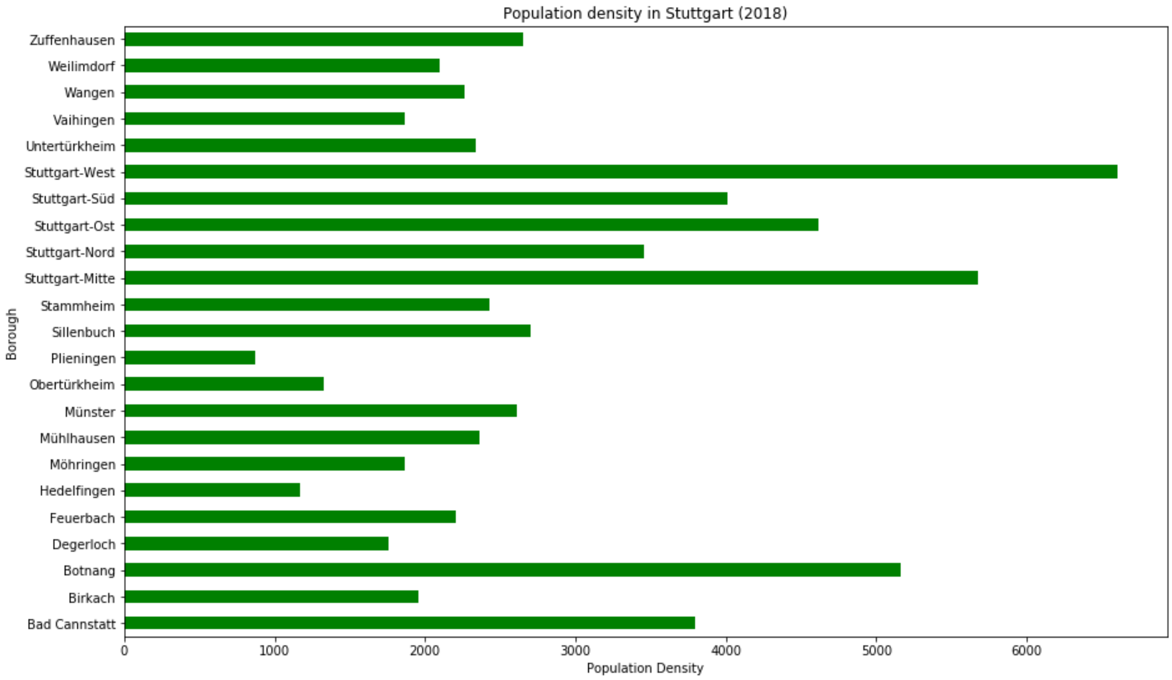


Fig. 3.3 Population Density in Stuttgart (2018)

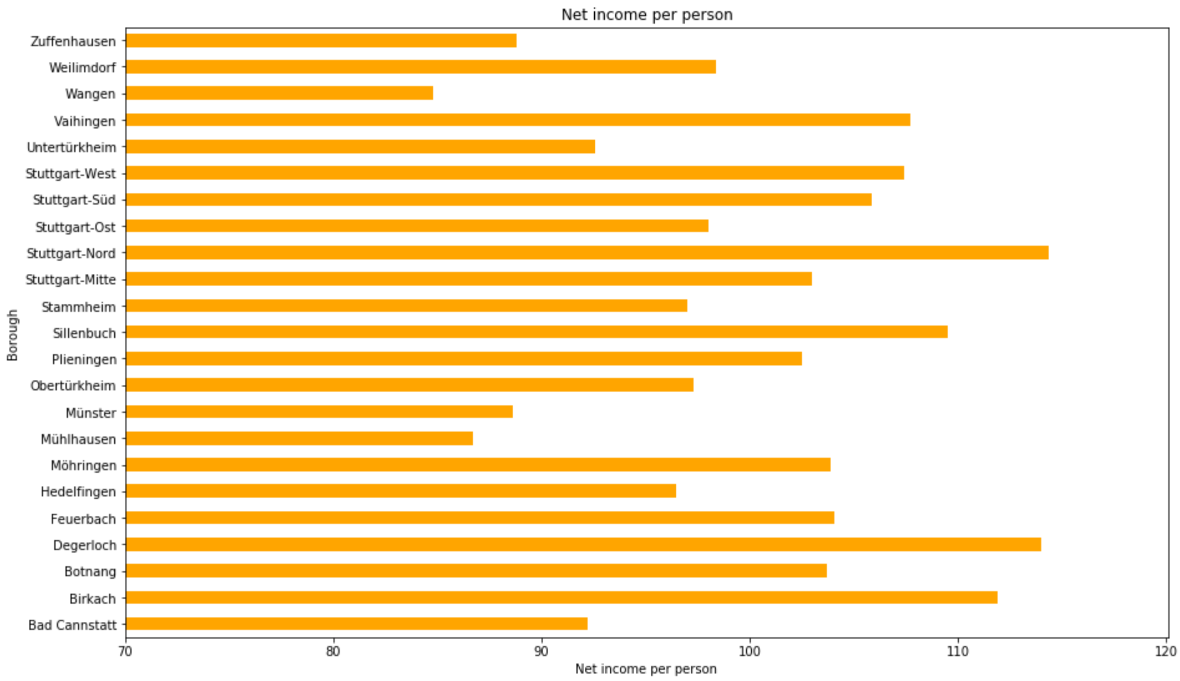


Fig. 3.4 Net Income per Person

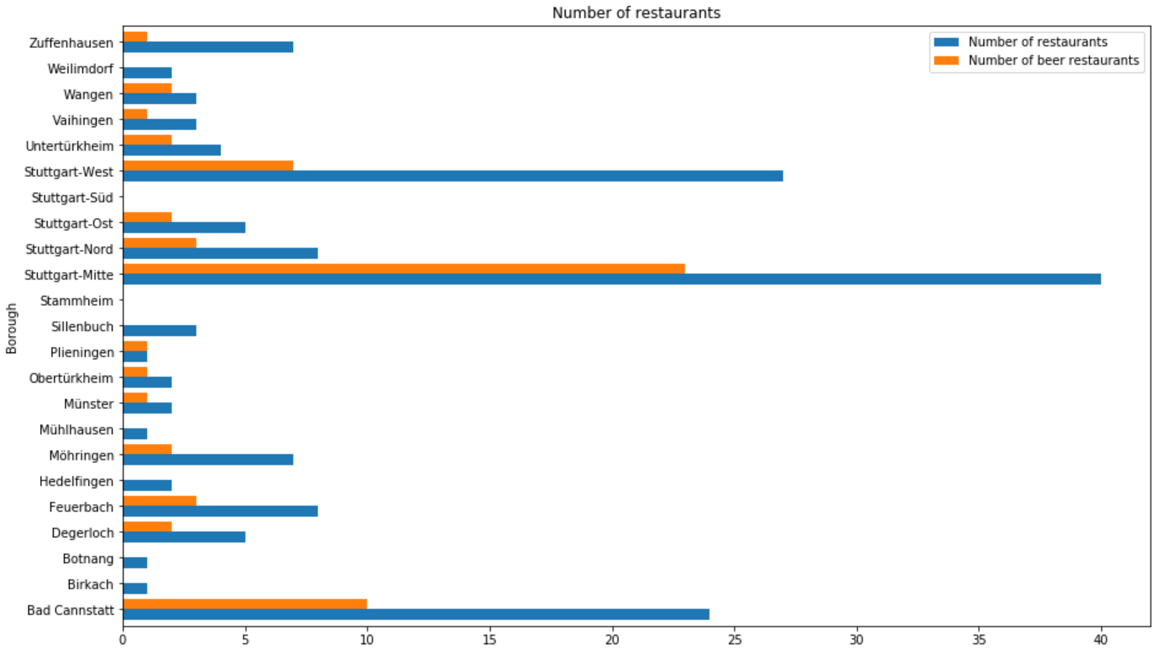


Fig. 3.5 Number of restaurants

Fig. 3.5 reveals a potential underdog of our analysis – Stuttgart-Mitte. This borough is littered with the high number of restaurants, especially with the beer concept ones.

3.2 Cluster Analysis

To identify groups (clusters) with similar characteristics, the unsupervised learning method to our data, namely K-Means algorithm, was applied to our data.

But before that, to reduce dimensionality of the problem the columns "Population", "Number of restaurants" and "Number of beer restaurants" were removed. These three columns were replaced with two new ones, namely, "Number of restaurants per thousand people" and "Number of beer restaurants per thousand people".

	Net income per person	Number of restaurants per 1000 people	Number of beer restaurants per 1000 people
0	92.2	0.402111	0.167546
1	111.9	0.165837	0.000000
2	103.7	0.090736	0.000000
3	114.0	0.355114	0.142045
4	104.1	0.313762	0.117661

Fig. 3.6 Modified dataframe

To identify the optimal number of clusters, the Elbow method is used:

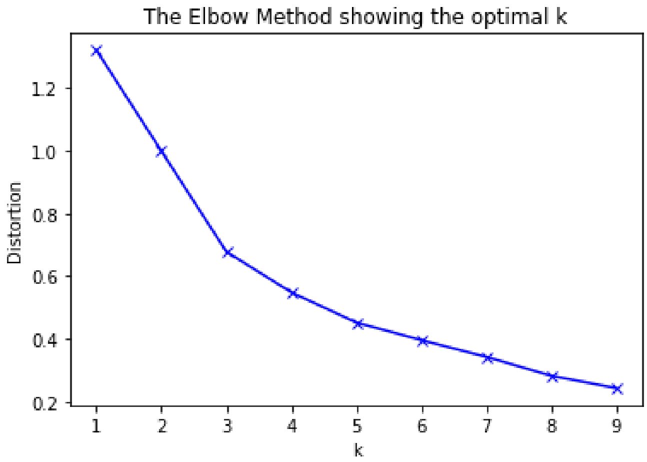


Fig. 3.7 The Elbow method showing the optimal number of clusters

It can be seen from the graph that three clusters are the best choice.

Based on clustering results, two maps are created. The first map illustrates the clusters where the colors are associated with clusters and radius of the Circle marker is proportional to a number of restaurants per 1000 people in each borough. The second map shows the clusters where the radius of the Circle marker is proportional to a Net income per person in each borough.

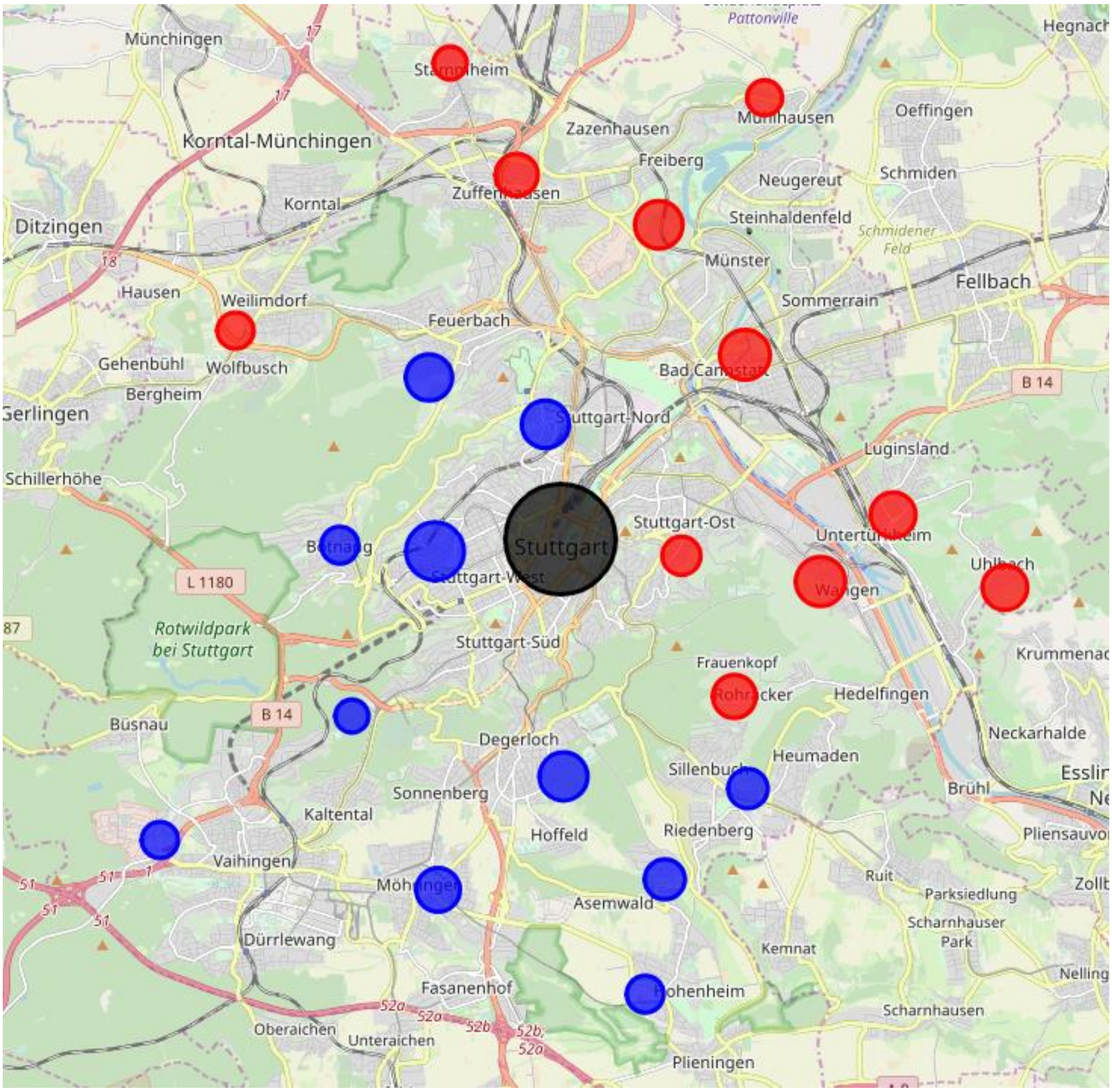


Fig. 3.8 Stuttgart Map with clustered boroughs (radius of the circle marker is proportional to a number of restaurants per 1000 people in each borough)

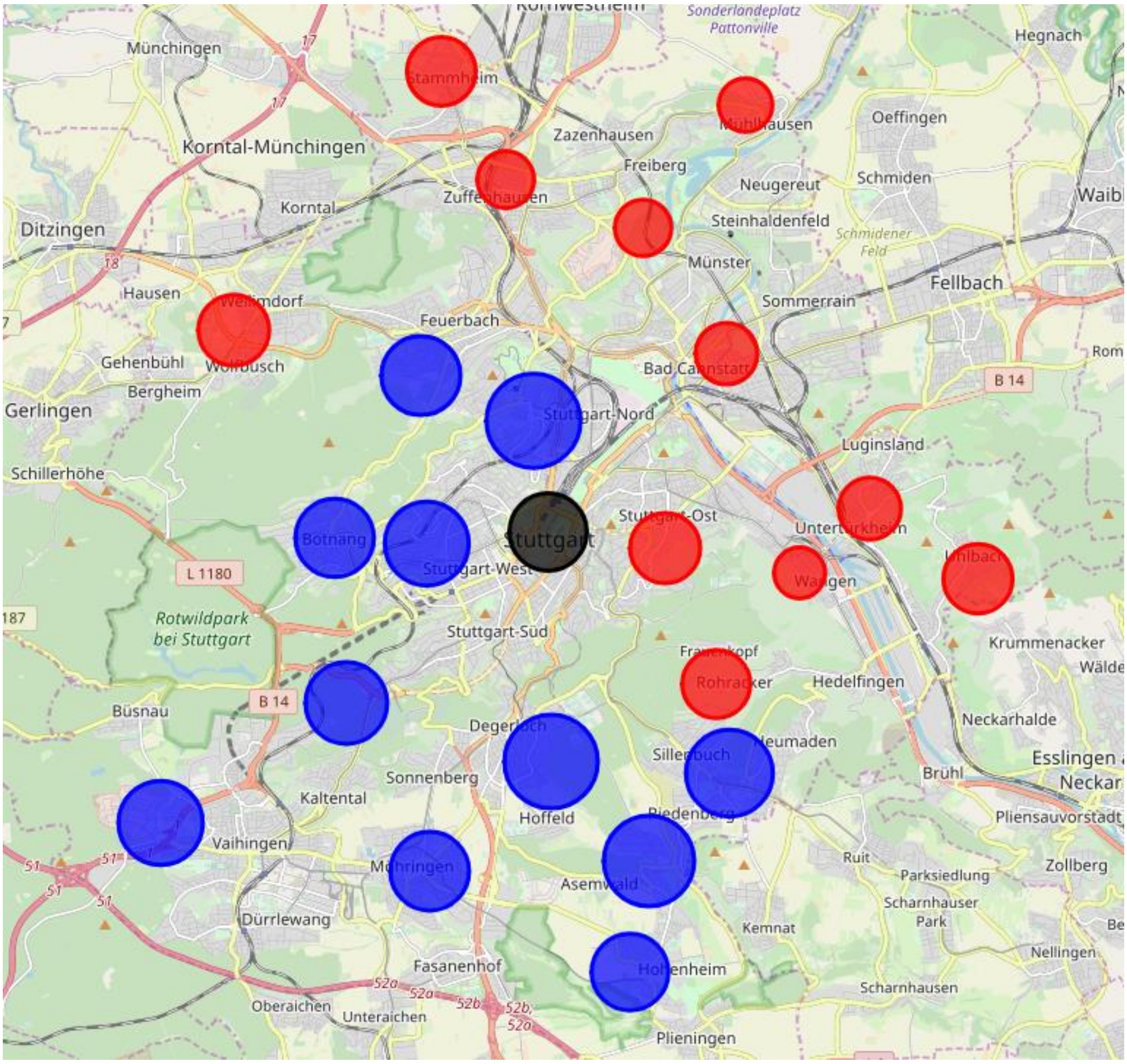
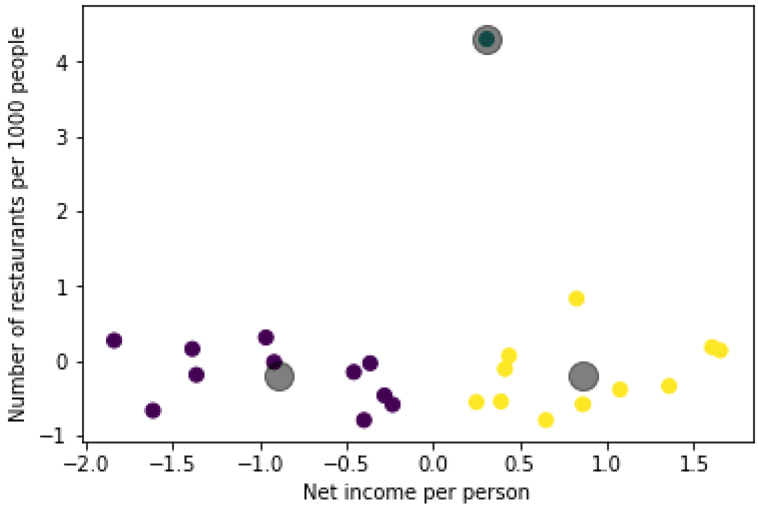


Fig. 3.9 Stuttgart Map with clustered boroughs (radius of the circle marker is proportional to a net income per person in each borough.)

We can see that one of the clusters (black circle) consists only of one borough - Stuttgart-Mitte.

Let's look at the scatter plots of our data and define our clusters with colors. The grey circle marker is representing the centroid of each cluster. Don't forget, that our data is normalized, so the axes do not deliver real values.



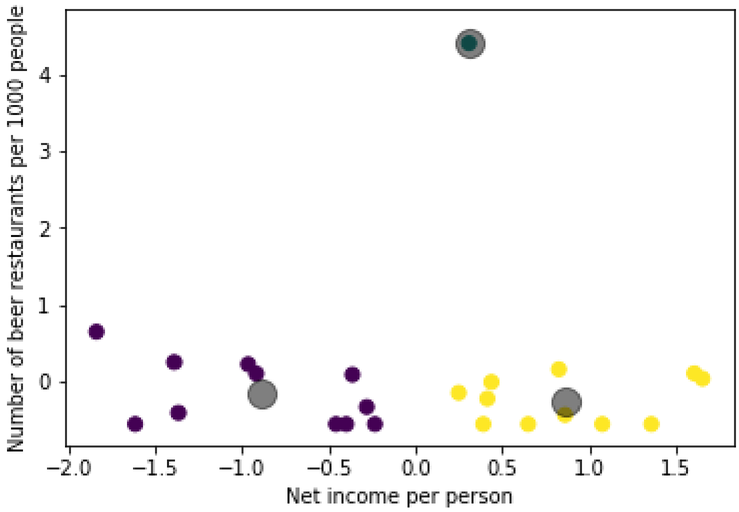


Fig. 3.10 Scatter plots of clustered boroughs

One can observe obvious outlier here. This borough has too high concentration of beer restaurants and restaurants in general. From maps above, we can easily say that it is Stuttgart-Mitte.

Two other clusters were defined according to a net income per person.

4. Results and discussion

During the analysis, three clusters were defined. One cluster, that consists of only one area, has been defined as the outsider, due to the high number of competitors, which means that the placement of beer restaurant in that area is too risky venture. Two other groups were clustered according to the amount income per person. It is obvious, that the cluster with highest average income per person has the highest priority for us (Cluster 2).

Stuttgart-Sud and Stuttgart-Nord are the most attractive options in terms of distances to the center of their own cluster and relatively high value of income per person. However, one can perform further analysis of this particular cluster with additional features, such as distance to the center of city or to the center of cluster.

After defining a borough, one can perform deeper analysis to find the best exact location of the restaurant taking into account factors such as number of parking places in the vicinity of the spot or distances to the main streets.

What could be done better?

Foursquare doesn't represent the full picture, since many venues are not on the list. For that reason, another map could be utilized such as Google map or Openstreet map.

Boroughs have too complex geometry, thus defining the closest venues within the certain radius brings additional error to our analysis.

5. Conclusion

To conclude, the basic data analysis was performed to identify the most optimal boroughs for the placement of the beer restaurant in the city of Stuttgart. During the analysis, several important statistical features of the boroughs were explored and visualized. Furthermore, clustering helped to highlight the group of optimal areas. Finally, Stuttgart Sud and Stuttgart Nord were chosen as the most attractive options for the further analysis.