Applied Data Science Capstone Project

The Best Places in Frankfurt to Open an Uzbek Restaurant: A Clustering Approach

Bekzod Makhmudov
LinkedIn | GitHub

October 10, 2019

Introduction and Business Problem

We are interested in opening an Uzbek restaurant in Frankfurt, which potentially can attract a high number of customers. Ideally, our stakeholders are interested in a narrowed-down list of potential locations to explore further. The choice of proper locations for an Uzbek restaurant in Frankfurt is very important, because Uzbek cuisine is not widely spread in Frankfurt unlike any other Asian or European cuisines, though it is very popular among European tourists visiting Uzbekistan due to its great potential in gastronomic tourism. Currently, there is only one Uzbek restaurant in Frankfurt and it is located in Wöllstädter Str. 11, 60385 Frankfurt am Main in Bornheim neighborhood.

Our stakeholders point out that a potentially good location for an Uzbek restaurant in Frankfurt should comply to the following criteria:

- Locations close to the city center
- Locations with low restaurant density
- Locations with few or no close Asian restaurants (competitors)
- Locations with high population density
- Locations close to tourist destinations/attractions

Our goal is to come up with a manageable list of optimal locations in Frankfurt that satisfy all of the above criteria.

Data

According to the criteria that we have defined in our business understanding section, the following factors will influence the choice of optimal locations for an Uzbek restaurant in Frankfurt:

- The closer a location to the city center, the better. The city center will be defined at **Hauptwache** (the street is called “An der Hauptwache”). **The distance between a potential location and Hauptwache** will determine the closeness criterion.
- **The number of restaurants** nearby a potential location will be counted towards the **restaurant density**. The less restaurants surround the location, the more suitable it will be.
- **The number of Asian restaurants** surrounding a given location will be counted to determine suitability of a location, as they are close competitors to the Uzbek cuisine.
- **Neighborhoods with high population density** will be preferable for an optimal location choice, because potentially more customers can be attracted there.
- Finally, **a location close to the major tourist attractions** will be taken into account as these places also attract many foreign tourists who can potentially become customers of a nearby Uzbek restaurant.

The following data sources will be used to extract required data:

- **Distances from the city center, from the main tourist attractions to potential locations** will be obtained using **Google Maps API geocoding and revers geocoding**.

- **The names of most popular tourist attractions** in Frankfurt will be obtained from **Tripadvisor.com using web scraping** as it does not provide API access for data analysis purposes.
- **The number of restaurants,** their **category** and **type** as well as their **geolocation** information will be obtained through **Foursquare API**.
- **The names of the neighbourhoods** in Frankfurt and their **population density** will be obtained using web scraping from Wikipedia through the link below: https://de.wikipedia.org/wiki/Liste_der_Stadtteile_von_Frankfurt_am_Main.

Methodology

After getting the names of the neighborhoods in Frankfurt (46 neighborhoods) and obtaining their geolocational coordinates, I will proceed to determine the **geographical centers** for each neighborhood, which will consist of **latitude and longitude** as well as **a precise street name**. I will do that using Google Map API geocoding and reverse geocoding. The decision to choose a geographical center rather than a popular neighborhood center is performed for exploratory reasons and for getting initial insights about the neighborhoods. This can be changed according to the requirements of stakeholders.

After obtaining the information about the "neighborhood centers", I will explore **the number of restaurants in the radius of 500 m around each neighborhood center** using Foursquare API. Choosing the radius of 500 m was done for exploratory reasons and it is meant to be a reasonable area around each neighborhood center to analyze the level of competition. Furthermore, among these restaurants within the radius **the number of Asian restaurants** will be of particular interest because they are considered to the main competitors for an Uzbek cuisine. It is important to note that **only category "restaurant" will be considered**, and all the other food venues will be ignored in order to produce a consistent analysis.

Next, the names of **top 100 tourist attractions in Frankfurt** will be obtained from Tripadvisor.com and their geolocational coordinates (longitude and latitude) will be identified using Google Map geocoding API. This geolocational information will be used to calculate **the average distance between a neighborhood center to these tourist attractions**, which will give us insight on how close each neighborhood is on average to the main tourist destinations.

Furthermore, each neighborhood center will be associated with **the population density of the whole neighborhood** to explore how attractive a particular neighborhood is to open a new restaurant. Also, **the distance of the neighborhood center to the center of Frankfurt (Hauptwache)** will be measured to explore how quickly can this location be reached from the city center.

For exploratory analysis I will use Python's Folium library to construct maps with circle markers for locations of neighborhoods, restaurants, and tourist attractions. These maps will be used to get the initial insights about each neighborhood in terms of their vicinity to Asian restaurants, and to the main tourist attractions.

Furthermore, the locations of the top 100 tourist location will be displayed on the map of Frankfurt to analyze their concentration level to a particular neighborhood. Also, various relationships between the key decision making criteria in choosing a potential location will be analyzed using scatter plot and simple regression analysis.

Next, I will proceed to cluster neighborhood center according to the following 5 features:

- Distance to the city center
- Average distance to the top 100 tourist attractions
- The number of restaurants within 500 m radius around the neighborhood center
- The number of Asian restaurants within 500 m radius around the neighborhood center
- Population density of a neighborhood.

I will use k-Means clustering algorithm because this is a widely applied method to cluster items based on their similarity. I will apply Python's Scikit-learn library to identify clusters. I will also use Elbow method to determine an optimal number of clusters using sum of squared errors technique. To give an equal weight to the numerical features, I will use standard scaling method to prepare the data for clustering analysis.

The purpose of using clustering of neighborhoods is to get insights about the nature of neighborhoods in terms of the five features and try to reveal commonalities of neighborhoods within each cluster. This will help me to preselect neighborhoods in the most suitable cluster and provide a finalized list of potential locations (neighborhood centers) to open an Uzbek restaurant in Frankfurt.

Another interesting contribution of this analysis will be to evaluate the location of the existing Uzbek restaurant in Frankfurt in terms of the key criteria provided by the stakeholders.

Analysis

Using the data sources described above, I have obtained all the necessary information on the five key decision variables, namely (1) the population density of each neighborhood, (2) distance from a neighborhood center to the city center, (3) total restaurants within a neighborhood center as well as (4) the number of potential competitor restaurants, and (5) the median distance from a neighborhood center to the top 100 attractions in Frankfurt.

The following table provides the main descriptive statistics of our five key decision-making variables.

	Population density	Distance to city center	Competitor	Total restaurants	Median distance to attractions
Count	45	45	45	45	45
Mean	4792.76	7818.09	0.53	4.13	8121.60
Std	3043.43	4772.23	0.94	4.46	4691.81
Min	564.00	1698.00	0.00	0.00	1811.50
25%	2512.00	4063.00	0.00	1.00	3967.50
50%	4644.00	6795.00	0.00	2.00	7381.00
75%	6384.00	9376.00	1.00	8.00	10274.50
Max	15031.00	22428.00	4.00	14.00	19858.50

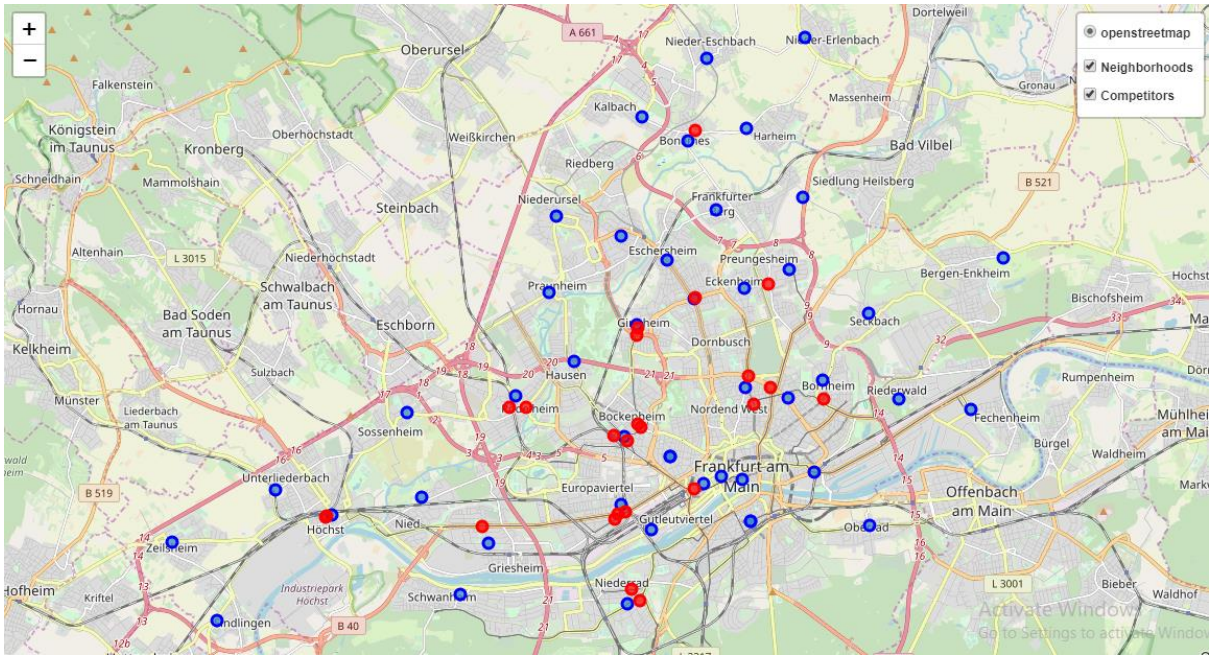
On average, a neighborhood has a population density of **4793 people per km^2**, it is located **7.8 km from the city center** and **8.1 km from top 100 tourist attractions**, has **4.1 restaurants** and **0.5 competitor restaurants**. These are **benchmark values of key decision variables** and will be referenced in our results section when we look at clustering of neighborhoods. In general, locations below these benchmark values will be preferred.

Using Foursquare API, I have obtained the names and locations of venues within 500 m radius of each neighborhood center. Since I am interested in the restaurant category only and the filtered list contains a total of **186 restaurants** in all neighborhoods (182 unique names). To determine a competitor restaurant type, it was revealed that not only Asian restaurant category is present, but there are also other types of restaurants that may well compete an Uzbek restaurant. The following list summarizes the competitor list.

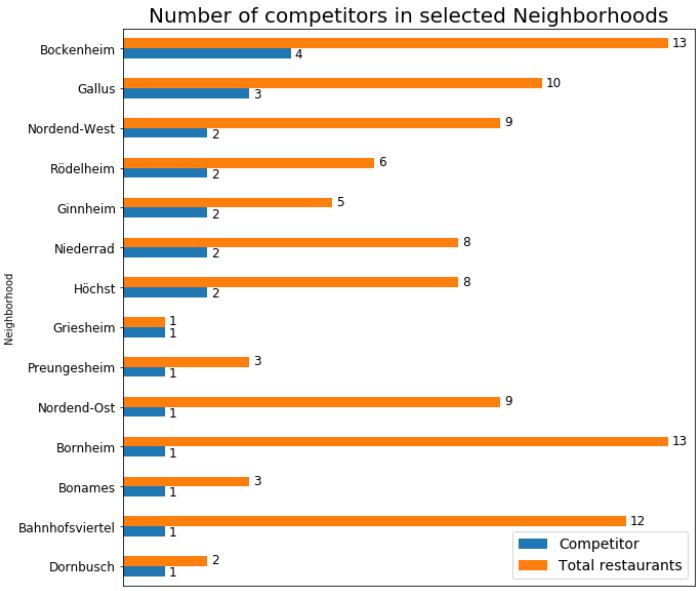
Competitor restaurant categories (24 competitors):

- Asian Restaurant – 12
- Doner Restaurant – 4
- Kebab Restaurant – 1 (The existing Uzbek restaurant in Bornheim is under this category)
- Persian Restaurant – 1
- Turkish Restaurant – 5
- Middle Eastern Restaurant – 1.

The following map shows the locations of 45 neighborhood centers (blue circles) and locations of 24 competitor restaurants (red circles) in the city of Frankfurt, Germany.



The locations of the competitor restaurants show that **they are relatively spread around the middle and north part of Frankfurt** and **one can hardly find any competitors in the south-eastern part of the city**.

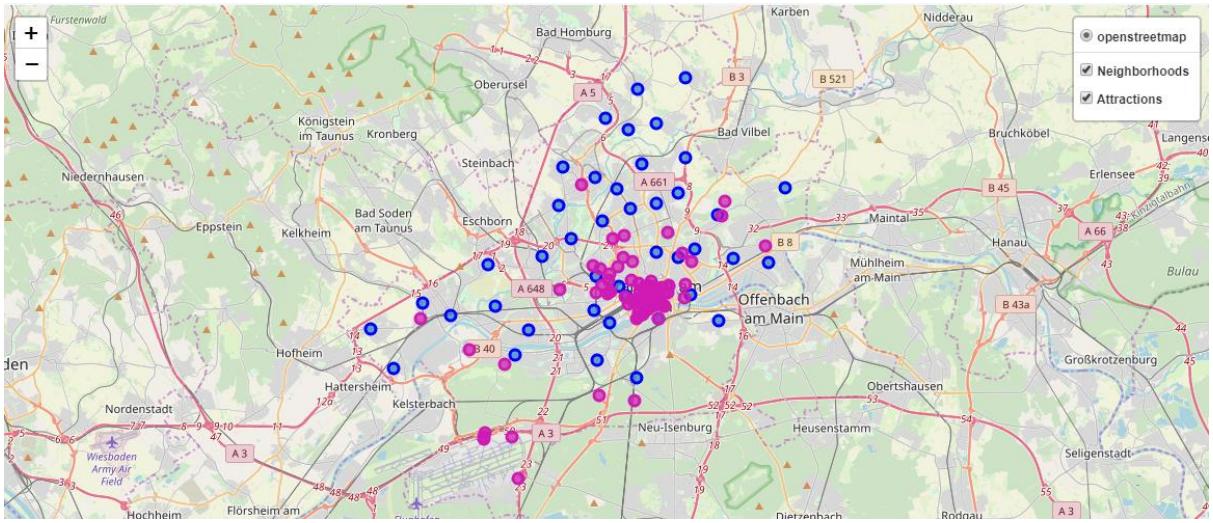


The 24 competitor restaurants are distributed across 14 neighborhood centers. The top neighborhoods with most competitors are **Bockenheim, and Gallus with 4 and 3 competitors** respectively.

In general, it is possible to classify neighborhoods by high (3 to 4), medium (2), low (1) and no competition.

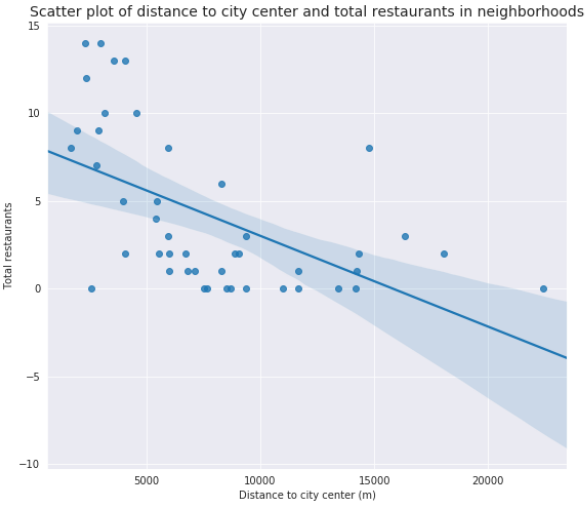
Neighborhoods with low and no competition are potentially good location candidates to open an Uzbek restaurant.

The following map shows the locations of top 100 attractions (purple circles) obtained from Tripadvisor.com and geocoded using Google Map API.



The locations of the top 100 attractions in Frankfurt show that t**hey are very concentrated around the center of Frankfurt**. This is not surprising because almost all historical buildings, squares and museums are located there.

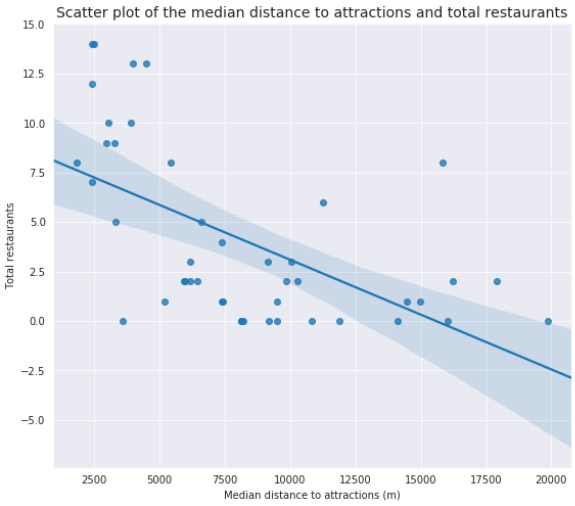
The correlation between the distance from a neighborhood center to the city center and the average distance from a neighborhood center to these attractions confirms this finding. **All in all, if a restaurant is located close to the city center, it will also be located very close to the main city attractions.**

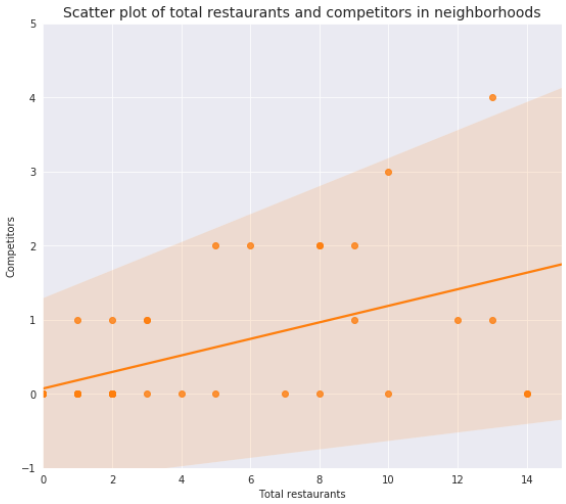


Now, I will test hypotheses of key location choice criteria with the data for all the neighborhoods. The scatter plot on the left above shows the correlation between the distance to city center and the total restaurants for all neighborhoods. The regression line is negative (correlation coefficient is -0.554) and significant at 95% level (p-value is 0.0001). **This means that neighborhood centers located closer to the city center have higher restaurant density. Distance to the city center is a significant determinant of location choice for restaurants.**

The next hypothesis concerns the population density of a neighborhood. The scatter plot and the regression line on the right on the previous page show that a correlation between population density and total number of restaurants in neighborhoods is positive (correlation coefficient is 0.634) and

statistically significant (p-value is zero). **This confirms the importance of population density in determining locations of restaurants in Frankfurt.**





We can also observe the linear relationship between a median distance to attractions and total restaurants (restaurant density) in neighborhoods. According to the Pearson correlation statistics, this relationship is negative (coefficient is -0.582) and statistically significant (p-value is zero). **This means that a median distance to top 100 attractions is a good predictor of restaurant density in Frankfurt. The lower the median distance to attractions from a neighborhood center, the higher is the density of restaurants in that neighborhood center**. In fact, here we can observe a curvature in the scatter plot, which means that in neighborhoods with median distance to attraction lower than 5 km, the rate of increase of the restaurant density is much higher than in neighborhoods with median distance higher than 5 km.

Is the restaurant density in neighborhoods a good predictor and determinant of competitors? According to the scatter plot and the regression line on the right, we can observe a positive linear relationship (correlation coefficient is 0.528) between the restaurant density and the number of competitors in neighborhoods and this relationship is statistically significant (p-value is 0.0002). **We can find more competitors in neighborhoods with higher levels of restaurant density. This relationship satisfies two decision criteria: it is possible to find locations with low restaurant density and low number of competitors.**

However, locations with low levels of competition, and hence with low levels of restaurant density, are found in areas with lower population density, farther from the main attractions and farther from the city center. Therefore, the two key decision criteria, which are levels of competition and restaurant density, must be either omitted or changed. **I would recommend changing the two criteria as follows: locations with higher levels of restaurant density and higher levels of competition are preferable.**

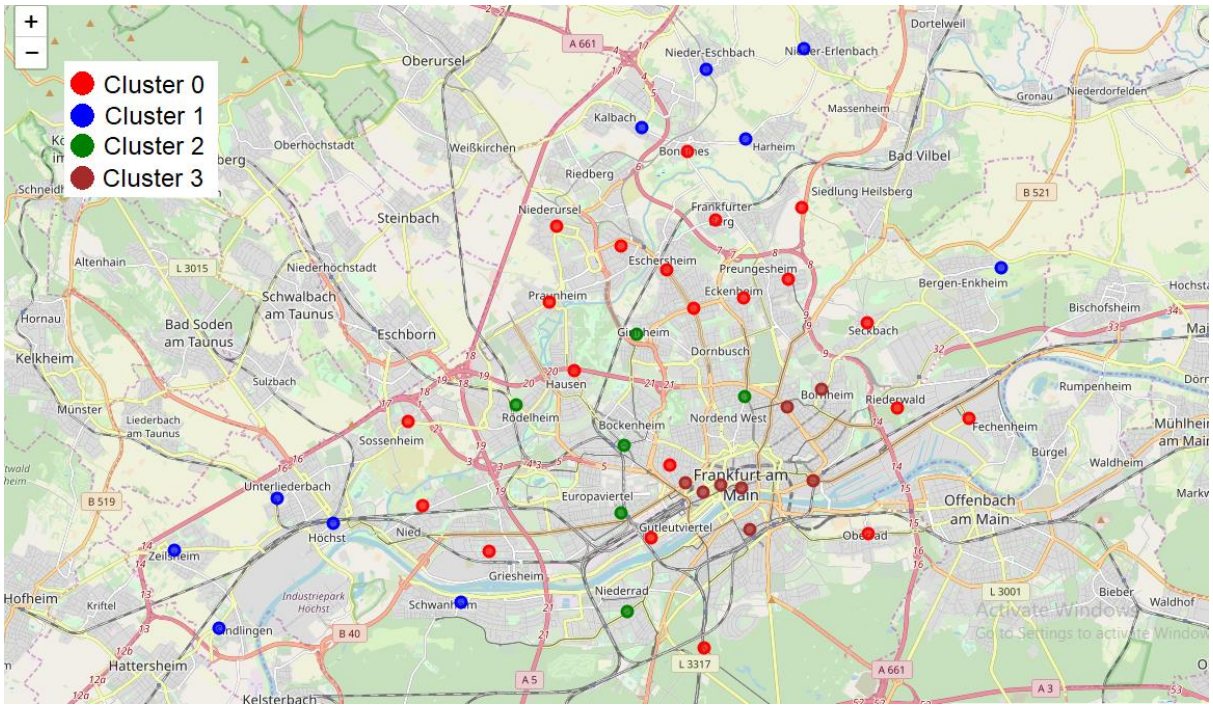
With the above analysis of neighborhood characteristics (features) we can conclude that all five decision criteria for choosing a potential location for an Uzbek restaurant are important determinants of restaurant locations in Frankfurt. **With some changes to the key decision criteria for choosing a suitable location for an Uzbek restaurant, we can clarify them again as follows**:

- Locations close to the city center
- Locations with high restaurant density
- Locations with some or many close Asian restaurants (competitors)
- Locations with high population density
- Locations close to tourist destinations/attractions

Results

The dataset on the five key decision-making criteria was prepared and preprocessed using Python's Sci-kit learn library. This library is also used to build clustering algorithm to group all neighborhood centers by their features. Using Elbow method, I have obtained an optimal number of k = 4 clusters based on the five key decision features of neighborhoods.

The folium map below shows the four clusters of neighborhoods displayed in different colors.



What insights can we get from this map? As can be seen, cluster 3 includes neighborhoods located in and around the center of the city. Their most important features are **close location to the city center and to the main attractions**.

Cluster 2 includes neighborhoods on the western part of the city. Their distinguishing characteristic can be **middle distance to the center of the city and to main attractions**.

Cluster 0 surrounds clusters 3 and 2 and includes **the greatest number of neighborhoods** compared to other clusters. The neighborhoods in this cluster are **mostly concentrated in northern part of the city**.

Finally, cluster 1 comprises of neighborhoods located on the western, northern and north-eastern borders of the city and their main features are **a long distance to the city center and hence to the main tourist attractions**.

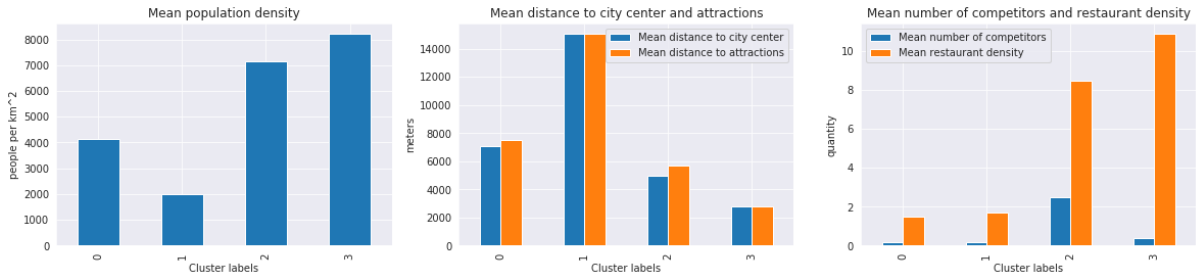
Next, we will compare the quantitative measures of our five key decision variables to the benchmark values and identify the most promising clusters and neighborhoods to open an Uzbek restaurant.

The table below shows the mean values of the key cluster features.

Cluster labels	Mean population density	Mean distance to city center	Mean number of competitors	Mean number of restaurants	Mean distance to attractions
Cluster 0	4147.57	7109.48	0.19	1.48	7549.29
Cluster 1	1996.90	15044.00	0.20	1.70	15041.85
Cluster 2	7161.50	4965.33	2.50	8.50	5689.08
Cluster 3	8204.63	2785.38	0.38	10.88	2798.00
Benchmark	4792.76	7818.09	0.53	4.13	8121.60

The table above summarizes clusters against the key decision variables according to the defined criteria and extreme values are highlighted in red color. Finally, it is possible to identify important insights from the clustering analysis of the neighborhoods:

- Insight #1: Cluster 3 is an ideal candidate to choose neighborhood locations from in order to open an Uzbek restaurant. Surprisingly, the mean number of competitors is even lower than was defined earlier for decision making purposes (0.38 against 2.50).
- Insight #2: The greatest number of competitors is included in Cluster 2, which is a middle-western part of Frankfurt. This cluster is characterized by a relatively high population and restaurant densities and average distance to the city center and main attractions according to our benchmark. The only tradeoff from cluster 3 is its remoteness from the city center and tourist attractions. The choice of this cluster for most competitors is probably driven by being close to their communities rather than to tourists visiting Frankfurt.
- Insight #3: Since there are no Uzbek communities in sizes comparable to other Asian communities, the choice of proximity to the city center and to the main tourist attractions in deciding on potential locations to open an Uzbek restaurant is absolutely reasonable.



Now, we can generate the final table with a list of potential locations from cluster 3 to review for our stakeholders and make the final decision about the desired location to open an Uzbek restaurant.

The following table shows optimal locations in Frankfurt to open an Uzbek restaurant.

Neighborhood	Neighborhood center	Population density	Distance to city center	Number of competitors	Number of restaurants	Median distance to attractions
Altstadt	Römerberg 16, 60311 Frankfurt am Main, Germany	8187	2792	0.0	7.0	2424.0
Bahnhofsviertel	Jürgen-Ponto-Platz 2, 60329 Frankfurt am Main,...	6558	2383	1.0	12.0	2405.5
Bornheim	Im Prüfling 63, 60389 Frankfurt am Main, Germany	10959	4065	1.0	13.0	4500.5
Innenstadt	Kaiserstraße 19-21, 60311 Frankfurt am Main, G...	4430	1698	0.0	8.0	1811.5
Nordend-Ost	Rothschildallee 20, 60389 Frankfurt am Main, G...	15031	2880	1.0	9.0	3288.5
Ostend	Hanauer Landstraße 89, 60314 Frankfurt am Main...	5246	3160	0.0	10.0	3026.0
Sachsenhausen-N	Stegstraße 77, 60594 Frankfurt am Main, Germany	7676	2982	0.0	14.0	2428.0
Westend-South	Bockenheimer Landstraße 69, 60325 Frankfurt am...	7550	2323	0.0	14.0	2500.0

Discussion of results

The analysis of location problem using spatial data was conducted to come up with a manageable list of potential locations in Frankfurt to open an Uzbek restaurant. The final result obtained from my clustering analysis of 45 neighborhood centers in Frankfurt consists of specific locational addresses with reference characteristics of those locations. All in all, there are 8 potential locations that can be presented to the stakeholders who wish to find a place that meet their five key requirements.

The stakeholders can now proceed further and inspect these locations on sight and consider other important factors that influence restaurant location decision, for example, cost of renting a space, its offered sizes in that particular neighborhood, neighborhood traffic, access to public transportation and so on.

In this particular cluster of interest (cluster 3), 4 out of 5 neighborhood features meet stakeholders' initially set expectations for a potential location. Although these locations have only one or no competitor restaurants, there are many other restaurants whose number can range from 7 to 14 within 500 m radius for a given location of this cluster. I think that having many restaurants around would not necessarily influence negatively the operation of a new restaurant, provided that this can lure potential customers who value a comfortable place to spend evenings and, therefore, the diversity of such places may add value to potential customers.

In fact, the only Uzbek restaurant one can find in Frankfurt is located in Bornheim, which is in a list of potential neighborhoods in the table above**. An interesting observation from my clustering analysis is that the decision to choose a location for any restaurant in Frankfurt can be based on either its proximity to the city center and tourist attractions or its ability to serve local communities.** In the case of an Uzbek restaurant, since there are no big Uzbek communities in Frankfurt, a decision to choose a location close to the city center and tourist attractions is a viable direction.

Conclusion

In this study, I used clustering analysis to solve a specific spatial problem for stakeholders. I used data science methodology and tools to come up with a manageable list of potential locations in Frankfurt that will be suitable to open an Uzbek restaurant. 45 neighborhoods were analyzed based on five decision variables, namely population density of neighborhoods, distance of neighborhoods to the city center, median distance of neighborhoods to top 100 tourist attractions, the number of restaurants within 500 m radius of a neighborhood center and the number of potential competitors in those areas. These five key variables were calculated using data from various open sources and used in k-Means clustering that divided neighborhoods into 4 clusters. These clusters were labeled based on key characteristics and one cluster was chosen to come up with potential locations to open an Uzbek restaurant in Frankfurt. A table of 8 potential locations with addresses was presented for stakeholders, which can help them to save resources in finding a suitable place.