The Battle of the Neighborhoods

Finding optimal location for a French Restaurant in Cologne

Iliyan Aleksiev

December 2020

1. Introduction

1.1 Background

Cologne is located in the state of North-Rhine Westphalia and is the fourth-most populous city in Germany [1]. It is also the largest city on the Rhine river, split into 9 districts and is most famous for its world heritage cathedral, as well as the carnival festivals which are attracting millions of tourists every year. In addition, the city of Cologne has a unique mixture of people with various cultural backgrounds and this contributes to a very rich gastronomy environment. Due to its proximity to neighboring countries of France, Belgium and the Netherlands, the French cuisine can be particularly popular among tourists and locals.

1.2 Business Problem

This project will focus on finding an optimal location for opening a new French restaurant in Cologne, Germany. The main goal of the project is to analyze how saturated each of the 9 districts are with regards to French restaurants and to support stakeholders who want to find the optimal district in Cologne for opening a new French restaurant.

2. Data

2.1 Data Sources

Given the nature of the business problem, we will consider the following factors for our analysis and decision:

- number of existing French restaurants in each of the 9 districts of Cologne
- population in each of the 9 districts
- average housing price within each district

The data was obtained from the following sources:

- To find information about the different districts in Cologne, data will be scraped from Wikipedia [2]
- Geopy geocoders library is used to find the geolocation for each of the districts
- Foursquare API is used to obtain the most common venues around each district and in particular the French restaurants
- To find data about housing prices, I used a report created by immobilienscout24, a leading real estate company for the German market [3]. It breaks down the housing price for each neighborhood in Cologne for the year 2019. This data was used to generate the average housing price for the 9 districts of Cologne.

2.2 Data Cleaning

After data was scraped from Wikipedia, it had to be put into a data frame and then further transformed, so that unnecessary columns and rows are dropped and column headers have a name that is easy to work with. Final data frame contained informed about each of the districts, along with the respective population.

To find the housing price for each district, a report from immobilienscout24 was used. It provided data for each neighborhood with its corresponding district. The average housing price per district was then calculated and put into a dataframe.

The Geopy geocoders library was then utilized to gather the coordinates of each district and this information was also put into a data frame.

Finally, the data from the 3 data frames is merged into a final one.

3. Methodology

The aim of the project is to find an optimal location for establishing a French restaurant within each of the 9 districts in the city of Cologne. To enable this the following steps were carried out.

Firstly, the data used for the analysis was collected from publicly available sources. The obtained information comprises of the names, geolocation, population & average housing price for each of the districts in Cologne. Fourspace API is utilized to find restaurants in a radius of

6000m from the coordinates of each district. Folium is then used to visualize the data, where barchart and histograms will be used to analyze the distribution of the number of restaurants, population & average housing price.

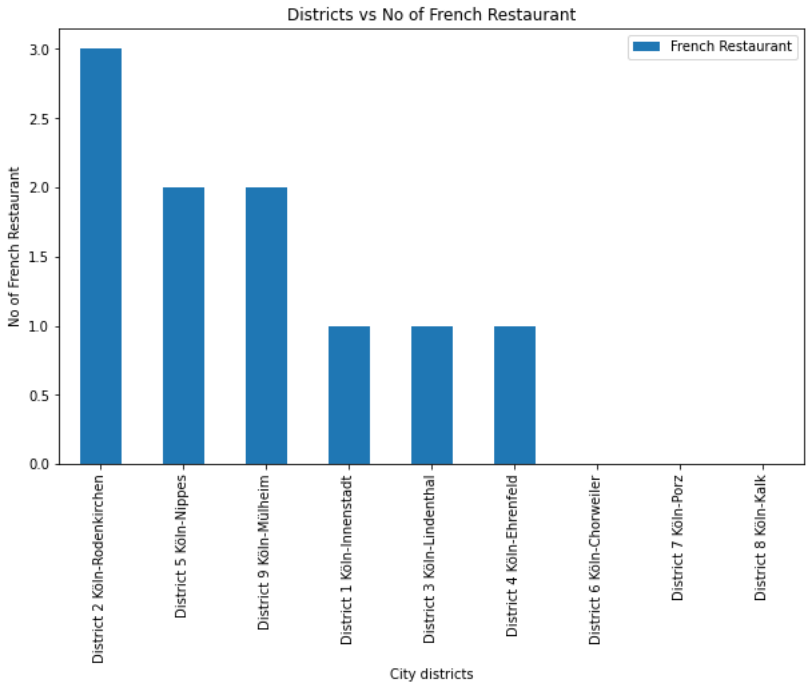
Next step performed is to cluster the districts with an unsupervised learning K-means algorithm. The elbow method is to be utilized to obtain the optimum value for K.

Finally, based on the obtained clusters, outcomes can be discussed and a conclusion can be drawn.

4. Results

4.1 Exploratory analysis

4.1.1 Relationship between city district & concentration of French restaurants

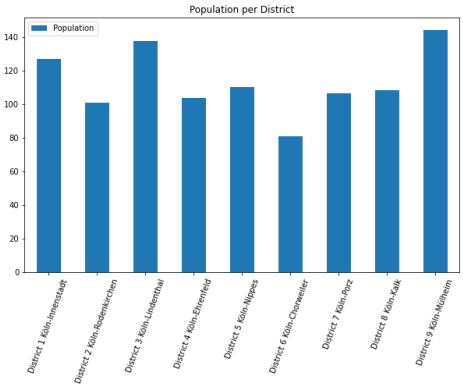


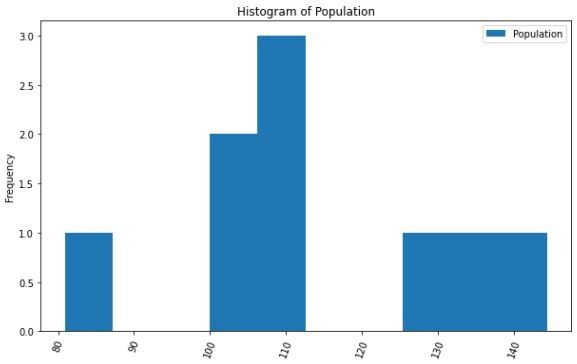
We easily see that District 2 Köln-Rodenkichen has the highest concentration of French restaurants - 3. This is followed by district 5 & 9 with 2 respectively. District 1, 3 and 4 have 1 French restaurant. Finally, District 6, 7 and 8 do not have any restaurants in the radius determined above.

We can then group the Districts in the following manner:

- Value 3: High concentration of French restaurants
- Value 2: Mid concentration of French restaurants
- Value 1: Low concentration of French restaurants
- Value 0: No concentration of French restaurants

4.1.2 Relationship between city district & population



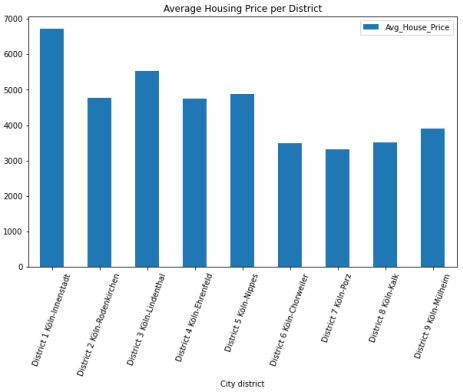


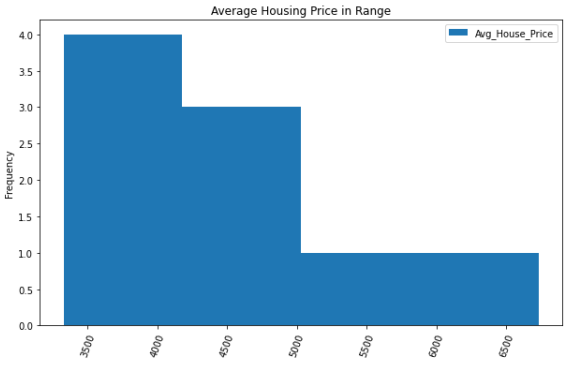
The barchart shows us that district 9 Köln-Mülheim is the most populous area, whereas district 6 Köln-Chorweiler is the least populous.

Based on the histogram we can define the following ranges for the distribution of the population per district:

- Up to 100k people = "Low Population District"
- Between 100k and 125k people = "Mid Population District"
- Above 125k people = "High Population District"

4.1.3 Relationship between city district & housing price





Housing prices are the highest in district 1 Köln-Innenstadt, whereas in district 7 Köln-Porz the average housing prices are the lowest.

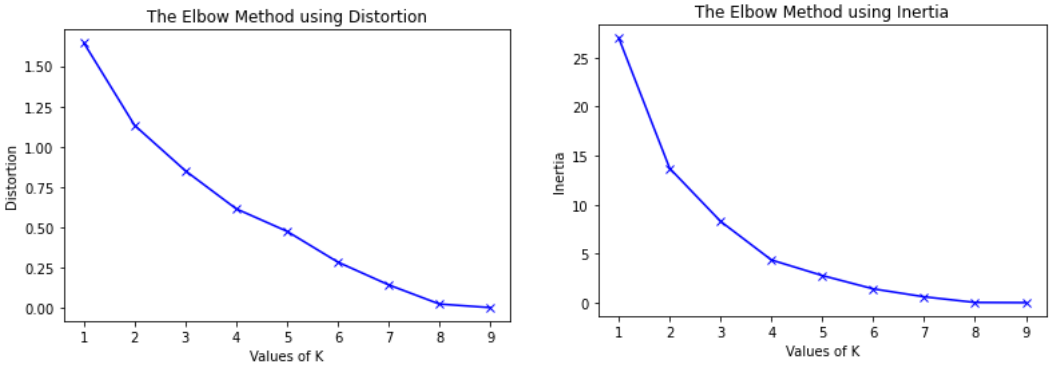
Based on the histogram we can define the following ranges for the distribution of the average housing price per district:

- Up to 4000 EUR/m2 = "Low Average Housing Price"
- Between 4000 and 5000 EUR/m2 = "Mid Average Housing Price"
- Above 5000 EUR/m2 = "High Average Housing Price"

4.2 K-means Clustering

4.2.1 Determine value for K

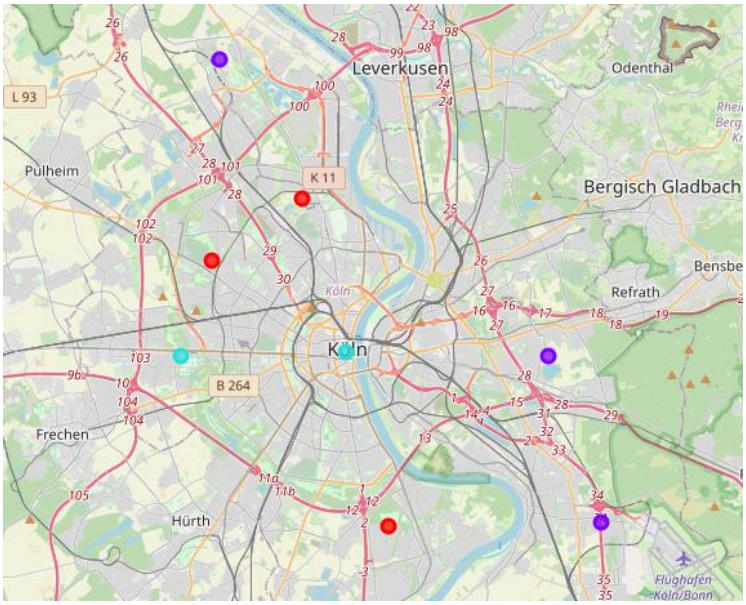
To determine the best value for K the elbow method was applied, once using distortion and once using inertia:



Based on analysis above value of k equal to 4 was chosen.

4.2.2 Model evaluation

After the model was fitted Folium was used to illustrate each of the 4 clusters:



The analysis shows that our collected data is suitable to be grouped into 4 clusters. If we examine the Cluster 0 (red color) we see that a mid-size population & mid to high average housing prices entails having a mixed environment for a French restaurant. Each of the 3 districts in the clusters have a different concentration - from low to high. This represents an interesting outlook for any stakeholder who might be willing to open a new French restaurant in the districts within this cluster that attain a low concentration - in particular District 4 Köln-Ehrenfeld. Next, Looking at Cluster 1 (purple color) we see that low income & low housing prices contribute to no French restaurants in those districts. This can be inferred due to lower purchasing power and lower population density in those areas, nevertheless further analysis with more variables is needed to come to a conclusion if this is indeed the case. In Cluster 2 (turquoise color) it can observed that the grouped districts have high population & high average housing prices. There is however low concentration of French restaurants here - this seems like another good opportunity for any future investor. Let's take a look at the final cluster number 3 (yellow color). It has a high population, low housing prices and a medium concentration of French restaurants. It would be interesting to enrich the analysis with more research as to why this district has a mid concentration. This could be a starting point to further understand whether Cluster 1 can be a potential location for instance.

5. Discussion

All in all, the results suggest that District 1 Köln-Innenstadt & District 3 Köln-Lindethal from the Cluster 2 can be a suitable location for opening a new French restaurant. The high population & high average housing prices infer that there can be a higher demand and higher purchasing power of the consumers in those areas. Given the already low concentration of incumbent restaurants, this may attract the attention of any future investors. Nevertheless, this by no means implies that these two districts are the optimal ones for opening a new French restaurant. It can be that population that lives in those areas has a different taste preferences that does not match the French cuisine, thus this can be a reason for the already low concentration of French restaurants there. Therefore, the recommended districts are meant to

serve as a start point for any additional research that can take into consideration further economical & sociological factors that will support any future investor for his decision.

6. Conclusion

The main goal of this project was to find an optimal location within the city of Cologne for opening a new French restaurant and by that help stakeholders that are interested to commence further analysis into this direction. The 9 districts of Cologne were chosen as a point for comparison and further data was gathered from immobilienscout24 on the average housing prices, as well as from Foursquare to obtain information about the nearest venues around those districts. At the end, the districts were clustered based on the population, housing price and concentration of existing French restaurants. This has created a good suggestion and basis for further research, which can enrich the analysis by including additional variables that can contribute to the final decision of the investor.

7. References

[1] Wikipedia: https://en.wikipedia.org/wiki/Cologne

[2] Wikipedia: https://en.wikipedia.org/wiki/Districts_of_Cologne

[3] Immobilienscout24: https://www.immobilienscout24.de/neubau/ratgeber/aktuelle-neubau-themen/neubau-kauf-map-wohnungen-koeln-2019.html