

Battle of the Neighborhoods

1. Introduction/ Business Problem

1.1 Background

This scenario is based around a restaurant that is located in Berlin, Germany. As the restaurant business goes well, the owner wants to open another branch. The owner wants to settle down in another area to reach more customers and chooses to reach out for Hamburg. Keeping an eye on the investments, the business strategy of the branch, e.g. the menu and the kind of service, should be copied from the original restaurant. Therefore, the new location should be as similar to the original restaurant as possible, to avoid changes in clients.

1.2 Problem

The owner doesn't know Hamburg and needs recommendation for the new branch's location. This project aims to recommend a neighborhood in Hamburg that is a good match with the current business strategy in Berlin.

1.3 Interest

By analyzing the neighborhoods of Hamburg, the restaurant owner increases the chances to find a good fitting neighborhood for the new branch. As the location is a crucial factor for a restaurant's profit, this project's outcome will have a major impact on the following economic development of the branch.

2. Data acquisition and cleaning

2.1 Data sources

The required data includes geographical data from Hamburg, which is scraped from https://de.wikipedia.org/wiki/Liste_der_Bezirke_und_Stadtteile_Hamburgs and put into a Dataframe. The website includes a complete list of the neighborhoods of Hamburg as well as their Latitude and Longitude and the local population density. Using the geopy library, the coordinates of the original restaurant in Berlin is extracted from a given address. Venues of each neighborhood are explored by using the Foursquare API.

Further information about the neighborhoods rental prices is scraped from <https://mietspiegeltabelle.de/mietspiegel-hamburg/>.

2.2 Data cleaning

The Data of Hamburg and the restaurant in Berlin are merged and grouped by neighborhood. In preparation of the following clustering processes one-hot-encoding is used.

2.3 Feature selection

A major feature in this analysis is the venue category. The city is clustered by this feature which allows an overall comparison of the neighborhoods. More specific information about the explored venues will be ignored, like their names and locations.

Besides the venues, the population density and the rental prices are used for further analysis. As the area and the absolute population is taken into account with the population density, those values will not be used for analysis.

In order to make a recommendation of a neighborhood, the list of their names is used in this project. The Boroughs will not be used, as the neighborhoods are more specific.

3. Exploratory data analysis

3.1 Common venues

A list of the top 10 most common venues per neighborhood helps to get a first impression of the neighborhoods. This is used as an additional information for the clustering process.

3.2 Clustering of Hamburg's neighborhoods

All neighborhoods of Hamburg as well as the neighborhood of the restaurant in Berlin are clustered, using the kmeans clustering method. The clustering is based on the venue categories per neighborhood and divides the neighborhoods in five clusters. The cluster that contains the neighborhood of the restaurant in Berlin and similar neighborhoods in Hamburg is shown in red in figure 1.

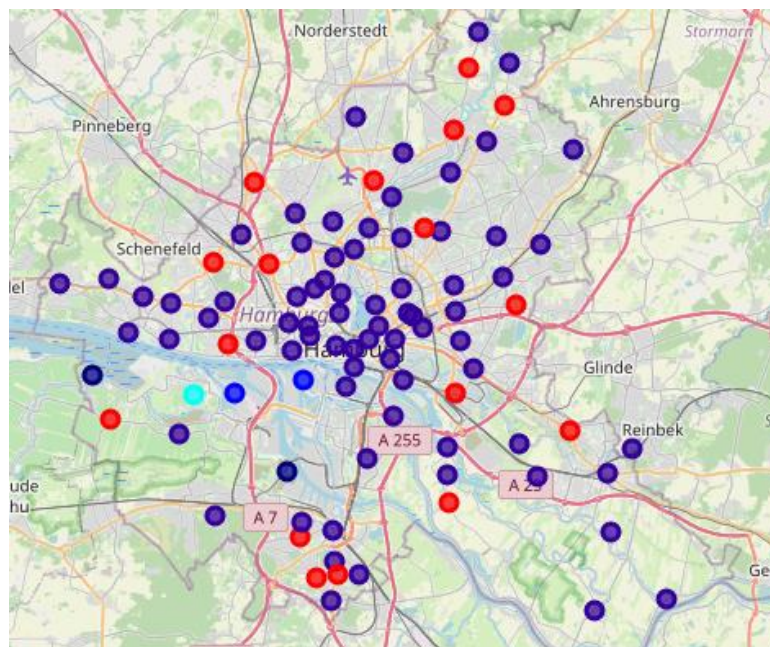


Figure 1: Clustered neighborhoods of Hamburg. Red: most relevant neighborhoods for the branch

Regarding this clustering process, neighborhoods close to the center seem to be less relevant for the new branch of the restaurant as they share less venues. Nevertheless, there are still 17 neighborhoods that are worth considering. To narrow these neighborhoods down another clustering process is done.

3.3 Refinement of clustering

The relevant cluster described in 3.2. and shown in figure 1 is used as a dataset that is split into 5 clusters one more time. The methods in this clustering step are copied from the previous one, but the used data only includes the 17 neighborhoods in Hamburg as well as the restaurant's neighborhood in Berlin. After splitting this dataset, we see that we are more likely to find similar neighborhoods in the north of Hamburg (figure 2, red dots). Considering this second clustering step, there are only six neighborhoods left, that are of interest for the stakeholder.

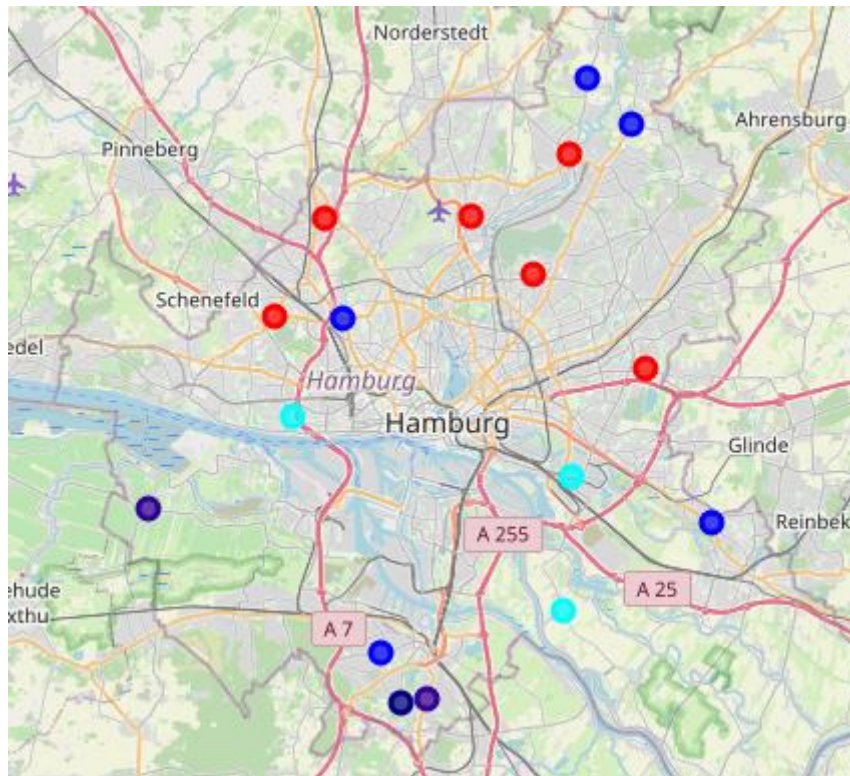


Figure 2: Further clustering of relevant neighborhoods in Hamburg. Red: most relevant neighborhoods for the branch

3.3 Demographical and economic comparison

With two sets of clustering, we found six neighborhoods in Hamburg that are comparable with the restaurant's neighborhood in Berlin. Regarding the clients they should be a good pick. The selection of a final neighborhood will therefore be made by further aspects than the comparability of the local venues.

The following diagram, figure 3, takes two more aspects into account: population density and rental prices. A high population density correlates with a potentially high number of customers, as there are more people nearby. The rent on the other hand is an economic factor that influences the costs and therefore the profit of the new branch.

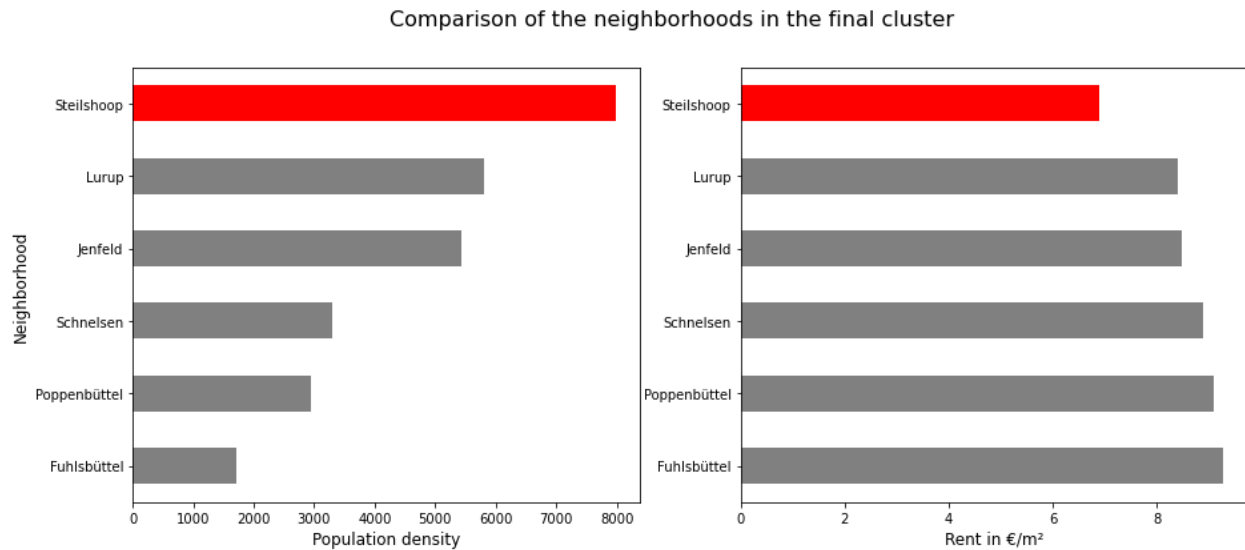


Figure 3: left: Population density of the neighborhoods in the final cluster, right: average rent per square meter of the neighborhoods in the final cluster

As shown in figure 3, there is one neighborhood with beneficial values for both factors. "Steilshoop" has the highest population density and at the same time the rental prices are lower than in the other relevant neighborhoods.

4. conclusion

The purpose of this project was to find a well-suited location in Hamburg for a new branch of a restaurant in Berlin. As the stakeholder owns a restaurant in Berlin and wants to keep a comparable business plan and clients, the goal is to find similar locations in Hamburg. The analysis shows, that a neighborhood outside of the city center is preferred and can be narrowed down to six neighborhoods in the north of Hamburg.

The final recommendation, based on the shown data, is to place the new branch in "Steilshoop". It is comparable with the restaurant's neighborhood in Berlin and has a high population density and therefore a large number of possible customers. Furthermore, the rental prices are lower than the prices in other comparable neighborhoods.