**Project report – “Finding a good neighborhood for opening a bar in Cologne, Germany”**

1. **Introduction**

The idea of my project is to collect data of the city of Cologne in Germany in order to give advice to people, that want to start a business in cologne and want to find an appropriate neighborhood based on specific criteria.

As an example I want to give advice to a client, who wants to open a bar in Cologne, Germany. The target group of the bar is younger people (about age 18 to 35). Therefore he plans to open the bar in an area, where a lot of young people live. Furthermore it should be an area where are yet not too many bars in order to avoid strong competition. I want to find a neighborhood in Cologne, that fits this criteria the most.

Here for I want to build a data frame that contains the top 10 neighborhoods in terms of the highest proportion of people in the targeted age-group. Within these top-10 neighborhoods I want to look for neighborhoods that do not have a lot of bars listed and at the same time have a high population-density.

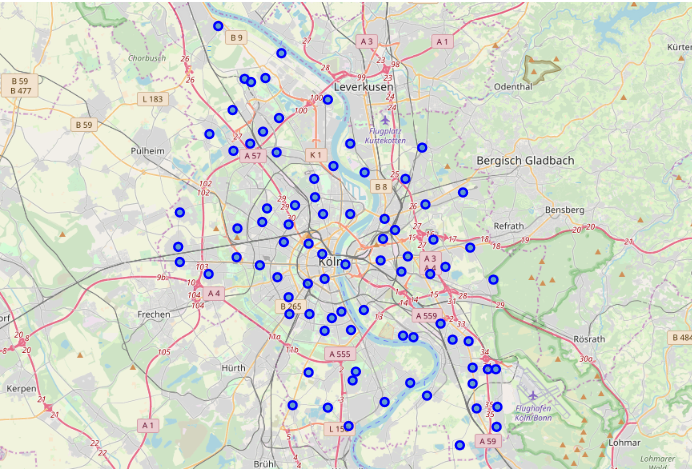
1. **Data Description**

**2.1. Data source**

The data I used for this project I downloaded from the webpage <https://www.offenedaten-koeln.de>. Here I found two tables, that are interesting for my research question. One table contains information about the population in each neighborhood by age (stadtteile\_alter14.xlsx) and one table contains information about the density of the population in each neighborhood (stadtteile\_dichte14.xlsx).

In total there are 86 neighborhoods in cologne. Three neighborhoods were excludes from the dataset, because of missing data for “citizens per square kilometers”. All three were small neighborhoods with only few young people and yet would not have been interesting for the analysis. The following map shows the remaining 83 neighborhoods in cologne, which were taken into the analysis.

Figure 1: Map of Cologne



1. **Methodology**

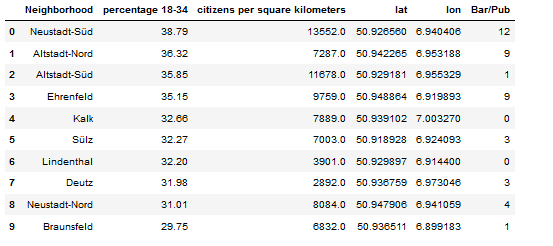
To get the longitude and latitude for each neighborhood in cologne I used “Nominatim” as a tool.

I selected the interesting columns from the above described dataframes and combined them to get my final database. As a last step I sorted the data by “Percentage of age group 18-34” and kept only the ten neighborhoods with the highest proportion of the interested age group.

Finally I found the nearby venues of Cologne with the Foursquare API. I chose a radius of 600 meters for each neighborhood, since increasing the radius led to double hits for found venues.

There were 1225 venues for all neighborhoods in cologne. 67 of the venues were listed as bars or Pubs. The number of bars/Pubs for each neighborhood was calculated and added to the dataframe (Table 1)

Table 1: Citizens per square kilometers, latitude and longitude, number of bars for each the top ten neighborhoods with the highest percentage of age group 18-34.



1. **Results**

In the dataframe we can see that it would be another method

1. **Discussion**

|  |  |
| --- | --- |
|  | Since the foursquare API does not provide information about the neighborhood each venue belongs to, I used an area of 600 meters of each neighborhood to search for the bars that are close by. With this method we cannot be sure to fin all the bars, tha are located in that area, especially in the bigger neighborhoods. Increaing the radius would on the other hand lead to double hits. To calculate the exact amount af bars per neighborhood it would be necessary to use a different API. But the collected data can be seen as a good estimation for the amount of bars in each area.  An object containing none, some, or all of address (street address), crossStreet, city, state, postalCode, country, lat, lng, and distance. All fields are strings, except for lat, lng, and distance. Distance is measured in meters. Some venues have their locations intentionally hidden for privacy reasons (such as private residences). If this is the case, the parameter isFuzzed will be set to true, and the lat/lng parameters will have reduced precision. |

Conclusion section where you conclude the report. 3. Your choice of a presentation or blogpost. (**10 marks**)

**Zweite idee:**

* Clustern von Stadtteilen in Köln und Berlin
* Zuordnen des Eigenen Stadtteils in Köln zu einer Kategorie
* Finden eines ähnliches Stadtteiles in Berlin im jeweiligen Cluster