

Capstone Project

The Battle of Neighborhoods in Cologne: Kölsch

1. Introduction and Problem

Cologne is the fourth-largest city in Germany and the largest in North Rhine-Westphalia, in many of the postals you can recognize it because of its remarkable skyline with the Cologne Cathedral as the main character. I am not exaggerating when I say that Cologne is one of the most fascinating cities I have the opportunity to visit and live in, so I can assure you that Cologne has a lot to offer, like its beautiful Parks, the Chocolate Museum where you can taste some of the best chocolates, the renowned Eau de Cologne perfume and all the many museums, art galleries and historic spots, but the most remarkable about Cologne is their joyful people, not for nothing Cologne is considered the Germany's capital of carnivals, and a Carnival will not be a Carnival without one of the treasures of Cologne, its famous Kölsch beer.

As one of the emblems of the city it is a must that you have to try a Kölsch and for that there are several places to consider apart from the restaurants, so for the travelers (who in some cases have few hours) it is important to know where are the places they can go in order to enjoy the experience of having a Kölsch, especially if they are interested in a non-tourist recommendation.

In order to help some tourists with their decision of having a Kölsch and find the appropriate neighborhood to go depending on their preferences we are going to explore the different bar categories in the city to observe where people meet to have a beer using Foursquare.

2. Data Requirements

The final dataset consists of two parts:

- Cologne data regarding to their districts (boroughs) and neighborhoods with their latitude and longitude.

https://en.wikipedia.org/wiki/Districts_of_Cologne

Although we can find information of interest about the districts of the city we are only interested in the main table that contains the relation between the boroughs and the neighborhoods.

- **Bars and German Restaurants in Cologne:**
The Foursquare API will supply us with this information through all venues in the city and later we can get only the information we have interest on.

3. Methodology

3.1 Data Preparation

3.1.1 Getting Cologne Boroughs and Neighborhoods information

The first step is to extract the table regarding of the information about the City districts and city parts, as you can see this table contains a lo of information that we will not use.

Map	Coat	City district		City parts	Area	Population1	Pop. density	District Councils	Town Hall
0	NaN	NaN	District 1 Köln-Innenstadt	Altstadt-Nord, Altstadt-Süd, Deutz, Neustadt-N...	16.4 km²	127.033	7.746/km²	Bezirksamt Innenstadt Brückenstraße 19, D-50...	NaN
1	NaN	NaN	District 2 Köln-Rodenkirchen	Bayenthal, Godorf, Hahnwald, Immendorf, Marien...	54.6 km²	100.936	1.850/km²	Bezirksamt Rodenkirchen Hauptstraße 85, D-5099...	NaN
2	NaN	NaN	District 3 Köln-Lindenthal	Braunsfeld, Junkersdorf, Klettenberg, Lindenth...	41.6 km²	137.552	3.308/km²	Bezirksamt Lindenthal Aachener Straße 220, 509...	NaN
3	NaN	NaN	District 4 Köln-Ehrenfeld	Bickendorf, Bocklemünd/Mengenich, Ehrenfeld, N...	23.8 km²	103.621	4.348/km²	Bezirksamt Ehrenfeld Venloer Straße 419 – 421,...	NaN
4	NaN	NaN	District 5 Köln-Nippes	Bilderstöckchen, Longerich, Mauenheim, Niehl, ...	31.8 km²	110.092	3.462/km²	Bezirksamt NippesNeusser Straße 450, D-50733 Köln	NaN

Before the last manipulating step we obtain the data-frame showed below, this is was an important step in the scrapping of the data because we needed to split the information contain in each row of the Neighborhoods column.

	Borough	Neighborhoods
0	Innenstadt	Altstadt-Nord, Altstadt-Süd, Deutz, Neustadt-N...
1	Rodenkirchen	Bayenthal, Godorf, Hahnwald, Immendorf, Marien...
2	Lindenthal	Braunsfeld, Junkersdorf, Klettenberg, Lindenth...
3	Ehrenfeld	Bickendorf, Bocklemünd/Mengenich, Ehrenfeld, N...
4	Nippes	Bilderstöckchen, Longerich, Mauenheim, Niehl, ...
5	Chorweiler	Blumenberg, Chorweiler, Esch/Auweiler, Fühling...
6	Porz	Eil, Elsdorf, Ensen, Finkenber, Gremberghoven...
7	Kalk	Brück, Höhenberg, Humboldt/Gremberg, Kalk, Mer...
8	Mülheim	Buchforst, Buchheim, Dellbrück, Dünnwald, Flit...

After some final manipulation we obtain the below data-frame, here is already a relation between the boroughs (districts) and their neighborhoods.

	Borough	Neighborhoods
0	Innenstadt	Altstadt-Nord
1	Innenstadt	Altstadt-Süd
2	Innenstadt	Deutz
3	Innenstadt	Neustadt-Nord
4	Innenstadt	Neustadt-Süd
5	Rodenkirchen	Bayenthal
6	Rodenkirchen	Godorf
7	Rodenkirchen	Hahnwald
8	Rodenkirchen	Immendorf
9	Rodenkirchen	Marienburg

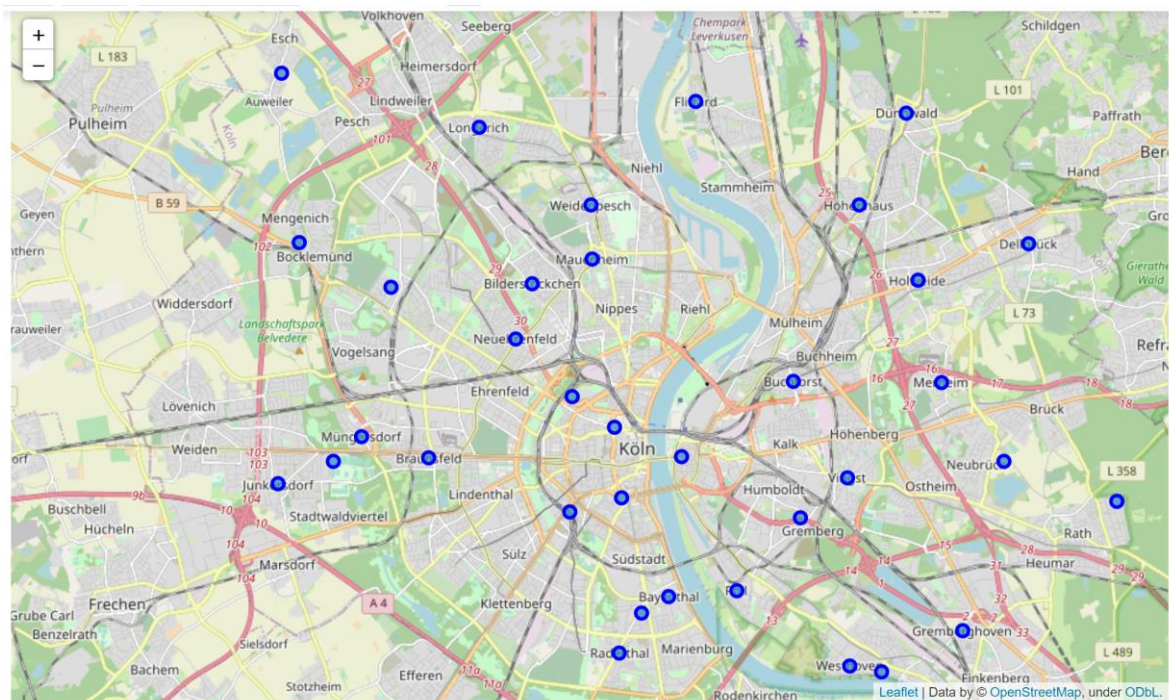
3.1.2 Getting coordinates of each neighborhood

After getting the data-frame with the boroughs and neighborhoods we can assign each one its latitude and longitude with the use of the Geopy Client.

```
cologne_coor = pd.DataFrame(columns = ['Latitude','Longitude'])
geolocator = Nominatim(user_agent="Cologne")
for row in df['Neighborhoods']:
    location = geolocator.geocode(row)
    cologne_coor = cologne_coor.append({'Latitude':location.latitude, 'Longitude':location.longitude}, ignore_index=True)
df = df.join(cologne_coor)
df
```

	Borough	Neighborhoods	Latitude	Longitude
0	Innenstadt	Altstadt-Nord	50.942265	6.953188
1	Innenstadt	Altstadt-Süd	50.929181	6.955329
2	Innenstadt	Deutz	50.936759	6.973046
3	Innenstadt	Neustadt-Nord	50.947906	6.941059
4	Innenstadt	Neustadt-Süd	50.926560	6.940406
5	Rodenkirchen	Bayenthal	50.911058	6.969238
6	Rodenkirchen	Godorf	50.849154	6.975350
7	Rodenkirchen	Hahnwald	49.662407	9.265221
8	Rodenkirchen	Immendorf	48.646985	16.127598
9	Rodenkirchen	Marienburg	54.035936	19.034882
10	Rodenkirchen	Meschenich	50.860585	6.927259
11	Rodenkirchen	Raderberg	50.907988	6.961169
12	Rodenkirchen	Raderthal	50.900683	6.954855
13	Rodenkirchen	Rodenkirchen	50.878750	6.981804

Finally in order to verify this information we used folium. Here we can see the center location of each neighborhood with remarked by the blue bubble.



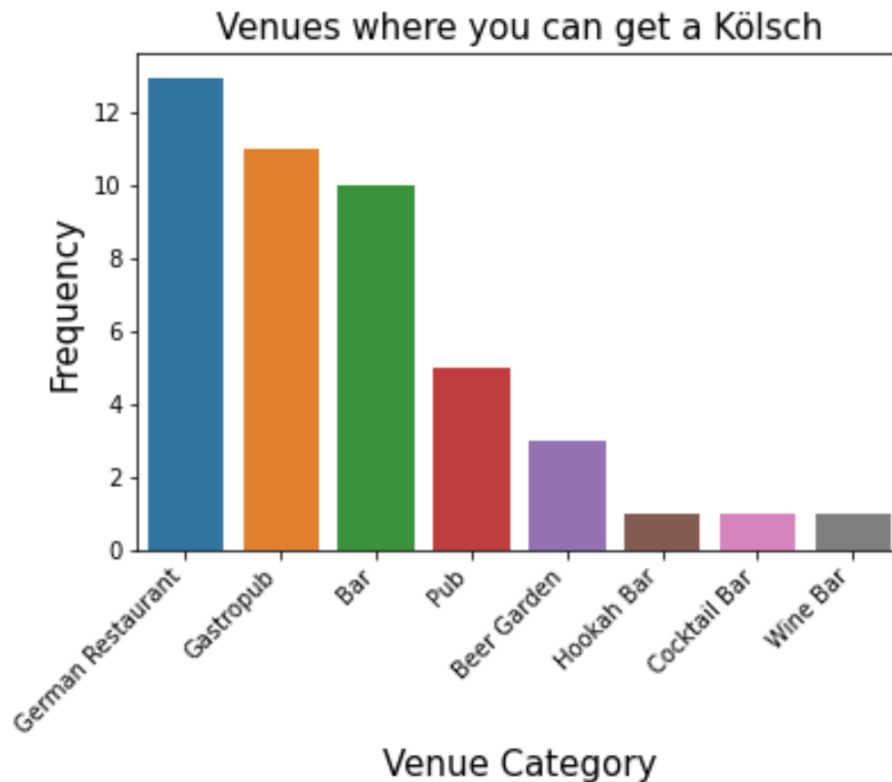
3.2 Exploration of the Cologne Neighborhoods and Data

3.2.1 Venues collection using Foursquare

With the use of Foursquare API we can get the Venues of each neighborhood, the data frame below is an example of the first dataset that contains all venues of Cologne but as we are only interested in venues where you can try a Kölsch we will manipulate it some more.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Altstadt-Nord	50.942265	6.953188	Excelsior Hotel Ernst	50.942222	6.956282	Hotel
1	Altstadt-Nord	50.942265	6.953188	KölnTourismus	50.941554	6.956278	Tourist Information Center
2	Altstadt-Nord	50.942265	6.953188	Kuchi Mami	50.941390	6.954036	Asian Restaurant
3	Altstadt-Nord	50.942265	6.953188	Punto Pasta	50.942301	6.955168	Italian Restaurant
4	Altstadt-Nord	50.942265	6.953188	Enoteca La Cantina	50.944522	6.955899	Gourmet Shop

As we explore the data, we identify that the venues where tourists will enjoy the best the experience of a Kölsch are the ones in the graph, where we can observe that the venue in first place are the German Restaurants where you definitely will find a Kölsch but also other specialties of the region, surprisingly in second place are the gastropubs that is not a German concept but where you will enjoy too the food of the city without losing the experience of a pub. Finally in fourth and fifth place are our points of most interest, the bars and pubs. In the results chapter we will see how these venues distribute on the city and where can we go to experience each of one, if they are near the city or in the surroundings and other interesting insights. Our final dataframe will only have the information about these venues.



On the other hand we also check how many venues per neighborhood we returned, we can see this information in the data frame below. This information will help us later in knowing where are the majority of the venues and how they are distributed.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Altstadt-Süd	1	1	1	1	1	1
Dellbrück	2	2	2	2	2	2
Deutz	2	2	2	2	2	2
Finkenberg	2	2	2	2	2	2
Junkersdorf	1	1	1	1	1	1
Libur	1	1	1	1	1	1
Lindenthal	1	1	1	1	1	1
Longerich	1	1	1	1	1	1
Mauenheim	1	1	1	1	1	1
Merheim	3	3	3	3	3	3
Mülheim	4	4	4	4	4	4
Neuehrenfeld	2	2	2	2	2	2
Neustadt-Nord	3	3	3	3	3	3
Neustadt-Süd	6	6	6	6	6	6
Pesch	1	1	1	1	1	1
Poll	1	1	1	1	1	1
Raderberg	1	1	1	1	1	1
Wahnheide	1	1	1	1	1	1
Weiden	4	4	4	4	4	4
Weidenpesch	2	2	2	2	2	2

3.3 Anlyze each Neighborhood

In order to categorize the venues we create a one hot encoding data frame.

	Neighborhood	Bar	Beer Garden	Cocktail Bar	Gastropub	German Restaurant	Hookah Bar	Pub	Wine Bar
1	Neuehrenfeld	0	1	0	0	0	0	0	0
2	Weidenpesch	0	1	0	0	0	0	0	0
3	Poll	0	1	0	0	0	0	0	0
4	Neustadt-Nord	0	0	0	0	0	0	1	0
5	Weiden	0	0	0	0	0	0	1	0

The next step is to group the venues by neighborhood by taking the mean of the frequency of ocurrence of each category. This will help us to know which are the mosto common venues per neighborhood. Like the example below.

```

---- Merheim----
              venue  freq
0      Gastropub  0.67
1  German Restaurant  0.33
2              Bar  0.00
3      Beer Garden  0.00
4      Cocktail Bar  0.00

```

The final result of this process will be a dataframe with the information of the most common venue in each neighborhood, since we are only analyzing eight venue categories we can put them all in the dataframe.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	Altstadt-Süd	Bar	Wine Bar	Pub	Hookah Bar	German Restaurant	Gastropub	Cocktail Bar	Beer Garden
1	Dellbrück	German Restaurant	Gastropub	Wine Bar	Pub	Hookah Bar	Cocktail Bar	Beer Garden	Bar
2	Deutz	German Restaurant	Gastropub	Wine Bar	Pub	Hookah Bar	Cocktail Bar	Beer Garden	Bar
3	Finkenbergr	Pub	Wine Bar	Hookah Bar	German Restaurant	Gastropub	Cocktail Bar	Beer Garden	Bar
4	Junkersdorf	German Restaurant	Wine Bar	Pub	Hookah Bar	Gastropub	Cocktail Bar	Beer Garden	Bar

3.4 Cluster Neighborhoods

Finally we cluster the neighborhoods into 4 clusters, so that we can know how the venues are distributed and where we can find most of them depending on our preferences.

```
# set number of clusters
kclusters = 4

cologne_grouped_clustering = cologne_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(cologne_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
array([1, 2, 2, 1, 0, 2, 2, 2, 0, 2])
```

We join our previous dataframe with the common venues and the resulting clusters for each neighborhood. We obtain the next dataframe.

	index	Borough	Neighborhoods	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	1	Innenstadt	Altstadt-Süd	50.929181	6.955329	1.0	Bar	Wine Bar	Pub	Hookah Bar	German Restaurant	Gastropub	Cocktail Bar	Beer Garden
1	2	Innenstadt	Deutz	50.936759	6.973046	2.0	German Restaurant	Gastropub	Wine Bar	Pub	Hookah Bar	Cocktail Bar	Beer Garden	Bar
2	3	Innenstadt	Neustadt-Nord	50.947906	6.941059	1.0	Pub	Gastropub	Bar	Wine Bar	Hookah Bar	German Restaurant	Cocktail Bar	Beer Garden
3	4	Innenstadt	Neustadt-Süd	50.926560	6.940406	1.0	Bar	Gastropub	Cocktail Bar	Wine Bar	Pub	Hookah Bar	German Restaurant	Beer Garden
4	5	Rodenkirchen	Bayenthal	50.911058	6.969238	0.0	German Restaurant	Wine Bar	Pub	Hookah Bar	Gastropub	Cocktail Bar	Beer Garden	Bar
5	11	Rodenkirchen	Raderberg	50.907988	6.961169	1.0	Hookah Bar	Wine Bar	Pub	German Restaurant	Gastropub	Cocktail Bar	Beer Garden	Bar
6	18	Lindenthal	Braunsfeld	50.936511	6.899183	0.0	German Restaurant	Wine Bar	Pub	Hookah Bar	Gastropub	Cocktail Bar	Beer Garden	Bar
7	19	Lindenthal	Junkersdorf	50.931753	6.855169	0.0	German Restaurant	Wine Bar	Pub	Hookah Bar	Gastropub	Cocktail Bar	Beer Garden	Bar
8	21	Lindenthal	Lindenthal	50.935935	6.871246	2.0	Gastropub	Wine Bar	Pub	Hookah Bar	German Restaurant	Cocktail Bar	Beer Garden	Bar
9	25	Lindenthal	Weiden	50.935679	6.832909	1.0	Bar	Wine Bar	Pub	Hookah Bar	German Restaurant	Gastropub	Cocktail Bar	Beer Garden
0	30	Ehrenfeld	Neuehrenfeld	50.958512	6.924394	3.0	Gastropub	Beer Garden	Wine Bar	Pub	Hookah Bar	German Restaurant	Cocktail Bar	Bar

	Borough	Neighborhoods	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
4	Rodenkirchen	Bayenthal	50.911058	6.969238	0.0	German Restaurant	Wine Bar	Pub	Hookah Bar	Gastropub	Cocktail Bar	Beer Garden	Bar
6	Lindenthal	Braunsfeld	50.936511	6.899183	0.0	German Restaurant	Wine Bar	Pub	Hookah Bar	Gastropub	Cocktail Bar	Beer Garden	Bar
7	Lindenthal	Junkersdorf	50.931753	6.855169	0.0	German Restaurant	Wine Bar	Pub	Hookah Bar	Gastropub	Cocktail Bar	Beer Garden	Bar
12	Nippes	Mauenheim	50.973039	6.946731	0.0	German Restaurant	Wine Bar	Pub	Hookah Bar	Gastropub	Cocktail Bar	Beer Garden	Bar
18	Porz	Wahnheide	50.859911	7.103522	0.0	German Restaurant	Wine Bar	Pub	Hookah Bar	Gastropub	Cocktail Bar	Beer Garden	Bar
19	Porz	Westhoven	50.898321	7.022011	0.0	German Restaurant	Wine Bar	Pub	Hookah Bar	Gastropub	Cocktail Bar	Beer Garden	Bar
20	Porz	Zündorf	50.865926	7.042982	0.0	German Restaurant	Wine Bar	Pub	Hookah Bar	Gastropub	Cocktail Bar	Beer Garden	Bar

	Borough	Neighborhoods	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	Innenstadt	Altstadt-Süd	50.929181	6.955329	1.0	Bar	Wine Bar	Pub	Hookah Bar	German Restaurant	Gastropub	Cocktail Bar	Beer Garden
2	Innenstadt	Neustadt-Nord	50.947906	6.941059	1.0	Pub	Gastropub	Bar	Wine Bar	Hookah Bar	German Restaurant	Cocktail Bar	Beer Garden
3	Innenstadt	Neustadt-Süd	50.926560	6.940406	1.0	Bar	Gastropub	Cocktail Bar	Wine Bar	Pub	Hookah Bar	German Restaurant	Beer Garden
5	Rodenkirchen	Raderberg	50.907988	6.961169	1.0	Hookah Bar	Wine Bar	Pub	German Restaurant	Gastropub	Cocktail Bar	Beer Garden	Bar
9	Lindenthal	Weiden	50.935679	6.832909	1.0	Bar	Wine Bar	Pub	Hookah Bar	German Restaurant	Gastropub	Cocktail Bar	Beer Garden
15	Porz	Finkenberg	50.896965	7.062007	1.0	Pub	Wine Bar	Hookah Bar	German Restaurant	Gastropub	Cocktail Bar	Beer Garden	Bar
23	Mülheim	Mülheim	50.961858	7.004074	1.0	Bar	Pub	German Restaurant	Wine Bar	Hookah Bar	Gastropub	Cocktail Bar	Beer Garden

3.4.3 Cluster 3 – Navy Blue

	Borough	Neighborhoods	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
1	Innenstadt	Deutz	50.936759	6.973046	2.0	German Restaurant	Gastropub	Wine Bar	Pub	Hookah Bar	Cocktail Bar	Beer Garden	Bar
8	Lindenthal	Lindenthal	50.935935	6.871246	2.0	Gastropub	Wine Bar	Pub	Hookah Bar	German Restaurant	Cocktail Bar	Beer Garden	Bar
11	Nippes	Longerich	50.997324	6.913713	2.0	Gastropub	Wine Bar	Pub	Hookah Bar	German Restaurant	Cocktail Bar	Beer Garden	Bar
14	Chorweiler	Pesch	51.001078	6.877718	2.0	Gastropub	Wine Bar	Pub	Hookah Bar	German Restaurant	Cocktail Bar	Beer Garden	Bar
16	Porz	Libur	50.838733	7.071327	2.0	Gastropub	Wine Bar	Pub	Hookah Bar	German Restaurant	Cocktail Bar	Beer Garden	Bar
21	Kalk	Merheim	50.950419	7.048598	2.0	Gastropub	German Restaurant	Wine Bar	Pub	Hookah Bar	Cocktail Bar	Beer Garden	Bar
22	Mülheim	Delbrück	50.975921	7.074106	2.0	German Restaurant	Gastropub	Wine Bar	Pub	Hookah Bar	Cocktail Bar	Beer Garden	Bar

3.4.4 Cluste 4 – Green

	Borough	Neighborhoods	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
10	Ehrenfeld	Neuehrenfeld	50.958512	6.924394	3.0	Gastropub	Beer Garden	Wine Bar	Pub	Hookah Bar	German Restaurant	Cocktail Bar	Bar
13	Nippes	Weidenpesch	50.983000	6.946341	3.0	German Restaurant	Beer Garden	Wine Bar	Pub	Hookah Bar	Gastropub	Cocktail Bar	Bar
17	Porz	Poll	50.912101	6.988889	3.0	Beer Garden	Wine Bar	Pub	Hookah Bar	German Restaurant	Gastropub	Cocktail Bar	Bar

4. Results and Discussion

We can get some interesting insights from the cluster:

- Cluster 1: Here we will find the majority of the german restaurants, surprisingly al lof them are in the surroundings of the city center, so we can asume the people who visit these venues are the locals ones who infering from the second venue are people looking for a more formal kind of place to enjoy their food and drinks.
- Cluster 2: We can find here our point of intereset because the mainly venues that we can find here are the bars and pubs. Something really interesting is that most of them are in the neighborhood of Neustadt-Süd where is the Zülpicher Strasse, a street and its surroundings like the University of Köln with a lot of offer in terms of bars, restaurants and clubs. With this point we can confirm that the clustiring is working in a good way.
- Cluster 3: Most of the gastropubs are in this cluster. By the distribution we find them most in teh surroundings of the city but some also in Deutz near from the TH Köln. The gastropubs are places where you can also enjoy food but without losing the experience of a pub but in a more formal way.
- Cluster 4: The tourist will fin the beer gardens in this cluster, this is a traditional german concept, as an example is the famous Oktober Fest, of course that is the extrem example, but in general the idea is of having a beer in an open space, like a garden.

5. Conclusion

The information is one of the most valuable resources nowadays, as the example of the project we can get a lot of guidance in which decision to take depending of what we want and solve a problem. Imagine using all this information in order to solve most important problems, or as we have seen in recently months with the pandemic, that a lot of information has been gathered in order to know the development of it but also in terms of knowing what to do. As you can see this is an extreme example but what about making better decisions in our jobs from the information of the processes, this will be a great opportunity to launch our career and help others and our companies in achieving the objectives.