Rivalry on the Rhine - The Battle of the Neighborhoods

Y. K.

(December 2019)

1. Introduction

Cologne (or Köln in German) and Düsseldorf are two major cities in western Germany and lie only about 40 kilometers apart on the Rhine. Despite of similarities in culture, social geography and many other aspects, a competitive relationship between the two cities, also called the rivalry on the Rhine, has been evolved during last decades, which is mainly from historical and economic facts and celebrated folklorically on a sporting and cultural level as "enmity". While the larger Cologne with a population of more than 1 million was developed from a Roman colony and later free imperial city, the modern residence city of Düsseldorf that emerged from a small medieval settlement is today the capital of the state of North Rhine-Westphalia.

By analyzing publicly available data with modern data science techniques, the primary objective of current project is to investigate/explore how similar and dissimilar both cities are. The results may shed lights on distinctions between the two cities in diverse social and cultural aspects, and help

- tourists who would like to spend a few days in one of the two cities or
- people/groups who want to choose a place to develop business

make their first decision.

2. Data Exploration

In order to identify similarities or distinctions between Cologne and Düsseldorf, it is desired to investigate and compare boroughs and neighborhoods in both cities.

A list of boroughs and neighborhoods containing their names and geospatial coordinates can be prepared by scraping wiki and related webpages, and then converted to Pandas dataframes. Figure 1 shows the first few rows of data frame obtained for the city Cologne with 9 boroughs (Stadtbezirke) and 86 neighborhoods (Stadtteile).

1

	Neighborhood	Borough	Area	Population	Density	Latitude	Longitude
0	Altstadt-Süd	Innenstadt	2.36	27071	11469	50.929181	6.955329
1	Neustadt-Süd	Innenstadt	3.50	37719	13343	50.926560	6.940406
2	Altstadt-Nord	Innenstadt	2.46	17960	7301	50.942265	6.953188
3	Neustadt-Nord	Innenstadt	2.83	28720	8213	50.947906	6.941059
4	Deutz	Innenstadt	5.24	15563	2968	50.936759	6.973046
...	...	...	...	...	...	...	...

Fig. 1 Dara frame containing Cologne neighborhoods and their geospatial coordinates

The Foursquare API will be then applied to explore boroughs and neighborhoods in both cities and segment them. A database/dataframe accumulated by the Foursquare explore function is composed of the most common venue categories in each neighborhood. Each data record corresponds to a venue containing its name, unique ID, location and category. A few data records from the data frame containing Düsseldorf venue data are shown in Figure 2.

	Neighborhood	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Altstadt	51.225912	6.773567	Casita Mexicana	51.226676	6.775478	Mexican Restaurant
1	Altstadt	51.225912	6.773567	Rösterei VIER	51.225940	6.772294	Coffee Shop
2	Altstadt	51.225912	6.773567	Elephant Bar	51.226851	6.772636	Cocktail Bar
3	Altstadt	51.225912	6.773567	Bar Chérie	51.226886	6.772424	Bar
4	Altstadt	51.225912	6.773567	Rösterei VIER	51.224536	6.773703	Coffee Shop
...	...	...	...	...	...	...	...
886	Wersten	51.189628	6.817777	Bäckerei Café Oebel	51.186723	6.821452	Bakery
887	Wittlaer	51.319439	6.742058	Gasthaus Peters	51.321216	6.739248	German Restaurant
888	Wittlaer	51.319439	6.742058	Bäcker Wolff	51.320635	6.744788	Bakery
889	Wittlaer	51.319439	6.742058	Im Jäger	51.320626	6.744812	German Restaurant
890	Wittlaer	51.319439	6.742058	U Wittlaer	51.318344	6.745337	Metro Station

Fig. 2 A few data records from the Düsseldorf venue data frame

Before any analysis can be performed by data science algorithms, the collected raw data must be cleaned and prepared.

3. Methodology

Various tools and/or libraries can be used to understand and analyze the data. For example, the neighborhoods and their distributions in both cities can be visualized using the Folium library. The word-cloud plots of venue categories in city neighborhoods can be

used to identify those categories that really stand out in each city, which might tell us intuitively how dissimilar the both cities are.
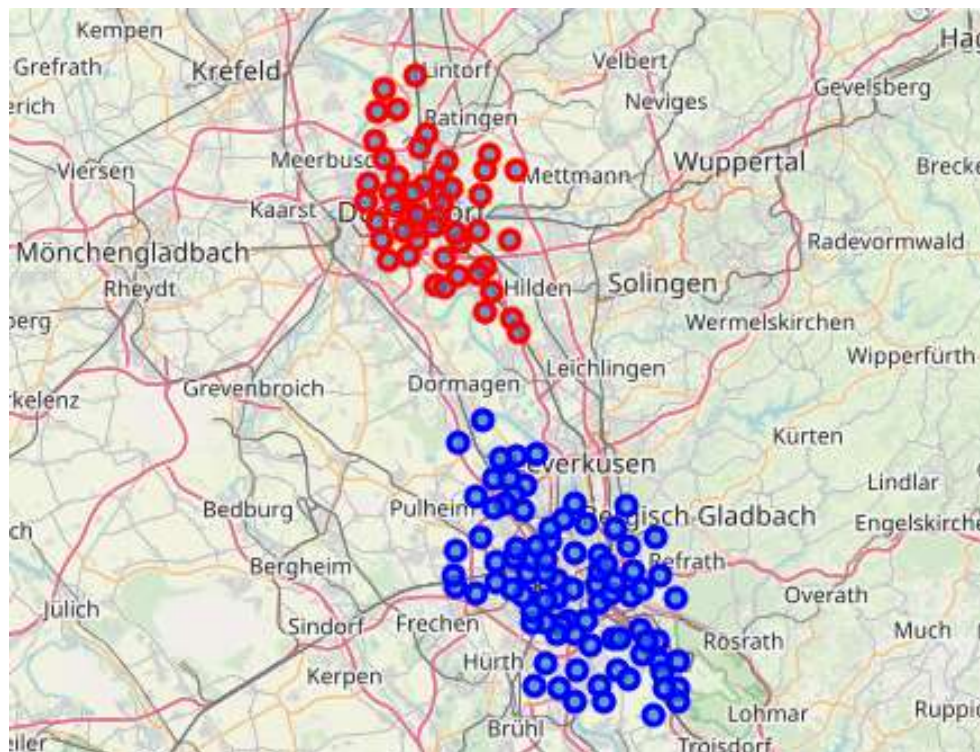
The k-means clustering algorithm is then applied to group the neighborhoods in each city into clusters, which will be compared between both cities for any similarity and distinction. Furthermore, the k-means clustering can be performed for the data set containing all the neighborhoods from both cities. The more the neighborhoods from both cities are assigned to the same groups, the more similar both cities are.

Finally, the neighborhood clusters will be visualized with the Folium library.

4. Comparison of neighborhoods between Cologne and Düsseldorf cities

4.1 Neighborhoods in Cologne and Düsseldorf

The city of Cologne is made up of 9 boroughs and 86 neighborhoods, which is the largest city in the state of North Rhine-Westphalia and the fourth largest city in Germany. Only lying ca. 30 km northern from Cologne, the capital city of Düsseldorf with 10 boroughs and

Fig. 3 Map of Cologne and Düsseldorf superimposed with their neighborhoods

50 neighborhoods is the second largest city in the state. Figure 3 shows the folium map of
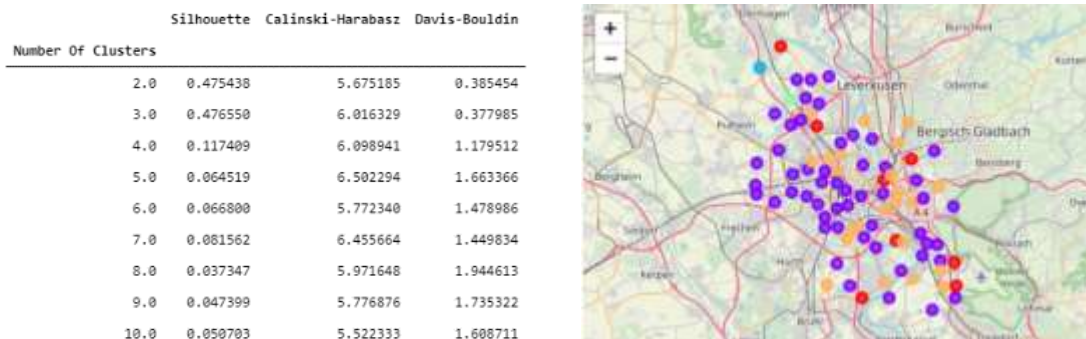
Cologne and Düsseldorf superimposed with their neighborhoods. The blue circles denote the neighborhoods of Cologne and the red ones the neighborhoods in Düsseldorf. A few basic information for both cities is summarized in Table 1.

Table 1. Basic information of Cologne and Düsseldorf

	Latitude	Longitude	Boroughs	Neighborhoods	Area (km2)	Population	Density (/km2)
Cologne	50.938837	6.970407	9	86	405.12	1019326	2516
Düsseldorf	51.225402	6.776314	10	50	217.20	591829	2724

4.2 Neighborhood comparison by most common venue categories

The venue categories in city neighborhoods, which were grouped by neighborhoods and sorted by frequency of occurrence, have been investigated to get an intuitive comparison of both cities. The mostly common venue categories in Cologne and Düsseldorf neighborhoods are illustrated by word-cloud plots as shown in Fig. 4. At a first glance, two

Fig. 4 Word-cloud plots showing venue categories in Cologne (*left*) and Düsseldorf (*right*) neighborhoods

word-cloud plots only show subtle difference. While the venue categories in both Cologne and Düsseldorf neighborhoods are dominated by restaurant, Café and supermarket, hotels and diverse shops also occupy similar weights among minority of categories in both cities.

Each dominant category of venues can be further studied to compare neighborhoods between Cologne and Düsseldorf. In the category of restaurant, for example, the sub-category of foreign restaurant can be analyzed by excluding German, Italian and non-categorized restaurants from venue data. Except dominant Italian restaurants, as shown in Fig. 5, Greek and Turkish restaurants are the mostly common foreign restaurants in Cologne, while the Asian restaurants like Japanese and Vietnamese dominate the foreign restaurants in Düsseldorf.

Fig. 5 Word-cloud plots showing most common foreign restaurants in Cologne (*left*) and Düsseldorf (*right*) neighborhoods

4.3 Neighborhood segmentation/clustering

The k-means clustering algorithm was applied to group the neighborhoods in Cologne and Düsseldorf into clusters. The one-hot prepared venue data have been fitted with various number of clusters from 2 to 10, and corresponding scores including Silhouette score, Calinski-Harabasz and Davis-Bouldin index were calculated to find an optimized number of clusters.

The k-means clustering evaluation results with the venue data for Cologne neighborhoods

Number Of Clusters	Silhouette	Calinski-Harabasz	Davis-Bouldin
2.0	0.475438	5.675185	0.385454
3.0	0.476550	6.016329	0.377985
4.0	0.117409	6.098941	1.179512
5.0	0.064519	6.502294	1.663366
6.0	0.066800	5.772340	1.478986
7.0	0.081562	6.455664	1.449834
8.0	0.037347	5.971648	1.944613
9.0	0.047399	5.776876	1.735322
10.0	0.050703	5.522333	1.608711

Fig. 6 Clustering neighborhoods of Cologne into groups. *Left*: Evaluation of k-means clustering algorithm. *Right*: neighborhoods groups by k-means clustering with 5 groups

are shown in the left panel of Fig. 6. The smaller Silhouette score and Davis-Bouldin index and larger Calinski-Harabasz index for smaller number of clusters suggest better clustering results can only be obtained with smaller number of clusters. In other words, the Cologne neighborhoods may not be well grouped by the venue category data. As shown in

the right panel of Fig. 6, the neighborhoods grouped by k-means clustering with 5 clusters (labeled with different colors on the folium map) are not well concentrated and separated.

Fig. 7 Clustering neighborhoods of Düsseldorf into groups. *Left*: Evaluation of k-means clustering algorithm. *Right*: neighborhoods groups by k-means clustering with 4 groups

Similarly, k-means clustering was applied to group the neighborhoods in Düsseldorf with the neighborhood-grouped venue data. In Fig. 7 are presented the clustering evaluations calculated for clustering with various number of clusters and neighborhood groups obtained by 4-group clustering. Compared to the neighborhoods in Cologne, the Düsseldorf neighborhoods are a little bit better concentrated but still not well separated by current venue data.

Furthermore, the Foursquare data collected for Cologne and Düsseldorf neighborhoods were merged into a new dataframe. And the k-means clustering has been performed for the new data set containing all the neighborhoods from both cities. The evaluation with various number of clusters and 9 neighborhood groups are presented in Fig. 8. Most of groups contain neighborhoods from both cities, suggesting that those neighborhoods are quite similar according to the venue data used in the fitting procedure.

Fig. 8 Clustering neighborhoods in both cities into groups. *Left*: Evaluation of k-means clustering algorithm. *Right*: neighborhoods groups by k-means clustering with 9 groups

5. Summary and conclusion

Based on the venue category data obtained by the Foursquare explore function, the neighborhoods in the cities of Cologne and Düsseldorf have been analyzed. The results of neighborhood segmentation indicated that the neighborhoods in Cologne and Düsseldorf are not well distinguished by their venue data.

In Summary, the city of Cologne and Düsseldorf are very similar, and there is no real rivalry in everyday life.