Introduction:

As a part of the IBM Data Science Professional Certificate, you will find in this post an overview of the final capstone project.

Germany is one of the western European country with a varied landscape of rivers, forests, mountain ranges and beaches in the northern part of the country. It is the most populous country in the current member state of the European union. Berlin and Munich are two of the top three largest cities in Germany and there is significant amount of attrition of people from one city to the other every year.

This project explores the neighborhoods in both Berlin and Munich on how similar and dissimilar they are based on the location data from Foursquare website. The aim of the project is to aid people who would like to move from one city to other and suggest where they should rent their new apartment in the respective city.

Data:

For this project, data from foursquare is used to analyze different neighborhoods in the above-mentioned cities Berlin and Munich. Venue data of these two cities will be analyzed to identify certain traits such as availability of schools, cafes, restaurants etc. to identify the neighborhoods.

Regularly spaced grid of locations centered around a defined city center is used in this analysis to define the neighborhoods in Berlin and Munich as the official neighborhoods in both cities are sized very differently.

The following data sources will be needed to generate the required information.

- Geopy Nominatim: to extract the data for city center. Bundestag building for Berlin and Marienplatz for Munich.
- With a fixed radius, the neighborhoods are generated algorithmically. Addresses of the above-mentioned city centers are obtained using geopy Nominatim reverse geocoding.
- Using the Foursquare API, number of venues and their category in every neighborhood are obtained.

Disclaimer:

The data that is used in this project for Germany might not be as extensive as it is compared to that of the U.S and the analysis therefore might not be comprehensive.

Methodology

Generate neighborhood dataset

To generate the dataset of the neighborhood locations, the city centers for Munich and Berlin were used as starting points for the purpose of the assignment. For Munich, Marienplatz and for Berlin the Bundestag building were used as the city center addresses. From these addresses the geocoordinates were derived using the geocode function from Nominatum.

The latitude and longitude coordinates were required to be transformed into UTM coordinates to enable a defining radius around the city center and a smaller radius for each neighborhood. For the cities this was set to 5 km and for the neighborhoods to 1.5 km. Looping over this radius in a hexagonal grid generated UTM coordinates for each neighborhood. These were transformed back into

geocoordinates resulting in a list of 39 neighborhoods for each city. To check whether the neighborhoods were generated correctly, folium was used to map their locations.
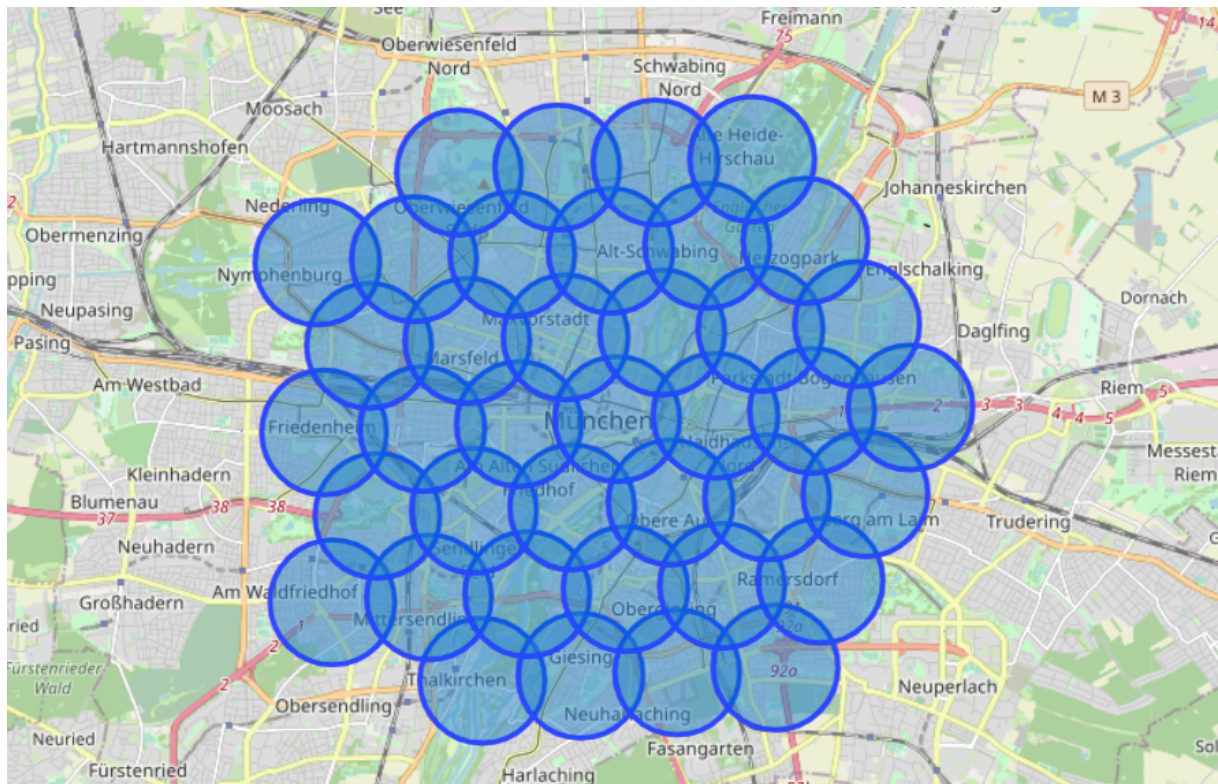
The Foursquare Places API uses addresses as search entries. The reverse function from Nominatim was used to generate neighborhood addresses from the latitude and longitude values. In addition, from the resulting addresses the borough names could be extracted. For Berlin this resulted in 06 boroughs vs. only one borough in Munich.

Cleaned up dataset includes the neighborhood address, latitude, longitude, borough and city.

The below mentioned screen shot shows the examples of the boroughs of Berlin and Munich

	Neighborhood	Latitude	Longitude	Borough	City
0	47, Hauptstraße, Rote Insel, Schöneberg, Tempe...	52.483193	13.348420	Tempelhof-Schöneberg	Berlin
1	Informationsort Schwerbelastungskörper, 100, G...	52.484143	13.371623	Tempelhof-Schöneberg	Berlin
2	P 3, Columbiadamm, Tempelhof, Tempelhof-Schöne...	52.483240	13.391364	Tempelhof-Schöneberg	Berlin
3	Freiluftkino Hasenheide, Columbiadamm, Neuköll...	52.483780	13.416432	Neukölln	Berlin
4	6, Wittelsbacherstraße, Wilmersdorf, Charlotte...	52.494681	13.314535	Charlottenburg-Wilmersdorf	Berlin

|	Neighborhood	Latitude	Longitude	Borough	City
0	Tierparkstraße (Isar-Eingang), Tierparkstraße,...	48.100620	11.551424	Bezirksteil Siebenbrunn	munich
1	90, Säbener Straße, Bezirksteil Giesing, Unter...	48.101930	11.571492	Bezirksteil Giesing	munich
2	22, Hohenschwangaustraße, Bezirksteil Obergies...	48.102484	11.591668	Bezirksteil Obergiesing	munich
3	107, Görzer Straße, Balanstraße-West, Bezirkst...	48.103028	11.611556	Bezirksteil Balanstraße-West	munich
4	11, Johann-Houis-Straße, Bezirksteil Mittersen...	48.111960	11.520543	Bezirksteil Mittersendling	munich

Get venue data from Foursquare

The Foursquare Places API provides venue recommendations using the "explore" functionality. It is used in this instance used to request the 100 most popular venues within 1.5 km of each neighborhood center. Along with the venue name, Foursquare provides many additional details on the venue such as the category (e.g. bakery or theater). Those neighborhoods for which less than 100 venues were found were excluded from the dataset. For each remaining neighborhood and borough the top 10 venue categories were identified. This was done using one hot encoding and calculation of the mean per category.

	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	9th Most Common Venue	10th Most Common Venue
0	0, 21, Arcisstraße, Bezirksteil Königsplatz, M...	Café	Plaza	Art Museum	Italian Restaurant	Ice Cream Shop	Steakhouse	Bar	Vietnamese Restaurant	Bakery	Coffee Shop
1	1, Burgunderstraße, Bezirksteil Am Luitpoldpar...	Park	Greek Restaurant	Drugstore	Italian Restaurant	Pizza Place	Bar	Bakery	Pool	Restaurant	Museum
2	101, Gartenstraße, Spandauer Vorstadt, Mitte, ...	Coffee Shop	Café	Hotel	Ice Cream Shop	Bookstore	Bakery	Park	Italian Restaurant	Playground	Theater
3	13, Hans-Otto-Straße, Bötzowviertel, Prenzlaue...	Café	Park	Italian Restaurant	Playground	Bakery	Coffee Shop	Vietnamese Restaurant	Indie Movie Theater	Ice Cream Shop	Pizza Place
4	13, Löwestraße, Hausburgviertel, Friedrichshai...	Café	Vegetarian / Vegan Restaurant	Coffee Shop	Pizza Place	Ice Cream Shop	Park	Middle Eastern Restaurant	Italian Restaurant	Cocktail Bar	Nightclub

For the following analysis the resulting tables were simplified, and geo-coordinates generated for each borough.

Cluster neighborhoods and boroughs

The neighborhoods and boroughs were clustered using the Kmeans algorithm from sklearn. Kmeans groups each observation (in this case each neighborhood or borough) into clusters minimizing the variance within each cluster. Important to note is that it converges to a local optimum, so each computation may lead to different results. A high number of clusters were specified (13 for the neighborhoods and 6 for the boroughs) with the aim of getting very specific clusters with an average of 4 neighborhoods or boroughs. The results for the neighborhood clustering will be discussed in more detail in the next section. The clustering of the boroughs, however, led to inconclusive results because they appear to be too large to be distinct from one another or base a recommendation on.
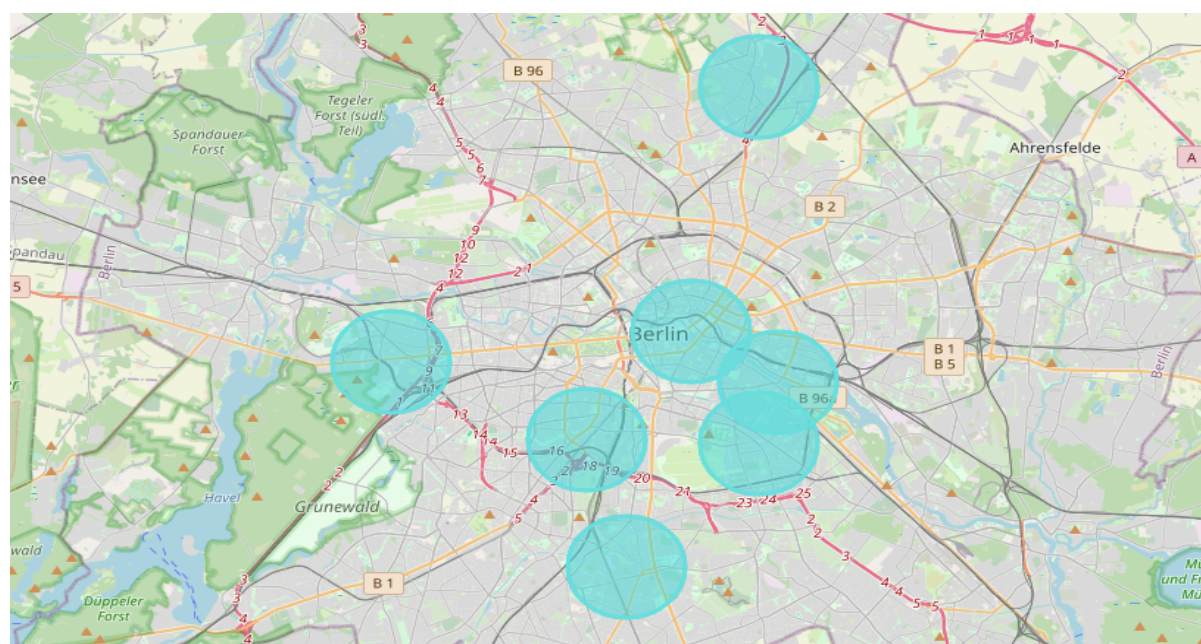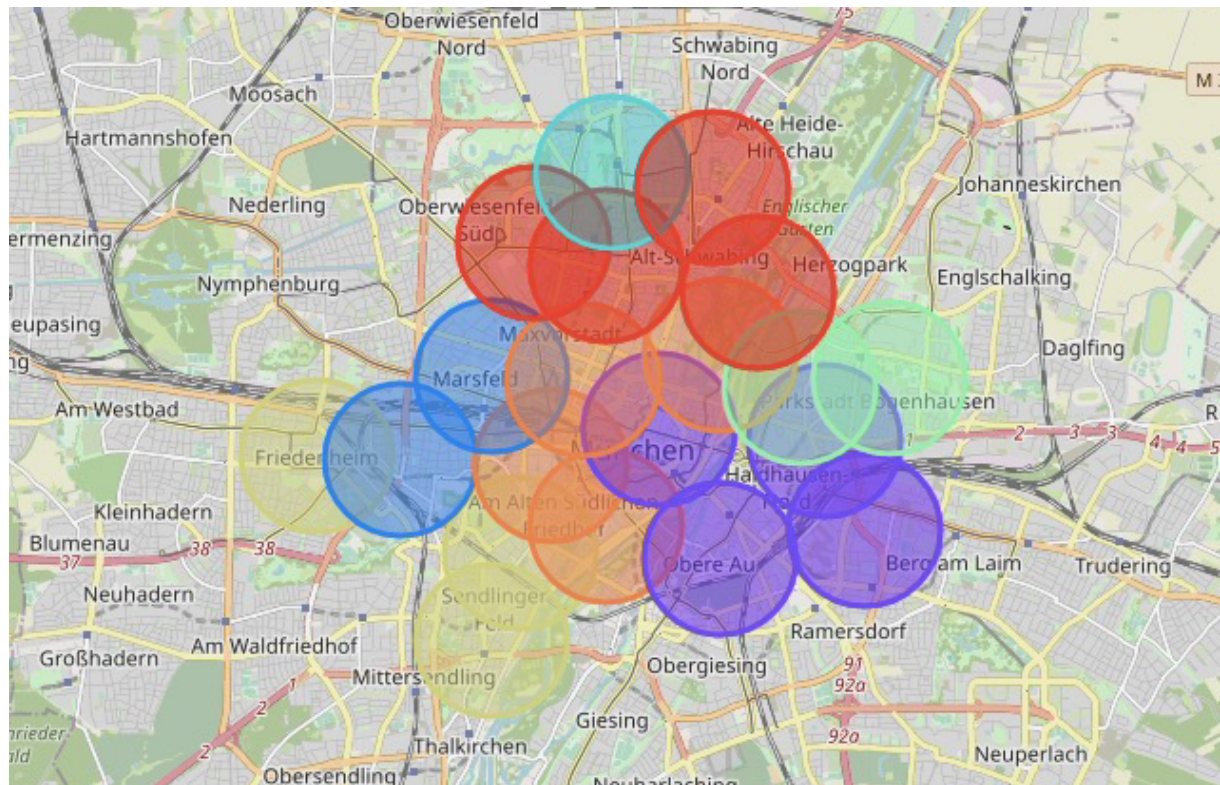
Results

The neighbourhoods in Berlin and Munich are very different from one another. Out of 14 clusters only two include neighbourhoods in both cities. All others are limited to one city.

In the case of Berlin and Munich, there is no regional effect which is visible. However, there are some clusters of neighborhoods which are adjacent to each other and most of them are in various different directions.
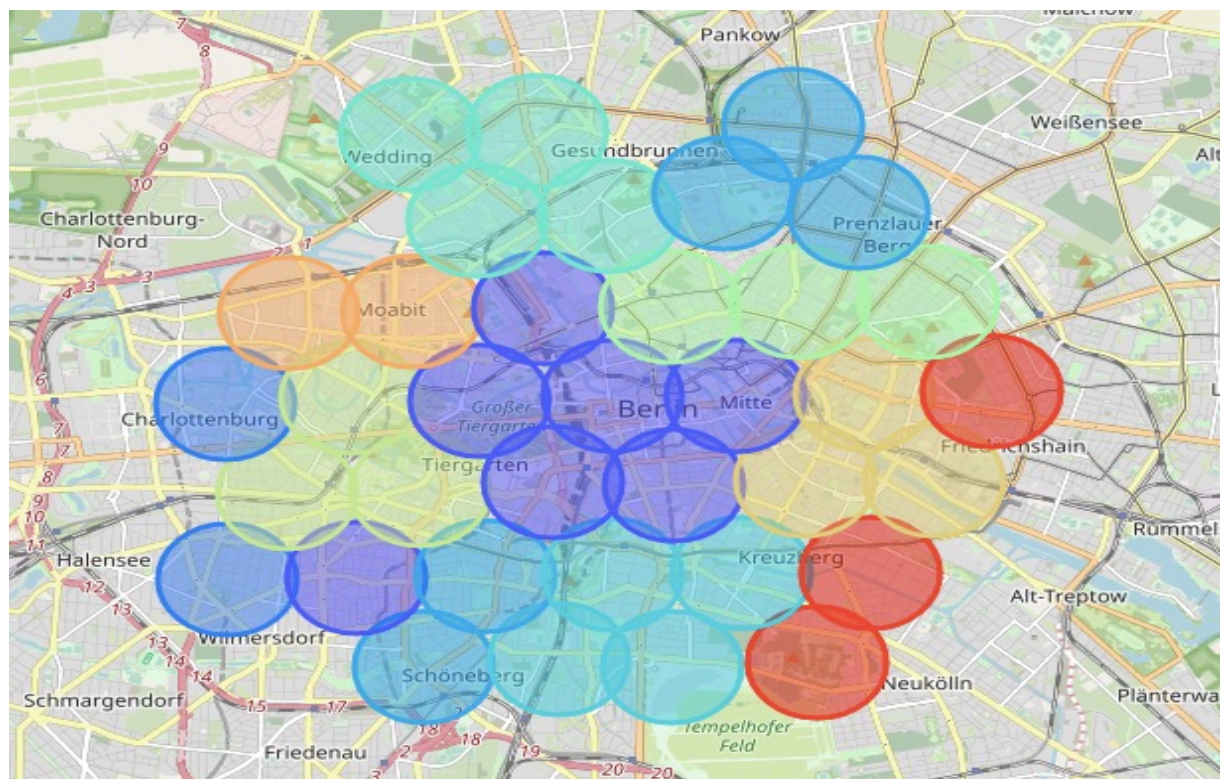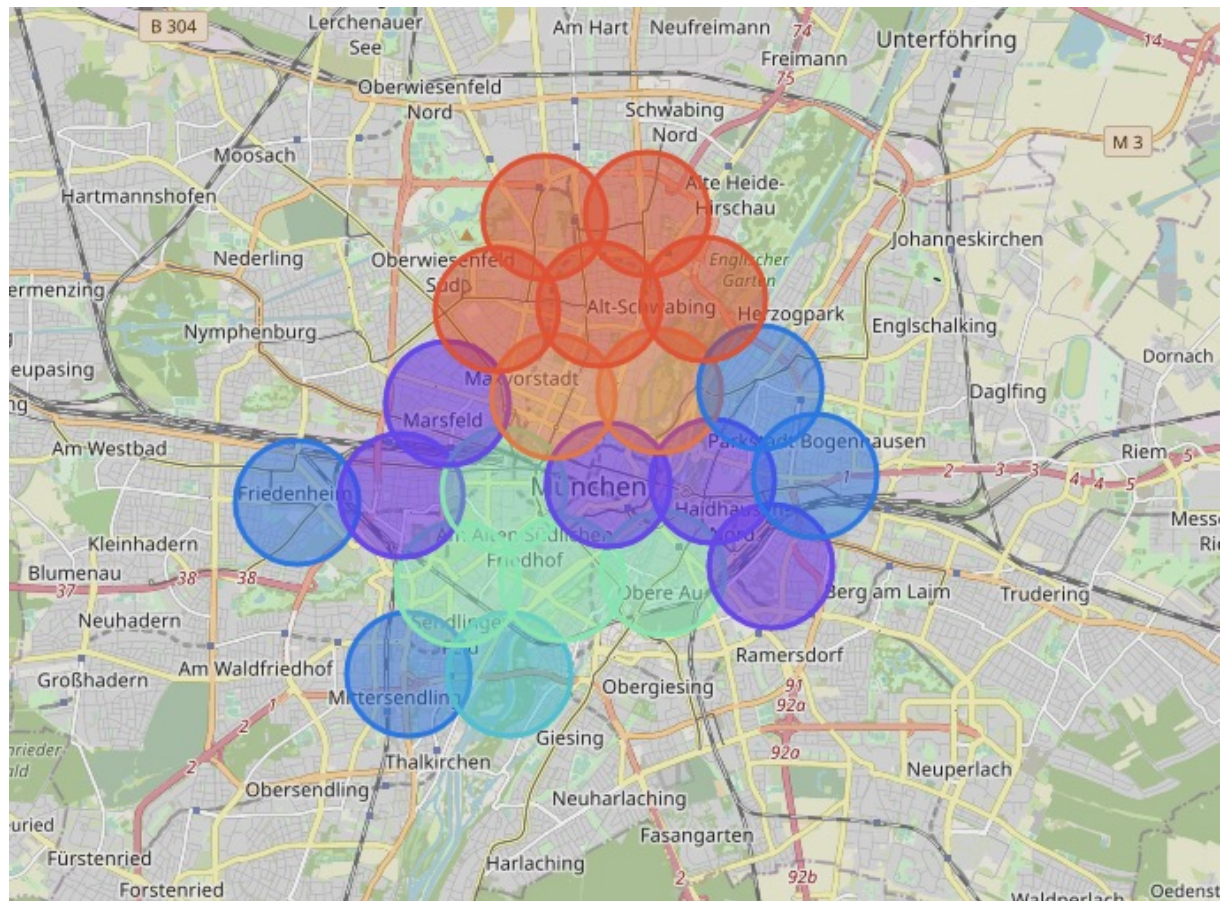
Therefore, for the original question of where to move to in the new city based on the current living neighborhood only two clusters seems to suitable for analysis.

```
Cluster 0 has 0 munich neighborhoods and 3 Berlin neighborhoods.
Cluster 1 has 5 munich neighborhoods and 0 Berlin neighborhoods.
Cluster 2 has 0 munich neighborhoods and 7 Berlin neighborhoods.
Cluster 3 has 4 munich neighborhoods and 2 Berlin neighborhoods.
Cluster 4 has 0 munich neighborhoods and 5 Berlin neighborhoods.
Cluster 5 has 1 munich neighborhoods and 4 Berlin neighborhoods.
Cluster 6 has 0 munich neighborhoods and 4 Berlin neighborhoods.
Cluster 7 has 4 munich neighborhoods and 0 Berlin neighborhoods.
Cluster 8 has 0 munich neighborhoods and 3 Berlin neighborhoods.
Cluster 9 has 0 munich neighborhoods and 3 Berlin neighborhoods.
Cluster 10 has 0 munich neighborhoods and 3 Berlin neighborhoods.
Cluster 11 has 0 munich neighborhoods and 2 Berlin neighborhoods.
Cluster 12 has 2 munich neighborhoods and 0 Berlin neighborhoods.
Cluster 13 has 5 munich neighborhoods and 0 Berlin neighborhoods.
```

...

	Cluster Labels	munich	Berlin
3	3	4	2
5	5	1	4

Cluster 3:

The most common venues categories in cluster 3 are Hotel, Italian Restaurant, Bakery, Supermarket and Cafe. In Munich this includes neighborhoods in Sendlingerfeld, Friedenheim, Parkstadt and Altbogenhausen. In Berlin this includes neighborhoods in Charlottenburg and Wilmersdorf.

Cluster 5:

The most common venues categories in cluster 5 are Italian Restaurant, Café, Bar, cocktail bar and bistro. In Munich this includes only the neighborhood in Sendlingerfeld. In Berlin this includes twice the neighborhoods in Tempelhof-Schöneberg and Friedrichshain-Kreuzberg.

Discussion:

The approach of going with algorithmically generated neighborhoods has proven to lead to much better results than going with the official boroughs. Neighbourhoods in German cities appear to be defined by a very small radius and the boroughs therefore include many neighborhoods that can be very different from one another.

As expected the data in Germany is not as extensive as it is for American cities. I had to exclude several neighbourhoods because I could not even get 100 venue recommendations for them. This also means that all results should be taken "with a grain of salt".

Conclusion

In this analysis we tried to identify neighborhoods in Berlin and Munich that are similar to one another so that I could use these as recommendations for people moving between the two cities.

I used a clustering algorithm and Foursquare venue recommendations to group 60 neighbourhoods.

Interestingly, I found that not a lot of neighbourhoods in each city share similar traits with neighbourhoods in the other city. Instead, only two clusters with 6 and 5 neighborhoods each are represented in both cities respectively.

Cluster 5 appears to be a cluster of neighborhoods where many young people live and therefore go out to eat in Cafés and bars. Cluster 3 on the other hand is more targeted towards families with many essential day to day services like Supermarkets, restaurants and Bakeries.

So if you're moving from Berlin to Munich you may want to consider Sendlingerfeld or Altbogenhausen for your next apartment.