Moving to Düsseldorf, Germany

Markus Gehring

26th May 2020

Contents

1 Introduction **2**

1.1 Background . . . 2

1.2 Reasons for choosing to move to Düsseldorf . . . 2

1.3 Problem statement . . . 2

2 Data used to approach the analysis **3**

2.1 Data and data sources . . . 3

2.1.1 Geospatial data and meta data of the city districts/neighbourhoods . . . 3

2.1.2 Infrastructure data on the neighbourhoods . . . 3

2.1.3 Information on apartments . . . 3

2.1.4 Data on public transport . . . 3

2.2 Data usage to address problem . . . 3

3 Methodology **4**

3.1 General idea . . . 4

3.2 Geolocation data of the city . . . 4

3.2.1 Map of Düsseldorf with its districts and their boroughs . . . 4

3.2.2 Apartment data . . . 4

3.3 Clustering the boroughs . . . 4

4 Results and Discussion **5**

4.1 Map of Düsseldorf . . . 5

4.2 Clustering results . . . 5

4.3 Apartment evaluation on location . . . 6

5 Conclusion **7**

1 Introduction

1.1 Background

Düsseldorf is a city in the west of Germany with a population of over 600.000. [1]. According to a 2012 quality of living survey it is ranked the sixth most liveable city in the world[2].

For a new job, I consider moving to the city of Düsseldorf, Germany. I have been there a couple of times, but mainly in the city centre and therefore do not know much about the living quality of the individual districts and neighbourhoods.

To make a choice on a new apartment, not solely based on the cold facts about said apartment, such as size or rent, but also on the living quality that comes with it in terms of infrastructure, I want to use this project to identify the most appropriate area to move to within the city. This question of living quality is potentially not only interesting to me, but to everyone, who considers moving to Düsseldorf.

1.2 Reasons for choosing to move to Düsseldorf

Apart from the aforementioned overall high living quality, there are more reasons to consider in favour of moving to Düsseldorf:

- It is a business hot-spot in the area, located in the the region North-Rhine Westphalia, Germany's most densely populated region, with a large number of job opportunities;
- it is also a very international city, featuring the third largest Japanese community in Europe, hosting branch offices of hundreds of Japanese companies.
- It is also located near the border to the Netherlands and Belgium.
- The large Düsseldorf Airport (as the name suggests) is located in close proximity to the city and can be reached using the public transport system.

Therefore an in-depth analysis of its rental market combined with clustered infrastructure and transportation system data, is an asset to many people and also businesses, which want to support employees moving to this vibrant city full of opportunities.

1.3 Problem statement

In this concrete case, an apartment is sought, that fulfils the following requirements:

1. Proximity to a tram or train station of less than 500 metre
2. Size at least 50 m^2
3. Rent of up to 800 € (not going into the details of German rental prices, this is an appropriate amount excluding additional cost for heat, gas, and water)
4. The neighbourhood should feature an adequate number of restaurants, cafés, and parks. (This requirement will be refined based on the identified clusters)

[1]https://en.wikipedia.org/wiki/D%C3%BCsseldorf
[2]https://mobilityexchange.mercer.com/insights/quality-of-living-rankings

2 Data used to approach the analysis

2.1 Data and data sources

2.1.1 Geospatial data and meta data of the city districts/neighbourhoods

Düsseldorf is administratively divided into 10 districts with 2 to 8 neighbourhoods each. For my approach I will use the following:

- Boundaries of the neighbourhoods[3] and districts[4] provided by the city of Düsseldorf as geojson data;
- the geolocation of neighbourhoods selected for analysis and clustering using the bing maps API.

2.1.2 Infrastructure data on the neighbourhoods

To analyse and cluster the neighbourhoods Foursquare data will be used via the Foursquare API.

2.1.3 Information on apartments

To get an idea of the rental prices of apartments in Düsseldorf, data will be scraped from a popular German rental service website *immobilienscout24.de*[5].

The data set will be reduced in size using two alternative approaches. As one approach, additional requirements will be applied to the data set within python. This approach resembles a real life application approach. The alternative approach is random sampling of the data, to ensure a large variety of the data set.

The number of analysed neighbourhoods will be limited, to stay within the scope of the Foursquare API requests limitations.

2.1.4 Data on public transport

Location data of long distance and local train stops and city train stops is provided by the city of Düsseldorf[6].

Location data of bus and tram stops is provided by the regional public transportation provider 'Verkehrsverbund Rhein-Ruhr'[7].

2.2 Data usage to address problem

To address the problem stated in the introduction, the location of the available apartments, and stations will be mapped using folium. The neighbourhoods will be analysed and clustered as was demonstrated in the previous courses. Combining these pieces of information suitable living quarters will be analysed.

[3] https://opendata.duesseldorf.de/dataset/stadtbezirksgrenzen-düsseldorf
[4] https://opendata.duesseldorf.de/dataset/stadtteilgrenzen-d%C3%BCsseldorf
[5] https://www.immobilienscout24.de/Suche/de/nordrhein-westfalen/duesseldorf/wohnung-mieten
[6] https://opendata.duesseldorf.de/dataset/bahnh%C3%B6fe-d%C3%BCsseldorf
[7] https://www.openvrr.de/dataset/haltestellen

3 Methodology

3.1 General idea

The overall methodology of this project is not overly complex. For the most part, data is scraped from various sources on geolocation data, venues, and apartments. The data is cleaned and prepared for mapping. The machine learning technique employed to cluster the boroughs of the city is k-means. In this section, the individual steps will be discusses in detail.

3.2 Geolocation data of the city

3.2.1 Map of Düsseldorf with its districts and their boroughs

The geolocation data of the city's districts and their boroughs will be read into the notebook using pandas and geopandas. At this point, the geolocation center of each borough will be determined for later analysis using Foursquare.

This map is enhanced by the location data of public transportation stops in the city. For this purpose the data provided by the city was used as is. However, the data from the transportation services provider, required some cleaning. First, any stops outside of the city of Düsseldorf were removed and the coordinates were transformed from an uncommon MRCV format to decimal coordinates and some missing data had to be removed.

3.2.2 Apartment data

The apartment data was scraped and the data types were prepared for analysis. 1314 apartments were found. They are located in eight boroughs distributed across six districts. All but one of the districts were connected, so only these five districts were further analysed. To ensure a high enough number of different elements to cluster, all 25 of the boroughs in the districts were analysed.

The apartment data set was analysed using descriptive statistics, especially in the 'real' approach, to visualise the distribution of price and size. Regression could have been used to correlate the rent with the size, but visualising the correlation using a scatter plot was thorough enough to get a feeling for the data set.

3.3 Clustering the boroughs

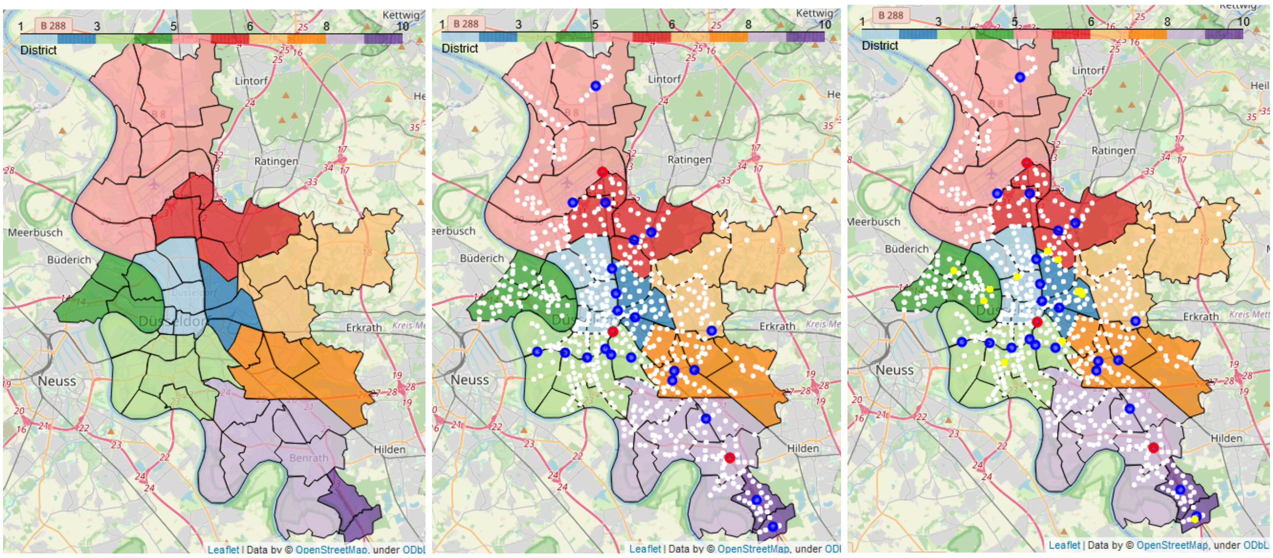
Using Foursquare data, venues around the center of each borough are identified. Following the examples of the previous courses and their assignments, the frequency of each type of venue is calculated for each borough. Based on the occurrence of the top 20 venue types the boroughs are clustered into 11 clusters using k-means. The number of clusters was chosen such that the largest cluster is less than 10 boroughs, which readily happened choosing smaller k values. A to large value created issues when mapping using a choropleth map with colour-coded clusters. The largest number of colours in one qualitative set is 12.

k-means clustering was an appropriate approach for this type of analysis, because the boroughs were compared using their venue type frequency as the only parameter and this is a typical application for clustering.

4 Results and Discussion

4.1 Map of Düsseldorf

Figure 1 shows the maps of Düsseldorf with various amount of additional information.



(a) Colour-coded districts and their boroughs.

(b) Map with indicated long distance train stations (red), city train stations (blue), and bus/tram/metro stops (white dots).

(c) Map with transportation data and apartment location (yellow dots).

Figure 1: Maps of Düsseldorf

Mapping out the districts of the city and colour-code them, makes the shape of the city quite apparent. Especially the west border of the city formed for the most part by the Rhine river becomes directly visible as the smooth curved line to the left hand side. The one district on the left-hand side of the Rhine (district 4 in dark green) can be identified.

Mapping out the transportation data visualises very well, that there are two main city train axes in the city, a north-south bound one, which includes the long distance train stations and a west-east connection, which continues to the neighbouring city of Neuss. This makes sense, as Neuss is city were many commuters live who work in Düsseldorf.

The apartments in our random sample of 200 out of the more than 1300 available apartments are mostly located in and around the city centre in districts 1 through 4, 6 and 10. However, district 10 is located in the far south of the city and is not connected to any of the other districts.

4.2 Clustering results

The results of the clustering are shown in Figure 2.

The clustering reveals two very large clusters of similar boroughs in light blue and dark green, while the surrounding boroughs are more or less unique with only one other cluster including more than one borough (cluster 10).

Cluster 3 (dark green) consists of boroughs with all but one having hotels as their most common venue. Other common venues include various types of restaurants, bars and cafés. They are also located towards the centre of the city and alongside the city train north-south axis.

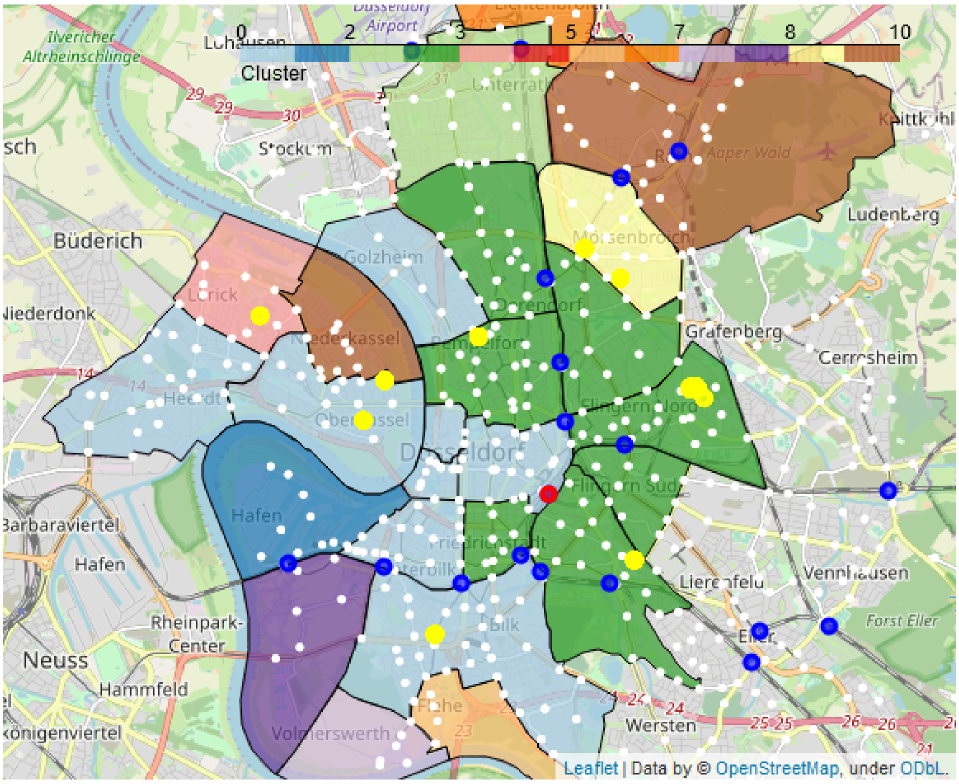


Figure 2: Clusters of districts 1, 2, 3, 4, and 6.

Cluster 0 (light blue) on the other hand is characterised mostly by shopping venues, such as clothing stores, shopping malls and boutiques. There is also quite a large number of restaurants and cafés. It includes the actual city centre of Düsseldorf.

Clusters 1 and 2 are very similar to each other and ignoring their lack of hotels are quite similar both to each other and to cluster 3 with many restaurants. What sets them apart is their higher frequency of sport venues such as dance and yoga studios.

Cluster 4 is a bit special because its number 2 of the top locations are actually bus stops. Notwithstanding a large number of restaurants there are also many cinemas (4th most common).

Cluster 5 (light orange in the south) features by far the most sport related venues with the top four being soccer fields, beaches, sports clubs and gyms. As well as yoga and dance studios in the top 20 most frequent venues. However, having a look on the map reveals that this part of the city is not connected very well with public transport in mind, likely related to the motorway that goes through here.

The other clusters will not be analysed in detail. Suffice to say, that they are actually very similar to each other still.

4.3 Apartment evaluation on location

When the randomly selected apartments are then analysed some patterns emerge. Apartments in district 4 (to the west) are quite large with at least 90 m^2 and very expensive with two of the three apartments having a monthly rent of over 2000 Euro.

Likewise apartments in district 2 are extremely expensive, even though they are not very large. This indicates a high cost of living in district 2. This fits the large number of restaurants that can be found in these boroughs. Similarly apartments in district 3 are equally expensive as the ones in district 2.

Apartments prices plummet when the border to district 6 is crossed. Both apartments identified here have a rent of below 800 Euro, while being of similar size to the ones in district 2. The apartment in district 10 is also similar.

These trends make sense considering district 1 is the city centre and living in close proximity to the city centre is very attractive leading to higher rents. An exception to this rule is the apartment located within district 1, in the borough of Pempelfort. It has three rooms, with a total size of 96 m^2 for 950.00 €.

5 Conclusion

In this simulation the city of Düsseldorf was analysed and its boroughs clustered to identify areas with optimal living conditions to help identify viable candidates of apartments.

To this end, publicly available information on the boroughs of the city of Düsseldorf and its public transportation network were used. Apartment locations were scraped off of a popular rental apartment search website.

Clustering the boroughs of the city of Düsseldorf using k-means, a machine learning algorithm for clustering data was employed. However, the results were only mediocre, as the central boroughs of the city of Düsseldorf were overall very similar.

Even so, possible feasible apartments were identified from a random sample of 200 apartments. Namely the one in Pempeldorf, a central borough in district 1 as well as two apartments in district 6. All of them are located in the north-western part of the city. Due to the very dense net of tram and bus stops, any requirements to the availability of public transport are fulfilled for all of the apartments.

The identification of apartments could be further enhanced by clustering the apartments using machine learning instead of random sampling or employing strict cut-off requirements.

The results may also be better if more advanced machine learning tools, such as SVM for clustering of arbitrarily shaped clusters was used instead of k-means. Although this was outside the scope of this assignment.