The Battle of Neighborhoods

- Moving to Frankfurt am Main -

Final Report as Part of the
IBM Capstone Project

Julian Oellrich
7 January 2021

Inhaltsverzeichnis

1 Introduction and Project Goal

1.1 Background

There are many people, who are working in various cities across the world. Let's say you live in the city center of Munich, Germany. You love your neighborhood mainly because of all the great amenities and other types of venues that exist in the neighborhood. Such as gourmet restaurants, pharmacies, parks, bars, and so on. Now say you receive a job offer from a great company in another city with great career prospects. However, given the far distance from your current place, you unfortunately must move if you decide to accept the offer. Wouldn't it be great if you're able to determine neighborhoods in the new city, which are exactly the same as your current neighborhood, and if not, perhaps similar neighborhoods that are at least closer to your new job.

1.2 Problem Scenario

That's exactly the challenge i am facing at the moment. My girlfriend and I are currently living in Munich. We really enjoy living in our borough because of all the great bars, parks and restaurants. All in all our borough offers us a great level of comfort and living quality. Unfortunately, because of job opportunities and a personal decision, we are planning to move to Frankfurt am Main medium term. - That's great and we really want to move and looking forward to it, but we do not want to sacrifice the great amenities of our current borough in munich.

1.3 Project Goal

The goal of this project is to find boroughs in Frankfurt that are similar to our current borough in munich in terms of venues nearby or a specific category of venues. Therefore, we want to compare different boroughs in both cities and find a list of the most similar in terms of living qualities for us.

1.4 Methodology

In order to do this, given the cities Munich and Frankfurt, I will segment them into different boroughs using the geographical coordinates of the center of each borough. And then, using a combination of location data and machine learning to group the neighborhoods into clusters. To compare the boroughs of both cities and to find the most similar boroughs to our current location I will use an item based recommender system algorithm. That will give me a list of boroughs that are the most similar to our current borough in Munich.

2 Data Structure

In order to segment the boroughs, explore venues in those boroughs and compare them, we will essentially need a dataset that contains all the boroughs and the neighborhoods of both Munich and Frankfurt as well as the the latitude and longitude coordinates of each borough.

Additionally it is interesting to gather some demographic data of each borough, in order to make a demographic comparison of the result boroughs. Data such as size, population, population density and number of immigrants are gathered.

Therefore, in a first important step, the borough and location data of Munich and Frankfurt has to be imported. If necessary needs to be cleaned and prepared as a dataframe with the following information:

- Borough
- Size in qkm
- Population
- Density in Inh/qkm
- Immigrants
- Latitude
- Longitude

This dataframe will be created for both cities, Munich and Frankfurt.

2.1 Frankfurt Data

The Frankfurt borough and population data is provided as XLSX File (filename: *Frankfurt_Borough_Data.xlsx*). Frankfurts's borough and demographic information can be imported through this Excel file. Latitude and longitude information has to be pulled by the python geolocator package.

An excerpt of the final Munich dataframe is shown below.

	Borough	Size in qkm	Population	Density in Inh/qkm	Immigrants	Latitude	Longitude
0	Altstadt	0.506	4218.0	8336.0	0.367	50.110644	8.682092
1	Innenstadt	1.491	6599.0	4426.0	0.464	50.108052	8.682161
2	Bahnhofsviertel	0.542	3552.0	6554.0	0.520	50.108411	8.668151
3	Westend-Süd	2.497	19314.0	7735.0	0.275	50.115245	8.662270
4	Westend-Nord	1.632	10373.0	6356.0	0.290	50.126356	8.667921
5	Nordend-West	3.100	30897.0	9967.0	0.219	50.124914	8.677950
6	Nordend-Ost	1.532	23182.0	15132.0	0.223	50.124920	8.692317
7	Ostend	5.564	29477.0	5298.0	0.285	50.112373	8.699967

2.1.1 Frankfurt Data Summary

To get a first idea of the Frankfurt data structure a few basic characteristics are described in the following section.

The Frankfurt dataframe has 45 rows and 7 columns. Frankfurt has 45 unique Boroughs and 758574 inhabitants in total across all boroughs.

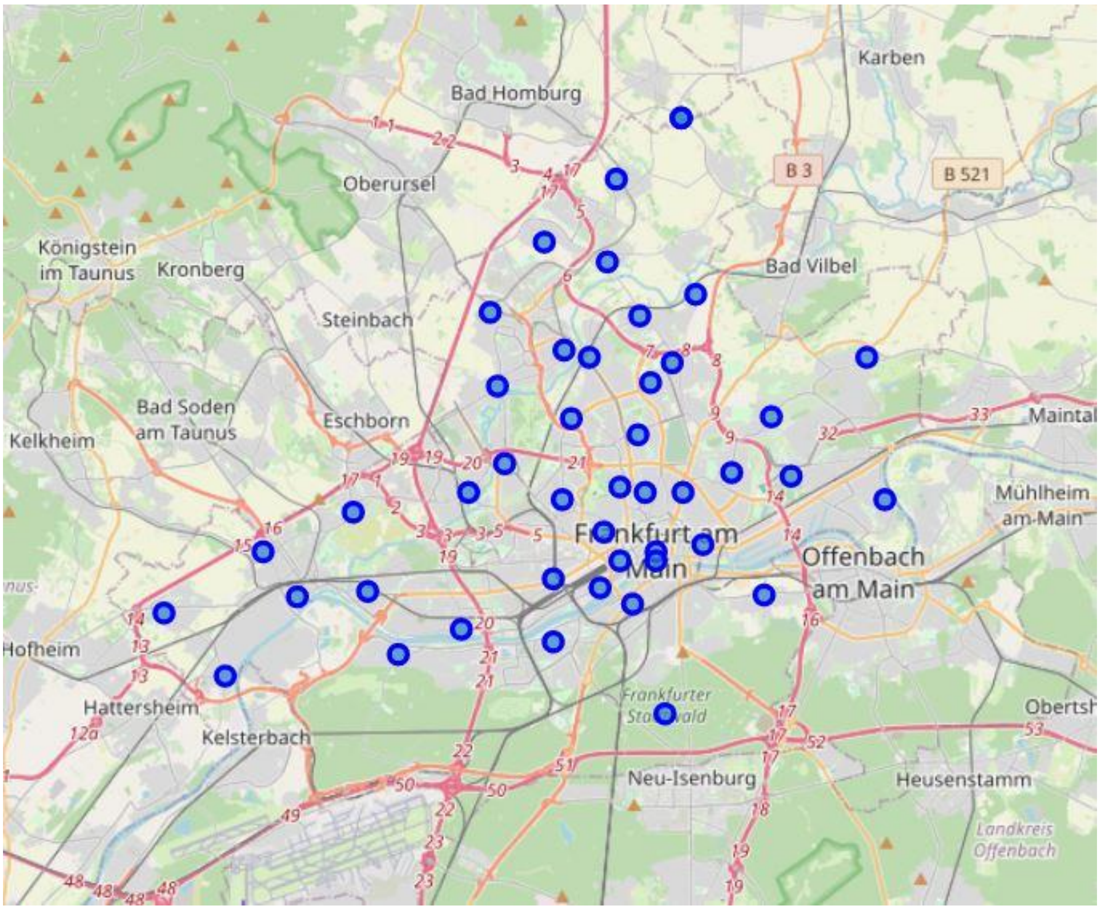
The biggest borough in terms of its absolute population is **Bockenheim** with 41904 inhabitants, where **Sachsenhausen-Süd** is the biggest in size with 30.535 square km.

An important borough characteristic, when discussing life quality can be population density. **Nordend-Ost** with 15132.0 inhabitants per square km has the highest population density of all boroughs in Frankfurt.

Finally yet importantly, in **Bahnhofsviertel** 52% of inhabitants have migration background, which makes this borough number one in terms of immigrants.

2.1.2 Frankfurt Borough Map

The following figure visualizes each borough based on the geospatial data (latitude and longitude) and superimposes the markers on top of the Frankfurt city map.



2.2 Munich Data

The Munich borough and population data cannot be download directly. To get the final dataframe several steps have to be executed.

Munich's borough and demographic information can be scrapped from this [wikipedia page](). Latitude and longitude information have to be pulled by the python geolocator package.

An excerpt of the final Munich dataframe is shown below.

	Borough	Size in qkm	Population	Density in Inh/qkm	Immigrants	Latitude	Longitude
0	Altstadt-Lehel	3.15	21126.0	6716.0	0.260	48.137828	11.574582
1	Ludwigsvorstadt-Isarvorstadt	4.40	51933.0	11799.0	0.283	48.130722	11.566526
2	Maxvorstadt	4.30	51834.0	12060.0	0.256	48.151092	11.562418
3	Schwabing-West	4.36	68935.0	15800.0	0.228	48.168271	11.569873
4	Au-Haidhausen	4.22	61654.0	14611.0	0.235	48.128753	11.590536
5	Sendling	3.94	41256.0	10475.0	0.272	48.118012	11.539083
6	Sendling-Westpark	7.81	60498.0	7742.0	0.295	48.118031	11.519333
7	Schwanthalerhöhe	2.07	29611.0	14303.0	0.326	48.133782	11.541057

2.2.1 Munich Data Summary

To get a first idea of the Munich data structure a few basic characteristics are described in the following section.

The Munich dataframe has 25 rows and 7 columns. Munich has 25 unique Boroughs and 1560042 inhabitants in total across all boroughs.

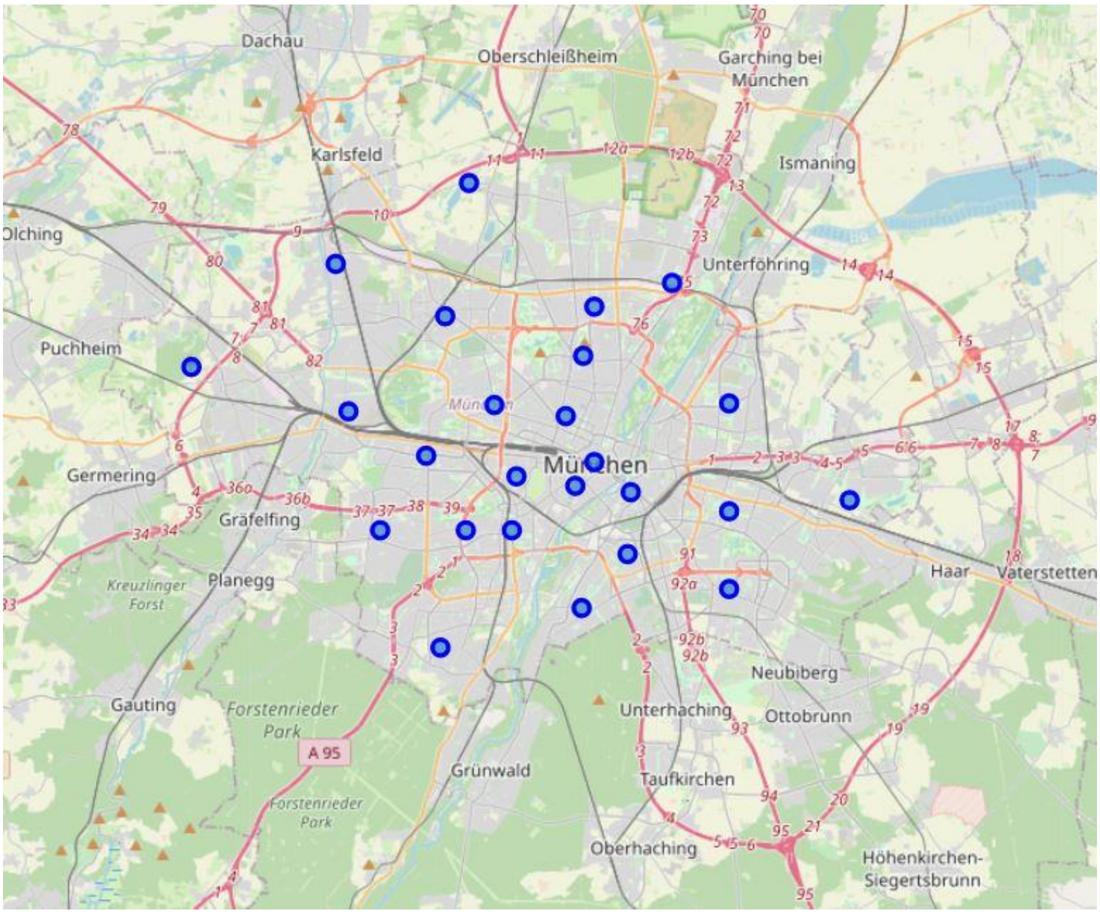
The biggest borough in terms of its absolute population is **Ramersdorf-Perlach** with 117918 inhabitants, where **Aubing-Lochhausen-Langwied** is the biggest in size with 34.06 square km.

An important borough characteristic, when discussing life quality can be population density. **Schwabing-West** with 15800 inhabitants per square km has the highest population density of all boroughs in Munich.

Finally yet importantly, in **Milbertshofen-Am Hart** 41.3% of inhabitants have migration background, which makes this borough number one in terms of immigrants.

2.2.2 Munich Borough Map

The following figure visualizes each borough based on the geospatial data (latitude and longitude) and superimposes the markers on top of the Munich city map.



Dachau
Oberschleißheim
Garching bei München
Karlsfeld
Ismaning
Olching
Unterföhring
Puchheim
München
Germering
Gräfelfing
Planegg
Kreuzlinger Forst
Haar
Vaterstetten
Neubiberg
Gauting
Forstenrieder Park
A 95
Unterhaching
Ottobrunn
Grünwald
Taufkirchen
Oberhaching
Höhenkirchen-Siegertsbrunn

3 Venue Exploration using Foursquare API

After getting above data, let’s find the nearby venues of each borough in both cities. To get the nearby venue information the Foursquare API was used. The result is one dataframe for each city, containing all nearby venue categories for each borough. A one hot encoding was performed and finally the dataframe was grouped by borough and by taking the mean of the frequency of occurrence of each venue category. This results in the following to dataframes, which are the base for our further analysis and clustering.

Munich grouped dataframe (exerpt):

	Borough	ATM	Accessories Store	Afghan Restaurant	American Restaurant	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Asian Restaurant	...	Tram Station	Trattoria/Osteria	Tunnel	Turkish Restaurant	Vegetarian / Vegan Restaurant	Vi R
0	Allach-Untermenzing	0.0	0.000000	0.00	0.0	0.0	0.00	0.00	0.00	0.000000	...	0.000000	0.062500	0.0	0.000000	0.00	
1	Altstadt-Lehel	0.0	0.000000	0.00	0.0	0.0	0.01	0.00	0.01	0.000000	...	0.000000	0.010000	0.0	0.000000	0.01	
2	Au-Haidhausen	0.0	0.000000	0.02	0.0	0.0	0.00	0.01	0.00	0.000000	...	0.000000	0.010000	0.0	0.020000	0.00	
3	Aubing-Lochhausen-Langwied	0.0	0.000000	0.00	0.0	0.0	0.00	0.00	0.00	0.000000	...	0.000000	0.000000	0.0	0.000000	0.00	
4	Berg am Laim	0.0	0.017857	0.00	0.0	0.0	0.00	0.00	0.00	0.053571	...	0.017857	0.035714	0.0	0.017857	0.00	

Frankfurt grouped dataframe (exerpt):

	Borough	ATM	Accessories Store	African Restaurant	American Restaurant	Apple Wine Pub	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	...	Trail	Train Station	Tram Station	Trattoria/Osteria	Turkish Restaurant	Vegetarian / Vegan Restaurant	Viet Res
0	Altstadt	0.0	0.0	0.00	0.01	0.04	0.0	0.0	0.0	0.05	...	0.000000	0.00	0.0	0.01	0.01	0.0	
1	Bahnhofsviertel	0.0	0.0	0.02	0.00	0.00	0.0	0.0	0.0	0.03	...	0.000000	0.00	0.0	0.01	0.01	0.0	
2	Bergen-Enkheim	0.0	0.0	0.00	0.00	0.00	0.0	0.0	0.0	0.00	...	0.090909	0.00	0.0	0.00	0.00	0.0	
3	Berkersheim	0.0	0.0	0.00	0.00	0.00	0.0	0.0	0.0	0.00	...	0.000000	0.05	0.0	0.00	0.00	0.0	
4	Bockenheim	0.0	0.0	0.00	0.01	0.00	0.0	0.0	0.0	0.00	...	0.000000	0.00	0.0	0.00	0.02	0.0	

For further exploration, the most common venue categories in each borough, we extracted the top 10 venue categories of each borough for Frankfurt and Munich. Those two dataframes can give us a better understanding of the characteristic of each borough in terms of venues nearby.

Munich top 10 venue categories (dataframe exerpt):

	Borough	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	Allach-Untermenzing	Bakery	Drugstore	Bavarian Restaurant	Supermarket	Hotel	Tennis Court	German Restaurant	Sporting Goods Shop	Italian Restaurant	Trattoria/Osteria
1	Altstadt-Lehel	Plaza	Café	Hotel	Coffee Shop	Bavarian Restaurant	Restaurant	Gourmet Shop	Department Store	German Restaurant	Boutique
2	Au-Haidhausen	Café	Plaza	Coffee Shop	Hotel	German Restaurant	Italian Restaurant	Beer Garden	Cocktail Bar	Turkish Restaurant	Beach
3	Aubing-Lochhausen-Langwied	Bakery	Bus Stop	Light Rail Station	Pharmacy	Soccer Field	Supermarket	Hotel	Bistro	German Restaurant	Electronics Store
4	Berg am Laim	Supermarket	Italian Restaurant	Asian Restaurant	Drugstore	German Restaurant	Bakery	Hotel	Doner Restaurant	Park	Bavarian Restaurant
5	Bogenhausen	Bus Stop	Italian Restaurant	Supermarket	Hotel	Drugstore	German Restaurant	Tram Station	Beer Garden	Plaza	Asian Restaurant

Frankfurt top10 venue categories (dataframe exerpt):

	Borough	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	Altstadt	Café	Plaza	Art Museum	Bar	Wine Bar	Apple Wine Pub	German Restaurant	Hotel	Scenic Lookout	Coffee Shop
1	Bahnhofsviertel	Hotel	Café	Indian Restaurant	Seafood Restaurant	Vietnamese Restaurant	Italian Restaurant	Restaurant	Asian Restaurant	Art Museum	Bar
2	Bergen-Enkheim	Italian Restaurant	Taverna	Plaza	German Restaurant	Business Service	Ice Cream Shop	Trail	Food & Drink Shop	Water Park	Supermarket
3	Berkersheim	Supermarket	Soccer Field	German Restaurant	Pharmacy	Pizza Place	Climbing Gym	River	Bus Stop	Boxing Gym	Light Rail Station
4	Bockenheim	Italian Restaurant	Café	Asian Restaurant	Botanical Garden	Wine Bar	Supermarket	Bakery	Bar	Japanese Restaurant	Pizza Place
5	Bonames	Supermarket	Café	Electronics Store	Italian Restaurant	Hotel	Athletics & Sports	Motorcycle Shop	Burger Joint	Garden Center	Chinese Restaurant

4 Analysis

In this section, we will describe the results of some preliminary analysis. This exploratory data analysis compares both cities in terms of demographic data. Further, we will analyze the venue category overlap between both cities and cluster the cities boroughs on the base of most common venues.

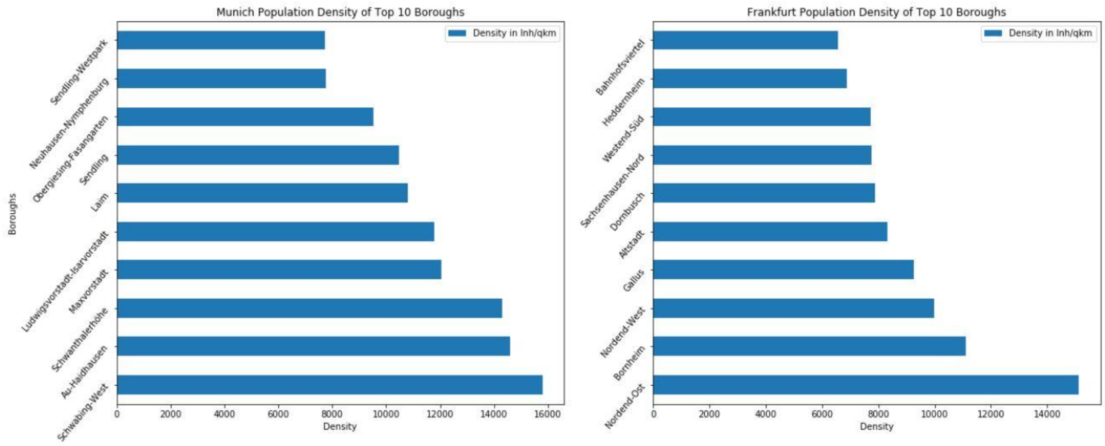
The second part of this section explains the clustering approach to answer our main question: Which boroughs in Frankfurt are most similar to our current borough in Munich (Ludwigsvorstadt-Isarvorstadt)?

4.1 Exploratory Data Analysis

In this first step we want to conduct some exploratory data analysis to further get a grasp of our data and the cities boroughs.

Demographics

The following graph compares both cities in terms of their top 10 boroughs, with the highest population density.



We found, that Munich overall has a higher population density in its boroughs. The mean population density of Munich is 7349 inhabitants per square km (inh/km^2) compared to 4768 inh/km^2 in Frankfurt. Schwabing-West has the highest density in Munich with 15800 inh/km^2. In comparison, Nordend-Ost counts 15132 inh/km^2 and marks the densest borough in Frankfurt.

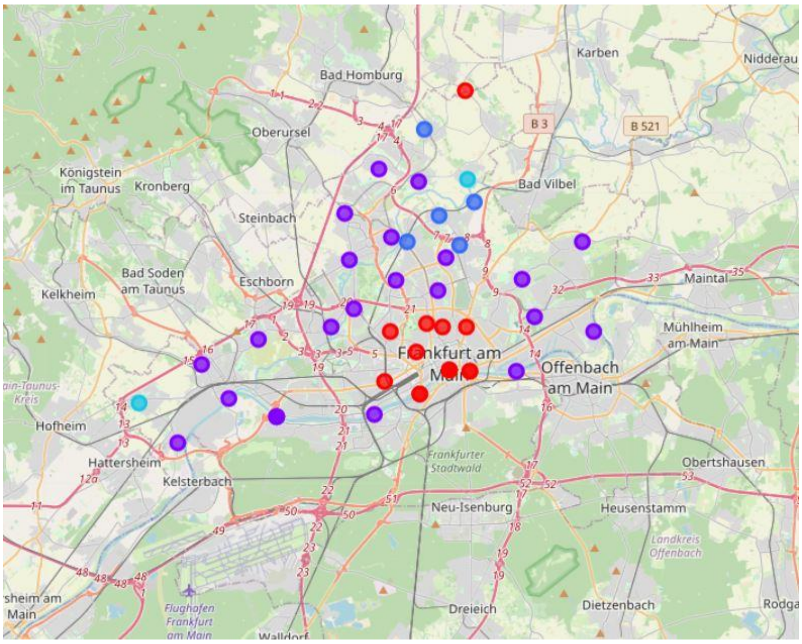
The current borough we are living in, Ludwigsvorstadt-Isarvorstadt, has a population density of 11799 inh/km^2, which is way above average of Munich and Frankfurt.

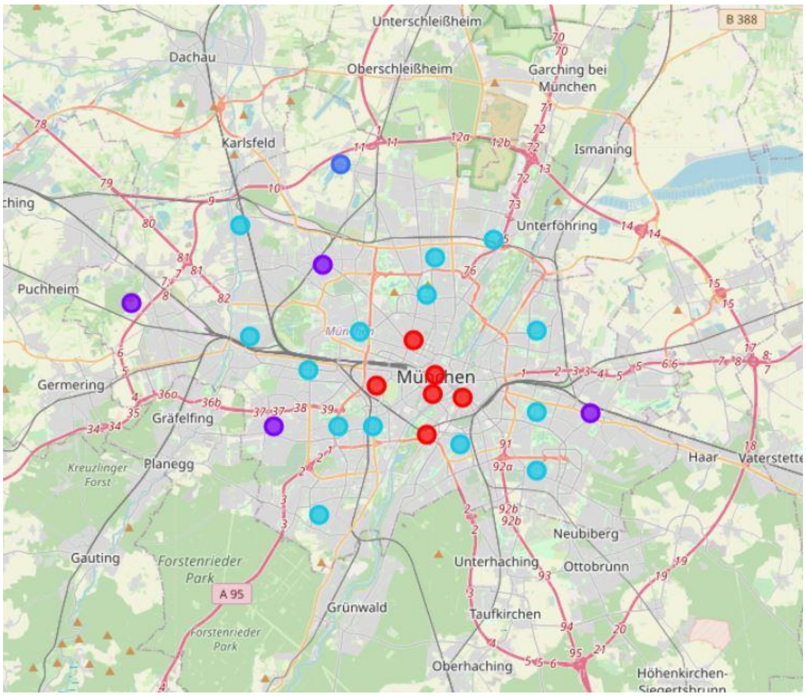
Let’s assume that population density is an additional decision characteristic to look for when discussing our result. Our goal is to find the most similar boroughs in terms of venue categories, but we can additionally look for the population density in comparison to our current borough. The assumption is a lower density means a better life quality. We will discuss this in the conclusion section.

Borough Clustering and Clustered Maps

To get a first understanding about the differences of the boroughs in terms of venues nearby, we clustered the boroughs of both cities separately. For clustering the grouped dataframes shown earlier were used. Those dataframes contain on hot encoded data and the mean of the frequency of occurrence of each venue category in every borough.

A k-means algorithm was modeled in order to cluster each the boroughs of each city into 4 clusters. To visualize the result, city maps of both cities with the boroughs super-imposed on top were plotted and shown below.

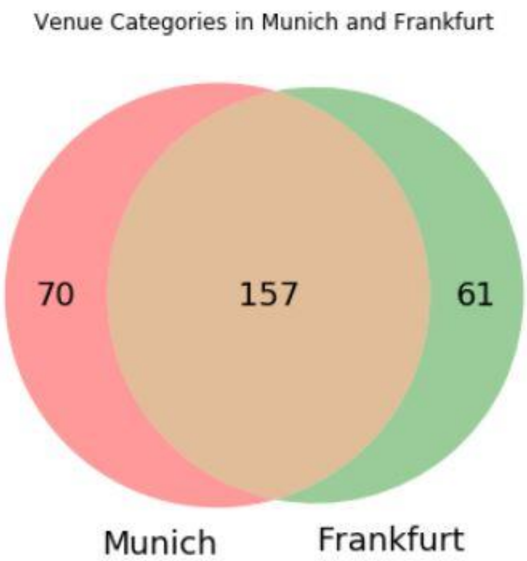




Venue Category Overlap

Both cities may yield in different number of venue categories, but we'll take only overlapping venue categories. This is required in order to run a clustering algorithm across all boroughs of both cities combined.

The following venn diagram shows the number of venue categories of each city and the overlap of categories, which are present in Frankfurt **and** Munich.



For the final clustering algorithm, we only take those 157 venue categories and therefore remove all other entries in both dataframes.

4.2 Most similar Boroughs in Frankfurt

In order to find the most similar boroughs to our current, we combined all boroughs of both cities and there venue category entries to on big dataframe.

For clustering we run a k-means algorithm with 8 clusters using this combined dataframe. This algorithm gives us 8 clusters of boroughs across both cities, where boroughs within the same cluster are very similar to each other in terms of venue categories.

Those clusters were added to the dataframe and the resulting dataframe looks as followed:

	Borough	Cluster Labels	ATM	Accessories Store	American Restaurant	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Asian Restaurant	...	Thai Restaurant	Theater	Trail	Tram Station	Trattoria/Osteria	Turk Restaura
0	Allach-Untermenzing	2	0.0	0.000000	0.00	0.0	0.00	0.00	0.00	0.000000	...	0.000000	0.00	0.0	0.000000	0.062500	0.000
1	Altstadt-Lehel	3	0.0	0.000000	0.00	0.0	0.01	0.00	0.01	0.000000	...	0.000000	0.01	0.0	0.000000	0.010000	0.000
2	Au-Haidhausen	3	0.0	0.000000	0.00	0.0	0.00	0.01	0.00	0.000000	...	0.010000	0.00	0.0	0.000000	0.010000	0.020
3	Aubing-Lochhausen-Langwied	6	0.0	0.000000	0.00	0.0	0.00	0.00	0.00	0.000000	...	0.000000	0.00	0.0	0.000000	0.000000	0.000
4	Berg am Laim	2	0.0	0.017857	0.00	0.0	0.00	0.00	0.00	0.053571	...	0.000000	0.00	0.0	0.017857	0.035714	0.017
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
65	Sossenheim	2	0.0	0.000000	0.00	0.0	0.00	0.00	0.00	0.000000	...	0.041667	0.00	0.0	0.000000	0.000000	0.000
66	Unterliederbach	2	0.0	0.015873	0.00	0.0	0.00	0.00	0.00	0.015873	...	0.031746	0.00	0.0	0.000000	0.000000	0.015
67	Westend-Nord	1	0.0	0.000000	0.01	0.0	0.00	0.00	0.00	0.010000	...	0.010000	0.00	0.0	0.000000	0.000000	0.000
68	Westend-Süd	3	0.0	0.000000	0.01	0.0	0.00	0.00	0.01	0.010000	...	0.000000	0.02	0.0	0.000000	0.000000	0.000
69	Zeilsheim	4	0.0	0.000000	0.00	0.0	0.00	0.00	0.00	0.000000	...	0.000000	0.00	0.0	0.000000	0.000000	0.000

The last step is now, finding all boroughs, that have the same cluster as our current borough Ludwigsvorstadt-Isarvorstadt and are located in Frankfurt.

The resulting list of the most similar boroughs in terms of venue categories is shown below:

- Altstadt
- Bahnhofsviertel
- Bornheim
- Gutleutviertel
- Innenstadt
- Ostend
- Sachsenhausen-Nord
- Sachsenhausen-Süd
- Westend-Süd

5 Conclusion

In this project, we solved our main goal: we will move to Frankfurt and want to find the most similar boroughs in Frankfurt in order to maintain our quality of living. With a clustering algorithm, it was possible to get a list of similar boroughs in terms of venue categories. We found that this resulting list was a good fit, after talking to some friends who are living in Frankfurt and know both cities.

As a further analysis, we could analyze this resulting list of boroughs in terms of its demographics or several other characteristics (such as apartment prices etc.) and compare them to Ludwigsvorstadt-Isarvorstadt. This could result in an even better collection of similar boroughs.