

Munich – Analysis of the neighbourhoods for newcomers

Applied Data Science Capstone Project

Jakub Čampa (Juli 2020)

1. Introduction

1.1 Background

The city of Munich, Germany is third biggest city in Germany and home to approximately 1.5 Million (Q1/2020) inhabitants [1]. With many national and international companies, Munich is one of the most important centres of economy in Germany and therefore interesting for employees from Germany as well as foreigners.

According to the official city website, there were 118.000 new (domestic and foreign) inhabitants registered in 2019 [2]. That means, that 118.000 people came last year to Munich and needed to search for accommodation and a place to live.

Geographically, Munich consists of 25 neighbourhoods, each with its own organisational structure and administration.

The real estate situation in Munich is tense; Munich is currently the most expensive city in Germany according to immobilienmarkt magazine 'immobilienmarkt.faz' [3].

1.2 Problem

Each of the 25 city neighbourhoods is unique and offers a different setup and environment for its new inhabitants. The requirements of each newcomer are individual and requires personal assessment.

The aim to this project is to provide a qualitative analysis of Munich neighbourhoods and help to answer the questions:

Which of the 25 neighbourhoods in Munich may be suitable for me/us?

Which of the 25 neighbourhoods are similar to each other?

1.3 Interest

According to the official city website, there were 118.000 new (domestic and foreign) inhabitants registered in 2019 [2]. That means, that 118.000 people came last year to Munich and needed to search for accommodation and a place to live.

2. Data acquisition

2.1 Data sources

All of the basic geographical and demographical data of the city of Munich can be scraped on Wikipedia page in German language [here](#) [4]. This data was used as a base for neighbourhood analysis as well as for part of a demographical analysis.

Further data about city criminality were used to analyse criminal rate occurrence for each neighbourhood. The data can be found [here](#) [5].

To be able to analyse one of the most important factors for newcomers (Real estate Price), data from miet.check.de were used. For purpose of this analysis, only average price/m2 for renting an apartment were used.

For visualisation purposes, .json data of Munich neighbourhoods can be downloaded and adjusted [here](#).

Foursquare data were used, to analyse Munich 25 neighbourhoods, cluster them and understand unique specifics of each of them.

2.2 Data cleaning/pre-processing

The downloaded and scraped data were combined into one table/dataframe. Several adjustments were made to be able to proceed with this data.

Since most of the website from which data originated are in German language, all data used were changed to English language.

JSON data were manipulated, for easier identification and pairing of the neighbourhoods.

Foursquare data were cleaned, merged with original data and processed to be able to obtain meaningful results.

2.3 Feature selection

Five main features were selected for the analysis, with focus on Price as a main and primary feature in the decision process of newcomers.

a) Price (EUR/m2/Neighbourhood)

Property rent Price is mostly a number one criterium for newcomers in every city, country. In this analysis, the price per m2 in EUR of rented flat was used to display correlation with other features and differentiate 25 neighbourhoods according to this feature.

b) Distance from the city centre (km)

Distance from the city may be one another important features and decision criteria as well. Commuting to work, closeness of shopping possibilities, freetime activities, restaurants, city centre and many more, may be important in decision making. Data were gained from [google.maps.com](https://www.google.com/maps) and measured as distance from neighbourhood centre to Munich city centre (Marienplatz).

c) Criminality (number of criminal acts)

The safety situation of chosen neighbourhood may play important role in decision making process. Criminal act statistics were used to display potentially dangerous neighbourhoods and put in comparison with other neighbourhoods.

d) Foreigners share (%)

Foreigners share feature may be seen as a individual decision criterium. For some foreign newcomers welcoming community, infrastructure and known language may be preposition for a good start in new country. Other may prefer less foreigners to be able to quicker integrate in German culture.

e) Density (Inhabitants/km2)

Density of each neighbourhood is measured in 'inhabitants per square kilometre', and provides crucial information about concentration of inhabitant as well as possible information about apartment availability.

f) Venues (Foursquare)

Venues data from Foursquare API were analysed to provide specific features and information about each neighbourhood. Many useful data for individual decision making as an amount of shopping possibilities, restaurants, galleries, museums, parks and gardens may be found here.

3. Exploratory Data Analysis

3.1 Methodology

At the beginning of this study, **statistical exploratory analysis** was used to determine possible **correlation** between chosen features and to better understand characteristics of the data using **numpy** and **seaborn** libraries. The strongest correlation between factors was further analysed to better understand the relationship. **Folium** libraries were used for **geographical visualization** of **choropleth** maps of each features and provide visual, interactive guidance for newcomers in Munich. To better understand individual neighbourhood specifics, **Foursquare** database were used. At the end of the analysis, **K-means clustering** was used to show similarities between neighbourhoods. **Pandas libraries** were used as a base for all coding in python.

	Nr.	Neighborhood_x	Inhabitants	Density(Inh/km2)	Foreigners(%)	Latitude	Longitude	PriceEURm2	Distance_from_citycentre_in_km	CriminalActs/Year
0	1.0	Altstadt-Lehel	21.100	6.708	261	48.137828	11.574582	24.55	1.9	8205.0
1	2.0	Ludwigsvorstadt-Isarvorstadt	51.644	11.734	284	48.131771	11.555809	24.15	3.2	14091.0
2	3.0	Maxvorstadt	51.402	11.960	254	48.151092	11.562418	25.17	3.0	4799.0
3	4.0	Schwabing-West	68.527	15.706	227	48.168271	11.569873	25.14	5.6	2310.0
4	5.0	Au-Haidhausen	61.356	14.541	235	48.128753	11.590536	22.68	1.6	3531.0

Figure 1 - Final Munich dataframe

Munich has 25 neighbourhoods spread around the area of 310 square kilometres. Each of them has different specifics, geographical position, and history. At the map below, 25 neighbourhoods with its centre points are displayed for first and basic orientation in Munich.

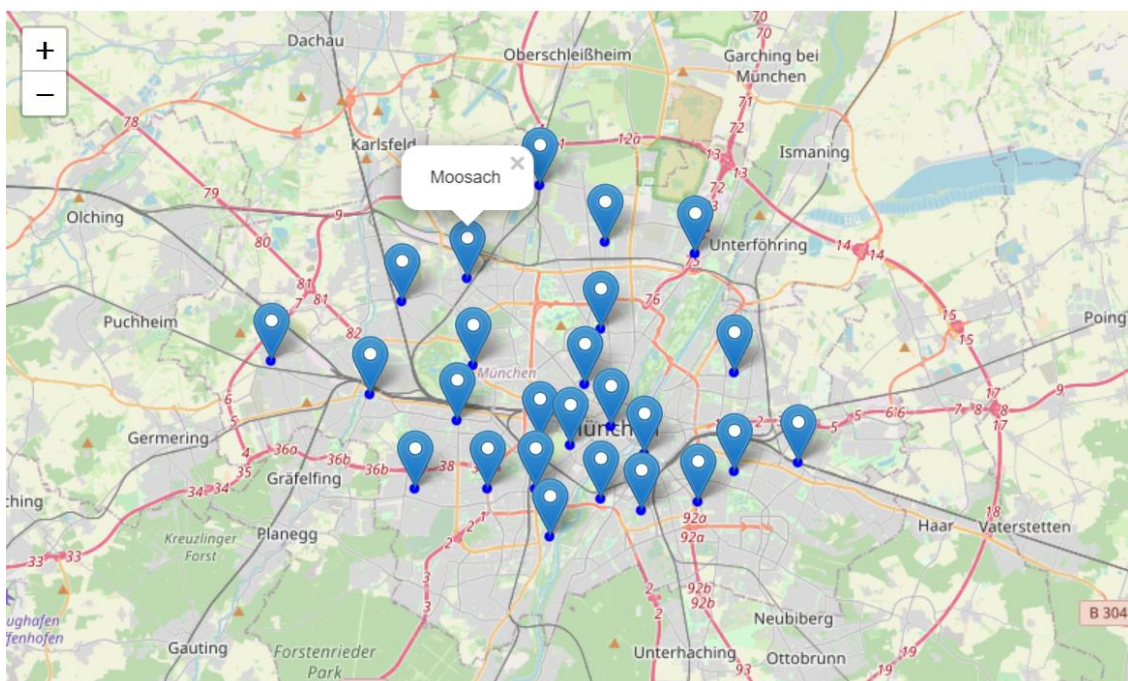


Figure 2 - Map of Munich neighbourhoods

3.2 Correlation analysis, relationships

Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate together. A positive correlation indicates the extent to which those variables increase or decrease in parallel; a negative correlation indicates the extent to which one variable increases as the other decreases[6].

In this analysis we will focus on correlation between any of five chosen factors with strong focus on main factor → Price of flat rent pro m2 in EUR.

	Density(Inh/km2)	Foreigners(%)	PriceEURm2	Distance_from_citycentre_in_km	CriminalActs/Year
Density(Inh/km2)	1.000000	-0.098530	0.697905	-0.791994	0.124203
Foreigners(%)	-0.098530	1.000000	-0.138102	0.064809	0.037091
PriceEURm2	0.697905	-0.138102	1.000000	-0.742745	0.541222
Distance_from_citycentre_in_km	-0.791994	0.064809	-0.742745	1.000000	-0.373864
CriminalActs/Year	0.124203	0.037091	0.541222	-0.373864	1.000000

Figure 3 – Correlation Matrix

Results:

- ➔ There is a significant correlation between Price/EUR/m2 & Distance from the city centre (-0.7427).
- ➔ There is a correlation between Price/EUR/m2 & Density (0.6979)
- ➔ There is a significant correlation between Density & Distance from the city centre (-0.7919)

These three correlations relationship will be examined closely.

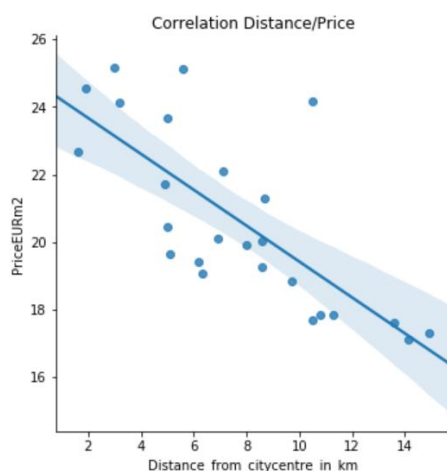


Figure 4 - Scatter Plot

-0.7427

Insights:

- further away from city centre, cheaper the prices for accommodation,
- cheapest and most expensive neighbourhoods are clearly defined,
- middle price segment is mixed.

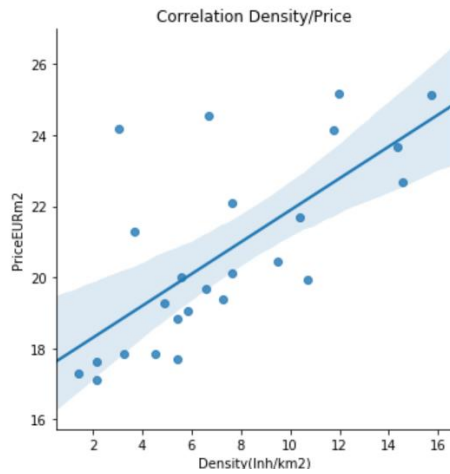


Figure 5 - Scatter Plot

0.6979

Insights:

- tendency: denser neighbourhoods are more expensive
- density around 2.000 Inh/km2 builds cheapest cluster

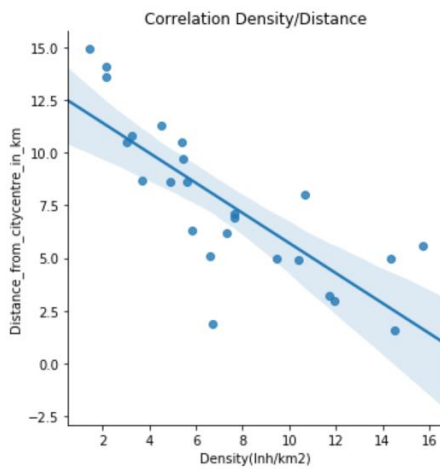


Figure 6 - Scatter Plot

-0.7919

Insights:

- further away from city centre, less dense
- probably more agriculture outside of the city
- more possibilities to find an accommodation outside of the city

3.2 Geographical analysis

Choropleth maps provide an easy way to visualize how a measurement varies across a geographic area or show the level of variability within a region.

Centre points were added for each map, to easily identify neighbourhood and see where the neighbourhood centre is located. All maps are interactive and each neighbourhood label consist of additional information.

3.2.1 Geographical analysis – Price

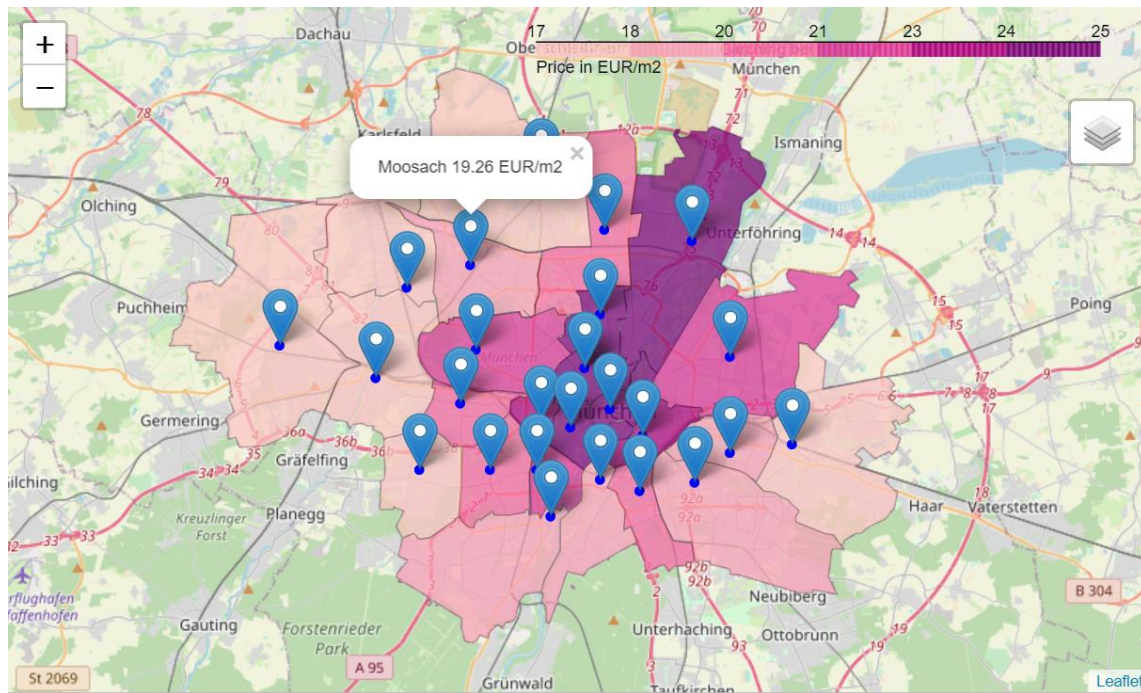


Figure 7 – Choropleth map, Price/m² in Munich

Insights:

- More central, more expensive
- West of Munich has several cheap neighbourhoods
- Schwabing Freiman, generally expensive also when further outside of the city
- All neighbourhoods close to the 'English garden' Park are expensive

3.2.2 Geographical analysis – Criminality

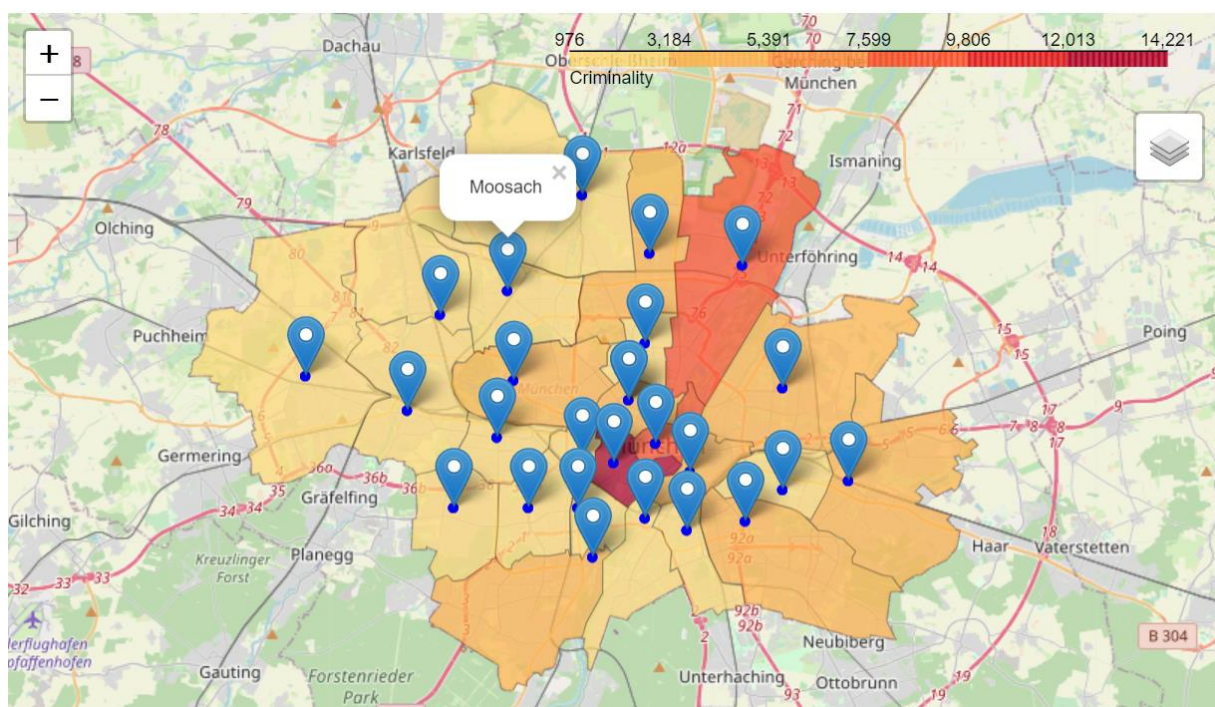


Figure 8 – Choropleth map, Criminality in Munich

Insights:

- Central train station area belongs to the most dangerous neighbourhood
- some of the most expensive neighbourhoods are also most dangerous
- the further out of the city you go, the safer are the neighbourhoods

3.2.3 Geographical analysis – Foreigners

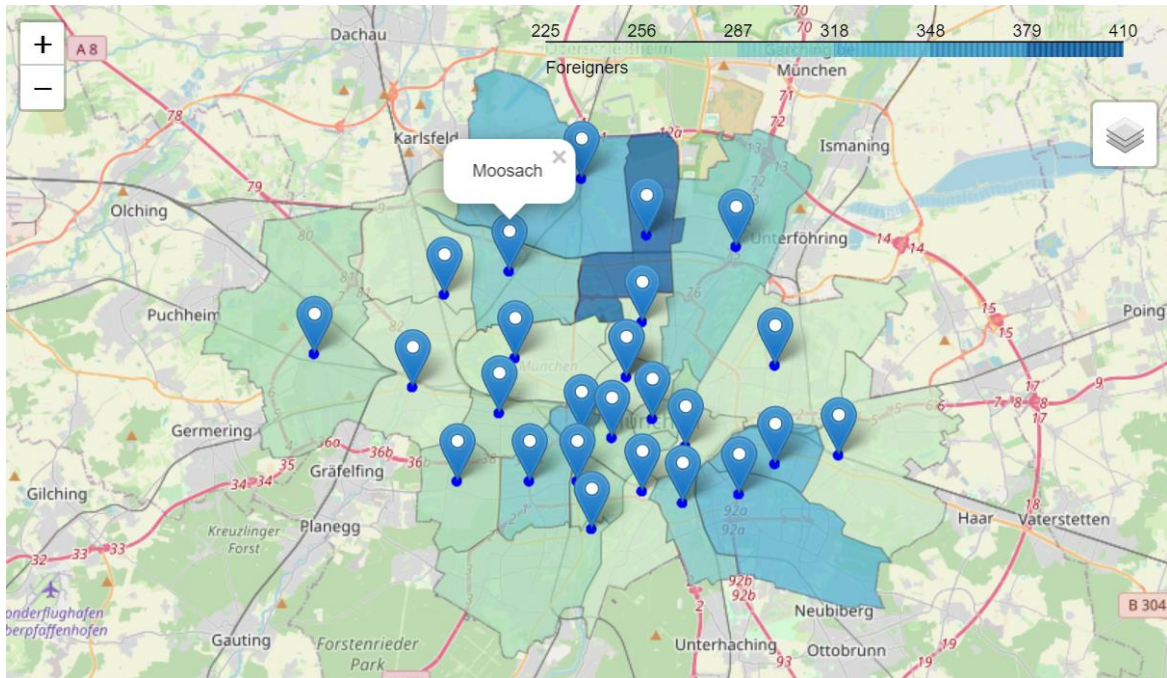


Figure 9 – Choropleth map, Foreigners proportion in Munich

Insights:

- Highest percentage of foreigners: Milbertshofen - am Hart, 41%
- Minimum of 22.5% of foreigners in each neighbourhood
- Western and Eastern outside neighbourhood have lowest amount of foreigners
- North Area has highest amount of foreigners

3.2.4 Geographical analysis – Density

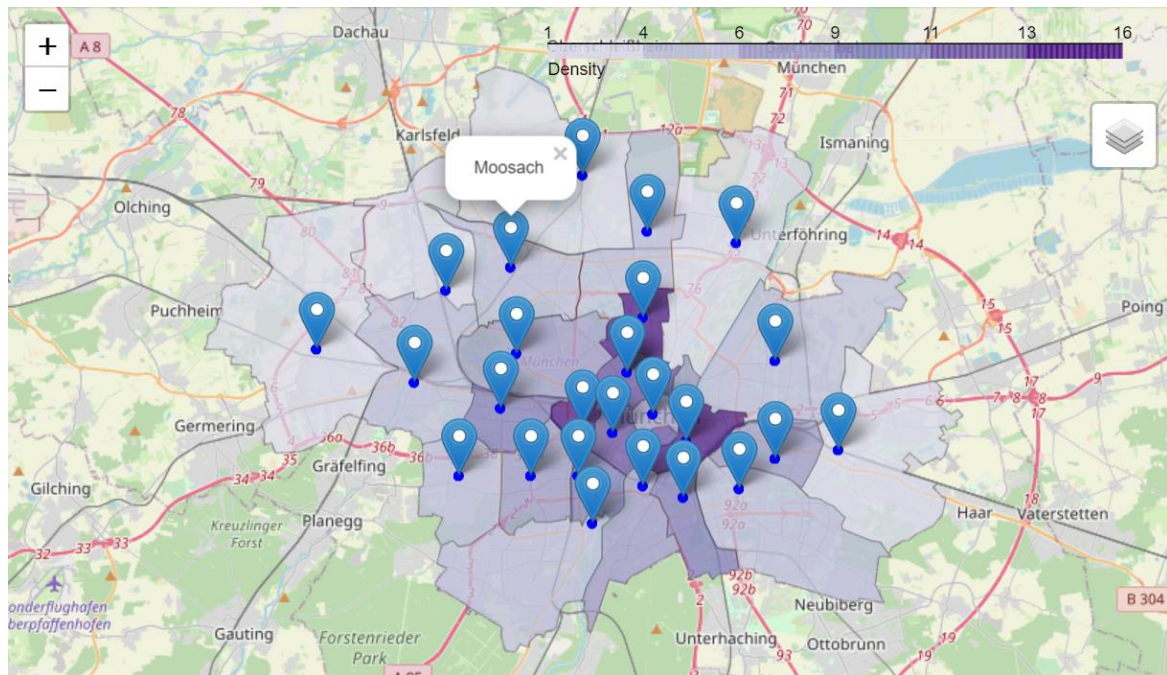


Figure 10 – Choropleth map, Population density in Munich

Insights:

- city centre neighbourhoods are denser populated as outside neighbourhoods
- Within outside ring, probably more 'village' feeling and agriculture land
- Schwabing West, most densed part of the city

3.3 Neighborhood 'Soft' Analysis (Foursquare)

Foursquare API provides free data and rich content about more than 100k venues. This data were used to analyse further Munich 25 neighbourhoods to better understand each individual characteristics and prepare the data for clustering.

	name	categories	lat	lng
0	Westpark	Park	48.122594	11.514015
1	Wöllinger	Brewery	48.113050	11.519748
2	Thalassa	Greek Restaurant	48.111216	11.517364
3	Seebühne im Westpark	Performing Arts Venue	48.123489	11.511789
4	Kino, Mond & Sterne	Indie Movie Theater	48.123530	11.511852

Figure 11 – Top 5 Venues – Sendling Westpark (Foursquare)

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Allach-Untermenzing	Bus Stop	Supermarket	Italian Restaurant	Home Service	Turkish Restaurant
1	Altstadt-Lehel	Café	Department Store	Boutique	Hotel	Clothing Store
2	Au-Haidhausen	Concert Hall	Thai Restaurant	Afghan Restaurant	Restaurant	Pool
3	Aubing-Lochhausen-Langwied	Bakery	Supermarket	Bus Stop	Soccer Field	Hotel
4	Berg am Laim	Bakery	Hotel	Supermarket	Metro Station	Asian Restaurant

Figure 12 – Top 5 Venues of each Neighbourhood (Foursquare)

4. Neighborhood clustering

The data from **foursquare API** were analysed, cleaned and prepared for clustering using combination of main Munich dataframe and new dataframe gained from foursquare API. Within the foursquare data, focused lied on first 30 venues within Radius of 1km of centre point of each neighbourhood.

Cluster column were added for further clustering of the neighbourhoods. K-means clustering and folium libraries were used to cluster and visualize the data.

According to 'toward data science blog', **K-means clustering** is one of the simplest and popular unsupervised machine learning algorithms.

A cluster refers to a collection of data points aggregated together because of certain similarities.

In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

Nr.	Neighborhood_x	Inhabitants	Density(Inh/km2)	Foreigners(%)	Latitude	Longitude	PriceEUR/m2	Distance_from_citycentre_in_km	CriminalActs/Year	Label	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0 1.0	Altstadt-Lehel	21.100	6.708	261	48.137828	11.574582	24.55	1.9	8205.0	Altstadt-Lehel/n24.55 EUR/m2	0	Café	Department Store	Boutique	Hotel	Clothing Store
1 2.0	Ludwigsvorstadt-Isarvorstadt	51.844	11.734	284	48.131771	11.555809	24.15	3.2	14291.0	Ludwigsvorstadt-Isarvorstadt/n24.15 EUR/m2	3	Café	Vietnamese Restaurant	German Restaurant	Italian Restaurant	Outdoor Sculpture
4 5.0	Au-Haidhausen	61.356	14.541	235	48.128793	11.590536	22.68	1.6	3531.0	Au-Haidhausen/n22.68 EUR/m2	0	Concert Hall	Thai Restaurant	Afghan Restaurant	Restaurant	Pool
5 6.0	Sending	40.983	10.405	269	48.118012	11.539083	21.71	4.9	1983.0	Sending/n21.71 EUR/m2	3	Café	German Restaurant	Hotel	Drugstore	Gym
6 7.0	Sending-Westpark	59.843	7.632	289	48.117986	11.516253	20.11	6.9	2234.0	Sending-Westpark/n20.11 EUR/m2	2	Bus Stop	Bakery	Garden	Intersection	Italian Restaurant

Figure 13 – Final Dataframe for clustering

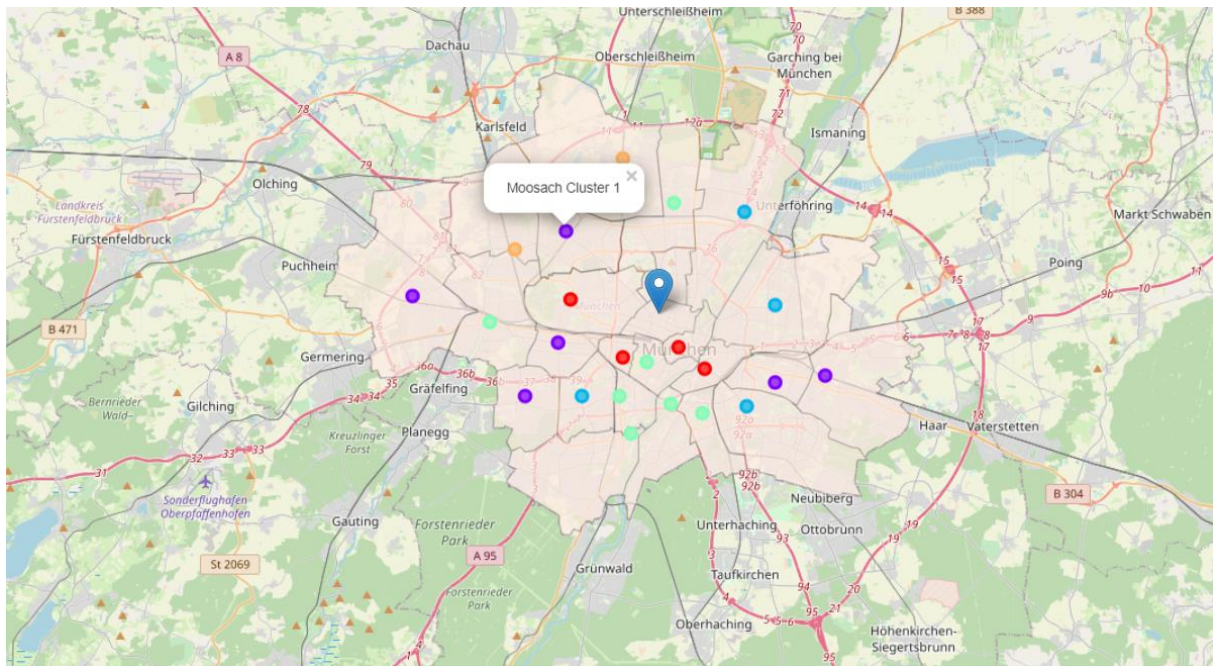


Figure 14 – Cluster map, displaying 5 neighbourhood clusters

5. Results:

5.1 Cluster No. 1

	Neighborhood_x	Latitude	Longitude	PriceEUR/m2	Distance_from_citycentre_in_km	CriminalActs/Year	Label	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
0	Altstadt-Lehel	48.137828	11.574582	24.55		1.9	8205.0	Altstadt-Lehel\n24.55 EUR/m2	0	Café	Department Store	Boutique	Hotel	Clothing Store
4	Au-Haidhausen	48.128753	11.590536	22.68		1.6	3531.0	Au-Haidhausen\n22.68 EUR/m2	0	Concert Hall	Thai Restaurant	Afghan Restaurant	Restaurant	Pool
7	Schwanthalerhöhe	48.133782	11.541057	23.67		5.0	1703.0	Schwanthalerhöhe\n23.67 EUR/m2	0	Café	Asian Restaurant	Italian Restaurant	Outdoor Sculpture	Museum
8	Neuhausen-Nymphenburg	48.157124	11.509703	22.09		7.1	3913.0	Neuhausen-Nymphenburg\n22.09 EUR/m2	0	Café	Italian Restaurant	Greek Restaurant	German Restaurant	Tram Station

Figure 15. – Cluster Nr.1

Cluster No. 1 Characteristics:

- Coffee and Restaurants cluster
- Distance from the city centre up to 7km
- Average renting price is more than 22EUR/per m2 – expensive neighbourhoods

5.2 Cluster No. 2

	Neighborhood_x	Latitude	Longitude	PriceEURm2	Distance_from_citycentre_in_km	CriminalActs/Year	Label	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
9	Moosach	48.184241	11.506496	19.26	8.6	2796.0	Moosach\n19.26 EUR/m2	1	Supermarket	Bakery	Drugstore	Plaza	Hotel
13	Berg am Laim	48.123483	11.633451	19.40	6.2	2322.0	Berg am Laim\n19.40 EUR/m2	1	Bakery	Hotel	Supermarket	Metro Station	Asian Restaurant
14	Trudering-Riem	48.126036	11.663338	17.84	10.8	3251.0	Trudering-Riem\n17.84 EUR/m2	1	Supermarket	Drugstore	Bakery	German Restaurant	Hotel
19	Hadern	48.118064	11.481842	17.69	10.5	1556.0	Hadern\n17.69 EUR/m2	1	Supermarket	Bus Stop	German Restaurant	Metro Station	Asian Restaurant
21	Aubing-Lochhausen-Langwied	48.158437	11.414066	17.30	14.9	1620.0	Aubing-Lochhausen-Langwied\n17.30 EUR/m2	1	Bakery	Supermarket	Bus Stop	Soccer Field	Hotel
24	Laim	48.139551	11.502166	19.93	8.0	2413.0	Laim\n19.93 EUR/m2	1	Supermarket	Greek Restaurant	Gastropub	Drugstore	Doner Restaurant

Figure 16 – Cluster Nr.2

Cluster No. 2 Characteristics:

- Supermarket and Bakeries neighbourhoods
- Distance from the city centre between 6.2 – 15km
- Average renting price is between 17.30 – 19.93EUR/per m2 – average neighbourhoods
- Low criminality rate

5.3 Cluster No. 3

	Neighborhood_x	Latitude	Longitude	PriceEUR/m2	Distance_from_citycentre_in_km	CriminalActs/Year	Label	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
6	Sendling-Westpark	48.117986	11.516253	20.11		6.9	2234.0	Sendling-Westpark\n20.11 EUR/m2	2	Bus Stop	Bakery	Garden	Intersection	Italian Restaurant
11	Schwabing-Freimann	48.191958	11.614275	24.18		10.5	8413.0	Schwabing-Freimann\n24.18 EUR/m2	2	Fast Food Restaurant	Concert Hall	Bakery	Bus Stop	Comedy Club
12	Bogenhausen	48.154782	11.633484	21.30		8.7	3225.0	Bogenhausen\n21.30 EUR/m2	2	Bus Stop	Park	Drugstore	Hotel	Italian Restaurant
15	Ramersdorf-Perlach	48.113916	11.615699	19.05		6.3	4712.0	Ramersdorf-Perlach\n19.05 EUR/m2	2	Bus Stop	Bakery	Chinese Restaurant	Park	Taverna

Figure 17 – Cluster No. 3

Cluster No. 3 Characteristics:

- Supermarket and Bakeries neighbourhoods
- Distance from the city centre between 6.3 – 10.5km
- Average renting price is between 19.05– 24.18EUR/per m2 – tendence expensive neighbourhoods
- Average to high criminality rate

Cluster No. 4

	Neighborhood_x	Latitude	Longitude	PriceEURm2	Distance_from_citycentre_in_km	CriminalActs/Year	Label	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Ludwigsvorstadt-Isarvorstadt	48.131771	11.555809	24.15	3.2	14091.0	Ludwigsvorstadt-Isarvorstadt\n24.15 EUR/m2	3	Café	Vietnamese Restaurant	German Restaurant	Italian Restaurant	Outdoor Sculpture
5	Sendling	48.118012	11.539083	21.71	4.9	1983.0	Sendling\n21.71 EUR/m2	3	Café	German Restaurant	Hotel	Drugstore	Gym
10	Milbertshofen-Am Hart	48.195925	11.571815	20.01	8.6	3987.0	Milbertshofen-Am Hart\n20.01 EUR/m2	3	Café	Hotel	Bus Stop	Bakery	Metro Station
16	Obergiesing-Fasangarten	48.111156	11.588909	20.45	5.0	2574.0	Obergiesing-Fasangarten\n20.45 EUR/m2	3	Greek Restaurant	Ice Cream Shop	Gym / Fitness Center	Café	Hotel
17	Untergiesing-Harlaching	48.114963	11.570189	19.66	5.1	1691.0	Untergiesing-Harlaching\n19.66 EUR/m2	3	German Restaurant	Italian Restaurant	Restaurant	Beach	Café
18	Thalkirchen-Obersendling-Forstennied-Fürstent...	48.102840	11.545979	18.84	9.7	3311.0	Thalkirchen-Obersendling-\n18.84 EUR/m2	3	Zoo Exhibit	German Restaurant	Beer Garden	Greek Restaurant	Italian Restaurant
20	Pasing-Obermenzing	48.147785	11.460701	17.83	11.3	3103.0	Pasing-Obermenzing\n17.83 EUR/m2	3	Coffee Shop	Café	German Restaurant	Drugstore	Spa

Figure 18 – Cluster No. 4

Cluster No. 4 Characteristics:

- Biggest cluster
- Distance from the city centre between 3 – 11km
- Average renting price is between 17.83 – 24.15EUR/per m2 – mixed from cheap to expensive
- Low criminality rate, except 'Ludwigsvorstadt-Isarvorstadt'

Cluster No. 5

	Neighborhood_x	Latitude	Longitude	PriceEURm2	Distance_from_citycentre_in_km	CriminalActs/Year	Label	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
22	Allach-Untermenzing	48.176884	11.476058	17.62	13.6	1106.0	Allach-Untermenzing\n17.62 EUR/m2	4	Bus Stop	Supermarket	Italian Restaurant	Home Service	Turkish Restaurant
23	Feldmoching-Hasenberg	48.213804	11.541275	17.12	14.1	2035.0	Feldmoching-Hasenberg\n17.12 EUR/m2	4	Bus Stop	Plaza	Supermarket	Bakery	Café

Figure 18 – Cluster Nr.5

Cluster Nr.5 Characteristics:

- Price Range around 17 EUR/m2
- Good connectivity
- Distance Range around 14 km
- Outside of the city

7. Discussion

Munich – Analysis of the neighbourhoods for newcomers, showed how different the neighbourhoods really are. I would like to highlight several key insights from this analysis:

1. Further from the city centre, cheaper, safer, and emptier the neighbourhoods are.
2. Higher foreigner proportion has nothing to do with criminality rate in Munich,
3. All neighbourhoods close to the 'English garden' Park are generally expensive,
4. Main train Station Neighbourhood (Ludwigsvorstadt-Isarvorstadt) ist most dangerous and also one of the most expensive places to live in Munich,
5. Density proportion may be a good key to assume accommodation availability.

8. Conclusion

In this study, I analysed Munich 25 neighbourhoods, their similarities and differences to help newcomers orientate and find suitable and appropriate place to start their live in Munich. I identified price as a main feature for decision making process. Along price there were another important feature as distance from the city centre, foreigners share, criminality, and density. I discovered the data using statistical approaches as a correlation matrix to show relationships between the features as well as machine learning model using k-mean clustering to show similarities between Munich neighbourhoods. These models can be helpful to compare neighbourhoods of newcomer choice with other maybe cheaper or safer neighbourhoods with same characteristics.

Sources

[1] <https://en.wikipedia.org/wiki/Munich>, 20.07.2020

[2] <https://www.muenchen.de/rathaus/Stadtfinfos/Statistik/Indikatoren-und-Monatszahlen/Indikatorenatlas.html>, 20.07.2020

[3] <https://immobilienmarkt.faz.net/immobilien-ratgeber/miete-teuerste-staedte-deutschland>, 20.07.2020

[4] https://de.wikipedia.org/wiki/Stadtbezirke_M%C3%BCnchens, 20.07.2020

[5] <https://www.muenchen.de/rathaus/Stadtfinfos/Statistik/Sipo.html>, 20.07.2020

[6] https://en.wikipedia.org/wiki/Correlation_and_dependence, 20.07.2020