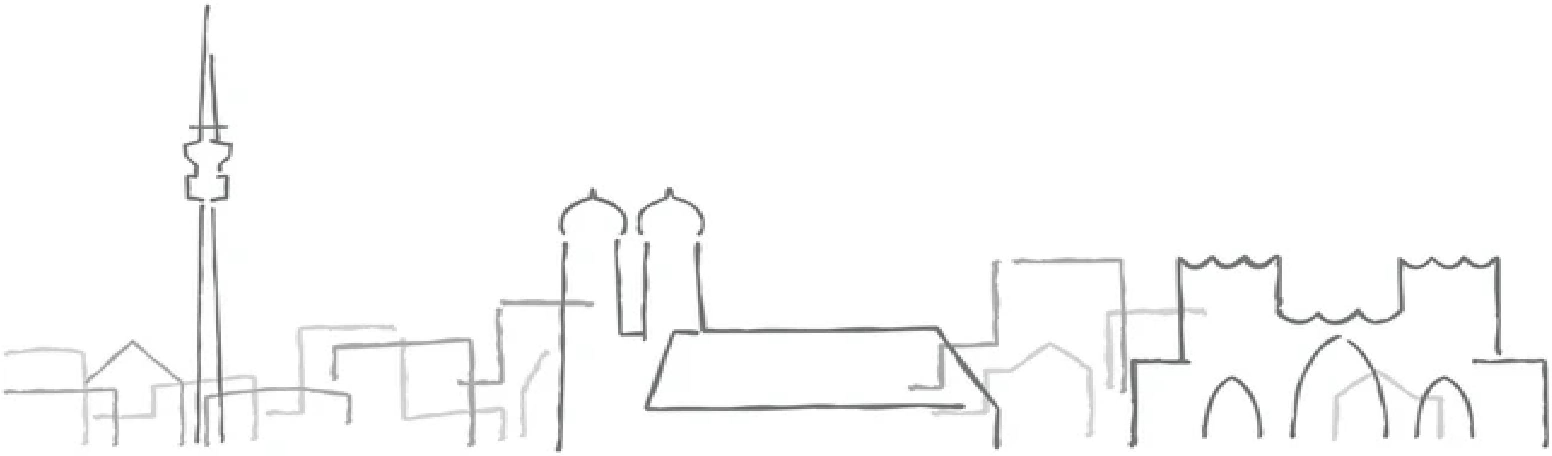
IBM Data Science
Applied Data Science Capstone.



The Battle of Neighborhoods – Munich

Find the best place to stay in Munich – Germany.

BY
FELIPE DEMENECH VASCONCELOS

FEBRUARY 2021

1. Introduction

1.1 Background

Munich is a popular German city, capital of Bavaria and land of the most famous beer festival, the Oktoberfest.
Munich has a population of 1,558,395 inhabitants and the third-largest city in Germany, after Berlin and Hamburg. Munich is a global centre of art, science, technology, finance, publishing, culture, innovation, education, business, and tourism and enjoys a very high standard and quality of living, reaching first in Germany and third worldwide according to the 2018 Mercer survey.

All these great attributes of Munich means that the city is very popular for tourism, mainly at the time of Oktoberfest, and it also means that there are a lot of places available for spend some nights and enjoy the city. But what is the best borough to stay? Wich are the attractives in these boroughs? Where is the best place to have as accomodation and to have the gratest time in munich?

Due to have a short budget, we will focus our accommodation search on Airbnb. Airbnb is an online community service for people to advertise, discover and book accommodation. Airbnb it can be a cheap way to find a place to stay, the offer of places is huge and the tourist can find a place that fits their conditions pretty easy.

1.2 Bussines Problem.

Find a good place to stay some nights in a big city like Munich can be sometimes very hard. The prices variation, the interest spots, restaurants, attractions, transport, localization, and many other attributes are hard to be filtred when we are looking for accommodation. Our problem here is to find the best place to stay in Munich accordingly with the following conditions:

- Price: With a low budget for accommodation, we need to find a place that fits the standards we want without costing much money. For that we have established a maximum charge of €50 per night.

- Proximity to public transport: Ease of mobility is one of the most important points when doing tourism. No One likes to spend lots of precious time inside of subways and buses every time when want to go somewhere, so, our Airbnb place must be near of a subway tation.

- Places of interests: In a trip the tourist must know what they want to do, what they want to see, what places to visit, what experiences to live, what is better to eat and so on. So, in order to find a realy good place to stay, our Airbnb host must be next to some beer gardens and german restaurants (what is better in Germany then drink a german beer and eat some german food?).

2. Data.

The data used in this project was collected from several internet public sources, from Foursquare API and from some python libraries used to get coordinates.

- The Munich bourough's list was obteined from: https://de.wikipedia.org/wiki/Stadtbezirke_Münchens

- The airbnb file with the hosts offer in Munich ware obteined from Inside Airbnb website: http://insideairbnb.com/get-the-data.html

- The Munich U-Bahn stations (subway) list was obteined from: https://en.wikipedia.org/wiki/List_of_Munich_U-Bahn_stations

- The munich boroughs coordinates and the subway stations coordinates were obtained using a python geocode web service called Geocoder (library documentation: https://geocoder.readthedocs.io/#).

- The Foursquare API was used in order to obtein the venues of each Munich borough accordingly with a according to a pre-established search radius (Foursquare API: https://developer.foursquare.com/)

3. Metodology

3.1 Data Acquisition and Transformation.

3.1.1 Munich Boroughs Data.

The Munich boroughs dataset (df_munich) was obtained using the Pandas python library, a very popular python tool used to work with datasets, the first result was the following table (Table 01).

	Nr.	Stadtbezirk	Fläche(km²)	Einwohner	Dichte(Einw./km²)	Ausländer(%)
0	1.0	Altstadt-Lehel	315	21.126	6.716	260
1	2.0	Ludwigsvorstadt-Isarvorstadt	440	51.933	11.799	283
2	3.0	Maxvorstadt	430	51.834	12.060	256
3	4.0	Schwabing-West	436	68.935	15.800	228
4	5.0	Au-Haidhausen	422	61.654	14.611	235

Table 01 – Initial dataframe (first 5 rows).

This first dataset had the columns names in german, who ware translated to english, and contained some unusable columns that were excluded, as well as the population density column, which have an “object” data type and had to be converted to a “float” data type, so they could be be used in further calculations. The following table show us the first five rows of the first dataframe after the first transformations.

	borough	area_km2	residents	density
0	Altstadt-Lehel	315	21.126	6.716
1	Ludwigsvorstadt-Isarvorstadt	440	51.933	11.799
2	Maxvorstadt	430	51.834	12.060
3	Schwabing-West	436	68.935	15.800
4	Au-Haidhausen	422	61.654	14.611

Table 02 – Initial dataframe after transformation (first 5 rows).

Using the Geocoder python library, the coordinates of Munich city and each Munich borough was obteined. After the geocoder process three new columns ware created into the first dataframe and filled with the latitude and longitude from each Munich borough and the search radius column, created in order to be used further in the Foursquare API to get the venues in each neighborhood, updating the first dataframe as showed below (Table 03).

	borough	area_km2	residents	density	lat	long	search_radius
0	Altstadt-Lehel	315	21126	6.716	48.137828	11.574582	10013.0
1	Ludwigsvorstadt-Isarvorstadt	440	51933	11.799	48.130340	11.573366	11835.0
2	Maxvorstadt	430	51834	12.060	48.151092	11.562418	11699.0
3	Schwabing-West	436	68935	15.800	48.168271	11.569873	11781.0
4	Au-Haidhausen	422	61654	14.611	48.128753	11.590536	11590.0

Table 03 – df_munich final version.

After this last transformation in the Munich borough dataframe, we get a dataset with 26 rows and 7 columns and we were able to plot the first map of Munich with the markers of all boroughs of our project.

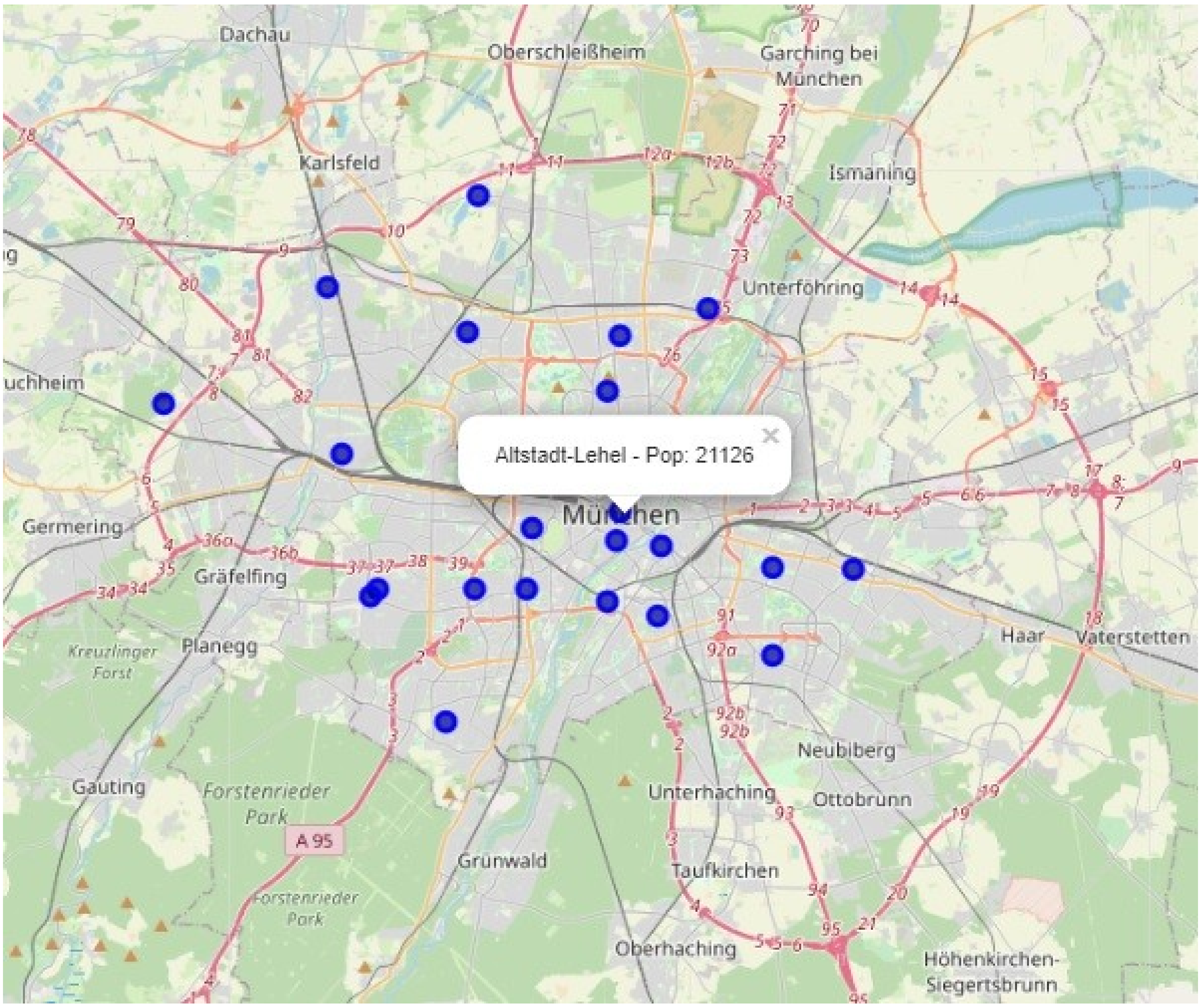


Figure 01 – Munich map with boroughs marks.

3.1.2 – Airbnb Locations Data.

The first Airbnb file ware also obteined using Pandas, and after excluded some unnecessary columns we ware left with a dataset with 5106 rows and 4 columns.

	neighbourhood	latitude	longitude	price
0	Hadern	48.11476	11.48782	80
1	Berg am Laim	48.11923	11.63726	95
2	Maxvorstadt	48.15198	11.56486	99
3	Pasing-Obermenzing	48.13898	11.46612	52
4	Sendling-Westpark	48.10751	11.52733	55

Table 04 – Initial Airbnb dataframe.

This first Airbnb dataframe had a maximum price of €10519 and a minimum of €0 per night of stay, so, to filter this dataframe accordingly to the projetct conditions, we have selected only the rows with prices bigger than €0 until a limit of €50. After that we have obteined a dataset with 1370 hosting options and a price distribution like showed at Graph 01.

Graph 01 – Munich Airbnb Price Disctribution.

After these transformations on the dataset we ware abble to plot a map from Munich with the 1370 hosting options. The Figure 02 figure shows us a center area from Munich with the accommodation options and the price of each one.

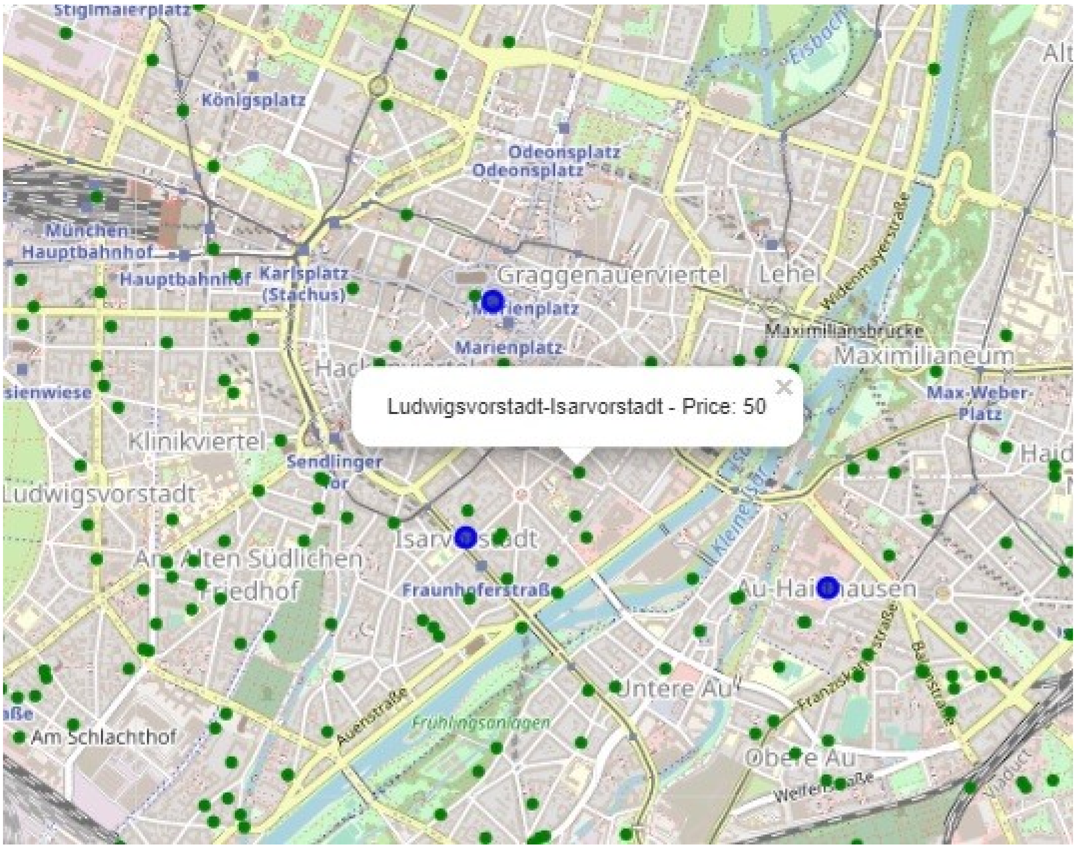


Figure 02 – Center Zone of Munich with Hosting Options.

3.1.3 – Foursquare API Data.

After we have obteined the Airbnb places locations, the next step would be to obtain the location of Beer Gardens and German Restaurants, according to the conditions imposed in the project. To do so we used the request python library and the Foursquare API in order to obtein the venues categoryes inside each borough, than we filtered the obteined dataset to show us only beer gardens and german restaurants. We let the original Fourquare API dataset saved in order to use later, for classification purposes. The next map (Figure 03) show us the distribution for this venues in Munich city, along with the Airbnb accomodations.

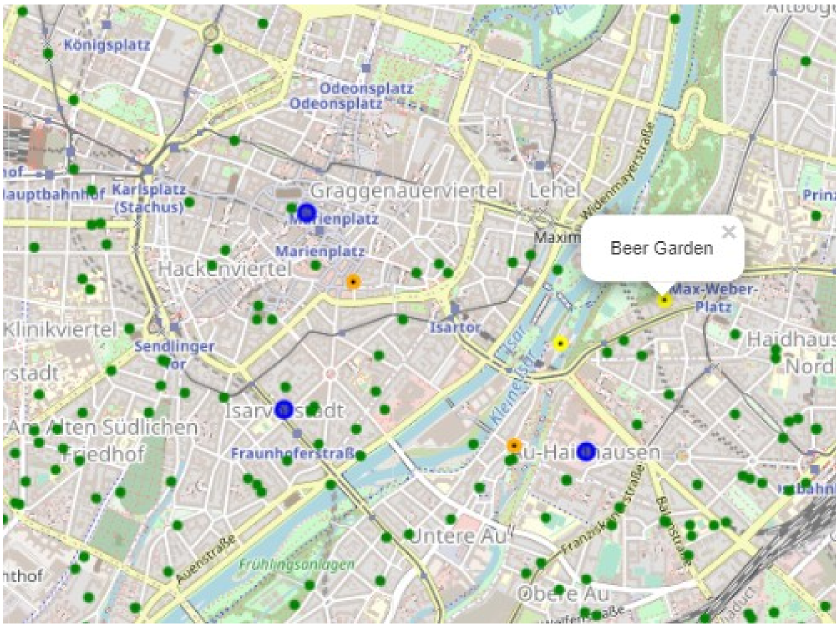


Figure 03 – Beer Garden and German Restaurants locations.

3.1.4 – U-Bahn Sattions (subway) Coordinates.

The U-Bahn stations coordinates ware obteined from wikpedia as well as the other datasets, using Pandas and Geocoder libraryes to obtain the first dataset with the U-Bahn Stations names and them the U-Bahn Stations coordinates. The result ware the following table (Table 05).

	Station	Lines	lat	long
0	Aidenbachstraße	U3	48.097887	11.525195
1	Alte Heide	U6	48.178552	11.602555
2	Am Hart	U2	48.195925	11.571815
3	Arabellapark†	U4	48.153555	11.622063
4	Basler Straße	U3	48.091273	11.491169

Table 05 – U-Bahn Stations Dataset.

With the U-Bahn stations coordinates, we ware abble to plot our last map, summing the Airbnb accomodations locations, the beer gardens and german restaurants locantions and the U-Bahn stations locations, allowing us to visually analyze which places to stay fit into the project criteria.

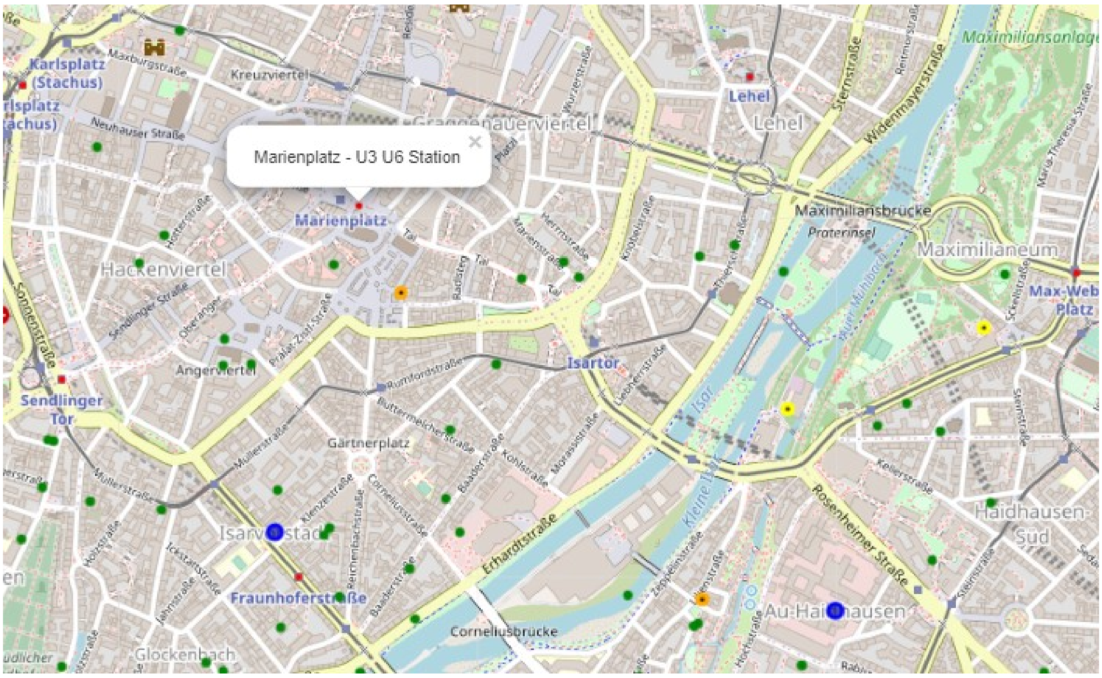


Figure 04 – U-Bahn Stations.

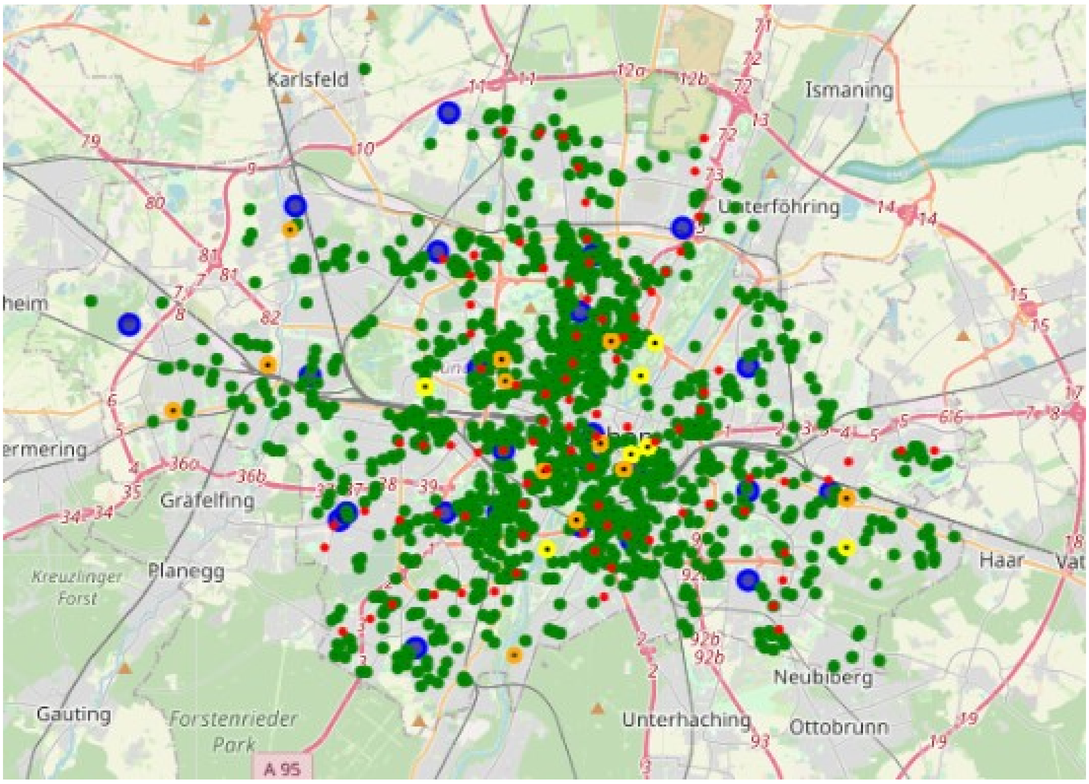
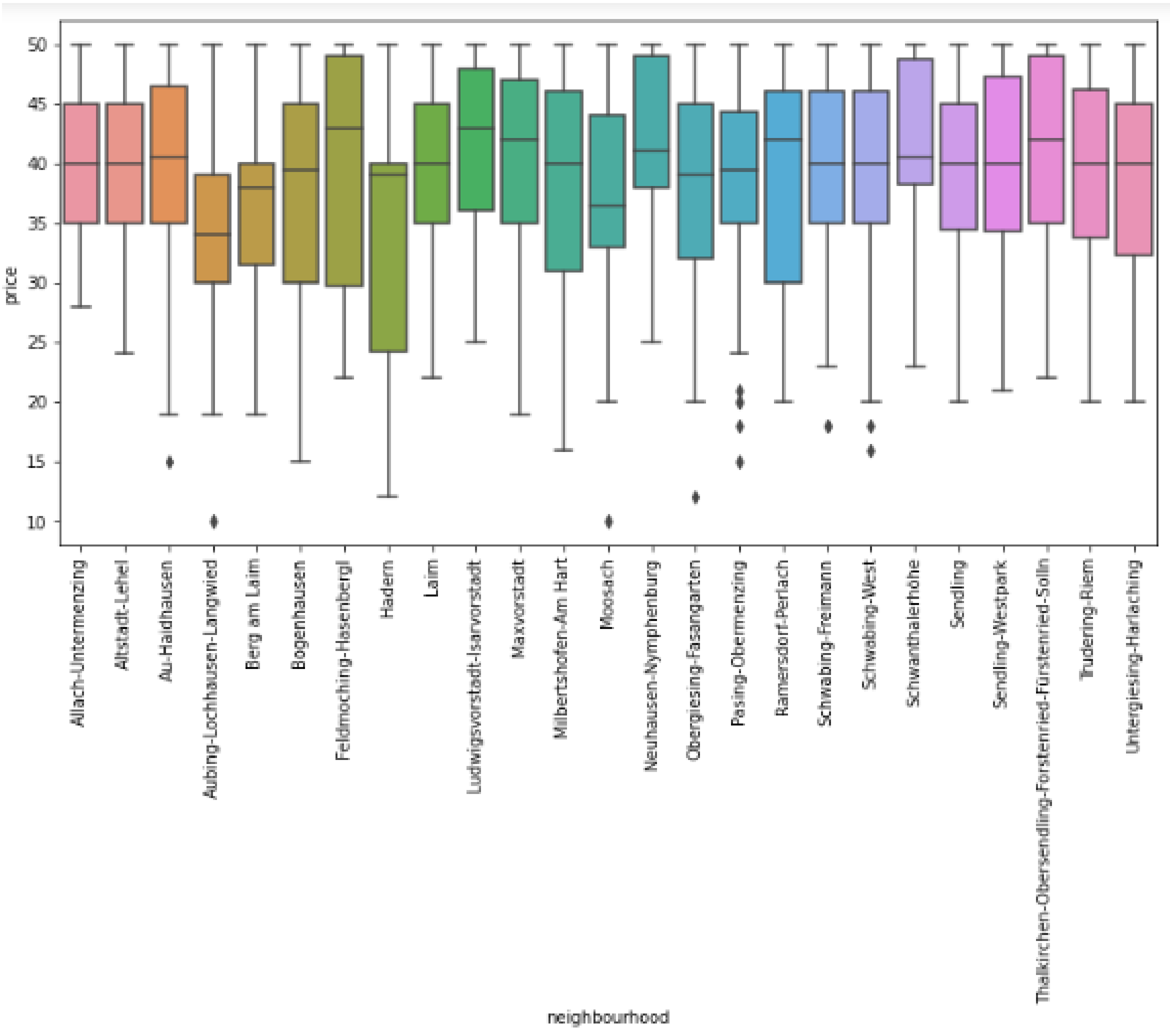


Figure 05 – Munich Map – Airbnb, Interesst Spots, U-Bahn Stations.

3.2 EDA – Exploratory Data Analysis.

First we took a look at the distribution of hosting prices for the entire filtered Airbnb dataset, as we can see, the maximum accommodation values per night were set at € 50, so we can evaluate the average values per district, the minimum values and also the existence of some outliers. We can see that “Aubing-Lochhausen-Langwied“ has the lowest Airbnb prices per night.

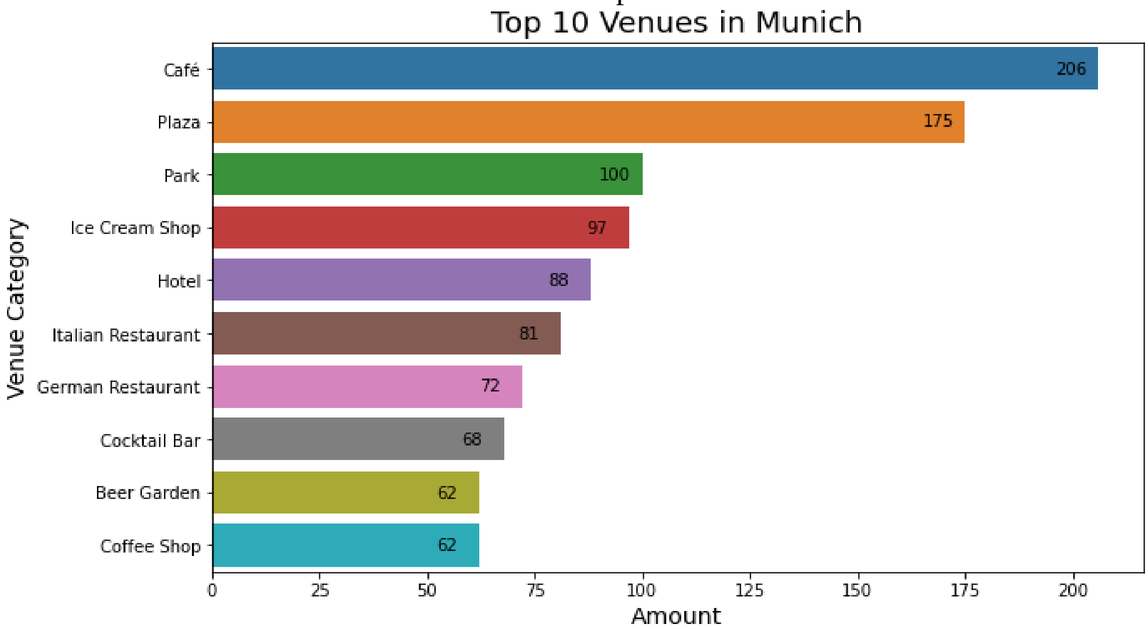


Graph 02 – Munich Airbnb Prices Distribution per Borough (filtered dataset).

The next table and graph shows us the amount of the top 10 venues in Munich, and how we can see the beer gardens and german restaurants, our venues of interests, are in the top 10 list.

	Venue Category	Count
0	Café	206
1	Plaza	175
2	Park	100
3	Ice Cream Shop	97
4	Hotel	88
5	Italian Restaurant	81
6	German Restaurant	72
7	Cocktail Bar	68
8	Beer Garden	62
9	Coffee Shop	62

Table 06 – Top 10 Venues.



Graph 03 – Top 10 Venues in Munich.

In the Graph 03 we can see that Cafés and plazas are the most common venues in Munich, but it is not new, what catches our attention is that the third position are Parks, Munich has a great amount of urban parks and it is very nice, it shows us that Munich has a great quantity of green.
But what really interest us are the beer gardens and the German restaurants, there are a good amount of each, being 72 German restaurants and 62 beer gardens.

3.3 K-means Clustering

K-means clustering is an unsupervised algorithm that divides data into k non-overlapping clusters without any internal cluster structure or labels. The objects in a cluster are similar to each other and different from other objects in other clusters.
The following 5 tables shows us the 5 clusters distributions accordingly with K-means algorithm.

	Cluster Labels	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	Untergiesing-Harlaching	Café	Plaza	Hotel	Beer Garden	Cocktail Bar	Beach	Ice Cream Shop	Italian Restaurant	German Restaurant	Concert Hall
1	0	Altstadt-Lehel	Café	Plaza	Cocktail Bar	Ice Cream Shop	Hotel	Coffee Shop	Beer Garden	Burger Joint	German Restaurant	Monument / Landmark
2	0	Au-Haidhausen	Plaza	Café	Cocktail Bar	Hotel	Ice Cream Shop	Gourmet Shop	Greek Restaurant	Beer Garden	Restaurant	Coffee Shop
3	0	Ramersdorf-Perlach	Plaza	Café	Beer Garden	Greek Restaurant	Gourmet Shop	Beach	Hotel	Cocktail Bar	German Restaurant	Park
4	0	Berg am Laim	Café	Plaza	Beer Garden	Gourmet Shop	Hotel	German Restaurant	Greek Restaurant	Restaurant	Beach	Cocktail Bar
5	0	Obergiesing-Fasangarten	Café	Plaza	Cocktail Bar	Hotel	Beach	Beer Garden	Park	Italian Restaurant	Ice Cream Shop	Gourmet Shop
6	0	Trudering-Riem	Plaza	Beer Garden	Café	Ice Cream Shop	Park	Greek Restaurant	Coffee Shop	German Restaurant	Restaurant	Gourmet Shop
7	0	Ludwigsvorstadt-Isarvorstadt	Plaza	Café	Ice Cream Shop	Hotel	Beach	Cocktail Bar	Greek Restaurant	Pastry Shop	Surf Spot	River

Table 07 – Cluster 0.

	Cluster Labels	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
8	1	Aubing-Lochhausen-Langwied	Park	Ice Cream Shop	Plaza	Café	Italian Restaurant	Coffee Shop	Pastry Shop	Beach	Cocktail Bar	Beer Garden
9	1	Feldmoching-Hasenbergl	Café	Italian Restaurant	Park	Plaza	Hotel	Ice Cream Shop	Beer Garden	German Restaurant	Bakery	Coffee Shop
10	1	Moosach	Café	Plaza	Park	Italian Restaurant	Ice Cream Shop	German Restaurant	Hotel	Coffee Shop	Bakery	River
11	1	Neuhausen-Nymphenburg	Café	Plaza	Park	Italian Restaurant	Ice Cream Shop	Hotel	Coffee Shop	Pastry Shop	Bistro	Beer Garden
12	1	Allach-Untermenzing	Park	Café	Italian Restaurant	Plaza	Ice Cream Shop	German Restaurant	Hotel	Pastry Shop	Coffee Shop	Greek Restaurant
13	1	Pasing-Obermenzing	Plaza	Park	Café	Ice Cream Shop	Italian Restaurant	Beach	Coffee Shop	Beer Garden	Hotel	Opera House

Table 08 – Cluster 1.

	Cluster Labels	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
14	2	Hadern	Café	Plaza	Park	Bakery	Italian Restaurant	German Restaurant	Cocktail Bar	Beach	Ice Cream Shop	Opera House
15	2	Schwanthalerhöhe	Café	Plaza	Ice Cream Shop	Hotel	Cocktail Bar	German Restaurant	Beach	Park	Bakery	Bar
16	2	Sendling	Café	Plaza	Hotel	German Restaurant	Cocktail Bar	Park	Bakery	Beach	Ice Cream Shop	Restaurant
17	2	Sendling-Westpark	Café	Plaza	Park	German Restaurant	Ice Cream Shop	Beach	Cocktail Bar	Bakery	Italian Restaurant	Hotel
18	2	Thalkirchen-Obersendling-Forstenried-Fürstenri...	Café	Park	German Restaurant	Plaza	Hotel	Beach	Cocktail Bar	Greek Restaurant	Ice Cream Shop	Bakery

Table 09 – Cluster 2.

	Cluster Labels	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
19	3	Laim	Italian Restaurant	Hotel	Café	Trentino Restaurant	Bar	Cocktail Bar	Diner	Restaurant	Pizza Place	Historic Site

Table 10 – Cluster 3.

	Cluster Labels	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
20	4	Bogenhausen	Plaza	Café	Park	Beer Garden	German Restaurant	Ice Cream Shop	Coffee Shop	Gourmet Shop	Hotel	Italian Restaurant
21	4	Schwabing-Freimann	Café	Plaza	Hotel	Park	Ice Cream Shop	Beer Garden	German Restaurant	Gourmet Shop	Coffee Shop	Pastry Shop
22	4	Schwabing-West	Plaza	Café	Park	Hotel	Ice Cream Shop	Coffee Shop	German Restaurant	Gourmet Shop	Italian Restaurant	Cocktail Bar
23	4	Milbertshofen-Am Hart	Plaza	Café	Park	Ice Cream Shop	Coffee Shop	Italian Restaurant	German Restaurant	Gourmet Shop	Hotel	Beer Garden
24	4	Maxvorstadt	Café	Plaza	Ice Cream Shop	Hotel	Coffee Shop	Park	Italian Restaurant	Cocktail Bar	German Restaurant	Vietnamese Restaurant

Table 11 – Cluster 4.

Based on the clusters distributions, proximity from city center and looking for our places of interests we have selected the best 2 boroughs to find an Airbnb accomodation. The Table 12, below, shows us the top 2 boroughs in Munich who meets our interests point criteria.

	Cluster Labels	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	Untergiesing-Harlaching	Café	Plaza	Hotel	Beer Garden	Cocktail Bar	Beach	Ice Cream Shop	Italian Restaurant	German Restaurant	Concert Hall
1	0	Altstadt-Lehel	Café	Plaza	Cocktail Bar	Ice Cream Shop	Hotel	Coffee Shop	Beer Garden	Burger Joint	German Restaurant	Monument / Landmark

Table 12 – Top 09 Munich Boroughs.

Filtering the Airbnb options dataset based on the top 2 boroughs to stay in Munich we have 55 accommodation options to choose from, as shown in the Table 13.

	neighbourhood	latitude	longitude	price
0	Untergiesing-Harlaching	48.10517	11.57388	36
1	Untergiesing-Harlaching	48.11846	11.57047	45
2	Untergiesing-Harlaching	48.10919	11.57485	40
3	Untergiesing-Harlaching	48.10096	11.58102	39
4	Untergiesing-Harlaching	48.10127	11.56051	35
5	Untergiesing-Harlaching	48.11826	11.57060	39
6	Untergiesing-Harlaching	48.10711	11.56805	49
7	Untergiesing-Harlaching	48.11555	11.57393	50
8	Untergiesing-Harlaching	48.09498	11.56250	45
9	Untergiesing-Harlaching	48.10505	11.57277	45
10	Untergiesing-Harlaching	48.11641	11.57342	28
11	Untergiesing-Harlaching	48.10163	11.55765	40
12	Untergiesing-Harlaching	48.10726	11.57238	30
13	Untergiesing-Harlaching	48.10856	11.57285	43
14	Untergiesing-Harlaching	48.08777	11.56230	36
15	Untergiesing-Harlaching	48.11471	11.56895	40
16	Untergiesing-Harlaching	48.10425	11.57501	50
17	Untergiesing-Harlaching	48.11025	11.57794	50
18	Untergiesing-Harlaching	48.10657	11.57246	20
19	Untergiesing-Harlaching	48.08584	11.54904	32
20	Untergiesing-Harlaching	48.11507	11.57042	44
21	Untergiesing-Harlaching	48.10537	11.55968	35
22	Untergiesing-Harlaching	48.11556	11.56715	44
23	Untergiesing-Harlaching	48.08812	11.56062	33
24	Untergiesing-Harlaching	48.10587	11.57585	49
25	Untergiesing-Harlaching	48.11096	11.57753	44
26	Untergiesing-Harlaching	48.08869	11.55841	25
27	Untergiesing-Harlaching	48.10168	11.55965	45
28	Untergiesing-Harlaching	48.10642	11.57177	21
29	Untergiesing-Harlaching	48.10848	11.57677	40
30	Untergiesing-Harlaching	48.09558	11.57498	45
31	Untergiesing-Harlaching	48.10649	11.57493	31
32	Untergiesing-Harlaching	48.11941	11.57483	30
33	Untergiesing-Harlaching	48.11385	11.56616	38
34	Untergiesing-Harlaching	48.10506	11.57951	43
35	Untergiesing-Harlaching	48.11631	11.57509	49
36	Untergiesing-Harlaching	48.10558	11.56383	27
37	Untergiesing-Harlaching	48.10583	11.57843	30
38	Altstadt-Lehel	48.13641	11.56998	43
39	Altstadt-Lehel	48.13623	11.58744	25
40	Altstadt-Lehel	48.13378	11.57262	50
41	Altstadt-Lehel	48.13599	11.58636	40
42	Altstadt-Lehel	48.13583	11.57515	24
43	Altstadt-Lehel	48.13568	11.58896	35
44	Altstadt-Lehel	48.13431	11.57183	30
45	Altstadt-Lehel	48.14850	11.59499	40
46	Altstadt-Lehel	48.13554	11.58092	40
47	Altstadt-Lehel	48.13824	11.56791	45
48	Altstadt-Lehel	48.14523	11.59566	45
49	Altstadt-Lehel	48.13555	11.58269	49
50	Altstadt-Lehel	48.13376	11.57179	39
51	Altstadt-Lehel	48.13589	11.58219	39
52	Altstadt-Lehel	48.14059	11.57046	45
53	Altstadt-Lehel	48.13586	11.56921	30
54	Altstadt-Lehel	48.13805	11.57374	50

Table 13 – 55 Airbnb Accomodation Options.

The following map shows us the distribution from the 55 Airbnb last options to choice. The 2 red circled areas are our main radius of search where the black circle represents the borough center, the green point represents the Airbnb accomodations, the orange markers are the points of interest and finaly, the red point represents the U-Bahn stations.

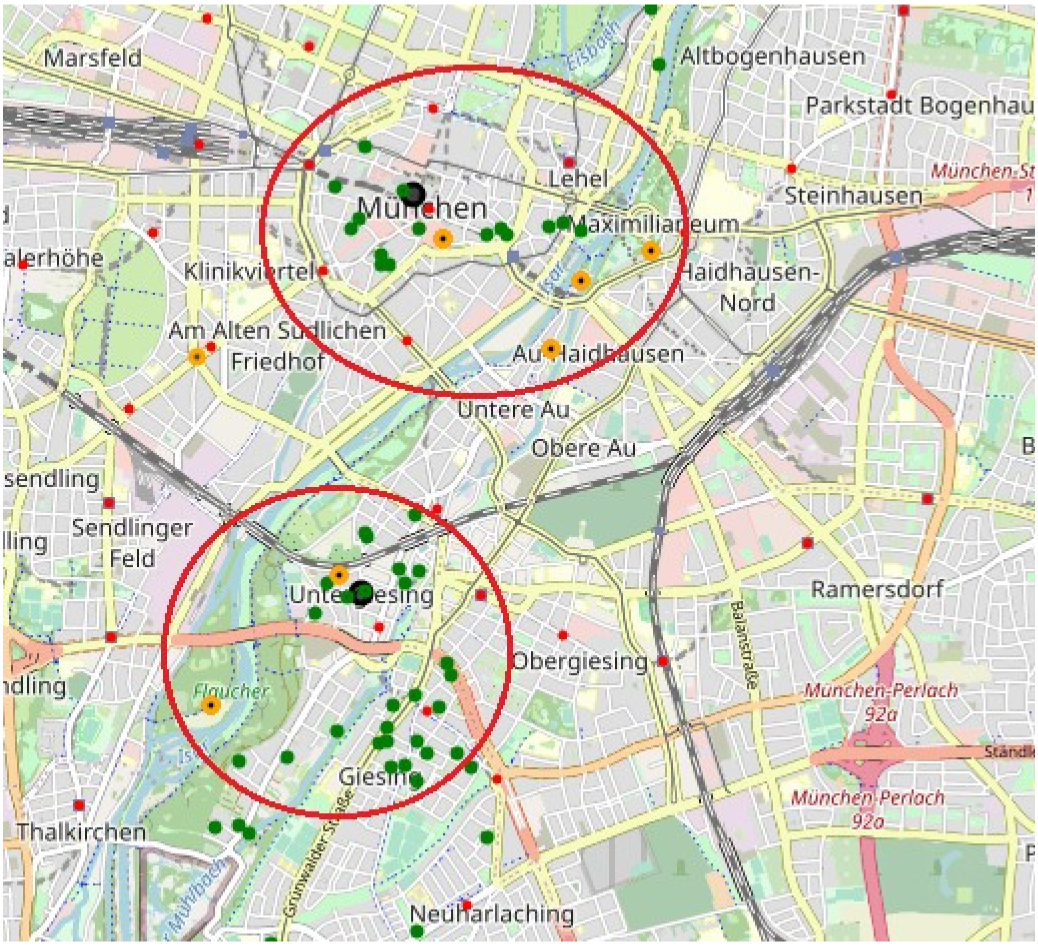


Figure 06 – Final Munich Airbnb Search Map.

4. Conclusion.

This capstone project was very challenging and very fun to work. In order to solve the problem we had to search and collect lots of information from the Internet doing web scrapping, getting coordinates using Geopy and learn how to use and request infos using the Foursquare API, and the most important, all of that using python tools.

After the load, cleaning and filtering the datasets and get a "feeling" of how we could solve our business problem. To do so we made use of excellent tools like Pandas, Numpy, Folium, Geopy and some Machine Learning algorithms, like K-means, which are very powerful python tools that made our work much easier.

With all the hard work done, the next fase of the project was to determine which are the best Airbnb accomodations to stay in Munich based on our project criterias. At the end we "filtered", using python and Machine Learning, the best 55 Airbnb options. Our initial Airbnb dataset started with 5106 rows and ended with just 55 options, making our job to find a best place very easyer.

In summary, the project objective was achieved, we were able to facilitate the Airbnb search using Machine Learning and python as tools, but it was not the main objective of the IBM Capstone Project, the main object was to get knolege about how this tools work, how we can use them, how it can facilitate our work and our life, how we can solve problems using it.

5. References.

Foursquare API. (23 February 2021). Foursquare/developers. Retrieved from: https://developer.foursquare.com

Geocoder. (23 February 2021). Geocoder: Simple, Consistent. Retrieved from: https://geocoder.readthedocs.io

Inside Airbnb. (23 February 2021). Get the Data. Retrieved from: insideairbnb.com/get-the-data.html

Wikipedia. (23 February 2021). Stadtbezirke Münchens. Retrieved from Wikipedia: https://de.wikipedia.org/wiki/Stadtbezirke_Münchens

Wikipedia. (23 February 2021). List of Munich U-Bahn stations. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/List_of_Munich_U-Bahn_stations