Battle of Neighbourhoods in Munich

Introduction and business problem

Munich ist the capital of the federal state of Bavaria, Germany. With its approximately 1.5 million inhabitants and many more in the metropolitan region it is the most densely populated area in Bavaria and one of the most densely populated areas in Germany.

Today, Munich is a global centre of art, science, technology, culture, innovation 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.

The Bavarian cuisine contributes to this ranking, most popular is the annual *Oktoberfest* where millions of people enjoy the Bavarian cuisine. In small, there are countless traditional Bavarian restaurants all over the city area, many of which also have small outside areas called beer garden. These are popular fixtures of Munich's gastronomic landscape. They are central to the city's culture and serve as a kind of melting pot for members of all walks of life, for locals and tourists alike.

The objective of this project is to use Foursquare location data and regional clustering of venue information as well as information about the population and the number of possible AirB&B apartments to determine what might be the 'best' neighbourhood in Munich to open a restaurant. Although there are many Bavarian restaurants in the city area, in 2019 tourism in Munich recorded 8.8 million arrivals and surely many of them want to taste the traditional Bavarian cuisine.

Through this project, we will find the most suitable location for an entrepreneur to open a new Bavarian restaurant in Munich.

Data Description

The data that will be required will be a combination of .csv, .xlsx and .json files that have been prepared for the purposes of the analysis from multiple sources:

1. list of neighbourhoods in Munich from wikipedia.org in combination with the geographical location of the neighbourhoods from https://geohack.toolforge.org/
2. zip codes were collected via https://www.muenchen.de/. Munich has for every neighbourhood more than one zip code, so to make things easier only one zip code for every neighbourhood was chosen.

Neighbourhood	Postal Code	Latitude	Longitude
Altstadt-Lehel	80331	48.136111	11.572222
Ludwigsvorstadt-Isarvorstadt	80335	48.127222	11.564722
Maxvorstadt	80333	48.150000	11.569444
Schwabing-West	80809	48.161111	11.568889
Au-Haidhausen	81543	48.131944	11.588889

3. the geographical data of the neighbourhood borders and apartments per neighbourhood via the AIR'B&B data base, http://insideairbnb.com/get-the-data.html
4. demographic information, e.g. population and density from https://www.opengov-muenchen.de/

Neighbourhood	Population	Area	Density	Percentage
Altstadt-Lehel	20422	314.57	65	1.39
Ludwigsvorstadt-Isarvorstadt	50620	440.14	115	3.46
Maxvorstadt	51642	429.79	120	3.53
Schwabing-West	65892	436.30	151	4.50
Au-Haidhausen	59752	421.96	142	4.08

5. venue data pertaining to Bavarian restaurants via Foursquare API. The venue data in combination with population data will help find which neighbourhood is best suitable to open a traditional Bavarian restaurant.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Altstadt-Lehel	48.136111	11.572222	Asamkirche (St. Johann Nepomuk)	48.135053	11.569746	Church
1	Altstadt-Lehel	48.136111	11.572222	Kustermann	48.136242	11.574897	Department Store
2	Altstadt-Lehel	48.136111	11.572222	Galeria Gourmet	48.137432	11.573217	Gourmet Shop
3	Altstadt-Lehel	48.136111	11.572222	Hirmer	48.138023	11.572046	Men's Store
4	Altstadt-Lehel	48.136111	11.572222	La Burrita	48.136143	11.574489	Burrito Place

6. listings from AIR'B&B database to get the number of theoretical available accomodations. This stands for the amount of tourist appearances in each neighbourhood. http://insideairbnb.com/get-the-data.html

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	reviews_per_month
0	36720	Beautiful 2 rooms flat, Glockenbach	158413	Gabriela	NaN	Ludwigsvorstadt-Isarvorstadt	48.13057	11.56929	Entire home/apt	95	2	25	2017-07-22	0.34
1	97945	Deluxw-Apartm. with roof terrace	517685	Angelika	NaN	Hadern	48.11476	11.48782	Entire home/apt	80	2	131	2019-10-03	1.23
2	114695	Apartment Munich/East with sundeck	581737	Stephan	NaN	Berg am Laim	48.11923	11.63726	Entire home/apt	95	2	53	2019-10-06	0.49
3	127383	City apartment next to Pinakothek	630556	Sonja	NaN	Maxvorstadt	48.15198	11.56486	Entire home/apt	120	2	84	2020-03-01	0.76
4	157808	Near Olympia,English Garden	759734	Christian	NaN	Schwabing-West	48.16381	11.56089	Private room	35	1	0	NaN	NaN

Afterwards the location data got merged with the population data for further analysis.

Neighbourhood	Population	Area	Density	Percentage	Postal Code	Latitude	Longitude
Altstadt-Lehel	20422	314.57	65	1.39	80331	48.136111	11.572222
Ludwigsvorstadt-Isarvorstadt	50620	440.14	115	3.46	80335	48.127222	11.564722
Maxvorstadt	51642	429.79	120	3.53	80333	48.150000	11.569444
Schwabing-West	65892	436.30	151	4.50	80809	48.161111	11.568889
Au-Haidhausen	59752	421.96	142	4.08	81543	48.131944	11.588889
Sendling	39953	393.87	101	2.73	80336	48.121389	11.541389
Sendling-Westpark	55405	781.45	71	3.78	81377	48.122222	11.531944
Schwanthalerhoehe	29663	207.02	143	2.02	80339	48.138900	11.541700
Neuhausen-Nymphenburg	95906	1291.45	74	6.55	80634	48.156944	11.516667
Moosach	51537	1109.36	46	3.52	80997	48.183333	11.516667
Milbertshofen-Am Hart	73617	1341.64	55	5.03	80937	48.198333	11.576389
Schwabing-Freimann	69676	2567.22	27	4.76	80939	48.180556	11.602778
Bogenhausen	82138	2370.97	35	5.61	81675	48.148056	11.616667
Berg am Laim	43068	631.46	68	2.94	81671	48.122222	11.627778
Trudering-Riem	67009	2245.05	30	4.57	81735	48.116667	11.658333
Ramersdorf-Perlach	108244	1989.50	54	7.39	81739	48.102800	11.625000
Obergiesing	51499	572.04	90	3.52	81547	48.111111	11.594444
Untergiesing-Harlaching	51937	805.67	64	3.55	81545	48.092222	11.560556
Thalkirchen-Obersendling-Forstenried-Fuerstenried-Solln	90790	1776.31	51	6.20	81477	48.089000	11.519000
Hadern	48945	922.37	53	3.34	80689	48.110556	11.465278
Pasing-Obermenzing	70783	1649.78	43	4.83	80687	48.145800	11.459700
Aubing-Lochhausen-Langwied	42305	3406.02	12	2.89	81243	48.158333	11.419444
Allach-Untermenzing	30737	1545.17	20	2.10	80995	48.200000	11.452778
Feldmoching-Hasenbergl	59391	2893.78	21	4.05	80933	48.211100	11.541700
Laim	54030	528.59	102	3.69	80686	48.140278	11.497222

All data, except the json File, were loaded into Pandas dataframes as shown above.

Methodology

The first step ist o preprocess the data to bring it in the desired form. Thus, index and column names are checked and equalized due to different labeling within the data. After that, AirB&B data with the number of apartments (Num_Listings) got merged with Postal Codes dataframe and Munich neighbourhood data.

Then Fouraquare API was used to get venue data in an radius of 1000 meters around the central point of each of the 25 neighbourhoods. Around Munich, there are 1362 venues separated in 215 unique venue categories.

To apply the later used Machine Learning Clustering algorithm KMeans, One hot encoding is used to convert categorical variables into a form that could be provided to Machine Learning algorithms to do a better job in prediction.

After one hot encoding, some venue categories were summarized due to similar meanings. This is an exception fort he city of munich, because beer gardens as well as German restaurants offer neraly the same as Bavarian restaurants. So Bavarian and German restaurants and beer gardens were summed up to get the absolute number of Bavarian restaurants per neighbourhood as well as the importance.

To get an better feeling for each neighbourhood and also the importance of Bavarian restuarants in all neighbourhoods, we put the top five venues per neighbourhood in tables with their frequency. Afterwards we created a table with the top ten venues per neighbourhood.

With these preprocessing steps and with a feeling oft he data, we apply the KMeans Clustering algorithm with three clusters due to the low number of neighbourhoods. Therefore, we merged the data with overall information about munich with the dataframe with the most common venues.

Results

Figure 1 and 2 show the city of Munich as a Choroplethmap divided by neighbourhood borders. There we can see that there is a kind of inner cluster and a kind of outer cluster like two circles.

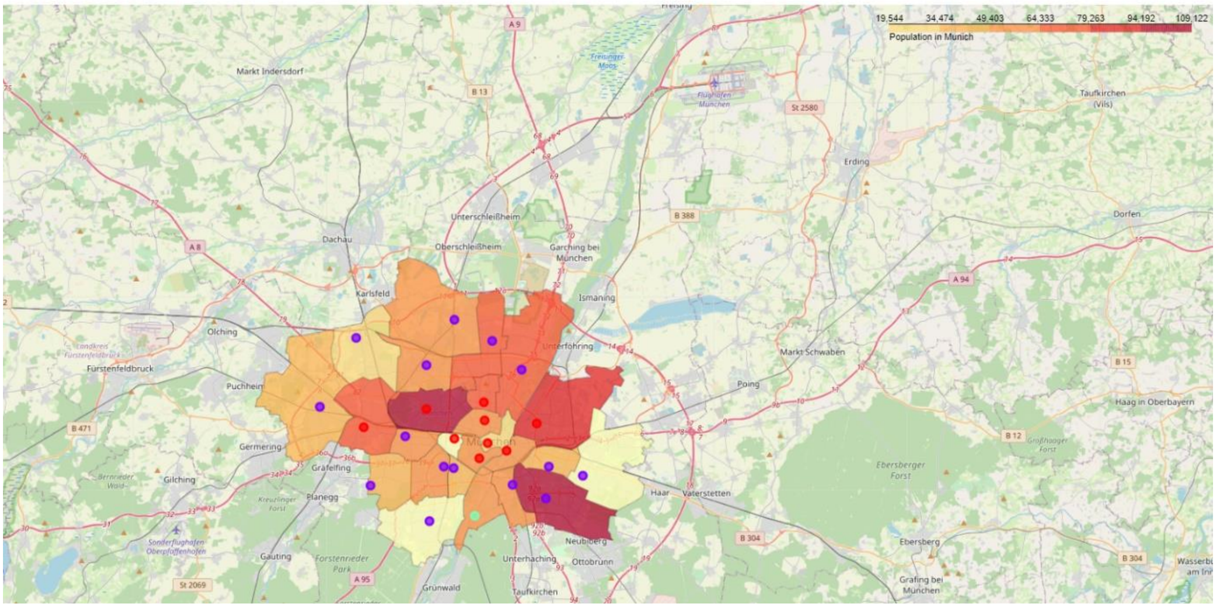

Figure 1: Choropleth Map of Munich showing the three clusters as circles and the heatmap of population

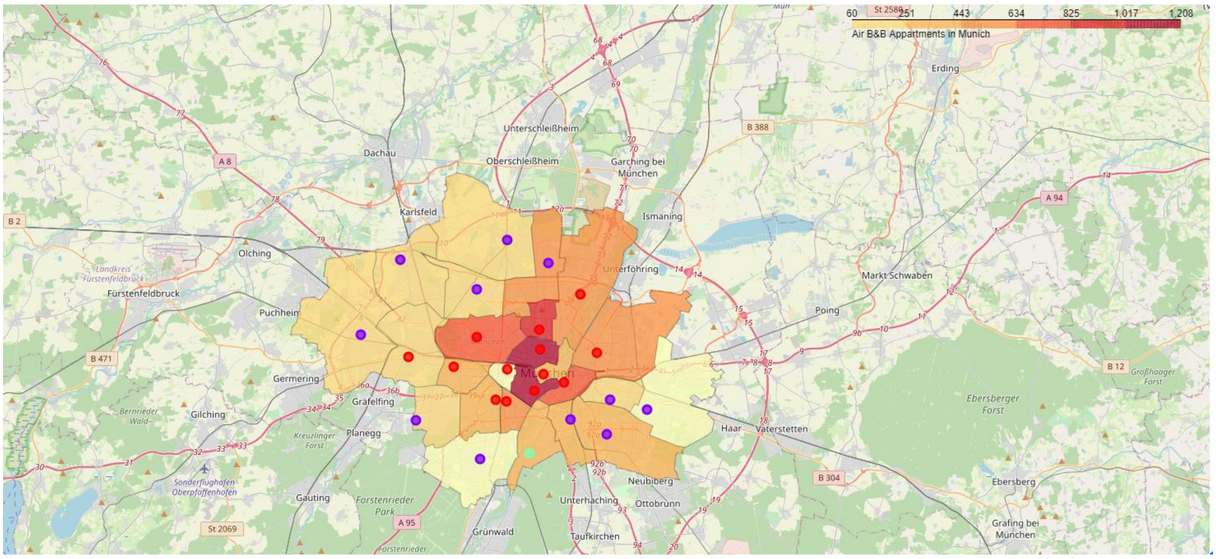

Figure 2: Choropleth Map of Munich showing the three clusters as circles and the heatmap of AirB&B apartments

Figure 2 shows the sorted neighbourhoods by the number of listings. The top three neighbourhoods are Ludwigvorstadt-Isarvorstadt, Maxvorstadt and Schwabing-West. These neighbourhoods have also the highst number of Bavarian restaurants. The goal is to consider if we take the population or the number of apartments which should be compared to the number of Bavarian restaurants and which cluster might fit better, the „inner“ or the „outer“.

The linear regression in Figure 3 shows a positive correlation between the number of listings of AirB&B apartments and Bavarian restuarants, although the variance is very high and isn’t a very good model due to low R^2.

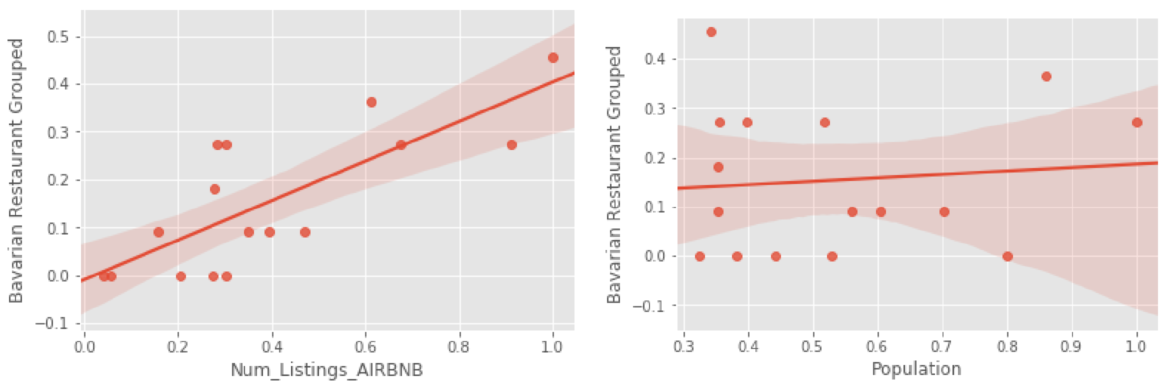

Figure 3: Linear Regression of number of apartments (left) and Population (right) against the number of Bavarian restaurants

So there is a stronger correlation between tourism and the number of Bavarian restaurants as the population. This leads to look at the lowest ratio of Bavarian restaurants and number of apartments

Neighbourhood	Population	Area	Density	Num_Listings_AIRBNB	Bavarian Restaurant Grouped	Ratio_AIRBNB	Ratio_Pop
Laim	54030	528.59	102	380	0	0.000000	0.000000
Hadern	48945	922.37	53	136	0	0.000000	0.000000
Thalkirchen-Obersendling-Forstenried-Fuerstenried-Solln	90790	1776.31	51	410	0	0.000000	0.000000
Trudering-Riem	67009	2245.05	30	302	0	0.000000	0.000000
Feldmoching-Hasenbergl	59391	2893.78	21	116	0	0.000000	0.000000
Schwabing-Freimann	69676	2567.22	27	601	1	0.001664	0.000014
Bogenhausen	82138	2370.97	35	516	1	0.001938	0.000012
Milbertshofen-Am Hart	73617	1341.64	55	465	1	0.002151	0.000014
Maxvorstadt	51642	429.79	120	1097	3	0.002735	0.000058
Schwabing-West	65892	436.30	151	831	3	0.003610	0.000046
Moosach	51537	1109.36	46	250	1	0.004000	0.000019
Ludwigsvorstadt-Isarvorstadt	50620	440.14	115	1197	5	0.004177	0.000099
Obergiesing	51499	572.04	90	384	2	0.005208	0.000039
Neuhausen-Nymphenburg	95906	1291.45	74	761	4	0.005256	0.000042
Ramersdorf-Perlach	108244	1989.50	54	413	3	0.007264	0.000028
Sendling-Westpark	55405	781.45	71	391	3	0.007673	0.000054
Sendling	39953	393.87	101	447	4	0.008949	0.000100
Aubing-Lochhausen-Langwied	42305	3406.02	12	91	1	0.010989	0.000024
Untergiesing-Harlaching	51937	805.67	64	347	4	0.011527	0.000077
Au-Haidhausen	59752	421.96	142	713	9	0.012623	0.000151
Berg am Laim	43068	631.46	68	210	3	0.014286	0.000070
Pasing-Obermenzing	70783	1649.78	43	227	4	0.017621	0.000057
Schwanthalerhoehe	29663	207.02	143	430	9	0.020930	0.000303
Allach-Untermenzing	30737	1545.17	20	71	2	0.028169	0.000065
Altstadt-Lehel	20422	314.57	65	386	11	0.028497	0.000539

Conclusion

To summarize the findings and give the entrepreneurs a help in decision making. If they take the number of listings of AirB&B apartments into consideration as example for the tourism sector in Munich, which was the key point in the business problem they should open their restaurant firstly in Laim, secondly in Hadern or thirdly in Thalkirchen-Obersendling-Forstenried-Fuerstenried-Solln, which are all in Cluster 1 and so in the outer areas of Munich. This gives the best ratio, because in the inner city are already many bavarian restaurants.

To conclude, with the help of multiple datasets from different platforms it was possible after the step of preprocessing the data to apply a Machine Learning algorithm as well as a simple linear regression to give data driven advice for someone who has a specific problem that can be solved through data science methods.