Battle of Neighborhoods (Week 1 & Week 2)

Introduction

Munich is the most important center of commerce and industry in Germany and among top in Europe. Not only because of that, it is expected that the population of Munich will increase by more than 10% within the next 20 years (source).

(By Reinald Kirchner - originally posted to Flickr as Frauenkirche in München, CC BY-SA 2.0)

This project will investigate the prospect of launching an Italian restaurant in the major business districts of Munich. On one hand, the target audience are indeed business people seeking to open a restaurant. On the other hand, results of the analysis also provide hints on where to find interesting restaurants in Munich.

Data

In this work, both data from Wikipedia, as well as, Foursquare location data will be used.

Web-scrapping

There is a dedicated Wikipedia page on the districts of Munich. It includes a table providing for each district additional information like area and population. Additionally, also the official district number is provided which will be important as we see later.

The BeautifulSoup package was used to scrap the data from the website. Unnecessary data was removed from the dataframe and the column names were translated from German to English.

Getting the coordinates of the districts

Simply passing the name of the district from Wikipedia to the geocoder of the geopy package lead to party wrong coordinates. For a distinct search result, the query name needed to be adapted to e.g. 'Stadtbezirk 01 Altstadt-Lehel'. 'Stadtbezirk' being German for district, '01' being the district number, and 'Alstadt-Lehel' the corresponding district name.

Finally, the following dataframe was obtained:

	No.	district	area	population	Latitude	Longitude
0	1	Altstadt-Lehel	315	21.100	48.143648	11.589579
1	2	Ludwigsvorstadt-Isarvorstadt	440	51.644	48.130722	11.566526
2	3	Maxvorstadt	430	51.402	48.146570	11.571445
3	4	Schwabing-West	436	68.527	48.166354	11.566191
4	5	Au-Haidhausen	422	61.356	48.130274	11.598334
5	6	Sendling	394	40.983	48.115921	11.548387
6	7	Sendling-Westpark	781	59.643	48.118031	11.519333
7	8	Schwanthalerhöhe	207	29.743	48.134589	11.538013
8	9	Neuhausen-Nymphenburg	1291	98.814	48.156626	11.518016
9	10	Moosach	1109	54.223	48.185426	11.515594
10	11	Milbertshofen-Am Hart	1342	75.094	48.197532	11.577300
11	12	Schwabing-Freimann	2567	77.936	48.189278	11.608583
12	13	Bogenhausen	2371	87.950	48.158487	11.636682
13	14	Berg am Laim	631	46.098	48.125613	11.629557
14	15	Trudering-Riem	2245	73.206	48.126678	11.685296
15	16	Ramersdorf-Perlach	1990	116.327	48.100894	11.633371
16	17	Obergiesing-Fasangarten	572	54.256	48.103934	11.590683
17	18	Untergiesing-Harlaching	806	53.184	48.095380	11.569129
18	19	Thalkirchen-Obersendling-Forstenried-Fürstenri...	1776	96.714	48.084213	11.508051
19	20	Hadern	922	49.898	48.113193	11.482850
20	21	Pasing-Obermenzing	1650	74.625	48.152363	11.468434
21	22	Aubing-Lochhausen-Langwied	3406	47.813	48.165059	11.400221
22	23	Allach-Untermenzing	1545	33.355	48.195157	11.462973
23	24	Feldmoching-Hasenbergl	2894	61.774	48.218462	11.520409
24	25	Laim	529	56.546	48.136310	11.505333

Getting the information on venues

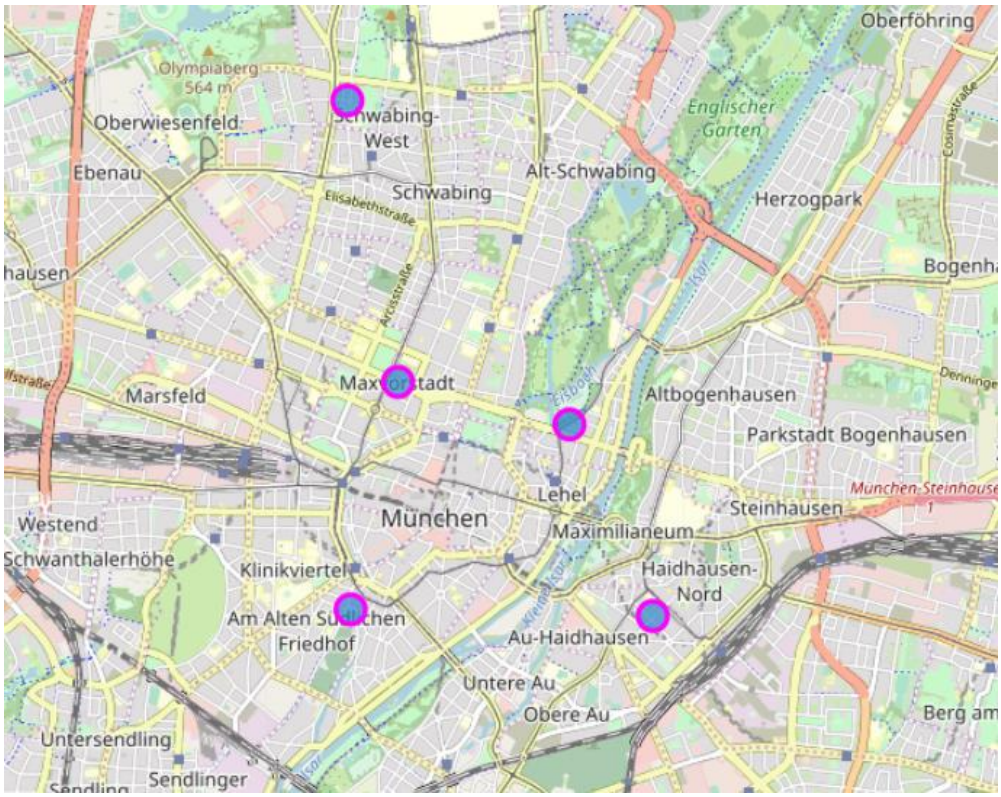
Foursquare data is used to explore the venues in the districts. Therefore, an area within a radius of 750 meters around the respective district coordinate is considered. Details on the procedure are described in the next section on data exploration.

Methodology

Not all the 25 districts of Munich will be considered in the following. The exploratory analysis will concentrate on the major business areas of Munich:

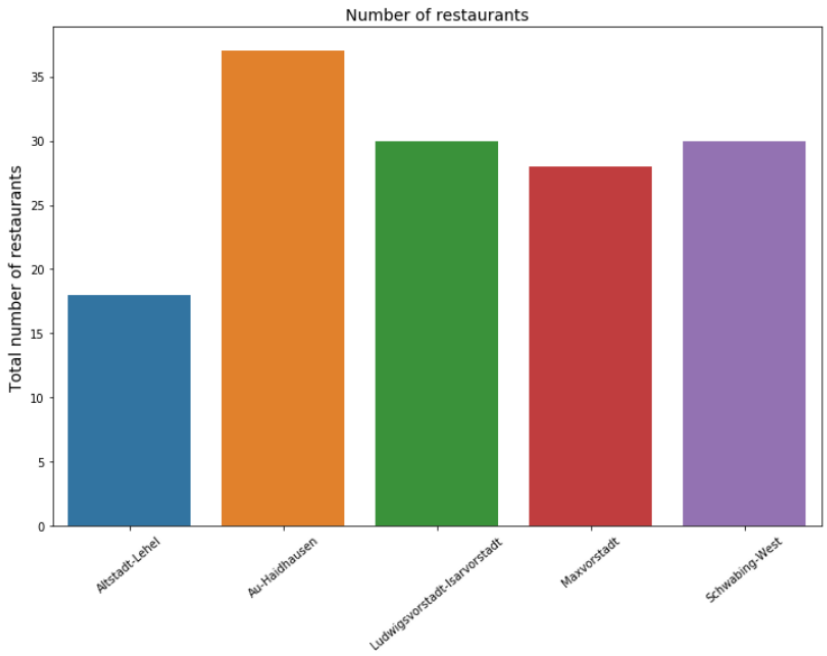
- Altstadt-Lehel
- Schwabing-West
- Au-Haidhausen

- Ludwigsvorstadt-Isarvorstadt
- Maxvorstadt

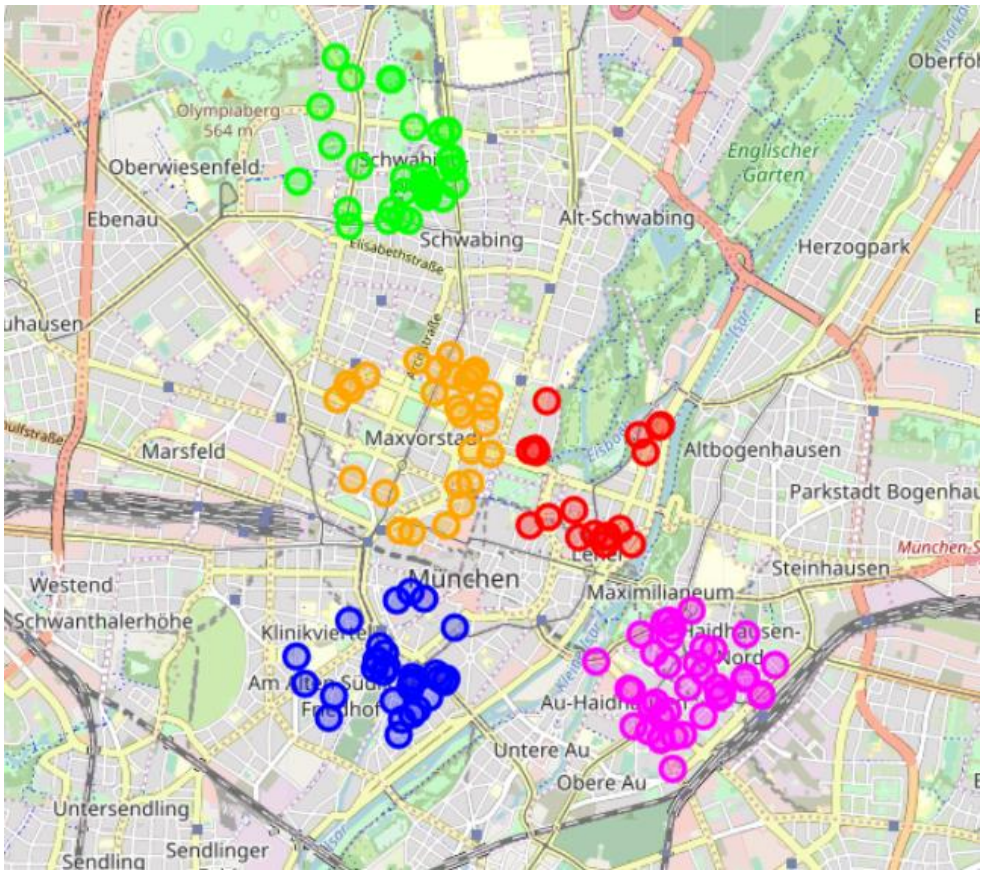


Getting the information on venues

With the five mentioned districts, Foursquare returned 416 venues in total. As only restaurants are of interest for the current investigation, the venues were filtered to categories having 'restaurant' in their name. Finally, 143 restaurants were obtained. The number of restaurants in *Ludwigvorstadt-Isarvorstadt*, *Maxvorstadt*, and *Schwabing-West* are close to the average value. In contrast, *Alstadt-Lehel* shows less than average number of restaurant and in turn *Au-Haidhausen* more than the average.

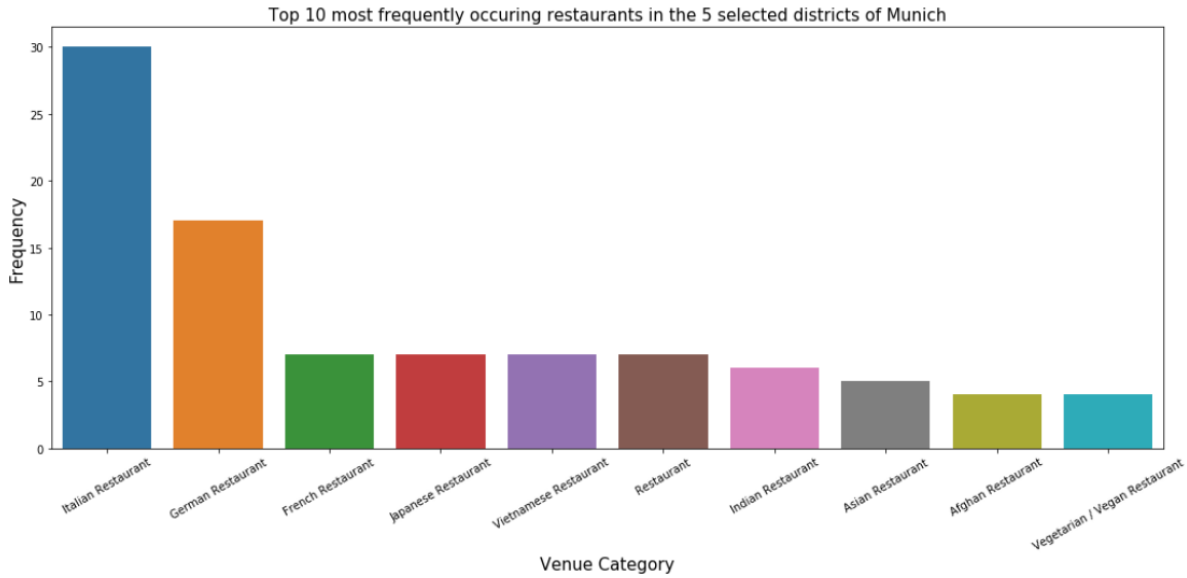


The location of the restaurants was visualized in order to confirm that the search radius has been defined small enough to prevent overlapping of the neighbouring districts:

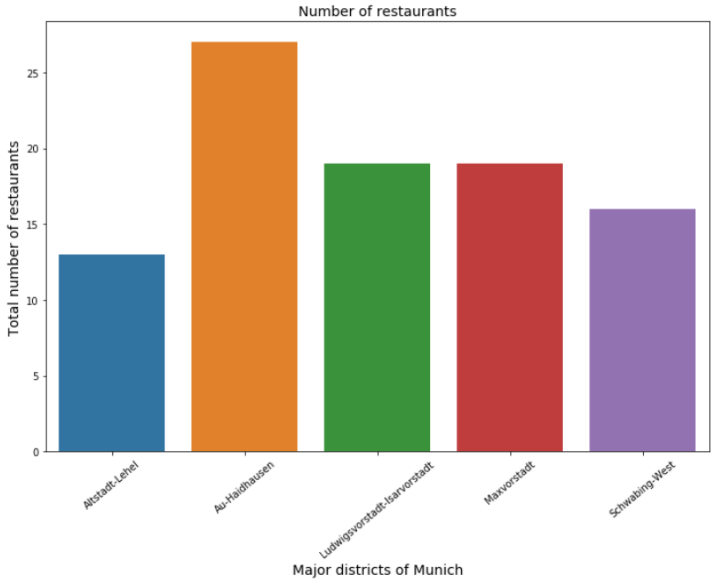


Exploratory analysis

The following investigation is restricted to the ten most frequently occurring restaurant categories:



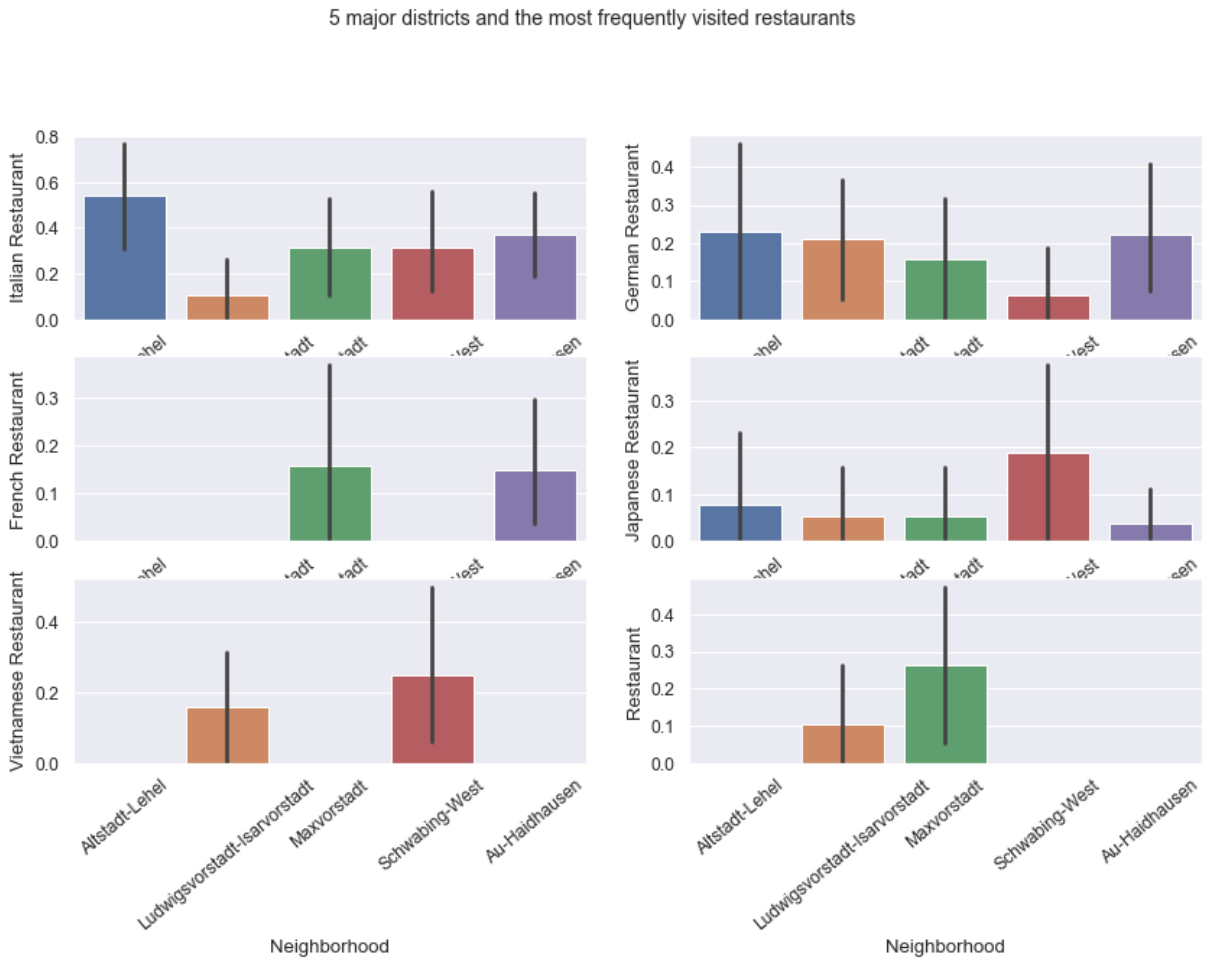
Finally, 94 restaurants remained in the data set. The initial distribution of the restaurants among the districts kept the characteristic as previously discussed:



Ludwigvorstadt-Isarvorstadt is the only district where Italian is not the major category of restaurants. Here, German, Vietnamese, Asian, and generic 'Restaurant' are the most frequent categories ahead of Italian restaurants.

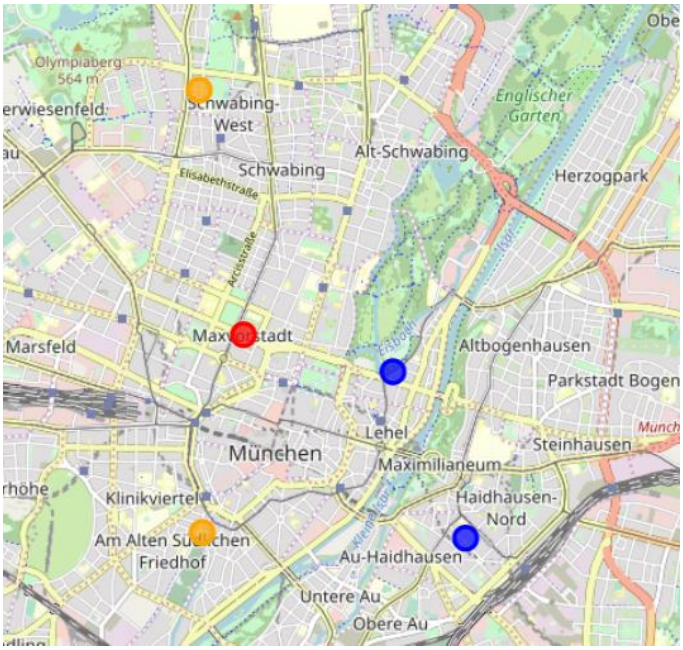
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Altstadt-Lehel	Italian Restaurant	German Restaurant	Japanese Restaurant	Indian Restaurant	Afghan Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Restaurant	French Restaurant	Asian Restaurant
2	Au-Haidhausen	Italian Restaurant	German Restaurant	French Restaurant	Vegetarian / Vegan Restaurant	Indian Restaurant	Japanese Restaurant	Asian Restaurant	Afghan Restaurant	Vietnamese Restaurant	Restaurant
3	Ludwigsvorstadt-Isarvorstadt	German Restaurant	Vietnamese Restaurant	Asian Restaurant	Restaurant	Italian Restaurant	Afghan Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Indian Restaurant	French Restaurant
4	Maxvorstadt	Italian Restaurant	Restaurant	German Restaurant	French Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Afghan Restaurant
5	Schwabing-West	Italian Restaurant	Vietnamese Restaurant	Japanese Restaurant	Indian Restaurant	German Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Restaurant	French Restaurant	Afghan Restaurant

Further, it is interesting to look at how the different categories are distributed among the districts:



Clustering of the districts

K-means clustering was applied to the frequency of the restaurant categories. Three cluster were obtained:



Ludwigsvorstadt-Isarvorstadt and *Schwabing-West* are considered similar by the algorithm, as is *Altstadt-Lehel* and *Au-Haidhausen*. *Maxvorstadt* ends up in a separate cluster.

Results

It turns out that Italian is by far the dominant category of restaurants in the selected districts; almost 30%. It is followed by German restaurant with a little more than 15% and several Asian and other restaurants with each about 5% of all restaurant.

Except for *Ludwigvorstadt-Isarvorstadt*, Italian restaurant dominate in all other selected districts.

Altstadt-Lehel and *Au-Haidhausen* share a cluster due to the similarity that Italian, German and Indian restaurants are in both districts among the five highest frequency categories. *Maxvorstadt* remains in a separate cluster, probably due to the large number of restaurants residing in the generic 'Restaurant' category. *Ludwigsvorstadt-Isarvorstadt* and *Schwabing-West* fall into the third cluster, presumably as those to are the only districts with Vietnamese restaurants within the ten highest frequency categories and among the ones without any French restaurants.

Discussion

The choice of opening an Italian restaurant in Munich may not be a very promising idea after all. Already now, almost one third of the highest frequency restaurants in the selected districts are Italian. The least competition in terms of similar type of restaurants can be expected in *Ludwigvorstadt-Isarvorstadt.*

Conclusion

Albeit the rather large similarities among the selected district, at least one could be identified as the most promising location for a new Italian restaurant: *Ludwigvorstadt-Isarvorstadt.*

In a next step, a more comprehensive data set could be used. The geocoder can return a polygon representing the actual district boundaries. It would be more accurate to uses these polygons as a searching area with Foursquare rather than using the district coordinates and a specified search radius around it.