Opening an Italian restaurant in the canton of Zurich, Switzerland

Stefan Gedamke

IBM Data Science capstone project

1. Introduction

A good friend would like to open an Italian restaurant in the canton of Zurich, Switzerland (a canton is like a state). As there is already a large number of restaurants in the area (some of them Italian), he asked me to help him to find out what the density of restaurants is**.**

He does not really care where in Zurich he opens the restaurant, but it should be a borough (based on ZIP code) that is less crowded with **a)** restaurants in general and **b)** specifically Italian restaurants. Additionally, he would like to not only serve people living in the area**,** but also people working there (walk-in customers for their lunch breaks and dinner meetings). Therefore, we also need to analyse the density of businesses**.**

So ideally, we would find a borough with a low restaurant population (and ideally very few Italian ones) and with many people working in the area.

If possible, he would also like to have customer ratings from the restaurants that might be competitors. This way he can see if it might be easier to break into the area and build a successful business.

2. Data acquisition and cleaning

Based on our problem description, I will need data for the following:

- overview of the existing restaurants (incl. types of restaurants) in the canton of Zurich, based on the ZIP codes of the cities. All cities in the same borough have the same ZIP code, therefore, I will group the data per borough
- population per ZIP code
- number of businesses per ZIP code

- optional: ratings for the focus areas that we will determine with the above data

There is an open data portal available for Switzerland (https://opendata.swiss/en/). This collects data from all official government sources and makes them available for download in various formats. I will get parts of my data from there.

2.1. Data sources

Specifically, this is the data (incl. source) that I will use:

- ZIP codes for Swiss boroughs (filtered for Zurich), this list already contains the geocoordinated of the ZIP codes (from Opendata, https://data.geo.admin.ch/ch.swisstopo-vd.ortschaftenverzeichnis_plz/PLZO_CSV_WGS84.zip)
- list of cities with population and number of business. This does not contain ZIP codes but so-called "BFS codes" (which are tax codes for the respective city). This data needs to be combined with the other list with the ZIP codes (which also includes the BFS code; this will be used for mapping), https://www.bfs.admin.ch/bfs/de/home/statistiken/regionalstatistik/regionale-portraets-kennzahlen/gemeinden.assetdetail.11587763.html
- overview of restaurants (per type) for the ZIP codes, based on a data extraction from Foursquare (via API)
- optional: restaurant ratings - as the ratings are only included in a Foursquare premium account, this will be attempted with the Google API with Google ratings, based on the restaurant names (for selected focus boroughs only)

I will then use the data to first create a Pandas dataframe of the combined data (Zurich ZIP codes with geocoordinated, population per ZIP code and number of businesses). Then, I will extract the restaurants from Foursquare and calculate the number of restaurants and Italian restaurants per ZIP code. I will then generate a map to show population density vs number of restaurants and business desity vs number of restaurants to determine the focus areas. For these, I will attempt to get the rating for the Italian restaurants from Google **to finally determine where to open the restaurant**.

2.2. Data cleaning and preparation

As all the column names were in German, I first renamed everything to English. I then created a new dataframe that only contained citied from the canton of Zurich (the list contained citied form Switzerland), grouped them into boroughs and dropped the city names. As each of the cities had a ZIP code and a precise location already included

(latitude and longitude), this information was partially lost upon grouping. I decided to always take the first ZIP code in the list and the mean of the geospacial values. The letter turned out to be a good-enough representation of the location of the borough (as can be seen in figure 1), except for the borough of Winterthur. This was not centered and I manually corrected that error.

Afterwards, I merged the two lists (one containing the cities, geospacial data and the other one the population, number of businesses and other economic data) based on the so-called “BFS-number”, which is a unique tax identifier for each borough.

Furthermore, I had to delete some of the boroughs, as they were missing information about the number of workplaces (this is an important criterium for the analysis). Finally, I had to remove some special signs in the numbers and convert some of the data from objects to Int.

I was now left with 142 boroughs and 16 columns, down from 4133 cities.

As one of the target customers are people at work for their lunch break, we determine that we need at least 500 workplaces in the borough in order to make the restaurant work. Therefore, I reduced the data to these down to these boroughs, first. The result was 36 boroughs in the canton of Zurich, a number I could work with.

3. Methodology

3.1. Exploratory data analysis

As my focus was now on the 36 boroughs in Zurich, I modified the Foursquare API in such a way that it would only retrieve the information for restaurants categorized as “Italian”. This way, I was able to first only focus on the important information and the direct competition.

After plotting the scope of my data retrieval with the help of the folium package on a map (similar to figure 1), it turned out that the radius of 2.5km was not representative of the city of Zurich, especially when combining the number of restaurants with the population that I retrieved from the OpenData homepage. Therefore, I deleted the city of Zurich from my dataset and plotted it again. The result can be seen in figure 1: all 35 remaining boroughs in the canton of Zurich and the corresponding radius of 2.5km. The blue-shaded circles are the areas I retrieved the information from Foursquare for. It is a good representation of the area and population for all boroughs in scope.

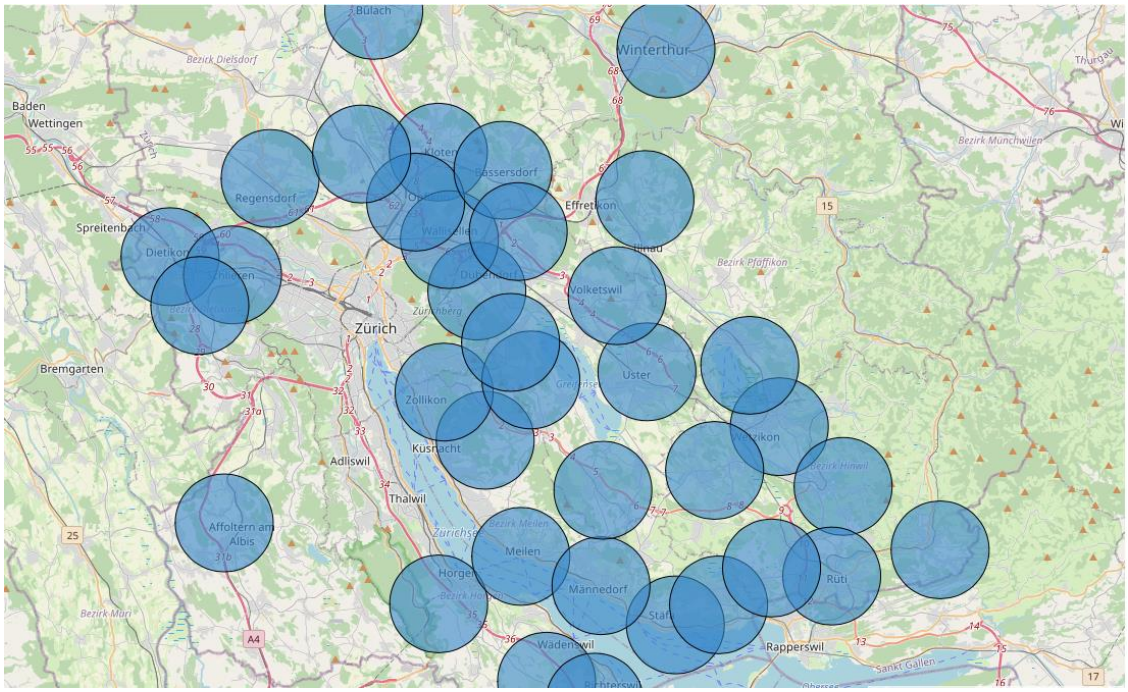

Figure 1: overview of the 35 boroughs in scope for my analysis, incl. radius of 2.5km

The data retrieval from Foursquare revealed the number of Italian restaurants for each of the circles in figure 1. The corresponding results can be seen in figure 2: some boroughs have a relatively high number of Italian restaurants, others only have a few. **But what exactly does this mean**? In order to really be able to tell, I have to compare the number of restaurants with the overall population and number of business in the area. Only then will I be able to say what the density of Italian restaurants is.

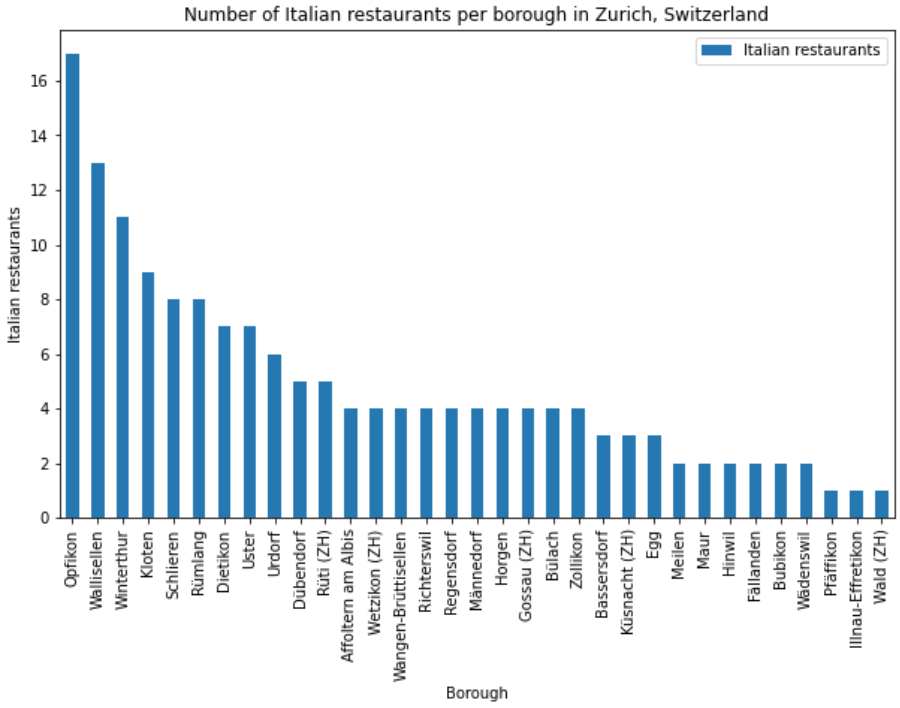

Figure 2: number of Italian restaurants per borough

The result can be seen in table 1: I calculated the amount of population per Italian restaurant and the same for the number of businesses in the respective borough and added the data to the Pandas dataframe. The dataframe was sorted from largest to smallest for population per Italian restaurant and I took the top 5 boroughs including all Italian restaurants in these boroughs for further processing.

	Borough	Population 2018	Employed persons	Number of workplaces	Italian restaurants	Italian Restaurants per X people	Italian Restaurants per X workplaces
17	Illnau-Effretikon	17068	6822	1052	1.0	17068.000000	1052.000000
2	Wädenswil	24341	9624	1867	2.0	12170.500000	933.500000
22	Pfäffikon	11935	5674	811	1.0	11935.000000	811.000000
0	Winterthur	111851	71832	7927	11.0	10168.272727	720.636364
26	Wald (ZH)	9949	3521	676	1.0	9949.000000	676.000000

Table 1: Italian restaurants per population and per workplaces – Top 5 boroughs

The overview of these restaurants is depicted in table 2. From a location perspective, all of the venues are distributed throughout the canton. I visualized this with a folium map in figure 3.

	Borough	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Winterthur	47.499503	8.737565	Molino	47.499554	8.727712	Italian Restaurant
1	Winterthur	47.499503	8.737565	Restaurant Al Giardino	47.494340	8.739810	Italian Restaurant
2	Winterthur	47.499503	8.737565	Cantinetta Bindella	47.499667	8.728742	Italian Restaurant
3	Winterthur	47.499503	8.737565	Santa Lucia	47.501045	8.725249	Italian Restaurant
4	Winterthur	47.499503	8.737565	La Pergola	47.500243	8.729235	Italian Restaurant
5	Winterthur	47.499503	8.737565	Ristorante La Torre	47.497981	8.723812	Italian Restaurant
6	Winterthur	47.499503	8.737565	San Remo	47.507193	8.725907	Italian Restaurant
7	Winterthur	47.499503	8.737565	Restaurant Gutschick	47.491842	8.750220	Italian Restaurant
8	Winterthur	47.499503	8.737565	Gerry's Bar	47.504525	8.763603	Italian Restaurant
9	Winterthur	47.499503	8.737565	La Vita	47.494019	8.710929	Italian Restaurant
10	Winterthur	47.499503	8.737565	Euro Pizza	47.515803	8.759626	Italian Restaurant
18	Wädenswil	47.211005	8.657193	Pizzeria Romantica	47.231575	8.668086	Italian Restaurant
19	Wädenswil	47.211005	8.657193	Gustav	47.230212	8.673345	Italian Restaurant
105	Illnau-Effretikon	47.430953	8.723401	Restaurant Rustica Effretikon ZH	47.427107	8.691886	Italian Restaurant
117	Pfäffikon	47.355366	8.794004	A Casa Di Luca	47.366308	8.781671	Italian Restaurant
134	Wald (ZH)	47.271079	8.922132	Salt & Pepper Restaurant	47.274221	8.913056	Italian Restaurant

Table 2: Italian restaurants in Top 5 boroughs

3.2. Clustering

As a comparison and additional information, I decided to get all the venues for the Top 5 boroughs from Foursquare. This way, I can see if there is any competition other than Italian restaurants in the area. I then clustered the venues with K-means clustering into three clusters (generating up to five clusters always resulted in three) and named them “Restaurants (cluster 0), “Food and drinks” (cluster 1) and “Groceries” (cluster 2). As all but three boroughs were in the “Food and drinks” cluster, I neglected the other ones.

The overview of the clusters can be seen in figure 4. Furthermore, this meant that the clustering did not help me to come to a final decision. Therefore, I created frequency tables of the top 5 venues for the top 5 boroughs in order to see, if anything out of the order shows up. What I found out and can be seen in figure 5 was that the borough of “Wädenswil” did not have any restaurants in the top 5 regarding frequency (only fast food). Therefore, I decided that this would be the final candidate for the location of the restaurant.

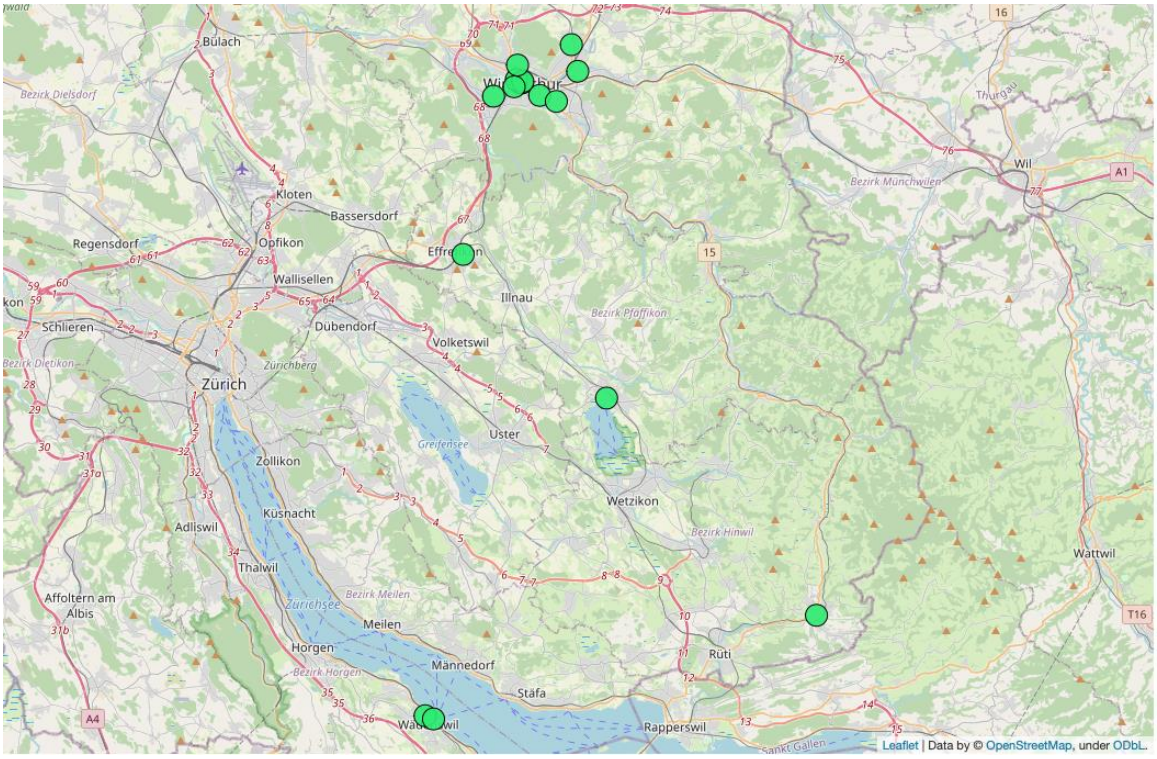

Figure 3: Geographic distribution of the restaurants

Despite the fact that the free Foursquare version is limited to receiving one rating per day only, I attempted to retrieve one rating from Foursquare. Unfortunately, this did not result in any data, as the venues in scope did not have any rating, yet. I double-checked with the Foursquare mobile app and it confirmed the result. I then checked Google maps manually and retrieved the ratings of 4.7 and 4.3 (out of 5) for the two Italian restaurants in Wädenswil ("Gustav": 4.7 and "Pizzaria Romantica": 4.3).

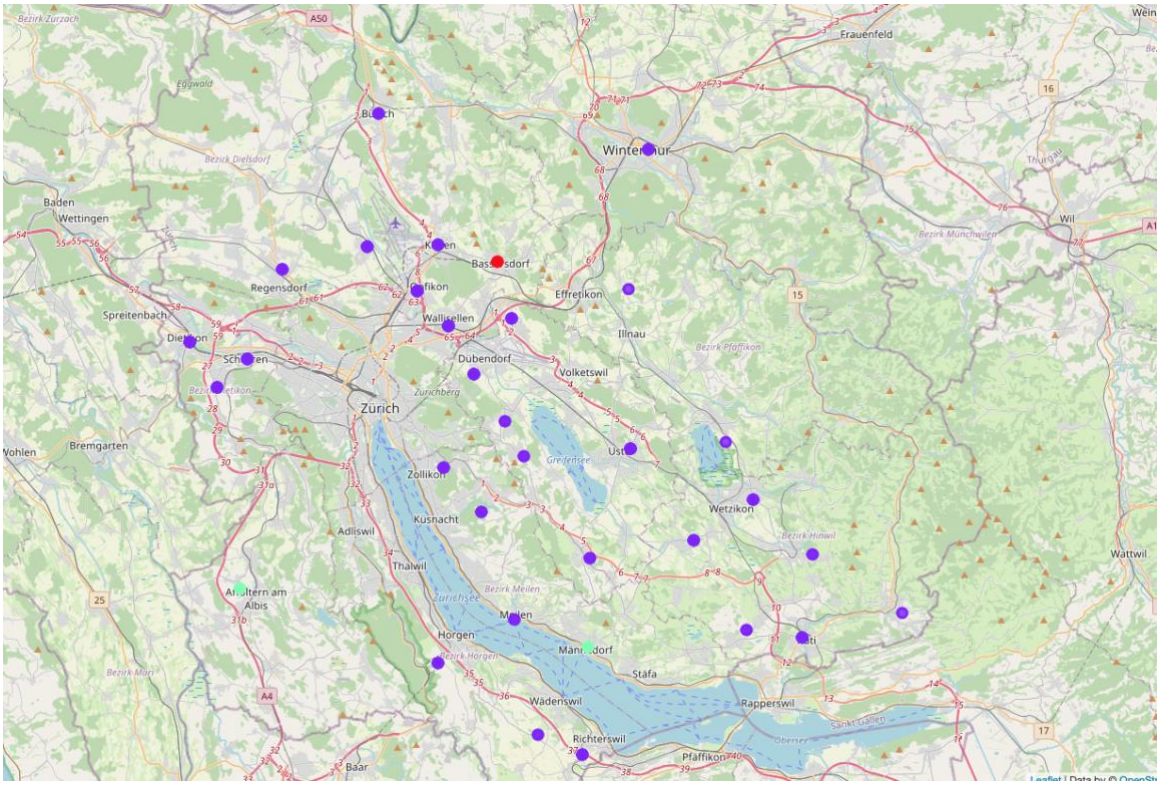

Figure 4: Three different clusters in Zurich

```
----Wädenswil----
                 venue  freq
0          Supermarket  0.18
1                Hotel  0.18
2           Food Court  0.09
3  Fast Food Restaurant  0.09
4         Train Station  0.09

----Illnau-Effretikon----
              venue  freq
0        Restaurant  0.33
1       Supermarket  0.17
2  Swiss Restaurant  0.17
3             Café  0.17
4            Lounge  0.17

----Pfäffikon----
                 venue  freq
0           Restaurant  0.13
1      Swiss Restaurant  0.13
2         Grocery Store  0.07
3         Train Station  0.07
4  Fast Food Restaurant  0.07

----Winterthur----
                 venue  freq
0                Café  0.09
1                 Bar  0.08
2          Restaurant  0.07
3  Italian Restaurant  0.05
4         Supermarket  0.04

----Wald (ZH)----
                          venue  freq
0                    Restaurant  0.29
1                        Bakery  0.14
2                   Gas Station  0.14
3                   Supermarket  0.14
4  Construction & Landscaping  0.14
```

Figure 5: Frequency tables of the Top 5 venues for the Top 5 boroughs

4. Results and Discussion

My analysis showed that there are 4133 cities in the canton of Zurich in Switzerland. Grouping these into boroughs makes a lot of sense in order to make them "economically viable" when analyzing for a valid location for an Italian restaurant. Additionally, I merged the location data with the economical data of the boroughs. The latter was not available on a city level. Once grouped, I was left with 162 borough in the canton of Zurich. Thus number was further reduced to 143, as some missed information about the population and number of businesses.

As the Italian resturant planned to be opened should also serve customers from local business (walk-in customers), I decided to only look at the boroughs that had at least 500 businesses in the area. This cut out an additional 107 boroughs, which left me with 36 in scope. I plotted these on a map to check if the location data makes sense (as I took the average location from all cities in the respective borough). I had to correct the location of the borough of "Winterthur" only, all others were ok. The plot includes a radius of 2.5km, which was a reasonable area to not have too much overlap between the remaining boroughs and large enough to make sure the relevant venues are include to provide good enough results of the search for restaurant. As the city of Zurich was too large for this and exceeded the Foursquare API data limit of 100 per category anyways, I also cut it out and ended up with 35 boroughs. Of course, one could play around with the radius, based on assumptions how mobile people are (walking, by food or with public transportation).

The same radius of 2.5k was then also used to retreive the data from the Foursquare API. I deliberately only focused on the category of "Italian restaurants" this time, including name and location of the restaurants. I then grouped the number of restaurants per borough. Afterwards, I calculated the number of restaurants per population and per number of businesses per borough and generated a Top 5 list, based on "Restaurants per population". The Top 5 borough were:

- Illnau-Effretikon
- Pfäffikon
- Wädenswil
- Wald (ZH)
- Winterthur

As an additional information for my decision I then took the 35 relevant borough mentioned earlier and retreived the venue information from Foursquare for all of them. I then generated clusters, which left me with three clusters overall, which were named “Restaurants (cluster 0), “Food and drinks” (cluster 1) and “Groceries” (cluster 2). All but three clusters were in cluster 1, including all Top 5 boroughs from above. As this was not helpful, I generated the frequency tables with the top 5 clusters per borough and checked the frequency of resturants or similar for each of them. It turned out that only "Wädenswil" had no real restaurants (only fast food) in the top 5. Therefore, I decided that this would be a good focus borough for an Italian restaurant.

Finally, I attenpted to receive a rating from Foursquare, but it turned out that both Italian restaurants in the borough do not have one. I then consulted Google to get at least a rating

from there and both already have a good rating ("Gustav": 4.7 and "Pizzaria Romantica": 4.3).

Wädenswil seems to be a good choice, as it has the second-highest number (of the 35 focus boroughs) for "Populations per Italian restaurant" and "Number of businesses per Italian restaurant" in the entire canton. Additionally to this, the social welfare quota of 2.3% is quite low (second-lowest of the Top 5 boroughs) and it has a nice location at the lake. Thus, there will possibily also be tourists visiting.

Of course, looking only at boroughs with at least 500 businesses as a first filter does not mean the other areas do not have any potential. If the focus would be low competition only in a promising borough then it would also make sense to look at the density of restaurants overall. Zurich is a high-income canton in Switzerland and most people can affort to dine out on a regaular basis. Also, excluding Zurich city from the analysis could raise some doubts. While the city was out of scope for this analysis, one could divide the city and look for low-density restaurant area there. The advantage of the city is that it has a dense public transportation network and people can move around easily. Another area I didn't touch is the potential rent of a restaurant: depending on the area, this can be very different. While the rent Zurich city is certainly very high, other boroughs are much cheaper. This of course has to be weighted against the tax rates, which change with the borough. This might be an area to look at for future analysis.

5. Conclusion

The purpose of this analysis was to find out where it would be feasible to open an Italian restaurant in the canton of Zurich, Switzerland. I first narrowed down the search to boroughs with a reasonable population and number of businesses and then gathered the Foursquare data for Italian restaurants. After generating the restaurant density per population and businesses, I used the Foursquare API to get additional data on the venues for the Top 5 boroughs. These were clustered and the venues with the highest frequency analysed. Based on the comparison to the density of Italian restaurants, the borough of Wädenswil was picked as a target area that should be further evaluated.

Overall, opening an Italian restaurant in Wädenswil has high potential with a relatively high population and customer potential, low competition and a nice location. Of course, one should get familiar further with the borough and visit it, to get more impressions. Talking to people living on site would also help to get more insights before making a final decision for investment.