The Battle of the Neighbourhoods

Ahmed B

1. Introduction: Business Problem

Zurich, Switzerland is home to a large number of financial institutions and banking companies, and the low tax rates attract overseas companies to set up their headquarters there. It is also a great tourist destination and is home to a few UNESCO world heritage sites. In this project we will determine the optimal location to open a restaurant in the Zurich district, and will be targeted at stakeholders interested in opening a restaurant in Zurich. Since we are not aware of what type of restaurant is popular in Zurich, we will also determine the most popular type(s) of restaurants in Zurich. We will utilize data science and machine learning techniques in order to do this. This project will be of interest to any one wanting to open a restaurant in Zurich, and as a secondary application can also be useful to anyone wanting to visit or live in Zurich.

2. Data

In this project we will utilize three datasets in order to solve our problem. The first dataset is a list of municipalities and information regarding them such as which district it is in, the date of its merger and its coat of arms. The latter two columns were dropped as they are not needed for this project. This data set was scraped from Wikipedia. The next data set is one contain the latitude and longitude for each municipality. This dataset was obtaining from taking the coordinates from each of the Wikipedia pages of the municipalities and pasting them into a CSV file using excel. Then we merge the two datasets to create a data frame with all the desired information. The file containing the list of venues was obtained using the Foursquare API. We selected all venues not just restaurants in order to see where restaurants were more popular.

Below are the first 5 rows of the cleaned datasets:

	Municipalities	District	Latitude	Longitude
0	Affoltern	11	47.418611	8.506111
1	Albisrieden	9	47.374350	8.483119
2	Altstetten	9	47.384969	8.477175
3	Aussersihl	4 and 5	47.375000	8.527000
4	Enge	2	47.363611	8.531111

Figure 1

	Municipalities	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Affoltern	47.418611	8.506111	Migros	47.420404	8.508402	Supermarket
1	Affoltern	47.418611	8.506111	Bahnhof Zürich Affoltern	47.420950	8.507619	Train Station
2	Affoltern	47.418611	8.506111	Coop	47.419237	8.505753	Supermarket
3	Affoltern	47.418611	8.506111	Einkaufszentrum Affoltern	47.419593	8.507790	Department Store
4	Affoltern	47.418611	8.506111	Pizzeria Piazza	47.419276	8.505978	Italian Restaurant

Figure 2 Foursquare API data

Some of the features extracted from the data was the district of the municipality, the latitude and longitude of each municipality, and from the foursquare API we extracted the venues, the venue category, and its geo-coordinates. All of this data can be used to help us in our analysis and problem solving, which will be discussed more thoroughly in the methodology and analysis sections.

3. Methodology

After cleaning the data, we began the analysis of the data by utilizing one hot encoding on the venues data, this makes it easier to work with categorical variables, in this case the venue categories, which are needed in order to solve this problem and do the analysis. After this we will do a frequency analysis using the mean of the frequency of occurrence of each category. This will aid us in when we utilize clustering, as well as in identifying the ideal locations for a restaurant. After doing this we sort the data into a data frame displaying the top 10 venues, in order to make it easier to work with.

The next step is to begin the clustering process, for this project we have chosen to use k-means clustering. We have chosen this over other methods for several reasons. First of all, it is easier to implement, and is relatively efficient for our data set size. Second of all we are not aware of any arbitrary clusters in our dataset, so we rule out using DBSCAN.

The first step in k-means clustering is to choose the optimal or best value for 'k', i.e. the number of clusters. For this project we have chosen to do so using the elbow method. Using this method, we plot the values for 'k' on the X axis and the distortion on the Y axis (the values calculated with the cost function). When K increases, the centroids are closer to the cluster's centroids. The improvements will decline, at some point rapidly, creating the elbow shape. The point located on the 'elbow' is our optimal 'k'[1].

After we choose the optimal 'k' we will create the model and fit it to the data and create a new data frame that includes the cluster and the top 10 venues. Using this we create a map to display the clusters and display the individual clusters.

[1] https://pythonprogramminglanguage.com/kmeans-elbow-method/

4. Results and Discussion

Municipalities	Antique Shop	Argentinian Restaurant	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bar	Bath House	Beach	Beer Garden	Bike Trail	Bistro
Affoltern	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Affoltern	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Affoltern	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Affoltern	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Affoltern	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 3 Example of encoded data

```
----Affoltern----
                venue  freq
0         Bus Station  0.21
1         Supermarket  0.14
2  Italian Restaurant  0.07
3               Diner  0.07
4    Department Store  0.07


----Albisrieden----
               venue  freq
0        Bus Station  0.29
1      Grocery Store  0.14
2        Supermarket  0.14
3  Trattoria/Osteria  0.14
4     Scenic Lookout  0.14


----Altstetten----
          venue  freq
0          Pool   0.5
1  Soccer Field   0.5
2  Antique Shop   0.0
3        Museum   0.0
4     Pool Hall   0.0
```

Figure 4 frequency analysis

```
----Aussersihl----
                venue  freq
0                 Bar  0.14
1  Italian Restaurant  0.09
2          Restaurant  0.07
3    Swiss Restaurant  0.06
4               Café  0.04


----Enge----
                venue  freq
0  Italian Restaurant  0.08
1              Hotel  0.08
2        Tram Station  0.06
3                 Bar  0.06
4              Plaza  0.04


----Fluntern----
           venue  freq
0   Tram Station  0.21
1  Grocery Store  0.14
2          Plaza  0.14
3      Gastropub  0.07
4      Cafeteria  0.07
```

Figure 5 Frequency Analysis con't

```
----Hirslanden----
                venue  freq
0  Italian Restaurant  0.12
1                Park  0.12
2         Bus Station  0.12
3               Hotel  0.12
4    Swiss Restaurant  0.12


----Höngg----
           venue  freq
0   Tram Station  0.14
1    Pizza Place  0.10
2          Plaza  0.10
3  Grocery Store  0.10
4    Gas Station  0.05


----Oberstrass----
                venue  freq
0  Italian Restaurant  0.20
1         Supermarket  0.13
2              Bakery  0.13
3        Tram Station  0.07
4                Park  0.07


----Oerlikon----
                  venue  freq
0  Gym / Fitness Center  0.11
1               Bakery  0.11
2            Laser Tag  0.05
3       Thai Restaurant  0.05
4                   Gym  0.05


----Riesbach----
                venue  freq
0    Swiss Restaurant  0.14
1          Restaurant  0.11
2  Italian Restaurant  0.07
3              Bakery  0.07
4        Tram Station  0.07
```

Figure 6 frequency analysis con't.

```
----Schwamendingen----
                venue  freq
0          Bus Station  0.15
1  Arts & Crafts Store  0.15
2           Restaurant  0.08
3     Asian Restaurant  0.08
4                Café  0.08


----Seebach----
                venue  freq
0  Italian Restaurant  0.24
1              Bakery  0.12
2         Supermarket  0.12
3          Restaurant  0.12
4           Gastropub  0.06


----Unterstrass----
                venue  freq
0  Italian Restaurant  0.08
1                Café  0.08
2              Bakery  0.08
3         Bus Station  0.08
4               Hotel  0.08


----Wiedikon----
                venue  freq
0  Italian Restaurant  0.11
1          Restaurant  0.09
2         Pizza Place  0.09
3        Tram Station  0.06
4         Supermarket  0.06


----Wipkingen----
                venue  freq
0                 Bar  0.09
1  Italian Restaurant  0.06
2              Bakery  0.06
3    Swiss Restaurant  0.06
4         Supermarket  0.06
```

```
----Witikon----
                  venue  freq
0    Athletics & Sports  0.33
1      Swiss Restaurant  0.33
2  Other Great Outdoors  0.33
3          Antique Shop  0.00
4                Museum  0.00


----Wollishofen----
                      venue  freq
0               Supermarket  0.25
1               Bus Station  0.08
2          Community Center  0.08
3  Mediterranean Restaurant  0.08
4                Restaurant  0.08
```

Figure 7 Frequency Analysis cont.

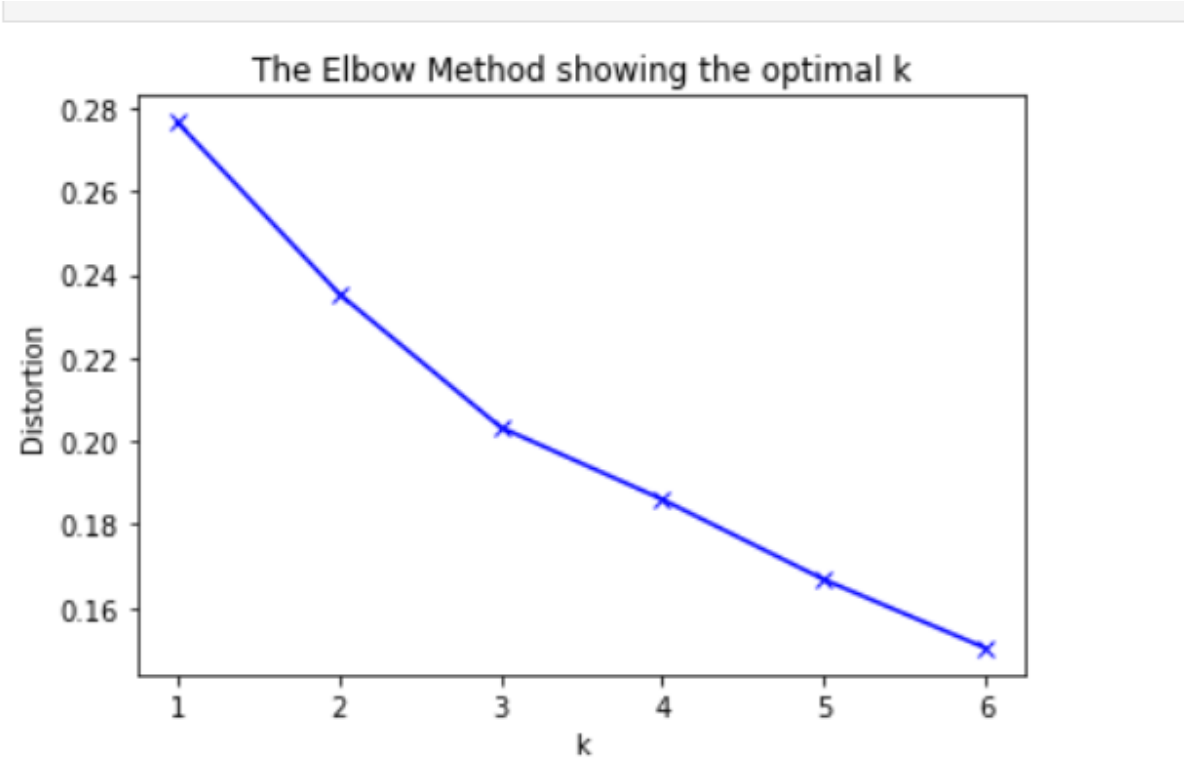

Figure 8 elbow method results

Based on the above, the optimal value for ‘k’ was chosen as 3.

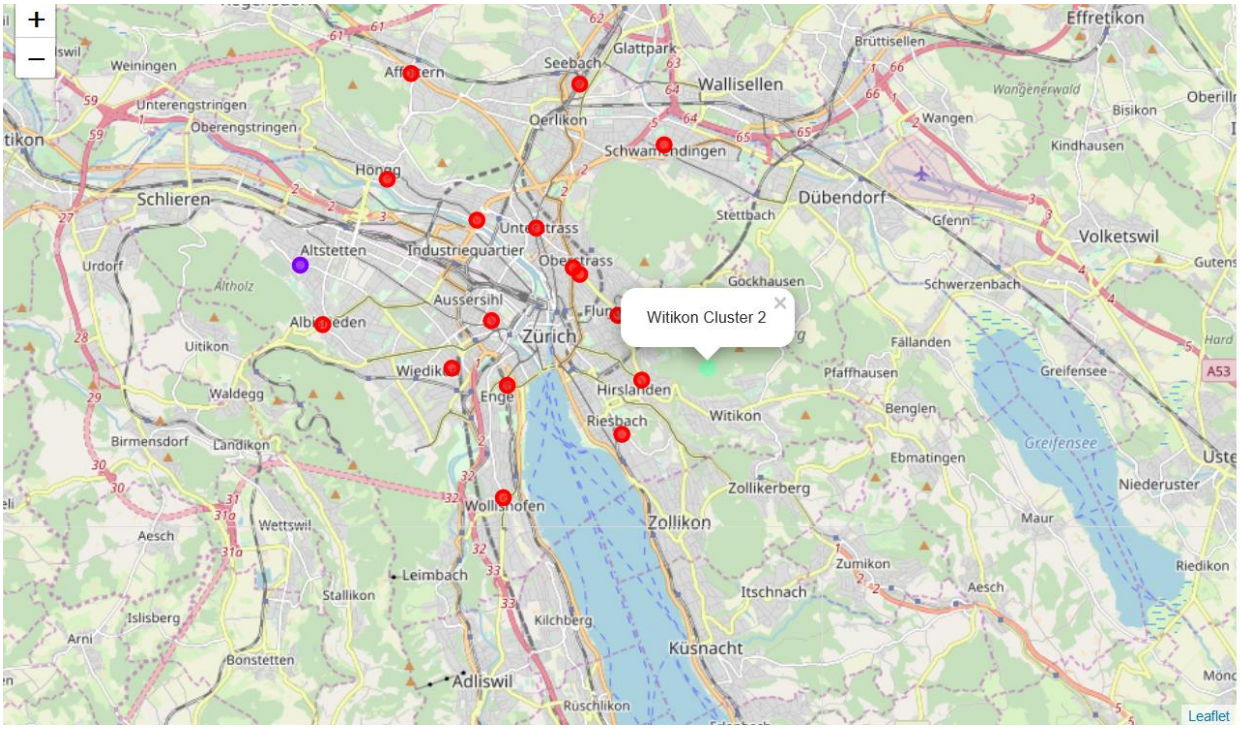

Figure 9 Map of Clusters

	Municipalities	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	Affoltern	Bus Station	Supermarket	Department Store	Train Station	Bike Trail	Music Venue	Miscellaneous Shop	Athletics & Sports	Hotel	Italian Restaurant
1	Albisrieden	Bus Station	Grocery Store	Swiss Restaurant	Trattoria/Osteria	Scenic Lookout	Supermarket	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Yoga Studio
3	Aussersihl	Bar	Italian Restaurant	Restaurant	Swiss Restaurant	Café	Thai Restaurant	Chinese Restaurant	Coffee Shop	Nightclub	Pub
4	Enge	Hotel	Italian Restaurant	Tram Station	Bar	Restaurant	French Restaurant	Plaza	Park	Supermarket	Yoga Studio
5	Fluntern	Tram Station	Plaza	Grocery Store	Bakery	Supermarket	Pizza Place	Bus Station	Gym Pool	Gastropub	Cafeteria
6	Hirslanden	Hotel	Park	Bus Station	Italian Restaurant	Swiss Restaurant	Bakery	Modern European Restaurant	Plaza	Grocery Store	Mediterranean Restaurant
7	Höngg	Tram Station	Plaza	Grocery Store	Pizza Place	Mexican Restaurant	Sporting Goods Shop	Gas Station	Café	Supermarket	Bus Station
8	Oberstrass	Italian Restaurant	Bakery	Supermarket	Swiss Restaurant	Cable Car	Diner	Park	Asian Restaurant	Tram Station	Hotel
9	Oerlikon	Bakery	Gym / Fitness Center	Plaza	Hookah Bar	Laser Tag	Korean Restaurant	Salad Place	Supermarket	Falafel Restaurant	Coffee Shop
10	Riesbach	Swiss Restaurant	Restaurant	Museum	Tram Station	Italian Restaurant	Bakery	Bath House	Mexican Restaurant	Café	Food Court
11	Schwamendingen	Bus Station	Arts & Crafts Store	Café	Thai Restaurant	Supermarket	Restaurant	Swiss Restaurant	Tram Station	Asian Restaurant	Plaza
12	Seebach	Italian Restaurant	Restaurant	Bakery	Supermarket	Swiss Restaurant	Cable Car	College Academic Building	Park	Tram Station	Sandwich Place
13	Unterstrass	Hotel	Italian Restaurant	Café	Bakery	Bus Station	Bistro	Park	Coffee Shop	Falafel Restaurant	Food & Drink Shop
14	Wiedikon	Italian Restaurant	Pizza Place	Restaurant	Fast Food Restaurant	Tram Station	Light Rail Station	Supermarket	Antique Shop	Mexican Restaurant	Burrito Place
15	Wipkingen	Bar	Supermarket	Art Museum	Restaurant	Italian Restaurant	Swiss Restaurant	Bakery	Gym / Fitness Center	Multiplex	Shopping Mall
17	Wollishofen	Supermarket	Hostel	Community Center	Restaurant	Cheese Shop	Tram Station	Train Station	Café	Mediterranean Restaurant	Bus Station

Figure 10 Cluster 1

	Municipalities	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
2	Altstetten	Pool	Soccer Field	Yoga Studio	College Academic Building	Cupcake Shop	Department Store	Diner	Discount Store	Eastern European Restaurant	Ethiopian Restaurant

Figure 11 Cluster 2

	Municipalities	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
16	Witikon	Athletics & Sports	Other Great Outdoors	Swiss Restaurant	Yoga Studio	French Restaurant	Department Store	Diner	Discount Store	Eastern European Restaurant	Ethiopian Restaurant

Figure 12 Cluster 2

Based on the above results, the most popular restaurant in Zurich is the Italian restaurant, being the most common venues restaurant for 3 districts and showing in the top 10 for most of the other districts in cluster 1. It should also be noted that restaurants in general are in the top 10 for all districts in all three clusters. Looking at the frequency analysis we see Italian restaurants appearing in the top 5 for ten districts out of 17. Based on these results the recommended type of restaurant to open is an Italian restaurant, based one where they are most common, we feel they would have most success in any district in cluster 1. If stakeholders wish to avoid competition, they can choose a district in cluster 1 where Italian restaurants do not show up in the top 10, such as Albisrieden or Höngg, as it can be assumed, they will have success there due the similarities they share with other clusters. We do not recommend opening an Italian restaurant in cluster 2 and 3, both containing a single district, as although restaurants do appear in the top 10, they are both in 9th and 10th place, and are not Italian restaurants. The decision however is ultimately left to the stakeholder's discretion.

5. Conclusion

The purpose of this project was to find the best type of restaurant to open in Zurich, Switzerland and a suitable location for it. Based on the analysis we conducted we recommend opening an Italian restaurant in any one of the districts located in cluster 1 (see figure 10), based on how common it is in this area. We leave the decision of which district specifically up to stakeholders. It should be noted however, that this report should not be the sole criterion for making this decision. Analyses such as these are only a tool to aid the decision-making process and are only as good as the data and methods (which are never perfect) used in them. We recommend using other methods, tools and forms of decision-making criteria before reaching a final decision.