

# Choosing The Optimal Location To Open A Restaurant

Ahmed Hourani

August 2, 2020

## 1. Introduction

Opening a restaurant requires a big investment and is often very risky. Before opening a restaurant, one should consider the most important factors that will lead to the success of one's restaurant. These factors include quality of service, quality and taste of the food, and most importantly, a good location. Choosing the restaurant's location can make or break your business.

Let's say that we are planning to open a restaurant in Toronto that sells Arabic food. We've perfected all the recipes and we have the branding, logo and restaurant name all planned out. The only thing left, which is the most important, is we have to find the optimal Neighborhood to open the restaurant.

## 2. Data

We will use the following information:

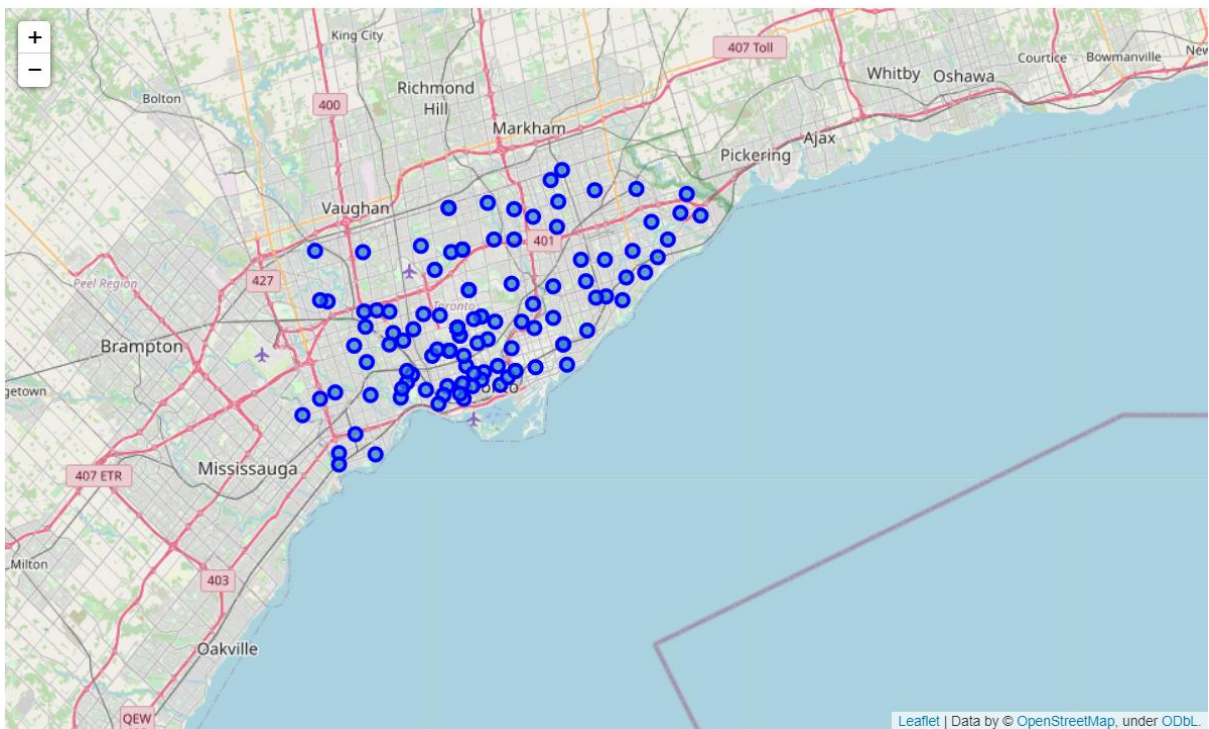
- List of neighborhoods in Toronto (Toronto Census)
- Population in each neighborhood (Toronto Census)
- Number of Arabic speaking people in each neighborhood (Toronto Census)
- Average income in each neighborhood (Toronto Census)
- Number of restaurants in each neighborhood (Foursquare API)
- Number of Arabic restaurants in each neighborhood (Foursquare API)

[Here](#) is a link to the Toronto Census dataset used. As for getting the longitudes and latitudes of each Neighborhood, forward geocoding was done using an API through a resource called [OpenCage](#).

## 3. Methodology:

The first thing we did is extract the list of neighborhoods, population, population of Arabic speakers, and average income from the Toronto Census dataset. The data was already pretty clean and ready to use. The next step would be to use the Foursquare api, however for that, we need to get the coordinates for each neighborhood. OpenCage was the api of choice for the job. Through forward geocoding, we get all the necessary coordinates and we plot them on a map using folium (figure 1).

**Figure 1: location of each neighborhood in Torono, Canada**



Once we are done getting the coordinates, we use the Foursquare search endpoint and enter the category ids for “food” and “middle eastern restaurants” and we search a radius of 700 from the coordinates, which returns a JSON format response. Using the Python len function, we can get the number of all restaurants and Middle Eastern restaurants in each neighborhood. All the data is then joined into one data frame (figure 2).

**Figure 2: Final dataset with all features**

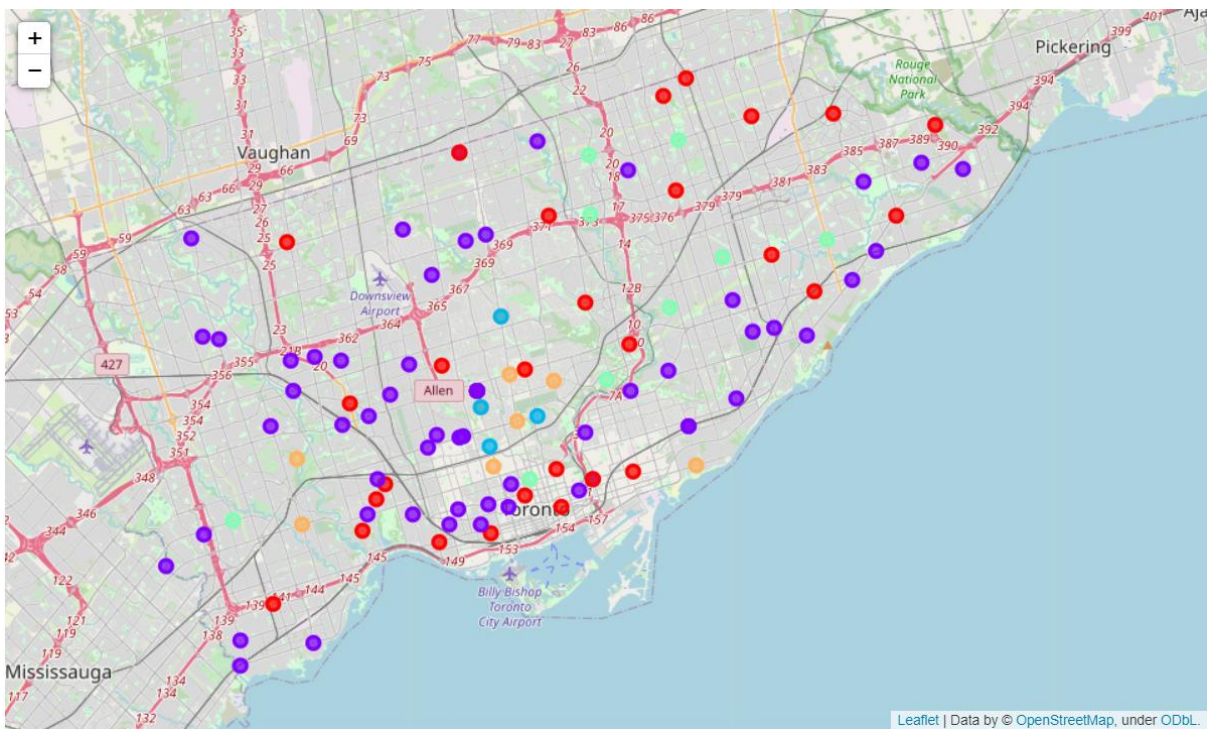
	Neighborhood	Population	Average Income	Arabic Speakers	lat	Ing	Nearby Restaurants	Nearby Middle Eastern
0	Agincourt North	29113	30414	250	43.808038	-79.266439	29.0	1.0
1	Agincourt South-Malvern West	23757	31825	200	43.663280	-79.466170	29.0	1.0
2	Alderwood	12054	47709	35	43.601717	-79.545232	27.0	0.0
3	Annex	30526	112766	225	43.670338	-79.407117	30.0	9.0
4	Banbury-Don Mills	27695	67757	395	43.734804	-79.357243	11.0	0.0
...	...	...	...	...	...	...	...	...
135	Wychwood	14349	54460	50	43.682122	-79.423839	29.0	5.0
136	Yonge-Eglinton	11817	89330	85	43.706748	-79.398327	30.0	9.0
137	Yonge-St.Clair	12528	114174	80	43.688078	-79.394396	28.0	1.0
138	York University Heights	27593	29958	250	43.758781	-79.519434	30.0	1.0
139	Yorkdale-Glen Park	14804	38527	65	43.710434	-79.453340	29.0	0.0

140 rows × 8 columns

Next we create a sub data frame consisting of columns 'Population', 'Arabic Speakers' and 'Average Income' and we normalize the data. Afterwards, we perform a Kmeans clustering with an n\_cluster value of 5, in order to help us see patterns between the 140 neighborhoods. The resulting cluster labels were then inserted into the original non-normalized data frame and folium was used to plot the clusters (figure 3)

### 3. Results

**Figure 3: Map of clustered data. Cluster 0: Red. Cluster 1: Purple. Cluster 2: Blue. Cluster 3: Turquoise. Cluster 4: Orange**



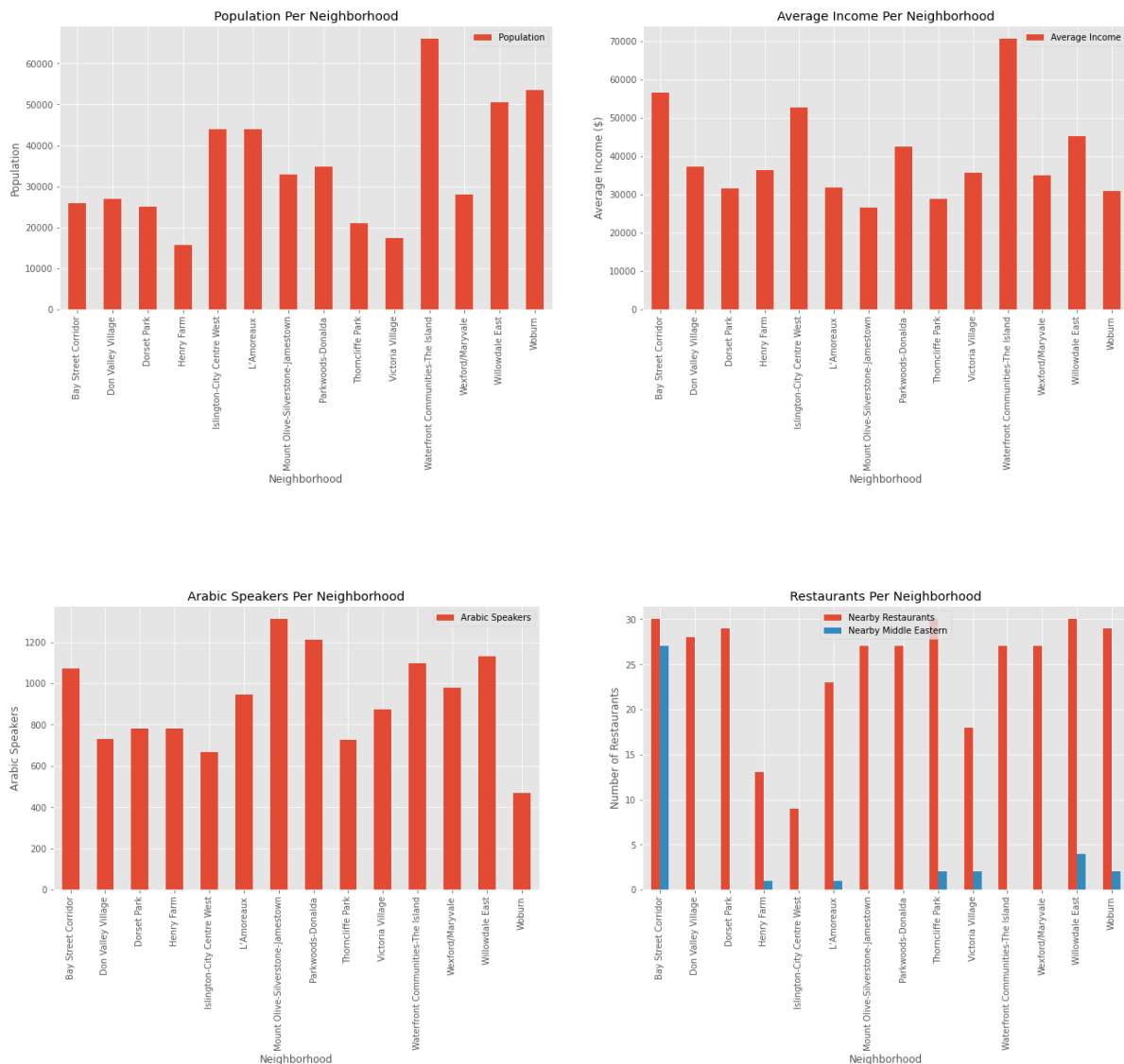
**Table 1: Characteristics of each cluster**

Cluster	Population	Arabic Speaking Population	Average Income (,000\$)
0	High	244	42
1	Low	88	45
2	Low	71	210
3	High	911	40
4	Low	81	106

Judging by the characteristics of each cluster, there is no clear answer as to which cluster is better than the rest. For this reason we need to go back and dig into the business knowledge part of a data science project. How classy or high end is the restaurant going to be? Let's say we are going for a fast casual concept. Based

on this fact, we focus our efforts into cluster 3 since neighborhoods in that cluster seem to have high general populations and high Arabic Speaking Populations.

In the next step, we need to study the different neighborhoods in order to choose the most optimal one. This is done by graphing bar charts for the different features.



#### 4. Discussion

Our last goal is to determine which of the cluster 3 neighborhoods is the optimal neighborhood to open our restaurant. We graph all the most important factors as a bar chart in order to visually determine which neighborhood is most complete in terms of all the factors. Again, there is no completely clear answer. There are however two neighborhoods that are more prominent than the rest.

- **Communtities-The Island:** This is our top choice. This neighborhood has the highest average income, highest population, high population of Arabic speakers, and zero Middle Eastern restaurants

- **Mount Olive-Silverstone-Jamestown:** This is our second choice. This neighborhood has the highest number of Arabic speakers, and no middle eastern restaurants. However this neighborhood has an average total population and a low average income as compared to other neighborhoods.

To pick the best out of these two restaurants, one should have a deep understanding of their business. These are some questions one will need to ask:

1. How economical is the pricing going to be? Will the people in the low income neighborhood afford to eat from the restaurant at a regular basis?
2. Do you think you will be able to get non Arab customers? If you are solely depending on Arab customers, it may make more sense to pick the neighborhood with the higher Arab population

## 5. Conclusion

Data science is a great tool which can be used in many business situations to help us make the best decisions possible given the resources that we have. In this scenario, we used data science, instead of just intuition in order to help us make our decision on which neighborhood to open our restaurants in. We learned that using data science by itself will not really get you anywhere. Instead, you need to apply data science with business knowledge. Business knowledge helps us when selecting the features that will be used. We also realize that we need to use business knowledge even when making decisions based on the results that we get from the algorithms.