
The Battle of Neighbourhoods: Opening a Middle Eastern Restaurant in Toronto, CA.

1. Introduction:

1.1: Description and Discussion of Background:

The objective of this project is to determine which neighbourhood in Toronto would be a good option for opening a Middle Eastern restaurant. Toronto is the most densely populated city in Canada and is well known for its diversity. Being a multicultural city provides countless opportunities for business owners to open such a business. A Middle Eastern restaurant would bring about Palestinians, Jordanians, Lebanese, Syrians as well as many others who enjoy the unique flavors of Middle Eastern food such as hummus, falafel, tabbouleh, manakeesh, shish tawooq, kofta, etc.



Opening a restaurant in the city of Toronto requires finding a neighbourhood that is densely populated and is positioned around numerous venues. These factors would help attract as many people as possible, for grabbing a bite or sitting down for lunch or dinner.

1.2: Data Overview:

The data utilized in this study include:

- A list of Toronto's boroughs, neighbourhoods, and postal codes on Wikipedia under: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
- A list of geographical location data, including postal codes as well as latitude and longitude coordinates, found under: https://cocl.us/Geospatial_data.
- A list of venue data, including the most common venues of a given neighbourhood in Toronto, via Foursquare API. These venues include restaurants, gyms, parks, etc.

2. Methodology:

2.1: Data Pre-processing:

- The list of Toronto's boroughs, neighbourhoods, and postal codes was scrapped from the Wikipedia page using *BeautifulSoup* library. The *html* table was converted into a *pandas* data frame and, then, cleaned:
 - Rows where the boroughs do not have an assigned value (i.e.: "Not Assigned") were dropped.
 - Rows where boroughs have an assigned value and neighbourhoods don't were edited, such that the name of the borough was assigned to its respective neighbourhood.
 - Neighborhoods having the same postal codes were combined into one row and separated by commas.

	Postal Code	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

- The list of geographical location data was read into a *pandas* data frame using the *read_csv()* method.

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

- The two data frames were merged together, based on the postal codes.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

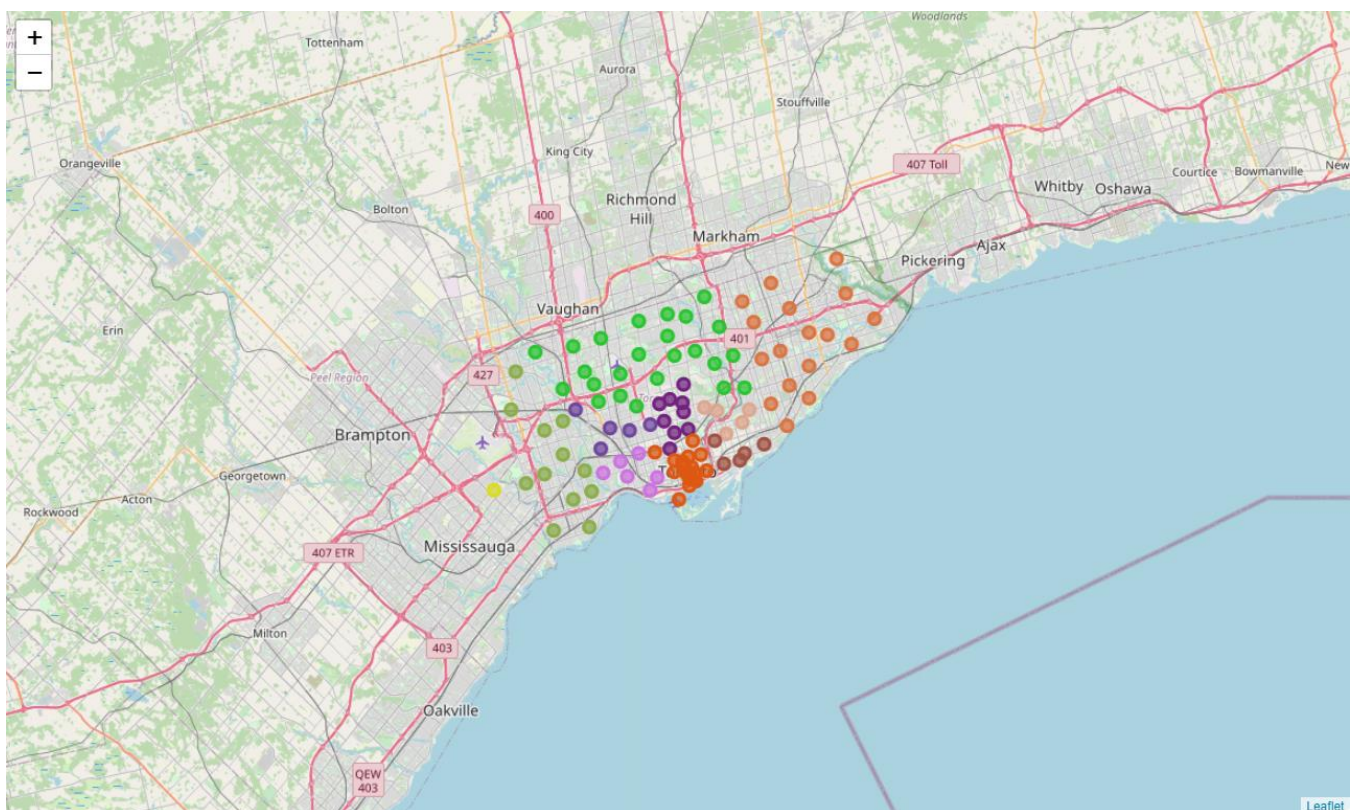
- The number of venues within an assigned radius from each neighbourhood center was retrieved using Foursquare API. The maximum number of venues and the radius were set to be 500 venues and 1000 meters, respectively.

A new data frame, containing neighbourhood and venue data, was created.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Allwyn's Bakery	43.759840	-79.324719	Caribbean Restaurant
1	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
2	Parkwoods	43.753259	-79.329656	Tim Hortons	43.760668	-79.326368	Café
3	Parkwoods	43.753259	-79.329656	A&W	43.760643	-79.326865	Fast Food Restaurant
4	Parkwoods	43.753259	-79.329656	Bruno's valu-mart	43.746143	-79.324630	Grocery Store

2.2: Data Exploration and Analysis:

- The neighbourhoods in the different boroughs of Toronto were visualized using *folium* library. Each borough was represented by a different color:



- The categorical values, representing the venues, were transformed into numerical values using the one-hot encoding technique. Then, the average of the frequency of each venue category that exist in each neighbourhood was calculated.

	Neighbourhood	ATM	Accessories Store	Afghan Restaurant	African Restaurant	Airport	American Restaurant	Amphitheater	A S
0	Agincourt	0.0	0.0	0.0	0.0	0.0	0.00000	0.0	
1	Alderwood, Long Branch	0.0	0.0	0.0	0.0	0.0	0.00000	0.0	
2	Bathurst Manor, Wilson Heights, Downsview North	0.0	0.0	0.0	0.0	0.0	0.00000	0.0	
3	Bayview Village	0.0	0.0	0.0	0.0	0.0	0.00000	0.0	
4	Bedford Park, Lawrence Manor East	0.0	0.0	0.0	0.0	0.0	0.02381	0.0	

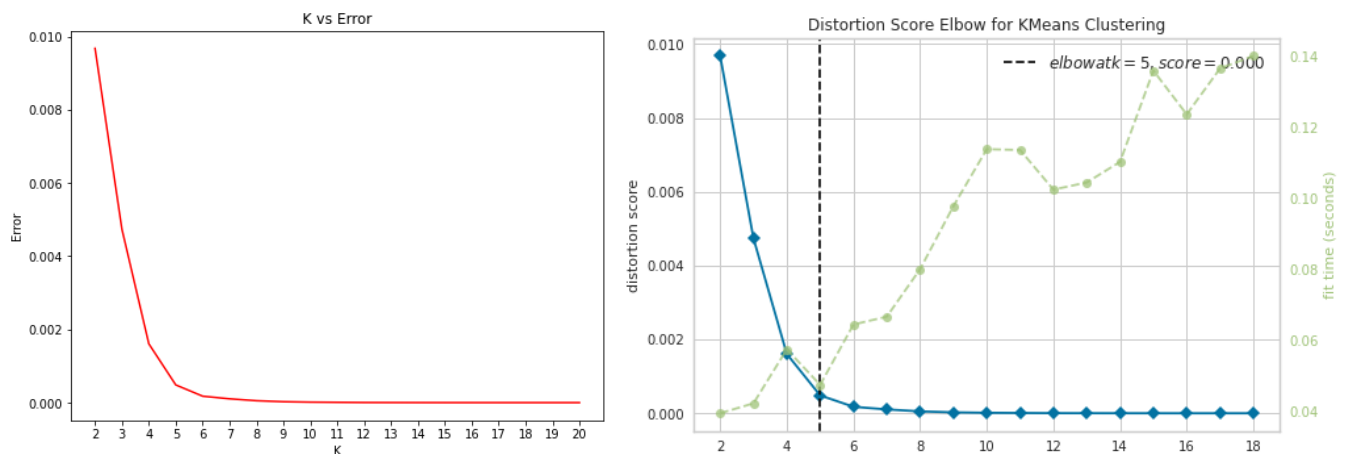
- A new data frame that stores the neighbourhood name and the mean frequency of the “Middle Eastern Restaurant” category in each neighbourhood was created.

	Neighbourhood	Middle Eastern Restaurant
0	Agincourt	0.000000
1	Alderwood, Long Branch	0.000000
2	Bathurst Manor, Wilson Heights, Downsview North	0.032258
3	Bayview Village	0.000000
4	Bedford Park, Lawrence Manor East	0.000000

- The “Neighbourhood” column was dropped to prepare the data for clustering.

	Middle Eastern Restaurant
0	0.000000
1	0.000000
2	0.032258
3	0.000000
4	0.000000

- Before the data was clustered, the optimum value of K was found using the elbow-point technique. In this method, the optimal K is chosen at the point where the line has the biggest downward turn, which is 5 in this case.



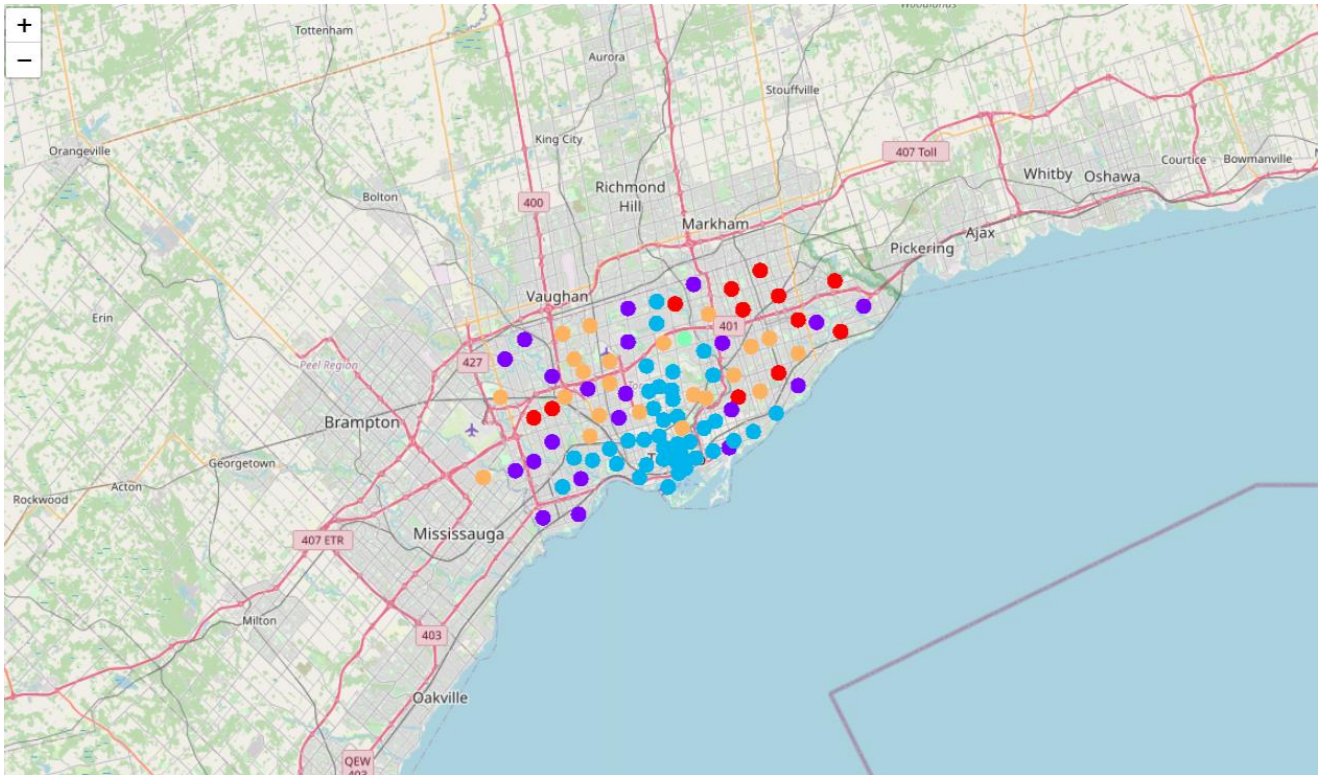
- The data was clustered into 5 distinct groups, based on the average of the frequency of venues in each neighbourhood. Neighbourhoods having similar averages were clustered together. Then, the cluster number for each neighborhood was added in a new column in the data frame.

	Borough	Neighbourhood
0	North York	Parkwoods
1	North York	Victoria Village
2	Downtown Toronto	Regent Park, Harbourfront
3	North York	Lawrence Manor, Lawrence Heights
4	Downtown Toronto	Queen's Park, Ontario Provincial Government

- The above data frame was merged wto include venue data as well. Then, it was sorted based on the assigned cluster numbers.

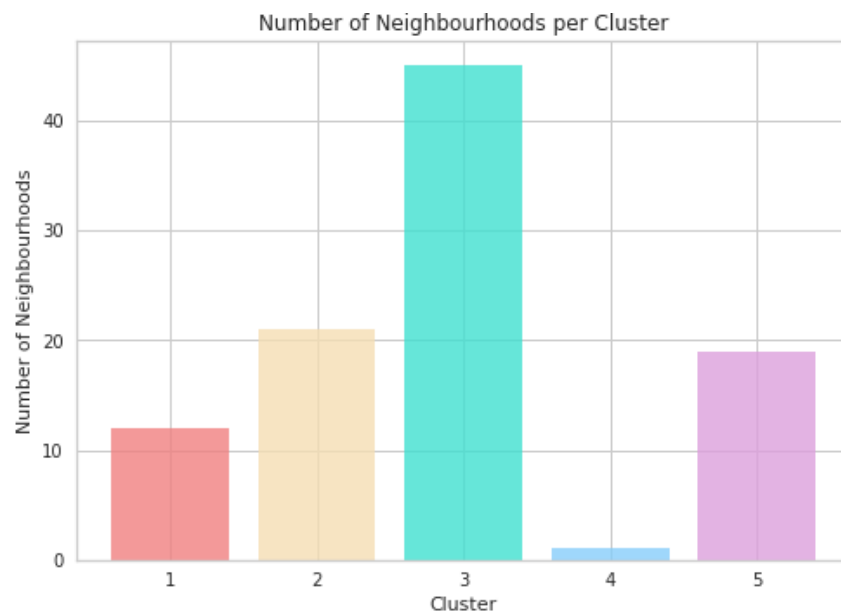
	Neighbourhood	Middle Eastern Restaurant	Cluster Labels	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Agincourt	0.0	0	43.794200	-79.262029	The Roti Hut	43.787277	-79.258724	Caribbean Restaurant
34	Guildwood, Morningside, West Hill	0.0	0	43.763573	-79.188711	Mr Mike's Pizza Company	43.771454	-79.186921	Pizza Place
34	Guildwood, Morningside, West Hill	0.0	0	43.763573	-79.188711	Kapit Bahay	43.769319	-79.184442	Filipino Restaurant
34	Guildwood, Morningside, West Hill	0.0	0	43.763573	-79.188711	First Choice Haircutters	43.769219	-79.184538	Salon / Barbershop
34	Guildwood, Morningside, West Hill	0.0	0	43.763573	-79.188711	Food Basics	43.770184	-79.184852	Supermarket

- The 5 different clusters were visualized on a map using *folium* library.



3. Results:

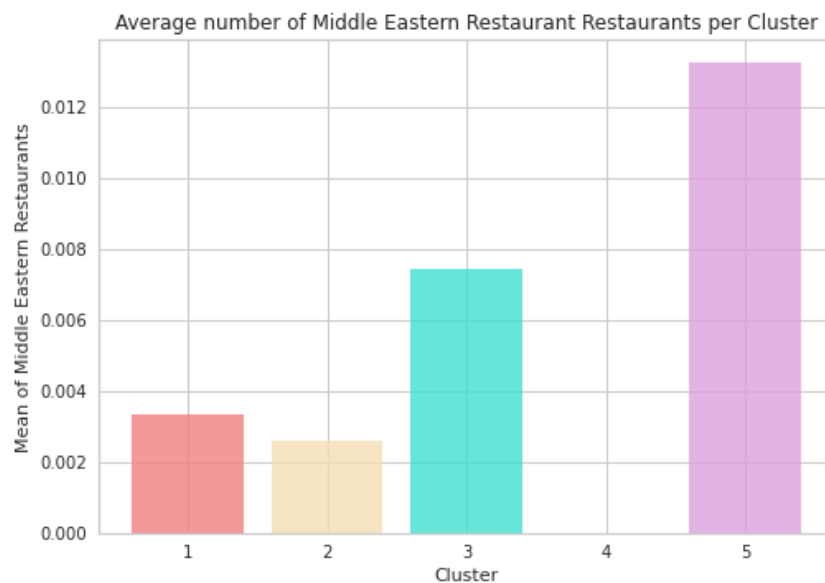
- The number of neighbourhoods in clusters 1, 2, 3, 4, and 5 are 12, 21, 45, 1, and 19, respectively. The majority of the neighbourhoods were combined under cluster 3. Whereas, cluster 4 consisted of only 1 neighbourhood. This means that no other neighborhood had a similar average frequency of venues to that of the neighbourhood in cluster 4.



- The number of venues in clusters 1, 2, 3, 4, and 5 are 300, 386, 3644, 4, and 831, respectively. This indicates that most of the venues are concentrated in neighbourhoods that fall under cluster 3. Also, cluster 4 has only 4 venues.



- The number of Middle Eastern restaurants in clusters 1, 2, 3, 4, and 5 are 1, 1, 27, 0, and 11, respectively. Moreover, the average number of Middle Eastern restaurants in clusters 1, 2, 3, 4, and 5 are 0.0033, 0.0026, 0.0074, 0.0, and 0.013, respectively. This is a very clear indicator that there is a scarcity of such cuisine in the city of Toronto, when compared to other venue categories.



4. Discussion

According to the results, the number of Middle Eastern restaurants in Toronto is considerably low; thus, decreasing the level of competition. Still, the optimal option would be a place in one of the neighbourhoods that fall under cluster 3. These areas had the highest number of venues and, at the same time, only the second highest number of Middle Eastern restaurants around. This means that opening the desired restaurant in such neighbourhoods would attract large numbers of people and have little to no competition at all – in terms of similar cuisine.

Neighbourhoods under clusters 1, 2, and 4 were not considered as a good option as the number of venues in these areas is extremely low. A new restaurant in one of these neighbourhoods would mostly attract people living nearby only. This would be optimal in the case where many Middle Eastern people live in these areas; however, this would require retrieving and analyzing further data including people demographics in Toronto.

Neighbourhoods under cluster 5 had a substantially smaller amount of venues when compared to those under cluster 3. Also, several Middle Eastern restaurants were located in those areas, making up the highest average of such cuisine compared to the rest of the clusters. Thus, this option was eliminated as well.

In a nutshell, neighborhoods under cluster 3 provide an optimum location for opening a Middle Eastern restaurant in Toronto. Of course, considering other details, such as rental or ownership prices of venues and/or demographic data of people, would definitely improve the accuracy of this study and result in a more precise recommendation.

5. Conclusion

All in all, large multicultural cities are attractive places for people to start their own business. In this study, the best neighbourhood in Toronto for opening a Middle Eastern restaurant was determined. This was done by using numerous Python libraries in order to fetch, analyze, and visualize the data. These included *pandas*, *BeautifulSoup*, *seaborn*, *matplotlib*, *scikit-learn*, and *folium*. Also, Foursquare API was utilized to explore the settings in the neighbourhoods of Toronto.

Determining the best place to open this restaurant was based on the average number of surrounding venues for each neighbourhood. The optimal neighbourhoods were ones having very little competition and a very high number of surrounding venues (i.e.: having a dense population).