Optimal Location to Open a Restaurant in Toronto

Govind Lipne

1. Introduction

1.1. Background

Toronto is the capital of the province of Ontario and a major Canadian city located along Lake Ontario's northwestern shore. It is highly international with various ethnicities, which accordingly presents entrepreneurs with various possible business opportunities, including opening ethnic restaurants.

1.2. Problem

A client seeks advice as to where to open a Persian restaurant in Toronto to have an optimal return on investment and future growth prospects. This project is aimed to enable the client to make an informed decision about this business venture by locating the optimal neighborhood(s) for such business operations. The following factors seem to be the most effective factors in determining the right neighborhood:

1. Competition: Existence of a multitude of Persian restaurant in a neighborhood most likely will act as a barrier to entry for new entrepreneurs.
2. Density of the Persian community in the neighborhood: It is reasonable to expect more sales and revenue in neighborhoods with a denser Persian community.
3. Spending power: Generally, the more money and buying power people have, the more they are likely to spend money on restaurants and the like.

1.3. Interest

Potential stakeholders/audience of this project who will benefit from the outcomes are entrepreneurs seeking to find an optimal neighborhood to open a Persian restaurant or intending to expand their franchise to other locations.

2. Data acquisition and cleaning

For information on the abovementioned factors, we use the census data available on Toronto's Open Data Portal, and after cleaning the data, we store them in the following files for easier usage in the future:

- negighbourhood_population.csv
- persian_population.csv
- income_avg.cs

The geographical coordinates of each neighborhood in Toronto in available at this link. The list of neighborhoods and boroughs along with their postal codes can be downloaded from this link. While gathering this list, we only process the cells that have an assigned borough and ignore cells with a borough that is Not assigned.

Since the popularity of the restaurant in a neighborhood depends on the density of the Persian community, i.e., the Persian community as compared to the total population in the neighborhood, the percentage of the Persian community was calculated and used as the density factor. Table 1 shows the first 5 rows of the final dataframe. As we see from the table, the data set is now clean and ready for the next steps. Now, it has information about two of the abovementioned factors. i.e., Persian density and buying power. For the third factor, we will be using the Foursquare API to find the competition in each area. Then, we will move to visualizing the data and learning from the data using machine learning.

Table 1. The first 5 rows of data after cleaning and transformation

	Neighbourhood	Latitude	Longitude	Average After-tax income	Percentage of Persian
0	Victoria Village	43.725882	-79.315572	30769.0	3.198172
1	Rouge	43.806686	-79.194353	33736.0	0.838782
2	Malvern	43.806686	-79.194353	26505.0	0.685025
3	Highland Creek	43.784535	-79.160497	34837.0	0.760365
4	Flemingdon Park	43.725900	-79.340923	25608.0	6.269092

3. Exploratory Data Analysis

1. Creating a Folium Map

We use Folium to visualize the neighborhoods in Toronto. The map is shown in Figure 1. Later on, when we cluster the neighborhoods, we will again show the clusters with different colors on a map, like what is shown in the figure.

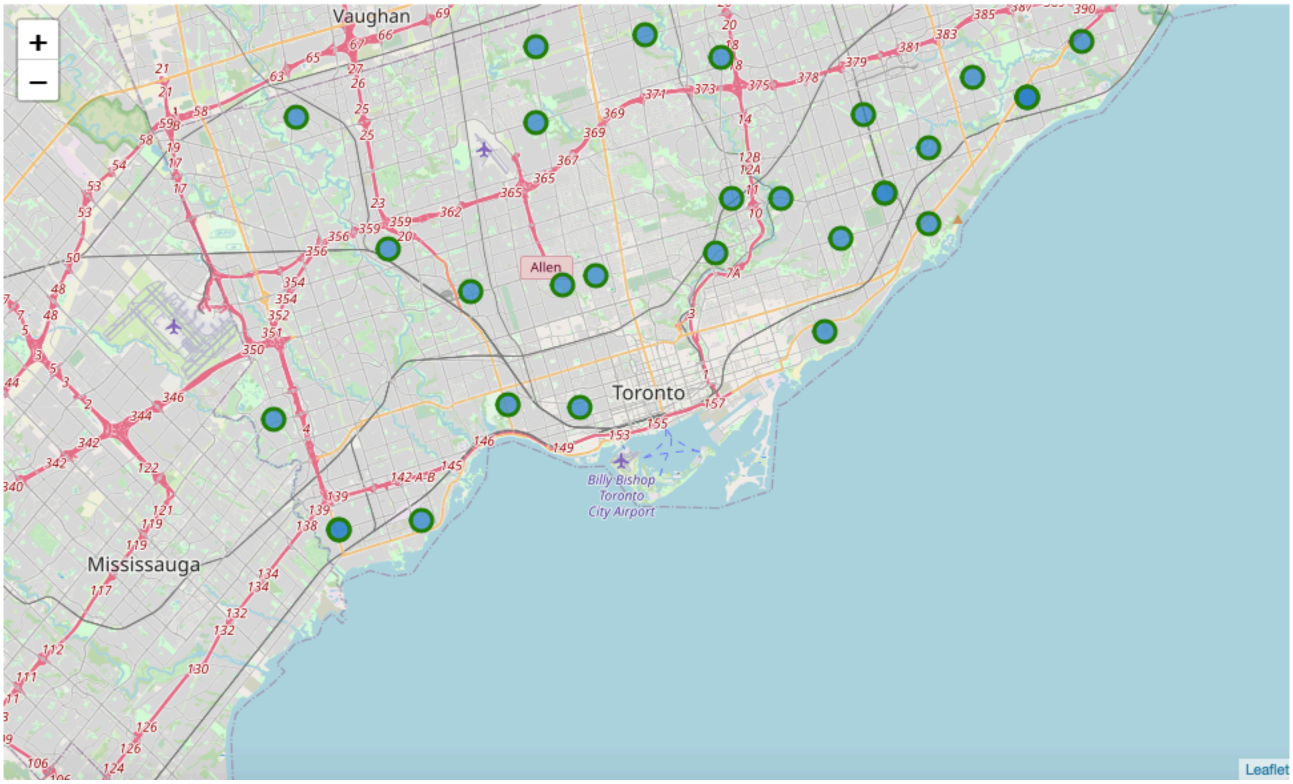


Figure 1. A view of the neighborhoods in a map generated by Folium

2. Exploring the Neighbourhood Using the Foursquare API

In this section, we utilize the Foursquare API to explore the neighborhoods and segment them. Table 2 shows the data extracted from the API, representing the top 100 venues in each neighborhood in Toronto within a radius of 1500 meters from the neighborhoods. As shown in Table 3, the chosen total number of venues for the search (i.e., 100) pretty much covers all the venues in a radius of 1500 meters from each neighborhood.

Table 2. The top 100 venues in each neighborhood and the relevant categories

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
1	Victoria Village	43.725882	-79.315572	Armenian Kitchen	43.731071	-79.305390	Middle Eastern Restaurant
2	Victoria Village	43.725882	-79.315572	Sultan Of Samosas	43.718823	-79.304350	Indian Restaurant
3	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
4	Victoria Village	43.725882	-79.315572	Aga Khan Museum	43.725105	-79.332076	History Museum
...	...	...	...	...	...	...	...
1794	Long Branch	43.602414	-79.543484	EB Games	43.611490	-79.555906	Video Game Store
1795	Long Branch	43.602414	-79.543484	Nordstrom	43.610807	-79.556781	Department Store
1796	Long Branch	43.602414	-79.543484	Nordstrom Ebar Artisan Coffee	43.610877	-79.556794	Café
1797	Long Branch	43.602414	-79.543484	Danish Pastry House	43.611361	-79.557073	Bakery
1798	Long Branch	43.602414	-79.543484	lululemon athletica	43.611894	-79.556627	Clothing Store

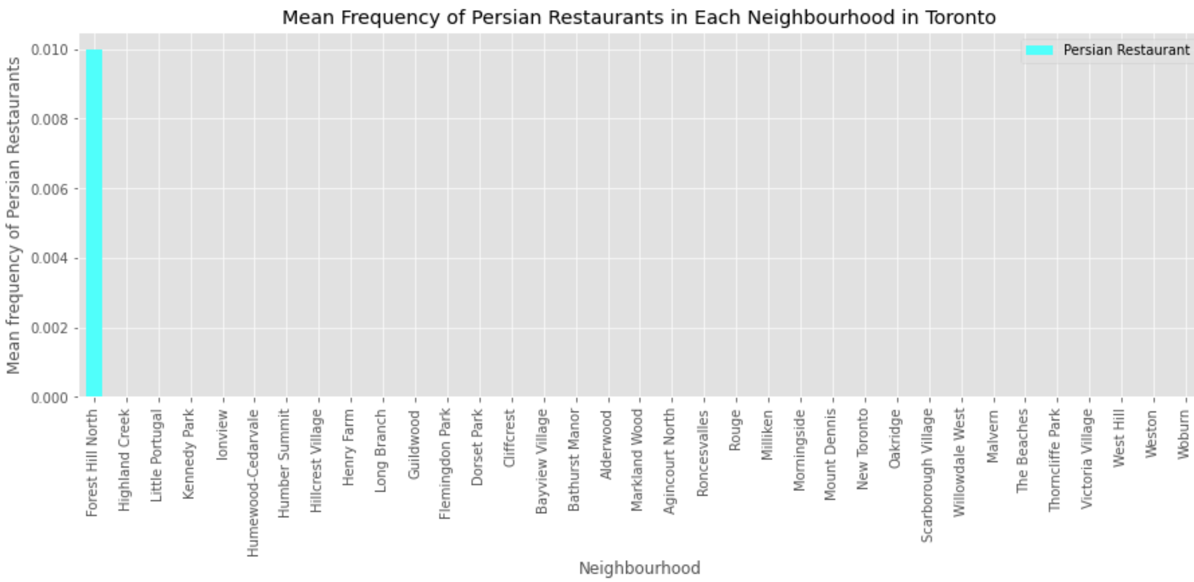
Table 3. Total count of the found data/venues in each neighborhood

Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Agincourt North	79	79	79	79	79	79
Alderwood	43	43	43	43	43	43
Bathurst Manor	40	40	40	40	40	40
Bayview Village	16	16	16	16	16	16
Cliffcrest	38	38	38	38	38	38
Dorset Park	57	57	57	57	57	57
Flemingdon Park	85	85	85	85	85	85
Forest Hill North	99	99	99	99	99	99
Guildwood	31	31	31	31	31	31
Henry Farm	64	64	64	64	64	64
Highland Creek	11	11	11	11	11	11
Hillcrest Village	54	54	54	54	54	54
Humber Summit	20	20	20	20	20	20
Humewood-Cedarvale	88	88	88	88	88	88
Ionview	36	36	36	36	36	36
Kennedy Park	36	36	36	36	36	36
Little Portugal	100	100	100	100	100	100
Long Branch	43	43	43	43	43	43
Malvern	32	32	32	32	32	32
Markland Wood	38	38	38	38	38	38
Milliken	79	79	79	79	79	79
Morningside	31	31	31	31	31	31
Mount Dennis	39	39	39	39	39	39
New Toronto	39	39	39	39	39	39
Oakridge	35	35	35	35	35	35
Roncesvalles	100	100	100	100	100	100
Rouge	32	32	32	32	32	32
Scarborough Village	35	35	35	35	35	35
The Beaches	100	100	100	100	100	100
Thorncliffe Park	94	94	94	94	94	94
Victoria Village	49	49	49	49	49	49
West Hill	31	31	31	31	31	31
Weston	51	51	51	51	51	51
Willowdale West	42	42	42	42	42	42
Woburn	32	32	32	32	32	32

3. Analyzing Each Neighborhood

3.1. Frequency Distribution of Persian Restaurants in Each Neighborhood

Figure 2 shows the mean frequency of Persian restaurants in each neighborhood. Results show that all the Persian restaurants are found in the Forest Hill North area. Hence, to avoid barriers to entry by these existing restaurants, it is not advised to open the new restaurant in this area.



3.2. Distribution of Persian Population in Each Neighborhood

Figure 3 shows the percentage of Persian population in each neighborhood. Results show that Bayview Village has the highest density of Persian population among the neighborhoods and, therefore, should be the most suitable for opening a Persian restaurant with regards to this factor.

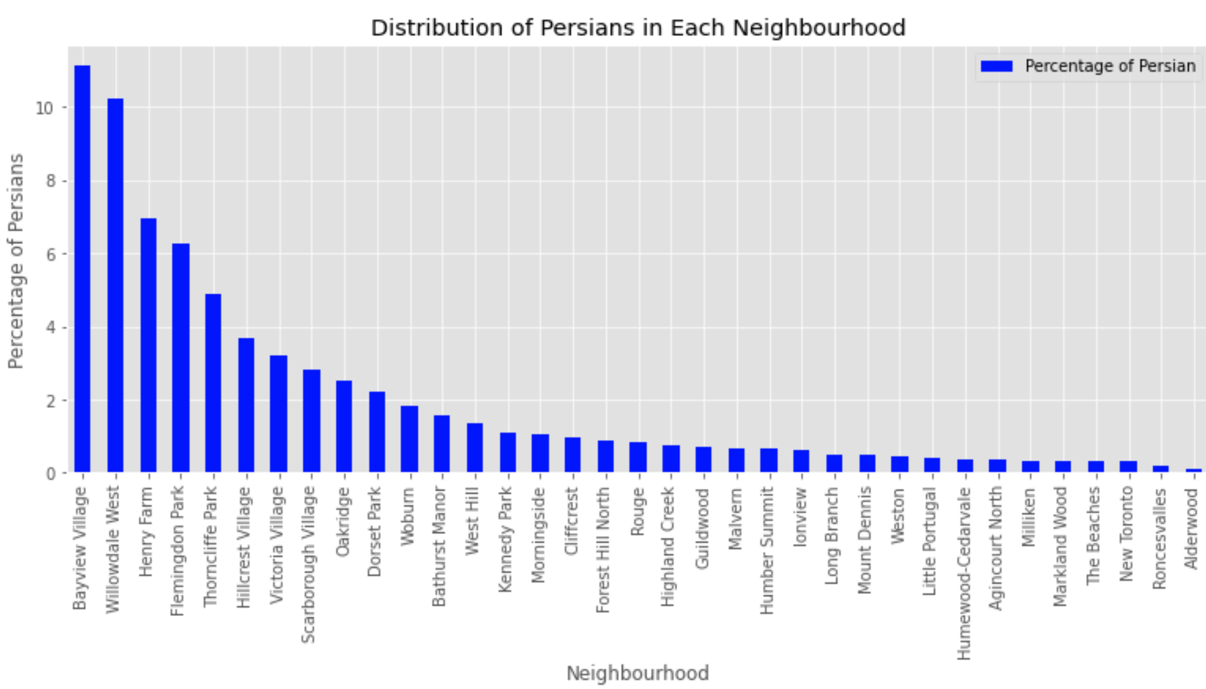
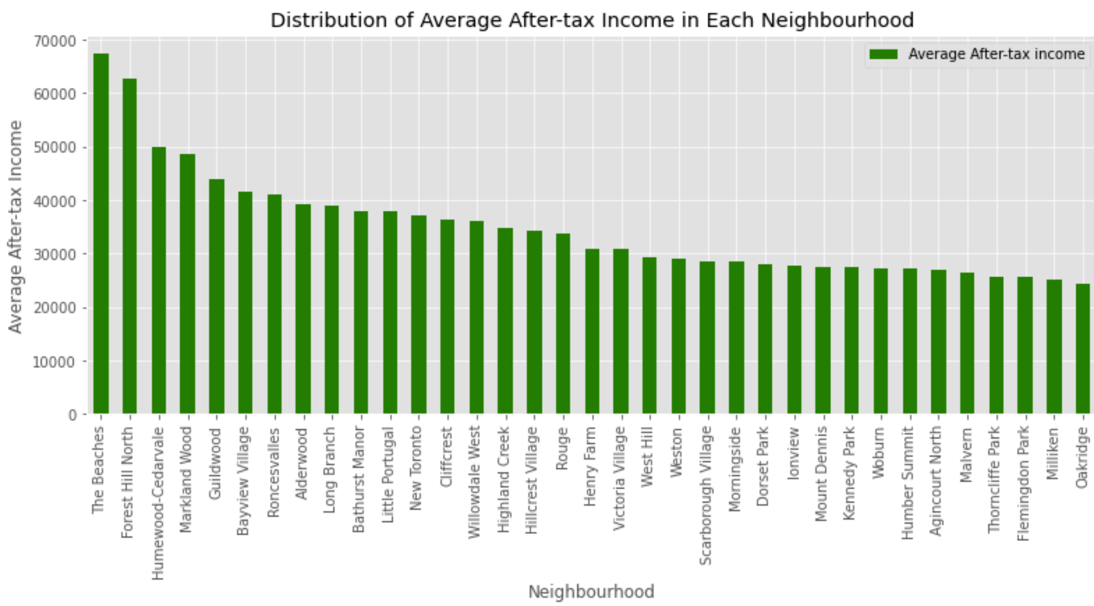


Figure 4 shows the distribution of average after-tax income in each neighborhood. Results show that The Beaches has the highest income rating. This neighborhood, however, had one of the lowest densities of Persian population. Therefore, it may not be a good candidate for opening the new restaurant. The next candidate with regards to income is, interestingly, the Forest Hill North area, which already contains all the existing Persian restaurants and is not a suitable choice due to being a barrier to entry. Among the next candidates, Bayview Village seems to be the best candidate as it also has the highest density of Persian population.



4. Clustering the Neighborhoods

We use k-means to cluster the data into 5 clusters. First, we need to make the data ready for the learning algorithm. We do this by normalizing the data, using mean and standard deviation, so that Kmeans treats all the variables the same in terms of variation. Table 4 shows the learned cluster labels along with other stats for each neighborhood. The updated map is shown in Figure 5.

	Neighbourhood	Latitude	Longitude	Average After-tax income	Percentage of Persian	Cluster Labels	Persian Restaurant Mean Freq.
0	Victoria Village	43.725882	-79.315572	30769.0	3.198172	0	0.0
1	Rouge	43.806686	-79.194353	33736.0	0.838782	3	0.0
2	Malvern	43.806686	-79.194353	26505.0	0.685025	0	0.0
3	Highland Creek	43.784535	-79.160497	34837.0	0.760365	3	0.0
4	Flemingdon Park	43.725900	-79.340923	25608.0	6.269092	4	0.0

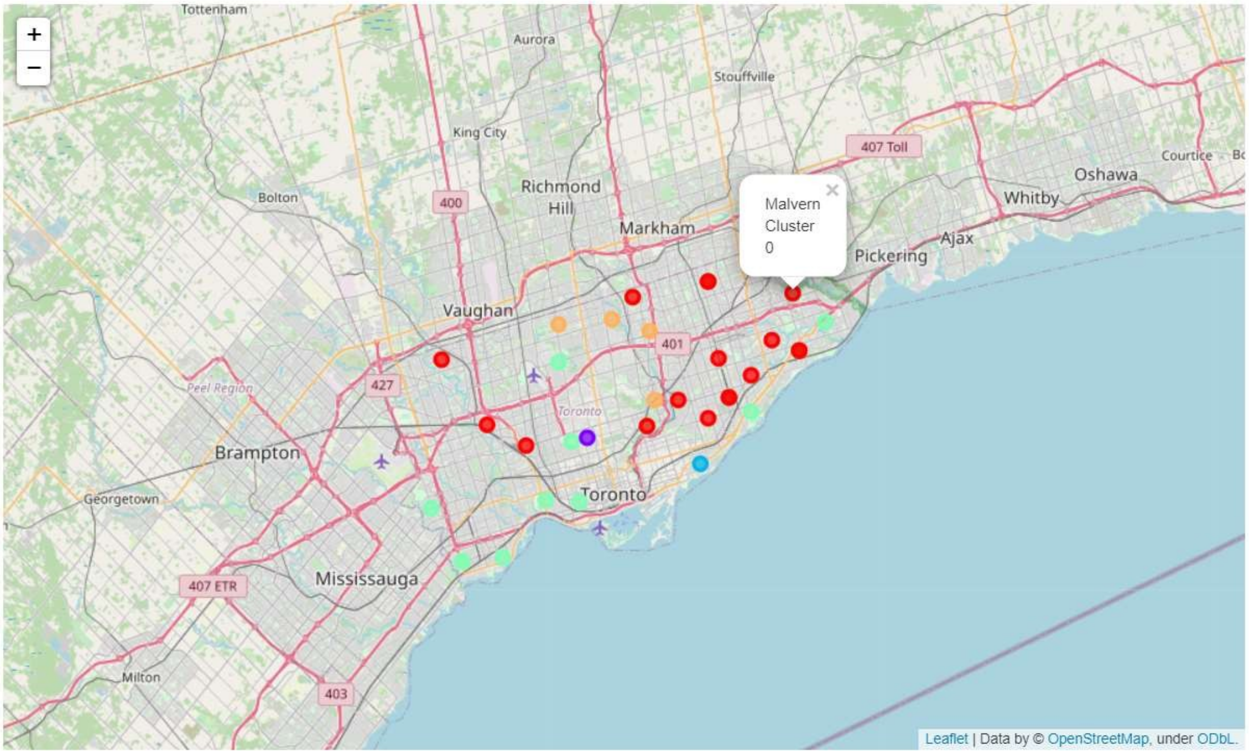


Figure 5. Updated map of neighborhoods showing different clusters with different colors

1. Examining and Analyzing the Clusters

Now, we can examine the clusters and determine the discriminating factors that distinguish each cluster. Tables 5 to 9 show the neighborhoods in clusters 1 to 5, respectively. By digging into these tables, we see that we can label them as follows:

1. Cluster 1: Neighborhoods with a medium Persian population, low income, and no local Persian Restaurant.
2. Cluster 2: Neighborhoods with a medium Persian population, high income, and with local Persian Restaurants (with competition).
3. Cluster 3: Neighborhoods with a low Persian population, high income, and no local Persian Restaurant.
4. Cluster 4: Neighborhoods with a low Persian population, medium income, and no local Persian Restaurant.

5. Cluster 5: Neighborhoods with a high Persian population, medium income, and no local Persian Restaurant.

Thus, Cluster 5 seems to have the most attractive profile with regards to out metrics (no competition, high density of Persian population, and a reasonable spending power). Among all the neighborhoods in Cluster 5, Bayview Village seems to be the best candidate due to having the highest income and Persian population (as compared to the other neighborhoods in Cluster 5).Interestingly, this result that we got from Kmeans clustering agrees well with our results from Data Exploration in section 3.

Table 5. Neighborhoods of cluster 1 and their attributes

	Neighbourhood	Latitude	Longitude	Average After-tax income	Percentage of Persian	Cluster Labels	Persian Restaurant Mean Freq.
0	Victoria Village	43.725882	-79.315572	30769.0	3.198172	0	0.0
1	Malvern	43.806686	-79.194353	26505.0	0.685025	0	0.0
2	Morningside	43.763573	-79.188711	28497.0	1.031223	0	0.0
3	West Hill	43.763573	-79.188711	29221.0	1.369013	0	0.0
4	Woburn	43.770992	-79.216917	27341.0	1.841638	0	0.0
5	Hillcrest Village	43.803762	-79.363452	34203.0	3.661273	0	0.0
6	Thorncliffe Park	43.705369	-79.349372	25675.0	4.903354	0	0.0
7	Scarborough Village	43.744734	-79.239476	28634.0	2.810332	0	0.0
8	Ionview	43.727929	-79.262029	27852.0	0.623121	0	0.0
9	Kennedy Park	43.727929	-79.262029	27439.0	1.109619	0	0.0
10	Oakridge	43.711112	-79.284577	24247.0	2.527988	0	0.0
11	Humber Summit	43.756303	-79.565963	27334.0	0.644330	0	0.0
12	Mount Dennis	43.691116	-79.476013	27461.0	0.478187	0	0.0
13	Weston	43.706876	-79.518188	28965.0	0.444642	0	0.0
14	Dorset Park	43.757410	-79.273304	28052.0	2.199736	0	0.0
15	Agincourt North	43.815252	-79.284577	26955.0	0.343489	0	0.0
16	Milliken	43.815252	-79.284577	25109.0	0.338702	0	0.0

Table 6. Neighborhoods of cluster 2 and their attributes

	Neighbourhood	Latitude	Longitude	Average After-tax income	Percentage of Persian	Cluster Labels	Persian Restaurant Mean Freq.
0	Forest Hill North	43.696948	-79.411307	62714.0	0.858972	1	0.010101

Table 7. Neighborhoods of cluster 3 and their attributes

	Neighbourhood	Latitude	Longitude	Average After-tax income	Percentage of Persian	Cluster Labels	Persian Restaurant Mean Freq.
0	The Beaches	43.676357	-79.293031	67490.0	0.32457	2	0.0

Table 8. Neighborhoods of cluster 4 and their attributes

	Neighbourhood	Latitude	Longitude	Average After-tax income	Percentage of Persian	Cluster Labels	Persian Restaurant Mean Freq.
0	Rouge	43.806686	-79.194353	33736.0	0.838782	3	0.0
1	Highland Creek	43.784535	-79.160497	34837.0	0.760365	3	0.0
2	Humewood-Cedarvale	43.693781	-79.428191	49925.0	0.382875	3	0.0
3	Markland Wood	43.643515	-79.577201	48586.0	0.331628	3	0.0
4	Guildwood	43.763573	-79.188711	43848.0	0.705859	3	0.0
5	Bathurst Manor	43.754328	-79.442259	37927.0	1.575002	3	0.0
6	Little Portugal	43.647927	-79.419750	37924.0	0.417765	3	0.0
7	Cliffcrest	43.716316	-79.239476	36364.0	0.972702	3	0.0
8	Roncesvalles	43.648960	-79.456325	40926.0	0.200347	3	0.0
9	New Toronto	43.605647	-79.501321	37209.0	0.305330	3	0.0
10	Alderwood	43.602414	-79.543484	39159.0	0.124440	3	0.0
11	Long Branch	43.602414	-79.543484	39000.0	0.495835	3	0.0

Table 9. Neighborhoods of cluster 5 and their attributes

	Neighbourhood	Latitude	Longitude	Average After-tax income	Percentage of Persian	Cluster Labels	Persian Restaurant Mean Freq.
0	Flemingdon Park	43.725900	-79.340923	25608.0	6.269092	4	0.0
1	Henry Farm	43.778517	-79.346556	30931.0	6.964320	4	0.0
2	Bayview Village	43.786947	-79.385975	41440.0	11.146943	4	0.0
3	Willowdale West	43.782736	-79.442259	36093.0	10.244450	4	0.0

5. Summary and Future Directions

In this project, data was gathered from different sources and consolidated into a dataframe to find the best location for opening a Persian restaurant in Toronto. Three factors were defined to narrow down the selections: competition, Persian population, and income. The results of both data exploration and Kmeans clustering showed Bayview Village to be the best area in terms of the aforementioned factors. For future explorations and refinements, it is recommended to define and examine other possible factors. For example, one interesting factor could be the density of hotels in each neighborhood, as hotel customers usually become customers of the restaurants in the neighborhood of their hotel of residence. Another factor of interest can be proximity to recreational venues and historical places where there is a high likelihood that people go to restaurant after visiting the recreational place. Even proximity to industrial sections of the city, where there is a high possibility that people go to a restaurant during their lunch time or for client meetings, can be a viable factor. Going deeper into the bolts and nuts of the business, factors like rates for renting the restaurant place or land price for buying the place in each neighborhood can be important factors, when considered in the context of target profitability and similar factors, although such factors should be clearly defined and customized by the client.