

Market Analysis for an Indian Restaurant in Toronto

Table of Contents

<i>Introduction/Business Problem</i>	2
Target Audience	2
<i>Data</i>	2
1. Collecting Data	2
2. Data Cleaning	3
3. Data Exploration / Analysis	3
4. Clustering	3
<i>Methodology</i>	3
Visualize Toronto Map	3
Venues in Toronto using FourSquare.....	3
Indian Demography from Statcan Portal	4
Frequency Distribution of Indian Restaurants / Competitor Landscape	4
Frequency Distribution of Indian Population / Demography.....	5
K Means Clustering.....	5
Find best k.....	5
Clustering using k = 6.....	6
<i>Results</i>	8
<i>Discussion</i>	8
<i>Conclusion</i>	9

Introduction/Business Problem

Ajay is a successful restaurateur based in Mumbai, India. Under his brand AJ hospitality group, he has a chain of fine dining restaurants across India. He intends to expand his business and plans to start an Indian fine dining restaurant in Toronto, Canada in 2021. To begin with, AJ hospitality group approaches an analytics company in Toronto and assigns them with the task of conducting a market feasibility study.

The objective for the company is to understand and analyze the neighborhoods in Toronto suitable for AJ hospitality group to start their operations. They have to collect information on Indian population in different neighborhoods in Toronto and the competitors in the area. Finally, they intend to segregate the neighborhoods to different clusters based on similarities of factors mentioned above. This will help AJ hospitality group to make an intelligent decision by January 2021.

Target Audience

Such a study would be beneficial for anyone planning to setup a restaurant in Toronto. The analytics company can make use of this model for their other restaurant clients who are interested in understanding the ethnicity-competitor landscape before setting up their base in Toronto.

Data

The different processes involved in this stage are mentioned below

1. Collecting Data

We will use the below wikipedia page to gather the details of all the neighbourhoods in Toronto.

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

We will use BeautifulSoup python library for web scraping.

Geographical co-ordinates for the neighborhoods are obtained from the below link.
https://cocl.us/Geospatial_data

Foursquare API will be used to get the list of Indian restaurants in each neighborhood. This can be used to establish a competitor landscape

We will use 'Statistics Canada' portal to find out the demography details (South Asian population)

<https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CSD&Code1=3520005&Geo2=PR&Code2=35&Data=Count&SearchType=Begins&SearchPR=01&B1>All>

2. Data Cleaning

We will bring all the data from step 1 together, into a single dataframe which consists of neighborhoods, postal codes, geo-coordinates and South Asian (or Indian) population details.

3. Data Exploration / Analysis

We will use Foursquare API to explore the Indian restaurants in each neighbourhood and then we will group rows by neighborhood and then take the mean of the frequency of occurrence of Indian restaurants in each neighborhood. A similar exercise is performed for population as well

4. Clustering

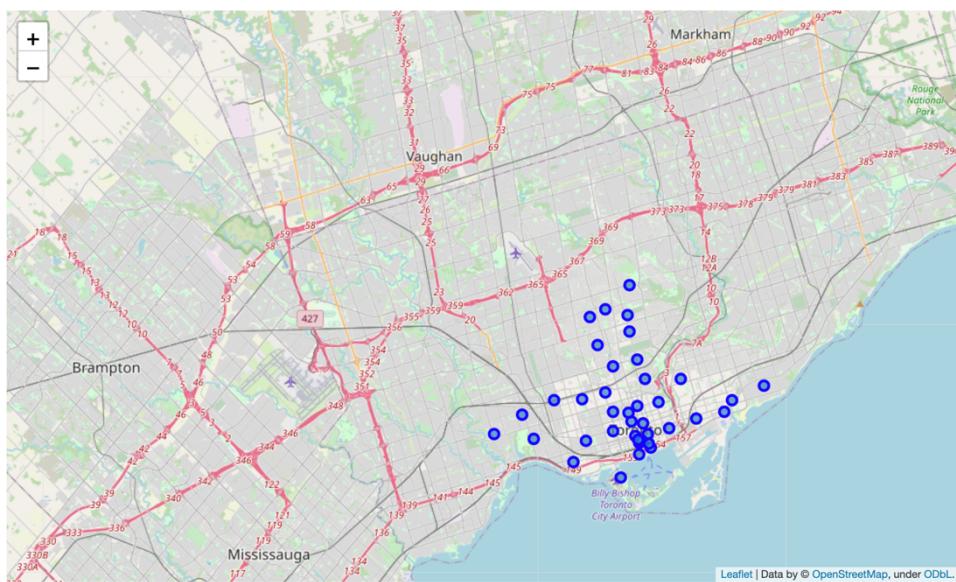
The final stage would be clustering where k-Means clustering technique is used to come up with multiple clusters based on competitor landscape and South Asian population.

This final stage will help AJ hospitality to make an intelligent decision as to where the restaurant should be located.

Methodology

Visualize Toronto Map

Initially, a map of Toronto was created with neighborhoods superimposed on top. I used Folium library for map visualization.



Venues in Toronto using Foursquare

I used FourSquare API to list all the venues in the Toronto neighborhoods. Further, only venues with the keyword 'Indian' was filtered out. A percentage of this was calculated and appended to the dataframe

	Neighborhood	Indian Restaurant
0	Berczy Park	0.000000
20	Moore Park, Summerhill East	0.000000
21	North Toronto West	0.000000
22	Parkdale, Roncesvalles	0.000000
23	Queen's Park, Ontario Provincial Government	0.000000
...
4	Central Bay Street	0.015873
30	St. James Town, Cabbagetown	0.020408
36	The Danforth West, Riverdale	0.023810
8	Davisville	0.030303
34	The Annex, North Midtown, Yorkville	0.043478

39 rows × 2 columns

Indian Demography from Statcan Portal

From the statcan portal (<https://www12.statcan.gc.ca/census-recensement/2011/dp-pd/hlt-fst/pd-pl/FullFile.cfm?T=1201&LANG=Eng&OFT=CSV&OFN=98-310-XWE2011002-1201.CSV>), the Indian dwelling information was retrieved as CSV. This is based on the postal codes.

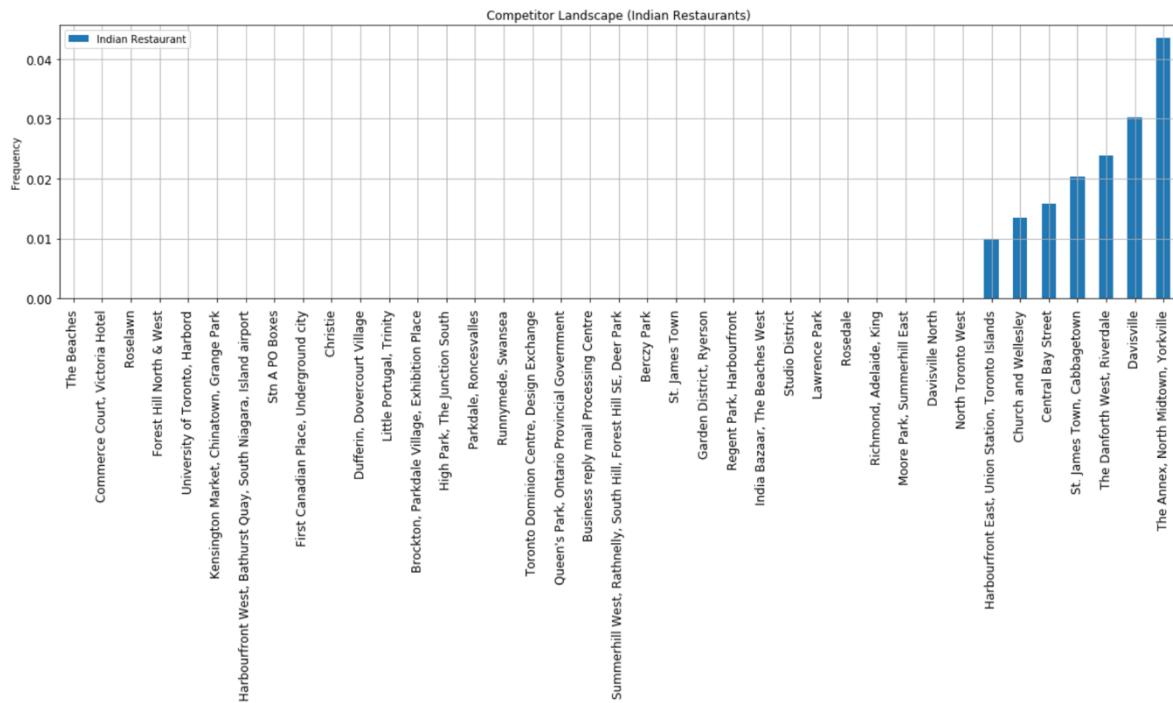
This was merged with the neighborhood dataframe to form the full dataset which includes 'neighborhood', 'Indian Restaurant Percentages' and 'Dwellings' information.

PostalCode	Dwellings	Borough	Neighborhood	Latitude	Longitude	Indian Restaurant	
0	M4E	11346.0	East Toronto	The Beaches	43.676357	-79.293031	0.00000
1	M4K	15058.0	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	0.02381
2	M4L	14420.0	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572	0.00000
3	M4M	9969.0	East Toronto	Studio District	43.659526	-79.340923	0.00000
4	M4N	6377.0	Central Toronto	Lawrence Park	43.728020	-79.388790	0.00000
...	
34	M6P	19322.0	West Toronto	High Park, The Junction South	43.661608	-79.464763	0.00000
35	M6R	9149.0	West Toronto	Parkdale, Roncesvalles	43.648960	-79.456325	0.00000
36	M6S	14240.0	West Toronto	Rannymede, Swansea	43.651571	-79.484450	0.00000
37	M7A	1.0	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0.00000
38	M7Y	1.0	East Toronto	Business reply mail Processing Centre	43.662744	-79.321558	0.00000

39 rows × 7 columns

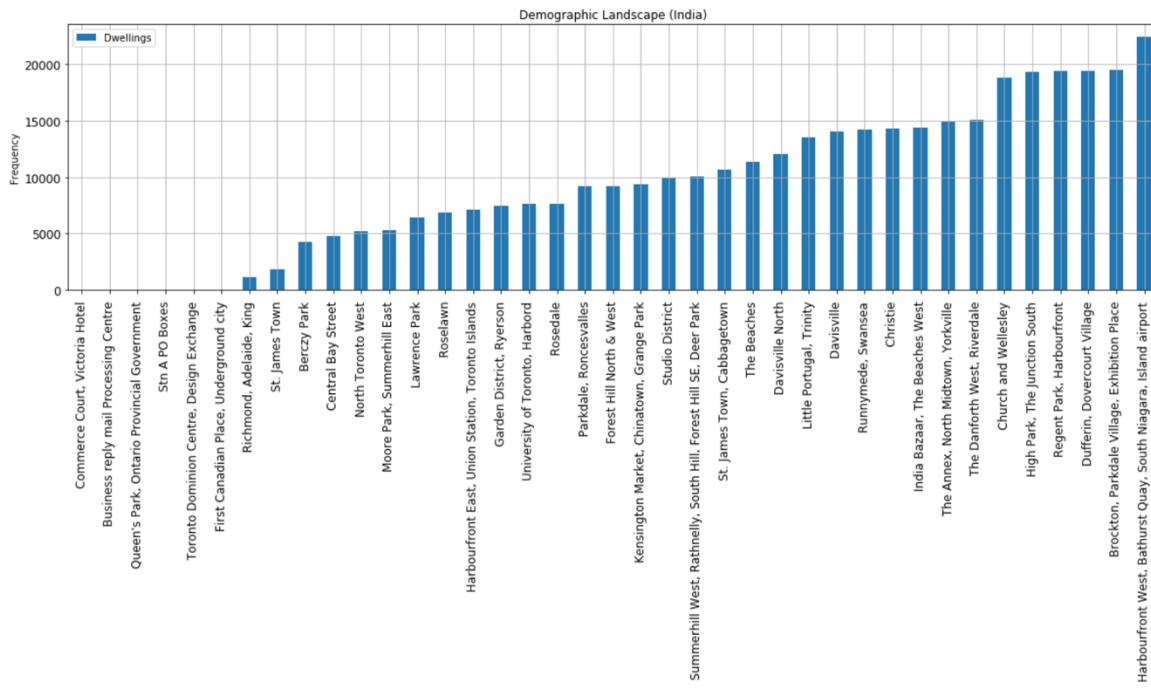
Frequency Distribution of Indian Restaurants / Competitor Landscape

The data from the dataframe above was used to plot the competitor landscape (as bar plot)



Frequency Distribution of Indian Population / Demography

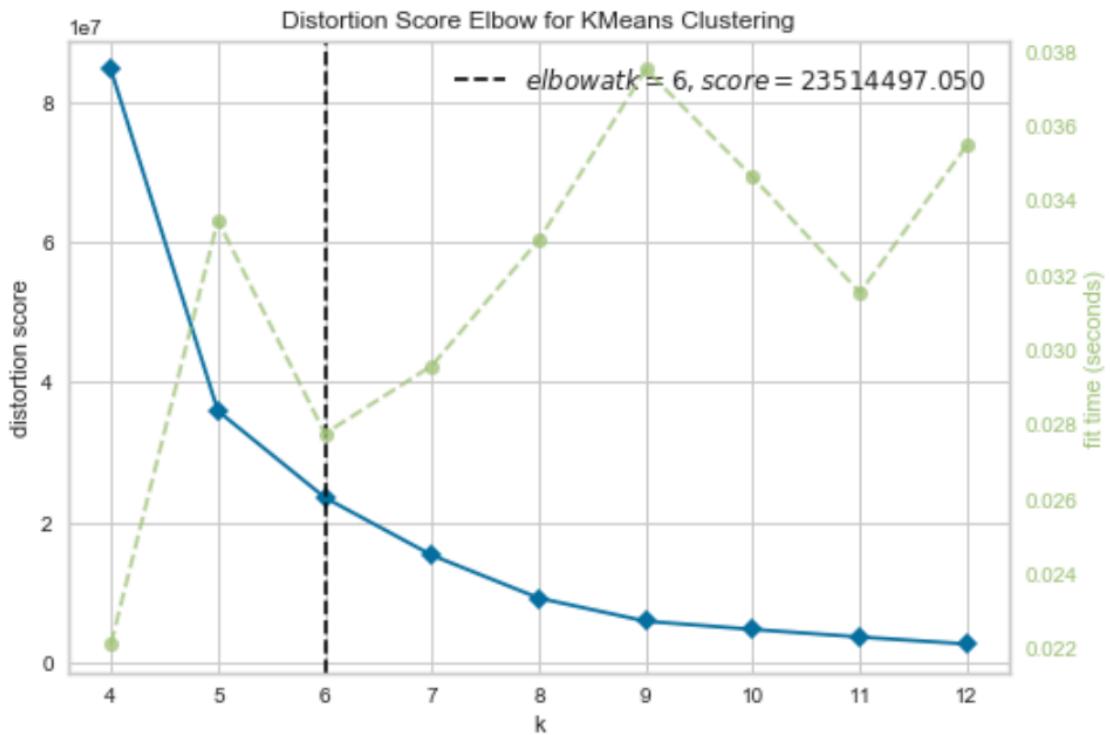
Similarly, a bar plot was plotted to visualize the Indian demography details in each neighborhood



K Means Clustering

Find best k

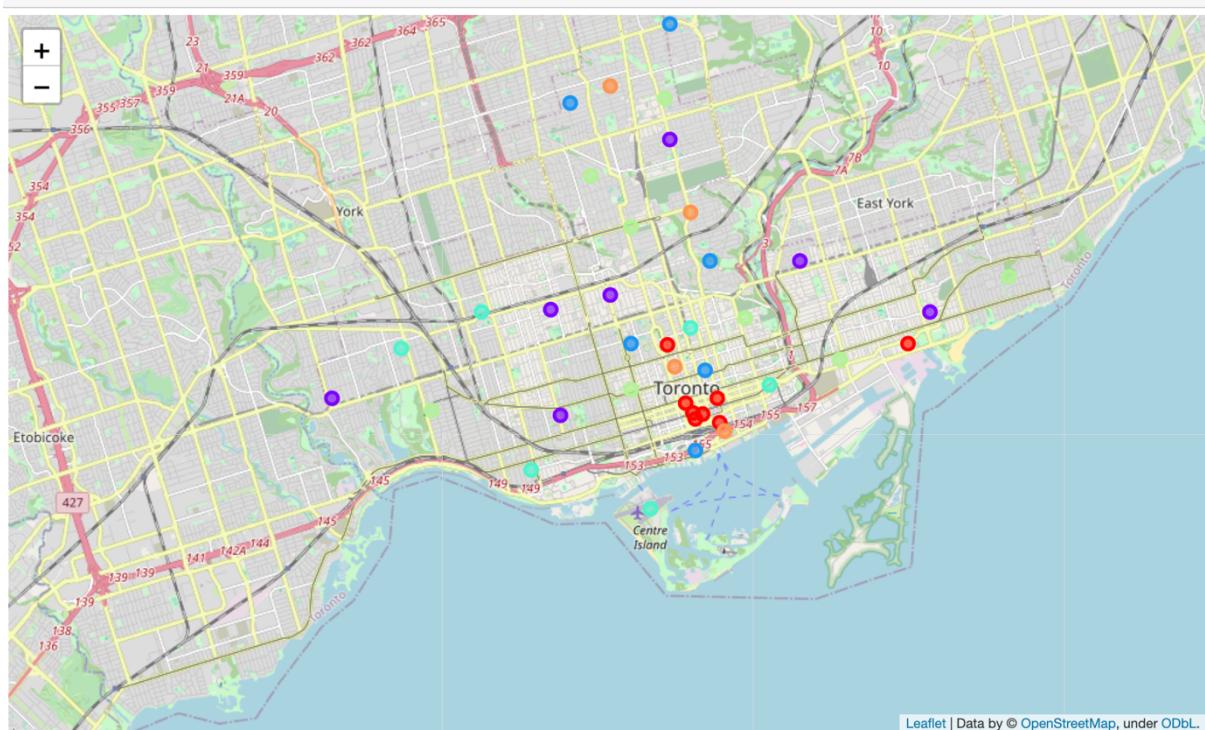
The best k was found out using the KElbowVisualizer from yellowbrick library



From here, we can see that $k = 6$

Clustering using $k = 6$

Now that the k is defined, I used k Means clustering to come up with 6 clusters. This was then merged to the intial dataframe to get latitude and longitude information which could then be plotted on a map. The different clusters formed are shown below:



Finally, we can examine the 6 different clusters.

Cluster Labels	Dwellings	Neighborhood	Indian Restaurant	PostalCode	Borough	Latitude	Longitude
0	0	1.0 Commerce Court, Victoria Hotel	0.0	M5L	Downtown Toronto	43.648198	-79.379817
1	0	1.0 Business reply mail Processing Centre	0.0	M7Y	East Toronto	43.662744	-79.321558
2	0	1.0 Queen's Park, Ontario Provincial Government	0.0	M7A	Downtown Toronto	43.662301	-79.389494
3	0	1.0 Stn A PO Boxes	0.0	M5W	Downtown Toronto	43.646435	-79.374846
4	0	2.0 Toronto Dominion Centre, Design Exchange	0.0	M5K	Downtown Toronto	43.647177	-79.381576
5	0	2.0 First Canadian Place, Underground city	0.0	M5X	Downtown Toronto	43.648429	-79.382280
6	0	1103.0 Richmond, Adelaide, King	0.0	M5H	Downtown Toronto	43.650571	-79.384568
7	0	1866.0 St. James Town	0.0	M5C	Downtown Toronto	43.651494	-79.375418

```
df_final.loc[df_final['Cluster Labels'] == 1]
```

Cluster Labels	Dwellings	Neighborhood	Indian Restaurant	PostalCode	Borough	Latitude	Longitude
26	1 13535.0	Little Portugal, Trinity	0.000000	M6J	West Toronto	43.647927	-79.419750
27	1 14039.0	Davisville	0.030303	M4S	Central Toronto	43.704324	-79.388790
28	1 14240.0	Runnymede, Swansea	0.000000	M6S	West Toronto	43.651571	-79.484450
29	1 14311.0	Christie	0.000000	M6G	Downtown Toronto	43.669542	-79.422564
30	1 14420.0	India Bazaar, The Beaches West	0.000000	M4L	East Toronto	43.668999	-79.315572
31	1 14907.0	The Annex, North Midtown, Yorkville	0.043478	M5R	Central Toronto	43.672710	-79.405678
32	1 15058.0	The Danforth West, Riverdale	0.023810	M4K	East Toronto	43.679557	-79.352188

```
df_final.loc[df_final['Cluster Labels'] == 2]
```

Cluster Labels	Dwellings	Neighborhood	Indian Restaurant	PostalCode	Borough	Latitude	Longitude
12	2 6377.0	Lawrence Park	0.00	M4N	Central Toronto	43.728020	-79.388790
13	2 6811.0	Roselawn	0.00	M5N	Central Toronto	43.711695	-79.416936
14	2 7137.0	Harbourfront East, Union Station, Toronto Isla...	0.01	M5J	Downtown Toronto	43.640816	-79.381752
15	2 7425.0	Garden District, Ryerson	0.00	M5B	Downtown Toronto	43.657162	-79.378937
16	2 7622.0	University of Toronto, Harbord	0.00	M5S	Downtown Toronto	43.662696	-79.400049
17	2 7645.0	Rosedale	0.00	M4W	Downtown Toronto	43.679563	-79.377529

```
df_final.loc[df_final['Cluster Labels'] == 3]
```

Cluster Labels	Dwellings	Neighborhood	Indian Restaurant	PostalCode	Borough	Latitude	Longitude
33	3	18779.0	Church and Wellesley	0.013514	M4Y	Downtown Toronto	43.665860 -79.383160
34	3	19322.0	High Park, The Junction South	0.000000	M6P	West Toronto	43.661608 -79.464763
35	3	19382.0	Regent Park, Harbourfront	0.000000	M5A	Downtown Toronto	43.654260 -79.360636
36	3	19383.0	Dufferin, Dovercourt Village	0.000000	M6H	West Toronto	43.669005 -79.442259
37	3	19536.0	Brockton, Parkdale Village, Exhibition Place	0.000000	M6K	West Toronto	43.636847 -79.428191
38	3	22424.0	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	M5V	Downtown Toronto	43.628947 -79.394420

```
df_final.loc[df_final['Cluster Labels'] == 4]
```

Cluster Labels	Dwellings	Neighborhood	Indian Restaurant	PostalCode	Borough	Latitude	Longitude
18	4	9149.0	Parkdale, Roncesvalles	0.000000	M6R	West Toronto	43.648960 -79.456325
19	4	9181.0	Forest Hill North & West	0.000000	M5P	Central Toronto	43.696948 -79.411307
20	4	9335.0	Kensington Market, Chinatown, Grange Park	0.000000	M5T	Downtown Toronto	43.653206 -79.400049
21	4	9969.0	Studio District	0.000000	M4M	East Toronto	43.659526 -79.340923
22	4	10096.0	Summerhill West, Rathnelly, South Hill, Forest...	0.000000	M4V	Central Toronto	43.686412 -79.400049
23	4	10689.0	St. James Town, Cabbagetown	0.020408	M4X	Downtown Toronto	43.667967 -79.367675
24	4	11346.0	The Beaches	0.000000	M4E	East Toronto	43.676357 -79.293031
25	4	12028.0	Davisville North	0.000000	M4P	Central Toronto	43.712751 -79.390197

```
df_final.loc[df_final['Cluster Labels'] == 5]
```

Cluster Labels	Dwellings	Neighborhood	Indian Restaurant	PostalCode	Borough	Latitude	Longitude
8	5	4280.0	Berczy Park	0.000000	M5E	Downtown Toronto	43.644771 -79.373306
9	5	4788.0	Central Bay Street	0.015873	M5G	Downtown Toronto	43.657952 -79.387383
10	5	5230.0	North Toronto West	0.000000	M4R	Central Toronto	43.715383 -79.405678
11	5	5293.0	Moore Park, Summerhill East	0.000000	M4T	Central Toronto	43.689574 -79.383160

Results

6 different clusters were formed

- Cluster 1 had low number of Indian dwellings and low number of Indian restaurants as well.
- Cluster 2 had relatively high Indian dwellings and a moderate number of competitors.
- Cluster 3 had relatively low to medium dwellings and low number of competitors.
- Cluster 4 had high number of dwellings and low number of competitors.
- Cluster 5 had medium dwellings and low competitors.
- Finally, Cluster 6 had low number of dwellings and low competitors.

Discussion

- Cluster 1 had low number of Indian dwellings and low number of Indian restaurants as well. Therefore, this cluster would be a **poor choice and shouldn't be considered**.
- Cluster 2 had relatively high Indian dwellings and a moderate number of competitors. So, this cluster could be a **possible choice**.
- Cluster 3 had relatively low to medium dwellings and low number of competitors. So, this cluster could be **considered; but not a good option though**.
- Cluster 4 had high number of dwellings and low number of competitors. So, this could be a **good option**.
- Cluster 5 had medium dwellings and low competitors. So, this cluster could be **considered**.
- Finally, Cluster 6 had low number of dwellings and low competitors. So, this cluster **shouldn't be considered**.

Additional parameters like income of the groups and population of other South Asian groups (Sri Lankan, Bangladeshi etc.) could increase the accuracy of the model. So, this can be considered as a future improvement to the model.

Conclusion

Based on the results from the k-Means clustering model, clusters 4,2 and 5 should be possible options for the client – AJ hospitality group. But, because of the fact that there are relatively low number of competitors in cluster 4, that could be the best option for AJ hospitality group to setup their operations in Toronto (Location shown below).

