Exploring Toronto neighbourhood for opening an Indian Restaurant

Abhijit Chatterjee

31-Mar-21

1. Introduction

1.1 Background
Toronto, It's one of the most popular Canadian cities among Indian and there is a thick population of Indian citizen observed in the city spread across different Borough. Indians usually have a completely different food habit as compared to other ethnic. Within Indian also, food habit, culture, language, tradition changes with different state of India and Indians are greatly diversified ethnic in the world. Indians usually prefer to have authentic Indian food across the world. Toronto is no exception in that where Indians often explore different venues and people enjoys mouth-watering traditional spicy Indian cuisine. This also gives Indian the feeling of home away from motherland. Also, Indian cuisines are becoming famous among other ethnic and nationalities due to rich taste of spices.

1.2 Problem Statement
In a city like Toronto, it is difficult to identify a suitable location where an entrepreneur can open an Indian Restaurant which will be able to attract and serve authentic traditional Indian food to maximum customer which in turn would help to expand the business and to make it profitable. Our Project is to explore Toronto neighbourhoods and recommend best suitable location to open an Indian Restaurant

1.3 Target audience
- ✓ The Investor/Businessman who wants to invest in the Indian Restaurant start-up and slowly expand their footprint in the segment to make it more profitable.

- ✓ The professionals in different fields who wants to own an Indian restaurant will be beneficial from this project by understanding different pros and cons, the competition in the field one can expect, analysing the expected customer footfall to generate profitability and to become a successful entrepreneur.

- ✓ The Indian crowd living in areas who wants to taste the authentic food of India but having limited option will be beneficial from this project.

2. Data Acquisition and Cleaning
2.1 Data Sources:

- ✓ **List of postal code and Neighborhood in Toronto Canada :**
  Data will be downloaded from wiki to get the complete list of postal codes in Toronto along with Borough and neighborhoods.
  https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- ✓ **Geospatial data :**
  In order to get the coordinates of different neighborhood downloaded above, geospatial data will be used.
  https://cocl.us/Geospatial_data

- ✓ **Ethnic population of Toronto:**
  Success of a ethnic food chain depends on the ethnic population residing in the area of food chain. Toronto Ethnic population data will be used to find out densely Indian populated neighborhood to use in analysis.
  https://en.m.wikipedia.org/wiki/Demographics_of_Toronto#Ethnic_diversity

- ✓ **Location data using Foursquare API:**
  To get different information, location about popular venues of Toronto, foursquare API will be used.

2.2 Data Cleaning:

- ✓ First, Toronto Postal codes data downloaded from wiki and converted into dataframe. The received in raw format where Postal codes, Borough, Neighbourhood information were clubbed together into single cell and not in tabular format. Based on the pattern of the data, it was converted into tabular format and loaded into dataframe.
  - o Postal code having Borough="Not Assigned" discarded
  - o Postal codes with Neighbourhood = "Not Assigned" populated with Borough data

Before: *Raw format*

	0	1	2
0	M1ANot assigned	M2ANot assigned	M3ANorth York(Parkwoods)
1	M1BScarborough(Malvern / Rouge)	M2BNot assigned	M3BNorth York(Don Mills)North

After: *Tabular Format*

	Postal Code	Borough	Neighborhood
1	M1B	Scarborough	Malvern,Rouge
2	M1C	Scarborough	Rouge Hill,Port Union,Highland Creek
3	M1E	Scarborough	Guildwood,Morningside,West Hill

- ✓ Geospatial data of latitude and longitude downloaded through API call and merged with above data frame

Location Data using API

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497

Merged Dataset

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern,Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill,Port Union,Highland Creek	43.784535	-79.160497

✓ Toronto population data by different ethnic downloaded from wiki and converted into data frame

	Riding	Population	Ethnic Origin #1	%	Ethnic Origin #2	%.1	Ethnic Origin #3	%.2	Ethnic Origin #4	%.3	Ethnic Origin #5	%.4	Ethnic Origin #6	%.5	Ethnic Origin #7	%.6	Ethnic Origin #8	%.7
0	Willowdale	117405	Chinese	25.9	Iranian	12.1	Korean	10.6	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Eglinton-Lawrence	112925	Canadian	14.7	English	12.6	Polish	12.0	Filipino	11.0	Scottish	9.7	Italian	9.5	Irish	9.2	Russian	8.4
2	Don Valley North	109060	Chinese	32.4	East Indian	7.3	Iranian	7.3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

✓ Location data retrieved using Foursquare API and converted into data frame

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern,Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge Hill,Port Union,Highland Creek	43.784535	-79.160497	Chris Effects Painting	43.784343	-79.163742	Construction & Landscaping

2.3 Feature Engineering and Feature Selection:

✓ Ethnic data had multiple Ethnic columns with different ethnic and corresponding percentage. Since we are interested in Indian Population, all the ethnic columns were iterated for East Indian ethnic and corresponding percentage and merged into single dataframe. Rest of the ethnic (Non Indian) data were discarded

Before

	Riding	Population	Ethnic Origin #1	%	Ethnic Origin #2	%.1	Ethnic Origin #3	%.2	Ethnic Origin #4	%.3	Ethnic Origin #5	%.4	Ethnic Origin #6	%.5	Ethnic Origin #7	%.6	Ethnic Origin #8	%.7
0	Scarborough Centre	110450	Filipino	13.1	East Indian	12.2	Canadian	11.2	Chinese	10.7	English	7.8	Sri Lankan	7.0	NaN	NaN	NaN	NaN
1	Scarborough Southwest	108295	Canadian	16.2	English	14.3	Irish	11.5	Scottish	10.9	Filipino	9.5	East Indian	8.2	Chinese	7.2	NaN	NaN
2	Scarborough-Agincourt	104225	Chinese	47.0	East Indian	7.4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Scarborough-Rouge Park	101445	East Indian	16.7	Canadian	11.8	Sri Lankan	11.1	English	9.8	Filipino	9.3	Jamaican	8.4	Scottish	7.2	Irish	7.0
4	Scarborough-Guildwood	101115	East Indian	18.0	Canadian	11.6	English	9.7	Filipino	8.5	Sri Lankan	7.8	Chinese	7.1	Scottish	7.0	NaN	NaN

After

	Riding	Population	Ethnic	Percentage
0	Don Valley East	93170	East Indian	10.6
1	Don Valley North	109060	East Indian	7.3
2	Humber River-Black Creek	107725	East Indian	9.2
3	Scarborough-Rouge Park	101445	East Indian	16.7

✓ Foursquare data retrieved was processed to keep only venue category = "Indian Restaurant" by Neighbourhood latitude, longitude.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
14	Woburn	43.770992	-79.216917	Al-Hamd Biryani & Pizza	43.767585	-79.219570	Indian Restaurant
50	Dorset Park,Wexford Heights,Scarborough Town C...	43.757410	-79.273304	Kairali	43.754915	-79.276945	Indian Restaurant
52	Dorset Park,Wexford Heights,Scarborough Town C...	43.757410	-79.273304	Karaikudi Chettinad South Indian Restaurant	43.756042	-79.276276	Indian Restaurant
344	Thorncliffe Park	43.705369	-79.349372	Iqbal Kebab & Sweet Centre	43.705923	-79.351521	Indian Restaurant

3. Exploratory Data Analysis

3.1 Statistical mean of Indian Restaurant by Neighbourhood and graphical View

✓ Indian restaurant data retrieved was transformed using Pandas one hot encoding and then calculated mean by neighbourhoods.
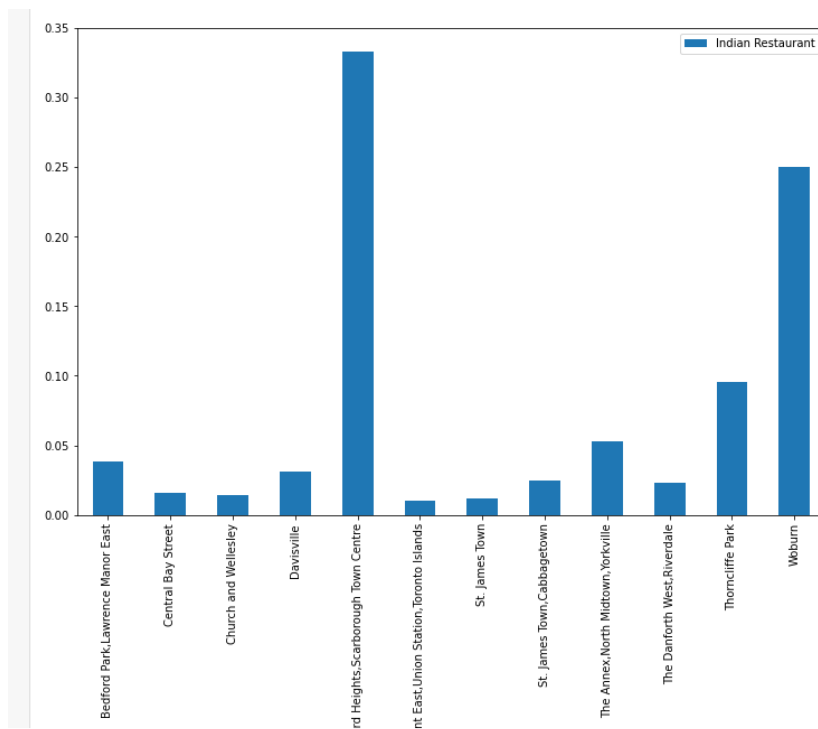
Code Snippet

```
Toronto_venues_onhot=pd.get_dummies(data=Toronto_venues[['Venue Category']],prefix='',prefix_sep='')

Toronto_venues_onhot['Neighborhood']=Toronto_venues['Neighborhood']
col=[Toronto_venues_onhot.columns[-1]]+list(Toronto_venues_onhot.columns[:-1])
Toronto_venues_onhot=Toronto_venues_onhot[col]
Toronto_group=Toronto_venues_onhot.groupby('Neighborhood').mean().reset_index()
Toronto_Indian_Restaurant=Toronto_group[['Neighborhood','Indian Restaurant']]
Toronto_Indian_Restaurant
```

Transformed Data

	Neighborhood	Indian Restaurant
0	Agincourt	0.000000
1	Alderwood,Long Branch	0.000000
2	Bathurst Manor,Wilson Heights,Downsview North	0.000000
3	Bayview Village	0.000000
4	Bedford Park,Lawrence Manor East	0.038462

✓ Plotted Bar graph of Indian Restaurant population by neighbourhoods:

3.2 Indian Population by Riding and graphical view

✓ Extracting Indian Ethnic population by Riding/Neighbourhood and combining into single data frame

Code Snippet

```python
Toronto_Indian_Ethnic=pd.DataFrame(columns=['Riding','Population','Ethnic','Percentage'])
for name in [df_pop_tor_eyrk,df_pop_nyrk,df_pop_scrb,df_pop_eto]:
    for i in range (8):
        col1='Ethnic Origin '+str(i+1)
        col2='Ethnic Origin '+str(i+1)+" Pct"
        dict={col1:"Ethnic",col2:"Percentage"}
        tmp_df=name.loc[name[col1]=='East Indian',['Riding','Population',col1,col2]]
        tmp_df.rename(columns=dict,inplace=True)
        #print("tmp df is:", tmp_df.iloc[:,0:4])
        Toronto_Indian_Ethnic=Toronto_Indian_Ethnic.append(tmp_df,ignore_index=True)
tmp_df=df_pop_tor_eyrk.loc[df_pop_tor_eyrk['Ethnic Origin 9']=='East Indian',['Riding','Population','Ethnic Origin 9','Ethnic Origin 9 Pct']]
tmp_df.rename(columns={"Ethnic Origin 9":"Ethnic","Ethnic Origin 9 Pct":"Percentage"},inplace=True)
Toronto_Indian_Ethnic=Toronto_Indian_Ethnic.append(tmp_df,ignore_index=True)
Toronto_Indian_Ethnic['Indian Population']=Toronto_Indian_Ethnic['Population']*Toronto_Indian_Ethnic['Percentage']/100
Toronto_Indian_Ethnic
```

Output

	Riding	Population	Ethnic	Percentage	Indian Population
0	Don Valley East	93170	East Indian	10.6	9876.02
1	Don Valley North	109060	East Indian	7.3	7961.38
2	Humber River-Black Creek	107725	East Indian	9.2	9910.7
3	Scarborough-Rouge Park	101445	East Indian	16.7	16941.3
4	Scarborough-Guildwood	101115	East Indian	18.0	18200.7
5	Scarborough Centre	110450	East Indian	12.2	13474.9
6	Scarborough-Agincourt	104225	East Indian	7.4	7712.65
7	Scarborough North	97610	East Indian	11.8	11518
8	Scarborough Southwest	108295	East Indian	8.2	8880.19
9	Etobicoke North	116960	East Indian	22.2	25965.1

3.3 Relationship between Indian Population and Indian Restaurant by Neighbourhoods/Riding

✓ Merged two data set of Indian Restaurant and Indian Ethnic population by Neighbourhood/Riding. No direct relationship visible as the Riding and Neighbourhood are not exactly same. However, for Indian population data, by riding names, most of the ridings are it quite identifiable which Borough it belongs to. Indian Restaurant data also populated by neighbourhood/Borough.

✓ With that understanding, it looks, Scarborough Borough has major Indian crowd as well as densely populated Indian restaurant

	Cluster Label	Neighborhood	Indian Restaurant	Riding	Indian Population
16	0.0	Cliffside,Cliffcrest,Scarborough Village West	0.000000	NaN	NaN
23	1.0	Dorset Park,Wexford Heights,Scarborough Town C...	0.333333	NaN	NaN
73	0.0	Scarborough Village	0.000000	NaN	NaN
102	NaN	NaN	NaN	Scarborough-Rouge Park	16941.3
103	NaN	NaN	NaN	Scarborough-Guildwood	18200.7
104	NaN	NaN	NaN	Scarborough Centre	13474.9
105	NaN	NaN	NaN	Scarborough-Agincourt	7712.65
106	NaN	NaN	NaN	Scarborough North	11518
107	NaN	NaN	NaN	Scarborough Southwest	8880.19

4. Classification Model
## 4.1 Finding Best K for KMeans Clustering using Elbow method

```python
In [36]: from sklearn.cluster import KMeans

Toronto_Clust=Toronto_Indian_Restaurant.drop('Neighborhood',1

cost=[]
for i in range(2,13):
    KM = KMeans(n_clusters=i, max_iter=100)
    KM.fit(Toronto_Clust)
    cost.append(KM.inertia_/100)
```

Plot K Values Against Cost

```python
In [37]: plt.plot (range(2,13),cost, color='r', linewidth='4')
plt.xlabel('K values')
plt.ylabel('Cost')
plt.show()
```
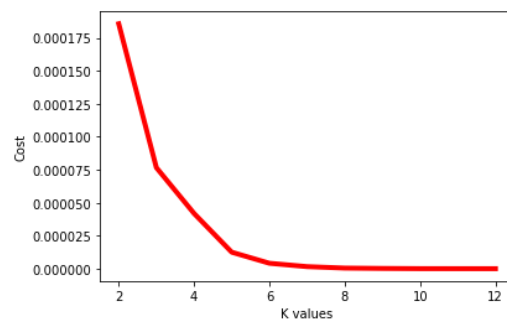

✓ From Elbow method it is evident , at k=6, the graph started flattening. So I took K=6 for Clustering.
✓

```python
k_clust=6
km=KMeans(n_clusters=k_clust, random_state=0).fit(Toronto_Clust)
print(km.labels_)

Toronto_Indian_Restaurant.insert(0,'Cluster Label',km.labels_)
Toronto_Indian_Restaurant.head()
```

```
[0 0 0 0 4 0 0 0 0 0 0 5 0 5 0 0 0 0 4 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 5 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 5 5 0 0 0 4 0 0 5 0 2 0 0 0 0 0 0 0 0 3 0 0 0]
```

8]:

	Cluster Label	Neighborhood	Indian Restaurant
0	0	Agincourt	0.000000
1	0	Alderwood,Long Branch	0.000000
2	0	Bathurst Manor,Wilson Heights,Downsview North	0.000000
3	0	Bayview Village	0.000000
4	4	Bedford Park,Lawrence Manor East	0.038462

✓ The Indian Restaurant dataset was assigned to different cluster.

4.2 Plotting cluster in Map

✓ With the help of Folium package and geocoder API, The clusters were plotted in Map.
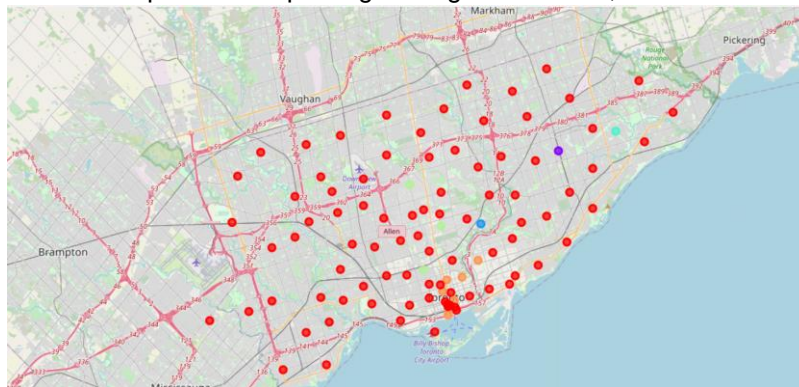
4.3 Evaluate the Cluster

✓ We have total 6 clusters. These are 0,1,2,3,4,5,6
✓ Custer 0 : This cluster contains the neighbourhood having the least number of Indian Restaurant

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Label	Indian Restaurant
0	M1B	Scarborough	Malvern,Rouge	43.806686	-79.194353	0.0	0.0
1	M1C	Scarborough	Rouge Hill,Port Union,Highland Creek	43.784535	-79.160497	0.0	0.0
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711	0.0	0.0
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	0.0	0.0
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476	0.0	0.0

✓ Custer 1 : This cluster contains the neighbourhood having the moderate to High number of Indian Restaurant

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Label	Indian Restaurant
10	M1P	Scarborough	Dorset Park,Wexford Heights,Scarborough Town C...	43.75741	-79.273304	1.0	0.333333

✓ Custer 2 : This cluster contains the neighbourhood having the low number of Indian Restaurant

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Label	Indian Restaurant
37	M4H	East York	Thorncliffe Park	43.705369	-79.349372	2.0	0.095238

✓ Custer 3 : This cluster contains the neighborhood having the moderate to low number of Indian Restaurant

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Label	Indian Restaurant
3	M1G	Scarborough	Woburn	43.770992	-79.216917	3.0	0.25

✓ Custer 4 : This cluster contains the neighborhood having the very low number of Indian Restaurant

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Label	Indian Restaurant
45	M4S	Central Toronto	Davisville	43.704324	-79.388790	4.0	0.031250
59	M5M	North York	Bedford Park,Lawrence Manor East	43.733283	-79.419750	4.0	0.038462
62	M5R	Central Toronto	The Annex,North Midtown,Yorkville	43.672710	-79.405678	4.0	0.052632

- ✓ Custer 5 : This cluster contains the neighborhood having very low number of Indian Restaurant

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Label	Indian Restaurant
39	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	5.0	0.023256
48	M4X	Downtown Toronto	St. James Town,Cabbagetown	43.667967	-79.367675	5.0	0.024390
49	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160	5.0	0.014286
52	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	5.0	0.012195
54	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	5.0	0.015625
56	M5J	Downtown Toronto	Harbourfront East,Union Station,Toronto Islands	43.640816	-79.381752	5.0	0.010000

- ✓ Custer 6 : This cluster contains the neighborhood having the moderate number of Indian
- ✓ Alal
- ✓ a

```
Toronto_Final[Toronto_Final['Cluster Label']==6]
```

2]:

Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Label	Indian Restaurant

- ✓ Overall Indian Restaurant density by Borough and Cluster

Borough	Cluster Label	Latitude	Longitude	Indian Restaurant
Central Toronto	0.0	262.251210	-476.412958	0.000000
	4.0	87.377034	-158.794468	0.083882
Downtown Toronto	0.0	523.844521	-952.635732	0.000000
	5.0	218.284089	-396.895388	0.076496
Downtown TorontoStn A PO Boxes25 The Esplanade	0.0	43.646435	-79.374846	0.000000
East Toronto	0.0	131.004881	-237.949526	0.000000
	5.0	43.679557	-79.352188	0.023256
East TorontoBusiness reply mail Processing Centre969 Eastern	0.0	43.662744	-79.321558	0.000000
East York	0.0	131.110802	-237.991777	0.000000
	2.0	43.705369	-79.349372	0.095238
East YorkEast Toronto	0.0	43.685347	-79.338106	0.000000
Etobicoke	0.0	436.545912	-795.378586	0.000000
EtobicokeNorthwest	0.0	43.706748	-79.594054	0.000000
MississaugaCanada Post Gateway Processing Centre	0.0	43.636966	-79.615819	0.000000
North York	0.0	962.495123	-1747.475877	0.000000
	4.0	43.733283	-79.419750	0.038462
Queen's Park	0.0	43.662301	-79.389494	0.000000
Scarborough	0.0	612.661365	-1109.538592	0.000000
	1.0	43.757410	-79.273304	0.333333
	3.0	43.770992	-79.216917	0.250000
West Toronto	0.0	261.915918	-476.695739	0.000000
York	0.0	218.453984	-397.363167	0.000000

5. Discussion

- ✓ The Boroughs like Central Toronto, Downtown Toronto,East Toronto,East York, North York, Scarborough are sparsly or densely populated with Indian Restaurant.
- ✓ From the Indian population bar chart, it is also clearly visible that scarborogh is also modertely populated with Indians.
- ✓ So opening a restaurant in Scarborough might attract good foot fall but competetion will be high
- ✓ Etobicocke is one borough which have most densly indian population.
- ✓ Also there is no Indian Restaurant in Etobicocke borough
- ✓ Other boroughs like West Toronto, York, Queen's Park, Etobicoke, MississaugaCanada Post Gateway Processing Centre neither have Indian population, nor have Indian Restaurant

6. Conclusion

With Above analysis, it is recommended to leave Scarborough to avoid Higher competition which would impact the lesser profitability. <u>Etobicoke is the Borough which should be considered to open Indian Restaurant due to heavy Indian Population and Limited to no competition.</u>