Finding the best place in Toronto to open an Indian restaurant

Amit Kumar

April 13, 2019

1. Introduction

1.1 Background

Indian cuisine consists of a wide variety of regional and traditional cuisines native to the Indian subcontinent. Given the range of diversity in soil type, climate, culture, ethnic groups, and occupations, these cuisines vary substantially from each other and use locally available spices, herbs, vegetables, and fruits. Indian migration has spread the traditions of the subcontinent throughout the world. These cuisines have been adapted to local tastes. Curry's international appeal has been compared to that of pizza. Indian tandoor dishes such as chicken tikka enjoy widespread popularity.

Indian cuisines are very popular around the world and it is gaining its popularity rapidly. The demand of these cuisines is growing in countries like United Kingdome, Middle East, Southeast Asia, Canada and United States. A survey by The Washington Post in 2007 stated that more than 1,200 Indian food products had been introduced into the United States since 2000. There are numerous Indian restaurants across the US, which vary based on regional culture and climate. North Indian and South Indian cuisines are especially well represented. Most Indian restaurants in the United States serve Americanized versions of North Indian food, which is generally less spicy than its Indian equivalents.

Indian cuisine is widely available in Canada, especially in the cities of Toronto, Vancouver, and Ottawa where the majority of Canadians of South Asian heritage live. Beginning in the 1960s and 1970s, emigration from South Asia increased and the first groceries and restaurants run by Sikhs and other Indians began to appear in Canada. As communities blossomed in Vancouver and Toronto in particular, grocers and other food establishments rushed to fill the gap. Although Indian restaurants generally retained a heavy emphasis on North Indian cuisine due to the preponderance of Punjabi immigrants, waves of Indians from other parts of the diaspora began to diversify the offerings, from the Northern-influenced traditions of Indo-Caribbean immigrants to the Southern-influenced traits of East African Indians. The result is the highly diverse contemporary offerings at present.

1.2 Problem

Finding the right market and targeted customers is very important to start a new business or business at new place and could be a challenging task if sufficient data is not available. Fortunately, the required data is available on different platforms. We will analyze these data and find the neighborhood in the Toronto which will have the most common Indian restaurants.

1.3 Interest

People who are interested in opening a new Indian restaurants would be very interested to know about the best place to open their restaurants. Food companies like KFC, McDonald's, Starbucks etc. would be interesting in knowing about Indian cuisine markets to include Indian cuisines in their existing menu to increase their market size and profit. Also the people who wants to taste Indian food would be interested in knowing the most common place where they can get good Indian food

2. Data acquisition and cleaning

2.1 Data Source

First we need the data of Toronto neighborhoods which can be found on https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. The extracted data contains the columns namely PostalCode, Borough and Neighborhood, where PostalCode column contain zip code of all the neighborhoods, Borough column contains the borough name to which a particular neighborhood belongs and Neighborhood column contains name of all the neighborhoods in the Toronto.

	PostalCode	Borough	Neighborhood
1	M1A	Not assigned	Not assigned
2	M2A	Not assigned	Not assigned
3	M3A	North York	Parkwoods
4	M4A	North York	Victoria Village
5	M5A	Downtown Toronto	Harbourfront

Fig.1 Extracted data from Wikipedia

We are using **Foursquare** location data To extract the detailed data about restaurants in each neighborhood, but to get detailed information from Foursquare we need to provide coordinates longitude and latitude of each neighborhood and this data can be extracted from https://cocl.us/Geospatial_data.

	PostalCode	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

Fig.2 Neighborhood coordinates

2.2 Data Cleaning

The data extracted from the Wikipedia page contains 'Not assigned' boroughs and neighborhood. All the rows which contains not assigned borough will be dropped from the dataset. In case the neighborhoods are having not assigned value but having assigned borough, the borough will be assigned to the corresponding neighborhood.

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M5A	Downtown Toronto	Regent Park
4	M6A	North York	Lawrence Heights

Fig.3 Data after dropping not assigned boroughs

2.3 Data Preprocessing

After the data is cleaned we will preprocess it before feeding it to the machine learning model. First we will combine the data from Wikipedia and coordinates in to a single dataset.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Fig.4 Combined data

2.3.1 Extracting information about neighborhoods using Foursquare

Now we will feed the combined data to the Foursquare to extract the detailed information about each neighborhood. The figure just shows snapshot not an entire details.

	PostalCode	Borough	Neighborhood	Borough Latitude	Borough Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
0	M4E	East Toronto	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
1	M4E	East Toronto	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
2	M4E	East Toronto	The Beaches	43.676357	-79.293031	Starbucks	43.678798	-79.298045	Coffee Shop
3	M4E	East Toronto	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant

Fig.5 Foursquare location data

3. Methodology

3.1 Exploratory Data Analysis

The data of venues in each neighborhood were extracted using foursquare. In exploring the neighborhood we limited the maximum number of venues to 100 and searched in the radius of 500m only which resulted in total of 2245 venues and 275 unique venues.

We plotted the bar graph of the number of venues present in each neighborhood in 500m radius.

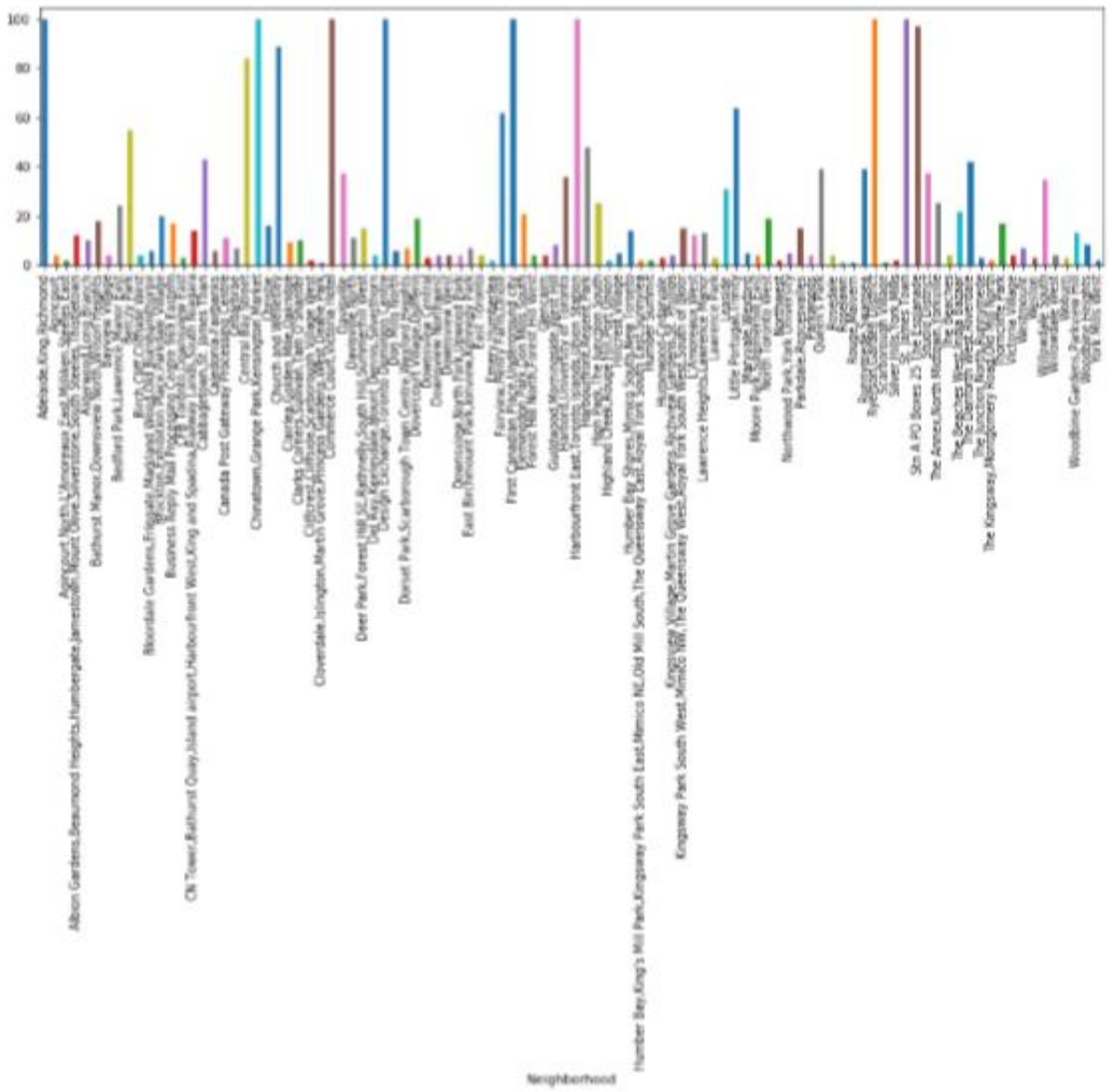

Fig.6 Bar graph of No of Venues per neighborhood

3.2 Machine Learning Technique used

We used KMeans clustering machine learning technique to group the neighborhoods into clusters based on their mean of frequency of occurrence.

```
cntr=np.array([0.2, 0.03, 0]).reshape(-1,1)# initilization of center coordinates
X=df_ind_rest.values.reshape(-1, 1)
kmeans=KMeans(init = cntr,n_clusters = 3, n_init = 12)
kmeans.fit(X)
```

```
C:\Users\mahe\Anaconda3\lib\site-packages\sklearn\cluster\k_means_.py:971: RuntimeWarning: Explicit i
nitial center position passed: performing only one init in k-means instead of n_init=12
  return_n_iter=True)
```

```
KMeans(algorithm='auto', copy_x=True,
    init=array([[0.2 ],
       [0.03],
       [0.  ]]), max_iter=300,
    n_clusters=3, n_init=12, n_jobs=None, precompute_distances='auto',
    random_state=None, tol=0.0001, verbose=0)
```

```
# getting labels for each neighborhood
kmeans_labels = kmeans.labels_
kmeans_labels
```

```
array([0, 0, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2])
```

```
# getting center coordinates each cluster
cl=kmeans.cluster_centers_
cl
```

```
array([[0.20168067],
       [0.02992809],
       [0.00044822]])
```

4. Results

The data is divided into three clusters with centers at 0.20168067, 0.02992809 and 0.00044822 top two neighborhoods are grouped into cluster label 0, next top 6 are grouped into cluster label 1 and rest are grouped into cluster label 2 as show in Fig.7. In this study it is found that Dorset Park,Scarborough Town Centre,Wexfordare are the neighborhoods where mean frequency is highest 0.285714 followed by Thorncliffe Park on the second number 0.117647.

34	0.285714
90	0.117647
7	0.041667
84	0.040000
28	0.027027
87	0.023810
19	0.023810
15	0.023256
22	0.011236
81	0.010000
0	0.010000
50	0.010000
38	0.000000
32	0.000000
27	0.000000
44	0.000000
29	0.000000
43	0.000000
30	0.000000
31	0.000000

Fig.7 mean of frequency of occurrence of Indian restaurant

5. Discussion & Conclusion

In this study it is found that neighborhoods in the cluster label 0 which contains Dorset Park, Scarborough Town Centre, Wexfordareand Thorncliffe Park have highest values of frequency of occurrence of Indian restaurant, this means the Indian restaurants are more popular in these neighborhoods than others.

According to the findings of this study it is advisable to open the Indian restaurants in the neighborhoods which belongs to the cluster label 0.