

Finding the best place in Toronto to open an Indian restaurant

The Battle of Neighborhoods

Problem Statement

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, for this study 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.

Who will be interested about this project

1. People who are interested in opening a new Indian restaurants.
2. Food companies like KFC, McDonald's, Starbucks etc.
3. The people who wants to taste Indian food

Data source

1. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2. https://cocl.us/Geospatial_data
3. Foursquare location Data.

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.

Data cleaning Contd...

	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

	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

Extracting information about neighborhoods using Foursquare

```
radius = 500
LIMIT = 100
CLIENT_ID='Removed'
CLIENT_SECRET='Removed'
VERSION='Removed'
venues = []
for lat, long, post, borough, neighborhood in zip(df5['Latitude'], df5['Longitude'], df5['PostalCode'], df5['Borough'], df5['Nei
    url = "https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}".format(
        CLIENT_ID,
        CLIENT_SECRET,
        VERSION,
        lat,
        long,
        radius,
        LIMIT)
    results = requests.get(url).json()
    results=results["response"]["groups"][0]['items']
    for venue in results:
        venues.append((
            post,
            borough,
            neighborhood,
            lat,
            venue['venue']['name'],
            venue['venue']['location']['lat'],
            venue['venue']['location']['lng'],
            venue['venue']['categories'][0]['name']))
```

Extracting information about neighborhoods using Foursquare contd...

Extracted Foursquare location data

	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

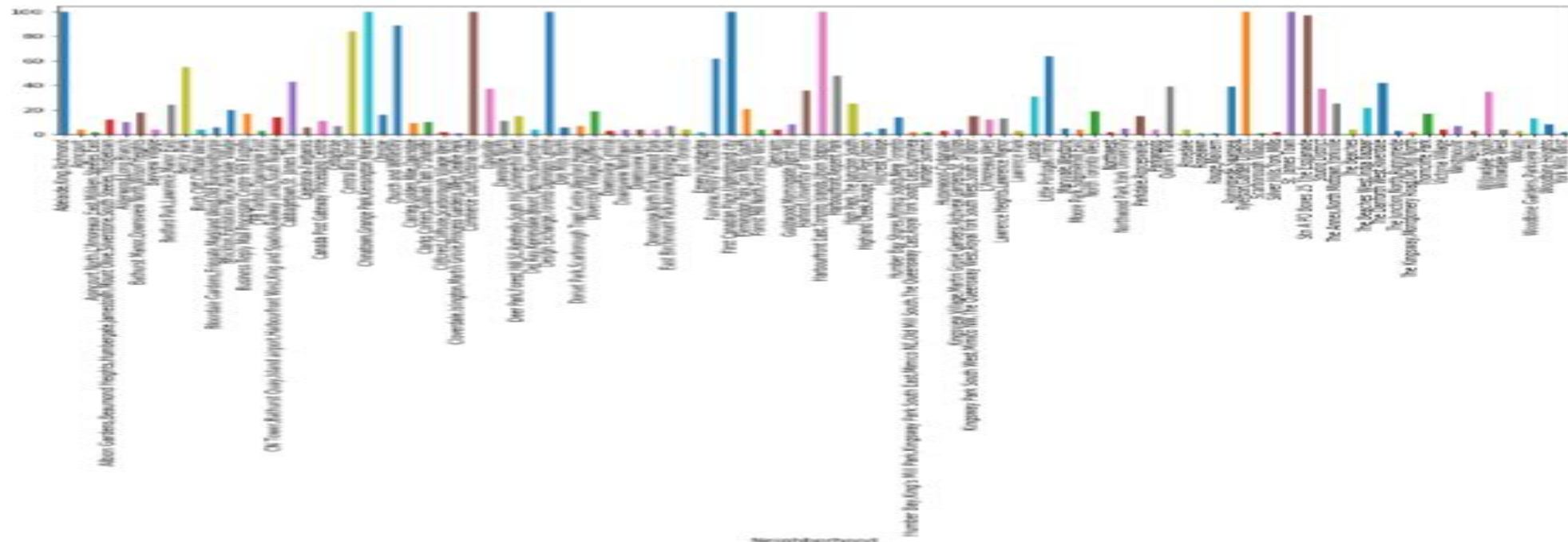
Methodology

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.

Methodology Contd...

Bar graph of No of Venues per neighborhood



Methodology Contd...

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

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, Wexford are the neighborhoods where mean frequency is highest 0.285714 followed by Thorncliffe Park on the second number 0.117647

Results Contd...

The mean of frequency of occurrence of Indian restaurant in each neighborhood in descending order

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

Discussion & Conclusion

- In this study it is found that neighborhoods in the cluster label 0 which contains Dorset Park, Scarborough Town Centre, Wexford and 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.