# Exploring the Neighbourhoods in Toronto, Canada.

Siddharaj Maramwar

## Introduction

As a part of the IBM Data Science professional program Capstone Project, I worked on the real datasets to get an experience of what a data scientist goes through in real life. Main objectives of this project were to define a business problem, look for data in the web and, use Foursquare location data to compare different neighborhoods of Toronto to figure out which neighborhood is suitable for starting a new restaurant business. In this project we will follow step by step methods to get the results.

## Problem Description

Consider a situation where a person who wants to open a new Indian restaurant. And the person is Indo-Canadian and lives in the most populated city of Canada Toronto. So he has some doubts in his mind whether it is a good idea to open a restaurant. And if it is good idea in which Neighbourhood he should open his new restaurant, such that it should be profitable for him.

## Benefits

There are different people who will have benefit of this project.

- Business Person who wants to open a new restaurant in the neighbourhood.
- Indian people who wants to move to the neighbourhoods which has ample Indian restaurants and culture.
- Data Analyst / Data Scientist who analyse the neighbourhood using statistical and Exploratory Data analysis.

# Data acquisition

There are different sources from which I will collect the data for different purpose.

1. List of postal Codes for Canada :-
    - I will get the postal code of the neighbourhoods in Canada from Wikipedia.
    - Link - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

2. Geographical Co-ordinates :-
    - I will use a csv file which consist latitude and longitude of the neighbourhoods in Canada.
    - We can use geocoder for the same purpose, but it is not persistent sometime. So I choose the csv file instead using geocoder.
    - Link for csv - http://cocl.us/Geospatial_data

3. Fetching Details of the venue :
    - I will use Foursquare API for fetching the details and location of the venues.
    - I will use venue ratings as a threshold. And finally visualize using Folium.

From Foursquare API (https://developer.foursquare.com/docs),

I retrieved the following for each venue:

a) Name: The name of the venue.
b) Category: The category type as defined by the API.
c) Latitude: The latitude value of the venue.
d) Longitude: The longitude value of the venue.
e) Likes: Likes of the venue, that the user liked the restaurant.
f) Rating: Rating of the venue.
g) Tips: Tips given by the users.

# Data Cleaning

## Cleaning the Postal Code data

- The dataframe will consist of three columns: PostalCode, Borough, and Neighborhood
- Only process the cells that have an assigned borough. Ignore cells with a borough that is **Not assigned.**
- More than one neighborhood can exist in one postal code area. For example, in the table on the Wikipedia page, you will notice that **M5A** is listed twice and has two neighborhoods: **Harbourfront** and **Regent Park**. These two rows will be combined into one row with the neighborhoods separated with a comma as shown in **row 11** in the above table.
- If a cell has a borough but a **Not assigned** neighborhood, then the neighborhood will be the same as the borough.

| | PostalCode | Borough | Neighborhood |
|---|---|---|---|
| 0 | M5G | Downtown Toronto | Central Bay Street |
| 1 | M2H | North York | Hillcrest Village |
| 2 | M4B | East York | Parkview Hill, Woodbine Gardens |
| 3 | M1J | Scarborough | Scarborough Village |
| 4 | M4G | East York | Leaside |
| 5 | M4M | East Toronto | Studio District |
| 6 | M1R | Scarborough | Wexford, Maryvale |
| 7 | M9V | Etobicoke | South Steeles, Silverstone, Humbergate, Jamest... |
| 8 | M9L | North York | Humber Summit |
| 9 | M5V | Downtown Toronto | CN Tower, King and Spadina, Railway Lands, Har... |
| 10 | M1B | Scarborough | Malvern, Rouge |
| 11 | M5A | Downtown Toronto | Regent Park, Harbourfront |

*Data Frame of Postal Codes in Canada.*

## Adding Geographical Co-ordinates

For this I will use a csv file which consist latitude and longitude of the neighbourhoods in Canada.

Link for csv - http://cocl.us/Geospatial_data

| | Postalcode | 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 |
| 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 |
| ... | ... | ... | ... | ... | ... |
| 98 | M9N | York | Weston | 43.706876 | -79.518188 |
| 99 | M9P | Etobicoke | Westmount | 43.696319 | -79.532242 |
| 100 | M9R | Etobicoke | Kingsview Village, St. Phillips, Martin Grove ... | 43.688905 | -79.554724 |

*Data Frame ~ Neighborhoods with Co-ordinates.*

Now we will work only with boroughs that contain Toronto

| Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|
| East Toronto | The Beaches | 43.676357 | -79.293031 |
| East Toronto | The Danforth West, Riverdale | 43.679557 | -79.352188 |
| East Toronto | India Bazaar, The Beaches West | 43.668999 | -79.315572 |
| East Toronto | Studio District | 43.659526 | -79.340923 |
| Central Toronto | Lawrence Park | 43.728020 | -79.388790 |
| Central Toronto | Davisville North | 43.712751 | -79.390197 |
| Central Toronto | North Toronto West, Lawrence Park | 43.715383 | -79.405678 |
| Central Toronto | Davisville | 43.704324 | -79.388790 |
| Central Toronto | Moore Park, Summerhill East | 43.689574 | -79.383160 |
| Central Toronto | Summerhill West, Rathnelly, South Hill, Forest... | 43.686412 | -79.400049 |
| Downtown Toronto | Rosedale | 43.679563 | -79.377529 |

*Data Frame ~ Consists Boroughs only containing Toronto.*

# Indian Restaurants in the Toronto

Now fetch all the Indian restaurants in the Toronto

```python
# prepare neighborhood list that contains indian resturants
column_names=['Borough', 'Neighborhood', 'ID','Name']
indian_rest_to=pd.DataFrame(columns=column_names)
count=1
for row in toronto_data.values.tolist():
    Borough, Neighborhood, Latitude, Longitude=row
    venues = get_venues(Latitude,Longitude)
    indian_resturants=venues[venues['Category']=='Indian Restaurant']
    print('(',count,'/',len(toronto_data),')','Indian Resturants in '+Neighborhood+', '+Borough+':'+str(len(indian_resturants)))
    for resturant_detail in indian_resturants.values.tolist():
        id, name , category=resturant_detail
        indian_rest_to = indian_rest_to.append({'Borough': Borough,
                                                 'Neighborhood': Neighborhood,
                                                 'ID': id,
                                                 'Name' : name
                                                }, ignore_index=True)
    count+=1
```

*Code Snippet*

```
: indian_rest_to.head()
```

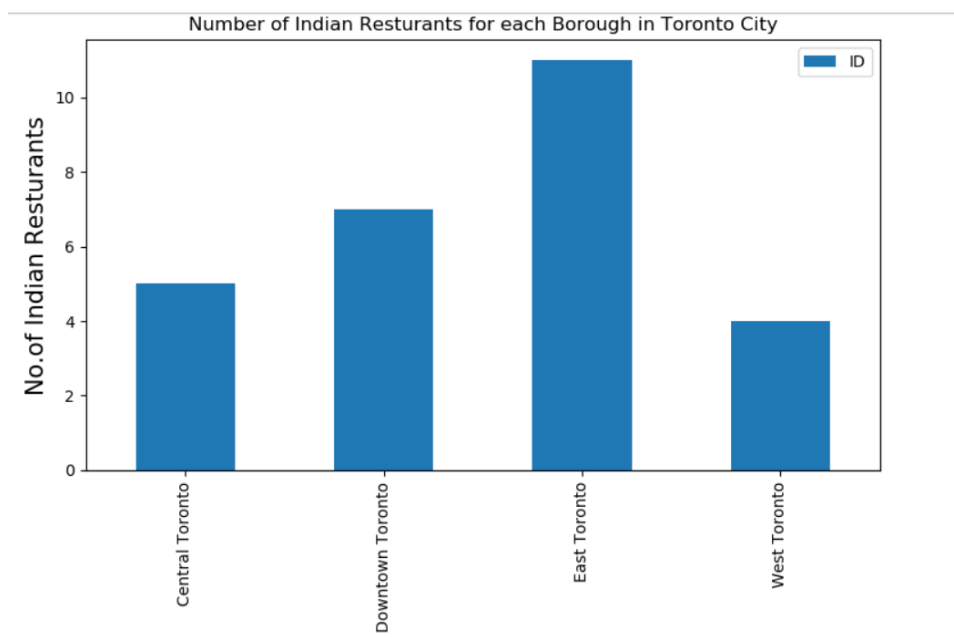| | Borough | Neighborhood | ID | Name |
|---|---|---|---|---|
| 0 | East Toronto | The Beaches | 4dcd7c6352b1f8915b7e7f7e | Delhi Bistro |
| 1 | East Toronto | The Danforth West, Riverdale | 4c1d5337eac020a1cb1048c2 | Sher-E-Punjab |
| 2 | East Toronto | India Bazaar, The Beaches West | 4ae0c7a8f964a520638221e3 | Udupi Palace |
| 3 | East Toronto | India Bazaar, The Beaches West | 4afc9816f964a520312422e3 | Motimahal |
| 4 | East Toronto | India Bazaar, The Beaches West | 4bac30a2f964a52018ea3ae3 | Bombay Chowpatty |

*Data Frame ~ Indian Restaurants in Toronto.*

# Exploratory Analysis

Let's analyse how many Indian restaurants are present in each Borough.

```python
plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Number of Indian Resturants for each Borough in Toronto City')
#On x-axis
plt.xlabel('Borough', fontsize = 15)
#On y-axis
plt.ylabel('No.of Indian Resturants', fontsize=15)
#giving a bar plot
indian_rest_to.groupby('Borough')['ID'].count().plot(kind='bar')
#legend
plt.legend()
#displays the plot
plt.show()
```

*Code Snippet*



*fig ~ Number of Indian Restaurants in each borough in Toronto.*

Let's also analyse how many Indian restaurants are present in each Neighborhood.

```
plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Number of Indian Resturants for each Neighborhood in Toronto City')
#On x-axis
plt.xlabel('Neighborhood', fontsize = 15)
#On y-axis
plt.ylabel('No.of Indian Resturants', fontsize=15)
#giving a bar plot
indian_rest_to.groupby('Neighborhood')['ID'].count().nlargest(5).plot(kind='bar')
#legend
plt.legend()
#displays the plot
plt.show()
```

*Code Snippet*
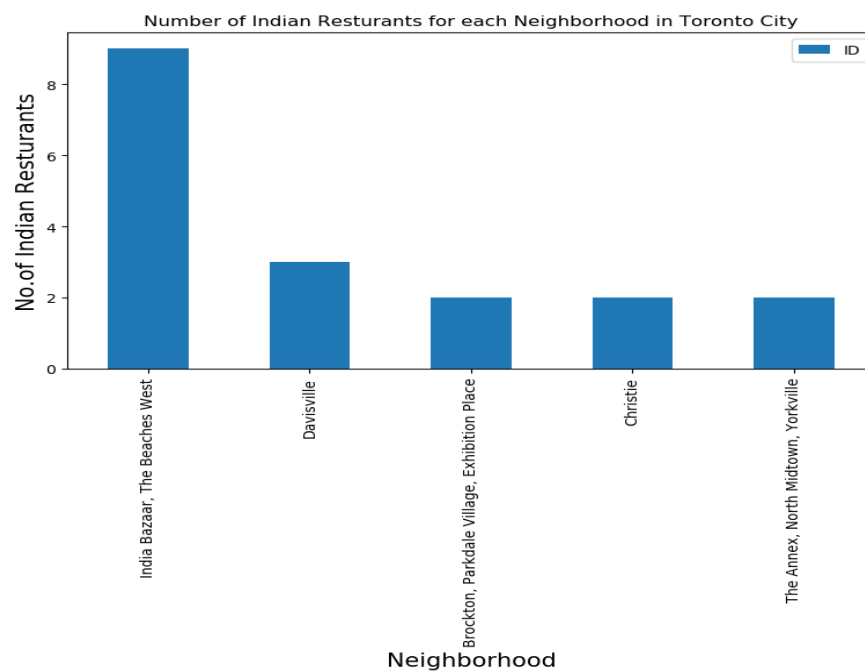


*fig ~ Number of Indian Restaurants in each neighborhoods in Toronto.*

# Get Ratings, Likes, Tips of the restaurants using Foursquare API

Fetching the Ratings, Likes, Tips of the restaurants using Foursquare API

```python
#prepare neighborhood list that contains indian resturants
column_names=['Borough', 'Neighborhood', 'ID','Name','Likes','Rating','Tips']
indian_rest_stats_to=pd.DataFrame(columns=column_names)
count=1


for row in indian_rest_to.values.tolist():
    Borough,Neighborhood,ID,Name=row
    try:
        venue_details=get_venue_details(ID)
        print(venue_details)
        id,name,likes,rating,tips=venue_details.values.tolist()[0]
    except IndexError:
        print('No data available for id=',ID)
        # we will assign 0 value for these resturants as they may have been
        #recently opened or details does not exist in FourSquare Database
        id,name,likes,rating,tips=[0]*5
    print('(',count,'/',len(indian_rest_to),')','processed')
    indian_rest_stats_to = indian_rest_stats_to.append({'Borough': Borough,
                                        'Neighborhood': Neighborhood,
                                        'ID': id,
                                        'Name' : name,
                                        'Likes' : likes,
                                        'Rating' : rating,
                                        'Tips' : tips
                                        }, ignore_index=True)
    count+=1
```

*Code Snippet*

| | Borough | Neighborhood | ID | Name | Likes | Rating | Tips |
|---|---|---|---|---|---|---|---|
| 0 | East Toronto | The Beaches | 4dcd7c6352b1f8915b7e7f7e | Delhi Bistro | 6 | 6.9 | 3 |
| 1 | East Toronto | The Danforth West, Riverdale | 4c1d5337eac020a1cb1048c2 | Sher-E-Punjab | 10 | 7.0 | 8 |
| 2 | East Toronto | India Bazaar, The Beaches West | 4ae0c7a8f964a520638221e3 | Udupi Palace | 79 | 8.7 | 31 |
| 3 | East Toronto | India Bazaar, The Beaches West | 4afc9816f964a520312422e3 | Motimahal | 25 | 8.1 | 13 |
| 4 | East Toronto | India Bazaar, The Beaches West | 4bac30a2f964a52018ea3ae3 | Bombay Chowpatty | 7 | 7.1 | 5 |

*Data Frame ~ Ratings, Tips, and Likes of the restaurants*

# The Average rating

Getting The Average rating of restaurants in particular Neighborhood

```
to_neighborhood_stats.columns=['Neighborhood','Average Rating']
to_neighborhood_stats=indian_rest_stats_to.groupby('Neighborhood',as_index=False).mean()[['Neighborhood','Rating']]
```

| | Neighborhood | Average Rating |
|---|---|---|
| 1 | Christie | 8.450000 |
| 4 | High Park, The Junction South | 8.300000 |
| 10 | The Annex, North Midtown, Yorkville | 8.250000 |
| 2 | Church and Wellesley | 8.000000 |
| 6 | Queen's Park, Ontario Provincial Government | 8.000000 |
| 9 | St. James Town, Cabbagetown | 8.000000 |
| 0 | Brockton, Parkdale Village, Exhibition Place | 7.850000 |
| 13 | University of Toronto, Harbord | 7.800000 |
| 3 | Davisville | 7.133333 |
| 8 | Runnymede, Swansea | 7.000000 |

*Data Frame Average ratings of restaurants*

Now we have the list top performing restaurants,

Lets visualize it.

```
to_map = folium.Map(location=geo_location('Toronto'), zoom_start=12)

to_map
```
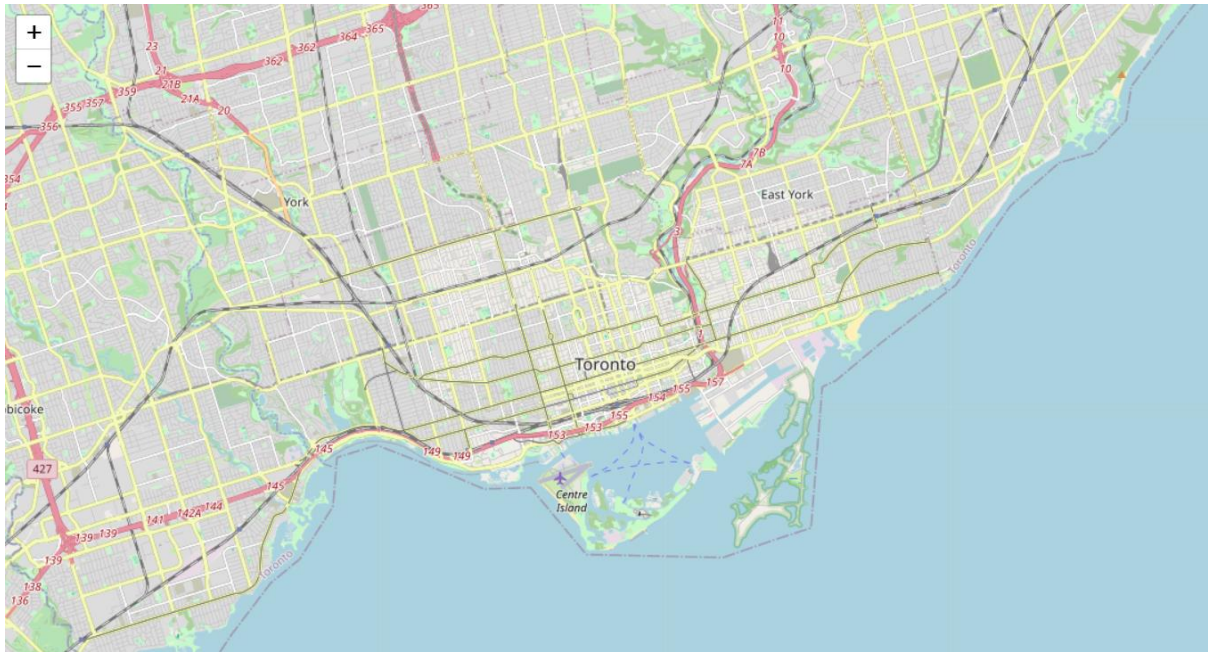
*Data Frame ~ Leaflet Map of Toronto*

```python
# instantiate a feature group for the incidents in the dataframe
incidents = folium.map.FeatureGroup()

# loop through the 100 crimes and add each to the incidents feature group
for lat, lng, in to_neighborhood_stats[['Latitude','Longitude']].values:
    incidents.add_child(
        folium.CircleMarker(
            [lat, lng],
            radius=10, # define how big you want the circle markers to be
            color='yellow',
            fill=True,
            fill_color='blue',
            fill_opacity=0.6
        )
    )
```

```python
to_map = folium.Map(location=geo_location('Toronto'), zoom_start=12)
to_geo = r'https://github.com/jasonicarter/toronto-geojson/blob/master/toronto_crs84.geojson'

to_map.choropleth(
    geo_data=to_geo,
    data=to_borough_stats,
    columns=['Borough', 'Average Rating'],
    key_on='feature.properties.boro_name',
    fill_color='YlOrRd',
    fill_opacity=0.7,
    line_opacity=0.2,
    legend_name='Average Rating'
)

# display map
# as this is huge map data , we will save it to a file
ny_map.save('borough_rating.html')
```
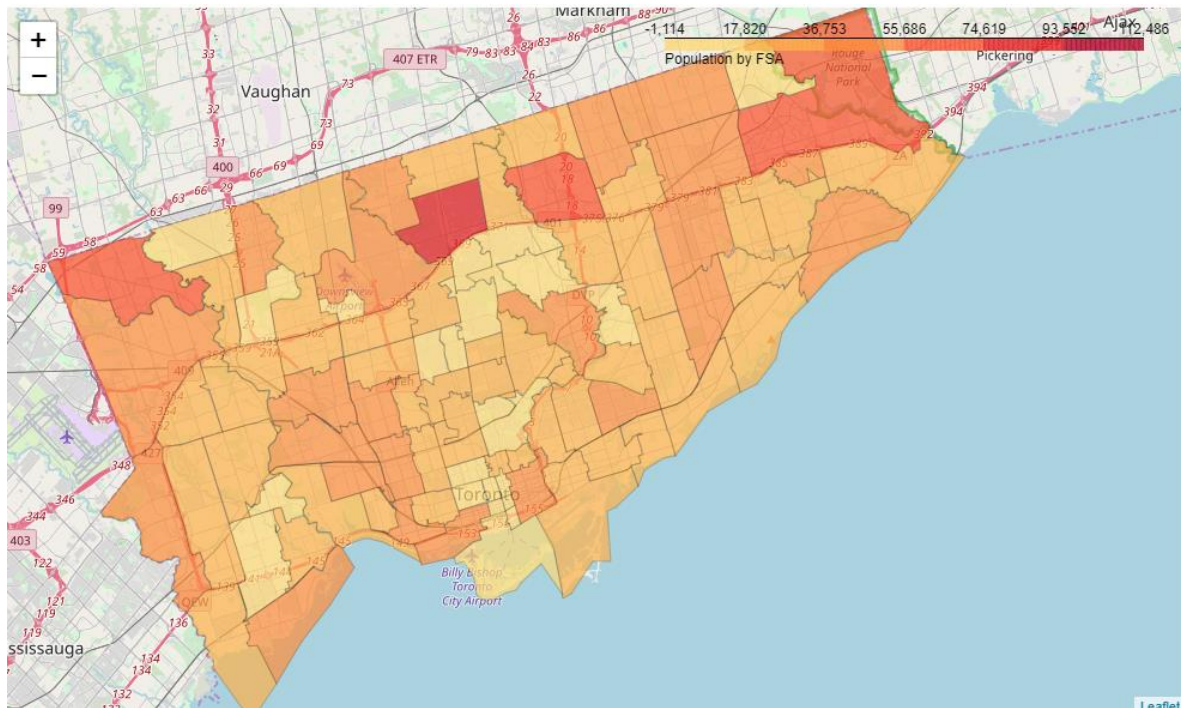
*Data Frame ~ Choropleth maps of Toronto based on avg Ratings.*

## Conclusion

Below are the Best Neighborhoods to open an Indian Restaurant

Christie, High Park, The Junction South, The Annex, North Midtown, Yorkville, Church and Wellesley, Queen's Park, Ontario Provincial Government, St. James Town, Cabbage town.

## Limitations

1. The Results are highly dependent on the ratings of the Restaurants.

2. The Rating Accuracy is highly dependent on Foursquare API.