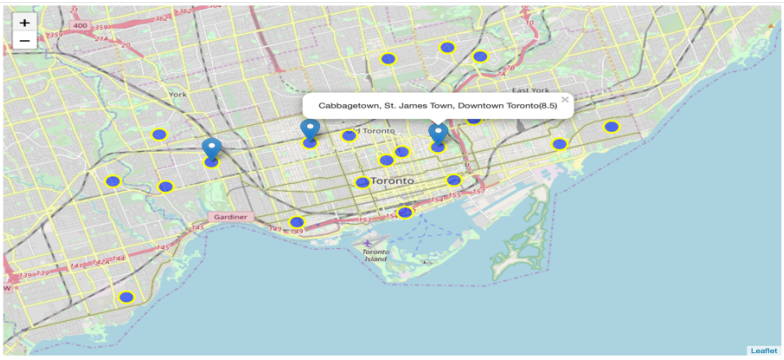
The Battle of Neighborhood – Capstone Project
Indian Restaurant analysis in Toronto



Introduction

Toronto is the largest city of Canada with the Greater Toronto population of over 6 million. It is ethnically very diverse city with various cultures with which come numerous cuisines including Chinese, Indian, Italian etc. Canada, in particular Toronto has a significant Indian population. The project is to analysis the neighborhoods of Toronto to find best places with Indian restaurants and potentially help find area where there are business opportunities to open an Indian restaurant.

Part of the project we will analyze and visualize major parts of Toronto with Indian restaurants.

We will try to get answers to following question

- Best places in Toronto to find good Indian restaurant.
- Areas having potential market for Indian restaurant.
- Neighborhood to stay with good Indian restaurant.

Also, same project can be extended to analyze other cities and other type of restaurants / business int the city

Following are the potential Target Audience who can get benefited from the analysis

- People who are looking to move or stay during Visit to Toronto and would like to have insight into where to find good Indian restaurants or the areas that are completely not suited for their taste.
- For audience who are looking to open an Indian restaurant and keen to know where the potential scope areas are

Moreover, the same project can be extended to other kind of business not necessarily Indian restaurants and it is it is easily extendable to other cities once we identify similar source data for those cities.

Data

As we need to explore neighborhoods in the city of Toronto, the Toronto neighborhoods data is key for this project. Unfortunately, the data is for the Toronto neighborhood data is not widely available on the Internet in the structured format, hence we need to scrap it through an existing Wikipedia page exists that has all the information we need to explore and analyze the neighborhoods in Toronto.

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

We also need the coordinates of the neighborhood to obtain venue details of the neighborhood also to visualize the data in the map.

We use http://cocl.us/Geospatial_data for getting the latitude and longitude of the postal codes of the neighborhood

We rely extensively on Foursquare for collecting Venue data pertaining to Indian restaurant and also to collect rating of each of the Indian restaurant.

We will also use the Geospacial data for getting the boundaries of Borough to visualize data on the map

Neighborhood, with Postal code and Coordinates of each postal code. Neighborhoods of same postal codes are treated as single line item to avoid data sparsity

	Postcode	Borough	Neighbourhood	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

Gather (Indian) restaurants in each of the neighborhood (with predefined radius from coordinates of the Neighborhood/Postal code). This is done using Foursquare

	Borough	Neighbourhood	ID	Name
13	East Toronto	The Beaches West, India Bazaar	4ae0c7a8f964a520638221e3	Udupi Palace
14	East Toronto	The Beaches West, India Bazaar	4afc9816f964a520312422e3	Motimahal
15	East Toronto	The Beaches West, India Bazaar	4bac30a2f964a52018ea3ae3	Bombay Chowpatty
16	East Toronto	The Beaches West, India Bazaar	4ad9052cf964a520301721e3	Regency Restaurant
17	East Toronto	The Beaches West, India Bazaar	4bbcc0efa0a0c9b60ebd1a0f	Haandi 2000
18	East Toronto	The Beaches West, India Bazaar	4d8d278a1d06b1f712942a3b	Gautama

The Above Restaurant Ids are then used to collect the restaurant ratings using venue details of Foursquare data.

	Borough	Neighbourhood	ID	Name	Likes	Rating	Tips
0	Central Toronto	Davisville	5169d445e4b07de190b5c3d6	Marigold Indian Bistro	11	7.9	11
1	Central Toronto	Davisville	4b7ccc72f964a520e3a52fe3	Banjara Indian Cuisine	24	6.7	19
2	Central Toronto	The Annex, North Midtown, Yorkville	4c62c59ce1621b8dd0332453	Roti Cuisine of India	39	8.6	18
3	Central Toronto	The Annex, North Midtown, Yorkville	4ad4c060f964a5204af720e3	The Host	31	7.4	13
4	Downtown Toronto	Cabbagetown, St. James Town	4c8c21fdf0ce236ab28e15ef	Butter Chicken Factory	19	8.5	10

Methodology:

GitHub was used as the code repository for this project/study. The master data for the project is the **Postcode, Borough, Neighborhood information of Toronto City**. The data was grouped at Postcode level and if there are multiple Neighborhood with same Postcode they are considered as one Neighborhood for the study. This is to avoid data sparsity if the Neighborhoods are so close and small.

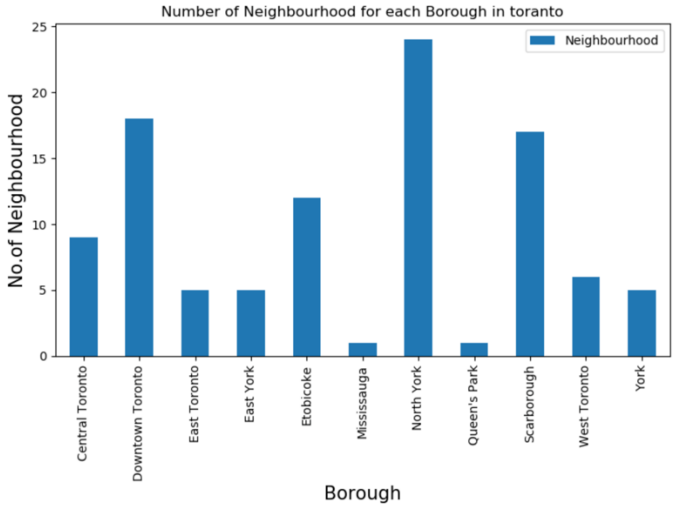
Data was cleansed for both Borough and/or Neighborhood are empty (Not Assigned)

	Postcode	Borough	Neighbourhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Postal code is then used to determine Latitude and longitude of each of this neighborhood. I used a simple method to get the Latitude, Longitude from http://cocl.us/Geospatial_data by passing the postal code. later this data is appended to the above master data to get the below format

	Borough	Neighbourhood	Latitude	Longitude
47	Central Toronto	Davisville	43.704324	-79.388790
45	Central Toronto	Davisville North	43.712751	-79.390197
49	Central Toronto	Deer Park, Forest Hill SE, Rathnelly, South Hi...	43.686412	-79.400049
64	Central Toronto	Forest Hill North, Forest Hill West	43.696948	-79.411307
44	Central Toronto	Lawrence Park	43.728020	-79.388790

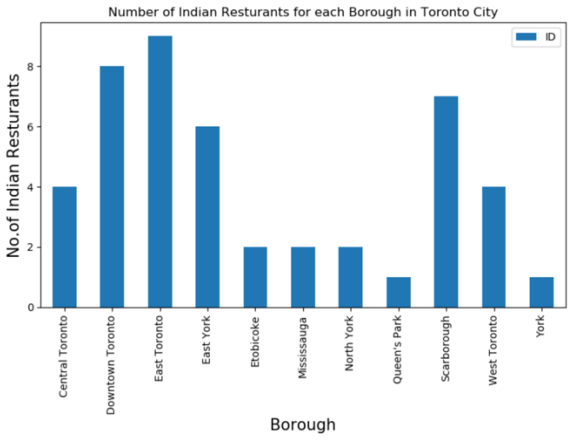
We also quickly visualize Boroughs and number of Neighborhood in each borough

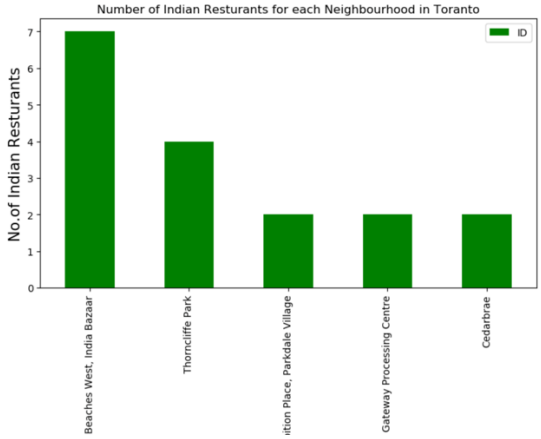


Next step is the key step to Implementation of Foursquare API to get all the Indian restaurant against each of the Neighborhood. This is achieved through calling the API for venues, with Latitude and Longitude of each of the Neighborhood, a radius of 1Km was used (which is optimal to stay within the neighborhood and avoid too many overlap), the data was filter for Category of "Indian Restaurant". The resulting data is as below.

	Borough	Neighbourhood	ID	Name
0	Central Toronto	Davisville	5169d445e4b07de190b5c3d6	Marigold Indian Bistro
1	Central Toronto	Davisville	4b7ccc72f964a520e3a52fe3	Banjara Indian Cuisine
2	Central Toronto	The Annex, North Midtown, Yorkville	4c62c59ce1621b8dd0332453	Roti Cuisine of India
3	Central Toronto	The Annex, North Midtown, Yorkville	4ad4c060f964a5204af720e3	The Host
4	Downtown Toronto	Cabbagetown, St. James Town	4c8c21fdf0ce236ab28e15ef	Butter Chicken Factory

Now since we have collected all the Indian restaurant and mapped to the Neighborhood, we can quickly visualize things like below (Boroughs vs Number of Indian restaurants, to Neighborhood by number of Indian restaurants.)



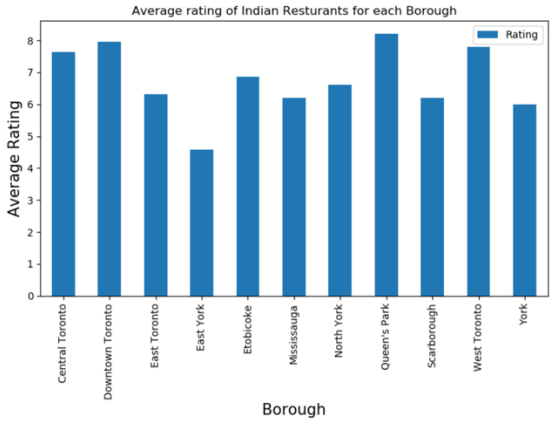


Next step is to collect the Indian restaurant details. We are interested mainly in Ratings and also, we collected number of likes as well.
Once again use Foursquare API to get the Venue Details by passing the restaurant IDs one by One. Once we have the ratings in place, we can get the average Indian restaurant rating across Boroughs and as well as across Neighborhood. This gives us very good insight into where the top Indian restaurants are located.

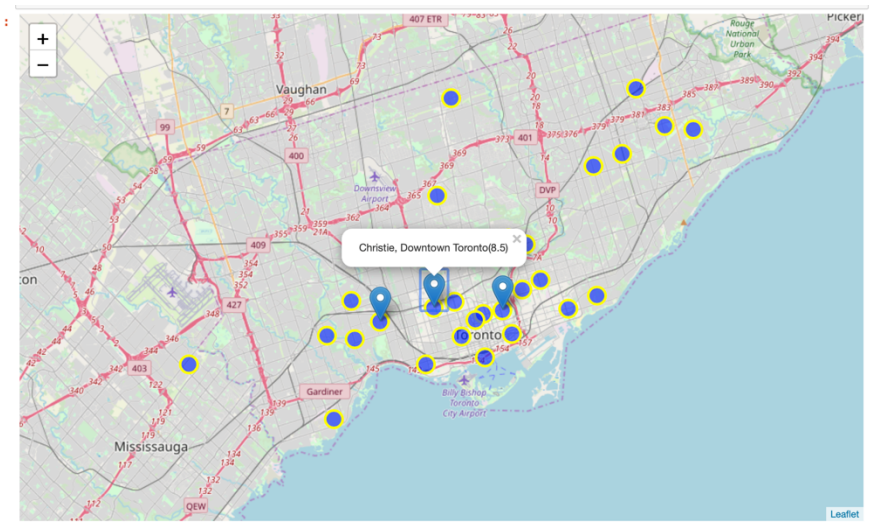
I used Data frame and various python methods to get average rating, number of restaurants, merging dfs

	Borough	Neighbourhood	ID	Name	Likes	Rating	Tips
13	East Toronto	The Beaches West, India Bazaar	4ae0c7a8f964a520638221e3	Udupi Palace	79.0	8.7	31.0
14	East Toronto	The Beaches West, India Bazaar	4afc9816f964a520312422e3	Motimahal	25.0	7.8	13.0
15	East Toronto	The Beaches West, India Bazaar	4bac30a2f964a52018ea3ae3	Bombay Chowpatty	7.0	7.5	5.0
16	East Toronto	The Beaches West, India Bazaar	4ad9052cf964a520301721e3	Regency Restaurant	6.0	6.4	4.0
17	East Toronto	The Beaches West, India Bazaar	4bbcc0efa0a0c9b60ebd1a0f	Haandi 2000	3.0	6.4	7.0
18	East Toronto	The Beaches West, India Bazaar	4d8d278a1d06b1f712942a3b	Gautama	15.0	5.9	15.0
19	East Toronto	The Beaches West, India Bazaar	0	0	0.0	0.0	0.0

I did plot the Boroughs against Average ratings



By sorting the data frame, we can easily identify Neighborhood with average top rating (3 of them). I used folium library to visualize these locations on the map on top of all Neighborhood with the Indian restaurant.



	Neighbourhood	Average Rating
3	Cabbagetown, St. James Town	8.5
7	Christie	8.5
14	High Park, The Junction South	8.3

Similar data can be pulled at Borough level as below.

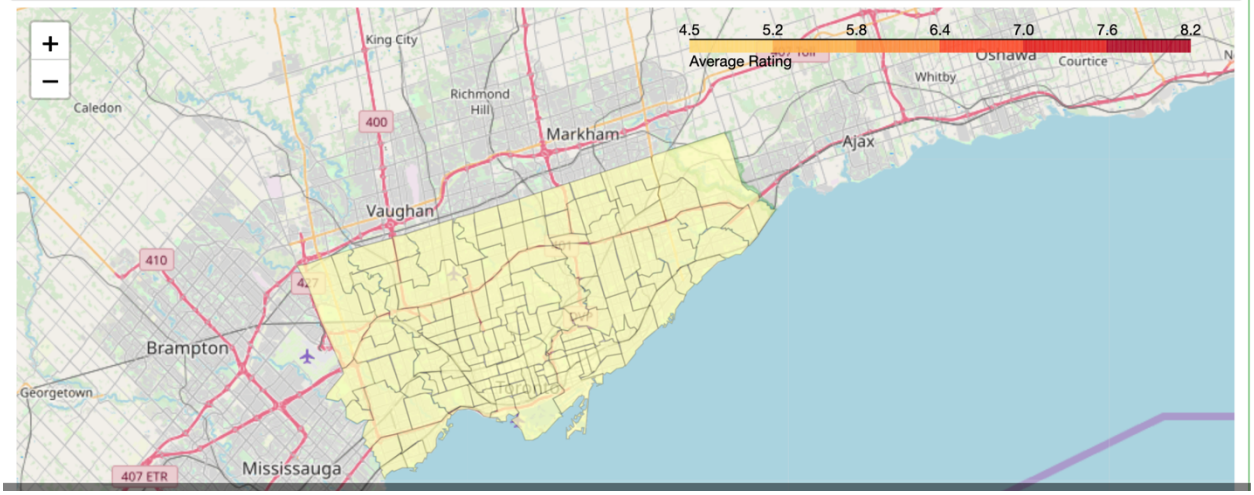
	Borough	Average Rating
1	Downtown Toronto	7.9625
7	Queen's Park	8.2000
9	West Toronto	7.8000

With Borough boundaries and average Indian restaurant rating for each Borough I created a choropleth map to indicate best and worst Borough for Indian restaurant.

```
Toronto_map = folium.Map(location=geo_location('Toronto'), zoom_start=12)
Toronto_geo = r'toronto_crs84.geojson'

Toronto_map.choropleth(
    geo_data=Toronto_geo,
    data=Toronto_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
Toronto_map.save('borough_rating.html')
Toronto_map
```



Results

From the above Analysis we can infer following details which would really help us in answering our questions when starting the project.

- East Toronto Borough is the one with largest number of Indian restaurants
- 'The Beaches West, India Bazaar' is the number-1 by number of Indian restaurants, followed by Thorncliffe Park
- When it comes to top rated Indian restaurants following neighborhoods
    - Cabbagetown, St. James Town
    - Christie
    - High Park, The Junction South
- With respects to ratings **Queen's Park** Borough tops
- This also gives insight into where there are gaps in terms of Indian Restaurant. and can be an input for business owners who want to open one.

Discussion

As mentioned in the intro, Toronto is the biggest city of Canada with majority of immigrant population. Indian population is significant and it a rapidly growing city. This exercise will provide valuable insight for people looking for locations with best Indian restaurant or easy access to one.

This result can extend with use of other powerful data like Demography to analysis population in each of the areas where there are gaps in Indian restaurants to help in determining best location to open one.

The same method can be applied to other type of business and extended to other cities as well

Conclusion

The Exercise / project greatly helped me in applying key concept I learnt throughout this course and helped gaining confidence in applying Python for real world problems. I expect the results to useful for people interested in Indian restaurant. Also, the analysis is a good as the data, several other data like population, demography, other nearby locations/Venues etc. can be used to enhance this project.