Battle of neighbourhood: A restaurant settlement problem in Toronto

Problem Statement:

I would like to explore the prospect of opening a restaurant in Toronto. Toronto is a city of diversity and hub of immigrant people. In this project, I will illustrate the decision-making process of establishing a restaurant in Toronto analyzing the neighbourhoods in Toronto to identifying the most profitable area. It is important to get an idea of the neighborhood before opening a restaurant business to have an idea about their ethnicity and taste.

Target Audience

Business personnel who wants to invest or open an Indian restaurant in Toronto. This analysis will be a comprehensive guide to start or expand restaurants targeting the Indian crowd.

Data Sources

The following dataset I am using to do this report,

a) (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) wiki page to get all the information about the neighbourhoods present in Toronto. This page has the postal code, borough & the name of all the neighbourhoods present in Toronto.
b) "https://cocl.us/Geospatial_data" csv file to get all the geographical coordinates of the neighbourhoods. To get information about the distribution of population by their ethnicity
c) "Demographics of Toronto.(https://en.m.wikipedia.org/wiki/Demographics_of_Toronto#Ethnic_diversity) wiki page. Using this page I'm going to identify the neighbourhoods which are densely populated with Indians as it might be helpful in identifying the suitable neighbourhood to open a new Indian restaurant.

To get location and other information about various venues in Toronto I'm using Foursquare's explore API. Using the Foursquare's explore API (which gives venues recommendations), I'm fetching details about the venues up present in Toronto and collected their names, categories and locations (latitude and longitude). From Foursquare API (https://developer.foursquare.com/docs), I retrieved the following for each venue:

- Name: The name of the venue.
- Category: The category type as defined by the API.
- Latitude: The latitude value of the venue.
- Longitude: The longitude value of the venue.

Methodology:

I have scraped the following Wikipedia page, “List of Postal code of Canada: M” in order to obtain the data about the Toronto & the Neighbourhoods in it. Then I have done some cleaning we got the proper data frame with the Postal code, Borough & Neighborhood information. Next important step is adding the geographical coordinates to these neighborhoods. To do so I’m extracting the data present in the Geospatial Data csv file and I’m combining it with the existing neighborhood data frame by merging them both based on the postal code. Now, the data frame has 11 boroughs and 103 neighborhoods.

Another factor that can help us in deciding which neighborhood would be best option to open a restaurant is, the distribution of population based on the ethnic diversity for each neighborhood. As this helps us in identifying the neighborhoods which are densely populated since that neighborhood would be an ideal place to open a restaurant.

Scraped the following Wikipedia page, “Demographics of Toronto” in order to obtain the data about the Toronto & the Neighborhoods in it. Compared to all the neighborhoods in Toronto below given neighborhoods only had considerable amount of Indian crowd. We are examining those neighborhood’s population to identify the densely populated neighborhoods with Indian population.

There were only six neighborhoods in Toronto which Indian population spread across, so we are gathering the population, it’s percentage in each riding in those neighborhoods.

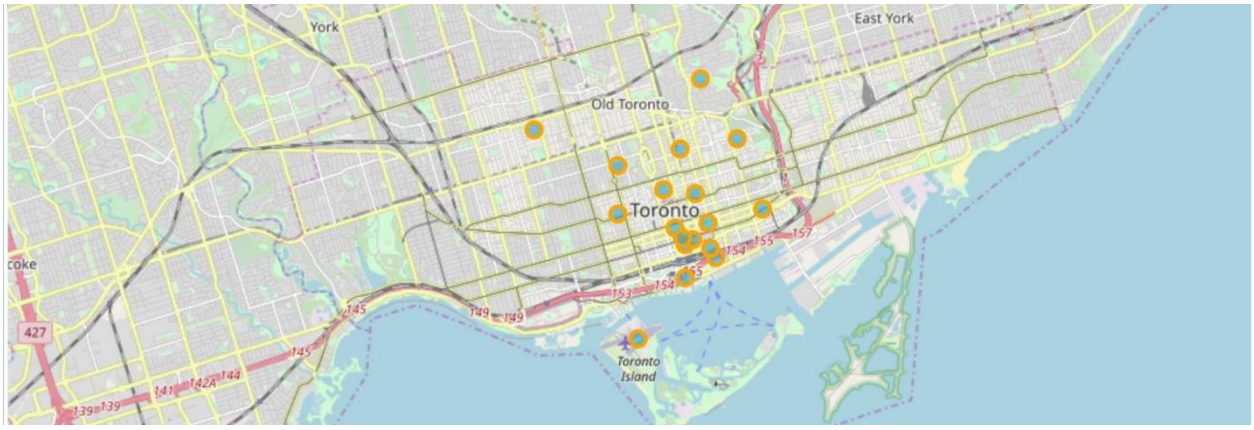
Get location data using Foursquare

Foursquare API is very useful online application used my many developers & other applications like Uber etc. In this project I have used it to retrieve information about the places present in the neighborhoods of Toronto. The API returns a JSON file and we need to turn that into a data-frame. Here I’ve chosen 100 popular spots for each neighborhood within a radius of 1km.

Exploratory Data Analysis:

Folium Library and Leaflet Map

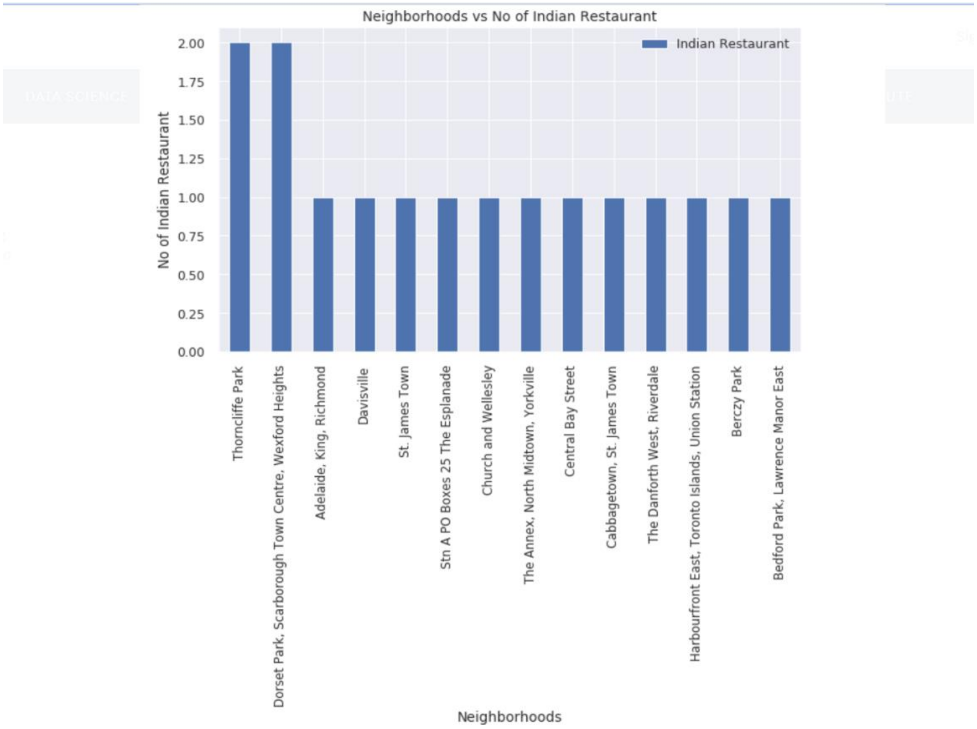
Folium is a python library, I’m using it to draw an interactive leaflet map using coordinate data.



Relationship between neighbourhood and Indian Restaurant

First we will extract the Neighbourhood and Indian Restaurant column from the above Toronto data frame for further analysis:

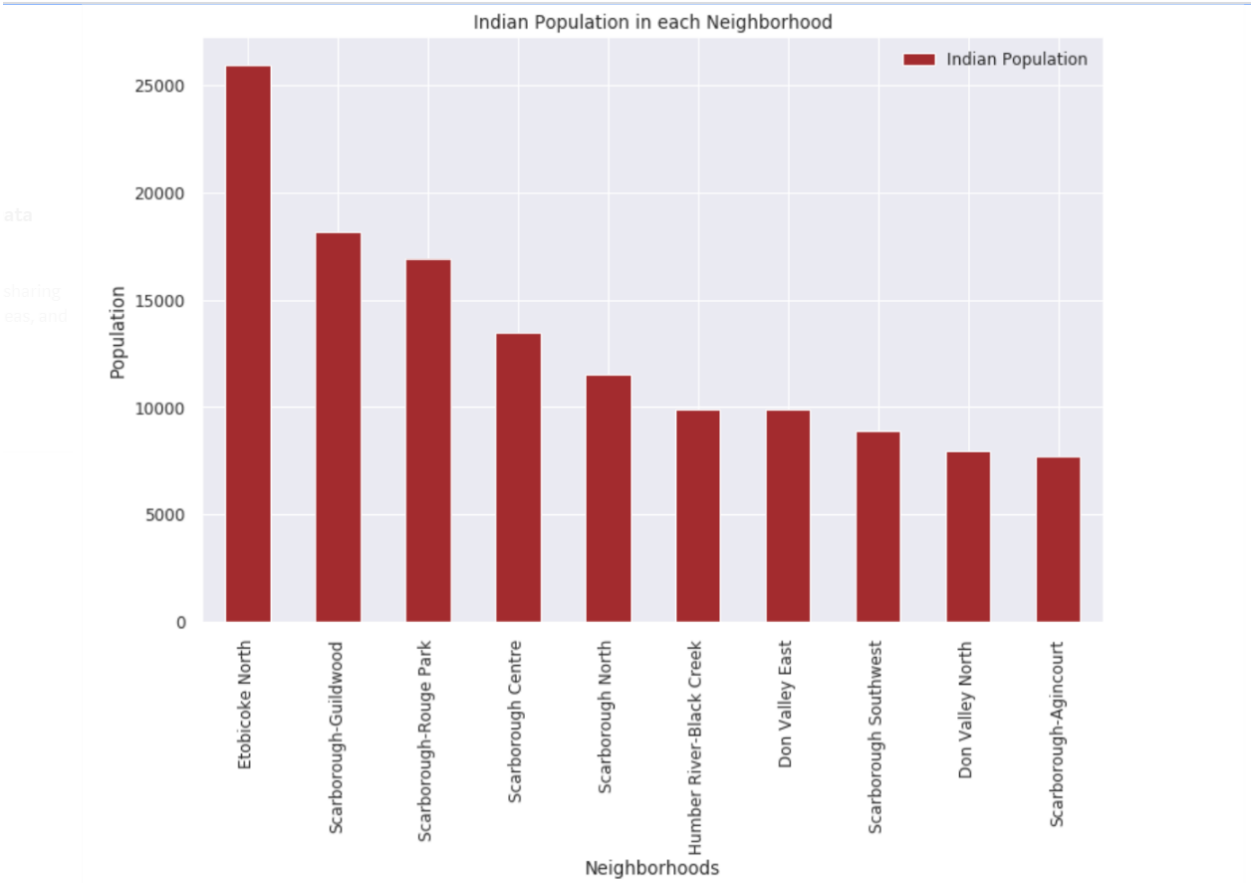
After performing pandas one hot encoding for the venue categories, let us merge this data frame with the Toronto Data Frame with latitude & longitude information on neighbourhood.



Relationship between neighbourhood and Indian population

Another key feature is the distribution of Indian crowd in each neighbourhoods. Let us analyse the neighbourhoods and identify the neighbourhoods with highest number of Indian population.

To achieve that we are joining all the neighbourhood's data frame from using the wiki page with ethnic population and in that we are extracting just the Indian population for each neighbourhood.



This analysis & visualization of the relationship between neighborhoods & Indian population present in those neighborhoods helps us in identifying the highly populated Indian neighborhoods. Once we identify those neighborhoods it helps us in deciding where to place the new Indian restaurant. Indian restaurant placed in an densely populated Indian neighborhood is more likely to get more Indian customers than a restaurant placed in a neighborhood with less or no Indian population. Thus this analysis helps in the determining the success of the new Indian restaurant.

Predictive Modelling:

Clustering Neighborhoods of Toronto:

First step in K-means clustering is to identify best K value meaning the number of clusters in a given dataset. To do so we are going to use the elbow method on the Toronto dataset with Indian restaurant percentage. After analysing using elbow method using distortion score & Squared error for each K value, looks like K = 6 is the best value.

Results and Discussion:

Results

We have found out that in those 11 boroughs we identified that only Central Toronto, Downtown Toronto, East Toronto, East York, North York & Scarborough boroughs have high amount of Indian restaurants with the help of Violin plots between Number of Indian restaurants in Borough of Toronto. . With the help of clusters examining & violin plots looks like Downtown Toronto, Central Toronto, East York are already densely populated with Indian restaurants. So it is better idea to leave those boroughs out and consider only Scarborough, East Toronto & North York for the new restaurant's location. After careful consideration it is a good idea to open a new Indian restaurant in Scarborough borough since it has high number of Indian population which gives a higher number of customers possibility and lower competition since very less Indian restaurants in the neighbourhoods.

Conclusion:

We have used many python libraries to fetch the data , to manipulate the contents & to analyse and visualize those datasets. We have made use of Foursquare API to explore the venues in neighbourhoods of Toronto, then get good amount of data from Wikipedia which we scraped with help of Wikipedia python library and visualized using various plots present in seaborn & matplotlib. We also applied machine learning technique to predict the output given the data and used Folium to visualize it on a map.