**Battle of the Neighbourhoods**

**Saloni Garg**

**June 22,2020**

**In a city, if someone is looking to open a restaurant, where would you recommend that they open it?**

***Introduction***

This final project explores the best locations for restaurants throughout the city of Toronto. In this project we would be able to know the correct places where a resturant is in high demand. This project will help people like if a contractor is trying to start their own business, then he would be able to find the best place where should they setup their office. A resturant should be at the most visited and popular places so that it can serve more and more people.

***Data***

In order to answer the above questions, data on Toronto neighbourhoods, boroughs to include boundaries, latitude, longitude, restaurants, and restaurant ratings and tips are required.

Toronto City data containing the neighbourhoods and boroughs will be obtained from the data source: '<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'>

Toronto City data containing neighbourhood boundaries will be obtained from the data source: <http://cocl.us/Geospatial_data>

All data related to locations of restaurants will be obtained via the FourSquare API utilized via the Request library in Python.

##### *Methodology*

• Data will be collected from '<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'> and cleaned and processed into a dataframe.

• Data will be sorted based on rankings

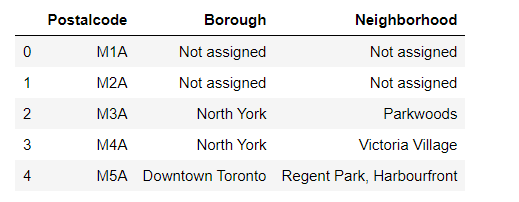
• Finally, the data be will be visually assessed using graphing from various Python libraries.

##### *Problem Statement*

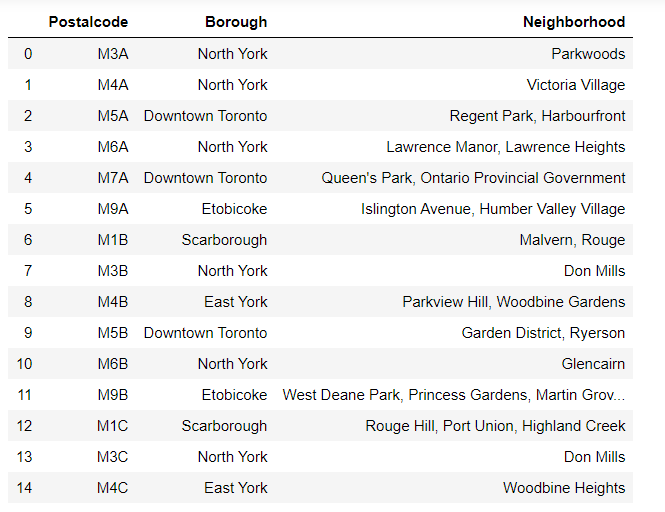
In what Neighbourhood and/or borough should I open a restaurant to have the best chance of being successful?

**Data cleaning**

Data downloaded or scraped from multiple sources were combined into one table. There were a lot of missing values because of lack of record keeping. We process the cells that have an assigned borough. Ignoring the cells with a borough that is Not assigned. Droping row where borough is "Not assigned". Then we sorted all the rows in descending order. We also verified the number of rows and column that we have now after data cleaning.

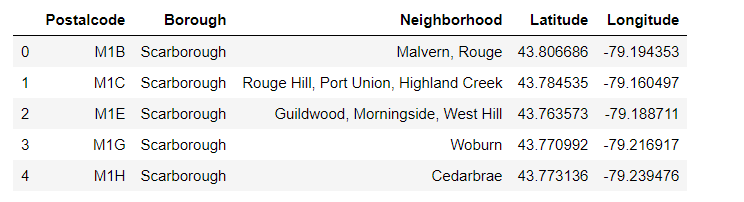
Our Table before data cleaning***:***

Our Table after data cleaning:



**Exploratory Data Analysis:**

After data cleaning, there were 103 rows and 3 columns as samples in the data. Upon examining the data, it was clear that there was some redundancy in the number of rows, for this we have used shape function. For further analysis we found the location of most visited or popular places so that restaurant run well at these places. For which we have used geospatial data, link of which I have provided above already. I have particularly found out the coordinates in latitudes and longitudes which looked like:



##### *Finding maps and clusters where the most populated places are:*

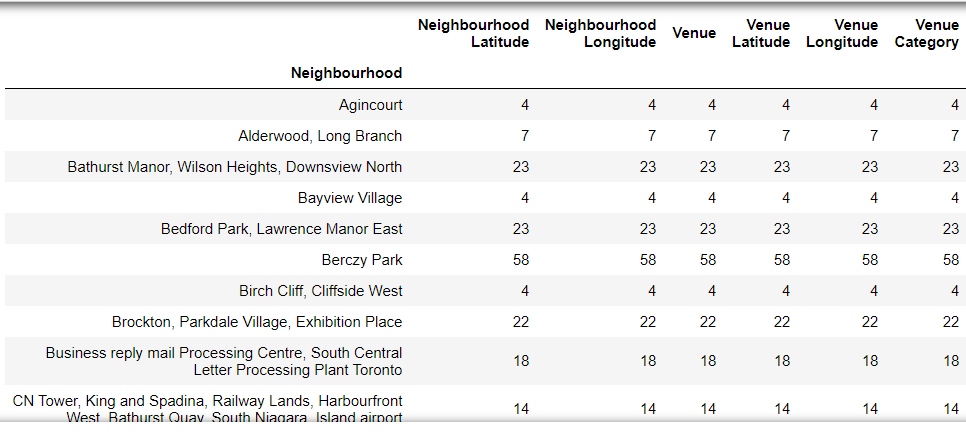
##### After finding out the coordinated of selected places I plotted them upon the map of Toronto. So as to take a quick look upon the places or area which is suitable for us. It is easier to select a location by this manner because we can easily look upon larger surroundings of that area. In this I took help of folium api . So here is the map:

##### 

Now, by using Foursquare Credentials and Version I tried to find out from the places which I have shortlisted above which are the venues which are popular and around which of the venues we can open the resturants . Here are some of the venues***:***



Now ,we will count the number of venues in each category, So that if more number of resturants are at the places we should not open the restaurant there because high competition would be there that’s why its important to count the number of those places and here it is:



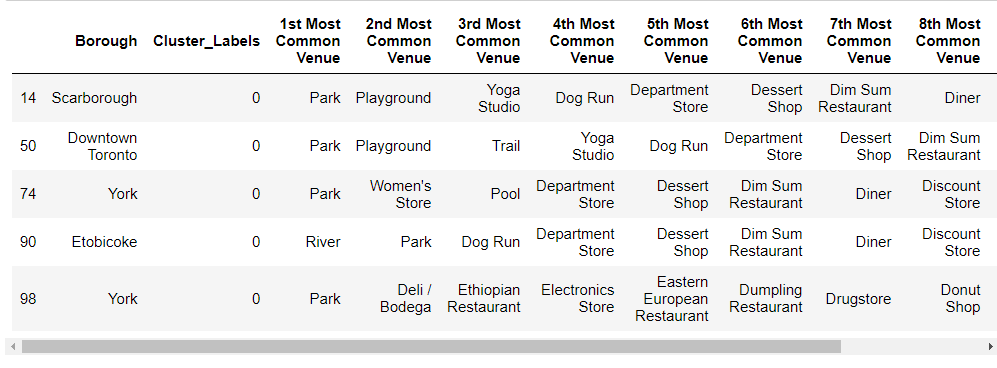
After that we will analyse each neighbourhood and try to sort places accordingly. So here are some of the shortlisted places:



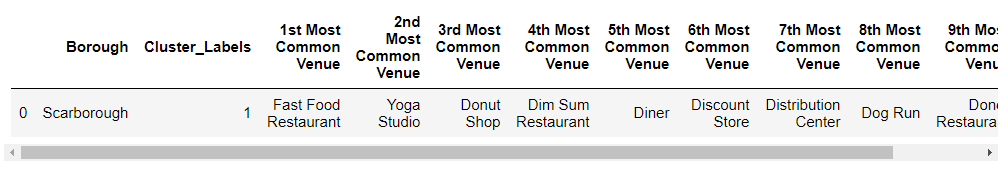
***Cluster Neighbourhoods:***

For clustering the above data I have used kmeans clustering. From the above shortlistings we can group the data into three clusters which indicates the suitable places to open a restaurant. It gives a concise idea of the whole neighbourhood. So here are the three clusters:

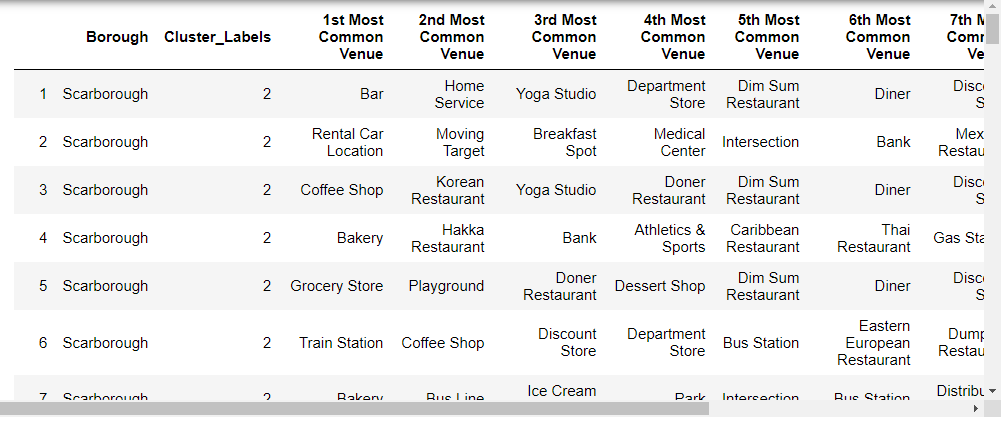
***Cluster 1:***

******

**Cluster 2:**



**Cluster 3:**



**Conclusion:**

##### *Cluster 1,2 & 3 represents the popular places of the city, therefore if someone is looking to open a restaurant, I would recommend that they open it amongst these places.*