**The Battle of Neighbourhoods:**

**Introduction Business Proposal:**

This Project aims to help people in exploring better facilities around their neighbourhood. It will help people in making right decisions while selecting best neighbourhood out of numbers of other neighbourhoods in Scarborough, Toronto.

As people are migrating to Canada they need a lot of research for a good housing price. This project can help those people who are looking for better neighbourhoods. For ease of accessing to Super market, medical shops, grocery shops, Cafe, School, mall, theatre, hospital, etc. This Project creates an analysis of features for a people migrating to Scarborough to search a best neighbourhood as a comparative analysis between neighbourhoods. The features include median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, and good management for emergency.

The major purpose of this project is to suggest a better neighbourhood in a new city for the people who are shifting there. Social presence in society in terms of similar minded people. Connectivity to the airport, city centre, bus stops, markets, hospitals, sorted list of house in terms of housing prices in ascending or descending order and sorted list of schools in terms of location, fees, rating and reviews.

**The Location:** Scarborough is a popular destination for new immigrants in Canada to reside. As a result, it is one of the most diverse and multicultural areas in the Greater Toronto Area, being home to various religious groups and places of worship. Although immigration has become a hot topic over the past few years with more governments seeking more restrictions on immigrants and refugees, the general trend of immigration into Canada has been one of on the rise.

**Work Flow:** Using credentials of Foursquare API features of near-by places of the neighbourhoods would be mined. Due to http request limitations the radius parameter would be set to 500 and the number of places per neighbourhood parameter would reasonably be set to 100.

In this project I am using Four-square API as its prime data gathering source has a database of millions of places, especially their places API which provides the ability to perform location sharing, location search and details about a business.

**Clustering Approach:** To compare the similarities of two cities, I decided to explore neighbourhoods, segment them, and group them into clusters to find similar neighbourhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

### Libraries used to develop the project are:

**Folium:** Python visualization library would be used to visualize the neighbourhoods cluster distribution of using interactive leaflet map.

**Pandas:** For creating and manipulating dataframes.

**Beautiful Soup and Requests:** To scrap and library to handle http requests.

**Matplotlib:** Python Plotting Module.

**XML:** To separate data from presentation and XML stores data in plain text format.

**Geocoder:** To retrieve Location Data.

**Scikit Learn:** For importing k-means clustering.

**JSON:** Library to handle JSON files.

### Data Description:

**Link for the data:** <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

I am using the Scarborough dataset which we scrapped from Wikipedia on Week 3 which includes the latitude, longitude, and zip codes.

**The Foursquare API Data:**

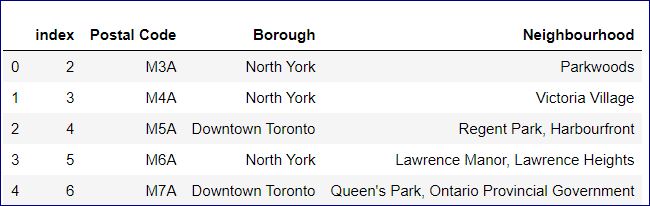
Data about different venues in different neighbourhoods of that specific borough. In order to gain the information we will use the "Foursquare" locational information. Foursquare is a location data provider with information about all types of venues and events within an area of interest. Such information includes venue names, locations, menus and photos. For each neighbourhood, we have chosen the radius to be 100 meter.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes.

**The information obtained per venue are:**

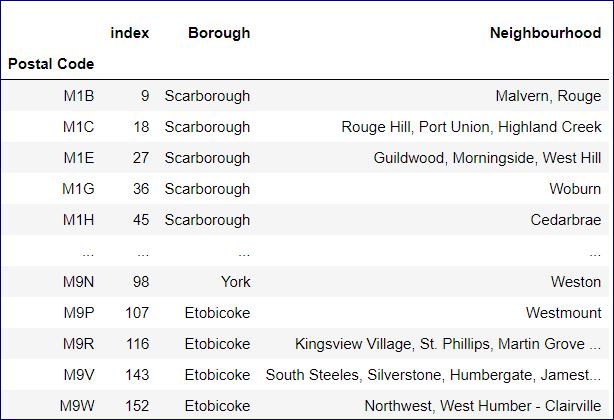
1. Neighbourhood
2. Neighbourhood Latitude
3. Neighbourhood Longitude
4. Venue
5. Name of the venue
6. Venue Latitude
7. Venue Longitude
8. Venue Category

Get the HTML page as link, using read\_html we convert the html data into list of Data frame objects and we are removing cells which have borrow not assigned.

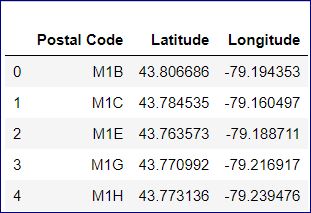


(first 5 rows of the data frame – df)

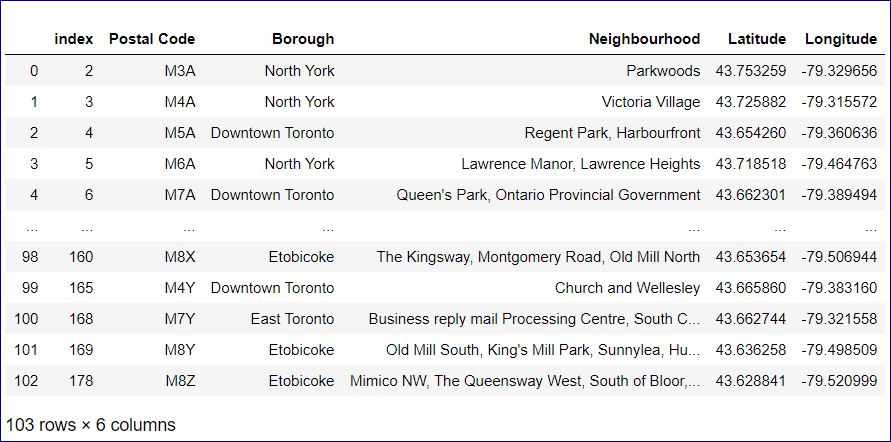
(After grouping by “Postal code”)



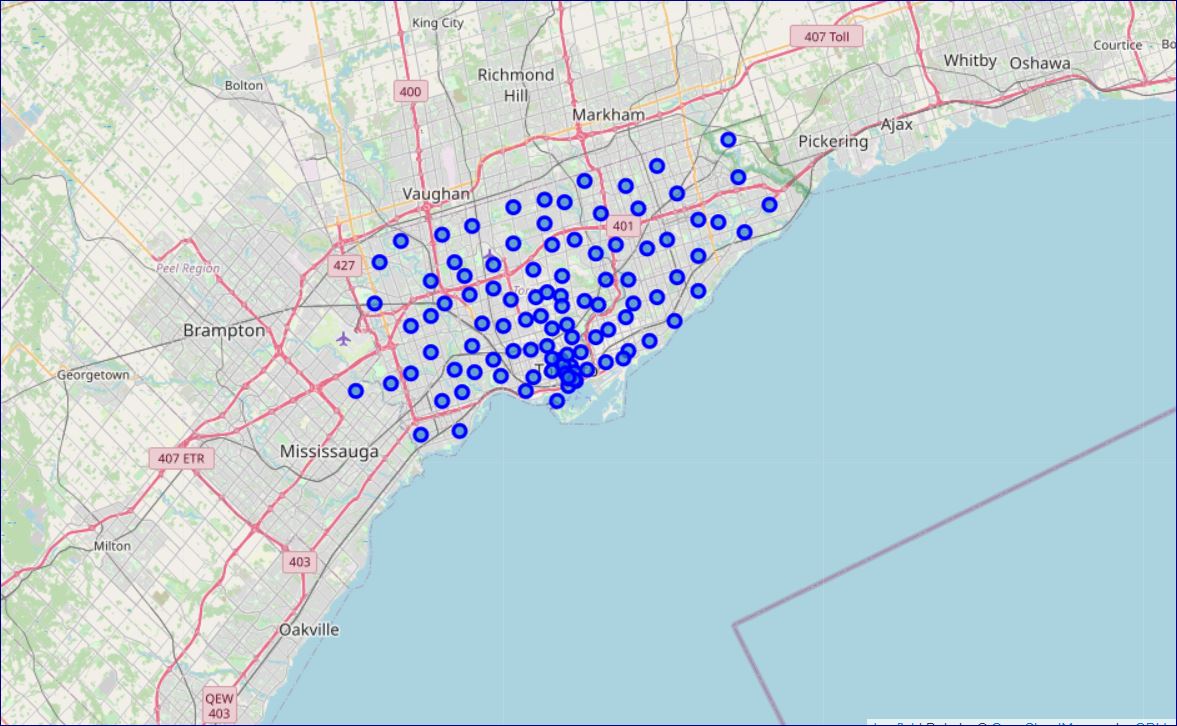
(first 5 rows of the data frame - df\_ (Geospatial\_data))



**Cleaning and joining both df and df\_**



Adding markers to map:



**Methodology:**

This Project aims to help people in exploring better facilities around their neighbourhood. It will help people in making right decisions while selecting best neighbourhood out of numbers of other neighbourhoods in Scarborough, Toronto.

The major purpose of this project is to suggest a better neighbourhood in a new city for the people who are shifting there. Social presence in society in terms of similar minded people. Connectivity to the airport, city center, bus stops, markets, hospitals, sorted list of house in terms of housing prices in ascending or descending order and sorted list of schools in terms of location, fees, rating and reviews.

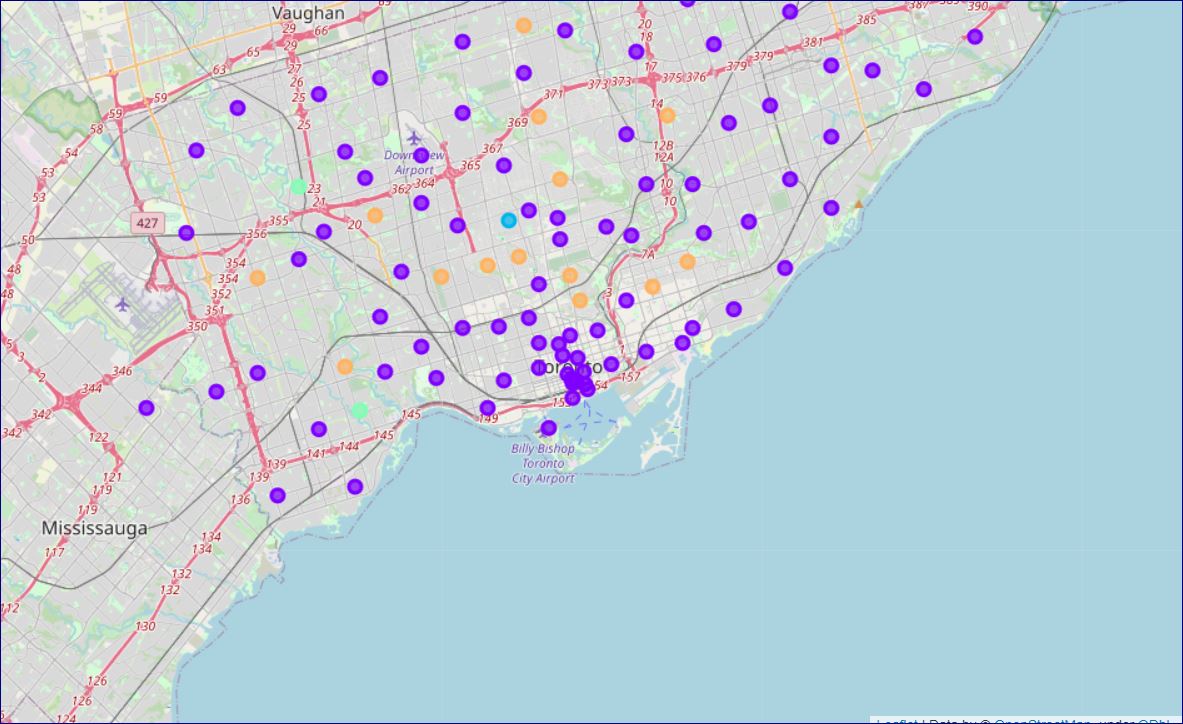
First we need to clean both the dataframes and join the two. Then we need to find the venues nearby by passing the location of interest.

In the map clusters of locations that meet some basic requirements are coloured and displayed.

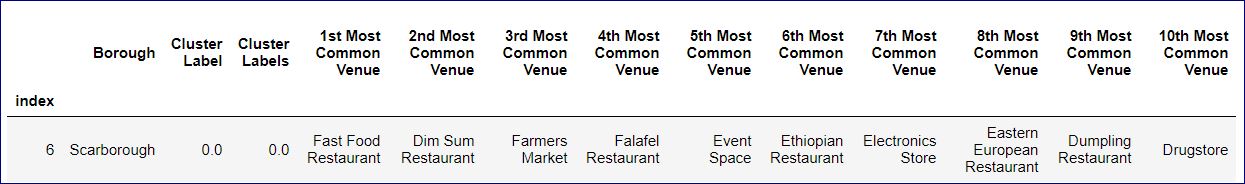
**Analysis:**

**Neighbourhood venues in Toronto:**

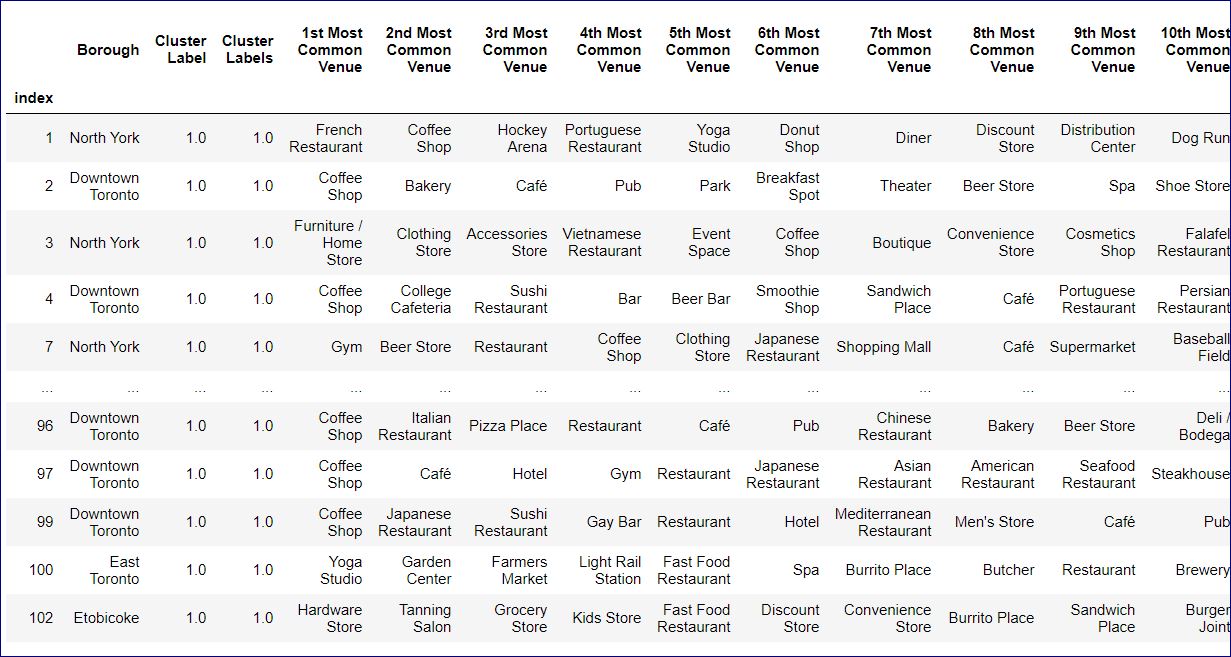
****

****

After merging the non NAN values for ‘Cluster label’ = 0



After merging the non NAN values for ‘Cluster label’ = 1



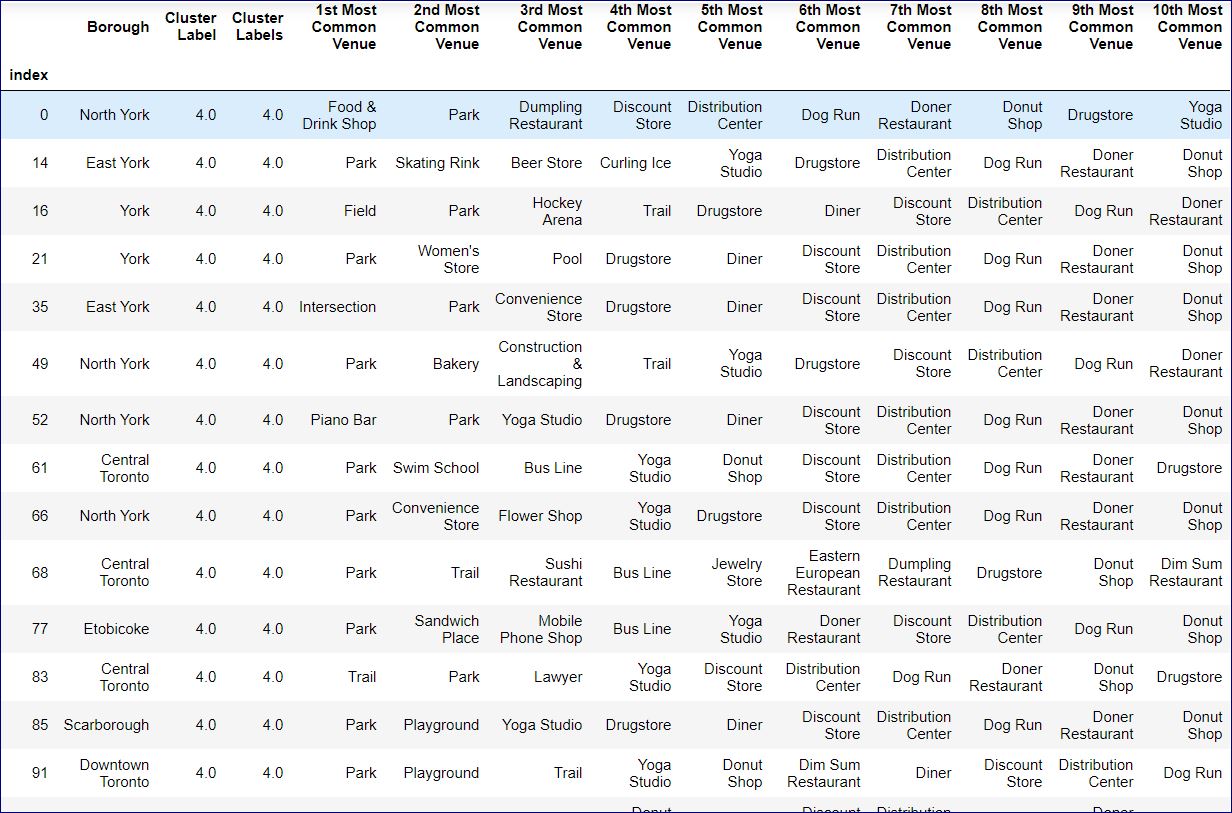
After merging the non NAN values for ‘Cluster label’ = 2



After merging the non NAN values for ‘Cluster label’ = 3



After merging the non NAN values for ‘Cluster label’ = 4



**Results:**

Similarities were compared and neighbourhoods were explored. The data was retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes.

**Conclusion:**

This project can help those people who are looking for better neighbourhoods. For ease of accessing to Super market, medical shops, grocery shops, Cafe, School, mall, multiplexes, hospital, etc. This Project creates an analysis of features for a people migrating to Scarborough to search a best neighbourhood as a comparative analysis between neighbourhoods. The features include median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, and good management for emergency.