1. **Introduction**

Toronto 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 country, 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.

The aim of this project is to help people in exploring better facilities around their neighbourhood. It will help people making smart and efficient decision on selecting great neighbourhood out of numbers of other neighbourhoods in Toronto, Canada.

It will help people to get awareness of the area and neighbourhood before moving to a new city, state, country or place for their work or to start a new fresh life.

1. **Data**

Data Link: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

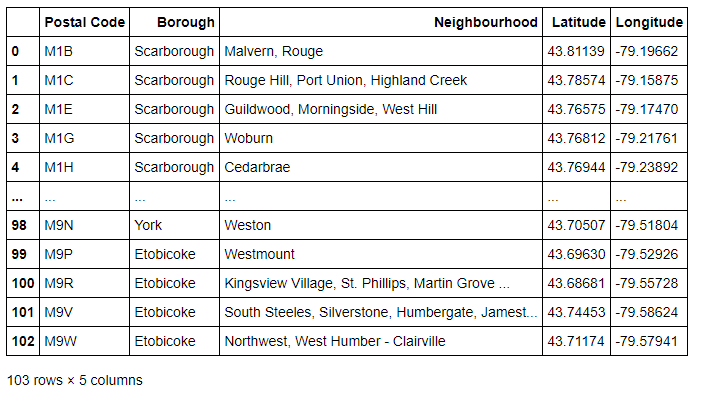
The data will be scraped from the link above. The dataset consists of latitude and longitude, zip codes.

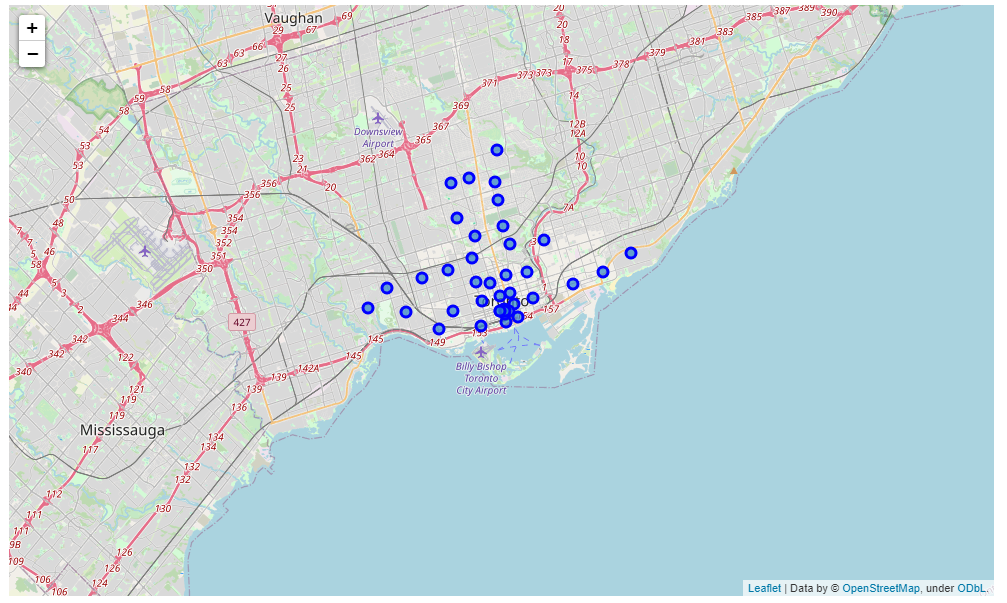
## Foursquare API Data:

Foursquare API is a platform where which we will be using to gain data about different venues in different neighbourhoods. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

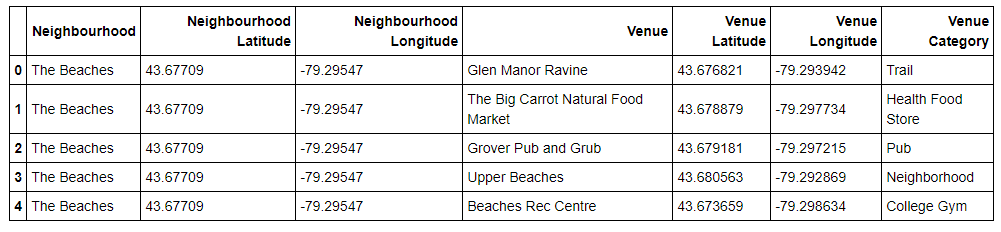
After finding the list of neighborhoods, we will then connect to the Foursquare API to gather information about venues in every neighbourhood.

1. **Methodology**

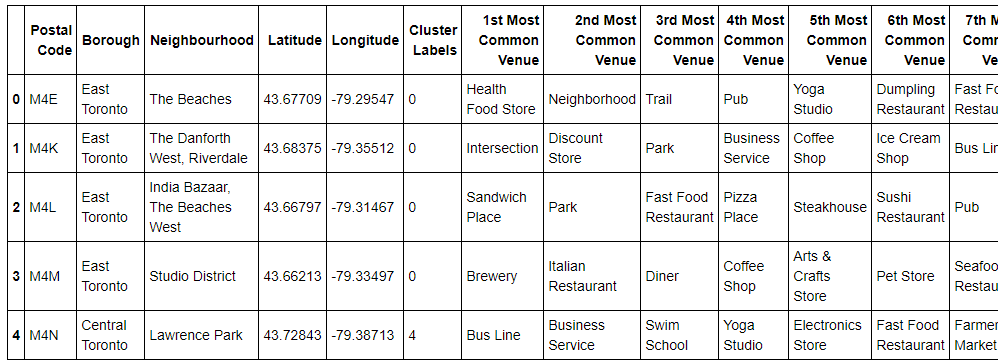
The data is first extracted from the above link mentioned above via web scraping. Then it is put into panda dataframe. Geocoder is used to collect latitude and longitude for each row in the data. 

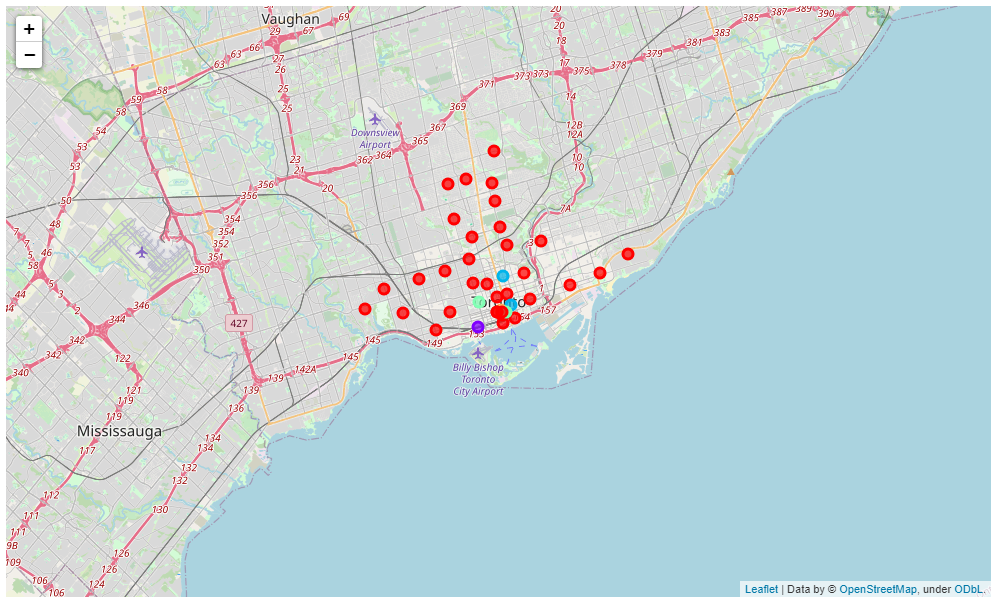
The neighbourhoods are visualized using Folium package. This is to ensure that the data returned by Geocoder are correctly plotted. 

Nearby venues around neighbourhoods are extracted using Foursquare API.



To compare the similarities of every neighbourhoods, k-means clustering algorithm is used based on the most common venues in each neigbourhood(frequency similarities). 5 cluster points are used for this project.



1. **Results**

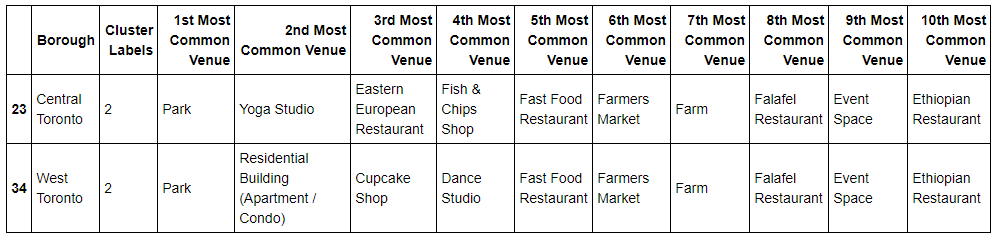
## Cluster 1:

## 

## Cluster 2:



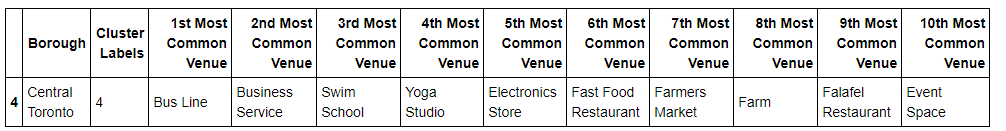
## Cluster 3:



## Cluster 4:



## Cluster 5:



**5) Discussion:**

Based on the results shown above, each of the clusters are segmented based on the similarity of ‘most common venue’. The major purpose of this project is to suggest a better neighborhood in a new city for the person who are shifting there. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

**6) Conclusion:**

In this project, using k-means cluster algorithm I separated the neighbourhood into 5 different clusters from the dataset, which have very-similar neighbourhoods around them.

Finally, this project provides recommendation to immigrants who are looking to find a suitable neighbourhood to live in. It also gives insights to business developers on which infrastructure is lacking in that particular neighbourhood.