**Final Report :**

**Capstone Project – The Battle of Neighborhoods (Finding a Better Place in Scarborough, Toronto)**

**Introduction:**

Lots of people are migrating to various states of Canada and needed lots of research for good housing prices and reputated schools for their children. This project is for those people who are looking for better neighborhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, etc.

The aim of this Project is to help people in exploring better facilities around their neighborhood. It will help people making decision on selecting great neighborhood out of numbers of other neighborhoods in Scarborough, Toranto.

This Project aim to create an analysis of features for a people migrating to Scarborough to search a best neighborhood as a comparative analysis between neighborhoods. The features include median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, good management for emergency, water resources.

It will help people to get awareness of the area and neighborhood before moving to a new city, state, country or place.

## Data Description:

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

Will use Scarborough dataset which we scrapped from wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes.

#### Foursquare API Data:

We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. 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 then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, 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 as follows:

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

## Map of Scarborough

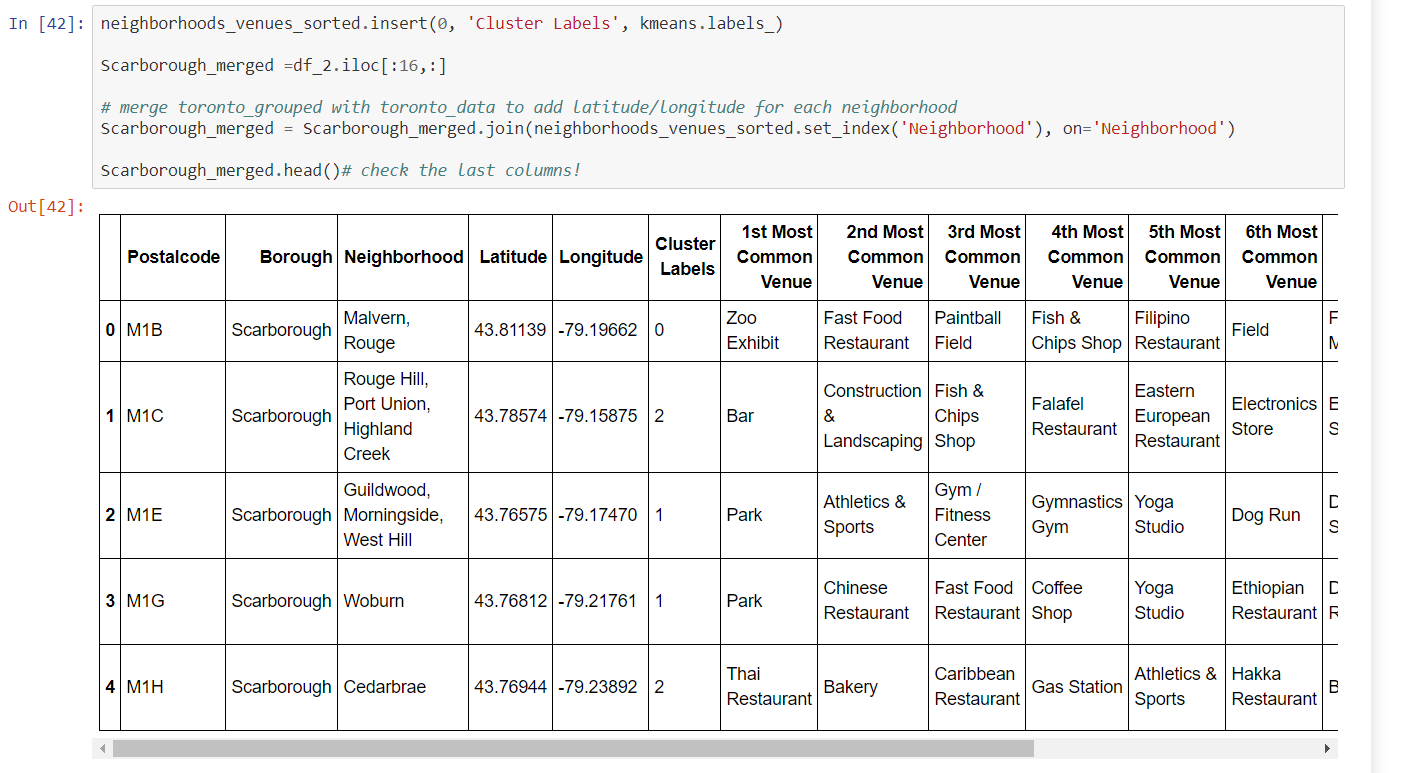
## C:\Users\Tanay\Desktop\1.png

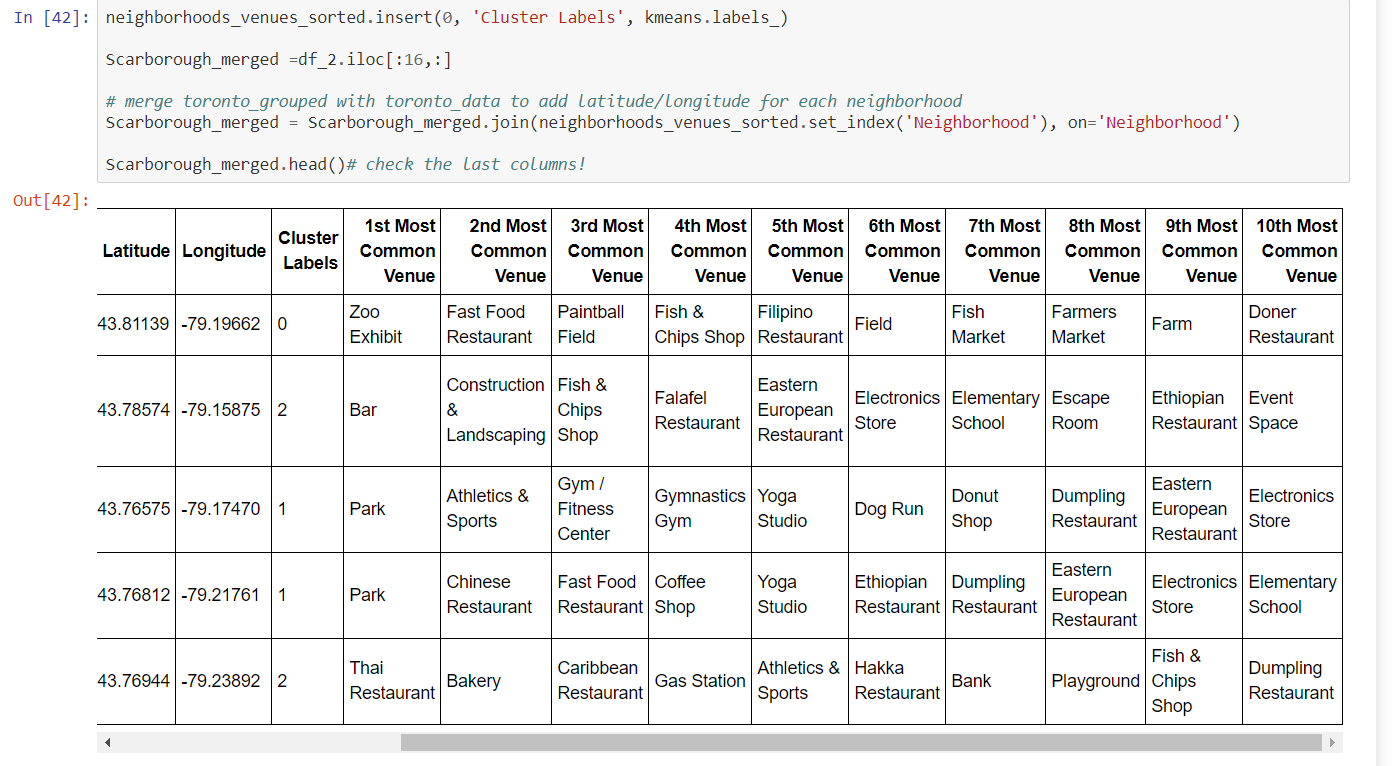
## Methodology Section:

#### Clustering Approach:

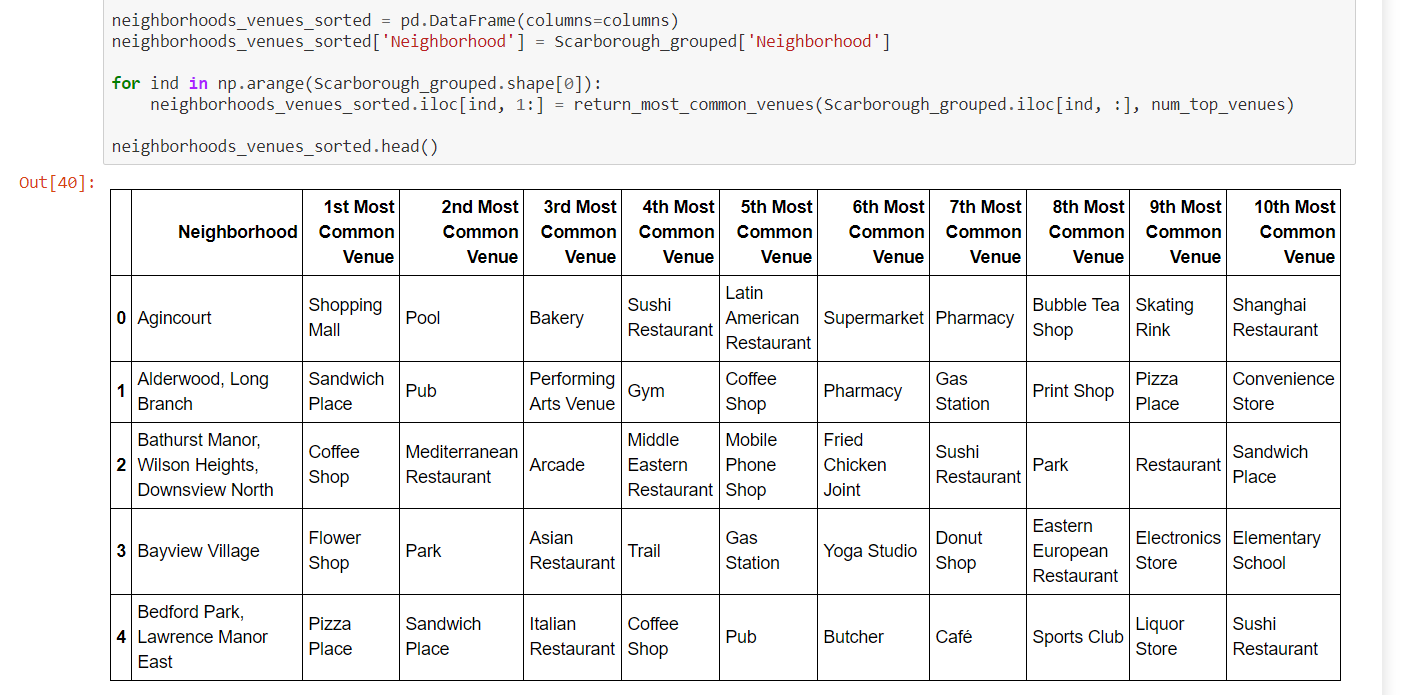
To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering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

#### Using K-Means Clustering Approach



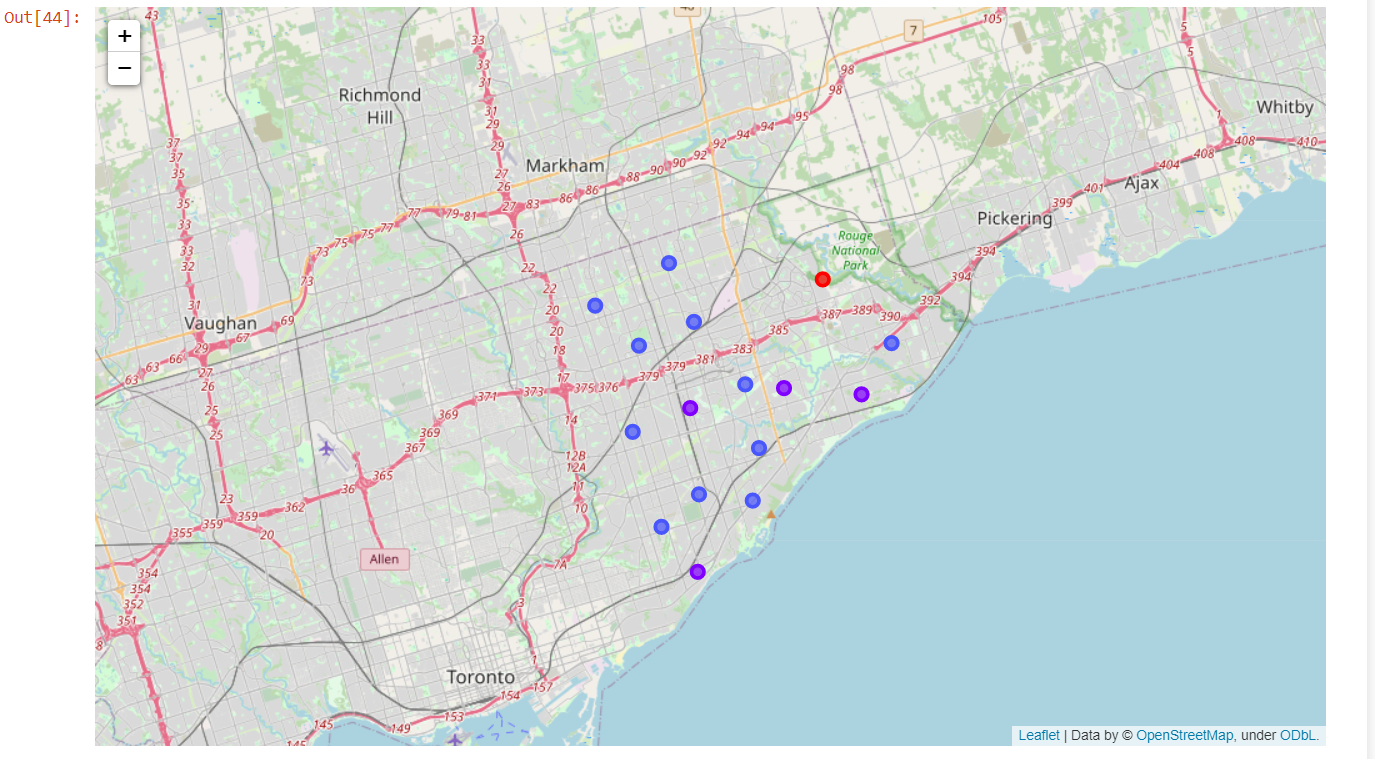


#### Most Common Venues:

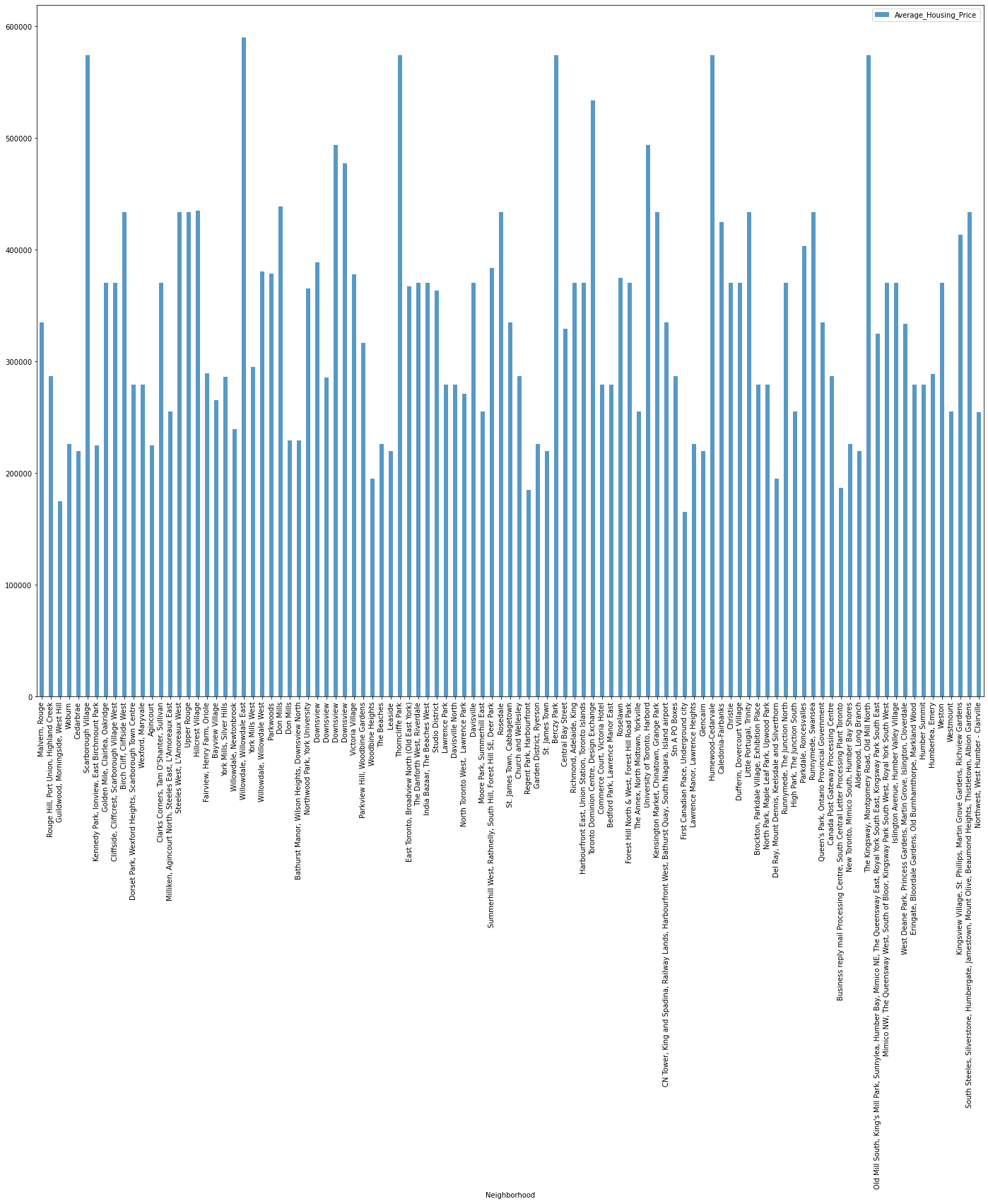


## Results:

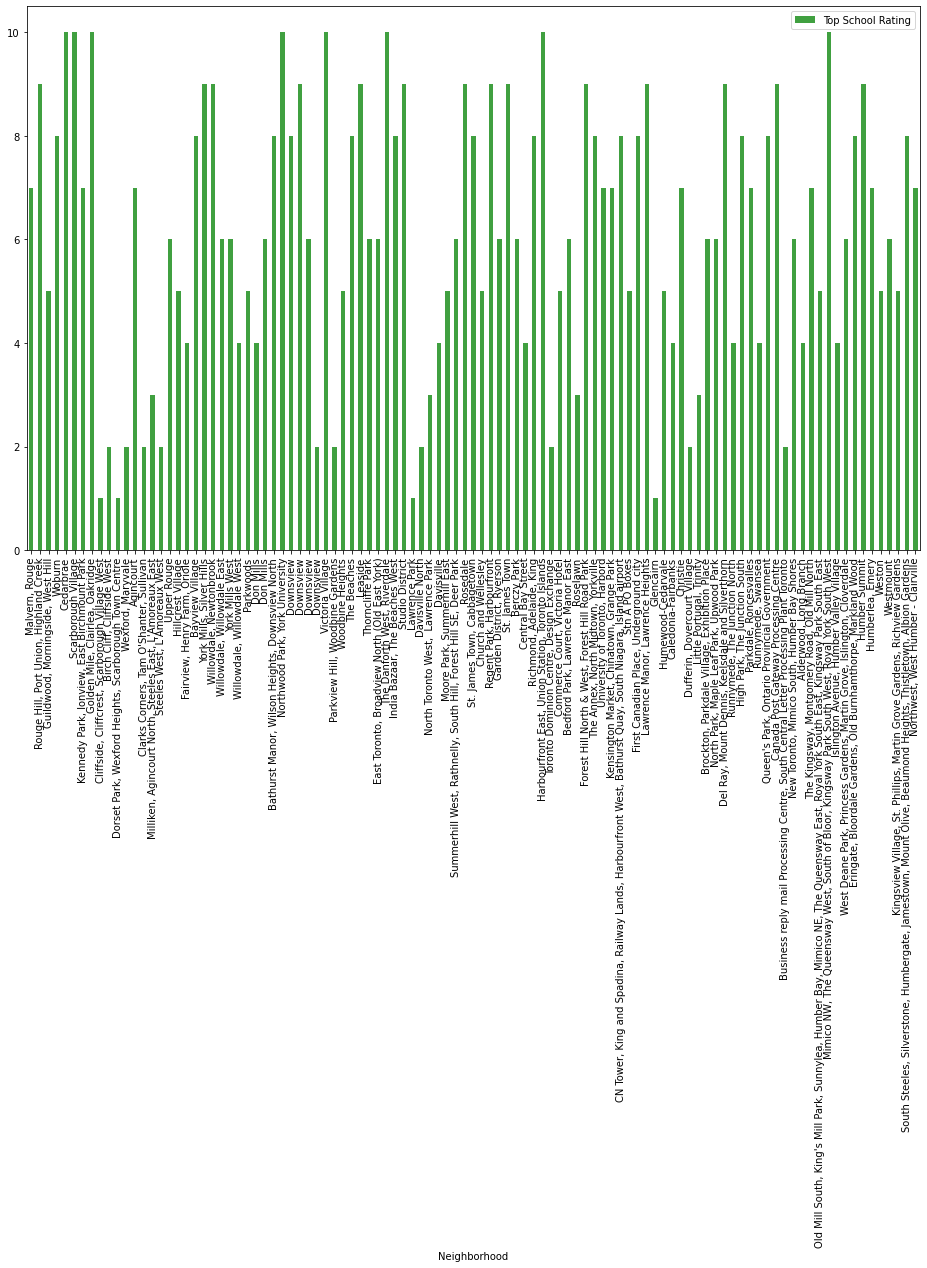
#### Map of Clusters in Scarborough



#### Average Housing Price by Clusters in Scarborough



#### School Ratings by Clusters in Scarborough



#### Location:

Scarborough is a popular destination for in Canada. As a result, it is one of the most diverse and multicultural areas in the Greater Toronto Area. 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.

#### Foursquare API:

This project would use Foursquare API as its prime data gathering source as it has a database of millions of places.

## Discussion Section:

Problem Which Tried to Solve: The major purpose of this project, is to suggest a better neighborhood in a new city. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

1. Sorted list of house in terms of housing prices in a ascending or descending order.
2. Sorted list of schools in terms of location, fees, rating and reviews.

## Conclusion:

In this project, using k-means cluster algorithm I separated the neighborhood into different clusters and for different lattitude and logitude from dataset. Using the charts above results presented to a particular neighborhood based on average house prices and school rating have been made.

Future Works: This project can be continued making it more precise in terms to find best house in Scarborough. Best means on the basis of all required things(daily needs or things we need to live a better life) around and also in terms of cost effective.

#### Libraries Which are Used to Develope the Project:

Pandas: For creating and manipulating dataframes.

Matplotlib: Python Plotting Module.

Scikit Learn: For importing k-means clustering.

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

JSON: Library to handle JSON files.

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

Geocoder: To retrieve Location Data.

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