**Final Report**

**Capstone Project – The Battle of Neighbourhoods Finding a Better Place in Scarborough, Toronto**

**1. Introduction:**

The main objective of this project is to help people in exploring better facilities around their neighbourhood. This helps in making proper, smart and efficient decision on selecting good neighbourhood out of other neighbourhoods that are present in Scarborough, Toronto.

There are a lot of people immigrating to Canada and they require to find good neighbourhood and also affordable houses and good educational institutes for education. This project helps in finding all these along with the ease of access to the facilities like grocery shops, medical store, hospital, parks and restaurants along with better community for living etc.

This Project aim to create 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, good management for emergency, water resources both fresh and waste water and excrement conveyed in sewers and recreational facilities.

This helps them with a better knowledge about the whole area and can find the required details about all the neighbourhood.

## 2. Data Section

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:**

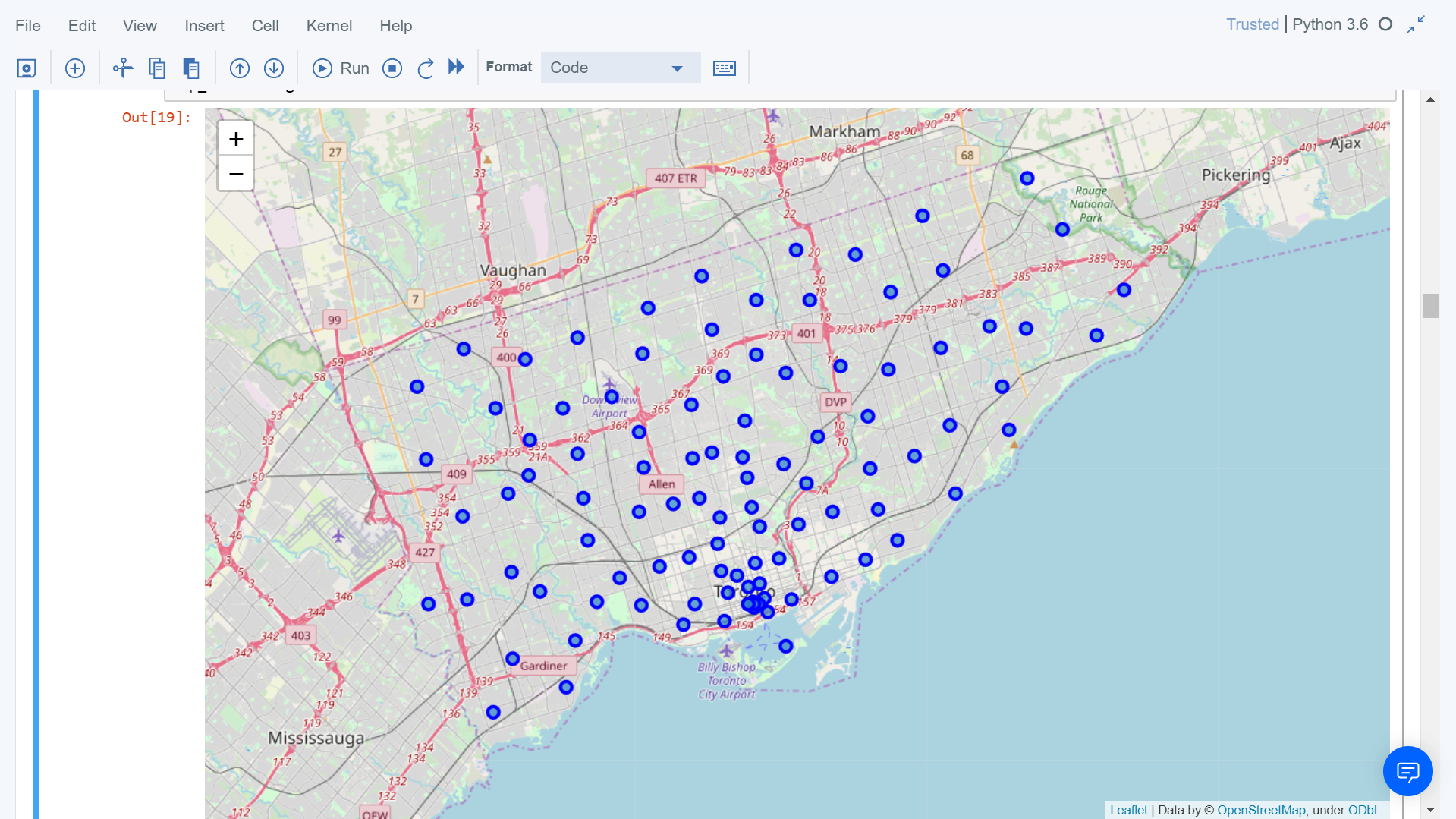
We will need data about different venues in different neighbourhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such 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 neighbourhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighbourhood. 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 as follows:

1. Neighbourhood
2. Neighbourhood Latitude
3. Neighbourhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

**Map of Scarborough**

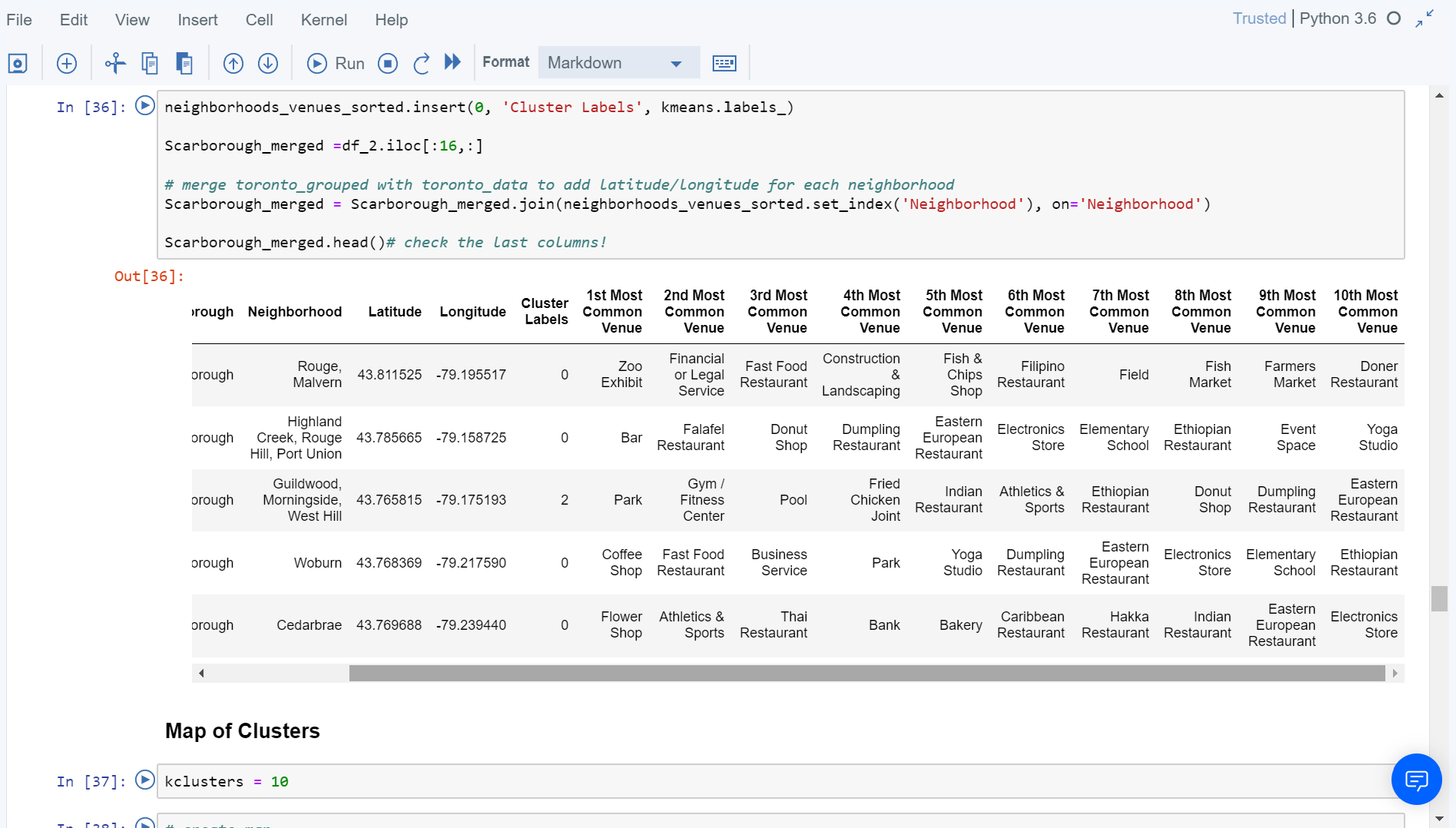


## 3. Methodology Section

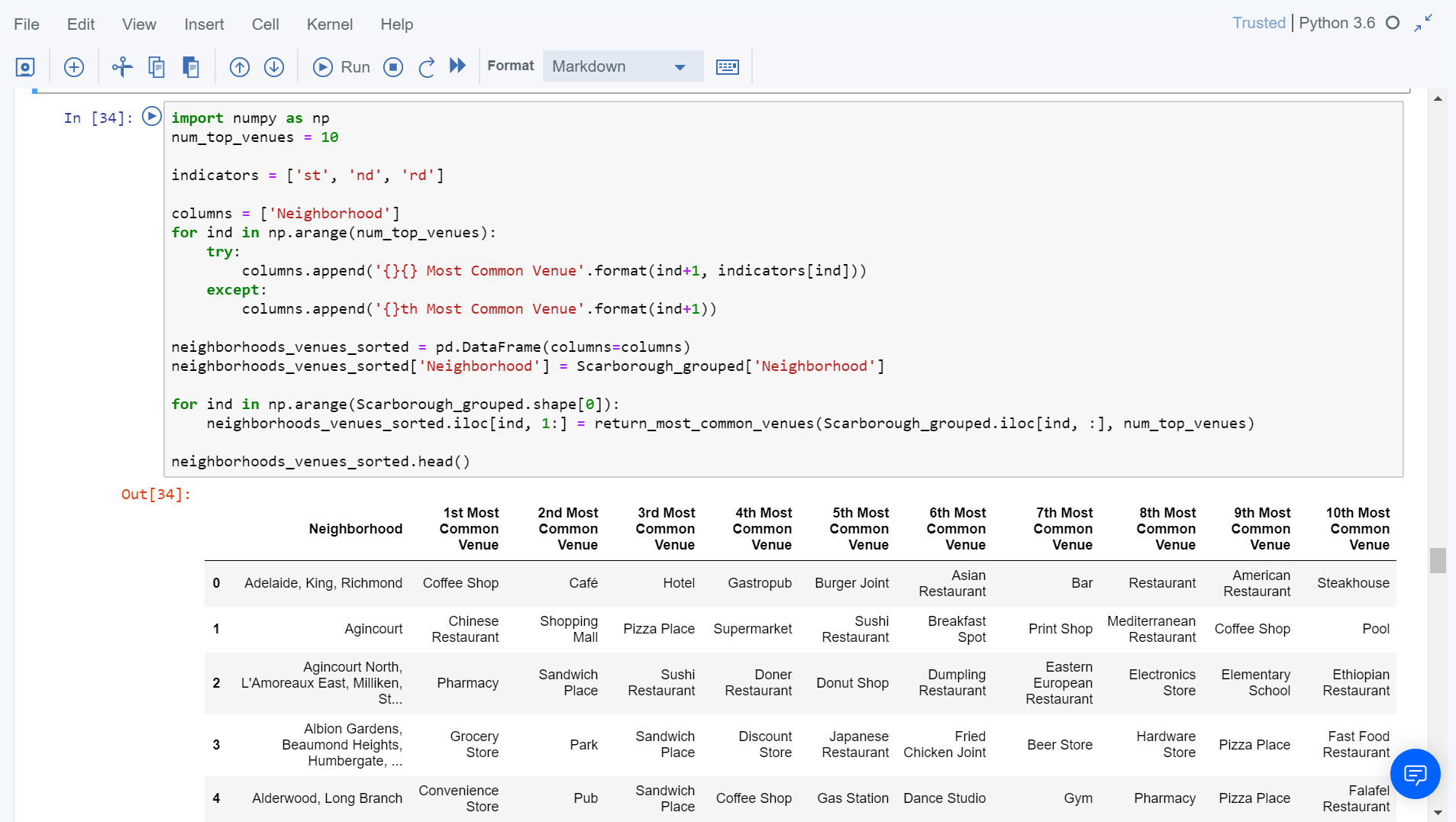
#### **Clustering Approach**:

To compare the similarities of two cities, we 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.

**Using K-Means Clustering Approach**

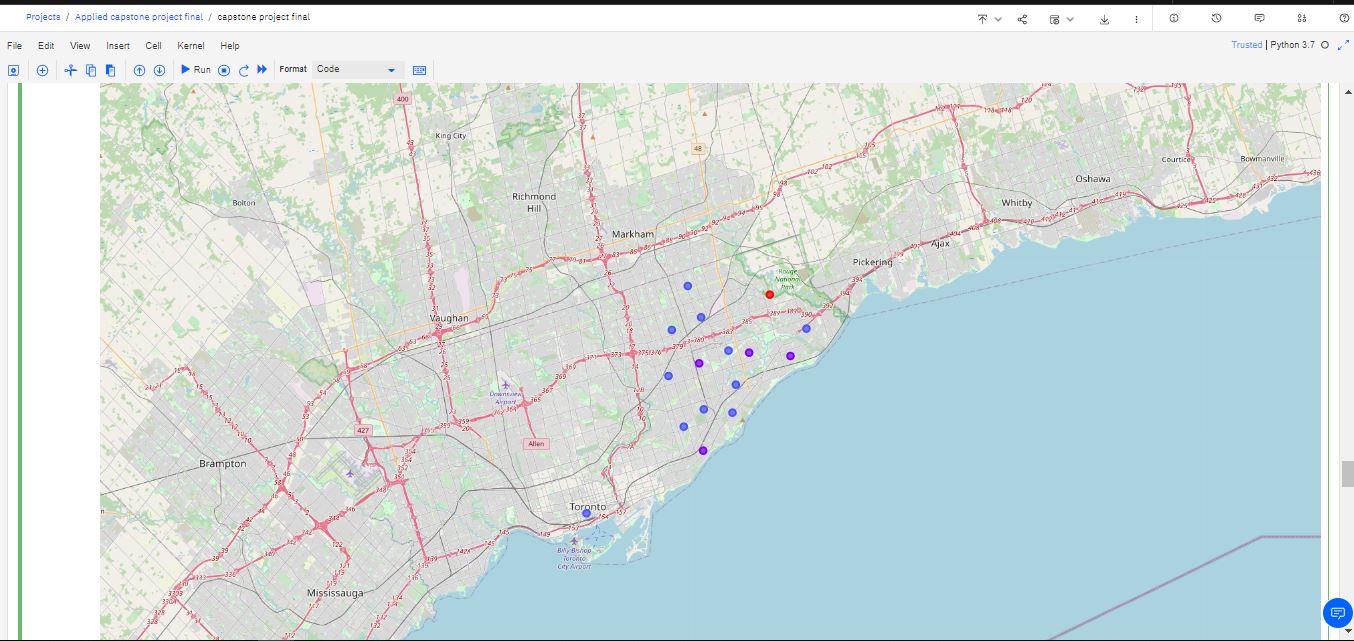


**Most Common venues near Neighborhood**

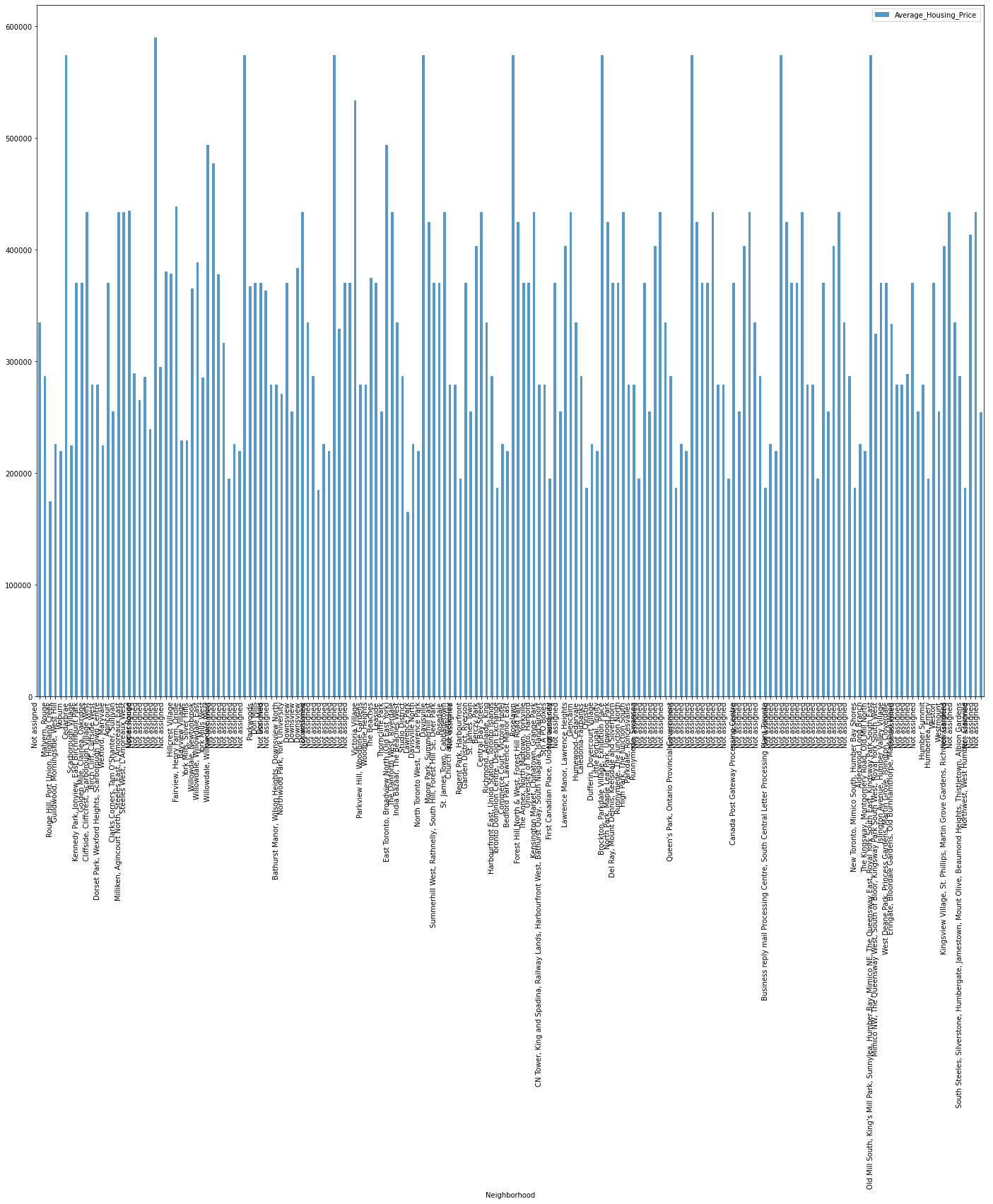


## 4. Results Section

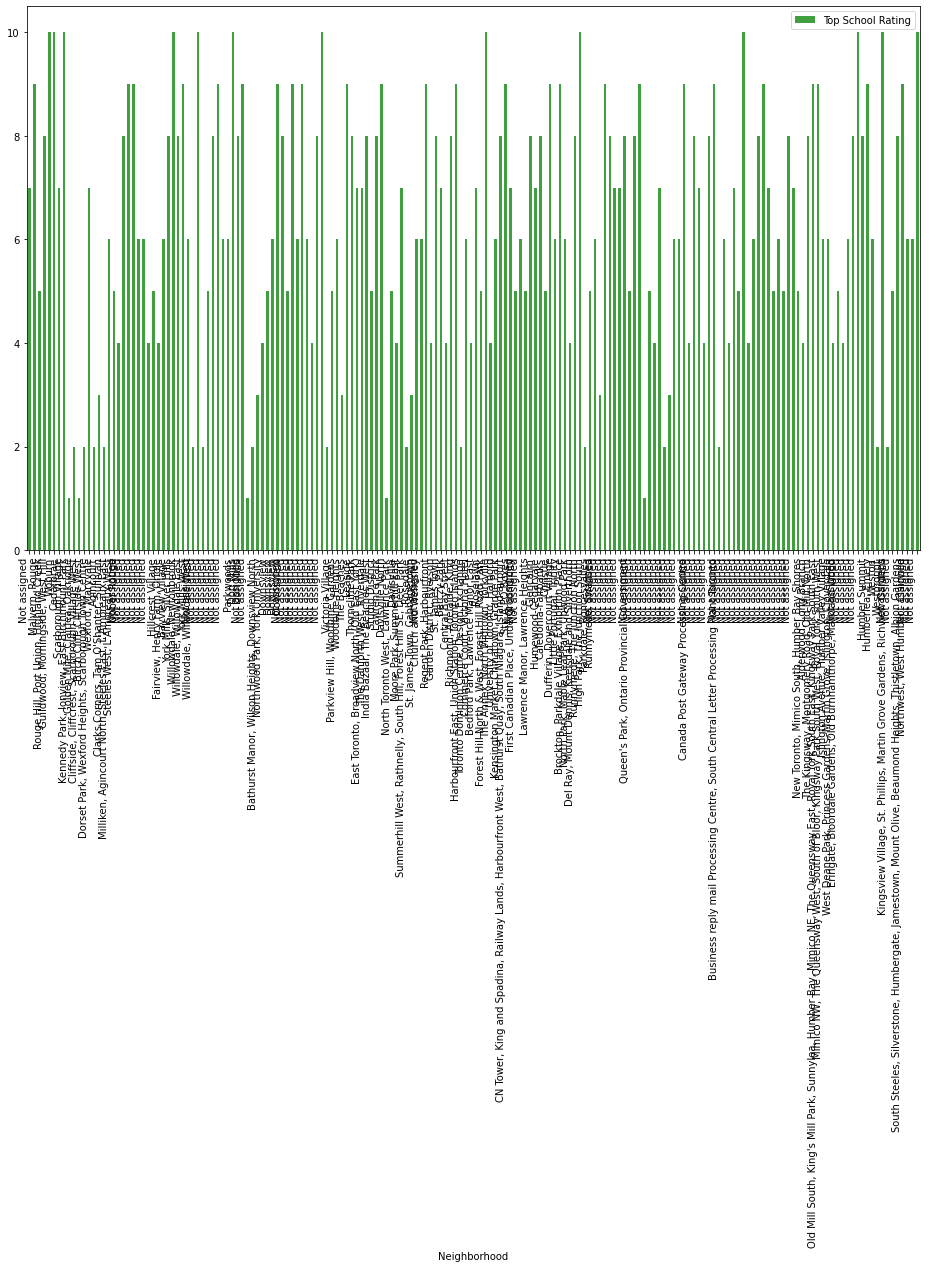
**Map of Clusters in Scarborough**



**Average Housing Price by Clusters in Scarborough**



**School Ratings by Clusters in Scarborough**



## Location

Scarborough is a popular destination for 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.

## Forsquare

This project would use Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

## 5. Discussion Section

This project objective is to help find the people with better neighbourhood, along with the neighbourhood with good community and ease of access to essential needs like grocery shops, school, market and transport facilities.

1.To sort the list of house in terms of housing prices in a ascending or descending order.

2.To sort the list of schools in terms of location, fees, rating and reviews.

## 6. Conclusion Section

In this project, using k-means cluster algorithm I separated the neighbourhood into 10(Ten) different clusters and for 103 different latitude and longitude from dataset, which have very-similar neighbourhoods around them. Using the charts above results presented to a particular neighbourhood based on average house prices and school rating have been made.

I feel rewarded with the efforts and believe this course with all the topics covered is well worthy of appreciation. This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools. The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.

#### **Future Works:**

This project can be continued for 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.

#### **List of libraries that are used in 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 neighbourhood 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.