# **Final Report | Capstone Project – The Battle of Neighborhoods**

# **Objective: For people planning to Shift, Finding a Better Place in Scarborough, Toronto.**

Introduction:

Project is suppose to help people in exploring better facilities around their neighborhood. It will help them take smart and efficient decisions on selecting great neighborhood in Scarborough, Toranto.

This Project aims to perform analysis of facilities for people migrating to Scarborough in search of a better neighborhood as a comparative analysis between neighborhoods. The facilities are median house pricing, better schools, crime rate, road connectivity, weather conditions, emergency help, water resources both fresh and waste water facilities.

Following are expected out of this project deliverable:

1. Sorted list of houses in terms of pricing

2. Sorted list of schools in terms of location, fees, rating and reviews

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:

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 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 will be as follows:

1. Neighborhood

2. Neighborhood Latitude

3. Neighborhood Longitude

4. Venue

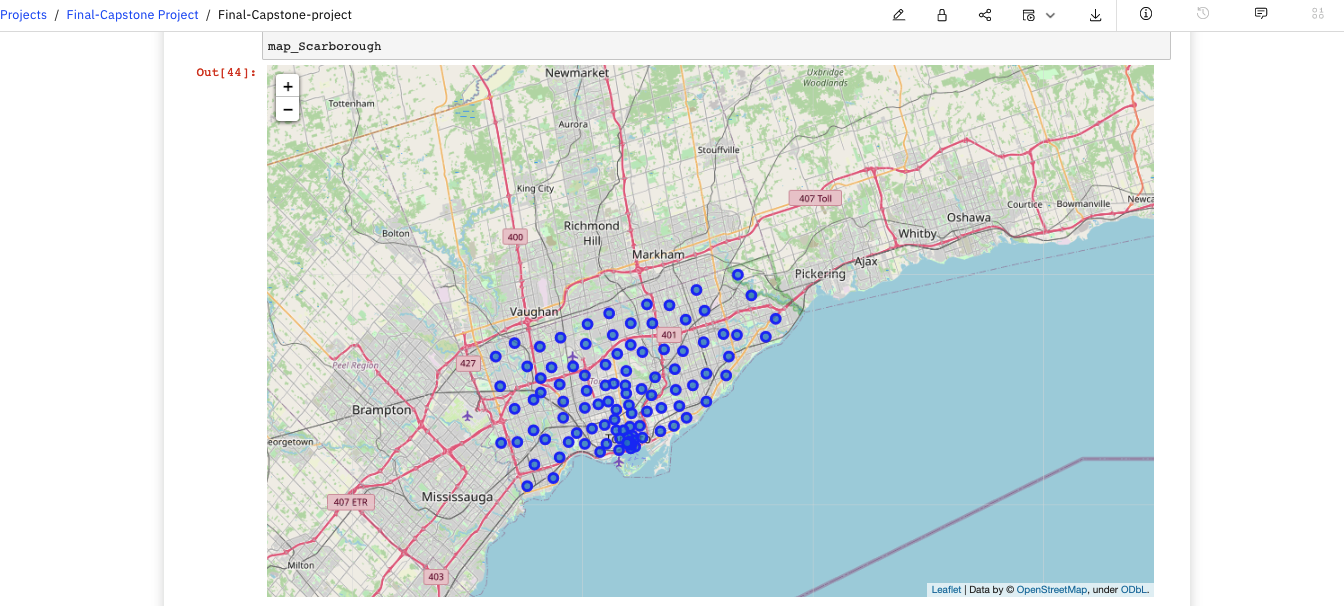
5. Name of the venue

6. Venue Latitude

7. Venue Longitude

8. Venue Category

****Scarborough Map****

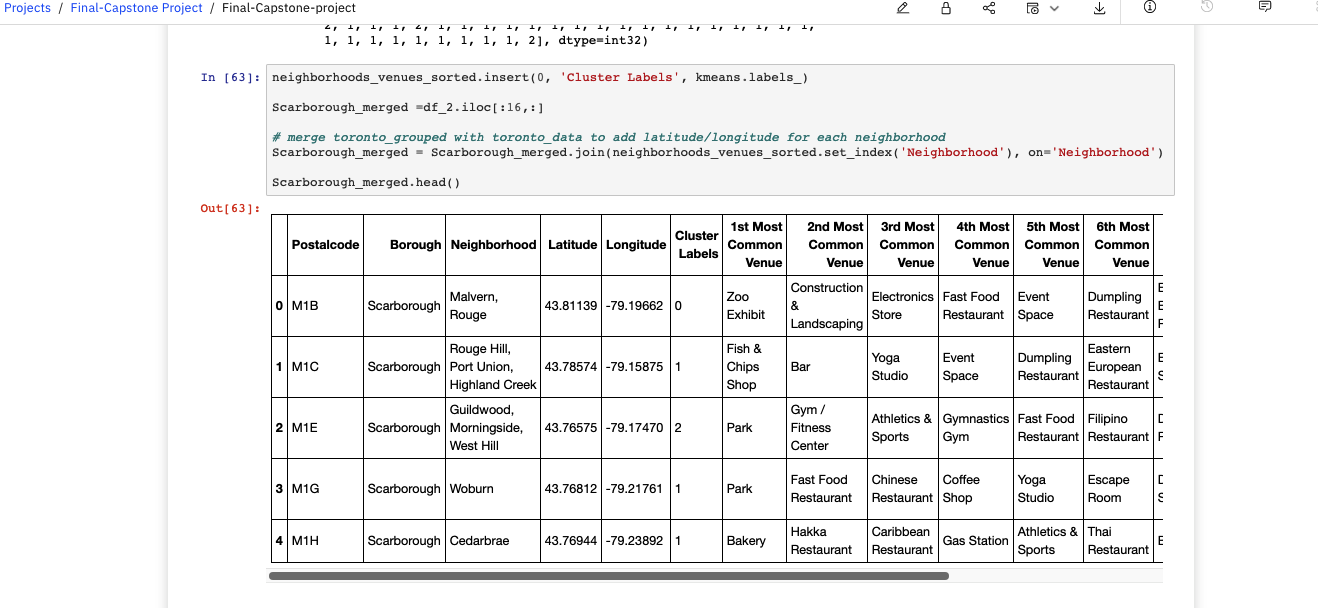


Methodology

Clustering approach:

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

K-means clustering approach:

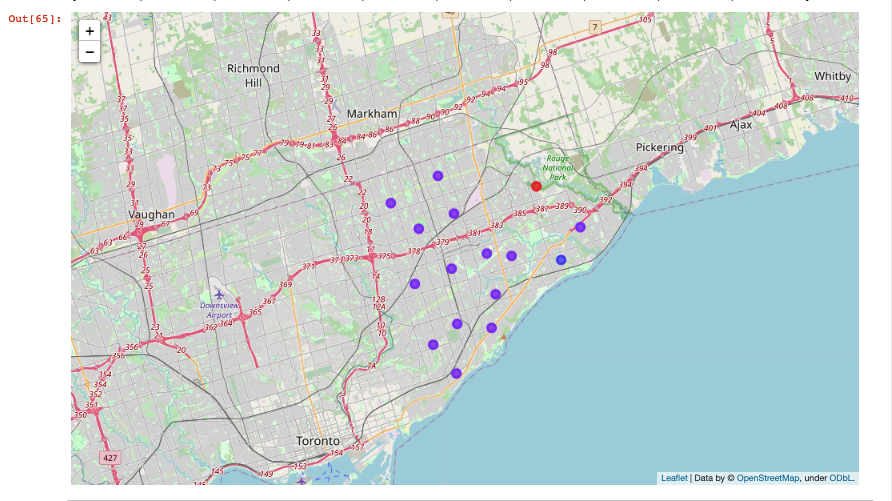


Most common values near neighborhood:

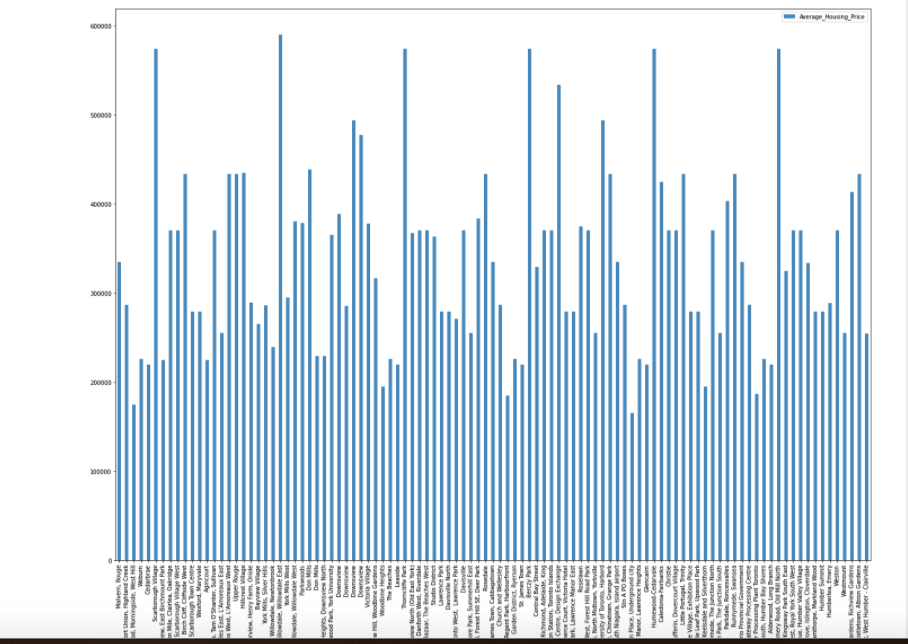


Results Section:

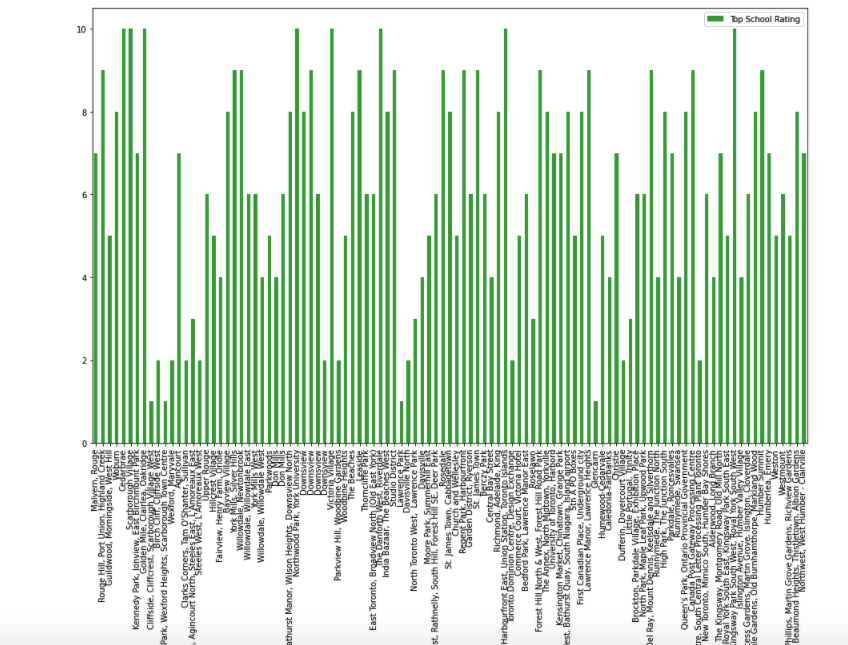
Map of clusters in Scarborough:



Average Housing price by cluster in Scarborough:



Scarborough school ratings by cluster



**Discussion section:**

Project is suppose to help people in exploring better facilities around their neighborhood. It will help them take smart and efficient decisions on selecting great neighborhood in Scarborough, Toranto.

This Project aims to perform analysis of facilities for people migrating to Scarborough in search of a better neighborhood as a comparative analysis between neighborhoods. The facilities are median house pricing, better schools, crime rate, road connectivity, weather conditions, emergency help, water resources both fresh and waste water facilities.

Following are expected out of this project deliverable:

1. Sorted list of houses in terms of pricing

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 10(Ten) different clusters and for 103 different lattitude and longitude from dataset, having very-similar neighborhoods around them. Using the charts above results presented to a particular neighborhood based on average house prices and school ratings.

This project can be continued to make it better for precise results.

**Libraries used to Develop the Project:**

Pandas: For creating and manipulating dataframes.

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

Scikit Learn: For importing k-means clustering.

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.

Matplotlib: Python Plotting Module.