Final Report

Capstone – The Battle of Neighbourhoods

Finding a Better Place in Scarborough, Toronto

1. Introduction:

The purpose of this Project is to help people to explore better facilities around their neighbourhood. It will also help people to make smart & efficient decision on selecting good neighbourhoods in Scarborough, Toronto.

Lots of people are migrating to various provinces in Canada and need lot of exploration to find better housing prices and good schools for their children. This project would help those who are looking for better neighbourhoods to accommodate themselves.

This Project aim to create an analysis of features for people relocating to Scarborough to search for a better neighbourhood as a comparative analysis between neighbourhoods.

The features Include:

- Median Housing Price
- Better Schools
- Lower Crime Rates
- Road Connectivity
- Water Resources
- Recreational Facilities

Data

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

Will use Scarborough dataset which we scrapped from Wikipedia. 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.

Once the list of neighbourhoods is found, we would then connect to the Foursquare API to gather information about venues inside each neighbourhood. For each neighbourhood, Chosen radius is 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

6. Venue Latitude

7. Venue Longitude

8. Venue Category

Map of Scarborough
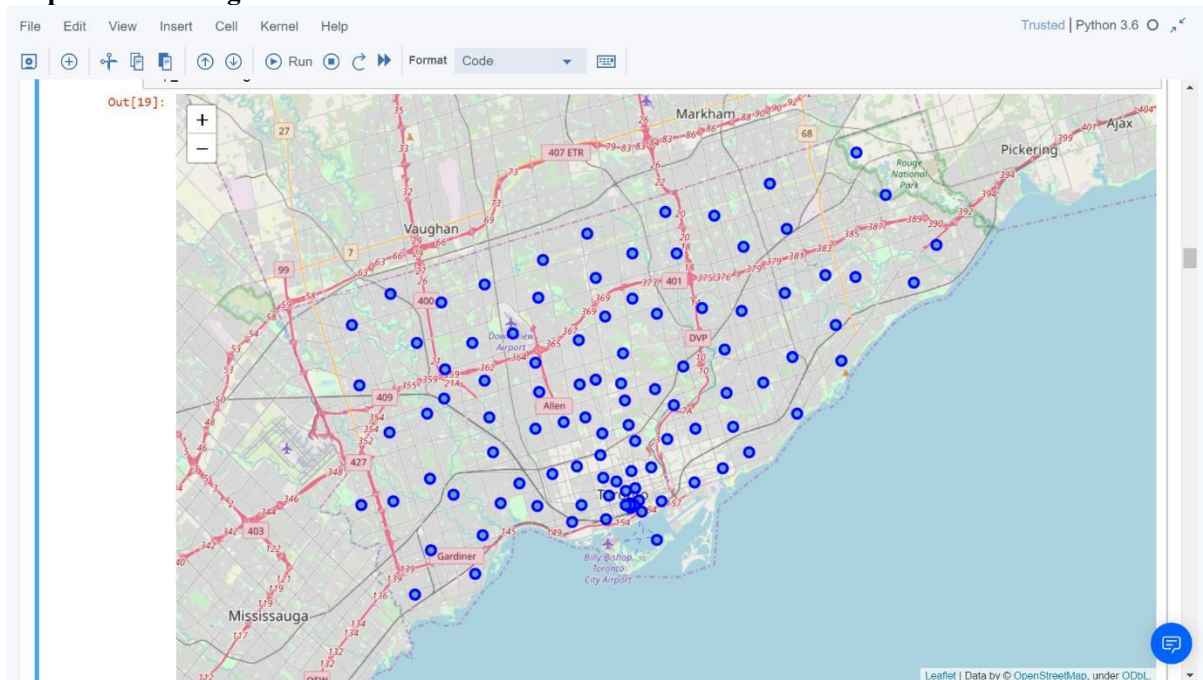
3. Methodology Section

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

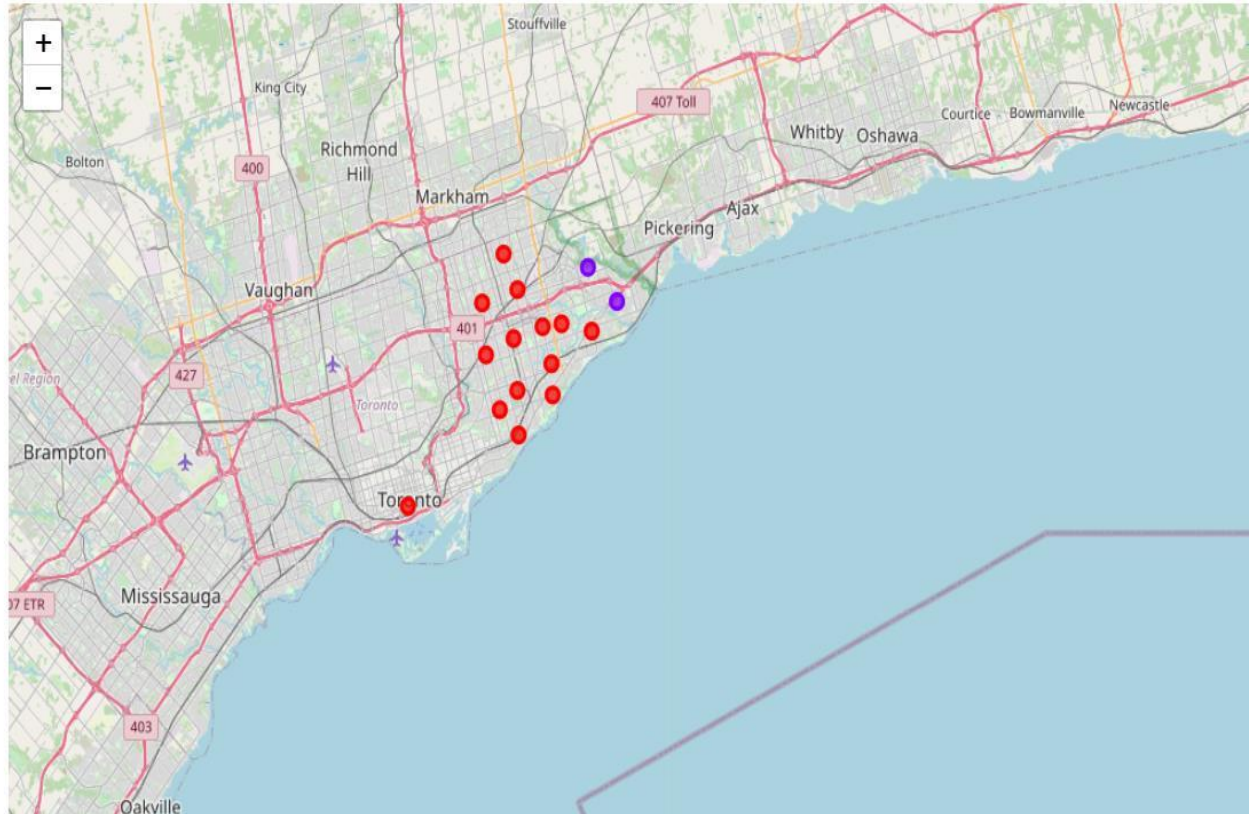**Most Common venues near Neighbourhood**

Work Flow:

Nearby places of neighbourhoods are mined using credentials of Foursquare API features. Number of places/neighbourhood parameter is set to 100 and the radius parameter is set to 500
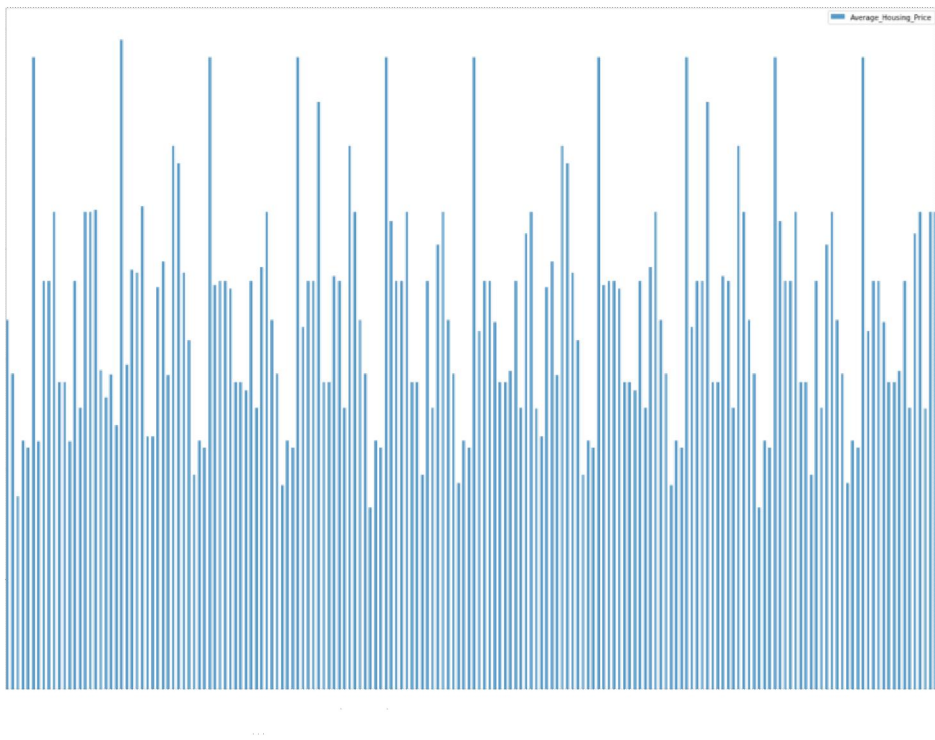
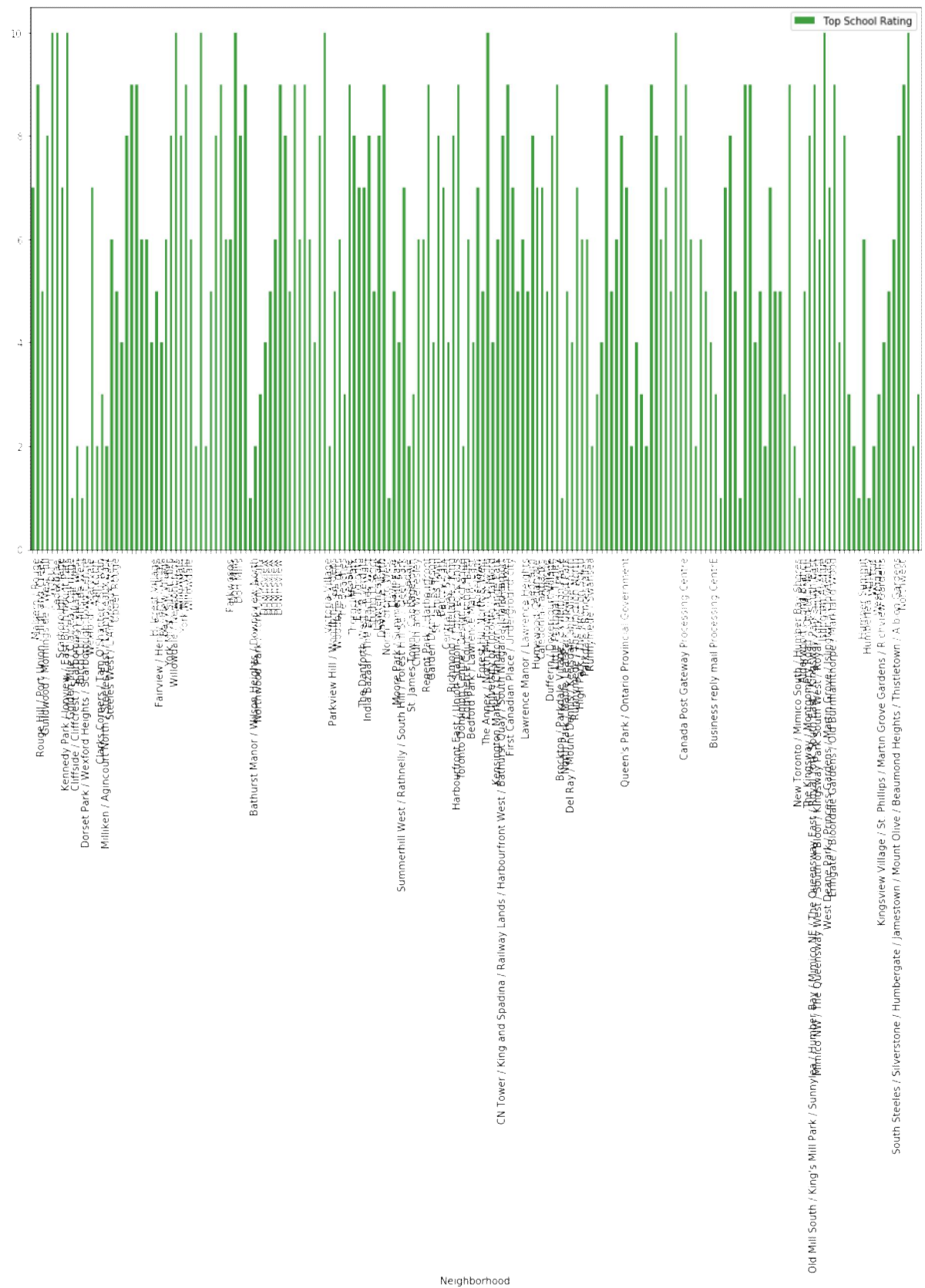# Results

Map of Clusters in Scarborough

[43]:

Average Housing Price by Clusters in Scarborough

School Ratings by Clusters in Scarborough

Legend: Top School Rating

X-axis: Neighborhood

The Location:

Scarborough is a popular destination for new immigrants in Canada to settle. As a result, it is one of the most diverse and multicultural areas in the Greater Toronto Area. It is home to various religious & ethnic groups and places of worships.

Foursquare API:

This project have used 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.

Discussion

The main purpose of this project is to suggest a better neighbourhood in a new city for someone to shift and settle there with better Connectivity to the airport, bus station, and easy access to markets and other essential things nearby.

1. List of house is sorted in terms of housing prices in an ascending or descending order
2. List of schools sorted in terms of location, fees, rating and reviews

Conclusion

In this project, using k-means cluster algorithm. Neighbourhood is separated into 10 different clusters and for 103 different latitude and longitude from the dataset, which have very-similar neighbourhoods around them.

Libraries used:

Pandas: For creating and manipulating data frames.

Folium: Python visualization library would be used to visualize the neighbourhoods 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.