BATTLE OF NEIGHBOURHOODS

PRESENTATION FINAL

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.

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:

- Neighbourhood

- Neighbourhood Latitude

- Neighbourhood Longitude

- Venue

- Name of the venue e.g. the name of a store or restaurant

- Venue Latitude

- Venue Longitude

- Venue Category

MAP OF SCARBOROUGH
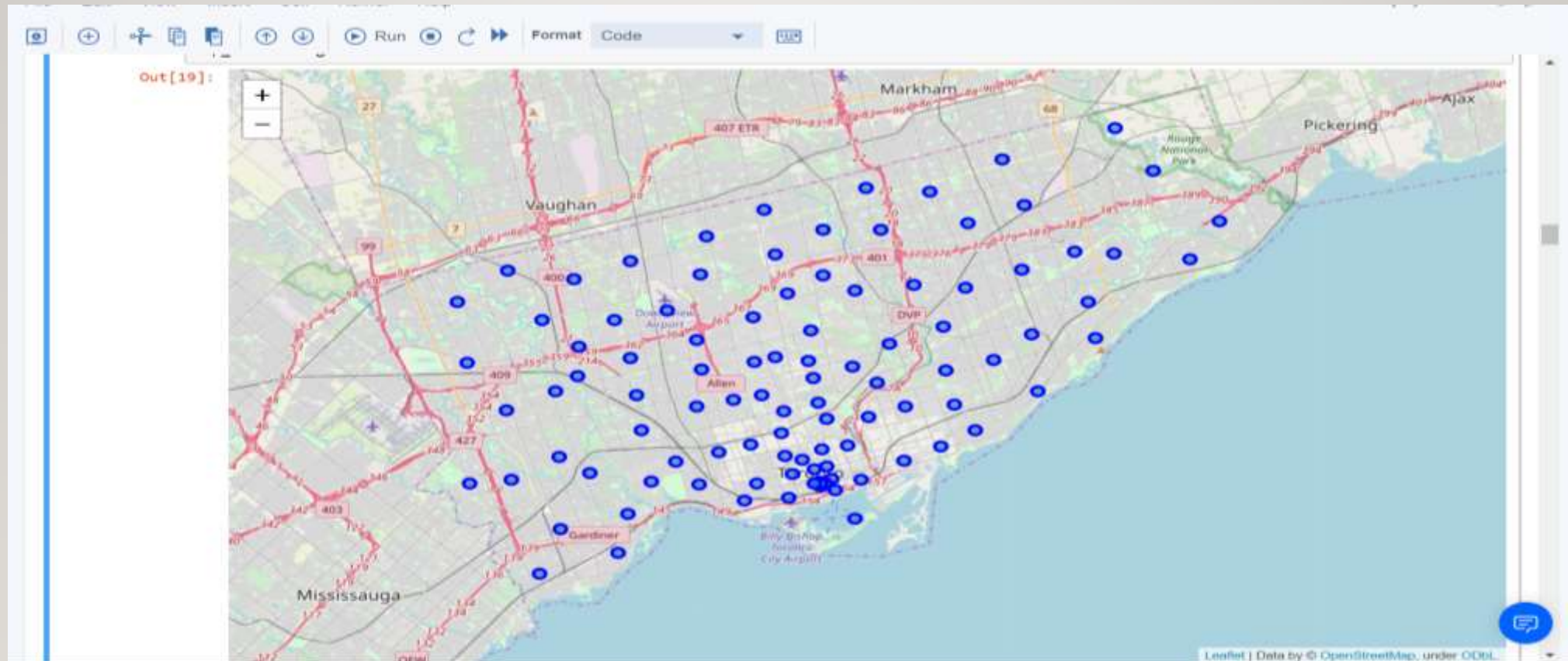
METHODOLOGY SECTION

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

```
In [36]:  ▶  neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

          Scarborough_merged =df_2.iloc[:16,:]

          # merge toronto_grouped with toronto_data to add Latitude/Longitude for each neighborhood
          Scarborough_merged = Scarborough_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

          Scarborough_merged.head()# check the last columns!
```

Out[36]:

rough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
rough	Rouge, Malvern	43.811525	-79.195517	0	Zoo Exhibit	Financial or Legal Service	Fast Food Restaurant	Construction & Landscaping	Fish & Chips Shop	Filipino Restaurant	Field	Fish Market	Farmers Market	Doner Restaurant
rough	Highland Creek, Rouge Hill, Port Union	43.785665	-79.158725	0	Bar	Falafel Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant	Event Space	Yoga Studio
rough	Guildwood, Morningside, West Hill	43.765815	-79.175193	2	Park	Gym / Fitness Center	Pool	Fried Chicken Joint	Indian Restaurant	Athletics & Sports	Ethiopian Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant
rough	Woburn	43.768369	-79.217590	0	Coffee Shop	Fast Food Restaurant	Business Service	Park	Yoga Studio	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant
rough	Cedarbrae	43.769688	-79.239440	0	Flower Shop	Athletics & Sports	Thai Restaurant	Bank	Bakery	Caribbean Restaurant	Hakka Restaurant	Indian Restaurant	Eastern European Restaurant	Electronics Store

MOST COMMON VENUES NEAR NEIGHBOURHOOD


```python
import numpy as np
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = Scarborough_grouped['Neighborhood']

for ind in np.arange(Scarborough_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(Scarborough_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

Out[34]:

|	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Café	Hotel	Gastropub	Burger Joint	Asian Restaurant	Bar	Restaurant	American Restaurant	Steakhouse
1	Agincourt	Chinese Restaurant	Shopping Mall	Pizza Place	Supermarket	Sushi Restaurant	Breakfast Spot	Print Shop	Mediterranean Restaurant	Coffee Shop	Pool
2	Agincourt North, L'Amoreaux East, Milliken, St...	Pharmacy	Sandwich Place	Sushi Restaurant	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant
3	Albion Gardens, Beaumond Heights, Humbergate, ...	Grocery Store	Park	Sandwich Place	Discount Store	Japanese Restaurant	Fried Chicken Joint	Beer Store	Hardware Store	Pizza Place	Fast Food Restaurant
4	Alderwood, Long Branch	Convenience Store	Pub	Sandwich Place	Coffee Shop	Gas Station	Dance Studio	Gym	Pharmacy	Pizza Place	Falafel Restaurant

RESULTS SECTION

Map of clusters in Scarborough
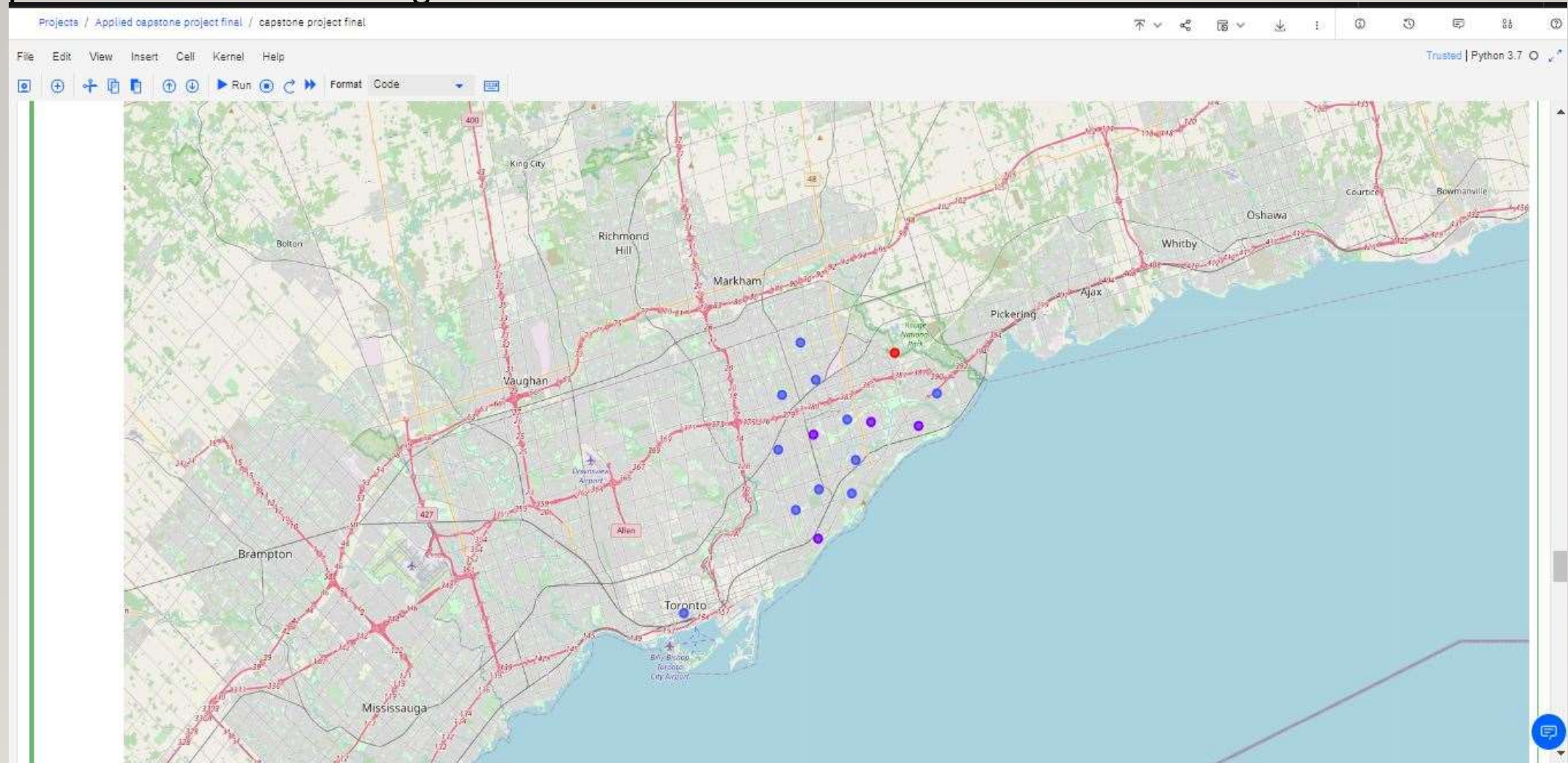
AVERAGE HOUSING PRICE BY CLUSTERS IN SCARBOROUGH
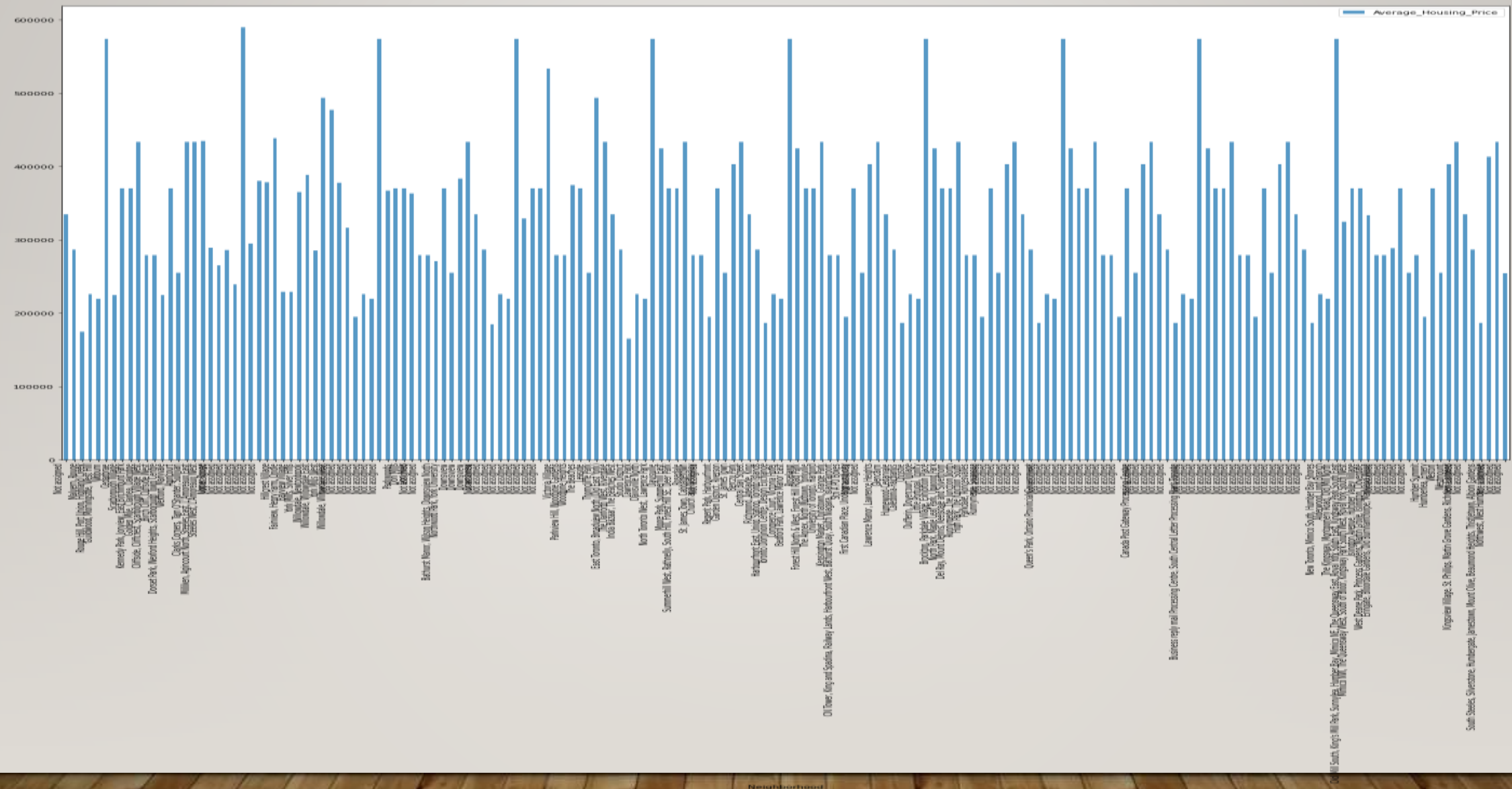
SCHOOL RATINGS BY CLUSTERS IN SCARBOROUGH
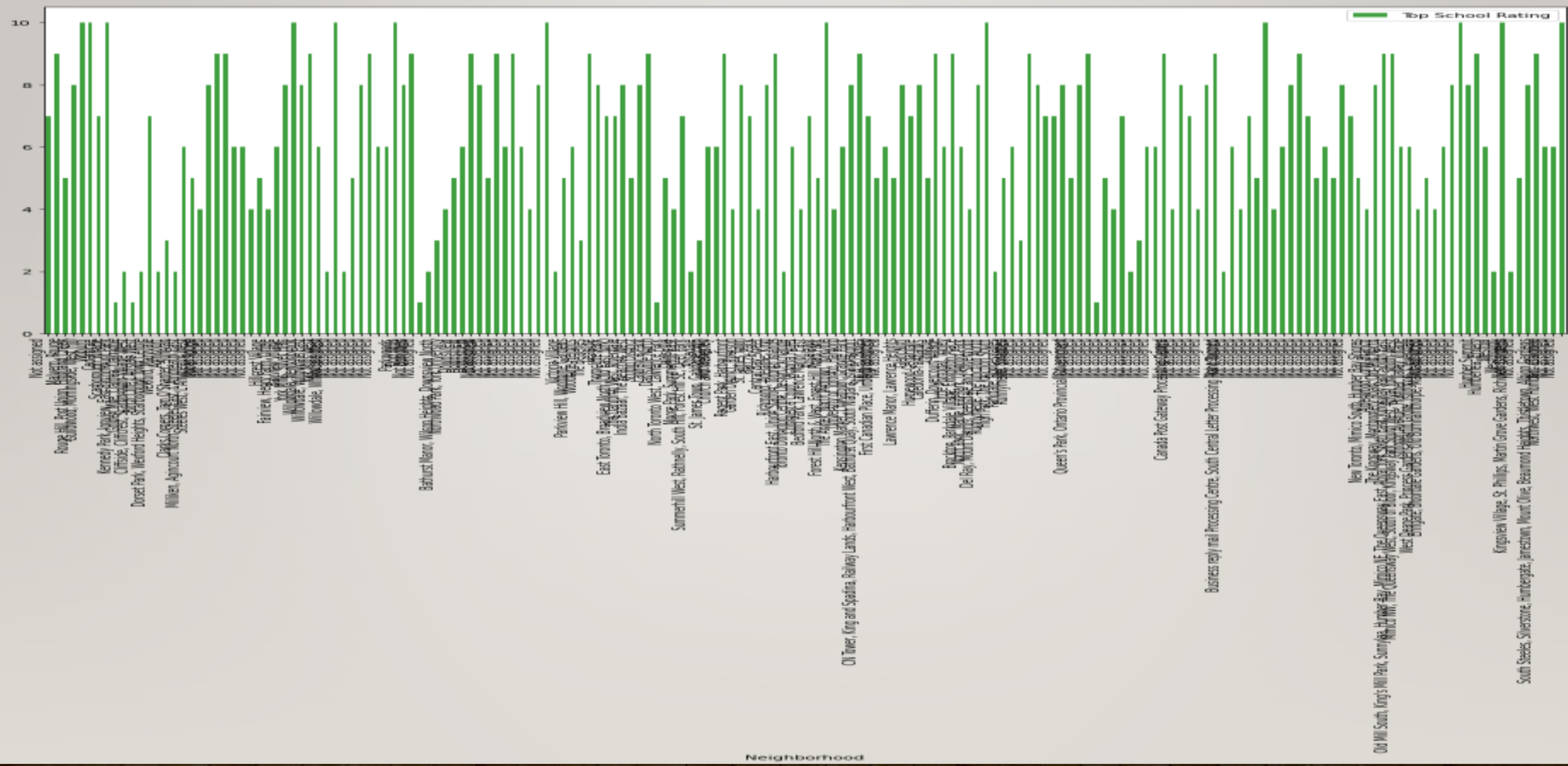
Location

- 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.

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.

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.

- 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.

- 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.