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.

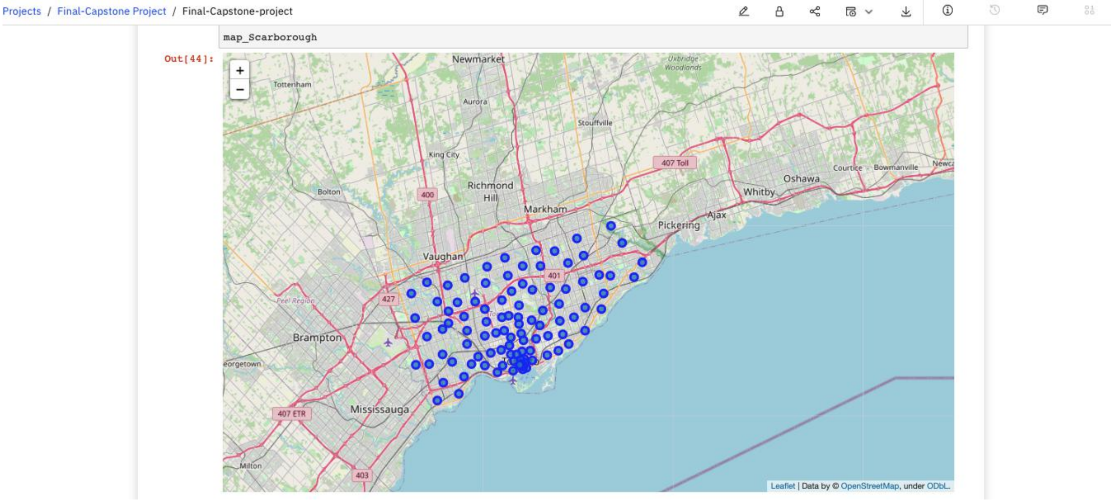
Foursquare API Data:
Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since 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



Projects / Final-Capstone Project / Final-Capstone-project

```
map_Scarborough
```

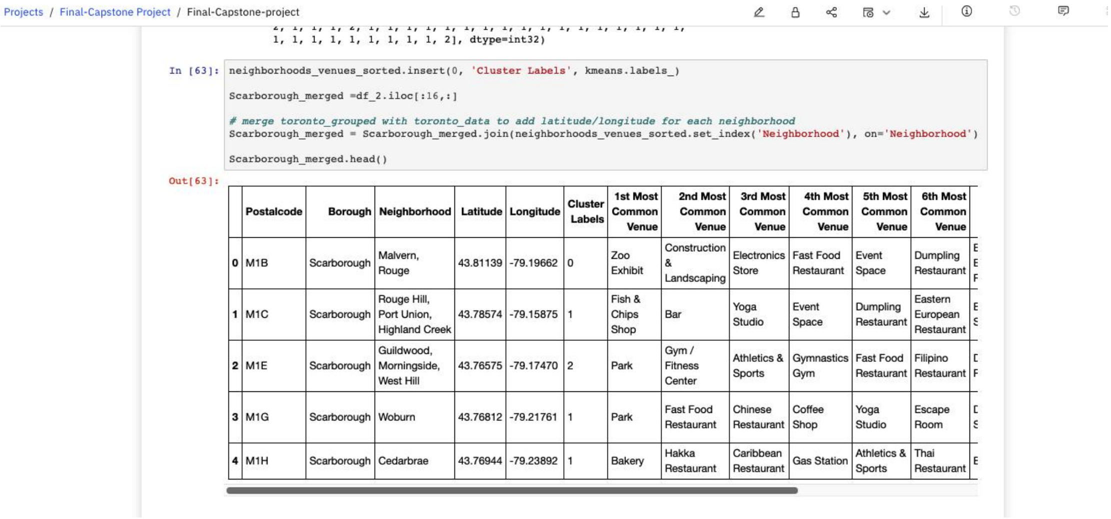
Out[44]:

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:



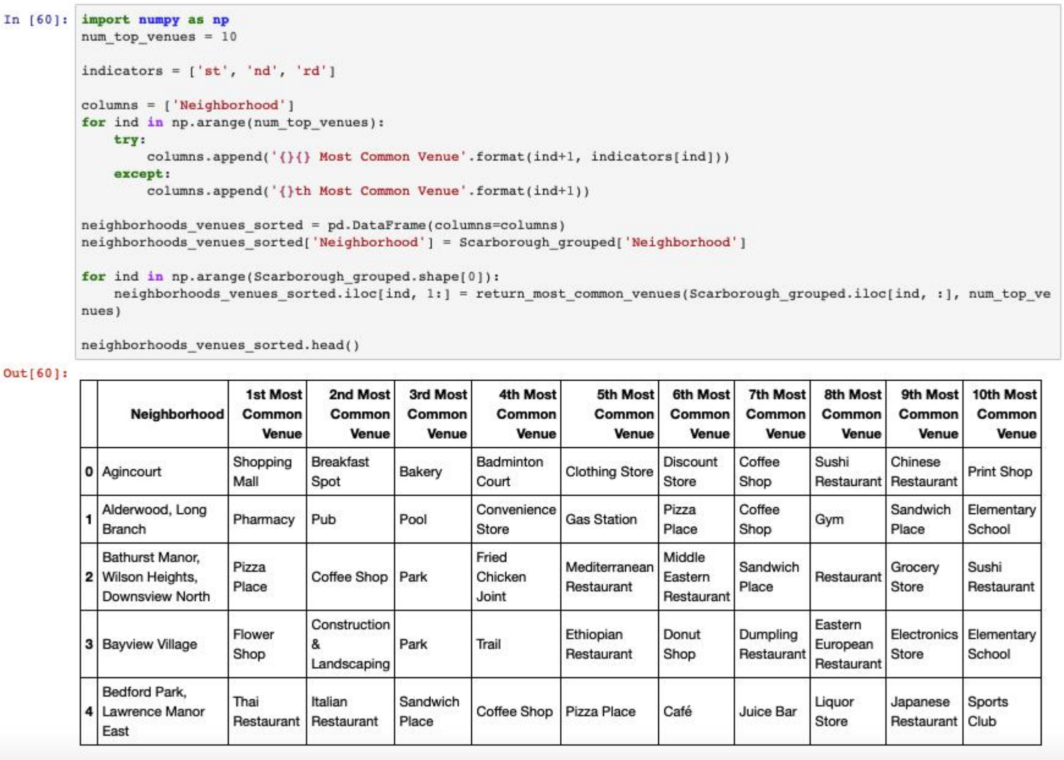
```
         2, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
         1, 1, 1, 1, 1, 1, 1, 1, 1, 2], dtype=int32)
```

```
In [63]: 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()
```

Out[63]:

	Postalcode	Borough	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
0	M1B	Scarborough	Malvern, Rouge	43.81139	-79.19662	0	Zoo Exhibit	Construction & Landscaping	Electronics Store	Fast Food Restaurant	Event Space	Dumpling Restaurant
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78574	-79.15875	1	Fish & Chips Shop	Bar	Yoga Studio	Event Space	Dumpling Restaurant	Eastern European Restaurant
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.76575	-79.17470	2	Park	Gym / Fitness Center	Athletics & Sports	Gymnastics Gym	Fast Food Restaurant	Filipino Restaurant
3	M1G	Scarborough	Woburn	43.76812	-79.21761	1	Park	Fast Food Restaurant	Chinese Restaurant	Coffee Shop	Yoga Studio	Escape Room
4	M1H	Scarborough	Cedarbrae	43.76944	-79.23892	1	Bakery	Hakka Restaurant	Caribbean Restaurant	Gas Station	Athletics & Sports	Thai Restaurant

Most common values near neighborhood:



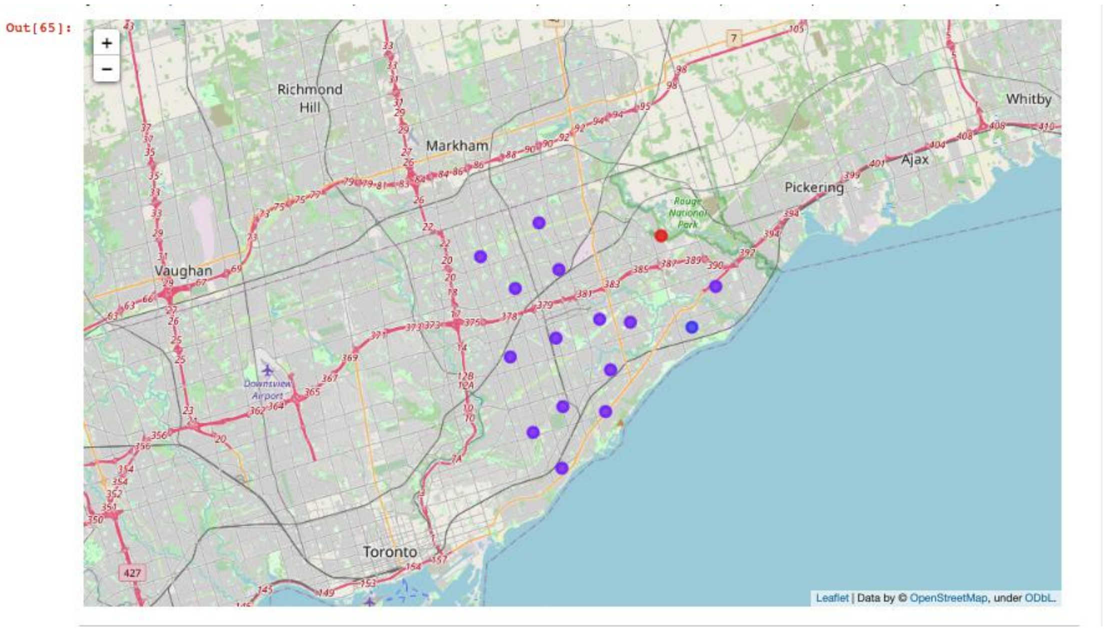
```
In [60]: import numpy as np
         num_top_venues = 10

         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_ve
         nues)

         neighborhoods_venues_sorted.head()
```

Out[60]:

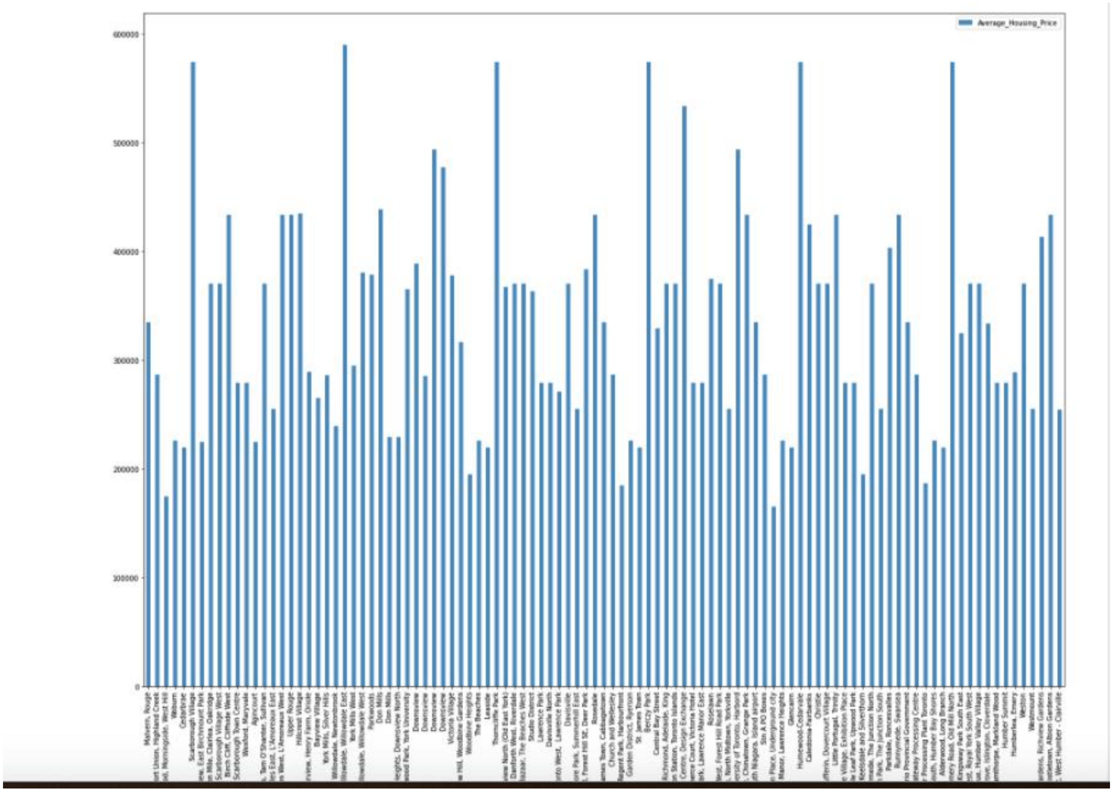
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Shopping Mall	Breakfast Spot	Bakery	Badminton Court	Clothing Store	Discount Store	Coffee Shop	Sushi Restaurant	Chinese Restaurant	Print Shop
1	Alderwood, Long Branch	Pharmacy	Pub	Pool	Convenience Store	Gas Station	Pizza Place	Coffee Shop	Gym	Sandwich Place	Elementary School
2	Bathurst Manor, Wilson Heights, Downsview North	Pizza Place	Coffee Shop	Park	Fried Chicken Joint	Mediterranean Restaurant	Middle Eastern Restaurant	Sandwich Place	Restaurant	Grocery Store	Sushi Restaurant
3	Bayview Village	Flower Shop	Construction & Landscaping	Park	Trail	Ethiopian Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School
4	Bedford Park, Lawrence Manor East	Thai Restaurant	Italian Restaurant	Sandwich Place	Coffee Shop	Pizza Place	Café	Juice Bar	Liquor Store	Japanese Restaurant	Sports Club

Results Section:

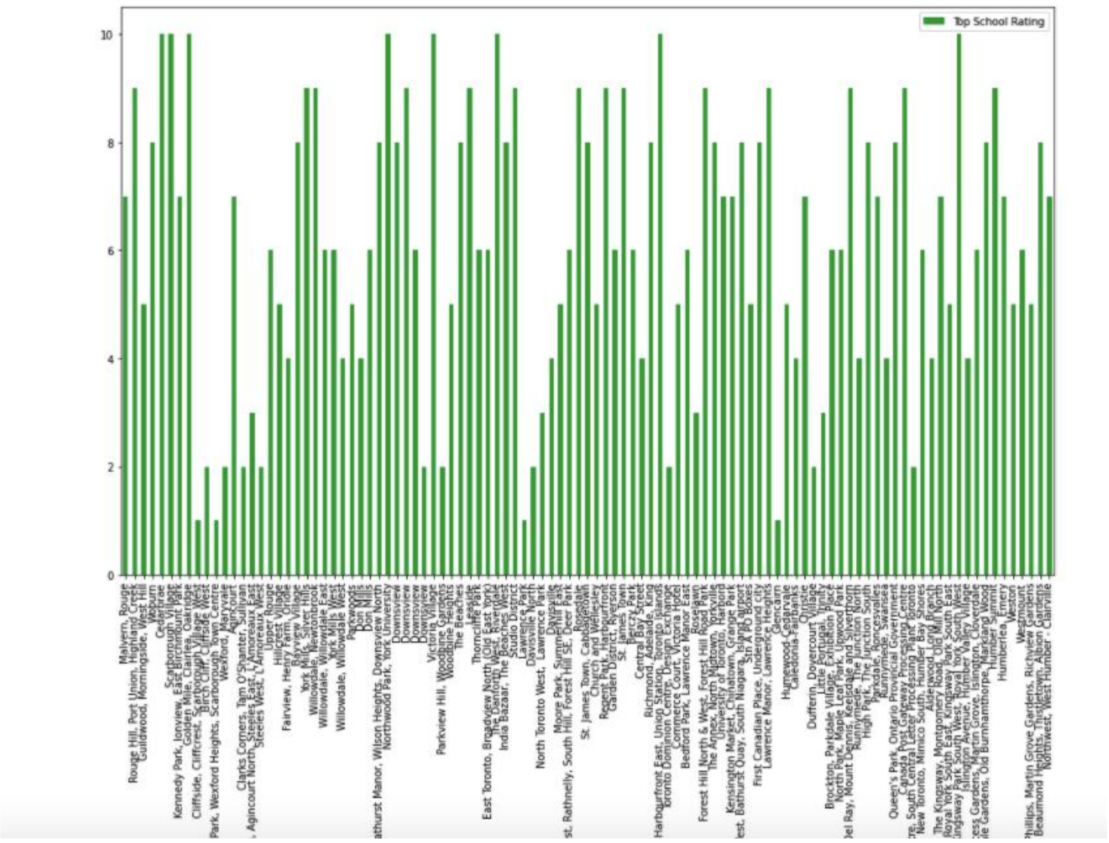
Map of clusters in Scarborough:



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.