# Capstone Project – The Battle of Neighborhoods

## Exploring the new neighborhood

# Awareness of the area and neighborhood to start a new fresh life

The purpose of this Project is to help people in exploring better facilities around their neighborhood. It will help people making smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in Scarborough, Toronto.

This Project aim to

- create an analysis of features to search a best neighborhood as a comparative analysis between neighborhoods.

The features

- median housing price,
- 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)
- excrement conveyed in sewers and recreational facilities.

# Data Extraction

**Foursquare API Data:**

- Data about different venues in different neighborhoods of that specific borough - use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest.
- 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.
- 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 as follows:
1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood 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

Maps of Scarborough

# Methodology Section



Using K-Means clustering Approach

# Most common venues near Neighborhood

File    Edit    View    Insert    Cell    Kernel    Help

Format   Markdown

```python
In [34]: 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 |

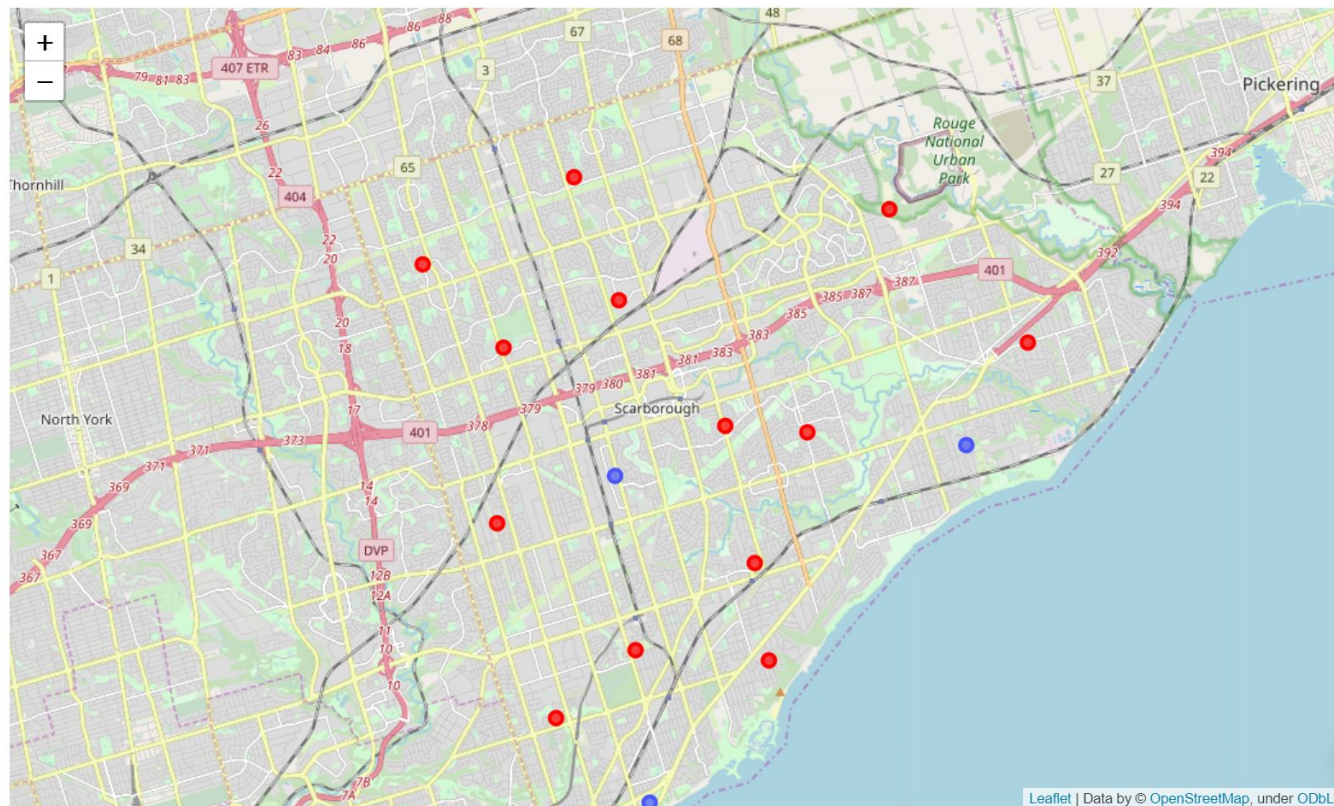# Result – Average housing prices
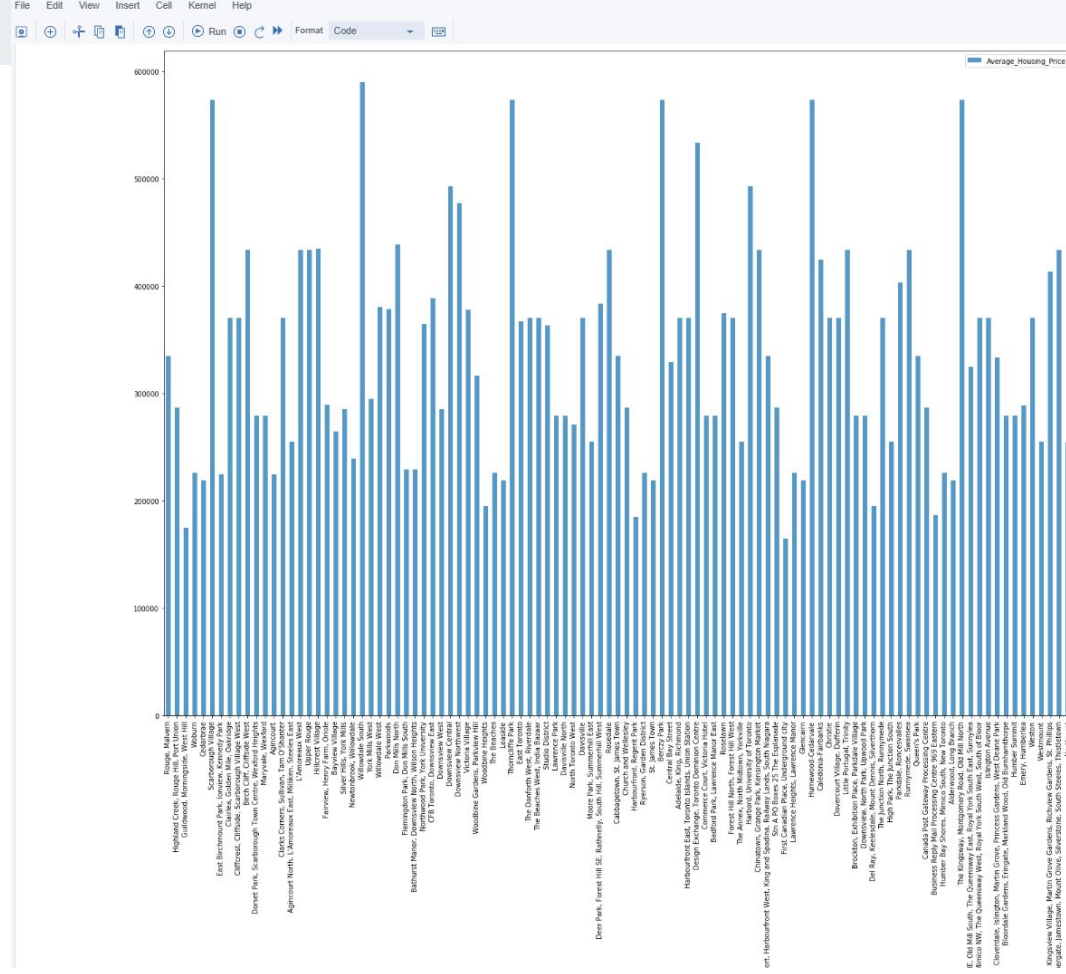
**School Ratings by clusters in Scarborough**
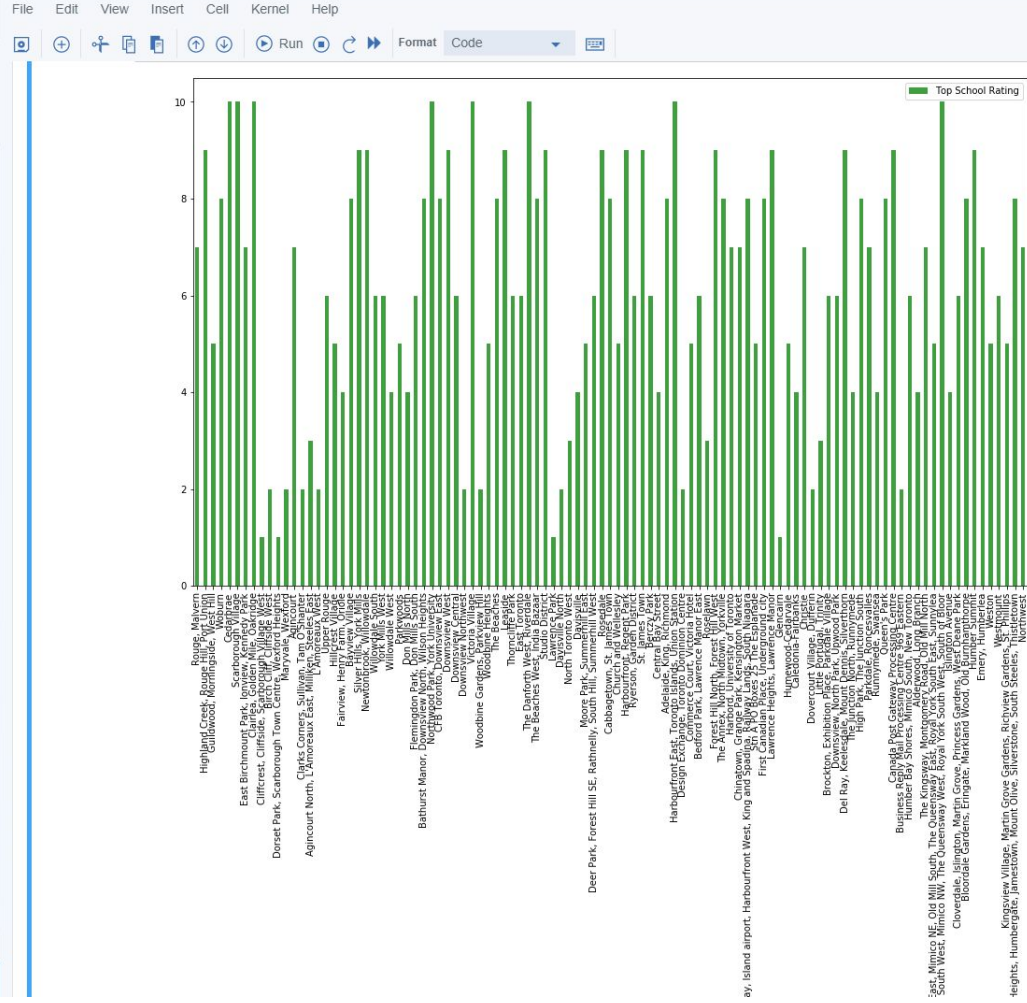
# Maps of clusters Scarborough

Format   Markdown

Out[38]:

# Conclusion

In this project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 103 different latitude and longitude from dataset, which have very-similar neighborhoods around them. Using the charts above results presented to a particular neighborhood 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.

# Libraries Used

- 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.