

## 1. Introduction:

This capstone project's goal is to assist locals in discovering better facilities in their community. It will assist individuals in choosing a wonderful neighbourhood in Scarborough, Toronto from among a multitude of other neighbourhoods in a shrewd and effective manner.

Many individuals are moving to other parts of Canada, so they had to do a lot of research to find affordable homes and reputable schools for their kids. Those looking for better neighbourhoods are the target audience for this project. For convenience of access to cafes, schools, supermarkets, pharmacies, grocery stores, theatres, hospitals, and other like-minded individuals, etc.

The goal of this capstone project is to develop a feature analysis that will help people moving to Scarborough find the best neighbourhood by comparing them to one another. The characteristics include the average cost of a home, an excellent school based on ratings, the area's crime rate, road accessibility, meteorological

conditions, effective emergency management, fresh and waste water resources, the conveyance of excrement through sewers, and recreational amenities. It will assist people in becoming familiar with the neighbourhood and area before relocating to a new city, state, or nation for employment or to begin a new life.

## 2. Data Section

Data Link: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](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 require information about various locations in several borough neighbourhoods.

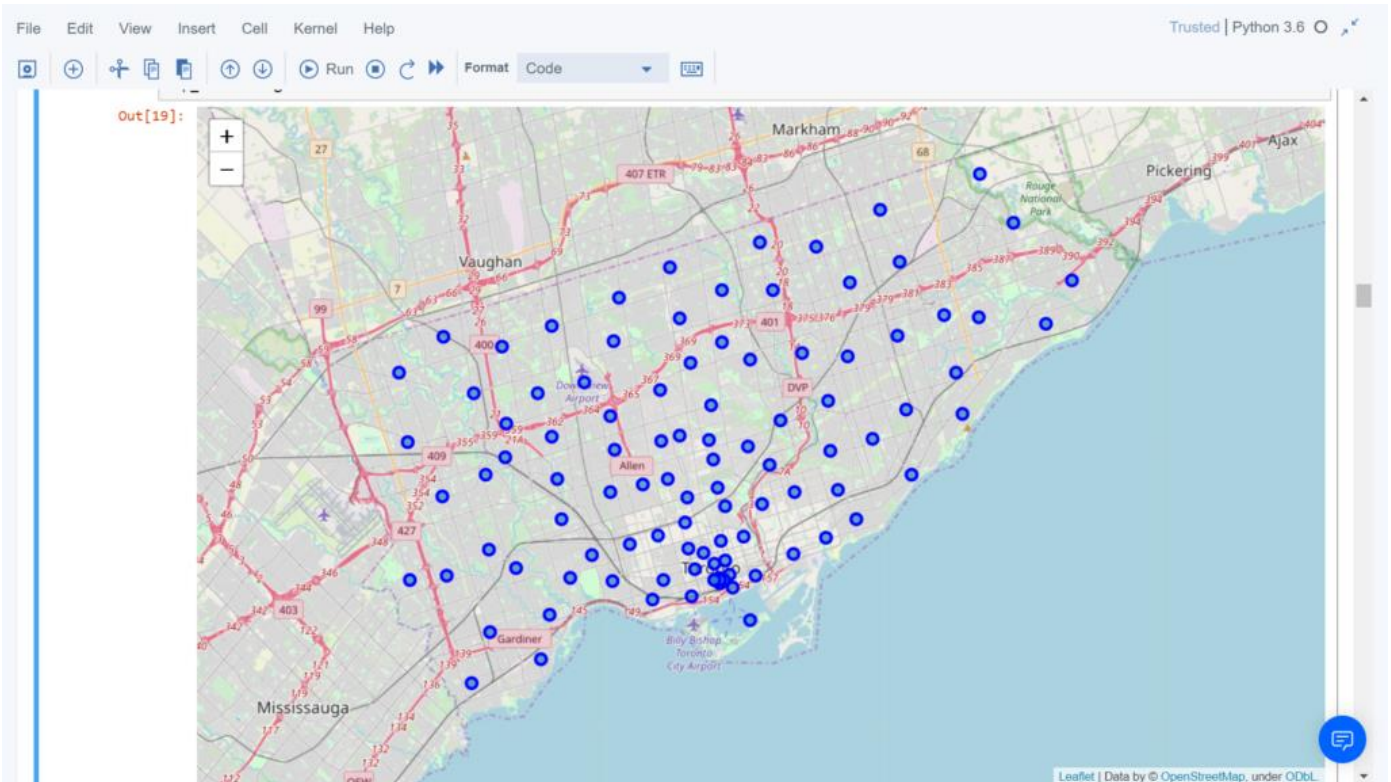
We will use "Foursquare" location data to obtain that information. Foursquare is a location data provider that offers details on a wide range of locations and events in a given area. These details include names of the venues, their locations, menus, and even pictures. Because the API makes it possible to acquire all of the specified required information, the foursquare location platform will be the only source of data.

Following the discovery of the neighbourhood list, we establish a connection with the Foursquare API in order to obtain venue details for each and every neighbourhood. We have selected a radius of 100 metres for each neighbourhood.

Venues within a given radius of the postcodes' latitude and longitude were included in the Foursquare data that was obtained. The data gathered for each location is 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

### **Map of Scarborough**



### 3. Methodology Section

#### Clustering Approach:

We made the decision to look for comparable neighbourhoods in large cities like New York and Toronto by exploring them, segmenting them, and putting them into clusters in order to analyse the similarities of the two cities. We need to cluster data using the k-means clustering algorithm, a type of unsupervised machine learning, in order to be able to accomplish that.

#### Using K-Means Clustering Approach | Most Common Venue

File Edit View Insert Cell Kernel Help Trusted | Python 3.6

```

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]:

	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
orough	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
orough	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
orough	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
orough	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
orough	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

Map of Clusters

```

In [37]: kclusters = 10

```

## Most Common Venues near Neighborhood | Using Clustering

File Edit View Insert Cell Kernel Help Trusted | Python 3.6

```

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	Ablon Gardens, Beaumont 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

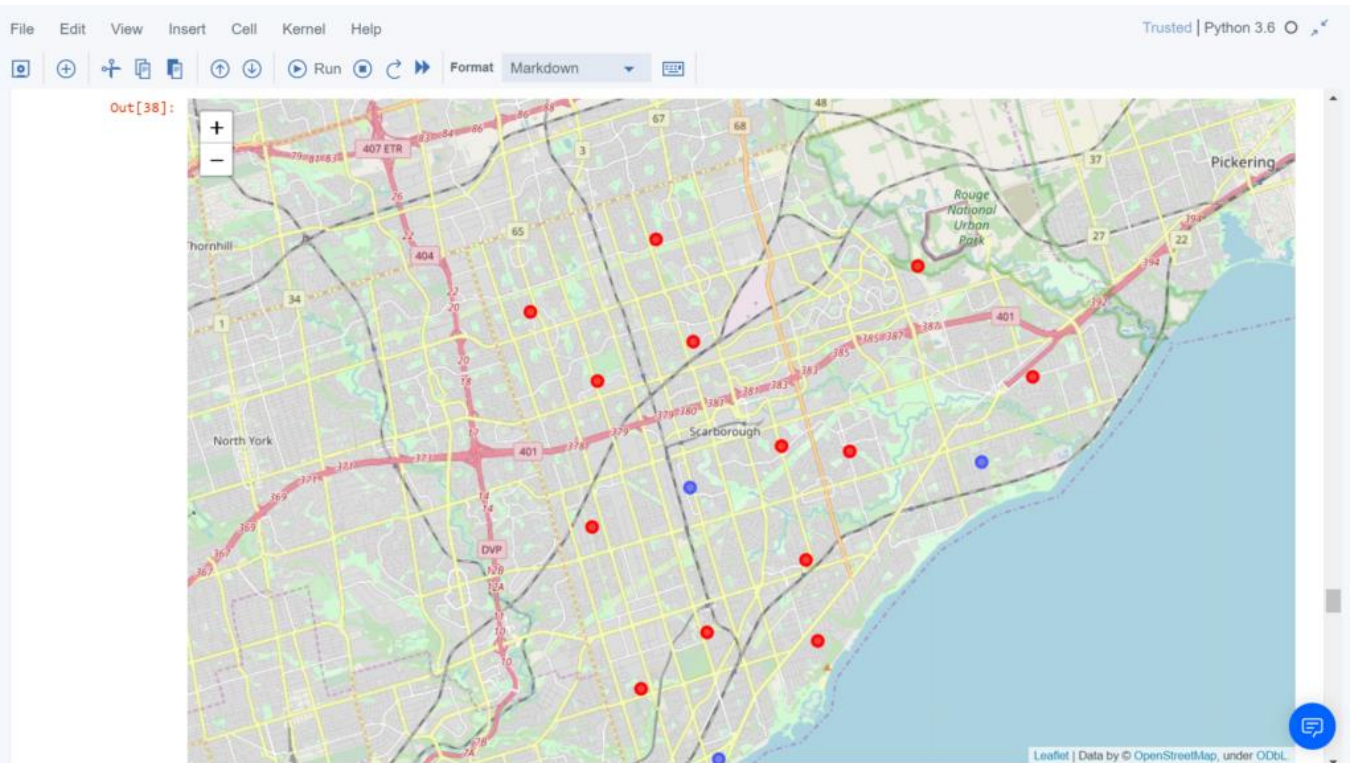
Work Flow:

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

would be set to 500.

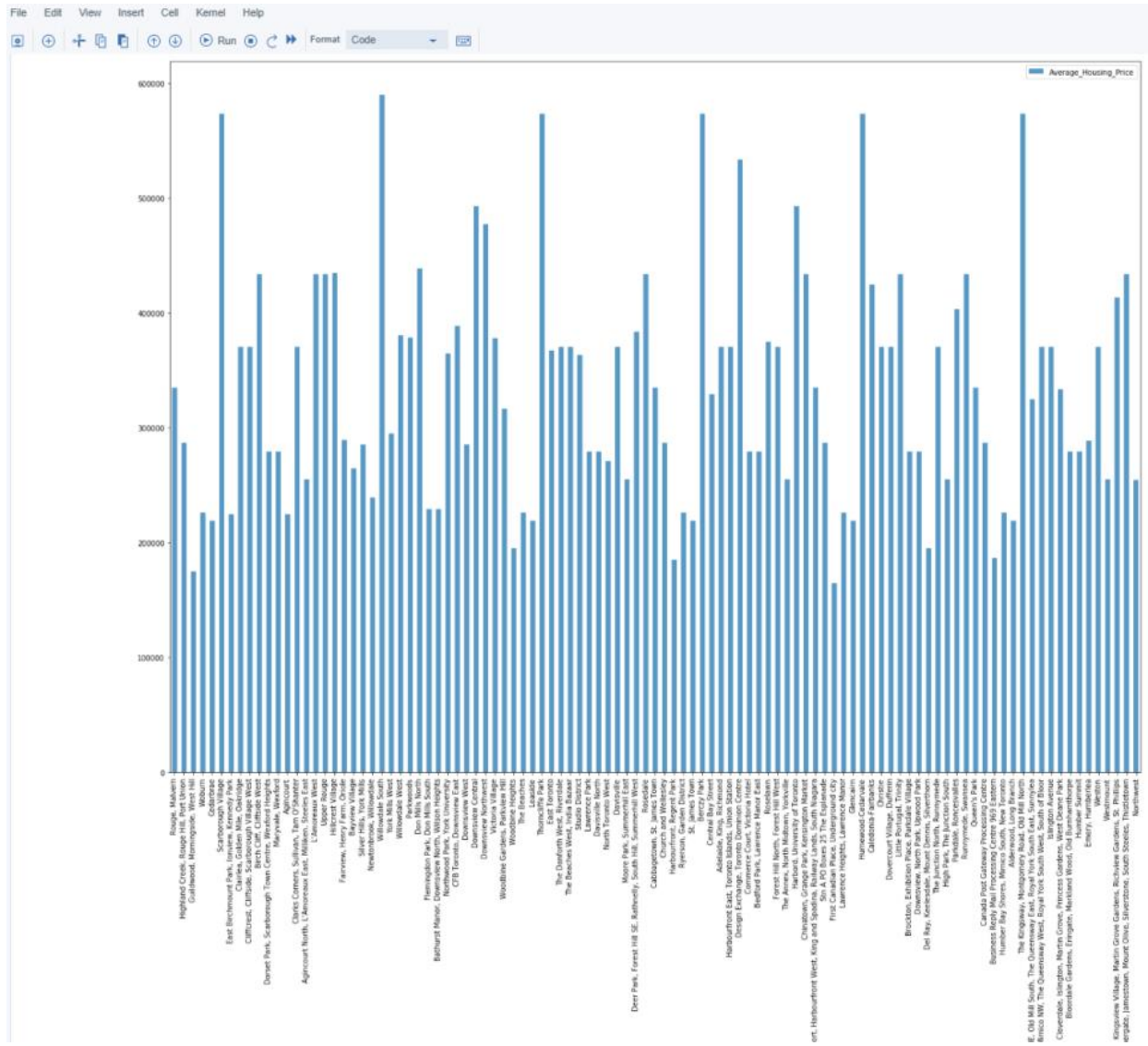
## 4. Results Section

### Map of Clusters in Scarborough



### Average Housing Price by Clusters in Scarborough





## School Ratings by Clusters in Scarborough



With its database of millions of locations, Four-square API has been the primary data source for our capstone project. In particular, its places API offers location sharing, location search, and company details.

## 5. Discussion Section

Problem Which Tried to Solve:

- This project's main goal is to recommend a better neighbourhood in a new city to those who are moving there. social presence in terms of like-minded individuals in society. accessibility to the city centre, bus stop, airport, stores, and other everyday necessities in the area.
- Sorted list of house in terms of housing prices in a ascending or descending order
- Sorted list of schools in terms of location, fees, rating and reviews

## 6. Conclusion Section

For my capstone project, I used the k-means cluster algorithm to divide the neighbourhood into 10 (ten) distinct clusters and for 103 different latitude and longitude values from the dataset, all of which had neighbourhoods that are extremely similar to one another. Based on average home prices and school ratings, a specific neighborhood's findings are displayed using the charts above.

I feel that my efforts have been appreciated, and this course, with all of the topics addressed, is definitely deserving of praise.

Through this project, I've learned how to use data science methods to tackle a real-world problem with potential financial and personal repercussions.

Folium mapping is an extremely effective method for combining data to improve analysis and decision-making with assurance.

Future Works:

The goal of this capstone project is to improve its precision in identifying the greatest home in Scarborough. The term "best" refers to the combination of all that is necessary (daily needs or things we need to live a better life) and affordability.

Libraries Which are 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.*