

# **Final Report | Capstone Project – The Battle of Neighborhoods Finding a Better Place in Scarborough, Toronto**

**By Amey Deshpande**

**March-2020**

## **1. Introduction:**

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.

Lots of people are migrating to various states of Canada and needed lots of research for good housing prices and better schools for their children. This project is for those people who are looking for better neighborhoods. For ease of accessing to Cafe, School, Supermarket, medical shops, grocery shops, mall, theatre, hospital, likeminded people, etc.

This Project aim to create an analysis of features for a people migrating to Scarborough to search a best neighborhood as a comparative analysis between neighborhoods. 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 wastewater and excrement conveyed in sewers and recreational facilities.

It will help people to get awareness of the area and neighborhood before moving to a new city, state, country or place for their work or to start a new fresh 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 need data about different venues in different neighborhoods 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

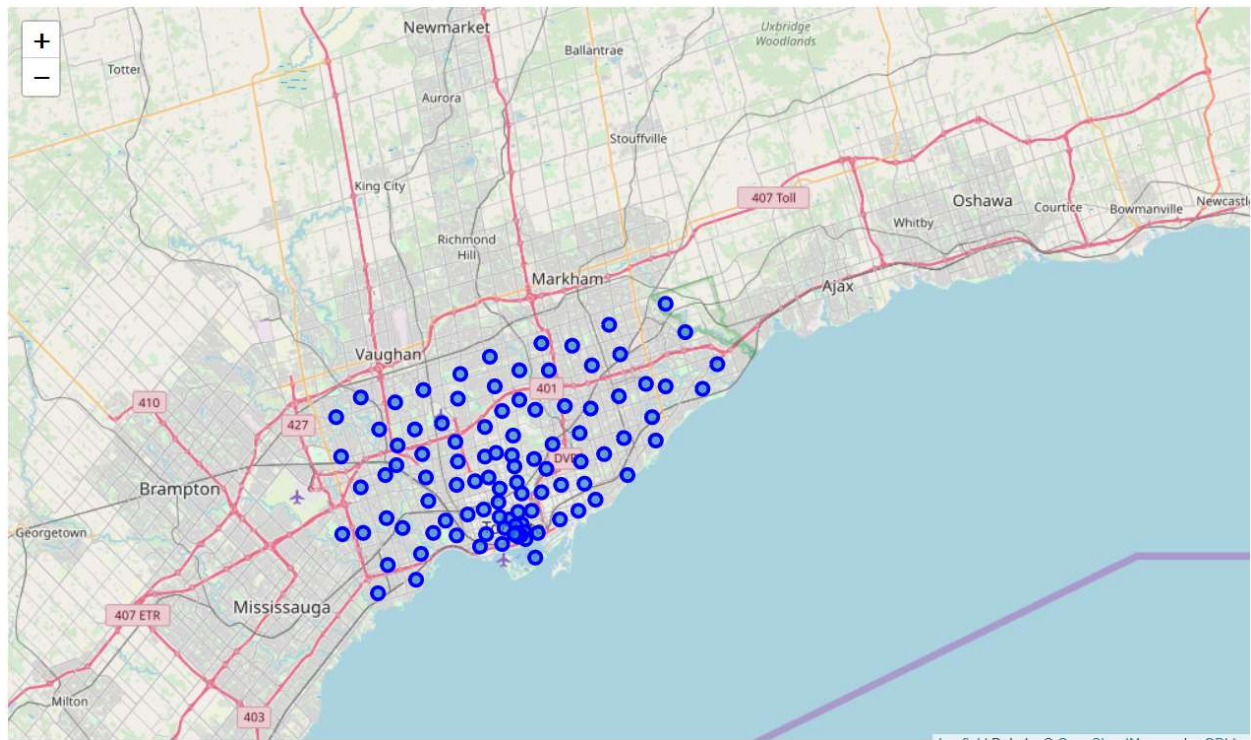
interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of 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 meters.

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

## Map of Scarborough



## 3. Methodology Section

### Clustering Approach:

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods 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

```
[32]: #neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
```

```
Scarborough_merged = toronto_data.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!
```

```
[32]:
```

|   | 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       | 7th Most Common Venue | 8th Most Common Venue |
|---|------------|--|--|-----------|------------|----------------|-----------------------|-----------------------|---------------------------|-----------------------|-----------------------|-----------------------------|-----------------------|-----------------------|
| 0 | M3A        | North York                                   | Parkwoods                                    | 43.752420 | -79.329242 | 1              | Park                  | Food & Drink Shop     | Bed & Breakfast           | Pet Store             | Yoga Studio           | Dive Bar                    | Dog Run               | Doner Restaurant      |
| 1 | M4A        | North York                                   | Victoria Village                             | 43.730600 | -79.313265 | 0              | Coffee Shop           | Pizza Place           | Middle Eastern Restaurant | Portuguese Restaurant | Intersection          | Eastern European Restaurant | Distribution Center   | Dive Bar              |
| 2 | M5A        | Downtown Toronto                             | Regent Park Harbourfront                     | 43.650295 | -79.359166 | 0              | Coffee Shop           | Pub                   | Theater                   | Bakery                | Park                  | Restaurant                  | Boat or Ferry         | Gym Pool              |
| 3 | M6A        | North York                                   | Lawrence Manor, Lawrence Heights             | 43.723270 | -79.451286 | 0              | Clothing Store        | Restaurant            | Dessert Shop              | Cosmetics Shop        | Toy / Game Store      | Men's Store                 | Bookstore             | Fried Chicken Joint   |
| 4 | M7A        | Queen's Park / Ontario Provincial Government | Queen's Park / Ontario Provincial Government | 43.661150 | -79.391715 | 0              | Coffee Shop           | Café                  | Sandwich Place            | Chinese Restaurant    | Italian Restaurant    | Creperie                    | Ice Cream Shop        | Burrito Place         |

## Most Common venues near Neighborhood

Most Common venues near neighborhood

```
[26]: 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]):
    neighborhood_venues_sorted.iloc[ind, 1:] = return_most_common_venues(Scarborough_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

```
[26]:
```

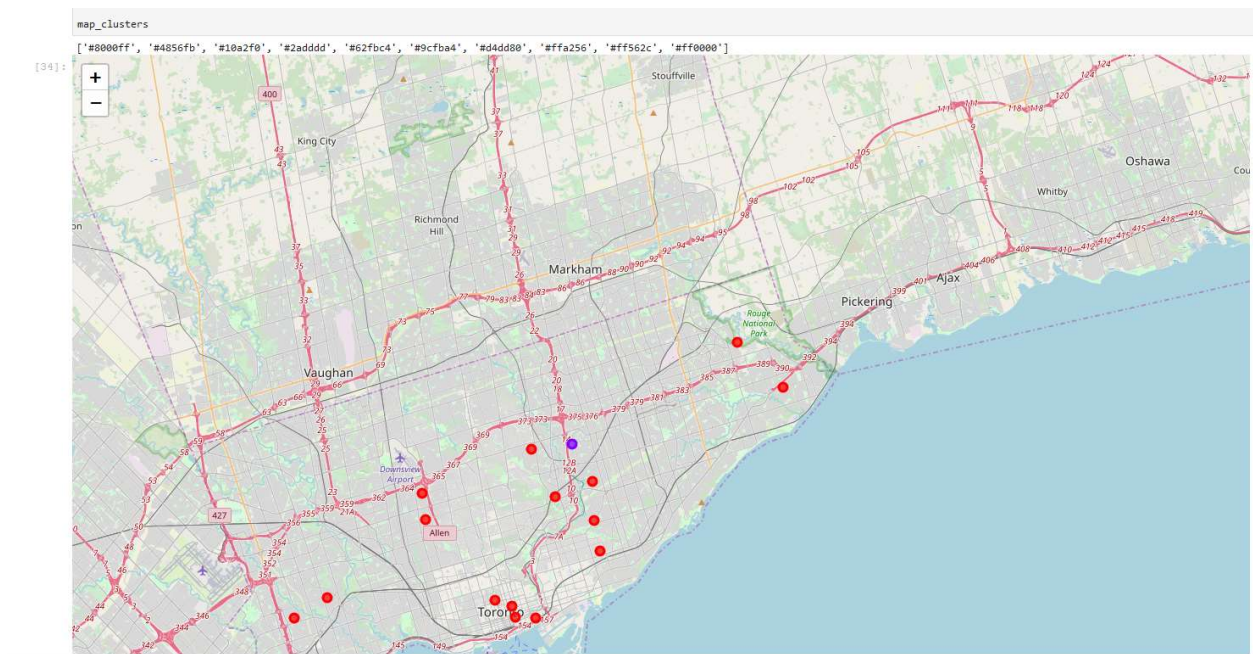
|   | Neighborhood                                    | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue    | 4th Most Common Venue | 5th Most Common Venue       | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|-----------------------|-----------------------|--------------------------|-----------------------|-----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 0 | Agincourt                                       | Chinese Restaurant    | Shopping Mall         | Pizza Place              | Malay Restaurant      | Sushi Restaurant            | Supermarket           | Japanese Restaurant   | Breakfast Spot        | Skating Rink          | Shanghai Restaurant    |
| 1 | Alderwood, Long Branch                          | Pizza Place           | Pharmacy              | Pool                     | Convenience Store     | Gas Station                 | Pub                   | Gym                   | Coffee Shop           | Sandwich Place        | Yoga Studio            |
| 2 | Bathurst Manor, Wilson Heights, Downsview North | Pizza Place           | Deli / Bodega         | Mediterranean Restaurant | Fried Chicken Joint   | Middle Eastern Restaurant   | Park                  | Coffee Shop           | Sushi Restaurant      | Sandwich Place        | Restaurant             |
| 3 | Bayview Village                                 | Park                  | Gas Station           | Trail                    | Asian Restaurant      | Eastern European Restaurant | Dive Bar              | Dog Run               | Doner Restaurant      | Donut Shop            | Dumpling Restaurant    |
| 4 | Bedford Park, Lawrence Manor East               | Italian Restaurant    | Coffee Shop           | Restaurant               | Sandwich Place        | Pizza Place                 | Fast Food Restaurant  | Sushi Restaurant      | Juice Bar             | Sports Club           | Liquor Store           |

## Workflow:

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.

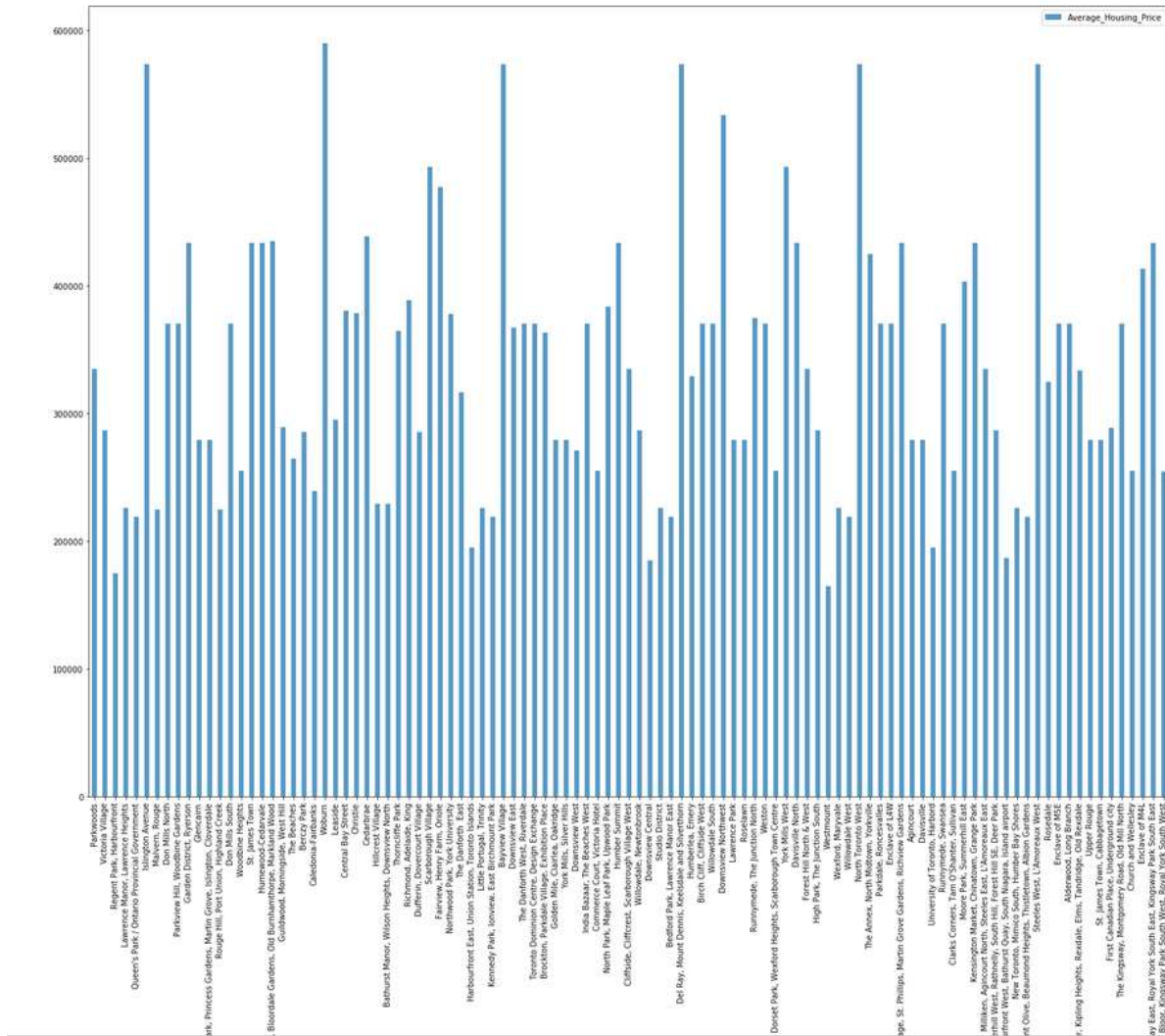
## 4. Results Section

### Map of Clusters in Scarborough



### Average Housing Price by Clusters in Scarborough

```
[40]: matplotlib.axes._subplots.AxesSubplot at 0x7fe354a7d6d0>
```



## School Ratings by Clusters in Scarborough





The major purpose of this project is to suggest a better neighborhood in a new city for the person who are shifting there. Social presence in society in terms of likeminded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

1. Sorted list of houses in terms of housing prices in a ascending or descending order
2. Sorted list of schools in terms of location, fees, rating and reviews

## **6. Conclusion Section**

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 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 Which are Used to Developed the Project:**

Pandas: For creating and manipulating data frames.

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.