

## **1. Introduction:**

The Project aims to determine the suitable neighborhood in Scarborough considering the various amenities therein. This will enable decision making for people looking out for a house in Scarborough, Toronto.

Much ground work is needed for choice of good housing prices and good schools for children. This project suggests a better neighborhood for nearness to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, like minded people, etc.

This project analyses the various features like 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, recreation etc.

Thus the project provides a broad overview of area and neighborhood for people who are looking at moving to a new city, state, country or place for their work.

## **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)

The Scarborough dataset consisting of latitude and longitude, postal codes of Canada has been chosen.

### **Foursquare API Data:**

Since we require data about different venues in different neighborhoods of that specific borough. "Foursquare" locational information is used. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. It contains information on venue names, locations, menus etc.

After finding the list of neighborhoods, we connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

The data retrieved from Foursquare contains information of venues within a specified distance of the longitude and latitude of the postal codes. The information obtained per venue as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude

4. Venue

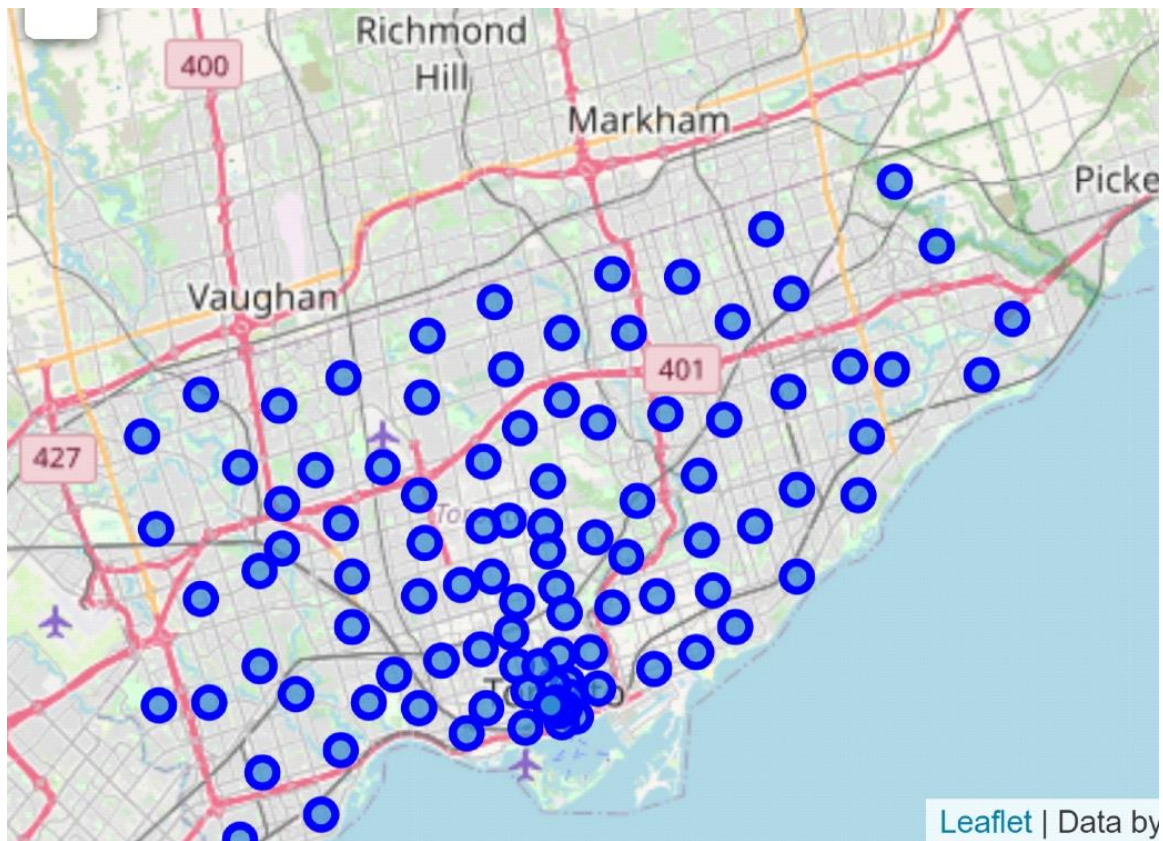
5. Name of the venue e.g. the name of a store or restaurant

6. Venue Latitude

7. Venue Longitude

8. Venue Category

**Map of Scarborough, Toronto.**



### 3. Methodology Section

Clustering Approach:

To compare the similarities of two cities, we 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. For this the k-means clustering algorithm is chosen.

```
In [119]: Scarborough_grouped_clustering = Scarborough_grouped.drop('Neighborhood', 1)
           kmeans = KMeans(n_clusters=3, random_state=0).fit(Scarborough_grouped_clustering)
           kmeans.labels_

Out[119]: array([[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                  0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                  0, 0, 0, 0, 0, 0, 2, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
                  0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                  1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1])
```

```
In [120]: neighborhoodsvenues_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(neighborhoodsvenues_sorted.set_index('Neighborhood'), on='Neighborhood')

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

```
Out[120]:
```

	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
0	M1A1n	Not assigned'n	Not assigned'n	43.64969	-79.38544	0	Coffee Shop	Hotel	Cafe	Beer Bar	Restaurant	Bar	Italian Restaurant
1	M1B1n	Scarborough'n	Malvern, Rouge	43.81139	-79.19682	2	Zoo Exhibit	Fast Food Restaurant	Paintball Field	Fish & Chips Shop	Filipino Restaurant	Field	Fish Market

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.

## Map of Clusters in Scarborough

```

kclusters = 10

map_clusters = folium.Map(location=[latitude_x, longitude_y], zoom_start=11)

# set color scheme for the clusters
x = np.arange(kclusters)
colors_array = cm.rainbow(np.linspace(0, 1, kclusters))
rainbow = [colors.rgb2hex(i) for i in colors_array]
print(rainbow)
# add markers to the map

markers_colors = []
for lat, lon, nei, cluster in zip(Scarborough_merged['Latitude'],
                                  Scarborough_merged['Longitude'],
                                  Scarborough_merged['Neighborhood'],
                                  Scarborough_merged['Cluster Labels']):
    label = folium.Popup(str(nei) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters

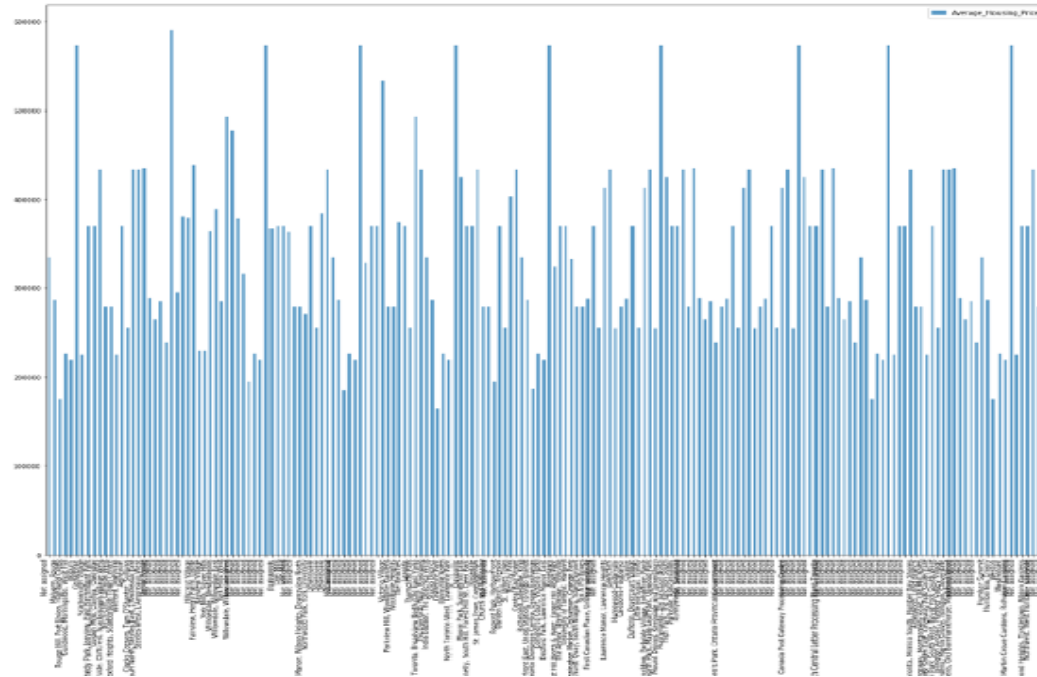
```

['#8a00ff', '#4056fb', '#10a2f0', '#2adddd', '#62fbc4', '#9cfba4', '#d4dd80', '#ffa256', '#ff562c', '#ff0000']



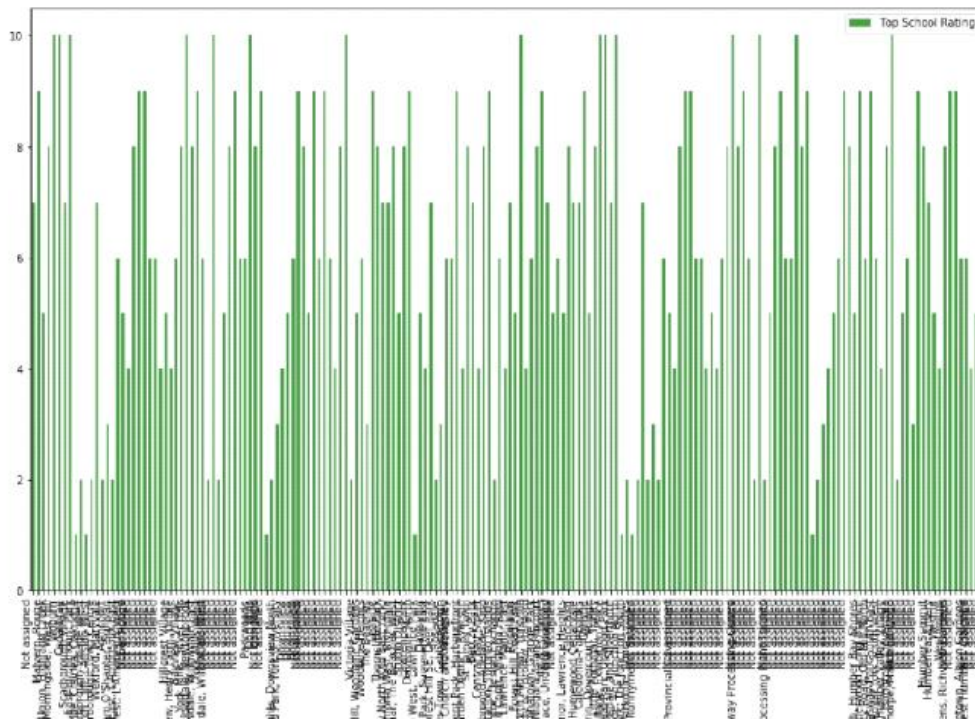
**Average Housing Price by Clusters in Scarborough**

```
<matplotlib.axes._subplots.AxesSubplot at 0x18edcd93d48>
```



## School Ratings by Clusters in Scarborough

```
Scarborough_school_ratings.set_index('Neighborhood', inplace=True, drop=True)
Scarborough_school_ratings.plot(kind='bar', figsize=(16,10), color='green', alpha=0.75);
```



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.

### Foursquare API:

In this project Four-square API is used 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.

## 5. Discussion Section

The major purpose of this project, is to suggest a better neighborhood in a city for the person who plans to shift the residence. Common factors that are to be scrutinized are Social presence in society in terms of like minded people, Connectivity to the airport, bus stand, city center, markets and access to daily needs stores.

Sorted list of house in terms of housing prices in ascending or descending order

Sorted list of schools in terms of location, fees, rating and reviews

## **6. Conclusion Section**

In this project, using k-means cluster algorithm neighborhood is divided into 10(Ten) different clusters and for 180 different latitude and longitude with similar neighborhoods around them. Neighborhood is suggested based on average house prices and school ratings.

### **Future Works:**

This project can be expanded in terms of other parameters such as nearest metro service etc.

### **Libraries used are as follows:**

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.