



# **Capstone Project: The Battle of Neighborhoods - Finding a Better PLace in Scarborough, Toronto**



# 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.*

*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.*

## Data Selection

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)

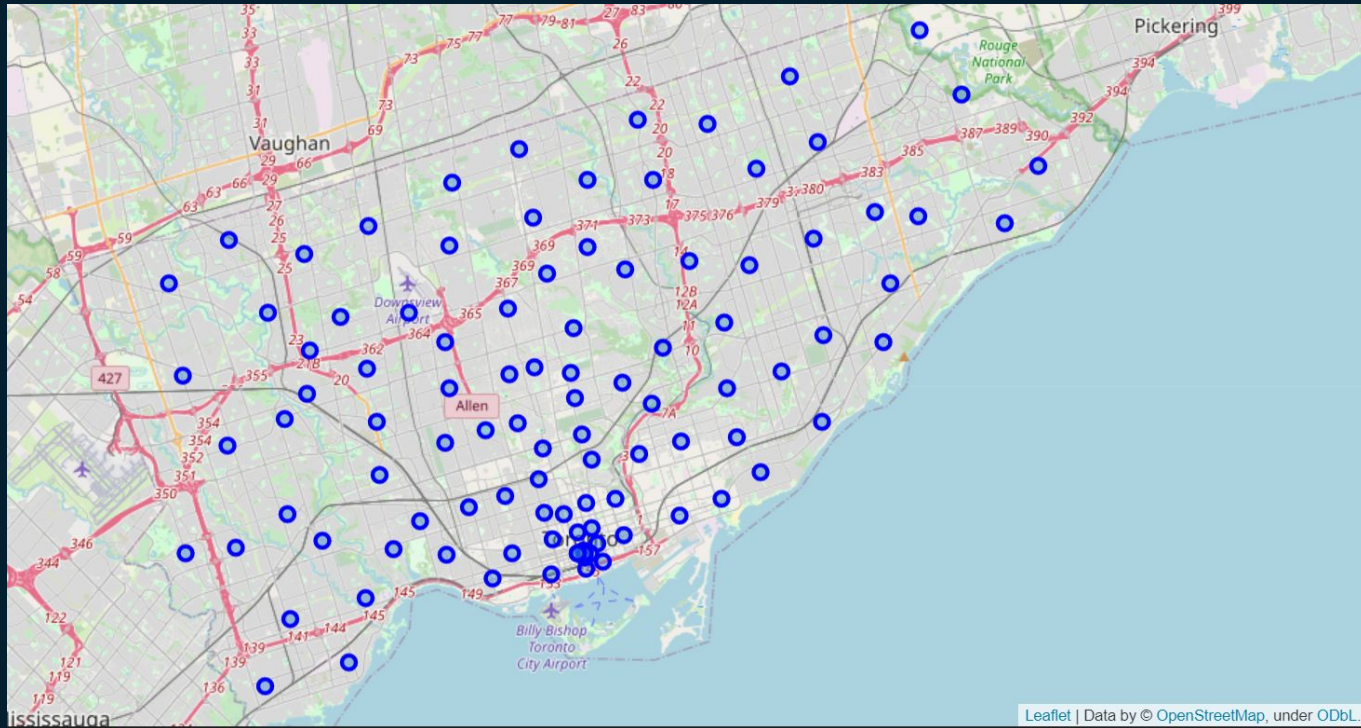
Will use Scarborough dataset which we scrapped from wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes.

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 and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

# Map of Scarborough



## Methodology

### 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:

```
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!
```

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

## Most Common venues near Neighborhood:

```
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()
```

	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	Pharmacy	Print Shop	Supermarket	Sandwich Place	Chinese Restaurant	Bank	Bakery	Dim Sum Restaurant	Clothing Store
1	Alderwood, Long Branch	Convenience Store	Pizza Place	Sandwich Place	Gym	Pharmacy	Coffee Shop	Print Shop	Pub	Gas Station	Event Space
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Sushi Restaurant	Mobile Phone Shop	Mediterranean Restaurant	Pizza Place	Middle Eastern Restaurant	Fried Chicken Joint	Deli / Bodega	Restaurant	Sandwich Place
3	Bayview Village	Flower Shop	Park	Trail	Asian Restaurant	Gas Station	Yoga Studio	Donut Shop	Eastern European Restaurant	Electronics Store	Elementary School
4	Bedford Park, Lawrence Manor East	Sandwich Place	Italian Restaurant	Coffee Shop	Pizza Place	Indian Restaurant	Liquor Store	Thai Restaurant	Sports Club	Restaurant	Sushi Restaurant

### Workflow:

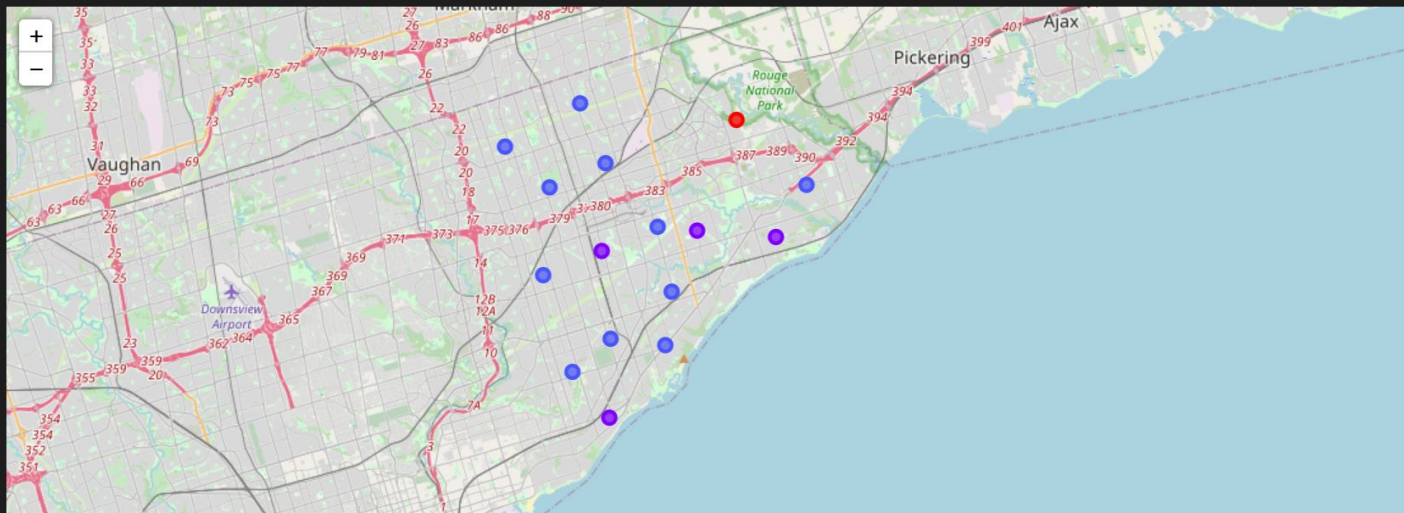
Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

## Results:

# Map of Clusters - Scarborough

```
popup=popup,  
color=rainbow[cluster-1],  
fill=True,  
fill_color=rainbow[cluster-1],  
fill_opacity=0.7).add_to(map_clusters)
```

map\_clusters



[illegible][illegible]



## The Location:

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:

This project have used Four-square API as its prime data gathering source 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.

## Discussion Section

Problem Which Tried to Solve:

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 like minded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

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

## Conclusion Section

In this project, using k-means cluster algorithm I separated the neighborhood into 10 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.

The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.

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