1. **Introduction:**

The purpose of this project is to help in exploring better facilities around the 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 reputed schools for their children. This project is for those people who are looking for better neighborhoods in order to have easy access to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital.

The aim of the project is to create an analysis of features for the people migrating to Scarborough to search a best neighborhood as a comparative analysis between neighborhoods. The features include median housing price and better school in terms of ratings, crime rates of that particular area, road connectivity, weather conditions, good management for emergency, water resources both fresh and waste water and excrement conveyed in sewers and recreational facilities.

It will help people in increasing their 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 Source : <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

We scrapped the Scarborough dataset in week 3 assignments. We will use that for the purpose of getting information about latitude, longitude and 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 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. Approach & Method

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.

**Libraries:**

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.

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



**Most Common Venues near Neighborhood** | Using Clustering



**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.

**4. Results Section**

**Map of Clusters in Scarborough**



**Average Housing Price by Clusters in Scarborough**



**School Ratings by Clusters in Scarborough**



**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.

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.

**5. Discussion**

As mentioned above, we will work on the project to solve the multiple problems faced by people in terms of how unaware they are about the surrounding places.The major purpose of this project, is to suggest a better neighborhood in a new city for the person who are shifting there. We will do it by clustering the views of like minded people, who have provided their reviews and insights about a location.

* Knowledge about housing prices are one of the ways of finding a location suitable for a new person's living. We will work on this.
* People want to know and think a lot about the schools in an area where their children can go for gaining education. We will analyse the location and fees of schools based on ratings and reviews provided by other like-minded people.

**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 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.