

**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.

**Project Description:**

Many people migrating to various states of Canada require search of a good housing prices as well as good rating schools for their children. The projects aim to create an analysis of features for a neighborhood as a comparative analysis between neighborhoods. The features include median house price and school ratings and recreational facilities. This would help people to get awareness of the places before moving to a new country, state, city or place for their work or to start a new life This Project would help people take a better decision on choosing the best neighborhood out of many neighborhoods in Scarborough city based on the distribution of various facilities in and around that neighborhood.

**Foursquare API:**

This project would use 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.

**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. Steps taken were:

1. Data acquisition and cleansing
2. Data preparation
3. Feature selection
4. Clustering

**Data acquisition and cleansing**

Data acquisition was a 2-step process:

1. Obtaining the postcodes for neighborhoods in Toronto
2. Obtaining venues within these neighborhoods

## **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

## **Conclusion:**

In this project, through a k-means cluster algorithm we separate the neighborhood into 03 clusters, which have similar neighborhoods around them. Using the charts above decision leading to a particular neighborhood based on average house prices and school rating can be made.