

Capstone

The Battle of Neighborhoods

## Introduction/Business Problem



## Project Description

Due to various reasons like Jobs, Business many people migrating to various states of Canada and it requires search of a good housing prices as well as good rating schools for their children.

This projects aim is to create an analysis of features for a neighbourhood as a comparative analysis between neighbourhoods. The features include median house price and school ratings, crime rates, weather conditions, 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

Also this project helps people explore different possibilities and take a better decision on choosing the best neighbourhood out of many neighbourhoods in **Scarborough** city based on the distribution of various facilities in and around that neighbourhood.

## Selection Criteria:

For the purposes of this project, the definition of a good neighbourhood is one that has an appreciable commercial presence within a given community as well as:

1. Compare median housing prices

## 2. Compare school ratings

### **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. It includes a number of natural landmarks, including the Toronto Zoo, Rouge Park and the Scarborough Bluffs.

### **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 neighbourhoods would be mined. Due to http request limitations the number of places per neighbourhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

### **Clustering Approach:**

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

*Libraries*

*Pandas*

*Folium*

*Scikit Learn*

*JSON*

*Geopy*

*Matplotlib*

*Requests*

*Description*

For creating and manipulating data frames

Python visualization library would be used to visualize the neighbourhoods cluster distribution of using interactive leaflet map.

For importing k-means clustering

Library to handle JSON files

To retrieve Location Data

Python Plotting Module

Library to handle http requests