

# Applied Data Science using Python

## Capstone

### The Battle of Neighborhoods: Part-1

#### Problem Introduction & Explanation

##### 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.
- 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. For ease of accessing to Cafe, School, Super market, medical shops, and grocery shops, mall, theatre, hospital, likeminded people, etc.
- This Project aims 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 waste water and excrement conveyed in sewers and recreational facilities.
- It will help people to get 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.

## 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 is shifting there. Social presence in society in terms of likeminded 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

## 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 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.

## 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

## Libraries which are used to developed the Project:

- i) **Pandas:** For creating and manipulating data frames.
- ii) **Folium:** Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
- iii) **Sickit Learn:** For importing k-means clustering.
- iv) **JSON:** Library to handle JSON files.
- v) **XML:** To separate data from presentation and XML stores data in plain text format.
- vi) **Decoder:** To retrieve Location Data.
- vii) **Beautiful Soup and Requests:** To scrap and library to handle http requests.
- viii) **Matplotlib:** Python Plotting Module.