Capstone Project - The Battle of Neighborhoods

Business Problem

Project Description:

Most 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. This would help people to get awareness of the places before moving to new state, city or place for their work or to start a new life.

The aim of this Project is to help people explore different possibilities and 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.

Selection criteria

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

- Compare median housing prices
- Compare school ratings

The Location:

Scarborough is a popular destination for new immigrants in Canada to reside.

Foursquare API:

This project would use Four-square API

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. 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 dataframe
- Matplotlib: Python Plotting Module
- Requests: Library to handle http requests
- 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
- Geopy: To retrieve Location Data