

Report - Battle of Neighborhood

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 Canada

Lots of people are migrating to Canada. This project is for those people who are looking for better neighborhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, like minded people, etc.

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 Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

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.

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 Develop the Project:

- Pandas: For creating and manipulating dataframes.
- Matplotlib: Python Plotting Module.
- 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.

Data Description

[Data Link:](#)

Will use Scarborough dataset which we scrapped from wikipedia in Week 3 . Dataset consisting of latitude and longitude, zip codes.

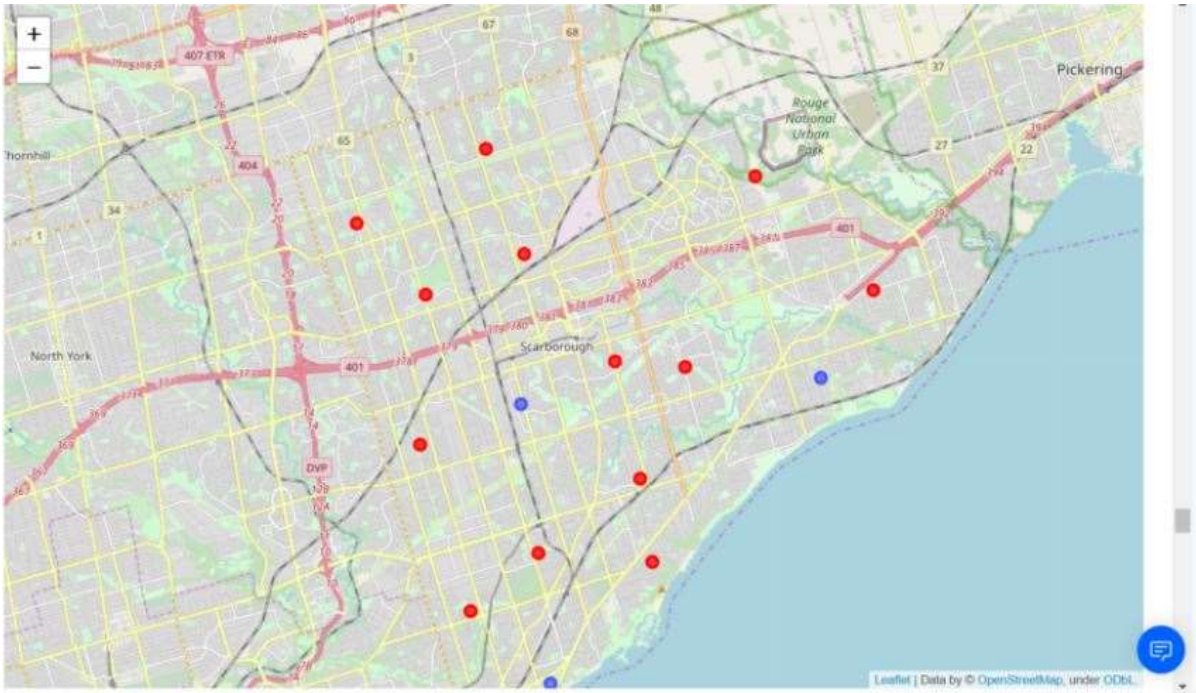
The data retrieved contained information of venues within a specified distance of the longitude and latitude of the postcodes.

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

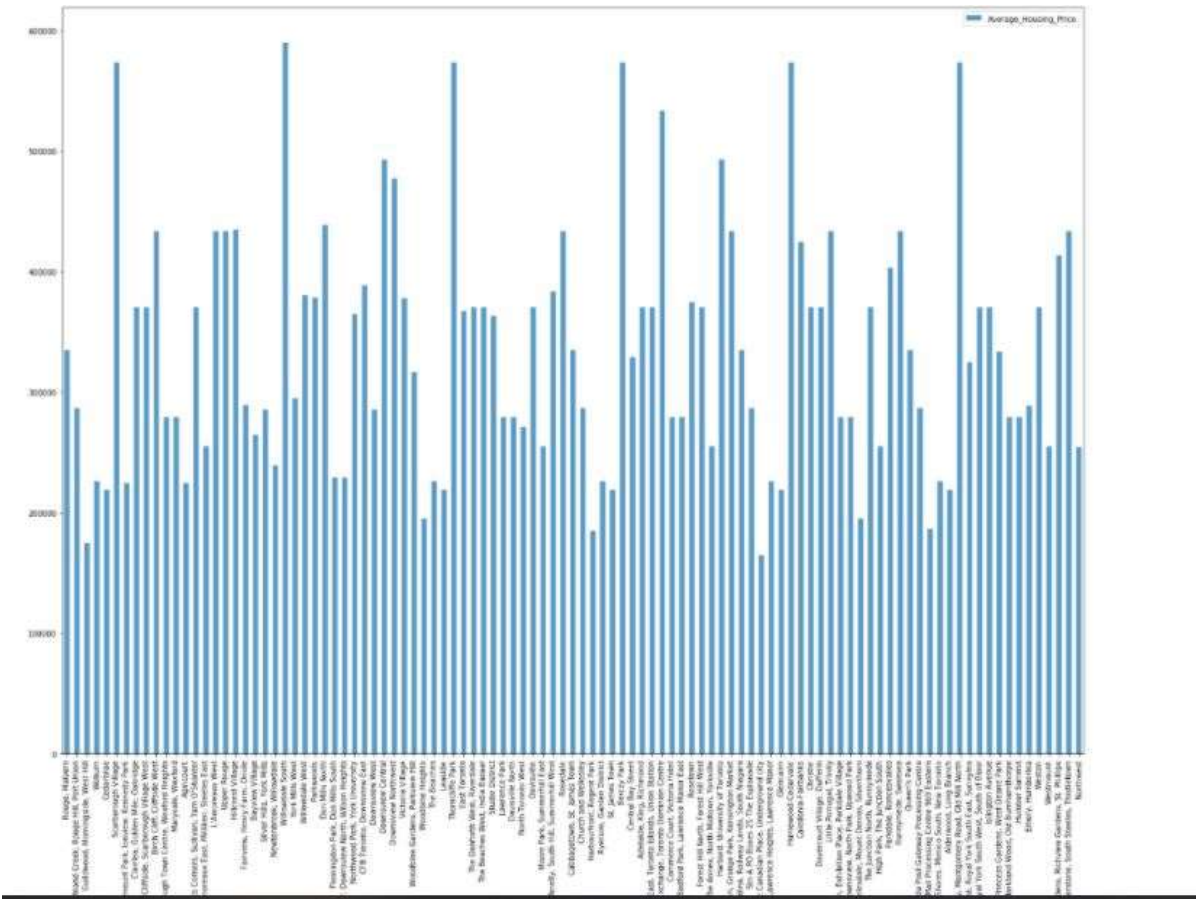
Map of scarborough



Result



Housing Price



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.

Future Works: This project can be continued for making it more precise in terms to find best house in Scarborough. Best means on the basis of all required things(daily needs or things we need to live a better life) around and also in terms of cost effective.

In []: