

Capstone Project: The Battle of Neighborhoods

Introduction:

The main objective of this project is to help people in exploring better facilities around their neighborhood. It will help people in making a smart and efficient decision on choosing excellent neighborhoods out of numbers of other neighborhoods in Scarborough, Toronto.

Many people are migrating to many different states of Canada and need a lot of research for sound 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, Supermarket, medical shops, grocery shops, mall, theatre, hospital, like-minded people, etc.

This project intends to create an analysis of features for people who are migrating to Scarborough in search of the best neighborhood as a relative analysis between neighborhoods. The highlights include median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, good management for an emergency, water resources both fresh and wastewater and excrement conveyed in sewers and recreational facilities.

It will help people to get the 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.

Data Description:

Data Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

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

Foursquare API Data:

We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information, we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

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

Methodology:

The major goal of this project is to recommend a better neighborhood in a new city for the person who is moving there. Social presence in society in terms of like-minded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

- 1) Sorted list of the house in terms of housing prices in an ascending or descending order
- 2) Sorted list of schools in terms of location, fees, rating, and reviews

The Location:

Scarborough is a favorite destination for new immigrants in Canada to reside in. As a result, it is one of the most distinct 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 constraints on immigrants and refugees, the general trend of immigration into Canada has been one of the rises.

Foursquare API:

This project would use Four-square API as its prime data collecting source as it has a database of millions of places, notably their places API which provides the capacity 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 restrictions, the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

Clustering Strategy:

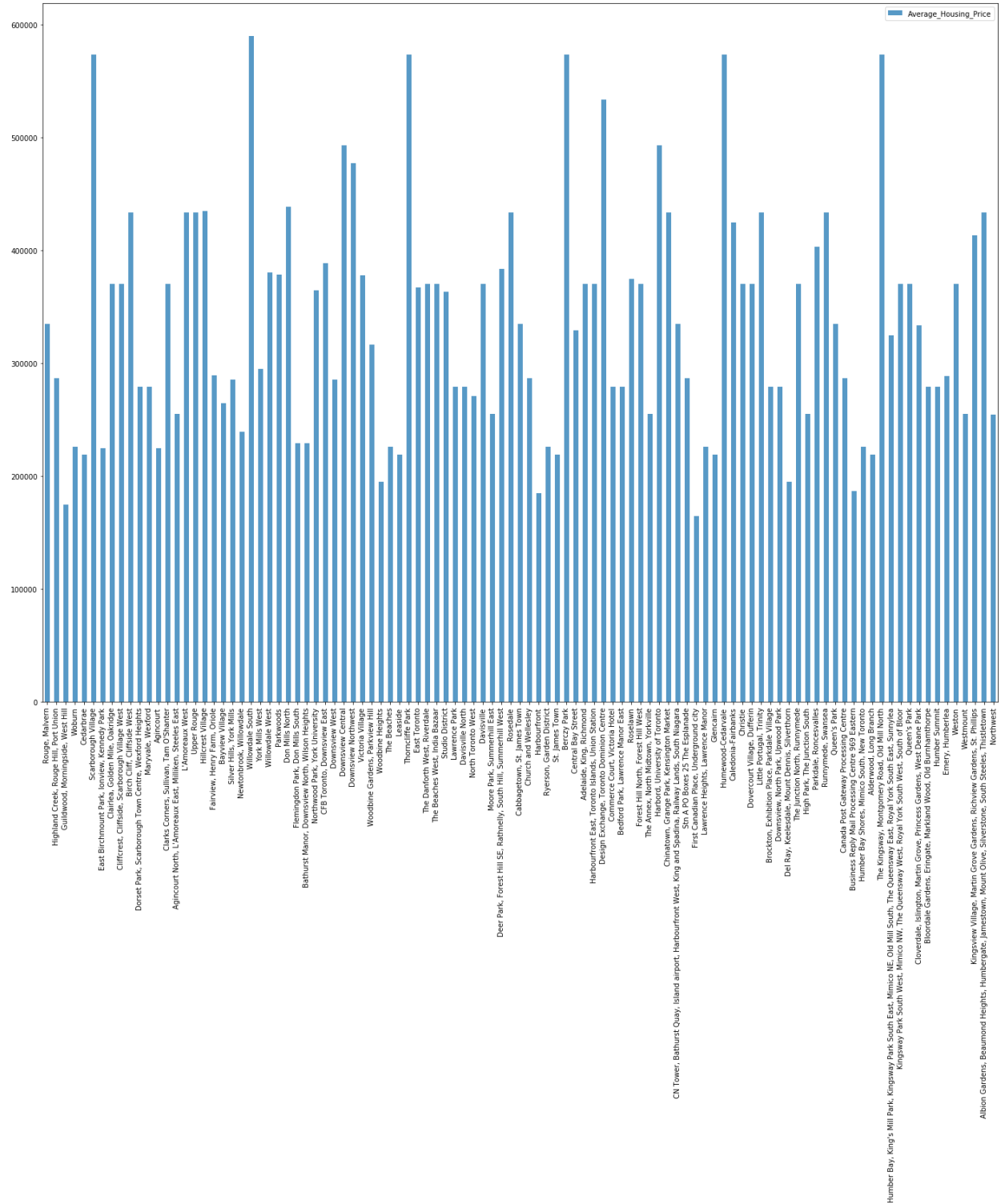
To compare the similarities of two cities, we decided to examine neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: a k-means clustering algorithm

Libraries Which are Used to Develop the Project:

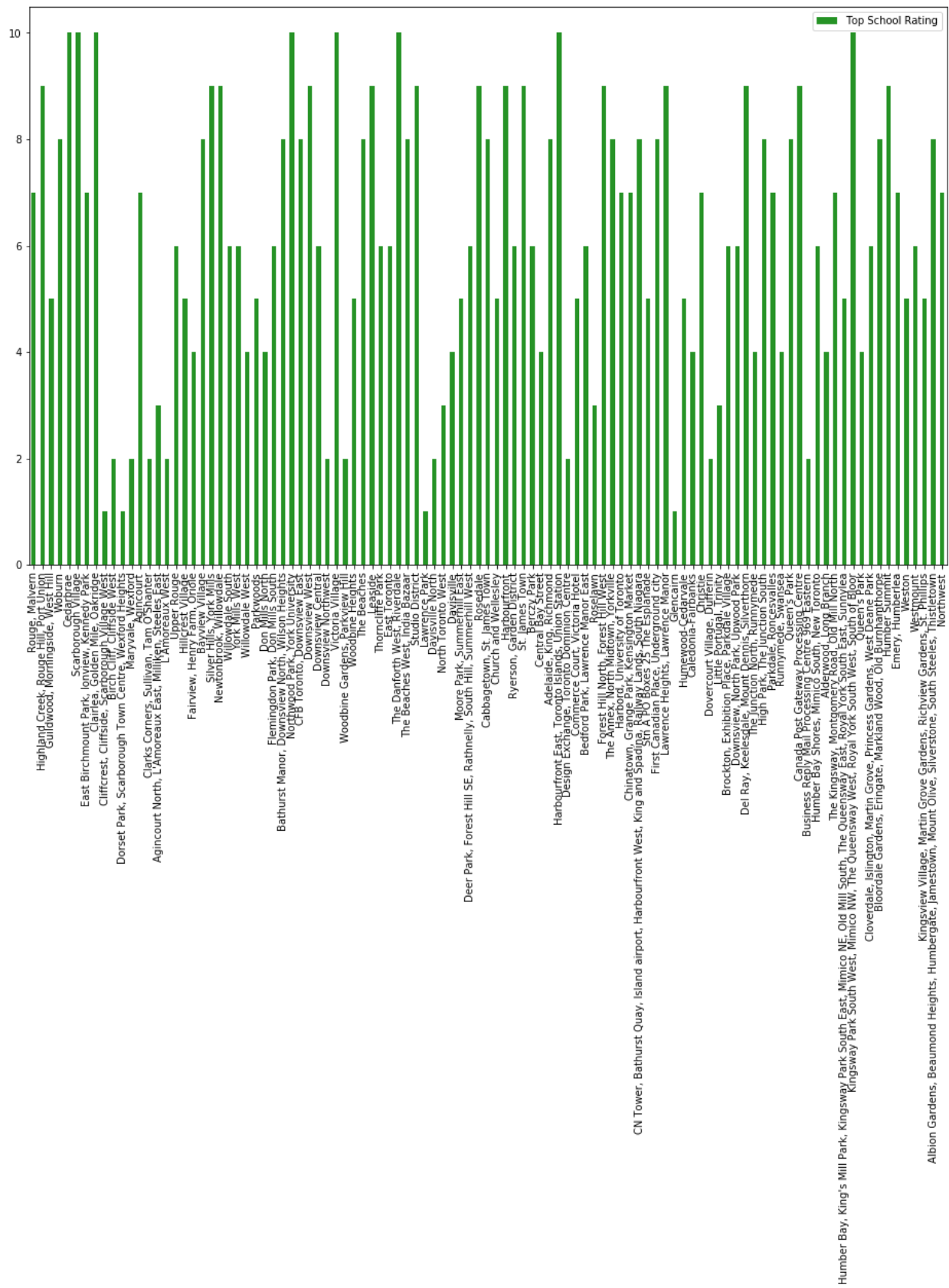
1. Pandas: To create and manipulate data frames.
2. Folium: Python visualization library would be used to visualize the neighborhood cluster distribution of using an interactive leaflet map.
3. Scikit Learn: To import k-means clustering.
4. JSON: Library to handle JSON files.
5. XML: To separate data from presentation and XML stores data in plain text format.
6. Geocoder: To retrieve Location Data.
7. Beautiful Soup and Requests: To scrap and library to handle HTTP requests.
8. Matplotlib: To Python Plotting Module.

Result:

The highlights include median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, good management for an emergency, water resources both fresh and wastewater and excrement conveyed in sewers and recreational facilities.



School Rating by clusters:



Discussion:

The major goal of this project is to recommend a better neighborhood in a new city for the person who is moving there. Social presence in society in terms of like-minded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby. We can predict more data like others features so that the person can migrate comfortably from one place to other and they do not have to face problems and also, they can live pleasantly.

Conclusion:

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.