

The Battle of Neighbourhoods Project

Introduction:

This Project can be considered an extension of what we have already been working on and is going to help people finding the best neighbourhoods in Toronto, Canada. It is the most populous city in Canada and the largest urban and metro area, with a population density of 4,149.5 people per square kilometre (World population review). So, clearly a lot of people move to this city every year and providing them with the results of this project is going to help them decide on the best neighbourhood in Toronto area. The analysis of this project is going to include the density and availability of schools, cafes, grocery shops, supermarkets, hospitals, malls etc.

This project is aimed at creating a detailed analysis report of the neighbourhoods in Toronto to help people with moving to Toronto. The features will include median housing prices and better schools according to ratings, crime rates of that particular area, road connectivity, weather conditions, good management for emergency, water resources and recreational facilities.

Problems that are assessed in this project:

The major purpose of this project, is to suggest a better neighbourhood in the city of Toronto, social presence in society in terms of likeminded people, connectivity to the airport, bus stand, city centre, markets, and proximity of other daily needs things.

> Sorted list of houses in terms of housing prices in a ascending or descending order

> Sorted list of schools in terms of location, fees, rating and reviews

Libraries and resources:

The Location:

Scarborough is turning to be a popular destination for new immigrants. 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 worships. 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 on the rise.

Foursquare API:

We'll be using the foursquare API because of it's huge database and the features it has which will help ease the search and analysis of neighbourhoods.

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 Which are Used to Develop the Project:

Pandas: For creating and manipulating data frames.

Folium: Python visualization library would be used to visualize the neighbourhoods 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 data and library to handle http requests.

Matplotlib: Python Plotting Module.

Data Source

We will use the dataset that we used in week 3 of this course.

Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada: M

Foursquare API:

We will need data about different venues in different neighbourhoods 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 neighbourhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighbourhood and the radius is limited to 100 mts.

The following data is retrieved from Foursquare API:

1. Neighbourhood
2. Neighbourhood Latitude
3. Neighbourhood Longitude
4. Venue and Name
5. Venue Latitude
6. Venue Longitude
7. Venue Category

Methodology Section

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.

Using K-Means Clustering Approach

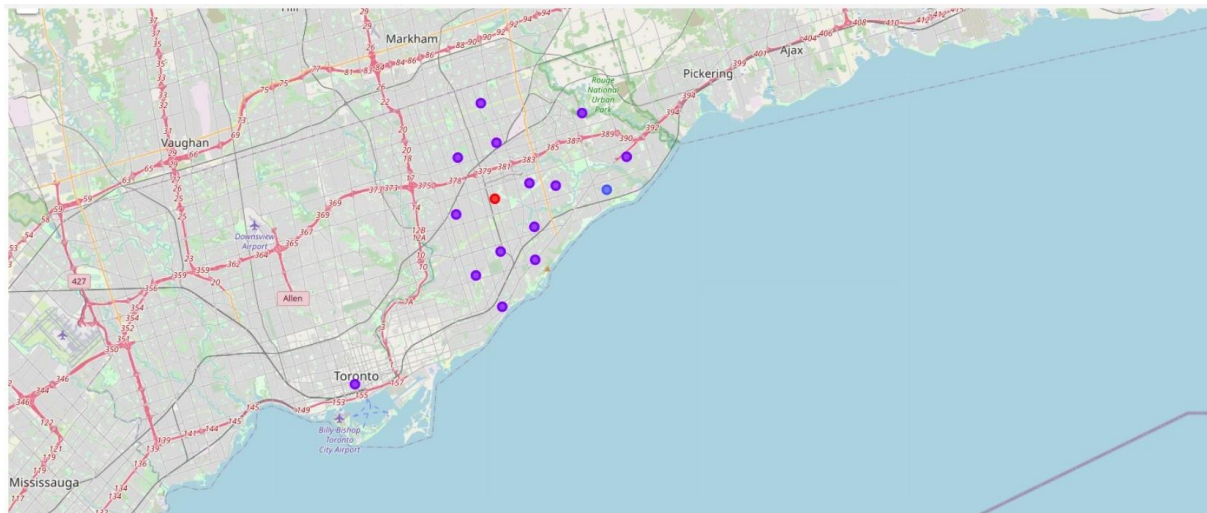
Most Common venues near Neighborhood

Work Flow:

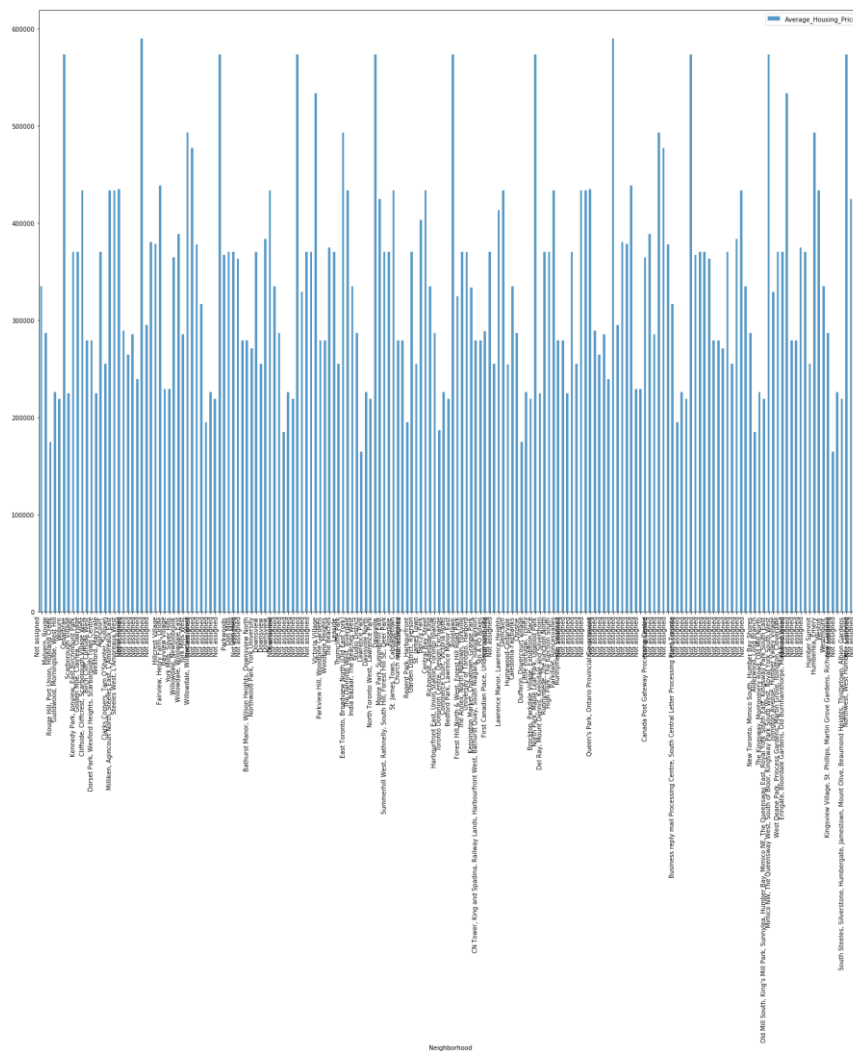
Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

Results Section

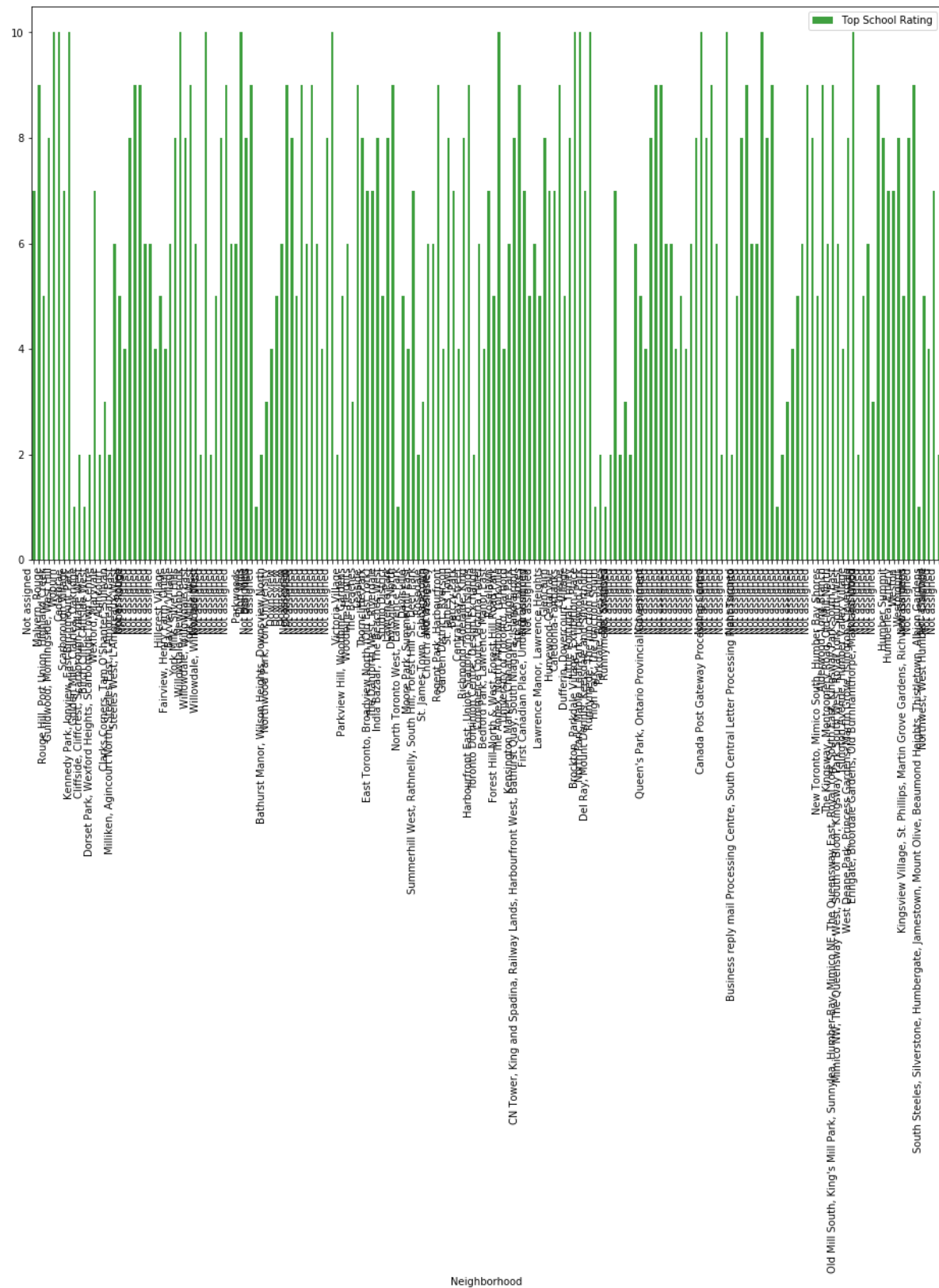
Map of Clusters in Scarborough



Average Housing Price by Clusters in Scarborough



School Ratings by Clusters in Scarborough



Conclusion

In this project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 180 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.