The Battle of Neighborhoods

Better Neighborhood in Toronto

Introduction:

The purpose of this Project is to help people in exploring better facilities around their neighborhood. It will help people making smart decision on selecting a great neighborhood in Scarborough, Toronto.

Lot of people migrate to various parts of Canada for jobs and other business reasons. They need to do a lot of research for finding good housing price and reputed school for their children. This project is for people who are looking for better neighborhoods with easy access to Cafe, School, Super market, medical shop & grocery shops, hospital, etc.

This Project aims to create an analysis of feature 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 , weather conditions, good management for emergency 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.

Data Description:

Link: https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=890001695
We 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.

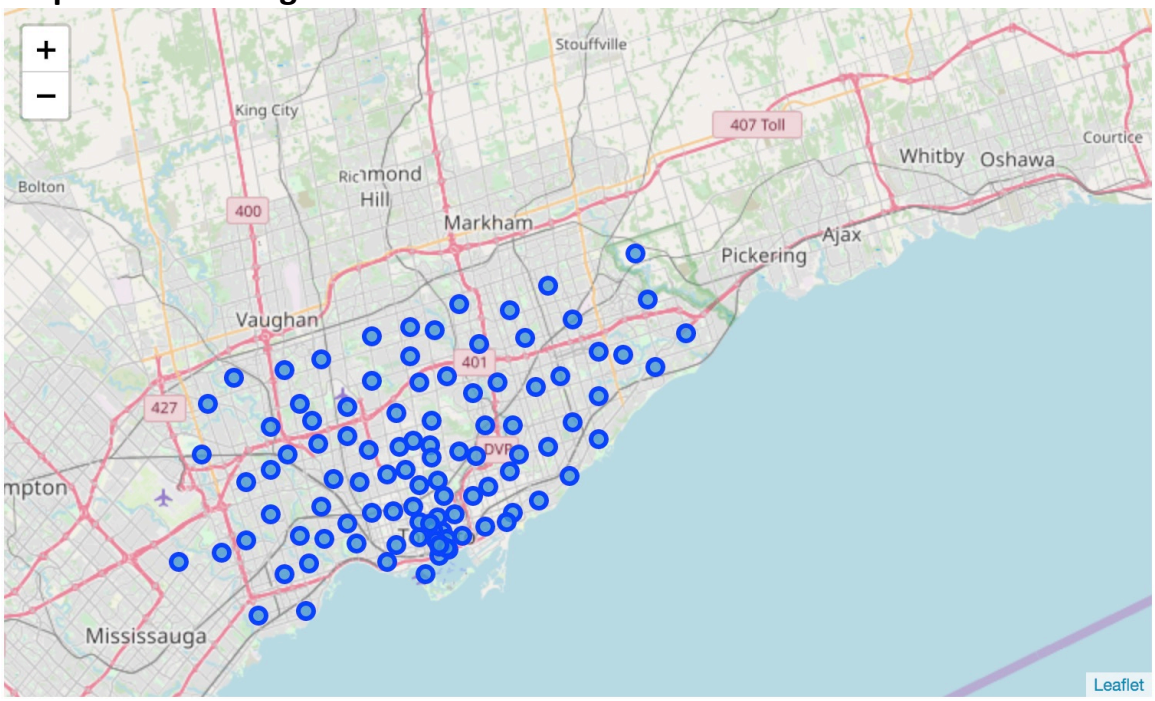
The data retrieved from foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

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

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.

Map of Scarborough



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 Toronto. To be able to do that, we need to cluster data, which is a form of unsupervised machine learning: k-means clustering algorithm.

Most Common Values

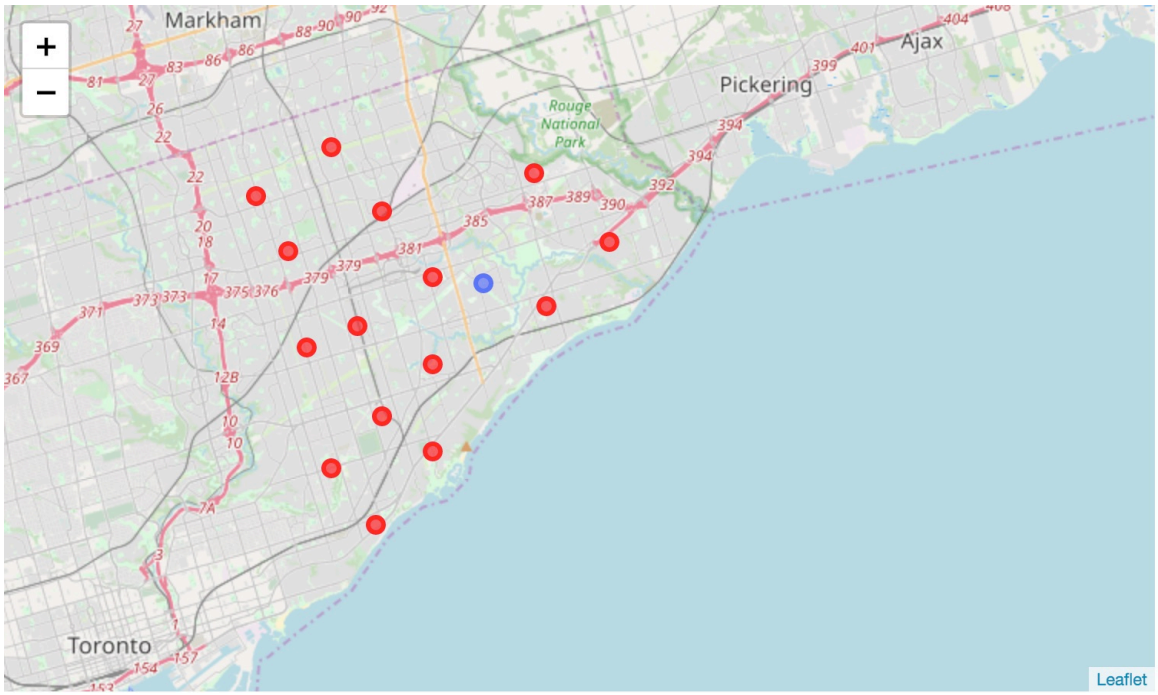
Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
Adelaide,King,Richmond	Coffee Shop	Restaurant	Café	Hotel	Theater	Gym	Breakfast Spot	Bakery
Agincourt	Breakfast Spot	Lounge	Shanghai Restaurant	Motorcycle Shop	Latin American Restaurant	Sandwich Place	Skating Rink	Pool Hall
Agincourt North,L'Amoreaux East,Milliken,Steel...	Pizza Place	Noodle House	BBQ Joint	Chinese Restaurant	Fast Food Restaurant	Pharmacy	Gym	Bakery
Albion Gardens,Beaumond Heights,Humbergate,Jam...	Grocery Store	Caribbean Restaurant	Beer Store	Fried Chicken Joint	Discount Store	Sandwich Place	Japanese Restaurant	Fast Food Restaurant
Alderwood,Long Branch	Pizza Place	Convenience Store	Skating Rink	Sandwich Place	Pharmacy	Pool	Coffee Shop	Gym

Work Flow

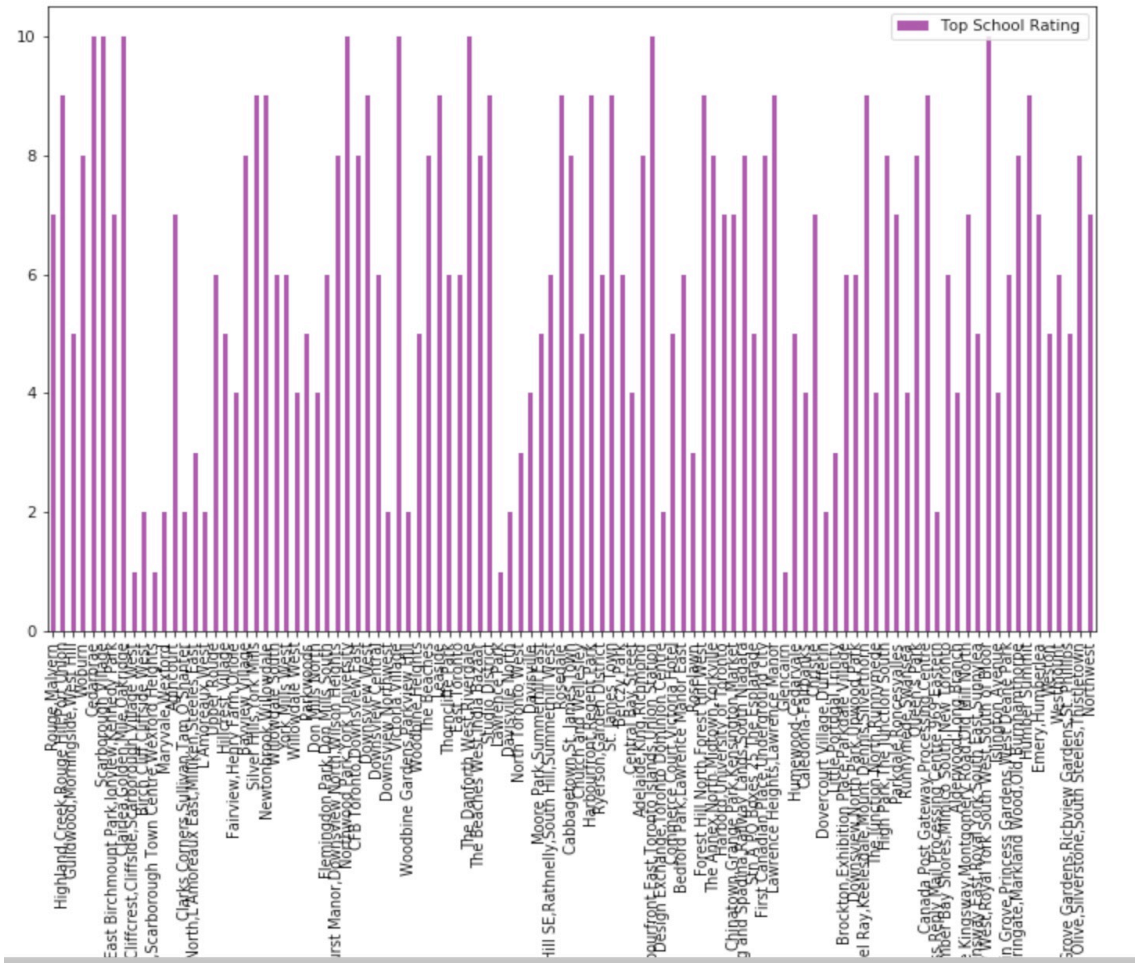
Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

Results Section

Map of Clusters in Scarborough



School ratings:



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 Capstone project have used 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.

Discussion Section:

Problem:

The major purpose of this project, is to suggest a better neighborhood in a new city for the person who are shifting there. Social presence in society in terms of like minded people. Sorted list of house in terms of housing prices in a ascending or descending order

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

Conclusion:

In this Capstone 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.

Libraries used:

Pandas: For creating and manipulating data frames.

Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Data.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

Matplotlib: Python Plotting Module.

Project:
https://github.com/kavsraj/Applied_Datascience_Capstone_Project/blob/master/The%20Battle%20of%20Neighborhoods.ipynb