

FINAL REPORT – THE BATTLE OF NEIGHBOURHOODS

[FINDING A BETTER PLACE IN SCARBOROUGH, TORONTO]

Introduction

The purpose of this project is to help people in exploring better facilities around their neighbourhood. It will help people making smart and efficient decision on selecting great neighbourhood out of numbers of other neighbourhoods in Scarborough, Toronto.

Lots of people are migrating to various states of Canada and needed lots of research for good housing prices and reputed schools for their children. This project is for those people who are looking for better neighbourhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, like-minded people, etc.

This Project aim to create an analysis of features for a people migrating to Scarborough to search a best neighbourhood as a comparative analysis between neighbourhoods. The features include median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, good management for emergency, water resources both fresh and waste water and excrement conveyed in sewers and recreational facilities.

It will help people to get awareness of the area and neighbourhood before moving to a new city, state, country or place for their work or to start a new fresh life.

Data Used

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

I will use Scarborough dataset, scrapped from Wikipedia in Week 3. This dataset consists of latitude and longitude, zip codes.

Foursquare API Data

We need data about different venues in different neighbourhoods of that specific borough. To retrieve this information, I'll 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, I'll connect to the Foursquare API to gather information about venues inside each and every neighbourhood. For each neighbourhood, we have chosen the radius to be 100 meters.

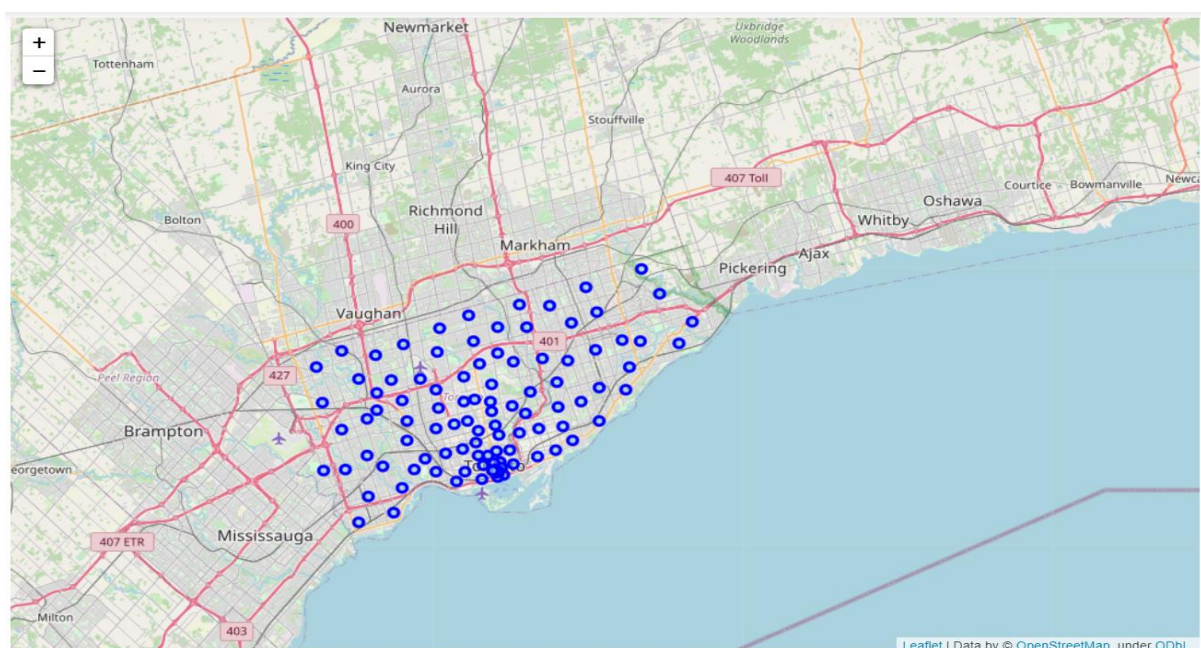
The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighbourhood
2. Neighbourhood Latitude
3. Neighbourhood 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

```
1 df.head()
```

	Postalcode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.81139	-79.19662
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78574	-79.15875
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.76575	-79.17470
3	M1G	Scarborough	Woburn	43.76812	-79.21761
4	M1H	Scarborough	Cedarbrae	43.76944	-79.23892

Map of Scarborough



Methodology

K-Means Clustering

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.

Most Common Venues

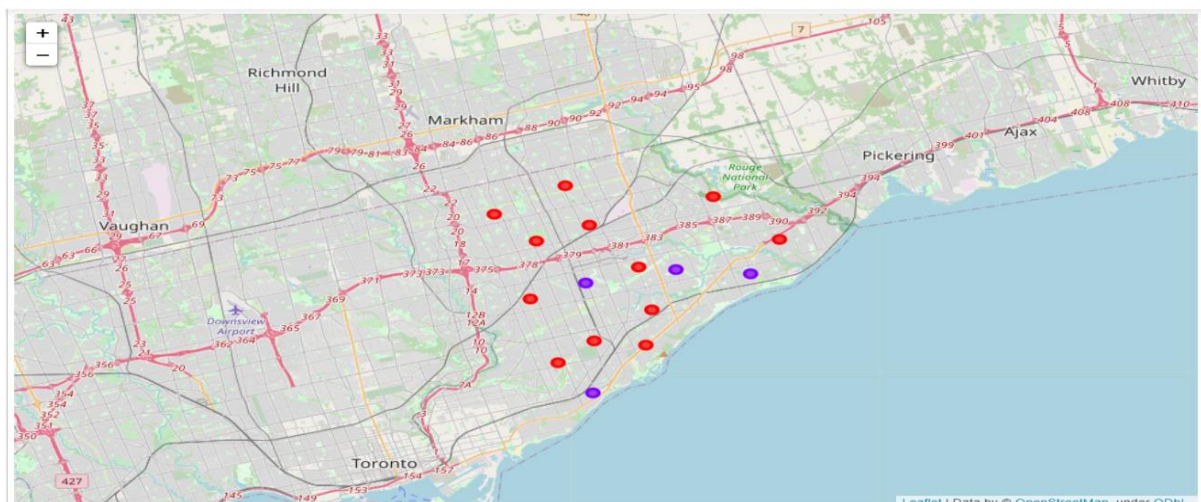
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Chinese Restaurant	Shopping Mall	Pizza Place	Grocery Store	Sandwich Place	Sushi Restaurant	Supermarket	Bank	Bakery	Japanese Restaurant
1	Alderwood, Long Branch	Print Shop	Pub	Convenience Store	Pizza Place	Coffee Shop	Gas Station	Pharmacy	Gym	Sandwich Place	Eastern European Restaurant
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Park	Pizza Place	Grocery Store	Sushi Restaurant	Sandwich Place	Intersection	Deli / Bodega	Restaurant	Fried Chicken Joint
3	Bayview Village	Park	Asian Restaurant	Trail	Gas Station	Dog Run	Farmers Market	Farm	Falafel Restaurant	Event Space	Ethiopian Restaurant
4	Bedford Park, Lawrence Manor East	Italian Restaurant	Coffee Shop	Sandwich Place	Boutique	Comfort Food Restaurant	Pharmacy	Pet Store	Pub	Restaurant	Café

Work Flow:

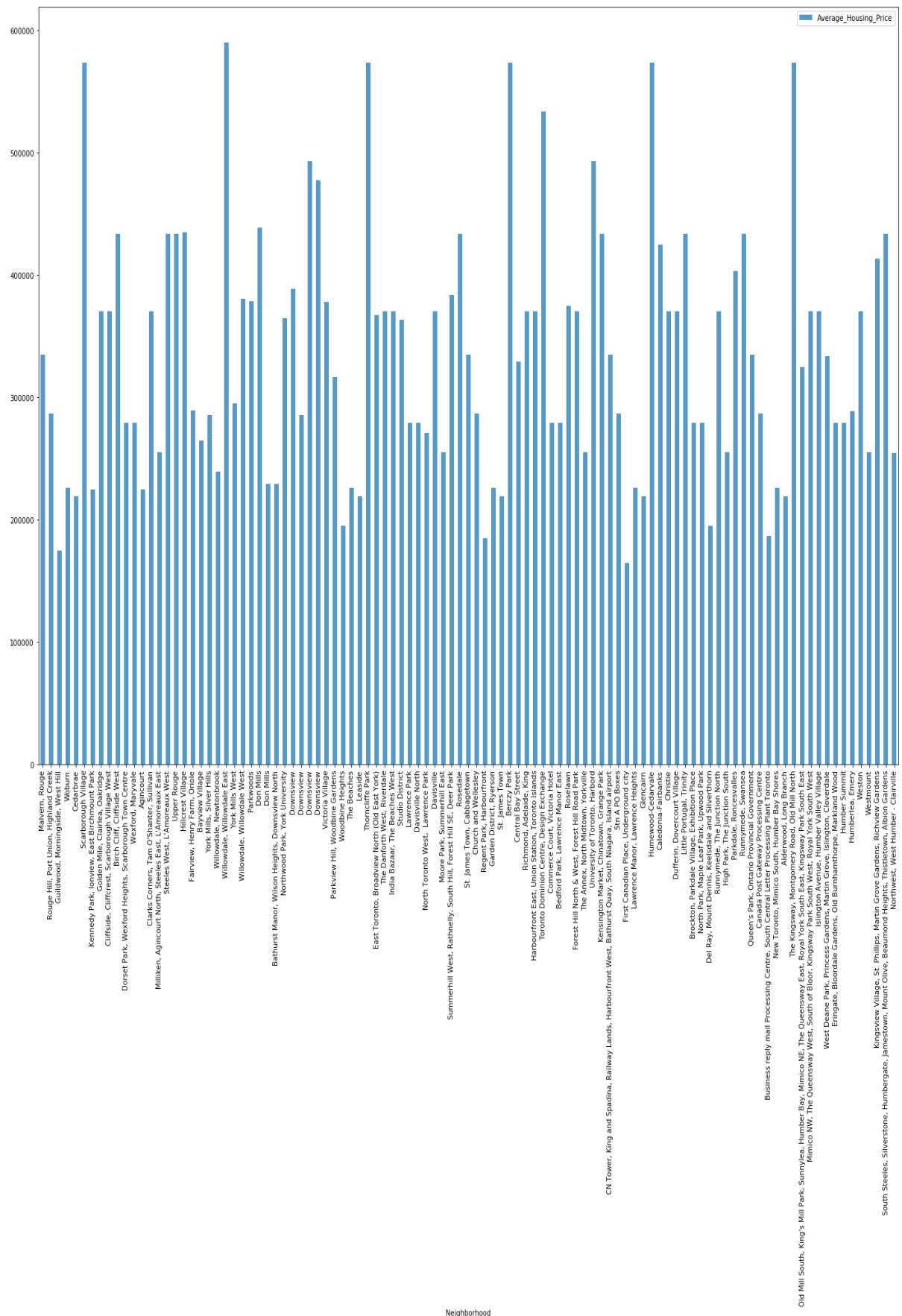
Using credentials of Foursquare API features of near-by places of the neighbourhoods would be mined. Due to http request limitations the number of places per neighbourhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

Results

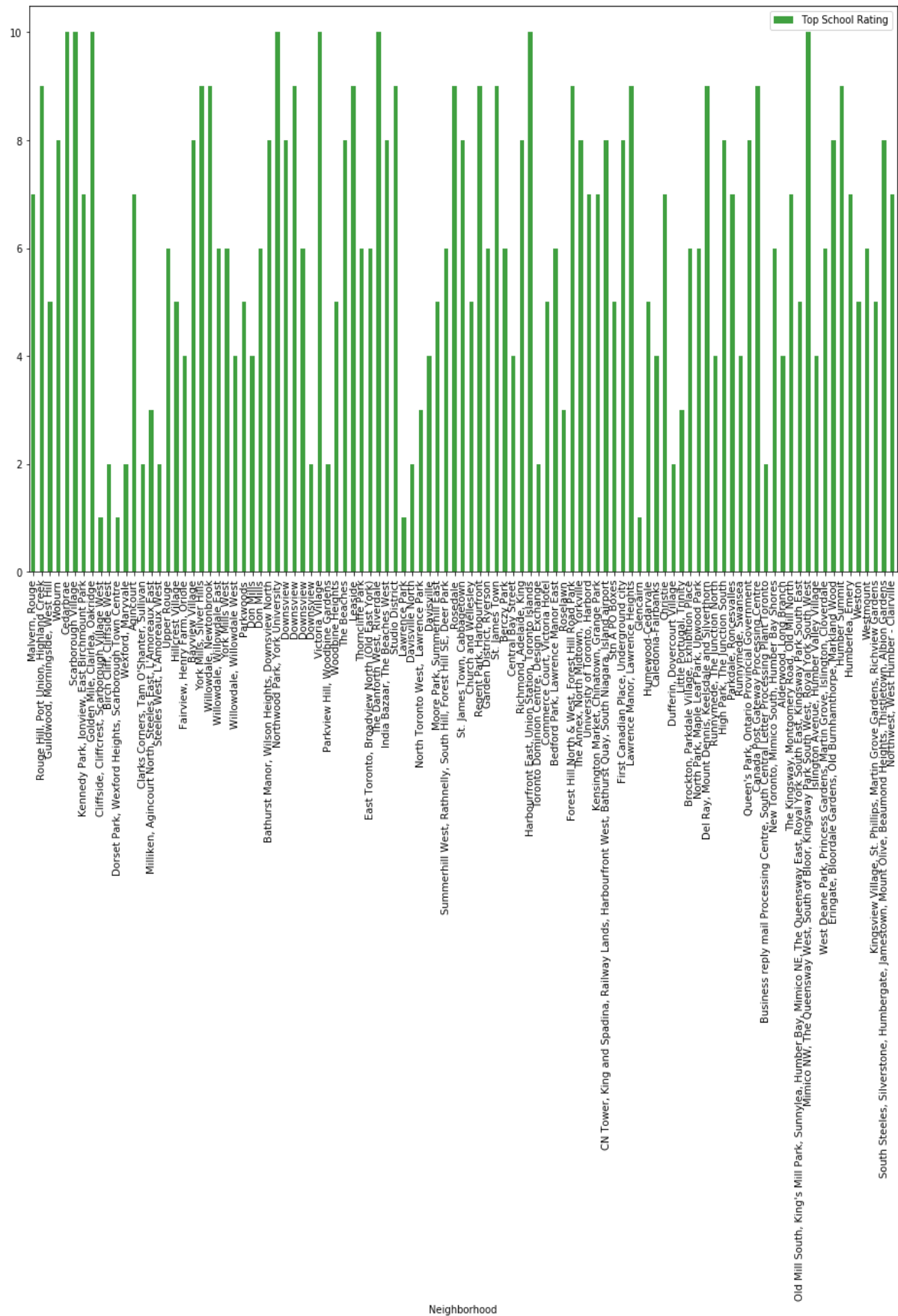
Map of Clusters in Scarborough



Average Housing Price by clusters in Scarborough



School Ratings by clusters in Scarborough



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 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

Problem which I tried to Solve

The major purpose of this project, is to suggest a better neighbourhood in a new city for the person who are shifting 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 houses in terms of housing prices in a ascending or descending order
2. Sorted list of schools in terms of location, fees, rating and reviews

Conclusion

In this project, using k-means cluster algorithm I separated the neighbourhood into 10(Ten) different clusters and for 103 different latitude and longitude from dataset, which have very-similar neighbourhoods around them. Using the charts above results presented to a particular neighbourhood 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.

Libraries Used

Pandas: For creating and manipulating dataframes.

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 and library to handle http requests.

Matplotlib: Python Plotting Module.