

Capstone Project – The Battle of Neighborhoods

**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.
- This project's 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.



TARGET GROUPS

- Fresh immigrants to Canada.
- People migrating to various states of Canada.
- People not familiar with their surroundings

DATA SECTION

- Data Link:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- I will use Scarborough dataset which is scrapped from Wikipedia.
- Dataset consisting of latitude and longitude, zip codes.
- Foursquare API data

FOURSQUARE API

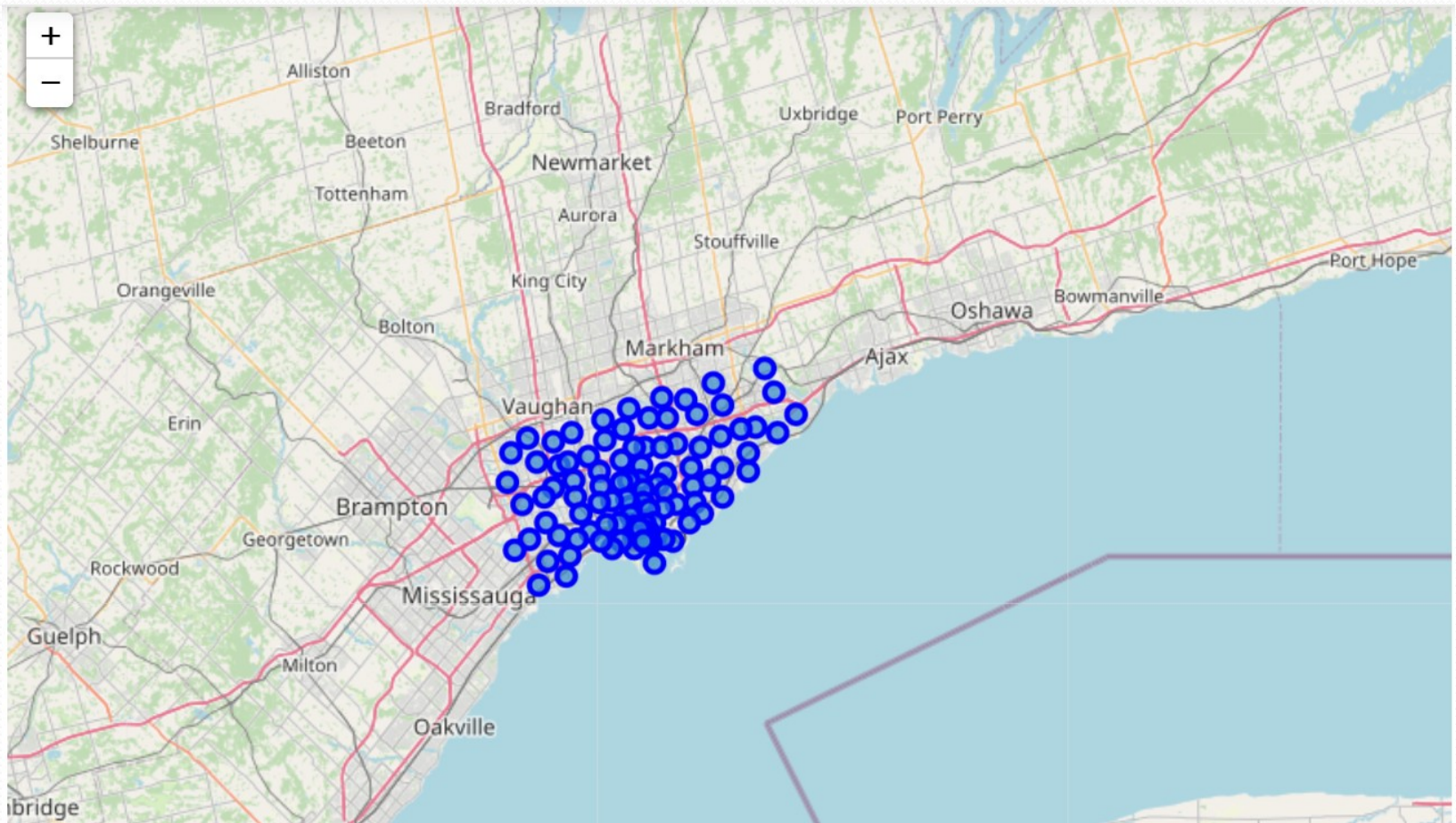
- Foursquare is a location data provider with information about all manner of venues and events within an area of interest.
- Such information includes venue names, locations, menus and even photos.
- As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

FOURSQUARE API DATA

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:

- Neighbourhood
- Neighbourhood Latitude
- Neighbourhood Longitude
- Venue
- Name of the venue e.g. the name of a store or restaurant
- Venue Latitude
- Venue Longitude
- Venue Category

Map of Scarborough



METHODOLOGY

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.

K-MEANS CLUSTERING

Out[119]:

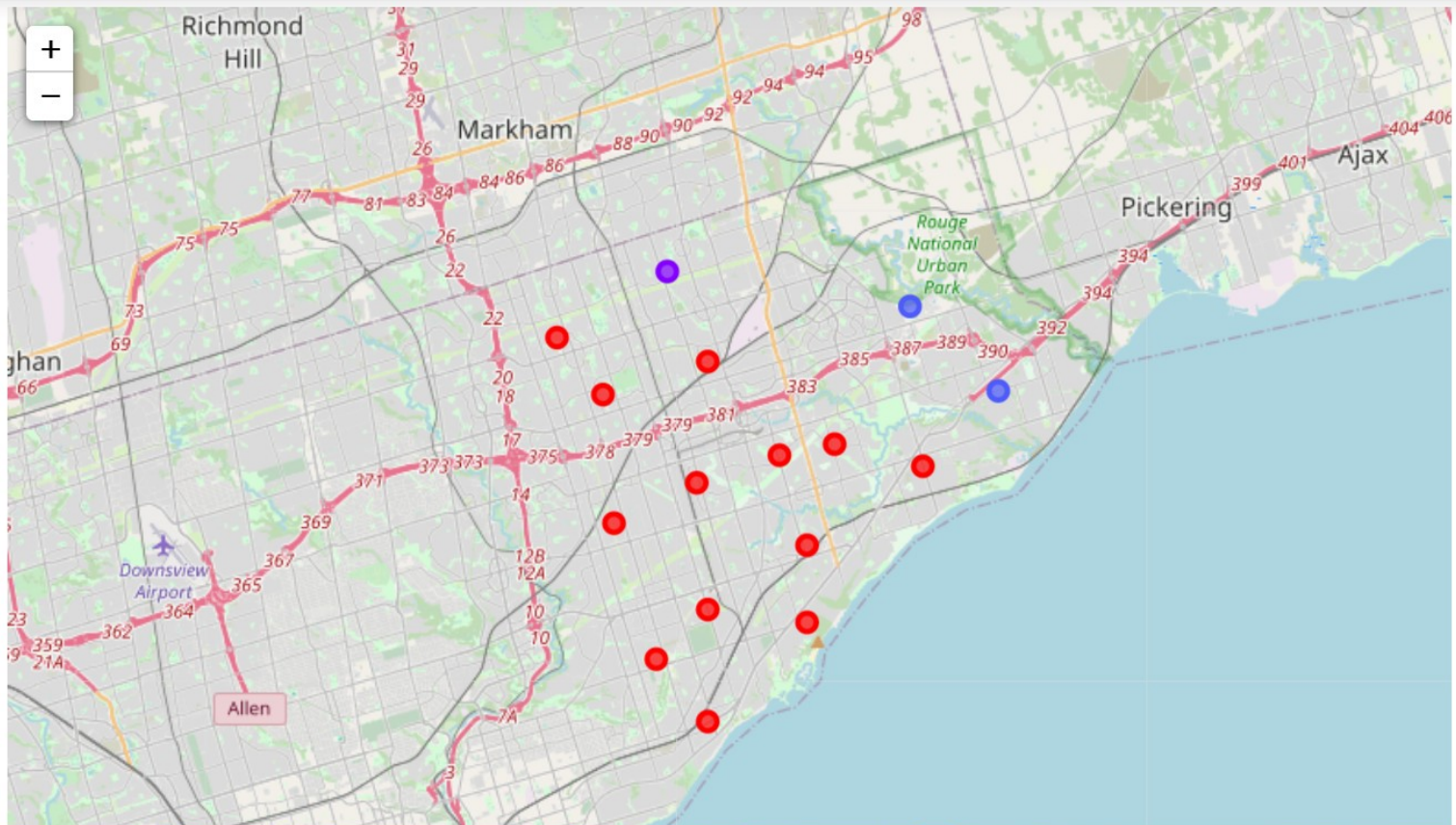
	Postalcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	M1B\n	Scarborough\n	Malvern, Rouge	43.808626	-79.189913	2	Park	Trail	Women's Store	Ethiopian Restaurant	Café
1	M1C\n	Scarborough\n	Rouge Hill, Port Union, Highland Creek	43.785779	-79.157368	2	Fish & Chips Shop	History Museum	Park	Bar	Worship
2	M1E\n	Scarborough\n	Guildwood, Morningside, West Hill	43.765806	-79.185284	0	Pizza Place	Fast Food Restaurant	Restaurant	Coffee Shop	
3	M1G\n	Scarborough\n	Woburn	43.771545	-79.218135	0	Coffee Shop	Insurance Office	Business Service	Park	Dumpling Restaurant
4	M1H\n	Scarborough\n	Cedarbrae	43.768791	-79.238813	0	Bank	Hakka Restaurant	Thai Restaurant	Athletics & Sports	Store

WORKFLOW

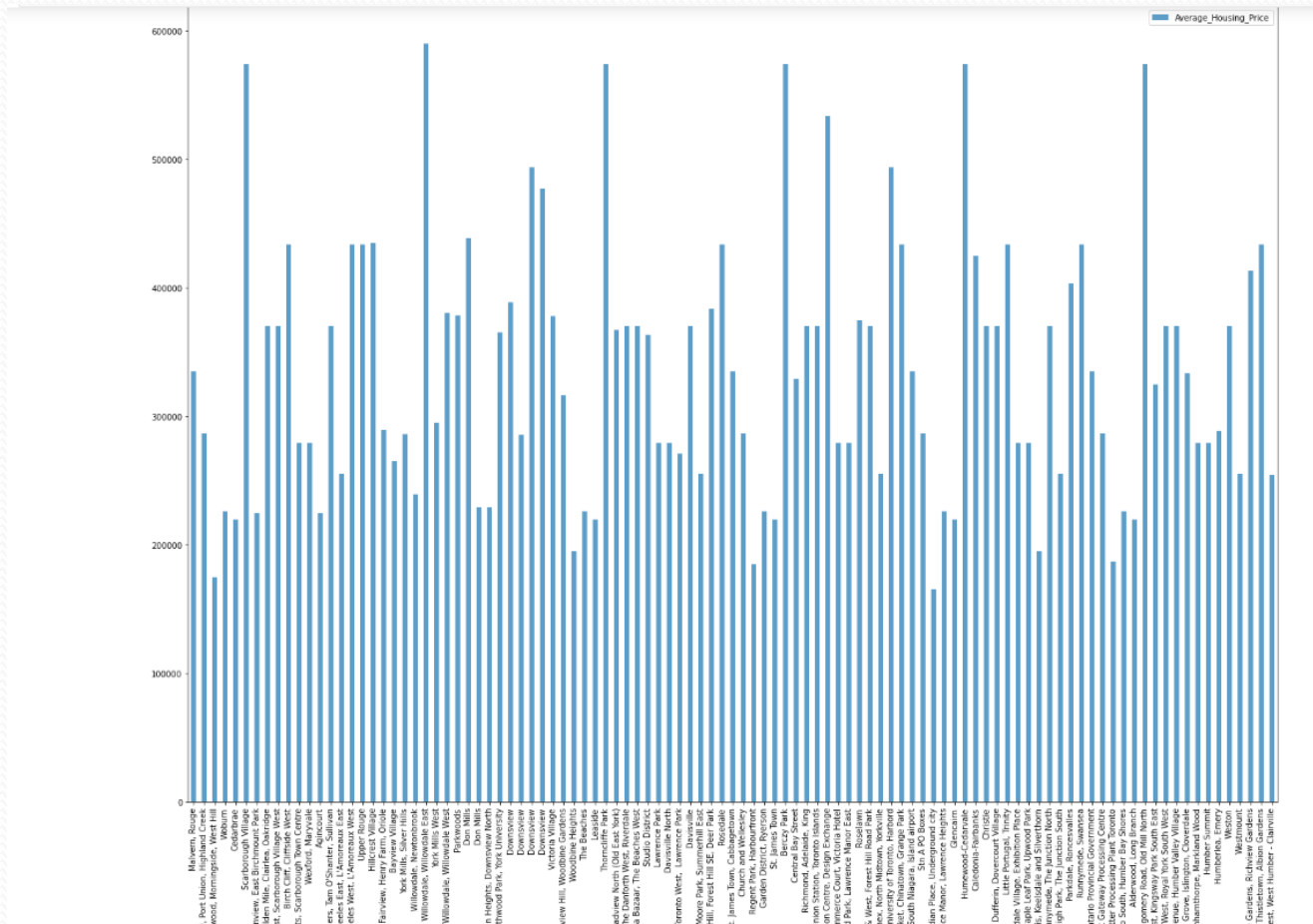
- Using credentials of Foursquare API features of near-by places of the neighbourhoods would be mined.
- Due to http request limitations the number of places per neighbourhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

RESULTS-Map of clusters in Scarborough

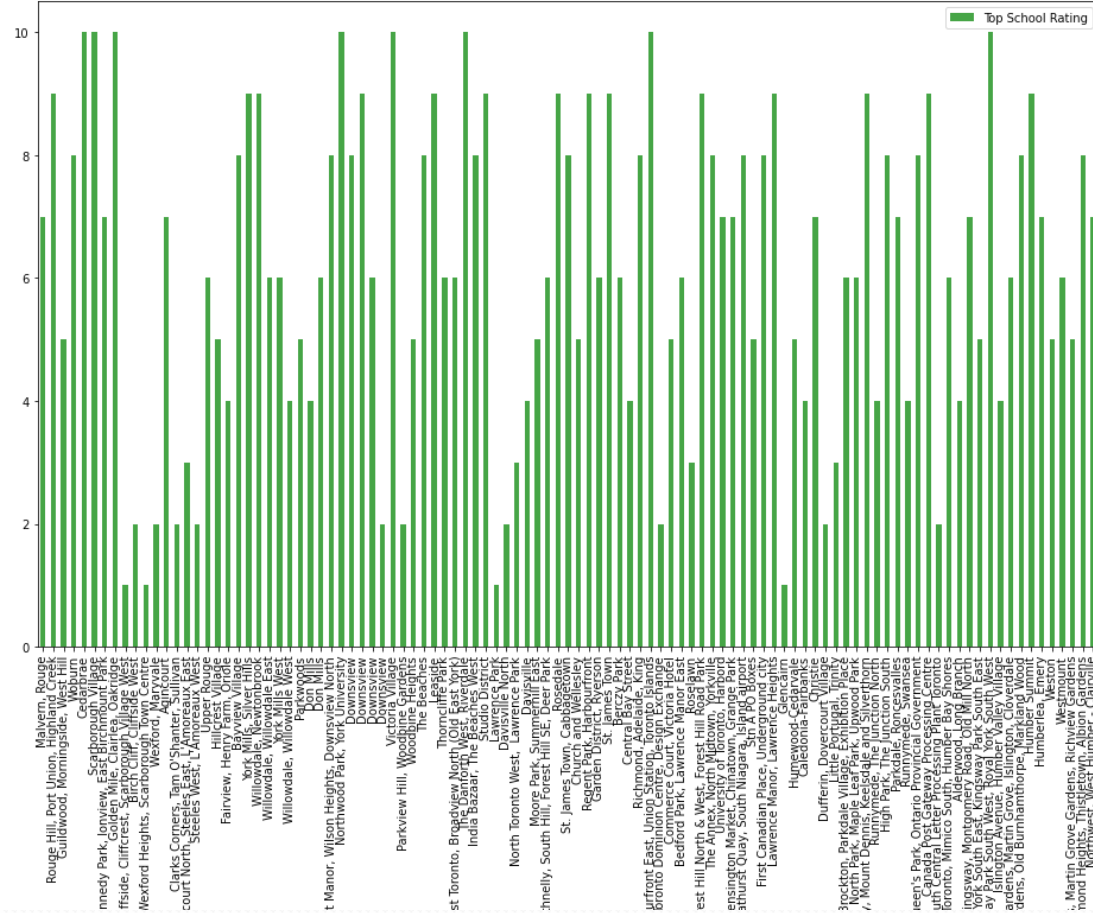
Out[120]:



RESULTS- Average Housing Price by Clusters in Scarborough



RESULTS- School Ratings by Clusters in Scarborough



DISCUSSION

The major purpose of this project, is to suggest a better neighbourhood in a new city for the people who are shifting there.

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

- 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 review

CONCLUSION

- In this Capstone 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 for a particular neighbourhood based on average house prices and school rating have been made.

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