

# **Report: Capstone Project - The Battle of Neighborhoods (Week 2)**

## **Finding a Better Place in Scarborough, Toronto**

### **1. Introduction:**

Toronto is the largest city in Canada with 2.6 million people and is the capital of the province of Ontario. About one third of the country's population growth has lived in this metropolitan area in recent years. The population of the metropolitan area (census area) increased from 4.1 million in 1992 to 5.6 million in 2011. The Greater Toronto Area had more than 6.2 million in 2010.

#### **Business Problems:**

The main purpose of this project is to propose a better neighborhood in a new city for people moving there. Social presence in society about like-minded people. Connect to the airport, bus stop, city center, markets, and other daily necessities nearby.

- List sorted of houses by housing price in ascending or descending order
- Sort the list of schools by location, fee, rating, and rating

### **2. Data Section**

#### **Data Source:**

I will reuse the data source already implemented in the project in Week 3

Data Link: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

#### **Foursquare API:**

I will need data about different locations in different areas of that particular county. To get that information, we will use the location information "Foursquare". Foursquare is a location data provider with information about all venues and events in an area of interest. This information includes location names, locations, menus, and even photos. Therefore, the positional platform of 5.0 will be used as the only data source since all of the required mandatory information can be obtained through the API.

After finding a list of neighborhoods, we then connected to the Foursquare API to collect information about the locations within each neighborhood. For each neighborhood, we chose a radius of 100 meters.

The data retrieved from Foursquare contains information about places within a specified distance of longitude and latitude of postal code. Information collected on each location is as follows:

### 1. Neighborhood

- Neighborhood latitude
- Neighborhood longitude

### 2. Venue

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

### Libraries Which are Used to Develop the Project:

Pandas: For creating and manipulating dataframes.

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

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

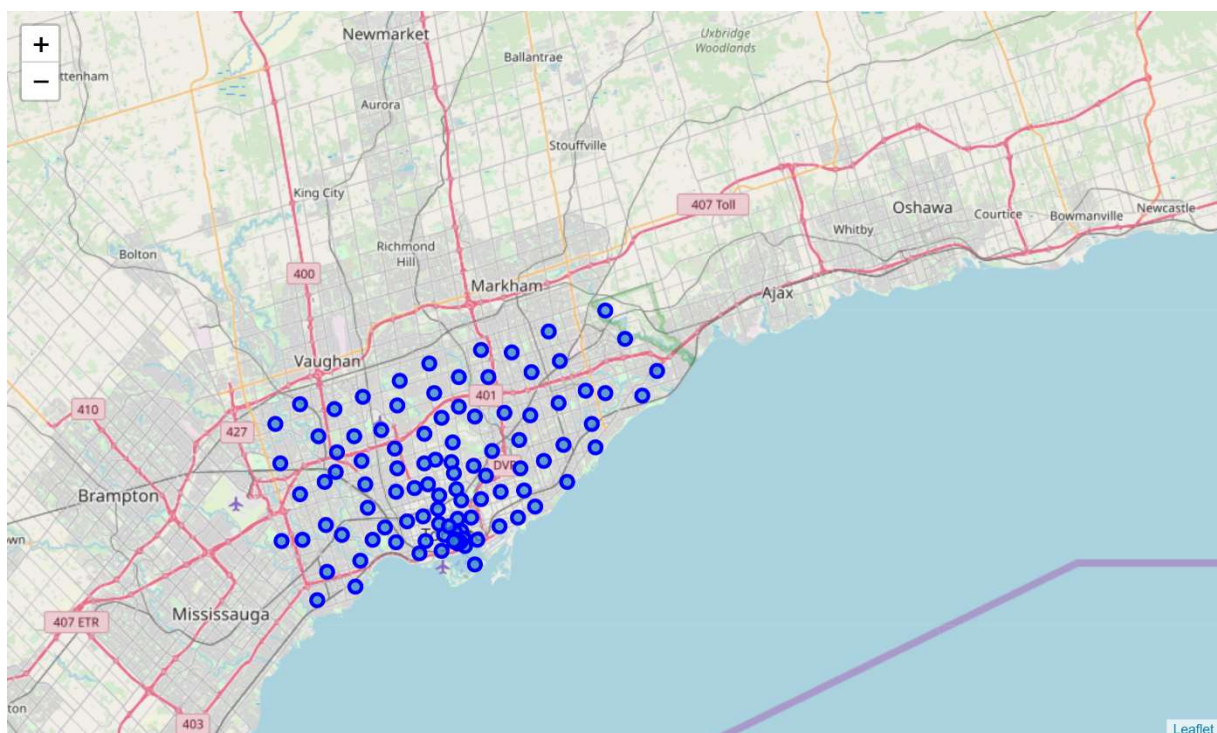
XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Data.

Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.

### Map of Scarborough



### 3. 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

	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	6th Most Common Venue
0	M1B	Scarborough	Rouge, Malvern	43.811525	-79.195517	2	Construction & Landscaping	Climbing Gym	Fast Food Restaurant	Zoo Exhibit	African Restaurant	Food Court
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.785665	-79.158725	0	Moving Target	History Museum	Bar	Yoga Studio	Doner Restaurant	Donut Shop
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.765815	-79.175193	1	Park	Gym / Fitness Center	Athletics & Sports	Yoga Studio	Doner Restaurant	Donut Shop
3	M1G	Scarborough	Woburn	43.768369	-79.217590	0	Coffee Shop	Park	Fast Food Restaurant	Business Service	Falafel Restaurant	Donut Shop
4	M1H	Scarborough	Cedarbrae	43.769688	-79.239440	0	Flower Shop	Indian Restaurant	Bank	Thai Restaurant	Bakery	Caribbean Restaurant

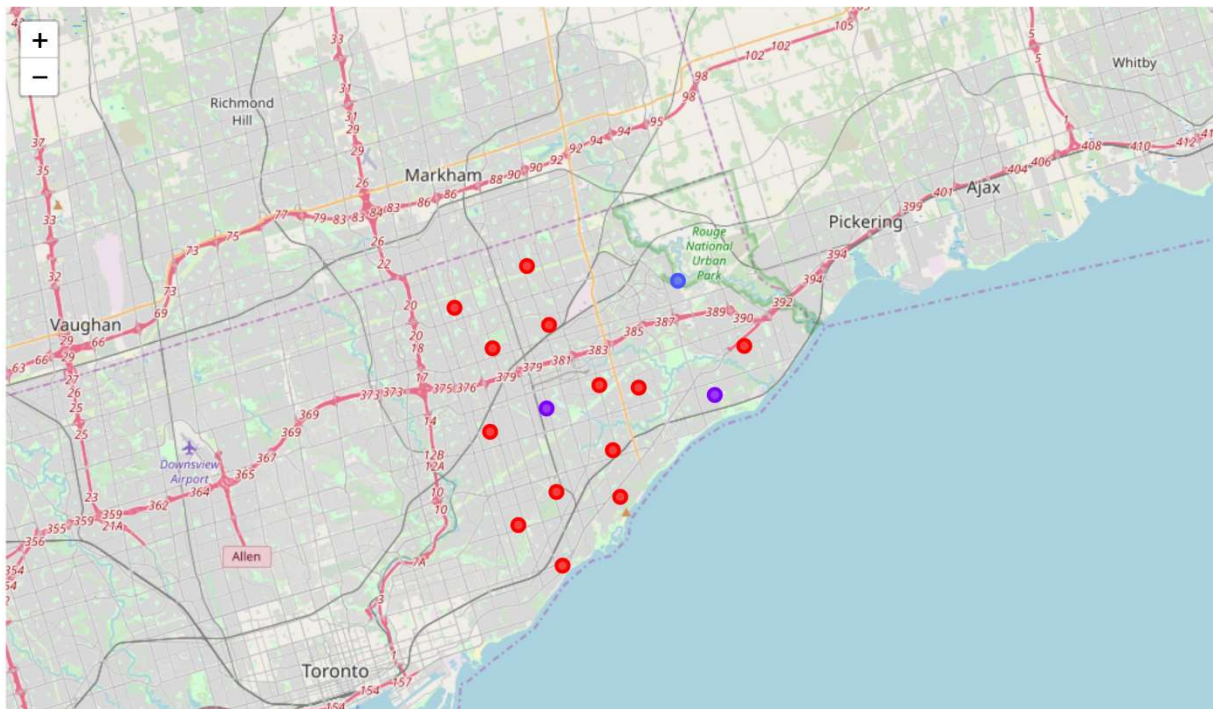
#### Most Common venues near Neighborhood

	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	Adelaide, King, Richmond	Coffee Shop	Restaurant	Café	Hotel	Breakfast Spot	Seafood Restaurant	Gastropub	Steakhouse	Gym	Sushi Restaurant
1	Agincourt	Chinese Restaurant	Pizza Place	Shopping Mall	Pool	Malay Restaurant	Breakfast Spot	Coffee Shop	Japanese Restaurant	Bank	Bakery
2	Agincourt North, L'Amoreaux East, Milliken, St...	Pharmacy	Sandwich Place	Yoga Studio	Event Space	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant
3	Albion Gardens, Beaumont Heights, Humbertgate, ...	Grocery Store	Beer Store	Japanese Restaurant	Auto Garage	Caribbean Restaurant	Gym Pool	Park	Hardware Store	Discount Store	Sandwich Place
4	Alderwood, Long Branch	Pizza Place	Pool	Sandwich Place	Coffee Shop	Gym	Print Shop	Pub	Pharmacy	Gas Station	Event Space

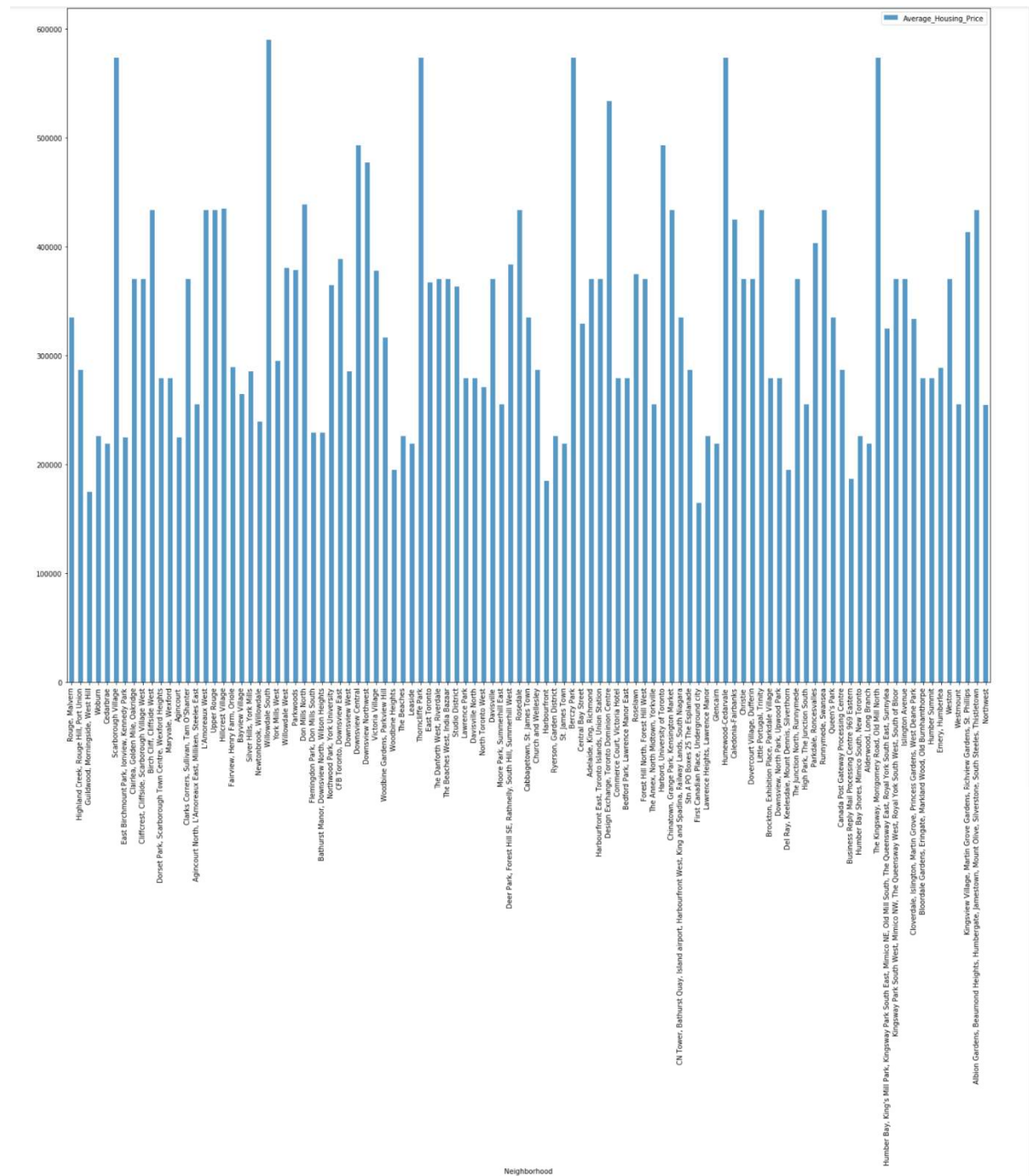
**Work Flow:** Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

# 4. Results Section

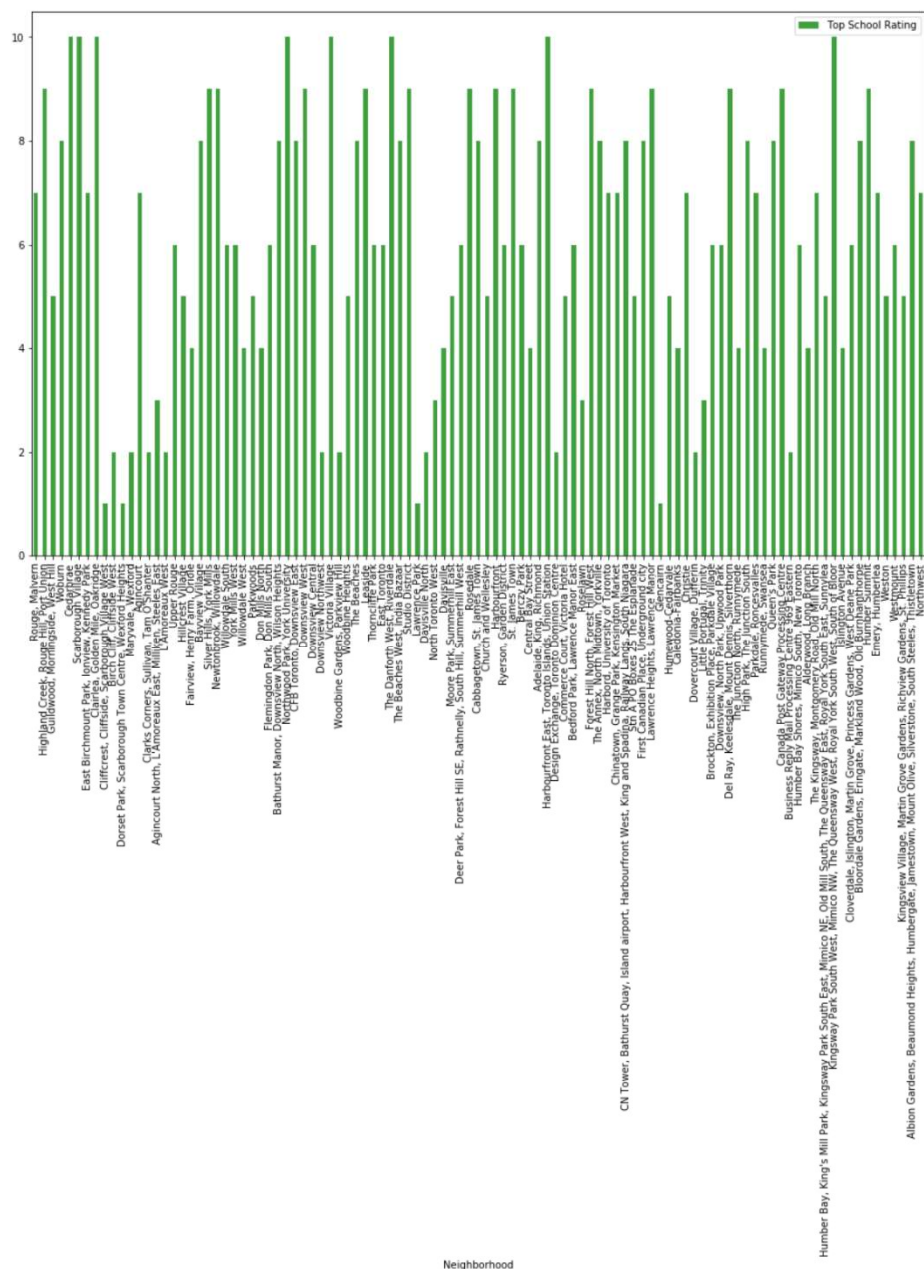
Map of Clusters in Scarborough



## Average Housing Price by Clusters in Scarborough



## School Ratings by Clusters in Scarborough



**Location:**

Scarborough is a popular destination for new immigrants in Canada to reside in. As a result, it is one of the most multicultural and multicultural areas in the Greater Toronto Area, home to various religious groups and places of worship. Although immigration has become a hot topic in the past few years with many governments seeking more restrictions on immigrants and refugees, the general trend of immigration into Canada is one of the increases.

## **5. Conclusion Section**

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.

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.