# Finding a Better Place in Scarborough, Toronto

July 14, 2021

## 1. Introduction:

The purpose of this project is to help people in exploring better facilities around their neighborhood. It will help people make an innovative and efficient decision to select a great community out of numbers of other areas in Scarborough, Toronto.

Many people migrate to various states of Canada and needed lots of research for reasonable housing prices and reputed schools for their children. This project is for better neighborhoods to ease access to cafes, schools, supermarkets, medical shops, grocery shops, malls, theatres, hospitals, like-minded people, etc.

This project analyzes features for people migrating to Scarborough to search for the best neighborhood comparative analysis between communities. 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 an emergency, water resources, fresh and wastewater, and excrement conveyed in sewers and recreational facilities.

It will help people understand the area and neighborhood before moving to a new city, state, country, or place for their work or starting a new fresh life.

## 2. Problem Description:

The primary purpose of this project is to suggest a better neighborhood in a new city for the person who is 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 nearby.

1. Sorted list of the house in terms of housing prices in an ascending or descending order
2. Sorted list of schools in terms of location, fees, rating, and reviews

## 3. Data Description:

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

We will use the Scarborough dataset, which we scrapped from Wikipedia on Week 3. Dataset consists of latitude and longitude, zip codes.

## Foursquare API Data:

We will need data about different venues in different neighborhoods of that specific borough. 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. The foursquare location platform will be used as the sole data source since I can obtain all the required information through the API.

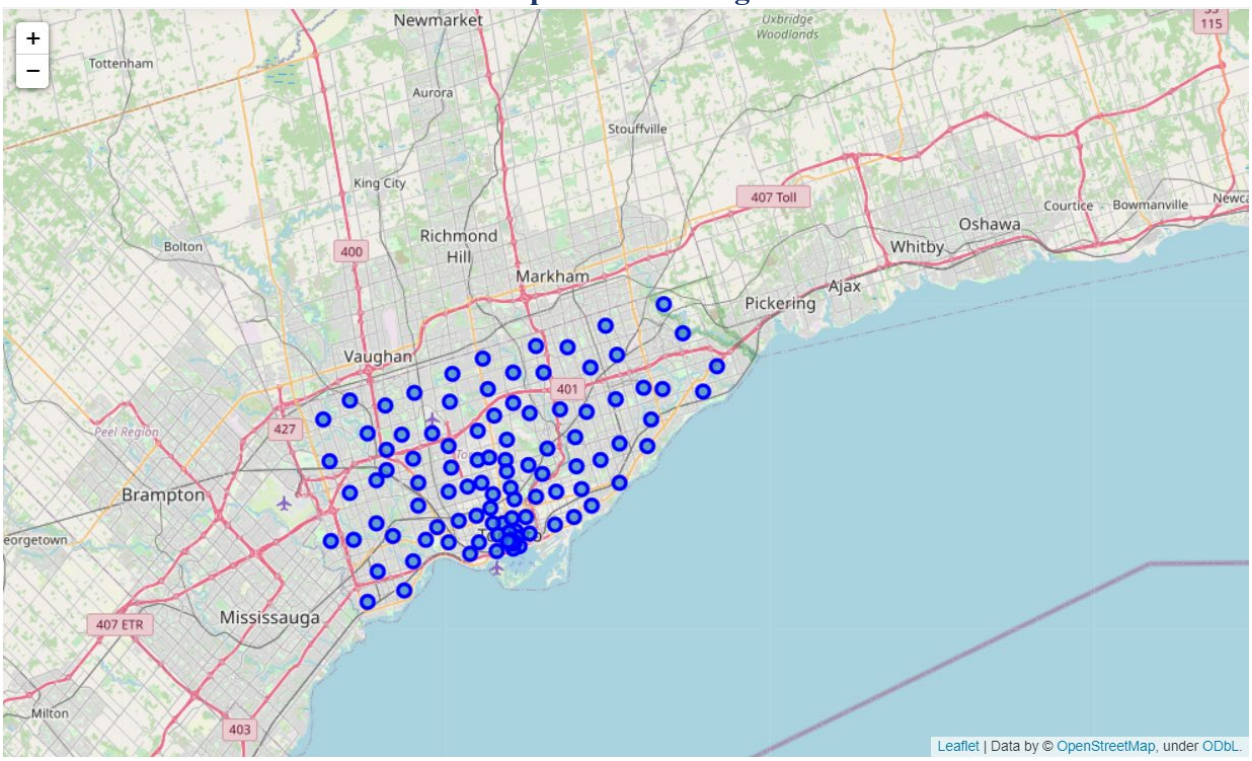
This project would use Four-square API as its prime data gathering source. It has a database of millions of places, especially their places API, which can perform location search, location sharing, and details about a business.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside every area. For each community, 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. 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

### Map of Scarborough




## 4. Methodology:

### Clustering Approach:

To compare the two cities' similarities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar areas in big cities like New York and Toronto. 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 in Scarborough

| 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 | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Malvern, Rouge | 43.81139 | -79.19662 | 0 | Zoo Exhibit | Fast Food Restaurant | Electronics Store | Dive Bar | Dog Run | Doner Restaurant | Donut Shop | Dry Cleaner | Dumpling Restaurant | Eastern European Restaurant |
| Rouge Hill, Port Union, Highland Creek | 43.78574 | -79.15875 | 0 | Fish & Chips Shop | Home Service | Bar | Electronics Store | Dog Run | Doner Restaurant | Donut Shop | Dry Cleaner | Dumpling Restaurant | Eastern European Restaurant |

**A Popular Zoo at Malvern, Scarborough**

**A Popular Fish & Chips Shop at Rouge Hill, Scarborough**

### Most Common Venues near Neighborhood

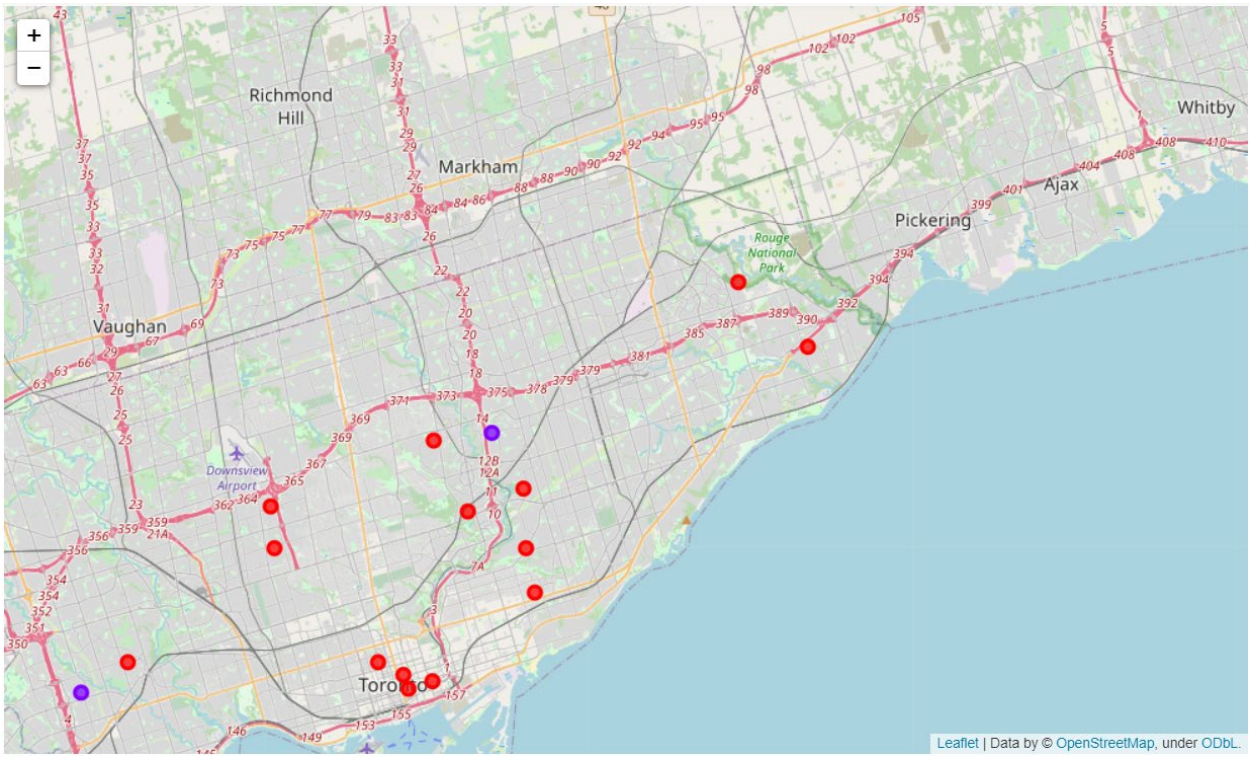
| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Agincourt | Shopping Mall | Coffee Shop | Pool | Bubble Tea Shop | Skating Rink | Latin American Restaurant | Sandwich Place | Supermarket | Sushi Restaurant | Bank |
| 1 | Alderwood, Long Branch | Gas Station | Pub | Sandwich Place | Dance Studio | Gym | Pizza Place | Pharmacy | Coffee Shop | Comic Shop | Dumpling Restaurant |
| 2 | Bathurst Manor, Wilson Heights, Downsview North | Park | Pizza Place | Coffee Shop | Mobile Phone Shop | Sandwich Place | Intersection | Deli / Bodega | Restaurant | Sushi Restaurant | Fried Chicken Joint |
| 3 | Bayview Village | Gas Station | Park | Flower Shop | Trail | Asian Restaurant | Dive Bar | Dog Run | Doner Restaurant | Donut Shop | Dry Cleaner |
| 4 | Bedford Park, Lawrence Manor East | Sandwich Place | Italian Restaurant | Coffee Shop | Liquor Store | Thai Restaurant | Juice Bar | Sushi Restaurant | Sports Club | Fast Food Restaurant | Pizza Place |

## Workflow:

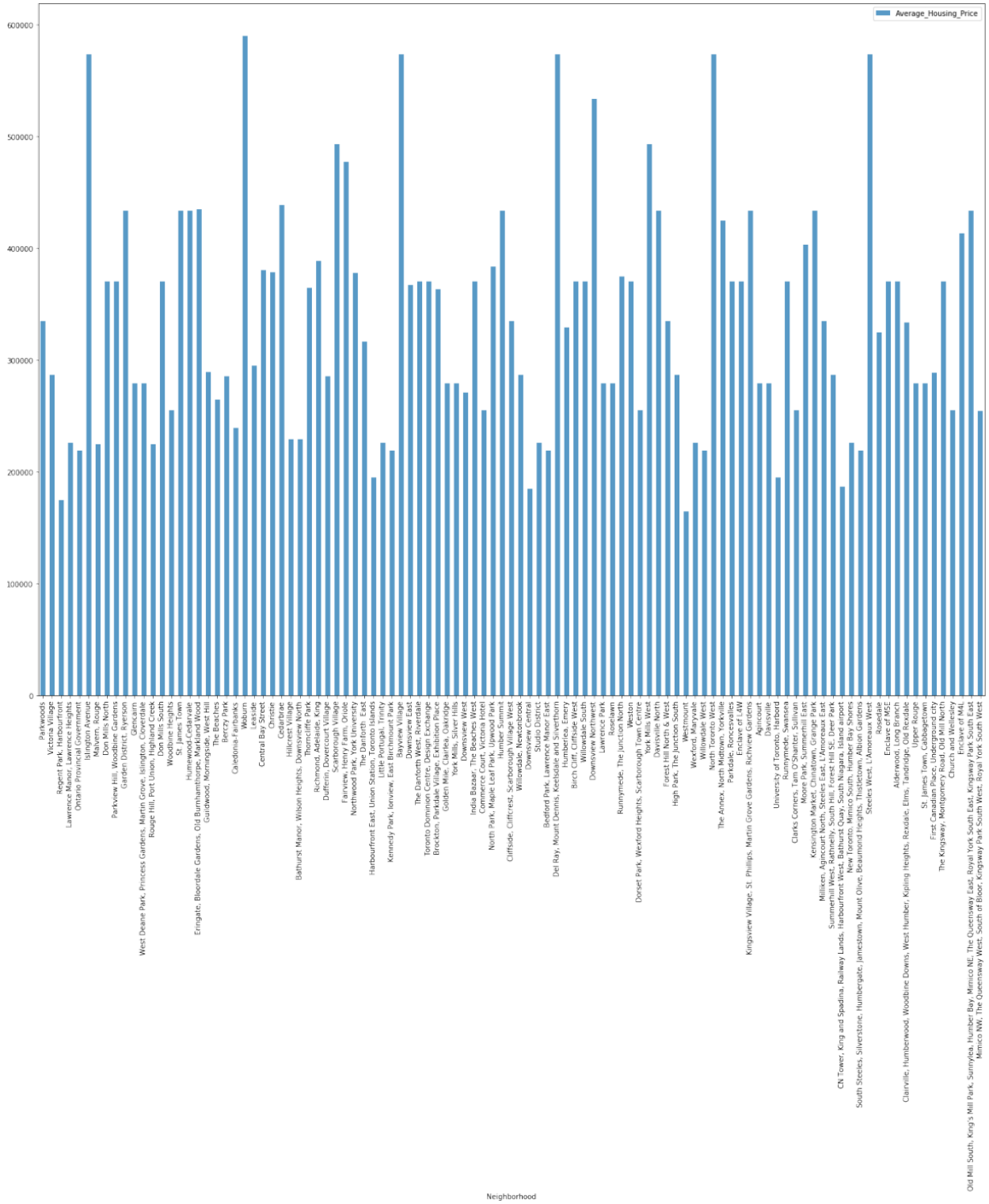
Using credentials of Foursquare API, I would mine features of nearby places of the neighborhoods. Due to HTTP request limitations, I would reasonably set the number of spots per neighborhood parameter to 100. The radius parameter would be set to 500.

# 5. Results:

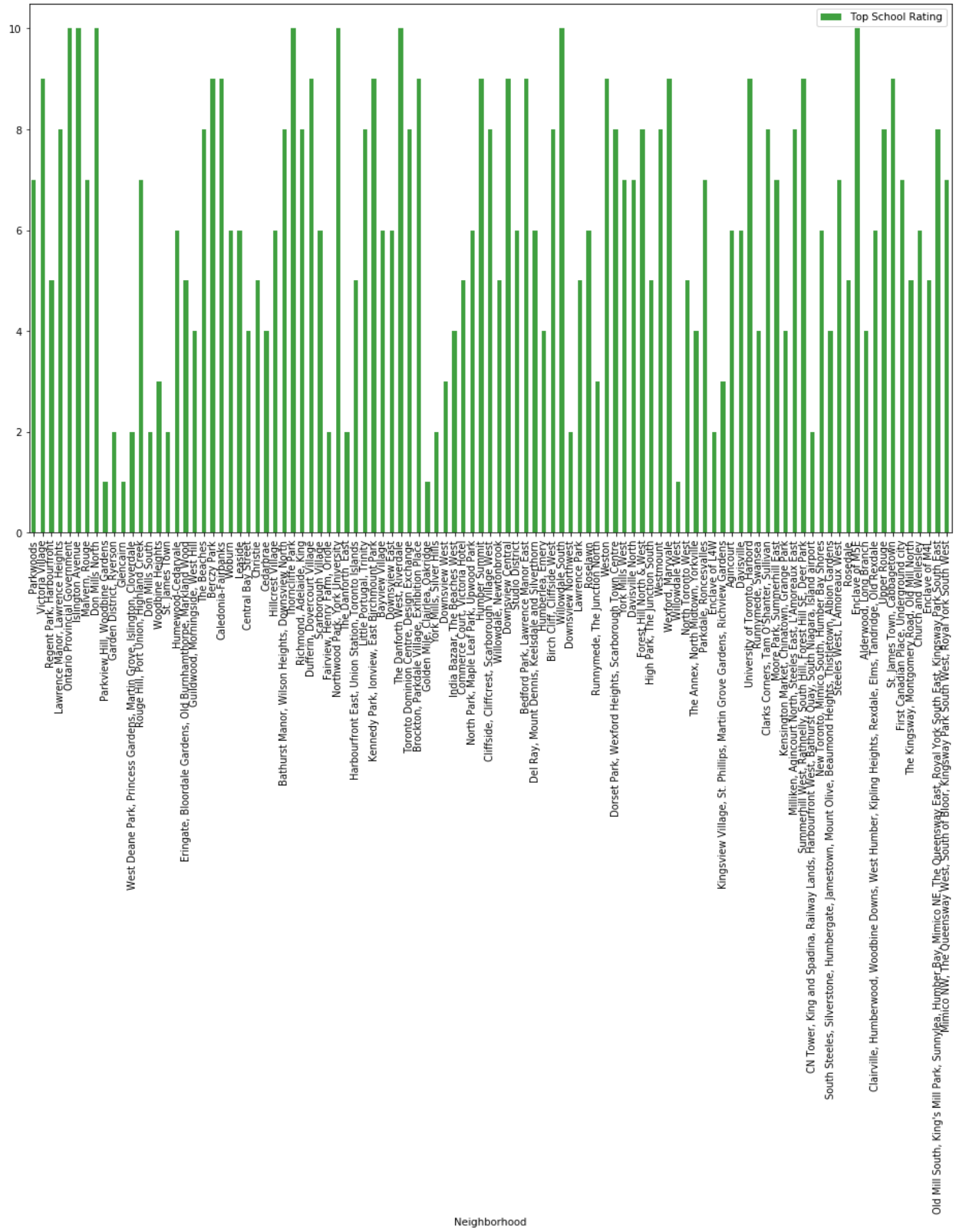
### Map of Clusters in Scarborough



# 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 on the rise.

## 6. Conclusion

In this Capstone project, using the k-means cluster algorithm, I separated the neighborhood into ten different clusters and 103 different latitudes and longitude from the dataset, which has very similar neighborhoods. Using the charts above, results presented to a particular neighborhood based on average house prices and school ratings 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 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 Capstone project can be continued to make it more precise to find the best house in Scarborough. Best means based on all required things(daily needs or things we need to live a better life) and cost-effectiveness.

### Libraries used to Develop the Project:

Pandas: For creating and manipulating data frames.
Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution using an 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.

LinkedIn Article: https://www.linkedin.com/pulse/finding-better-place-scarborough-toronto-kush-adhvaryu/