



# Capstone Project

## The Battle of Neighborhoods

NARGES FAROKHSHAD

# Introduction

- ▶ The purpose of this Project is to help people in exploring better facilities around their neighborhoods in Scarborough, Toronto.
- ▶ This Project aims to create an analysis of features for people migrating to Scarborough to search for a best neighborhood as a comparative analysis between neighborhoods.
- ▶ The features include median housing price and better schools 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.

# Data Section

- ▶ Will use Scarborough dataset which we scrapped from wikipedia on Week 3.
- ▶ Dataset consisting of latitude and longitude, zip codes.
- ▶ 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 Data

- ▶ We will need data about different venues in different neighborhoods of that specific borough.
- ▶ Foursquare is a location data provider with information about all manner of venues and events within an area of interest.
- ▶ 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

- ▶ 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

100



# Methodology Section

- ▶ **Clustering Approach**
- ▶ **Using K-Means Clustering Approach**
- ▶ **Most Common venues near Neighborhood**

# Workflow

- ▶ Using credentials of Foursquare API features of nearby 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



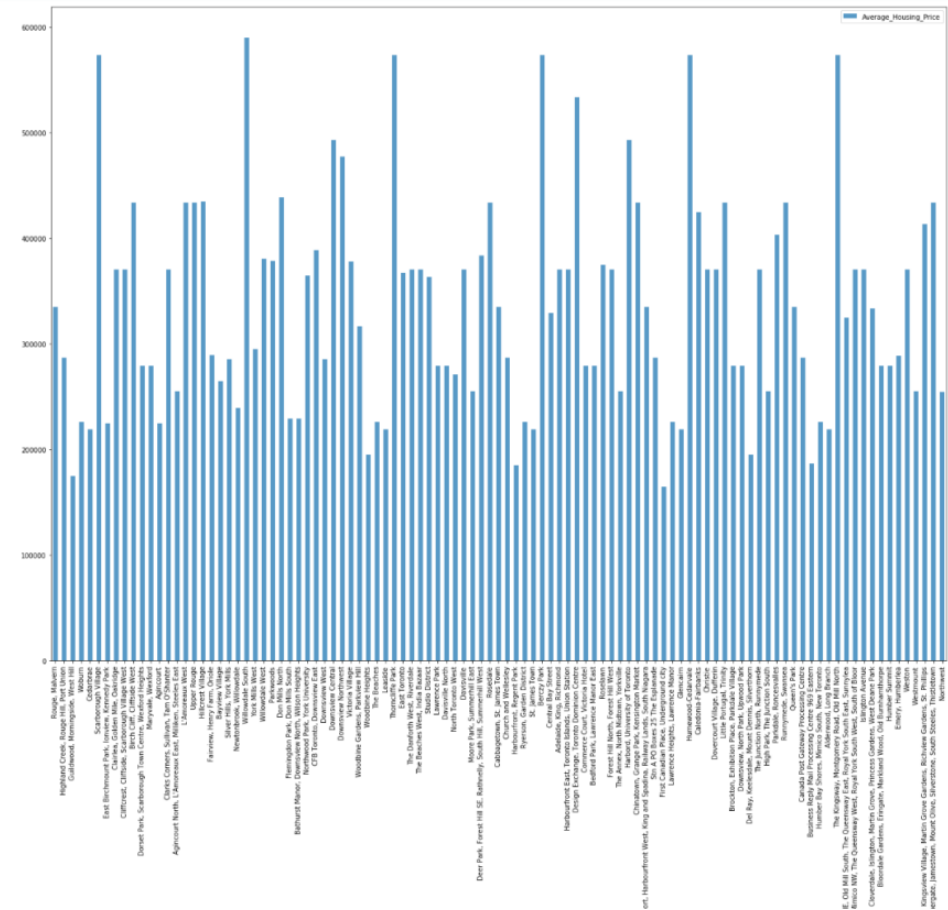
# Results Section

## ► Map of Clusters in Scarborough



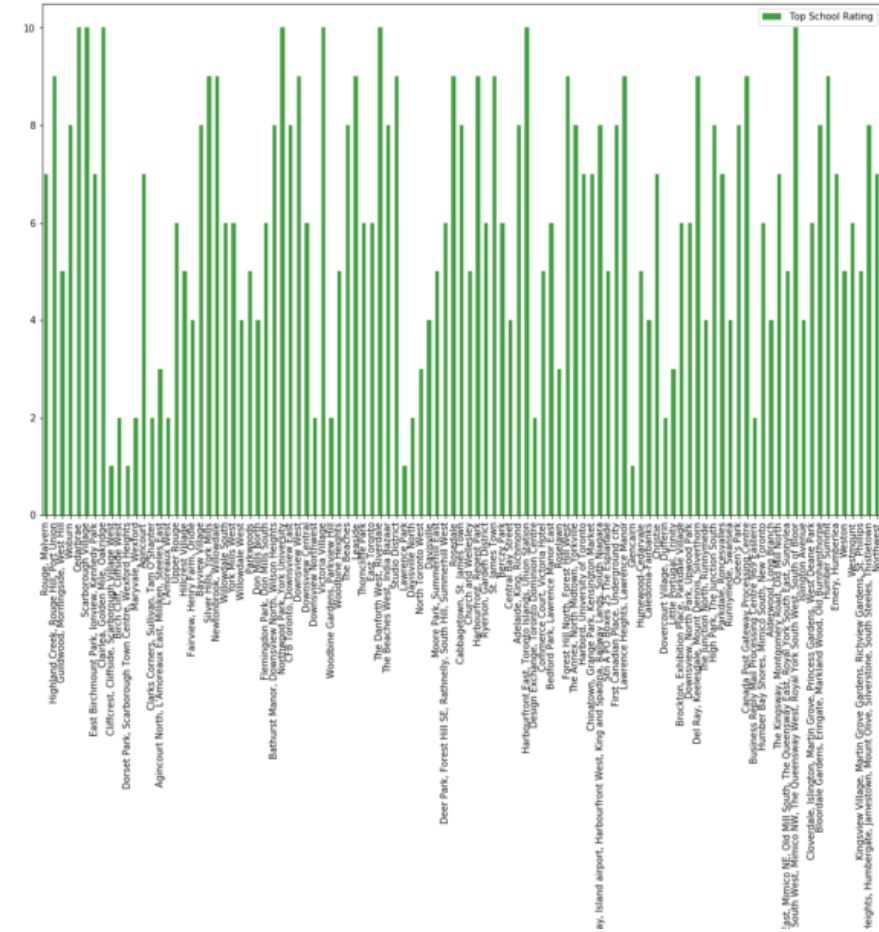
# Result Section

## ► Average Housing Price by Clusters in Scarborough



# Result Section

## ► 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.

# Problem Which Tried to Solve

- ▶ 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 Section

- ▶ In this project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 103 different latitude and logitude 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

- ▶ **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.