

# The Battle of Neighborhoods

Finding the best neighborhood in  
Scarborough, Toronto





# Introduction

The purpose of this Project is to help people explore better facilities around their neighborhood. It will help people in making smarter and more efficient decisions on selecting good neighborhoods out of the other neighborhoods in Scarborough, Toronto.

A lot of people are migrating to various states of Canada and need lots of research regarding good housing prices and reputed schools for their children. This project is for those people who are looking for better neighborhoods, for ease of accessing public places like Cafes, Schools, Super markets, medical shops, grocery shops, malls, theatres, hospitals etc.

This Project aims to create an analysis of features for people migrating to Scarborough to find the best neighborhood as a comparative analysis between neighborhoods. The features include median housing price and best 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 recreational facilities.

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

The major purpose of this project, is to suggest the best neighborhood in a new city for the people shifting there. Social presence in society in terms of like minded people, connectivity to the airport, bus stand, city center, markets and other daily things needed nearby.

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



# Data Description

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)

We will use Scarborough dataset which we scraped from wikipedia in Week 3. Dataset consists of latitude and longitude values, zip codes.



# Data Description

## Foursquare API data:

We will need data about different venues in different neighborhoods of a specific borough. In order to obtain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all types of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. Hence, 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 neighborhoods, we connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we choose the radius to be 100 meters.

The data retrieved from Foursquare contains information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue is 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



# Methodology

## Clustering

To compare the similarities of two cities, it's a good option 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

## Workflow

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



# Methodology

## Libraries used

Pandas: For creating and manipulating dataframes.

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

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

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

Geocoder: To retrieve Location Data.

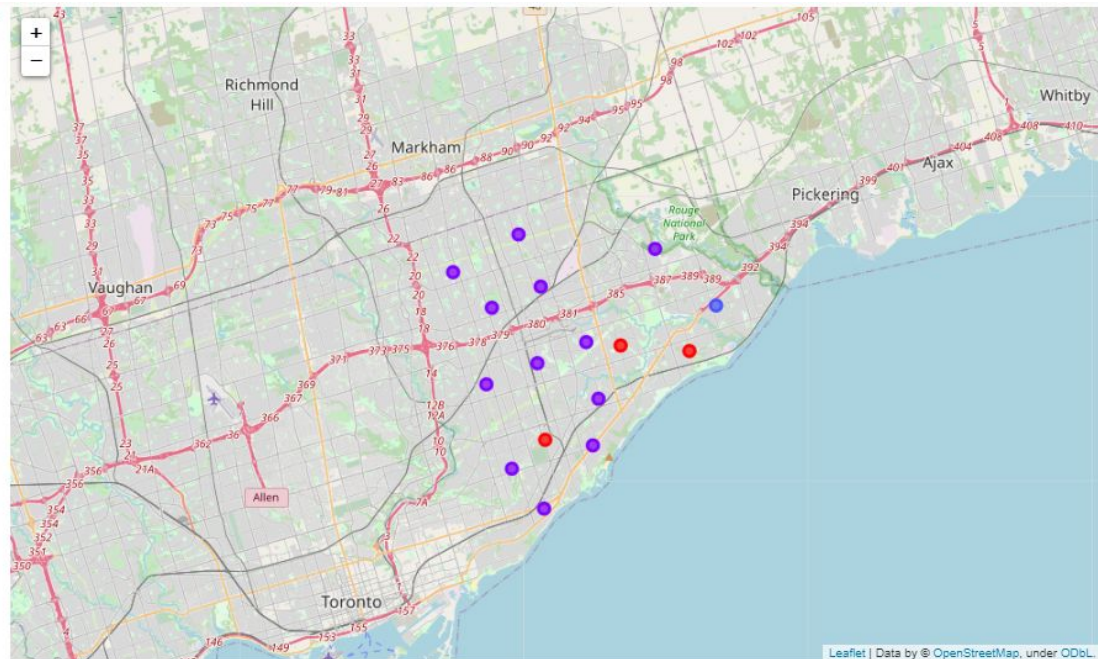
Beautiful Soup and Requests: To scrape data and to handle http requests.

Matplotlib: Python Plotting Module.



# Results

## Map of clusters in Scarborough

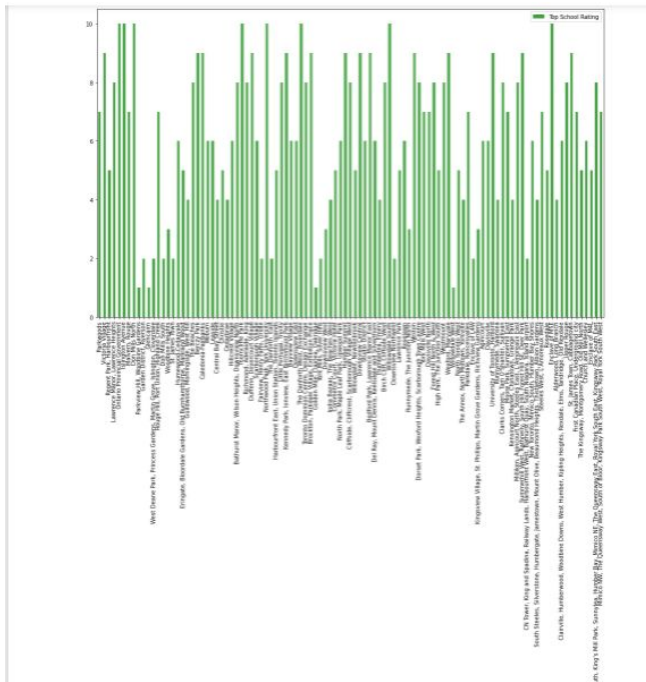


[illegible]





## School Ratings by Clusters in Scarborough





## Conclusion

In this project, using k-means cluster algorithm the neighborhood has been divided into 10 clusters for 103 different latitudes and longitudes from the dataset, which have very-similar neighborhoods around them. The plots above represent results of a particular neighborhood based on average house prices and school rating.