FINAL REPORT

The Battle of Neighbourhoods - Finding a better place in Scarborough Toronto

1. Introduction

We know that people migrate to different cities in search of jobs, education or due to some business requirements. While shifting many people are concerned about the neighbourhood around the particular location. They always aware of how easily can access hospital, groceries, and schools for children, medical shops, shopping mall, and theatre and also about the crime rates and housing prices. The need of this project is to explore the facilities around the neighbourhood for people who are moving out. This project focussed on the neighbourhood dataset of Scarborough and its features. Scarborough is one of the popular destination for new immigrants in Canada to reside.

2. Objective

The major objective of this project is to suggest a better neighborhood for people who are moving in to new city.

a) To sort list of house in terms of housing price

b) To sort list of schools in terms of ratings, reviews and fees

3. Data description

This project use Scarborough dataset scrapped from wikipedia. This data consists of

1. Postal codes
2. Borough
3. Neighbourhood

Data link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

After scrapping data from Wikipedia we include its latitude and longitude details. For specific borough we need data about different venues in different neighbourhoods. To gather that information we use Four Square API. Foursquare is a location data provider with information about all manner of venues and events within an area of interest.

After finding the list of neighbourhoods around Scarborough, connect it to the Foursquare API to gather information about venues inside each and every neighbourhood. For each neighbourhood, we have chosen the radius to be 100 meter. 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. Neighbourhood
2. Neighbourhood Latitude
3. Neighbourhood 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

4. Work flow

Four Square API

This project make use of Four Square API for data gathering. It has a collection of millions of places and its API which helps to perform location search, location sharing and business details. Due to some constraints http request for number of places per neighbourhood parameter be set to 100 and radius set to 500.

Clustering

To compare similarities of two cities we need to explore its neighbourhood, segment and finally group them to clusters. From this clusters we can find the feature of neighbourhoods. Unsupervised machine learning K-means clustering technique is used to cluster the data.

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.

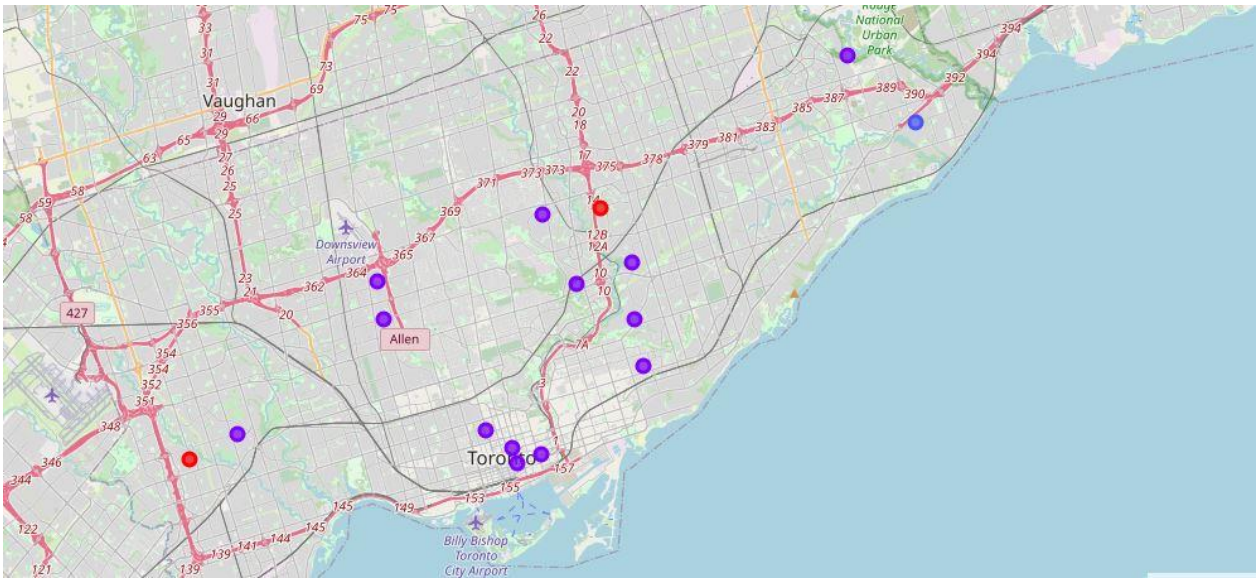
Geocoder: To retrieve Location Data.

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

Matplotlib: Python Plotting Module.

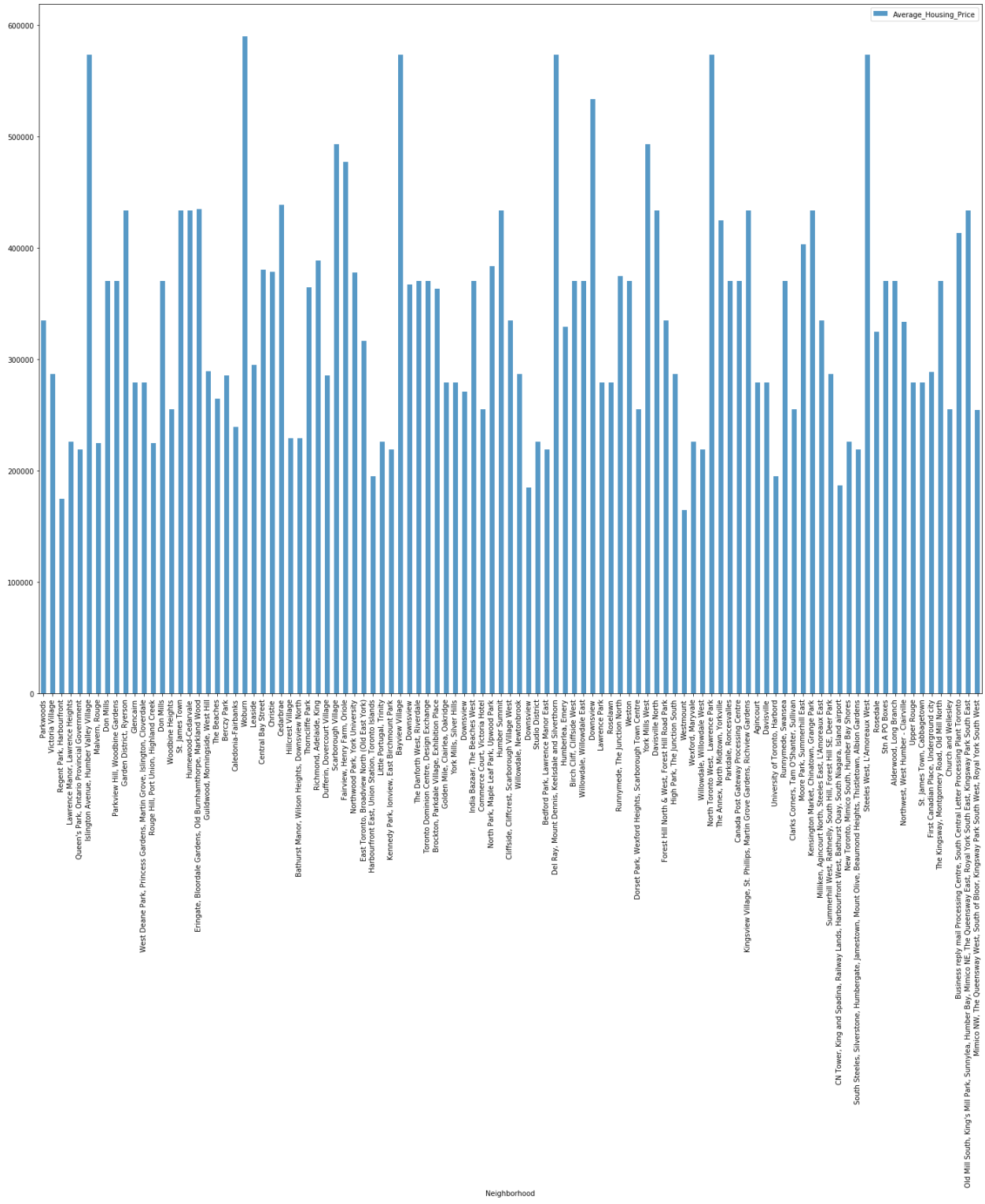
5. Results

Maps of clusters in Scarborough



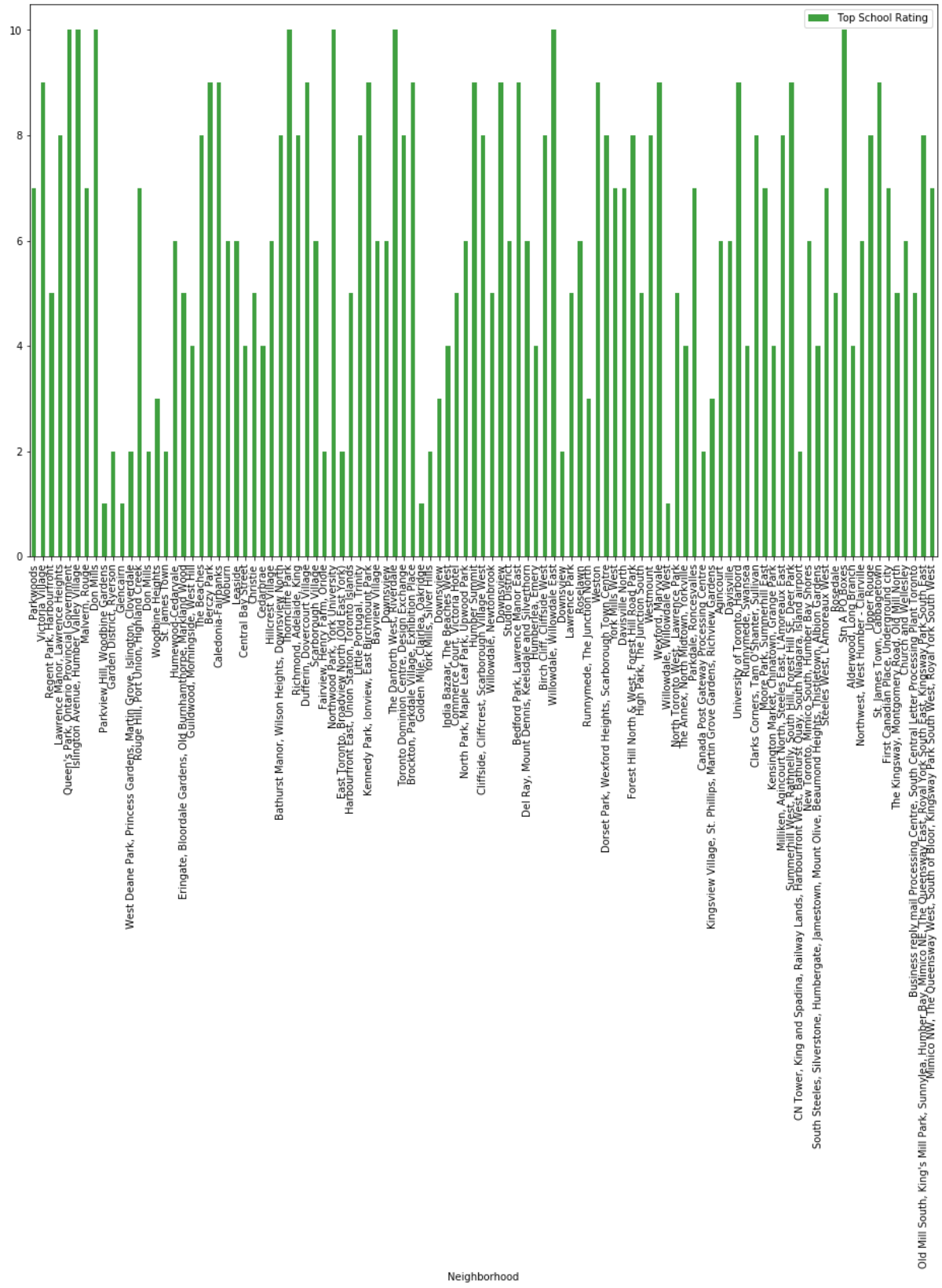
In kmeans approach given 10 clusters maximum and found out that there are majorly 3 clusters in the Scarborough. It’s shown in red, violet and blue

Average Housing Price by Clusters in Scarborough



In this graph its shows the list of house clusters and its prices

School Ratings by Clusters in Scarborough



In this graph its shows the list of school clusters and its rating

6. Discussion

Problem Which Tried to Solve:

The major purpose of this project, is to suggest a better neighbourhood in a new city for the person who are shifting there. Social presence in society in terms of likeminded people. Connectivity to the airport, bus stand, city centre, markets and other daily needs things 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 reviews.

7. Conclusion and future work

In this 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 presented to a particular neighbourhood 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.

This project can be continued for making it more precise in terms to find best house in Scarborough. Best means on the basis of all required things (daily needs or things we need to live a better life) around and also in terms of cost effective.