The Battle of Neighbourhoods

Swapnil Gaikwad

November 21, 2020

1. Introduction

1.1 Background

Lots of people are migrating to various states of Canada and needed lots of research for good housing prices and reputed schools for their children. This project will help to those people who are looking for better neighbourhoods.

1.2 Problem

Data that might contribute to determining a better neighbourhood might include - for ease of accessing to Cafe, School, Supermarket, medical shops, grocery shops, mall, theatre, hospital, likeminded people, etc. The goal of this project is to help people in exploring available facilities around their neighbourhood. It will help people in making smart and efficient decision on selecting great neighbourhood in Scarborough, Toronto.

This Project aim to create an analysis of features for a people migrating to Scarborough to search a best neighbourhood as a comparative analysis between neighbourhoods. 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 emergency, water resources both fresh and waste water and excrement conveyed in sewers and recreational facilities.

1.3 Interest

Obviously, it will help people to get awareness of the area and neighbourhood before moving to a new city, state, country, or place for their work or to start a new fresh life. Others who are finding different kind of facilities nearby may also be interested to know neighbourhood.

2. Data acquisition and cleaning

2.1 Data sources

For the Toronto neighbourhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighbourhoods in Toronto.

2.2 Data cleaning

Here we will be required to scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas data frame so that it is in a structured format like the New York dataset.

Once the data is in a structured format, you can replicate the analysis that we did to the New York City dataset to explore and cluster the neighbourhoods in the city of Toronto.

2.3 Feature selection

Dataset consisting of latitude and longitude, zip codes.

Table 1. Scraped dataset structure

Columns Names	Meaning	Absent Values
Postal Code	It is compulsory and should be present.	NA
Borough	It is present or absent.	Not assigned
Neighbourhood	It is present or absent.	Not assigned

2.4 Data extraction using Foursquare API Data

We will need data about different venues in different neighbourhoods 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. As such, 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 neighbourhoods, we then connect to the Foursquare API to gather information about venues inside each neighbourhood. For each neighbourhood, 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. 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

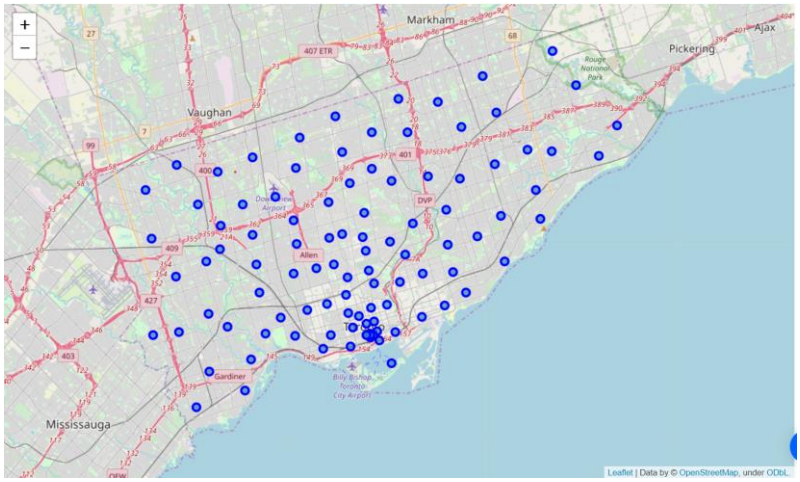

Figure 1. Map of Scarborough

3. Exploratory Data Analysis

3.1 Foursquare API features and limitations

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

3.2 Clustering Approach

To compare the similarities of two cities, we decided to explore neighbourhoods, segment them, and group them into clusters to find similar neighbourhoods 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.

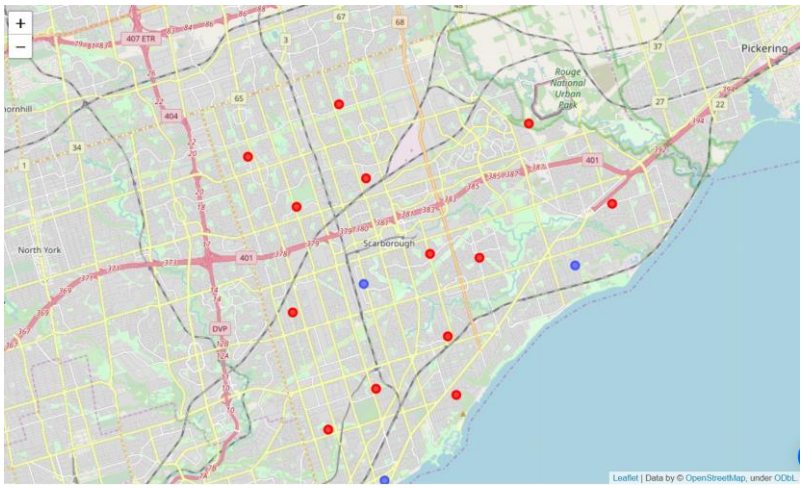

Figure 2. Map of Clusters in Scarborough

3.3 Average Housing Price by Clusters

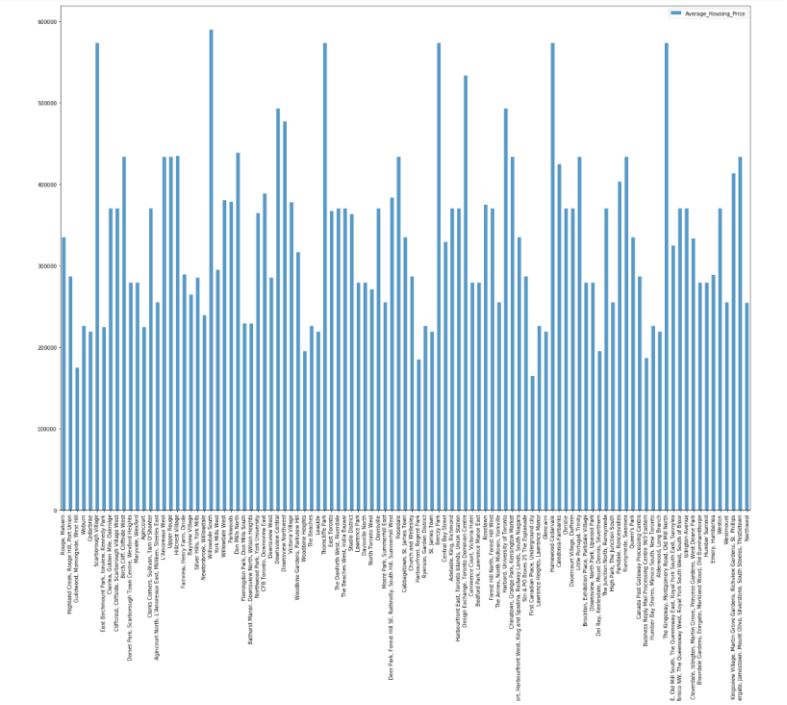

Figure 3. Average Housing Price by Clusters in Scarborough

3.4 School Ratings by Clusters

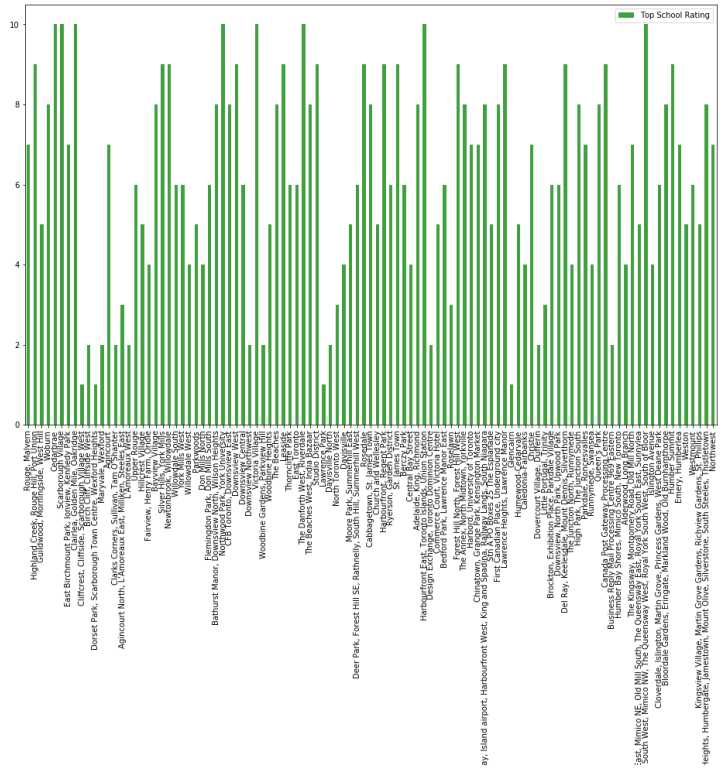

Figure 4. School Ratings by Clusters in Scarborough

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.

Foursquare API

This project have used Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

4. Predictive Modelling

4.1 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.

1. Sorted list of houses in terms of housing prices in an ascending or descending order

2. Sorted list of schools in terms of location, fees, rating, and reviews

5. Conclusion

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.

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.

6. Future directions

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