Neighbourhoods of Canada

Vanshaj Jain

July 1st, 2020

Introduction

This project has been created to help people in exploring better facilities in and around their neighbourhood. By using this project, we can make smart and effective decisions on selecting preferred neighbourhood out of numbers of other neighbourhoods in Scarborough, Toronto.

High migration rates to Canada prove the project's utility because the migrants need to research for affordable housing prices and great schools for their children. This project will also provide useful data to help people in finding and accessing hospitals, grocery shops, medical shops, cafe etc.

Aim of this project is to analyse features for migrants of Scarborough to search the most preferable neighbourhood by a comparative analysis. The features include affordable housing prices and better school rating, crime rates of that area, weather conditions, water resources and excrement conveyed in sewers and recreational facilities.

PROBLEMS which the project tried to solve:
The main aim of this project is to provide better preferences among different neighbourhoods in Scarborough, Toronto.
1) A sorted list of houses with their prices arranged in ascending or descending order.
2) A sorted list of schools with their location, rating, fees and reviews.

The Location:
Scarborough being a popular location with new immigrants, also happens to be one the most diverse are in Greater Toronto area. It's also a place for religious groups to perform worship in certain places. Albeit governments of some country are restricting immigrants and refugees, the trend of immigration in Canada has been on rise.

Foursquare API:
Foursquare API has a database of millions of locations, it's API provides us the ability to perform location search, sharing and details of a certain location or business. This project will mainly use Foursquare to gather data for locations.

Methodology

Work Flow:
Features of all nearby locations will be gathered using credentials of Foursquare API. The number of places per neighborhood parameter will be limited to 100 and the radius parameter to 500.

Clustering Approach:
In order to compare two cities, we have decided to explore neighbourhoods, segment them and group them in clusters so that we can find similarities in the neighbourhoods in an enormous city like Los Angeles and Toronto. To enable that, we need to cluster data using the k-means clustering algorithm.

Libraries used to develop this project:
Pandas - Used to make python data frames
Scikit Learn - Used to import the K-means clustering algorithm
Folium - Used to visualize the cluster distribution of neighbourhood of using leaflet map
JSON - Used to handle JSON files
Geocoder - Used to retrieve data
Beautiful Soup and requests - To scrap and to handle http requests
Matplotlib - Python plotting module

Data Description -

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

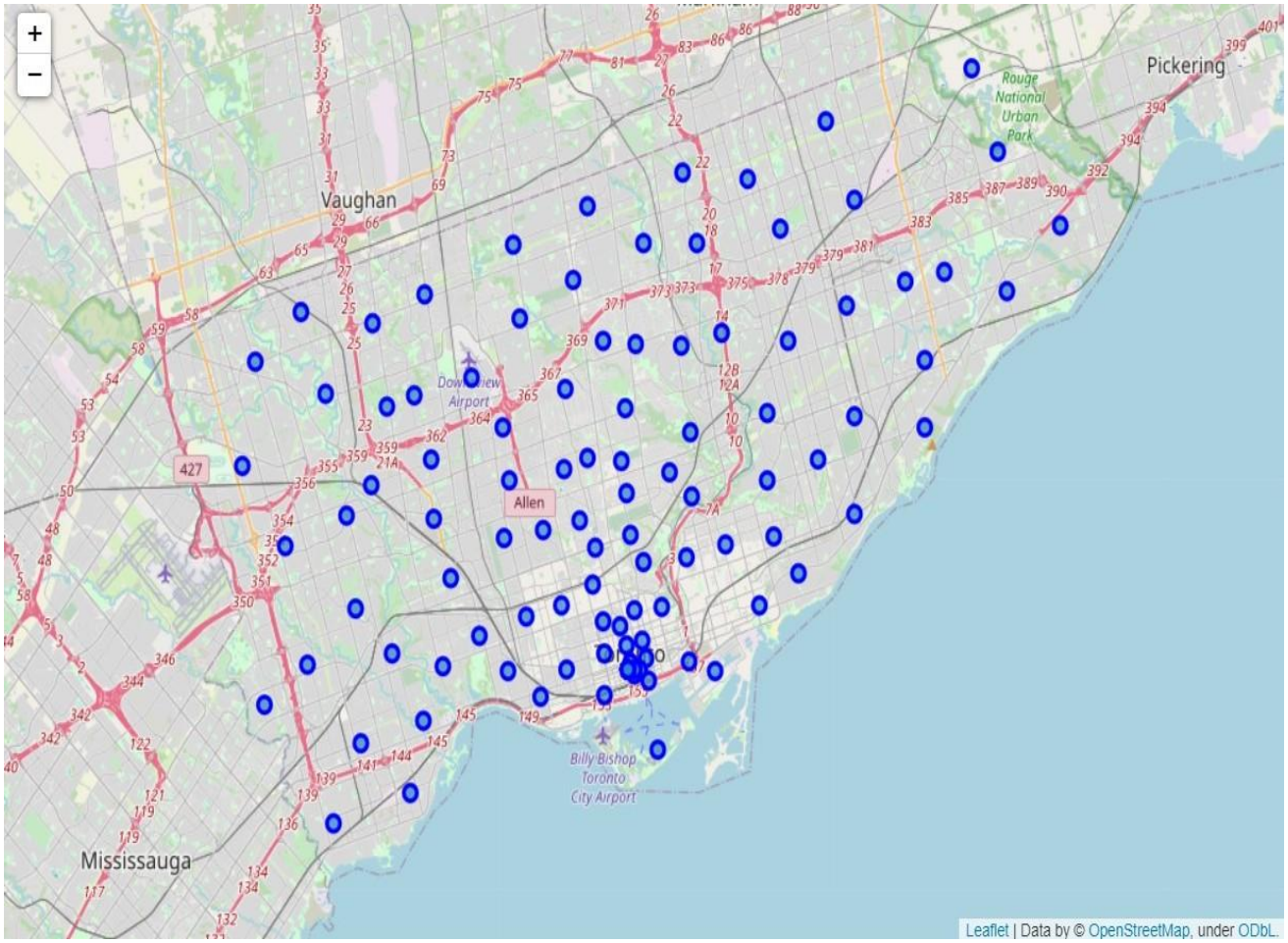
Foursquare API data:
Data about different venues in various neighbourhoods will be needed. To gather this information, we will be using Foursquare location information.
Foursquare API has a database of millions of locations, it's API provides us the ability to perform location search, sharing and details of a certain location or business. It will provide us detailed information of a venue, for example - venue names, locations, photos and menus.

After getting the list of neighbourhoods, we will connect to the Foursquare API to gather detailed information about venues in each neighbourhood. We will choose the radius to be 100 meters for each neighbourhood.
The information obtained per venue is -
1)Neighbourhood
2)Neighbourhood Latitude
3)Neighbourhood Longitude
4)Venue
5) Name of venue
6) Venue latitude
7) Venue longitude
8) Venue category

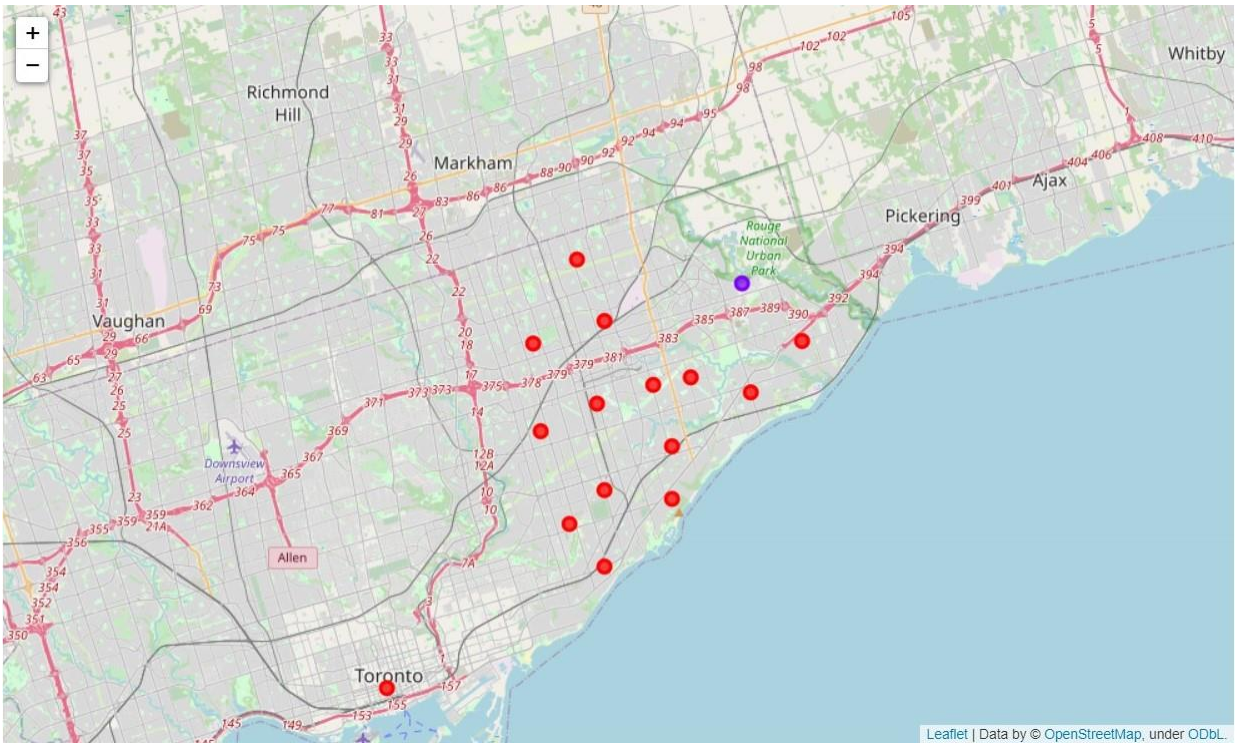
Map of Scarborough

1)This map of Scarborough was generated using Foursquare API.

2)Blue dots indicate all the neighbourhoods of Scarborough.



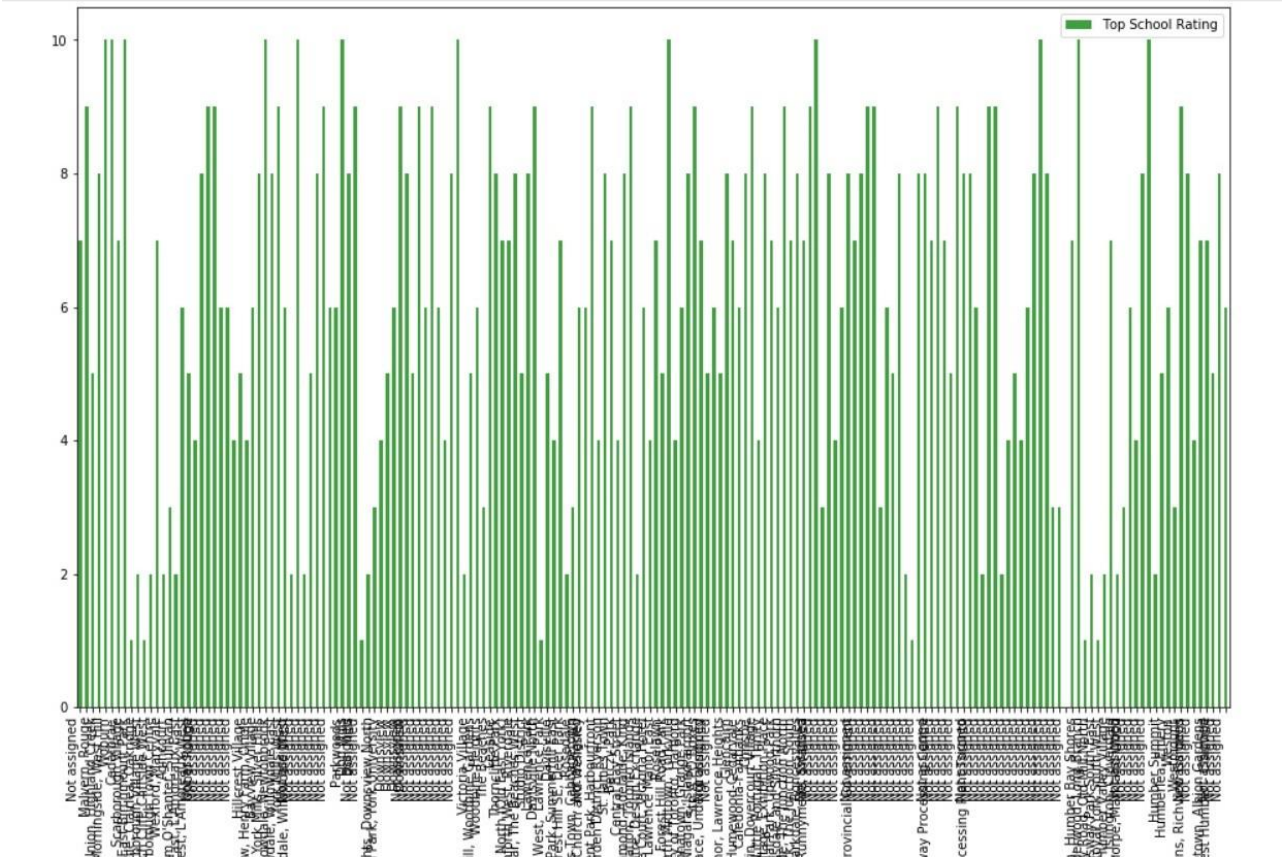
Map of Scarborough showing clusters of neighbourhoods



Result

This graph is a box plot showing school names against their rating.

Rating of school shows the overall quality of a school helping users to make a better judgement.



Conclusion

This project has been useful – 1) To form clusters of neighbourhoods in the city of Scarborough, Canada.

2) To find out schools and their ratings.

Ideas include: 1) Using K means Algorithm to cluster Data.

2) Using Foursquare API to gather location data

3)To use a box plot to visualize ratings against school names.

Discussion

This project can be enhanced and more tools can be added to get information about the broader area.

More information like venues near the neighbourhoods like café, restaurants, movie theatres, hospitals and parks etc.

We can also provide details about these venues, like location, reviews and menu of the place.