Neighbourhood Analysis

Niharika Reddy Kota

1. Introduction

The purpose of the project is to explore different facilities around the neighborhood. This project helps people explore their neighborhood. It helps people in finding similar neighborhoods to theirs in Scarborough, Toronto.
Many People transfer from one state or contry to another. It is important to find a neighbourhood which is similar to the one you stayed till date in Scarborough, Toronto.
This Project aim to create an analysis of features for a people migrating to Scarborough to search a best neighborhood as a comparative analysis between neighborhoods. 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 freash and waste water and excrement conveyed in sewers and recreational facilities.

It will help people to get 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.

1.2 Problem

The Main problem is tring to find a better neighborhood which includes many facilities like airport, bus stand, city center, markets and other daily needs things nearby.
having these facilities nearby with appropriate house rates and a safe enivronment in the neighborhood.

2. Data acquisition and cleaning

2.1 Data sources

The data related to Canada is acquired from the link https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and is retrieved using BeautifulSoup. The dataset related to New York is from IBM Json file. The dataset contains information about the neighbourhood, postal code, latitude, longitude.

We need data about the locations of the neighbourhood. Using FourSquare, we provide the latitude, longitude and neighbourhood to access the facilities nearby.
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.

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

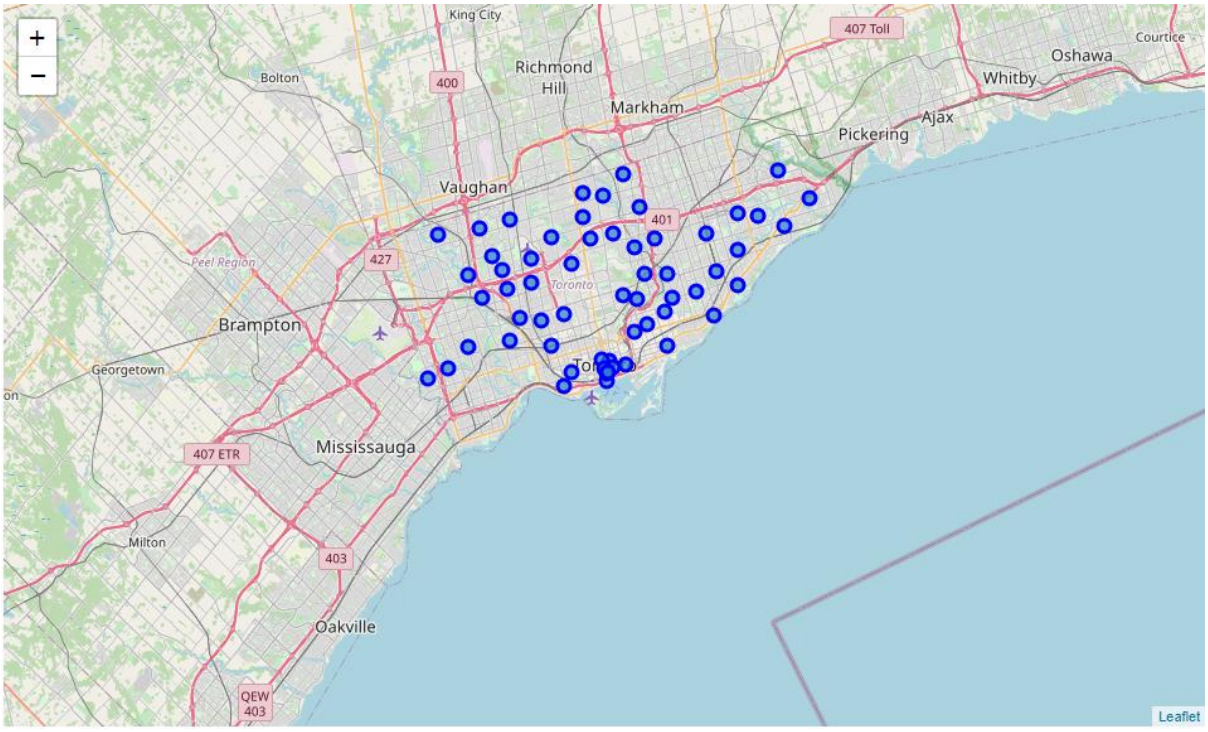
2.2 Data Cleaning

The Canada data from the link is retrieved through the use of BeautifulSoup libraries. We extract the required fields postal code, borough, neighbourhood from the link. The rows which contain nan values are dropped from the table. A csv file is loaded which contains the latitude and longitudes to the neighbourhoods.

Acquiring the latitude and longitude of Toronto, we use foursquare to extract the venue and its details in JSON format. The data is cleaned to extract name, categories, lat and log from the FourSquare site. Same format is done to New York data and all the required details are extracted. All the nan values after combining the both Toronto and Manhattan data together are replaced with 0.0 for the process of clustering.

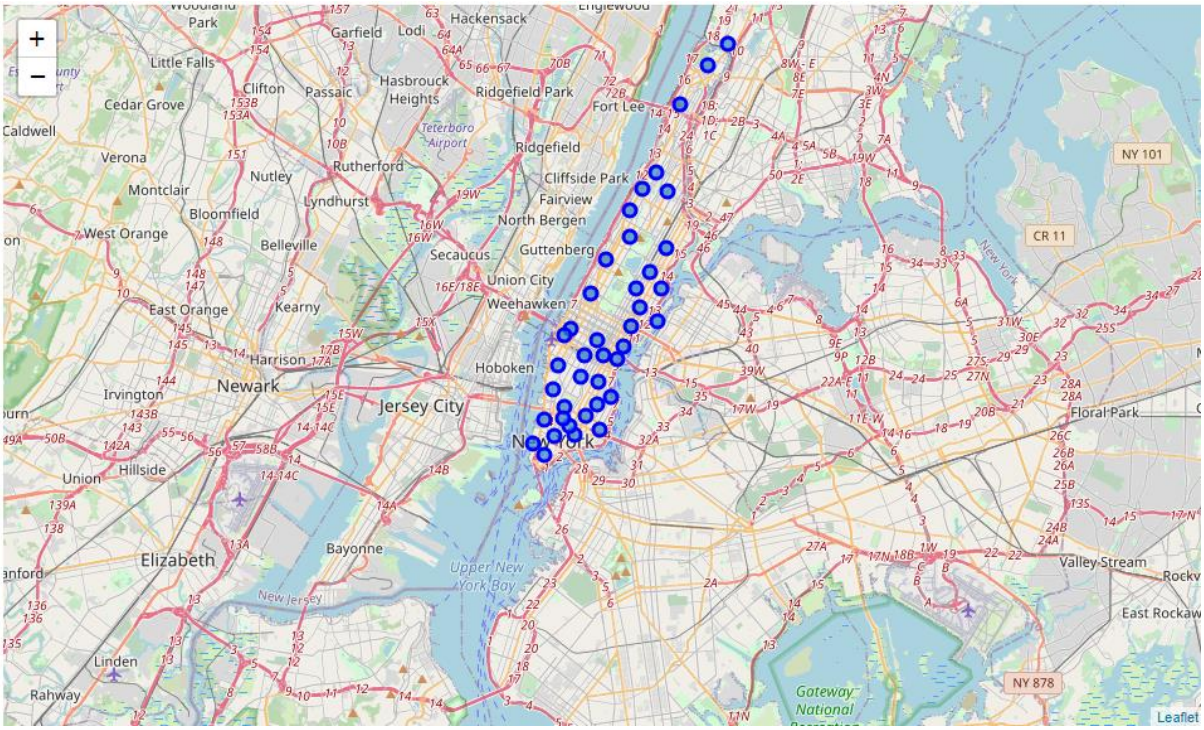
3. Exploratory Data Analysis

3.1 Toronto neighbourhood



All the neighbourhoods are marked in blue. This neighbourhoods belong to city Toronto, Canada.

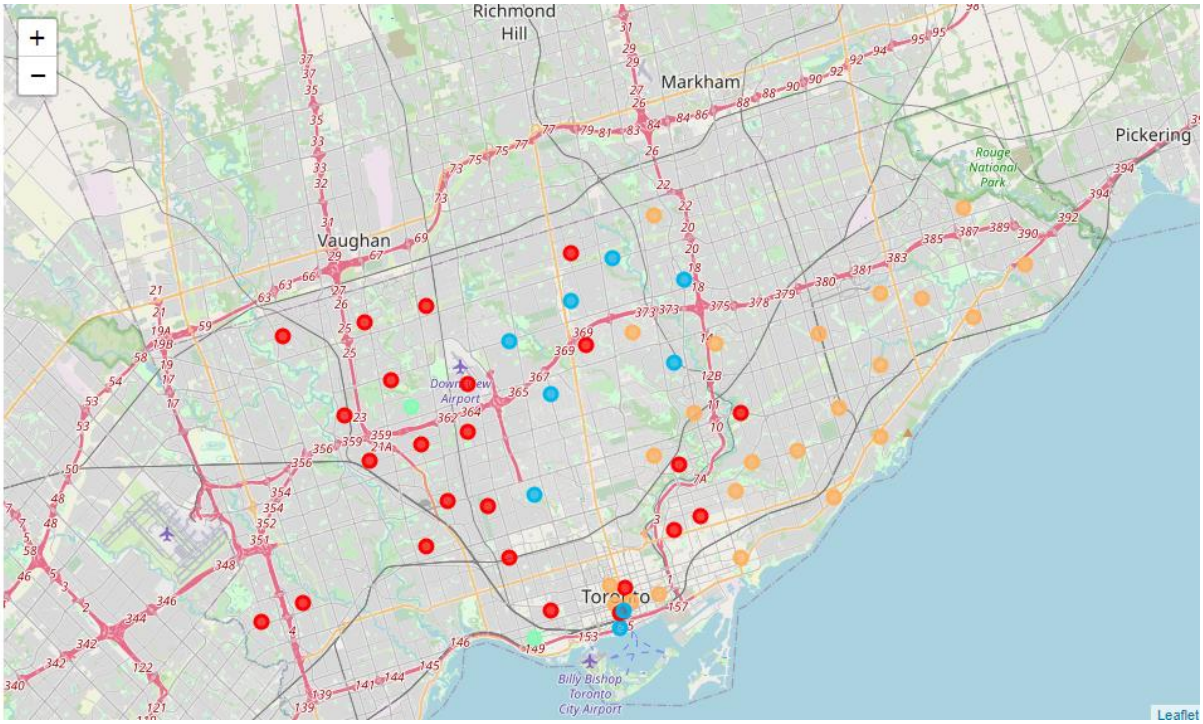
3.2 Manhattan neighbourhood



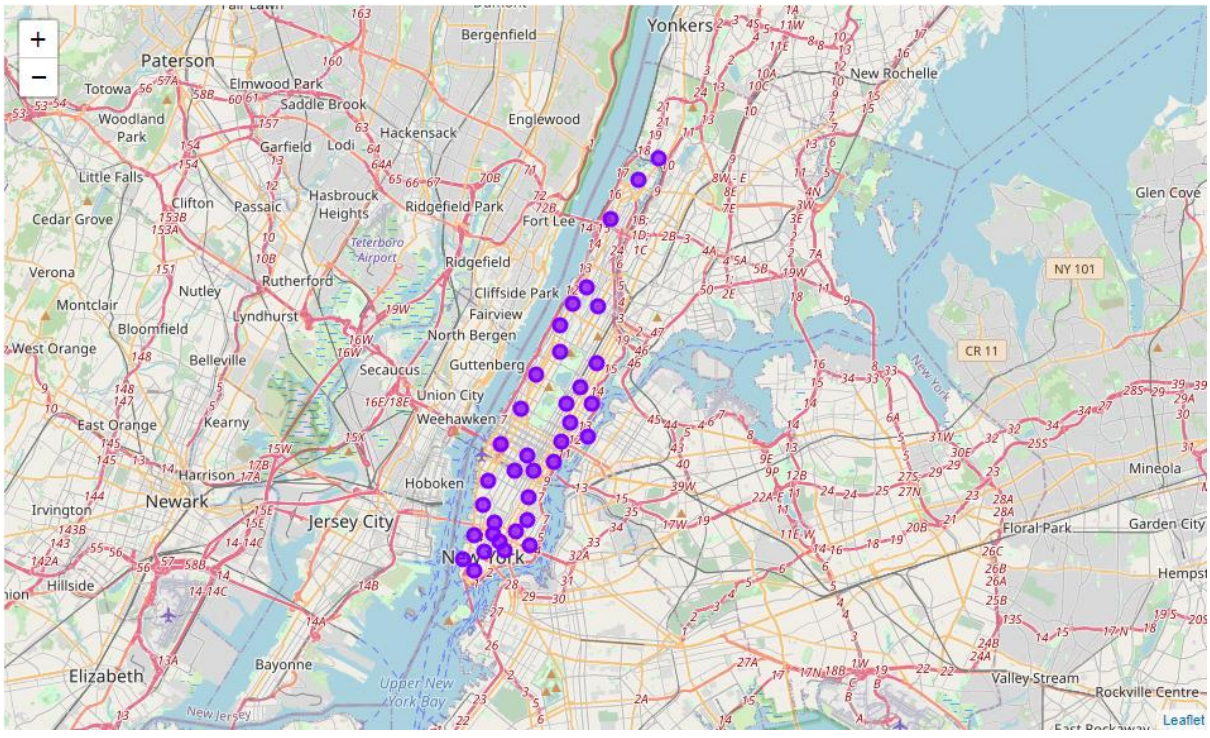
The blue dots in the figure represents the neighbours of the city Manhattan, New York.

4. Cluster Analysis

In order to know which neighbourhood is similar to the user neighbourhood, we perform cluster analysis on the merged data of both cities. Here, we have performed K-means algorithm on the merged dataset. The dataset contains all the nearby venues which are used as parameters for finding a suitable location. After checking in which number of clusters provide the most fit, the no of cluster value is 5 and the efficiency is high. From the cluster analysis of KMeans, it is determined that the among the city Toronto we can observe that there are similar neighbourhood in each cluster. In case of New York, the city has all similar neighbourhoods but it is not similar to other clusters.



In the above figure, it can be observed that each colour represents a cluster, similar neighbourhoods also it represents city Toronto in Canada.



In the above figure, it is observed that all neighbourhoods in New York, Manhattan are all similar and yet different to the Toronto neighbourhoods.

5. Conclusion
   It is observed that the neighbourhoods in Toronto don't have similar neighbourhoods in New York. Moreover, the city Toronto has many similar cluster of neighbourhoods which can be observed in the figure above. We can conclude that the if a person wants to shift to New York, it is not likely to find a place which is similar to a neighbourhood in Toronto.

6. Future directions
   It would be better to include not only the venues as parameters but also other parameters like crime rate, property lease, etc. These extra parameters help the clients in make a more suitable decision and work on their planning well.