The Battle of Neighborhoods

Problem Definition:

The objective is to help people explore better facilities around their neighborhood to help them make good and efficient choices. We will use the neighborhoods in Scarborough, Toronto from Week 3 assignment for this study.

People moving across different neighborhoods require careful consideration for good housing prices, reputed schools for their children, community engagement, crime rate, supermarkets, cafes, hospitals, etc. The intent of this study is to help such people.

Problem Which Tried to Solve:¶

This study aims to create a comparative analysis of features for people migrating to Scarborough to search for the best neighborhood based on the factors mentioned above.

The Location:

Scarborough is an administrative division of Toronto, Ontario, Canada. Situated atop the Scarborough Bluffs, it occupies the eastern part of the city. Scarborough is contained within the borders of Victoria Park Avenue on the west, Steeles Avenue to the north, Rouge River and the city of Pickering to the east, and Lake Ontario to the south. It borders Old Toronto, East York and North York in the west and the city of Markham in the north. Scarborough was named after the English town of Scarborough, North Yorkshire.

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. It includes a number of natural landmarks, including the Toronto Zoo, Rouge Park and the Scarborough Bluffs. The northeast corner of Scarborough is largely rural with some of Toronto's last remaining farms, earning Scarborough its reputation of being greener than any other part of Toronto.[3]

Foursquare API:

This project would use Four-square API as its data gathering source which provides the ability to perform location search, location sharing and details about any business.

Clustering Approach:

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm

Libraries Which are Used to Develop the Project:

- Pandas: For creating and manipulating data frames.
- Folium: Python visualization library would be used to visualize the neighborhoods 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.

The Battle of Neighborhoods | Data Description
Data Description:

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

Will use Scarborough dataset which we scrapped from wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes.

Foursquare API Data:

We will need data about different venues in different neighborhoods of that specific borough. In order 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 neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, 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. 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