Finding a neighborhood in New York City to feel exactly like home

Introduction

Both Toronto and New York are great cities to live. As the economic center of United States, obviously New York City offers more career opportunities. Many people are moving from Canada to United Sates for better career opportunities. Assuming a person is moving from his home in the Caledonia- Fairbanks neighborhood in Toronto to New York City. He loves the neighborhood he lives in in Toronto because of the venues (parks, gyms, shopping places, restaurants, coffee shops, etc.) the neighborhood provides, and would like to find a neighborhood in New York city where he feels like being home.

This project provide a solution to find the New York City neighborhood(s) that are very alike to the neighborhood in Toronto, using data mining techniques. Characteristics of a neighborhood was quantitatively described using the types of numbers of venues (parks, stores, restaurants, etc.) in the neighborhood. These information/data of the neighborhood in Toronto (i.e., “Home Neighborhood”) was put into the information/data of all New York City neighborhoods, and k-means clustering technique was used to build up clusters of neighborhoods based on the degree of similarities regarding the neighborhood venues. The New York City neighborhoods that fell into the same cluster as the “Home Neighborhood” were identified as the target neighborhoods. The value of K was selected such that the number of identified target neighborhoods is between 1 to 5.

Data

List of the name, zip codes of Toronto neighborhoods and New York City neighborhoods were collected from the online database as below.

<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

https://cocl.us/new\_york\_dataset

Geographical coordinates were collected from an online database

https://cocl.us/Geospatial\_data

Methods

The neighborhood lists and the geographical coordinates were linked together by matching the zip code of the neighborhood. Foursquare API was called to collect information of venues in each of the neighborhood. Data cleaning and transformation/standardization was applied to the venue data before the k-means clustering model was ran. A Toronto neighborhood called “CALEDONIS- FAIRBANKS” was used as the “Home Neighborhood”. Different value of K (number of clusters) was applied. The final pool of neighborhoods were selected as the solution pool when the number of New York City falling into the same cluster as the “Home Neighborhood” is between 0 and 5.

Conclusion

The neighborhoods of New York City that can offer a life style that is most alike to the Toronton neighborhood “CALEDONIS- FAIRBANKS” are the following 4-

Bayswater, Clason Point, Somerville and Todt Hill.

