Capstone Project - Comparing NYC and Toronto

Introduction

Background

Toronto is similar to New York city in many ways. They are both the largest city in their own countries. Besides, both cities are centers for immigrations. According to Quora (John Smith, 2017):

Both the largest cities in their respective countries.

Both the financial and tourist capitals of their countries

Both have a high cost of living

Both have a network of subways, trains and buses serving their greater areas

Both are known for their skyscrapers (albeit, Toronto's are limited to its Financial District; New York's cover all of Manhattan)

Both are extremely multi-ethnic, although Toronto is more so (49% of TOs citizens were born somewhere else, compared to 36% of NYC's)

Both are theatre capitals of the English-speaking world

Description of Problem

In week 3, we explored New York City and the city of Toronto and segmented and clustered their neighborhoods. We noticed that both cities have financial centers and a lot of diversified neighborhoods. Giving that the two cities are very similar, it is interesting for us to examine how similar the two cities are regarding neighborhoods and venues. Through data analysis, we can have better understanding on the whether New York city or Toronto is more multicultural.

Data

Datasets

In order to make comparison between NYC and Toronto, we make use of datasets from both cities.

First, we downloaded the data set from "lab:Segmenting and Clustering Neighborhoods in New York City". The data is presented on https://geo.nyu.edu/catalog/nyu_2451_34572. After cleaning the data, we export it into the nyc.csv file for further research.

Besides, we also make use of the Toronto data from "peer assignment: Explore and cluster the neighborhoods in Toronto". The data is produced by merging two datasets from http://cocl.us/Geospatial_data and https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.

Both data contains neighborhoods geographical information. The columns in the data sets are:

'Borough','Neighborhood','Latitude','Longitude'

We can use the datasets above to get Venues information through Foursquare API.

Foursquare API
Foursquare is a location technology platform dedicated to improving how people move through the real world. In this project, we will used the Foursquare developer API to gather information of venues through the geographic information we gather from our two datasets (New York city and Toronto).

This API provides every specific Venue's information respect to given Latitude and Longitude. Foursquare API return the following information of each venue:

1. Neighborhoods
2. Neighborhoods Latitude
3. Neighborhoods Longitude
4. Venue
5. Name of Venue
6. Venue Latitude
7. Venue Longitude
8. Venue Catagory

The information can help us make comparision between venues in NYC and venues in Toronto. Also, the catagories of venues can reflect the degree of multiculture, which we are interetsed in explore.