The Battle of Neighborhoods

(Toronto Vs New York)

Introduction:

Canada's immigration destination has been increasing from past decade with a total of 313,580 immigrants in 2019 (Erin Duffin, 2019). Number of people immigrating from United States has also increased drastically and the main reason being politics. So let's see if moving to Canada from USA is a good option or not for a political neutral person and for a person who likes to explore places and meet new people.

The project's main focus will be on comparing the financial capitals of two countries being Toronto, CA and New York City, US. Let's draw some insights about the neighborhoods of two cities and see which city has rich culture, interesting places to explore and has wide variety of cuisines.

Problem Statement:

Analyze the data regarding venues in Toronto and New York City. Cluster their neighborhoods and draw insights on which city is better to live for a person who likes to explore and try different cuisines.

Data Acquisition

Data Source:

The data set for New York City Neighborhood will be downloaded from NYC OpenData. A csv file is downloaded from "https://data.cityofnewyork.us/City-Government/Neighborhood-Names-GIS/99bc-9p23". The csv file is converted to a data frame which will be transformed/cleaned. The cleaned up data frame is expected to contain Borough, Neighborhood, Latitude, Longitude.

The data set for Toronto will be downloaded from Wikipedia. The tables containing the neighborhood information of Toronto is scraped from "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M". The scraping of the data from web is done using Beautiful soup and html parser and the data frame for Toronto's neighborhood is created and cleaned according to the requirements. A csv file containing data about latitudes and longitudes will be downloaded from "https://cocl.us/Geospatial_data". The cleaned up data framed is expected to contain Postal Code, Borough, Neighborhood, Latitude, Longitude.

Using Foursquare, we will find latitudes, longitudes and venues for each neighborhood in both the cities. This information will form a dataset which will be used in drawing insights about the problem.

Using all the above data, neighborhoods are clustered and grouped to understand cities better.

References:

1) Immigration Statistics: (Duffin, 2019)