Project: *Exploring and Comparing Neighbourhood Clusters in Toronto and New York City to Select the Most Similar Neighbourhood*

Introduction/Business Problem

One issue that people must face in a globalized world is moving far away from home. Whether they are pursuing a dream career, moving closer to family and friends, or just wanting a change of environment, it is inevitable that some people will make big changes in their life and move elsewhere. People would rather move to areas where they feel the most comfortable so they can adjust to their new surroundings a little easier; some may want to move somewhere that has similar food culture, shops, recreational spots, etc. However, when you're moving somewhere new, it can be overwhelming to search through the whole city and find these areas.

Both New York City (NYC) and Toronto are major metropolitan cities renown for being financial capitals of their respective countries with incredible multiculturalism. However, they are in different countries and have different histories; their environments, though similar in some ways, are vastly different in other ways. When one is moving from Toronto to NYC, it would be ideal to move to a neighbourhood that had the greatest similarity to their home city to ease the adjustment process.

This Capstone project aims to analyze neighbourhoods in Toronto and NYC and determine their similarity (this will mostly be based on similarity in venues using Foursquare API), using these insights as a reference for recommending which neighbourhood one should choose in NYC when moving there from Toronto. Following data science methodology, machine learning techniques such as k-means clustering will be used to provide answers to the business question: What similar neighbourhood in NYC would you recommend an individual move to if they were moving from Toronto?

Data

The following data will be needed to answer the business question:

- List of neighbourhoods in Toronto and their latitude and longitude coordinates
- List of neighbourhoods in NYC and their latitude and longitude coordinates
- Information of the venues in each of the neighbourhoods

The Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) "List of postal codes of Canada: M" will be used to get all the names of each neighbourhood, their boroughs, and postal codes. Web scraping techniques will be used to extract the data, using BeautifulSoup and other Python libraries.

Luckily, the NYC data is available on IBM cloud (https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json), which will be used to get all the names of each neighbourhood and their boroughs.

The geographical coordinates (latitude, longitude) of each neighbourhood will be retrieved from using Python's geocoder package. Then, Foursquare's API will be used to get the venue data for each neighbourhood.

The machine learning technique k-Means clustering will be used to cluster the neighbourhoods, and the Python folium library will be used to visualize these clusters on a map. Then, the clusters of neighbourhoods that have the most similarity in terms of types of venues and size will be chosen for the individual to move to.