**Applied Data Science Capstone Final Project**

**Differences between New York and Toronto: Which city is more like the other?**

**Presented by** Marcus Miles

**Project Description:**

This project will examine the differences between the two major cities of New York and Toronto in order to distinguish which city is more like the other based on the characteristics of their neighborhoods. New York is one of the most densely populated cities in the East Coast of the United States and is considered one of the financial capitals of the world. Toronto is one of the most popular destinations to visit in Canada and is also the most populated city within Canada with a population of over 2.8 million people.

I would make the assumptions that both New York and Toronto are very similar in their characteristics given how they are the most populated cities in their respective areas and how many businesses have locations within the city. This analysis will focus on key areas such as parks, museums, etc. to show how similar/different the two cities are from each other.

**Description of Data:**

The data will be from three different sources:

1. Toronto, Canada data: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

2. New York Data: <https://geo.nyu.edu/catalog/nyu_2451_34572>

3. Foursquare venues data: <https://foursquare.com/>

The data that will be gathered will be used to showcase the differences/similarities between the two cities. The venues that will be gathered from the Foursquare website will show which amenities and businesses that may be appealing to companies that want to set up shop in the area. This information will also be beneficial to people who are looking to move to a major city and are trying to find out what each city has to offer. Job recruiters can use this analysis as a selling point to any prospective client who they are trying to recruit for companies in the New York/Toronto area.

**Methodology:**

The New York data that was used for analysis is the geographical data for cities within the New York, NY area. The data was gathered from a JSON file and was formatted into a pandas dataframe to make it easier manipulate the data. The Toronto, Canada data was gathered from a Wikipedia page that had the postal codes for the neighborhoods within Toronto. This data was also formatted into a pandas dataframe for easier data manipulation. The following bullet points describe the process of the analysis:

* Gathered data from New York and Toronto and formatted into pandas dataframes.
* Pull the geographical coordinates for each city by using the address of the city and imported the latitude and longitude.
* Created a map with data points to map out each neighborhood within the cities.
* Pick a single city/borough within New York/Toronto to narrow our analysis down for comparison.
* Grabbed the latitude and longitude for each city/borough.
* Created a map with data points that represent neighborhoods within the city/borough.
* Using the latitude and longitude of a neighborhood, we used the Foursquare API to gather 100 venues within the area.
* After formatting and encoding the data, we gather the top 5 venues within each neighborhood within the city.
* Then the data was put into a dataframe that showcased the top 10 most common venues with the neighborhoods.

**Results:**

The results showed that the neighborhoods within New York and Toronto had similar venues between each city. Given that these two cities are major cities within the east coast, I expected that would be the case. There were slight differences between the most common venues within the neighborhoods but for the most part, the venues were similar on average.

**Discussion:**

The analysis showed that New York and Toronto are very similar cities based on the venue data gathered from the Foursquare API. Any realtor or business could use this analysis as a selling point for any buyers looking for a place in either city. There are a lot of attractions that could pull in a younger crowd or growing businesses looking to be in high traffic areas.

With a lot of different venues within the city, a realtor could use the data to gather clients who are looking for a more upbeat city lifestyle. The options available to them would be a great benefit to their life if they decided to live within the city limits.

As for businesses, it is always a plus to be where the action is. Many potential clients as well as workers who can help grow the business would be a huge benefit to the company by relocating to these areas.