Introduction/Business Problem

I am a small business owner, currently running my business from an office in Los Angeles. I have decided to relocate to either Toronto or New York City to open another office, but need help in deciding which location is the better fit. In order to make the decision, I will need insight into the neighborhoods and local businesses in the cities - things to do, places to eat. This project will explore the similarities and differences between neighborhoods of the two cities and determine which best fit the lifestyle of my employees.

Data

The data for this project will be scraped from the website pages below. This data consists of postal codes, neighborhood names, latitudes and longitudes for each neighborhood. Foursquare API search feature will be used to collect neighborhood venue data, while details about local venues and businesses will provide insight into the qualities of a neighborhood. In addition to Foursquare, various Python packages will be used to create maps and machine learning modeling techniques will be used for deeper insight into each neighborhood.

Website pages:

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

Toronto Latitude and Longitude: [http://cocl.us/Geospatial\_data](https://cocl.us/Geospatial_data)

New York City Neighborhoods: <https://geo.nyu.edu/catalog/nyu_2451_34572>

New York City Latitude and Longitude: Python Geopy

Methodology

* Using postal codes, Foursquare API server will pull location (latitude and longitude) information.
* Foursquare API search feature will be enabled to collect the nearby venues of each neighborhood. Due to call request limitations, the number of places per neighborhood parameter would be set to 100, while the radius parameter would be set to 500.
* Folium - Python visualization library would be used as the visualization tool to visualize neighborhood cluster distribution of Toronto over an interactive map.
* Unsupervised machine learning algorithm K-mean clustering would be applied to form clusters of different categories of places residing in and around the neighborhoods.These clusters would be analyzed individually, collectively and comparatively to compile conclusions.