Introduction/Business Problem

When moving from one city to another, it's often difficult to decide which neighbourhood to move to because there's no way to know what these neighbourhoods are really like. To solve this problem, I would like to analyze the neighbourhoods of 4 cities: Toronto, Montreal, Vancouver, and New York City to find which neighbourhoods in different cities are most similar to each other.

With my solution, people will be able to base their moving decisions on what they already know about neighbourhoods in their area. If, for example, I wanted to move to Vancouver and I knew already that I really liked the Church and Wellesley neighbourhood of Toronto, I can find which neighbourhoods in Vancouver are most similar to help me decide where to live in Vancouver.

Data

To solve this problem, I'll be using the New York neighbourhoods dataset we used in the course, as well as the Toronto neighbourhood data we scraped from Wikipedia. As for Montreal and Vancouver, I found these two web pages which contain postal code and neighbourhood data for all of British Columbia and Quebec, respectively. I plan to scrape the tables from these pages and then drop all rows that refer to neighbourhoods other than in the Vancouver or Montreal area.

After collecting the neighbourhood and postal code data, I plan to use foursquare to identify the types of venues in these neighbourhoods. I plan to use the distribution of types of venues as the feature to determine similarity between neighbourhoods in the different cities.

Data sources:
New York City - https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json
Toronto - https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M
Vancouver - http://www.geonames.org/postal-codes/CA/BC/british-columbia.html
Montreal - http://www.geonames.org/postal-codes/CA/QC/quebec.html