Compare the neighborhoods of two cities of New York and Toronto.

Background:

The City of New York, usually called either New York City (NYC) or simply New York (NY), is the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over a land area of about 302.6 square miles (784 km2), New York is also the most densely populated major city in the United States. A global power city, New York City has been described as the cultural, financial, and media capital of the world, and exerts a significant impact upon commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports.

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 2016. Current to 2016, the Toronto census metropolitan area (CMA), of which the majority is within the Greater Toronto Area (GTA), held a population of 5,928,040, making it Canada's most populous CMA. Its economy is highly diversified with strengths in technology, design, financial services, life sciences, education, arts, fashion, business services, environmental innovation, food services, and tourism.

Problem:
New York City and the city of Toronto are financial capitals of their respective countries of USA and Canada. Both of the city host multi-cultural environments.
New York and Toronto are huge cities. We are going to use data to find out how similar or dissimilar they are to each other.

Data acquisition and cleaning

Data Sources

These two cities are vast. We are going to make select of what we are going to work with.
We will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.
Luckily, this dataset exists for free on the web. here is the link to the dataset:
https://geo.nyu.edu/catalog/nyu_2451_34572
For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto. We will scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas dataframe so that it is in a structured format like the New York dataset.
We will use the Foursquare API to explore neighborhoods in New York City and city of Toronto.
We will use the explore function to get the most common venue and other categories in each neighborhood, and then use these features to group the neighborhoods into clusters. We will use the k-means clustering algorithm to complete this task. Finally, we will use the Folium library to visualize the neighborhoods in New York City, City of Toronto and their emerging clusters. We will compare the two city and come up with their similarities and dissimilarities.