Comparing Neighborhoods of New York City and Toronto

Anuj Chaudhary

May 11, 2020

1. Introduction

1.1 Background

New York City (NYC), is the most populous city in the United States. With an estimated 2018 population of 8,398,748. With almost 20 million people in its metropolitan statistical area and approximately 23 million in its combined statistical area, it is one of the world's most populous megacities. New York City has been described as the cultural, financial, and media capital of the world, significantly influencing 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 as of 2016. Current to 2016, the Toronto census metropolitan area (CMA) held a population of 5,928,040, making it Canada's most populous CMA. Toronto is also an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada.

Therefore, it is advantageous to compare the neighborhoods in two big cities and understand how similar or diverse are the two cities. This information will be used by cross-cultural researchers and economists to understand how diversity could play a big role in a city's success.

1.2 Problem Statement

We explored New York city and the city of Toronto and segmented and clustered their neighborhoods. Both cities are very diverse and are the financial capitals of their respective countries. The idea is to compare the neighborhoods of the two cities and determine how similar or dissimilar they are.

1.3 Interest

It will be interesting for the world to see how the different neighborhoods to two world's biggest cities could be similar or dissimilar.

2. Data acquisition and cleaning

2.1 Data Sources

We will start by looking at the different neighborhoods in both the cities. The data can be found for New York City and Toronto city here and here respectively. Within each neighborhood, we will explore different venues to form our core data frame for further analysis. For example, Marble Hill neighborhood in New York City has various venues like American restaurants, Donut shops, Bagel places, Yoga centers, etc.

2.2 Data Cleaning

Data downloaded from the sources were extracted into two tables. I used the Foursquare developer's tool to fetch all the venues of all neighborhoods for each city. Next, to figure out the top ten venues under each neighborhood, I count the number of times a venue appears under a category. This way we could use a mean to figure out the top ten neighborhoods. Now for each neighborhood, the top ten venues were captured in separate data frames. Now we used the clustering algorithms to figure out the various clusters within each city and compared them against each other to know how similar or dissimilar they are against each other.