IBM Applied Data Science Capstone - The Battle of Neighborhoods

Similarities Between Manhattan, New York, and Downtown, Toronto

1. Introduction

Tourism is a complex set of industries, including accommodation, recreation & entertainment, and food & beverage services. These days, people would like to visit different places like New York and Toronto to experience and explore their lives, heritage, and culture. Tourists need information regarding the nearby coffee shops, restaurants, hotels/accommodation, places to visit, and many more. Manhattan, New York, and downtown Toronto are very popular places people visit. This project aims to determine the similarity of these two cities are when it comes to tourism using data science. It will also assist tourists by providing them with the information needed to travel, comparing two places to make their decision depending on their choices.

The results of this project will guide then potential tourists who are planning a trip to visit Manhattan, New York, or downtown Toronto. A Data Scientist in this project would like to compare the similarities two places and identify tourist areas where to stay, popular places to visits etc.

2. Data sources

The data used in this project is public domain. To solve the problem, we need the following data:

Toronto data: we web scraped data from
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Manhattan, New York city data is available from https://geo.nyu.edu/catalog/nyu_2451_34572

Geospatial data is available from http://cocl.us/Geospatial_data

3. Methodology

The data extracted from the above webpages will have neighborhood data of different places in Toronto and New York. This project focuses on Manhattan, New York, and Downtown Toronto using the Foursquare API services. Foursquare API will provide the nearest venue data including their name, location, and venue type. We will apply a K-means Clustering machine learning algorithm to compare the two places Manhattan, New York, and Downtown, Toronto, and find the similarities.