The Battle of Neighborhoods

Comparing Toronto, Canada to US Cities

James Walsh

February 12, 2020
Coursera Capstone Project

Introduction

The focus of my project is to use Foursquare data for Toronto, Canada and three US cities: Chicago, Los Angeles and New York, and determine which US city Toronto is most similar to. The United States is very diverse and each of these three cities have markedly different histories and cultural influences as well as large spaces of land separating them. Determining which US is most like Toronto would be of value to anyone seeking to travel, move or open / expand their business in a new area.

Data

To start, I need to gather the neighborhoods and their respective coordinate data so I can make use of the Foursquare API. Methods for retrieving this data for Toronto and New York already been outlined in previous exercises, so I will need to find a way of collecting similar data for Los Angeles and Chicago. Once I have the neighborhoods and their coordinates for all three US cities and Toronto, I'll be able to pull the venue data for all neighborhoods in each of the four cities from Foursquare which includes up to 100 venues within a 500 meter radius of the geographic coordinates gathered for each neighborhood. After this data is cleaned and any duplicates removed, I'll be able to perform EDA and ML in order to determine which US city is most similar to Toronto. Creating a vector of the different venues in each city's neighborhoods will be the central feature engineering component to be used in ML.