Introduction/Business Problem

Two of the world's most popular tourist locations are Toronto and New York City. In many respects, they are diverse, with each neighborhood within those cities presenting different types of food and culture. The two cities are multicultural in nature, and they are their respective countries ' financial hubs as well. Our desire is to explore how similar or different these two cities are in terms of food, lodging, beautiful locations/attractions, and several other tourist criteria.

Today, tourism is one of the economic pillars of the world for most countries. People most frequently visit countries that are rich in heritage and well established from an international viewpoint, either economically or environmentally. Every city is unique and offers something different and exciting, especially for tourists. And nowadays, data pertaining to the sights and locations of everything on the planet is just a few clicks of a button away, making it easier and more accessible to explore than ever before. This crucial travel data is one of the most powerful tools that a tourist can have when it comes deciding where they would like to travel, as well as comparing multiple different locations to each other in order to see the unique qualities of each destination. But the data must be gathered and compiled first if the tourists hope to use it one day.

Data

We will use the Foursquare API service to gather and compile data about Toronto and New York. We will group this data in terms of each of the neighborhoods within the cities. This data will include information about the places around each neighborhood such as restaurants, hotels, coffee shops, parks, theaters, art galleries, museums, etc. Within each city, we will select one borough from each city to narrow down the area even more, and then we will move on to analyzing the neighborhoods within that borough. I have chosen to go with Manhattan from New York and Downtown Toronto from Toronto. We will use the machine learning technique known as “Clustering” to cluster the neighborhoods with similar amenities based on the neighborhood data we gather. These amenities will be given priority on the basis of amount of foot traffic (activity) in their respective neighborhoods. This will help to locate the tourist’s areas and hubs and where people tend to go the most. Then we can judge the similarity or dissimilarity between the two cities based on that.