1. A description of the problem and a discussion of the background:

Introduction / Business Problem:

This project addresses the need to identify the location for a new hotel in the center of one of the two most important financial districts in North America: New York City & Toronto. The problem is to figure out which of the two locations would be preferable to attract people and justify the high cost of buying the property for the hotel. This project will attempt to compare the neighborhoods of downtowns of New York and Toronto and showcase the similarities and dissimilarities between the cities and justify the reasons for preferring one or the other location for starting a new hotel business.

The final decision based on the most optimal location of the new hotel will be decided based on the number of interesting venues identified by the process of clustering of neighborhoods. A larger number of interesting venues in a cluster should indicate a happening area where the location of the new hotel would be preferred.

2. A description of the data and how it will be used to solve the problem:

Exploration of Dataset:

In this project, Foursquare location data will be leveraged to show the location or coordinates i.e. latitudes and longitudes of interests in each neighborhood and assist in the decision-making process of choosing the best neighborhood. The New York city dataset will be obtained from the geo.nyc.edu site while the Toronto dataset will be obtained form Wikipedia. From each of these datasets, only the location of interests pertaining to the borough of Manhattan in New York city and a downtown area of Toronto city will be explored and compared. Top 10 venues from each neighborhood will be clustered and explored in a greater detail.

Segmenting and clustering neighborhoods in New York and Toronto:

The explore function will be used to get the most common venue categories in each neighborhood, and to group the neighborhoods into clusters. K-means clustering algorithm will be used to complete the process of clustering. Finally, the Folium library will be used to visualize the neighborhoods in New York City and Toronto and their emerging clusters.