I. Objective

The purpose of this document is to describe the background of problem & data acquisition methods used for IBM data science capstone project.

II. Introduction/Business Problem

Singapore is an island nation with most number of visitors at any given time. Considering a well-known tourist destination, Singapore already has a lot websites in place to recommend places to visit or stay. However, all these available websites focus mostly on usual tourist attractions which may not be of interest for backpackers. The purpose of this project is to,

- Choose one of the recently opened MRT lines in Singapore, Downtown line (DT) and analyze the most common venues located along the MRT line so that visitors can choose a place to stay near any DT MRT station and explore all the venues.
- The project also intends to cluster the MRT stations based on the venues so that the visitors can choose to spend more time on the stations with venues of their interest.
- The project not only provides the top most common venues but also provides a table of 10 most common venues

For the project, MRT station names are extracted from the Wikipedia page on Singapore MRT lines and common venues in the neighborhood of MRT stations in downtown line.

III. Data Acquisition

This project will use the following data source:
MRT station names and associated details are retrieved from the open data source in Wikipedia page on all the MRT line in Singapore.
Website: https://en.wikipedia.org/wiki/List_of_Singapore_MRT_stations

The original data source contains stations for all MRT lines with names in multiple languages, alpha numeric codes, interchanges and other details like opening date etc. For the purpose of this project, only the alpha numeric code and respective station names in English has been chosen. Then the data has been modified to choose locations only in downtown line.

For the MRT stations of interest, location data has been retrieved using geopy package. The top venues recommendations are retrieved from Foursquare API. FourSquare API to explore neighborhoods in selected towns in Singapore. The Foursquare API was also used to explore the most common venues along the downtown MRT line.