**IBM Data Science Capstone Project**

**By Rajeswaran Srinivasan**

**1. Introduction**

Machine and Data Science are now at the heart of data. With enormous amount of data available these days, it is easily possible that the people are misguided. So it is imperative that the data are analyzed by experts and provide organizations, businesses with essential data in the best possible format, which can help them achieve greater success and at the same time provide people, customers with best experiences possible. Data Analyst bridges this gap and provide a smooth ride for all the parties.

For the purpose of this capstone project, I have chosen “LONDON CITY”, which attracts large tourists each year. London is one of the most popular cities in the world, which has plenty of interesting places to visit and things to do.

**2. Business Problem**

With lot of travelers each year to the city of London, it is important that they enjoy the best places and stay in the best hotels available. The choice of selection of hotels is extremely important such that it provides the travelers with the best comforts for relaxing and at the same time maximize the places of visits during their stay.

This document is aimed at helping an online travel agency (organization), which gives the best hotel suggestions for visitors to the City of London by providing the list of top hotels in their website according to the analysis done with the tools available for a data scientist.

**3. Data**

For this project the Foursquare API is used along with additional data taken from some of the websites. A list of hotels in London is downloaded using their respective location in longitude and latitude coordinates. The sources are the following:

* **Places of interest in London:**

https://data.london.gov.uk/dataset/cultural-infrastructure-map

* **Hotels in London:**

https://api.foursquare.com/v2/venues/search?client\_id={}&client\_secret={}&ll={},{}&v={}&query={}&radius={}&limit={}'.format(CLIENT\_ID, CLIENT\_SECRET, latitude, longitude, VERSION, search\_query, radius, LIMIT)

Dataset from the places of interest in London has information on different regions within the London City, showing Cinemas, Art Gallery, Museums, Libraries, Pubs, Music Studios, etc. Foursquare provide me the list of hotels around the city of London along with other details like ratings and total number of people who liked the hotel.

I have used both these datasets to identify the top rated hotels with most likes across regions and used them to compare it with the number of places of interests per region that would light up the travelers and keep them entertained throughout their stay.

This is how the sample data looks like from both the datasets. I have removed some of the unwanted columns for the purpose of the data analysis.

**Dataset 1 for Hotels in London through Foursquare:**



**Dataset 2 for places of interests in London:**

