**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.

**3. Methodology**

**3.1. Data Extraction**

For the purpose of this study, it is important to have ratings of each of the hotels and the total likes against each hotel by different users. All of these data are retrieved using the Foursquare API against the venue of interest around the City of London.

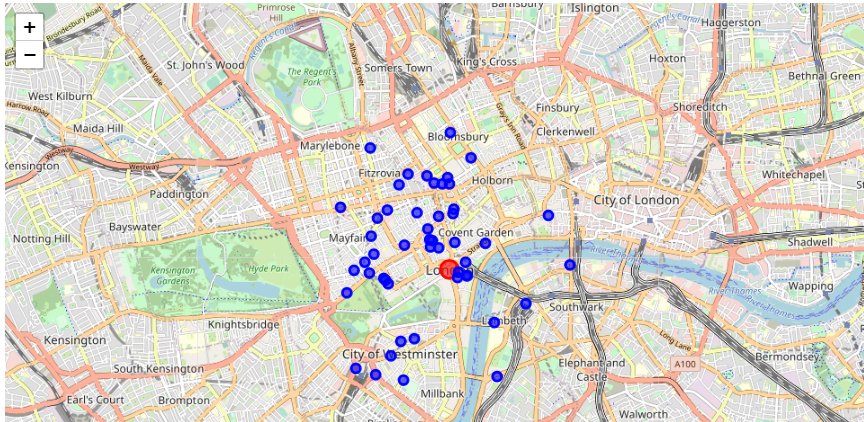
Another set of data containing the places of interest is obtained from the csv which is available in the web portal mentioned earlier.

**3.2. Data Cleaning**

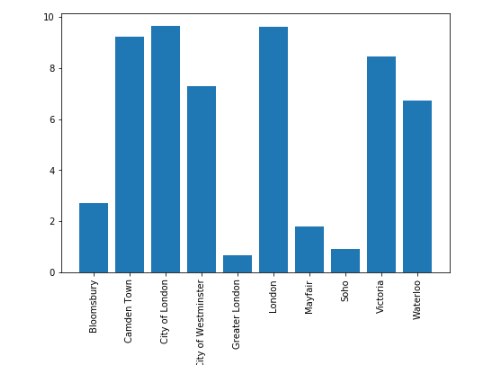
Having a complete data set without any missing or wrong data is clinical to the quality analysis of the Data Scientist. Removing unwanted columns of interest from the dataset is key and replacing NaN with mean values or getting rid of those rows of data is highly necessary. For the purpose of my research work here, I have removed the rows where the values are missing.

**4. Results**

Firstly, location of different hotels around the city of London is plotted in a geographical map using the Folium packages available.

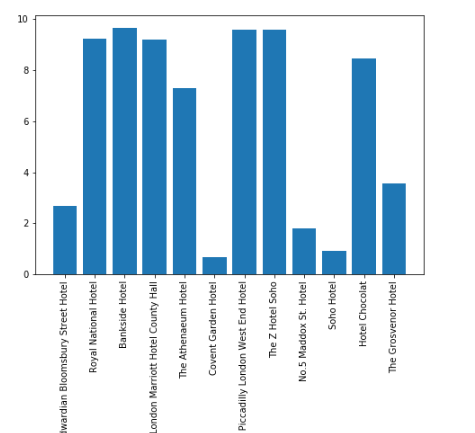


**Fig 1:** Hotels around London

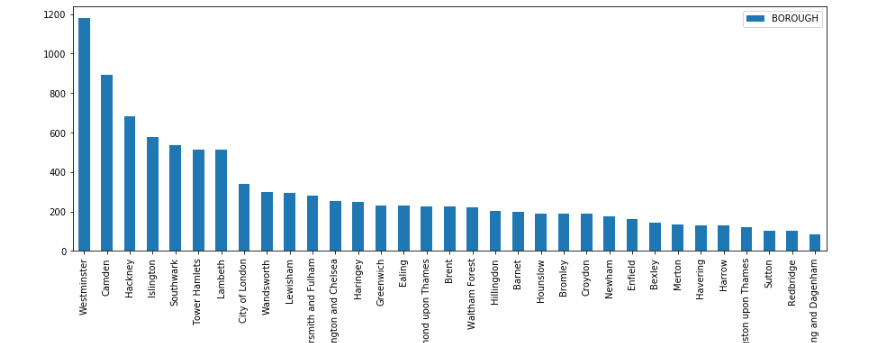
Secondly, bar charts showing the hotels with highest rating and highest numbers of likes are pitted against each other to show the most sought after hotels to stay in London.

**Fig 2:** Areas in London against the rating of the hotels

The hotels in and around the greater London city has the highest ratings, which is pretty obvious. Places like Victoria and Camden are not too far away as well.

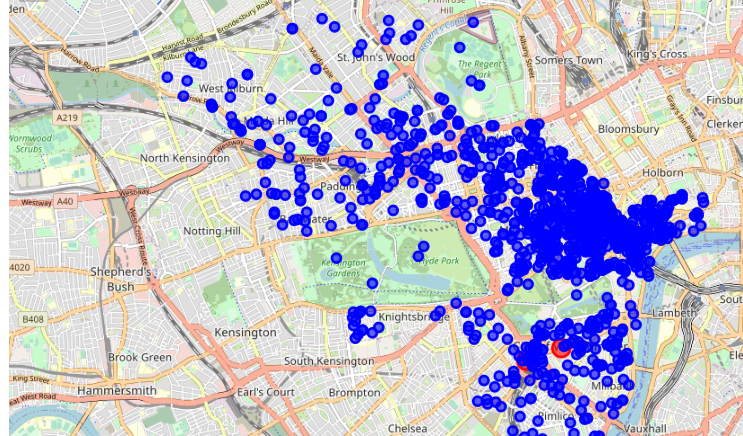


**Fig 3:** Hotels against the ratings



The above charts gives a clear indication on the number of areas of interest that tourists gets attracted to when they travel to London. The bar chart above shows interesting places against different areas/boroughs in London. Westminster has the most number of such places close to 1200, while the next best is only around 900, while the City of London has not more than 400 though.

Below two maps shows how close the hotels are to the different areas of interests for both Westminster and City of London (as mentioned only these 2 places are considered for this study).



**Fig 4:** Map of Westminster showing the different areas of interests as blue dots, while some of red dots are the hotels in this area

In the above map, Most of the red dots are swallowed inside the congested blue dots which gives the clear indication, how good the hotels are in terms of its close proximity to the areas of interests around the City of Westminster.

Same is not the case, with respect to the hotels in City of London, as the hotels are slightly away when you consider the below map.



**Fig 5:** Map of City of London showing the different areas of interests as blue dots, while some of red dots are the hotels in this area

**5. Discussion**

This is only useful if any one visiting the city of London for tourism. Hotel rating and likes along with the places of interest around these hotels are mainly targeted for the tourist audiences and not business people or other kind travelers. Moreover, for the purpose of this work, only the area around City of London and City of Westminster are considered. Though the same logic used can be implemented against other places as well. Given these constraints or assumptions in place, the model works pretty well to suggest the best hotel to stay for the visiting tourists.

**6. Conclusion**

The model can be used to a greater extent by the online travel agencies to help them in predicting/suggesting the best hotel to stay in London. Given the details are taken from Foursquare and other web portals mentioned earlier, with larger and more live datasets, the accuracy can be greatly improved to giving customers the best experiences that they carve for. This model can be widely used for other countries too. So large online travel booking portals can greatly benefit.