**Data Science Capstone Project - Full Project**

**Introduction**

The goal in this project is to recommend a best place to stay during the holiday in London

**Business case**

London is one of the largest cities in Europe - thousands of tourists visit this city every day. The first and obvious step in holiday-planning is - "Where we will stay" ? Then the real problem begins. Hundreds of places to sleep and hundreds of places to visit. Searching through the web might be inefficient as we would need to check tens of pages...

...and here is where we are coming with help

**Problems**

We are here to recommend the best choice of stay in London, but how we are going to do it ?

Couple of questions come into consideration

* How do we find hotels ?
* How do we grade them ?
* From where we obtain the data

**Resolution**

**First, lets describe the data source:**

All the data comes from Foursquare API - powerful application which contains geospatial data from all over the world. By using it, we will have access to all the hotel data in London as well as nearby venues which will help with recommendation process

**Second - grading method:**

We are going to assume that the best hotel would be the one, which has the highest online rating and has the most venues nearby.

Having those 2 questions answered - we have the fundamentals to resolve that 'hotel issue in London'

The final form of the recommendation will be the list of 5 best hotels using the assumptions above - This will surely help the people coming here :)

**Data Analysis**

Using GeoPy Python's library we are going to obtain the geographic localization of London. Once its done, utlizing Foursquare API, we are going to get full list of the hotels in the range of 10 km from the central city point. Once done, we will get online ratings for each hotel - this will help us to choose 5 best hotels in London. Once hotels are chosen, we will find nearby venues in the range of 1km from the hotel. The best hotel would be with highest online rating + with highest number of venues in a neighbourhood. All this data will be obtained using Foursquare API

Having in mind that venues might be in range of more than 1 hotel - we will use K-Means method to cluster them.   
  
First case is to generate the map of hotels in London

#### After that we have obtained data about ratings and then selected top 5 hotels in London Given the ratings and hotel data, we have captured venues around them and plotted on a map of London Looking at the map we can clearly see that 'Z' family hotels have most concentrated venues within the range. Given that 'Gloucester place' belongs to the 'Z' family and it has the highest rating - then we have our best choice of hotel in London! For the user information we can create quick handbook about top5 hotels and the most popular venue within the range

# Results and Discussion

This project has started from capturing the data about the hotels within the 10km range. Next using this data we have managed to obtain the ratings for them - this allowed us to select top 5 hotels within the range. As we know rating is not the only requirement for tourists, but a venues around. Given the hotel data we have managed to get the venues in a range of 1 km and succesfully visualized them on a map of London. When examining the map we can clearly see the biggest concentration of venues in the 'Z' family of hotels. The hotel with the highest rating overall is 'The Z Hotel Gloucester Place' - it does belong to the 'Z' family of hotels. Assuming that rating + venues concentration are these 2 main requirements we can state that 'The Z Hotel Gloucester Place' was chosen as the best place to stay in London.

**Future improvements**

* We can add to analysis the average price for a room and customer's preferences for the venue type, to prepare deep personalized analysis
* We can enhance this project to analyze every big city in the world

**Conclusion**

The purpose of this project was to help tourists to find the best place to stay - this is one of the most important factors when deciding about the trip. Preferences greatly differ for every individual, but there is one common value - "What is others opinion about this place" ? This has placed the fundamentals for this project, but we have gone 1 step ahead and tried to answer the second question "What i can do there once im placed in hotel" ? Combining these two charactersitics allows us to navigate to the best hotel in city - with highest rating and big concentration of venues nearby. We have made our choice in terms of which hotel is the best as per dataset, but we know that it still might not be the best for the person who is planning a trip. In order to satisfy the most people, we have created a small leaflet about the most popular venues around the top 5 places to stay - I'm sure this will be helpful to anyone going to London this year :)

**Thank you for reaching to the end! - have a great day**