**Business Problem:**

Booking.com collected **~515k** reviews given by various customers to **1493** luxurious European hotels across Spain, Netherlands, UK, France, Italy, and Austria. We were given this dataset of reviews to analyze thoroughly and find out meaningful insights that tell a nuanced story about data. After analyzing data, we had to write an **executive summary** consisting of around **250 words** which gives an overview of our analysis and also will help management to make the decision in a particular direction if they find insights meaningful. We presented 3 main static figures which highlighted different stories about data from different perspectives.

**Data:**

Dataset for the problem consists of 515k reviews which included information like positive words in a review, negative words, positive word count, negative word count, date review was given, reviewer's country, reviewer rating, rating to the hotel, family type, room type, address of hotel, no of nights visitor stayed, booking source, etc.

**Prime Objective:**

The main objective of doing this in-depth analysis was to find out meaningful insights about data and tell a nuanced story about findings written as executive summary. Executive Summary will help management to look at an overview of findings and make the decision rather than going in-depth to check analysis.

**Approach:**

Our main approach was doing analysis of data around country-wise performances of hotels giving more importance to UK hotels though as more than 50% of reviews consist of hotels from it. We wanted to look at hotels in which countries are getting visited most, which has good reviews, which has bad reviews, etc. Our steps involved data cleaning, data visualizations, and an executive summary. While we were finalizing particular charts for visualization, we also considered why, how and what of the plot as well as refined our choices a few times before finalizing particular charts and figures. The whole analysis was reproducible and performed in Jupyter Notebook using Python as prime language and matplotlib as the main plotting library.

**Steps:**

1. **Data Cleaning**

Our first step started with data cleaning which consists of loading data, removing unwanted rows that had many NaNs and derivation of new attributes from existing data to do a detailed analysis. As a part of data loading, we also tried to find basic stats like country-wise reviews, country-wise hotels, local customer percentages, etc.

2. **Visualizations**

The second step in the analysis consisted of debating plots, curating plots and creating figures which tell a unique story about data. Data Visualizations step involved plotting various types of charts and combining them to create big figure which highlights one or other way of looking at data. It also involved making a list of selection about charts, which color combinations to use, color bar, attribute, presentation theme, etc. After debating on the list of charts to present data, we ended up selecting a pie chart, bar chart, stacked bar chart, choropleth map and word cloud for our purpose.

Our first figure consisted of pie charts, bar charts, and stacked bar charts. The pie chart was useful to show the distribution of one attributes whereas Bar charts were useful to show distribution of one attribute compared to others. We also used the bar chart to show the trend of yearly visits. We used proper annotations in graphs to highlight important points along with proper titles and axis labels. Color selection was done in away from the color pallet that it helps someone easily distinguish items. All our plots had proper legends to highlight attributes. We also selected themes from other plotting libraries to improve the look of visualizations. All plots had grids on them for users to easily track particular values based on axes.

The second figure is choropleth graphs which were very useful to show the intensity of attribute in a particular region on the whole world map. We had used 2 choropleth graph in one figure to show a negative/positive relation. A positive relation was shown with various shades of green whereas negative relation was shown with various shades of red. Graphs were also annotated with top-n countries with the highest intensity for that attribute to draw attention to them. Both the graphs had their separate color bars which were used to show how different color shades are related to different intensity of attribute.

Our third choice of figure consisted of word clouds. They were used to highlight positive and negative words written in a review. Word clouds are best suited to highlight commonly occurring words from big text corpus given big size to frequently occurring words and less size to least occurring. We used a dark background for word clouds and light colors for text in word cloud so that word looks popping out of the chart. Like choropleth, word clouds were also used to analyze positive/negative relation by use of text as the main attribute. Word clouds were used to highlight which words were commonly used by customers when they were happy/angry/sad/frustrated with the experience of the hotel.

Above mentioned 3 figures were used to tell a story about reviews from 3 different perspectives.

## 3. Executive Summary

Once we were done with detailed analysis and visualizations, we condensed the contents of the jupyter notebook to create an executive summary. Executive Summary consisted of 250 words along with 3 main figures. It started with giving an introduction, explaining each figure and giving a few recommendations at the end based on findings.

**Conclusion:**

Detailed analysis and visualizing data helped us find out meaningful insights that won't be possible with just scrolling data and looking at it in excel. We were able to find out the insightful relationships between various attributes of data which won't be possible otherwise. Condensing such detailed analysis as an executive summary of 250 words helped us learn how to highlight meaningful insights with fewer words and make it to the point without wasting someone's precious time.