Capstone Project Submission

Instructions:
i) Please fill in all the required information.
ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Member Name:
- Mohammed Rafid Deshmukh – rafiddeshmukh@gmail.com

Contribution:
- Exploring Data
- Data Wrangling
- Data Cleaning
- Checking for null and duplicate values
- Removed outliers.
- Analyzed type of hotel booked
- Analyzed Distribution of Booking (Total, Confirmed, Canceled)
- Analyzed number of special requests
- Analyzed Moth and year wise book
- Analyzed countries of origin
- Analyzed arrival period and duration of stay
- Analyzed ADR (average daily rate)
- Performed EDA on data

Please paste the GitHub Repo link.

GitHub Link: - https://github.com/rafidddeshmukh/EDA-Hotel_Booking

---

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Problem Statement:

The Hotel data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things. All personally identifying information has been removed from the data.

This hotel data set can help to answer so many questions like best time of the year to book a hotel room, optimal length of stay in order to get the best daily rates, predicting whether or not a hotel is likely to receive huge amounts of booking requests which factor affect the booking.

Approaches:

The first step imported all the necessary libraries like NumPy, pandas etc. and then explored data and figured out all columns, all value data type, data shape and descriptive information about data. After that second step is data processing. Data processing is a process where raw data is converted or processed into clean and or readable format.

The data is based upon two hotels: resort hotel and city hotel. This data set has 1,19,390 rows and 32 columns. We can say this is a huge amount of data but we used the panda's package so that it is easier. Firstly, I removed all the null values, outlier and undefined values.

Then we moved on to the visualization part where I tried to extract some insights from the given data set. I created bar graphs, pie charts, box plots, subplots and count plots. I also plotted a heat map to get correlation between variables. From this visualization I get the answer which we are looking for.

Conclusions:

Almost 61% of the total guests opt for a city hotel instead of a resort hotel. In both hotels 19% to 24% bookings were canceled. Summer months are more expensive for resort hotels while the prices for city hotels don't really fluctuate throughout the year.

Special requests increase in summer time when booking is high. Almost all of the guest's holiday without any children or babies. Majority of the guests booked the hotels through both online and offline travel agents and tour operators

To get the most bang for your buck you should book at least 175 days in advance, make the booking directly with the hotel and also join any loyalty program.