

EDA Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

Name: Gauri Agarwal

Email: gauriagarwal4444@gmail.com

Contributions:

- Data Preparation and Cleaning
 1. Read the data from csv file.
 2. Removed outliers, visualize it using boxplots.
- Exploratory Data Analysis and visualizations
 1. Analysis on market_segment column.
 2. Analysis on lead_time column.
 3. Analysis on is_canceled column.
 4. Relation between market_segment, month and year column.
 5. Calculate which market segment have more retention rate.
 6. Various types of Distribution Channels used percentage.
 7. Finding top agents and agent those are not very active.
 8. Finding room type preferred by guests.
 9. Division of deposit type offered by hotels.

Name: Saurav Kumar

Email: bhadanisaurav7@gmail.com

Contributions:

- Data Preparation and Cleaning
 1. Removed and replaced all the null values.
 2. Added two required columns in the data set.
- Exploratory Data Analysis and visualizations
 1. Analysis on hotels column.
 2. Analysis on country column.
 3. Visualize how various factors effect adr.
 4. Calculated ADR per month of each hotel.
 5. Calculated monthly adr for each month of both types of hotel.
 6. Calculated which market_segment generates more revenue.
 7. Finding meal preferred by guests in both types of hotels.
 8. Analysis of Customer_type column.

Please paste the GitHub Repo link.

Github Link:- https://github.com/gauriagarwal18/Hotel_Booking_Analysis.git

Drive Link:- <https://drive.google.com/drive/folders/1wFfPPqSWW4Mj7oS4-mmaH0GemD7Dq7v8?usp=sharing>

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

Hotels Booking data set have the record of the bookings done between July-2015 and August-2017, from the data set we note that hotels are categorized into two categories one is city hotel and the other is resort hotel. Each hotel have different attributes like is_canceled which show whether a booking is canceled or not, lead time which shows the time between booking is made and guests actually arrived to hotel, market segment which shows the strategy which drives the customer to a hotel, adr which shows the revenue generated and many other important attributes.

After understating the data we loaded the data into our colab notebook in pandas dataframe, after loading the data we removed noise from the data, removed or replaced all the null values and removed outliers using upper and lower quantiles and then we performed feature engineering and then add two columns total_guests and days_spent.

Then after cleaning, data is ready for visualization, we first selected some columns which seems to be more important like type of hotel to note the difference between facilities available and their requirements in different countries,

is_canceled to check what are the conditions when the booking made is canceled, lead_time to visualize what is the most probable lead time, in what conditions lead time is more, next column that we visualized was market segment it is one the most important one and we make out what are the most used market segments and the revenue generated by them and their retention rate,

Next very important column was adr we visualized what is revenue generated by both the hotels, revenue generated by various distribution channels.

We also visualized all these graphs month wise and year wise as well to see a pattern and to figure out some ups or downs in past three years, we also look for various deposit types, meal preferred, type of room demanded most, and try to connect all the visualizations and draw some worthy conclusions.

Through these visualizations we drew some useful conclusions like we analyzed what will be the best time for a guest to make a booking depending on the adr and the chances when the bookings are available in the hotel, we checked for the factors where resort hotels are lagging behind and receiving less bookings by various market_segment and by analyzing facilities provided by the two, we also checked how we can reduce our loss due to cancelation by taking some extra bookings in advance depending the chances of booking get canceled and so we try to visualize some traits and features of data which may prove to be beneficial for data guest and hotels as well.