

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

Contributor Roles:

1. Sai Krishna Vamshi Devarasetty - krishnavamshidevarasetty@gmail.com
 1. Data Wrangling
 2. Analysis on reserved room vs assigned room
 3. Correlation Analysis between Repeated Guest and Cancellation Status
 4. Distribution Analysis based on customer type
2. Anshita Gawade - anshitagawade@gmail.com
 1. Data Wrangling
 2. Booking Analysis based on month of booking.
 3. Analysis based on the preferred meal of a booking.
 4. Country wise Analysis and distribution.
 5. Weekday vs Weekend booking analysis
3. Mrudula S Prasad - mrudulamrudu1808@gmail.com
 1. Data Wrangling
 2. Correlation Analysis between meal type chosen and Cancellation status
 3. Analysis based on market segment
 4. Correlation Analysis between deposit type and cancellation

Please paste the GitHub Repo link.

Github Link:- <https://github.com/anshitagawade/Hotel-Booking-Analysis>

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

Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests?

Exploratory Data Analysis helps us in understanding the data and in exploring some questions related to users on the data provided.

This Dataset contains data that compares various booking information between two hotels, City Hotel and Resort Hotel. We will be using the data to analyze the factors affecting the hotel bookings. These factors can be used for reporting trends and predicting future bookings.

Each observation represents hotel bookings. Bookings arrive between the 1st of July of 2015 and the 31st of August 2017, including bookings that effectively arrived and bookings that were canceled. Since this is hotel real data, all data elements personal data or customer identification was deleted.

The dataset contains over 119,390 rows and 32 columns. Columns have the data type object, int64 and float64. It appears that a few columns contain some empty values since the Non-Null count for a few columns is lower than the total number of rows (119390).

The tools for data analysis used in this project are the packages Numpy and Pandas, and to visualize and explore the data: Matplotlib and Seaborn.

Then we've cleaned up and prepared the dataset for analysis by removing duplicates and working on null values.

Then we did some analysis by putting some questions which conclude as follows:

- Majority of hotel bookings are from city hotels.
- The cancellation rate for hotels is 27.5%
- We should target July and August as most of the bookings. Those are peak months due to the summer period.
- The number of bookings seems to be high in 2016 while the bookings seem to be less in 2015 and 2017. This is majorly due to more data points being available in 2016.
- BB meal is the most preferred meal by customers.
- It appears that a disproportionately high number of bookings are from Portugal, probably because the hotel is located in Portugal itself.
- Most people do not seem to prefer to stay at the hotel for more than 1 week.
- We observe that Weekday bookings are higher than the Weekend numbers. That is an interesting finding.
- When we find a correlation between assigning different rooms for having children/babies there is a low probability of getting a room in a resort hotel as compared to a city hotel.
- The percentage of cancellation for Non Repeated User is higher when compared to repeated users for both City and Resort Hotels.

- There's some positive correlation between lead_time and cancellation status i.e., the higher the lead time the more chances of cancellation and avg lead time of canceled booking is 40 days more when compared to not canceled bookings.
- Having children/babies have less special requests in Resort Hotel when compared to City hotel.
- Even though overall bookings of City Hotel are higher when compared to Resort Hotel the percent of growth in number of booking from 2015 to 2016 is much higher in City Hotel.
- Customer Type distribution has the same trend in both types of Hotels, as transient type of customers are higher in both hotels.
- Type A room is assigned to most of the customers in both the hotels.
- Agent 9 is made most bookings in both the hotels. So we can focus more on agent 9 for more profitable business.
- Most bookings are through Online TA. Also 'Online T A' has the highest cancellation rate of both types of hotels. To avoid this the hotels can give additional discounts or offers to people who book through Online TA's.
- Repeated guests are very few which only 3.9 %. we should target our advertisement on guests to increase returning guests. Also should take feedback and reviews and should work on that.
- We also realize that the high rate of cancellations can be due high no deposit policies
- We notice that most of the records require no or at most 1 car parking spaces, though there are records that show more number of parking spaces. This could be families traveling together and larger business trips.