

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

1) Bindu Kovvada (bindukovvada187@gmail.com):

Getting the data

Data Cleaning

Data Analysis:

- 1) Percentage of bookings in each hotel type?
- 2) Which is the busiest month for hotels?
- 3) How many bookings were cancelled?
- 4) From which country most guests come?

2) Saksham Tripathi (saksham757474@gmail.com):

Getting the data

Data Cleaning

Exploratory Data Analysis:

Working on Outliers

Heatmap

Data Analysis:

- 1) Which room type is in most demand?
- 2) Which room type generate highest adr?
- 3) Let's see does not getting same room affects the adr?
- 4) Which is the most common channel for booking hotels?

3) Shivangi Mishra:

Getting the data

Data Cleaning

Data Analysis:

- 1) Booking cancellation and repeated guest
- 2) Distribution channel and days on waiting list?
- 3) Which distribution channel has highest cancellation percentage?
- 4) Days on waiting list and market segment?

4) Deepak Kumar Gautam:

Getting the data

Data Cleaning

Data Analysis:

- 1) Comparing different variables month wise?
- 2) What is the preferred stay length in each hotel?
- 3) Which meal type is most preferred meal of customers?

5) Satyajit Sahoo:

Getting the data

Data Cleaning

Data Analysis:

- 1) Checking whether not getting allotted the same type as demanded is the cause of cancellation of bookings?
- 2) Effect on lead time on cancellation?

3) Which type of customers are most repeated?

Please paste the GitHub Repo link.

<https://github.com/SakshamTripathi1996?tab=repositories>

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 in order 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? This hotel booking dataset can help you explore those questions!

This data set contains booking information for a city hotel and a resort hotel including dates of booking, length of stay, number of adults, children, and/or babies, and available parking spaces and other things. All personally identifying information has been removed from the data. Exploring and analyzing data to identify key elements that influence bookings.

We started with importing all the necessary data analysis libraries such as NumPy, pandas, seaborn, and matplotlib. Then after, we mounted our drive, specified the file path and read the CSV file.

Then we moved on to the data exploration part where we looked at many aspects of our data set such as shape, head, tail, information and description. Then we moved on to find and remove all duplicate rows in the data set. Missing values are also identified and filled. We then updated the datatypes of the columns.

Then we moved on to the visualization part where we tried to extract some insights from the given data set. We started with outliers and then used a heat map to look for correlation.

Then we talked about some important questions which the hotel manager must face before making decisions. We created bar graphs, pie charts, box plots, subplots, count plots, scatter plot.

Finally, we came to some important conclusion such as: higher lead time has

higher chance of cancellation, July and August are the busiest months, most people prefer to stay in hotels for at least 5 days, most repeat guests are transient type, most guests are coming from Portugal, bookings are not affected if customers do not receive same room reserved and many more conclusions.

Thank you