

# Capstone Project Submission

## Instructions:

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

### **Team Member's Name, Email and Contribution:**

**Name:** Arshad Aafaq D

**Email:** [aafaqclassic@gmail.com](mailto:aafaqclassic@gmail.com)

#### **Contribution:**

- Cleaning of **Data Set** removing null values.
- Contributed to the **Exploratory Data Analysis**
- Analyzed the correlation between both the data sets using heatmaps.
- Contributed in the Data Wrangling.
- Data Visualization.
- Prepared the presentation
- Prepared the summary
- Contributed in the collab notebook
- Fetched the insights from the data set.
- Contributed in writing inferences and conclusions of the EDA.
- Contributed in technical documentation.

Name- Sakshi Chaturvedi

Email-[sakshichaturvedi0207@gmail.com](mailto:sakshichaturvedi0207@gmail.com)

#### **Contribution: -**

- Importing all the important libraries.
- Importing data from the local drive.
- Understanding the dataset (like shape, head, data information)
- Removing the null values and duplicates.
- Prepared the Presentation.
- Contributed in Technical documentation.
- Data pre-processing (like data cleaning handles null values).
- Done 7 questions solved (example -What is the booking rate according to the population?)
- Data Visualization
- Conclusion.

Name: Vikas Kumar

Email: [er.vikas2268@gmail.com](mailto:er.vikas2268@gmail.com)

#### **Contribution:**

- Prepared Technical Documentation.
- Contribution to Presentation.
- Contribution to team colab.
- Find the missing value, then delete it from the dataframe.

- Used heatmaps to analyze the relationship between the two data sets.
- Examination of both hotel and reservation data year-wise.
- Hotels lead time analysis.
- Based on Deposit type analysis.
- Contribution to writing inferences and conclusions of EDA.

Name : Kaveri Shende

Email : [kaverishende@gmail.com](mailto:kaverishende@gmail.com)

**Contribution:**

- Analyzed the correlation between the Data Sets
- Prepared the presentation.
- Contribution in Team Collaboratory
- Data cleaning and manipulation
- Contributed to technical documentation.
- Contributed to writing inferences and conclusions of the EDA
- Prepared individual collab
- Prepared the video.

Name : Yogesh Shivraj Agre

Email : [yogeshagre62@gmail.com](mailto:yogeshagre62@gmail.com)

**Contribution:**

- In the data cleansing section, look for null values. Instead of using the 'isnull' method
- directly, I created a separate function called "null details" to find all of the data details,
- including null values, unique values, total values, and datatype.
- Recognized the variety of hotels that appear in our data.
- Determining which hotels are preferred by which age group.
- Focus is primarily on the presentation and colab portions.
- Investigated cancellation data and made several attempts to find answers.
- Participated in the writing of the EDA's inferences and conclusions..

**Please paste the GitHub Repo link.**

**GitHub Link:-**

- Arshad Aafaq D- [https://github.com/arshadaafaq/Hotel\\_Booking\\_EDA-Capstone\\_Project.git](https://github.com/arshadaafaq/Hotel_Booking_EDA-Capstone_Project.git)

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

## **SUMMARY:**

- The purpose of our project was to gather and analyse detailed information about hotels in order to provide insights and estimate the profit.
- We have given two hotel data sets. i.e., the resort hotel is one of the hotels, and the city hotel is the other. There are 32 columns and 119390 rows.
- The majority of Revenue Management research on demand forecasting and prediction issues is conducted in the tourism and travel-related industries.
- It would be fantastic if the hotel management team could identify the root cause and develop a better strategy.
- The goal of our project was to collect and analyse detailed hotel information in order to provide insights and estimate profit.

## **CLEANING OF DATA:**

- We focused more on data uniformity to ensure that we explored the best results from our analysis. We took a few steps with the data, such as removing NAN and duplicate values. The mean value was used to replace a few NULL values. We ensured that data consistency is maintained throughout a feature. We used the cleaned data to conduct Exploratory Data Analysis to better understand our dataset. Using heatmaps, we investigated the correlation between various features. We were able to gather some useful information by extrapolating graphs and plots.

## **ANALYSING:**

- After cleaning the data, we tried to find some insights from the given data, using which we could increase the profit. And few of them are listed below.
- we found that, from the above 2 hotels, the city hotel was most preferred by the customer, most of the returning customers are from the resort hotel, and TA/TO are the most chosen distribution channel by the customers.
- In comparison to City Hotels, the ADR for Resort Hotel is higher in the months of June, July, and August. Perhaps clients/people wish to vacation in resort hotels this summer
- More than 25000 people, or the majority of the attendees, are from Portugal.
- And we tried to find the root cause, and we found that customers had cancelled 72.51% of their reservations. Now, more investigation is required to determine the potential cause of the cancellation.

## **CONCLUSION:**

The datasets provided have tremendous key performance indicators to improve business opportunities while having a huge positive impact on both users and stakeholders. The project objectives of analysing hotel data and determining key factors responsible for success were met as a result of the EDA process that I followed. Companies and developers can use these insights to strategically plan their business moves.

This 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.

The Hotel booking data set includes Numerical, Categorical and Binary data. The data set has columns like the hotel type, is\_canceled, arrival\_date\_year, arrival\_date\_month, stays\_in\_weekend\_nights, stays\_in\_week\_nights, country, market\_segment, distribution\_channel,

etc. which helped us draw major insights from the data set. Our aim here is to understand the important factors that govern hotel bookings.

Hotel Booking is governed by many factors including the time of the year, number of guests, distribution channel, hotel type, etc. The problem statement for this project is to perform Exploratory data analysis and draw insights to understand all the important factors that govern the Hotel bookings.

We have performed analysis such as univariate analysis, Hotel wise analysis, Time wise analysis and also tried to answer other important questions to solve the problem statement.

We tried to answer the questions such as:

- Which type of hotel is mostly preferred by the guests.
- Adults traveling with kids or without kids?
- What are the most preferred distribution channels for booking hotels?
- What is the percentage of cancellation?
- Which hotel has higher bookings cancellation rate?
- What is the percentage distribution of “customer Type”?
- What is the Percentage of repeated guests?
- What is the percentage distribution of required car parking spaces?
- What is the percentage of booking changes made by the customer?
- Which type of food is mostly preferred by the guests?
- From which country the most guests are coming?
- In which month most of the booking happened?
- Which year and hotel had highest booking?

We used different libraries to form tables and graphs in order to understand and answer these questions. We go to know that:

- City Hotel is most preferred hotel by guests thus city hotel has maximum bookings.
- The guest visit in the hotel without kids is more than with kids.
- TA/TO is mostly (82.2%) used for booking hotels.
- 37.1% bookings were cancelled.
- Resort has higher cancellation rate.
- Percentage of repeated guests.
- Repeated guests are very few which only 3.2% it means in order to retained the guests and try to improve the services.
- 93.9% guests did not require the parking space only 6.1% guests required only 1 parking space.
- Almost 82% of the bookings were not changed by the guests.
- The most preferred meal type by the guests is BB (Bed and Breakfast), HB- (Half Board) and SC-(Self Catering) are equally preferred.
- Most of the guests are coming from Portugal that's 4800 guests are from Portugal.
- July and August months had the highest number of bookings.