

# Capstone Project Submission

## Instructions:

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

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

#### **1. Jayesh Yadav ([yjayesh1@gmail.com](mailto:yjayesh1@gmail.com))**

- 1.1 Data Processing
- 1.2 Data Wrangling
  - 1.1.2 Data Cleaning
    - 1.1.1.1 Fixing all the missing values in the data set.
    - 1.1.1.2 Percentage of null values in data set and total null values in data set.
    - 1.1.1.3 Heatmap for null values.
- 1.1 EDA - Hotel wise analysis
  - 1.1.1. Bar graph to check which hotel was more engaging and in demand.
  - 1.1.2 Pie chart to show percentage of customer of hotel in 3 years
  - 1.1.3 Finding out the number of reservations cancelled in the last 3 Years.
  - 1.1.4 Finding the most profitable year for hotel in the given time span of 2015-2017 .
  - 1.1.5 Checking the total cancellation of booking between 2015 -2017 for both hotels.
  - 1.1.6 Number of Cancelation in individual years from 2015-2017
- 1.3 Conclusion

#### **2. Piyush Kutemate ([kutematepiyush@gmail.com](mailto:kutematepiyush@gmail.com))**

- 3.1 Data Wrangling.
- 3.2 Exploratory Data Analysis-Univariate Analysis.
  - 3.2.1 Finding out numbers of travelers stayed on weekend night on the basis of their arrival in given years (2015-2017).
  - 3.2.2 Finding out numbers of travelers staying on week night on the basis of their arrival in given years (2015-2017).
  - 3.2.3 Figuring Top 3 ways Travelers preferred to travel with the help of Graph.
  - 3.2.4 Finding out and showing the percentage of travelers who travel without Kids with the help of pie-chart.
  - 3.2.5 Finding out what type of hotel provides what type of Meal with table.
  - 3.2.6 Finding out Top 5 Countries Guests/Travelers Visiting From.
  - 3.2.7 Finding out which month has more number of guests in all 3 years.
- 3.3 Observations.

#### **3. Prince Chauhan ([princechauhan608@gmail.com](mailto:princechauhan608@gmail.com))**

- 4.1 Data Wrangling.
- 4.2 Exploratory Data Analysis -Bivariate Analysis.
  - 4.2.1 Finding out how much hotel bookings were made by which

- market segment.
- 4.2.2 Finding out which distribution channel is used for most of the hotel bookings.
- 4.2.3 Checking the month which has the highest number of arrivals for individual year.
- 4.3 Exploratory Data Analysis - Hotel wise Analysis
  - 4.3.1 Finding out the number of repeated guests counted by hotel type.
  - 4.3.2 Finding out the number of bookings with no deposit and refundable deposits.
  - 4.3.3 Finding out the most preferred room type.
- 4.4 Exploratory Data Analysis - Some other important questions
  - 4.4.1 Finding out the number of guests that change their bookings.
  - 4.4.2 Finding out the types of customers.
  - 4.4.3 Finding out the requirements of car parking.
  - 4.4.4 Finding out the number of special requests.
  - 4.4.5 Finding the most profitable year for individual hotel i.e for City hotel and Resort hotel
  - 4.4.6 Finding out which month is busiest and profitable for the hotels.
- 4.5 Conclusion

#### **4. Mangal Lokhande (mangallokhande2016@gmail.com)**

- 5.1 Data Wrangling
- 5.2 Exploratory Data Analysis-Multivariate Analysis.
  - 5.2.1 Brief Picture of column trends with the help of Histogram.
  - 5.2.2 Show the requirement of car parking spaces year and month wise by creating a line chart graph.
  - 5.2.3 Plot bar graph of booking made through market segments by different countries.
  - 5.2.4 Correlation Analysis.
    - 5.2.4.1 Correlation graph of the data using heatmap.
    - 5.2.4.2 Brief of insight drawn from the exploratory data analysis.

**Please paste the GitHub Repo link.**

GitHub Link:- <https://github.com/PrinceChauhan608/HOTELDBA-EDA-CAPSTONE>

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

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.