

# Capstone Project Submission

## Instructions:

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

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

- Karan Rawat ([rawatkarankr99@gmail.com](mailto:rawatkarankr99@gmail.com))
  - **Data Wrangling**
    - The Data file consists of some null values "Nan".
    - Replacing those null values with zero, median and mode.
    - Dropping some of the rows which don't have values.
    - Convert data types from float to int.
  - **EDA-Booking wise analysis**
    - How many booking cancelled each year?
    - What is the booking difference between weekends and week day's night?
    - From which market segment bookings done the most?
    - What is the bookings percentage each year?
  - **EDA-Type of visitors wise analysis**
    - Which is the most booked accommodation type?
    - From which country visitors comes the most?
  - **EDA-Month wise analysis**
    - Which is the most occupied month for hotels?
  - **EDA-Room wise analysis**
    - Which room type has the most demand?
    - How many rooms wrongly assigned to with respect to booked room type by each hotel?

Other than this I prepared summary, PPT report and Technical document for this

- Abhash Jain ([abhashjain9@gmail.com](mailto:abhashjain9@gmail.com))
  - **Data Wrangling**
    - The Data file consists of some null values "Nan".
    - Replacing those null values with zero, median and mode.
    - Dropping some of the rows which don't have values.
    - Convert data types from float to int.
  - **EDA-Booking wise analysis**
    - How many booking cancelled each year?
    - What is the booking difference between weekends and week day's night?
    - From which market segment bookings done the most?
    - What is the bookings percentage each year?
  - **EDA-Type of visitors wise analysis**
    - Which is the most booked accommodation type?
    - From which country visitors comes the most?
  - **EDA-Month wise analysis**

- Which is the most occupied month for hotels?
- **EDA-Room wise analysis**
  - Which room type has the most demand?
  - How many rooms wrongly assigned to with respect to booked room type by each hotel?
- Other than this I prepared summary, PPT report and Technical document for this study.
- Mounika Dontula ([mounikadontula2795@gmail.com](mailto:mounikadontula2795@gmail.com))
  - **Data Wrangling**
    - The Data file consists of some null values "Nan".
    - Replacing those null values with zero, median and mode.
    - Dropping some of the rows which don't have values.
    - Convert data types from float to int.
  - **EDA-Booking wise analysis**
    - How many bookings canceled each year?
    - From which market segment bookings done the most?
    - What is the booking percentage between city hotel and resort hotel?
    - What is the bookings percentage each year
  - **EDA-Guest wise analysis**
    - How many guests repeated each year?
  - **EDA-Type of visitors wises analysis**
    - From which country visitors comes the most?
  - **EDA-Month wise analysis**
    - Which is the most occupied month for hotels?
    - What is the average daily rate for each month for hotel type?
  - **EDA-Room wise analysis**
    - Which room type has the most demand?

Other than this I prepared summary, PPT report and Technical document for this study.

**Please paste the GitHub Repo link.**

Github Link:- <https://github.com/rawatkarankr99/Capstone-Project-1>

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

We first decided to take up this project solely due to our mutual interest in hotel booking. Then we decided first work individually gaining insights doing some eda etc. I started to form the questions and we discussed with team member. After doing some random EDAs. I gained some confidence.

We used the given dataset that contains data about hotel bookings between 2015-2017 like booking cancelled, booking channels, booking date etc. We cleaned and pre-processed the data and then we performed the exploratory data analysis to extract information from the data as per below:

#### **Booking wise analysis**

- The total bookings got cancelled 44199 i.e. 37% of total booking. Most of the booking cancelled for City Hotel during the year of 2016 and 2017 that is 61% of total booking cancelled.
- The total bookings got canceled 44199 i.e. 37% of total booking. Most of the booking cancelled for City Hotel during the year of 2016 and 2017 that is 61% of total booking cancelled.
- More than 60% of the population booked the City hotel
- More bookings were made in 2016, compared to the previous year. But the bookings decreased by almost 14% the next year.
- As we can see from analysis most bookings were made by the online TA market segment. And the least bookings were made by aviation and complementary market segments.

#### **Guest wise analysis**

- In the resort hotel 7334 rooms were assigned wrong i.e 25.4% of the total reserved room type in the resort hotel. In the city hotel 6661 rooms were assigned wrong i.e 14.5% of total reserved room type in city hotel.
- The 280 guests were repeated in the year 2015 , 1619 guests were repeated in the year 2016 and 1306 guests were repeated in the year 2017.

#### **Type of visitors**

- Majority of visitors travel with family. The visitors with babies prefer mostly the resort hotel
- The majority of the bookings are from Portugal. The countries UK, France, Spain and Germany also hold a great portion in bookings. The approx 70% comes from these 5 countries.

#### **Month wise analysis**

- In the month of August, the maximum number of bookings was made. Then the second number of majority of bookings was made in the month of July. In the month of January there are the least number of bookings.
- The adr for City Hotel is highest for the months May and August, the adr for Resort Hotel is highest for the August month and the adr for City Hotel is more expensive than Resort Hotel for each month.

#### **Room wise analysis**

- The room types 'A' have the most demand and the adr is also high.

- In the resort hotel 7334 rooms were assigned wrong i.e 25.4% of total reserved room type in resort hotel. In the city hotel 6661 rooms were assigned wrong i.e 14.5% of the total reserved room type in city hotel.
- The room type 'A' has the highest adr.

**Challenges**

- (1) Data was present in the wrong data type format.
- (2) Choosing appropriate visualization techniques to use was difficult.
- (3) A lot of null values were there in the dataset.