

EDA Capstone Project :

Hotel Booking Analysis

Team Members :

Prasad Patil

[Email: prasadpatila99@gmail.com]

[Github: <https://github.com/prasadpatil1999/EDA-Hotel-Booking-Data-Analysis>]

Prasad Deore

[Email: prasaddeore07@gmail.com]

[Github: https://github.com/prasaddeore007/Hotel_Booking_Data_Analysis]

Kiran Wagh

[Email: patlanchagopal119@gmail.com]

[Github: <https://github.com/Kiran2611/HOTEL-BOOKING-DATA-ANALYSIS>]

Index of content

- Introduction
- Methodology
- Analysing the data using EDA
- histogram for all the columns in the dataset
- Room Booking by the Agent
- Maximum booking over the year and month
- Analysis based on Cancellation
- Cancellations made in each Year, Cancellation made in each hotel type
- Cancellations made in each months, Analysis of cancellations on deposits
- Cancellation based on customer type
- Which significant distribution channel has highest cancellation percentage
- Preferred Room type
- Percentage of Repeated Guests in hotels
- From where the most guests are coming
- Observation
- Conclusion

INTRODUCTION

The purpose of this analysis is to understand the customer behavior while booking a hotel room. The analysis will help in understanding what factors influence the customers while booking and what are their preferences. This will help the hoteliers in providing better services to their customers and improving their business.

The dataset used for this analysis is the Hotel Booking Demand dataset. It contains information on hotel bookings made by customers from 2015 to 2017. The data includes information on the customer id, country of origin, hotel, arrival date, departure date, adults, children, infants, lead time, stays in weekend nights, stays in weekday nights, meal plan, market segment, distribution channel, is repeated guest, previous cancellations, previous bookings not cancelled, reserved room type, assigned room type and booking changes.

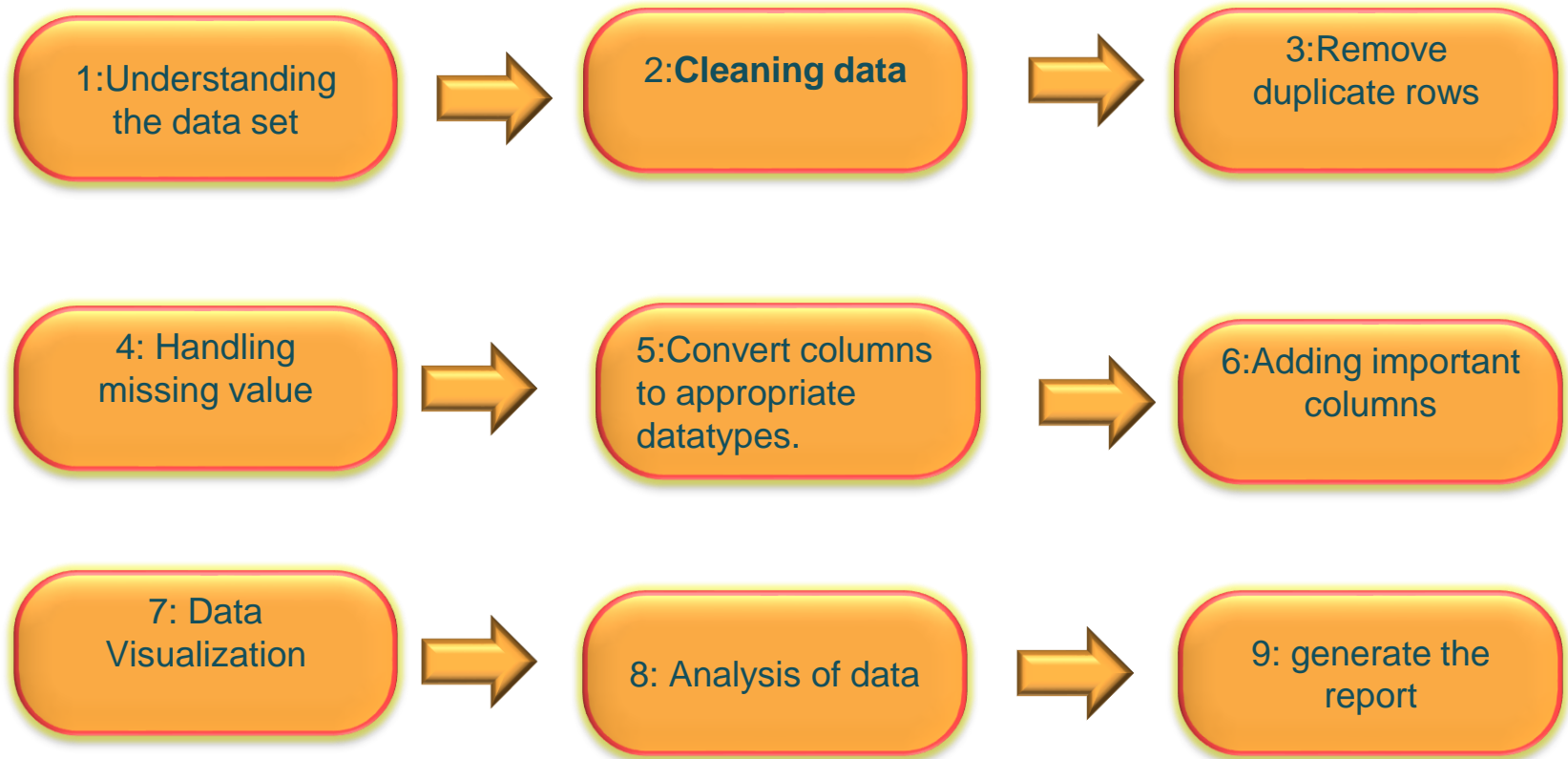
The first step in the analysis is to understand the customer behavior while booking a hotel room. The analysis will help in understanding what factors influence the customers while booking and what are their preferences. This will help the hoteliers in providing better services to their customers and improving their business.

Attributes in the study

→ This data has 119390 rows and 32 columns. Here are the columns

- Hotel
- Is_cancelled
- Arrival_date_year
- Arrival_date_month
- Is_repeated_guest
- Arrival_date_week_number
- Day_of_the_month
- Stays_in_weekend_nights
- Previous_cancel
- Stays_in_week_nights
- Adults
- Children
- Babies
- Meals
- Country
- Market_segment
- Distribution_Channel

METHODOLOGY

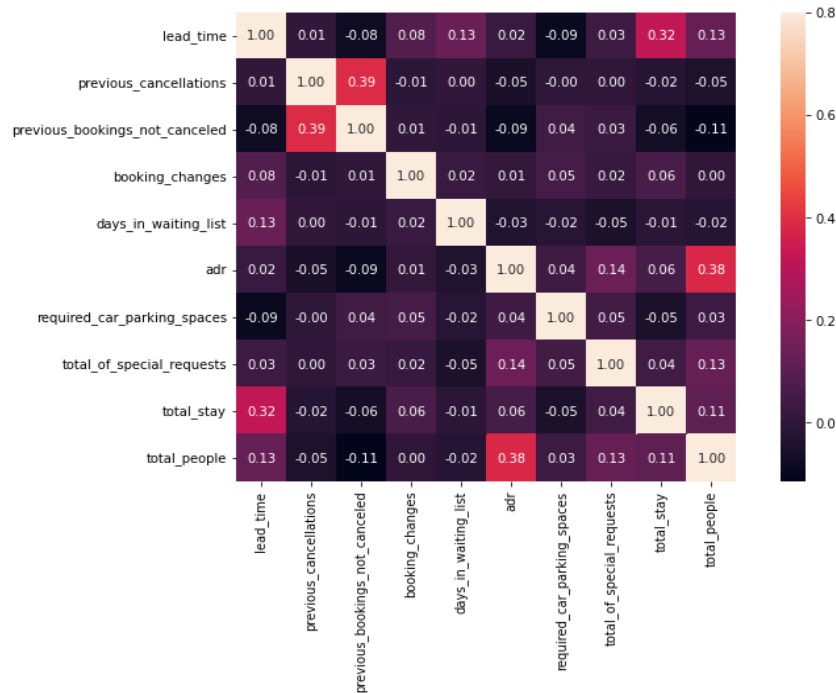


Correlation Heatmap

1) Total stay length and lead time have slight correlation. This may mean that for longer hotel stays people generally plan little before the actual arrival.

2) adr is slightly correlated with total_people, which makes sense as more no. of people means more revenue, therefore more adr.

Lets see does length of stay affects the adr.



Analysing the data using EDA

Data Preparation and

Cleaning: Importing the libraries needed in this project for data manipulation, cleaning and visualization. Example – Column agents and column companies has many missing values making the columns irrelevant for our analysis, hence removing them. Finding all Null and Duplicate values and dropping them.

Handling Outliers: By plotting boxplot, we see some outliers. We have used median for removing outliers. Seaborn boxplot are used for visualizing the outliers in the present data.

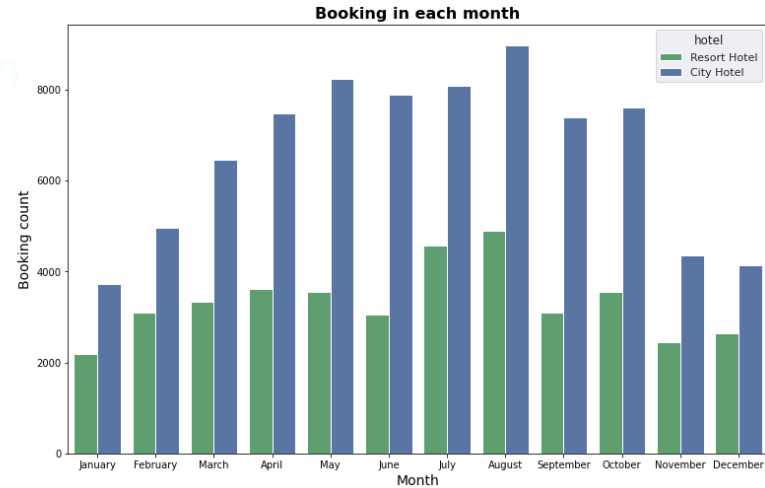
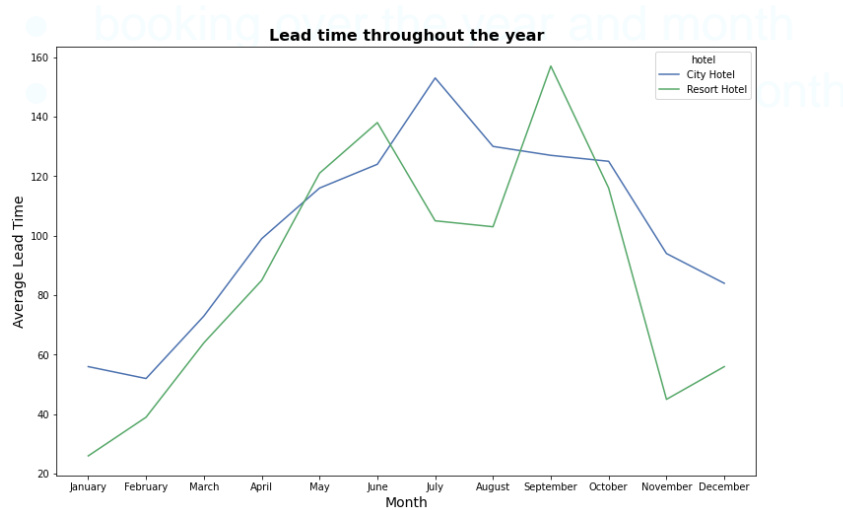


Room Booking by the Agent:



- 9.0 No. agent has maximum booking and 8.0 has minimum booking

Maximum booking over the year and month:

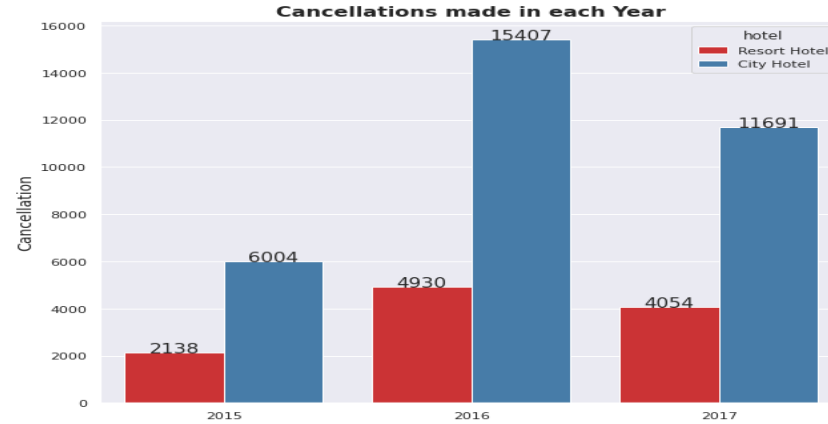


- From the above graph we can see in city hotels there is a peak in lead time from April to July and the booking is high in August. And in resort hotels we can see two peaks in lead time, first in June and second in September and booking is high in July, August and October, so people usually book hotel 30-60 days in advance.

Analysis based on Cancellation:

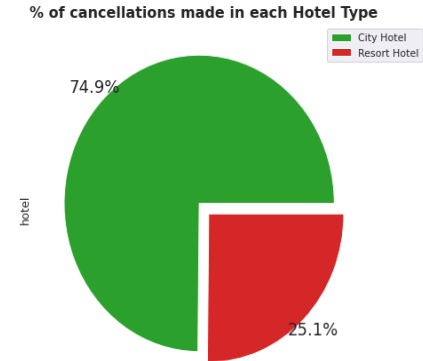
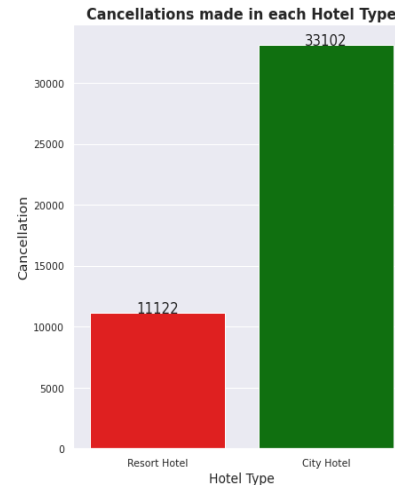
Cancellations made in each Year:

We can see maximum cancellations were done in the year 2016 in both the hotel types



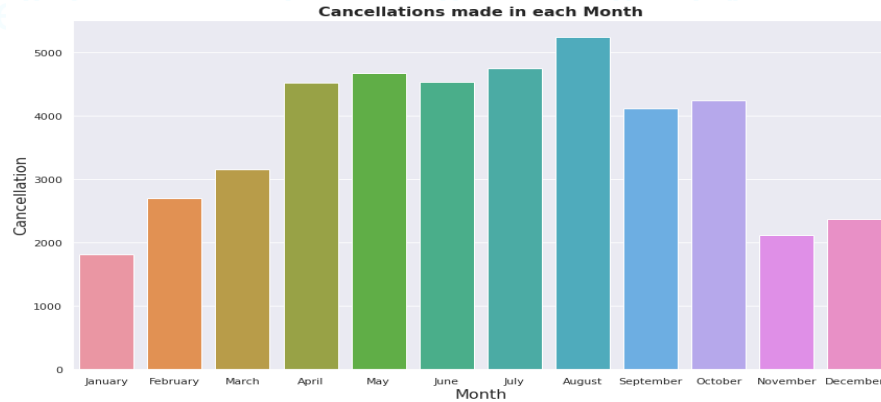
Cancellation made in each hotel type:

- Cancellation is less in Resort hotels as compared to City hotels.
- Every year there is 25-30% cancellation for resort hotels and 40-45% cancellation for city hotels.



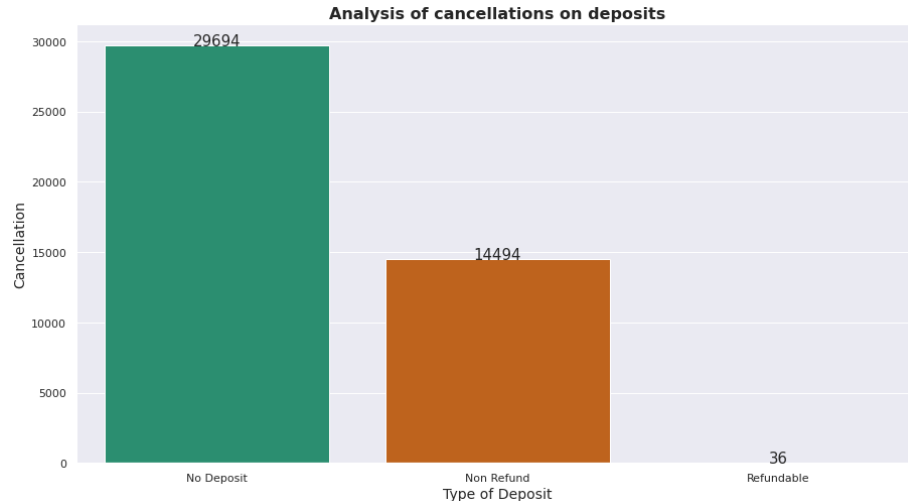
Cancellations made in each months:

- We can see a huge number of cancellations are done in the month of August.

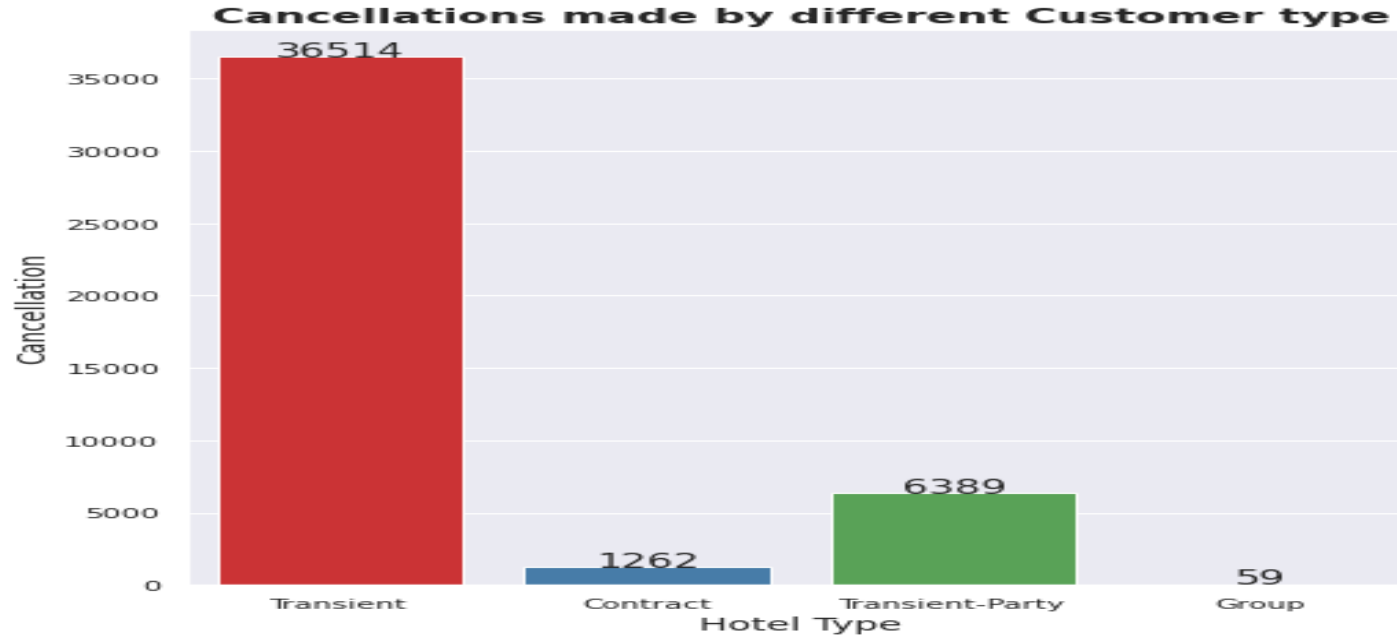


Analysis of cancellations on deposits:

- Chances of cancellation is high when there is no deposits taken by hotels. So minimum deposits should be taken by hotels to decrease the rate of cancellation

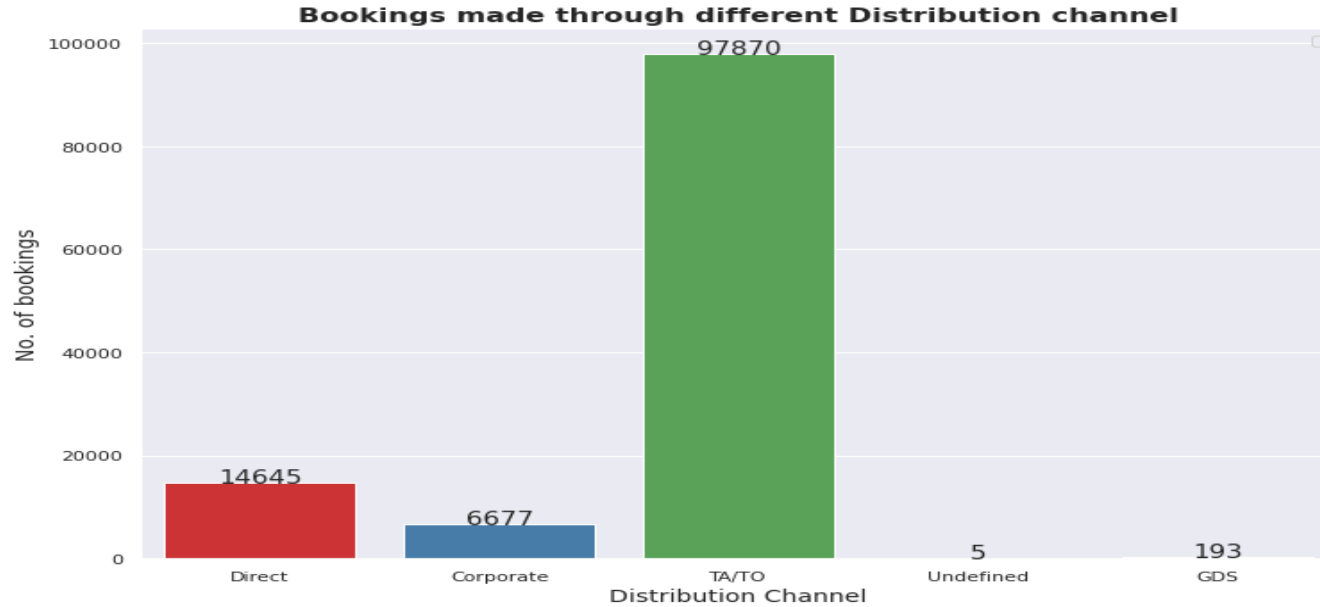


Cancellation based on customer type:



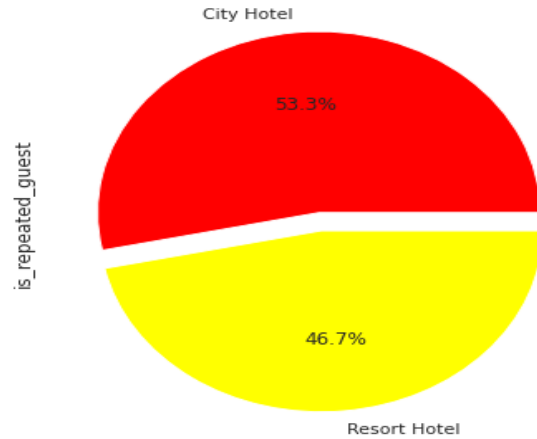
Transient customers cancels more often but when people book in group it leads to lesser cancellations.

Which significant distribution channel has highest cancellation percentage Preferred Room type:



- TA/TO has highest booking cancellation %. Therefore, a booking via TA/TO is 30% likely to get cancelled.

Percentage of Repeated Guests in hotels:



City hotel has more number of repeated guest then Resort hotel

Observations

- 1 Since the bookings are double for City hotel but revenue difference is less so Resort hotels are more expensive as compared to City hotels.
- 2 Best time to book a resort hotel is from October to April and best time to book a City hotel is from November to March as we are getting lower ADR in this period.
- 3 Maximum number of bookings and cancellations are done in the month of August.
- 4 Both hotel types are getting higher revenue between June to September as ADR is also high in this period.
- 5 Hotel bookings and cancellations are highest in year 2016.
- 6 Every year there is 25-30% cancellation for resort hotels and 40-45% cancellation for city hotels.
- 7 As lead time increases ADR decreases. This means if a customer book a hotel in advance, he can get hotels at lower price.
- 8 As length of total stay increases ADR decreases. This means for longer stays, customers can get a better deal.
- 9 People prefer city hotels for shorter stay and resort hotels for longer stay
- 10 Hotels having refundable policy are expensive whereas hotel with no-refund policy are cheaper.
- 11 Most bookings are done by transient customer types. Transient customers cancels more often but when people book in group it leads to lesser cancellations.
- 12 Majority of the bookings and cancellations are made through Travel agencies (Online/Offline) and Tour Operators.
- 13 Cancellation is more in City hotels as compared to Resort hotels.
- 14 Chances of cancellation is high when there are no deposits taken by hotels.
- 15 77% of the people prefer the BB (bed & breakfast) meal type in both the hotel types.
- 16 Maximum bookings and revenue are generated from country Portugal.

Conclusions

Majority of people prefer A-room type so hotels should increase their numbers to get more revenue.

Chances of cancellation is high when there are no deposits taken by hotels, so hotels should take minimum deposits to minimize the rate of cancellation.

Transient customers cancels more often but when people book in group it leads to lesser cancellations, hence hotels should provide some offers to decrease cancellations.

Maximum number of bookings are in the month May to August, so hotels should provide exciting offers to customers to increase their booking in off season.

As hotels are getting less repeated customers so management should take customer's feedback and improve the hotel facilities to increase count of their repeated guests.

We have a huge number of visitors from western Europe namely Portugal, France, UK and same countries generate highest revenue. Marketing teams should target these countries to get

THANK YOU