

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

1. Chinmay Rojatka: chinmayrojatkar4@gmail.com
Time Wise Analysis
 - Which Months are the most busiest months for the hotels?
 - In which year the Highest Number of Bookings were Done?
 - What are the busiest days of the months in staying the Hotels?
 - Trends of booking Changes and Months of Arrivals?
2. Kunal Gawande: gkunal8019@gmail.com
Data Cleaning – Imputing null values.
Cancellation Analysis
 - Which distribution channel has highest the cancellation percentage?
 - Reasons behind the Cancellation Occurs?
 - Which channels is affecting the most revenue of the Hotels?
3. Deepali Mahajan: deepali2062@gmail.com
Bivariate Analysis
 - Which room type is in the demand?
 - Which room type generated highest Revenue?
 - In Which Year the Highest Number of Rooms booked by the each type of the Hotel?
 - Maximum Number of the Rooms booked by ?
 - In which Month the Highest Number of Customers came?
4. Bipasha Zade: bipashazade527@gmail.com
Distribution Analysis
 - What is the most common channel for making the booking of the hotels?
 - Which distribution channel generates more revenue deals for the hotels?
 - Which channel does have more repeated Customers?
5. Nikhil Aggarwal: nickagg30899@gmail.com
Hotel Wise Analysis
 - Percentage of the bookings in each hotels?
 - Which hotel makes more revenue?
 - Which hotel has higher booking cancellations rates?
 - What is most preferred stay length in each hotel?
 - What is the preferred stay time by the Customers?

Please paste the GitHub Repo link.

Kunal Gawande Link:- <https://github.com/gkunal8019>

Chinmay Rojarka Link:- <https://github.com/ChinmayRojatkar>

Bipasha Zade Link:- <https://github.com/Bipashazade>

Nikhil Aggarwal Link:- <https://github.com/Nikhil8815>

Deepali Mahajan Link:- <https://github.com/deepali2062>

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

Through this analysis, we were able to answer key business questions which are stated below:

Is there a difference in volume of bookings and revenue across the two type of hotels?

Is the change driven by a time period such as a year or a month?

Is the cancellation rate higher than the booking rate?

Are the cancellations driven by any factors such as time period or hotel type?

What is the average revenue generated per booking?

Is the revenue and booking driven by specific customer type such as New Customers or Existing Customers?

Are there more new customers coming in as years progress? How is the growth rate?

What type of meal is preferred?

What type of rooms are preferred?

Are the bookings driven by the week of day?

For a new booking, based on certain attributes, can we predict if the booking will go through or get cancelled?

What is the revenue expected in the coming 12 months? When do I need to run promotion?

Methodology

Majority of the analysis was EDA which was digging one level deeper and getting the data to answer the above questions

To answer few questions I sliced the data across various cuts

Another methodology included using different lenses to view data across segments

I used classification tree to predict if the booking will go through

For estimating the revenue over the next 12 months, I used time-series modelling

Insights

The number of bookings seems to be high in 2016 while the bookings seem to be less in 2015 and 2017. This is majorly due to more data points being available in 2016

The ratio of bookings cancelled to the confirmed bookings seems high for City Hotels

We observe that the avg. # of bookings in a week is growing year over year

The growth is higher from 2015 to 2016 as compared to the jump from 2016 to 2017

The growth is higher from 2015 to 2016 as compared to the jump from 2016 to 2017

The revenue almost doubled in 2016 but it did not grow at the same rate from 2016 to 2017

We observe that the number of New Customers and Existing Customers increases drastically in 2016 and the rise is not that high in 2017

We observe that the New Customers tend to spend significantly lower than the Existing Customers

Majority of bookings are accompanied with a breakfast or dinner

There are few Full Board bookings. This gives an opportunity to the hotels to increase revenue from these type of meals by offering some discounts

The hotels need to have more of Type A room followed by other types

We can investigate the Type L rooms to conclude about the 4 bookings in these rooms

We observe that Weekday bookings are higher than the Weekend numbers. That is an interesting finding.

We need to investigate our corporate bookings!

Implication to business

The data for 2015 and 2017 is for different months. Even though we have converted them to same base line of weekly numbers, there are chances that some weeks perform differently as compared to other weeks

The definition of new customers is not very well described. A new customer this year will be existing next year, or they can be existing customer from the 2nd booking. A deeper analysis is required based on definition

The weekday vs weekend analysis can be further drilled for the type of bookings

The classification model uses few variables. We can tune the model with new variables and adjusting the cost of misclassification. Additionally, we need to split the data into train and validation to avoid overfitting

The forecasting aspect can be further drilled to analyze the residuals and split the model across years and various factors such as customer type or hotel type