Technical document for EDA project

Hotel booking Analysis

Prepared By

Nitin Solanki

(solankinitin1210@gmail.com)

1. Purpose of Document
   - The purpose of this document is to understand the technical details of the analysis and explain each step followed to get the conclusion

2. Our Approach

   We cleaned and preprocessed the data, categorized each column and then we performed the exploratory data analysis important column-wise to extract information from the data

Steps performed in this EDA Projects

Handling this dataset with the fundamental steps to unveil the factors:

- Imports Libraries, Dataset, and Other modules
- Analyzing Data: Reading & Inspection of Data.
  - Data Structure (Head, Tail or Body)
  - Data information (basic schema)
  - Arithmetic measurement (data descriptions)
- Data Preparation & Processing
  - Identifying Different Variable
  - Shortlisting Dependent Variables
  - Handling Null Values
  - Adding, combining &renaming of columns if required.
- Analysis & Visualization
  - Identifying relation of dependent variables with Problem set.
  - Visualization using different plots in relation to a question.
- Conclusion: Answer to the Questions

3. Analyze each column in the given data

1st of all we checked the relevancy of the columns with our problem which is "To discover the important factors that govern the Bookings". then defined the Possibility of the relevance of each Column one by one for that we use the spreadsheet and define the following Column for each variable

- Type of Variable- Identified if it is continuous or Categorical by Continuous we mean those non Repeating values mainly in integer or number format. Examples are leadtime, adults, adr etc. Categorical - Variables that can be subcategorized like month variable can be subcategorized in 12 different months.

- Expectation - Relevancy of the variable with the Problem set which is important factor that govern the bookings well we define the expectation as high (Directly relevant to the problem we explore), Medium (may affect the analysis will check the relevancy by some visualization and explore if required), Low (which is irrelevant or may have mild effect will remove the column for our analysis after confirmation)

- Analysis to be done - Identified Analysis with the different variables

- Conclusion - Outcomes of the analysis

Well after brainstorming we conclude and we have divided all the 32 columns/variables into 2 categories **Continuous** and **Categorical**

Continuous variables are those non Repeating values mainly in integer or number format. We put the following variable in this list

Categorical variables that can be subcategorized like month variable can be subcategorized in 12 different months. We put the following variable in this list Here is the table which shows the category and expectation of all the variable

Variable Name	Type	Expectation
hotel	categorical	High
is_canceled	categorical	Low
lead_time	Continuous	Medium
arrival_date_year	categorical	Medium
arrival_date_month	categorical	High
arrival_date_week_number	Continuous	Medium
arrival_date_day_of_month	Continuous	Medium
stays_in_weekend_nights	Continuous	High
stays_in_week_nights	Continuous	Medium
adults	Continuous	Medium
children	categorical	Medium
babies	categorical	Medium
meal	categorical	Low
country	categorical	High
market_segment	categorical	High
distribution_channel	categorical	High
is_repeated_guest	categorical	High
previous_cancellations	Continuous	Low
previous_bookings_not_canceled	Continuous	Low
reserved_room_type	categorical	Low
assigned_room_type	categorical	Low
booking_changes	Continuous	Low
deposit_type	categorical	Low
agent	categorical	High
company	categorical	High
days_in_waiting_list	Continuous	Low
customer_type	categorical	Medium
adr	Continuous	Medium
required_car_parking_spaces	categorical	Low
total_of_special_requests	categorical	Low
reservation_status	categorical	Low
reservation_status_date	Continuous	Low

4. Technicalities Details of each analysis

- For Every analysis, we used variables mostly with high and medium expectation
- We took one by one variable and apply the group by function to get a table that we can use for plotting
- We use the plot function to plot the chart of resulted table

Hotel type-wise Analysis:

- **Column Used for analysis**: hotel
- **Function Used:** value_count ()
- **Columns of Resulted Table:** Hotel(type), No. of booking
- **Chart Type**: Pie

Market segment-wise Analysis:

- **Column Used:** market_sagment, hotel
- **Function Used:** groupby (), rename ()
- **Columns of Resulted Table:** distribution_channel, No of booking, Percentage
- **Chart Type:** Bar
- **Plotted Values:** X=Market Segment, Y=percentage

Distributed channel-wise Analysis:

- **Column Used:** month-year, hotel, meal (meal is used for counting total no.)
- **Function Used:** groupby (), unstack ()
- **Columns of Resulted Table:** Month-year, City hotel (no. of booking), resort hotel (No. of booking)
- **Chart Type:** Line
- **Plotted Values:** X= (Month-year, City hotel, resort hotel), Y=No. of bookings

Distributed channel-wise Analysis:

- **Column Used:** distributed_channel, hotel
- **Function Used:** groupby (), rename ()
- **Columns of Resulted Table:** distribution_channel, No of booking, Percentage

- **Chart Type:** Bar
- **Plotted Values:** X=Distributed channel=percentage

Month wise Analysis:

- **Column Used:** month-year, hotel, meal (meal is used for counting total no.)
- **Function Used:** groupby (), unstack ()
- **Columns of Resulted Table:** Month-year, City hotel (no. of booking), resort hotel (No. of booking)
- **Chart Type:** Line
- **Plotted Values:** X= (Month-year, City hotel, resort hotel), Y=No. of bookings

Country-wise Analysis:

- **Column Used:** country, hotel
- **Function Used:** groupby (), rename ()
- **Columns of Resulted Table:** country, No of booking, Percentage
- **Chart Type:** Bar
- **Plotted Values:** X=country=percentage

Accommodation type-wise analysis:

- **Column Used:** total_guest, hotel
- **Function Used:** groupby (), rename ()
- **Columns of Resulted Table: Total_guest, No of booking, Percentage**
- **Chart Type:** Bar
- **Plotted Values:** X= **Total_guest**, Y=percentage

Customer type-wise analysis:

- **Column Used:** customer_type, hotel
- **Function Used:** groupby (),
- **Columns of Resulted Table:** customer_type, City_hotel (Total no. of booking), resort_hotel (Total no. of booking)
- **Chart Type:** Bar
- **Plotted Values:** X= (Customer type, City hotel, resort hotel), Y=No. of bookings

Cancellation analysis

- **Column Used:** Is_cancelled, hotel
- **The table used from other analysis:** hotel_type_wise_analysis

- **Function Used:** groupby (), merge () (Inner joins)
- **Columns of Resulted Table:** hotel(type), No. of cancellation, No. of booking, percentage
- **Chart Type:** Bar
- **Plotted Values:** X= hotel(type), Y= Percentage