EDA CAPSTONE PROJECT

Topic

HOTEL BOOKING TRENDS

ANALYSIS

Technical Document

Presented By :-

Souvik Karmakar

1.Abstract:-

This project investigates a rich hotel booking dataset, extracting actionable insights to enhance revenue and customer satisfaction. Focusing on factors like lead time, guest composition, meal preferences, origin country, and booking details, the analysis aims to uncover trends that empower strategic decision-making.

Lead time emerges as a crucial factor, guiding hotels in adjusting pricing strategies to incentivize early bookings or capture last-minute reservations. Understanding guest composition enables tailored room and meal offerings, while meal type influences daily rates, aligning with market preferences.

Exploring distribution channels and deposit types reveals strategic opportunities, such as offering preferential rates for direct bookings or non-refundable deposits. Additionally, insights into reservation status and customer types inform loyalty programs and optimize room inventory.

In summary, this analysis transforms raw data into actionable insights, guiding hotels to refine pricing, tailor offerings, and optimize marketing, ultimately unlocking new revenue streams and elevating customer satisfaction. This project serves as a blueprint for hotels navigating the competitive landscape of the hospitality industry.

2.Introduction:-

Within the hotel sector, demand forecasting and customer demands prediction are the main areas of study attention.

There are two different kinds of hotels in this dataset. The city and the resort hotel

lodging. It is gathered to forecast hotel reservations and the likelihood that they will be canceled.

Understanding the elements that contribute to the business's revenue is one of these qualities. Certain variables reflect the preferences of the clients while making reservations, while other features reflect the causes of cancellations.

3.Project Goal:-

The main aim of looking at the hotel booking data is to better understand how customers behave. We want to find out trends and patterns in the bookings that can help us make smarter decisions based on data.

By doing this, we hope to improve how well the hotel is doing and make more money. We'll be looking at things like how far in advance people book, how many people are staying, what kind of meals they prefer, where they're from, and what types of bookings they make. We'll also check patterns like how people pay, the status of their reservation, and what kind of customers they are.

The insights we get from all of this will help us set better prices, offer things people want, and improve how we reach customers. Ultimately, our goal is to make more money and make sure customers are happy with their stay at the hotel.

4.Attributes:-

Here is the detailed description about all columns:-

❖ hotel: Indicate different type of hotel present
  ● City hotel
  ● Resort type
❖ is_canceled: Indicates the cancellation of the hotel booking
  ● Cancellation = 1
  ● No Cancellation = 0
❖ lead_time: Time (in days) between booking transaction and actual arrival.
❖ arrival_date_year: Year of arrival.
❖ arrival_date_month: Month of arrival
❖ arrival_date_week_number: week number of arrival date.
❖ arrival_date_day_of_month: Day of month of arrival date
❖ stays_in_weekend_nights: No. of weekend nights stayed in a hotel
❖ stays_in_week_nights: No. of weeknights stayed in a hotel
❖ adults: No. of adults in a single booking record.
❖ children: No. of children in a single booking record.
❖ babies: No. of babies in single booking record.
❖ meal: Type of meal chosen
  ● BB:- bed and breakfast
  ● HB:-Half board (Breakfast and dinner)
  ● FB:- Full Board (All meals included)
  ● SC:- Self catering (No meals Included)
❖ country: Country of origin.
❖ market_segment: market segment for booking
  ● Aviation
  ● Complimentary
  ● Corporate
  ● Direct
  ● Groups
  ● Online (TA)
  ● Offline (TA/TO)

- ❖ distribution_channel: Via which medium booking
    - Corporate
    - Direct
    - GDS: - Global Distribution System
    - TA/TO: - Travel Agent/Operator
- ❖ is_repeated_guest:
    - 0 for new customer
    - 1 for repeated customer
- ❖ previous_cancellations: No. of previous cancelled bookings.
- ❖ previous_bookings_not_canceled: No. of previous non-cancelled bookings.
- ❖ reserved_room_type: Room type reserved by a customer.
- ❖ assigned_room_type: Room type assigned to the customer.
- ❖ booking_changes: No. of booking changes done by customers
- ❖ deposit_type: Type of deposit at the time of making a booking
    - No deposit
    - Refundable
    - No refund
- ❖ agent: Id of agent for booking
- ❖ company: Id of the company making a booking
- ❖ days_in_waiting_list: No. of days in waiting to book
- ❖ customer_type: Type of customer
    - Contract: - bookings done by the contract
    - Group: - Group booking
    - Transient: - Customer staying for shorter period
    - Transient-Party: - Group of customers staying for a shorter period
- ❖ adr: Average Daily rate of hotels.
- ❖ required_car_parking_spaces: No. of car parking preferred by customers at   the time of booking.
- ❖ total_of_special_requests: total no. of special request.
- ❖ reservation_status:
    - checked out
    - canceled
    - not showed
- ❖ reservation_status_date: Date of making reservation status.

5.Exploratory Data Analysis:-

The process of examining and analyzing record sets in order to recognize patterns, find outliers, and determine the relationships between variables is known as exploratory data analysis, or EDA.

Importing libraries like pandas, matplotlib, seaborn, and NumPy is the first step. Import the raw dataset next. There are numerous raw numbers in this data that are not suitable for the

investigation. We have followed some process to rectify our unprocessed data before visualization.

a) Data Cleaning: -

Cleaning data is crucial step before EDA as it will remove the ambiguous data that can affect the outcome of EDA.

Steps used for data Cleaning:

Remove duplicate rows:- We have removed all duplicate rows by using drop() method also we have dropped company column as it contains more than 90 % missing values also we have dropped babies , children, and adults columns .

Handling missing values: - We have imputed the missing values with their mean for agent column.

b) **Data Manipulation**: -
   Creating new columns
   - Total_Guest= children +babies+adults
   - **Total_Duration**= stays_in_weekend_nights+ stays_in_week_nights
   - **Revenue**= Total_Duration * ADR

6.Data Study:-

i) UNIVARIATE ANALYSIS:-

The most basic type of data analysis, or the examination of a single variable, is univariate analysis. Finding patterns in the data and describing the distribution of individual data are its main goals.

ii) BIVARIATE ANALYSIS:-

Analysis of bivariates using two variables. There will be one dependent variable and one independent variable. To determine the magnitude of the change, the study is evaluated between the two variables.

iii) MULTIVARIATE ANALYSIS: -

Multivariate data analysis involves grouping the gathered samples into homogeneous groups, analyzing the correlations between the qualities, and extrapolating conclusions about the underlying populations from the sample.

7.Data Visualization: -

Data visualization is a graphical representation of quantitative information and data by using visual elements like graphs, charts, and maps. Data visualization convert large and small data sets into visuals, which is easy to understand and process for humans. Data visualization tools provide accessible ways to understand outliers, patterns, and trends in the data. We have used the below graphs or plots for our analysis :-

- Box Plot.
- Histogram.
- Pie Chart.
- Bar Plot.
- Line Plot.
- Scatter Plot.
- Pair Plot
- Heat Map

8.Conclusion: -

Now after this in-depth analysis of the hotel booking data we can conclude below mentioned points which will help business to take meaningful decision, so that they can grow more :-

1. The most popular type of hotel among visitors is a city hotel.
2. Despite having nearly twice as many reservations as a resort hotel, the city hotel makes less money overall.
3. We can observe that reservations for city hotels peak in August and peak from April to July. Additionally, there are two peak months for resort hotels: June and September. July, August, and October see the highest booking volumes, therefore reservations are typically made 30 to 60 days in advance.
4. From June to September, revenue at both resort and city hotels increased. This is also due to the fact that, as the previous slide illustrates, ADR is high for both kinds of hotels at the same time. Therefore this period is best for hotels to generate more revenue.
5. Here, we can observe that the ADR falls with increasing lead time. This implies that a customer can obtain a better rate if he books a hotel in advance.
6. The majority of reservations are made by sporadic clientele.
7. The majority of reservations are made by transient clientele.
8. The majority of reservations and cancellations are handled by tour operators and online and offline travel agents.
9. Compared to resort hotels, there are more cancellations at city hotels.
10. When hotels don't collect deposits, there is a greater likelihood of cancellation. Therefore, hotels ought to require minimum deposits in order to lower the cancellation rate.

11.ADR falls as overall stay duration rises. This implies that a better offer for the clients might be arranged for a longer stay.

the comprehensive exploratory data analysis (EDA) conducted on the hotel booking trend dataset has unveiled key patterns and insights crucial for strategic decision-making in the hospitality industry. The data strongly suggests that minimizing cancellations can be achieved by implementing a policy of requiring minimum deposits, especially given the high cancellation rates observed when no deposits are taken.

Furthermore, understanding customer segmentation is vital, as transient customers exhibit higher cancellation rates. Offering enticing group booking incentives can significantly reduce cancellations, promoting customer loyalty and satisfaction.

Seasonal trends indicate peak booking months from May to August. Capitalizing on this insight, hotels can strategically introduce special offers and promotions during off-peak seasons to attract more bookings and maximize revenue.

The analysis also underscores the significance of targeting specific geographical regions, with Western European countries—Portugal, France, and the UK—emerging as major contributors to bookings and revenue. A targeted marketing approach towards these regions is advised to further amplify customer acquisition.

In essence, this EDA serves as a robust foundation for strategic decision-making, offering actionable insights that, when implemented, can drive positive business outcomes. The client is well-positioned to leverage these findings for increased revenue, improved customer satisfaction, and operational excellence. Continuous monitoring and adaptation to changing market conditions will be key to sustaining and enhancing the success achieved through these recommendations. The insights derived from this analysis provide a roadmap for the hotel industry to navigate evolving trends and stay ahead in a competitive landscape.