Hotel Harmony: Data Insights for Optimized Operations

Background:

Company: Elite Hotels International

Scenario: Elite Hotels International, a prominent player in the hospitality industry, has been facing challenges in managing its booking patterns and customer satisfaction. With hotels spread across various locations, understanding booking trends and guest behavior is crucial for optimizing operations and enhancing guest experience. The management has decided to conduct a comprehensive data analysis of their booking data to uncover insights that can drive strategic decisions.

Understanding of the Project:
https://www.sciencedirect.com/science/article/pii/S2352340918315191

Dataset:

https://www.kaggle.com/jessemostipak/hotel-booking-demand

Data Dictionary

1. **hotel:** Name of the hotel (Resort Hotel or City Hotel).
2. **is_canceled:** Indicates if the booking was canceled (1) or not (0).
3. **lead_time:** Number of days between booking date and arrival date.
4. **arrival_date_year:** Year of arrival date.
5. **arrival_date_month:** Month of arrival date.
6. **arrival_date_week_number:** Week number of year for arrival date.
7. **arrival_date_day_of_month:** Day of arrival date.
8. **stays_in_weekend_nights:** Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel.
9. **stays_in_week_nights:** Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel.
10. **adults:** Number of adults.
11. **children:** Number of children.
12. **babies:** Number of babies.
13. **meal:** Type of meal booked.
14. **country:** Country of origin.
15. **market_segment:** Market segment designation.
16. **distribution_channel:** Booking distribution channel.
17. **is_repeated_guest:** Indicates if the booking was from a repeated guest (1) or not (0).
18. **previous_cancellations:** Number of previous bookings that were canceled by the customer prior to the current booking.
19. **previous_bookings_not_canceled:** Number of previous bookings not canceled by the customer prior to the current booking.
20. **reserved_room_type:** Code of room type reserved.
21. **assigned_room_type:** Code for the type of room assigned to the booking.

22. **booking_changes:** Number of changes made to the booking from the moment the booking was entered to the moment of check-in or cancellation.
23. **deposit_type:** Type of deposit made for the booking.
24. **agent:** ID of the travel agency that made the booking.
25. **company:** ID of the company/entity that made the booking.
26. **days_in_waiting_list:** Number of days the booking was in the waiting list before it was confirmed to the customer.
27. **customer_type:** Type of customer (e.g., transient, contract, group, etc.).
28. **adr:** Average Daily Rate, as defined by dividing the sum of all lodging transactions by the total number of staying nights.
29. **required_car_parking_spaces:** Number of car parking spaces required by the customer.
30. **total_of_special_requests:** Number of special requests made by the customer (e.g., twin bed or high floor).
31. **reservation_status:** Reservation last status (e.g., Canceled, Check-Out, No-Show).
32. **reservation_status_date:** Date at which the last status was set.

Stakeholder Involvement

Internal Stakeholders:

1. **Management Team:** Responsible for strategic decision-making.
2. **Operations Team:** Handles day-to-day hotel operations.
3. **Marketing Team:** Plans and executes marketing campaigns.
4. **Customer Service Team:** Manages guest relations and satisfaction.

External Stakeholders:

1. **Guests:** Provide feedback and influence hotel reputation.
2. **Travel Agencies:** Facilitate bookings and provide a significant portion of hotel clientele.
3. **Suppliers:** Provide goods and services essential for hotel operations.

Problem Definition

Problem: Elite Hotels International needs to optimize its booking patterns and enhance customer satisfaction to maintain a competitive edge in the hospitality industry.

Importance: Solving this problem will help in:

- Improving operational efficiency.
- Enhancing guest satisfaction and loyalty.
- Increasing revenue through better resource allocation and marketing strategies.

Data Requirements

- **Booking Information:** Includes lead time, arrival date, stay duration, etc.
- **Guest Demographics:** Information such as country of origin, number of adults, children, and babies.
- **Booking Channels:** Includes market segment, distribution channel, and agent details.
- **Booking Modifications:** Number of changes, special requests, and deposit type.

● **Financial Data:** Average Daily Rate (ADR) and revenue metrics.

Metric Development

● **Booking Lead Time:** Average number of days guests book in advance.
● **Cancellation Rate:** Percentage of bookings that are canceled.
● **Occupancy Rate:** Average number of nights stayed by guests.
● **Special Requests Frequency:** Average number of special requests per booking.
● **Revenue Metrics:** ADR, total revenue generated, etc.

Insights & Actions

● **Booking Trends:** Identify peak booking periods and adjust marketing strategies accordingly.
● **Cancellation Insights:** Develop policies to reduce cancellations and manage overbooking.
● **Guest Preferences:** Tailor services and amenities based on guest demographics and preferences.
● **Revenue Management:** Optimize pricing strategies based on booking patterns and ADR.

Communication

Final Conclusion:

● Summarize key findings from the data analysis.
● Provide actionable recommendations for management.
● Highlight potential areas for further investigation.

Data Preprocessing and EDA Steps

1. **Data Cleaning:**
   ○ Handle missing values in columns such as `agent`, `company`, and `children`.
   ○ Convert date columns to appropriate datetime formats.
   ○ Check for and handle duplicate entries.
2. **Exploratory Data Analysis (EDA):**
   ○ Analyze the distribution of bookings by hotel type, month, and market segment.
   ○ Investigate the correlation between lead time and cancellation rate.
   ○ Visualize the trend of average daily rate (ADR) over time.
   ○ Examine the frequency of special requests and its impact on guest satisfaction.

Basic Level Questions with Solution Ideas

1. **What is the average lead time for bookings?**
   ○ **Hint:** Calculate the mean lead time using the 'lead_time' column.
   ○ **Solution Idea:** Calculate the average number of days between the booking date and the arrival date.

- ○ **How it helps:** Identifies how early customers tend to book.
- ○ **Business Impact:** Aids in predicting booking trends and planning marketing campaigns accordingly.
2. **What is the distribution of bookings by hotel type?**
    - ○ **Hint:** Count the number of bookings for each hotel type and visualize it.
    - ○ **Solution Idea:** Create a bar chart showing the number of bookings for Resort Hotels and City Hotels.
    - ○ **How it helps:** Shows the popularity of each hotel type.
    - ○ **Business Impact:** Helps in targeting marketing efforts appropriately.
3. **How many bookings were canceled?**
    - ○ **Hint:** Count the number of canceled bookings using the 'is_canceled' column.
    - ○ **Solution Idea:** Sum the values in the 'is_canceled' column to get the total number of cancellations.
    - ○ **How it helps:** Provides insight into booking reliability.
    - ○ **Business Impact:** Helps in developing strategies to reduce cancellations.
4. **What is the most common arrival month for bookings?**
    - ○ **Hint:** Identify the month with the highest number of arrivals using the 'arrival_date_month' column.
    - ○ **Solution Idea:** Determine the mode of the 'arrival_date_month' column to find the most frequent month of arrival.
    - ○ **How it helps:** Identifies peak booking periods.
    - ○ **Business Impact:** Aids in resource and staff planning for peak seasons.
5. **What is the average number of special requests per booking?**
    - ○ **Hint:** Calculate the mean of the 'total_of_special_requests' column.
    - ○ **Solution Idea:** Compute the average number of special requests made by guests.
    - ○ **How it helps:** Understands customer service requirements.
    - ○ **Business Impact:** Improves customer satisfaction by meeting their requests.
6. **Which country has the highest number of bookings?**
    - ○ **Hint:** Identify the country with the highest count of bookings using the 'country' column.
    - ○ **Solution Idea:** Find the country with the most entries in the 'country' column.
    - ○ **How it helps:** Identifies major customer demographics.
    - ○ **Business Impact:** Focuses marketing efforts on key regions.
7. **What is the average daily rate (ADR) for each hotel type?**
    - ○ **Hint:** Calculate the mean ADR for each hotel type using the 'adr' column.
    - ○ **Solution Idea:** Group the data by 'hotel' and calculate the average ADR for each group.
    - ○ **How it helps:** Compares financial performance of different hotel types.
    - ○ **Business Impact:** Informs pricing and revenue management strategies.
8. **What percentage of guests required car parking spaces?**
    - ○ **Hint:** Calculate the percentage of bookings that required car parking spaces using the 'required_car_parking_spaces' column.
    - ○ **Solution Idea:** Determine the proportion of bookings with 'required_car_parking_spaces' greater than 0.
    - ○ **How it helps:** Assesses the demand for parking facilities.
    - ○ **Business Impact:** Ensures adequate parking availability for guests.
9. **What is the average stay duration in week nights and weekend nights?**
    - ○ **Hint:** Calculate the mean stay duration using the 'stays_in_week_nights' and 'stays_in_weekend_nights' columns.
    - ○ **Solution Idea:** Compute the average number of week nights and weekend nights stayed by guests.
    - ○ **How it helps:** Understands booking patterns.

- ○ **Business Impact:** Helps in optimizing room occupancy and pricing strategies.
10. **How many bookings were made through travel agents?**
    - ○ **Hint:** Count the number of bookings made through travel agents using the 'agent' column.
    - ○ **Solution Idea:** Count the non-null values in the 'agent' column.
    - ○ **How it helps:** Identifies the importance of travel agents in booking processes.
    - ○ **Business Impact:** Helps in managing relationships with travel agents effectively.

Medium Level Questions with Solution Ideas

1. **What is the cancellation rate for each hotel type?**
   - ○ **Hint:** Calculate the mean of the 'is_canceled' column for each hotel type.
   - ○ **Solution Idea:** Group the data by 'hotel' and calculate the average cancellation rate for each group.
   - ○ **How it helps:** Identifies which hotel type has a higher cancellation rate.
   - ○ **Business Impact:** Helps in developing strategies to reduce cancellations for specific hotel types.
2. **What is the average ADR per market segment?**
   - ○ **Hint:** Calculate the mean ADR for each market segment using the 'adr' column.
   - ○ **Solution Idea:** Group the data by 'market_segment' and calculate the average ADR for each segment.
   - ○ **How it helps:** Identifies revenue performance across different market segments.
   - ○ **Business Impact:** Helps in tailoring pricing strategies to different market segments.
3. **What is the relationship between lead time and cancellation rate?**
   - ○ **Hint:** Visualize the relationship using a scatter plot of 'lead_time' vs. 'is_canceled'.
   - ○ **Solution Idea:** Use a scatter plot to explore the correlation between lead time and cancellation likelihood.
   - ○ **How it helps:** Helps understand if bookings made with longer lead times are more likely to be canceled.
   - ○ **Business Impact:** Aids in developing strategies to manage long lead-time bookings.
4. **Which distribution channel has the highest number of bookings?**
   - ○ **Hint:** Count the number of bookings for each distribution channel.
   - ○ **Solution Idea:** Identify the distribution channel with the highest booking count.
   - ○ **How it helps:** Identifies the most effective booking channels.
   - ○ **Business Impact:** Helps in focusing efforts on the most productive channels.
5. **What is the average number of previous cancellations by hotel type?**
   - ○ **Hint:** Calculate the mean of the 'previous_cancellations' column for each hotel type.
   - ○ **Solution Idea:** Group the data by 'hotel' and calculate the average number of previous cancellations for each group.
   - ○ **How it helps:** Understands past cancellation behavior by hotel type.
   - ○ **Business Impact:** Helps in developing targeted strategies to reduce cancellations.
6. **What is the trend of ADR over the years?**

- **Hint:** Plot the average ADR per year using the 'arrival_date_year' and 'adr' columns.
- **Solution Idea:** Create a line plot showing the trend of average ADR over different years.
- **How it helps:** Identifies changes in ADR over time.
- **Business Impact:** Informs long-term pricing and revenue management strategies.

7. **Which month has the highest revenue?**
   - **Hint:** Calculate the total revenue for each month using the 'arrival_date_month' and 'adr' columns.
   - **Solution Idea:** Group the data by 'arrival_date_month' and sum the ADR values to find the month with the highest revenue.
   - **How it helps:** Identifies the most profitable months.
   - **Business Impact:** Helps in planning promotional activities and resource allocation.

8. **What is the impact of special requests on ADR?**
   - **Hint:** Analyze the relationship between 'total_of_special_requests' and 'adr'.
   - **Solution Idea:** Explore the correlation between the number of special requests and ADR.
   - **How it helps:** Understands how special requests influence pricing.
   - **Business Impact:** Helps in optimizing pricing strategies based on customer preferences.

9. **What is the average stay duration for repeated guests versus new guests?**
   - **Hint:** Calculate the mean stay duration for repeated and new guests using the 'is_repeated_guest', 'stays_in_week_nights', and 'stays_in_weekend_nights' columns.
   - **Solution Idea:** Compare the average stay duration between repeated and new guests.
   - **How it helps:** Understands guest behavior and loyalty.
   - **Business Impact:** Helps in developing targeted loyalty programs and improving guest retention.

10. **Which room type has the highest number of bookings?**
    - **Hint:** Count the number of bookings for each room type using the 'reserved_room_type' column.
    - **Solution Idea:** Identify the room type with the highest booking count.
    - **How it helps:** Identifies the most popular room types.
    - **Business Impact:** Helps in optimizing room availability and marketing strategies.

Advanced Level Questions with Solution Ideas

1. **What factors significantly impact the cancellation rate?**
   - **Hint:** Perform a logistic regression analysis using relevant features.
   - **Solution Idea:** Identify significant predictors of cancellations by analyzing features such as lead time, booking changes, and special requests.
   - **How it helps:** Understands the key drivers behind cancellations.
   - **Business Impact:** Helps in developing targeted policies to reduce cancellations.

2. **How does the ADR vary with the number of adults, children, and babies?**
   - **Hint:** Use a multiple regression analysis to explore the relationship.
   - **Solution Idea:** Analyze the impact of the number of adults, children, and babies on the ADR.

- ○ **How it helps:** Understands how different guest compositions affect pricing.
- ○ **Business Impact:** Helps in optimizing room rates based on guest demographics.

3. **What is the impact of booking changes on guest satisfaction as indicated by special requests?**
   - ○ **Hint:** Analyze the correlation between 'booking_changes' and 'total_of_special_requests'.
   - ○ **Solution Idea:** Investigate how the number of booking changes correlates with the number of special requests made by guests.
   - ○ **How it helps:** Understands the relationship between booking changes and guest needs.
   - ○ **Business Impact:** Helps in improving booking flexibility and customer satisfaction.

4. **What is the seasonal impact on booking cancellations?**
   - ○ **Hint:** Analyze the cancellation rate across different seasons using the 'arrival_date_month' column.
   - ○ **Solution Idea:** Calculate the cancellation rate for each month and compare the rates across different seasons.
   - ○ **How it helps:** Identifies seasonal patterns in cancellations.
   - ○ **Business Impact:** Helps in developing strategies to mitigate seasonal cancellation spikes.

5. **How does the booking lead time distribution vary between different market segments?**
   - ○ **Hint:** Visualize the lead time distribution for each market segment using a box plot.
   - ○ **Solution Idea:** Create box plots to compare the distribution of lead times across different market segments.
   - ○ **How it helps:** Understands the booking behavior of different market segments.
   - ○ **Business Impact:** Helps in tailoring marketing and promotional strategies to different market segments.

Additional Considerations for Advanced Questions:

- ● Ethical and Privacy Concerns: While developing predictive models and handling patient data, it's crucial to consider the ethical implications and ensure privacy and data protection standards are met.
- ● Interdisciplinary Collaboration: Engage with clinical experts, healthcare providers, and patients to validate findings and refine intervention strategies.
- ● Continuous Improvement: Consider these analyses as part of an ongoing effort to improve healthcare delivery. Regularly update models and strategies based on new data and outcomes.

Deliverables

- ● Case Study Document: Includes problem statement, data dictionary, and questions.
- ● Solution Guide: Detailed answers and explanations for each question.
- ● Additional Resources: References for further exploration.

Desired Outcome

The trainees will develop an analytical and logical mindset, understanding the importance of various factors in loan analysis. They will learn to apply different data analysis techniques to uncover insights and make data-driven decisions.