Hotel Booking Analysis

Name - Bhumika singh
Email - asurewaps@gmail.com

Short Summary of capstone project (Hotel Booking Analysis)

The purpose of this hotel booking research was to investigate a hotel's customer data and uncover any relevant trends or correlations. The goal of this exploratory data analysis (EDA) was to look at the hotel booking data set and see whether there were any potential links between important factors.

Customer booking information was included in the data set. Descriptive statistics were computed for each variable as part of the analysis, and visualizations were created to investigate the correlations between various variables. We created a number of charts to get understanding from the dataset, including a bar plot, pie chart, heatmap, pairplot, and barplot.

The data set was composed of over 119390 hotel bookings, each containing several variables such as

- Hotel : Kind of hotel (Resort or City)
- is_cancelled : Whether the reservation was canceled(1) or not (0)
- Lead_time: Days that are left before the guests actually arrive
- arrival_date_year: Year of arrival date
- arrival_date_month: Month of arrival date
- arrival_date_week_number : The week and year of the arrival
- arrival_date_day_of_month: Day of arrival date
- stays_in_weekend_nights:How many weekends (Saturday or Sunday) do customers stay at the hotel?
- stays_in_week_nights: The number of weeknights (Monday through Friday) that visitors stay at the hotel.
- adults: Number of adults among the guests
- children: Number of children
- babies: Number of babies
- meal: Type of meal booked
- country: country of the guests

- market_segment: Segmentation of the market
- distribution_channel: The channel name for booking distribution
- is_repeated_guest: If the reservation came from a returning customer(1) or not (0)
- previous_cancellation: The number of earlier reservations that the client canceled before the current reservation
- previous_bookings_not_cancelled: The number of prior bookings that the customer did not cancel before the current booking
- reserved_room_type: Code from room type reserved
- assigned_room_type: Code of room type assigned
- booking_changes: Number of changes made to the booking
- deposit_type: Type of deposit made by the guest
- agent: ID of travel agent who made the booking
- company: ID of the company that made the booking
- days_in_waiting_list: Number of the days the booking was in the waiting list
- customer_type: Type of customer, assuming one of four categories
- adr: Average daily rate
- required_car_parking_spaces : The quantity of parking spaces needed, but the client
- total_of_special_requesrs: The quantity of unique requests the client has made
- reservation_status: Status of reservation (canceled, checked out, or no-show)
- reservation_status_date: Date of the most recent update to the reservation status

There were 31994 duplicate values deleted. Country had 452 null values, children had 4, agents had 12193, and the company had 82137 null values. For these variables, we substituted the null value with the mode of each variable (country, children, agent), but the variable "company" had more than 50% null value, so we eliminated it. We also eliminated outliers from lead_time and adr. The total number of observations in the final dataset was 87396.

The data types of variables children, agent, and reservation_status_date were also modified to int64, int64, and datetime64, respectively. To improve convenience, we included new variables: total_stays, total_people, total_childrens, reserved_room_assigned, guest_category, and lead_time_category. We transformed total_people and total_childrens from the floating 64 data type to int64.

Following data cleansing, exploratory data analysis yielded the following intriguing findings:

1. The majority of visitors prefer the city hotel.
2. The most reservations were made by agent no. 9.
3. The percentage of returning visitors is lower, at 3.86%.
4. The majority of people prefer room type A.
5. Food of the BB type is generally favored.
6. July has the most reservations after August, with August having the most.
7. The majority of distribution channels, or 79.13%, are TA/TO.
8. Hotel City has the highest ADR. Revenue is increased by the highest ADR.
9. The year 2016 saw the highest number of bookings, totaling 42313.
10. Longer wait times at city hotels indicate that they are busier.
11. The GDS distribution channel made the largest contribution to ADR in city hotels, but not at all in resort hotels.
12. Less than seven days is the ideal stay duration for both types of hotels.
13. Not only do returning visitors not cancel their reservations, but so do regular visitors.
14. There is a -0.51 negative correlation between the arrival date week number and arrival date year columns.
15. The correlation between stays_in_week_nights and total_stays is positive (0.95).

In conclusion, the hotel booking analysis system is a reliable and comprehensive approach to understanding, optimizing, and enhancing hotel operations inside the competitive and expeditious hospitality industry. By employing hotel booking analytic solutions, lodging organizations may improve revenue production, tailor marketing strategies, and increase occupancy rates. The insights provided by the system are a helpful resource for hotel management and stakeholders that strive for operational excellence.