Hotel Reservation Data Analysis

This project consists of two main parts: Exploratory Data Analysis (EDA) and a results-based conclusion.

The goal of this project was to imagine myself as an intern for a company that was tasked with analyzing company data and providing a report to the client to give them information on what is going well for them, what was not going well for them, and how could they capitalize on certain information.

This Project consisted of a public dataset on hotel reservation data from booking.com that was analyzed to perform the EDA and conclusion. The dataset consists of the following columns:

1. hotel- The dataset contains the booking information of two hotels. One of the hotels is a resort hotel, and the other is a city
2. is_canceled- Value indicating if the booking was canceled (1) or not (0).
3. arrival_date_year- Year of arrival date
4. arrival_date_month- Month of arrival date with 12 categories: “January” to “December.”
5. arrival_date_week_number- The week number of the arrival date
6. arrival_date_day_of_month- Day of the month of the arrival date
7. stays_in_weekend_nights- Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
8. stays_in_week_nights- Number of weeknights (Monday to Friday) the guest stayed or booked to stay at the hotel BO and BL/Calculated by counting
9. adults- Number of adults
10. children- Number of children
11. babies- Number of babies
12. meal- BB – Bed & Breakfast, FB-full board, HB-half board
13. country- Country of origin.
14. market_segment- Market segment designation. In categories, the term “TA” means “Travel Agents,” and “TO” means “Tour Operators.”
15. distribution_channel- Booking distribution channel. The term “TA” means “Travel Agents,” and “TO” means “Tour Operators.”
16. is_repeated_guest- Value indicating if the booking name was from a repeated guest (1) or not (0)
17. previous_cancellations- Number of previous bookings that the customer canceled prior to the current booking
18. previous_bookings_not_canceled- Number of previous bookings not canceled by the customer prior to the current booking
19. reserved_room_type- Code of room type reserved. Code is presented instead of designation for anonymity reasons
20. assigned_room_type- Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due

21. booking_changes- Number of changes/amendments made to the booking from the moment the booking was entered on the PMS
22. deposit_type- No Deposit – no deposit was made; Non-Refund – a deposit was made in the value of the total stay cost; Refundable
23. days_in_waiting_list- Number of days the booking was on the waiting list before it was confirmed to the customer
24. customer_type- Group – when the booking is associated with a group;
25. adr- Average Daily Rate (Calculated by dividing the sum of all lodging transactions by the total number of staying nights)
26. required_car_parking_spaces- Number of car parking spaces required by the customer
27. total_of_special_requests- Number of special requests made by the customer (e.g., twin bed or high floor)
28. reservation_status- Check-Out – customer has checked in but already departed; No-Show – the customer did not check in and did inform
29. reservation_status_date- The date at which the last status was set. This variable can be used in conjunction with the ReservationStatus 30. name- Name of the Guest (Not Real)

Exploratory Data Analysis (EDA)

All quantitative and categorical variables are analyzed using R. All statistics and visuals are created within R.

Project Process

1. Exploratory Data Analysis (EDA): Observations of the dataset and descriptions of all variables, categorical or quantitative. All variables are summarized.

Data Analysis

2. Variable “hotel” (that means city hotel and resort hotel).
3. previous_cancellations and previous_bookings_not_canceled with the variable hotel.
4. Summarize the hotel variable based on year.
5. Check the number of bookings by month.
6. Find out the unique countries.
7. Number of bookings based on countries (a plot is great).
8. Check outliers for average daily rate(adr) based on hotel types.
9. Check the average daily rate(adr) vs. hotel.
10. Customer type vs. hotel type.
11. Customer type vs. special request.
12. Hotel Preference by customer type.
13. Discuss the correlation of the dataset.

14. Pick any two variables and fit a regression line. For this situation, you may use the above results in order to select which two variables show the “best” relationship.
15. Average daily rate trend over three years

Reflection/Conclusion

The reflection component was to explain how completing this project helped my understanding of statistical concepts. Part of the reflection was meant to simulate a brief initial overview that would be given to the client before seeing the statistical details.

Also included in the conclusion were the following:

a. A section summarizing the results from the EDA reports. Any variables or pairs of variables that were surprising to me, or possible reasons why variables may or may not be associated.

b. A critique of my study by citing its good and bad points. Suggest how you could improve the study if you could do it over again.

Appendix

R code