Hotel Bookings Exploratory Data Analysis

Contributor Role:

Shivam R. Mistry (mistryshivam619@gmail.com)

Data Cleaning and Feature Engineering

(1) Removing Duplicate rows

(2) Handling null values

(3) Converting columns to appropriate data types

(4) Removing outliers

(5) Creating new columns

Exploratory Data Analysis:

Q1) Which agent makes the most no. of bookings?
Q2) Which room type is in most demand and which room type generates the highest adr?
Q3) Which meal type is the most preferred meal of customers?
Q4) What is the percentage of bookings in each hotel?
Q5) Which is the most common channel for booking hotels?
Q6) Which are the busiest months?
Q7) From which country most of the guests are cumin?
Q8) How long do people stay at the hotels?
Q9) Which hotel seems to make more revenue?
Q10) Which hotel has a higher lead time?
Q11) What is preferred stay length in each hotel?
Q12) Which hotel has higher bookings cancellation rate.
Q13) Which hotel has a high chance that its customer will return for another stay?
Q14) Which channel is mostly used for the early booking of hotels?
Q15) Which channel has a longer average waiting time?
Q16) Which distribution channel brings better revenue-generating deals for hotels?
Q17) Which significant distribution channel has the highest cancellation percentage?
Q18) Does longer waiting period or longer lead time causes the cancellation of bookings?

Q19) Whether not getting allotted the same room type as demand is the main cause of cancellation for bookings?
Q20) Does not allotting the same room as demanded affect adr?
Q21) What is the trend of bookings within a month?
 Q22) Which types of customers mostly make bookings?

Conclusion:

(1) Around 60% bookings are for City hotel and 40% bookings are for Resort hotel, therefore City Hotel is busier than Resort hotel. Also, the overall adr of City hotel is slightly higher than Resort hotel.
(2) Mostly guests stay for less than 5 days in hotel and for longer stays Resort hotel is preferred.
(3) Both hotels have significantly higher booking cancellation rates and very few guests less than 3 % return for another booking in City hotel. 5% guests return for stay in Resort hotel.
(4) Most of the guests came from European countries, with most of guests coming from Portugal.
(5) Guests use different channels for making bookings out of which most preferred way is TA/TO.
(6) For hotels higher adr deals come via GDS channel, so hotels should increase their popularity on this channel.
(7) Almost 30% of bookings via TA/TO are cancelled.
(8) Not getting same room as reserved, longer lead time and waiting time do not affect cancellation of bookings. Although different room allotment does lowers the adr.
(9) July- August are the busier and profitable months for both of hotels.
(10) Within a month, adr gradually increases as month ends, with small sudden rise on weekends.
(11) Couples are the most common guests for hotels, hence hotels can plan services according to couples needs to increase revenue.
(12) More number of people in guests results in a greater number of special requests.
(13) Bookings made via complementary market segment and adults have on average high no. of special request.
(14) For customers, generally the longer stays (more than 15 days) can result in better deals in terms of low adr.

 Challenges:

(1) There was a lot of duplicate data.

(2) Data was present in wrong datatype format.

(3) Choosing appropriate visualization techniques to use was difficult.

(4) A lot of null values were there in the dataset.

GitHub Link:

https://github.com/shivamm007/Hotel_booking_Eda

Drive

https://drive.google.com/drive/folders/1ChqDecdKnvA0T0QZ4IHsH5dD5pl6Y83B?usp=share _link