

Hotel Bookings Analysis

Analysis is based on customer data for a hotel. This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things. The data has 119,390 entries across 32 columns. The aim of this project is to predict whether a hotel booking is going to be canceled or not.

EXPLORING THE DATA

CHART 1: LEAD TIME

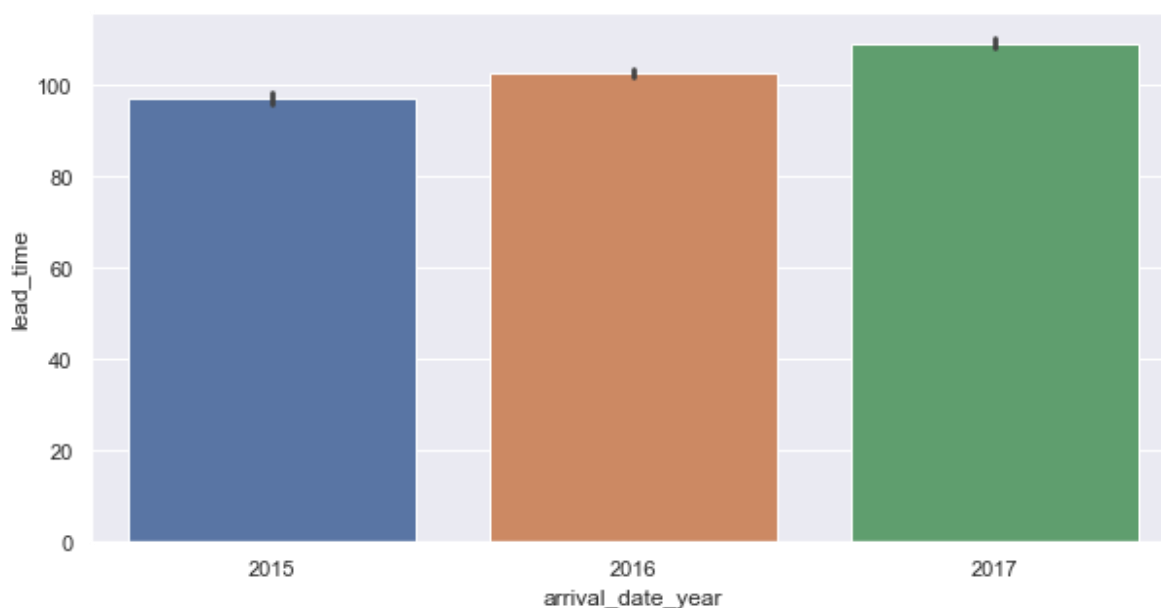


Table 1 shows the average lead time which is the amount of time (in hours) between a booking and the guest arrival. This will help the hotel in managing room availability in order to ensure that profit is being maximized.

CHART 2: DEPOSIT TYPE

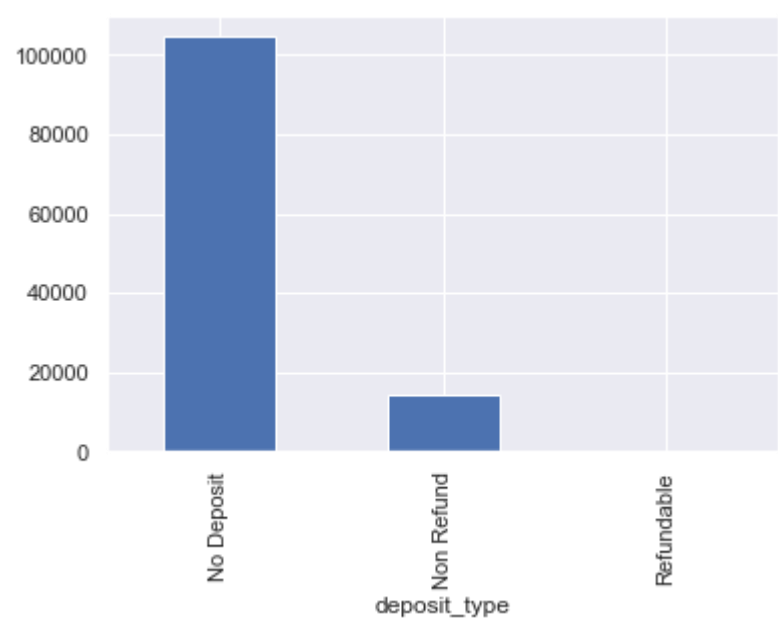


Table 2 shows that there are no cash deposits for most hotel bookings. Only in few instances are non-refundable deposits made.

CHART 3: Reservation Status

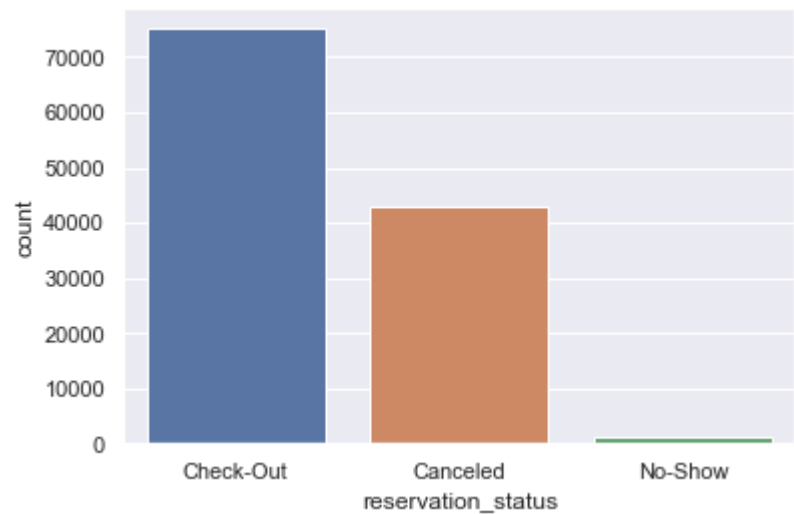
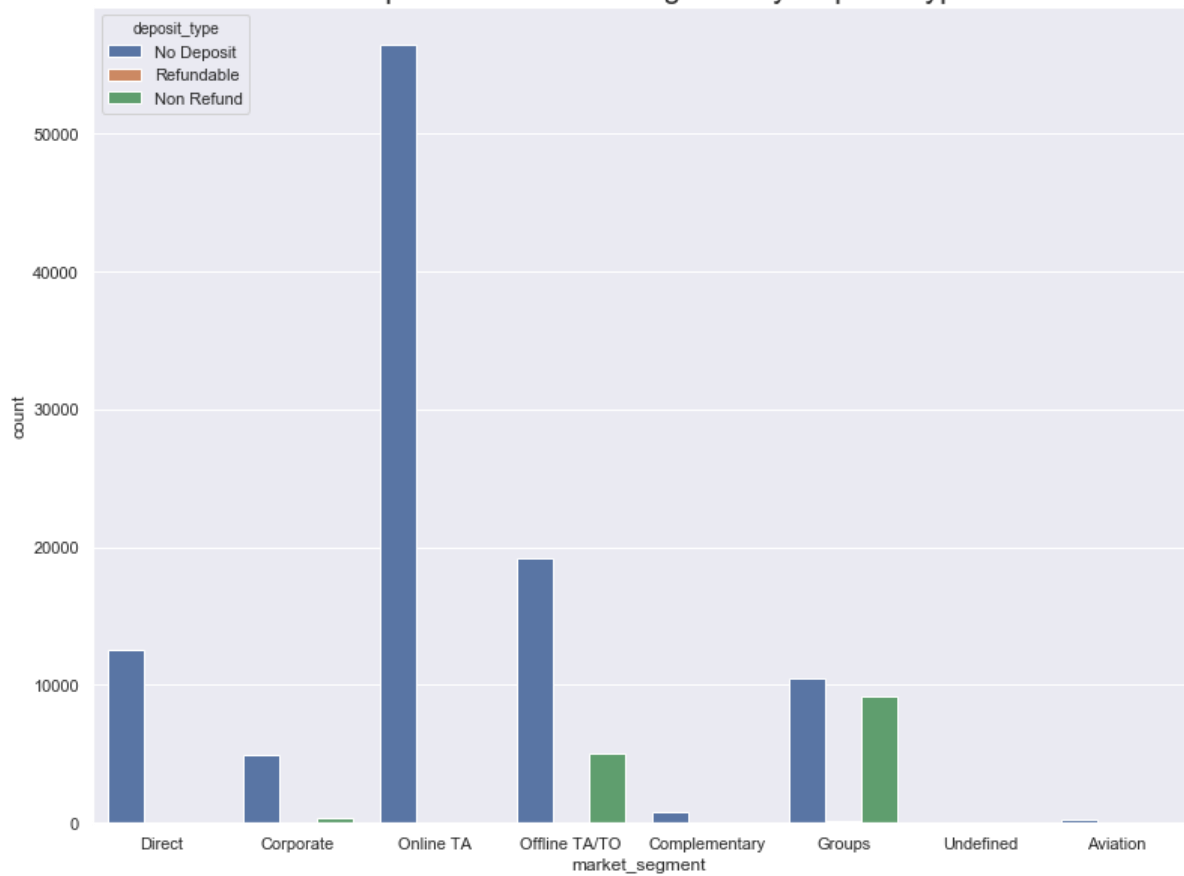


Table 3 shows the relative amount of bookings which showed up (check-out), those who canceled and those who didn't show up at all.

CHART 4: Segment by Deposit Type
Countplot Distrubiton of Segment by Deposit Type



In table 2, there were instances of those who made non-refundable deposits. Table 4 shows the segments that are paying the non-refundable deposits. These are the offline TA/TO market segment and groups.

BUILDING THE MODEL

The dependent variable that is predicted is whether a booking will be canceled or not.

The independent variables used to predict the models are: lead time, total of special requests, required car parking spaces, booking changes, previous cancellations and average daily rate(adr).

A logistic regression was used to build the model.

Rationale behind the chosen independent variables:

Lead time: The more time passes before a customer checking in, the more likelihood they are going to cancel.

Total of special requests: if a customer has a high number of special requests, he/she will want to ascertain that the hotel is able to meet its demands. Therefore, such customers are likely to cancel if they don't think the hotel can meet their request.

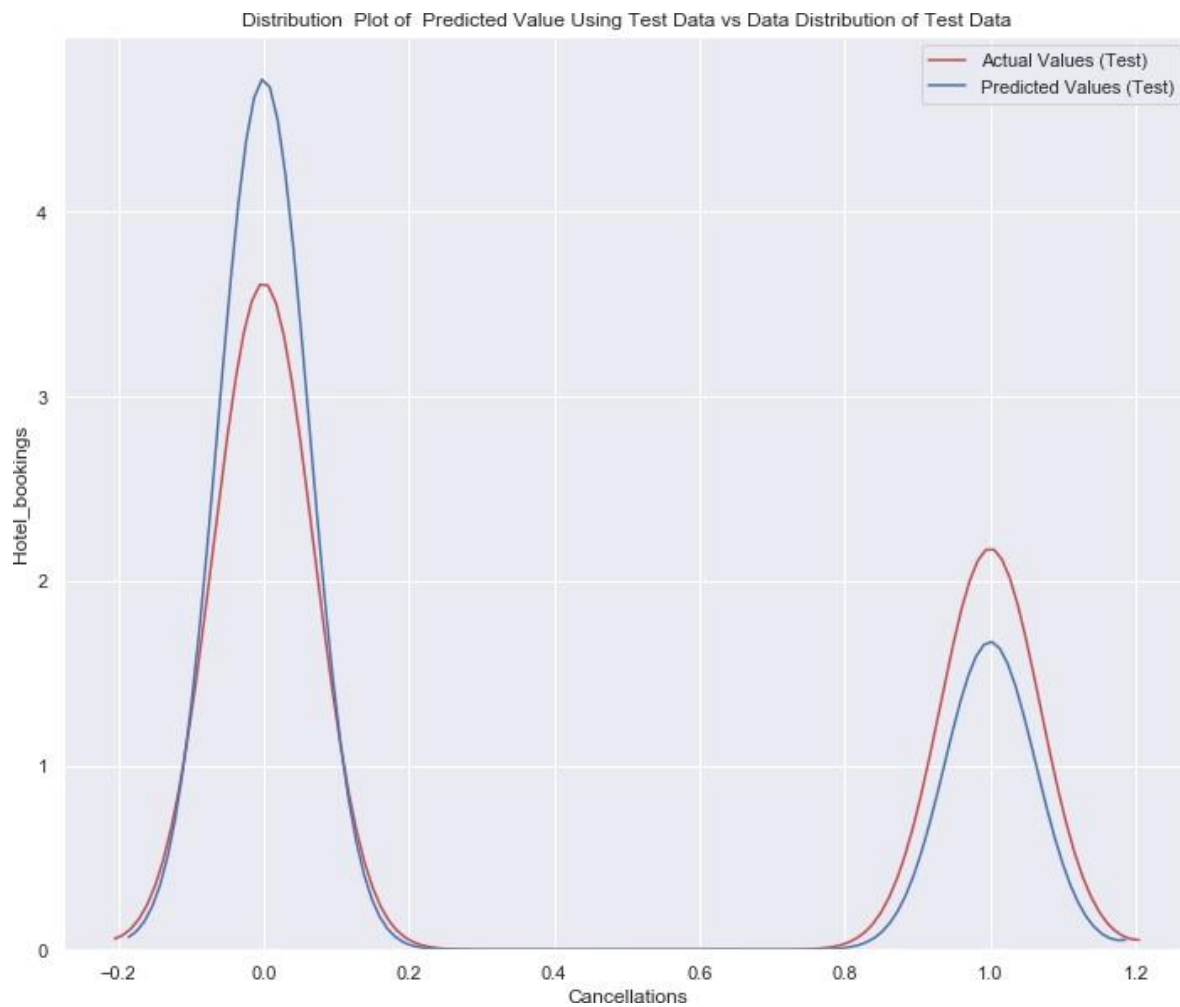
Required car parking spaces: This is another factor a customer might consider, for example, it is a corporate booking for a conference and they are expecting a lot of guests, they will likely cancel if they feel the parking would not be enough.

Booking changes: customers making booking changes have reason to believe their previous booking does not satisfy their requirements, this might lead to cancellation.

Previous cancellations: if a customer has history of previous cancellations, there is a chance they might cancel again.

Average daily rate(adr): if the average daily rate is high, a customer might cancel their booking if they find cheaper alternative elsewhere.

CHART 5: Regression Plot



The chart above is a visual representation of the predictions of the created model against the actual values. Bookings not cancelled are 0 while cancelled bookings are 1.

The accuracy of the prediction is 73 percent.

Coefficient of the Independent Variables

Independent Variable	Coefficient	ODDS RATIO
Previous cancellations	1.445636	4.244550
Lead time	0.563101	1.756110
adr	0.384834	1.469371
Booking changes	-0.449718	0.637808
Total of special requests	-0.571787	0.564516
Intercept	-1.620975	0.197706
Required car parking spaces	-4.417063	0.012070

Mathematical Representation

Cancellation = $-1.620975 + 1.445636(\text{previous cancellations}) + 0.563101(\text{lead time}) + 0.384834(\text{adr}) + 0.0449718(\text{booking changes}) - 0.571787(\text{total of special requests}) - 4.417063(\text{required car parking spaces})$