MACHINE LEARNING – 1 PROJECT

Coded Project Report

INN GROUP OF HOTELS

Submitted By – Yash Juneja

Submitted to – Great Learning

Table of Contents

Context

A significant number of hotel bookings are called off due to cancellations or no-shows. The typical reasons for cancellations include change of plans, scheduling conflicts, etc. This is often made easier by the option to do so free of charge or preferably at a low cost which is beneficial to hotel guests but it is a less desirable and possibly revenue-diminishing factor for hotels to deal with. Such losses are particularly high on last-minute cancellations.

The new technologies involving online booking channels have dramatically changed customers' booking possibilities and behavior. This adds a further dimension to the challenge of how hotels handle cancellations, which are no longer limited to traditional booking and guest characteristics.

The cancellation of bookings impact a hotel on various fronts:
1. Loss of resources (revenue) when the hotel cannot resell the room.
2. Additional costs of distribution channels by increasing commissions or paying for publicity to help sell these rooms.
3. Lowering prices last minute, so the hotel can resell a room, resulting in reducing the profit margin.
4. Human resources to make arrangements for the guests.

Objective

The increasing number of cancellations calls for a Machine Learning based solution that can help in predicting which booking is likely to be canceled. INN Hotels Group has a chain of hotels in Portugal, they are facing problems with the high number of booking cancellations and have reached out to your firm for data-driven solutions. You as a data scientist have to analyze the data provided to find which factors have a high influence on booking cancellations, build a predictive model that can predict which booking is going to be canceled in advance, and help in formulating profitable policies for cancellations and refunds.

Data Description

The data contains the different attributes of customers' booking details. The detailed data dictionary is given below.

Data Dictionary:

- Booking_ID: the unique identifier of each booking
- no_of_adults: Number of adults

- no_of_children: Number of Children
- no_of_weekend_nights: Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
- no_of_week_nights: Number of weeknights (Monday to Friday) the guest stayed or booked to stay at the hotel
- type_of_meal_plan: Type of meal plan booked by the customer:
    - Not Selected – No meal plan selected
    - Meal Plan 1 – Breakfast
    - Meal Plan 2 – Half board (breakfast and one other meal)
    - Meal Plan 3 – Full board (breakfast, lunch, and dinner)
- required_car_parking_space: Does the customer require a car parking space? (0 - No, 1- Yes)
- room_type_reserved: Type of room reserved by the customer. The values are ciphered (encoded) by INN Hotels Group
- lead_time: Number of days between the date of booking and the arrival date
- arrival_year: Year of arrival date
- arrival_month: Month of arrival date
- arrival_date: Date of the month
- market_segment_type: Market segment designation.
- repeated_guest: Is the customer a repeated guest? (0 - No, 1- Yes)
- no_of_previous_cancellations: Number of previous bookings that were canceled by the customer prior to the current booking
- no_of_previous_bookings_not_canceled: Number of previous bookings not canceled by the customer prior to the current booking
- avg_price_per_room: Average price per day of the reservation; prices of the rooms are dynamic. (in euros)
- no_of_special_requests: Total number of special requests made by the customer (e.g. high floor, view from the room, etc)
- booking_status: Flag indicating if the booking was canceled or not.

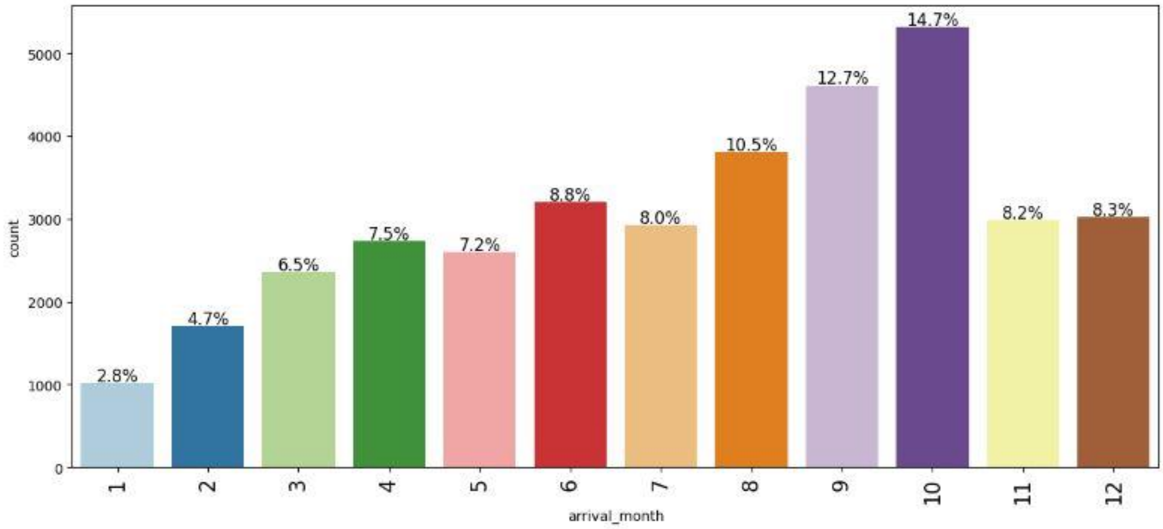
EDA Questions

1. What are the busiest months in the hotel?
2. Which market segment do most of the guests come from?
3. Hotel rates are dynamic and change according to demand and customer demographics. What are the differences in room prices in different market segments?
4. What percentage of bookings are canceled?
5. Repeating guests are the guests who stay in the hotel often and are important to brand equity. What percentage of repeating guests cancel?

6. Many guests have special requirements when booking a hotel room. Do these requirements affect booking cancellation?

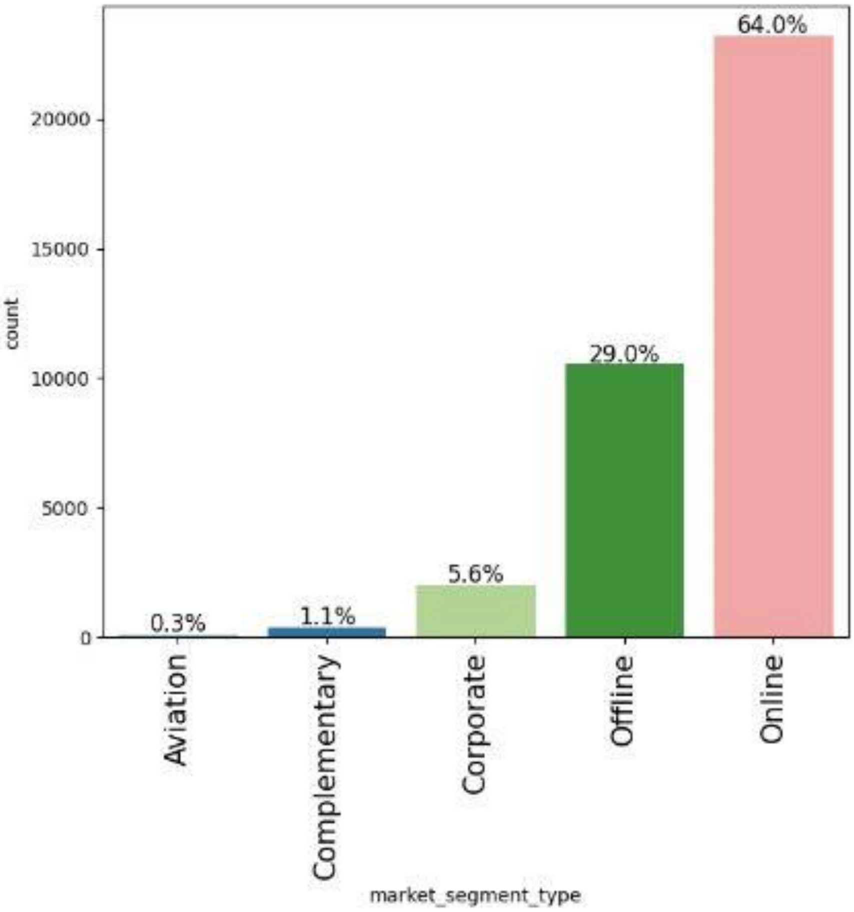
Exploratory Data Analysis

- **What are the busiest months in the hotel?**



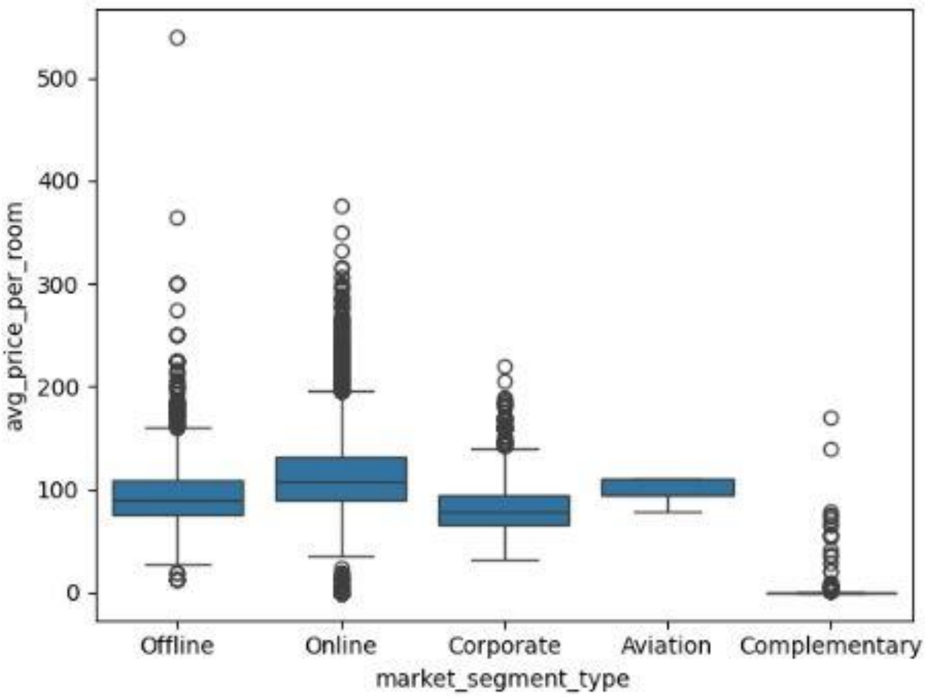
Month 10 = October with 14.7% of the total booking for the year

- **Which market segment do most of the guests come from?**



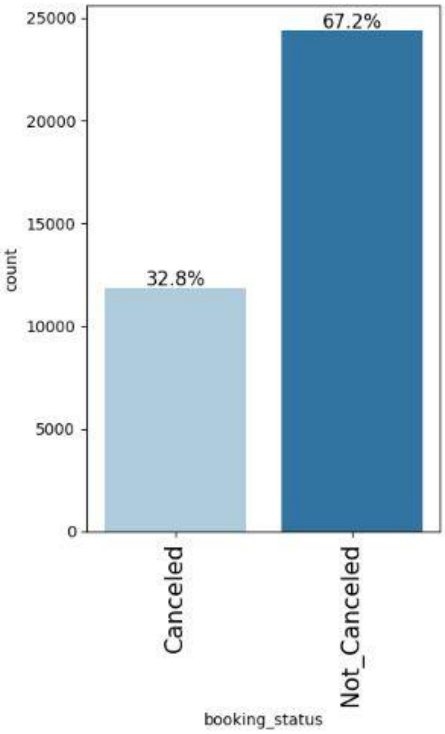
Online 23214 or 64% of the bookings come via the internet

- **Hotel rates are dynamic and change according to demand and customer demographics. What are the differences in room prices in different market segments?**



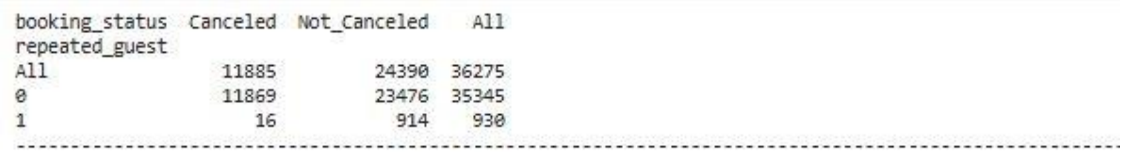
Online booking are the highest despite also having the highest amount of free rooms, Offline and Corporate are generally slightly lower priced with Corporate edging out for the lowest. Complimentary are of course free

- **What percentage of bookings are cancelled?**

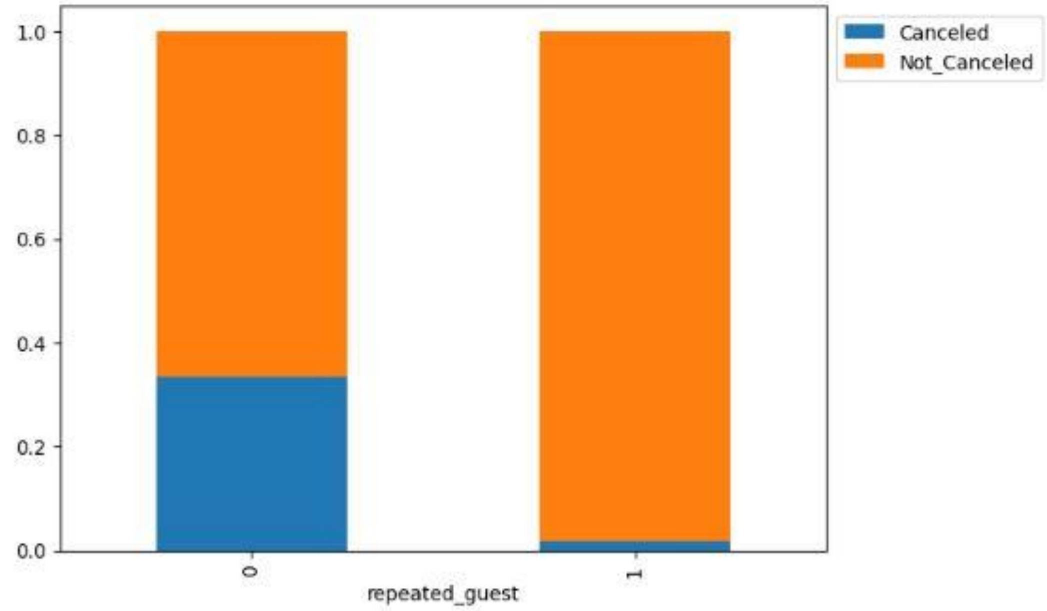


about 32.8% of bookings are canceled in the sample data.

- **Repeating guests are the guests who stay in the hotel often and are important to brand equity. What percentage of repeating guests cancel?**

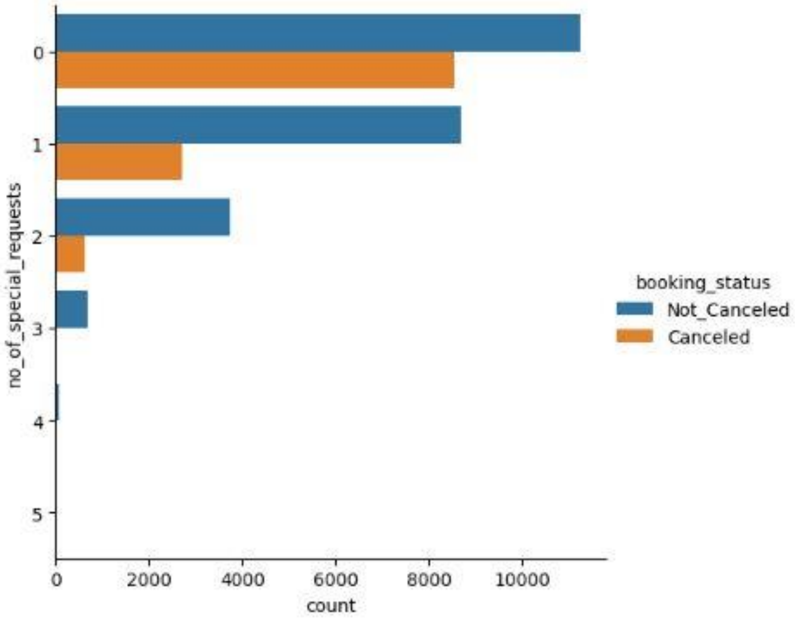


```
booking_status  Canceled  Not_Canceled    All
repeated_guest
All                11885         24390  36275
0                  11869         23476  35345
1                     16           914    930
------------------------------------------------------------------------------------------------------------------------
```



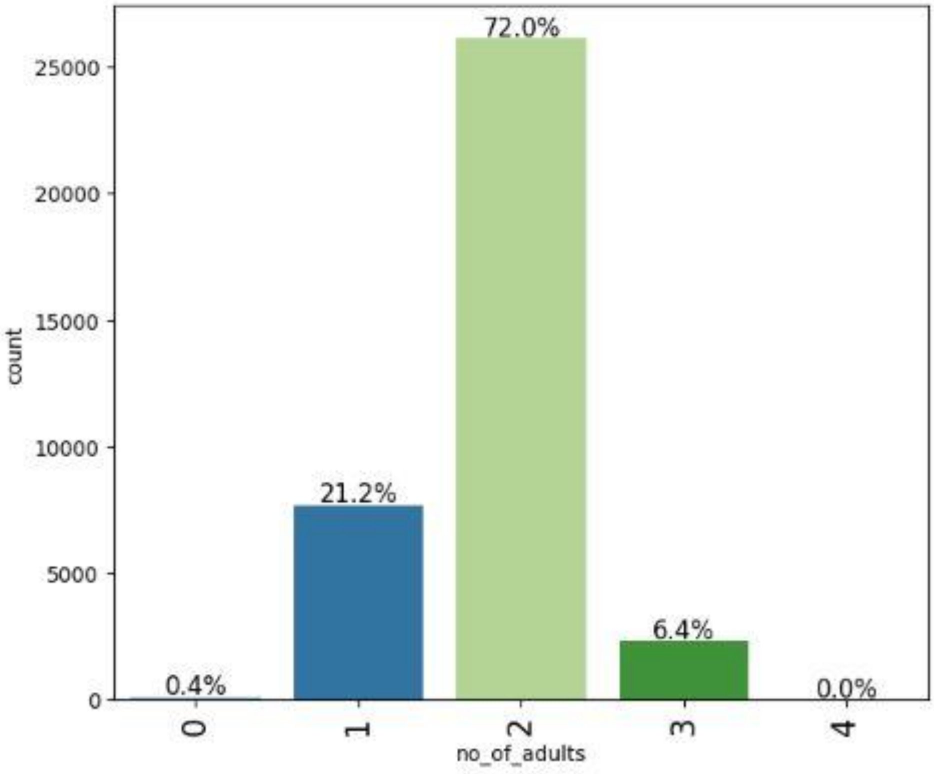
Repeating guest rarely cancel (1.75%)

- **Many guests have special requirements when booking a hotel room. Do these requirements affect booking cancellation?**

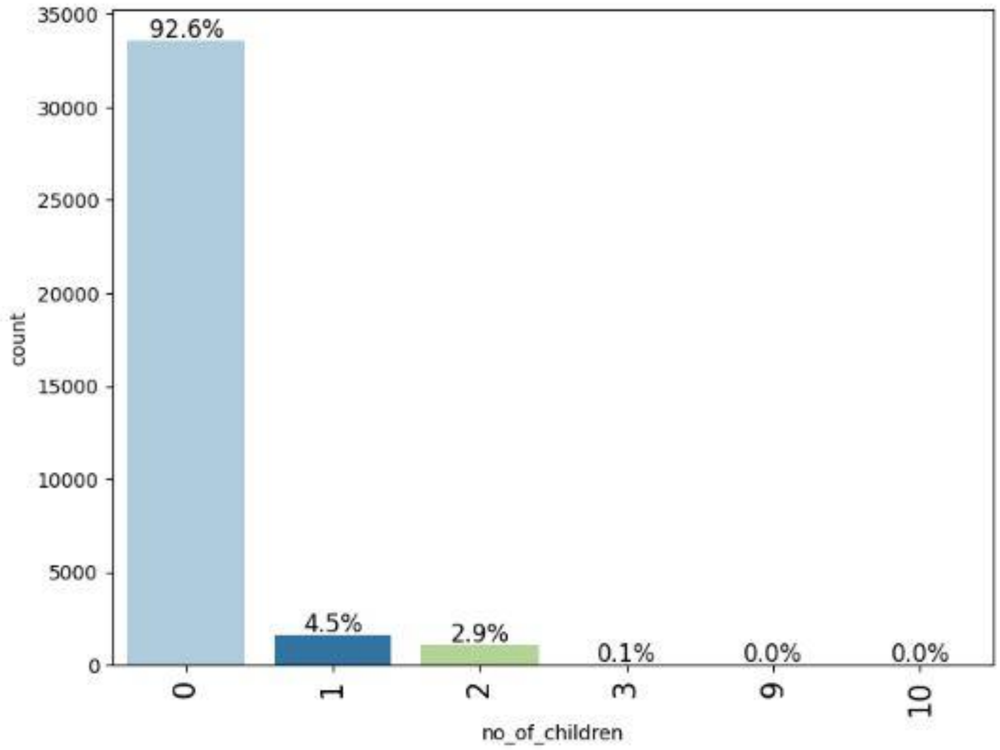


The absence of special request increases the likelihood of cancellation, the addition of special request begins to reduce the likelihood of cancellation at one and progressively reduces cancellation to Zero on the instance of a third request.

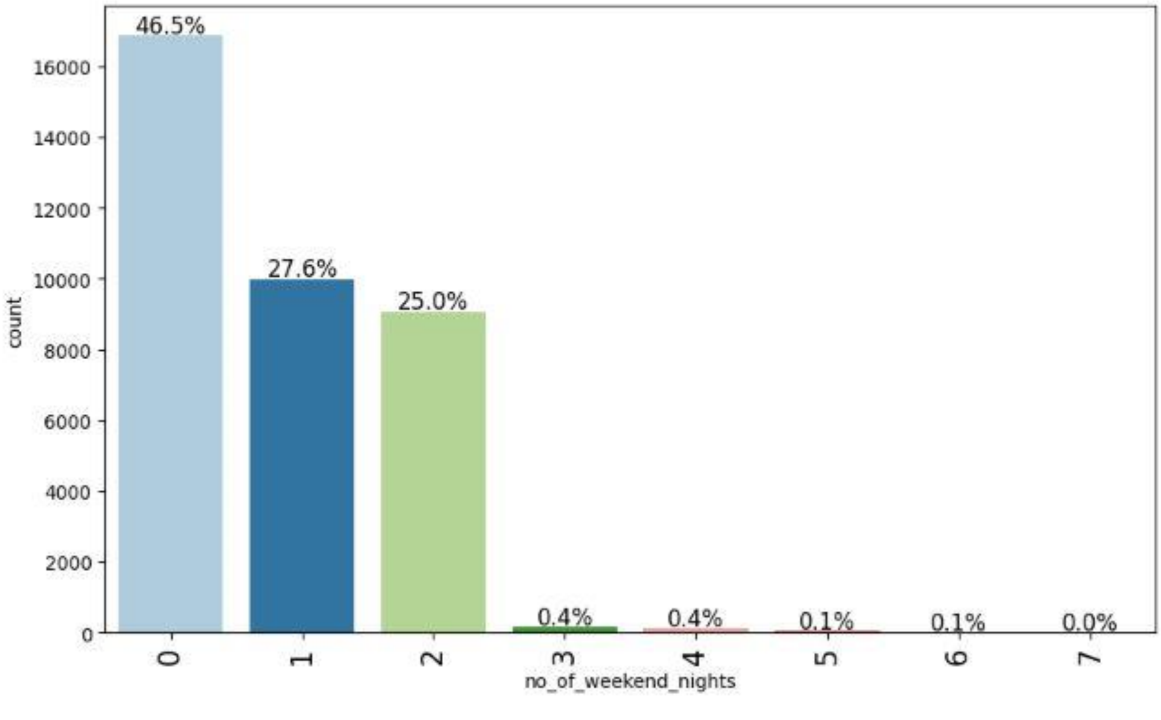
Univariate Analysis



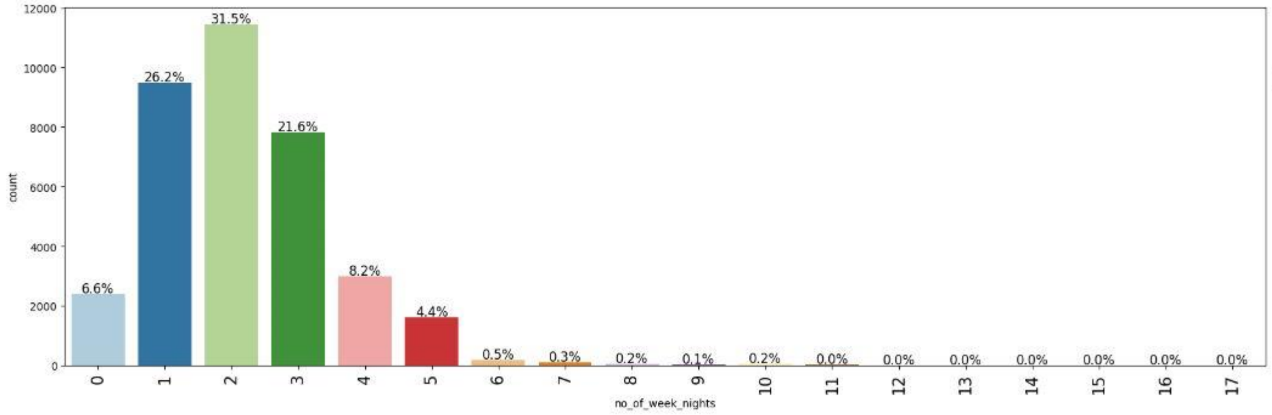
72% of bookings were made for 2 adults, whereas 21.2% bookings were made for 1 adult, It means hotel may be attractive place for couples



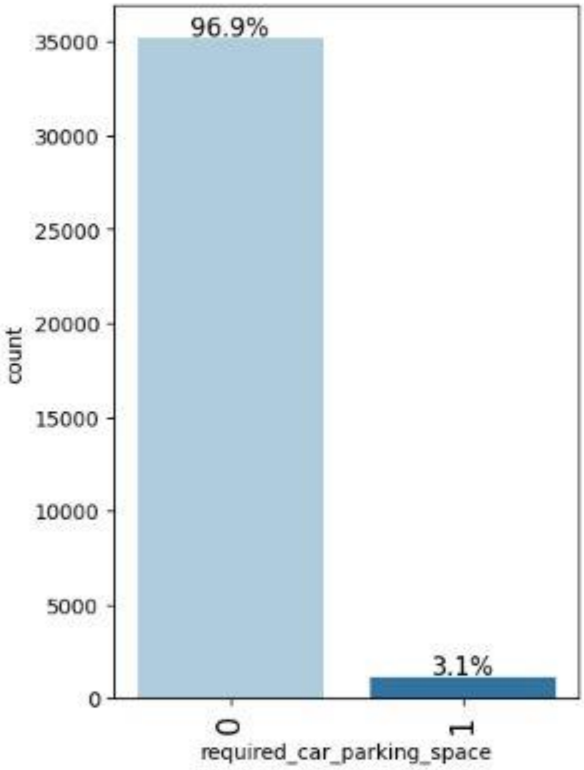
93% of the customer's didn't make reservations for children or Booking is done for those having no children.



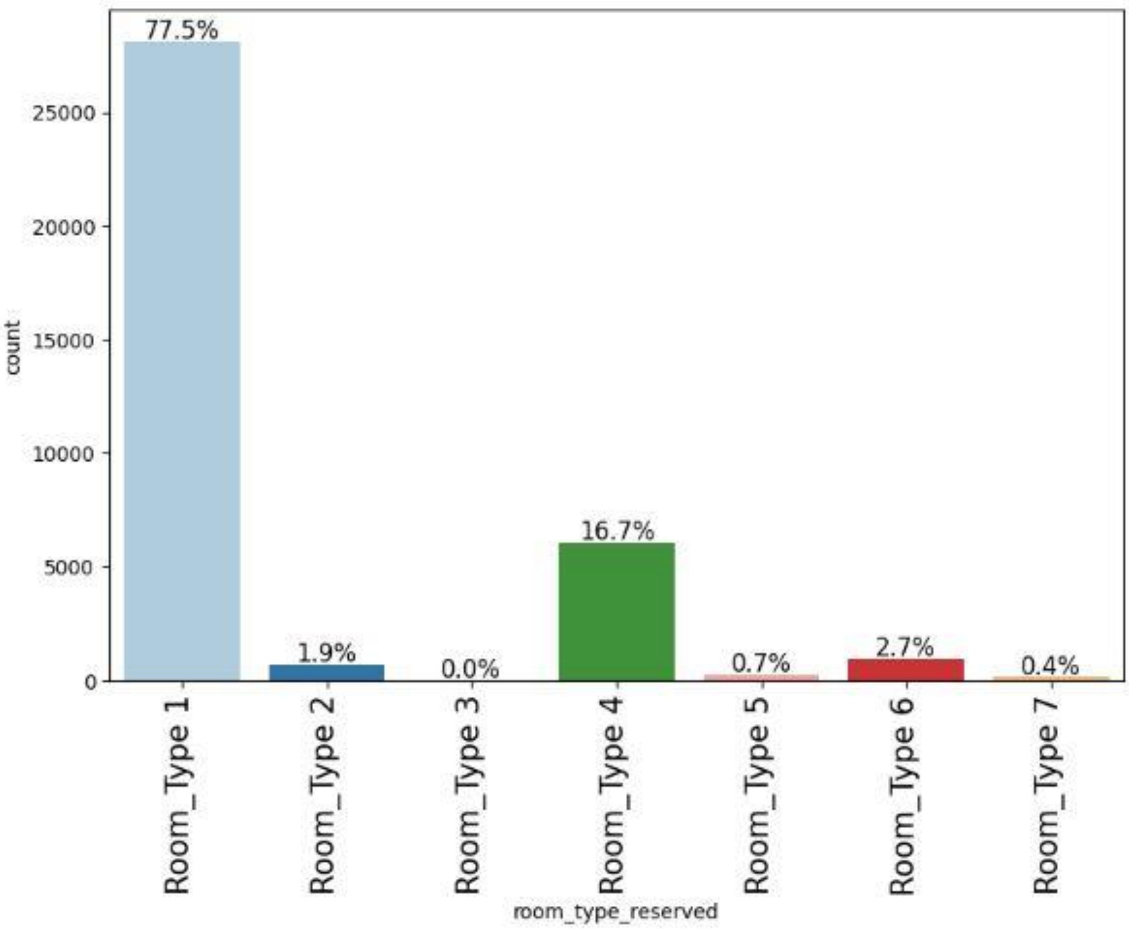
46.5% of the customers do not plan to spend the weekend in the hotel. The percentage of the customer's planning to spend 1 or 2 weekends in the hotel is almost same.



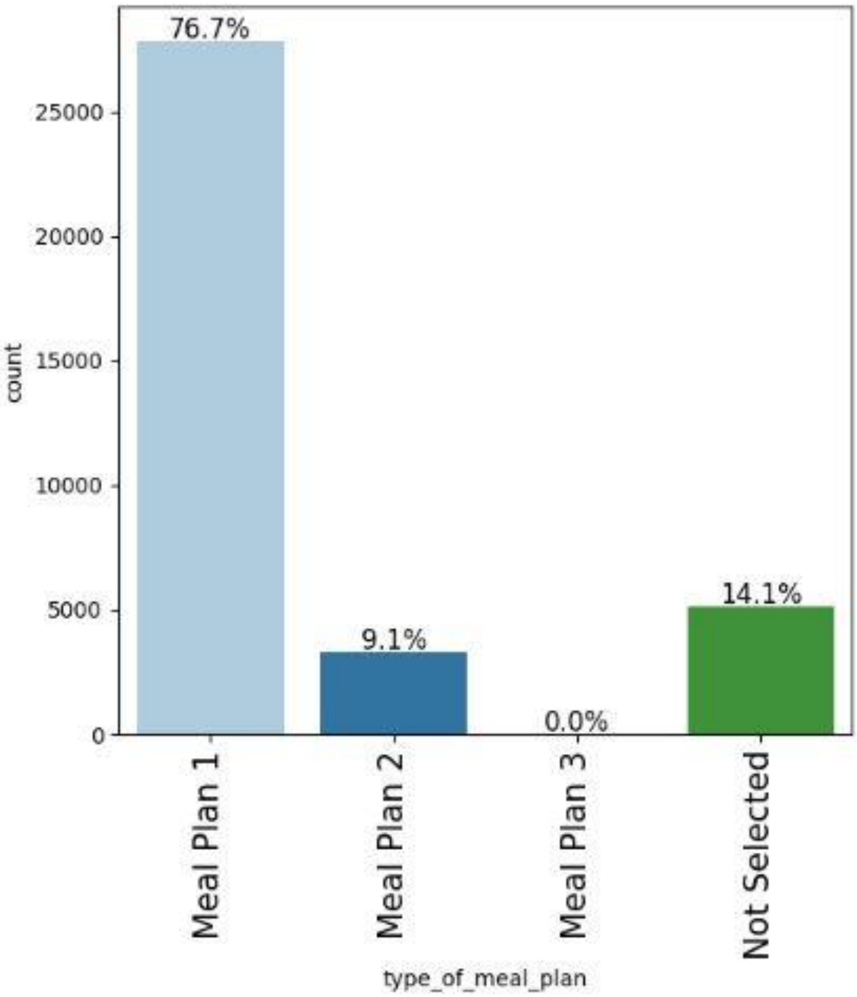
Most bookings are made for 2 Nights which is 31.5% and 1 Night that is 26.2%



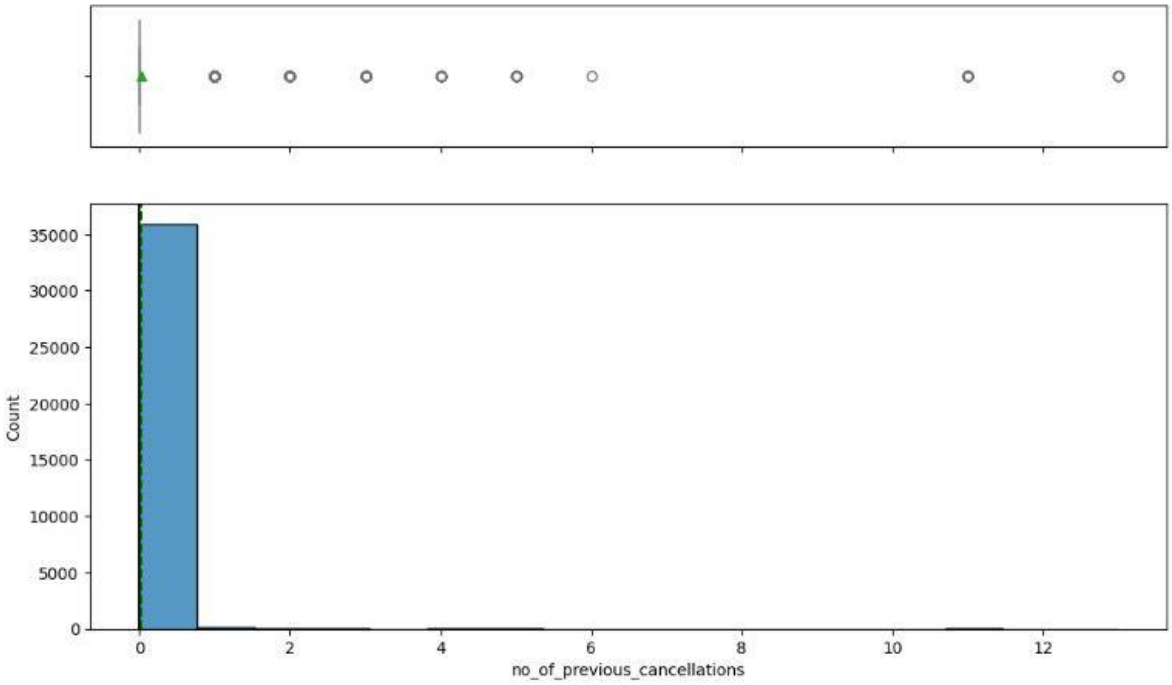
96.9% of the customers don’t need Car Parking space.



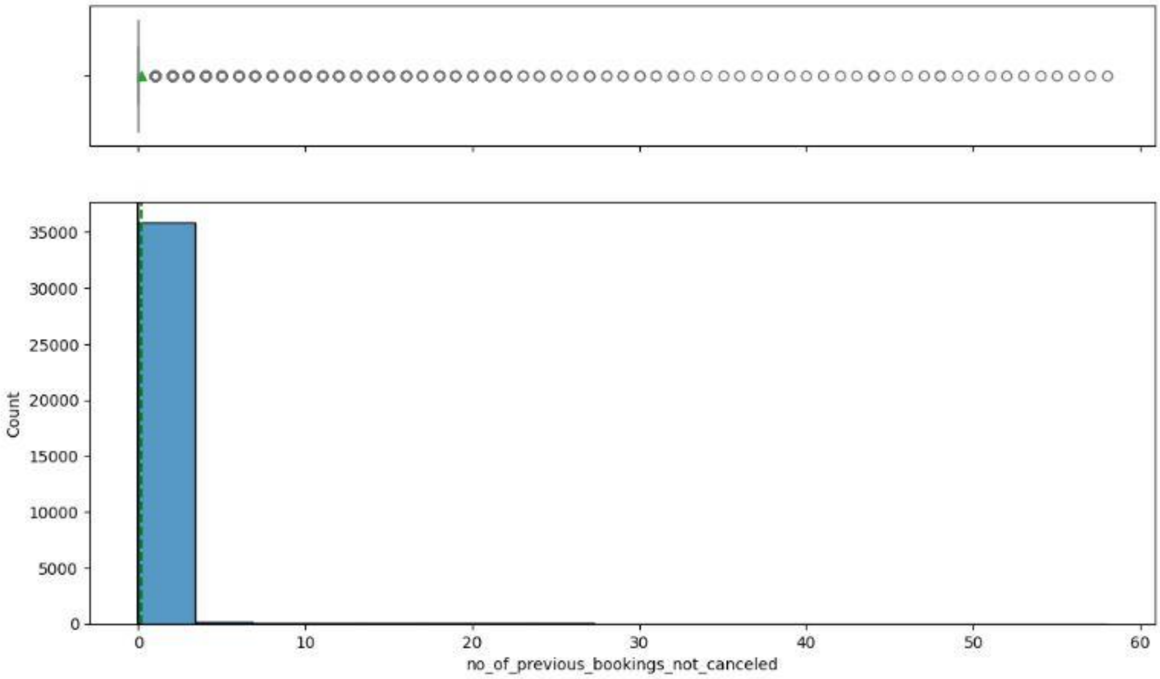
As per graph Most of the customers are preferring Room Type 1 followed by Room Type 4



Majority of the customers which is 76.7% are Preferring Mean Plan 1 And 14.1% of the customers are not selecting any plan

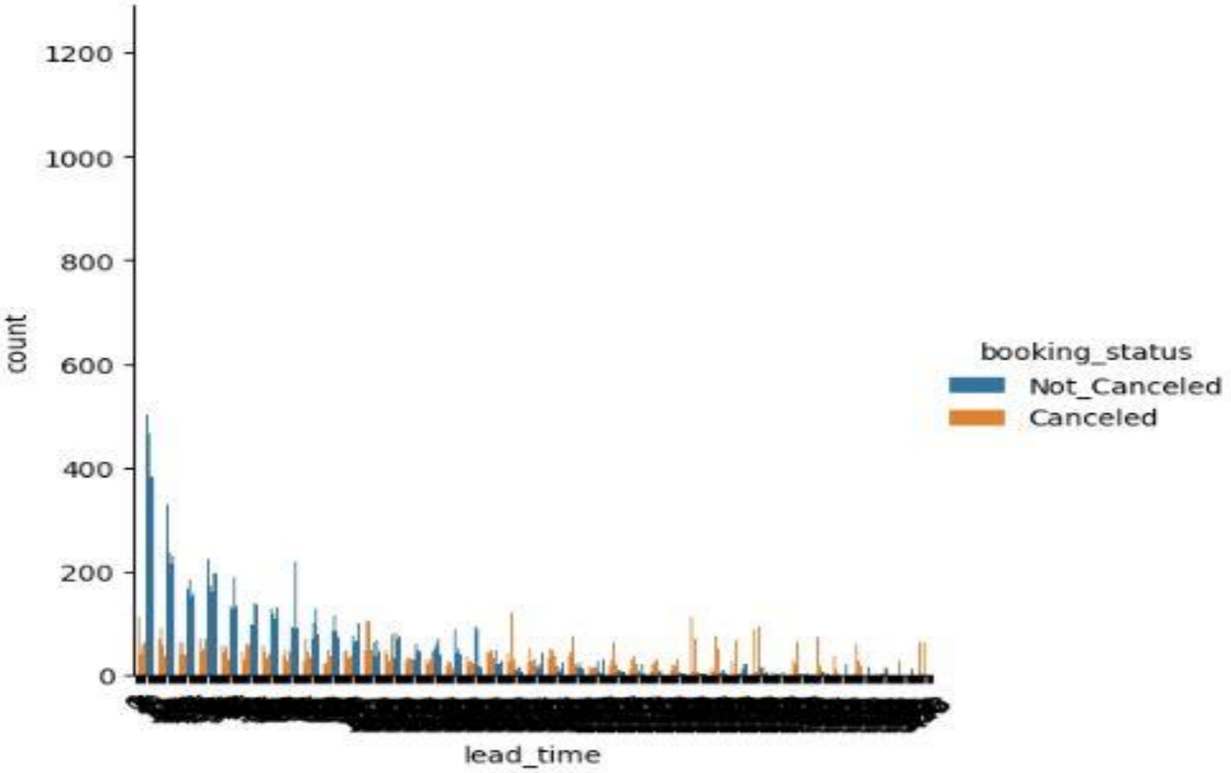


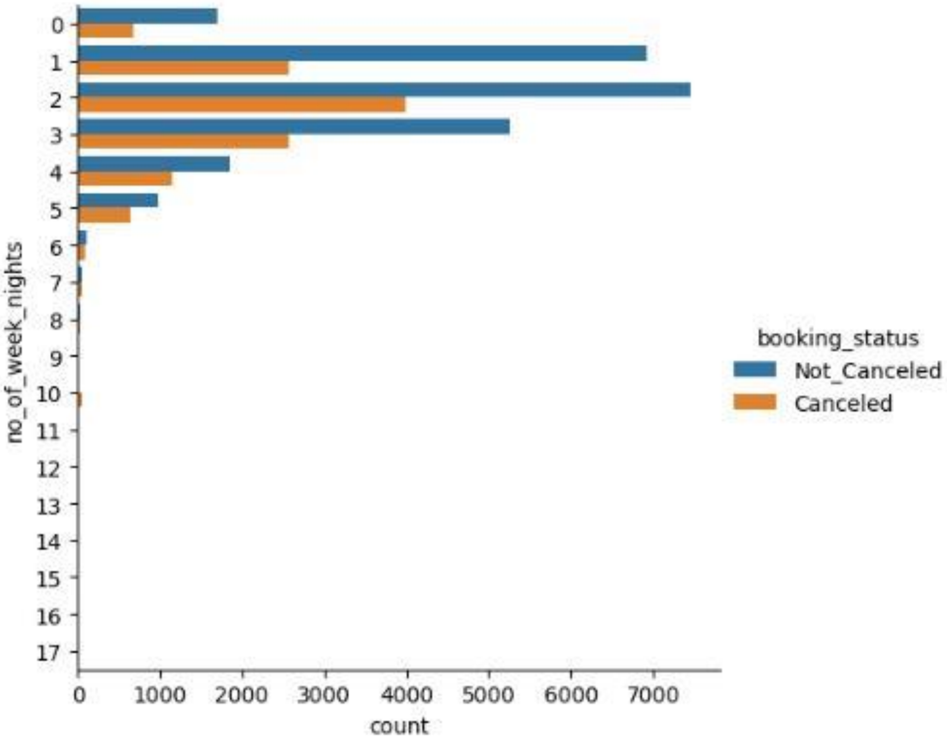
Above Graph shows the Number of Previous Cancellations is around 35,000

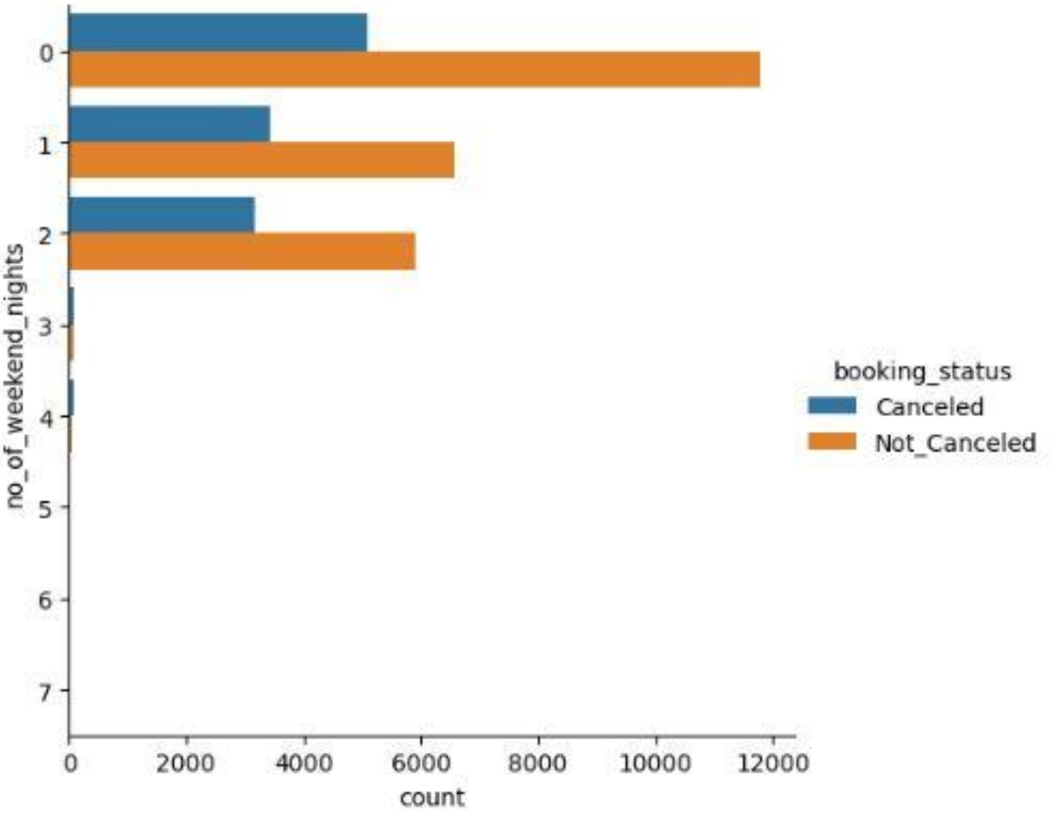


Above Graph shows the Number of Previous Cancellations Not Cancelled is around 35,000

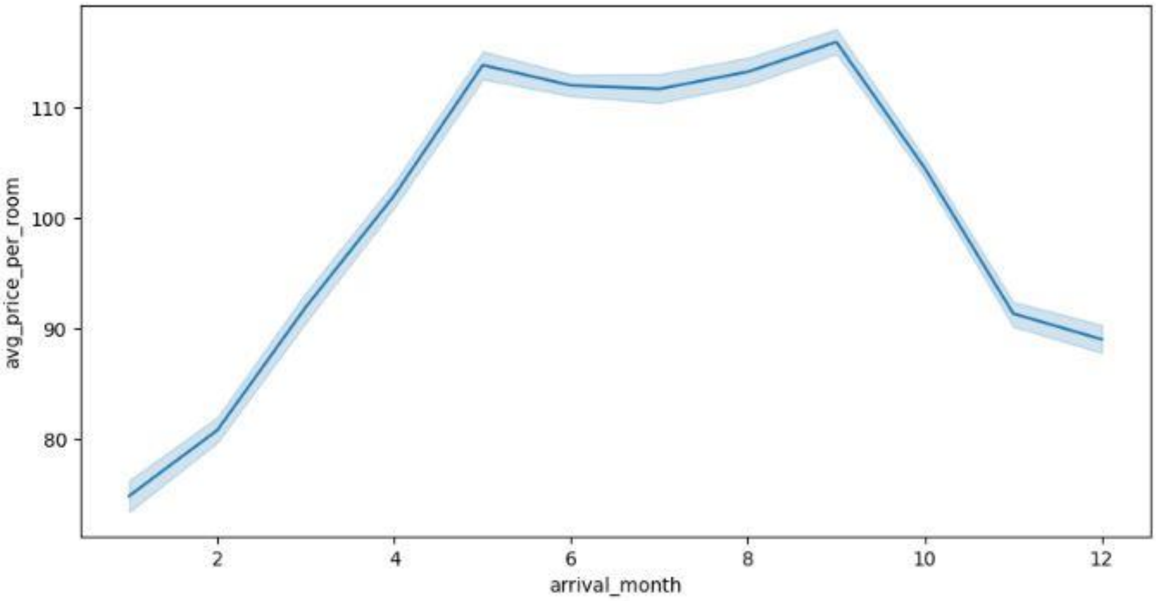
Above Graph shows that Number of Person Spending 2 nights have less chance of cancellation followed by 1 and 3 nights



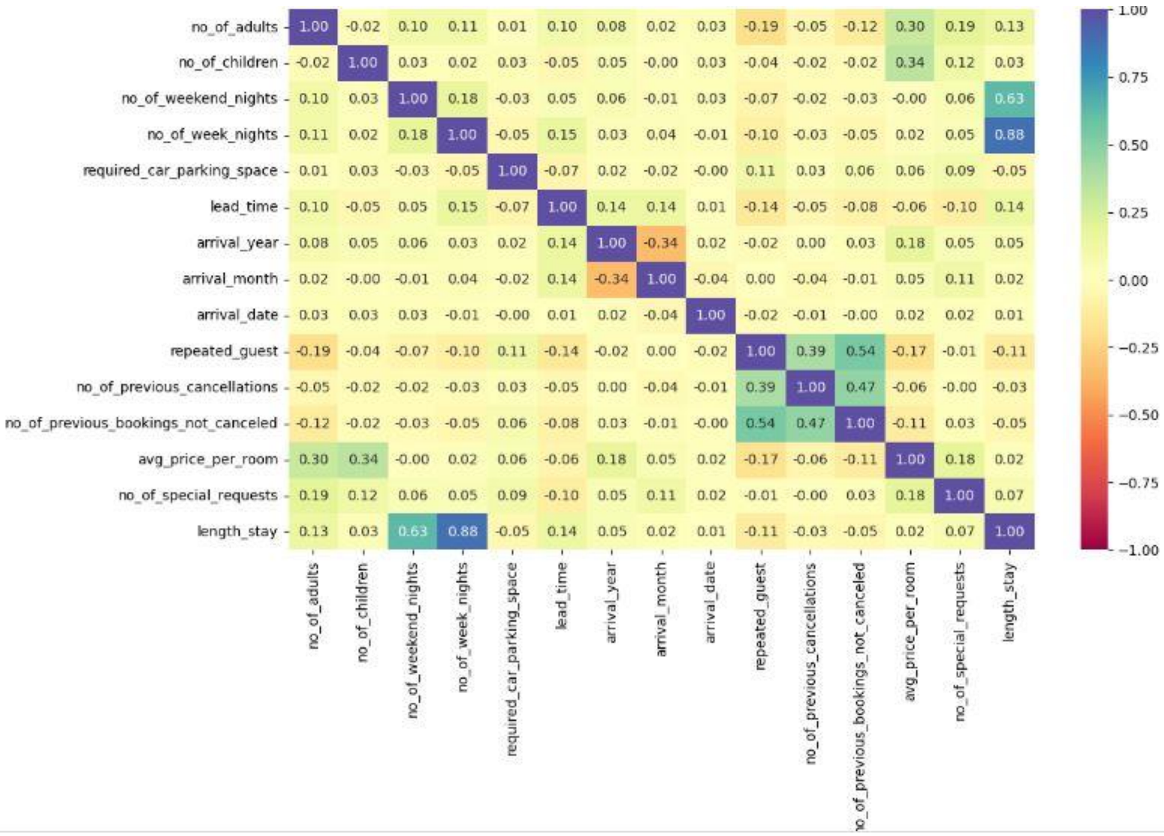




The above graph shows that Customer spending 2 weekend having higher chances of cancellation and spending 1 weekend having lower chance of cancellation.

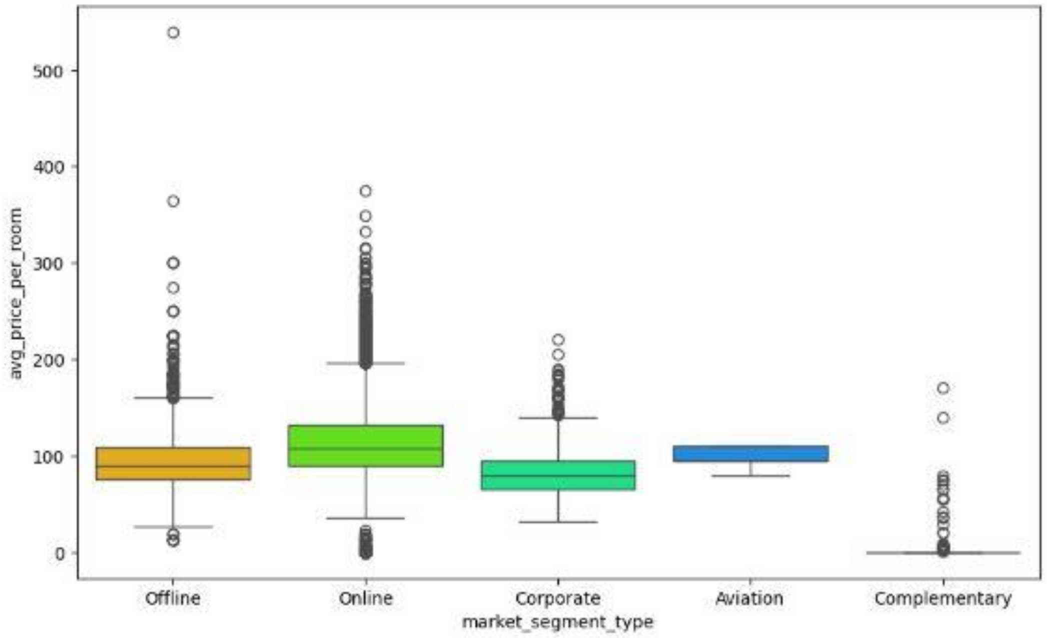


Bivariate Analysis

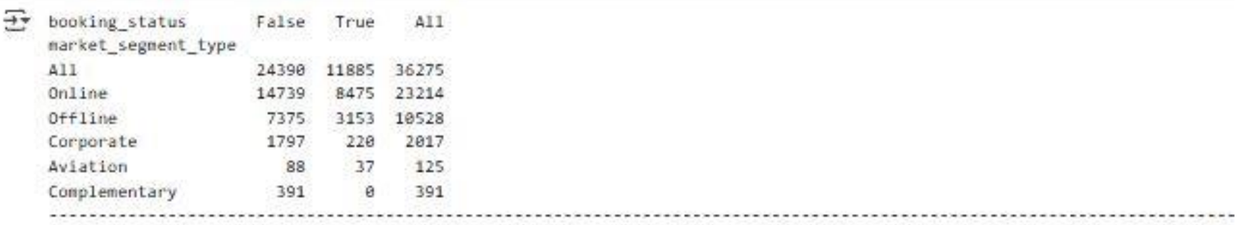


	no_of_adults	no_of_children	no_of_weekend_nights	no_of_week_nights	required_car_parking_space	lead_time	arrival_year	arrival_month	arrival_date	repeated_guest	no_of_previous_cancellations	no_of_previous_bookings_not_canceled	avg_price_per_room	no_of_special_requests	length_stay
no_of_adults	1.00	-0.02	0.10	0.11	0.01	0.10	0.08	0.02	0.03	-0.19	-0.05	-0.12	0.30	0.19	0.13
no_of_children	-0.02	1.00	0.03	0.02	0.03	-0.05	0.05	-0.00	0.03	-0.04	-0.02	-0.02	0.34	0.12	0.03
no_of_weekend_nights	0.10	0.03	1.00	0.18	-0.03	0.05	0.06	-0.01	0.03	-0.07	-0.02	-0.03	-0.00	0.06	0.63
no_of_week_nights	0.11	0.02	0.18	1.00	-0.05	0.15	0.03	0.04	-0.01	-0.10	-0.03	-0.05	0.02	0.05	0.88
required_car_parking_space	0.01	0.03	-0.03	-0.05	1.00	-0.07	0.02	-0.02	-0.00	0.11	0.03	0.06	0.06	0.09	-0.05
lead_time	0.10	-0.05	0.05	0.15	-0.07	1.00	0.14	0.14	0.01	-0.14	-0.05	-0.08	-0.06	-0.10	0.14
arrival_year	0.08	0.05	0.06	0.03	0.02	0.14	1.00	-0.34	0.02	-0.02	0.00	0.03	0.18	0.05	0.05
arrival_month	0.02	-0.00	-0.01	0.04	-0.02	0.14	-0.34	1.00	-0.04	0.00	-0.04	-0.01	0.05	0.11	0.02
arrival_date	0.03	0.03	0.03	-0.01	-0.00	0.01	0.02	-0.04	1.00	-0.02	-0.01	-0.00	0.02	0.02	0.01
repeated_guest	-0.19	-0.04	-0.07	-0.10	0.11	-0.14	-0.02	0.00	-0.02	1.00	0.39	0.54	-0.17	-0.01	-0.11
no_of_previous_cancellations	-0.05	-0.02	-0.02	-0.03	0.03	-0.05	0.00	-0.04	-0.01	0.39	1.00	0.47	-0.06	-0.00	-0.03
no_of_previous_bookings_not_canceled	-0.12	-0.02	-0.03	-0.05	0.06	-0.08	0.03	-0.01	-0.00	0.54	0.47	1.00	-0.11	0.03	-0.05
avg_price_per_room	0.30	0.34	-0.00	0.02	0.06	-0.06	0.18	0.05	0.02	-0.17	-0.06	-0.11	1.00	0.18	0.02
no_of_special_requests	0.19	0.12	0.06	0.05	0.09	-0.10	0.05	0.11	0.02	-0.01	-0.00	0.03	0.18	1.00	0.07
length_stay	0.13	0.03	0.63	0.88	-0.05	0.14	0.05	0.02	0.01	-0.11	-0.03	-0.05	0.02	0.07	1.00

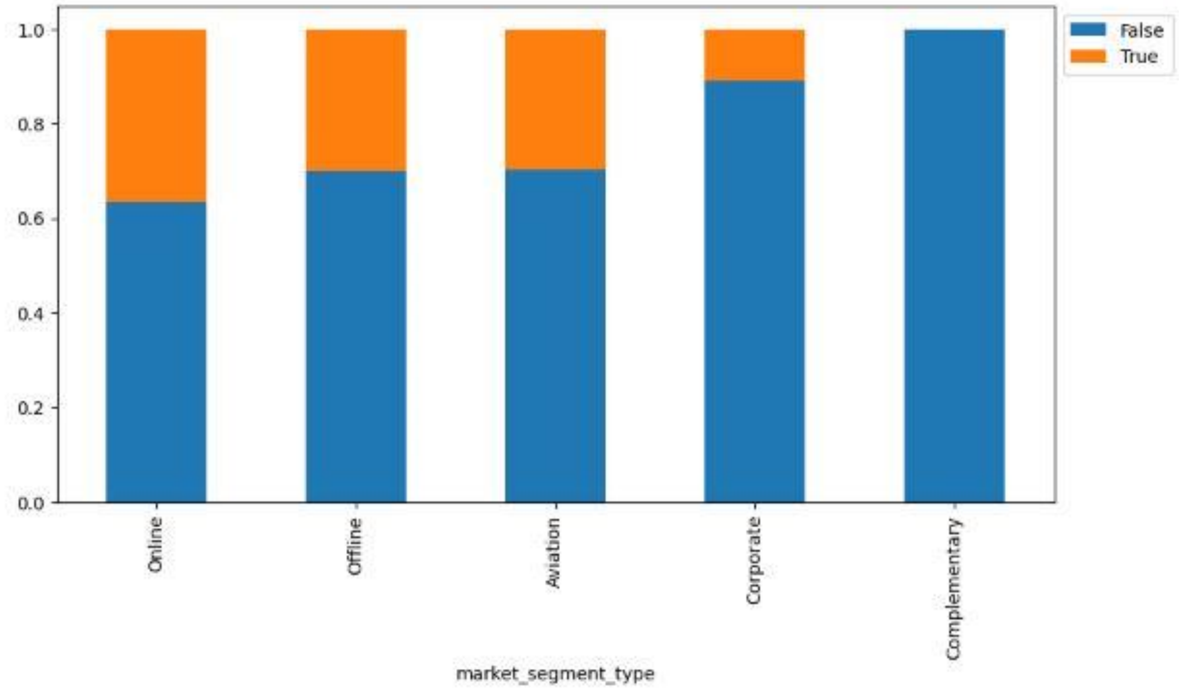
No quantitative variables are correlated with each other besides repeated guests and number of previous bookings not cancelled, which may indicate that there is some brand loyalty when it comes to repeated guests



As hotel prices are very demand heavy as mentioned, online seems to have the highest median price, since online is where the most booking takes place

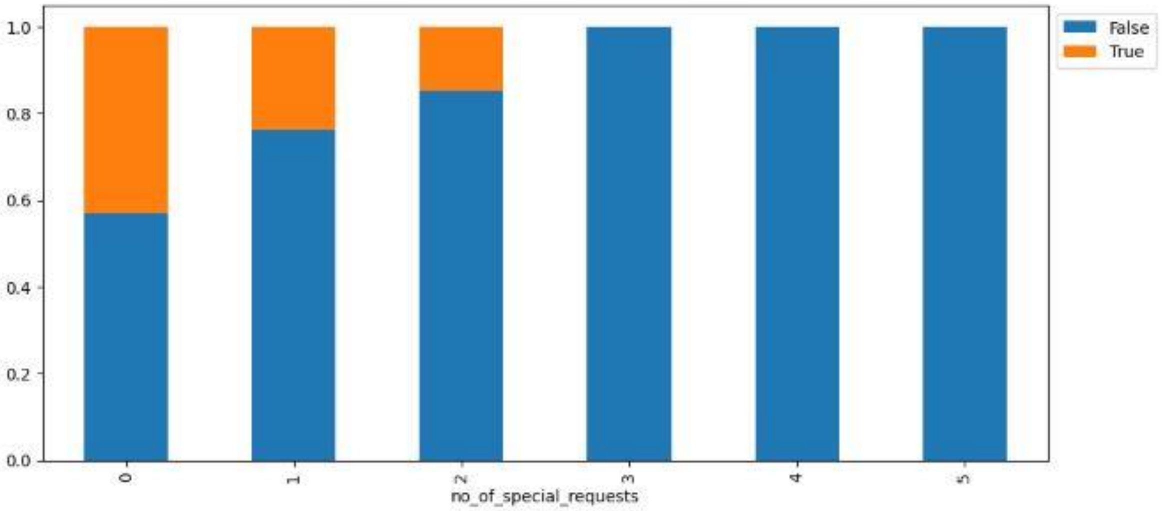


```
booking_status      False  True    All
market_segment_type
All                 24390  11885  36275
Online              14739   8475  23214
Offline              7375   3153  10528
Corporate            1797    220   2017
Aviation               88     37    125
Complementary         391      0    391
------------------------------------------------------------------------------------------------
```

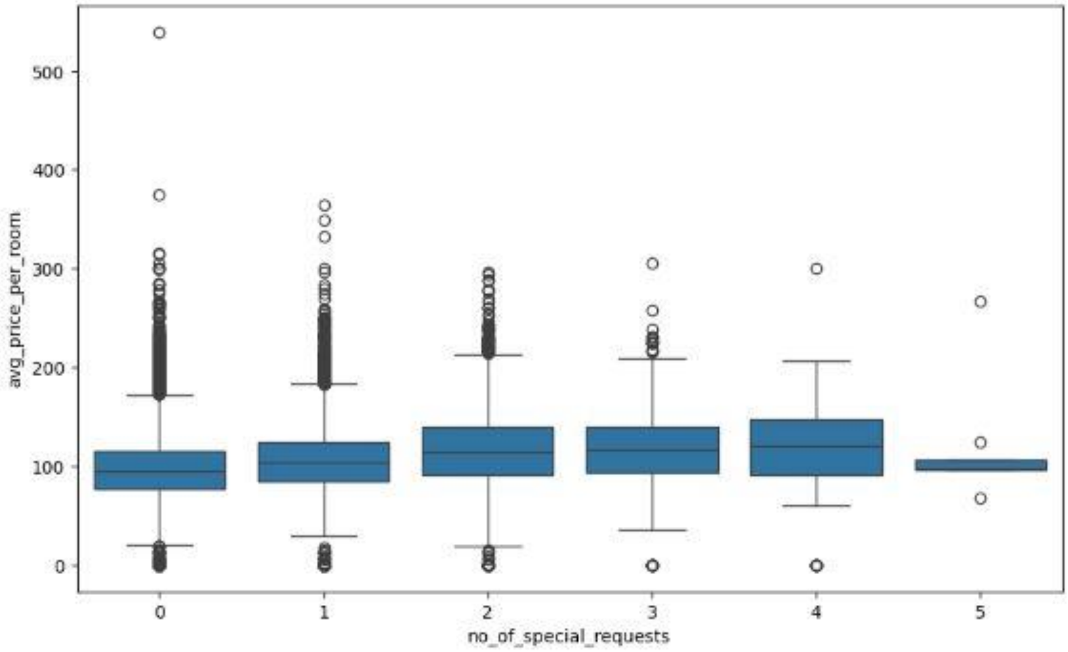


Most of the cancellations when customers book the hotel online and almost no cancellations happen when it is complementary

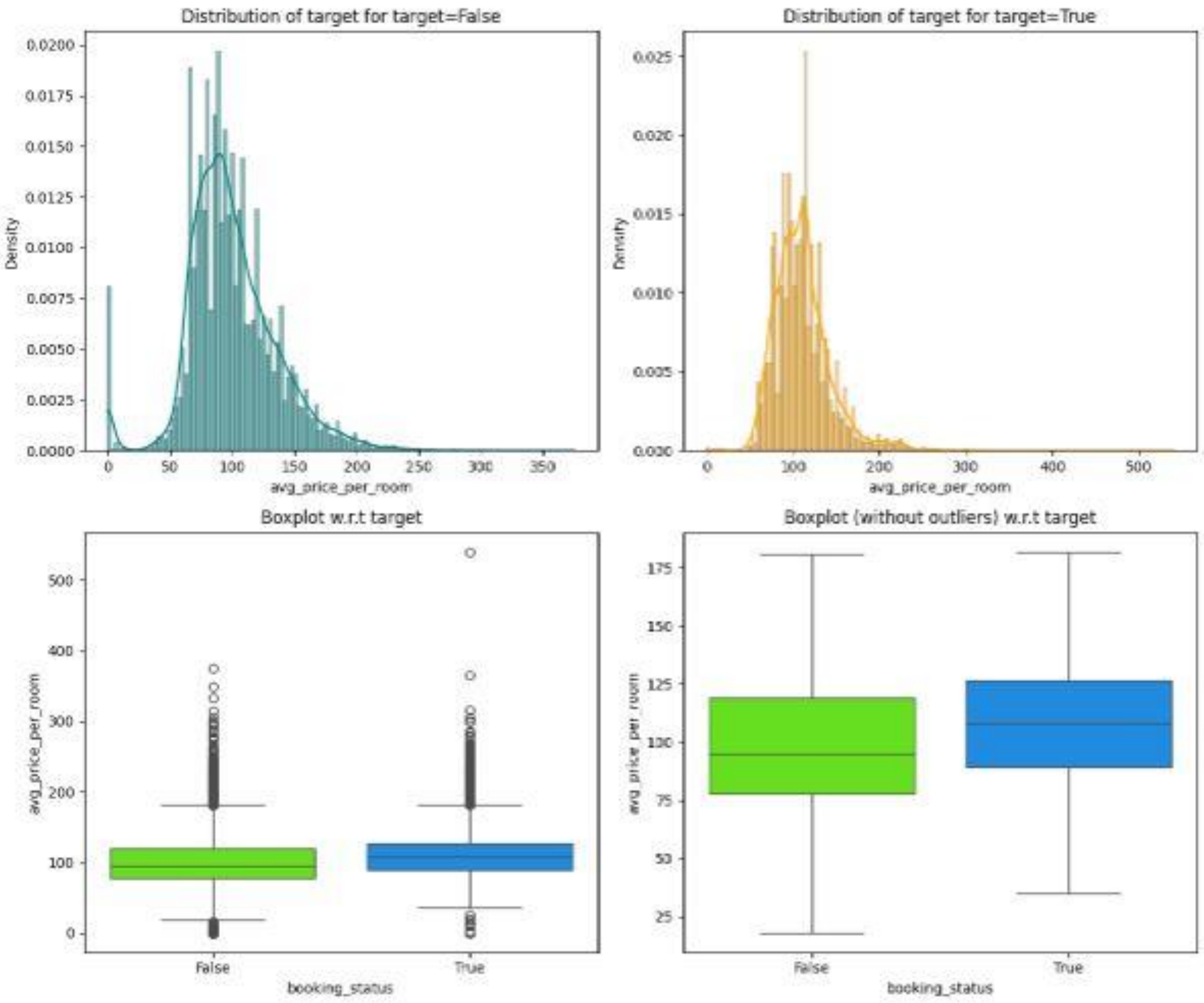
```
booking_status           False   True    All
no_of_special_requests
All                      24390  11885  36275
0                        11232   8545  19777
1                         8670   2703  11373
2                         3727    637   4364
3                          675      0    675
4                           78      0     78
5                            8      0      8
------------------------------------------------------------------------------------------------------------------------
```

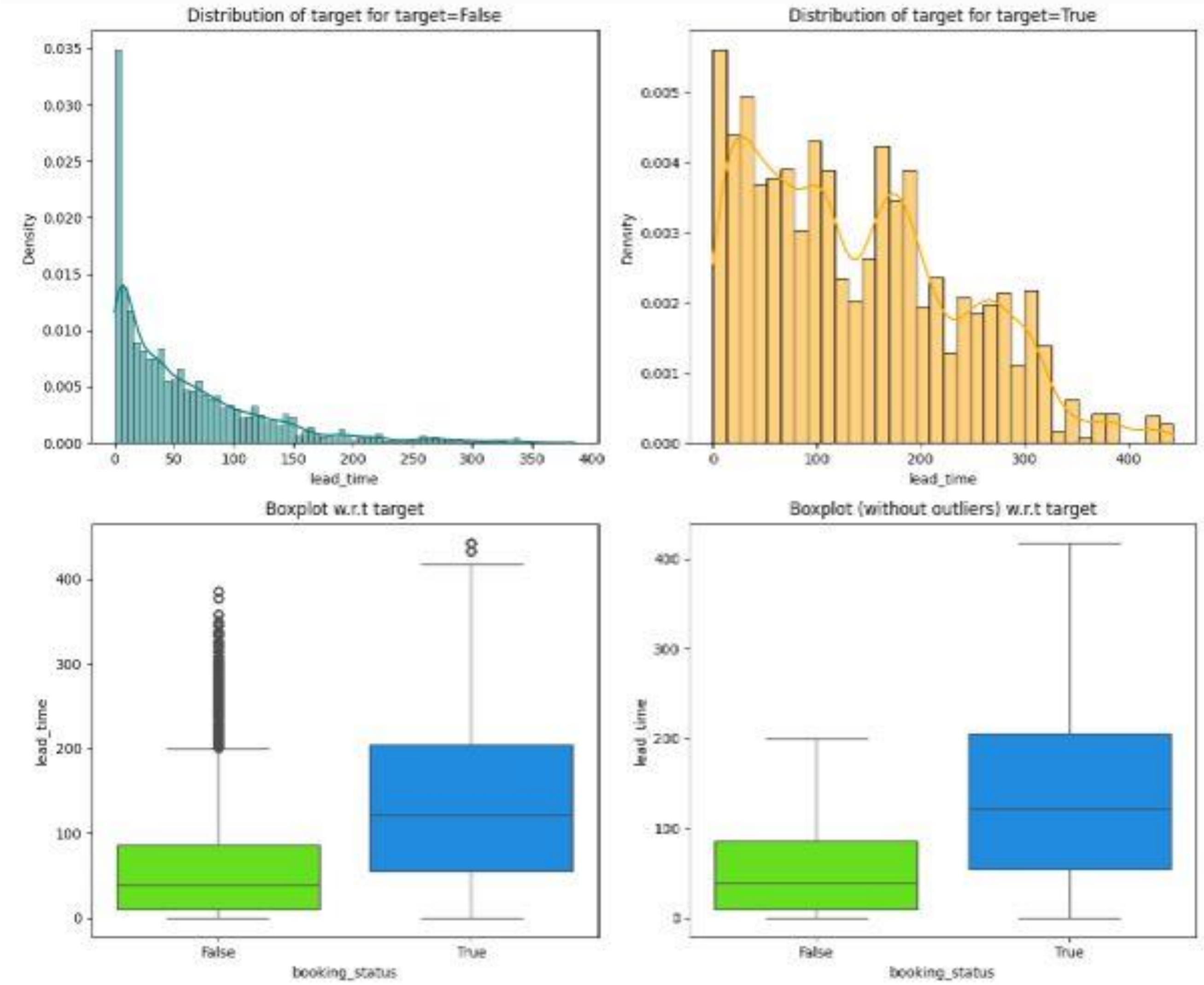


When there is no special requests, you can see that there are more more cancellations. When there is more special requests, there is less likelihood of cancellation

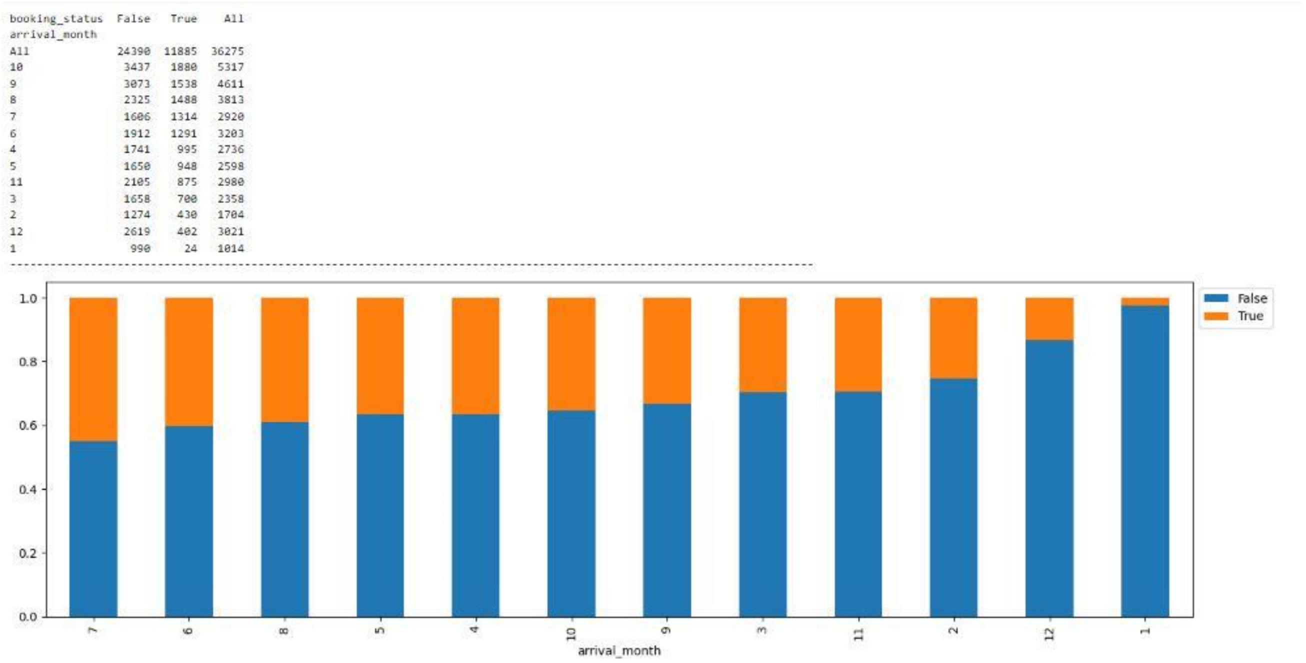


The price of the room does not seem to vary too much if there is little requests given, but when there is about 4 or more special requests the price seems to suffer variation

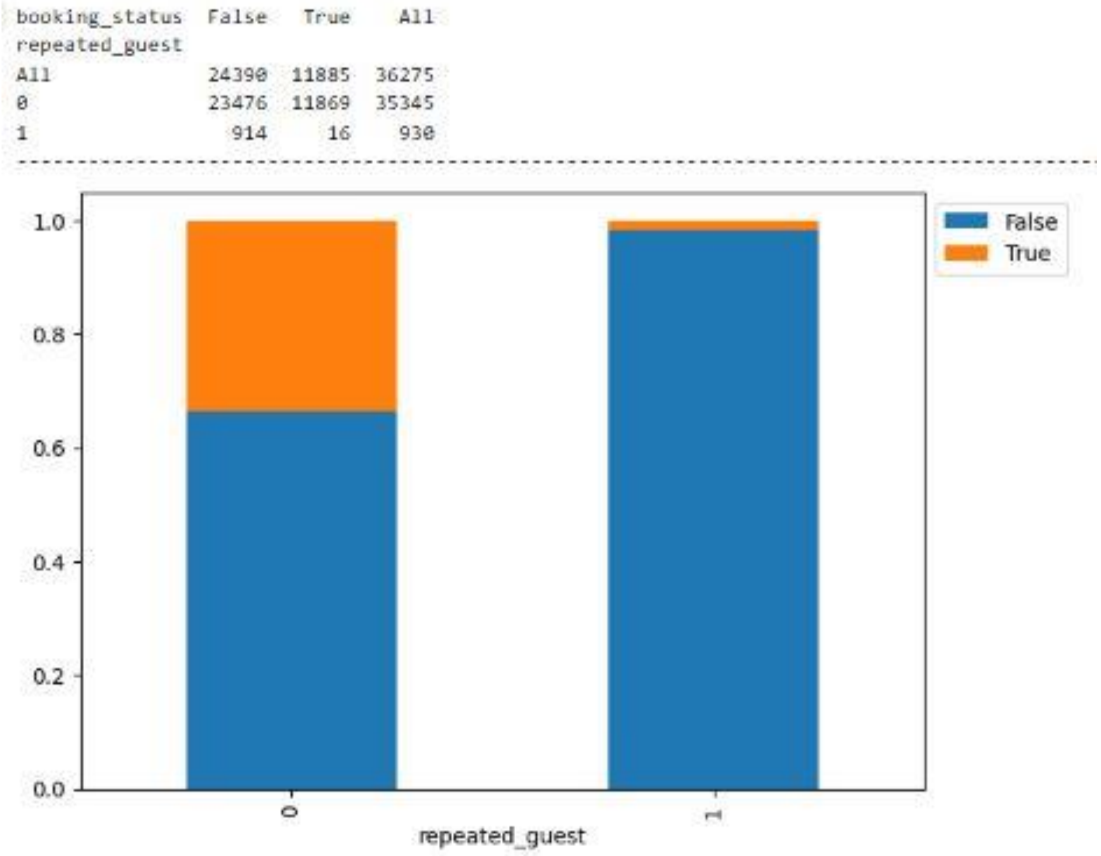

Distribution of target for target=False
Density
0.0000
0.0025
0.0050
0.0075
0.0100
0.0125
0.0150
0.0175
0.0200
0
50
100
150
200
250
300
350
avg_price_per_room
Distribution of target for target=True
Density
0.000
0.005
0.010
0.015
0.020
0.025
0
100
200
300
400
500
avg_price_per_room
Boxplot w.r.t target
avg_price_per_room
0
100
200
300
400
500
False
True
booking_status
Boxplot (without outliers) w.r.t target
avg_price_per_room
25
50
75
100
125
150
175
False
True
booking_status



Generally people travel with their spouse and children for vacations or other activities. Let's create a new dataframe of the customers who traveled with their families and analyze the impact on booking status

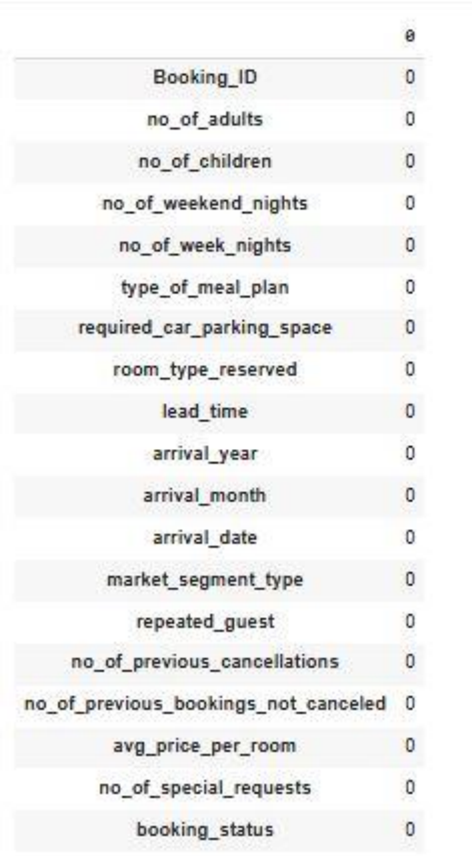


```
booking_status  False   True    All
arrival_month
All             24390  11885  36275
10               3437   1880   5317
9                3073   1538   4611
8                2325   1488   3813
7                1606   1314   2920
6                1912   1291   3203
4                1741    995   2736
5                1650    948   2598
11               2105    875   2980
3                1658    700   2358
2                1274    430   1704
12               2619    402   3021
1                 990     24   1014
------------------------------------------------------------------------------------------------------------------------
```



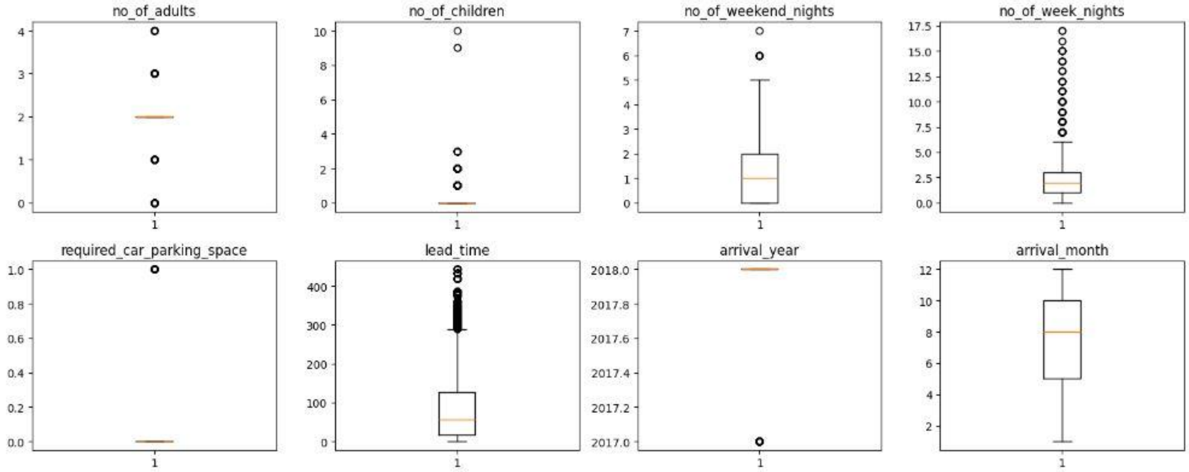
```
booking_status  False   True    All
repeated_guest
All             24390  11885  36275
0               23476  11869  35345
1                 914     16    930
------------------------------------------------------------------------------------------------------------------------
```

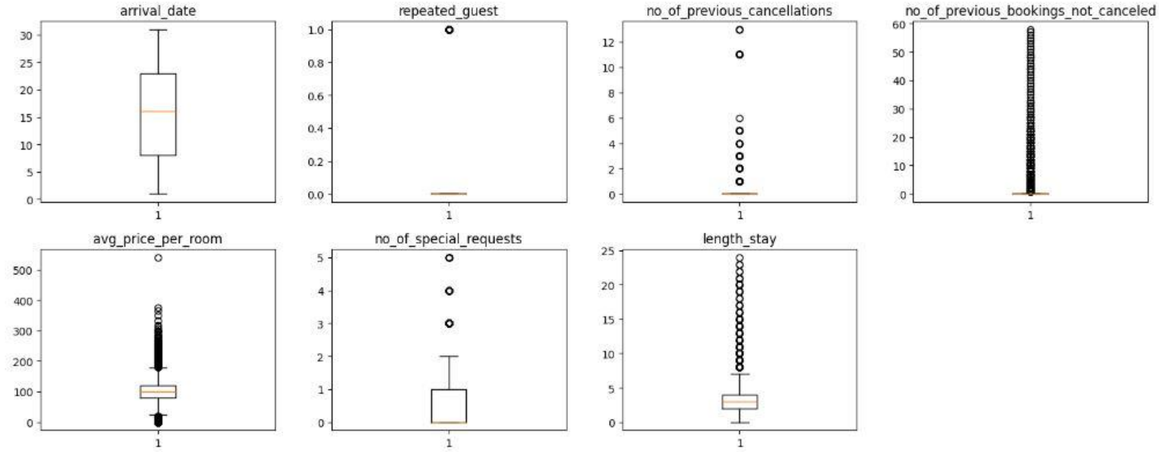
Data Preprocessing



	0
Booking_ID	0
no_of_adults	0
no_of_children	0
no_of_weekend_nights	0
no_of_week_nights	0
type_of_meal_plan	0
required_car_parking_space	0
room_type_reserved	0
lead_time	0
arrival_year	0
arrival_month	0
arrival_date	0
market_segment_type	0
repeated_guest	0
no_of_previous_cancellations	0
no_of_previous_bookings_not_canceled	0
avg_price_per_room	0
no_of_special_requests	0
booking_status	0

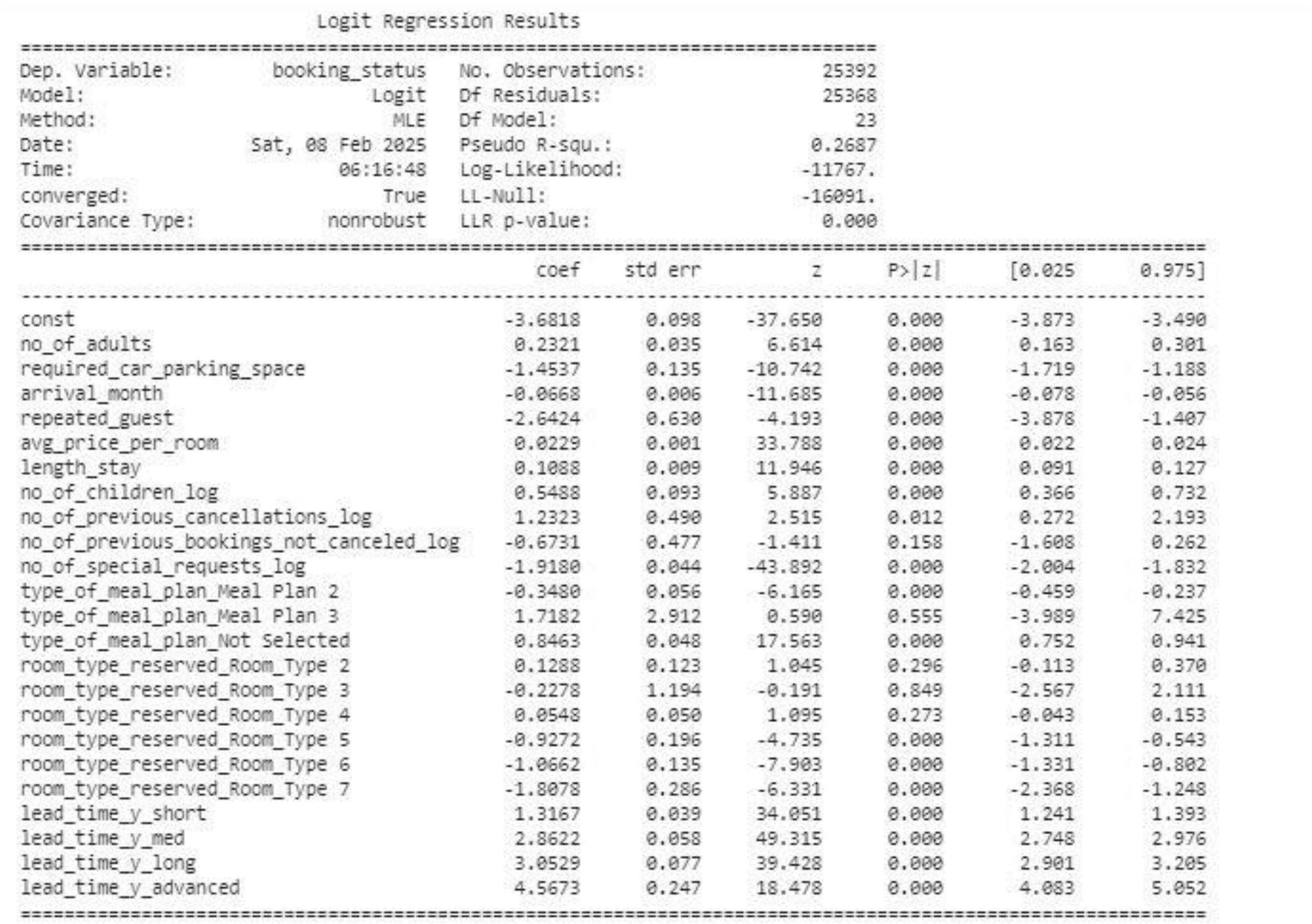
There are No Missing Values in Data





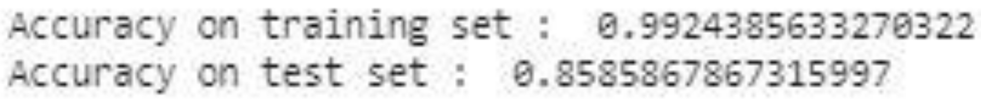
Model Building

Logistic Regression model



```
                           Logit Regression Results
==============================================================================
Dep. Variable:         booking_status   No. Observations:                25392
Model:                          Logit   Df Residuals:                    25368
Method:                           MLE   Df Model:                           23
Date:                Sat, 08 Feb 2025   Pseudo R-squ.:                  0.2687
Time:                        06:16:48   Log-Likelihood:                -11767.
converged:                       True   LL-Null:                       -16091.
Covariance Type:            nonrobust   LLR p-value:                     0.000
====================================================================================================
                                              coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------------------
const                                      -3.6818      0.098    -37.650      0.000      -3.873      -3.490
no_of_adults                                0.2321      0.035      6.614      0.000       0.163       0.301
required_car_parking_space                 -1.4537      0.135    -10.742      0.000      -1.719      -1.188
arrival_month                              -0.0668      0.006    -11.685      0.000      -0.078      -0.056
repeated_guest                             -2.6424      0.630     -4.193      0.000      -3.878      -1.407
avg_price_per_room                          0.0229      0.001     33.788      0.000       0.022       0.024
length_stay                                 0.1088      0.009     11.946      0.000       0.091       0.127
no_of_children_log                          0.5488      0.093      5.887      0.000       0.366       0.732
no_of_previous_cancellations_log            1.2323      0.490      2.515      0.012       0.272       2.193
no_of_previous_bookings_not_canceled_log   -0.6731      0.477     -1.411      0.158      -1.608       0.262
no_of_special_requests_log                 -1.9180      0.044    -43.892      0.000      -2.004      -1.832
type_of_meal_plan_Meal Plan 2              -0.3480      0.056     -6.165      0.000      -0.459      -0.237
type_of_meal_plan_Meal Plan 3               1.7182      2.912      0.590      0.555      -3.989       7.425
type_of_meal_plan_Not Selected              0.8463      0.048     17.563      0.000       0.752       0.941
room_type_reserved_Room_Type 2              0.1288      0.123      1.045      0.296      -0.113       0.370
room_type_reserved_Room_Type 3             -0.2278      1.194     -0.191      0.849      -2.567       2.111
room_type_reserved_Room_Type 4              0.0548      0.050      1.095      0.273      -0.043       0.153
room_type_reserved_Room_Type 5             -0.9272      0.196     -4.735      0.000      -1.311      -0.543
room_type_reserved_Room_Type 6             -1.0662      0.135     -7.903      0.000      -1.331      -0.802
room_type_reserved_Room_Type 7             -1.8078      0.286     -6.331      0.000      -2.368      -1.248
lead_time_y_short                           1.3167      0.039     34.051      0.000       1.241       1.393
lead_time_y_med                             2.8622      0.058     49.315      0.000       2.748       2.976
lead_time_y_long                            3.0529      0.077     39.428      0.000       2.901       3.205
lead_time_y_advanced                        4.5673      0.247     18.478      0.000       4.083       5.052
====================================================================================================
```

Accuracy of Training and Test set



```
Accuracy on training set :  0.9924385633270322
Accuracy on test set :  0.8585867867315997
```

- Negative values of the coefficient show that the probability of customers cancelling the booking decreases with the increase of the corresponding attribute value.
- Positive values of the coefficient show that the probability of customer cancelling increases with the increase of corresponding attribute value.
- p-value of a variable indicates if the variable is significant or not. If we consider the significance level to be 0.05 (5%), then any variable with a p-value less than 0.05 would be considered significant.
- But these variables might contain multicollinearity, which will affect the p-values.
- We will have to remove multicollinearity from the data to get reliable coefficients and p-values.
- There are different ways of detecting (or testing) multi-collinearity, one such way is the Variation Inflation Factor.

Model evaluation criterion

Model can make wrong predictions as:

1. Predicting a customer will not cancel their booking but in reality, the customer will cancel their booking.

2. Predicting a customer will cancel their booking but in reality, the customer will not cancel their booking.

Which case is more important?

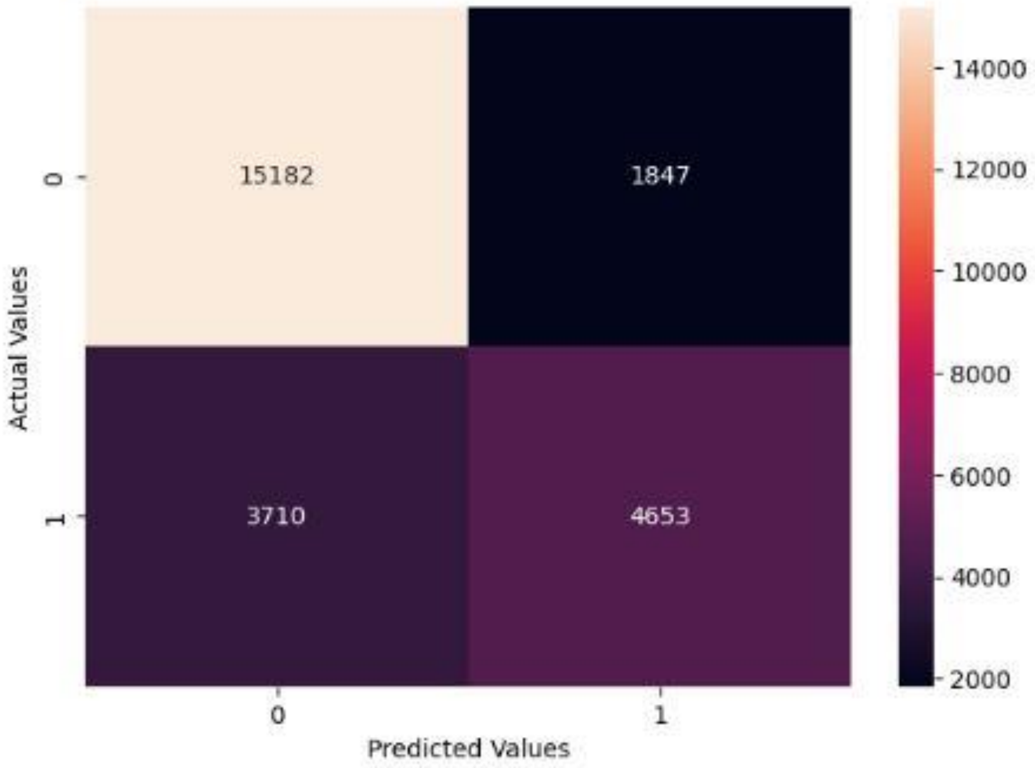
Both the cases are important as:

- If we predict that a booking will not be cancelled and the booking gets cancelled then the hotel will lose resources and will have to bear additional costs of distribution channels.
- If we predict that a booking will get cancelled and the booking doesn't get cancelled the hotel might not be able to provide satisfactory services to the customer by assuming that this booking will be cancelled. This might damage the Brand Quality.

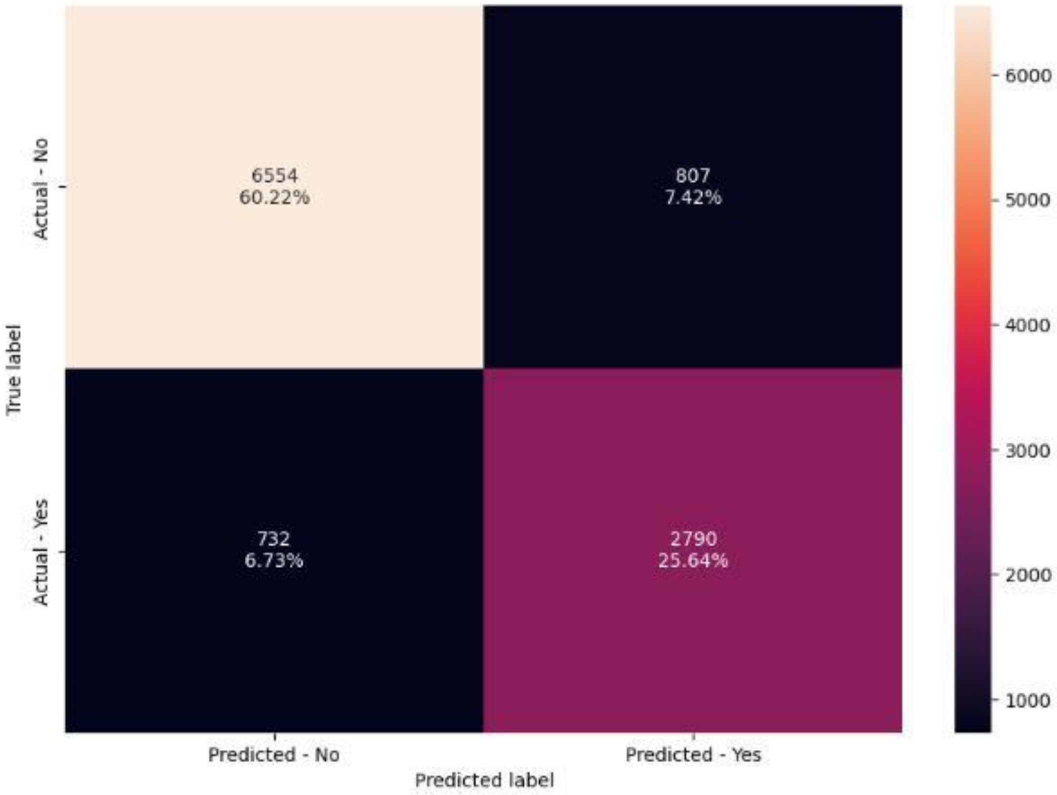
Model Performance Evaluation

Confusion Matrix

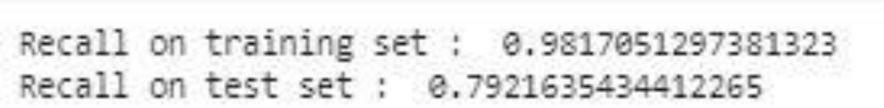
Checking Performance of Training Set



Checking Performance of Test Set



Recall on Training and Test set

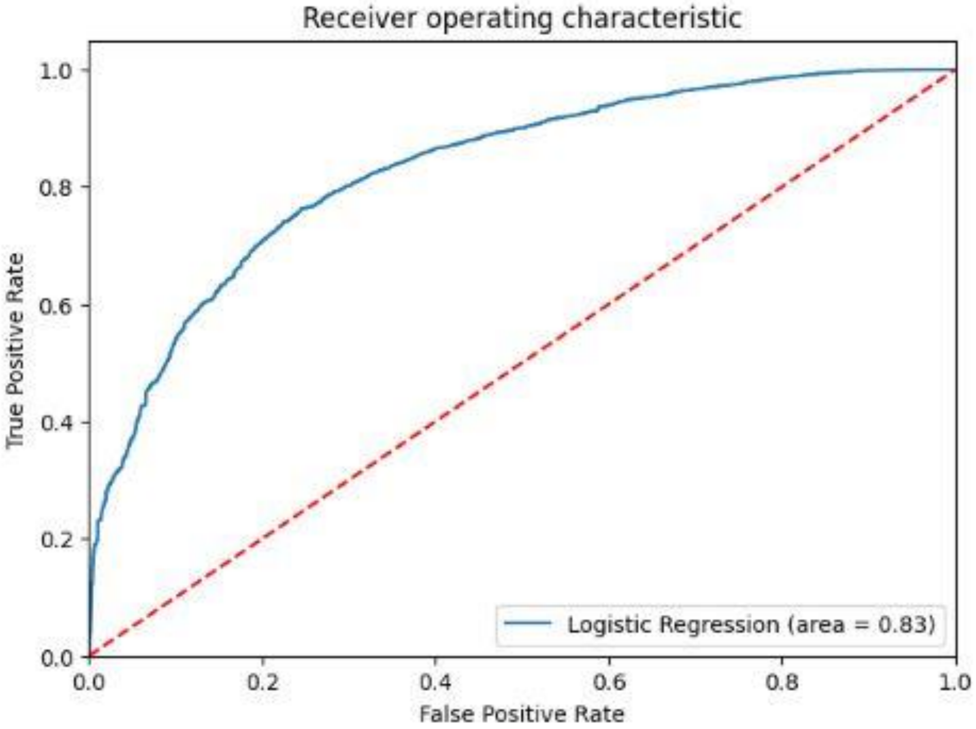


```
Recall on training set :  0.9817051297381323
Recall on test set :  0.7921635434412265
```

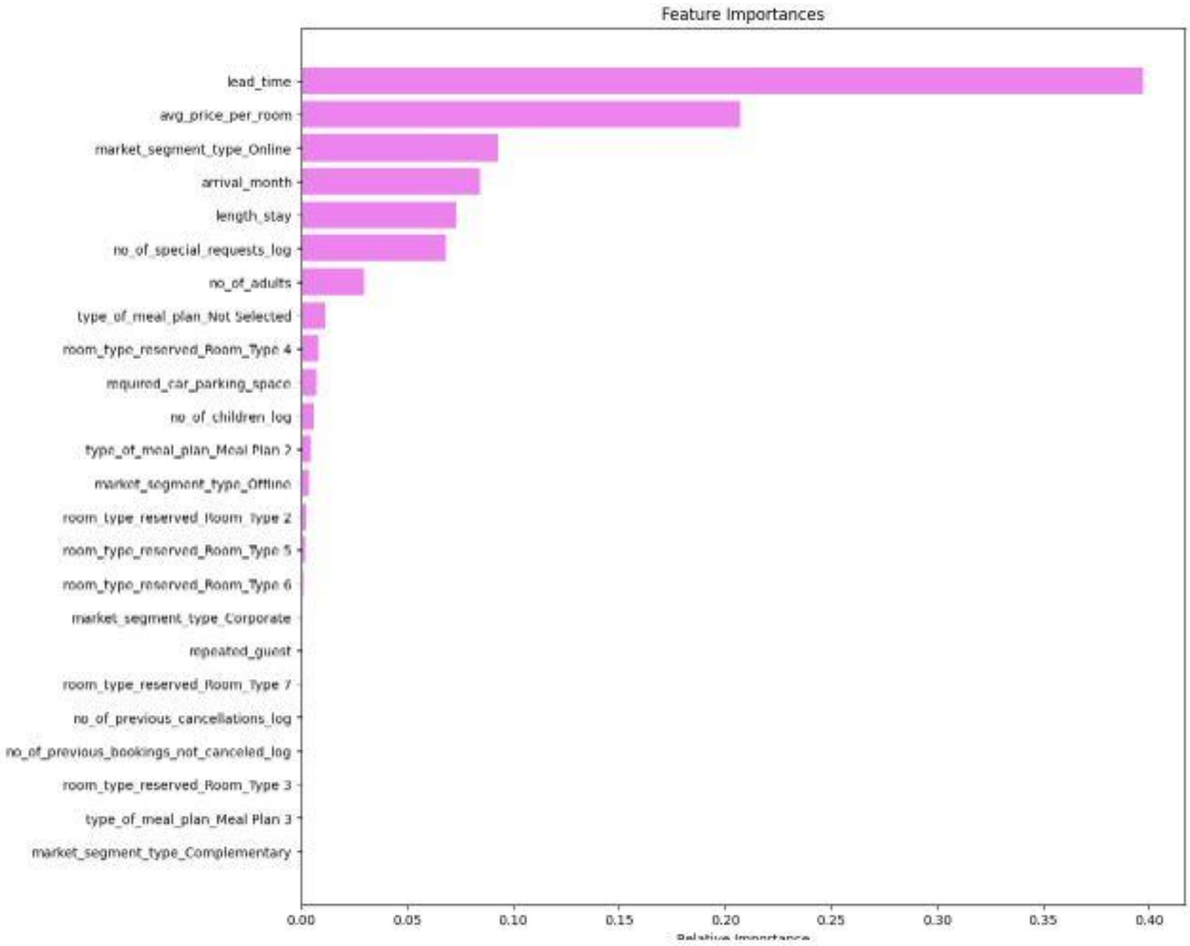
Coefficient interpretations

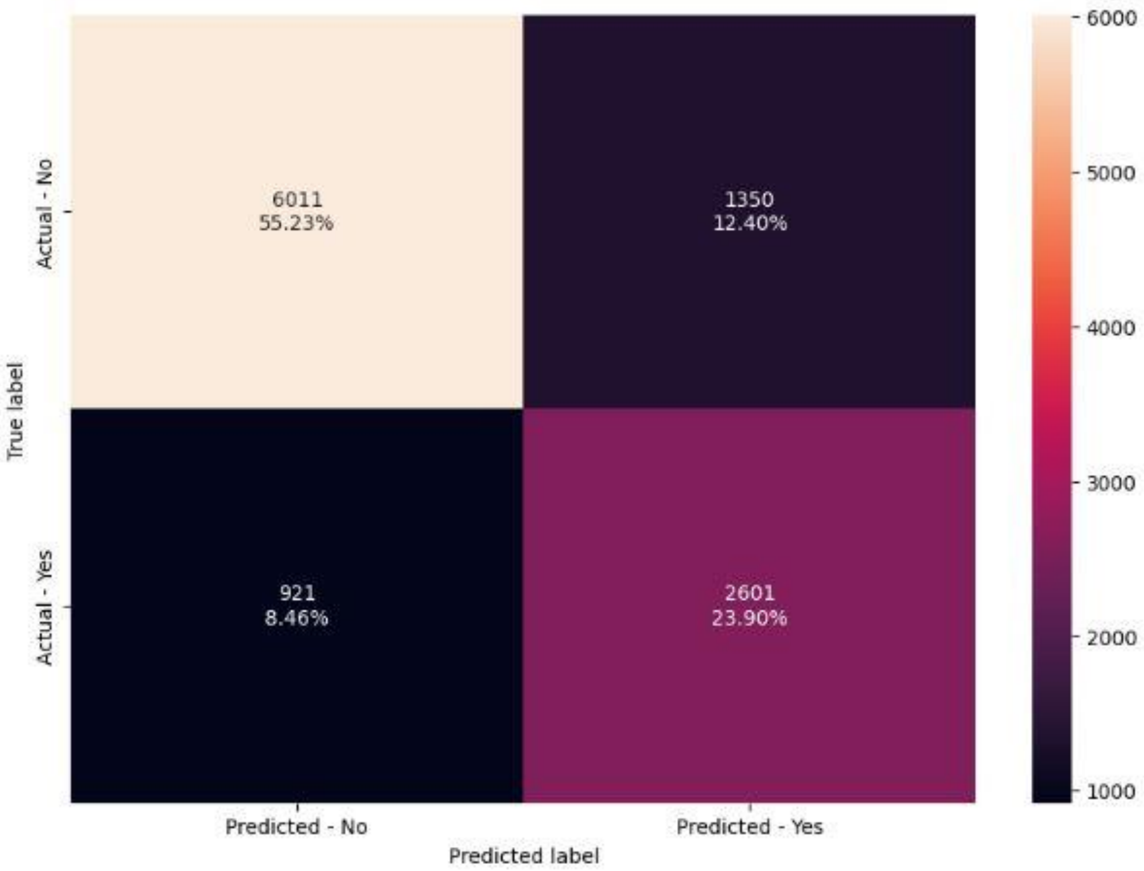
- no of adults: Holding all other features constant a 1 unit change in the number of children will increase the odds of a booking getting cancelled by 1.11 times or a 11.49% increase in the odds of a booking getting cancelled.
- no of children: Holding all other features constant a 1 unit change in the number of children will increase the odds of a booking getting cancelled by 1.16 times or a 16.54% increase in the odds of a booking getting cancelled.
- no of weekend nights: Holding all other features constant a 1 unit change in the number of weeknights a customer stays at the hotel will increase the odds of a booking getting cancelled by 1.11 times or a 11.46% increase in the odds of a booking getting cancelled.
- no of week nights: Holding all other features constant a 1 unit change in the number of weeknights a customer stays at the hotel will increase the odds of a booking getting cancelled by 1.04 times or a 4.25% increase in the odds of a booking getting cancelled.
- required car parking space: The odds of a customer who requires a car parking space are 0.2 times less than a customer who doesn't require a car parking space or a 79.70% fewer odds of a customer canceling their booking.
- lead time: Holding all other features constant a 1 unit change in the lead time will increase the odds of a booking getting cancelled by 1.01 times or a 1.58% increase in the odds of a booking getting cancelled.
- no of special requests: Holding all other features constant a 1 unit change in the number of special requests made by the customer will decrease the odds of a booking getting cancelled by 0.22 times or a 77% decrease in the odds of a booking getting cancelled.

ROC-AUC (Test Set)

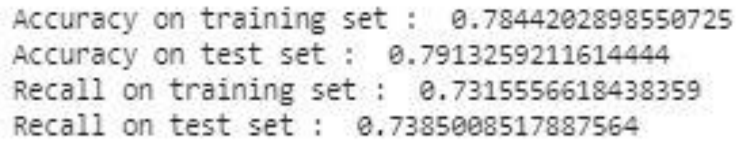


Building a Decision Tree model


Feature Importances
lead_time
avg_price_per_room
market_segment_type_Online
arrival_month
length_stay
no_of_special_requests_log
no_of_adults
type_of_meal_plan_Not Selected
room_type_reserved_Room_Type 4
required_car_parking_space
no_of_children_log
type_of_meal_plan_Meal Plan 2
market_segment_type_Offline
room_type_reserved_Room_Type 2
room_type_reserved_Room_Type 5
room_type_reserved_Room_Type 6
market_segment_type_Corporate
repeated_guest
room_type_reserved_Room_Type 7
no_of_previous_cancellations_log
no_of_previous_bookings_not_canceled_log
room_type_reserved_Room_Type 3
type_of_meal_plan_Meal Plan 3
market_segment_type_Complementary
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
Relative Importance



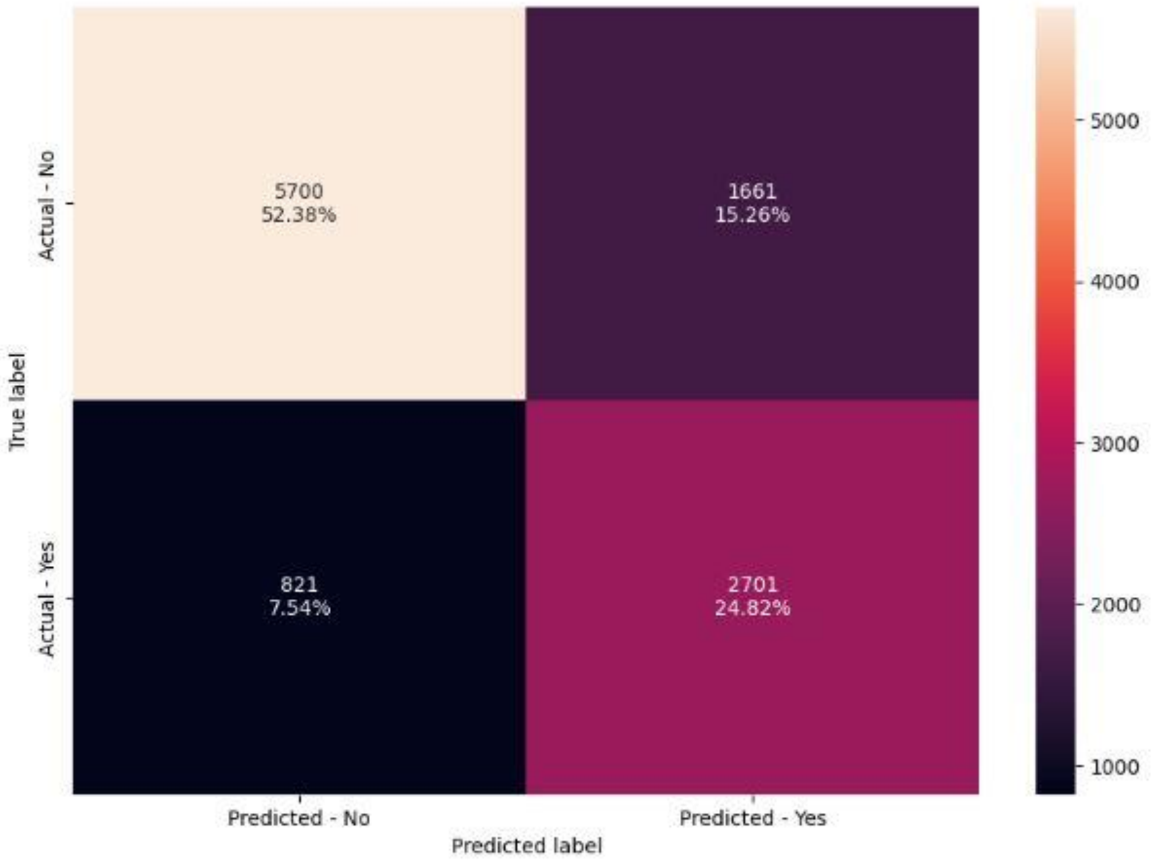
Accuracy and recall on Training set and Test Set



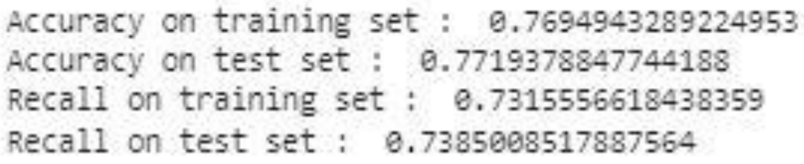
```
Accuracy on training set :  0.7844202898550725
Accuracy on test set :  0.7913259211614444
Recall on training set :  0.7315556618438359
Recall on test set :  0.7385008517887564
```

Lead time is the most important feature followed by average price per room

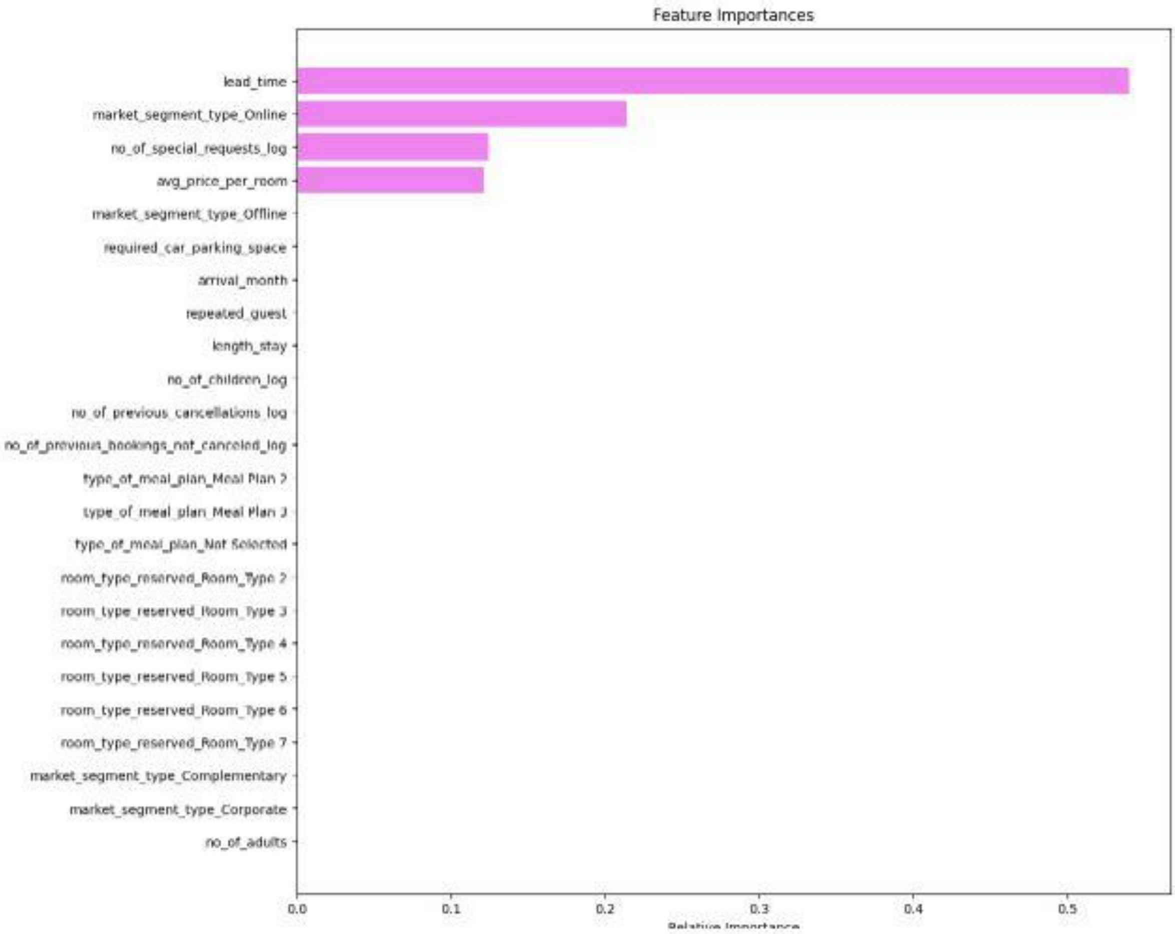
Now Lets prune the tree to see if we can reduce the complexity

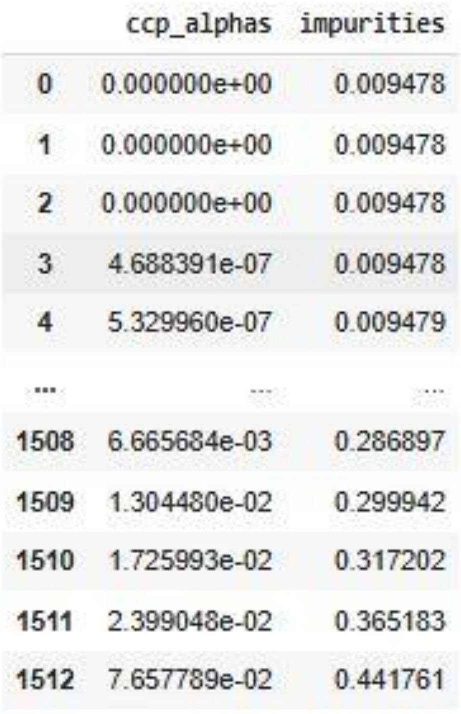


Accuracy and Recall on Training Set and Test Set

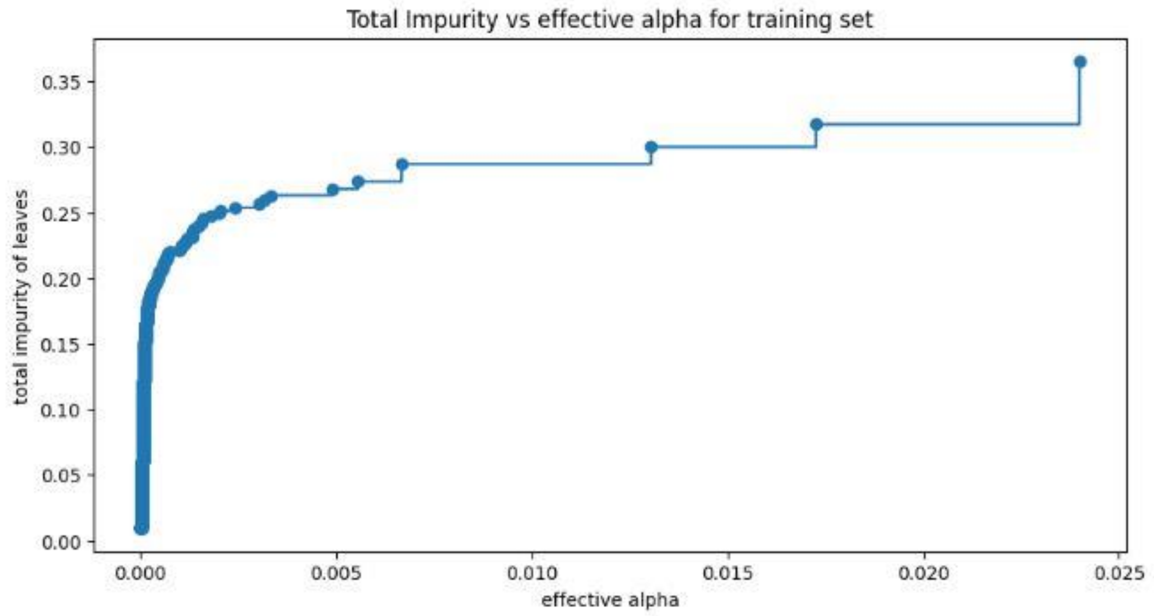


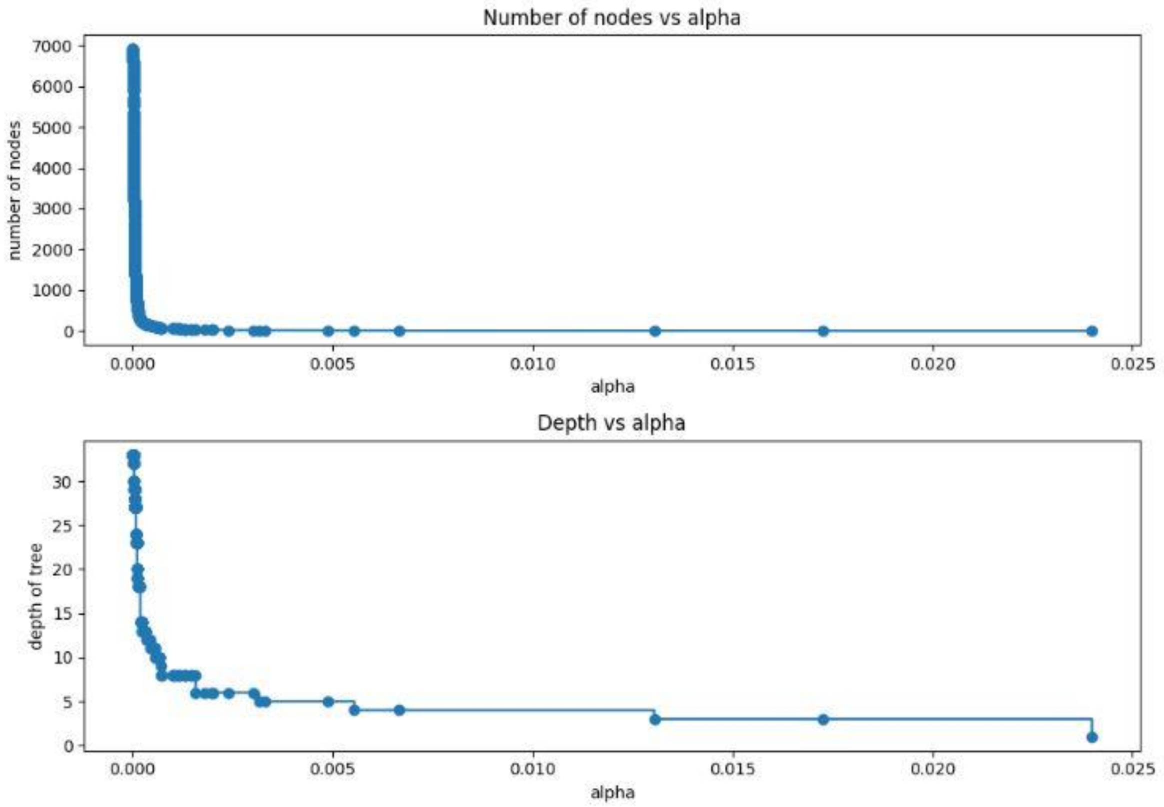
```
Accuracy on training set :  0.7694943289224953
Accuracy on test set :  0.7719378847744188
Recall on training set :  0.7315556618438359
Recall on test set :  0.7385008517887564
```

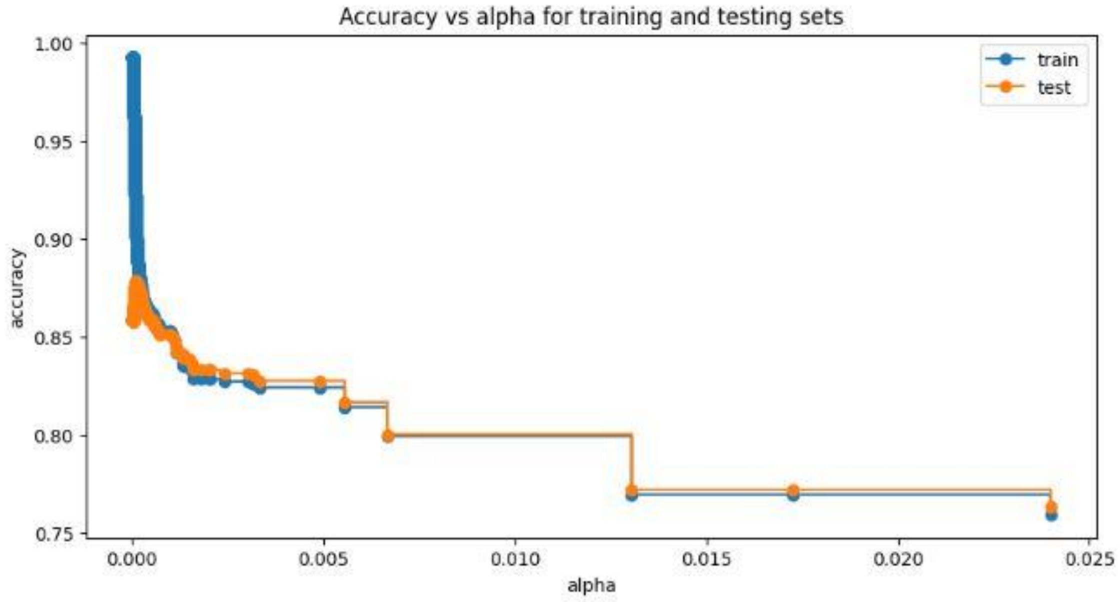


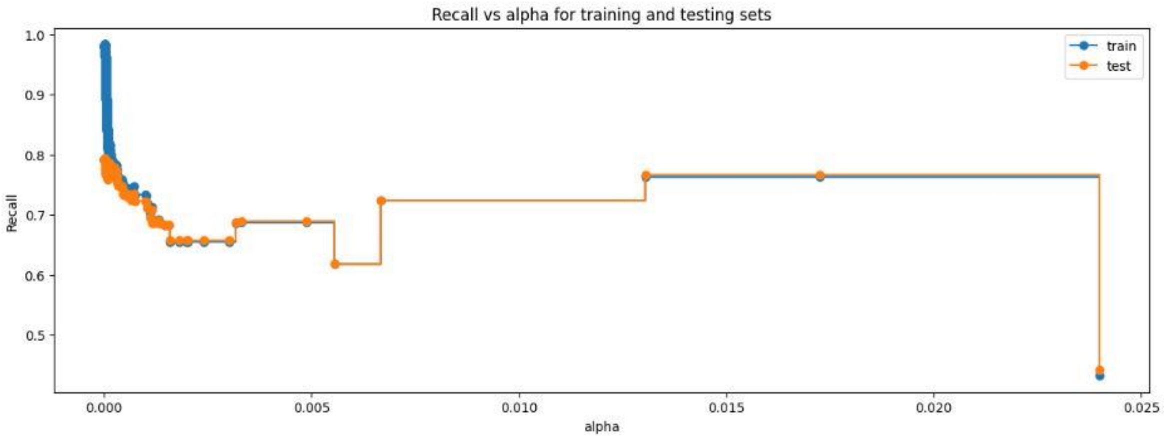


	ccp_alphas	impurities
0	0.000000e+00	0.009478
1	0.000000e+00	0.009478
2	0.000000e+00	0.009478
3	4.688391e-07	0.009478
4	5.329960e-07	0.009479
...	...	...
1508	6.665684e-03	0.286897
1509	1.304480e-02	0.299942
1510	1.725993e-02	0.317202
1511	2.399048e-02	0.365183
1512	7.657789e-02	0.441761


Total Impurity vs effective alpha for training set
total impurity of leaves
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.000
0.005
0.010
0.015
0.020
0.025
effective alpha


Number of nodes vs alpha
number of nodes
0
1000
2000
3000
4000
5000
6000
7000
0.000
0.005
0.010
0.015
0.020
0.025
alpha
Depth vs alpha
depth of tree
0
5
10
15
20
25
30
0.000
0.005
0.010
0.015
0.020
0.025
alpha





Observations from decision tree

- We can see that the tree has become simpler and the rules of the trees are readable.
- The model performance of the model has been generalized.

We observe that the most important features are:

- Lead Time
- Market Segment-Online
- Number of special requests
- Average price per room

The rules obtained from the decision tree can be interpreted as:

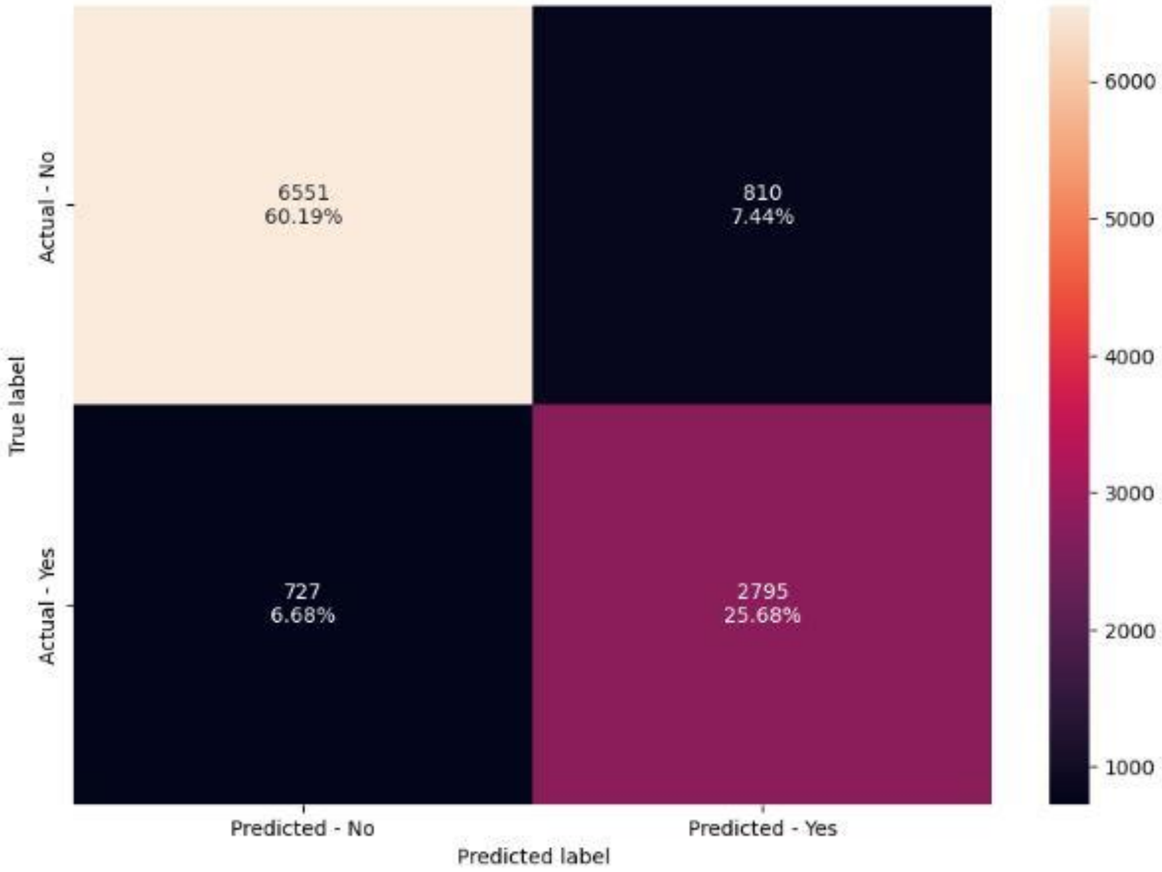
- The rules show that lead time plays a key role in identifying if a booking will be cancelled or not. 151 days has been considered as a threshold value by the model to make the first split.

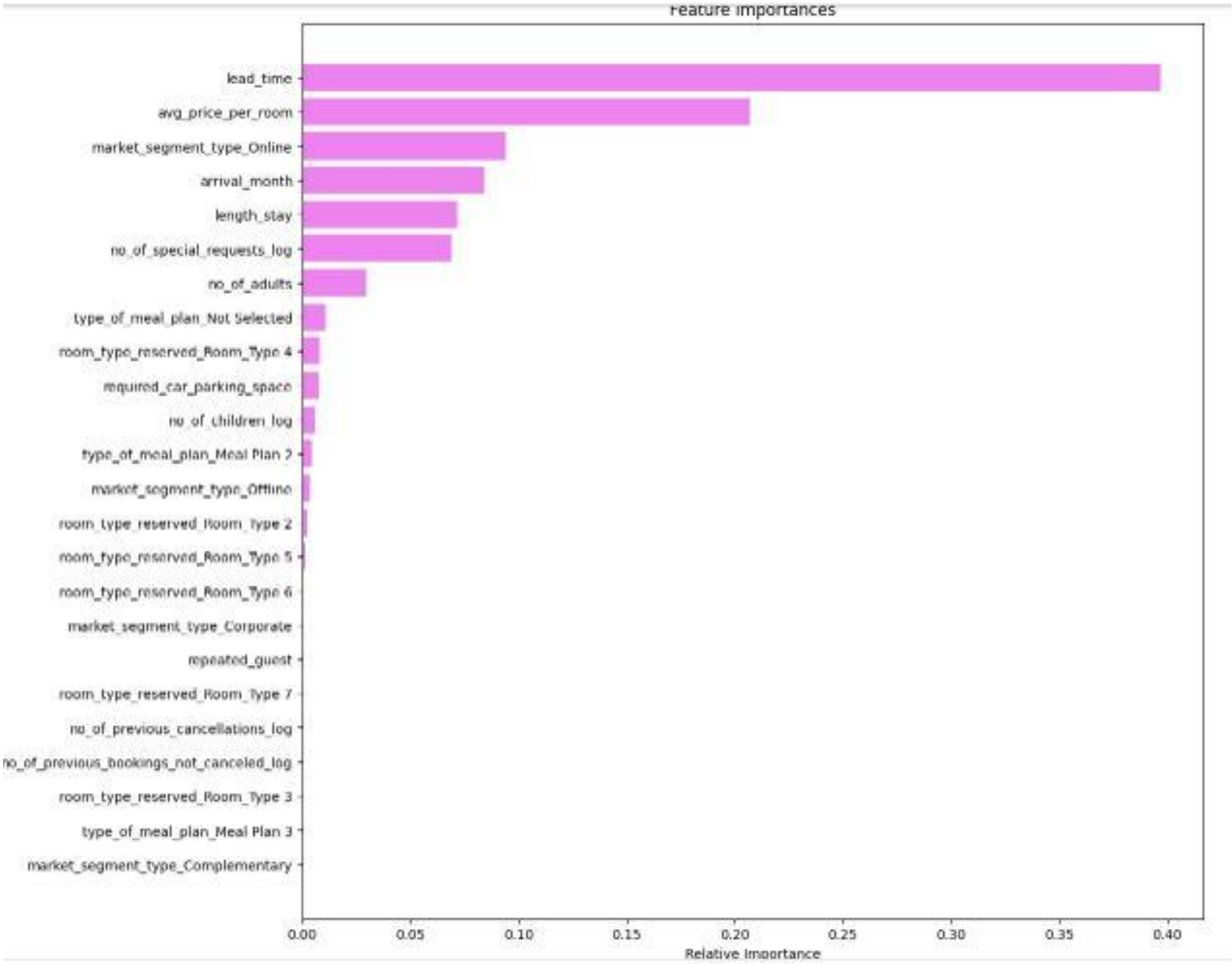
Bookings made more than 151 days before the date of arrival:

- If the average price per room is greater than 100 euros and the arrival month is December, then the the booking is less likely to be cancelled.
- If the average price per room is less than or equal to 100 euros and the number of special request is 0, then the booking is likely to get canceled.

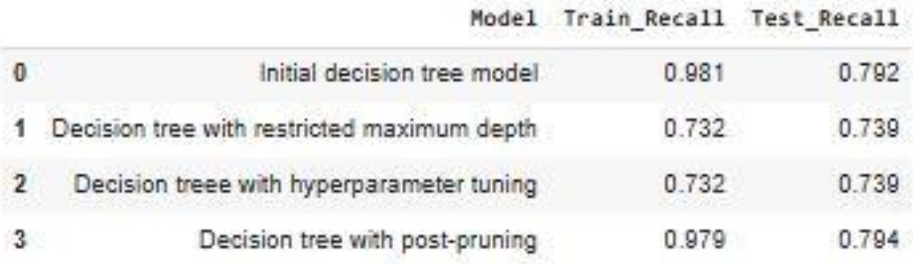
Bookings made under 151 days before the date of arrival:

- If a customer has at least 1 special request the booking is less likely to be cancelled.
- If the customer didn't make any special requests and the booking was done Online it is more likely to get cancelled, if the booking was not done online, it is less likely to be cancelled.





Train and Test Recall



	Model	Train_Recall	Test_Recall
0	Initial decision tree model	0.981	0.792
1	Decision tree with restricted maximum depth	0.732	0.739
2	Decision treee with hyperparameter tuning	0.732	0.739
3	Decision tree with post-pruning	0.979	0.794

Observations

Decision tree model with default parameters is overfitting the training data and is not able to generalize well.

Pre-pruned tree has given a generalized performance with balanced values of precision and recall.

Post-pruned tree is giving a high F1 score as compared to other models but the difference between precision and recall is high.

The hotel will be able to maintain a balance between resources and brand equity using the pre-pruned decision tree model.

Actionable Insights and Recommendations

The three most important variables in terms of cancellations were the lead time, meaning how far in advance they booked the room(s), special request for the stay, and average price of the room. Rooms booked in advance of 151 days (5 months) or less were much less likely to cancel the reservation. Those who made a special request on top of that were very unlikely to cancel. This I believe is an opportunity. Rooms booked over 151 days were more likely to cancel. Price was the determining factor for those cancellations. As the likelihood of a cancelation was increased if the room was priced over 100.04 Euros. Leading me to believe that booked early and then subsequently found a better deal

Overall we can see that the Decision Tree model performs better on the dataset.

Looking at important variables based on p-values in Logistic regression and feature importance in the Decision Tree model

Lead Time, Number of special requests, Average price per room are important in both model

From the Logistic Regression model we observe that Lead Time, and Average price per room have a positive relation with bookings getting cancelled. And the number of special requests has negative relation with bookings getting cancelled.

My Recommendations

- Offer your best room rates before 5 months ahead. After that you may be able to increase your prices slightly and increase profit.
- Require a nonrefundable deposit on all rooms in advance of over 5 months.
- Replace the 'Full Board' option on your booking with a menu of special requests available, instead of waiting for them to come in sell them even if they are no charge.
    - VIP a champagne toast at sunset your first night.
    - Room upgrades
    - WiFi
    - Laundry Bag
    - Slippers

- I believe that seasonal high prices may peak to early in OCT.

Business Recommendations

1.) The lead time and the number of special requests made by the customer play a key role in identifying if a booking will be cancelled or not. Bookings where a customer has made a special request and the booking was done under 151 days to the date of arrival are less likely to be canceled.

Using this information, the hotel can take the following actions:

- Set up a system that can send a prompt like an automated email to the customers before the arrival date asking for a re-confirmation of their booking and any changes they would like to make in their bookings.
- Remind guests about imminent deadlines.

The response given by the customer will give the hotel ample time to re-sell the room or make preparations for the customers' requests.

2.) Stricter cancellation policies can be adopted by the hotel.

- The bookings where the average price per room is high, and there were special requests associated should not get a full refund as the loss of resources will be high in these cases.
- Ideally the cancellation policies should be consistent across all market segments but as noticed in our analysis high percentage of bookings done online are cancelled. The booking cancelled online should yield less percentage of refund to the customers.

The refunds, cancellation fee, etc should be highlighted on the website/app before a customer confirms their booking to safeguard guests' interest.

3. The length of stay at the hotel can be restricted.

- We saw in our analysis that bookings, where the total length of stay was more than 5 days, had higher chances of getting cancelled.
- Hotel can allow bookings up to 5 days only and then customers should be asked to re-book if they wish to stay longer. These policies can be relaxed for corporate and Aviation market segments. For other market segments, the process should be fairly easy to not hamper their experience with the hotel.

Such restrictions can be strategized by the hotel to generate additional revenue.

4. In the months of December and January cancellation to non- cancellation ratio is low. Customers might travel to celebrate Christmas and New Year. The hotel should ensure that enough human resources are available to cater to the needs of the guests.

5. October and September saw the highest number of bookings but also high number of cancellations. This should be investigated further by the hotel.

6. Post-booking interactions can be initiated with the customers.

- Post-booking interactions will show the guests the level of attention and care they would receive at the hotel.
- To give guests a personalized experience, information about local events, nearby places to explore, etc can be shared from time to time.

7. Improving the experience of repeated customers.

- Our analysis shows that there are very few repeated customers and the cancellation among them is very less which is a good indication as repeat customers are important for the hospitality industry as they can help in spreading the word of mouth.
- A loyal guest is usually more profitable for the business because they are more familiar with offerings from the hotel they have visited before.
- Attracting new customers is tedious and costs more as compared to a repeated guest.
- A loyalty program that offers - special discounts, access to services in hotels, etc for these customers can help in improving their experience.