

# INN Hotels Project

## 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:

- Loss of resources (revenue) when the hotel cannot resell the room.
- Additional costs of distribution channels by increasing commissions or paying for publicity to help sell these rooms.
- Lowering prices last minute, so the hotel can resell a room, resulting in reducing the profit margin.
- 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: 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 week nights (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.
- 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.

## Importing necessary libraries and data

In [314...]

```
# this will help in making the Python code more structured automatically (help adhere
%load_ext nb_black

import warnings

warnings.filterwarnings("ignore")

# Libraries to help with reading and manipulating data
import pandas as pd
import numpy as np

# Libraries to help with data visualization
import matplotlib.pyplot as plt
import seaborn as sns
```

```

# Removes the limit for the number of displayed columns
pd.set_option("display.max_columns", None)
# Sets the limit for the number of displayed rows
pd.set_option("display.max_rows", 200)
# setting the precision of floating numbers to 5 decimal points
pd.set_option("display.float_format", lambda x: "%.5f" % x)

# Library to split data
from sklearn.model_selection import train_test_split

# To build model for prediction
from sklearn.tree import DecisionTreeClassifier
from sklearn import tree

# To tune different models
from sklearn.model_selection import GridSearchCV

# To get different metric scores
from sklearn.metrics import (
    f1_score,
    accuracy_score,
    recall_score,
    precision_score,
    confusion_matrix,
    make_scorer,
)

```

The nb\_black extension is already loaded. To reload it, use:  
`%reload_ext nb_black`

In [315...]: `df_main = pd.read_csv("INNHotelsGroup.csv")`

In [316...]: `# copying data to another variable to avoid any changes to original data  
data = df_main.copy()`

## Data Overview

- Observations
- Sanity checks

In [317...]: `## Overview of the dataset  
data.head()`

Out[317]:

	Booking_ID	no_of_adults	no_of_children	no_of_weekend_nights	no_of_week_nights	type_of_meal_plan
0	INN00001	2	0	1	2	Meal Plan A
1	INN00002	2	0	2	3	Not Selected
2	INN00003	1	0	2	1	Meal Plan B
3	INN00004	2	0	0	2	Meal Plan C
4	INN00005	2	0	1	1	Not Selected

In [318...]:

```
data.tail()
```

Out[318]:

	Booking_ID	no_of_adults	no_of_children	no_of_weekend_nights	no_of_week_nights	type_of_meal_plan
36270	INN36271	3	0	2	6	Meal Plan D
36271	INN36272	2	0	1	3	Meal Plan E
36272	INN36273	2	0	2	6	Meal Plan F
36273	INN36274	2	0	0	3	Not Selected
36274	INN36275	2	0	1	2	Meal Plan G

In [319...]:

```
data.shape
```

Out[319]:

```
(36275, 19)
```

The dataset has 36275 rows and 19 columns.

In [320...]:

```
data.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36275 entries, 0 to 36274
Data columns (total 19 columns):
 #   Column           Non-Null Count Dtype  
--- 
 0   Booking_ID       36275 non-null  object  
 1   no_of_adults     36275 non-null  int64   
 2   no_of_children   36275 non-null  int64   
 3   no_of_weekend_nights 36275 non-null  int64   
 4   no_of_week_nights 36275 non-null  int64   
 5   type_of_meal_plan 36275 non-null  object  
 6   required_car_parking_space 36275 non-null  int64   
 7   room_type_reserved 36275 non-null  object  
 8   lead_time         36275 non-null  int64   
 9   arrival_year      36275 non-null  int64   
 10  arrival_month     36275 non-null  int64   
 11  arrival_date      36275 non-null  int64   
 12  market_segment_type 36275 non-null  object  
 13  repeated_guest    36275 non-null  int64   
 14  no_of_previous_cancellations 36275 non-null  int64   
 15  no_of_previous_bookings_not_canceled 36275 non-null  int64   
 16  avg_price_per_room 36275 non-null  float64  
 17  no_of_special_requests 36275 non-null  int64   
 18  booking_status     36275 non-null  object  
dtypes: float64(1), int64(13), object(5)
memory usage: 5.3+ MB

```

- a. There are 5 columns which are of the type object. b. There are 14 columns that are numeric. c. The columns booking\_status is the dependent variable. d. All other columns are independent and would be used to predict the value of booking\_status.

```
In [321]: # checking for null values
data.isnull().sum()
```

```
Out[321]: 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
dtype: int64
```

There are no nulls in the data. All columns in all rows are populated.

In [322]:

```
# checking for duplicate values  
data.duplicated().sum()
```

Out[322]:

0

There are no duplicates in the dataset.

In [323]:

```
data.describe(include="all").T
```

Out[323]:

		count	unique	top	freq	mean	std
	<b>Booking_ID</b>	36275	36275	INN00001	1	NaN	NaN
	<b>no_of_adults</b>	36275.00000	NaN	NaN	NaN	1.84496	0.51871
	<b>no_of_children</b>	36275.00000	NaN	NaN	NaN	0.10528	0.40265
	<b>no_of_weekend_nights</b>	36275.00000	NaN	NaN	NaN	0.81072	0.87064
	<b>no_of_week_nights</b>	36275.00000	NaN	NaN	NaN	2.20430	1.41090
	<b>type_of_meal_plan</b>	36275	4	Meal Plan 1	27835	NaN	NaN
	<b>required_car_parking_space</b>	36275.00000	NaN	NaN	NaN	0.03099	0.17328
	<b>room_type_reserved</b>	36275	7	Room_Type 1	28130	NaN	NaN
	<b>lead_time</b>	36275.00000	NaN	NaN	NaN	85.23256	85.93082
	<b>arrival_year</b>	36275.00000	NaN	NaN	NaN	2017.82043	0.38384
	<b>arrival_month</b>	36275.00000	NaN	NaN	NaN	7.42365	3.06989
	<b>arrival_date</b>	36275.00000	NaN	NaN	NaN	15.59700	8.74045
	<b>market_segment_type</b>	36275	5	Online	23214	NaN	NaN
	<b>repeated_guest</b>	36275.00000	NaN	NaN	NaN	0.02564	0.15805
	<b>no_of_previous_cancellations</b>	36275.00000	NaN	NaN	NaN	0.02335	0.36833
	<b>no_of_previous_bookings_not_canceled</b>	36275.00000	NaN	NaN	NaN	0.15341	1.75417
	<b>avg_price_per_room</b>	36275.00000	NaN	NaN	NaN	103.42354	35.08942
	<b>no_of_special_requests</b>	36275.00000	NaN	NaN	NaN	0.61966	0.78624
	<b>booking_status</b>	36275	2	Not_Canceled	24390	NaN	NaN

In [324]:

```
# checking for unique values in ID column  
data["Booking_ID"].nunique()
```

Out[324]:

36275

```
In [325...]: data.drop(["Booking_ID"], axis=1, inplace=True)
```

Booking ID has been dropped.

```
In [326...]: data.isnull().sum()
```

```
Out[326]: 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  
dtype: int64
```

## Exploratory Data Analysis (EDA)

- EDA is an important part of any project involving data.
- It is important to investigate and understand the data better before building a model with it.
- A few questions have been mentioned below which will help you approach the analysis in the right manner and generate insights from the data.
- A thorough analysis of the data, in addition to the questions mentioned below, should be done.

### Leading 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?

# Univariate Analysis

In [327...]

```
def histogram_boxplot(data, feature, figsize=(15, 10), kde=False, bins=None):
    """
    Boxplot and histogram combined

    data: dataframe
    feature: dataframe column
    figsize: size of figure (default (15,10))
    kde: whether to show the density curve (default False)
    bins: number of bins for histogram (default None)
    """

    f2, (ax_box2, ax_hist2) = plt.subplots(
        nrows=2, # Number of rows of the subplot grid= 2
        sharex=True, # x-axis will be shared among all subplots
        gridspec_kw={"height_ratios": (0.25, 0.75)},
        figsize=figsize,
    ) # creating the 2 subplots
    sns.boxplot(
        data=data, x=feature, ax=ax_box2, showmeans=True, color="violet"
    ) # boxplot will be created and a triangle will indicate the mean value of the column
    sns.histplot(
        data=data, x=feature, kde=kde, ax=ax_hist2, bins=bins
    ) if bins else sns.histplot(
        data=data, x=feature, kde=kde, ax=ax_hist2
    ) # For histogram
    ax_hist2.axvline(
        data[feature].mean(), color="green", linestyle="--"
    ) # Add mean to the histogram
    ax_hist2.axvline(
        data[feature].median(), color="black", linestyle="-"
    ) # Add median to the histogram
```

In [328...]

```
# function to create Labeled barplots

def labeled_barplot(data, feature, perc=False, n=None):
    """
    Barplot with percentage at the top

    data: dataframe
    feature: dataframe column
    perc: whether to display percentages instead of count (default is False)
    n: displays the top n category levels (default is None, i.e., display all levels)
    """

    total = len(data[feature]) # length of the column
    count = data[feature].nunique()
    if n is None:
        plt.figure(figsize=(count + 2, 6))
    else:
        plt.figure(figsize=(n + 2, 6))

    plt.xticks(rotation=90, fontsize=15)
    ax = sns.countplot(
        data=data,
```

```

        x=feature,
        palette="Paired",
        order=data[feature].value_counts().index[:n],
    )

    for p in ax.patches:
        if perc == True:
            label = "{:.1f}%".format(
                100 * p.get_height() / total
            ) # percentage of each class of the category
        else:
            label = p.get_height() # count of each level of the category

        x = p.get_x() + p.get_width() / 2 # width of the plot
        y = p.get_height() # height of the plot

        ax.annotate(
            label,
            (x, y),
            ha="center",
            va="center",
            size=12,
            xytext=(0, 5),
            textcoords="offset points",
        ) # annotate the percentage

plt.show() # show the plot

```

In [329...]

```

def stacked_barplot(data, predictor, target):
    """
    Print the category counts and plot a stacked bar chart

    data: dataframe
    predictor: independent variable
    target: target variable
    """
    count = data[predictor].nunique()
    sorter = data[target].value_counts().index[-1]
    tab1 = pd.crosstab(data[predictor], data[target], margins=True).sort_values(
        by=sorter, ascending=False
    )
    print(tab1)
    print("-" * 120)
    tab = pd.crosstab(data[predictor], data[target], normalize="index").sort_values(
        by=sorter, ascending=False
    )
    tab.plot(kind="bar", stacked=True, figsize=(count + 5, 5))
    plt.legend(
        loc="lower left", frameon=False,
    )
    plt.legend(loc="upper left", bbox_to_anchor=(1, 1))
    plt.show()

```

In [330...]

```

### function to plot distributions wrt target

def distribution_plot_wrt_target(data, predictor, target):

```

```
fig, axs = plt.subplots(2, 2, figsize=(12, 10))

target_uniq = data[target].unique()

axs[0, 0].set_title("Distribution of target for target=" + str(target_uniq[0]))
sns.histplot(
    data=data[data[target] == target_uniq[0]],
    x=predictor,
    kde=True,
    ax=axs[0, 0],
    color="teal",
    stat="density",
)

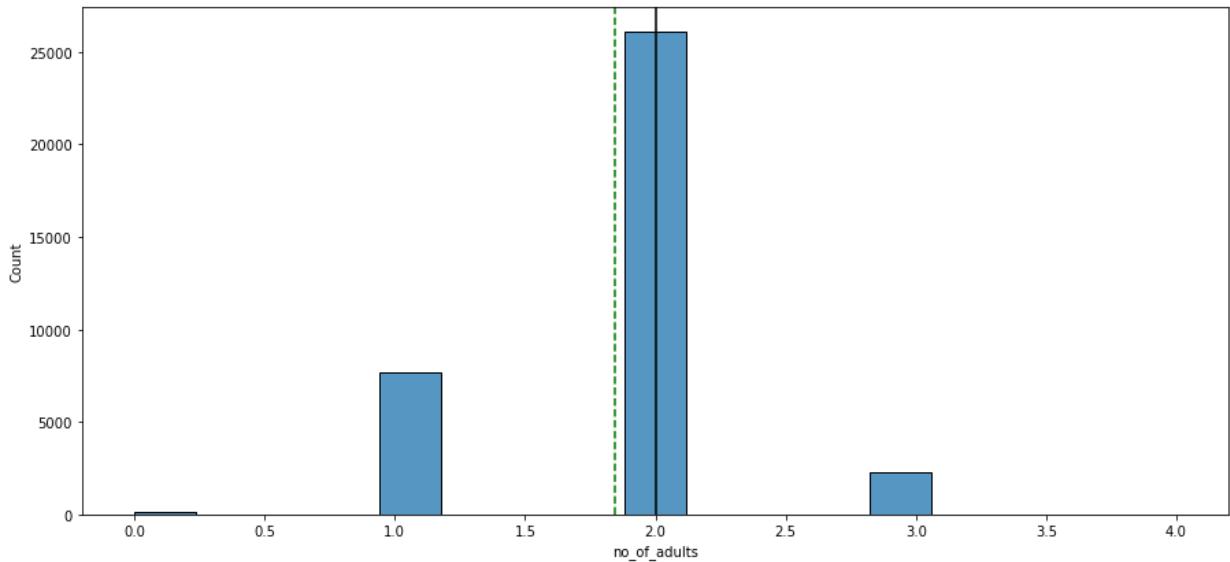
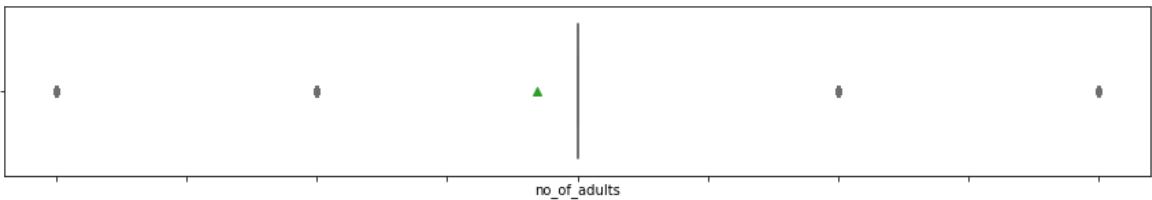
axs[0, 1].set_title("Distribution of target for target=" + str(target_uniq[1]))
sns.histplot(
    data=data[data[target] == target_uniq[1]],
    x=predictor,
    kde=True,
    ax=axs[0, 1],
    color="orange",
    stat="density",
)

axs[1, 0].set_title("Boxplot w.r.t target")
sns.boxplot(data=data, x=target, y=predictor, ax=axs[1, 0], palette="gist_rainbow")

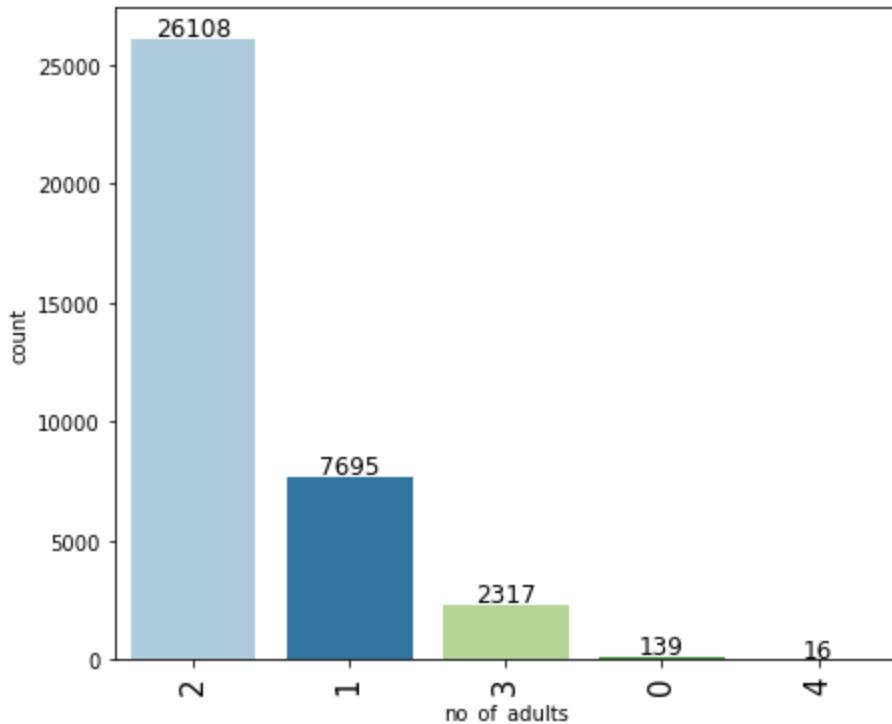
axs[1, 1].set_title("Boxplot (without outliers) w.r.t target")
sns.boxplot(
    data=data,
    x=target,
    y=predictor,
    ax=axs[1, 1],
    showfliers=False,
    palette="gist_rainbow",
)

plt.tight_layout()
plt.show()
```

In [331...]: histogram\_boxplot(data, "no\_of\_adults")



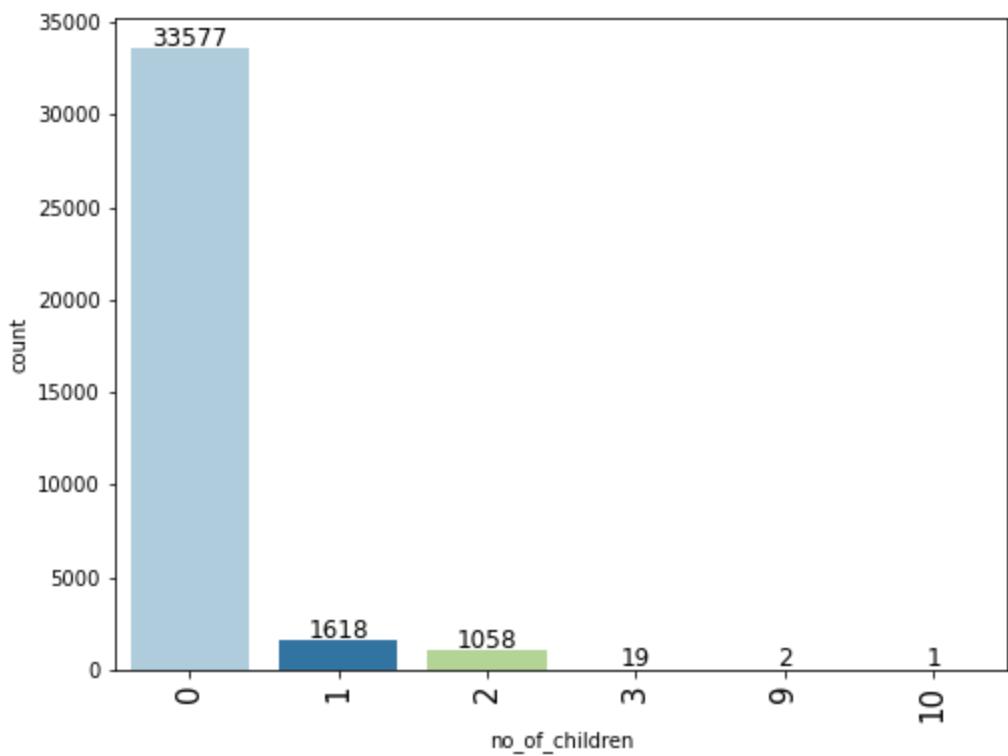
```
In [332]: labeled_barplot(data, "no_of_adults")
```



Most customers make reservations for 2 adults.

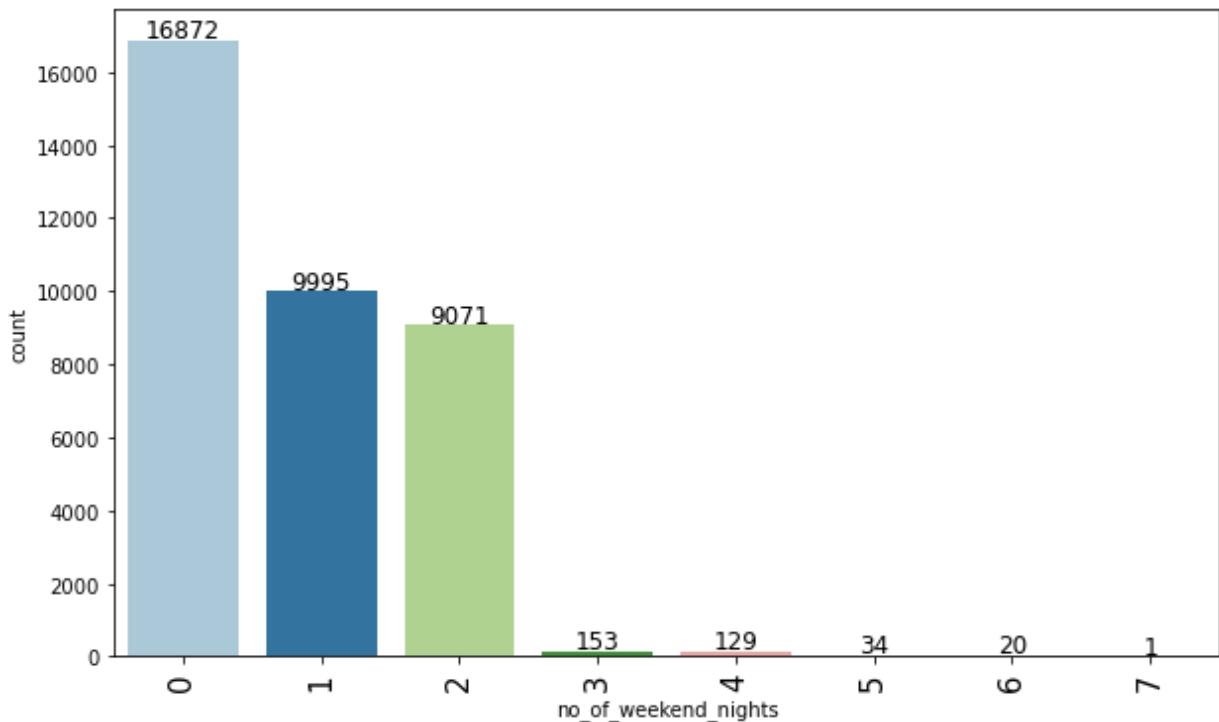
There are few with 3 and even fewer with 4 adults in the group.

```
In [333]: labeled_barplot(data, "no_of_children")
```



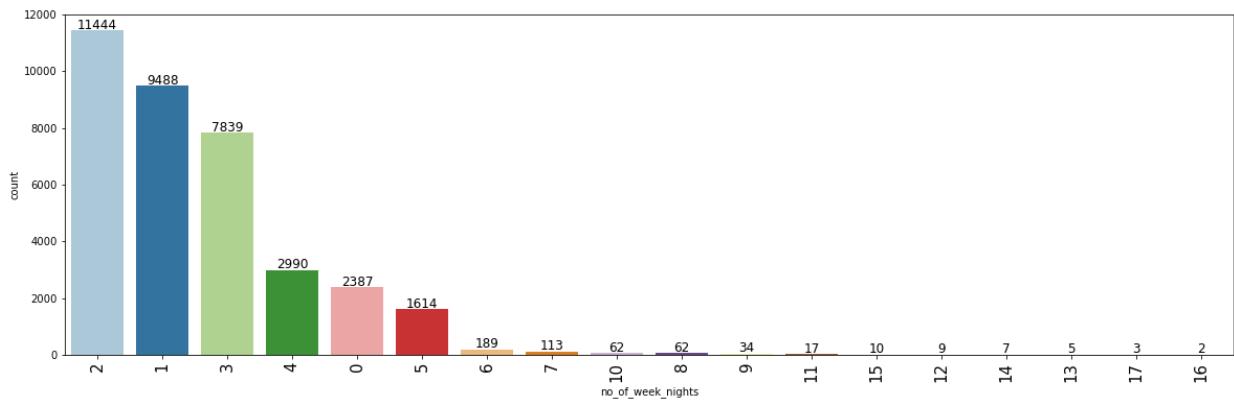
Most customers made reservations with 2 kids.

```
In [334]: labeled_barplot(data, "no_of_weekend_nights")
```



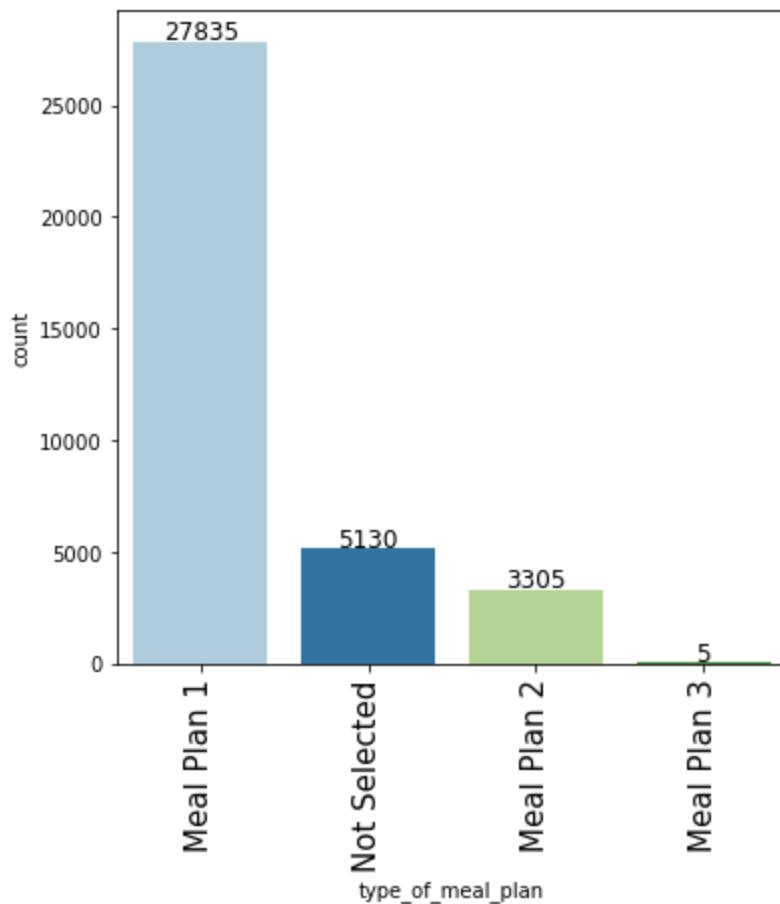
Majority of reservations were not for weekend night

```
In [335]: labeled_barplot(data, "no_of_week_nights")
```



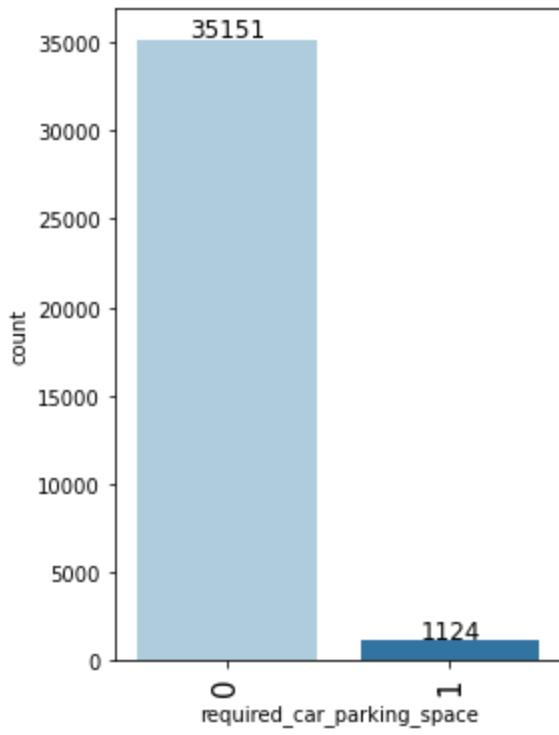
Most reservations are made for 2 week nights.

```
In [339]: labeled_barplot(data, "type_of_meal_plan")
```



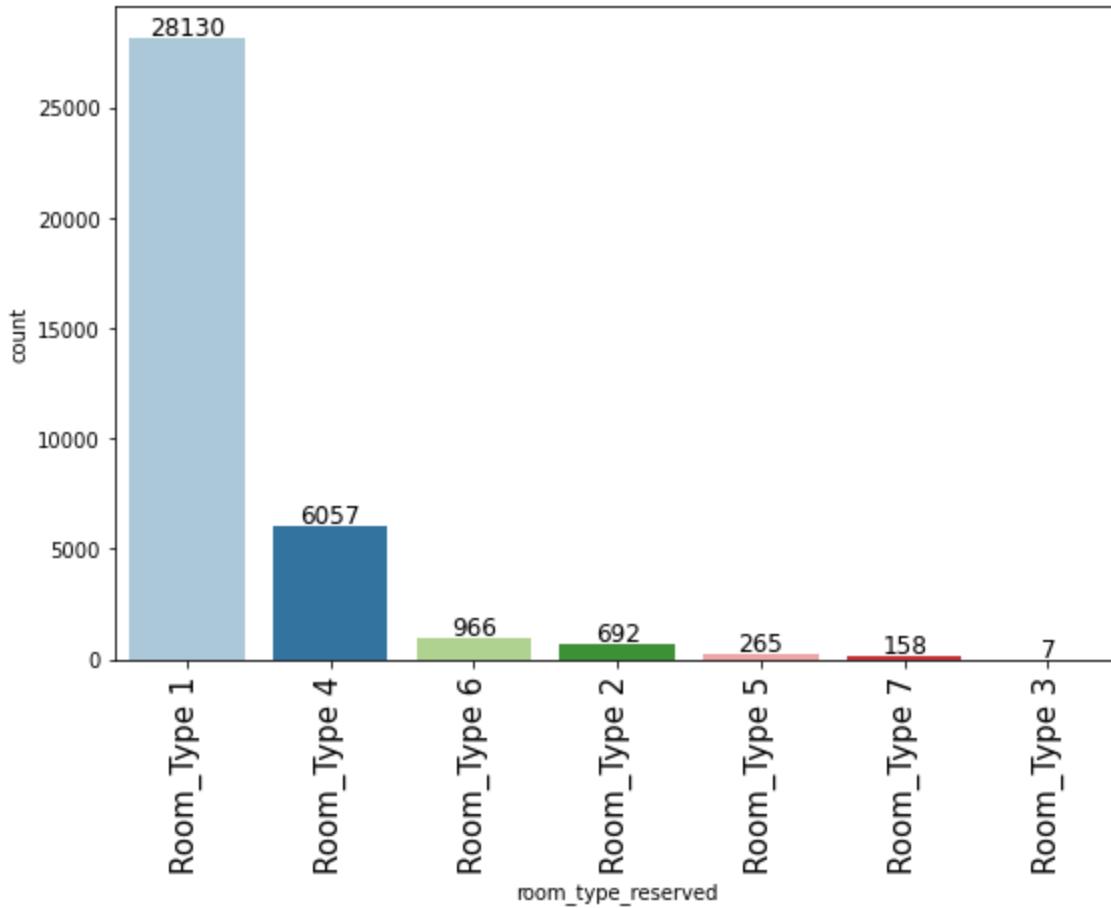
Most customers opted for Breakfast next morning as part of their reservation.

```
In [340]: labeled_barplot(data, "required_car_parking_space")
```



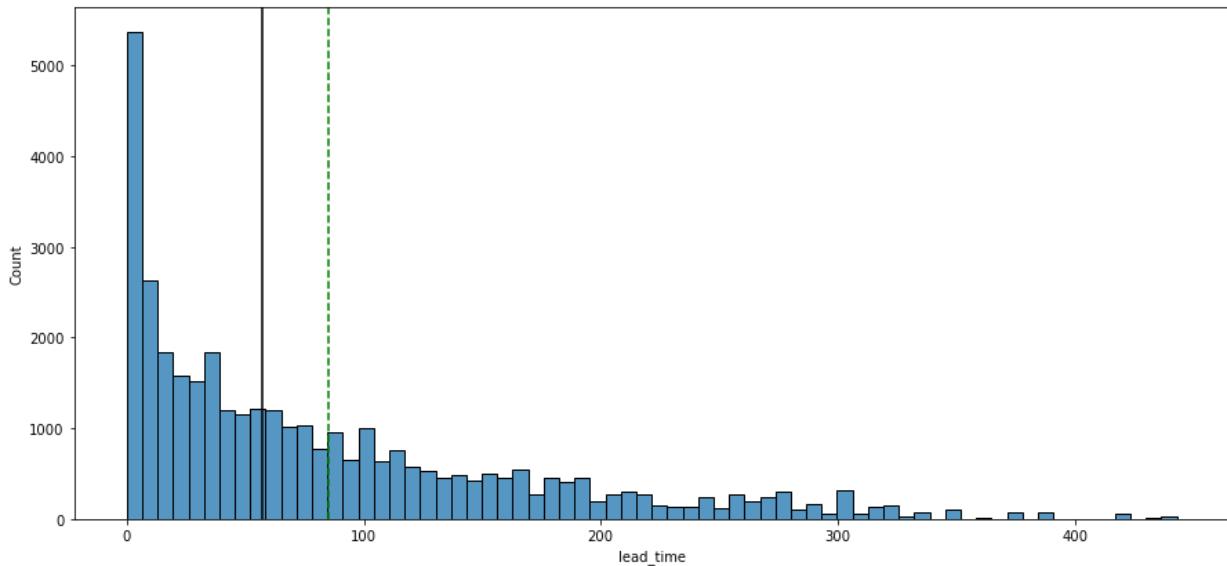
Most customers making reservation did not request a car parking be made available to them.

```
In [341]: labeled_barplot(data, "room_type_reserved")
```



Room type 1 is a very popular choice among the customers. Room Type 3 is the least popular.

```
In [342...]: # labeled_barplot(data, "lead_time")
histogram_boxplot(data, "lead_time")
```

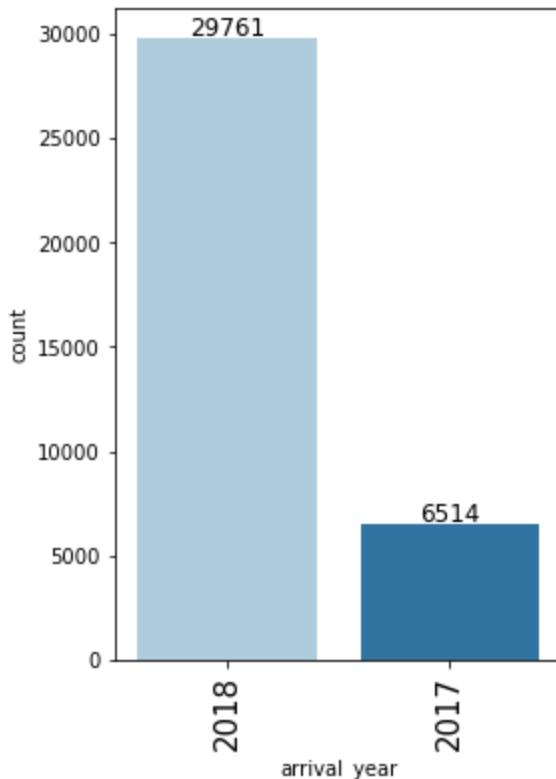


```
In [343...]: data.groupby("lead_time")["lead_time"].count().head(20)
```

```
Out[343]: lead_time
0    1297
1    1078
2     643
3     630
4     628
5     577
6     519
7     429
8     436
9     332
10    317
11    371
12    412
13    332
14    384
15    298
16    255
17    288
18    302
19    316
Name: lead_time, dtype: int64
```

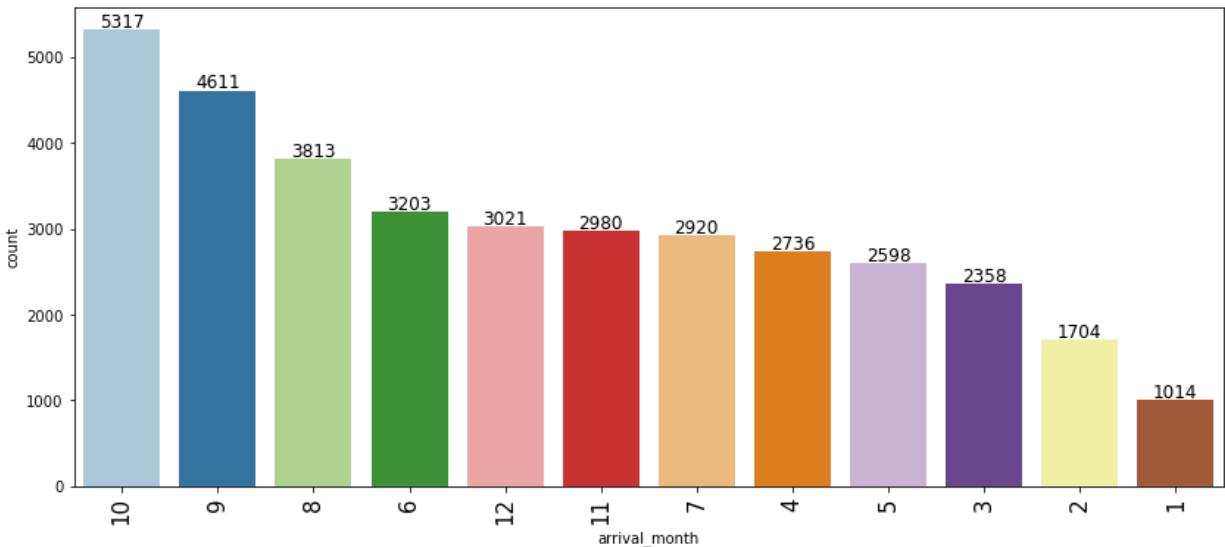
Most customers make instant bookings, i.e. with a lead time of 1 day or less.

```
In [344...]: labeled_barplot(data, "arrival_year")
```



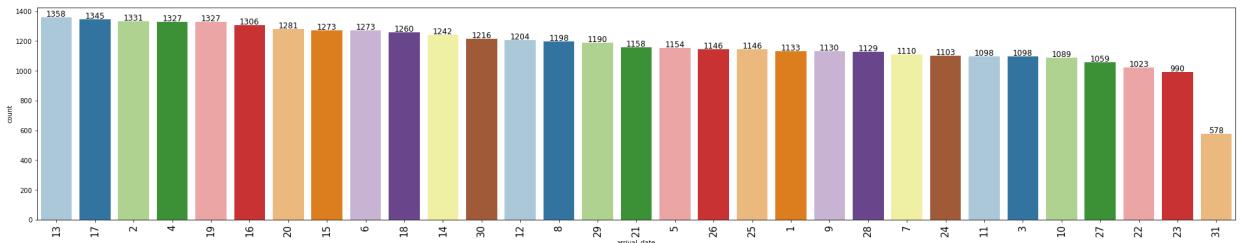
The dataset contains details about reservations made in 2017 and 2018. Most reservations in the dataset are from 2018.

```
In [345...]: labeled_barplot(data, "arrival_month")
```



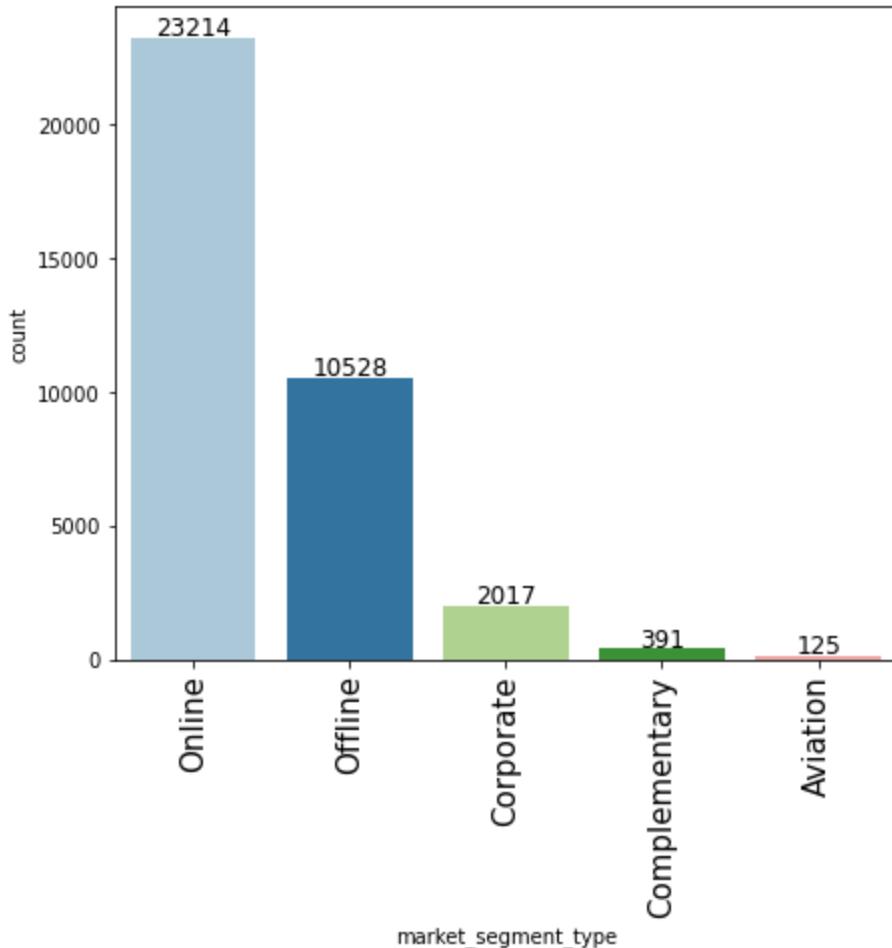
**October is the most popular month followed by September and August. February and January are the least popular.**

```
In [346...]: labeled_barplot(data, "arrival_date")
```



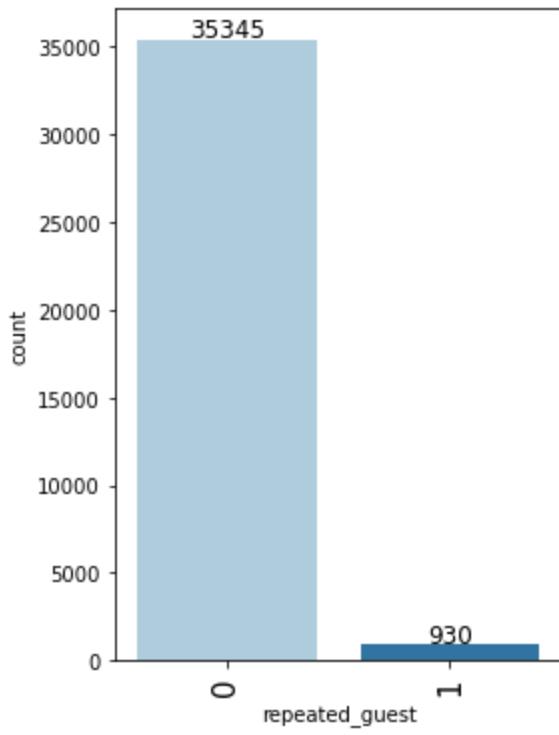
There is not a very significant difference on the booking date for each month.

In [347...]: `labeled_barplot(data, "market_segment_type")`



**Most reservations were made online by the customers and 'Aviation' was the least popular market segment used for making reservations.**

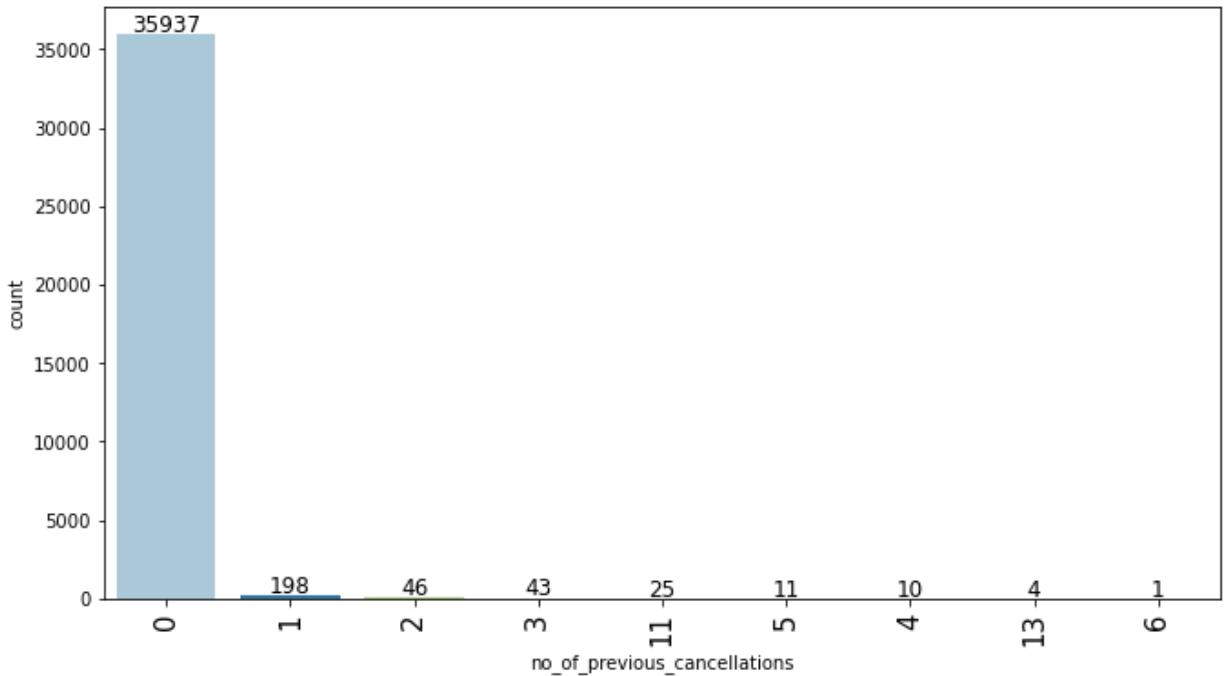
In [348...]: `labeled_barplot(data, "repeated_guest")`



Most customers present in the dataset and not repeating or returning customers. These are just one time bookings.

In [349...]

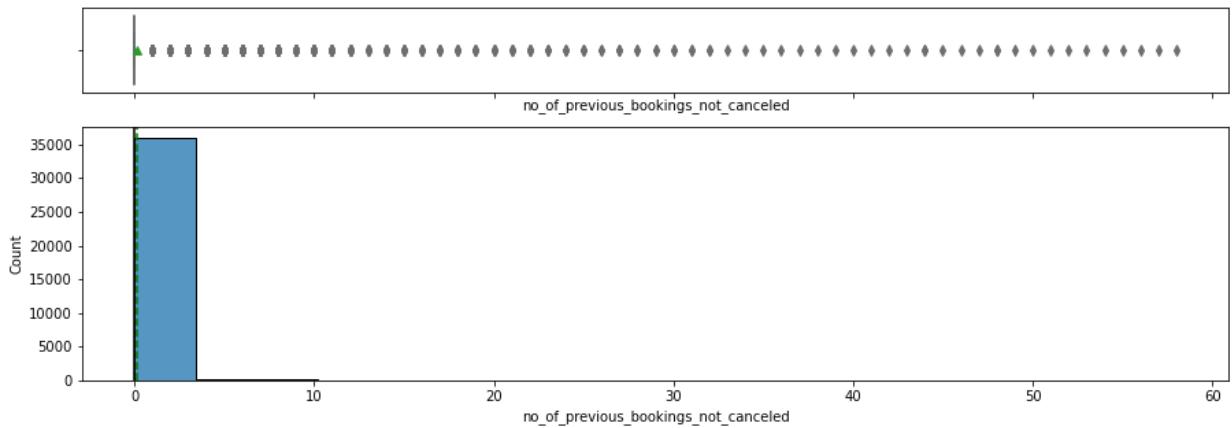
```
labeled_barplot(data, "no_of_previous_cancellations")
```



Most customers have not made any prior cancellations.

In [350...]

```
# labeled_barplot(data, "no_of_previous_bookings_not_canceled")
histogram_boxplot(data, "no_of_previous_bookings_not_canceled", figsize=(15, 5))
```

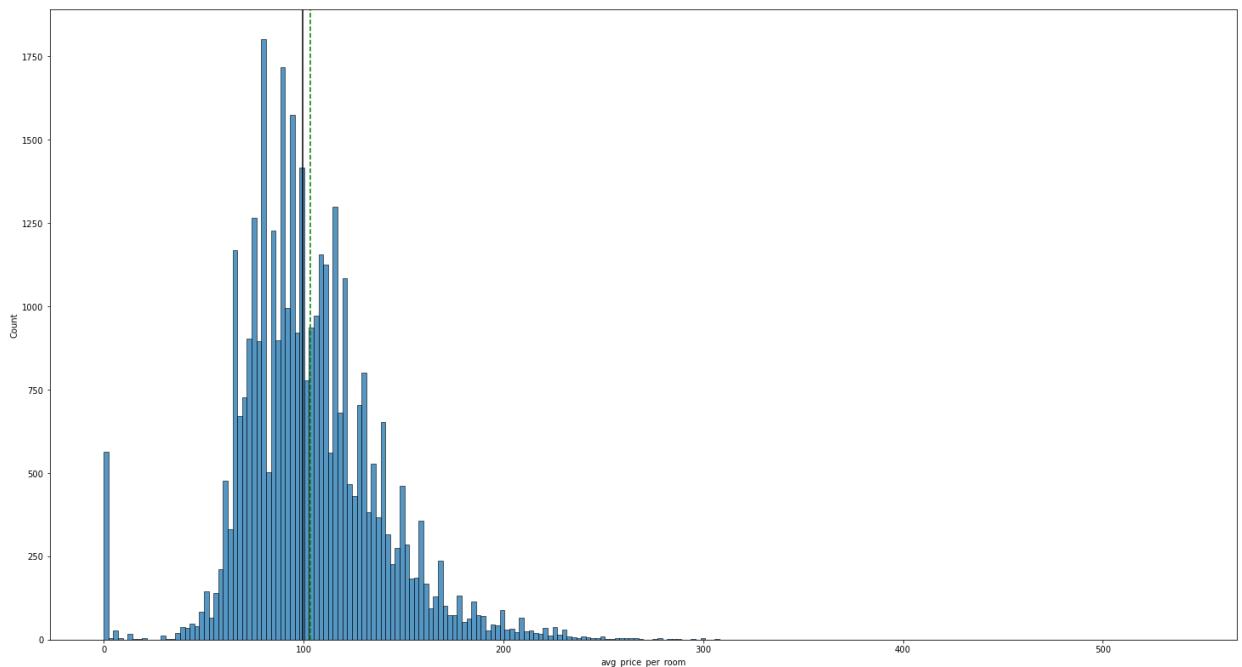
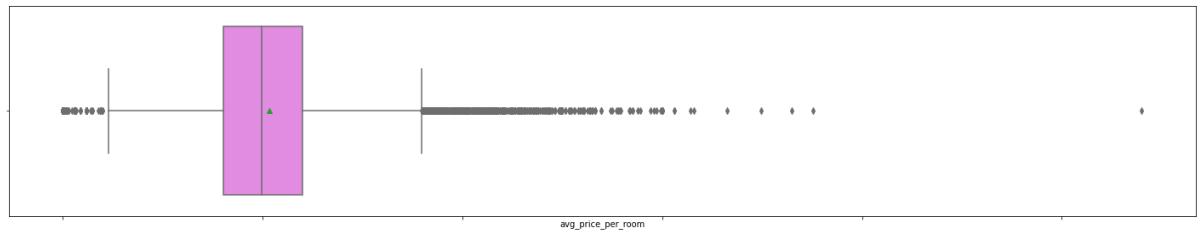


```
In [351]: data.groupby("no_of_previous_bookings_not_canceled")[
    "no_of_previous_bookings_not_canceled"
].count().head(20)
```

```
Out[351]: no_of_previous_bookings_not_canceled
0      35463
1       228
2       112
3        80
4        65
5        60
6        36
7        24
8        23
9        19
10      19
11      15
12      12
13       7
14       9
15       8
16       7
17       6
18       6
19       6
Name: no_of_previous_bookings_not_canceled, dtype: int64
```

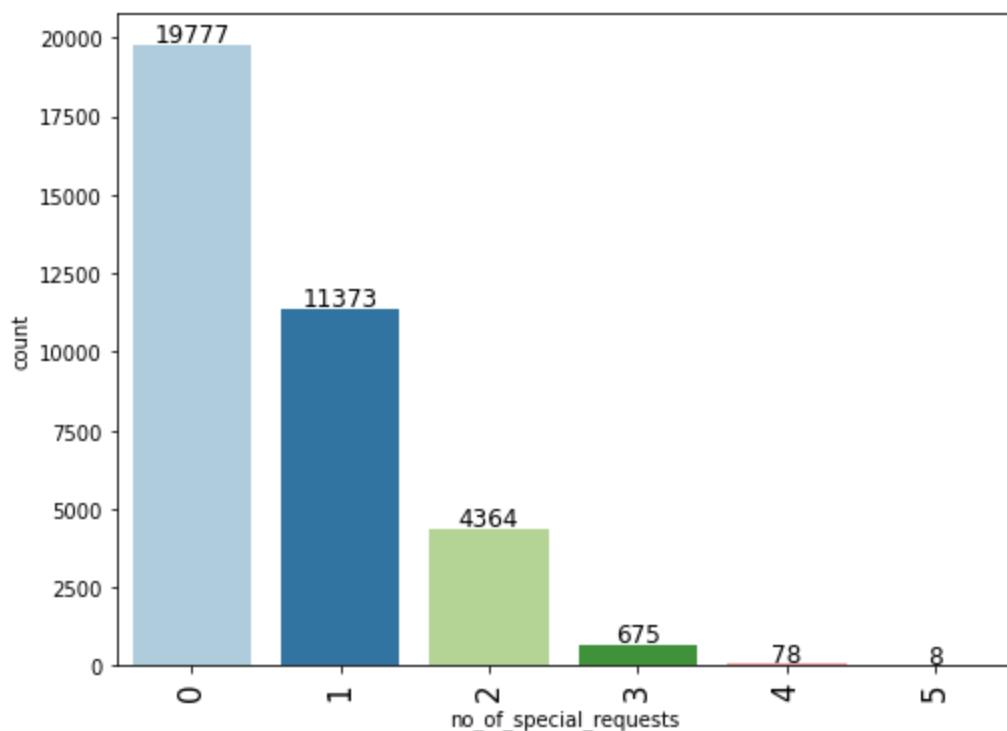
Most customers have not cancelled any previous booking.

```
In [352]: # Labeled_barPlot(data, "avg_price_per_room")
histogram_boxplot(data, "avg_price_per_room", figsize=(25, 20))
```



Average price per room varies with the reservation but usually has a median of close to 100 EU.

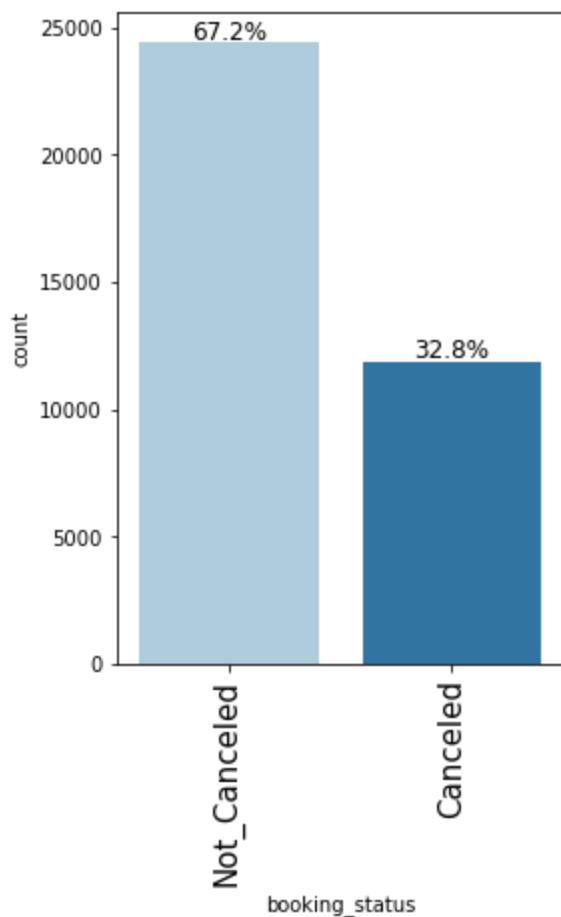
```
In [353]: labeled_barplot(data, "no_of_special_requests")
```



Most customers have not made any special request.

In [354...]

```
labeled_barplot(data, "booking_status", perc=True, n=None)
```

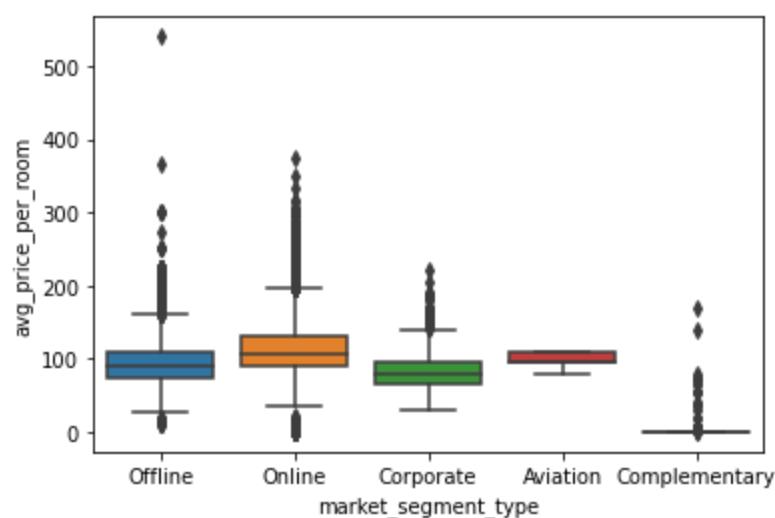


The percentage of cancelled bookings is 32.8 percent.

## Bivariate Analysis

In [356...]

```
sns.boxplot(data=data, y="avg_price_per_room", x="market_segment_type")
```

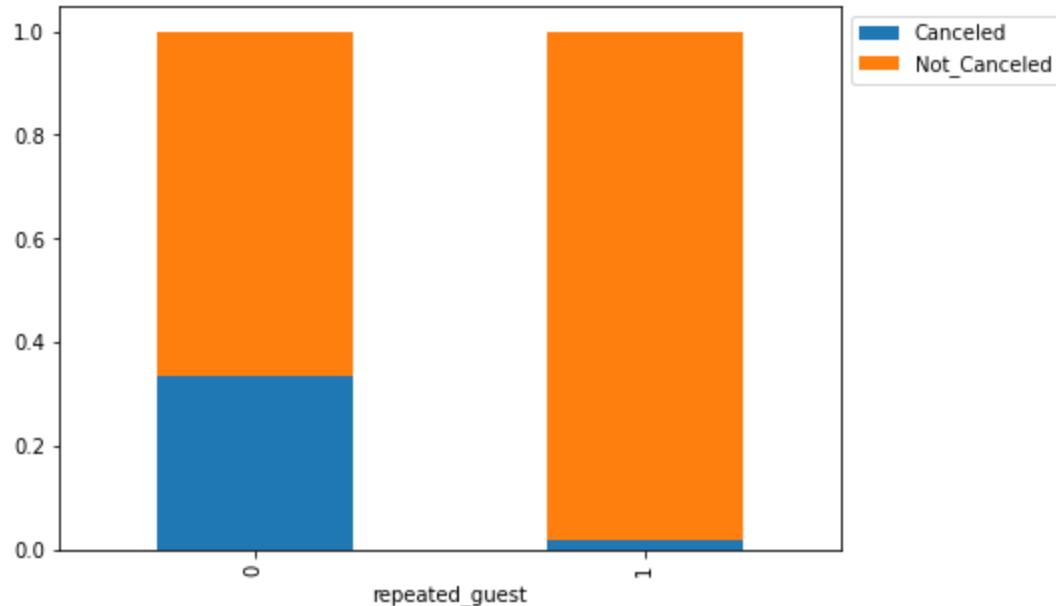


There's not a major difference in the median of all the booking methods. Although there's a slightly higher median for online bookings. There are outliers though in all the methods except Aviation.

```
In [358...]: # sns.catplot(data=data, x="repeated_guest", hue="booking_status", kind="count")
```

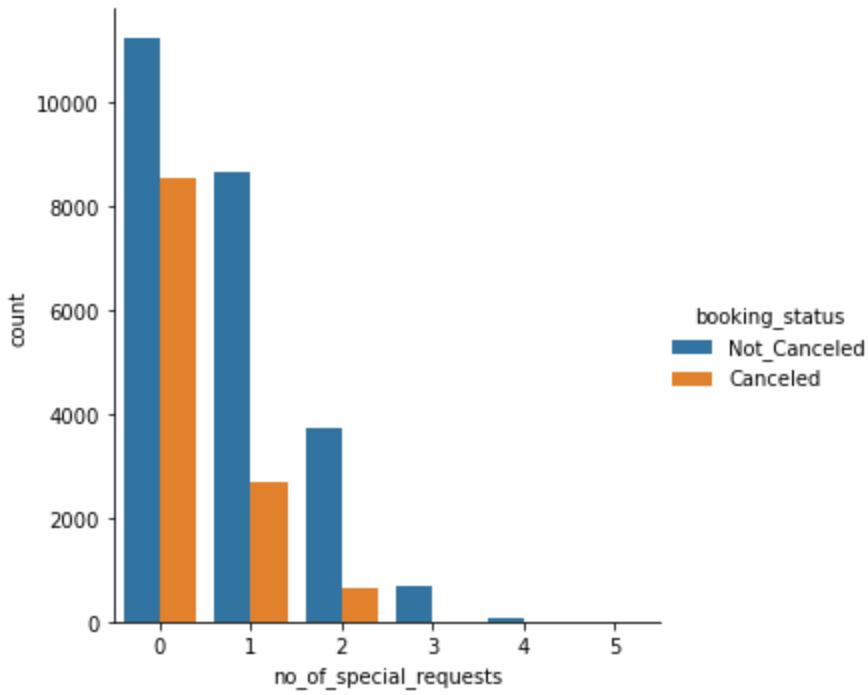
```
In [359...]: stacked_barplot(data, "repeated_guest", "booking_status")
```

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



Repeating guests have cancelled very few bookings (16).

```
In [361...]: sns.catplot(data=data, x="no_of_special_requests", hue="booking_status", kind="count")
```



It does appear that there is a high probability of cancellation if there are no special requests. There's a downward trend in cancellations as the number of special requests increases.

In [363...]

```
plt.figure(figsize=(20, 10))
sns.heatmap(data.corr(), annot=True, vmin=-1, vmax=1, fmt=".2f", cmap="Spectral")
```



In [364...]

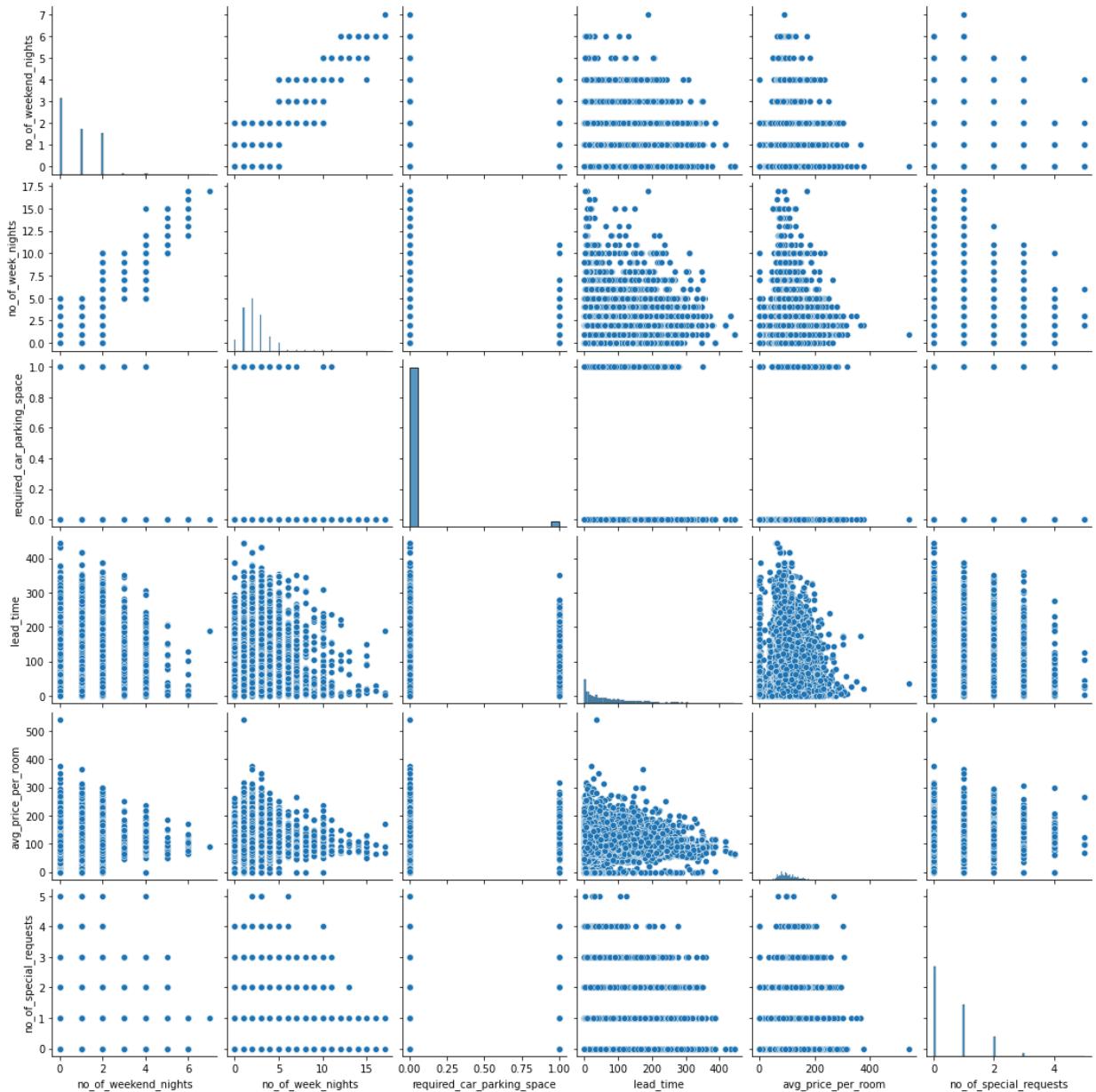
```
sns.pairplot(
    data[
        [

```

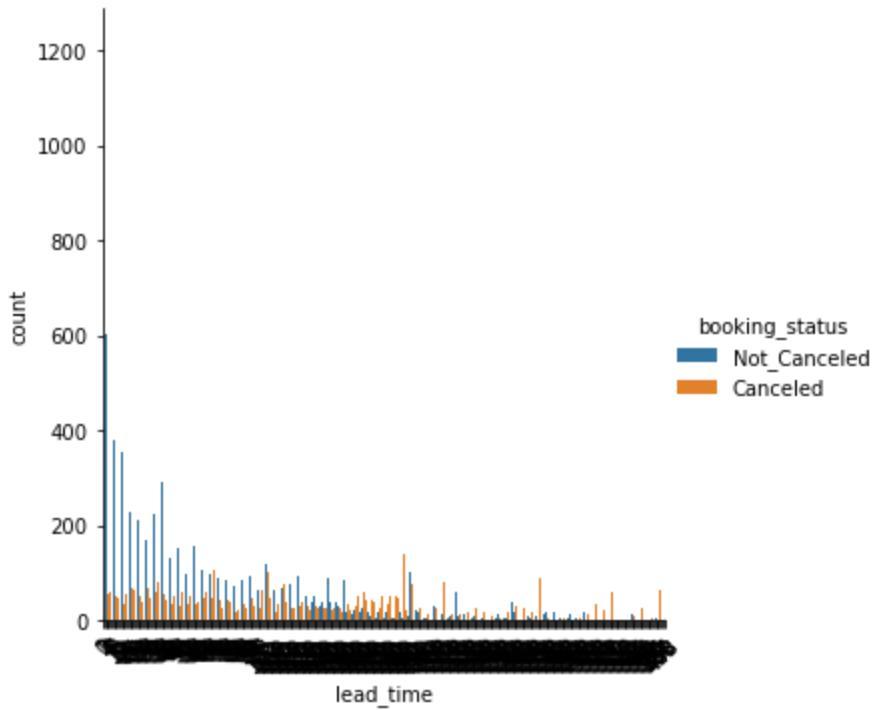
```

        "no_of_weekend_nights",
        "no_of_week_nights",
        "required_car_parking_space",
        "lead_time",
        "avg_price_per_room",
        "no_of_special_requests",
        "type_of_meal_plan",
        "room_type_reserved",
        "market_segment_type",
        "booking_status",
    ]
)
]
)
```

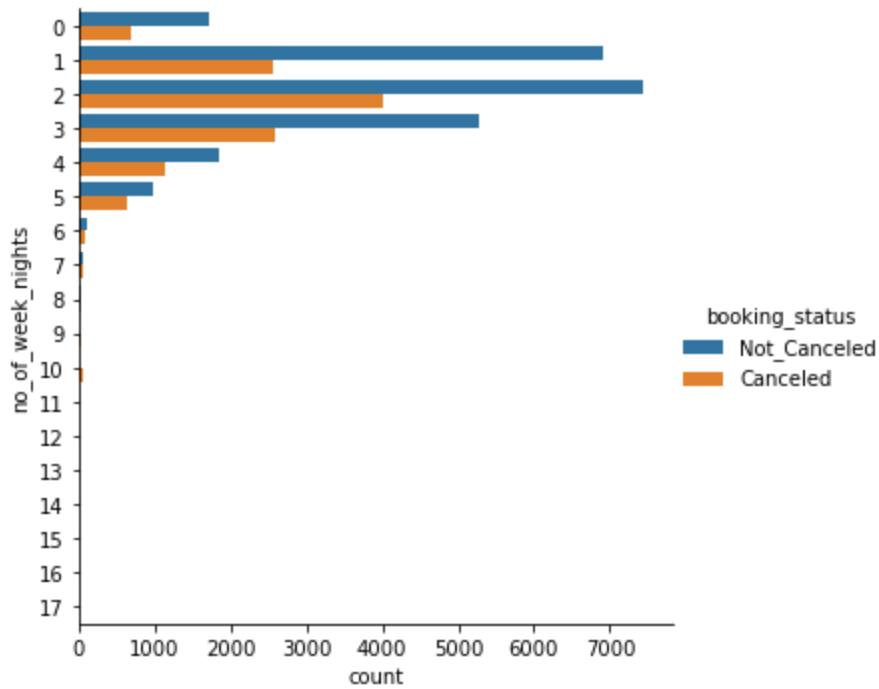
Out[364]: <seaborn.axisgrid.PairGrid at 0x1e373263130>



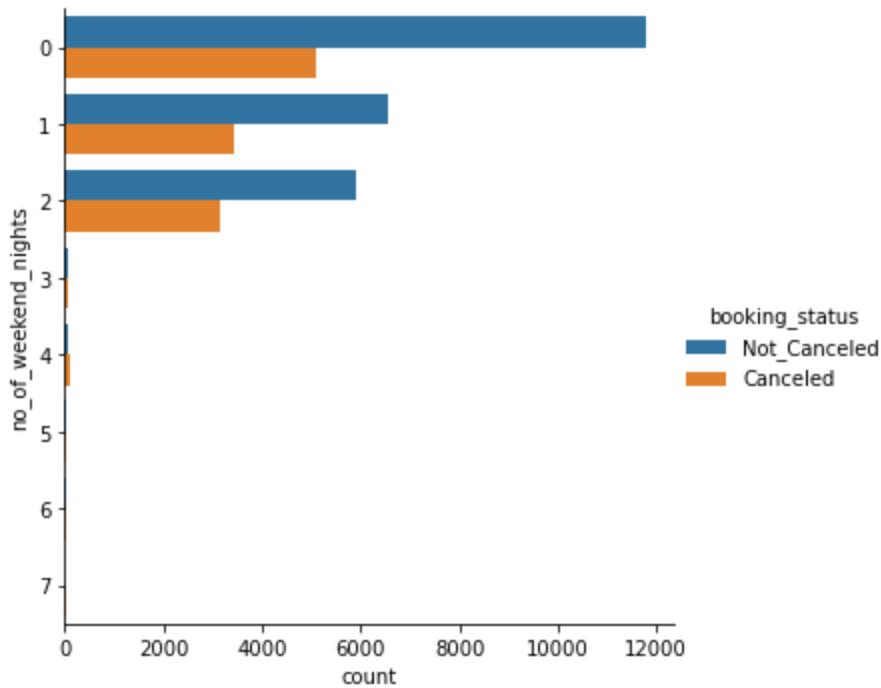
In [365...]  
`sns.catplot(data=data, x="lead_time", hue="booking_status", kind="count")  
plt.xticks(rotation=45)  
plt.show()`



```
In [366]: sns.catplot(data=data, y="no_of_week_nights", hue="booking_status", kind="count")
plt.show()
```

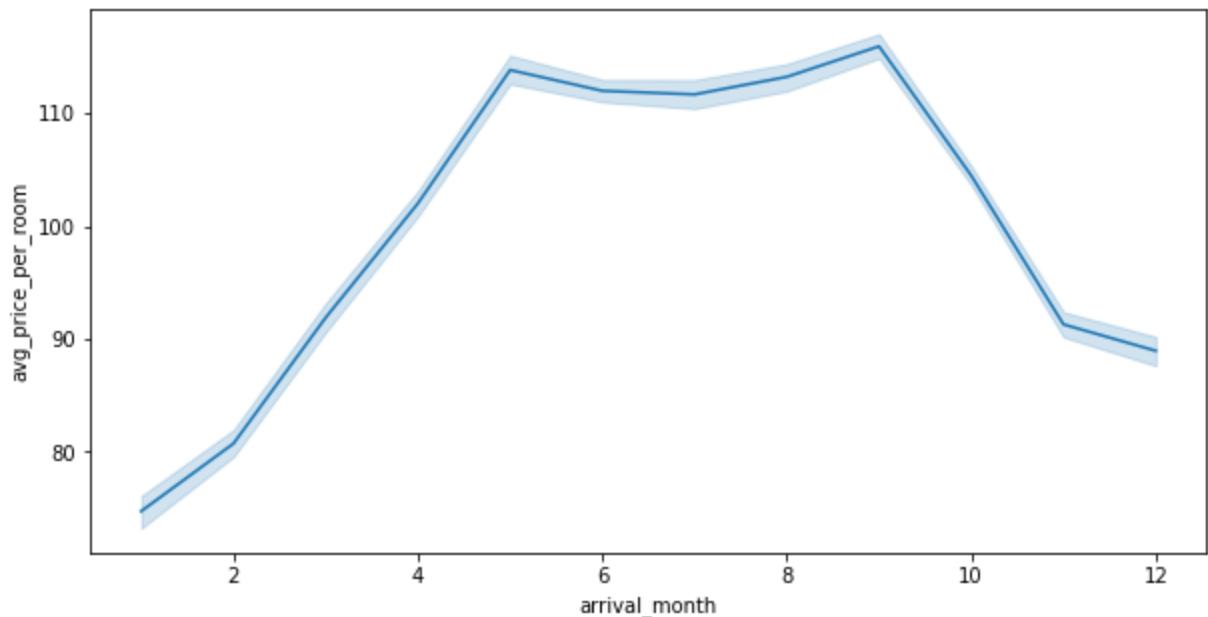


```
In [368]: sns.catplot(data=data, y="no_of_weekend_nights", hue="booking_status", kind="count")
plt.show()
```



In [374]:

```
plt.figure(figsize=(10, 5))
sns.lineplot(data=data, x="arrival_month", y="avg_price_per_room")
plt.show()
```

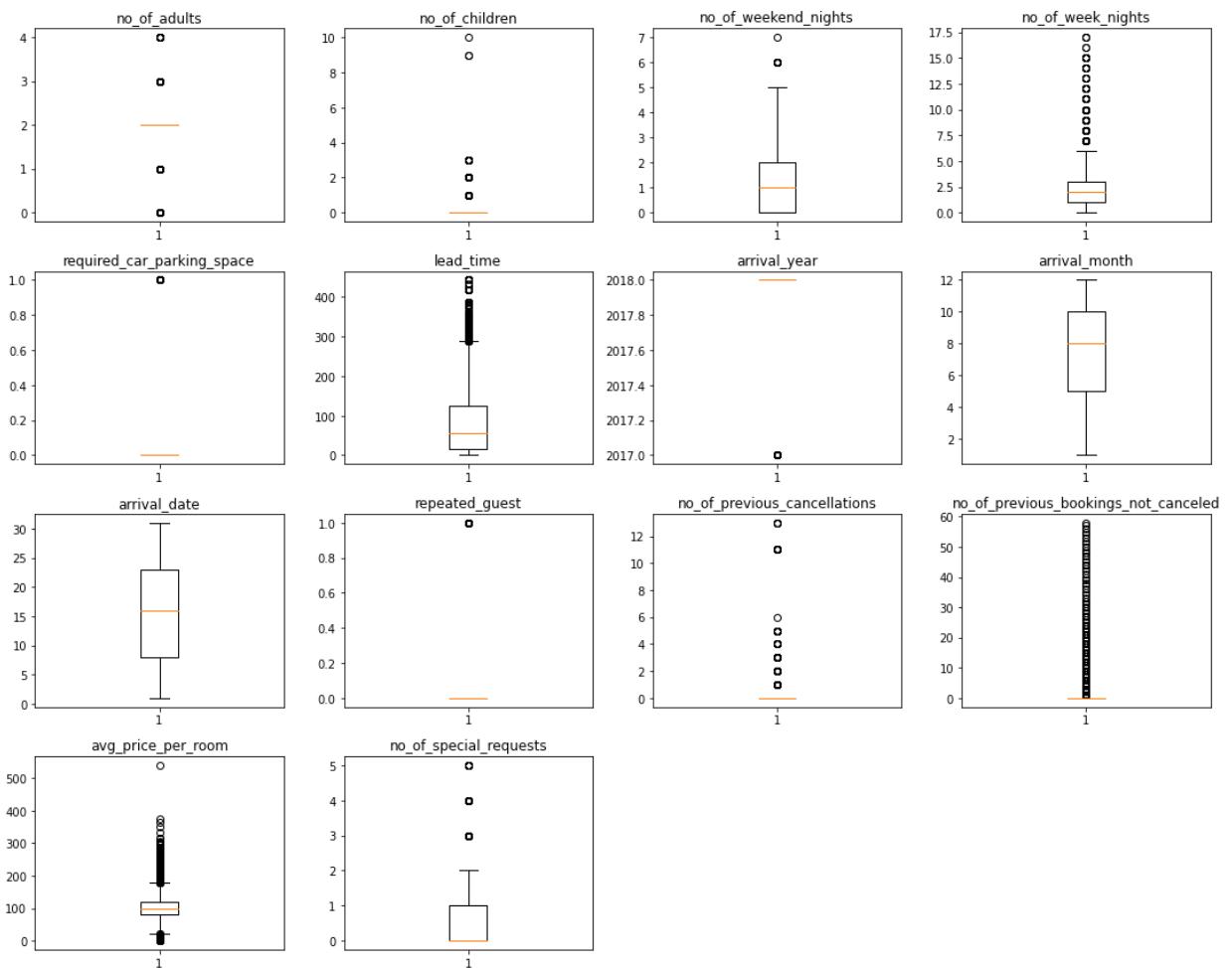


## Data Preprocessing

- Missing value treatment (if needed)
- Feature engineering (if needed)
- Outlier detection and treatment (if needed)
- Preparing data for modeling
- Any other preprocessing steps (if needed)

## Checking for outliers in the data.

```
In [375...]  
# outlier detection using boxplot  
numeric_columns = data.select_dtypes(include=np.number).columns.tolist()  
  
plt.figure(figsize=(15, 12))  
  
for i, variable in enumerate(numeric_columns):  
    plt.subplot(4, 4, i + 1)  
    plt.boxplot(data[variable], whis=1.5)  
    plt.tight_layout()  
    plt.title(variable)  
  
plt.show()
```



There are quite a few outliers in the columns 'avg\_price\_per\_room', 'lead\_time', 'no\_of\_previous\_bookings\_not\_canceled'. But we are NOT going to treat them as they are actual values that have been provided.

```
In [376...]  
data
```

Out[376]:

	no_of_adults	no_of_children	no_of_weekend_nights	no_of_week_nights	type_of_meal_plan	req
0	2	0	1	2	Meal Plan 1	
1	2	0	2	3	Not Selected	
2	1	0	2	1	Meal Plan 1	
3	2	0	0	2	Meal Plan 1	
4	2	0	1	1	Not Selected	
...	...	...	...	...	...	...
36270	3	0	2	6	Meal Plan 1	
36271	2	0	1	3	Meal Plan 1	
36272	2	0	2	6	Meal Plan 1	
36273	2	0	0	3	Not Selected	
36274	2	0	1	2	Meal Plan 1	

36275 rows × 18 columns

### Creating dummy columns for categorical variables.

In [377...]

```
data["booking_status"] = data["booking_status"].apply(
    lambda x: 1 if x == "Canceled" else 0
)
```

In [378...]

```
df = pd.get_dummies(
    data,
    columns=["type_of_meal_plan", "room_type_reserved", "market_segment_type"],
    drop_first=True,
)
df.head()
```

Out[378]:

	no_of_adults	no_of_children	no_of_weekend_nights	no_of_week_nights	required_car_parking_space
0	2	0	1	2	0
1	2	0	2	3	0
2	1	0	2	1	0
3	2	0	0	2	0
4	2	0	1	1	0

Year is probably a column that is not of a major significance and can be dropped.

```
In [379... df = df.drop(["arrival_year"], axis=1)
## Arrival Year has been successfully dropped.
```

```
In [380... df.head()
```

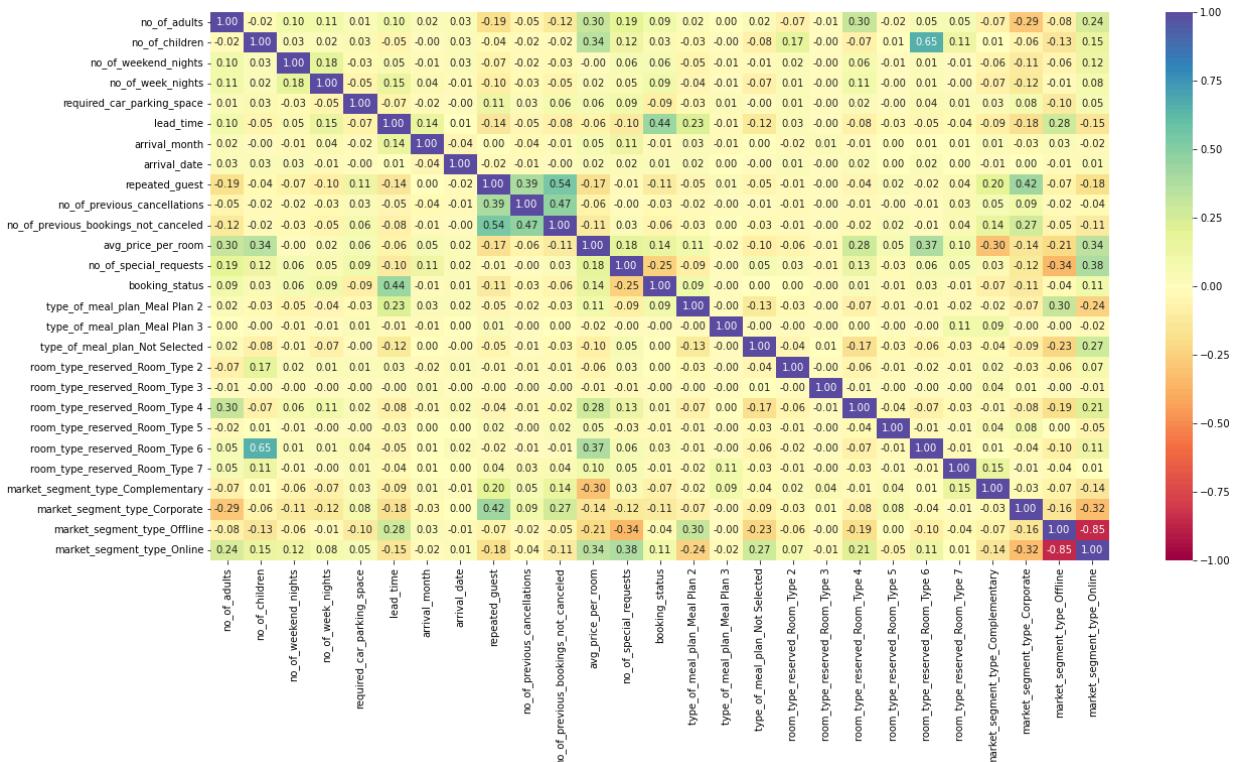
Out[380]:

	no_of_adults	no_of_children	no_of_weekend_nights	no_of_week_nights	required_car_parking_space	
0	2	0		1	2	0
1	2	0		2	3	0
2	1	0		2	1	0
3	2	0		0	2	0
4	2	0		1	1	0

## EDA

- It is a good idea to explore the data once again after manipulating it.

```
In [382... plt.figure(figsize=(20, 10))
sns.heatmap(df.corr(), annot=True, vmin=-1, vmax=1, fmt=".2f", cmap="Spectral")
```

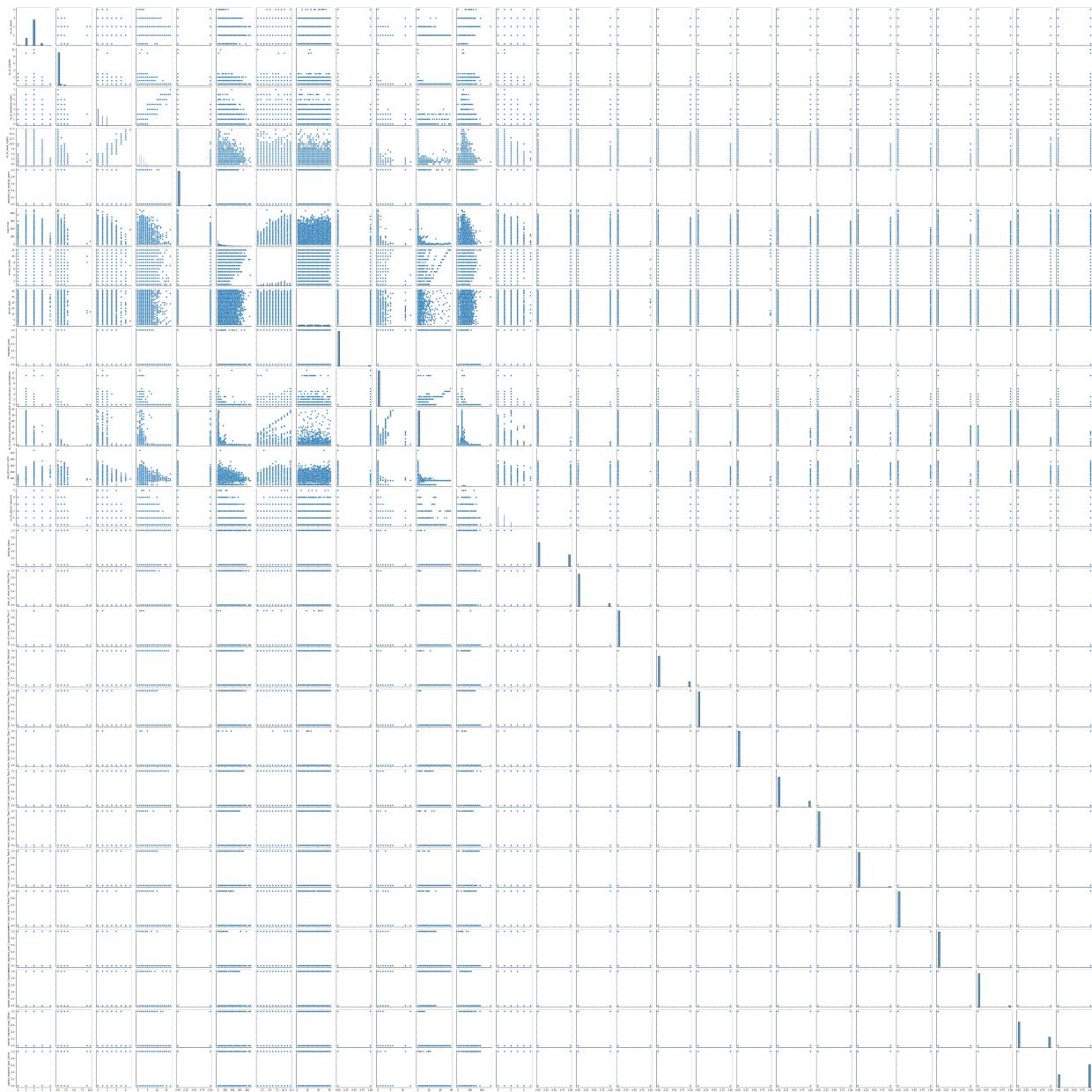


We do not see a very high correlation except:

- a. Number of children and room type 6
  - b. Corporate guests are repeating
  - c. Repeated guests have cancelled fewer bookings
  - d. More the number of special requests, fewer are the cancellations

```
In [383]: sns.pairplot(df)
```

Out[383]: <seaborn.axisgrid.PairGrid at 0x1e350da0f70>



In [384...]

```
df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36275 entries, 0 to 36274
Data columns (total 27 columns):
 #   Column           Non-Null Count Dtype  
--- 
 0   no_of_adults    36275 non-null  int64  
 1   no_of_children  36275 non-null  int64  
 2   no_of_weekend_nights 36275 non-null  int64  
 3   no_of_week_nights 36275 non-null  int64  
 4   required_car_parking_space 36275 non-null  int64  
 5   lead_time        36275 non-null  int64  
 6   arrival_month    36275 non-null  int64  
 7   arrival_date     36275 non-null  int64  
 8   repeated_guest   36275 non-null  int64  
 9   no_of_previous_cancellations 36275 non-null  int64  
 10  no_of_previous_bookings_not_canceled 36275 non-null  int64  
 11  avg_price_per_room 36275 non-null  float64 
 12  no_of_special_requests 36275 non-null  int64  
 13  booking_status   36275 non-null  int64  
 14  type_of_meal_plan_Meal Plan 2 36275 non-null  uint8  
 15  type_of_meal_plan_Meal Plan 3 36275 non-null  uint8  
 16  type_of_meal_plan_Not Selected 36275 non-null  uint8  
 17  room_type_reserved_Room_Type 2 36275 non-null  uint8  
 18  room_type_reserved_Room_Type 3 36275 non-null  uint8  
 19  room_type_reserved_Room_Type 4 36275 non-null  uint8  
 20  room_type_reserved_Room_Type 5 36275 non-null  uint8  
 21  room_type_reserved_Room_Type 6 36275 non-null  uint8  
 22  room_type_reserved_Room_Type 7 36275 non-null  uint8  
 23  market_segment_type_Complementary 36275 non-null  uint8  
 24  market_segment_type_Corporate 36275 non-null  uint8  
 25  market_segment_type_Offline 36275 non-null  uint8  
 26  market_segment_type_Online 36275 non-null  uint8  
dtypes: float64(1), int64(13), uint8(13)
memory usage: 4.3 MB

```

## Checking Multicollinearity

- In order to make statistical inferences from a logistic regression model, it is important to ensure that there is no multicollinearity present in the data.

In [386...]

```

import statsmodels.api as sm

X = df.drop(["booking_status"], axis=1)
y = df["booking_status"]

X = sm.add_constant(X)

```

In [387...]

```

# Splitting data in train and test sets
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.30, random_state=1
)

```

```
In [388...]
print("Shape of Training set : ", X_train.shape)
print("Shape of test set : ", X_test.shape)
print("Percentage of classes in training set:")
print(y_train.value_counts(normalize=True))
print("Percentage of classes in test set:")
print(y_test.value_counts(normalize=True))
```

Shape of Training set : (25392, 27)  
 Shape of test set : (10883, 27)  
 Percentage of classes in training set:  
 0 0.67064  
 1 0.32936  
 Name: booking\_status, dtype: float64  
 Percentage of classes in test set:  
 0 0.67638  
 1 0.32362  
 Name: booking\_status, dtype: float64

```
In [389...]
from statsmodels.stats.outliers_influence import variance_inflation_factor
```

```
vif_series = pd.Series(
    [variance_inflation_factor(X_train.values, i) for i in range(X_train.shape[1])],
    index=X_train.columns,
    dtype=float,
)
print("VIF values: \n\n{}\n".format(vif_series))
```

VIF values:

const	329.50535
no_of_adults	1.34659
no_of_children	1.97764
no_of_weekend_nights	1.06813
no_of_week_nights	1.09561
required_car_parking_space	1.03975
lead_time	1.24630
arrival_month	1.05366
arrival_date	1.00673
repeated_guest	1.78138
no_of_previous_cancellations	1.39456
no_of_previous_bookings_not_canceled	1.64730
avg_price_per_room	1.93621
no_of_special_requests	1.24533
type_of_meal_plan_Meal Plan 2	1.20349
type_of_meal_plan_Meal Plan 3	1.02521
type_of_meal_plan_Not Selected	1.23669
room_type_reserved_Room_Type 2	1.10068
room_type_reserved_Room_Type 3	1.00324
room_type_reserved_Room_Type 4	1.35809
room_type_reserved_Room_Type 5	1.02735
room_type_reserved_Room_Type 6	1.96495
room_type_reserved_Room_Type 7	1.11492
market_segment_type_Complementary	4.49960
market_segment_type_Corporate	16.90216
market_segment_type_Offline	63.95133
market_segment_type_Online	71.05438
dtype: float64	

In [209...]

```
#Dropping columns that have a high VIF value
X_train1 = X_train.drop(['market_segment_type_Online',
                        'market_segment_type_Offline',
                        'market_segment_type_Corporate',
                        'market_segment_type_Complementary'],
                       axis=1)
```

As the columns with higher VIF have been dropped, multicollinearity has been removed.

## Building a Logistic Regression model

In [390...]

```
logit = sm.Logit(y_train, X_train1.astype(float))
lg = logit.fit()

Optimization terminated successfully.
    Current function value: 0.454308
    Iterations 11
```

In [391...]

```
# print the Logistic regression summary
print(lg.summary())
```

### Logit Regression Results

Dep. Variable:	booking_status	No. Observations:	25392
Model:	Logit	Df Residuals:	25369
Method:	MLE	Df Model:	22
Date:	Sat, 10 Dec 2022	Pseudo R-squ.:	0.2831
Time:	00:01:07	Log-Likelihood:	-11536.
converged:	True	LL-Null:	-16091.
Covariance Type:	nonrobust	LLR p-value:	0.000

		coef	std err	z	P> z
[0.025	0.975]				
-----	-----	-----	-----	-----	-----
const		-4.2201	0.106	-39.938	0.000
-4.427	-4.013				
no_of_adults		0.1594	0.035	4.492	0.000
0.090	0.229				
no_of_children		0.2972	0.054	5.477	0.000
0.191	0.404				
no_of_weekend_nights		0.1630	0.019	8.617	0.000
0.126	0.200				
no_of_week_nights		0.0683	0.012	5.771	0.000
0.045	0.092				
required_car_parking_space		-1.4152	0.135	-10.480	0.000
-1.680	-1.151				
lead_time		0.0143	0.000	61.099	0.000
0.014	0.015				
arrival_month		-0.0727	0.006	-12.465	0.000
-0.084	-0.061				
arrival_date		0.0017	0.002	0.915	0.360
-0.002	0.005				
repeated_guest		-2.1130	0.514	-4.114	0.000
-3.120	-1.106				
no_of_previous_cancellations		0.2408	0.091	2.635	0.008
0.062	0.420				
no_of_previous_bookings_not_canceled		-0.1479	0.136	-1.085	0.278
-0.415	0.119				
avg_price_per_room		0.0249	0.001	35.962	0.000
0.024	0.026				
no_of_special_requests		-1.1372	0.027	-42.217	0.000
-1.190	-1.084				
type_of_meal_plan_Meal Plan 2		-0.5700	0.059	-9.715	0.000
-0.685	-0.455				
type_of_meal_plan_Meal Plan 3		1.9398	3.471	0.559	0.576
-4.863	8.743				
type_of_meal_plan_Not Selected		0.9195	0.048	18.973	0.000
0.824	1.014				
room_type_reserved_Room_Type 2		0.1496	0.124	1.210	0.226
-0.093	0.392				
room_type_reserved_Room_Type 3		0.0280	1.131	0.025	0.980
-2.190	2.246				
room_type_reserved_Room_Type 4		0.0324	0.050	0.647	0.517
-0.066	0.131				
room_type_reserved_Room_Type 5		-0.8380	0.199	-4.214	0.000
-1.228	-0.448				
room_type_reserved_Room_Type 6		-1.1504	0.140	-8.201	0.000
-1.425	-0.875				
room_type_reserved_Room_Type 7		-1.8374	0.290	-6.340	0.000

```
-2.405      -1.269
=====
=====
```

In [392...]

```
# Let's check the VIF of the predictors
vif_series = pd.Series(
    [variance_inflation_factor(X_train1.values, i) for i in range(X_train1.shape[1])],
    index=X_train1.columns,
    dtype=float,
)
print("VIF values: \n\n{}\n".format(vif_series))
```

VIF values:

```
const                      32.69202
no_of_adults                1.28890
no_of_children               1.96840
no_of_weekend_nights         1.05398
no_of_week_nights             1.08659
required_car_parking_space   1.03539
lead_time                     1.19662
arrival_month                  1.04880
arrival_date                   1.00576
repeated_guest                 1.57986
no_of_previous_cancellations  1.37613
no_of_previous_bookings_not_canceled 1.63196
avg_price_per_room            1.57829
no_of_special_requests        1.12703
type_of_meal_plan_Meal Plan 2 1.12788
type_of_meal_plan_Meal Plan 3 1.01862
type_of_meal_plan_Not Selected 1.10978
room_type_reserved_Room_Type 2 1.08911
room_type_reserved_Room_Type 3 1.00080
room_type_reserved_Room_Type 4 1.31233
room_type_reserved_Room_Type 5 1.01290
room_type_reserved_Room_Type 6 1.93863
room_type_reserved_Room_Type 7 1.07588
dtype: float64
```

## Observations

- Negative values of the coefficient show that the probability of a cancellation decreases with the increase of the corresponding attribute value.
- Positive values of the coefficient show that the probability of a cancellation increases with the increase of the 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.

Dropping columns with a higher p-value room\_type\_reserved\_Room\_Type 3 ,  
type\_of\_meal\_plan\_Meal Plan 3 , room\_type\_reserved\_Room\_Type 4

```
In [393... X_train2 = X_train1.drop(
```

```
[  
    "room_type_reserved_Room_Type 3",  
    "type_of_meal_plan_Meal Plan 3",  
    "room_type_reserved_Room_Type 4",  
,  
    axis=1,  
)
```

```
In [394... logit1 = sm.Logit(y_train, X_train2.astype(float))
```

```
lg1 = logit1.fit(disp=False)  
print(lg1.summary())
```

### Logit Regression Results

Dep. Variable:	booking_status	No. Observations:	25392
Model:	Logit	Df Residuals:	25372
Method:	MLE	Df Model:	19
Date:	Sat, 10 Dec 2022	Pseudo R-squ.:	0.2831
Time:	00:02:01	Log-Likelihood:	-11536.
converged:	True	LL-Null:	-16091.
Covariance Type:	nonrobust	LLR p-value:	0.000
<hr/>		<hr/>	
<hr/>		coef	std err
<hr/>		z	P> z
[0.025	0.975]		
<hr/>		<hr/>	
const		-4.2384	0.102
-4.438	-4.038	-41.542	0.000
no_of_adults		0.1647	0.035
0.097	0.232	4.774	0.000
no_of_children		0.2948	0.054
0.189	0.401	5.447	0.000
no_of_weekend_nights		0.1635	0.019
0.127	0.201	8.656	0.000
no_of_week_nights		0.0690	0.012
0.046	0.092	5.854	0.000
required_car_parking_space		-1.4154	0.135
-1.680	-1.151	-10.483	0.000
lead_time		0.0142	0.000
0.014	0.015	61.201	0.000
arrival_month		-0.0728	0.006
-0.084	-0.061	-12.495	0.000
arrival_date		0.0017	0.002
-0.002	0.005	0.932	0.351
repeated_guest		-2.1108	0.513
-3.117	-1.105	-4.111	0.000
no_of_previous_cancellations		0.2406	0.091
0.061	0.420	2.631	0.009
no_of_previous_bookings_not_canceled		-0.1477	0.136
-0.415	0.120	-1.083	0.279
avg_price_per_room		0.0251	0.001
0.024	0.026	38.119	0.000
no_of_special_requests		-1.1365	0.027
-1.189	-1.084	-42.241	0.000
type_of_meal_plan_Meal Plan 2		-0.5758	0.058
-0.690	-0.462	-9.920	0.000
type_of_meal_plan_Not Selected		0.9128	0.047
0.820	1.006	19.273	0.000
room_type_reserved_Room_Type 2		0.1472	0.124
-0.095	0.389	1.191	0.234
room_type_reserved_Room_Type 5		-0.8486	0.198
-1.237	-0.460	-4.281	0.000
room_type_reserved_Room_Type 6		-1.1670	0.138
-1.437	-0.897	-8.460	0.000
room_type_reserved_Room_Type 7		-1.8610	0.287
-2.424	-1.298	-6.475	0.000
<hr/>		<hr/>	
<hr/>		<hr/>	

Dropping other columns with higher p-value

```
In [395...]: X_train3 = X_train2.drop(  
    [  
        "arrival_date",  
        "no_of_previous_bookings_not_canceled",  
        "room_type_reserved_Room_Type_2",  
    ],  
    axis=1,  
)
```

```
In [396...]: logit2 = sm.Logit(y_train, X_train3.astype(float))  
lg2 = logit2.fit(disp=False)  
print(lg2.summary())
```

### Logit Regression Results

Dep. Variable:	booking_status	No. Observations:	25392			
Model:	Logit	Df Residuals:	25375			
Method:	MLE	Df Model:	16			
Date:	Sat, 10 Dec 2022	Pseudo R-squ.:	0.2829			
Time:	00:02:42	Log-Likelihood:	-11539.			
converged:	True	LL-Null:	-16091.			
Covariance Type:	nonrobust	LLR p-value:	0.000			
<hr/>						
<hr/>						
		coef	std err	z		
5	0.975]				P> z  [0.02	
<hr/>						
<hr/>						
const		-4.2017	0.097	-43.231	0.000	-4.39
2	-4.011					
no_of_adults		0.1636	0.034	4.748	0.000	0.09
6	0.231					
no_of_children		0.3106	0.053	5.874	0.000	0.20
7	0.414					
no_of_weekend_nights		0.1642	0.019	8.695	0.000	0.12
7	0.201					
no_of_week_nights		0.0688	0.012	5.841	0.000	0.04
6	0.092					
required_car_parking_space		-1.4114	0.135	-10.464	0.000	-1.67
6	-1.147					
lead_time		0.0143	0.000	61.327	0.000	0.01
4	0.015					
arrival_month		-0.0733	0.006	-12.602	0.000	-0.08
5	-0.062					
repeated_guest		-2.3936	0.475	-5.039	0.000	-3.32
5	-1.463					
no_of_previous_cancellations		0.1963	0.077	2.546	0.011	0.04
5	0.347					
avg_price_per_room		0.0250	0.001	38.150	0.000	0.02
4	0.026					
no_of_special_requests		-1.1356	0.027	-42.245	0.000	-1.18
8	-1.083					
type_of_meal_plan_Meal Plan 2		-0.5779	0.058	-9.962	0.000	-0.69
2	-0.464					
type_of_meal_plan_Not Selected		0.9111	0.047	19.257	0.000	0.81
8	1.004					
room_type_reserved_Room_Type 5		-0.8511	0.198	-4.296	0.000	-1.23
9	-0.463					
room_type_reserved_Room_Type 6		-1.1915	0.137	-8.711	0.000	-1.46
0	-0.923					
room_type_reserved_Room_Type 7		-1.8728	0.288	-6.511	0.000	-2.43
7	-1.309					
<hr/>						
<hr/>						

**Now no feature has p-value greater than 0.05, so we'll consider the features in  $X_{train3}$  as the final ones and  $lg2$  as final model.**

**Converting coefficients to odds**

- The coefficients of the logistic regression model are in terms of log(odd), to find the odds we have to take the exponential of the coefficients.
- Therefore, **odds = exp(b)**
- The percentage change in odds is given as **odds = (exp(b) - 1) \* 100**

In [397...]

```
# converting coefficients to odds
odds = np.exp(lg2.params)

# finding the percentage change
perc_change_odds = (np.exp(lg2.params) - 1) * 100

# removing limit from number of columns to display
pd.set_option("display.max_columns", None)

# adding the odds to a dataframe
pd.DataFrame({"Odds": odds, "Change_odd%": perc_change_odds}, index=X_train3.columns).
```

Out[397]:

	const	no_of_adults	no_of_children	no_of_weekend_nights	no_of_week_nights	reqt
Odds	0.01497	1.17779	1.36422	1.17846	1.07125	
Change_odd%	-98.50294	17.77932	36.42164	17.84617	7.12522	

In [ ]:

## Model performance evaluation

In [398...]

```
# defining a function to compute different metrics to check performance of a classifier
def model_performance_classification_statsmodels(
    model, predictors, target, threshold=0.5
):
    """
    Function to compute different metrics to check classification model performance

    model: classifier
    predictors: independent variables
    target: dependent variable
    threshold: threshold for classifying the observation as class 1
    """

    # checking which probabilities are greater than threshold
    pred_temp = model.predict(predictors) > threshold
    # rounding off the above values to get classes
    pred = np.round(pred_temp)

    acc = accuracy_score(target, pred) # to compute Accuracy
    recall = recall_score(target, pred) # to compute Recall
    precision = precision_score(target, pred) # to compute Precision
    f1 = f1_score(target, pred) # to compute F1-score

    # creating a dataframe of metrics
    df_perf = pd.DataFrame(
```

```

        {"Accuracy": acc, "Recall": recall, "Precision": precision, "F1": f1, },
        index=[0],
    )

return df_perf

```

In [399]: # defining a function to plot the confusion\_matrix of a classification model

```

def confusion_matrix_statsmodels(model, predictors, target, threshold=0.5):
    """
    To plot the confusion_matrix with percentages

    model: classifier
    predictors: independent variables
    target: dependent variable
    threshold: threshold for classifying the observation as class 1
    """

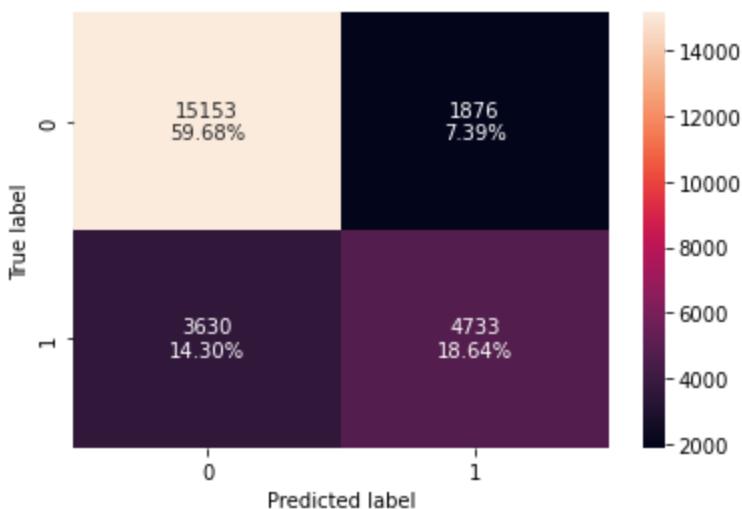
    y_pred = model.predict(predictors) > threshold
    cm = confusion_matrix(target, y_pred)
    labels = np.asarray(
        [
            ["{0:0.0f} \n {0:.2%}".format(item) + "\n" for item in cm.flatten()]
        ]
    ).reshape(2, 2)

    plt.figure(figsize=(6, 4))
    sns.heatmap(cm, annot=labels, fmt="")
    plt.ylabel("True label")
    plt.xlabel("Predicted label")

```

In [ ]:

In [400...]: # creating confusion matrix  
confusion\_matrix\_statsmodels(lg2, X\_train3, y\_train)



In [401...]: log\_reg\_model\_train\_perf = model\_performance\_classification\_statsmodels(  
lg2, X\_train3, y\_train)

```
)  
  
print("Training performance:")  
log_reg_model_train_perf
```

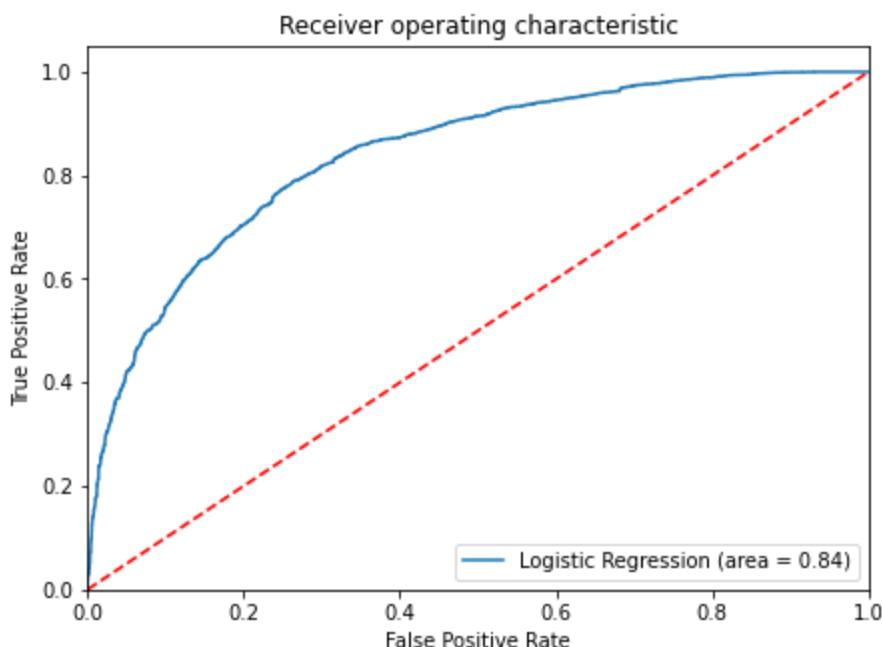
Training performance:

Out[401]:

	Accuracy	Recall	Precision	F1
0	0.78316	0.56595	0.71614	0.63225

In [402...]

```
from sklearn.metrics import (  
    f1_score,  
    accuracy_score,  
    recall_score,  
    precision_score,  
    confusion_matrix,  
    plot_confusion_matrix,  
    make_scorer,  
    roc_auc_score,  
    roc_curve,  
)  
  
logit_roc_auc_train = roc_auc_score(y_train, lg2.predict(X_train3))  
fpr, tpr, thresholds = roc_curve(y_train, lg2.predict(X_train3))  
plt.figure(figsize=(7, 5))  
plt.plot(fpr, tpr, label="Logistic Regression (area = %0.2f)" % logit_roc_auc_train)  
plt.plot([0, 1], [0, 1], "r--")  
plt.xlim([0.0, 1.0])  
plt.ylim([0.0, 1.05])  
plt.xlabel("False Positive Rate")  
plt.ylabel("True Positive Rate")  
plt.title("Receiver operating characteristic")  
plt.legend(loc="lower right")  
plt.show()
```



## Model Performance Improvement

In [403...]

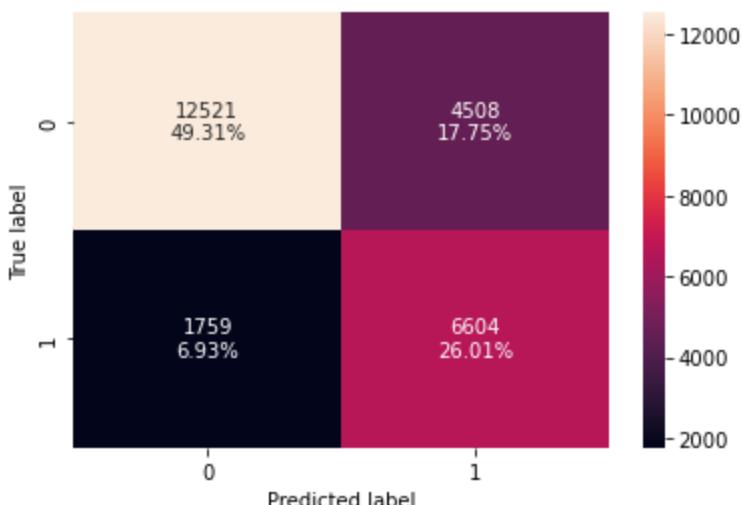
```
# Optimal threshold as per AUC-ROC curve
# The optimal cut off would be where tpr is high and fpr is low
fpr, tpr, thresholds = roc_curve(y_train, lg2.predict(X_train3))

optimal_idx = np.argmax(tpr - fpr)
optimal_threshold_auc_roc = thresholds[optimal_idx]
print(optimal_threshold_auc_roc)
```

0.298245547374405

In [404...]

```
# creating confusion matrix
confusion_matrix_statsmodels(
    lg2, X_train3, y_train, threshold=optimal_threshold_auc_roc
)
```



In [405...]

```
# checking model performance for this model
log_reg_model_train_perf_threshold_auc_roc = model_performance_classification_statsmodels(
    lg2, X_train3, y_train, threshold=optimal_threshold_auc_roc
)
print("Training performance:")
log_reg_model_train_perf_threshold_auc_roc
```

Training performance:

Out[405]:

	Accuracy	Recall	Precision	F1
0	0.75319	0.78967	0.59431	0.67820

- Model performance has improved significantly.
- Model is giving a recall of 0.78 as compared to initial model which was giving a recall of 0.56
- Precision has decreased from 0.71 to 0.59.

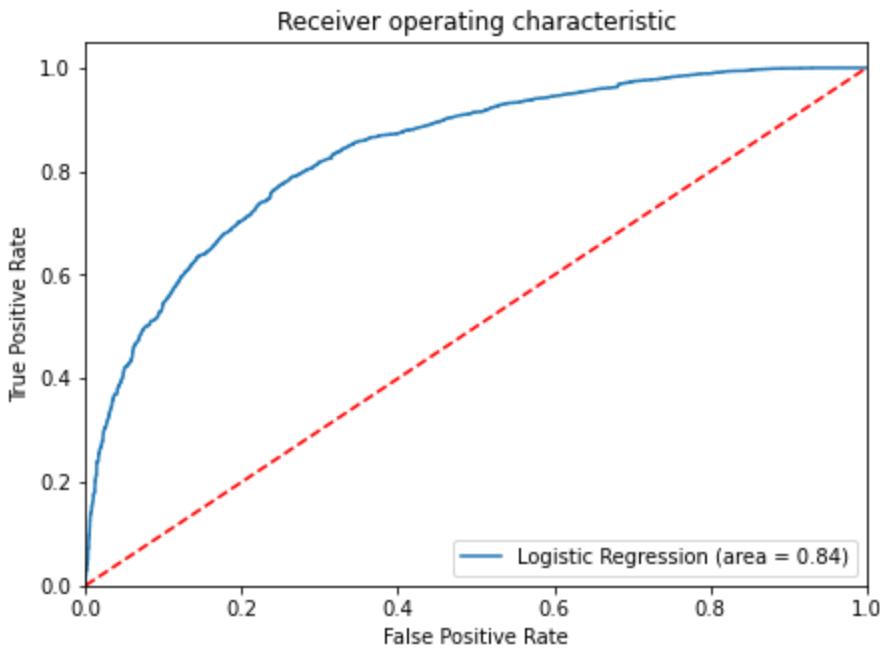
In [406...]

```
logit_roc_auc_train = roc_auc_score(y_train, lg2.predict(X_train3))
fpr, tpr, thresholds = roc_curve(y_train, lg2.predict(X_train3))
```

```

plt.figure(figsize=(7, 5))
plt.plot(fpr, tpr, label="Logistic Regression (area = %0.2f)" % logit_roc_auc_train)
plt.plot([0, 1], [0, 1], "r--")
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel("False Positive Rate")
plt.ylabel("True Positive Rate")
plt.title("Receiver operating characteristic")
plt.legend(loc="lower right")
plt.show()

```



Let's use Precision-Recall curve and see if we can find a better threshold

```

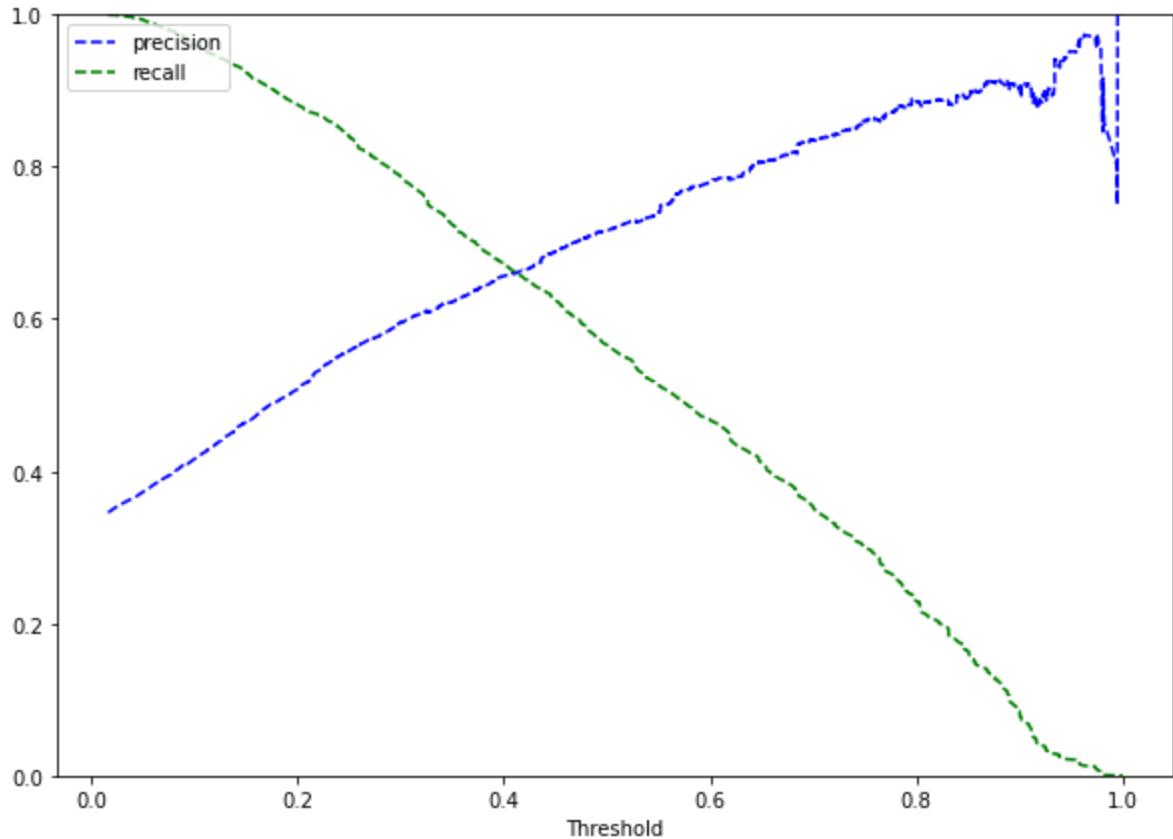
In [407...]: from sklearn.metrics import (
    f1_score,
    accuracy_score,
    recall_score,
    precision_score,
    confusion_matrix,
    roc_auc_score,
    plot_confusion_matrix,
    precision_recall_curve,
    roc_curve,
)

y_scores = lg2.predict(X_train3)
prec, rec, tre = precision_recall_curve(y_train, y_scores)

def plot_prec_recall_vs_thresh(precisions, recalls, thresholds):
    plt.plot(thresholds, precisions[:-1], "b--", label="precision")
    plt.plot(thresholds, recalls[:-1], "g--", label="recall")
    plt.xlabel("Threshold")
    plt.legend(loc="upper left")
    plt.ylim([0, 1])

```

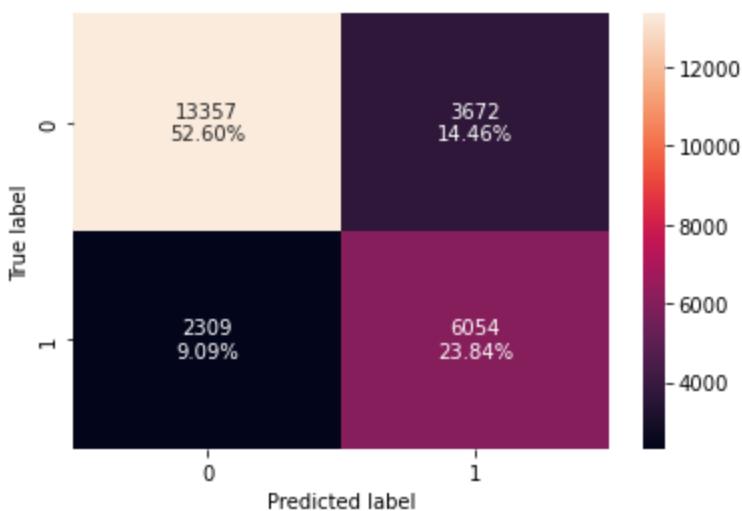
```
plt.figure(figsize=(10, 7))
plot_prec_recall_vs_thresh(prec, rec, tre)
plt.show()
```



- At threshold around 0.41 we will get equal precision and recall but taking a step back and selecting value around 0.35 will provide a higher recall and a good precision.

```
In [408...]: # setting the threshold
optimal_threshold_curve = 0.35
```

```
In [409...]: # creating confusion matrix
confusion_matrix_statsmodels(lg2, X_train3, y_train, threshold=optimal_threshold_curve)
```



```
In [410...]: log_reg_model_train_perf_threshold_curve = model_performance_classification_statsmodel
    lg2, X_train3, y_train, threshold=optimal_threshold_curve
)
print("Training performance:")
log_reg_model_train_perf_threshold_curve
```

Training performance:

```
Out[410]: Accuracy Recall Precision F1
0 0.76445 0.72390 0.62246 0.66936
```

- Recall has improved as compared to the initial model.
- Model is giving a better performance with threshold found using AUC-ROC curve.

## Model Performance Summary

```
In [411...]: # training performance comparison

models_train_comp_df = pd.concat(
    [
        log_reg_model_train_perf.T,
        log_reg_model_train_perf_threshold_auc_roc.T,
        log_reg_model_train_perf_threshold_curve.T,
    ],
    axis=1,
)
models_train_comp_df.columns = [
    "Logistic Regression statsmodel",
    "Logistic Regression-0.29 Threshold",
    "Logistic Regression-0.35 Threshold",
]
print("Training performance comparison:")
models_train_comp_df
```

Training performance comparison:

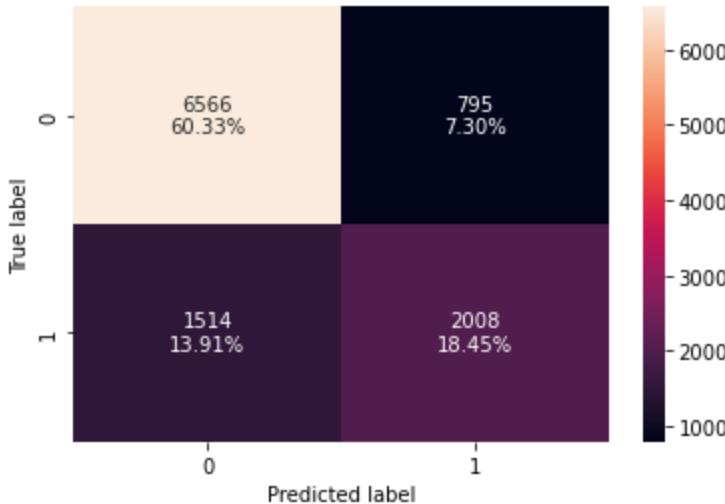
	Logistic Regression statsmodel	Logistic Regression-0.29 Threshold	Logistic Regression-0.35 Threshold
<b>Accuracy</b>	0.78316	0.75319	0.76445
<b>Recall</b>	0.56595	0.78967	0.72390
<b>Precision</b>	0.71614	0.59431	0.62246
<b>F1</b>	0.63225	0.67820	0.66936

## Let's check the performance on the test set

Dropping the columns from the test set that were dropped from the training set

```
In [412]: X_test3 = X_test[X_train3.columns].astype(float)
```

```
In [413]: # creating confusion matrix  
confusion_matrix_statsmodels(lg2, X_test3, y_test)
```



```
In [414]: log_reg_model_test_perf = model_performance_classification_statsmodels(  
    lg2, X_test3, y_test  
)  
  
print("Test performance:")  
log_reg_model_test_perf
```

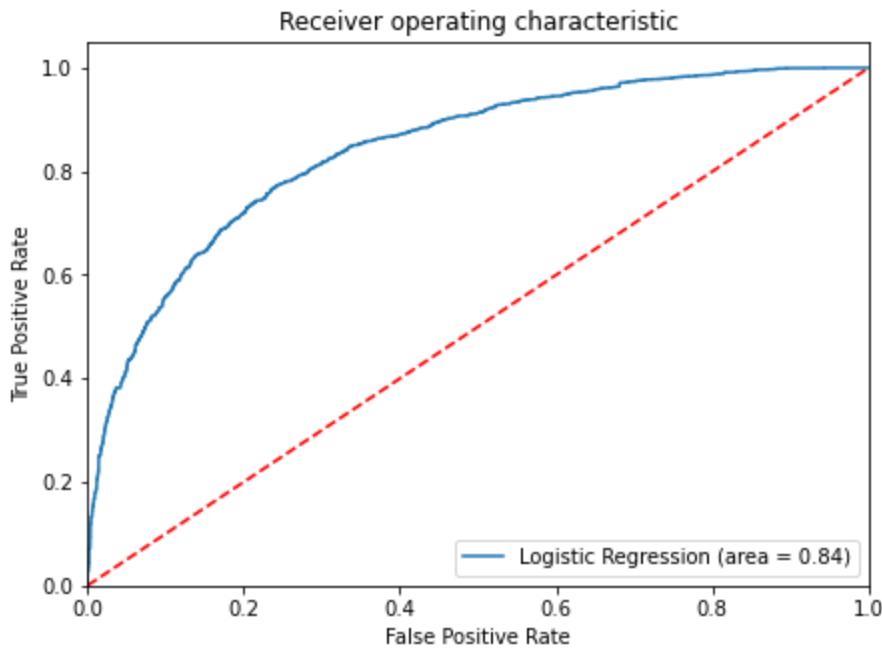
Test performance:

```
Out[414]:
```

	Accuracy	Recall	Precision	F1
0	0.78783	0.57013	0.71638	0.63494

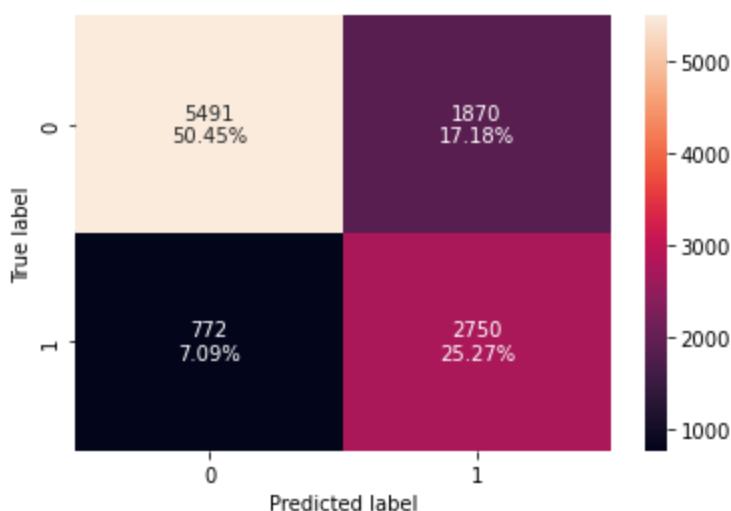
- ROC curve on test set

```
In [415]: logit_roc_auc_train = roc_auc_score(y_test, lg2.predict(X_test3))  
fpr, tpr, thresholds = roc_curve(y_test, lg2.predict(X_test3))  
plt.figure(figsize=(7, 5))  
plt.plot(fpr, tpr, label="Logistic Regression (area = %0.2f)" % logit_roc_auc_train)  
plt.plot([0, 1], [0, 1], "r--")  
plt.xlim([0.0, 1.0])  
plt.ylim([0.0, 1.05])  
plt.xlabel("False Positive Rate")  
plt.ylabel("True Positive Rate")  
plt.title("Receiver operating characteristic")  
plt.legend(loc="lower right")  
plt.show()
```



In [416]:

```
# creating confusion matrix
confusion_matrix_statsmodels(lg2, X_test3, y_test, threshold=optimal_threshold_auc_roc)
```



In [417]:

```
# checking model performance for this model
log_reg_model_test_perf_threshold_auc_roc = model_performance_classification_statsmodels(
    lg2, X_test3, y_test, threshold=optimal_threshold_auc_roc
)
print("Test performance:")
log_reg_model_test_perf_threshold_auc_roc
```

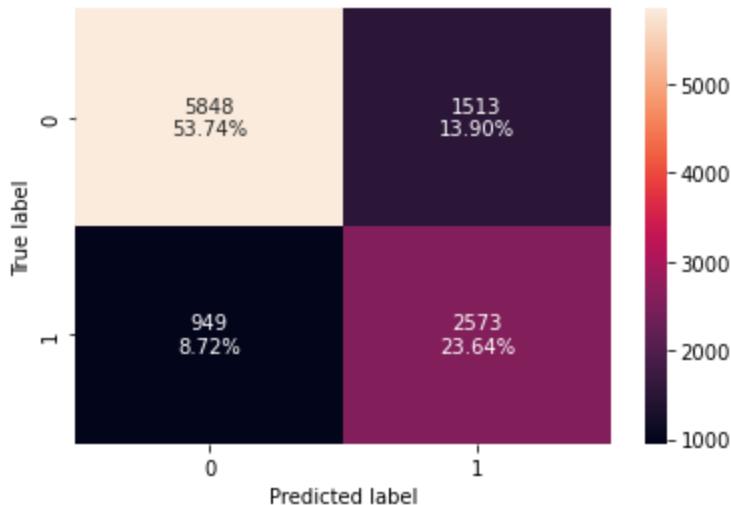
Test performance:

Out[417]:

	Accuracy	Recall	Precision	F1
0	0.75724	0.78081	0.59524	0.67551

In [418]:

```
# creating confusion matrix
confusion_matrix_statsmodels(lg2, X_test3, y_test, threshold=optimal_threshold_curve)
```



```
In [419]: log_reg_model_test_perf_threshold_curve = model_performance_classification_statsmodels
          lg2, X_test3, y_test, threshold=optimal_threshold_curve
      )
print("Test performance:")
log_reg_model_test_perf_threshold_curve
```

Test performance:

	Accuracy	Recall	Precision	F1
0	0.77378	0.73055	0.62971	0.67639

## Model performance summary

```
In [420]: # training performance comparison

models_train_comp_df = pd.concat(
    [
        log_reg_model_train_perf.T,
        log_reg_model_train_perf_threshold_auc_roc.T,
        log_reg_model_train_perf_threshold_curve.T,
    ],
    axis=1,
)
models_train_comp_df.columns = [
    "Logistic Regression statsmodel",
    "Logistic Regression-0.29 Threshold",
    "Logistic Regression-0.35 Threshold",
]
print("Training performance comparison:")
models_train_comp_df
```

Training performance comparison:

Out[420]:

	Logistic Regression statsmodel	Logistic Regression-0.29 Threshold	Logistic Regression-0.35 Threshold
<b>Accuracy</b>	0.78316	0.75319	0.76445
<b>Recall</b>	0.56595	0.78967	0.72390
<b>Precision</b>	0.71614	0.59431	0.62246
<b>F1</b>	0.63225	0.67820	0.66936

In [422...]

```
# testing performance comparison

models_test_comp_df = pd.concat(
    [
        log_reg_model_test_perf.T,
        log_reg_model_test_perf_threshold_auc_roc.T,
        log_reg_model_test_perf_threshold_curve.T,
    ],
    axis=1,
)
models_test_comp_df.columns = [
    "Logistic Regression statsmodel",
    "Logistic Regression-0.29 Threshold",
    "Logistic Regression-0.35 Threshold",
]

print("Test set performance comparison:")
models_test_comp_df
```

Test set performance comparison:

Out[422]:

	Logistic Regression statsmodel	Logistic Regression-0.29 Threshold	Logistic Regression-0.35 Threshold
<b>Accuracy</b>	0.78783	0.75724	0.77378
<b>Recall</b>	0.57013	0.78081	0.73055
<b>Precision</b>	0.71638	0.59524	0.62971
<b>F1</b>	0.63494	0.67551	0.67639

## Final Model Summary

The model with the optimal\_threshold\_auc\_roc of 0.29 has the best F1 score. So it can be selected as the best model.

## Building a Decision Tree model

In [423...]

```
# defining a function to compute different metrics to check performance of a classifier
def model_performance_classification_sklearn(model, predictors, target):
    """
    Function to compute different metrics to check classification model performance
    """

    # Compute metrics
```

```

model: classifier
predictors: independent variables
target: dependent variable
"""

# predicting using the independent variables
pred = model.predict(predictors)

acc = accuracy_score(target, pred) # to compute Accuracy
recall = recall_score(target, pred) # to compute Recall
precision = precision_score(target, pred) # to compute Precision
f1 = f1_score(target, pred) # to compute F1-score

# creating a dataframe of metrics
df_perf = pd.DataFrame(
    {"Accuracy": acc, "Recall": recall, "Precision": precision, "F1": f1},
    index=[0],
)
)

return df_perf

```

In [424]:

```

def confusion_matrix_sklearn(model, predictors, target):
    """
    To plot the confusion_matrix with percentages

    model: classifier
    predictors: independent variables
    target: dependent variable
    """

    y_pred = model.predict(predictors)
    cm = confusion_matrix(target, y_pred)
    labels = np.asarray([
        ["{0:0.0f}" .format(item) + "\n{0:.2%}" .format(item / cm.flatten().sum())]
        for item in cm.flatten()
    ])
    ).reshape(2, 2)

    plt.figure(figsize=(6, 4))
    sns.heatmap(cm, annot=labels, fmt="")
    plt.ylabel("True label")
    plt.xlabel("Predicted label")

```

In [426]:

```

model = DecisionTreeClassifier(criterion="gini", random_state=1)
model.fit(X_train3, y_train)

```

Out[426]:

```
DecisionTreeClassifier(random_state=1)
```

In [427]:

```

decision_tree_perf_train = model_performance_classification_sklearn(
    model, X_train3, y_train
)
decision_tree_perf_train

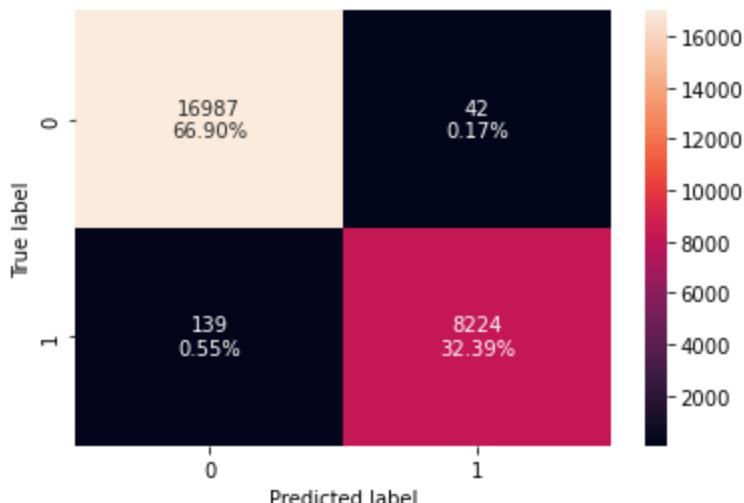
```

	Accuracy	Recall	Precision	F1
0	0.99287	0.98338	0.99492	0.98912

```
In [429...]: print("Accuracy on training set : ", model.score(X_train3, y_train))
print("Accuracy on test set : ", model.score(X_test3, y_test))
```

Accuracy on training set : 0.992871770636421  
 Accuracy on test set : 0.8598731967288431

```
In [430...]: confusion_matrix_sklearn(model, X_train3, y_train)
```

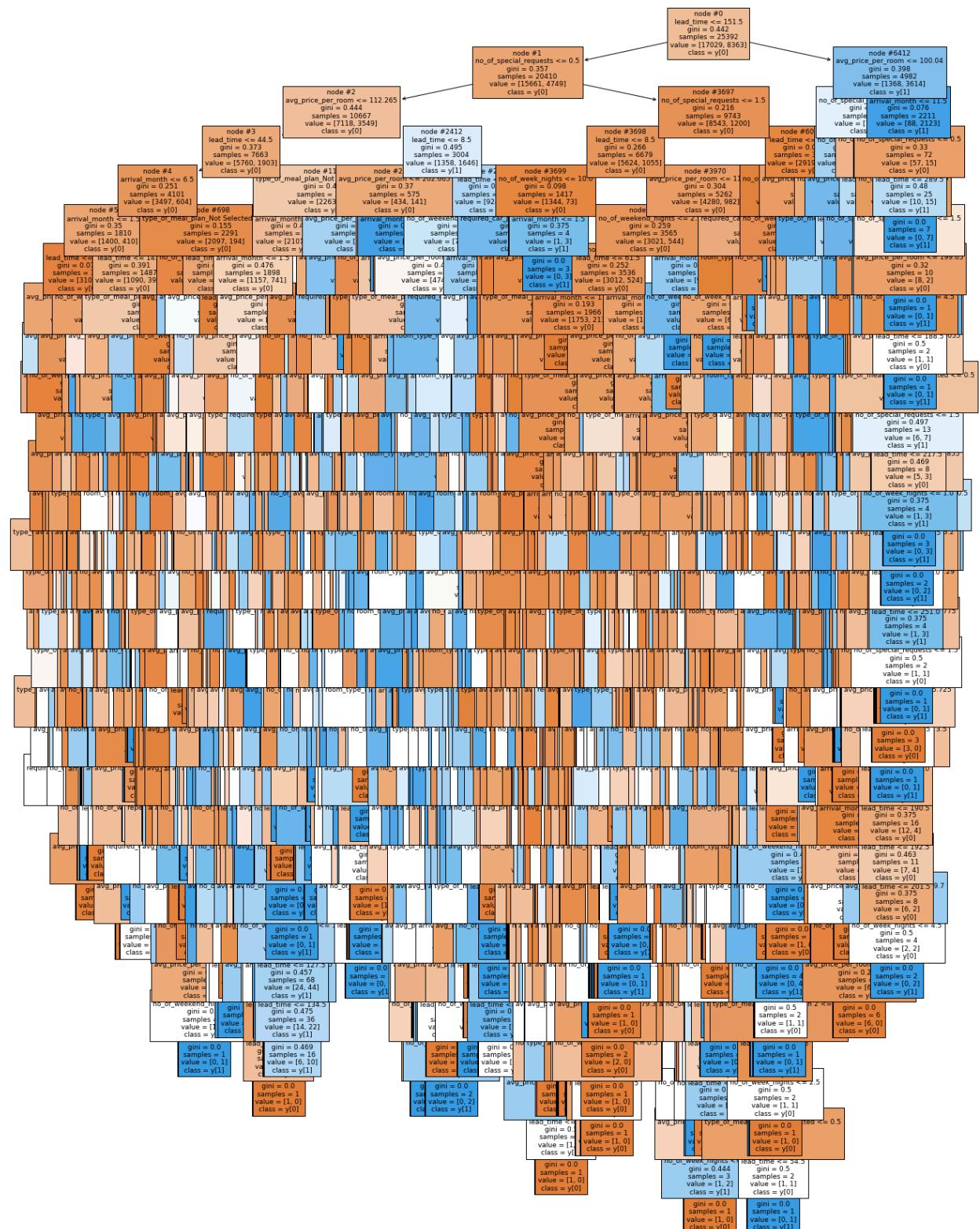


```
In [431...]: column_names = list(X_train3.columns)
feature_names = column_names
print(feature_names)
```

['const', 'no\_of\_adults', 'no\_of\_children', 'no\_of\_weekend\_nights', 'no\_of\_week\_nights', 'required\_car\_parking\_space', 'lead\_time', 'arrival\_month', 'repeated\_guest', 'no\_of\_previous\_cancellations', 'avg\_price\_per\_room', 'no\_of\_special\_requests', 'type\_of\_meal\_plan\_Meal Plan 2', 'type\_of\_meal\_plan\_Not Selected', 'room\_type\_reserved\_Room\_Type 5', 'room\_type\_reserved\_Room\_Type 6', 'room\_type\_reserved\_Room\_Type 7']

```
In [432...]: plt.figure(figsize=(20, 30))

out = tree.plot_tree(
    model,
    feature_names=feature_names,
    filled=True,
    fontsize=9,
    node_ids=True,
    class_names=True,
)
for o in out:
    arrow = o.arrow_patch
    if arrow is not None:
        arrow.set_edgecolor("black")
        arrow.set linewidth(1)
plt.show()
```



```
|--- lead_time <= 151.50
|--- no_of_special_requests <= 0.50
|   |--- avg_price_per_room <= 112.26
|   |--- lead_time <= 44.50
|   |   |--- arrival_month <= 6.50
|   |   |--- arrival_month <= 1.50
|   |   |   |--- lead_time <= 28.50
|   |   |   |   |--- weights: [268.00, 0.00] class: 0
|   |   |--- lead_time > 28.50
|   |   |   |--- avg_price_per_room <= 75.95
|   |   |   |   |--- avg_price_per_room <= 63.00
|   |   |   |   |   |--- avg_price_per_room <= 60.15
|   |   |   |   |   |   |--- weights: [17.00, 0.00] class: 0
|   |   |   |   |   |--- avg_price_per_room > 60.15
|   |   |   |   |   |   |--- weights: [0.00, 6.00] class: 1
|   |   |   |   |--- avg_price_per_room > 63.00
|   |   |   |   |   |--- weights: [20.00, 0.00] class: 0
|   |   |--- avg_price_per_room > 75.95
|   |   |   |--- avg_price_per_room <= 89.22
|   |   |   |   |--- no_of_adults <= 1.50
|   |   |   |   |   |--- weights: [1.00, 0.00] class: 0
|   |   |   |   |--- no_of_adults > 1.50
|   |   |   |   |   |--- weights: [0.00, 7.00] class: 1
|   |   |   |--- avg_price_per_room > 89.22
|   |   |   |   |--- weights: [4.00, 0.00] class: 0
|--- arrival_month > 1.50
|   |--- lead_time <= 14.50
|   |   |--- no_of_weekend_nights <= 1.50
|   |   |   |--- avg_price_per_room <= 78.90
|   |   |   |   |--- no_of_weekend_nights <= 0.50
|   |   |   |   |   |--- lead_time <= 8.50
|   |   |   |   |   |   |--- truncated branch of depth 7
|   |   |   |   |   |--- lead_time > 8.50
|   |   |   |   |   |   |--- truncated branch of depth 5
|   |   |   |--- no_of_weekend_nights > 0.50
|   |   |   |   |--- arrival_month <= 3.50
|   |   |   |   |   |--- truncated branch of depth 8
|   |   |   |   |--- arrival_month > 3.50
|   |   |   |   |   |--- weights: [32.00, 0.00] class: 0
|   |--- avg_price_per_room > 78.90
|   |   |--- no_of_adults <= 1.50
|   |   |   |--- avg_price_per_room <= 92.85
|   |   |   |   |--- truncated branch of depth 10
|   |   |   |   |--- avg_price_per_room > 92.85
|   |   |   |   |   |--- truncated branch of depth 11
|   |--- no_of_adults > 1.50
|   |   |--- lead_time <= 4.50
|   |   |   |--- truncated branch of depth 12
|   |   |--- lead_time > 4.50
|   |   |   |   |--- truncated branch of depth 11
|--- no_of_weekend_nights > 1.50
|   |--- avg_price_per_room <= 61.88
|   |   |--- avg_price_per_room <= 57.50
|   |   |   |--- weights: [5.00, 0.00] class: 0
|   |   |--- avg_price_per_room > 57.50
|   |   |   |--- weights: [0.00, 42.00] class: 1
|   |--- avg_price_per_room > 61.88
|   |   |--- no_of_week_nights <= 5.50
|   |   |   |--- avg_price_per_room <= 88.63
|   |   |   |   |--- truncated branch of depth 7
```

```

    |--- avg_price_per_room > 88.63
    |   |--- truncated branch of depth 9
    --- no_of_week_nights > 5.50
    |--- no_of_weekend_nights <= 4.50
    |   |--- truncated branch of depth 3
    |--- no_of_weekend_nights > 4.50
    |   |--- weights: [0.00, 7.00] class: 1
--- lead_time > 14.50
|--- type_of_meal_plan_Not Selected <= 0.50
|   |--- avg_price_per_room <= 100.87
|       |--- avg_price_per_room <= 75.15
|           |--- no_of_adults <= 1.50
|               |--- truncated branch of depth 5
|           |--- no_of_adults > 1.50
|               |--- truncated branch of depth 8
|       |--- avg_price_per_room > 75.15
|           |--- avg_price_per_room <= 85.17
|               |--- truncated branch of depth 11
|           |--- avg_price_per_room > 85.17
|               |--- truncated branch of depth 13
|       |--- avg_price_per_room > 100.87
|           |--- lead_time <= 19.50
|               |--- type_of_meal_plan_Meal Plan 2 <= 0.50
|                   |--- weights: [0.00, 13.00] class: 1
|               |--- type_of_meal_plan_Meal Plan 2 > 0.50
|                   |--- weights: [2.00, 0.00] class: 0
|           |--- lead_time > 19.50
|               |--- arrival_month <= 5.50
|                   |--- truncated branch of depth 8
|               |--- arrival_month > 5.50
|                   |--- truncated branch of depth 3
|       |--- type_of_meal_plan_Not Selected > 0.50
|           |--- avg_price_per_room <= 85.99
|               |--- arrival_month <= 2.50
|                   |--- avg_price_per_room <= 80.00
|                       |--- truncated branch of depth 6
|                   |--- avg_price_per_room > 80.00
|                       |--- weights: [3.00, 0.00] class: 0
|           |--- arrival_month > 2.50
|               |--- arrival_month <= 3.50
|                   |--- truncated branch of depth 8
|               |--- arrival_month > 3.50
|                   |--- truncated branch of depth 3
|       |--- avg_price_per_room > 85.99
|           |--- no_of_weekend_nights <= 0.50
|               |--- lead_time <= 27.00
|                   |--- truncated branch of depth 3
|               |--- lead_time > 27.00
|                   |--- truncated branch of depth 5
|           |--- no_of_weekend_nights > 0.50
|               |--- lead_time <= 27.50
|                   |--- weights: [0.00, 13.00] class: 1
|               |--- lead_time > 27.50
|                   |--- truncated branch of depth 5
--- arrival_month > 6.50
|--- type_of_meal_plan_Not Selected <= 0.50
|   |--- no_of_week_nights <= 7.50
|       |--- avg_price_per_room <= 109.73
|           |--- no_of_weekend_nights <= 0.50
|               |--- arrival_month <= 10.50

```

```
    |--- avg_price_per_room <= 75.30
    |   |--- truncated branch of depth 12
    |--- avg_price_per_room >  75.30
    |   |--- truncated branch of depth 7
    --- arrival_month >  10.50
    |--- avg_price_per_room <= 91.62
    |   |--- truncated branch of depth 9
    |--- avg_price_per_room >  91.62
    |   |--- truncated branch of depth 8
    --- no_of_weekend_nights >  0.50
    |--- avg_price_per_room <= 101.68
    |   |--- arrival_month <= 8.50
    |       |--- truncated branch of depth 10
    |--- arrival_month >  8.50
    |       |--- truncated branch of depth 16
    --- avg_price_per_room >  101.68
    |--- lead_time <= 24.50
    |   |--- truncated branch of depth 6
    |--- lead_time >  24.50
    |   |--- truncated branch of depth 5
    --- avg_price_per_room >  109.73
    |--- arrival_month <= 9.50
    |   |--- lead_time <= 3.50
    |       |--- weights: [12.00, 0.00] class: 0
    |--- lead_time >  3.50
    |   |--- lead_time <= 31.00
    |       |--- truncated branch of depth 5
    |--- lead_time >  31.00
    |       |--- truncated branch of depth 3
    --- arrival_month >  9.50
    |--- no_of_adults <= 1.50
    |   |--- avg_price_per_room <= 111.50
    |       |--- truncated branch of depth 7
    |--- avg_price_per_room >  111.50
    |       |--- weights: [0.00, 1.00] class: 1
    --- no_of_adults >  1.50
    |   |--- weights: [22.00, 0.00] class: 0
    --- no_of_week_nights >  7.50
    |--- avg_price_per_room <= 92.66
    |   |--- lead_time <= 16.50
    |       |--- weights: [2.00, 0.00] class: 0
    |--- lead_time >  16.50
    |       |--- weights: [0.00, 1.00] class: 1
    |--- avg_price_per_room >  92.66
    |       |--- weights: [0.00, 3.00] class: 1
    --- type_of_meal_plan_Not Selected >  0.50
    |--- lead_time <= 19.50
    |   |--- arrival_month <= 11.50
    |       |--- lead_time <= 2.50
    |           |--- no_of_week_nights <= 0.50
    |               |--- avg_price_per_room <= 78.83
    |                   |--- truncated branch of depth 2
    |--- avg_price_per_room >  78.83
    |       |--- truncated branch of depth 2
    |--- no_of_week_nights >  0.50
    |       |--- weights: [60.00, 0.00] class: 0
    --- lead_time >  2.50
    |--- arrival_month <= 7.50
    |   |--- weights: [0.00, 4.00] class: 1
    |--- arrival_month >  7.50
```

```

|--- no_of_week_nights <= 3.50
|   |--- truncated branch of depth 6
|--- no_of_week_nights >  3.50
|   |--- truncated branch of depth 4
|--- arrival_month >  11.50
|   |--- weights: [68.00, 0.00] class: 0
|--- lead_time >  19.50
|   |--- avg_price_per_room <= 64.65
|   |   |--- weights: [9.00, 0.00] class: 0
|--- avg_price_per_room >  64.65
|   |--- no_of_week_nights <= 1.50
|   |   |--- avg_price_per_room <= 86.75
|   |   |   |--- avg_price_per_room <= 81.04
|   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |--- avg_price_per_room >  81.04
|   |   |   |   |--- truncated branch of depth 4
|   |   |--- avg_price_per_room >  86.75
|   |   |--- lead_time <= 22.00
|   |   |   |--- weights: [1.00, 0.00] class: 0
|   |   |--- lead_time >  22.00
|   |   |   |--- truncated branch of depth 4
|--- no_of_week_nights >  1.50
|   |--- avg_price_per_room <= 100.10
|   |   |--- lead_time <= 42.00
|   |   |   |--- truncated branch of depth 11
|   |   |--- lead_time >  42.00
|   |   |   |--- weights: [0.00, 3.00] class: 1
|--- avg_price_per_room >  100.10
|   |--- weights: [4.00, 0.00] class: 0
|--- lead_time >  44.50
|--- type_of_meal_plan_Not Selected <= 0.50
|   |--- arrival_month <= 8.50
|   |--- arrival_month <= 1.50
|   |   |--- lead_time <= 51.50
|   |   |   |--- no_of_adults <= 1.50
|   |   |   |   |--- weights: [0.00, 2.00] class: 1
|   |   |   |--- no_of_adults >  1.50
|   |   |   |   |--- weights: [10.00, 0.00] class: 0
|--- lead_time >  51.50
|   |--- weights: [124.00, 0.00] class: 0
|--- arrival_month >  1.50
|--- avg_price_per_room <= 86.38
|   |--- avg_price_per_room <= 80.17
|   |   |--- avg_price_per_room <= 60.80
|   |   |   |--- arrival_month <= 6.50
|   |   |   |   |--- truncated branch of depth 6
|   |   |   |   |--- arrival_month >  6.50
|   |   |   |   |--- truncated branch of depth 5
|--- avg_price_per_room >  60.80
|   |--- avg_price_per_room <= 61.47
|   |   |--- truncated branch of depth 2
|--- avg_price_per_room >  61.47
|   |   |--- truncated branch of depth 15
|--- avg_price_per_room >  80.17
|   |--- avg_price_per_room <= 81.78
|   |   |--- no_of_weekend_nights <= 2.50
|   |   |   |--- truncated branch of depth 8
|--- no_of_weekend_nights >  2.50
|   |   |   |--- weights: [0.00, 1.00] class: 1
|--- avg_price_per_room >  81.78

```

```
    |--- avg_price_per_room <= 85.44
    |   |--- truncated branch of depth 8
    |--- avg_price_per_room >  85.44
    |   |--- truncated branch of depth 6
    --- avg_price_per_room >  86.38
    --- avg_price_per_room <= 105.92
        |--- no_of_weekend_nights <= 0.50
            |--- type_of_meal_plan_Meal Plan 2 <= 0.50
                |--- truncated branch of depth 14
            |--- type_of_meal_plan_Meal Plan 2 >  0.50
                |--- weights: [57.00, 0.00] class: 0
        |--- no_of_weekend_nights >  0.50
            |--- required_car_parking_space <= 0.50
                |--- truncated branch of depth 17
            |--- required_car_parking_space >  0.50
                |--- weights: [7.00, 0.00] class: 0
    --- avg_price_per_room >  105.92
        |--- lead_time <= 116.50
            |--- lead_time <= 81.50
                |--- truncated branch of depth 7
            |--- lead_time >  81.50
                |--- truncated branch of depth 7
        |--- lead_time >  116.50
            |--- arrival_month <= 5.50
                |--- truncated branch of depth 3
            |--- arrival_month >  5.50
                |--- truncated branch of depth 6
    --- arrival_month >  8.50
        |--- lead_time <= 93.50
            |--- avg_price_per_room <= 86.20
                |--- no_of_weekend_nights <= 0.50
                    |--- avg_price_per_room <= 69.99
                        |--- avg_price_per_room <= 68.99
                            |--- weights: [70.00, 0.00] class: 0
                        |--- avg_price_per_room >  68.99
                            |--- weights: [0.00, 1.00] class: 1
                    |--- avg_price_per_room >  69.99
                            |--- weights: [127.00, 0.00] class: 0
                |--- no_of_weekend_nights >  0.50
                    |--- arrival_month <= 10.50
                        |--- no_of_adults <= 1.50
                            |--- truncated branch of depth 3
                        |--- no_of_adults >  1.50
                            |--- truncated branch of depth 9
        |--- arrival_month >  10.50
            |--- avg_price_per_room <= 79.10
                |--- truncated branch of depth 5
            |--- avg_price_per_room >  79.10
                |--- weights: [0.00, 1.00] class: 1
    --- avg_price_per_room >  86.20
        |--- arrival_month <= 9.50
            |--- lead_time <= 86.50
                |--- type_of_meal_plan_Meal Plan 2 <= 0.50
                    |--- truncated branch of depth 11
                |--- type_of_meal_plan_Meal Plan 2 >  0.50
                    |--- weights: [123.00, 0.00] class: 0
            |--- lead_time >  86.50
                |--- lead_time <= 90.50
                    |--- weights: [0.00, 5.00] class: 1
                |--- lead_time >  90.50
```

```
| | | --- weights: [3.00, 0.00] class: 0
| --- arrival_month > 9.50
| | --- avg_price_per_room <= 86.85
| | | --- no_of_week_nights <= 1.50
| | | | --- weights: [0.00, 7.00] class: 1
| | | | --- no_of_week_nights > 1.50
| | | | | --- truncated branch of depth 2
| | --- avg_price_per_room > 86.85
| | | --- avg_price_per_room <= 93.12
| | | | --- truncated branch of depth 7
| | | | --- avg_price_per_room > 93.12
| | | | | --- truncated branch of depth 12
--- lead_time > 93.50
| --- lead_time <= 106.50
| | --- avg_price_per_room <= 108.50
| | | --- avg_price_per_room <= 91.72
| | | | --- no_of_week_nights <= 2.50
| | | | | --- truncated branch of depth 7
| | | | --- no_of_week_nights > 2.50
| | | | | --- truncated branch of depth 6
| | --- avg_price_per_room > 91.72
| | | --- arrival_month <= 10.50
| | | | --- truncated branch of depth 2
| | | --- arrival_month > 10.50
| | | | | --- truncated branch of depth 2
| --- avg_price_per_room > 108.50
| | --- avg_price_per_room <= 109.40
| | | --- weights: [40.00, 0.00] class: 0
| --- avg_price_per_room > 109.40
| | --- avg_price_per_room <= 110.49
| | | --- truncated branch of depth 2
| | --- avg_price_per_room > 110.49
| | | | --- truncated branch of depth 2
--- lead_time > 106.50
| --- avg_price_per_room <= 100.12
| | --- arrival_month <= 11.50
| | | --- arrival_month <= 10.50
| | | | --- truncated branch of depth 7
| | | --- arrival_month > 10.50
| | | | | --- truncated branch of depth 4
| --- arrival_month > 11.50
| | --- avg_price_per_room <= 77.50
| | | --- truncated branch of depth 2
| | --- avg_price_per_room > 77.50
| | | | --- truncated branch of depth 4
| --- avg_price_per_room > 100.12
| | --- arrival_month <= 11.00
| | | --- lead_time <= 122.00
| | | | --- truncated branch of depth 2
| | | --- lead_time > 122.00
| | | | | --- weights: [0.00, 2.00] class: 1
| | --- arrival_month > 11.00
| | | | --- weights: [0.00, 4.00] class: 1
--- type_of_meal_plan_Not Selected > 0.50
| --- avg_price_per_room <= 66.49
| | --- avg_price_per_room <= 61.58
| | | --- lead_time <= 47.00
| | | | --- weights: [0.00, 1.00] class: 1
| | --- lead_time > 47.00
| | | --- no_of_weekend_nights <= 3.50
```

```
|--- lead_time <= 99.50
|   |--- weights: [22.00, 0.00] class: 0
|--- lead_time > 99.50
|   |--- lead_time <= 109.00
|       |--- truncated branch of depth 2
|       |--- lead_time > 109.00
|           |--- weights: [5.00, 0.00] class: 0
|--- no_of_weekend_nights > 3.50
|   |--- weights: [0.00, 1.00] class: 1
|--- avg_price_per_room > 61.58
|   |--- avg_price_per_room <= 64.01
|       |--- arrival_month <= 5.50
|           |--- lead_time <= 90.00
|               |--- weights: [0.00, 9.00] class: 1
|           |--- lead_time > 90.00
|               |--- weights: [1.00, 0.00] class: 0
|       |--- arrival_month > 5.50
|           |--- lead_time <= 147.00
|               |--- lead_time <= 76.50
|                   |--- weights: [6.00, 0.00] class: 0
|               |--- lead_time > 76.50
|                   |--- truncated branch of depth 2
|               |--- lead_time > 147.00
|                   |--- weights: [0.00, 2.00] class: 1
|       |--- avg_price_per_room > 64.01
|           |--- weights: [8.00, 0.00] class: 0
|--- avg_price_per_room > 66.49
|   |--- avg_price_per_room <= 105.28
|       |--- required_car_parking_space <= 0.50
|           |--- lead_time <= 63.50
|               |--- arrival_month <= 5.50
|                   |--- arrival_month <= 4.50
|                       |--- truncated branch of depth 8
|                   |--- arrival_month > 4.50
|                       |--- truncated branch of depth 2
|               |--- arrival_month > 5.50
|                   |--- lead_time <= 47.50
|                       |--- truncated branch of depth 4
|                   |--- lead_time > 47.50
|                       |--- truncated branch of depth 5
|           |--- lead_time > 63.50
|               |--- avg_price_per_room <= 72.93
|                   |--- lead_time <= 73.50
|                       |--- weights: [0.00, 4.00] class: 1
|                   |--- lead_time > 73.50
|                       |--- truncated branch of depth 8
|               |--- avg_price_per_room > 72.93
|                   |--- no_of_adults <= 1.50
|                       |--- truncated branch of depth 4
|                   |--- no_of_adults > 1.50
|                       |--- truncated branch of depth 14
|--- required_car_parking_space > 0.50
|   |--- weights: [6.00, 0.00] class: 0
|--- avg_price_per_room > 105.28
|   |--- lead_time <= 47.50
|       |--- no_of_weekend_nights <= 1.00
|           |--- avg_price_per_room <= 107.55
|               |--- weights: [0.00, 1.00] class: 1
|           |--- avg_price_per_room > 107.55
|               |--- weights: [1.00, 0.00] class: 0
```





```
    |--- lead_time <= 31.50
    |   |--- truncated branch of depth 8
    |--- lead_time > 31.50
    |   |--- truncated branch of depth 15
    --- type_of_meal_plan_Not Selected > 0.50
    |--- no_of_adults <= 1.50
    |   |--- lead_time <= 62.50
    |       |--- no_of_week_nights <= 3.50
    |           |--- truncated branch of depth 4
    |       |--- no_of_week_nights > 3.50
    |           |--- weights: [0.00, 1.00] class: 1
    |--- lead_time > 62.50
    |   |--- weights: [0.00, 3.00] class: 1
    --- no_of_adults > 1.50
    |--- lead_time <= 11.50
    |   |--- no_of_week_nights <= 1.50
    |       |--- weights: [0.00, 1.00] class: 1
    |   |--- no_of_week_nights > 1.50
    |       |--- weights: [2.00, 0.00] class: 0
    |--- lead_time > 11.50
    |   |--- avg_price_per_room <= 117.50
    |       |--- weights: [0.00, 25.00] class: 1
    |   |--- avg_price_per_room > 117.50
    |       |--- truncated branch of depth 8
    --- avg_price_per_room > 140.26
    |--- arrival_month <= 9.50
    |   |--- avg_price_per_room <= 194.50
    |       |--- avg_price_per_room <= 159.65
    |           |--- no_of_week_nights <= 4.50
    |               |--- truncated branch of depth 17
    |           |--- no_of_week_nights > 4.50
    |               |--- weights: [4.00, 0.00] class: 0
    |--- avg_price_per_room > 159.65
    |   |--- lead_time <= 54.00
    |       |--- truncated branch of depth 8
    |--- lead_time > 54.00
    |   |--- truncated branch of depth 6
    --- avg_price_per_room > 194.50
    |--- weights: [0.00, 44.00] class: 1
    |--- arrival_month > 9.50
    |--- lead_time <= 39.50
    |   |--- arrival_month <= 11.50
    |       |--- no_of_week_nights <= 1.50
    |           |--- truncated branch of depth 3
    |       |--- no_of_week_nights > 1.50
    |           |--- truncated branch of depth 4
    |--- arrival_month > 11.50
    |   |--- weights: [12.00, 0.00] class: 0
    |--- lead_time > 39.50
    |--- arrival_month <= 11.50
    |   |--- avg_price_per_room <= 167.50
    |       |--- weights: [3.00, 0.00] class: 0
    |   |--- avg_price_per_room > 167.50
    |       |--- truncated branch of depth 2
    |--- arrival_month > 11.50
    |   |--- weights: [0.00, 4.00] class: 1
    |--- no_of_weekend_nights > 0.50
    |--- arrival_month <= 7.50
    |   |--- required_car_parking_space <= 0.50
    |       |--- room_type_reserved_Room_Type 5 <= 0.50
```

```
    |--- arrival_month <= 2.50
    |   |--- no_of_weekend_nights <= 1.50
    |   |   |--- truncated branch of depth 4
    |   |--- no_of_weekend_nights >  1.50
    |   |   |--- truncated branch of depth 4
    |--- arrival_month >  2.50
    |   |--- avg_price_per_room <= 162.33
    |   |   |--- truncated branch of depth 17
    |   |--- avg_price_per_room >  162.33
    |   |   |--- truncated branch of depth 8
    |--- room_type_reserved_Room_Type 5 >  0.50
    |--- lead_time <= 41.50
    |   |--- weights: [6.00, 0.00] class: 0
    |--- lead_time >  41.50
    |   |--- arrival_month <= 4.50
    |   |   |--- weights: [1.00, 0.00] class: 0
    |   |--- arrival_month >  4.50
    |   |   |--- weights: [0.00, 1.00] class: 1
    |--- required_car_parking_space >  0.50
    |   |--- weights: [9.00, 0.00] class: 0
    --- arrival_month >  7.50
    |--- avg_price_per_room <= 203.04
    |   |--- avg_price_per_room <= 181.97
    |   |   |--- room_type_reserved_Room_Type 5 <= 0.50
    |   |   |   |--- avg_price_per_room <= 117.10
    |   |   |   |   |--- truncated branch of depth 6
    |   |   |   |--- avg_price_per_room >  117.10
    |   |   |   |   |--- truncated branch of depth 16
    |   |--- room_type_reserved_Room_Type 5 >  0.50
    |   |   |--- no_of_week_nights <= 2.50
    |   |   |   |--- weights: [9.00, 0.00] class: 0
    |   |   |--- no_of_week_nights >  2.50
    |   |   |   |--- truncated branch of depth 3
    |   |--- avg_price_per_room >  181.97
    |   |--- no_of_week_nights <= 3.50
    |   |   |--- arrival_month <= 8.50
    |   |   |   |--- truncated branch of depth 3
    |   |   |--- arrival_month >  8.50
    |   |   |   |--- weights: [33.00, 0.00] class: 0
    |   |--- no_of_week_nights >  3.50
    |   |   |   |--- weights: [0.00, 3.00] class: 1
    |--- avg_price_per_room >  203.04
    |   |--- arrival_month <= 11.00
    |   |   |--- weights: [0.00, 21.00] class: 1
    |--- arrival_month >  11.00
    |   |--- no_of_week_nights <= 1.50
    |   |   |--- weights: [0.00, 1.00] class: 1
    |   |--- no_of_week_nights >  1.50
    |   |   |--- weights: [3.00, 0.00] class: 0
    --- lead_time >  91.50
    |--- required_car_parking_space <= 0.50
    |   |--- arrival_month <= 7.50
    |   |   |--- arrival_month <= 4.50
    |   |   |   |--- lead_time <= 118.00
    |   |   |   |   |--- weights: [0.00, 36.00] class: 1
    |   |   |--- lead_time >  118.00
    |   |   |   |   |--- weights: [1.00, 0.00] class: 0
    |--- arrival_month >  4.50
    |   |--- avg_price_per_room <= 141.97
    |   |   |--- avg_price_per_room <= 139.72
```

```
|--- type_of_meal_plan_Meal Plan 2 <= 0.50
|   |--- truncated branch of depth 16
|--- type_of_meal_plan_Meal Plan 2 >  0.50
|   |--- truncated branch of depth 3
|--- avg_price_per_room >  139.72
|   |--- arrival_month <= 5.50
|       |--- weights: [6.00, 0.00] class: 0
|--- arrival_month >  5.50
|   |--- truncated branch of depth 5
|--- avg_price_per_room >  141.97
|   |--- no_of_adults <= 1.50
|       |--- arrival_month <= 6.00
|           |--- weights: [2.00, 0.00] class: 0
|--- arrival_month >  6.00
|   |--- weights: [0.00, 2.00] class: 1
|--- no_of_adults >  1.50
|   |--- no_of_weekend_nights <= 1.50
|       |--- truncated branch of depth 7
|--- no_of_weekend_nights >  1.50
|   |--- truncated branch of depth 4
|--- arrival_month >  7.50
|   |--- avg_price_per_room <= 113.58
|       |--- no_of_adults <= 2.50
|           |--- weights: [6.00, 0.00] class: 0
|--- no_of_adults >  2.50
|   |--- weights: [0.00, 1.00] class: 1
|--- avg_price_per_room >  113.58
|   |--- arrival_month <= 10.50
|       |--- no_of_week_nights <= 3.50
|           |--- no_of_adults <= 2.50
|               |--- truncated branch of depth 9
|--- no_of_adults >  2.50
|   |--- truncated branch of depth 6
|--- no_of_week_nights >  3.50
|   |--- no_of_weekend_nights <= 1.50
|       |--- weights: [0.00, 54.00] class: 1
|--- no_of_weekend_nights >  1.50
|   |--- truncated branch of depth 5
|--- arrival_month >  10.50
|   |--- no_of_adults <= 2.50
|       |--- no_of_weekend_nights <= 1.50
|           |--- weights: [0.00, 2.00] class: 1
|--- no_of_weekend_nights >  1.50
|   |--- truncated branch of depth 3
|--- no_of_adults >  2.50
|   |--- lead_time <= 102.50
|       |--- weights: [1.00, 0.00] class: 0
|--- lead_time >  102.50
|   |--- weights: [0.00, 6.00] class: 1
|--- required_car_parking_space >  0.50
|   |--- no_of_week_nights <= 8.00
|       |--- weights: [13.00, 0.00] class: 0
|--- no_of_week_nights >  8.00
|   |--- weights: [0.00, 1.00] class: 1
|--- no_of_special_requests >  0.50
|--- no_of_special_requests <= 1.50
|   |--- lead_time <= 8.50
|       |--- no_of_week_nights <= 10.00
|           |--- lead_time <= 4.50
|               |--- avg_price_per_room <= 219.86
```

```
    |--- avg_price_per_room <= 157.64
    |   |--- avg_price_per_room <= 78.71
    |   |   |--- weights: [230.00, 0.00] class: 0
    |   |--- avg_price_per_room > 78.71
    |   |   |--- avg_price_per_room <= 88.22
    |   |   |   |--- no_of_children <= 1.50
    |   |   |       |--- truncated branch of depth 10
    |   |   |   |--- no_of_children > 1.50
    |   |   |       |--- truncated branch of depth 2
    |   |--- avg_price_per_room > 88.22
    |   |   |--- no_of_week_nights <= 4.50
    |   |   |       |--- truncated branch of depth 13
    |   |   |   |--- no_of_week_nights > 4.50
    |   |   |       |--- truncated branch of depth 2
    |--- avg_price_per_room > 157.64
    |   |--- avg_price_per_room <= 158.50
    |   |   |--- weights: [0.00, 1.00] class: 1
    |   |--- avg_price_per_room > 158.50
    |   |   |--- no_of_week_nights <= 3.50
    |   |   |       |--- type_of_meal_plan_Meal Plan 2 <= 0.50
    |   |   |           |--- truncated branch of depth 6
    |   |   |           |--- type_of_meal_plan_Meal Plan 2 > 0.50
    |   |   |               |--- truncated branch of depth 2
    |   |   |--- no_of_week_nights > 3.50
    |   |   |       |--- lead_time <= 2.50
    |   |   |           |--- weights: [4.00, 0.00] class: 0
    |   |   |       |--- lead_time > 2.50
    |   |   |           |--- truncated branch of depth 2
    |--- avg_price_per_room > 219.86
    |   |--- arrival_month <= 6.00
    |   |   |--- weights: [0.00, 1.00] class: 1
    |--- arrival_month > 6.00
    |   |--- avg_price_per_room <= 237.25
    |   |   |--- weights: [5.00, 0.00] class: 0
    |   |--- avg_price_per_room > 237.25
    |   |   |--- no_of_week_nights <= 0.50
    |   |   |       |--- weights: [0.00, 1.00] class: 1
    |   |   |       |--- no_of_week_nights > 0.50
    |   |   |           |--- avg_price_per_room <= 244.12
    |   |   |               |--- weights: [0.00, 1.00] class: 1
    |   |   |           |--- avg_price_per_room > 244.12
    |   |   |               |--- weights: [4.00, 0.00] class: 0
    |--- lead_time > 4.50
    |   |--- avg_price_per_room <= 123.60
    |   |   |--- no_of_children <= 5.50
    |   |   |       |--- no_of_adults <= 0.50
    |   |   |           |--- weights: [0.00, 1.00] class: 1
    |   |   |       |--- no_of_adults > 0.50
    |   |   |           |--- arrival_month <= 8.50
    |   |   |               |--- avg_price_per_room <= 85.90
    |   |   |                   |--- truncated branch of depth 6
    |   |   |                   |--- avg_price_per_room > 85.90
    |   |   |                       |--- truncated branch of depth 10
    |   |   |               |--- arrival_month > 8.50
    |   |   |                   |--- weights: [131.00, 0.00] class: 0
    |--- no_of_children > 5.50
    |   |--- weights: [0.00, 1.00] class: 1
    |--- avg_price_per_room > 123.60
    |   |--- type_of_meal_plan_Not Selected <= 0.50
    |       |--- avg_price_per_room <= 124.05
```

```
|--- arrival_month <= 6.00
|   |--- weights: [2.00, 0.00] class: 0
|--- arrival_month > 6.00
|   |--- weights: [0.00, 2.00] class: 1
--- avg_price_per_room > 124.05
--- arrival_month <= 2.50
|   |--- avg_price_per_room <= 128.50
|   |   |--- truncated branch of depth 2
|--- avg_price_per_room > 128.50
|   |   |--- weights: [2.00, 0.00] class: 0
|--- arrival_month > 2.50
|   |--- arrival_month <= 6.50
|   |   |--- weights: [40.00, 0.00] class: 0
|--- arrival_month > 6.50
|   |   |--- truncated branch of depth 13
--- type_of_meal_plan_Not Selected > 0.50
|--- arrival_month <= 9.50
|   |--- no_of_week_nights <= 1.50
|   |   |--- avg_price_per_room <= 128.50
|   |   |   |--- weights: [3.00, 0.00] class: 0
|   |   |--- avg_price_per_room > 128.50
|   |   |   |--- truncated branch of depth 4
|   |--- no_of_week_nights > 1.50
|   |   |--- weights: [11.00, 0.00] class: 0
|--- arrival_month > 9.50
|   |--- lead_time <= 6.50
|   |   |--- weights: [1.00, 0.00] class: 0
|--- lead_time > 6.50
|   |--- no_of_week_nights <= 2.00
|   |   |--- truncated branch of depth 2
|   |--- no_of_week_nights > 2.00
|   |   |--- weights: [0.00, 2.00] class: 1
--- no_of_week_nights > 10.00
|--- arrival_month <= 1.50
|   |--- weights: [1.00, 0.00] class: 0
|--- arrival_month > 1.50
|   |--- weights: [0.00, 3.00] class: 1
--- lead_time > 8.50
|--- avg_price_per_room <= 118.55
|--- no_of_weekend_nights <= 2.50
|   |--- lead_time <= 61.50
|   |   |--- arrival_month <= 11.50
|   |   |   |--- avg_price_per_room <= 66.30
|   |   |   |   |--- no_of_week_nights <= 2.50
|   |   |   |   |   |--- weights: [97.00, 0.00] class: 0
|   |   |   |   |--- no_of_week_nights > 2.50
|   |   |   |   |   |--- lead_time <= 16.50
|   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |   |--- lead_time > 16.50
|   |   |   |   |   |   |--- truncated branch of depth 5
|--- avg_price_per_room > 66.30
|   |--- type_of_meal_plan_Not Selected <= 0.50
|   |   |--- avg_price_per_room <= 90.99
|   |   |   |--- truncated branch of depth 19
|   |   |--- avg_price_per_room > 90.99
|   |   |   |--- truncated branch of depth 17
|--- type_of_meal_plan_Not Selected > 0.50
|   |--- no_of_week_nights <= 4.50
|   |   |--- truncated branch of depth 18
|--- no_of_week_nights > 4.50
```

```
|   |   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |--- arrival_month > 11.50
|   |   |   |   |--- weights: [233.00, 0.00] class: 0
|--- lead_time > 61.50
|   |   |   |--- arrival_month <= 6.50
|   |   |   |   |--- avg_price_per_room <= 71.98
|   |   |   |   |   |--- lead_time <= 72.50
|   |   |   |   |   |   |--- lead_time <= 71.50
|   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |--- lead_time > 71.50
|   |   |   |   |   |--- truncated branch of depth 2
|   |   |   |   |--- lead_time > 72.50
|   |   |   |   |--- weights: [76.00, 0.00] class: 0
|--- avg_price_per_room > 71.98
|   |   |   |--- avg_price_per_room <= 72.37
|   |   |   |   |--- no_of_week_nights <= 4.00
|   |   |   |   |   |--- weights: [0.00, 7.00] class: 1
|   |   |   |   |--- no_of_week_nights > 4.00
|   |   |   |   |   |--- weights: [3.00, 0.00] class: 0
|--- avg_price_per_room > 72.37
|   |   |   |--- lead_time <= 126.50
|   |   |   |   |--- truncated branch of depth 15
|   |   |   |   |--- lead_time > 126.50
|   |   |   |   |--- truncated branch of depth 9
|--- arrival_month > 6.50
|   |   |   |--- no_of_week_nights <= 1.50
|   |   |   |   |--- no_of_adults <= 1.50
|   |   |   |   |   |--- lead_time <= 135.00
|   |   |   |   |   |   |--- truncated branch of depth 4
|   |   |   |   |--- lead_time > 135.00
|   |   |   |   |   |--- truncated branch of depth 2
|--- no_of_adults > 1.50
|   |   |   |--- type_of_meal_plan_Not Selected <= 0.50
|   |   |   |   |--- truncated branch of depth 12
|   |   |   |--- type_of_meal_plan_Not Selected > 0.50
|   |   |   |   |--- truncated branch of depth 9
|--- no_of_week_nights > 1.50
|   |   |   |--- avg_price_per_room <= 76.60
|   |   |   |   |--- arrival_month <= 7.50
|   |   |   |   |   |--- truncated branch of depth 9
|   |   |   |   |--- arrival_month > 7.50
|   |   |   |   |   |--- truncated branch of depth 11
|--- avg_price_per_room > 76.60
|   |   |   |--- arrival_month <= 8.50
|   |   |   |   |--- truncated branch of depth 9
|--- arrival_month > 8.50
|   |   |   |   |--- truncated branch of depth 13
|--- no_of_weekend_nights > 2.50
|   |   |--- arrival_month <= 1.50
|   |   |   |--- weights: [3.00, 0.00] class: 0
|--- arrival_month > 1.50
|   |   |--- no_of_week_nights <= 8.50
|   |   |   |--- avg_price_per_room <= 106.66
|   |   |   |   |--- arrival_month <= 2.50
|   |   |   |   |   |--- weights: [1.00, 0.00] class: 0
|--- arrival_month > 2.50
|   |   |   |--- arrival_month <= 11.50
|   |   |   |   |--- truncated branch of depth 2
|--- arrival_month > 11.50
|   |   |   |   |--- weights: [1.00, 0.00] class: 0
```

```
|   |   |   --- avg_price_per_room > 106.66
|   |   |   |   --- weights: [3.00, 0.00] class: 0
|   |   |   --- no_of_week_nights > 8.50
|   |   |   |   --- weights: [0.00, 11.00] class: 1
--- avg_price_per_room > 118.55
--- required_car_parking_space <= 0.50
|   --- arrival_month <= 8.50
|   |   --- no_of_week_nights <= 7.50
|   |   |   --- no_of_adults <= 2.50
|   |   |   |   --- lead_time <= 138.50
|   |   |   |   |   --- avg_price_per_room <= 127.80
|   |   |   |   |   |   --- truncated branch of depth 12
|   |   |   |   --- avg_price_per_room > 127.80
|   |   |   |   |   --- truncated branch of depth 20
|   |   |   |   --- lead_time > 138.50
|   |   |   |   |   --- no_of_children <= 0.50
|   |   |   |   |   |   --- truncated branch of depth 4
|   |   |   |   --- no_of_children > 0.50
|   |   |   |   |   --- truncated branch of depth 2
|   |   --- no_of_adults > 2.50
|   |   |   --- avg_price_per_room <= 131.55
|   |   |   |   --- arrival_month <= 2.50
|   |   |   |   |   --- truncated branch of depth 4
|   |   |   |   --- arrival_month > 2.50
|   |   |   |   |   --- truncated branch of depth 6
|   |   |   --- avg_price_per_room > 131.55
|   |   |   |   --- avg_price_per_room <= 182.20
|   |   |   |   |   --- truncated branch of depth 10
|   |   |   |   --- avg_price_per_room > 182.20
|   |   |   |   |   --- truncated branch of depth 4
|   |   --- no_of_week_nights > 7.50
|   |   |   --- weights: [0.00, 4.00] class: 1
--- arrival_month > 8.50
|   --- arrival_month <= 11.50
|   |   --- lead_time <= 40.50
|   |   |   --- no_of_weekend_nights <= 0.50
|   |   |   |   --- arrival_month <= 10.50
|   |   |   |   |   --- truncated branch of depth 13
|   |   |   |   --- arrival_month > 10.50
|   |   |   |   |   --- truncated branch of depth 5
|   |   |   --- no_of_weekend_nights > 0.50
|   |   |   |   --- type_of_meal_plan_Not Selected <= 0.50
|   |   |   |   |   --- truncated branch of depth 15
|   |   |   |   --- type_of_meal_plan_Not Selected > 0.50
|   |   |   |   |   --- truncated branch of depth 8
|   |   --- lead_time > 40.50
|   |   |   --- no_of_weekend_nights <= 0.50
|   |   |   |   --- arrival_month <= 9.50
|   |   |   |   |   --- truncated branch of depth 14
|   |   |   |   --- arrival_month > 9.50
|   |   |   |   |   --- truncated branch of depth 9
|   |   |   --- no_of_weekend_nights > 0.50
|   |   |   |   --- avg_price_per_room <= 119.53
|   |   |   |   |   --- truncated branch of depth 5
|   |   |   |   --- avg_price_per_room > 119.53
|   |   |   |   |   |   --- truncated branch of depth 20
|   |   --- arrival_month > 11.50
|   |   |   --- lead_time <= 100.00
|   |   |   |   --- weights: [80.00, 0.00] class: 0
|   |   |   --- lead_time > 100.00
```

```

|--- room_type_reserved_Room_Type 7 <= 0.50
|--- weights: [0.00, 11.00] class: 1
|--- room_type_reserved_Room_Type 7 > 0.50
|--- lead_time <= 129.50
|--- weights: [2.00, 0.00] class: 0
|--- lead_time > 129.50
|--- weights: [0.00, 1.00] class: 1
--- required_car_parking_space > 0.50
|--- room_type_reserved_Room_Type 7 <= 0.50
|--- weights: [103.00, 0.00] class: 0
|--- room_type_reserved_Room_Type 7 > 0.50
|--- weights: [0.00, 1.00] class: 1
--- no_of_special_requests > 1.50
|--- lead_time <= 90.50
|--- no_of_week_nights <= 3.50
|--- weights: [2126.00, 0.00] class: 0
|--- no_of_week_nights > 3.50
|--- no_of_week_nights <= 9.50
|--- no_of_special_requests <= 2.50
|--- lead_time <= 6.50
|--- weights: [43.00, 0.00] class: 0
|--- lead_time > 6.50
|--- arrival_month <= 11.50
|--- avg_price_per_room <= 93.09
|--- lead_time <= 54.00
|--- truncated branch of depth 8
|--- lead_time > 54.00
|--- truncated branch of depth 6
|--- avg_price_per_room > 93.09
|--- avg_price_per_room <= 107.29
|--- truncated branch of depth 3
|--- avg_price_per_room > 107.29
|--- truncated branch of depth 13
|--- arrival_month > 11.50
|--- weights: [26.00, 0.00] class: 0
|--- no_of_special_requests > 2.50
|--- weights: [70.00, 0.00] class: 0
|--- no_of_week_nights > 9.50
|--- weights: [0.00, 2.00] class: 1
--- lead_time > 90.50
|--- avg_price_per_room <= 202.95
|--- arrival_month <= 8.50
|--- lead_time <= 150.50
|--- lead_time <= 141.50
|--- lead_time <= 108.50
|--- lead_time <= 107.50
|--- no_of_week_nights <= 1.50
|--- weights: [27.00, 0.00] class: 0
|--- no_of_week_nights > 1.50
|--- truncated branch of depth 9
|--- lead_time > 107.50
|--- avg_price_per_room <= 125.53
|--- truncated branch of depth 3
|--- avg_price_per_room > 125.53
|--- weights: [3.00, 0.00] class: 0
|--- lead_time > 108.50
|--- no_of_children <= 0.50
|--- no_of_week_nights <= 1.50
|--- truncated branch of depth 5
|--- no_of_week_nights > 1.50

```

```
|   |   |   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 5
|--- no_of_children >  0.50
|   |--- no_of_week_nights <= 1.50
|   |   |--- weights: [12.00, 0.00] class: 0
|   |--- no_of_week_nights >  1.50
|   |   |--- truncated branch of depth 4
|--- lead_time >  141.50
|   |--- avg_price_per_room <= 92.72
|   |--- avg_price_per_room <= 74.38
|   |   |--- weights: [3.00, 0.00] class: 0
|   |--- avg_price_per_room >  74.38
|   |   |--- required_car_parking_space <= 0.50
|   |   |   |--- weights: [0.00, 3.00] class: 1
|   |--- required_car_parking_space >  0.50
|   |   |   |--- weights: [1.00, 0.00] class: 0
|--- avg_price_per_room >  92.72
|   |--- lead_time <= 144.50
|   |   |--- no_of_children <= 0.50
|   |   |   |--- truncated branch of depth 2
|--- no_of_children >  0.50
|   |   |--- weights: [0.00, 2.00] class: 1
|--- lead_time >  144.50
|   |--- arrival_month <= 7.50
|   |   |--- weights: [16.00, 0.00] class: 0
|   |--- arrival_month >  7.50
|   |   |--- truncated branch of depth 2
|--- lead_time >  150.50
|   |--- avg_price_per_room <= 103.50
|   |   |--- weights: [2.00, 0.00] class: 0
|--- avg_price_per_room >  103.50
|   |--- weights: [0.00, 3.00] class: 1
|--- arrival_month >  8.50
|--- no_of_special_requests <= 2.50
|   |--- avg_price_per_room <= 90.42
|   |   |--- arrival_month <= 11.50
|   |   |   |--- lead_time <= 107.00
|   |   |   |   |--- avg_price_per_room <= 70.52
|   |   |   |   |   |--- weights: [3.00, 0.00] class: 0
|   |   |   |   |--- avg_price_per_room >  70.52
|   |   |   |   |--- truncated branch of depth 5
|--- lead_time >  107.00
|   |--- arrival_month <= 9.50
|   |   |--- weights: [0.00, 1.00] class: 1
|   |--- arrival_month >  9.50
|   |   |--- truncated branch of depth 4
|--- arrival_month >  11.50
|   |--- lead_time <= 101.00
|   |   |--- weights: [11.00, 0.00] class: 0
|--- lead_time >  101.00
|   |--- lead_time <= 104.00
|   |   |--- weights: [0.00, 1.00] class: 1
|--- lead_time >  104.00
|   |--- truncated branch of depth 5
|--- avg_price_per_room >  90.42
|--- no_of_adults <= 1.50
|   |--- weights: [11.00, 0.00] class: 0
|--- no_of_adults >  1.50
|   |--- avg_price_per_room <= 153.15
|   |   |--- no_of_weekend_nights <= 1.50
|   |   |   |--- truncated branch of depth 10
```



```
    |    |    |    |    |    --- avg_price_per_room >  68.79
    |    |    |    |    |    --- avg_price_per_room <= 93.56
    |    |    |    |    |    |--- weights: [0.00, 15.00] class: 1
    |    |    |    |    |    --- avg_price_per_room >  93.56
    |    |    |    |    |    |--- weights: [1.00, 1.00] class: 0
    |    |    |    |    --- type_of_meal_plan_Meal Plan 2 >  0.50
    |    |    |    |    |--- weights: [1.00, 0.00] class: 0
    |    |    --- no_of_adults >  1.50
    |    |    --- avg_price_per_room <= 82.47
    |    |    |--- lead_time <= 170.50
    |    |    |    |--- type_of_meal_plan_Not Selected <= 0.50
    |    |    |    |    |--- no_of_weekend_nights <= 2.50
    |    |    |    |    |    |--- avg_price_per_room <= 24.50
    |    |    |    |    |    |--- weights: [0.00, 1.00] class: 1
    |    |    |    |    |    --- avg_price_per_room >  24.50
    |    |    |    |    |    |--- arrival_month <= 9.50
    |    |    |    |    |    |    |--- avg_price_per_room <= 77.23
    |    |    |    |    |    |    |    |--- truncated branch of depth 6
    |    |    |    |    |    |    |    --- avg_price_per_room >  77.23
    |    |    |    |    |    |    |    |--- weights: [10.00, 0.00] class: 0
    |    |    |    |    |    --- arrival_month >  9.50
    |    |    |    |    |    |--- weights: [26.00, 0.00] class: 0
    |    |    |    |    |    --- no_of_weekend_nights >  2.50
    |    |    |    |    |    |--- weights: [0.00, 2.00] class: 1
    |    |    |    --- type_of_meal_plan_Not Selected >  0.50
    |    |    |    --- lead_time <= 155.50
    |    |    |    |--- weights: [1.00, 0.00] class: 0
    |    |    |    --- lead_time >  155.50
    |    |    |    |--- weights: [0.00, 6.00] class: 1
    |    |    --- lead_time >  170.50
    |    |    |--- type_of_meal_plan_Not Selected <= 0.50
    |    |    |    |--- arrival_month <= 6.50
    |    |    |    |    |--- lead_time <= 187.00
    |    |    |    |    |    |--- weights: [9.00, 0.00] class: 0
    |    |    |    |    --- lead_time >  187.00
    |    |    |    |    |--- avg_price_per_room <= 77.62
    |    |    |    |    |    |--- no_of_weekend_nights <= 0.50
    |    |    |    |    |    |    |--- weights: [0.00, 73.00] class: 1
    |    |    |    |    |    |    --- no_of_weekend_nights >  0.50
    |    |    |    |    |    |    |--- truncated branch of depth 5
    |    |    |    |    |    --- avg_price_per_room >  77.62
    |    |    |    |    |    |--- no_of_week_nights <= 3.50
    |    |    |    |    |    |    |--- truncated branch of depth 5
    |    |    |    |    |    |    --- no_of_week_nights >  3.50
    |    |    |    |    |    |    |--- weights: [8.00, 0.00] class: 0
    |    |    --- arrival_month >  6.50
    |    |    |--- lead_time <= 337.00
    |    |    |    |--- lead_time <= 314.50
    |    |    |    |    |--- lead_time <= 289.50
    |    |    |    |    |    |--- truncated branch of depth 15
    |    |    |    |    |    --- lead_time >  289.50
    |    |    |    |    |    |--- truncated branch of depth 3
    |    |    |    |    --- lead_time >  314.50
    |    |    |    |    |--- avg_price_per_room <= 64.62
    |    |    |    |    |    |--- weights: [33.00, 0.00] class: 0
    |    |    |    |    |    --- avg_price_per_room >  64.62
    |    |    |    |    |    |--- truncated branch of depth 2
    |    |    |    |    --- lead_time >  337.00
    |    |    |    |    |--- weights: [0.00, 39.00] class: 1
    |    |    --- type_of_meal_plan_Not Selected >  0.50
```

```
    |--- avg_price_per_room <= 54.57
    |   |--- arrival_month <= 10.00
    |   |   |--- weights: [0.00, 1.00] class: 1
    |   |--- arrival_month > 10.00
    |   |   |--- weights: [1.00, 0.00] class: 0
    |--- avg_price_per_room > 54.57
    |   |--- no_of_week_nights <= 1.50
    |   |   |--- lead_time <= 214.50
    |   |   |   |--- weights: [0.00, 1.00] class: 1
    |   |   |--- lead_time > 202.00
    |   |   |   |--- weights: [1.00, 0.00] class: 0
    |   |   |--- lead_time > 214.50
    |   |   |   |--- weights: [0.00, 6.00] class: 1
    |   |   |--- no_of_week_nights > 1.50
    |   |   |   |--- weights: [0.00, 62.00] class: 1
--- avg_price_per_room > 82.47
|--- no_of_adults <= 2.50
|   |--- lead_time <= 324.50
|       |--- arrival_month <= 11.50
|           |--- avg_price_per_room <= 90.78
|               |--- no_of_week_nights <= 4.50
|                   |--- no_of_weekend_nights <= 1.50
|                       |--- weights: [0.00, 280.00] class: 1
|                   |--- no_of_weekend_nights > 1.50
|                       |--- truncated branch of depth 4
|               |--- no_of_week_nights > 4.50
|                   |--- type_of_meal_plan_Not Selected <= 0.50
|                       |--- truncated branch of depth 2
|                   |--- type_of_meal_plan_Not Selected > 0.50
|                       |--- weights: [0.00, 11.00] class: 1
|--- avg_price_per_room > 90.78
|   |--- avg_price_per_room <= 93.04
|       |--- lead_time <= 297.50
|           |--- truncated branch of depth 8
|       |--- lead_time > 297.50
|           |--- weights: [4.00, 0.00] class: 0
|--- avg_price_per_room > 93.04
|   |--- lead_time <= 192.50
|       |--- truncated branch of depth 4
|   |--- lead_time > 192.50
|       |--- weights: [0.00, 178.00] class: 1
--- arrival_month > 11.50
|--- no_of_week_nights <= 2.50
|   |--- avg_price_per_room <= 87.00
|       |--- weights: [0.00, 1.00] class: 1
|   |--- avg_price_per_room > 87.00
|       |--- weights: [6.00, 0.00] class: 0
|--- no_of_week_nights > 2.50
|   |--- no_of_weekend_nights <= 0.50
|       |--- lead_time <= 285.00
|           |--- weights: [1.00, 0.00] class: 0
|       |--- lead_time > 285.00
|           |--- weights: [0.00, 2.00] class: 1
|   |--- no_of_weekend_nights > 0.50
|       |--- weights: [0.00, 10.00] class: 1
|--- lead_time > 324.50
|--- avg_price_per_room <= 89.00
|   |--- weights: [8.00, 0.00] class: 0
|--- avg_price_per_room > 89.00
```

```
| | | | | --- no_of_week_nights <= 4.00
| | | | | --- weights: [0.00, 2.00] class: 1
| | | | | --- no_of_week_nights > 4.00
| | | | | --- weights: [1.00, 7.00] class: 1
| | | | | --- no_of_adults > 2.50
| | | | | --- weights: [7.00, 0.00] class: 0
--- no_of_special_requests > 0.50
| --- no_of_weekend_nights <= 0.50
| | --- lead_time <= 180.50
| | | --- lead_time <= 159.50
| | | | --- arrival_month <= 8.50
| | | | | --- weights: [8.00, 0.00] class: 0
| | | | | --- arrival_month > 8.50
| | | | | --- no_of_week_nights <= 2.50
| | | | | | --- weights: [0.00, 4.00] class: 1
| | | | | --- no_of_week_nights > 2.50
| | | | | | --- lead_time <= 156.50
| | | | | | | --- weights: [0.00, 1.00] class: 1
| | | | | | --- lead_time > 156.50
| | | | | | | --- weights: [2.00, 0.00] class: 0
| | | | | --- lead_time > 159.50
| | | | | | --- no_of_adults <= 0.50
| | | | | | | --- weights: [0.00, 1.00] class: 1
| | | | | --- no_of_adults > 0.50
| | | | | | --- lead_time <= 178.50
| | | | | | | --- weights: [44.00, 0.00] class: 0
| | | | | | --- lead_time > 178.50
| | | | | | | --- avg_price_per_room <= 93.62
| | | | | | | | --- weights: [5.00, 0.00] class: 0
| | | | | | | --- avg_price_per_room > 93.62
| | | | | | | | --- avg_price_per_room <= 96.92
| | | | | | | | | --- weights: [0.00, 2.00] class: 1
| | | | | | | | --- avg_price_per_room > 96.92
| | | | | | | | | --- weights: [1.00, 0.00] class: 0
--- lead_time > 180.50
| | --- no_of_special_requests <= 2.50
| | | --- avg_price_per_room <= 72.58
| | | | --- lead_time <= 238.00
| | | | | --- lead_time <= 198.00
| | | | | | --- weights: [2.00, 0.00] class: 0
| | | | | --- lead_time > 198.00
| | | | | | | --- avg_price_per_room <= 57.75
| | | | | | | | --- weights: [1.00, 0.00] class: 0
| | | | | | | --- avg_price_per_room > 57.75
| | | | | | | | --- weights: [0.00, 9.00] class: 1
| | | | | --- lead_time > 238.00
| | | | | | --- lead_time <= 304.50
| | | | | | | --- weights: [7.00, 0.00] class: 0
| | | | | --- lead_time > 304.50
| | | | | | | --- weights: [0.00, 1.00] class: 1
| | | | | --- avg_price_per_room > 72.58
| | | | | | --- arrival_month <= 11.50
| | | | | | | --- type_of_meal_plan_Meal Plan 2 <= 0.50
| | | | | | | --- lead_time <= 322.50
| | | | | | | | --- arrival_month <= 6.50
| | | | | | | | | --- truncated branch of depth 3
| | | | | | | | --- arrival_month > 6.50
| | | | | | | | | --- truncated branch of depth 3
| | | | | | | --- lead_time > 322.50
| | | | | | | | --- weights: [2.00, 1.00] class: 0
```

```
    |--- type_of_meal_plan_Meal Plan 2 > 0.50
    |--- avg_price_per_room <= 93.00
    |   |--- weights: [0.00, 3.00] class: 1
    |--- avg_price_per_room > 93.00
    |   |--- weights: [3.00, 0.00] class: 0
    --- arrival_month > 11.50
    |--- lead_time <= 272.00
    |   |--- type_of_meal_plan_Not Selected <= 0.50
    |   |   |--- weights: [6.00, 0.00] class: 0
    |--- type_of_meal_plan_Not Selected > 0.50
    |   |--- weights: [0.00, 2.00] class: 1
    --- lead_time > 272.00
    |--- avg_price_per_room <= 78.90
    |   |--- lead_time <= 300.50
    |       |--- truncated branch of depth 2
    |--- lead_time > 300.50
    |   |--- weights: [0.00, 2.00] class: 1
    --- avg_price_per_room > 78.90
    |   |--- weights: [0.00, 5.00] class: 1
    --- no_of_special_requests > 2.50
    |   |--- weights: [12.00, 0.00] class: 0
--- no_of_weekend_nights > 0.50
|--- no_of_week_nights <= 8.50
|   |--- arrival_month <= 11.50
|       |--- avg_price_per_room <= 76.48
|       |--- no_of_special_requests <= 1.50
|           |--- lead_time <= 245.50
|           |--- weights: [81.00, 0.00] class: 0
|       |--- lead_time > 245.50
|           |--- lead_time <= 249.00
|           |--- weights: [0.00, 1.00] class: 1
|       |--- lead_time > 249.00
|           |--- weights: [18.00, 0.00] class: 0
|--- no_of_special_requests > 1.50
|   |--- lead_time <= 285.50
|       |--- arrival_month <= 7.50
|           |--- no_of_week_nights <= 3.00
|           |--- weights: [1.00, 0.00] class: 0
|           |--- no_of_week_nights > 3.00
|           |--- weights: [2.00, 1.00] class: 0
|       |--- arrival_month > 7.50
|           |--- weights: [19.00, 0.00] class: 0
|   |--- lead_time > 285.50
|       |--- weights: [0.00, 1.00] class: 1
--- avg_price_per_room > 76.48
|--- no_of_week_nights <= 6.50
|   |--- arrival_month <= 9.50
|       |--- no_of_children <= 1.50
|           |--- lead_time <= 152.50
|               |--- truncated branch of depth 3
|           |--- lead_time > 152.50
|               |--- truncated branch of depth 14
|       |--- no_of_children > 1.50
|           |--- weights: [0.00, 3.00] class: 1
|   |--- arrival_month > 9.50
|       |--- no_of_weekend_nights <= 1.50
|           |--- avg_price_per_room <= 90.92
|           |--- weights: [24.00, 0.00] class: 0
|       |--- avg_price_per_room > 90.92
|           |--- truncated branch of depth 3
```

```
    |--- no_of_weekend_nights >  1.50
    |   |--- no_of_week_nights <= 1.50
    |   |   |--- truncated branch of depth 7
    |   |--- no_of_week_nights >  1.50
    |   |   |--- truncated branch of depth 7
    |--- no_of_week_nights >  6.50
    |   |--- lead_time <= 201.00
    |   |   |--- weights: [0.00, 2.00] class: 1
    |--- lead_time >  201.00
    |   |--- lead_time <= 204.50
    |   |   |--- weights: [3.00, 0.00] class: 0
    |--- lead_time >  204.50
    |   |--- arrival_month <= 7.50
    |   |   |--- truncated branch of depth 2
    |   |--- arrival_month >  7.50
    |   |   |--- weights: [0.00, 2.00] class: 1
    |--- arrival_month >  11.50
    |--- lead_time <= 316.00
    |   |--- avg_price_per_room <= 55.92
    |   |   |--- weights: [5.00, 0.00] class: 0
    |--- avg_price_per_room >  55.92
    |   |--- avg_price_per_room <= 61.06
    |   |   |--- weights: [0.00, 2.00] class: 1
    |--- avg_price_per_room >  61.06
    |   |--- type_of_meal_plan_Not Selected <= 0.50
    |   |   |--- avg_price_per_room <= 77.19
    |   |   |   |--- weights: [8.00, 0.00] class: 0
    |   |--- avg_price_per_room >  77.19
    |   |   |--- truncated branch of depth 7
    |--- type_of_meal_plan_Not Selected >  0.50
    |   |--- no_of_special_requests <= 1.50
    |   |   |--- truncated branch of depth 2
    |   |--- no_of_special_requests >  1.50
    |   |   |--- truncated branch of depth 3
    |--- lead_time >  316.00
    |   |--- no_of_weekend_nights <= 2.50
    |   |   |--- weights: [0.00, 4.00] class: 1
    |--- no_of_weekend_nights >  2.50
    |   |--- weights: [1.00, 0.00] class: 0
    |--- no_of_week_nights >  8.50
    |   |--- no_of_special_requests <= 2.50
    |   |   |--- weights: [0.00, 8.00] class: 1
    |--- no_of_special_requests >  2.50
    |   |--- weights: [2.00, 0.00] class: 0
--- avg_price_per_room >  100.04
|--- arrival_month <= 11.50
|--- no_of_special_requests <= 2.50
|   |--- weights: [0.00, 2108.00] class: 1
|--- no_of_special_requests >  2.50
|   |--- weights: [31.00, 0.00] class: 0
--- arrival_month >  11.50
|--- no_of_special_requests <= 0.50
|   |--- weights: [47.00, 0.00] class: 0
|--- no_of_special_requests >  0.50
|--- lead_time <= 289.50
|   |--- no_of_special_requests <= 1.50
|   |   |--- avg_price_per_room <= 114.59
|   |   |   |--- weights: [2.00, 0.00] class: 0
|   |--- avg_price_per_room >  114.59
|   |   |--- weights: [0.00, 6.00] class: 1
```

```

|   |   |   |   |   |   | --- no_of_special_requests > 1.50
|   |   |   |   |   |   | --- avg_price_per_room <= 199.65
|   |   |   |   |   |   | --- no_of_week_nights <= 4.50
|   |   |   |   |   |   | --- weights: [7.00, 0.00] class: 0
|   |   |   |   |   |   | --- no_of_week_nights > 4.50
|   |   |   |   |   |   | --- lead_time <= 188.50
|   |   |   |   |   |   |   | --- weights: [1.00, 0.00] class: 0
|   |   |   |   |   |   | --- lead_time > 188.50
|   |   |   |   |   |   |   | --- weights: [0.00, 1.00] class: 1
|   |   |   |   |   |   | --- avg_price_per_room > 199.65
|   |   |   |   |   |   |   | --- weights: [0.00, 1.00] class: 1
|   |   |   |   |   | --- lead_time > 289.50
|   |   |   |   |   | --- weights: [0.00, 7.00] class: 1

```

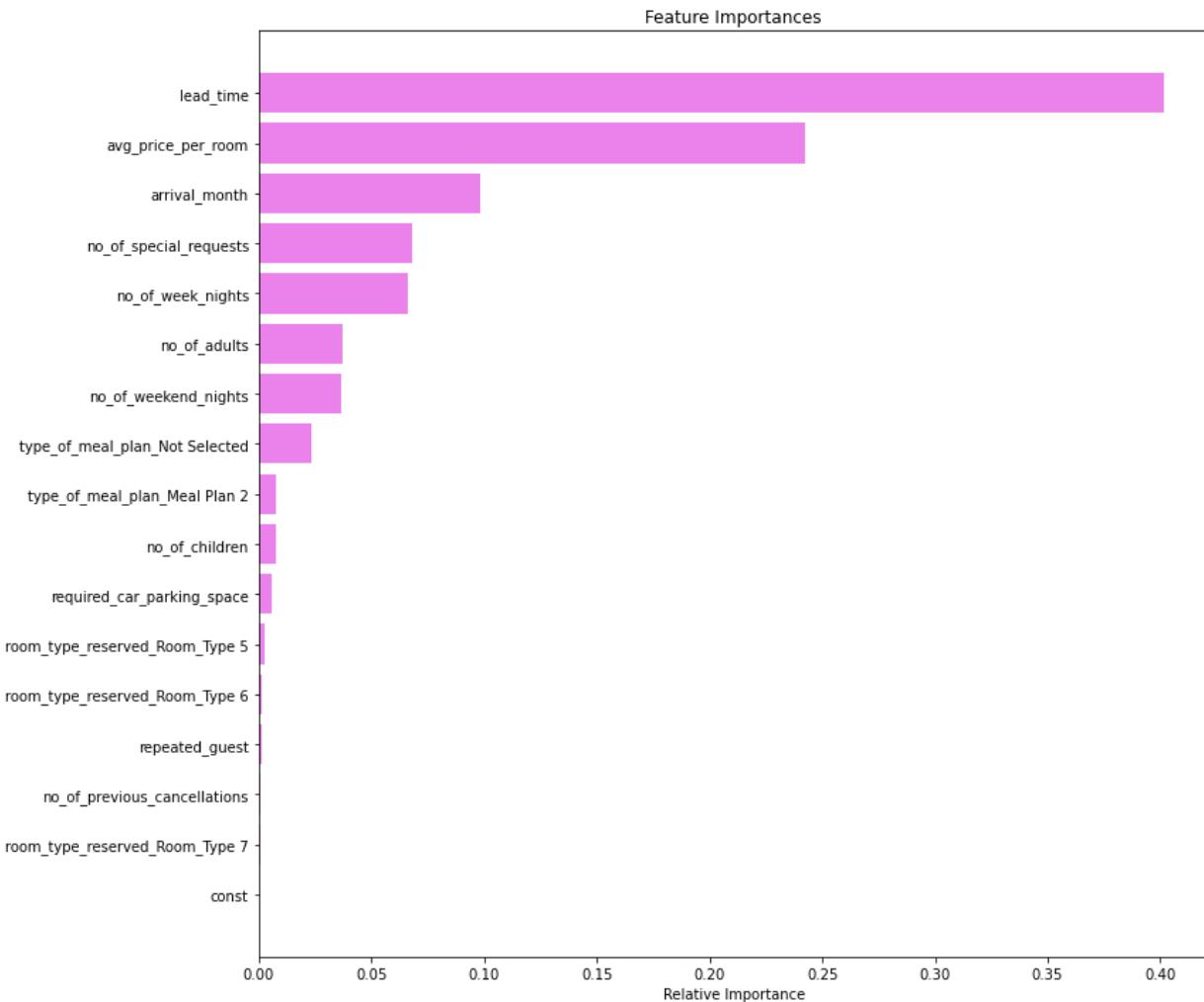
In [434...]

```

importances = model.feature_importances_
indices = np.argsort(importances)

plt.figure(figsize=(12, 12))
plt.title("Feature Importances")
plt.barh(range(len(indices)), importances[indices], color="violet", align="center")
plt.yticks(range(len(indices)), [feature_names[i] for i in indices])
plt.xlabel("Relative Importance")
plt.show()

```



Using GridSearch for Hyperparameter tuning of our tree model

- Let's see if we can improve our model performance even more.

```
In [435...]:
# Choose the type of classifier.
estimator = DecisionTreeClassifier(random_state=1)

# Grid of parameters to choose from

parameters = {
    "max_depth": [np.arange(2, 50, 5), None],
    "criterion": ["entropy", "gini"],
    "splitter": ["best", "random"],
    "min_impurity_decrease": [0.000001, 0.00001, 0.0001],
}

# Type of scoring used to compare parameter combinations
acc_scoring = make_scorer(recall_score)

# Run the grid search
grid_obj = GridSearchCV(estimator, parameters, scoring=acc_scoring, cv=5)
grid_obj = grid_obj.fit(X_train3, y_train)

# Set the clf to the best combination of parameters
estimator = grid_obj.best_estimator_

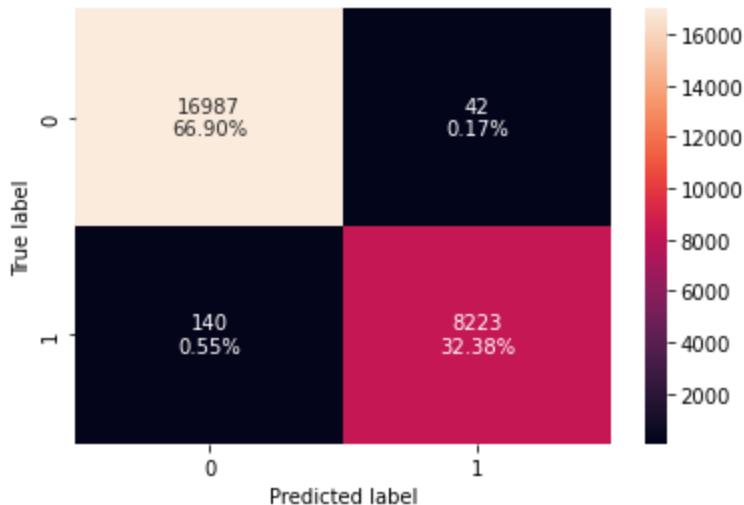
# Fit the best algorithm to the data.
estimator.fit(X_train3, y_train)
```

```
Out[435]: DecisionTreeClassifier(criterion='entropy', min_impurity_decrease=1e-05,
                                    random_state=1)
```

```
In [436...]:
decision_tree_tune_perf_train = model_performance_classification_sklearn(
    estimator, X_train3, y_train
)
decision_tree_tune_perf_train
```

	<b>Accuracy</b>	<b>Recall</b>	<b>Precision</b>	<b>F1</b>
<b>0</b>	0.99283	0.98326	0.99492	0.98905

```
In [437...]:
confusion_matrix_sklearn(estimator, X_train3, y_train)
```



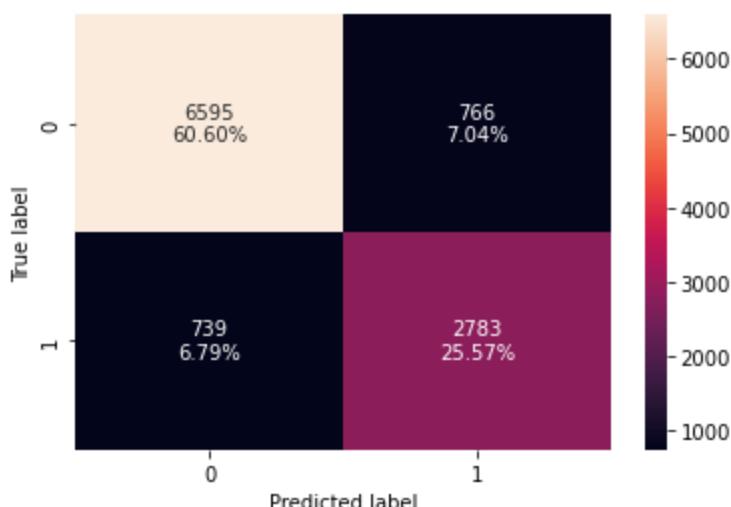
```
In [438]: decision_tree_tune_perf_test = model_performance_classification_sklearn(
    estimator, X_test3, y_test
)

decision_tree_tune_perf_test
```

```
Out[438]:
```

	Accuracy	Recall	Precision	F1
<b>0</b>	0.86171	0.79018	0.78416	0.78716

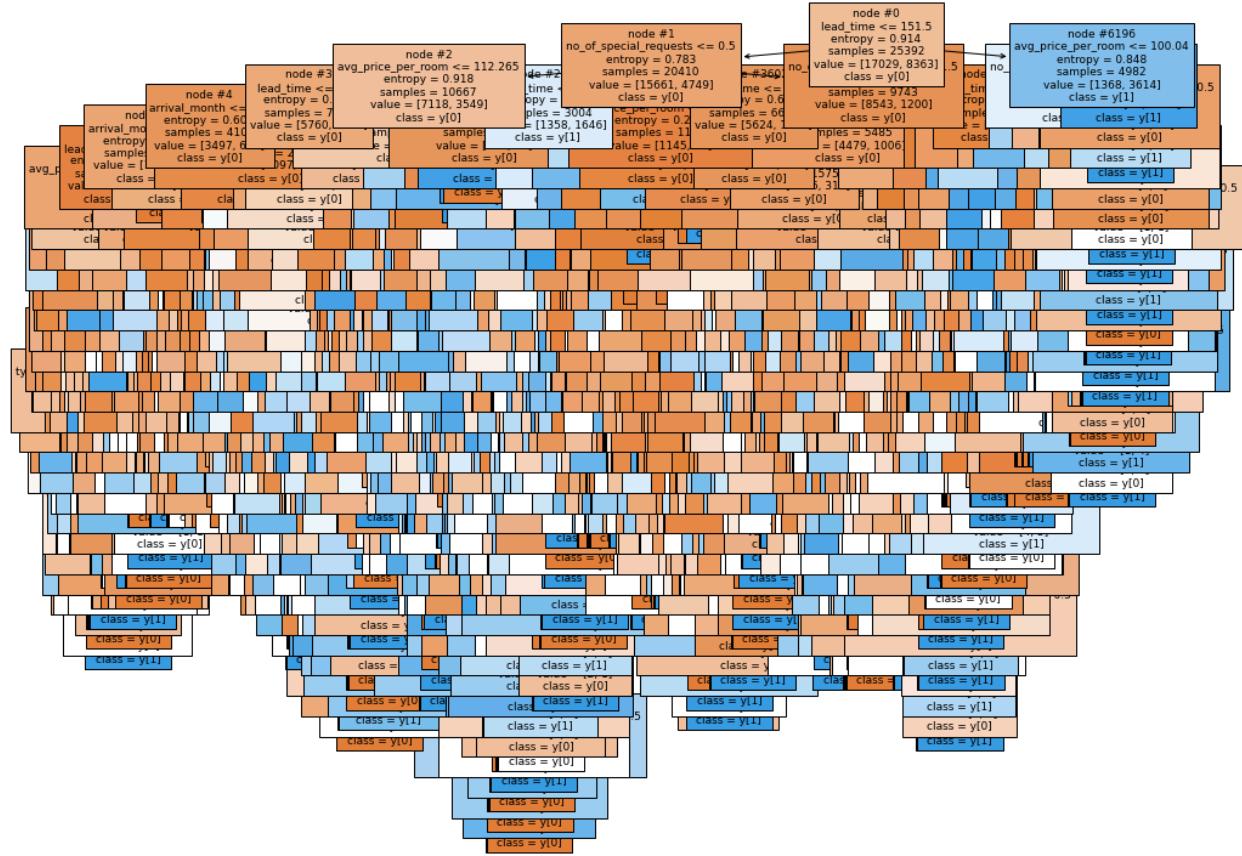
```
In [439]: confusion_matrix_sklearn(estimator, X_test3, y_test)
```



```
In [440]: plt.figure(figsize=(15, 12))

tree.plot_tree(
    estimator,
    feature_names=feature_names,
    filled=True,
    fontsize=9,
    node_ids=True,
    class_names=True,
```

```
)  
plt.show()
```



## Do we need to prune the tree?

### Cost Complexity Pruning

```
In [441...]  
clf = DecisionTreeClassifier(random_state=1)  
path = clf.cost_complexity_pruning_path(X_train3, y_train)  
ccp_alphas, impurities = path ccp_alphas, path impurities
```

```
In [442...]  
pd.DataFrame(path)
```

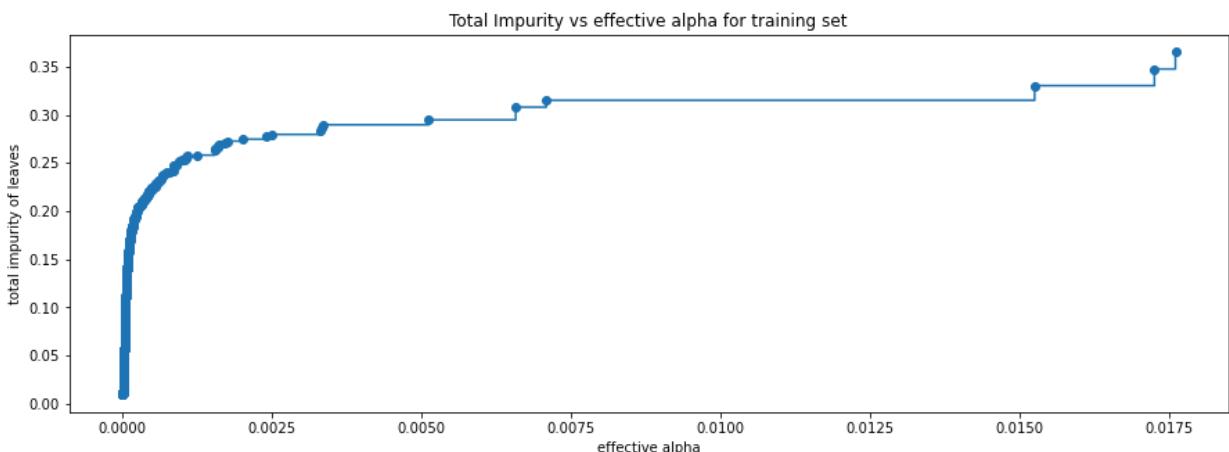
Out[442]:

	ccp_alphas	impurities
0	0.00000	0.00912
1	0.00000	0.00912
2	0.00000	0.00912
3	0.00000	0.00912
4	0.00000	0.00913
...	...	...
1568	0.00710	0.31506
1569	0.01526	0.33031
1570	0.01726	0.34757
1571	0.01761	0.36518
1572	0.07658	0.44176

1573 rows × 2 columns

In [443...]

```
fig, ax = plt.subplots(figsize=(15, 5))
ax.plot(ccp_alphas[:-1], impurities[:-1], marker="o", drawstyle="steps-post")
ax.set_xlabel("effective alpha")
ax.set_ylabel("total impurity of leaves")
ax.set_title("Total Impurity vs effective alpha for training set")
plt.show()
```



Next, we train a decision tree using the effective alphas. The last value in `ccp_alphas` is the alpha value that prunes the whole tree, leaving the tree, `clfs[-1]`, with one node.

In [444...]

```
clfs = []
for ccp_alpha in ccp_alphas:
    clf = DecisionTreeClassifier(random_state=1, ccp_alpha=ccp_alpha)
    clf.fit(X_train3, y_train)
    clfs.append(clf)
print(
    "Number of nodes in the last tree is: {} with ccp_alpha: {}".format(
        clfs[-1].tree_.node_count, ccp_alphas[-1]
    )
)
```

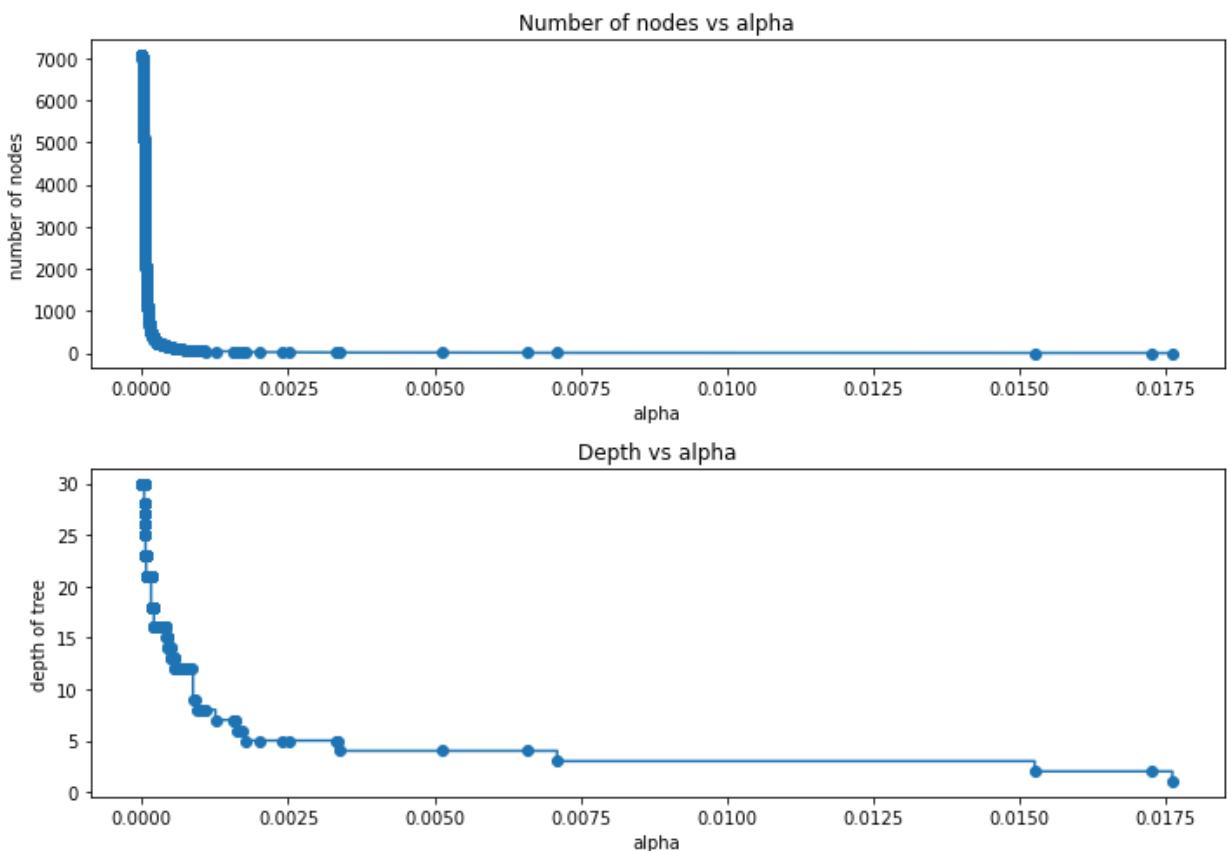
```
)  
)
```

```
Number of nodes in the last tree is: 1 with ccp_alpha: 0.07657789477371357
```

For the remainder, we remove the last element in `clfs` and `ccp_alphas`, because it is the trivial tree with only one node. Here we show that the number of nodes and tree depth decreases as alpha increases.

```
In [445...]
```

```
clfs = clfs[:-1]  
ccp_alphas = ccp_alphas[:-1]  
  
node_counts = [clf.tree_.node_count for clf in clfs]  
depth = [clf.tree_.max_depth for clf in clfs]  
fig, ax = plt.subplots(2, 1, figsize=(10, 7))  
ax[0].plot(ccp_alphas, node_counts, marker="o", drawstyle="steps-post")  
ax[0].set_xlabel("alpha")  
ax[0].set_ylabel("number of nodes")  
ax[0].set_title("Number of nodes vs alpha")  
ax[1].plot(ccp_alphas, depth, marker="o", drawstyle="steps-post")  
ax[1].set_xlabel("alpha")  
ax[1].set_ylabel("depth of tree")  
ax[1].set_title("Depth vs alpha")  
fig.tight_layout()
```



```
In [446...]
```

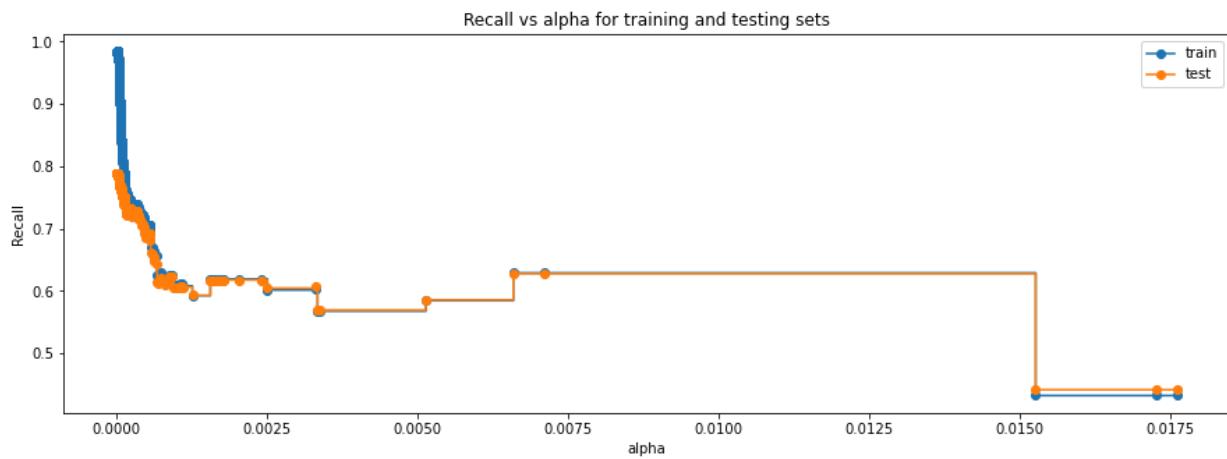
```
recall_train = []  
for clf in clfs:  
    pred_train = clf.predict(X_train3)  
    values_train = recall_score(y_train, pred_train)  
    recall_train.append(values_train)
```

```
In [447]:
```

```
recall_test = []
for clf in clfs:
    pred_test = clf.predict(X_test3)
    values_test = recall_score(y_test, pred_test)
    recall_test.append(values_test)
```

```
In [448]:
```

```
fig, ax = plt.subplots(figsize=(15, 5))
ax.set_xlabel("alpha")
ax.set_ylabel("Recall")
ax.set_title("Recall vs alpha for training and testing sets")
ax.plot(ccp_alphas, recall_train, marker="o", label="train", drawstyle="steps-post")
ax.plot(ccp_alphas, recall_test, marker="o", label="test", drawstyle="steps-post")
ax.legend()
plt.show()
```



```
In [449]:
```

```
index_best_model = np.argmax(recall_test)
best_model = clfs[index_best_model]
print(best_model)
```

```
DecisionTreeClassifier(ccp_alpha=1.9691241335853813e-05, random_state=1)
```

```
In [450]:
```

```
decision_tree_postpruned_perf_train = model_performance_classification_sklearn(
    best_model, X_train3, y_train
)
decision_tree_postpruned_perf_train
```

```
Out[450]:
```

	Accuracy	Recall	Precision	F1
0	0.99228	0.98350	0.99300	0.98823

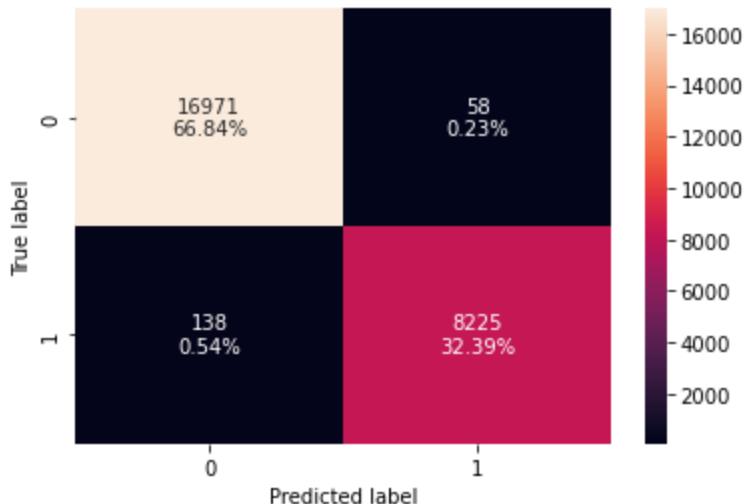
```
In [451]:
```

```
decision_tree_postpruned_perf_test = model_performance_classification_sklearn(
    best_model, X_test3, y_test
)
decision_tree_postpruned_perf_test
```

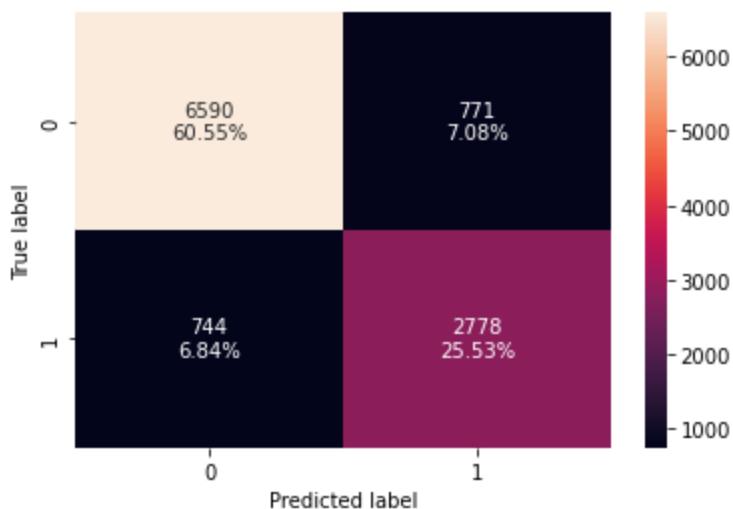
```
Out[451]:
```

	Accuracy	Recall	Precision	F1
0	0.86079	0.78876	0.78276	0.78574

```
In [452...]: confusion_matrix_sklearn(best_model, X_train3, y_train)
```



```
In [453...]: confusion_matrix_sklearn(best_model, X_test3, y_test)
```



- With post-pruning we are getting good and generalized model performance on both training and test set.
- The recall has improved further.

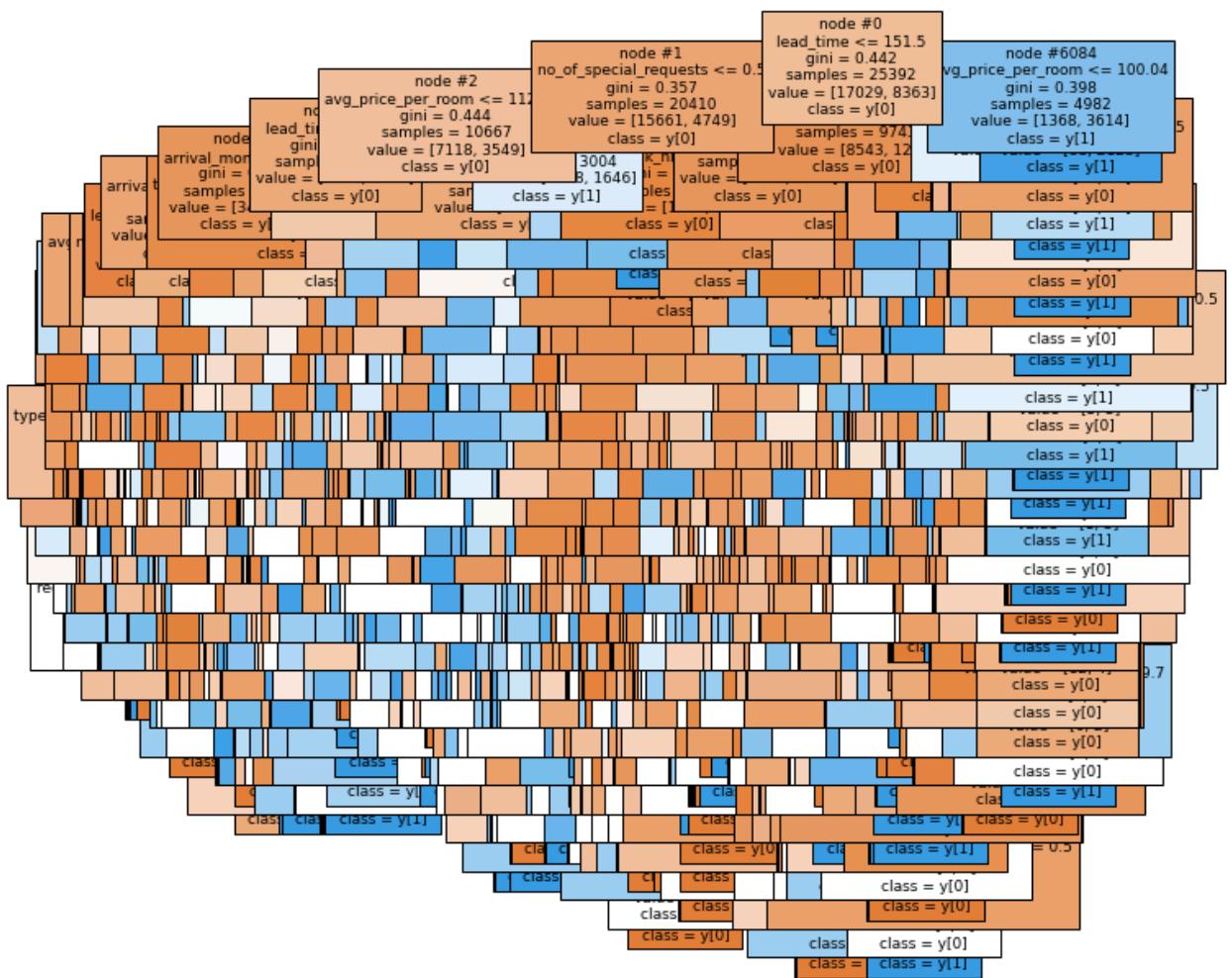
```
In [454...]: plt.figure(figsize=(10, 10))
```

```
out = tree.plot_tree(  
    best_model,  
    feature_names=feature_names,  
    filled=True,  
    fontsize=9,  
    node_ids=True,  
    class_names=True,  
)  
for o in out:  
    arrow = o.arrow_patch
```

```

if arrow is not None:
    arrow.set_edgecolor("black")
    arrow.set linewidth(1)
plt.show()
plt.show()

```



In [455...]: # Text report showing the rules of a decision tree -

```
print(tree.export_text(best_model, feature_names=feature_names, show_weights=True))
```

```
|--- lead_time <= 151.50
|--- no_of_special_requests <= 0.50
|   --- avg_price_per_room <= 112.26
|   |--- lead_time <= 44.50
|   |   |--- arrival_month <= 6.50
|   |   |   |--- lead_time <= 1.50
|   |   |   |   |--- weights: [268.00, 0.00] class: 0
|   |   |   |--- lead_time > 28.50
|   |   |   |   |--- avg_price_per_room <= 75.95
|   |   |   |   |   |--- avg_price_per_room <= 63.00
|   |   |   |   |   |   |--- avg_price_per_room <= 60.15
|   |   |   |   |   |   |   |--- weights: [17.00, 0.00] class: 0
|   |   |   |   |   |   |--- avg_price_per_room > 60.15
|   |   |   |   |   |   |   |--- weights: [0.00, 6.00] class: 1
|   |   |   |   |   |--- avg_price_per_room > 63.00
|   |   |   |   |   |   |--- weights: [20.00, 0.00] class: 0
|   |   |   |--- avg_price_per_room > 75.95
|   |   |   |   |--- avg_price_per_room <= 89.22
|   |   |   |   |   |--- no_of_adults <= 1.50
|   |   |   |   |   |   |--- weights: [1.00, 0.00] class: 0
|   |   |   |   |   |--- no_of_adults > 1.50
|   |   |   |   |   |   |--- weights: [0.00, 7.00] class: 1
|   |   |   |   |--- avg_price_per_room > 89.22
|   |   |   |   |   |--- weights: [4.00, 0.00] class: 0
|--- arrival_month > 1.50
|   |--- lead_time <= 14.50
|   |   |--- no_of_weekend_nights <= 1.50
|   |   |   |--- avg_price_per_room <= 78.90
|   |   |   |   |--- no_of_weekend_nights <= 0.50
|   |   |   |   |   |--- lead_time <= 8.50
|   |   |   |   |   |   |--- weights: [123.00, 1.00] class: 0
|   |   |   |   |   |--- lead_time > 8.50
|   |   |   |   |   |   |--- truncated branch of depth 5
|--- no_of_weekend_nights > 0.50
|   |--- arrival_month <= 3.50
|   |   |--- truncated branch of depth 8
|   |   |--- arrival_month > 3.50
|   |   |   |--- weights: [32.00, 0.00] class: 0
|--- avg_price_per_room > 78.90
|   |--- no_of_adults <= 1.50
|   |   |--- avg_price_per_room <= 92.85
|   |   |   |--- truncated branch of depth 10
|   |   |   |--- avg_price_per_room > 92.85
|   |   |   |   |--- truncated branch of depth 10
|--- no_of_adults > 1.50
|   |--- lead_time <= 4.50
|   |   |--- truncated branch of depth 11
|   |   |--- lead_time > 4.50
|   |   |   |--- truncated branch of depth 11
|--- no_of_weekend_nights > 1.50
|   |--- avg_price_per_room <= 61.88
|   |   |--- avg_price_per_room <= 57.50
|   |   |   |--- weights: [5.00, 0.00] class: 0
|   |   |   |--- avg_price_per_room > 57.50
|   |   |   |   |--- weights: [0.00, 42.00] class: 1
|--- avg_price_per_room > 61.88
|   |--- no_of_week_nights <= 5.50
|   |   |--- avg_price_per_room <= 88.63
|   |   |   |--- truncated branch of depth 7
```

```
|--- avg_price_per_room > 88.63
|   |--- truncated branch of depth 9
|--- no_of_week_nights > 5.50
|   |--- no_of_weekend_nights <= 4.50
|       |--- truncated branch of depth 3
|       |--- no_of_weekend_nights > 4.50
|           |--- weights: [0.00, 7.00] class: 1
--- lead_time > 14.50
|--- type_of_meal_plan_Not Selected <= 0.50
|   |--- avg_price_per_room <= 100.87
|       |--- avg_price_per_room <= 75.15
|           |--- no_of_adults <= 1.50
|               |--- truncated branch of depth 5
|           |--- no_of_adults > 1.50
|               |--- truncated branch of depth 8
|--- avg_price_per_room > 75.15
|   |--- avg_price_per_room <= 85.17
|       |--- truncated branch of depth 11
|   |--- avg_price_per_room > 85.17
|       |--- truncated branch of depth 13
|--- avg_price_per_room > 100.87
|--- lead_time <= 19.50
|   |--- type_of_meal_plan_Meal Plan 2 <= 0.50
|       |--- weights: [0.00, 13.00] class: 1
|   |--- type_of_meal_plan_Meal Plan 2 > 0.50
|       |--- weights: [2.00, 0.00] class: 0
|--- lead_time > 19.50
|   |--- arrival_month <= 5.50
|       |--- truncated branch of depth 8
|   |--- arrival_month > 5.50
|       |--- truncated branch of depth 3
|--- type_of_meal_plan_Not Selected > 0.50
|--- avg_price_per_room <= 85.99
|   |--- arrival_month <= 2.50
|       |--- avg_price_per_room <= 80.00
|           |--- truncated branch of depth 6
|       |--- avg_price_per_room > 80.00
|           |--- weights: [3.00, 0.00] class: 0
|--- arrival_month > 2.50
|   |--- arrival_month <= 3.50
|       |--- truncated branch of depth 8
|   |--- arrival_month > 3.50
|       |--- truncated branch of depth 3
|--- avg_price_per_room > 85.99
|--- no_of_weekend_nights <= 0.50
|   |--- lead_time <= 27.00
|       |--- truncated branch of depth 3
|   |--- lead_time > 27.00
|       |--- truncated branch of depth 5
|--- no_of_weekend_nights > 0.50
|   |--- lead_time <= 27.50
|       |--- weights: [0.00, 13.00] class: 1
|   |--- lead_time > 27.50
|       |--- truncated branch of depth 5
--- arrival_month > 6.50
|--- type_of_meal_plan_Not Selected <= 0.50
|   |--- no_of_week_nights <= 7.50
|       |--- avg_price_per_room <= 109.73
|           |--- no_of_weekend_nights <= 0.50
|               |--- arrival_month <= 10.50
```

```
    |--- avg_price_per_room <= 75.30
    |   |--- truncated branch of depth 8
    |--- avg_price_per_room >  75.30
    |   |--- truncated branch of depth 7
    --- arrival_month >  10.50
    |--- avg_price_per_room <= 91.62
    |   |--- truncated branch of depth 7
    |--- avg_price_per_room >  91.62
    |   |--- truncated branch of depth 7
    --- no_of_weekend_nights >  0.50
    |--- avg_price_per_room <= 101.68
    |   |--- arrival_month <= 8.50
    |       |--- truncated branch of depth 10
    |--- arrival_month >  8.50
    |       |--- truncated branch of depth 15
    --- avg_price_per_room >  101.68
    |--- lead_time <= 24.50
    |   |--- truncated branch of depth 6
    |--- lead_time >  24.50
    |       |--- truncated branch of depth 5
    --- avg_price_per_room >  109.73
    |--- arrival_month <= 9.50
    |   |--- lead_time <= 3.50
    |       |--- weights: [12.00, 0.00] class: 0
    |--- lead_time >  3.50
    |   |--- lead_time <= 31.00
    |       |--- truncated branch of depth 5
    |--- lead_time >  31.00
    |       |--- truncated branch of depth 3
    --- arrival_month >  9.50
    |--- no_of_adults <= 1.50
    |   |--- avg_price_per_room <= 111.50
    |       |--- truncated branch of depth 7
    |--- avg_price_per_room >  111.50
    |       |--- weights: [0.00, 1.00] class: 1
    --- no_of_adults >  1.50
    |   |--- weights: [22.00, 0.00] class: 0
    --- no_of_week_nights >  7.50
    |--- avg_price_per_room <= 92.66
    |   |--- lead_time <= 16.50
    |       |--- weights: [2.00, 0.00] class: 0
    |--- lead_time >  16.50
    |       |--- weights: [0.00, 1.00] class: 1
    |--- avg_price_per_room >  92.66
    |       |--- weights: [0.00, 3.00] class: 1
    --- type_of_meal_plan_Not Selected >  0.50
    |--- lead_time <= 19.50
    |   |--- arrival_month <= 11.50
    |       |--- lead_time <= 2.50
    |           |--- no_of_week_nights <= 0.50
    |               |--- avg_price_per_room <= 78.83
    |                   |--- truncated branch of depth 2
    |                   |--- avg_price_per_room >  78.83
    |                       |--- truncated branch of depth 2
    |--- no_of_week_nights >  0.50
    |       |--- weights: [60.00, 0.00] class: 0
    --- lead_time >  2.50
    |--- arrival_month <= 7.50
    |       |--- weights: [0.00, 4.00] class: 1
    |--- arrival_month >  7.50
```

```
    |--- no_of_week_nights <= 3.50
    |   |--- truncated branch of depth 6
    |--- no_of_week_nights >  3.50
    |   |--- truncated branch of depth 4
    |--- arrival_month >  11.50
    |   |--- weights: [68.00, 0.00] class: 0
    --- lead_time >  19.50
    |--- avg_price_per_room <= 64.65
    |   |--- weights: [9.00, 0.00] class: 0
    --- avg_price_per_room >  64.65
    |--- no_of_week_nights <= 1.50
    |   |--- avg_price_per_room <= 86.75
    |       |--- avg_price_per_room <= 81.04
    |           |--- truncated branch of depth 3
    |       |--- avg_price_per_room >  81.04
    |           |--- truncated branch of depth 4
    |--- avg_price_per_room >  86.75
    |   |--- lead_time <= 22.00
    |       |--- weights: [1.00, 0.00] class: 0
    |--- lead_time >  22.00
    |   |--- truncated branch of depth 4
    |--- no_of_week_nights >  1.50
    |   |--- avg_price_per_room <= 100.10
    |       |--- lead_time <= 42.00
    |           |--- truncated branch of depth 11
    |       |--- lead_time >  42.00
    |           |--- weights: [0.00, 3.00] class: 1
    |--- avg_price_per_room >  100.10
    |   |--- weights: [4.00, 0.00] class: 0
|--- lead_time >  44.50
|--- type_of_meal_plan_Not Selected <= 0.50
|   |--- arrival_month <= 8.50
|       |--- arrival_month <= 1.50
|           |--- lead_time <= 51.50
|               |--- no_of_adults <= 1.50
|                   |--- weights: [0.00, 2.00] class: 1
|               |--- no_of_adults >  1.50
|                   |--- weights: [10.00, 0.00] class: 0
|--- lead_time >  51.50
|   |--- weights: [124.00, 0.00] class: 0
|--- arrival_month >  1.50
|--- avg_price_per_room <= 86.38
|   |--- avg_price_per_room <= 80.17
|       |--- avg_price_per_room <= 60.80
|           |--- arrival_month <= 6.50
|               |--- truncated branch of depth 6
|           |--- arrival_month >  6.50
|               |--- truncated branch of depth 5
|--- avg_price_per_room >  60.80
|   |--- avg_price_per_room <= 61.47
|       |--- truncated branch of depth 2
|       |--- avg_price_per_room >  61.47
|           |--- truncated branch of depth 15
|--- avg_price_per_room >  80.17
|   |--- avg_price_per_room <= 81.78
|       |--- no_of_weekend_nights <= 2.50
|           |--- truncated branch of depth 8
|       |--- no_of_weekend_nights >  2.50
|           |--- weights: [0.00, 1.00] class: 1
|--- avg_price_per_room >  81.78
```

```
    |--- avg_price_per_room <= 85.44
    |   |--- truncated branch of depth 8
    |--- avg_price_per_room >  85.44
    |   |--- truncated branch of depth 6
    --- avg_price_per_room >  86.38
    --- avg_price_per_room <= 105.92
        |--- no_of_weekend_nights <= 0.50
            |--- type_of_meal_plan_Meal Plan 2 <= 0.50
                |--- truncated branch of depth 14
            |--- type_of_meal_plan_Meal Plan 2 >  0.50
                |--- weights: [57.00, 0.00] class: 0
        |--- no_of_weekend_nights >  0.50
            |--- required_car_parking_space <= 0.50
                |--- truncated branch of depth 15
            |--- required_car_parking_space >  0.50
                |--- weights: [7.00, 0.00] class: 0
    --- avg_price_per_room >  105.92
        |--- lead_time <= 116.50
            |--- lead_time <= 81.50
                |--- truncated branch of depth 7
            |--- lead_time >  81.50
                |--- truncated branch of depth 7
        |--- lead_time >  116.50
            |--- arrival_month <= 5.50
                |--- truncated branch of depth 3
            |--- arrival_month >  5.50
                |--- truncated branch of depth 6
    --- arrival_month >  8.50
        |--- lead_time <= 93.50
            |--- avg_price_per_room <= 86.20
                |--- no_of_weekend_nights <= 0.50
                    |--- avg_price_per_room <= 69.99
                        |--- avg_price_per_room <= 68.99
                            |--- weights: [70.00, 0.00] class: 0
                        |--- avg_price_per_room >  68.99
                            |--- weights: [0.00, 1.00] class: 1
                    |--- avg_price_per_room >  69.99
                            |--- weights: [127.00, 0.00] class: 0
                |--- no_of_weekend_nights >  0.50
                    |--- arrival_month <= 10.50
                        |--- no_of_adults <= 1.50
                            |--- truncated branch of depth 3
                        |--- no_of_adults >  1.50
                            |--- truncated branch of depth 8
        |--- arrival_month >  10.50
            |--- avg_price_per_room <= 79.10
                |--- weights: [98.00, 1.00] class: 0
            |--- avg_price_per_room >  79.10
                |--- weights: [0.00, 1.00] class: 1
    --- avg_price_per_room >  86.20
        |--- arrival_month <= 9.50
            |--- lead_time <= 86.50
                |--- type_of_meal_plan_Meal Plan 2 <= 0.50
                    |--- truncated branch of depth 11
                |--- type_of_meal_plan_Meal Plan 2 >  0.50
                    |--- weights: [123.00, 0.00] class: 0
            |--- lead_time >  86.50
                |--- lead_time <= 90.50
                    |--- weights: [0.00, 5.00] class: 1
                |--- lead_time >  90.50
```

```
| | | --- weights: [3.00, 0.00] class: 0
| --- arrival_month > 9.50
| | --- avg_price_per_room <= 86.85
| | | --- no_of_week_nights <= 1.50
| | | | --- weights: [0.00, 7.00] class: 1
| | | | --- no_of_week_nights > 1.50
| | | | | --- truncated branch of depth 2
| | --- avg_price_per_room > 86.85
| | | --- avg_price_per_room <= 93.12
| | | | --- truncated branch of depth 7
| | | | --- avg_price_per_room > 93.12
| | | | | --- truncated branch of depth 12
--- lead_time > 93.50
| --- lead_time <= 106.50
| | --- avg_price_per_room <= 108.50
| | | --- avg_price_per_room <= 91.72
| | | | --- no_of_week_nights <= 2.50
| | | | | --- truncated branch of depth 7
| | | | --- no_of_week_nights > 2.50
| | | | | --- truncated branch of depth 6
| | --- avg_price_per_room > 91.72
| | | --- arrival_month <= 10.50
| | | | --- truncated branch of depth 2
| | | --- arrival_month > 10.50
| | | | | --- truncated branch of depth 2
| --- avg_price_per_room > 108.50
| | --- avg_price_per_room <= 109.40
| | | --- weights: [40.00, 0.00] class: 0
| --- avg_price_per_room > 109.40
| | --- avg_price_per_room <= 110.49
| | | --- truncated branch of depth 2
| | --- avg_price_per_room > 110.49
| | | | --- truncated branch of depth 2
--- lead_time > 106.50
| --- avg_price_per_room <= 100.12
| | --- arrival_month <= 11.50
| | | --- arrival_month <= 10.50
| | | | --- truncated branch of depth 5
| | | --- arrival_month > 10.50
| | | | | --- truncated branch of depth 4
| --- arrival_month > 11.50
| | --- avg_price_per_room <= 77.50
| | | --- truncated branch of depth 2
| | --- avg_price_per_room > 77.50
| | | | --- truncated branch of depth 4
| --- avg_price_per_room > 100.12
| | --- arrival_month <= 11.00
| | | --- lead_time <= 122.00
| | | | --- truncated branch of depth 2
| | | --- lead_time > 122.00
| | | | | --- weights: [0.00, 2.00] class: 1
| | --- arrival_month > 11.00
| | | | --- weights: [0.00, 4.00] class: 1
--- type_of_meal_plan_Not Selected > 0.50
| --- avg_price_per_room <= 66.49
| | --- avg_price_per_room <= 61.58
| | | --- lead_time <= 47.00
| | | | --- weights: [0.00, 1.00] class: 1
| | --- lead_time > 47.00
| | | --- no_of_weekend_nights <= 3.50
```

```
|--- lead_time <= 99.50
|   |--- weights: [22.00, 0.00] class: 0
|--- lead_time > 99.50
|   |--- lead_time <= 109.00
|       |--- truncated branch of depth 2
|       |--- lead_time > 109.00
|           |--- weights: [5.00, 0.00] class: 0
|--- no_of_weekend_nights > 3.50
|   |--- weights: [0.00, 1.00] class: 1
|--- avg_price_per_room > 61.58
|   |--- avg_price_per_room <= 64.01
|       |--- arrival_month <= 5.50
|           |--- lead_time <= 90.00
|               |--- weights: [0.00, 9.00] class: 1
|           |--- lead_time > 90.00
|               |--- weights: [1.00, 0.00] class: 0
|       |--- arrival_month > 5.50
|           |--- lead_time <= 147.00
|               |--- lead_time <= 76.50
|                   |--- weights: [6.00, 0.00] class: 0
|               |--- lead_time > 76.50
|                   |--- truncated branch of depth 2
|               |--- lead_time > 147.00
|                   |--- weights: [0.00, 2.00] class: 1
|       |--- avg_price_per_room > 64.01
|           |--- weights: [8.00, 0.00] class: 0
|--- avg_price_per_room > 66.49
|   |--- avg_price_per_room <= 105.28
|       |--- required_car_parking_space <= 0.50
|           |--- lead_time <= 63.50
|               |--- arrival_month <= 5.50
|                   |--- arrival_month <= 4.50
|                       |--- truncated branch of depth 8
|                   |--- arrival_month > 4.50
|                       |--- truncated branch of depth 2
|               |--- arrival_month > 5.50
|                   |--- lead_time <= 47.50
|                       |--- truncated branch of depth 4
|                   |--- lead_time > 47.50
|                       |--- truncated branch of depth 5
|           |--- lead_time > 63.50
|               |--- avg_price_per_room <= 72.93
|                   |--- lead_time <= 73.50
|                       |--- weights: [0.00, 4.00] class: 1
|                   |--- lead_time > 73.50
|                       |--- truncated branch of depth 8
|               |--- avg_price_per_room > 72.93
|                   |--- no_of_adults <= 1.50
|                       |--- truncated branch of depth 4
|                   |--- no_of_adults > 1.50
|                       |--- truncated branch of depth 14
|--- required_car_parking_space > 0.50
|   |--- weights: [6.00, 0.00] class: 0
|--- avg_price_per_room > 105.28
|   |--- lead_time <= 47.50
|       |--- no_of_weekend_nights <= 1.00
|           |--- avg_price_per_room <= 107.55
|               |--- weights: [0.00, 1.00] class: 1
|           |--- avg_price_per_room > 107.55
|               |--- weights: [1.00, 0.00] class: 0
```





```
|--- lead_time <= 31.50
|   |--- truncated branch of depth 8
|--- lead_time >  31.50
|   |--- truncated branch of depth 15
--- type_of_meal_plan_Not Selected >  0.50
--- no_of_adults <= 1.50
|--- lead_time <= 62.50
|   |--- no_of_week_nights <= 3.50
|   |--- truncated branch of depth 3
|--- no_of_week_nights >  3.50
|   |--- weights: [0.00, 1.00] class: 1
--- lead_time >  62.50
|   |--- weights: [0.00, 3.00] class: 1
--- no_of_adults >  1.50
|--- lead_time <= 11.50
|   |--- no_of_week_nights <= 1.50
|   |--- weights: [0.00, 1.00] class: 1
|--- no_of_week_nights >  1.50
|   |--- weights: [2.00, 0.00] class: 0
--- lead_time >  11.50
|--- avg_price_per_room <= 117.50
|   |--- weights: [0.00, 25.00] class: 1
|--- avg_price_per_room >  117.50
|   |--- truncated branch of depth 8
--- avg_price_per_room >  140.26
--- arrival_month <= 9.50
|--- avg_price_per_room <= 194.50
|   |--- avg_price_per_room <= 159.65
|   |--- no_of_week_nights <= 4.50
|   |--- truncated branch of depth 17
|--- no_of_week_nights >  4.50
|   |--- weights: [4.00, 0.00] class: 0
--- avg_price_per_room >  159.65
|--- lead_time <= 54.00
|   |--- truncated branch of depth 8
|--- lead_time >  54.00
|   |--- truncated branch of depth 6
--- avg_price_per_room >  194.50
|--- weights: [0.00, 44.00] class: 1
--- arrival_month >  9.50
|--- lead_time <= 39.50
|--- arrival_month <= 11.50
|   |--- no_of_week_nights <= 1.50
|   |--- truncated branch of depth 3
|--- no_of_week_nights >  1.50
|   |--- truncated branch of depth 4
--- arrival_month >  11.50
|   |--- weights: [12.00, 0.00] class: 0
--- lead_time >  39.50
|--- arrival_month <= 11.50
|   |--- avg_price_per_room <= 167.50
|   |--- weights: [3.00, 0.00] class: 0
|--- avg_price_per_room >  167.50
|   |--- truncated branch of depth 2
--- arrival_month >  11.50
|   |--- weights: [0.00, 4.00] class: 1
--- no_of_weekend_nights >  0.50
--- arrival_month <= 7.50
|--- required_car_parking_space <= 0.50
|   |--- room_type_reserved_Room_Type 5 <= 0.50
```

```

    |--- arrival_month <= 2.50
    |   |--- no_of_weekend_nights <= 1.50
    |   |   |--- truncated branch of depth 4
    |   |--- no_of_weekend_nights >  1.50
    |   |   |--- truncated branch of depth 4
    |--- arrival_month >  2.50
    |   |--- avg_price_per_room <= 162.33
    |   |   |--- truncated branch of depth 17
    |   |--- avg_price_per_room >  162.33
    |   |   |--- truncated branch of depth 8
    |--- room_type_reserved_Room_Type 5 >  0.50
    |--- lead_time <= 41.50
    |   |--- weights: [6.00, 0.00] class: 0
    |--- lead_time >  41.50
    |   |--- arrival_month <= 4.50
    |   |   |--- weights: [1.00, 0.00] class: 0
    |   |--- arrival_month >  4.50
    |   |   |--- weights: [0.00, 1.00] class: 1
    |--- required_car_parking_space >  0.50
    |   |--- weights: [9.00, 0.00] class: 0
    --- arrival_month >  7.50
    |--- avg_price_per_room <= 203.04
    |   |--- avg_price_per_room <= 181.97
    |   |   |--- room_type_reserved_Room_Type 5 <= 0.50
    |   |   |   |--- avg_price_per_room <= 117.10
    |   |   |   |   |--- truncated branch of depth 6
    |   |   |   |--- avg_price_per_room >  117.10
    |   |   |   |   |--- truncated branch of depth 16
    |   |--- room_type_reserved_Room_Type 5 >  0.50
    |   |   |--- no_of_week_nights <= 2.50
    |   |   |   |--- weights: [9.00, 0.00] class: 0
    |   |   |--- no_of_week_nights >  2.50
    |   |   |   |--- truncated branch of depth 3
    |   |--- avg_price_per_room >  181.97
    |   |--- no_of_week_nights <= 3.50
    |   |   |--- arrival_month <= 8.50
    |   |   |   |--- truncated branch of depth 3
    |   |   |--- arrival_month >  8.50
    |   |   |   |--- weights: [33.00, 0.00] class: 0
    |   |--- no_of_week_nights >  3.50
    |   |   |   |--- weights: [0.00, 3.00] class: 1
    |--- avg_price_per_room >  203.04
    |   |--- arrival_month <= 11.00
    |   |   |--- weights: [0.00, 21.00] class: 1
    |--- arrival_month >  11.00
    |   |--- no_of_week_nights <= 1.50
    |   |   |--- weights: [0.00, 1.00] class: 1
    |   |--- no_of_week_nights >  1.50
    |   |   |--- weights: [3.00, 0.00] class: 0
    --- lead_time >  91.50
    |--- required_car_parking_space <= 0.50
    |   |--- arrival_month <= 7.50
    |   |   |--- arrival_month <= 4.50
    |   |   |   |--- lead_time <= 118.00
    |   |   |   |   |--- weights: [0.00, 36.00] class: 1
    |   |   |--- lead_time >  118.00
    |   |   |   |   |--- weights: [1.00, 0.00] class: 0
    |--- arrival_month >  4.50
    |   |--- avg_price_per_room <= 141.97
    |   |   |--- avg_price_per_room <= 139.72

```

```
|--- type_of_meal_plan_Meal Plan 2 <= 0.50
|   |--- truncated branch of depth 16
|--- type_of_meal_plan_Meal Plan 2 >  0.50
|   |--- truncated branch of depth 3
|--- avg_price_per_room >  139.72
|   |--- arrival_month <= 5.50
|       |--- weights: [6.00, 0.00] class: 0
|--- arrival_month >  5.50
|   |--- truncated branch of depth 5
|--- avg_price_per_room >  141.97
|   |--- no_of_adults <= 1.50
|       |--- arrival_month <= 6.00
|           |--- weights: [2.00, 0.00] class: 0
|--- arrival_month >  6.00
|   |--- weights: [0.00, 2.00] class: 1
|--- no_of_adults >  1.50
|   |--- no_of_weekend_nights <= 1.50
|       |--- truncated branch of depth 7
|--- no_of_weekend_nights >  1.50
|   |--- truncated branch of depth 4
|--- arrival_month >  7.50
|   |--- avg_price_per_room <= 113.58
|       |--- no_of_adults <= 2.50
|           |--- weights: [6.00, 0.00] class: 0
|--- no_of_adults >  2.50
|   |--- weights: [0.00, 1.00] class: 1
|--- avg_price_per_room >  113.58
|   |--- arrival_month <= 10.50
|       |--- no_of_week_nights <= 3.50
|           |--- no_of_adults <= 2.50
|               |--- truncated branch of depth 9
|--- no_of_adults >  2.50
|   |--- truncated branch of depth 6
|--- no_of_week_nights >  3.50
|   |--- no_of_weekend_nights <= 1.50
|       |--- weights: [0.00, 54.00] class: 1
|--- no_of_weekend_nights >  1.50
|   |--- truncated branch of depth 5
|--- arrival_month >  10.50
|   |--- no_of_adults <= 2.50
|       |--- no_of_weekend_nights <= 1.50
|           |--- weights: [0.00, 2.00] class: 1
|--- no_of_weekend_nights >  1.50
|   |--- truncated branch of depth 3
|--- no_of_adults >  2.50
|   |--- lead_time <= 102.50
|       |--- weights: [1.00, 0.00] class: 0
|--- lead_time >  102.50
|   |--- weights: [0.00, 6.00] class: 1
|--- required_car_parking_space >  0.50
|   |--- no_of_week_nights <= 8.00
|       |--- weights: [13.00, 0.00] class: 0
|--- no_of_week_nights >  8.00
|   |--- weights: [0.00, 1.00] class: 1
|--- no_of_special_requests >  0.50
|--- no_of_special_requests <= 1.50
|   |--- lead_time <= 8.50
|       |--- no_of_week_nights <= 10.00
|           |--- lead_time <= 4.50
|               |--- avg_price_per_room <= 219.86
```

```
    |--- avg_price_per_room <= 157.64
    |   |--- avg_price_per_room <= 78.71
    |   |   |--- weights: [230.00, 0.00] class: 0
    |   |--- avg_price_per_room > 78.71
    |   |   |--- avg_price_per_room <= 88.22
    |   |   |   |--- no_of_children <= 1.50
    |   |   |       |--- truncated branch of depth 9
    |   |   |   |--- no_of_children > 1.50
    |   |   |       |--- truncated branch of depth 2
    |   |--- avg_price_per_room > 88.22
    |   |   |--- no_of_week_nights <= 4.50
    |   |   |       |--- truncated branch of depth 13
    |   |   |   |--- no_of_week_nights > 4.50
    |   |   |       |--- truncated branch of depth 2
    |--- avg_price_per_room > 157.64
    |   |--- avg_price_per_room <= 158.50
    |   |   |--- weights: [0.00, 1.00] class: 1
    |   |--- avg_price_per_room > 158.50
    |   |   |--- no_of_week_nights <= 3.50
    |   |   |       |--- type_of_meal_plan_Meal Plan 2 <= 0.50
    |   |   |       |--- truncated branch of depth 5
    |   |   |       |--- type_of_meal_plan_Meal Plan 2 > 0.50
    |   |   |           |--- truncated branch of depth 2
    |   |   |--- no_of_week_nights > 3.50
    |   |   |       |--- lead_time <= 2.50
    |   |   |           |--- weights: [4.00, 0.00] class: 0
    |   |   |       |--- lead_time > 2.50
    |   |   |           |--- truncated branch of depth 2
    |--- avg_price_per_room > 219.86
    |   |--- arrival_month <= 6.00
    |   |   |--- weights: [0.00, 1.00] class: 1
    |--- arrival_month > 6.00
    |   |--- avg_price_per_room <= 237.25
    |   |   |--- weights: [5.00, 0.00] class: 0
    |   |--- avg_price_per_room > 237.25
    |   |   |--- no_of_week_nights <= 0.50
    |   |   |       |--- weights: [0.00, 1.00] class: 1
    |   |   |   |--- no_of_week_nights > 0.50
    |   |   |       |--- avg_price_per_room <= 244.12
    |   |   |           |--- weights: [0.00, 1.00] class: 1
    |   |   |       |--- avg_price_per_room > 244.12
    |   |   |           |--- weights: [4.00, 0.00] class: 0
    |--- lead_time > 4.50
    |   |--- avg_price_per_room <= 123.60
    |   |   |--- no_of_children <= 5.50
    |   |   |   |--- no_of_adults <= 0.50
    |   |   |       |--- weights: [0.00, 1.00] class: 1
    |   |   |   |--- no_of_adults > 0.50
    |   |   |       |--- arrival_month <= 8.50
    |   |   |           |--- avg_price_per_room <= 85.90
    |   |   |           |--- truncated branch of depth 6
    |   |   |       |--- avg_price_per_room > 85.90
    |   |   |           |--- truncated branch of depth 10
    |   |   |       |--- arrival_month > 8.50
    |   |   |           |--- weights: [131.00, 0.00] class: 0
    |--- no_of_children > 5.50
    |   |--- weights: [0.00, 1.00] class: 1
    |--- avg_price_per_room > 123.60
    |   |--- type_of_meal_plan_Not Selected <= 0.50
    |       |--- avg_price_per_room <= 124.05
```

```
    |--- arrival_month <= 6.00
    |   |--- weights: [2.00, 0.00] class: 0
    |--- arrival_month >  6.00
    |   |--- weights: [0.00, 2.00] class: 1
    --- avg_price_per_room > 124.05
    |--- arrival_month <= 2.50
        |--- avg_price_per_room <= 128.50
            |--- truncated branch of depth 2
        |--- avg_price_per_room > 128.50
            |--- weights: [2.00, 0.00] class: 0
    |--- arrival_month >  2.50
        |--- arrival_month <= 6.50
            |--- weights: [40.00, 0.00] class: 0
        |--- arrival_month >  6.50
            |--- truncated branch of depth 13
    |--- type_of_meal_plan_Not Selected >  0.50
        |--- arrival_month <= 9.50
            |--- no_of_week_nights <= 1.50
                |--- avg_price_per_room <= 128.50
                    |--- weights: [3.00, 0.00] class: 0
                |--- avg_price_per_room > 128.50
                    |--- truncated branch of depth 4
            |--- no_of_week_nights >  1.50
                |--- weights: [11.00, 0.00] class: 0
        |--- arrival_month >  9.50
            |--- lead_time <= 6.50
                |--- weights: [1.00, 0.00] class: 0
            |--- lead_time >  6.50
                |--- weights: [1.00, 4.00] class: 1
    |--- no_of_week_nights >  10.00
        |--- arrival_month <= 1.50
            |--- weights: [1.00, 0.00] class: 0
        |--- arrival_month >  1.50
            |--- weights: [0.00, 3.00] class: 1
--- lead_time >  8.50
    |--- avg_price_per_room <= 118.55
        |--- no_of_weekend_nights <= 2.50
            |--- lead_time <= 61.50
                |--- arrival_month <= 11.50
                    |--- avg_price_per_room <= 66.30
                        |--- no_of_week_nights <= 2.50
                            |--- weights: [97.00, 0.00] class: 0
                        |--- no_of_week_nights >  2.50
                            |--- lead_time <= 16.50
                                |--- truncated branch of depth 2
                            |--- lead_time >  16.50
                                |--- weights: [81.00, 1.00] class: 0
                    |--- avg_price_per_room > 66.30
                        |--- type_of_meal_plan_Not Selected <= 0.50
                            |--- avg_price_per_room <= 90.99
                                |--- truncated branch of depth 19
                            |--- avg_price_per_room > 90.99
                                |--- truncated branch of depth 17
                        |--- type_of_meal_plan_Not Selected >  0.50
                            |--- no_of_week_nights <= 4.50
                                |--- truncated branch of depth 18
                            |--- no_of_week_nights >  4.50
                                |--- truncated branch of depth 2
    |--- arrival_month >  11.50
        |--- weights: [233.00, 0.00] class: 0
```

```
    |--- lead_time >  61.50
    |--- arrival_month <= 6.50
    |--- avg_price_per_room <= 71.98
    |--- lead_time <= 72.50
    |--- lead_time <= 71.50
    |--- truncated branch of depth 3
    |--- lead_time >  71.50
    |--- truncated branch of depth 2
    |--- lead_time >  72.50
    |--- weights: [76.00, 0.00] class: 0
    |--- avg_price_per_room >  71.98
    |--- avg_price_per_room <= 72.37
    |--- no_of_week_nights <= 4.00
    |--- weights: [0.00, 7.00] class: 1
    |--- no_of_week_nights >  4.00
    |--- weights: [3.00, 0.00] class: 0
    |--- avg_price_per_room >  72.37
    |--- lead_time <= 126.50
    |--- truncated branch of depth 15
    |--- lead_time >  126.50
    |--- truncated branch of depth 9
    |--- arrival_month >  6.50
    |--- no_of_week_nights <= 1.50
    |--- no_of_adults <= 1.50
    |--- lead_time <= 135.00
    |--- truncated branch of depth 4
    |--- lead_time >  135.00
    |--- truncated branch of depth 2
    |--- no_of_adults >  1.50
    |--- type_of_meal_plan_Not Selected <= 0.50
    |--- truncated branch of depth 12
    |--- type_of_meal_plan_Not Selected >  0.50
    |--- truncated branch of depth 9
    |--- no_of_week_nights >  1.50
    |--- avg_price_per_room <= 76.60
    |--- arrival_month <= 7.50
    |--- truncated branch of depth 9
    |--- arrival_month >  7.50
    |--- truncated branch of depth 11
    |--- avg_price_per_room >  76.60
    |--- arrival_month <= 8.50
    |--- truncated branch of depth 9
    |--- arrival_month >  8.50
    |--- truncated branch of depth 13
    |--- no_of_weekend_nights >  2.50
    |--- arrival_month <= 1.50
    |--- weights: [3.00, 0.00] class: 0
    |--- arrival_month >  1.50
    |--- no_of_week_nights <= 8.50
    |--- avg_price_per_room <= 106.66
    |--- arrival_month <= 2.50
    |--- weights: [1.00, 0.00] class: 0
    |--- arrival_month >  2.50
    |--- arrival_month <= 11.50
    |--- truncated branch of depth 2
    |--- arrival_month >  11.50
    |--- weights: [1.00, 0.00] class: 0
    |--- avg_price_per_room >  106.66
    |--- weights: [3.00, 0.00] class: 0
    |--- no_of_week_nights >  8.50
```

```
| | | | --- weights: [0.00, 11.00] class: 1
--- avg_price_per_room > 118.55
| --- required_car_parking_space <= 0.50
| | --- arrival_month <= 8.50
| | | --- no_of_week_nights <= 7.50
| | | | --- no_of_adults <= 2.50
| | | | | --- lead_time <= 138.50
| | | | | | --- avg_price_per_room <= 127.80
| | | | | | | --- truncated branch of depth 12
| | | | | | --- avg_price_per_room > 127.80
| | | | | | | --- truncated branch of depth 20
| | | | | --- lead_time > 138.50
| | | | | | --- no_of_children <= 0.50
| | | | | | | --- truncated branch of depth 4
| | | | | | --- no_of_children > 0.50
| | | | | | | --- truncated branch of depth 2
| | | | | --- no_of_adults > 2.50
| | | | | | --- avg_price_per_room <= 131.55
| | | | | | | --- arrival_month <= 2.50
| | | | | | | | --- truncated branch of depth 4
| | | | | | | --- arrival_month > 2.50
| | | | | | | | --- truncated branch of depth 6
| | | | | | --- avg_price_per_room > 131.55
| | | | | | | --- avg_price_per_room <= 182.20
| | | | | | | | --- truncated branch of depth 10
| | | | | | | --- avg_price_per_room > 182.20
| | | | | | | | --- truncated branch of depth 4
| | | | | --- no_of_week_nights > 7.50
| | | | | | --- weights: [0.00, 4.00] class: 1
--- arrival_month > 8.50
| --- arrival_month <= 11.50
| | --- lead_time <= 40.50
| | | --- no_of_weekend_nights <= 0.50
| | | | --- arrival_month <= 10.50
| | | | | --- truncated branch of depth 13
| | | | --- arrival_month > 10.50
| | | | | --- truncated branch of depth 5
| | | | --- no_of_weekend_nights > 0.50
| | | | | --- type_of_meal_plan_Not Selected <= 0.50
| | | | | | --- truncated branch of depth 15
| | | | | --- type_of_meal_plan_Not Selected > 0.50
| | | | | | --- truncated branch of depth 8
| | | --- lead_time > 40.50
| | | | --- no_of_weekend_nights <= 0.50
| | | | | --- arrival_month <= 9.50
| | | | | | --- truncated branch of depth 11
| | | | | --- arrival_month > 9.50
| | | | | | --- truncated branch of depth 9
| | | | --- no_of_weekend_nights > 0.50
| | | | | --- avg_price_per_room <= 119.53
| | | | | | --- truncated branch of depth 5
| | | | | --- avg_price_per_room > 119.53
| | | | | | --- truncated branch of depth 20
| | --- arrival_month > 11.50
| | | --- lead_time <= 100.00
| | | | --- weights: [80.00, 0.00] class: 0
| | | --- lead_time > 100.00
| | | | --- room_type_reserved_Room_Type 7 <= 0.50
| | | | | --- weights: [0.00, 11.00] class: 1
| | | | --- room_type_reserved_Room_Type 7 > 0.50
```

```
|--- lead_time <= 129.50
|   |--- weights: [2.00, 0.00] class: 0
|   |--- lead_time > 129.50
|   |   |--- weights: [0.00, 1.00] class: 1
|--- required_car_parking_space > 0.50
|   |--- room_type_reserved_Room_Type 7 <= 0.50
|   |   |--- weights: [103.00, 0.00] class: 0
|   |   |--- room_type_reserved_Room_Type 7 > 0.50
|   |   |   |--- weights: [0.00, 1.00] class: 1
--- no_of_special_requests > 1.50
|--- lead_time <= 90.50
|   |--- no_of_week_nights <= 3.50
|   |   |--- weights: [2126.00, 0.00] class: 0
|   |--- no_of_week_nights > 3.50
|   |   |--- no_of_week_nights <= 9.50
|   |   |   |--- no_of_special_requests <= 2.50
|   |   |   |   |--- lead_time <= 6.50
|   |   |   |   |   |--- weights: [43.00, 0.00] class: 0
|   |   |   |   |--- lead_time > 6.50
|   |   |   |   |   |--- arrival_month <= 11.50
|   |   |   |   |   |   |--- avg_price_per_room <= 93.09
|   |   |   |   |   |   |--- lead_time <= 54.00
|   |   |   |   |   |   |   |--- truncated branch of depth 8
|   |   |   |   |   |   |--- lead_time > 54.00
|   |   |   |   |   |   |   |--- truncated branch of depth 6
|   |   |   |   |   |--- avg_price_per_room > 93.09
|   |   |   |   |   |   |--- avg_price_per_room <= 107.29
|   |   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |   |--- avg_price_per_room > 107.29
|   |   |   |   |   |   |   |--- truncated branch of depth 13
|   |   |   |   |   |--- arrival_month > 11.50
|   |   |   |   |   |   |--- weights: [26.00, 0.00] class: 0
|   |   |   |   |--- no_of_special_requests > 2.50
|   |   |   |   |   |--- weights: [70.00, 0.00] class: 0
|   |   |   |--- no_of_week_nights > 9.50
|   |   |   |   |--- weights: [0.00, 2.00] class: 1
|--- lead_time > 90.50
|   |--- avg_price_per_room <= 202.95
|   |   |--- arrival_month <= 8.50
|   |   |   |--- lead_time <= 150.50
|   |   |   |   |--- lead_time <= 141.50
|   |   |   |   |   |--- lead_time <= 108.50
|   |   |   |   |   |   |--- lead_time <= 107.50
|   |   |   |   |   |   |--- no_of_week_nights <= 1.50
|   |   |   |   |   |   |   |--- weights: [27.00, 0.00] class: 0
|   |   |   |   |   |--- no_of_week_nights > 1.50
|   |   |   |   |   |   |--- truncated branch of depth 9
|   |   |   |   |--- lead_time > 107.50
|   |   |   |   |   |--- avg_price_per_room <= 125.53
|   |   |   |   |   |   |--- truncated branch of depth 3
|   |   |   |   |   |--- avg_price_per_room > 125.53
|   |   |   |   |   |   |--- weights: [3.00, 0.00] class: 0
|   |   |   |--- lead_time > 108.50
|   |   |   |--- no_of_children <= 0.50
|   |   |   |   |--- no_of_week_nights <= 1.50
|   |   |   |   |   |--- truncated branch of depth 5
|   |   |   |   |   |--- no_of_week_nights > 1.50
|   |   |   |   |   |   |--- truncated branch of depth 5
|   |   |   |--- no_of_children > 0.50
|   |   |   |   |--- no_of_week_nights <= 1.50
```

```
|--- weights: [12.00, 0.00] class: 0
|--- no_of_week_nights > 1.50
|--- truncated branch of depth 4
--- lead_time > 141.50
    |--- avg_price_per_room <= 92.72
    |--- avg_price_per_room <= 74.38
        |--- weights: [3.00, 0.00] class: 0
    --- avg_price_per_room > 74.38
        |--- required_car_parking_space <= 0.50
            |--- weights: [0.00, 3.00] class: 1
        |--- required_car_parking_space > 0.50
            |--- weights: [1.00, 0.00] class: 0
    --- avg_price_per_room > 92.72
        |--- lead_time <= 144.50
            |--- no_of_children <= 0.50
                |--- truncated branch of depth 2
            --- no_of_children > 0.50
                |--- weights: [0.00, 2.00] class: 1
        --- lead_time > 144.50
            |--- arrival_month <= 7.50
                |--- weights: [16.00, 0.00] class: 0
            --- arrival_month > 7.50
                |--- truncated branch of depth 2
    --- lead_time > 150.50
        |--- avg_price_per_room <= 103.50
            |--- weights: [2.00, 0.00] class: 0
        --- avg_price_per_room > 103.50
            |--- weights: [0.00, 3.00] class: 1
--- arrival_month > 8.50
    |--- no_of_special_requests <= 2.50
        |--- avg_price_per_room <= 90.42
            |--- arrival_month <= 11.50
                |--- lead_time <= 107.00
                    |--- avg_price_per_room <= 70.52
                        |--- weights: [3.00, 0.00] class: 0
                    --- avg_price_per_room > 70.52
                        |--- truncated branch of depth 5
                --- lead_time > 107.00
                    |--- arrival_month <= 9.50
                        |--- weights: [0.00, 1.00] class: 1
                    --- arrival_month > 9.50
                        |--- truncated branch of depth 4
                --- arrival_month > 11.50
                    |--- lead_time <= 101.00
                        |--- weights: [11.00, 0.00] class: 0
                --- lead_time > 101.00
                    |--- lead_time <= 104.00
                        |--- weights: [0.00, 1.00] class: 1
                    --- lead_time > 104.00
                        |--- truncated branch of depth 5
    --- avg_price_per_room > 90.42
        |--- no_of_adults <= 1.50
            |--- weights: [11.00, 0.00] class: 0
        --- no_of_adults > 1.50
            |--- avg_price_per_room <= 153.15
                |--- no_of_weekend_nights <= 1.50
                    |--- truncated branch of depth 9
                --- no_of_weekend_nights > 1.50
                    |--- truncated branch of depth 6
            --- avg_price_per_room > 153.15
```

```
| | | | | | | --- lead_time <= 100.00
| | | | | | | --- truncated branch of depth 2
| | | | | | | --- lead_time > 100.00
| | | | | | | --- truncated branch of depth 3
| | | | | | | --- no_of_special_requests > 2.50
| | | | | | | --- weights: [52.00, 0.00] class: 0
| | | | | | | --- avg_price_per_room > 202.95
| | | | | | | --- weights: [0.00, 7.00] class: 1
--- lead_time > 151.50
| --- avg_price_per_room <= 100.04
| | --- no_of_special_requests <= 0.50
| | | --- no_of_adults <= 1.50
| | | | --- no_of_week_nights <= 4.50
| | | | | --- lead_time <= 163.50
| | | | | | --- lead_time <= 162.50
| | | | | | | --- lead_time <= 160.50
| | | | | | | | --- weights: [4.00, 0.00] class: 0
| | | | | | | --- lead_time > 160.50
| | | | | | | | --- weights: [1.00, 1.00] class: 0
| | | | | | | --- lead_time > 162.50
| | | | | | | | --- weights: [0.00, 15.00] class: 1
| | | | | | | --- lead_time > 163.50
| | | | | | | | --- type_of_meal_plan_Not Selected <= 0.50
| | | | | | | | --- no_of_children <= 0.50
| | | | | | | | --- lead_time <= 341.00
| | | | | | | | | --- lead_time <= 255.50
| | | | | | | | | | --- lead_time <= 226.00
| | | | | | | | | | | --- truncated branch of depth 7
| | | | | | | | | | --- lead_time > 226.00
| | | | | | | | | | | --- truncated branch of depth 6
| | | | | | | | | | --- lead_time > 255.50
| | | | | | | | | | | --- avg_price_per_room <= 98.00
| | | | | | | | | | | | --- truncated branch of depth 4
| | | | | | | | | | | --- avg_price_per_room > 98.00
| | | | | | | | | | | | | --- weights: [0.00, 1.00] class: 1
| | | | | | | | | | --- lead_time > 341.00
| | | | | | | | | | | --- arrival_month <= 10.50
| | | | | | | | | | | | --- lead_time <= 402.00
| | | | | | | | | | | | | --- truncated branch of depth 3
| | | | | | | | | | | | --- lead_time > 402.00
| | | | | | | | | | | | | --- weights: [0.00, 3.00] class: 1
| | | | | | | | | | | --- arrival_month > 10.50
| | | | | | | | | | | | --- weights: [0.00, 4.00] class: 1
| | | | | | | | | | | --- no_of_children > 0.50
| | | | | | | | | | | | --- avg_price_per_room <= 55.53
| | | | | | | | | | | | | --- weights: [1.00, 0.00] class: 0
| | | | | | | | | | | | --- avg_price_per_room > 55.53
| | | | | | | | | | | | | --- weights: [0.00, 9.00] class: 1
| | | | | | | | | | | --- type_of_meal_plan_Not Selected > 0.50
| | | | | | | | | | | | --- weights: [0.00, 8.00] class: 1
--- no_of_week_nights > 4.50
| --- type_of_meal_plan_Meal Plan 2 <= 0.50
| | --- avg_price_per_room <= 68.79
| | | --- lead_time <= 188.00
| | | | --- weights: [0.00, 1.00] class: 1
| | | --- lead_time > 188.00
| | | | --- weights: [1.00, 0.00] class: 0
| | | --- avg_price_per_room > 68.79
| | | | --- avg_price_per_room <= 93.56
| | | | | --- weights: [0.00, 15.00] class: 1
```

```
| | | | | --- avg_price_per_room > 93.56
| | | | | --- weights: [1.00, 1.00] class: 0
| | | | --- type_of_meal_plan_Meal Plan 2 > 0.50
| | | | | --- weights: [1.00, 0.00] class: 0
| | | --- no_of_adults > 1.50
| | | --- avg_price_per_room <= 82.47
| | | | --- lead_time <= 170.50
| | | | | --- type_of_meal_plan_Not Selected <= 0.50
| | | | | --- no_of_weekend_nights <= 2.50
| | | | | | --- avg_price_per_room <= 24.50
| | | | | | --- weights: [0.00, 1.00] class: 1
| | | | | --- avg_price_per_room > 24.50
| | | | | | --- arrival_month <= 9.50
| | | | | | | --- avg_price_per_room <= 77.23
| | | | | | | --- truncated branch of depth 5
| | | | | | | --- avg_price_per_room > 77.23
| | | | | | | --- weights: [10.00, 0.00] class: 0
| | | | | --- arrival_month > 9.50
| | | | | | | --- weights: [26.00, 0.00] class: 0
| | | | | --- no_of_weekend_nights > 2.50
| | | | | | --- weights: [0.00, 2.00] class: 1
| | | | --- type_of_meal_plan_Not Selected > 0.50
| | | | | --- lead_time <= 155.50
| | | | | | --- weights: [1.00, 0.00] class: 0
| | | | | --- lead_time > 155.50
| | | | | | --- weights: [0.00, 6.00] class: 1
| | | --- lead_time > 170.50
| | | | --- type_of_meal_plan_Not Selected <= 0.50
| | | | | --- arrival_month <= 6.50
| | | | | | --- lead_time <= 187.00
| | | | | | | --- weights: [9.00, 0.00] class: 0
| | | | | | --- lead_time > 187.00
| | | | | | | --- avg_price_per_room <= 77.62
| | | | | | | | --- no_of_weekend_nights <= 0.50
| | | | | | | | --- weights: [0.00, 73.00] class: 1
| | | | | | | --- no_of_weekend_nights > 0.50
| | | | | | | | --- truncated branch of depth 4
| | | | | | | --- avg_price_per_room > 77.62
| | | | | | | | --- no_of_week_nights <= 3.50
| | | | | | | | --- truncated branch of depth 4
| | | | | | | --- no_of_week_nights > 3.50
| | | | | | | | --- weights: [8.00, 0.00] class: 0
| | | | --- arrival_month > 6.50
| | | | | --- lead_time <= 337.00
| | | | | | --- lead_time <= 314.50
| | | | | | | --- lead_time <= 289.50
| | | | | | | | --- truncated branch of depth 15
| | | | | | | --- lead_time > 289.50
| | | | | | | | --- truncated branch of depth 3
| | | | | | | --- lead_time > 314.50
| | | | | | | | --- avg_price_per_room <= 64.62
| | | | | | | | | --- weights: [33.00, 0.00] class: 0
| | | | | | | | --- avg_price_per_room > 64.62
| | | | | | | | | --- truncated branch of depth 2
| | | | | | | --- lead_time > 337.00
| | | | | | | | --- weights: [0.00, 39.00] class: 1
| | | | | --- type_of_meal_plan_Not Selected > 0.50
| | | | | --- avg_price_per_room <= 54.57
| | | | | | --- arrival_month <= 10.00
| | | | | | | --- weights: [0.00, 1.00] class: 1
```

```
    |--- arrival_month > 10.00
    |   |--- weights: [1.00, 0.00] class: 0
    |--- avg_price_per_room > 54.57
    |   |--- no_of_week_nights <= 1.50
    |       |--- lead_time <= 214.50
    |           |--- lead_time <= 202.00
    |               |--- weights: [0.00, 1.00] class: 1
    |               |--- lead_time > 202.00
    |                   |--- weights: [1.00, 0.00] class: 0
    |--- lead_time > 214.50
    |   |--- weights: [0.00, 6.00] class: 1
    |--- no_of_week_nights > 1.50
    |   |--- weights: [0.00, 62.00] class: 1
    --- avg_price_per_room > 82.47
    |--- no_of_adults <= 2.50
    |--- lead_time <= 324.50
        |--- arrival_month <= 11.50
            |--- avg_price_per_room <= 90.78
                |--- no_of_week_nights <= 4.50
                    |--- weights: [1.00, 321.00] class: 1
                |--- no_of_week_nights > 4.50
                    |--- type_of_meal_plan_Not Selected <= 0.50
                        |--- truncated branch of depth 2
                    |--- type_of_meal_plan_Not Selected > 0.50
                        |--- weights: [0.00, 11.00] class: 1
            |--- avg_price_per_room > 90.78
                |--- avg_price_per_room <= 93.04
                    |--- lead_time <= 297.50
                        |--- truncated branch of depth 8
                    |--- lead_time > 297.50
                        |--- weights: [4.00, 0.00] class: 0
            |--- avg_price_per_room > 93.04
                |--- lead_time <= 192.50
                    |--- truncated branch of depth 4
                |--- lead_time > 192.50
                    |--- weights: [0.00, 178.00] class: 1
            |--- arrival_month > 11.50
                |--- no_of_week_nights <= 2.50
                    |--- avg_price_per_room <= 87.00
                        |--- weights: [0.00, 1.00] class: 1
                    |--- avg_price_per_room > 87.00
                        |--- weights: [6.00, 0.00] class: 0
                |--- no_of_week_nights > 2.50
                    |--- no_of_weekend_nights <= 0.50
                        |--- lead_time <= 285.00
                            |--- weights: [1.00, 0.00] class: 0
                        |--- lead_time > 285.00
                            |--- weights: [0.00, 2.00] class: 1
                    |--- no_of_weekend_nights > 0.50
                        |--- weights: [0.00, 10.00] class: 1
                |--- lead_time > 324.50
                    |--- avg_price_per_room <= 89.00
                        |--- weights: [8.00, 0.00] class: 0
                    |--- avg_price_per_room > 89.00
                        |--- weights: [1.00, 9.00] class: 1
                |--- no_of_adults > 2.50
                    |--- weights: [7.00, 0.00] class: 0
--- no_of_special_requests > 0.50
|--- no_of_weekend_nights <= 0.50
|--- lead_time <= 180.50
```

```

|--- lead_time <= 159.50
|   |--- arrival_month <= 8.50
|   |   |--- weights: [8.00, 0.00] class: 0
|   |--- arrival_month >  8.50
|   |   |--- no_of_week_nights <= 2.50
|   |   |   |--- weights: [0.00, 4.00] class: 1
|   |   |--- no_of_week_nights >  2.50
|   |   |   |--- lead_time <= 156.50
|   |   |       |--- weights: [0.00, 1.00] class: 1
|   |   |--- lead_time >  156.50
|   |   |       |--- weights: [2.00, 0.00] class: 0
|--- lead_time >  159.50
|   |--- no_of_adults <= 0.50
|   |   |--- weights: [0.00, 1.00] class: 1
|   |--- no_of_adults >  0.50
|   |   |--- lead_time <= 178.50
|   |   |       |--- weights: [44.00, 0.00] class: 0
|   |   |--- lead_time >  178.50
|   |   |       |--- avg_price_per_room <= 93.62
|   |   |           |--- weights: [5.00, 0.00] class: 0
|   |   |       |--- avg_price_per_room >  93.62
|   |   |           |--- avg_price_per_room <= 96.92
|   |   |               |--- weights: [0.00, 2.00] class: 1
|   |   |           |--- avg_price_per_room >  96.92
|   |   |               |--- weights: [1.00, 0.00] class: 0
|--- lead_time >  180.50
|   |--- no_of_special_requests <= 2.50
|   |   |--- avg_price_per_room <= 72.58
|   |   |   |--- lead_time <= 238.00
|   |   |       |--- lead_time <= 198.00
|   |   |           |--- weights: [2.00, 0.00] class: 0
|   |   |       |--- lead_time >  198.00
|   |   |           |--- avg_price_per_room <= 57.75
|   |   |               |--- weights: [1.00, 0.00] class: 0
|   |   |           |--- avg_price_per_room >  57.75
|   |   |               |--- weights: [0.00, 9.00] class: 1
|   |   |--- lead_time >  238.00
|   |   |       |--- lead_time <= 304.50
|   |   |           |--- weights: [7.00, 0.00] class: 0
|   |   |       |--- lead_time >  304.50
|   |   |           |--- weights: [0.00, 1.00] class: 1
|--- avg_price_per_room >  72.58
|   |--- arrival_month <= 11.50
|   |   |--- type_of_meal_plan_Meal Plan 2 <= 0.50
|   |   |   |--- lead_time <= 322.50
|   |   |       |--- arrival_month <= 6.50
|   |   |           |--- truncated branch of depth 3
|   |   |       |--- arrival_month >  6.50
|   |   |           |--- truncated branch of depth 3
|   |   |--- lead_time >  322.50
|   |   |       |--- weights: [2.00, 1.00] class: 0
|--- type_of_meal_plan_Meal Plan 2 >  0.50
|   |--- avg_price_per_room <= 93.00
|   |   |--- weights: [0.00, 3.00] class: 1
|   |--- avg_price_per_room >  93.00
|   |   |--- weights: [3.00, 0.00] class: 0
|--- arrival_month >  11.50
|   |--- lead_time <= 272.00
|       |--- type_of_meal_plan_Not Selected <= 0.50
|           |   |--- weights: [6.00, 0.00] class: 0

```

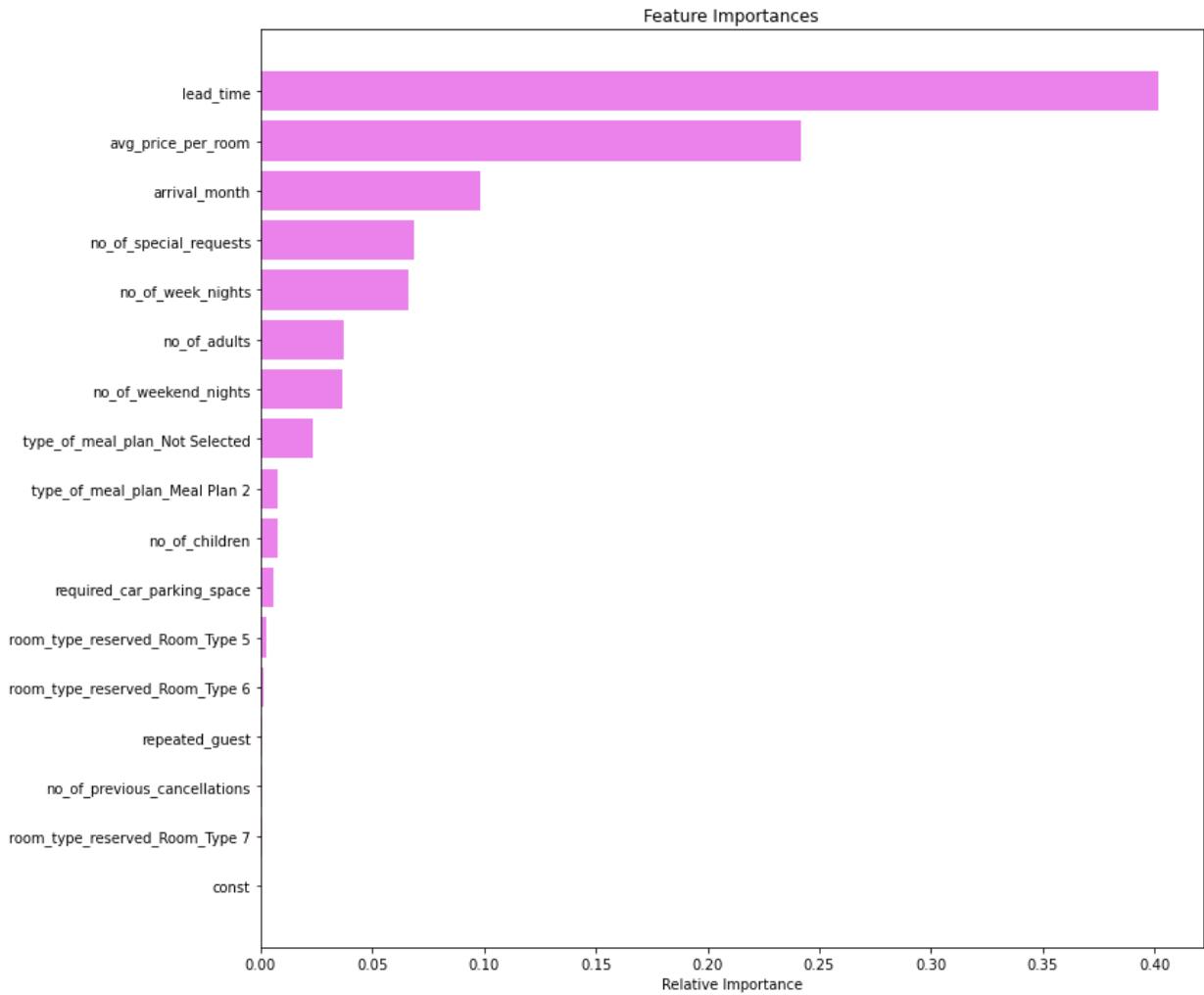
```
|--- type_of_meal_plan_Not Selected > 0.50
|   |--- weights: [0.00, 2.00] class: 1
|--- lead_time > 272.00
|   |--- avg_price_per_room <= 78.90
|   |   |--- lead_time <= 300.50
|   |       |--- truncated branch of depth 2
|   |       |--- lead_time > 300.50
|   |           |--- weights: [0.00, 2.00] class: 1
|   |           |--- avg_price_per_room > 78.90
|   |               |--- weights: [0.00, 5.00] class: 1
|--- no_of_special_requests > 2.50
|   |--- weights: [12.00, 0.00] class: 0
--- no_of_weekend_nights > 0.50
|--- no_of_week_nights <= 8.50
|   |--- arrival_month <= 11.50
|       |--- avg_price_per_room <= 76.48
|       |   |--- no_of_special_requests <= 1.50
|       |       |--- lead_time <= 245.50
|       |           |--- weights: [81.00, 0.00] class: 0
|       |       |--- lead_time > 245.50
|       |           |--- lead_time <= 249.00
|       |               |--- weights: [0.00, 1.00] class: 1
|       |           |--- lead_time > 249.00
|       |               |--- weights: [18.00, 0.00] class: 0
|       |       |--- no_of_special_requests > 1.50
|       |           |--- lead_time <= 285.50
|       |               |--- weights: [22.00, 1.00] class: 0
|       |           |--- lead_time > 285.50
|       |               |--- weights: [0.00, 1.00] class: 1
|--- avg_price_per_room > 76.48
|   |--- no_of_week_nights <= 6.50
|       |--- arrival_month <= 9.50
|           |--- no_of_children <= 1.50
|               |--- lead_time <= 152.50
|                   |--- truncated branch of depth 3
|                   |--- lead_time > 152.50
|                       |--- truncated branch of depth 14
|           |--- no_of_children > 1.50
|               |--- weights: [0.00, 3.00] class: 1
|--- arrival_month > 9.50
|   |--- no_of_weekend_nights <= 1.50
|       |--- avg_price_per_room <= 90.92
|           |--- weights: [24.00, 0.00] class: 0
|       |--- avg_price_per_room > 90.92
|           |--- truncated branch of depth 3
|   |--- no_of_weekend_nights > 1.50
|       |--- no_of_week_nights <= 1.50
|           |--- truncated branch of depth 6
|       |--- no_of_week_nights > 1.50
|           |--- truncated branch of depth 7
|--- no_of_week_nights > 6.50
|   |--- lead_time <= 201.00
|       |--- weights: [0.00, 2.00] class: 1
|--- lead_time > 201.00
|   |--- lead_time <= 204.50
|       |--- weights: [3.00, 0.00] class: 0
|   |--- lead_time > 204.50
|       |--- arrival_month <= 7.50
|           |--- truncated branch of depth 2
|       |--- arrival_month > 7.50
```

```
| | | | | | | --- weights: [0.00, 2.00] class: 1
| --- arrival_month > 11.50
| --- lead_time <= 316.00
| | --- avg_price_per_room <= 55.92
| | | --- weights: [5.00, 0.00] class: 0
| | --- avg_price_per_room > 55.92
| | | --- avg_price_per_room <= 61.06
| | | | --- weights: [0.00, 2.00] class: 1
| | | --- avg_price_per_room > 61.06
| | | | --- type_of_meal_plan_Not Selected <= 0.50
| | | | | --- avg_price_per_room <= 77.19
| | | | | | --- weights: [8.00, 0.00] class: 0
| | | | --- avg_price_per_room > 77.19
| | | | | | --- truncated branch of depth 7
| | | | | | --- type_of_meal_plan_Not Selected > 0.50
| | | | | | | --- no_of_special_requests <= 1.50
| | | | | | | | --- truncated branch of depth 2
| | | | | | | --- no_of_special_requests > 1.50
| | | | | | | | --- truncated branch of depth 3
| | | --- lead_time > 316.00
| | | | --- no_of_weekend_nights <= 2.50
| | | | | | --- weights: [0.00, 4.00] class: 1
| | | | | | --- no_of_weekend_nights > 2.50
| | | | | | | | --- weights: [1.00, 0.00] class: 0
| | --- no_of_week_nights > 8.50
| | | --- no_of_special_requests <= 2.50
| | | | | --- weights: [0.00, 8.00] class: 1
| | | | --- no_of_special_requests > 2.50
| | | | | | --- weights: [2.00, 0.00] class: 0
| --- avg_price_per_room > 100.04
| --- arrival_month <= 11.50
| | --- no_of_special_requests <= 2.50
| | | | --- weights: [0.00, 2108.00] class: 1
| | --- no_of_special_requests > 2.50
| | | | --- weights: [31.00, 0.00] class: 0
| --- arrival_month > 11.50
| | --- no_of_special_requests <= 0.50
| | | | --- weights: [47.00, 0.00] class: 0
| | --- no_of_special_requests > 0.50
| | | --- lead_time <= 289.50
| | | | --- no_of_special_requests <= 1.50
| | | | | --- avg_price_per_room <= 114.59
| | | | | | --- weights: [2.00, 0.00] class: 0
| | | | --- avg_price_per_room > 114.59
| | | | | | --- weights: [0.00, 6.00] class: 1
| | | --- no_of_special_requests > 1.50
| | | | --- avg_price_per_room <= 199.65
| | | | | | --- no_of_week_nights <= 4.50
| | | | | | | --- weights: [7.00, 0.00] class: 0
| | | | | --- no_of_week_nights > 4.50
| | | | | | --- lead_time <= 188.50
| | | | | | | --- weights: [1.00, 0.00] class: 0
| | | | | --- lead_time > 188.50
| | | | | | | --- weights: [0.00, 1.00] class: 1
| | | | --- avg_price_per_room > 199.65
| | | | | | --- weights: [0.00, 1.00] class: 1
| --- lead_time > 289.50
| | | --- weights: [0.00, 7.00] class: 1
```

```
In [456...]  
# importance of features in the tree building ( The importance of a feature is compute  
# (normalized) total reduction of the 'criterion' brought by that feature. It is also  
  
print(  
    pd.DataFrame(  
        best_model.feature_importances_, columns=["Imp"], index=X_train3.columns  
    ).sort_values(by="Imp", ascending=False)  
)
```

	Imp
lead_time	0.40148
avg_price_per_room	0.24194
arrival_month	0.09817
no_of_special_requests	0.06841
no_of_week_nights	0.06604
no_of_adults	0.03722
no_of_weekend_nights	0.03648
type_of_meal_plan_Not Selected	0.02315
type_of_meal_plan_Meal Plan 2	0.00762
no_of_children	0.00738
required_car_parking_space	0.00583
room_type_reserved_Room_Type 5	0.00235
room_type_reserved_Room_Type 6	0.00152
repeated_guest	0.00097
no_of_previous_cancellations	0.00076
room_type_reserved_Room_Type 7	0.00069
const	0.00000

```
In [457...]  
importances = best_model.feature_importances_  
indices = np.argsort(importances)  
  
plt.figure(figsize=(12, 12))  
plt.title("Feature Importances")  
plt.barh(range(len(indices)), importances[indices], color="violet", align="center")  
plt.yticks(range(len(indices)), [feature_names[i] for i in indices])  
plt.xlabel("Relative Importance")  
plt.show()
```



## Model Performance Comparison and Conclusions

In [458]:

```
# training performance comparison

models_train_comp_df = pd.concat(
    [
        decision_tree_perf_train.T,
        decision_tree_tune_perf_train.T,
        decision_tree_postpruned_perf_train.T,
    ],
    axis=1,
)
models_train_comp_df.columns = [
    "Decision Tree sklearn",
    "Decision Tree (Pre-Pruning)",
    "Decision Tree (Post-Pruning)",
]
print("Training performance comparison:")
models_train_comp_df
```

Training performance comparison:

Out[458]:

	Decision Tree sklearn	Decision Tree (Pre-Pruning)	Decision Tree (Post-Pruning)
<b>Accuracy</b>	0.99287	0.99283	0.99228
<b>Recall</b>	0.98338	0.98326	0.98350
<b>Precision</b>	0.99492	0.99492	0.99300
<b>F1</b>	0.98912	0.98905	0.98823

In [459...]

```
# test performance comparison

decision_tree_perf_test = model_performance_classification_sklearn(
    model, X_test3, y_test
)
decision_tree_perf_test

models_train_comp_df = pd.concat(
    [
        decision_tree_perf_test.T,
        decision_tree_tune_perf_test.T,
        decision_tree_postpruned_perf_test.T,
    ],
    axis=1,
)
models_train_comp_df.columns = [
    "Decision Tree sklearn",
    "Decision Tree (Pre-Pruning)",
    "Decision Tree (Post-Pruning)",
]
print("Test set performance comparison:")
models_train_comp_df
```

Test set performance comparison:

Out[459]:

	Decision Tree sklearn	Decision Tree (Pre-Pruning)	Decision Tree (Post-Pruning)
<b>Accuracy</b>	0.85987	0.86171	0.86079
<b>Recall</b>	0.78819	0.79018	0.78876
<b>Precision</b>	0.78087	0.78416	0.78276
<b>F1</b>	0.78451	0.78716	0.78574

## Actionable Insights and Recommendations

- What profitable policies for cancellations and refunds can the hotel adopt?
- What other recommendations would you suggest to the hotel?

### Observations

- The more lead time is there in a reservation, the less probability of its cancellation.
- The more special requests that are part of a reservation, the less probability of its cancellation.

- If customers are booking way in advance, they should get some discounted price so that they do not get a better deal later and there's a less probability of reservation cancellation in future.

## Recommendations

- Hotel can start segregating rooms that are discounted but have a non-refundable deposit clause. This would ensure that these reservations are not cancelled. Even if they are cancelled, hotel has a deposit that would cover the losses.
- Do share all the special services that the hotel provides and ask customers at the time of the booking that which of these services would make their stay more comfortable. Adding special services has been known to lead to fewer cancellations.
- The months in which there's less reservations, special discounts could be rolled out to attract customers.
- Complimentary services like Laundry, WiFi, GYM etc could be provided that gives INN hotels an edge over their competitors.