

# gradient-boosting-1

November 15, 2024

```
[5]: !pip install folium
```

```
Defaulting to user installation because normal site-packages is not writeable
Collecting folium
  Downloading folium-0.18.0-py2.py3-none-any.whl (108 kB)
Requirement already satisfied: requests in c:\programdata\anaconda3\lib\site-packages (from folium) (2.27.1)
Collecting branca>=0.6.0
  Downloading branca-0.8.0-py3-none-any.whl (25 kB)
Collecting xyzservices
  Downloading xyzservices-2024.9.0-py3-none-any.whl (85 kB)
Requirement already satisfied: jinja2>=2.9 in c:\programdata\anaconda3\lib\site-packages (from folium) (2.11.3)
Requirement already satisfied: numpy in c:\programdata\anaconda3\lib\site-packages (from folium) (1.21.5)
Collecting jinja2>=2.9
  Downloading jinja2-3.1.4-py3-none-any.whl (133 kB)
Requirement already satisfied: MarkupSafe>=2.0 in c:\programdata\anaconda3\lib\site-packages (from jinja2>=2.9->folium) (2.0.1)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\programdata\anaconda3\lib\site-packages (from requests->folium) (1.26.9)
Requirement already satisfied: certifi>=2017.4.17 in c:\programdata\anaconda3\lib\site-packages (from requests->folium) (2021.10.8)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\programdata\anaconda3\lib\site-packages (from requests->folium) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\programdata\anaconda3\lib\site-packages (from requests->folium) (3.3)
Installing collected packages: jinja2, xyzservices, branca, folium
Successfully installed branca-0.8.0 folium-0.18.0 jinja2-3.1.4 xyzservices-2024.9.0

ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of the following dependency conflicts.
anaconda-project 0.10.2 requires ruamel-yaml, which is not installed.
jupyter-server 1.13.5 requires pywinpty<2; os_name == "nt", but you have pywinpty 2.0.2 which is incompatible.
```

```
[6]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import missingno as msno

import warnings
warnings.filterwarnings('ignore')

from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score, confusion_matrix, \
    classification_report

from sklearn.ensemble import GradientBoostingClassifier
from xgboost import XGBClassifier

import folium
from folium.plugins import HeatMap
import plotly.express as px

plt.style.use('fivethirtyeight')
%matplotlib inline
pd.set_option('display.max_columns', 32)
```

```
[60]: df = pd.read_csv('C:\\Users\\ANUSHA\\Downloads\\archive (7)\\hotel_booking.csv')
df.head()
```

```
[60]:
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	\
0	Resort Hotel	0	342	2015	July	
1	Resort Hotel	0	737	2015	July	
2	Resort Hotel	0	7	2015	July	
3	Resort Hotel	0	13	2015	July	
4	Resort Hotel	0	14	2015	July	

  

	arrival_date_week_number	arrival_date_day_of_month	\
0	27	1	
1	27	1	
2	27	1	
3	27	1	
4	27	1	

  

	stays_in_weekend_nights	stays_in_week_nights	adults	children	babies	\
0	0	0	2	0.0	0	
1	0	0	2	0.0	0	
2	0	1	1	0.0	0	
3	0	1	1	0.0	0	

4		0		2	2	0.0	0
---	--	---	--	---	---	-----	---

  

	meal	country	market_segment	distribution_channel	...	assigned_room_type	\
0	BB	PRT	Direct	Direct	...	C	
1	BB	PRT	Direct	Direct	...	C	
2	BB	GBR	Direct	Direct	...	C	
3	BB	GBR	Corporate	Corporate	...	A	
4	BB	GBR	Online TA	TA/T0	...	A	

  

	booking_changes	deposit_type	agent	company	days_in_waiting_list	\
0	3	No Deposit	NaN	NaN	0	
1	4	No Deposit	NaN	NaN	0	
2	0	No Deposit	NaN	NaN	0	
3	0	No Deposit	304.0	NaN	0	
4	0	No Deposit	240.0	NaN	0	

  

	customer_type	adr	required_car_parking_spaces	total_of_special_requests	\
0	Transient	0.0	0	0	
1	Transient	0.0	0	0	
2	Transient	75.0	0	0	
3	Transient	75.0	0	0	
4	Transient	98.0	0	1	

  

	reservation_status	reservation_status_date	name	\
0	Check-Out	2015-07-01	Ernest Barnes	
1	Check-Out	2015-07-01	Andrea Baker	
2	Check-Out	2015-07-02	Rebecca Parker	
3	Check-Out	2015-07-02	Laura Murray	
4	Check-Out	2015-07-03	Linda Hines	

  

	email	phone-number	credit_card
0	Ernest.Barnes31@outlook.com	669-792-1661	*****4322
1	Andrea_Baker94@aol.com	858-637-6955	*****9157
2	Rebecca_Parker@comcast.net	652-885-2745	*****3734
3	Laura_M@gmail.com	364-656-8427	*****5677
4	LHines@verizon.com	713-226-5883	*****5498

[5 rows x 36 columns]

```
[61]: df.describe()
```

```
[61]:
```

	is_canceled	lead_time	arrival_date_year	\
count	119390.000000	119390.000000	119390.000000	
mean	0.370416	104.011416	2016.156554	
std	0.482918	106.863097	0.707476	
min	0.000000	0.000000	2015.000000	
25%	0.000000	18.000000	2016.000000	

50%	0.000000	69.000000	2016.000000
75%	1.000000	160.000000	2017.000000
max	1.000000	737.000000	2017.000000

	arrival_date_week_number	arrival_date_day_of_month	\
count	119390.000000	119390.000000	
mean	27.165173	15.798241	
std	13.605138	8.780829	
min	1.000000	1.000000	
25%	16.000000	8.000000	
50%	28.000000	16.000000	
75%	38.000000	23.000000	
max	53.000000	31.000000	

	stays_in_weekend_nights	stays_in_week_nights	adults	\
count	119390.000000	119390.000000	119390.000000	
mean	0.927599	2.500302	1.856403	
std	0.998613	1.908286	0.579261	
min	0.000000	0.000000	0.000000	
25%	0.000000	1.000000	2.000000	
50%	1.000000	2.000000	2.000000	
75%	2.000000	3.000000	2.000000	
max	19.000000	50.000000	55.000000	

	children	babies	is_repeated_guest	\
count	119386.000000	119390.000000	119390.000000	
mean	0.103890	0.007949	0.031912	
std	0.398561	0.097436	0.175767	
min	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.000000	
50%	0.000000	0.000000	0.000000	
75%	0.000000	0.000000	0.000000	
max	10.000000	10.000000	1.000000	

	previous_cancellations	previous_bookings_not_canceled	\
count	119390.000000	119390.000000	
mean	0.087118	0.137097	
std	0.844336	1.497437	
min	0.000000	0.000000	
25%	0.000000	0.000000	
50%	0.000000	0.000000	
75%	0.000000	0.000000	
max	26.000000	72.000000	

	booking_changes	agent	company	days_in_waiting_list	\
count	119390.000000	103050.000000	6797.000000	119390.000000	
mean	0.221124	86.693382	189.266735	2.321149	

std	0.652306	110.774548	131.655015	17.594721
min	0.000000	1.000000	6.000000	0.000000
25%	0.000000	9.000000	62.000000	0.000000
50%	0.000000	14.000000	179.000000	0.000000
75%	0.000000	229.000000	270.000000	0.000000
max	21.000000	535.000000	543.000000	391.000000

  

	adr	required_car_parking_spaces	total_of_special_requests
count	119390.000000	119390.000000	119390.000000
mean	101.831122	0.062518	0.571363
std	50.535790	0.245291	0.792798
min	-6.380000	0.000000	0.000000
25%	69.290000	0.000000	0.000000
50%	94.575000	0.000000	0.000000
75%	126.000000	0.000000	1.000000
max	5400.000000	8.000000	5.000000

```
[62]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 36 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   hotel                                119390 non-null  object
1   is_canceled                          119390 non-null  int64
2   lead_time                           119390 non-null  int64
3   arrival_date_year                   119390 non-null  int64
4   arrival_date_month                  119390 non-null  object
5   arrival_date_week_number            119390 non-null  int64
6   arrival_date_day_of_month           119390 non-null  int64
7   stays_in_weekend_nights             119390 non-null  int64
8   stays_in_week_nights                119390 non-null  int64
9   adults                              119390 non-null  int64
10  children                             119386 non-null  float64
11  babies                              119390 non-null  int64
12  meal                                119390 non-null  object
13  country                             118902 non-null  object
14  market_segment                      119390 non-null  object
15  distribution_channel                 119390 non-null  object
16  is_repeated_guest                   119390 non-null  int64
17  previous_cancellations               119390 non-null  int64
18  previous_bookings_not_canceled       119390 non-null  int64
19  reserved_room_type                   119390 non-null  object
20  assigned_room_type                   119390 non-null  object
21  booking_changes                      119390 non-null  int64
22  deposit_type                         119390 non-null  object
```

```

23 agent                103050 non-null float64
24 company              6797 non-null float64
25 days_in_waiting_list 119390 non-null int64
26 customer_type        119390 non-null object
27 adr                  119390 non-null float64
28 required_car_parking_spaces 119390 non-null int64
29 total_of_special_requests 119390 non-null int64
30 reservation_status    119390 non-null object
31 reservation_status_date 119390 non-null object
32 name                 119390 non-null object
33 email                119390 non-null object
34 phone-number          119390 non-null object
35 credit_card           119390 non-null object
dtypes: float64(4), int64(16), object(16)
memory usage: 32.8+ MB

```

```

[63]: null = pd.DataFrame({'Null Values' : df.isna().sum(), 'Percentage Null Values' :
    ↪ (df.isna().sum()) / (df.shape[0]) * (100)})
null

```

```

[63]:
Null Values  Percentage Null Values
hotel                0             0.000000
is_canceled         0             0.000000
lead_time           0             0.000000
arrival_date_year    0             0.000000
arrival_date_month   0             0.000000
arrival_date_week_number 0             0.000000
arrival_date_day_of_month 0             0.000000
stays_in_weekend_nights 0             0.000000
stays_in_week_nights 0             0.000000
adults              0             0.000000
children            4             0.003350
babies              0             0.000000
meal                0             0.000000
country             488            0.408744
market_segment      0             0.000000
distribution_channel 0             0.000000
is_repeated_guest    0             0.000000
previous_cancellations 0             0.000000
previous_bookings_not_canceled 0             0.000000
reserved_room_type   0             0.000000
assigned_room_type    0             0.000000
booking_changes      0             0.000000
deposit_type         0             0.000000
agent              16340           13.686238
company            112593           94.306893
days_in_waiting_list 0             0.000000

```

customer_type	0	0.000000
adr	0	0.000000
required_car_parking_spaces	0	0.000000
total_of_special_requests	0	0.000000
reservation_status	0	0.000000
reservation_status_date	0	0.000000
name	0	0.000000
email	0	0.000000
phone-number	0	0.000000
credit_card	0	0.000000

```
[64]: df.fillna(0, inplace = True)
```

```
[65]: filter = (df.children == 0) & (df.adults == 0) & (df.babies == 0)
df[filter]
```

```
[65]:
```

	hotel	is_canceled	lead_time	arrival_date_year	\
2224	Resort Hotel	0	1	2015	
2409	Resort Hotel	0	0	2015	
3181	Resort Hotel	0	36	2015	
3684	Resort Hotel	0	165	2015	
3708	Resort Hotel	0	165	2015	
...	...	...	...	...	
115029	City Hotel	0	107	2017	
115091	City Hotel	0	1	2017	
116251	City Hotel	0	44	2017	
116534	City Hotel	0	2	2017	
117087	City Hotel	0	170	2017	

	arrival_date_month	arrival_date_week_number	\
2224	October	41	
2409	October	42	
3181	November	47	
3684	December	53	
3708	December	53	
...	...	...	
115029	June	26	
115091	June	26	
116251	July	28	
116534	July	28	
117087	July	30	

	arrival_date_day_of_month	stays_in_weekend_nights	\
2224	6	0	
2409	12	0	
3181	20	1	
3684	30	1	

3708	30	2
...	...	...
115029	27	0
115091	30	0
116251	15	1
116534	15	2
117087	27	0

	stays_in_week_nights	adults	children	babies	meal	country	\
2224	3	0	0.0	0	SC	PRT	
2409	0	0	0.0	0	SC	PRT	
3181	2	0	0.0	0	SC	ESP	
3684	4	0	0.0	0	SC	PRT	
3708	4	0	0.0	0	SC	PRT	
...	...	...	...	...	...	...	
115029	3	0	0.0	0	BB	CHE	
115091	1	0	0.0	0	SC	PRT	
116251	1	0	0.0	0	SC	SWE	
116534	5	0	0.0	0	SC	RUS	
117087	2	0	0.0	0	BB	BRA	

	market_segment	distribution_channel	...	assigned_room_type	\
2224	Corporate	Corporate	...	I	
2409	Corporate	Corporate	...	I	
3181	Groups	TA/TO	...	C	
3684	Groups	TA/TO	...	A	
3708	Groups	TA/TO	...	C	
...	...	...	...	...	
115029	Online TA	TA/TO	...	A	
115091	Complementary	Direct	...	K	
116251	Online TA	TA/TO	...	K	
116534	Online TA	TA/TO	...	K	
117087	Offline TA/TO	TA/TO	...	A	

	booking_changes	deposit_type	agent	company	days_in_waiting_list	\
2224	1	No Deposit	0.0	174.0	0	
2409	0	No Deposit	0.0	174.0	0	
3181	0	No Deposit	38.0	0.0	0	
3684	1	No Deposit	308.0	0.0	122	
3708	1	No Deposit	308.0	0.0	122	
...	...	...	...	...	...	
115029	1	No Deposit	7.0	0.0	0	
115091	0	No Deposit	0.0	0.0	0	
116251	2	No Deposit	425.0	0.0	0	
116534	1	No Deposit	9.0	0.0	0	
117087	0	No Deposit	52.0	0.0	0	



	customer_type	adr	required_car_parking_spaces	\
2224	Transient-Party	0.00	0	
2409	Transient	0.00	0	
3181	Transient-Party	0.00	0	
3684	Transient-Party	0.00	0	
3708	Transient-Party	0.00	0	
...	...	...	...	
115029	Transient	100.80	0	
115091	Transient	0.00	1	
116251	Transient	73.80	0	
116534	Transient-Party	22.86	0	
117087	Transient	0.00	0	

	total_of_special_requests	reservation_status	reservation_status_date	\
2224	0	Check-Out	2015-10-06	
2409	0	Check-Out	2015-10-12	
3181	0	Check-Out	2015-11-23	
3684	0	Check-Out	2016-01-04	
3708	0	Check-Out	2016-01-05	
...	...	...	...	
115029	0	Check-Out	2017-06-30	
115091	1	Check-Out	2017-07-01	
116251	0	Check-Out	2017-07-17	
116534	1	Check-Out	2017-07-22	
117087	0	Check-Out	2017-07-29	

	name	email	phone-number	\
2224	Colleen Hernandez	Colleen.H25@yandex.com	863-349-5397	
2409	Mark Martinez	Martinez.Mark@yahoo.com	969-893-1222	
3181	Carrie Carney	CarrieCarney@yandex.com	944-344-7905	
3684	Summer Stewart	Stewart_Summer84@aol.com	125-574-6642	
3708	Brian Buchanan	Brian.B@yahoo.com	946-176-5726	
...	...	...	...	
115029	Gregory Ashley	GregoryAshley51@outlook.com	784-717-9035	
115091	Christina McDonald	Mcdonald.Christina@protonmail.com	487-056-2442	
116251	Megan Johnson	MJohnson@protonmail.com	501-374-4690	
116534	Alexandra Riley	Alexandra.Riley55@mail.com	318-771-9720	
117087	Mark Thomas	Mark.T@zoho.com	663-347-2891	

	credit_card
2224	*****8042
2409	*****9670
3181	*****8964
3684	*****1069
3708	*****9836
...	...
115029	*****9561

```

115091 *****8308
116251 *****3910
116534 *****6778
117087 *****7175

```

[180 rows x 36 columns]

```

[66]: df = df[~filter]
      df

```

```

[66]:          hotel  is_canceled  lead_time  arrival_date_year  \
0      Resort Hotel           0        342             2015
1      Resort Hotel           0        737             2015
2      Resort Hotel           0         7             2015
3      Resort Hotel           0        13             2015
4      Resort Hotel           0        14             2015
...
119385  City Hotel           0         23             2017
119386  City Hotel           0        102             2017
119387  City Hotel           0         34             2017
119388  City Hotel           0        109             2017
119389  City Hotel           0        205             2017

```

```

          arrival_date_month  arrival_date_week_number  \
0                July                27
1                July                27
2                July                27
3                July                27
4                July                27
...
119385            August                35
119386            August                35
119387            August                35
119388            August                35
119389            August                35

```

```

          arrival_date_day_of_month  stays_in_weekend_nights  \
0                1                0
1                1                0
2                1                0
3                1                0
4                1                0
...
119385            30                2
119386            31                2
119387            31                2
119388            31                2

```

119389

29

2

	stays_in_week_nights	adults	children	babies	meal	country	\
0	0	2	0.0	0	BB	PRT	
1	0	2	0.0	0	BB	PRT	
2	1	1	0.0	0	BB	GBR	
3	1	1	0.0	0	BB	GBR	
4	2	2	0.0	0	BB	GBR	
...	...	...	...	...	...	...	
119385	5	2	0.0	0	BB	BEL	
119386	5	3	0.0	0	BB	FRA	
119387	5	2	0.0	0	BB	DEU	
119388	5	2	0.0	0	BB	GBR	
119389	7	2	0.0	0	HB	DEU	

	market_segment	distribution_channel	...	assigned_room_type	\
0	Direct	Direct	...	C	
1	Direct	Direct	...	C	
2	Direct	Direct	...	C	
3	Corporate	Corporate	...	A	
4	Online TA	TA/TO	...	A	
...	...	...	...	...	
119385	Offline TA/TO	TA/TO	...	A	
119386	Online TA	TA/TO	...	E	
119387	Online TA	TA/TO	...	D	
119388	Online TA	TA/TO	...	A	
119389	Online TA	TA/TO	...	A	

	booking_changes	deposit_type	agent	company	days_in_waiting_list	\
0	3	No Deposit	0.0	0.0	0	
1	4	No Deposit	0.0	0.0	0	
2	0	No Deposit	0.0	0.0	0	
3	0	No Deposit	304.0	0.0	0	
4	0	No Deposit	240.0	0.0	0	
...	...	...	...	...	...	
119385	0	No Deposit	394.0	0.0	0	
119386	0	No Deposit	9.0	0.0	0	
119387	0	No Deposit	9.0	0.0	0	
119388	0	No Deposit	89.0	0.0	0	
119389	0	No Deposit	9.0	0.0	0	

	customer_type	adr	required_car_parking_spaces	\
0	Transient	0.00	0	
1	Transient	0.00	0	
2	Transient	75.00	0	
3	Transient	75.00	0	
4	Transient	98.00	0	

...	...	...	...
119385	Transient	96.14	0
119386	Transient	225.43	0
119387	Transient	157.71	0
119388	Transient	104.40	0
119389	Transient	151.20	0

  

	total_of_special_requests	reservation_status	reservation_status_date \
0	0	Check-Out	2015-07-01
1	0	Check-Out	2015-07-01
2	0	Check-Out	2015-07-02
3	0	Check-Out	2015-07-02
4	1	Check-Out	2015-07-03
...	...	...	...
119385	0	Check-Out	2017-09-06
119386	2	Check-Out	2017-09-07
119387	4	Check-Out	2017-09-07
119388	0	Check-Out	2017-09-07
119389	2	Check-Out	2017-09-07

  

	name	email	phone-number \
0	Ernest Barnes	Ernest.Barnes31@outlook.com	669-792-1661
1	Andrea Baker	Andrea_Baker94@aol.com	858-637-6955
2	Rebecca Parker	Rebecca_Parker@comcast.net	652-885-2745
3	Laura Murray	Laura_M@gmail.com	364-656-8427
4	Linda Hines	LHines@verizon.com	713-226-5883
...	...	...	...
119385	Claudia Johnson	Claudia.J@yahoo.com	403-092-5582
119386	Wesley Aguilar	WAguilar@xfinity.com	238-763-0612
119387	Mary Morales	Mary_Morales@hotmail.com	395-518-4100
119388	Caroline Conley MD	MD_Caroline@comcast.net	531-528-1017
119389	Ariana Michael	Ariana_M@xfinity.com	422-804-6403

  

	credit_card
0	*****4322
1	*****9157
2	*****3734
3	*****5677
4	*****5498
...	...
119385	*****8647
119386	*****4333
119387	*****1821
119388	*****7860
119389	*****4482

[119210 rows x 36 columns]

```
[67]: country_wise_guests = df[df['is_canceled'] == 0]['country'].value_counts().
      ↪reset_index()
      country_wise_guests.columns = ['country', 'No of guests']
      country_wise_guests
```

```
[67]:      country  No of guests
0      PRT      20977
1      GBR      9668
2      FRA      8468
3      ESP      6383
4      DEU      6067
..      ...      ...
161     BHR          1
162     DJI          1
163     MLI          1
164     NPL          1
165     FRO          1
```

[166 rows x 2 columns]

```
[68]: data = df[df['is_canceled'] == 0]
      px.box(data_frame = data, x = 'reserved_room_type', y = 'adr', color = 'hotel',
      ↪template = 'plotly_dark')
```

```
[69]: data_resort = df[(df['hotel'] == 'Resort Hotel') & (df['is_canceled'] == 0)]
      data_city = df[(df['hotel'] == 'City Hotel') & (df['is_canceled'] == 0)]
```

```
[70]: resort_hotel = data_resort.groupby(['arrival_date_month'])['adr'].mean().
      ↪reset_index()
      resort_hotel
```

```
[70]:      arrival_date_month      adr
0      April      75.867816
1      August     181.205892
2      December     68.410104
3      February     54.147478
4      January     48.761125
5      July      150.122528
6      June      107.974850
7      March      57.056838
8      May       76.657558
9      November    48.706289
10     October    61.775449
11     September    96.416860
```

```
[71]: city_hotel=data_city.groupby(['arrival_date_month'])['adr'].mean().reset_index()
      city_hotel
```

```
[71]: arrival_date_month    adr
0      April    111.962267
1      August    118.674598
2      December    88.401855
3      February    86.520062
4      January    82.330983
5      July    115.818019
6      June    117.874360
7      March    90.658533
8      May    120.669827
9      November    86.946592
10     October    102.004672
11     September    112.776582
```

```
[72]: final_hotel = resort_hotel.merge(city_hotel, on = 'arrival_date_month')
final_hotel.columns = ['month', 'price_for_resort', 'price_for_city_hotel']
final_hotel
```

```
[72]:      month  price_for_resort  price_for_city_hotel
0     April      75.867816      111.962267
1     August     181.205892     118.674598
2  December      68.410104      88.401855
3  February      54.147478      86.520062
4   January      48.761125      82.330983
5      July     150.122528     115.818019
6      June     107.974850     117.874360
7      March      57.056838      90.658533
8      May      76.657558     120.669827
9  November      48.706289      86.946592
10   October      61.775449     102.004672
11  September      96.416860     112.776582
```

```
[73]: !pip install sort-dataframeby-monthorweek
```

```
!pip install sorted-months-weekdays
```

Defaulting to user installation because normal site-packages is not writeable  
Requirement already satisfied: sort-dataframeby-monthorweek in  
c:\users\anusha\appdata\roaming\python\python39\site-packages (0.4)  
Defaulting to user installation because normal site-packages is not writeable  
Requirement already satisfied: sorted-months-weekdays in  
c:\users\anusha\appdata\roaming\python\python39\site-packages (0.2)

```
[74]: import sort_dataframeby_monthorweek as sd

def sort_month(df, column_name):
    return sd.Sort_Dataframeby_Month(df, column_name)
```

```
[75]: final_prices = sort_month(final_hotel, 'month')
final_prices
```

```
[75]:
```

	month	price_for_resort	price_for_city_hotel
0	January	48.761125	82.330983
1	February	54.147478	86.520062
2	March	57.056838	90.658533
3	April	75.867816	111.962267
4	May	76.657558	120.669827
5	June	107.974850	117.874360
6	July	150.122528	115.818019
7	August	181.205892	118.674598
8	September	96.416860	112.776582
9	October	61.775449	102.004672
10	November	48.706289	86.946592
11	December	68.410104	88.401855

```
[76]: plt.figure(figsize = (17, 8))
px.line(final_prices, x = 'month', y =_
↳['price_for_resort','price_for_city_hotel'],
title = 'Room price per night over the Months', template =_
↳'plotly_dark')
```

<Figure size 1224x576 with 0 Axes>

```
[77]: resort_guests = data_resort['arrival_date_month'].value_counts().reset_index()
resort_guests.columns=['month','no of guests']
resort_guests
```

```
[77]:
```

	month	no of guests
0	August	3257
1	July	3137
2	October	2575
3	March	2571
4	April	2550
5	May	2535
6	February	2308
7	September	2102
8	June	2037
9	December	2014
10	November	1975
11	January	1866

```
[78]: city_guests = data_city['arrival_date_month'].value_counts().reset_index()
city_guests.columns=['month','no of guests']
city_guests
```

```
[78]:      month  no of guests
      0      August      5367
      1        July      4770
      2         May      4568
      3         June      4358
      4      October      4326
      5    September      4283
      6        March      4049
      7         April      4010
      8    February      3051
      9     November      2676
     10    December      2377
     11     January      2249
```

```
[79]: final_guests = resort_guests.merge(city_guests,on='month')
      final_guests.columns=['month','no of guests in resort','no of guest in city_
      ↪hotel']
      final_guests
```

```
[79]:      month  no of guests in resort  no of guest in city hotel
      0      August                3257                5367
      1         July                3137                4770
      2      October                2575                4326
      3         March                2571                4049
      4         April                2550                4010
      5          May                2535                4568
      6    February                2308                3051
      7    September                2102                4283
      8          June                2037                4358
      9    December                2014                2377
     10   November                1975                2676
     11   January                 1866                2249
```

```
[80]: final_guests = sort_month(final_guests,'month')
      final_guests
```

```
[80]:      month  no of guests in resort  no of guest in city hotel
      0     January                 1866                2249
      1    February                 2308                3051
      2      March                 2571                4049
      3      April                 2550                4010
      4          May                 2535                4568
      5          June                 2037                4358
      6          July                 3137                4770
      7      August                 3257                5367
      8    September                 2102                4283
      9     October                 2575                4326
```



10	November	1975	2676
11	December	2014	2377

```
[81]: px.line(final_guests, x = 'month', y = ['no of guests in resort', 'no of guest_
      ↪in city hotel'],
      title='Total no of guests per Months', template = 'plotly_dark')
```

```
[82]: filter = df['is_canceled'] == 0
data = df[filter]
data.head()
```

```
[82]:
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	\
0	Resort Hotel	0	342	2015	July	
1	Resort Hotel	0	737	2015	July	
2	Resort Hotel	0	7	2015	July	
3	Resort Hotel	0	13	2015	July	
4	Resort Hotel	0	14	2015	July	

	arrival_date_week_number	arrival_date_day_of_month	\
0	27	1	
1	27	1	
2	27	1	
3	27	1	
4	27	1	

	stays_in_weekend_nights	stays_in_week_nights	adults	children	babies	\
0	0	0	2	0.0	0	
1	0	0	2	0.0	0	
2	0	1	1	0.0	0	
3	0	1	1	0.0	0	
4	0	2	2	0.0	0	

	meal	country	market_segment	distribution_channel	...	assigned_room_type	\
0	BB	PRT	Direct	Direct	...	C	
1	BB	PRT	Direct	Direct	...	C	
2	BB	GBR	Direct	Direct	...	C	
3	BB	GBR	Corporate	Corporate	...	A	
4	BB	GBR	Online TA	TA/TO	...	A	

	booking_changes	deposit_type	agent	company	days_in_waiting_list	\
0	3	No Deposit	0.0	0.0	0	
1	4	No Deposit	0.0	0.0	0	
2	0	No Deposit	0.0	0.0	0	
3	0	No Deposit	304.0	0.0	0	
4	0	No Deposit	240.0	0.0	0	

	customer_type	adr	required_car_parking_spaces	total_of_special_requests	\
--	---------------	-----	-----------------------------	---------------------------	---

0	Transient	0.0	0	0
1	Transient	0.0	0	0
2	Transient	75.0	0	0
3	Transient	75.0	0	0
4	Transient	98.0	0	1

	reservation_status	reservation_status_date	name \
0	Check-Out	2015-07-01	Ernest Barnes
1	Check-Out	2015-07-01	Andrea Baker
2	Check-Out	2015-07-02	Rebecca Parker
3	Check-Out	2015-07-02	Laura Murray
4	Check-Out	2015-07-03	Linda Hines

	email	phone-number	credit_card
0	Ernest.Barnes31@outlook.com	669-792-1661	*****4322
1	Andrea_Baker94@aol.com	858-637-6955	*****9157
2	Rebecca_Parker@comcast.net	652-885-2745	*****3734
3	Laura_M@gmail.com	364-656-8427	*****5677
4	LHines@verizon.com	713-226-5883	*****5498

[5 rows x 36 columns]

```
[83]: data['total_nights'] = data['stays_in_weekend_nights'] +
      ↪data['stays_in_week_nights']
data.head()
```

```
[83]:
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month \
0	Resort Hotel	0	342	2015	July
1	Resort Hotel	0	737	2015	July
2	Resort Hotel	0	7	2015	July
3	Resort Hotel	0	13	2015	July
4	Resort Hotel	0	14	2015	July

	arrival_date_week_number	arrival_date_day_of_month \
0	27	1
1	27	1
2	27	1
3	27	1
4	27	1

	stays_in_weekend_nights	stays_in_week_nights	adults	children	babies \
0	0	0	2	0.0	0
1	0	0	2	0.0	0
2	0	1	1	0.0	0
3	0	1	1	0.0	0
4	0	2	2	0.0	0

	meal	country	market_segment	distribution_channel	...	booking_changes	\
0	BB	PRT	Direct	Direct	...	3	
1	BB	PRT	Direct	Direct	...	4	
2	BB	GBR	Direct	Direct	...	0	
3	BB	GBR	Corporate	Corporate	...	0	
4	BB	GBR	Online TA	TA/TO	...	0	

  

	deposit_type	agent	company	days_in_waiting_list	customer_type	adr	\
0	No Deposit	0.0	0.0	0	Transient	0.0	
1	No Deposit	0.0	0.0	0	Transient	0.0	
2	No Deposit	0.0	0.0	0	Transient	75.0	
3	No Deposit	304.0	0.0	0	Transient	75.0	
4	No Deposit	240.0	0.0	0	Transient	98.0	

  

	required_car_parking_spaces	total_of_special_requests	reservation_status	\
0	0	0	Check-Out	
1	0	0	Check-Out	
2	0	0	Check-Out	
3	0	0	Check-Out	
4	0	1	Check-Out	

  

	reservation_status_date	name	email	\
0	2015-07-01	Ernest Barnes	Ernest.Barnes31@outlook.com	
1	2015-07-01	Andrea Baker	Andrea_Baker94@aol.com	
2	2015-07-02	Rebecca Parker	Rebecca_Parker@comcast.net	
3	2015-07-02	Laura Murray	Laura_M@gmail.com	
4	2015-07-03	Linda Hines	LHines@verizon.com	

  

	phone-number	credit_card	total_nights
0	669-792-1661	*****4322	0
1	858-637-6955	*****9157	0
2	652-885-2745	*****3734	1
3	364-656-8427	*****5677	1
4	713-226-5883	*****5498	2

[5 rows x 37 columns]

```
[84]: stay = data.groupby(['total_nights', 'hotel']).agg('count').reset_index()
      stay = stay.iloc[:, :3]
      stay = stay.rename(columns={'is_canceled': 'Number of stays'})
      stay
```

```
[84]:   total_nights   hotel  Number of stays
0         0   City Hotel             251
1         0  Resort Hotel             371
2         1   City Hotel            9155
3         1  Resort Hotel            6579
```

4	2	City Hotel	10983
..	...	...	...
57	46	Resort Hotel	1
58	48	City Hotel	1
59	56	Resort Hotel	1
60	60	Resort Hotel	1
61	69	Resort Hotel	1

[62 rows x 3 columns]

```
[85]: px.bar(data_frame = stay, x = 'total_nights', y = 'Number of stays', color = 'hotel',
           barmode = 'group',
           template = 'plotly_dark')
```

```
[86]: correlation = df.corr()['is_canceled'].abs().sort_values(ascending = False)
correlation
```

```
[86]: is_canceled          1.000000
lead_time              0.292876
total_of_special_requests 0.234877
required_car_parking_spaces 0.195701
booking_changes        0.144832
previous_cancellations  0.110139
is_repeated_guest       0.083745
company                0.083594
adults                 0.058182
previous_bookings_not_canceled 0.057365
days_in_waiting_list   0.054301
agent                  0.046770
adr                    0.046492
babies                 0.032569
stays_in_week_nights    0.025542
arrival_date_year       0.016622
arrival_date_week_number 0.008315
arrival_date_day_of_month 0.005948
children                0.004851
stays_in_weekend_nights 0.001323
Name: is_canceled, dtype: float64
```

```
[87]: useless_col = ['days_in_waiting_list', 'arrival_date_year',
                    'arrival_date_year', 'assigned_room_type', 'booking_changes',
                    'reservation_status', 'country', 'days_in_waiting_list']

df.drop(useless_col, axis = 1, inplace = True)
```

```
[88]: df.head()
```

```

[88]:      hotel  is_canceled  lead_time arrival_date_month  \
0  Resort Hotel          0        342             July
1  Resort Hotel          0        737             July
2  Resort Hotel          0         7             July
3  Resort Hotel          0        13             July
4  Resort Hotel          0        14             July

      arrival_date_week_number  arrival_date_day_of_month  \
0                             27                         1
1                             27                         1
2                             27                         1
3                             27                         1
4                             27                         1

      stays_in_weekend_nights  stays_in_week_nights  adults  children  babies  \
0                             0                     0       2        0.0       0
1                             0                     0       2        0.0       0
2                             0                     1       1        0.0       0
3                             0                     1       1        0.0       0
4                             0                     2       2        0.0       0

      meal market_segment distribution_channel  is_repeated_guest  \
0  BB          Direct          Direct              0
1  BB          Direct          Direct              0
2  BB          Direct          Direct              0
3  BB    Corporate    Corporate              0
4  BB    Online TA    TA/TO              0

      previous_cancellations  previous_bookings_not_canceled  reserved_room_type  \
0                             0                             0                  C
1                             0                             0                  C
2                             0                             0                  A
3                             0                             0                  A
4                             0                             0                  A

      deposit_type  agent  company customer_type  adr  \
0  No Deposit    0.0    0.0    Transient    0.0
1  No Deposit    0.0    0.0    Transient    0.0
2  No Deposit    0.0    0.0    Transient   75.0
3  No Deposit   304.0    0.0    Transient   75.0
4  No Deposit   240.0    0.0    Transient   98.0

      required_car_parking_spaces  total_of_special_requests  \
0                                 0                          0
1                                 0                          0
2                                 0                          0
3                                 0                          0

```

4

0

1

	reservation_status_date	name	email \
0	2015-07-01	Ernest Barnes	Ernest.Barnes31@outlook.com
1	2015-07-01	Andrea Baker	Andrea_Baker94@aol.com
2	2015-07-02	Rebecca Parker	Rebecca_Parker@comcast.net
3	2015-07-02	Laura Murray	Laura_M@gmail.com
4	2015-07-03	Linda Hines	LHines@verizon.com

	phone-number	credit_card
0	669-792-1661	*****4322
1	858-637-6955	*****9157
2	652-885-2745	*****3734
3	364-656-8427	*****5677
4	713-226-5883	*****5498

```
[89]: cat_cols = [col for col in df.columns if df[col].dtype == 'O']
cat_cols
```

```
[89]: ['hotel',
'arrival_date_month',
'meal',
'market_segment',
'distribution_channel',
'reserved_room_type',
'deposit_type',
'customer_type',
'reservation_status_date',
'name',
'email',
'phone-number',
'credit_card']
```

```
[90]: cat_df = df[cat_cols]
cat_df.head()
```

```
[90]:      hotel arrival_date_month meal market_segment distribution_channel \
0  Resort Hotel          July    BB          Direct          Direct
1  Resort Hotel          July    BB          Direct          Direct
2  Resort Hotel          July    BB          Direct          Direct
3  Resort Hotel          July    BB    Corporate    Corporate
4  Resort Hotel          July    BB    Online TA    TA/TO

reserved_room_type deposit_type customer_type reservation_status_date \
0          C    No Deposit    Transient    2015-07-01
1          C    No Deposit    Transient    2015-07-01
2          A    No Deposit    Transient    2015-07-02
```

3	A	No Deposit	Transient	2015-07-02
4	A	No Deposit	Transient	2015-07-03

	name	email	phone-number	credit_card
0	Ernest Barnes	Ernest.Barnes31@outlook.com	669-792-1661	*****4322
1	Andrea Baker	Andrea_Baker94@aol.com	858-637-6955	*****9157
2	Rebecca Parker	Rebecca_Parker@comcast.net	652-885-2745	*****3734
3	Laura Murray	Laura_M@gmail.com	364-656-8427	*****5677
4	Linda Hines	LHines@verizon.com	713-226-5883	*****5498

```
[91]: cat_df['reservation_status_date'] = pd.
      ↪to_datetime(cat_df['reservation_status_date'])

cat_df['year'] = cat_df['reservation_status_date'].dt.year
cat_df['month'] = cat_df['reservation_status_date'].dt.month
cat_df['day'] = cat_df['reservation_status_date'].dt.day
```

```
[92]: cat_df.head()
```

```
[92]:      hotel arrival_date_month meal market_segment distribution_channel \
0  Resort Hotel          July    BB          Direct          Direct
1  Resort Hotel          July    BB          Direct          Direct
2  Resort Hotel          July    BB          Direct          Direct
3  Resort Hotel          July    BB    Corporate    Corporate
4  Resort Hotel          July    BB    Online TA      TA/TO
```

	reserved_room_type	deposit_type	customer_type	reservation_status_date
0	C	No Deposit	Transient	2015-07-01
1	C	No Deposit	Transient	2015-07-01
2	A	No Deposit	Transient	2015-07-02
3	A	No Deposit	Transient	2015-07-02
4	A	No Deposit	Transient	2015-07-03

	name	email	phone-number
0	Ernest Barnes	Ernest.Barnes31@outlook.com	669-792-1661
1	Andrea Baker	Andrea_Baker94@aol.com	858-637-6955
2	Rebecca Parker	Rebecca_Parker@comcast.net	652-885-2745
3	Laura Murray	Laura_M@gmail.com	364-656-8427
4	Linda Hines	LHines@verizon.com	713-226-5883

	credit_card	year	month	day
0	*****4322	2015	7	1
1	*****9157	2015	7	1
2	*****3734	2015	7	2
3	*****5677	2015	7	2
4	*****5498	2015	7	3

```
[93]: cat_df.  
      ↪drop(['reservation_status_date', 'arrival_date_month', 'name', 'email', 'credit_card', 'phone-num  
      ↪, axis = 1, inplace = True)  
      cat_df.head()
```

```
[93]:          hotel meal market_segment distribution_channel reserved_room_type \  
0  Resort Hotel   BB          Direct          Direct          C  
1  Resort Hotel   BB          Direct          Direct          C  
2  Resort Hotel   BB          Direct          Direct          A  
3  Resort Hotel   BB      Corporate      Corporate          A  
4  Resort Hotel   BB      Online TA      TA/TO          A  
  
   deposit_type customer_type  year  month  day  
0   No Deposit      Transient  2015     7    1  
1   No Deposit      Transient  2015     7    1  
2   No Deposit      Transient  2015     7    2  
3   No Deposit      Transient  2015     7    2  
4   No Deposit      Transient  2015     7    3
```

```
[94]: for col in cat_df.columns:  
      print(f"{col}: \n{cat_df[col].unique()}\n")
```

hotel:

```
['Resort Hotel' 'City Hotel']
```

meal:

```
['BB' 'FB' 'HB' 'SC' 'Undefined']
```

market\_segment:

```
['Direct' 'Corporate' 'Online TA' 'Offline TA/TO' 'Complementary' 'Groups'  
'Undefined' 'Aviation']
```

distribution\_channel:

```
['Direct' 'Corporate' 'TA/TO' 'Undefined' 'GDS']
```

reserved\_room\_type:

```
['C' 'A' 'D' 'E' 'G' 'F' 'H' 'L' 'B']
```

deposit\_type:

```
['No Deposit' 'Refundable' 'Non Refund']
```

customer\_type:

```
['Transient' 'Contract' 'Transient-Party' 'Group']
```

year:

```
[2015 2014 2016 2017]
```



month:

[ 7 5 4 6 3 8 9 1 11 10 12 2]

day:

[ 1 2 3 6 22 23 5 7 8 11 15 16 29 19 18 9 13 4 12 26 17 10 20 14  
30 28 25 21 27 24 31]

```
[95]: cat_df['hotel'] = cat_df['hotel'].map({'Resort Hotel' : 0, 'City Hotel' : 1})

cat_df['meal'] = cat_df['meal'].map({'BB' : 0, 'FB': 1, 'HB': 2, 'SC': 3,
↳ 'Undefined': 4})

cat_df['market_segment'] = cat_df['market_segment'].map({'Direct': 0,
↳ 'Corporate': 1, 'Online TA': 2, 'Offline TA/TO': 3,
                                                                    'Complementary': 4,
↳ 'Groups': 5, 'Undefined': 6, 'Aviation': 7})

cat_df['distribution_channel'] = cat_df['distribution_channel'].map({'Direct':
↳ 0, 'Corporate': 1, 'TA/TO': 2, 'Undefined': 3,
                                                                    'GDS':
↳ 4})

cat_df['reserved_room_type'] = cat_df['reserved_room_type'].map({'C': 0, 'A':
↳ 1, 'D': 2, 'E': 3, 'G': 4, 'F': 5, 'H': 6,
                                                                    'L': 7, 'B':
↳ 8})

cat_df['deposit_type'] = cat_df['deposit_type'].map({'No Deposit': 0,
↳ 'Refundable': 1, 'Non Refund': 3})

cat_df['customer_type'] = cat_df['customer_type'].map({'Transient': 0,
↳ 'Contract': 1, 'Transient-Party': 2, 'Group': 3})

cat_df['year'] = cat_df['year'].map({2015: 0, 2014: 1, 2016: 2, 2017: 3})
```

```
[96]: cat_df.head()
```

```
[96]:   hotel  meal  market_segment  distribution_channel  reserved_room_type  \
0      0     0                0                0                0
1      0     0                0                0                0
2      0     0                0                0                1
3      0     0                1                1                1
4      0     0                2                2                1

   deposit_type  customer_type  year  month  day
0              0              0    0     7    1
1              0              0    0     7    1
```

2	0	0	0	7	2
3	0	0	0	7	2
4	0	0	0	7	3

```
[97]: num_df = df.drop(columns = cat_cols, axis = 1)
num_df.drop('is_canceled', axis = 1, inplace = True)
num_df
```

```
[97]:      lead_time  arrival_date_week_number  arrival_date_day_of_month \
0          342                        27                        1
1          737                        27                        1
2           7                        27                        1
3          13                        27                        1
4          14                        27                        1
...
119385      23                        35                       30
119386     102                        35                       31
119387      34                        35                       31
119388     109                        35                       31
119389     205                        35                       29
```

	stays_in_weekend_nights	stays_in_week_nights	adults	children	\
0	0	0	2	0.0	
1	0	0	2	0.0	
2	0	1	1	0.0	
3	0	1	1	0.0	
4	0	2	2	0.0	
...	...	...	...	...	
119385	2	5	2	0.0	
119386	2	5	3	0.0	
119387	2	5	2	0.0	
119388	2	5	2	0.0	
119389	2	7	2	0.0	

	babies	is_repeated_guest	previous_cancellations	\
0	0	0	0	
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	
...	...	...	...	
119385	0	0	0	
119386	0	0	0	
119387	0	0	0	
119388	0	0	0	
119389	0	0	0	

	previous_bookings_not_canceled	agent	company	adr	\
0	0	0.0	0.0	0.00	
1	0	0.0	0.0	0.00	
2	0	0.0	0.0	75.00	
3	0	304.0	0.0	75.00	
4	0	240.0	0.0	98.00	
...	...	...	...	...	
119385	0	394.0	0.0	96.14	
119386	0	9.0	0.0	225.43	
119387	0	9.0	0.0	157.71	
119388	0	89.0	0.0	104.40	
119389	0	9.0	0.0	151.20	

	required_car_parking_spaces	total_of_special_requests
0	0	0
1	0	0
2	0	0
3	0	0
4	0	1
...	...	...
119385	0	0
119386	0	2
119387	0	4
119388	0	0
119389	0	2

[119210 rows x 16 columns]

```
[98]: num_df.var()
```

```
[98]: lead_time          11422.361808
      arrival_date_week_number    184.990111
      arrival_date_day_of_month    77.107192
      stays_in_weekend_nights    0.990258
      stays_in_week_nights    3.599010
      adults    0.330838
      children    0.159070
      babies    0.009508
      is_repeated_guest    0.030507
      previous_cancellations    0.713887
      previous_bookings_not_canceled    2.244415
      agent    11485.169679
      company    2897.684308
      adr    2543.589039
      required_car_parking_spaces    0.060201
      total_of_special_requests    0.628652
      dtype: float64
```

```
[113]: num_df['lead_time'] = np.log(num_df['lead_time'] + 1)
num_df['arrival_date_week_number'] = np.log(num_df['arrival_date_week_number'] + 1)
num_df['arrival_date_day_of_month'] = np.
    log(num_df['arrival_date_day_of_month'] + 1)
num_df['agent'] = np.log(num_df['agent'] + 1)
num_df['company'] = np.log(num_df['company'] + 1)
num_df['adr'] = np.log(num_df['adr'] + 1)
```

```
[114]: num_df.var()
```

```
[114]: lead_time          0.228853
arrival_date_week_number  0.035169
arrival_date_day_of_month 0.052872
stays_in_weekend_nights  0.990258
stays_in_week_nights     3.599010
adults                   0.330838
children                 0.159070
babies                   0.009508
is_repeated_guest        0.030507
previous_cancellations    0.713887
previous_bookings_not_canceled 2.244415
agent                    0.363545
company                  0.168993
adr                      0.051156
required_car_parking_spaces 0.060201
total_of_special_requests 0.628652
dtype: float64
```

```
[115]: num_df['adr'] = num_df['adr'].fillna(value = num_df['adr'].mean())
num_df.head()
```

```
[115]:   lead_time  arrival_date_week_number  arrival_date_day_of_month \
0    1.922456                1.466077                0.526589
1    2.028667                1.466077                0.526589
2    1.124748                1.466077                0.526589
3    1.291725                1.466077                0.526589
4    1.310506                1.466077                0.526589

   stays_in_weekend_nights  stays_in_week_nights  adults  children  babies \
0                        0                      0       2         0.0       0
1                        0                      0       2         0.0       0
2                        0                      1       1         0.0       0
3                        0                      1       1         0.0       0
4                        0                      2       2         0.0       0

   is_repeated_guest  previous_cancellations  previous_bookings_not_canceled \
```

0	0	0	0
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0

	agent	company	adr	required_car_parking_spaces	\
0	0.000000	0.0	0.000000	0	
1	0.000000	0.0	0.000000	0	
2	0.000000	0.0	1.673489	0	
3	1.905135	0.0	1.673489	0	
4	1.869461	0.0	1.721895	0	

	total_of_special_requests
0	0
1	0
2	0
3	0
4	1

```
[116]: X = pd.concat([cat_df, num_df], axis = 1)
y = df['is_canceled']
```

```
[117]: X.shape, y.shape
```

```
[117]: ((119210, 26), (119210,))
```

```
[118]: from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.30)
```

```
[119]: X_train.head()
```

```
[119]:
```

	hotel	meal	market_segment	distribution_channel	reserved_room_type	\
88129	1	0	0	0	1	
4958	0	2	5	0	2	
106714	1	0	2	2	2	
17040	0	0	2	2	3	
20395	0	0	2	2	1	

	deposit_type	customer_type	year	month	day	lead_time	\
88129	0	0	2	5	1	1.457646	
4958	0	0	2	4	17	1.890156	
106714	0	0	3	3	2	1.640720	
17040	0	0	0	9	23	1.710658	
20395	0	0	2	1	27	0.959135	

	arrival_date_week_number	arrival_date_day_of_month	\
88129	1.372307	1.474144	
4958	1.343703	1.223156	
106714	1.194706	1.457646	
17040	1.539779	1.372307	
20395	0.959135	1.419607	

	stays_in_weekend_nights	stays_in_week_nights	adults	children	\
88129	0	3	2	0.0	
4958	2	5	2	0.0	
106714	2	2	2	0.0	
17040	2	3	2	0.0	
20395	2	3	1	0.0	

	babies	is_repeated_guest	previous_cancellations	\
88129	0	0	0	
4958	0	0	0	
106714	0	0	0	
17040	0	0	0	
20395	0	0	0	

	previous_bookings_not_canceled	agent	company	adr	\
88129	0	0.000000	0.0	1.742137	
4958	0	1.889057	0.0	1.672125	
106714	0	1.194706	0.0	1.723743	
17040	0	1.869461	0.0	1.723689	
20395	0	1.869461	0.0	1.545194	

	required_car_parking_spaces	total_of_special_requests
88129	0	0
4958	0	0
106714	0	0
17040	0	1
20395	0	1

```
[120]: X_test.head()
```

```
[120]:
```

	hotel	meal	market_segment	distribution_channel	reserved_room_type	\
107224	1	3	2	2	1	
86430	1	0	5	2	1	
65955	1	0	1	1	1	
28058	0	0	3	2	2	
18103	0	0	1	1	1	

  

	deposit_type	customer_type	year	month	day	lead_time	\
107224	0	0	3	3	10	1.539779	
86430	0	2	2	4	3	1.595709	

65955	0	0	3	4	6	1.026672
28058	0	0	2	9	19	1.248441
18103	0	2	0	11	6	0.000000

	arrival_date_week_number	arrival_date_day_of_month	\
107224	1.223156	1.124748	
86430	1.310506	1.496434	
65955	1.327761	1.248441	
28058	1.545194	1.372307	
18103	1.574565	0.959135	

	stays_in_weekend_nights	stays_in_week_nights	adults	children	\
107224	0	3	2	0.0	
86430	0	3	2	0.0	
65955	0	1	1	0.0	
28058	1	0	2	0.0	
18103	0	2	1	0.0	

	babies	is_repeated_guest	previous_cancellations	\
107224	0	0	0	
86430	0	0	0	
65955	0	0	0	
28058	0	0	0	
18103	0	0	0	

	previous_bookings_not_canceled	agent	company	adr	\
107224	0	1.194706	0.000000	1.702680	
86430	0	0.526589	0.000000	1.646667	
65955	0	0.000000	1.708702	1.716380	
28058	0	1.876320	0.000000	1.680725	
18103	0	0.000000	0.000000	1.550446	

	required_car_parking_spaces	total_of_special_requests
107224	0	1
86430	0	0
65955	0	0
28058	0	0
18103	0	0

```
[121]: y_train.head(), y_test.head()
```

```
[121]: (88129    0
        4958    0
        106714  0
        17040    0
        20395    0
        Name: is_canceled, dtype: int64,
```

```

107224    0
86430     0
65955     1
28058     0
18103     0
Name: is_canceled, dtype: int64)

```

```

[122]: xgb = XGBClassifier(booster = 'gbtree', learning_rate = 0.1, max_depth = 5,
    ↪n_estimators = 180)
xgb.fit(X_train, y_train)

y_pred_xgb = xgb.predict(X_test)

acc_xgb = accuracy_score(y_test, y_pred_xgb)
conf = confusion_matrix(y_test, y_pred_xgb)
clf_report = classification_report(y_test, y_pred_xgb)

print(f"Accuracy Score : {acc_xgb}")
print(f"Confusion Matrix : \n{conf}")
print(f"Classification Report : \n{clf_report}")

```

Accuracy Score : 0.9829991891060593

Confusion Matrix :

```

[[22584    13]
 [  595 12571]]

```

Classification Report :

	precision	recall	f1-score	support
0	0.97	1.00	0.99	22597
1	1.00	0.95	0.98	13166
accuracy			0.98	35763
macro avg	0.99	0.98	0.98	35763
weighted avg	0.98	0.98	0.98	35763

[ ]: