importig the libraries

# EDA of Hotel Bookings Demand

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

## importing and understanding the dataset

```
df=pd.read_csv("hotel_bookings.csv")
df.shape
df.head
```

```
<bound method NDFrame.head of              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  ...  deposit_type  agent company  \
0                          0       2  ...    No Deposit    NaN     NaN
1                          0       2  ...    No Deposit    NaN     NaN
2                          1       1  ...    No Deposit    NaN     NaN
3                          1       1  ...    No Deposit  304.0     NaN
4                          2       2  ...    No Deposit  240.0     NaN
...                      ...     ...  ...           ...    ...     ...
119385                     5       2  ...    No Deposit  394.0     NaN
119386                     5       3  ...    No Deposit    9.0     NaN
119387                     5       2  ...    No Deposit    9.0     NaN
119388                     5       2  ...    No Deposit   89.0     NaN
119389                     7       2  ...    No Deposit    9.0     NaN

       days_in_waiting_list customer_type     adr  \
0                         0     Transient    0.00
1                         0     Transient    0.00
2                         0     Transient   75.00
```

```
3                       0     Transient   75.00
```

###

```
pd.set_option('display.max_columns',32)
```

Double-click (or enter) to edit

```
df.columns
```

```
df.nunique()
```

```
hotel                             2
is_canceled                       2
lead_time                       479
arrival_date_year                 3
arrival_date_month               12
arrival_date_week_number         53
arrival_date_day_of_month        31
stays_in_weekend_nights          17
stays_in_week_nights             35
adults                           14
children                          5
babies                            5
meal                              5
country                         177
market_segment                    8
distribution_channel              5
is_repeated_guest                 2
previous_cancellations           15
previous_bookings_not_canceled   73
reserved_room_type               10
assigned_room_type               12
booking_changes                  21
```

```
deposit_type                      3
agent                           333
company                         352
days_in_waiting_list            128
customer_type                     4
adr                            8879
required_car_parking_spaces       5
total_of_special_requests         6
reservation_status                3
reservation_status_date         926
dtype: int64
```

Double-click (or enter) to edit

```
df['hotel'].value_counts()
```

```
City Hotel      79330
Resort Hotel    40060
Name: hotel, dtype: int64
```

Double-click (or enter) to edit

```
df['meal'].value_counts()
```

```
BB           92310
HB           14463
SC           10650
Undefined     1169
FB             798
Name: meal, dtype: int64
```

Double-click (or enter) to edit

```
df['market_segment'].value_counts()
```

```
Online TA        56477
Offline TA/TO    24219
Groups           19811
Direct           12606
Corporate         5295
Complementary      743
Aviation           237
Undefined            2
Name: market_segment, dtype: int64
```

Double-click (or enter) to edit

```
df['distribution_channel'].value_counts()
```

```
TA/TO        97870
Direct       14645
Corporate     6677
GDS            193
Undefined        5
Name: distribution_channel, dtype: int64
```

Double-click (or enter) to edit

```
df['deposit_type'].value_counts()
```

```
No Deposit    104641
Non Refund     14587
Refundable       162
Name: deposit_type, dtype: int64
```

Double-click (or enter) to edit

```
df['customer_type'].value_counts()
```

```
Transient          89613
Transient-Party    25124
Contract            4076
Group                577
Name: customer_type, dtype: int64
```

Double-click (or enter) to edit

```
df['total_of_special_requests'].value_counts()
```

```
0    70318
1    33226
2    12969
3     2497
4      340
5       40
Name: total_of_special_requests, dtype: int64
```

Double-click (or enter) to edit

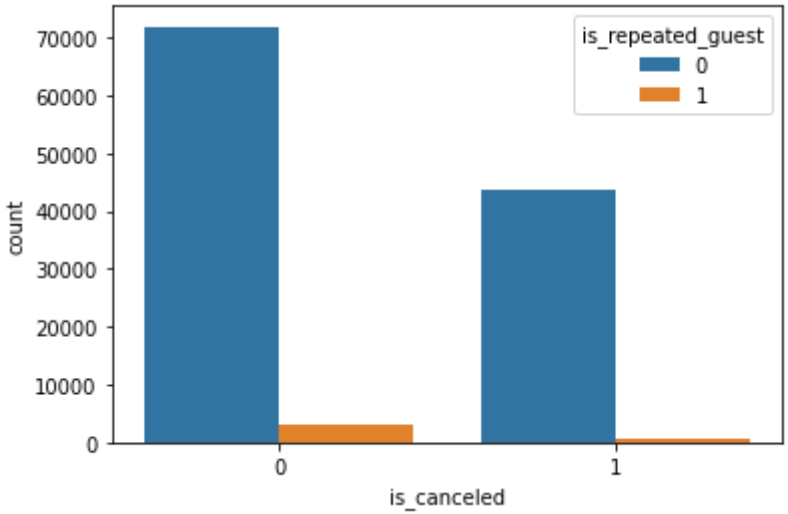
```
sns.countplot(data=df,x='hotel')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fe70e4c2a50>
```


Double-click (or enter) to edit

```
sns.countplot(data=df,x='is_canceled',hue='is_repeated_guest')
```

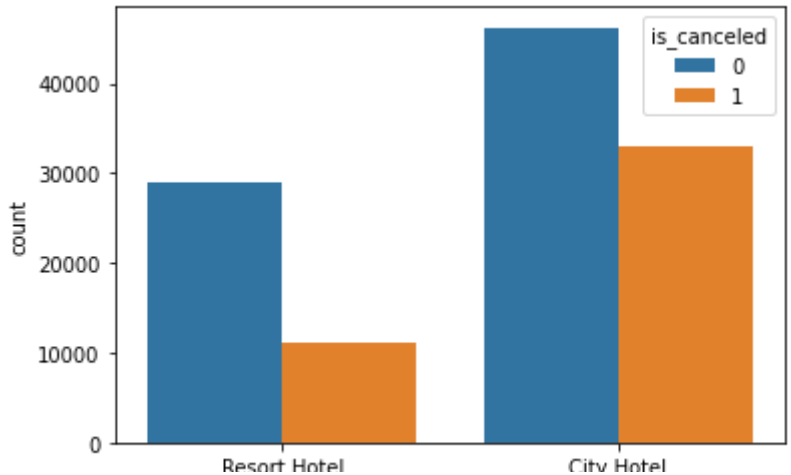
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fe70e3d0f90>
```




Double-click (or enter) to edit

```
sns.countplot(data=df,x='hotel',hue='is_canceled')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fe70decb050>
```




# Data Preparation

## Missing Data

```
df.isnull().values.any()
```

```
True
```

```
df.isnull().sum()
```

```
hotel                           0
is_canceled                     0
lead_time                       0
arrival_date_year               0
arrival_date_month              0
arrival_date_week_number        0
arrival_date_day_of_month       0
```

```
stays_in_weekend_nights               0
stays_in_week_nights                  0
adults                                0
children                              4
babies                                0
meal                                  0
country                             488
market_segment                        0
distribution_channel                  0
is_repeated_guest                     0
previous_cancellations                0
previous_bookings_not_canceled        0
reserved_room_type                    0
assigned_room_type                    0
booking_changes                       0
deposit_type                          0
agent                             16340
company                          112593
days_in_waiting_list                  0
customer_type                         0
adr                                   0
required_car_parking_spaces           0
total_of_special_requests             0
reservation_status                    0
reservation_status_date               0
dtype: int64
```

```
df.fillna(0,inplace=True)
```

```
df.isnull().sum()
```

```
hotel                           0
is_canceled                     0
lead_time                       0
arrival_date_year               0
arrival_date_month              0
arrival_date_week_number        0
arrival_date_day_of_month       0
```

```
stays_in_weekend_nights           0
stays_in_week_nights              0
adults                            0
children                          0
babies                            0
meal                              0
country                           0
market_segment                    0
distribution_channel              0
is_repeated_guest                 0
previous_cancellations            0
previous_bookings_not_canceled    0
reserved_room_type                0
assigned_room_type                0
booking_changes                   0
deposit_type                      0
agent                             0
company                           0
days_in_waiting_list              0
customer_type                     0
adr                               0
required_car_parking_spaces       0
total_of_special_requests         0
reservation_status                0
reservation_status_date           0
dtype: int64
```

```
df["meal"].replace("Undefined","SC", inplace=True)
```

Double-click (or enter) to edit

```
df["meal"].unique()
```

```
array(['BB', 'FB', 'HB', 'SC'], dtype=object)
```

```
Subset=df[(df['children']==0) & (df['adults']==0) & (df['babies']==0)]
```

Double-click (or enter) to edit

```
Subset[['adults','babies','children']]
```

| | adults | babies | children |
|---|---|---|---|
| **2224** | 0 | 0 | 0.0 |
| **2409** | 0 | 0 | 0.0 |
| **3181** | 0 | 0 | 0.0 |
| **3684** | 0 | 0 | 0.0 |
| **3708** | 0 | 0 | 0.0 |
| **...** | ... | ... | ... |
| **115029** | 0 | 0 | 0.0 |
| **115091** | 0 | 0 | 0.0 |
| **116251** | 0 | 0 | 0.0 |
| **116534** | 0 | 0 | 0.0 |
| **117087** | 0 | 0 | 0.0 |

180 rows × 3 columns

Double-click (or enter) to edit

```
type(Subset)
```

```
pandas.core.frame.DataFrame
```

```
Delete=(df['children']==0) & (df['adults']==0) & (df['babies']==0)
```

Double-click (or enter) to edit

```
type(Delete)
```

```
pandas.core.series.Series
```

Double-click (or enter) to edit

```
Delete
```

```
0         False
1         False
2         False
3         False
4         False
          ...
119385    False
119386    False
119387    False
119388    False
119389    False
Length: 119390, dtype: bool
```

```
data=df[~Delete]
```

Double-click (or enter) to edit

```
data.head()
```

| | hotel | is_canceled | lead_time | arrival_date_year | arrival_date_month | arrival_date_week_number | arrival_date_day_of_month | s |
|---|---|---|---|---|---|---|---|---|
| **0** | Resort Hotel | 0 | 342 | 2015 | July | 27 | 1 | |
| **1** | Resort Hotel | 0 | 737 | 2015 | July | 27 | 1 | |
| **2** | Resort Hotel | 0 | 7 | 2015 | July | 27 | 1 | |
| **3** | Resort Hotel | 0 | 13 | 2015 | July | 27 | 1 | |
| **4** | Resort Hotel | 0 | 14 | 2015 | July | 27 | 1 | |

```
Subset=data[(data['children']==0) & (data['adults']==0) & (data['babies']==0)]
```

Double-click (or enter) to edit

```
Subset
```

| | hotel | is_canceled | lead_time | arrival_date_year | arrival_date_month | arrival_date_week_number | arrival_date_day_of_month | st |
|---|---|---|---|---|---|---|---|---|

```
data.shape
```

```
(119210, 32)
```

Double-click (or enter) to edit

```
119390-119210
```

```
180
```

Double-click (or enter) to edit

```
data.to_csv('Updataed_Hotel_Booking.csv', index=False)
```

## Home country of Guests

```
guest_country=data[data['is_canceled']==0  ]['country'].value_counts().reset_index()
guest_country.columns=['country','Number of guests']
```

Double-click (or enter) to edit

```
guest_country
```

| | country | Number of guests |
|---|---|---|
| **0** | PRT | 20977 |
| **1** | GBR | 9668 |
| **2** | FRA | 8468 |

```
import plotly.offline as pyo
import plotly.graph_objs as go
import plotly.express as px
```

Double-click (or enter) to edit

```
total_guests = guest_country["Number of guests"].sum()
print(total_guests)
```

```
75011
```

Double-click (or enter) to edit

```
guest_country["Guests in %"] = round(guest_country["Number of guests"] / total_guests * 100, 2)
guest_country
```

| | country | Number of guests | Guests in % |
|---|---|---|---|
| **0** | PRT | 20977 | 27.97 |
| **1** | GBR | 9668 | 12.89 |
| **2** | FRA | 8468 | 11.29 |
| **3** | ESP | 6383 | 8.51 |
| **4** | DEU | 6067 | 8.09 |
| **...** | ... | ... | ... |

Double-click (or enter) to edit

| | | | |
|---|---|---|---|
| **162** | DJI | 1 | 0.00 |

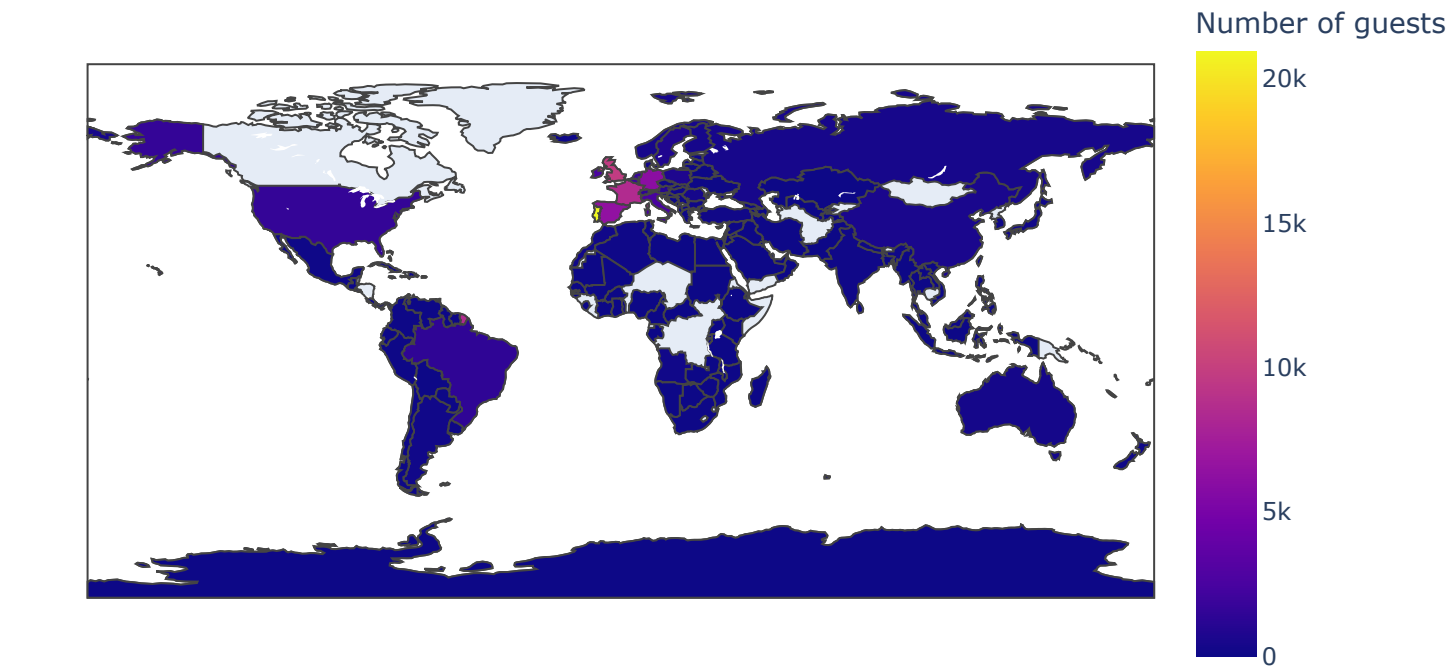
```
trace= go.Bar(
    x=guest_country["country"],
    y=guest_country['Number of guests'],
    marker=dict(color='#CD7F32')
)
data1 = [ trace]
layout = go.Layout(
    title='Guests by Country'
)
fig = go.Figure(data=data1, layout=layout)
pyo.plot(fig)
```

```
'temp-plot.html'
```

Double-click (or enter) to edit

```
map_guest = px.choropleth(guest_country,
                    locations=guest_country['country'],
                    color=guest_country['Number of guests'],
                    hover_name=guest_country['country'],
```

```
                      title="Home country of guests")
map_guest.show()
```




## Misinterpreting Data

```
resort = data[(data["hotel"] == "Resort Hotel") & (data["is_canceled"] == 0)]
city = data[(data["hotel"] == "City Hotel") & (data["is_canceled"] == 0)]
```

Double-click (or enter) to edit

```
resort
```

| | hotel | is_canceled | lead_time | arrival_date_year | arrival_date_month | arrival_date_week_number | arrival_date_day_of_mont |
|---|---|---|---|---|---|---|---|
| 0 | Resort | 0 | 342 | 2015 | July | 27 | |

```
resort_hotel=resort.groupby(['arrival_date_month'])['adr'].mean().reset_index()
resort_hotel
```

| | 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 |

```
city_hotel=city.groupby(['arrival_date_month'])['adr'].mean().reset_index()
city_hotel
```

| | 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 |

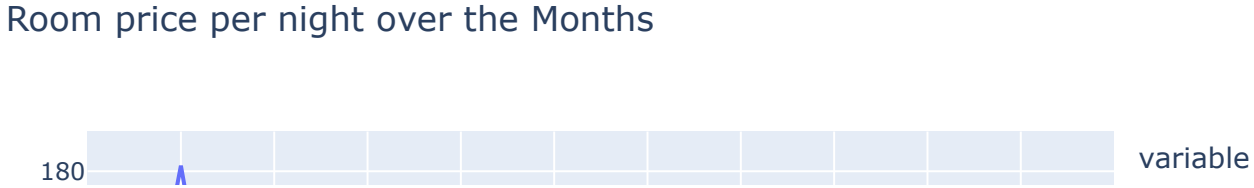
```
final=resort_hotel.merge(city_hotel,on='arrival_date_month')
final.columns=['month','price_for_resort','price_for_city_hotel']
final
```

| | 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 |

## Room price per night over the months

| | | | |
|---|---|---|---|
| **5** | July | 150.122528 | 115.818019 |

```
px.line(final, x='month',
        y=['price_for_resort','price_for_city_hotel'],
        title='Room price per night over the Months')
```




## Guests Pay For A Room Per Night

```
df['reserved_room_type'].unique()
```

```
array(['C', 'A', 'D', 'E', 'G', 'F', 'H', 'L', 'P', 'B'], dtype=object)
```

## Plotting the graph

```
data["adr_Updated"]=data["adr"]/(data["adults"]+data["children"])
data
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ve
```

| | hotel | is_canceled | lead_time | arrival_date_year | arrival_date_month | arrival_date_week_number | arrival_date_day_of_mon |
|---|---|---|---|---|---|---|---|
| **0** | Resort Hotel | 0 | 342 | 2015 | July | 27 | |
| **1** | Resort Hotel | 0 | 737 | 2015 | July | 27 | |
| **2** | Resort Hotel | 0 | 7 | 2015 | July | 27 | |
| **3** | Resort Hotel | 0 | 13 | 2015 | July | 27 | |
| **4** | Resort Hotel | 0 | 14 | 2015 | July | 27 | |
| **...** | ... | ... | ... | ... | ... | ... | |
| **119385** | City Hotel | 0 | 23 | 2017 | August | 35 | |
| | City | | | | | | |

```
data["adr_Updated"]=data["adr"]/(data["adults"]+data["children"])
valid_guest= data.loc[data["is_canceled"] == 0]
prices = valid_guest[["hotel", "reserved_room_type", "adr_Updated"]].sort_values("reserved_room_type")

plt.figure(figsize=(12, 8))
sns.boxplot(x="reserved_room_type",
            y="adr_Updated",
            hue="hotel",
            data=prices
```

```
            )
    plt.title("Price of room types per night and person", fontsize=16)
    plt.xlabel("Room type", fontsize=16)
    plt.ylabel("Price [EUR]", fontsize=16)

    plt.ylim(0, 160)
    plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
prices_C=prices[prices['reserved_room_type']=='C']
prices_C
```

| | hotel | reserved_room_type | adr_Updated |
|---|---|---|---|
| **119278** | City Hotel | C | 104.000000 |
| **99432** | City Hotel | C | 65.000000 |
| **99431** | City Hotel | C | 65.000000 |
| **101137** | City Hotel | C | 55.000000 |
| **0** | Resort Hotel | C | 0.000000 |
| **...** | ... | ... | ... |
| **27668** | Resort Hotel | C | 88.500000 |
| **39525** | Resort Hotel | C | 87.500000 |
| **15642** | Resort Hotel | C | 62.500000 |
| **15609** | Resort Hotel | C | 40.500000 |
| **38758** | Resort Hotel | C | 80.666667 |

623 rows × 3 columns

```
prices_City=prices_C[prices_C['hotel']=='City Hotel']
prices_Resort=prices_C[prices_C['hotel']=='Resort Hotel']
prices_Resort
```

| | hotel | reserved_room_type | adr_Updated |
|---|---|---|---|
| **0** | Resort Hotel | C | 0.000000 |
| **34951** | Resort Hotel | C | 30.250000 |
| **15955** | Resort Hotel | C | 47.366667 |
| **25990** | Resort Hotel | C | 62.166667 |
| **15960** | Resort Hotel | C | 90.000000 |
| **...** | ... | ... | ... |
| **27668** | Resort Hotel | C | 88.500000 |
| **39525** | Resort Hotel | C | 87.500000 |
| **15642** | Resort Hotel | C | 62.500000 |
| **15609** | Resort Hotel | C | 40.500000 |
| **38758** | Resort Hotel | C | 80.666667 |

615 rows × 3 columns

```
prices_City
```

| | hotel | reserved_room_type | adr_Updated |
|---|---|---|---|

```
prices_Resort.describe()
```

| | adr_Updated |
|---|---|
| **count** | 615.000000 |
| **mean** | 50.549085 |
| **std** | 27.821397 |
| **min** | 0.000000 |
| **25%** | 31.000000 |
| **50%** | 49.000000 |
| **75%** | 65.562500 |
| **max** | 171.000000 |

## Stay time

```
df3=data[data['is_canceled']==0]
df3["total_nights"] = df3["stays_in_weekend_nights"] + df3["stays_in_week_nights"]
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ve
```

```
df3
```

| | hotel | is_canceled | lead_time | arrival_date_year | arrival_date_month | arrival_date_week_number | arrival_date_day_of_mon |
|---|---|---|---|---|---|---|---|
| **0** | Resort Hotel | 0 | 342 | 2015 | July | 27 | |
| **1** | Resort Hotel | 0 | 737 | 2015 | July | 27 | |
| **2** | Resort Hotel | 0 | 7 | 2015 | July | 27 | |
| **3** | Resort Hotel | 0 | 13 | 2015 | July | 27 | |
| **4** | Resort Hotel | 0 | 14 | 2015 | July | 27 | |
| **...** | ... | ... | ... | ... | ... | ... | |
| **119385** | City Hotel | 0 | 23 | 2017 | August | 35 | |
| **119386** | City Hotel | 0 | 102 | 2017 | August | 35 | |
| **119387** | City Hotel | 0 | 34 | 2017 | August | 35 | |
| **119388** | City Hotel | 0 | 109 | 2017 | August | 35 | |
| **119389** | City Hotel | 0 | 205 | 2017 | August | 35 | |

75011 rows × 34 columns

```
df4=df3[['total_nights','hotel','is_canceled']]
df4
```

| | total_nights | hotel | is_canceled |
|---|---|---|---|
| **0** | 0 | Resort Hotel | 0 |
| **1** | 0 | Resort Hotel | 0 |
| **2** | 1 | Resort Hotel | 0 |
| **3** | 1 | Resort Hotel | 0 |
| **4** | 2 | Resort Hotel | 0 |
| **...** | ... | ... | ... |
| **119385** | 7 | City Hotel | 0 |
| **119386** | 7 | City Hotel | 0 |
| **119387** | 7 | City Hotel | 0 |
| **119388** | 7 | City Hotel | 0 |
| **119389** | 9 | City Hotel | 0 |

75011 rows × 3 columns

```
hotel_stay=df4.groupby(['total_nights','hotel']).agg('count').reset_index()

hotel_stay
```

| | total_nights | hotel | is_canceled |
|---|---|---|---|
| **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 |

```
hotel_stay=hotel_stay.rename(columns={'is_canceled':'Number of stays'})
hotel_stay.head()
```

| | 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 |

```
hotel_stay_r=hotel_stay[hotel_stay['hotel']=='Resort Hotel']
hotel_stay_r
```

| | total_nights | hotel | Number of stays |
|---|---|---|---|
| **1** | 0 | Resort Hotel | 371 |
| **3** | 1 | Resort Hotel | 6579 |
| **5** | 2 | Resort Hotel | 4488 |
| **7** | 3 | Resort Hotel | 3828 |
| **9** | 4 | Resort Hotel | 3321 |
| **11** | 5 | Resort Hotel | 1899 |
| **13** | 6 | Resort Hotel | 1205 |
| **15** | 7 | Resort Hotel | 4434 |
| **17** | 8 | Resort Hotel | 509 |
| **19** | 9 | Resort Hotel | 408 |
| **21** | 10 | Resort Hotel | 699 |
| **23** | 11 | Resort Hotel | 240 |
| **25** | 12 | Resort Hotel | 89 |
| **27** | 13 | Resort Hotel | 75 |
| **29** | 14 | Resort Hotel | 630 |
| **31** | 15 | Resort Hotel | 23 |
| **33** | 16 | Resort Hotel | 12 |
| **35** | 17 | Resort Hotel | 11 |
| **37** | 18 | Resort Hotel | 5 |
| **39** | 19 | Resort Hotel | 4 |
| **42** | 21 | Resort Hotel | 35 |
| **44** | 22 | Resort Hotel | 7 |

| | | | |
|---|---|---|---|
| **46** | 23 | Resort Hotel | 1 |
| **48** | 25 | Resort Hotel | 14 |
| **49** | 28 | Resort Hotel | 22 |
| **50** | 29 | Resort Hotel | 2 |
| **51** | 30 | Resort Hotel | 2 |
| **53** | 35 | Resort Hotel | 5 |
| **54** | 38 | Resort Hotel | 1 |

```
hotel_stay_c=hotel_stay[hotel_stay['hotel']=='City Hotel']
hotel_stay_c
```

| | total_nights | hotel | Number of stays |
|---|---|---|---|
| **0** | 0 | City Hotel | 251 |
| **2** | 1 | City Hotel | 9155 |
| **4** | 2 | City Hotel | 10983 |
| **6** | 3 | City Hotel | 11889 |
| **8** | 4 | City Hotel | 7694 |
| **10** | 5 | City Hotel | 3210 |
| **12** | 6 | City Hotel | 1111 |
| **14** | 7 | City Hotel | 1245 |
| **16** | 8 | City Hotel | 205 |
| **18** | 9 | City Hotel | 119 |
| **20** | 10 | City Hotel | 80 |
| **22** | 11 | City Hotel | 33 |
| **24** | 12 | City Hotel | 33 |

```
trace = go.Bar(
    x=hotel_stay_r["total_nights"],
    y=hotel_stay_r["Number of stays"],
    name='Resort Stay'
    )

trace1=go.Bar(
    x=hotel_stay_c["total_nights"],
    y=hotel_stay_c["Number of stays"],
    name='City stay'
    )
```

```
data5 = [trace,trace1]
layout = go.Layout(
    title='Total Number of stays by Guest'
)
fig = go.Figure(data=data5, layout=layout)
pyo.plot(fig)
```

```
'temp-plot.html'
```

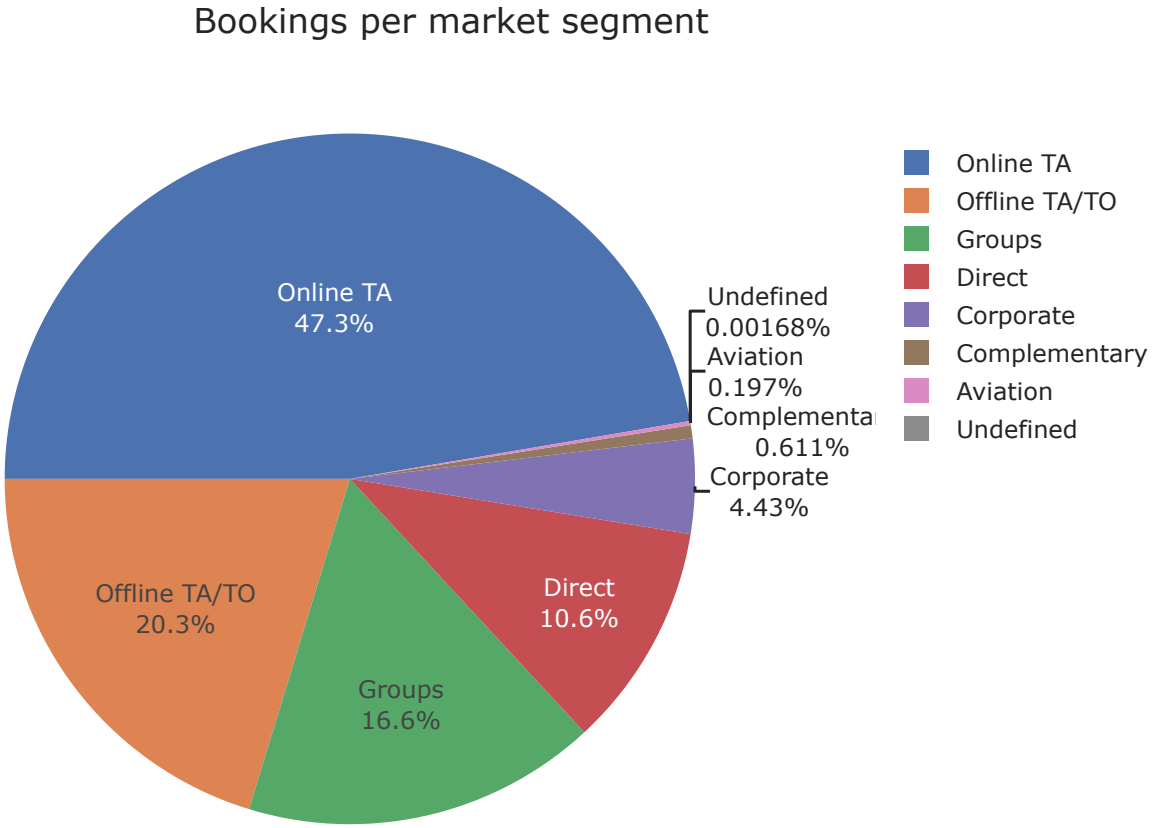
## Bookings by market segment

```
segments=data["market_segment"].value_counts()
segments
```

```
Online TA        56408
Offline TA/TO    24182
Groups           19791
Direct           12582
Corporate         5282
Complementary      728
Aviation           235
Undefined            2
Name: market_segment, dtype: int64
```

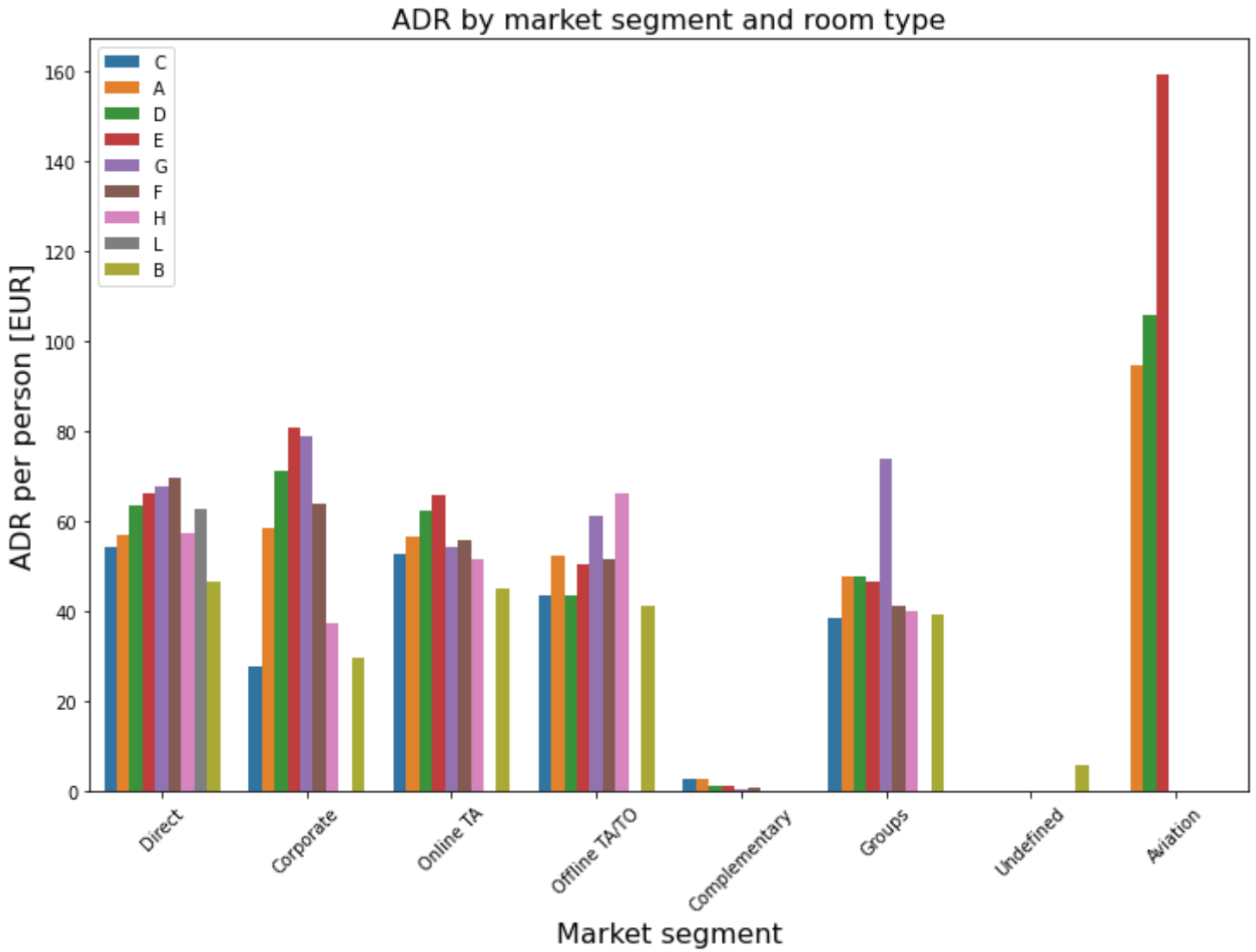
```
segments=data["market_segment"].value_counts()

# pie plot
fig = px.pie(segments,
             values=segments.values,
             names=segments.index,
             title="Bookings per market segment",
             template="seaborn")
fig.update_traces(rotation=-90, textinfo="percent+label")
fig.show()
```




```
plt.figure(figsize=(12, 8))
sns.barplot(x="market_segment",
            y="adr_Updated",
            hue="reserved_room_type",
            data=data,
            ci=None)
plt.title("ADR by market segment and room type", fontsize=16)
plt.xlabel("Market segment", fontsize=16)
```

```
plt.xticks(rotation=45)
plt.ylabel("ADR per person [EUR]", fontsize=16)
plt.legend(loc="upper left")
plt.show()
```




## Number of bookings get canceled

```
Cancel=data['is_canceled']==1
```

```
cancel=Cancel.sum()
```

```
resort_cancelation = data.loc[data["hotel"] == "Resort Hotel"]["is_canceled"].sum()
city_cancelation = data.loc[data["hotel"] == "City Hotel"]["is_canceled"].sum()
```

```
resort_cancelation
```

```
11120
```

```
city_cancelation
```

```
33079
```

```
print(f"Total Booking Cancelled : {cancel} . ")
print(f"Total Resort Hotel Booking Cancelled : {resort_cancelation} . ")
print(f"Total City Hotel Booking Cancelled : {city_cancelation} . ")
```

```
Total Booking Cancelled : 44199 .
Total Resort Hotel Booking Cancelled : 11120 .
Total City Hotel Booking Cancelled : 33079 .
```

## Month having the highest number of cancelations

```
res_book_per_month = data.loc[(data["hotel"] == "Resort Hotel")].groupby("arrival_date_month")["hotel"].count()
res_cancel_per_month = data.loc[(data["hotel"] == "Resort Hotel")].groupby("arrival_date_month")["is_canceled"].sum()

cty_book_per_month = data.loc[(data["hotel"] == "City Hotel")].groupby("arrival_date_month")["hotel"].count()
cty_cancel_per_month = data.loc[(data["hotel"] == "City Hotel")].groupby("arrival_date_month")["is_canceled"].sum()
```

```
res_cancel_data = pd.DataFrame({"Hotel": "Resort Hotel",
                                "Month": list(res_book_per_month.index),
                                "Bookings": list(res_book_per_month.values),
                                "Cancelations": list(res_cancel_per_month.values)})
cty_cancel_data = pd.DataFrame({"Hotel": "City Hotel",
                                "Month": list(cty_book_per_month.index),
                                "Bookings": list(cty_book_per_month.values),
                                "Cancelations": list(cty_cancel_per_month.values)})
```

```
res_cancel_data
```

| | Hotel | Month | Bookings | Cancelations |
|---|---|---|---|---|
| **0** | Resort Hotel | April | 3609 | 1059 |
| **1** | Resort Hotel | August | 4894 | 1637 |
| **2** | Resort Hotel | December | 2645 | 631 |
| **3** | Resort Hotel | February | 3102 | 794 |
| **4** | Resort Hotel | January | 2191 | 325 |
| **5** | Resort Hotel | July | 4573 | 1436 |
| **6** | Resort Hotel | June | 3044 | 1007 |
| **7** | Resort Hotel | March | 3334 | 763 |
| **8** | Resort Hotel | May | 3559 | 1024 |
| **9** | Resort Hotel | November | 2435 | 460 |
| **10** | Resort Hotel | October | 3553 | 978 |
| **11** | Resort Hotel | September | 3108 | 1006 |

```
plt.figure(figsize=(12, 8))

trace = go.Bar(
    x=res_cancel_data["Month"],
    y=res_cancel_data["Cancelations"],
    name="Rst Cancelled"
    )
trace1 = go.Bar(
    x=cty_cancel_data["Month"],
    y=cty_cancel_data["Cancelations"],
    name="Cty Cancelled"
    )


data6 = [trace,trace1]
layout = go.Layout(
    title='Total Number of stays by Guest'
)
fig = go.Figure(data=data6, layout=layout)
pyo.plot(fig)
```

```
 'temp-plot.html'
 <Figure size 864x576 with 0 Axes>
```

Colab paid products - Cancel contracts here