In [1]:
```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
```

In [2]:
```
hotel=pd.read_csv("hotel_booking.csv")
hotel
```

Out[2]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_wee
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	
...	...	...	...	...	...	
119385	City Hotel	0	23	2017	August	
119386	City Hotel	0	102	2017	August	
119387	City Hotel	0	34	2017	August	
119388	City Hotel	0	109	2017	August	
119389	City Hotel	0	205	2017	August	

119390 rows × 32 columns

In [3]:
```
hotel.head()
```

Out[3]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_nun
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	

5 rows × 32 columns

In [4]:
```
hotel.isnull().sum()
```

Out[4]:
```
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
```

In [5]:
```
hotel.drop(["agent","company"],axis=1,inplace=True)
```

In [6]:
```
hotel.isna().sum()
```

Out[6]:

```
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
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
```

In [7]:

```
hotel.dropna(inplace=True)
```

In [8]:

```
hotel.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 118898 entries, 0 to 119389
Data columns (total 30 columns):
 #   Column                          Non-Null Count   Dtype  
---  ------                          --------------   -----  
 0   hotel                           118898 non-null  object 
 1   is_canceled                     118898 non-null  int64  
 2   lead_time                       118898 non-null  int64  
 3   arrival_date_year               118898 non-null  int64  
 4   arrival_date_month              118898 non-null  object 
 5   arrival_date_week_number        118898 non-null  int64  
 6   arrival_date_day_of_month       118898 non-null  int64  
 7   stays_in_weekend_nights         118898 non-null  int64  
 8   stays_in_week_nights            118898 non-null  int64  
 9   adults                          118898 non-null  int64  
 10  children                        118898 non-null  float64
 11  babies                          118898 non-null  int64  
 12  meal                            118898 non-null  object 
 13  country                         118898 non-null  object 
 14  market_segment                  118898 non-null  object 
 15  distribution_channel            118898 non-null  object 
 16  is_repeated_guest               118898 non-null  int64  
 17  previous_cancellations          118898 non-null  int64  
 18  previous_bookings_not_canceled  118898 non-null  int64  
 19  reserved_room_type              118898 non-null  object 
 20  assigned_room_type              118898 non-null  object 
 21  booking_changes                 118898 non-null  int64  
 22  deposit_type                    118898 non-null  object 
 23  days_in_waiting_list            118898 non-null  int64  
 24  customer_type                   118898 non-null  object 
 25  adr                             118898 non-null  float64
 26  required_car_parking_spaces     118898 non-null  int64  
 27  total_of_special_requests       118898 non-null  int64  
 28  reservation_status              118898 non-null  object 
 29  reservation_status_date         118898 non-null  object 
dtypes: float64(2), int64(16), object(12)
memory usage: 28.1+ MB
```

In [9]:
```
hotel.describe()
```

Out[9]:

	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date
count	118898.000000	118898.000000	118898.000000	118898.000000	
mean	0.371352	104.311435	2016.157656	27.166555	
std	0.483168	106.903309	0.707459	13.589971	
min	0.000000	0.000000	2015.000000	1.000000	
25%	0.000000	18.000000	2016.000000	16.000000	
50%	0.000000	69.000000	2016.000000	28.000000	
75%	1.000000	161.000000	2017.000000	38.000000	
max	1.000000	737.000000	2017.000000	53.000000	

In [10]:
```
hotel.columns
```

Out[10]:
```
Index(['hotel', 'is_canceled', 'lead_time', 'arrival_date_year',
       'arrival_date_month', 'arrival_date_week_number',
       'arrival_date_day_of_month', 'stays_in_weekend_nights',
       'stays_in_week_nights', 'adults', 'children', 'babies', 'meal',
       'country', 'market_segment', 'distribution_channel',
       'is_repeated_guest', 'previous_cancellations',
       'previous_bookings_not_canceled', 'reserved_room_type',
       'assigned_room_type', 'booking_changes', 'deposit_type',
       'days_in_waiting_list', 'customer_type', 'adr',
       'required_car_parking_spaces', 'total_of_special_requests',
       'reservation_status', 'reservation_status_date'],
      dtype='object')
```

In [11]:
```
hotel["reservation_status_date"]
```

Out[11]:
```
0         1/7/2015
1         1/7/2015
2         2/7/2015
3         2/7/2015
4         3/7/2015
           ...
119385    6/9/2017
119386    7/9/2017
119387    7/9/2017
119388    7/9/2017
119389    7/9/2017
Name: reservation_status_date, Length: 118898, dtype: object
```

In [12]:
```
hotel.rename(columns={"is_canceled":"canceled"},inplace=True)
```

In [13]:
```
hotel["canceled"].replace(1,"Yes",inplace=True)
```

In [14]:
```
hotel["canceled"].replace(0,"No",inplace=True)
```

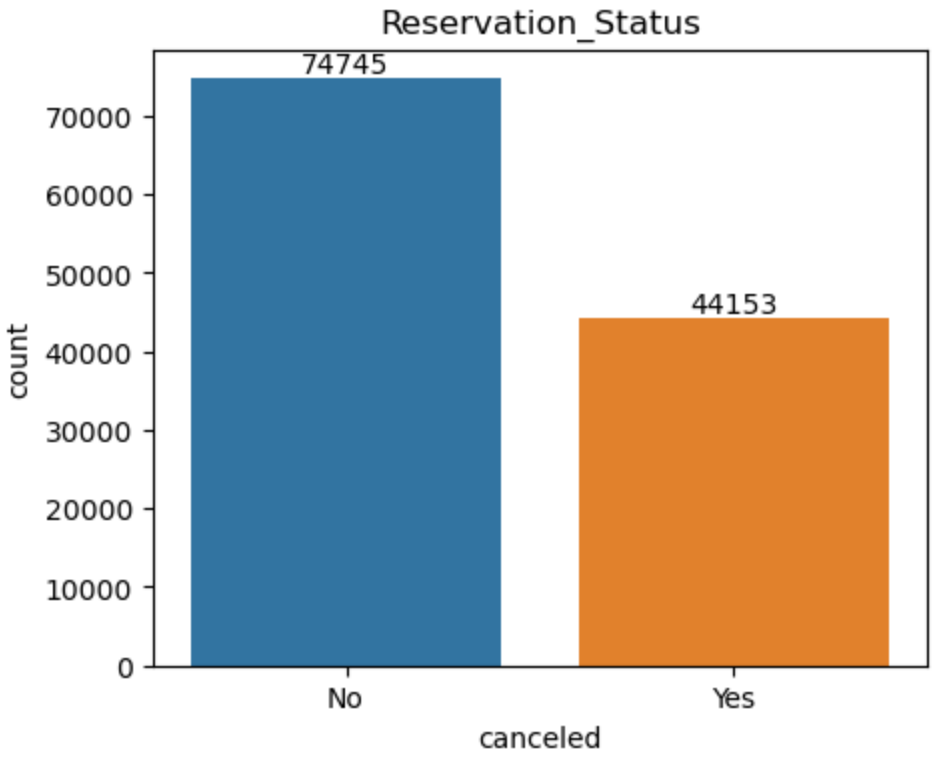
In [15]:
```
cancel_status=hotel['canceled'].value_counts()
cancel_status
```

Out[15]:
```
No     74745
Yes    44153
Name: canceled, dtype: int64
```

In [16]:
```
hotel.shape
```

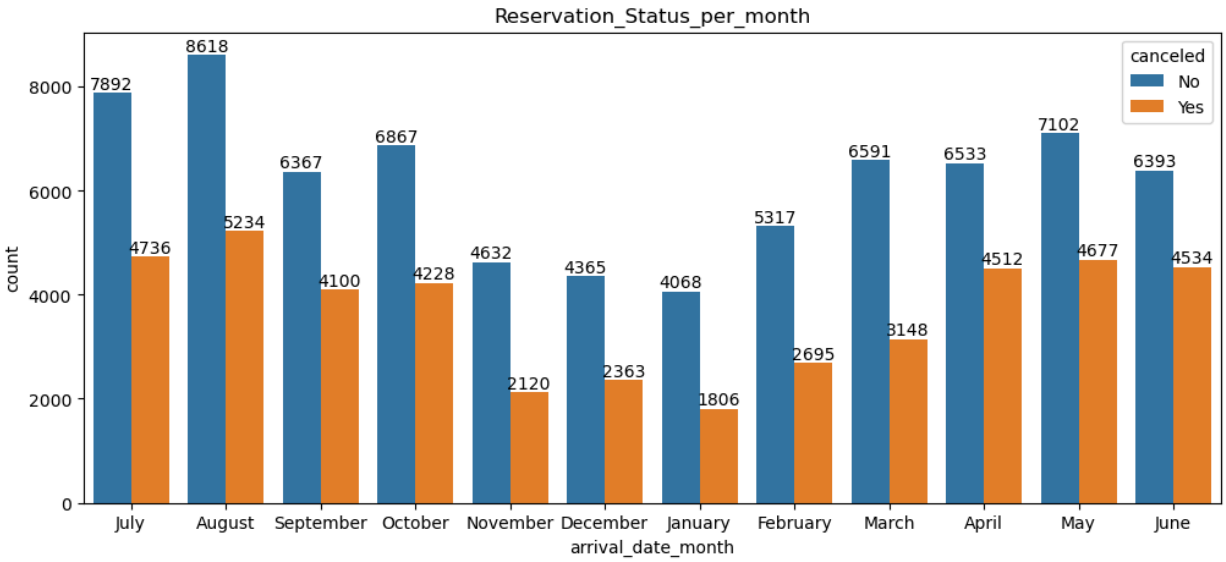
Out[16]:
```
(118898, 30)
```

In [19]:
```
plt.figure(figsize=(5,4))
plt.title("Reservation_Status")
x=sns.countplot(x='canceled',data=hotel)
x
for bars in x.containers:
 x.bar_label(bars)
```



In [21]:
```
plt.title("Reservation_for_different_hotels")
y=sns.countplot(x="hotel",hue="canceled",data=hotel)
plt.ylabel("No_of_Reservation")
y
for bars in y.containers:
 y.bar_label(bars)
```

In [23]:
```
plt.figure(figsize=(12,5))
plt.title("Reservation_Status_per_month")
m=sns.countplot(x="arrival_date_month",hue="canceled",data=hotel)
m
for bars in m.containers:
 m.bar_label(bars)
```

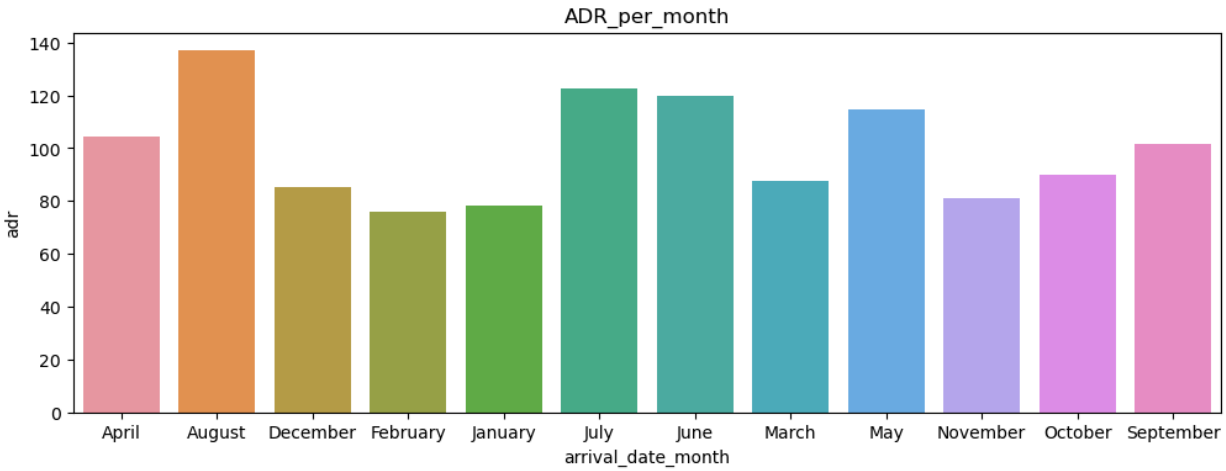


In [32]:
```
data=hotel[hotel["canceled"]=="Yes"].groupby(["arrival_date_month"],as_index=Fa
data
```

Out[32]:

	arrival_date_month	adr
0	April	104.198958
1	August	136.929135
2	December	85.249399
3	February	75.927506
4	January	78.037580
5	July	122.660245
6	June	119.619559
7	March	87.759257
8	May	114.556427
9	November	81.114538
10	October	89.787394
11	September	101.706537

In [33]:
```
plt.figure(figsize=(12,4))
sns.barplot(x="arrival_date_month",y="adr",data=data)
plt.title("ADR_per_month")
plt.show()
```

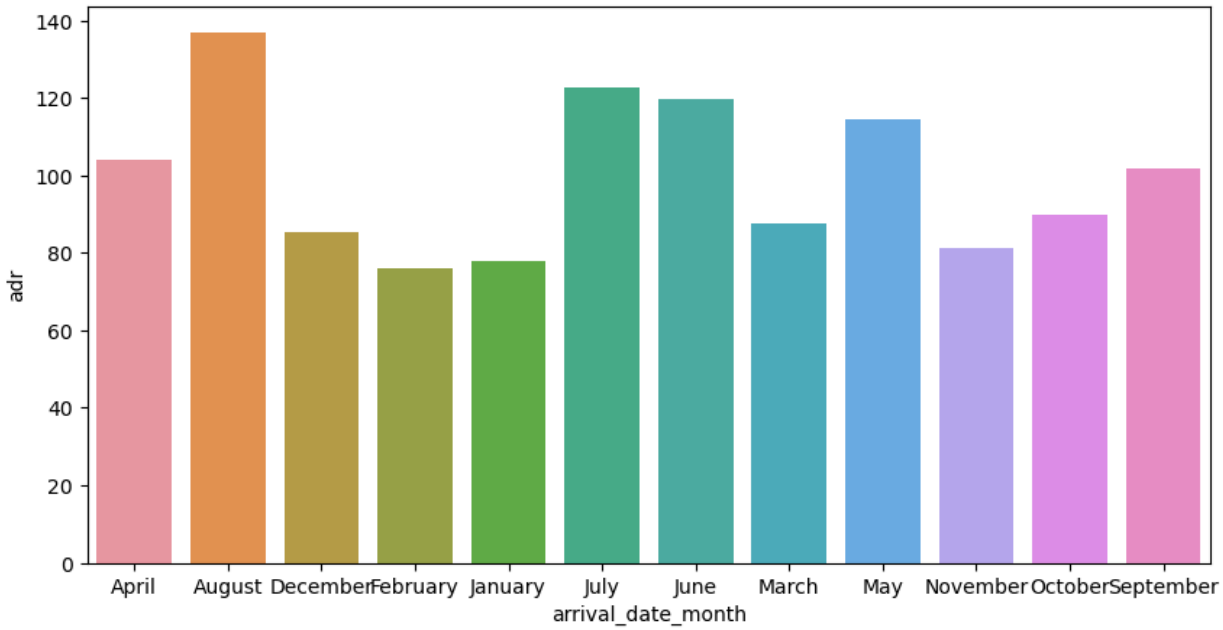


In [34]:
```
m=hotel[hotel["canceled"]=="Yes"].groupby(["arrival_date_month"],as_index=Fals
m
```

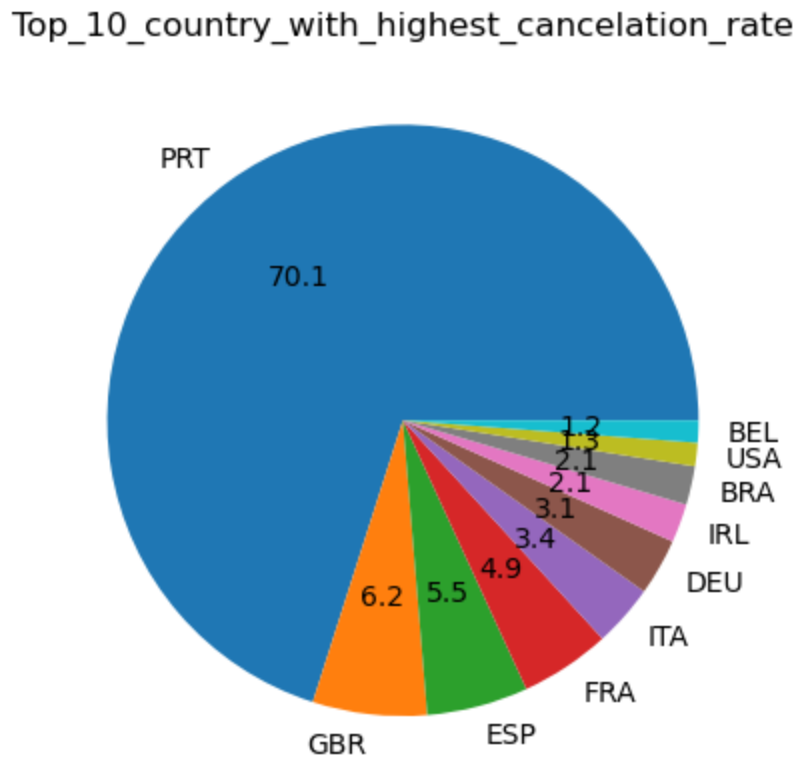
Out[34]:

	arrival_date_month	adr
0	April	104.198958
1	August	136.929135
2	December	85.249399
3	February	75.927506
4	January	78.037580
5	July	122.660245
6	June	119.619559
7	March	87.759257
8	May	114.556427
9	November	81.114538
10	October	89.787394
11	September	101.706537

In [35]:
```
plt.figure(figsize=(10,5))
sns.barplot(x="arrival_date_month",y="adr",data=m)
plt.show()
```



In [36]:
```
canceled_data=hotel[hotel["canceled"]=="Yes"]
top_10_country=canceled_data["country"].value_counts()[ :10]
plt.pie(top_10_country,labels=top_10_country.index,autopct="%.1f")
plt.title("Top_10_country_with_highest_cancelation_rate")
plt.show()
```



In [37]:
```
hotel["market_segment"].value_counts()
```

Out[37]:
```
Online TA        56402
Offline TA/TO    24160
Groups           19806
Direct           12448
Corporate         5111
Complementary      734
Aviation           237
Name: market_segment, dtype: int64
```

In [38]:
```
canceled_data["market_segment"].value_counts()
```

Out[38]:
```
Online TA        20738
Groups           12097
Offline TA/TO     8278
Direct            1920
Corporate          978
Complementary       90
Aviation            52
Name: market_segment, dtype: int64
```

In [ ]:
```

```