hotel_booking

September 5, 2024

[1]:

```
!pip install numpy
```

```
Requirement already satisfied: numpy in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (1.24.3)
```

[2]:

```
!pip install pandas
!pip install seaborn
```

```
Requirement already satisfied: pandas in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (2.0.3)
Requirement already satisfied: python-dateutil>=2.8.2 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from pandas)
(2.8.2)
Requirement already satisfied: pytz>=2020.1 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from pandas)
(2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from pandas)
(2023.3)
Requirement already satisfied: numpy>=1.21.0 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from pandas)
(1.24.3)
Requirement already satisfied: six>=1.5 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from python-
dateutil>=2.8.2->pandas) (1.16.0)
Requirement already satisfied: seaborn in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (0.12.2)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from seaborn)
(1.24.3)
Requirement already satisfied: pandas>=0.25 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from seaborn)
(2.0.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from seaborn)
(3.7.2)
Requirement already satisfied: contourpy>=1.0.1 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (1.0.5)
```

```
Requirement already satisfied: cycler>=0.10 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (23.1)
Requirement already satisfied: pillow>=6.2.0 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (10.2.0)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from
pandas>=0.25->seaborn) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from
pandas>=0.25->seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from python-
dateutil>=2.7->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)
```

[3]:

```
!pip install matplotlib
```

```
Requirement already satisfied: matplotlib in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (3.7.2)
Requirement already satisfied: contourpy>=1.0.1 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from matplotlib)
(1.0.5)
Requirement already satisfied: cycler>=0.10 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from matplotlib)
(0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from matplotlib)
(4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from matplotlib)
(1.4.4)
Requirement already satisfied: numpy>=1.20 in
```

```
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from matplotlib)
(1.24.3)
Requirement already satisfied: packaging>=20.0 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from matplotlib)
(23.1)
Requirement already satisfied: pillow>=6.2.0 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from matplotlib)
(10.2.0)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from matplotlib)
(3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from matplotlib)
(2.8.2)
Requirement already satisfied: six>=1.5 in
c:\users\dskho410\appdata\local\anaconda3\lib\site-packages (from python-
dateutil>=2.7->matplotlib) (1.16.0)
```

[4]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

[5]:
```python
df = pd.read_csv('C:\\Users\\DSKHO410\\Downloads\\hotel_bookings 2.csv')
df
```

[5]:
```
               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
4                         0     Transient   98.00
…                     …           …     …
119385                    0     Transient   96.14
119386                    0     Transient  225.43
119387                    0     Transient  157.71
119388                    0     Transient  104.40
119389                    0     Transient  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

        reservation_status reservation_status_date
0                Check-Out                1/7/2015
1                Check-Out                1/7/2015
2                Check-Out                2/7/2015
3                Check-Out                2/7/2015
4                Check-Out                3/7/2015
…                    …                      …
119385           Check-Out                6/9/2017
119386           Check-Out                7/9/2017
119387           Check-Out                7/9/2017
119388           Check-Out                7/9/2017
119389           Check-Out                7/9/2017

[119390 rows x 32 columns]
```

[6]:

```
df.head()
```

[6]:

```
          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  …  deposit_type  \
0                        0                     0       2  …    No Deposit
1                        0                     0       2  …    No Deposit
2                        0                     1       1  …    No Deposit
```

```
3                        0                     1       1  ...    No Deposit
4                        0                     2       2  ...    No Deposit

   agent company days_in_waiting_list customer_type   adr  \
0    NaN     NaN                    0     Transient   0.0
1    NaN     NaN                    0     Transient   0.0
2    NaN     NaN                    0     Transient  75.0
3  304.0     NaN                    0     Transient  75.0
4  240.0     NaN                    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
0                1/7/2015
1                1/7/2015
2                2/7/2015
3                2/7/2015
4                3/7/2015

[5 rows x 32 columns]
```

[7]:

```
df.shape
```

[7]:

```
(119390, 32)
```

[8]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 32 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
dtypes: float64(4), int64(16), object(12)
memory usage: 29.1+ MB
```

[9]:

```
df['reservation_status_date'] = pd.
 ↪to_datetime(df['reservation_status_date'],format = '%d/%m/%Y')
```

[10]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 32 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  datetime64[ns]
dtypes: datetime64[ns](1), float64(4), int64(16), object(11)
memory usage: 29.1+ MB
```

[11]:

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

[11]:

```
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
```

[12]:

```
df.drop(columns=['agent','company'],axis = 1,inplace =True)
df.dropna(inplace = True)
```

[13]:

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

[13]:

```
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
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
```

[14]:
```
df.describe()
```

[14]:
```
         is_canceled      lead_time  arrival_date_year  \
count  118898.000000  118898.000000      118898.000000
mean        0.371352     104.311435        2016.157656
min         0.000000       0.000000        2015.000000
25%         0.000000      18.000000        2016.000000
50%         0.000000      69.000000        2016.000000
75%         1.000000     161.000000        2017.000000
max         1.000000     737.000000        2017.000000
std         0.483168     106.903309           0.707459

       arrival_date_week_number  arrival_date_day_of_month  \
count             118898.000000              118898.000000
mean                  27.166555                  15.800880
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
std                   13.589971                   8.780324

       stays_in_weekend_nights  stays_in_week_nights         adults  \
count            118898.000000         118898.000000  118898.000000
mean                  0.928897              2.502145       1.858391
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                  16.000000             41.000000      55.000000
std                   0.996216              1.900168       0.578576

            children         babies  is_repeated_guest  \
count  118898.000000  118898.000000      118898.000000
mean        0.104207       0.007948           0.032011
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
std         0.399172       0.097380           0.176029

       previous_cancellations  previous_bookings_not_canceled  \
count           118898.000000                   118898.000000
mean                 0.087142                        0.131634
min                  0.000000                        0.000000
25%                  0.000000                        0.000000
```

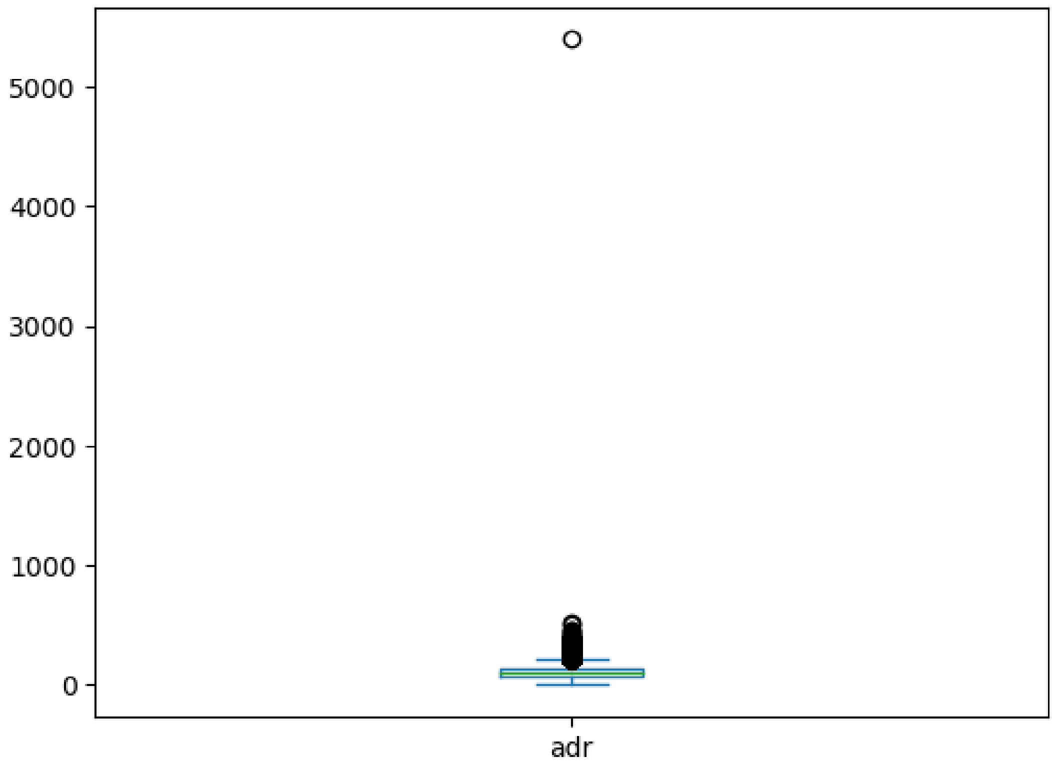
```
50%                  0.000000                        0.000000
75%                  0.000000                        0.000000
max                 26.000000                       72.000000
std                  0.845869                        1.484672

       booking_changes  days_in_waiting_list            adr  \
count    118898.000000         118898.000000  118898.000000
mean          0.221181              2.330754     102.003243
min           0.000000              0.000000      -6.380000
25%           0.000000              0.000000      70.000000
50%           0.000000              0.000000      95.000000
75%           0.000000              0.000000     126.000000
max          21.000000            391.000000    5400.000000
std           0.652785             17.630452      50.485862

       required_car_parking_spaces  total_of_special_requests  \
count                118898.000000              118898.000000
mean                      0.061885                   0.571683
min                       0.000000                   0.000000
25%                       0.000000                   0.000000
50%                       0.000000                   0.000000
75%                       0.000000                   1.000000
max                       8.000000                   5.000000
std                       0.244172                   0.792678

             reservation_status_date
count                         118898
mean   2016-07-30 07:37:53.336809984
min              2014-10-17 00:00:00
25%              2016-02-02 00:00:00
50%              2016-08-08 00:00:00
75%              2017-02-09 00:00:00
max              2017-09-14 00:00:00
std                              NaN
```

[15]:

```
df['adr'].plot(kind='box')
```

[15]:

```
<Axes: >
```



```
[16]: df = df[df['adr'] < 5000]
```

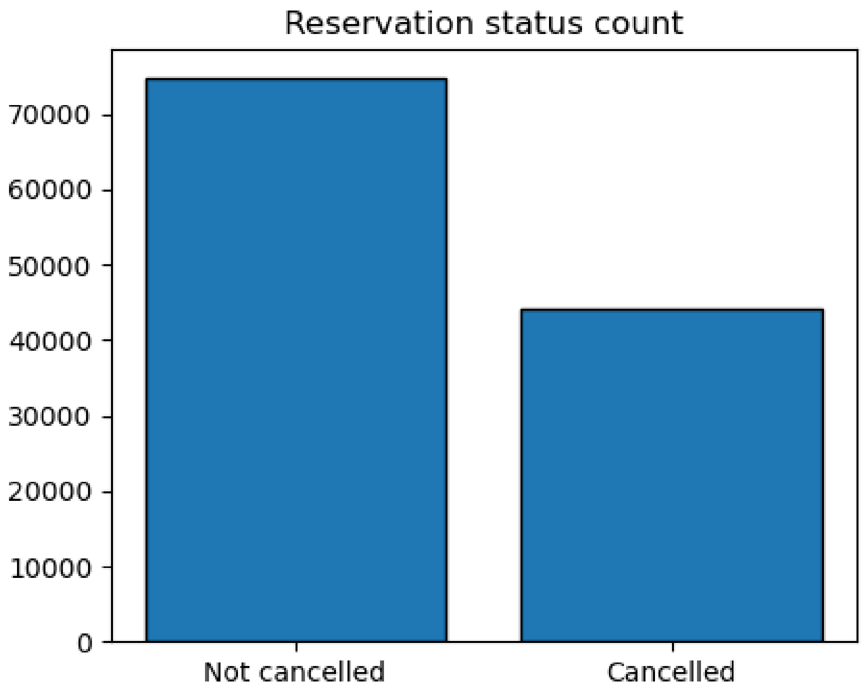
```
[17]: df['is_canceled'].value_counts(normalize = True)
```

```
[17]: is_canceled
      0    0.628653
      1    0.371347
      Name: proportion, dtype: float64
```

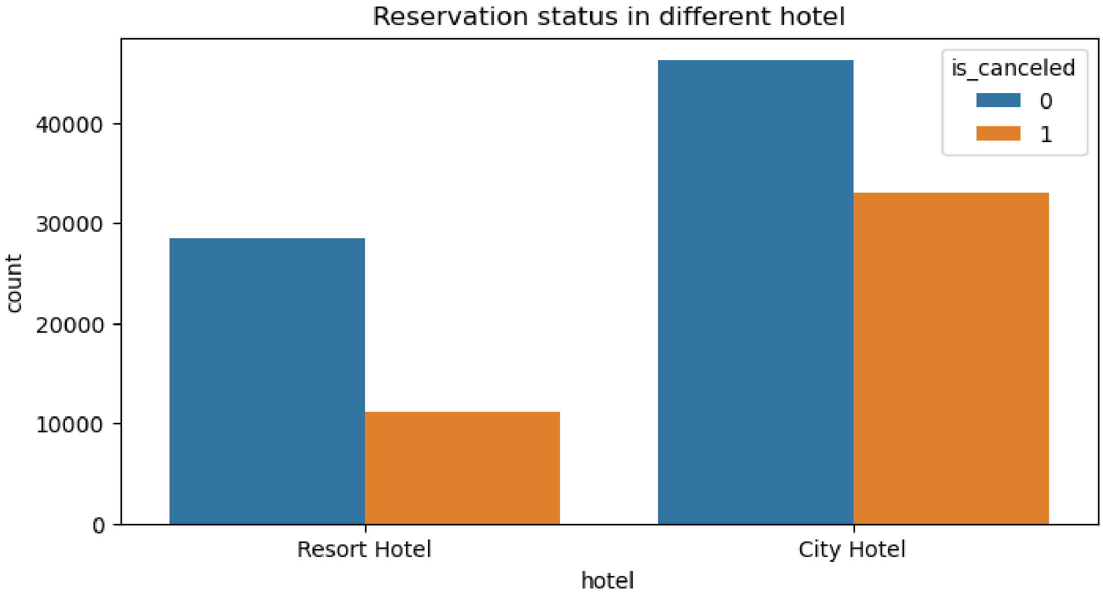
```
[18]: cancelled_perc = df['is_canceled'].value_counts(normalize = True)
      print(cancelled_perc)
```

```
is_canceled
0    0.628653
1    0.371347
Name: proportion, dtype: float64
```

```
[19]: plt.figure(figsize=(5,4))
      plt.title('Reservation status count')
      plt.bar(['Not cancelled','Cancelled'],df['is_canceled'].
        ↪value_counts(),edgecolor='k')
      plt.show()
```



[20]:
```
plt.figure(figsize=(8,4))
sns.countplot(x = 'hotel',hue ='is_canceled',data = df)
plt.title('Reservation status in different hotel')
plt.show()
```



[21]:

```
resort_hotel = df[df['hotel'] == 'Resort Hotel']
resort_hotel['is_canceled'].value_counts(normalize = True)
```

[21]:

```
is_canceled
0    0.72025
1    0.27975
Name: proportion, dtype: float64
```

[22]:

```
citytel = df[df['hotel'] == 'City Hotel']
citytel['is_canceled'].value_counts(normalize = True)
```

[22]:

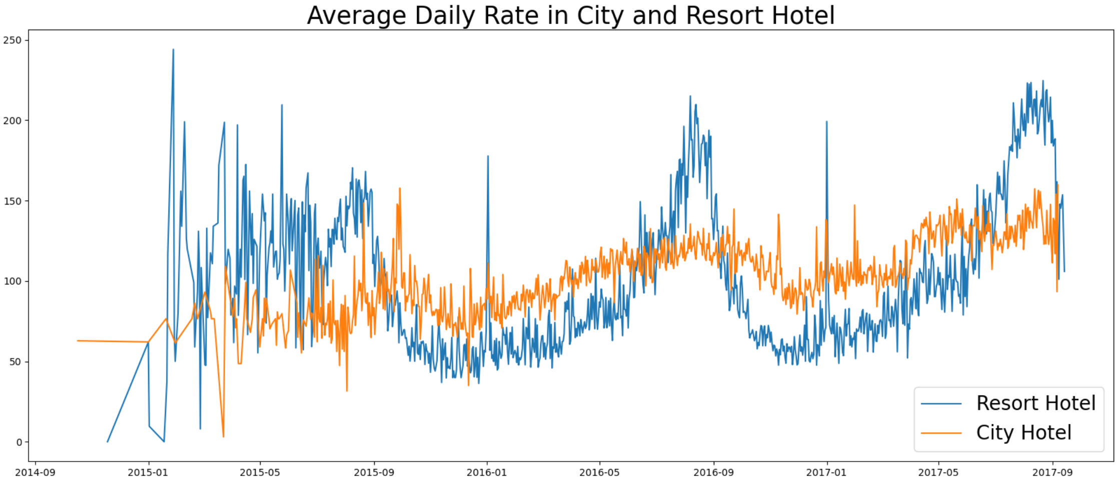
```
is_canceled
0    0.582918
1    0.417082
Name: proportion, dtype: float64
```

[23]:

```
resort_hotel = resort_hotel.groupby('reservation_status_date')[['adr']].mean()
citytel = citytel.groupby('reservation_status_date')[['adr']].mean()
```

[24]:

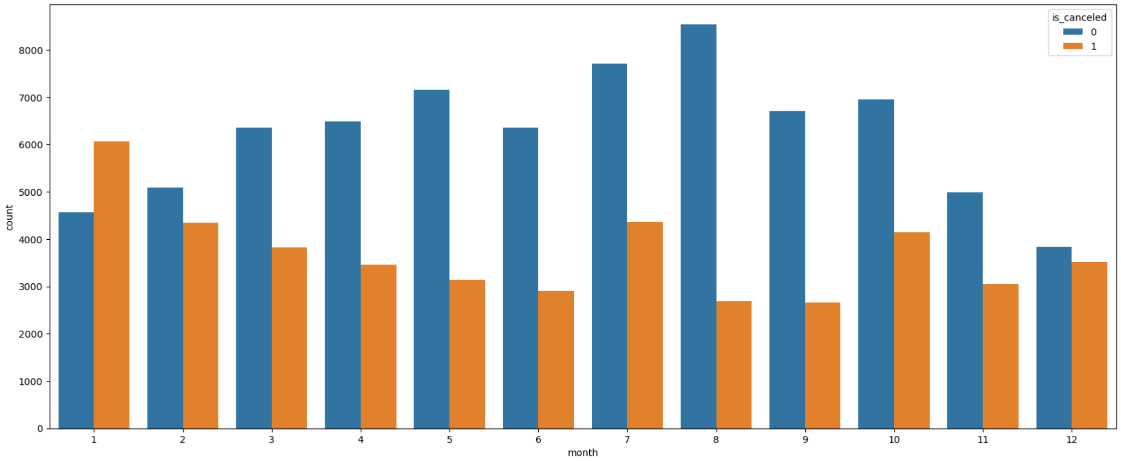
```
plt.figure(figsize=(20,8))
plt.title('Average Daily Rate in City and Resort Hotel',fontsize=25)
plt.plot(resort_hotel.index,resort_hotel['adr'],label = 'Resort Hotel')
plt.plot(citytel.index, citytel['adr'],label = 'City Hotel')
plt.legend(fontsize=20)
plt.show()
```



[25]:

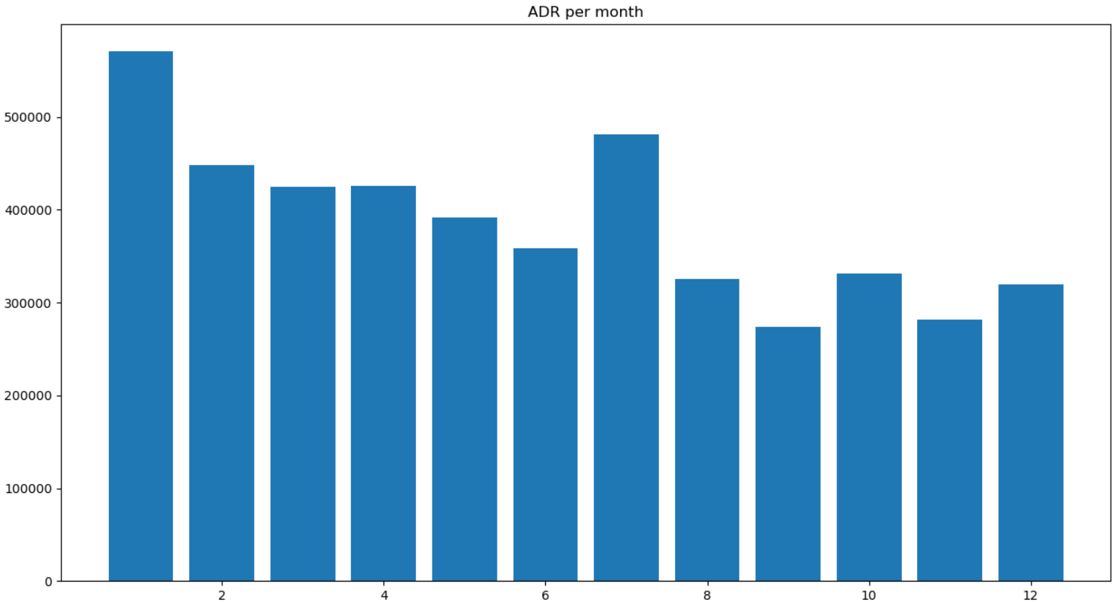
```
df['month'] = df['reservation_status_date'].dt.month
plt.figure(figsize=(20,8))
```

```
sns.countplot(x = 'month',hue ='is_canceled',data = df)
plt.show()
```



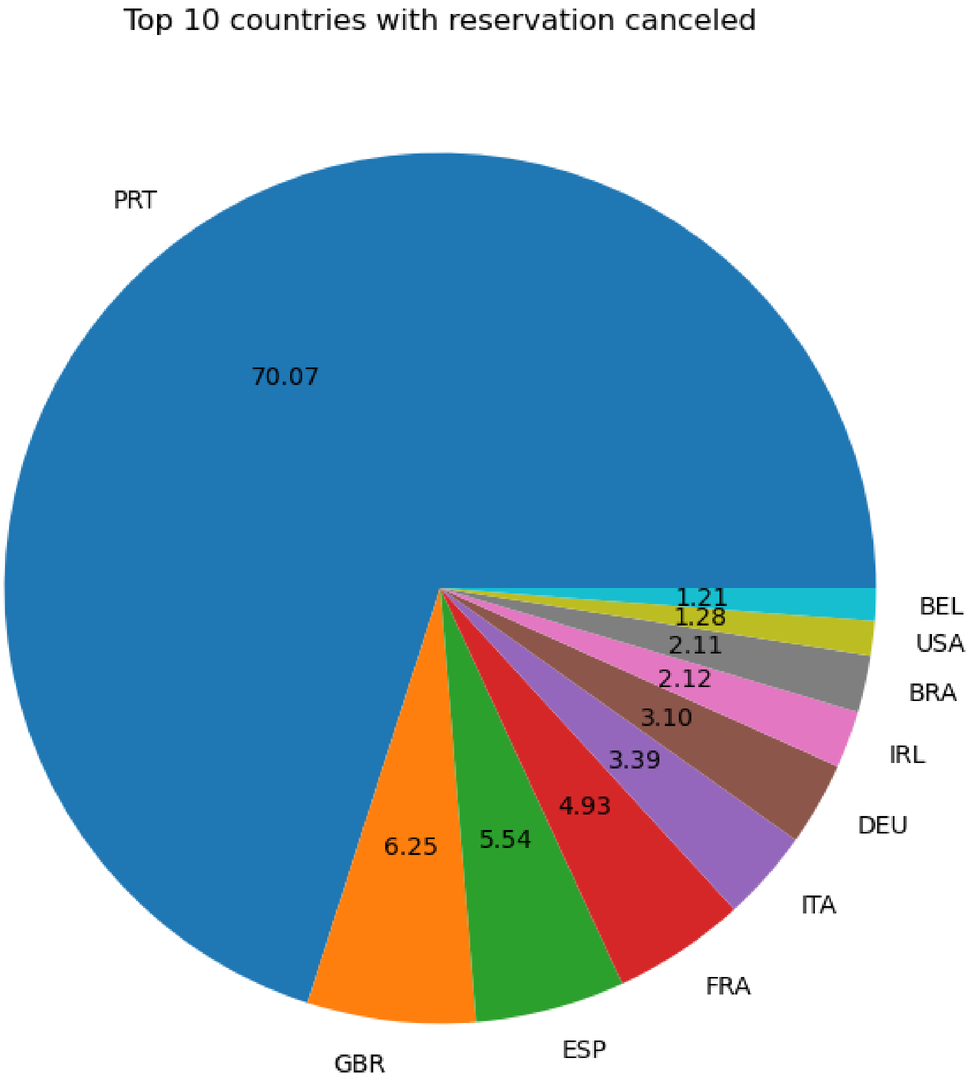
[26]:

```
plt.figure(figsize=(15,8))
plt.title('ADR per month')
plt.bar('month','adr', data =df[df['is_canceled'] ==1].
  ↪groupby('month')[['adr']].sum().reset_index())
plt.show()
```



[27]:

```
cancelled_data = df[df['is_canceled'] == 1]
top_10_country = cancelled_data['country'].value_counts()[:10]
plt.figure(figsize=(8,8))
plt.title('Top 10 countries with reservation canceled')
plt.pie(top_10_country,autopct='%.2f',labels = top_10_country.index)
plt.show()
```



[28]:

```
df['market_segment'].value_counts(normalize =True)
```

[28]:

```
market_segment
Online TA        0.474377
Offline TA/TO    0.203193
```

```
Groups           0.166581
Direct           0.104696
Corporate        0.042987
Complementary    0.006173
Aviation         0.001993
Name: proportion, dtype: float64
```

[29]:

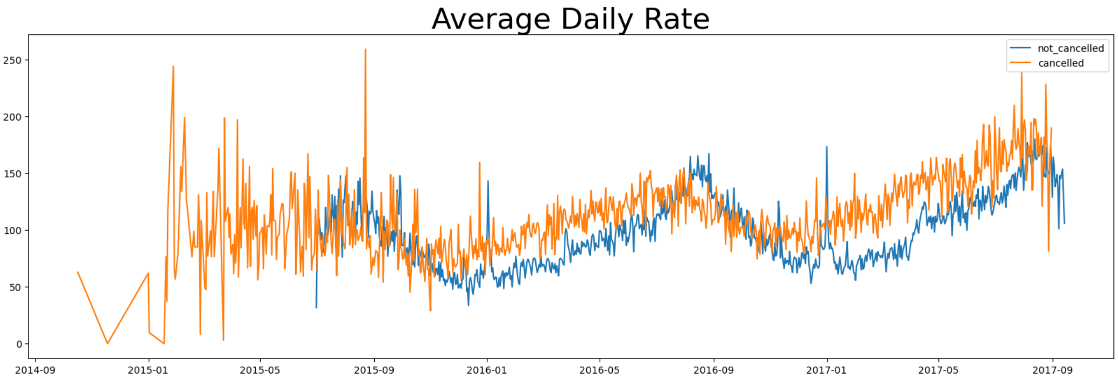
```
cancelled_data['market_segment'].value_counts(normalize =True)
```

[29]:

```
market_segment
Online TA        0.469696
Groups           0.273985
Offline TA/TO    0.187466
Direct           0.043486
Corporate        0.022151
Complementary    0.002038
Aviation         0.001178
Name: proportion, dtype: float64
```

[30]:

```
cancelled_df_adr = cancelled_data.groupby('reservation_status_date')[['adr']].
  ↪mean()
cancelled_df_adr.reset_index(inplace=True)
cancelled_df_adr.sort_values('reservation_status_date',inplace=True)

not_cancelled_data = df[df['is_canceled']==0]
not_cancelled_adr = not_cancelled_data.
  ↪groupby('reservation_status_date')[['adr']].mean()
not_cancelled_adr.reset_index(inplace=True)
not_cancelled_adr.sort_values('reservation_status_date',inplace=True)

plt.figure(figsize=(20,6))
plt.title('Average Daily Rate',fontsize=30)
plt.
  ↪plot(not_cancelled_adr['reservation_status_date'],not_cancelled_adr['adr'],label='not_cance
plt.
  ↪plot(cancelled_df_adr['reservation_status_date'],cancelled_df_adr['adr'],label='cancelled')
plt.legend()
plt.show()
```


Average Daily Rate
not_cancelled
cancelled
250
200
150
100
50
0
2014-09
2015-01
2015-05
2015-09
2016-01
2016-05
2016-09
2017-01
2017-05
2017-09