**Hotel data set**

Importing liabiers

In [1]:

**import** pandas **as** pd

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

**import** warnings

warnings.filterwarnings('ignore')

**loading of dataset**

In [2]:

**import** pandas **as** pd

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

**import** warnings

warnings.filterwarnings('ignore')

df **=** pd.read\_csv("D:\python\PANDAS\Hotel\_data\hotel\_booking.csv\hotel\_booking.csv")

df.head(2)

Out[2]:

|  | **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** | **...** | **customer\_type** | **adr** | **required\_car\_parking\_spaces** | **total\_of\_special\_requests** | **reservation\_status** | **reservation\_status\_date** | **name** | **email** | **phone-number** | **credit\_card** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | Resort Hotel | 0 | 342 | 2015 | July | 27 | 1 | 0 | 0 | 2 | ... | Transient | 0.0 | 0 | 0 | Check-Out | 2015-07-01 | Ernest Barnes | Ernest.Barnes31@outlook.com | 669-792-1661 | \*\*\*\*\*\*\*\*\*\*\*\*4322 |
| **1** | Resort Hotel | 0 | 737 | 2015 | July | 27 | 1 | 0 | 0 | 2 | ... | Transient | 0.0 | 0 | 0 | Check-Out | 2015-07-01 | Andrea Baker | Andrea\_Baker94@aol.com | 858-637-6955 | \*\*\*\*\*\*\*\*\*\*\*\*9157 |

2 rows × 36 columns

**Data exploration**

In [3]:

df.shape

Out[3]:

(119390, 36)

In [4]:

df.columns

Out[4]:

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', 'agent',

'company', 'days\_in\_waiting\_list', 'customer\_type', 'adr',

'required\_car\_parking\_spaces', 'total\_of\_special\_requests',

'reservation\_status', 'reservation\_status\_date', 'name', 'email',

'phone-number', 'credit\_card'],

dtype='object')

In [5]:

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

In [6]:

*## Formatting Datetime*

df['reservation\_status\_date'] **=** pd.to\_datetime(df['reservation\_status\_date'])

In [7]:

df.describe(include **=** 'object')

Out[7]:

|  | **hotel** | **arrival\_date\_month** | **meal** | **country** | **market\_segment** | **distribution\_channel** | **reserved\_room\_type** | **assigned\_room\_type** | **deposit\_type** | **customer\_type** | **reservation\_status** | **name** | **email** | **phone-number** | **credit\_card** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **count** | 119390 | 119390 | 119390 | 118902 | 119390 | 119390 | 119390 | 119390 | 119390 | 119390 | 119390 | 119390 | 119390 | 119390 | 119390 |
| **unique** | 2 | 12 | 5 | 177 | 8 | 5 | 10 | 12 | 3 | 4 | 3 | 81503 | 115889 | 119390 | 9000 |
| **top** | City Hotel | August | BB | PRT | Online TA | TA/TO | A | A | No Deposit | Transient | Check-Out | Michael Johnson | Michael.C@gmail.com | 669-792-1661 | \*\*\*\*\*\*\*\*\*\*\*\*4923 |
| **freq** | 79330 | 13877 | 92310 | 48590 | 56477 | 97870 | 85994 | 74053 | 104641 | 89613 | 75166 | 48 | 6 | 1 | 28 |

In [8]:

*## To find unique characters of columns*

**for** col **in** df.describe(include **=** 'object').columns:

print(col)

print(df[col].unique())

hotel

['Resort Hotel' 'City Hotel']

arrival\_date\_month

['July' 'August' 'September' 'October' 'November' 'December' 'January'

'February' 'March' 'April' 'May' 'June']

meal

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

country

['PRT' 'GBR' 'USA' 'ESP' 'IRL' 'FRA' nan 'ROU' 'NOR' 'OMN' 'ARG' 'POL'

'DEU' 'BEL' 'CHE' 'CN' 'GRC' 'ITA' 'NLD' 'DNK' 'RUS' 'SWE' 'AUS' 'EST'

'CZE' 'BRA' 'FIN' 'MOZ' 'BWA' 'LUX' 'SVN' 'ALB' 'IND' 'CHN' 'MEX' 'MAR'

'UKR' 'SMR' 'LVA' 'PRI' 'SRB' 'CHL' 'AUT' 'BLR' 'LTU' 'TUR' 'ZAF' 'AGO'

'ISR' 'CYM' 'ZMB' 'CPV' 'ZWE' 'DZA' 'KOR' 'CRI' 'HUN' 'ARE' 'TUN' 'JAM'

'HRV' 'HKG' 'IRN' 'GEO' 'AND' 'GIB' 'URY' 'JEY' 'CAF' 'CYP' 'COL' 'GGY'

'KWT' 'NGA' 'MDV' 'VEN' 'SVK' 'FJI' 'KAZ' 'PAK' 'IDN' 'LBN' 'PHL' 'SEN'

'SYC' 'AZE' 'BHR' 'NZL' 'THA' 'DOM' 'MKD' 'MYS' 'ARM' 'JPN' 'LKA' 'CUB'

'CMR' 'BIH' 'MUS' 'COM' 'SUR' 'UGA' 'BGR' 'CIV' 'JOR' 'SYR' 'SGP' 'BDI'

'SAU' 'VNM' 'PLW' 'QAT' 'EGY' 'PER' 'MLT' 'MWI' 'ECU' 'MDG' 'ISL' 'UZB'

'NPL' 'BHS' 'MAC' 'TGO' 'TWN' 'DJI' 'STP' 'KNA' 'ETH' 'IRQ' 'HND' 'RWA'

'KHM' 'MCO' 'BGD' 'IMN' 'TJK' 'NIC' 'BEN' 'VGB' 'TZA' 'GAB' 'GHA' 'TMP'

'GLP' 'KEN' 'LIE' 'GNB' 'MNE' 'UMI' 'MYT' 'FRO' 'MMR' 'PAN' 'BFA' 'LBY'

'MLI' 'NAM' 'BOL' 'PRY' 'BRB' 'ABW' 'AIA' 'SLV' 'DMA' 'PYF' 'GUY' 'LCA'

'ATA' 'GTM' 'ASM' 'MRT' 'NCL' 'KIR' 'SDN' 'ATF' 'SLE' 'LAO']

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' 'P' 'B']

assigned\_room\_type

['C' 'A' 'D' 'E' 'G' 'F' 'I' 'B' 'H' 'P' 'L' 'K']

deposit\_type

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

customer\_type

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

reservation\_status

['Check-Out' 'Canceled' 'No-Show']

name

['Ernest Barnes' 'Andrea Baker' 'Rebecca Parker' ... 'Wesley Aguilar'

'Caroline Conley MD' 'Ariana Michael']

email

['Ernest.Barnes31@outlook.com' 'Andrea\_Baker94@aol.com'

'Rebecca\_Parker@comcast.net' ... 'Mary\_Morales@hotmail.com'

'MD\_Caroline@comcast.net' 'Ariana\_M@xfinity.com']

phone-number

['669-792-1661' '858-637-6955' '652-885-2745' ... '395-518-4100'

'531-528-1017' '422-804-6403']

credit\_card

['\*\*\*\*\*\*\*\*\*\*\*\*4322' '\*\*\*\*\*\*\*\*\*\*\*\*9157' '\*\*\*\*\*\*\*\*\*\*\*\*3734' ...

'\*\*\*\*\*\*\*\*\*\*\*\*9170' '\*\*\*\*\*\*\*\*\*\*\*\*6349' '\*\*\*\*\*\*\*\*\*\*\*\*7959']

In [9]:

*## Column wise null value finding*

​

df.isnull().sum()

Out[9]:

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

name 0

email 0

phone-number 0

credit\_card 0

dtype: int64

In [10]:

*## Remove agent and company beacsue of irrelevent for analysis*

​

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

df.dropna(inplace**=True**)

In [11]:

df.isnull().sum()

Out[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 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

name 0

email 0

phone-number 0

credit\_card 0

dtype: int64

In [12]:

​

df.describe()

Out[12]:

|  | **is\_canceled** | **lead\_time** | **arrival\_date\_year** | **arrival\_date\_week\_number** | **arrival\_date\_day\_of\_month** | **stays\_in\_weekend\_nights** | **stays\_in\_week\_nights** | **adults** | **children** | **babies** | **is\_repeated\_guest** | **previous\_cancellations** | **previous\_bookings\_not\_canceled** | **booking\_changes** | **days\_in\_waiting\_list** | **adr** | **required\_car\_parking\_spaces** | **total\_of\_special\_requests** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **count** | 118898.000000 | 118898.000000 | 118898.000000 | 118898.000000 | 118898.000000 | 118898.000000 | 118898.000000 | 118898.000000 | 118898.000000 | 118898.000000 | 118898.000000 | 118898.000000 | 118898.000000 | 118898.000000 | 118898.000000 | 118898.000000 | 118898.000000 | 118898.000000 |
| **mean** | 0.371352 | 104.311435 | 2016.157656 | 27.166555 | 15.800880 | 0.928897 | 2.502145 | 1.858391 | 0.104207 | 0.007948 | 0.032011 | 0.087142 | 0.131634 | 0.221181 | 2.330754 | 102.003243 | 0.061885 | 0.571683 |
| **std** | 0.483168 | 106.903309 | 0.707459 | 13.589971 | 8.780324 | 0.996216 | 1.900168 | 0.578576 | 0.399172 | 0.097380 | 0.176029 | 0.845869 | 1.484672 | 0.652785 | 17.630452 | 50.485862 | 0.244172 | 0.792678 |
| **min** | 0.000000 | 0.000000 | 2015.000000 | 1.000000 | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | -6.380000 | 0.000000 | 0.000000 |
| **25%** | 0.000000 | 18.000000 | 2016.000000 | 16.000000 | 8.000000 | 0.000000 | 1.000000 | 2.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 70.000000 | 0.000000 | 0.000000 |
| **50%** | 0.000000 | 69.000000 | 2016.000000 | 28.000000 | 16.000000 | 1.000000 | 2.000000 | 2.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 95.000000 | 0.000000 | 0.000000 |
| **75%** | 1.000000 | 161.000000 | 2017.000000 | 38.000000 | 23.000000 | 2.000000 | 3.000000 | 2.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 126.000000 | 0.000000 | 1.000000 |
| **max** | 1.000000 | 737.000000 | 2017.000000 | 53.000000 | 31.000000 | 16.000000 | 41.000000 | 55.000000 | 10.000000 | 10.000000 | 1.000000 | 26.000000 | 72.000000 | 21.000000 | 391.000000 | 5400.000000 | 8.000000 | 5.000000 |

In [13]:

*## delete outliares*

df **=** df[df['adr']**<**5000]

In [14]:

df.describe()

Out[14]:

|  | **is\_canceled** | **lead\_time** | **arrival\_date\_year** | **arrival\_date\_week\_number** | **arrival\_date\_day\_of\_month** | **stays\_in\_weekend\_nights** | **stays\_in\_week\_nights** | **adults** | **children** | **babies** | **is\_repeated\_guest** | **previous\_cancellations** | **previous\_bookings\_not\_canceled** | **booking\_changes** | **days\_in\_waiting\_list** | **adr** | **required\_car\_parking\_spaces** | **total\_of\_special\_requests** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **count** | 118897.000000 | 118897.000000 | 118897.000000 | 118897.000000 | 118897.000000 | 118897.000000 | 118897.000000 | 118897.000000 | 118897.000000 | 118897.000000 | 118897.000000 | 118897.000000 | 118897.000000 | 118897.000000 | 118897.000000 | 118897.000000 | 118897.000000 | 118897.000000 |
| **mean** | 0.371347 | 104.312018 | 2016.157657 | 27.166674 | 15.800802 | 0.928905 | 2.502157 | 1.858390 | 0.104208 | 0.007948 | 0.032011 | 0.087143 | 0.131635 | 0.221175 | 2.330774 | 101.958683 | 0.061885 | 0.571688 |
| **std** | 0.483167 | 106.903570 | 0.707462 | 13.589966 | 8.780321 | 0.996217 | 1.900171 | 0.578578 | 0.399174 | 0.097381 | 0.176030 | 0.845872 | 1.484678 | 0.652784 | 17.630525 | 48.091199 | 0.244173 | 0.792680 |
| **min** | 0.000000 | 0.000000 | 2015.000000 | 1.000000 | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | -6.380000 | 0.000000 | 0.000000 |
| **25%** | 0.000000 | 18.000000 | 2016.000000 | 16.000000 | 8.000000 | 0.000000 | 1.000000 | 2.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 70.000000 | 0.000000 | 0.000000 |
| **50%** | 0.000000 | 69.000000 | 2016.000000 | 28.000000 | 16.000000 | 1.000000 | 2.000000 | 2.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 95.000000 | 0.000000 | 0.000000 |
| **75%** | 1.000000 | 161.000000 | 2017.000000 | 38.000000 | 23.000000 | 2.000000 | 3.000000 | 2.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 126.000000 | 0.000000 | 1.000000 |
| **max** | 1.000000 | 737.000000 | 2017.000000 | 53.000000 | 31.000000 | 16.000000 | 41.000000 | 55.000000 | 10.000000 | 10.000000 | 1.000000 | 26.000000 | 72.000000 | 21.000000 | 391.000000 | 510.000000 | 8.000000 | 5.000000 |

**Data analysis and visualization**

In [15]:

*## Finding percentage of cancellation*

*## (normalize = True has been use to percentage calcullation)*

perecentage\_cenecel **=** df['is\_canceled'].value\_counts(normalize **=** **True**)

perecentage\_cenecel

Out[15]:

0 0.628653

1 0.371347

Name: is\_canceled, dtype: float64

In [16]:

*## 62% is not cancelled*

*## 37% booking has been canacelled*

In [17]:

plt.figure(figsize **=** (6,4))

plt.title("Reservataion status")

plt.bar(['Not cancelled', 'Cancelled'], df['is\_canceled'].value\_counts(), edgecolor **=** 'k', width **=** 0.7, color **=** 'Green' )

plt.show()



In [18]:

*## Hotel wise cancellation and non-cancellation count*

​

plt.figure(figsize **=** (8,4))

ax1 **=** sns.countplot(x**=** 'hotel', hue **=** 'is\_canceled', data**=**df, palette **=** "viridis")

legend\_labels,\_ **=** ax1.get\_legend\_handles\_labels()

ax1.legend(bbox\_to\_anchor**=**(1,1))

plt.title('Reservation status of diffrent hotels', size **=** 20)

plt.xlabel('Hotel')

plt.ylabel('Number of reservation')

plt.legend(['Not canceled', 'Canceled'])

plt.show()



**Resort\_hotel cancellation and not cancellation percentage**

In [19]:

df.head(2)

Out[19]:

|  | **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** | **...** | **customer\_type** | **adr** | **required\_car\_parking\_spaces** | **total\_of\_special\_requests** | **reservation\_status** | **reservation\_status\_date** | **name** | **email** | **phone-number** | **credit\_card** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | Resort Hotel | 0 | 342 | 2015 | July | 27 | 1 | 0 | 0 | 2 | ... | Transient | 0.0 | 0 | 0 | Check-Out | 2015-07-01 | Ernest Barnes | Ernest.Barnes31@outlook.com | 669-792-1661 | \*\*\*\*\*\*\*\*\*\*\*\*4322 |
| **1** | Resort Hotel | 0 | 737 | 2015 | July | 27 | 1 | 0 | 0 | 2 | ... | Transient | 0.0 | 0 | 0 | Check-Out | 2015-07-01 | Andrea Baker | Andrea\_Baker94@aol.com | 858-637-6955 | \*\*\*\*\*\*\*\*\*\*\*\*9157 |

2 rows × 34 columns

In [20]:

resort\_hotel **=** df[df['hotel']**==** 'Resort Hotel']

resort\_hotel['is\_canceled'].value\_counts(normalize **=** **True**)

Out[20]:

0 0.72025

1 0.27975

Name: is\_canceled, dtype: float64

In [21]:

city\_hotel **=**df[df['hotel']**==**'City Hotel']

city\_hotel['is\_canceled'].value\_counts(normalize **=True**)

Out[21]:

0 0.582918

1 0.417082

Name: is\_canceled, dtype: float64

**Using of GROUPBY function to identifaction of mean value of adr**

In [22]:

*## Extract month from status\_date*

​

df['month']**=**df['reservation\_status\_date'].dt.month

​

plt.figure(figsize **=** (8,4))

ax1 **=** sns.countplot(x**=** 'month', hue **=** 'is\_canceled', data**=**df, palette **=** "bright")

legend\_labels,\_ **=** ax1.get\_legend\_handles\_labels()

ax1.legend(bbox\_to\_anchor**=**(1,1))

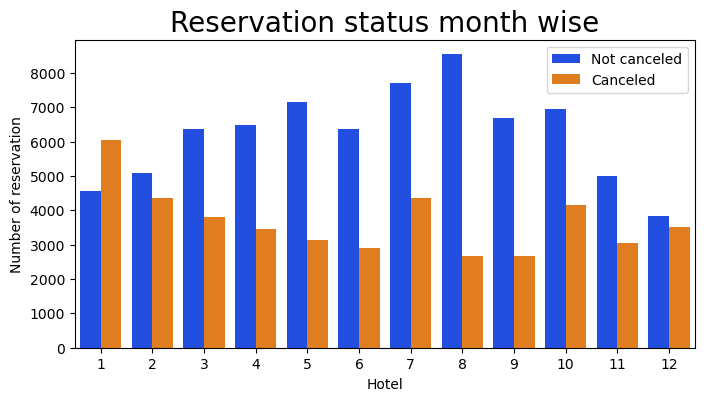
plt.title('Reservation status month wise', size **=** 20)

plt.xlabel('Hotel')

plt.ylabel('Number of reservation')

plt.legend(['Not canceled', 'Canceled'])

plt.show()



In [23]:

*## sum of cancellation of reservation monthwise*

canceled\_r**=** df[df['is\_canceled'] **==** 1].groupby('month')[['adr']].sum().reset\_index()

canceled\_r

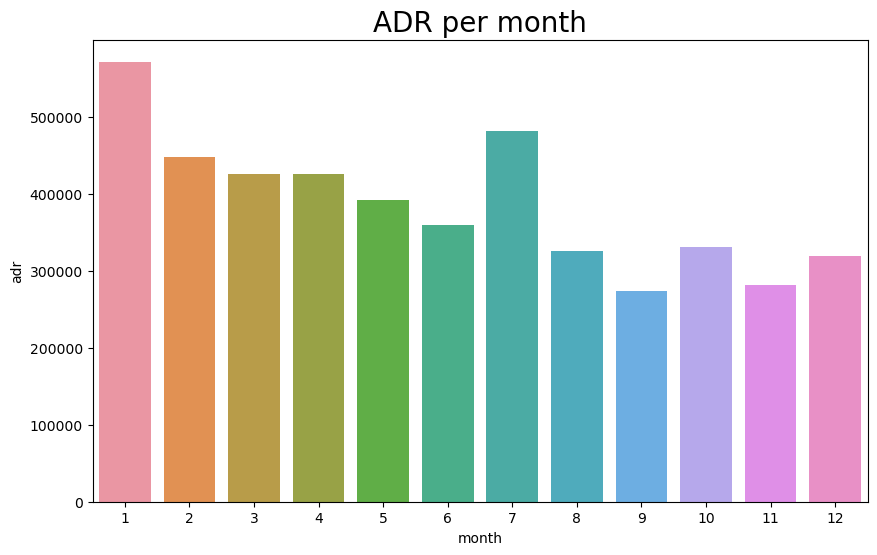
*## Ploting graphs*

plt.figure(figsize **=** (10,6))

sns.barplot(x**=**'month', y**=**'adr',data**=** canceled\_r)

plt.title('ADR per month', size **=** 20)

plt.show()



In [24]:

*## Hypothesis has been established*

*## Higher cancellation has been identified in terms of high "adr"*

**Top 10 countries**

In [38]:

canceled\_c**=** df[df['is\_canceled'] **==**1]

​

top\_10 **=** canceled\_c["country"].value\_counts()[:10]

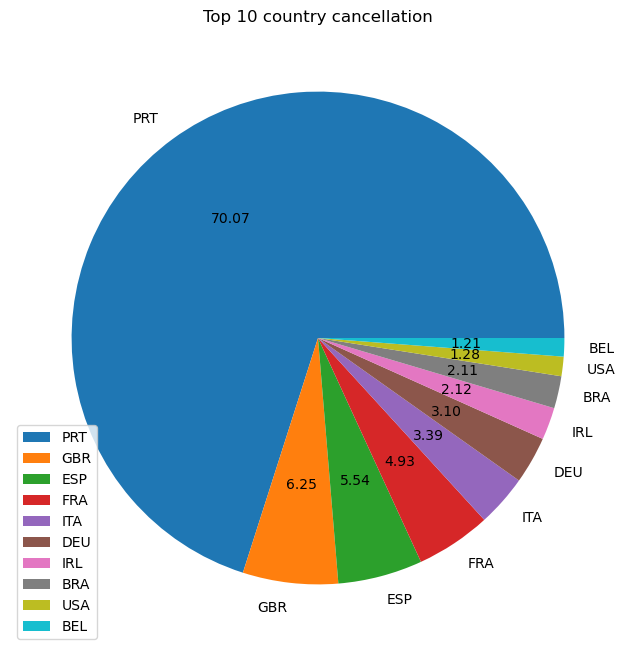
plt.figure(figsize **=** (8,8))

plt.title('Top 10 country cancellation')

plt.pie(top\_10, autopct **=** '%.2f', labels **=** top\_10.index)

plt.legend()

plt.show()



In [26]:

*## Highest cancellation is been identifed in "PRT"*

*## Around 70% cacellation loacted in "PRT"*

In [166]:

df['market\_segment'].value\_counts()

Out[166]:

Online TA 56402

Offline TA/TO 24159

Groups 19806

Direct 12448

Corporate 5111

Complementary 734

Aviation 237

Name: market\_segment, dtype: int64

In [170]:

df['market\_segment'].value\_counts(normalize**=True**)

Out[170]:

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: market\_segment, dtype: float64

## Maximum customer come through online TA

**Cancellation based on market segment**

In [48]:

mrk\_sg **=**canceled\_c['market\_segment'].value\_counts(normalize**=True**)

​

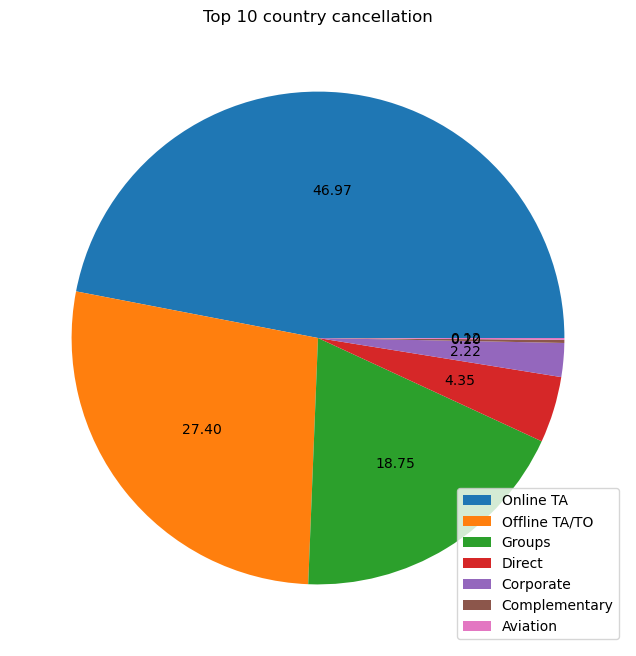
plt.figure(figsize **=** (8,8))

plt.title('Top 10 country cancellation')

plt.pie(mrk\_sg, autopct **=** '%.2f')

plt.legend(['Online TA','Offline TA/TO','Groups', 'Direct', 'Corporate', 'Complementary', 'Aviation' ])

plt.show()



In [57]:

customers\_sg **=**canceled\_c['customer\_type'].value\_counts(normalize**=True**)

customers\_sg

Out[57]:

Transient 0.825761

Transient-Party 0.144342

Contract 0.028583

Group 0.001314

Name: customer\_type, dtype: float64

In [60]:

​

​

plt.figure(figsize **=** (8,8))

plt.title('Cancellation analysis by customer segment ')

plt.pie(customers\_sg, autopct **=** '%.2f')

plt.legend(['Transient', 'Transient-Party', 'Contract', 'Group' ])

plt.show()

