In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import missingno as msno
import os
import warnings
warnings.filterwarnings('ignore')
```

In [5]:
```
# To access files and directories:
os.listdir()
```

Out[5]:
```
['.ipynb_checkpoints',
 'cleaned_data(Hotel_booking).csv',
 'Data Analysis(Hotel_Booking).ipynb',
 'hotel_booking.csv']
```

In [6]:
```
df = pd.read_csv('hotel_booking.csv')
```

In [7]:
```
df.shape
```

Out[7]:
```
(119390, 36)
```

In [8]:
```
df.head()
```

Out[8]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week
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 × 36 columns

In [9]:
```
df.tail()
```

Out[9]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_
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	

5 rows × 36 columns

In [10]:
```
df.isnull()
```

Out[10]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_
0	False	False	False	False	False	
1	False	False	False	False	False	
2	False	False	False	False	False	
3	False	False	False	False	False	
4	False	False	False	False	False	
...	...	...	...	...	...	
119385	False	False	False	False	False	
119386	False	False	False	False	False	
119387	False	False	False	False	False	
119388	False	False	False	False	False	
119389	False	False	False	False	False	

119390 rows × 36 columns

In [77]:
```
df.isnull().sum()
```

Out[77]:
```
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 [78]:
```
pd.isnull(df).sum()
```

Out[78]:
```
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 [79]:
```
df.columns
```

Out[79]:
```
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 [11]:
```
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 [12]:
```
df['reservation_status_date'] = pd.to_datetime(df['reservation_status_date']
```

In [13]:
```
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  datetime64[ns]
 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: datetime64[ns](1), float64(4), int64(16), object(15)
memory usage: 32.8+ MB
```

In [14]:
```
df.drop(['name','agent','company'],axis = 1 ,inplace = True)
df.dropna(inplace = True)
```

In [85]:
```
pd.isnull(df).sum()
```

Out[85]:
```
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
email                           0
phone-number                    0
credit_card                     0
dtype: int64
```

In [15]:
```
df.isnull().any()
```

Out[15]:
```
hotel                           False
is_canceled                     False
lead_time                       False
arrival_date_year               False
arrival_date_month              False
arrival_date_week_number        False
arrival_date_day_of_month       False
stays_in_weekend_nights         False
stays_in_week_nights            False
adults                          False
children                        False
babies                          False
meal                            False
country                         False
market_segment                  False
distribution_channel            False
is_repeated_guest               False
previous_cancellations          False
previous_bookings_not_canceled  False
reserved_room_type              False
assigned_room_type              False
booking_changes                 False
deposit_type                    False
days_in_waiting_list            False
customer_type                   False
adr                             False
required_car_parking_spaces     False
total_of_special_requests       False
reservation_status              False
reservation_status_date         False
email                           False
phone-number                    False
credit_card                     False
dtype: bool
```

In [22]:
```
df.isnull().all()
```

Out[22]:
```
hotel                           False
is_canceled                     False
lead_time                       False
arrival_date_year               False
arrival_date_month              False
arrival_date_week_number        False
arrival_date_day_of_month       False
stays_in_weekend_nights         False
stays_in_week_nights            False
adults                          False
children                        False
babies                          False
meal                            False
country                         False
market_segment                  False
distribution_channel            False
is_repeated_guest               False
previous_cancellations          False
previous_bookings_not_canceled  False
reserved_room_type              False
assigned_room_type              False
booking_changes                 False
deposit_type                    False
days_in_waiting_list            False
customer_type                   False
adr                             False
required_car_parking_spaces     False
total_of_special_requests       False
reservation_status              False
reservation_status_date         False
email                           False
phone-number                    False
credit_card                     False
dtype: bool
```

In [16]:
```
df.describe()
```

Out[16]:

	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival
count	118898.000000	118898.000000	118898.000000	118898.000000	
mean	0.371352	104.311435	2016.157656	27.166555	
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	
std	0.483168	106.903309	0.707459	13.589971	

In [17]:
```
df.describe(include = 'object')
```

Out[17]:

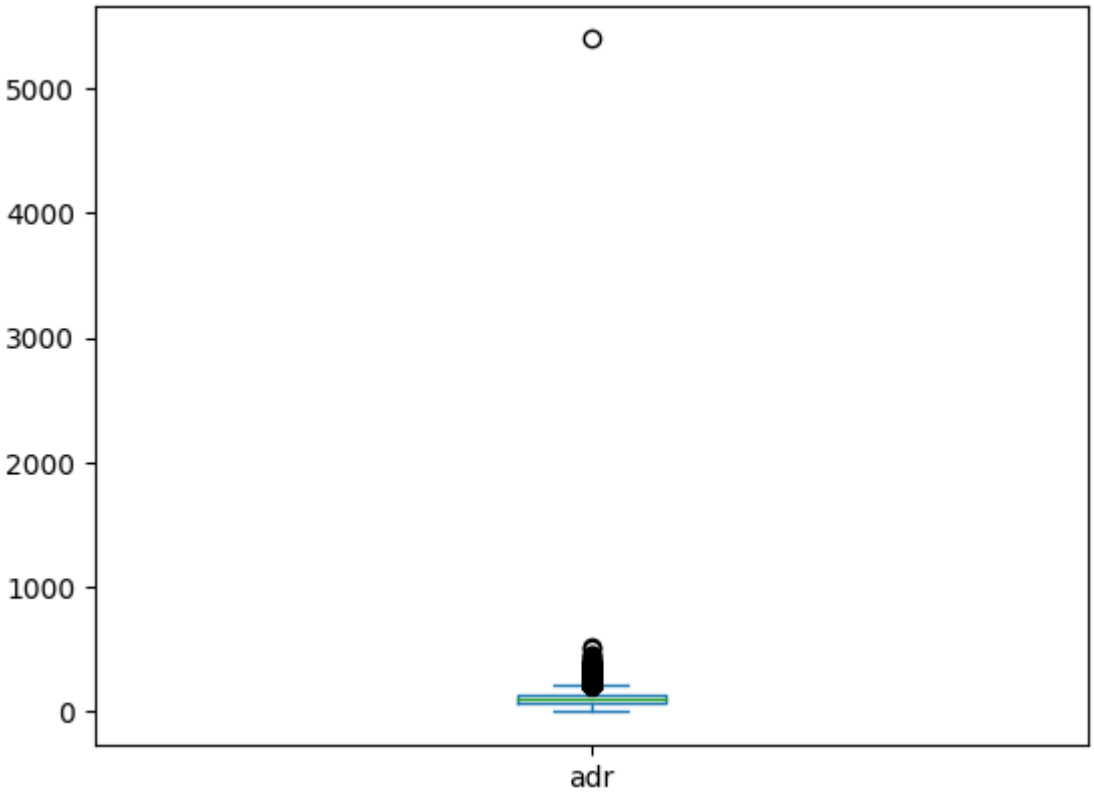
	hotel	arrival_date_month	meal	country	market_segment	distribution_chanr
count	118898	118898	118898	118898	118898	1188
unique	2	12	5	177	7	
top	City Hotel	August	BB	PRT	Online TA	TA/T
freq	79302	13852	91863	48586	56402	977

In [26]:

```
#name of each column that is of type object, followed by the unique values f
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' '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'
 'Aviation']
distribution_channel
['Direct' 'Corporate' 'TA/TO' 'Undefined' 'GDS']
reserved_room_type
['C' 'A' 'D' 'E' 'G' 'F' 'H' 'L' 'B' 'P']
assigned_room_type
['C' 'A' 'D' 'E' 'G' 'F' 'I' 'B' 'H' 'L' 'K' 'P']
deposit_type
['No Deposit' 'Refundable' 'Non Refund']
customer_type
['Transient' 'Contract' 'Transient-Party' 'Group']
reservation_status
['Check-Out' 'Canceled' 'No-Show']
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 [18]:
```
# Plot a box plot of the adr column
df['adr'].plot(kind='box')
plt.show()
```



In [19]:
```
df=df[df['adr']<5000]
```

In [20]:
```
df['adr'].plot(kind = 'box')
```

Out[20]: <Axes: >

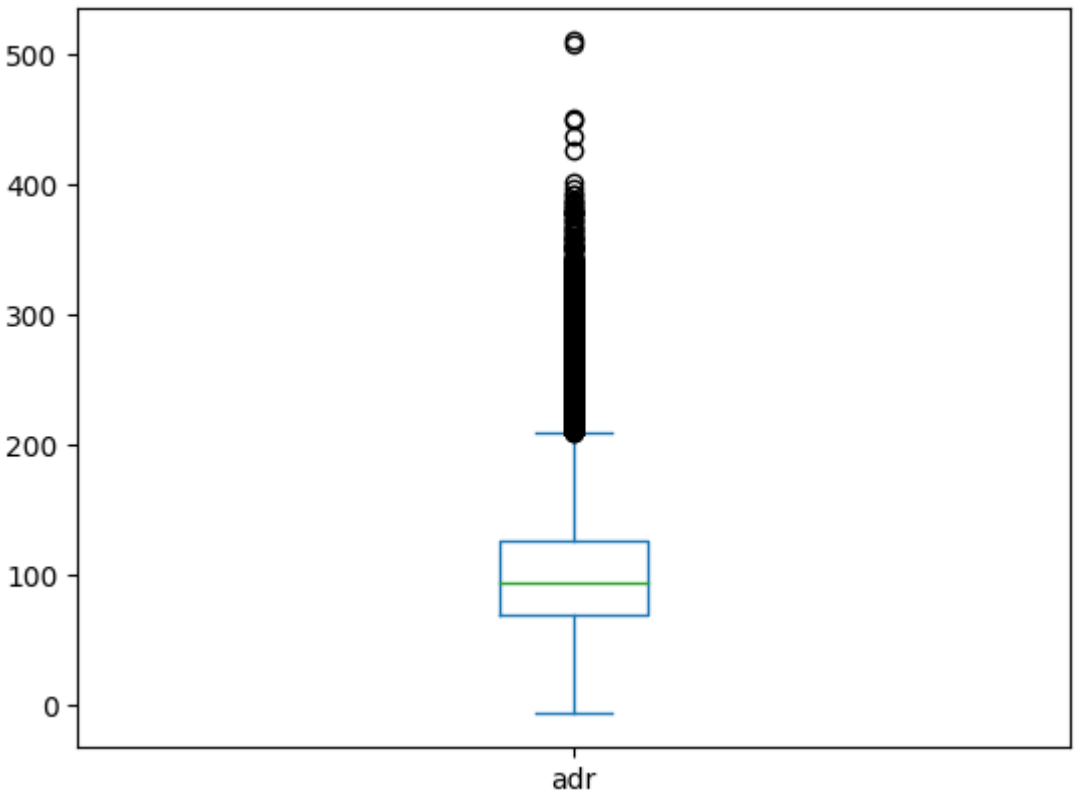


Table of Contents

EDA

* From where the most guests are coming ?
* How much do guests pay for a room per night?
* How does the price vary per night over the year?
* Which are the most busy months?

In [21]:
```
# Get the top 5 countries where the guests are coming from
top_5_countries = df["country"].value_counts().head(5)

# Print the top 5 countries
print(top_5_countries)
```

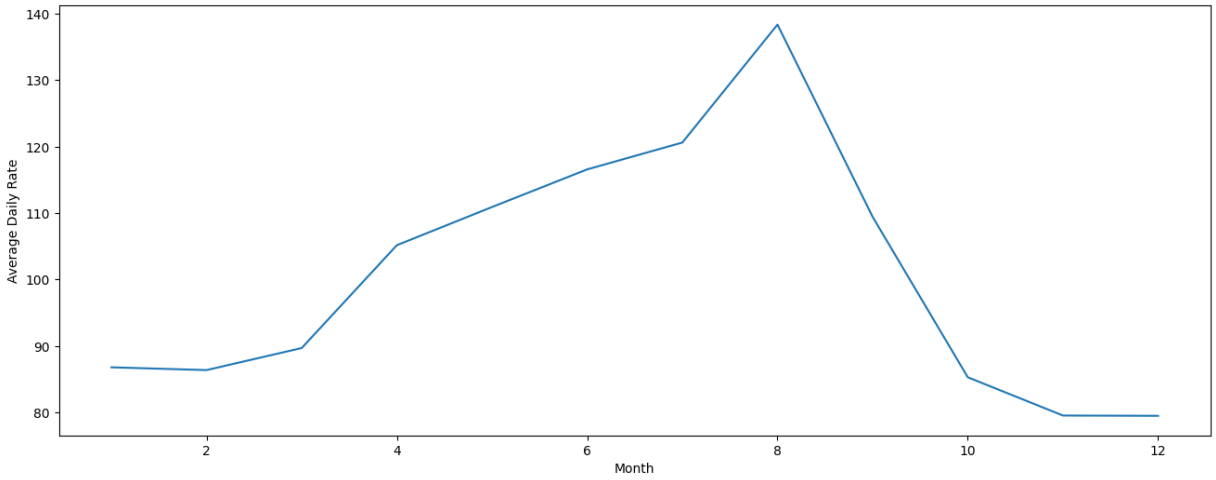
```
country
PRT    48585
GBR    12129
FRA    10415
ESP     8568
DEU     7287
Name: count, dtype: int64
```

In [27]:
```
# Calculate the average daily rate (ADR)
adr = df["adr"].mean()

# Print the ADR
print(adr)
```

```
101.9586829777034
```

In [43]:
```
# Create a new column in the df DataFrame called month.
# The month column contains the month of the reservation status date.
df['month'] = df['reservation_status_date'].dt.month
```

In [42]:
```
# Calculate the average daily rate (ADR) for each month
monthly_adr = df.groupby("month")["adr"].mean()

# Plot the monthly ADR
plt.figure(figsize=(16, 6))
plt.plot(monthly_adr.index, monthly_adr.values)
plt.xlabel("Month")
plt.ylabel("Average Daily Rate")

plt.show()
```



In [41]:
```
# Get the month with the most bookings
most_busy_month = df["month"].value_counts().sort_values(ascending=False).ir

# Print the most busy month
print(most_busy_month)
```

```
7
```

Research Question

- What are the variables that affect hotel reservation cancellations?
- how can we make hotel reservation cancellation better?
- how will be hotel be assisted in making pricing and promotions decisions ?

Hypothesis

- More cancellations occur when price are higher .
- When there is longer waiting list, customer tend to cancel more frequently.
- The majority of cliens are coming from online travel agents to make thier reservations.

Analysis and Finding

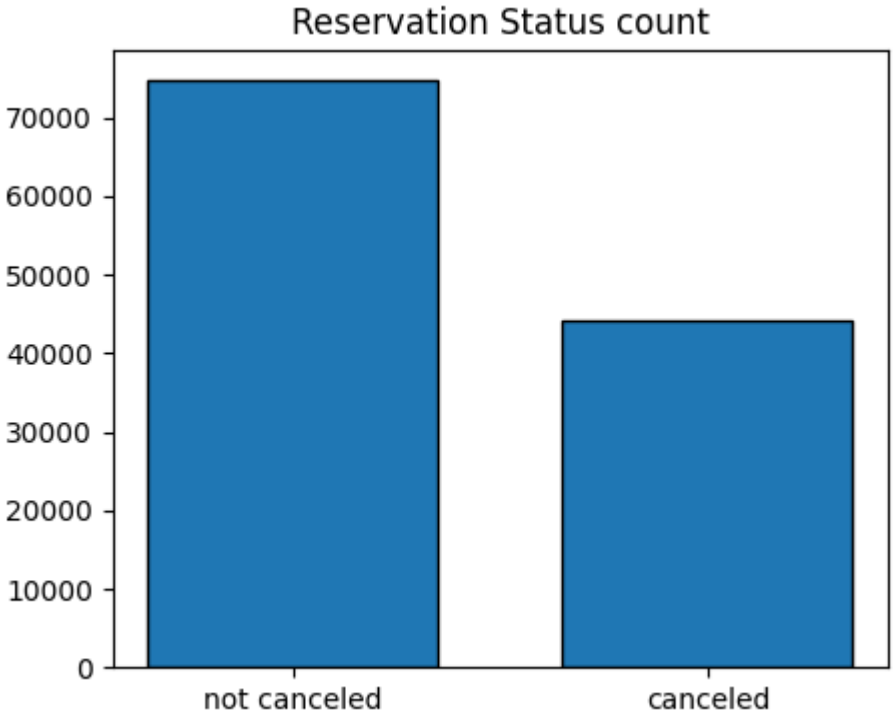
In [91]:
```
cancelled_perc = df['is_canceled'].value_counts()
print(cancelled_perc)
```

```
is_canceled
0    74745
1    44152
Name: count, dtype: int64
```

In [92]:
```
cancelled_perc = df['is_canceled'].value_counts(normalize = True)
print(cancelled_perc)
```

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

In [28]:
```
cancelled_perc = df["is_canceled"].value_counts(normalize=True)

# Plot the bar graph
plt.figure(figsize=(5, 4))
plt.title("Reservation Status count")
plt.bar(["not canceled", "canceled"], df["is_canceled"].value_counts(), edge
plt.show()

# Print the percentage of canceled bookings
print(cancelled_perc)
```

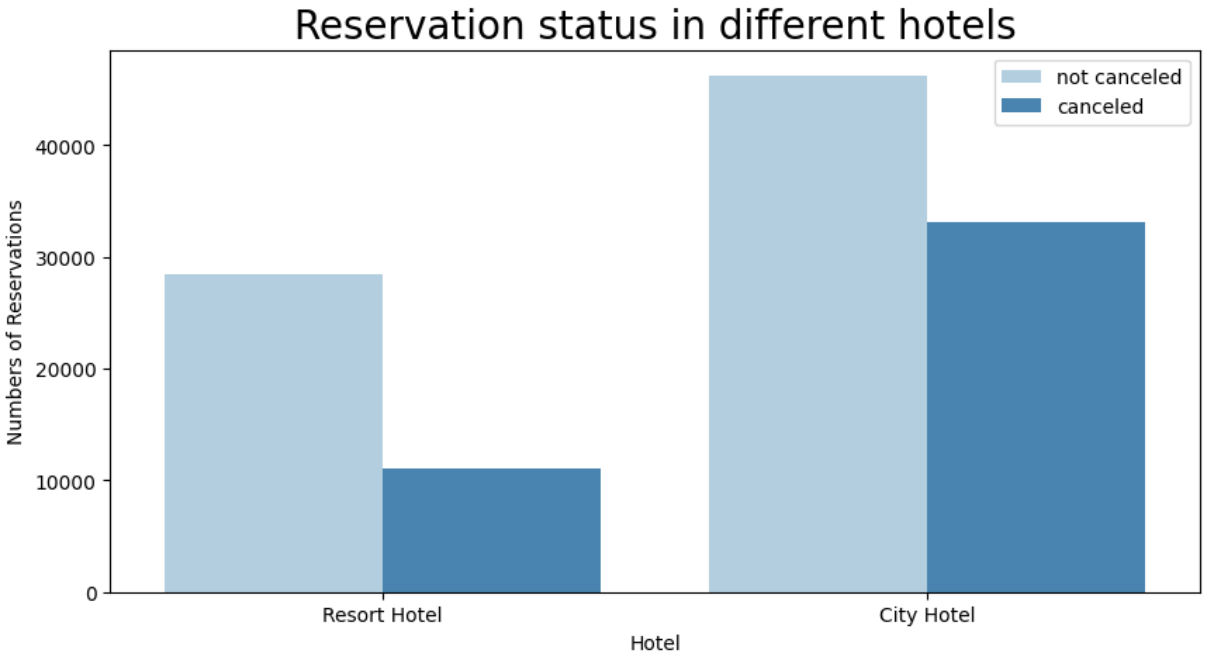



```
is_canceled
0    0.628653
1    0.371347
Name: proportion, dtype: float64
```

The accompanying bar graph shows that the percentage of reservation that are cancelled and those that are not. It is obvious that there are still a significant number og reservations that have not been cancelled . There are still 37% of clients who canceled thier reservations ,which has significant impact on the hotel's earnings.

In [30]:
```
# Create a figure with a size of 10x5 inches
plt.figure(figsize=(10, 5))

# Create a countplot of the hotel column, with the is_canceled column as the
ax1 = sns.countplot(x='hotel', hue='is_canceled', data=df, palette='Blues')

ndles and labels for the legend
```

```
legend_labels, _ = ax1.get_legend_handles_labels()

# Add the legend to the plot
ax1.legend(legend_labels, bbox_to_anchor=(1, 1))

# Add a title, x-axis label, and y-axis label to the plot
plt.title('Reservation status in different hotels', size=20)
plt.xlabel('Hotel')
plt.ylabel('Numbers of Reservations')

# Set the legend labels
plt.legend(['not canceled', 'canceled'])

# Show the plot
plt.show()
```



In comprision to resort hotels,city hotel have more booking . It's possible that resort hotels are more expensive than those in cities.

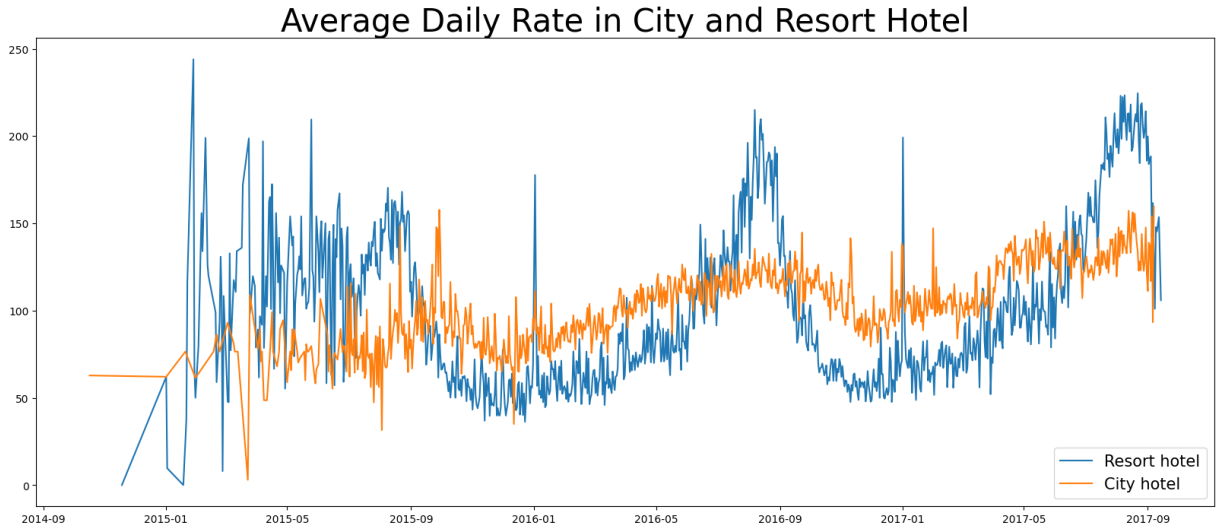
In [31]:
```
# code to calculate the percentage of canceled bookings for Resort Hotel

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

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

In [147…
```
# code to calculate the percentage of canceled bookings for  Hotel

city_hotel = df[df["hotel"] == "City Hotel"]
cancelled_perc = city_hotel["is_canceled"].value_counts(normalize=True)
print(cancelled_perc)
```

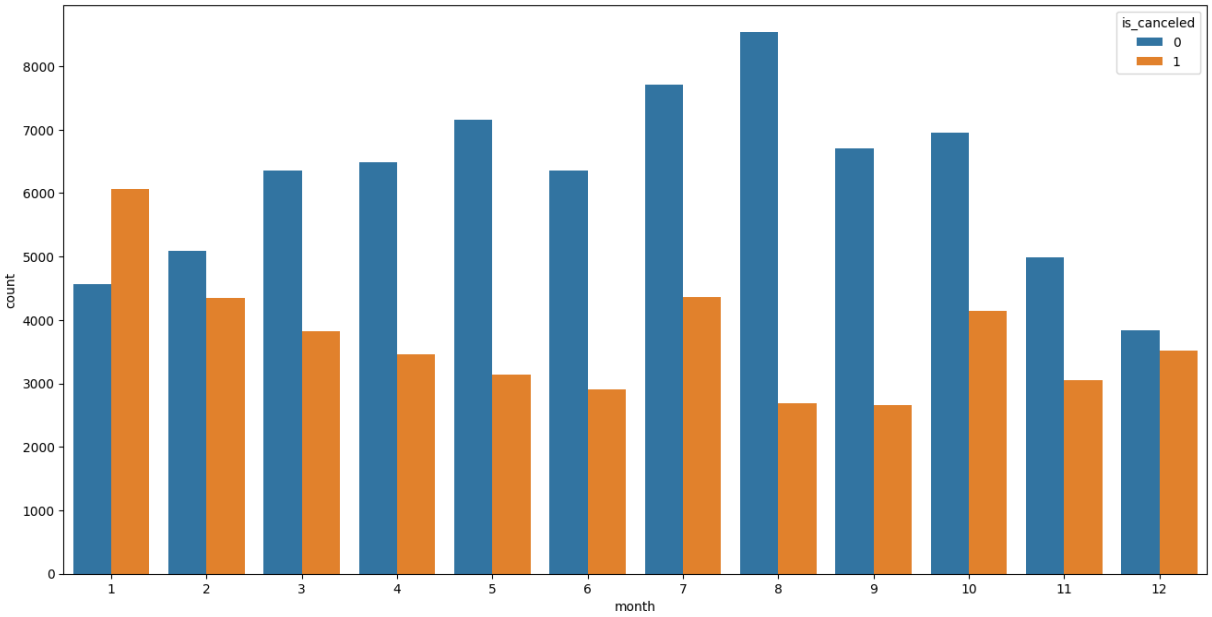
```
is_canceled
0    0.582918
1    0.417082
Name: proportion, dtype: float64
```

In [32]:
```
resort_hotel = df[df["hotel"] == "Resort Hotel"]
city_hotel = df[df["hotel"] == "City Hotel"]

resort_hotel = resort_hotel.groupby("reservation_status_date")[["adr"]].mean
city_hotel = city_hotel.groupby("reservation_status_date")[["adr"]].mean()

plt.figure(figsize=(20, 8))
plt.title("Average Daily Rate in City and Resort Hotel", fontsize=30)
plt.plot(resort_hotel.index, resort_hotel["adr"], label="Resort hotel")
plt.plot(city_hotel.index, city_hotel["adr"], label="City hotel")
plt.legend(fontsize=15)
plt.show()
```



The line Graph above shows that , on certain days, the average daily rate for a city hotel is less than that of a resort hotel, and on other days. It is even less . It goes without saying that weekends and holidays msy see a rise in resort hotel rates.

In [36]:
```
# Create a new column in the df DataFrame called month.
# The month column contains the month of the reservation status date.
df['month'] = df['reservation_status_date'].dt.month
plt.figure(figsize = (16,8))
ax1 = sns.countplot(x= 'month', hue = 'is_canceled', data = df)
plt.show()
```

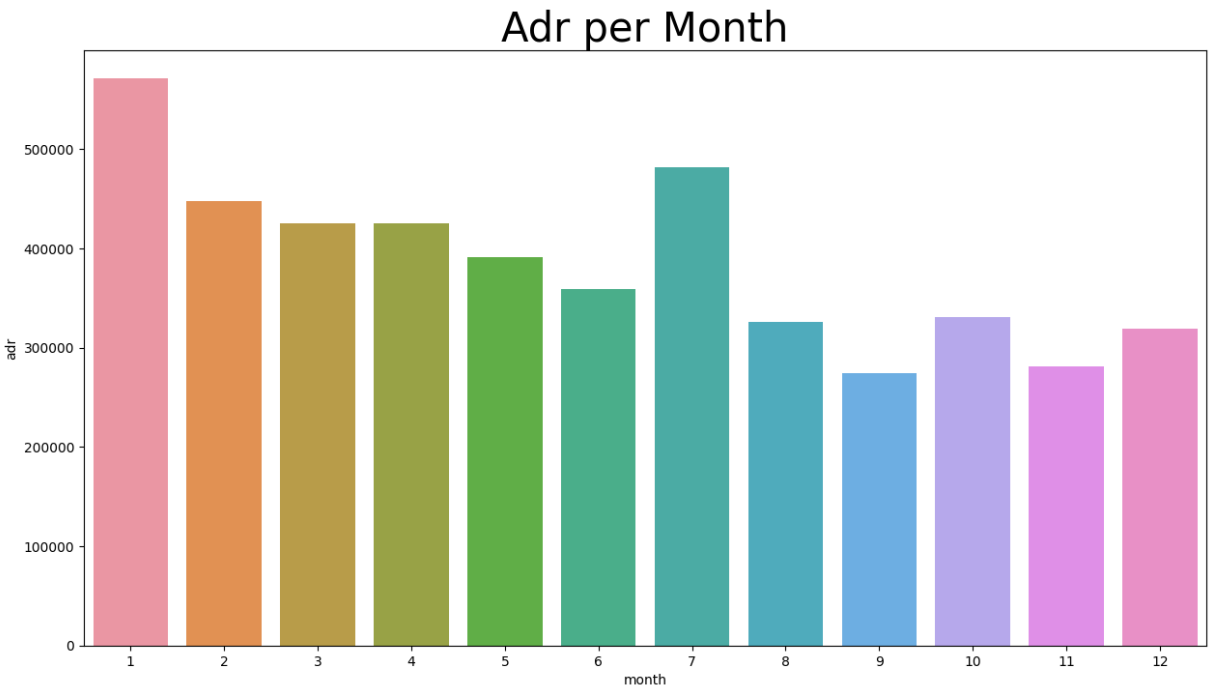


We have developed the grouped bar graph to analyze the months with highest and lowest reservation levels according to reservation status. As can v=be seen. Both the number of confirmed reservations and the number of cancelled reservation are largest in the month of august ,whereas January is the month with the most canceled reservations.

In [37]:
```
df['month'].value_counts()
```

Out[37]:
```
month
7     12074
8     11223
10    11095
1     10622
5     10294
3     10177
4      9957
2      9435
9      9359
6      9255
11     8052
12     7354
Name: count, dtype: int64
```

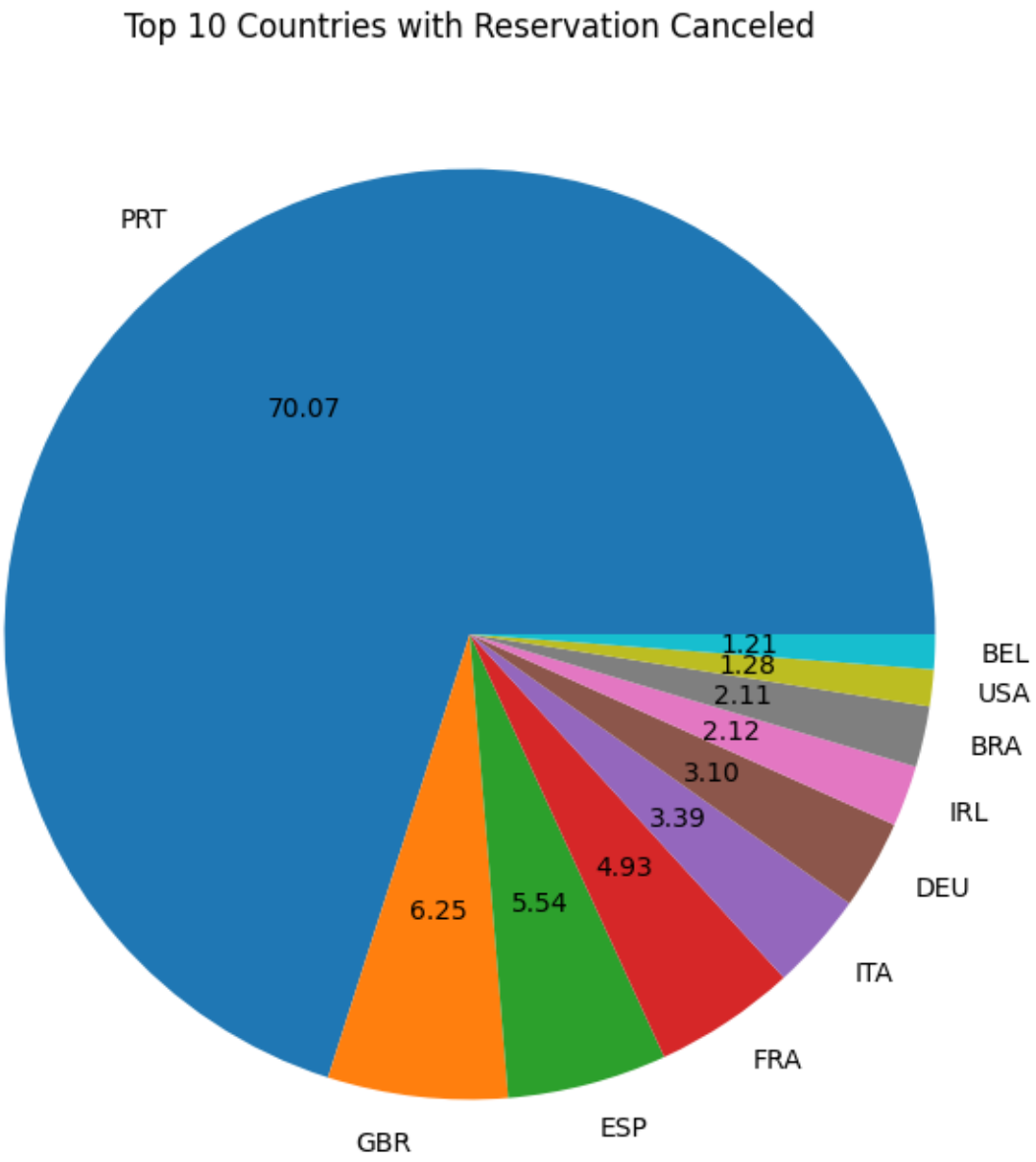
In [38]:
```
# Create a barplot of ADR per month for canceled bookings
plt.figure(figsize=(15, 8))
plt.title("Adr per Month", fontsize=30)
sns.barplot(x="month", y="adr", data=df[df["is_canceled"] == 1].groupby("mor
plt.show()
```



This bar demostates that cancellation are mosst common when prices are greatest and the least common when they are lowest. Therfore, the cost of the acccommodation is solely responsible for the cancellation.

In [141…

```
# Get the top 10 countries with canceled reservations
cancelled_data = df[df["is_canceled"] == 1]
top_10_country = cancelled_data["country"].value_counts()[:10]

# Create a pie chart of the top 10 countries
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()
```



Let's check the area from where guests are visitings the hotels and making reservations . is it comming from Direct and Groups,Online or Offline Travel Agents? Around 46% of the clients come from online a= travel agencies,whereas 27% of come from groups .Only 4% of clients book hotels directly by visiting them and making reservations.

In [135…

```
# market segment distribution for hotel bookings

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

Out[135]:
```
market_segment
Online TA        56402
Offline TA/TO    24159
Groups           19806
Direct           12448
Corporate         5111
Complementary      734
Aviation           237
Name: count, dtype: int64
```

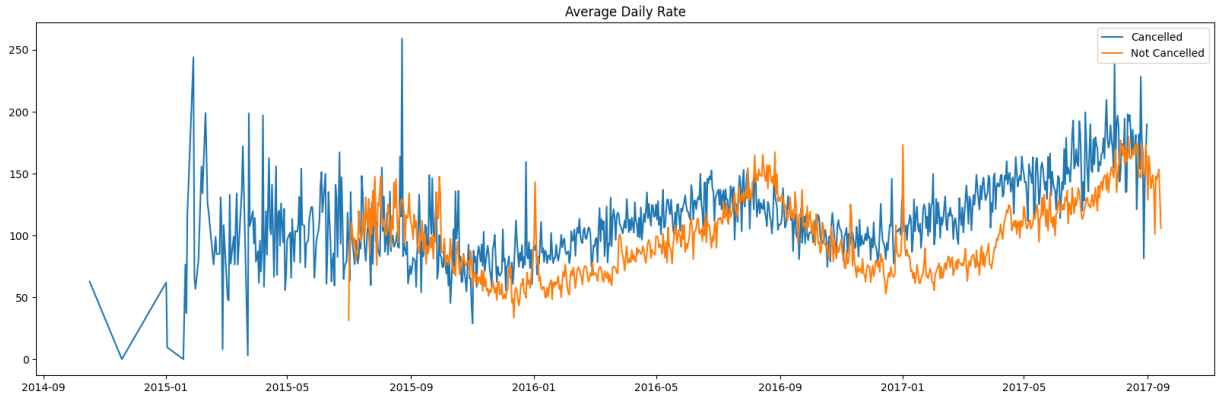
In [136…
```
# % of market segment distribution for hotel bookings
df['market_segment'].value_counts(normalize = True)
```

Out[136]:
```
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
```

In [142…
```
# market segment distribution for canceled bookings
cancelled_data = df[df["is_canceled"] == 1]

market_segment_distribution = cancelled_data["market_segment"].value_counts(

print(market_segment_distribution)
```

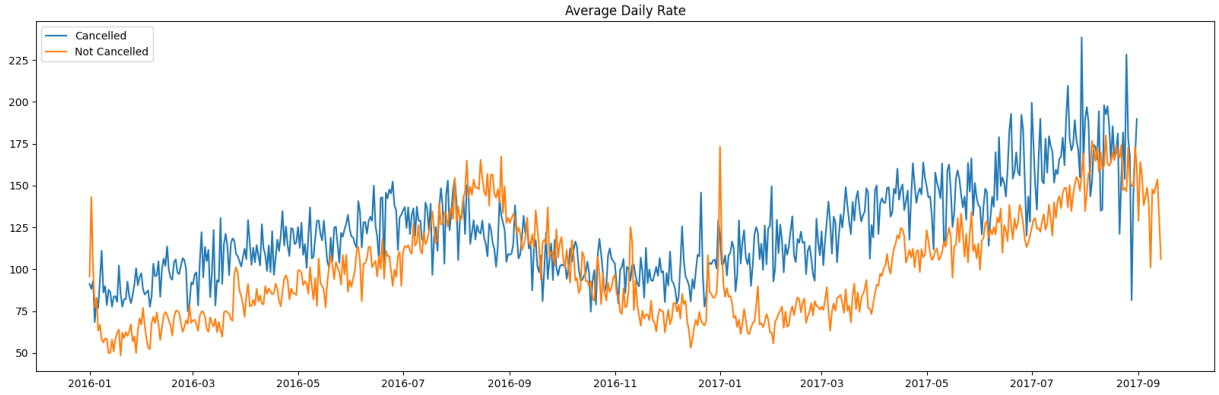
```
market_segment
Online TA        0.469696
Groups           0.273985
Offline TA/TO    0.187466
Direct           0.043486
Corporate        0.022151
Complementary    0.002038
Aviation         0.001178
Name: proportion, dtype: float64
```

In [39]:
```
# Create two DataFrames for canceled and not canceled bookings
cancelled_data = df[df["is_canceled"] == 1]
not_cancelled_data = df[df["is_canceled"] == 0]

# Calculate the average daily rate (ADR) for each DataFrame
cancelled_df_adr = cancelled_data.groupby("reservation_status_date")["adr"].
not_cancelled_df_adr = not_cancelled_data.groupby("reservation_status_date")

# Plot the ADR for each DataFrame
plt.figure(figsize=(20, 6))
plt.title("Average Daily Rate")
plt.plot(cancelled_df_adr.index, cancelled_df_adr.values, label="Cancelled")
plt.plot(not_cancelled_df_adr.index, not_cancelled_df_adr.values, label="Not
plt.legend()

plt.show()
```



In [158…

```
# Set the start and end dates for the plot
start_date = "2016-01-01"
end_date = "2017-09-30"

# Filter the data for the specified dates
cancelled_data = cancelled_data[cancelled_data["reservation_status_date"] >=
cancelled_data = cancelled_data[cancelled_data["reservation_status_date"] <=
not_cancelled_data = not_cancelled_data[not_cancelled_data["reservation_stat
not_cancelled_data = not_cancelled_data[not_cancelled_data["reservation_stat

# Calculate the average daily rate (ADR) for each DataFrame
cancelled_df_adr = cancelled_data.groupby("reservation_status_date")["adr"].
not_cancelled_df_adr = not_cancelled_data.groupby("reservation_status_date")

# Plot the ADR for each DataFrame
plt.figure(figsize=(20, 6))
plt.title("Average Daily Rate")
plt.plot(cancelled_df_adr.index, cancelled_df_adr.values, label="Cancelled")
plt.plot(not_cancelled_df_adr.index, not_cancelled_df_adr.values, label="Not
plt.legend()

plt.show()
```



As seen in the graph, reservations are canceled when the average daily rate is higher than when it is not cancelled.It clearly proves all the above analysis,that the higher price leads to higher cancellations.

Suggestions

1. Cancellation rates rise as the price does. In order to prevent canccellations of reservations, hotels could work on thier stragies and try ti lower the specific hotel based on locations. They can also provide some discount to the consumers.
2. As the ratio of the cancellation and not cancellations of the resort hotel is higher in the resort hotel than the city hotels. So the hotels should provide a reasonable discount on the room prices on weekend or on holidays.
3. In the month of january , hotel can starts campaigns or marketing with a resonable amount to increase thier revenue as the cancellation is the highest in this month.
4. They can also increase the quality of thier hotels and thier servies mainly in portugal to reduce the cancellation rate.