Out[4]:

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

5 rows × 32 columns

In [5]:
```
df.shape
```

Out[5]:
```
(119390, 32)
```

In [7]:
```
#df.drop(['company','agent','Unnamed: 32'],axis=1,inplace=True)
```

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

Out[8]:
```
hotel                               2
is_canceled                         2
lead_time                         479
arrival_date_year                   3
arrival_date_month                 12
arrival_date_week_number           53
arrival_date_day_of_month          31
stays_in_weekend_nights            17
stays_in_week_nights               35
adults                             14
children                            5
babies                              5
meal                                5
country                           177
market_segment                      8
distribution_channel                5
is_repeated_guest                   2
previous_cancellations             15
previous_bookings_not_canceled     73
reserved_room_type                 10
assigned_room_type                 12
booking_changes                    21
deposit_type                        3
agent                             333
company                           352
days_in_waiting_list              128
customer_type                       4
adr                              8879
required_car_parking_spaces         5
total_of_special_requests           6
reservation_status                  3
reservation_status_date           926
dtype: int64
```

In [9]:
```
df.loc[df['reserved_room_type']==df['assigned_room_type'],'room_assigned']='Desired'
df.loc[df['reserved_room_type']!=df['assigned_room_type'],'room_assigned']='Undesired'
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 33 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  room_assigned                   119390 non-null  object 
dtypes: float64(4), int64(16), object(13)
memory usage: 30.1+ MB
```

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                             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
room_assigned                        0
dtype: int64
```

In [12]:
```
df.dropna(inplace= True)
```

In [13]:
```
df.describe()
```

Out[13]:

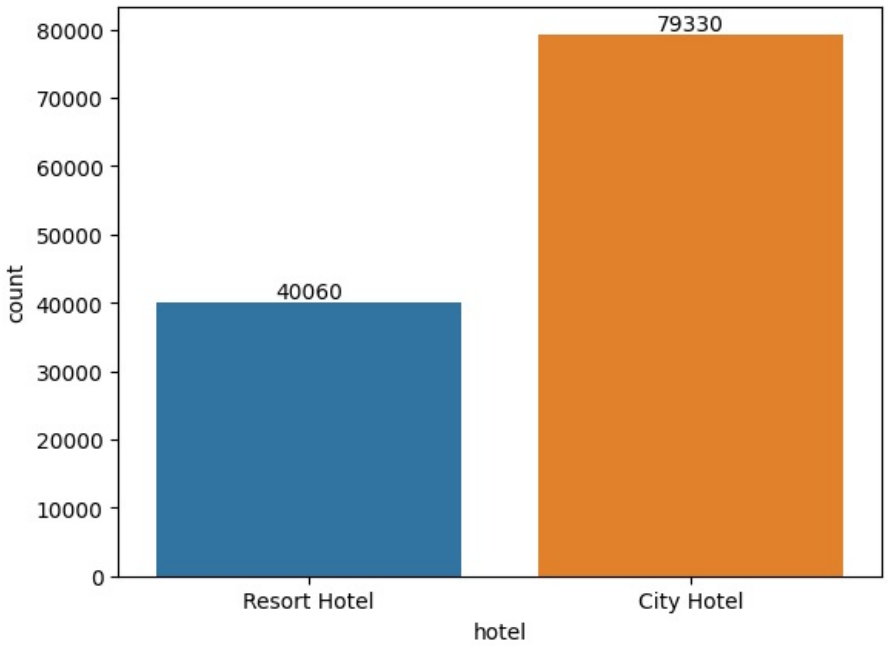
| | is_canceled | lead_time | arrival_date_year | arrival_date_week_number | arrival_date_day_of_month | stays_in_weekend_nights |
|---|---|---|---|---|---|---|
| **count** | 217.000000 | 217.000000 | 217.000000 | 217.000000 | 217.000000 | 217.00000 |
| **mean** | 0.078341 | 40.520737 | 2015.465438 | 38.198157 | 10.824885 | 1.56682 |
| **std** | 0.269329 | 61.748375 | 0.720053 | 12.890292 | 7.582065 | 1.49270 |
| **min** | 0.000000 | 0.000000 | 2015.000000 | 1.000000 | 1.000000 | 0.00000 |
| **25%** | 0.000000 | 12.000000 | 2015.000000 | 33.000000 | 6.000000 | 0.00000 |
| **50%** | 0.000000 | 27.000000 | 2015.000000 | 45.000000 | 9.000000 | 2.00000 |
| **75%** | 0.000000 | 36.000000 | 2016.000000 | 46.000000 | 13.000000 | 2.00000 |
| **max** | 1.000000 | 364.000000 | 2017.000000 | 53.000000 | 31.000000 | 9.00000 |

In [14]:
```
df['reservation_status_date']= pd.to_datetime(df['reservation_status_date'])
```

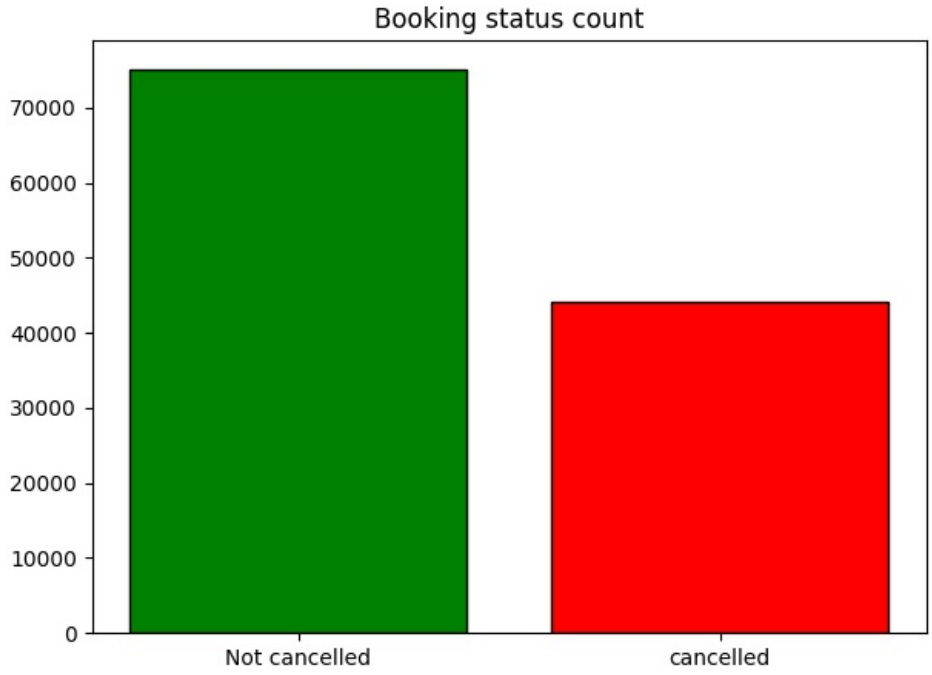
In [15]:
```
df.columns
```

Out[15]:
```
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', 'room_assigned'],
      dtype='object')
```

In [10]:
```
ax1=sns.countplot(x='hotel',data=df,hue='hotel')
for bars in ax1.containers:
    ax1.bar_label(bars)
```




In [73]:
```
cancelled_count= df['is_canceled'].value_counts()
plt.figure(figsize=(7,5))
plt.title('Booking status count')
colors=('Green','Red')
plt.bar(['Not cancelled','cancelled'],df['is_canceled'].value_counts(),edgecolor='black',color=colors)
plt.show()
```



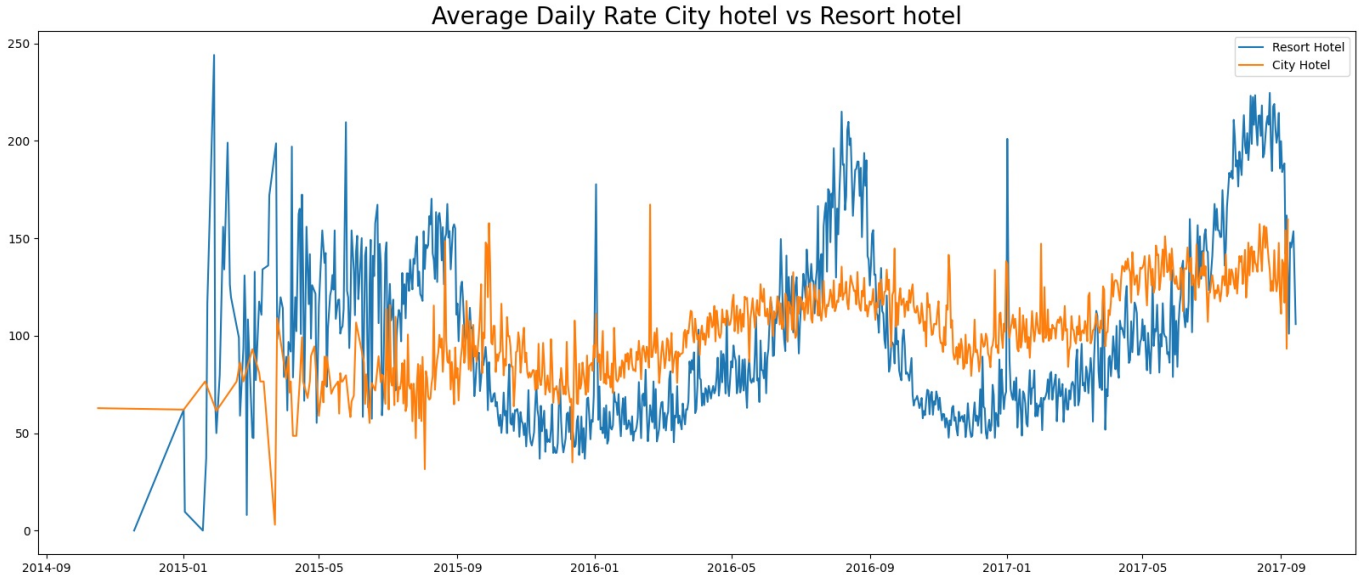

In [59]:
```
cancelled_booking= df[df['is_canceled']==1]
Not_cancelled=df[df['is_canceled']==0]
resort_hotel=df[df['hotel']=='Resort Hotel']
resort_hotel['is_canceled'].value_counts()
city_hotel=df[df['hotel']=='City Hotel']
city_hotel['is_canceled'].value_counts()
df['month']= df['reservation_status_date'].dt.month
```

In [20]:
```
ax1= sns.countplot(x= 'hotel',hue='is_canceled',data=df,palette='viridis')
plt.title('Booking status Resort hotel vs city hotel')
plt.xlabel('Hotel')
plt.ylabel('No of Booking')
```

Out[20]:
```
Text(0, 0.5, 'No of Booking')
```


In [62]:
```
resort_hotel= resort_hotel.groupby('reservation_status_date')[['adr']].mean()
city_hotel= city_hotel.groupby('reservation_status_date')[['adr']].mean()
```

In [63]:
```
plt.figure(figsize=(20,8))
plt.plot(resort_hotel.index, resort_hotel.values,label='Resort Hotel')
plt.plot(city_hotel.index, city_hotel.values,label='City Hotel')
plt.title('Average Daily Rate City hotel vs Resort hotel', size= 20)
plt.legend()
plt.show()
```

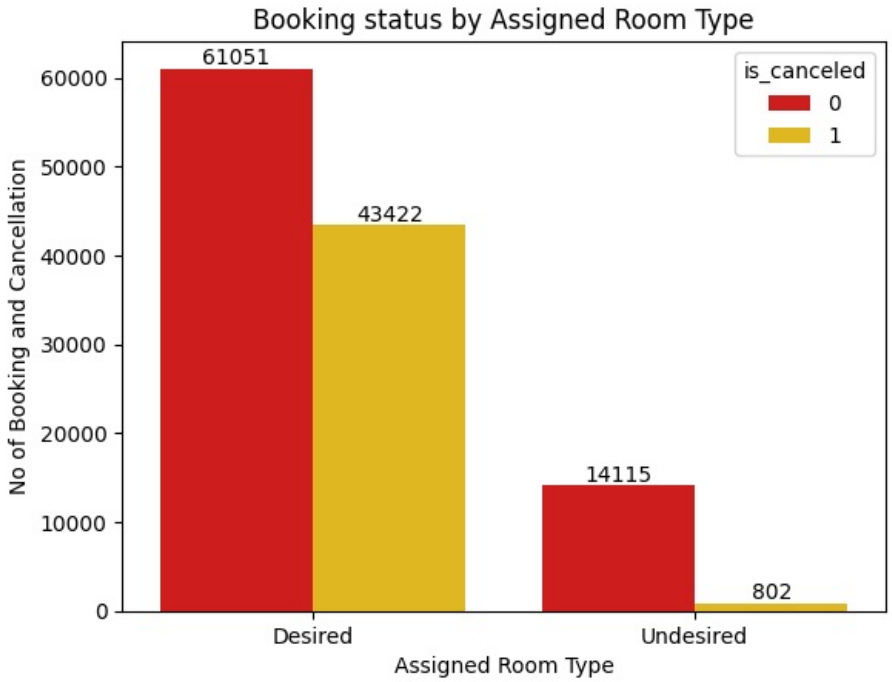



In [15]:
```
df['room_assigned'].unique()
Desired_room= df[df['room_assigned']=='Desired']
undesired_room= df[df['room_assigned']=='Undesired']
Desired_room['is_canceled'].value_counts(normalize=True)
undesired_room['is_canceled'].value_counts(normalize=True)
```

Out[15]:
```
is_canceled
0    0.946236
1    0.053764
Name: proportion, dtype: float64
```

In [17]:
```
ax= sns.countplot(x='room_assigned',hue='is_canceled', data=df,palette='hot')
for bars in ax.containers:
    ax.bar_label(bars)
plt.title('Booking status by Assigned Room Type')
plt.xlabel('Assigned Room Type')
plt.ylabel('No of Booking and Cancellation')
```

Out[17]:
```
Text(0, 0.5, 'No of Booking and Cancellation')
```




In [17]:
```
df['market_segment'].value_counts()
```

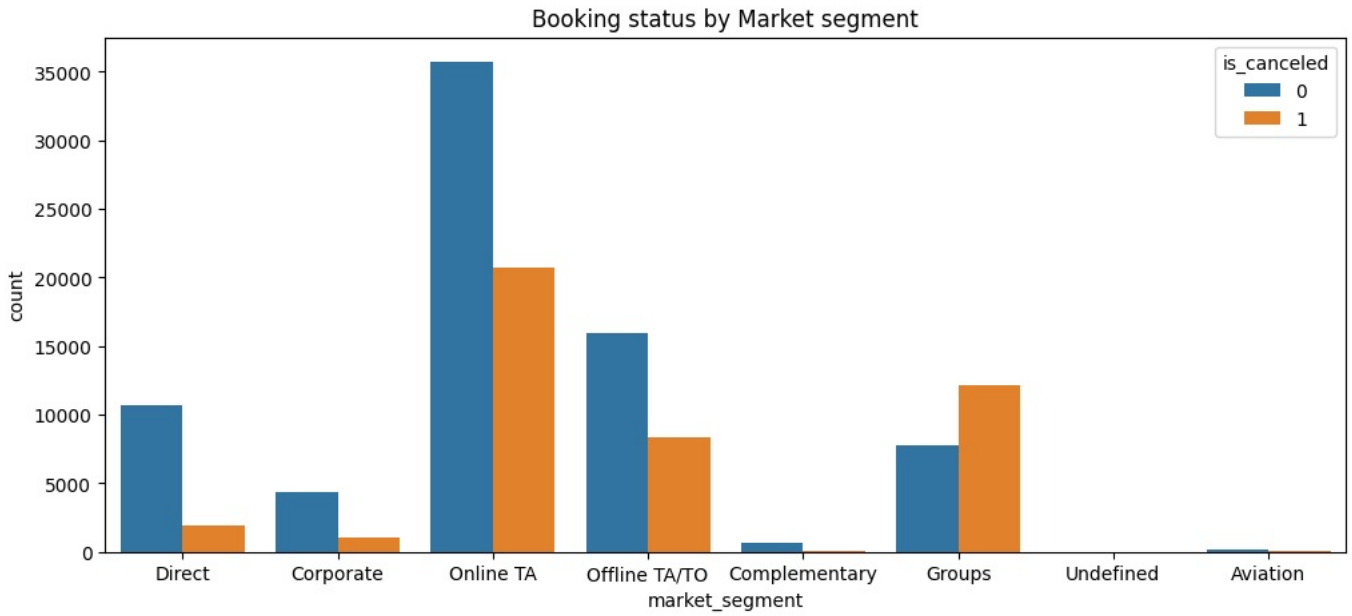
Out[17]:
```
market_segment
Online TA        56477
Offline TA/TO    24219
Groups           19811
Direct           12606
Corporate         5295
Complementary      743
Aviation           237
Undefined            2
Name: count, dtype: int64
```

In [24]:
```
cancelled_booking['market_segment'].value_counts()
```

Out[24]:
```
market_segment
Online TA        50967
Offline TA/TO    20598
Groups           17918
Direct           10468
Corporate         3805
Complementary      498
Aviation           217
Undefined            2
Name: count, dtype: int64
```

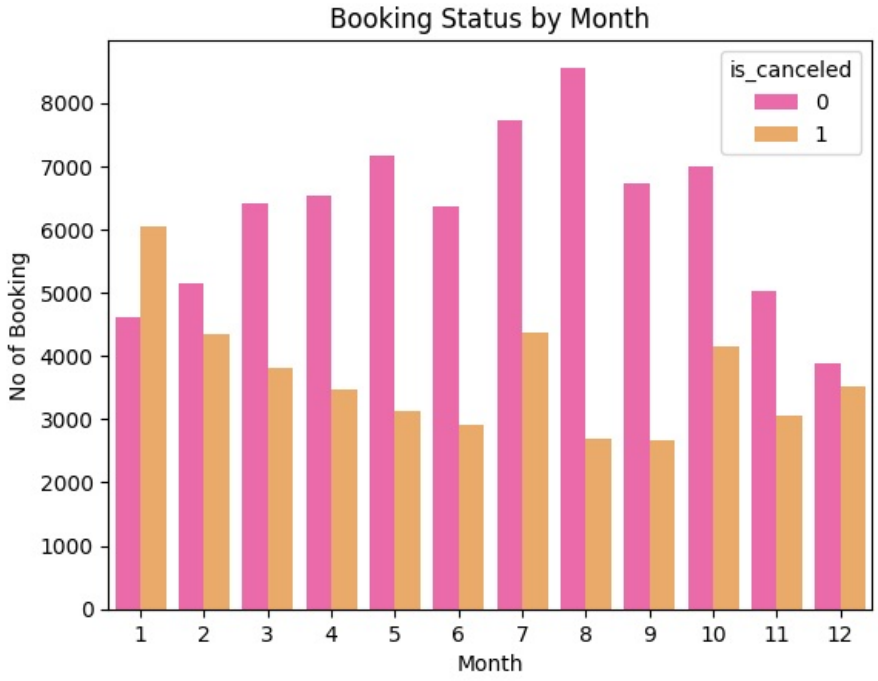
In [58]:
```
plt.figure(figsize=(12,5))
sns.countplot(x='market_segment',data=df,hue='is_canceled')
plt.title('Booking status by Market segment')
```

Out[58]:
```
Text(0.5, 1.0, 'Booking status by Market segment')
```



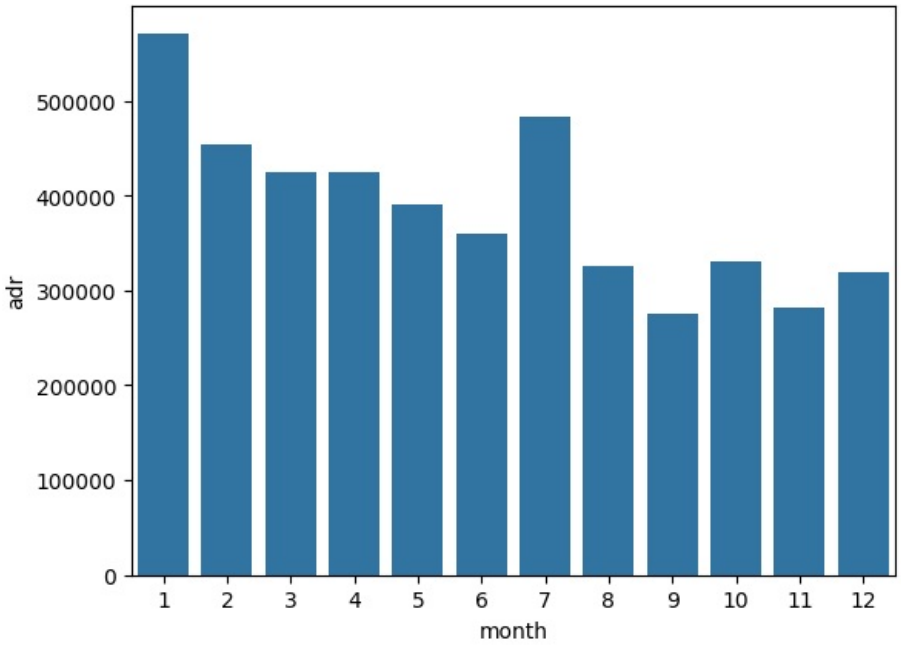

In [21]:
```
ax =sns.countplot(x='month',hue='is_canceled',data= df,palette='spring')
plt.title('Booking Status by Month')
plt.xlabel('Month')
plt.ylabel('No of Booking')
```

Out[21]:
```
Text(0, 0.5, 'No of Booking')
```

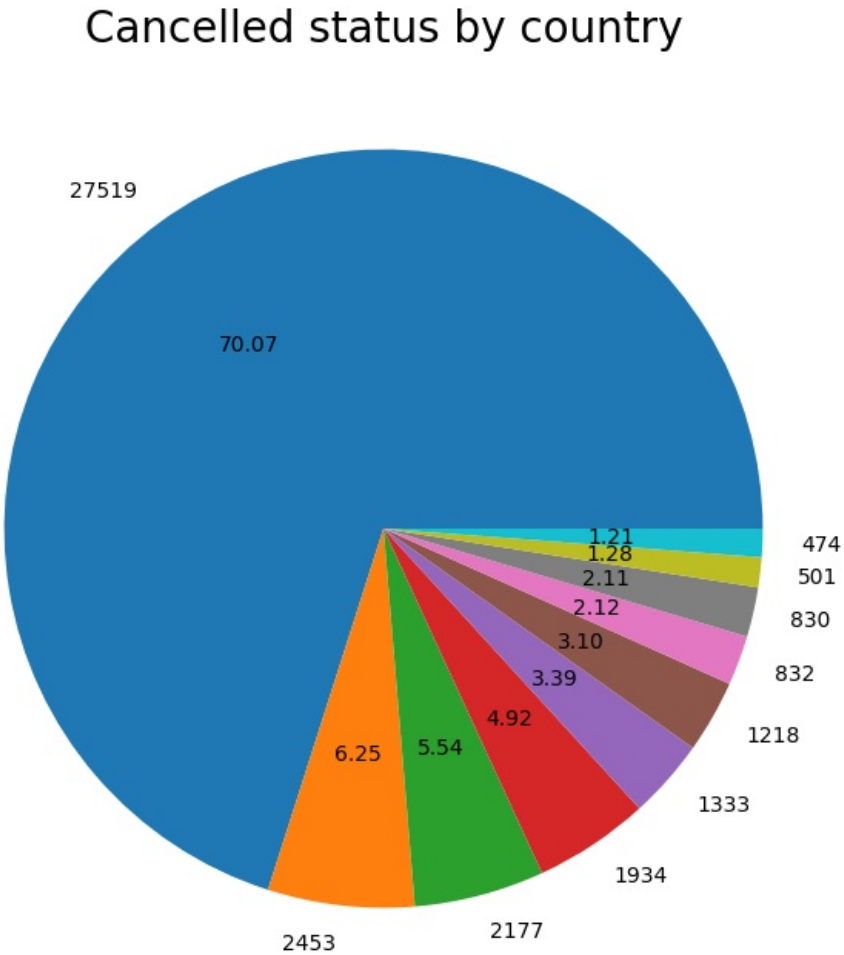



In [24]:
```
ss= cancelled_booking.groupby('month')[['adr']].sum()
ss
sns.barplot(x='month',y='adr',data= ss)
```

Out[24]:
```
<Axes: xlabel='month', ylabel='adr'>
```




In [53]:
```
cb= cancelled_booking['country'].value_counts()[:10]
plt.figure(figsize= (8,8))
plt.pie(cb,autopct= '%.2f',labels=cb)
plt.title('Cancelled status by country',size= 20)
plt.show()
```




In [ ]: