Importing Libraries

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

In [5]:
```
df=pd.read_csv(r"C:\Users\hp\Desktop\Jatin\EXCEL\Data\hotel_booking.csv")
df
```

Out[5]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	a
0	Resort Hotel	0	342	2015	July	27	
1	Resort Hotel	0	737	2015	July	27	
2	Resort Hotel	0	7	2015	July	27	
3	Resort Hotel	0	13	2015	July	27	
4	Resort Hotel	0	14	2015	July	27	
...	...	...	...	...	...	...	
119385	City Hotel	0	23	2017	August	35	
119386	City Hotel	0	102	2017	August	35	
119387	City Hotel	0	34	2017	August	35	
119388	City Hotel	0	109	2017	August	35	
119389	City Hotel	0	205	2017	August	35	

119390 rows × 36 columns

Eploratory Data Anlaysis And Data Cleaning

In [7]:
```
df.head()
```

Out[7]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival
0	Resort Hotel	0	342	2015	July	27	
1	Resort Hotel	0	737	2015	July	27	
2	Resort Hotel	0	7	2015	July	27	
3	Resort Hotel	0	13	2015	July	27	
4	Resort Hotel	0	14	2015	July	27	

5 rows × 36 columns

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

Out[8]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	ar
119385	City Hotel	0	23	2017	August	35	
119386	City Hotel	0	102	2017	August	35	
119387	City Hotel	0	34	2017	August	35	
119388	City Hotel	0	109	2017	August	35	
119389	City Hotel	0	205	2017	August	35	

5 rows × 36 columns

In [10]:
```
df.shape
```

Out[10]:
```
(119390, 36)
```

In [11]:
```
df.columns
```

Out[11]:
```
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 [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  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 [14]:
```
df['reservation_status_date']=pd.to_datetime(df['reservation_status_date'])
```

In [15]:
```
df.describe(include='object')
```

Out[15]:

	hotel	arrival_date_month	meal	country	market_segment	distribution_channel	reserved_roor
count	119390	119390	119390	118902	119390	119390	
unique	2	12	5	177	8	5	
top	City Hotel	August	BB	PRT	Online TA	TA/TO	
freq	79330	13877	92310	48590	56477	97870	

In [16]:
```
df.isnull().sum()
```

Out[16]:

```
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 [17]:
```
df.drop(['company','agent'],axis=1,inplace=True)
```

In [18]:
```
df.dropna(inplace=True)
```

In [20]:
```
df.isnull().sum()
```

Out[20]:
```
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 [21]:
```
df.describe()
```

Out[21]:

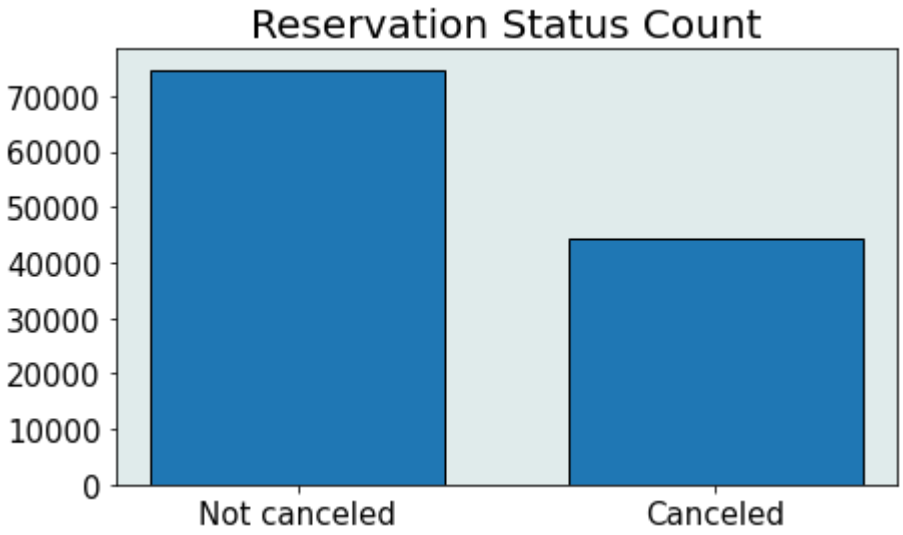
	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date_day_of_mo
count	118898.000000	118898.000000	118898.000000	118898.000000	118898.000
mean	0.371352	104.311435	2016.157656	27.166555	15.800
std	0.483168	106.903309	0.707459	13.589971	8.780
min	0.000000	0.000000	2015.000000	1.000000	1.000
25%	0.000000	18.000000	2016.000000	16.000000	8.000
50%	0.000000	69.000000	2016.000000	28.000000	16.000
75%	1.000000	161.000000	2017.000000	38.000000	23.000
max	1.000000	737.000000	2017.000000	53.000000	31.000

In [22]:
```
df=df[df['adr']<5000]
```

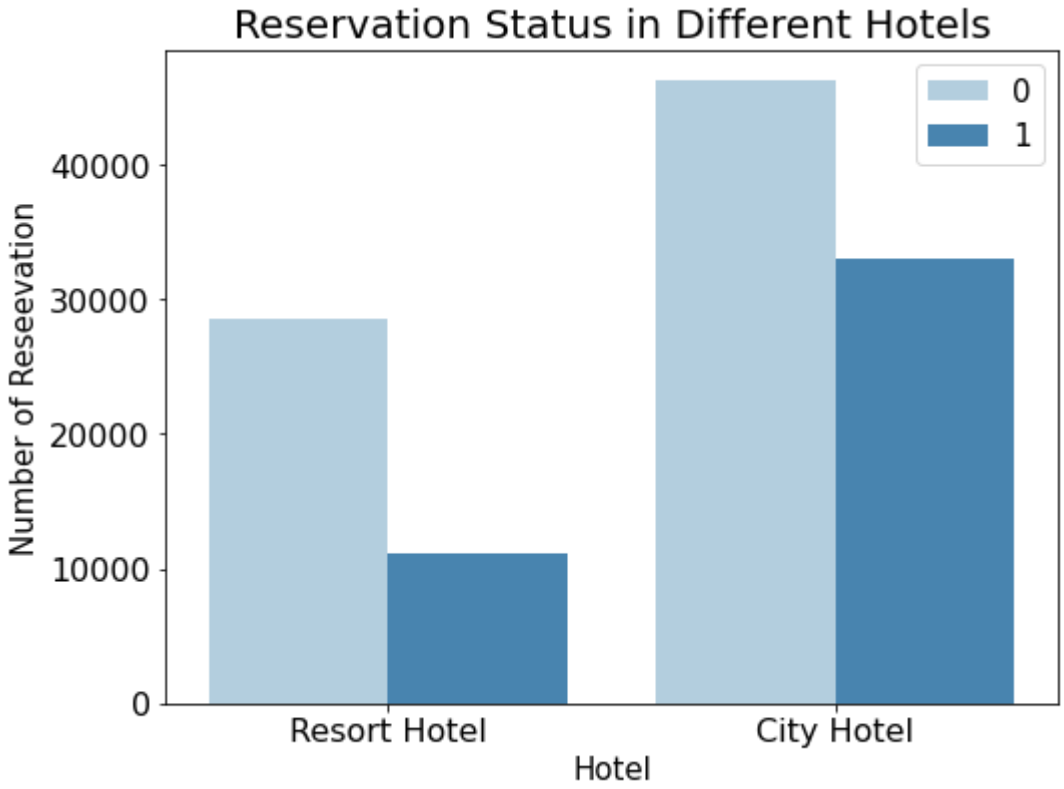
Data Analysis And Visualization

In [125]:
```
cancelled_percentage = df['is_canceled'].value_counts(normalize=True)
print(cancelled_percentage)

plt.figure(figsize=(7,4))

ax=plt.axes()
ax.set_facecolor("#e0ebeb")

plt.title("Reservation Status Count",fontsize=20)
plt.bar(['Not canceled','Canceled'],df['is_canceled'].value_counts(),edgecolor="k", w
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)
plt.show()
```

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



In [126]:
```
plt.figure(figsize=(8,6))

ax1=sns.countplot(x='hotel',hue="is_canceled",data=df,palette='Blues')
legend_labels,_=ax1.get_legend_handles_labels()

plt.title("Reservation Status in Different Hotels",size=20)
plt.xlabel("Hotel",fontsize=15)
plt.ylabel("Number of Reseevation",fontsize=15)
plt.legend(fontsize=15)

plt.xticks(fontsize=16)
plt.yticks(fontsize=16)
plt.show()
```



In [127]:
```
resort_hotel=df[df['hotel']=='Resort Hotel']
resort_hotel['is_canceled'].value_counts(normalize=True)
```

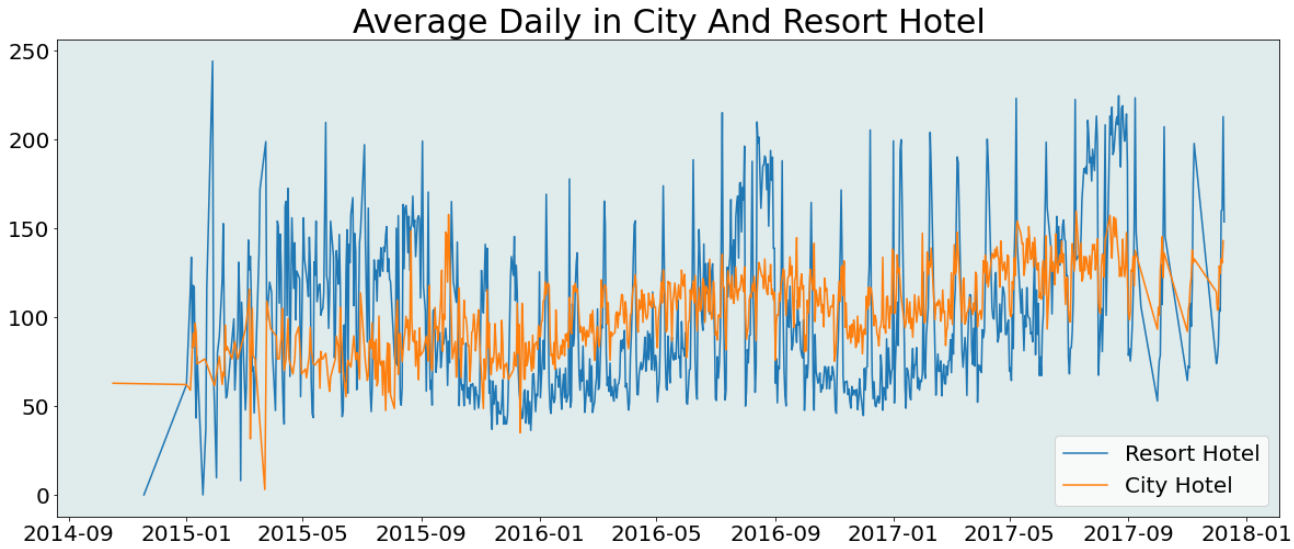
Out[127]:
```
0    0.72025
1    0.27975
Name: is_canceled, dtype: float64
```

In [128]:
```
city_hotel=df[df['hotel']=='City Hotel']
city_hotel['is_canceled'].value_counts(normalize=True)
```

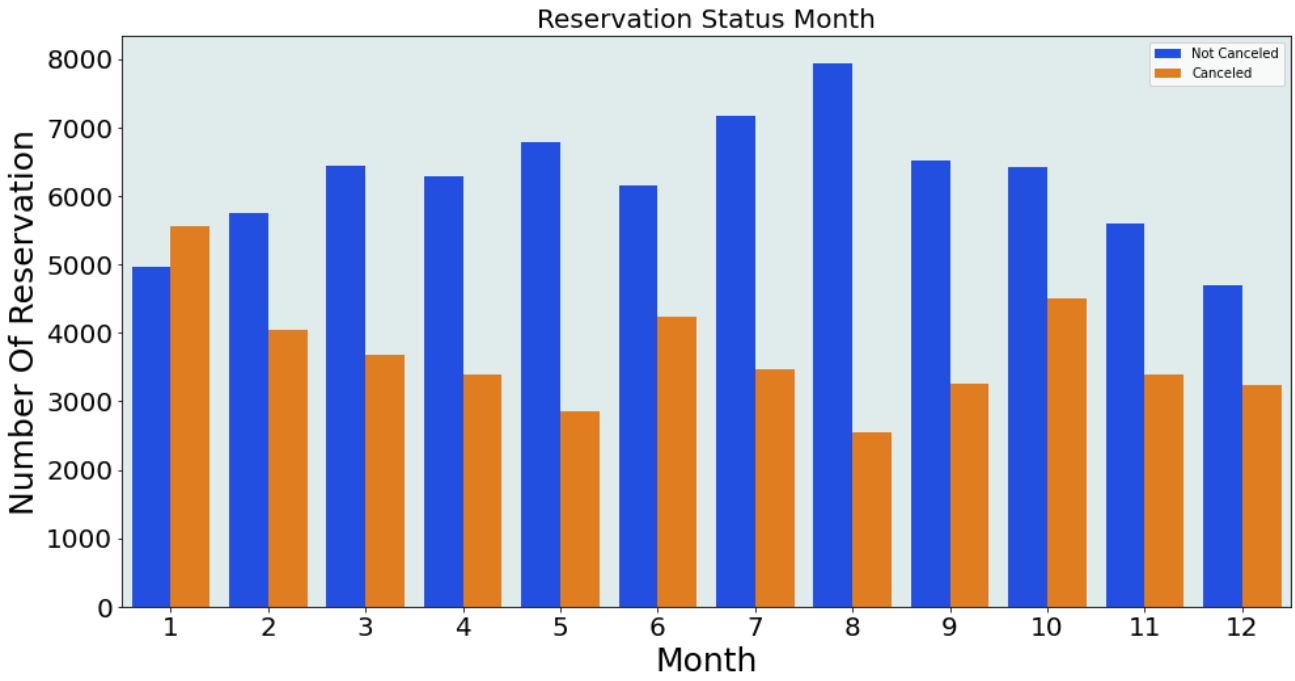
Out[128]:
```
0    0.582918
1    0.417082
Name: is_canceled, dtype: float64
```

In [129]:
```
resort_hotel=resort_hotel.groupby('reservation_status_date')[['adr']].mean()
city_hotel=city_hotel.groupby('reservation_status_date')[['adr']].mean()
```

In [130]:
```
plt.figure(figsize=(20,8))
ax=plt.axes()
ax.set_facecolor("#e0ebeb")
plt.title("Average Daily 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=20)
plt.xticks(fontsize=20)
plt.yticks(fontsize=20)
plt.show()
```

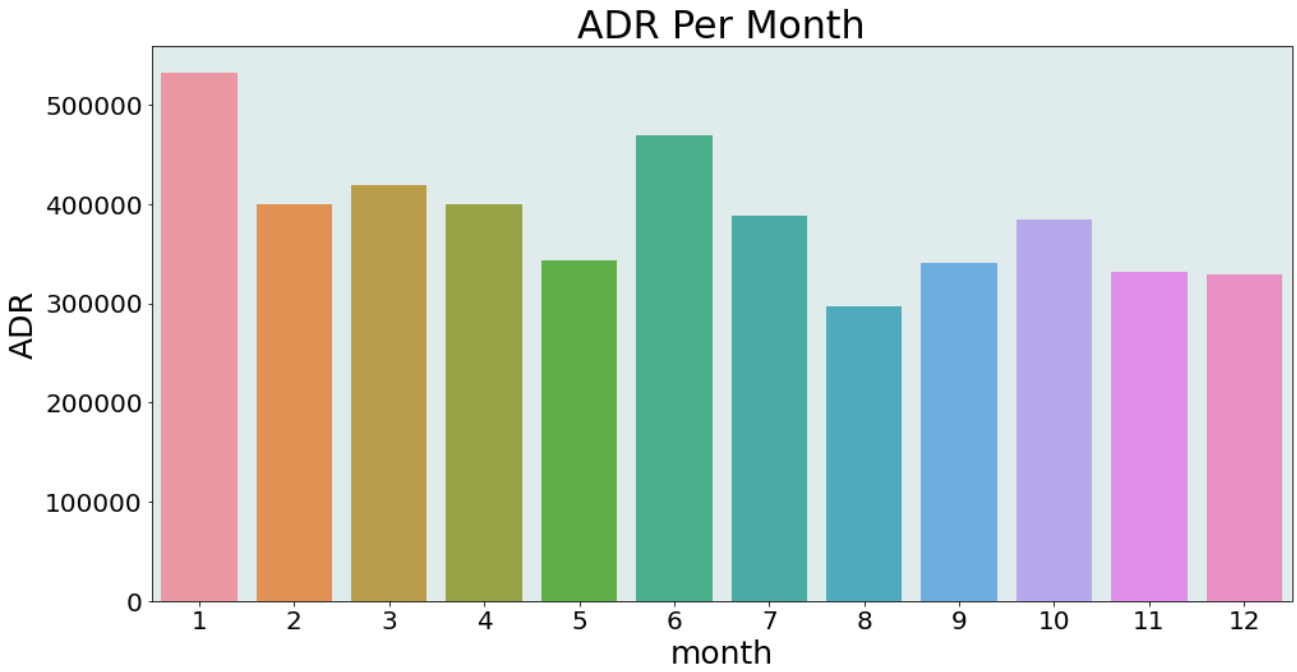


In [131]:
```
df['month']=df['reservation_status_date'].dt.month

plt.figure(figsize=(16,8))
ax=plt.axes()
ax.set_facecolor("#e0ebeb")

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",size=20)
plt.xlabel("Month",fontsize=25)
plt.ylabel("Number Of Reservation",fontsize=25)
plt.legend(['Not Canceled','Canceled'])
plt.xticks(fontsize=20)
plt.yticks(fontsize=20)
plt.show()
```

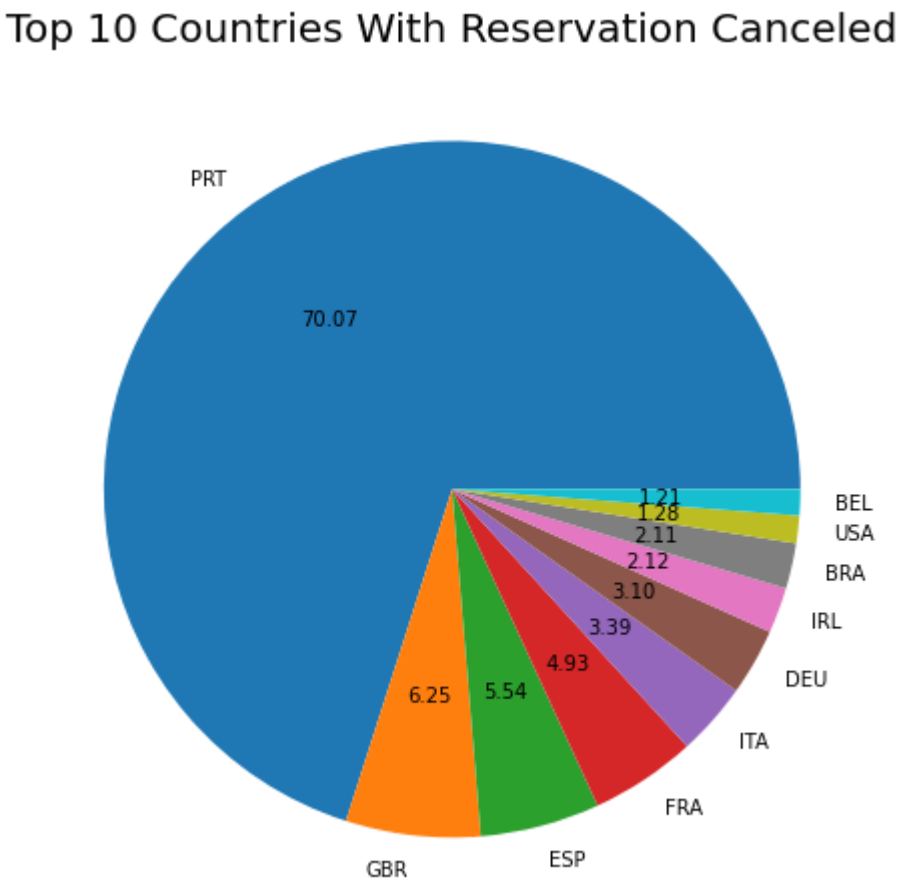


In [132]:

```
plt.figure(figsize=(16,8))
ax=plt.axes()
ax.set_facecolor("#e0ebeb")
plt.title('ADR Per Month',fontsize=30)
sns.barplot('month','adr',data=df[df['is_canceled']==1].groupby('month')[['adr']].sum
plt.xlabel("month",fontsize=25)
plt.ylabel("ADR",fontsize=25)
plt.xticks(fontsize=20)
plt.yticks(fontsize=20)
plt.show()
```



In [135]:
```
cancelled_data=df[df['is_canceled']==1]
top10_country=cancelled_data['country'].value_counts()[:10]
plt.figure(figsize=(8,8))
plt.title("Top 10 Countries With Reservation Canceled",fontsize=20)
plt.pie(top10_country,autopct='%.2f',labels=top10_country.index)
plt.show()
```



In [136]:
```
df['market_segment'].value_counts()
```

Out[136]:
```
Online TA        56402
Offline TA/TO    24159
Groups           19806
Direct           12448
Corporate         5111
Complementary      734
Aviation           237
Name: market_segment, dtype: int64
```

In [137]:
```
df['market_segment'].value_counts(normalize=True)
```

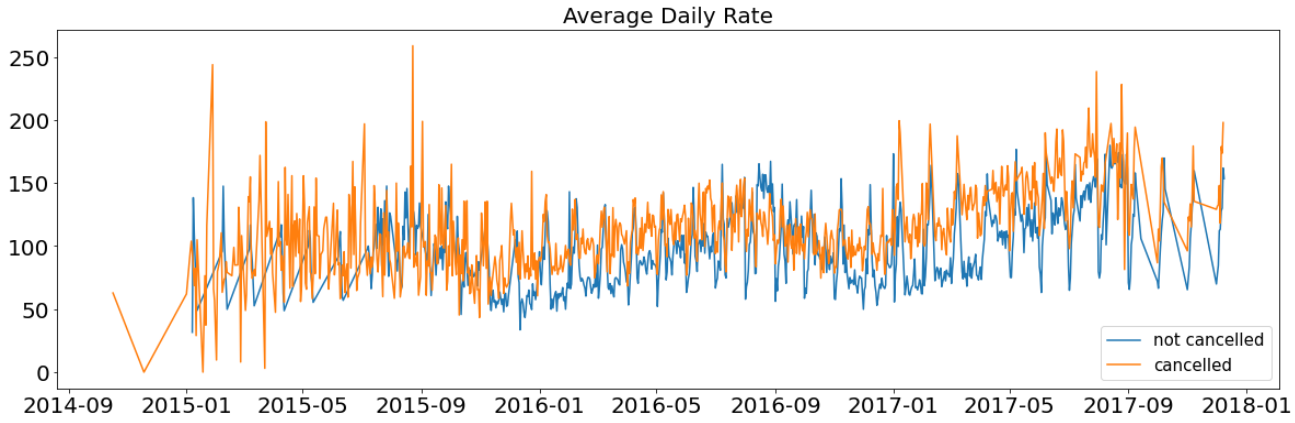
Out[137]:
```
Online TA        0.474377
Offline TA/TO    0.203193
Groups           0.166581
Direct           0.104696
Corporate        0.042987
Complementary    0.006173
Aviation         0.001993
Name: market_segment, dtype: float64
```

In [138]:
```
cancelled_data['market_segment'].value_counts(normalize=True)
```

Out[138]:
```
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: market_segment, dtype: float64
```

In [150]:
```
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_df_adr=not_cancelled_data.groupby('reservation_status_date')[['adr']].m
not_cancelled_df_adr.reset_index(inplace=True)
not_cancelled_df_adr.sort_values('reservation_status_date',inplace=True)
plt.figure(figsize=(20,6))
plt.title("Average Daily Rate",fontsize=20)
plt.plot(not_cancelled_df_adr['reservation_status_date'],not_cancelled_df_adr['adr'],
plt.plot(cancelled_df_adr['reservation_status_date'],cancelled_df_adr['adr'],label='c
plt.xticks(fontsize=20)
plt.yticks(fontsize=20)
plt.legend(fontsize=15)
```

Out[150]:
```
<matplotlib.legend.Legend at 0x1420cd47b80>
```



In [ ]: