PROBLEM STATEMENT REPORT

Library Used:

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import warnings

warnings.filterwarnings('ignore')
```

Exploratory data analysis and Data Cleaning

- Checking number of Rows and Columns.
- Changing object to datetime
- Checking missing value
- Checking Null value
- Removing Outliers

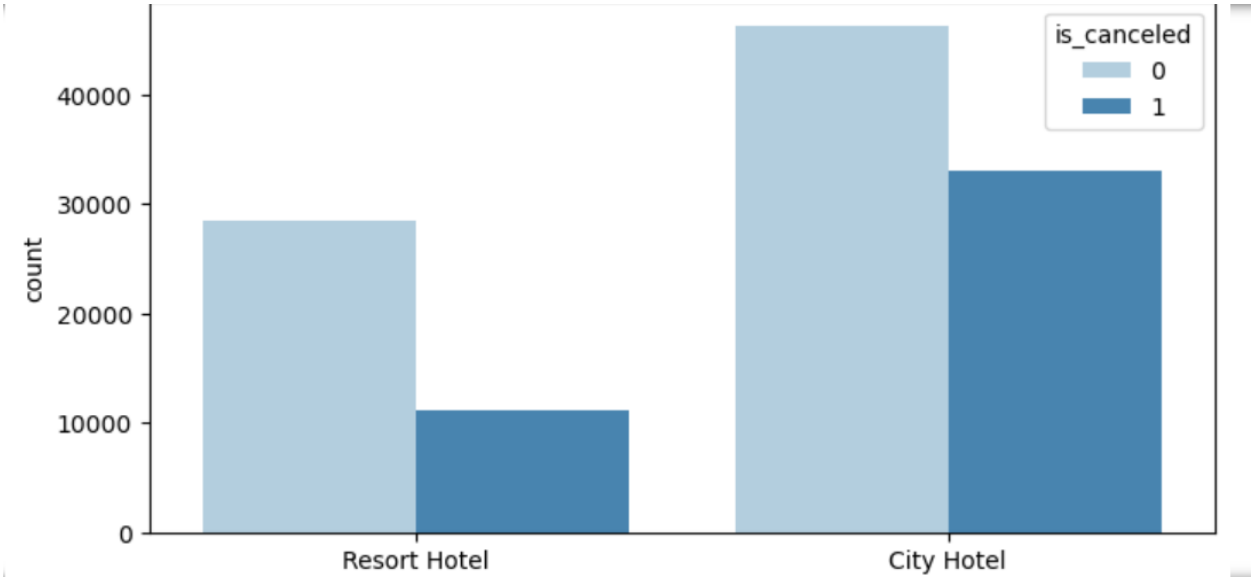
What is the amount reservation has ben canceled and not canceled

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

can_per

plt.figure(figsize=(5,4))

plt.title('Reservation status counts')

plt.bar(['Not canceled','Canceled'],df['is_canceled'].value_counts(),edgecolor='k',width=0.8)

plt.show()
```

As we saw that near about 63 % reservation are not cancled and 37% reservation are cancled

Which category of hotel having more canceled?

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

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

legend_labels,_=ax1.get_legend_handles_labels()

ax1.legend(bbox_to_anchor(1,1))

plt.title("Reservation status in different hotelss",size=25)

plt.xlabel('Hotel')

plt.ylabel('Numbers of reservations')
```

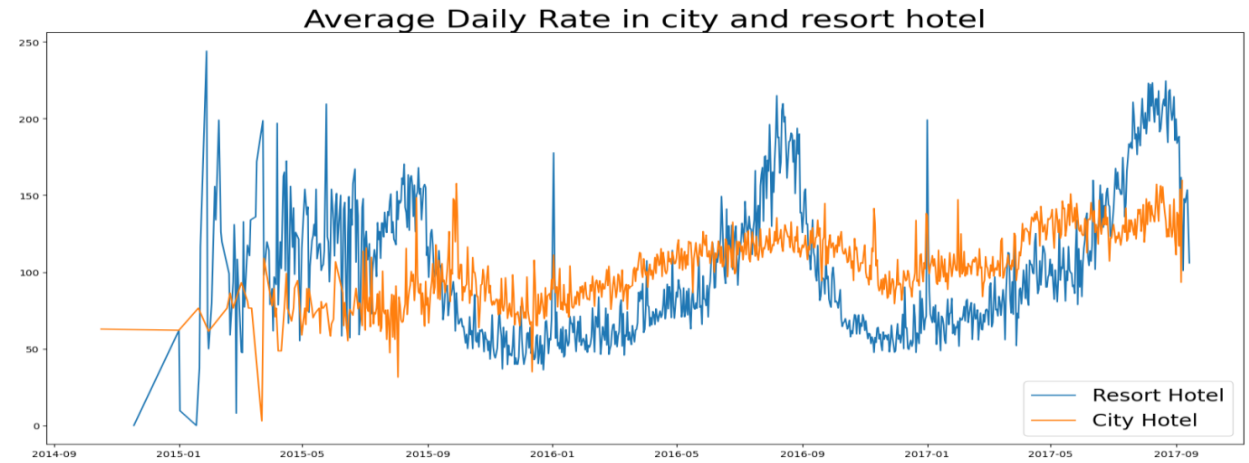


in the case of Resort hotel, we saw that near about 72% is not canceled but 30% get canceled

In the case of City hotels we can see that about 58% booking is not canceled but 41% booking is canceled¶

Q Average Daily Rate in city and resort hotel

```
plt.figure(figsize=(20,9))

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=20)

plt.show()
```



Q Reservation status per month

df['month']=df['reservation_status_date'].dt.month

plt.figure(figsize=(16,8))

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))

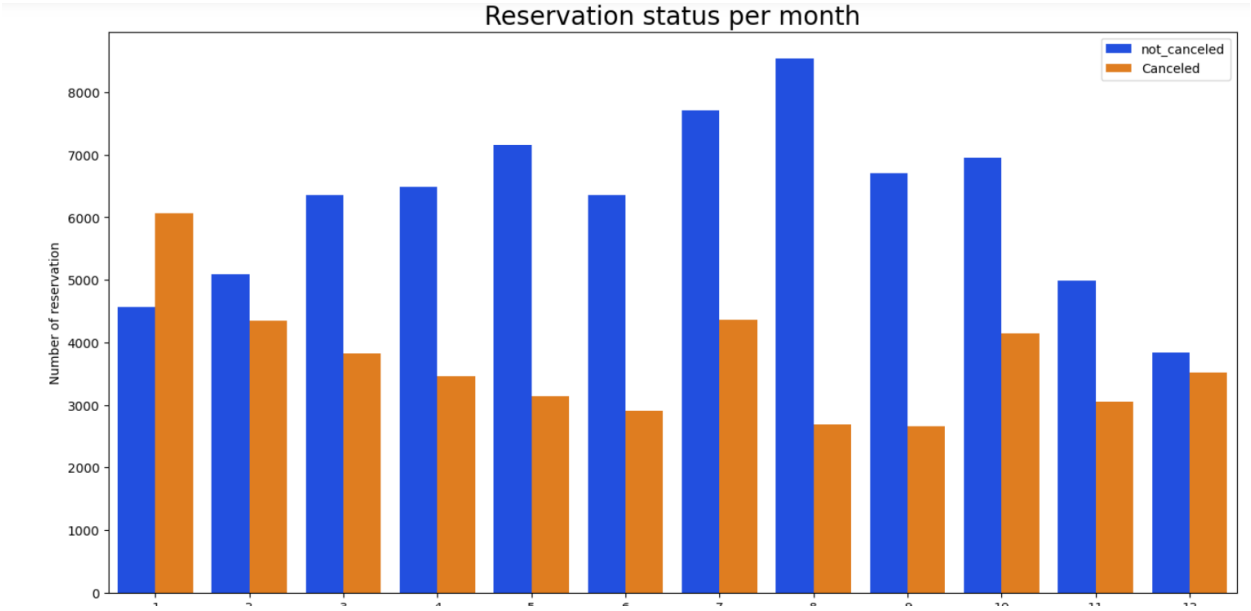
plt.title("Reservation status per month",size=20)

plt.xlabel('month')

plt.ylabel('Number of reservation')

plt.legend(['not_canceled','Canceled'])

plt.show()



Q.ADR Per Month

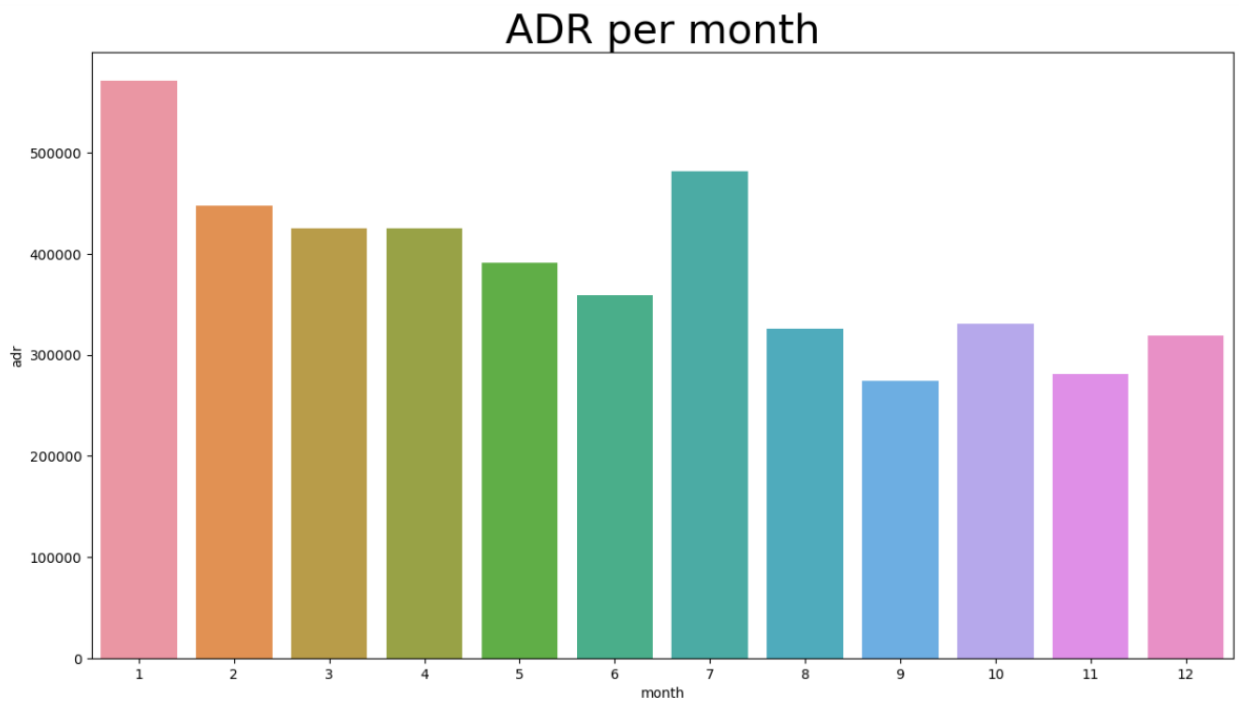
canceled_data=df[df['is_canceled']==1]

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('month')[['adr']].sum().reset_index())

plt.show()



Q. Top country with reservation canceled

cancelled_data=df[df['is_canceled']==1]

top_10_country=cancelled_data['country'].value_counts()[:10]

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

plt.title("Top country with reversavation canceled")

plt.pie(top_10_country,autopct='%.2f',labels=top_10_country.index)

plt.show()