Title: Hotel Booking Cancellation Analysis: Maximizing Revenue Efficiency

Introduction:

In this report, we will analyze the issue of high cancellation rates in City Hotel and Resort Hotel, aiming to provide valuable insights and recommendations to address this challenge. The focus is on optimizing revenue generation and improving operational efficiency by reducing cancellation rates.

Objective:

The primary objective of this analysis is to understand the factors contributing to high cancellation rates in both City Hotel and Resort Hotel. By identifying these factors, we can develop targeted strategies to minimize cancellations, increase revenue, and optimize hotel room utilization.

Methodology:

To achieve our objective, we will perform a comprehensive analysis of hotel booking cancellations, examining various factors that influence cancellation rates. We will consider factors such as pricing strategies, seasonal variations, hotel location, and quality of service. The analysis will involve studying historical data, industry trends, and customer behavior patterns.

Key Deliverables:

1. Identification of key factors influencing cancellation rates in City Hotel and Resort Hotel.
2. Evaluation of the impact of pricing strategies, seasonal variations, and location on cancellation rates.
3. Assessment of the quality of service and its correlation with cancellations.
4. Recommendations for targeted strategies to reduce cancellation rates and increase revenue efficiency.

5. Suggestions for improving hotel room utilization and overall operational efficiency.

Conclusion:

By analyzing hotel booking cancellations and related factors, this report aims to provide actionable insights for City Hotel and Resort Hotel. By implementing the recommended strategies, both hotels can minimize cancellations, increase revenue, and improve overall business performance. With a focus on revenue efficiency and optimized hotel room utilization, the hotels can enhance customer satisfaction and drive long-term success in the hospitality industry.

LINKEDIN LINK :

https://www.linkedin.com/in/asad-ali-044240262/ (https://www.linkedin.com/in/asad-ali-044240262/)

GIT HUB LINK :

https://github.com/Asadxio (https://github.com/Asadxio)

Importing Libraries

In [1]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

Loading The Data Set

In [2]:

```
df = pd.read_csv("C:/Users/Asad/Downloads/Compressed/Hotel booking/hotel_booking.csv")
```

Exploratory Data Analysis and Data Cleaning

In [3]:

```
df.head()
```

Out[3]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_nur
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 × 32 columns

In [4]:

```
df.tail()
```

Out[4]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week
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 × 32 columns

In [5]:

```
df.shape
```

Out[5]:

```
(119390, 32)
```

In [6]:

```
df.columns
```

Out[6]:

```
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'],
      dtype='object')
```

In [7]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 32 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 
dtypes: float64(4), int64(16), object(12)
memory usage: 29.1+ MB
```

In [8]:

```
df["reservation_status_date"] = pd.to_datetime(df["reservation_status_date"])
```

In [9]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 32 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]
dtypes: datetime64[ns](1), float64(4), int64(16), object(11)
memory usage: 29.1+ MB
```

In [10]:

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

Out[10]:

	hotel	arrival_date_month	meal	country	market_segment	distribution_channel	r
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 [11]:

```
for col in df.describe(include = "object").columns:
    print(col)
    print(df[col].unique())
    print('-'*70)
```

```
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']
----------------------------------------------------------------------
```

In [12]:

```
df.isnull().sum()
```

Out[12]:

```
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
dtype: int64
```

In [13]:

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

In [14]:

```
df.isnull().sum()
```

Out[14]:

```
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
dtype: int64
```

In [15]:

```
df.describe()
```

Out[15]:

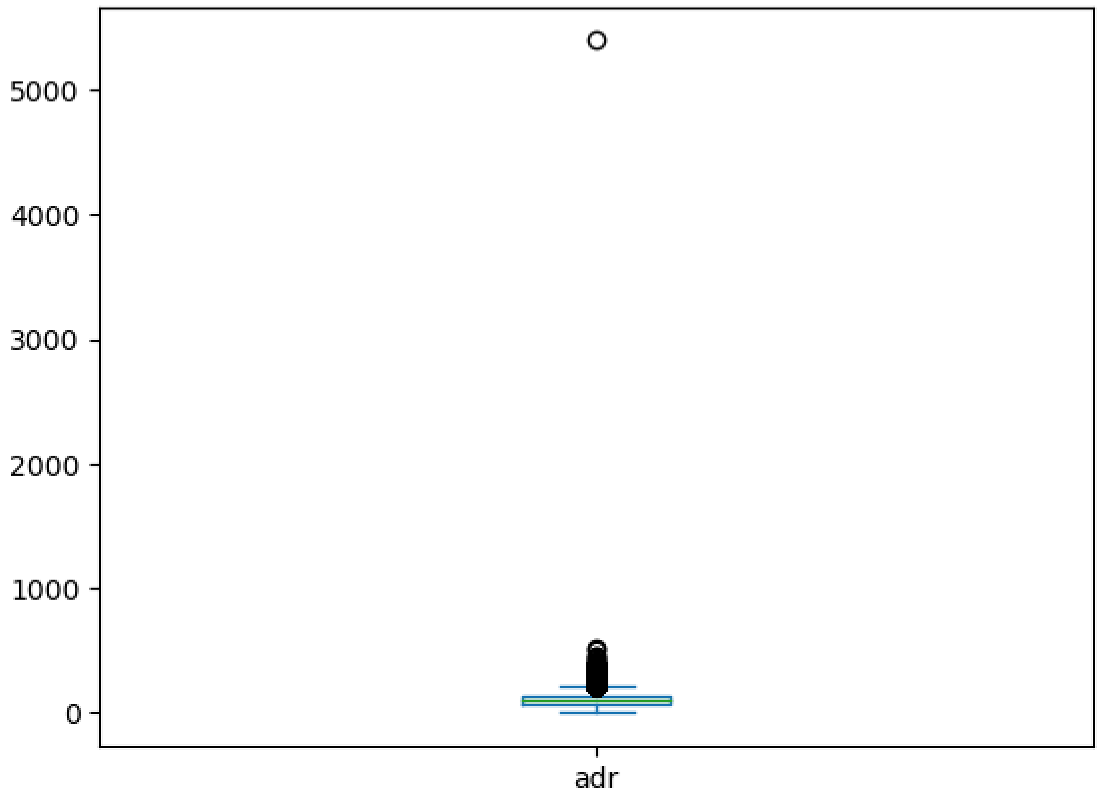
	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date
count	118898.000000	118898.000000	118898.000000	118898.000000	
mean	0.371352	104.311435	2016.157656	27.166555	
std	0.483168	106.903309	0.707459	13.589971	
min	0.000000	0.000000	2015.000000	1.000000	
25%	0.000000	18.000000	2016.000000	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	

In [16]:

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

Out[16]:

```
<Axes: >
```



In [17]:

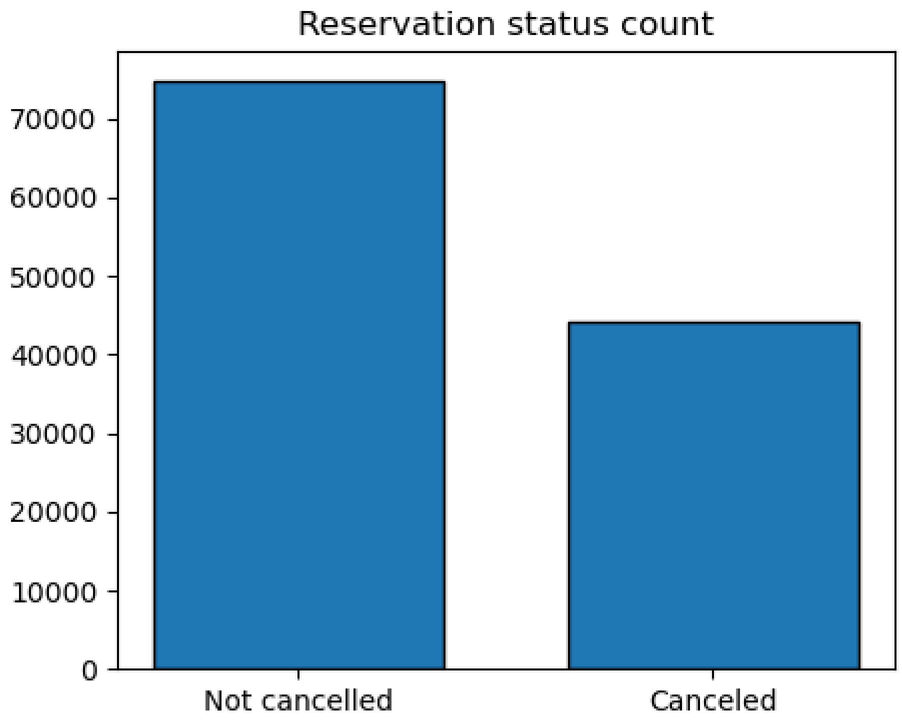
```
df= df[df['adr']<5000]
```

Data Analysis and Visualization

In [18]:

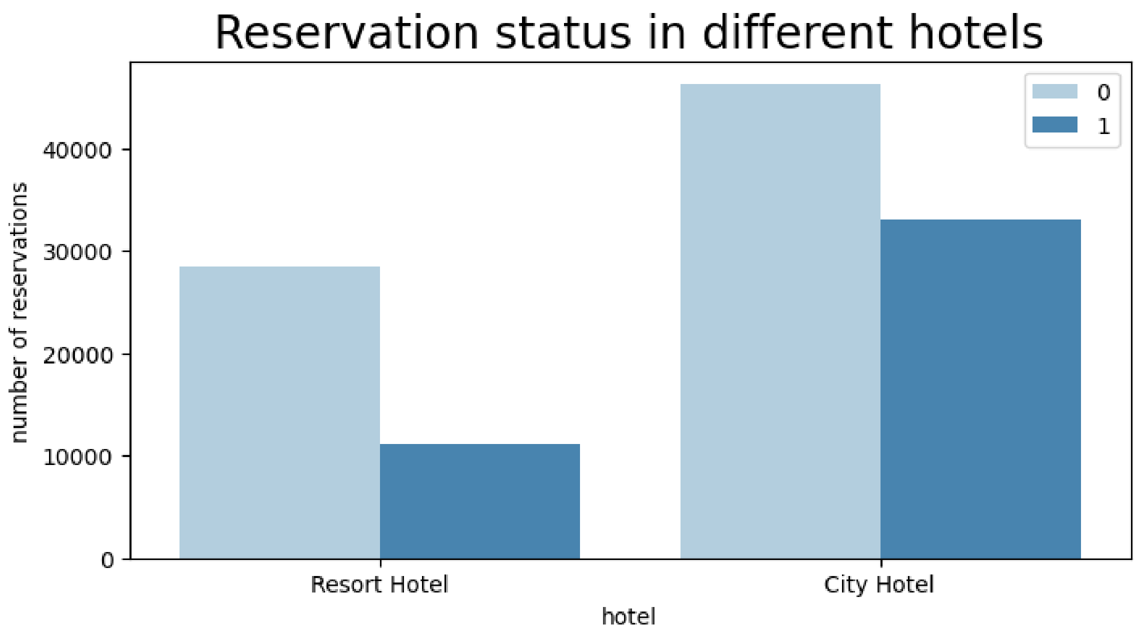
```
cancelled_prec = df['is_canceled'].value_counts(normalize=True)
print(cancelled_prec)

plt.figure(figsize=(5, 4))
plt.title('Reservation status count')
plt.bar(['Not cancelled', 'Canceled'], df['is_canceled'].value_counts(), edgecolor='k', w
plt.show()
```

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



In [19]:

```
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 hotels', size=20)
plt.xlabel('hotel')
plt.ylabel('number of reservations')
plt.show()
```



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_canceled_prec = city_hotel['is_canceled'].value_counts(normalize=True)
print(city_canceled_prec)
```

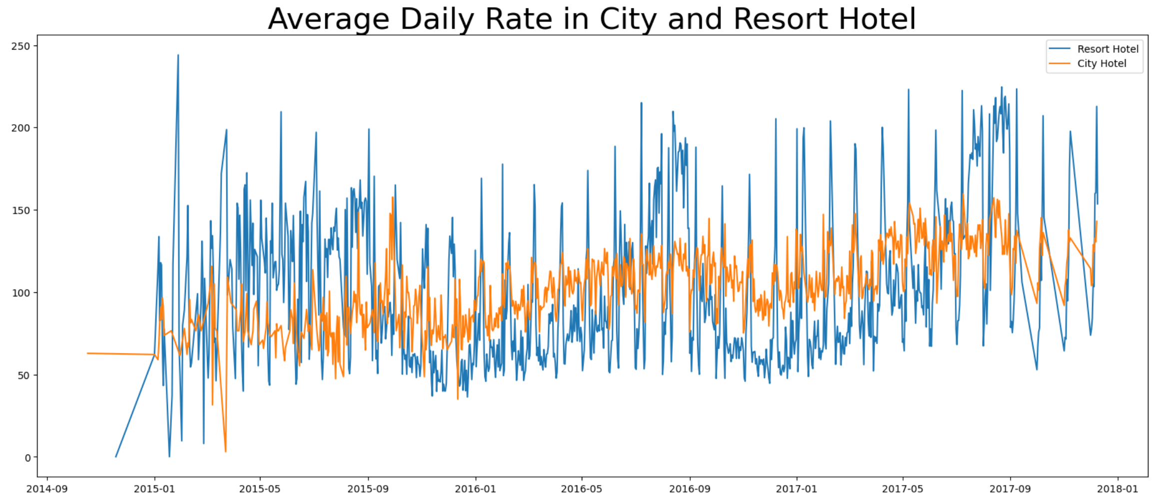
```
0    0.582918
1    0.417082
Name: is_canceled, dtype: float64
```

In [22]:

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

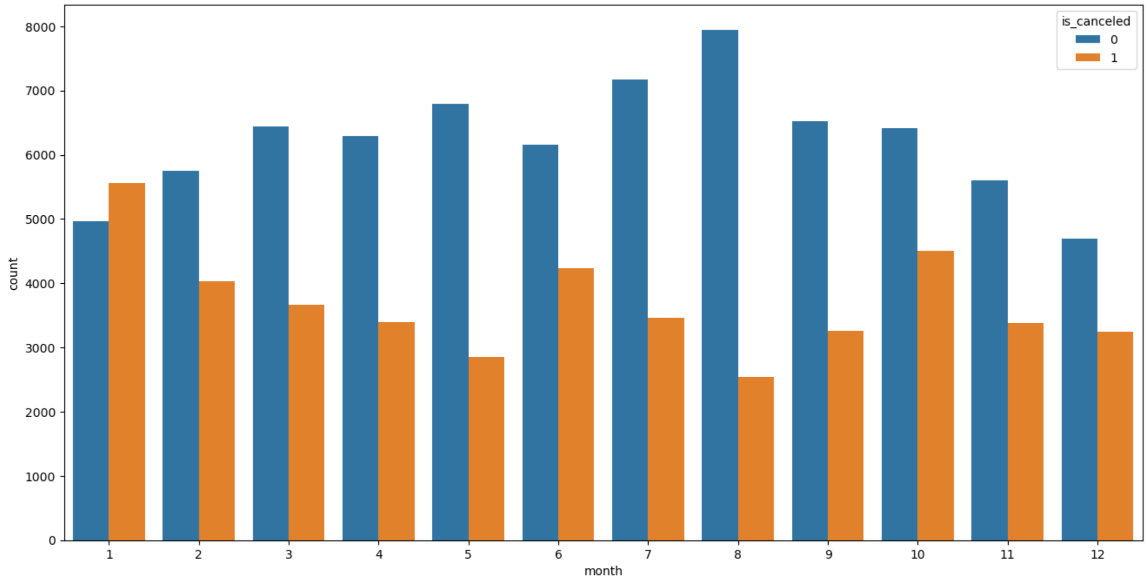
In [23]:

```
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()
plt.show()
```



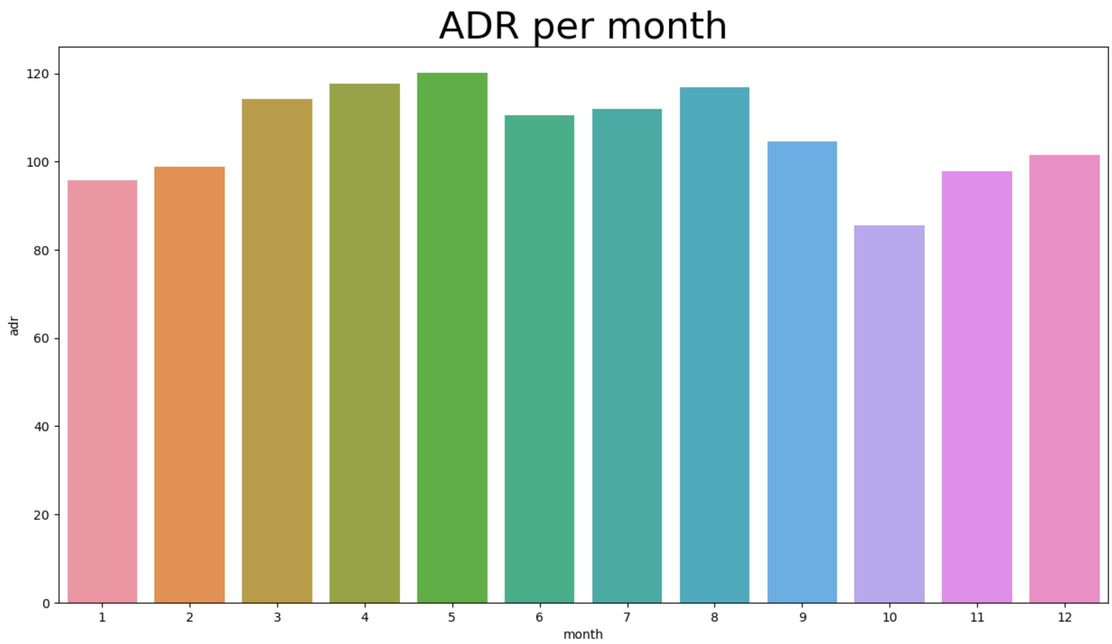
In [24]:

```
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()
```



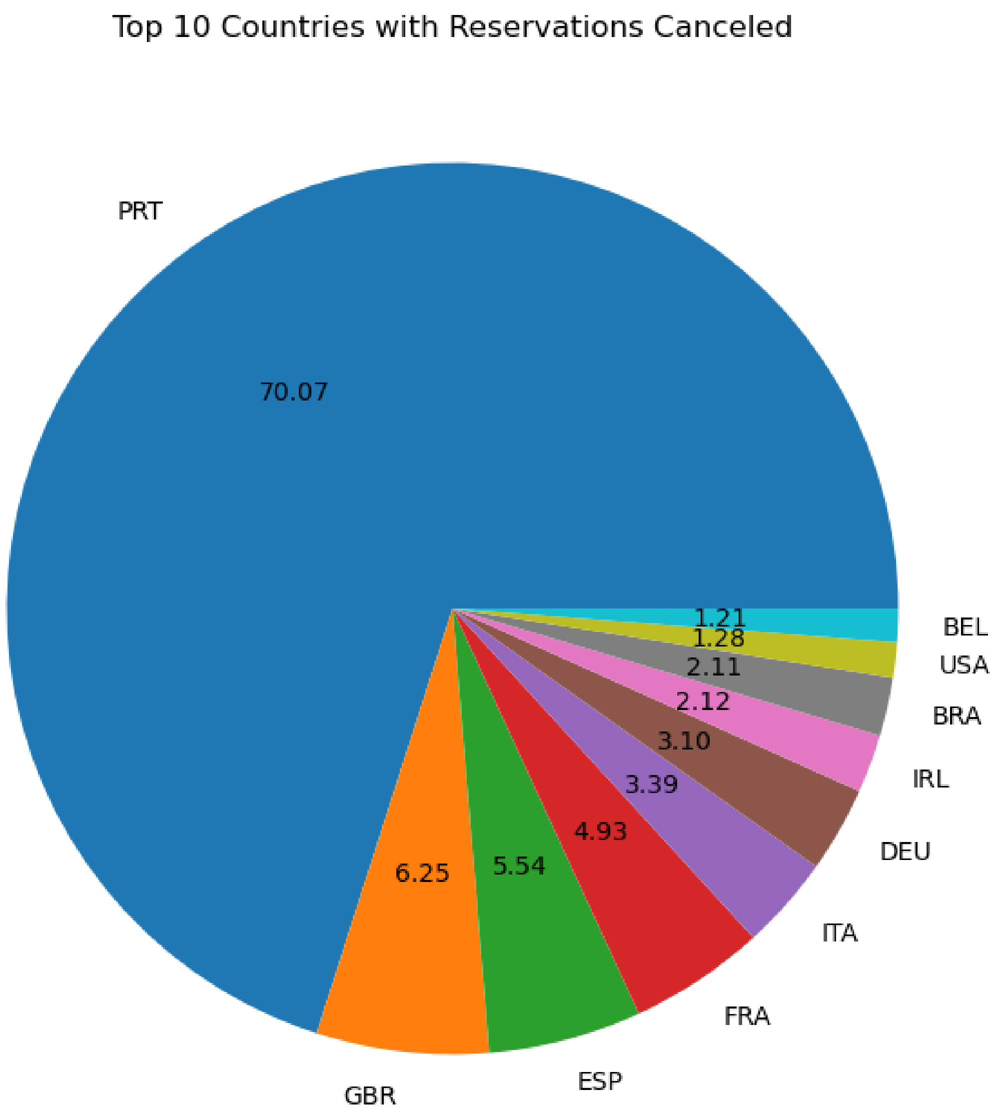
In [25]:

```
plt.figure(figsize=(15, 8))
plt.title('ADR per month', fontsize=30)

canceled_reservations = df[df['is_canceled'] == 1]

# Group the data by month and calculate the mean of ADR
monthly_adr = canceled_reservations.groupby('month')['adr'].mean().reset_index()

# Plot the bar chart
sns.barplot(x='month', y='adr', data=monthly_adr)

plt.show()
```



In [26]:

```
cancelled_data = df[df['is_canceled'] == 1]
top_10_country = cancelled_data['country'].value_counts().head(10)

plt.figure(figsize=(8, 8))
plt.title('Top 10 Countries with Reservations Canceled')
plt.pie(top_10_country, autopct='%.2f', labels=top_10_country.index)
plt.show()
```



In [27]:

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

Out[27]:

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

In [28]:

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

Out[28]:

```
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 [29]:

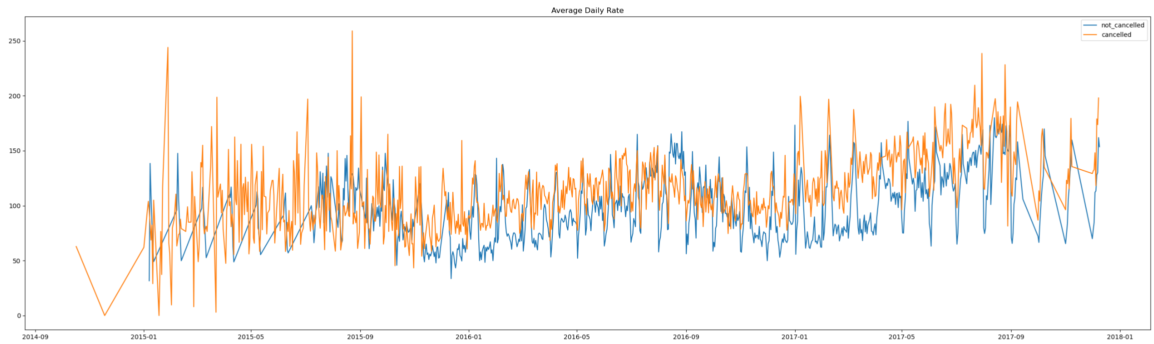
```
cancelled_data['market_segment'].value_counts(normalize = True)
```

Out[29]:

```
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 [30]:

```
import matplotlib.pyplot as plt

cancelled_df_adr = cancelled_data.groupby("reservation_status_date")["adr"].mean().reset_
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"].mean(
not_cancelled_df_adr.sort_values("reservation_status_date", inplace=True)

plt.figure(figsize=(32, 9))
plt.title('Average Daily Rate')
plt.plot(not_cancelled_df_adr['reservation_status_date'], not_cancelled_df_adr['adr'], la
plt.plot(cancelled_df_adr['reservation_status_date'], cancelled_df_adr['adr'], label='car
plt.legend()
plt.show()
```



Suggestions :

1 Pricing strategies: Hotels can analyze the relationship between cancellation rates and pricing. By adjusting their pricing strategies, they can offer lower rates for specific hotels based on their locations or seasons. This can attract more bookings and potentially reduce cancellations.

2 Weekend and holiday discounts: Since resort hotels have a higher cancellation ratio compared to city hotels, offering reasonable discounts on room prices during weekends or holidays can incentivize guests to book and reduce the likelihood of cancellations. Special promotions or packages tailored to these periods can also be effective.

3 January campaigns: Given that January has the highest cancellation rate, hotels can launch targeted marketing campaigns during this month. Offering attractive deals, such as

discounted rates or value-added services, can encourage guests to maintain their bookings and increase revenue during this period.

4 Quality improvements: Focusing on improving the quality of hotels and services, particularly in Portugal where cancellations are prominent, can enhance guest satisfaction and reduce cancellations. This can include upgrading facilities, enhancing customer service, and actively addressing guest feedback to create a positive experience that encourages guests to keep their reservations.

Implementing these suggestions can help hotels minimize