Data Discription

The dataset consists of hotel booking data with 36 columns and 119,390 rows. It contains all the necessary details required for booking a room in a hotel, including information such as guest demographics, booking dates, room type, number of adults and children, special requests, and more.

Additionally, the dataset includes the cancellation status of each booking, indicating whether the reservation was canceled or not. This information can provide insights into booking trends and cancellation rates within the hotel industry.

Furthermore, the dataset also captures the source of each booking, which indicates the channel or platform through which the reservation was made. This data can help analyze the effectiveness of different booking sources and identify the most popular platforms for hotel bookings.

Overall, the dataset provides a comprehensive view of hotel booking activities, including reservation details, cancellation status, and the sources through which bookings are made. It can be used for various analytical purposes, such as understanding booking patterns, predicting cancellations, and evaluating the performance of different booking channels.

Column descriptions

- hotel: The name of the hotel (non-null object) *is_canceled: Indicates whether the booking was canceled (non-null int64)*lead_time: Number of days between booking and arrival (non-null int64) *arrival_date_year: Year of arrival (non-null int64)*arrival_date_month: Month of arrival (non-null object) *arrival_date_week_number: Week number of arrival (non-null int64)*arrival_date_day_of_month: Day of the month of arrival (non-null int64) *stays_in_weekend_nights: Number of weekend nights stayed (non-null int64)*stays_in_week_nights: Number of weekday nights stayed (non-null int64) *adults: Number of adults (non-null int64)*children: Number of children (non-null float64) *babies: Number of babies (non-null int64)*meal: Type of meal booked (non-null object) *country: Country of origin (non-null object)*market_segment: Market segment of the booking (non-null object) *distribution_channel: Booking distribution channel (non-null object)*is_repeated_guest: Indicates if the guest is a repeated guest (non-null int64) *previous_cancellations: Number of previous cancellations by the guest (non-null int64)*previous_bookings_not_canceled: Number of previous bookings not canceled by the guest (non-null int64) *reserved_room_type: Type of room reserved (non-null object)*assigned_room_type: Type of room assigned (non-null object) *booking_changes: Number of changes made to the booking (non-null int64)*deposit_type: Type of deposit made (non-null object) *agent: ID of the travel agency that made the booking (non-null float64)*company: ID of the company that made the booking (non-null float64) *days_in_waiting_list: Number of days the booking was on the waiting list (non-null int64)*customer_type: Type of customer (non-null object) *adr: Average Daily Rate (non-null float64)*required_car_parking_spaces: Number of car parking spaces requested (non-null int64) *total_of_special_requests: Number of special requests made (non-null int64)*reservation_status: Current status of the reservation (non-null object) *reservation_status_date: Date at which the reservation status was last updated (non-null object)*name: Name of the guest (non-null object) *email: Email address of the guest (non-null object)*phone-number: Phone number of the guest (non-null object) *credit_card: Credit card information of the guest (non-null object)

Importing Libraries

In [30]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings
```

Out[30]:
```
<function warnings.filterwarnings(action, message='', category=<class 'Warning'>, module='', lineno=0, append=False)>
```

Loading Dataset

In [31]:
```python
df=pd.read_csv('hotel_booking.csv')
```

In [32]:
```python
df.head()
```

Out[32]:

	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	...	customer_type
0	Resort Hotel	0	342	2015	July	27	1	0	0	2	...	Transient
1	Resort Hotel	0	737	2015	July	27	1	0	0	2	...	Transient
2	Resort Hotel	0	7	2015	July	27	1	0	1	1	...	Transient
3	Resort Hotel	0	13	2015	July	27	1	0	1	1	...	Transient
4	Resort Hotel	0	14	2015	July	27	1	0	2	2	...	Transient

5 rows × 36 columns

Data Analysis

In [33]:
```python
df.shape
```

Out[33]:
```
(119390, 36)
```

In [34]:
```python
df.columns
```

Out[34]:
```
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 [35]:
```python
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 [36]:
```python
df['reservation_status_date'] =pd.to_datetime(df['reservation_status_date' ])
```

In [37]:
```python
df.describe(include='object')
```

Out[37]:

	hotel	arrival_date_month	meal	country	market_segment	distribution_channel	reserved_room_type	assigned_room_type	deposit_type	customer_type	reservation_status	name	
count	119390	119390	119390	118902	119390	119390	119390	119390	119390	119390	119390	119390	
unique	2	12	5	177	8	5	10	12	3	4	3	81503	
top	City Hotel	August	BB	PRT	Online TA	TA/TO	A	A	No Deposit	Transient	Check-Out	Michael Johnson	Michael.
freq	79330	13877	92310	48590	56477	97870	85994	74053	104641	89613	75166	48	

In [38]:
```python
df.isnull().sum()
```

Out[38]:
```
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 [39]:
```python
df.drop(['company','agent'], axis=1 , inplace=True )
```

In [40]:
```python
df.dropna(inplace=True)
```

In [41]:
```python
df.isnull().sum()
```

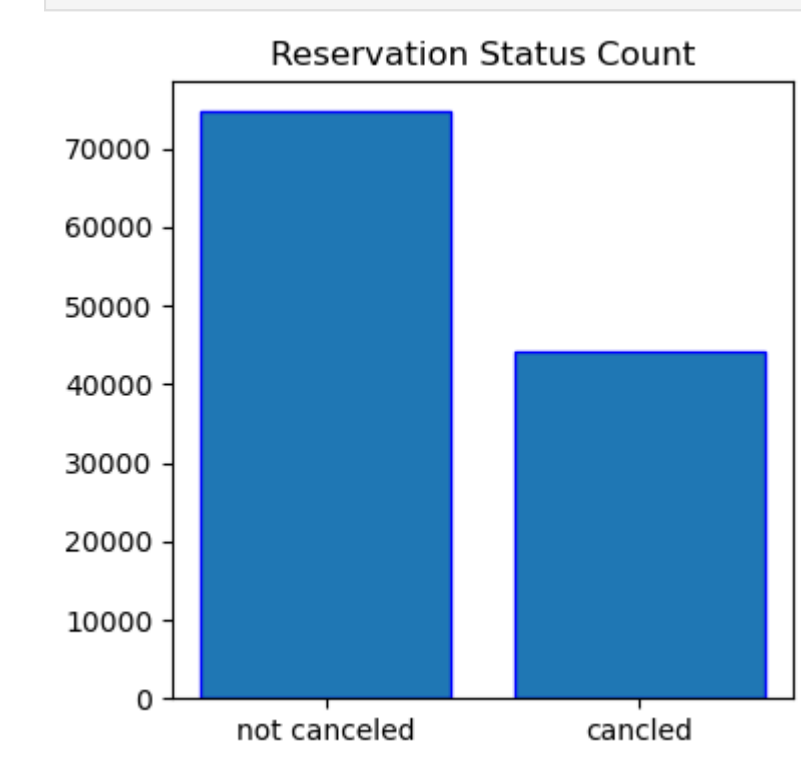
Out[41]:
```
hotel                             0
is_canceled                       0
lead_time                         0
arrival_date_year                 0
arrival_date_month                0
arrival_date_week_number          0
arrival_date_day_of_month         0
stays_in_weekend_nights           0
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
```

Data Visualization

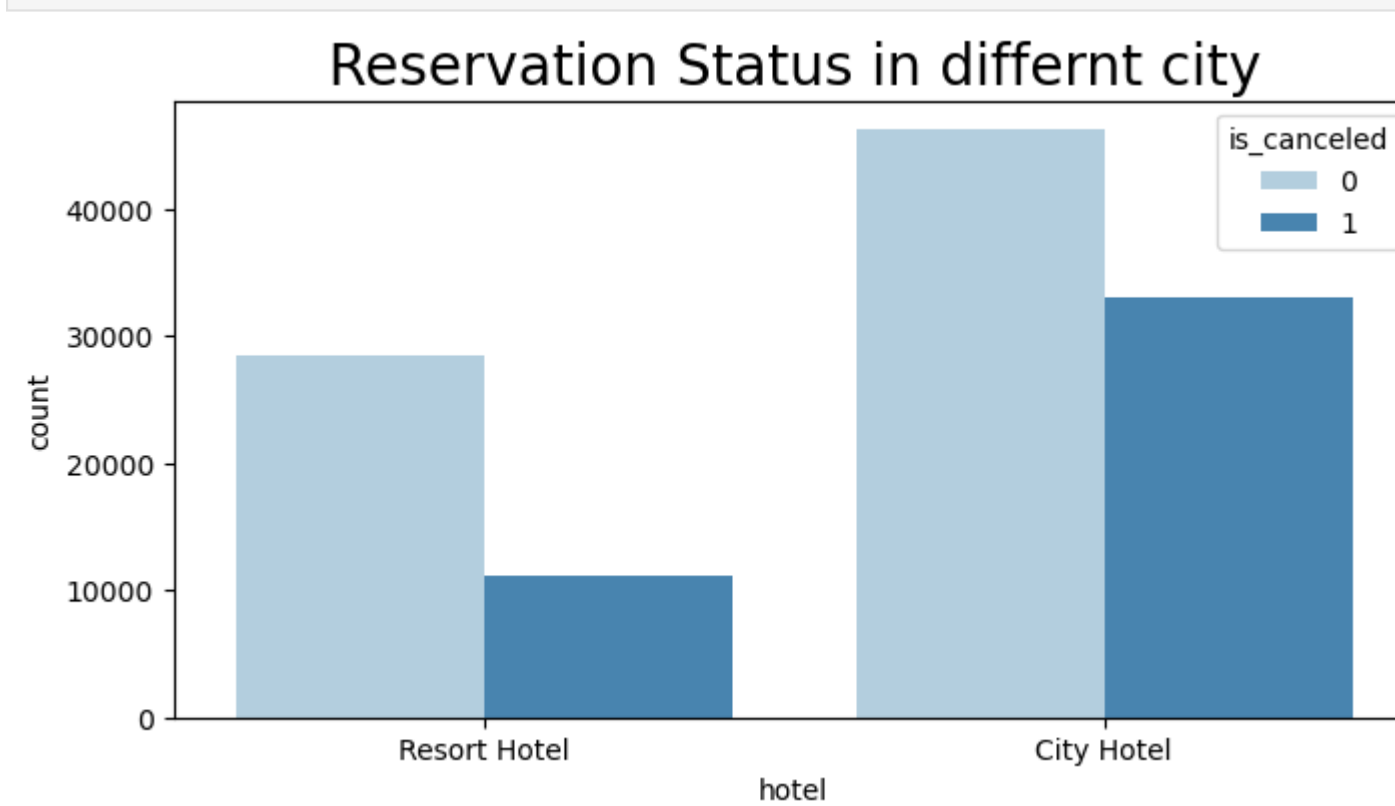
In [42]:
```python
cancelled_perc =df['is_canceled'].value_counts(normalize=True)
cancelled_perc
```

Out[42]:
```
0    0.628648
1    0.371352
Name: is_canceled, dtype: float64
```

In [43]:
```python
plt.figure(figsize=(4,4))
plt.title('Reservation Status Count')
plt.bar(['not canceled','cancled'],df['is_canceled'].value_counts(),edgecolor='b')
plt.show()
```



In [44]:
```python
plt.figure(figsize=(8,4))
ax1=sns.countplot(x='hotel',hue='is_canceled',data=df,palette='Blues')
plt.title('Reservation Status in differnt city',size=20)
plt.xlabel=('hotel')
plt.ylabel=('Num of reservation')
```



In [45]:
```python
resort_hotel=df[df['hotel']=='Resort Hotel']
resort_hotel['is_canceled'].value_counts(normalize = True)
```

Out[45]:
```
0    0.72025
1    0.27975
Name: is_canceled, dtype: float64
```

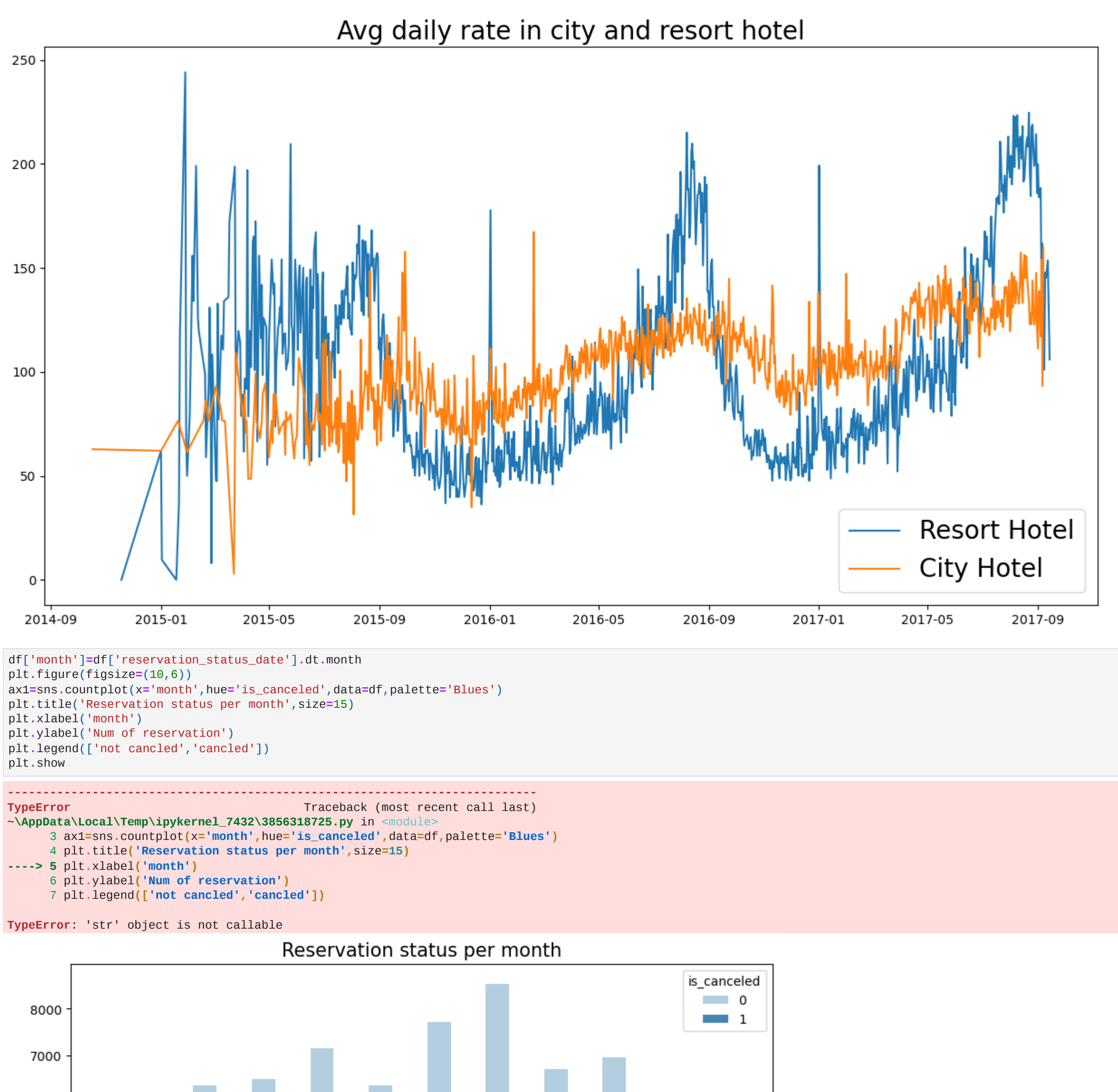
In [46]:
```python
city_hotel=df[df['hotel']=='City Hotel']
city_hotel['is_canceled'].value_counts(normalize = True)
```

Out[46]:
```
0    0.582911
1    0.417089
Name: is_canceled, dtype: float64
```

In [47]:
```python
resort_hotel=resort_hotel.groupby('reservation_status_date')[['adr']].mean()
city_hotel=city_hotel.groupby('reservation_status_date')[['adr']].mean()
```

In [48]:
```python
plt.figure(figsize=(15,8))
plt.title('Avg daily rate in city and resort hotel',size=20)
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
```

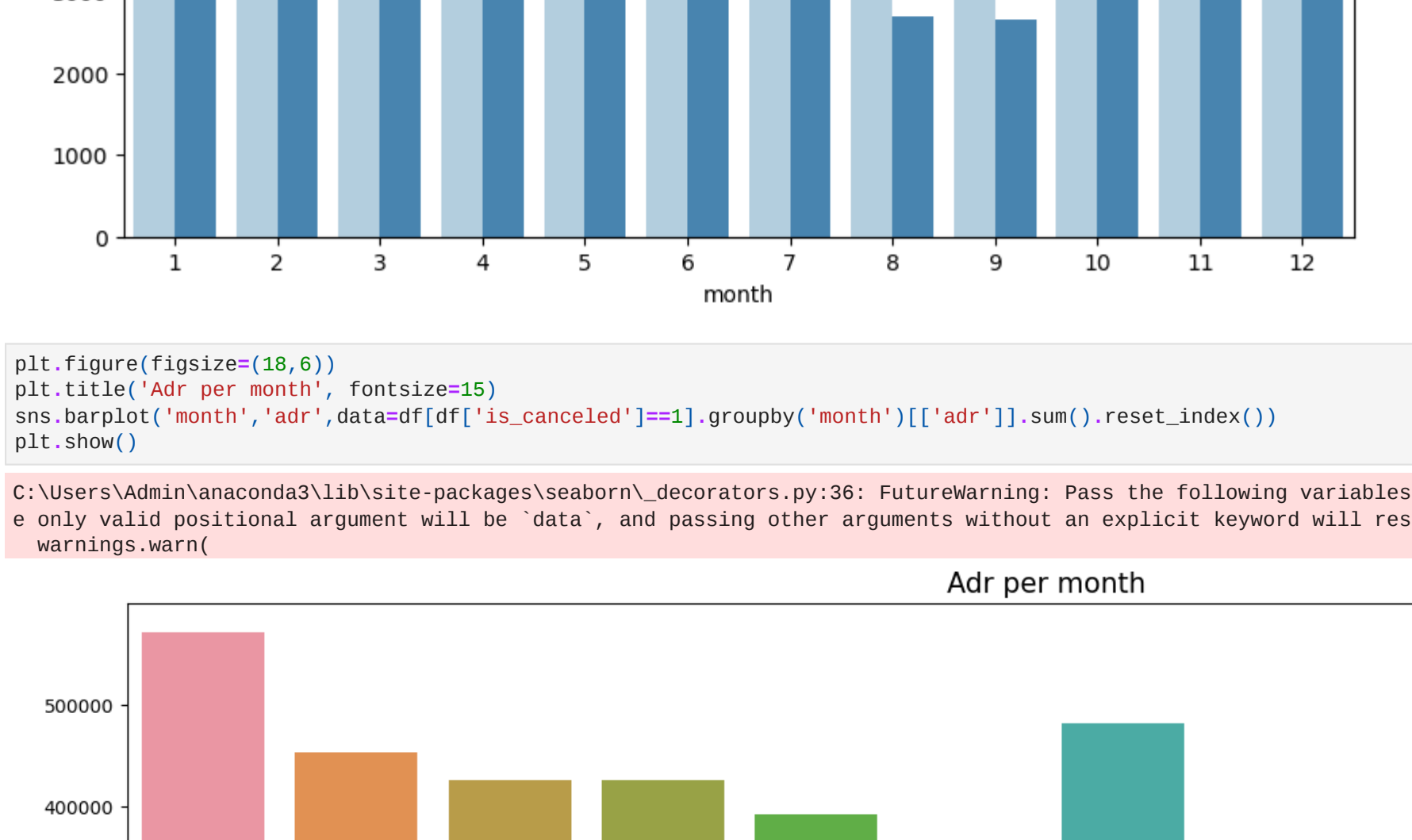
Out[48]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



In [76]:
```python
df['month']=df['reservation_status_date'].dt.month
plt.figure(figsize=(10,6))
ax1=sns.countplot(x='month',hue='is_canceled',data=df,palette='Blues')
plt.title('Reservation status per month',size=15)
plt.xlabel('month')
plt.ylabel('Num of reservation')
plt.legend(['not cancled','cancled'])
plt.show
```

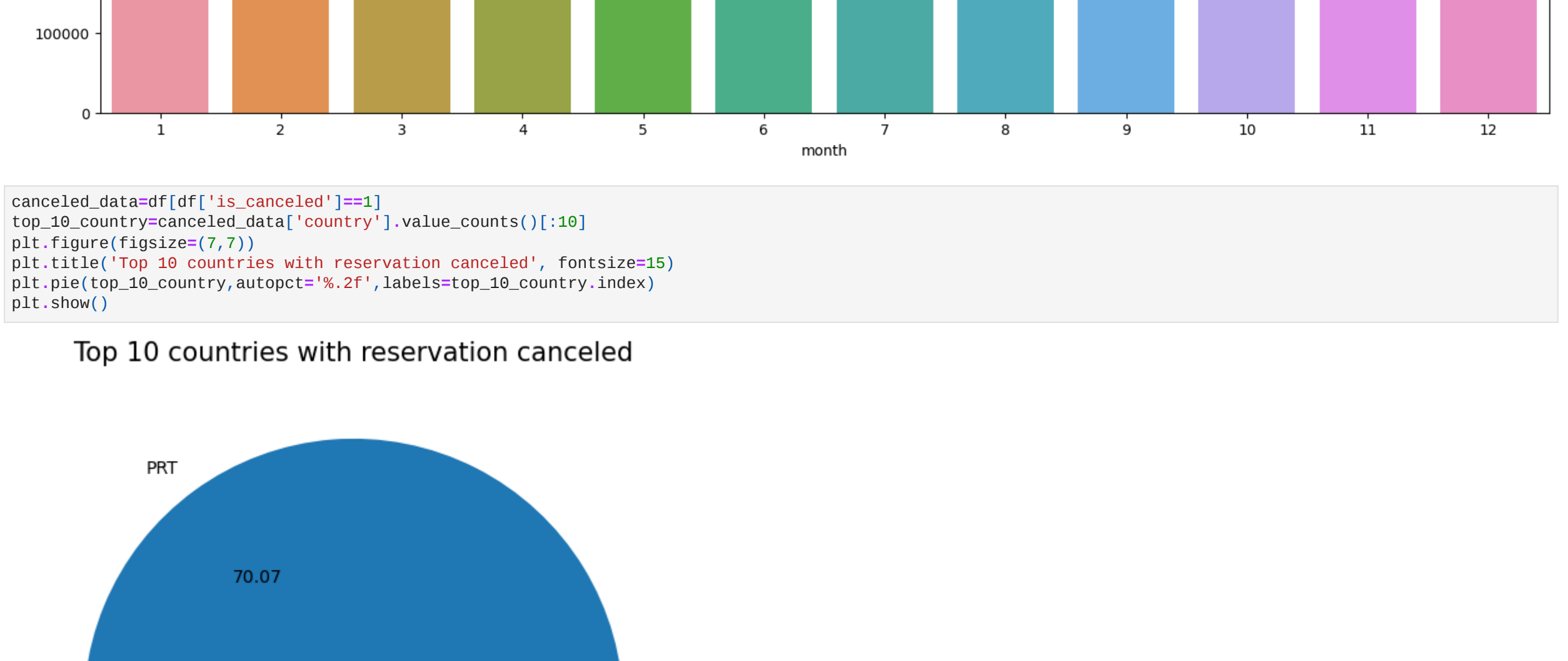
```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_7432\3856318725.py in <module>
      3 ax1=sns.countplot(x='month',hue='is_canceled',data=df,palette='Blues')
      4 plt.title('Reservation status per month',size=15)
----> 5 plt.xlabel('month')
      6 plt.ylabel('Num of reservation')
      7 plt.legend(['not cancled','cancled'])

TypeError: 'str' object is not callable
```

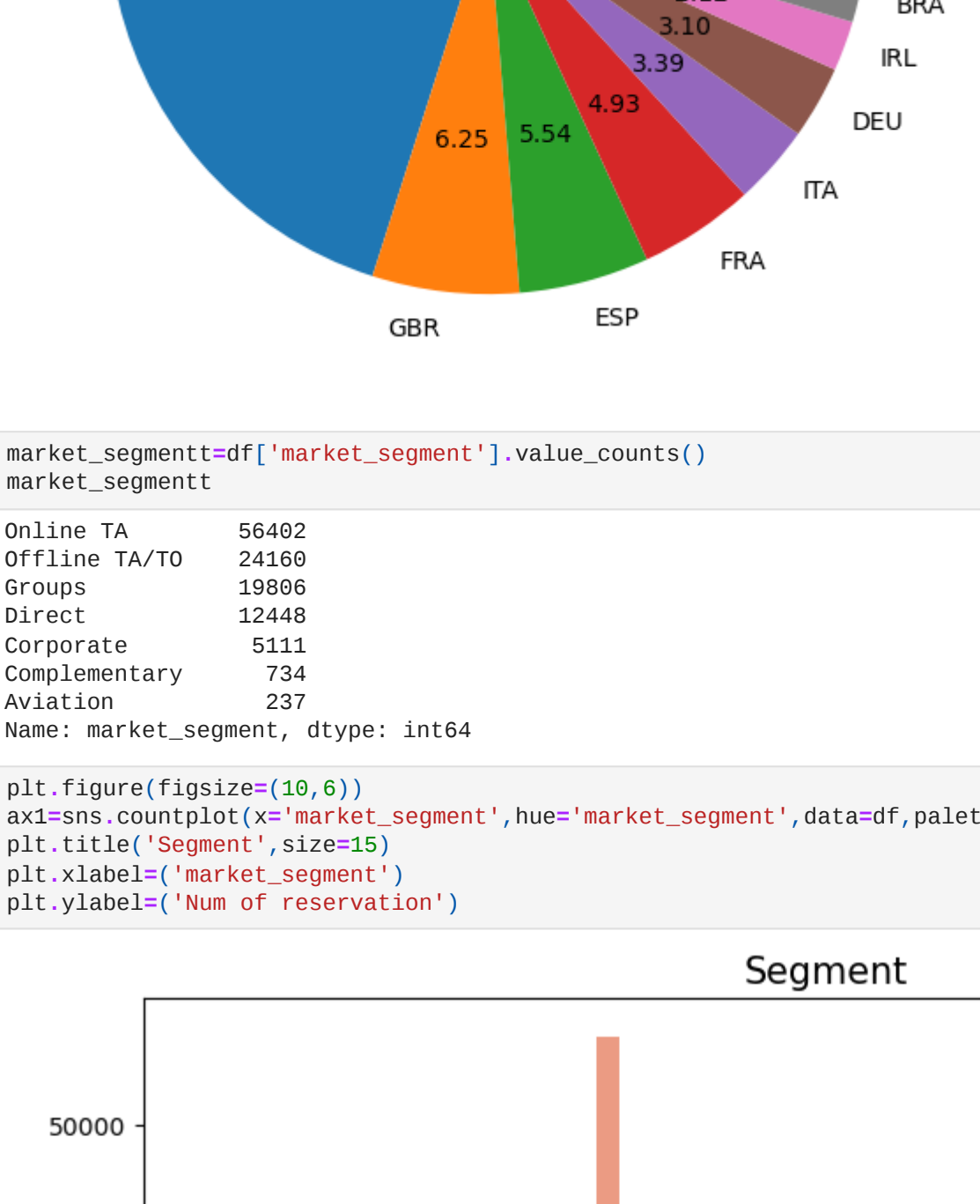


In [52]:
```python
plt.figure(figsize=(18,6))
plt.title('Adr per month', fontsize=15)
sns.barplot('month','adr',data=df[df['is_canceled']==1].groupby('month')[['adr']].sum().reset_index())
plt.show()
```

```
C:\Users\Admin\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```



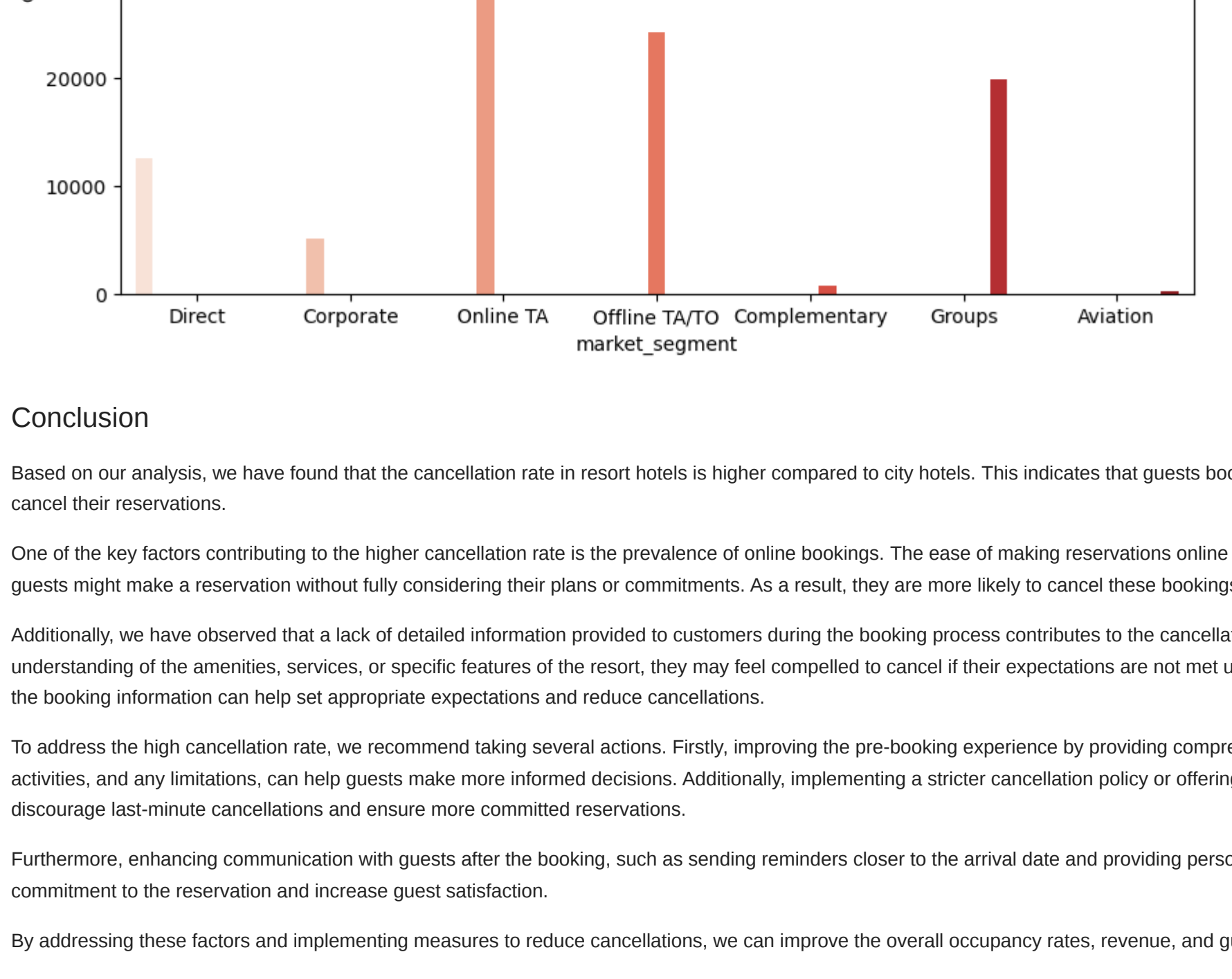
In [59]:
```python
canceled_data=df[df['is_canceled']==1]
top_10_country=canceled_data['country'].value_counts()[:10]
plt.figure(figsize=(7,7))
plt.title('Top 10 countries with reservation canceled', fontsize=15)
plt.pie(top_10_country,autopct='%.2f',labels=top_10_country.index)
plt.show()
```



In [63]:
```python
market_segmentt=df['market_segment'].value_counts()
market_segmentt
```

Out[63]:
```
Online TA        56402
Offline TA/TO    24160
Groups           19806
Direct           12448
Corporate         5111
Complementary      734
Aviation           237
Name: market_segment, dtype: int64
```

In [77]:
```python
plt.figure(figsize=(10,6))
ax1=sns.countplot(x='market_segment',hue='market_segment',data=df,palette='Reds')
plt.title('Segment',size=15)
plt.xlabel=('market_segment')
plt.ylabel=('Num of reservation')
```



Conclusion

Based on our analysis, we have found that the cancellation rate in resort hotels is higher compared to city hotels. This indicates that guests booking stays at resort properties are more likely to cancel their reservations.

One of the key factors contributing to the higher cancellation rate is the prevalence of online bookings. The ease of making reservations online has led to an increase in impulsive bookings, where guests might make a reservation without fully considering their plans or commitments. As a result, they are more likely to cancel these bookings later on.

Additionally, we have observed that a lack of detailed information provided to customers during the booking process contributes to the cancellation trend. When guests do not have a clear understanding of the amenities, services, or specific features of the resort, they may feel compelled to cancel if their expectations are not met upon arrival. Enhancing the clarity and transparency of the booking information can help set appropriate expectations and reduce cancellations.

To address the high cancellation rate, we recommend taking several actions. Firstly, improving the pre-booking experience by providing comprehensive details about the resort, including facilities, activities, and any limitations, can help guests make more informed decisions. Additionally, implementing a stricter cancellation policy or offering incentives for non-refundable bookings can discourage last-minute cancellations and ensure more committed reservations.

Furthermore, enhancing communication with guests after the booking, such as sending reminders closer to the arrival date and providing personalized recommendations, can help reinforce their commitment to the reservation and increase guest satisfaction.

By addressing these factors and implementing measures to reduce cancellations, we can improve the overall occupancy rates, revenue, and guest experience at our resort properties.

In [ ]: