

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

In [157]:

```
df = pd.read_csv('E:\PORTFOLIO-PROJECTS\hotel_booking.csv')
```

In [135]:

```
df.head(3)
```

Out[135]:

	hotel	is_cancelled	lead_time	arrival_date	arrival_date_week_number	arrival_date_of_month	stays_in_weekend_nights	stays_in_week_nights	adults	children	couple	required_car_spaces	total_of_special_requests	reservation_status	reservation_status_date	name	email	phone_number	credit_card
0	Resort Hotel	0	342	2015 Jul y	27	1	0	0	2	0	0	0	0	Checked Out	2015-07-01	Ernest Barnes31@outlook.com		6697921661	** ** ** ** ** 22
1	Resort Hotel	0	737	2015 Jul y	27	1	0	0	2	0	0	0	0	Checked Out	2015-07-01	Andrea Baker94@ao1.com		858-637-6955	** ** ** ** ** 9157
2	Reso	0	7	2015 Jul y	27	1	0	1	1	0	0	0	0	Checked -	2015-	Rebecca Parker		652-	** ** ** **

3 rows \times 36 columns

Out[13]:

In [136]:

Data columns (total 36 columns):

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 [15]:

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

In [16]:

```
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

	hotel	arrival_date_month	meal	country	market_segment	distribution_channel	reserved_room_type	assigned_room_type	deposit_type	customer_type	reservation_status	reservation_status_date	name	email	phone_number	credit_card
nt	90		90	02									90		90	
unique	2	12	5	177	8	5	10	12	3	4	3	926	81503	115889	119390	9000
top	City Hotel	August	BB	PR	Online TA	TA/TO	A	A	No Deposit	Transient	Check-Out	2015-10-21	Michael Johnson	Michael.C@ gmail.com	667921661	**** **** **** 4923
frequency	79330	13877	92310	4859	56477	97870	85994	74053	104641	89613	75166	1461	48	6	1	28

In [138]:

```
for col in df.describe(include = 'object').columns:
    print(col)
    print(df[col].unique())
    print('_'*50)
```

```
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']

reservation_status_date

['2015-07-01' '2015-07-02' '2015-07-03' '2015-05-06' '2015-04-22'
'2015-06-23' '2015-07-05' '2015-07-06' '2015-07-07' '2015-07-08'
'2015-05-11' '2015-07-15' '2015-07-16' '2015-05-29' '2015-05-19'
'2015-06-19' '2015-05-23' '2015-05-18' '2015-07-09' '2015-06-02'
'2015-07-13' '2015-07-04' '2015-06-29' '2015-06-16' '2015-06-18'
'2015-06-12' '2015-06-09' '2015-05-26' '2015-07-11' '2015-07-12'
'2015-07-17' '2015-04-15' '2015-05-13' '2015-07-10' '2015-05-20'
'2015-05-12' '2015-07-14' '2015-06-17' '2015-05-01' '2015-03-30'
'2015-07-19' '2015-06-03' '2015-06-26' '2015-05-14' '2015-07-20'
'2015-05-07' '2015-05-28' '2015-04-13' '2015-03-25' '2015-07-21'
'2015-06-27' '2015-07-18' '2015-07-23' '2015-06-08' '2015-06-22']

'2015-06-24'	'2015-03-05'	'2015-06-01'	'2015-04-24'	'2015-07-22'
'2015-05-27'	'2015-04-06'	'2015-04-11'	'2015-07-25'	'2015-07-28'
'2015-07-29'	'2015-06-25'	'2015-07-24'	'2015-06-05'	'2015-06-30'
'2015-06-13'	'2015-06-11'	'2015-07-30'	'2015-07-27'	'2015-04-29'
'2015-06-04'	'2015-07-26'	'2015-08-01'	'2015-08-02'	'2015-06-15'
'2015-04-23'	'2015-07-31'	'2015-05-25'	'2015-08-03'	'2015-04-17'
'2015-08-04'	'2015-08-06'	'2015-05-15'	'2015-05-09'	'2015-03-17'
'2015-05-22'	'2015-08-07'	'2015-04-04'	'2015-08-05'	'2015-08-08'
'2015-08-10'	'2015-05-04'	'2015-06-06'	'2015-08-09'	'2015-08-15'
'2015-08-11'	'2015-03-28'	'2015-08-14'	'2015-08-12'	'2015-08-16'
'2015-05-16'	'2015-08-21'	'2015-08-13'	'2015-08-17'	'2015-04-20'
'2015-08-18'	'2015-08-23'	'2015-08-22'	'2015-08-19'	'2015-08-20'
'2015-08-29'	'2015-03-31'	'2015-05-30'	'2015-08-25'	'2015-04-14'
'2015-08-24'	'2015-03-24'	'2015-05-21'	'2015-08-28'	'2015-08-26'
'2015-08-27'	'2015-08-30'	'2015-08-31'	'2015-09-06'	'2015-09-03'
'2015-09-04'	'2015-09-02'	'2015-09-01'	'2015-09-05'	'2015-06-20'
'2015-09-07'	'2015-09-10'	'2015-09-11'	'2015-09-08'	'2015-09-09'
'2015-09-13'	'2015-09-15'	'2015-04-10'	'2015-01-02'	'2014-11-18'
'2015-09-12'	'2015-09-17'	'2015-09-14'	'2015-04-07'	'2015-09-19'
'2015-09-16'	'2015-09-20'	'2015-01-18'	'2015-10-23'	'2015-01-22'
'2015-01-01'	'2015-09-22'	'2015-09-24'	'2015-09-18'	'2015-09-21'
'2015-09-30'	'2015-09-25'	'2015-09-27'	'2015-09-28'	'2015-10-12'
'2015-09-29'	'2015-09-23'	'2015-10-01'	'2015-09-26'	'2015-04-18'
'2015-10-02'	'2015-10-04'	'2015-10-08'	'2015-10-03'	'2015-10-07'
'2015-10-09'	'2015-10-11'	'2015-10-05'	'2015-10-06'	'2015-10-10'
'2015-10-14'	'2015-10-15'	'2015-10-18'	'2015-10-13'	'2015-10-20'
'2015-10-19'	'2015-10-31'	'2015-10-16'	'2015-10-21'	'2015-10-22'
'2015-10-17'	'2015-10-24'	'2015-10-25'	'2015-10-28'	'2015-10-27'
'2015-10-26'	'2015-10-30'	'2015-11-05'	'2015-10-29'	'2015-11-03'
'2015-11-07'	'2015-11-04'	'2015-11-01'	'2015-11-02'	'2015-11-17'
'2015-11-06'	'2015-11-10'	'2015-11-08'	'2015-11-09'	'2015-11-15'
'2015-11-16'	'2015-11-11'	'2015-11-12'	'2015-11-14'	'2015-11-13'
'2015-11-18'	'2015-11-22'	'2015-11-19'	'2015-11-21'	'2015-11-20'
'2015-11-24'	'2015-11-25'	'2015-11-23'	'2015-11-28'	'2015-11-26'
'2015-11-27'	'2015-11-29'	'2015-12-04'	'2015-12-01'	'2015-12-06'
'2015-12-08'	'2015-12-02'	'2015-12-03'	'2015-12-31'	'2015-12-05'
'2015-12-10'	'2015-12-17'	'2015-11-30'	'2015-12-12'	'2015-12-07'
'2016-01-05'	'2015-12-11'	'2015-12-13'	'2015-12-15'	'2015-12-16'
'2015-12-19'	'2015-12-18'	'2015-12-26'	'2015-12-27'	'2015-12-22'
'2015-12-23'	'2015-12-24'	'2015-12-29'	'2015-12-28'	'2015-12-20'
'2015-12-30'	'2016-01-02'	'2016-01-01'	'2015-12-25'	'2016-01-03'
'2016-01-04'	'2016-01-11'	'2016-01-07'	'2015-12-21'	'2016-01-09'
'2016-01-10'	'2016-01-08'	'2016-01-06'	'2016-01-12'	'2016-01-13'
'2016-01-23'	'2016-02-09'	'2016-01-15'	'2016-01-16'	'2016-01-17'
'2016-01-19'	'2016-01-18'	'2016-01-21'	'2016-01-24'	'2016-01-22'
'2016-01-29'	'2016-01-27'	'2016-01-25'	'2016-03-08'	'2016-01-26'
'2016-01-20'	'2016-01-30'	'2016-02-01'	'2016-02-02'	'2016-02-08'
'2016-02-07'	'2016-01-28'	'2016-02-05'	'2016-02-03'	'2016-02-13'
'2016-02-10'	'2016-02-04'	'2016-02-12'	'2016-02-11'	'2016-02-16'
'2016-02-14'	'2016-02-15'	'2016-02-20'	'2016-02-06'	'2016-01-14'
'2016-02-17'	'2016-02-21'	'2016-02-24'	'2016-02-25'	'2016-02-19'
'2016-02-18'	'2016-02-26'	'2016-02-23'	'2016-03-05'	'2016-02-22'
'2016-02-27'	'2016-03-03'	'2016-03-24'	'2016-03-04'	'2016-02-29'

'2016-03-01'	'2016-03-02'	'2016-03-30'	'2016-03-07'	'2016-03-14'
'2016-03-21'	'2016-03-09'	'2016-03-12'	'2016-03-22'	'2016-03-10'
'2016-03-11'	'2016-03-20'	'2016-03-15'	'2016-03-17'	'2016-03-16'
'2016-03-19'	'2016-03-27'	'2016-03-18'	'2016-03-26'	'2016-03-31'
'2016-03-28'	'2016-03-29'	'2016-04-01'	'2016-03-23'	'2016-04-02'
'2016-03-25'	'2016-03-13'	'2016-04-04'	'2016-04-03'	'2016-04-05'
'2016-04-08'	'2016-04-06'	'2016-04-09'	'2016-04-12'	'2016-04-16'
'2016-04-17'	'2016-04-27'	'2016-04-14'	'2016-04-18'	'2016-04-21'
'2016-04-19'	'2016-04-20'	'2016-04-10'	'2016-04-13'	'2016-04-11'
'2016-04-07'	'2016-04-15'	'2016-04-22'	'2016-04-23'	'2016-04-26'
'2016-04-28'	'2016-04-24'	'2016-04-25'	'2016-04-29'	'2016-04-30'
'2016-05-01'	'2016-05-10'	'2016-05-02'	'2016-05-07'	'2016-05-08'
'2016-05-12'	'2016-05-04'	'2016-05-06'	'2016-05-03'	'2016-05-09'
'2016-05-05'	'2016-05-13'	'2016-05-14'	'2016-05-18'	'2016-05-19'
'2016-05-15'	'2016-05-16'	'2016-05-11'	'2016-05-21'	'2016-05-22'
'2016-05-20'	'2016-05-24'	'2016-05-25'	'2016-05-26'	'2016-05-23'
'2016-05-27'	'2016-05-17'	'2016-05-29'	'2016-05-28'	'2016-05-30'
'2016-05-31'	'2016-06-01'	'2016-06-03'	'2016-06-08'	'2016-06-02'
'2016-06-05'	'2016-06-06'	'2016-06-13'	'2016-06-07'	'2016-06-10'
'2016-06-11'	'2016-06-16'	'2016-06-12'	'2016-06-14'	'2016-06-17'
'2016-06-04'	'2016-06-18'	'2016-06-21'	'2016-06-09'	'2016-06-24'
'2016-06-20'	'2016-06-25'	'2016-06-22'	'2016-06-26'	'2016-06-23'
'2016-07-01'	'2016-06-15'	'2016-06-28'	'2016-07-02'	'2016-06-19'
'2016-06-27'	'2016-07-04'	'2016-06-30'	'2016-07-05'	'2016-07-08'
'2016-07-09'	'2016-07-07'	'2016-07-12'	'2016-06-29'	'2016-07-10'
'2016-07-15'	'2016-07-03'	'2016-07-16'	'2016-07-14'	'2016-07-18'
'2016-07-13'	'2016-07-06'	'2016-07-20'	'2016-07-21'	'2016-07-23'
'2016-07-19'	'2016-07-11'	'2016-07-28'	'2016-07-17'	'2016-07-25'
'2016-07-22'	'2016-07-29'	'2016-08-03'	'2016-08-02'	'2016-08-04'
'2016-08-08'	'2016-08-10'	'2016-08-01'	'2016-08-06'	'2016-03-06'
'2016-08-05'	'2016-07-26'	'2016-08-07'	'2016-07-30'	'2016-07-24'
'2016-08-12'	'2016-07-27'	'2016-08-13'	'2016-08-18'	'2016-08-16'
'2016-08-15'	'2016-08-17'	'2016-08-11'	'2016-07-31'	'2016-08-19'
'2016-09-01'	'2016-08-23'	'2016-08-26'	'2016-08-20'	'2016-08-21'
'2016-09-04'	'2016-08-22'	'2016-08-27'	'2016-08-25'	'2016-08-09'
'2016-09-05'	'2016-08-24'	'2016-09-10'	'2016-08-29'	'2016-09-09'
'2016-08-30'	'2016-09-13'	'2016-08-31'	'2016-09-14'	'2016-09-12'
'2016-09-15'	'2016-08-14'	'2016-09-02'	'2016-09-08'	'2016-09-19'
'2016-09-16'	'2016-09-07'	'2016-09-21'	'2016-09-06'	'2016-09-22'
'2016-09-17'	'2016-09-20'	'2016-09-03'	'2016-09-26'	'2016-09-23'
'2016-09-18'	'2016-09-29'	'2016-10-02'	'2016-10-01'	'2016-09-27'
'2016-09-25'	'2016-10-05'	'2016-09-11'	'2016-09-30'	'2016-10-09'
'2016-10-03'	'2016-10-06'	'2016-10-11'	'2016-09-24'	'2016-10-13'
'2016-09-28'	'2016-10-08'	'2016-10-07'	'2016-10-16'	'2016-08-28'
'2016-10-17'	'2016-10-18'	'2016-10-10'	'2016-10-04'	'2016-10-15'
'2016-10-19'	'2016-10-21'	'2016-10-12'	'2016-10-24'	'2016-10-26'
'2016-10-23'	'2016-10-20'	'2016-10-25'	'2016-10-27'	'2016-10-28'
'2016-10-30'	'2016-10-29'	'2016-11-01'	'2016-11-04'	'2016-10-14'
'2016-11-07'	'2016-11-03'	'2016-11-10'	'2016-11-14'	'2016-11-02'
'2016-10-31'	'2016-11-11'	'2016-11-08'	'2016-11-05'	'2016-11-25'
'2016-11-09'	'2016-11-20'	'2016-11-21'	'2016-10-22'	'2016-11-22'
'2016-11-16'	'2016-11-23'	'2016-11-17'	'2016-11-06'	'2016-11-15'
'2016-11-13'	'2016-11-12'	'2016-11-27'	'2016-11-19'	'2016-11-30'

'2016-11-18'	'2016-12-02'	'2016-12-04'	'2016-11-29'	'2016-12-07'
'2016-11-28'	'2016-12-03'	'2016-12-06'	'2016-11-24'	'2016-12-08'
'2016-12-05'	'2016-12-10'	'2016-12-13'	'2016-12-14'	'2016-12-16'
'2016-12-15'	'2016-12-17'	'2016-12-19'	'2016-12-21'	'2016-12-20'
'2016-12-22'	'2016-12-23'	'2016-12-24'	'2016-12-01'	'2016-12-27'
'2016-12-29'	'2016-12-30'	'2016-12-12'	'2017-01-02'	'2016-12-11'
'2017-01-03'	'2017-01-04'	'2017-01-01'	'2016-12-26'	'2017-01-06'
'2016-12-28'	'2016-12-18'	'2017-01-10'	'2017-01-11'	'2017-01-07'
'2017-01-12'	'2017-01-16'	'2017-01-14'	'2017-01-13'	'2017-01-05'
'2017-01-17'	'2017-01-20'	'2016-12-09'	'2017-01-26'	'2016-12-31'
'2017-01-23'	'2017-01-27'	'2017-01-28'	'2017-01-19'	'2017-01-25'
'2017-01-24'	'2017-01-29'	'2017-01-18'	'2016-12-25'	'2017-01-15'
'2017-01-21'	'2017-02-01'	'2017-02-02'	'2017-01-31'	'2017-02-03'
'2017-02-04'	'2017-02-06'	'2017-02-07'	'2017-02-08'	'2017-01-30'
'2017-02-09'	'2017-01-09'	'2017-02-11'	'2017-02-10'	'2017-02-12'
'2017-02-13'	'2017-02-14'	'2017-02-16'	'2017-02-17'	'2017-02-18'
'2017-02-19'	'2017-02-20'	'2017-02-15'	'2017-02-21'	'2017-02-22'
'2017-02-26'	'2017-02-23'	'2017-02-24'	'2017-02-25'	'2017-02-28'
'2017-03-05'	'2017-02-27'	'2017-03-03'	'2017-03-06'	'2017-03-02'
'2017-03-08'	'2017-03-09'	'2017-03-10'	'2017-03-07'	'2017-03-12'
'2017-03-13'	'2017-03-14'	'2017-03-01'	'2017-03-18'	'2017-03-17'
'2017-03-24'	'2017-03-22'	'2017-03-26'	'2017-03-27'	'2017-03-11'
'2017-03-28'	'2017-03-29'	'2017-03-30'	'2017-03-31'	'2017-03-19'
'2017-01-22'	'2017-04-02'	'2017-03-20'	'2017-04-03'	'2017-01-08'
'2017-03-23'	'2017-04-05'	'2017-02-05'	'2017-04-04'	'2017-03-15'
'2017-04-07'	'2017-03-25'	'2017-04-08'	'2017-04-06'	'2017-03-21'
'2017-04-10'	'2017-04-01'	'2017-04-11'	'2017-04-13'	'2017-04-15'
'2017-04-12'	'2017-03-04'	'2017-04-19'	'2017-04-22'	'2017-04-20'
'2017-05-02'	'2017-04-09'	'2017-04-23'	'2017-04-24'	'2017-04-16'
'2017-04-28'	'2017-04-18'	'2017-04-26'	'2017-04-25'	'2017-04-17'
'2017-04-21'	'2017-05-03'	'2017-05-04'	'2017-03-16'	'2017-05-05'
'2017-04-29'	'2017-04-14'	'2017-05-08'	'2017-04-27'	'2017-05-11'
'2017-05-01'	'2017-05-10'	'2017-05-13'	'2017-05-06'	'2017-05-14'
'2017-05-16'	'2017-04-30'	'2017-05-15'	'2017-05-07'	'2017-05-09'
'2017-05-17'	'2017-05-21'	'2017-05-12'	'2017-05-22'	'2017-05-24'
'2017-05-23'	'2017-05-25'	'2017-05-26'	'2017-05-28'	'2017-05-27'
'2017-05-29'	'2017-05-19'	'2017-05-31'	'2017-05-20'	'2017-06-01'
'2017-05-30'	'2017-06-02'	'2016-11-26'	'2017-06-04'	'2017-06-05'
'2017-06-06'	'2017-06-07'	'2017-05-18'	'2017-06-09'	'2017-06-10'
'2017-06-11'	'2017-06-12'	'2017-06-14'	'2017-06-08'	'2017-06-16'
'2017-06-13'	'2017-06-03'	'2017-06-24'	'2017-06-20'	'2017-06-19'
'2017-06-21'	'2017-06-26'	'2017-06-27'	'2017-06-22'	'2017-06-28'
'2017-06-15'	'2017-06-29'	'2017-06-30'	'2017-06-18'	'2017-07-04'
'2017-07-08'	'2017-07-05'	'2017-07-03'	'2017-07-07'	'2017-07-01'
'2017-07-06'	'2017-07-11'	'2017-07-12'	'2017-06-23'	'2017-07-13'
'2017-07-02'	'2017-07-10'	'2017-07-14'	'2017-07-15'	'2017-07-16'
'2017-07-18'	'2017-07-17'	'2017-07-19'	'2017-07-20'	'2017-07-21'
'2017-06-25'	'2017-06-17'	'2017-07-24'	'2017-07-26'	'2017-07-09'
'2017-07-27'	'2017-07-28'	'2017-07-31'	'2017-07-29'	'2017-07-22'
'2017-08-02'	'2017-08-01'	'2017-08-03'	'2017-08-04'	'2017-07-25'
'2017-07-23'	'2017-08-09'	'2017-08-10'	'2017-07-30'	'2017-08-07'
'2017-08-13'	'2017-08-05'	'2017-08-14'	'2017-08-08'	'2017-08-16'
'2017-08-17'	'2017-08-15'	'2017-08-18'	'2017-08-20'	'2017-08-22'

```
'2017-08-06' '2017-08-25' '2017-08-26' '2017-08-23' '2017-08-11'
'2017-08-27' '2017-08-21' '2017-08-29' '2017-08-31' '2017-08-12'
'2017-08-19' '2016-01-31' '2017-09-01' '2017-08-28' '2015-04-03'
'2015-01-21' '2015-01-28' '2015-01-29' '2015-01-30' '2015-02-02'
'2015-02-05' '2015-02-06' '2015-02-09' '2015-02-10' '2015-02-11'
'2015-02-12' '2015-02-19' '2015-02-20' '2015-02-23' '2015-02-24'
'2015-02-25' '2015-02-26' '2015-02-27' '2015-03-03' '2015-03-04'
'2015-03-06' '2015-03-09' '2015-03-11' '2015-03-12' '2015-03-18'
'2015-04-02' '2015-06-14' '2015-04-08' '2015-04-16' '2015-04-25'
'2015-04-28' '2015-05-08' '2017-09-06' '2016-02-28' '2015-12-09'
'2015-12-14' '2017-09-09' '2017-09-02' '2017-08-24' '2017-08-30'
'2017-09-03' '2017-09-04' '2017-09-05' '2017-09-07' '2017-09-08'
'2017-09-10' '2017-09-12' '2017-09-14' '2015-04-30' '2015-04-21'
'2015-04-05' '2015-03-13' '2015-05-05' '2015-03-29' '2015-06-10'
'2015-04-27' '2014-10-17' '2015-01-20' '2015-02-17' '2015-03-10'
'2015-03-23']
```

name

```
['Ernest Barnes' 'Andrea Baker' 'Rebecca Parker' ... 'Wesley Aguilar'
 'Caroline Conley MD' 'Ariana Michael']
```

email

```
['Ernest.Barnes31@outlook.com' 'Andrea_Baker94@aol.com'
 'Rebecca_Parker@comcast.net' ... 'Mary_Morales@hotmail.com'
 'MD_Caroline@comcast.net' 'Ariana_M@xfinity.com']
```

phone-number

```
['669-792-1661' '858-637-6955' '652-885-2745' ... '395-518-4100'
 '531-528-1017' '422-804-6403']
```

credit_card

```
['*****4322' '*****9157' '*****3734' ...
 '*****9170' '*****6349' '*****7959']
```

df.isnull().sum()

In [140]:

```
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
```

Out[140]:

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 [141]:

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

In [142]:

```
df.dropna(inplace = True)
```

In [25]:

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

Out[25]:

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 [143]:

```
df.describe()
```

Out[143]:

	is_canceled	length_of_stay	arrival_date_year	arrival_date_week_number	arrival_date_month	stays_in_weekend_nights	stays_in_week_nights	adults	children	babies	is_repeat_guest	previous_cancellations	previous_bookings_not_canceled	bookings_in_changes	days_in_waiting_list	adr	required_car_parking_spaces	total_of_special_requests
count	1188	1188	1188	1188	1188	1188	1188	1188	1188	1188	1188	1188	1188	1188	1188	1188	1188	1188
mean	0.000000	10.35	16.76	27.155	15.880	0.927	2.545	1.8591	0.1007	0.0048	0.0321	0.87142	0.131634	0.2281	2.304	0.0600	0.061885	0.571683
std	0.000000	6.903309	0.70459	13.589971	8.780324	0.996216	1.900168	0.578576	0.399172	0.097380	0.176029	0.845869	1.484672	0.652785	17.630452	50.485862	0.244172	0.792678
min	0.000000	0.000000	15.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-6.380000	0.000000	0.000000

	is_canceled	lead_time	arrival_date	arrival_date_week_number	arrival_date_of_month	stay_in_week_nights	stay_in_week_nights	adults	children	babies	is_repeat_guest	previous_cancellations	previous_bookings_not_cancelled	bookings_in_changes	days_in_waiting_list	adr	required_car_parking_spaces	total_of_special_requests
25%	0.0000	18.0000	20.1600	16.0000	8.0000	0.0000	1.0000	2.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	70.0000	0.0000	0.0000
50%	0.0000	69.0000	20.1600	28.0000	16.0000	1.0000	2.0000	2.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	95.0000	0.0000	0.0000
75%	1.0000	16.0000	20.1700	38.0000	23.0000	2.0000	3.0000	2.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	12.6000	0.0000	1.0000
max	1.0000	73.0000	20.1700	53.0000	31.0000	16.0000	41.0000	55.0000	10.0000	10.0000	1.0000	26.0000	72.0000	21.0000	39.0000	54.0000	8.0000	5.0000

```
In [144]:
df = df[df['adr']<5000]
```

Data Analysis and Visualizations

```
In [145]:
cancelled_perc = df['is_canceled'].value_counts(normalize = True)
cancelled_perc
```

```
Out[145]:
0    0.628653
1    0.371347
Name: is_canceled, dtype: float64
```

```
In [146]:
cancelled_perc = df['is_canceled'].value_counts(normalize = True)
print(cancelled_perc)
```

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

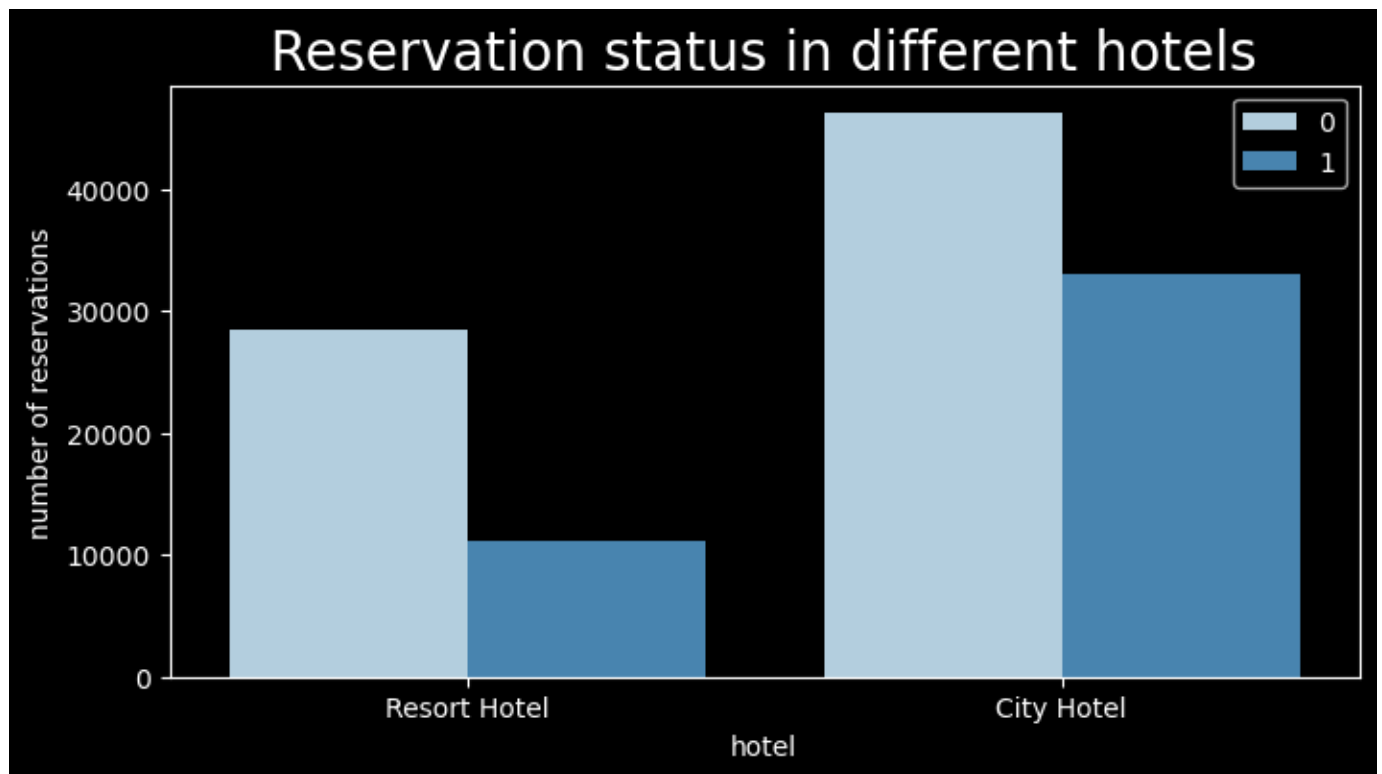
In [158]:

```
plt.figure(figsize = (5,4))
plt.title('Reservation Status Count')
plt.bar(['Not canceled', 'canceled'],
df['is_canceled'].value_counts(),edgecolor = 'white', width = 0.7)
plt.show()
```



In [148]:

```
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 [149]:

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

Out[149]:

```
0    0.72025
1    0.27975
Name: is_canceled, dtype: float64
```

In [150]:

```
City_hotel = df[df['hotel'] == 'City hotel']
City_hotel['is_canceled'].value_counts(normalize = True)
```

Out[150]:

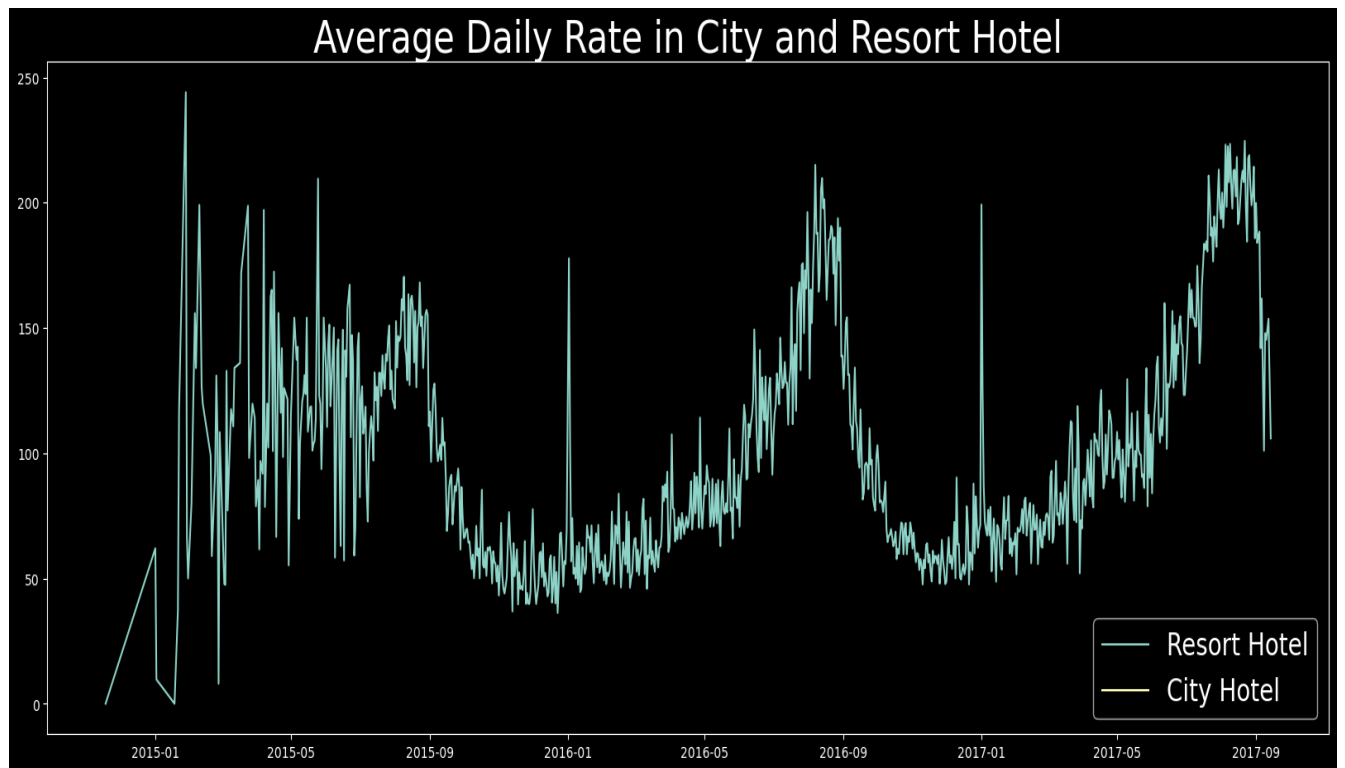
```
Series([], Name: is_canceled, dtype: float64)
```

In [151]:

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

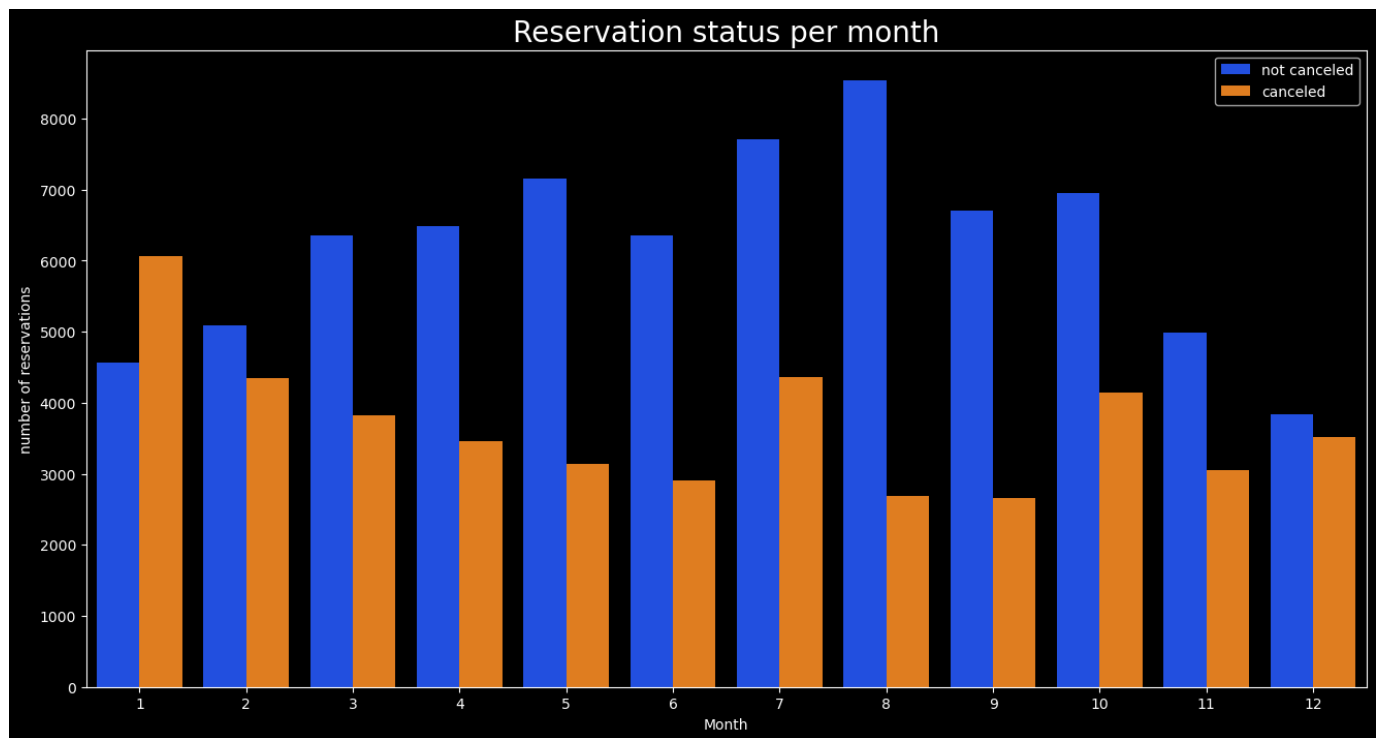
In [118]:

```
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(fontsize = 20)
plt.show()
```



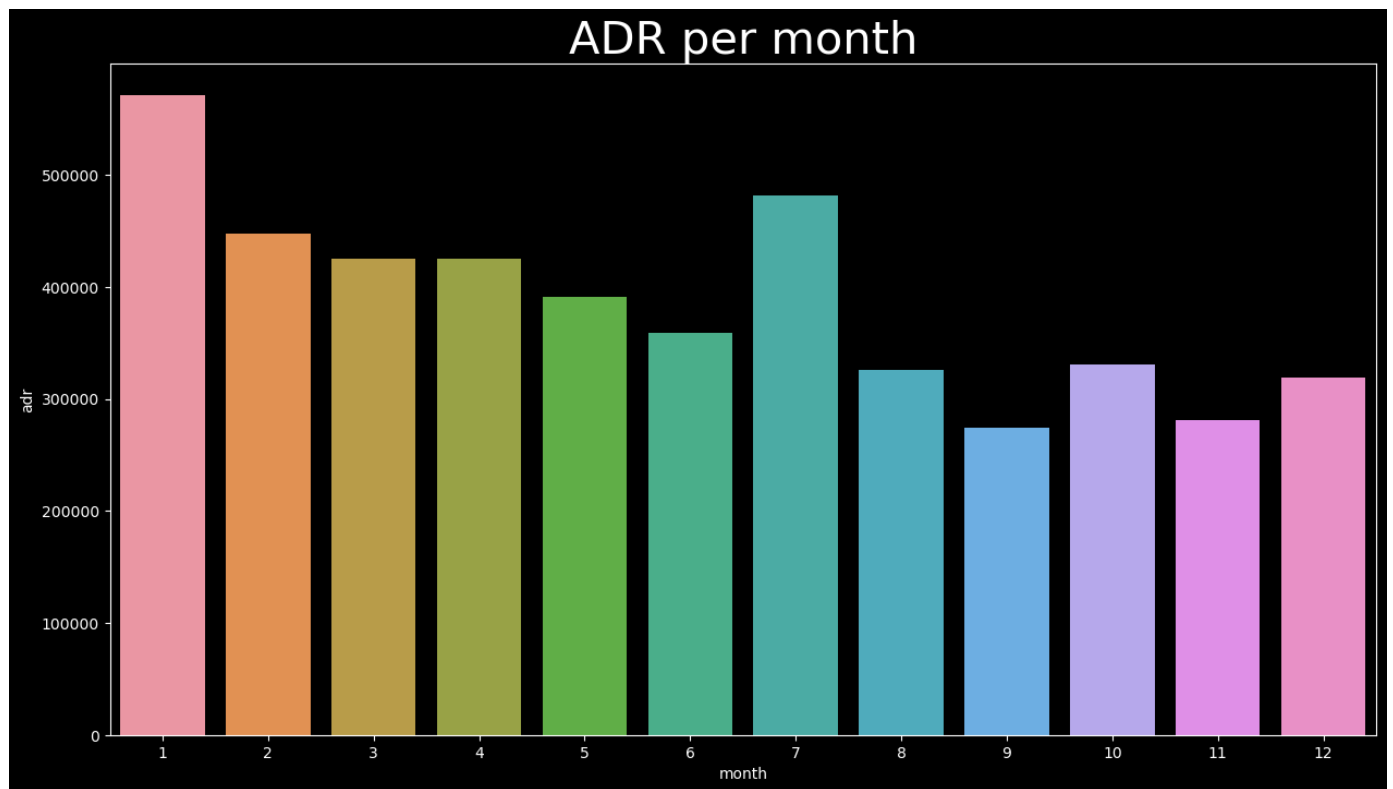
In [125]:

```
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 reservations')
plt.legend(['not canceled', 'canceled'])
plt.show()
```

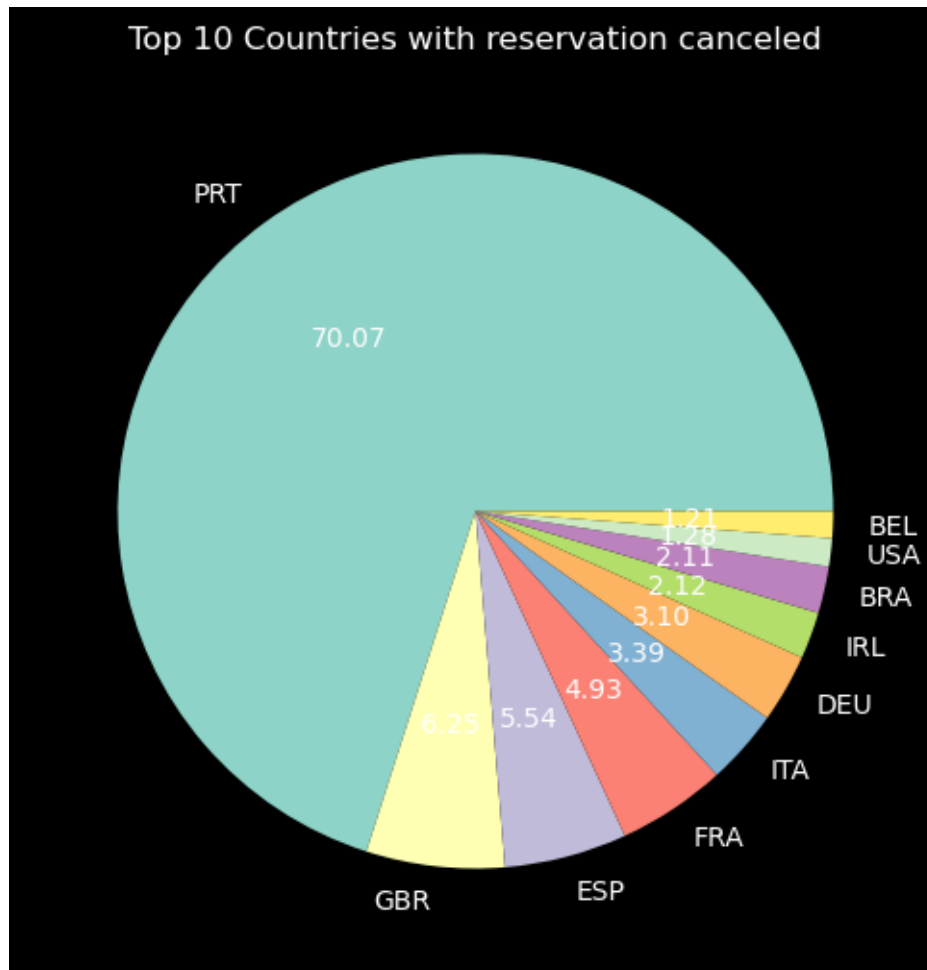
In [126]:

```
plt.figure(figsize = (15,8))
plt.title('ADR per month', fontsize = 30)
sns.barplot('month', 'adr', data = df[df['is_canceled'] ==
1].groupby('month')[['adr']].sum().reset_index())
plt.show()
```



In [155]:

```
cancelled_data = df[df['is_canceled'] == 1]
top_10_country = cancelled_data['country'].value_counts()[:10]
plt.figure(figsize = (6,8))
plt.title('Top 10 Countries with reservation canceled')
plt.pie(top_10_country, autopct = '%.2f', labels = top_10_country.index)
plt.show()
```



In [75]:

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

Out[75]:

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

In [76]:

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

Out[76]:

```
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 [130]:

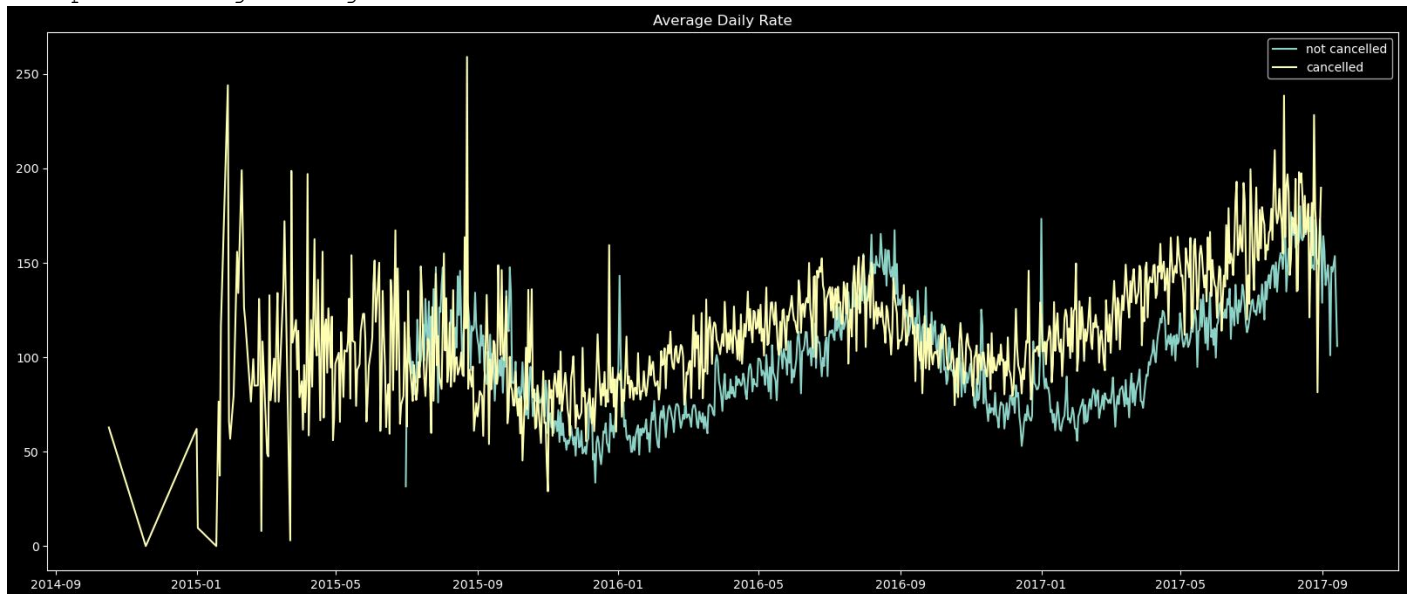
```
cancelled_df_adr =
cancelled_data.groupby('reservation_status_date')[['adr']].mean()
cancelled_df_adr.reset_index(inplace=True)
cancelled_df_adr.sort_values('reservation_status_date', inplace=True)

not_cancelled_data = df[df['is_canceled'] == 0]
not_cancelled_df_adr =
not_cancelled_data.groupby('reservation_status_date')[['adr']].mean()
not_cancelled_df_adr.reset_index(inplace = True)
not_cancelled_df_adr.sort_values('reservation_status_date', inplace=True)

plt.figure(figsize=(20, 8))
plt.title('Average Daily Rate')
plt.plot(not_cancelled_df_adr['reservation_status_date'],
not_cancelled_df_adr['adr'], label='not cancelled')
plt.plot(cancelled_df_adr['reservation_status_date'],
cancelled_df_adr['adr'], label='cancelled')
plt.legend()
```

Out[130]:

<matplotlib.legend.Legend at 0x25782f1cdf0>



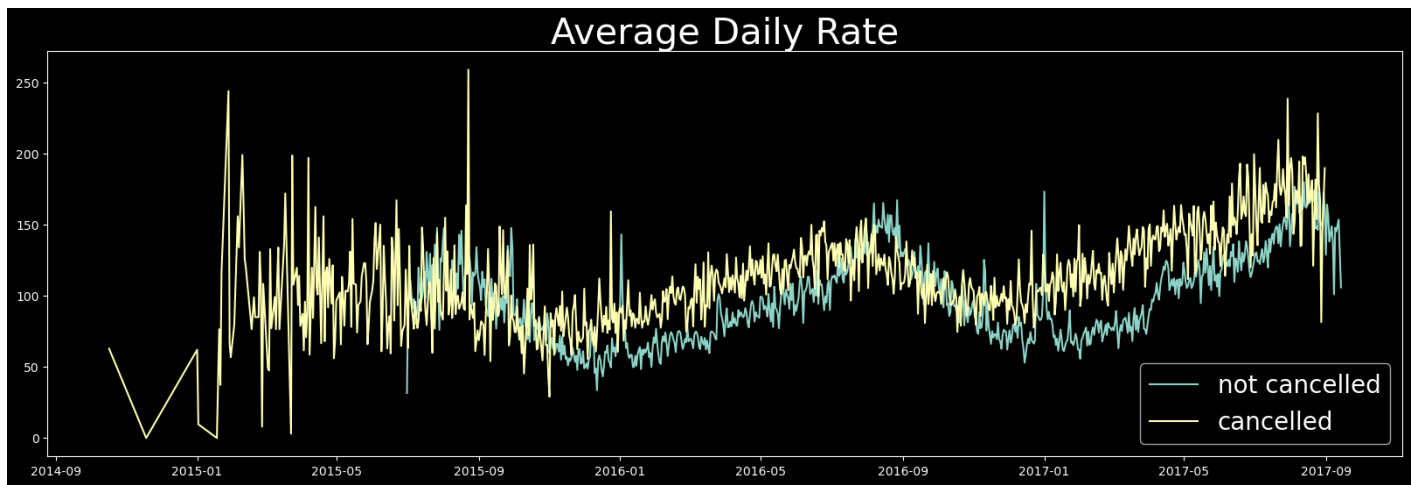
In [108]:

```
cancelled_df_adr =
cancelled_df_adr[(cancelled_df_adr['reservation_status_date']>'2016') &
(cancelled_df_adr['reservation_status_date']<'2017-09')]
not_cancelled_df_adr =
not_cancelled_df_adr[(not_cancelled_df_adr['reservation_status_date']>'2016')
& (not_cancelled_df_adr['reservation_status_date']<'2017-09')]
```

In [131]:

```
plt.figure(figsize = (20, 6))
plt.title('Average Daily Rate', fontsize = 30)
plt.plot(not_cancelled_df_adr['reservation_status_date'],not_cancelled_df_adr
['adr'],label = 'not cancelled')
```

```
plt.plot(cancelled_df_adr['reservation_status_date'],cancelled_df_adr['adr'],  
label = 'cancelled')  
plt.legend(fontsize = 20)  
plt.show()
```



In []: