

# Hotel Business Analytics

## Business Problem

Higher Cancellation rates.

## Assumptions

1. Hotels are not currently using any suggested solutions.
2. We assume that the data which we are using to analyse the issue is efficient.
3. The biggest factor affecting the business is booking cancellation.
4. Clients make hotel reservations the same year they make cancellations.

## Research Question

1. What are the variables that affect hotel reservation cancellations?
2. How to assist hotels in making pricing and promotional decisions?

## Hypothesis

1. More cancellations occur when prices are higher.
2. Longer waiting list also increasing cancellations.

## Analysis

Libraries we are using for this analysis: pandas,numpy,matplotlib & seaborn.

```
In [1]: pip install pandas
```

```
Requirement already satisfied: pandas in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (2.0.1)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: tzdata>=2022.1 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: numpy>=1.21.0 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from pandas) (1.24.3)
Requirement already satisfied: six>=1.5 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

```
In [2]: pip install numpy
```

```
Requirement already satisfied: numpy in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (1.24.3)
Note: you may need to restart the kernel to use updated packages.
```

```
In [3]: pip install matplotlib
```

Requirement already satisfied: matplotlib in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (3.7.1)  
Requirement already satisfied: contourpy>=1.0.1 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib) (1.0.7)  
Requirement already satisfied: cycler>=0.10 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib) (0.11.0)  
Requirement already satisfied: fonttools>=4.22.0 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib) (4.39.4)  
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib) (1.4.4)  
Requirement already satisfied: numpy>=1.20 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib) (1.24.3)  
Requirement already satisfied: packaging>=20.0 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib) (23.1)  
Requirement already satisfied: pillow>=6.2.0 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib) (9.5.0)  
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib) (3.0.9)  
Requirement already satisfied: python-dateutil>=2.7 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib) (2.8.2)  
Requirement already satisfied: six>=1.5 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)  
Note: you may need to restart the kernel to use updated packages.

In [4]: `pip install seaborn`

Requirement already satisfied: seaborn in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (0.12.2)  
Requirement already satisfied: numpy!=1.24.0,>=1.17 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from seaborn) (1.24.3)  
Requirement already satisfied: pandas>=0.25 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from seaborn) (2.0.1)  
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from seaborn) (3.7.1)  
Requirement already satisfied: contourpy>=1.0.1 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.0.7)  
Requirement already satisfied: cycler>=0.10 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (0.11.0)  
Requirement already satisfied: fonttools>=4.22.0 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (4.39.4)  
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.4.4)  
Requirement already satisfied: packaging>=20.0 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (23.1)  
Requirement already satisfied: pillow>=6.2.0 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (9.5.0)  
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (3.0.9)  
Requirement already satisfied: python-dateutil>=2.7 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (2.8.2)  
Requirement already satisfied: pytz>=2020.1 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from pandas>=0.25->seaborn) (2023.3)  
Requirement already satisfied: tzdata>=2022.1 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from pandas>=0.25->seaborn) (2023.3)  
Requirement already satisfied: six>=1.5 in c:\users\hites\appdata\local\programs\python\python311\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)  
Note: you may need to restart the kernel to use updated packages.

In [5]: `import pandas as pd  
import numpy as np  
import seaborn as sns  
import matplotlib.pyplot as plt`

## Loadind the dataset

In [6]: `df = pd.read_csv(r"C:\Users\hites\Downloads\archive\hotel_booking.csv",encoding='unicode escape')`

## Data Cleaning

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

Out[7]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stay
--	-------	-------------	-----------	-------------------	--------------------	--------------------------	---------------------------	------

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

5 rows × 36 columns

In [8]: df.shape

Out[8]: (119390, 36)

In [9]: df.columns

Out[9]: 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 [10]: df.drop(['credit\_card', 'email', 'name', 'phone-number'], axis=1, inplace=True)

In [11]: df

Out[11]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month
0	Resort Hotel	0	342	2015	July	27	1
1	Resort Hotel	0	737	2015	July	27	1
2	Resort Hotel	0	7	2015	July	27	1
3	Resort Hotel	0	13	2015	July	27	1
4	Resort Hotel	0	14	2015	July	27	1
...	...	...	...	...	...	...	...
119385	City Hotel	0	23	2017	August	35	30
119386	City Hotel	0	102	2017	August	35	31
119387	City Hotel	0	34	2017	August	35	31
119388	City Hotel	0	109	2017	August	35	31
119389	City Hotel	0	205	2017	August	35	29

119390 rows × 32 columns

In [12]: df.shape

Out[12]: (119390, 32)

In [13]: df.columns

Out[13]: 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 [14]: 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 [15]: import datetime  
df['reservation\_status\_date'] = pd.to\_datetime(df['reservation\_status\_date'])

In [16]: df.describe(include = 'object')

Out[16]:

	hotel	arrival_date_month	meal	country	market_segment	distribution_channel	reserved_room_type	assigned_room_type
count	119390	119390	119390	118902	119390	119390	119390	119390
unique	2	12	5	177	8	5	10	10
top	City Hotel	August	BB	PRT	Online TA	TA/TO	A	A
freq	79330	13877	92310	48590	56477	97870	85994	74330

In [17]: for i in (df.describe(include = 'object')):  
print(i)  
print(df[i].unique())  
print('-----')

```

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 [18]: pd.isnull(df).sum()
```

```
Out[18]: 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 [19]: df.drop(['agent', 'company'], axis=1, inplace=True)
```

```
In [20]: df.dropna(subset = ['children'], inplace = True)
```

```
In [21]: df.dropna(subset = ['country'], inplace = True)
```

```
In [22]: pd.isnull(df).sum()
```

```
Out[22]: 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 [23]: df.describe()
```

Out[23]:

	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_ni
count	118898.000000	118898.000000	118898.000000	118898.000000	118898.000000	118898.00
mean	0.371352	104.311435	2016.157656	27.166555	15.800880	0.92
min	0.000000	0.000000	2015.000000	1.000000	1.000000	0.00
25%	0.000000	18.000000	2016.000000	16.000000	8.000000	0.00
50%	0.000000	69.000000	2016.000000	28.000000	16.000000	1.00
75%	1.000000	161.000000	2017.000000	38.000000	23.000000	2.00
max	1.000000	737.000000	2017.000000	53.000000	31.000000	16.00
std	0.483168	106.903309	0.707459	13.589971	8.780324	0.99

--	--	--

In [24]:

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

In [25]:

```
df= df[df['adr']>0]
```

In [26]:

```
df.describe()
```

Out[26]:

	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_ni
count	116958.000000	116958.000000	116958.000000	116958.000000	116958.000000	116958.00
mean	0.375767	105.376879	2016.161443	27.138109	15.805050	0.93
min	0.000000	0.000000	2015.000000	1.000000	1.000000	0.00
25%	0.000000	19.000000	2016.000000	16.000000	8.000000	0.00
50%	0.000000	71.000000	2016.000000	27.000000	16.000000	1.00
75%	1.000000	162.000000	2017.000000	38.000000	23.000000	2.00
max	1.000000	709.000000	2017.000000	53.000000	31.000000	16.00
std	0.484322	106.944356	0.706509	13.561162	8.783049	0.99

--	--	--

## Data Analysis & Visualization

In [27]:

```
df['is_canceled']
```

Out[27]:

```
2      0
3      0
4      0
5      0
6      0
..
119385  0
119386  0
119387  0
119388  0
119389  0
Name: is_canceled, Length: 116958, dtype: int64
```

In [28]:

```
m = 0
n = 0
for i in df['is_canceled']:
    if i == 0:
        m = m+1
    elif i == 1:
        n = n+1

print(m , n )
print('Not Cancellation %: ',m/(m+n)*100,'%')
print('Cancellation %: ',n/(m+n)*100,'%')
```

73009 43949  
Not Cancellation %: 62.423263051693766 %  
Cancellation %: 37.57673694830623 %

Not Cancellation %: 62.423263051693766 % || Cancellation %: 37.57673694830623 %

In [29]:

```
x = ['Not Cancelled','Cancelled']
y = [m,n]

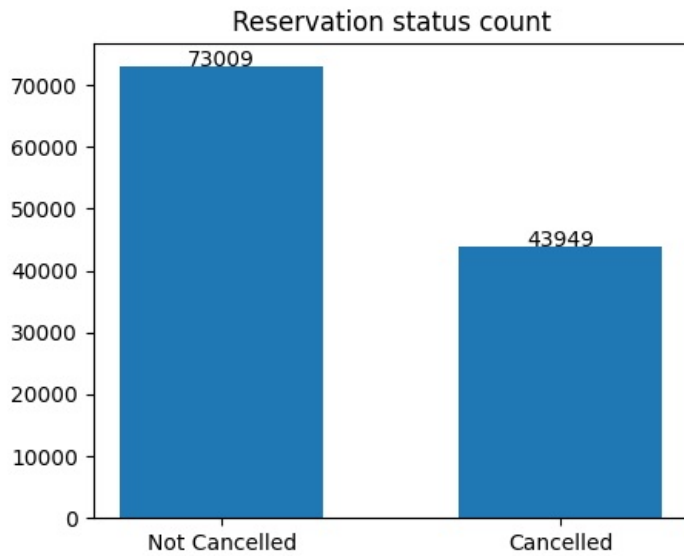
def addlabels(x,y):
```

```

for i in range(len(x)):
    plt.text(i,y[i],y[i], ha = 'center')

plt.figure(figsize = (5,4))
plt.bar(x,y,width = 0.6)
addlabels(x,y)
plt.title('Reservation status count')
plt.show()

```



```

In [30]: sns.countplot(x = 'hotel',hue = 'is_canceled',data =df)
plt.legend(['not cancelled','cancelled'])
plt.title('Hotel wise cancellation')
plt.show()

```



```

In [31]: Resort_hotel = df[df['hotel']=='Resort Hotel']
Resort_hotel['is_canceled'].value_counts()

```

```

Out[31]: is_canceled
0      27862
1      10995
Name: count, dtype: int64

```

```

In [32]: City_hotel = df[df['hotel']=='City Hotel']
City_hotel['is_canceled'].value_counts()

```

```

Out[32]: is_canceled
0      45147
1      32954
Name: count, dtype: int64

```

```

In [33]: Resort_hotel['is_canceled'].value_counts(normalize = True)

```



```
Out[33]: is_canceled
0      0.717039
1      0.282961
Name: proportion, dtype: float64
```

Resort Hotel cancellation % is around 29%

```
In [34]: City_hotel['is_canceled'].value_counts(normalize = True)
```

```
Out[34]: is_canceled
0      0.578059
1      0.421941
Name: proportion, dtype: float64
```

City Hotel cancellation % is around 42%

```
In [35]: Resort_hotel['adr'].describe()
```

```
Out[35]: count      38857.000000
mean         97.161082
std          60.641131
min           0.260000
25%          52.400000
50%          77.500000
75%         127.000000
max          508.000000
Name: adr, dtype: float64
```

```
In [36]: City_hotel['adr'].describe()
```

```
Out[36]: count      78101.000000
mean        106.876989
std          37.383940
min           0.500000
25%          80.000000
50%         100.000000
75%         126.000000
max          510.000000
Name: adr, dtype: float64
```

Max & Min price of both the hotels is almost equal and mean price is also almost equal.

```
In [37]: x = df.groupby(['is_canceled'])
y = x['adr'].describe()
print(y)
```

	count	mean	std	min	25%	50%	75%	max
is_canceled								
0	73009.0	102.593502	47.193325	0.26	70.0	94.5	126.00	510.0
1	43949.0	105.402601	45.670160	0.50	73.8	96.4	127.93	450.0

Overall Mean price was high when cancellation done.

```
In [38]: x = City_hotel.groupby(['is_canceled'])
y = x['adr'].describe()
print(y)
```

	count	mean	std	min	25%	50%	75%	max
is_canceled								
0	45147.0	108.274054	37.600913	1.0	80.75	100.2	126.9	510.0
1	32954.0	104.963010	36.999653	0.5	76.93	100.0	125.8	352.5

```
In [39]: x = Resort_hotel.groupby(['is_canceled'])
y = x['adr'].describe()
print(y)
```

	count	mean	std	min	25%	50%	75%
is_canceled							
0	27862.0	93.388857	58.380442	0.26	50.050	74.25	120.00 \
1	10995.0	106.720132	65.054706	4.00	57.265	84.00	146.05

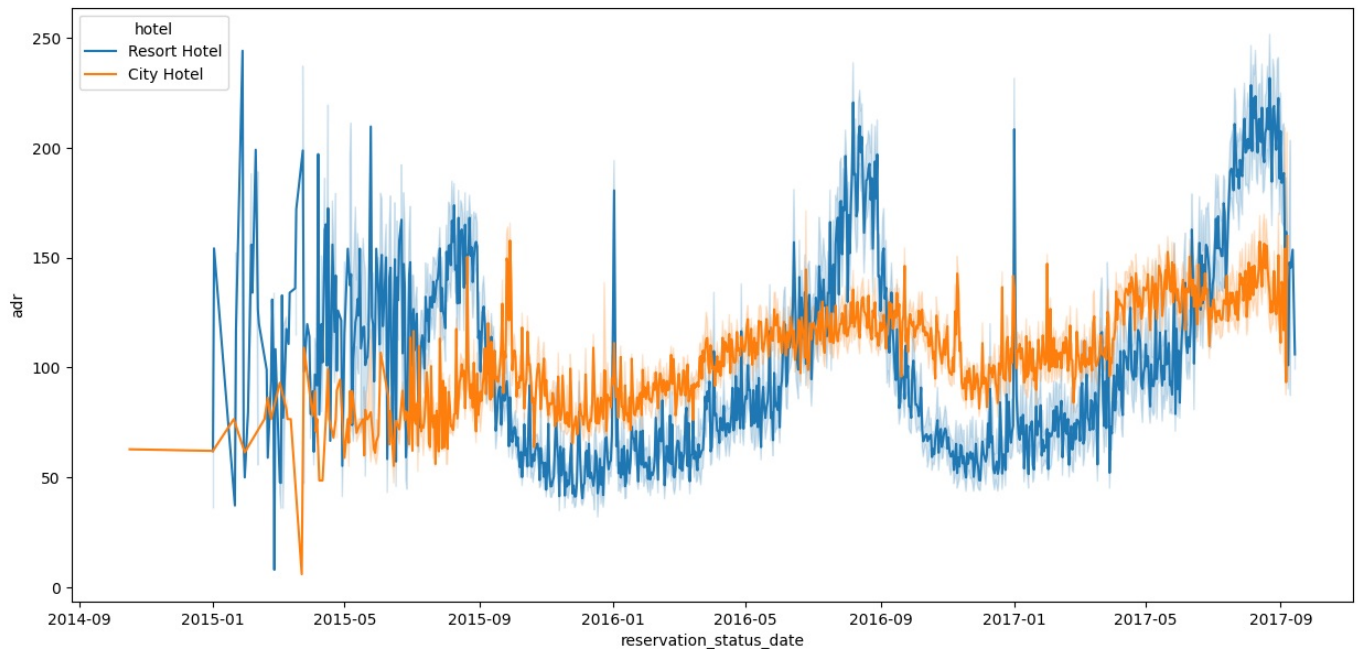
  

	max
is_canceled	
0	508.0
1	450.0

It is found that Cancellation occurred when overall mean price was high. But no such relation found when we compared hotel wise mean price.

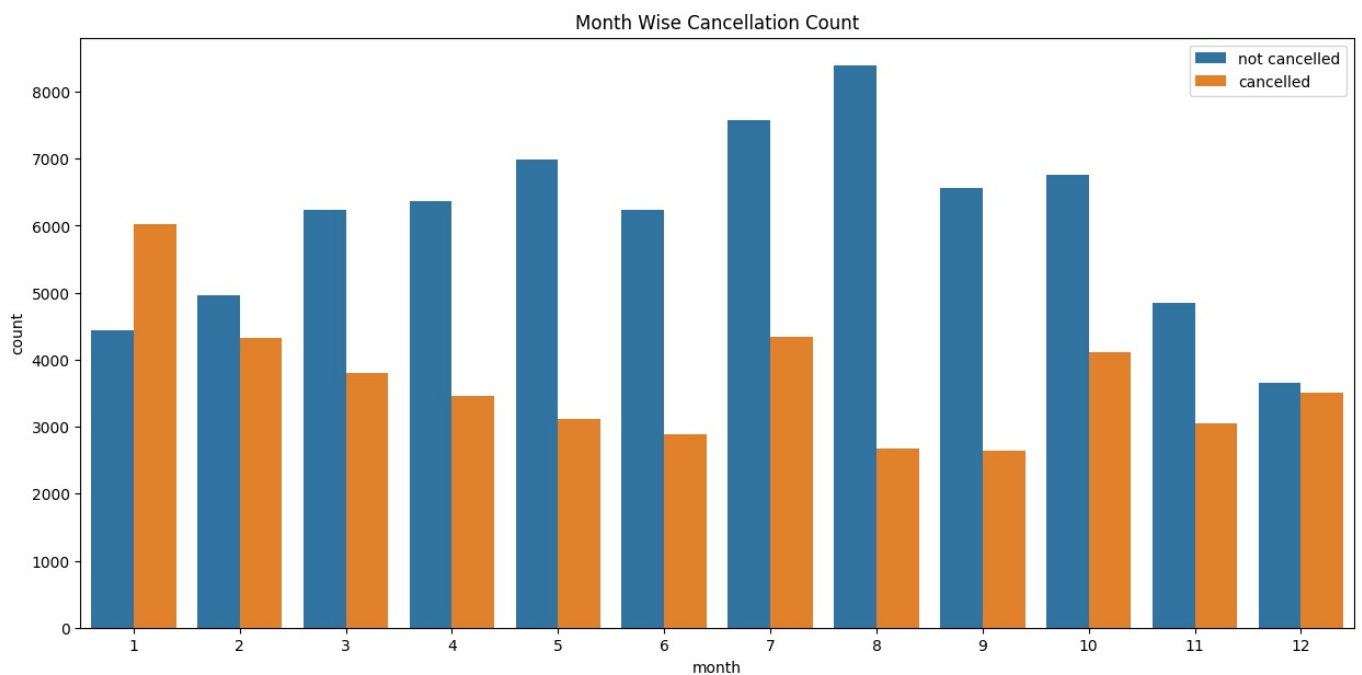
```
In [66]: plt.figure(figsize = (15,7))
sns.lineplot(x = 'reservation_status_date', y = 'adr', hue='hotel', data = df)
```

```
plt.show()
```



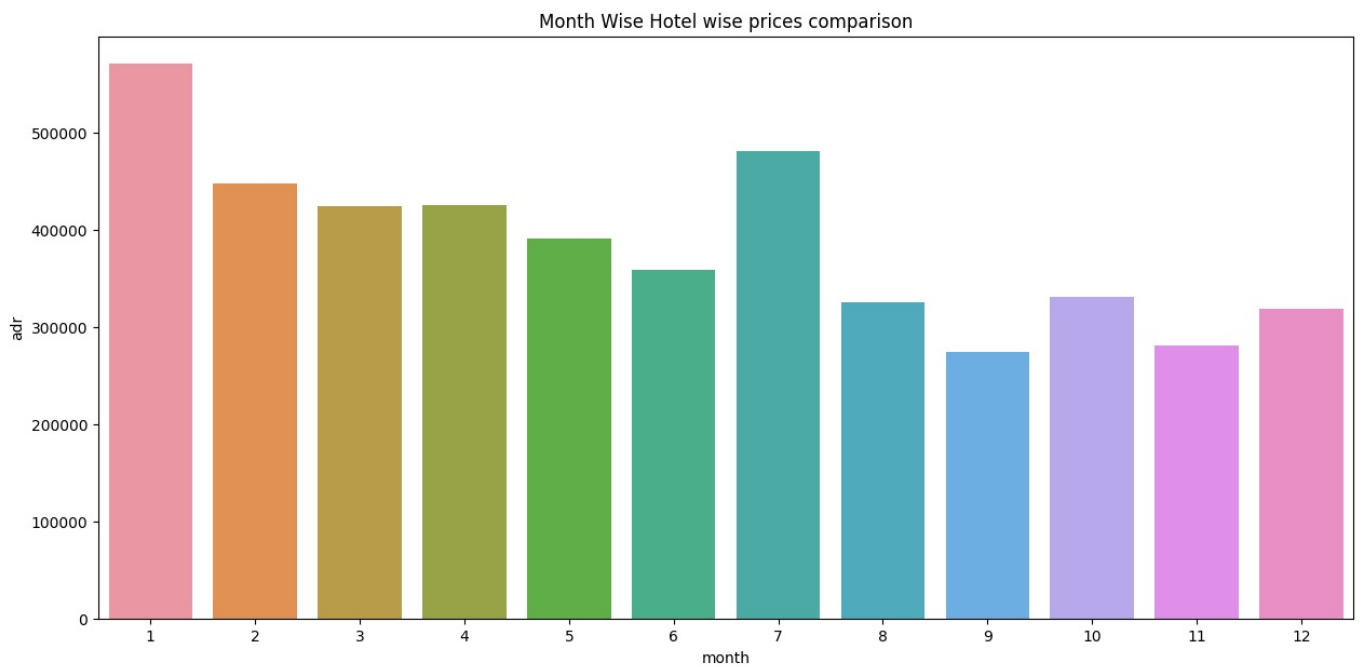
Resort Hotel Price Fluctuation is too much.

```
In [46]: df['month'] = df['reservation_status_date'].dt.month
plt.figure(figsize = (15,7))
sns.countplot(x = 'month',hue='is_canceled',data = df)
plt.legend(['not cancelled','cancelled'])
plt.title('Month Wise Cancellation Count')
plt.show()
```



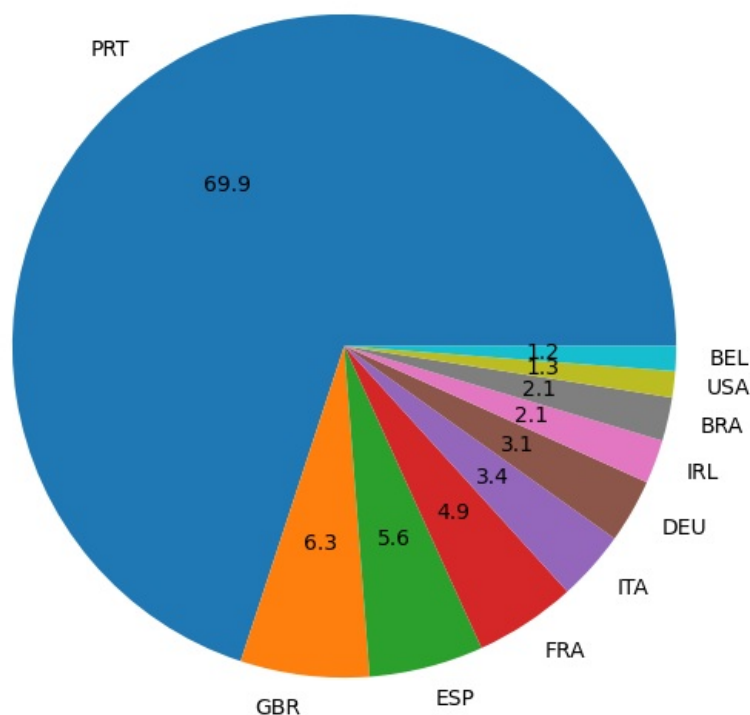
As per above chart it is clear that the cancellation in January month was highest and in August month was lowest.

```
In [47]: plt.figure(figsize = (15,7))
sns.barplot(x = 'month',y='adr',data = df[df['is_canceled']==1].groupby('month')[['adr']].sum().reset_index())
plt.title('Month Wise Hotel wise prices comparison')
plt.show()
```



As per above graph it is clear that the price of hotels in January was highest as compare to other months.

```
In [61]: df1 = df[df['is_canceled']==1]
df2 = df1['country'].value_counts()[:10]
plt.figure(figsize = (7,7))
plt.pie(df2,autopct='%.1f',labels=df2.index)
plt.show()
```

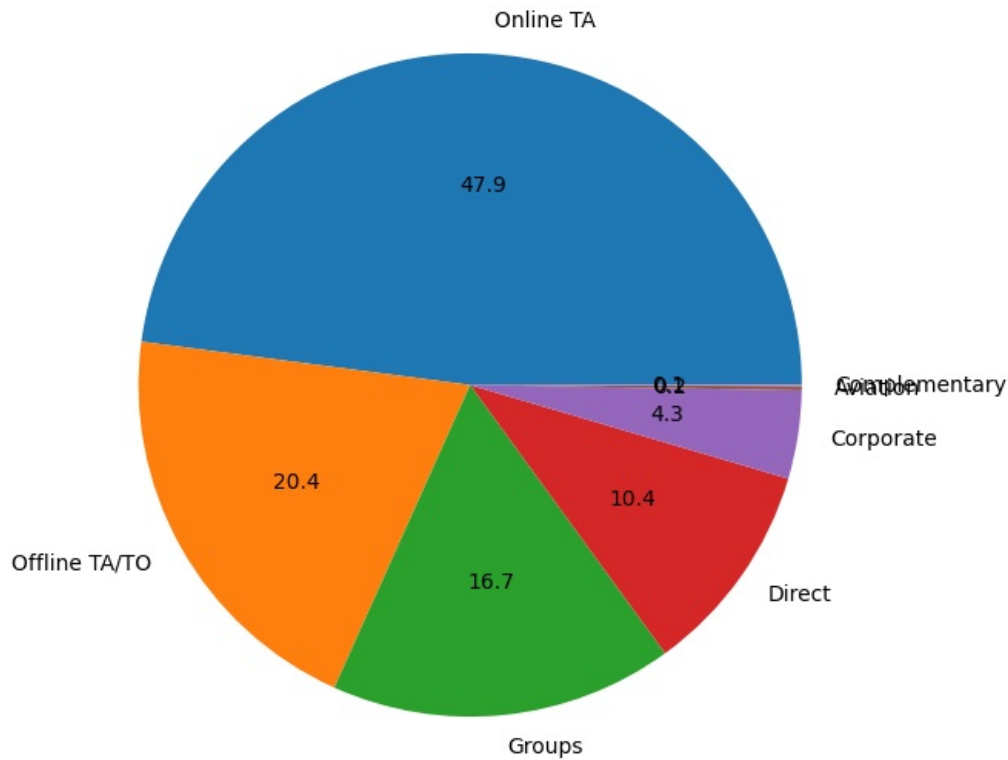


As per above pie chart it is clear that the cancellation rate in Portugal is highest.

```
In [64]: df3=df['market_segment'].value_counts()
df3
```

```
Out[64]: market_segment
Online TA      56036
Offline TA/TO  23829
Groups         19554
Direct         12210
Corporate       5035
Aviation        231
Complementary    63
Name: count, dtype: int64
```

```
In [65]: plt.figure(figsize = (7,7))
plt.pie(df3,autopct='%.1f',labels=df3.index)
plt.show()
```

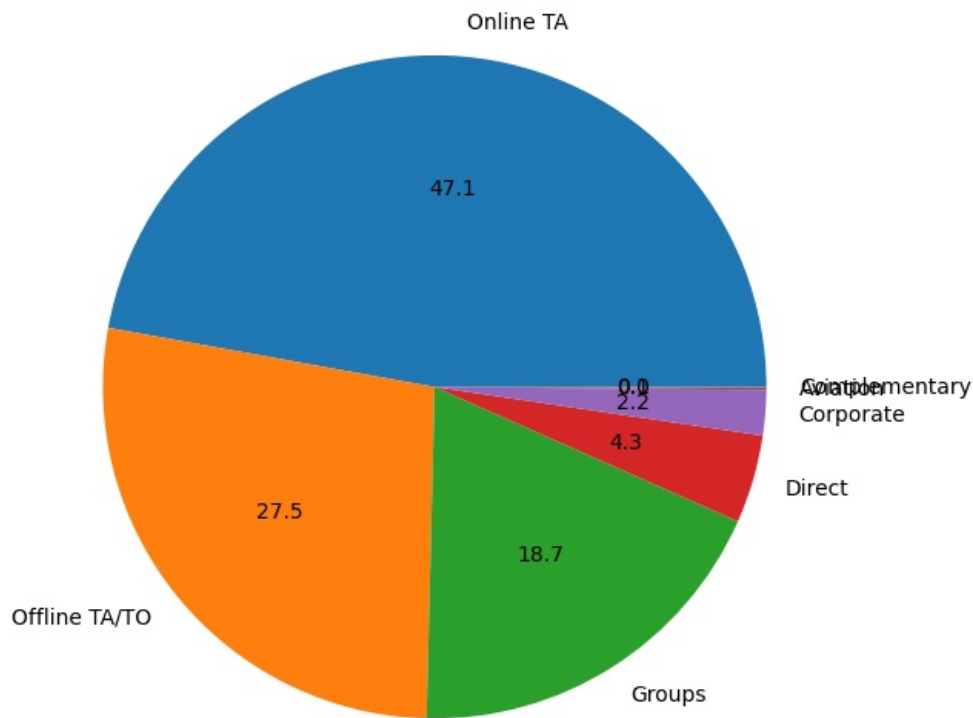


Mostly people are booking their hotels through online travel agency.

```
In [68]: df4= df1['market_segment'].value_counts()
df4
```

```
Out[68]: market_segment
Online TA      20718
Groups         12073
Offline TA/TO   8239
Direct          1889
Corporate        974
Aviation         51
Complementary     5
Name: count, dtype: int64
```

```
In [69]: plt.figure(figsize = (7,7))
plt.pie(df4,autopct='%.1f',labels=df3.index)
plt.show()
```



47% people doing cancellation are from Online Travel Agency segment.

## Findings:

As per my analysis:

1. Total cancelled Bookings: 43949 and Total Non cancelled Bookings: 73009.
2. City Hotels cancellation % was higher (42% approx) as compare to Resort Hotels cancellation % (28% approx).
3. Overall mean price of hotels was high when cancellation done.
4. In January month, cancellation rate was highest and Hotels prices were also at their peak in January month.
5. In Portugal country, cancellation rate was highest (70%) as compare to other countries.
6. Online travel agency booking and cancellation rate both were highest.

## Suggestions:

1. Need more focus on City Hotels. Provide better services in City Hotels.
2. Try to provide best in lowest prices.
3. Main focus required on Portugal country's hotels.
4. Give the customer what you are showing him online.