hotel-booking-analysis

June 5, 2024

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```
ds=pd.read_csv(r'C:\Users\ART\Downloads\hotel_booking.csv')
```

[3]:
```
ds.shape
```

[3]:
```
(119390, 36)
```

[4]:
```
ds.head()
```

[4]:

	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	adr	required_car_parking_spaces	total_of_special_requests
0	Resort Hotel	0	342	2015	July	27	1	0	0	2	…	Transient	0.0	0	0
1	Resort Hotel	0	737	2015	July	27	1	0	0	2	…	Transient	0.0	0	0
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			

```
2  75.0                            0                          0
3  75.0                            0                          0
4  98.0                            0                          1

  reservation_status reservation_status_date            name  \
0          Check-Out              2015-07-01    Ernest Barnes
1          Check-Out              2015-07-01     Andrea Baker
2          Check-Out              2015-07-02   Rebecca Parker
3          Check-Out              2015-07-02    Laura Murray
4          Check-Out              2015-07-03     Linda Hines

                          email  phone-number       credit_card
0  Ernest.Barnes31@outlook.com  669-792-1661  ************4322
1       Andrea_Baker94@aol.com  858-637-6955  ************9157
2   Rebecca_Parker@comcast.net  652-885-2745  ************3734
3            Laura_M@gmail.com  364-656-8427  ************5677
4           LHines@verizon.com  713-226-5883  ************5498

[5 rows x 36 columns]
```

[5]:
```
ds.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
```

[6]:
```
pd.isnull(ds).sum()
```

[6]:
```
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
```

[7]:
```
ds.drop(['company'], axis=1, inplace=True)
```

[8]:
```
ds.describe()
```

[8]:
```
         is_canceled      lead_time  arrival_date_year  \
count  119390.000000  119390.000000      119390.000000
mean        0.370416     104.011416        2016.156554
std         0.482918     106.863097           0.707476
min         0.000000       0.000000        2015.000000
25%         0.000000      18.000000        2016.000000
50%         0.000000      69.000000        2016.000000
75%         1.000000     160.000000        2017.000000
max         1.000000     737.000000        2017.000000

       arrival_date_week_number  arrival_date_day_of_month  \
count             119390.000000              119390.000000
mean                  27.165173                  15.798241
std                   13.605138                   8.780829
min                    1.000000                   1.000000
25%                   16.000000                   8.000000
50%                   28.000000                  16.000000
75%                   38.000000                  23.000000
max                   53.000000                  31.000000

       stays_in_weekend_nights  stays_in_week_nights         adults  \
count            119390.000000         119390.000000  119390.000000
mean                  0.927599              2.500302       1.856403
std                   0.998613              1.908286       0.579261
min                   0.000000              0.000000       0.000000
25%                   0.000000              1.000000       2.000000
50%                   1.000000              2.000000       2.000000
75%                   2.000000              3.000000       2.000000
max                  19.000000             50.000000      55.000000

            children         babies  is_repeated_guest  \
```

```
count  119386.000000  119390.000000      119390.000000
mean        0.103890       0.007949           0.031912
std         0.398561       0.097436           0.175767
min         0.000000       0.000000           0.000000
25%         0.000000       0.000000           0.000000
50%         0.000000       0.000000           0.000000
75%         0.000000       0.000000           0.000000
max        10.000000      10.000000           1.000000

       previous_cancellations  previous_bookings_not_canceled  \
count           119390.000000                   119390.000000
mean                 0.087118                        0.137097
std                  0.844336                        1.497437
min                  0.000000                        0.000000
25%                  0.000000                        0.000000
50%                  0.000000                        0.000000
75%                  0.000000                        0.000000
max                 26.000000                       72.000000

       booking_changes          agent  days_in_waiting_list            adr  \
count    119390.000000  103050.000000         119390.000000  119390.000000
mean          0.221124      86.693382              2.321149     101.831122
std           0.652306     110.774548             17.594721      50.535790
min           0.000000       1.000000              0.000000      -6.380000
25%           0.000000       9.000000              0.000000      69.290000
50%           0.000000      14.000000              0.000000      94.575000
75%           0.000000     229.000000              0.000000     126.000000
max          21.000000     535.000000            391.000000    5400.000000

       required_car_parking_spaces  total_of_special_requests
count                119390.000000              119390.000000
mean                      0.062518                   0.571363
std                       0.245291                   0.792798
min                       0.000000                   0.000000
25%                       0.000000                   0.000000
50%                       0.000000                   0.000000
75%                       0.000000                   1.000000
max                       8.000000                   5.000000
```

[9]:
```
Con=ds['country'].mode()
```

[10]:
```
ds.fillna({'country':'Con'},inplace=True)
```

[11]:
```
pd.isnull(ds).sum()
```

[11]:
```
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                             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
agent                           16340
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
```

[12]:
```
ds.fillna({'children':0.103890},inplace=True)
```

[13]:
```
pd.isnull(ds).sum()
```

[13]:
```
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
agent                           16340
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
```

[14]:
```
ds['agent'].fillna(-1, inplace=True)
```

```
C:\Users\ART\AppData\Local\Temp\ipykernel_7160\570396003.py:1: FutureWarning: A
value is trying to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work
because the intermediate object on which we are setting values always behaves as
a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using
'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value)
instead, to perform the operation inplace on the original object.


  ds['agent'].fillna(-1, inplace=True)
```

[15]:
```
pd.isnull(ds).sum()
```

[15]:
```
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
agent                           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
```

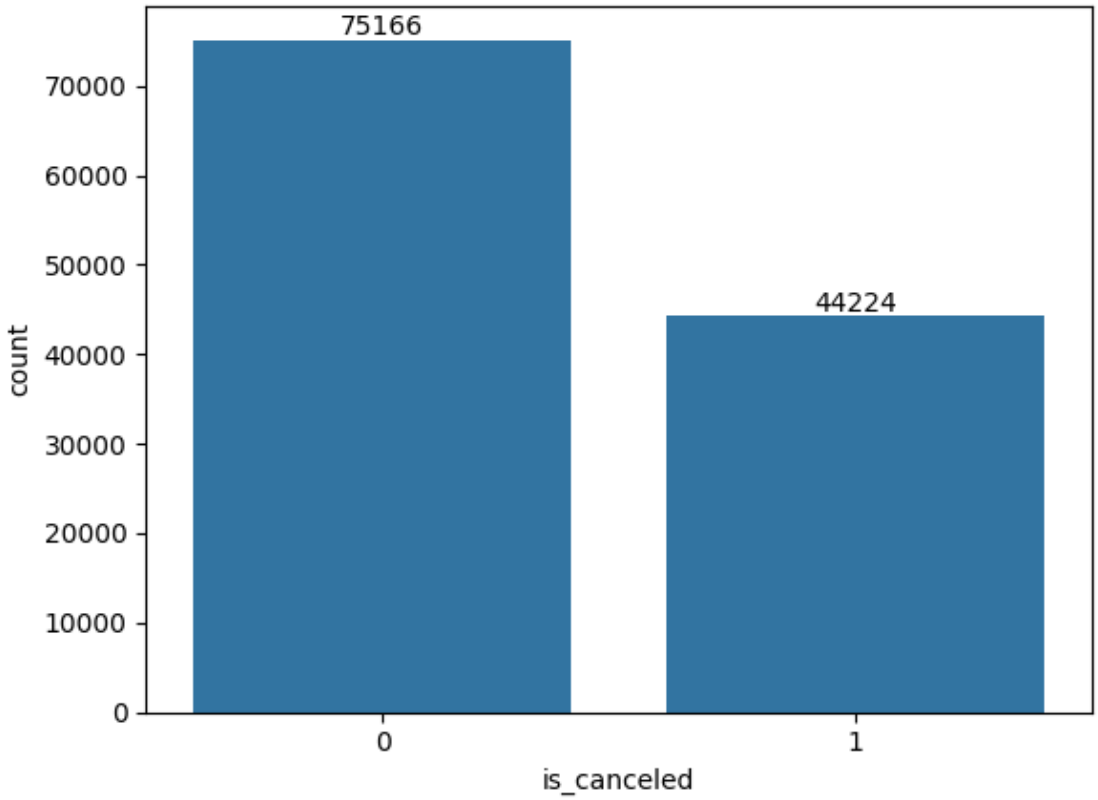
[16]:
```
ds.describe(include='object')
```

[16]:

	hotel	arrival_date_month	meal	country	market_segment	distribution_channel	reserved_room_type	assigned_room_type
count	119390	119390	119390	119390	119390	119390	119390	119390
unique	2	12	5	178	8			
top	City Hotel	August	BB	PRT	Online TA			
freq	79330	13877	92310	48590	56477			

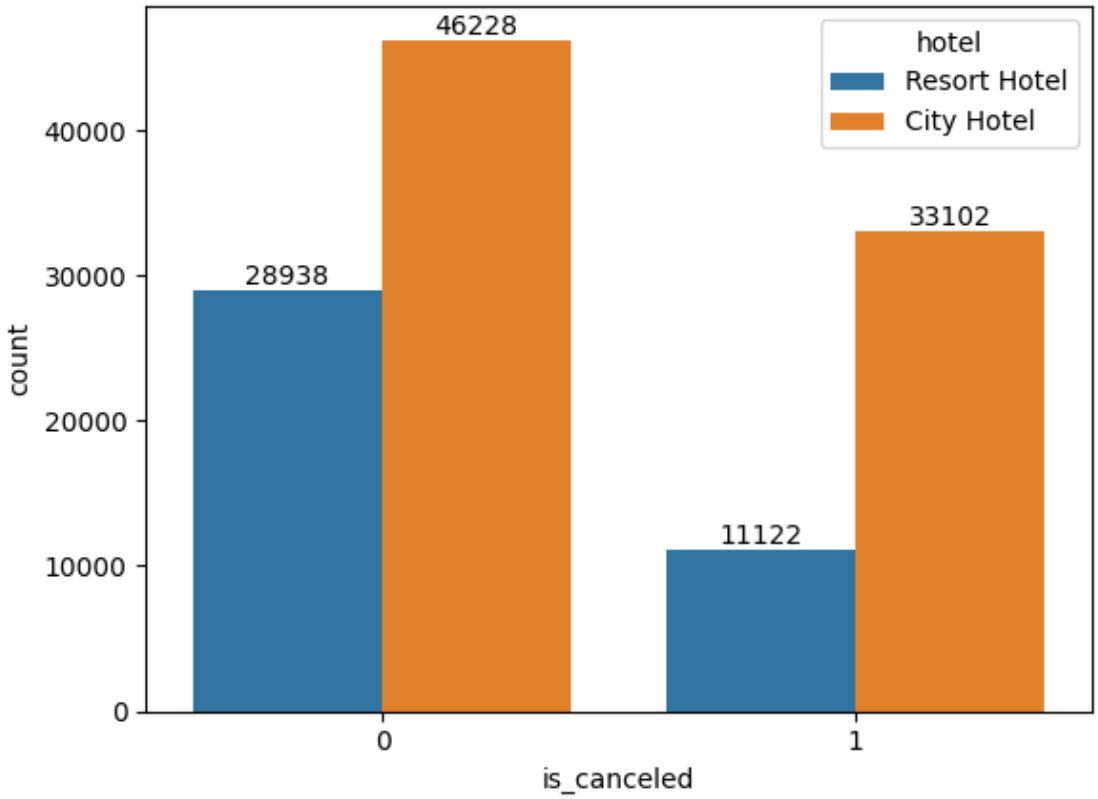
```
unique                    5                 10                 12
top                   TA/TO                  A                  A
freq                  97870              85994              74053

       deposit_type customer_type reservation_status reservation_status_date  \
count        119390        119390             119390                   119390
unique            3             4                  3                      926
top      No Deposit     Transient          Check-Out               2015-10-21
freq         104641         89613              75166                     1461

                    name                email  phone-number       credit_card
count             119390               119390        119390            119390
unique             81503               115889        119390              9000
top     Michael Johnson  Michael.C@gmail.com  669-792-1661  ************4923
freq                  48                    6             1                28
```

[17]:
```
ax=sns.countplot(x='is_canceled',data=ds)
for bars in ax.containers: ax.bar_label(bars)
```



[18]:
```
count_plot=sns.countplot(x='is_canceled', hue='hotel', data=ds)
for container in count_plot.containers:
```

```
    count_plot.bar_label(container)
plt.show()
```



[19]:
```
hotel_name=ds['hotel'].unique()
```

[20]:
```
unique=ds.hotel.value_counts()
```

[21]:
```
meal_count=ds.meal.value_counts()
```

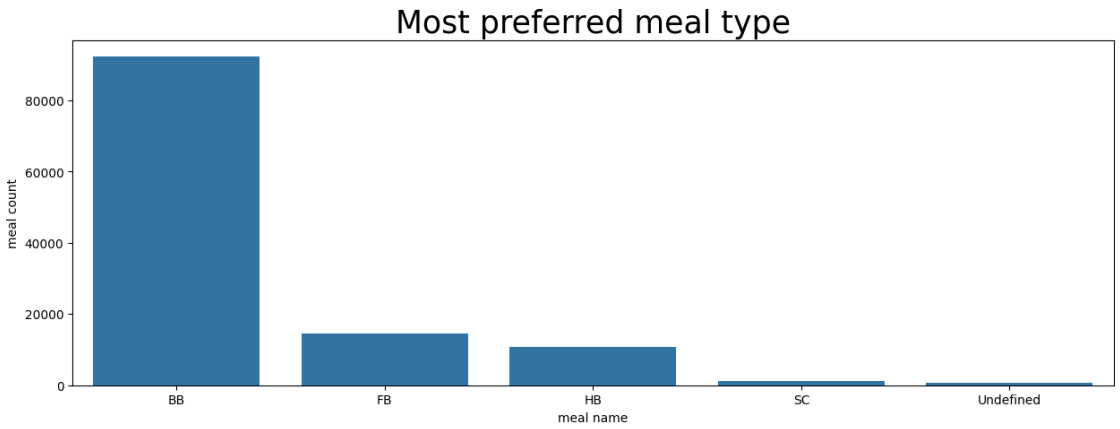
[22]:
```
meal_name=ds['meal'].unique()
```

[23]:
```
meal_ds=pd.DataFrame(zip(meal_name,meal_count),columns=['meal name', 'meal␣
↪count'])
```

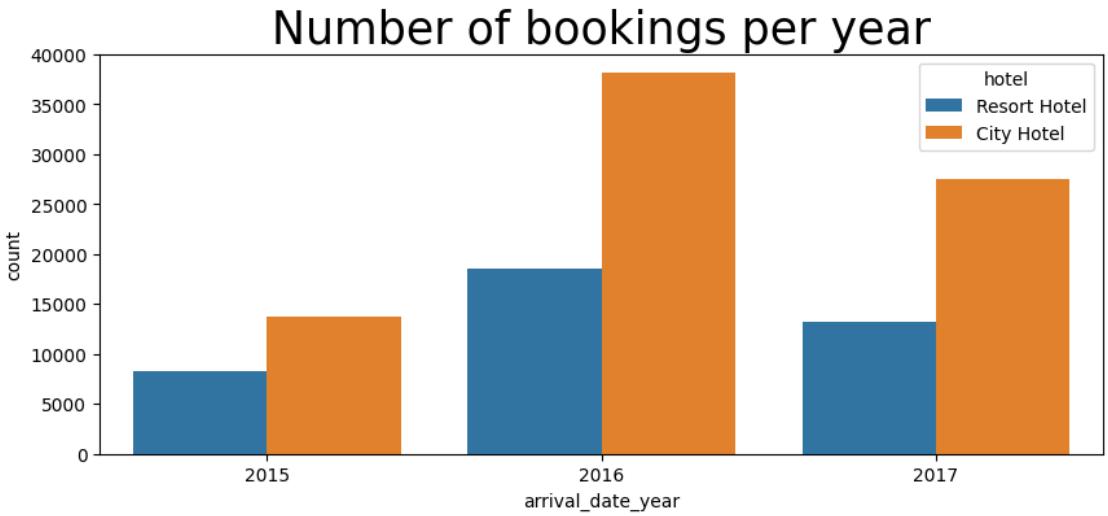
[24]:
```
plt.figure(figsize=(15,5))
g=sns.barplot(data=meal_ds,x='meal name',y='meal count')
g.set_xticklabels(meal_ds['meal name'])
plt.title('Most preferred meal type', fontsize=25)
plt.show()
```

```
C:\Users\ART\AppData\Local\Temp\ipykernel_7160\2935601102.py:3: UserWarning:
```

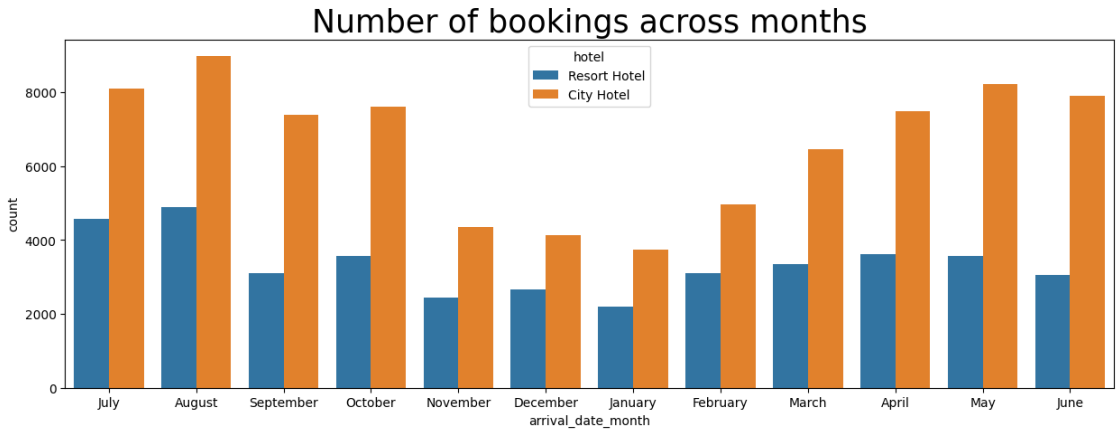
```
set_ticklabels() should only be used with a fixed number of ticks, i.e. after
set_ticks() or using a FixedLocator.
  g.set_xticklabels(meal_ds['meal name'])
```



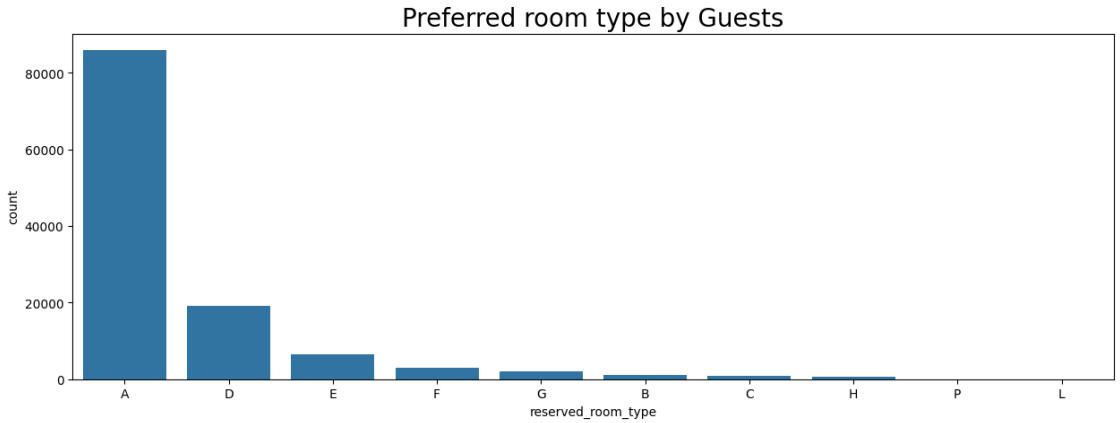
[25]:
```
plt.figure(figsize=(10,4))
sns.countplot(x=ds['arrival_date_year'],hue=ds['hotel'])
plt.title("Number of bookings per year", fontsize=25)
plt.show()
```



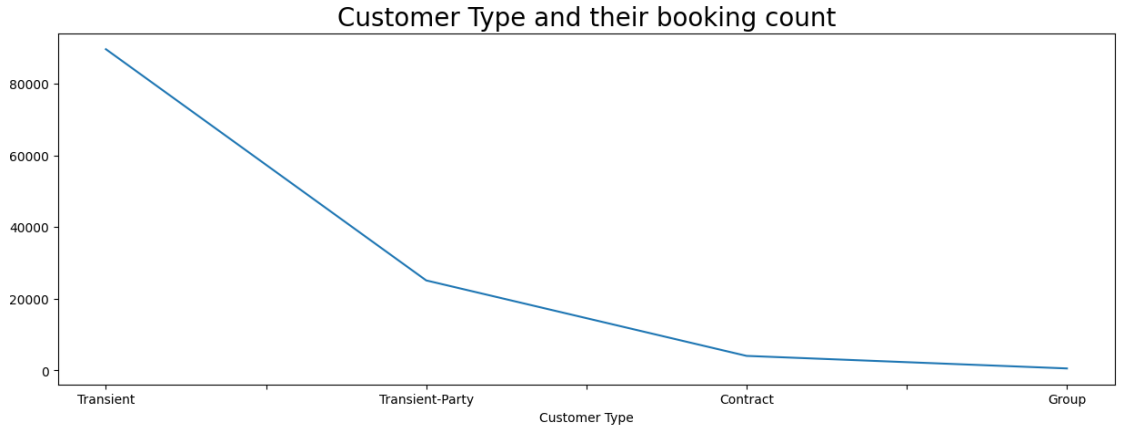
[26]:
```
plt.figure(figsize=(15,5))
sns.countplot(x=ds['arrival_date_month'],hue=ds['hotel'])
plt.title("Number of bookings across months", fontsize=25)
plt.show()
```



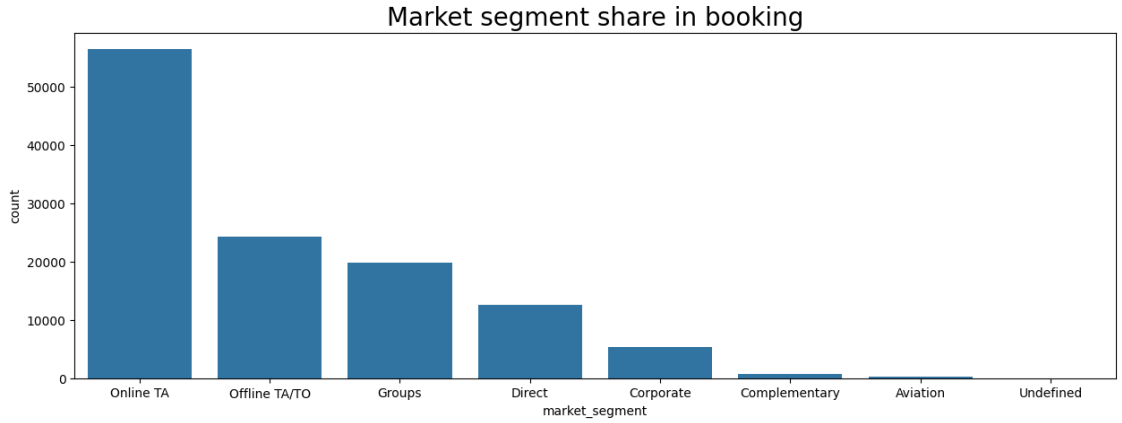
[27]:
```
plt.figure(figsize=(15,5))
sns.countplot(x=ds['reserved_room_type'],order=ds['reserved_room_type'].
 ↪value_counts().index)
plt.title('Preferred room type by Guests',fontsize=20)
plt.show()
```



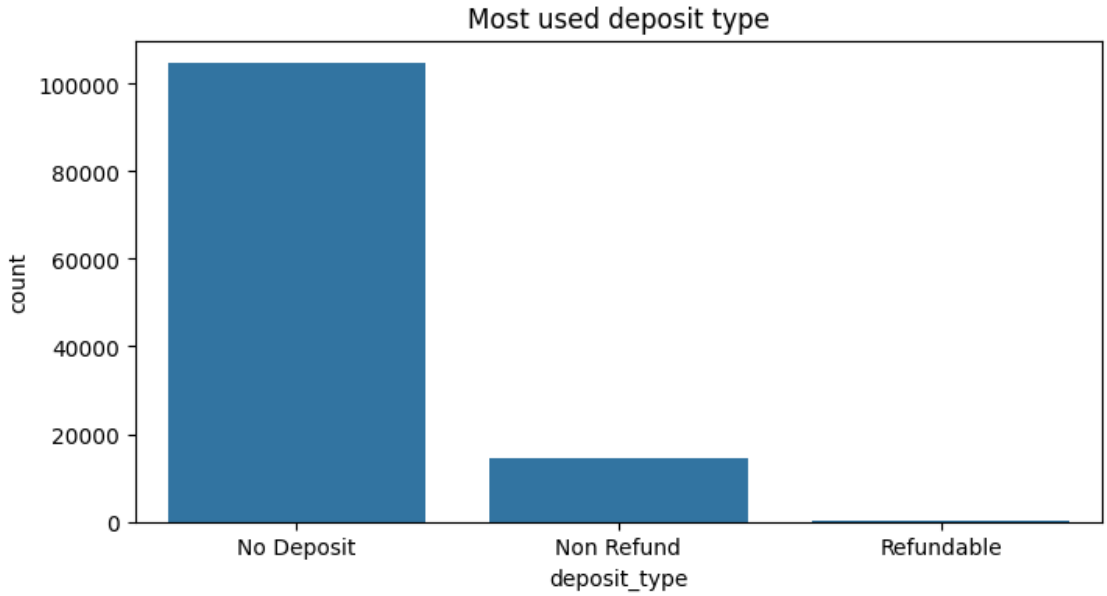
[28]:
```
cust_type = ds['customer_type'].value_counts()
cust_type.plot(figsize=(15,5))
plt.xlabel('Count',fontsize=8)
plt.xlabel('Customer Type',fontsize=10)
plt.title('Customer Type and their booking count',fontsize=20)
plt.show()
```



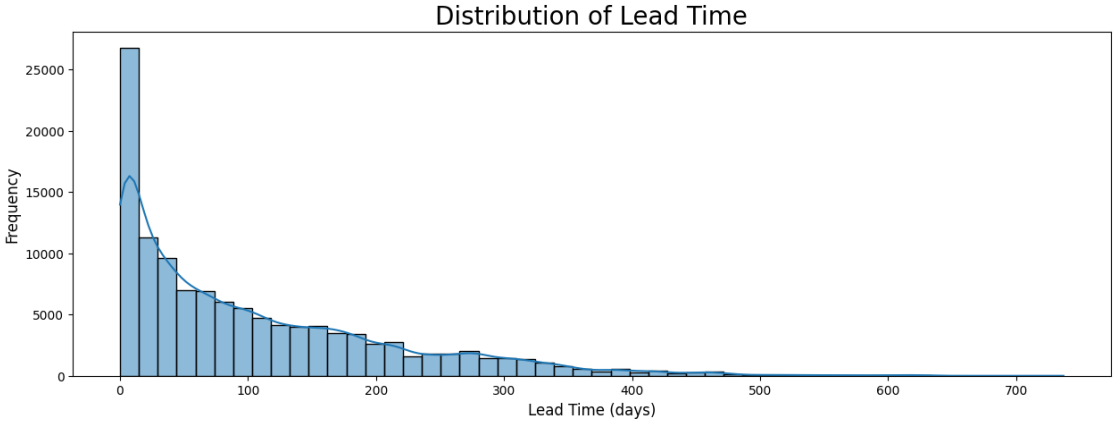
[29]:
```
plt.figure(figsize=(15,5))
sns.countplot(x=ds['market_segment'],order=ds['market_segment'].value_counts().
↪index)
plt.title('Market segment share in booking', fontsize=20)
plt.show()
```



[30]:
```
deposit=ds['deposit_type'].value_counts().index
plt.figure(figsize=(8,4))
sns.countplot(x=ds['deposit_type'],order=deposit)
plt.title('Most used deposit type')
plt.show()
```

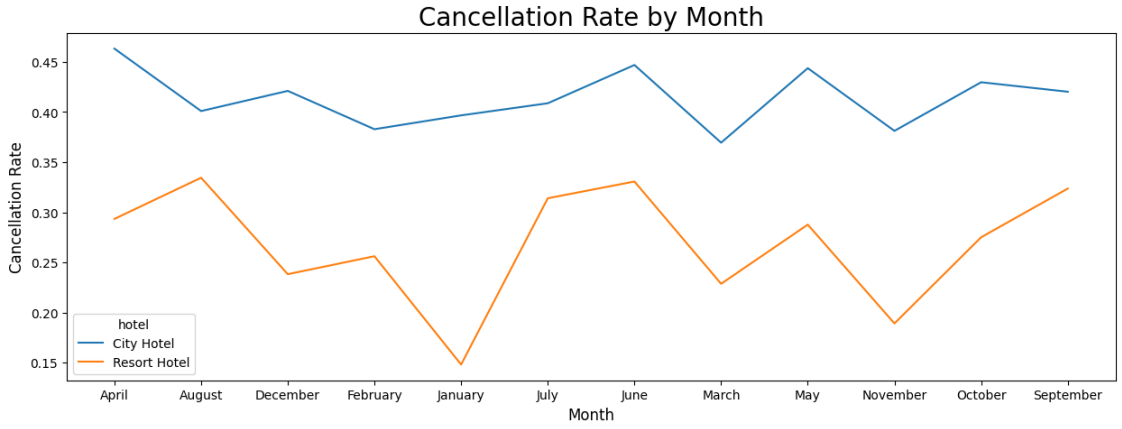


[31]:
```
plt.figure(figsize=(15, 5))
sns.histplot(ds['lead_time'], bins=50, kde=True)
plt.title('Distribution of Lead Time', fontsize=20)
plt.xlabel('Lead Time (days)', fontsize=12)
plt.ylabel('Frequency', fontsize=12)
plt.show()
```

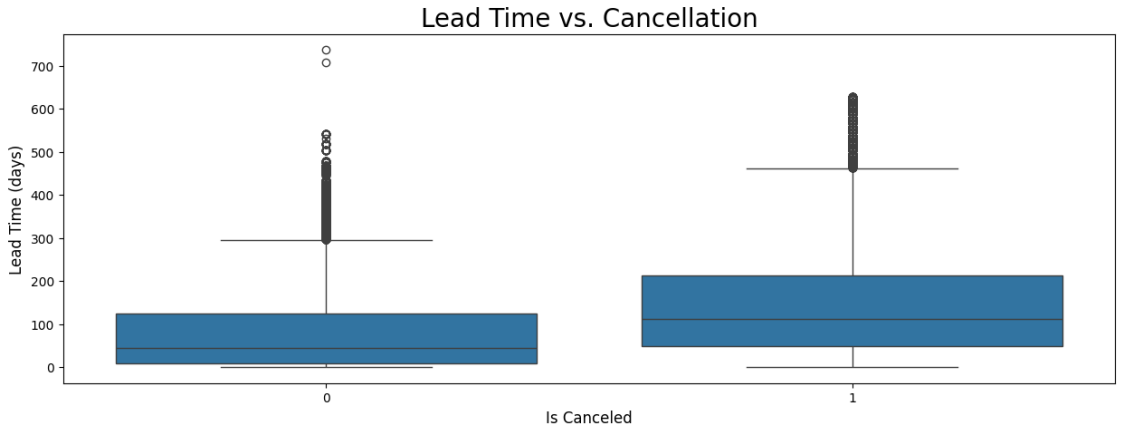


[32]:
```
plt.figure(figsize=(15, 5))
sns.lineplot(x='arrival_date_month', y='is_canceled', hue='hotel', data=ds.
  ↪groupby(['arrival_date_month', 'hotel'])['is_canceled'].mean().reset_index())
plt.title('Cancellation Rate by Month', fontsize=20)
plt.xlabel('Month', fontsize=12)
```

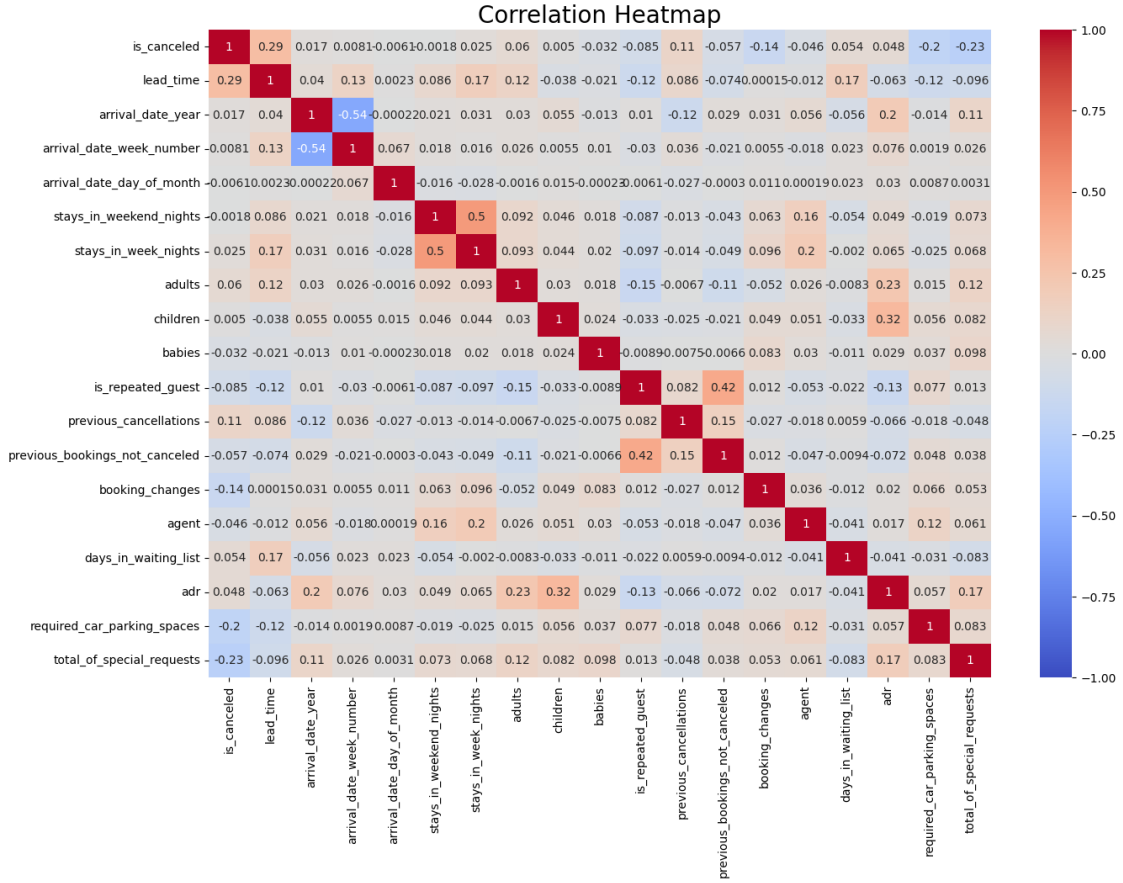
```
plt.ylabel('Cancellation Rate', fontsize=12)
plt.show()
```



[33]:
```
plt.figure(figsize=(15, 5))
sns.boxplot(x='is_canceled', y='lead_time', data=ds)
plt.title('Lead Time vs. Cancellation', fontsize=20)
plt.xlabel('Is Canceled', fontsize=12)
plt.ylabel('Lead Time (days)', fontsize=12)
plt.show()
```



[36]:
```
numerical_ds = ds.select_dtypes(include=[np.number])
corr = numerical_ds.corr()
plt.figure(figsize=(15, 10))
sns.heatmap(corr, annot=True, cmap='coolwarm', vmin=-1, vmax=1)
plt.title('Correlation Heatmap', fontsize=20)
plt.show()
```


Correlation Heatmap
is_canceled 1 0.29 0.017 0.0081 -0.0061 -0.0018 0.025 0.06 0.005 -0.032 -0.085 0.11 -0.057 -0.14 -0.046 0.054 0.048 -0.2 -0.23
lead_time 0.29 1 0.04 0.13 0.0023 0.086 0.17 0.12 -0.038 -0.021 -0.12 0.086 -0.074 0.00015 -0.012 0.17 -0.063 -0.12 -0.096
arrival_date_year 0.017 0.04 1 -0.54 0.00022 0.021 0.031 0.03 0.055 -0.013 0.01 -0.12 0.029 0.031 0.056 -0.056 0.2 -0.014 0.11
arrival_date_week_number 0.0081 0.13 -0.54 1 0.067 0.018 0.016 0.026 0.0055 0.01 -0.03 0.036 -0.021 0.0055 -0.018 0.023 0.076 0.0019 0.026
arrival_date_day_of_month -0.0061 0.0023 0.00022 0.067 1 -0.016 -0.028 -0.0016 0.015 -0.00023 0.0061 -0.027 -0.0003 0.011 0.00019 0.023 0.03 0.0087 0.0031
stays_in_weekend_nights -0.0018 0.086 0.021 0.018 -0.016 1 0.5 0.092 0.046 0.018 -0.087 -0.013 -0.043 0.063 0.16 -0.054 0.049 -0.019 0.073
stays_in_week_nights 0.025 0.17 0.031 0.016 -0.028 0.5 1 0.093 0.044 0.02 -0.097 -0.014 -0.049 0.096 0.2 -0.002 0.065 -0.025 0.068
adults 0.06 0.12 0.03 0.026 -0.0016 0.092 0.093 1 0.03 0.018 -0.15 -0.0067 -0.11 -0.052 0.026 -0.0083 0.23 0.015 0.12
children 0.005 -0.038 0.055 0.0055 0.015 0.046 0.044 0.03 1 0.024 -0.033 -0.025 -0.021 0.049 0.051 -0.033 0.32 0.056 0.082
babies -0.032 -0.021 -0.013 0.01 -0.00023 0.018 0.02 0.018 0.024 1 -0.0089 -0.0075 -0.0066 0.083 0.03 -0.011 0.029 0.037 0.098
is_repeated_guest -0.085 -0.12 0.01 -0.03 -0.0061 -0.087 -0.097 -0.15 -0.033 -0.0089 1 0.082 0.42 0.012 -0.053 -0.022 -0.13 0.077 0.013
previous_cancellations 0.11 0.086 -0.12 0.036 -0.027 -0.013 -0.014 -0.0067 -0.025 -0.0075 0.082 1 0.15 -0.027 -0.018 0.0059 -0.066 -0.018 -0.048
previous_bookings_not_canceled -0.057 -0.074 0.029 -0.021 -0.0003 -0.043 -0.049 -0.11 -0.021 -0.0066 0.42 0.15 1 0.012 -0.047 -0.0094 -0.072 0.048 0.038
booking_changes -0.14 0.00015 0.031 0.0055 0.011 0.063 0.096 -0.052 0.049 0.083 0.012 -0.027 0.012 1 0.036 -0.012 0.02 0.066 0.053
agent -0.046 -0.012 0.056 -0.018 0.00019 0.16 0.2 0.026 0.051 0.03 -0.053 -0.018 -0.047 0.036 1 -0.041 0.017 0.12 0.061
days_in_waiting_list 0.054 0.17 -0.056 0.023 0.023 -0.054 -0.002 -0.0083 -0.033 -0.011 -0.022 0.0059 -0.0094 -0.012 -0.041 1 -0.041 -0.031 -0.083
adr 0.048 -0.063 0.2 0.076 0.03 0.049 0.065 0.23 0.32 0.029 -0.13 -0.066 -0.072 0.02 0.017 -0.041 1 0.057 0.17
required_car_parking_spaces -0.2 -0.12 -0.014 0.0019 0.0087 -0.019 -0.025 0.015 0.056 0.037 0.077 -0.018 0.048 0.066 0.12 -0.031 0.057 1 0.083
total_of_special_requests -0.23 -0.096 0.11 0.026 0.0031 0.073 0.068 0.12 0.082 0.098 0.013 -0.048 0.038 0.053 0.061 -0.083 0.17 0.083 1
is_canceled
lead_time
arrival_date_year
arrival_date_week_number
arrival_date_day_of_month
stays_in_weekend_nights
stays_in_week_nights
adults
children
babies
is_repeated_guest
previous_cancellations
previous_bookings_not_canceled
booking_changes
agent
days_in_waiting_list
adr
required_car_parking_spaces
total_of_special_requests
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

[ ]: