```
import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

hdf = pd.read_csv("C:\my files\hotel_bookings.csv")

hdf.head()
```

```
          hotel  is_canceled  lead_time  arrival_date_year
arrival_date_month  \
0  Resort Hotel            0        342               2015
July
1  Resort Hotel            0        737               2015
July
2  Resort Hotel            0          7               2015
July
3  Resort Hotel            0         13               2015
July
4  Resort Hotel            0         14               2015
July

   arrival_date_week_number  arrival_date_day_of_month  \
0                        27                          1
1                        27                          1
2                        27                          1
3                        27                          1
4                        27                          1

   stays_in_weekend_nights  stays_in_week_nights  adults  ...
deposit_type  \
0                        0                     0       2  ...    No
Deposit
1                        0                     0       2  ...    No
Deposit
2                        0                     1       1  ...    No
Deposit
3                        0                     1       1  ...    No
Deposit
4                        0                     2       2  ...    No
Deposit

   agent company days_in_waiting_list customer_type   adr  \
0    NaN     NaN                    0     Transient   0.0
1    NaN     NaN                    0     Transient   0.0
2    NaN     NaN                    0     Transient  75.0
3  304.0     NaN                    0     Transient  75.0
4  240.0     NaN                    0     Transient  98.0

   required_car_parking_spaces  total_of_special_requests
```

```
reservation_status  \
0                               0                               0
Check-Out
1                               0                               0
Check-Out
2                               0                               0
Check-Out
3                               0                               0
Check-Out
4                               0                               1
Check-Out

  reservation_status_date
0              2015-07-01
1              2015-07-01
2              2015-07-02
3              2015-07-02
4              2015-07-03

[5 rows x 32 columns]
```

```
hdf.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
```

```
hdf.isnull().sum()
```

```
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
```

```
hdf.dropna(inplace = True)
hdf.isnull().sum()
```

```
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
company                           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
```

```
hdf.describe(include = "all")
```

```
                hotel  is_canceled   lead_time  arrival_date_year  \
count             217   217.000000  217.000000         217.000000
unique              2          NaN         NaN                NaN
top     Resort Hotel          NaN         NaN                NaN
freq              170          NaN         NaN                NaN
mean              NaN     0.078341   40.520737        2015.465438
std               NaN     0.269329   61.748375           0.720053
min               NaN     0.000000    0.000000        2015.000000
25%               NaN     0.000000   12.000000        2015.000000
50%               NaN     0.000000   27.000000        2015.000000
75%               NaN     0.000000   36.000000        2016.000000
```

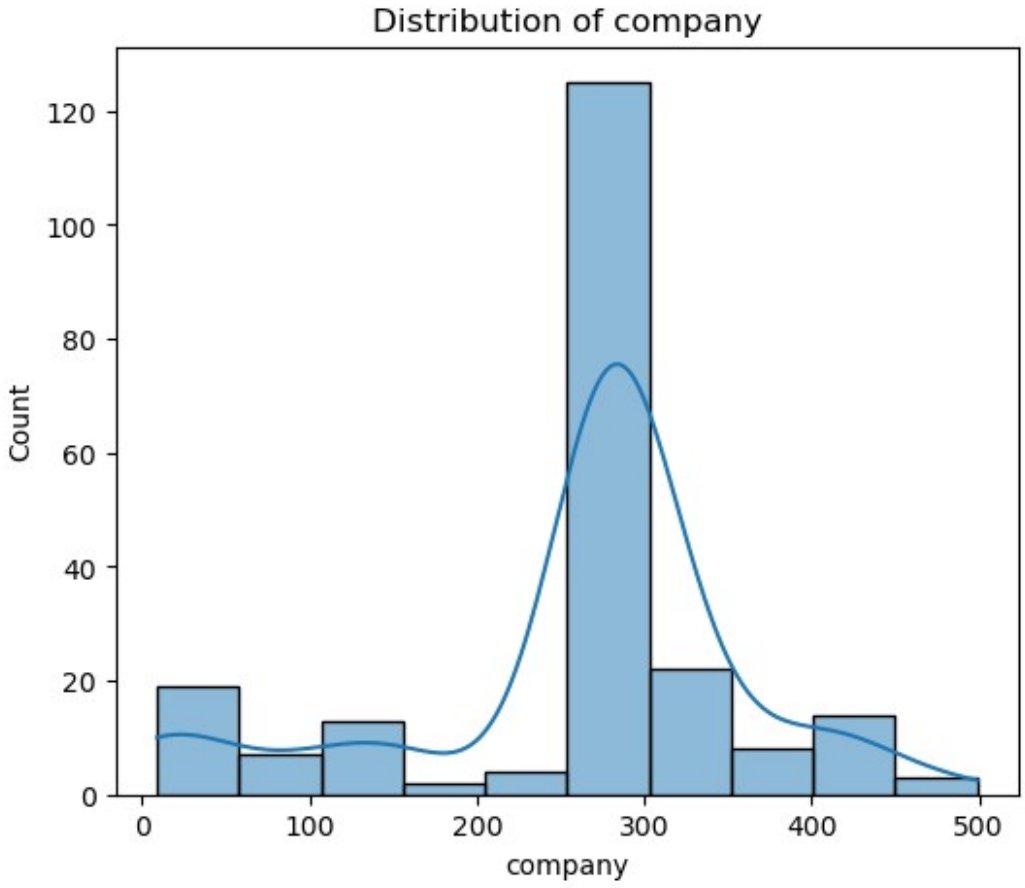
```
max                NaN     1.000000  364.000000        2017.000000

       arrival_date_month  arrival_date_week_number  \
count                 217                217.000000
unique                 12                       NaN
top                November                       NaN
freq                  129                       NaN
mean                  NaN                 38.198157
std                   NaN                 12.890292
min                   NaN                  1.000000
25%                   NaN                 33.000000
50%                   NaN                 45.000000
75%                   NaN                 46.000000
max                   NaN                 53.000000

        arrival_date_day_of_month  stays_in_weekend_nights  \
count                  217.000000                217.00000
unique                        NaN                      NaN
top                           NaN                      NaN
freq                          NaN                      NaN
mean                    10.824885                  1.56682
std                      7.582065                  1.49270
min                      1.000000                  0.00000
25%                      6.000000                  0.00000
50%                      9.000000                  2.00000
75%                     13.000000                  2.00000
max                     31.000000                  9.00000

        stays_in_week_nights      adults  ...  deposit_type
agent  \
count             217.000000  217.000000  ...           217
217.000000
unique                   NaN         NaN  ...             3
NaN
top                      NaN         NaN  ...    No Deposit
NaN
freq                     NaN         NaN  ...           213
NaN
mean                4.631336    1.410138  ...           NaN
202.617512
std                 3.552846    0.520406  ...           NaN
111.487444
min                 0.000000    1.000000  ...           NaN
5.000000
25%                 2.000000    1.000000  ...           NaN
185.000000
50%                 4.000000    1.000000  ...           NaN
185.000000
75%                 6.000000    2.000000  ...           NaN
324.000000
```

```
max                 21.000000    3.000000  ...           NaN
444.000000

           company days_in_waiting_list    customer_type
adr  \
count   217.000000                217.0              217  217.000000

unique         NaN                  NaN                4         NaN

top            NaN                  NaN  Transient-Party         NaN

freq           NaN                  NaN              147         NaN

mean    262.105991                  0.0              NaN   63.793733

std     103.602708                  0.0              NaN   38.823568

min       9.000000                  0.0              NaN    0.000000

25%     281.000000                  0.0              NaN   40.000000

50%     281.000000                  0.0              NaN   48.000000

75%     281.000000                  0.0              NaN   85.000000

max     499.000000                  0.0              NaN  246.000000


        required_car_parking_spaces  total_of_special_requests  \
count                    217.000000                 217.000000
unique                          NaN                        NaN
top                             NaN                        NaN
freq                            NaN                        NaN
mean                       0.092166                   0.198157
std                        0.289929                   0.546365
min                        0.000000                   0.000000
25%                        0.000000                   0.000000
50%                        0.000000                   0.000000
75%                        0.000000                   0.000000
max                        1.000000                   3.000000

        reservation_status reservation_status_date
count                  217                     217
unique                   3                      76
top              Check-Out             2015-11-15
freq                   200                      76
mean                   NaN                     NaN
std                    NaN                     NaN
min                    NaN                     NaN
25%                    NaN                     NaN
50%                    NaN                     NaN
```

```
75%                          NaN                          NaN
max                          NaN                          NaN

[11 rows x 32 columns]
```

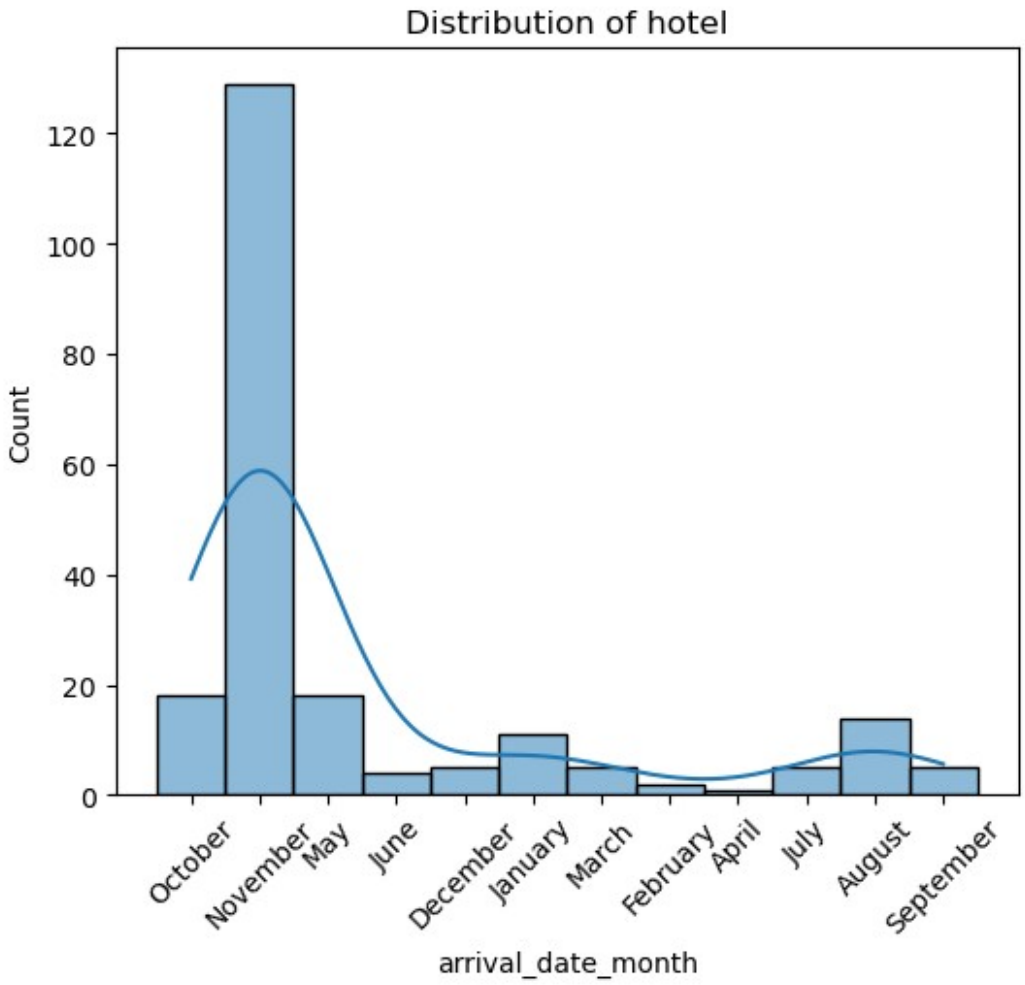
```
plt.figure(figsize = (6,5))
sns.histplot(hdf["company"],kde = True, bins =10)
plt.title("Distribution of company")
plt.show()
```

```
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

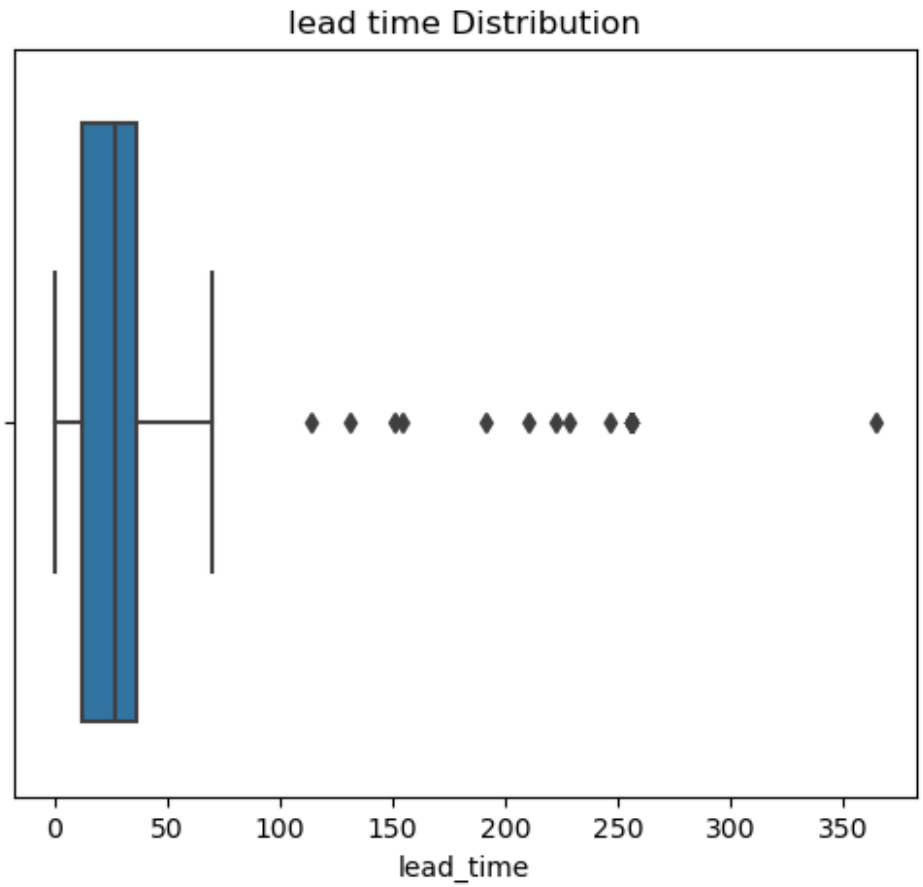



```
plt.figure(figsize = (6,5))
sns.histplot(hdf["arrival_date_month"],kde = True, bins =10)
plt.xticks(rotation=45)
plt.title("Distribution of hotel")
plt.show()
```

```
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
```




```
plt.figure(figsize = (6,5))
sns.boxplot(x=hdf["lead_time"])
plt.title("lead time Distribution")
plt.show()
```



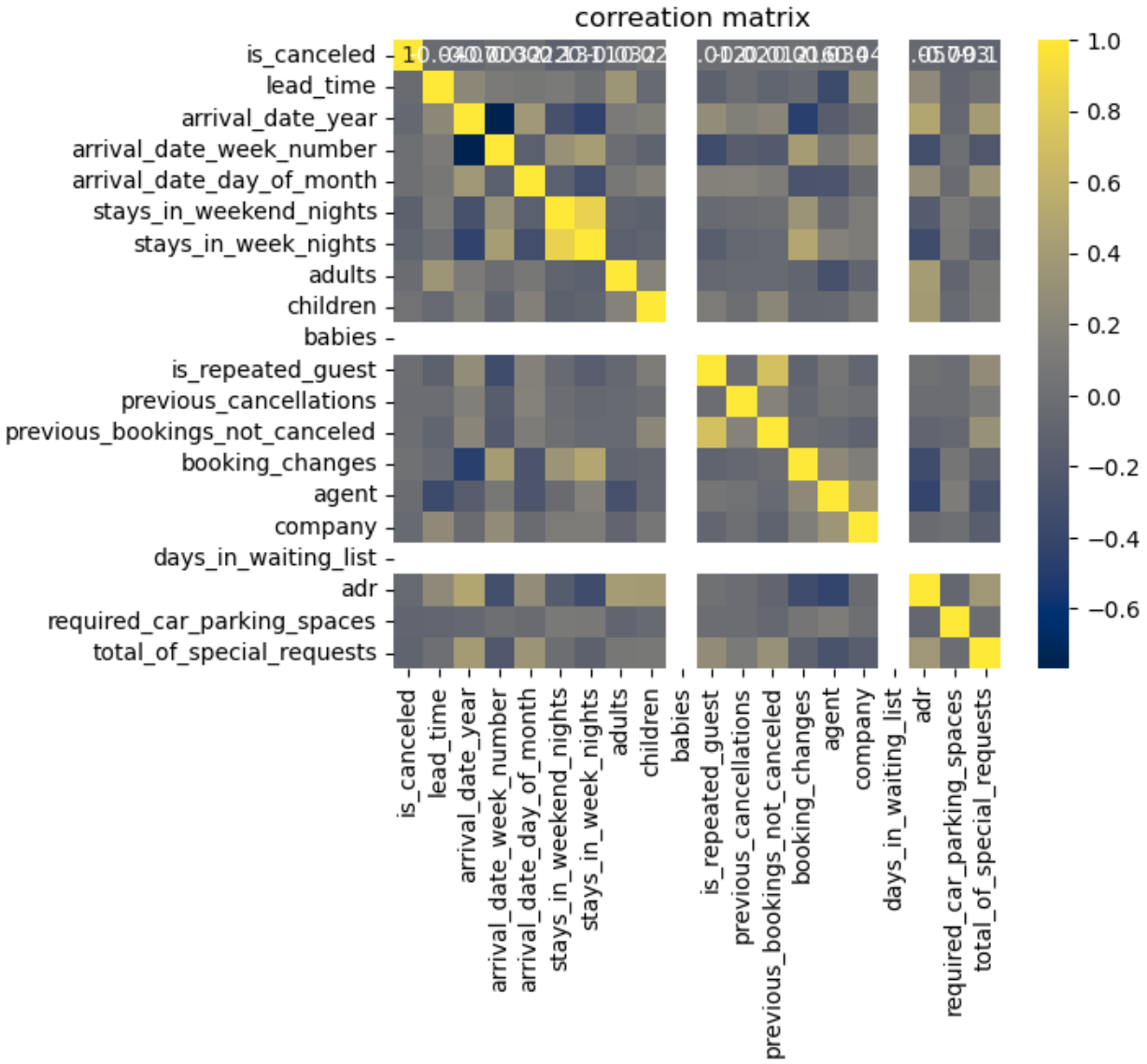

```
hdf1= hdf.select_dtypes(include = ["number"])
hdf1.head()
```

```
      is_canceled  lead_time  arrival_date_year
arrival_date_week_number  \
2392            0          6               2015
42
2697            0         24               2015
44
2867            0         24               2015
45
2877            0         24               2015
45
2878            0         24               2015
45

      arrival_date_day_of_month  stays_in_weekend_nights  \
2392                         11                        2
2697                         26                        7
2867                          3                        0
2877                          3                        2
2878                          3                        3
```

```
      stays_in_week_nights  adults  children  babies
is_repeated_guest  \
2392                     0       2       0.0       0
0
2697                    15       1       0.0       0
0
2867                     3       2       0.0       0
0
2877                    10       1       0.0       0
0
2878                    10       2       0.0       0
0

      previous_cancellations  previous_bookings_not_canceled
booking_changes  \
2392                       0                               0
1
2697                       0                               0
2
2867                       0                               0
1
2877                       0                               0
2
2878                       0                               0
2

      agent  company  days_in_waiting_list   adr
required_car_parking_spaces  \
2392  240.0    113.0                     0  82.0
1
2697  185.0    281.0                     0  52.2
0
2867  334.0    281.0                     0  48.0
0
2877  328.0    281.0                     0  40.0
0
2878  326.0    281.0                     0  48.0
0

      total_of_special_requests
2392                           1
2697                           0
2867                           0
2877                           0
2878                           0

plt.figure(figsize = (6,5))
sns.heatmap(hdf1.corr(),cmap = "cividis",annot = True)
```

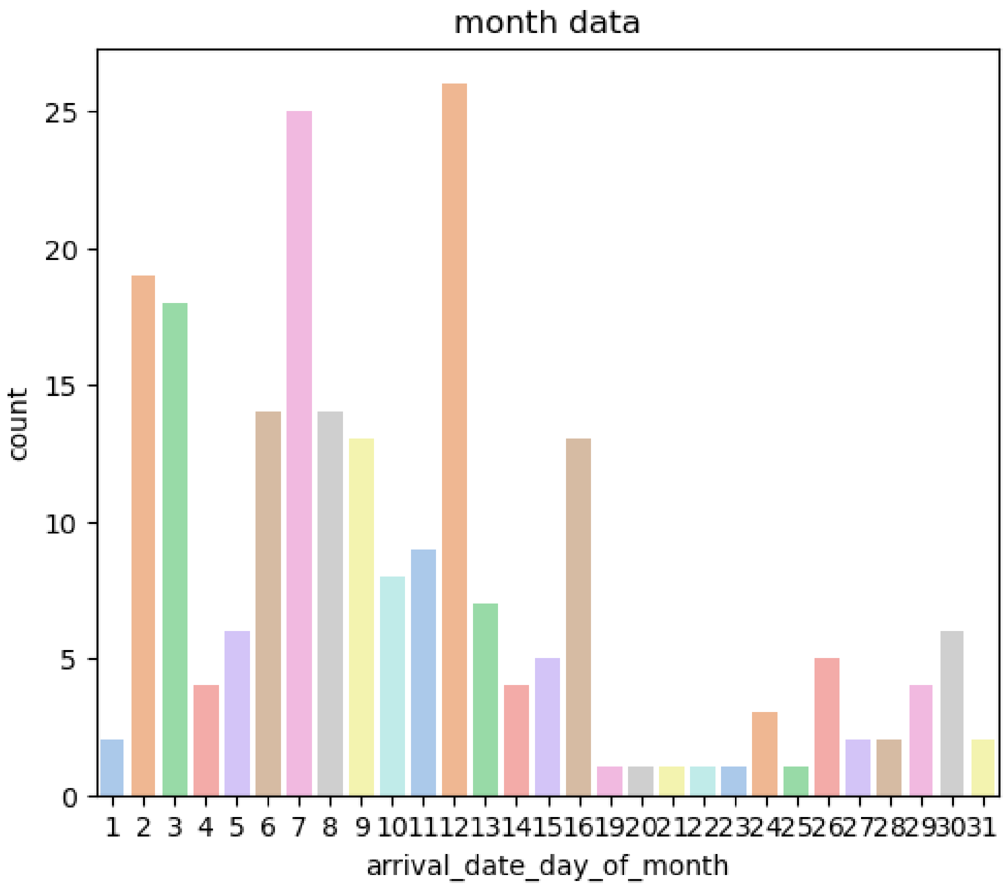
```
plt.title("correation matrix ")
plt.show()
```

```
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\matrix.py:260:
FutureWarning: Format strings passed to MaskedConstant are ignored,
but in future may error or produce different behavior
  annotation = ("{:" + self.fmt + "}").format(val)
```




```
plt.figure(figsize = (6,5))
sns.countplot(x=hdf["arrival_date_day_of_month"],palette = "pastel")
plt.title("month data")
```

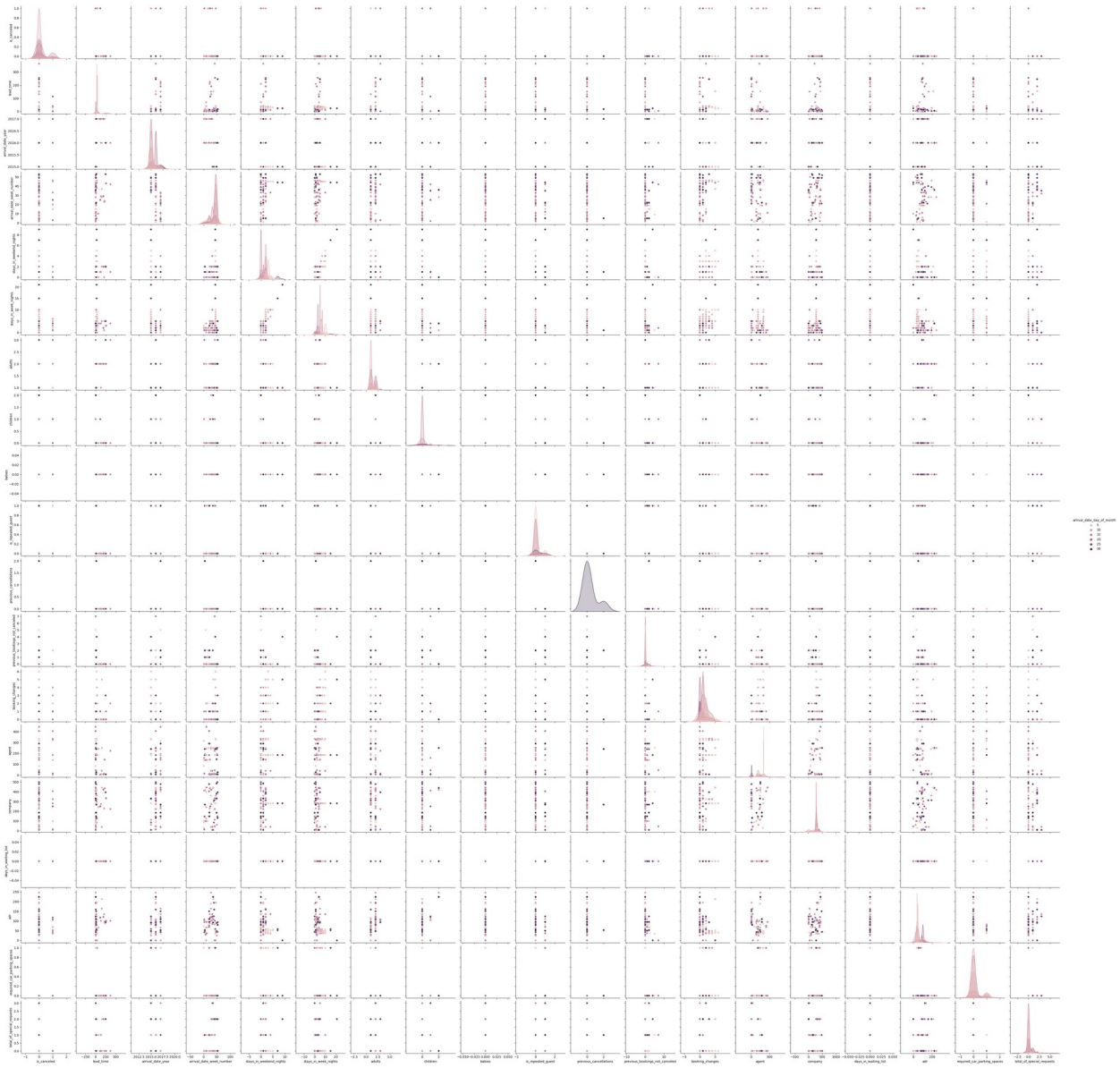
```
Text(0.5, 1.0, 'month data')
```




```
sns.pairplot(hdf,hue="arrival_date_day_of_month")
plt.show()

C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
```

```
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
```

```
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
```



```
hdf.groupby('hotel')['arrival_date_week_number'].mean()
```

```
hotel
City Hotel      30.765957
Resort Hotel    40.252941
Name: arrival_date_week_number, dtype: float64
```