

y73kb7as4

February 9, 2025

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

```
[ ]: df=pd.read_csv("/content/hotel_bookings.csv")
df.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]

```
[ ]: 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

```

```
[ ]: df.describe()
```

```

[ ]:      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	company	days_in_waiting_list	\
count	119390.000000	103050.000000	6797.000000	119390.000000	
mean	0.221124	86.693382	189.266735	2.321149	
std	0.652306	110.774548	131.655015	17.594721	
min	0.000000	1.000000	6.000000	0.000000	
25%	0.000000	9.000000	62.000000	0.000000	
50%	0.000000	14.000000	179.000000	0.000000	
75%	0.000000	229.000000	270.000000	0.000000	
max	21.000000	535.000000	543.000000	391.000000	

	adr	required_car_parking_spaces	total_of_special_requests
count	119390.000000	119390.000000	119390.000000
mean	101.831122	0.062518	0.571363
std	50.535790	0.245291	0.792798
min	-6.380000	0.000000	0.000000
25%	69.290000	0.000000	0.000000
50%	94.575000	0.000000	0.000000
75%	126.000000	0.000000	1.000000
max	5400.000000	8.000000	5.000000

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

```
[ ]: 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	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

```
[ ]: df.shape
```

```
[ ]: (119390, 32)
```

```
[ ]: df.columns
```

```
[ ]: 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')
```

```
[ ]: missing_values=df.isnull().sum()
missing_columns=missing_values[missing_values>0]
print(missing_columns)
```

```

children      4
country       488
agent        16340
company      112593
dtype: int64

```

```
[ ]: df["children"]=df["children"].fillna(df["children"].mode()[0])
df["country"]=df["country"].fillna(df["country"].mode()[0])
df["agent"]=df["agent"].fillna(df["agent"].mode()[0])
```

```
[ ]: missing_values=df.isnull().sum()
missing_columns=missing_values[missing_values>0]
print(missing_columns)
```

```

company      112593
dtype: int64

```

```
[ ]: df=df.drop("company",axis=1)
```

```
[ ]: num=df.select_dtypes(include=['int64','float64'])
num
```

```
[ ]:
      is_canceled  lead_time  arrival_date_year  arrival_date_week_number \
0                0       342             2015                27
1                0       737             2015                27
2                0         7             2015                27
3                0        13             2015                27
4                0        14             2015                27
...            ...      ...             ...                ...
119385           0         23             2017                35
119386           0        102             2017                35
119387           0         34             2017                35
119388           0        109             2017                35
119389           0        205             2017                35
```

```

      arrival_date_day_of_month  stays_in_weekend_nights \
0                             1                        0
1                             1                        0
2                             1                        0
3                             1                        0
4                             1                        0
...                          ...                      ...
119385                       30                        2
119386                       31                        2
119387                       31                        2
119388                       31                        2
119389                       29                        2

```

	stays_in_week_nights	adults	children	babies	is_repeated_guest	\
0	0	2	0.0	0	0	
1	0	2	0.0	0	0	
2	1	1	0.0	0	0	
3	1	1	0.0	0	0	
4	2	2	0.0	0	0	
...	
119385	5	2	0.0	0	0	
119386	5	3	0.0	0	0	
119387	5	2	0.0	0	0	
119388	5	2	0.0	0	0	
119389	7	2	0.0	0	0	

	previous_cancellations	previous_bookings_not_canceled	\
0	0	0	
1	0	0	
2	0	0	
3	0	0	
4	0	0	
...	
119385	0	0	
119386	0	0	
119387	0	0	
119388	0	0	
119389	0	0	

	booking_changes	agent	days_in_waiting_list	adr	\
0	3	9.0	0	0.00	
1	4	9.0	0	0.00	
2	0	9.0	0	75.00	
3	0	304.0	0	75.00	
4	0	240.0	0	98.00	
...	
119385	0	394.0	0	96.14	
119386	0	9.0	0	225.43	
119387	0	9.0	0	157.71	
119388	0	89.0	0	104.40	
119389	0	9.0	0	151.20	

	required_car_parking_spaces	total_of_special_requests
0	0	0
1	0	0
2	0	0
3	0	0
4	0	1
...

119385	0	0
119386	0	2
119387	0	4
119388	0	0
119389	0	2

[119390 rows x 19 columns]

```
[ ]: num.shape
```

```
[ ]: (119390, 19)
```

```
[ ]: num.corr()
```

```
[ ]:
```

	is_canceled	lead_time	arrival_date_year	\
is_canceled	1.000000	0.293123	0.016660	
lead_time	0.293123	1.000000	0.040142	
arrival_date_year	0.016660	0.040142	1.000000	
arrival_date_week_number	0.008148	0.126871	-0.540561	
arrival_date_day_of_month	-0.006130	0.002268	-0.000221	
stays_in_weekend_nights	-0.001791	0.085671	0.021497	
stays_in_week_nights	0.024765	0.165799	0.030883	
adults	0.060017	0.119519	0.029635	
children	0.005036	-0.037613	0.054636	
babies	-0.032491	-0.020915	-0.013192	
is_repeated_guest	-0.084793	-0.124410	0.010341	
previous_cancellations	0.110133	0.086042	-0.119822	
previous_bookings_not_canceled	-0.057358	-0.073548	0.029218	
booking_changes	-0.144381	0.000149	0.030872	
agent	-0.049857	-0.017975	0.056905	
days_in_waiting_list	0.054186	0.170084	-0.056497	
adr	0.047557	-0.063077	0.197580	
required_car_parking_spaces	-0.195498	-0.116451	-0.013684	
total_of_special_requests	-0.234658	-0.095712	0.108531	

	arrival_date_week_number	\
is_canceled	0.008148	
lead_time	0.126871	
arrival_date_year	-0.540561	
arrival_date_week_number	1.000000	
arrival_date_day_of_month	0.066809	
stays_in_weekend_nights	0.018208	
stays_in_week_nights	0.015558	
adults	0.025909	
children	0.005515	
babies	0.010395	
is_repeated_guest	-0.030131	

previous_cancellations	0.035501
previous_bookings_not_canceled	-0.020904
booking_changes	0.005508
agent	-0.019320
days_in_waiting_list	0.022933
adr	0.075791
required_car_parking_spaces	0.001920
total_of_special_requests	0.026149

	arrival_date_day_of_month \
is_canceled	-0.006130
lead_time	0.002268
arrival_date_year	-0.000221
arrival_date_week_number	0.066809
arrival_date_day_of_month	1.000000
stays_in_weekend_nights	-0.016354
stays_in_week_nights	-0.028174
adults	-0.001566
children	0.014553
babies	-0.000230
is_repeated_guest	-0.006145
previous_cancellations	-0.027011
previous_bookings_not_canceled	-0.000300
booking_changes	0.010613
agent	0.000321
days_in_waiting_list	0.022728
adr	0.030245
required_car_parking_spaces	0.008683
total_of_special_requests	0.003062

	stays_in_weekend_nights	stays_in_week_nights \
is_canceled	-0.001791	0.024765
lead_time	0.085671	0.165799
arrival_date_year	0.021497	0.030883
arrival_date_week_number	0.018208	0.015558
arrival_date_day_of_month	-0.016354	-0.028174
stays_in_weekend_nights	1.000000	0.498969
stays_in_week_nights	0.498969	1.000000
adults	0.091871	0.092976
children	0.045794	0.044203
babies	0.018483	0.020191
is_repeated_guest	-0.087239	-0.097245
previous_cancellations	-0.012775	-0.013992
previous_bookings_not_canceled	-0.042715	-0.048743
booking_changes	0.063281	0.096209
agent	0.158992	0.193228
days_in_waiting_list	-0.054151	-0.002020

adr	0.049342	0.065237
required_car_parking_spaces	-0.018554	-0.024859
total_of_special_requests	0.072671	0.068192

	adults	children	babies \
is_canceled	0.060017	0.005036	-0.032491
lead_time	0.119519	-0.037613	-0.020915
arrival_date_year	0.029635	0.054636	-0.013192
arrival_date_week_number	0.025909	0.005515	0.010395
arrival_date_day_of_month	-0.001566	0.014553	-0.000230
stays_in_weekend_nights	0.091871	0.045794	0.018483
stays_in_week_nights	0.092976	0.044203	0.020191
adults	1.000000	0.030440	0.018146
children	0.030440	1.000000	0.024030
babies	0.018146	0.024030	1.000000
is_repeated_guest	-0.146426	-0.032858	-0.008943
previous_cancellations	-0.006738	-0.024729	-0.007501
previous_bookings_not_canceled	-0.107983	-0.021072	-0.006550
booking_changes	-0.051673	0.048952	0.083440
agent	0.019754	0.049641	0.030696
days_in_waiting_list	-0.008283	-0.033271	-0.010621
adr	0.230641	0.324853	0.029186
required_car_parking_spaces	0.014785	0.056255	0.037383
total_of_special_requests	0.122884	0.081736	0.097889

	is_repeated_guest	previous_cancellations \
is_canceled	-0.084793	0.110133
lead_time	-0.124410	0.086042
arrival_date_year	0.010341	-0.119822
arrival_date_week_number	-0.030131	0.035501
arrival_date_day_of_month	-0.006145	-0.027011
stays_in_weekend_nights	-0.087239	-0.012775
stays_in_week_nights	-0.097245	-0.013992
adults	-0.146426	-0.006738
children	-0.032858	-0.024729
babies	-0.008943	-0.007501
is_repeated_guest	1.000000	0.082293
previous_cancellations	0.082293	1.000000
previous_bookings_not_canceled	0.418056	0.152728
booking_changes	0.012092	-0.026993
agent	-0.045230	-0.017475
days_in_waiting_list	-0.022235	0.005929
adr	-0.134314	-0.065646
required_car_parking_spaces	0.077090	-0.018492
total_of_special_requests	0.013050	-0.048384

previous_bookings_not_canceled \

is_canceled	-0.057358
lead_time	-0.073548
arrival_date_year	0.029218
arrival_date_week_number	-0.020904
arrival_date_day_of_month	-0.000300
stays_in_weekend_nights	-0.042715
stays_in_week_nights	-0.048743
adults	-0.107983
children	-0.021072
babies	-0.006550
is_repeated_guest	0.418056
previous_cancellations	0.152728
previous_bookings_not_canceled	1.000000
booking_changes	0.011608
agent	-0.041177
days_in_waiting_list	-0.009397
adr	-0.072144
required_car_parking_spaces	0.047653
total_of_special_requests	0.037824

	booking_changes	agent \
is_canceled	-0.144381	-0.049857
lead_time	0.000149	-0.017975
arrival_date_year	0.030872	0.056905
arrival_date_week_number	0.005508	-0.019320
arrival_date_day_of_month	0.010613	0.000321
stays_in_weekend_nights	0.063281	0.158992
stays_in_week_nights	0.096209	0.193228
adults	-0.051673	0.019754
children	0.048952	0.049641
babies	0.083440	0.030696
is_repeated_guest	0.012092	-0.045230
previous_cancellations	-0.026993	-0.017475
previous_bookings_not_canceled	0.011608	-0.041177
booking_changes	1.000000	0.038935
agent	0.038935	1.000000
days_in_waiting_list	-0.011634	-0.042391
adr	0.019618	0.012805
required_car_parking_spaces	0.065620	0.122846
total_of_special_requests	0.052833	0.058029

	days_in_waiting_list	adr \
is_canceled	0.054186	0.047557
lead_time	0.170084	-0.063077
arrival_date_year	-0.056497	0.197580
arrival_date_week_number	0.022933	0.075791
arrival_date_day_of_month	0.022728	0.030245

stays_in_weekend_nights	-0.054151	0.049342
stays_in_week_nights	-0.002020	0.065237
adults	-0.008283	0.230641
children	-0.033271	0.324853
babies	-0.010621	0.029186
is_repeated_guest	-0.022235	-0.134314
previous_cancellations	0.005929	-0.065646
previous_bookings_not_canceled	-0.009397	-0.072144
booking_changes	-0.011634	0.019618
agent	-0.042391	0.012805
days_in_waiting_list	1.000000	-0.040756
adr	-0.040756	1.000000
required_car_parking_spaces	-0.030600	0.056628
total_of_special_requests	-0.082730	0.172185

	required_car_parking_spaces	\
is_canceled	-0.195498	
lead_time	-0.116451	
arrival_date_year	-0.013684	
arrival_date_week_number	0.001920	
arrival_date_day_of_month	0.008683	
stays_in_weekend_nights	-0.018554	
stays_in_week_nights	-0.024859	
adults	0.014785	
children	0.056255	
babies	0.037383	
is_repeated_guest	0.077090	
previous_cancellations	-0.018492	
previous_bookings_not_canceled	0.047653	
booking_changes	0.065620	
agent	0.122846	
days_in_waiting_list	-0.030600	
adr	0.056628	
required_car_parking_spaces	1.000000	
total_of_special_requests	0.082626	

	total_of_special_requests
is_canceled	-0.234658
lead_time	-0.095712
arrival_date_year	0.108531
arrival_date_week_number	0.026149
arrival_date_day_of_month	0.003062
stays_in_weekend_nights	0.072671
stays_in_week_nights	0.068192
adults	0.122884
children	0.081736
babies	0.097889

```

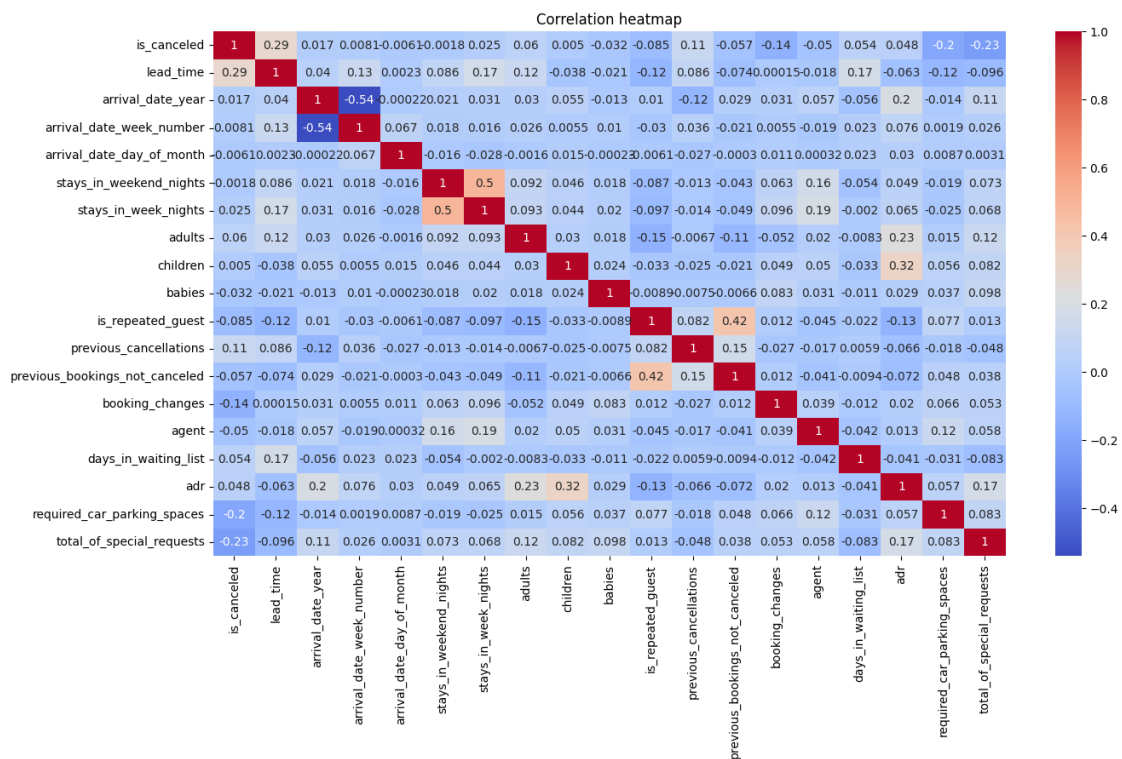
is_repeated_guest          0.013050
previous_cancellations     -0.048384
previous_bookings_not_canceled 0.037824
booking_changes            0.052833
agent                     0.058029
days_in_waiting_list      -0.082730
adr                       0.172185
required_car_parking_spaces 0.082626
total_of_special_requests   1.000000

```

```

[ ]: plt.figure(figsize=(15,8))
sns.heatmap(num.corr(),annot=True,cmap="coolwarm")
plt.title("Correlation heatmap")
plt.show()

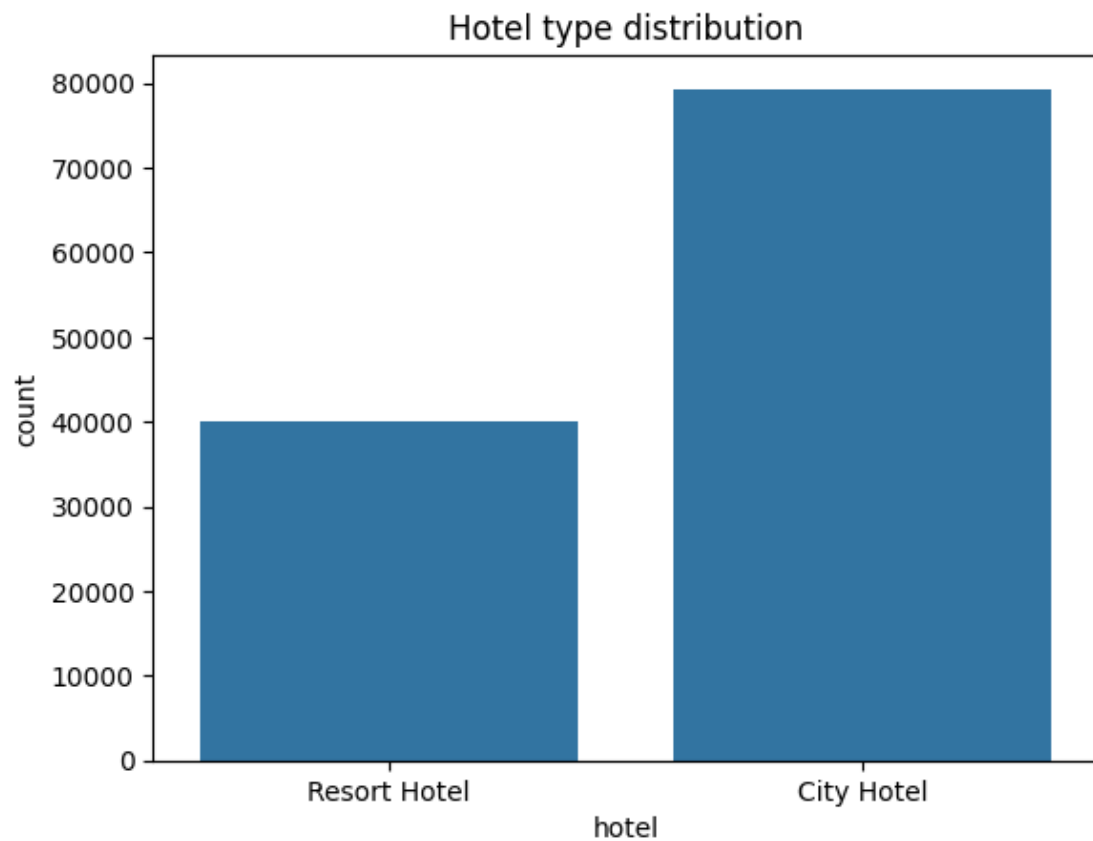
```



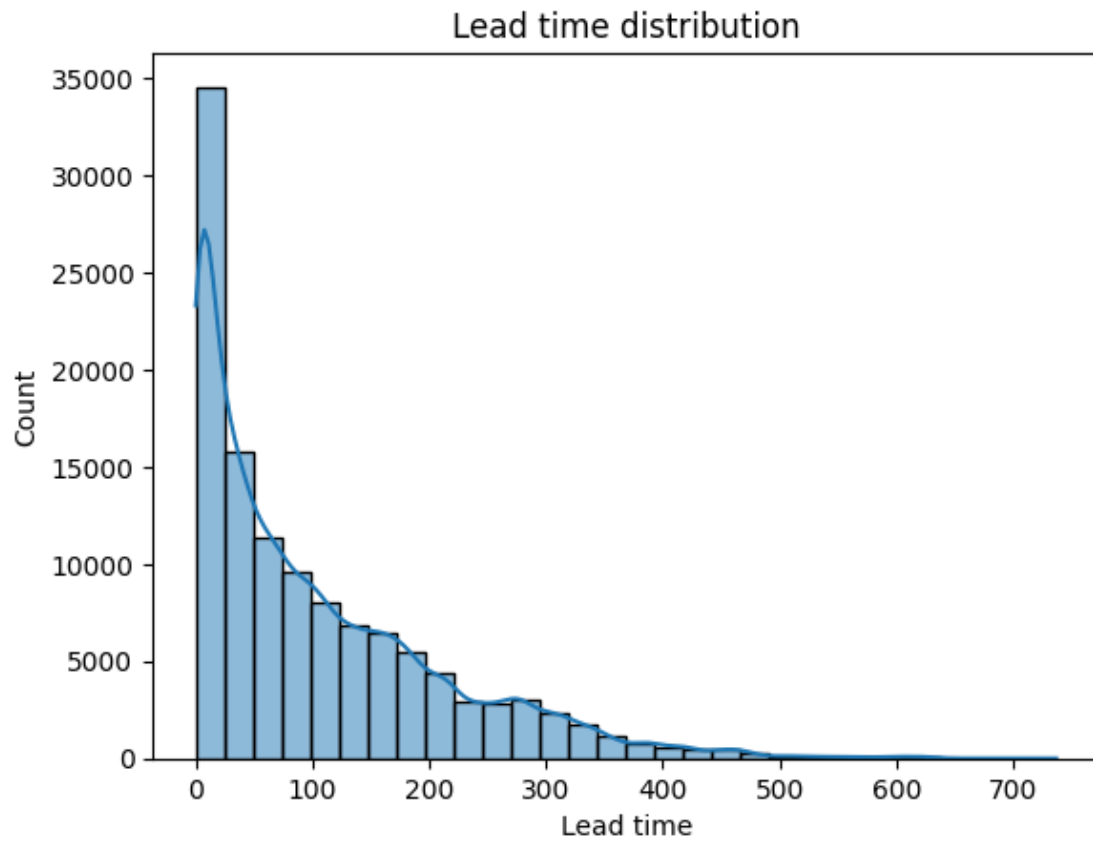
```

[ ]: sns.countplot(x="hotel",data=df)
plt.title("Hotel type distribution")
plt.show()

```



```
[ ]: sns.histplot(df["lead_time"],bins=30,kde=True)  
plt.xlabel("Lead time")  
plt.title("Lead time distribution")  
plt.show()
```



```
[ ]: sns.boxplot(x='is_canceled',y='adr',data=df)
plt.title("ADR vs Cancellation")
plt.show()
```

