**INPUT CODE:**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Step 1: Load the Dataset

df = pd.read\_csv('/content/drive/MyDrive/hotel\_bookings.csv')  # Replace 'your\_dataset.csv' with the actual file path

# Step 2: Check the Structure

print("Dimensions of the dataset:", df.shape)

print("\nFirst few rows of the dataset:")

print(df.head())

# Step 3: Understand the Variables

print("\nData types of variables:")

print(df.dtypes)

# Step 4: Summary Statistics

print("\nSummary statistics for numerical variables:")

print(df.describe())

# Step 5: Distribution of Numerical Variables

numerical\_variables = df.select\_dtypes(include=['int64', 'float64']).columns

for var in numerical\_variables:

    plt.figure(figsize=(8, 5))

    sns.histplot(df[var], kde=True)

    plt.title(f'Distribution of {var}')

    plt.xlabel(var)

    plt.ylabel('Frequency')

    plt.show()

# Step 6: Explore Categorical Variables

categorical\_variables = df.select\_dtypes(include=['object']).columns

for var in categorical\_variables:

    plt.figure(figsize=(8, 5))

    sns.countplot(data=df, x=var)

    plt.title(f'Distribution of {var}')

    plt.xlabel(var)

    plt.ylabel('Count')

    plt.xticks(rotation=45)

    plt.show()

# Step 7: Correlation Analysis

correlation\_matrix = df.corr()

plt.figure(figsize=(10, 8))

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', fmt=".2f")

plt.title('Correlation Matrix')

plt.show()

# Step 8: Handle Missing Values

print("\nNumber of missing values in each column:")

print(df.isnull().sum())

**OUTPUT:**

Dimensions of the dataset: (119390, 32)

First few rows of the dataset:

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 01-07-2015

1 01-07-2015

2 02-07-2015

3 02-07-2015

4 03-07-2015

[5 rows x 32 columns]

Data types of variables:

hotel object

is\_canceled int64

lead\_time int64

arrival\_date\_year int64

arrival\_date\_month object

arrival\_date\_week\_number int64

arrival\_date\_day\_of\_month int64

stays\_in\_weekend\_nights int64

stays\_in\_week\_nights int64

adults int64

children float64

babies int64

meal object

country object

market\_segment object

distribution\_channel object

is\_repeated\_guest int64

previous\_cancellations int64

previous\_bookings\_not\_canceled int64

reserved\_room\_type object

assigned\_room\_type object

booking\_changes int64

deposit\_type object

agent float64

company float64

days\_in\_waiting\_list int64

customer\_type object

adr float64

required\_car\_parking\_spaces int64

total\_of\_special\_requests int64

reservation\_status object

reservation\_status\_date object

dtype: object

Summary statistics for numerical variables:

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

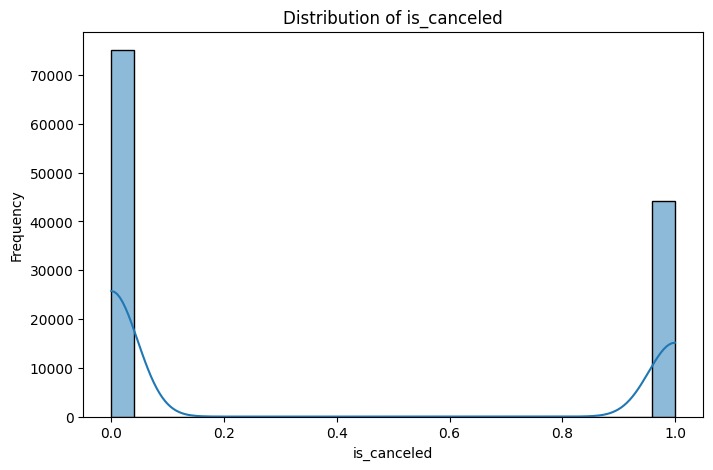
min -6.380000 0.000000 0.000000

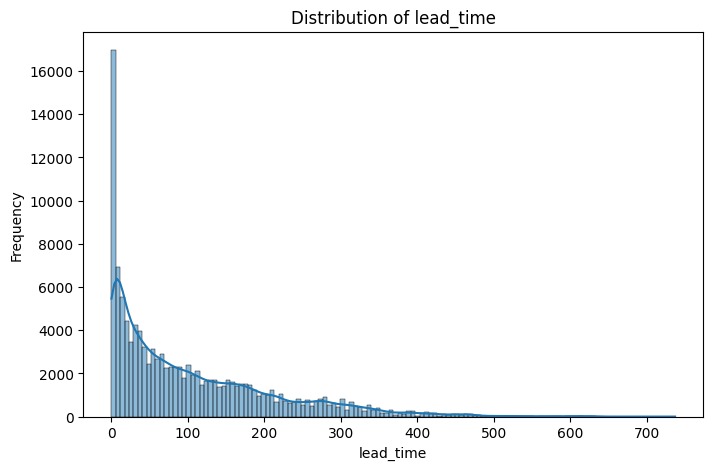
25% 69.290000 0.000000 0.000000

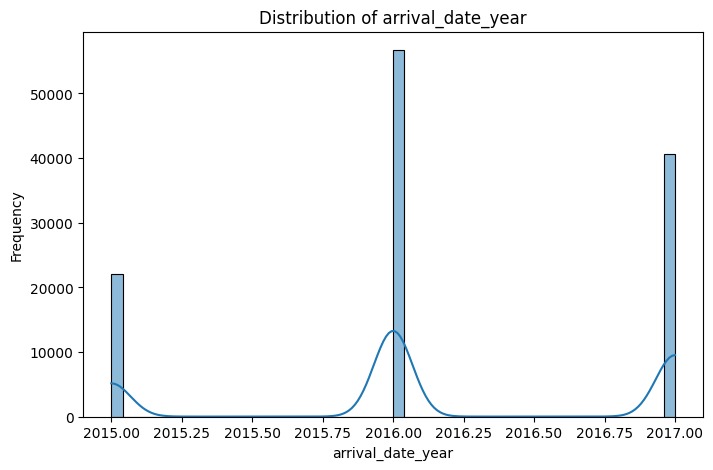
50% 94.575000 0.000000 0.000000

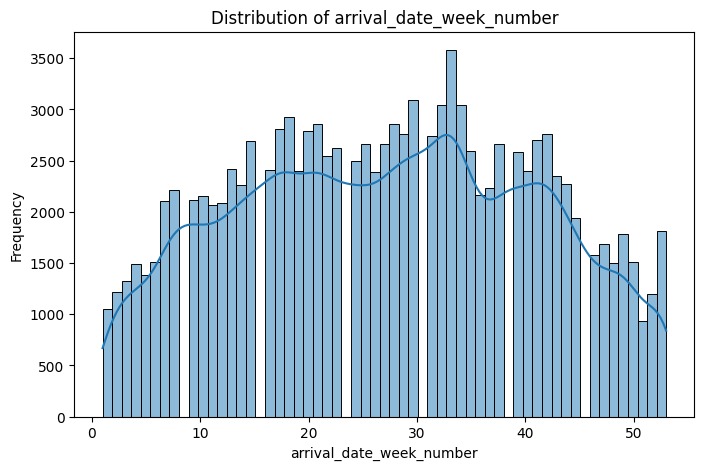
75% 126.000000 0.000000 1.000000

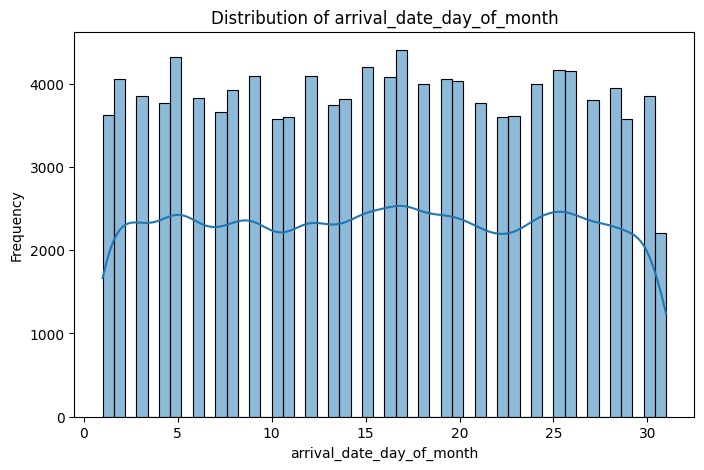
max 5400.000000 8.000000 5.000000

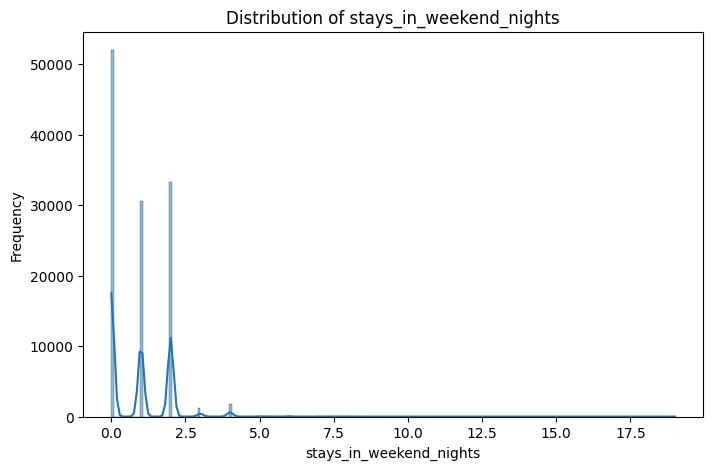
****

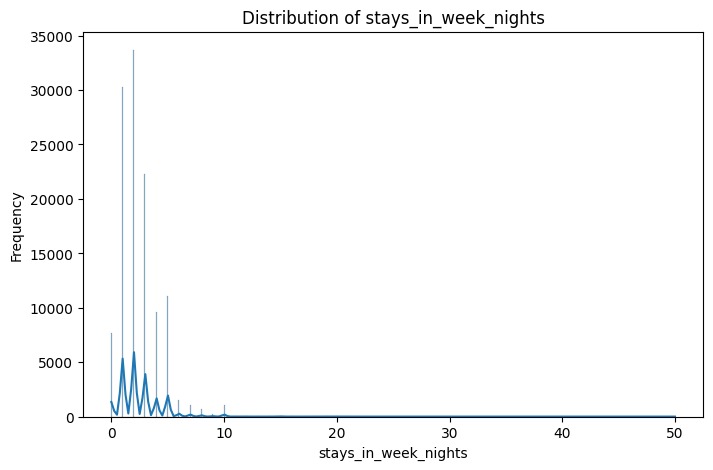
****

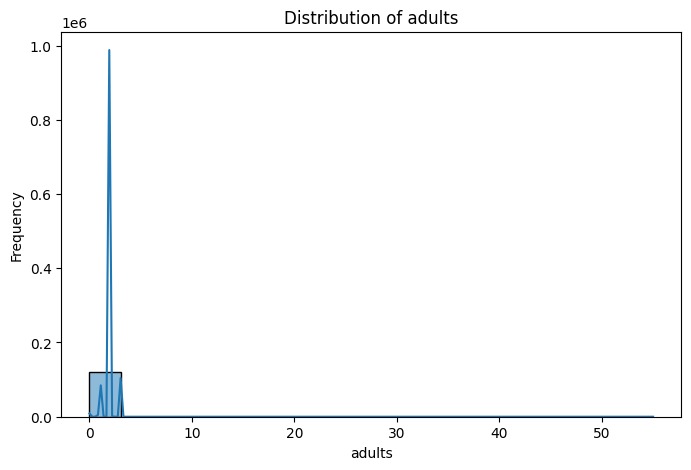
****

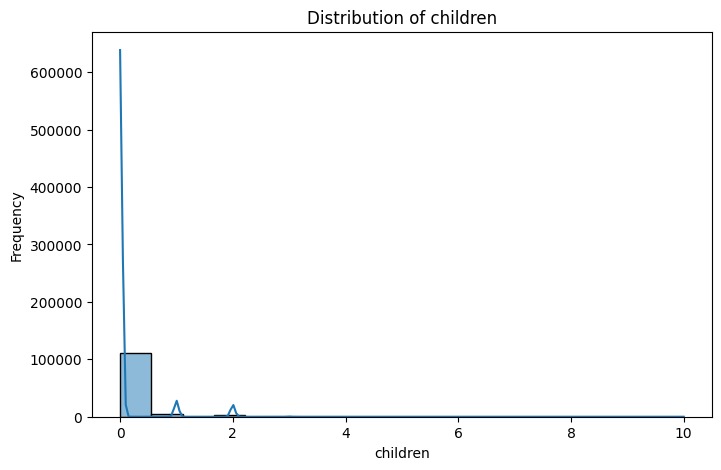
****

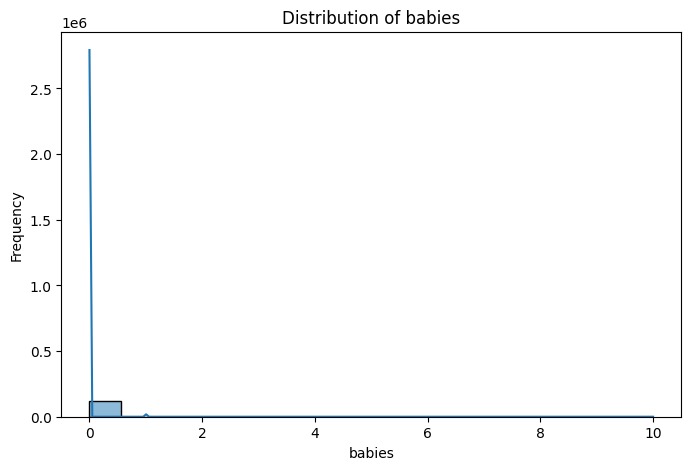
****

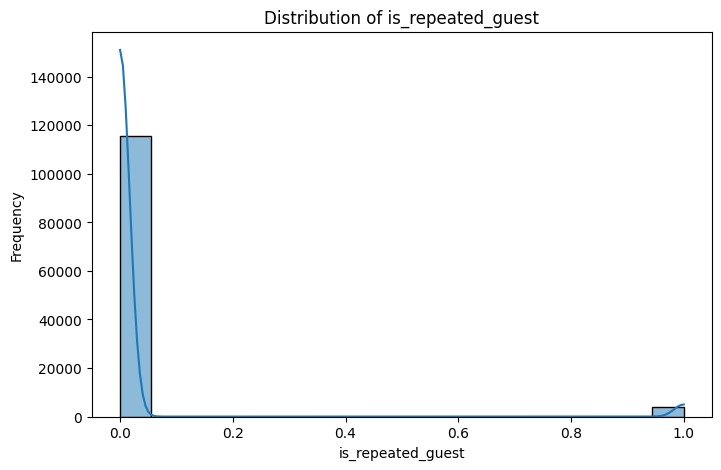
****

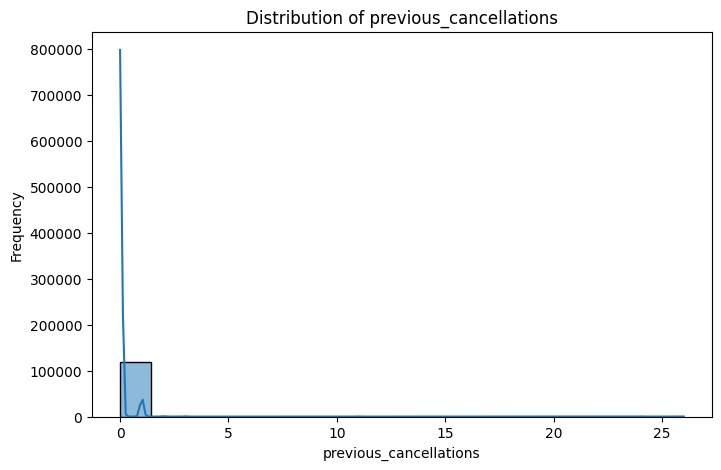
****

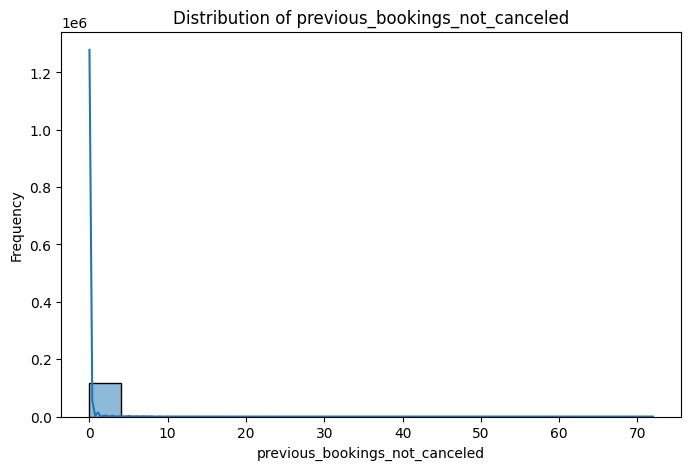
****

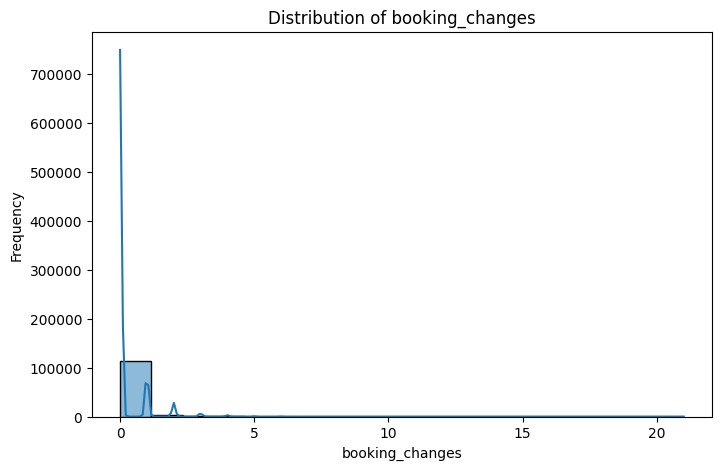
****

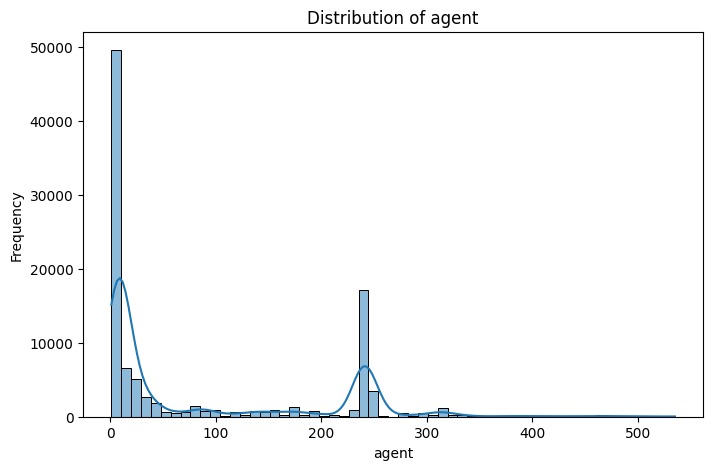
****

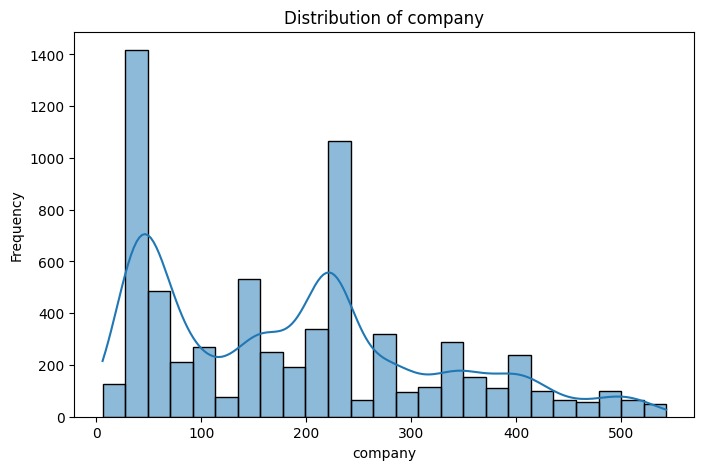
****

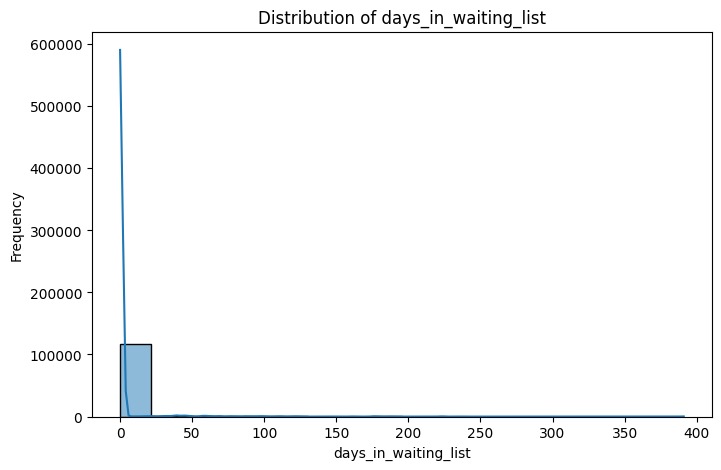
****

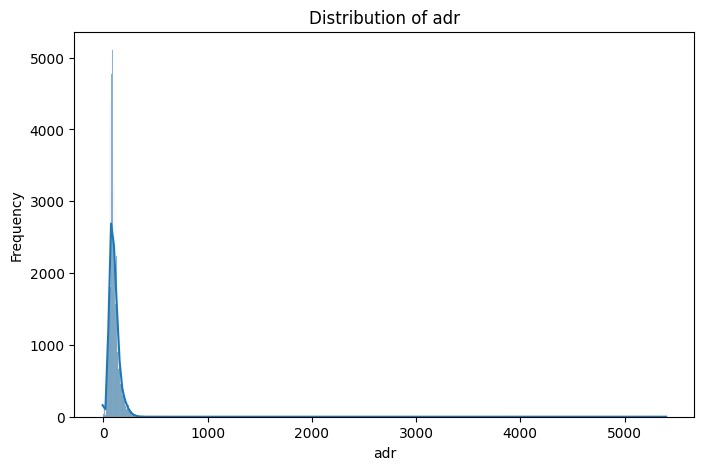
****

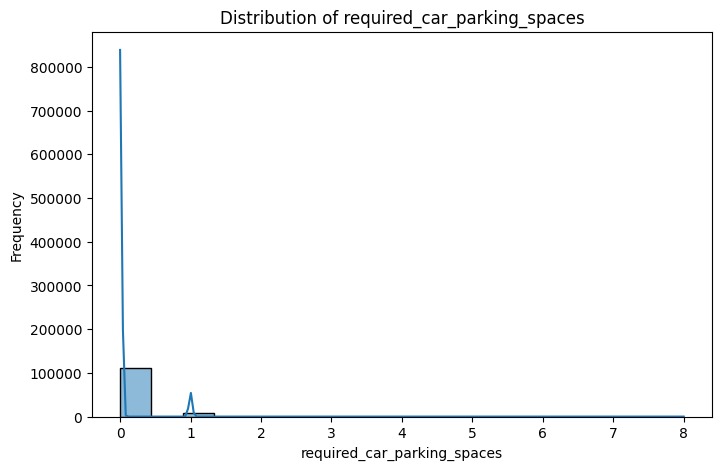
****

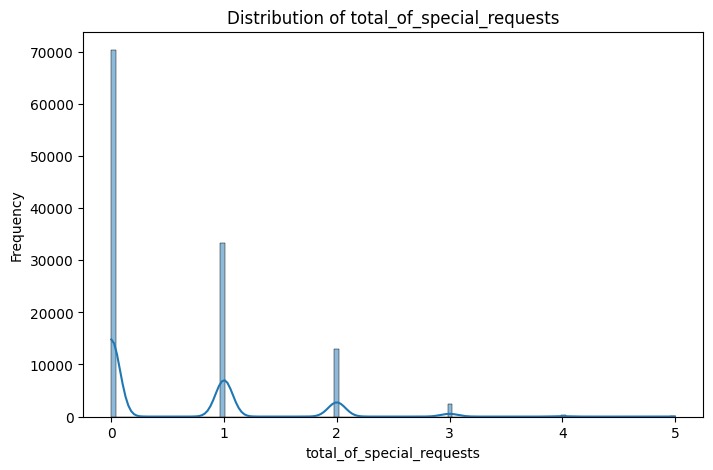
****

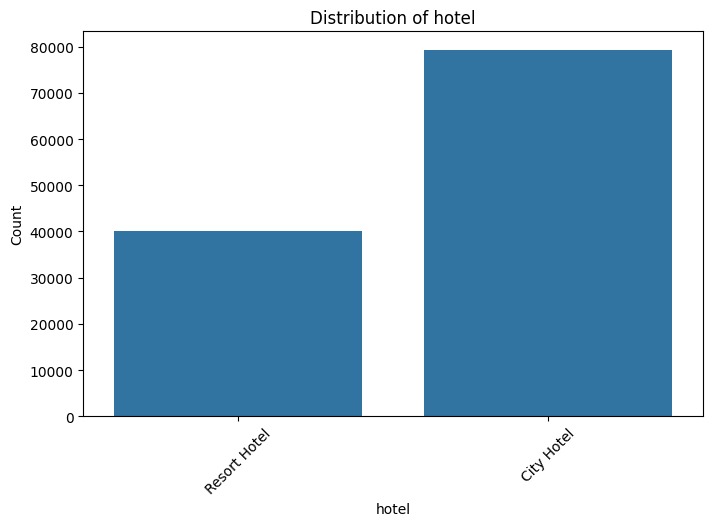
****

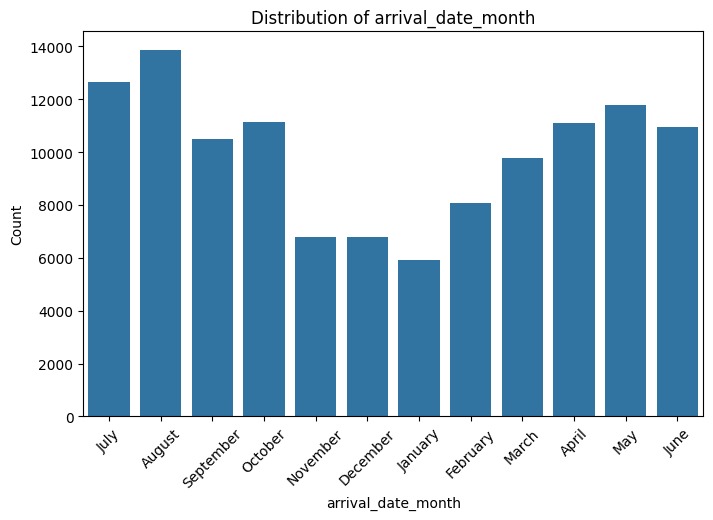
****

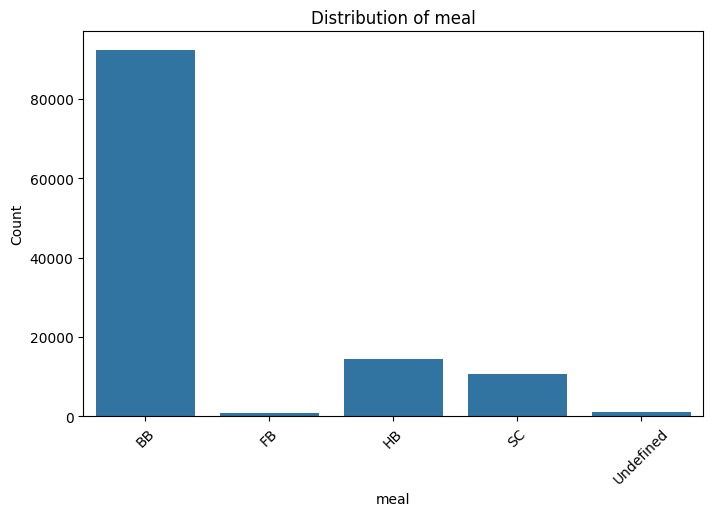
****

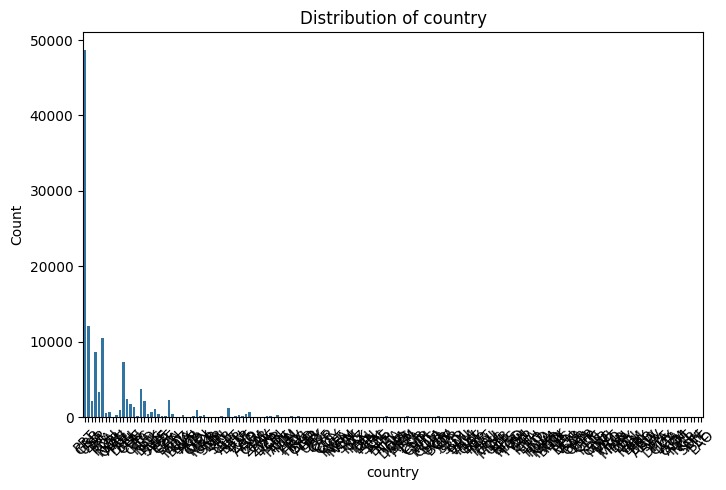
****

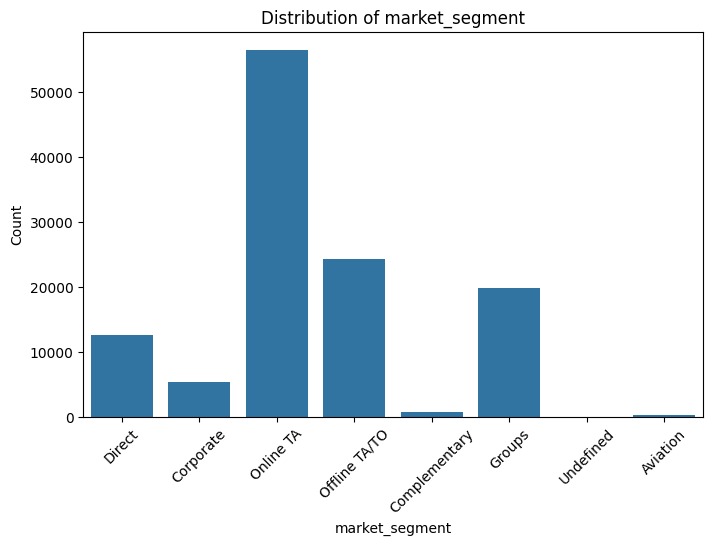
****

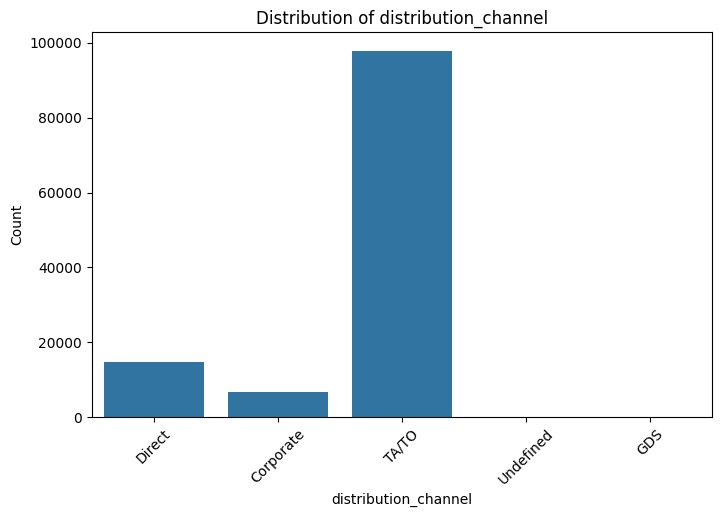
****

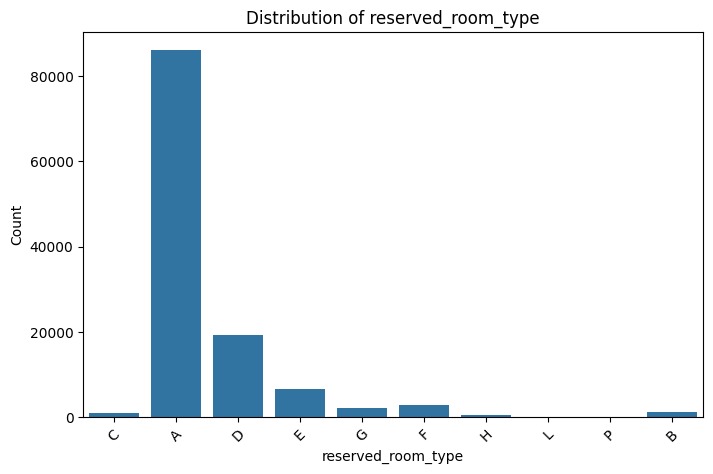
****

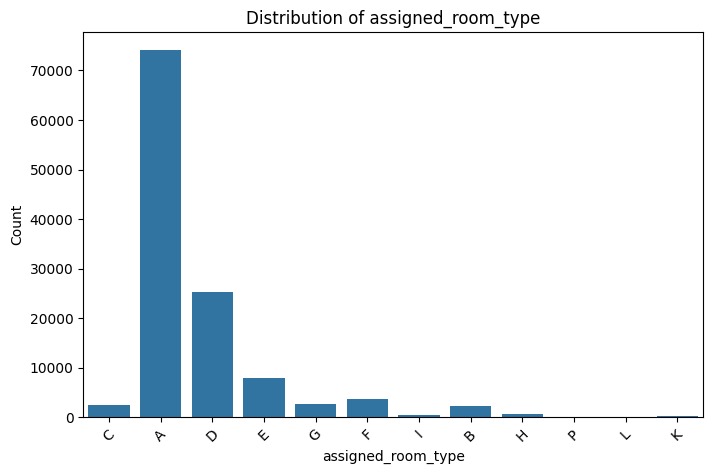
****

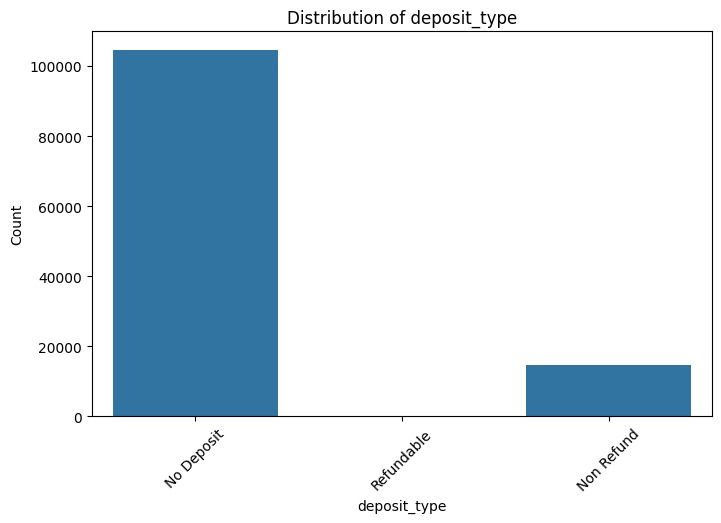
****

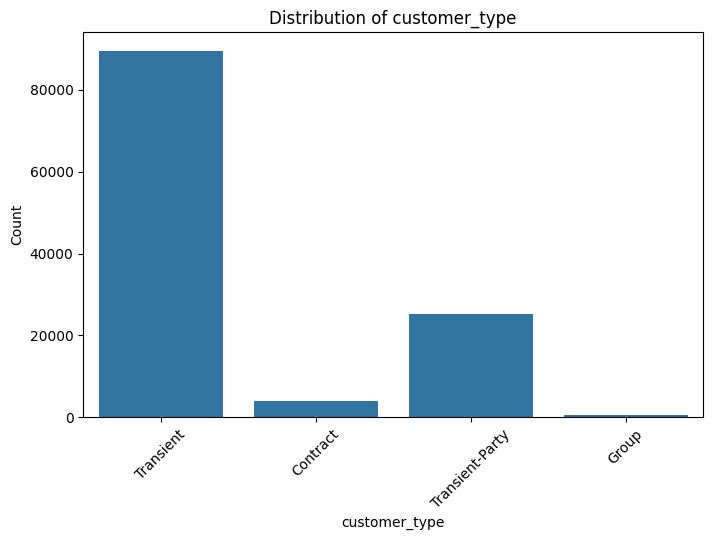
****

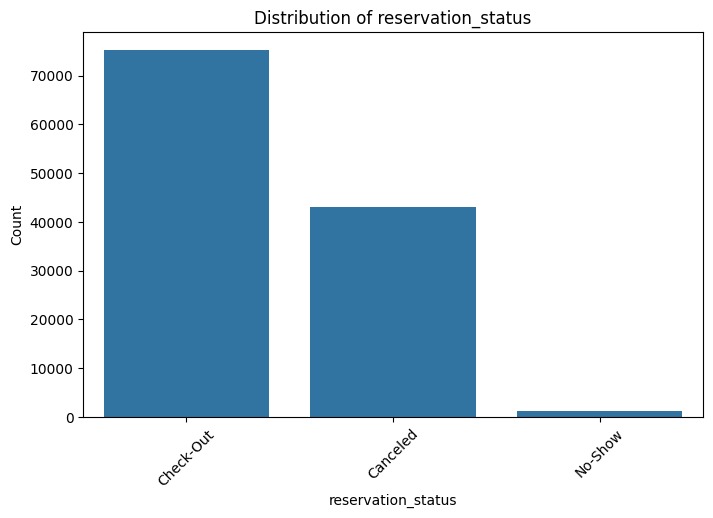
****

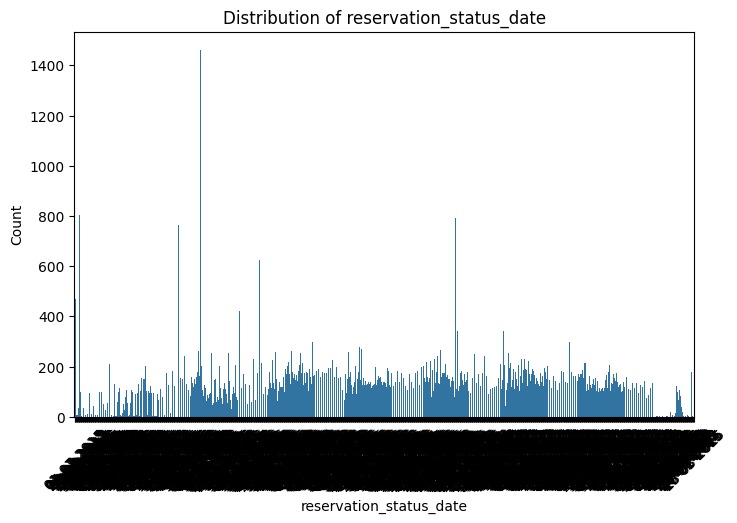
****

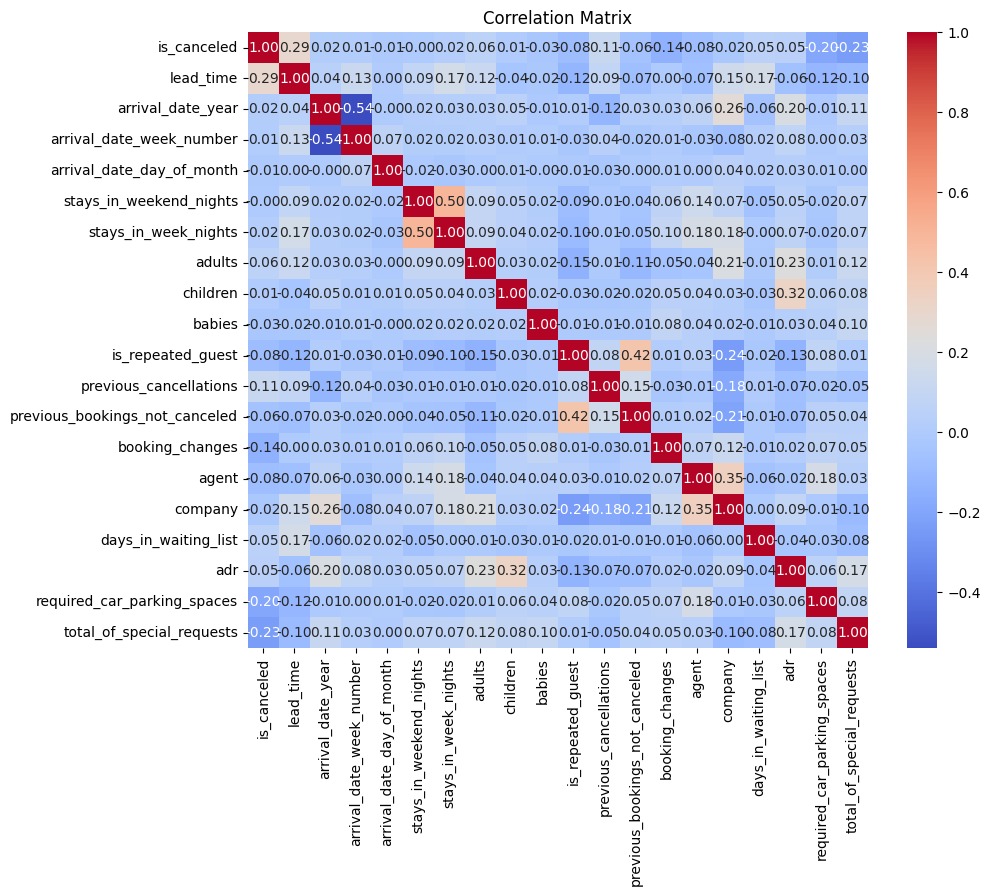
****

****

****

****

****

****