

```
In [131... import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

```
In [133... df_air = pd.read_csv("air.csv")
df_air.head()
```

```
Out[133]:
```

	Unnamed: 0	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure time
0	0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	
1	1	5047	Male	disloyal Customer	25	Business travel	Business	235	3	
2	2	110028	Female	Loyal Customer	26	Business travel	Business	1142	2	
3	3	24026	Female	Loyal Customer	25	Business travel	Business	562	2	
4	4	119299	Male	Loyal Customer	61	Business travel	Business	214	3	

5 rows x 25 columns

```
In [135... df_air.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103904 entries, 0 to 103903
Data columns (total 25 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Unnamed: 0                               103904 non-null  int64
1   id                                         103904 non-null  int64
2   Gender                                   103904 non-null  object
3   Customer Type                            103904 non-null  object
4   Age                                       103904 non-null  int64
5   Type of Travel                           103904 non-null  object
6   Class                                    103904 non-null  object
7   Flight Distance                          103904 non-null  int64
8   Inflight wifi service                    103904 non-null  int64
9   Departure/Arrival time convenient        103904 non-null  int64
10  Ease of Online booking                   103904 non-null  int64
11  Gate location                            103904 non-null  int64
12  Food and drink                           103904 non-null  int64
13  Online boarding                          103904 non-null  int64
14  Seat comfort                             103904 non-null  int64
15  Inflight entertainment                   103904 non-null  int64
16  On-board service                         103904 non-null  int64
17  Leg room service                         103904 non-null  int64
18  Baggage handling                         103904 non-null  int64
19  Checkin service                         103904 non-null  int64
20  Inflight service                         103904 non-null  int64
21  Cleanliness                             103904 non-null  int64
22  Departure Delay in Minutes               103904 non-null  int64
23  Arrival Delay in Minutes                 103594 non-null  float64
24  satisfaction                             103904 non-null  object
dtypes: float64(1), int64(19), object(5)
memory usage: 19.8+ MB
```

In [137... `df_air.describe()`

Out[137]:

	Unnamed: 0	id	Age	Flight Distance	Inflight wifi service	Departure Delay in Minutes
<b>count</b>	103904.000000	103904.000000	103904.000000	103904.000000	103904.000000	103594.000000
<b>mean</b>	51951.500000	64924.210502	39.379706	1189.448375	2.729683	15.114964
<b>std</b>	29994.645522	37463.812252	15.114964	997.147281	1.327829	15.114964
<b>min</b>	0.000000	1.000000	7.000000	31.000000	0.000000	0.000000
<b>25%</b>	25975.750000	32533.750000	27.000000	414.000000	2.000000	0.000000
<b>50%</b>	51951.500000	64856.500000	40.000000	843.000000	3.000000	0.000000
<b>75%</b>	77927.250000	97368.250000	51.000000	1743.000000	4.000000	0.000000
<b>max</b>	103903.000000	129880.000000	85.000000	4983.000000	5.000000	0.000000

In [139... `df_air.isnull().sum()`

```
Out[139]: Unnamed: 0      0
          id          0
          Gender       0
          Customer Type 0
          Age          0
          Type of Travel 0
          Class        0
          Flight Distance 0
          Inflight wifi service 0
          Departure/Arrival time convenient 0
          Ease of Online booking 0
          Gate location 0
          Food and drink 0
          Online boarding 0
          Seat comfort 0
          Inflight entertainment 0
          On-board service 0
          Leg room service 0
          Baggage handling 0
          Checkin service 0
          Inflight service 0
          Cleanliness 0
          Departure Delay in Minutes 0
          Arrival Delay in Minutes 310
          satisfaction 0
          dtype: int64
```

```
In [141]: df_air.head()
```

```
Out[141]:
```

	Unnamed: 0	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure time
0	0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	
1	1	5047	Male	disloyal Customer	25	Business travel	Business	235	3	
2	2	110028	Female	Loyal Customer	26	Business travel	Business	1142	2	
3	3	24026	Female	Loyal Customer	25	Business travel	Business	562	2	
4	4	119299	Male	Loyal Customer	61	Business travel	Business	214	3	

5 rows x 25 columns

```
In [143]: df_air.dropna().head()
```

Out[143]:

	Unnamed: 0	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure time
0	0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	
1	1	5047	Male	disloyal Customer	25	Business travel	Business	235	3	
2	2	110028	Female	Loyal Customer	26	Business travel	Business	1142	2	
3	3	24026	Female	Loyal Customer	25	Business travel	Business	562	2	
4	4	119299	Male	Loyal Customer	61	Business travel	Business	214	3	

5 rows x 25 columns

In [145... df\_air["Arrival Delay in Minutes"] = df\_air["Arrival Delay in Minutes"].fill

In [147... df\_air.isnull().sum()

Out[147]:

Unnamed: 0	0
id	0
Gender	0
Customer Type	0
Age	0
Type of Travel	0
Class	0
Flight Distance	0
Inflight wifi service	0
Departure/Arrival time convenient	0
Ease of Online booking	0
Gate location	0
Food and drink	0
Online boarding	0
Seat comfort	0
Inflight entertainment	0
On-board service	0
Leg room service	0
Baggage handling	0
Checkin service	0
Inflight service	0
Cleanliness	0
Departure Delay in Minutes	0
Arrival Delay in Minutes	0
satisfaction	0
dtype:	int64

In [149... df = df\_air.drop(columns = ["id"], inplace = True, axis = 1)

In [151... df\_air.head()

Out[151]:

	Unnamed: 0	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arr time conven
0	0	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	
1	1	Male	disloyal Customer	25	Business travel	Business	235	3	
2	2	Female	Loyal Customer	26	Business travel	Business	1142	2	
3	3	Female	Loyal Customer	25	Business travel	Business	562	2	
4	4	Male	Loyal Customer	61	Business travel	Business	214	3	

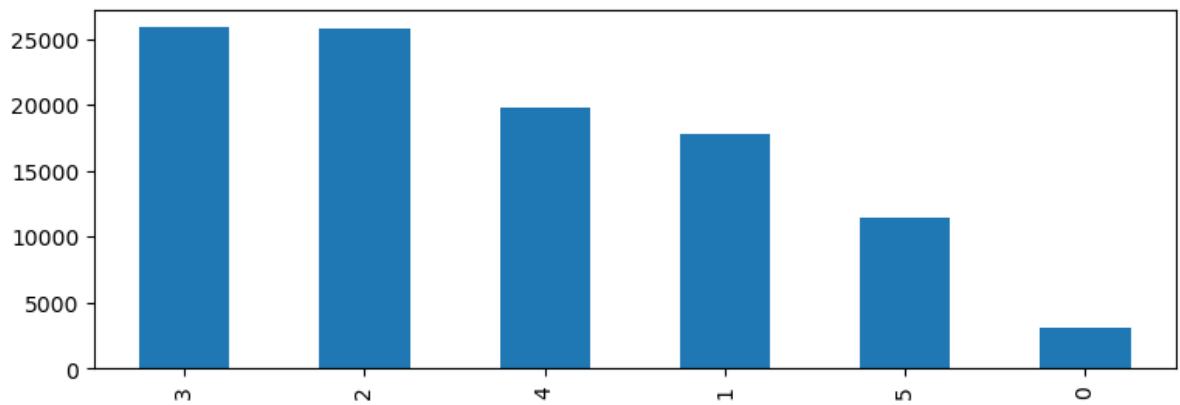
5 rows × 24 columns

In [153... df\_air.isnull().sum()

```
Out[153]: Unnamed: 0      0
Gender      0
Customer Type      0
Age      0
Type of Travel      0
Class      0
Flight Distance      0
Inflight wifi service      0
Departure/Arrival time convenient      0
Ease of Online booking      0
Gate location      0
Food and drink      0
Online boarding      0
Seat comfort      0
Inflight entertainment      0
On-board service      0
Leg room service      0
Baggage handling      0
Checkin service      0
Inflight service      0
Cleanliness      0
Departure Delay in Minutes      0
Arrival Delay in Minutes      0
satisfaction      0
dtype: int64
```

```
In [155... var = df_air["Inflight wifi service"].value_counts()

plt.figure(figsize=(9, 3))
var.plot(kind='bar')
plt.show()
```



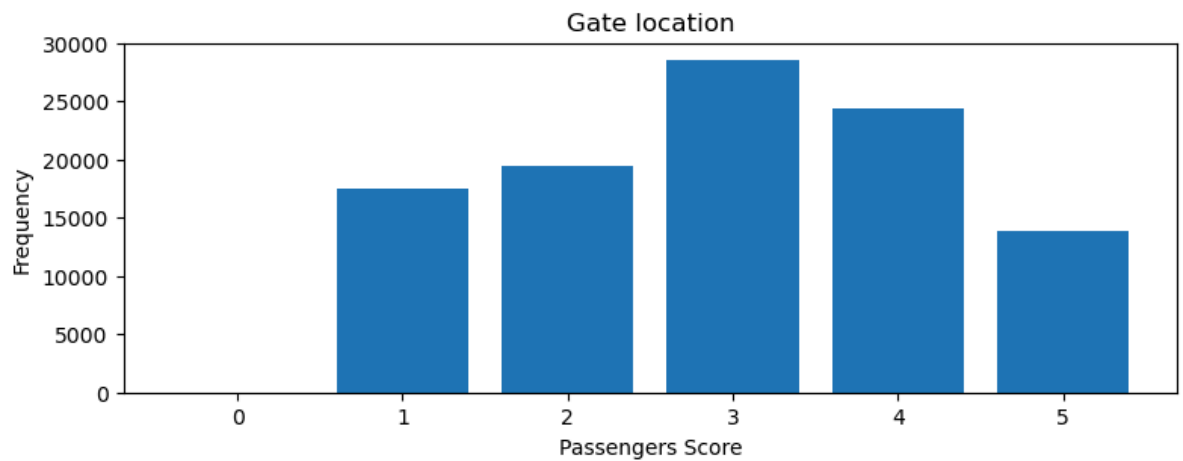
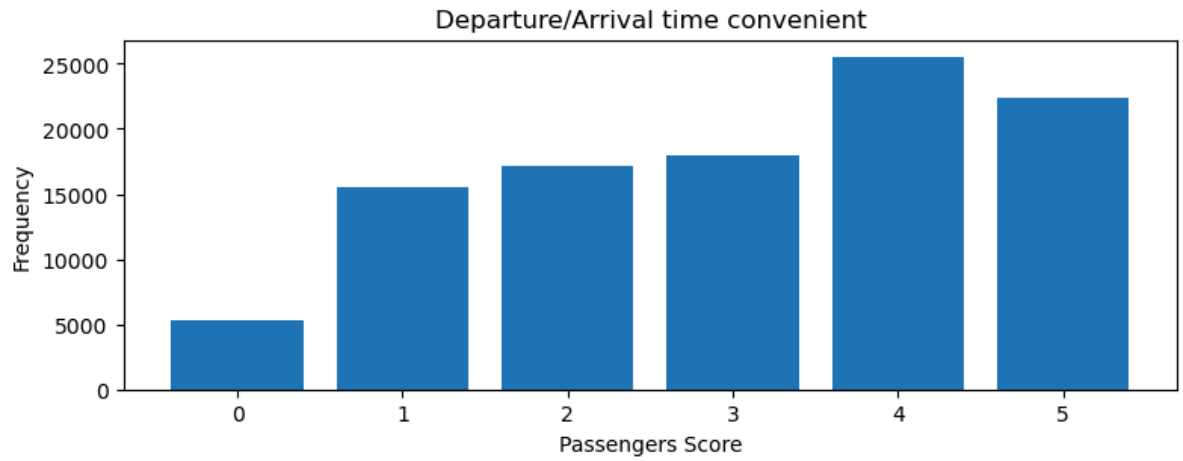
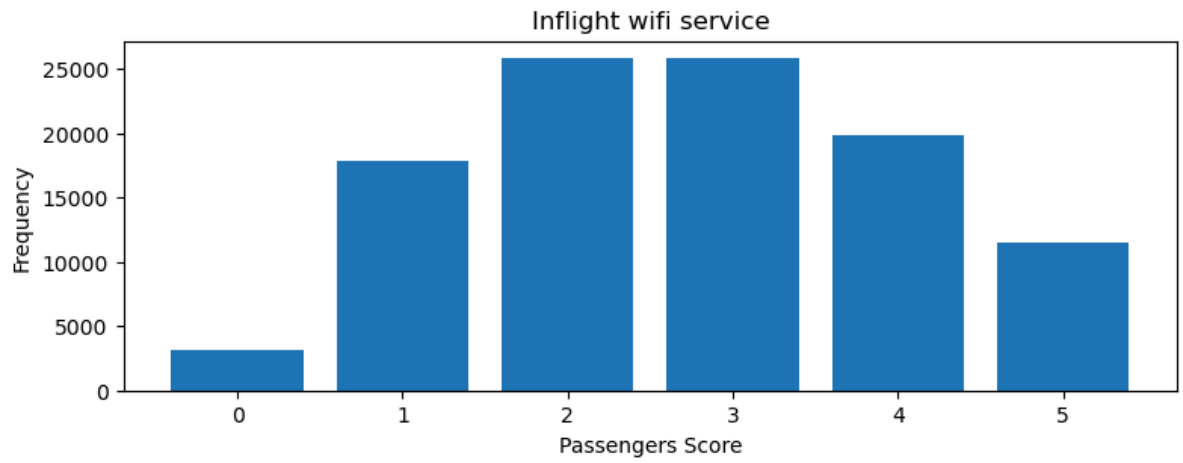
```
In [157... df_air.columns.to_list()
```

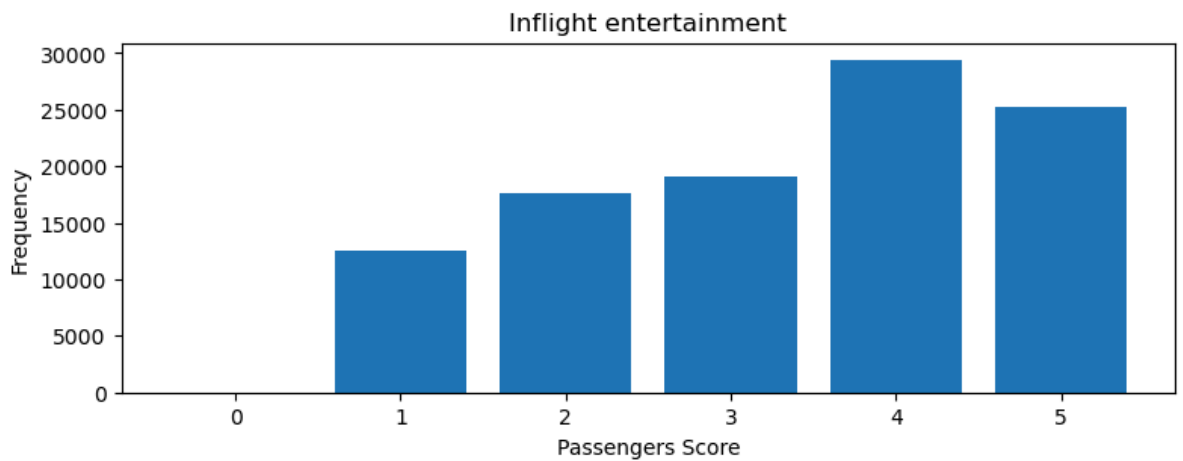
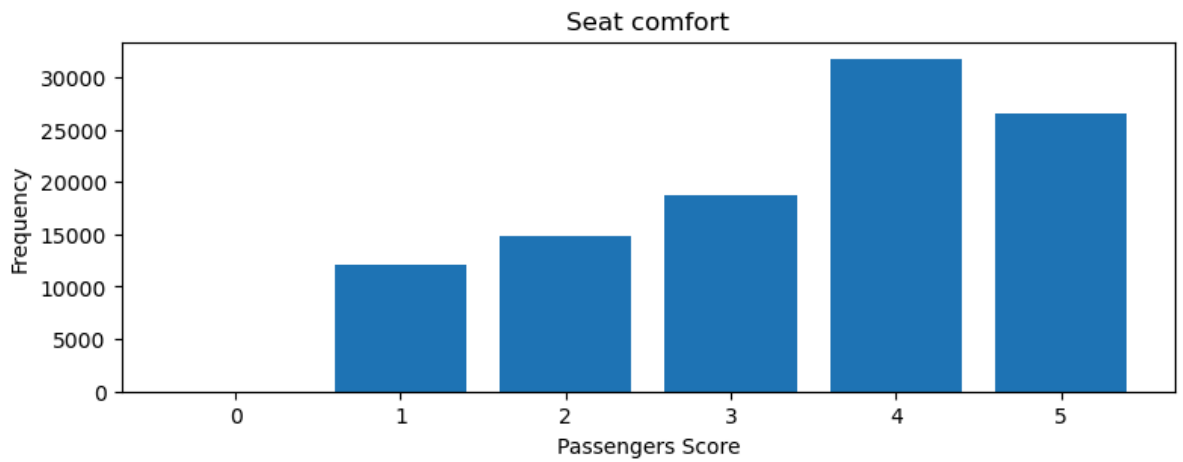
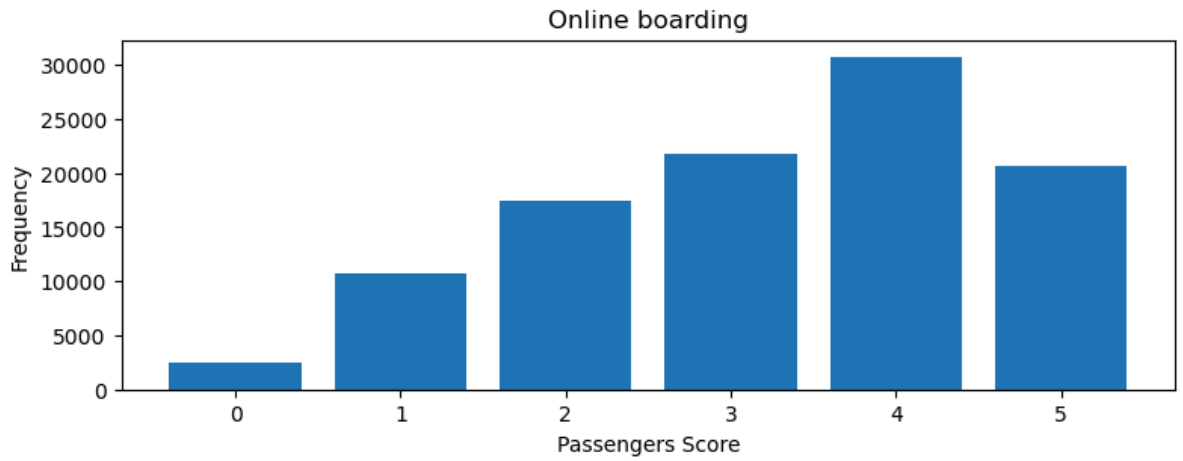
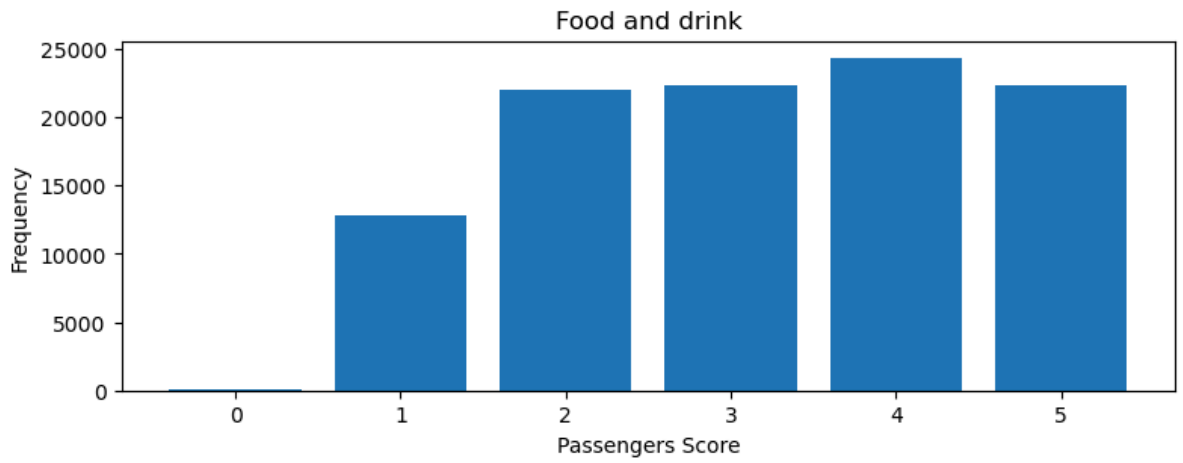
```
Out[157]: ['Unnamed: 0',
'Gender',
'Customer Type',
'Age',
'Type of Travel',
'Class',
'Flight Distance',
'Inflight wifi service',
'Departure/Arrival time convenient',
'Ease of Online booking',
'Gate location',
'Food and drink',
'Online boarding',
'Seat comfort',
'Inflight entertainment',
'On-board service',
'Leg room service',
'Baggage handling',
'Checkin service',
'Inflight service',
'Cleanliness',
'Departure Delay in Minutes',
'Arrival Delay in Minutes',
'satisfaction']
```

```
In [ ]:
```

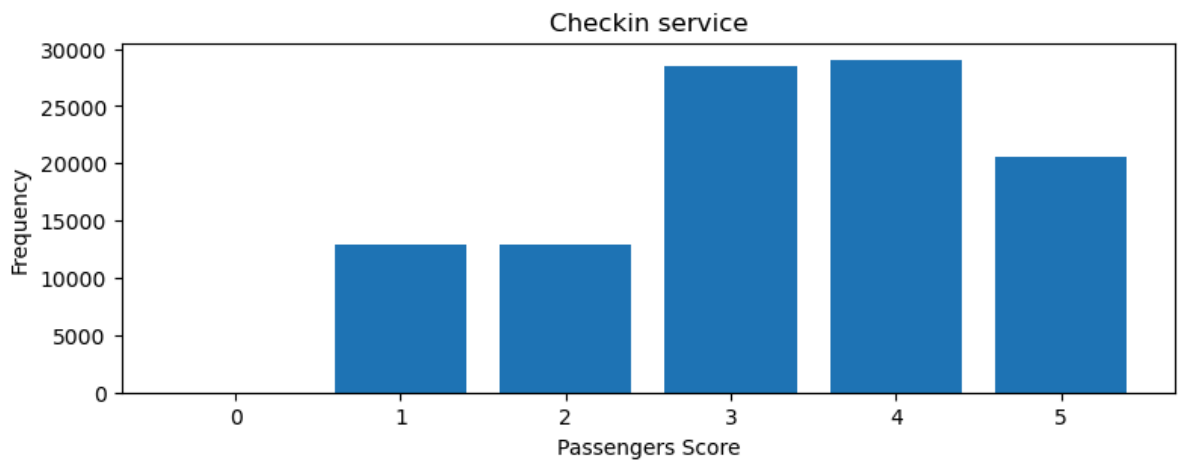
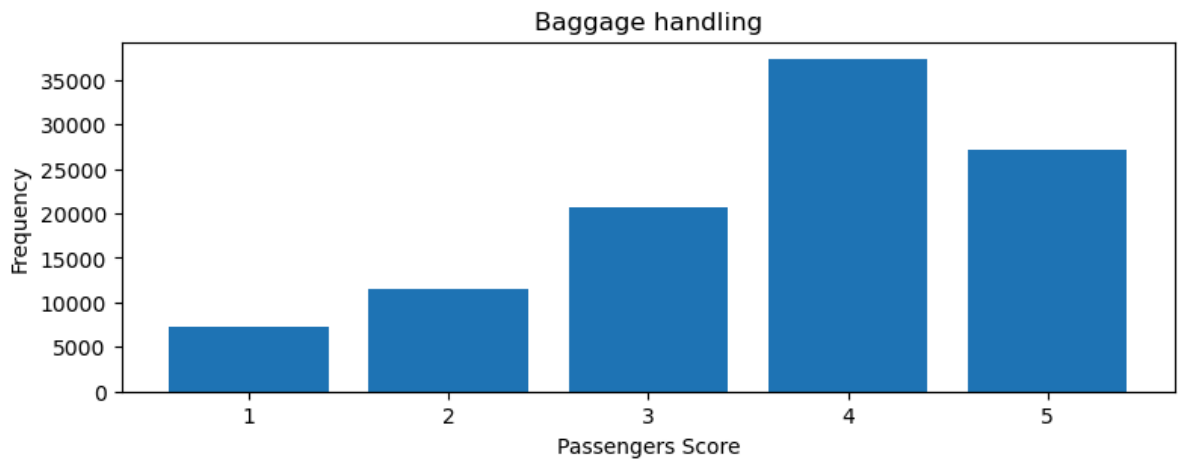
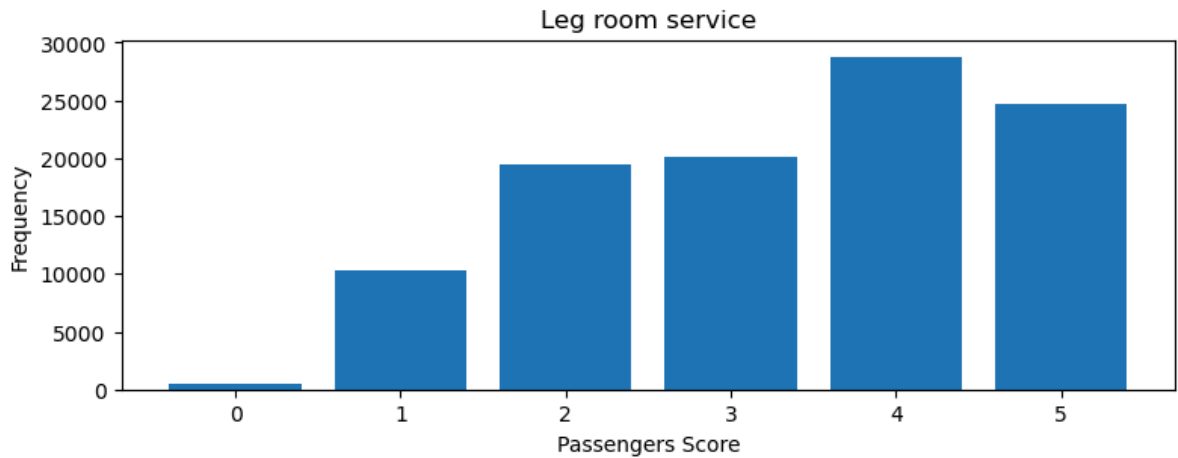
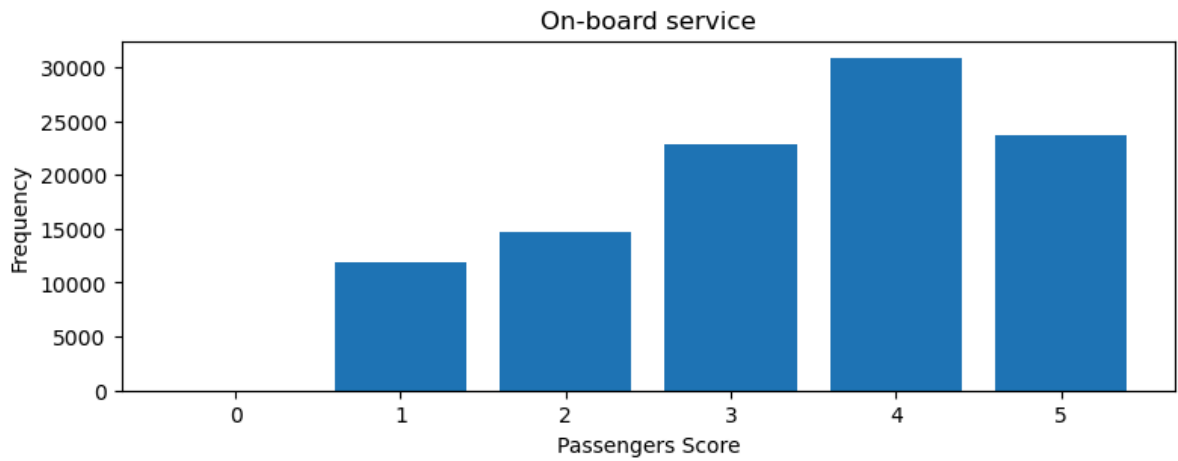
```
In [160... def bar_plot(variable):
    var = df_air[variable]
    var_Value = var.value_counts()
    plt.figure(figsize = (9, 3))
    plt.bar(var_Value.index, var_Value.values)
    plt.xlabel("Passengers Score")
    plt.ylabel("Frequency")
    plt.title(variable)
    plt.show()
```

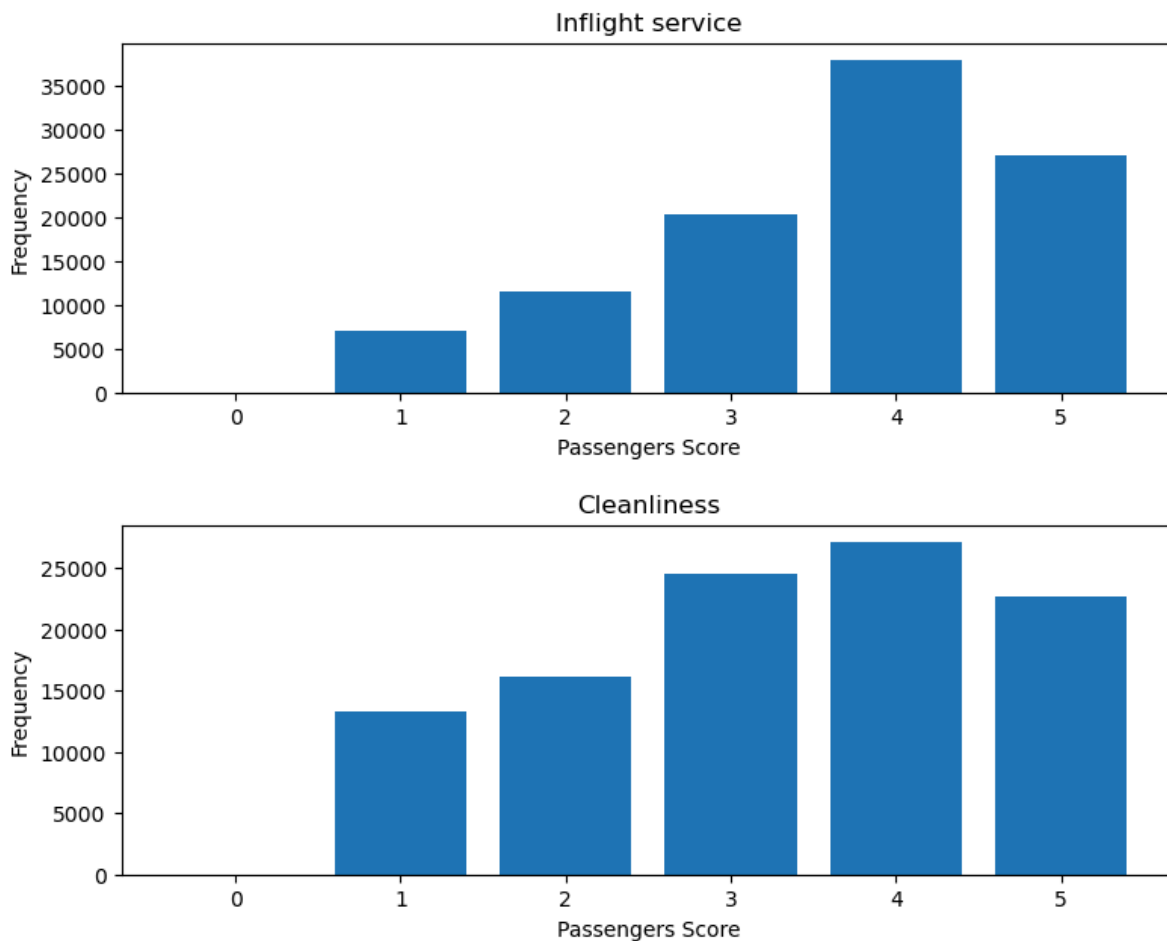
```
In [162... category1 = ['Inflight wifi service', 'Departure/Arrival time convenient', 'Ease
                'Gate location', 'Food and drink', 'Online boarding', 'Seat comfort
                'On-board service', 'Leg room service', 'Baggage handling', 'Checki
for c in category1:
    bar_plot(c)
```





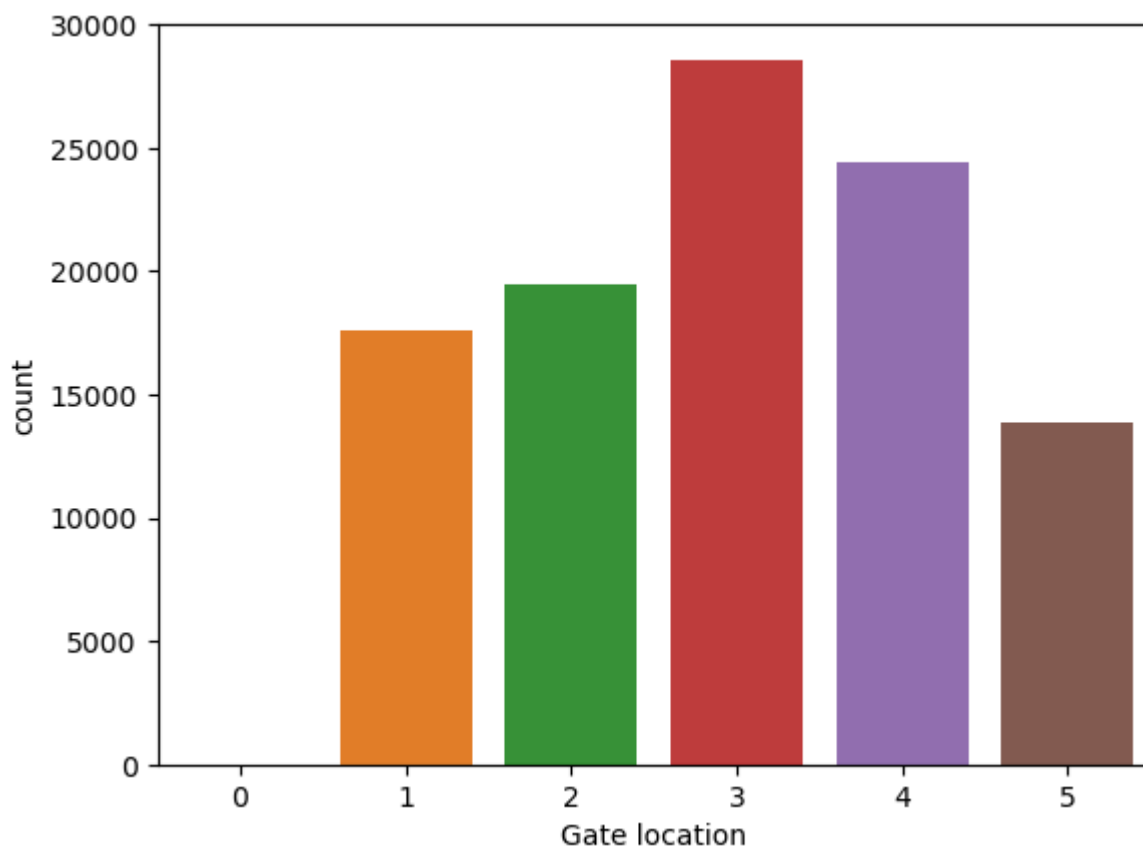






```
In [226...] sns.countplot(data = df_air, x = "Gate location")#, color = "blue")
```

```
Out[226]: <Axes: xlabel='Gate location', ylabel='count'>
```



```
In [230...] category = ['Gender', 'Customer Type', 'Type of Travel', 'Class', 'satisfaction']  
for c in category:
```

```
print(df_air[c].value_counts())
print(".....")
print()
```

```
Female      52727
Male        51177
Name: Gender, dtype: int64
.....

Loyal Customer      84923
disloyal Customer   18981
Name: Customer Type, dtype: int64
.....

Business travel      71655
Personal Travel      32249
Name: Type of Travel, dtype: int64
.....

Business      49665
Eco            46745
Eco Plus       7494
Name: Class, dtype: int64
.....

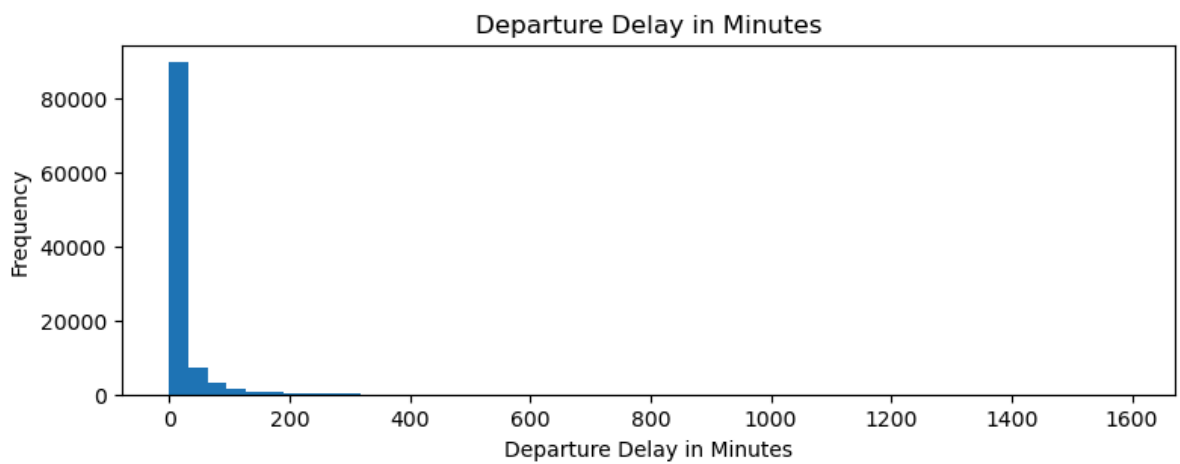
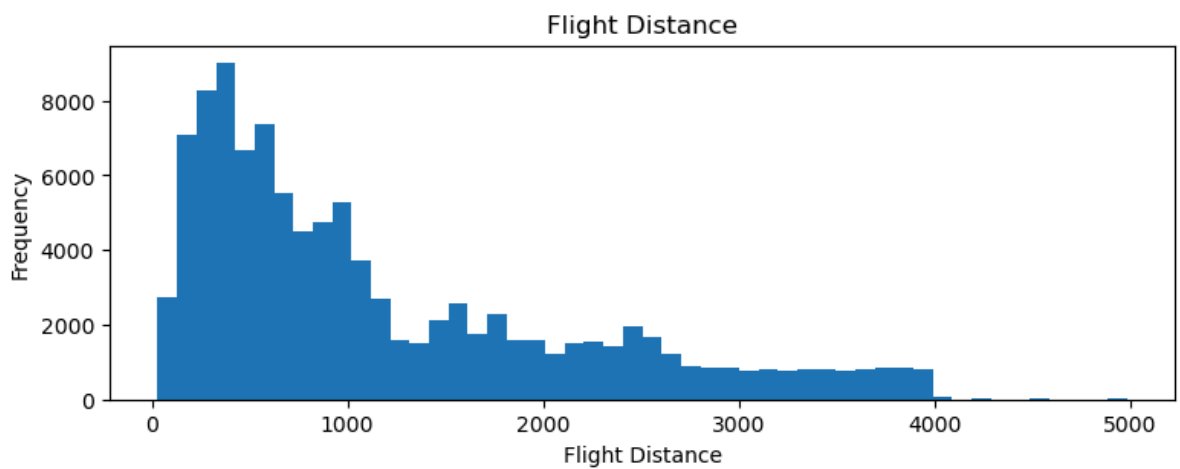
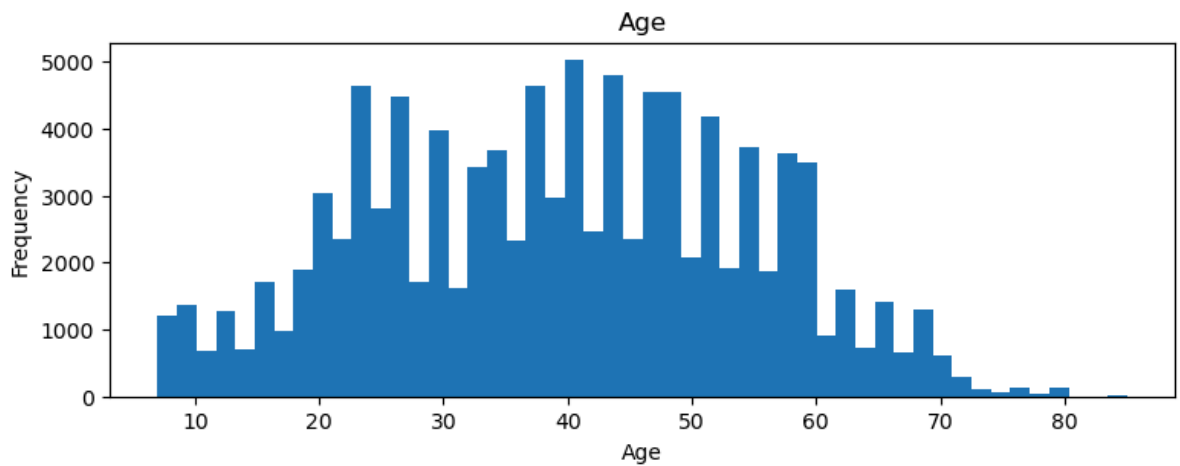
0      58879
1      45025
Name: satisfaction, dtype: int64
.....
```

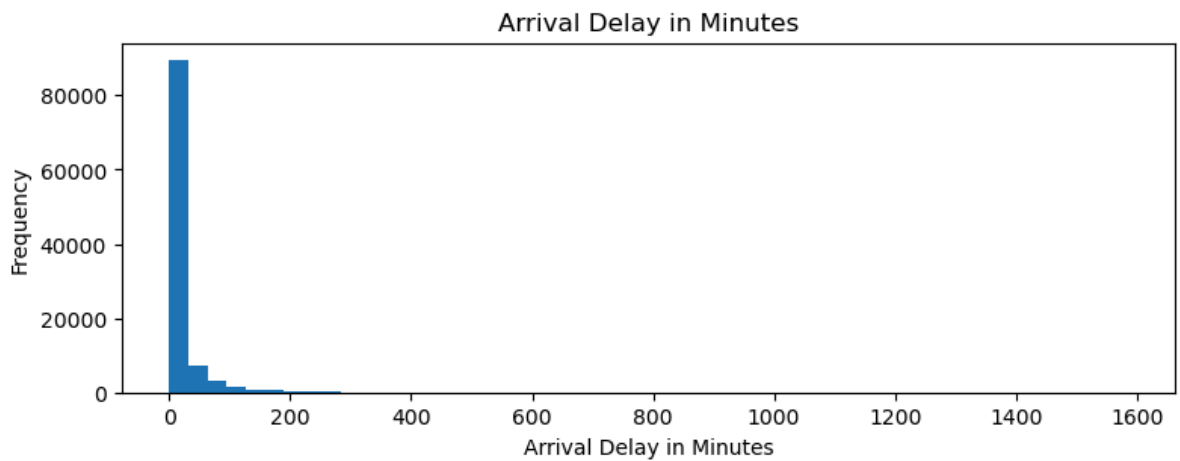
In [172... df\_air.columns.to\_list()

```
Out[172]: ['Unnamed: 0',
'Gender',
'Customer Type',
'Age',
'Type of Travel',
'Class',
'Flight Distance',
'Inflight wifi service',
'Departure/Arrival time convenient',
'Ease of Online booking',
'Gate location',
'Food and drink',
'Online boarding',
'Seat comfort',
'Inflight entertainment',
'On-board service',
'Leg room service',
'Baggage handling',
'Checkin service',
'Inflight service',
'Cleanliness',
'Departure Delay in Minutes',
'Arrival Delay in Minutes',
'satisfaction']
```

```
In [174... def plot_hist(variable):
plt.figure(figsize = (9, 3))
plt.hist(df_air[variable], bins = 50)
plt.xlabel(variable)
plt.ylabel("Frequency")
plt.title(vaAriable)
plt.show()
```

```
In [176... numericVar = ["Age", "Flight Distance", "Departure Delay in Minutes", "Arriv  
for n in numericVar:  
    plot_hist(n)
```





```
In [178... df_air.head()
```

```
Out[178]:
```

	Unnamed: 0	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arr time conven
0	0	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	
1	1	Male	disloyal Customer	25	Business travel	Business	235	3	
2	2	Female	Loyal Customer	26	Business travel	Business	1142	2	
3	3	Female	Loyal Customer	25	Business travel	Business	562	2	
4	4	Male	Loyal Customer	61	Business travel	Business	214	3	

5 rows × 24 columns

```
In [180... df_air["satisfaction"].replace({"satisfied": 1, "neutral or dissatisfied": 0})
df_air.head()
```

```
Out[180]:
```

	Unnamed: 0	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arr time conven
0	0	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	
1	1	Male	disloyal Customer	25	Business travel	Business	235	3	
2	2	Female	Loyal Customer	26	Business travel	Business	1142	2	
3	3	Female	Loyal Customer	25	Business travel	Business	562	2	
4	4	Male	Loyal Customer	61	Business travel	Business	214	3	

5 rows × 24 columns

```
In [182... df_air[["Gender", "satisfaction"]].groupby(["Gender"], as_index = False).mea
```

Out[182]:

	Gender	satisfaction
1	Male	0.439475
0	Female	0.427371

In [184...]

```
df_sat = df_air[["Age", "satisfaction"]].groupby(["Age"], as_index = False)
df_sat.head()
```

Out[184]:

	Age	satisfaction
34	41	0.607899
39	46	0.599819
44	51	0.597249
50	57	0.590102
45	52	0.586207

In [186...]

```
df_sat[50:]
```

Out[186]:

	Age	satisfaction
12	19	0.245575
10	17	0.244919
64	71	0.244681
11	18	0.244376
57	64	0.214485
55	62	0.212337
9	16	0.204672
73	80	0.192308
59	66	0.188513
58	65	0.187332
70	77	0.183908
8	15	0.183374
74	85	0.176471
61	68	0.174383
5	12	0.166929
4	11	0.160767
60	67	0.156061
6	13	0.154818
62	69	0.145511
7	14	0.144272
2	9	0.137283
63	70	0.136143
3	10	0.131772
1	8	0.107813
0	7	0.099644

In [188...

```
df_air[["Cleanliness", "satisfaction"]].groupby(["Cleanliness"], as_index =
```

Out[188]:

	Cleanliness	satisfaction
5	5	0.608797
4	4	0.535340
3	3	0.431879
2	2	0.212993
1	1	0.196201
0	0	0.000000

In [190...

```
df_air[["Inflight wifi service", "satisfaction"]].groupby(["Inflight wifi se
```

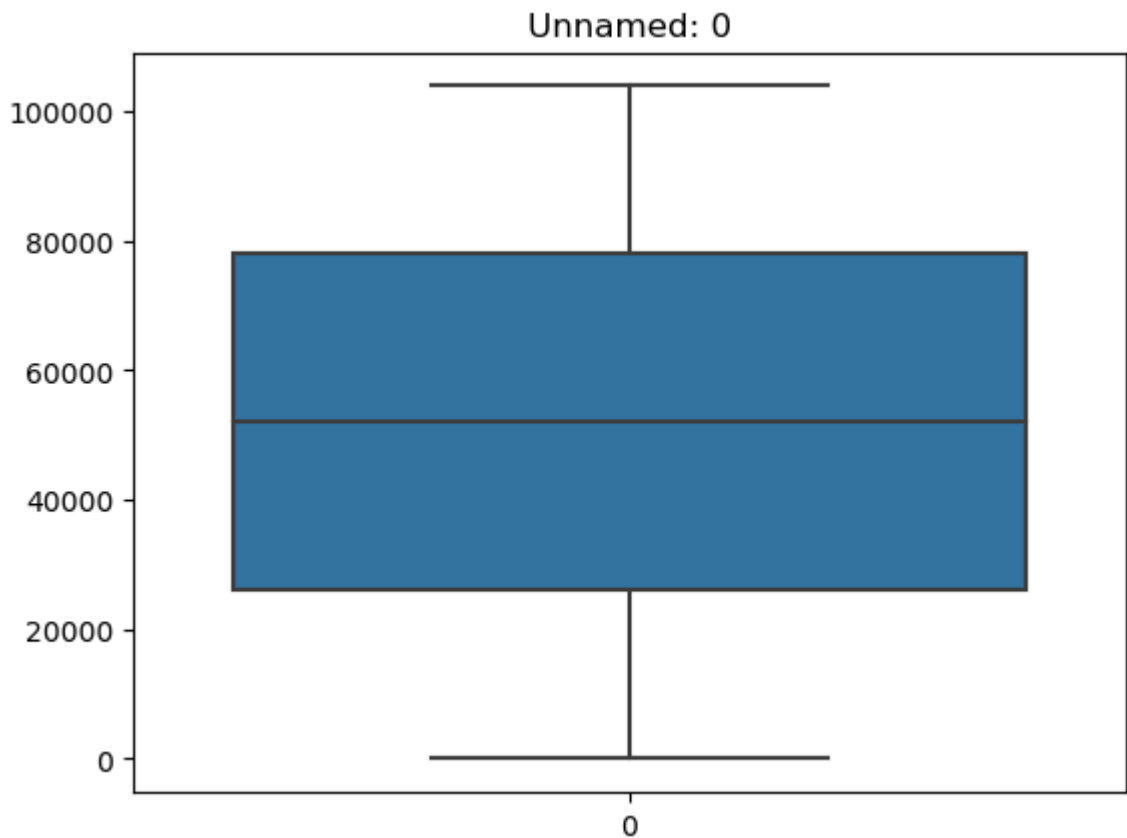
Out[190]:

	Inflight wifi service	satisfaction
0	0	0.997422
5	5	0.990758
4	4	0.598969
1	1	0.325448
3	3	0.250580
2	2	0.248664

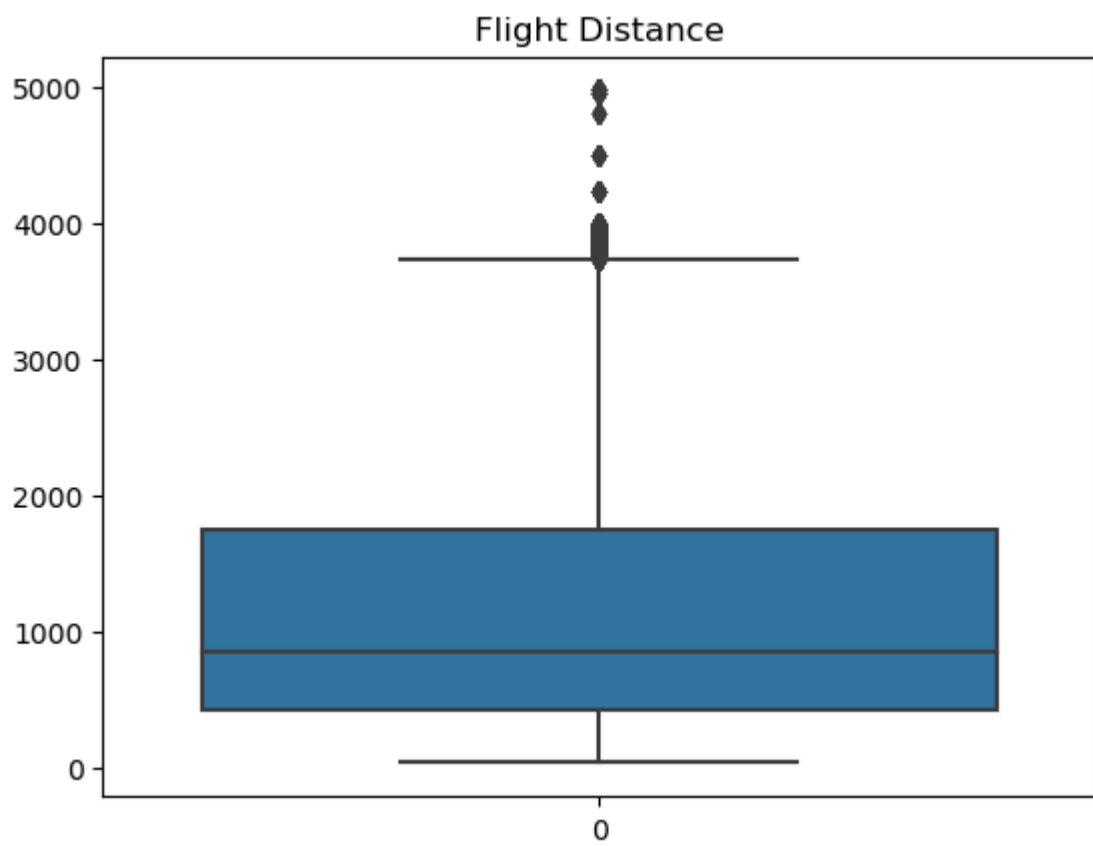
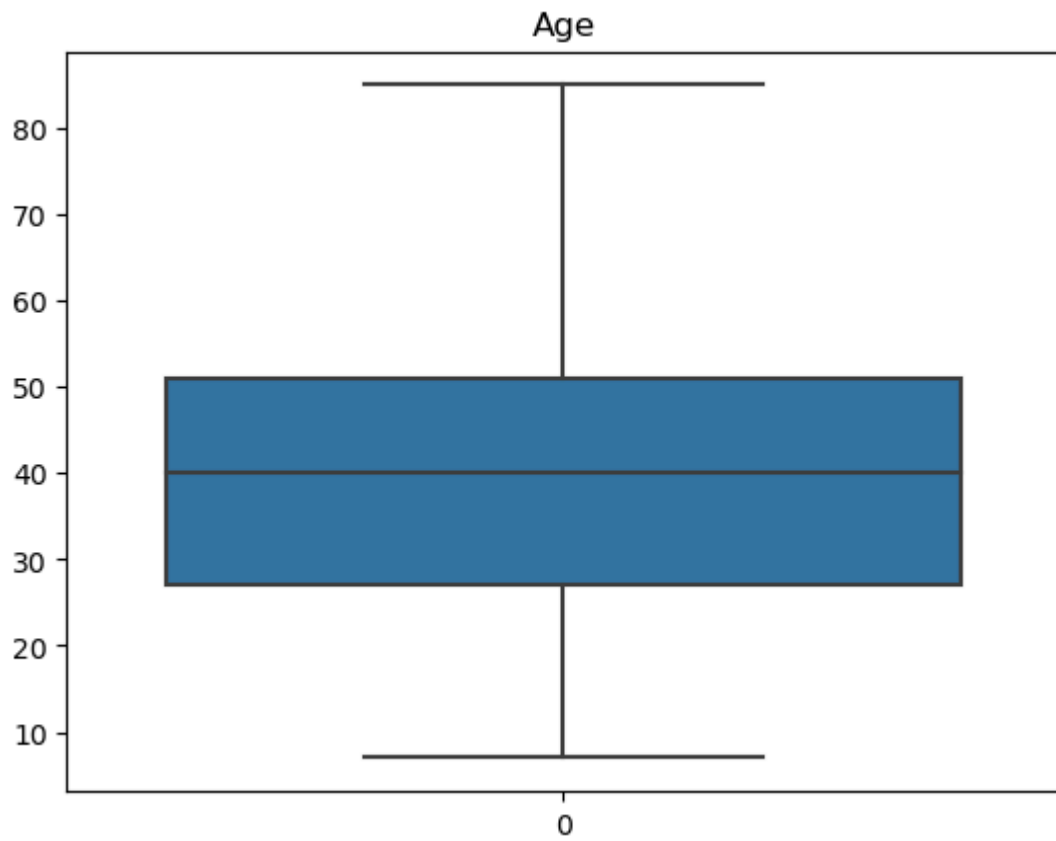
0	0	0.997422
5	5	0.990758
4	4	0.598969
1	1	0.325448
3	3	0.250580
2	2	0.248664

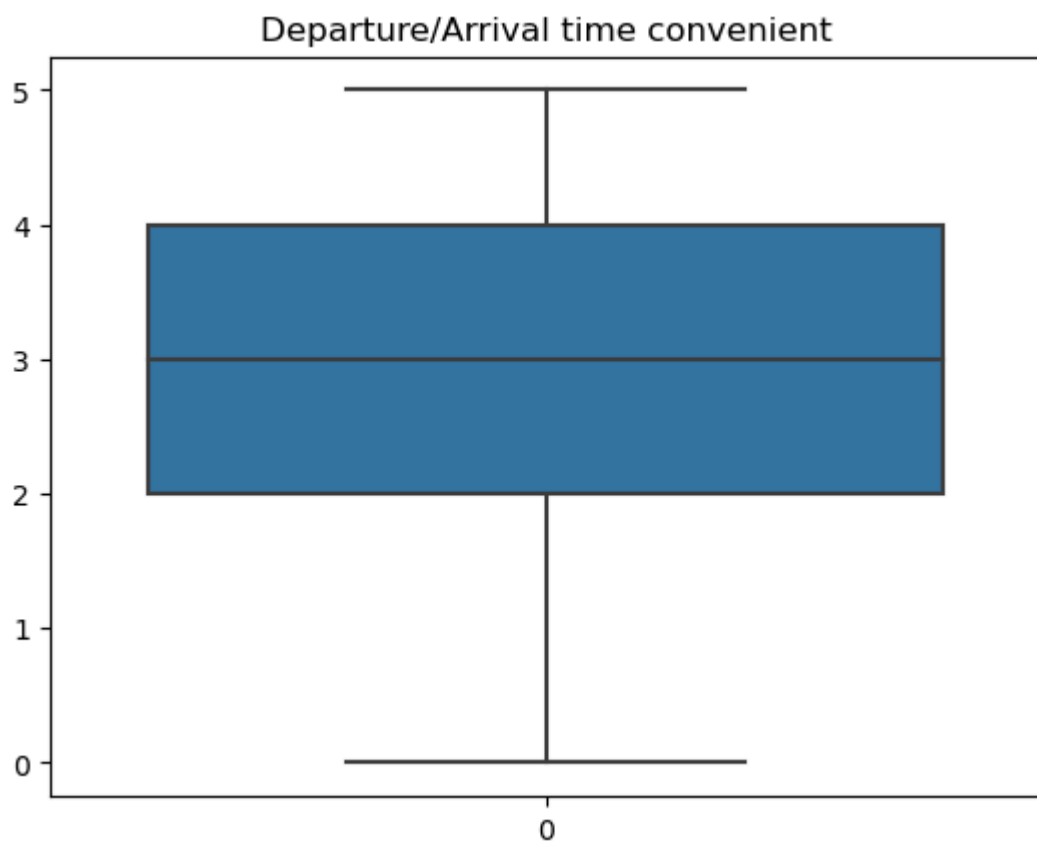
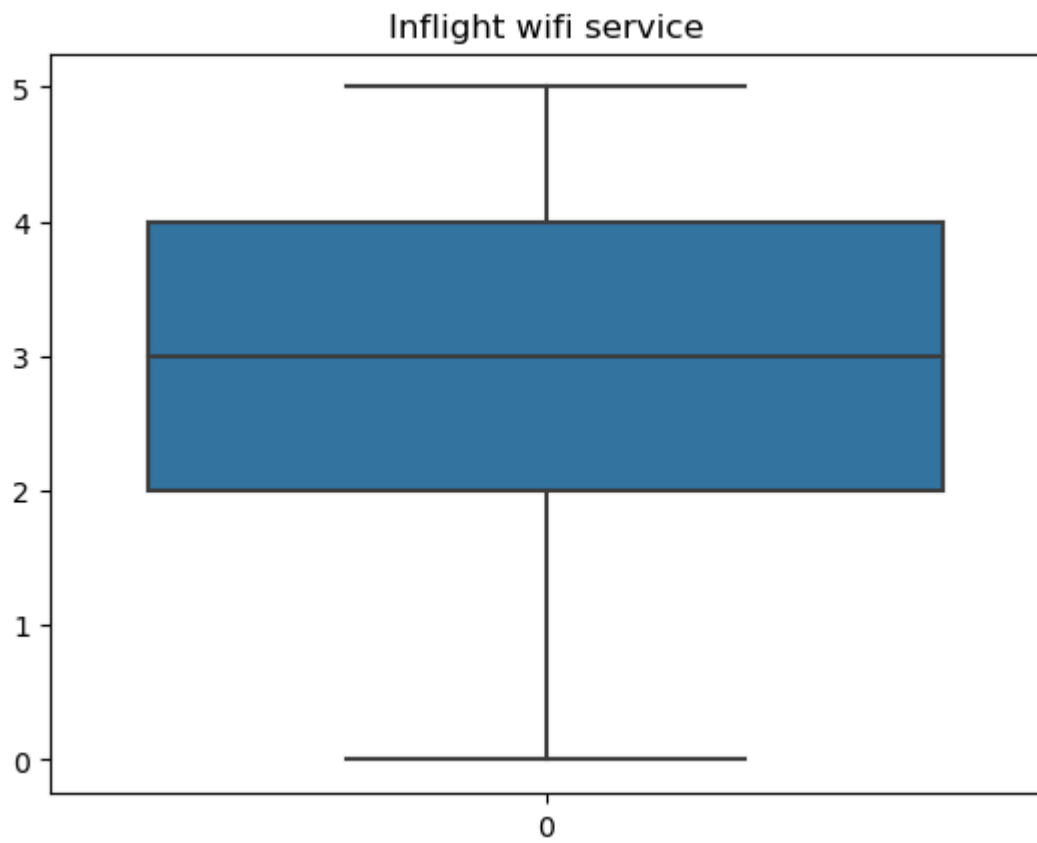
```
In [232... numeric_features = df_air.select_dtypes(include = ["int64", "float64"])
num_list = numeric_features.columns.to_list()
```

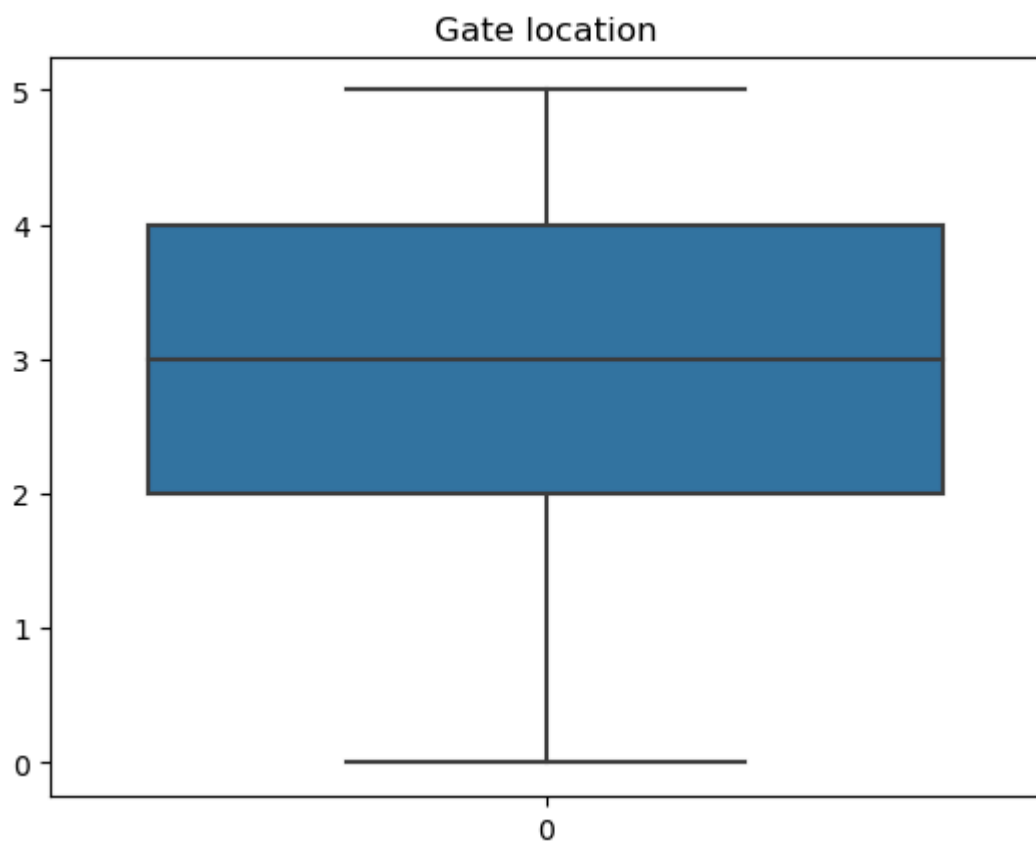
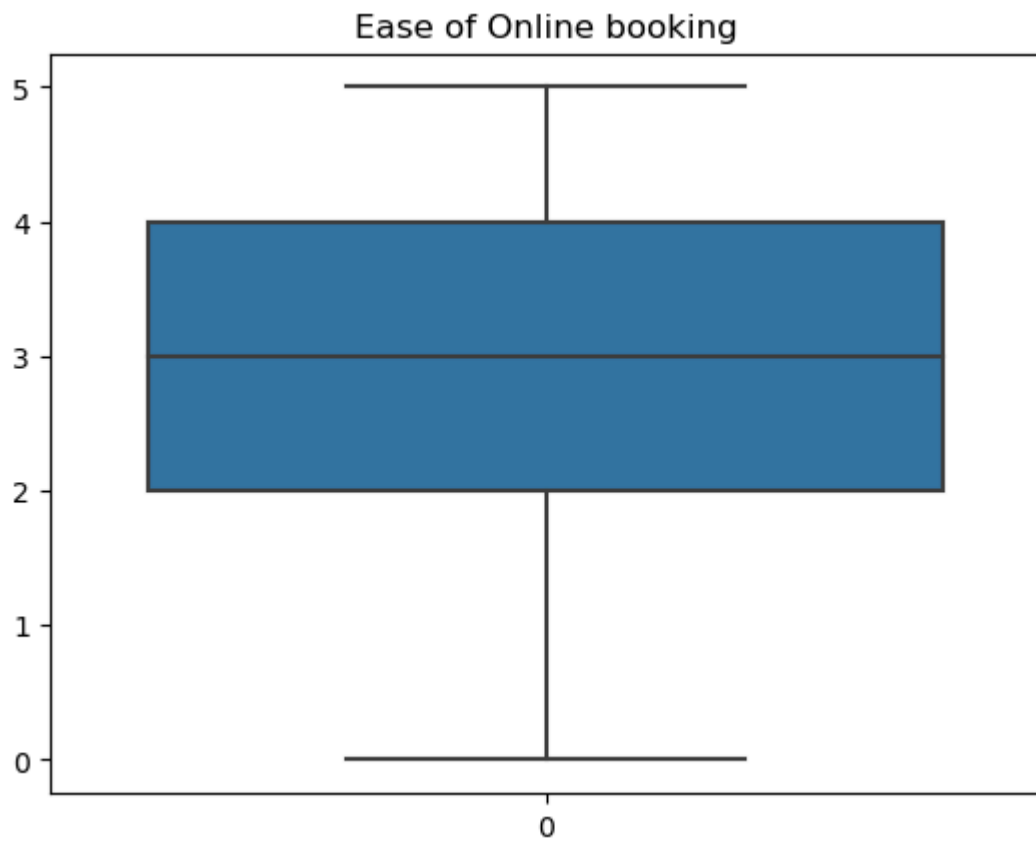
```
In [194... for cat in num_list:
    sns.boxplot(df_air[cat])
    plt.title(cat)
    plt.show()
```

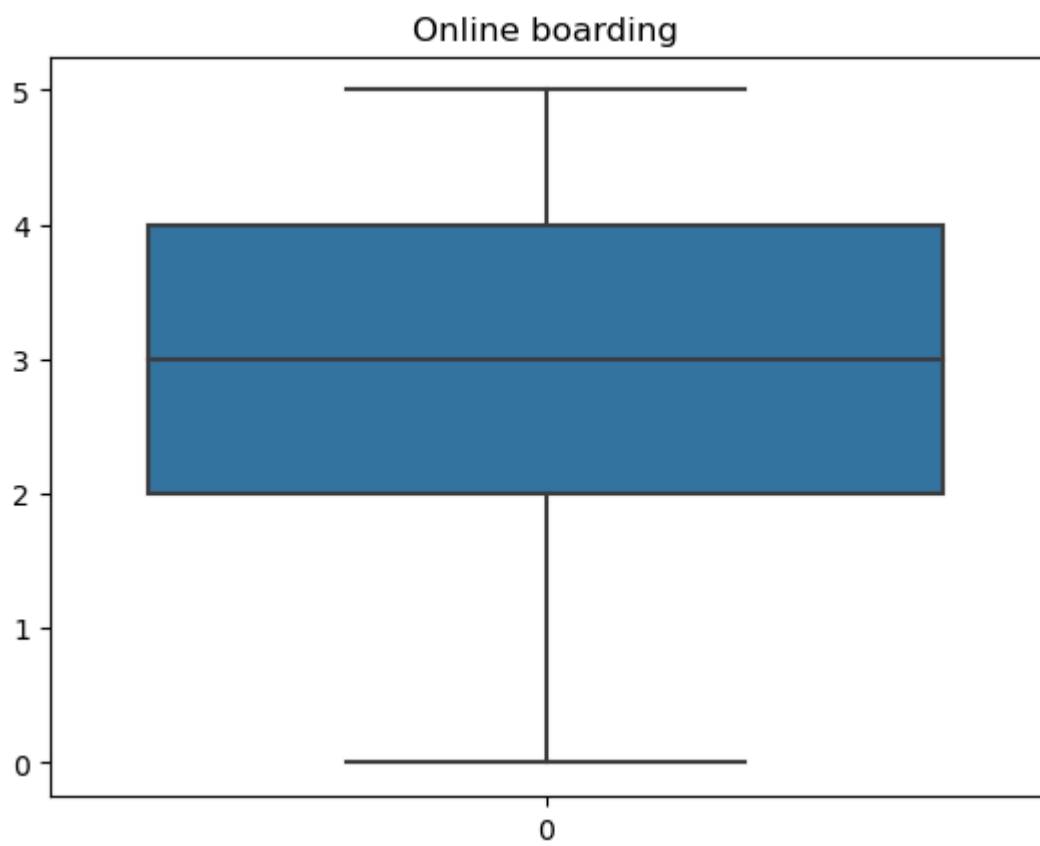
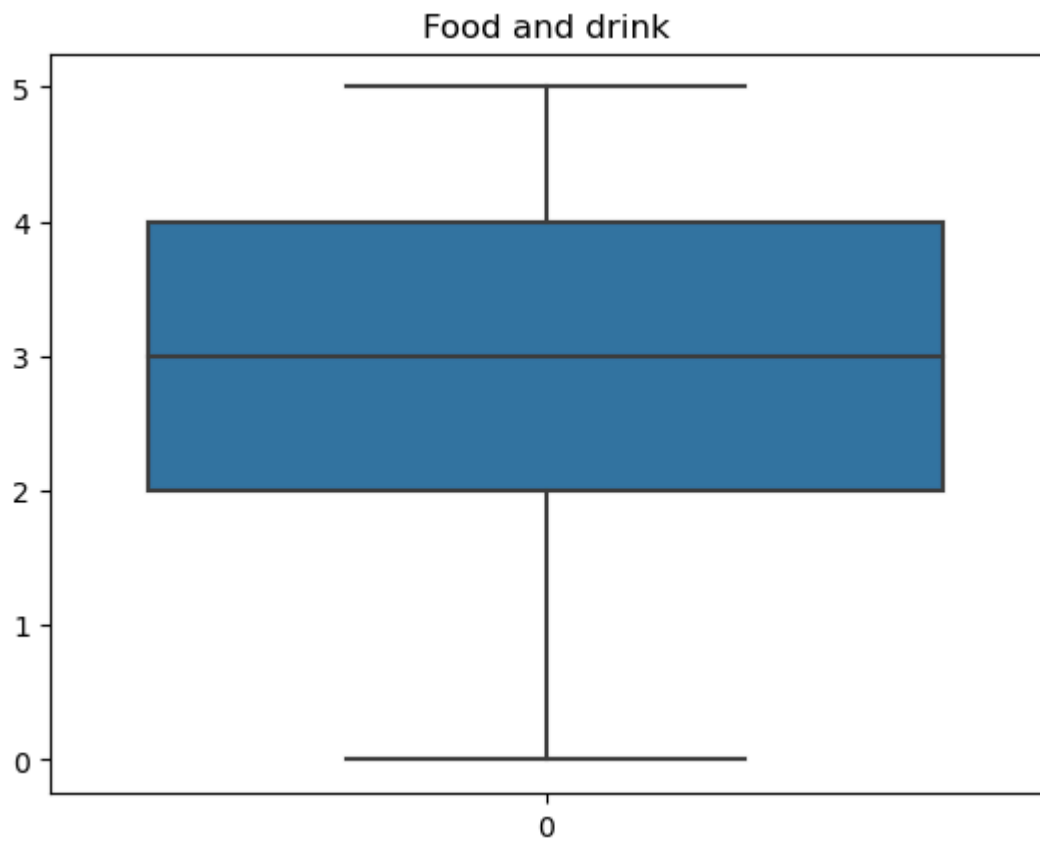


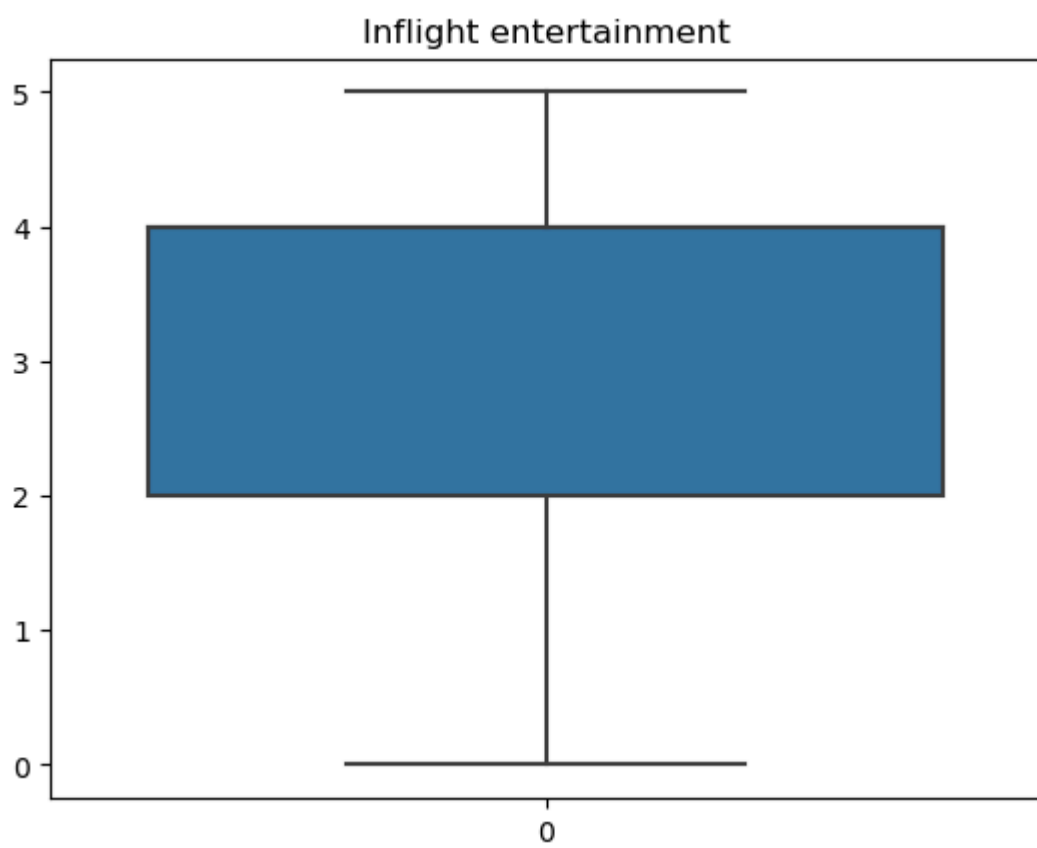
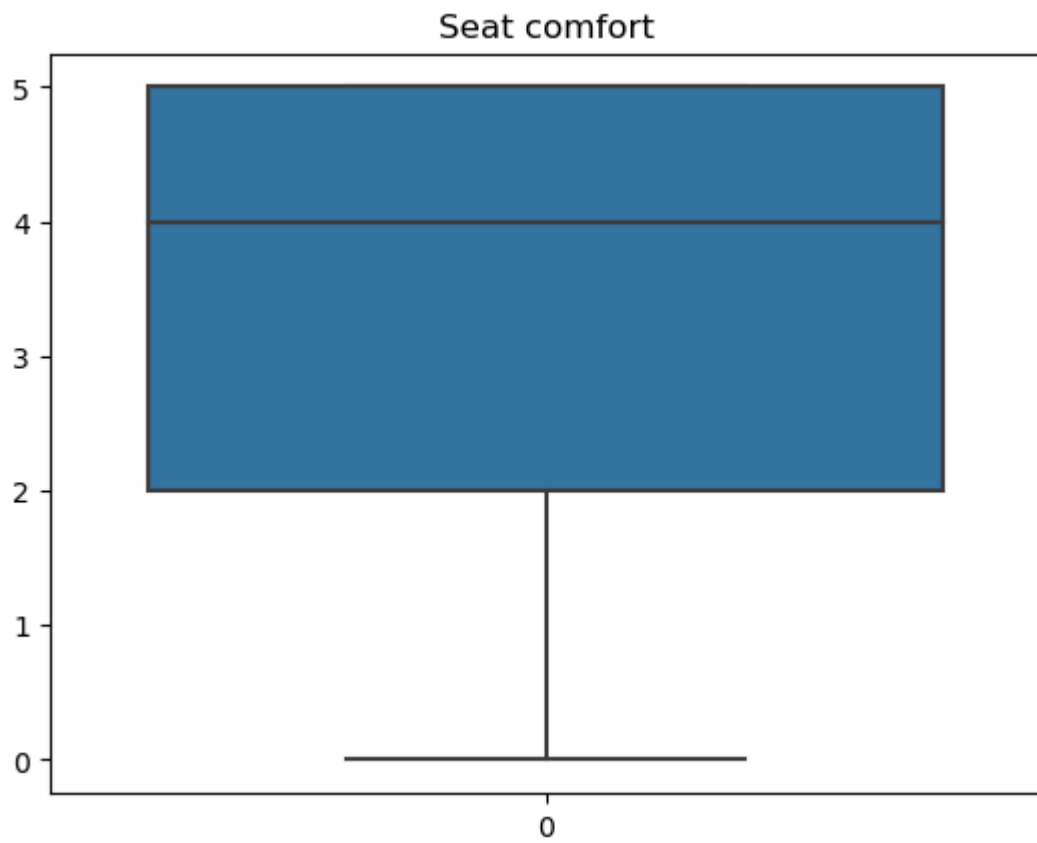


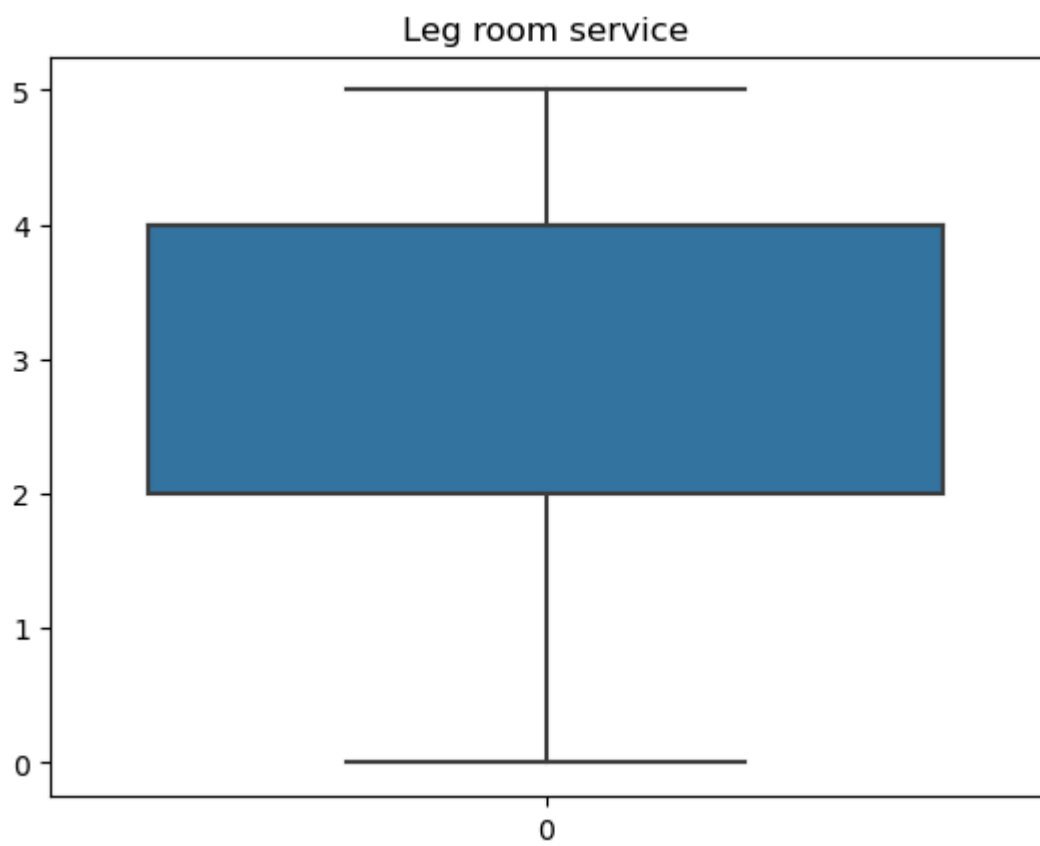
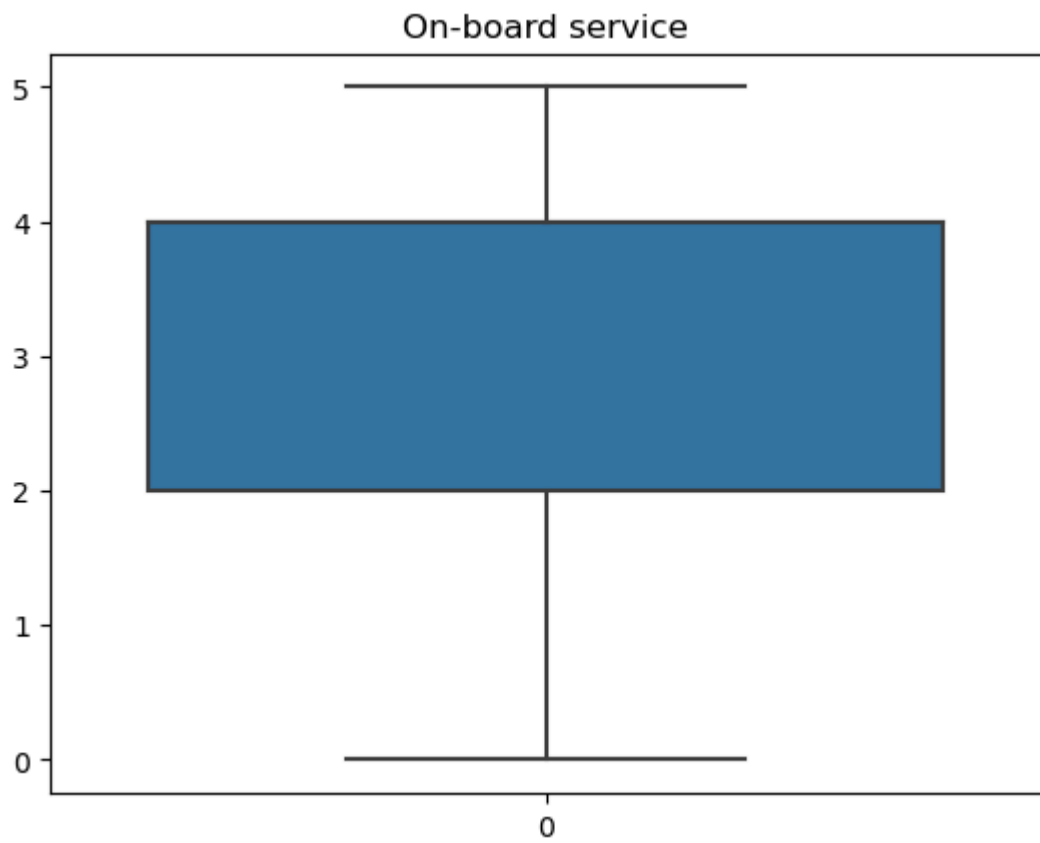


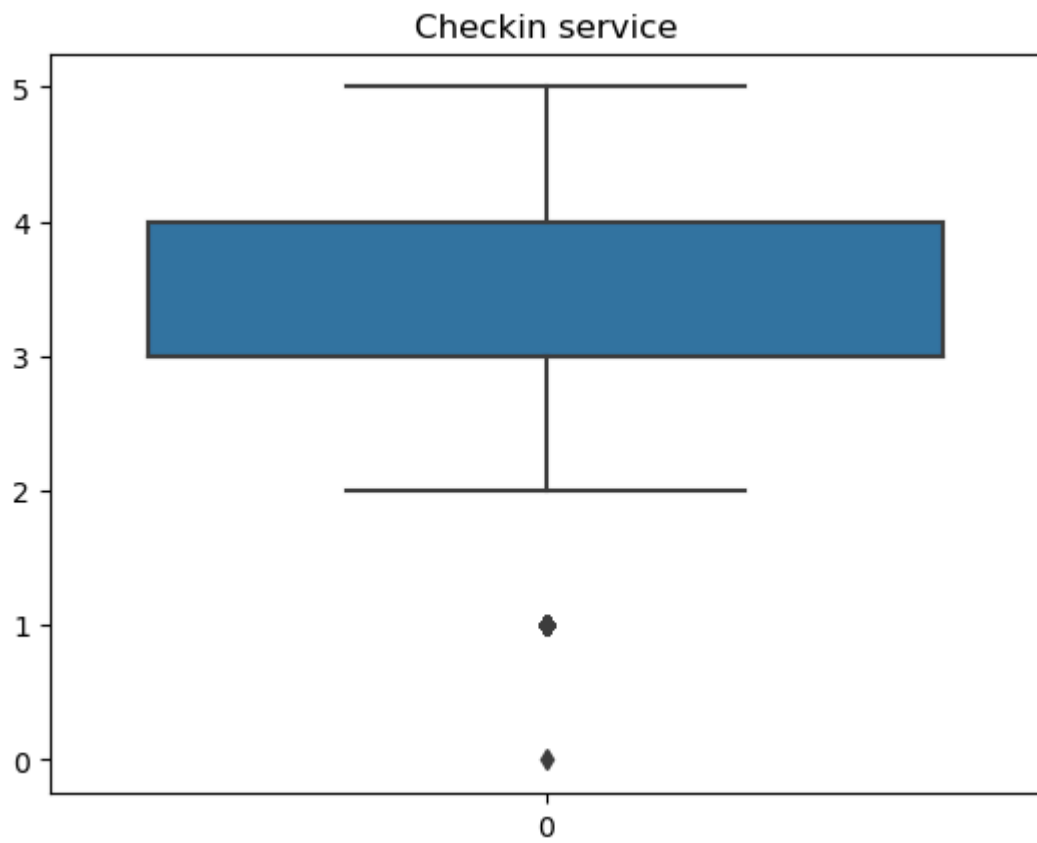
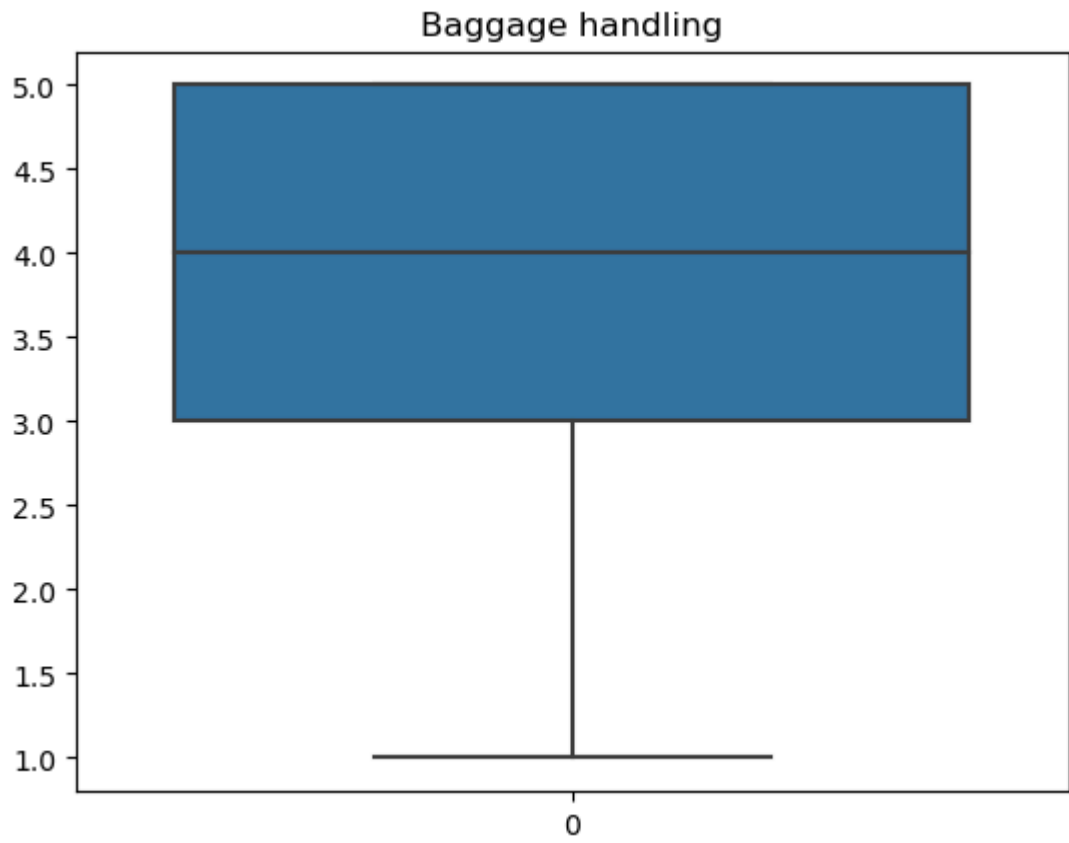


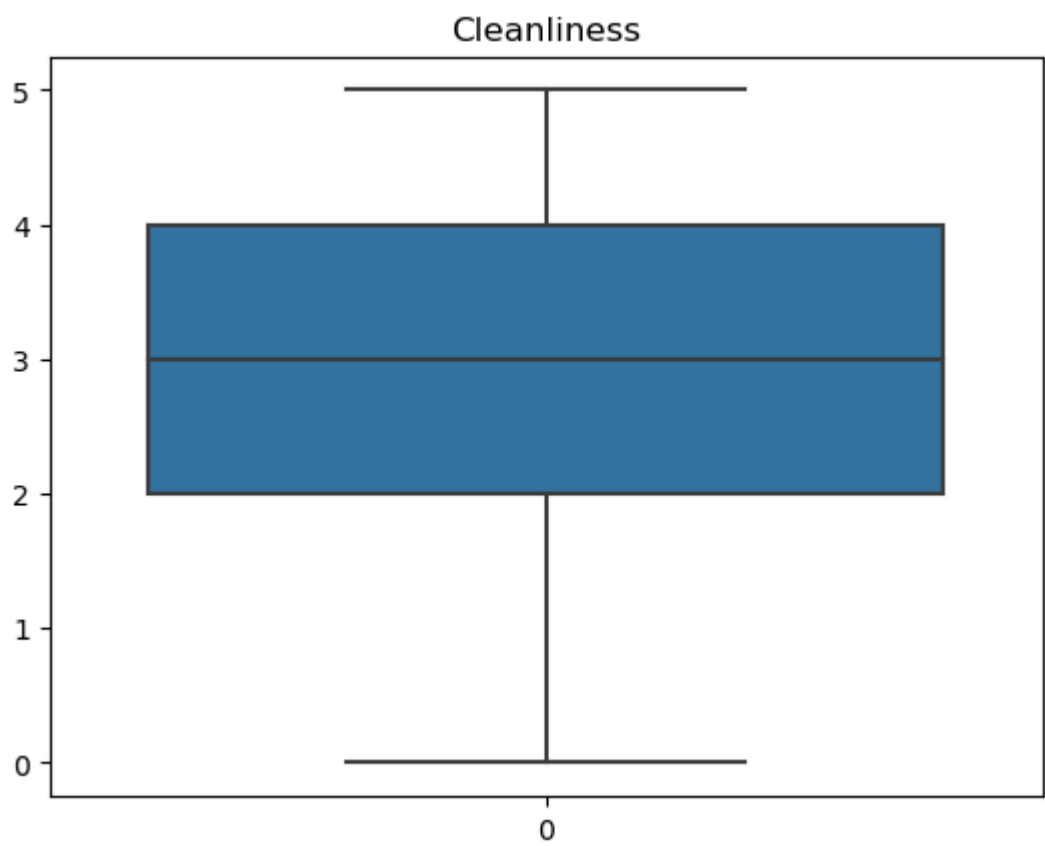
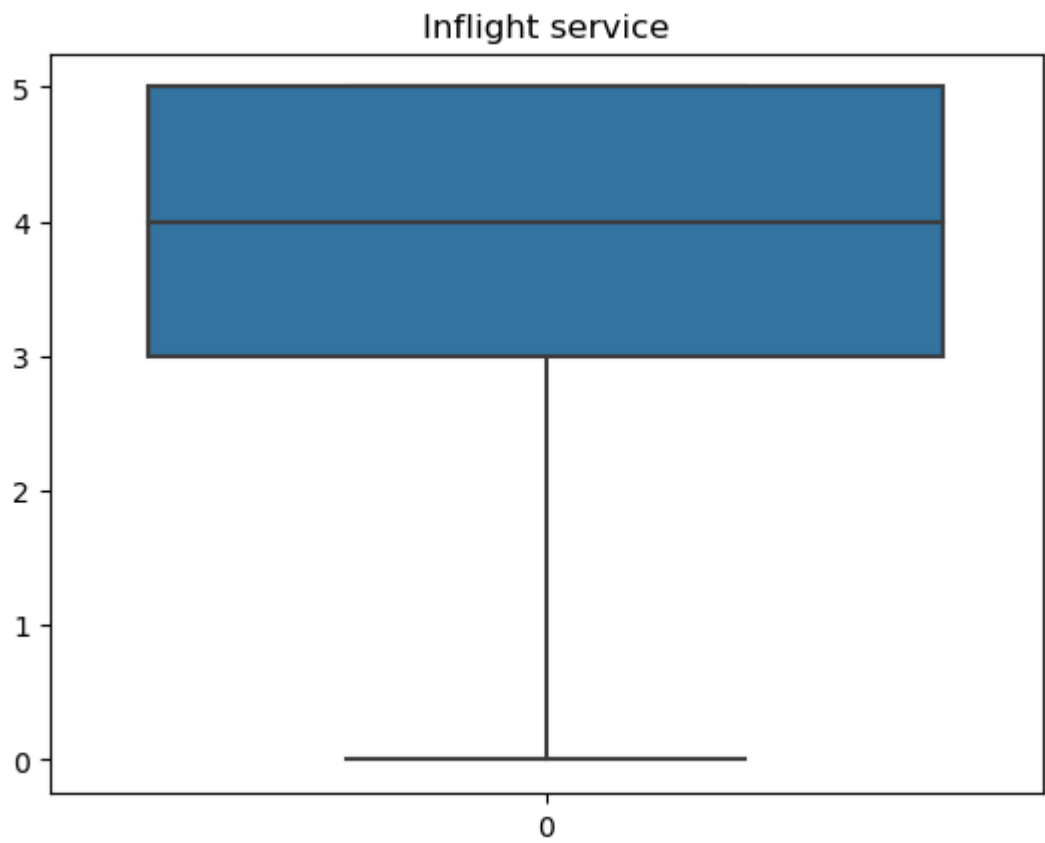




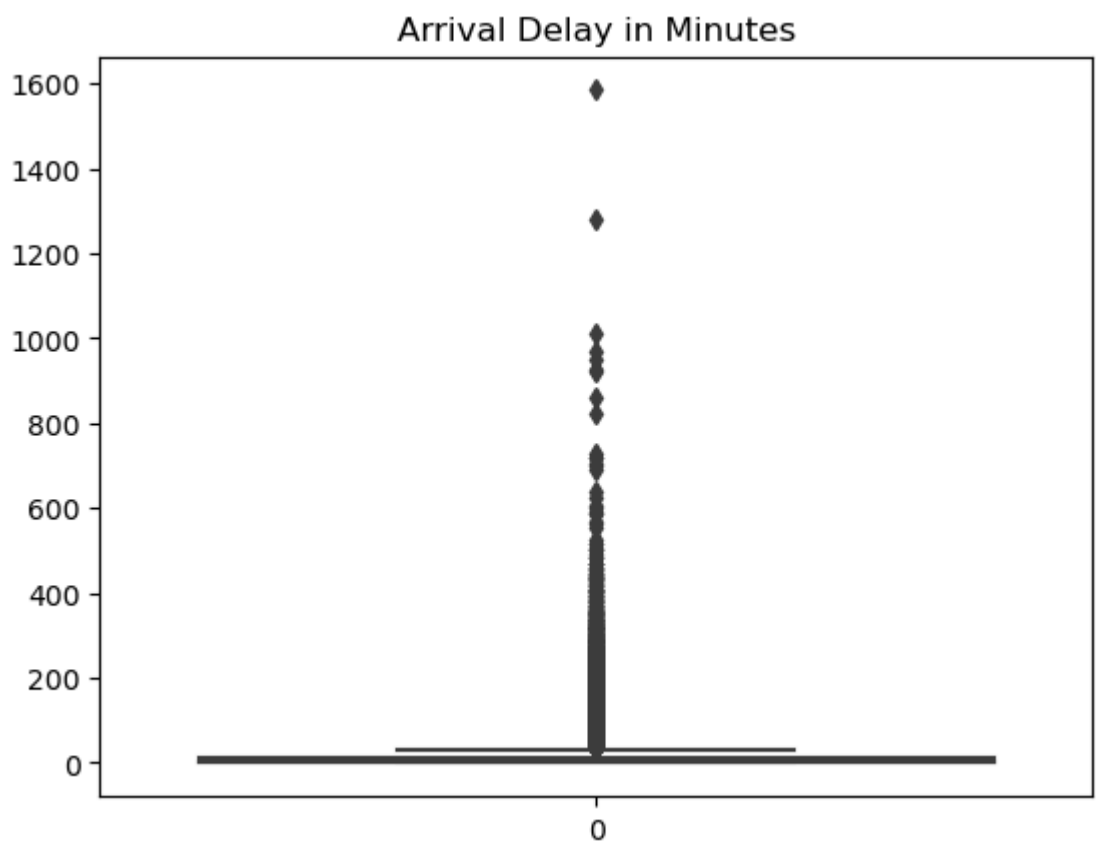
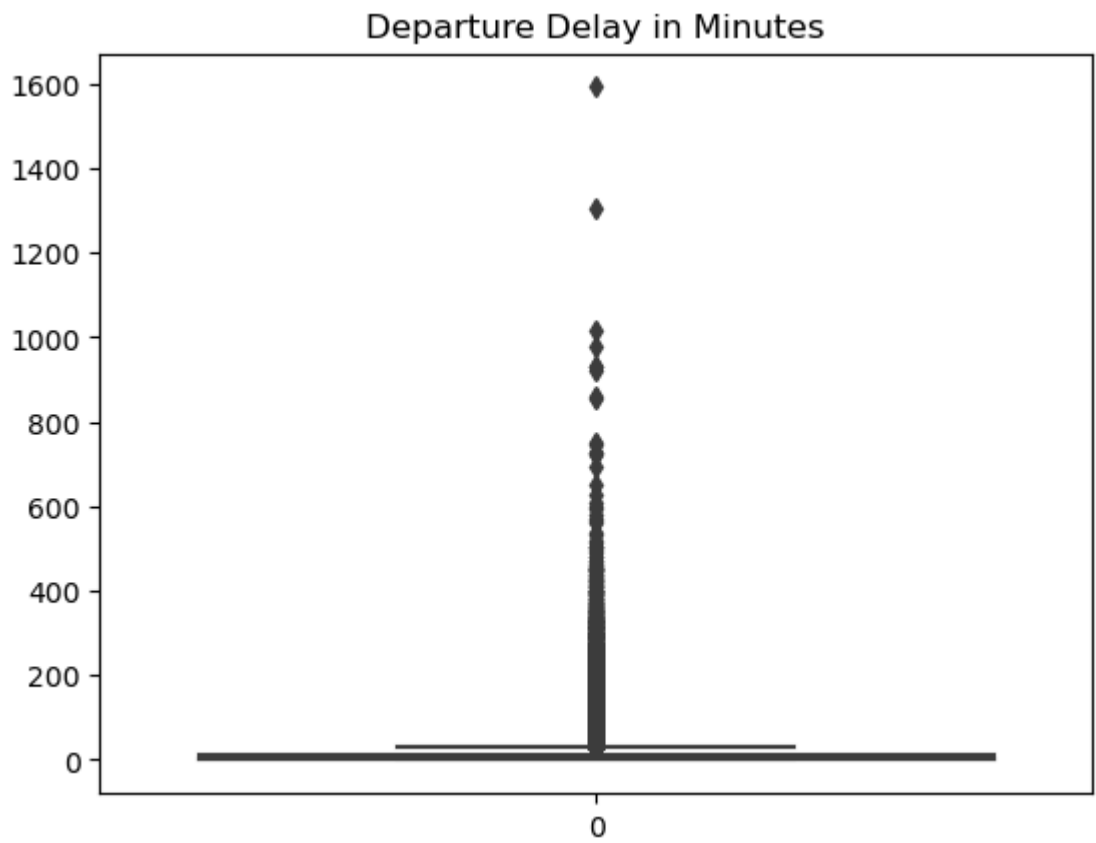


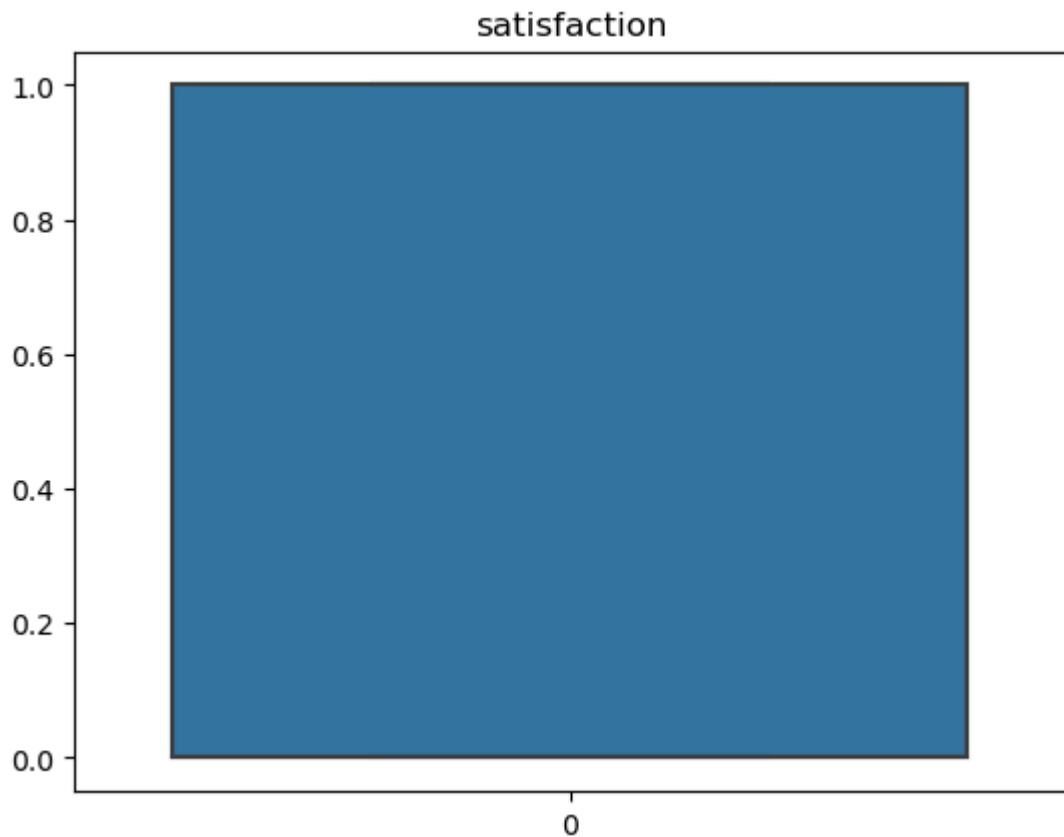








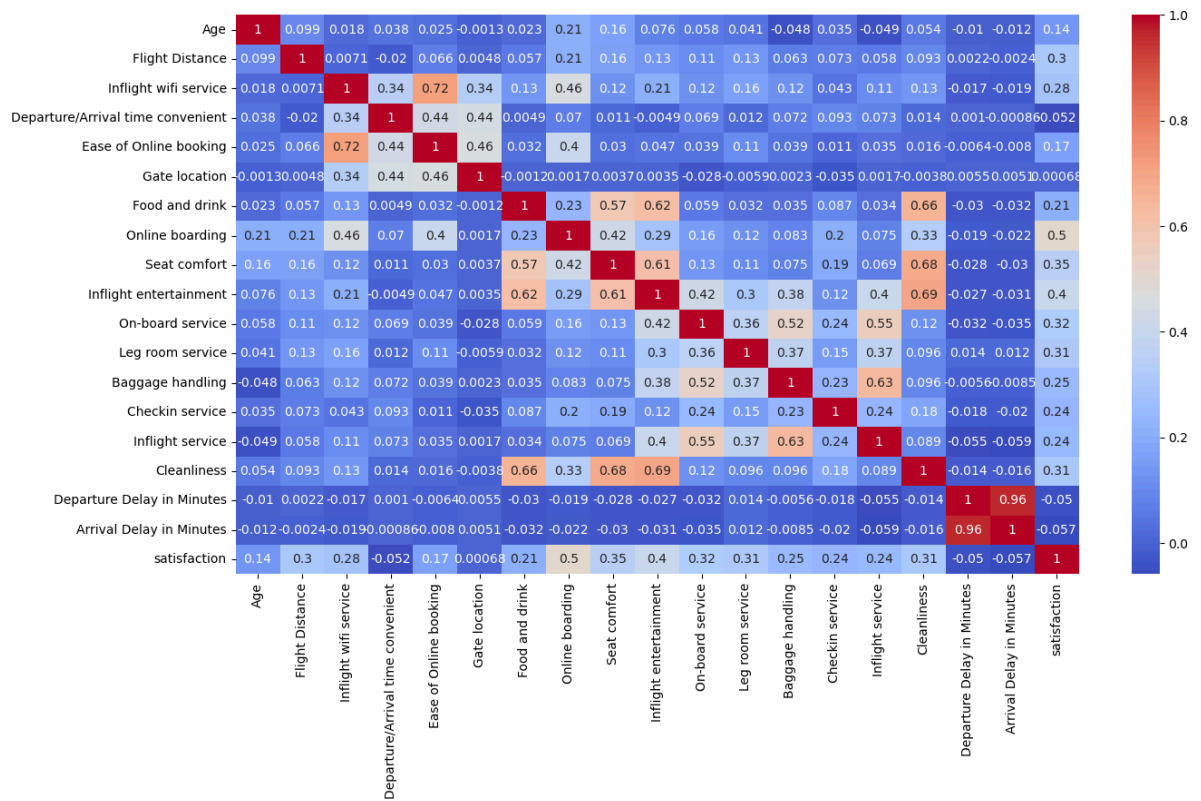




```
In [195]: df_air.columns.to_list()
```

```
Out[195]: ['Unnamed: 0',
           'Gender',
           'Customer Type',
           'Age',
           'Type of Travel',
           'Class',
           'Flight Distance',
           'Inflight wifi service',
           'Departure/Arrival time convenient',
           'Ease of Online booking',
           'Gate location',
           'Food and drink',
           'Online boarding',
           'Seat comfort',
           'Inflight entertainment',
           'On-board service',
           'Leg room service',
           'Baggage handling',
           'Checkin service',
           'Inflight service',
           'Cleanliness',
           'Departure Delay in Minutes',
           'Arrival Delay in Minutes',
           'satisfaction']
```

```
In [196]: corr = df_air[['Age', 'Flight Distance', 'Inflight wifi service',
                          'Departure/Arrival time convenient', 'Ease of Online booking',
                          'Gate location', 'Food and drink', 'Online boarding', 'Seat comfort',
                          'Inflight entertainment', 'On-board service', 'Leg room service',
                          'Baggage handling', 'Checkin service', 'Inflight service', 'Cleanliness',
                          'Departure Delay in Minutes', 'Arrival Delay in Minutes', 'satisfaction']]
plt.figure(figsize = (15, 8))
sns.heatmap(corr, annot = True, cmap = "coolwarm")
plt.show()
```



```
In [234... personal = df_air[df_air["Type of Travel"] == "Personal Travel"]
personal.head()
```

```
Out[234]:
```

	Unnamed: 0	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient
0	0	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	
5	5	Female	Loyal Customer	26	Personal Travel	Eco	1180	3	
6	6	Male	Loyal Customer	47	Personal Travel	Eco	1276	2	
11	11	Female	Loyal Customer	12	Personal Travel	Eco Plus	308	2	
13	13	Male	Loyal Customer	33	Personal Travel	Eco	946	4	

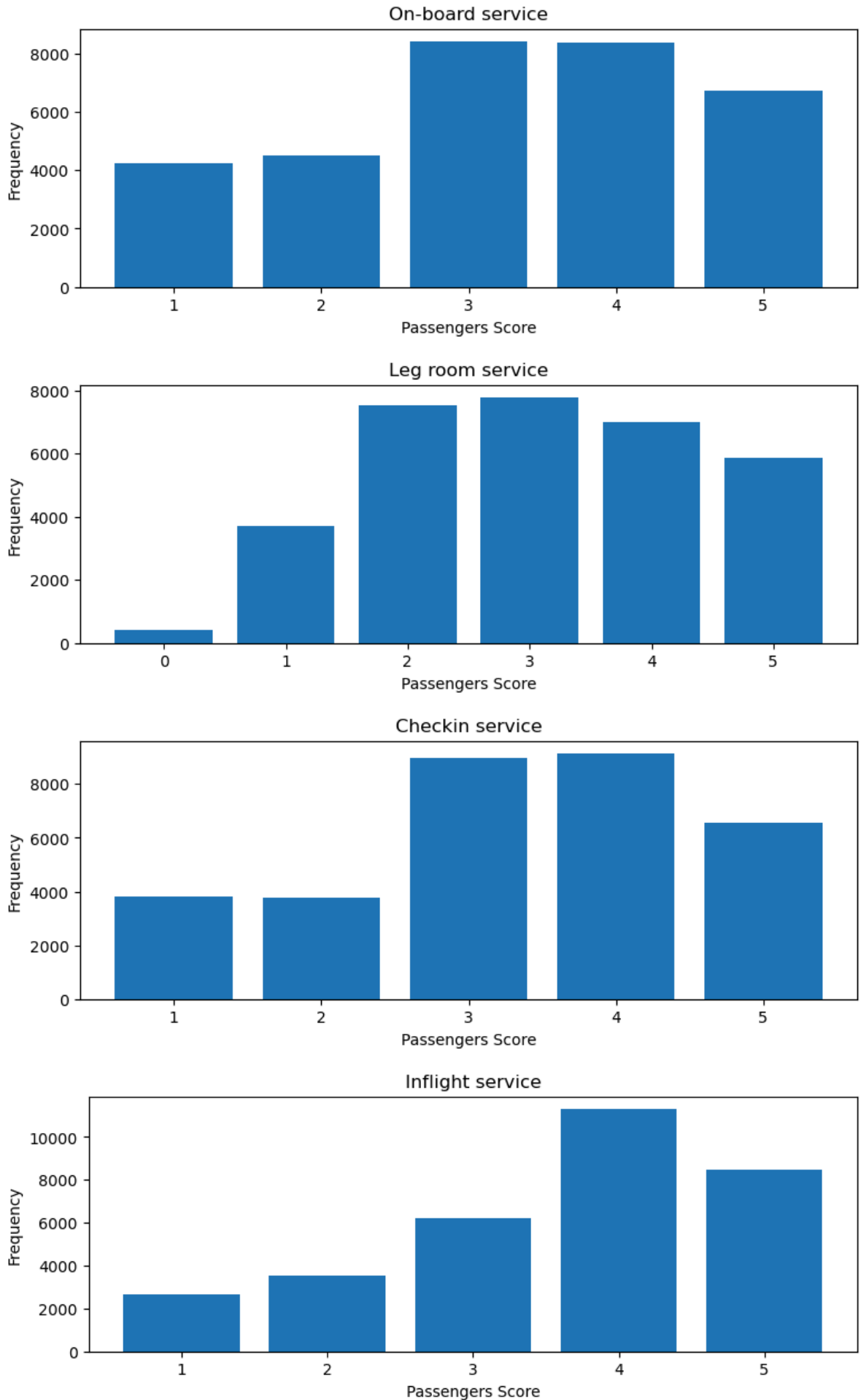
5 rows x 24 columns

```
In [ ]:
```

```
In [203... def service_plot(variable):
    var = personal[variable]
    var_Value = var.value_counts()
    plt.figure(figsize = (9, 3))
    plt.bar(var_Value.index, var_Value.values)
    plt.xlabel("Passengers Score")
    plt.ylabel("Frequency")
    plt.title(variable)
    plt.show()
```

```
In [205... service = ["On-board service", "Leg room service", "Checkin service", "Infl
for c in service:
```

```
service_plot(c)
```



```
In [207... service = ['Food and drink', 'Online boarding', 'Seat comfort', 'Inflight service']  
for c in service:
```

service\_plot(c)

