In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
#warnings
import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```
#load and check the data
df= pd.read_excel(r"airline_passenger_satisfaction.xlsx")
df.head()
```

Out[2]:

	ID	Gender	Age	Customer Type	Type of Travel	Class	Flight Distance	Departure Delay	Arrival Delay	Departure and Arrival Time Convenience	...	On-board Service	Se Comfc
0	1	Male	48	First-time	Business	Business	821	2	5.0	3	...	3	
1	2	Female	35	Returning	Business	Business	821	26	39.0	2	...	5	
2	3	Male	41	Returning	Business	Business	853	0	0.0	4	...	3	
3	4	Male	50	Returning	Business	Business	1905	0	0.0	2	...	5	
4	5	Female	49	Returning	Business	Business	3470	0	1.0	3	...	3	

5 rows × 24 columns

In [3]:
```
from dataprep.datasets import get_dataset_names
from dataprep.eda import create_report
from dataprep.clean import clean_df,clean_date
import pandas as pd
import seaborn as sns
df= pd.read_excel(r"airline_passenger_satisfaction.xlsx")
report=create_report(df)
```

```
  0%|
      | 0/2679 [00:00<…
```

In [4]:
```
report
```

Out[4]:

DataPrep Report | Overview | Variables ≡ | Interactions | Correlations | Missing Values

Overview

Dataset Statistics

Number of Variables	24
Number of Rows	129880

Dataset Insights

`ID` is uniformly distributed	Uniform
`Departu…` and `Arrival…` have	Similar

Missing Cells	393
Missing Cells (%)	0.0%
Duplicate Rows	0
Duplicate Rows (%)	0.0%
Total Size in Memory	59.9 MB
Average Row Size in Memory	484.0 B
Variable Types	Numerical: 5 Categorical: 19

similar distributions — Distribution

Departu… is skewed — Skewed

Arrival… is skewed — Skewed

Type of… has constant length 8 — Constant Length

Departu… has constant length 1 — Constant Length

Ease of… has constant length 1 — Constant Length

Check-i… has constant length 1 — Constant Length

Online … has constant length 1 — Constant Length

Gate Lo… has constant length 1 — Constant Length

1 2 3

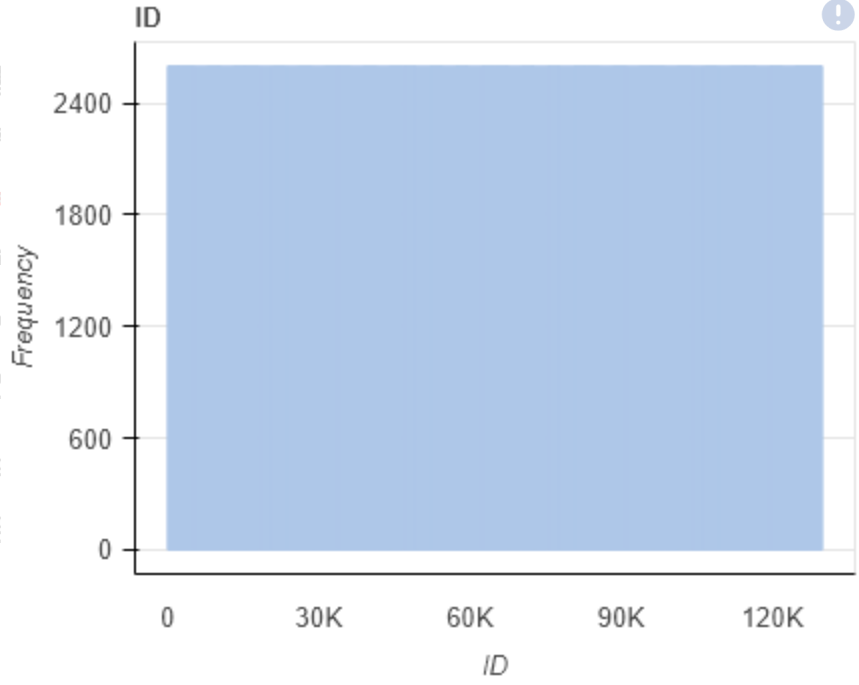
Variables

Sort by Feature order ▼ Reverse order

ID
numerical

Show Details

Approximate Distinct Count
Approximate Unique (%)
Missing
Missing (%)
Infinite
Infinite (%)
Memory



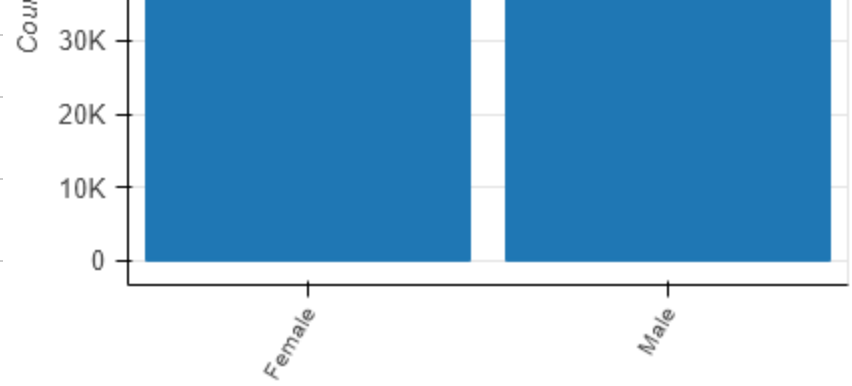
Gender

Approximate Distinct Count 2
Approximate Unique

Gender
categorical

Show Details

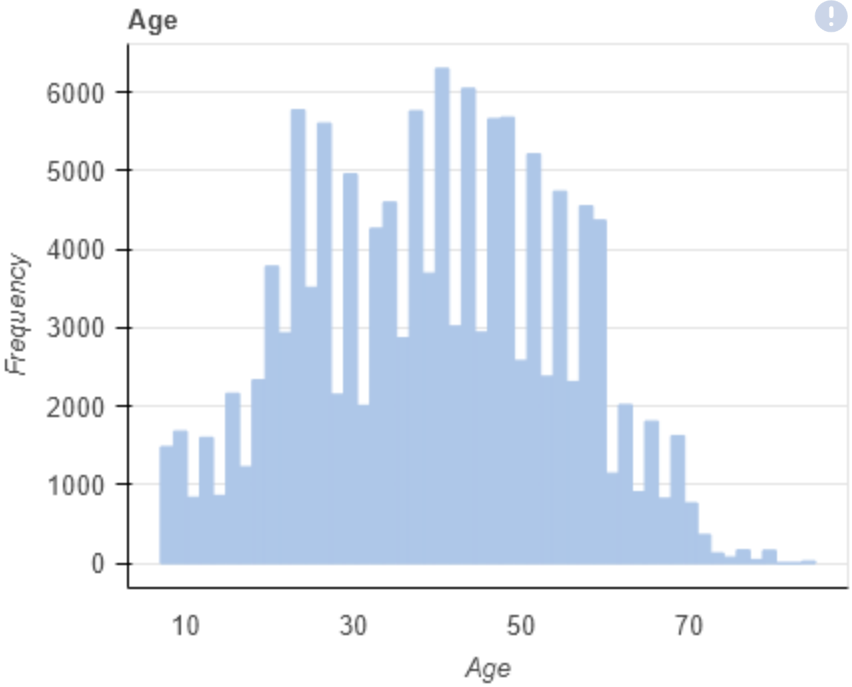
Missing (%)	0
Missing (%)	0.0%
Memory Size	8.7 MB



Age
numerical

Show Details

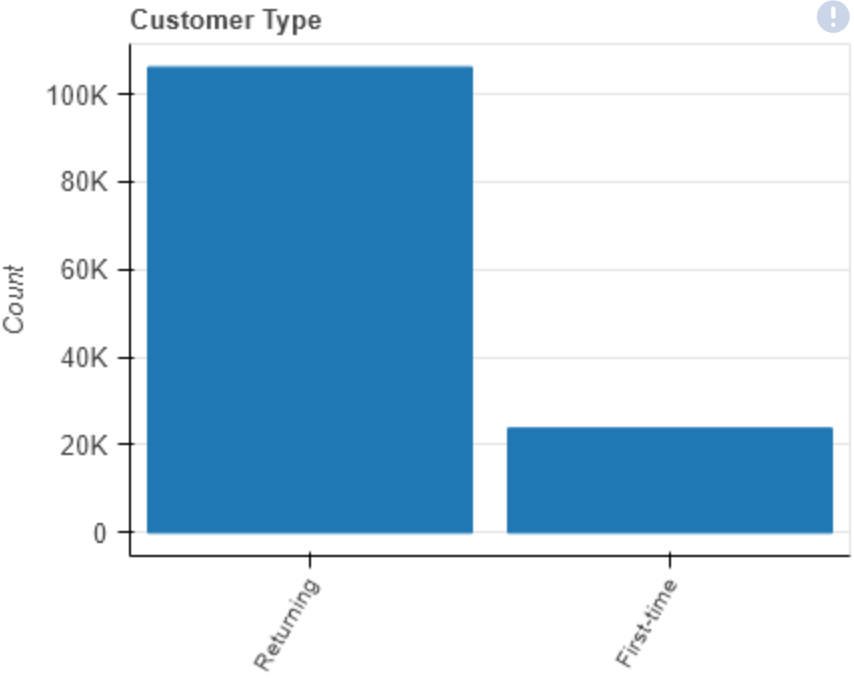
Approximate Distinct Count	75
Approximate Unique (%)	0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Memory	2.0

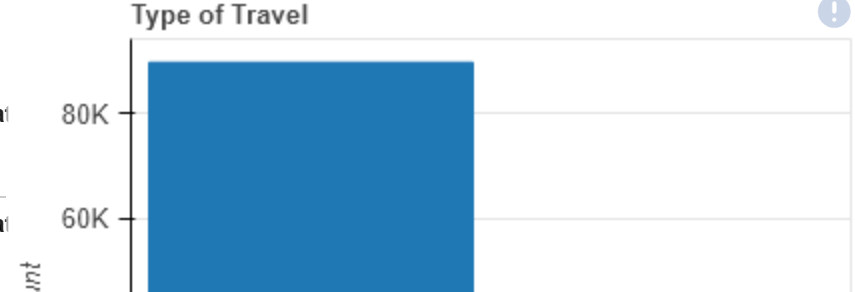


Customer Type
categorical

Show Details

Approximate Distinct Count	2
Approximate Unique (%)	0.0%
Missing	0
Missing (%)	0.0%
Memory Size	9.2 MB

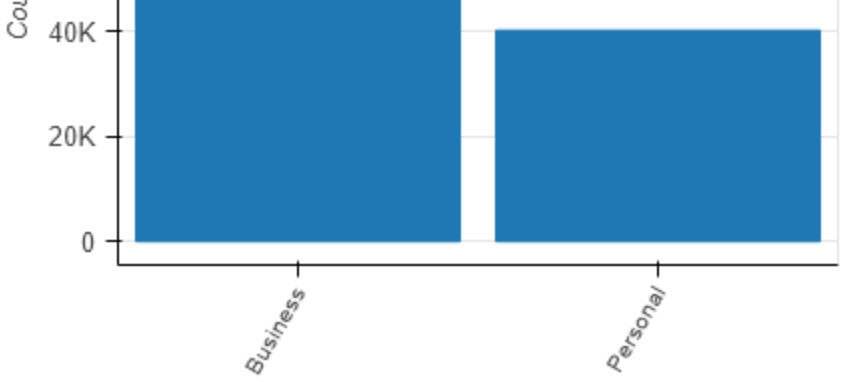




Type of Travel
categorical

Show Details

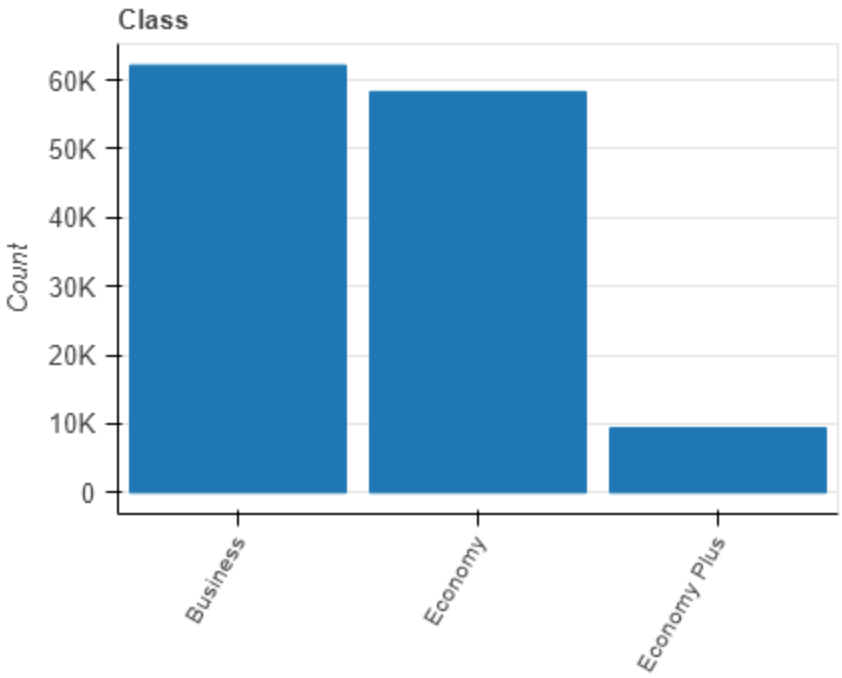
Missing	0
Missing (%)	0.0%
Memory Size	9.0 MB



Class
categorical

Show Details

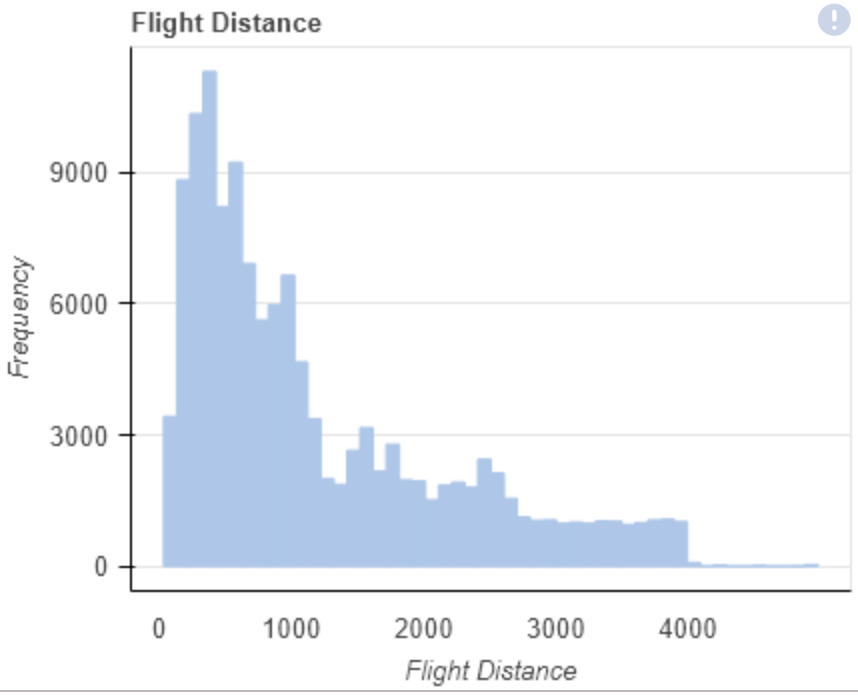
Approximate Distinct Count	3
Approximate Unique (%)	0.0%
Missing	0
Missing (%)	0.0%
Memory Size	9.0 MB

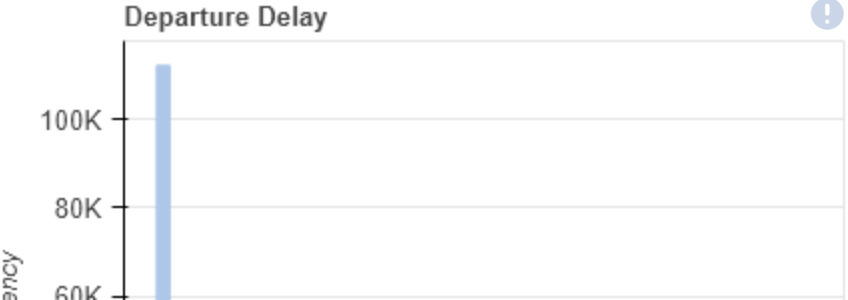


Flight Distance
numerical

Show Details

Approximate Distinct Count	1821
Approximate Unique (%)	2.9%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Memory	2.0

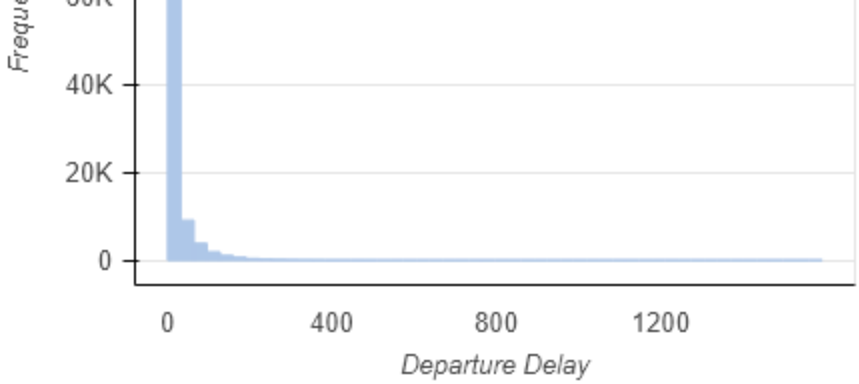




Departure Delay
numerical

Show Details

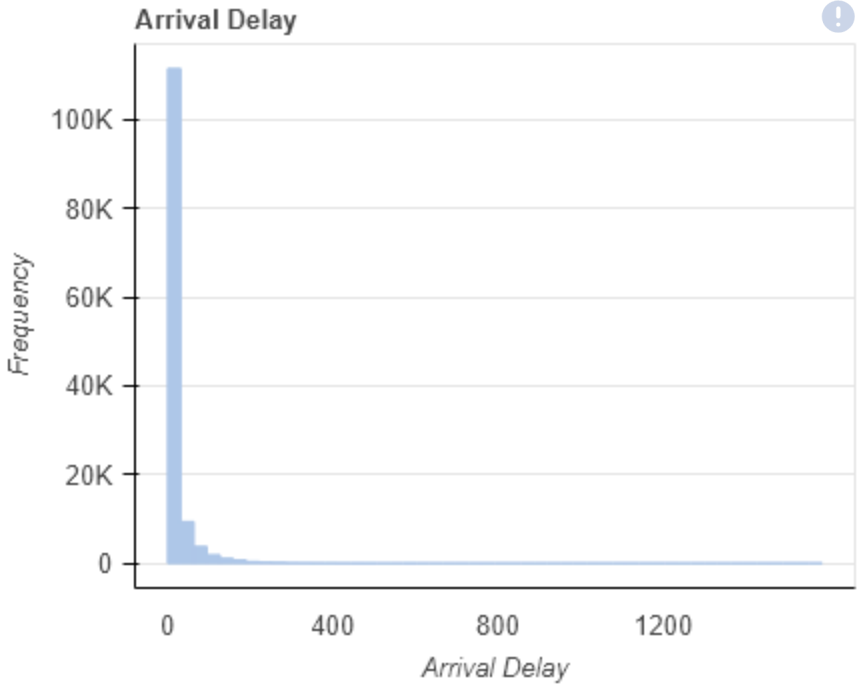
(%)
Zer
Memory Size MB
Zer
(%)
Mean 14.7137
Ne
Minimum 0
Ne
(%)
Maximum 1592



Arrival Delay
numerical

Show Details

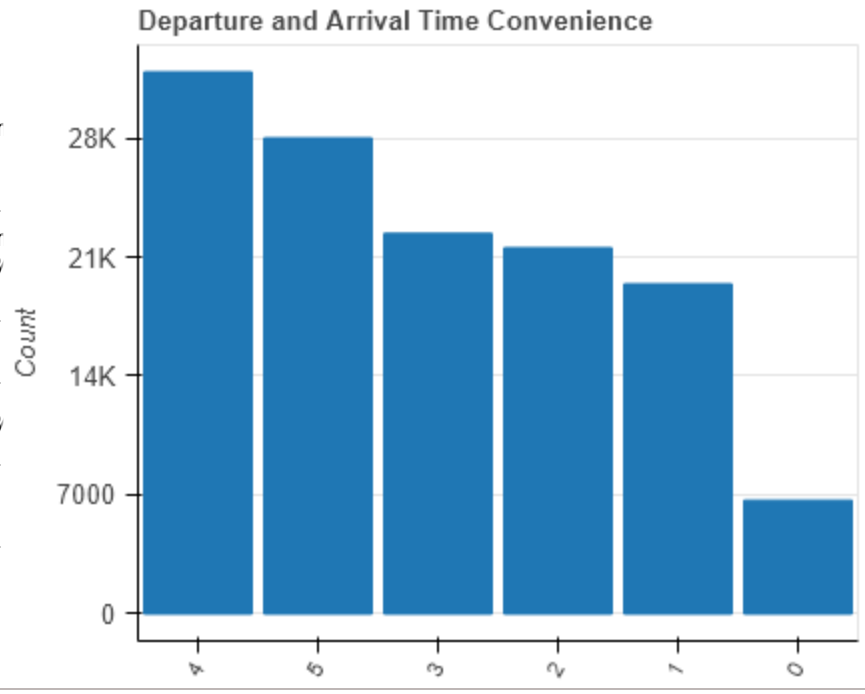
Approximate Distinct Count 472
Approximate Unique (%) 0.4%
Missing 393
Missing (%) 0.3%
Infinite 0
Infinite (%) 0.0%
Memory
Si
Me
Mi
Ma
Zer
Zer
(%)
Ne
Ne
(%)



Departure and A…
categorical

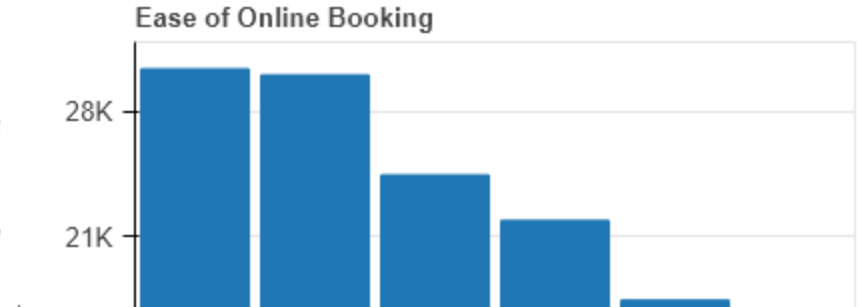
Show Details

Approximate Distinct Count
Approximate Unique (%)
Missing 0
Missing (%) 0.0%
Memory Size MB



Ease of Online B…

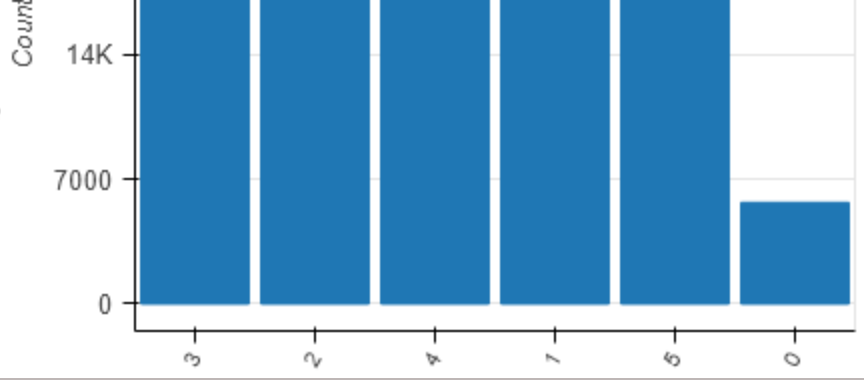
Approximate Distinct Count
Approximate Unique (%)



categorical

Show Details

Missing	0
Missing (%)	0.0%
Memory Size	8.2 MB

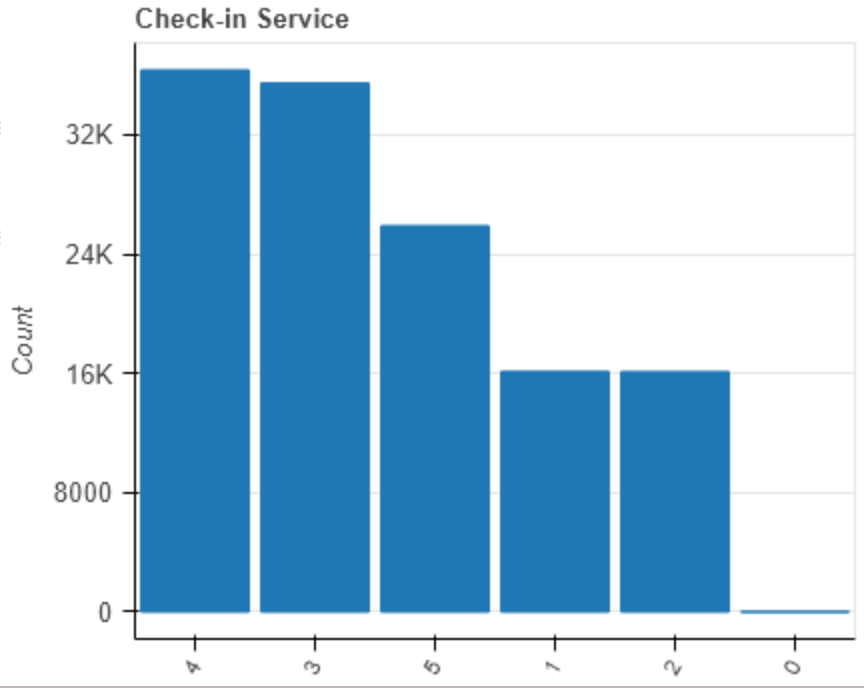


Check-in Service

categorical

Show Details

Approximate Distinct Count	6
Approximate Unique (%)	0.0%
Missing	0
Missing (%)	0.0%
Memory Size	8.2 MB

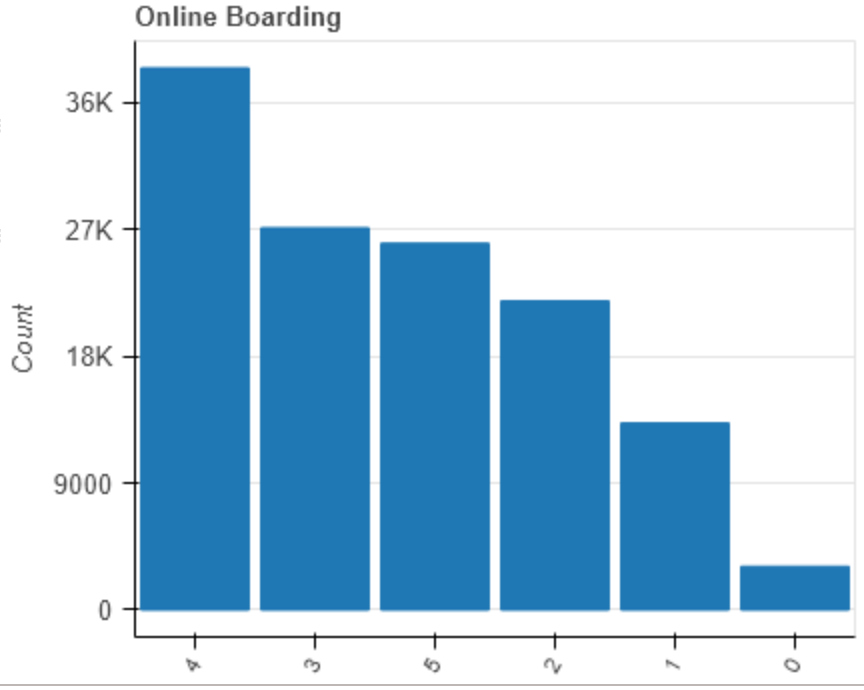


Online Boarding

categorical

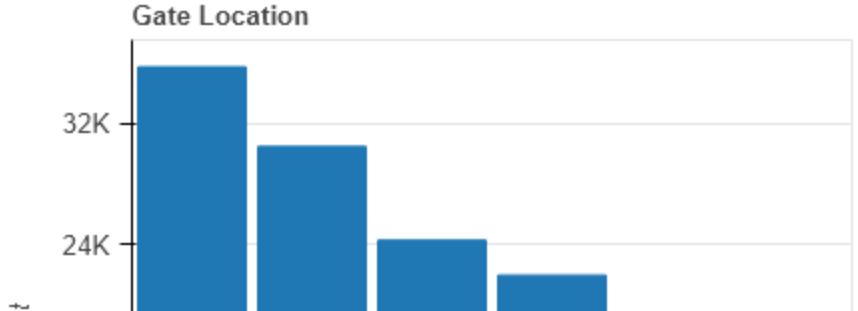
Show Details

Approximate Distinct Count	6
Approximate Unique (%)	0.0%
Missing	0
Missing (%)	0.0%
Memory Size	8.2 MB



Gate Location

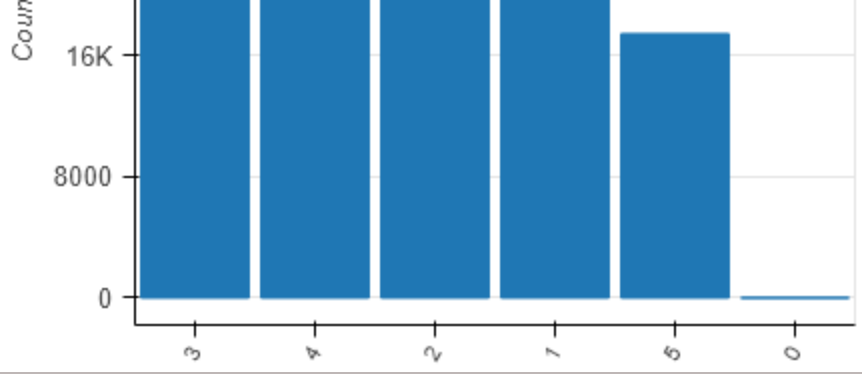
Approximate Distinct Count	6
Approximate Unique (%)	0.0%



categorical

Show Details

Missing	0
Missing (%)	0.0%
Memory Size	8.2 MB

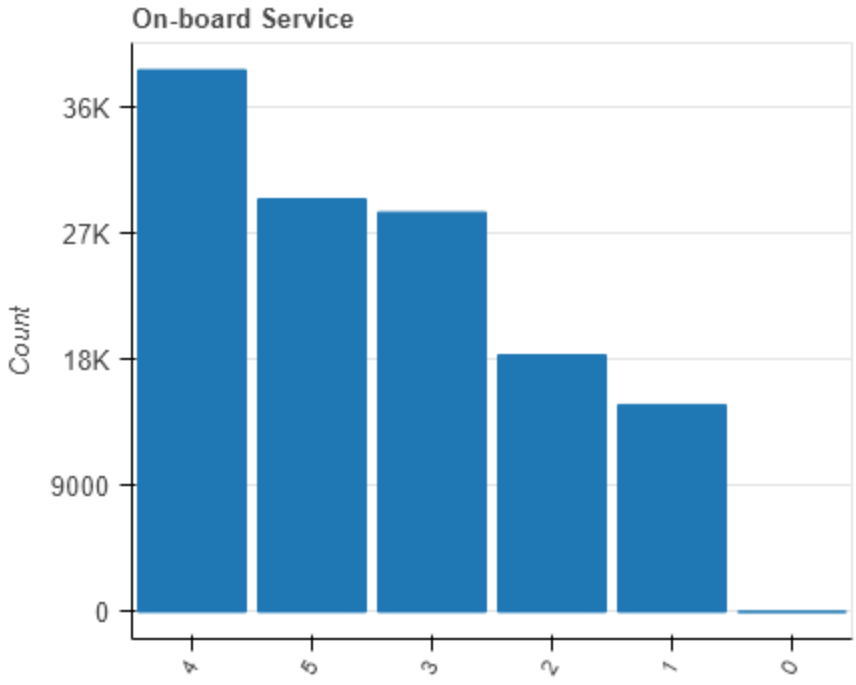


On-board Service

categorical

Show Details

Approximate Distinct Count	6
Approximate Unique (%)	0.0%
Missing	0
Missing (%)	0.0%
Memory Size	8.2 MB

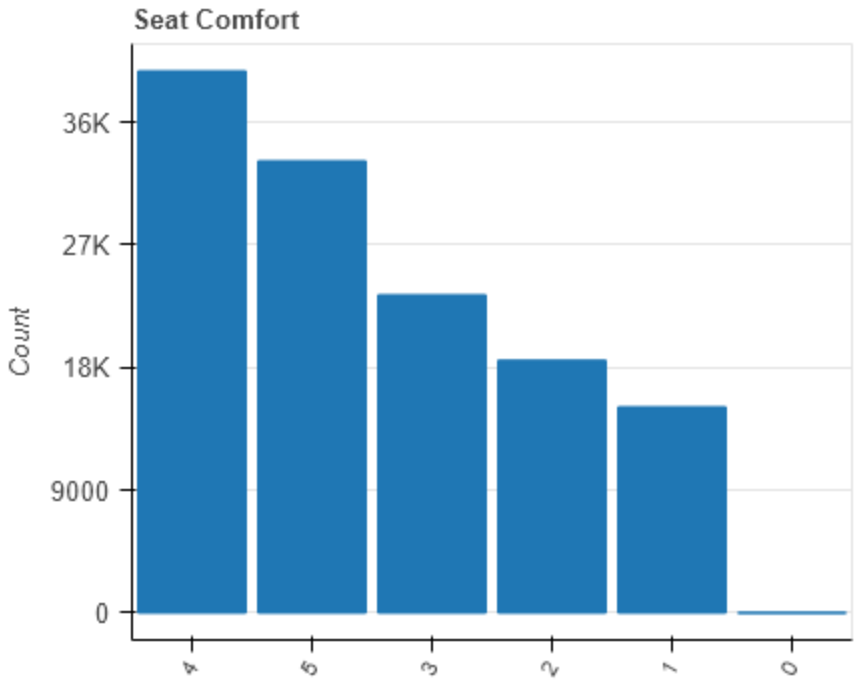


Seat Comfort

categorical

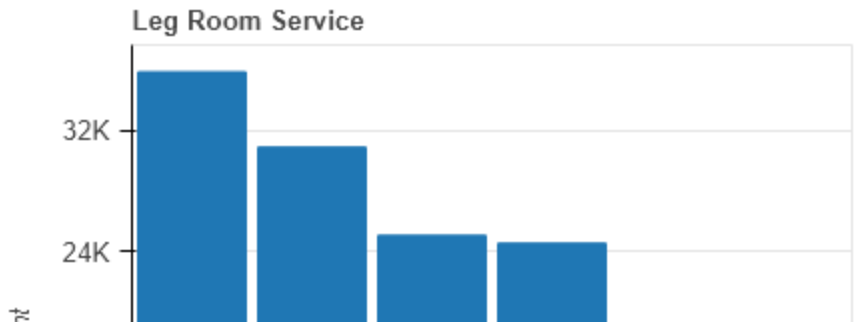
Show Details

Approximate Distinct Count	6
Approximate Unique (%)	0.0%
Missing	0
Missing (%)	0.0%
Memory Size	8.2 MB



Leg Room Service

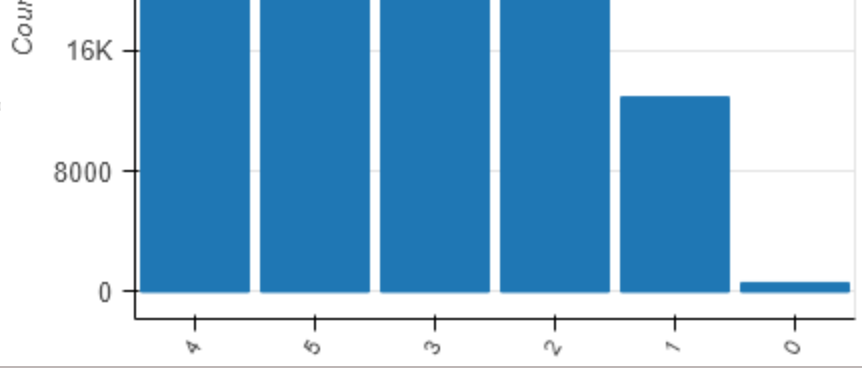
Approximate Distinct Count	6
Approximate Unique (%)	0.0%



categorical

Show Details

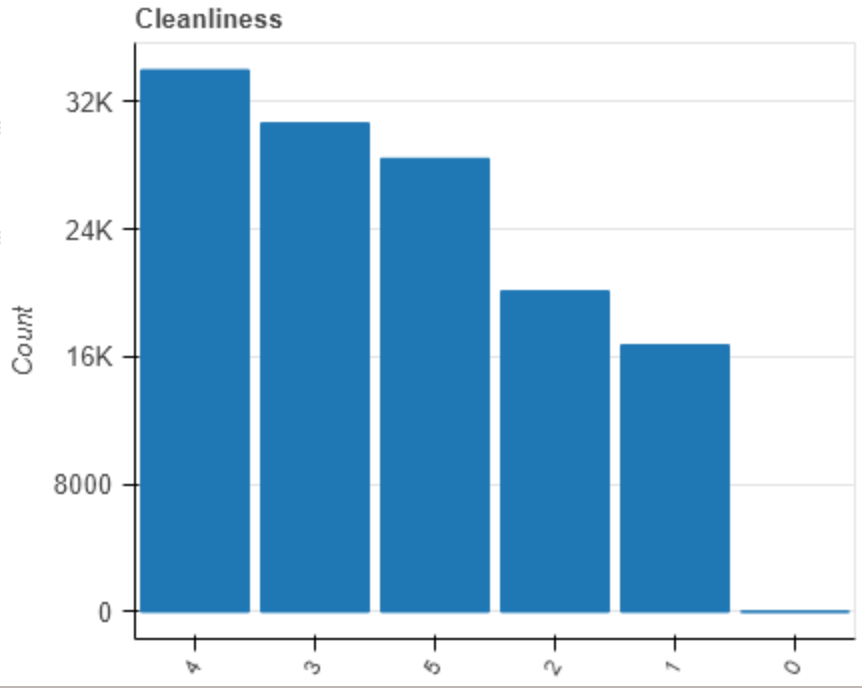
Missing	0
Missing (%)	0.0%
Memory Size	8.2 MB



Cleanliness
categorical

Show Details

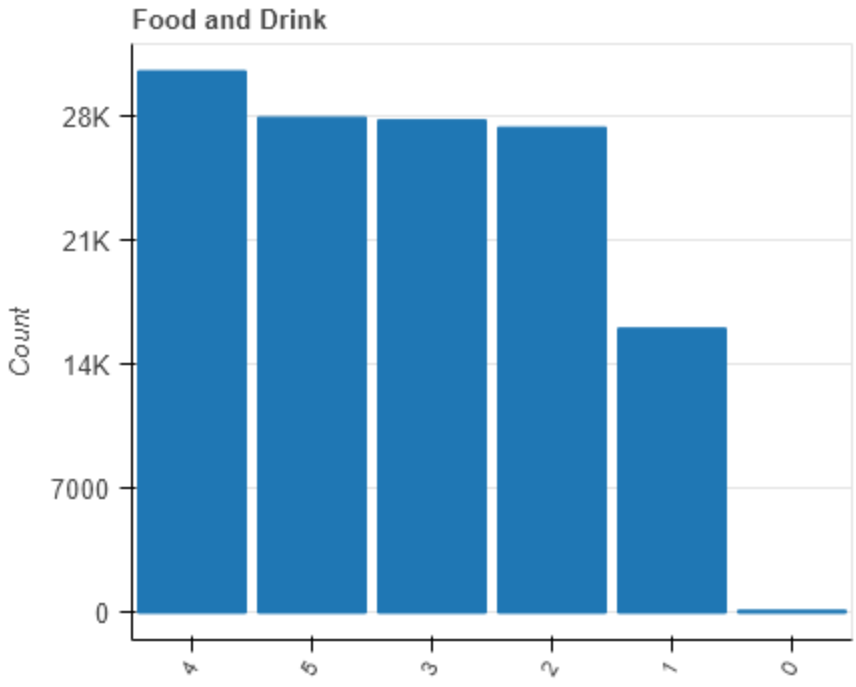
Approximate Distinct Count	6
Approximate Unique (%)	0.0%
Missing	0
Missing (%)	0.0%
Memory Size	8.2 MB



Food and Drink
categorical

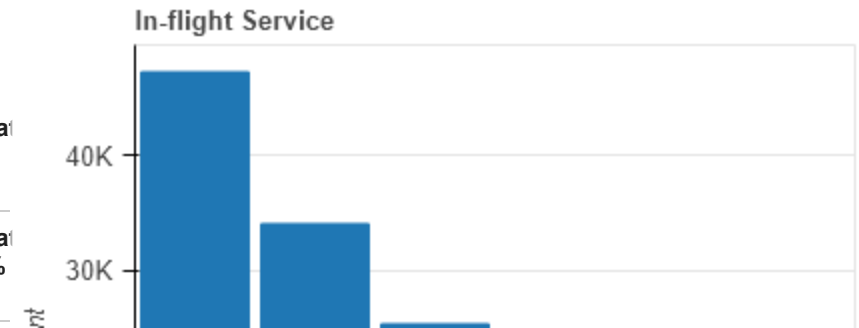
Show Details

Approximate Distinct Count	6
Approximate Unique (%)	0.0%
Missing	0
Missing (%)	0.0%
Memory Size	8.2 MB



In-flight Service

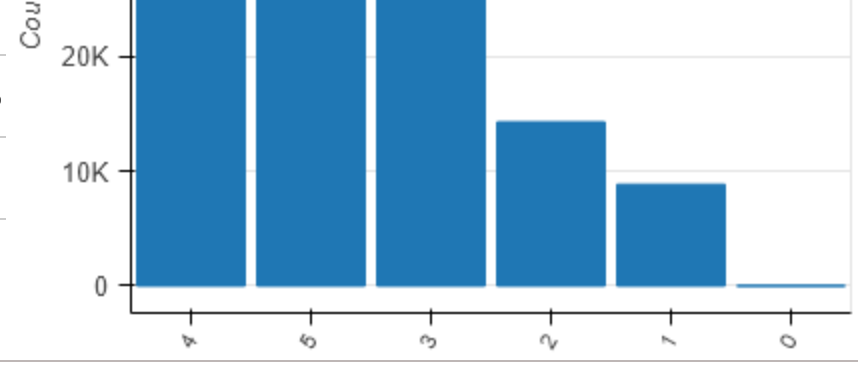
Approximate Distinct Count	6
Approximate Unique (%)	0.0%



categorical

Show Details

Missing	0
Missing (%)	0.0%
Memory Size	3.2 MB

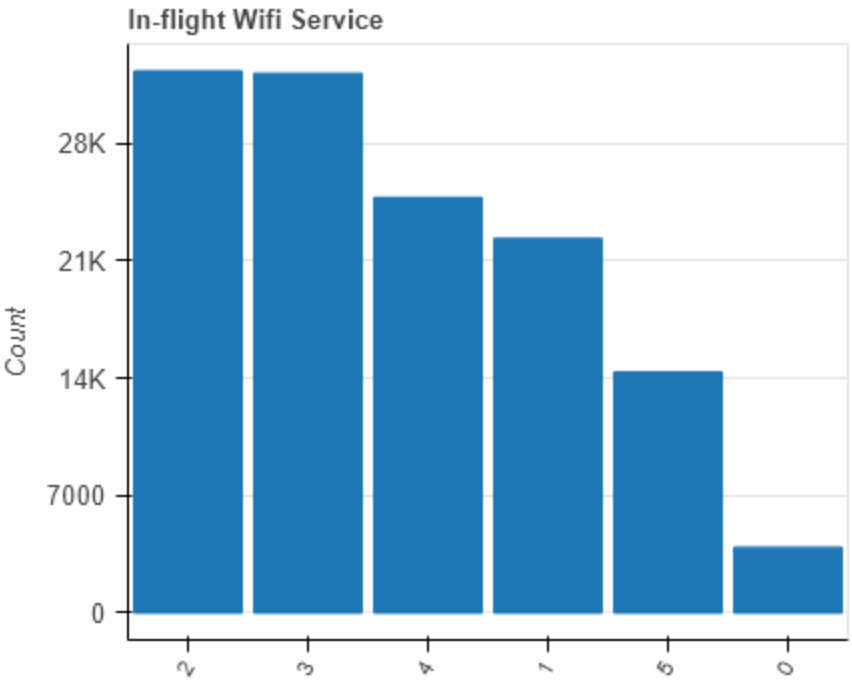


In-flight Wifi Serv…

categorical

Show Details

Approxi Distinct Count	6
Approxi Unique (%)	0.0%
Missing	0
Missing (%)	0.0%
Memory Size	3.2 MB

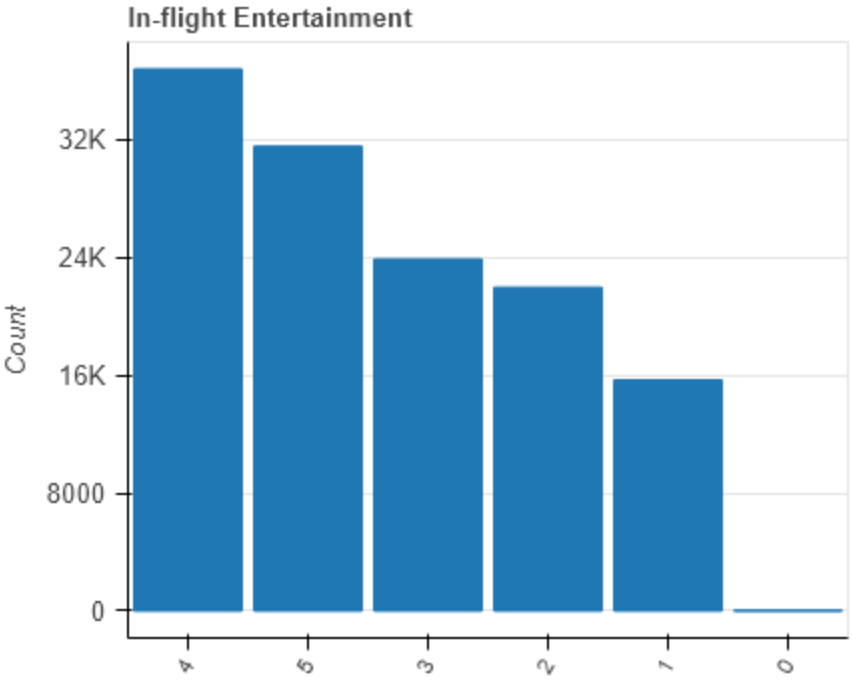


In-flight Entertain…

categorical

Show Details

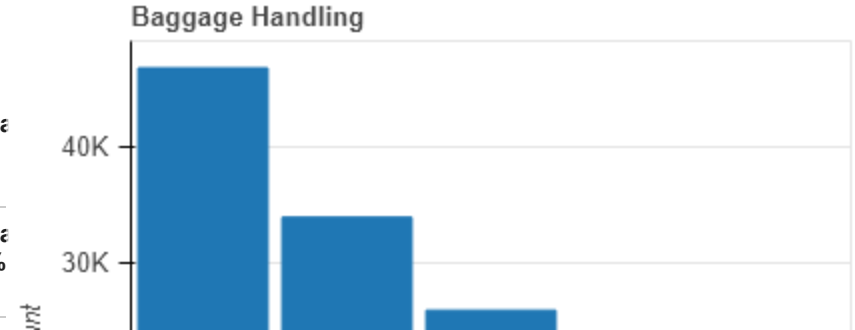
Approxi Distinct Count	6
Approxi Unique (%)	0.0%
Missing	0
Missing (%)	0.0%
Memory Size	3.2 MB



Baggage Handling

categorical

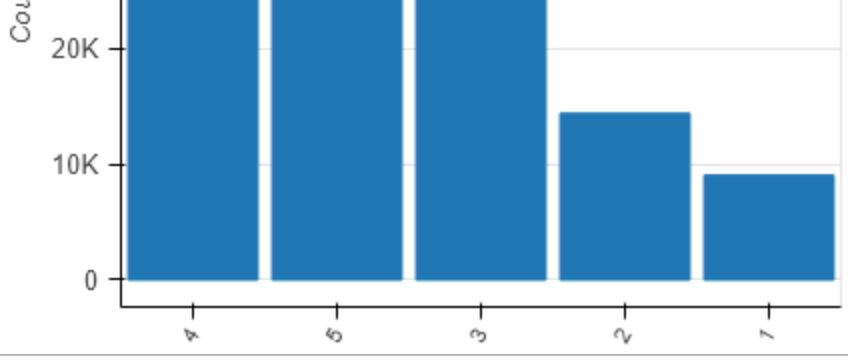
Approxima Distinct Count	5
Approxima Unique (%)	0.0%



categorical

Show Details

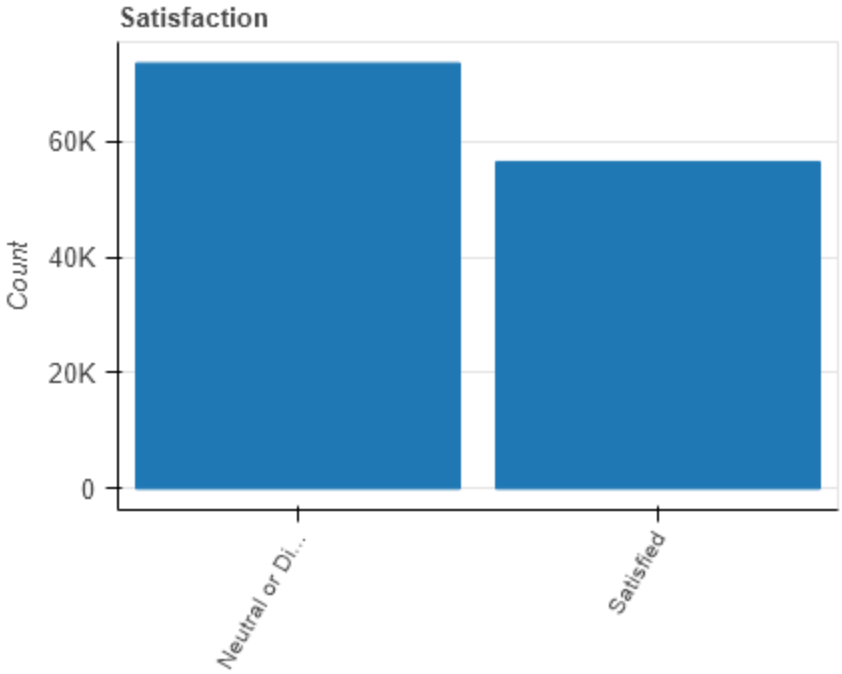
Missing	0
Missing (%)	0.0%
Memory Size	8.2 MB



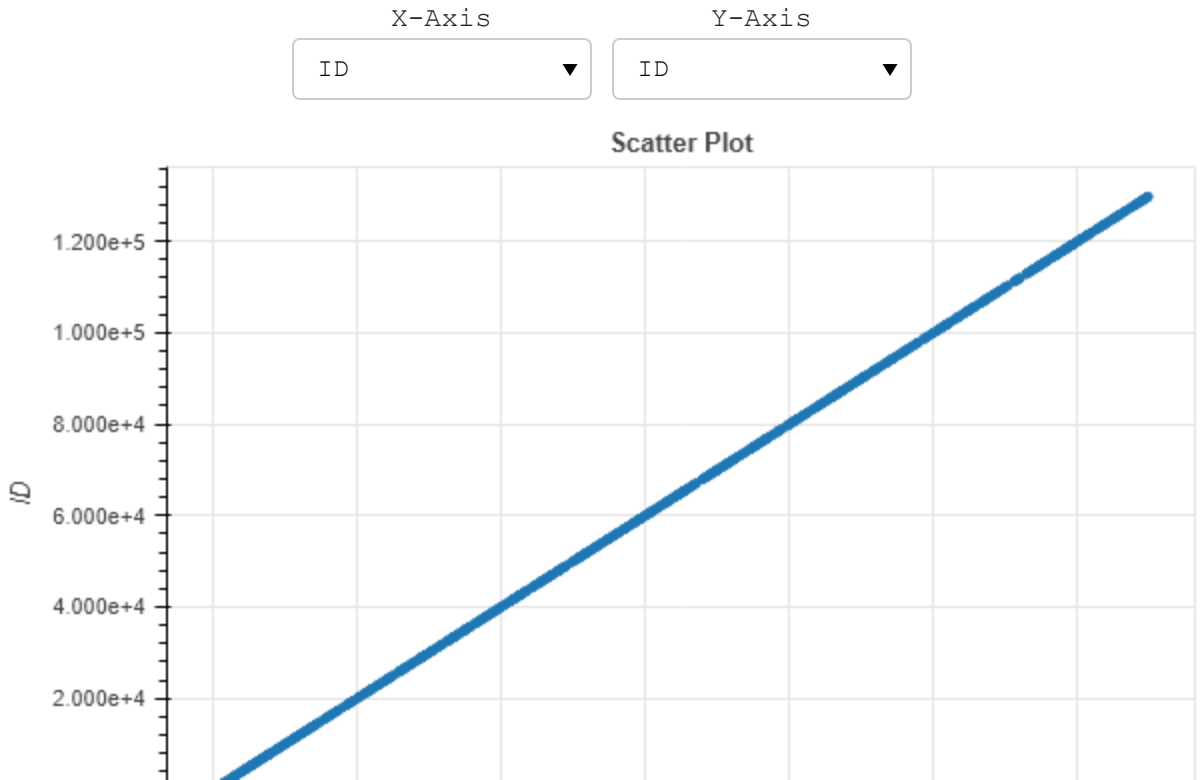
Satisfaction
categorical

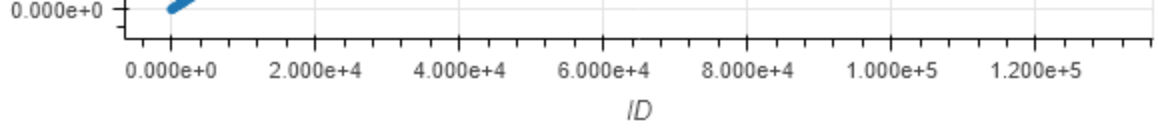
Show Details

Approximate Distinct Count	2
Approximate Unique (%)	0.0%
Missing	0
Missing (%)	0.0%
Memory Size	10.1 MB

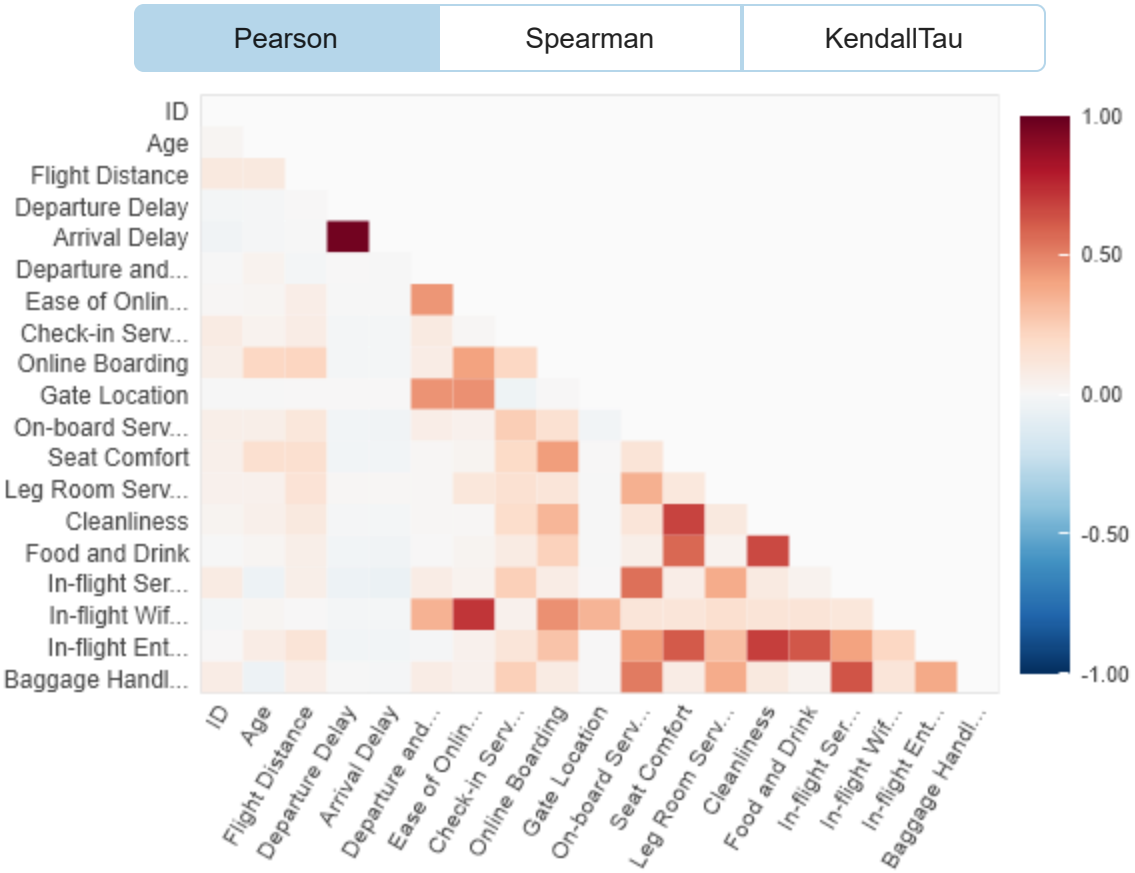


Interactions

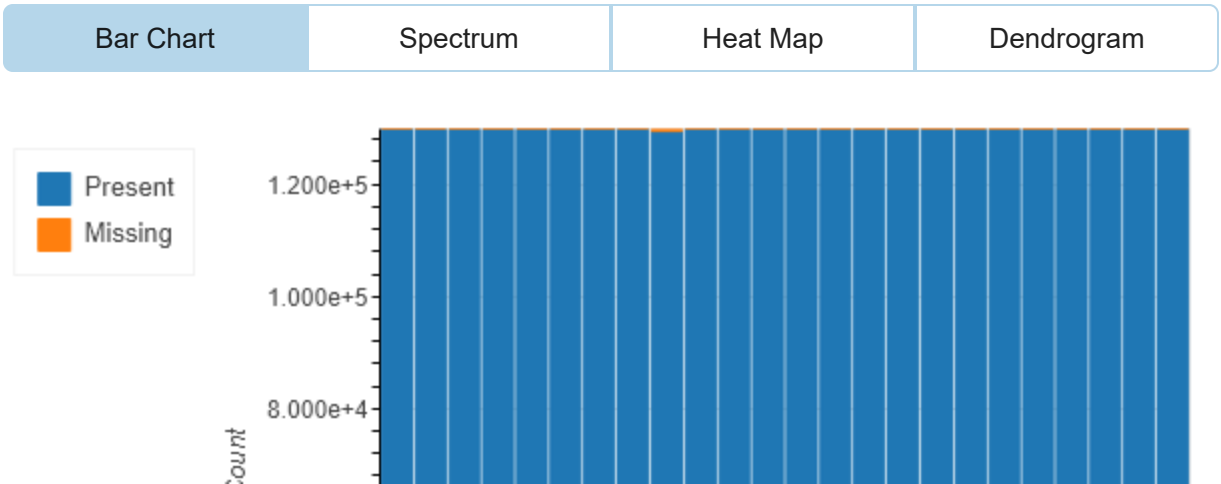


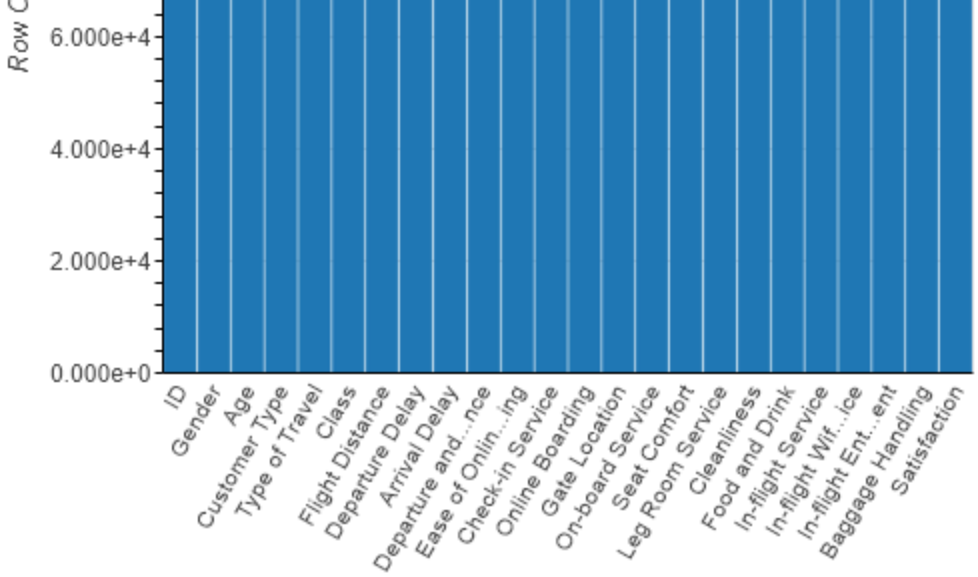


Correlations



Missing Values





Report generated with DataPrep

In [5]:
```
# Get the Total Passengers
num_unique_ids = df['ID'].nunique()

# Print Total Passengers
print("Total Passengers:", num_unique_ids)
```

```
Total Passengers: 129880
```

In [6]:
```
# Get the count of passengers who were satisfied (assuming 'Satisfaction' column contain
num_satisfied = df['Satisfaction'].value_counts()['Satisfied']

# Print the Total satisfied passengers
print("Total satisfied passengers:", num_satisfied)
```

```
Total satisfied passengers: 56428
```

In [7]:
```
# Get the count of passengers who were unsatisfied (assuming 'Satisfaction' column conta
num_satisfied = df['Satisfaction'].value_counts()['Neutral or Dissatisfied']

# Print the Total satisfied passengers
print("Total unsatisfied passengers:", num_satisfied)
```

```
Total unsatisfied passengers: 73452
```

In [8]:
```
# Get the count of male and female passengers
gender_counts = df['Gender'].value_counts()

# Extract the counts for male and female passengers
num_males = gender_counts['Male']
num_females = gender_counts['Female']

# Print the counts of male and female passengers
print("Total male passengers:", num_males)
print("Total female passengers:", num_females)
```

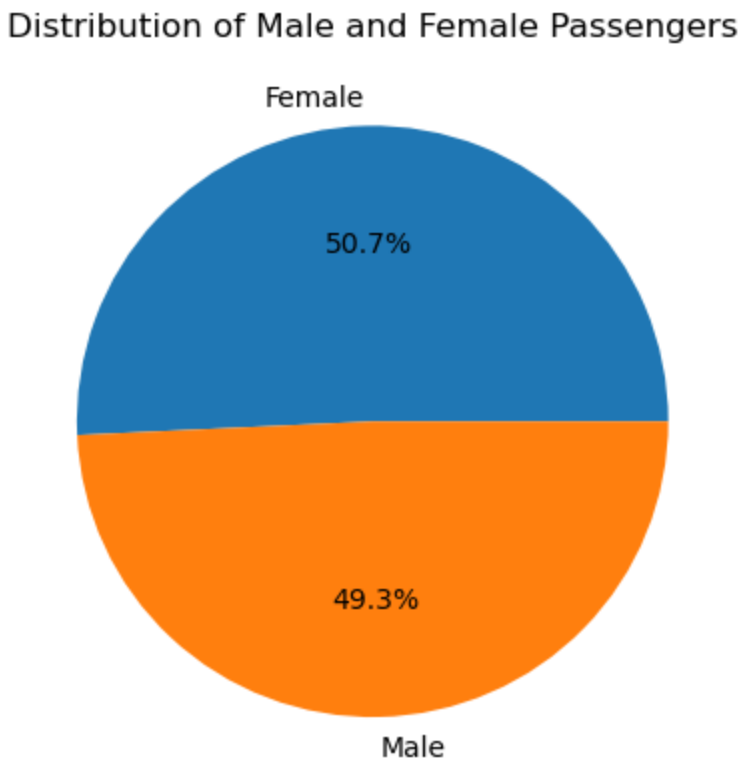
```
Total male passengers: 63981
Total female passengers: 65899
```

In [9]:
```
# Create a pie chart to visualize the counts
```

```
plt.pie(gender_counts.values, labels=gender_counts.index, autopct='%1.1f%%')

# Set the chart title
plt.title('Distribution of Male and Female Passengers')

# Display the chart
plt.show()
```



In [10]:
```
# Calculate the mean rating for each service category
mean_ratings = df[['Ease of Online Booking','Check-in Service','Online Boarding','Gate L

# Print the mean ratings for each service category
print(mean_ratings)
```

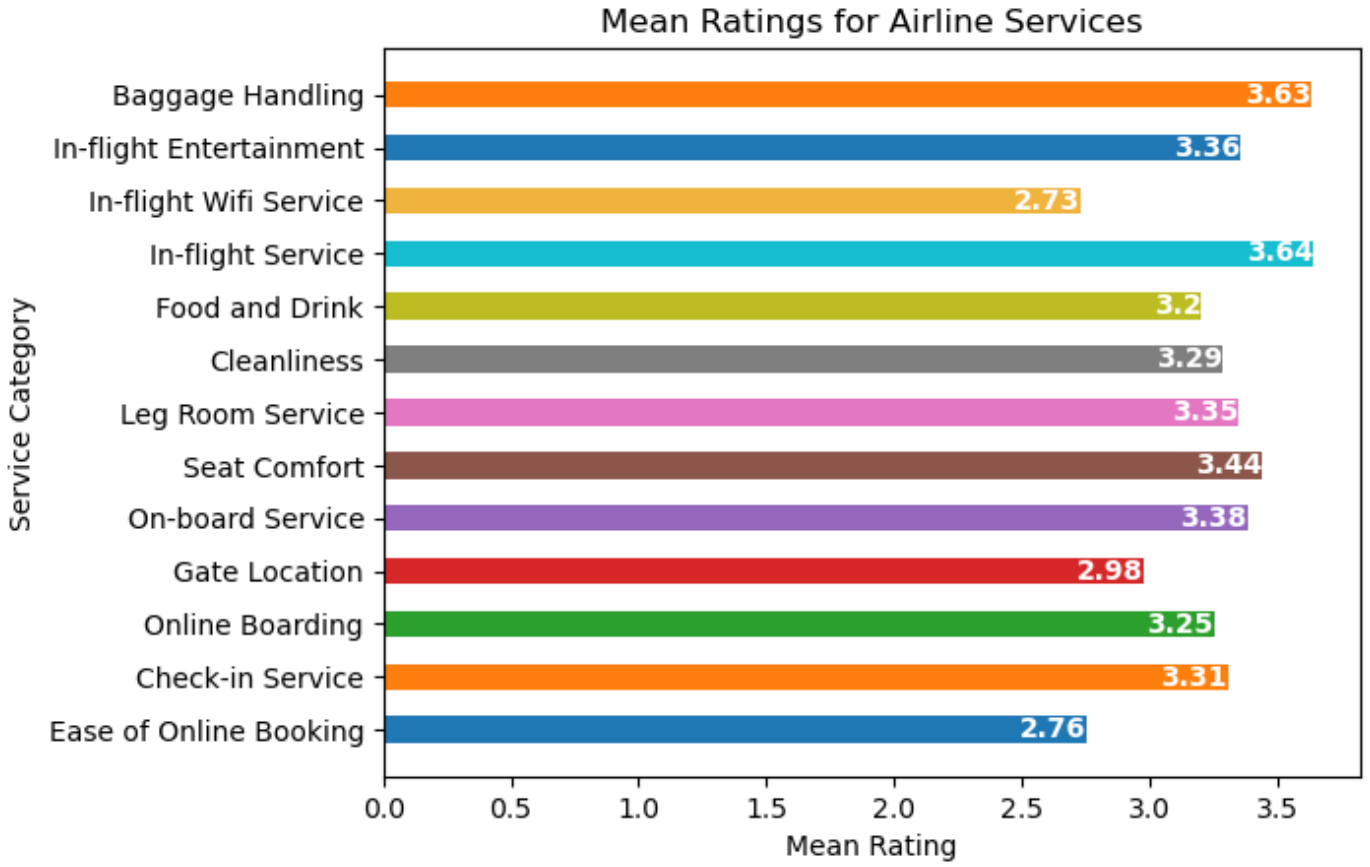
```
Ease of Online Booking     2.756876
Check-in Service           3.306267
Online Boarding            3.252633
Gate Location              2.976925
On-board Service           3.383023
Seat Comfort               3.441361
Leg Room Service           3.350878
Cleanliness                3.286326
Food and Drink             3.204774
In-flight Service          3.642193
In-flight Wifi Service     2.728696
In-flight Entertainment    3.358077
Baggage Handling           3.632114
dtype: float64
```

In [11]:
```
# Define a list of colors for each service category
colors = ['#1f77b4', '#ff7f0e', '#2ca02c', '#d62728', '#9467bd', '#8c564b', '#e377c2', '
# Create a horizontal bar chart of the mean ratings
plt.barh(mean_ratings.index, mean_ratings.values, height=0.5, color=colors)

# Set the title and axis labels
plt.title('Mean Ratings for Airline Services')
plt.xlabel('Mean Rating')
plt.ylabel('Service Category')
```

```
# Display the number of ratings above each bar
for i, v in enumerate(mean_ratings.values):
    plt.text(v, i, str(round(v, 2)), color='white', fontweight='bold', ha='right', va='c

# Display the chart
plt.show()
```



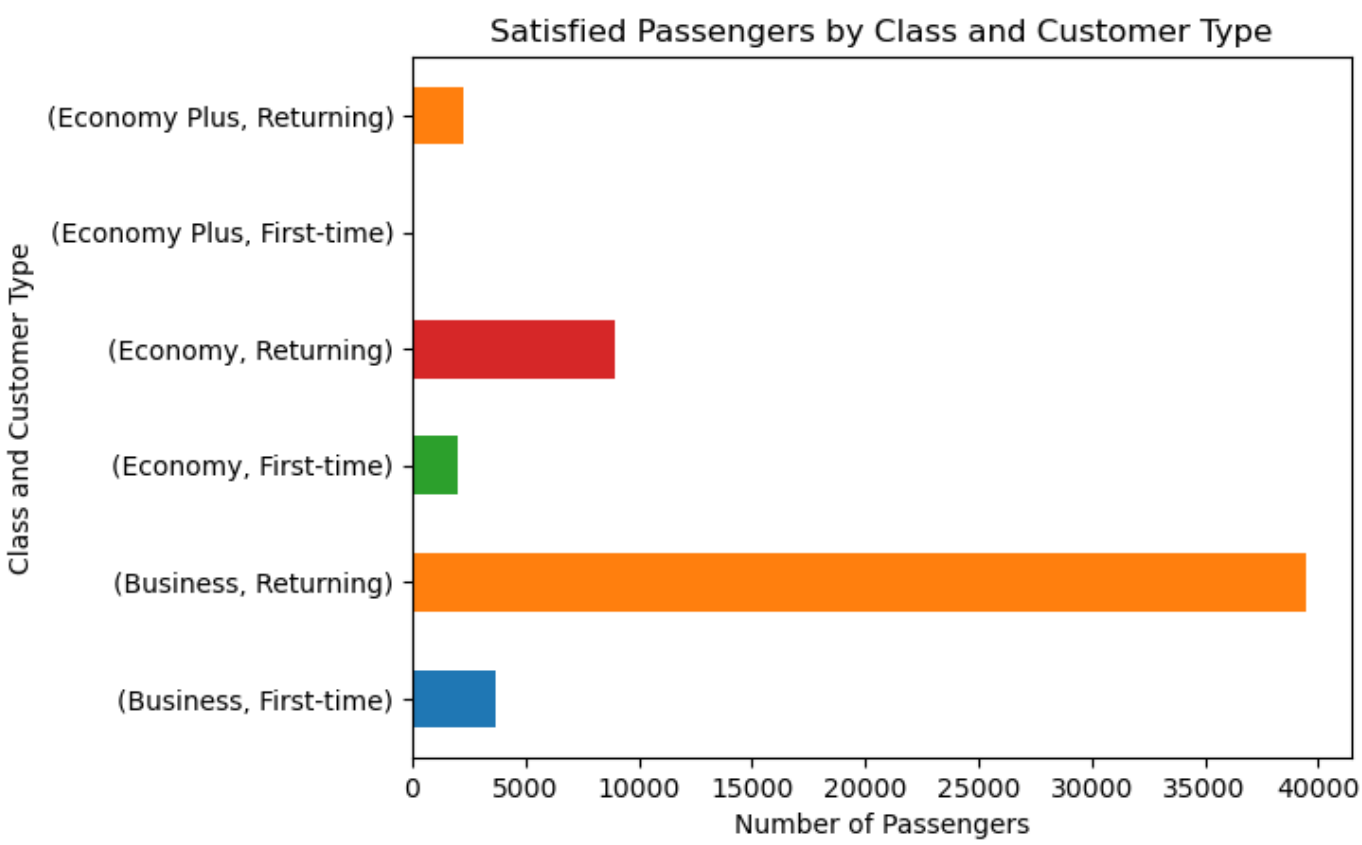
In [12]:

```
# Get the number of satisfied passengers by class and customer type
satisfied = df[df['Satisfaction'] == 'Satisfied'].groupby(['Class', 'Customer Type']).si

# Create a horizontal bar chart of the satisfied passengers by class and customer type
satisfied.plot(kind='barh', color=['#1f77b4', '#ff7f0e', '#2ca02c', '#d62728'])

# Set the title and axis labels
plt.title('Satisfied Passengers by Class and Customer Type')
plt.xlabel('Number of Passengers')
plt.ylabel('Class and Customer Type')

# Display the chart
plt.show()
```



In [13]:

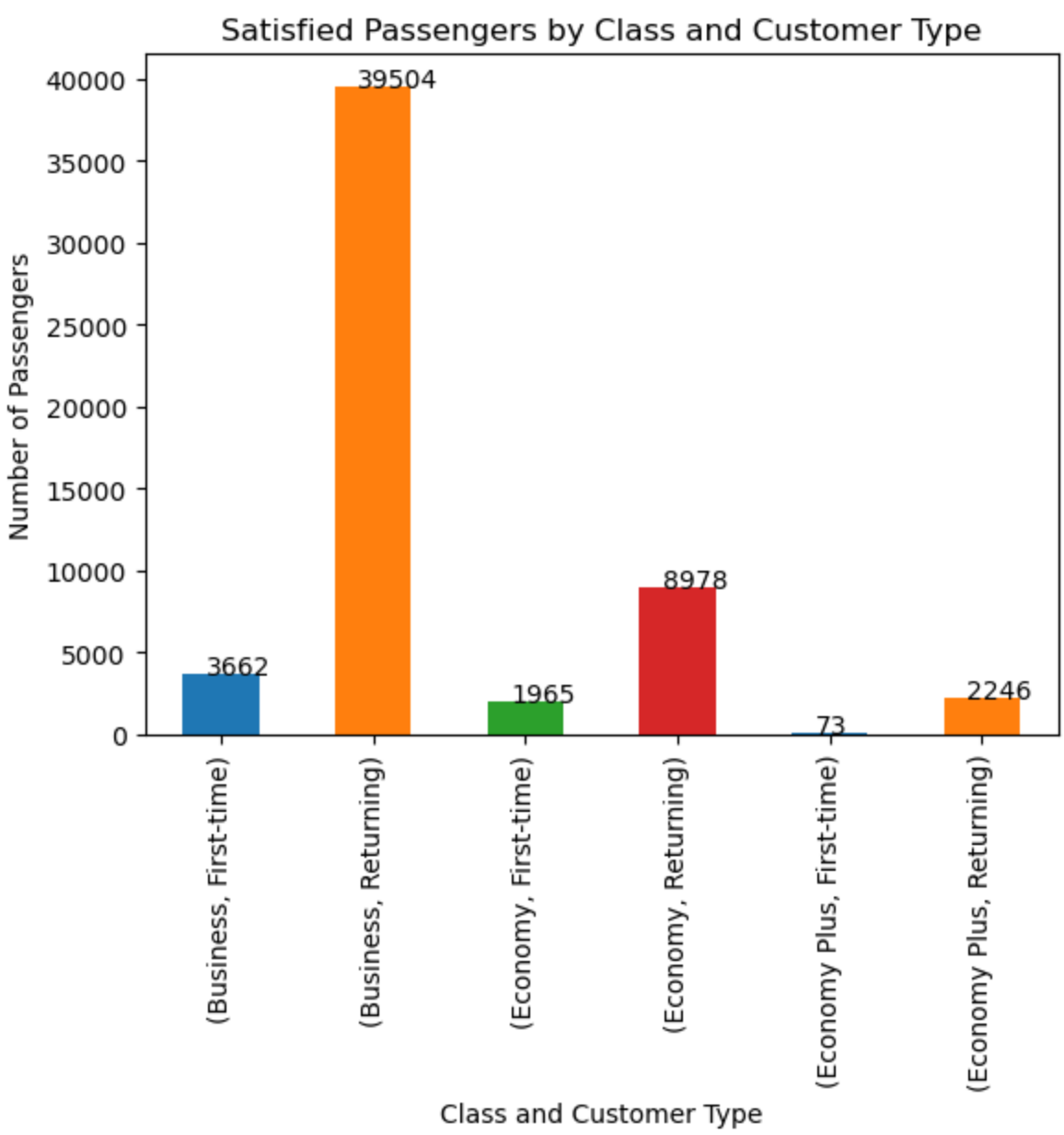
```
# Get the number of satisfied passengers by class and customer type
satisfied = df[df['Satisfaction'] == 'Satisfied'].groupby(['Class', 'Customer Type']).si

# Create a vertical bar chart of the satisfied passengers by class and customer type
satisfied.plot(kind='bar', color=['#1f77b4', '#ff7f0e', '#2ca02c', '#d62728'])

# Set the title and axis labels
plt.title('Satisfied Passengers by Class and Customer Type')
plt.xlabel('Class and Customer Type')
plt.ylabel('Number of Passengers')

# Add the numbers above the bars
for i, v in enumerate(satisfied):
    plt.text(i - 0.1, v + 1, str(v))

# Display the chart
plt.show()
```



In [14]:
```
from subprocess import call
call(['python','-m','nbconvert',r'airline_passenger_satisfaction.xlsx'])
```

Out[14]:
```
1
```