```python
import micropip
await micropip.install('seaborn')
# 1. Importing Libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Set Seaborn style
sns.set(style="whitegrid")
%matplotlib inline

# 2. Load the Dataset
df = pd.read_csv("train.csv")  # Adjust path if needed

# 3. Basic Info
print("Shape of data:", df.shape)
print("\nData Types:")
print(df.dtypes)
print("\nMissing values:")
print(df.isnull().sum())
```

```
Shape of data: (891, 12)

Data Types:
PassengerId      int64
Survived         int64
Pclass           int64
Name            object
Sex             object
Age            float64
SibSp            int64
Parch            int64
Ticket          object
Fare           float64
Cabin           object
Embarked        object
dtype: object

Missing values:
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
```

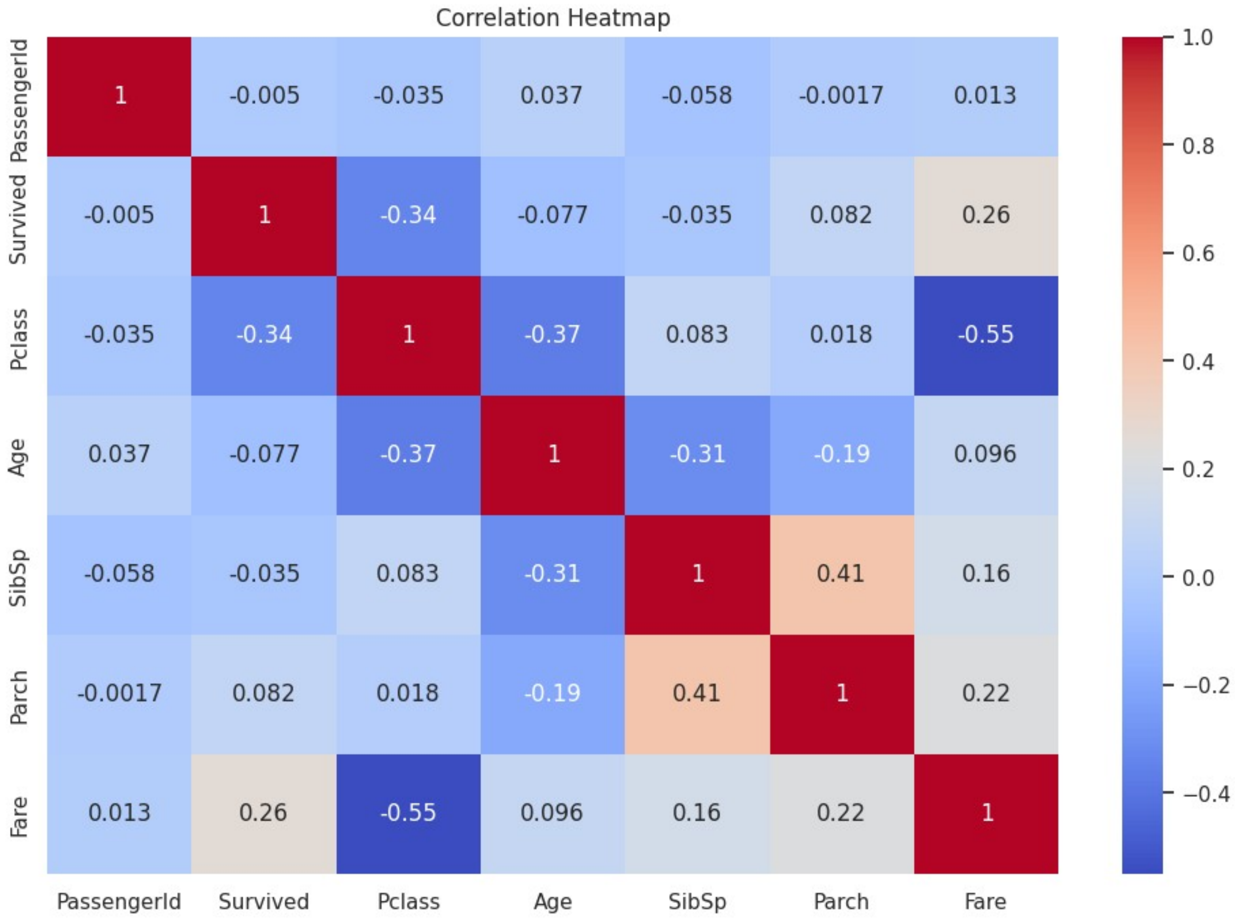
```
Embarked         2
dtype: int64

# 4. Summary Statistics
print("\nDescriptive Statistics:")
print(df.describe(include='all'))


Descriptive Statistics:
        PassengerId    Survived      Pclass                     Name
Sex  \
count    891.000000  891.000000  891.000000                      891
891
unique          NaN         NaN         NaN                      891
2
top             NaN         NaN         NaN  Braund, Mr. Owen Harris
male
freq            NaN         NaN         NaN                        1
577
mean     446.000000    0.383838    2.308642                      NaN
NaN
std      257.353842    0.486592    0.836071                      NaN
NaN
min        1.000000    0.000000    1.000000                      NaN
NaN
25%      223.500000    0.000000    2.000000                      NaN
NaN
50%      446.000000    0.000000    3.000000                      NaN
NaN
75%      668.500000    1.000000    3.000000                      NaN
NaN
max      891.000000    1.000000    3.000000                      NaN
NaN

               Age       SibSp       Parch  Ticket        Fare
Cabin  \
count   714.000000  891.000000  891.000000     891  891.000000
204
unique         NaN         NaN         NaN     681         NaN
147
top            NaN         NaN         NaN  347082         NaN  B96
B98
freq           NaN         NaN         NaN       7         NaN
4
mean     29.699118    0.523008    0.381594     NaN   32.204208
NaN
std      14.526497    1.102743    0.806057     NaN   49.693429
NaN
min       0.420000    0.000000    0.000000     NaN    0.000000
NaN
```

```
25%      20.125000    0.000000    0.000000     NaN    7.910400
NaN
50%      28.000000    0.000000    0.000000     NaN   14.454200
NaN
75%      38.000000    1.000000    0.000000     NaN   31.000000
NaN
max      80.000000    8.000000    6.000000     NaN  512.329200
NaN

       Embarked
count       889
unique        3
top           S
freq        644
mean        NaN
std         NaN
min         NaN
25%         NaN
50%         NaN
75%         NaN
max         NaN
```

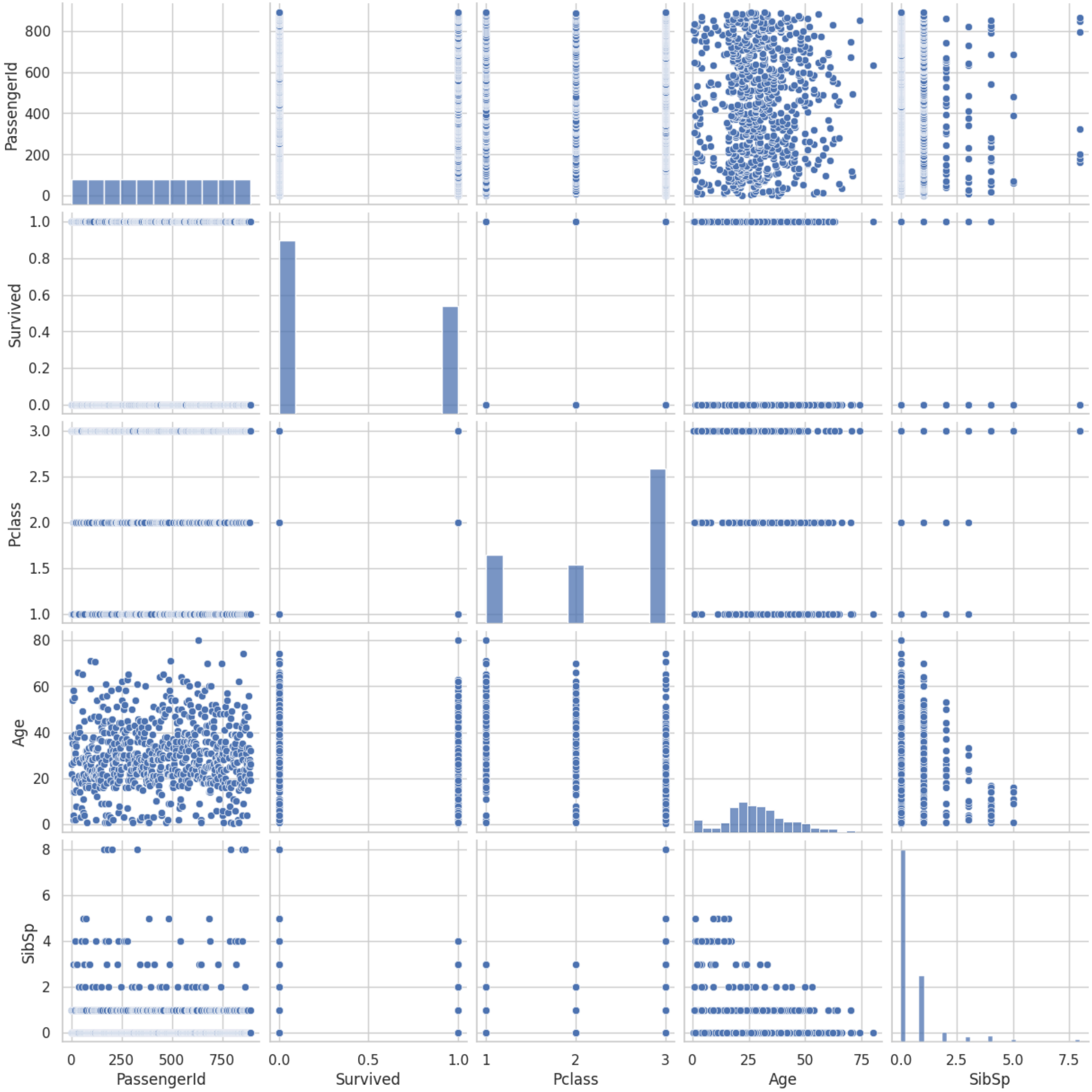
```python
# 5. Value Counts for Categorical Columns
cat_cols = df.select_dtypes(include='object').columns
for col in cat_cols:
    print(f"\nValue Counts for {col}:")
    print(df[col].value_counts())
```

```

Value Counts for Name:
Name
Braund, Mr. Owen Harris                     1
Boulos, Mr. Hanna                           1
Frolicher-Stehli, Mr. Maxmillian            1
Gilinski, Mr. Eliezer                       1
Murdlin, Mr. Joseph                         1
                                           ..
Kelly, Miss. Anna Katherine "Annie Kate"    1
McCoy, Mr. Bernard                          1
Johnson, Mr. William Cahoone Jr             1
Keane, Miss. Nora A                         1
Dooley, Mr. Patrick                         1
Name: count, Length: 891, dtype: int64

Value Counts for Sex:
Sex
male      577
female    314
Name: count, dtype: int64
```

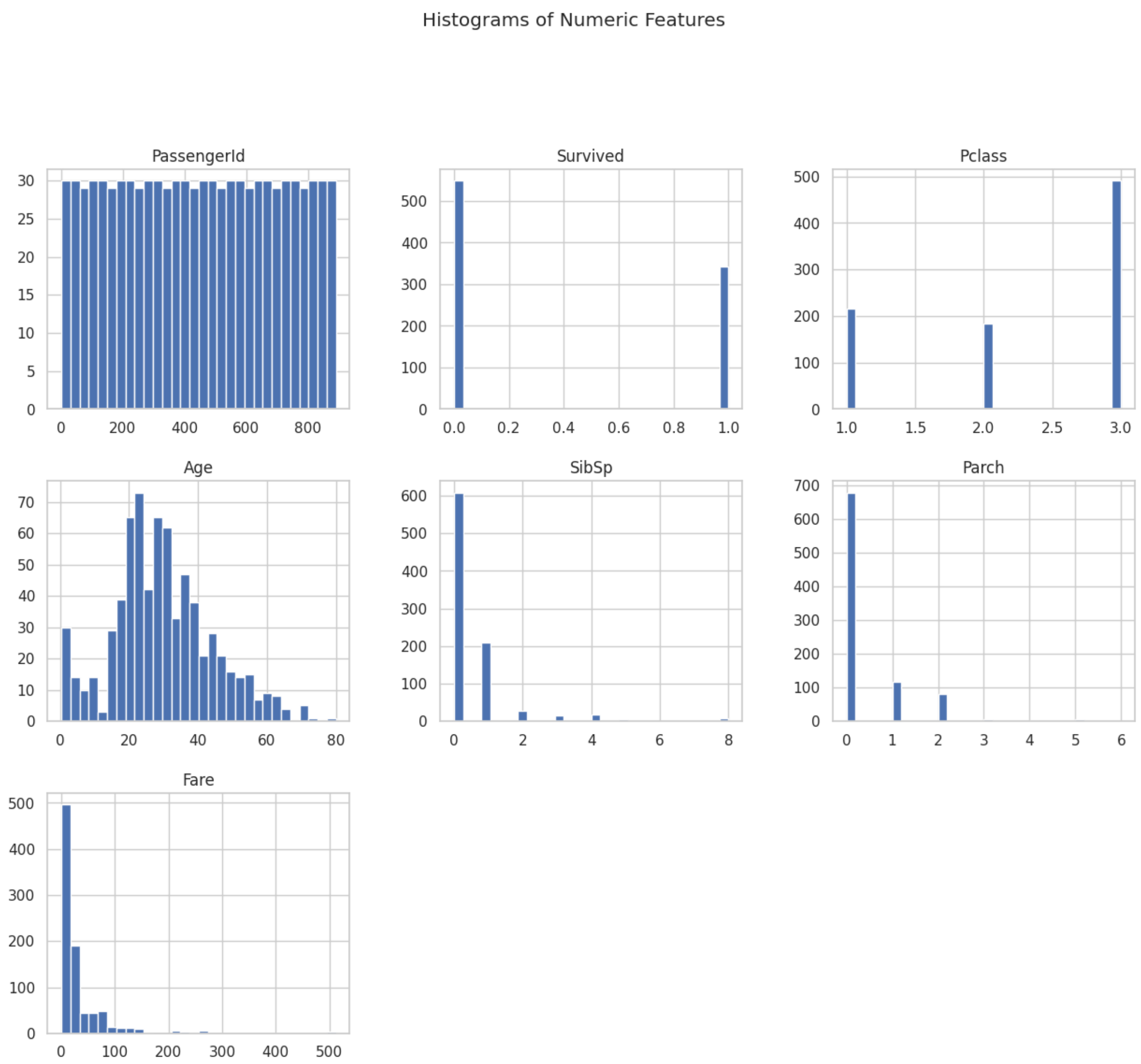
```
Value Counts for Ticket:
Ticket
347082      7
CA. 2343    7
1601        7
3101295     6
CA 2144     6
           ..
9234        1
19988       1
2693        1
PC 17612    1
370376      1
Name: count, Length: 681, dtype: int64

Value Counts for Cabin:
Cabin
B96 B98        4
G6             4
C23 C25 C27    4
C22 C26        3
F33            3
              ..
E34            1
C7             1
C54            1
E36            1
C148           1
Name: count, Length: 147, dtype: int64

Value Counts for Embarked:
Embarked
S    644
C    168
Q     77
Name: count, dtype: int64

# 6. Correlation Heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm')
plt.title("Correlation Heatmap")
plt.show()
```



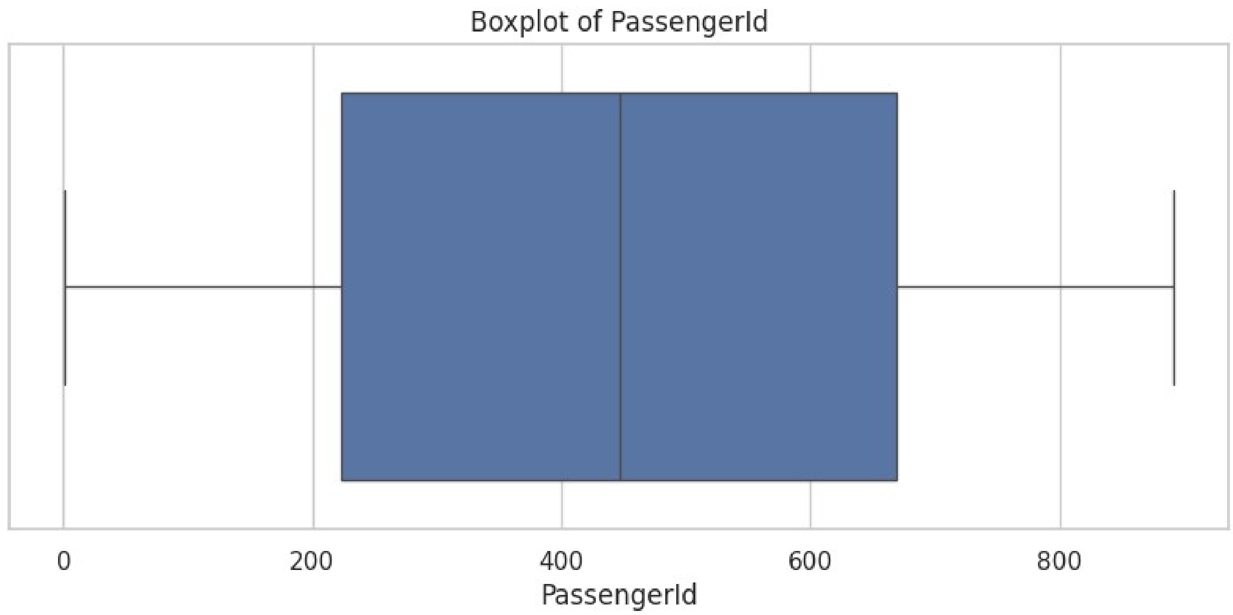
```
# 7. Pairplot (only on numeric subset to avoid overload)
numeric_cols = df.select_dtypes(include='number').columns[:5]
sns.pairplot(df[numeric_cols])
plt.show()
```

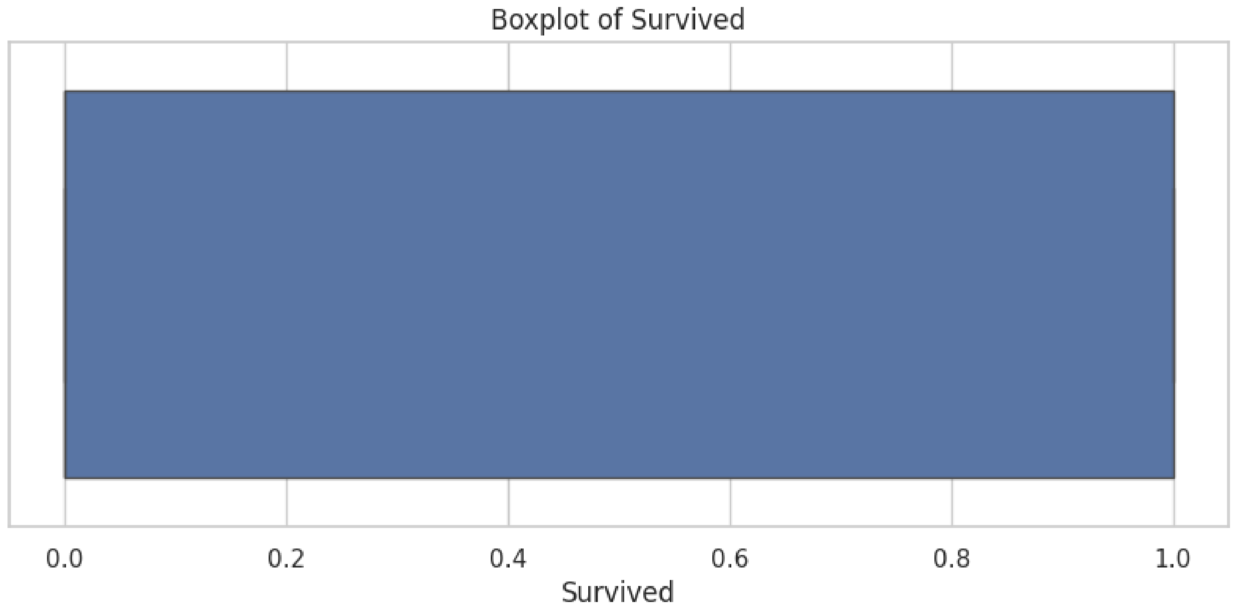


```
# 8. Histograms
df.hist(figsize=(15, 12), bins=30)
plt.suptitle("Histograms of Numeric Features", y=1.02)
plt.show()
```

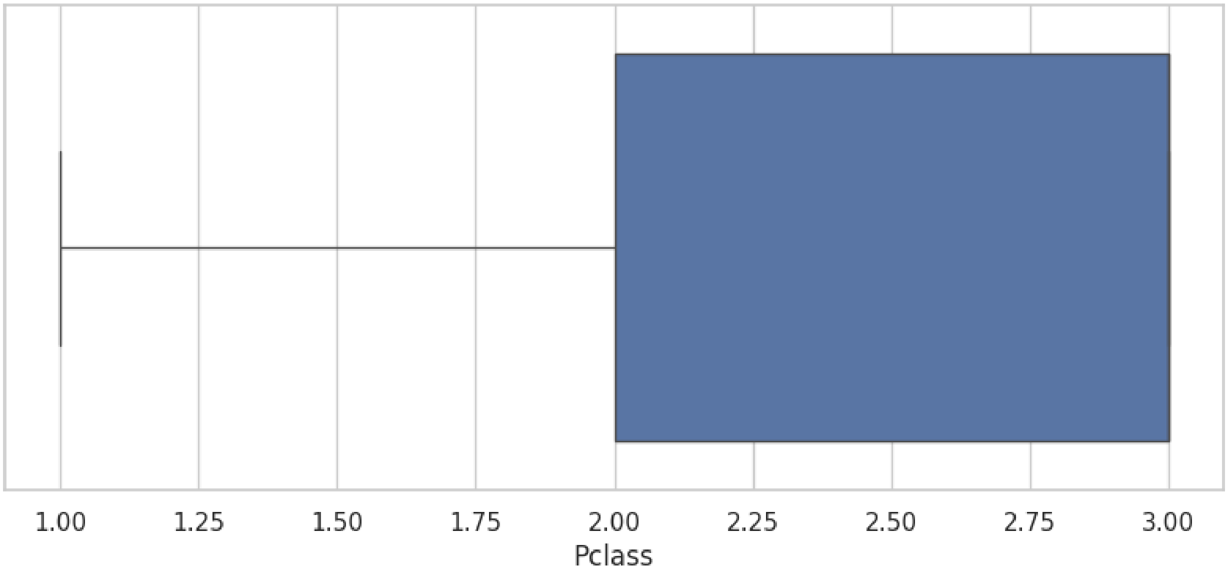


```
# 9. Boxplots for Outliers
for col in numeric_cols:
    plt.figure(figsize=(10, 4))
    sns.boxplot(x=df[col])
    plt.title(f"Boxplot of {col}")
    plt.show()
```

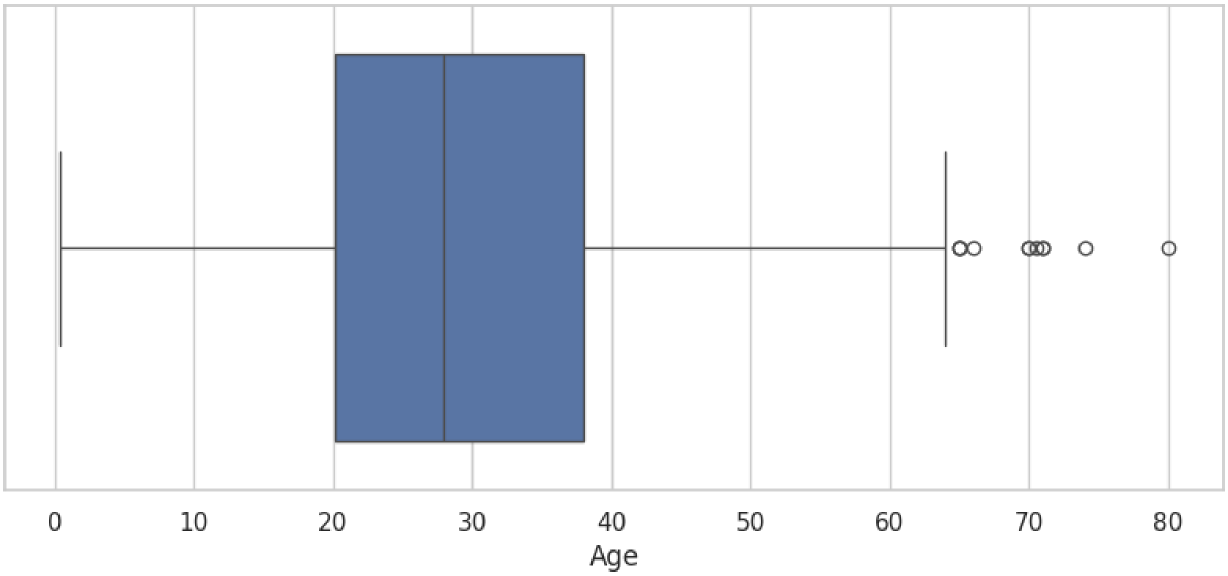

Boxplot of PassengerId

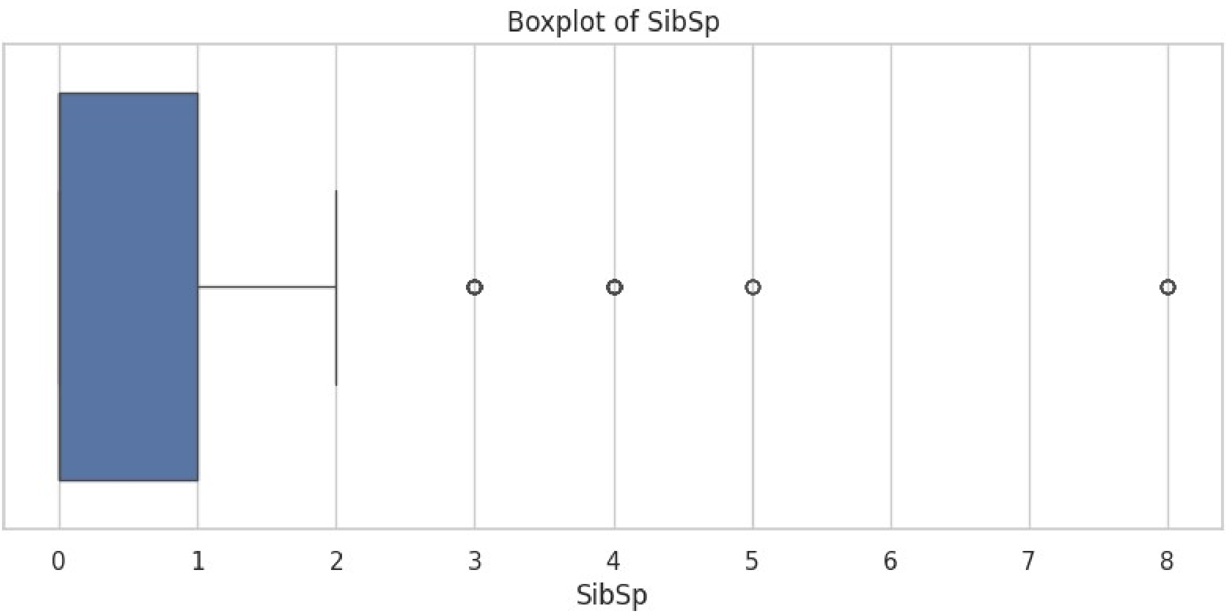

Boxplot of Survived

Boxplot of Pclass



Boxplot of Age




Boxplot of SibSp
0
1
2
3
4
5
6
7
8
SibSp