

NAME:K.R.LAHARI

## TASK -05

### CODE:

```
•[1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

sns.set(style='whitegrid')
plt.rcParams['figure.figsize'] = (10, 6)

df = pd.read_csv(r"C:\Users\julug\OneDrive\Documents\Titanic - Machine Learning from Disaster.csv")

print("First 5 rows of the dataset:")
print(df.head())

print("\nDataset Information:")
print(df.info())

print("\nSummary Statistics:")
print(df.describe())

print("\nMissing Values:")
print(df.isnull().sum())

print("\nValue Counts for Categorical Variables:")
for col in df.select_dtypes(include='object').columns:
    print(f"\n{col}: \n{df[col].value_counts()}")

plt.figure(figsize=(10, 8))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Correlation Heatmap")
plt.show()

sns.pairplot(df.select_dtypes(include='number'))
plt.suptitle("Pairwise Relationships", y=1.02)
plt.show()
```

```

sns.pairplot(df.select_dtypes(include='number'))
plt.suptitle("Pairwise Relationships", y=1.02)
plt.show()

df.hist(bins=30, figsize=(15, 10), color='skyblue', edgecolor='black')
plt.suptitle("Histograms of Numeric Features")
plt.tight_layout()
plt.show()

for col in df.select_dtypes(include='number').columns:
    sns.boxplot(x=df[col])
    plt.title(f"Boxplot of {col}")
    plt.show()

if 'feature1' in df.columns and 'feature2' in df.columns:
    sns.scatterplot(data=df, x='feature1', y='feature2', hue='target_column')
    plt.title("Scatterplot: feature1 vs feature2")
    plt.show()

print("\n🔍 Summary of Findings:")
print("- Dataset contains X rows and Y columns.")
print("- No missing values found / Missing values handled.")
print("- Correlation found between A and B.")
print("- Outliers detected in column XYZ.")
print("- Positive trend observed between Feature1 and Feature2.")

```

## OUTPUT:

First 5 rows of the dataset:

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	

	Name	Sex	Age	SibSp	\
0	Braund, Mr. Owen Harris	male	22.0	1	
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	
2	Heikkinen, Miss. Laina	female	26.0	0	
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	
4	Allen, Mr. William Henry	male	35.0	0	

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S

Dataset Information:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 891 entries, 0 to 890

Data columns (total 12 columns):

Dataset Information:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 891 entries, 0 to 890

Data columns (total 12 columns):

#	Column	Non-Null Count	Dtype
0	PassengerId	891 non-null	int64
1	Survived	891 non-null	int64
2	Pclass	891 non-null	int64
3	Name	891 non-null	object
4	Sex	891 non-null	object
5	Age	714 non-null	float64
6	SibSp	891 non-null	int64
7	Parch	891 non-null	int64
8	Ticket	891 non-null	object
9	Fare	891 non-null	float64
10	Cabin	204 non-null	object
11	Embarked	889 non-null	object

10 Cabin 204 non-null object

11 Embarked 889 non-null object

dtypes: float64(2), int64(5), object(5)

memory usage: 83.7+ KB

None

Summary Statistics:

	PassengerId	Survived	Pclass	Age	SibSp \
count	891.000000	891.000000	891.000000	714.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008
std	257.353842	0.486592	0.836071	14.526497	1.102743
min	1.000000	0.000000	1.000000	0.420000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000
50%	446.000000	0.000000	3.000000	28.000000	0.000000
75%	668.500000	1.000000	3.000000	38.000000	1.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000

	Parch	Fare
count	891.000000	891.000000
mean	0.381594	32.204208
std	0.806057	49.693429
min	0.000000	0.000000
25%	0.000000	7.910400
50%	0.000000	14.454200
75%	0.000000	31.000000
max	6.000000	512.329200

#### Missing Values:

PassengerId	0
Survived	0
Pclass	0
Name	0
Sex	0
Age	177
SibSp	0

Sex	0
Age	177
SibSp	0
Parch	0
Ticket	0
Fare	0
Cabin	687
Embarked	2

dtype: int64

#### Value Counts for Categorical Variables:

##### Name:

Name	
Dooley, Mr. Patrick	1
Braund, Mr. Owen Harris	1
Cumings, Mrs. John Bradley (Florence Briggs Thayer)	1
Heikkinen, Miss. Laina	1

```

name .
Name
Dooley, Mr. Patrick 1
Braund, Mr. Owen Harris 1
Cumings, Mrs. John Bradley (Florence Briggs Thayer) 1
Heikkinen, Miss. Laina 1
Futrelle, Mrs. Jacques Heath (Lily May Peel) 1
..
Hewlett, Mrs. (Mary D Kingcome) 1
Vestrom, Miss. Hulda Amanda Adolfina 1
Andersson, Mr. Anders Johan 1
Saundercock, Mr. William Henry 1
Bonnell, Miss. Elizabeth 1
Name: count, Length: 891, dtype: int64

```

```

Sex:
Sex
male 577

```

```

Sex:
Sex
male 577
female 314
Name: count, dtype: int64

```

```

Ticket:
Ticket
347082 7
1601 7
CA. 2343 7
3101295 6
CA 2144 6
..
PC 17590 1
17463 1
330877 1
373450 1

```

```
373450      1
STON/O2. 3101282      1
Name: count, Length: 681, dtype: int64
```

Cabin:

Cabin

```
G6      4
C23 C25 C27      4
B96 B98      4
F2      3
D      3
..
E17      1
A24      1
C50      1
B42      1
C148      1
```

---

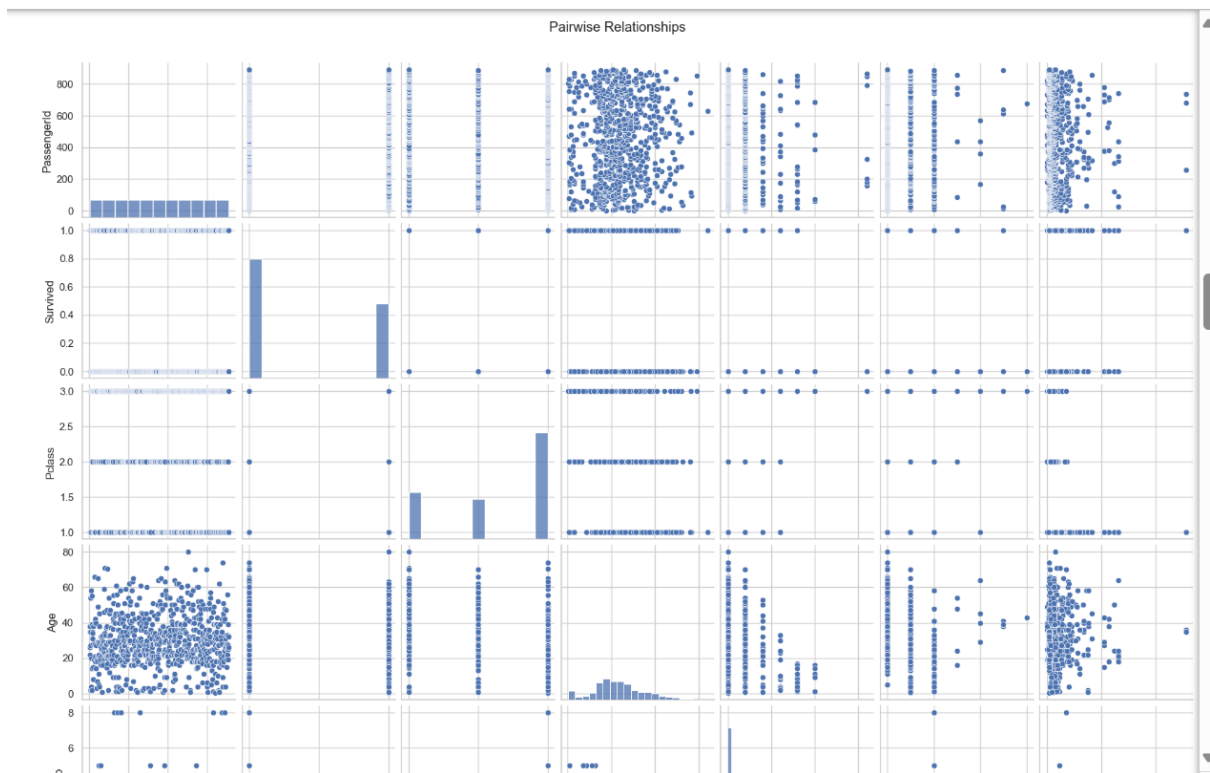
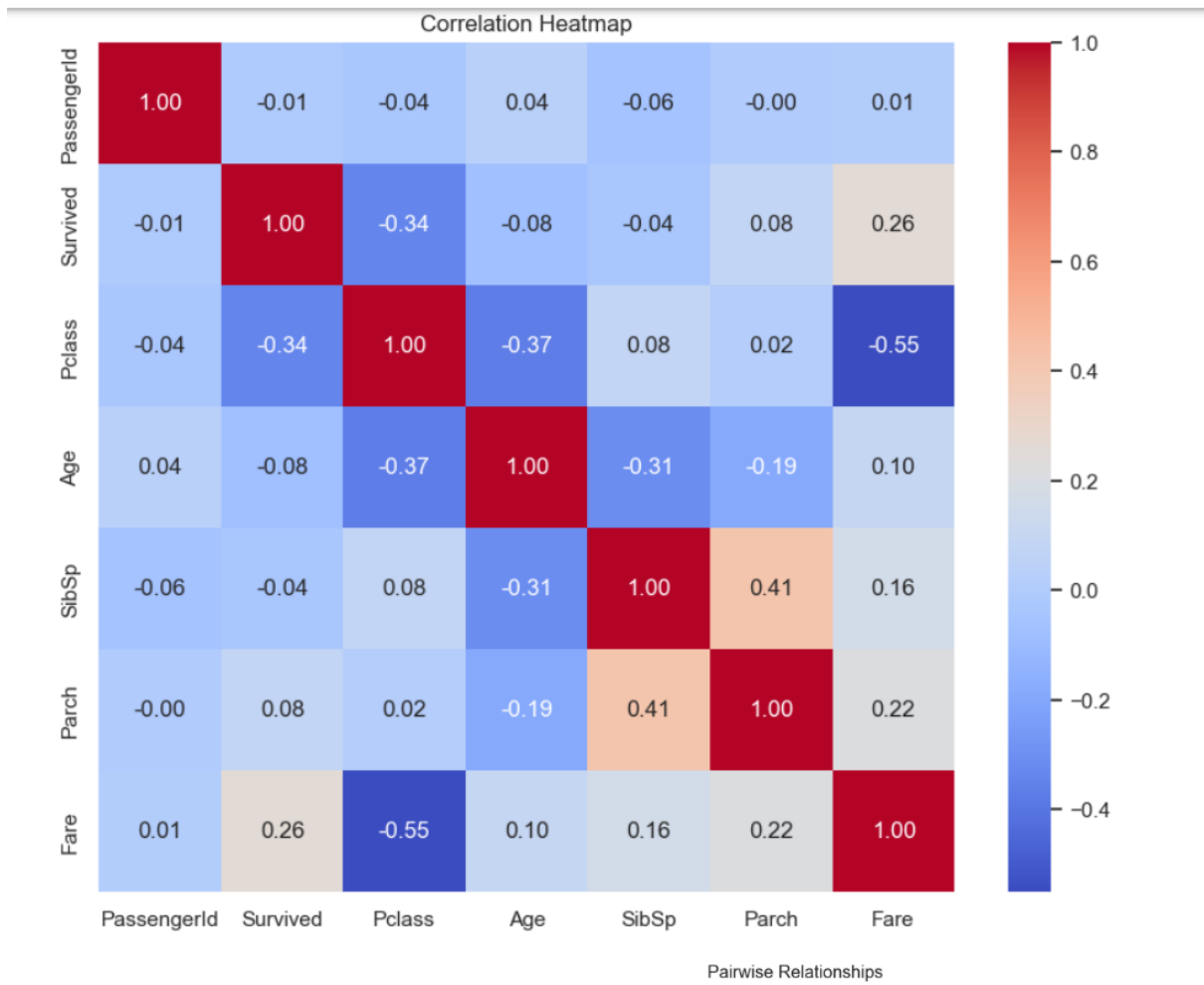
```
C148      1
Name: count, Length: 147, dtype: int64
```

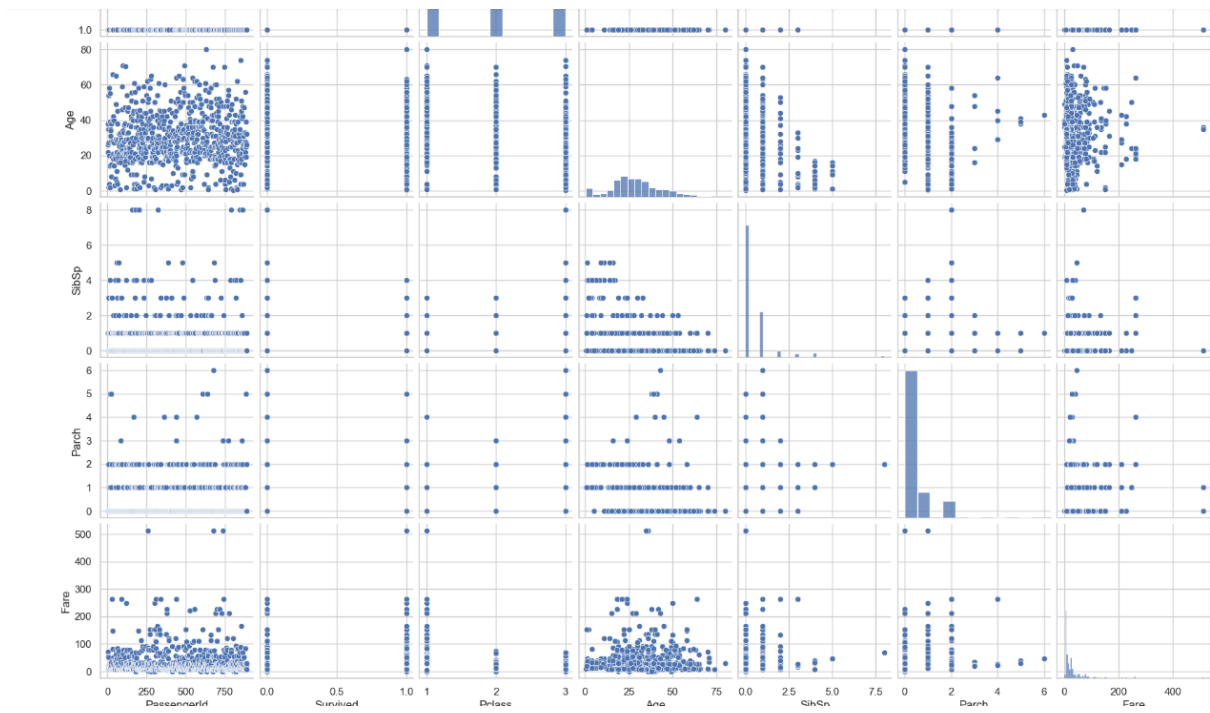
Embarked:

Embarked

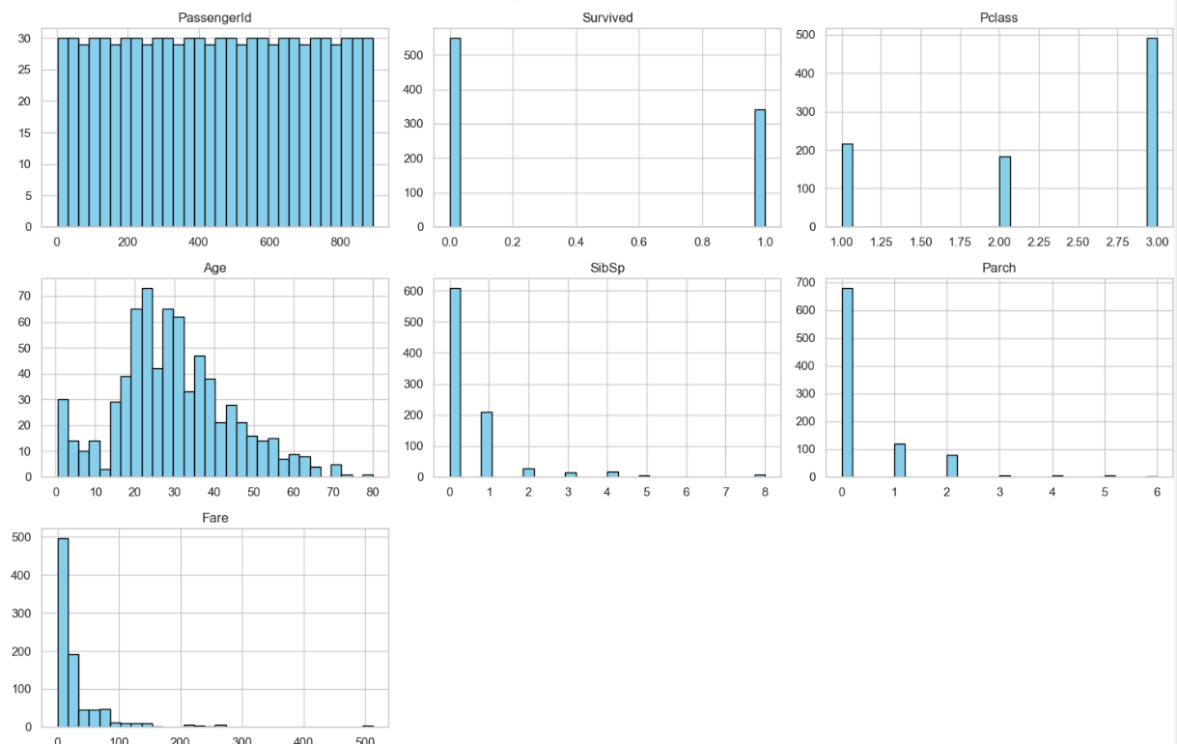
```
S      644
C      168
Q       77
```

```
Name: count, dtype: int64
```

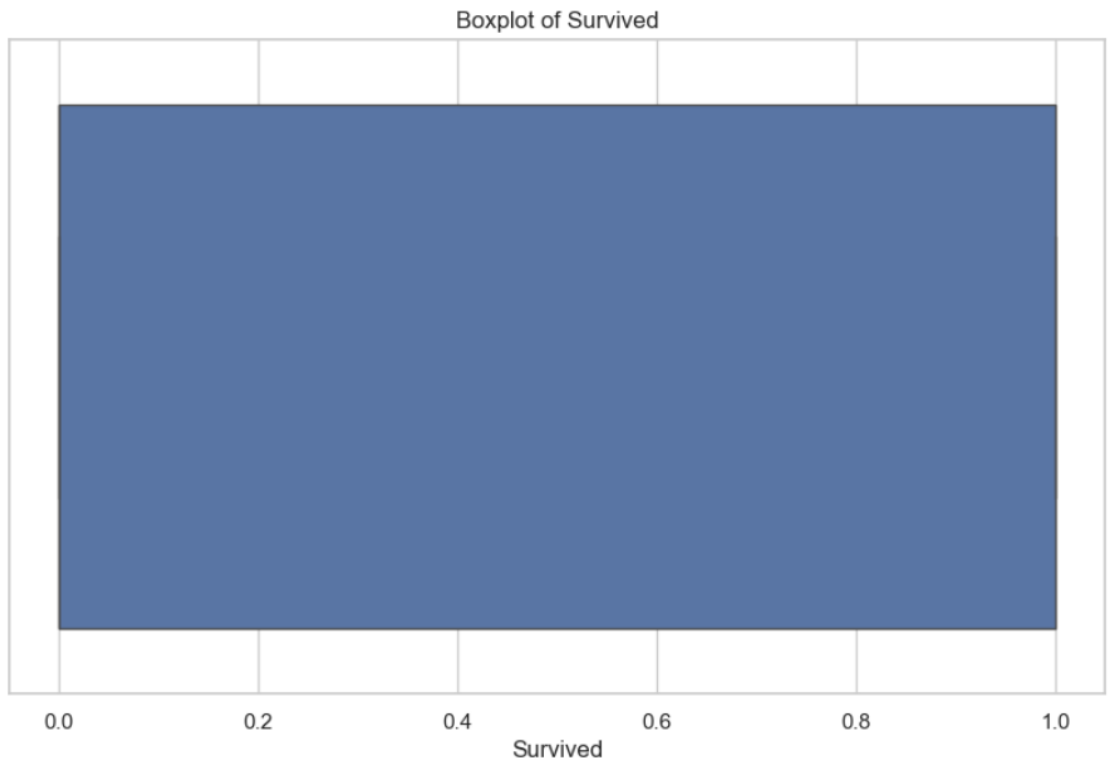
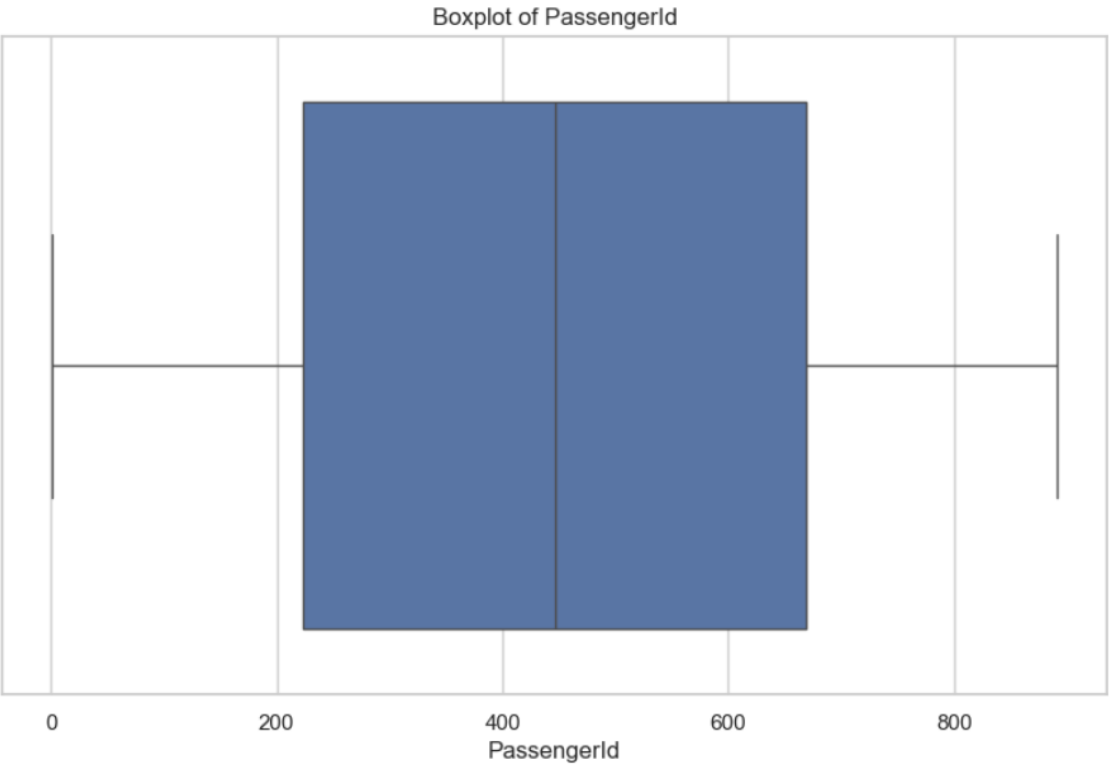


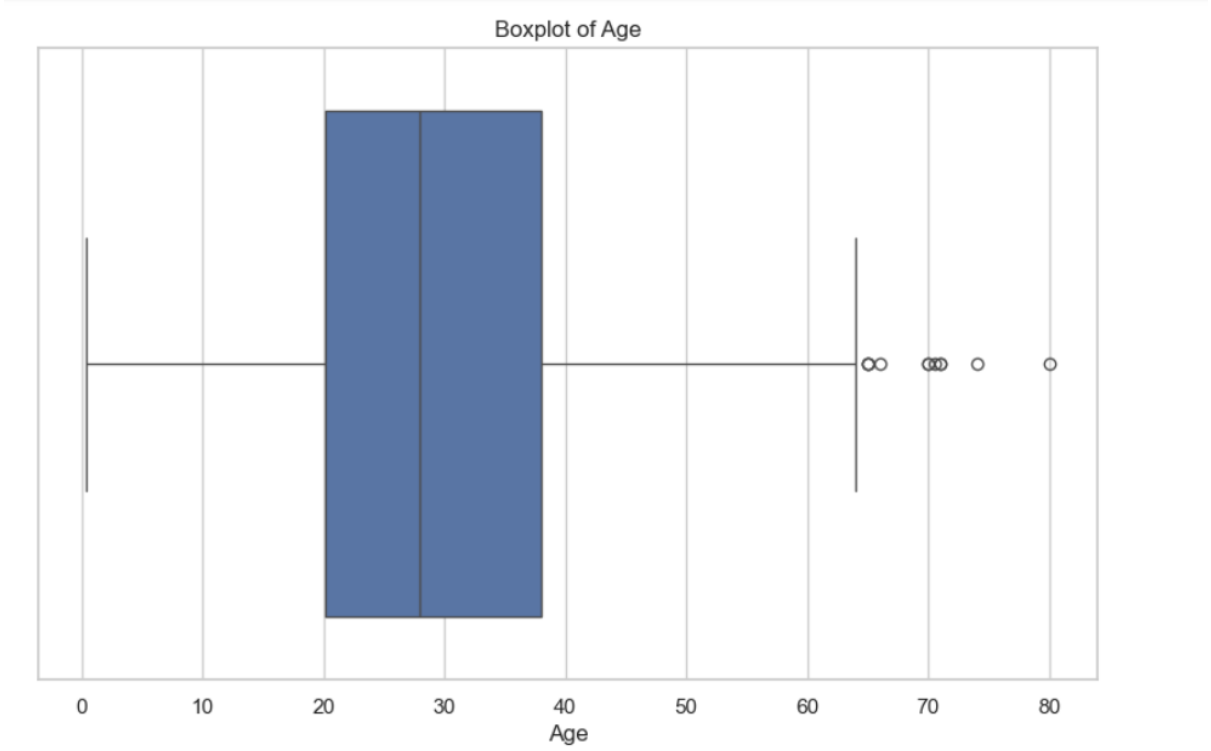
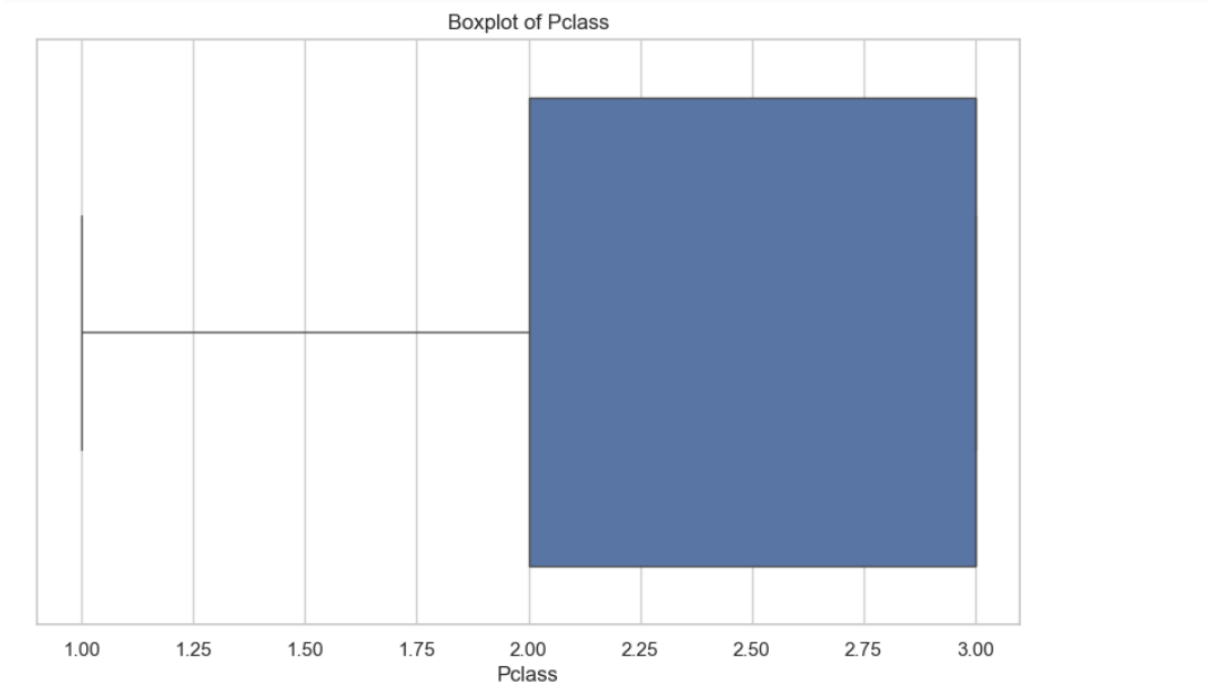


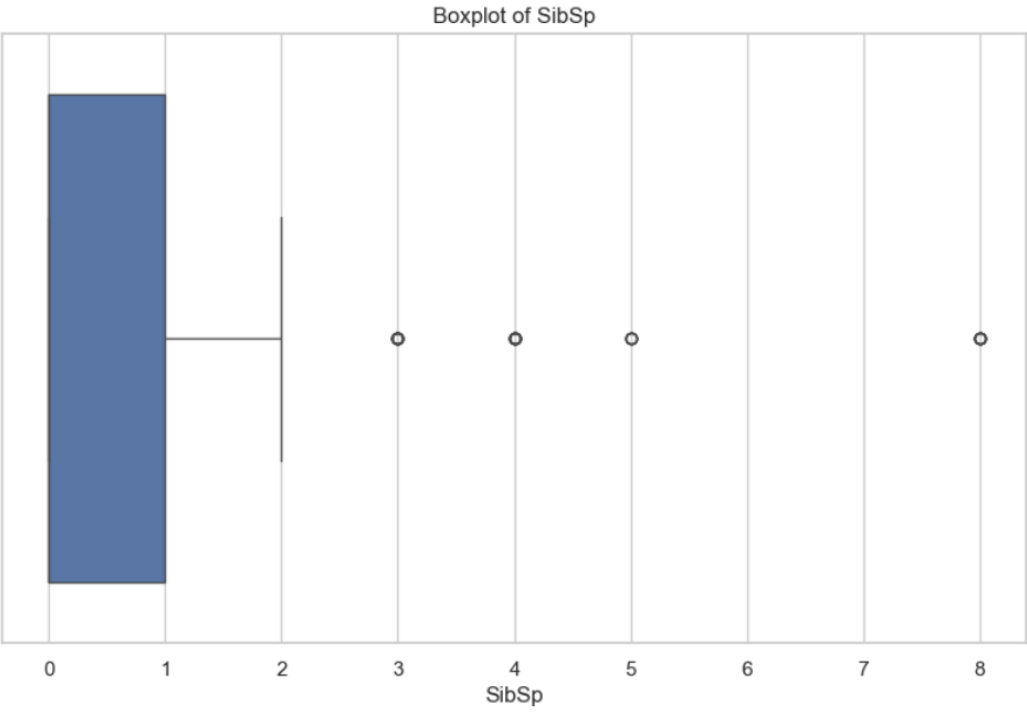
Histograms of Numeric Features





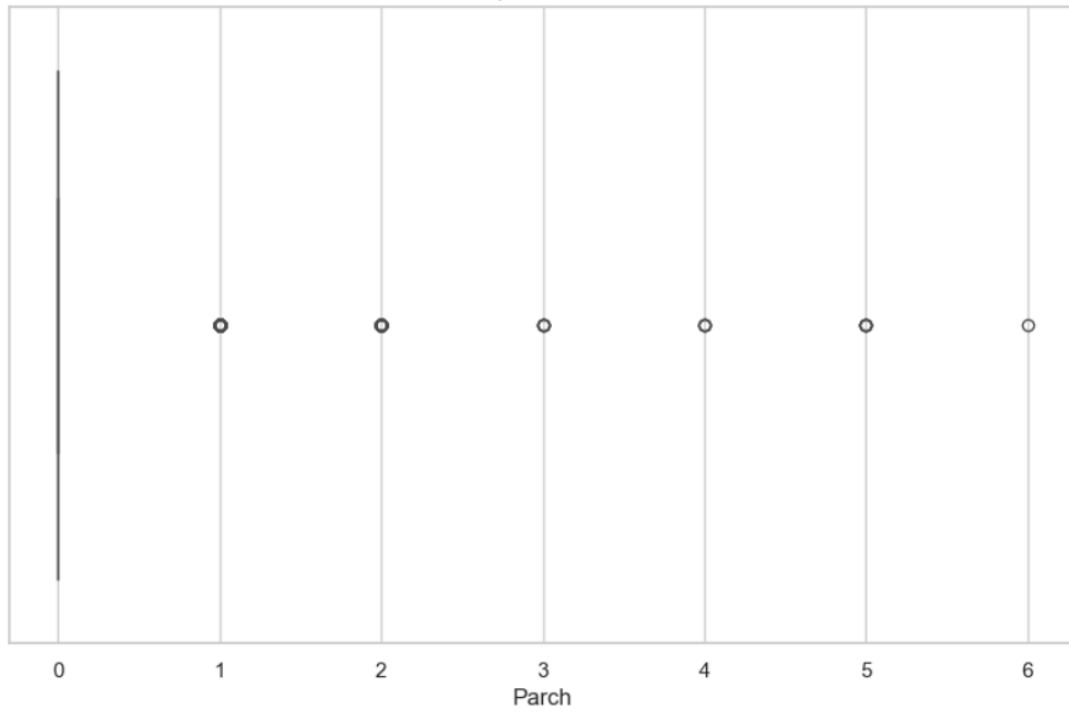




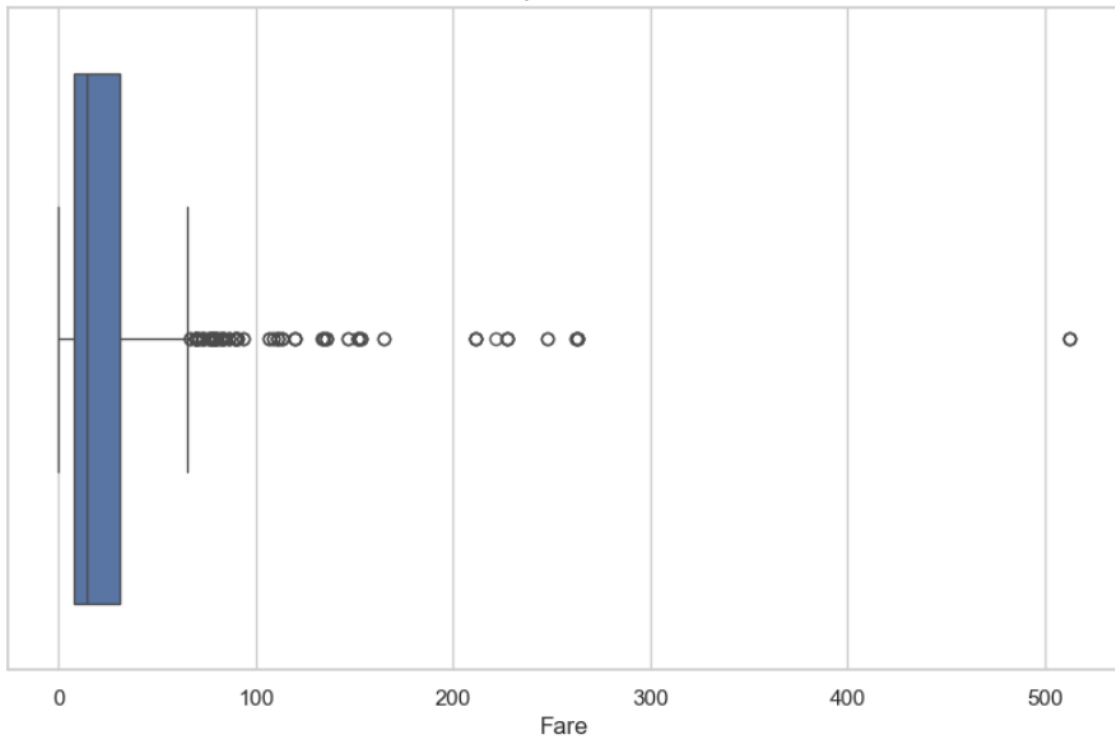


SibSp

Boxplot of Parch



Boxplot of Fare



#### Summary of Findings:

- Dataset contains X rows and Y columns.
- No missing values found / Missing values handled.
- Correlation found between A and B.
- Outliers detected in column XYZ.
- Positive trend observed between Feature1 and Feature2.