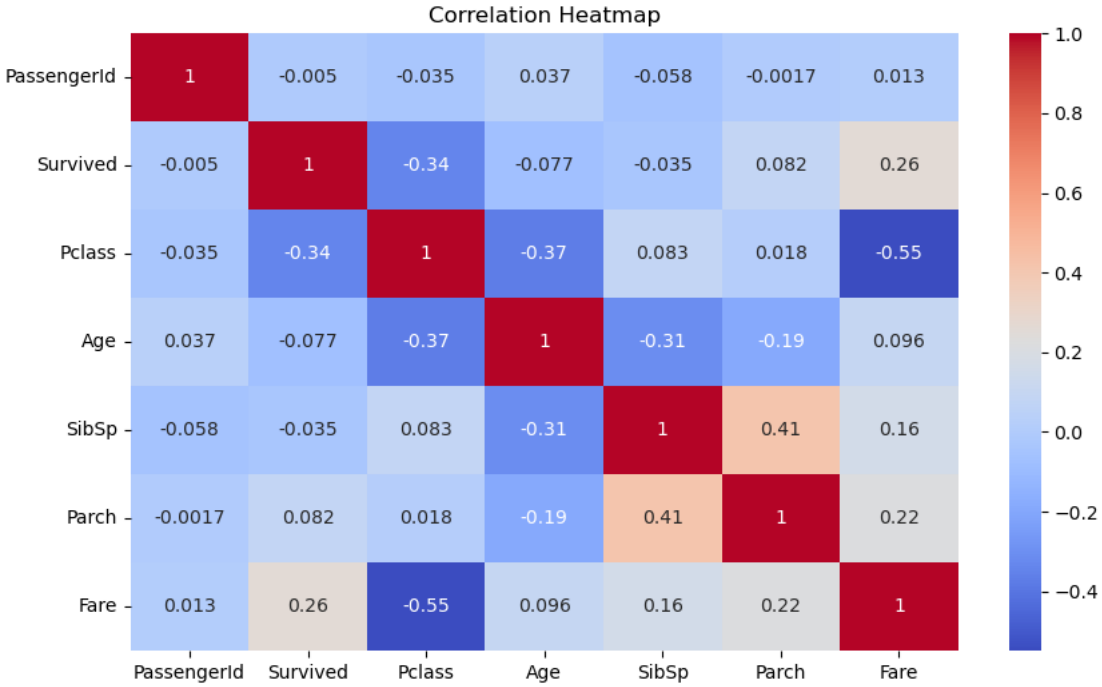
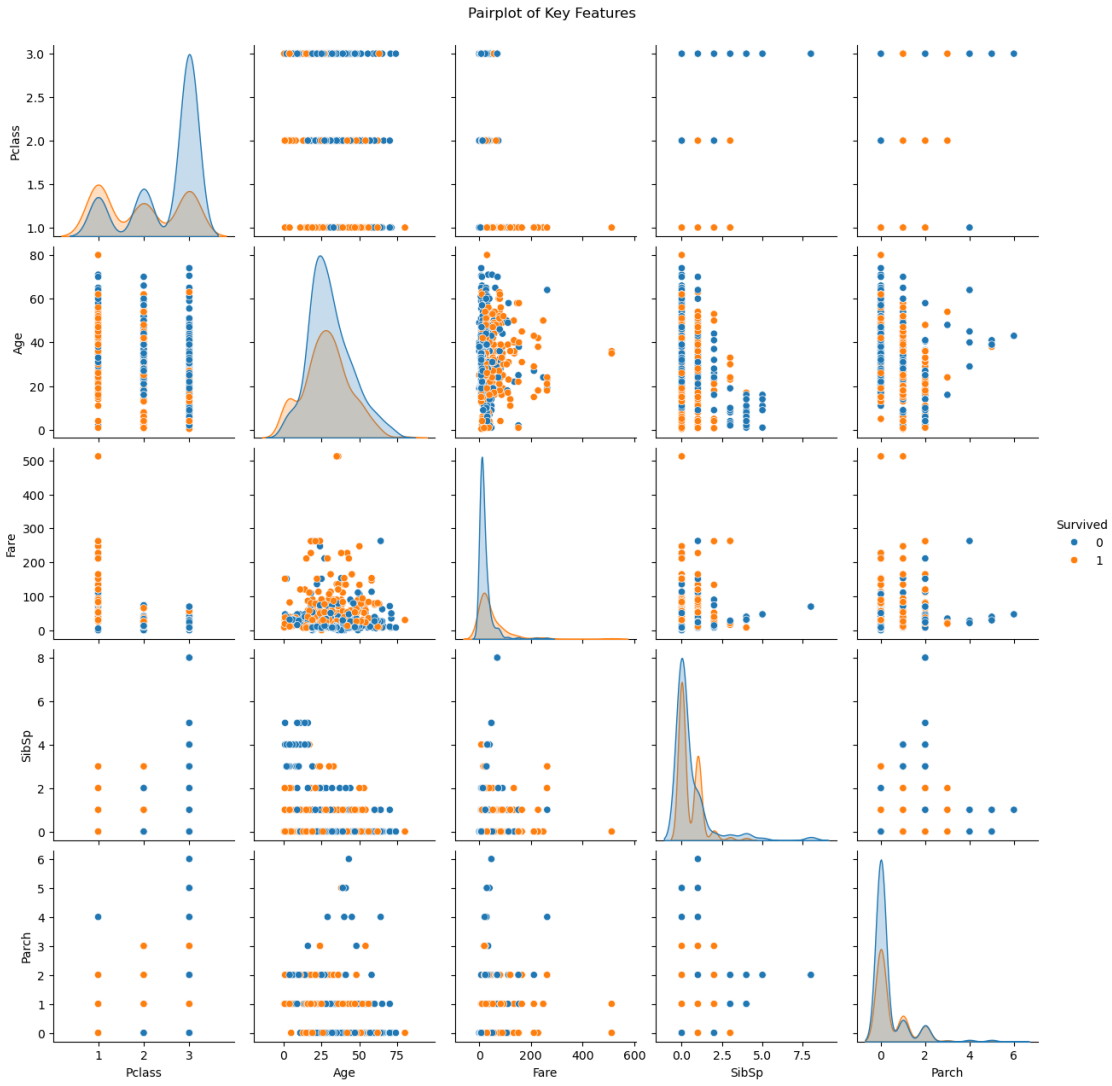
```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv(r'C:\Users\K.s.Santhos\Downloads\train.csv')

print("Data Info:")
print(df.info())

print("\nSummary Statistics:")
print(df.describe(include='all'))

print("\nSurvived value counts:")
print(df['Survived'].value_counts())

print("\nEmbarked value counts:")
print(df['Embarked'].value_counts())

print("\nPclass value counts:")
print(df['Pclass'].value_counts())


plt.figure(figsize=(10, 6))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm')
plt.title("Correlation Heatmap")
plt.show()


sns.pairplot(df[['Survived', 'Pclass', 'Age', 'Fare', 'SibSp', 'Parch']],
hue='Survived')
plt.suptitle("Pairplot of Key Features", y=1.02)
plt.show()


df.hist(column=['Age', 'Fare'], bins=30, figsize=(10, 5), color='skyblue')
plt.suptitle("Histograms of Age and Fare")
plt.show()


plt.figure(figsize=(10, 5))
sns.boxplot(x='Pclass', y='Fare', data=df)
plt.title("Boxplot: Fare vs Passenger Class")
plt.show()

plt.figure(figsize=(10, 5))
sns.boxplot(x='Sex', y='Age', data=df)
plt.title("Boxplot: Age vs Sex")
plt.show()
```

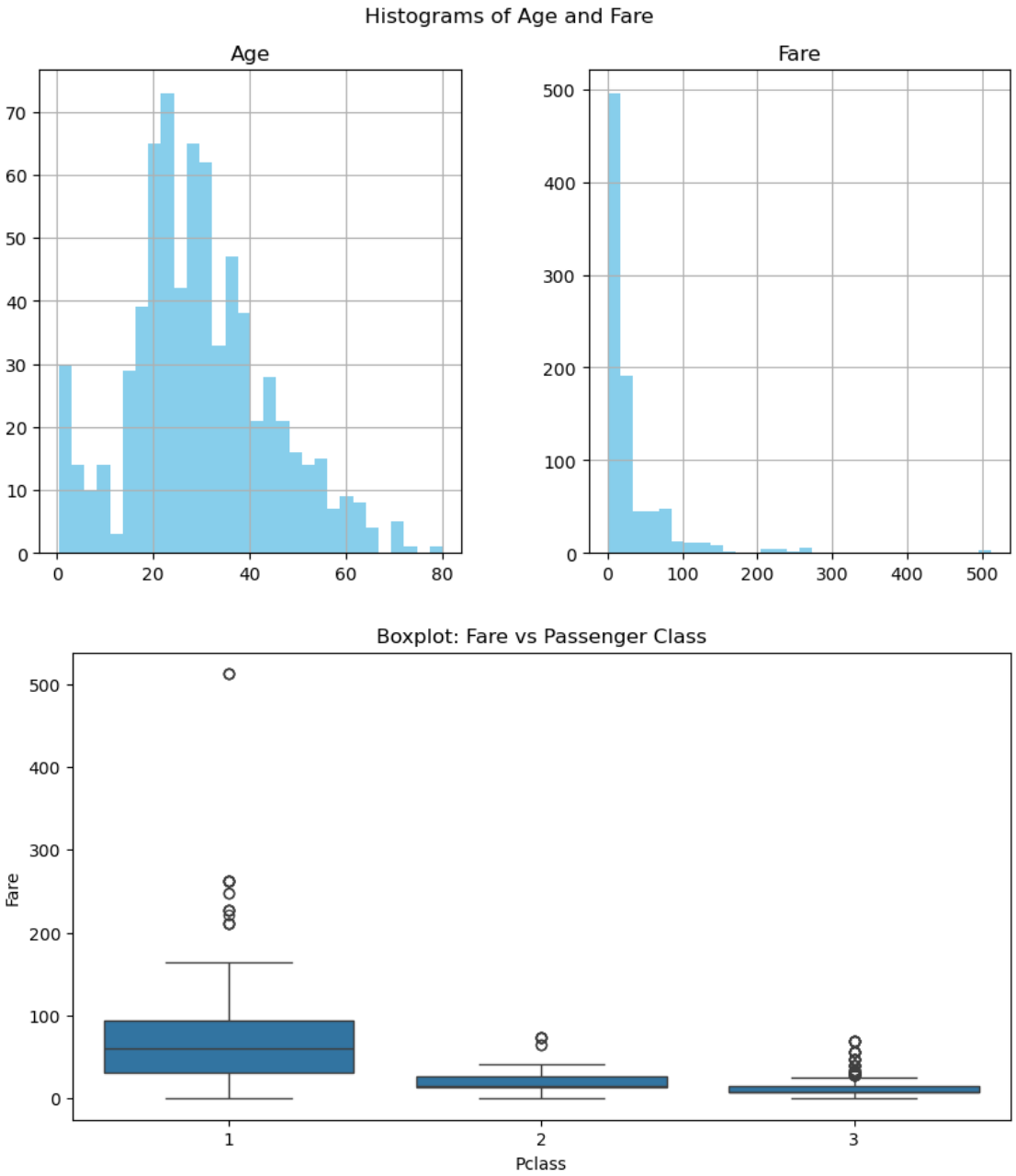
```
plt.figure(figsize=(10, 5))
sns.scatterplot(x='Age', y='Fare', hue='Survived', data=df)
plt.title("Scatterplot: Age vs Fare by Survival")
plt.show()


sns.countplot(x='Survived', hue='Sex', data=df)
plt.title("Survival Count by Sex")
plt.show()

sns.countplot(x='Embarked', hue='Survived', data=df)
plt.title("Survival by Embarkation Port")
plt.show()


plt.figure(figsize=(10, 5))
sns.heatmap(df.isnull(), cbar=False, cmap='viridis')
plt.title("Missing Values Heatmap")
plt.show()
```

```
Data Info:
<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
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
None

Summary 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  Dooley, Mr. Patrick  male
freq            NaN         NaN         NaN                    1   577
```

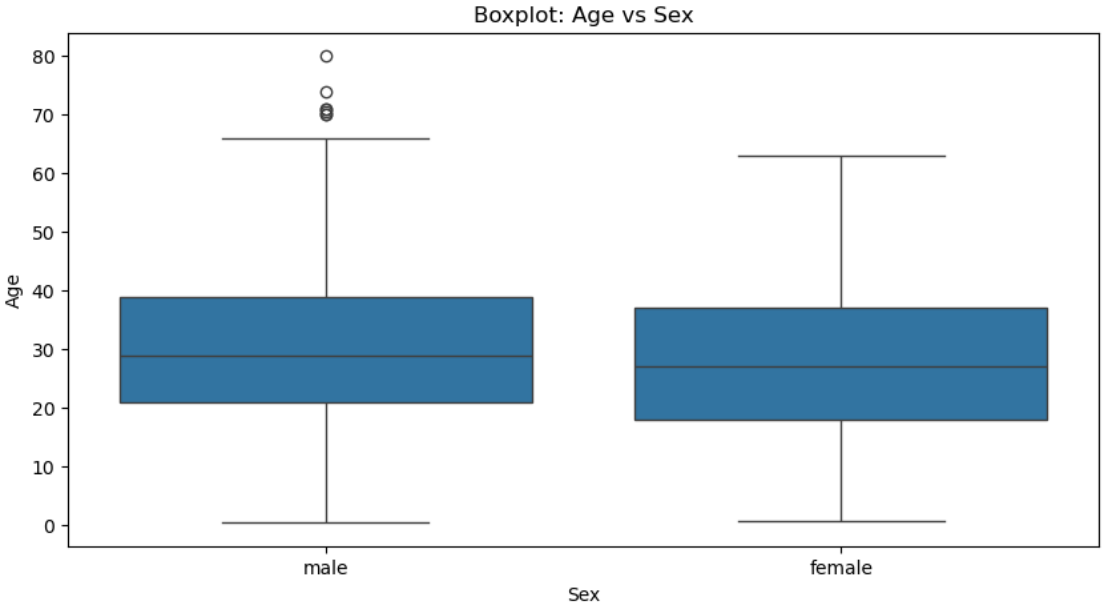
```
mean     446.000000    0.383838    2.308642                NaN   NaN
std      257.353842    0.486592    0.836071                NaN   NaN
min        1.000000    0.000000    1.000000                NaN   NaN
25%      223.500000    0.000000    2.000000                NaN   NaN
50%      446.000000    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 Embarked
count   714.000000  891.000000  891.000000     891  891.000000   204      889
unique         NaN         NaN         NaN     681         NaN   147        3
top            NaN         NaN         NaN  347082         NaN    G6        S
freq           NaN         NaN         NaN       7         NaN     4      644
mean     29.699118    0.523008    0.381594     NaN   32.204208   NaN      NaN
std      14.526497    1.102743    0.806057     NaN   49.693429   NaN      NaN
min       0.420000    0.000000    0.000000     NaN    0.000000   NaN      NaN
25%      20.125000    0.000000    0.000000     NaN    7.910400   NaN      NaN
50%      28.000000    0.000000    0.000000     NaN   14.454200   NaN      NaN
75%      38.000000    1.000000    0.000000     NaN   31.000000   NaN      NaN
max      80.000000    8.000000    6.000000     NaN  512.329200   NaN      NaN

Survived value counts:
Survived
0    549
1    342
Name: count, dtype: int64

Embarked value counts:
Embarked
S    644
C    168
Q     77
Name: count, dtype: int64

Pclass value counts:
Pclass
3    491
1    216
2    184
Name: count, dtype: int64
```

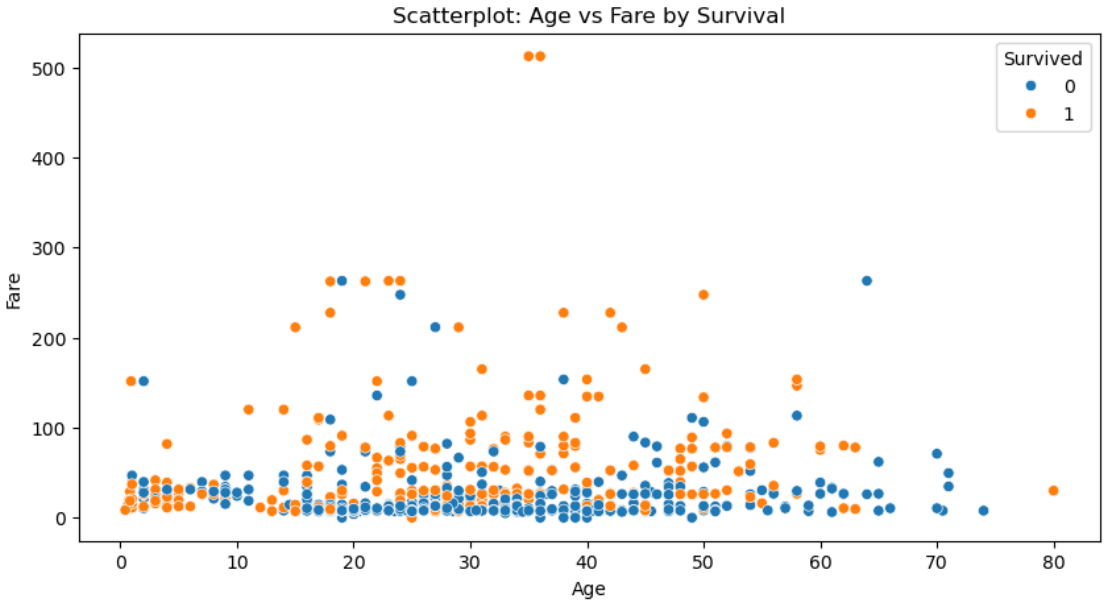

Correlation Heatmap
PassengerId
Survived
Pclass
Age
SibSp
Parch
Fare
1 -0.005 -0.035 0.037 -0.058 -0.0017 0.013
-0.005 1 -0.34 -0.077 -0.035 0.082 0.26
-0.035 -0.34 1 -0.37 0.083 0.018 -0.55
0.037 -0.077 -0.37 1 -0.31 -0.19 0.096
-0.058 -0.035 0.083 -0.31 1 0.41 0.16
-0.0017 0.082 0.018 -0.19 0.41 1 0.22
0.013 0.26 -0.55 0.096 0.16 0.22 1
PassengerId
Survived
Pclass
Age
SibSp
Parch
Fare
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

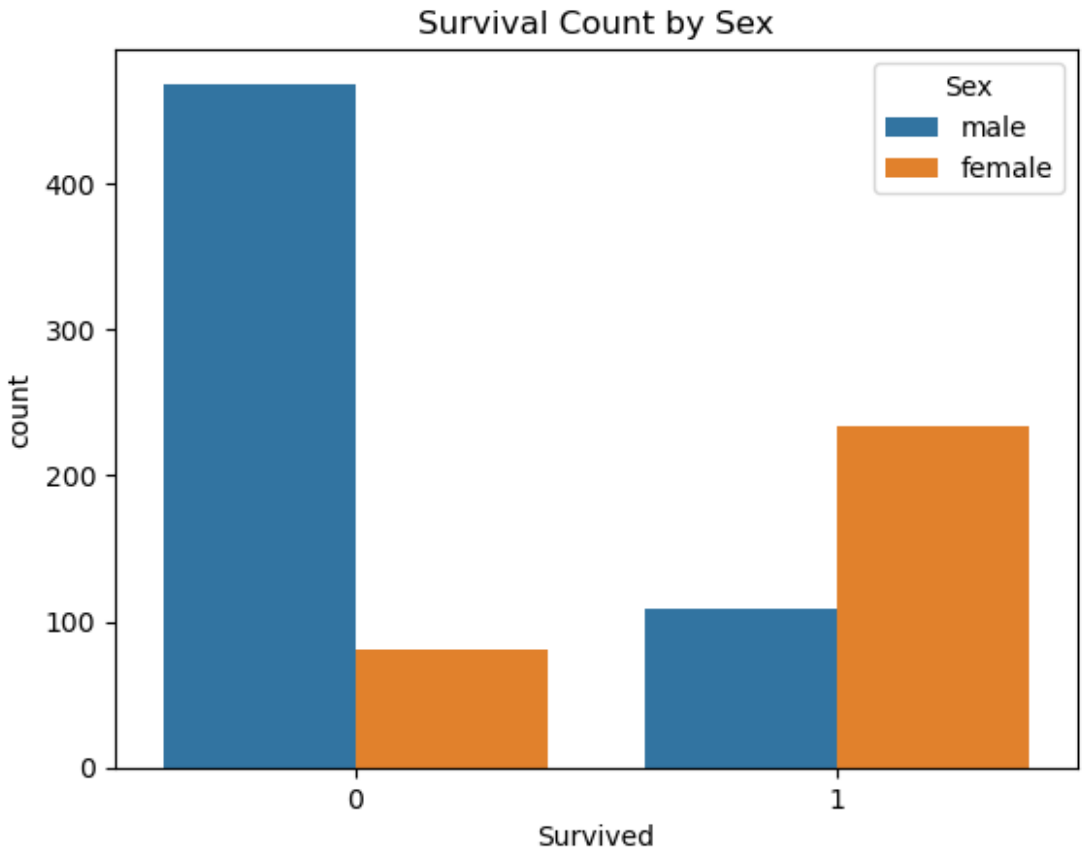

Pairplot of Key Features
Pclass
Age
Fare
SibSp
Parch
Survived
0
1

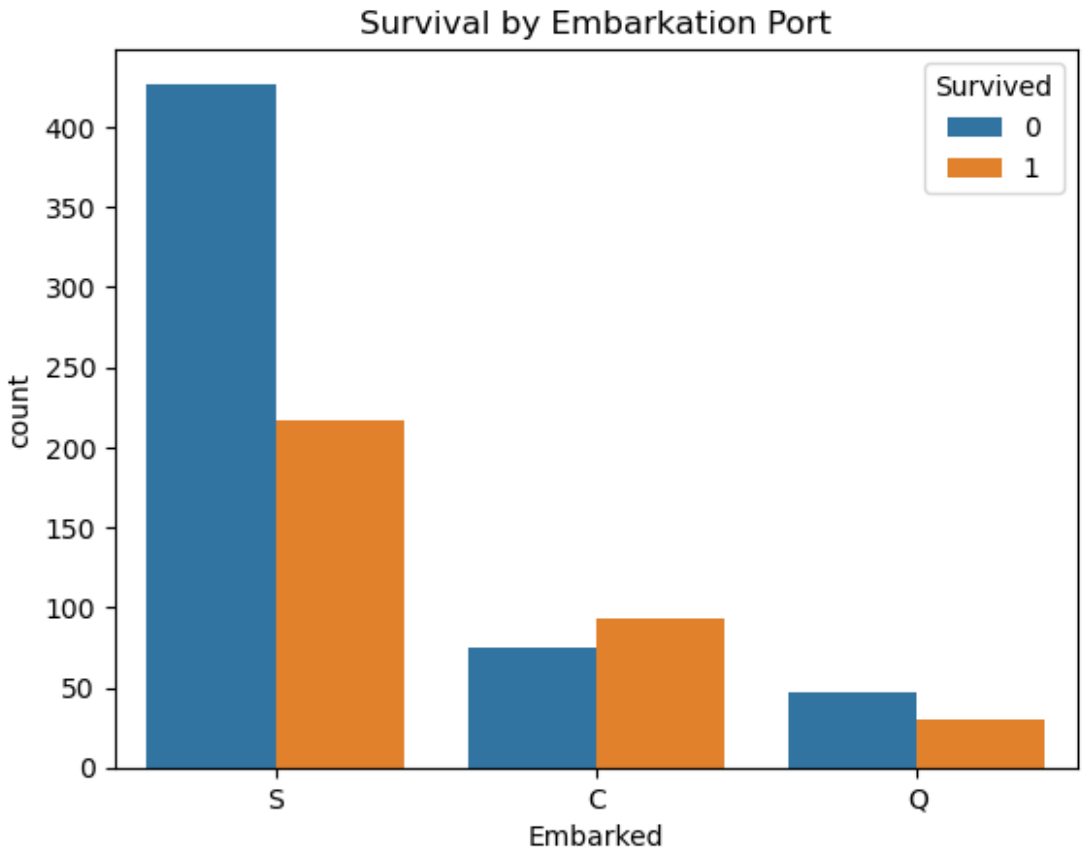

Histograms of Age and Fare
Age
Fare
0
10
20
30
40
50
60
70
0
20
40
60
80
0
100
200
300
400
500
0
100
200
300
400
500

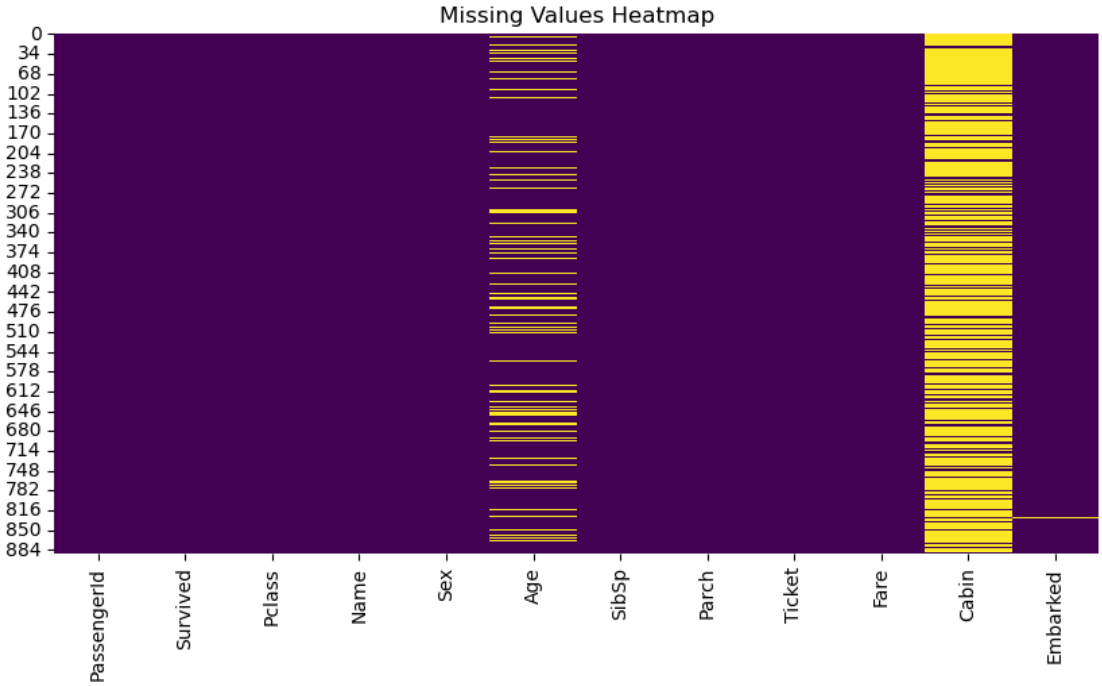
Boxplot: Fare vs Passenger Class
Fare
Pclass
0
100
200
300
400
500
1
2
3


Boxplot: Age vs Sex
Age
Sex
male
female
0
10
20
30
40
50
60
70
80


Scatterplot: Age vs Fare by Survival
Survived
0
1
Fare
Age
0
100
200
300
400
500
0
10
20
30
40
50
60
70
80


Survival Count by Sex
Sex
male
female
400
300
200
100
0
count
0
1
Survived


Survival by Embarkation Port
Survived
0
1
count
400
350
300
250
200
150
100
50
0
S
C
Q
Embarked


Missing Values Heatmap
0
34
68
102
136
170
204
238
272
306
340
374
408
442
476
510
544
578
612
646
680
714
748
782
816
850
884
PassengerId
Survived
Pclass
Name
Sex
Age
SibSp
Parch
Ticket
Fare
Cabin
Embarked