```
!pip install pandas numpy matplotlib seaborn
```

```
Requirement already satisfied: pandas in c:\users\gv862\appdata\local\packages\pythonsoftwarefoundation.python.3.11_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (2.2.3)
Requirement already satisfied: numpy in c:\users\gv862\appdata\local\packages\pythonsoftwarefoundation.python.3.11_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (2.2.4)
Requirement already satisfied: matplotlib in c:\users\gv862\appdata\local\packages\pythonsoftwarefoundation.python.3.11_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (3.10.1)
Requirement already satisfied: seaborn in c:\users\gv862\appdata\local\packages\pythonsoftwarefoundation.python.3.11_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (0.13.2)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\gv862\appdata\local\packages\pythonsoftwarefoundation.python.3.11_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\gv862\appdata\local\packages\pythonsoftwarefoundation.python.3.11_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in c:\users\gv862\appdata\local\packages\pythonsoftwarefoundation.python.3.11_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from pandas) (2025.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\gv862\appdata\local\packages\pythonsoftwarefoundation.python.3.11_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from matplotlib) (1.3.1)
Requirement already satisfied: cycler>=0.10 in c:\users\gv862\appdata\local\packages\pythonsoftwarefoundation.python.3.11_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\gv862\appdata\local\packages\pythonsoftwarefoundation.python.3.11_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from matplotlib) (4.57.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\gv862\appdata\local\packages\pythonsoftwarefoundation.python.3.11_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from matplotlib) (1.4.8)
Requirement already satisfied: packaging>=20.0 in c:\users\gv862\appdata\local\packages\pythonsoftwarefoundation.python.3.11_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from matplotlib) (24.2)
Requirement already satisfied: pillow>=8 in c:\users\gv862\appdata\local\packages\pythonsoftwarefoundation.python.3.11_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from matplotlib) (11.2.1)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\gv862\appdata\local\packages\pythonsoftwarefoundation.python.3.11_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from matplotlib) (3.2.3)
Requirement already satisfied: six>=1.5 in c:\users\gv862\appdata\local\packages\pythonsoftwarefoundation.python.3.11_qbz5n2kfra8p0\localcache\local-packages\python311\site-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
[notice] A new release of pip is available: 24.0 -> 25.0.1
[notice] To update, run: C:\Users\GV862\AppData\Local\Microsoft\WindowsApps\PythonSoftwareFoundation.Python.3.11_qbz5n2kfra8p0\python.exe -m pip install --upgrade pip
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# For better visual aesthetics
sns.set(style="whitegrid")
```

```
import os
print(os.getcwd())
```

```
c:\Users\GV862
```

```
train = pd.read_csv(r"C:\Users\GV862\OneDrive\Desktop\coding\python\train.csv")
test = pd.read_csv(r"C:\Users\GV862\OneDrive\Desktop\coding\python\test.csv")
```

```
import os
print(os.getcwd())
```

```
c:\Users\GV862
```

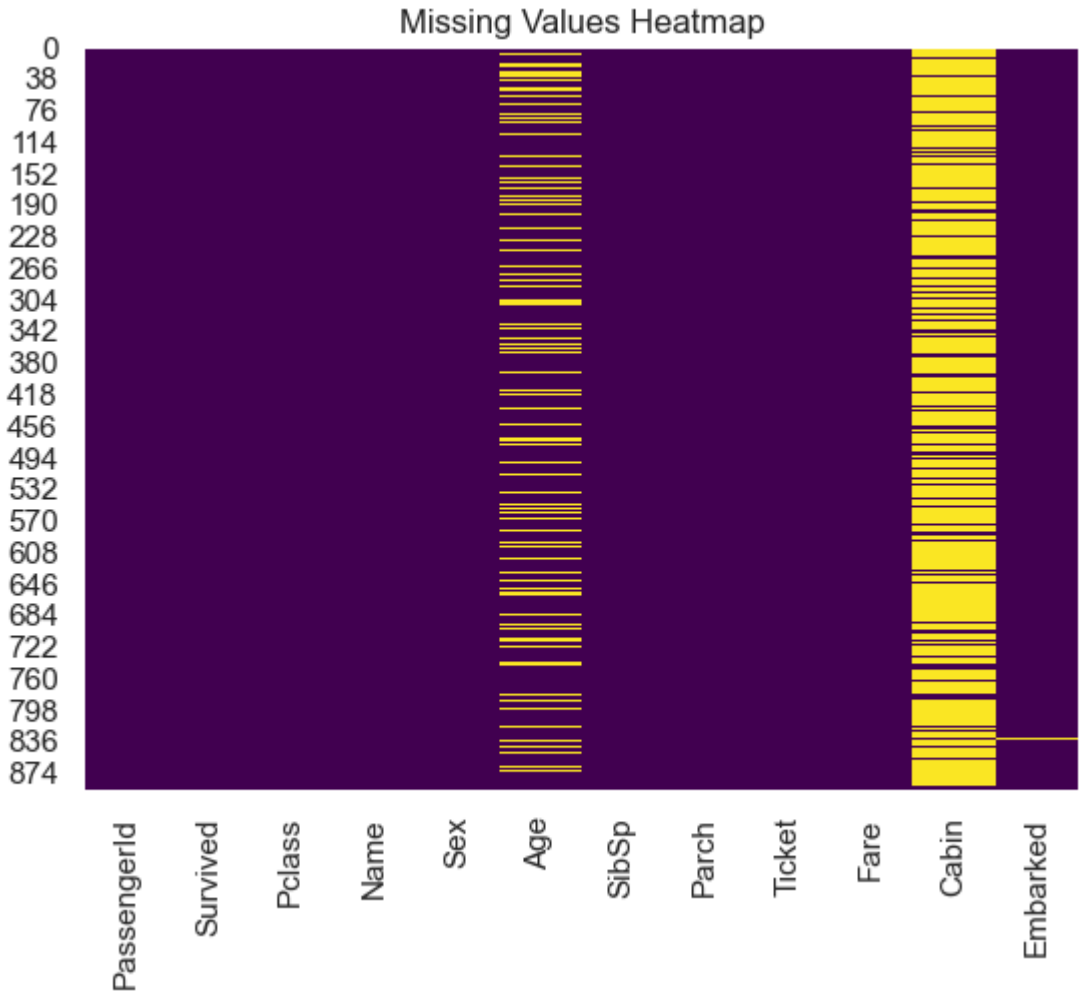
```
train = pd.read_csv(r"C:\Users\GV862\OneDrive\Desktop\coding\python\train.csv")
test = pd.read_csv(r"C:\Users\GV862\OneDrive\Desktop\coding\python\test.csv")
```

```
# Shape of the dataset
print(train.shape)
print(test.shape)

# Preview
train.head()

# Info about data types and missing values
train.info()

# Statistical Summary
train.describe()

# Value counts of categorical features
train['Survived'].value_counts()
train['Pclass'].value_counts()
train['Sex'].value_counts()
```

```
(891, 12)
(418, 11)
<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
```

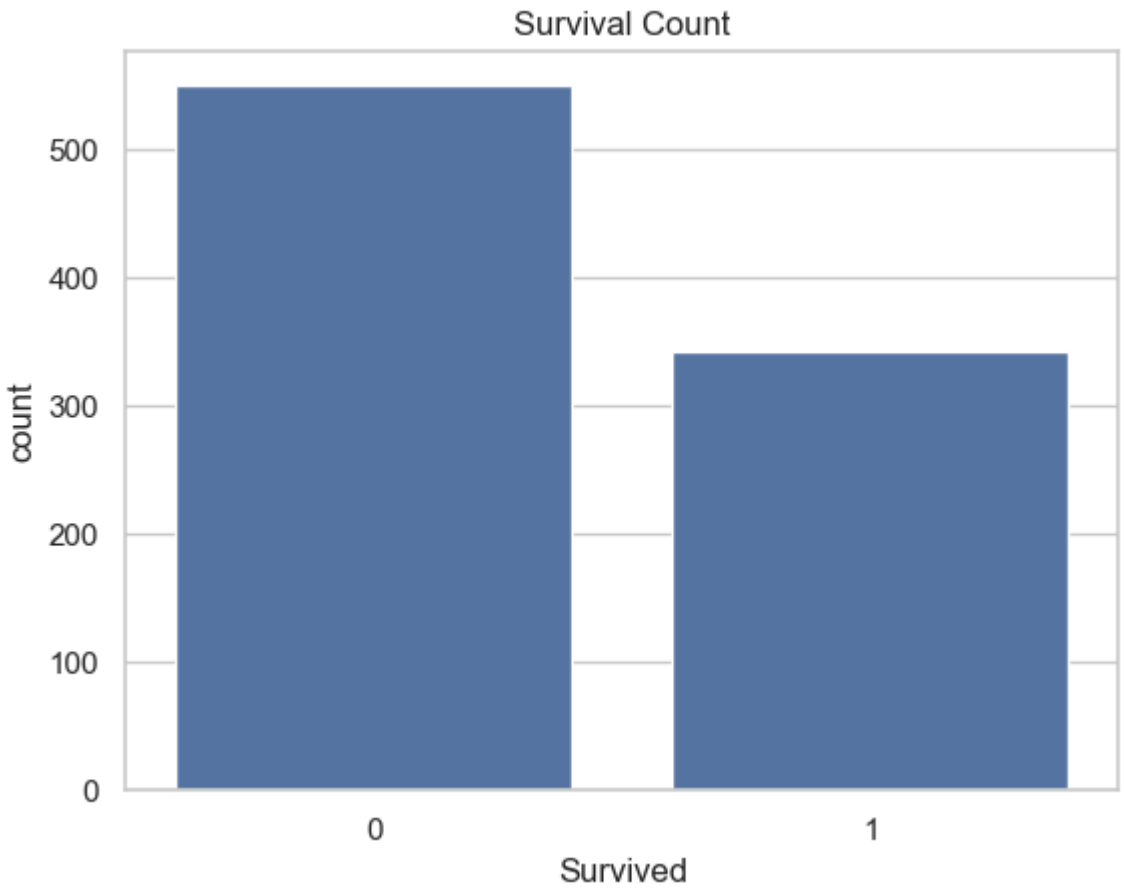
```
Sex
male      577
female    314
Name: count, dtype: int64
```

```
# Total missing values
train.isnull().sum()
sns.heatmap(train.isnull(), cbar=False, cmap="viridis")
plt.title("Missing Values Heatmap")
plt.show()
```



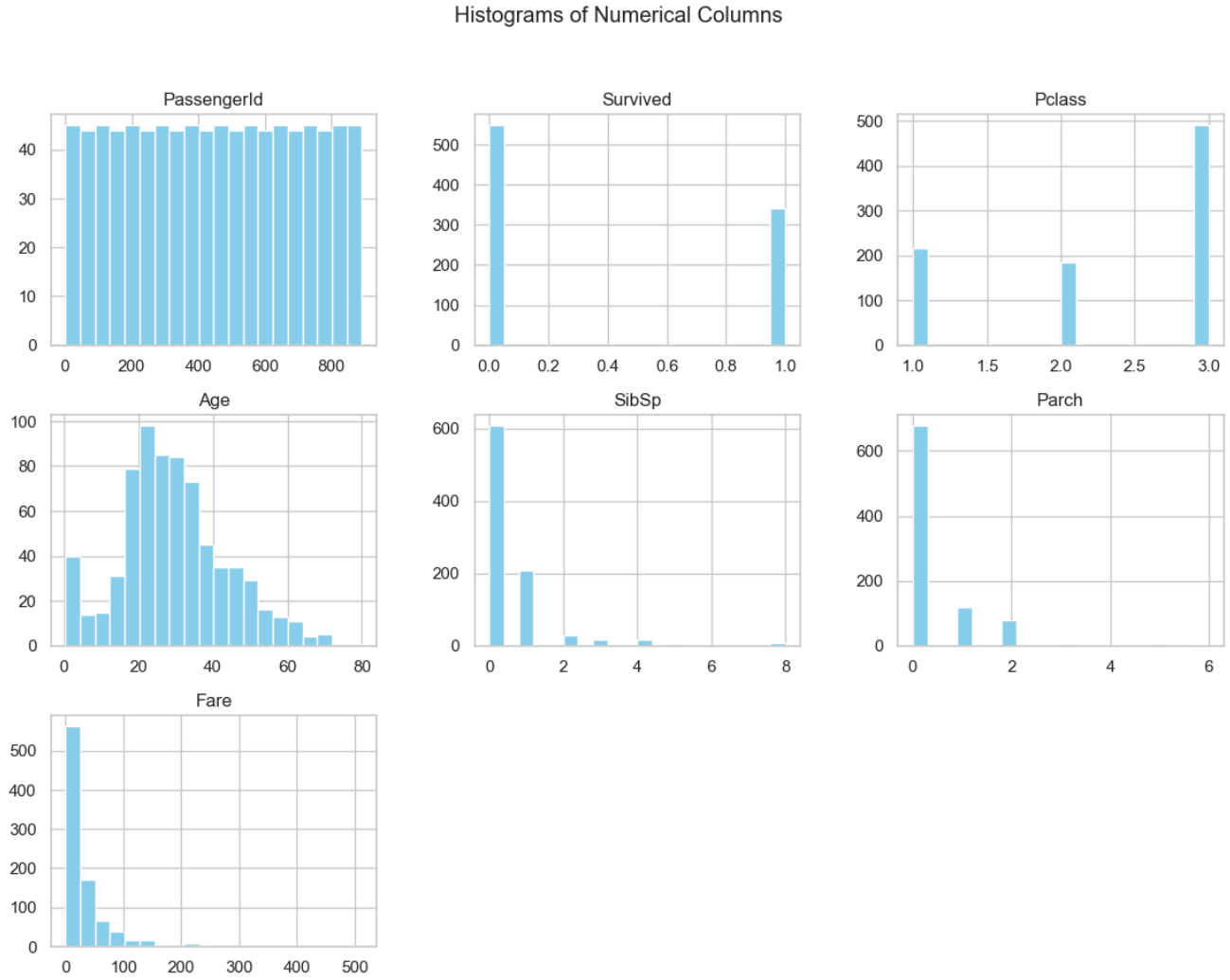

```
sns.countplot(x='Survived', data=train)
plt.title("Survival Count")
```

```
Text(0.5, 1.0, 'Survival Count')
```



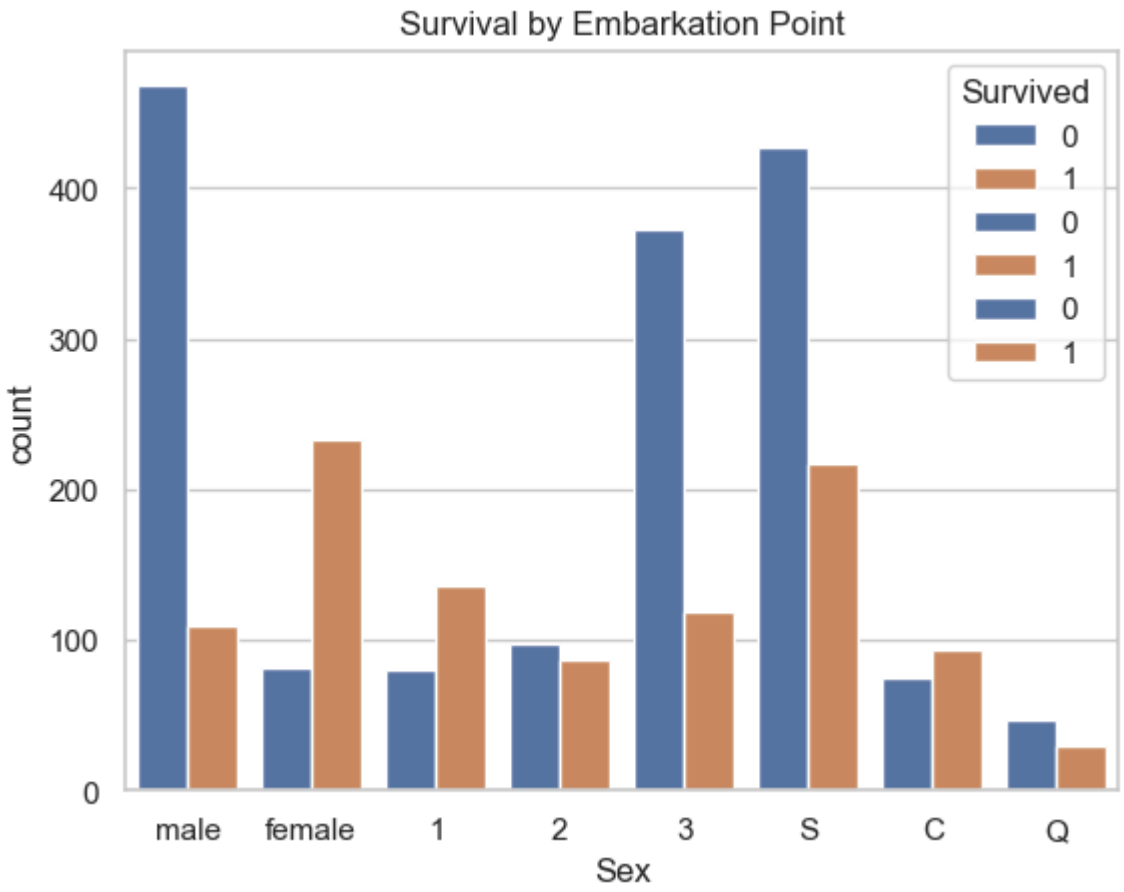

```
train.hist(bins=20, figsize=(14,10), color='skyblue')
plt.suptitle("Histograms of Numerical Columns")
```

```
Text(0.5, 0.98, 'Histograms of Numerical Columns')
```

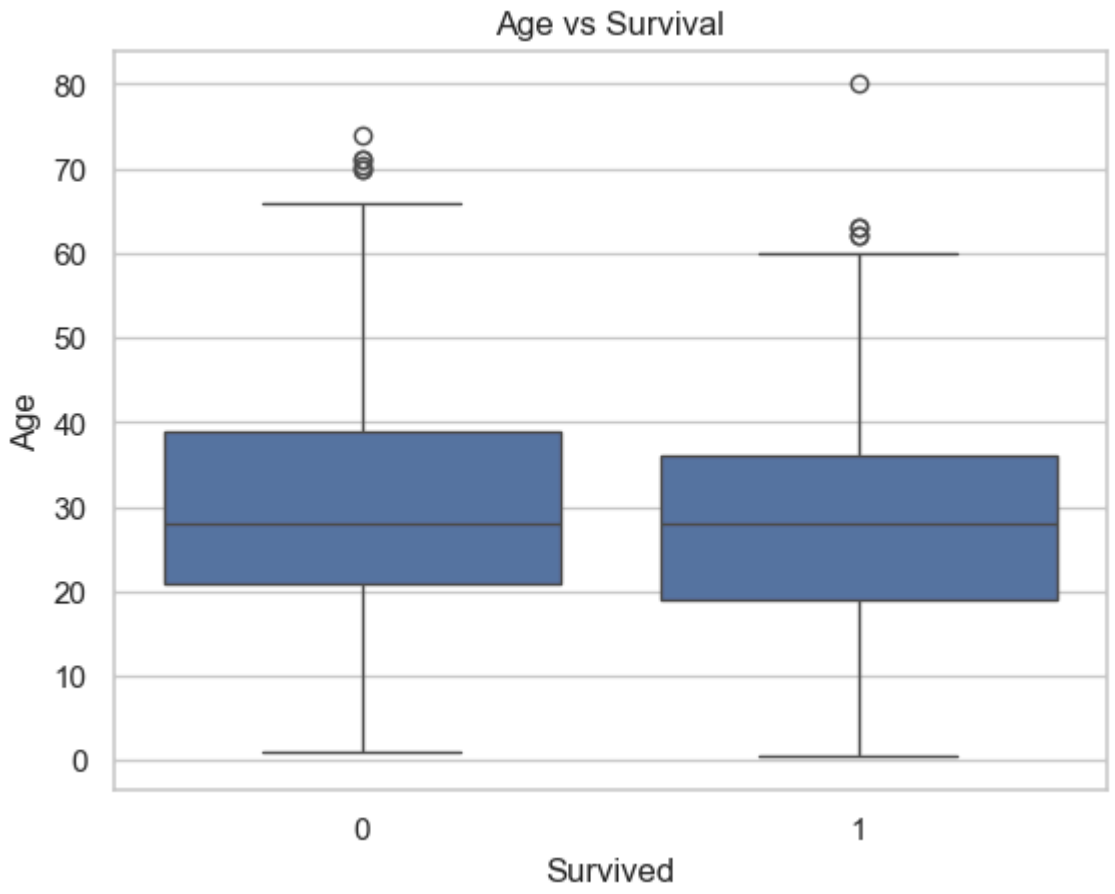


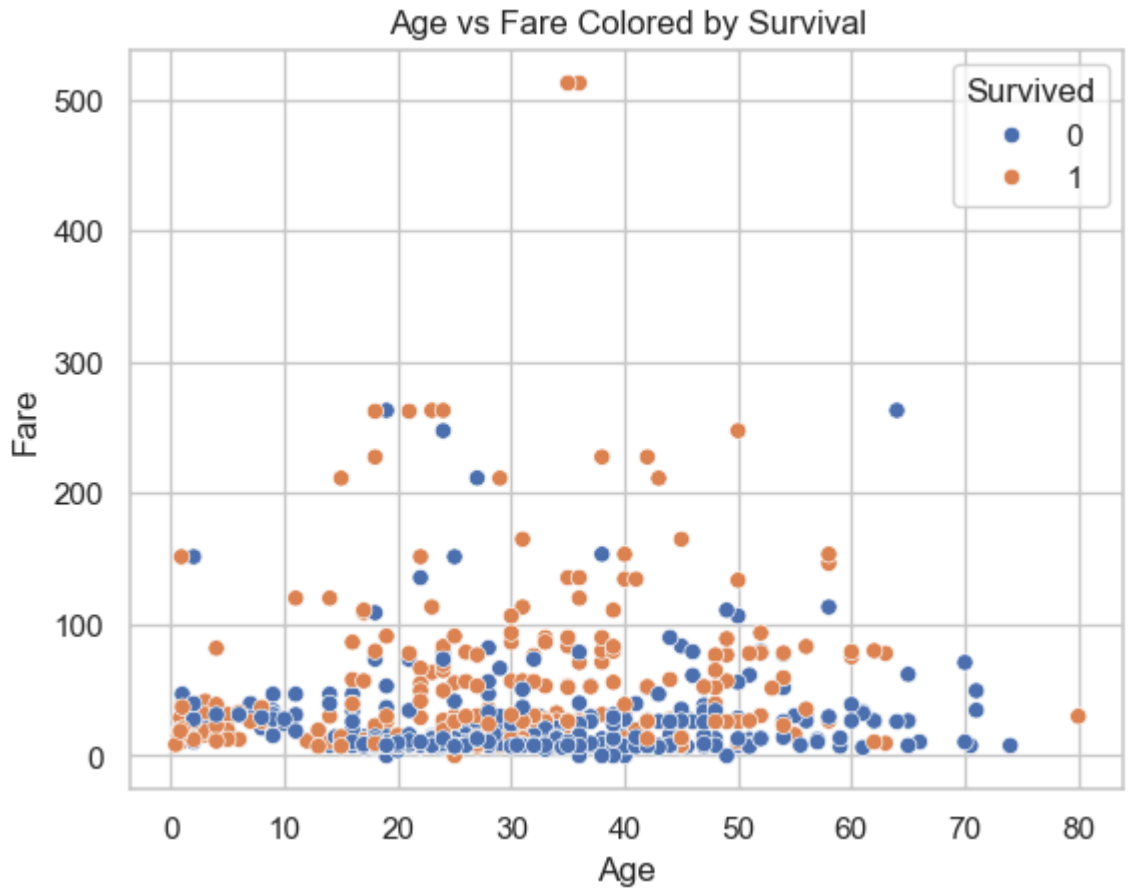

```
sns.countplot(x='Sex', hue='Survived', data=train)
plt.title("Survival by Gender")

sns.countplot(x='Pclass', hue='Survived', data=train)
plt.title("Survival by Class")

sns.countplot(x='Embarked', hue='Survived', data=train)
plt.title("Survival by Embarkation Point")
```

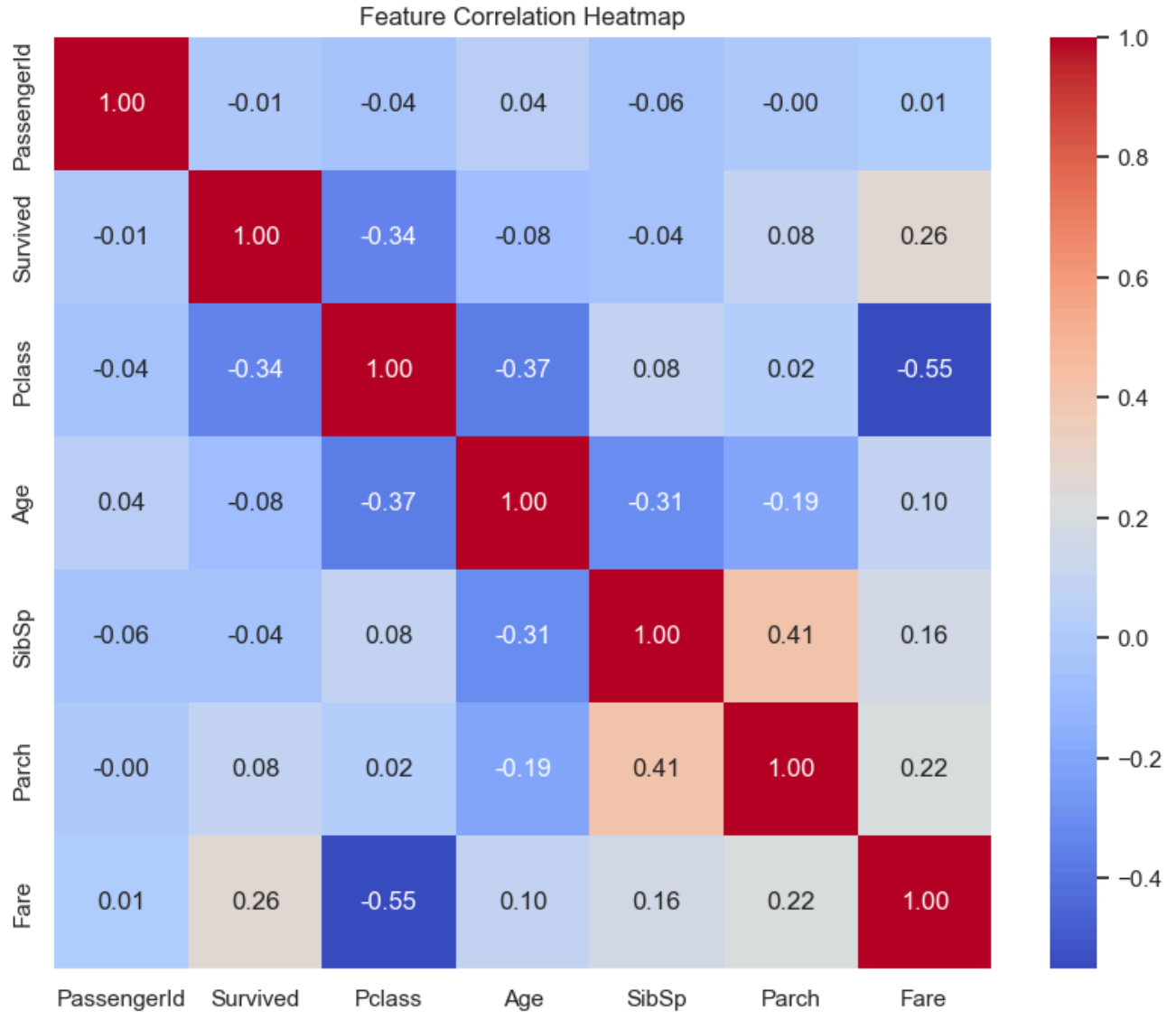
```
Text(0.5, 1.0, 'Survival by Embarkation Point')
```




```
# Boxplot - Age vs Survived
sns.boxplot(x='Survived', y='Age', data=train)
plt.title("Age vs Survival")
plt.show()

# Scatterplot - Fare vs Age
sns.scatterplot(x='Age', y='Fare', hue='Survived', data=train)
plt.title("Age vs Fare Colored by Survival")

# Heatmap of Correlation (numeric columns only)
plt.figure(figsize=(10,8))
numeric_data = train.select_dtypes(include=['int64', 'float64'])
sns.heatmap(numeric_data.corr(), annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Feature Correlation Heatmap")
plt.show()
```

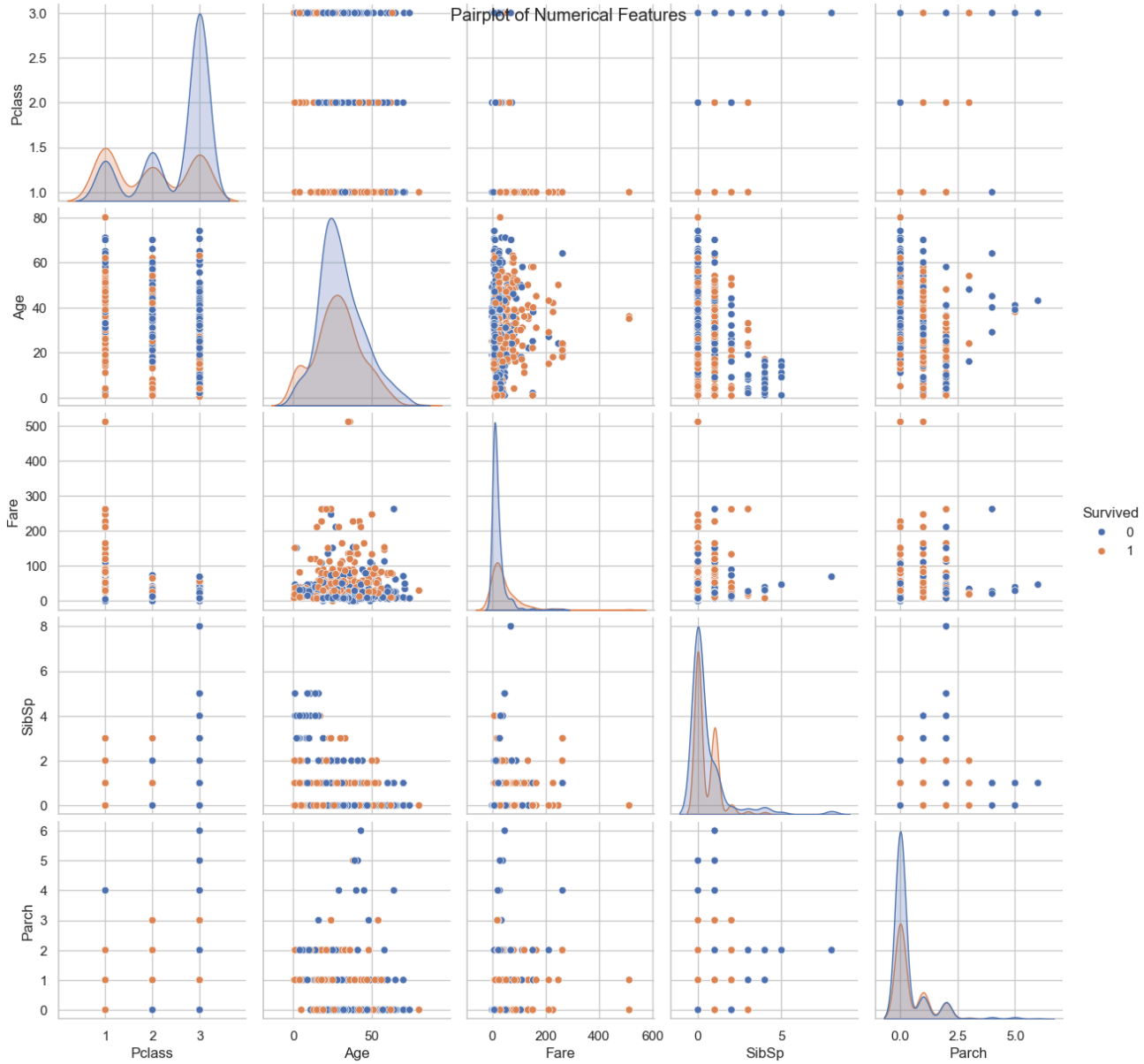

Age vs Survival
Age
Survived
0
1
0
10
20
30
40
50
60
70
80


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




```
sns.pairplot(train[['Survived', 'Pclass', 'Age', 'Fare', 'SibSp', 'Parch']], hue='Survived
plt.suptitle("Pairplot of Numerical Features")
```

```
Text(0.5, 0.98, 'Pairplot of Numerical Features')
```


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