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

# Load Titanic dataset (adjust path as necessary)
df = pd.read_csv("titanic1.csv")

# Display the first few rows
df.head()
```

```
   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
```

```
# Check for missing values
df.isnull().sum()
```

```
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
```

```
# Impute missing values for 'Age' with the median
df['Age'] = df['Age'].fillna(df['Age'].median())

# Impute missing 'Embarked' with the most frequent value
df['Embarked'] = df['Embarked'].fillna(df['Embarked'].mode()[0])

# Drop rows with missing 'Cabin' (due to many missing values)
df = df.drop(columns=['Cabin'])

# Check again for missing values
df.isnull().sum()
```

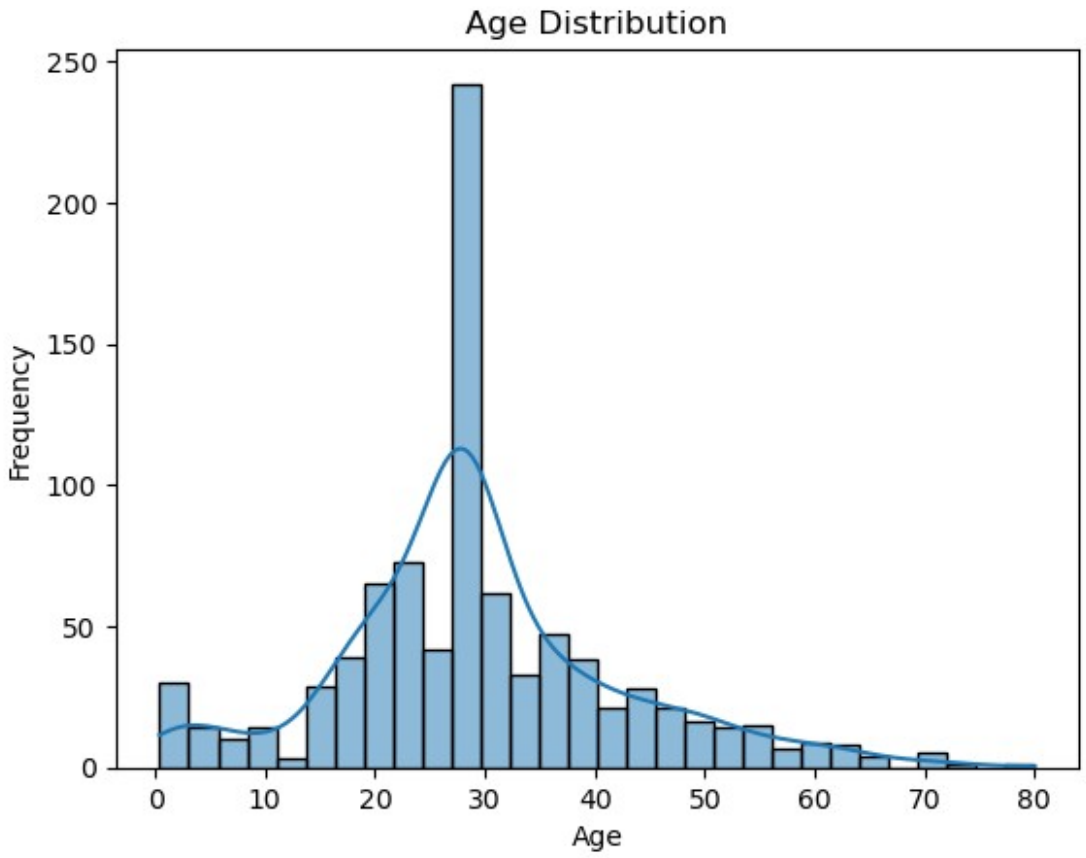
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       0
dtype: int64
```

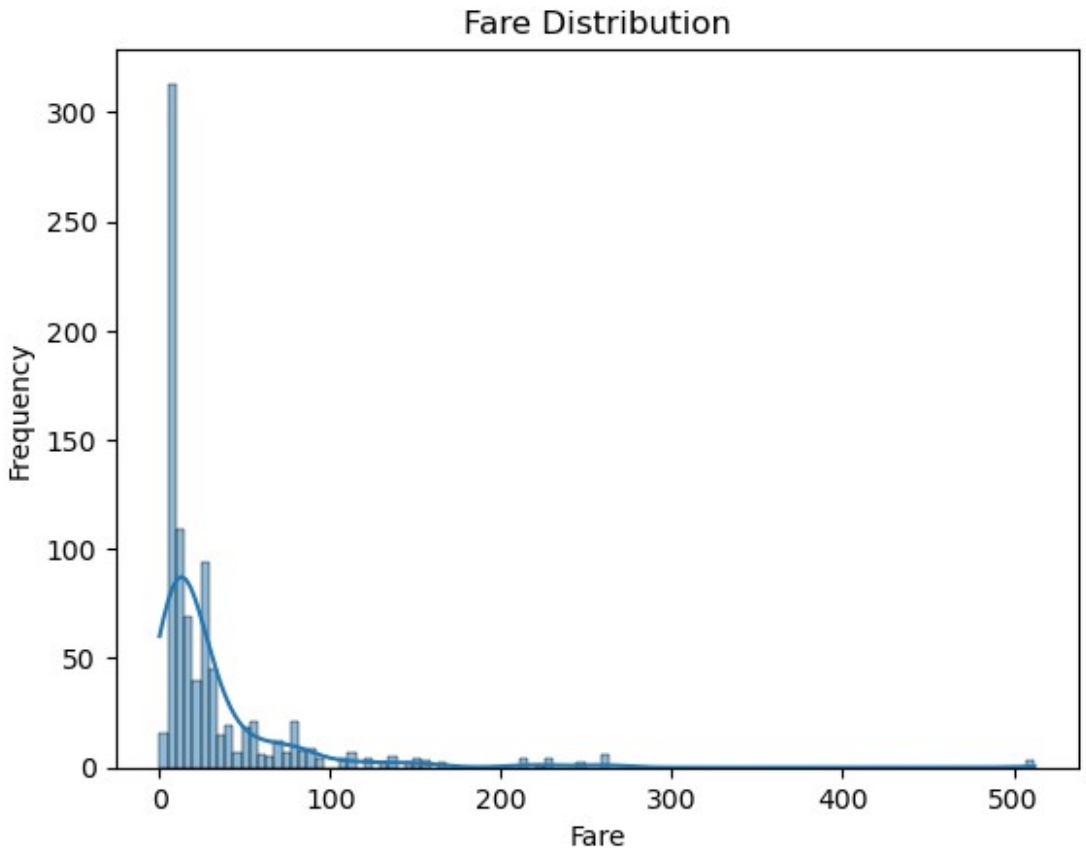
```
# Convert categorical features to numeric codes
df['Sex'] = df['Sex'].map({'male': 0, 'female': 1})
df['Embarked'] = df['Embarked'].map({'C': 0, 'Q': 1, 'S': 2})

# Check the transformed data
df.head()
```

```
   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
Parch  \
0                            Braund, Mr. Owen Harris    0  22.0      1
0
1  Cumings, Mrs. John Bradley (Florence Briggs Th...    1  38.0      1
0
2                             Heikkinen, Miss. Laina    1  26.0      0
0
```

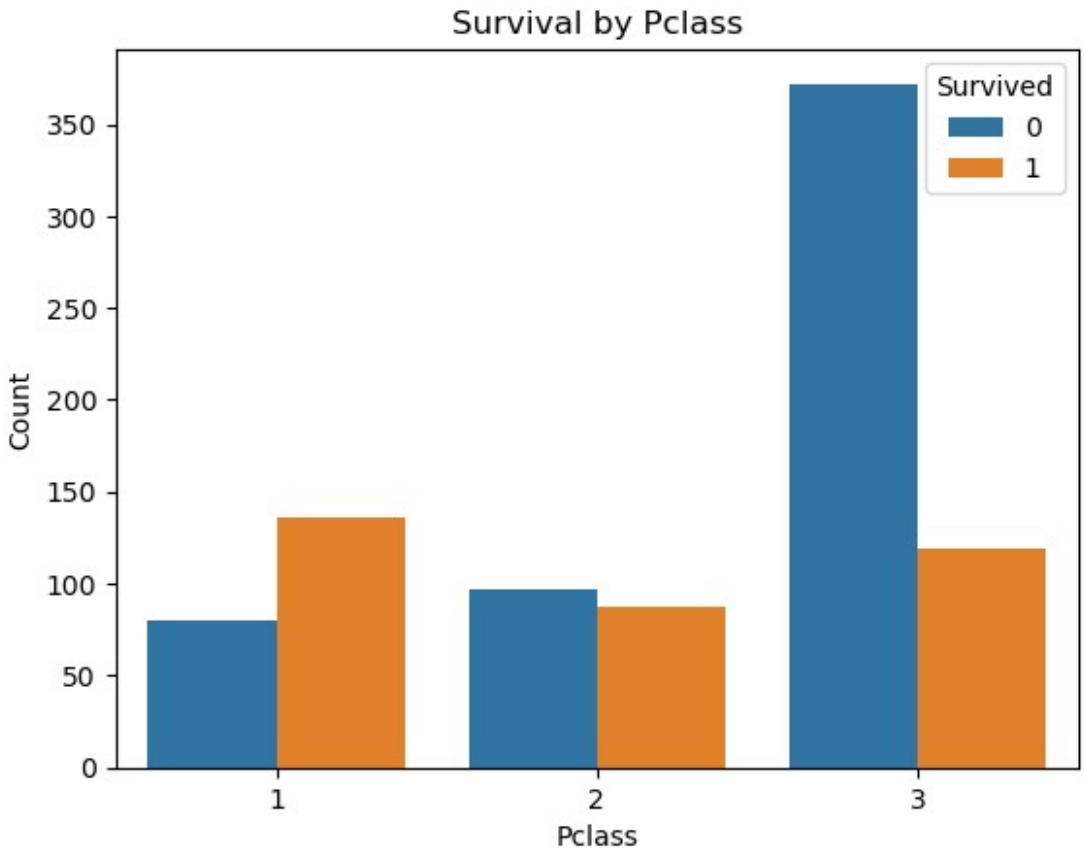
```
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)    1  35.0      1
0
4                            Allen, Mr. William Henry    0  35.0      0
0

             Ticket     Fare  Embarked
0         A/5 21171   7.2500         2
1          PC 17599  71.2833         0
2  STON/O2. 3101282   7.9250         2
3            113803  53.1000         2
4            373450   8.0500         2
```

```python
# Age distribution
sns.histplot(df['Age'], kde=True)
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()

# Fare distribution
sns.histplot(df['Fare'], kde=True)
plt.title('Fare Distribution')
plt.xlabel('Fare')
plt.ylabel('Frequency')
plt.show()
```

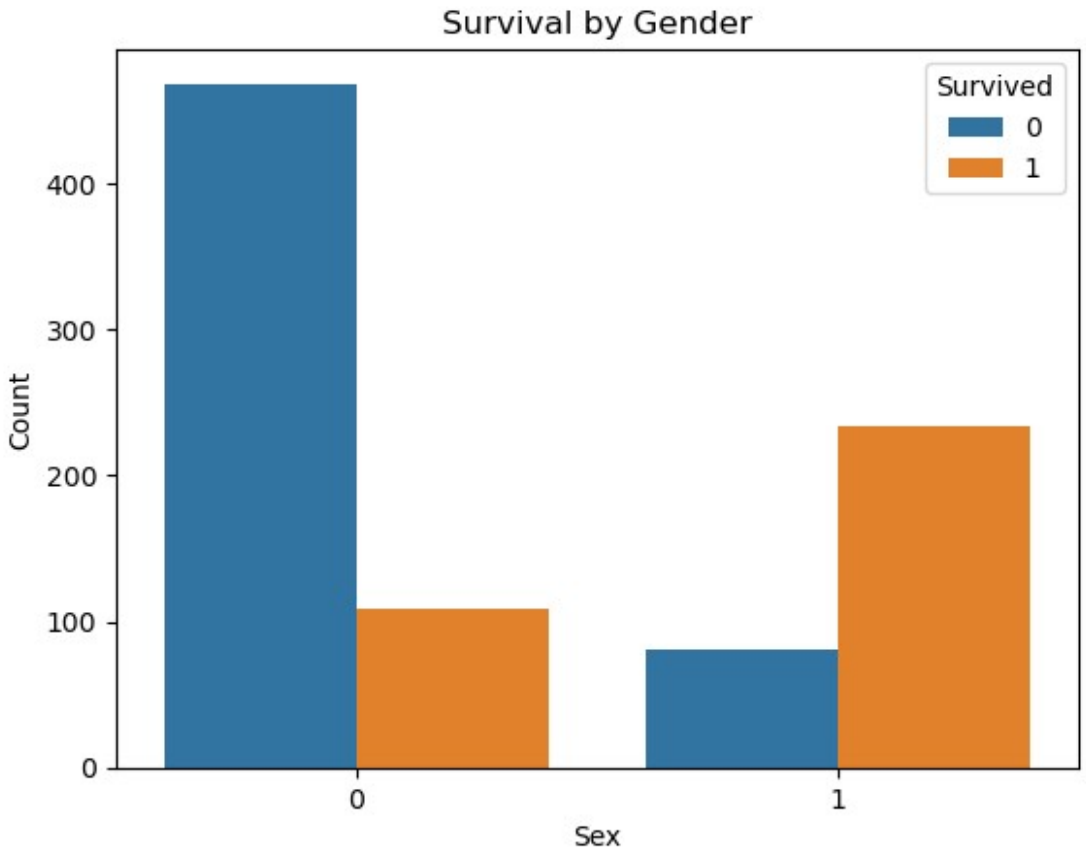

Age Distribution
Frequency
Age
0
50
100
150
200
250
0
10
20
30
40
50
60
70
80



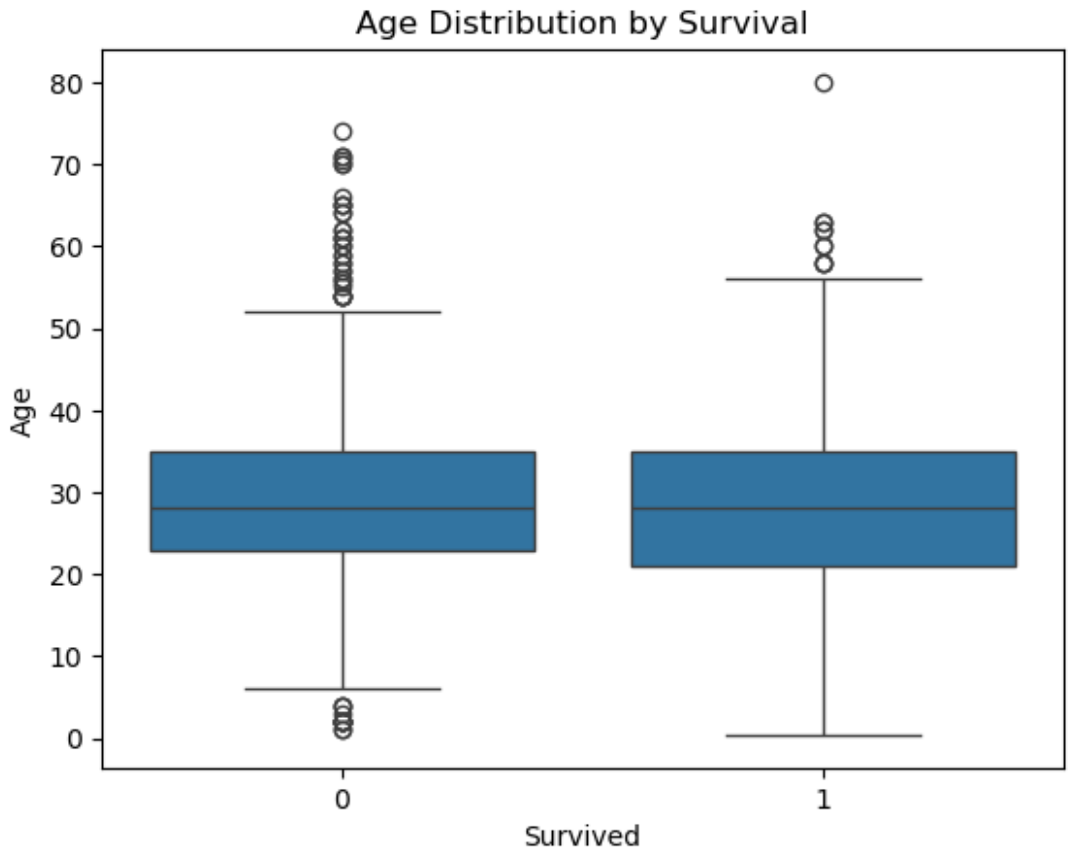

```
sns.countplot(data=df, x='Pclass', hue='Survived')
plt.title('Survival by Pclass')
plt.xlabel('Pclass')
plt.ylabel('Count')
plt.show()
```



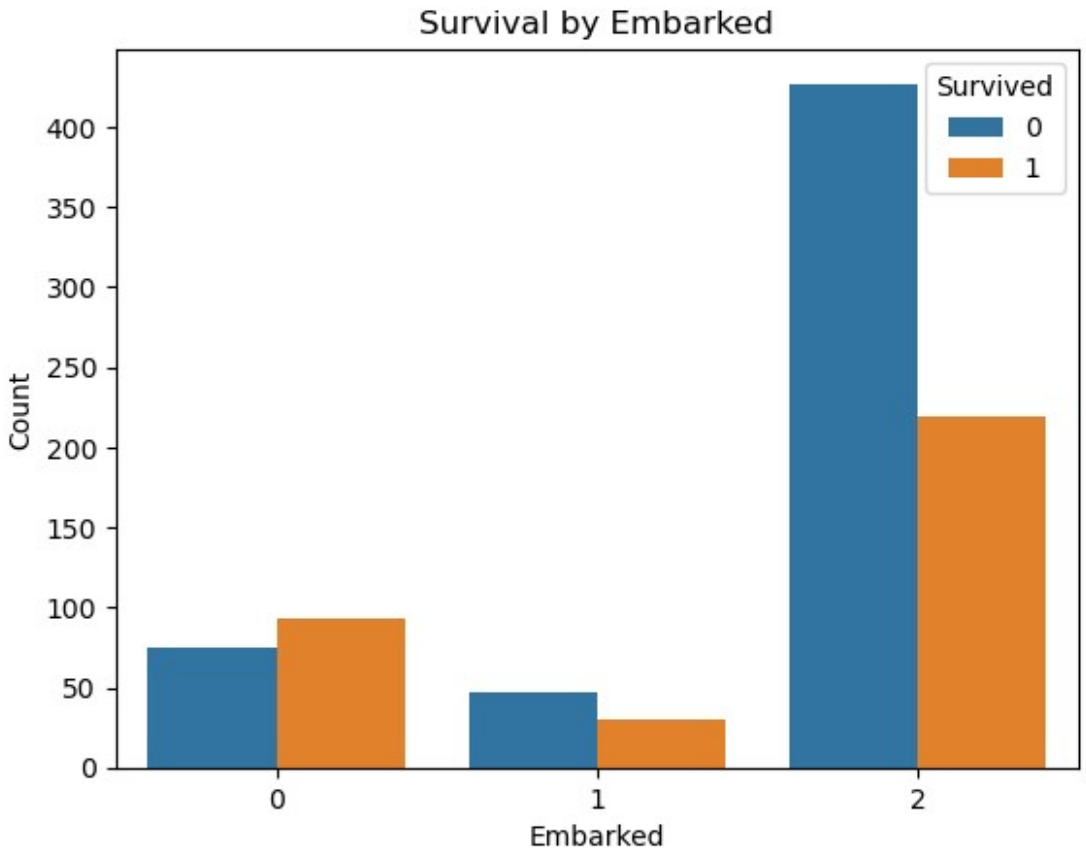

```
sns.countplot(data=df, x='Sex', hue='Survived')
plt.title('Survival by Gender')
plt.xlabel('Sex')
plt.ylabel('Count')
plt.show()
```



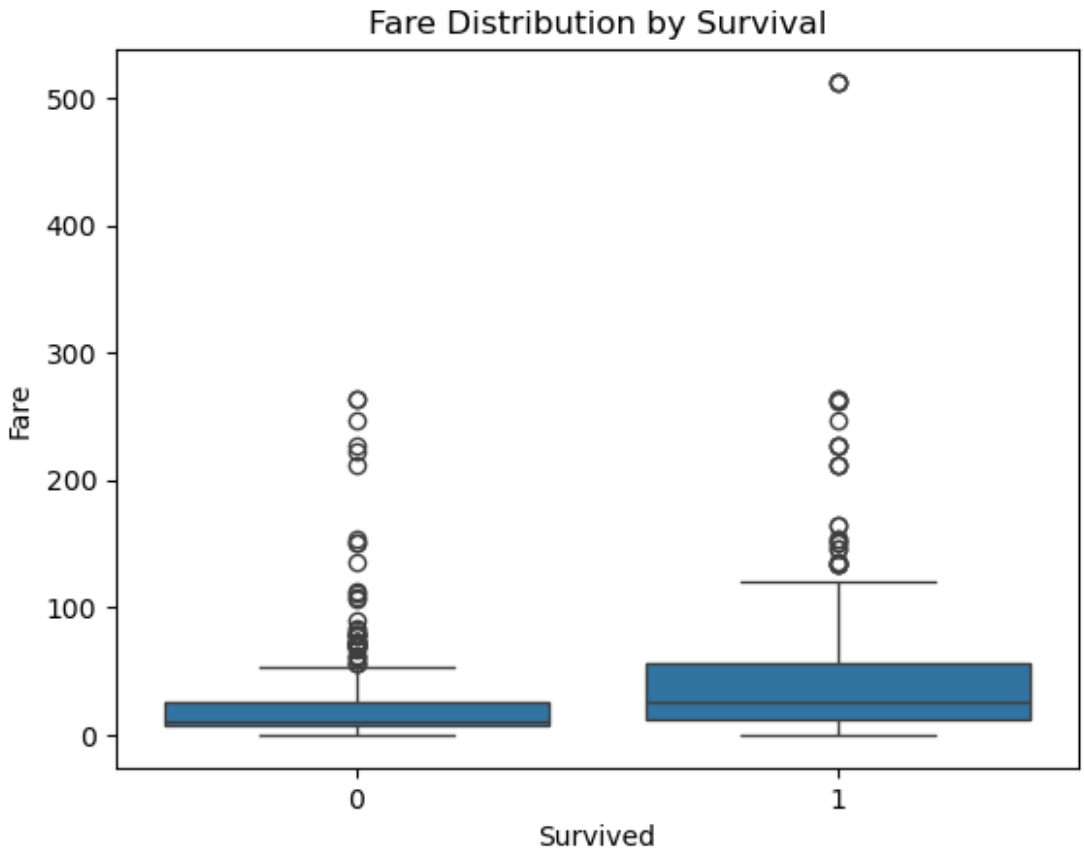

```
sns.boxplot(data=df, x='Survived', y='Age')
plt.title('Age Distribution by Survival')
plt.xlabel('Survived')
plt.ylabel('Age')
plt.show()
```




```
sns.countplot(data=df, x='Embarked', hue='Survived')
plt.title('Survival by Embarked')
plt.xlabel('Embarked')
plt.ylabel('Count')
plt.show()
```




```
sns.boxplot(data=df, x='Survived', y='Fare')
plt.title('Fare Distribution by Survival')
plt.xlabel('Survived')
plt.ylabel('Fare')
plt.show()
```


Fare Distribution by Survival
500
400
300
200
100
0
Fare
0
1
Survived