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

```
Matplotlib is building the font cache; this may take a moment.
```

In [2]:
```
df = pd.read_csv("train.csv")
```

In [3]:
```
df.head()
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	ST 3
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	

In [4]:
```
df.shape
```

Out[4]:
```
(891, 12)
```

In [5]:
```
df.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
```

In [6]:
```
df.describe()
```

Out[6]:

	PassengerId	Survived	Pclass	Age	SibSp	
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.0
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.3
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.8
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.0
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.0
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.0
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.0
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.0

In [7]:
```
df.isnull().sum()
```

Out[7]:
```
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
```

In [11]:
```
df['Age']=df['Age'].fillna(df['Age'].median())
```

In [12]:
```
df['Embarked'] = df['Embarked'].fillna(df['Embarked'].mode()[0])
```

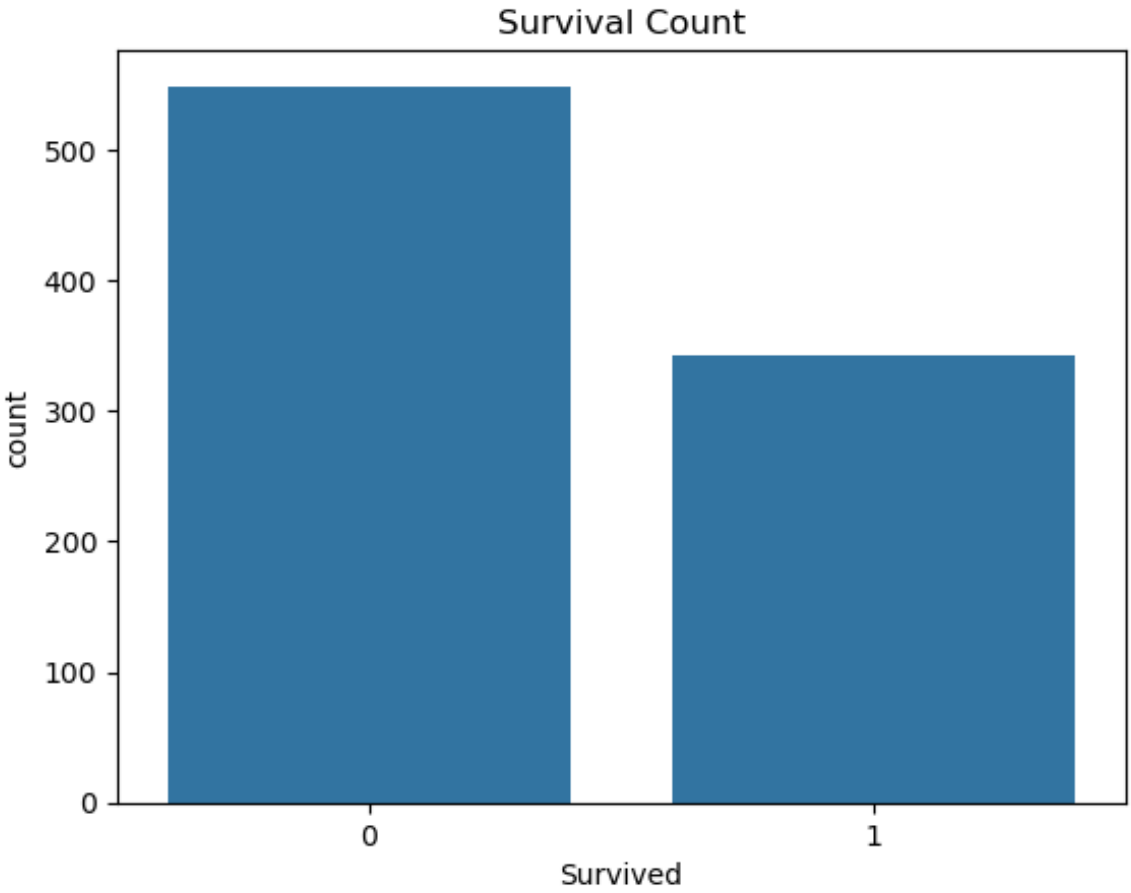
In [13]:
```
df.drop('Cabin' , axis=1, inplace=True)
```

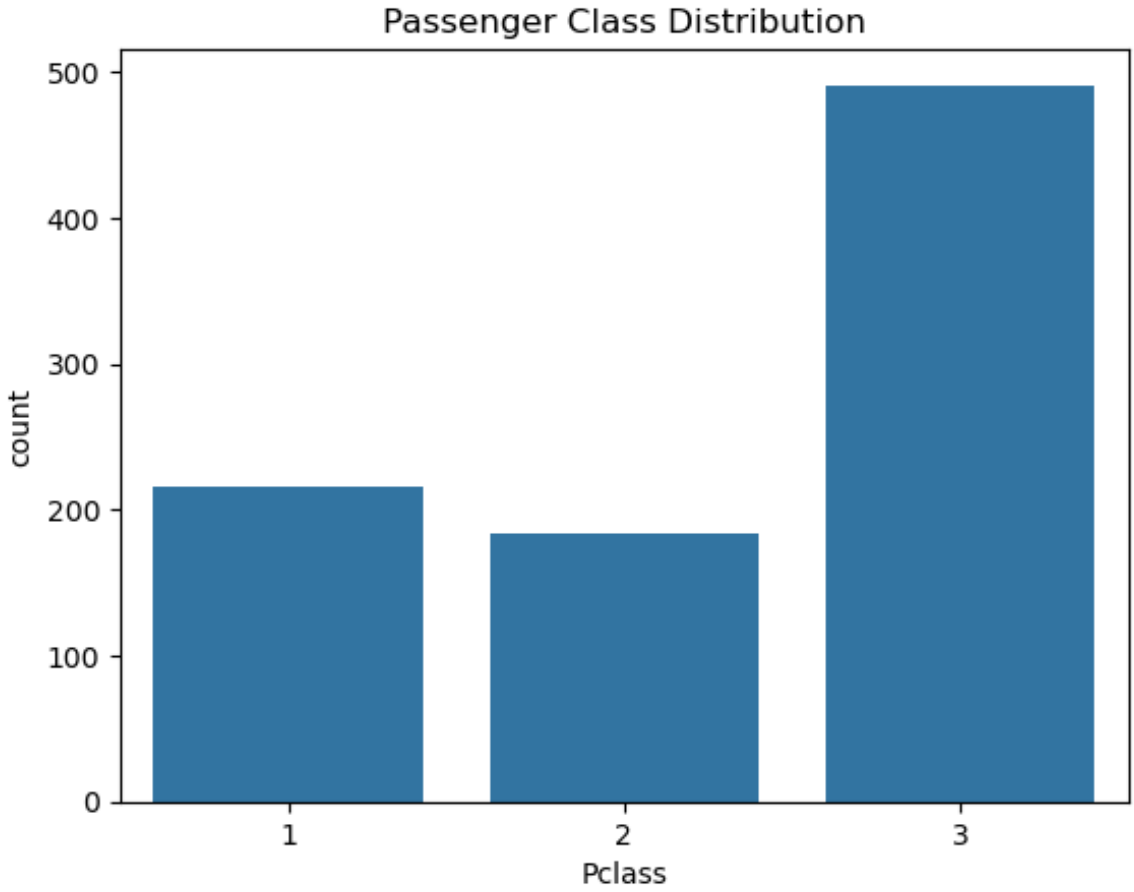
In [16]:
```
df.isnull().sum()
```

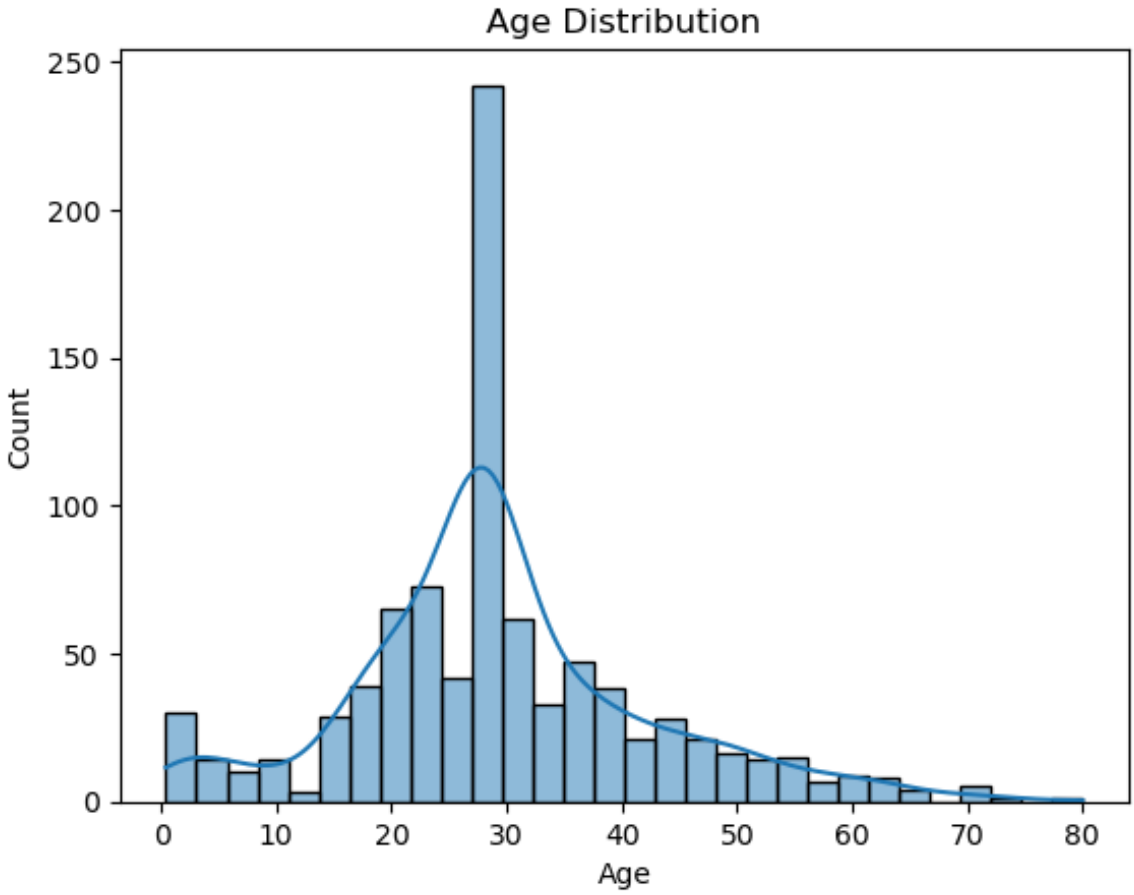
Out[16]:
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       0
dtype: int64
```

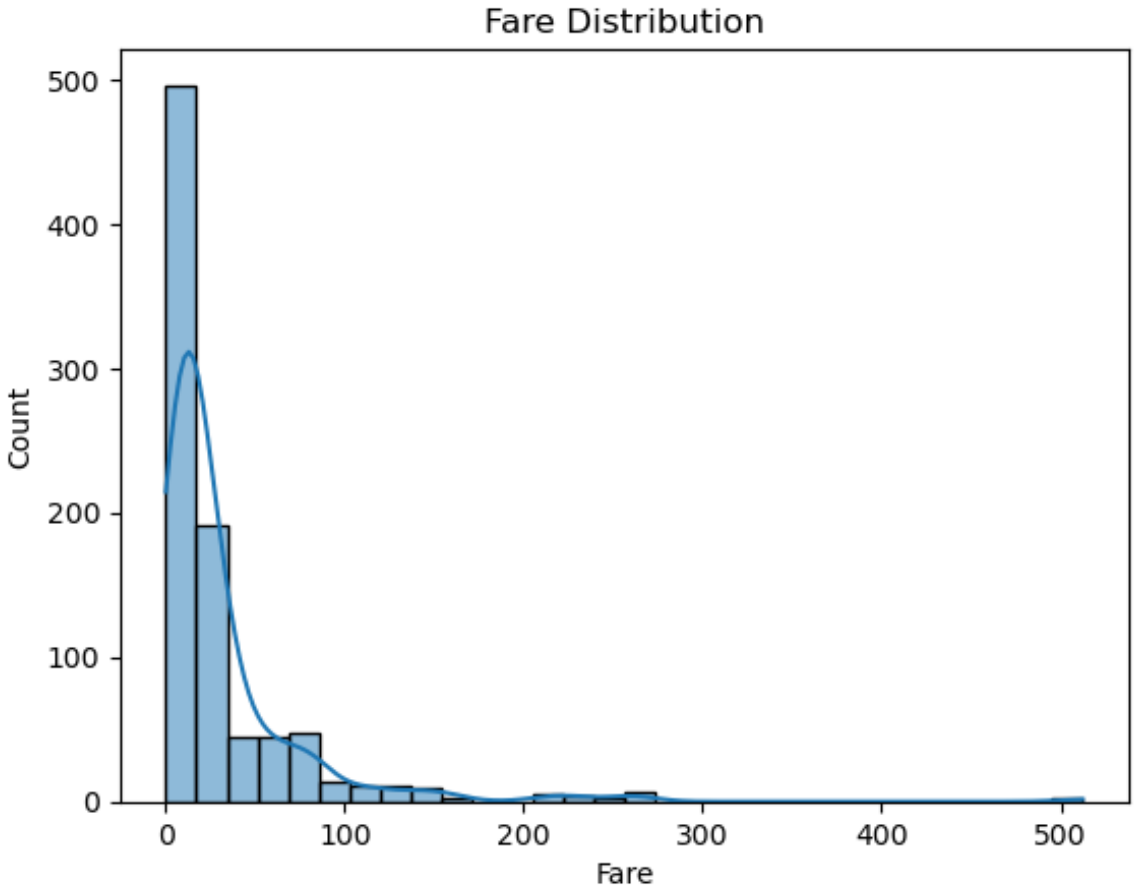
In [18]:
```

```




Passenger Class Distribution
count
Pclass
0
100
200
300
400
500
1
2
3


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

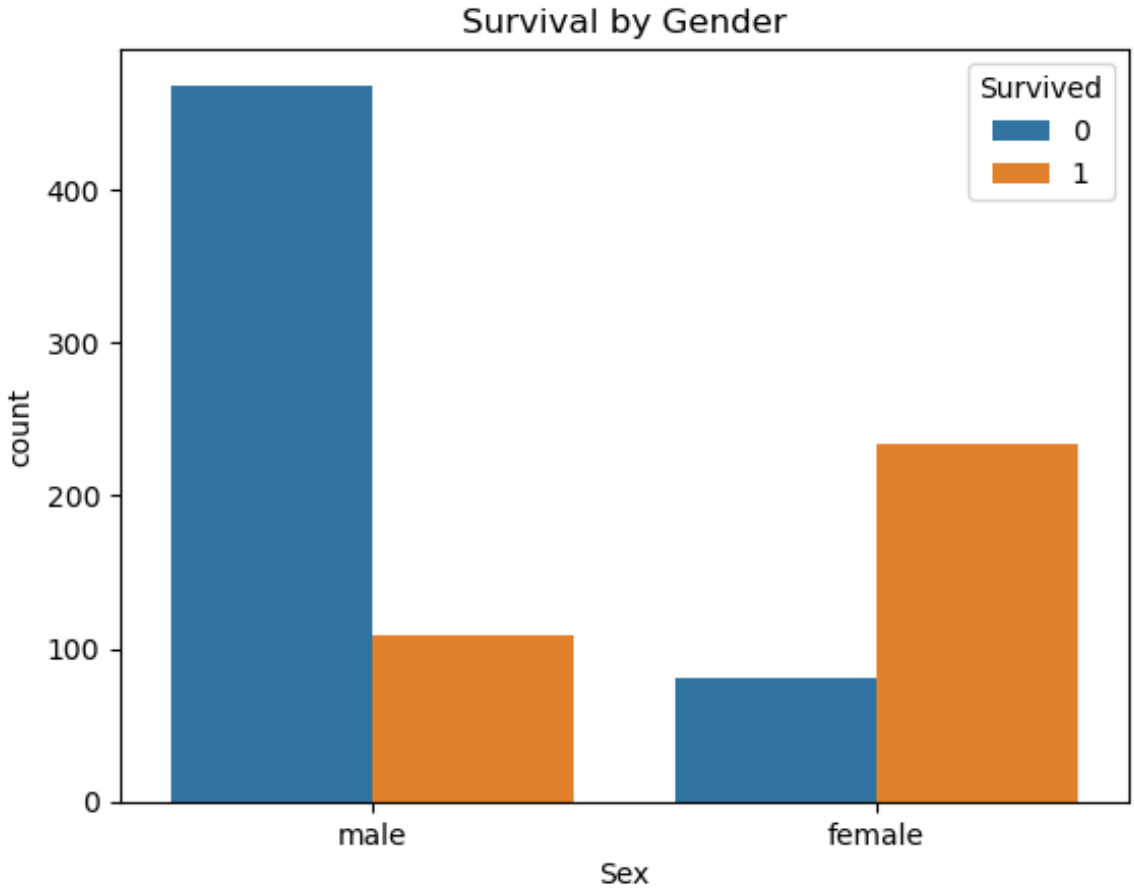


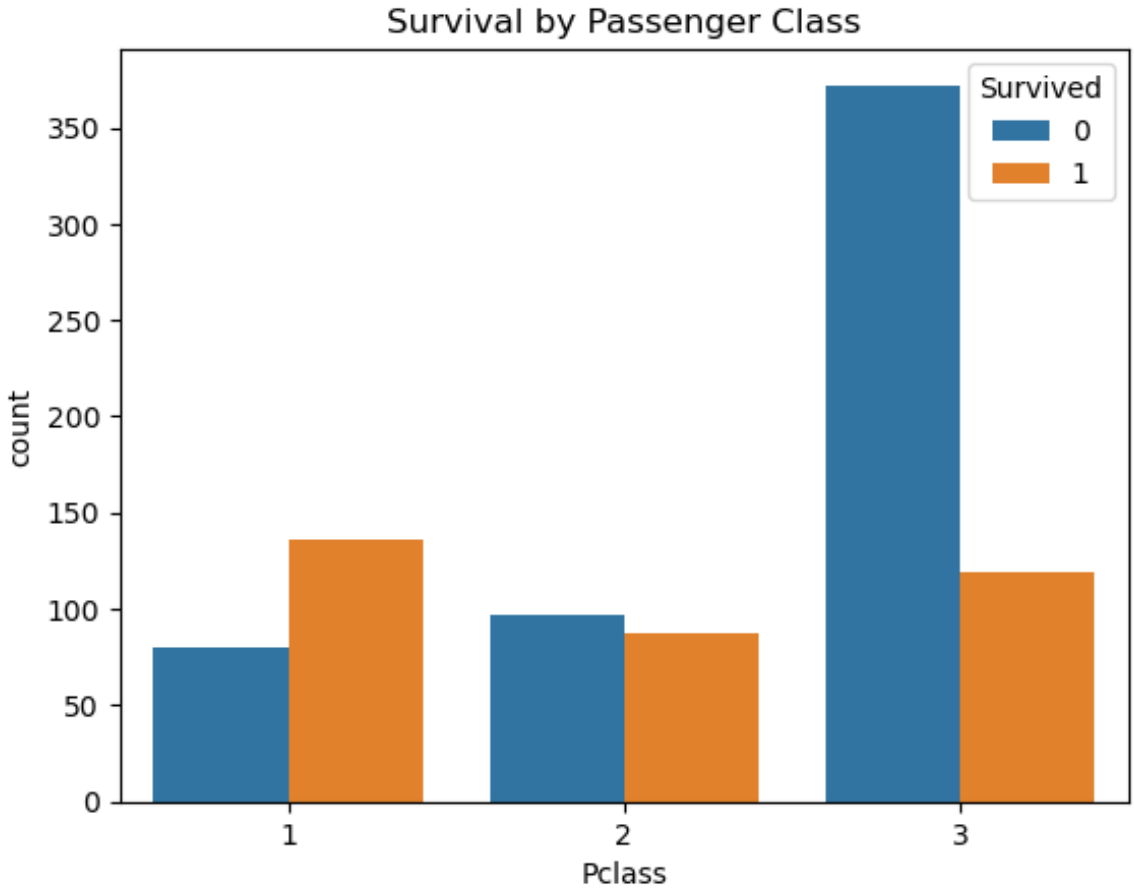
Observation: Most passengers were in 3rd class. Survival rate seems low overall. Age distribution is roughly normal, centered around 30.

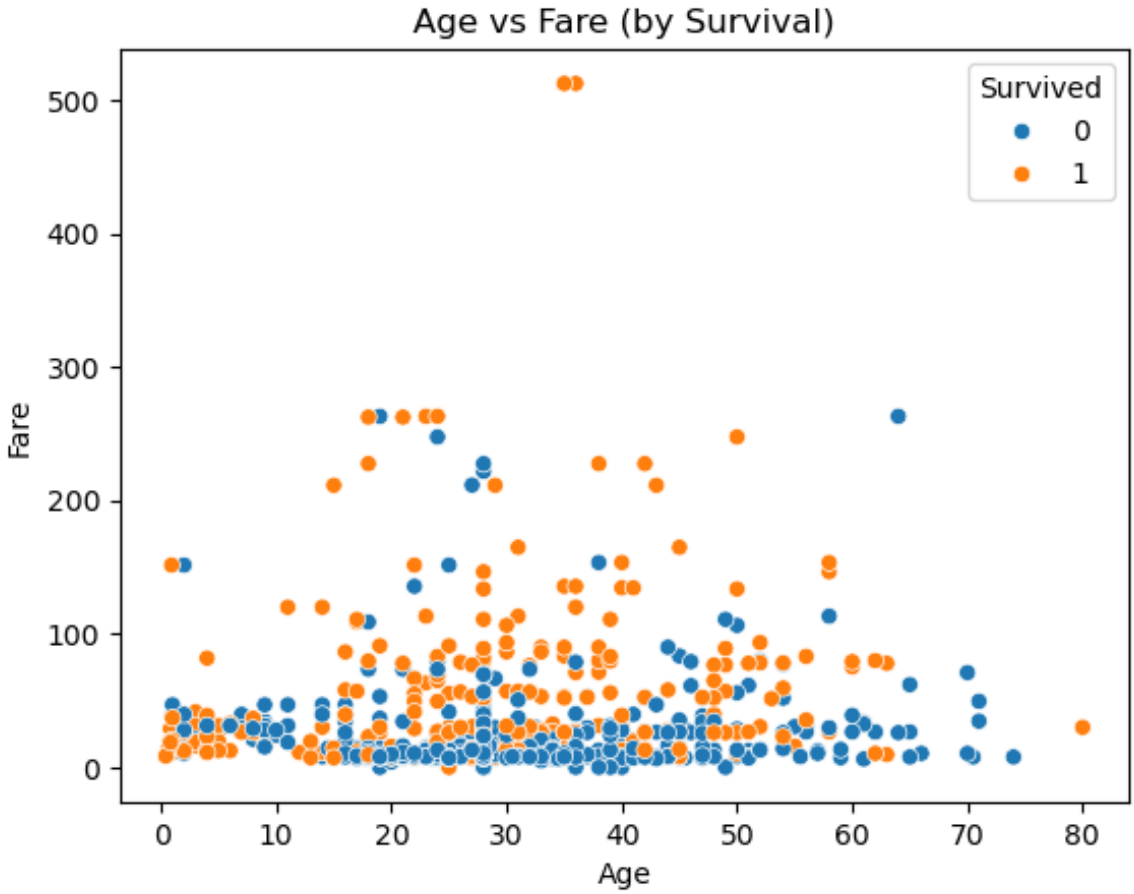
In [19]:
```
# Gender vs Survival
sns.countplot(x='Sex', hue='Survived', data=df)
plt.title('Survival by Gender')
plt.show()

# Class vs Survival
sns.countplot(x='Pclass', hue='Survived', data=df)
plt.title('Survival by Passenger Class')
plt.show()

# Age vs Fare
sns.scatterplot(x='Age', y='Fare', hue='Survived', data=df)
plt.title('Age vs Fare (by Survival)')
plt.show()
```

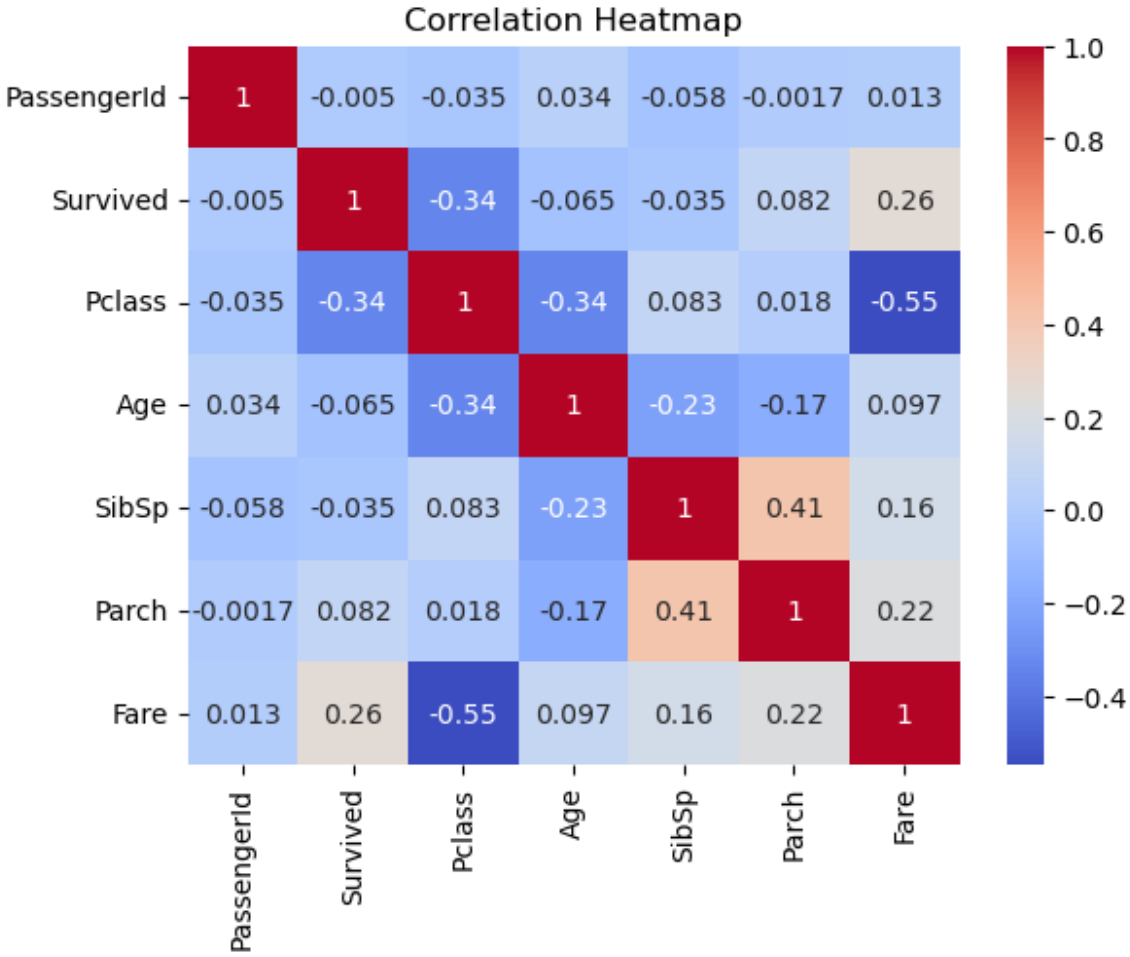

Survival by Gender
Survived
0
1
count
0
100
200
300
400
male
female
Sex


Survival by Passenger Class
Survived
0
1
count
0
50
100
150
200
250
300
350
1
2
3
Pclass



Observation: Female passengers had higher survival rate than males.
Passengers in 1st class survived more often than those in 3rd

In [20]:
```
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm'
plt.title('Correlation Heatmap')
plt.show()
```



Observation: Fare and passenger class are negatively correlated (higher fare → higher class).

Summary of Insights: Most passengers were from 3rd class. Females and 1st class passengers had higher survival rates. Younger passengers had slightly better survival chances. Fare had a positive relationship with survival.