Missing values in the dataset:

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

Dataset shape: (891, 12)

Dataset 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 int32

5 Age 891 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 891 non-null int32

dtypes: float64(2), int32(2), int64(5), object(3)

memory usage: 76.7+ KB

None

Dataset summary statistics:

PassengerId Survived Pclass Sex Age \

count 891.000000 891.000000 891.000000 891.000000 891.000000

mean 446.000000 0.383838 2.308642 0.647587 29.361582

std 257.353842 0.486592 0.836071 0.477990 13.019697

min 1.000000 0.000000 1.000000 0.000000 0.420000

25% 223.500000 0.000000 2.000000 0.000000 22.000000

50% 446.000000 0.000000 3.000000 1.000000 28.000000

75% 668.500000 1.000000 3.000000 1.000000 35.000000

max 891.000000 1.000000 3.000000 1.000000 80.000000

SibSp Parch Fare Embarked

count 891.000000 891.000000 891.000000 891.000000

mean 0.523008 0.381594 32.204208 1.536476

std 1.102743 0.806057 49.693429 0.791503

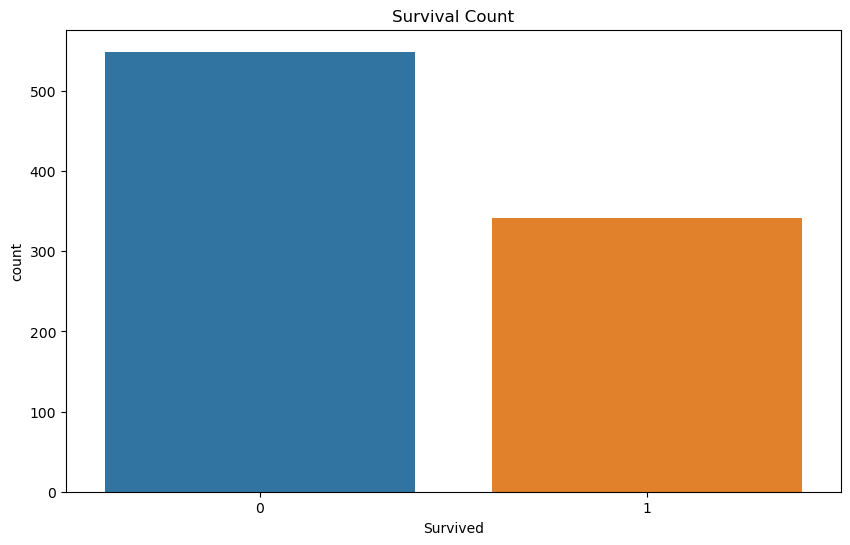
min 0.000000 0.000000 0.000000 0.000000

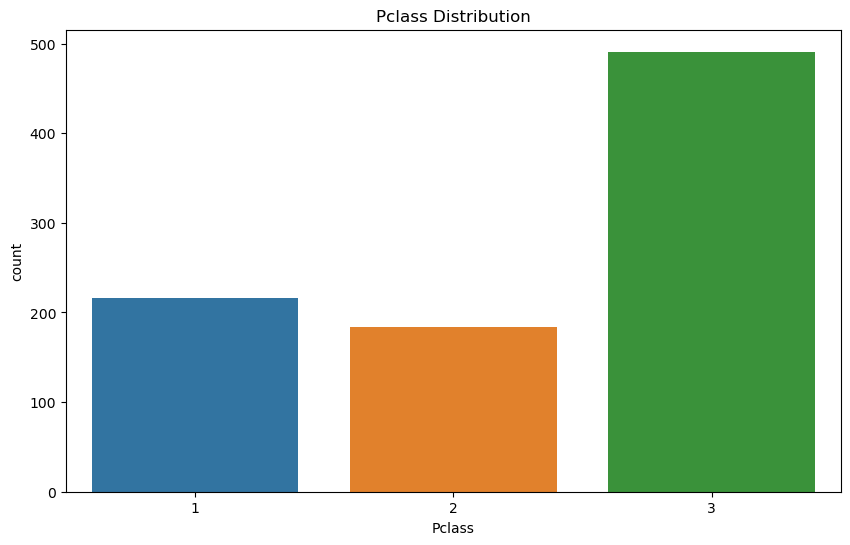
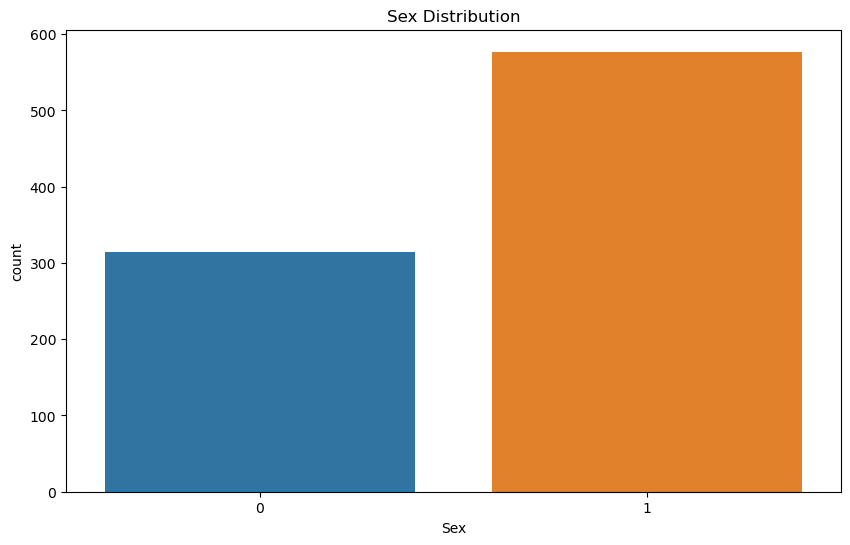
25% 0.000000 0.000000 7.910400 1.000000

50% 0.000000 0.000000 14.454200 2.000000

75% 1.000000 0.000000 31.000000 2.000000

max 8.000000 6.000000 512.329200 2.000000





Out[6]:Text(0.5, 1.0, 'Embarked Distribution')

