```
import os
os.listdir('/kaggle/input/titanic-task5')

import pandas as pd

df = pd.read_csv('/kaggle/input/titanic-task5/train.csv')
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
```

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

plt.rcParams['figure.figsize'] = (10,6)
sns.set(style='whitegrid')



df.shape
```

```
(891, 12)
```

```
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
```

```
df.describe(include='all').T
```

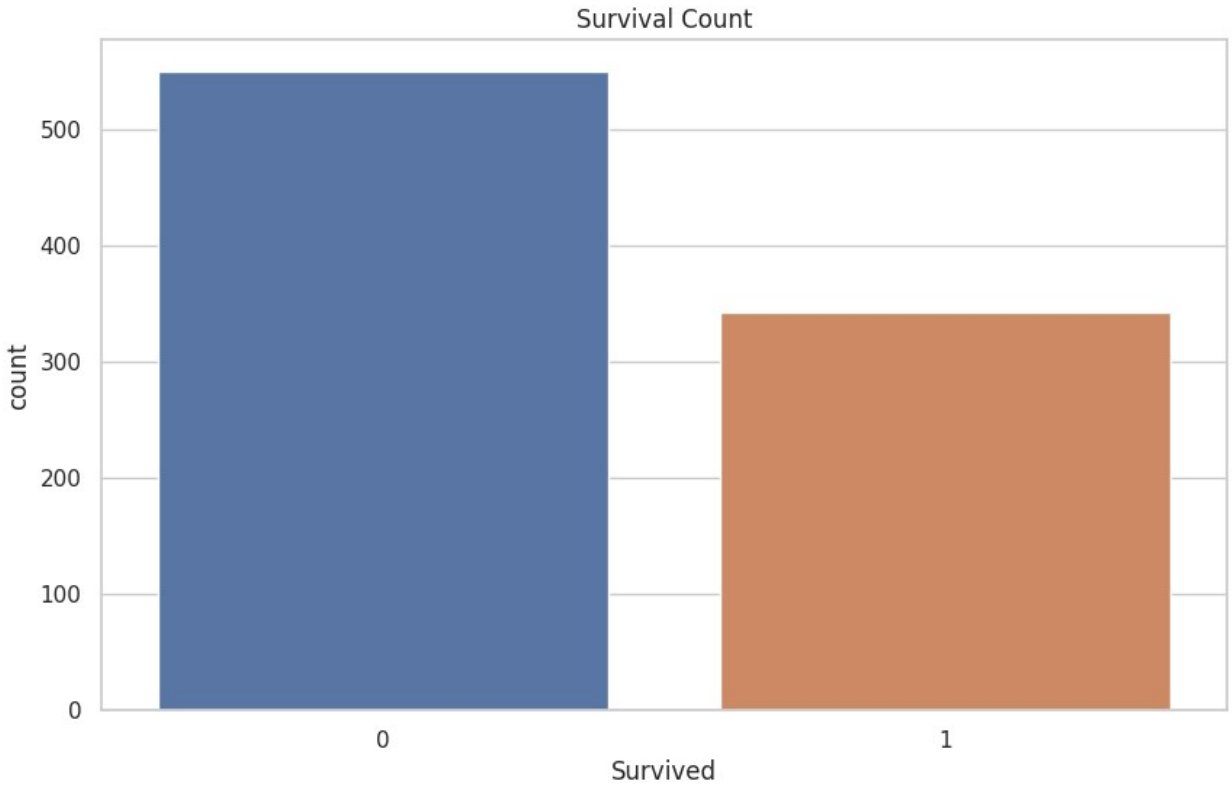
```
             count unique                      top freq       mean  \
PassengerId  891.0    NaN                      NaN  NaN      446.0
Survived     891.0    NaN                      NaN  NaN   0.383838
Pclass       891.0    NaN                      NaN  NaN   2.308642
Name           891    891  Braund, Mr. Owen Harris    1        NaN
Sex            891      2                     male  577        NaN
Age          714.0    NaN                      NaN  NaN  29.699118
SibSp        891.0    NaN                      NaN  NaN   0.523008
Parch        891.0    NaN                      NaN  NaN   0.381594
Ticket         891    681                   347082    7        NaN
Fare         891.0    NaN                      NaN  NaN  32.204208
Cabin          204    147                  B96 B98    4        NaN
Embarked       889      3                        S  644        NaN

                    std   min     25%      50%    75%       max
PassengerId  257.353842   1.0   223.5    446.0  668.5     891.0
Survived       0.486592   0.0     0.0      0.0    1.0       1.0
Pclass         0.836071   1.0     2.0      3.0    3.0       3.0
Name                NaN   NaN     NaN      NaN    NaN       NaN
Sex                 NaN   NaN     NaN      NaN    NaN       NaN
Age           14.526497  0.42  20.125     28.0   38.0      80.0
SibSp          1.102743   0.0     0.0      0.0    1.0       8.0
Parch          0.806057   0.0     0.0      0.0    0.0       6.0
Ticket              NaN   NaN     NaN      NaN    NaN       NaN
Fare          49.693429   0.0  7.9104  14.4542   31.0  512.3292
Cabin               NaN   NaN     NaN      NaN    NaN       NaN
Embarked            NaN   NaN     NaN      NaN    NaN       NaN
```

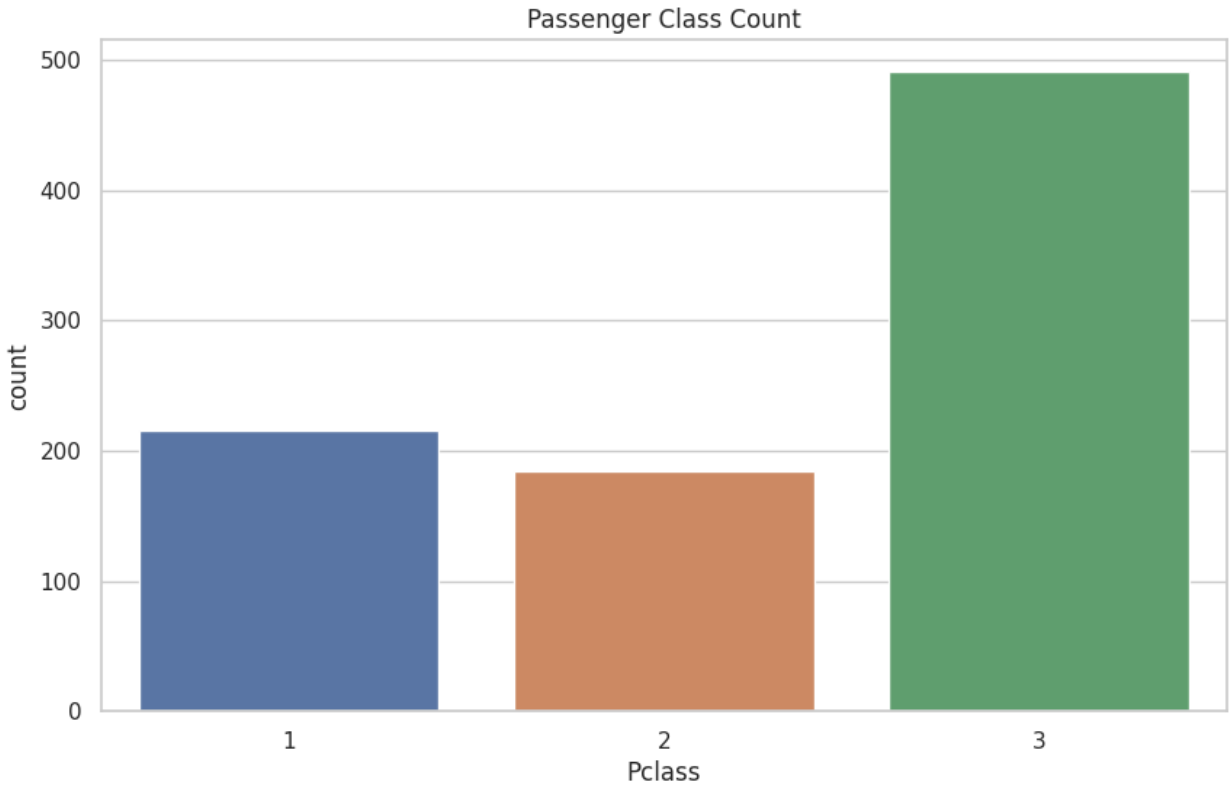
```
df.isnull().sum()
```

```
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
```

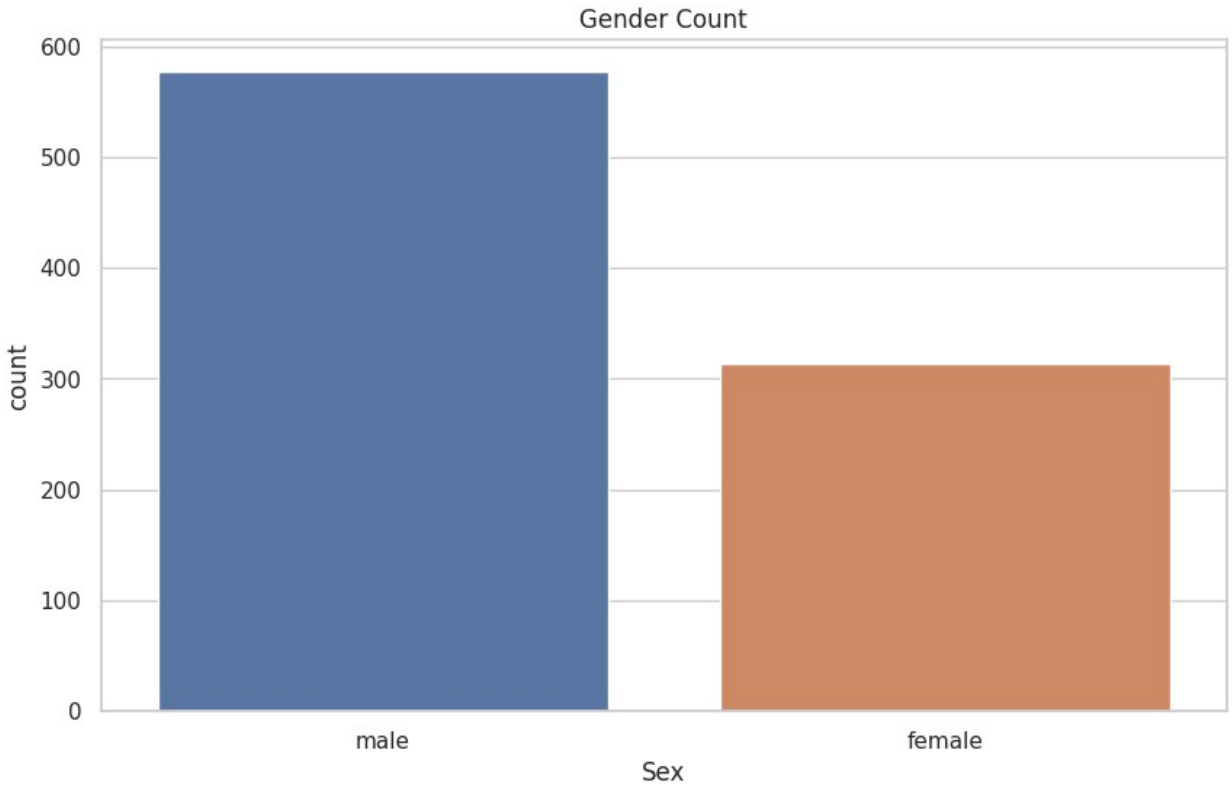
```
sns.countplot(x='Survived', data=df)
plt.title("Survival Count")
plt.show()
```



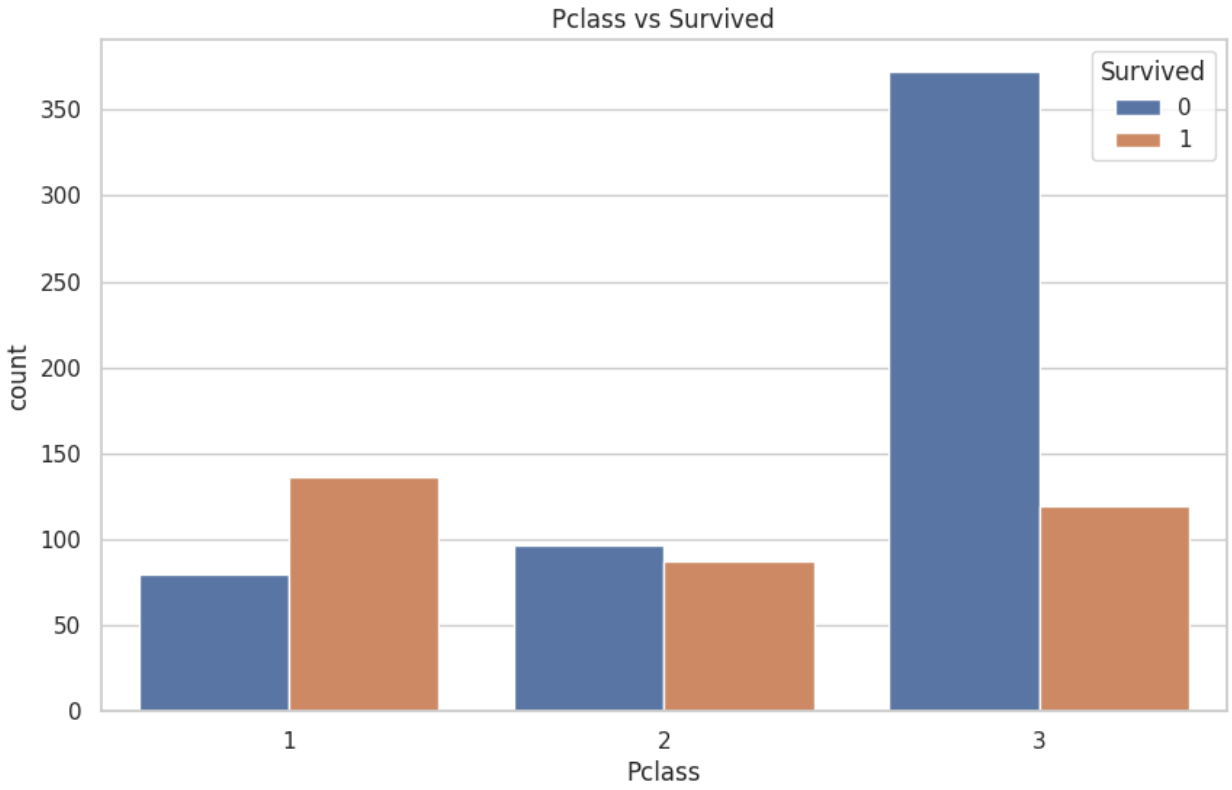

```
sns.countplot(x='Pclass', data=df)
plt.title("Passenger Class Count")
plt.show()
```



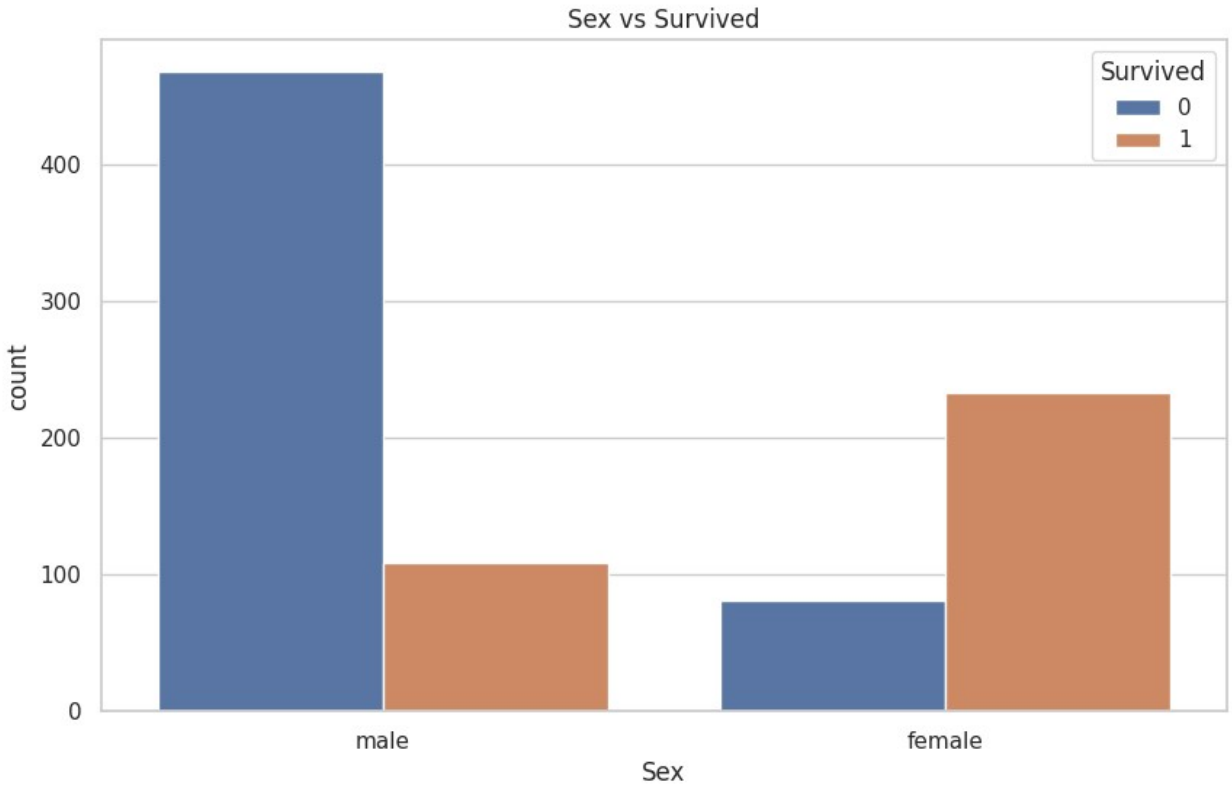

```
sns.countplot(x='Sex', data=df)
plt.title("Gender Count")
plt.show()
```



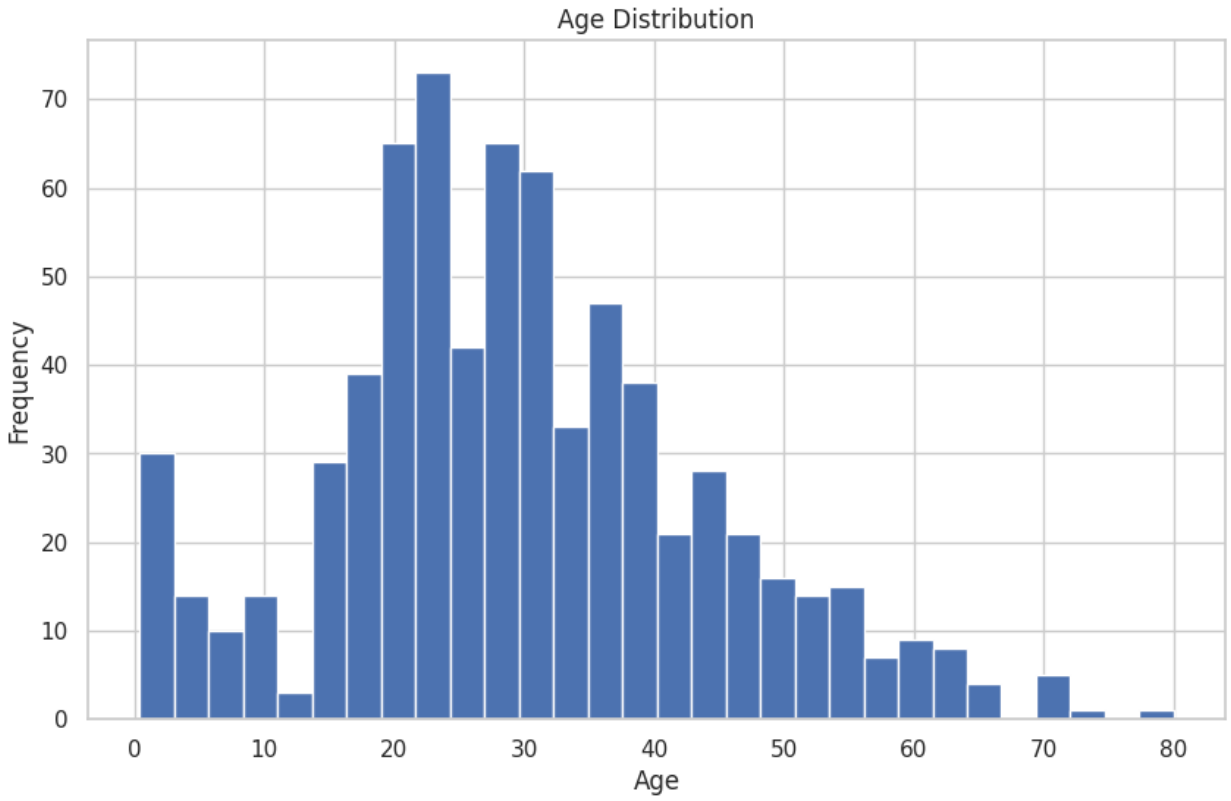

```
sns.countplot(x='Pclass', hue='Survived', data=df)
plt.title("Pclass vs Survived")
plt.show()
```



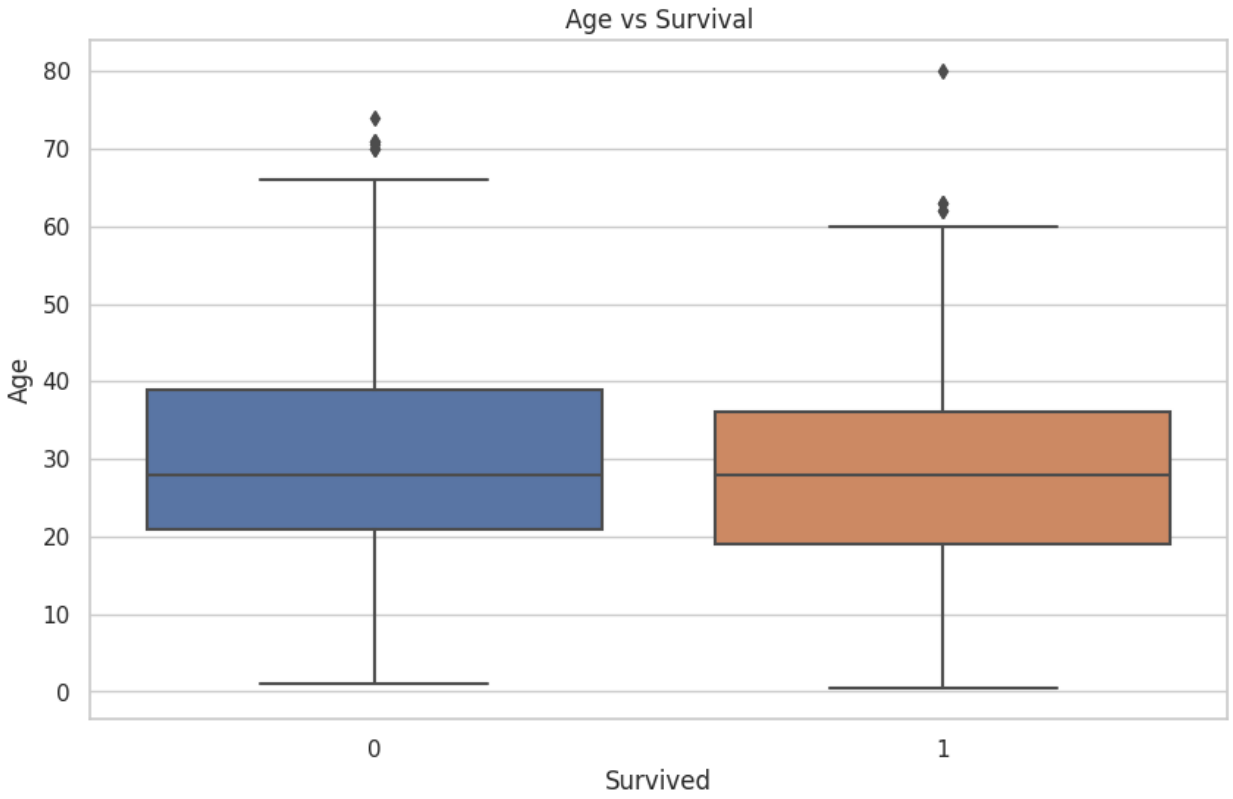

```
sns.countplot(x='Sex', hue='Survived', data=df)
plt.title("Sex vs Survived")
plt.show()
```




```
df['Age'].hist(bins=30)
plt.title("Age Distribution")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.show()
```




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




```
numeric_df = df.select_dtypes(include=['float64', 'int64'])
sns.heatmap(numeric_df.corr(), annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Correlation Heatmap")
plt.show()
```


```
Females survived more compared to males.

First class passengers had better survival chances.

Age has missing values → might require imputation.

Fare is positively correlated with survival.
```