Titanic Dataset: Exploratory Data Analysis

1. Overview
   a. Dataset: https://www.kaggle.com/competitions/titanic/data
   b. Goal: To identify patterns in survival using EDA

2. Code summary & Outputs

```
3.  a. DESCRIBE, INFO, VALUE_COUNTS
4.  print("=== Describe ===")
5.  print(df_train.describe(), "\n")
6.
7.  print("=== Value Counts ===")
8.  print("Survived:\n", df_train['Survived'].value_counts(), "\n")
9.  print("Pclass:\n", df_train['Pclass'].value_counts(), "\n")
10. print("Sex:\n", df_train['Sex'].value_counts(), "\n")
```

Output:

```
<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
None
```
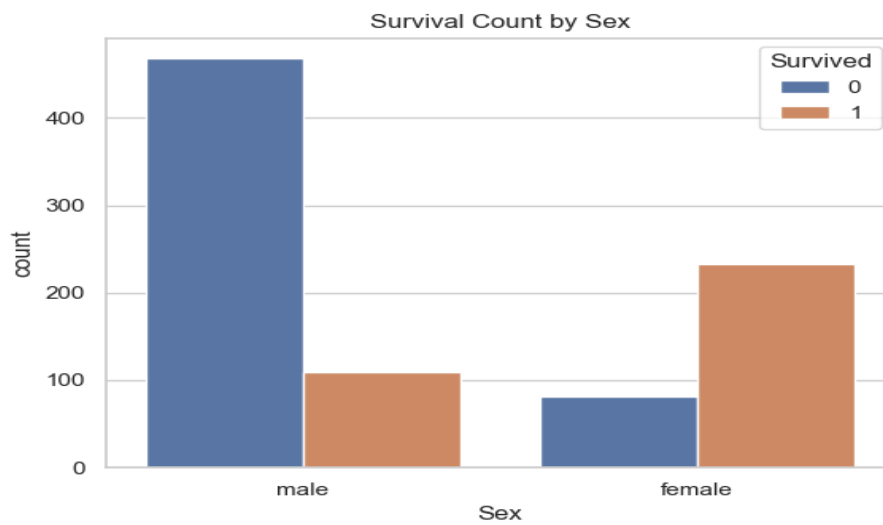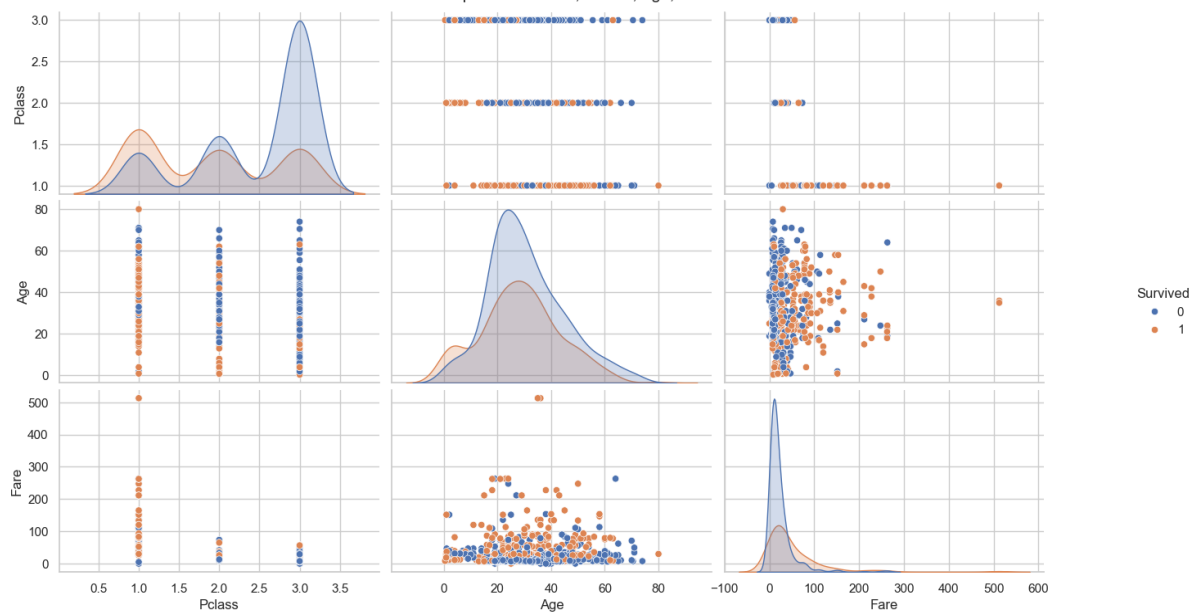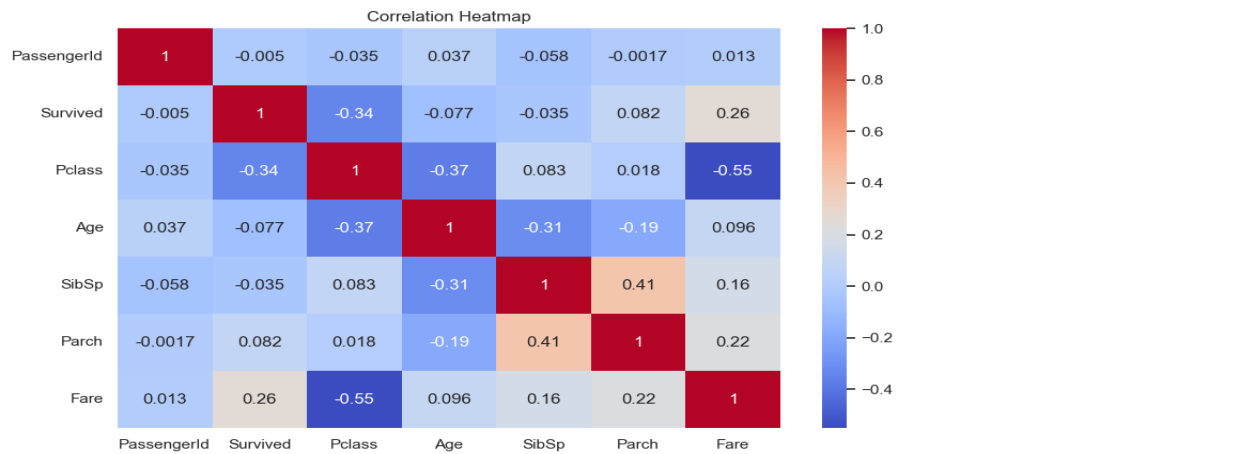
```
=== Describe ===
       PassengerId    Survived      Pclass         Age       SibSp       Parch        Fare
count   891.000000  891.000000  891.000000  714.000000  891.000000  891.000000  891.000000
mean    446.000000    0.383838    2.308642   29.699118    0.523008    0.381594   32.204208
std     257.353842    0.486592    0.836071   14.526497    1.102743    0.806057   49.693429
min       1.000000    0.000000    1.000000    0.420000    0.000000    0.000000    0.000000
25%     223.500000    0.000000    2.000000   20.125000    0.000000    0.000000    7.910400
50%     446.000000    0.000000    3.000000   28.000000    0.000000    0.000000   14.454200
75%     668.500000    1.000000    3.000000   38.000000    1.000000    0.000000   31.000000
max     891.000000    1.000000    3.000000   80.000000    8.000000    6.000000  512.329200

=== Value Counts ===
Survived:
 Survived
0    549
1    342
Name: count, dtype: int64

Pclass:
 Pclass
3    491
1    216
2    184
Name: count, dtype: int64

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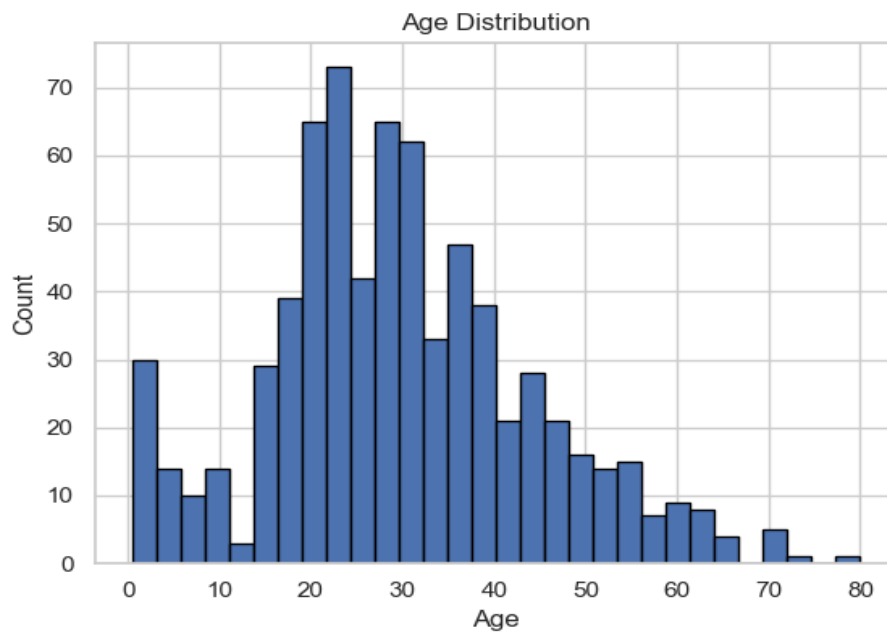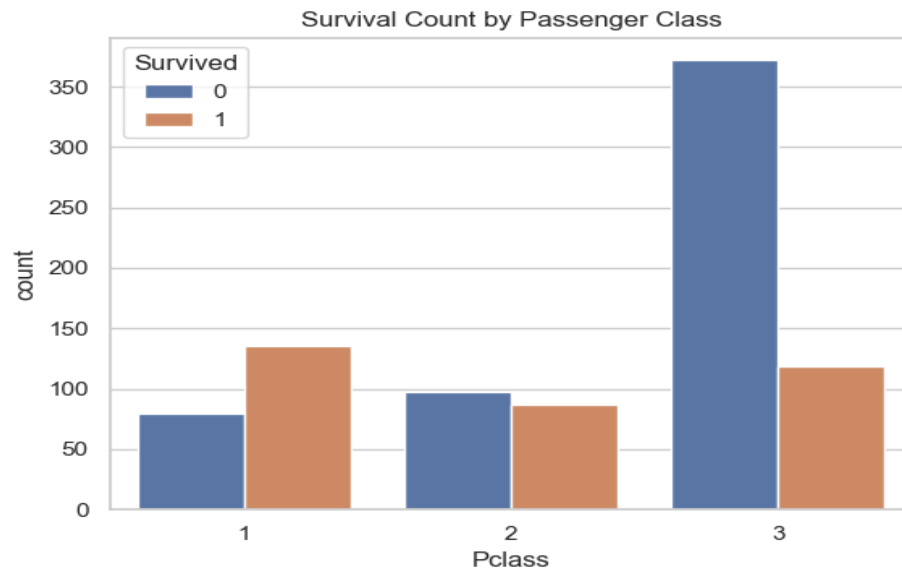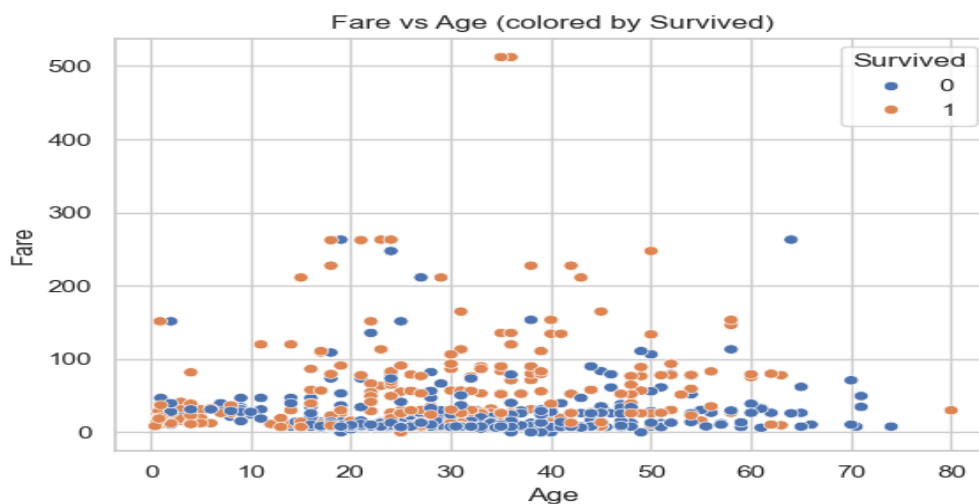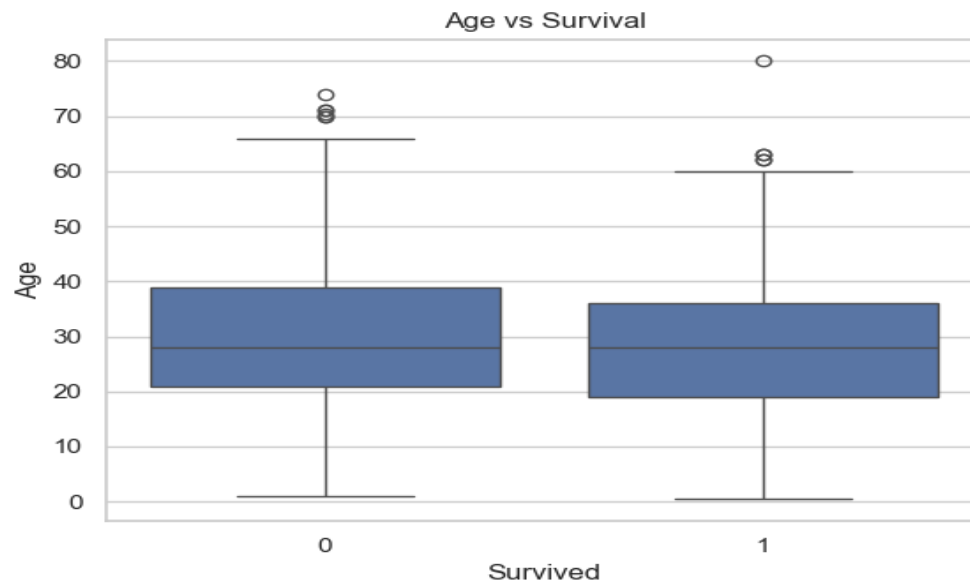
3. Visualizations

Correlation Heatmap

Survival Count by Sex

Survival Count by Passenger Class

Age Distribution

Age vs Survival

Fare vs Age (colored by Survived)

4. Observations

```
=== Observations ===

1. Age and Fare are positively correlated with survival probability.
2. Females had much higher survival rates than males.
3. 1st class passengers were more likely to survive.
4. Younger passengers (children) had better survival rates.
5. Many missing values in 'Age' and 'Cabin' columns.
```

5. Summary Of Findings

```
=== Summary of Findings ===

- Survival Rate: 38.38%
- Majority of passengers did not survive.
- High Fare, Female Gender, and 1st Class are associated with better survival chances.
- Age, Fare, and Class show meaningful trends with survival.
```