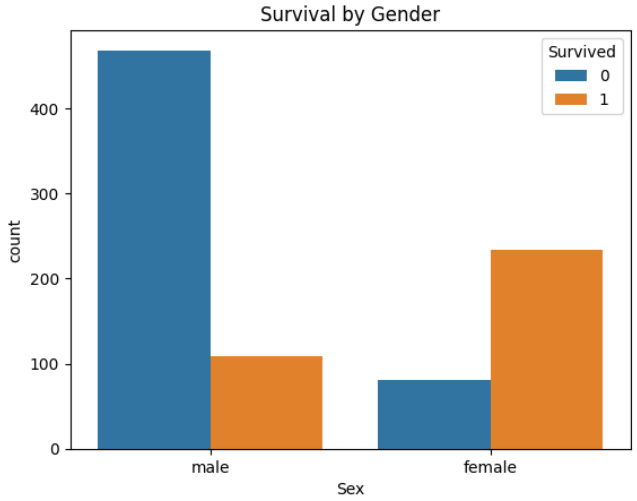
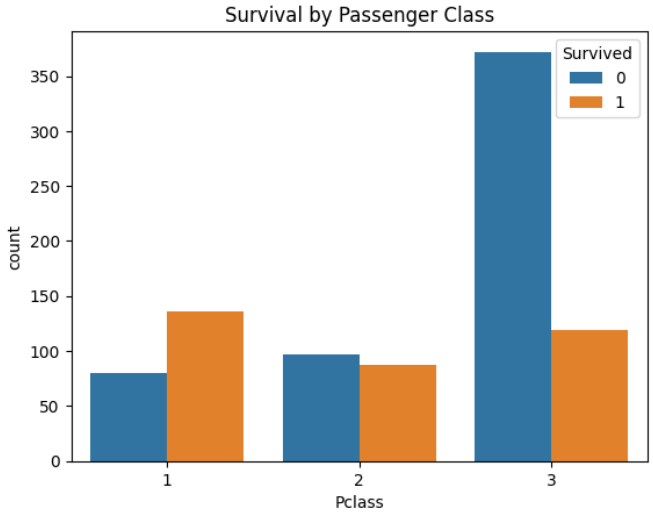
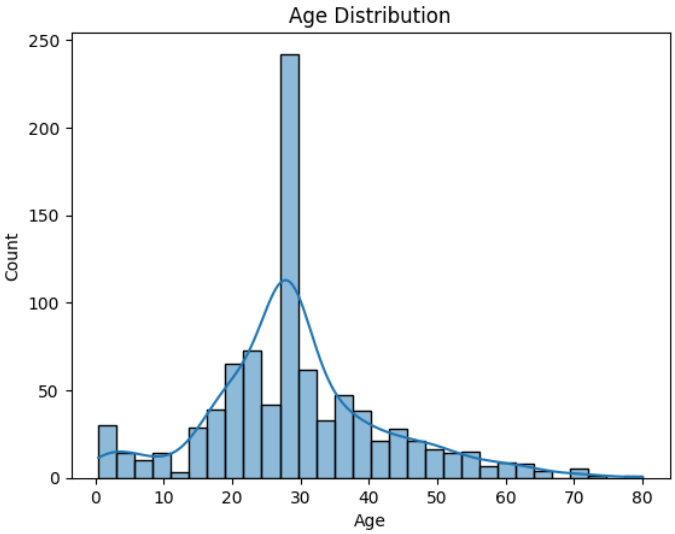
TASK – 02

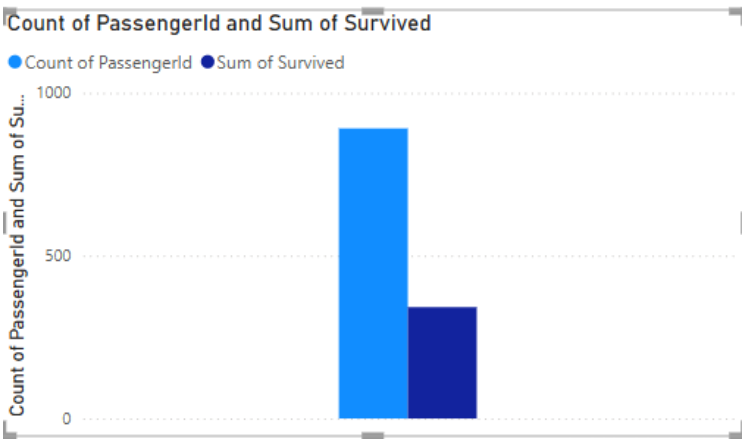
Titanic dataset by using EDA

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
df = pd.read_csv("Titanic-Dataset.csv")
print(df.head())
print(df.info())
print(df.describe())
print(df.isnull().sum())
df['Age'].fillna(df['Age'].median(), inplace=True)
df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)
if 'Cabin' in df.columns:
    df.drop(columns=['Cabin'], inplace=True)
sns.countplot(x='Survived', data=df)
plt.title("Survival Count")
plt.show()
sns.countplot(x='Sex', hue='Survived', data=df)
plt.title("Survival by Gender")
plt.show()
sns.countplot(x='Pclass', hue='Survived', data=df)
plt.title("Survival by Passenger Class")
plt.show()
sns.histplot(df['Age'], kde=True)
plt.title("Age Distribution")
plt.show()
sns.heatmap(df.corr(), annot=True, cmap="coolwarm")
plt.title("Correlation Heatmap")
plt.show()
```


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


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


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


Count of PassengerId and Sum of Survived
Count of PassengerId
Sum of Survived
Count of PassengerId and Sum of Su...
1000
500
0