

```
1  # Import necessary libraries
2  import pandas as pd
3  import numpy as np
4  import seaborn as sns
5  import matplotlib.pyplot as plt
6
7  # Load the dataset
8
9  titanic_data = pd.read_csv("/content/tested.csv")
10
11 # Display the first few rows of the dataset to understand its structure
12 print(titanic_data.head())
13
14 # Check for missing values
15 print(titanic_data.isnull().sum())
16
17 # Handle missing values
18
19 median_age = titanic_data['Age'].median()
20 titanic_data['Age'].fillna(median_age, inplace=True)
21
22 # Summary statistics
23 print(titanic_data.describe())
24
25 # Visualize the distribution of numerical variables
26 sns.histplot(titanic_data['Age'], kde=True)
27 plt.title('Distribution of Age')
28 plt.show()
29
30 sns.countplot(x='Sex', data=titanic_data)
31 plt.title('Count of Passengers by Sex')
32 plt.show()
33
34 # Explore relationships between variables
35 sns.boxplot(x='Pclass', y='Fare', data=titanic_data)
36 plt.title('Fare Distribution by Passenger Class')
37 plt.show()
38
39 sns.violinplot(x='Pclass', y='Age', hue='Survived', data=titanic_data, split=True)
40 plt.title('Age Distribution by Passenger Class and Survival')
41 plt.show()
42
43 # Correlation matrix
44 correlation_matrix = titanic_data.corr()
45 sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
46 plt.title('Correlation Matrix')
47 plt.show()
48
49 # Pairplot for exploring relationships between numerical variables
50 sns.pairplot(titanic_data, vars=['Age', 'Fare'], hue='Survived')
51 plt.title('Pairplot of Age and Fare with Survival')
52 plt.show()
```

	PassengerId	Survived	Pclass	\
0	892	0	3	
1	893	1	3	
2	894	0	2	
3	895	0	3	
4	896	1	3	

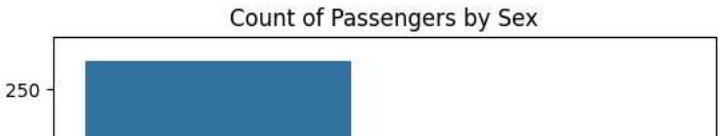
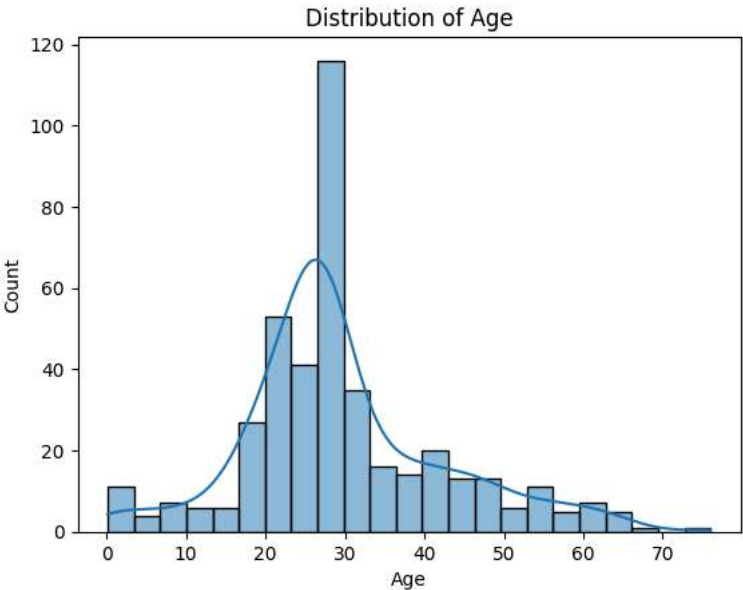
	Name	Sex	Age	SibSp	Parch	\
0	Kelly, Mr. James	male	34.5	0	0	
1	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	
2	Myles, Mr. Thomas Francis	male	62.0	0	0	
3	Wirz, Mr. Albert	male	27.0	0	0	
4	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	

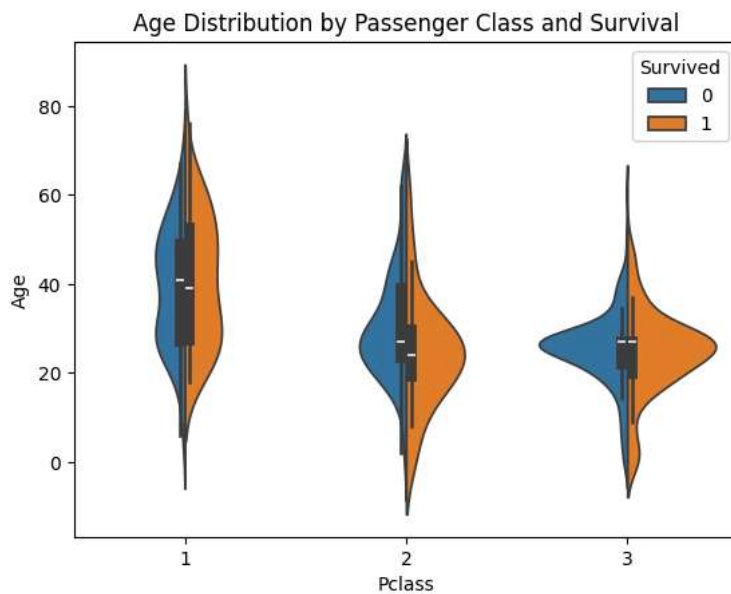
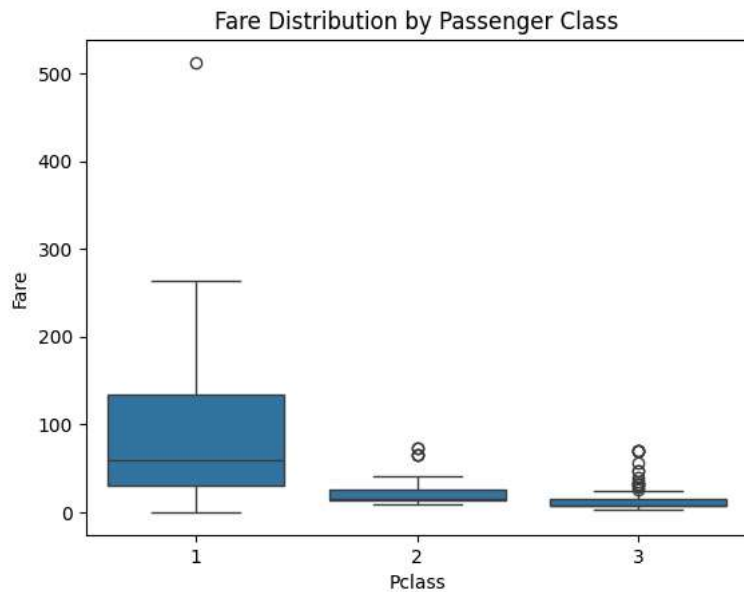
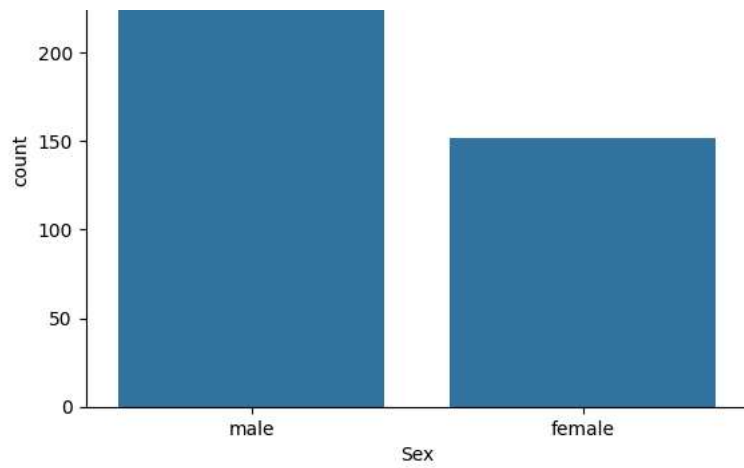
	Ticket	Fare	Cabin	Embarked
0	330911	7.8292	NaN	Q
1	363272	7.0000	NaN	S
2	240276	9.6875	NaN	Q
3	315154	8.6625	NaN	S
4	3101298	12.2875	NaN	S

PassengerId 0
Survived 0
Pclass 0
Name 0
Sex 0
Age 86
SibSp 0
Parch 0
Ticket 0
Fare 1
Cabin 327
Embarked 0
dtype: int64

	PassengerId	Survived	Pclass	Age	SibSp	\
count	418.000000	418.000000	418.000000	418.000000	418.000000	
mean	1100.500000	0.363636	2.265550	29.599282	0.447368	
std	120.810458	0.481622	0.841838	12.703770	0.896760	
min	892.000000	0.000000	1.000000	0.170000	0.000000	
25%	996.250000	0.000000	1.000000	23.000000	0.000000	
50%	1100.500000	0.000000	3.000000	27.000000	0.000000	
75%	1204.750000	1.000000	3.000000	35.750000	1.000000	
max	1309.000000	1.000000	3.000000	76.000000	8.000000	

	Parch	Fare
count	418.000000	417.000000
mean	0.392344	35.627188
std	0.981429	55.907576
min	0.000000	0.000000
25%	0.000000	7.895800
50%	0.000000	14.454200
75%	0.000000	31.500000
max	9.000000	512.329200





```
<ipython-input-3-0b919a52e660>:61: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version correlation_matrix = titanic_data.corr()
```

