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

dataset =pd.read_csv('train.csv')

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    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
```

```
dataset.shape
```

```
(891, 12)
```

```
dataset.head()
```


```
{"summary":"{\n  \"name\": \"dataset\",\n  \"rows\": 891,\n
\"fields\": [\n    {\n      \"column\": \"PassengerId\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
257,\n        \"min\": 1,\n        \"max\": 891,\n
\"num_unique_values\": 891,\n        \"samples\": [\n          710,\n
440,\n          841\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Survived\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 0,\n        \"min\": 0,\n
\"max\": 1,\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          1,\n          0\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"Pclass\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 0,\n        \"min\": 1,\n
\"max\": 3,\n        \"num_unique_values\": 3,\n        \"samples\":
[\n          3,\n          1\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
```

\"column\": \"Name\",\n   \"properties\": {\n   \"dtype\":
\"string\",\n   \"num_unique_values\": 891,\n   \"samples\":
[\n   \"Moubarek, Master. Halim Gonios (\\\"William
George\\\")\",\n   \"Kvillner, Mr. Johan Henrik Johannesson\"\n
],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n
}\n   },\n   {\n   \"column\": \"Sex\",\n   \"properties\": {\
n   \"dtype\": \"category\",\n   \"num_unique_values\": 2,\n
\"samples\": [\n   \"female\",\n   \"male\"\n   ],\
n   \"semantic_type\": \"\",\n   \"description\": \"\"\n
}\n   },\n   {\n   \"column\": \"Age\",\n   \"properties\": {\
n   \"dtype\": \"number\",\n   \"std\": 14.526497332334044,\
n   \"min\": 0.42,\n   \"max\": 80.0,\n
\"num_unique_values\": 88,\n   \"samples\": [\n   0.75,\n
22.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"SibSp\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 1,\n   \"min\": 0,\n   \"max\": 8,\n
\"num_unique_values\": 7,\n   \"samples\": [\n   1,\n
0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"Parch\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 0,\n   \"min\": 0,\n   \"max\": 6,\n
\"num_unique_values\": 7,\n   \"samples\": [\n   0,\n
1\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"Ticket\",\n   \"properties\": {\n   \"dtype\": \"string\",\n
\"num_unique_values\": 681,\n   \"samples\": [\n
\"11774\",\n   \"248740\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"Fare\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 49.693428597180905,\n
\"min\": 0.0,\n   \"max\": 512.3292,\n
\"num_unique_values\": 248,\n   \"samples\": [\n
11.2417,\n   51.8625\n   ],\n   \"semantic_type\":
\"\",\n   \"description\": \"\"\n   }\n   },\n   {\n
\"column\": \"Cabin\",\n   \"properties\": {\n   \"dtype\":
\"category\",\n   \"num_unique_values\": 147,\n
\"samples\": [\n   \"D45\",\n   \"B49\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"Embarked\",\n   \"properties\":
{\n   \"dtype\": \"category\",\n   \"num_unique_values\":
3,\n   \"samples\": [\n   \"S\",\n   \"C\"\n
],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n
}\n   }\n]\n}","type":"dataframe","variable_name":"dataset"}

```
dataset.describe()
```

{"summary":"{\n \"name\": \"dataset\",\n \"rows\": 8,\n \"fields\":
[\n   {\n   \"column\": \"PassengerId\",\n   \"properties\": {\
n   \"dtype\": \"number\",\n   \"std\": 320.8159711429856,\n

\"min\": 1.0,\n   \"max\": 891.0,\n   \"num_unique_values\": 6,\n   \"samples\": [\n   891.0,\n   446.0,\n   668.5\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"Survived\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 314.8713661874558,\n   \"min\": 0.0,\n   \"max\": 891.0,\n   \"num_unique_values\": 5,\n   \"samples\": [\n   0.3838383838383838,\n   1.0,\n   0.4865924542648585\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"Pclass\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 314.2523437079693,\n   \"min\": 0.8360712409770513,\n   \"max\": 891.0,\n   \"num_unique_values\": 6,\n   \"samples\": [\n   891.0,\n   2.308641975308642,\n   3.0\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"Age\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 242.9056731818781,\n   \"min\": 0.42,\n   \"max\": 714.0,\n   \"num_unique_values\": 8,\n   \"samples\": [\n   29.69911764705882,\n   28.0,\n   714.0\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"SibSp\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 314.4908277465442,\n   \"min\": 0.0,\n   \"max\": 891.0,\n   \"num_unique_values\": 6,\n   \"samples\": [\n   891.0,\n   0.5230078563411896,\n   8.0\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"Parch\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 314.65971717879,\n   \"min\": 0.0,\n   \"max\": 891.0,\n   \"num_unique_values\": 5,\n   \"samples\": [\n   0.38159371492704824,\n   6.0,\n   0.8060572211299559\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"Fare\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 330.6256632228577,\n   \"min\": 0.0,\n   \"max\": 891.0,\n   \"num_unique_values\": 8,\n   \"samples\": [\n   32.204207968574636,\n   14.4542,\n   891.0\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   }\n]\n}","type":"dataframe"}

```
dataset.isna().sum()
```

```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
```

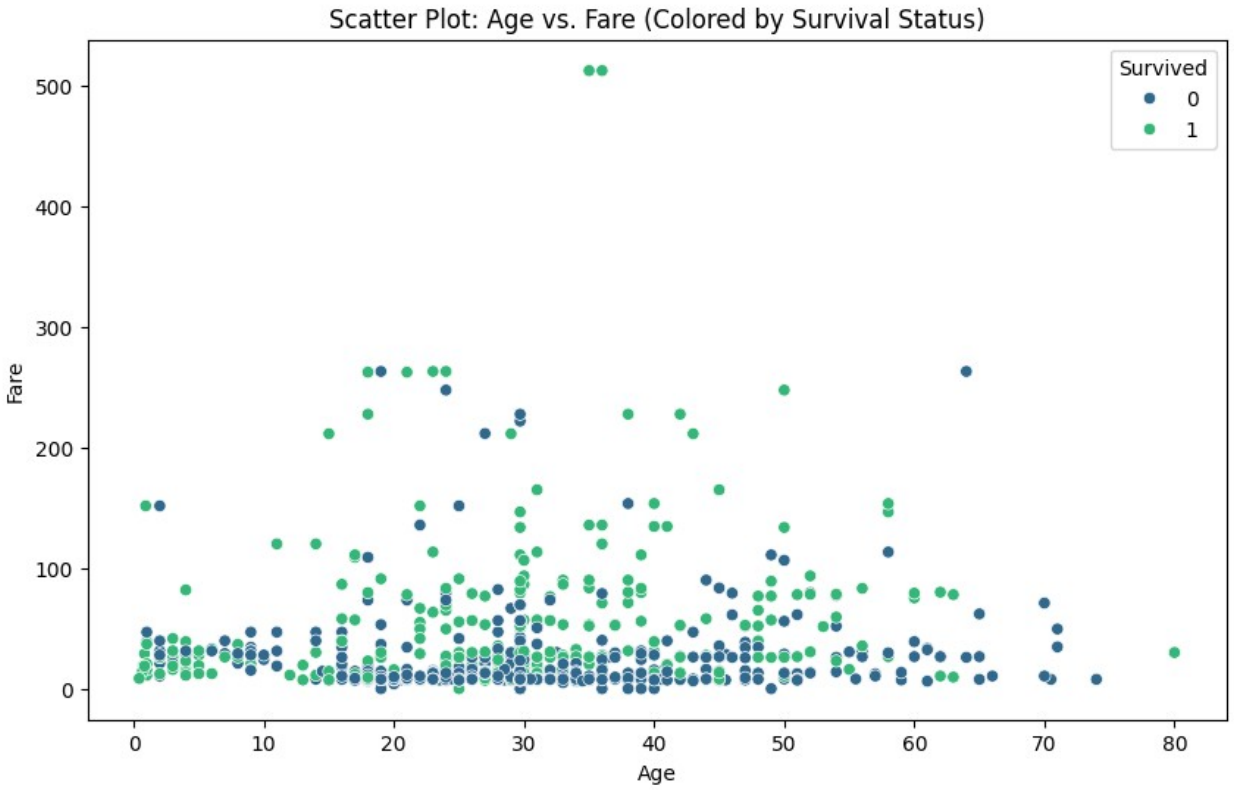
```
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

```
dataset["Age"] = dataset["Age"].fillna(dataset["Age"].mean())
```

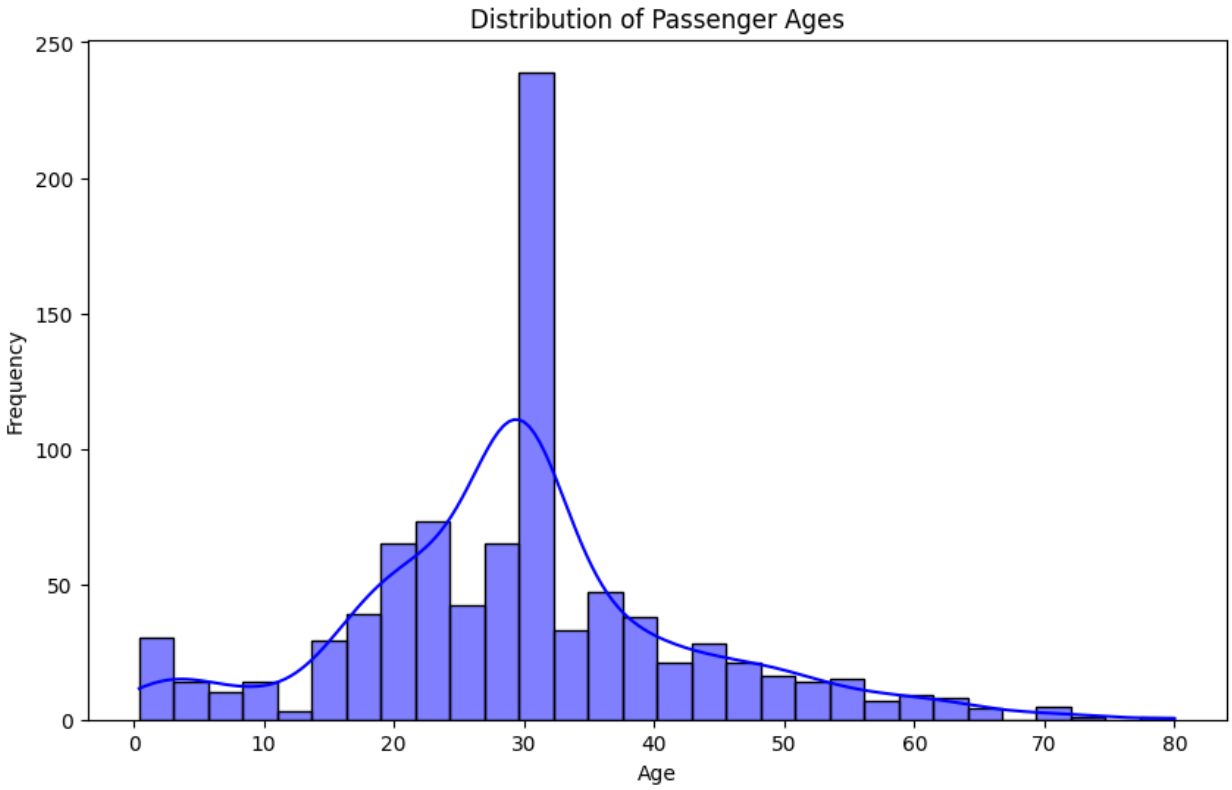
```
dataset.isna().sum()
```

```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age              0
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

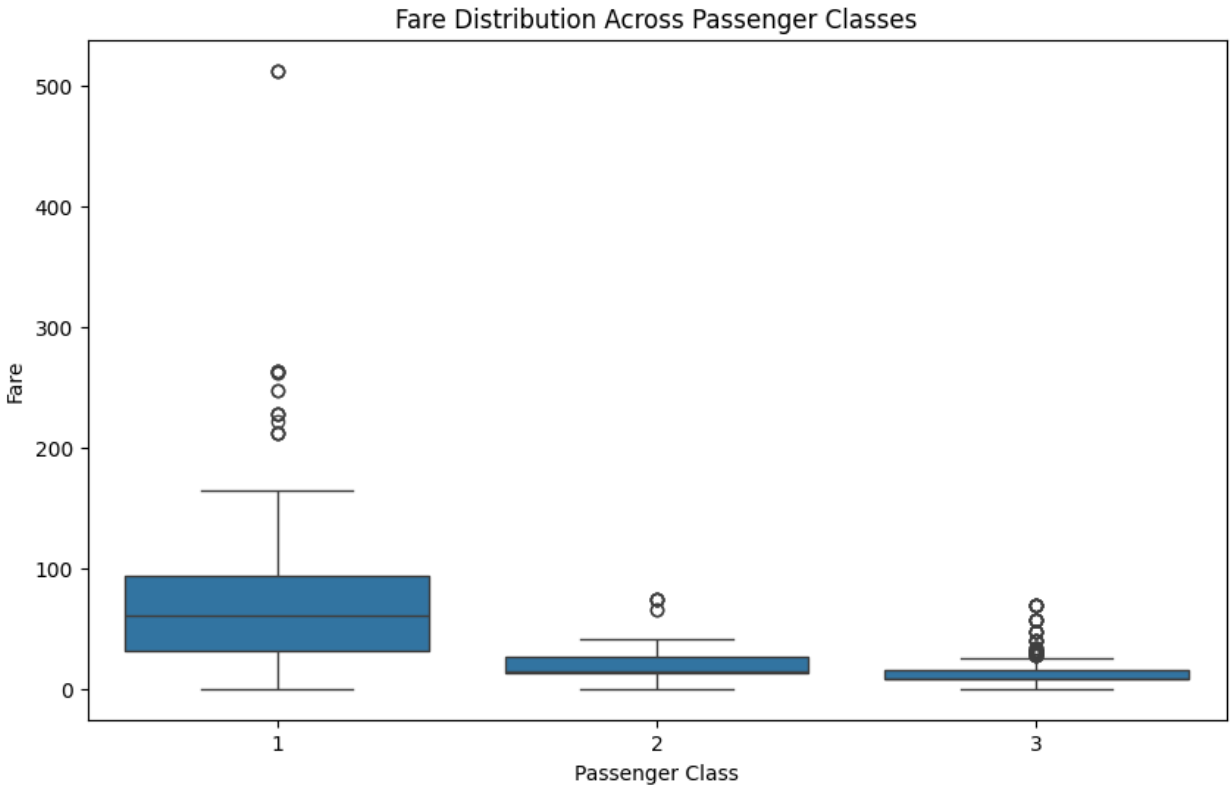
```
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Age', y='Fare', hue='Survived', data=dataset,
palette='viridis')
plt.title('Scatter Plot: Age vs. Fare (Colored by Survival Status)')
plt.xlabel('Age')
plt.ylabel('Fare')
plt.legend(title='Survived')
plt.show()
```



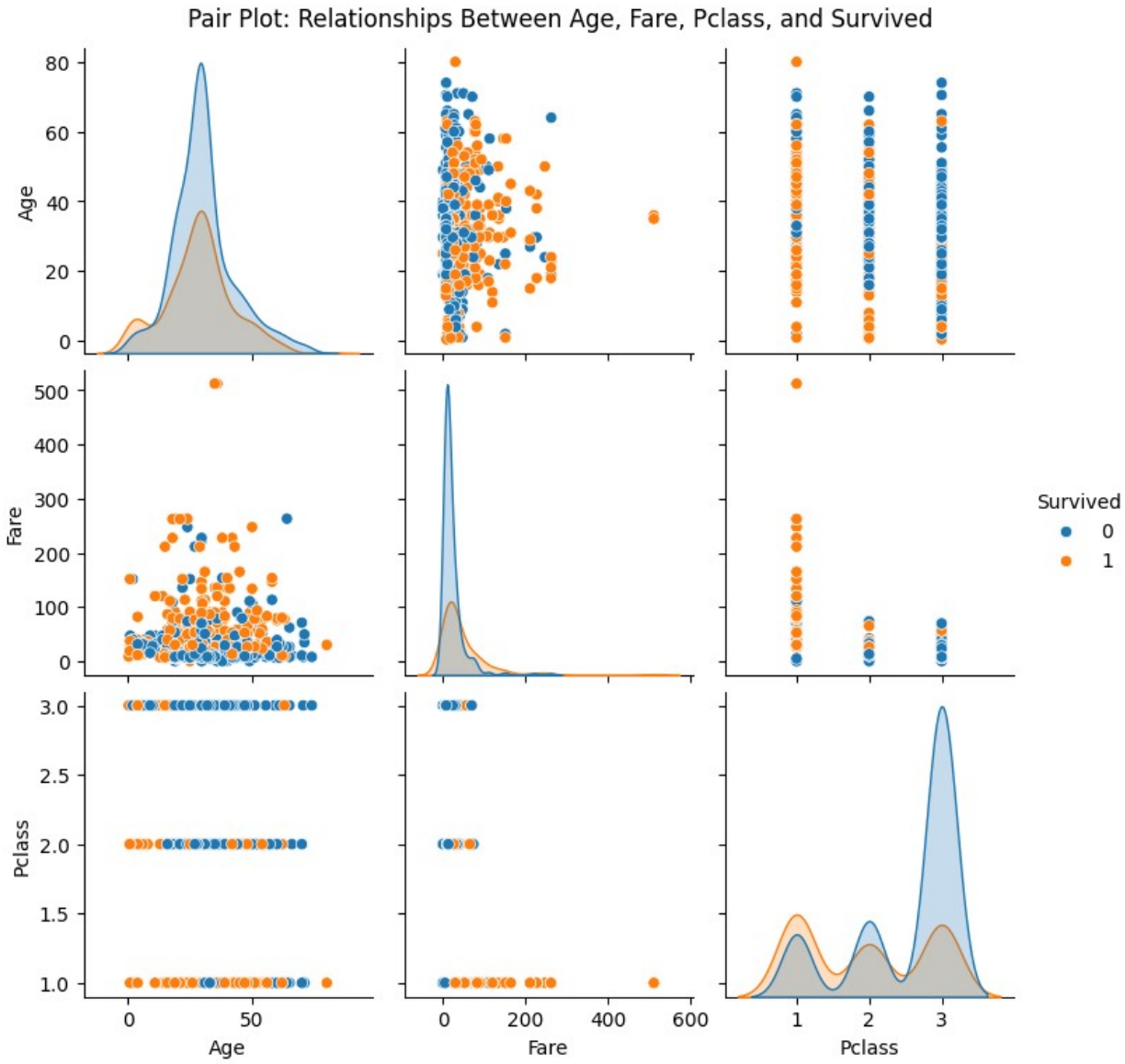

```
plt.figure(figsize=(10, 6))
sns.histplot(dataset['Age'].dropna(), bins=30, kde=True, color='blue')
plt.title('Distribution of Passenger Ages')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```



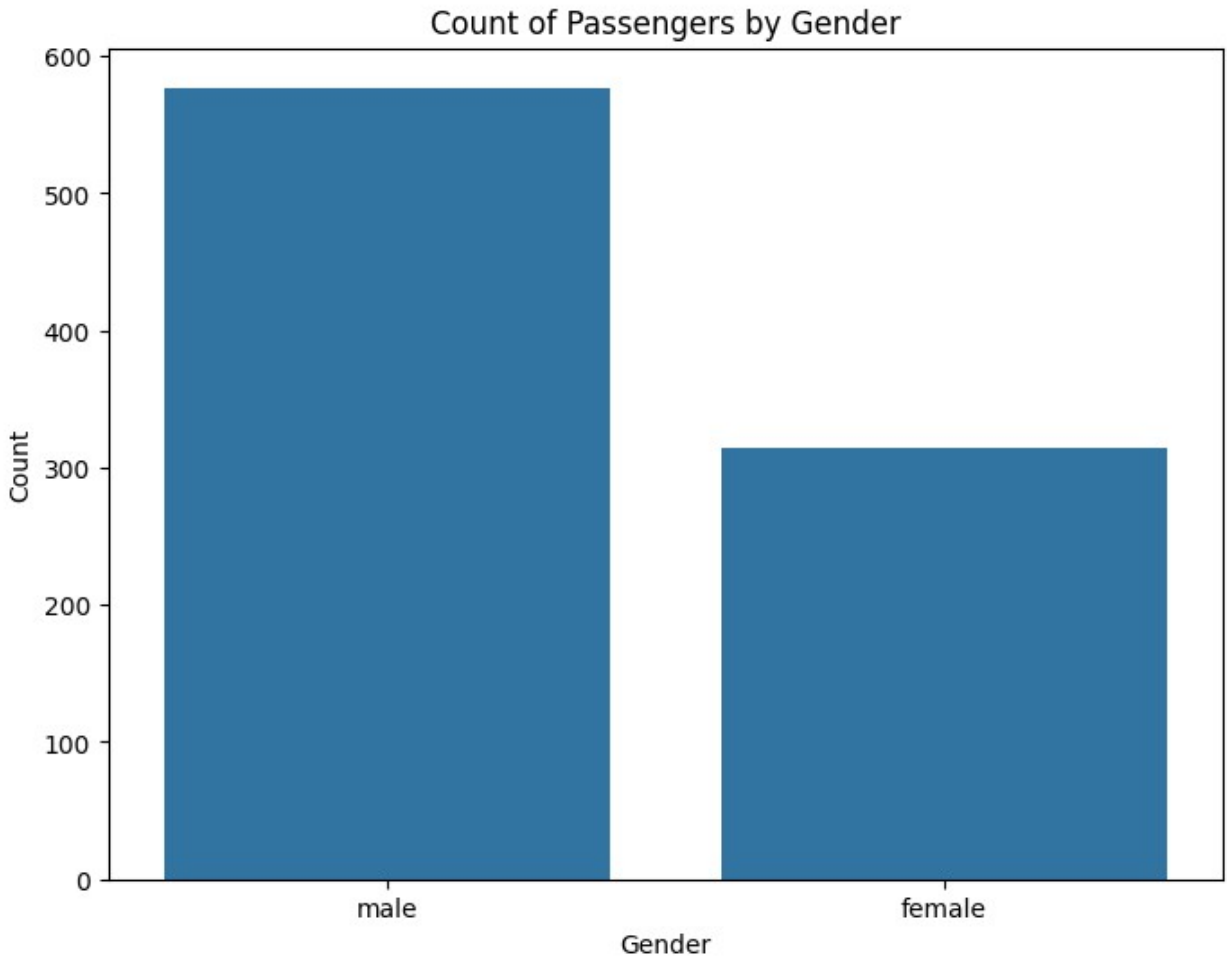

```
plt.figure(figsize=(10, 6))
sns.boxplot(x='Pclass', y='Fare', data=dataset)
plt.title('Fare Distribution Across Passenger Classes')
plt.xlabel('Passenger Class')
plt.ylabel('Fare')
plt.show()
```



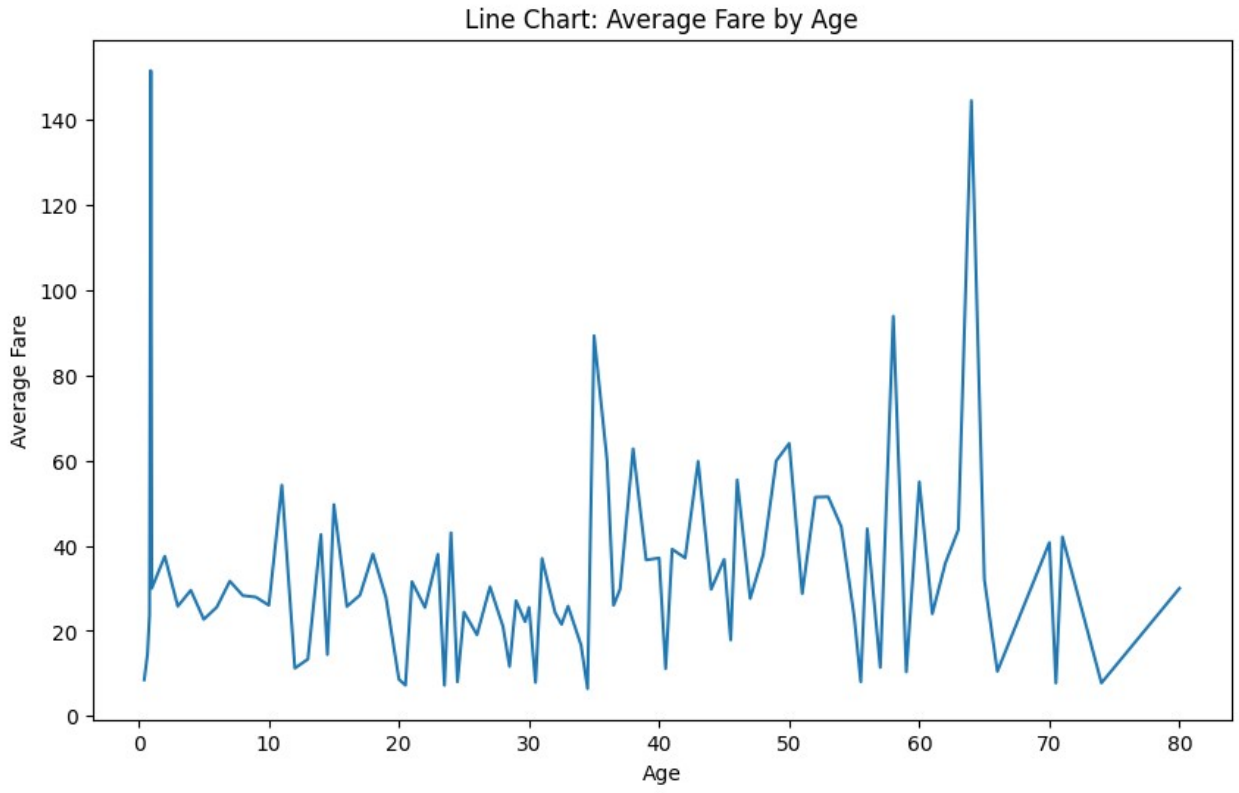

```
sns.pairplot(dataset[['Age', 'Fare', 'Pclass', 'Survived']],
hue='Survived')
plt.suptitle('Pair Plot: Relationships Between Age, Fare, Pclass, and
Survived', y=1.02)
plt.show()
```



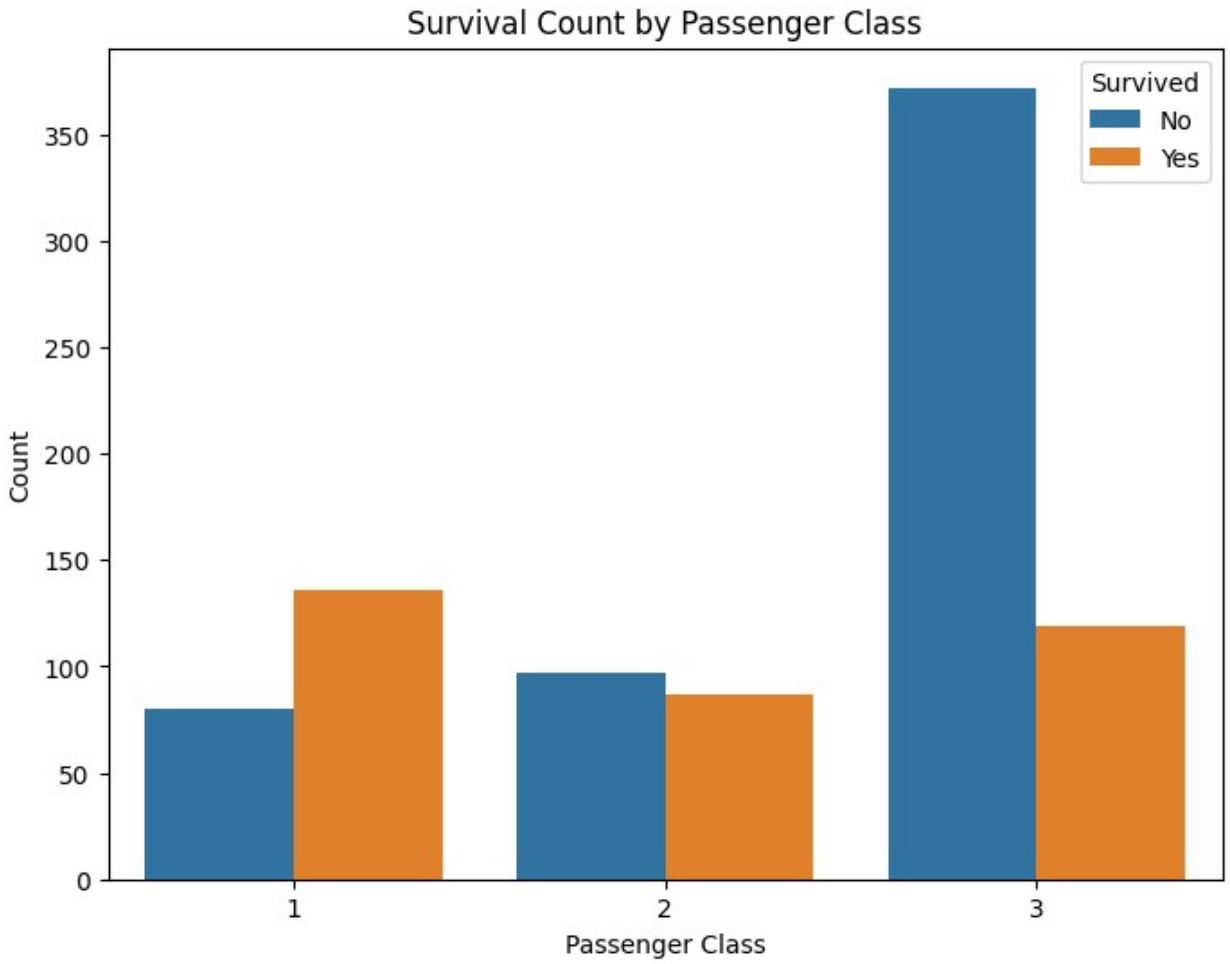

```
plt.figure(figsize=(8, 6))
sns.countplot(x='Sex', data=dataset)
plt.title('Count of Passengers by Gender')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.show()
```




```
plt.figure(figsize=(10, 6))
dataset.groupby('Age')['Fare'].mean().plot(kind='line')
plt.title('Line Chart: Average Fare by Age')
plt.xlabel('Age')
plt.ylabel('Average Fare')
plt.show()
```




```
plt.figure(figsize=(8, 6))
sns.countplot(x='Pclass', hue='Survived', data=dataset)
plt.title('Survival Count by Passenger Class')
plt.xlabel('Passenger Class')
plt.ylabel('Count')
plt.legend(title='Survived', labels=['No', 'Yes'])
plt.show()
```


Survival Count by Passenger Class
Survived
No
Yes
Count
0
50
100
150
200
250
300
350
1
2
3
Passenger Class