

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

df=pd.read_csv(r"C:\Users\dell\Downloads\Titanic-Dataset.csv")
df
```

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	
..	
886	887	0	2	
887	888	1	1	
888	889	0	3	
889	890	1	1	
890	891	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			
..
...			
886	Montvila, Rev. Juozas	male	27.0
0			
887	Graham, Miss. Margaret Edith	female	19.0
0			
888	Johnston, Miss. Catherine Helen "Carrie"	female	NaN
1			
889	Behr, Mr. Karl Howell	male	26.0
0			
890	Dooley, Mr. Patrick	male	32.0
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
...
886	0	211536	13.0000	NaN	S
887	0	112053	30.0000	B42	S
888	2	W./C. 6607	23.4500	NaN	S
889	0	111369	30.0000	C148	C
890	0	370376	7.7500	NaN	Q

[891 rows x 12 columns]

Top five rows and columns of dataset

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	Cummings, 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

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()

	PassengerId	Survived	Pclass	Age	SibSp	\
count	891.000000	891.000000	891.000000	714.000000	891.000000	
mean	446.000000	0.383838	2.308642	29.699118	0.523008	
std	257.353842	0.486592	0.836071	14.526497	1.102743	
min	1.000000	0.000000	1.000000	0.420000	0.000000	
25%	223.500000	0.000000	2.000000	20.125000	0.000000	
50%	446.000000	0.000000	3.000000	28.000000	0.000000	
75%	668.500000	1.000000	3.000000	38.000000	1.000000	
max	891.000000	1.000000	3.000000	80.000000	8.000000	

	Parch	Fare
count	891.000000	891.000000
mean	0.381594	32.204208
std	0.806057	49.693429
min	0.000000	0.000000
25%	0.000000	7.910400
50%	0.000000	14.454200
75%	0.000000	31.000000
max	6.000000	512.329200

df.value_counts()

PassengerId	Survived	Pclass	Name
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)
7	0	1	McCarthy, Mr. Timothy J
11	1	3	Sandstrom, Miss. Marguerite Rut


```
False
890      False      False      False      False      False      False      False      False
False
```

	Fare	Cabin	Embarked
0	False	True	False
1	False	False	False
2	False	True	False
3	False	False	False
4	False	True	False
..
886	False	True	False
887	False	False	False
888	False	True	False
889	False	False	False
890	False	True	False

```
[891 rows x 12 columns]
```

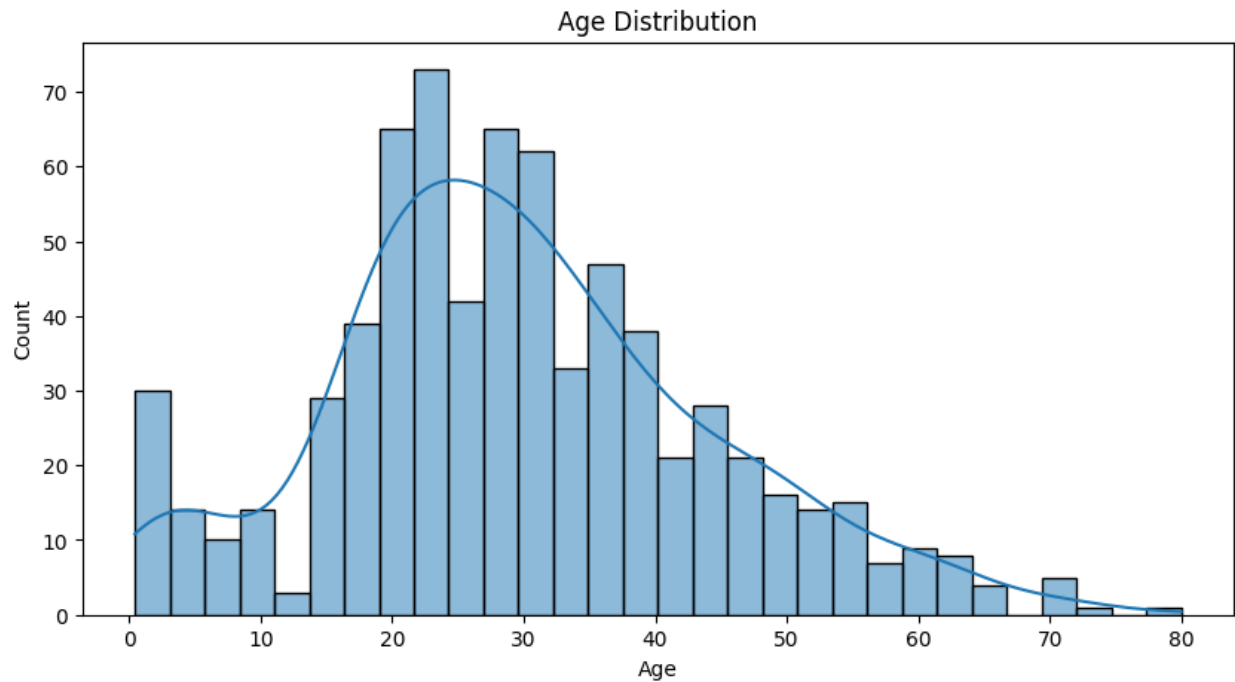
helping us to find the total number of null values in each column

```
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
```

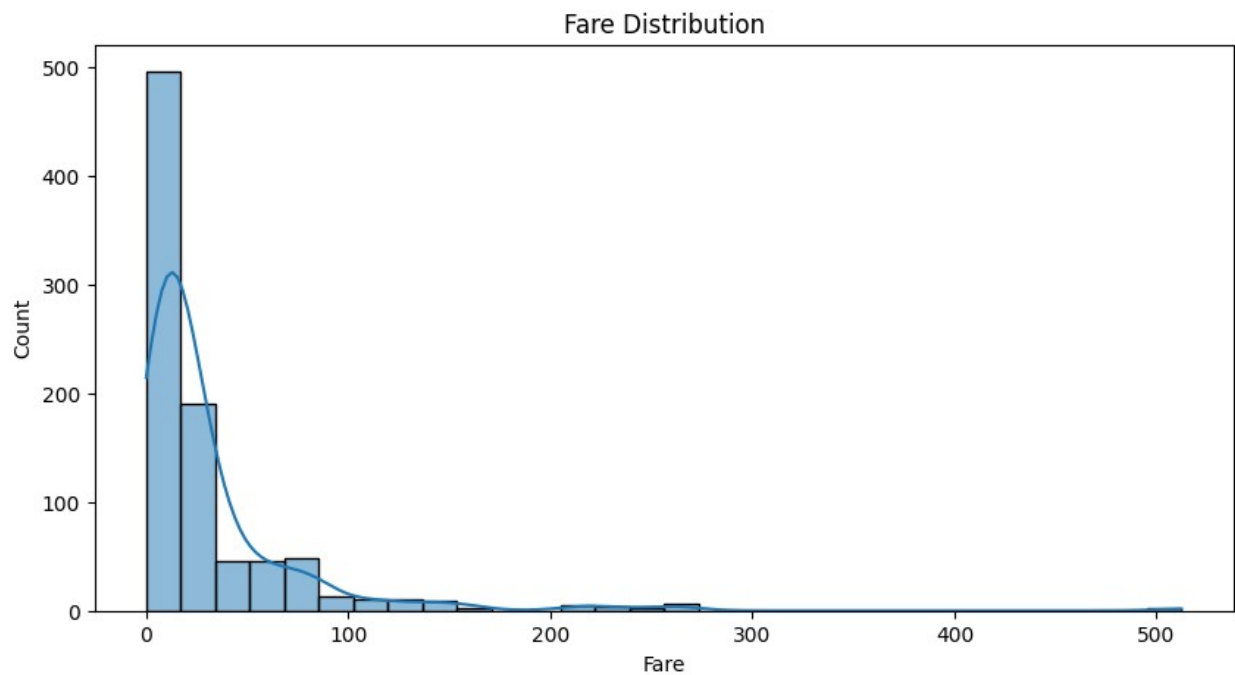
Age distribution

```
plt.figure(figsize=(10,5))
sns.histplot(df['Age'].dropna(), bins=30, kde=True)
plt.title("Age Distribution")
plt.show()
```



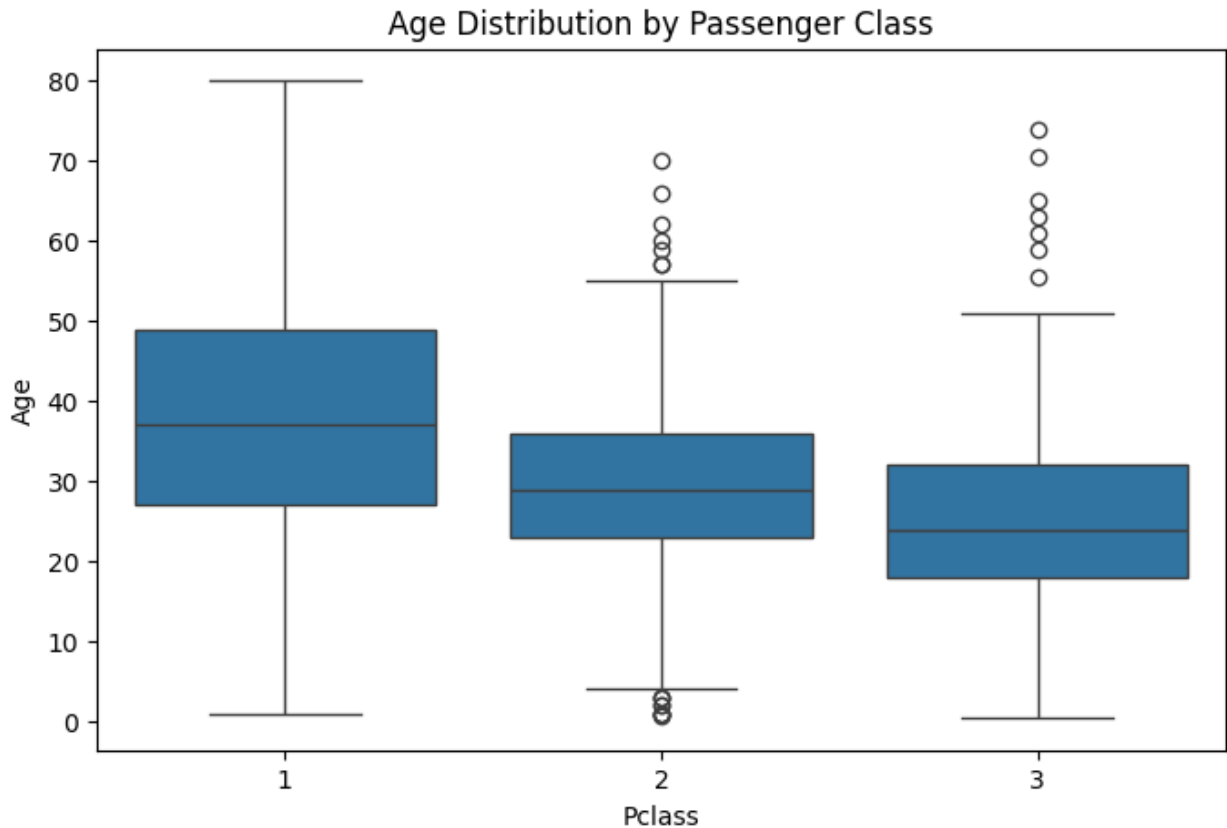
Fare distribution

```
plt.figure(figsize=(10,5))
sns.histplot(df['Fare'], bins=30, kde=True)
plt.title("Fare Distribution")
plt.show()
```



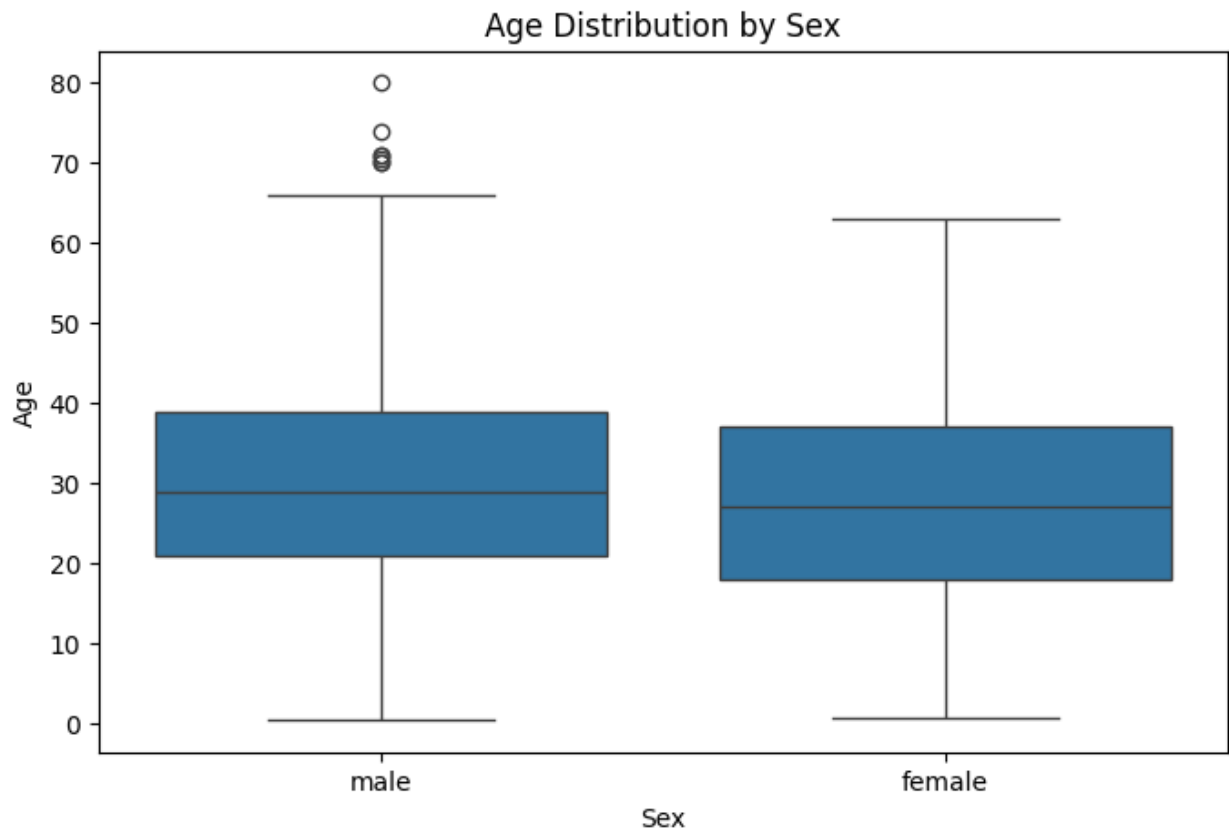
Age Distribution by passenger class

```
plt.figure(figsize=(8,5))
sns.boxplot(x='Pclass', y='Age', data=df)
plt.title("Age Distribution by Passenger Class")
plt.show()
```

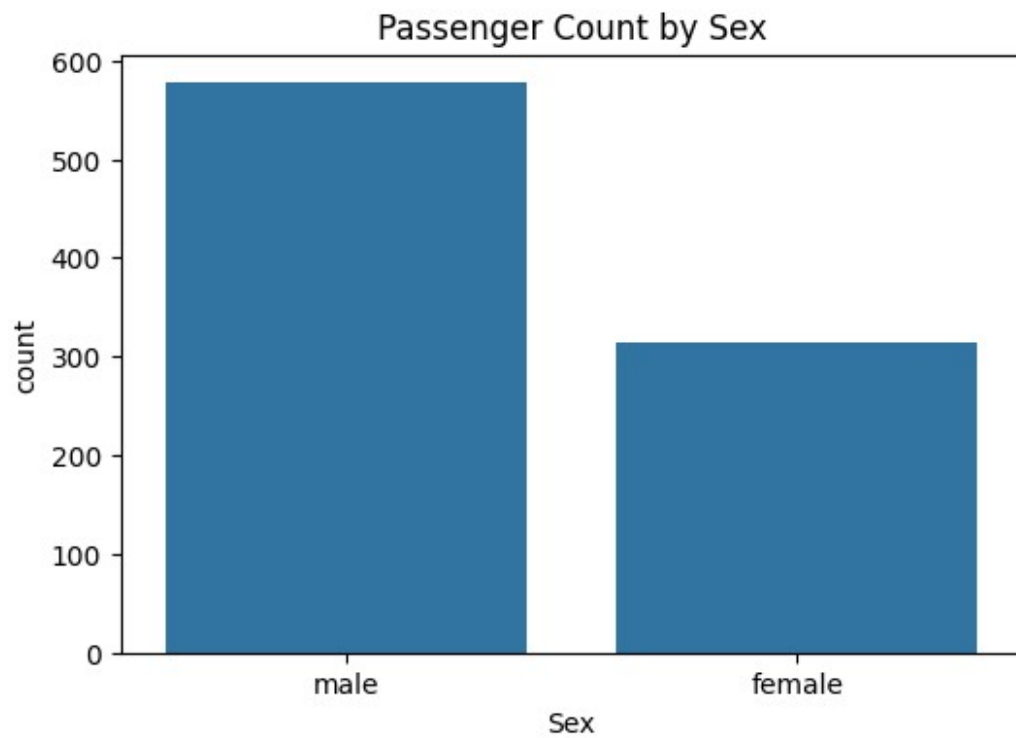


Age distribution by Sex

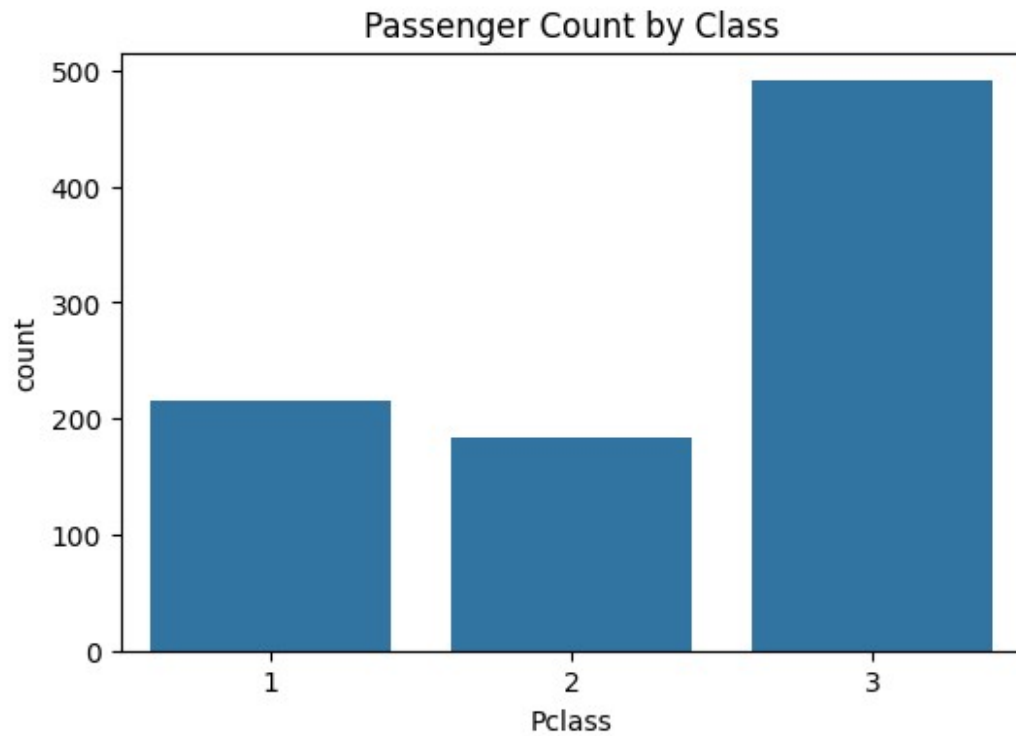
```
plt.figure(figsize=(8,5))
sns.boxplot(x='Sex', y='Age', data=df)
plt.title("Age Distribution by Sex")
plt.show()
```



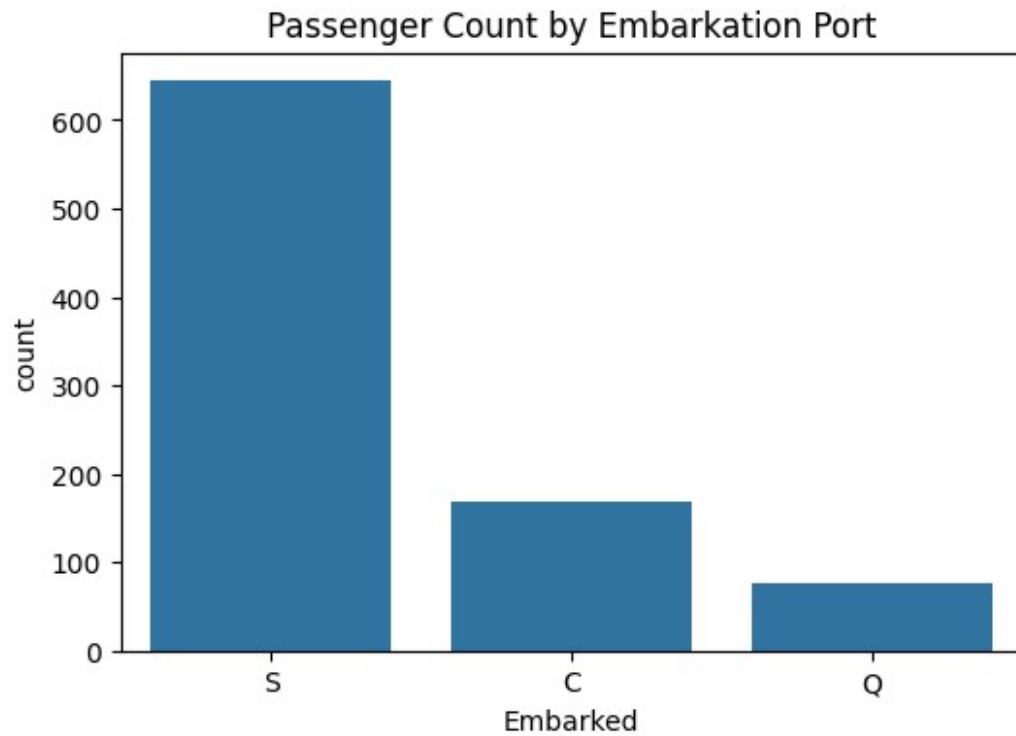
```
# Countplot of Sex
plt.figure(figsize=(6,4))
sns.countplot(x='Sex', data=df)
plt.title("Passenger Count by Sex")
plt.show()
```

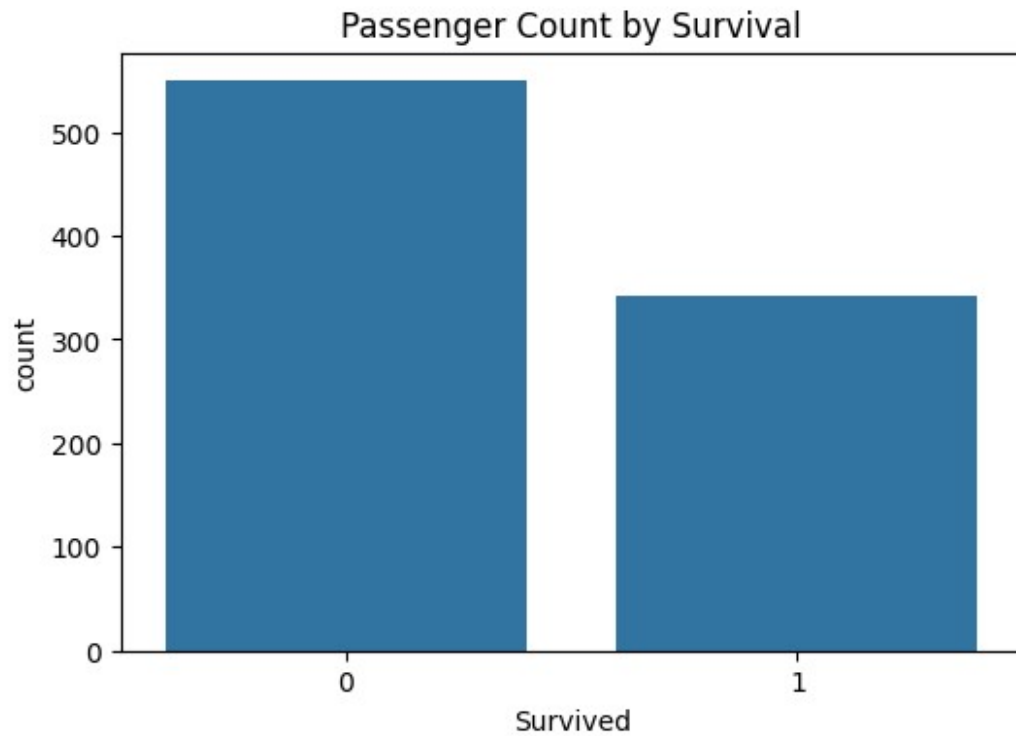
```
# Countplot of Pclass
plt.figure(figsize=(6,4))
sns.countplot(x='Pclass', data=df)
plt.title("Passenger Count by Class")
plt.show()
```



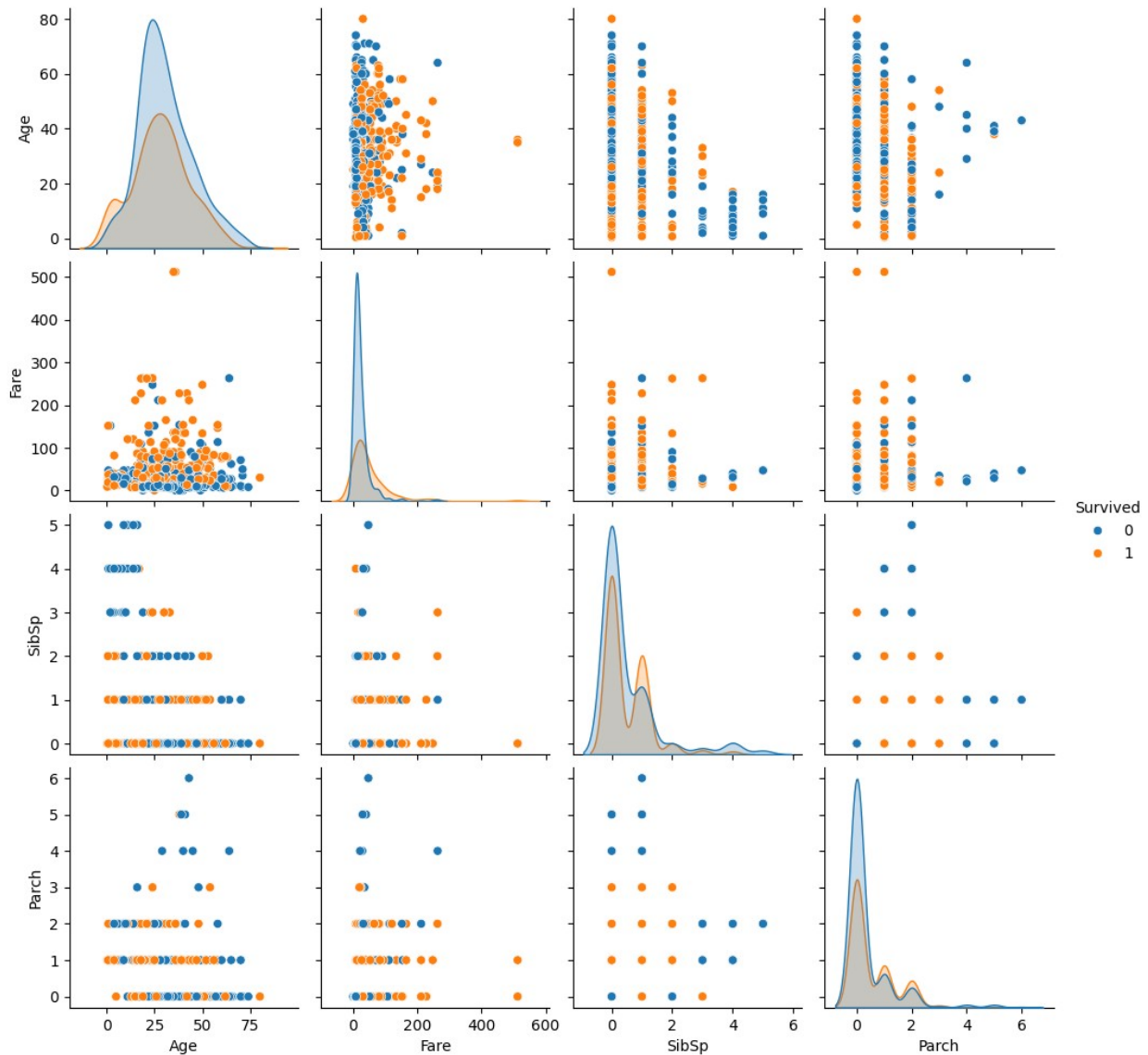
```
# Countplot of Embarked  
plt.figure(figsize=(6,4))  
sns.countplot(x='Embarked', data=df)  
plt.title("Passenger Count by Embarkation Port")  
plt.show()
```



```
# Countplot of Survived  
plt.figure(figsize=(6,4))  
sns.countplot(x='Survived', data=df)  
plt.title("Passenger Count by Survival")  
plt.show()
```

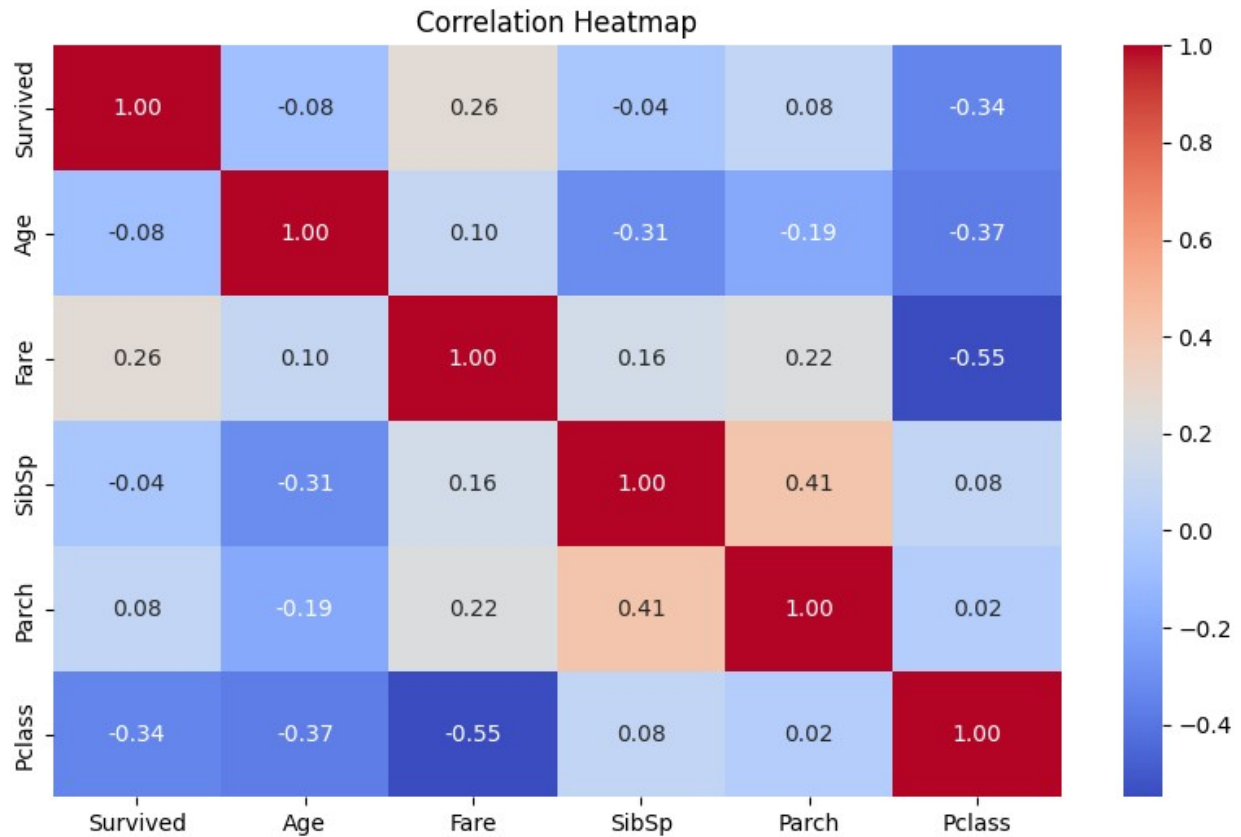


```
# Pairplot for numeric features with hue=Survived  
sns.pairplot(  
    df[['Survived', 'Age', 'Fare', 'SibSp', 'Parch']].dropna(),  
    hue='Survived',  
    diag_kind='kde')  
plt.show()
```



```
# Correlation heatmap
plt.figure(figsize=(10,6))
corr = df[['Survived', 'Age', 'Fare', 'SibSp', 'Parch', 'Pclass']].corr()

sns.heatmap(corr, annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Correlation Heatmap")
plt.show()
```



```
# Survival rate by Sex
print("Survival rate by Sex:")
print(df.groupby('Sex')['Survived'].mean())

# Survival rate by Pclass
print("\nSurvival rate by Passenger Class:")
print(df.groupby('Pclass')['Survived'].mean())

# Survival rate by Embarked
print("\nSurvival rate by Embarked Port:")
print(df.groupby('Embarked')['Survived'].mean())

# Average Age and Fare by Pclass and Sex
print("\nAverage Age and Fare by Class and Sex:")
print(df.groupby(['Pclass', 'Sex'])[['Age', 'Fare']].mean())
```

Survival rate by Sex:

```
Sex
female    0.742038
male      0.188908
Name: Survived, dtype: float64
```

Survival rate by Passenger Class:

```
Pclass
```

```
1    0.629630
2    0.472826
3    0.242363
```

Name: Survived, dtype: float64

Survival rate by Embarked Port:
Embarked

```
C    0.553571
Q    0.389610
S    0.336957
```

Name: Survived, dtype: float64

Average Age and Fare by Class and Sex:

		Age	Fare
Pclass	Sex		
1	female	34.611765	106.125798
	male	41.281386	67.226127
2	female	28.722973	21.970121
	male	30.740707	19.741782
3	female	21.750000	16.118810
	male	26.507589	12.661633