```
import pandas as pd
df = pd.read_csv("Titanic-Dataset.csv")
```

```
print(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  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

   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
```

```
print(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
None
```

```
print(df.describe())
```

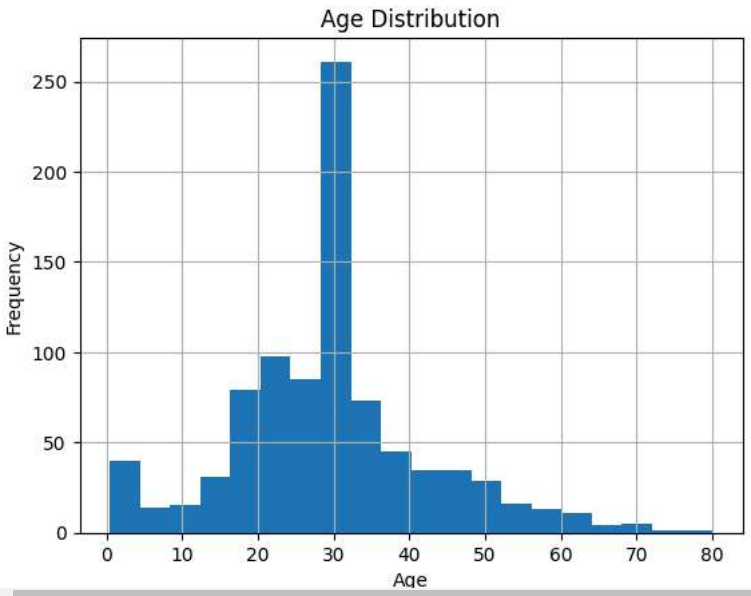
```
       PassengerId    Survived      Pclass         Age       SibSp  \
count   891.000000  891.000000  891.000000  714.000000  891.000000
mean    446.000000    0.383838    2.308642   29.699118    0.523008
std     257.353842    0.486592    0.836071   14.526497    1.102743
min       1.000000    0.000000    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['Age'] = df['Age'].fillna(df['Age'].mean())
```

```
df= df.drop(['Ticket','Cabin'],axis=1)
```

```
import matplotlib.pyplot as plt
df['Age'].hist(bins=20)
```

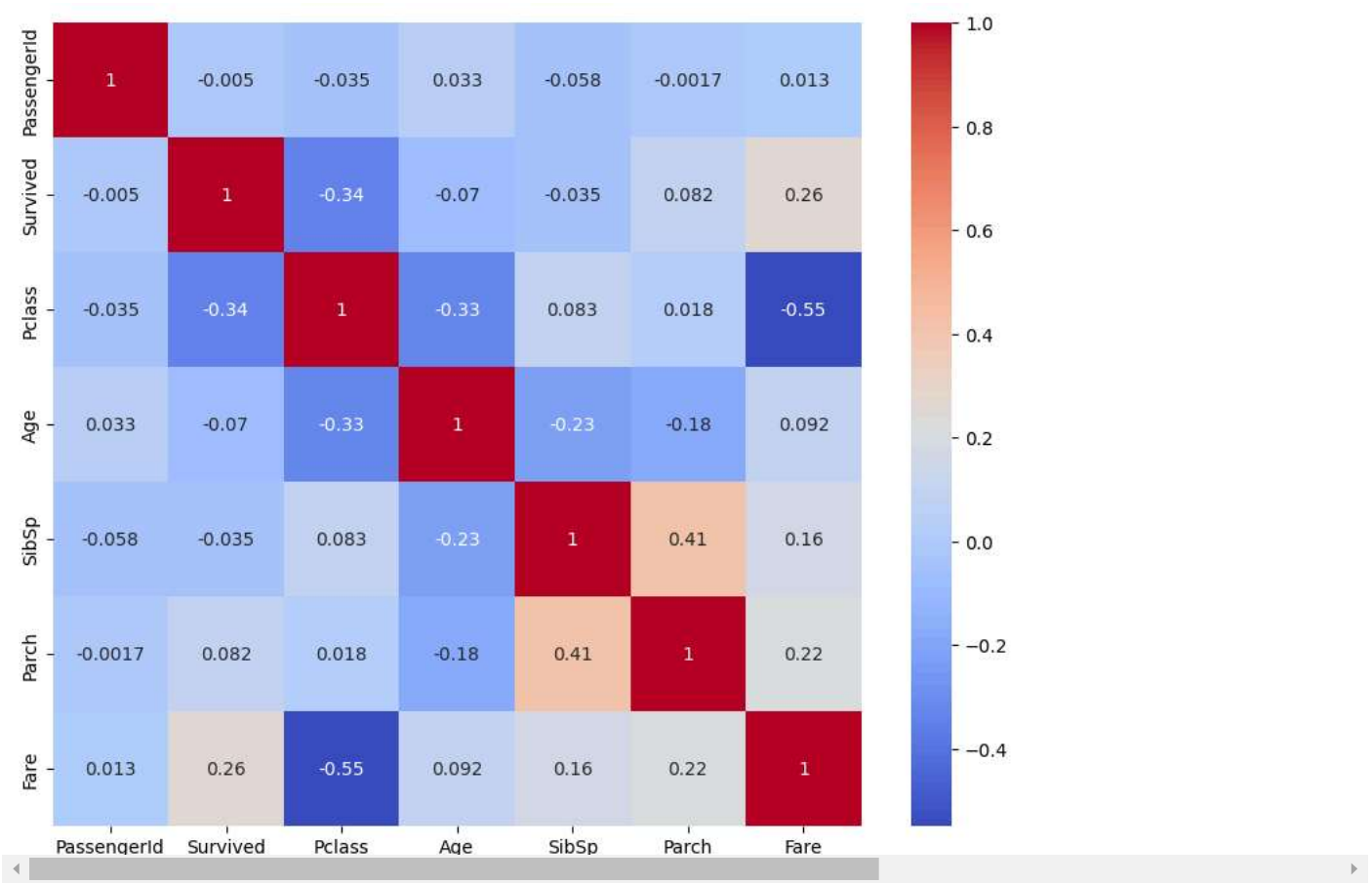
```
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.title('Age Distribution')
plt.show()
```



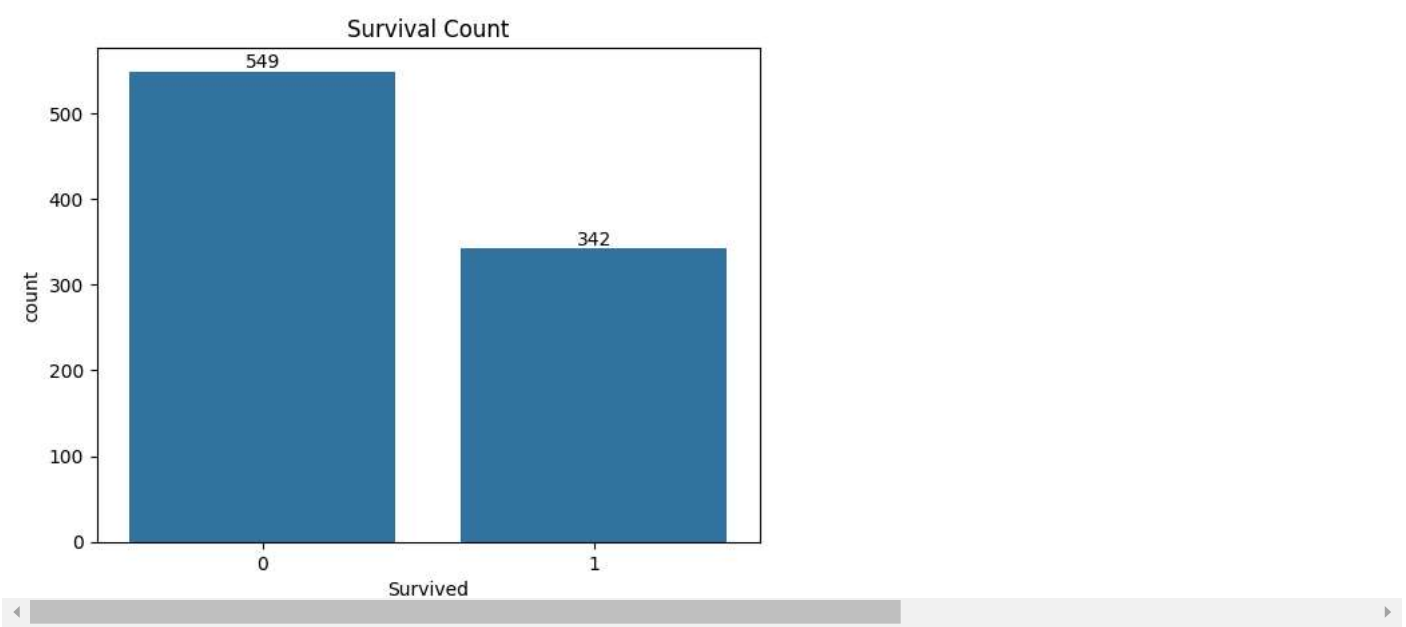

```
#Survival rate by Gender
print(df.groupby('Sex')['Survived'].mean())
```

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

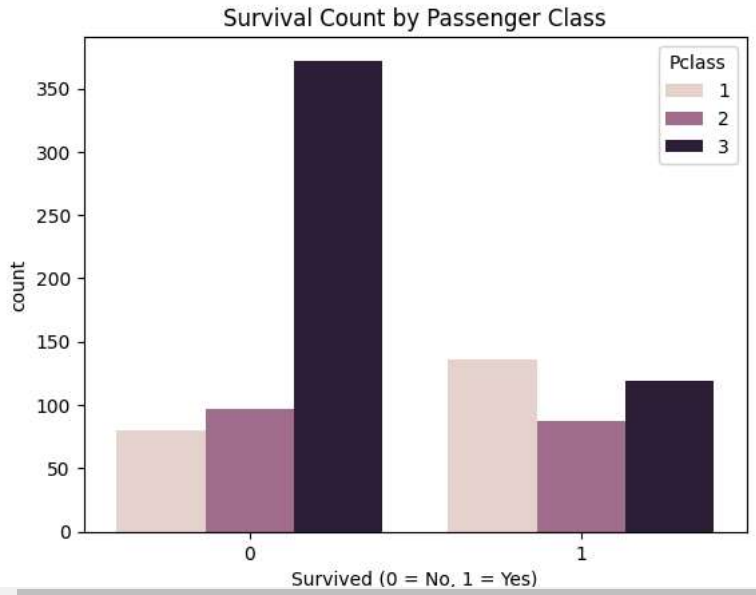
```
import seaborn as sns
import matplotlib.pyplot as plt

numeric_df=df.select_dtypes(include='number')

plt.figure(figsize=(10,8))
sns.heatmap(numeric_df.corr(),annot=True , cmap='coolwarm')
plt.show()
```



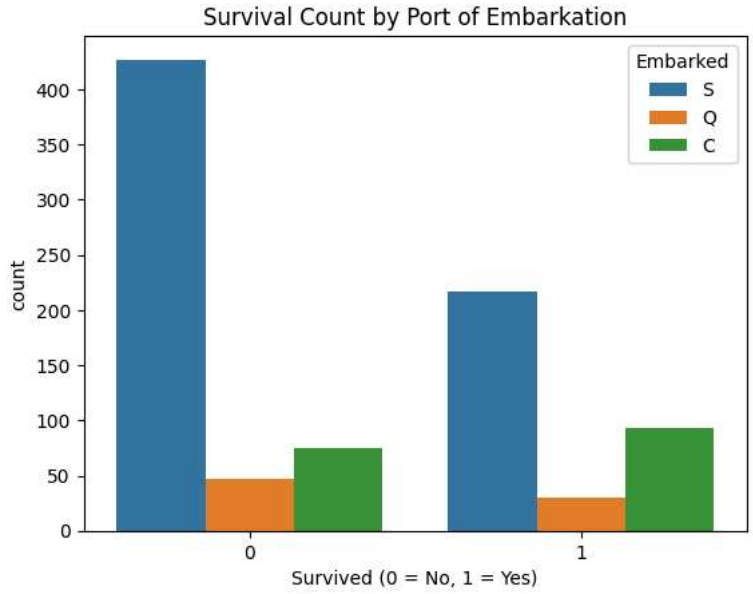

```
# Create a count plot for the 'Survived'
sns.countplot(x='Survived',data=df)


plt.title('Survival Count')

# Label the bars with the count values
for p in plt.gca().patches:
    plt.text(p.get_x() + p.get_width() / 2., p.get_height(),
             int(p.get_height()), ha='center', va='bottom')

plt.show()
```



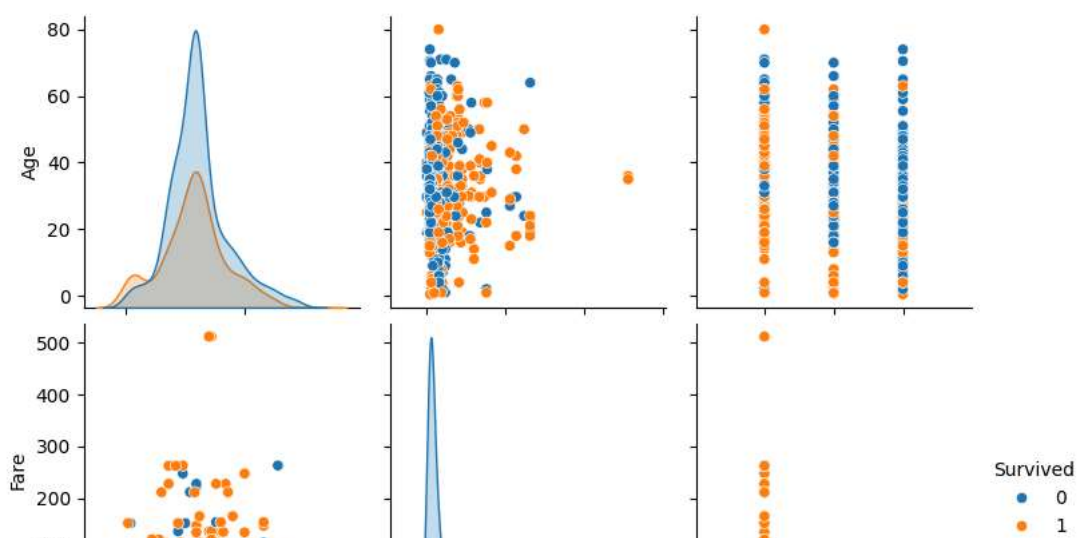
```
sns.countplot(x='Survived', hue='Pclass', data=df)

plt.title('Survival Count by Passenger Class')

plt.xlabel('Survived (0 = No, 1 = Yes)')

plt.show()
```




```
sns.countplot(x='Survived', hue='Embarked', data=df)
plt.title('Survival Count by Port of Embarkation')
plt.xlabel('Survived (0 = No, 1 = Yes)')
plt.show()
```




```
sns.pairplot(df, vars=['Age', 'Fare', 'Pclass'], hue='Survived', diag_kind='kde')
plt.show()
```


Age
Fare
Survived
0
1