[11]:
```
import pandas as pd

train = pd.read_csv('train.csv')
test = pd.read_csv('test.csv')
train.head()
```

[11]:

	PassengerId	Survived	Pclass	Name	Sex	...	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	...	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	...	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	...	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	...	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	...	0	373450	8.0500	NaN	S

5 rows × 12 columns

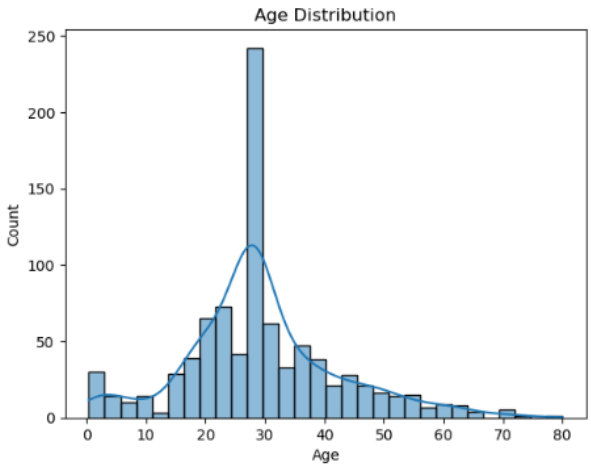
[19]:
```
train.describe(include='all')
```

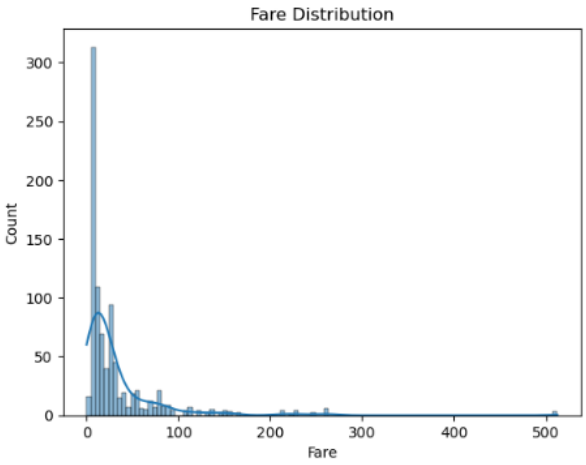
[19]:

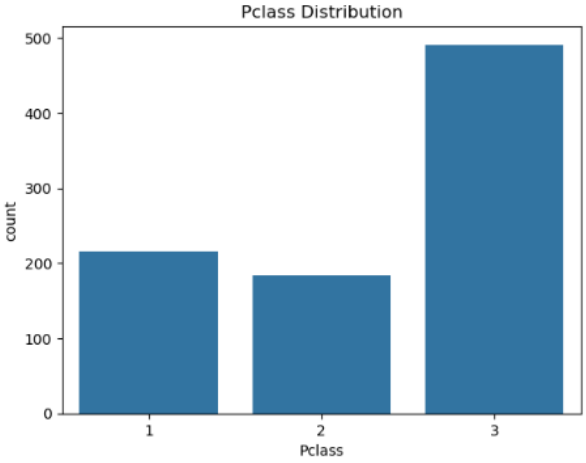
	PassengerId	Survived	Pclass	Name	Sex	...	SibSp	Parch	Ticket	Fare	Embarked
count	891.000000	891.0	891.0	891	891	...	891.000000	891.000000	891	891.000000	891
unique	NaN	2.0	3.0	891	2	...	NaN	NaN	681	NaN	3
top	NaN	0.0	3.0	Braund, Mr. Owen Harris	male	...	NaN	NaN	347082	NaN	S
freq	NaN	549.0	491.0	1	577	...	NaN	NaN	7	NaN	646
mean	446.000000	NaN	NaN	NaN	NaN	...	0.523008	0.381594	NaN	32.204208	NaN
std	257.353842	NaN	NaN	NaN	NaN	...	1.102743	0.806057	NaN	49.693429	NaN
min	1.000000	NaN	NaN	NaN	NaN	...	0.000000	0.000000	NaN	0.000000	NaN
25%	223.500000	NaN	NaN	NaN	NaN	...	0.000000	0.000000	NaN	7.910400	NaN
50%	446.000000	NaN	NaN	NaN	NaN	...	0.000000	0.000000	NaN	14.454200	NaN
75%	668.500000	NaN	NaN	NaN	NaN	...	1.000000	0.000000	NaN	31.000000	NaN
max	891.000000	NaN	NaN	NaN	NaN	...	8.000000	6.000000	NaN	512.329200	NaN

11 rows × 11 columns

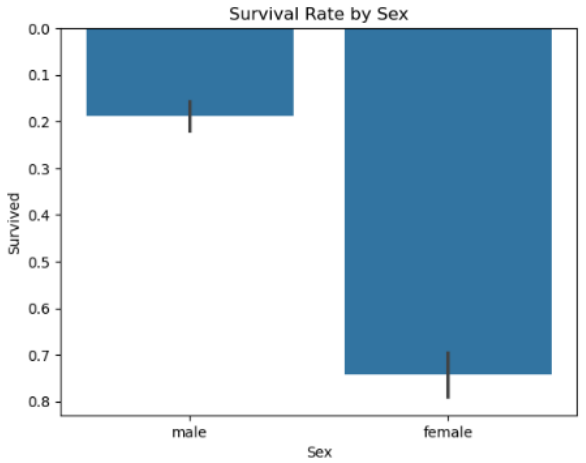
[21]:
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.histplot(train['Age'], kde=True)
plt.title('Age Distribution')
plt.show()

sns.histplot(train['Fare'], kde=True)
plt.title('Fare Distribution')
plt.show()

sns.countplot(x='Pclass', data=train)
plt.title('Pclass Distribution')
plt.show()
```

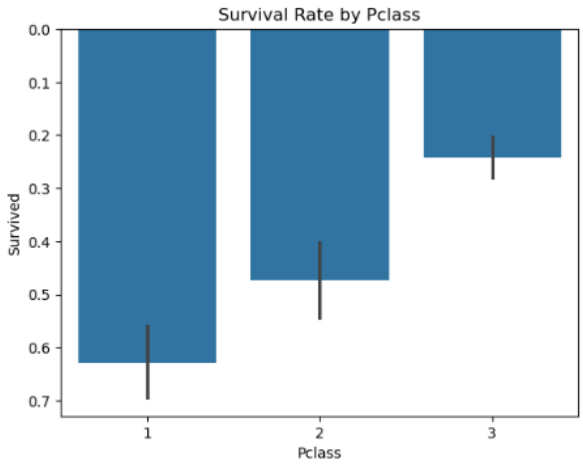


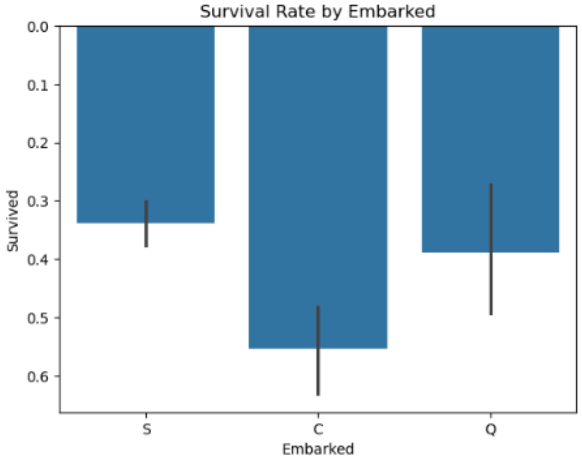




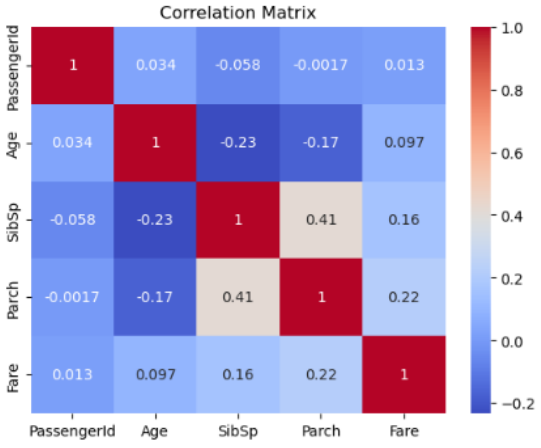
[23]:
```
sns.barplot(x='Sex', y='Survived', data=train)
plt.title('Survival Rate by Sex')
plt.show()

sns.barplot(x='Pclass', y='Survived', data=train)
plt.title('Survival Rate by Pclass')
plt.show()

sns.barplot(x='Embarked', y='Survived', data=train)
plt.title('Survival Rate by Embarked')
plt.show()
```







[27]:
```
corr_matrix = train.select_dtypes(include=['float64', 'int64']).corr()

sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```



[ ]: