In [1]:
```
import pandas as pd
```

In [3]:
```
df = pd.read_csv("D:\\IPSITA\\titanic.csv")
```

In [6]:
```
df.isnull().sum()
```

Out[6]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
Desc             0
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

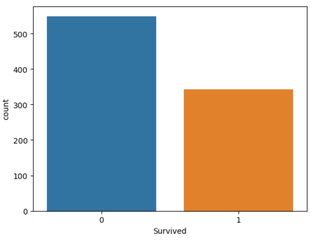
In [8]:
```
df['Age'].fillna(df['Age'].mean(), inplace=True)
df['Cabin'].fillna('Unknown', inplace=True)
```

In [9]:
```
df.head()
```

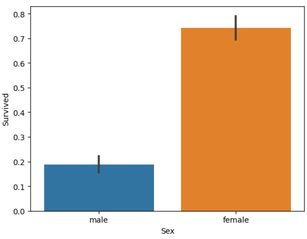
Out[9]:

	PassengerId	Survived	Pclass	Name	Sex	Age	Desc	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	Adult	1	0	A/5 21171	7.2500	Unknown	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	Adult	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	Adult	0	0	STON/O2. 3101282	7.9250	Unknown	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	Adult	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	Adult	0	0	373450	8.0500	Unknown	S

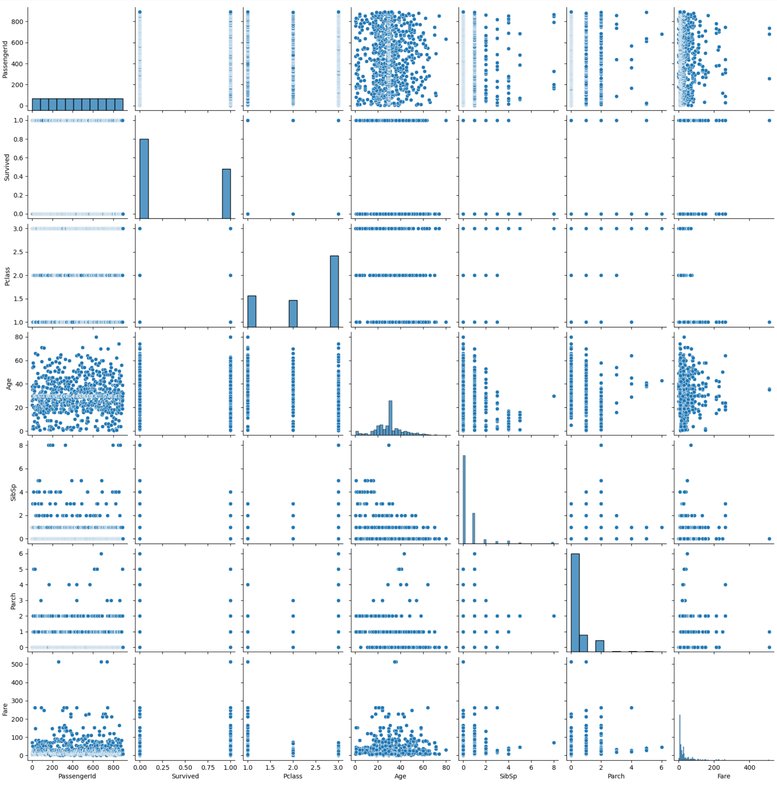
In [13]:
```
#UNIVARIATE MEANS ONLY SINGLE FEATURE(LIKE HERE SURVIVED)
import seaborn as sns
import matplotlib.pyplot as plt
sns.countplot(x='Survived', data=df)
plt.show()
```



In [14]:
```
#BIVARIATE REPLATION BETWEEN TWO FEATURES (SURVIVED VS SEX )
sns.barplot(x='Sex', y='Survived', data=df)
plt.show()
```



In [15]:
```
import warnings
warnings.filterwarnings("ignore")
sns.pairplot(df)
plt.show()
```



In [16]:
```
numerical_features = df.select_dtypes(include=['int64', 'float64'])
print("Distribution of numerical features:")
print(numerical_features.describe())
```

```
Distribution of numerical features:
       PassengerId    Survived      Pclass         Age       SibSp  \
count   891.000000  891.000000  891.000000  891.000000  891.000000
mean    446.000000    0.383838    2.308642   29.699118    0.523008
std     257.353842    0.486592    0.836071   13.002015    1.102743
min       1.000000    0.000000    1.000000    0.420000    0.000000
25%     223.500000    0.000000    2.000000   22.000000    0.000000
50%     446.000000    0.000000    3.000000   29.699118    0.000000
75%     668.500000    1.000000    3.000000   35.000000    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
```

In [17]:
```
categorical_features = df.select_dtypes(include=['object'])
for col in categorical_features:
    print("Distribution of", col, "feature:")
    print(df[col].value_counts())
```

```
Distribution of Name feature:
Braund, Mr. Owen Harris                     1
Boulos, Mr. Hanna                           1
Frolicher-Stehli, Mr. Maxmillian            1
Gilinski, Mr. Eliezer                       1
Murdlin, Mr. Joseph                         1
                                           ..
Kelly, Miss. Anna Katherine "Annie Kate"    1
McCoy, Mr. Bernard                          1
Johnson, Mr. William Cahoone Jr             1
Keane, Miss. Nora A                         1
Dooley, Mr. Patrick                         1
Name: Name, Length: 891, dtype: int64
Distribution of Sex feature:
male      577
female    314
Name: Sex, dtype: int64
Distribution of Desc feature:
Adult    675
Teen     133
Child     83
Name: Desc, dtype: int64
Distribution of Ticket feature:
347082      7
CA. 2343    7
1601        7
3101295     6
CA 2144     6
           ..
9234        1
19988       1
2693        1
PC 17612    1
370376      1
Name: Ticket, Length: 681, dtype: int64
Distribution of Cabin feature:
Unknown        687
C23 C25 C27      4
G6               4
B96 B98          4
C22 C26          3
              ...
E34              1
C7               1
C54              1
E36              1
C148             1
Name: Cabin, Length: 148, dtype: int64
Distribution of Embarked feature:
S    644
C    168
Q     77
Name: Embarked, dtype: int64
```

In [18]:
```
df.to_csv('cleaned_titanic_data.csv', index=False)
```

In [ ]:
```

```