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

In [2]:
```
import numpy as np
```

In [3]:
```
import seaborn as sns
#graphical package
```

In [5]:
```
df=pd.read_csv("C:\\Users\\praha\\OneDrive\\Desktop\\titanic_dataset.csv")
```

In [6]:
```
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
```

In [7]:
```
df.head()
```

Out[7]:

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

In [8]:
```
df.isnull().sum()
```

Out[8]:
```
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
```

In [10]:
```
df.drop(columns=['Cabin'],inplace=True)
```

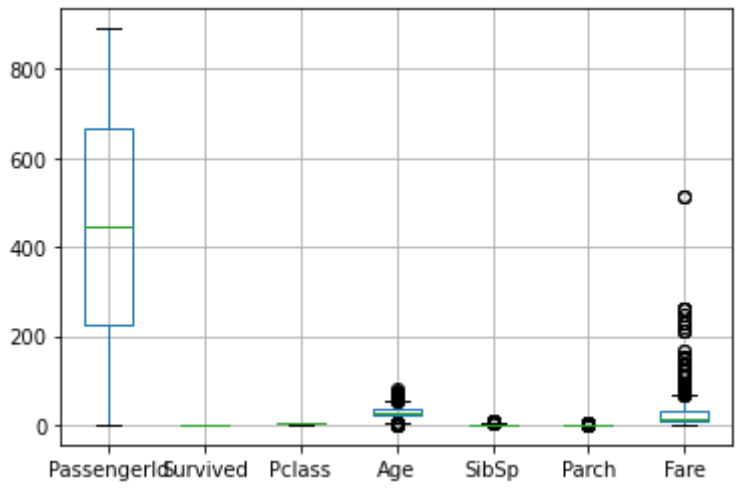
In [11]:
```
df.isnull().sum()
```

Out[11]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Embarked         2
dtype: int64
```

In [13]:
```
df['Age']=df['Age'].fillna(df['Age'].median())
```

In [14]:
```
df.boxplot()
```

Out[14]: <AxesSubplot:>



In [15]:
```
df.isnull().sum()
```

Out[15]:
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       2
dtype: int64
```

In [16]:
```
df['Embarked']=df['Embarked'].fillna(df['Embarked'].mode()[0])
#Non numerical data so we have used mode operation
```

In [17]:
```
df['Embarked'].value_counts()
```

Out[17]:
```
S    646
C    168
Q     77
Name: Embarked, dtype: int64
```

In [18]:
```
df['Pclass'].value_counts()
```

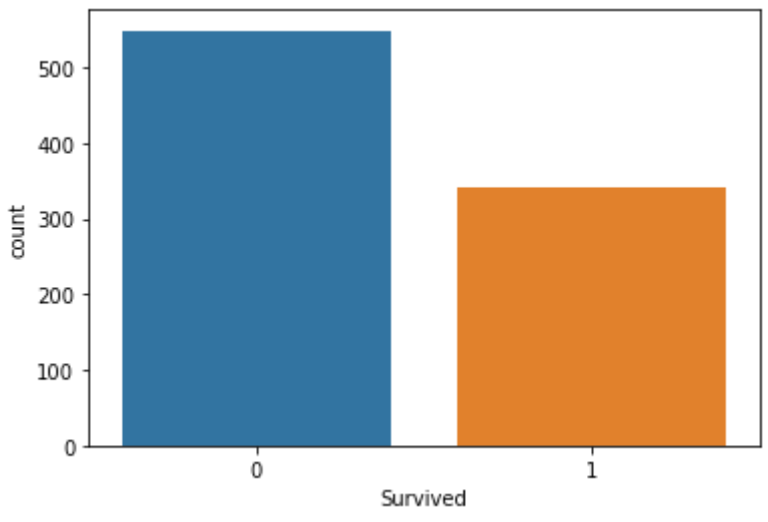
Out[18]:
```
3    491
1    216
2    184
Name: Pclass, dtype: int64
```

In [19]:
```
df['Survived'].value_counts()
#analyzing the data
```

Out[19]:
```
0    549
1    342
Name: Survived, dtype: int64
```

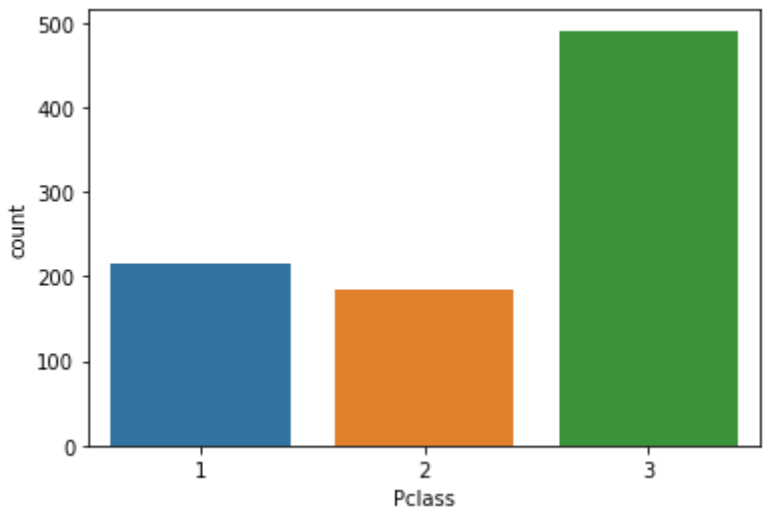
In [20]:
```
sns.countplot(x='Survived',data=df)
```

Out[20]: <AxesSubplot:xlabel='Survived', ylabel='count'>



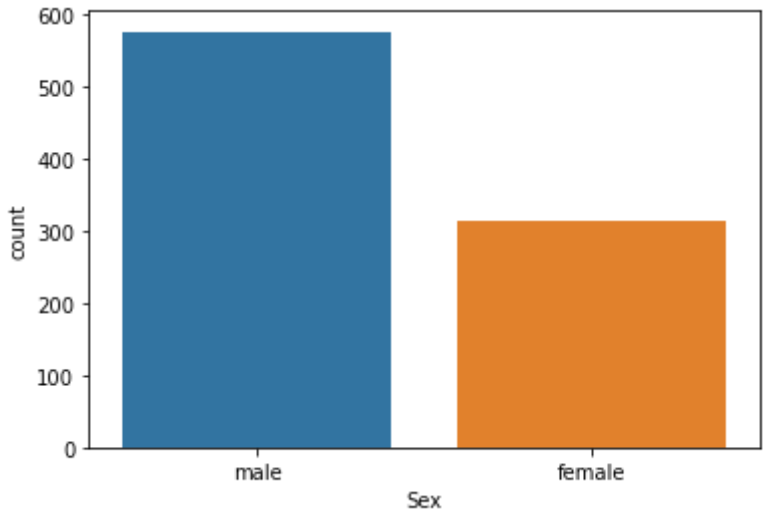
In [21]:
```
sns.countplot(x='Pclass',data=df)
```

Out[21]: <AxesSubplot:xlabel='Pclass', ylabel='count'>



In [22]:
```
sns.countplot(x='Sex',data=df)
```

Out[22]: <AxesSubplot:xlabel='Sex', ylabel='count'>

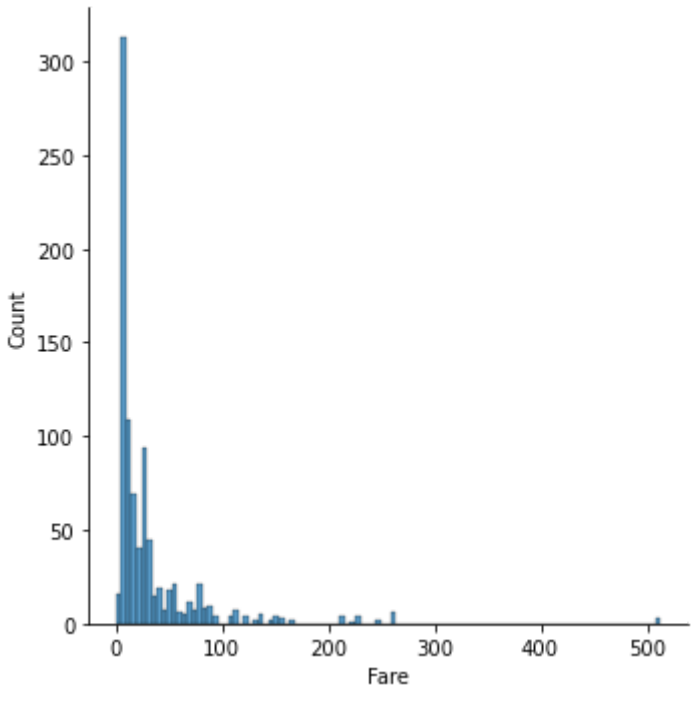


In [23]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 11 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          891 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  Embarked     891 non-null    object 
dtypes: float64(2), int64(5), object(4)
memory usage: 76.7+ KB
```

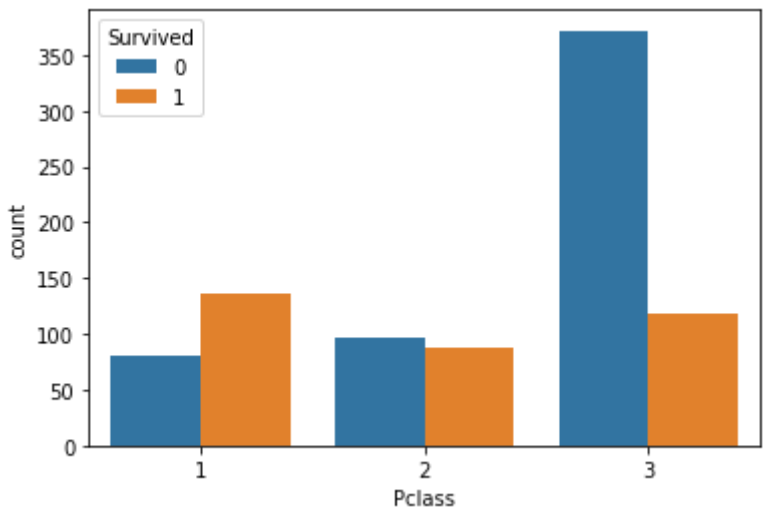
In [24]:
```
sns.displot(df['Fare'])
```

Out[24]: <seaborn.axisgrid.FacetGrid at 0x1da4ffb3070>



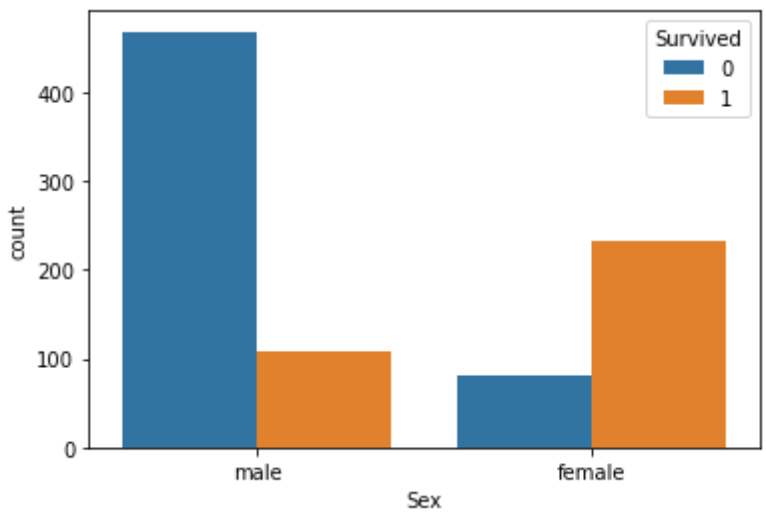
In [25]:
```
sns.countplot(x='Pclass',hue='Survived',data=df)
```

Out[25]: <AxesSubplot:xlabel='Pclass', ylabel='count'>



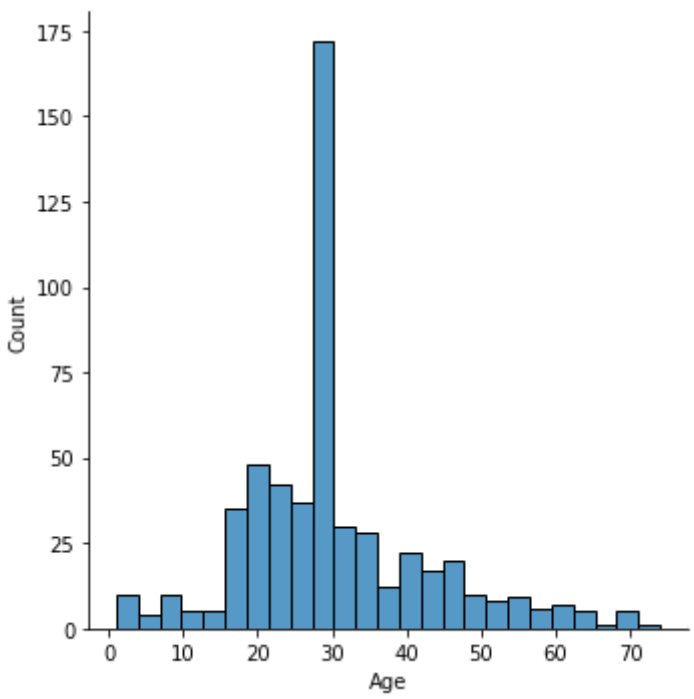
In [26]:
```
sns.countplot(x='Sex',hue='Survived',data=df)
```

Out[26]: <AxesSubplot:xlabel='Sex', ylabel='count'>



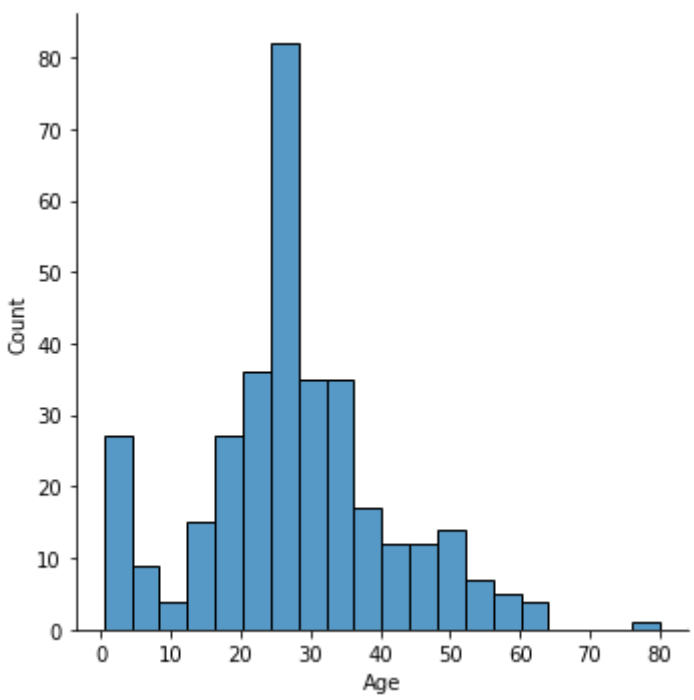
In [27]:
```
sns.displot(df[df['Survived']==0]['Age'])
```

Out[27]: <seaborn.axisgrid.FacetGrid at 0x1da4ffb3bb0>



In [28]:
```
sns.displot(df[df['Survived']==1]['Age'])
```

Out[28]: <seaborn.axisgrid.FacetGrid at 0x1da502add60>



In [29]:
```
pd.crosstab(df['Pclass'],df['Survived'])
```

Out[29]:

Survived	0	1
Pclass		
1	80	136
2	97	87
3	372	119

In [30]:
```
pd.crosstab(df['Sex'],df['Survived'])
```

Out[30]:

Survived	0	1
Sex		
female	81	233
male	468	109

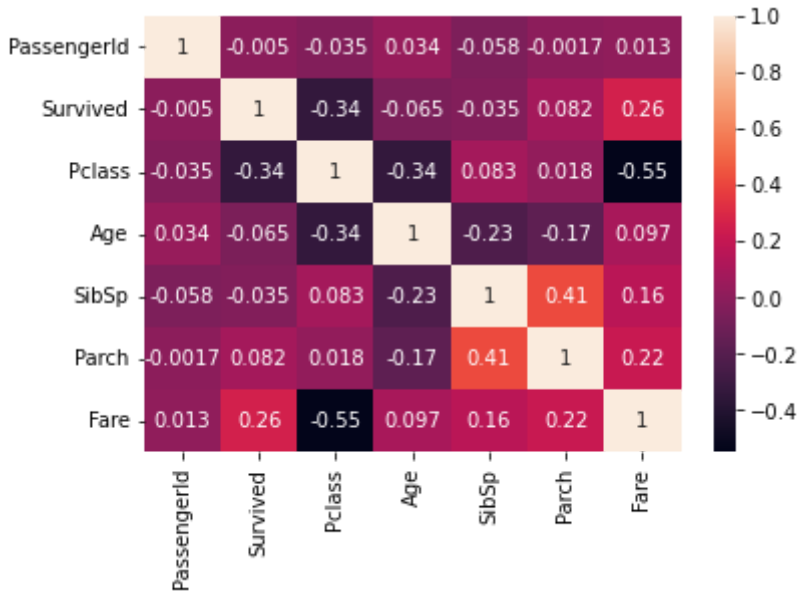
In [31]:
```
df.corr()
```

Out[31]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
PassengerId	1.000000	-0.005007	-0.035144	0.034212	-0.057527	-0.001652	0.012658
Survived	-0.005007	1.000000	-0.338481	-0.064910	-0.035322	0.081629	0.257307
Pclass	-0.035144	-0.338481	1.000000	-0.339898	0.083081	0.018443	-0.549500
Age	0.034212	-0.064910	-0.339898	1.000000	-0.233296	-0.172482	0.096688
SibSp	-0.057527	-0.035322	0.083081	-0.233296	1.000000	0.414838	0.159651
Parch	-0.001652	0.081629	0.018443	-0.172482	0.414838	1.000000	0.216225
Fare	0.012658	0.257307	-0.549500	0.096688	0.159651	0.216225	1.000000

In [32]:
```
sns.heatmap(df.corr(),annot=True)
```

Out[32]:
```
<AxesSubplot:>
```



In [ ]: