

```
In [15]: import pandas as pd
```

```
In [16]: import numpy as np
```

```
In [17]: import seaborn as sns
```

```
In [18]: df=pd.read_csv("titanic.csv")
df
```

```
Out[18]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN
...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
<b>889</b>	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148
<b>890</b>	891	0	3	Dooley, Mr.	male	32.0	0	0	370376	7.7500	NaN

In [19]:

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 [20]:

df.head()

Out[20]:

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

```
In [21]: df.isnull().sum()
```

```
Out[21]: 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 [22]: df.drop("Cabin",axis=1,inplace=True)
```

```
In [23]: df
```

```
Out[23]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	
...	...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	

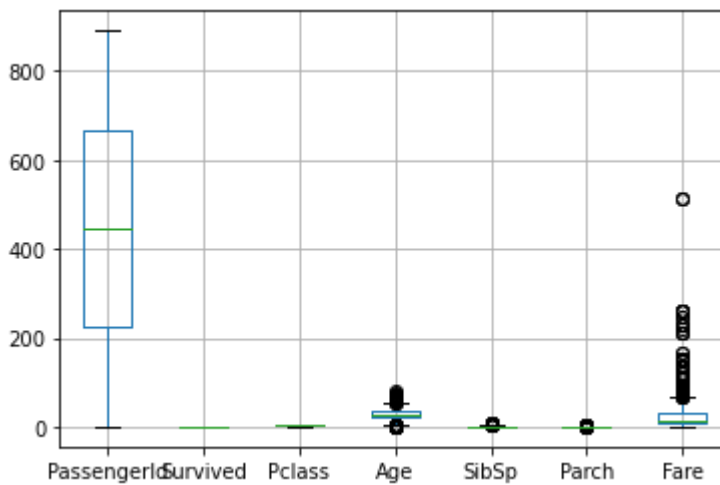
```
In [24]: df['Age']=df['Age'].fillna(df['Age'].median())
```

```
In [32]: df.isnull().sum()
```

```
Out[32]: PassengerId    0
Survived            0
Pclass              0
Name                0
Sex                 0
Age                 0
SibSp               0
Parch               0
Ticket              0
Fare                0
Embarked            0
dtype: int64
```

```
In [33]: df.boxplot()
```

```
Out[33]: <AxesSubplot:>
```



```
In [27]: df['Embarked']=df['Embarked'].fillna(df['Embarked'].mode()[0])
```

```
In [28]: df.isnull().sum()
```

```
Out[28]: PassengerId      0  
Survived      0  
Pclass      0  
Name      0  
Sex      0  
Age      0  
SibSp      0  
Parch      0  
Ticket      0  
Fare      0  
Embarked      0  
dtype: int64
```

```
In [34]: df["Embarked"].value_counts()
```

```
Out[34]: S      646  
C      168  
Q       77  
Name: Embarked, dtype: int64
```

```
In [35]: df["Pclass"].value_counts()
```

```
Out[35]: 3      491  
1      216  
2      184  
Name: Pclass, dtype: int64
```

```
In [36]: df["Survived"].value_counts()
```

```
Out[36]: 0      549  
1      342  
Name: Survived, dtype: int64
```

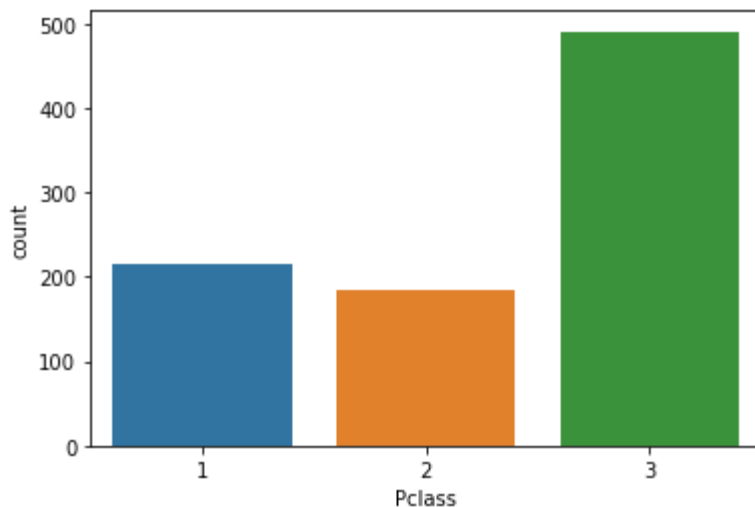
```
In [37]: sns.countplot(x="Survived",data=df)
```

```
Out[37]: <AxesSubplot:xlabel='Survived', ylabel='count'>
```



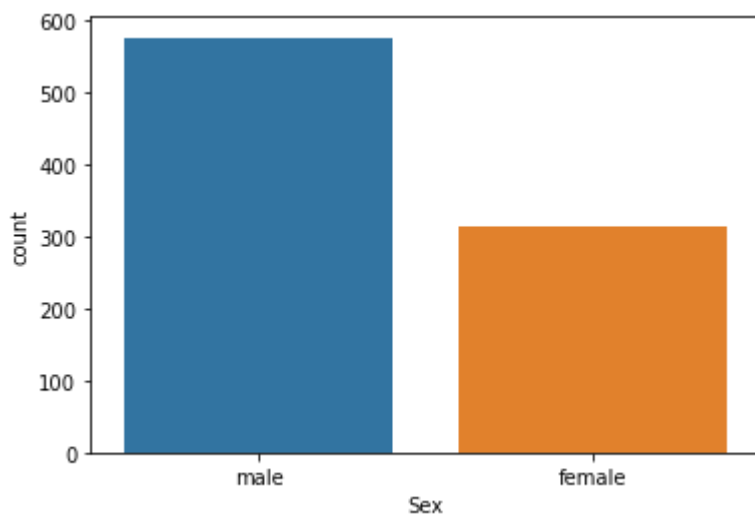
```
In [38]: sns.countplot(x="Pclass", data=df)
```

```
Out[38]: <AxesSubplot:xlabel='Pclass', ylabel='count'>
```



```
In [39]: sns.countplot(x="Sex", data=df)
```

```
Out[39]: <AxesSubplot:xlabel='Sex', ylabel='count'>
```



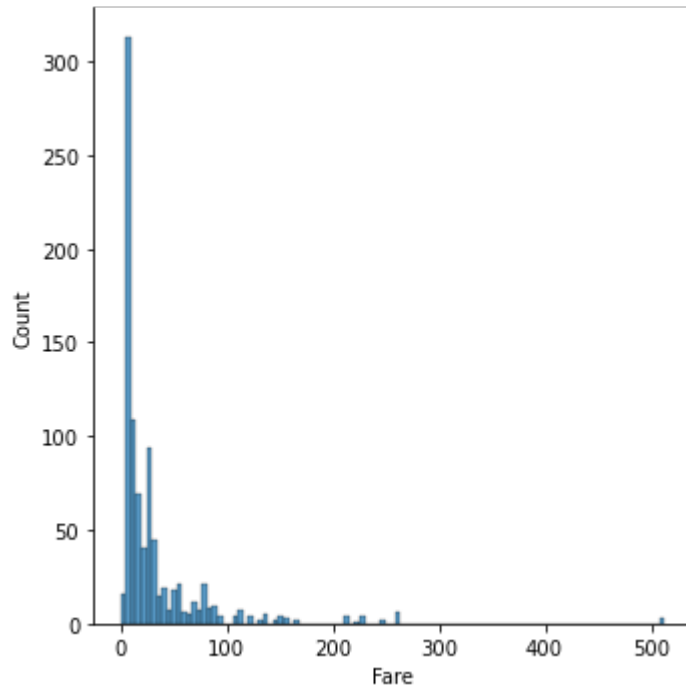
```
In [40]: df.info()
```

```
<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
```

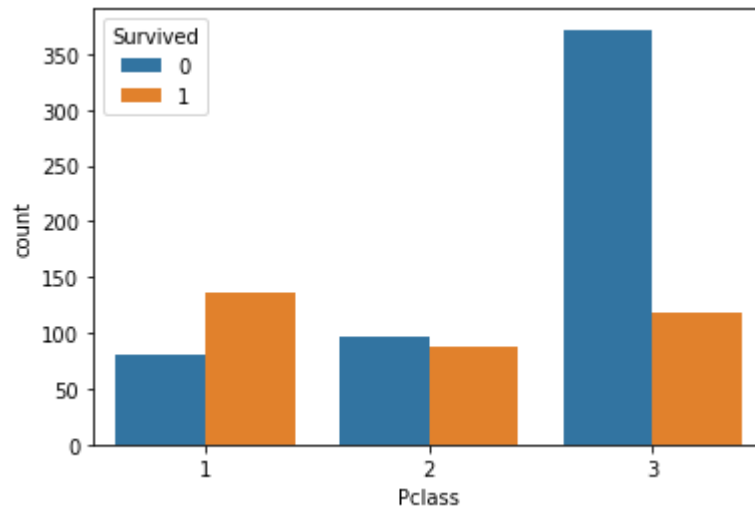
```
In [41]: sns.displot(df["Fare"])
```

```
Out[41]: <seaborn.axisgrid.FacetGrid at 0x275fceb5e0>
```



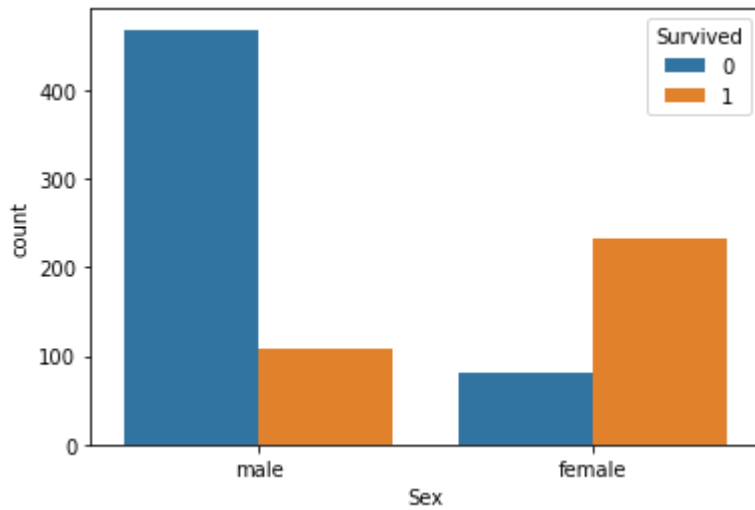
```
In [42]: sns.countplot(x="Pclass", hue="Survived", data=df)
```

```
Out[42]: <AxesSubplot:xlabel='Pclass', ylabel='count'>
```



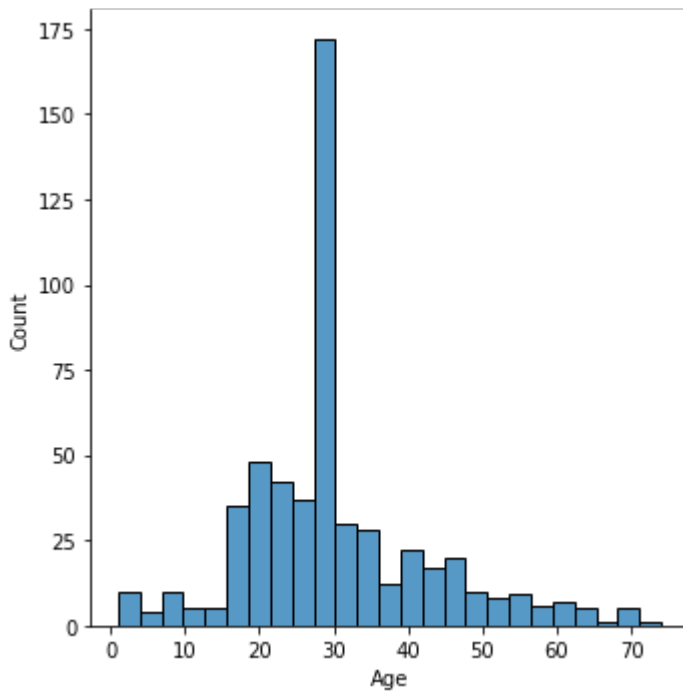
```
In [43]: sns.countplot(x="Sex", hue="Survived", data=df)
```

```
Out[43]: <AxesSubplot:xlabel='Sex', ylabel='count'>
```



```
In [44]: sns.displot(df[df["Survived"]==0]["Age"])
```

```
Out[44]: <seaborn.axisgrid.FacetGrid at 0x275fd290640>
```



```
In [45]: pd.crosstab(df["Pclass"], df["Survived"])
```

```
Out[45]: Survived    0    1
Pclass
1         80   136
```



**Survived**    0    1

```
In [46]: pd.crosstab(df["Sex"],df["Survived"])
```

Out[46]: **Survived**    0    1

	Sex	
female	81	233
male	468	109

```
In [47]: df.corr()
```

Out[47]:

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

Out[48]: <AxesSubplot:>

