In [3]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import math
data=pd.read_csv(r'C:\Users\nitya\Downloads\train.csv')
data.head()
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
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

In [3]:

```
print(data.shape)
```

```
(891, 12)
```

In [4]:

```
print(len(data))  #no.of passengers travelling in the ship
```

```
891
```

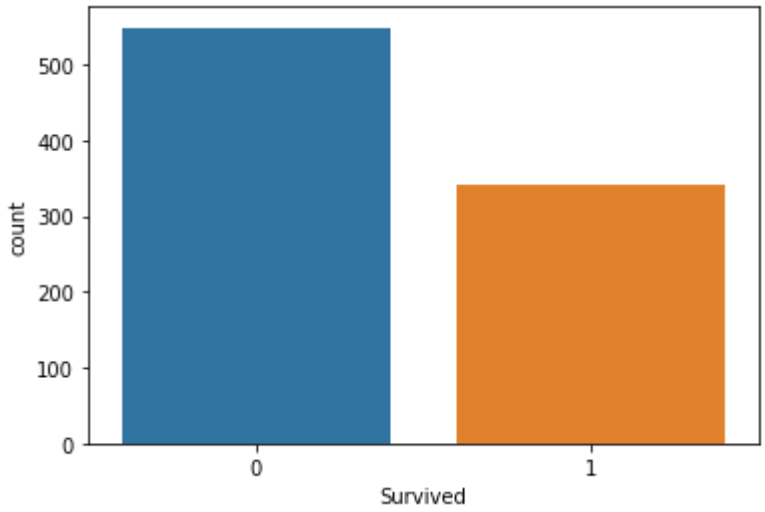
Analysing data

In [5]:

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

Out[5]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x15917a77630>
```

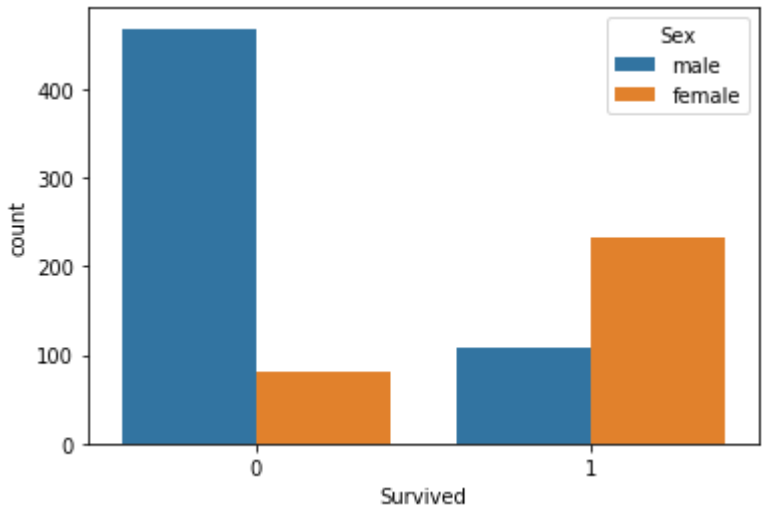


In [6]:

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

Out[6]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1591ca60a20>
```

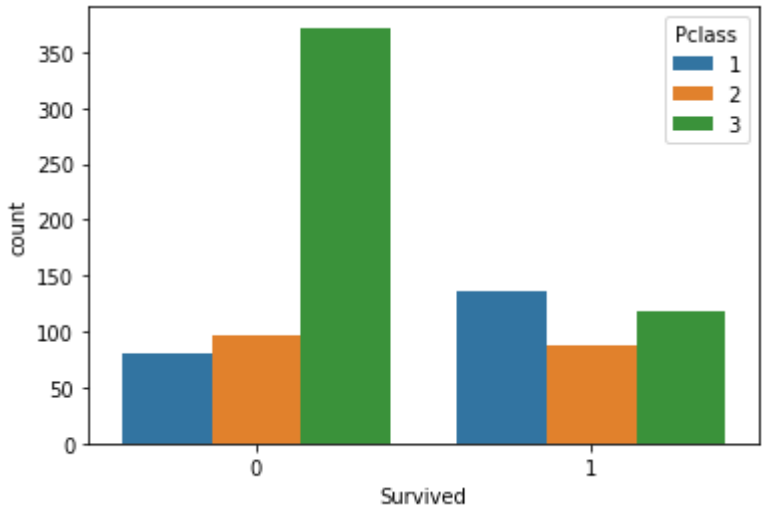


In [7]:

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

Out[7]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1591cac9d68>
```

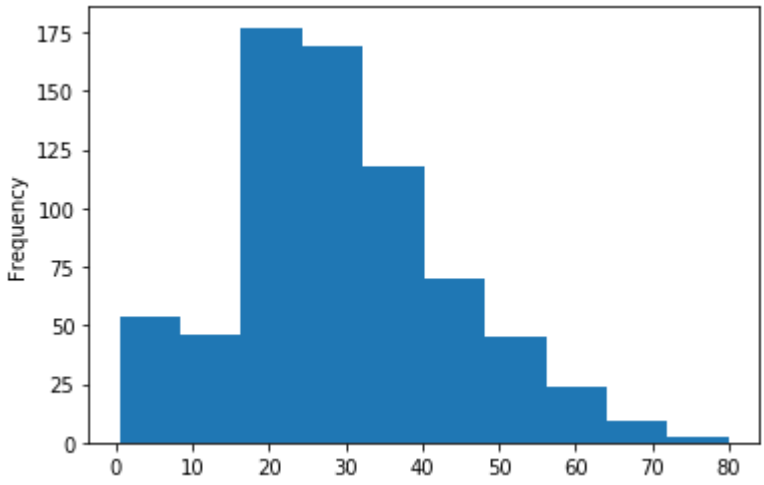


In [8]:

```
data['Age'].plot.hist()
```

Out[8]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1591cb4c9b0>
```

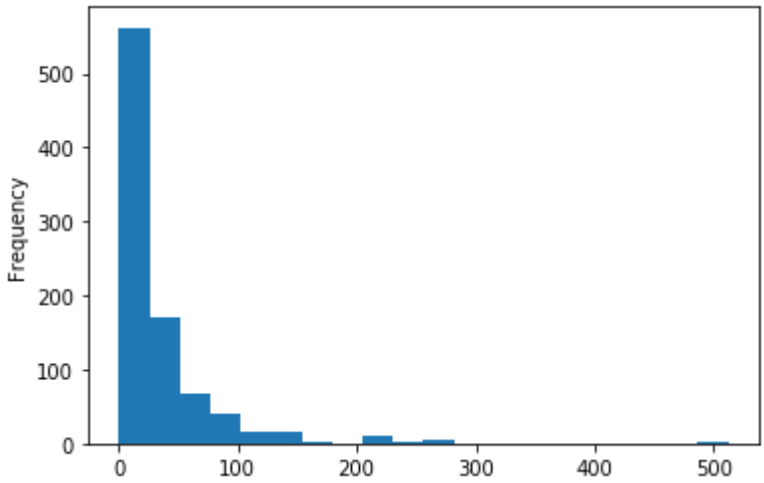


In [12]:

```
data['Fare'].plot.hist(bins=20)
```

Out[12]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1591ccd1f60>
```



In [14]:

```
data.info()
```

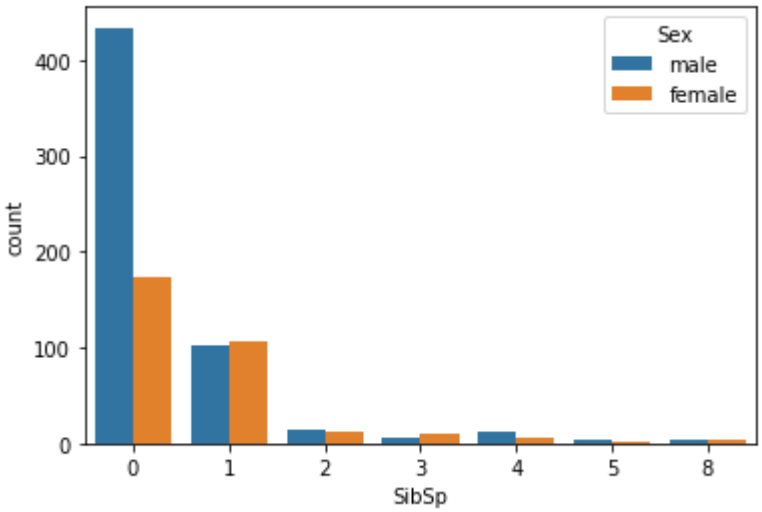
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
PassengerId    891 non-null int64
Survived       891 non-null int64
Pclass         891 non-null int64
Name           891 non-null object
Sex            891 non-null object
Age            714 non-null float64
SibSp          891 non-null int64
Parch          891 non-null int64
Ticket         891 non-null object
Fare           891 non-null float64
Cabin          204 non-null object
Embarked       889 non-null object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.6+ KB
```

In [16]:

```
sns.countplot(x='SibSp',hue='Sex',data=data)
```

Out[16]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1591ce65048>
```



Data wrangling

In [4]:

```
data.isnull()
```

Out[4]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	False	False	False	False	False	False	False	False	False	False	True
1	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	True
3	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	True
5	False	False	False	False	False	True	False	False	False	False	True
6	False	False	False	False	False	False	False	False	False	False	False
7	False	False	False	False	False	False	False	False	False	False	True
8	False	False	False	False	False	False	False	False	False	False	True
9	False	False	False	False	False	False	False	False	False	False	True
10	False	False	False	False	False	False	False	False	False	False	False
11	False	False	False	False	False	False	False	False	False	False	False
12	False	False	False	False	False	False	False	False	False	False	True
13	False	False	False	False	False	False	False	False	False	False	True
14	False	False	False	False	False	False	False	False	False	False	True
15	False	False	False	False	False	False	False	False	False	False	True
16	False	False	False	False	False	False	False	False	False	False	True
17	False	False	False	False	False	True	False	False	False	False	True
18	False	False	False	False	False	False	False	False	False	False	True
19	False	False	False	False	False	True	False	False	False	False	True
20	False	False	False	False	False	False	False	False	False	False	True
21	False	False	False	False	False	False	False	False	False	False	False
22	False	False	False	False	False	False	False	False	False	False	True
23	False	False	False	False	False	False	False	False	False	False	False
24	False	False	False	False	False	False	False	False	False	False	True
25	False	False	False	False	False	False	False	False	False	False	True
26	False	False	False	False	False	True	False	False	False	False	True
27	False	False	False	False	False	False	False	False	False	False	False
28	False	False	False	False	False	True	False	False	False	False	True
29	False	False	False	False	False	True	False	False	False	False	True
...	...	...	...	...	...	...	...	...	...	...	...
861	False	False	False	False	False	False	False	False	False	False	True
862	False	False	False	False	False	False	False	False	False	False	False
863	False	False	False	False	False	True	False	False	False	False	True
864	False	False	False	False	False	False	False	False	False	False	True
865	False	False	False	False	False	False	False	False	False	False	True
866	False	False	False	False	False	False	False	False	False	False	True

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
867	False	False	False	False	False	False	False	False	False	False	False
868	False	False	False	False	False	True	False	False	False	False	True
869	False	False	False	False	False	False	False	False	False	False	True
870	False	False	False	False	False	False	False	False	False	False	True
871	False	False	False	False	False	False	False	False	False	False	False
872	False	False	False	False	False	False	False	False	False	False	False
873	False	False	False	False	False	False	False	False	False	False	True
874	False	False	False	False	False	False	False	False	False	False	True
875	False	False	False	False	False	False	False	False	False	False	True
876	False	False	False	False	False	False	False	False	False	False	True
877	False	False	False	False	False	False	False	False	False	False	True
878	False	False	False	False	False	True	False	False	False	False	True
879	False	False	False	False	False	False	False	False	False	False	False
880	False	False	False	False	False	False	False	False	False	False	True
881	False	False	False	False	False	False	False	False	False	False	True
882	False	False	False	False	False	False	False	False	False	False	True
883	False	False	False	False	False	False	False	False	False	False	True
884	False	False	False	False	False	False	False	False	False	False	True
885	False	False	False	False	False	False	False	False	False	False	True
886	False	False	False	False	False	False	False	False	False	False	True
887	False	False	False	False	False	False	False	False	False	False	False
888	False	False	False	False	False	True	False	False	False	False	True
889	False	False	False	False	False	False	False	False	False	False	False
890	False	False	False	False	False	False	False	False	False	False	True

891 rows × 12 columns

In [5]:

```
data.isnull().sum()   #to have total no.of rows containing null values
```

Out[5]:

```
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 [6]:

```
data.head(5)
```

Out[6]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
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

In [7]:

```
data.drop("Cabin",axis=1,inplace=True)
```

In [8]:

```
data.head(5)
```

Out[8]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
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

In [9]:

```
data.dropna(inplace=True)
```

In [10]:

```
data.isnull().sum()
```

Out[10]:

```
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 [12]:

```
sex=pd.get_dummies(data['Sex'])  #convert into categorical data
sex.head(2)
```

Out[12]:

	female	male
0	0	1
1	1	0

In [13]:

```
sex=pd.get_dummies(data['Sex'],drop_first=True)
sex.head(2)       #donot require 2 column as male=0 means female hai
```

Out[13]:

	male
0	1
1	0

In [14]:

```
embark=pd.get_dummies(data['Embarked'],drop_first=True)
embark.head(5)    #in embark there are 3 values ,so q=0 and s=0 means c hai
```

Out[14]:

	Q	S
0	0	1
1	0	0
2	0	1
3	0	1
4	0	1

In [15]:

```
plc=pd.get_dummies(data['Pclass'],drop_first=True)
plc.head(5)
```

Out[15]:

	2	3
0	0	1
1	0	0
2	0	1
3	0	0
4	0	1

In [16]:

```
data=pd.concat([data,sex,embark,plc],axis=1)
data.head(5)
```

Out[16]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
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

In [17]:

```
data.drop(['PassengerId','Pclass','Name','Sex','Ticket','Embarked'],axis=1,inplace=True
)
data.head()
```

Out[17]:

	Survived	Age	SibSp	Parch	Fare	male	Q	S	2	3
0	0	22.0	1	0	7.2500	1	0	1	0	1
1	1	38.0	1	0	71.2833	0	0	0	0	0
2	1	26.0	0	0	7.9250	0	0	1	0	1
3	1	35.0	1	0	53.1000	0	0	1	0	0
4	0	35.0	0	0	8.0500	1	0	1	0	1

Train data

In [22]:

```
X=data.drop('Survived',axis=1)
y=data['Survived']
```

In [21]:

```
from sklearn.model_selection import train_test_split
```

In [23]:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1
) #testsize means test size is 30% and train is 70%
```

In [24]:

```
from sklearn.linear_model import LogisticRegression
```

In [25]:

```
logmodel=LogisticRegression()
```

In [26]:

```
logmodel.fit(X_train,y_train)
```

```
C:\Users\nitya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.p
y:432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. S
pecify a solver to silence this warning.
  FutureWarning)
```

Out[26]:

```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=Tru
e,
                   intercept_scaling=1, l1_ratio=None, max_iter=100,
                   multi_class='warn', n_jobs=None, penalty='l2',
                   random_state=None, solver='warn', tol=0.0001, verbose=
0,
                   warm_start=False)
```

In [27]:

```
prediction=logmodel.predict(X_test)
```

In [28]:

```
from sklearn.metrics import classification_report
```

In [29]:

```
classification_report(y_test,prediction)
```

Out[29]:

```
'              precision    recall  f1-score   support\n\n           0
0.81      0.83      0.82       126\n           1       0.75      0.72
0.73        88\n\n    accuracy                           0.79       214\n
macro avg       0.78      0.77      0.78       214\nweighted avg       0.7
8      0.79      0.78       214\n'
```

In [30]:

```
from sklearn.metrics import confusion_matrix
```

In [31]:

```
confusion_matrix(y_test,prediction)
```

Out[31]:

```
array([[105,  21],
       [ 25,  63]], dtype=int64)
```

In [33]:

```
from sklearn.metrics import accuracy_score
```

In [34]:

```
accuracy_score(y_test,prediction)
```

Out[34]:

```
0.7850467289719626
```

In [ ]:

```

```