In [71]:

```
# Collect Data : Import libraries
```

In [72]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In [73]:

```
titanic_data = pd.read_csv("titanic_dataset.csv")
```

In [74]:

```
titanic_data.head(10)
```

Out[74]:

	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
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708

In [75]:

```
print("# of passengers in original data :"+ str((len(titanic_data.index))))
```

```
# of passengers in original data :891
```

In [76]:

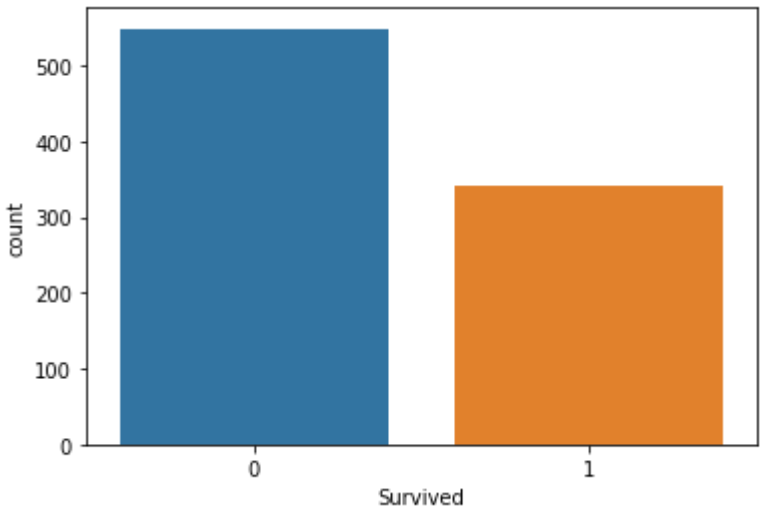
```
# Anaylsing Data
#Creating different plot to check relationship between variables
```

In [77]:

```
sns.countplot(x= "Survived", data=titanic_data)
```

Out[77]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c02bb54b38>
```

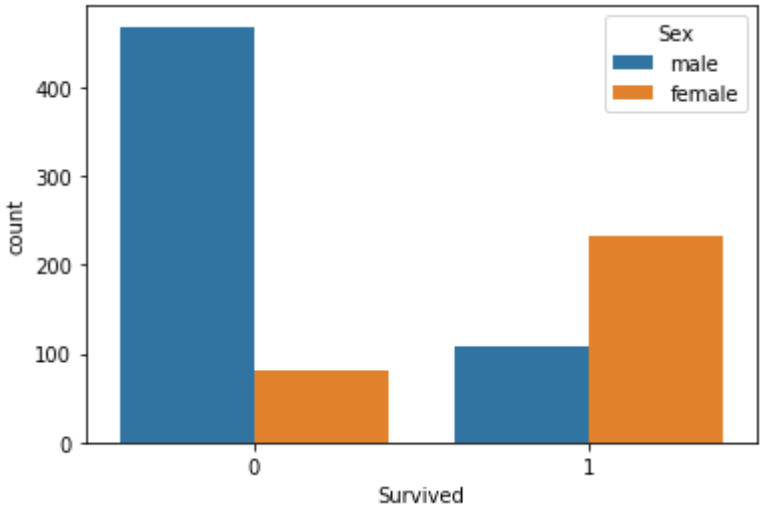


In [78]:

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

Out[78]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c02bb95f60>
```

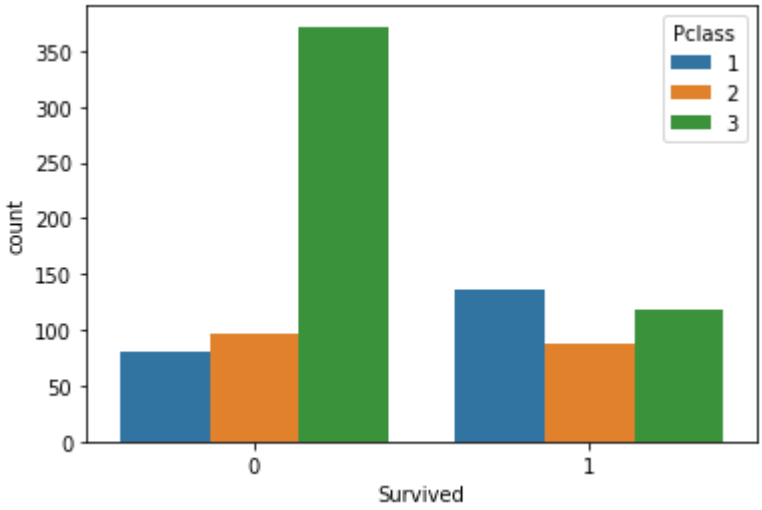


In [79]:

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

Out[79]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c02bbe50b8>
```



In [81]:

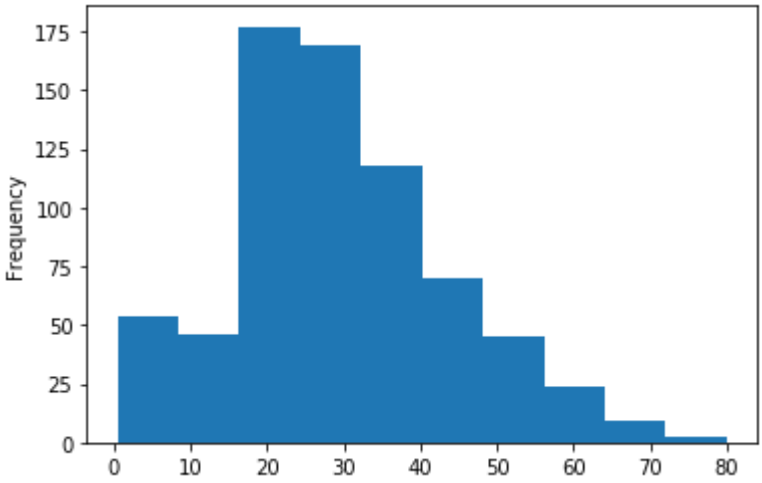
```
#titanic_data["Age"].plot.hist(bin= 20,fig.size(10,5))
```

In [85]:

```
titanic_data["Age"].plot.hist()
```

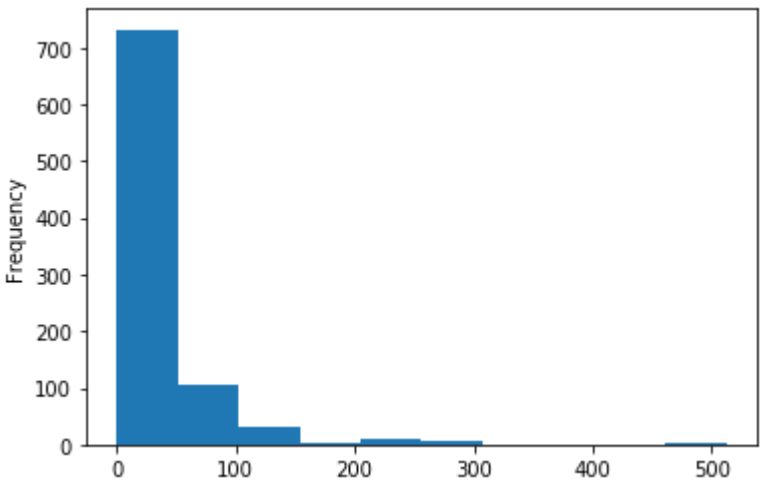
Out[85]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c02bcd0a58>
```



In [86]:

```
titanic_data["Fare"].plot.hist()
plt.show()
```



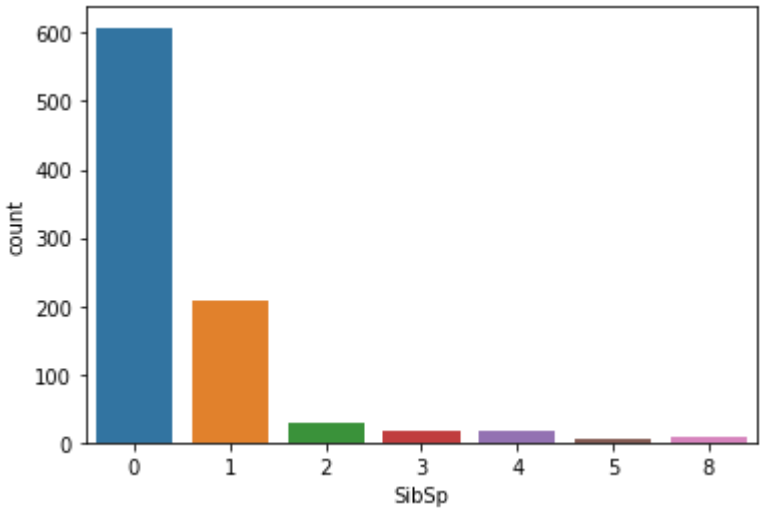
In [87]:

```
titanic_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 [88]:

```
sns.countplot(x = "SibSp", data = titanic_data)
plt.show()
```



In [89]:

```
# Data Wrangling
# Clean the data by removing the Nan Values and unnecessary columns in the dataset
```

In [90]:

```
titanic_data.isnull()#check Null Value
```

Out[90]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	I
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	

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
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	
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 [91]:

```
titanic_data.isnull().sum()
```

Out[91]:

```
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 [92]:

```
titanic_data.head(5)
```

Out[92]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	C
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 [93]:

```
titanic_data.drop("Cabin", axis = 1 , inplace = True)
titanic_data.head(5)
```

Out[93]:

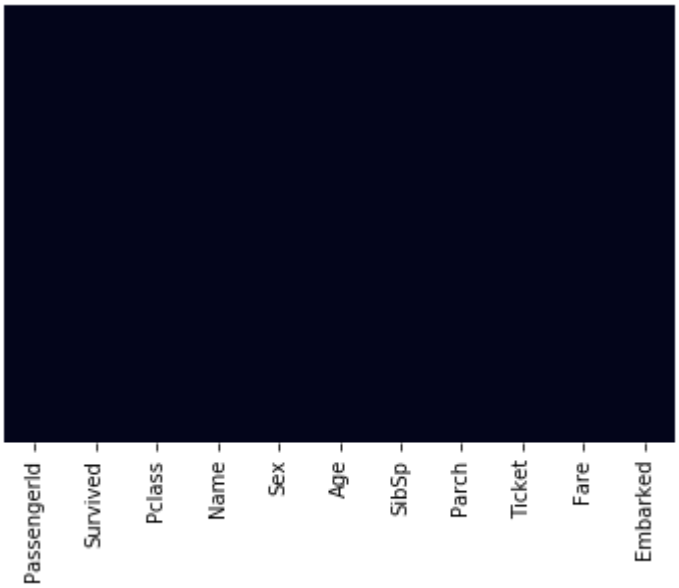
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle, 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 [94]:

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

In [95]:

```
sns.heatmap(titanic_data.isnull(), yticklabels = False , cbar = False)
plt.show()
```



In [96]:

```
titanic_data.isnull().sum()
```

Out[96]:

```
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 [97]:

```
pd.get_dummies(titanic_data["Sex"] , drop_first = True)
```

Out[97]:

	male
0	1
1	0
2	0
3	0
4	1
6	1
7	1
8	0
9	0
10	0
11	0
12	1
13	1
14	0
15	0
16	1
18	0
20	1
21	1
22	0
23	1
24	0
25	0
27	1
30	1
33	1
34	1
35	1
37	1
38	0
...	...
856	0
857	1
858	0

	male
860	1
861	1
862	0
864	1
865	0
866	0
867	1
869	1
870	1
871	0
872	1
873	1
874	0
875	0
876	1
877	1
879	0
880	0
881	1
882	0
883	1
884	1
885	0
886	1
887	0
889	1
890	1

712 rows × 1 columns

In [98]:

```
embark = pd.get_dummies(titanic_data["Embarked"], drop_first = True)
```

In [99]:

```
embark.head(5)
```

Out[99]:

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

In [100]:

```
pcl = pd.get_dummies(titanic_data["Pclass"], drop_first = True)
pcl.head(5)
```

Out[100]:

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

In [101]:

```
titanic_data.drop(["Sex", "Embarked", "PassengerId", "Name", "Ticket"], axis = 1, inplace =
```

In [102]:

```
titanic_data.head()
```

Out[102]:

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

In [103]:

```
titanic_data.drop(["Pclass"], axis =1 , inplace = True)
```

In [104]:

```
#Train  & test Data
```

In [105]:

```
# Build the model on the train data and predict the output on the test data
```

In [106]:

```
x = titanic_data.drop("Survived",  axis = 1)
y = titanic_data["Survived"]
```

In [107]:

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

In [108]:

```
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.3, random_state=7)
```

In [109]:

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

In [110]:

```
logmodel = LogisticRegression()
```

In [113]:

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

```
C:\Users\Hello\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:
432: FutureWarning: Default solver will be changed to 'lbfgs' in 0.22. Speci
fy a solver to silence this warning.
  FutureWarning)
```

Out[113]:

```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                   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 [114]:

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

In [115]:

```
# Accuracy Check
```

In [128]:

```
from sklearn.metrics import classification_report,confusion_matrix, accuracy_score
```

In [129]:

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

Out[129]:

```
'              precision    recall  f1-score   support\n\n           0
0.68      0.92      0.79       129\n           1       0.75      0.35
0.48        85\n\n    accuracy                           0.70       214\n
macro avg       0.72      0.64      0.63       214\nweighted avg       0.71
0.70      0.66       214\n'
```

In [130]:

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

Out[130]:

```
array([[119,  10],
       [ 55,  30]], dtype=int64)
```

In [131]:

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

Out[131]:

```
0.6962616822429907
```

In [ ]:

```

```