# UNIT: - Logistic Regression

In [1]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as pmt
import math

%matplotlib inline
```

In [2]:

```
titanic_data=pd.read_csv("C:/Users/Prathyu Lachireddy/Desktop/BP/titanic.CSV")
titanic_data
```

Out[2]:

| | 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 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **886** | 887 | 0 | 2 | Montvila, Rev. Juozas | male | 27.0 | 0 | 0 | 211536 | 13.0000 |
| **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 |

891 rows × 12 columns

In [3]:

```
titanic_data.head()
```

Out[3]:

| | 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 [4]:

```
print('no of passengers:'+str(len(titanic_data.index)))
```

```
no of passengers:891
```

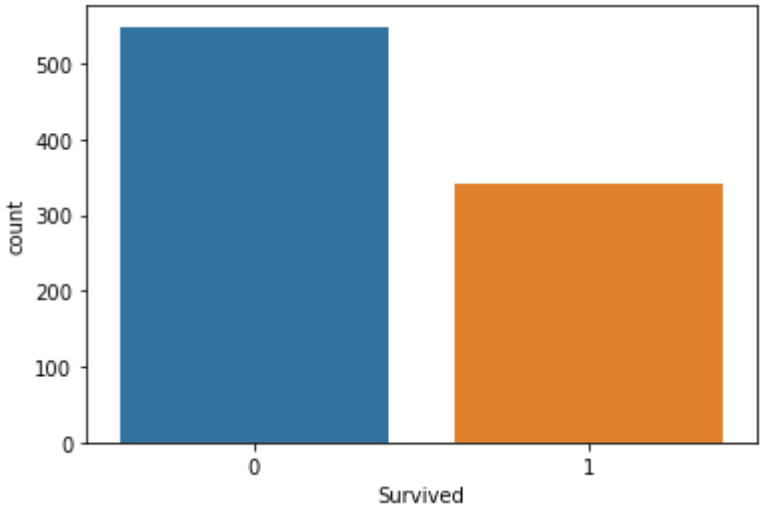
# Analyzing the data

In [5]:

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

Out[5]:

```
<AxesSubplot:xlabel='Survived', ylabel='count'>
```

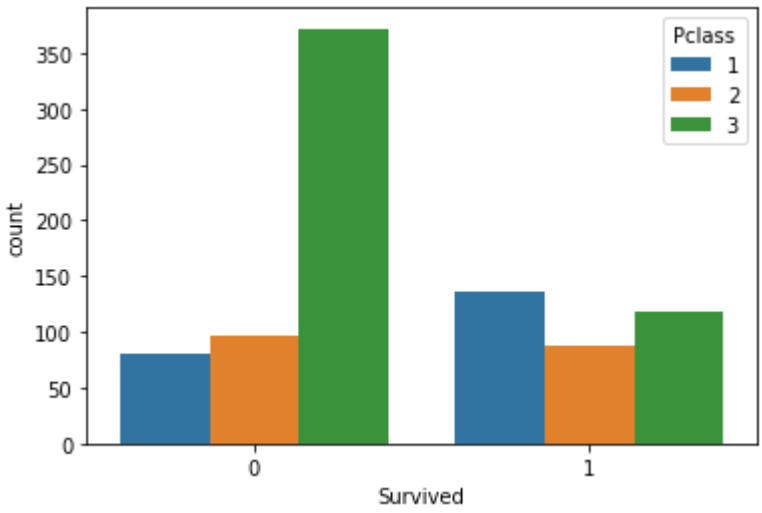



In [6]:

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

Out[6]:

```
<AxesSubplot:xlabel='Survived', ylabel='count'>
```

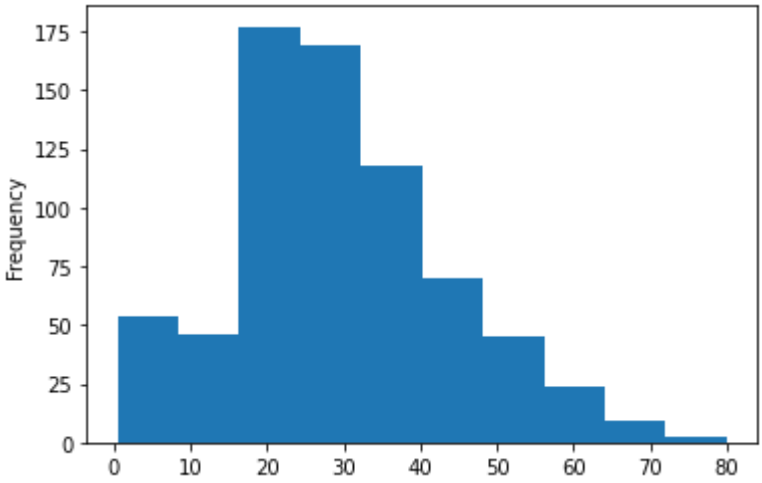


In [7]:

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

Out[7]:

```
<AxesSubplot:ylabel='Frequency'>
```

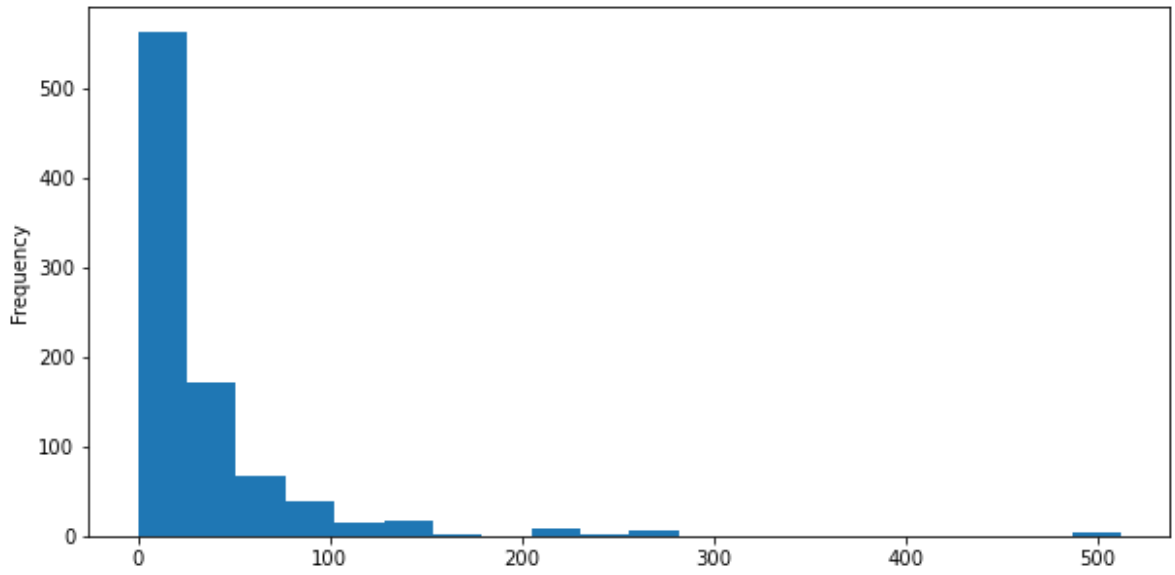


In [8]:

```
titanic_data['Fare'].plot.hist(bins=20,figsize=(10,5))
```

Out[8]:

```
<AxesSubplot:ylabel='Frequency'>
```



In [9]:

```
titanic_data.info
```

Out[9]:

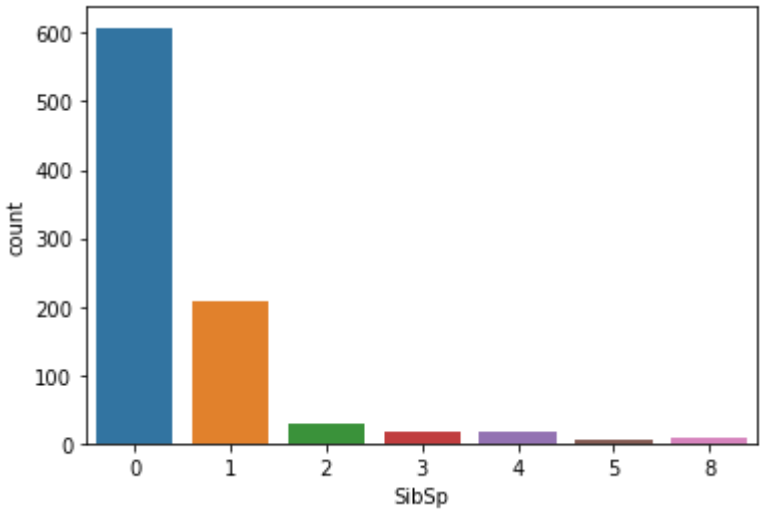
```
<bound method DataFrame.info of      PassengerId  Survived  Pclass  \
0              1         0       3
1              2         1       1
2              3         1       3
3              4         1       1
4              5         0       3
..           ...       ...     ...
886          887         0       2
887          888         1       1
888          889         0       3
889          890         1       1
890          891         0       3

                                                 Name     Sex   Age  SibSp
\
0                             Braund, Mr. Owen Harris    male  22.0      1
1    Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2                              Heikkinen, Miss. Laina  female  26.0      0
3        Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                            Allen, Mr. William Henry    male  35.0      0
..                                                ...     ...   ...    ...
886                            Montvila, Rev. Juozas    male  27.0      0
887                     Graham, Miss. Margaret Edith  female  19.0      0
888         Johnston, Miss. Catherine Helen "Carrie"  female   NaN      1
889                            Behr, Mr. Karl Howell    male  26.0      0
890                              Dooley, Mr. Patrick    male  32.0      0

     Parch            Ticket     Fare Cabin Embarked
0        0         A/5 21171   7.2500   NaN        S
1        0          PC 17599  71.2833   C85        C
2        0  STON/O2. 3101282   7.9250   NaN        S
3        0            113803  53.1000  C123        S
4        0            373450   8.0500   NaN        S
..     ...               ...      ...   ...      ...
886      0            211536  13.0000   NaN        S
887      0            112053  30.0000   B42        S
888      2        W./C. 6607  23.4500   NaN        S
889      0            111369  30.0000  C148        C
890      0            370376   7.7500   NaN        Q

[891 rows x 12 columns]>
```

In [10]:

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

Out[10]:

```
<AxesSubplot:xlabel='SibSp', ylabel='count'>
```




# 3. Data Wrangling

In [11]:

```
titanic_data.isnull()
```

Out[11]:

| | 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 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **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 [12]:

```
# Count of Nulls
titanic_data.isnull().sum()
```

Out[12]:

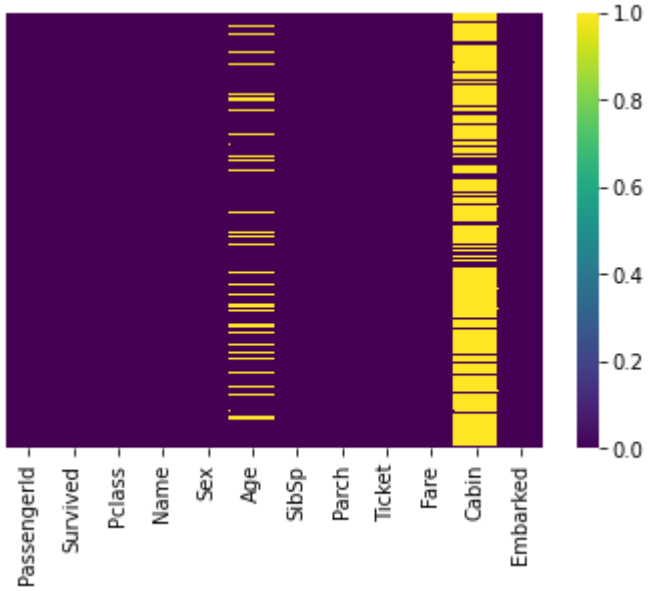
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

In [13]:

```
#Heatmap for nulls
sns.heatmap(titanic_data.isnull(),yticklabels=False,cmap='viridis')
```

Out[13]:

```
<AxesSubplot:>
```

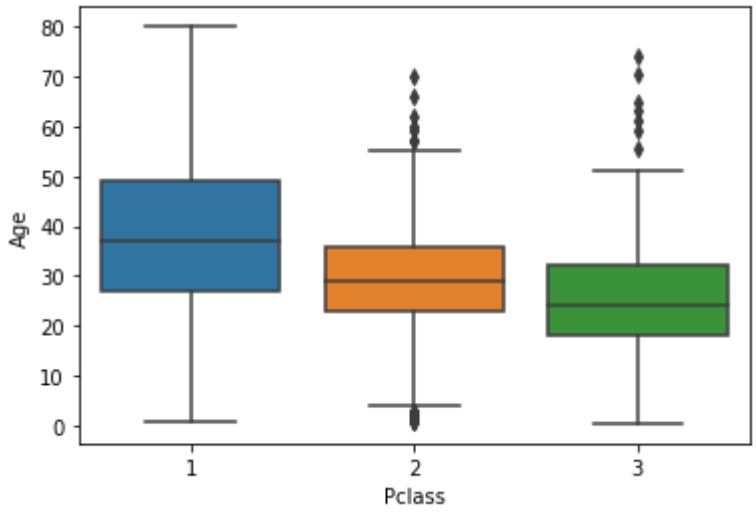


In [14]:

```
#Box plot

sns.boxplot(x='Pclass',y='Age',data=titanic_data)
```

Out[14]:

```
<AxesSubplot:xlabel='Pclass', ylabel='Age'>
```



In [15]:

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

Out[15]:

| | 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 [16]:

```
titanic_data.drop('Cabin',axis=1,inplace=True)
```

In [17]:

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

Out[17]:

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare | I |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **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 [18]:

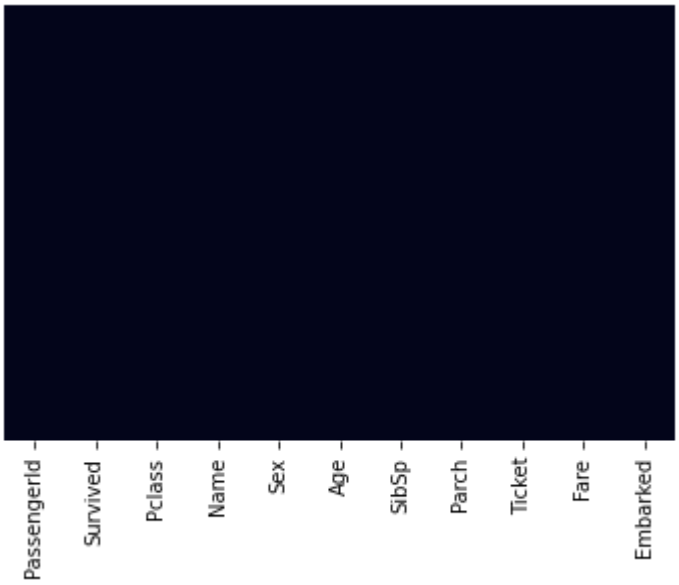
```
titanic_data.dropna(inplace=True)
```

In [19]:

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

Out[19]:

```
<AxesSubplot:>
```




In [20]:

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

Out[20]:

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

DATA WRANGLING CLEAN THE DATA BY REMOVING THE NULL AND UNNECCESSAY VALUES

In [21]:

```
titanic_data.head(2)
```

Out[21]:

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare | Emba |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **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 | |

In [22]:

```
pd.get_dummies(titanic_data['Sex'])
```

Out[22]:

| | female | male |
|---|---|---|
| **0** | 0 | 1 |
| **1** | 1 | 0 |
| **2** | 1 | 0 |
| **3** | 1 | 0 |
| **4** | 0 | 1 |
| **...** | ... | ... |
| **885** | 1 | 0 |
| **886** | 0 | 1 |
| **887** | 1 | 0 |
| **889** | 0 | 1 |
| **890** | 0 | 1 |

712 rows × 2 columns

In [23]:

```
sex=pd.get_dummies(titanic_data['Sex'],drop_first=True)
sex.head(5)
```

Out[23]:

| | male |
|---|---|
| **0** | 1 |
| **1** | 0 |
| **2** | 0 |
| **3** | 0 |
| **4** | 1 |

In [24]:

```
embarked=pd.get_dummies(titanic_data['Embarked'])
embarked.head(5)
```

Out[24]:

| | C | Q | S |
|---|---|---|---|
| **0** | 0 | 0 | 1 |
| **1** | 1 | 0 | 0 |
| **2** | 0 | 0 | 1 |
| **3** | 0 | 0 | 1 |
| **4** | 0 | 0 | 1 |

In [25]:

```
embarked=pd.get_dummies(titanic_data['Embarked'],drop_first=True)
embarked.head(5)
```

Out[25]:

| | Q | S |
|---|---|---|
| **0** | 0 | 1 |
| **1** | 0 | 0 |
| **2** | 0 | 1 |
| **3** | 0 | 1 |
| **4** | 0 | 1 |

In [26]:

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

Out[26]:

| | 2 | 3 |
|---|---|---|
| **0** | 0 | 1 |
| **1** | 0 | 0 |
| **2** | 0 | 1 |
| **3** | 0 | 0 |
| **4** | 0 | 1 |

In [27]:

```
titanic_data=pd.concat([titanic_data,sex,embarked,pcl])
titanic_data.head(5)
```

Out[27]:

| | Age | Embarked | Fare | Name | Parch | PassengerId | Pclass | Q | S | Sex | SibSp |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 22.0 | S | 7.2500 | Braund, Mr. Owen Harris | 0.0 | 1.0 | 3.0 | NaN | NaN | male | 1.0 |
| **1** | 38.0 | C | 71.2833 | Cumings, Mrs. John Bradley (Florence Briggs Th... | 0.0 | 2.0 | 1.0 | NaN | NaN | female | 1.0 |
| **2** | 26.0 | S | 7.9250 | Heikkinen, Miss. Laina | 0.0 | 3.0 | 3.0 | NaN | NaN | female | 0.0 |
| **3** | 35.0 | S | 53.1000 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | 0.0 | 4.0 | 1.0 | NaN | NaN | female | 1.0 |
| **4** | 35.0 | S | 8.0500 | Allen, Mr. William Henry | 0.0 | 5.0 | 3.0 | NaN | NaN | male | 0.0 |

In [28]:

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

Out[28]:

| | Age | Embarked | Fare | Name | Parch | PassengerId | Pclass | Q | S | Sex | SibSp |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 22.0 | S | 7.2500 | Braund, Mr. Owen Harris | 0.0 | 1.0 | 3.0 | NaN | NaN | male | 1.0 |
| **1** | 38.0 | C | 71.2833 | Cumings, Mrs. John Bradley (Florence Briggs Th... | 0.0 | 2.0 | 1.0 | NaN | NaN | female | 1.0 |
| **2** | 26.0 | S | 7.9250 | Heikkinen, Miss. Laina | 0.0 | 3.0 | 3.0 | NaN | NaN | female | 0.0 |
| **3** | 35.0 | S | 53.1000 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | 0.0 | 4.0 | 1.0 | NaN | NaN | female | 1.0 |
| **4** | 35.0 | S | 8.0500 | Allen, Mr. William Henry | 0.0 | 5.0 | 3.0 | NaN | NaN | male | 0.0 |

# 4. Train and test

train data

In [29]:

```
!pip install -U scikit-learn
```

```
Requirement already up-to-date: scikit-learn in c:\users\prathyu lachireddy
\anaconda3\lib\site-packages (0.24.2)
Requirement already satisfied, skipping upgrade: threadpoolctl>=2.0.0 in
c:\users\prathyu lachireddy\anaconda3\lib\site-packages (from scikit-learn)
(2.1.0)
Requirement already satisfied, skipping upgrade: scipy>=0.19.1 in c:\users\p
rathyu lachireddy\anaconda3\lib\site-packages (from scikit-learn) (1.5.2)
Requirement already satisfied, skipping upgrade: joblib>=0.11 in c:\users\pr
athyu lachireddy\anaconda3\lib\site-packages (from scikit-learn) (0.17.0)
Requirement already satisfied, skipping upgrade: numpy>=1.13.3 in c:\users\p
rathyu lachireddy\anaconda3\lib\site-packages (from scikit-learn) (1.19.2)
```

In [30]:

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

In [31]:

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

In [33]:

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=1)
```

In [34]:

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

In [42]:

```
logmodel=LogisticRegression(solver='Ibfgs',max_iter=10000)
```

In [56]:

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

```
---------------------------------------------------------------------------
-
ValueError                                Traceback (most recent call las
t)
<ipython-input-56-4ec4121cdd45> in <module>
----> 1 logmodel.fit(x_test,y_test)

~\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py in fit(sel
f, X, y, sample_weight)
   1304         The SAGA solver supports both float64 and float32 bit arra
ys.
   1305         """
-> 1306         solver = _check_solver(self.solver, self.penalty, self.dua
l)
   1307
   1308         if not isinstance(self.C, numbers.Number) or self.C < 0:

~\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py in _check_
solver(solver, penalty, dual)
```

In [44]:

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

```
---------------------------------------------------------------------------
NotFittedError                            Traceback (most recent call last)
<ipython-input-44-33c4de5fc33c> in <module>
----> 1 prediction=logmodel.predict(x_test)

~\anaconda3\lib\site-packages\sklearn\linear_model\_base.py in predict(self,
X)
    307             Predicted class label per sample.
    308         """
--> 309         scores = self.decision_function(X)
    310         if len(scores.shape) == 1:
    311             indices = (scores > 0).astype(int)

~\anaconda3\lib\site-packages\sklearn\linear_model\_base.py in decision_func
tion(self, X)
    280             class would be predicted.
    281         """
--> 282         check_is_fitted(self)
    283
    284         X = check_array(X, accept_sparse='csr')

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*args,
 **kwargs)
     61             extra_args = len(args) - len(all_args)
     62             if extra_args <= 0:
---> 63                 return f(*args, **kwargs)
     64
     65             # extra_args > 0

~\anaconda3\lib\site-packages\sklearn\utils\validation.py in check_is_fitted
(estimator, attributes, msg, all_or_any)
   1096
   1097     if not attrs:
-> 1098         raise NotFittedError(msg % {'name': type(estimator).__name__
})
   1099
   1100

NotFittedError: This LogisticRegression instance is not fitted yet. Call 'fi
t' with appropriate arguments before using this estimator.
```

In [45]:

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

In [46]:

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

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-46-e4ea99fb2866> in <module>
----> 1 classification_report(y_test,prediction)

NameError: name 'prediction' is not defined
```

In [52]:

```
from sklearn.metrics import confusion_matrix
confusion_matrix(y_test,prediction)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-52-941734aadd12> in <module>
      1 from sklearn.metrics import confusion_matrix
----> 2 confusion_matrix(y_test,prediction)

NameError: name 'prediction' is not defined
```

In [53]:

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

In [54]:

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

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-54-7117e5c868fb> in <module>
----> 1 accuracy_score(y_test,prediction)

NameError: name 'prediction' is not defined
```

In [55]:

```
(105+64)/(105+21+24+64)
```

Out[55]:

```
0.7897196261682243
```

In [59]:

```
# Natural Language Processing (NLP)
```

In [57]:

```
import nltk
```

In [60]:

```
nltk.download()
```

```
showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/inde
x.xml (https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml)
```

Out[60]:

```
True
```

In [ ]:

```

```