

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LogisticRegression
```

```
In [2]: !pip install scikit-learn
```

Defaulting to user installation because normal site-packages is not writeable

Requirement already satisfied: scikit-learn in c:\programdata\anaconda3\lib\site-packages (1.0.2)

Requirement already satisfied: threadpoolctl>=2.0.0 in c:\programdata\anaconda3\lib\site-packages (from scikit-learn) (2.2.0)

Requirement already satisfied: numpy>=1.14.6 in c:\programdata\anaconda3\lib\site-packages (from scikit-learn) (1.21.5)

Requirement already satisfied: scipy>=1.1.0 in c:\programdata\anaconda3\lib\site-packages (from scikit-learn) (1.9.1)

Requirement already satisfied: joblib>=0.11 in c:\programdata\anaconda3\lib\site-packages (from scikit-learn) (1.1.0)

```
In [3]: titanic = pd.read_csv(r'C:\Users\786\Downloads\Titanic-Dataset.csv')
titanic
```

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  |
| ... | ...         | ...      | ...    | ...   | ...    | ...  | ...   | ...   | ...              | ...     |
| 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 [4]: titanic.head()
```

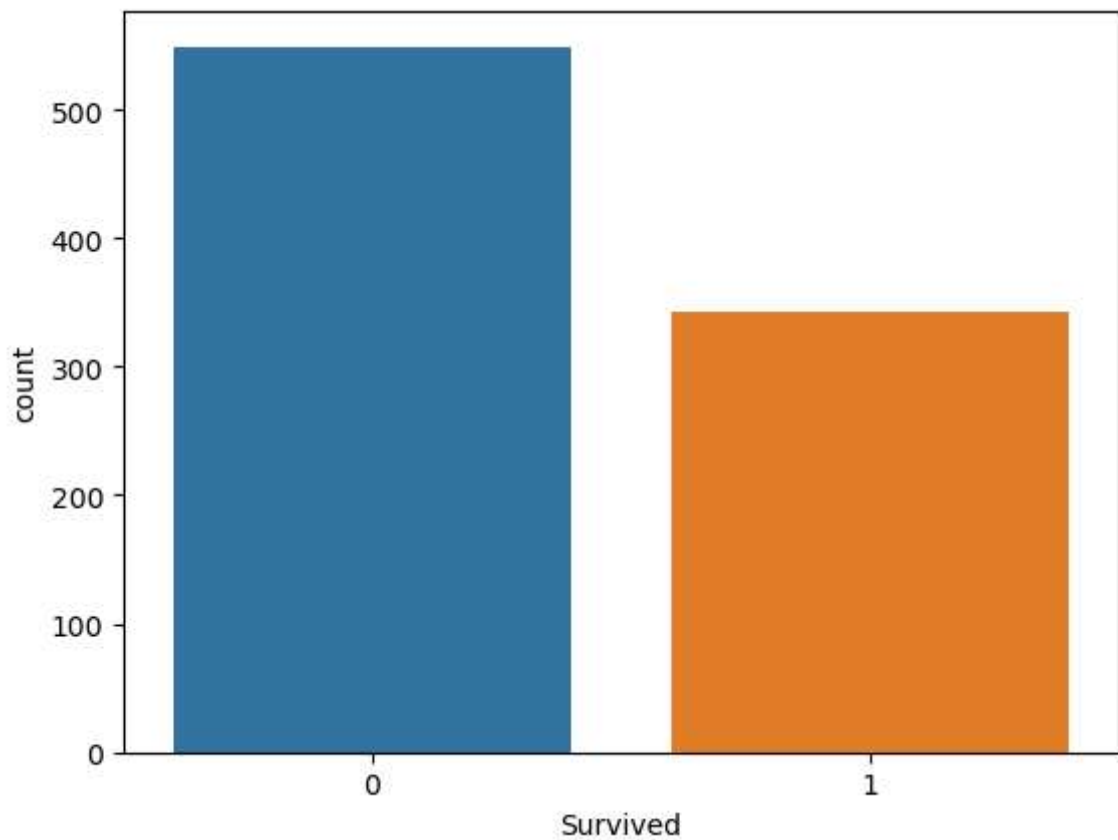
```
Out[4]:
```

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

◀ ————— ▶

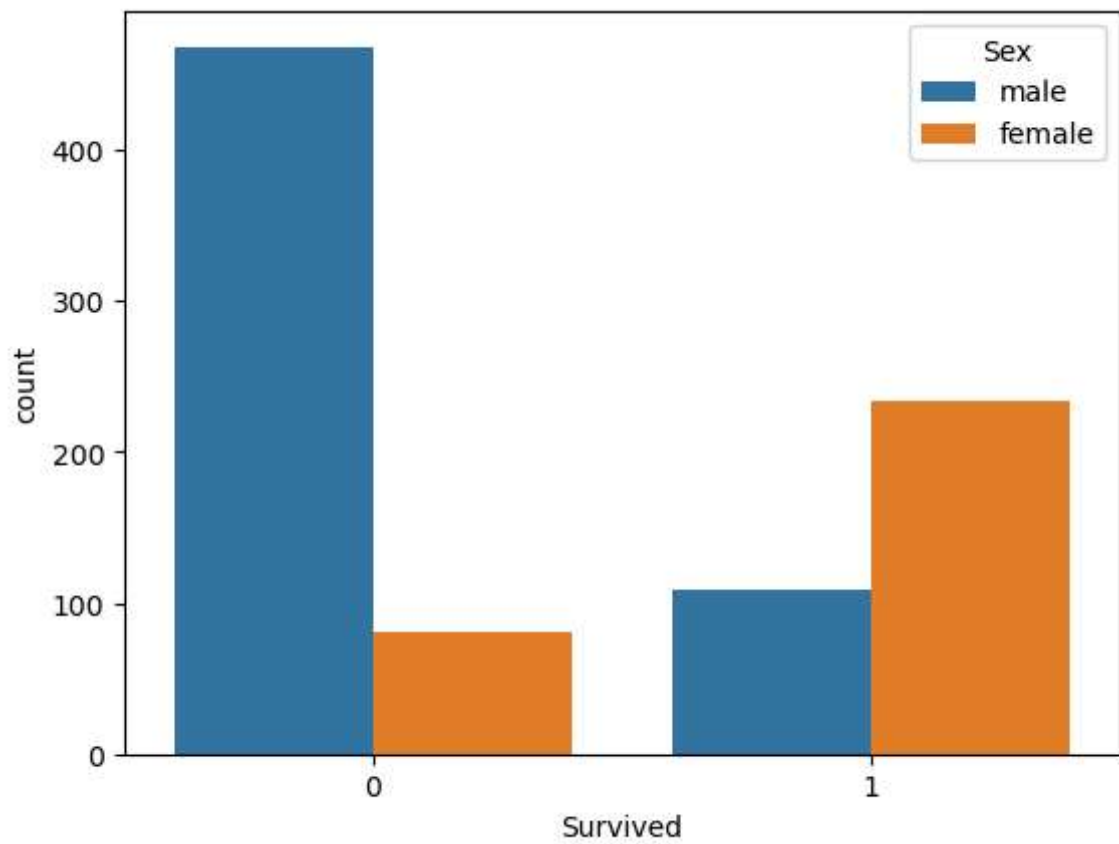
```
In [5]: sns.countplot(x='Survived',data=titanic)
```

```
Out[5]: <AxesSubplot:xlabel='Survived', ylabel='count'>
```



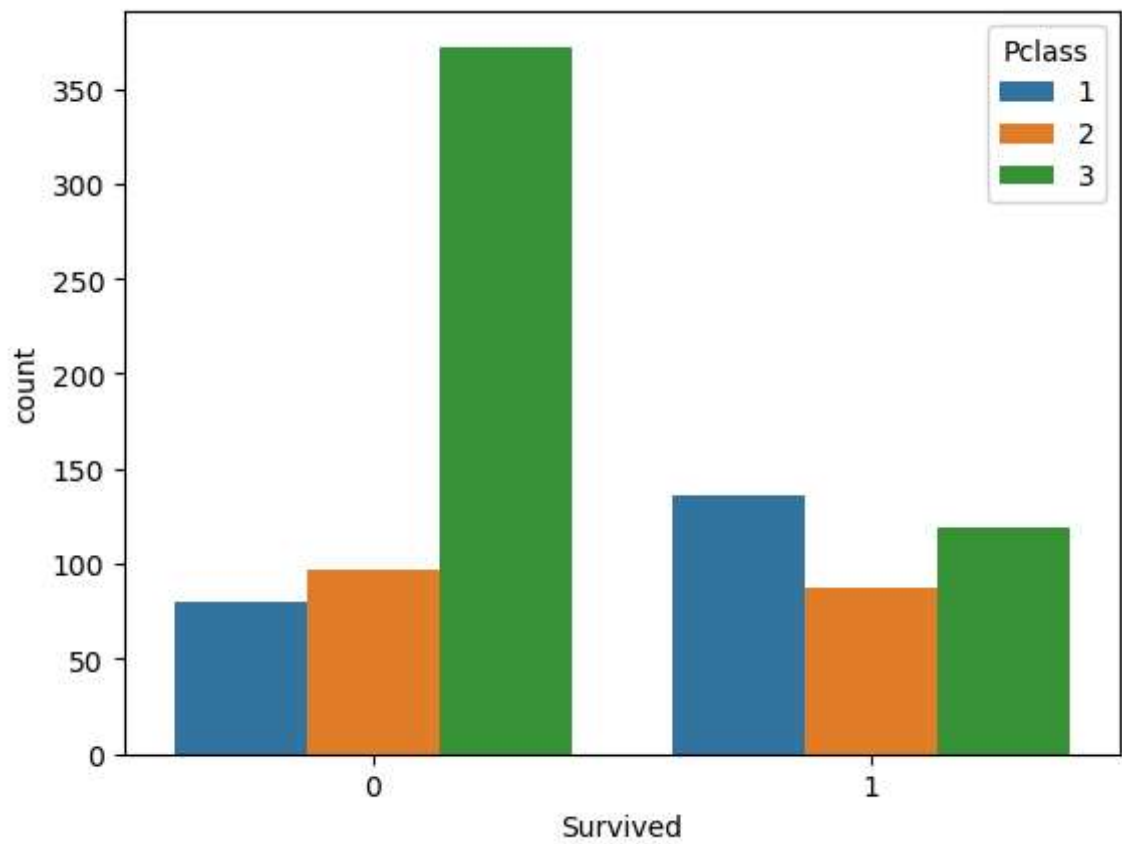
```
In [6]: sns.countplot(x='Survived',hue='Sex',data=titanic)
```

```
Out[6]: <AxesSubplot:xlabel='Survived', ylabel='count'>
```



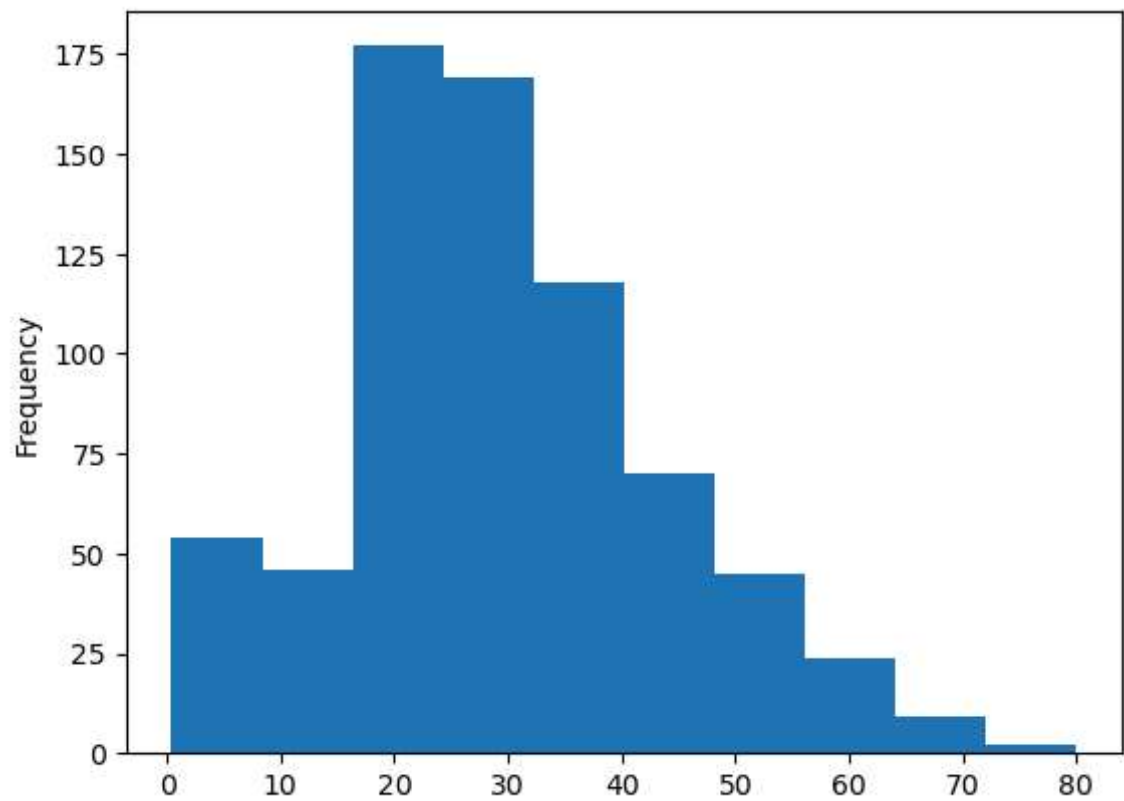
```
In [7]: sns.countplot(x='Survived',hue='Pclass',data=titanic)
```

```
Out[7]: <AxesSubplot:xlabel='Survived', ylabel='count'>
```



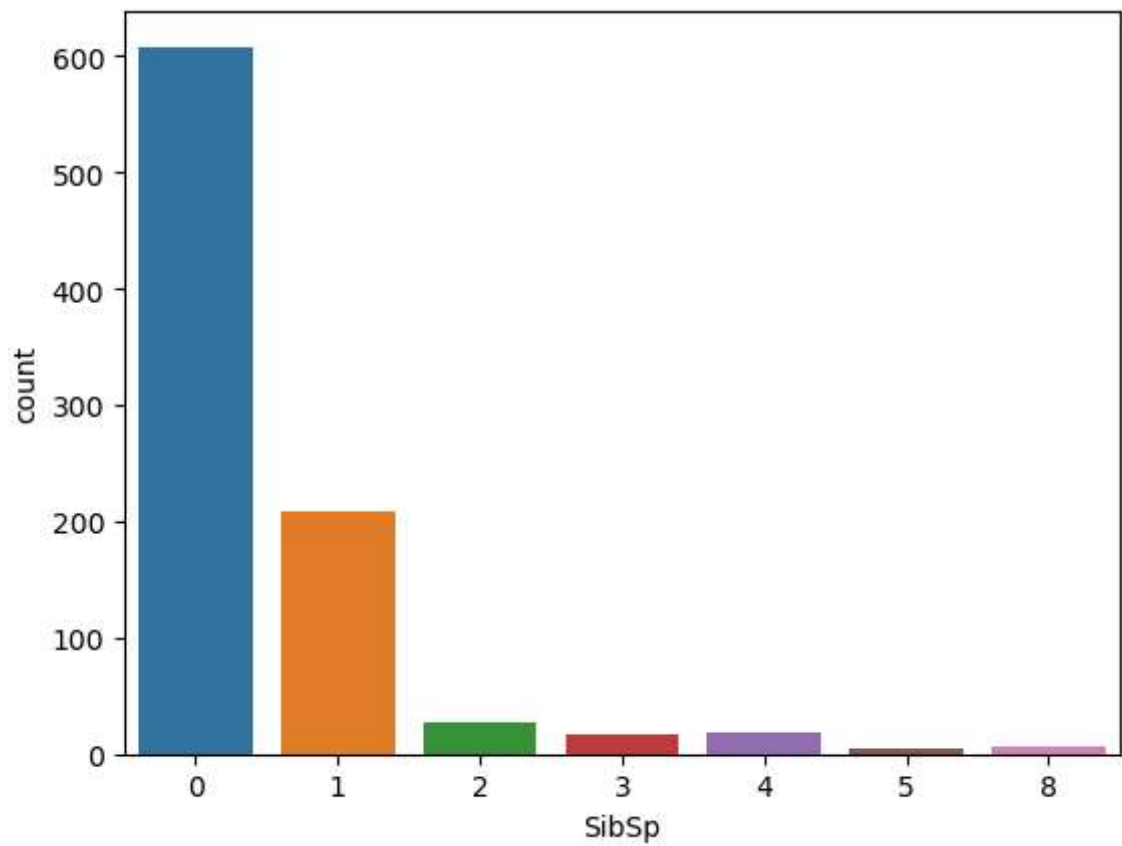
```
In [8]: titanic['Age'].plot.hist()
```

```
Out[8]: <AxesSubplot:ylabel='Frequency'>
```



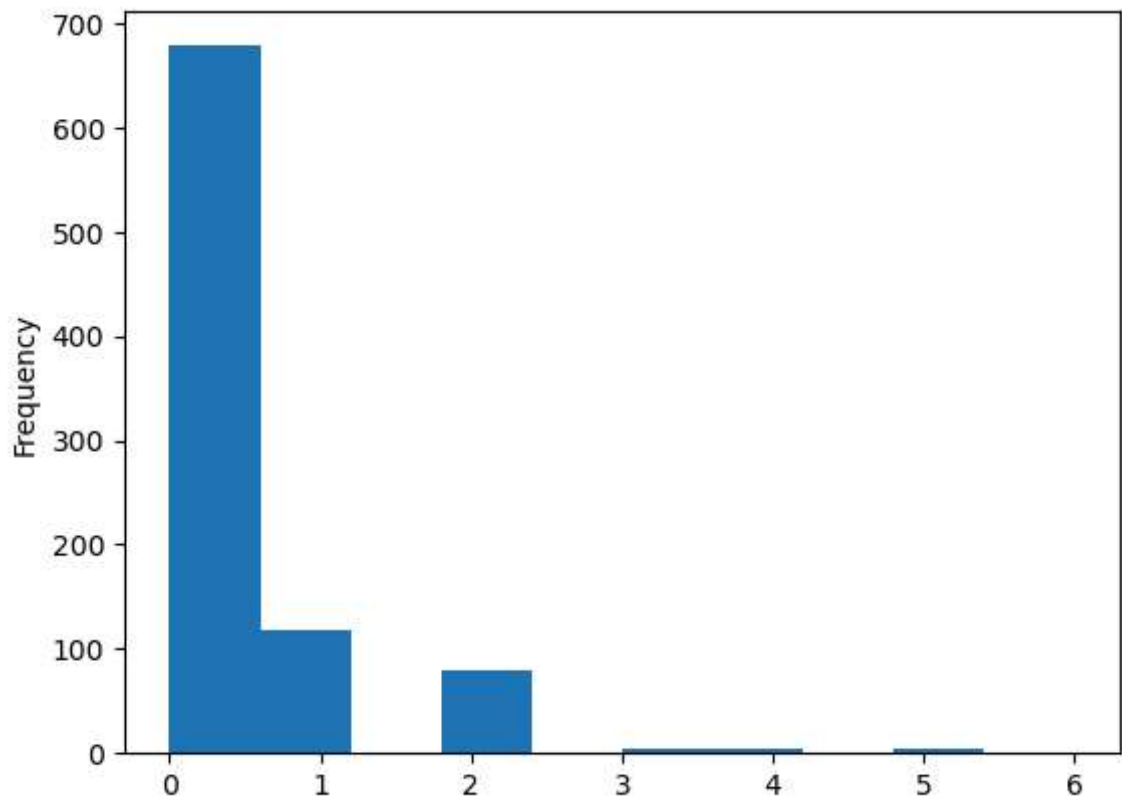
```
In [9]: sns.countplot(x='SibSp',data=titanic)
```

```
Out[9]: <AxesSubplot:xlabel='SibSp', ylabel='count'>
```



```
In [10]: titanic['Parch'].plot.hist()
```

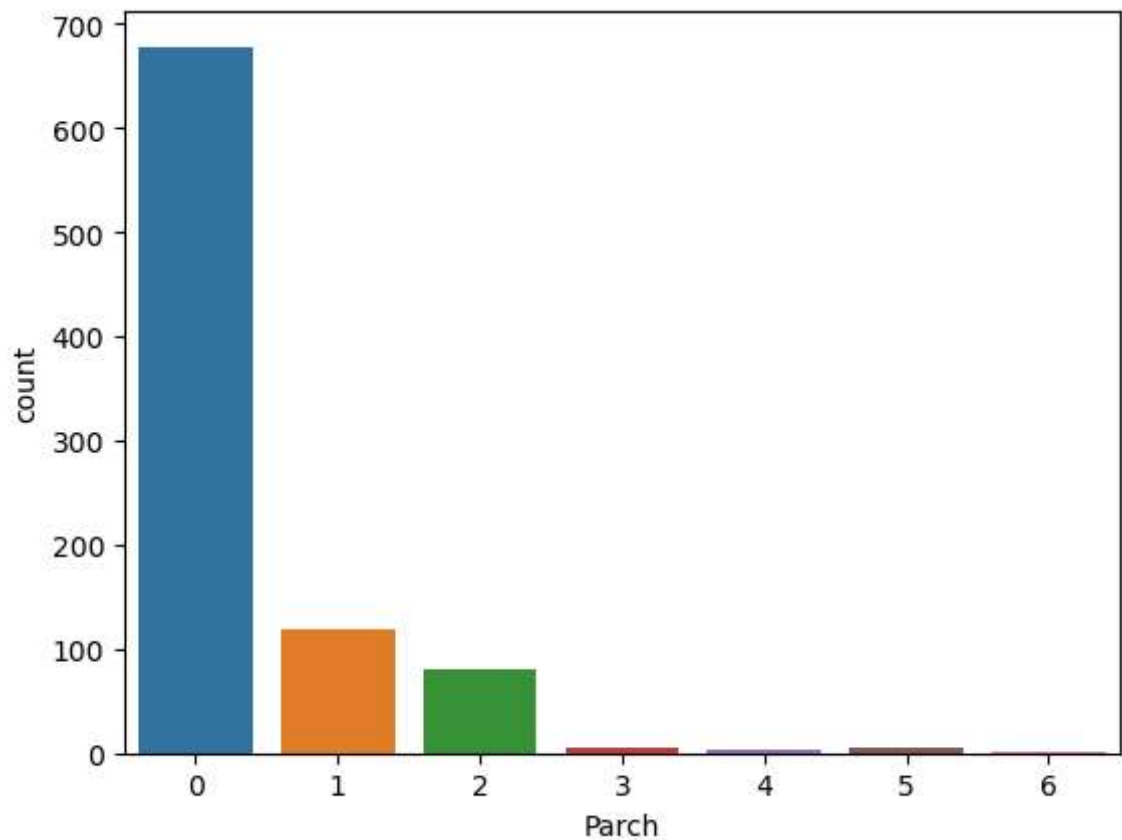
```
Out[10]: <AxesSubplot:ylabel='Frequency'>
```





```
In [11]: sns.countplot(x='Parch',data=titanic)
```

```
Out[11]: <AxesSubplot:xlabel='Parch', ylabel='count'>
```



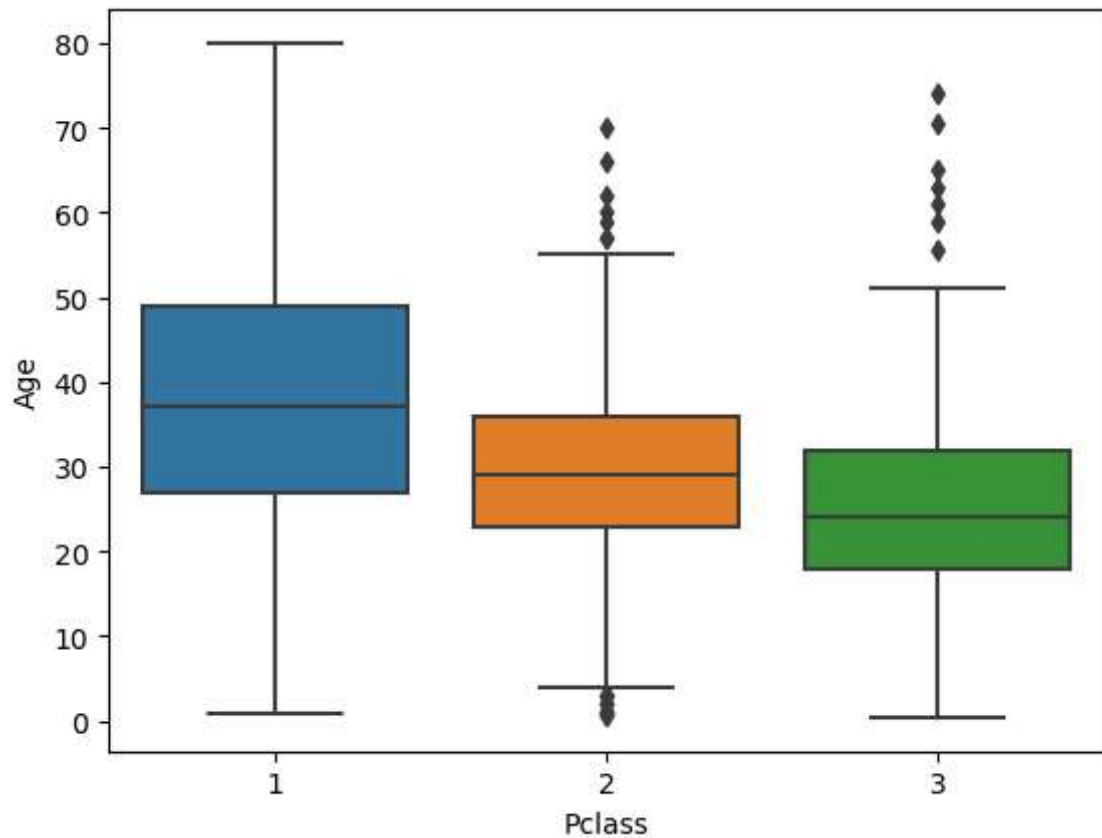
## Data Wrangling

```
In [12]: titanic.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]: sns.boxplot(x='Pclass',y='Age',data=titanic)
```

```
Out[13]: <AxesSubplot:xlabel='Pclass', ylabel='Age'>
```



```
In [14]: titanic.head()
```

```
Out[14]:
```


|   | 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   |

```
In [15]: titanic.drop(columns='Cabin',inplace=True)
```

```
In [16]: titanic.head()
```

```
Out[16]:
```

|   | 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  | S        |
| 1 | 2           | 1        | 1      | Cumings, Mrs. John Bradley (Florence Briggs Th...) | female | 38.0 | 1     | 0     | PC 17599         | 71.2833 | C        |
| 2 | 3           | 1        | 3      | Heikkinen, Miss. Laina                             | female | 26.0 | 0     | 0     | STON/O2. 3101282 | 7.9250  | S        |
| 3 | 4           | 1        | 1      | Futrelle, Mrs. Jacques Heath (Lily May Peel)       | female | 35.0 | 1     | 0     | 113803           | 53.1000 | C        |
| 4 | 5           | 0        | 3      | Allen, Mr. William Henry                           | male   | 35.0 | 0     | 0     | 373450           | 8.0500  | S        |



```
In [17]: titanic.dropna(inplace=True)
```


```
In [18]: titanic.isnull().sum()
```

```
Out[18]: 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 [19]: titanic.head(2)
```

```
Out[19]:
```

|   | PassengerId | Survived | Pclass | Name  | Sex    | Age  | SibSp | Parch | Ticket    | Fare    | Embar |
|---|-------------|----------|--------|---|--------|------|-------|-------|-----------|---------|-------|
| 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 [20]: pd.get_dummies(titanic['Sex']).head()
```

```
Out[20]:
```

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

```
In [21]: sex=pd.get_dummies(titanic['Sex'],drop_first=True)  
sex.head(3)
```

```
Out[21]:
```

|   | male |
|---|------|
| 0 | 1    |
| 1 | 0    |
| 2 | 0    |

```
In [22]: embark=pd.get_dummies(titanic['Embarked'])  
embark.head(3)
```

```
Out[22]:
```

|   | c | q | s |
|---|---|---|---|
| 0 | 0 | 0 | 1 |
| 1 | 1 | 0 | 0 |
| 2 | 0 | 0 | 1 |

```
In [23]: embark=pd.get_dummies(titanic['Embarked'],drop_first=True)
embark.head(3)
```

```
Out[23]:
```

|   | Q | S |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 0 | 0 |
| 2 | 0 | 1 |

```
In [24]: embark1=pd.get_dummies(titanic['Pclass'],drop_first=True)
embark1.head(3)
```

```
Out[24]:
```

|   | 2 | 3 |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 0 | 0 |
| 2 | 0 | 1 |

```
In [25]: titanic=pd.concat([titanic, sex, embark, embark1], axis=1)
```

```
In [26]: titanic.head(3)
```

```
Out[26]:
```

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

```
In [27]: titanic.drop(['Name', 'PassengerId', 'Pclass', 'Ticket', 'Sex', 'Embarked'], axis=1)
```

```
In [28]: titanic.head(3)
```

```
Out[28]:
```

|   | Survived | Age  | SibSp | Parch | Fare    | male | Q | S | 2 | 3 |
|---|----------|------|-------|-------|---------|------|---|---|---|---|
| 0 | 0        | 22.0 | 1     | 0     | 7.2500  | 1    | 0 | 1 | 0 | 0 |
| 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 |

## Train Data

```
In [29]: x=titanic.drop('Survived',axis=1)
        y=titanic['Survived']
```

```
In [30]: from sklearn.model_selection import train_test_split
```

```
In [31]: x_train,x_test, y_train, y_test=train_test_split(x,y, test_size=0.33, random
```

```
In [32]: model = LogisticRegression()
```

```
In [33]: model.fit(x_train,y_train)
```

C:\ProgramData\Anaconda3\lib\site-packages\sklearn\utils\validation.py:168: FutureWarning: Feature names only support names that are all strings. Got feature names with dtypes: ['int', 'str']. An error will be raised in 1.2.

warnings.warn(  
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\linear\_model\\_logistic.py:814: ConvergenceWarning: lbfgs failed to converge (status=1):  
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max\_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html> (<https://scikit-learn.org/stable/modules/preprocessing.html>)

Please also refer to the documentation for alternative solver options:

[https://scikit-learn.org/stable/modules/linear\\_model.html#logistic-regression](https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression) ([https://scikit-learn.org/stable/modules/linear\\_model.html#logistic-regression](https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression))

n\_iter\_i = \_check\_optimize\_result(  
Out[33]: LogisticRegression()

```
In [34]: y_pred = model.predict(x_test)
```

C:\ProgramData\Anaconda3\lib\site-packages\sklearn\utils\validation.py:168: FutureWarning: Feature names only support names that are all strings. Got feature names with dtypes: ['int', 'str']. An error will be raised in 1.2.

warnings.warn(  
In [35]: from sklearn.metrics import confusion\_matrix

```
In [36]: confusion_matrix(y_test,y_pred)
```

```
Out[36]: array([[127,  18],
               [ 30,  60]], dtype=int64)
```

```
In [37]: from sklearn.metrics import accuracy_score
```

```
In [38]: accuracy_score(y_test,y_pred)
```

```
Out[38]: 0.7957446808510639
```

```
In [39]: from sklearn.metrics import classification_report
```

```
In [40]: classification_report(y_test,y_pred)
```

```
Out[40]: '          precision    recall  f1-score   support\n\n 0.81          0.88          0.84        145\n 0.71          90\n\n accuracy              0.80        235\n macro avg          0.79          0.77          0.78        235\n weighted avg          0.80          0.79          0.79        235'
```

```
In [ ]:
```