```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as ply
import seaborn as sns

titanic_df=pd.read_csv(r"C:\Users\HP\Downloads\titanic train.csv")

titanic_df.head(5)
```

```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         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

   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
```

```python
titanic_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
```
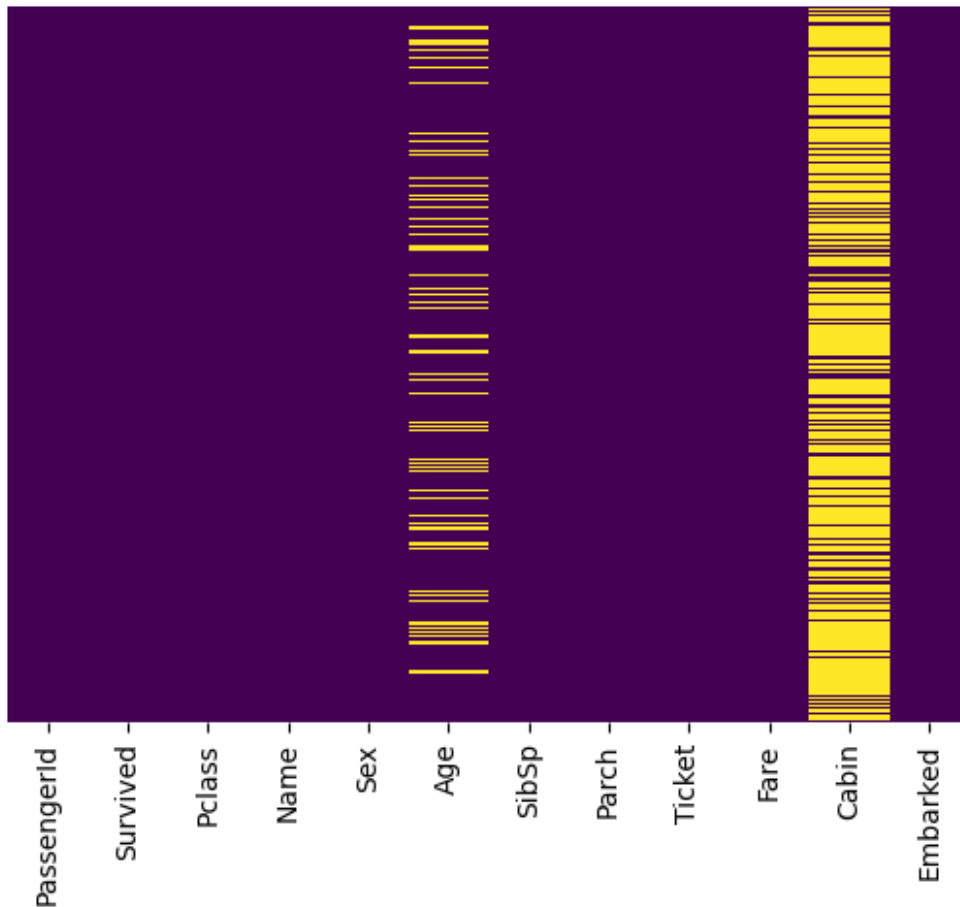
```
 8   Ticket        891 non-null    object
 9   Fare          891 non-null    float64
 10  Cabin         204 non-null    object
 11  Embarked      889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

```
sns.heatmap(titanic_df.isnull(),yticklabels=False,cbar=False,cmap='vir
idis');
```



```
titanic_df.describe()
```

	PassengerId	Survived	Pclass	Age	SibSp \
count	891.000000	891.000000	891.000000	714.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008
std	257.353842	0.486592	0.836071	14.526497	1.102743
min	1.000000	0.000000	1.000000	0.420000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000
50%	446.000000	0.000000	3.000000	28.000000	0.000000
75%	668.500000	1.000000	3.000000	38.000000	1.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000

```
           Parch          Fare
count  891.000000    891.000000
mean     0.381594     32.204208
std      0.806057     49.693429
min      0.000000      0.000000
25%      0.000000      7.910400
50%      0.000000     14.454200
75%      0.000000     31.000000
max      6.000000    512.329200
```
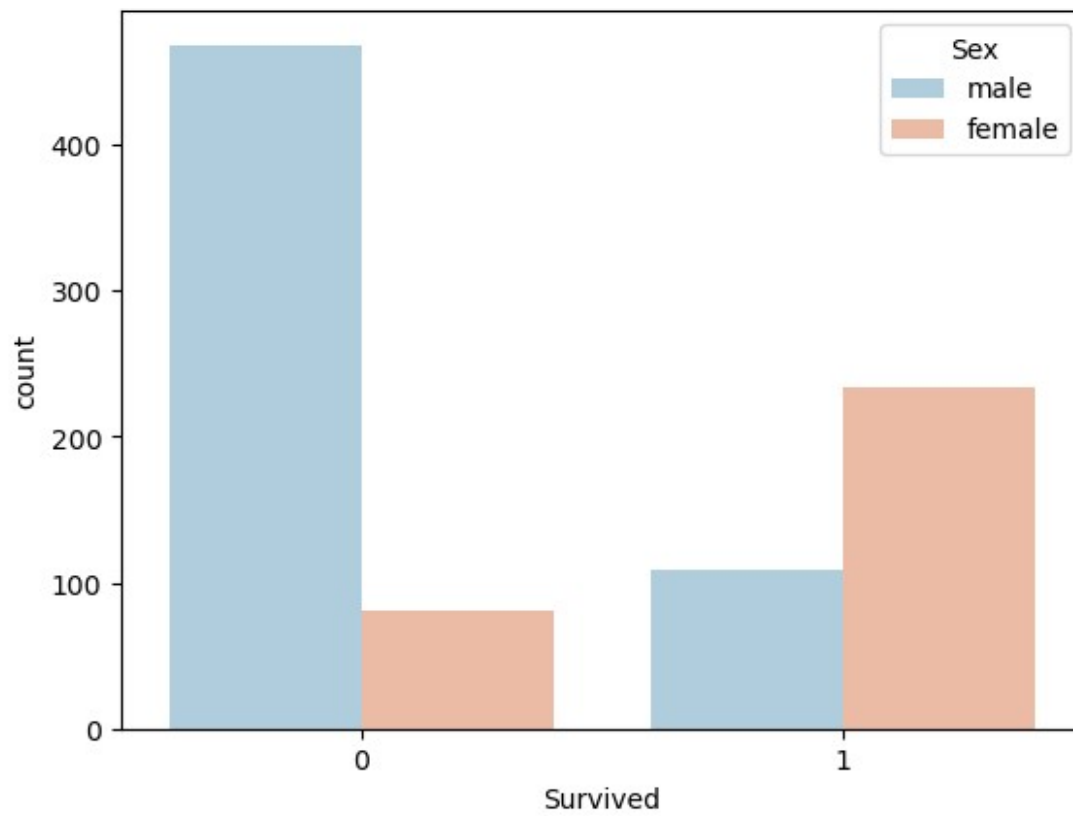
```python
sns.countplot(x='Survived',data=titanic_df);
```
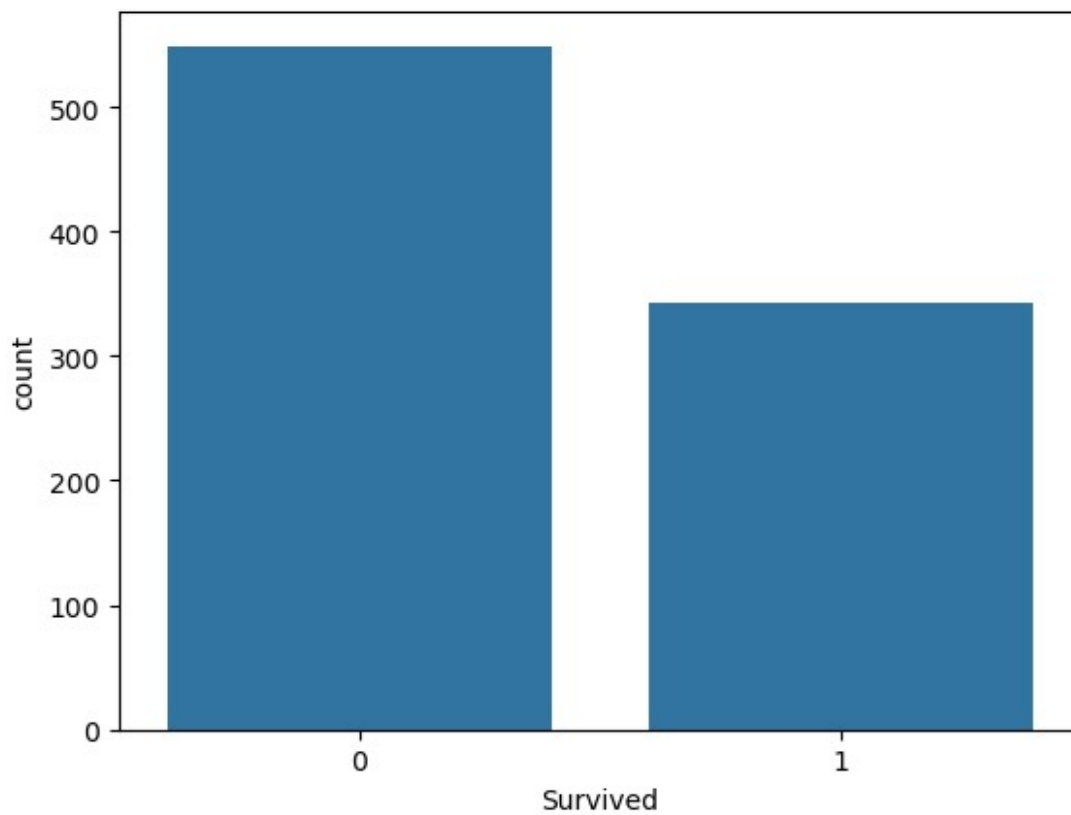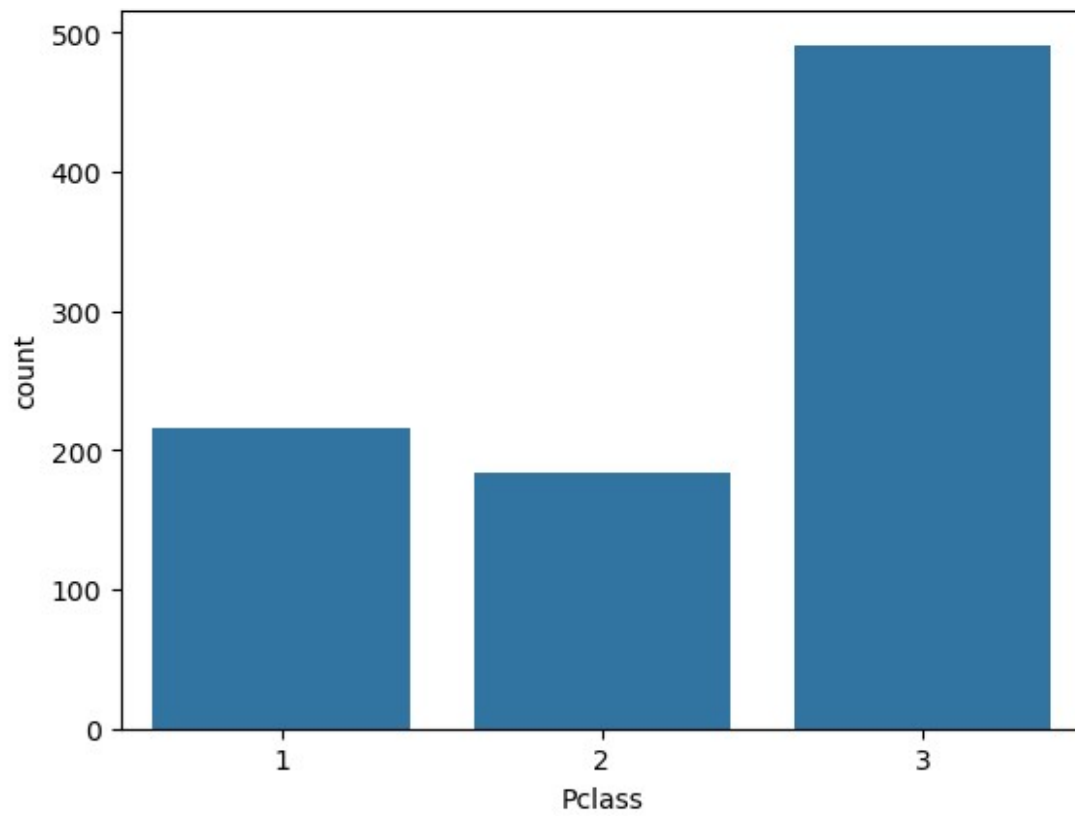

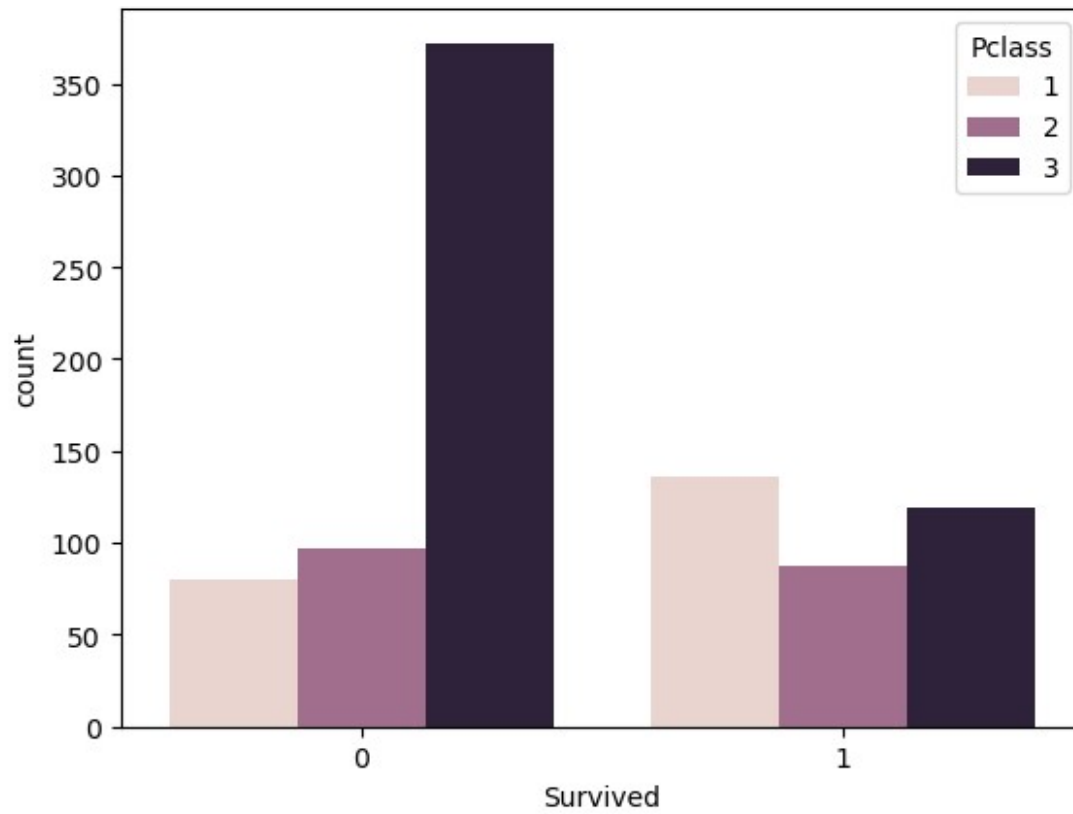```python
sns.countplot(x='Sex',data=titanic_df);
```
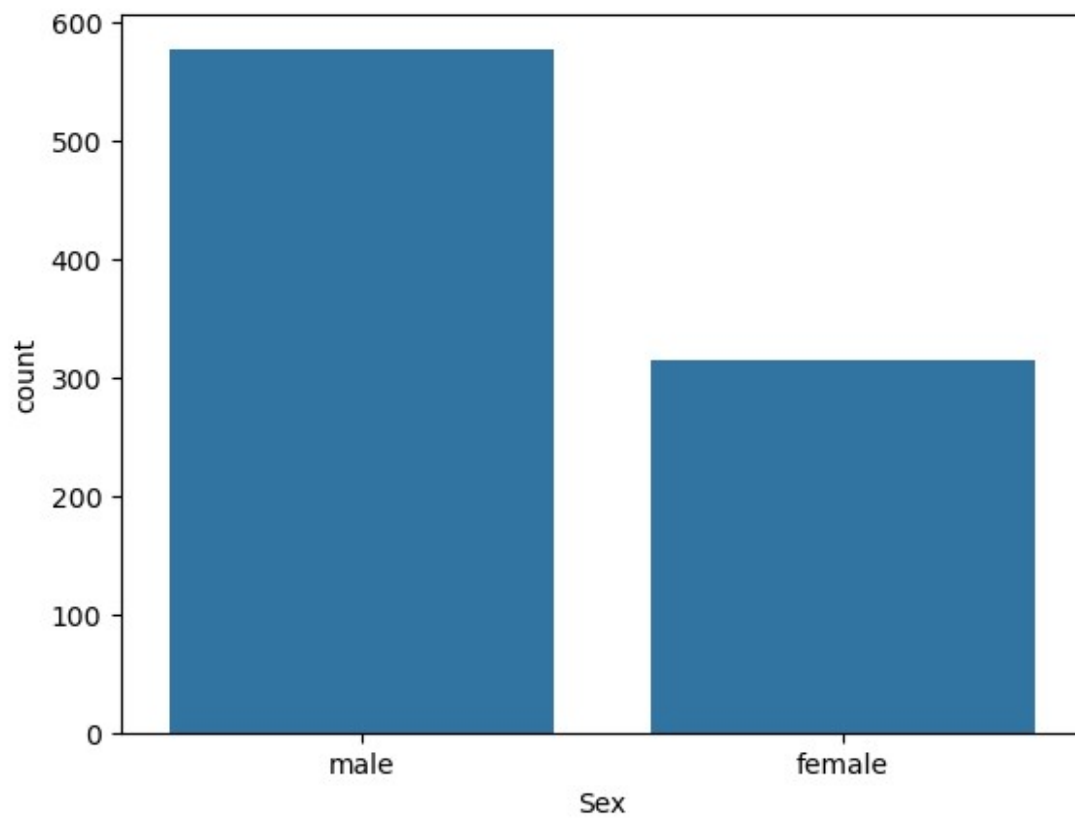
```
sns.countplot(x='Survived',hue='Sex',data=titanic_df,palette='RdBu_r')
;
```
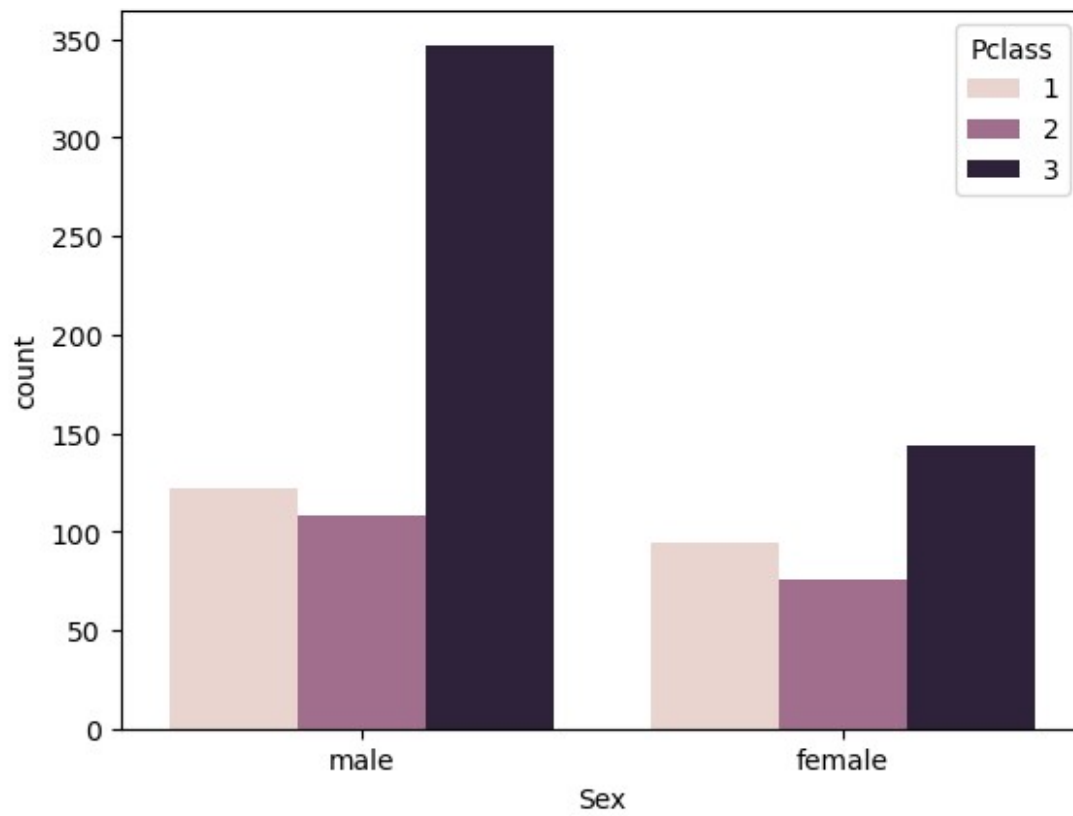
```
sns.countplot(x='Pclass',data=titanic_df);
```
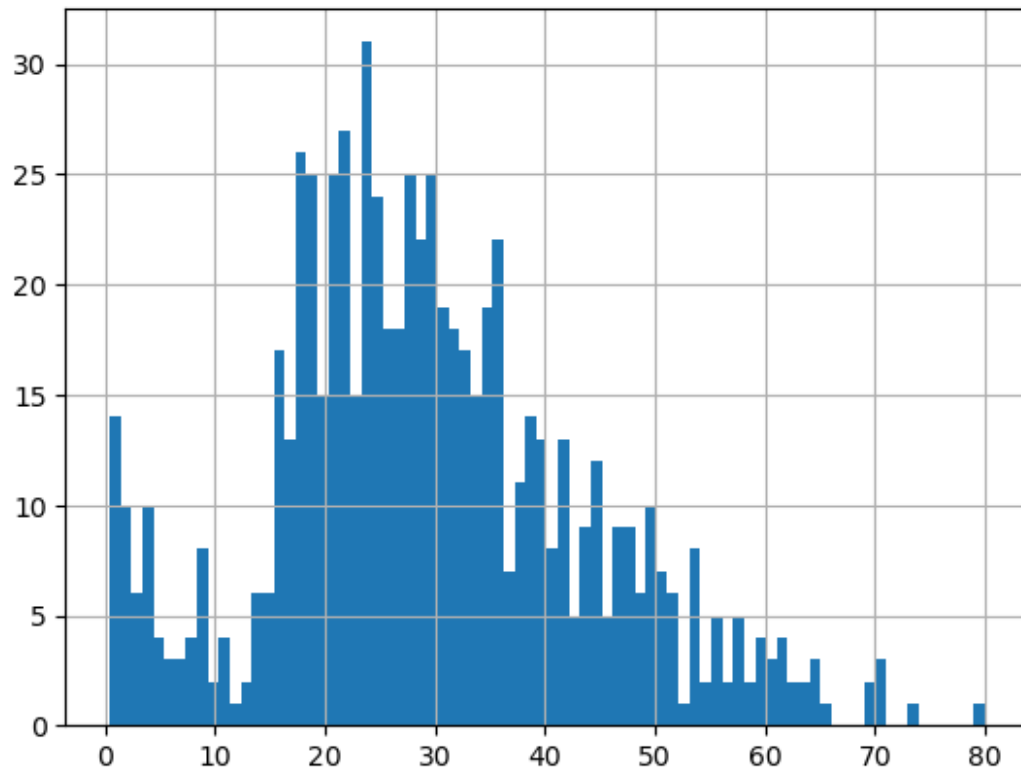
```python
sns.countplot(x='Survived',data=titanic_df,hue='Pclass');
```
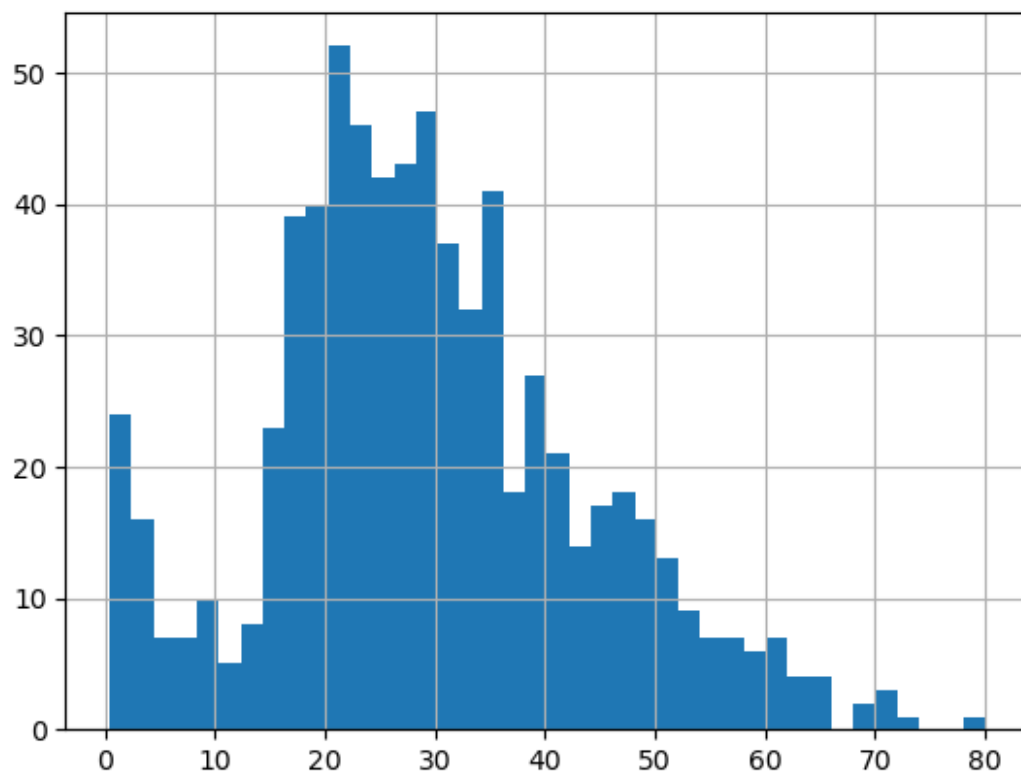
```
sns.countplot(x='Sex',data=titanic_df,hue='Pclass');
```

```
titanic_df['Age'].hist(bins=80);
```

```
titanic_df['Age'].hist(bins=40);
```

```
sns.countplot(x='Survived',data=titanic_df,hue='SibSp');
```
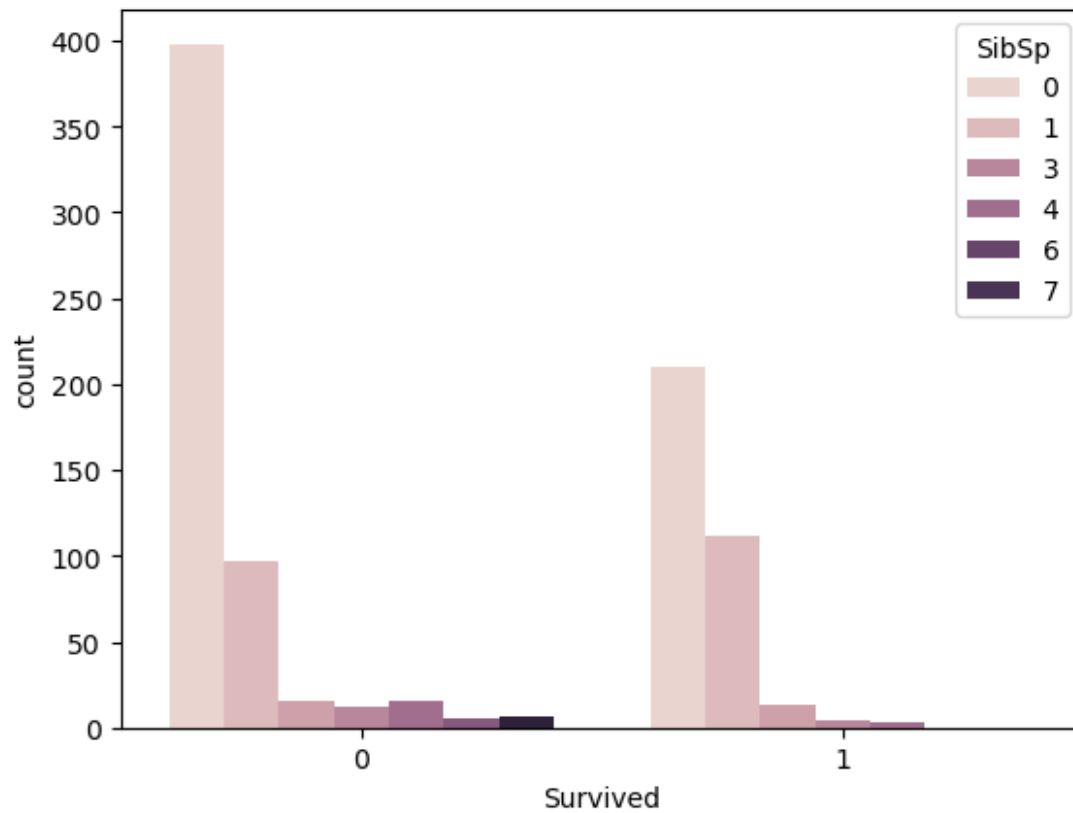


```
sns.countplot(x='Embarked',data=titanic_df,hue='Pclass');
```
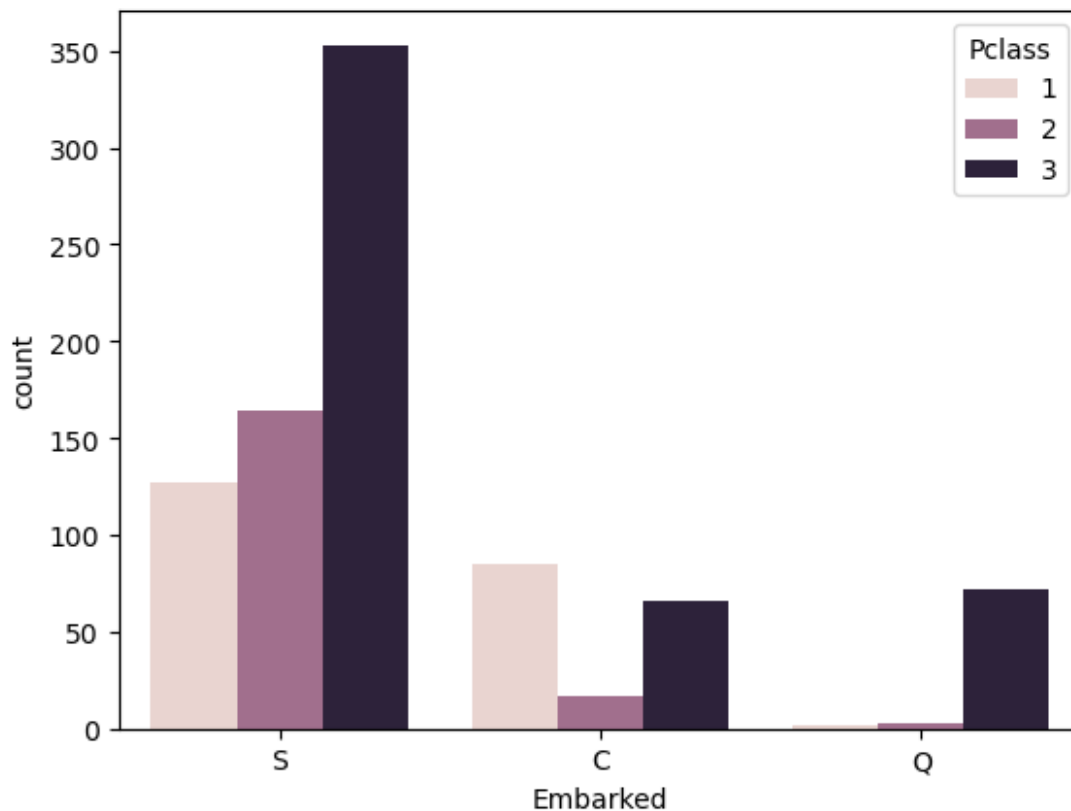
Data Cleanning

```python
titanic_df['Age'].mean()
```

```
29.69911764705882
```

```python
titanic_df.groupby(by='Pclass')['Age'].mean()
```

```
Pclass
1    38.233441
2    29.877630
3    25.140620
Name: Age, dtype: float64
```

```python
def mod_age(m):
    Age = m.iloc[0]
    Pclass = m.iloc[1]

    if pd.isnull(Age):  # Check if Age is null
        if Pclass == 1:
            return 38
        elif Pclass == 2:
            return 29
        else:
            return 25
```

```
    else:
        return Age  # If Age is not null, return it

titanic_df['Age']=titanic_df[['Age','Pclass']].apply(mod_age,axis=1)

titanic_df['Age']

0        22.0
1        38.0
2        26.0
3        35.0
4        35.0
         ...
886      27.0
887      19.0
888      25.0
889      26.0
890      32.0
Name: Age, Length: 891, dtype: float64

sns.heatmap(titanic_df.isnull(),yticklabels=False,cbar=False,cmap='vir
idis');
```
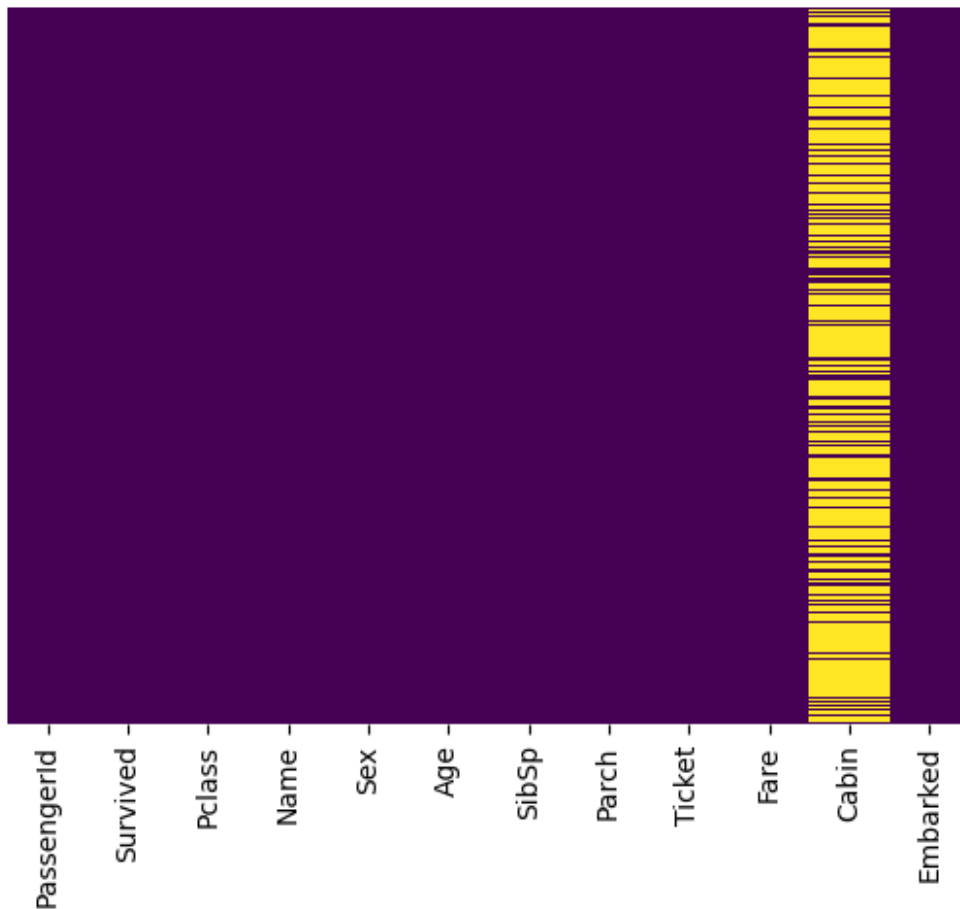
```
titanic_df.drop('Cabin',axis=1,inplace=True)

sns.heatmap(titanic_df.isnull(),yticklabels=False,cbar=False,cmap='vir
idis')
```

<Axes: >
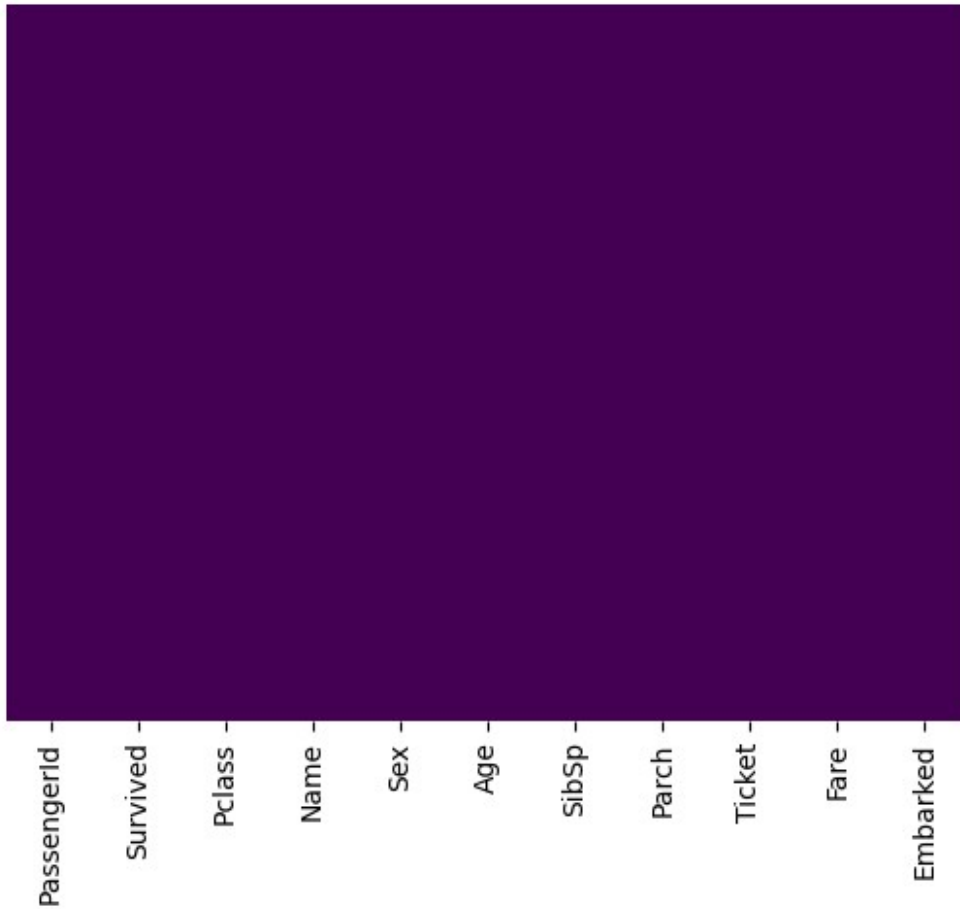

```
titanic_df.dropna

<bound method DataFrame.dropna 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   25.0
1
889                              Behr, Mr. Karl Howell    male   26.0
0
890                               Dooley, Mr. Patrick    male   32.0
0

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

[891 rows x 11 columns]>

titanic_df.head()

   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         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

   Parch            Ticket     Fare Embarked
0      0         A/5 21171   7.2500        S
1      0          PC 17599  71.2833        C
2      0  STON/O2. 3101282   7.9250        S
3      0            113803  53.1000        S
4      0            373450   8.0500        S
```

```python
sex_test=pd.get_dummies(titanic_df['Sex'],drop_first=True)
```

```python
sex_test
```

```
      male
0     True
1    False
2    False
3    False
4     True
..     ...
886   True
887  False
888  False
889   True
890   True

[891 rows x 1 columns]
```

```python
sex=sex_test.astype(int)
```

```python
embark_test=pd.get_dummies(titanic_df['Embarked'],drop_first=True)
```

```python
embark_test
```

```
         Q      S
0    False   True
1    False  False
2    False   True
3    False   True
```

```
4      False    True
..       ...      ...
886    False    True
887    False    True
888    False    True
889    False   False
890     True   False

[891 rows x 2 columns]

embrk=embark_test.astype(int)

titanic_df.drop(['Sex','Embarked','Name','Ticket'],axis=1,inplace=True
)

titanic_df.head()

    PassengerId  Survived  Pclass   Age  SibSp  Parch      Fare
0             1         0       3  22.0      1      0    7.2500
1             2         1       1  38.0      1      0   71.2833
2             3         1       3  26.0      0      0    7.9250
3             4         1       1  35.0      1      0   53.1000
4             5         0       3  35.0      0      0    8.0500

titanic_df=pd.concat([titanic_df,sex_test,embark_test],axis=1)

titanic_df

      PassengerId  Survived  Pclass   Age  SibSp  Parch      Fare   male
Q  \
0               1         0       3  22.0      1      0    7.2500   True
False
1               2         1       1  38.0      1      0   71.2833  False
False
2               3         1       3  26.0      0      0    7.9250  False
False
3               4         1       1  35.0      1      0   53.1000  False
False
4               5         0       3  35.0      0      0    8.0500   True
False
..            ...       ...     ...   ...    ...    ...       ...    ...
...
886           887         0       2  27.0      0      0   13.0000   True
False
887           888         1       1  19.0      0      0   30.0000  False
False
888           889         0       3  25.0      1      2   23.4500  False
False
889           890         1       1  26.0      0      0   30.0000   True
False
890           891         0       3  32.0      0      0    7.7500   True
```

```
True

          S
0      True
1     False
2      True
3      True
4      True
..      ...
886    True
887    True
888    True
889   False
890   False

[891 rows x 10 columns]

titanic_df['male'] = titanic_df['male'].astype(int)

titanic_df['Q'] = titanic_df['Q'].astype(int)

titanic_df['S'] = titanic_df['S'].astype(int)

titanic_df
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	male	Q	S
0	1	0	3	22.0	1	0	7.2500	1	0	1
1	2	1	1	38.0	1	0	71.2833	0	0	0
2	3	1	3	26.0	0	0	7.9250	0	0	1
3	4	1	1	35.0	1	0	53.1000	0	0	1
4	5	0	3	35.0	0	0	8.0500	1	0	1
..	...	...	...	...	...	...	...	...	..	..
886	887	0	2	27.0	0	0	13.0000	1	0	1
887	888	1	1	19.0	0	0	30.0000	0	0	1
888	889	0	3	25.0	1	2	23.4500	0	0	1
889	890	1	1	26.0	0	0	30.0000	1	0	0
890	891	0	3	32.0	0	0	7.7500	1	1	0

```
[891 rows x 10 columns]
```

Logistic regression

```python
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix

X_train, X_test, Y_train, Y_test=
train_test_split(titanic_df.drop(('Survived'),axis=1),titanic_df['Surv
ived'],test_size=0.30, random_state=101)

model=LogisticRegression(max_iter=1000)
model.fit(X_train, Y_train)
```

```
LogisticRegression(max_iter=1000)
```

```python
predictions=model.predict(X_test)
```

```python
predictions
```

```
array([0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0,
0,
       1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1,
0,
       0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0,
0,
       1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0,
1,
       0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1,
0,
       0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1,
1,
       1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0,
0,
       0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0,
0,
       1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
1,
       0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
1,
       1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1,
1,
       0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0,
0,
       1, 0, 0, 1], dtype=int64)
```

```python
accuracy=accuracy_score(Y_test, predictions)
```

```python
accuracy
```

```
0.7873134328358209
```

```
cm=confusion_matrix(Y_test, predictions)

cm

array([[134,  20],
       [ 37,  77]], dtype=int64)
```