

```
In [2]: import pandas as pd
data =pd.read_csv("C:\\Users\\sriha\\OneDrive\\Desktop\\pb excel\\Titanic.csv")
df=pd.DataFrame(data)
df
```

```
Out[2]:
```

	PassengerId	Name	Class	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	Braund	3	male	22	1	0	21171	7.2500	
1	2	Cumings	1	female	38	1	0	17599	71.2833	C85
2	3	Heikkinen	3	female	26	0	0	3101282	7.9250	
3	4	Futrelle	1	female	35	1	0	113803	53.1000	C123
4	5	Allen	3	male	35	0	0	373450	8.0500	
5	6	Moran	3	male	0	0	0	330877	8.4583	
6	7	McCarthy	1	male	54	0	0	17463	51.8625	E46
7	8	Palsson	3	male	2	3	1	349909	21.0750	
8	9	Johnson	3	female	27	0	2	347742	11.1333	
9	10	Nasser	2	female	14	1	0	237736	30.0708	

```
In [3]: df.drop(['PassengerId','Name','SibSp','Parch','Ticket','Cabin','Embarked'], axis=1)
df
```

```
Out[3]:
```

	Class	Sex	Age	Fare	Survived
0	3	male	22	7.2500	0
1	1	female	38	71.2833	1
2	3	female	26	7.9250	1
3	1	female	35	53.1000	1
4	3	male	35	8.0500	0
5	3	male	0	8.4583	0
6	1	male	54	51.8625	0
7	3	male	2	21.0750	0
8	3	female	27	11.1333	1
9	2	female	14	30.0708	1

```
In [4]: target = df.Survived
target
```

```
Out[4]: 0    0
        1    1
        2    1
        3    1
        4    0
        5    0
        6    0
        7    0
        8    1
        9    1
        Name: Survived, dtype: int64
```

```
In [8]: df1 = df.drop('Survived', axis = 'columns')
        df1
```

```
Out[8]:
```

	Class	Sex	Age	Fare
0	3	male	22	7.2500
1	1	female	38	71.2833
2	3	female	26	7.9250
3	1	female	35	53.1000
4	3	male	35	8.0500
5	3	male	0	8.4583
6	1	male	54	51.8625
7	3	male	2	21.0750
8	3	female	27	11.1333
9	2	female	14	30.0708

```
In [9]: # sex converted into integers
```

```
In [11]: dummy = pd.get_dummies(df1.Sex)
         dummy.head()
```

```
Out[11]:
```

	female	male
0	0	1
1	1	0
2	1	0
3	1	0
4	0	1

```
In [15]: df1 = pd.concat([df, dummy], axis = 'columns')
         df1
```

Out[15]:

	Class	Sex	Age	Fare	Survived	female	male
0	3	male	22	7.2500	0	0	1
1	1	female	38	71.2833	1	1	0
2	3	female	26	7.9250	1	1	0
3	1	female	35	53.1000	1	1	0
4	3	male	35	8.0500	0	0	1
5	3	male	0	8.4583	0	0	1
6	1	male	54	51.8625	0	0	1
7	3	male	2	21.0750	0	0	1
8	3	female	27	11.1333	1	1	0
9	2	female	14	30.0708	1	1	0

In [16]: `df1.drop(['Sex','Survived'], axis='columns', inplace=True)`
`df1`

Out[16]:

	Class	Age	Fare	female	male
0	3	22	7.2500	0	1
1	1	38	71.2833	1	0
2	3	26	7.9250	1	0
3	1	35	53.1000	1	0
4	3	35	8.0500	0	1
5	3	0	8.4583	0	1
6	1	54	51.8625	0	1
7	3	2	21.0750	0	1
8	3	27	11.1333	1	0
9	2	14	30.0708	1	0

train_test_split

In [17]: `X= df1`
`y= target`

In [18]: `from sklearn.model_selection import train_test_split`
`X_train,X_test,y_train,y_test = train_test_split (X,y,test_size= 0.3)`

In [20]: `len(X_train)`

Out[20]: 7

In [21]: `len (X_test)`

Out[21]: 3

Naive Bayes

In [35]: `from sklearn.naive_bayes import CategoricalNB,GaussianNB
nb= CategoricalNB(min_categories=5)
nb1=GaussianNB()
nb1.fit(X_train,y_train)`

Out[35]: GaussianNB()

In [33]: `nb1.score(X_test,y_test)`

Out[33]: 0.3333333333333333

In [36]: `X_test[:10]`

Out[36]:

	Class	Age	Fare	female	male
8	3	27	11.1333	1	0
5	3	0	8.4583	0	1
4	3	35	8.0500	0	1

In [37]: `y_test[:10]`

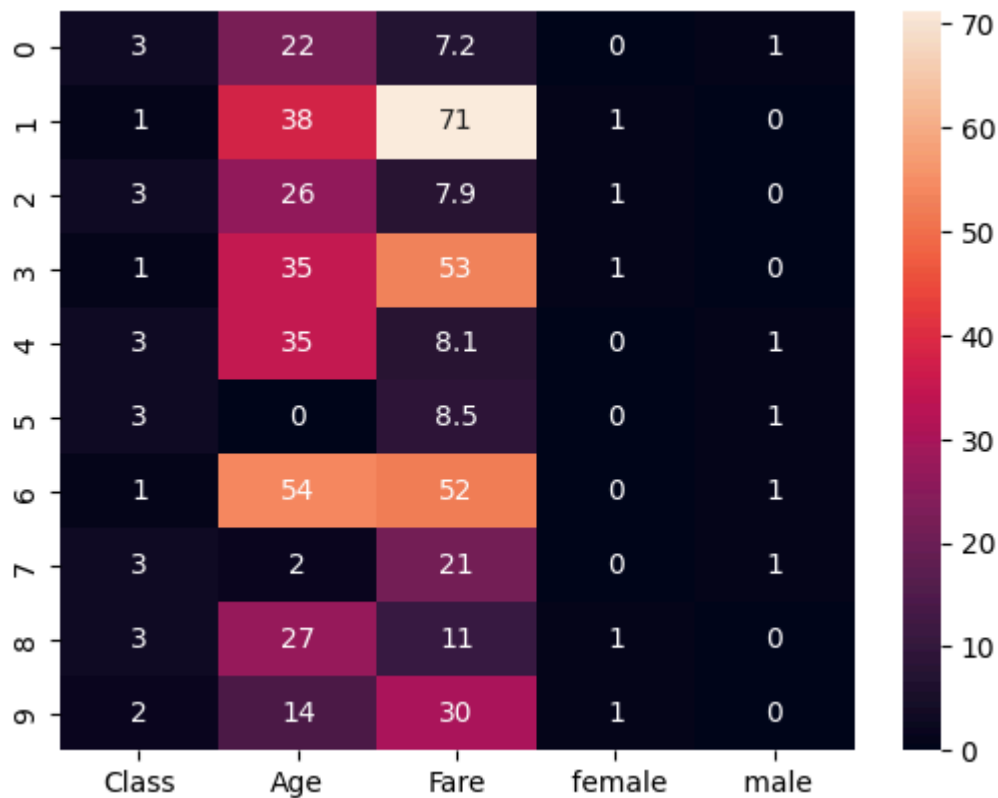
Out[37]:

8	1
5	0
4	0

Name: Survived, dtype: int64

In [39]: `import matplotlib.pyplot as pp
import seaborn as sn
sn.heatmap(df1, annot=True)`

Out[39]: <AxesSubplot:>



```
In [45]: y_pred=nb1.predict(X_test)
```

```
In [46]: from sklearn.metrics import confusion_matrix
cm=confusion_matrix(y_test,y_pred)
cm
```

```
Out[46]: array([[2, 0],
               [0, 1]], dtype=int64)
```

```
In [47]: import matplotlib.pyplot as pp
import seaborn as sn
sn.heatmap(cm, annot=True)
```

```
Out[47]: <AxesSubplot:>
```

