

```
import pandas as pd
```

```
data=pd.read_csv('/content/sample_data/Titanic.csv')
data.head()
```

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	I
0	1	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	
2	3	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	

```
data.drop(['PassengerId', 'Name', 'SibSp', 'Parch', 'Ticket', 'Cabin', 'Embarked'],axis='columns',inplace=True)
data.head()
```

	Pclass	Sex	Age	Fare	Survived	
0	3	male	22.0	7.2500	0	
1	1	female	38.0	71.2833	1	
2	3	female	26.0	7.9250	1	
3	1	female	35.0	53.1000	1	
4	3	male	35.0	8.0500	0	

```
target=data.Survived
inputs=data.drop('Survived',axis='columns')
```

```
dummies=pd.get_dummies(inputs.Sex)
dummies.head(5)
```

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

```
inputs=pd.concat([inputs,dummies],axis='columns')
inputs.head(3)
```

	Pclass	Sex	Age	Fare	female	male
0	3	male	22.0	7.2500	0	1
1	1	female	38.0	71.2833	1	0
2	3	female	26.0	7.9250	1	0

```
inputs.drop('Sex',axis='columns',inplace=True)
inputs.head()
```

	Pclass	Age	Fare	female	male
0	3	22.0	7.2500	0	1
1	1	38.0	71.2833	1	0
2	3	26.0	7.9250	1	0
3	1	35.0	53.1000	1	0
4	3	35.0	8.0500	0	1

```
inputs.columns[inputs.isna().any()]

Index(['Age'], dtype='object')
```

```
inputs.Age[:10]
```

```
0    22.0
1    38.0
2    26.0
3    35.0
4    35.0
5     NaN
6    54.0
7     2.0
8    27.0
9    14.0
Name: Age, dtype: float64
```

```
inputs.Age=inputs.Age.fillna(inputs.Age.mean())
inputs.head(6)
```

	Pclass	Age	Fare	female	male
0	3	22.000000	7.2500	0	1
1	1	38.000000	71.2833	1	0
2	3	26.000000	7.9250	1	0
3	1	35.000000	53.1000	1	0
4	3	35.000000	8.0500	0	1
5	3	29.699118	8.4583	0	1

**\*\* Splitting the dataset into the Training set and Test set \*\***

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(inputs,target,test_size=0.2)
```

```
len(x_train)
```

```
712
```

```
len(x_test)
```

```
179
```

```
len(inputs)
```

```
891
```

```
from sklearn.naive_bayes import GaussianNB
model=GaussianNB()
```

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

```
▼ GaussianNB
GaussianNB()
```

```
model.score(x_test,y_test)
```

```
0.8044692737430168
```

```
y_test[:10]
```

```
853    1
536    0
696    0
744    1
202    0
22     1
114    0
166    1
246    0
390    1
Name: Survived, dtype: int64
```

```
model.predict(x_test[:10])
```

```
array([1, 0, 0, 0, 0, 1, 1, 1, 1, 1])
```

```
model.predict_proba(x_test[:10])
```

```
array([[0.00610201, 0.99389799],
       [0.93261767, 0.06738233],
       [0.98970595, 0.01029405],
       [0.98917437, 0.01082563],
       [0.98943578, 0.01056422],
       [0.04677741, 0.95322259],
       [0.05062631, 0.94937369],
       [0.00577684, 0.99422316],
       [0.06104856, 0.93895144],
       [0.17648459, 0.82351541]])
```