In [2]:
```
import pandas as pd
```

In [4]:
```
df = pd.read_csv("titanic.csv")
df.head()
```

Out[4]:

	PassengerId	Name	Pclass	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
0	1	Braund, Mr. Owen Harris	3	male	22.0	1	0	A/5 21171	7.2500	NaN	
1	2	Cumings, Mrs. John Bradley (Florence Briggs Th...	1	female	38.0	1	0	PC 17599	71.2833	C85	
2	3	Heikkinen, Miss. Laina	3	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	
3	4	Futrelle, Mrs. Jacques Heath (Lily May Peel)	1	female	35.0	1	0	113803	53.1000	C123	
4	5	Allen, Mr. William Henry	3	male	35.0	0	0	373450	8.0500	NaN	

In [6]:
```
df.drop(['PassengerId','Name','SibSp','Parch','Ticket','Cabin','Embarked'],axis='co
df.head()
```

Out[6]:

	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

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

In [10]:
```
#inputs.Sex = inputs.Sex.map({'male': 1, 'female': 2})
```

In [12]:
```
dummies = pd.get_dummies(inputs.Sex)
dummies.head(3)
```

Out[12]:

	female	male
0	False	True
1	True	False
2	True	False

In [14]:
```
inputs = pd.concat([inputs,dummies],axis='columns')
inputs.head(3)
```

Out[14]:

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

In [17]:
```
inputs.drop(['Sex','male'],axis='columns',inplace=True)
inputs.head(3)
```

Out[17]:

	Pclass	Age	Fare	female
0	3	22.0	7.2500	False
1	1	38.0	71.2833	True
2	3	26.0	7.9250	True

In [19]:
```
inputs.columns[inputs.isna().any()]
```

Out[19]:
```
Index(['Age'], dtype='object')
```

In [21]:
```
inputs.Age[:10]
```

Out[21]:
```
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
```

In [23]:
```
inputs.Age = inputs.Age.fillna(inputs.Age.mean())
inputs.head()
```

Out[23]:

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

In [25]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(inputs,target,test_size=0.3)
```

In [26]:
```
from sklearn.naive_bayes import GaussianNB
model = GaussianNB()
```

In [27]:
```
model.fit(X_train,y_train)
```

Out[27]:

GaussianNB

```
GaussianNB()
```

In [28]:
```
model.score(X_test,y_test)
```

Out[28]:
```
0.746268656716418
```

In [29]:
```
X_test[0:10]
```

Out[29]:

	Pclass	Age	Fare	female
37	3	21.000000	8.0500	False
527	1	29.699118	221.7792	False
570	2	62.000000	10.5000	False
528	3	39.000000	7.9250	False
209	1	40.000000	31.0000	False
175	3	18.000000	7.8542	False
767	3	30.500000	7.7500	True
2	3	26.000000	7.9250	True
584	3	29.699118	8.7125	False
457	1	29.699118	51.8625	True

In [30]:
```
y_test[0:10]
```

Out[30]:
```
37     0
527    0
570    1
528    0
209    1
175    0
767    0
2      1
584    0
457    1
Name: Survived, dtype: int64
```

In [31]:
```
model.predict(X_test[0:10])
```

Out[31]:
```
array([0, 1, 0, 0, 0, 0, 0, 1, 0, 1], dtype=int64)
```

In [32]:
```
model.predict_proba(X_test[:10])
```

Out[32]:
```
array([[9.70320893e-01, 2.96791068e-02],
       [1.39498433e-13, 1.00000000e+00],
       [9.02849425e-01, 9.71505755e-02],
       [9.75013075e-01, 2.49869248e-02],
       [7.38808427e-01, 2.61191573e-01],
       [9.67825229e-01, 3.21747706e-02],
       [5.12781608e-01, 4.87218392e-01],
       [4.98861161e-01, 5.01138839e-01],
       [9.74563739e-01, 2.54362605e-02],
       [3.31940494e-02, 9.66805951e-01]])
```