In [12]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings("ignore")
```

In [13]:
```
df=pd.read_csv("Titanic-Dataset.csv")
df
```

Out[13]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cal
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	N
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	N
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C1
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	N
...	...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	N
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	E
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	N
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C1
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	N

891 rows × 12 columns

In [14]:
```
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
```

In [15]:
```
df.isnull().sum()
```

Out[15]:
```
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 [16]:
```
a=df.Age.mean()
a=round(a,1)
a
```

Out[16]:
```
29.7
```

In [17]:
```
df.Age.fillna(a,inplace=True)
```

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

Out[18]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age              0
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

In [19]:
```
df.drop(["Name","SibSp","Parch","Embarked","Pclass","Ticket","Cabin"],axis="columns
```

In [20]:
```
df
```

Out[20]:

	PassengerId	Survived	Sex	Age	Fare
0	1	0	male	22.0	7.2500
1	2	1	female	38.0	71.2833
2	3	1	female	26.0	7.9250
3	4	1	female	35.0	53.1000
4	5	0	male	35.0	8.0500
...	...	...	...	...	...
886	887	0	male	27.0	13.0000
887	888	1	female	19.0	30.0000
888	889	0	female	29.7	23.4500
889	890	1	male	26.0	30.0000
890	891	0	male	32.0	7.7500

891 rows × 5 columns

In [22]:
```
df.isnull().sum()
```

Out[22]:
```
PassengerId    0
Survived       0
Sex            0
Age            0
Fare           0
dtype: int64
```

In [23]:
```
df.head()
```

Out[23]:

	PassengerId	Survived	Sex	Age	Fare
0	1	0	male	22.0	7.2500
1	2	1	female	38.0	71.2833
2	3	1	female	26.0	7.9250
3	4	1	female	35.0	53.1000
4	5	0	male	35.0	8.0500

In [25]:
```
df["Sex"]=df.Sex.apply(lambda i:1 if i=="male" else 0)
```

In [26]:
```
df
```

Out[26]:

	PassengerId	Survived	Sex	Age	Fare
0	1	0	1	22.0	7.2500
1	2	1	0	38.0	71.2833
2	3	1	0	26.0	7.9250
3	4	1	0	35.0	53.1000
4	5	0	1	35.0	8.0500
...	...	...	...	...	...
886	887	0	1	27.0	13.0000
887	888	1	0	19.0	30.0000
888	889	0	0	29.7	23.4500
889	890	1	1	26.0	30.0000
890	891	0	1	32.0	7.7500

891 rows × 5 columns

In [27]:
```
inputs=df.drop("Survived",axis="columns")
target=df.Survived
target
```

Out[27]:
```
0      0
1      1
2      1
3      1
4      0
      ..
886    0
887    1
888    0
889    1
890    0
Name: Survived, Length: 891, dtype: int64
```

In [28]:
```
x=df.iloc[:,1:]
x
```

Out[28]:

	Survived	Sex	Age	Fare
0	0	1	22.0	7.2500
1	1	0	38.0	71.2833
2	1	0	26.0	7.9250
3	1	0	35.0	53.1000
4	0	1	35.0	8.0500
...	...	...	...	...
886	0	1	27.0	13.0000
887	1	0	19.0	30.0000
888	0	0	29.7	23.4500
889	1	1	26.0	30.0000
890	0	1	32.0	7.7500

891 rows × 4 columns

In [29]:
```
y=df["Survived"]
y
```

Out[29]:
```
0      0
1      1
2      1
3      1
4      0
      ..
886    0
887    1
888    0
889    1
890    0
Name: Survived, Length: 891, dtype: int64
```

In [30]:
```
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 [31]:
```
print(x_train.shape)
print(y_train.shape)
print(x_test.shape)
print(x_train.shape)
```

```
(623, 4)
(623,)
(268, 4)
(623, 4)
```

In [34]:
```
from sklearn.naive_bayes import GaussianNB
model=GaussianNB()
```

In [36]:
```
model.fit(x_train,y_train)
```

Out[36]:
```
▾ GaussianNB

GaussianNB()
```

In [37]:
```
model.score(x_test,y_test)
```

Out[37]:
```
1.0
```

In [38]:
```
k=model.predict([[1,0,25,10]])
k
```

Out[38]:
```
array([1], dtype=int64)
```

In [39]:
```
if k[0]==0:
    print("the person has not survived")
else:
    print("the person has survived")
```

```
the person has survived
```

In [ ]:
```

```