In [1]:
```
import pandas as pd
data=pd.read_csv("/home/placement/Downloads/TelecomCustomerChurn.csv")
```

In [2]:
```
data['TotalCharges'] = pd.to_numeric(data['TotalCharges'],errors='coerce')
```

In [3]:
```
data.describe()
```

Out[3]:

	SeniorCitizen	tenure	MonthlyCharges	TotalCharges
count	7043.000000	7043.000000	7043.000000	7032.000000
mean	0.162147	32.371149	64.761692	2283.300441
std	0.368612	24.559481	30.090047	2266.771362
min	0.000000	0.000000	18.250000	18.800000
25%	0.000000	9.000000	35.500000	401.450000
50%	0.000000	29.000000	70.350000	1397.475000
75%	0.000000	55.000000	89.850000	3794.737500
max	1.000000	72.000000	118.750000	8684.800000

In [4]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   customerID        7043 non-null   object 
 1   gender            7043 non-null   object 
 2   SeniorCitizen     7043 non-null   int64  
 3   Partner           7043 non-null   object 
 4   Dependents        7043 non-null   object 
 5   tenure            7043 non-null   int64  
 6   PhoneService      7043 non-null   object 
 7   MultipleLines     7043 non-null   object 
 8   InternetService   7043 non-null   object 
 9   OnlineSecurity    7043 non-null   object 
 10  OnlineBackup      7043 non-null   object 
 11  DeviceProtection  7043 non-null   object 
 12  TechSupport       7043 non-null   object 
 13  StreamingTV       7043 non-null   object 
 14  StreamingMovies   7043 non-null   object 
 15  Contract          7043 non-null   object 
 16  PaperlessBilling  7043 non-null   object 
 17  PaymentMethod     7043 non-null   object 
 18  MonthlyCharges    7043 non-null   float64
 19  TotalCharges      7032 non-null   float64
 20  Churn             7043 non-null   object 
dtypes: float64(2), int64(2), object(17)
memory usage: 1.1+ MB
```

In [5]:
```
list(data)
```

Out[5]:
```
['customerID',
 'gender',
 'SeniorCitizen',
 'Partner',
 'Dependents',
 'tenure',
 'PhoneService',
 'MultipleLines',
 'InternetService',
 'OnlineSecurity',
 'OnlineBackup',
 'DeviceProtection',
 'TechSupport',
 'StreamingTV',
 'StreamingMovies',
 'Contract',
 'PaperlessBilling',
 'PaymentMethod',
 'MonthlyCharges',
 'TotalCharges',
 'Churn']
```

In [6]:
```
data
```

Out[6]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DevicePro
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	
...	...	...	...	...	...	...	...	...	...	...	...	
7038	6840-RESVB	Male	0	Yes	Yes	24	Yes	Yes	DSL	Yes	...	
7039	2234-XADUH	Female	0	Yes	Yes	72	Yes	Yes	Fiber optic	No	...	
7040	4801-JZAZL	Female	0	Yes	Yes	11	No	No phone service	DSL	Yes	...	
7041	8361-LTMKD	Male	1	Yes	No	4	Yes	Yes	Fiber optic	No	...	
7042	3186-AJIEK	Male	0	No	No	66	Yes	No	Fiber optic	Yes	...	

7043 rows × 21 columns

In [7]:
```
data.shape
```

Out[7]:
```
(7043, 21)
```

In [8]:
```
data1=data.drop([ 'SeniorCitizen','Partner', 'Dependents','tenure','StreamingTV','StreamingMovies', 'Paperle
```

In [9]:
```
data1
```

Out[9]:

	gender	PhoneService	MultipleLines	InternetService	TechSupport	Contract	MonthlyCharges	TotalCharges	Churn
0	Female	No	No phone service	DSL	No	Month-to-month	29.85	29.85	No
1	Male	Yes	No	DSL	No	One year	56.95	1889.50	No
2	Male	Yes	No	DSL	No	Month-to-month	53.85	108.15	Yes
3	Male	No	No phone service	DSL	Yes	One year	42.30	1840.75	No
4	Female	Yes	No	Fiber optic	No	Month-to-month	70.70	151.65	Yes
...	...	...	...	...	...	...	...	...	...
7038	Male	Yes	Yes	DSL	Yes	One year	84.80	1990.50	No
7039	Female	Yes	Yes	Fiber optic	No	One year	103.20	7362.90	No
7040	Female	No	No phone service	DSL	No	Month-to-month	29.60	346.45	No
7041	Male	Yes	Yes	Fiber optic	No	Month-to-month	74.40	306.60	Yes
7042	Male	Yes	No	Fiber optic	Yes	Two year	105.65	6844.50	No

7043 rows × 9 columns

In [17]:
```
data2=data1.fillna(data1.median())
```

```
/tmp/ipykernel_5669/3414091449.py:1: FutureWarning: The default value of numeric_only in DataFrame.median i
s deprecated. In a future version, it will default to False. In addition, specifying 'numeric_only=None' is
deprecated. Select only valid columns or specify the value of numeric_only to silence this warning.
  data2=data1.fillna(data1.median())
```

In [18]:
```
data2.isna().sum()
```

Out[18]:
```
gender             0
PhoneService       0
MultipleLines      0
InternetService    0
TechSupport        0
Contract           0
MonthlyCharges     0
TotalCharges       0
Churn              0
dtype: int64
```

In [21]:
```
data2['Churn']=data2['Churn'].map({'Yes':1,'No':0})
```

In [22]:
```
data3=pd.get_dummies(data2)
```

In [23]:
```
data3
```

Out[23]:

	MonthlyCharges	TotalCharges	Churn	gender_Female	gender_Male	PhoneService_No	PhoneService_Yes	MultipleLines_No	MultipleLines_N phone servi
0	29.85	29.85	0	1	0	1	0	0	
1	56.95	1889.50	0	0	1	0	1	1	
2	53.85	108.15	1	0	1	0	1	1	
3	42.30	1840.75	0	0	1	1	0	0	
4	70.70	151.65	1	1	0	0	1	1	
...	...	...	...	...	...	...	...	...	
7038	84.80	1990.50	0	0	1	0	1	0	
7039	103.20	7362.90	0	1	0	0	1	0	
7040	29.60	346.45	0	1	0	1	0	0	
7041	74.40	306.60	1	0	1	0	1	0	
7042	105.65	6844.50	0	0	1	0	1	1	

7043 rows × 19 columns

In [24]:
```
y=data3['Churn']
x=data3.drop('Churn',axis=1)
```

In [25]:
```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.33,random_state=42)
```

from sklearn.linear_model import LogisticRegression reg=LogisticRegression() reg.fit(x_train,y_train)

In [27]:
```
y_pred=reg.predict(x_test)
```

In [28]:
```
y_pred
```

Out[28]:
```
array([0, 0, 0, ..., 1, 1, 0])
```

In [29]:
```
from sklearn.metrics import confusion_matrix
confusion_matrix(y_test,y_pred)from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

Out[29]:
```
array([[1517,  180],
       [ 299,  329]])
```

In [31]:
```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

Out[31]:
```
0.7939784946236559
```

In [ ]:
```

```