Analysis of Data of Telecom Company

Atharv Dusane

200356

Given is a dataset on whether customers of a telecom service provider will churn or not. We are going to perform basic data understanding and try to make some interpretations.

Data set info:

The type of data set we have is a relational record data.

The number of data objects and attributes are (7043, 21) respectively.

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

Now we will check the data type of each attribute.

```
customerID          object
gender              object
SeniorCitizen        int64
Partner             object
Dependents          object
tenure               int64
PhoneService        object
MultipleLines       object
InternetService     object
OnlineSecurity      object
OnlineBackup        object
DeviceProtection    object
TechSupport         object
StreamingTV         object
StreamingMovies     object
Contract            object
PaperlessBilling    object
PaymentMethod       object
MonthlyCharges     float64
TotalCharges        object
Churn               object
```

Here we can see that SeniorCitizens is of numeric data type but should be Nominal/Categorical. So, we convert it to that data type (object data type).

TotalCharges is of object data type here, so we need to convert it to numeric type (float64).

We find that 11 entries in TotalCharges are missing. The number of entries missing is less compared to the total data set so we can safely drop out these entries.

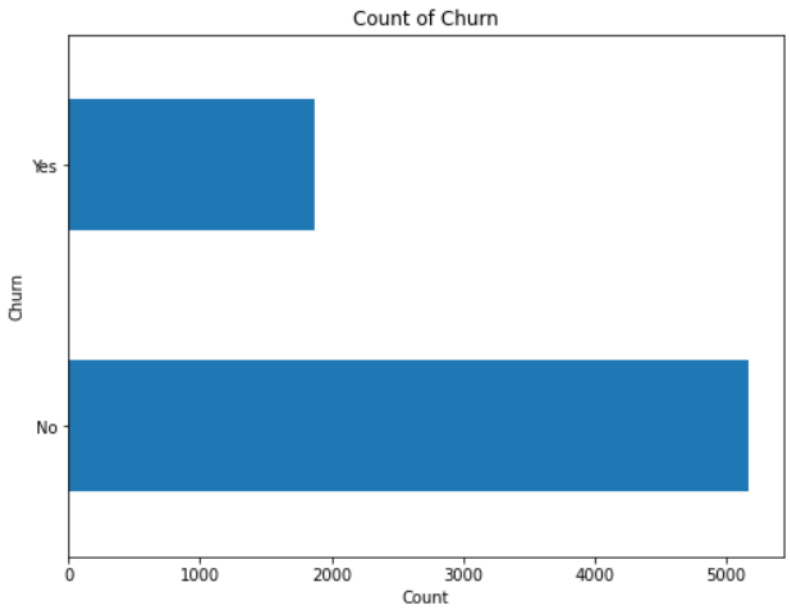
Statistics:

Now we will see the statistics of our data to get some basic insights.

	tenure	MonthlyCharges	TotalCharges
count	7032.000000	7032.000000	7032.000000
mean	32.421786	64.798208	2283.300441
std	24.545260	30.085974	2266.771362
min	1.000000	18.250000	18.800000
25%	9.000000	35.587500	401.450000
50%	29.000000	70.350000	1397.475000
75%	55.000000	89.862500	3794.737500
max	72.000000	118.750000	8684.800000

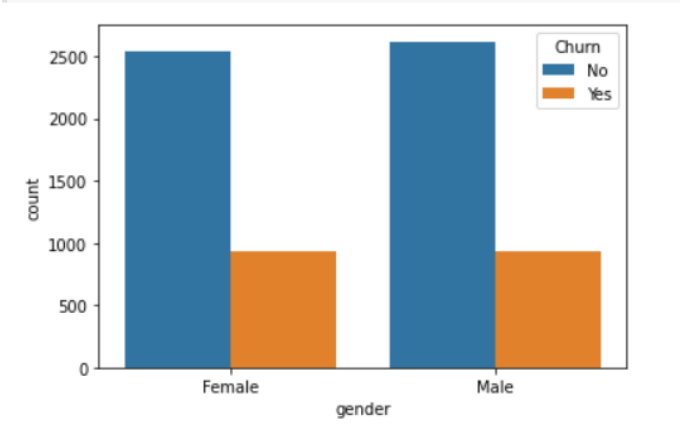
We can see that average monthly charge is 64.79, 50% people pay less than 70.35

Count of churn:

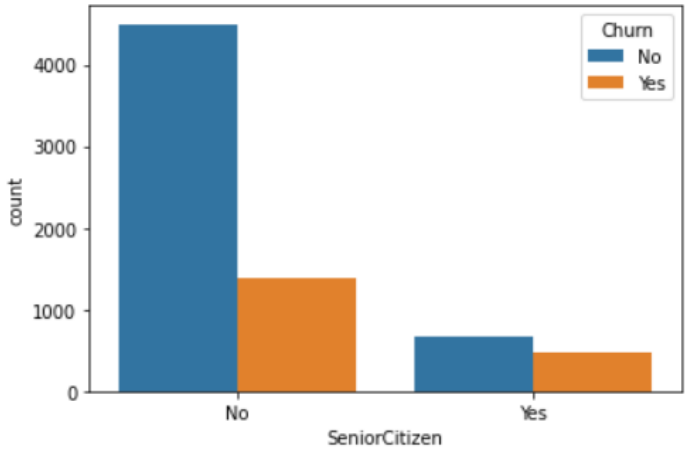


We can see that around 73.5% of people are not going to churn while 26.5% will churn.

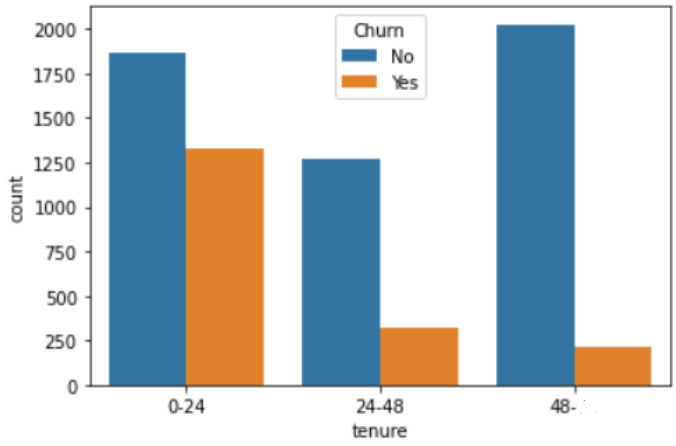
Count of churn depending on the attributes:



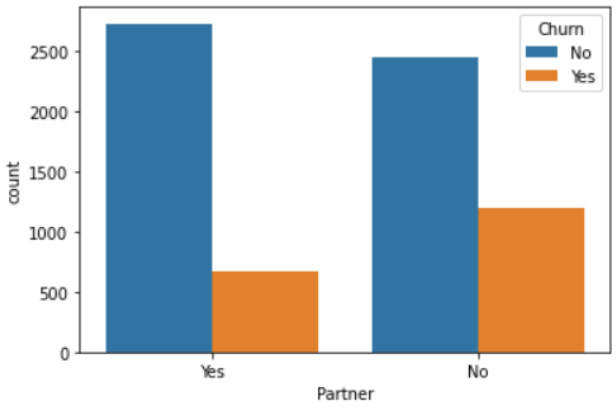
We can see that churn doesn't depend much on gender.

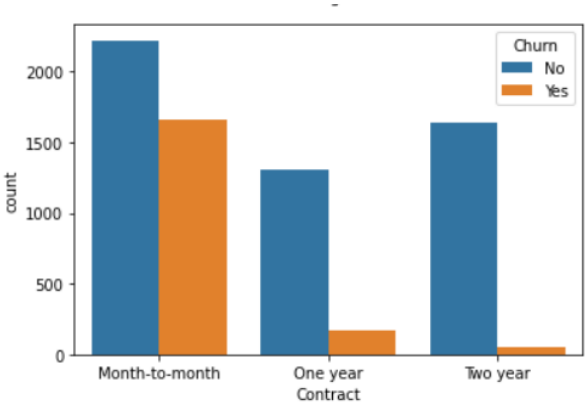


We can see that if you are a senior citizen then you are more likely to churn.

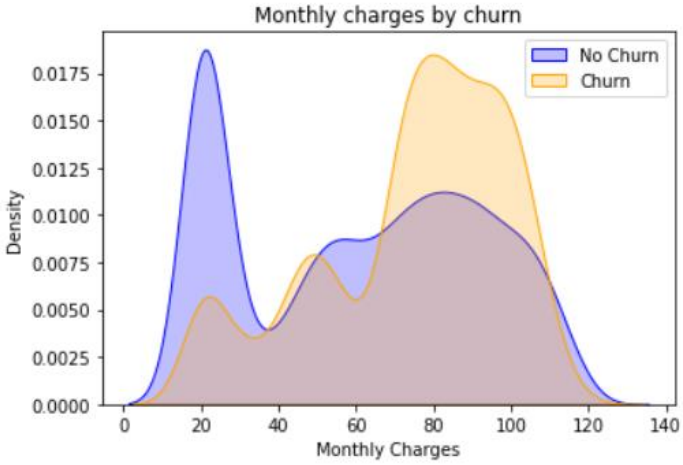


We have divided tenure into 3 groups: 0-24, 24-48, 48-72 months. We can see that customers with tenure of 48-72 are not likely to churn.

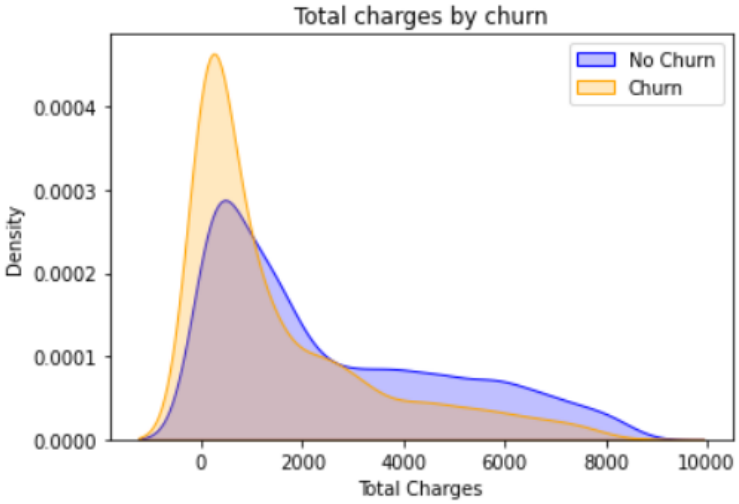




(All the other graphs are plotted in the code)



Here in this density function we can see that churn is high when monthly charges are high.



Here we can see that the customers that have paid higher amount of total money to company are less likely to churn.