

Customer Churn Analysis

Submitted by:

Vivek Singh

Batch NO:1827

**ACKNOWLEDGMENT**

Through this project I would like to Acknowledge my view towards The Data Trained Institute for their continuous support and Guidance which Give me a Confidence to complete the Assigned Project.

Through this I would like to thanks each and every Members of Data Trained Institute in completion of the project.

**INTRODUCTION**

(Customer Churn Analysis)

It mainly Occurs when a customer or a any subscriber stop doing business with company or any service. This is also known as Customer Attrition.

After losing a customer it might be very difficult for a service and or any company to deal with new customers.

This might not be easy to deal with new costumers because it will include in learning the business and all the steps again to bring everything on right path as earlier to run the Business.

Its very important to stop the Customers Attrition. As we can see nowadays in real-world lots of business services are rapidly growing up due to high Competition.

For a Customer it’s very easy to switch the services with low barriers.

**Data Description**: here we are examining customer data from IBM Sample Data Sets with the aim of building and comparing several customer churn prediction models. With the given link

<https://github.com/dsrscientist/DSData/blob/master/Telecom_customer_churn.csv>

* Importing the required libraries:

import pandas as pd

import numpy as np

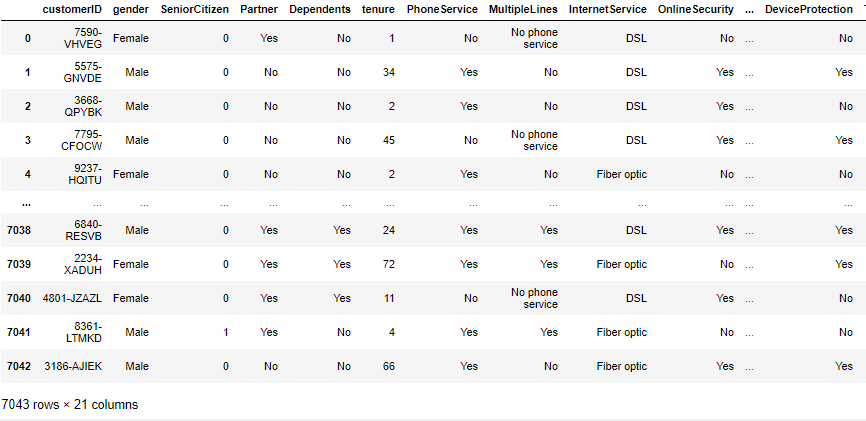
import seaborn as sns

import matplotlib.pyplot as plt

* Loading the data set:

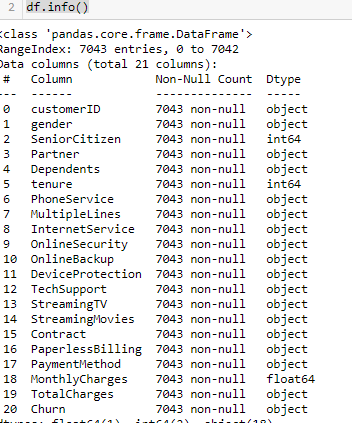
df=pd.read\_csv('Evaluation Project - 9Customer Churn Analysis')

df



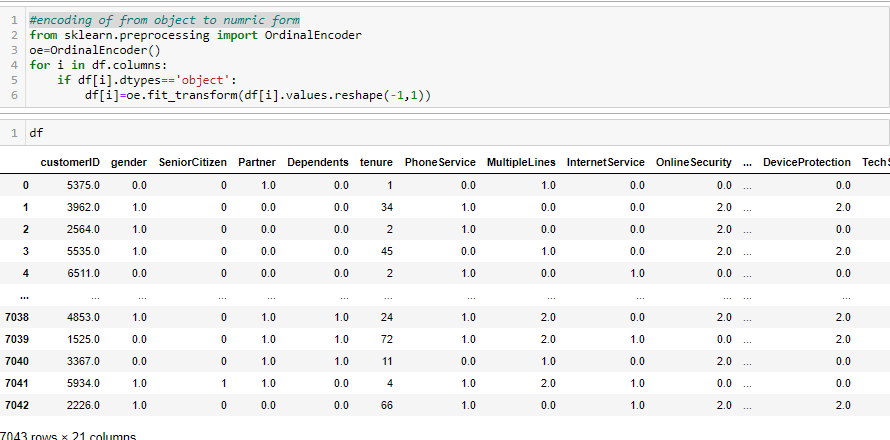


* we can clearly see our data set is consist of 7043 rows × 21 columns
* columns:'customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents','tenure', 'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport','StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn'
* **next we will check data details**

 from above we can clearly see the and its information either it is having null values or not and checking the data types its Object ,float or Int type data

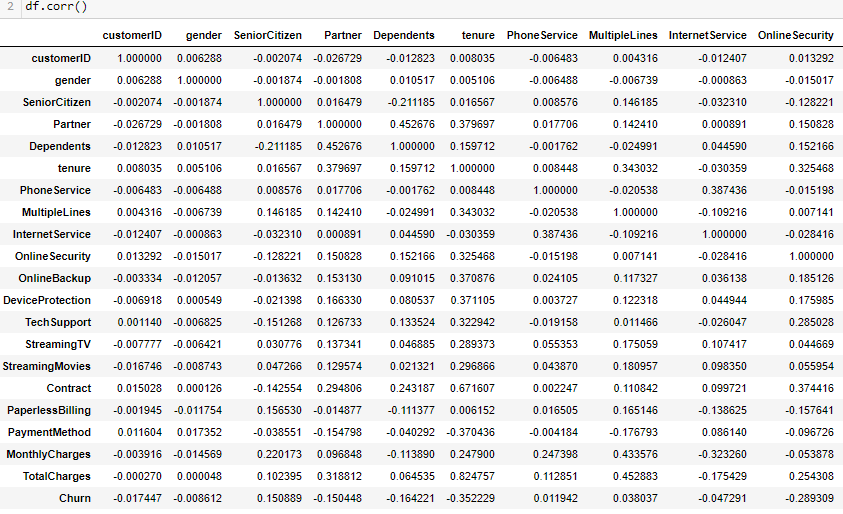
* **Data Encoding**

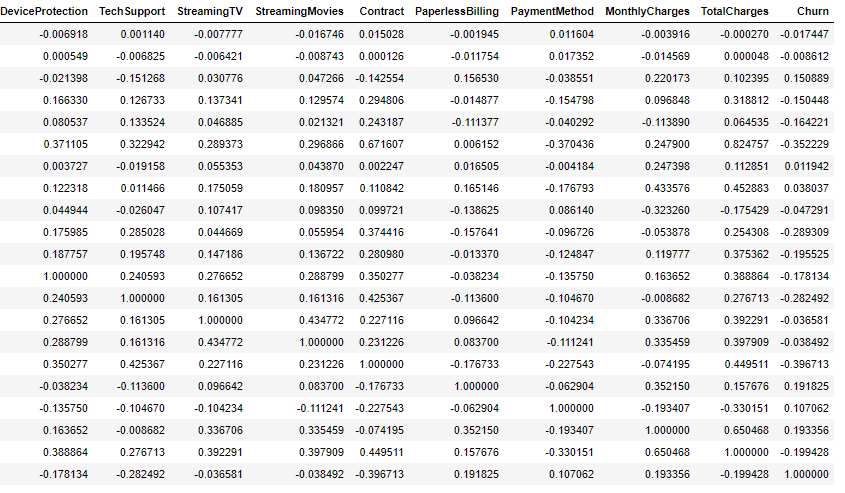
In this method we will change the our object data into numeric form



**OBSERVATION:** From we can clearly see that our object columns are changed into numeric form.

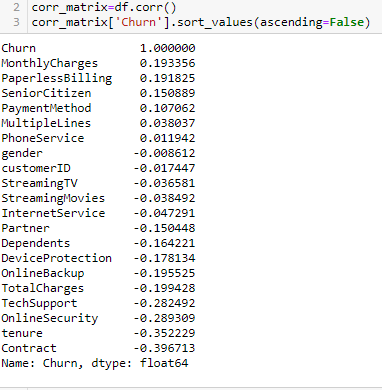
* **Checking the Correlation of all the data :**

****

****

**Observation:**

* from above we can clearly see how the data correlated with each other some of positively correlated and some are negatively correlated.
* We can also find out the correlation of every columns with others remaining column via using correlation matrix

****

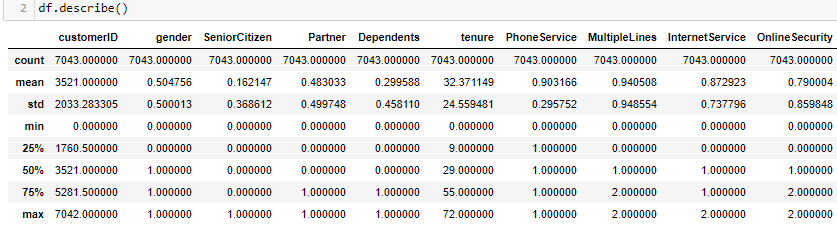
* From above we can clearly see the correlation of the column ‘Churn’ with the remaining columns
* ‘Churn’ is having highly correlated with Monthly Charges

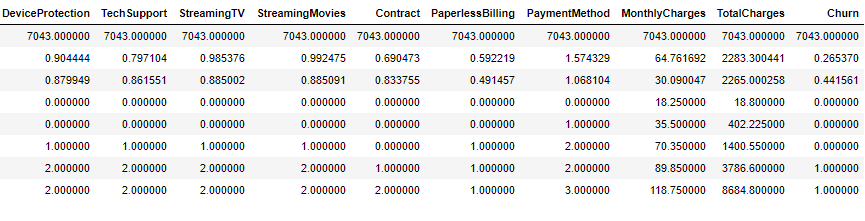
with value -0.193356

* ‘Churn’ Least correlated with Contract with value-0.396713

* **Statistical Summary:**

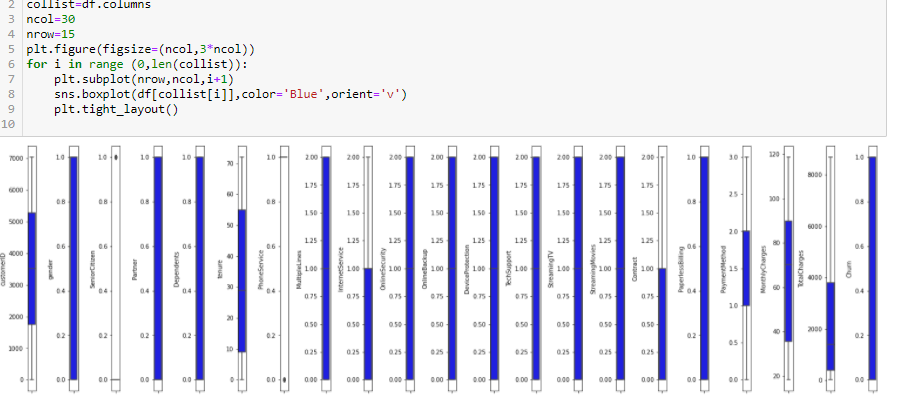
in this method we well see the statistical summary of each columns of our data set



****

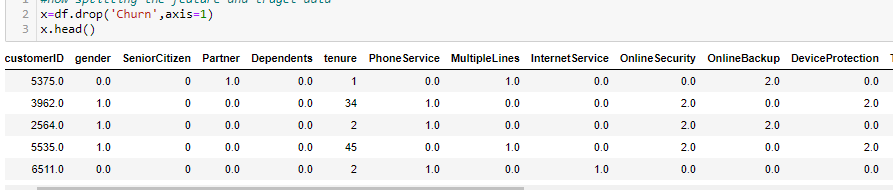
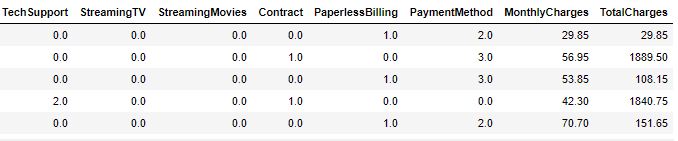
* From above we can easily analyse the out lairs skewness, mean, median, Standard deviation of our data set.

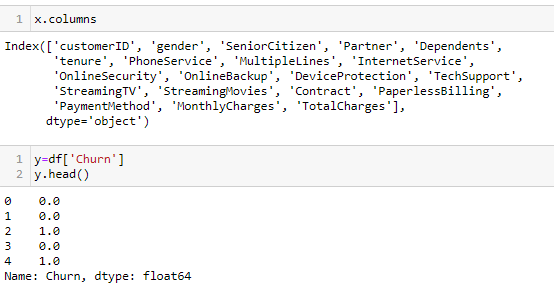
**Out lairs Detection :**



From above figure we can clear see there are no outliars in our data set except columns  {"senioir Citizen",phones service} which is object type.column

* **Features and Target data Split :**
* In the the case of dta split we will sepreate the feature and target data.
* We can mainly do splitting by using locking , method or column dropping.

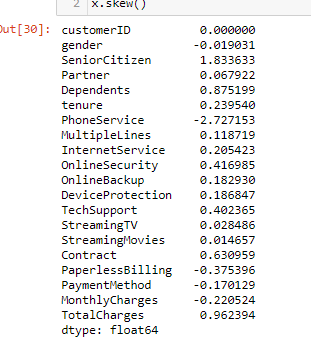
 

****

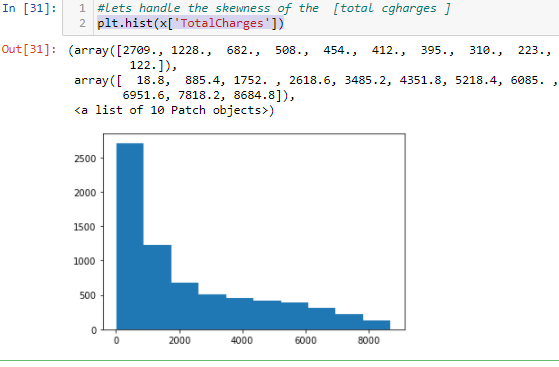
**Observation:**

Here for feature data selection we dropped the target columns. And we can easily split our feature and target data which his denoted by x and y.

* **Skewness:** before training and testing our data set we need to check the skewness of our data set.



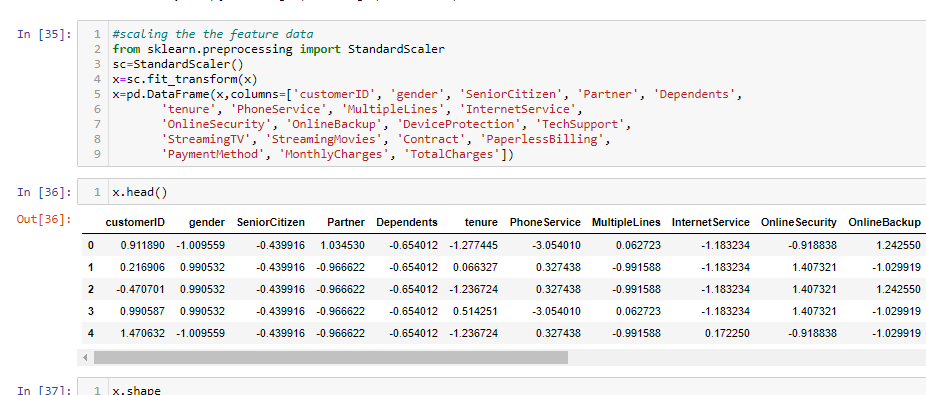
* Here we will check the skewness value of the our x data set i.e(±0.5)
* Here we can clearly see most the values of numerical data type lies in range i.e(±0.5)

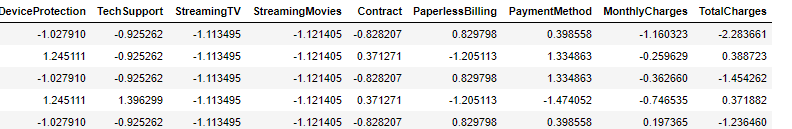


We can clearly see this right side skewed data. We can handle the skewness by log transformation method.

* **Scaling:**

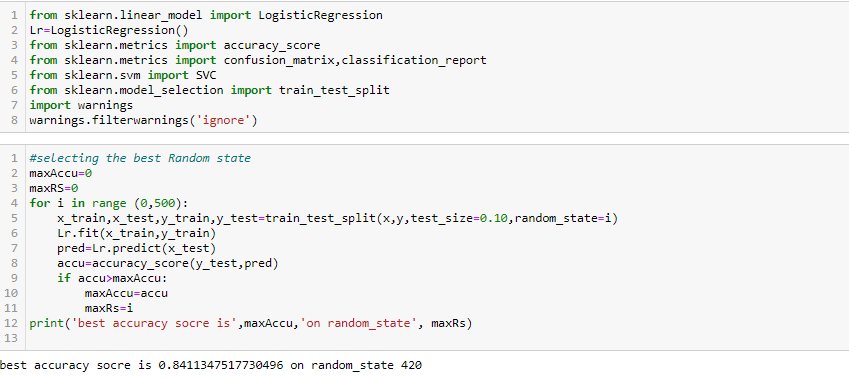
In this case before training we have to scale the our data in same unit .





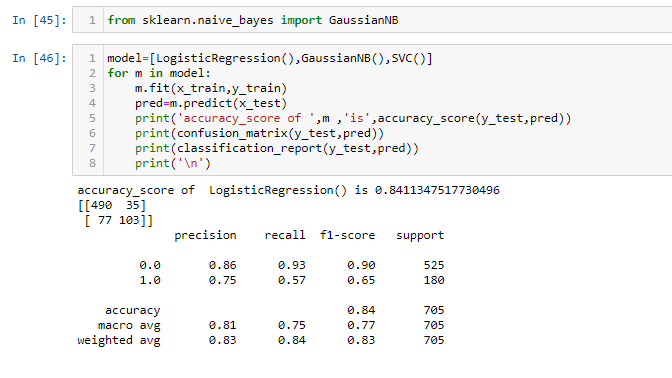
* From above we can Cleary see how we have done scaling in our data set before training.
* Before training we must have to check both our x and y data must be into 2D from. we change data into 2d form by using **reshape method** .
* After all the process we have to load the model. For training purpose of the data. As we can see our target data is into Classification form that is Binary i.e,yes or no . we will train it on the Logistic regression

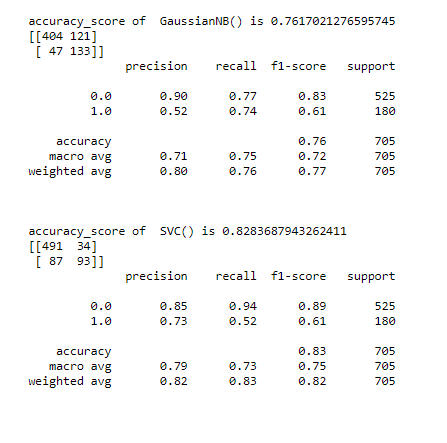
**Best random state selection:**

****

* From we can clearly see best random state is 420 because accuracy score is higher at this point

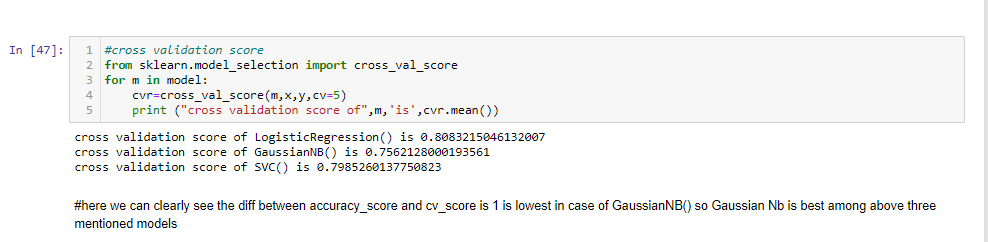
Now we will train our data in other classification model and check its accuracy score in each.





* We have done training on Logistic Regression, GaussianNB ,SVC
* Accuracy score and classification reports every thing is mentioned in above figure.

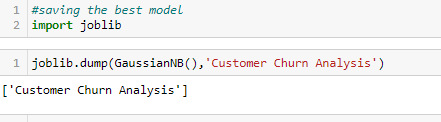
**Cross validation score:**



* Accuracy score before may be high or low due to underfitting and overfitting.
* Best model is always decided by after computing the difference between the accuracy score and Cross validation score which is also known as cv score
* here we can clearly see the diff between accuracy score and cv\_score is 1 is lowest in case of GaussianNB() so GaussianNb is best among above three mentioned models.
* **Saving the best model:**

After the training selection of the best model we have to save it for predicting the values .

We can save our model by using pickle **pickle or joblib**



From above we can clearly see joblib method the save the model.

We can recall it for prediction the values.

Thank you.