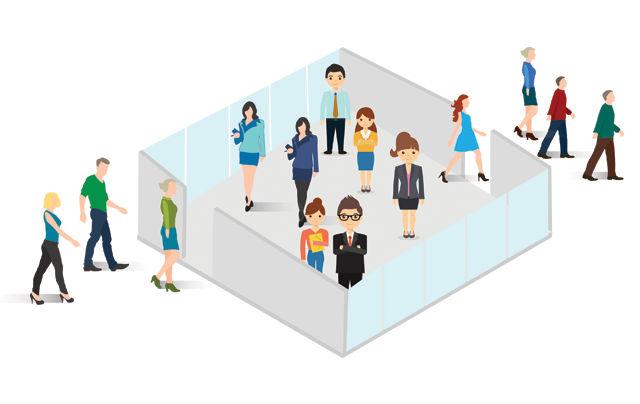
**Customer Churn Analysis Model**

**INTRODUCTION**

In this model, we are going TO predict churn of customer.

Used customer churn dataset which contains feature columns of customer and at the end, one column of target for which we will train model.

This is an classification column, as we have found target variable as categorical variable



**OVERVIEW OF DATASET:**

* This dataset is having 7043 rows and 21 columns.
* Dataset is having total 21 columns in which 18 columns are of object type, 1 column is of float type and 2 column are of int type

**PROBLEM STATEMENT:**

Customer churn is when a company’s customers stop doing business with that company. Business are very keen on measuring churn because keeping on existing customer is far less expensive then acquiring a new customers. Now business involves working leads through a sales funnel, using marketing and sales budgets to gain additional customers. Existing customers will often have a higher volume of service consumption and can generate additional customer referrals.

Customer retention can be achieved with good customer service and products. But the most effective way for a company to prevent attrition of customers is to truly know them. The vast volumes of data collected about customers can be used to build churn prediction models. Knowing who is most likely to defect means that a company can priorities focused marketing efforts on that subset of their customer base.

Preventing customer churn is critically important to the telecommunications sector, as the barriers to entry for switching services are so low.

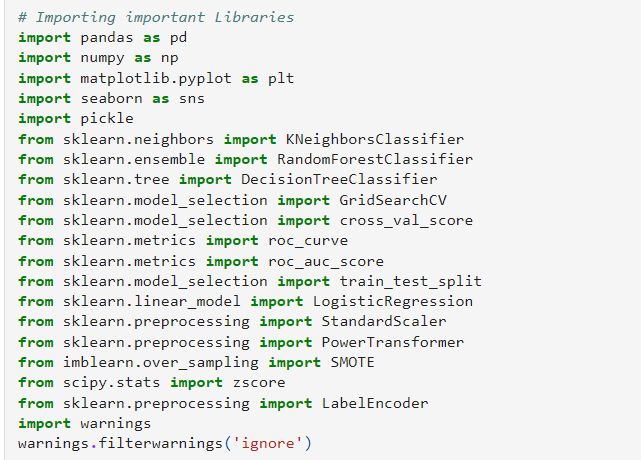
We will examine customer data from IBM Sample Data Sets with the aim of building and comparing several customer churn prediction models.

**FEATURES OF DATASET:**

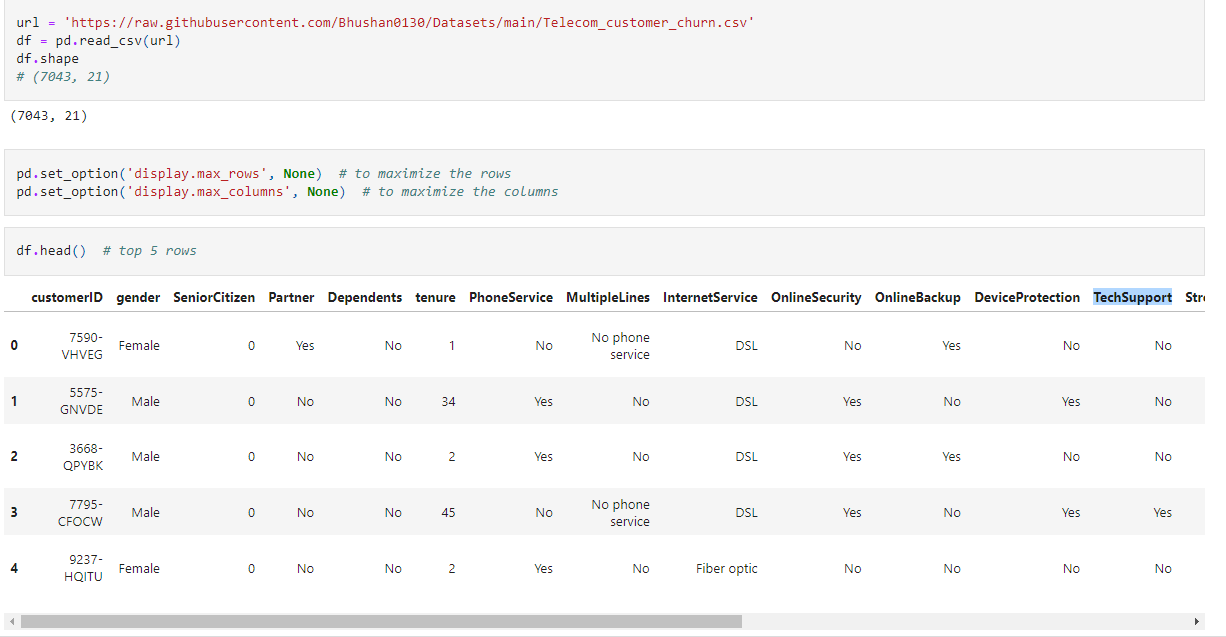
* **customerID**: Name of the customer
* **gender**: gender of customer
* **SeniorCitizen**: Senior Citizen yes or no
* **Partner**: Having partner Yes or No
* **Dependents**: Having dependent Yes or No.
* **tenure**: Tenure of customer
* **PhoneService**: Having Phone services Yes or No.
* **MultipleLines**: Having Multiple Lines service
* **InternetService**: Which internet services one is using.
* **OnlineSecurity**: Having online security Yes or No
* **OnlineBackup**: Having online Backup Yes or No
* **DeviceProtection:** Having device protection Yes or No
* **TechSupport:** Getting Technical Support or not
* **StreamingTV:** Gettingstreaming TV Yes or not
* **StreamingMovies:** Getting Streaming Movies Yes or not
* **Contract:** Type of contract
* **PaperlessBilling:** Getting paperless billing Yes or Not
* **PaymentMethod:** Mode of payment one is choosing
* **MonthlyCharges:** Monthly Charges one is paying
* **TotalCharges:** Total Charges one is paying
* **Churn : Customer Churn Analysis**

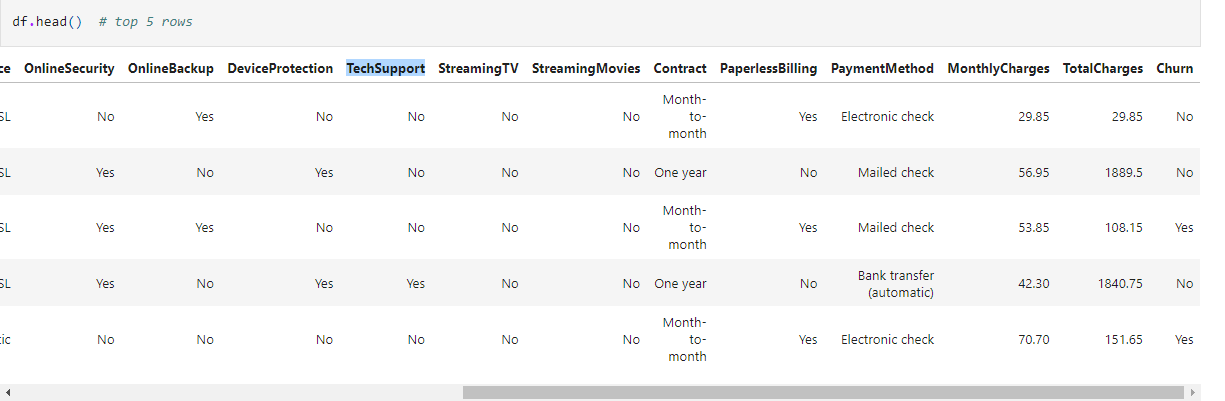
**MODEL BUILDIING:**

**Used Libraries:**

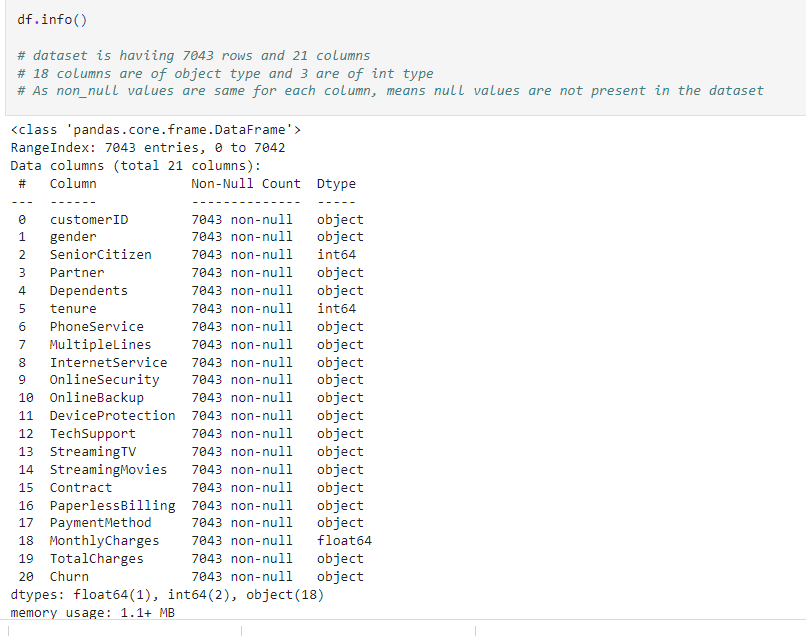
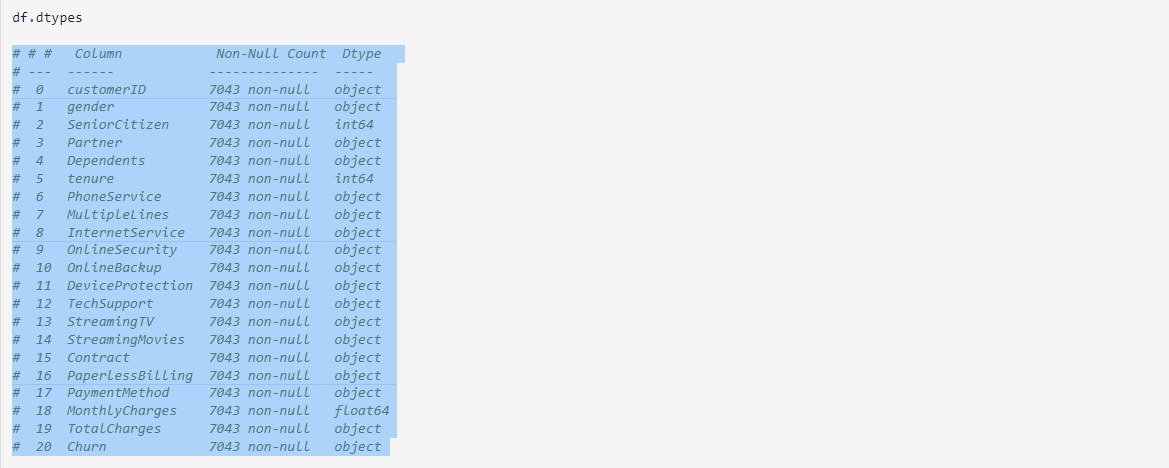


**LOADING DATASET**



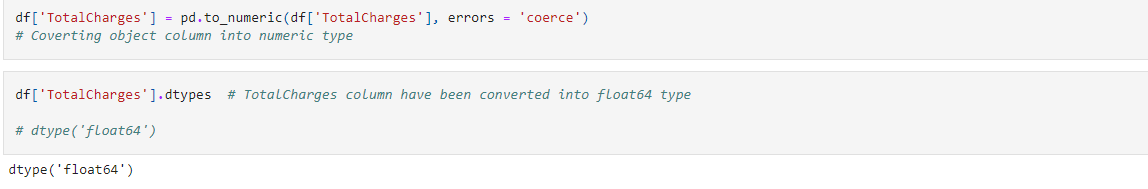


**INFORMATION ABOUT DATASET:**

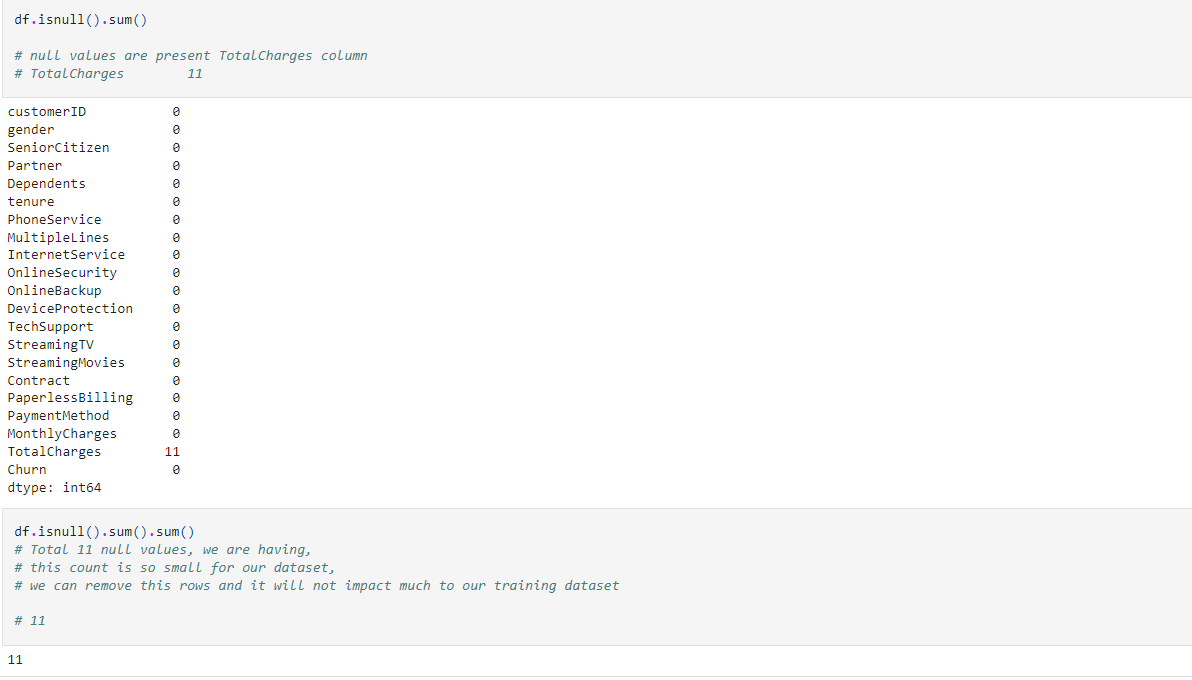
 

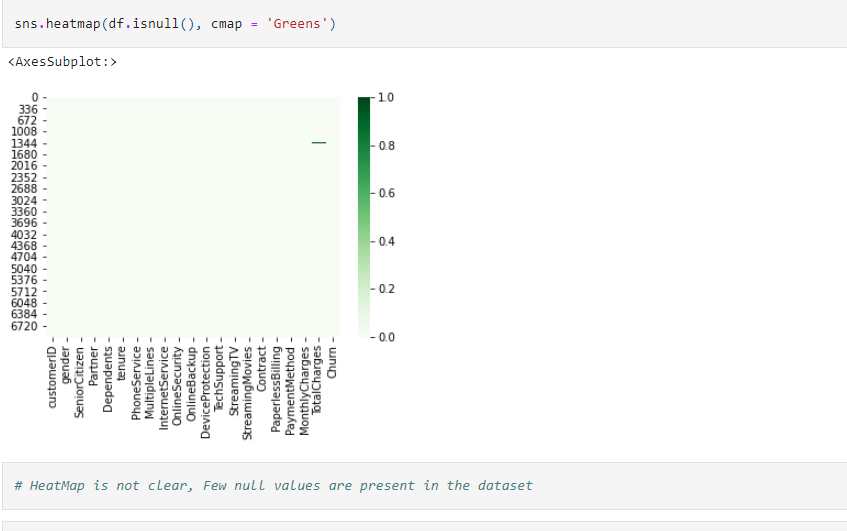
**PREPROCESSING PART:**

# ***Note****: Here one hack found that, Total Charges column containing value of int type but its column type is object, need to convert this column into numeric type*



***NULL VALUE CHECKING*:**

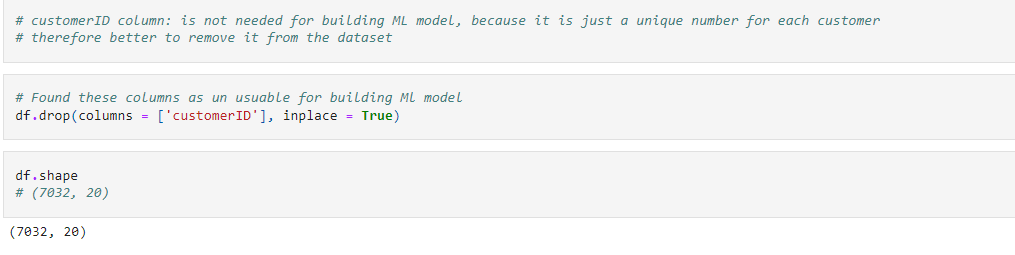




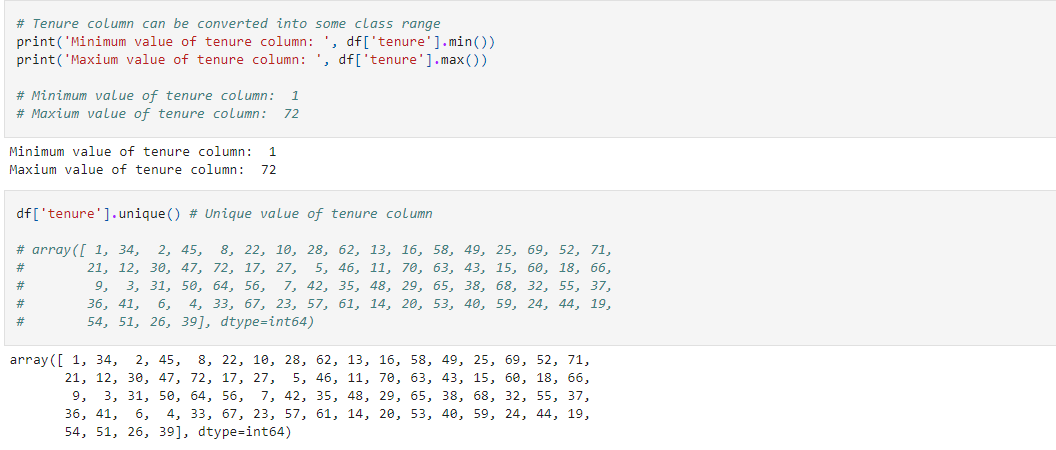


Few null values were found in the dataset and count of that null value was to small, therefore we have drop that rows instead performing mean, mode imputation

***DELETING UN-NECESSARY COLUMNS:***



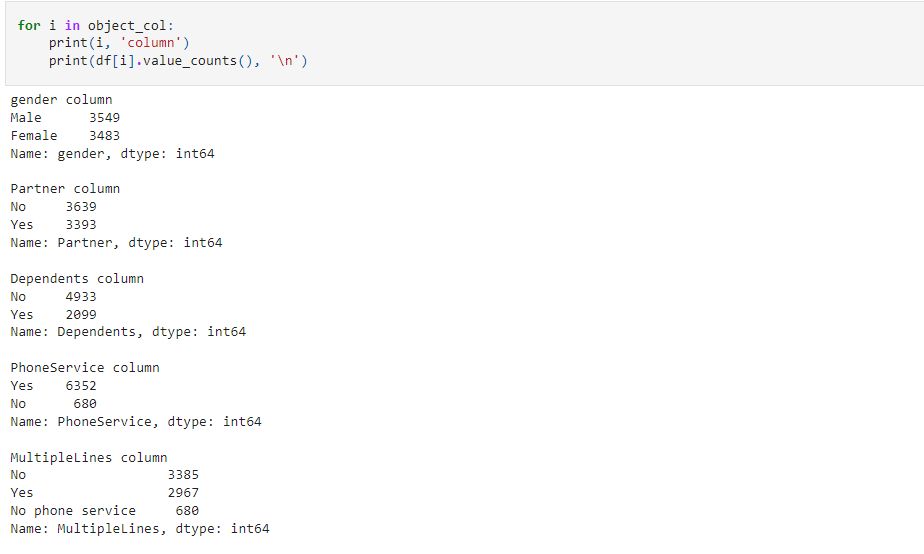
***CONVERTING TENURE COLUMN INTO CLASSES:***

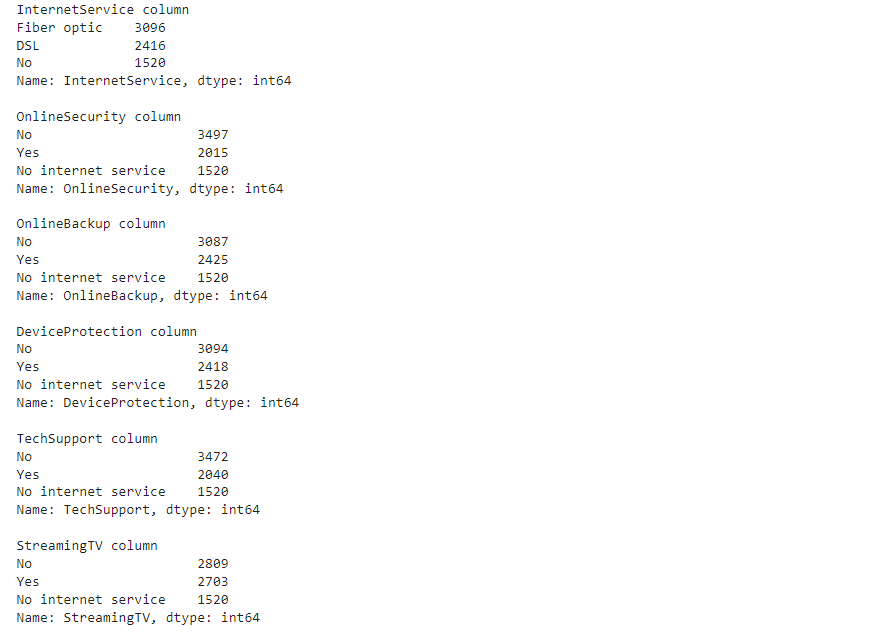


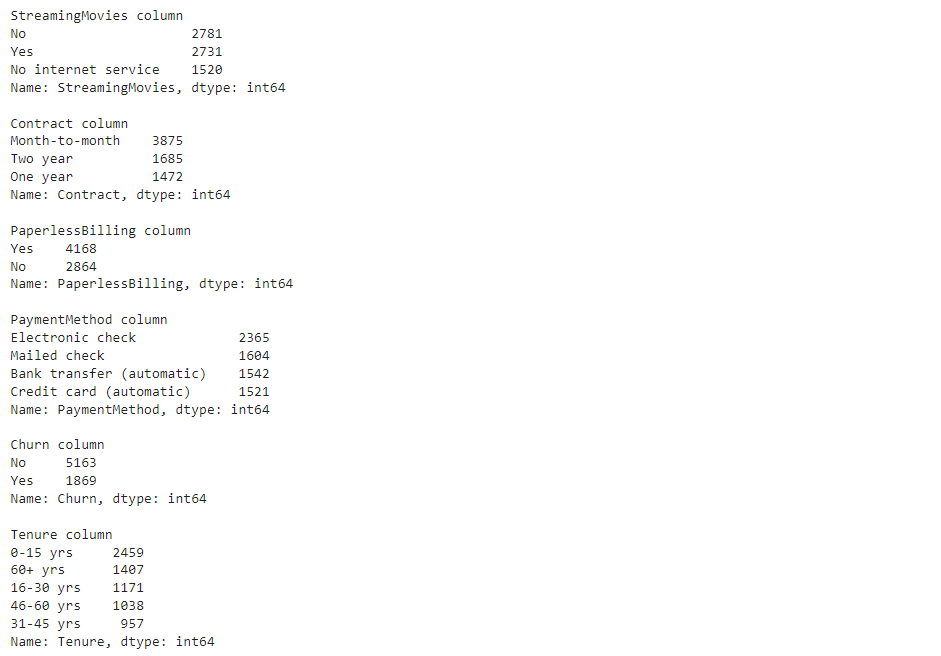


***VALUE COUNT FOR OBJECT TYPE COLUMNS:***





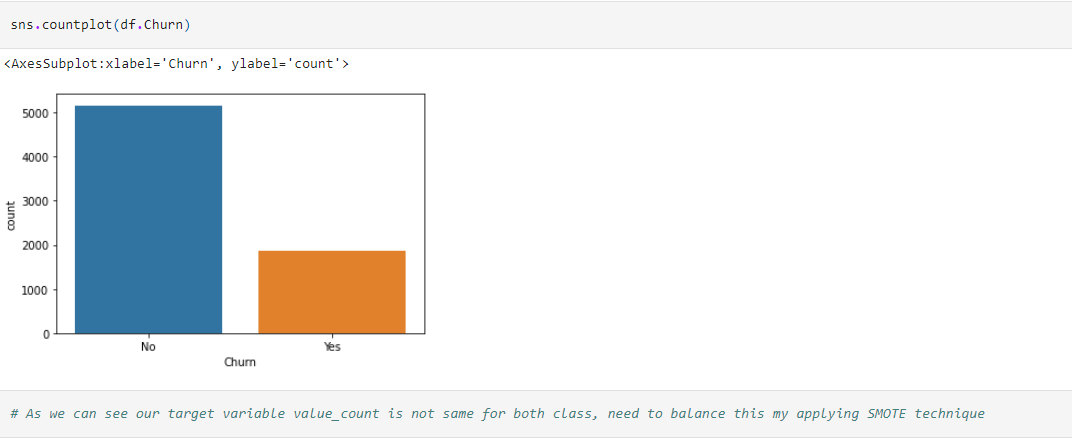




**VISUALIZATION OF CATEGORICAL COLUMNS:**

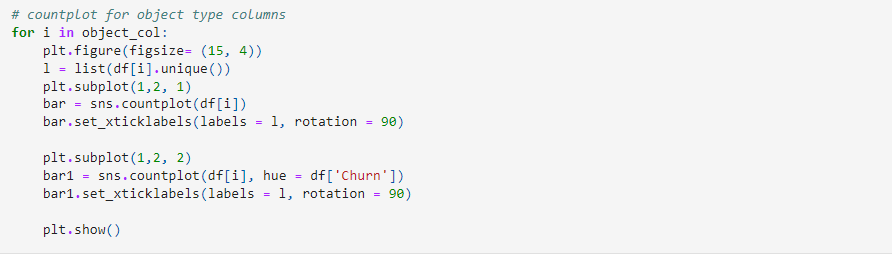


**COUNT PLOT FOR TARGET COLUMN:**



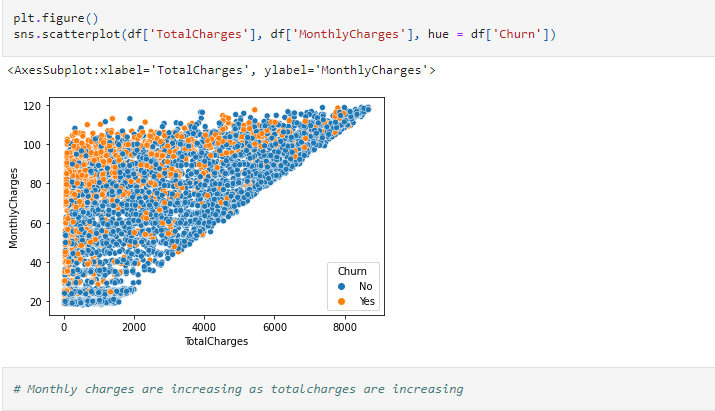
**COUNT PLOT FOR FEATURE COLUMNS**

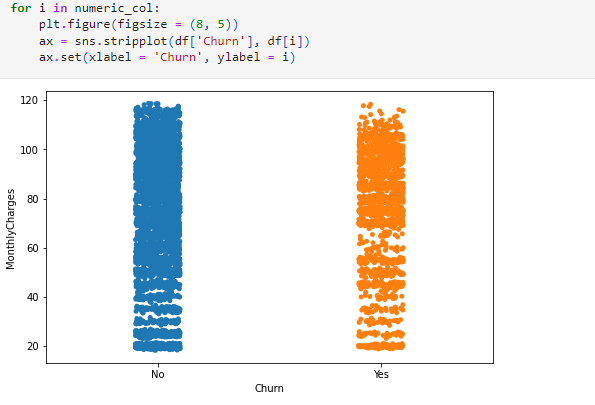
* Code for creating multiple count plots for categorical column of dataset

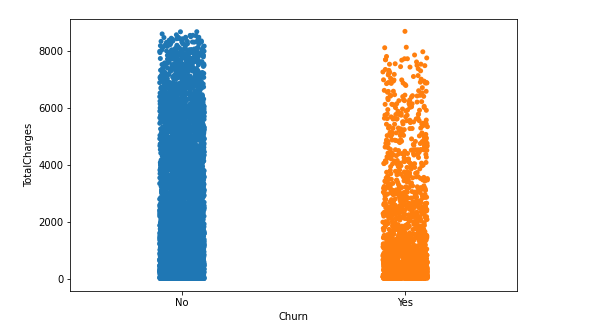


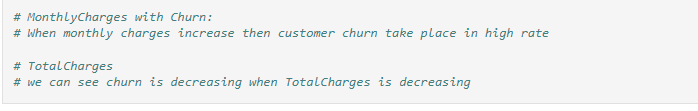
|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| ***KEY OBSERVATION FOR FEATURE COUNT PLOT:***  *#* ***Gender:***  *# for both gender, value seems almost same, therefore we can say, gender not impacting much to churn*  *#* ***OnlineBackup:***  *# Due online Backup issue many customer have churn*  *#* ***Device protection***  *# may be due to mobile protection, people churn the telecome, therefore company needs to work on it*  *#* ***Patner column:***  *# We can say if person is not having partner, then he is churn*  *#* ***Dependent column***  *# majority for population do not churn if he is not having dependents*  *#* ***Phone service***  *# Majority of population who churn, basically having phone services, need to improve phone service*  *#* ***MultipleLines column***  *# Very less people churn if person is not having phone service*  *#* ***Internet Service***  *# Telcom compnay need to work highly on fiber optice, by seeing it we can say, due fiber optice problem many person have churn*  *#* ***Online Security***  *# Majority of customer have churn due to internet security factor*  *#* ***TechSupport***  *# As we can see many of custerm have churn due to No tech support, need to work on it.*  *#* ***Streaming TV and Streamiing Movies:***  *# this is also leading for churn*  *#* ***Contract column:***  *# Majority of customer have churn, who are having month to month contract, need to find what problem they facing ,*  *# and how can be fixed that*  *#* ***Paper Less billing*** *is also leading for churn*  *#* ***ElectronicMethod*** *is main reason out of remaining, for churn the telecom*  *# Churn column, data is not balanced, need to balance it*  *#* ***Tenure column:***  *# customer having age more than 60 they are more loyal to the company as its churn rate is least as compare to other*  *# age group class*  *#* ***SeniorCitizen****:*  *# When seniorCitizen is 0 then its churn rate less as compare to 1*  **CODE FOR CREATING PIE CHART:**    ***PERCENTAGE OF VALUES OF COLUMN:***  ***For Target Variable:***    ***For feature categorical columns:***      ***CHECKING OUTLIERS:***      ***CHECKING SKEWNESS OF THE COLUMNS*** |  |  |
| **CHECKING DISTRIBUTION OF THE COLUMNS**      **SPREAD OF COLUMNS*:*** |  |  |
|  |  |  |

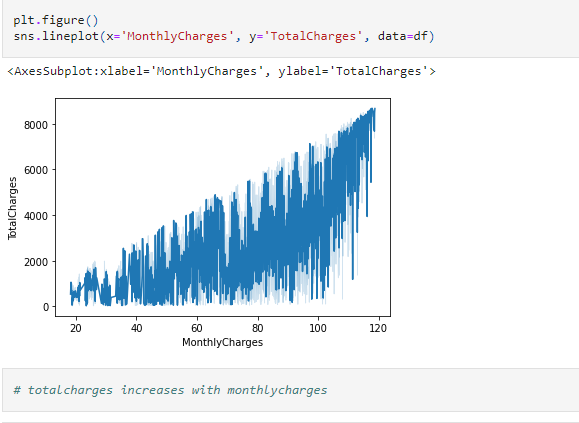
**KNOWING THE PATTERN OF DATA:**

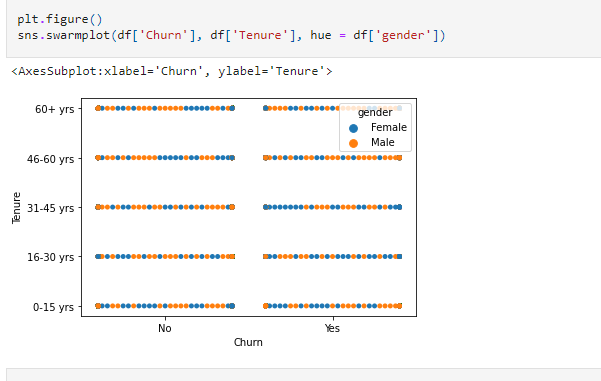




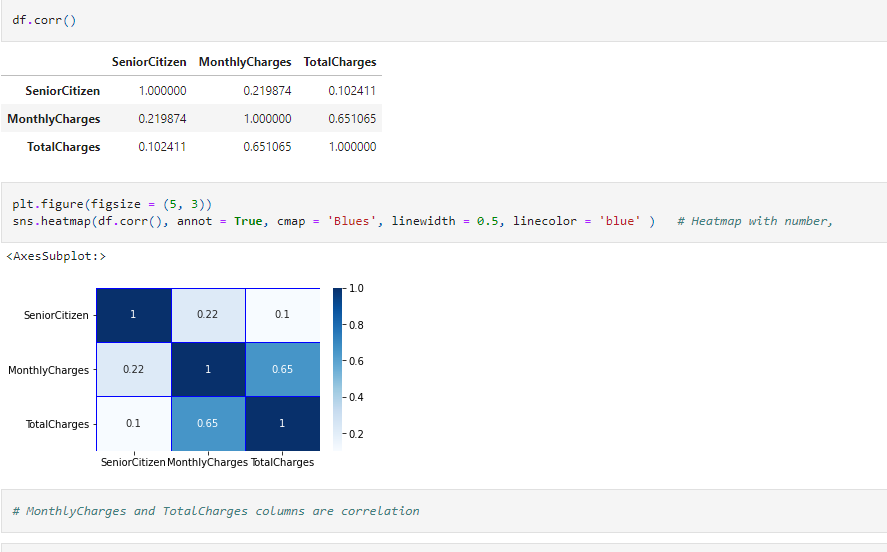




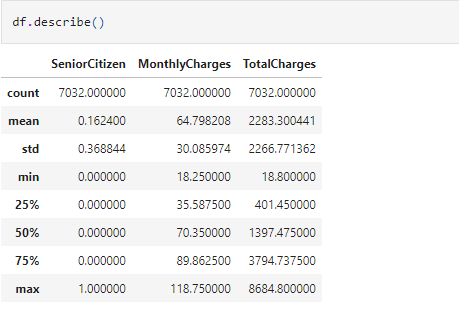




**CORRELATION OF COLUMNS:**



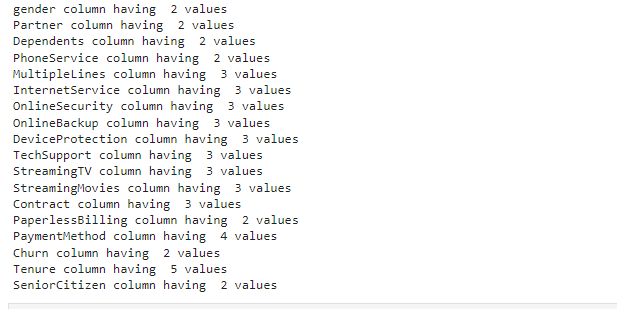
**DESCRIBE DATASET:**



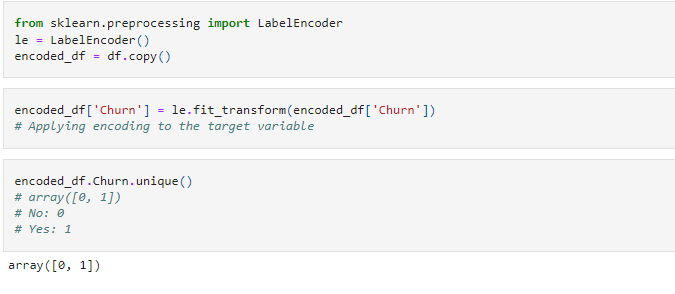


**ENCODING OPERATION:**

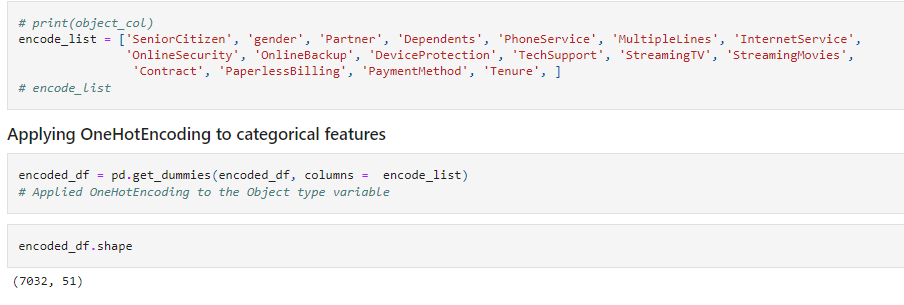




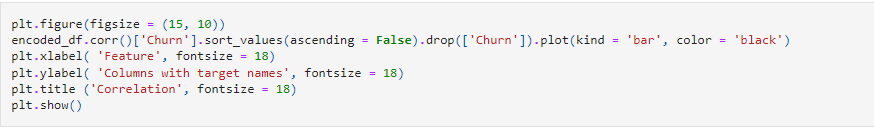
**LABEL ENCODING TO THE TARGET VARIABLE:**

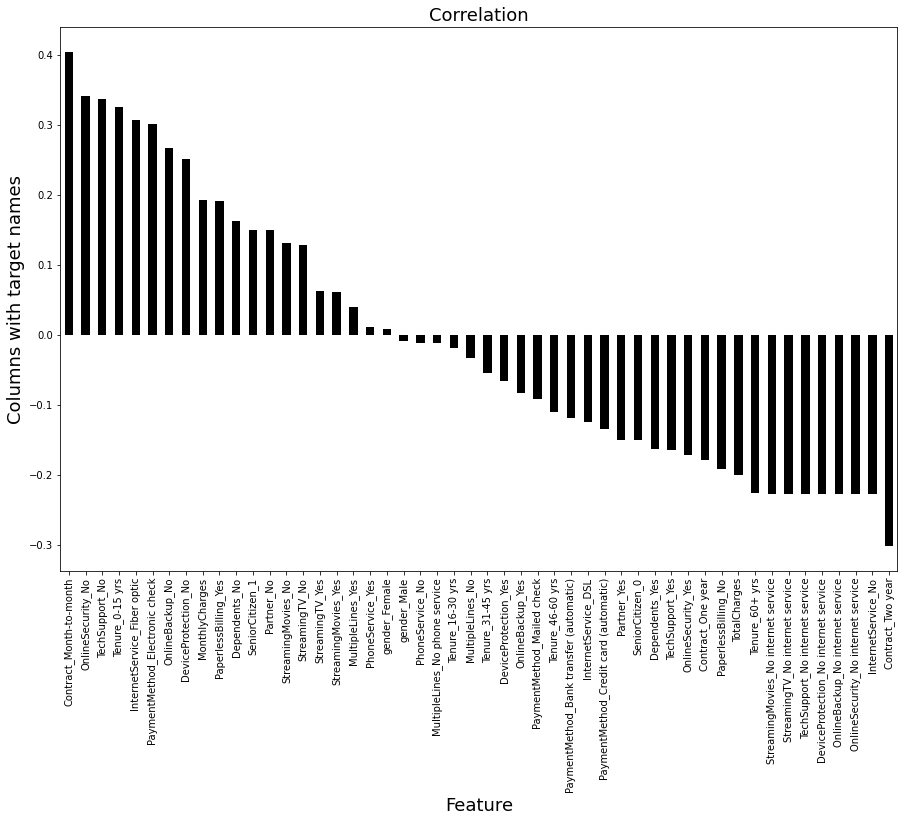


**ONE HOT ENCODING FOR FEATURE COLUMNS:**

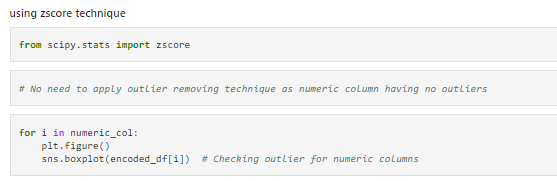


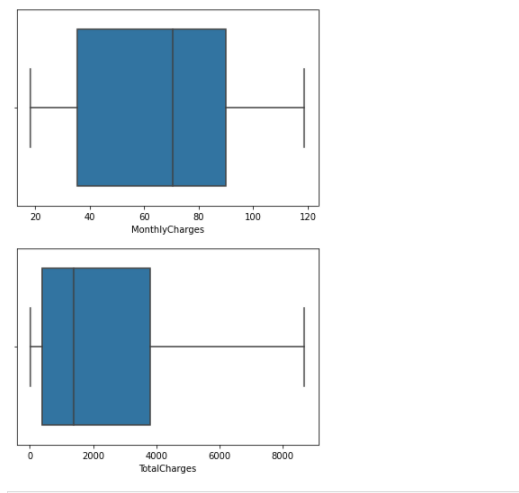
**IMPACT OF FEATURES ON TARGET:**





**REMOVINF OUTLIERS:**

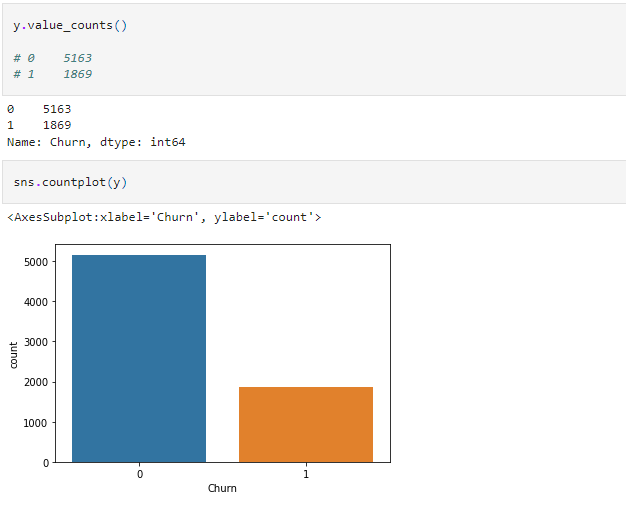




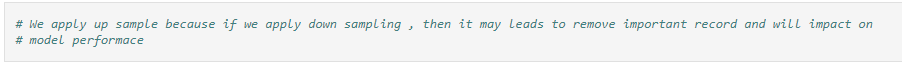
**SEOERATING DATASET INTO X AND Y FORM:**



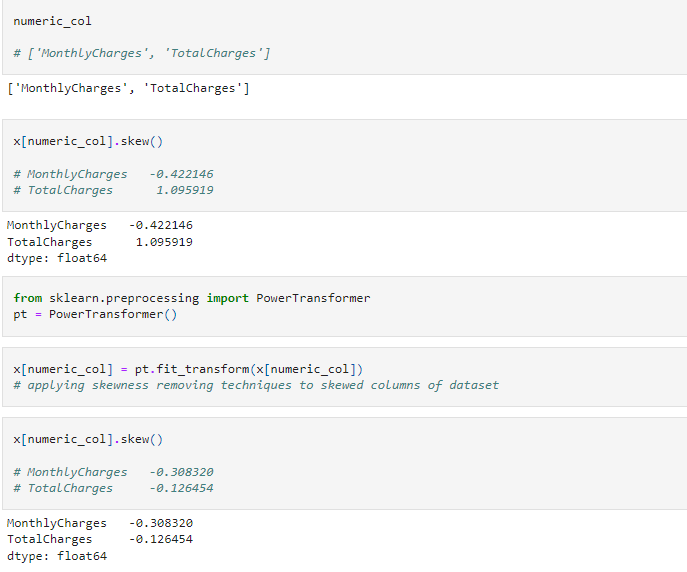
**SMOTE Technique to balance to dataset:**





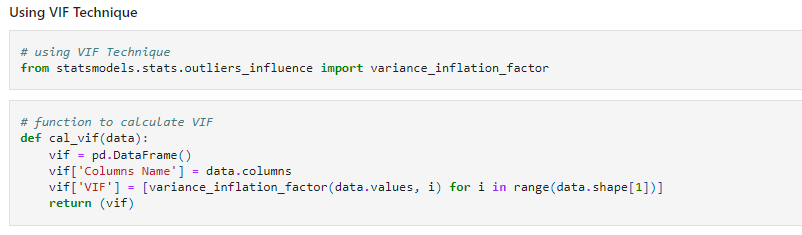


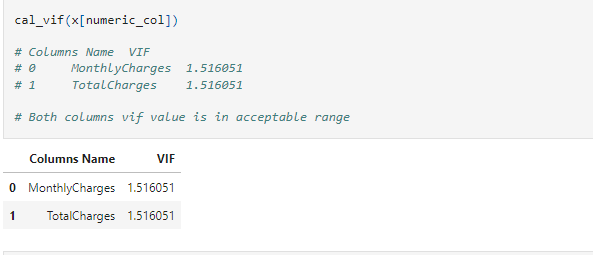
**REMOVE SKEWNESS:**



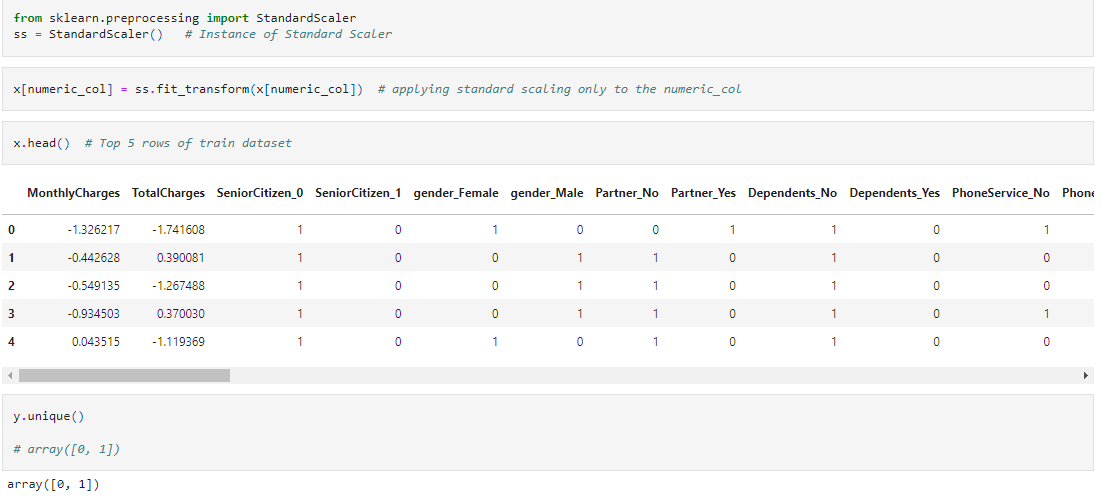


**REMOVING MULTICOLLINEARITY:**



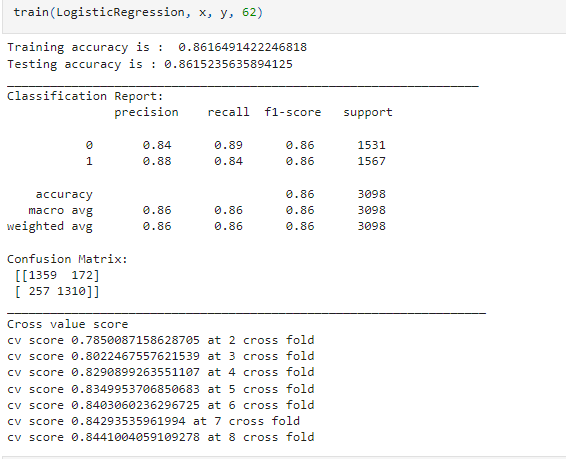


**STANDARD SCALING:**

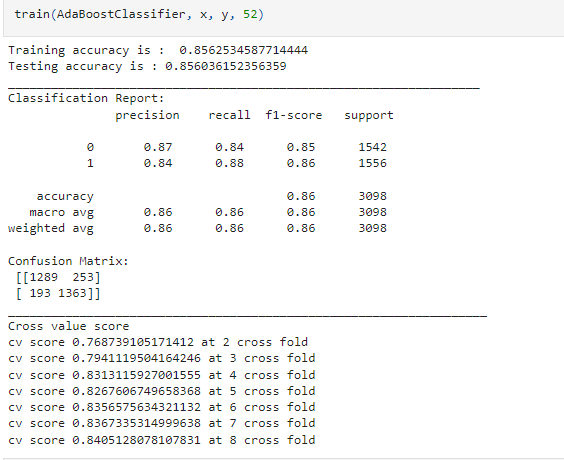


**TRAINING MODEL:**

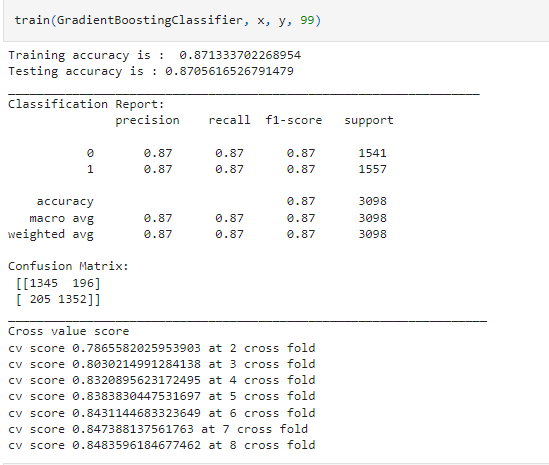
1. ***LogisticRegression:***



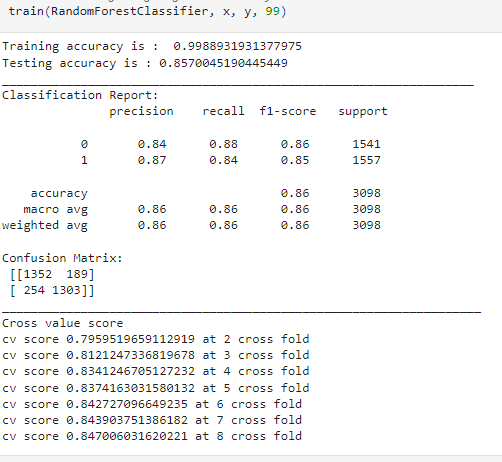
1. ***AdaBoostClassifier:***



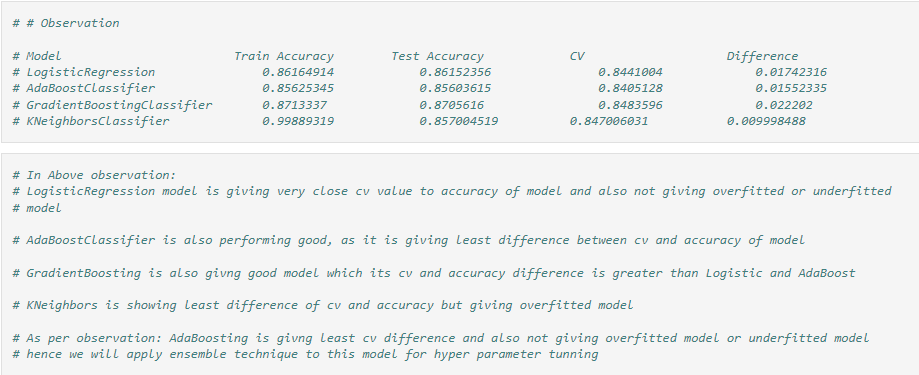
1. ***GradientBoostingClassifier:***



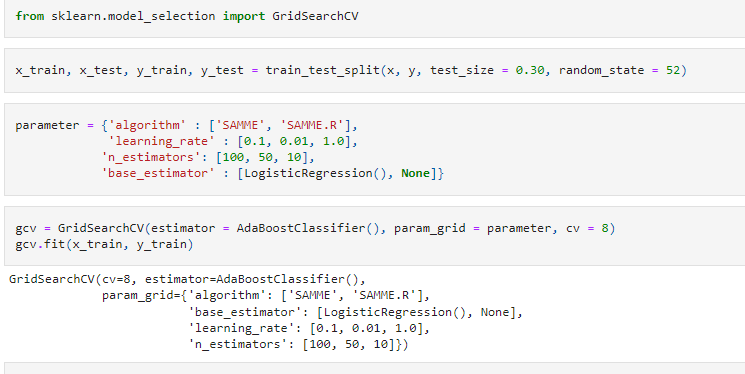
**4. *RandomForestClassifier*:**

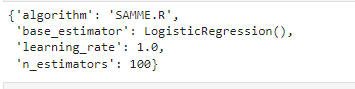


**OBSERVATION FOUND OF MACHINE LEARNING MODELS:**

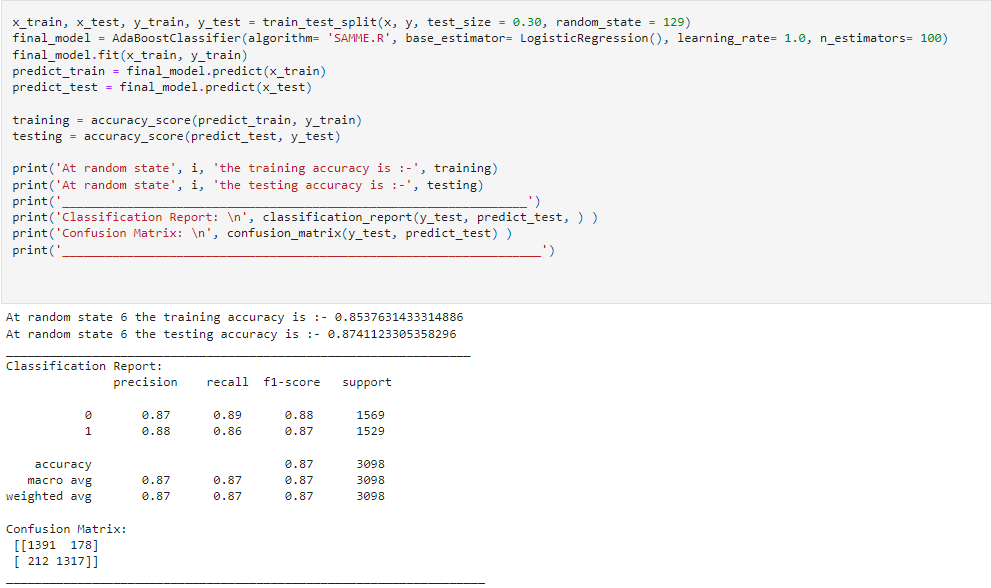


***Hyper Parameter Tuning for AdaBoostClassifier:***

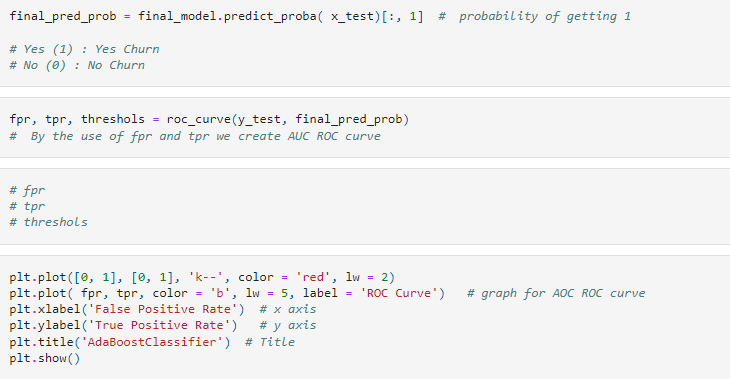


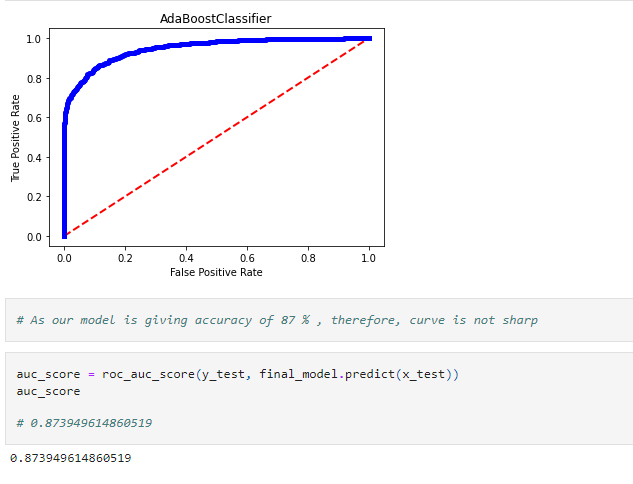


***Final Model of AdaBoostClassifier:***



***AOC – ROC Curve (Churn Status: Yes):***

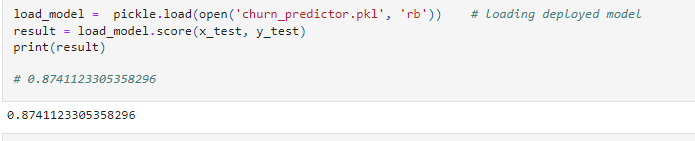




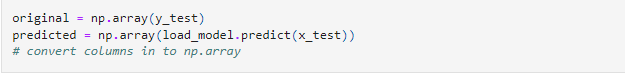
**DEPLOY MODEL:**

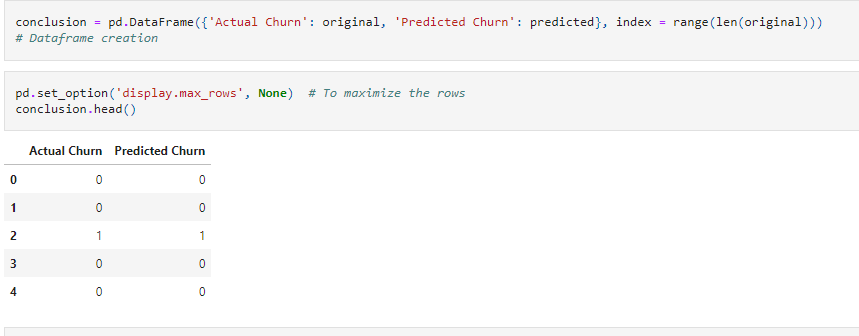


**LOADING MODEL:**



**CONCLUSION:**





In this way you can create your machine learning model for predicting churn of a customer,

By using this model one can growth his business, by focusing to overcome from the main reason of churn

Thank for giving time on this page, I hope it helped you!

**Submitted by:**

**Bhushan Kumar Sharma**

**(Batch 1834)**