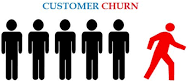
Customer Churn Analysis:

Data science Project



Presented by: Nilesh Kumar Singh

Topicscovered:

1. ProblemDefinition
2. DataAnalysis
3. EDAConcludingRemark
4. Pre-ProcessingPipeline
5. BuildingMachineLearningModels
6. ConcludingRemarks.

## WhatIsCustomerChurn?

Customer churn is the percentage of customers that stopped using company'sproductorserviceduringacertain timeframe.

## WhyIsCustomerChurnRateImportant?

An [increase in customer retention](http://www2.bain.com/Images/BB_Prescription_cutting_costs.pdf)of just 5% can create at least a 25% increasein profit. This is because returning customers will likely spend [67% more](http://www2.bain.com/Images/Value_online_customer_loyalty_you_capture.pdf)oncompany's products and services. As a result, company can spend less on theoperatingcosts ofhavingto acquirenewcustomers.

## HowtoCalculateChurn Rate?

1. Determineatimeperiod:monthly,annual,orquarterly.
2. Determine the number of customers we had at the beginning of the timeperiod.
3. Determinethenumberofcustomersthatchurnedbytheendofthetimeperiod.
4. Dividethenumberoflostcustomersbythenumberofcustomers wehadpriorto thechurn.
5. Multiplythatnumberby 100.

## Formula:

Customer Churn Rate = (Lost Customers ÷ Total Customers at the Start of TimePeriod)x100

# ProblemDefinition

Customer churn is when a company’s customers stop doing business with thatcompany. Businesses are very keen on measuring churn because keeping anexisting customer is far less expensive than acquiring a new customer. Newbusiness involves working leads through a sales funnel, using marketing andsales budgets to gain additional customers. Existing customers will often have ahigher volume of service consumption and can generate additional customerreferrals.

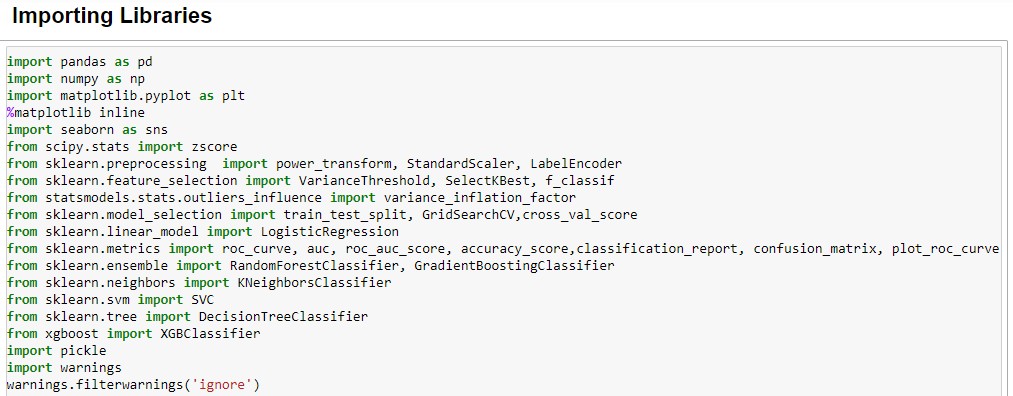
Customer retention can be achieved with good customer service and products.Butthemosteffectivewayforacompanytopreventattritionofcustomersistotruly know them. The vast volumes of data collected about customers can beused to build churn prediction models. Knowing who is most likely to defectmeans that a company can prioritise focused marketing efforts on that subsetoftheir customerbase.

Preventing customer churn is critically important to the telecommunicationssector,as thebarrierstoentryforswitchingservices areso low.

Here we are examining customer data from IBM Sample Data Sets with the aimofbuildingandcomparingseveralcustomerchurnpredictionmodels.

# DataAnalysis.

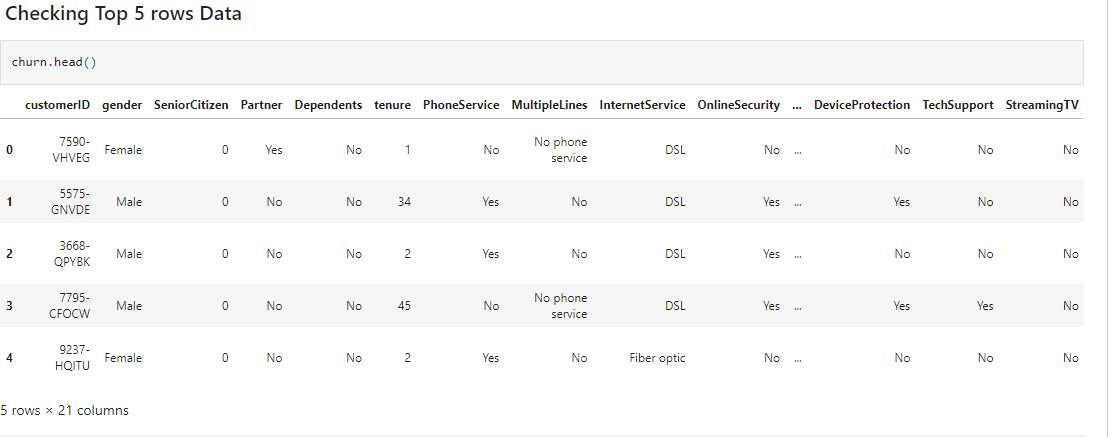
## Firstthingbeforeanalysisistoimportnecessarylibraries



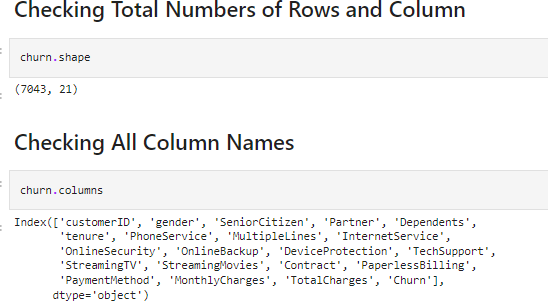
* + **SecondforanalysisimportDataset(Telecom\_customer\_churn.csv)**



## CheckingTop5rowsofDatasettoanalysethedata/valuesofdataset



* **Analysing TotalNumbersofRowsandColumnandAllColumnNames**



Weobservedthatthisdatasetcontains7043rowsand21columns.Namesofcolumnwith brief descriptionare:

## Featurescolumns:

* **CustomerID:**CustomerIDuniqueforeachcustomer
* **gender:**Whether thecustomeris amaleor afemale
* **SeniorCitizen:**Whetherthecustomerisaseniorcitizenornot(1,0)
* **Partner:**Whetherthecustomerhasapartnerornot(Yes,No)
* **Dependent:**Whetherthecustomerhasdependentsornot(Yes,No)
* **PhoneService:**Whetherthecustomerhasaphoneserviceornot(Yes,No)
* **MultipeLines:**Whetherthecustomerhasmultiplelinesornot(Yes,No,Nophoneservice)
* **InternetService:**Customer’sinternetserviceprovider(DSL,Fiberoptic,

No)

* **OnlineSecurity:**Whetherthecustomerhasonlinesecurityornot(Yes,No,Nointernetservice)
* **OnlineBackup:**Whetherthecustomerhasanonlinebackupornot(Yes,No,Nointernetservice)
* **DeviceProtection:**Whetherthecustomerhasdeviceprotectionornot(Yes,No,Nointernetservice)
* **TechSupport:**Whetherthecustomerhastechsupportornot(Yes,No,Nointernetservice)
* **StreamingTV:**WhetherthecustomerhasstreamingTVornot(Yes,No,Nointernetservice)
* **StreamingMovies:**Whetherthecustomerhasstreamingmoviesornot(Yes,No,Nointernetservice)
* **Contract:**Thecontracttermofthecustomer(Month-to-month,Oneyear,Twoyears)
* **PaperlessBilling:**Thecontracttermofthecustomer(Month-to-month,Oneyear,Twoyears)
* **PaymentMethod:**Thecustomer’spaymentmethod(Electroniccheck,

Mailedcheck,Banktransfer (automatic),Creditcard(automatic))

* **Tenure:**Numberofmonthsthecustomerhasstayedwiththecompany
* **MonthlyCharges:**Theamountchargedtothecustomermonthly
* **TotalCharges:**Thetotalamountchargedtothecustomer

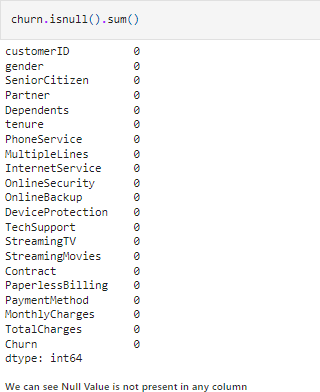
**Targetcolumn:**

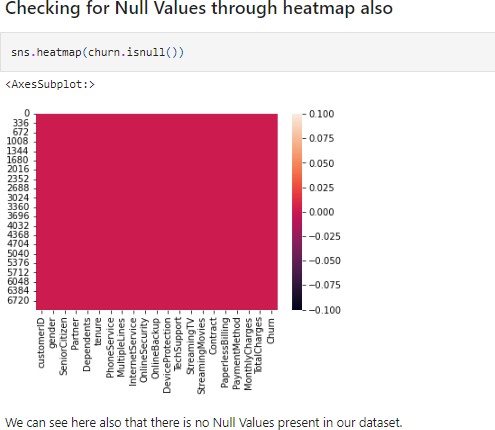
* **Churn**:Whether thecustomerchurnedornot(Yes orNo)

**Thesefeaturescanalsobesubdividedinto:**

* **Demographiccustomerinformation:**gender,SeniorCitizen,Partner,Dependents
* **ServicesInformation:**PhoneService,MultipleLines,InternetService,OnlineSecurity,OnlineBackup,DeviceProtection,TechSupport,StreamingTV,StreamingMovies
* **Customeraccountinformation:**tenure,Contract,PaperlessBilling,PaymentMethod,MonthlyCharges,TotalCharges

## AnalysingNullValues

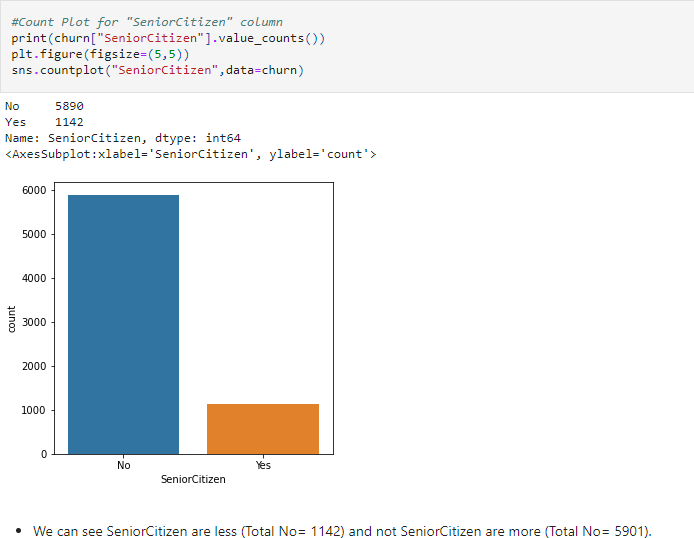
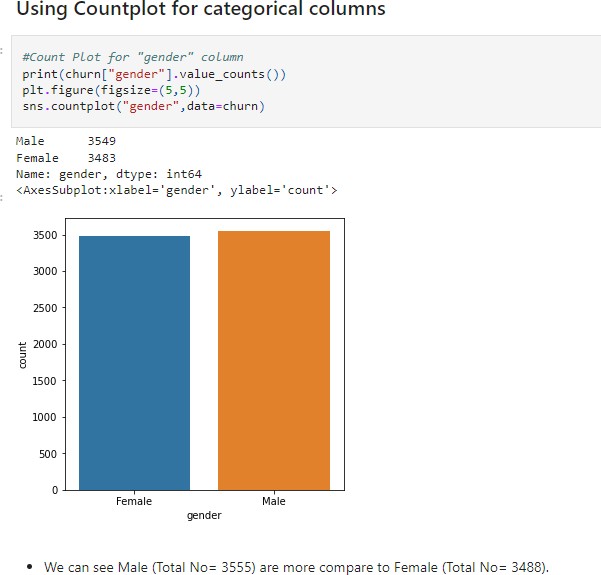


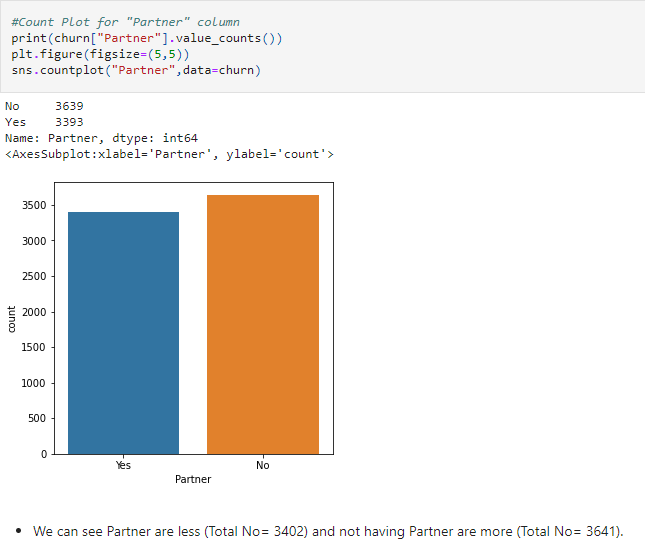


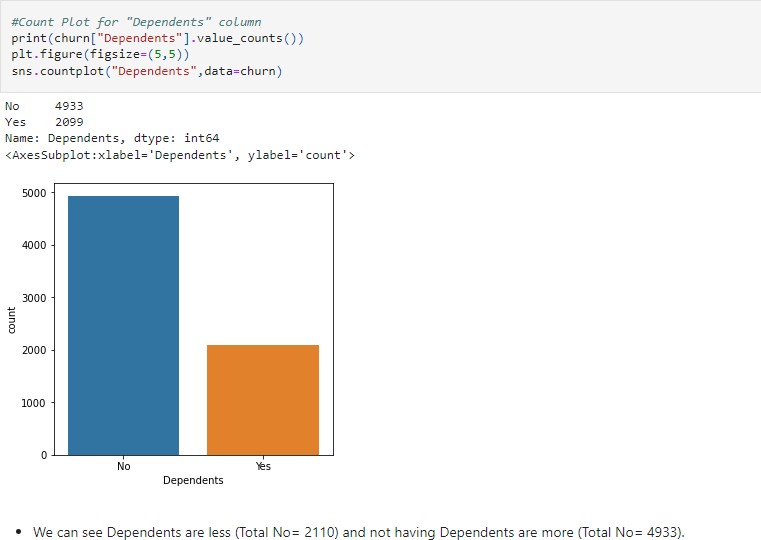
WecanobservetherearenoNullValuesinDataset.

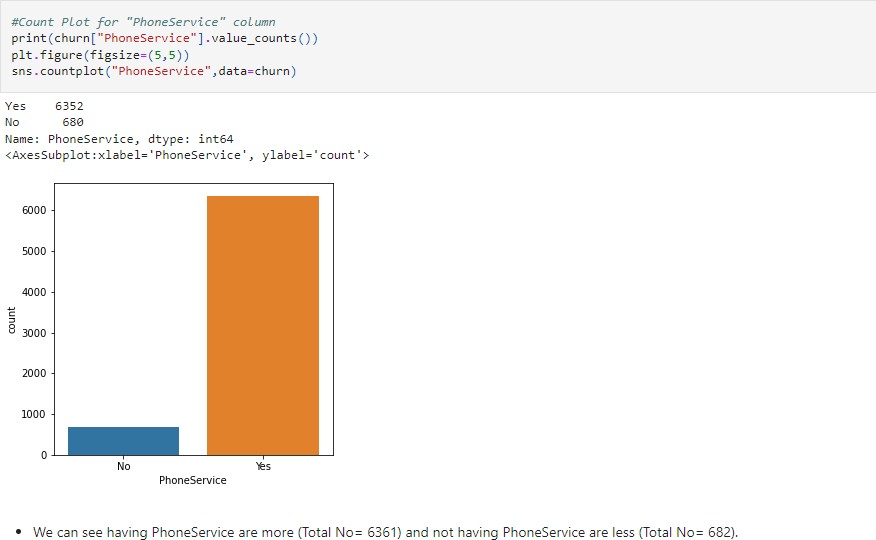
## Let’splottoAnalyseeachcolumn’sdatathroughDataVisualization

* + **UnivariantAnalysis**
    - Usingcountplot(forCategoricalColumns)



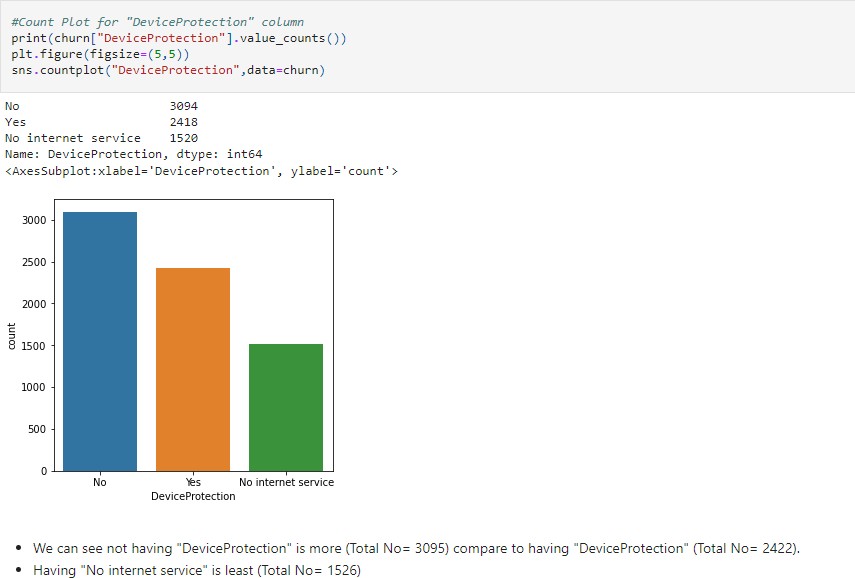




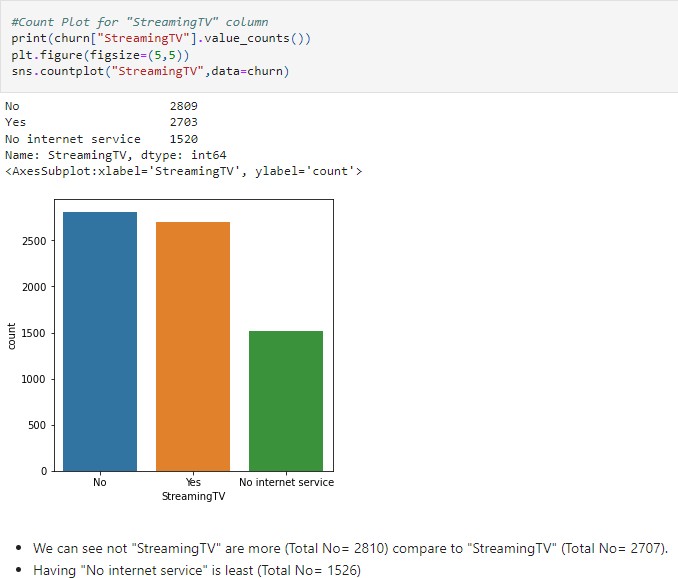




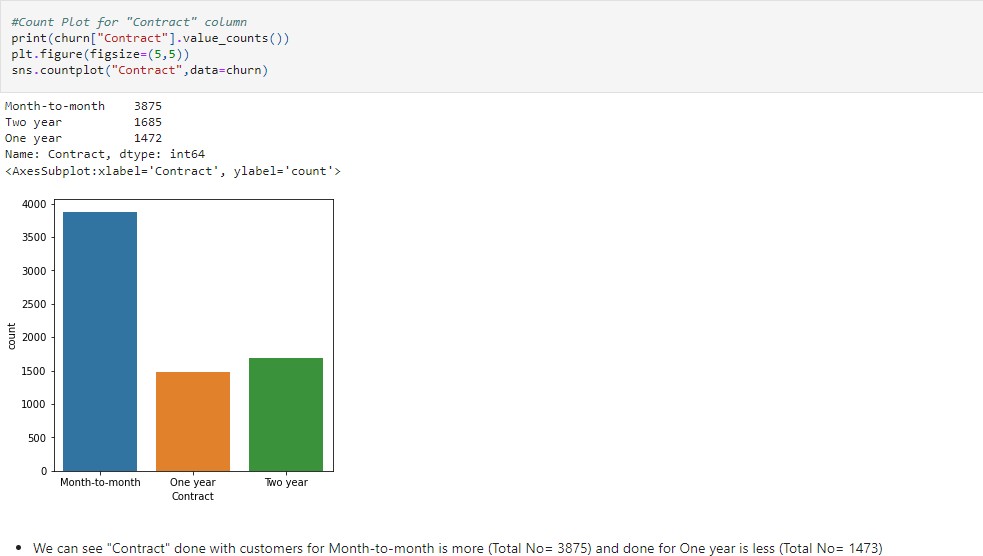




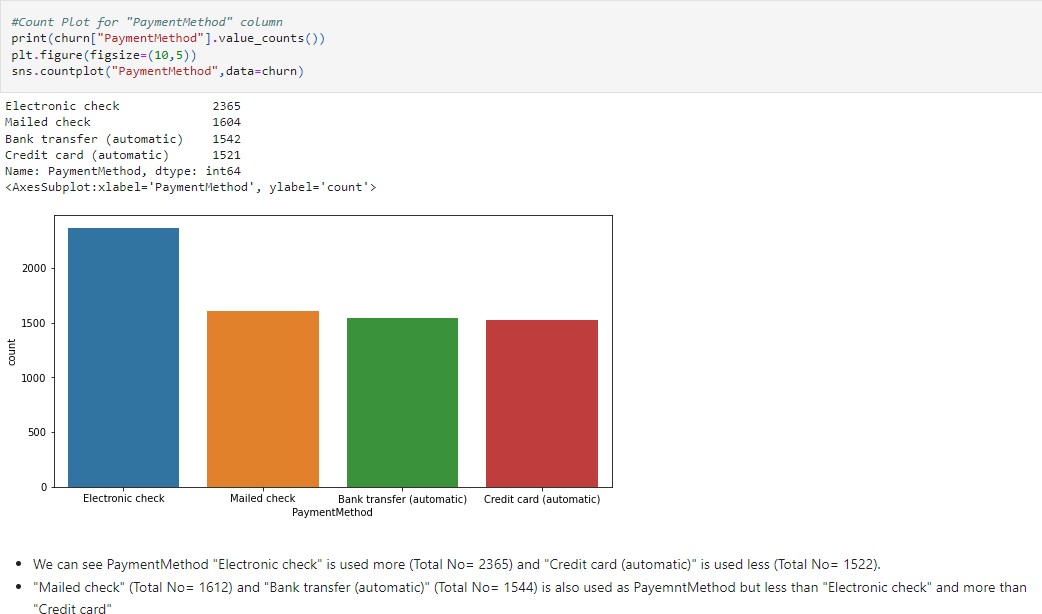






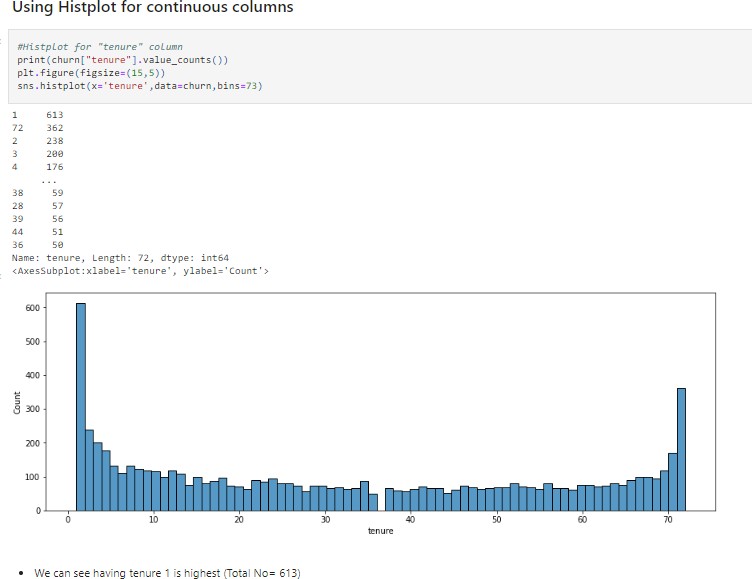


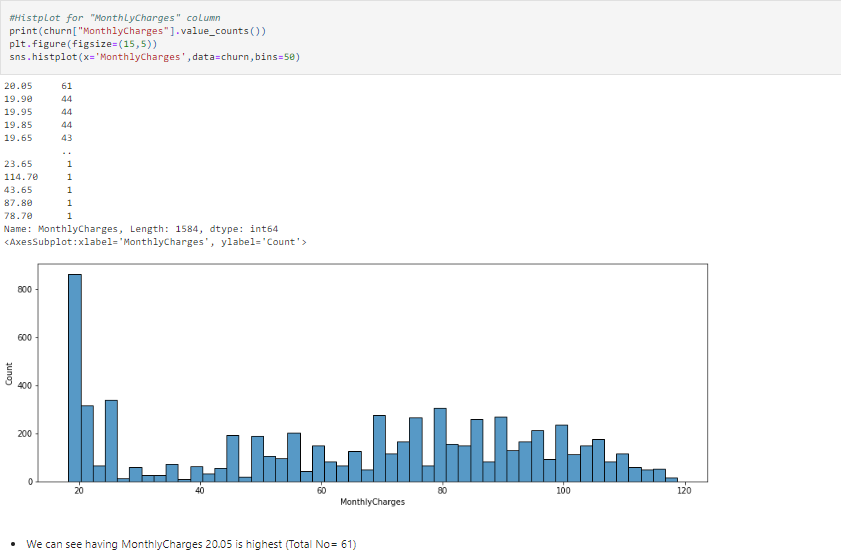


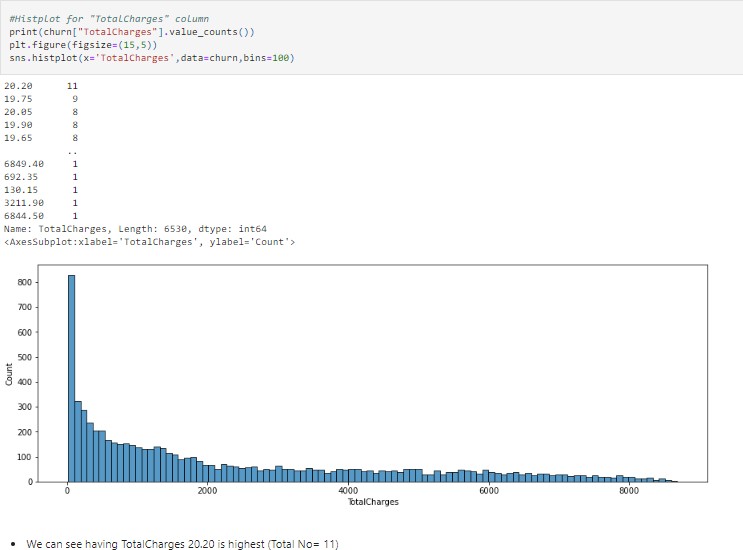




* + - UsingHistplot(forContinuousColumns)

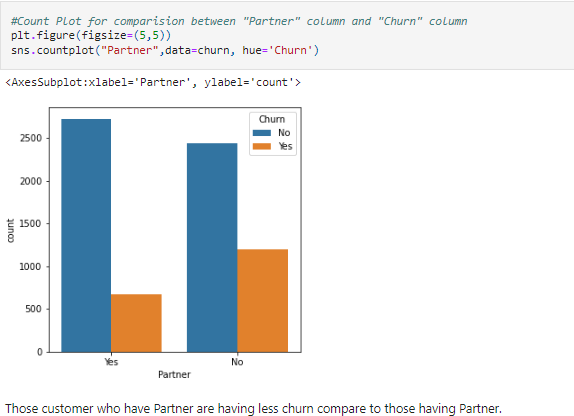


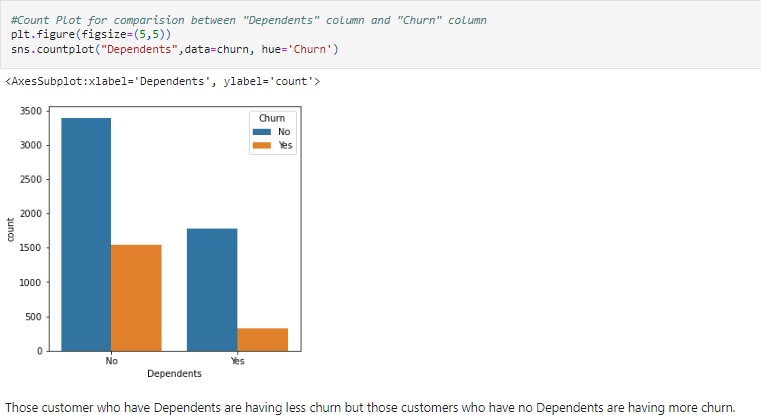




* + BivariantAnalysis(forcomparison)
    - Usingcountplot(forcomparisonbetweencategoricalcolumnandtarget)





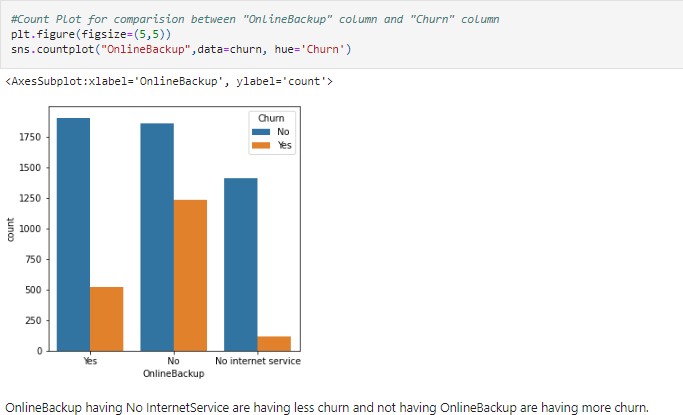


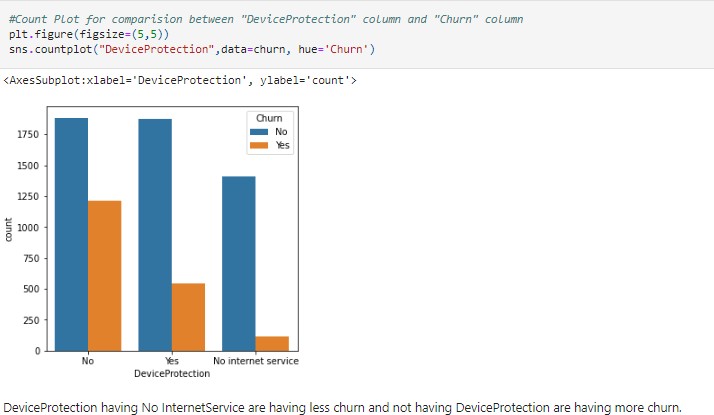


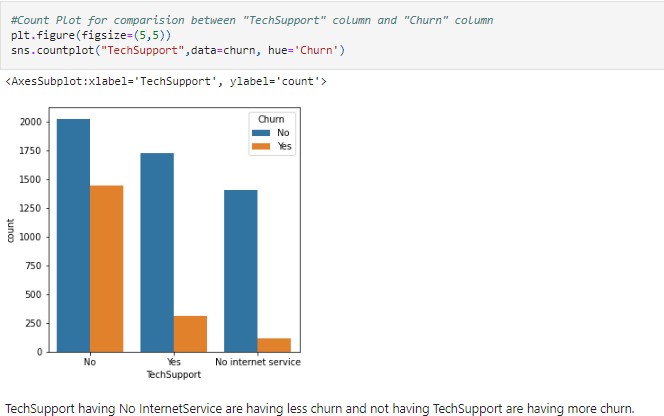




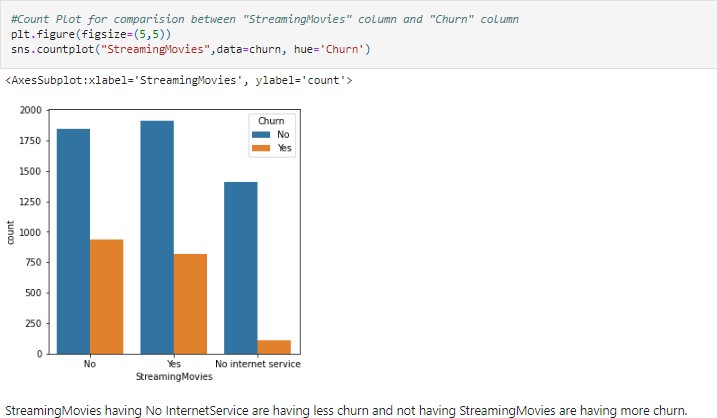




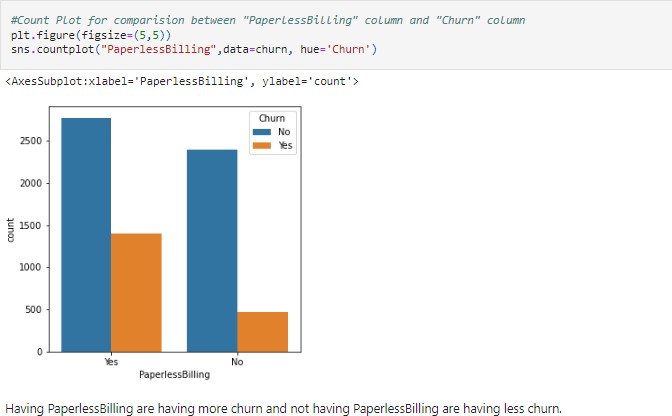


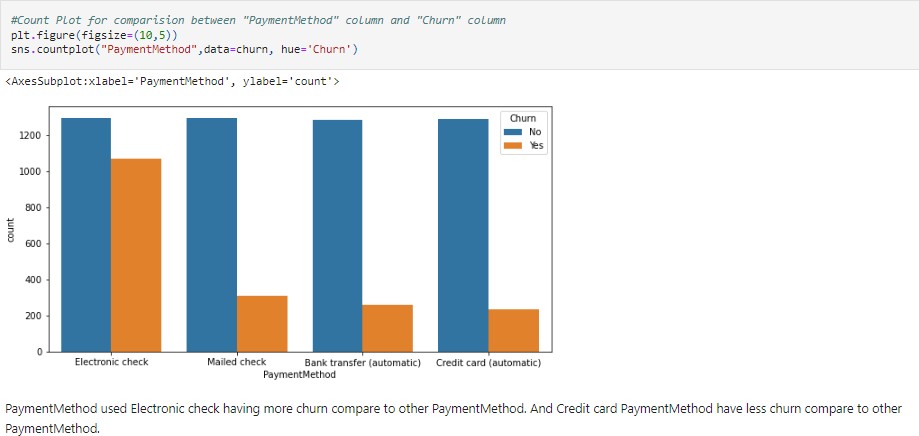




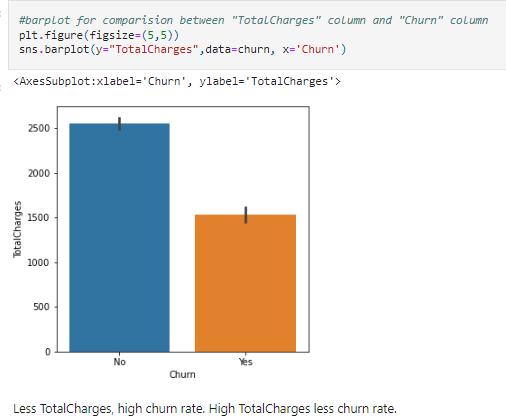






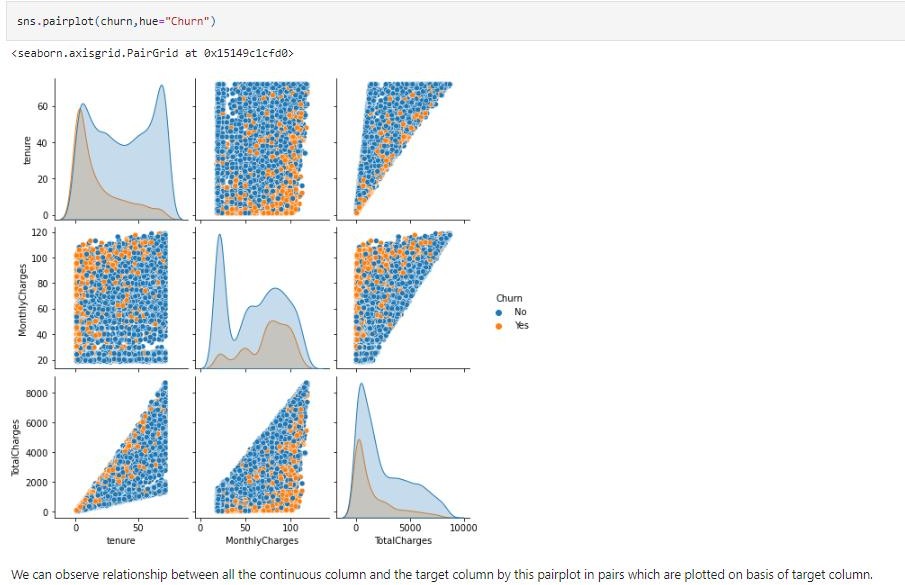


* + - UsingBarplot(forcomparisonbetweencontinuouscolumnandtarget)



* + **MultivariantAnalysis:**UsingPairplot

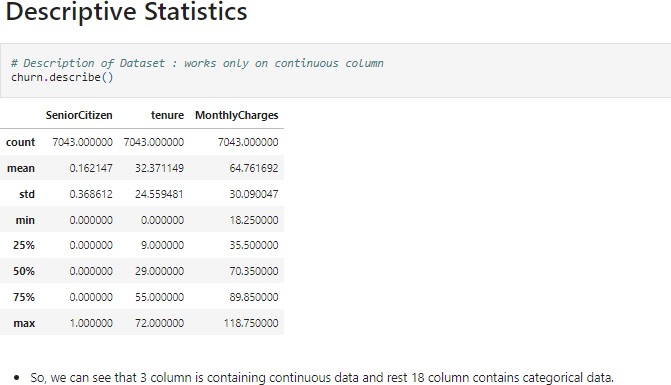
(Relationshipbetweencontinuousandtargetcolumn)

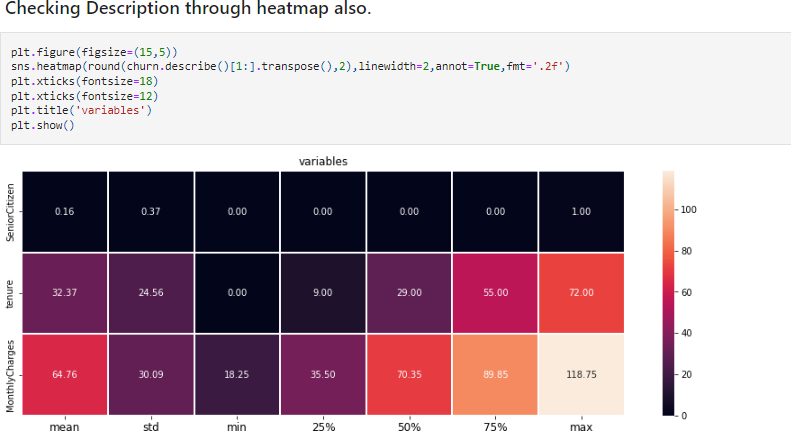


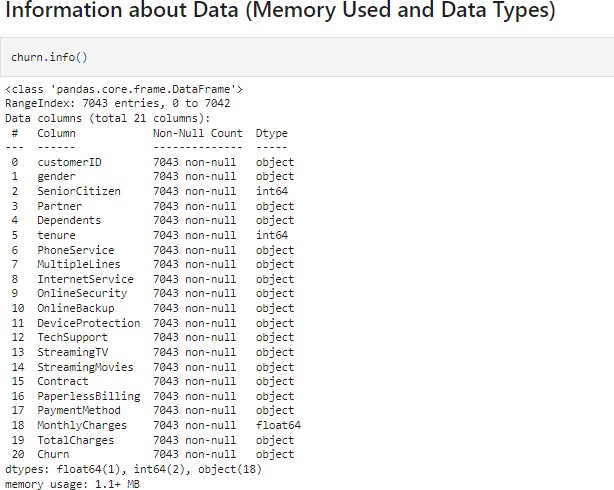
## ObservationofDataVisualisation

* Thechurnrateofseniorcitizensis approx. doubleofyoungcitizens.
* We cannot determine churn rate through Gender as both male and female areequal.
* Customerswithpartnerishavinglesschurncomparetocustomerswithnopartner.
* Customerswithmonth-to-monthcontractshavehigherchurnratescomparedtoCustomershavingyearlycontracts.
* Customerswhooptedforanelectroniccheckaspayingmethodarehavingmorechurnrate.
* Customerssubscribedtopaperlessbillingchurnmore
* Thechurnrate ishighwhenmonthlychargesarehigh.
* Newcustomerschurnmore.
* Clientswithhightotalchargeshavinglesschurn.
* We do not expect phone attributes (PhoneService and MultipleLines) to havesignificant predictive power. The percentage of churn for all classes in bothindependentvariablesisnearlythesame.
* Clientswithonlinesecuritychurnless.
* Customerswithnotechsupporttendtochurnmoreoftenthanthosewithtechsupport.

# EDAConcludingRemark







There are 18 categorical features including Target Feature and 3 continuousfeaturesin Dataset.

We also observed through visualization that column TotalCharges is wronglydetectedas**objectdatatype.**Thiscolumnrepresentsthetotalamountchargedtothecustomeranditshouldbeanumericvariable. So,forfurtheranalysis,weneedtotransformthiscolumnintoa **numericdatatype**throughfunctionpd.to\_numeric.



AfterconvertingDatatypeofcolumnTotalCharges,wecanseethereare11nullvalues present in column TotalCharges. So handled null values through meanmode.



Andalsoobservedthatcolumntenureishaving0values.Sohavetohandlethistoo.So,first checkforindexnumberwhere0valuesarepresent.



Wegettheseindex numbers:



Therefore,tohandlethiswearedroppingtheserowsasitisonly11anditwillnotaffectprediction.



# Pre-ProcessingPipeline

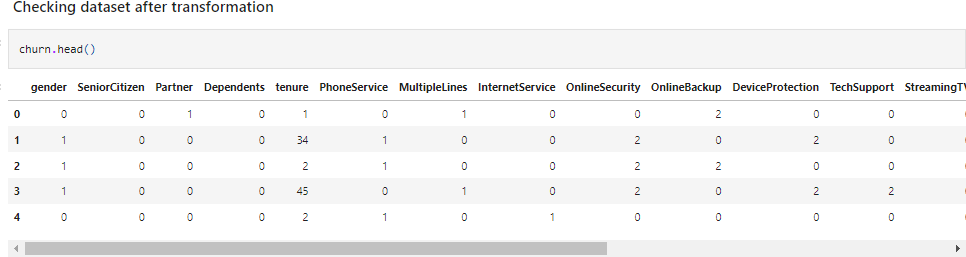
Column customerID is having unique value for each data so it has no relationwithTarget.Hence,willdrop this column.



## LabelEncoding

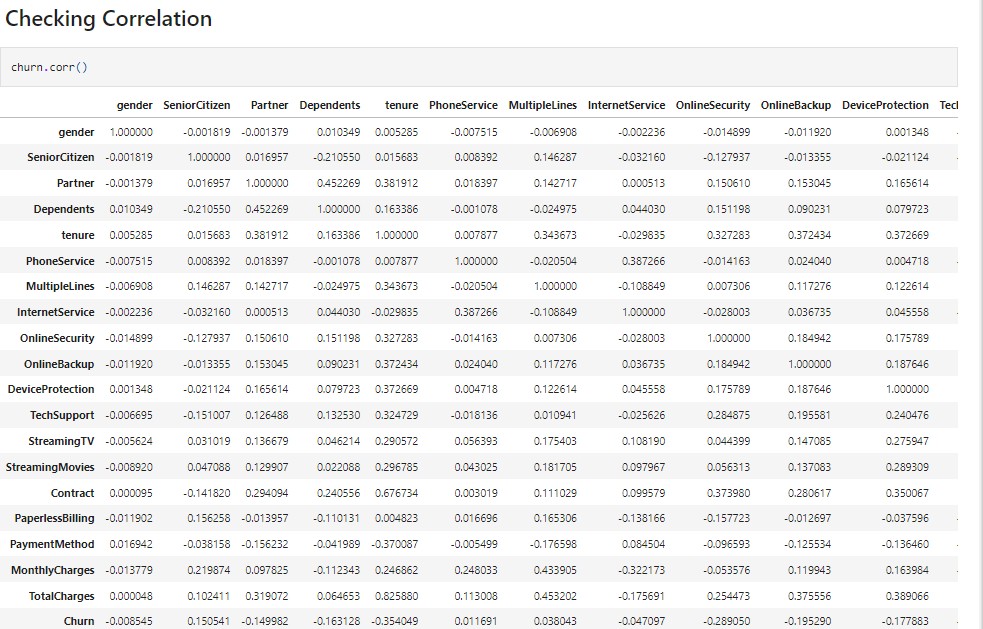
Label Encoding refers to converting the labels into a numeric form to convertthem into the machine-readable form. Machine learning algorithms can thendecide in a better way how those labels must be operated. It is an importantpre-processingstepforthestructureddatasetin supervisedlearning.

So,weareencodingthedatabytransformingfromcategoricaltocontinuous.

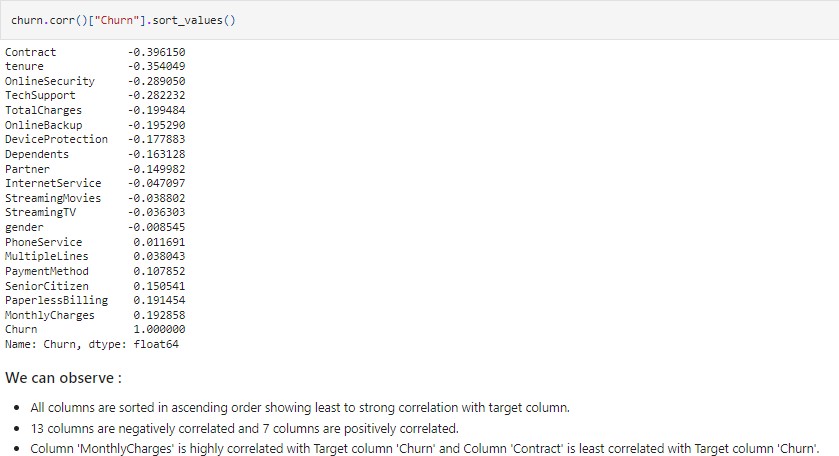


## CheckingCorrelation

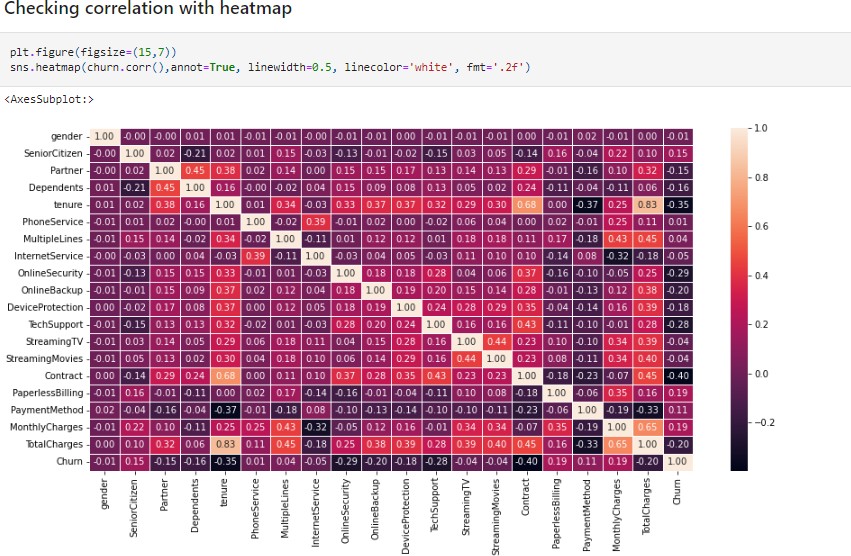
Data correlation is the way in which one set of data may correspond to anotherset means measures the linear relationship between two variables. It describeshowoneormorevariablesarerelatedtoeachother.Thesevariablescanbeinputdatafeatureswhichhavebeenusedtoforecastourtargetvariable.

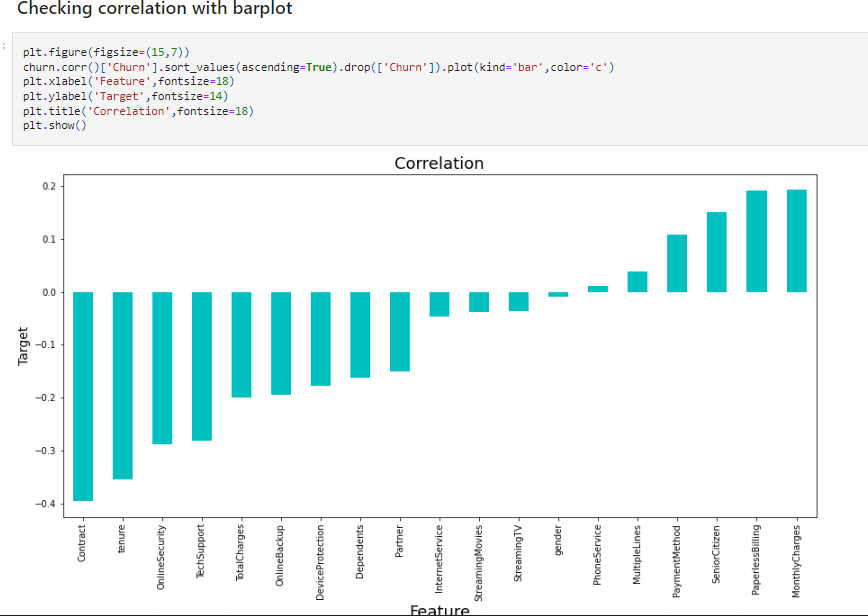


* + Checkingcorrelationinascendingorderalsotogethighestandlowestcorrelatedvalues.



* + Observingcorrelationthroughdatavisualizationalso:



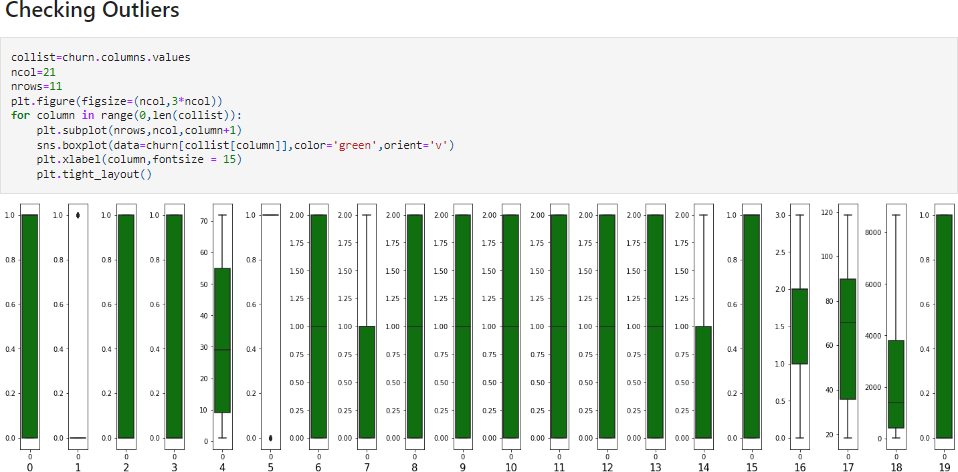


## ObservationofCorrelation:

* genderhas-1percentcorrelationwiththetargetcolumnwhichcanbeconsideredas goodcorrelation andnegativelycorrelated.
* SeniorCitizen has 15 percent correlation with the target columnwhichcanbeconsideredasgoodcorrelationandpositivelycorrelated.
* Partner has -15 percent correlation with the target column whichcanbeconsideredasgoodcorrelationandnegativelycorrelated.
* Dependents has -16 percent correlation with the target columnwhichcanbeconsideredasgoodcorrelationandnegativelycorrelated.
* tenurehas-35percentcorrelationwiththetargetcolumnwhichcanbeconsideredas goodcorrelation and negativelycorrelated.
* PhoneServicehas1percentcorrelationwiththetargetcolumnwhichcanbeconsideredasgoodcorrelationandpositivelycorrelated.
* MultipleLineshas4percentcorrelationwiththetargetcolumnwhichcanbeconsideredasgoodcorrelationandpositivelycorrelated.
* InternetService has -5 percent correlation with the target columnwhichcanbeconsideredasgoodcorrelationandnegativelycorrelated.
* OnlineSecurity has -29 percent correlation with the target columnwhichcanbeconsideredasgoodcorrelationandnegativelycorrelated.
* OnlineBackup has -20 percent correlation with the target columnwhichcanbeconsideredasgoodcorrelationandnegativelycorrelated.
* DeviceProtectionhas-18percentcorrelationwiththetargetcolumnwhichcanbeconsideredasgoodcorrelationandnegativelycorrelated.
* TechSupport has -28 percent correlation with the target columnwhichcanbeconsideredasgoodcorrelationandnegativelycorrelated.
* StreamingTVhas-4percentcorrelationwiththetargetcolumnwhichcanbeconsideredasgoodcorrelationandnegativelycorrelated.
* StreamingMovies has -4 percent correlation with the target columnwhichcanbeconsideredasgoodcorrelationandnegativelycorrelated.
* Contract has -40 percent correlation with the target column whichcanbeconsideredas weakcorrelation andnegativelycorrelated.
* PaperlessBilling has 19 percent correlation with the target columnwhichcanbeconsideredasgoodcorrelationandpositivelycorrelated.
* PaymentMethod has 11 percent correlation with the target columnwhichcanbeconsideredasgoodcorrelationandpositivelycorrelated.
* MonthlyCharges has 19 percent correlation with the target columnwhichcanbeconsideredasstrongcorrelationandpositivelycorrelated.
* TotalCharges has -20 percent correlation with the target columnwhichcanbeconsideredasgoodcorrelationandnegativelycorrelated.
* **Maxcorrelation**iswithMonthlyCharges
* **Mincorrelation**iswithContract

## FindingOutliers:

Anoutlierisadatapointthatisnoticeablydifferentfromtherest.Theyrepresent errors in measurement, bad data collection, or simply show variablesnotconsidered whencollectingthedata.

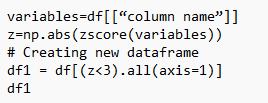


## Observation:

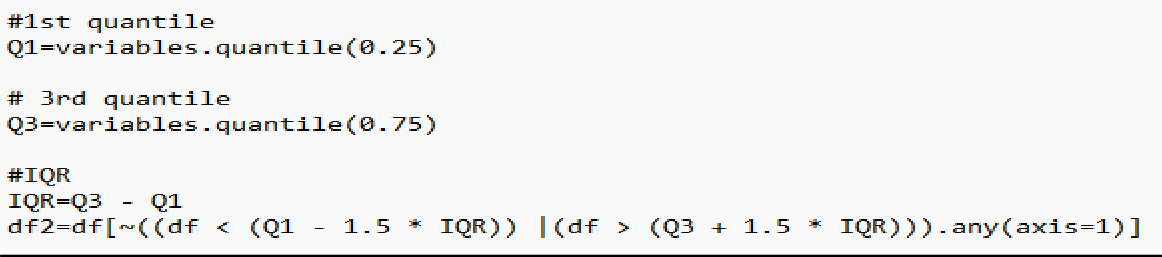
We can see Outliers are present only in 2 columns: "SeniorCitizen" and"PhoneService". But both columns are categorical, so we will not removeoutliers.

If Outliers were present then we have to remove through Zscore methodusing Scipy or IQR (Inter Quantile Range) method. Every time we shouldcheck both method for removal and then compare between both whichmethodisgivinglessdatalossthenusethatmethodforfurtherprocess.

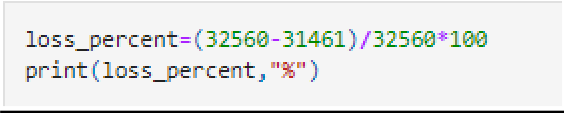
## FormulaforZscoremethodusingScipy:



**FormulaforIQR(InterQuantileRange)method:**



## FormulaforcheckingLossPercent:



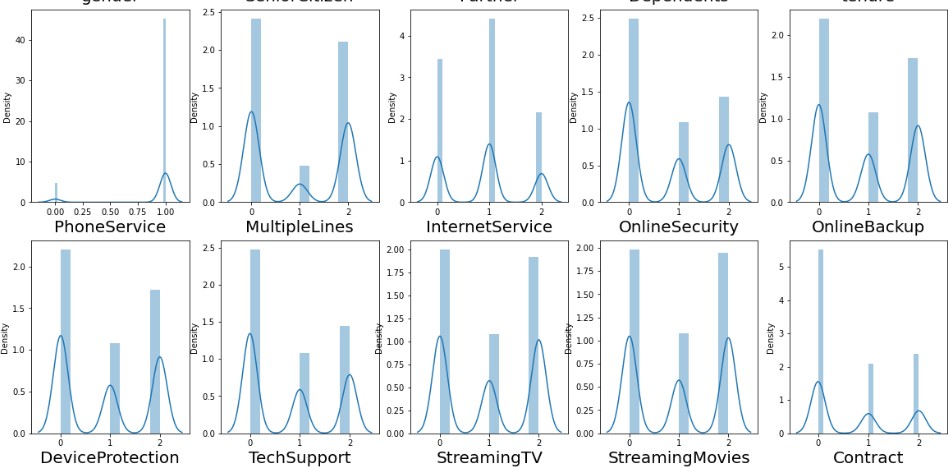
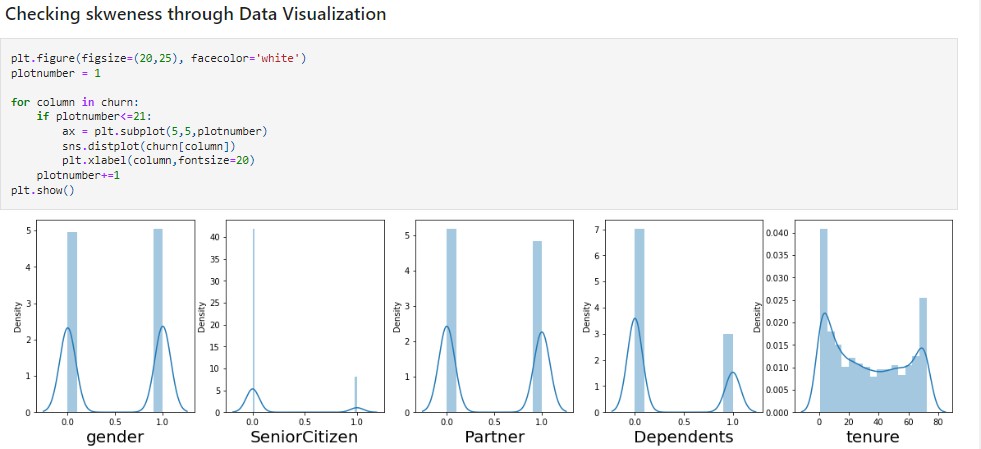
* **FindingSkewness:**

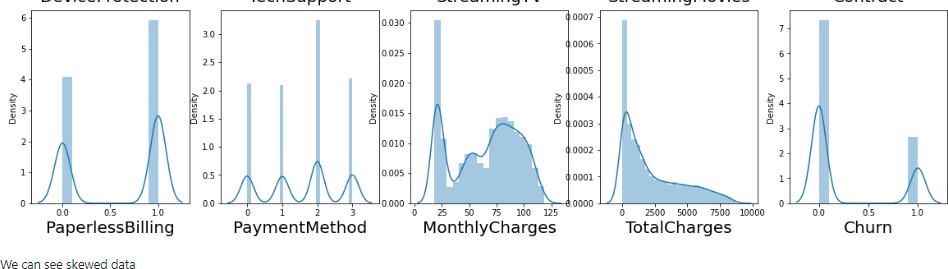
Skewness is asymmetry in a statistical distribution, in which thecurveappearsdistortedorskewedeithertotheleftortotheright.Skewnesscanbequantifiedtodefinetheextenttowhichadistributiondiffersfrom anormaldistribution.

## Thethreetypesofskewnessare:

* Right skew (also called positive skew). A right-skeweddistributionislongerontherightsideofitspeakthanonitsleft.
* Left skew (also called negative skew). A left-skewed distributionislonger on theleftsideof itspeakthan on its right.
* Zeroskew.

## SkewnesscheckedthroughDataVisualizationalso:





**Observation:**

* Skewnessthresholdtakenis+/-0.25
* Columns which are having skewness: SeniorCitizen, Dependents,PhoneService, OnlineSecurity, TechSupport, Contract, PaperlessBillingandTotalCharges.

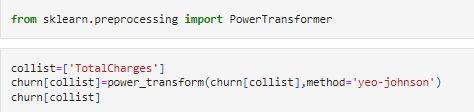
## RemovalofSkewness

For removing skewness Power Transformer method is used. Power transformsare a technique for transforming numerical input or output variables to have auniform or a Gaussian probability distribution. A power transform will maketheprobabilitydistribution of avariablemoreGaussian.

Currently, Power Transformer supports the Box-Cox transform and the Yeo-Johnsontransform.Theoptimalparameterforstabilizingvarianceandminimizingskewnessis estimatedthrough maximumlikelihood.

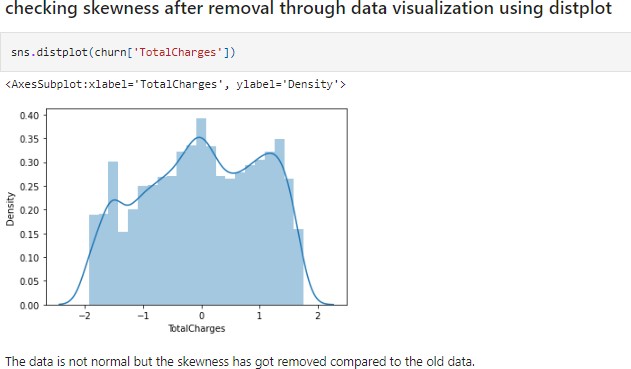
Box-Cox requires input data to be strictly positive, while Yeo-Johnson supportsbothpositiveand negativedata.

So we are using Yeo-Johnson method for removing skewness. SinceSeniorCitizen, Dependents, PhoneService, OnlineSecurity, TechSupport,Contract and PaperlessBilling are categorical column so we will not removeskewness from them. Only we will remove skewness from TotalCharges as thiscolumncontainscontinuous data.



Comparisonbetweenskewedandafterremovalofskewness

|  |  |
| --- | --- |
| Databeforeskewnessremoval | Dataafterskewnessremoval |
|  |  |



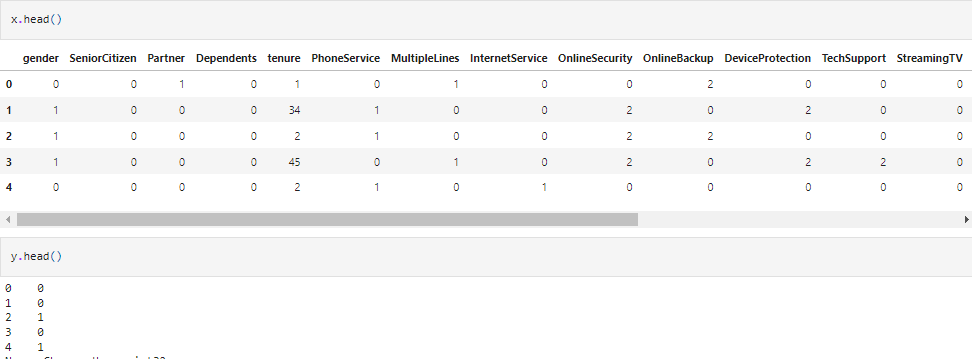
* **SplittingdataintoTargetandFeatures**

Split the data set into two set:a training set and a testing set. This consists ofrandomsamplingwithoutreplacementabout75percentoftherowsandputting them into our training set and the remaining 25 percent is put into ourtestset.

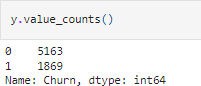
Separating data into training and testing sets is an important part of evaluatingdata mining models. Typically, when we separate a data set into a training setand testing set, most of the data is used for training, and a smaller portion ofthedataisused for testing.



* Dataaftersplitting:



* Totalvaluecountof ycolumn:



Hereitisobservedthatdataisnotbalanced.So,wewilluseoversamplingmethodto balanceit.

## OversamplingMethod

Oversampling methods duplicate examples in the minority class or synthesizenewexamplesfromtheexamplesintheminorityclass.Someofthemorewidelyusedand implemented oversamplingmethods include:

* + RandomOversampling
  + SyntheticMinorityOversamplingTechnique(SMOTE)
  + Borderline-SMOTE
  + BorderlineOversamplingwithSVM
  + AdaptiveSyntheticSampling(ADASYN)

## RandomOversampling:

The simplest oversampling method involves randomly duplicating examplesfromtheminorityclassin thetrainingdataset

## SMOTE:

The most popular and most successful oversampling method is SMOTE; that isanacronymforSynthetic MinorityOversamplingTechnique.

SMOTEworksbyselectingexamplesthatarecloseinthefeaturespace,drawinga line between the examples in the feature space and drawing a new sample asapointalongthatline.

So,to balancethedatahereusingSMOTEmethod

## Borderline-SMOTE:

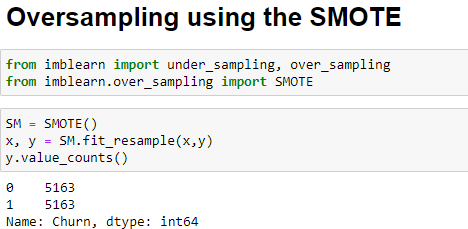
Selectingthoseinstancesoftheminorityclassthataremisclassified,suchaswithak-nearestneighborclassificationmodel,andonlygeneratingsyntheticsamplesthatare“difficult”toclassifyisBorderline-SMOTE.

## BorderlineOversamplingwithSVM:

ItisanextensiontoSMOTEthatfitsanSVMtothedatasetandusesthedecisionboundaryasdefinedbythesupportvectorsasthebasisforgeneratingsyntheticexamples,againbasedontheideathatthedecisionboundaryistheareawheremoreminorityexamplesarerequired.

## AdaptiveSyntheticSampling(ADASYN):

(ADASYN) is another extension to SMOTE that generates synthetic samplesinversely proportional to the density of the examples in the minority class. It isdesignedtocreatesyntheticexamplesinregionsofthefeaturespacewherethedensityofminorityexamplesislow,andfewerornonewherethedensityishigh.



## ScalingData:

It is a step of Data Pre-Processing that is applied to independent variables orfeatures of data. It basically helps to normalize the data within a particularrange.Sometimes,italso helpsin speedingupthecalculationsin an algorithm.

Normalization and Standardization are the two main methods for the scaling ofthedatawhicharewidelyusedinthealgorithmswherescalingisrequired.Bothofthemcanbeimplementedbythescikit-learnlibrariespre-processpackage.

**Normalization**isusedwhenwewanttoboundourvaluesbetweentwonumbers,typically,between[0,1]or[-1,1].Itisalsoknown asMin-Max scaling.

**Formulafornormalization:**

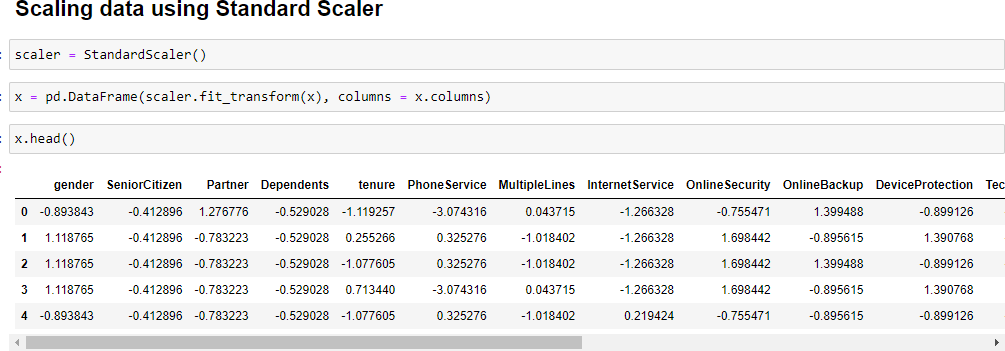


While**Standardization**transformsthedatatohavezeromeanandavarianceof1, they make our data unitless. The values are centred around the mean with aunitstandarddeviation.Thismeansthatthemeanoftheattributebecomeszeroandtheresultantdistributionhas aunitstandard deviation.

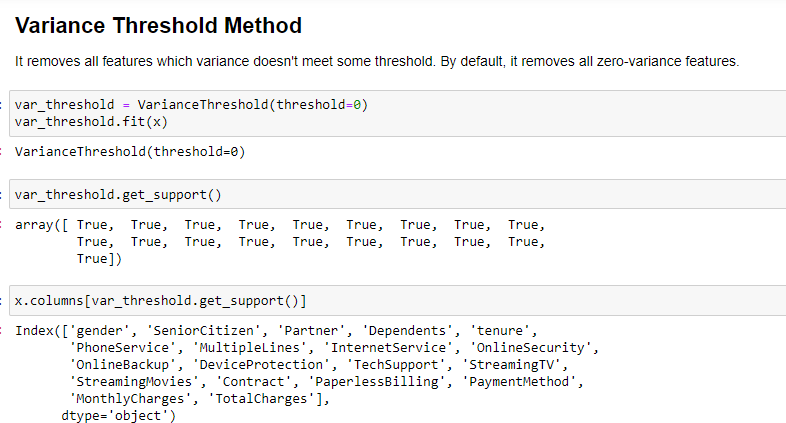
**Formulafornormalization:**

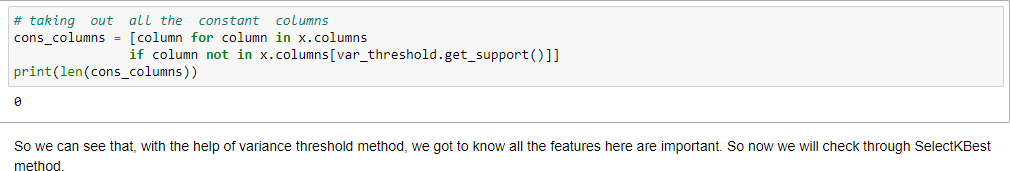
Standardization equation

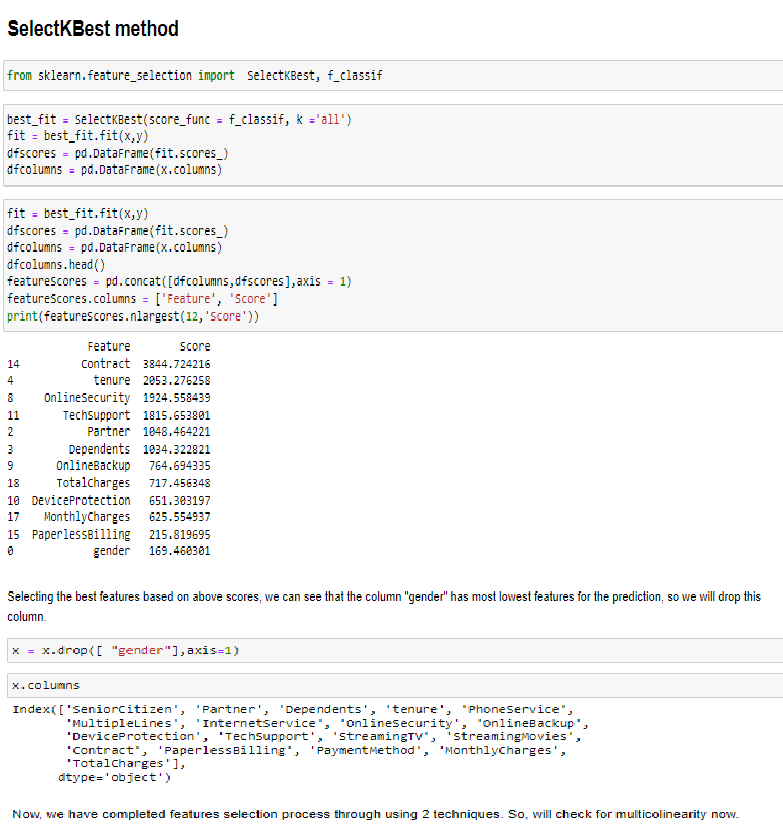
HereweareusingStandardizationtoscalethedata.



Before model creation, we need to know which features are important forpredictionandwhicharenot,SowewillcheckusingVarianceThresholdMethod,SelectKBestmethodandMulticollinearityusingVarianceInflationFactor.





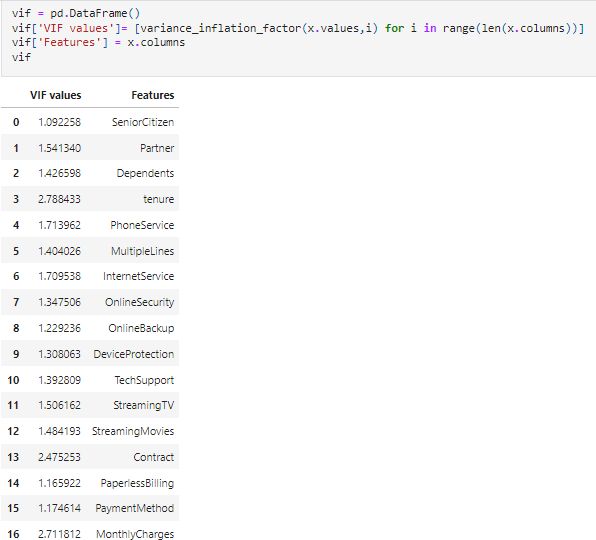


## CheckingforMulticollinearityusingVarianceInflationFactor

* **VIF(VarianceInflationfactor)**



We observe that the VIF value is more than 10 in the columns 'tenure' and'TotalCharges'.Butcolumn'TotalCharges'ishavinghighestVIFvalue.So,wewilldropcolumn'TotalCharges'andafterdroppingwewillagaincheckmulticollinearity.



Now, we can check Multicollinearity is removed from the columns as VIF valueofall columns arelessthan 10.So,wewillcreatemodelnow.

# BuildingMachineLearningModels

ClassificationModelwillbebuiltforTargetVariable“Churn”asitcontains

categoricaldata.Wearegoingto usemodels:

* LogisticRegression
* RandomForestClassifier
* DecisionTreeClassifier
* SupportVectorMachineClassifier
* KNeighborsClassifier
* GradientBoostingClassifier
* XGBClassifier

## LogisticRegression:

Itisaclassificationalgorithm,thatisusedwherethetargetvariableiscategorical. Logistic Regression is used to find a relationship between featuresandprobabilityofparticularoutcome.

## RandomForestClassifier:

Random forest is basically the combination of multiple individual decision treestoactasanensemble.Ensemblelearningcanbedefinedasaparadigmwherebymultiple learners are trained to solve the same problem. Ensemble learningactuallyhasbeenusedinseveralapplicationssuchasopticalcharacterrecognition,medicalpurpose,etc.Infact,ensemblelearningcanbeusedwherevermachinelearningtechniques can beused.

When it comes to classification using Random Forests, the idea is that thecombinationofoutputsofmutuallyexclusivenodeswilloutperformanyindividualmodelswhicharethensaidthepredictedoutput.Combiningmultipletrees(learner) maybeabetter choiceif thelearnersareperformingwell.

## DecisionTreeClassifier:

Decision Tree is a Supervised learning technique that can be used for bothclassification and Regression problems, but mostly it is preferred for solvingClassification problems. It is a tree-structured classifier, where internal nodesrepresent the features of a dataset, branches represent the decision rules andeachleaf noderepresentstheoutcome.

In a Decision tree, there are two nodes, which are the Decision Node and LeafNode.Decisionnodesareusedtomakeanydecisionandhavemultiplebranches, whereas Leaf nodes are the output of those decisions and do notcontain anyfurtherbranches.

The decisions or the test are performed on the basis of features of the givendataset.

## SupportVectorMachine:

SVMisasupervisedMachineLearningalgorithmthatisusedinmanyclassificationsandregressionproblems.Itstillpresentsasoneofthemostusedrobust prediction methods that can be applied to many use cases involvingclassifications.

It works by finding an optimal separation line called a hyperplane to accuratelyseparate 2 or more different classes. The goal is to find the optimal hyperplaneseparationthroughtrainingthelinearlyseparabledatawiththeSVMalgorithm.

## KNeighborsClassifier:

Thek-nearestneighborsalgorithm,alsoknownasKNNork-NN,isanon-parametric,supervisedlearningclassifier,whichusesproximitytomakeclassifications or predictions about the grouping of an individual data point.TheKinthenameofthisclassifierrepresentstheknearestneighbors,wherekisaninteger value specified by the user. Hence as the name suggests, this classifierimplements learning based on the k nearest neighbors. The choice of the valueof kis dependentondata.

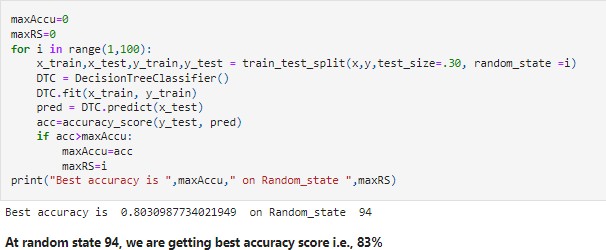
## GradientBoostingClassifier

Gradient boosting classifiers are a group of machine learning algorithms thatcombine many weak learning models together to create a strong predictivemodel. Decision trees are usually used when doing gradient boosting. Gradientboostingmodelsarebecomingpopularbecauseoftheireffectivenessatclassifyingcomplexdatasets.

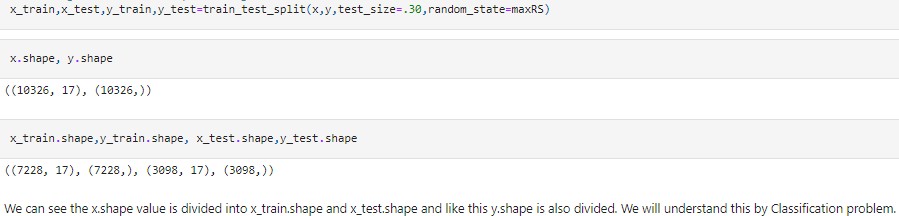
## XGBoostClassifier:

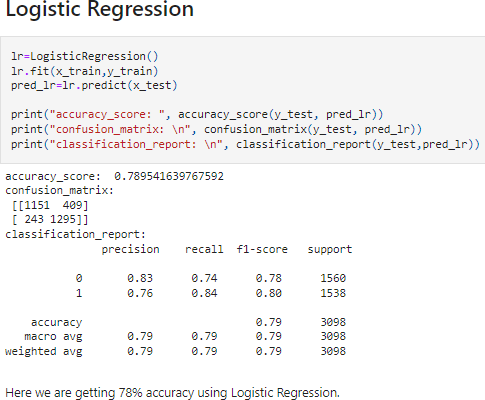
eXtreme Gradient Boosting (XGBoost) is a scalable and improved version of thegradientboostingalgorithm(terminologyalert)designedforefficacy,computationalspeedandmodelperformance.Itisanopen-sourcelibraryandapartoftheDistributedMachineLearningCommunity.XGBoostisaperfectblendof software and hardware capabilities designed to enhance existing boostingtechniqueswith accuracyin theshortestamountof time.

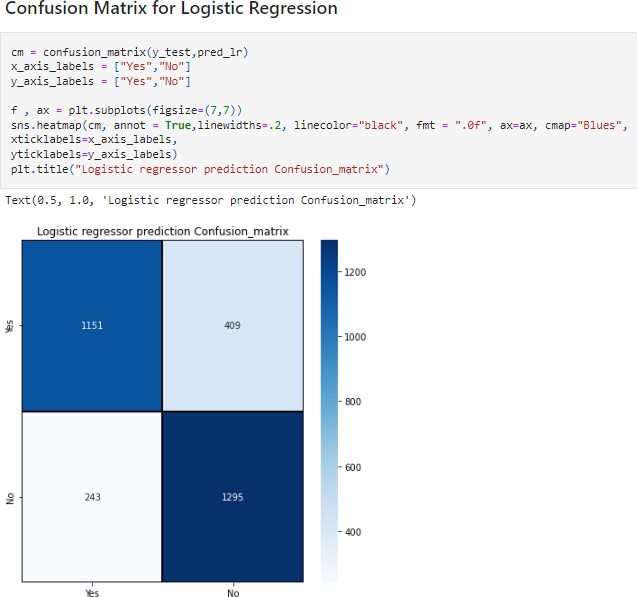
## Findingthebestrandomstateamongallthemodels

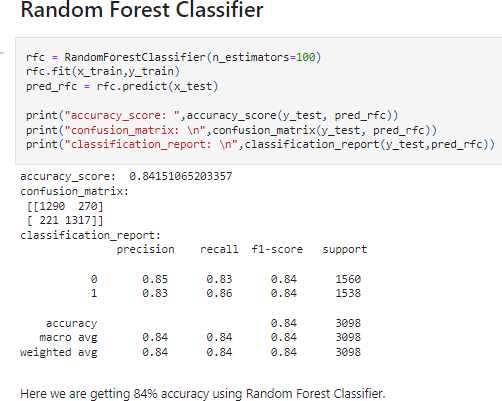


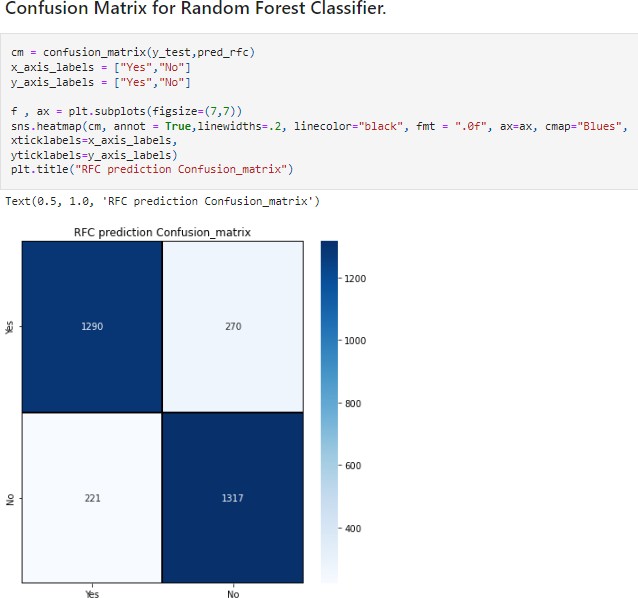
**Creatingnewtraintestsplitusingtherandomstate**

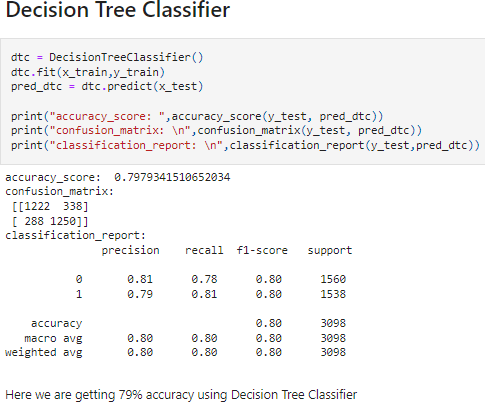


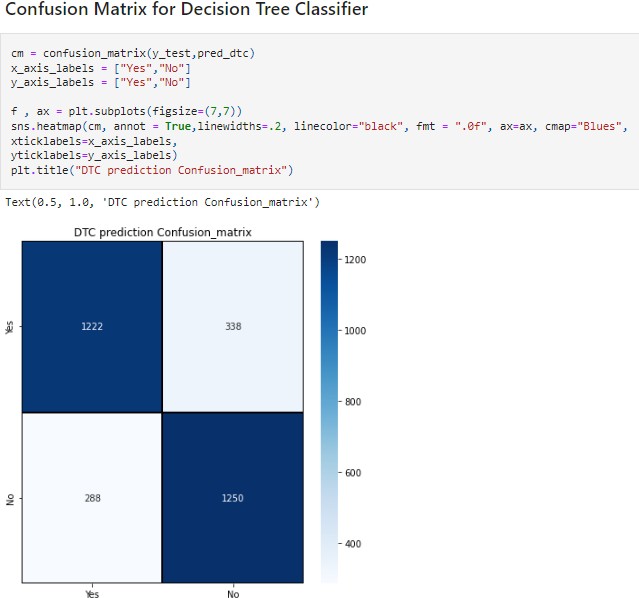


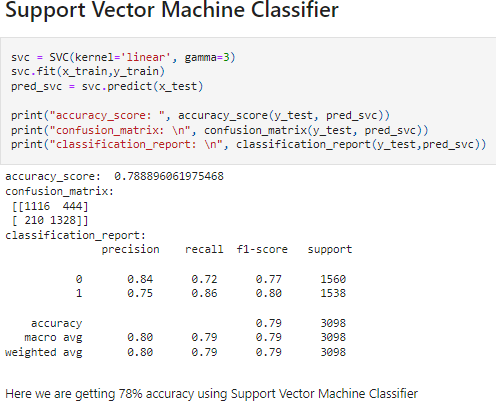


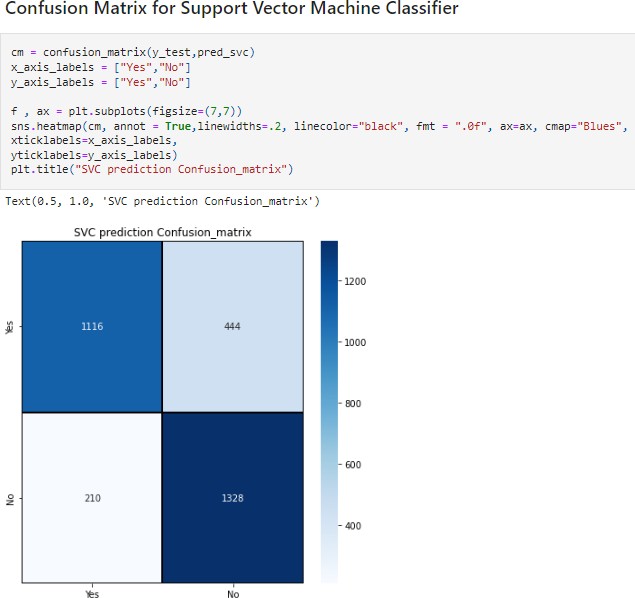


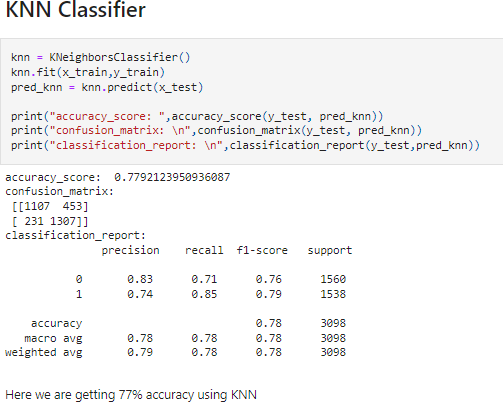


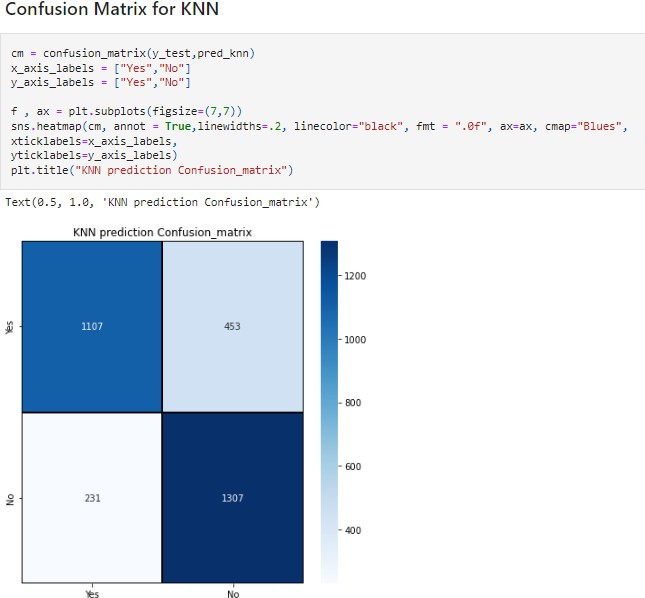


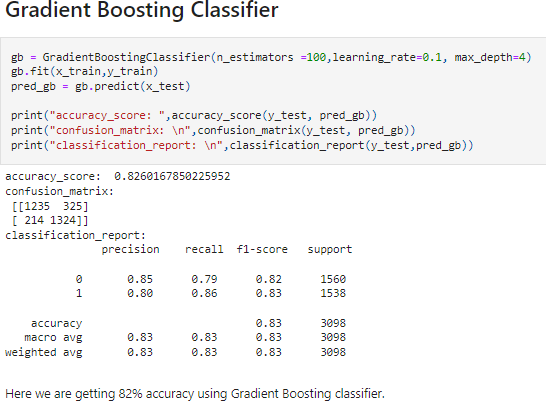


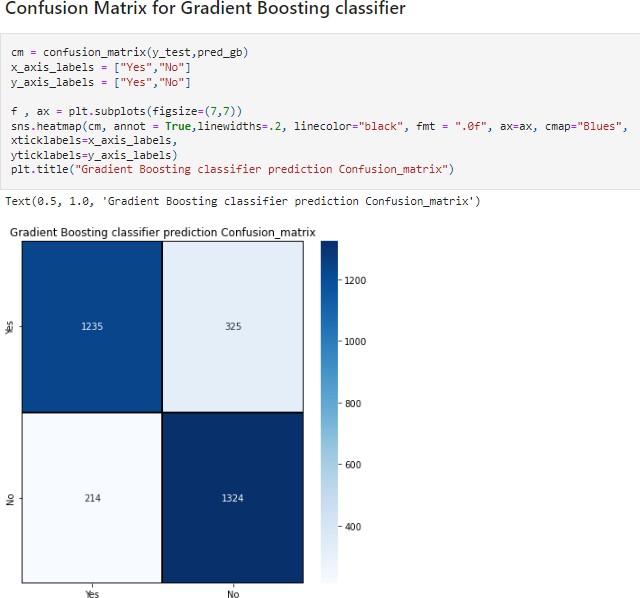


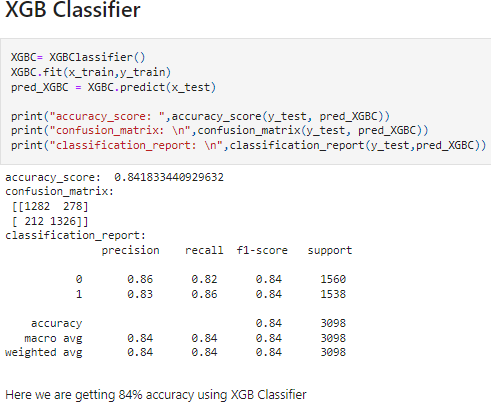


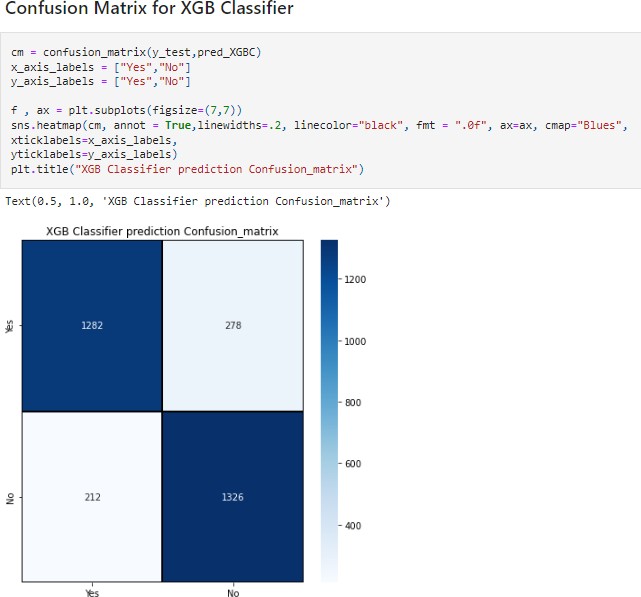












## CrossValidationScoreforallthemodel

Cross-validation is primarily used in applied machine learning to estimate theskillofamachinelearningmodelonunseendata.Thatis,tousealimitedsampleinordertoestimatehowthemodelisexpectedtoperformingeneralwhenusedtomakepredictionson datanotusedduringthetrainingof themodel.

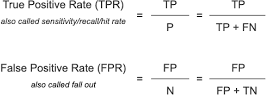


Fromtheobservationofaccuracyandcrossvalidationscoreandtheirdifferencewecan predictthatXGBClassifier isthebestmodel.

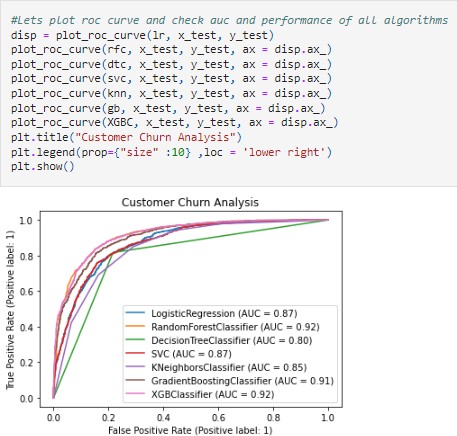
* **ROC&AUCCurvefor allmodel**

An ROC curve (receiver operating characteristic curve) is a graph showing theperformance of a classification model at all classification thresholds. This curveplotstwoparameters:

* TruePositiveRate
* FalsePositiveRate



An ROC curve plots TPR vs. FPR at different classification thresholds. Loweringthe classification threshold classifies more items as positive, thus increasingbothFalsePositives andTruePositives.



## Hyperparametertuning

Models can have many hyperparameters and finding the best combination ofparameters can be treated as a search problem. These parameters expressimportant properties of the model such as its complexity or how fast it shouldlearn.ThetwobeststrategiesforHyperparametertuningare:

* GridSearchCV
* RandomizedSearchCV

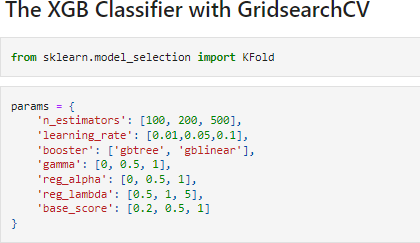
## GridSearchCV

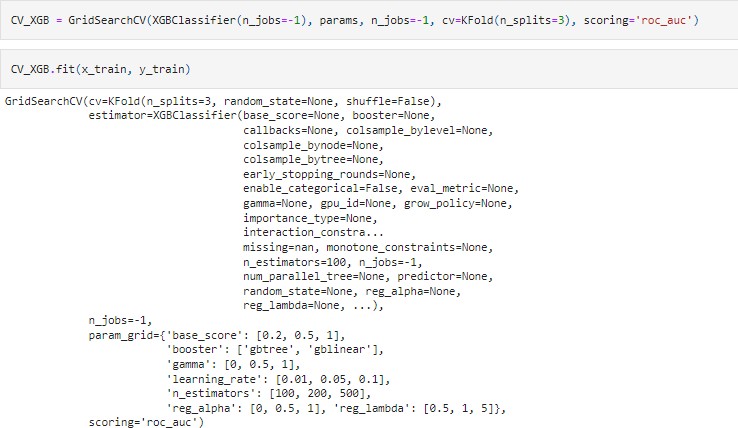
In GridSearchCV approach, the machine learning model is evaluated for a rangeof hyperparameter values. This approach is called GridSearchCV, because itsearches for the best set of hyperparameters from a grid of hyperparametersvalues.Wewillusethishyperparametersfor tuningthedata.

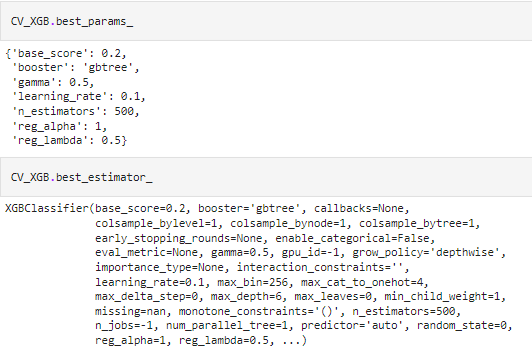
**Drawback of GridSearchCV:** It go through all the intermediate combinations ofhyperparameterswhichmakesgridsearchcomputationallyveryexpensive.

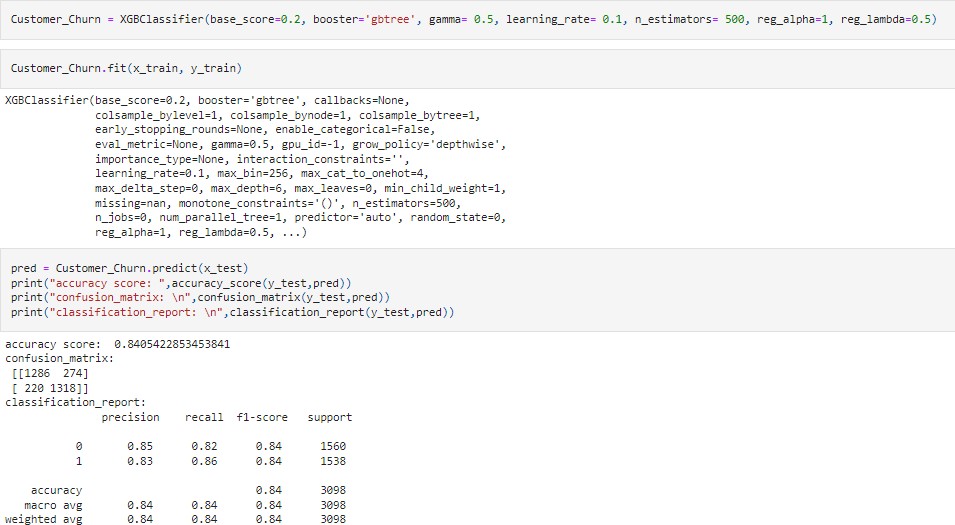
## RandomizedSearchCV

RandomizedSearchCVsolvesthedrawbacksofGridSearchCV,asitgoesthroughonly a fixed number of hyperparameter settings. It moves within the grid in arandom fashion to find the best set of hyperparameters. This approach reducesunnecessarycomputation.



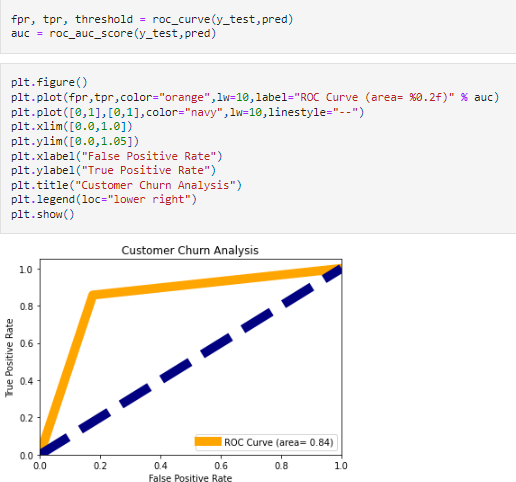






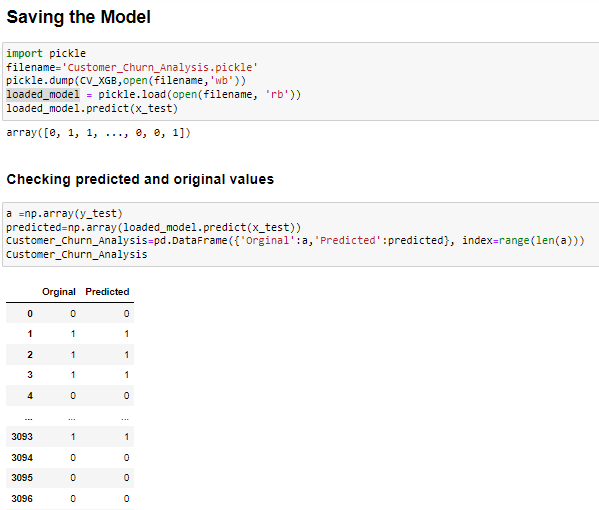


Thefinalmodelgives84%accuracyaftertuningandwillplotROC-AUCcurveforbestmodel.

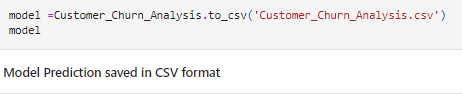


ThisistheAUC-ROCcurveforthemodelswhichisplottedFalsepositiverateagainstTruepositiverate. So,thebestmodelhas theareaundercurveas0.84.

Wegotourbestmodelso willsavethismodelandwilldoPrediction.



Finally,we cansee,predictedandoriginalvalues matches approx100%.So, wewillsavethismodelin csv.



# ConcludingRemarks

Churnrateisanimportantindicatorforcompanies.So,weneedto:

* + Analysechurntoimprovecustomerserviceteam.
  + Revamponboardingplanfornewcustomers.
  + Investinmoretrainingfor supportand salesreps.
  + Askforfeedbackatkeymomentsandrespondpromptly.
  + Communicateproactivelywithcustomers.
  + Offerexclusiveperkstoexistingcustomers.
  + Leveragefeedbackfromfreetrialcustomers.
  + ToReduceCustomerChurn:
    - Focusattentiononbestcustomers.
    - Analysechurnasitoccurs.
    - Showcustomersthatwecare.