Telecom Customer Churn

Ravindra Neralla  
DSC680- Spring 2021

Bellevue University

Milestone-3

Professor: Fadi Alsaleem

**Abstract**

Customer churn is a major problem and one of the most important concern for every company. Retaining an existing customer is many times more effective than gaining new customers. [2] Companies usually have a greater focus on customer acquisition and keep retention as a secondary priority. However, it can cost [five times more to attract a new customer](https://www.invespcro.com/blog/customer-acquisition-retention/) than it does to retain an existing one. Increasing customer retention rates by 5% can increase profits by 25% to 95%, according to [research done by Bain & Company](http://www2.bain.com/Images/BB_Prescription_cutting_costs.pdf). Generally, people only switch to a different company only when the service is not to the level of expectation in one or the other factor when compared to competitors, more than getting attracted to the specials and offers which are offered by the other companies. [1] It is pretty common to hear people express frustration with some aspect of their communications provider — be it convoluted billing, unwanted marketing emails, hard-to-navigate customer service or high plan prices. As the number of peoples who use a phone or a telecom product does not increase in huge volume, any company looking to improve the profits is trying to attract other company customers. At the same time retaining existing customers is very important. As retaining a current customer is ten times more productive than gaining a new one.

**Objective:**

As customer churn directly effects the revenues of the companies, companies are seeking to develop means to predict potential customer to churn. Therefore, finding factors that increase customer churn is important to take necessary actions to reduce this churn and retain the customers.

**Data:**

The dataset, I am planning to work on in this project is [3] [Telco Customer Churn](https://www.kaggle.com/blastchar/telco-customer-churn). Dataset consists of 21 variables with 7043 observations. Each Observation in the dataset represents a customer. Below is the detailed list of variables.

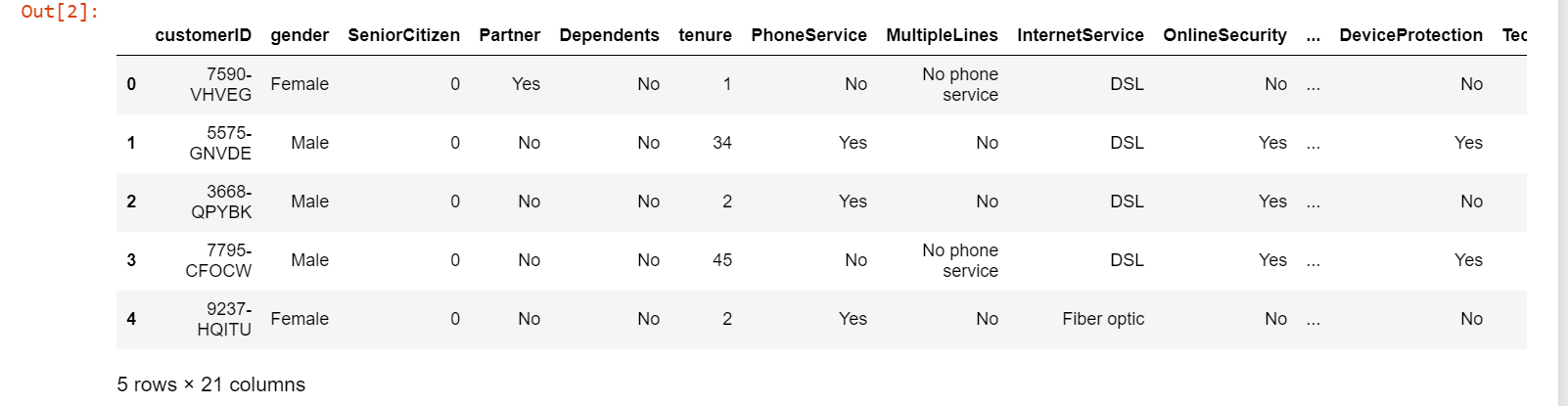
1. customerID
2. gender
3. SeniorCitizen
4. Partner
5. Dependents
6. tenure
7. PhoneService
8. MultipleLines
9. InternetService
10. OnlineSecurity
11. OnlineBackup
12. DeviceProtection
13. TechSupport
14. StreamingTV
15. StreamingMovies
16. Contract
17. PaperlessBilling
18. PaymentMethod
19. MonthlyCharges
20. TotalCharges
21. Churn

#### Method

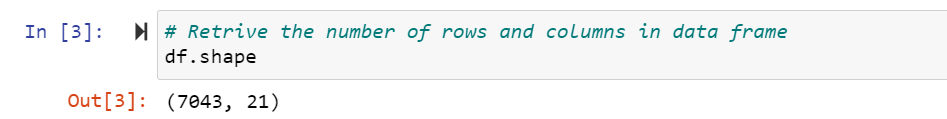
Data Analysis:

Customer churn data is loaded into a data frame using python and performed initial analysis to check if any duplicate observations or null values present in the dataset. Generate multiple charts on the variables to see the spread of the values in the variables. Perform Univariant analysis by using histograms and box plots. Perform Bivariant analysis by using correlation matrix and heat maps.

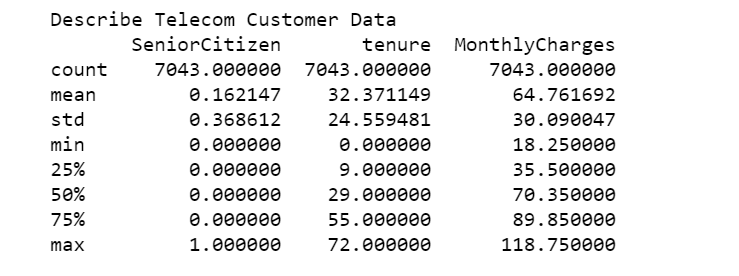
* 1. Load data into data frame:



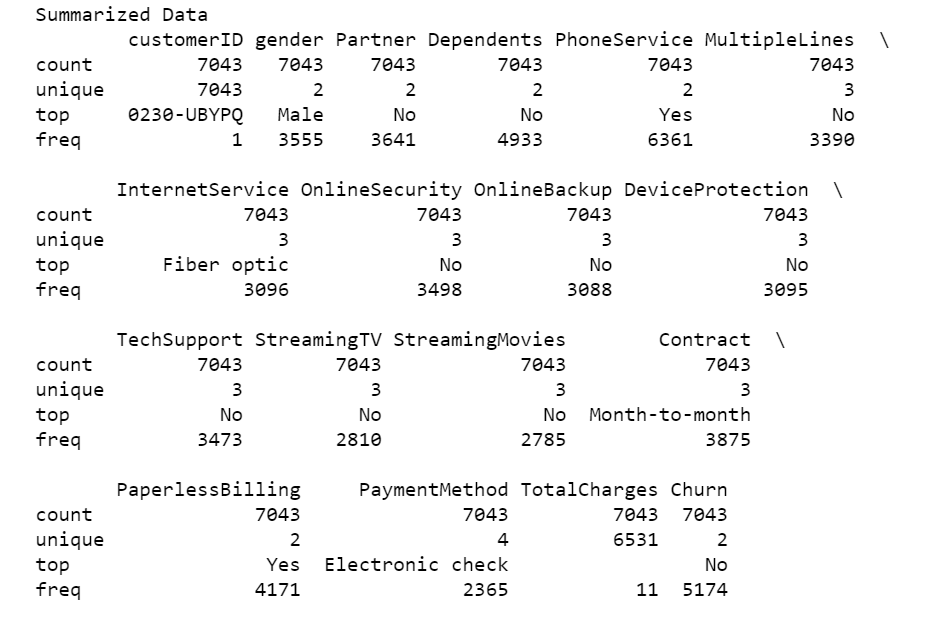
* 1. Verify the number of variables and observations:



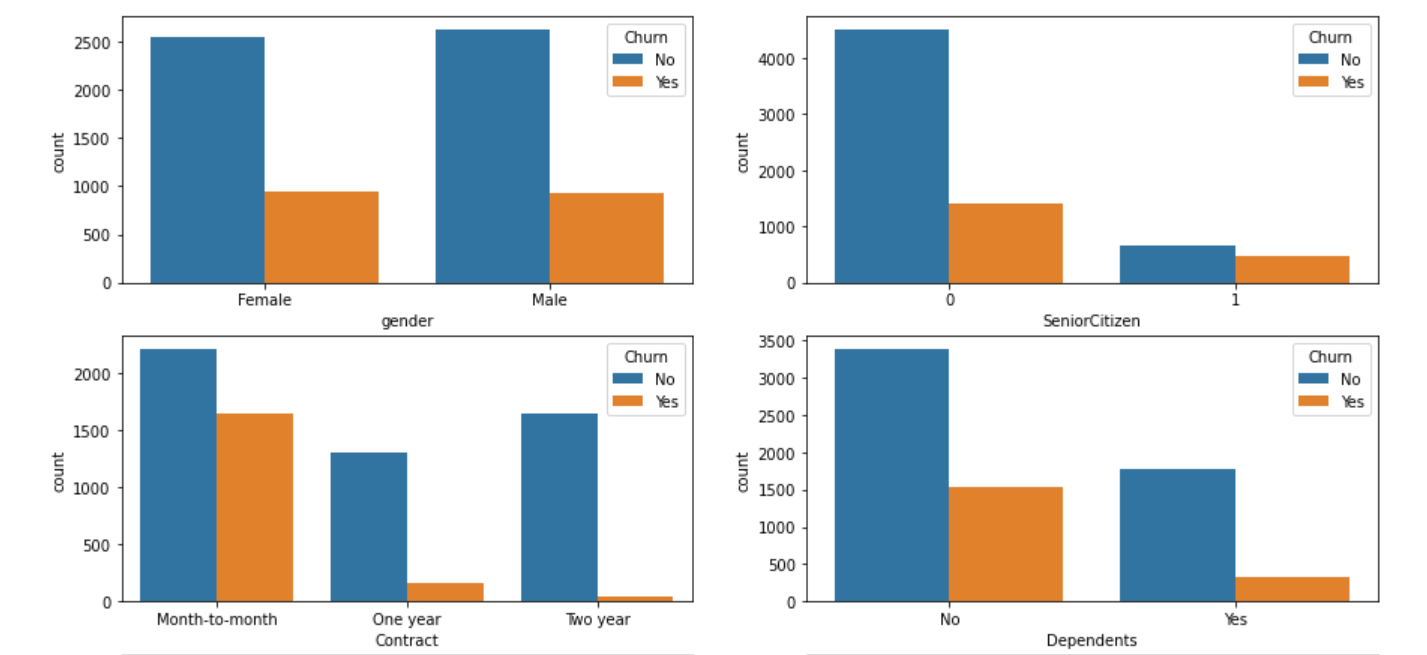
* 1. Get the stats on numerical variables:

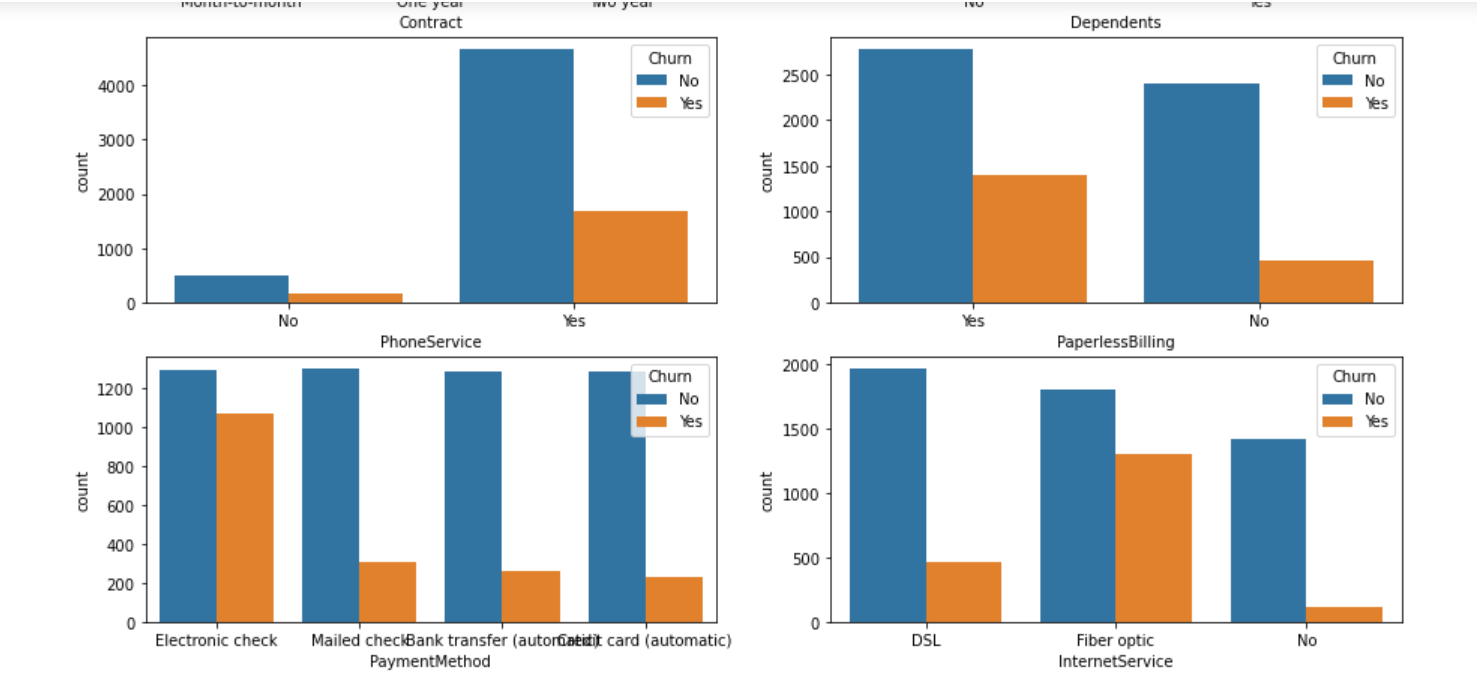


* 1. Get the stats on categorical variables:



g) Generate Charts on Churn vs Categorical Variables:





**Observations:**

Gender vs Customer\_Churn:

We do not see any difference in Male vs Female customers in terms of Customer Churn.

Contract\_Type vs Customer\_Churn :

'Month-on-month' type Contract has highest Customer Churn compared to other contract Types.

Payment\_Method vs Customer\_Churn :

'Electronic Check' payment method has the highest Customer Churn.

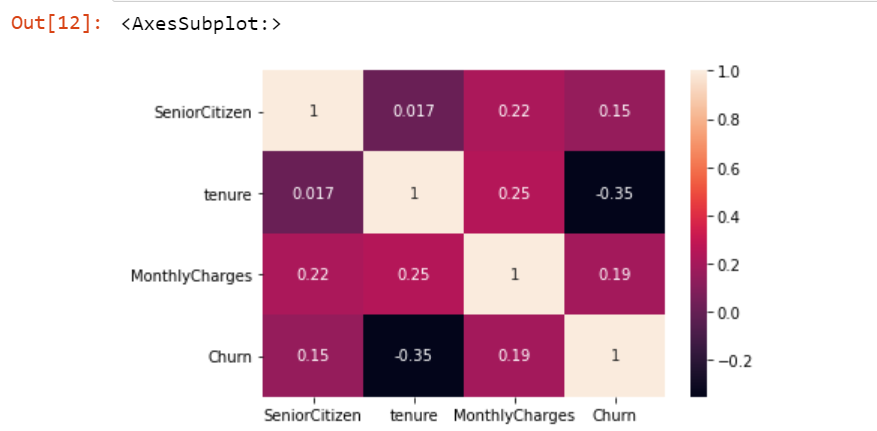
Paperless\_Billing vs Customer\_Churn : 'Paperless Billing' has highest Customer Churn.

Type\_Of\_Internet\_Service vs Customer\_Churn : 'Fiber optic' Internet service has highest Customer Churn.

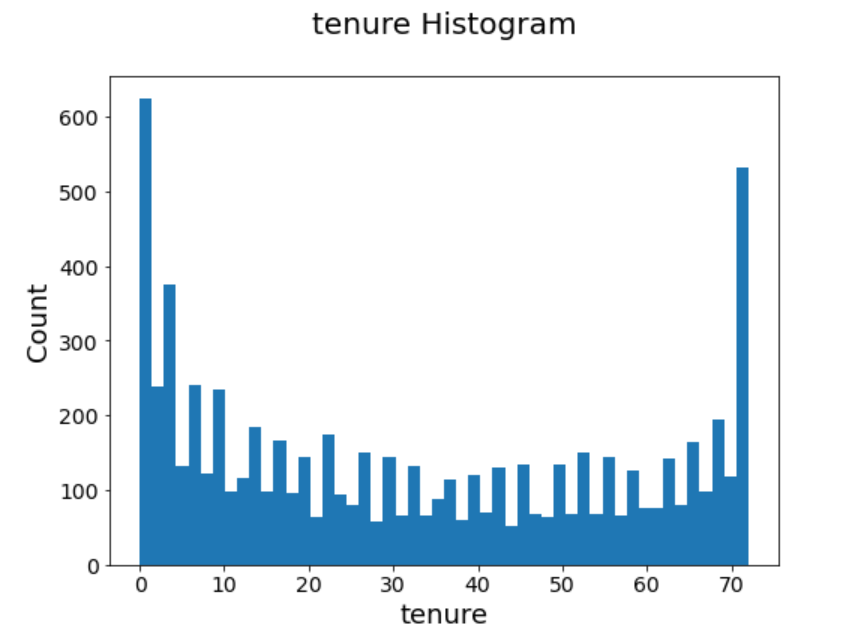
Phone\_Service vs Customer\_Churn : Customer who has Phone Service has highest Customer Churn.

h) Heatmap Analysis:

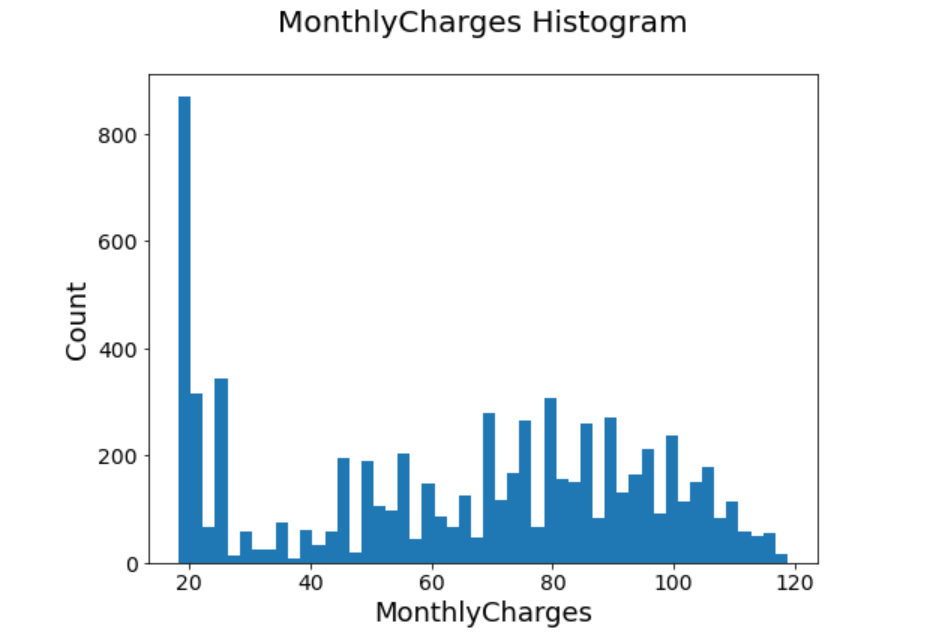
From the below heatmap we can observe that tenure and monthly variables are better correlated with churn. So, a histogram is built to understand the spread of tenure data.



Tenure Histogram:



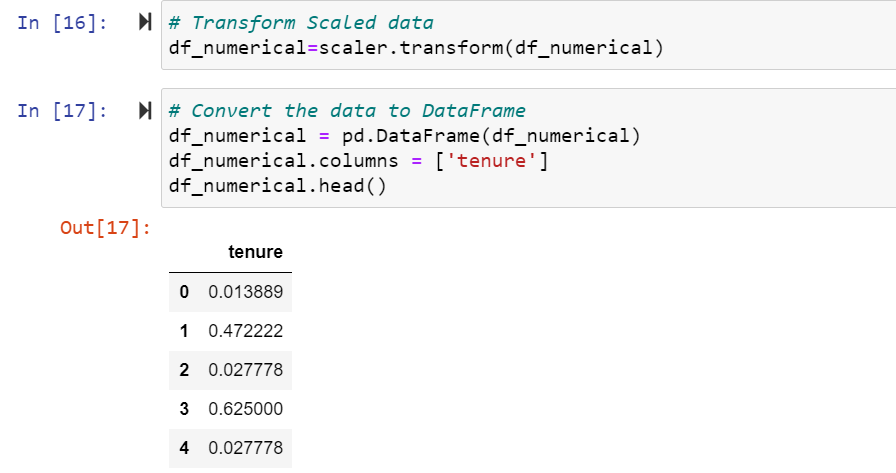
Monthly Charges Histogram:



* + 1. Encoding Method:

Applied Encoding method to convert Categorical variables to numerical values.

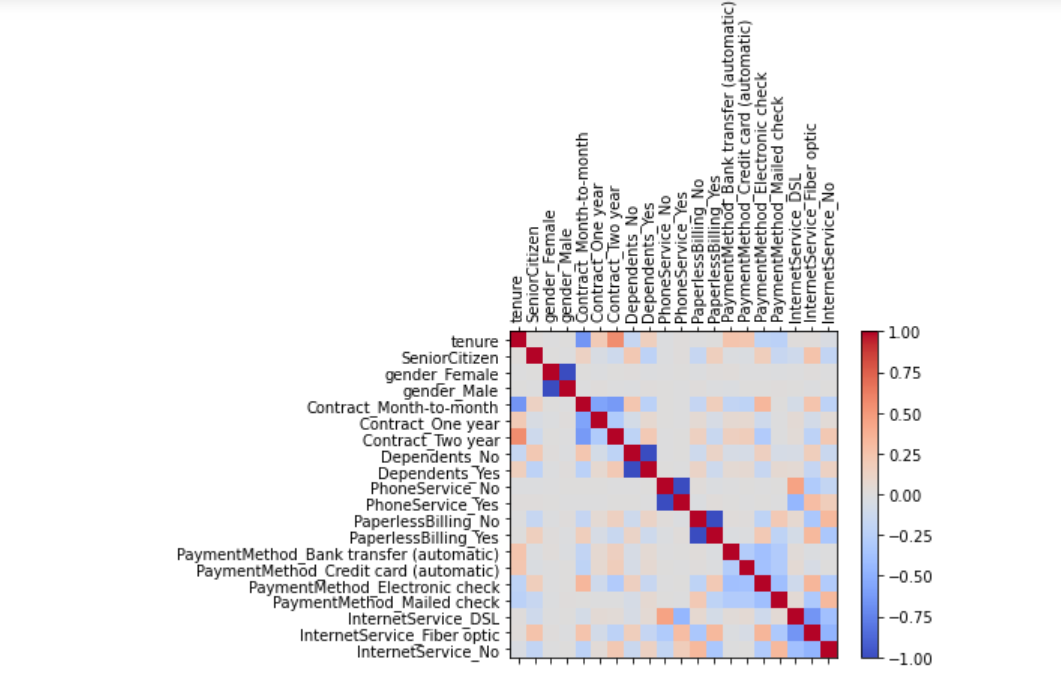
Applied Scalar method to convert the Numerical variable to reduce the value and maintain consistent value range, i.e., Tenure in months value ranging from 10 to 70 by applying encoding, values are in consistent range.

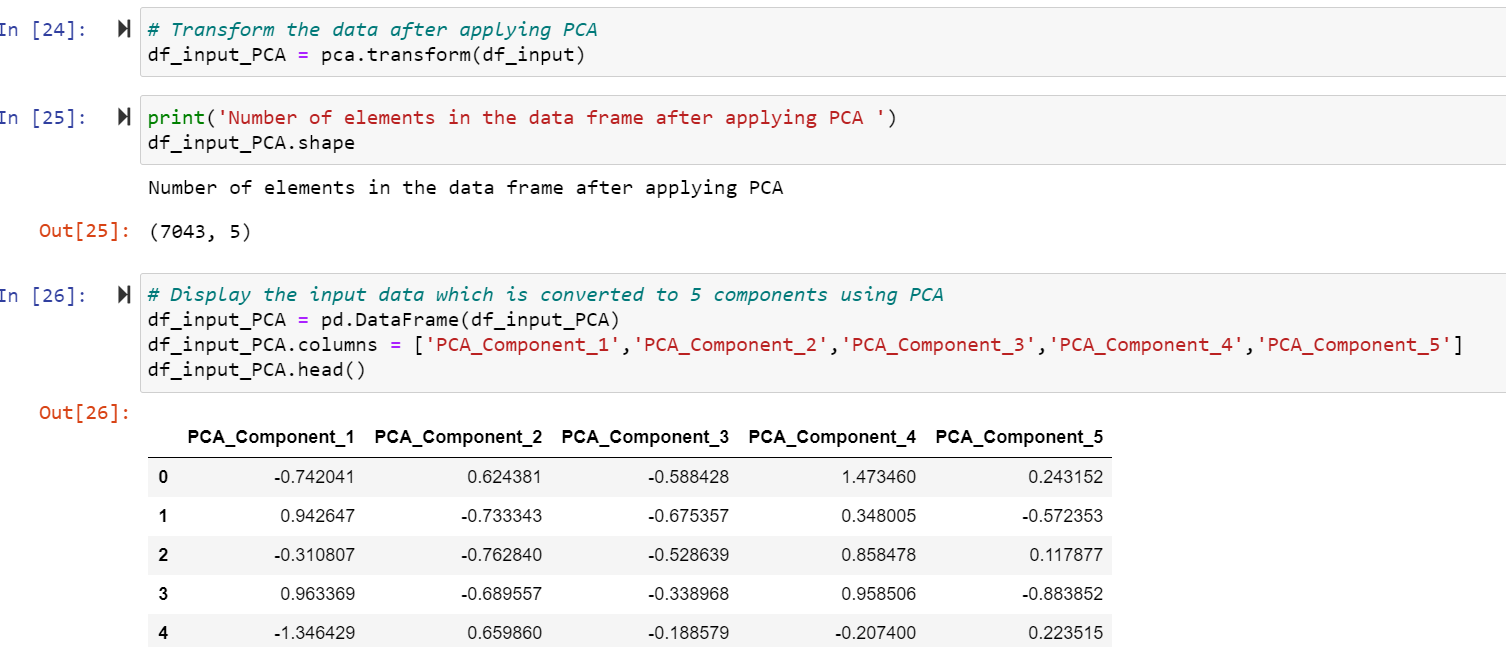




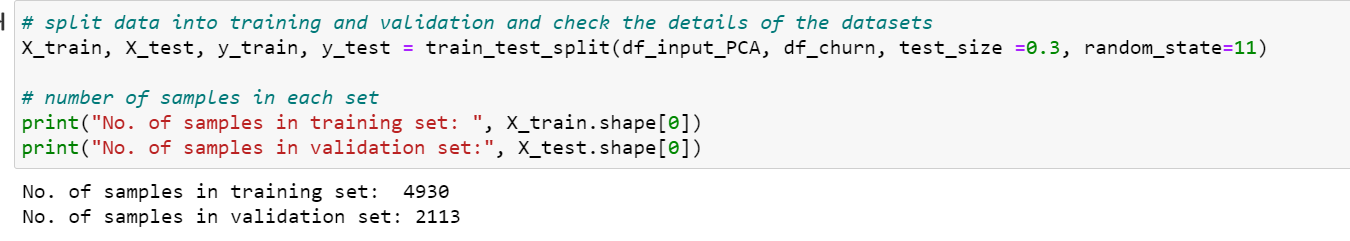
j) Dimensionality Reduction:

After converting categorical variables to numeric values, the number of input variables have increased. So, dimensionality reduction was performed by applying Principal Component Analysis [4], reducing the number of variables to 5.



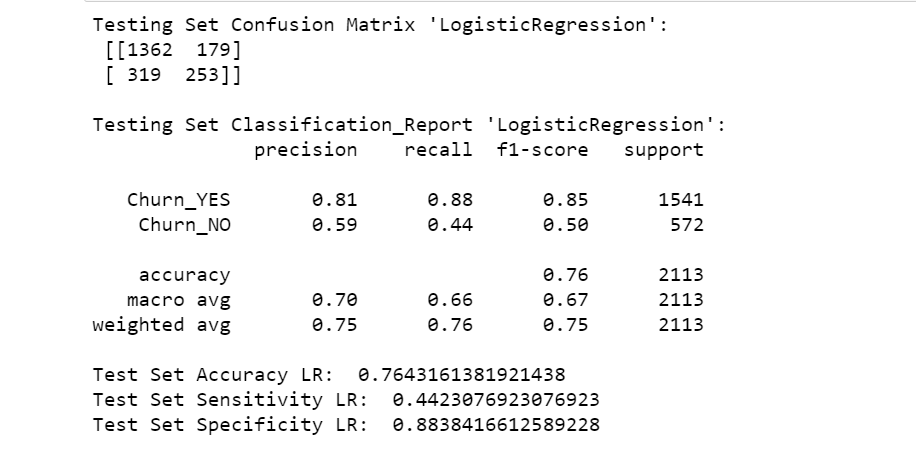


k) Split the data:

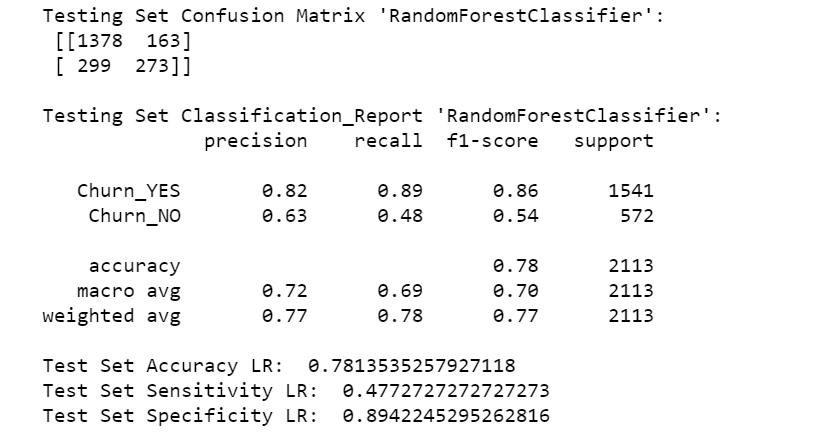


Modeling:

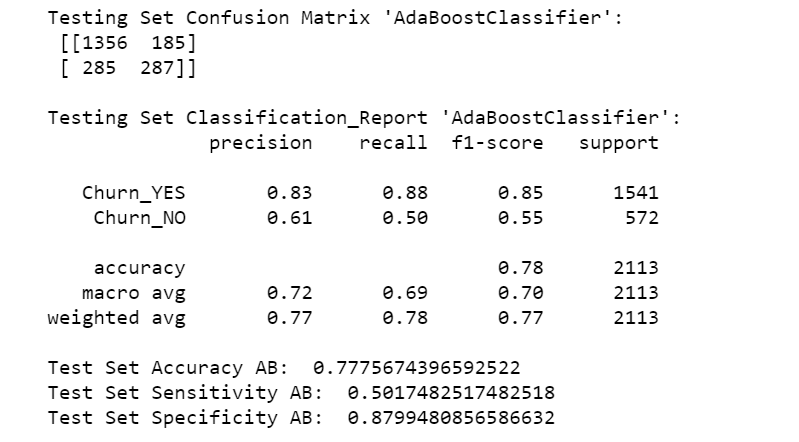
A baseline model was established by training a **Logistic Regression Model**, after splitting the data into test and train in the ratio of 70:30



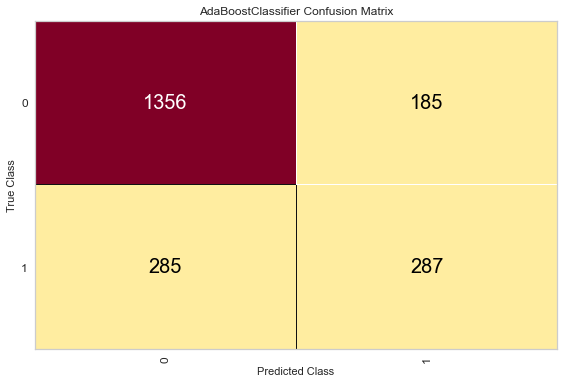
**Random Forest Classifier:**

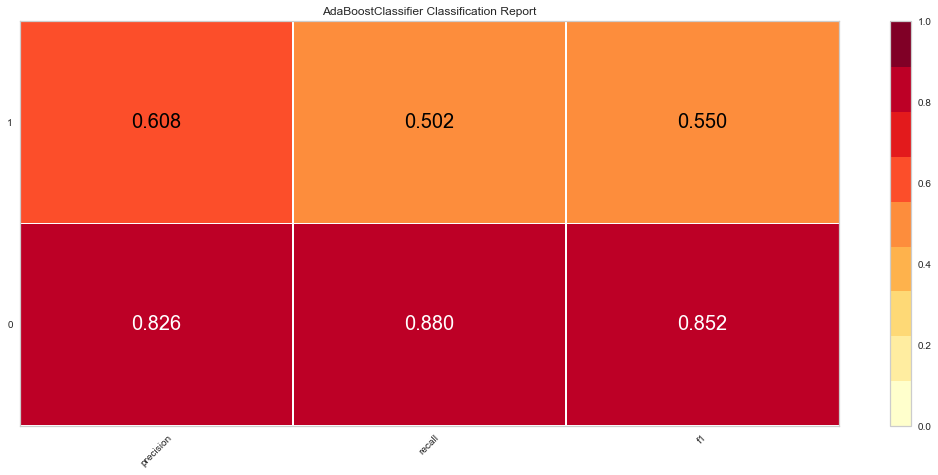


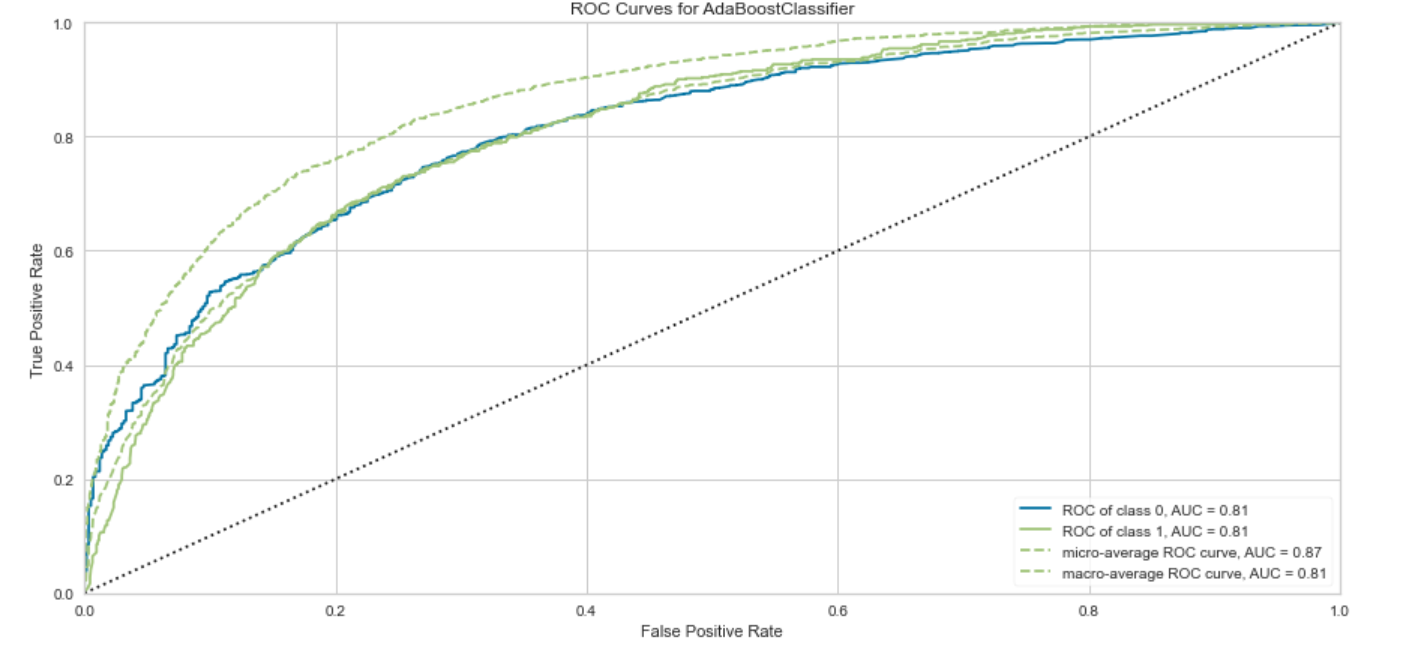
**AdaBoost:**



**Results:**

****

****



**Conclusion:**

Logistic Regression Model gave accuracy of 76% and Random Forest and ADA Boost gave 78% accuracy. Area Under the Curve (AUC) is 0.81, which indicates that the model is reliable and is accurately working in 81% of cases.

References:

[1]: [How Costly Is Customer Churn in the Telecom Industry? - EBR](https://www.citationmachine.net/apa/cite-a-website/search?q=https%3A%2F%2Fwww.europeanbusinessreview.com%2Fhow-costly-is-customer-churn-in-the-telecom-industry%2F%3F__cf_chl_jschl_tk__%3Dec58e0210ef4250a8414742b7aaa8d8e28384753-1611086044-0-AXu6Mldw8Cpmi6L6s5SgV1PXJkIGPIbNtM2k3mk8dG4GwrfiMZ4fT3rVf0CPVbws4ipH6NuFN2fITmeBiQb8uSJYkXQKBcXB-e3XBH6IJWYli7zOM4eqnO9BCh0JE11v9IHHQZeKKMNRTUecYHq5Mahj6uNBBJ23n_LWFAdEF677ZQEQpPYruho0OLOLbdjBj-oKjPGHAZ52v2QMgPAHjyXRIDb-FEUu_sFhz8CW6q0KxFhOgoPWW_R8KZhMo4AOUlv2fbzwIi3oYhYXMSMOdAjpE9u7gRwD-0RQeFCEtTEzQqRmjC2QEbIGAHZKgbkn99TTuu5kBJLOuOoRLFhaD9d-oHKuQ27yx7WrU_mDPxYdp7qkByeE3o25LzRg460JyA)

[2]: Customer Churn in Telecom Segment - Curi

<https://towardsdatascience.com/customer-churn-in-telecom-segment-5e49356f39e5>

[3]: Telco Customer Churn - BlastChar

<https://www.kaggle.com/blastchar/telco-customer-churn>

[4]: PCA using Python (scikit-learn) - Galarnyk

<https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60>

[5]: Churn Reduction in Telecom Industry – Arthur Middleton Hughes

<http://www.dbmarketing.com/telecom/churnreduction.html>