Churn Prediction (Harvard PH125.9)

Humayun Akram

09 March 2019

1. Introduction

1.1 Overview

A common expression in industry is that it costs five times more to acquire a new customer than it does to retain an existing one (Kirui et al., 2013). In sectors ranging from finance and telecom to cable service and insurance, significant sums of money are spent in reducing customer attrition. To prevent such attrition (churn), it is critical to be able to identify the early warning signs of churn (Van den Poel & Larivie 2004; Larivie`re & Van den Poel 2005).

Thus, a key part of any business these days is to manage customer churn; that is, to avoid losing its customers and build its customer base (Mattison 2005; Tsai & Lu 2010; Nie et al., 2011). A common approach to combat customer churn relies on identifying customers who have a high likelihood of churning. These customers are then given incentives or a customized service plan in order to convince them to stay (Ngai et al., 2009; Khan et al., 2010).

As part of second project in Harvard PH125.9 Capstone course, we will be building churn prediction models to address this classifcation problem. These models will predict potential churn customers for a sample of users based on training dataset. At the end, we will verify the performance of our models using Sensitivity, Accuracy and AUC.

1.2 DataSet

This project is based on sample data from IBM Sample DataSets for customer retention programs which will be used to create churn prediction models. The dataset is available at below location:

- [Telco CHurn dataset] https://community.watsonanalytics.com/wp-content/uploads/2015/03/WA_Fn-UseC_-Telco-Customer-Churn.csv?
cm_mc_uid=58920755505115141495567&cm_mc_sid_50200000=1514149556&cm_mc_sid_52640000=1514149556
(https://community.watsonanalytics.com/wp-content/uploads/2015/03/WA_Fn-UseC_-Telco-Customer-Churn.csv?
cm_mc_uid=58920755505115141495567&cm_mc_sid_50200000=1514149556&cm_mc_sid_52640000=1514149556)"

The dataset comprised of multiple rows/columns where each row represents a customer while each column contains a customer's attribute.

1.3 Target

The target of this project is to predict behavior of customer attrition i.e. whether customer will churn or not churn, based on churn prediction model which will, in turn, help to improve customer loyalty and retention.

As the focus of this project is to improve customer loyalty and retention, we will focus majorly on **Sensitivity** (Total Postive Rate) metric in addition to AUC & Accuracy. On the other hand, an average value of **Specificity** (True Negative Rate) will not cause any negative problem since the goal is to engage and communicate with the customers to prevent them from churning. Thus, its acceptable to engage with those who are mistakenly tagged as 'not churned'.

1.4 Key Steps

In this project, we will use the above mnetioned telco churn dataset. Major steps for this project is detailed as below:

- Load the data and do initial exploration
- Insight analysis and vizualization
- Data Cleansing & Preprocessing
- Build 5 models based on the churn training dataset
- Validate the models performance for test dataset.

2. Analysis

2.1 Data Loading & Exploration

We will start off by loading the data code provided by the IBM smaple churn dataset.

2.1.1 Data Check

Once loaded, we can see that churn dataset has **7043** rows and **21** columns.

Let's have an overview of churn dataset and verify if it has any missing/null values

```
## [1] 7043   21
```

```
## [1] TRUE
```

```
##    customerID            gender         SeniorCitizen       Partner
## Length:7043        Length:7043       Min.   :0.0000   Length:7043
## Class :character   Class :character  1st Qu.:0.0000   Class :character
## Mode  :character   Mode  :character  Median :0.0000   Mode  :character
##                                      Mean   :0.1621
##                                      3rd Qu.:0.0000
##                                      Max.   :1.0000
##
##    Dependents            tenure       PhoneService      MultipleLines
## Length:7043        Min.   : 0.00   Length:7043       Length:7043
## Class :character   1st Qu.: 9.00   Class :character  Class :character
## Mode  :character   Median :29.00   Mode  :character  Mode  :character
##                    Mean   :32.37
##                    3rd Qu.:55.00
##                    Max.   :72.00
##
## InternetService    OnlineSecurity    OnlineBackup
## Length:7043        Length:7043       Length:7043
## Class :character   Class :character  Class :character
## Mode  :character   Mode  :character  Mode  :character
##
##
##
##
## DeviceProtection   TechSupport       StreamingTV
## Length:7043        Length:7043       Length:7043
## Class :character   Class :character  Class :character
## Mode  :character   Mode  :character  Mode  :character
##
##
##
##
## StreamingMovies      Contract        PaperlessBilling
## Length:7043        Length:7043       Length:7043
## Class :character   Class :character  Class :character
## Mode  :character   Mode  :character  Mode  :character
##
##
##
##
## PaymentMethod      MonthlyCharges    TotalCharges         Churn
## Length:7043        Min.   : 18.25   Min.   :  18.8   Length:7043
## Class :character   1st Qu.: 35.50   1st Qu.: 401.4   Class :character
## Mode  :character   Median : 70.35   Median :1397.5   Mode  :character
##                    Mean   : 64.76   Mean   :2283.3
##                    3rd Qu.: 89.85   3rd Qu.:3794.7
##                    Max.   :118.75   Max.   :8684.8
##                                     NA's   :11
```

It can be seen from summary that TotalCharges have 11 NAs.However, being very low count, we can remove these rows from the dataset.

```
##           Attributes unique_values
## 1        customerID          7032
## 2      TotalCharges          6530
## 3    MonthlyCharges          1584
## 4            tenure            72
## 5     PaymentMethod             4
## 6     MultipleLines             3
## 7   InternetService             3
## 8    OnlineSecurity             3
## 9      OnlineBackup             3
## 10 DeviceProtection             3
## 11      TechSupport             3
## 12      StreamingTV             3
## 13  StreamingMovies             3
## 14         Contract             3
## 15           gender             2
## 16    SeniorCitizen             2
## 17          Partner             2
## 18       Dependents             2
## 19     PhoneService             2
## 20 PaperlessBilling             2
## 21            Churn             2
```

Next up, we can see from dataset that there are 3 contnuous features while rest are all categorical attributes.

2.1.2 Data Summarization

Summarization shows that Monthly Charges have pretty much uniform distribution with Median value at 70. Hoewver, TotalCharges distribution appeared to be highly skewed.

An overview of variables are given as below:

```
## Observations: 7,032
## Variables: 21
## $ customerID       <chr> "7590-VHVEG", "5575-GNVDE", "3668-QPYBK", "77...
## $ gender           <chr> "Female", "Male", "Male", "Male", "Female", "...
## $ SeniorCitizen    <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ Partner          <chr> "Yes", "No", "No", "No", "No", "No", "No", "N...
## $ Dependents       <chr> "No", "No", "No", "No", "No", "No", "Yes", "N...
## $ tenure           <int> 1, 34, 2, 45, 2, 8, 22, 10, 28, 62, 13, 16, 5...
## $ PhoneService     <chr> "No", "Yes", "Yes", "No", "Yes", "Yes", "Yes"...
## $ MultipleLines    <chr> "No phone service", "No", "No", "No phone ser...
## $ InternetService  <chr> "DSL", "DSL", "DSL", "DSL", "Fiber optic", "F...
## $ OnlineSecurity   <chr> "No", "Yes", "Yes", "Yes", "No", "No", "No", ...
## $ OnlineBackup     <chr> "Yes", "No", "Yes", "No", "No", "No", "Yes", ...
## $ DeviceProtection <chr> "No", "Yes", "No", "Yes", "No", "Yes", "No", ...
## $ TechSupport      <chr> "No", "No", "No", "Yes", "No", "No", "No", "N...
## $ StreamingTV      <chr> "No", "No", "No", "No", "No", "Yes", "Yes", "...
## $ StreamingMovies  <chr> "No", "No", "No", "No", "No", "Yes", "No", "N...
## $ Contract         <chr> "Month-to-month", "One year", "Month-to-month...
## $ PaperlessBilling <chr> "Yes", "No", "Yes", "No", "Yes", "Yes", "Yes"...
## $ PaymentMethod    <chr> "Electronic check", "Mailed check", "Mailed c...
## $ MonthlyCharges   <dbl> 29.85, 56.95, 53.85, 42.30, 70.70, 99.65, 89....
## $ TotalCharges     <dbl> 29.85, 1889.50, 108.15, 1840.75, 151.65, 820....
## $ Churn            <chr> "No", "No", "Yes", "No", "Yes", "Yes", "No", ...
```

2.2 Insights Analysis & Vizualization

2.2.1 Churn Customers Distribution

From Churn column, we can see the customers who have churned and it's showing us that ~26% of customers have churned based on given dataset.
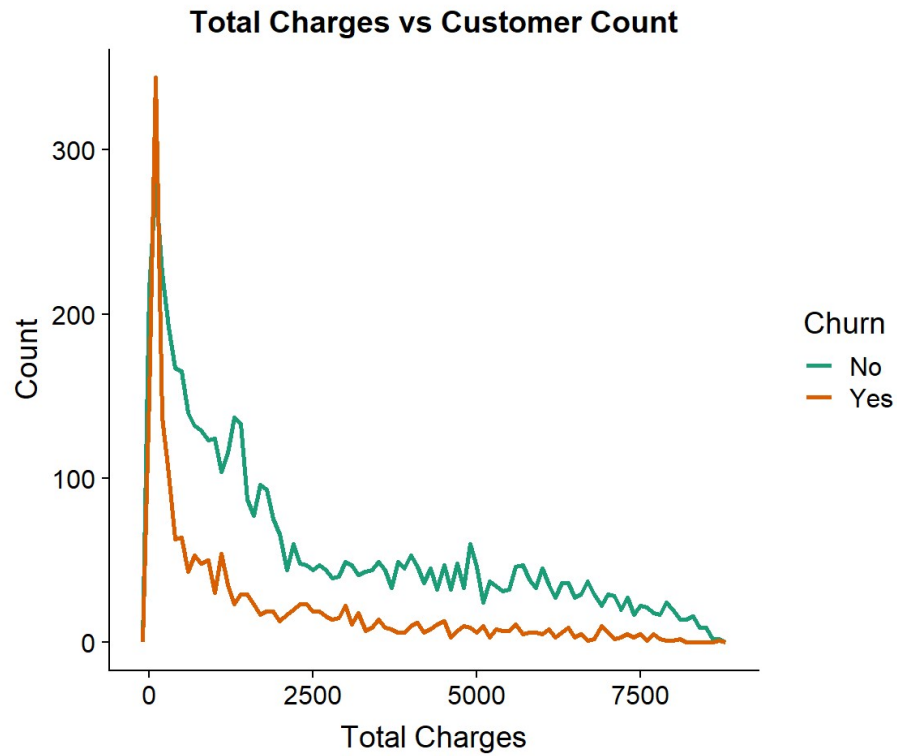
Distribution of Churn Customers
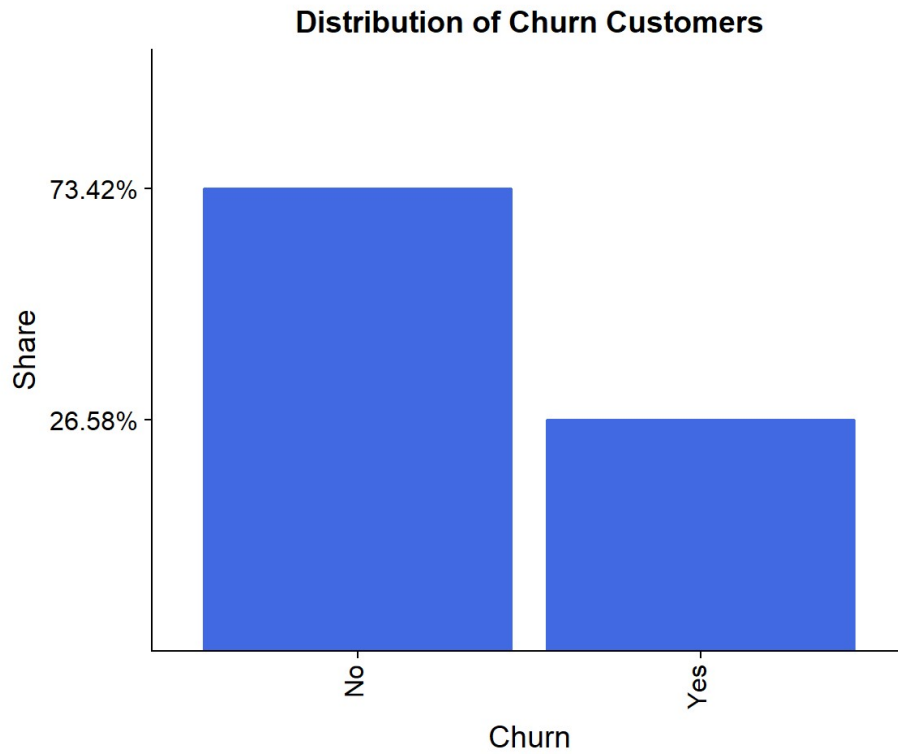

2.2.2 Continuous Features Distribution

Starting our data exploration with continuous features, we can see a clear diference between churn and non-churn segment for MonthlyCharges below $25 as churn customers don't have much presence in that section of MonthlyCharges. For rest of bins, the count is pretty much same between churn and non-churn customers.

Monthly Charges vs Customer Count
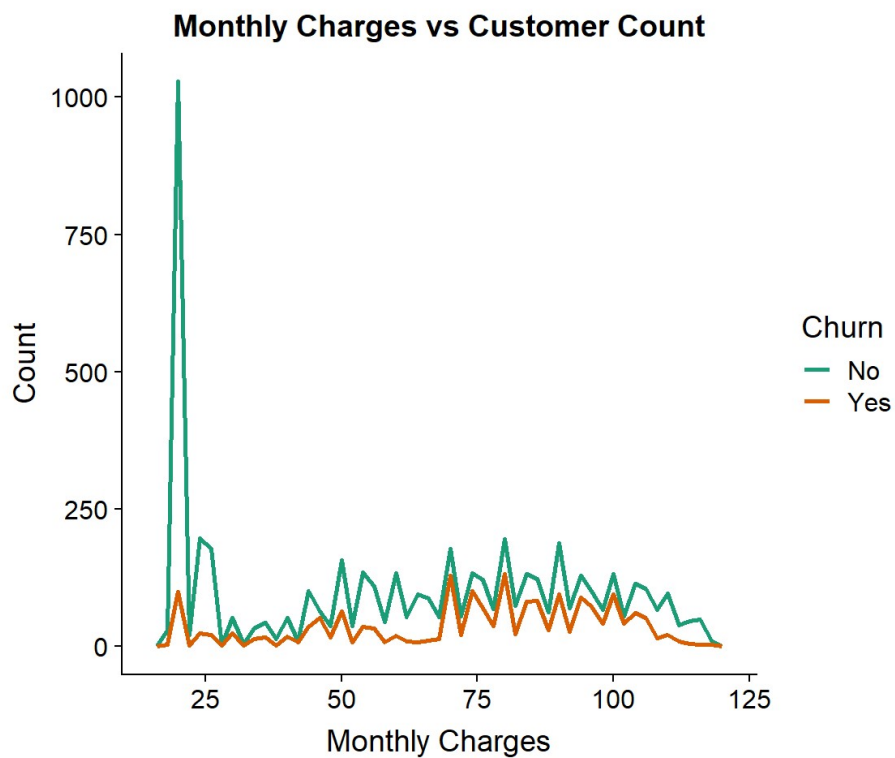
For TotalCharges, the distribution is almost same for churn and non-churn segment as it's right skewed for both the cases.

Total Charges vs Customer Count

For tenure, we can see a very interesting distribution. For new joiners, we can see a high count for both churn and non-churn customers, indicating that chances of attrition is high for new-joiners (which is typical for many service industries). However, the second spike can be seen in >60 months segment where majority of nonchurn chustomers are present indicating that service provider has pretty much loyal customer base.

Tenure vs Customer Count
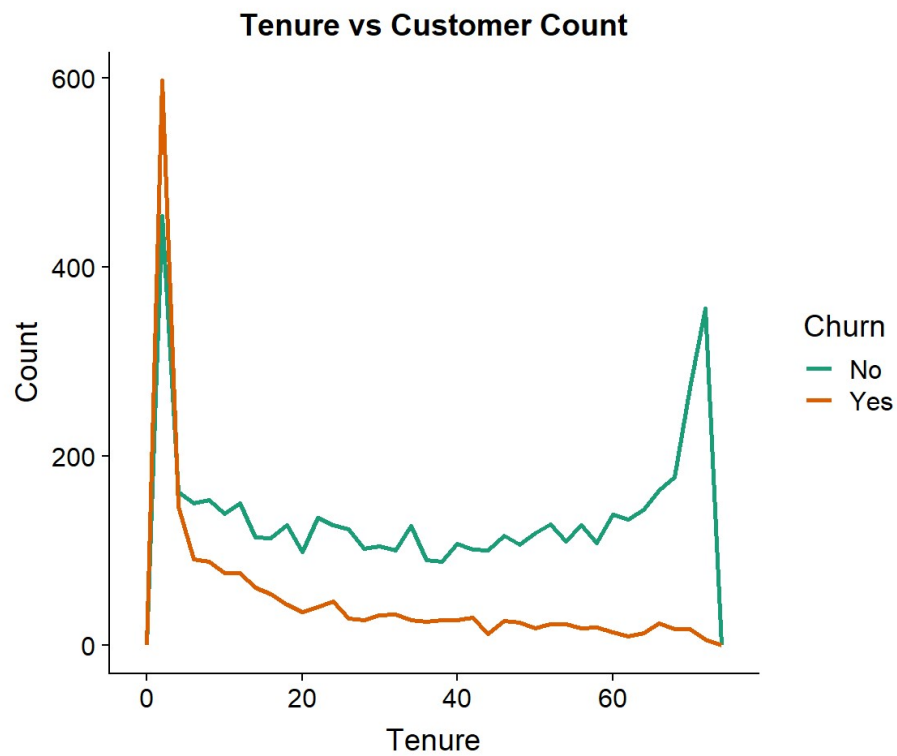
2.2.3 Data Correlation

The correlation plot shows TotalCharges are positively correlated with Tenure & MonthlyCharges

Correlation
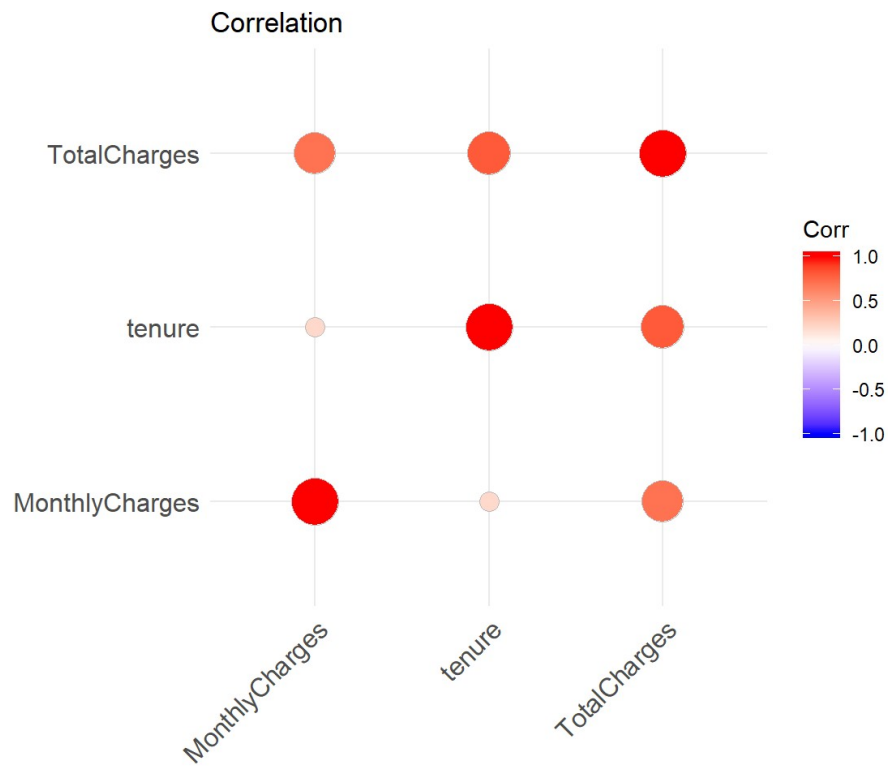
2.2.4 Categorical Features Distribution

Moving on to categorical features, we can see that for certain variables like Gender, PhoneService & MultipleLines have almost no impact on churn rate. However, high churn can be seen for case of senior citizens. Similarly, it can also be observed that customers with Partners & Dependents are less prone to churn. Thus, Partners & Dependants have significance in terms of customers loyalty.
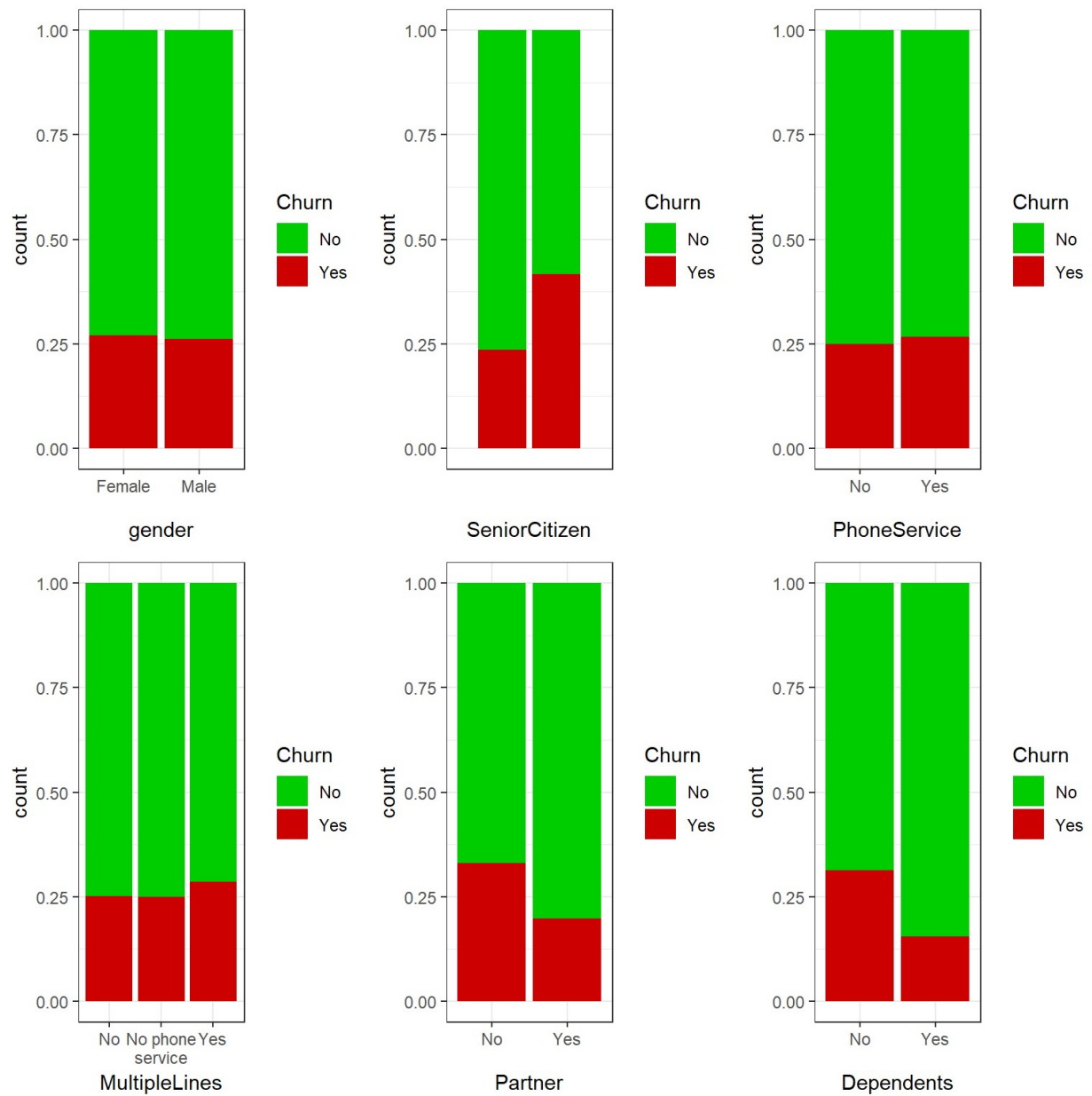
For service attributes, we can see that Fiber Optic for InternetServices is an important feature as churn rate is very high for these customers.Further, customers with "No" for services like OnlineSecurity & DeviceProtection are more prone to churn.
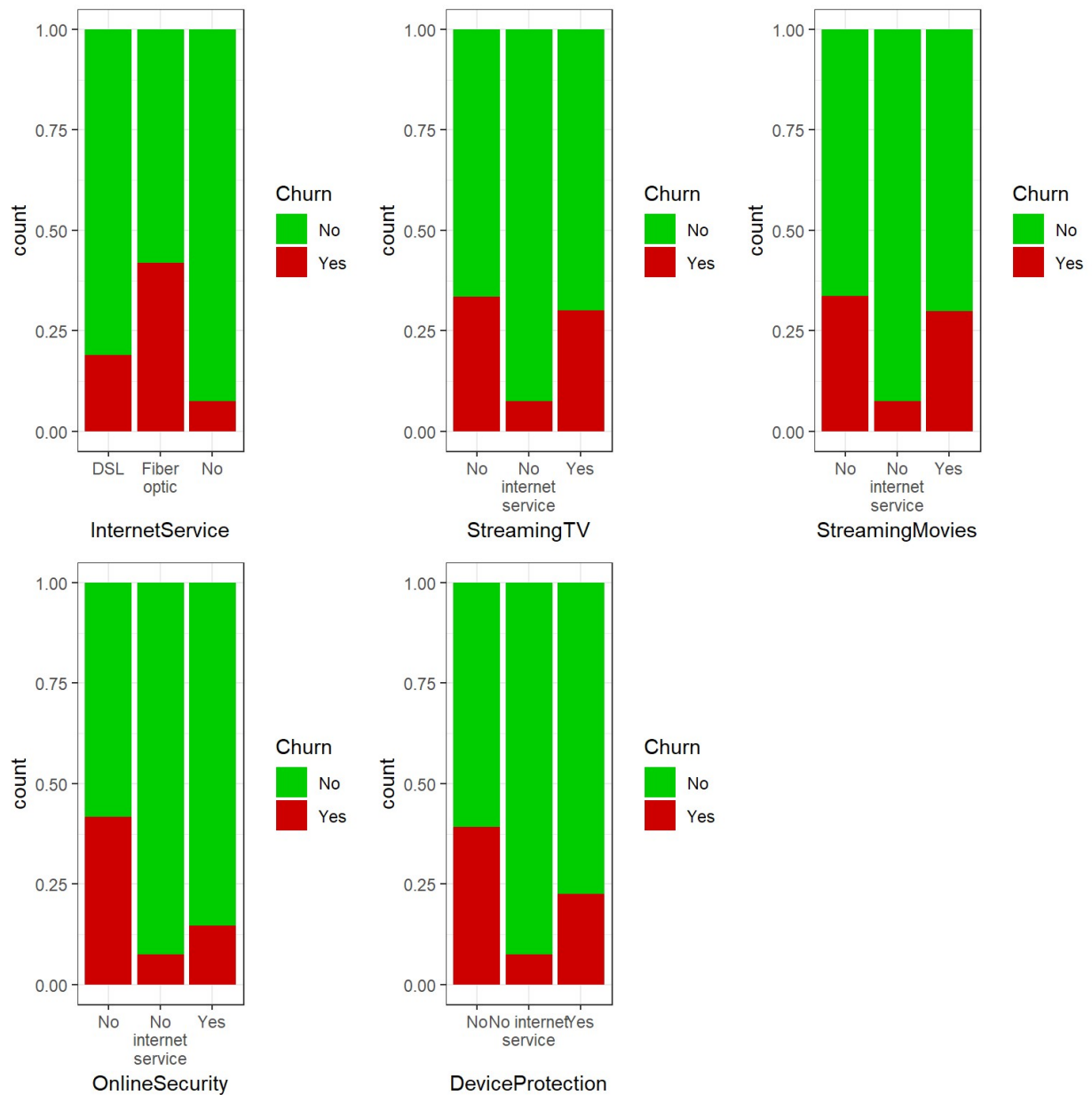
For contract feature, a clear distinction can be seen between short term & long term contracts i.e. monthly subscription customers have much higher churn rate as compared to yearly subscriptions. Secondly, customers with "No" for services like TechSupport & OnlineBackup are also more prone to churn. Finally, we can see that customers with paperless billing & with ElectronicCheck Payment Method tend to have a high churn rate as compared to others.
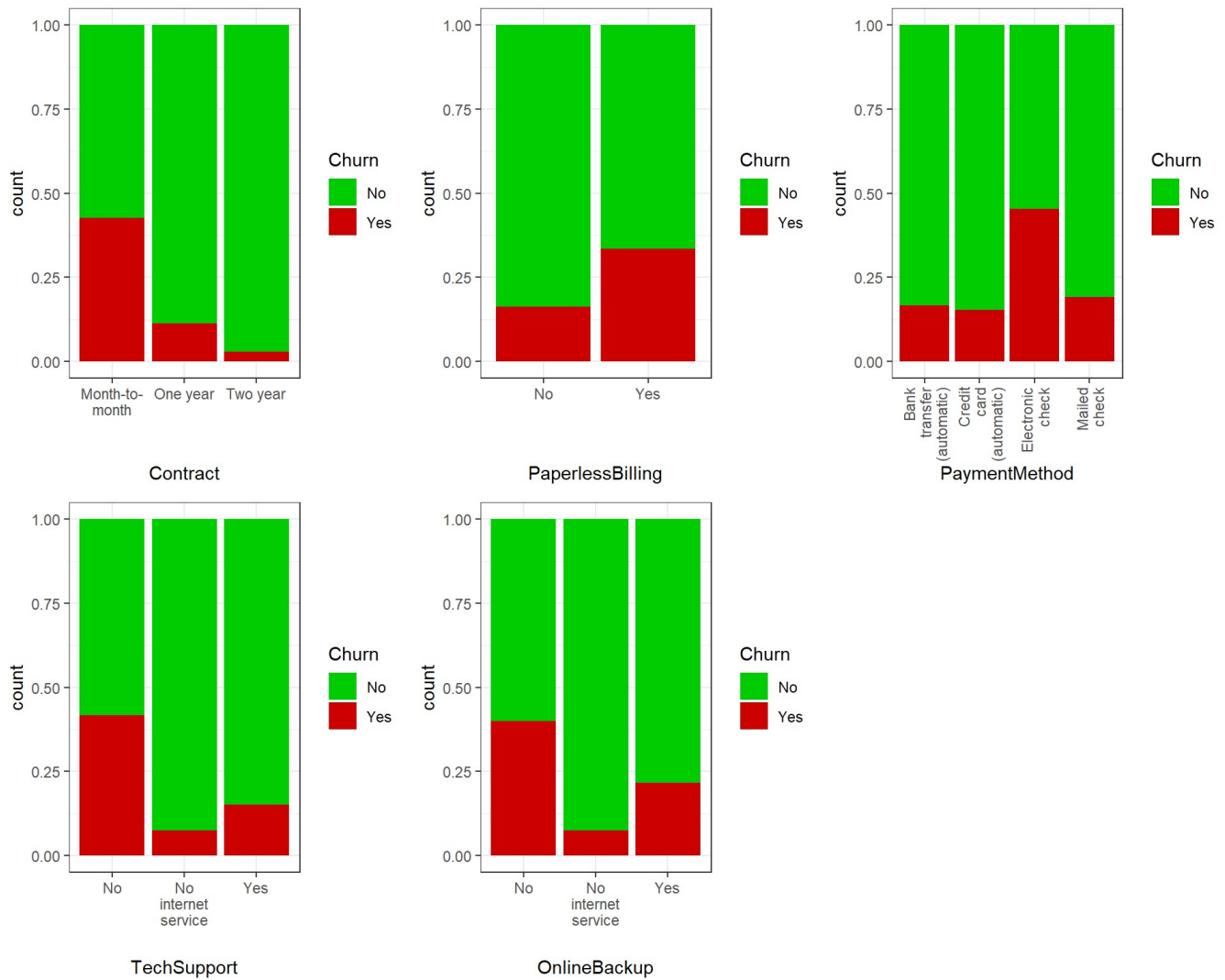
2.3 **Data Preprocessing**

Before Modeling, we will perform data preprocessing so that it can be readily used.

- Firstly, since Customer ID has no impact on churn, we can remove it from the dataset

```
####################################################################
######################### Data Proprocessing ######################
####################################################################

# Remove Unwanted Variables
churn_dataset <- churn_dataset[, -1]
```

- Secondly, we will cleanse the categorical attributes and recode them accordingly.

- After that, we will encode all other categorical variables in the dataset.

- Next step is to convert column to factors.

After above mentioned modifications, below is the final shape of churn dataset.

```
## Observations: 7,032
## Variables: 20
## $ gender          <fct> 0, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, ...
## $ SeniorCitizen   <fct> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ Partner         <fct> 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, ...
## $ Dependents      <fct> 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, ...
## $ tenure          <int> 1, 34, 2, 45, 2, 8, 22, 10, 28, 62, 13, 16, 5...
## $ PhoneService    <fct> 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, ...
## $ MultipleLines   <fct> 3, 0, 0, 3, 0, 1, 1, 3, 1, 0, 0, 0, 1, 1, 0, ...
## $ InternetService <fct> 1, 1, 1, 1, 2, 2, 2, 1, 2, 1, 1, 0, 2, 2, 2, ...
## $ OnlineSecurity  <fct> 0, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 2, 0, 0, 1, ...
## $ OnlineBackup    <fct> 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 2, 0, 1, 0, ...
## $ DeviceProtection <fct> 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 2, 1, 1, 1, ...
## $ TechSupport     <fct> 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 2, 0, 0, 1, ...
## $ StreamingTV     <fct> 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 2, 1, 1, 1, ...
## $ StreamingMovies <fct> 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 2, 1, 1, 1, ...
## $ Contract        <fct> 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 2, 1, 0, 0, ...
## $ PaperlessBilling <fct> 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, ...
## $ PaymentMethod   <fct> 1, 2, 2, 3, 1, 1, 4, 2, 1, 3, 2, 4, 4, 3, 1, ...
## $ MonthlyCharges  <dbl> 29.85, 56.95, 53.85, 42.30, 70.70, 99.65, 89....
## $ TotalCharges    <dbl> 29.85, 1889.50, 108.15, 1840.75, 151.65, 820....
## $ Churn           <fct> 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, ...
```

2.4 **Data Modeling**

For data modeling, we will start off by creating test and train datasets by using 75-25 split.

2.4.1 **Logistic Regression Model**

To start, we will use logistic regression to predict customers churn for this dataset.

Regression is considered to be a good technique for identifying and predicting customer satisfaction. For each of the variables in a regression model, the standard error rate is calculated using SPSS. Then the variables with the most significance in respect to linear regressions for churn prediction are obtained and a regression model is constructed.

Here, we will be using forward (step up) selection procedure to select the best features to be used in regression.

Next up, we will use selected variables for logistic regression modeling.

The confusion matrix, thus, formed is shown as below.

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |          N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  1757
##
##
##              | PREDICTED
##      ACTUAL |     FALSE |      TRUE | Row Total |
## -------------|-----------|-----------|-----------|
##           0 |      1140 |       150 |      1290 |
##             |     0.649 |     0.085 |           |
## -------------|-----------|-----------|-----------|
##           1 |       201 |       266 |       467 |
##             |     0.114 |     0.151 |           |
## -------------|-----------|-----------|-----------|
## Column Total |      1341 |       416 |      1757 |
## -------------|-----------|-----------|-----------|
##
##
```

Let's evaluate the performance of this model by using cutoff of 0.5

```
##            pred.churn
## actual.churn   No  Yes
##          No 1140  150
##          Yes  201  266
```

```
##   Accuracy
## 0.8002277
```

```
## Sensitivity
##   0.5695931
```

```
## Specificity
##   0.8837209
```

As can be seen that with cutoff of 0.50, although we are achieving good accuracy and specificity, the sensitivity appeared to be very less. Let's tweak the cutoff value to 0.4 to improve sensitivity.

```
##            pred.churn
## actual.churn   No  Yes
##          No 1051  239
##          Yes  155  312
```

```
##   Accuracy
## 0.7757541
```

```
## Sensitivity
##   0.6680942
```

```
## Specificity
##   0.8147287
```

```
## [1] 0.8408745
```

Summarizing the results of 1st model in below table.

Method	Accuracy	Sensitivity	Specificity	AUC
Logistic_Regression	0.7757541	0.6680942	0.8147287	0.8408745

2.4.2 LDA Model

Let's try Linear Discriminant Analysis(LDA) model to perform churn prediction. LDA is a very common technique used for supervised classification problems. Basically, a dimensionality reduction technique, the main goal of LDA is to reduce the dimensions by removing the reduntant and dependent features by transforming the features from higher dimensional space to a space with lower dimensions.

Training LDA model and testing on 25% data.

The confusion matrix, thus, formed is shown as below.

```
##
##
##    Cell Contents
## |-----------------------|
## |                     N |
## |         N / Table Total |
## |-----------------------|
##
##
## Total Observations in Table:  1757
##
##
##             | PREDICTED
##     ACTUAL |        0 |        1 | Row Total |
## ------------|----------|----------|----------|
##          0 |     1115 |      175 |     1290 |
##            |    0.635 |    0.100 |          |
## ------------|----------|----------|----------|
##          1 |      214 |      253 |      467 |
##            |    0.122 |    0.144 |          |
## ------------|----------|----------|----------|
## Column Total |     1329 |      428 |     1757 |
## ------------|----------|----------|----------|
##
##
```

For model performance, it can be seen that we are getting much better sensitivity values as compared to Logistic Regression

```
##  Accuracy
## 0.7785999
```

```
## Sensitivity
##   0.8643411
```

```
## Specificity
##   0.5417559
```

Method	Accuracy	Sensitivity	Specificity	AUC
Logistic_Regression	0.7757541	0.6680942	0.8147287	0.8408745
LDA Model	0.7785999	0.8643411	0.5417559	0.8335234

2.4.3 Decision Tree Model

The decision tree is the most prominent predictive model that is used for the purpose of classification. The decision tree consists of two steps, tree building and tree pruning. In tree building, the training set data is recursively partitioned in accordance with the values of the attributes.

This process goes on until there is no one partition is left to have identical values. In this process some values may be removed from the data due to noisy data.The largest estimated error rate branches are selected and then removed in pruning.

Let's use decision tree to predict customers churn
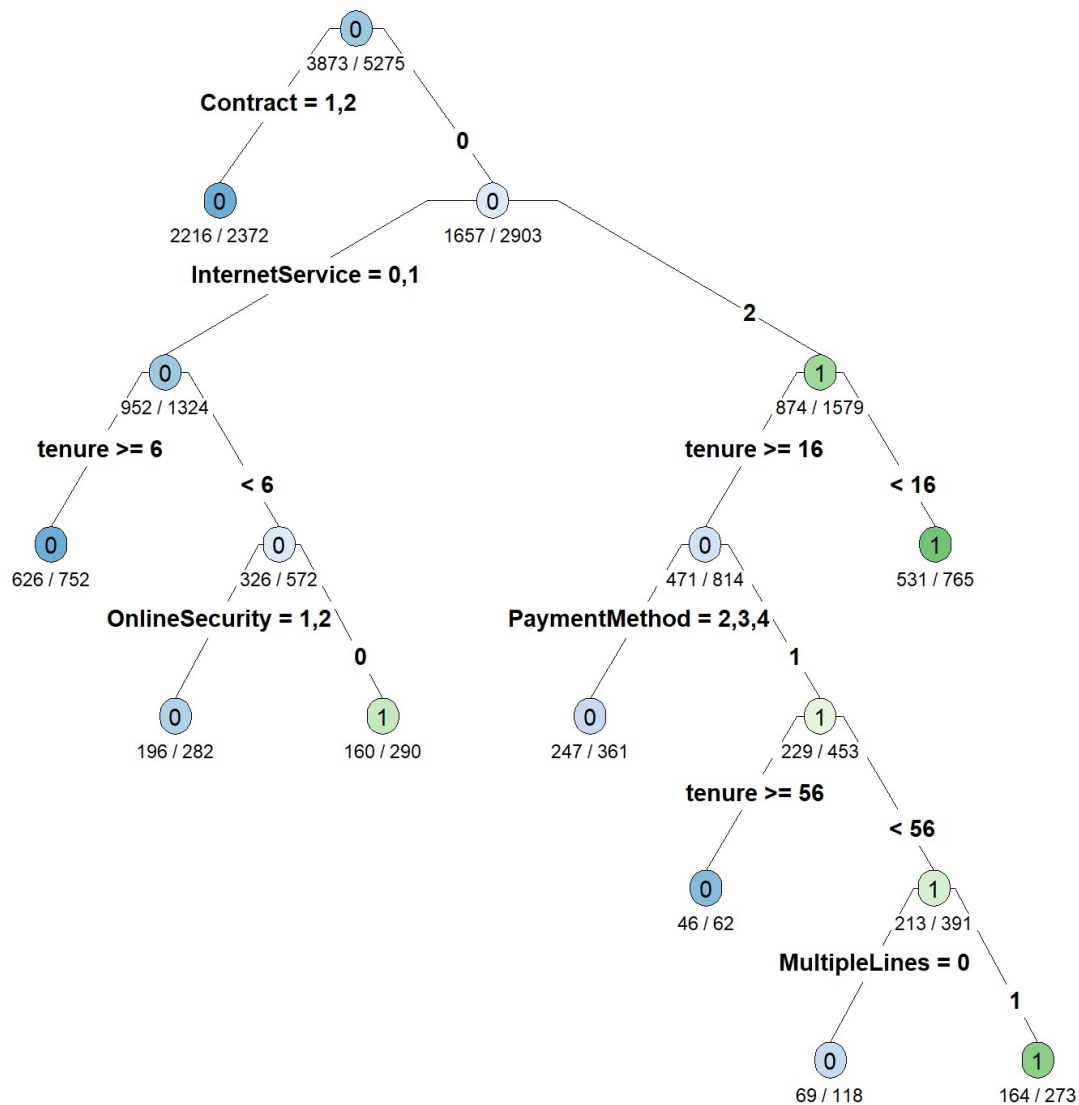
For model performance, decision tree model got an Accuracy of 0.783 with Sesnsitvity of 0.85.

```
# Prediction/Probability
treepred <- predict(tree_fit,test,type = "prob")
tree.pred.prob.val <- prediction(treepred[,2],test$Churn)
tree.pred.prob.perf <- performance(tree.pred.prob.val,"auc")


#Model Performance
dt_accuracy <- dt_conf$overall[1]
dt_sensitivity <- dt_conf$byClass[1]
dt_specificity <- dt_conf$byClass[2]
dt_accuracy
```

```
##   Accuracy
## 0.7831531
```

```
dt_sensitivity
```

```
## Sensitivity
##   0.8573643
```

```
dt_specificity
```

```
## Specificity
##   0.5781585
```

Results can be summarized as below.

Method	Accuracy	Sensitivity	Specificity	AUC
Logistic_Regression	0.7757541	0.6680942	0.8147287	0.8408745
LDA Model	0.7785999	0.8643411	0.5417559	0.8335234
Decision Tree Model	0.7831531	0.8573643	0.5781585	0.7982521

2.4.4 **Random Forest Model**

Next, we will use Random Forest i.e. use multiple decision trees to make predictions.

Single decision trees on their own can be very effective at learning non-linear relationships (low bias, but high variance). Due to their high variance, they can tend to over-fit. Random forest reduces this variance by averaging many trees (at the sacrifice of a slight increase in the bias).

The confusion matrix, thus, formed is shown as below.

```
##
##
##    Cell Contents
## |-----------------------|
## |                     N |
## |         N / Table Total |
## |-----------------------|
##
##
## Total Observations in Table:  1757
##
##
##              | PREDICTED
##      ACTUAL  |         0 |         1 | Row Total |
## -------------|-----------|-----------|-----------|
##           0  |      1153 |       137 |      1290 |
##              |     0.656 |     0.078 |           |
## -------------|-----------|-----------|-----------|
##           1  |       217 |       250 |       467 |
##              |     0.124 |     0.142 |           |
## -------------|-----------|-----------|-----------|
## Column Total |      1370 |       387 |      1757 |
## -------------|-----------|-----------|-----------|
##
##
```

```
##   Accuracy
## 0.7985202
```

```
## Sensitivity
##   0.8937984
```

```
## Specificity
##   0.5353319
```

Results can be summarized as below.

Method	Accuracy	Sensitivity	Specificity	AUC

Method	Accuracy	Sensitivity	Specificity	AUC
Logistic_Regression	0.7757541	0.6680942	0.8147287	0.8408745
LDA Model	0.7785999	0.8643411	0.5417559	0.8335234
Decision Tree Model	0.7831531	0.8573643	0.5781585	0.7982521
Random Forest Model	0.7985202	0.8937984	0.5353319	0.8287237

2.4.5 **Support Vector Machine (SVM) Model**

The SVM classifier deals with linear permutation of subset of the training set by finding a maximum edge over energized plane. SVM algorithm works on finding a hyperplane which will maximize the distance between the hyperplane and the support vectors. The formulation of the hyperplane can be both linear and non-linear.

Applying SVM on churn dataset.

The confusion matrix and model performance can be shown as below.

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |          N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  1757
##
##
##              | PREDICTED
##      ACTUAL  |        0 |        1 | Row Total |
## -------------|----------|----------|-----------|
##           0  |     1138 |      152 |      1290 |
##              |    0.648 |    0.087 |           |
## -------------|----------|----------|-----------|
##           1  |      215 |      252 |       467 |
##              |    0.122 |    0.143 |           |
## -------------|----------|----------|-----------|
## Column Total |     1353 |      404 |      1757 |
## -------------|----------|----------|-----------|
##
##
```

```
##  Accuracy
## 0.7911212
```

```
## Sensitivity
##   0.8821705
```
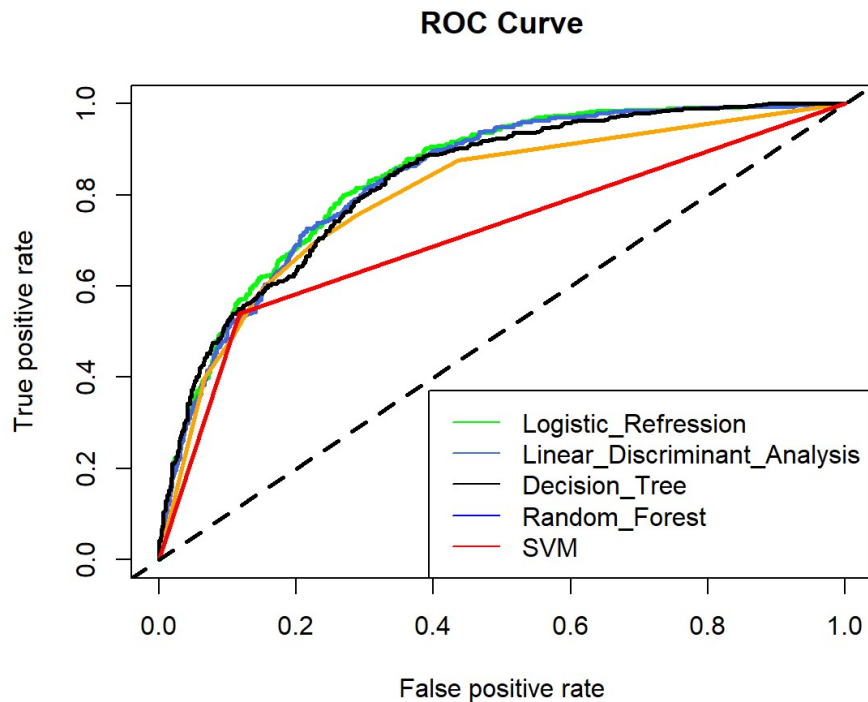
```
## Specificity
##   0.5396146
```

Results can be summarized as below.

Method	Accuracy	Sensitivity	Specificity	AUC
Logistic_Regression	0.7757541	0.6680942	0.8147287	0.8408745
LDA Model	0.7785999	0.8643411	0.5417559	0.8335234
Decision Tree Model	0.7831531	0.8573643	0.5781585	0.7982521

Method	Accuracy	Sensitivity	Specificity	AUC
Random Forest Model	0.7985202	0.8937984	0.5353319	0.8287237
SVM Model	0.7911212	0.8821705	0.5396146	0.7108926

3. **Results**

Let's plot ROC curves for each of these models.

It's clear that SVM has the worst AUC as compared to other models.

The results and comparison of above mentioned models can be summarized in below table.

Method	Accuracy	Sensitivity	Specificity	AUC
Logistic_Regression	0.7757541	0.6680942	0.8147287	0.8408745
LDA Model	0.7785999	0.8643411	0.5417559	0.8335234
Decision Tree Model	0.7831531	0.8573643	0.5781585	0.7982521
Random Forest Model	0.7985202	0.8937984	0.5353319	0.8287237
SVM Model	0.7911212	0.8821705	0.5396146	0.7108926

It's clear that Random Forest provides best result for Sensitivity as compared to other models. Although, Specificity value is quite average for Random Forest. However, being churn prediction cae, where major focus is to engage potential churn customers, low Specificity can be tolerated.

4. **Conclusion**

To conclude, below is the summary and highlights of this capstone project:

- After loading dataset, We started with an exploratory analysis and key insights about the data.
- First observation was that new-joiners are more prone to churn than old customers.
- TotalCharges are positively correlated with Tenure & MonthlyCharges
- It's observed that Partners & Dependants have significance in terms of customers loyalty.
- It' also seen that monthly subscription customers have much higher churn rate as compared to the ones with yearly subscriptions.

- As majority of dataset has categorical variables, we performed encoding to use for Modeling.
- For modeling, we split the data in Train & Test datasets and trained models using Logistic Regression, LDA, Decision Tree, Random Forest and SVM.
- We validate the models by testing it on test dataset and summarizing the whole results as below:

Method	Sensitivity	AUC
Logistic_Regression	0.6680942	0.8408745
LDA Model	0.8643411	0.8335234
Decision Tree Model	0.8573643	0.7982521
Random Forest Model	0.8937984	0.8287237
SVM Model	0.8821705	0.7108926

5. References

- Kirui, C., Hong, L., Cheruiyot, W., & Kirui, H. (2013). Predicting Customer Churn in Mobile Telephony Industry using Probabilistic Classifiers in Data Mining. IJCSI International Journal of Computer Science Issues, 10, 165-172.
- Lariviere, B., & Van den Poel, D. (2005). Predicting Customer Retention and Profitability by using Random Forests and Regression Forests Techniques. Expert Systems with Applications
- Mattison, R. (2005). The Telco Churn Management Handbook. Oakwood Hills, Illinois,USA: XiT Press.
- Nie, G., Rowe, W., Zhang, L., Tian, Y., & Shi, Y. (2011). Credit Card Churn Forecasting by Logistic Regression and Decision Tree. Expert Systems with Applications
- Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification. Expert Systems with Applications
- Khan, A. A., Jamwal, S., & Sepehri, M. M. (2010). Applying data mining to customer churn prediction in an internet service provider. International Journal of Computers and Applications