Telecom Data Chrun Analysis

Dhruv Cairae & Others

2/24/2022

# Introduction

## Overview

For this project our team decided to look at customer data to preform logistic regression. Our data set concerns customer retention data and our dependent variable is a dummy variable of whether a customer is retained. In this report we will be going through our data (including visualizations), exploratory data analysis, model building, and conclusions.

## The Data

The data selected for this project is a data set from kaggle called [“Telco Customer Churn”](https://www.kaggle.com/blastchar/telco-customer-churn/version/1). It is originally an IBM sample data set. As stated in the overview the data is concerned with whether customers are retained, this variable is a dummy variable called churn. Churn takes a value of 1 if the customer churned (left) or 0 if the customer is retained. The data set contains 19 more predictor varibles and a identification column. The predictor varibles fall into three broad categories: services each customer has, customer account information, and customer demographics. These variables will be explored in greater depth below.

# Exploratory Data Analysis

## Initial Wrangling

In this section of the report we will be exploring and explaining the data.

# load necessary packages  
library(dplyr)  
library(tidyverse)  
library(fastDummies)  
library(skimr)  
library(lares)  
library(caTools)  
library(caret)

# read in the data  
file<-read.csv("C:/Users/Dhruv Cairae/Desktop/WA\_Fn-UseC\_-Telco-Customer-Churn.csv",header=T) # reads in the data using read.csv()  
summary(file) # summary Statistics

## customerID gender SeniorCitizen Partner Dependents  
## 0002-ORFBO: 1 Female:3488 Min. :0.0000 No :3641 No :4933   
## 0003-MKNFE: 1 Male :3555 1st Qu.:0.0000 Yes:3402 Yes:2110   
## 0004-TLHLJ: 1 Median :0.0000   
## 0011-IGKFF: 1 Mean :0.1621   
## 0013-EXCHZ: 1 3rd Qu.:0.0000   
## 0013-MHZWF: 1 Max. :1.0000   
## (Other) :7037   
## tenure PhoneService MultipleLines InternetService  
## Min. : 0.00 No : 682 No :3390 DSL :2421   
## 1st Qu.: 9.00 Yes:6361 No phone service: 682 Fiber optic:3096   
## Median :29.00 Yes :2971 No :1526   
## Mean :32.37   
## 3rd Qu.:55.00   
## Max. :72.00   
##   
## OnlineSecurity OnlineBackup   
## No :3498 No :3088   
## No internet service:1526 No internet service:1526   
## Yes :2019 Yes :2429   
##   
##   
##   
##   
## DeviceProtection TechSupport   
## No :3095 No :3473   
## No internet service:1526 No internet service:1526   
## Yes :2422 Yes :2044   
##   
##   
##   
##   
## StreamingTV StreamingMovies Contract   
## No :2810 No :2785 Month-to-month:3875   
## No internet service:1526 No internet service:1526 One year :1473   
## Yes :2707 Yes :2732 Two year :1695   
##   
##   
##   
##   
## PaperlessBilling PaymentMethod MonthlyCharges   
## No :2872 Bank transfer (automatic):1544 Min. : 18.25   
## Yes:4171 Credit card (automatic) :1522 1st Qu.: 35.50   
## Electronic check :2365 Median : 70.35   
## Mailed check :1612 Mean : 64.76   
## 3rd Qu.: 89.85   
## Max. :118.75   
##   
## TotalCharges Churn   
## Min. : 18.8 No :5174   
## 1st Qu.: 401.4 Yes:1869   
## Median :1397.5   
## Mean :2283.3   
## 3rd Qu.:3794.7   
## Max. :8684.8   
## NA's :11

str(file) # simple structure

## 'data.frame': 7043 obs. of 21 variables:  
## $ customerID : Factor w/ 7043 levels "0002-ORFBO","0003-MKNFE",..: 5376 3963 2565 5536 6512 6552 1003 4771 5605 4535 ...  
## $ gender : Factor w/ 2 levels "Female","Male": 1 2 2 2 1 1 2 1 1 2 ...  
## $ SeniorCitizen : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Partner : Factor w/ 2 levels "No","Yes": 2 1 1 1 1 1 1 1 2 1 ...  
## $ Dependents : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 2 1 1 2 ...  
## $ tenure : int 1 34 2 45 2 8 22 10 28 62 ...  
## $ PhoneService : Factor w/ 2 levels "No","Yes": 1 2 2 1 2 2 2 1 2 2 ...  
## $ MultipleLines : Factor w/ 3 levels "No","No phone service",..: 2 1 1 2 1 3 3 2 3 1 ...  
## $ InternetService : Factor w/ 3 levels "DSL","Fiber optic",..: 1 1 1 1 2 2 2 1 2 1 ...  
## $ OnlineSecurity : Factor w/ 3 levels "No","No internet service",..: 1 3 3 3 1 1 1 3 1 3 ...  
## $ OnlineBackup : Factor w/ 3 levels "No","No internet service",..: 3 1 3 1 1 1 3 1 1 3 ...  
## $ DeviceProtection: Factor w/ 3 levels "No","No internet service",..: 1 3 1 3 1 3 1 1 3 1 ...  
## $ TechSupport : Factor w/ 3 levels "No","No internet service",..: 1 1 1 3 1 1 1 1 3 1 ...  
## $ StreamingTV : Factor w/ 3 levels "No","No internet service",..: 1 1 1 1 1 3 3 1 3 1 ...  
## $ StreamingMovies : Factor w/ 3 levels "No","No internet service",..: 1 1 1 1 1 3 1 1 3 1 ...  
## $ Contract : Factor w/ 3 levels "Month-to-month",..: 1 2 1 2 1 1 1 1 1 2 ...  
## $ PaperlessBilling: Factor w/ 2 levels "No","Yes": 2 1 2 1 2 2 2 1 2 1 ...  
## $ PaymentMethod : Factor w/ 4 levels "Bank transfer (automatic)",..: 3 4 4 1 3 3 2 4 3 1 ...  
## $ MonthlyCharges : num 29.9 57 53.9 42.3 70.7 ...  
## $ TotalCharges : num 29.9 1889.5 108.2 1840.8 151.7 ...  
## $ Churn : Factor w/ 2 levels "No","Yes": 1 1 2 1 2 2 1 1 2 1 ...

head(file) # first 6 rows

## customerID gender SeniorCitizen Partner Dependents tenure PhoneService  
## 1 7590-VHVEG Female 0 Yes No 1 No  
## 2 5575-GNVDE Male 0 No No 34 Yes  
## 3 3668-QPYBK Male 0 No No 2 Yes  
## 4 7795-CFOCW Male 0 No No 45 No  
## 5 9237-HQITU Female 0 No No 2 Yes  
## 6 9305-CDSKC Female 0 No No 8 Yes  
## MultipleLines InternetService OnlineSecurity OnlineBackup DeviceProtection  
## 1 No phone service DSL No Yes No  
## 2 No DSL Yes No Yes  
## 3 No DSL Yes Yes No  
## 4 No phone service DSL Yes No Yes  
## 5 No Fiber optic No No No  
## 6 Yes Fiber optic No No Yes  
## TechSupport StreamingTV StreamingMovies Contract PaperlessBilling  
## 1 No No No Month-to-month Yes  
## 2 No No No One year No  
## 3 No No No Month-to-month Yes  
## 4 Yes No No One year No  
## 5 No No No Month-to-month Yes  
## 6 No Yes Yes Month-to-month Yes  
## PaymentMethod MonthlyCharges TotalCharges Churn  
## 1 Electronic check 29.85 29.85 No  
## 2 Mailed check 56.95 1889.50 No  
## 3 Mailed check 53.85 108.15 Yes  
## 4 Bank transfer (automatic) 42.30 1840.75 No  
## 5 Electronic check 70.70 151.65 Yes  
## 6 Electronic check 99.65 820.50 Yes

Based on the summary, head, and structure functions we can see that most of the data is numerical or character. We can see that many of the character variables appear to actually be categorical (we will need to transform these) and we also need to check for missing values.

table(is.na(file)) # check for missing values in the data frame

##   
## FALSE TRUE   
## 147892 11

Our table of missing values shows that only 11 are missing out of nearly 150,000 data points. This is negligible and should not interfere with our model. Based on the data dictionary we have access to we know that the customerID variable is unique for every row. Our next step is to remove this from our data set.

file\_1<-file[-c(1)] # dataframe of the last column  
file\_2<-na.omit(file\_1) # dataframe of everything except the last column

Now that we have removed our first column, the next step is to transform the character variables into factors. This will make it easier to preform our logistic regression in the model building section, the numeric variables are left alone.

file\_3 <- dummy\_cols(file\_2, select\_columns = c('gender','Partner','Dependents','PhoneService','MultipleLines',  
 'InternetService','OnlineSecurity','OnlineBackup','DeviceProtection',  
 'TechSupport','StreamingTV','StreamingMovies',  
 'PaperlessBilling'),remove\_selected\_columns = TRUE) # creates factors of variables

After transforming these variables to factors, we further transform some of these into dummy variables to make our analysis easier and remove unecessary variables.

dataset<-subset(file\_3,select = c(SeniorCitizen,tenure,MonthlyCharges,TotalCharges,gender\_Male,Partner\_Yes,Dependents\_Yes,  
 PhoneService\_Yes,MultipleLines\_Yes,InternetService\_DSL,InternetService\_No,OnlineSecurity\_Yes,OnlineBackup\_Yes,DeviceProtection\_Yes,TechSupport\_Yes,StreamingTV\_Yes,  
 StreamingMovies\_Yes,PaperlessBilling\_Yes)) # subset of only necessary data

dataset$creditcard<- ifelse(file\_3$PaymentMethod=="Credit card (automatic)", 1, 0) # Creates new columns using ifelse()   
dataset$banktransfer<- ifelse(file\_3$PaymentMethod=="Bank transfer (automatic)", 1, 0)  
dataset$ec<- ifelse(file\_3$PaymentMethod=="Electronic check", 1, 0)  
dataset$monthlycontract<- ifelse(file\_3$Contract=="Month-to-month", 1, 0)  
dataset$annual<- ifelse(file\_3$Contract=="One year", 1, 0)  
dataset$Churn\_Yes<- ifelse(file\_3$Churn=="Yes", 1, 0)  
skim(dataset) # summary statistics similar to summary()

Data summary

|  |  |
| --- | --- |
| Name | dataset |
| Number of rows | 7032 |
| Number of columns | 24 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| numeric | 24 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: numeric**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| SeniorCitizen | 0 | 1 | 0.16 | 0.37 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | ▇▁▁▁▂ |
| tenure | 0 | 1 | 32.42 | 24.55 | 1.00 | 9.00 | 29.00 | 55.00 | 72.00 | ▇▃▃▃▅ |
| MonthlyCharges | 0 | 1 | 64.80 | 30.09 | 18.25 | 35.59 | 70.35 | 89.86 | 118.75 | ▇▅▆▇▅ |
| TotalCharges | 0 | 1 | 2283.30 | 2266.77 | 18.80 | 401.45 | 1397.47 | 3794.74 | 8684.80 | ▇▂▂▂▁ |
| gender\_Male | 0 | 1 | 0.50 | 0.50 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | ▇▁▁▁▇ |
| Partner\_Yes | 0 | 1 | 0.48 | 0.50 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | ▇▁▁▁▇ |
| Dependents\_Yes | 0 | 1 | 0.30 | 0.46 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | ▇▁▁▁▃ |
| PhoneService\_Yes | 0 | 1 | 0.90 | 0.30 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | ▁▁▁▁▇ |
| MultipleLines\_Yes | 0 | 1 | 0.42 | 0.49 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | ▇▁▁▁▆ |
| InternetService\_DSL | 0 | 1 | 0.34 | 0.47 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | ▇▁▁▁▅ |
| InternetService\_No | 0 | 1 | 0.22 | 0.41 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | ▇▁▁▁▂ |
| OnlineSecurity\_Yes | 0 | 1 | 0.29 | 0.45 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | ▇▁▁▁▃ |
| OnlineBackup\_Yes | 0 | 1 | 0.34 | 0.48 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | ▇▁▁▁▅ |
| DeviceProtection\_Yes | 0 | 1 | 0.34 | 0.48 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | ▇▁▁▁▅ |
| TechSupport\_Yes | 0 | 1 | 0.29 | 0.45 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | ▇▁▁▁▃ |
| StreamingTV\_Yes | 0 | 1 | 0.38 | 0.49 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | ▇▁▁▁▅ |
| StreamingMovies\_Yes | 0 | 1 | 0.39 | 0.49 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | ▇▁▁▁▅ |
| PaperlessBilling\_Yes | 0 | 1 | 0.59 | 0.49 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | ▆▁▁▁▇ |
| creditcard | 0 | 1 | 0.22 | 0.41 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | ▇▁▁▁▂ |
| banktransfer | 0 | 1 | 0.22 | 0.41 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | ▇▁▁▁▂ |
| ec | 0 | 1 | 0.34 | 0.47 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | ▇▁▁▁▅ |
| monthlycontract | 0 | 1 | 0.55 | 0.50 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 | ▆▁▁▁▇ |
| annual | 0 | 1 | 0.21 | 0.41 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | ▇▁▁▁▂ |
| Churn\_Yes | 0 | 1 | 0.27 | 0.44 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | ▇▁▁▁▃ |

The summary of our new data set demonstrates how our transformations have affected our original data. This summary can be compared to the summary created in the first part of this section.

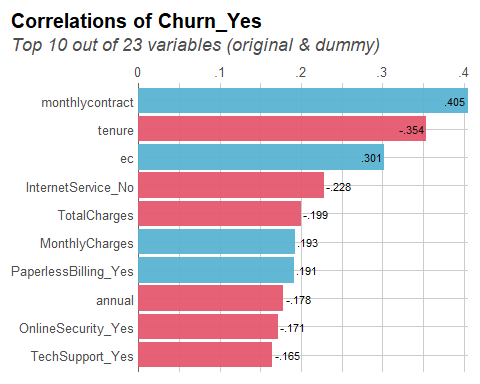
# Initial Model Building

## Correlation Analysis & Visualizations of top 3 Variables of Interest

We preform a correlation analysis to get the top 10 correlated values from there we will narrow it down to the top 3 variables of interest.

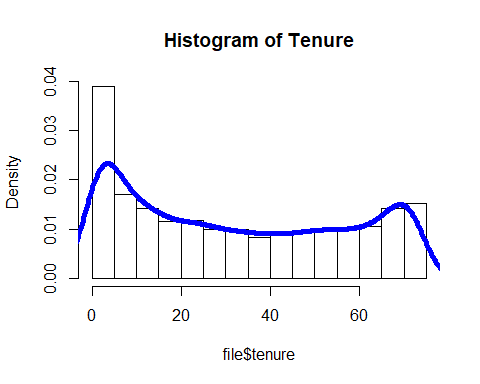
corr\_var(dataset, # name of dataset  
 Churn\_Yes, # name of variable to focus on  
 top = 10 # display top 10 correlations  
) # correlation analysis

## Warning in .font\_global(font, quiet = FALSE): Font 'Arial Narrow' is not  
## installed, has other name, or can't be found

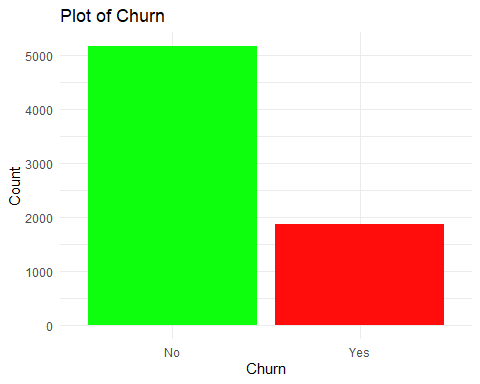


Based on the correlation analysis the most important variables in relation to churn\_yes (the customer has churned) are monthlycontract (the customers contract is month to month), tenure (The number of months the customer has stayed with the company), and ec (the customer pays with an electric check). EC and monthlycontract are highly positively correlated while tenure is highly negatively correlated. This makes logical sense for the tenure variable, the longer you are with the company the less likely you are to leave. We have graphed the three most important variables to make viewing them easier.

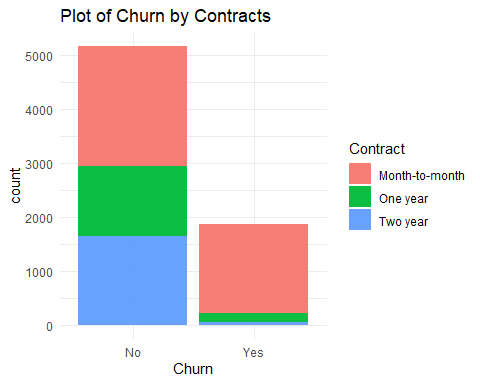
hist(file$tenure,main="Histogram of Tenure",freq = FALSE) #histogram of tenure  
lines(density(file$tenure), lwd=5, col='blue')



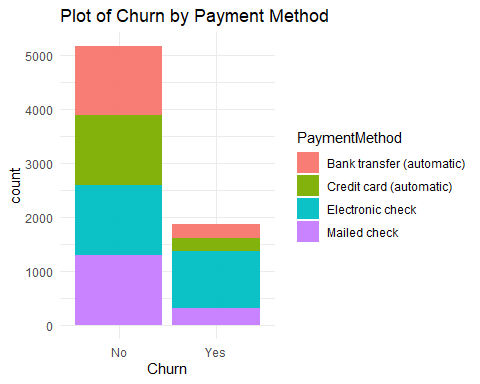
ggplot(file, aes(x = Churn)) + #ggplot of churn  
 geom\_bar(fill=c('green','red'))+  
 theme\_minimal()+  
 ggtitle("Plot of Churn")+  
 labs(x = "Churn", y = "Count")



ggplot(file, aes(x = Churn,fill=Contract)) + # ggplot of churn by contracts  
 geom\_bar()+  
 ggtitle("Plot of Churn by Contracts")+  
 theme\_minimal()+  
 labs(x = "Churn")



ggplot(file, aes(x = Churn,fill=PaymentMethod)) + # ggplot of churn by payment method  
 geom\_bar()+  
 ggtitle("Plot of Churn by Payment Method")+  
 theme\_minimal()+  
 labs(x = "Churn")



## Model Development

In this section we will be developing a model to predict whether a customer leaves the company or not. First we split the data into training and testing sets. Then we fit XGBoost to the training set and use K-fold cross validation to validate our predicted results.

set.seed(100)  
split = sample.split(dataset$Churn\_Yes, SplitRatio = 0.8) # create testing and training datasets  
training\_set = subset(dataset, split == TRUE)  
test\_set = subset(dataset, split == FALSE)  
  
# Fitting XGBoost to the Training set & K-Fold Cross Validation  
library(xgboost)  
classifier = xgboost(data = as.matrix(training\_set[-24]), label = training\_set$Churn\_Yes, nrounds = 10)

## [1] train-rmse:0.434445   
## [2] train-rmse:0.395380   
## [3] train-rmse:0.372658   
## [4] train-rmse:0.357991   
## [5] train-rmse:0.348013   
## [6] train-rmse:0.341597   
## [7] train-rmse:0.335460   
## [8] train-rmse:0.331614   
## [9] train-rmse:0.327296   
## [10] train-rmse:0.324736

# Predicting the Test set results  
y\_pred = predict(classifier, newdata = as.matrix(test\_set[-24]))  
y\_pred = (y\_pred >= 0.5)  
cm = table(test\_set[, 24], y\_pred)  
cm

## y\_pred  
## FALSE TRUE  
## 0 942 91  
## 1 174 200

library(caret)  
folds = createFolds(training\_set$Churn\_Yes, k = 5)  
cv = lapply(folds, function(x) {  
 training\_fold = training\_set[-x, ]  
 test\_fold = training\_set[x, ]  
 classifier = xgboost(data = as.matrix(training\_fold[-24]), label = training\_fold$Churn\_Yes, nrounds = 5)  
 y\_pred = predict(classifier, newdata = as.matrix(test\_fold[-24]))  
 y\_pred = (y\_pred >= 0.5)  
 cm = table(test\_fold[, 24], y\_pred)  
 accuracy = (cm[1,1] + cm[2,2]) / (cm[1,1] + cm[2,2] + cm[1,2] + cm[2,1])  
 return(accuracy)  
})

## [1] train-rmse:0.433763   
## [2] train-rmse:0.394571   
## [3] train-rmse:0.371047   
## [4] train-rmse:0.355262   
## [5] train-rmse:0.345286   
## [1] train-rmse:0.433647   
## [2] train-rmse:0.394210   
## [3] train-rmse:0.370405   
## [4] train-rmse:0.355088   
## [5] train-rmse:0.345524   
## [1] train-rmse:0.433397   
## [2] train-rmse:0.393363   
## [3] train-rmse:0.369611   
## [4] train-rmse:0.355079   
## [5] train-rmse:0.345474   
## [1] train-rmse:0.432931   
## [2] train-rmse:0.393024   
## [3] train-rmse:0.369668   
## [4] train-rmse:0.354796   
## [5] train-rmse:0.344006   
## [1] train-rmse:0.431996   
## [2] train-rmse:0.392834   
## [3] train-rmse:0.368289   
## [4] train-rmse:0.353862   
## [5] train-rmse:0.343629

accuracy = mean(as.numeric(cv))  
accuracy

## [1] 0.7946667

Based on the test results table and the accuracy measurement this appears to be an excellent model for our predictive purposes.

## Another modeling attempt using glm (Logistic Regression)

In this section we will test different glm models using the full model, the null model, the model using the 3 predictor variables from the previous section, stepwise AIC, and stepwise BIC.

## glm Model on all variables  
full\_model <- glm(Churn\_Yes ~ ., family = binomial, data = training\_set)  
full\_model\_summary <- summary(full\_model)  
  
full\_model\_summary$deviance/full\_model\_summary$df.residual # in-sample model mean residual deviance

## [1] 0.8298638

AIC(full\_model) #AIC

## [1] 4696.067

BIC(full\_model) #BIC

## [1] 4855.307

## glm Model on no variables  
null\_model <- glm(Churn\_Yes ~ 1, family = binomial, data = training\_set)  
null\_model\_summary <- summary(null\_model)  
  
null\_model\_summary$deviance/null\_model\_summary$df.residual # in-sample model mean residual deviance

## [1] 1.158234

AIC(null\_model)

## [1] 6515.907

BIC(null\_model)

## [1] 6522.542

## glm Model on Education and PAY\_0 variables  
glm\_model <- glm(Churn\_Yes ~monthlycontract+tenure+ec, family = binomial, data = training\_set)  
glm\_model\_summary <- summary(glm\_model)  
  
glm\_model\_summary$deviance/glm\_model\_summary$df.residual # in-sample model mean residual deviance

## [1] 0.9191821

AIC(glm\_model)

## [1] 5174.722

BIC(glm\_model)

## [1] 5201.262

glm\_model\_summary$df.residual

## [1] 5621

## Start: AIC=4696.07  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## gender\_Male + Partner\_Yes + Dependents\_Yes + PhoneService\_Yes +   
## MultipleLines\_Yes + InternetService\_DSL + InternetService\_No +   
## OnlineSecurity\_Yes + OnlineBackup\_Yes + DeviceProtection\_Yes +   
## TechSupport\_Yes + StreamingTV\_Yes + StreamingMovies\_Yes +   
## PaperlessBilling\_Yes + creditcard + banktransfer + ec + monthlycontract +   
## annual  
##   
## Df Deviance AIC  
## - banktransfer 1 4648.1 4694.1  
## - OnlineBackup\_Yes 1 4648.1 4694.1  
## - Partner\_Yes 1 4648.1 4694.1  
## - gender\_Male 1 4648.1 4694.1  
## - PhoneService\_Yes 1 4648.2 4694.2  
## - DeviceProtection\_Yes 1 4648.6 4694.6  
## - OnlineSecurity\_Yes 1 4648.9 4694.9  
## - creditcard 1 4648.9 4694.9  
## - MonthlyCharges 1 4649.3 4695.3  
## - TechSupport\_Yes 1 4649.3 4695.3  
## <none> 4648.1 4696.1  
## - StreamingTV\_Yes 1 4650.4 4696.4  
## - StreamingMovies\_Yes 1 4650.6 4696.6  
## - Dependents\_Yes 1 4651.0 4697.0  
## - InternetService\_DSL 1 4651.6 4697.6  
## - SeniorCitizen 1 4651.6 4697.6  
## - InternetService\_No 1 4651.7 4697.7  
## - MultipleLines\_Yes 1 4652.6 4698.6  
## - ec 1 4658.0 4704.0  
## - annual 1 4661.1 4707.1  
## - PaperlessBilling\_Yes 1 4663.4 4709.4  
## - TotalCharges 1 4671.3 4717.3  
## - monthlycontract 1 4714.9 4760.9  
## - tenure 1 4740.9 4786.9  
##   
## Step: AIC=4694.08  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## gender\_Male + Partner\_Yes + Dependents\_Yes + PhoneService\_Yes +   
## MultipleLines\_Yes + InternetService\_DSL + InternetService\_No +   
## OnlineSecurity\_Yes + OnlineBackup\_Yes + DeviceProtection\_Yes +   
## TechSupport\_Yes + StreamingTV\_Yes + StreamingMovies\_Yes +   
## PaperlessBilling\_Yes + creditcard + ec + monthlycontract +   
## annual  
##   
## Df Deviance AIC  
## - OnlineBackup\_Yes 1 4648.1 4692.1  
## - Partner\_Yes 1 4648.1 4692.1  
## - gender\_Male 1 4648.2 4692.2  
## - PhoneService\_Yes 1 4648.2 4692.2  
## - DeviceProtection\_Yes 1 4648.6 4692.6  
## - OnlineSecurity\_Yes 1 4648.9 4692.9  
## - MonthlyCharges 1 4649.3 4693.3  
## - TechSupport\_Yes 1 4649.3 4693.3  
## - creditcard 1 4649.4 4693.4  
## <none> 4648.1 4694.1  
## - StreamingTV\_Yes 1 4650.4 4694.4  
## - StreamingMovies\_Yes 1 4650.6 4694.6  
## - Dependents\_Yes 1 4651.0 4695.0  
## - InternetService\_DSL 1 4651.7 4695.7  
## - SeniorCitizen 1 4651.7 4695.7  
## - InternetService\_No 1 4651.8 4695.8  
## - MultipleLines\_Yes 1 4652.6 4696.6  
## - annual 1 4661.1 4705.1  
## - ec 1 4663.0 4707.0  
## - PaperlessBilling\_Yes 1 4663.4 4707.4  
## - TotalCharges 1 4671.4 4715.4  
## - monthlycontract 1 4714.9 4758.9  
## - tenure 1 4742.6 4786.6  
##   
## Step: AIC=4692.1  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## gender\_Male + Partner\_Yes + Dependents\_Yes + PhoneService\_Yes +   
## MultipleLines\_Yes + InternetService\_DSL + InternetService\_No +   
## OnlineSecurity\_Yes + DeviceProtection\_Yes + TechSupport\_Yes +   
## StreamingTV\_Yes + StreamingMovies\_Yes + PaperlessBilling\_Yes +   
## creditcard + ec + monthlycontract + annual  
##   
## Df Deviance AIC  
## - Partner\_Yes 1 4648.2 4690.2  
## - gender\_Male 1 4648.2 4690.2  
## - PhoneService\_Yes 1 4649.2 4691.2  
## - creditcard 1 4649.5 4691.5  
## - OnlineSecurity\_Yes 1 4649.8 4691.8  
## - DeviceProtection\_Yes 1 4650.1 4692.1  
## <none> 4648.1 4692.1  
## - TechSupport\_Yes 1 4650.8 4692.8  
## - Dependents\_Yes 1 4651.0 4693.0  
## - SeniorCitizen 1 4651.7 4693.7  
## - MonthlyCharges 1 4655.7 4697.7  
## - StreamingTV\_Yes 1 4659.3 4701.3  
## - StreamingMovies\_Yes 1 4660.3 4702.3  
## - annual 1 4661.1 4703.1  
## - MultipleLines\_Yes 1 4662.4 4704.4  
## - ec 1 4663.0 4705.0  
## - PaperlessBilling\_Yes 1 4663.4 4705.4  
## - InternetService\_No 1 4668.1 4710.1  
## - InternetService\_DSL 1 4668.2 4710.2  
## - TotalCharges 1 4671.4 4713.4  
## - monthlycontract 1 4714.9 4756.9  
## - tenure 1 4742.7 4784.7  
##   
## Step: AIC=4690.16  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## gender\_Male + Dependents\_Yes + PhoneService\_Yes + MultipleLines\_Yes +   
## InternetService\_DSL + InternetService\_No + OnlineSecurity\_Yes +   
## DeviceProtection\_Yes + TechSupport\_Yes + StreamingTV\_Yes +   
## StreamingMovies\_Yes + PaperlessBilling\_Yes + creditcard +   
## ec + monthlycontract + annual  
##   
## Df Deviance AIC  
## - gender\_Male 1 4648.2 4688.2  
## - PhoneService\_Yes 1 4649.3 4689.3  
## - creditcard 1 4649.5 4689.5  
## - OnlineSecurity\_Yes 1 4649.9 4689.9  
## - DeviceProtection\_Yes 1 4650.2 4690.2  
## <none> 4648.2 4690.2  
## - TechSupport\_Yes 1 4650.8 4690.8  
## - Dependents\_Yes 1 4651.3 4691.3  
## - SeniorCitizen 1 4651.9 4691.9  
## - MonthlyCharges 1 4655.8 4695.8  
## - StreamingTV\_Yes 1 4659.4 4699.4  
## - StreamingMovies\_Yes 1 4660.4 4700.4  
## - annual 1 4661.2 4701.2  
## - MultipleLines\_Yes 1 4662.5 4702.5  
## - ec 1 4663.1 4703.1  
## - PaperlessBilling\_Yes 1 4663.5 4703.5  
## - InternetService\_No 1 4668.3 4708.3  
## - InternetService\_DSL 1 4668.4 4708.4  
## - TotalCharges 1 4671.5 4711.5  
## - monthlycontract 1 4715.0 4755.0  
## - tenure 1 4743.2 4783.2  
##   
## Step: AIC=4688.23  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## Dependents\_Yes + PhoneService\_Yes + MultipleLines\_Yes + InternetService\_DSL +   
## InternetService\_No + OnlineSecurity\_Yes + DeviceProtection\_Yes +   
## TechSupport\_Yes + StreamingTV\_Yes + StreamingMovies\_Yes +   
## PaperlessBilling\_Yes + creditcard + ec + monthlycontract +   
## annual  
##   
## Df Deviance AIC  
## - PhoneService\_Yes 1 4649.3 4687.3  
## - creditcard 1 4649.6 4687.6  
## - OnlineSecurity\_Yes 1 4649.9 4687.9  
## - DeviceProtection\_Yes 1 4650.2 4688.2  
## <none> 4648.2 4688.2  
## - TechSupport\_Yes 1 4650.9 4688.9  
## - Dependents\_Yes 1 4651.4 4689.4  
## - SeniorCitizen 1 4652.0 4690.0  
## - MonthlyCharges 1 4655.8 4693.8  
## - StreamingTV\_Yes 1 4659.4 4697.4  
## - StreamingMovies\_Yes 1 4660.5 4698.5  
## - annual 1 4661.2 4699.2  
## - MultipleLines\_Yes 1 4662.6 4700.6  
## - ec 1 4663.2 4701.2  
## - PaperlessBilling\_Yes 1 4663.6 4701.6  
## - InternetService\_No 1 4668.3 4706.3  
## - InternetService\_DSL 1 4668.4 4706.4  
## - TotalCharges 1 4671.5 4709.5  
## - monthlycontract 1 4715.0 4753.0  
## - tenure 1 4743.3 4781.3  
##   
## Step: AIC=4687.31  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## Dependents\_Yes + MultipleLines\_Yes + InternetService\_DSL +   
## InternetService\_No + OnlineSecurity\_Yes + DeviceProtection\_Yes +   
## TechSupport\_Yes + StreamingTV\_Yes + StreamingMovies\_Yes +   
## PaperlessBilling\_Yes + creditcard + ec + monthlycontract +   
## annual  
##   
## Df Deviance AIC  
## - DeviceProtection\_Yes 1 4650.3 4686.3  
## - creditcard 1 4650.7 4686.7  
## <none> 4649.3 4687.3  
## - Dependents\_Yes 1 4652.5 4688.5  
## - SeniorCitizen 1 4653.0 4689.0  
## - OnlineSecurity\_Yes 1 4654.5 4690.5  
## - TechSupport\_Yes 1 4656.5 4692.5  
## - annual 1 4662.3 4698.3  
## - MultipleLines\_Yes 1 4664.0 4700.0  
## - ec 1 4664.1 4700.1  
## - PaperlessBilling\_Yes 1 4664.6 4700.6  
## - StreamingTV\_Yes 1 4665.9 4701.9  
## - MonthlyCharges 1 4666.6 4702.6  
## - StreamingMovies\_Yes 1 4668.4 4704.4  
## - TotalCharges 1 4672.1 4708.1  
## - InternetService\_DSL 1 4690.9 4726.9  
## - InternetService\_No 1 4708.3 4744.3  
## - monthlycontract 1 4715.9 4751.9  
## - tenure 1 4745.4 4781.4  
##   
## Step: AIC=4686.27  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## Dependents\_Yes + MultipleLines\_Yes + InternetService\_DSL +   
## InternetService\_No + OnlineSecurity\_Yes + TechSupport\_Yes +   
## StreamingTV\_Yes + StreamingMovies\_Yes + PaperlessBilling\_Yes +   
## creditcard + ec + monthlycontract + annual  
##   
## Df Deviance AIC  
## - creditcard 1 4651.7 4685.7  
## <none> 4650.3 4686.3  
## - Dependents\_Yes 1 4653.4 4687.4  
## - SeniorCitizen 1 4654.1 4688.1  
## - OnlineSecurity\_Yes 1 4656.3 4690.3  
## - TechSupport\_Yes 1 4657.9 4691.9  
## - annual 1 4663.1 4697.1  
## - MultipleLines\_Yes 1 4664.1 4698.1  
## - ec 1 4664.9 4698.9  
## - PaperlessBilling\_Yes 1 4665.5 4699.5  
## - StreamingTV\_Yes 1 4666.2 4700.2  
## - MonthlyCharges 1 4666.8 4700.8  
## - StreamingMovies\_Yes 1 4668.6 4702.6  
## - TotalCharges 1 4673.5 4707.5  
## - InternetService\_DSL 1 4692.0 4726.0  
## - InternetService\_No 1 4709.5 4743.5  
## - monthlycontract 1 4716.1 4750.1  
## - tenure 1 4746.1 4780.1  
##   
## Step: AIC=4685.66  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## Dependents\_Yes + MultipleLines\_Yes + InternetService\_DSL +   
## InternetService\_No + OnlineSecurity\_Yes + TechSupport\_Yes +   
## StreamingTV\_Yes + StreamingMovies\_Yes + PaperlessBilling\_Yes +   
## ec + monthlycontract + annual  
##   
## Df Deviance AIC  
## <none> 4651.7 4685.7  
## - Dependents\_Yes 1 4654.8 4686.8  
## - SeniorCitizen 1 4655.4 4687.4  
## - OnlineSecurity\_Yes 1 4657.7 4689.7  
## - TechSupport\_Yes 1 4659.2 4691.2  
## - annual 1 4664.5 4696.5  
## - MultipleLines\_Yes 1 4665.3 4697.3  
## - PaperlessBilling\_Yes 1 4666.4 4698.4  
## - StreamingTV\_Yes 1 4667.5 4699.5  
## - MonthlyCharges 1 4668.2 4700.2  
## - StreamingMovies\_Yes 1 4669.9 4701.9  
## - ec 1 4673.9 4705.9  
## - TotalCharges 1 4675.5 4707.5  
## - InternetService\_DSL 1 4693.1 4725.1  
## - InternetService\_No 1 4710.3 4742.3  
## - monthlycontract 1 4717.6 4749.6  
## - tenure 1 4749.9 4781.9

AIC\_step\_summary <- summary(AIC\_step)  
  
AIC\_step\_summary$deviance/AIC\_step\_summary$df.residual

## [1] 0.8294691

AIC(AIC\_step)

## [1] 4685.663

BIC(AIC\_step)

## [1] 4798.458

## Stepwise (BIC)  
n <- dim(training\_set)[1]

## Start: AIC=4855.31  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## gender\_Male + Partner\_Yes + Dependents\_Yes + PhoneService\_Yes +   
## MultipleLines\_Yes + InternetService\_DSL + InternetService\_No +   
## OnlineSecurity\_Yes + OnlineBackup\_Yes + DeviceProtection\_Yes +   
## TechSupport\_Yes + StreamingTV\_Yes + StreamingMovies\_Yes +   
## PaperlessBilling\_Yes + creditcard + banktransfer + ec + monthlycontract +   
## annual  
##   
## Df Deviance AIC  
## - banktransfer 1 4648.1 4846.7  
## - OnlineBackup\_Yes 1 4648.1 4846.7  
## - Partner\_Yes 1 4648.1 4846.7  
## - gender\_Male 1 4648.1 4846.7  
## - PhoneService\_Yes 1 4648.2 4846.8  
## - DeviceProtection\_Yes 1 4648.6 4847.2  
## - OnlineSecurity\_Yes 1 4648.9 4847.5  
## - creditcard 1 4648.9 4847.6  
## - MonthlyCharges 1 4649.3 4847.9  
## - TechSupport\_Yes 1 4649.3 4847.9  
## - StreamingTV\_Yes 1 4650.4 4849.0  
## - StreamingMovies\_Yes 1 4650.6 4849.2  
## - Dependents\_Yes 1 4651.0 4849.6  
## - InternetService\_DSL 1 4651.6 4850.3  
## - SeniorCitizen 1 4651.6 4850.3  
## - InternetService\_No 1 4651.7 4850.3  
## - MultipleLines\_Yes 1 4652.6 4851.2  
## <none> 4648.1 4855.3  
## - ec 1 4658.0 4856.6  
## - annual 1 4661.1 4859.7  
## - PaperlessBilling\_Yes 1 4663.4 4862.0  
## - TotalCharges 1 4671.3 4869.9  
## - monthlycontract 1 4714.9 4913.5  
## - tenure 1 4740.9 4939.5  
##   
## Step: AIC=4846.69  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## gender\_Male + Partner\_Yes + Dependents\_Yes + PhoneService\_Yes +   
## MultipleLines\_Yes + InternetService\_DSL + InternetService\_No +   
## OnlineSecurity\_Yes + OnlineBackup\_Yes + DeviceProtection\_Yes +   
## TechSupport\_Yes + StreamingTV\_Yes + StreamingMovies\_Yes +   
## PaperlessBilling\_Yes + creditcard + ec + monthlycontract +   
## annual  
##   
## Df Deviance AIC  
## - OnlineBackup\_Yes 1 4648.1 4838.1  
## - Partner\_Yes 1 4648.1 4838.1  
## - gender\_Male 1 4648.2 4838.1  
## - PhoneService\_Yes 1 4648.2 4838.2  
## - DeviceProtection\_Yes 1 4648.6 4838.6  
## - OnlineSecurity\_Yes 1 4648.9 4838.9  
## - MonthlyCharges 1 4649.3 4839.2  
## - TechSupport\_Yes 1 4649.3 4839.3  
## - creditcard 1 4649.4 4839.4  
## - StreamingTV\_Yes 1 4650.4 4840.4  
## - StreamingMovies\_Yes 1 4650.6 4840.6  
## - Dependents\_Yes 1 4651.0 4841.0  
## - InternetService\_DSL 1 4651.7 4841.6  
## - SeniorCitizen 1 4651.7 4841.6  
## - InternetService\_No 1 4651.8 4841.7  
## - MultipleLines\_Yes 1 4652.6 4842.6  
## <none> 4648.1 4846.7  
## - annual 1 4661.1 4851.1  
## - ec 1 4663.0 4852.9  
## - PaperlessBilling\_Yes 1 4663.4 4853.4  
## - TotalCharges 1 4671.4 4861.4  
## - monthlycontract 1 4714.9 4904.9  
## - tenure 1 4742.6 4932.6  
##   
## Step: AIC=4838.07  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## gender\_Male + Partner\_Yes + Dependents\_Yes + PhoneService\_Yes +   
## MultipleLines\_Yes + InternetService\_DSL + InternetService\_No +   
## OnlineSecurity\_Yes + DeviceProtection\_Yes + TechSupport\_Yes +   
## StreamingTV\_Yes + StreamingMovies\_Yes + PaperlessBilling\_Yes +   
## creditcard + ec + monthlycontract + annual  
##   
## Df Deviance AIC  
## - Partner\_Yes 1 4648.2 4829.5  
## - gender\_Male 1 4648.2 4829.5  
## - PhoneService\_Yes 1 4649.2 4830.5  
## - creditcard 1 4649.5 4830.8  
## - OnlineSecurity\_Yes 1 4649.8 4831.2  
## - DeviceProtection\_Yes 1 4650.1 4831.4  
## - TechSupport\_Yes 1 4650.8 4832.1  
## - Dependents\_Yes 1 4651.0 4832.4  
## - SeniorCitizen 1 4651.7 4833.0  
## - MonthlyCharges 1 4655.7 4837.0  
## <none> 4648.1 4838.1  
## - StreamingTV\_Yes 1 4659.3 4840.6  
## - StreamingMovies\_Yes 1 4660.3 4841.6  
## - annual 1 4661.1 4842.5  
## - MultipleLines\_Yes 1 4662.4 4843.7  
## - ec 1 4663.0 4844.3  
## - PaperlessBilling\_Yes 1 4663.4 4844.8  
## - InternetService\_No 1 4668.1 4849.5  
## - InternetService\_DSL 1 4668.2 4849.6  
## - TotalCharges 1 4671.4 4852.8  
## - monthlycontract 1 4714.9 4896.3  
## - tenure 1 4742.7 4924.0  
##   
## Step: AIC=4829.5  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## gender\_Male + Dependents\_Yes + PhoneService\_Yes + MultipleLines\_Yes +   
## InternetService\_DSL + InternetService\_No + OnlineSecurity\_Yes +   
## DeviceProtection\_Yes + TechSupport\_Yes + StreamingTV\_Yes +   
## StreamingMovies\_Yes + PaperlessBilling\_Yes + creditcard +   
## ec + monthlycontract + annual  
##   
## Df Deviance AIC  
## - gender\_Male 1 4648.2 4820.9  
## - PhoneService\_Yes 1 4649.3 4822.0  
## - creditcard 1 4649.5 4822.2  
## - OnlineSecurity\_Yes 1 4649.9 4822.6  
## - DeviceProtection\_Yes 1 4650.2 4822.9  
## - TechSupport\_Yes 1 4650.8 4823.5  
## - Dependents\_Yes 1 4651.3 4824.0  
## - SeniorCitizen 1 4651.9 4824.6  
## - MonthlyCharges 1 4655.8 4828.5  
## <none> 4648.2 4829.5  
## - StreamingTV\_Yes 1 4659.4 4832.1  
## - StreamingMovies\_Yes 1 4660.4 4833.1  
## - annual 1 4661.2 4833.9  
## - MultipleLines\_Yes 1 4662.5 4835.2  
## - ec 1 4663.1 4835.8  
## - PaperlessBilling\_Yes 1 4663.5 4836.2  
## - InternetService\_No 1 4668.3 4841.0  
## - InternetService\_DSL 1 4668.4 4841.1  
## - TotalCharges 1 4671.5 4844.2  
## - monthlycontract 1 4715.0 4887.7  
## - tenure 1 4743.2 4915.9  
##   
## Step: AIC=4820.93  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## Dependents\_Yes + PhoneService\_Yes + MultipleLines\_Yes + InternetService\_DSL +   
## InternetService\_No + OnlineSecurity\_Yes + DeviceProtection\_Yes +   
## TechSupport\_Yes + StreamingTV\_Yes + StreamingMovies\_Yes +   
## PaperlessBilling\_Yes + creditcard + ec + monthlycontract +   
## annual  
##   
## Df Deviance AIC  
## - PhoneService\_Yes 1 4649.3 4813.4  
## - creditcard 1 4649.6 4813.7  
## - OnlineSecurity\_Yes 1 4649.9 4814.0  
## - DeviceProtection\_Yes 1 4650.2 4814.3  
## - TechSupport\_Yes 1 4650.9 4815.0  
## - Dependents\_Yes 1 4651.4 4815.4  
## - SeniorCitizen 1 4652.0 4816.1  
## - MonthlyCharges 1 4655.8 4819.9  
## <none> 4648.2 4820.9  
## - StreamingTV\_Yes 1 4659.4 4823.5  
## - StreamingMovies\_Yes 1 4660.5 4824.5  
## - annual 1 4661.2 4825.3  
## - MultipleLines\_Yes 1 4662.6 4826.6  
## - ec 1 4663.2 4827.2  
## - PaperlessBilling\_Yes 1 4663.6 4827.6  
## - InternetService\_No 1 4668.3 4832.4  
## - InternetService\_DSL 1 4668.4 4832.5  
## - TotalCharges 1 4671.5 4835.6  
## - monthlycontract 1 4715.0 4879.1  
## - tenure 1 4743.3 4907.3  
##   
## Step: AIC=4813.37  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## Dependents\_Yes + MultipleLines\_Yes + InternetService\_DSL +   
## InternetService\_No + OnlineSecurity\_Yes + DeviceProtection\_Yes +   
## TechSupport\_Yes + StreamingTV\_Yes + StreamingMovies\_Yes +   
## PaperlessBilling\_Yes + creditcard + ec + monthlycontract +   
## annual  
##   
## Df Deviance AIC  
## - DeviceProtection\_Yes 1 4650.3 4805.7  
## - creditcard 1 4650.7 4806.1  
## - Dependents\_Yes 1 4652.5 4807.9  
## - SeniorCitizen 1 4653.0 4808.4  
## - OnlineSecurity\_Yes 1 4654.5 4810.0  
## - TechSupport\_Yes 1 4656.5 4811.9  
## <none> 4649.3 4813.4  
## - annual 1 4662.3 4817.7  
## - MultipleLines\_Yes 1 4664.0 4819.4  
## - ec 1 4664.1 4819.5  
## - PaperlessBilling\_Yes 1 4664.6 4820.0  
## - StreamingTV\_Yes 1 4665.9 4821.4  
## - MonthlyCharges 1 4666.6 4822.1  
## - StreamingMovies\_Yes 1 4668.4 4823.8  
## - TotalCharges 1 4672.1 4827.5  
## - InternetService\_DSL 1 4690.9 4846.4  
## - InternetService\_No 1 4708.3 4863.7  
## - monthlycontract 1 4715.9 4871.3  
## - tenure 1 4745.4 4900.8  
##   
## Step: AIC=4805.7  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## Dependents\_Yes + MultipleLines\_Yes + InternetService\_DSL +   
## InternetService\_No + OnlineSecurity\_Yes + TechSupport\_Yes +   
## StreamingTV\_Yes + StreamingMovies\_Yes + PaperlessBilling\_Yes +   
## creditcard + ec + monthlycontract + annual  
##   
## Df Deviance AIC  
## - creditcard 1 4651.7 4798.5  
## - Dependents\_Yes 1 4653.4 4800.2  
## - SeniorCitizen 1 4654.1 4800.9  
## - OnlineSecurity\_Yes 1 4656.3 4803.1  
## - TechSupport\_Yes 1 4657.9 4804.6  
## <none> 4650.3 4805.7  
## - annual 1 4663.1 4809.9  
## - MultipleLines\_Yes 1 4664.1 4810.9  
## - ec 1 4664.9 4811.7  
## - PaperlessBilling\_Yes 1 4665.5 4812.3  
## - StreamingTV\_Yes 1 4666.2 4813.0  
## - MonthlyCharges 1 4666.8 4813.6  
## - StreamingMovies\_Yes 1 4668.6 4815.4  
## - TotalCharges 1 4673.5 4820.3  
## - InternetService\_DSL 1 4692.0 4838.8  
## - InternetService\_No 1 4709.5 4856.3  
## - monthlycontract 1 4716.1 4862.9  
## - tenure 1 4746.1 4892.9  
##   
## Step: AIC=4798.46  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## Dependents\_Yes + MultipleLines\_Yes + InternetService\_DSL +   
## InternetService\_No + OnlineSecurity\_Yes + TechSupport\_Yes +   
## StreamingTV\_Yes + StreamingMovies\_Yes + PaperlessBilling\_Yes +   
## ec + monthlycontract + annual  
##   
## Df Deviance AIC  
## - Dependents\_Yes 1 4654.8 4793.0  
## - SeniorCitizen 1 4655.4 4793.6  
## - OnlineSecurity\_Yes 1 4657.7 4795.9  
## - TechSupport\_Yes 1 4659.2 4797.4  
## <none> 4651.7 4798.5  
## - annual 1 4664.5 4802.6  
## - MultipleLines\_Yes 1 4665.3 4803.5  
## - PaperlessBilling\_Yes 1 4666.4 4804.6  
## - StreamingTV\_Yes 1 4667.5 4805.6  
## - MonthlyCharges 1 4668.2 4806.4  
## - StreamingMovies\_Yes 1 4669.9 4808.0  
## - ec 1 4673.9 4812.0  
## - TotalCharges 1 4675.5 4813.7  
## - InternetService\_DSL 1 4693.1 4831.2  
## - InternetService\_No 1 4710.3 4848.5  
## - monthlycontract 1 4717.6 4855.8  
## - tenure 1 4749.9 4888.1  
##   
## Step: AIC=4793.01  
## Churn\_Yes ~ SeniorCitizen + tenure + MonthlyCharges + TotalCharges +   
## MultipleLines\_Yes + InternetService\_DSL + InternetService\_No +   
## OnlineSecurity\_Yes + TechSupport\_Yes + StreamingTV\_Yes +   
## StreamingMovies\_Yes + PaperlessBilling\_Yes + ec + monthlycontract +   
## annual  
##   
## Df Deviance AIC  
## - SeniorCitizen 1 4659.9 4789.5  
## - OnlineSecurity\_Yes 1 4661.0 4790.6  
## - TechSupport\_Yes 1 4662.3 4791.9  
## <none> 4654.8 4793.0  
## - annual 1 4668.0 4797.5  
## - MultipleLines\_Yes 1 4668.7 4798.2  
## - PaperlessBilling\_Yes 1 4670.1 4799.6  
## - StreamingTV\_Yes 1 4670.5 4800.1  
## - MonthlyCharges 1 4671.9 4801.4  
## - StreamingMovies\_Yes 1 4673.5 4803.0  
## - ec 1 4677.7 4807.2  
## - TotalCharges 1 4679.7 4809.2  
## - InternetService\_DSL 1 4697.3 4826.9  
## - InternetService\_No 1 4715.0 4844.5  
## - monthlycontract 1 4722.7 4852.2  
## - tenure 1 4756.4 4886.0  
##   
## Step: AIC=4789.46  
## Churn\_Yes ~ tenure + MonthlyCharges + TotalCharges + MultipleLines\_Yes +   
## InternetService\_DSL + InternetService\_No + OnlineSecurity\_Yes +   
## TechSupport\_Yes + StreamingTV\_Yes + StreamingMovies\_Yes +   
## PaperlessBilling\_Yes + ec + monthlycontract + annual  
##   
## Df Deviance AIC  
## - OnlineSecurity\_Yes 1 4666.3 4787.2  
## - TechSupport\_Yes 1 4668.1 4789.0  
## <none> 4659.9 4789.5  
## - annual 1 4673.5 4794.4  
## - MultipleLines\_Yes 1 4674.8 4795.7  
## - StreamingTV\_Yes 1 4675.9 4796.8  
## - PaperlessBilling\_Yes 1 4676.0 4796.9  
## - MonthlyCharges 1 4678.0 4798.9  
## - StreamingMovies\_Yes 1 4679.7 4800.6  
## - ec 1 4684.4 4805.3  
## - TotalCharges 1 4684.9 4805.8  
## - InternetService\_DSL 1 4705.2 4826.1  
## - InternetService\_No 1 4723.8 4844.7  
## - monthlycontract 1 4731.3 4852.2  
## - tenure 1 4760.0 4880.9  
##   
## Step: AIC=4787.21  
## Churn\_Yes ~ tenure + MonthlyCharges + TotalCharges + MultipleLines\_Yes +   
## InternetService\_DSL + InternetService\_No + TechSupport\_Yes +   
## StreamingTV\_Yes + StreamingMovies\_Yes + PaperlessBilling\_Yes +   
## ec + monthlycontract + annual  
##   
## Df Deviance AIC  
## - TechSupport\_Yes 1 4673.4 4785.6  
## <none> 4666.3 4787.2  
## - annual 1 4680.7 4793.0  
## - PaperlessBilling\_Yes 1 4683.7 4796.0  
## - MultipleLines\_Yes 1 4685.0 4797.3  
## - StreamingTV\_Yes 1 4688.8 4801.1  
## - TotalCharges 1 4690.9 4803.2  
## - StreamingMovies\_Yes 1 4692.2 4804.5  
## - ec 1 4692.3 4804.5  
## - MonthlyCharges 1 4696.1 4808.3  
## - InternetService\_DSL 1 4734.9 4847.1  
## - monthlycontract 1 4741.4 4853.6  
## - InternetService\_No 1 4748.6 4860.9  
## - tenure 1 4768.2 4880.4  
##   
## Step: AIC=4785.64  
## Churn\_Yes ~ tenure + MonthlyCharges + TotalCharges + MultipleLines\_Yes +   
## InternetService\_DSL + InternetService\_No + StreamingTV\_Yes +   
## StreamingMovies\_Yes + PaperlessBilling\_Yes + ec + monthlycontract +   
## annual  
##   
## Df Deviance AIC  
## <none> 4673.4 4785.6  
## - annual 1 4690.0 4793.6  
## - PaperlessBilling\_Yes 1 4690.7 4794.3  
## - MultipleLines\_Yes 1 4697.1 4800.7  
## - TotalCharges 1 4697.3 4800.9  
## - StreamingTV\_Yes 1 4700.7 4804.3  
## - ec 1 4700.8 4804.5  
## - StreamingMovies\_Yes 1 4704.2 4807.8  
## - MonthlyCharges 1 4719.4 4823.1  
## - monthlycontract 1 4758.2 4861.8  
## - tenure 1 4774.4 4878.0  
## - InternetService\_DSL 1 4775.5 4879.1  
## - InternetService\_No 1 4779.5 4883.2

BIC\_step\_summary <- summary(BIC\_step)  
BIC\_step\_summary$deviance/BIC\_step\_summary$df.residual

## [1] 0.8327492

AIC(BIC\_step)

## [1] 4699.389

BIC(BIC\_step)

## [1] 4785.643

Based on the various general linear model (logistic) the step wise AIC performs the best in terms of MSE, AIC, and BIC criteria. The stepwise AIC chose the following variables to be included in the final logistic model:SeniorCitizen + tenure + MonthlyCharges + TotalCharges + Dependents\_Yes + MultipleLines\_Yes + InternetService\_DSL + InternetService\_No + OnlineSecurity\_Yes + TechSupport\_Yes + StreamingTV\_Yes + StreamingMovies\_Yes + PaperlessBilling\_Yes + ec + monthlycontract + annual. We will test this model to see if it is more accurate then the model produced earlier.

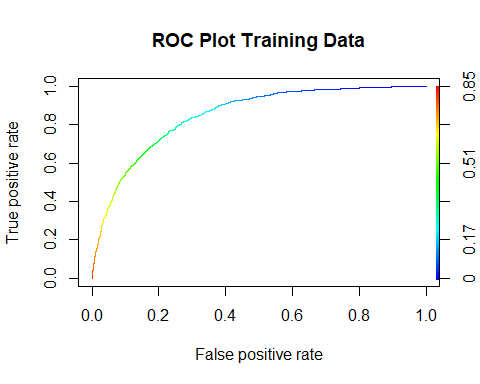
## Testing the best model from the logistic regression

### ROC, AUC, and Asymmetric Cost of the Stepwise AIC model

# ROC curve, in sample prediction  
AIC\_step\_train<- predict(AIC\_step, type="response")  
  
# ROC Curve  
library(ROCR)

## Warning: package 'ROCR' was built under R version 3.6.3

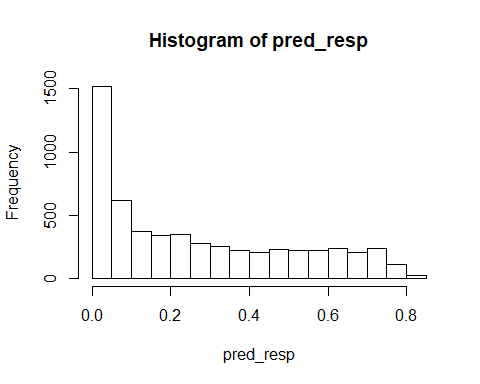
pred <- prediction(AIC\_step\_train, training\_set$Churn\_Yes)  
perf <- performance(pred, "tpr", "fpr")  
plot(perf, colorize=TRUE, main = "ROC Plot Training Data")



#Get the AUC  
unlist(slot(performance(pred, "auc"), "y.values"))

## [1] 0.8490154

# 2X2 misclassification table  
pred\_resp <- predict(AIC\_step,type="response")  
hist(pred\_resp)



table(training\_set$Churn\_Yes, (pred\_resp > 0.5)\*1, dnn=c("Truth","Predicted"))

## Predicted  
## Truth 0 1  
## 0 3705 425  
## 1 667 828

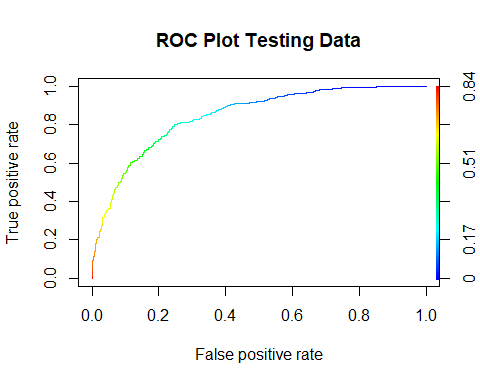
## Symetric cost (misclassification rate) function  
pcut <- 1/2 #prespecify pcut value  
cost1 <- function(r, pi){  
 mean(((r==0)&(pi>pcut)) | ((r==1)&(pi<pcut)))  
}  
  
#Symmetric cost  
cost1(r = training\_set$Churn\_Yes, pi = AIC\_step\_train)

## [1] 0.1941333

The AUC and cost show that this is an effective determiner of whether a customer is retained. It is more effective than the model created earlier and just guessing based on percentage of churn\_yes = 1.

Now we have confirmed the stepwise AIC performs well on the training data, we must now confirm using the testing data set.

# Out-of-sample Testing   
AIC\_step\_test<- predict(AIC\_step, newdata = test\_set, type="response")  
  
# Get ROC curve  
pred <- prediction(AIC\_step\_test, test\_set$Churn\_Yes)  
perf <- performance(pred, "tpr", "fpr")  
plot(perf, colorize=TRUE, main = "ROC Plot Testing Data")



#Get the AUC  
unlist(slot(performance(pred, "auc"), "y.values"))

## [1] 0.8445755

#Asymmetric cost  
cost1(r = test\_set$Churn\_Yes, pi = AIC\_step\_test)

## [1] 0.1933191

The AUC and cost are close to the results from the training set and they are good enough to show this is an effective model.

Next we use cross validation using AUC as cost.

#AUC as cost  
costfunc1 = function(obs, pred.p){  
 pred <- prediction(pred.p, obs)  
 perf <- performance(pred, "tpr", "fpr")  
 cost =unlist(slot(performance(pred, "auc"), "y.values"))  
 return(cost)  
}   
  
library(boot)  
library(ROCR)  
  
## Attempt using glm set to stepwise  
glm1<- glm(Churn\_Yes~SeniorCitizen + tenure + MonthlyCharges + TotalCharges + Dependents\_Yes + MultipleLines\_Yes + InternetService\_DSL + InternetService\_No + OnlineSecurity\_Yes + TechSupport\_Yes + StreamingTV\_Yes + StreamingMovies\_Yes + PaperlessBilling\_Yes + ec + monthlycontract + annual, family=binomial, data=dataset);   
cv\_result <- cv.glm(data=dataset, glmfit=glm1, cost=costfunc1, K=10)   
cv\_result$delta[2]

## [1] 0.8461742

The cross validation confirms the strong results above.

# Conclusion

This project showed the process of importing Telco customer data, performing exploratory data analysis, fitting a variety of models, and then comparing said models to determine best fit. Key indicators like monthly contract and tenure were explained as logically becoming important features. Models included XGBoost and GLM, with stepwise AIC being the best GLM model on multiple metrics, surprassing XGBoost. This model had high AUC (> 0.7) and low symmetric cost, and was further confirmed in cross-validation.

# Appendix

* xgboost() XGboost stands for Xtreme Gradient Boosting, and is a machine-learning model that creates decision trees and then prunes them in iterations until reaching a stopping point. XGboost’s power comes from its ability to generate trees much more quickly than a human could, allowing for the machine to continuously decrease prediction error until a stopping condition is reached. *dummy\_cols()* subset() *skim()* corr\_var() *xgboost()* as.matrix() *createfolds()* lapply() *AIC()* BIC() *step()* performance() *unlist()* cv.glm()

## Functions used that were covered in class

* c()
* ggplot()
* summary()
* str()
* aes()
* function()
* ifelse()
* set.seed()
* hist()
* predict()
* table()
* head()
* is.na()
* glm()