Churn

Faith Taylor

2025-04-09

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tinytex)
library(ggplot2)
library(readr)
```

```
Churn_data <- read.csv("C:/Users/Faith/OneDrive/Grad School/Semester 1 Working Files/Churn_Trai
n.csv") #read the data file
summary(Churn_data) #give summary statistics
```

```
##     state           account_length     area_code         international_plan
##  Length:3333        Min.   :-209.00   Length:3333        Length:3333
##  Class :character   1st Qu.:  72.00   Class :character   Class :character
##  Mode  :character   Median : 100.00   Mode  :character   Mode  :character
##                     Mean   :  97.32
##                     3rd Qu.: 127.00
##                     Max.   : 243.00
##                     NA's   :501
##  voice_mail_plan    number_vmail_messages total_day_minutes total_day_calls
##  Length:3333        Min.   :-10.000       Min.   :   0.0    Min.   :  0.0
##  Class :character   1st Qu.:  0.000       1st Qu.: 149.3    1st Qu.: 87.0
##  Mode  :character   Median :  0.000       Median : 190.5    Median :101.0
##                     Mean   :  7.333       Mean   : 418.9    Mean   :100.3
##                     3rd Qu.: 16.000       3rd Qu.: 237.8    3rd Qu.:114.0
##                     Max.   : 51.000       Max.   :2185.1    Max.   :165.0
##                     NA's   :200           NA's   :200       NA's   :200
##  total_day_charge total_eve_minutes total_eve_calls total_eve_charge
##  Min.   : 0.00    Min.   :   0.0    Min.   :  0.0   Min.   : 0.00
##  1st Qu.:24.45    1st Qu.: 170.5    1st Qu.: 87.0   1st Qu.:14.14
##  Median :30.65    Median : 209.9    Median :100.0   Median :17.09
##  Mean   :30.63    Mean   : 324.3    Mean   :100.1   Mean   :17.08
##  3rd Qu.:36.84    3rd Qu.: 257.6    3rd Qu.:114.0   3rd Qu.:20.00
##  Max.   :59.64    Max.   :1244.2    Max.   :170.0   Max.   :30.91
##  NA's   :200      NA's   :301       NA's   :200     NA's   :200
##  total_night_minutes total_night_calls total_night_charge total_intl_minutes
##  Min.   : 23.2       Min.   : 33.0     Min.   : 1.040     Min.   : 0.00
##  1st Qu.:167.3       1st Qu.: 87.0     1st Qu.: 7.530     1st Qu.: 8.50
##  Median :201.4       Median :100.0     Median : 9.060     Median :10.30
##  Mean   :201.2       Mean   :100.1     Mean   : 9.054     Mean   :10.23
##  3rd Qu.:235.3       3rd Qu.:113.0     3rd Qu.:10.590     3rd Qu.:12.10
##  Max.   :395.0       Max.   :175.0     Max.   :17.770     Max.   :20.00
##  NA's   :200                           NA's   :200        NA's   :200
##  total_intl_calls total_intl_charge number_customer_service_calls
##  Min.   : 0.00    Min.   :0.000     Min.   :0.000
##  1st Qu.: 3.00    1st Qu.:2.300     1st Qu.:1.000
##  Median : 4.00    Median :2.780     Median :1.000
##  Mean   : 4.47    Mean   :2.762     Mean   :1.561
##  3rd Qu.: 6.00    3rd Qu.:3.270     3rd Qu.:2.000
##  Max.   :20.00    Max.   :5.400     Max.   :9.000
##  NA's   :301      NA's   :200       NA's   :200
##     churn
##  Length:3333
##  Class :character
##  Mode  :character
##
##
##
##
```

#PART 1: Cleaning the data

```
colSums(is.na(Churn_data))/nrow(Churn_data) * 100 #calculating the percent of na values in each column in the dataset
```

```
##                         state                equ_length
##                      0.000000                 15.031503
##                     area_code            international_plan
##                      0.000000                  0.000000
##                voice_mail_plan         number_vmail_messages
##                      0.000000                  6.000600
##              total_day_minutes               total_day_calls
##                      6.000600                  6.000600
##               total_day_charge             total_eve_minutes
##                      6.000600                  9.030903
##                total_eve_calls              total_eve_charge
##                      6.000600                  6.000600
##            total_night_minutes             total_night_calls
##                      6.000600                  0.000000
##             total_night_charge            total_intl_minutes
##                      6.000600                  6.000600
##               total_intl_calls             total_intl_charge
##                      9.030903                  6.000600
## number_customer_service_calls                     churn
##                      6.000600                  0.000000
```

```
sum(Churn_data$number_vmail_messages < 0, na.rm = TRUE) #Number of voicemail messages should not be negative. Finding the sum of negative values for this variable
```

```
## [1] 201
```

```
sum(Churn_data$account_length < 0, na.rm = TRUE) #Account length should not be negative. Finding the sum of negative values for this variable
```

```
## [1] 51
```

#we have the option to replace negative values with 0 or mark them as na and then remove.

```
Churn_data$number_vmail_messages[Churn_data$number_vmail_messages < 0] #want to see the actual values that are negative in this column. Due to the range in the data, don't want to assume that these are meant to be 0. They may be a typo of positive values, so will exclude instead.
```

```
##   [1]  NA  -4  NA  NA  -2  NA  -7  -5  -7  -3  -8  NA  NA  NA  -5  -5  -3  NA
##  [19]  -3  -1  NA  -3  NA  -2  -6  -4  NA  NA  -6  NA -10  NA  NA  -9  NA  NA
##  [37]  -3  NA  -8  NA -10  NA  -5  NA  -7  NA  -3  -2  -8  -5  -2  NA  NA -10
##  [55]  NA  NA  NA -10  NA  -5  NA  -7  -9  -4  NA  NA  -1  NA  NA  NA  -7  NA
##  [73]  NA  -8  -1  NA  -2  -2  -2  -3  NA  NA  -1  NA  NA  -1  NA  -6  -6  NA
##  [91]  -5  NA  -5  NA  -1  -9  NA  -8  NA  -3  NA  NA  -6  NA  -8  NA  NA  NA
## [109]  NA  NA  -3  NA  NA  NA  NA -10  -9  -7 -10  NA  NA  NA  NA  -4  NA  -6
## [127]  NA -10  NA  NA  NA  -4  NA  -5  NA -10 -10 -10  -8  NA  NA  -1  NA  NA
## [145]  NA  NA -10  NA  NA  NA  NA  NA  -1  NA  NA  -8  -8  -6  NA  -6  NA  -5
## [163]  -3  -4  -9  -2  -3  NA  NA  NA  -5  -8  -3  -9  -2  NA  NA  -7  -8  -3
## [181]  -4  NA  NA  NA  NA  -6  NA  NA  NA  -9  -6  -7  NA -10  -8  NA  NA  -6
## [199]  NA  -8  NA  NA  NA  NA  -3  NA  NA  NA -10  NA  -8  NA  NA  NA -10  NA
## [217]  -5  NA  -9  NA  NA  -8  NA  -9  NA -10  NA  -2  NA  -8  NA  -6  -8  -4
## [235]  NA  -1  NA  NA  -9  NA  -7  -7  -9  NA  NA  NA  NA  NA  NA  NA  -1  NA
## [253]  NA  NA  NA  -7  NA  -3  -6  -2  NA  -9  -7  -2  -6  NA  -6  NA  NA  NA
## [271]  -5  NA  -1  NA -10  NA  NA  NA  NA  NA  -3  -2  NA  NA  -7  NA  NA  NA
## [289]  NA  -2  NA  -1  -1  -4  -3  -8 -10  NA  -3  -9  -9  -7  NA  NA  -3  NA
## [307]  NA  NA  -9  -5  -6  -9  NA  -7  NA  -2  -7  -8  NA  NA  NA  NA  -5  NA
## [325]  -6 -10 -10 -10  NA  NA  NA  -7  -6  -3  -1  NA -10 -10  NA  -5  -3  -3
## [343]  NA  NA  -3  -1  -1  -2 -10  NA  -1  NA  NA  -4  NA  NA  NA  -9  -2  -6
## [361]  -6  -5  -1  NA  -3  -5  NA  NA  NA  -4  -8  -8  NA -10  NA  -6  -9  NA
## [379]  -3  NA  -2  -8  -8  NA  -4  -3  NA  NA  -1  NA  -6  NA  NA  -2  -8  NA
## [397]  -8  NA  NA  -7  NA
```

```
Churn_data$account_length[Churn_data$account_length < 0] #want to see the actual values that are negative in this column. Due to the range in the data, don't want to assume that these are meant to be 0. They may be a typo of positive values, so will exclude instead.
```

```
##   [1]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
##  [16]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA  -80  -72   NA   NA   NA
##  [31]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
##  [46]  -12   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
##  [61]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
##  [76]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
##  [91]   NA   NA  -53   NA   NA   NA   NA   NA   NA   NA  -59   NA   NA   NA   NA
## [106]   NA   NA   NA   NA   NA   NA   NA   NA   NA  -78   NA   NA   NA -128   NA
## [121]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA -164   NA   NA   NA   NA
## [136]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## [151]   NA   NA   NA   NA -111   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## [166]   NA   NA   NA   NA   NA   NA   NA   NA   NA -132   NA   NA   NA -121   NA
## [181]   NA   NA   NA -209   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## [196]   NA   NA   NA   NA   NA   NA  -75   NA   NA   NA   NA   NA   NA   NA   NA
## [211]   NA   NA   NA -101   NA   NA   NA   NA   NA   NA -126   NA   NA   NA   NA
## [226]   NA   NA  -45  -95   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## [241]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## [256]   NA   NA   NA   NA   NA -138   NA   NA   NA   NA   NA   NA   NA   NA   NA
## [271]   NA  -93   NA   NA  -44   NA   NA   NA   NA -108 -104   NA   NA   NA   NA
## [286]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA  -78   NA
## [301]   NA   NA   NA   NA   NA   NA   NA   NA -121   NA   NA   NA   NA   NA   NA
## [316]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## [331]   NA   NA   NA  -68   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA  -82
## [346]   NA   NA  -50   NA  -68   NA   NA   NA   NA   NA   NA   NA -103   NA -142
## [361]   NA   NA  -74   NA -145   NA   NA -107 -121   NA   NA   NA   NA   NA   NA
## [376]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## [391]   NA   NA   NA   NA   NA   NA   NA -148   NA   NA   NA   NA   NA   NA   NA
## [406]  -73   NA   NA   NA   NA   NA   NA -115   NA   NA -105   NA   NA   NA   NA
## [421]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA  -69   NA   NA   NA   NA
## [436]   NA   NA -130   NA  -80   NA  -92   NA   NA   NA   NA   NA   NA   NA   NA
## [451]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA  -42   NA
## [466]   NA   NA   NA   NA   NA   NA  -95   NA   NA   NA   NA   NA   NA   NA -174
## [481]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
## [496]   NA   NA   NA   NA  -16   NA   NA -127   NA   NA   NA   NA   NA   NA   NA
## [511] -137 -132   NA   NA   NA   NA   NA   NA   NA   NA  -67   NA   NA   NA   NA
## [526]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA  -74   NA   NA   NA   NA
## [541]   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
```

```
Churn_data2 <- Churn_data #created an additonal dataset to start cleaning the data.
Churn_data2[Churn_data2 <0] <- NA #removing all negative values in the dataset
Churndata_cleaned <- na.omit(Churn_data2) #all NA values removed in the dataset
summary(Churndata_cleaned) #summary statistics of clean dataset
```

```
##     state           account_length   area_code         international_plan
##  Length:2378        Min.   :  1.0   Length:2378        Length:2378
##  Class :character   1st Qu.: 73.0   Class :character   Class :character
##  Mode  :character   Median :101.0   Mode  :character   Mode  :character
##                     Mean   :100.6
##                     3rd Qu.:127.0
##                     Max.   :243.0
##  voice_mail_plan    number_vmail_messages total_day_minutes total_day_calls
##  Length:2378        Min.   : 0.00         Min.   :   0.0    Min.   :  0.00
##  Class :character   1st Qu.: 0.00         1st Qu.: 151.6    1st Qu.: 87.00
##  Mode  :character   Median : 0.00         Median : 194.6    Median :100.00
##                     Mean   : 8.09         Mean   : 495.7    Mean   : 99.95
##                     3rd Qu.:19.00         3rd Qu.: 252.0    3rd Qu.:114.00
##                     Max.   :51.00         Max.   :2185.1    Max.   :165.00
##  total_day_charge total_eve_minutes total_eve_calls total_eve_charge
##  Min.   : 0.00    Min.   :  31.2    Min.   : 12.0   Min.   : 2.65
##  1st Qu.:24.58    1st Qu.: 172.9    1st Qu.: 87.0   1st Qu.:14.20
##  Median :30.84    Median : 213.1    Median :100.5   Median :17.12
##  Mean   :30.79    Mean   : 358.5    Mean   :100.2   Mean   :17.12
##  3rd Qu.:36.91    3rd Qu.: 267.3    3rd Qu.:114.0   3rd Qu.:20.00
##  Max.   :59.64    Max.   :1244.2    Max.   :170.0   Max.   :30.91
##  total_night_minutes total_night_calls total_night_charge total_intl_minutes
##  Min.   : 43.7       Min.   : 33.00    Min.   : 1.970     Min.   : 0.0
##  1st Qu.:167.2       1st Qu.: 86.00    1st Qu.: 7.522     1st Qu.: 8.4
##  Median :200.4       Median :100.00    Median : 9.020     Median :10.2
##  Mean   :200.9       Mean   : 99.74    Mean   : 9.039     Mean   :10.2
##  3rd Qu.:235.0       3rd Qu.:113.00    3rd Qu.:10.578     3rd Qu.:12.0
##  Max.   :367.7       Max.   :164.00    Max.   :16.550     Max.   :20.0
##  total_intl_calls total_intl_charge number_customer_service_calls
##  Min.   : 0.000   Min.   :0.000     Min.   :0.000
##  1st Qu.: 3.000   1st Qu.:2.270     1st Qu.:1.000
##  Median : 4.000   Median :2.750     Median :1.000
##  Mean   : 4.489   Mean   :2.753     Mean   :1.556
##  3rd Qu.: 6.000   3rd Qu.:3.240     3rd Qu.:2.000
##  Max.   :19.000   Max.   :5.400     Max.   :8.000
##     churn
##  Length:2378
##  Class :character
##  Mode  :character
##
##
##
```

#chose to remove na values rather than replace (impute) them since the mean and median are not in alignment for some of the variables - number_vmail_messages and total_dat_minutes

```
#Identify unique values in non-numerical columns; we can remove state and areacode from the data set & make international plan, voicemail plan, and churn categorical before putting into models
non_numeric_columns <- !sapply(Churndata_cleaned, is.numeric)
unique_values <- sapply(Churndata_cleaned[, non_numeric_columns], unique)
unique_values
```

```
## $state
##  [1] "NV" "HI" "DC" "OH" "NC" "PA" "IA" "DE" "KY" "MS" "NY" "AR" "AZ" "MT" "OR"
## [16] "IN" "FL" "MD" "TN" "AL" "SD" "WV" "MA" "VA" "WA" "NE" "AK" "MN" "NM" "GA"
## [31] "UT" "LA" "KS" "WI" "OK" "ME" "TX" "NJ" "WY" "ID" "VT" "MI" "RI" "CT" "CA"
## [46] "CO" "MO" "IL" "ND" "NH" "SC"
##
## $area_code
## [1] "area_code_510" "area_code_415" "area_code_408"
##
## $international_plan
## [1] "no"  "yes"
##
## $voice_mail_plan
## [1] "no"  "yes"
##
## $churn
## [1] "no"  "yes"
```

```
Churndata_cleaned2 <- Churndata_cleaned[, -c(1,3)] #remove area code and state from the dataset
summary(Churndata_cleaned2)
```

```
##  account_length  international_plan voice_mail_plan    number_vmail_messages
##  Min.   :  1.0   Length:2378        Length:2378        Min.   : 0.00
##  1st Qu.: 73.0   Class :character   Class :character   1st Qu.: 0.00
##  Median :101.0   Mode  :character   Mode  :character   Median : 0.00
##  Mean   :100.6                                         Mean   : 8.09
##  3rd Qu.:127.0                                         3rd Qu.:19.00
##  Max.   :243.0                                         Max.   :51.00
##  total_day_minutes total_day_calls  total_day_charge total_eve_minutes
##  Min.   :   0.0    Min.   :  0.00   Min.   : 0.00    Min.   :  31.2
##  1st Qu.: 151.6    1st Qu.: 87.00   1st Qu.:24.58    1st Qu.: 172.9
##  Median : 194.6    Median :100.00   Median :30.84    Median : 213.1
##  Mean   : 495.7    Mean   : 99.95   Mean   :30.79    Mean   : 358.5
##  3rd Qu.: 252.0    3rd Qu.:114.00   3rd Qu.:36.91    3rd Qu.: 267.3
##  Max.   :2185.1    Max.   :165.00   Max.   :59.64    Max.   :1244.2
##  total_eve_calls total_eve_charge total_night_minutes total_night_calls
##  Min.   : 12.0   Min.   : 2.65    Min.   : 43.7       Min.   : 33.00
##  1st Qu.: 87.0   1st Qu.:14.20    1st Qu.:167.2       1st Qu.: 86.00
##  Median :100.5   Median :17.12    Median :200.4       Median :100.00
##  Mean   :100.2   Mean   :17.12    Mean   :200.9       Mean   : 99.74
##  3rd Qu.:114.0   3rd Qu.:20.00    3rd Qu.:235.0       3rd Qu.:113.00
##  Max.   :170.0   Max.   :30.91    Max.   :367.7       Max.   :164.00
##  total_night_charge total_intl_minutes total_intl_calls total_intl_charge
##  Min.   : 1.970     Min.   : 0.0       Min.   : 0.000   Min.   :0.000
##  1st Qu.: 7.522     1st Qu.: 8.4       1st Qu.: 3.000   1st Qu.:2.270
##  Median : 9.020     Median :10.2       Median : 4.000   Median :2.750
##  Mean   : 9.039     Mean   :10.2       Mean   : 4.489   Mean   :2.753
##  3rd Qu.:10.578     3rd Qu.:12.0       3rd Qu.: 6.000   3rd Qu.:3.240
##  Max.   :16.550     Max.   :20.0       Max.   :19.000   Max.   :5.400
##  number_customer_service_calls    churn
##  Min.   :0.000                 Length:2378
##  1st Qu.:1.000                 Class :character
##  Median :1.000                 Mode  :character
##  Mean   :1.556
##  3rd Qu.:2.000
##  Max.   :8.000
```

```
Churndata_cleaned2[sapply(Churndata_cleaned2, is.character)] <- lapply(Churndata_cleaned2[sapply
(Churndata_cleaned2, is.character)], factor) #make remaining variables factors
summary(Churndata_cleaned2)
```

```
##  account_length  international_plan voice_mail_plan number_vmail_messages
##  Min.   :  1.0   no :2156           no :1718        Min.   : 0.00
##  1st Qu.: 73.0   yes: 222           yes: 660        1st Qu.: 0.00
##  Median :101.0                                      Median : 0.00
##  Mean   :100.6                                      Mean   : 8.09
##  3rd Qu.:127.0                                      3rd Qu.:19.00
##  Max.   :243.0                                      Max.   :51.00
##  total_day_minutes total_day_calls  total_day_charge total_eve_minutes
##  Min.   :   0.0    Min.   :  0.00   Min.   : 0.00    Min.   :  31.2
##  1st Qu.: 151.6    1st Qu.: 87.00   1st Qu.:24.58    1st Qu.: 172.9
##  Median : 194.6    Median :100.00   Median :30.84    Median : 213.1
##  Mean   : 495.7    Mean   : 99.95   Mean   :30.79    Mean   : 358.5
##  3rd Qu.: 252.0    3rd Qu.:114.00   3rd Qu.:36.91    3rd Qu.: 267.3
##  Max.   :2185.1    Max.   :165.00   Max.   :59.64    Max.   :1244.2
##  total_eve_calls total_eve_charge total_night_minutes total_night_calls
##  Min.   : 12.0   Min.   : 2.65    Min.   : 43.7       Min.   : 33.00
##  1st Qu.: 87.0   1st Qu.:14.20    1st Qu.:167.2       1st Qu.: 86.00
##  Median :100.5   Median :17.12    Median :200.4       Median :100.00
##  Mean   :100.2   Mean   :17.12    Mean   :200.9       Mean   : 99.74
##  3rd Qu.:114.0   3rd Qu.:20.00    3rd Qu.:235.0       3rd Qu.:113.00
##  Max.   :170.0   Max.   :30.91    Max.   :367.7       Max.   :164.00
##  total_night_charge total_intl_minutes total_intl_calls total_intl_charge
##  Min.   : 1.970     Min.   : 0.0       Min.   : 0.000   Min.   :0.000
##  1st Qu.: 7.522     1st Qu.: 8.4       1st Qu.: 3.000   1st Qu.:2.270
##  Median : 9.020     Median :10.2       Median : 4.000   Median :2.750
##  Mean   : 9.039     Mean   :10.2       Mean   : 4.489   Mean   :2.753
##  3rd Qu.:10.578     3rd Qu.:12.0       3rd Qu.: 6.000   3rd Qu.:3.240
##  Max.   :16.550     Max.   :20.0       Max.   :19.000   Max.   :5.400
##  number_customer_service_calls churn
##  Min.   :0.000                 no :2035
##  1st Qu.:1.000                 yes: 343
##  Median :1.000
##  Mean   :1.556
##  3rd Qu.:2.000
##  Max.   :8.000
```

#PART 2 - Trying different methods for modeling

```
library(ISLR)
library(caret)
```

```
## Loading required package: lattice
```

```
set.seed(123)
Index_Train <- createDataPartition(Churndata_cleaned2$churn, p=.7, list = FALSE)
Churn_Train <- Churndata_cleaned2[Index_Train,]
Churn_Test <-  Churndata_cleaned2[-Index_Train,]
Model_Regression <- glm(churn ~ ., Churn_Train, family = "binomial" ) #glm used here since the r
esponding variable (churn) is not continuous
summary(Model_Regression) #variables with lowest z scores are international_plan, and number_cus
tomer_service_calls
```

```
##
## Call:
## glm(formula = churn ~ ., family = "binomial", data = Churn_Train)
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -8.8274266  1.0375700  -8.508  < 2e-16 ***
## account_length               0.0008747  0.0019986   0.438  0.66162
## international_planyes        2.0723188  0.2095638   9.889  < 2e-16 ***
## voice_mail_planyes           -2.6069518  0.8571117  -3.042  0.00235 **
## number_vmail_messages         0.0485062  0.0265152   1.829  0.06734 .
## total_day_minutes            -0.0016526  0.0029843  -0.554  0.57973
## total_day_calls               0.0012566  0.0040234   0.312  0.75479
## total_day_charge              0.0861433  0.0180262   4.779 1.76e-06 ***
## total_eve_minutes             0.0029187  0.0059135   0.494  0.62161
## total_eve_calls              -0.0006223  0.0040174  -0.155  0.87691
## total_eve_charge              0.0739873  0.0713344   1.037  0.29965
## total_night_minutes          -0.0971300  1.2639217  -0.077  0.93874
## total_night_calls             0.0030187  0.0041054   0.735  0.46216
## total_night_charge            2.2351208 28.0857145   0.080  0.93657
## total_intl_minutes           -2.7870725  7.6424988  -0.365  0.71535
## total_intl_calls             -0.1041456  0.0367619  -2.833  0.00461 **
## total_intl_charge            10.6931064 28.3029682   0.378  0.70557
## number_customer_service_calls  0.4727469  0.0557582   8.479  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1377.2  on 1665  degrees of freedom
## Residual deviance: 1063.4  on 1648  degrees of freedom
## AIC: 1099.4
##
## Number of Fisher Scoring iterations: 6
```

```
library(pROC) #add library for roc function
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
Predicted_Values <- predict(Model_Regression, Churn_Test, type = 'response') #gives us the AUC, or area under the curve, which at .8264 means the model is pretty accurate
roc(Churn_Test$churn, Predicted_Values)
```

```
## Setting levels: control = no, case = yes
```

```
## Setting direction: controls < cases
```

```
##
## Call:
## roc.default(response = Churn_Test$churn, predictor = Predicted_Values)
##
## Data: Predicted_Values in 610 controls (Churn_Test$churn no) < 102 cases (Churn_Test$churn ye
s).
## Area under the curve: 0.8264
```

```
library(rpart)
library(rpart.plot)
```

```
## Warning: package 'rpart.plot' was built under R version 4.4.2
```

```
Model_Decision <- rpart(churn ~ ., Churn_Train, method = 'class') #model for decision tree
summary(Model_Decision) #summary which includes order of importance of variables
```

```
## Call:
## rpart(formula = churn ~ ., data = Churn_Train, method = "class")
##   n= 1666
##
##            CP nsplit rel error    xerror       xstd
## 1  0.07468880      0 1.0000000 1.0000000 0.05957464
## 2  0.07261411      2 0.8506224 0.9377593 0.05799371
## 3  0.04979253      4 0.7053942 0.8049793 0.05432506
## 4  0.04149378      7 0.5311203 0.6473029 0.04933967
## 5  0.03319502      8 0.4896266 0.6141079 0.04818504
## 6  0.02489627      9 0.4564315 0.6016598 0.04774128
## 7  0.02282158     10 0.4315353 0.5767635 0.04683520
## 8  0.02074689     12 0.3858921 0.5643154 0.04637251
## 9  0.01659751     13 0.3651452 0.5643154 0.04637251
## 10 0.01000000     15 0.3319502 0.4896266 0.04344822
##
## Variable importance
##              total_day_charge number_customer_service_calls
##                            20                            12
##              total_eve_charge             total_intl_charge
##                            11                             7
##            total_intl_minutes             total_day_minutes
##                             7                             7
##              total_intl_calls             total_eve_minutes
##                             6                             6
##            international_plan         number_vmail_messages
##                             6                             5
##               voice_mail_plan           total_night_minutes
##                             4                             3
##            total_night_charge             total_night_calls
##                             2                             1
##               total_day_calls
##                             1
##
## Node number 1: 1666 observations,    complexity param=0.0746888
##   predicted class=no   expected loss=0.1446579  P(node) =1
##     class counts:  1425   241
##    probabilities: 0.855 0.145
##   left son=2 (1465 obs) right son=3 (201 obs)
##   Primary splits:
##       total_day_charge              < 41.64   to the left,  improve=41.999570, (0 missing)
##       number_customer_service_calls < 3.5     to the left,  improve=39.956020, (0 missing)
##       international_plan            splits as  LR,          improve=28.846700, (0 missing)
##       total_day_minutes             < 223.45  to the left,  improve=13.634400, (0 missing)
##       voice_mail_plan               splits as  RL,          improve= 5.925315, (0 missing)
##
## Node number 2: 1465 observations,    complexity param=0.07261411
##   predicted class=no   expected loss=0.1030717  P(node) =0.8793517
##     class counts:  1314   151
##    probabilities: 0.897 0.103
##   left son=4 (1351 obs) right son=5 (114 obs)
##   Primary splits:
```

```
##       number_customer_service_calls < 3.5     to the left,  improve=44.289320, (0 missing)
##       international_plan            splits as  LR,          improve=23.000150, (0 missing)
##       total_eve_charge              < 28.545  to the left,  improve= 4.972699, (0 missing)
##       total_intl_calls              < 3.5     to the right, improve= 4.616694, (0 missing)
##       total_day_charge              < 37.99   to the left,  improve= 3.756133, (0 missing)
##
## Node number 3: 201 observations,    complexity param=0.0746888
##   predicted class=no   expected loss=0.4477612  P(node) =0.1206483
##     class counts:   111    90
##    probabilities: 0.552 0.448
##   left son=6 (139 obs) right son=7 (62 obs)
##   Primary splits:
##       total_eve_charge      < 18.895  to the left,  improve=21.04165, (0 missing)
##       voice_mail_plan       splits as  RL,          improve=17.34439, (0 missing)
##       number_vmail_messages < 5.5     to the right, improve=17.34439, (0 missing)
##       total_eve_minutes     < 199.6   to the left,  improve=15.07462, (0 missing)
##       total_day_charge      < 49.495  to the left,  improve=12.29145, (0 missing)
##   Surrogate splits:
##       total_eve_minutes < 222.3   to the left,  agree=0.896, adj=0.661, (0 split)
##       total_day_charge  < 41.82   to the right, agree=0.711, adj=0.065, (0 split)
##       total_day_minutes < 246     to the right, agree=0.706, adj=0.048, (0 split)
##       total_day_calls   < 134.5   to the left,  agree=0.697, adj=0.016, (0 split)
##       total_night_calls < 131.5   to the left,  agree=0.697, adj=0.016, (0 split)
##
## Node number 4: 1351 observations,    complexity param=0.04979253
##   predicted class=no   expected loss=0.06735751  P(node) =0.8109244
##     class counts:  1260    91
##    probabilities: 0.933 0.067
##   left son=8 (1229 obs) right son=9 (122 obs)
##   Primary splits:
##       international_plan splits as  LR,          improve=21.801800, (0 missing)
##       total_day_charge   < 37.99   to the left,  improve= 5.496487, (0 missing)
##       total_eve_charge   < 25.665  to the left,  improve= 5.454307, (0 missing)
##       total_intl_minutes < 13.15   to the left,  improve= 3.174802, (0 missing)
##       total_intl_charge  < 3.55    to the left,  improve= 3.174802, (0 missing)
##
## Node number 5: 114 observations,    complexity param=0.07261411
##   predicted class=yes  expected loss=0.4736842  P(node) =0.06842737
##     class counts:    54    60
##    probabilities: 0.474 0.526
##   left son=10 (47 obs) right son=11 (67 obs)
##   Primary splits:
##       total_day_charge  < 29.88   to the right, improve=17.930700, (0 missing)
##       total_day_minutes < 161.8   to the right, improve=15.048870, (0 missing)
##       total_eve_minutes < 199.8   to the right, improve= 7.814086, (0 missing)
##       total_eve_charge  < 17.025  to the right, improve= 6.456920, (0 missing)
##       total_intl_calls  < 4.5     to the right, improve= 4.994808, (0 missing)
##   Surrogate splits:
##       total_day_minutes  < 175.75  to the right, agree=0.904, adj=0.766, (0 split)
##       international_plan splits as  RL,          agree=0.632, adj=0.106, (0 split)
##       total_eve_minutes  < 318.25  to the right, agree=0.632, adj=0.106, (0 split)
##       total_intl_minutes < 14.25   to the right, agree=0.632, adj=0.106, (0 split)
```

```
##       total_intl_charge  < 3.85    to the right, agree=0.632, adj=0.106, (0 split)
##
## Node number 6: 139 observations,    complexity param=0.04149378
##   predicted class=no   expected loss=0.294964  P(node) =0.08343337
##     class counts:    98    41
##    probabilities: 0.705 0.295
##   left son=12 (109 obs) right son=13 (30 obs)
##   Primary splits:
##       total_day_charge    < 48.415  to the left,  improve=10.571360, (0 missing)
##       total_night_minutes < 212.7   to the left,  improve= 8.855730, (0 missing)
##       total_night_charge  < 9.57    to the left,  improve= 8.855730, (0 missing)
##       total_day_minutes   < 275.5   to the left,  improve= 7.988065, (0 missing)
##       voice_mail_plan     splits as  RL,          improve= 5.434667, (0 missing)
##   Surrogate splits:
##       total_day_minutes < 284.8   to the left,  agree=0.878, adj=0.433, (0 split)
##       account_length    < 15.5    to the right, agree=0.799, adj=0.067, (0 split)
##       total_night_calls < 137     to the left,  agree=0.799, adj=0.067, (0 split)
##       total_day_calls   < 58.5    to the right, agree=0.791, adj=0.033, (0 split)
##       total_intl_calls  < 1.5     to the right, agree=0.791, adj=0.033, (0 split)
##
## Node number 7: 62 observations,    complexity param=0.03319502
##   predicted class=yes  expected loss=0.2096774  P(node) =0.03721489
##     class counts:    13    49
##    probabilities: 0.210 0.790
##   left son=14 (12 obs) right son=15 (50 obs)
##   Primary splits:
##       voice_mail_plan       splits as  RL,          improve=11.575050, (0 missing)
##       number_vmail_messages < 5.5     to the right, improve=11.575050, (0 missing)
##       total_day_charge      < 44.425  to the left,  improve= 2.470080, (0 missing)
##       total_day_minutes     < 261.3   to the left,  improve= 2.138863, (0 missing)
##       total_night_minutes   < 159.55  to the left,  improve= 1.178822, (0 missing)
##   Surrogate splits:
##       number_vmail_messages         < 5.5     to the right, agree=1.000, adj=1.000, (0 split)
##       number_customer_service_calls < 4.5     to the right, agree=0.855, adj=0.250, (0 split)
##       total_night_calls             < 74      to the left,  agree=0.823, adj=0.083, (0 split)
##
## Node number 8: 1229 observations,    complexity param=0.01659751
##   predicted class=no   expected loss=0.03905614  P(node) =0.7376951
##     class counts:  1181    48
##    probabilities: 0.961 0.039
##   left son=16 (1100 obs) right son=17 (129 obs)
##   Primary splits:
##       total_day_charge  < 38.105  to the left,  improve=3.3766000, (0 missing)
##       total_eve_charge  < 25.665  to the left,  improve=3.1861680, (0 missing)
##       total_day_minutes < 209.3   to the left,  improve=1.0038510, (0 missing)
##       total_eve_minutes < 1204.9  to the left,  improve=0.8696579, (0 missing)
##       account_length    < 208.5   to the left,  improve=0.8566425, (0 missing)
##
## Node number 9: 122 observations,    complexity param=0.04979253
##   predicted class=no   expected loss=0.352459  P(node) =0.07322929
##     class counts:    79    43
##    probabilities: 0.648 0.352
```

```
##   left son=18 (98 obs) right son=19 (24 obs)
##   Primary splits:
##       total_intl_calls    < 2.5     to the right, improve=25.055870, (0 missing)
##       total_intl_minutes  < 13.1    to the left,  improve=21.272680, (0 missing)
##       total_intl_charge   < 3.535   to the left,  improve=21.272680, (0 missing)
##       total_eve_charge    < 14.11   to the left,  improve= 3.820409, (0 missing)
##       total_night_minutes < 203.75  to the right, improve= 2.977559, (0 missing)
##   Surrogate splits:
##       total_day_calls  < 49.5    to the right, agree=0.820, adj=0.083, (0 split)
##       total_eve_charge < 25.37   to the left,  agree=0.811, adj=0.042, (0 split)
##
## Node number 10: 47 observations
##   predicted class=no   expected loss=0.1914894  P(node) =0.02821128
##     class counts:    38     9
##    probabilities: 0.809 0.191
##
## Node number 11: 67 observations,    complexity param=0.02282158
##   predicted class=yes  expected loss=0.238806  P(node) =0.04021609
##     class counts:    16    51
##    probabilities: 0.239 0.761
##   left son=22 (29 obs) right son=23 (38 obs)
##   Primary splits:
##       total_eve_charge  < 18.08   to the right, improve=7.928082, (0 missing)
##       total_eve_minutes < 216.35  to the right, improve=6.929894, (0 missing)
##       total_day_charge  < 22.975  to the right, improve=3.968099, (0 missing)
##       total_day_minutes < 135.15  to the right, improve=3.152495, (0 missing)
##       total_intl_calls  < 4.5     to the right, improve=3.049118, (0 missing)
##   Surrogate splits:
##       total_eve_minutes             < 212.85  to the right, agree=0.940, adj=0.862, (0 split)
##       total_day_minutes             < 1991.9  to the right, agree=0.657, adj=0.207, (0 split)
##       total_intl_calls              < 4.5     to the right, agree=0.627, adj=0.138, (0 split)
##       number_customer_service_calls < 5.5     to the right, agree=0.627, adj=0.138, (0 split)
##       number_vmail_messages         < 16.5    to the right, agree=0.612, adj=0.103, (0 split)
##
## Node number 12: 109 observations,    complexity param=0.02489627
##   predicted class=no   expected loss=0.1926606  P(node) =0.06542617
##     class counts:    88    21
##    probabilities: 0.807 0.193
##   left son=24 (95 obs) right son=25 (14 obs)
##   Primary splits:
##       total_night_minutes < 270.6   to the left,  improve=8.741340, (0 missing)
##       total_night_charge  < 12.18   to the left,  improve=8.741340, (0 missing)
##       total_eve_minutes   < 185.95  to the left,  improve=3.251308, (0 missing)
##       total_eve_charge    < 15.805  to the left,  improve=3.190570, (0 missing)
##       international_plan  splits as  LR,          improve=3.034573, (0 missing)
##   Surrogate splits:
##       total_night_charge < 12.18   to the left,  agree=1, adj=1, (0 split)
##
## Node number 13: 30 observations,    complexity param=0.02074689
##   predicted class=yes  expected loss=0.3333333  P(node) =0.0180072
##     class counts:    10    20
##    probabilities: 0.333 0.667
```

```
##   left son=26 (7 obs) right son=27 (23 obs)
##   Primary splits:
##       voice_mail_plan       splits as  RL,          improve=5.010352, (0 missing)
##       number_vmail_messages < 10.5    to the right, improve=5.010352, (0 missing)
##       total_eve_charge      < 14.2    to the left,  improve=4.444444, (0 missing)
##       account_length        < 67      to the left,  improve=2.650104, (0 missing)
##       total_eve_calls       < 101     to the right, improve=2.500000, (0 missing)
##   Surrogate splits:
##       number_vmail_messages         < 10.5    to the right, agree=1.000, adj=1.000, (0 split)
##       number_customer_service_calls < 2.5     to the right, agree=0.867, adj=0.429, (0 split)
##       total_night_calls             < 68.5    to the left,  agree=0.833, adj=0.286, (0 split)
##
## Node number 14: 12 observations
##   predicted class=no   expected loss=0.1666667  P(node) =0.007202881
##     class counts:    10     2
##    probabilities: 0.833 0.167
##
## Node number 15: 50 observations
##   predicted class=yes  expected loss=0.06  P(node) =0.030012
##     class counts:     3    47
##    probabilities: 0.060 0.940
##
## Node number 16: 1100 observations
##   predicted class=no   expected loss=0.02636364  P(node) =0.6602641
##     class counts:  1071    29
##    probabilities: 0.974 0.026
##
## Node number 17: 129 observations,    complexity param=0.01659751
##   predicted class=no   expected loss=0.1472868  P(node) =0.07743097
##     class counts:   110    19
##    probabilities: 0.853 0.147
##   left son=34 (113 obs) right son=35 (16 obs)
##   Primary splits:
##       total_eve_charge  < 22.775  to the left,  improve=13.270360, (0 missing)
##       total_eve_minutes < 267.95  to the left,  improve= 5.395525, (0 missing)
##       total_day_charge  < 39.065  to the right, improve= 2.501667, (0 missing)
##       total_day_minutes < 227.05  to the right, improve= 2.179177, (0 missing)
##       voice_mail_plan   splits as  RL,          improve= 1.576611, (0 missing)
##   Surrogate splits:
##       total_eve_minutes < 1200.25 to the left,  agree=0.891, adj=0.125, (0 split)
##
## Node number 18: 98 observations,    complexity param=0.04979253
##   predicted class=no   expected loss=0.1938776  P(node) =0.05882353
##     class counts:    79    19
##    probabilities: 0.806 0.194
##   left son=36 (80 obs) right son=37 (18 obs)
##   Primary splits:
##       total_intl_minutes    < 13.1    to the left,  improve=28.657650, (0 missing)
##       total_intl_charge     < 3.535   to the left,  improve=28.657650, (0 missing)
##       total_day_calls       < 81.5    to the left,  improve= 1.771894, (0 missing)
##       total_intl_calls      < 5.5     to the left,  improve= 1.271554, (0 missing)
##       number_vmail_messages < 25      to the left,  improve= 1.236257, (0 missing)
```

```
##   Surrogate splits:
##       total_intl_charge     < 3.535   to the left,  agree=1.000, adj=1.000, (0 split)
##       number_vmail_messages < 40.5    to the left,  agree=0.837, adj=0.111, (0 split)
##
## Node number 19: 24 observations
##   predicted class=yes  expected loss=0  P(node) =0.01440576
##     class counts:     0    24
##    probabilities: 0.000 1.000
##
## Node number 22: 29 observations,    complexity param=0.02282158
##   predicted class=no   expected loss=0.4827586  P(node) =0.01740696
##     class counts:    15    14
##    probabilities: 0.517 0.483
##   left son=44 (17 obs) right son=45 (12 obs)
##   Primary splits:
##       total_day_charge   < 22.48   to the right, improve=7.708249, (0 missing)
##       total_day_minutes  < 132.25  to the right, improve=4.304981, (0 missing)
##       total_intl_calls   < 3.5     to the right, improve=3.867374, (0 missing)
##       total_intl_minutes < 9.4     to the right, improve=2.271648, (0 missing)
##       total_intl_charge  < 2.54    to the right, improve=2.271648, (0 missing)
##   Surrogate splits:
##       total_day_minutes   < 132.25  to the right, agree=0.897, adj=0.75, (0 split)
##       total_day_calls     < 89      to the right, agree=0.690, adj=0.25, (0 split)
##       total_eve_calls     < 92      to the right, agree=0.690, adj=0.25, (0 split)
##       total_night_minutes < 216.15  to the left,  agree=0.690, adj=0.25, (0 split)
##       total_night_calls   < 95      to the left,  agree=0.690, adj=0.25, (0 split)
##
## Node number 23: 38 observations
##   predicted class=yes  expected loss=0.02631579  P(node) =0.02280912
##     class counts:     1    37
##    probabilities: 0.026 0.974
##
## Node number 24: 95 observations
##   predicted class=no   expected loss=0.1157895  P(node) =0.05702281
##     class counts:    84    11
##    probabilities: 0.884 0.116
##
## Node number 25: 14 observations
##   predicted class=yes  expected loss=0.2857143  P(node) =0.008403361
##     class counts:     4    10
##    probabilities: 0.286 0.714
##
## Node number 26: 7 observations
##   predicted class=no   expected loss=0.1428571  P(node) =0.004201681
##     class counts:     6     1
##    probabilities: 0.857 0.143
##
## Node number 27: 23 observations
##   predicted class=yes  expected loss=0.173913  P(node) =0.01380552
##     class counts:     4    19
##    probabilities: 0.174 0.826
##
```

```
## Node number 34: 113 observations
##   predicted class=no   expected loss=0.0619469  P(node) =0.06782713
##     class counts:   106     7
##    probabilities: 0.938 0.062
##
## Node number 35: 16 observations
##   predicted class=yes  expected loss=0.25  P(node) =0.009603842
##     class counts:     4    12
##    probabilities: 0.250 0.750
##
## Node number 36: 80 observations
##   predicted class=no   expected loss=0.0125  P(node) =0.04801921
##     class counts:    79     1
##    probabilities: 0.988 0.012
##
## Node number 37: 18 observations
##   predicted class=yes  expected loss=0  P(node) =0.01080432
##     class counts:     0    18
##    probabilities: 0.000 1.000
##
## Node number 44: 17 observations
##   predicted class=no   expected loss=0.1764706  P(node) =0.01020408
##     class counts:    14     3
##    probabilities: 0.824 0.176
##
## Node number 45: 12 observations
##   predicted class=yes  expected loss=0.08333333  P(node) =0.007202881
##     class counts:     1    11
##    probabilities: 0.083 0.917
```

```
library(rattle) #library for fancyRpartplot function
```

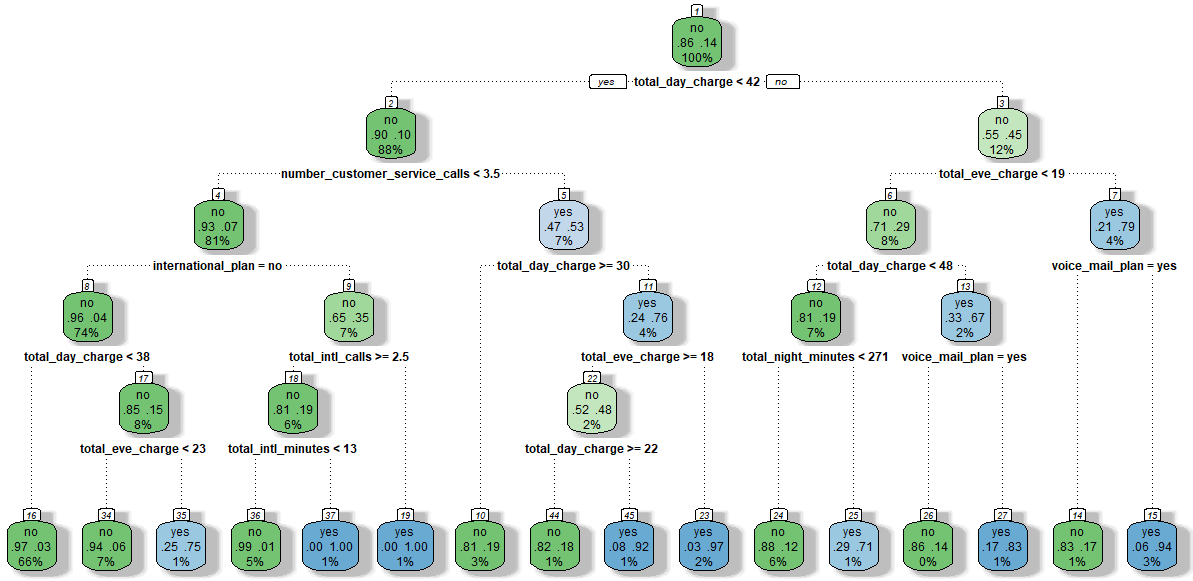
```
## Warning: package 'rattle' was built under R version 4.4.2
```

```
## Loading required package: tibble
```

```
## Loading required package: bitops
```

```
## Rattle: A free graphical interface for data science with R.
## Version 5.5.1 Copyright (c) 2006-2021 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.
```

```
fancyRpartPlot(Model_Decision) #model that brings more clarity and aesthetically more pleasing
```



Rattle 2025-Apr-09 09:37:03 Faith

```
Decision_Tree_Predictions <- predict(Model_Decision, Churn_Test, method = 'class') #predict churn using Churn_Test dataset
roc(Churn_Test$churn, Decision_Tree_Predictions[,2]) #checking accuracy of the decision tree model by finding AUC, which at .9222 is very accurate.
```

```
## Setting levels: control = no, case = yes
```

```
## Setting direction: controls < cases
```

```
##
## Call:
## roc.default(response = Churn_Test$churn, predictor = Decision_Tree_Predictions[,     2])
##
## Data: Decision_Tree_Predictions[, 2] in 610 controls (Churn_Test$churn no) < 102 cases (Churn_Test$churn yes).
## Area under the curve: 0.9222
```

```
Model_Decision2 <- rpart(churn ~ ., Churn_Train, method = 'class', control = rpart.control(minsplit = 60)) #model to make decision tree less complex
summary(Model_Decision2) #summary which includes order of importance of variables
```

```
## Call:
## rpart(formula = churn ~ ., data = Churn_Train, method = "class",
##     control = rpart.control(minsplit = 60))
##   n= 1666
##
##           CP nsplit rel error    xerror       xstd
## 1 0.07468880      0 1.0000000 1.0000000 0.05957464
## 2 0.07261411      2 0.8506224 0.9460581 0.05820929
## 3 0.04979253      4 0.7053942 0.7634855 0.05308581
## 4 0.04149378      7 0.5477178 0.5933610 0.04744205
## 5 0.01244813      8 0.5062241 0.5767635 0.04683520
## 6 0.01000000     10 0.4813278 0.6016598 0.04774128
##
## Variable importance
##              total_day_charge number_customer_service_calls
##                            25                            15
##              total_eve_charge              total_intl_calls
##                            12                             8
##             total_intl_charge            total_intl_minutes
##                             8                             8
##            international_plan             total_day_minutes
##                             8                             6
##             total_eve_minutes               total_day_calls
##                             6                             1
##                account_length         number_vmail_messages
##                             1                             1
##             total_night_calls
##                             1
##
## Node number 1: 1666 observations,    complexity param=0.0746888
##   predicted class=no   expected loss=0.1446579  P(node) =1
##     class counts:  1425   241
##    probabilities: 0.855 0.145
##   left son=2 (1465 obs) right son=3 (201 obs)
##   Primary splits:
##       total_day_charge              < 41.64   to the left,  improve=41.999570, (0 missing)
##       number_customer_service_calls < 3.5     to the left,  improve=39.956020, (0 missing)
##       international_plan            splits as  LR,          improve=28.846700, (0 missing)
##       total_day_minutes             < 223.45  to the left,  improve=13.634400, (0 missing)
##       voice_mail_plan               splits as  RL,          improve= 5.925315, (0 missing)
##
## Node number 2: 1465 observations,    complexity param=0.07261411
##   predicted class=no   expected loss=0.1030717  P(node) =0.8793517
##     class counts:  1314   151
##    probabilities: 0.897 0.103
##   left son=4 (1351 obs) right son=5 (114 obs)
##   Primary splits:
##       number_customer_service_calls < 3.5     to the left,  improve=44.289320, (0 missing)
##       international_plan            splits as  LR,          improve=23.000150, (0 missing)
##       total_intl_calls              < 3.5     to the right, improve= 4.616694, (0 missing)
##       total_eve_charge              < 25.665  to the left,  improve= 4.248699, (0 missing)
##       total_day_charge              < 37.99   to the left,  improve= 3.756133, (0 missing)
```

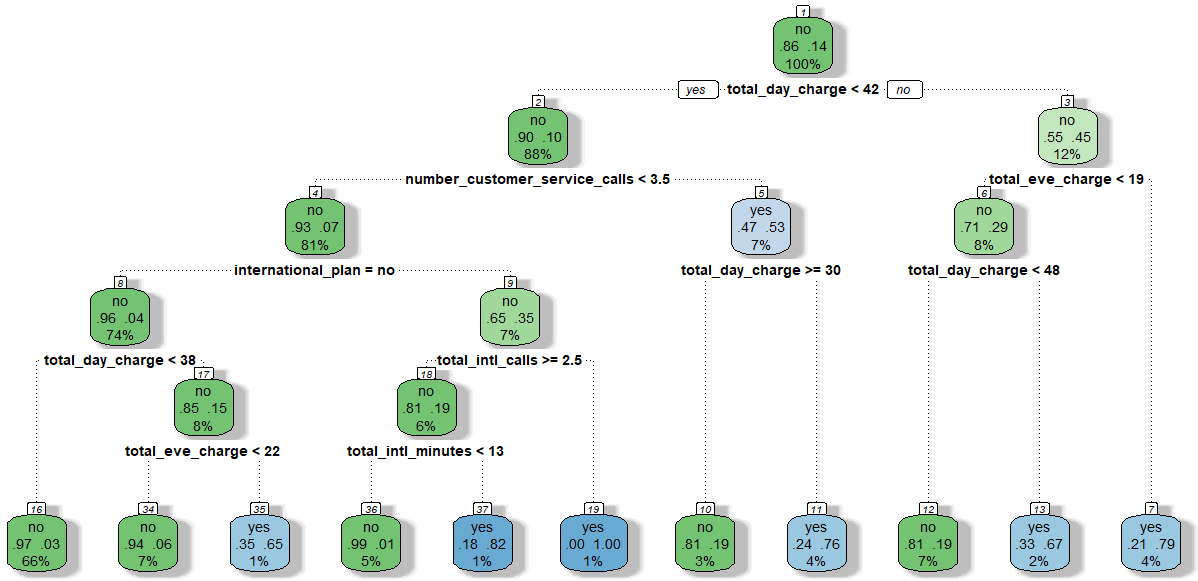
```
##
## Node number 3: 201 observations,    complexity param=0.0746888
##   predicted class=no   expected loss=0.4477612  P(node) =0.1206483
##     class counts:   111    90
##    probabilities: 0.552 0.448
##   left son=6 (139 obs) right son=7 (62 obs)
##   Primary splits:
##       total_eve_charge      < 18.895  to the left,  improve=21.04165, (0 missing)
##       voice_mail_plan       splits as  RL,          improve=17.34439, (0 missing)
##       number_vmail_messages < 5.5     to the right, improve=17.34439, (0 missing)
##       total_eve_minutes     < 199.6   to the left,  improve=15.07462, (0 missing)
##       total_day_charge      < 49.495  to the left,  improve=12.29145, (0 missing)
##   Surrogate splits:
##       total_eve_minutes < 222.3   to the left,  agree=0.896, adj=0.661, (0 split)
##       total_day_charge  < 41.82   to the right, agree=0.711, adj=0.065, (0 split)
##       total_day_minutes < 246     to the right, agree=0.706, adj=0.048, (0 split)
##       total_day_calls   < 134.5   to the left,  agree=0.697, adj=0.016, (0 split)
##       total_night_calls < 131.5   to the left,  agree=0.697, adj=0.016, (0 split)
##
## Node number 4: 1351 observations,    complexity param=0.04979253
##   predicted class=no   expected loss=0.06735751  P(node) =0.8109244
##     class counts:  1260    91
##    probabilities: 0.933 0.067
##   left son=8 (1229 obs) right son=9 (122 obs)
##   Primary splits:
##       international_plan splits as  LR,          improve=21.801800, (0 missing)
##       total_day_charge   < 37.99   to the left,  improve= 5.496487, (0 missing)
##       total_eve_charge   < 25.665  to the left,  improve= 5.454307, (0 missing)
##       total_intl_minutes < 13.15   to the left,  improve= 3.174802, (0 missing)
##       total_intl_charge  < 3.55    to the left,  improve= 3.174802, (0 missing)
##
## Node number 5: 114 observations,    complexity param=0.07261411
##   predicted class=yes  expected loss=0.4736842  P(node) =0.06842737
##     class counts:    54    60
##    probabilities: 0.474 0.526
##   left son=10 (47 obs) right son=11 (67 obs)
##   Primary splits:
##       total_day_charge  < 29.88   to the right, improve=17.930700, (0 missing)
##       total_day_minutes < 161.8   to the right, improve=15.048870, (0 missing)
##       total_eve_minutes < 199.8   to the right, improve= 7.814086, (0 missing)
##       total_eve_charge  < 17.025  to the right, improve= 6.456920, (0 missing)
##       total_intl_calls  < 4.5     to the right, improve= 4.994808, (0 missing)
##   Surrogate splits:
##       total_day_minutes  < 175.75  to the right, agree=0.904, adj=0.766, (0 split)
##       international_plan splits as  RL,          agree=0.632, adj=0.106, (0 split)
##       total_eve_minutes  < 318.25  to the right, agree=0.632, adj=0.106, (0 split)
##       total_intl_minutes < 14.25   to the right, agree=0.632, adj=0.106, (0 split)
##       total_intl_charge  < 3.85    to the right, agree=0.632, adj=0.106, (0 split)
##
## Node number 6: 139 observations,    complexity param=0.04149378
##   predicted class=no   expected loss=0.294964  P(node) =0.08343337
##     class counts:    98    41
```

```
##    probabilities: 0.705 0.295
##   left son=12 (109 obs) right son=13 (30 obs)
##   Primary splits:
##       total_day_charge    < 48.415  to the left,  improve=10.571360, (0 missing)
##       total_night_minutes < 212.7   to the left,  improve= 8.855730, (0 missing)
##       total_night_charge  < 9.57    to the left,  improve= 8.855730, (0 missing)
##       total_day_minutes   < 275.5   to the left,  improve= 7.988065, (0 missing)
##       voice_mail_plan     splits as  RL,          improve= 5.434667, (0 missing)
##   Surrogate splits:
##       total_day_minutes < 284.8   to the left,  agree=0.878, adj=0.433, (0 split)
##       account_length    < 15.5    to the right, agree=0.799, adj=0.067, (0 split)
##       total_night_calls < 137     to the left,  agree=0.799, adj=0.067, (0 split)
##       total_day_calls   < 58.5    to the right, agree=0.791, adj=0.033, (0 split)
##       total_intl_calls  < 1.5     to the right, agree=0.791, adj=0.033, (0 split)
##
## Node number 7: 62 observations
##   predicted class=yes  expected loss=0.2096774  P(node) =0.03721489
##     class counts:    13    49
##    probabilities: 0.210 0.790
##
## Node number 8: 1229 observations,    complexity param=0.01244813
##   predicted class=no   expected loss=0.03905614  P(node) =0.7376951
##     class counts:  1181    48
##    probabilities: 0.961 0.039
##   left son=16 (1100 obs) right son=17 (129 obs)
##   Primary splits:
##       total_day_charge  < 38.105  to the left,  improve=3.3766000, (0 missing)
##       total_eve_charge  < 25.665  to the left,  improve=3.1861680, (0 missing)
##       total_day_minutes < 209.3   to the left,  improve=1.0038510, (0 missing)
##       total_eve_minutes < 1193.55 to the left,  improve=0.6999996, (0 missing)
##       account_length    < 190.5   to the left,  improve=0.5004862, (0 missing)
##
## Node number 9: 122 observations,    complexity param=0.04979253
##   predicted class=no   expected loss=0.352459  P(node) =0.07322929
##     class counts:    79    43
##    probabilities: 0.648 0.352
##   left son=18 (98 obs) right son=19 (24 obs)
##   Primary splits:
##       total_intl_calls    < 2.5     to the right, improve=25.055870, (0 missing)
##       total_intl_minutes  < 13.1    to the left,  improve=21.272680, (0 missing)
##       total_intl_charge   < 3.535   to the left,  improve=21.272680, (0 missing)
##       total_eve_charge    < 14.11   to the left,  improve= 3.820409, (0 missing)
##       total_night_minutes < 203.75  to the right, improve= 2.977559, (0 missing)
##   Surrogate splits:
##       total_day_calls  < 49.5    to the right, agree=0.820, adj=0.083, (0 split)
##       total_eve_charge < 25.37   to the left,  agree=0.811, adj=0.042, (0 split)
##
## Node number 10: 47 observations
##   predicted class=no   expected loss=0.1914894  P(node) =0.02821128
##     class counts:    38     9
##    probabilities: 0.809 0.191
##
```

```
## Node number 11: 67 observations
##   predicted class=yes  expected loss=0.238806  P(node) =0.04021609
##     class counts:    16    51
##    probabilities: 0.239 0.761
##
## Node number 12: 109 observations
##   predicted class=no   expected loss=0.1926606  P(node) =0.06542617
##     class counts:    88    21
##    probabilities: 0.807 0.193
##
## Node number 13: 30 observations
##   predicted class=yes  expected loss=0.3333333  P(node) =0.0180072
##     class counts:    10    20
##    probabilities: 0.333 0.667
##
## Node number 16: 1100 observations
##   predicted class=no   expected loss=0.02636364  P(node) =0.6602641
##     class counts:  1071    29
##    probabilities: 0.974 0.026
##
## Node number 17: 129 observations,    complexity param=0.01244813
##   predicted class=no   expected loss=0.1472868  P(node) =0.07743097
##     class counts:   110    19
##    probabilities: 0.853 0.147
##   left son=34 (109 obs) right son=35 (20 obs)
##   Primary splits:
##       total_eve_charge  < 22.145  to the left,  improve=11.963650, (0 missing)
##       total_eve_minutes < 267.95  to the left,  improve= 5.395525, (0 missing)
##       total_day_charge  < 39.065  to the right, improve= 2.501667, (0 missing)
##       total_day_minutes < 229.8   to the right, improve= 2.151207, (0 missing)
##       voice_mail_plan   splits as  RL,          improve= 1.576611, (0 missing)
##   Surrogate splits:
##       total_eve_minutes < 261.55  to the left,  agree=0.884, adj=0.25, (0 split)
##       total_night_calls < 65      to the right, agree=0.853, adj=0.05, (0 split)
##
## Node number 18: 98 observations,    complexity param=0.04979253
##   predicted class=no   expected loss=0.1938776  P(node) =0.05882353
##     class counts:    79    19
##    probabilities: 0.806 0.194
##   left son=36 (76 obs) right son=37 (22 obs)
##   Primary splits:
##       total_intl_minutes    < 12.8    to the left,  improve=22.113510, (0 missing)
##       total_intl_charge     < 3.455   to the left,  improve=22.113510, (0 missing)
##       total_intl_calls      < 5.5     to the left,  improve= 1.271554, (0 missing)
##       number_vmail_messages < 25      to the left,  improve= 1.236257, (0 missing)
##       total_eve_charge      < 14.11   to the left,  improve= 1.175510, (0 missing)
##   Surrogate splits:
##       total_intl_charge     < 3.455   to the left,  agree=1.000, adj=1.000, (0 split)
##       account_length        < 174     to the left,  agree=0.796, adj=0.091, (0 split)
##       number_vmail_messages < 40.5    to the left,  agree=0.796, adj=0.091, (0 split)
##       total_eve_charge      < 11.16   to the right, agree=0.786, adj=0.045, (0 split)
##       total_night_minutes   < 123     to the right, agree=0.786, adj=0.045, (0 split)
```

```
##
## Node number 19: 24 observations
##   predicted class=yes  expected loss=0  P(node) =0.01440576
##     class counts:     0    24
##    probabilities: 0.000 1.000
##
## Node number 34: 109 observations
##   predicted class=no   expected loss=0.05504587  P(node) =0.06542617
##     class counts:   103     6
##    probabilities: 0.945 0.055
##
## Node number 35: 20 observations
##   predicted class=yes  expected loss=0.35  P(node) =0.0120048
##     class counts:     7    13
##    probabilities: 0.350 0.650
##
## Node number 36: 76 observations
##   predicted class=no   expected loss=0.01315789  P(node) =0.04561825
##     class counts:    75     1
##    probabilities: 0.987 0.013
##
## Node number 37: 22 observations
##   predicted class=yes  expected loss=0.1818182  P(node) =0.01320528
##     class counts:     4    18
##    probabilities: 0.182 0.818
```

```
fancyRpartPlot(Model_Decision2) #model that brings more clarity and aesthetically more pleasing
```



Rattle 2025-Apr-09 09:37:04 Faith

```
Decision_Tree_Predictions2 <- predict(Model_Decision2, Churn_Test, method = 'class') #predict ch
urn using Churn_Test dataset
roc(Churn_Test$churn, Decision_Tree_Predictions2[,2]) #checking accuracy of the decision tree mo
del by finding AUC, which at .9168, which showed pruning did not make more accurate. However, ac
curacy is still high.
```

```
## Setting levels: control = no, case = yes
```

```
## Setting direction: controls < cases
```

```
##
## Call:
## roc.default(response = Churn_Test$churn, predictor = Decision_Tree_Predictions2[,     2])
##
## Data: Decision_Tree_Predictions2[, 2] in 610 controls (Churn_Test$churn no) < 102 cases (Chur
n_Test$churn yes).
## Area under the curve: 0.9168
```

```
Churndata_cleaned3 <- Churndata_cleaned2 #using this data frame to change all factors to numeri
c; this data frame needs to be all numeric variables so we can use for regression modeling.
Churndata_cleaned3$international_plan <- as.numeric(Churndata_cleaned3$international_plan == "ye
s") #changing variable from factor to numeric
Churndata_cleaned3$voice_mail_plan <- as.numeric(Churndata_cleaned3$voice_mail_plan == "yes")
Churndata_cleaned3$churn <- as.numeric(Churndata_cleaned3$churn == "yes")
summary(Churndata_cleaned3) #summary of data frame with all numeric variables
```

```
##  account_length  international_plan voice_mail_plan  number_vmail_messages
##  Min.   :  1.0   Min.   :0.00000    Min.   :0.0000   Min.   : 0.00
##  1st Qu.: 73.0   1st Qu.:0.00000    1st Qu.:0.0000   1st Qu.: 0.00
##  Median :101.0   Median :0.00000    Median :0.0000   Median : 0.00
##  Mean   :100.6   Mean   :0.09336    Mean   :0.2775   Mean   : 8.09
##  3rd Qu.:127.0   3rd Qu.:0.00000    3rd Qu.:1.0000   3rd Qu.:19.00
##  Max.   :243.0   Max.   :1.00000    Max.   :1.0000   Max.   :51.00
##  total_day_minutes total_day_calls  total_day_charge total_eve_minutes
##  Min.   :   0.0    Min.   :  0.00   Min.   : 0.00    Min.   :  31.2
##  1st Qu.: 151.6    1st Qu.: 87.00   1st Qu.:24.58    1st Qu.: 172.9
##  Median : 194.6    Median :100.00   Median :30.84    Median : 213.1
##  Mean   : 495.7    Mean   : 99.95   Mean   :30.79    Mean   : 358.5
##  3rd Qu.: 252.0    3rd Qu.:114.00   3rd Qu.:36.91    3rd Qu.: 267.3
##  Max.   :2185.1    Max.   :165.00   Max.   :59.64    Max.   :1244.2
##  total_eve_calls total_eve_charge total_night_minutes total_night_calls
##  Min.   : 12.0   Min.   : 2.65    Min.   : 43.7       Min.   : 33.00
##  1st Qu.: 87.0   1st Qu.:14.20    1st Qu.:167.2       1st Qu.: 86.00
##  Median :100.5   Median :17.12    Median :200.4       Median :100.00
##  Mean   :100.2   Mean   :17.12    Mean   :200.9       Mean   : 99.74
##  3rd Qu.:114.0   3rd Qu.:20.00    3rd Qu.:235.0       3rd Qu.:113.00
##  Max.   :170.0   Max.   :30.91    Max.   :367.7       Max.   :164.00
##  total_night_charge total_intl_minutes total_intl_calls total_intl_charge
##  Min.   : 1.970     Min.   : 0.0       Min.   : 0.000   Min.   :0.000
##  1st Qu.: 7.522     1st Qu.: 8.4       1st Qu.: 3.000   1st Qu.:2.270
##  Median : 9.020     Median :10.2       Median : 4.000   Median :2.750
##  Mean   : 9.039     Mean   :10.2       Mean   : 4.489   Mean   :2.753
##  3rd Qu.:10.578     3rd Qu.:12.0       3rd Qu.: 6.000   3rd Qu.:3.240
##  Max.   :16.550     Max.   :20.0       Max.   :19.000   Max.   :5.400
##  number_customer_service_calls     churn
##  Min.   :0.000                 Min.   :0.0000
##  1st Qu.:1.000                 1st Qu.:0.0000
##  Median :1.000                 Median :0.0000
##  Mean   :1.556                 Mean   :0.1442
##  3rd Qu.:2.000                 3rd Qu.:0.0000
##  Max.   :8.000                 Max.   :1.0000
```

```
Model_Regression2 <- lm(churn ~ ., Churndata_cleaned3) #multiple R squared is incredibly low her
e at .1831 or 18.31%. This is not the model to use.
summary(Model_Regression2)
```

```
##
## Call:
## lm(formula = churn ~ ., data = Churndata_cleaned3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.65264 -0.17150 -0.08444  0.02636  1.12073
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -4.664e-01  7.962e-02  -5.857 5.36e-09 ***
## account_length                7.200e-05  1.660e-04   0.434 0.664568
## international_plan            3.010e-01  2.257e-02  13.337  < 2e-16 ***
## voice_mail_plan              -1.276e-01  5.006e-02  -2.549 0.010871 *
## number_vmail_messages         1.338e-03  1.641e-03   0.815 0.415130
## total_day_minutes            -6.161e-06  2.127e-04  -0.029 0.976897
## total_day_calls               2.992e-04  3.244e-04   0.922 0.356414
## total_day_charge              7.876e-03  1.253e-03   6.287 3.84e-10 ***
## total_eve_minutes            -1.488e-05  4.263e-04  -0.035 0.972164
## total_eve_calls              -9.396e-05  3.308e-04  -0.284 0.776385
## total_eve_charge              9.129e-03  5.213e-03   1.751 0.080053 .
## total_night_minutes           4.077e-02  1.041e-01   0.392 0.695198
## total_night_calls             1.808e-04  3.389e-04   0.534 0.593623
## total_night_charge           -9.009e-01  2.312e+00  -0.390 0.696865
## total_intl_minutes           -4.634e-01  6.191e-01  -0.748 0.454266
## total_intl_calls             -9.077e-03  2.704e-03  -3.357 0.000802 ***
## total_intl_charge             1.746e+00  2.293e+00   0.761 0.446625
## number_customer_service_calls  5.334e-02  5.027e-03  10.610  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3188 on 2360 degrees of freedom
## Multiple R-squared:  0.1831, Adjusted R-squared:  0.1772
## F-statistic: 31.11 on 17 and 2360 DF,  p-value: < 2.2e-16
```

```
Model_Decision3 <- rpart(churn ~ international_plan + number_customer_service_calls + total_day_
charge + total_intl_calls + voice_mail_plan, Churn_Train, method = 'class', control = rpart.cont
rol(minsplit = 60)) #model to make decision tree less complex; trying only variables with very l
ow p-value.
summary(Model_Decision3) #summary which includes order of importance of variables
```

```
## Call:
## rpart(formula = churn ~ international_plan + number_customer_service_calls +
##     total_day_charge + total_intl_calls + voice_mail_plan, data = Churn_Train,
##     method = "class", control = rpart.control(minsplit = 60))
##   n= 1666
##
##           CP nsplit rel error    xerror       xstd
## 1 0.05809129      0 1.0000000 1.0000000 0.05957464
## 2 0.04979253      5 0.7012448 0.7883817 0.05383509
## 3 0.02074689      7 0.6016598 0.6763485 0.05031746
## 4 0.01000000      8 0.5809129 0.6721992 0.05017956
##
## Variable importance
##              total_day_charge number_customer_service_calls
##                            41                            24
##              total_intl_calls            international_plan
##                            14                            13
##               voice_mail_plan
##                             9
##
## Node number 1: 1666 observations,    complexity param=0.05809129
##   predicted class=no   expected loss=0.1446579  P(node) =1
##     class counts:  1425   241
##    probabilities: 0.855 0.145
##   left son=2 (1465 obs) right son=3 (201 obs)
##   Primary splits:
##       total_day_charge              < 41.64  to the left,  improve=41.999570, (0 missing)
##       number_customer_service_calls < 3.5    to the left,  improve=39.956020, (0 missing)
##       international_plan            splits as  LR,         improve=28.846700, (0 missing)
##       voice_mail_plan               splits as  RL,         improve= 5.925315, (0 missing)
##       total_intl_calls              < 3.5    to the right, improve= 5.014111, (0 missing)
##
## Node number 2: 1465 observations,    complexity param=0.05809129
##   predicted class=no   expected loss=0.1030717  P(node) =0.8793517
##     class counts:  1314   151
##    probabilities: 0.897 0.103
##   left son=4 (1351 obs) right son=5 (114 obs)
##   Primary splits:
##       number_customer_service_calls < 3.5    to the left,  improve=44.2893200, (0 missing)
##       international_plan            splits as  LR,         improve=23.0001500, (0 missing)
##       total_intl_calls              < 3.5    to the right, improve= 4.6166940, (0 missing)
##       total_day_charge              < 37.99  to the left,  improve= 3.7561330, (0 missing)
##       voice_mail_plan               splits as  RL,         improve= 0.9322517, (0 missing)
##
## Node number 3: 201 observations,    complexity param=0.05809129
##   predicted class=no   expected loss=0.4477612  P(node) =0.1206483
##     class counts:   111    90
##    probabilities: 0.552 0.448
##   left son=6 (52 obs) right son=7 (149 obs)
##   Primary splits:
##       voice_mail_plan               splits as  RL,         improve=17.3443900, (0 missing)
##       total_day_charge              < 49.495 to the left,  improve=12.2914500, (0 missing)
```

```
##       international_plan            splits as  LR,         improve= 3.7010080, (0 missing)
##       number_customer_service_calls < 2.5    to the right, improve= 1.6513510, (0 missing)
##       total_intl_calls              < 3.5    to the right, improve= 0.5486585, (0 missing)
##
## Node number 4: 1351 observations,    complexity param=0.04979253
##   predicted class=no   expected loss=0.06735751  P(node) =0.8109244
##     class counts:  1260    91
##    probabilities: 0.933 0.067
##   left son=8 (1229 obs) right son=9 (122 obs)
##   Primary splits:
##       international_plan            splits as  LR,         improve=21.8018000, (0 missing)
##       total_day_charge              < 37.99  to the left,  improve= 5.4964870, (0 missing)
##       total_intl_calls              < 2.5    to the right, improve= 2.7912450, (0 missing)
##       number_customer_service_calls < 0.5    to the right, improve= 0.5212856, (0 missing)
##       voice_mail_plan               splits as  RL,         improve= 0.1353860, (0 missing)
##
## Node number 5: 114 observations,    complexity param=0.05809129
##   predicted class=yes  expected loss=0.4736842  P(node) =0.06842737
##     class counts:    54    60
##    probabilities: 0.474 0.526
##   left son=10 (47 obs) right son=11 (67 obs)
##   Primary splits:
##       total_day_charge              < 29.88  to the right, improve=17.930700, (0 missing)
##       total_intl_calls              < 4.5    to the right, improve= 4.994808, (0 missing)
##       number_customer_service_calls < 4.5    to the left,  improve= 1.672029, (0 missing)
##       voice_mail_plan               splits as  RL,         improve= 1.392105, (0 missing)
##   Surrogate splits:
##       international_plan splits as  RL, agree=0.632, adj=0.106, (0 split)
##
## Node number 6: 52 observations
##   predicted class=no   expected loss=0.09615385  P(node) =0.03121248
##     class counts:    47     5
##    probabilities: 0.904 0.096
##
## Node number 7: 149 observations,    complexity param=0.05809129
##   predicted class=yes  expected loss=0.4295302  P(node) =0.08943577
##     class counts:    64    85
##    probabilities: 0.430 0.570
##   left son=14 (98 obs) right son=15 (51 obs)
##   Primary splits:
##       total_day_charge              < 46.835 to the left,  improve=13.247830, (0 missing)
##       total_intl_calls              < 3.5    to the right, improve= 1.362688, (0 missing)
##       number_customer_service_calls < 1.5    to the right, improve= 1.176761, (0 missing)
##   Surrogate splits:
##       total_intl_calls < 9.5    to the left,  agree=0.664, adj=0.02, (0 split)
##
## Node number 8: 1229 observations
##   predicted class=no   expected loss=0.03905614  P(node) =0.7376951
##     class counts:  1181    48
##    probabilities: 0.961 0.039
##
## Node number 9: 122 observations,    complexity param=0.04979253
```

```
##   predicted class=no   expected loss=0.352459  P(node) =0.07322929
##     class counts:    79    43
##    probabilities: 0.648 0.352
##   left son=18 (98 obs) right son=19 (24 obs)
##   Primary splits:
##       total_intl_calls             < 2.5    to the right, improve=25.05587000, (0 missing)
##       total_day_charge             < 37.58  to the left,  improve= 1.48956200, (0 missing)
##       number_customer_service_calls < 0.5    to the right, improve= 0.62741350, (0 missing)
##       voice_mail_plan              splits as  LR,         improve= 0.03532262, (0 missing)
##
## Node number 10: 47 observations
##   predicted class=no   expected loss=0.1914894  P(node) =0.02821128
##     class counts:    38     9
##    probabilities: 0.809 0.191
##
## Node number 11: 67 observations
##   predicted class=yes  expected loss=0.238806  P(node) =0.04021609
##     class counts:    16    51
##    probabilities: 0.239 0.761
##
## Node number 14: 98 observations,    complexity param=0.02074689
##   predicted class=no   expected loss=0.4183673  P(node) =0.05882353
##     class counts:    57    41
##    probabilities: 0.582 0.418
##   left son=28 (71 obs) right son=29 (27 obs)
##   Primary splits:
##       total_day_charge             < 42.815 to the right, improve=2.2624740, (0 missing)
##       total_intl_calls             < 3.5    to the right, improve=1.0762440, (0 missing)
##       number_customer_service_calls < 1.5    to the right, improve=0.4129411, (0 missing)
##
## Node number 15: 51 observations
##   predicted class=yes  expected loss=0.1372549  P(node) =0.03061224
##     class counts:     7    44
##    probabilities: 0.137 0.863
##
## Node number 18: 98 observations
##   predicted class=no   expected loss=0.1938776  P(node) =0.05882353
##     class counts:    79    19
##    probabilities: 0.806 0.194
##
## Node number 19: 24 observations
##   predicted class=yes  expected loss=0  P(node) =0.01440576
##     class counts:     0    24
##    probabilities: 0.000 1.000
##
## Node number 28: 71 observations
##   predicted class=no   expected loss=0.3521127  P(node) =0.04261705
##     class counts:    46    25
##    probabilities: 0.648 0.352
##
## Node number 29: 27 observations
##   predicted class=yes  expected loss=0.4074074  P(node) =0.01620648
```

```
##     class counts:    11    16
##    probabilities: 0.407 0.593
```

```
Decision_Tree_Predictions3 <- predict(Model_Decision3, Churn_Test, method = 'class') #predict churn using Churn_Test dataset
roc(Churn_Test$churn, Decision_Tree_Predictions3[,2]) #checking accuracy of the decision tree model by finding AUC; accuracy is lower than the previous Decision Tree models that include all of the variables at .8753
```

```
## Setting levels: control = no, case = yes
```

```
## Setting direction: controls < cases
```

```
##
## Call:
## roc.default(response = Churn_Test$churn, predictor = Decision_Tree_Predictions3[,     2])
##
## Data: Decision_Tree_Predictions3[, 2] in 610 controls (Churn_Test$churn no) < 102 cases (Churn_Test$churn yes).
## Area under the curve: 0.8753
```