knnChurn.R

sigsp

2022-03-20

```
## Author: Stephen E. Porter
## Title: k-Nearest-Neighbor (Task 1)
## Course: WGU D209: Data Mining I
## Instructor: Dr. Festus Elleh
options(warn=-1)

# Libraries
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(fastDummies)
library(class)
library(caret)
```

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

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```
library(pROC)
```

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

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

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

```
# Import CSV as data frame
df <- read.csv(file = 'C:/WGU/D209 Data Mining I/churn_clean.csv')
```

```
# Checking for nulls
sapply(df, function(x) sum(is.na(x)))
```

```
##            CaseOrder          Customer_id          Interaction
##                    0                    0                    0
##                  UID                 City                State
##                    0                    0                    0
##               County                  Zip                  Lat
##                    0                    0                    0
##                  Lng           Population                 Area
##                    0                    0                    0
##             TimeZone                  Job             Children
##                    0                    0                    0
##                  Age               Income              Marital
##                    0                    0                    0
##               Gender                Churn   Outage_sec_perweek
##                    0                    0                    0
##                Email            Contacts Yearly_equip_failure
##                    0                    0                    0
##               Techie             Contract          Port_modem
##                    0                    0                    0
##               Tablet      InternetService               Phone
##                    0                    0                    0
##             Multiple       OnlineSecurity         OnlineBackup
##                    0                    0                    0
##     DeviceProtection          TechSupport          StreamingTV
##                    0                    0                    0
##      StreamingMovies     PaperlessBilling        PaymentMethod
##                    0                    0                    0
##               Tenure        MonthlyCharge    Bandwidth_GB_Year
##                    0                    0                    0
##                Item1                Item2                Item3
##                    0                    0                    0
##                Item4                Item5                Item6
##                    0                    0                    0
##                Item7                Item8
##                    0                    0
```

```
dim(df)
```

```
## [1] 10000    50
```

```
str(df)
```

```
## 'data.frame':    10000 obs. of  50 variables:
##  $ CaseOrder          : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Customer_id        : chr  "K409198" "S120509" "K191035" "D90850" ...
##  $ Interaction        : chr  "aa90260b-4141-4a24-8e36-b04ce1f4f77b" "fb76459f-c047-4a9d-8af9-e0f7d4
##  $ UID                : chr  "e885b299883d4f9fb18e39c75155d990" "f2de8bef964785f41a2959829830fb8a"
##  $ City               : chr  "Point Baker" "West Branch" "Yamhill" "Del Mar" ...
##  $ State              : chr  "AK" "MI" "OR" "CA" ...
##  $ County             : chr  "Prince of Wales-Hyder" "Ogemaw" "Yamhill" "San Diego" ...
##  $ Zip                : int  99927 48661 97148 92014 77461 31030 37847 73109 34771 45237 ...
##  $ Lat                : num  56.3 44.3 45.4 33 29.4 ...
##  $ Lng                : num  -133.4 -84.2 -123.2 -117.2 -95.8 ...
```

```
##  $ Population          : int  38 10446 3735 13863 11352 17701 2535 23144 17351 20193 ...
##  $ Area                : chr  "Urban" "Urban" "Urban" "Suburban" ...
##  $ TimeZone            : chr  "America/Sitka" "America/Detroit" "America/Los_Angeles" "America/Los_A
##  $ Job                 : chr  "Environmental health practitioner" "Programmer, multimedia" "Chief Fi
##  $ Children            : int  0 1 4 1 0 3 0 2 2 1 ...
##  $ Age                 : int  68 27 50 48 83 83 79 30 49 86 ...
##  $ Income              : num  28562 21705 9610 18925 40074 ...
##  $ Marital             : chr  "Widowed" "Married" "Widowed" "Married" ...
##  $ Gender              : chr  "Male" "Female" "Female" "Male" ...
##  $ Churn               : chr  "No" "Yes" "No" "No" ...
##  $ Outage_sec_perweek  : num  7.98 11.7 10.75 14.91 8.15 ...
##  $ Email               : int  10 12 9 15 16 15 10 16 20 18 ...
##  $ Contacts            : int  0 0 0 2 2 3 0 0 2 1 ...
##  $ Yearly_equip_failure: int  1 1 1 0 1 1 1 0 3 0 ...
##  $ Techie              : chr  "No" "Yes" "Yes" "Yes" ...
##  $ Contract            : chr  "One year" "Month-to-month" "Two Year" "Two Year" ...
##  $ Port_modem          : chr  "Yes" "No" "Yes" "No" ...
##  $ Tablet              : chr  "Yes" "Yes" "No" "No" ...
##  $ InternetService     : chr  "Fiber Optic" "Fiber Optic" "DSL" "DSL" ...
##  $ Phone               : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ Multiple            : chr  "No" "Yes" "Yes" "No" ...
##  $ OnlineSecurity      : chr  "Yes" "Yes" "No" "Yes" ...
##  $ OnlineBackup        : chr  "Yes" "No" "No" "No" ...
##  $ DeviceProtection    : chr  "No" "No" "No" "No" ...
##  $ TechSupport         : chr  "No" "No" "No" "No" ...
##  $ StreamingTV         : chr  "No" "Yes" "No" "Yes" ...
##  $ StreamingMovies     : chr  "Yes" "Yes" "Yes" "No" ...
##  $ PaperlessBilling    : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ PaymentMethod       : chr  "Credit Card (automatic)" "Bank Transfer(automatic)" "Credit Card (aut
##  $ Tenure              : num  6.8 1.16 15.75 17.09 1.67 ...
##  $ MonthlyCharge       : num  172 243 160 120 150 ...
##  $ Bandwidth_GB_Year   : num  905 801 2055 2165 271 ...
##  $ Item1               : int  5 3 4 4 4 3 6 2 5 2 ...
##  $ Item2               : int  5 4 4 4 4 3 5 2 4 2 ...
##  $ Item3               : int  5 3 2 4 4 3 6 2 4 2 ...
##  $ Item4               : int  3 3 4 2 3 2 4 5 3 2 ...
##  $ Item5               : int  4 4 4 5 4 4 1 2 4 5 ...
##  $ Item6               : int  4 3 3 4 4 3 5 3 3 2 ...
##  $ Item7               : int  3 4 3 3 4 3 5 4 4 3 ...
##  $ Item8               : int  4 4 3 3 5 3 5 5 4 3 ...
```

```r
# Renaming unclear columns named Item1 through Item8 for improved readability &
# confirming they have been renamed correctly

df <- df %>%
  rename(
    Response = Item1,
    Fix = Item2,
    Replacement = Item3,
    Reliability = Item4,
    Options = Item5,
    Respectful = Item6,
    Courteous = Item7,
    Listening = Item8
```

```
  )

colnames(df)
```

```
##  [1] "CaseOrder"            "Customer_id"          "Interaction"
##  [4] "UID"                  "City"                 "State"
##  [7] "County"               "Zip"                  "Lat"
## [10] "Lng"                  "Population"           "Area"
## [13] "TimeZone"             "Job"                  "Children"
## [16] "Age"                  "Income"               "Marital"
## [19] "Gender"               "Churn"                "Outage_sec_perweek"
## [22] "Email"                "Contacts"             "Yearly_equip_failure"
## [25] "Techie"               "Contract"             "Port_modem"
## [28] "Tablet"               "InternetService"      "Phone"
## [31] "Multiple"             "OnlineSecurity"       "OnlineBackup"
## [34] "DeviceProtection"     "TechSupport"          "StreamingTV"
## [37] "StreamingMovies"      "PaperlessBilling"     "PaymentMethod"
## [40] "Tenure"               "MonthlyCharge"        "Bandwidth_GB_Year"
## [43] "Response"             "Fix"                  "Replacement"
## [46] "Reliability"          "Options"              "Respectful"
## [49] "Courteous"            "Listening"
```

```
# Several columns will not be useful in analysis and therefore will be dropped.
to_drop <- c('CaseOrder', 'Customer_id', 'Interaction', 'UID', 'City',
             'County', 'Zip', 'Lat', 'Lng', 'TimeZone', 'Job')

dfDropped = df[,!(names(df) %in% to_drop)]
str(dfDropped)
```

```
## 'data.frame':    10000 obs. of  39 variables:
##  $ State               : chr  "AK" "MI" "OR" "CA" ...
##  $ Population          : int  38 10446 3735 13863 11352 17701 2535 23144 17351 20193 ...
##  $ Area                : chr  "Urban" "Urban" "Urban" "Suburban" ...
##  $ Children            : int  0 1 4 1 0 3 0 2 2 1 ...
##  $ Age                 : int  68 27 50 48 83 83 79 30 49 86 ...
##  $ Income              : num  28562 21705 9610 18925 40074 ...
##  $ Marital             : chr  "Widowed" "Married" "Widowed" "Married" ...
##  $ Gender              : chr  "Male" "Female" "Female" "Male" ...
##  $ Churn               : chr  "No" "Yes" "No" "No" ...
##  $ Outage_sec_perweek  : num  7.98 11.7 10.75 14.91 8.15 ...
##  $ Email               : int  10 12 9 15 16 15 10 16 20 18 ...
##  $ Contacts            : int  0 0 0 2 2 3 0 0 2 1 ...
##  $ Yearly_equip_failure: int  1 1 1 0 1 1 1 0 3 0 ...
##  $ Techie              : chr  "No" "Yes" "Yes" "Yes" ...
##  $ Contract            : chr  "One year" "Month-to-month" "Two Year" "Two Year" ...
##  $ Port_modem          : chr  "Yes" "No" "Yes" "No" ...
##  $ Tablet              : chr  "Yes" "Yes" "No" "No" ...
##  $ InternetService     : chr  "Fiber Optic" "Fiber Optic" "DSL" "DSL" ...
##  $ Phone               : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ Multiple            : chr  "No" "Yes" "Yes" "No" ...
##  $ OnlineSecurity      : chr  "Yes" "Yes" "No" "Yes" ...
##  $ OnlineBackup        : chr  "Yes" "No" "No" "No" ...
##  $ DeviceProtection    : chr  "No" "No" "No" "No" ...
##  $ TechSupport         : chr  "No" "No" "No" "No" ...
##  $ StreamingTV         : chr  "No" "Yes" "No" "Yes" ...
```

```
##  $ StreamingMovies     : chr  "Yes" "Yes" "Yes" "No" ...
##  $ PaperlessBilling    : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ PaymentMethod       : chr  "Credit Card (automatic)" "Bank Transfer(automatic)" "Credit Card (aut
##  $ Tenure              : num  6.8 1.16 15.75 17.09 1.67 ...
##  $ MonthlyCharge       : num  172 243 160 120 150 ...
##  $ Bandwidth_GB_Year   : num  905 801 2055 2165 271 ...
##  $ Response            : int  5 3 4 4 4 3 6 2 5 2 ...
##  $ Fix                 : int  5 4 4 4 4 3 5 2 4 2 ...
##  $ Replacement         : int  5 3 2 4 4 3 6 2 4 2 ...
##  $ Reliability         : int  3 3 4 2 3 2 4 5 3 2 ...
##  $ Options             : int  4 4 4 5 4 4 1 2 4 5 ...
##  $ Respectful          : int  4 3 3 4 4 3 5 3 3 2 ...
##  $ Courteous           : int  3 4 3 3 4 3 5 4 4 3 ...
##  $ Listening           : int  4 4 3 3 5 3 5 5 4 3 ...
```

```
# Creating dummy variables for categorical columns
dfReg <- dummy_cols(dfDropped, remove_selected_columns = TRUE)
names(dfReg) <- gsub(" ", "_", names(dfReg))
names(dfReg) <- gsub("-", "_", names(dfReg))
names(dfReg) <- gsub("[()]", "_", names(dfReg))
str(dfReg)
```

```
## 'data.frame':    10000 obs. of  118 variables:
##  $ Population                          : int  38 10446 3735 13863 11352 17701 2535 23144 17351 201
##  $ Children                            : int  0 1 4 1 0 3 0 2 2 1 ...
##  $ Age                                 : int  68 27 50 48 83 83 79 30 49 86 ...
##  $ Income                              : num  28562 21705 9610 18925 40074 ...
##  $ Outage_sec_perweek                  : num  7.98 11.7 10.75 14.91 8.15 ...
##  $ Email                               : int  10 12 9 15 16 15 10 16 20 18 ...
##  $ Contacts                            : int  0 0 0 2 2 3 0 0 2 1 ...
##  $ Yearly_equip_failure                : int  1 1 1 0 1 1 1 0 3 0 ...
##  $ Tenure                              : num  6.8 1.16 15.75 17.09 1.67 ...
##  $ MonthlyCharge                       : num  172 243 160 120 150 ...
##  $ Bandwidth_GB_Year                   : num  905 801 2055 2165 271 ...
##  $ Response                            : int  5 3 4 4 4 3 6 2 5 2 ...
##  $ Fix                                 : int  5 4 4 4 4 3 5 2 4 2 ...
##  $ Replacement                         : int  5 3 2 4 4 3 6 2 4 2 ...
##  $ Reliability                         : int  3 3 4 2 3 2 4 5 3 2 ...
##  $ Options                             : int  4 4 4 5 4 4 1 2 4 5 ...
##  $ Respectful                          : int  4 3 3 4 4 3 5 3 3 2 ...
##  $ Courteous                           : int  3 4 3 3 4 3 5 4 4 3 ...
##  $ Listening                           : int  4 4 3 3 5 3 5 5 4 3 ...
##  $ State_AK                            : int  1 0 0 0 0 0 0 0 0 0 ...
##  $ State_AL                            : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_AR                            : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_AZ                            : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_CA                            : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ State_CO                            : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_CT                            : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_DC                            : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_DE                            : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_FL                            : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ State_GA                            : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ State_HI                            : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_IA                            : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ State_ID                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_IL                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_IN                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_KS                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_KY                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_LA                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_MA                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_MD                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_ME                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_MI                      : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ State_MN                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_MO                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_MS                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_MT                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_NC                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_ND                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_NE                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_NH                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_NJ                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_NM                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_NV                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_NY                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_OH                      : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ State_OK                      : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ State_OR                      : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ State_PA                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_PR                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_RI                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_SC                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_SD                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_TN                      : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ State_TX                      : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ State_UT                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_VA                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_VT                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_WA                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_WI                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_WV                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ State_WY                      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Area_Rural                    : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ Area_Suburban                 : int  0 0 0 1 1 0 1 1 1 0 ...
##  $ Area_Urban                    : int  1 1 1 0 0 1 0 0 0 0 ...
##  $ Marital_Divorced              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Marital_Married               : int  0 1 0 1 0 0 0 1 0 1 ...
##  $ Marital_Never_Married         : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ Marital_Separated             : int  0 0 0 0 1 0 0 0 1 0 ...
##  $ Marital_Widowed               : int  1 0 1 0 0 0 1 0 0 0 ...
##  $ Gender_Female                 : int  0 1 1 0 0 1 0 1 0 1 ...
##  $ Gender_Male                   : int  1 0 0 1 1 0 1 0 0 0 ...
##  $ Gender_Nonbinary              : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ Churn_No                      : int  1 0 1 1 0 1 0 0 1 1 ...
##  $ Churn_Yes                     : int  0 1 0 0 1 0 1 1 0 0 ...
##  $ Techie_No                     : int  1 0 0 0 1 1 0 0 1 1 ...
##  $ Techie_Yes                    : int  0 1 1 1 0 0 1 1 0 0 ...
```

```
##  $ Contract_Month_to_month              : int  0 1 0 0 1 0 1 1 1 0 ...
##  $ Contract_One_year                    : int  1 0 0 0 0 1 0 0 0 0 ...
##  $ Contract_Two_Year                    : int  0 0 1 1 0 0 0 0 0 1 ...
##  $ Port_modem_No                        : int  0 1 0 1 0 0 1 1 0 0 ...
##  $ Port_modem_Yes                       : int  1 0 1 0 1 1 0 0 1 1 ...
##  $ Tablet_No                            : int  0 0 1 1 1 1 1 1 1 1 ...
##  $ Tablet_Yes                           : int  1 1 0 0 0 0 0 0 0 0 ...
##  $ InternetService_DSL                  : int  0 0 1 1 0 0 1 1 1 0 ...
##  $ InternetService_Fiber_Optic          : int  1 1 0 0 1 0 0 0 0 1 ...
##  $ InternetService_None                 : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ Phone_No                             : int  0 0 0 0 1 0 0 1 0 0 ...
##  $ Phone_Yes                            : int  1 1 1 1 0 1 1 0 1 1 ...
##  $ Multiple_No                          : int  1 0 0 1 1 0 1 1 1 1 ...
##   [list output truncated]
```

```
dim(dfReg)
```

```
## [1] 10000   118
```

```
# Split dfReg into training and testing subsets
set.seed(22)
trainId = createDataPartition(dfReg$Churn_Yes, times = 1, p = 0.7, list = FALSE)

dfTrain = dfReg[trainId,]
dfTest = dfReg[-trainId,]

# Summary Statistics
summary(dfTrain)
```

```
##    Population          Children          Age             Income
##  Min.   :     0.0   Min.   : 0.00   Min.   :18.00   Min.   :   643.2
##  1st Qu.:   733.8   1st Qu.: 0.00   1st Qu.:35.00   1st Qu.: 19349.8
##  Median :  2867.0   Median : 1.00   Median :53.00   Median : 33304.9
##  Mean   :  9683.6   Mean   : 2.06   Mean   :52.89   Mean   : 40080.0
##  3rd Qu.: 12933.0   3rd Qu.: 3.00   3rd Qu.:71.00   3rd Qu.: 53249.4
##  Max.   :111850.0   Max.   :10.00   Max.   :89.00   Max.   :256998.4
##  Outage_sec_perweek     Email          Contacts     Yearly_equip_failure
##  Min.   : 0.09975   Min.   : 1.00   Min.   :0.000   Min.   :0.0000
##  1st Qu.: 8.02346   1st Qu.:10.00   1st Qu.:0.000   1st Qu.:0.0000
##  Median : 9.98095   Median :12.00   Median :1.000   Median :0.0000
##  Mean   : 9.97626   Mean   :12.02   Mean   :1.004   Mean   :0.3939
##  3rd Qu.:11.93163   3rd Qu.:14.00   3rd Qu.:2.000   3rd Qu.:1.0000
##  Max.   :21.20723   Max.   :23.00   Max.   :7.000   Max.   :6.0000
##      Tenure       MonthlyCharge    Bandwidth_GB_Year    Response
##  Min.   : 1.000   Min.   : 79.98   Min.   : 155.5    Min.   :1.00
##  1st Qu.: 7.851   1st Qu.:139.98   1st Qu.:1233.9    1st Qu.:3.00
##  Median :28.095   Median :169.94   Median :3018.9    Median :3.00
##  Mean   :34.282   Mean   :172.84   Mean   :3377.1    Mean   :3.49
##  3rd Qu.:61.354   3rd Qu.:202.44   3rd Qu.:5570.3    3rd Qu.:4.00
##  Max.   :71.994   Max.   :290.16   Max.   :7159.0    Max.   :7.00
##       Fix         Replacement     Reliability       Options         Respectful
##  Min.   :1.00   Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:3.00   1st Qu.:3.000   1st Qu.:3.000   1st Qu.:3.000   1st Qu.:3.000
##  Median :4.00   Median :3.000   Median :4.000   Median :3.000   Median :4.000
##  Mean   :3.51   Mean   :3.491   Mean   :3.507   Mean   :3.485   Mean   :3.516
```

```
##  3rd Qu.:4.00   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :7.00   Max.   :8.000   Max.   :7.000   Max.   :7.000   Max.   :8.000
##    Courteous       Listening        State_AK           State_AL
##  Min.   :1.000   Min.   :1.000   Min.   :0.000000   Min.   :0.00000
##  1st Qu.:3.000   1st Qu.:3.000   1st Qu.:0.000000   1st Qu.:0.00000
##  Median :4.000   Median :4.000   Median :0.000000   Median :0.00000
##  Mean   :3.517   Mean   :3.508   Mean   :0.008143   Mean   :0.01857
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:0.000000   3rd Qu.:0.00000
##  Max.   :7.000   Max.   :8.000   Max.   :1.000000   Max.   :1.00000
##     State_AR          State_AZ          State_CA          State_CO
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01843   Mean   :0.01214   Mean   :0.05486   Mean   :0.01586
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##     State_CT           State_DC           State_DE           State_FL
##  Min.   :0.000000   Min.   :0.000000   Min.   :0.000000   Min.   :0.000
##  1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000
##  Median :0.000000   Median :0.000000   Median :0.000000   Median :0.000
##  Mean   :0.007286   Mean   :0.001429   Mean   :0.002143   Mean   :0.033
##  3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.000
##  Max.   :1.000000   Max.   :1.000000   Max.   :1.000000   Max.   :1.000
##     State_GA          State_HI         State_IA          State_ID
##  Min.   :0.00000   Min.   :0.000   Min.   :0.00000   Min.   :0.000000
##  1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.00000   1st Qu.:0.000000
##  Median :0.00000   Median :0.000   Median :0.00000   Median :0.000000
##  Mean   :0.02314   Mean   :0.003   Mean   :0.02714   Mean   :0.008714
##  3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.00000   3rd Qu.:0.000000
##  Max.   :1.00000   Max.   :1.000   Max.   :1.00000   Max.   :1.000000
##     State_IL          State_IN          State_KS          State_KY
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.04143   Mean   :0.02414   Mean   :0.01929   Mean   :0.02357
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##     State_LA          State_MA          State_MD         State_ME
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.000   Median :0.00000
##  Mean   :0.01386   Mean   :0.01586   Mean   :0.012   Mean   :0.01171
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.000   Max.   :1.00000
##     State_MI         State_MN         State_MO          State_MS
##  Min.   :0.000   Min.   :0.000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.000   Median :0.000   Median :0.00000   Median :0.00000
##  Mean   :0.028   Mean   :0.027   Mean   :0.02929   Mean   :0.01229
##  3rd Qu.:0.000   3rd Qu.:0.000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.000   Max.   :1.000   Max.   :1.00000   Max.   :1.00000
##     State_MT           State_NC          State_ND         State_NE
##  Min.   :0.000000   Min.   :0.00000   Min.   :0.000   Min.   :0.00000
##  1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.00000
```

```
##  Median :0.000000   Median :0.00000   Median :0.000   Median :0.00000
##  Mean   :0.009429   Mean   :0.02843   Mean   :0.012   Mean   :0.01986
##  3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.00000
##  Max.   :1.000000   Max.   :1.00000   Max.   :1.000   Max.   :1.00000
##     State_NH          State_NJ       State_NM          State_NV
##  Min.   :0.000000   Min.   :0.00   Min.   :0.00000   Min.   :0.000000
##  1st Qu.:0.000000   1st Qu.:0.00   1st Qu.:0.00000   1st Qu.:0.000000
##  Median :0.000000   Median :0.00   Median :0.00000   Median :0.000000
##  Mean   :0.008857   Mean   :0.02   Mean   :0.01157   Mean   :0.005429
##  3rd Qu.:0.000000   3rd Qu.:0.00   3rd Qu.:0.00000   3rd Qu.:0.000000
##  Max.   :1.000000   Max.   :1.00   Max.   :1.00000   Max.   :1.000000
##     State_NY          State_OH          State_OK        State_OR
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.000   Min.   :0.00
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.00
##  Median :0.00000   Median :0.00000   Median :0.000   Median :0.00
##  Mean   :0.05471   Mean   :0.03514   Mean   :0.021   Mean   :0.01
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.00
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.000   Max.   :1.00
##     State_PA          State_PR        State_RI        State_SC
##  Min.   :0.00000   Min.   :0.000   Min.   :0.000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.00000
##  Median :0.00000   Median :0.000   Median :0.000   Median :0.00000
##  Mean   :0.05471   Mean   :0.003   Mean   :0.002   Mean   :0.01314
##  3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.000   Max.   :1.000   Max.   :1.00000
##     State_SD          State_TN          State_TX          State_UT
##  Min.   :0.000000   Min.   :0.00000   Min.   :0.00000   Min.   :0.000000
##  1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000000
##  Median :0.000000   Median :0.00000   Median :0.00000   Median :0.000000
##  Mean   :0.008571   Mean   :0.01943   Mean   :0.05971   Mean   :0.006714
##  3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000000
##  Max.   :1.000000   Max.   :1.00000   Max.   :1.00000   Max.   :1.000000
##     State_VA          State_VT        State_WA          State_WI
##  Min.   :0.00000   Min.   :0.000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.000   Median :0.00000   Median :0.00000
##  Mean   :0.02857   Mean   :0.007   Mean   :0.01714   Mean   :0.02329
##  3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.000   Max.   :1.00000   Max.   :1.00000
##     State_WV        State_WY       Area_Rural     Area_Suburban
##  Min.   :0.000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.000   Median :0.000   Median :0.0000   Median :0.0000
##  Mean   :0.024   Mean   :0.004   Mean   :0.3391   Mean   :0.3324
##  3rd Qu.:0.000   3rd Qu.:0.000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000
##    Area_Urban     Marital_Divorced Marital_Married  Marital_Never_Married
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.3284   Mean   :0.2136   Mean   :0.1889   Mean   :0.1943
##  3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  Marital_Separated Marital_Widowed  Gender_Female     Gender_Male
```

```
##  Min.   :0.0000    Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000    1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.0000    Median :0.0000   Median :0.0000   Median :0.000
##  Mean   :0.2024    Mean   :0.2009   Mean   :0.4981   Mean   :0.479
##  3rd Qu.:0.0000    3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.0000    Max.   :1.0000   Max.   :1.0000   Max.   :1.000
##  Gender_Nonbinary     Churn_No        Churn_Yes        Techie_No
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:1.000
##  Median :0.00000   Median :1.0000   Median :0.0000   Median :1.000
##  Mean   :0.02286   Mean   :0.7267   Mean   :0.2733   Mean   :0.833
##  3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000
##    Techie_Yes    Contract_Month_to_month Contract_One_year Contract_Two_Year
##  Min.   :0.000   Min.   :0.000           Min.   :0.0000    Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.000           1st Qu.:0.0000    1st Qu.:0.0000
##  Median :0.000   Median :1.000           Median :0.0000    Median :0.0000
##  Mean   :0.167   Mean   :0.548           Mean   :0.2051    Mean   :0.2469
##  3rd Qu.:0.000   3rd Qu.:1.000           3rd Qu.:0.0000    3rd Qu.:0.0000
##  Max.   :1.000   Max.   :1.000           Max.   :1.0000    Max.   :1.0000
##  Port_modem_No    Port_modem_Yes     Tablet_No        Tablet_Yes
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :0.0000   Median :1.0000   Median :0.0000
##  Mean   :0.5173   Mean   :0.4827   Mean   :0.7063   Mean   :0.2937
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  InternetService_DSL InternetService_Fiber_Optic InternetService_None
##  Min.   :0.0000      Min.   :0.0000              Min.   :0.0000
##  1st Qu.:0.0000      1st Qu.:0.0000              1st Qu.:0.0000
##  Median :0.0000      Median :0.0000              Median :0.0000
##  Mean   :0.3544      Mean   :0.4344              Mean   :0.2111
##  3rd Qu.:1.0000      3rd Qu.:1.0000              3rd Qu.:0.0000
##  Max.   :1.0000      Max.   :1.0000              Max.   :1.0000
##     Phone_No         Phone_Yes       Multiple_No      Multiple_Yes
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.00000   Median :1.0000   Median :1.0000   Median :0.0000
##  Mean   :0.09586   Mean   :0.9041   Mean   :0.5407   Mean   :0.4593
##  3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  OnlineSecurity_No OnlineSecurity_Yes OnlineBackup_No  OnlineBackup_Yes
##  Min.   :0.0000    Min.   :0.0000     Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000    1st Qu.:0.0000     1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000    Median :0.0000     Median :1.0000   Median :0.0000
##  Mean   :0.6446    Mean   :0.3554     Mean   :0.5426   Mean   :0.4574
##  3rd Qu.:1.0000    3rd Qu.:1.0000     3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000    Max.   :1.0000     Max.   :1.0000   Max.   :1.0000
##  DeviceProtection_No DeviceProtection_Yes TechSupport_No   TechSupport_Yes
##  Min.   :0.0000      Min.   :0.0000       Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000      1st Qu.:0.0000       1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000      Median :0.0000       Median :1.0000   Median :0.0000
##  Mean   :0.5626      Mean   :0.4374       Mean   :0.6237   Mean   :0.3763
##  3rd Qu.:1.0000      3rd Qu.:1.0000       3rd Qu.:1.0000   3rd Qu.:1.0000
```

```
##  Max.   :1.0000      Max.   :1.0000       Max.   :1.0000   Max.   :1.0000
##  StreamingTV_No  StreamingTV_Yes StreamingMovies_No StreamingMovies_Yes
##  Min.   :0.000   Min.   :0.000   Min.   :0.0000     Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.0000     1st Qu.:0.0000
##  Median :1.000   Median :0.000   Median :1.0000     Median :0.0000
##  Mean   :0.503   Mean   :0.497   Mean   :0.5107     Mean   :0.4893
##  3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.0000     3rd Qu.:1.0000
##  Max.   :1.000   Max.   :1.000   Max.   :1.0000     Max.   :1.0000
##  PaperlessBilling_No PaperlessBilling_Yes
##  Min.   :0.000       Min.   :0.000
##  1st Qu.:0.000       1st Qu.:0.000
##  Median :0.000       Median :1.000
##  Mean   :0.414       Mean   :0.586
##  3rd Qu.:1.000       3rd Qu.:1.000
##  Max.   :1.000       Max.   :1.000
##  PaymentMethod_Bank_Transfer_automatic_ PaymentMethod_Credit_Card__automatic_
##  Min.   :0.000                          Min.   :0.000
##  1st Qu.:0.000                          1st Qu.:0.000
##  Median :0.000                          Median :0.000
##  Mean   :0.225                          Mean   :0.202
##  3rd Qu.:0.000                          3rd Qu.:0.000
##  Max.   :1.000                          Max.   :1.000
##  PaymentMethod_Electronic_Check PaymentMethod_Mailed_Check
##  Min.   :0.0000                 Min.   :0.0000
##  1st Qu.:0.0000                 1st Qu.:0.0000
##  Median :0.0000                 Median :0.0000
##  Mean   :0.3399                 Mean   :0.2331
##  3rd Qu.:1.0000                 3rd Qu.:0.0000
##  Max.   :1.0000                 Max.   :1.0000
```

```r
# Normalize Function
normalize = function(x) {
  result = (x - min(x)) / (max(x) - min(x))
  return(result)
}

# Normalize training set
dfTrainNorm <- dfTrain
for (i in 1:19) {
  dfTrainNorm[i] <- normalize(dfTrainNorm[i])
}
# Normalize testing set
dfTestNorm <- dfTest

for (i in 1:19) {
  dfTestNorm[i] <- normalize(dfTestNorm[i])
}

# Summary Statistics
summary(dfTrainNorm)
```

```
##    Population         Children          Age              Income
##  Min.   :0.00000   Min.   :0.000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00656   1st Qu.:0.000   1st Qu.:0.2394   1st Qu.:0.07297
##  Median :0.02563   Median :0.100   Median :0.4930   Median :0.12741
```

```
##  Mean   :0.08658   Mean   :0.206   Mean   :0.4914   Mean   :0.15384
##  3rd Qu.:0.11563   3rd Qu.:0.300   3rd Qu.:0.7465   3rd Qu.:0.20521
##  Max.   :1.00000   Max.   :1.000   Max.   :1.0000   Max.   :1.00000
##  Outage_sec_perweek     Email           Contacts      Yearly_equip_failure
##  Min.   :0.0000     Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.3754     1st Qu.:0.4091   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.4681     Median :0.5000   Median :0.1429   Median :0.00000
##  Mean   :0.4679     Mean   :0.5008   Mean   :0.1434   Mean   :0.06564
##  3rd Qu.:0.5606     3rd Qu.:0.5909   3rd Qu.:0.2857   3rd Qu.:0.16667
##  Max.   :1.0000     Max.   :1.0000   Max.   :1.0000   Max.   :1.00000
##      Tenure        MonthlyCharge    Bandwidth_GB_Year    Response
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000    Min.   :0.0000
##  1st Qu.:0.0965   1st Qu.:0.2855   1st Qu.:0.1540    1st Qu.:0.3333
##  Median :0.3817   Median :0.4280   Median :0.4088    Median :0.3333
##  Mean   :0.4688   Mean   :0.4418   Mean   :0.4600    Mean   :0.4150
##  3rd Qu.:0.8501   3rd Qu.:0.5827   3rd Qu.:0.7732    3rd Qu.:0.5000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000    Max.   :1.0000
##       Fix           Replacement      Reliability        Options
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.3333   1st Qu.:0.2857   1st Qu.:0.3333   1st Qu.:0.3333
##  Median :0.5000   Median :0.2857   Median :0.5000   Median :0.3333
##  Mean   :0.4183   Mean   :0.3558   Mean   :0.4179   Mean   :0.4141
##  3rd Qu.:0.5000   3rd Qu.:0.4286   3rd Qu.:0.5000   3rd Qu.:0.5000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##    Respectful       Courteous        Listening         State_AK
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000000
##  1st Qu.:0.2857   1st Qu.:0.3333   1st Qu.:0.2857   1st Qu.:0.000000
##  Median :0.4286   Median :0.5000   Median :0.4286   Median :0.000000
##  Mean   :0.3595   Mean   :0.4195   Mean   :0.3582   Mean   :0.008143
##  3rd Qu.:0.4286   3rd Qu.:0.5000   3rd Qu.:0.4286   3rd Qu.:0.000000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000000
##     State_AL          State_AR          State_AZ          State_CA
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01857   Mean   :0.01843   Mean   :0.01214   Mean   :0.05486
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##     State_CO          State_CT           State_DC           State_DE
##  Min.   :0.00000   Min.   :0.000000   Min.   :0.000000   Min.   :0.000000
##  1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000
##  Median :0.00000   Median :0.000000   Median :0.000000   Median :0.000000
##  Mean   :0.01586   Mean   :0.007286   Mean   :0.001429   Mean   :0.002143
##  3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.000000
##  Max.   :1.00000   Max.   :1.000000   Max.   :1.000000   Max.   :1.000000
##     State_FL         State_GA          State_HI         State_IA
##  Min.   :0.000   Min.   :0.00000   Min.   :0.000   Min.   :0.00000
##  1st Qu.:0.000   1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.00000
##  Median :0.000   Median :0.00000   Median :0.000   Median :0.00000
##  Mean   :0.033   Mean   :0.02314   Mean   :0.003   Mean   :0.02714
##  3rd Qu.:0.000   3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.00000
##  Max.   :1.000   Max.   :1.00000   Max.   :1.000   Max.   :1.00000
##     State_ID           State_IL          State_IN          State_KS
##  Min.   :0.000000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
```

```
##  1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.000000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.008714   Mean   :0.04143   Mean   :0.02414   Mean   :0.01929
##  3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.000000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##     State_KY          State_LA          State_MA          State_MD
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.000
##  Mean   :0.02357   Mean   :0.01386   Mean   :0.01586   Mean   :0.012
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.000
##     State_ME          State_MI         State_MN         State_MO
##  Min.   :0.00000   Min.   :0.000   Min.   :0.000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.00000
##  Median :0.00000   Median :0.000   Median :0.000   Median :0.00000
##  Mean   :0.01171   Mean   :0.028   Mean   :0.027   Mean   :0.02929
##  3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.000   Max.   :1.000   Max.   :1.00000
##     State_MS          State_MT           State_NC          State_ND
##  Min.   :0.00000   Min.   :0.000000   Min.   :0.00000   Min.   :0.000
##  1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.000
##  Median :0.00000   Median :0.000000   Median :0.00000   Median :0.000
##  Mean   :0.01229   Mean   :0.009429   Mean   :0.02843   Mean   :0.012
##  3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.000
##  Max.   :1.00000   Max.   :1.000000   Max.   :1.00000   Max.   :1.000
##     State_NE          State_NH           State_NJ       State_NM
##  Min.   :0.00000   Min.   :0.000000   Min.   :0.00   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.00   1st Qu.:0.00000
##  Median :0.00000   Median :0.000000   Median :0.00   Median :0.00000
##  Mean   :0.01986   Mean   :0.008857   Mean   :0.02   Mean   :0.01157
##  3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.00   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.000000   Max.   :1.00   Max.   :1.00000
##     State_NV           State_NY          State_OH          State_OK
##  Min.   :0.000000   Min.   :0.00000   Min.   :0.00000   Min.   :0.000
##  1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000
##  Median :0.000000   Median :0.00000   Median :0.00000   Median :0.000
##  Mean   :0.005429   Mean   :0.05471   Mean   :0.03514   Mean   :0.021
##  3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000
##  Max.   :1.000000   Max.   :1.00000   Max.   :1.00000   Max.   :1.000
##     State_OR       State_PA          State_PR        State_RI
##  Min.   :0.00   Min.   :0.00000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.00   1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.000
##  Median :0.00   Median :0.00000   Median :0.000   Median :0.000
##  Mean   :0.01   Mean   :0.05471   Mean   :0.003   Mean   :0.002
##  3rd Qu.:0.00   3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.000
##  Max.   :1.00   Max.   :1.00000   Max.   :1.000   Max.   :1.000
##     State_SC          State_SD           State_TN          State_TX
##  Min.   :0.00000   Min.   :0.000000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.000000   Median :0.00000   Median :0.00000
##  Mean   :0.01314   Mean   :0.008571   Mean   :0.01943   Mean   :0.05971
##  3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.000000   Max.   :1.00000   Max.   :1.00000
```

```
##     State_UT           State_VA          State_VT        State_WA
##  Min.   :0.000000   Min.   :0.00000   Min.   :0.000   Min.   :0.00000
##  1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.00000
##  Median :0.000000   Median :0.00000   Median :0.000   Median :0.00000
##  Mean   :0.006714   Mean   :0.02857   Mean   :0.007   Mean   :0.01714
##  3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.00000
##  Max.   :1.000000   Max.   :1.00000   Max.   :1.000   Max.   :1.00000
##     State_WI          State_WV        State_WY       Area_Rural
##  Min.   :0.00000   Min.   :0.000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :0.00000   Median :0.000   Median :0.000   Median :0.0000
##  Mean   :0.02329   Mean   :0.024   Mean   :0.004   Mean   :0.3391
##  3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:0.000   3rd Qu.:1.0000
##  Max.   :1.00000   Max.   :1.000   Max.   :1.000   Max.   :1.0000
##  Area_Suburban      Area_Urban     Marital_Divorced Marital_Married
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.3324   Mean   :0.3284   Mean   :0.2136   Mean   :0.1889
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##  Marital_Never_Married Marital_Separated Marital_Widowed  Gender_Female
##  Min.   :0.0000        Min.   :0.0000    Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000        1st Qu.:0.0000    1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000        Median :0.0000    Median :0.0000   Median :0.0000
##  Mean   :0.1943        Mean   :0.2024    Mean   :0.2009   Mean   :0.4981
##  3rd Qu.:0.0000        3rd Qu.:0.0000    3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :1.0000        Max.   :1.0000    Max.   :1.0000   Max.   :1.0000
##   Gender_Male    Gender_Nonbinary     Churn_No        Churn_Yes
##  Min.   :0.000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.000   Median :0.00000   Median :1.0000   Median :0.0000
##  Mean   :0.479   Mean   :0.02286   Mean   :0.7267   Mean   :0.2733
##  3rd Qu.:1.000   3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.000   Max.   :1.00000   Max.   :1.0000   Max.   :1.0000
##    Techie_No       Techie_Yes    Contract_Month_to_month Contract_One_year
##  Min.   :0.000   Min.   :0.000   Min.   :0.000           Min.   :0.0000
##  1st Qu.:1.000   1st Qu.:0.000   1st Qu.:0.000           1st Qu.:0.0000
##  Median :1.000   Median :0.000   Median :1.000           Median :0.0000
##  Mean   :0.833   Mean   :0.167   Mean   :0.548           Mean   :0.2051
##  3rd Qu.:1.000   3rd Qu.:0.000   3rd Qu.:1.000           3rd Qu.:0.0000
##  Max.   :1.000   Max.   :1.000   Max.   :1.000           Max.   :1.0000
##  Contract_Two_Year Port_modem_No    Port_modem_Yes     Tablet_No
##  Min.   :0.0000    Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000    1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000    Median :1.0000   Median :0.0000   Median :1.0000
##  Mean   :0.2469    Mean   :0.5173   Mean   :0.4827   Mean   :0.7063
##  3rd Qu.:0.0000    3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000    Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##    Tablet_Yes     InternetService_DSL InternetService_Fiber_Optic
##  Min.   :0.0000   Min.   :0.0000      Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000      1st Qu.:0.0000
##  Median :0.0000   Median :0.0000      Median :0.0000
##  Mean   :0.2937   Mean   :0.3544      Mean   :0.4344
```

```
##  3rd Qu.:1.0000   3rd Qu.:1.0000      3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000      Max.   :1.0000
##  InternetService_None    Phone_No         Phone_Yes       Multiple_No
##  Min.   :0.0000       Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000       1st Qu.:0.00000   1st Qu.:1.0000   1st Qu.:0.0000
##  Median :0.0000       Median :0.00000   Median :1.0000   Median :1.0000
##  Mean   :0.2111       Mean   :0.09586   Mean   :0.9041   Mean   :0.5407
##  3rd Qu.:0.0000       3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000       Max.   :1.00000   Max.   :1.0000   Max.   :1.0000
##   Multiple_Yes    OnlineSecurity_No OnlineSecurity_Yes OnlineBackup_No
##  Min.   :0.0000   Min.   :0.0000    Min.   :0.0000     Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000    1st Qu.:0.0000     1st Qu.:0.0000
##  Median :0.0000   Median :1.0000    Median :0.0000     Median :1.0000
##  Mean   :0.4593   Mean   :0.6446    Mean   :0.3554     Mean   :0.5426
##  3rd Qu.:1.0000   3rd Qu.:1.0000    3rd Qu.:1.0000     3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000    Max.   :1.0000     Max.   :1.0000
##  OnlineBackup_Yes DeviceProtection_No DeviceProtection_Yes TechSupport_No
##  Min.   :0.0000   Min.   :0.0000      Min.   :0.0000       Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000      1st Qu.:0.0000       1st Qu.:0.0000
##  Median :0.0000   Median :1.0000      Median :0.0000       Median :1.0000
##  Mean   :0.4574   Mean   :0.5626      Mean   :0.4374       Mean   :0.6237
##  3rd Qu.:1.0000   3rd Qu.:1.0000      3rd Qu.:1.0000       3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000      Max.   :1.0000       Max.   :1.0000
##  TechSupport_Yes  StreamingTV_No  StreamingTV_Yes StreamingMovies_No
##  Min.   :0.0000   Min.   :0.000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :0.0000   Median :1.000   Median :0.000   Median :1.0000
##  Mean   :0.3763   Mean   :0.503   Mean   :0.497   Mean   :0.5107
##  3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.000   Max.   :1.0000
##  StreamingMovies_Yes PaperlessBilling_No PaperlessBilling_Yes
##  Min.   :0.0000      Min.   :0.000       Min.   :0.000
##  1st Qu.:0.0000      1st Qu.:0.000       1st Qu.:0.000
##  Median :0.0000      Median :0.000       Median :1.000
##  Mean   :0.4893      Mean   :0.414       Mean   :0.586
##  3rd Qu.:1.0000      3rd Qu.:1.000       3rd Qu.:1.000
##  Max.   :1.0000      Max.   :1.000       Max.   :1.000
##  PaymentMethod_Bank_Transfer_automatic_ PaymentMethod_Credit_Card__automatic_
##  Min.   :0.000                          Min.   :0.000
##  1st Qu.:0.000                          1st Qu.:0.000
##  Median :0.000                          Median :0.000
##  Mean   :0.225                          Mean   :0.202
##  3rd Qu.:0.000                          3rd Qu.:0.000
##  Max.   :1.000                          Max.   :1.000
##  PaymentMethod_Electronic_Check PaymentMethod_Mailed_Check
##  Min.   :0.0000                 Min.   :0.0000
##  1st Qu.:0.0000                 1st Qu.:0.0000
##  Median :0.0000                 Median :0.0000
##  Mean   :0.3399                 Mean   :0.2331
##  3rd Qu.:1.0000                 3rd Qu.:0.0000
##  Max.   :1.0000                 Max.   :1.0000
```

```
# Export prepared data sets
write.csv(dfTrainNorm, "C:\\WGU\\D209 Data Mining I\\PA Task 1\\D209_dfTrainNorm.csv", row.names = FALSE
```

```r
write.csv(dfTestNorm, "C:\\WGU\\D209 Data Mining I\\PA Task 1\\D209_dfTestNorm.csv", row.names = FALSE)

dfTrainChurn <- dfTrainNorm[, -which(names(dfTrainNorm) %in% c("Churn_No"))]
dfTestChurn <- dfTestNorm[, -which(names(dfTestNorm) %in% c("Churn_No"))]

# Finding the best k-value for accuracy

# Initial guess
kVal <- sqrt(nrow(dfTrainChurn)) # kval = 83.666

predFloor <- knn(train = dfTrainChurn, test = dfTestChurn, cl = dfTrainChurn$Churn_Yes, k = floor(kVal))
outcomesFloor <- table(predFloor, dfTestChurn$Churn_Yes)
confFloor <- confusionMatrix(outcomesFloor)
confFloor # Accuracy 0.94
```

```
## Confusion Matrix and Statistics
##
##
## predFloor    0    1
##         0 2263  180
##         1    0  557
##
##                Accuracy : 0.94
##                  95% CI : (0.9309, 0.9482)
##     No Information Rate : 0.7543
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8236
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 1.0000
##             Specificity : 0.7558
##          Pos Pred Value : 0.9263
##          Neg Pred Value : 1.0000
##              Prevalence : 0.7543
##          Detection Rate : 0.7543
##    Detection Prevalence : 0.8143
##       Balanced Accuracy : 0.8779
##
##        'Positive' Class : 0
##
```

```r
predCeil <- knn(train = dfTrainChurn, test = dfTestChurn, cl = dfTrainChurn$Churn_Yes, k = ceiling(kVal))
outcomesCeil <- table(predCeil, dfTestChurn$Churn_Yes)
confCeil <- confusionMatrix(outcomesCeil)
confCeil # Accuracy 0.939
```

```
## Confusion Matrix and Statistics
##
##
## predCeil    0    1
##        0 2263  186
##        1    0  551
```

```
##
##                Accuracy : 0.938
##                  95% CI : (0.9288, 0.9464)
##     No Information Rate : 0.7543
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8172
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 1.0000
##             Specificity : 0.7476
##          Pos Pred Value : 0.9241
##          Neg Pred Value : 1.0000
##              Prevalence : 0.7543
##          Detection Rate : 0.7543
##    Detection Prevalence : 0.8163
##       Balanced Accuracy : 0.8738
##
##        'Positive' Class : 0
##
```

```
# Checking other values
pred100 <- knn(train = dfTrainChurn, test = dfTestChurn, cl = dfTrainChurn$Churn_Yes, k = 100)
outcomes100 <- table(pred100, dfTestChurn$Churn_Yes)
conf100 <- confusionMatrix(outcomes100)
conf100 # Accuracy 0.938
```

```
## Confusion Matrix and Statistics
##
##
## pred100    0    1
##       0 2262  188
##       1    1  549
##
##                Accuracy : 0.937
##                  95% CI : (0.9277, 0.9454)
##     No Information Rate : 0.7543
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8141
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.9996
##             Specificity : 0.7449
##          Pos Pred Value : 0.9233
##          Neg Pred Value : 0.9982
##              Prevalence : 0.7543
##          Detection Rate : 0.7540
##    Detection Prevalence : 0.8167
##       Balanced Accuracy : 0.8722
##
##        'Positive' Class : 0
##
```

```r
pred75 <- knn(train = dfTrainChurn, test = dfTestChurn, cl = dfTrainChurn$Churn_Yes, k = 75)
outcomes75 <- table(pred75, dfTestChurn$Churn_Yes)
conf75 <- confusionMatrix(outcomes75)
conf75 # Accuracy 0.9403
```

```
## Confusion Matrix and Statistics
##
##
## pred75    0    1
##      0 2263  179
##      1    0  558
##
##                Accuracy : 0.9403
##                  95% CI : (0.9313, 0.9485)
##     No Information Rate : 0.7543
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8247
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 1.0000
##             Specificity : 0.7571
##          Pos Pred Value : 0.9267
##          Neg Pred Value : 1.0000
##              Prevalence : 0.7543
##          Detection Rate : 0.7543
##    Detection Prevalence : 0.8140
##       Balanced Accuracy : 0.8786
##
##        'Positive' Class : 0
##
```

```r
pred50 <- knn(train = dfTrainChurn, test = dfTestChurn, cl = dfTrainChurn$Churn_Yes, k = 50)
outcomes50 <- table(pred50, dfTestChurn$Churn_Yes)
conf50 <- confusionMatrix(outcomes50)
conf50 # Accuracy 0.9513
```

```
## Confusion Matrix and Statistics
##
##
## pred50    0    1
##      0 2261  148
##      1    2  589
##
##                Accuracy : 0.95
##                  95% CI : (0.9416, 0.9575)
##     No Information Rate : 0.7543
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8554
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
```

```
##             Sensitivity : 0.9991
##             Specificity : 0.7992
##          Pos Pred Value : 0.9386
##          Neg Pred Value : 0.9966
##              Prevalence : 0.7543
##          Detection Rate : 0.7537
##    Detection Prevalence : 0.8030
##       Balanced Accuracy : 0.8992
##
##        'Positive' Class : 0
##
```

```r
pred25 <- knn(train = dfTrainChurn, test = dfTestChurn, cl = dfTrainChurn$Churn_Yes, k = 25)
outcomes25 <- table(pred25, dfTestChurn$Churn_Yes)
conf25 <- confusionMatrix(outcomes25)
conf25 # Accuracy 0.948
```

```
## Confusion Matrix and Statistics
##
##
## pred25    0    1
##      0 2258  151
##      1    5  586
##
##                Accuracy : 0.948
##                  95% CI : (0.9394, 0.9557)
##     No Information Rate : 0.7543
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8497
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.9978
##             Specificity : 0.7951
##          Pos Pred Value : 0.9373
##          Neg Pred Value : 0.9915
##              Prevalence : 0.7543
##          Detection Rate : 0.7527
##    Detection Prevalence : 0.8030
##       Balanced Accuracy : 0.8965
##
##        'Positive' Class : 0
##
```

```r
# k=50 gave the best for a random guess. Checking the 10 on either side of 50.

## The code below was borrowed in part from LearnByMarketing.com
## https://www.learnbymarketing.com/tutorials/k-nearest-neighbors-in-r-example/#:~:text=K-Nearest-Neigh

churn_acc <- numeric() #Holding variable

# This takes about 2.5-3 minutes to complete.
for(i in 40:60){
  #Apply knn with k = i
```

```
  predict <- knn(dfTrainChurn,dfTestChurn,dfTrainChurn$Churn_Yes,k=i)
  churn_acc <- c(churn_acc, mean(predict==dfTestChurn$Churn_Yes))
}

## end third party block
knnAcc <- tibble(x=40:60, y=churn_acc)
knnAcc[which.max(knnAcc$y),]
```

```
## # A tibble: 1 x 2
##       x     y
##   <int> <dbl>
## 1    43 0.953
```

```
# k=43 is the best for accuracy.
pred43 <- knn(train = dfTrainChurn, test = dfTestChurn, cl = dfTrainChurn$Churn_Yes, k = 43, prob = TRU
outcomes43 <- table(pred43, dfTestChurn$Churn_Yes)
conf43 <- confusionMatrix(outcomes43)
conf43 # Accuracy 0.9527
```

```
## Confusion Matrix and Statistics
##
##
## pred43    0    1
##      0 2261  140
##      1    2  597
##
##                Accuracy : 0.9527
##                  95% CI : (0.9444, 0.96)
##     No Information Rate : 0.7543
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.8637
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.9991
##             Specificity : 0.8100
##          Pos Pred Value : 0.9417
##          Neg Pred Value : 0.9967
##              Prevalence : 0.7543
##          Detection Rate : 0.7537
##    Detection Prevalence : 0.8003
##       Balanced Accuracy : 0.9046
##
##        'Positive' Class : 0
##
```

```
pred43Numeric <- as.numeric(as.character(pred43))

rocPlot <- roc(response = dfTestChurn$Churn_Yes,
               predictor = pred43Numeric,
               percent = TRUE,
               ci=TRUE,
               boot.n=100,
               ci.alpha=-0.9,
```

```
                stratified=FALSE,
                plot=TRUE,
                grid=TRUE,
                print.auc=TRUE,
                show.thres=TRUE)
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```