

Linear Regression on Telecom Churn Data

Alexander Vaillant

9/7/2021

Environment Setup

Import Necessary Libraries

```
# Load in necessary libraries using library()
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(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.1.1
## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.4      v stringr 1.4.0
## v tidyr   1.1.3      v forcats 0.5.1
## v readr   2.0.1

## Warning: package 'tibble' was built under R version 4.1.1
## Warning: package 'tidyr' was built under R version 4.1.1
## Warning: package 'readr' was built under R version 4.1.1
## Warning: package 'purrr' was built under R version 4.1.1
## Warning: package 'forcats' was built under R version 4.1.1

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

library(caret)

## Warning: package 'caret' was built under R version 4.1.1
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
## The following object is masked from 'package:purrr':
##
##      lift
library(leaps)

## Warning: package 'leaps' was built under R version 4.1.1
library(reshape2)

## Warning: package 'reshape2' was built under R version 4.1.1
##
## Attaching package: 'reshape2'
## The following object is masked from 'package:tidyr':
##
##      smiths
library(fastDummies) #Create dummy columns easily

## Warning: package 'fastDummies' was built under R version 4.1.1
library(plyr) # Rename columns

## Warning: package 'plyr' was built under R version 4.1.1
## -----
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## -----
##
## Attaching package: 'plyr'
## The following object is masked from 'package:purrr':
##
##      compact
## The following objects are masked from 'package:dplyr':
##
##      arrange, count, desc, failwith, id, mutate, rename, summarise,
##      summarize
```

Data Gathering

Load Dataset into Dataframe using read.csv()

```
# Import the raw dataset by using read.csv()
url <- "C:/Users/tedda/Desktop/Data Science Portfolio/Machine Learning/Supervised Learning/Regression/L
churndata <- read.csv(url, header = TRUE)
```

Data Preparation

```
# Remove any customer demographic data by indexing
churn_indexed <- churndata[c(20:50)]

# Filter the dataset to only Month-to-Month customers by using dplyr::filter()
churn_filtered <- churn_indexed %>% filter(Contract=='Month-to-month')

# Transform the categorical variables into dummy variable columns by using fastDummies::dummy_cols()
churn_dummies <- dummy_cols(churn_filtered, remove_first_dummy = TRUE, remove_selected_columns = TRUE)

# Index again to remove any unnecessary columns (Contract)
churn_index <- churn_dummies[c(1:17,19:34)]

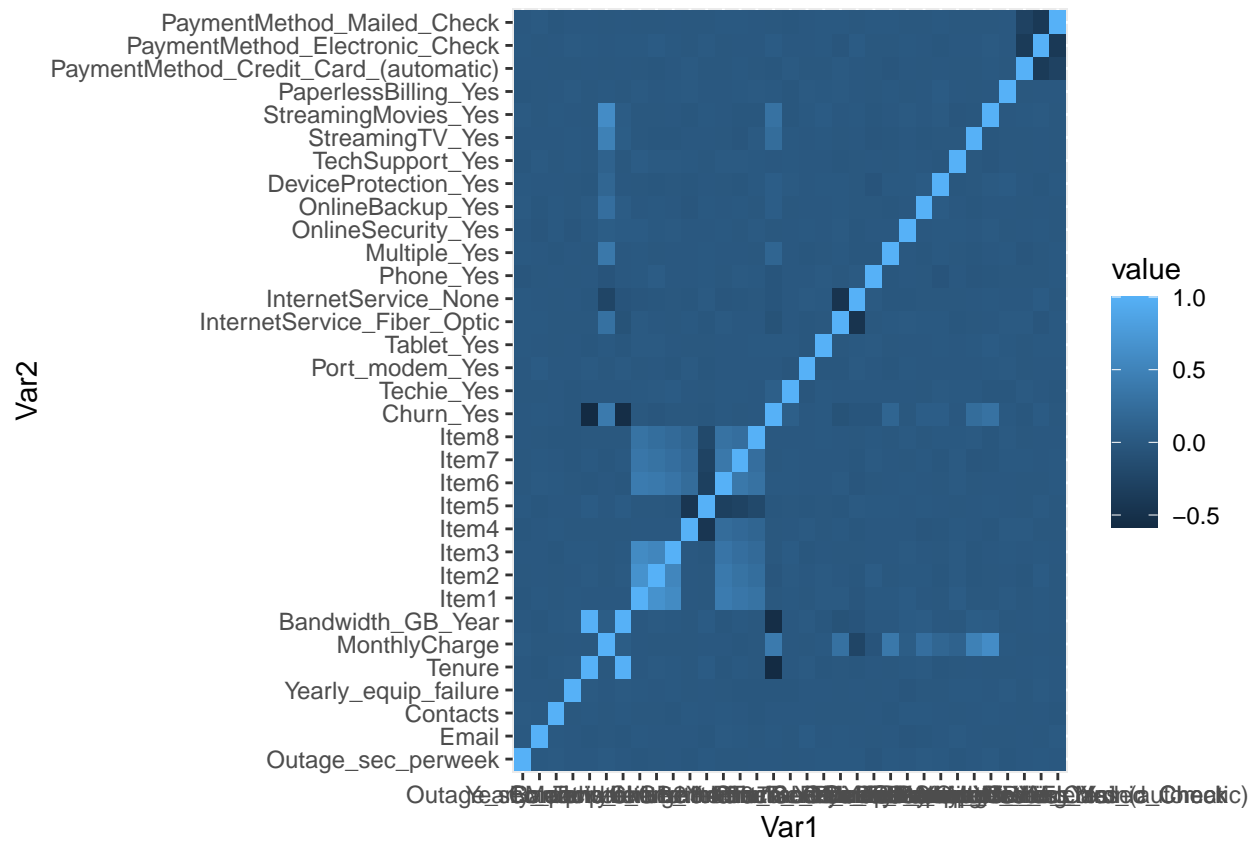
# Rename any variables with spaces in their name by using plyr::rename()
churn_renamed <- rename(churn_index, replace = c("InternetService_Fiber Optic" = "InternetService_Fiber",
churn_renamed <- rename(churn_renamed, replace = c("PaymentMethod_Credit Card (automatic)" = "PaymentMethod_Credit_Card_Automatic",
churn_renamed <- rename(churn_renamed, replace = c("PaymentMethod_Electronic Check" = "PaymentMethod_Electronic_Check",
churn_renamed <- rename(churn_renamed, replace = c("PaymentMethod_Mailed Check" = "PaymentMethod_Mailed_Check",

# Create a Summary Output of the dataset by using summary()
summary(churn_renamed)
```

```
## Outage_sec_perweek      Email      Contacts      Yearly_equip_failure
## Min.   : 0.2323      Min.   : 1.00      Min.   :0.0000      Min.   :0.0000
## 1st Qu.: 7.9702      1st Qu.:10.00      1st Qu.:0.0000      1st Qu.:0.0000
## Median : 9.9938      Median :12.00      Median :1.0000      Median :0.0000
## Mean   : 9.9784      Mean   :12.04      Mean   :0.9901      Mean   :0.3961
## 3rd Qu.:11.9439      3rd Qu.:14.00      3rd Qu.:2.0000      3rd Qu.:1.0000
## Max.   :21.2072      Max.   :23.00      Max.   :7.0000      Max.   :4.0000
##      Tenure      MonthlyCharge      Bandwidth_GB_Year      Item1
## Min.   : 1.019      Min.   : 79.98      Min.   : 169.4      Min.   :1.000
## 1st Qu.: 7.900      1st Qu.:139.98      1st Qu.:1211.0      1st Qu.:3.000
## Median :26.965      Median :169.94      Median :2914.5      Median :3.000
## Mean   :34.241      Mean   :172.14      Mean   :3363.4      Mean   :3.486
## 3rd Qu.:61.303      3rd Qu.:200.12      3rd Qu.:5566.0      3rd Qu.:4.000
## Max.   :71.999      Max.   :290.16      Max.   :7159.0      Max.   :7.000
##      Item2      Item3      Item4      Item5
## Min.   :1.000      Min.   :1.000      Min.   :1.000      Min.   :1.000
## 1st Qu.:3.000      1st Qu.:3.000      1st Qu.:3.000      1st Qu.:3.000
## Median :3.000      Median :3.000      Median :4.000      Median :3.000
## Mean   :3.505      Mean   :3.473      Mean   :3.503      Mean   :3.481
## 3rd Qu.:4.000      3rd Qu.:4.000      3rd Qu.:4.000      3rd Qu.:4.000
## Max.   :7.000      Max.   :8.000      Max.   :7.000      Max.   :7.000
##      Item6      Item7      Item8      Churn_Yes
## Min.   :1.000      Min.   :1.000      Min.   :1.000      Min.   :0.0000
## 1st Qu.:3.000      1st Qu.:3.000      1st Qu.:3.000      1st Qu.:0.0000
## Median :3.000      Median :3.000      Median :3.000      Median :0.0000
## Mean   :3.491      Mean   :3.505      Mean   :3.507      Mean   :0.3728
## 3rd Qu.:4.000      3rd Qu.:4.000      3rd Qu.:4.000      3rd Qu.:1.0000
## Max.   :8.000      Max.   :7.000      Max.   :7.000      Max.   :1.0000
##      Techie_Yes      Port_modem_Yes      Tablet_Yes      InternetService_Fiber_Optic
## 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   :0.1683    Mean   :0.4872    Mean   :0.2942    Mean   :0.4434
## 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
## InternetService_None Phone_Yes Multiple_Yes OnlineSecurity_Yes
## Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   :0.0000
## 1st Qu.:0.0000    1st Qu.:1.0000    1st Qu.:0.0000    1st Qu.:0.0000
## Median :0.0000    Median :1.0000    Median :0.0000    Median :0.0000
## Mean   :0.2124    Mean   :0.9036    Mean   :0.4514    Mean   :0.3512
## 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
## OnlineBackup_Yes DeviceProtection_Yes TechSupport_Yes StreamingTV_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 :0.0000    Median :0.0000    Median :0.0000    Median :0.0000
## Mean   :0.4529    Mean   :0.438    Mean   :0.3766    Mean   :0.4839
## 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
## StreamingMovies_Yes PaperlessBilling_Yes PaymentMethod_Credit_Card_(automatic)
## 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 :1.0000    Median :0.0000
## Mean   :0.4925    Mean   :0.5854    Mean   :0.2084
## 3rd Qu.:1.0000    3rd Qu.:1.0000    3rd Qu.:0.0000
## Max.   :1.0000    Max.   :1.0000    Max.   :1.0000
## 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   :0.3435    Mean   :0.2275
## 3rd Qu.:1.0000    3rd Qu.:0.0000
## Max.   :1.0000    Max.   :1.0000
```

```
# Create a correlation matrix to help identify potential issues for multicollinearity
cormatrix <- round(cor(churn_renamed),2)
melted_cormatrix <- melt(cormatrix)
ggplot(melted_cormatrix, aes(x = Var1, y = Var2, fill = value)) + geom_tile()
```



```
cormatrix[, "Tenure"]
```

```
##          Outage_sec_perweek          Email
##          0.00          -0.02
##          Contacts          Yearly_equip_failure
##          0.01          0.02
##          Tenure          MonthlyCharge
##          1.00          -0.01
##          Bandwidth_GB_Year          Item1
##          0.99          0.00
##          Item2          Item3
##          0.02          0.01
##          Item4          Item5
##          -0.01          0.03
##          Item6          Item7
##          -0.02          0.01
##          Item8          Churn_Yes
##          -0.01          -0.58
##          Techie_Yes          Port_modem_Yes
##          0.00          0.02
##          Tablet_Yes          InternetService_Fiber_Optic
##          -0.02          -0.01
##          InternetService_None          Phone_Yes
##          -0.02          0.00
##          Multiple_Yes          OnlineSecurity_Yes
##          -0.01          0.01
```

```
##           OnlineBackup_Yes           DeviceProtection_Yes
##                0.02                -0.02
##           TechSupport_Yes           StreamingTV_Yes
##                -0.01                0.00
##           StreamingMovies_Yes           PaperlessBilling_Yes
##                -0.01                -0.01
## PaymentMethod_Credit_Card_(automatic) PaymentMethod_Electronic_Check
##                0.00                0.01
##           PaymentMethod_Mailed_Check
##                0.00

write.csv(cormatrix,"C:/Users/tedda/Desktop/Data Science Portfolio/Machine Learning/Supervised Learning/

# Remove Bandwidth_GB_Year from the analysis and dataset by indexing as it is highly correlated with Tenure
churn_nomult <- churn_renamed[c(1:6,8:33)]

# Export the prepared dataset
write.csv(churn_nomult, "C:/Users/tedda/Desktop/Data Science Portfolio/Machine Learning/Supervised Learning/
```

Exploratory Data Analysis on Initial Model

```
# Create the gross/initial model w/ Multicollinearity
MLR_GrossModel_Mlt <- lm(Tenure ~ ., churn_renamed)
summary(MLR_GrossModel_Mlt)

##
## Call:
## lm(formula = Tenure ~ ., data = churn_renamed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.8458 -0.7777  0.0821  0.8736  2.8542
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -3.012e+00  2.483e-01  -12.130 < 2e-16
## Outage_sec_perweek -6.073e-03  5.539e-03   -1.096  0.272973
## Email          -7.792e-03  5.475e-03   -1.423  0.154776
## Contacts         2.815e-02  1.673e-02    1.683  0.092459
## Yearly_equip_failure -1.273e-02  2.612e-02   -0.487  0.625967
## MonthlyCharge    -3.188e-02  1.898e-03  -16.800 < 2e-16
## Bandwidth_GB_Year  1.216e-02  1.009e-05 1205.282 < 2e-16
## Item1           3.053e-02  2.373e-02    1.286  0.198349
## Item2          -2.185e-02  2.219e-02   -0.985  0.324785
## Item3          -1.249e-03  2.024e-02   -0.062  0.950777
## Item4           3.879e-02  1.834e-02    2.115  0.034468
## Item5          -1.277e-02  1.906e-02   -0.670  0.502898
## Item6          -1.015e-02  1.960e-02   -0.518  0.604515
## Item7          -1.342e-02  1.846e-02   -0.727  0.467181
## Item8           1.851e-02  1.742e-02    1.062  0.288090
## Churn_Yes       -1.620e-01  5.055e-02   -3.204  0.001363
## Techie_Yes       5.020e-02  4.475e-02    1.122  0.262039
## Port_modem_Yes   -1.379e-02  3.328e-02   -0.414  0.678719
## Tablet_Yes       3.820e-02  3.649e-02    1.047  0.295304
```

```

## InternetService_Fiber_Optic      5.672e+00  5.433e-02  104.413 < 2e-16
## InternetService_None              4.650e+00  5.296e-02   87.813 < 2e-16
## Phone_Yes                        4.525e-03  5.646e-02   0.080 0.936117
## Multiple_Yes                     2.478e-01  6.977e-02   3.551 0.000387
## OnlineSecurity_Yes               -8.987e-01  3.513e-02 -25.579 < 2e-16
## OnlineBackup_Yes                 -3.897e-01  5.440e-02  -7.164 8.86e-13
## DeviceProtection_Yes             -6.409e-01  4.104e-02 -15.618 < 2e-16
## TechSupport_Yes                  3.511e-01  4.151e-02   8.459 < 2e-16
## StreamingTV_Yes                  -1.374e+00  8.690e-02 -15.814 < 2e-16
## StreamingMovies_Yes              -8.148e-01  1.042e-01  -7.822 6.19e-15
## PaperlessBilling_Yes              7.924e-04  3.375e-02   0.023 0.981268
## `PaymentMethod_Credit_Card_(automatic)` -6.822e-03  5.082e-02  -0.134 0.893224
## PaymentMethod_Electronic_Check   -2.087e-03  4.540e-02  -0.046 0.963333
## PaymentMethod_Mailed_Check       -8.873e-03  4.968e-02  -0.179 0.858241
##
## (Intercept)                      ***
## Outage_sec_perweek
## Email
## Contacts                          .
## Yearly_equip_failure
## MonthlyCharge                     ***
## Bandwidth_GB_Year                 ***
## Item1
## Item2
## Item3
## Item4                             *
## Item5
## Item6
## Item7
## Item8
## Churn_Yes                         **
## Techie_Yes
## Port_modem_Yes
## Tablet_Yes
## InternetService_Fiber_Optic      ***
## InternetService_None              ***
## Phone_Yes
## Multiple_Yes                      ***
## OnlineSecurity_Yes                ***
## OnlineBackup_Yes                  ***
## DeviceProtection_Yes              ***
## TechSupport_Yes                   ***
## StreamingTV_Yes                   ***
## StreamingMovies_Yes               ***
## PaperlessBilling_Yes
## `PaymentMethod_Credit_Card_(automatic)`
## PaymentMethod_Electronic_Check
## PaymentMethod_Mailed_Check
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.225 on 5423 degrees of freedom
## Multiple R-squared:  0.9979, Adjusted R-squared:  0.9979
## F-statistic: 7.921e+04 on 32 and 5423 DF,  p-value: < 2.2e-16

```

```
# Create the gross/initial model w/o Multicollinearity
MLR_GrossModel <- lm(Tenure ~ ., churn_nomult)
summary(MLR_GrossModel)
```

```
##
## Call:
## lm(formula = Tenure ~ ., data = churn_nomult)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -52.26 -12.15   1.73  14.68  59.51
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      28.09944     4.04932   6.939 4.40e-12
## Outage_sec_perweek    0.01780     0.09082   0.196 0.844654
## Email             -0.14171     0.08976  -1.579 0.114446
## Contacts           0.25496     0.27429   0.929 0.352673
## Yearly_equip_failure  0.50857     0.42819   1.188 0.234994
## MonthlyCharge       0.12073     0.03105   3.889 0.000102
## Item1             -0.25434     0.38910  -0.654 0.513352
## Item2              0.06211     0.36386   0.171 0.864458
## Item3              0.20971     0.33183   0.632 0.527425
## Item4             -0.15721     0.30072  -0.523 0.601161
## Item5              0.06278     0.31254   0.201 0.840820
## Item6             -0.24804     0.32137  -0.772 0.440257
## Item7              0.37956     0.30258   1.254 0.209751
## Item8             -0.16990     0.28559  -0.595 0.551929
## Churn_Yes         -39.71274     0.63047 -62.989 < 2e-16
## Techie_Yes         4.02859     0.73178   5.505 3.86e-08
## Port_modem_Yes     1.12960     0.54551   2.071 0.038430
## Tablet_Yes        -1.07429     0.59814  -1.796 0.072543
## InternetService_Fiber_Optic -7.95298     0.87122  -9.129 < 2e-16
## InternetService_None -4.60237     0.85908  -5.357 8.79e-08
## Phone_Yes         -1.11803     0.92555  -1.208 0.227114
## Multiple_Yes       1.52815     1.14389   1.336 0.181631
## OnlineSecurity_Yes -0.27281     0.57598  -0.474 0.635773
## OnlineBackup_Yes    0.67800     0.89181   0.760 0.447135
## DeviceProtection_Yes -0.17177     0.67283  -0.255 0.798505
## TechSupport_Yes    -1.25016     0.68025  -1.838 0.066149
## StreamingTV_Yes     4.45133     1.42260   3.129 0.001763
## StreamingMovies_Yes  4.82578     1.70616   2.828 0.004695
## PaperlessBilling_Yes  0.01922     0.55331   0.035 0.972287
## `PaymentMethod_Credit_Card_(automatic)` 1.17385     0.83315   1.409 0.158912
## PaymentMethod_Electronic_Check 2.15344     0.74380   2.895 0.003804
## PaymentMethod_Mailed_Check 1.59492     0.81421   1.959 0.050182
##
## (Intercept)      ***
## Outage_sec_perweek
## Email
## Contacts
## Yearly_equip_failure
## MonthlyCharge      ***
## Item1
```



```

## Item2
## Item3
## Item4
## Item5
## Item6
## Item7
## Item8
## Churn_Yes ***
## Techie_Yes ***
## Port_modem_Yes *
## Tablet_Yes .
## InternetService_Fiber_Optic ***
## InternetService_None ***
## Phone_Yes
## Multiple_Yes
## OnlineSecurity_Yes
## OnlineBackup_Yes
## DeviceProtection_Yes
## TechSupport_Yes .
## StreamingTV_Yes **
## StreamingMovies_Yes **
## PaperlessBilling_Yes
## `PaymentMethod_Credit_Card_(automatic)`
## PaymentMethod_Electronic_Check **
## PaymentMethod_Mailed_Check .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.09 on 5424 degrees of freedom
## Multiple R-squared:  0.426, Adjusted R-squared:  0.4227
## F-statistic: 129.8 on 31 and 5424 DF, p-value: < 2.2e-16
# Use Regression Subsets to find the top 3 variables which impact Tenure
MLR_subsets <- regsubsets(Tenure ~ ., churn_nomult, nvmax = 3)
summary(MLR_subsets)

## Subset selection object
## Call: regsubsets.formula(Tenure ~ ., churn_nomult, nvmax = 3)
## 31 Variables (and intercept)
##
## Forced in Forced out
## Outage_sec_perweek FALSE FALSE
## Email FALSE FALSE
## Contacts FALSE FALSE
## Yearly_equip_failure FALSE FALSE
## MonthlyCharge FALSE FALSE
## Item1 FALSE FALSE
## Item2 FALSE FALSE
## Item3 FALSE FALSE
## Item4 FALSE FALSE
## Item5 FALSE FALSE
## Item6 FALSE FALSE
## Item7 FALSE FALSE
## Item8 FALSE FALSE
## Churn_Yes FALSE FALSE
## Techie_Yes FALSE FALSE

```

```

## Port_modem_Yes                FALSE    FALSE
## Tablet_Yes                    FALSE    FALSE
## InternetService_Fiber_Optic  FALSE    FALSE
## InternetService_None         FALSE    FALSE
## Phone_Yes                     FALSE    FALSE
## Multiple_Yes                  FALSE    FALSE
## OnlineSecurity_Yes           FALSE    FALSE
## OnlineBackup_Yes             FALSE    FALSE
## DeviceProtection_Yes         FALSE    FALSE
## TechSupport_Yes              FALSE    FALSE
## StreamingTV_Yes              FALSE    FALSE
## StreamingMovies_Yes          FALSE    FALSE
## PaperlessBilling_Yes         FALSE    FALSE
## `PaymentMethod_Credit_Card_(automatic)` FALSE    FALSE
## PaymentMethod_Electronic_Check FALSE    FALSE
## PaymentMethod_Mailed_Check   FALSE    FALSE
## 1 subsets of each size up to 3
## Selection Algorithm: exhaustive
##      Outage_sec_perweek Email Contacts Yearly equip_failure MonthlyCharge
## 1 ( 1 ) " "                " "      " "                " "
## 2 ( 1 ) " "                " "      " "                "*"
## 3 ( 1 ) " "                " "      " "                "*"
##      Item1 Item2 Item3 Item4 Item5 Item6 Item7 Item8 Churn_Yes Techie_Yes
## 1 ( 1 ) " "      " "      " "      " "      " "      " "      " "      "*"      " "
## 2 ( 1 ) " "      " "      " "      " "      " "      " "      " "      "*"      " "
## 3 ( 1 ) " "      " "      " "      " "      " "      " "      " "      "*"      " "
##      Port_modem_Yes Tablet_Yes InternetService_Fiber_Optic
## 1 ( 1 ) " "                " "                " "
## 2 ( 1 ) " "                " "                " "
## 3 ( 1 ) " "                " "                "*"
##      InternetService_None Phone_Yes Multiple_Yes OnlineSecurity_Yes
## 1 ( 1 ) " "                " "                " "                " "
## 2 ( 1 ) " "                " "                " "                " "
## 3 ( 1 ) " "                " "                " "                " "
##      OnlineBackup_Yes DeviceProtection_Yes TechSupport_Yes StreamingTV_Yes
## 1 ( 1 ) " "                " "                " "                " "
## 2 ( 1 ) " "                " "                " "                " "
## 3 ( 1 ) " "                " "                " "                " "
##      StreamingMovies_Yes PaperlessBilling_Yes
## 1 ( 1 ) " "                " "
## 2 ( 1 ) " "                " "
## 3 ( 1 ) " "                " "
##      `PaymentMethod_Credit_Card_(automatic)` PaymentMethod_Electronic_Check
## 1 ( 1 ) " "                " "
## 2 ( 1 ) " "                " "
## 3 ( 1 ) " "                " "
##      PaymentMethod_Mailed_Check
## 1 ( 1 ) " "
## 2 ( 1 ) " "
## 3 ( 1 ) " "

```

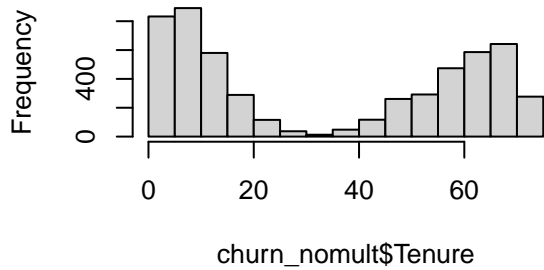
```

# Create Univariate visualizations of top 3 independent variables & dependent variable
par(mfrow = c(2,2))
Tenure_hist <- hist(churn_nomult$Tenure)

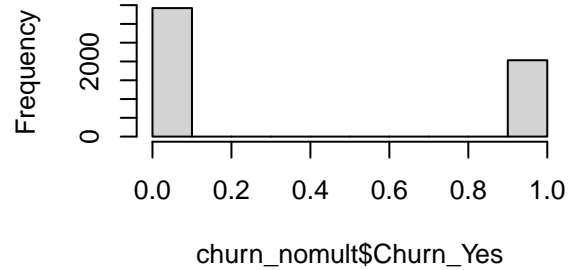
```

```
ChurnYes_hist <- hist(churn_nomult$Churn_Yes)
MonthlyCharge_hist <- hist(churn_nomult$MonthlyCharge)
InternetServiceFiberOptic_hist <- hist(churn_nomult$InternetService_Fiber_Optic)
```

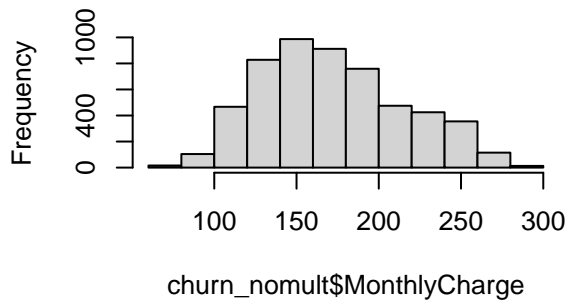
Histogram of churn_nomult\$Tenure



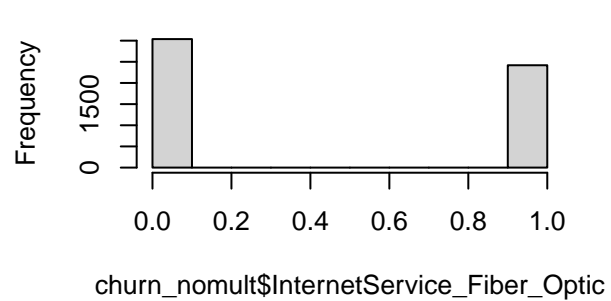
Histogram of churn_nomult\$Churn_Yes



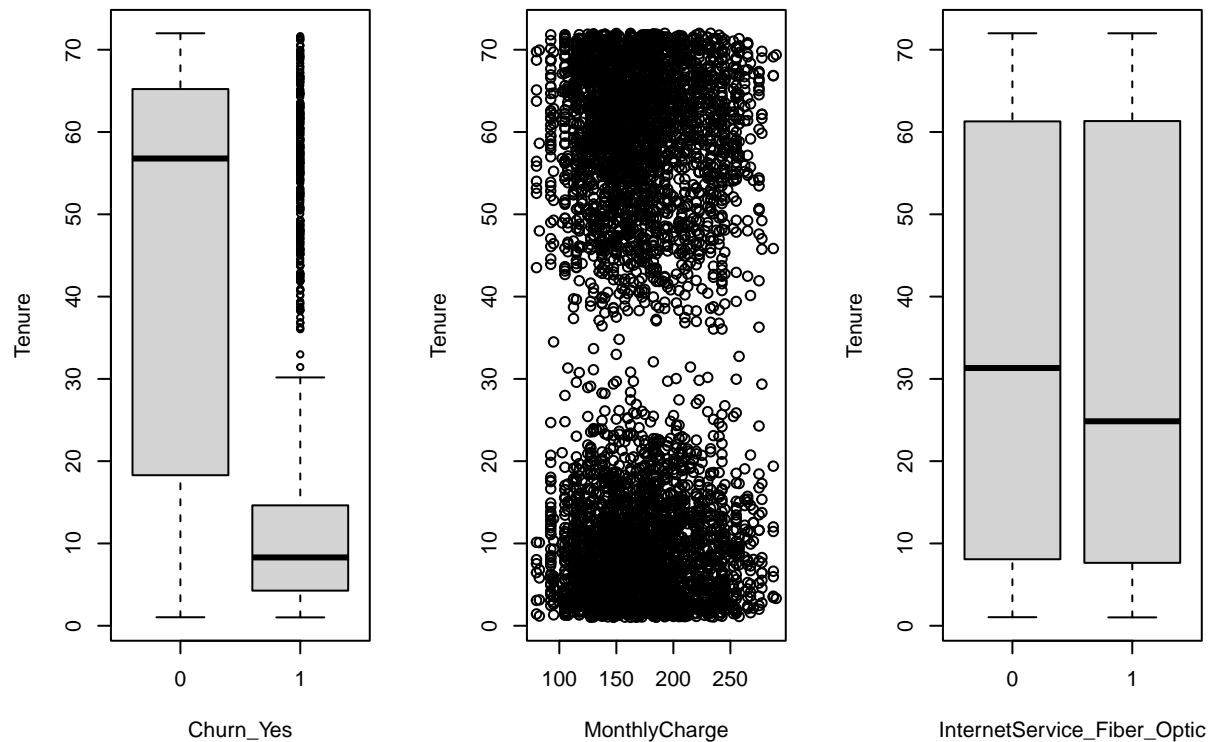
Histogram of churn_nomult\$MonthlyCharge



Histogram of churn_nomult\$InternetService_Fiber_Optic



```
# Create Bivariate visualizations of top 3 independent variables
par(mfrow=c(1,3))
ChurnYes_boxplot <- boxplot(Tenure ~ Churn_Yes, data = churn_nomult)
MonthlyCharge_boxplot <- plot(Tenure ~ MonthlyCharge, data = churn_nomult)
InternetServiceFiberOptic_boxplot <- boxplot(Tenure ~ InternetService_Fiber_Optic, data = churn_nomult)
```



Model Building

Create the adjusted/reduced model

```
MLR_AdjustedModel <- lm(Tenure ~ Churn_Yes + MonthlyCharge + InternetService_Fiber_Optic, churn_nomult)
summary(MLR_AdjustedModel)
```

```
##
## Call:
## lm(formula = Tenure ~ Churn_Yes + MonthlyCharge + InternetService_Fiber_Optic,
##     data = churn_nomult)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -53.121 -12.048   1.679  15.085  63.755
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    18.943622   1.182123   16.02  <2e-16 ***
## Churn_Yes      -38.918326   0.627638  -62.01  <2e-16 ***
## MonthlyCharge    0.193749   0.007494   25.85  <2e-16 ***
## InternetService_Fiber_Optic -7.998853   0.589442  -13.57  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.25 on 5452 degrees of freedom
## Multiple R-squared:  0.4137, Adjusted R-squared:  0.4134
```

```
## F-statistic: 1282 on 3 and 5452 DF, p-value: < 2.2e-16
# Identify the coefficients of the Adjusted Model
coef(MLR_AdjustedModel)

##              (Intercept)              Churn_Yes
##          18.9436216          -38.9183258
##      MonthlyCharge InternetService_Fiber_Optic
##          0.1937494          -7.9988533

# Create the reduced model with only 2 variables (just for testing)
MLR_Reduced2Model <- lm(Tenure ~ Churn_Yes + MonthlyCharge, churn_nomult)
summary(MLR_Reduced2Model)

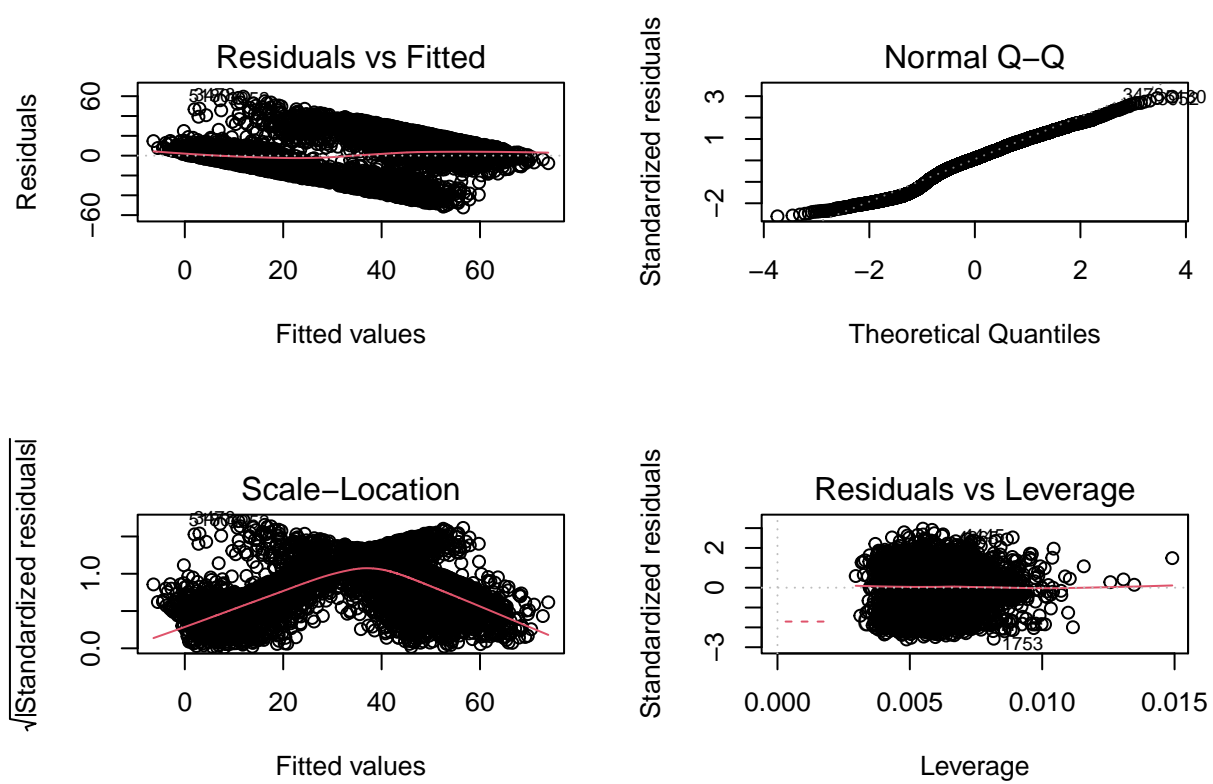
##
## Call:
## lm(formula = Tenure ~ Churn_Yes + MonthlyCharge, data = churn_nomult)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -54.98 -11.78   1.59  15.44  58.22
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   20.808672   1.193661   17.43  <2e-16 ***
## Churn_Yes     -37.180971   0.624675  -59.52  <2e-16 ***
## MonthlyCharge   0.158551   0.007148   22.18  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.58 on 5453 degrees of freedom
## Multiple R-squared:  0.3939, Adjusted R-squared:  0.3936
## F-statistic: 1772 on 2 and 5453 DF, p-value: < 2.2e-16
```

Save and Load Model

```
# Save and Load Model
model_url <- "C:/Users/tedda/Desktop/Data Science Portfolio/Machine Learning/Supervised Learning/Regression"
saveRDS(MLR_Reduced2Model, model_url)
MLR_model <- readRDS(model_url)
```

View residual plots of Trained Model

```
# Residual Plots
par(mfrow = c(2,2))
plot(MLR_GrossModel)
```



```
plot(MLR_model)
```

