Internet-Firewall-Data-Analysis.R

rstudio-user

2021-01-25

```
# Internet Firewall Data Analysis
# Mia Abrams

#Load Libraries
library(readr)
library (ggplot2)
install.packages("corrplot")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
## (as 'lib' is unspecified)
```

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
#Upload Firewall data logging file
log2 <- read_csv("log2.csv")
```

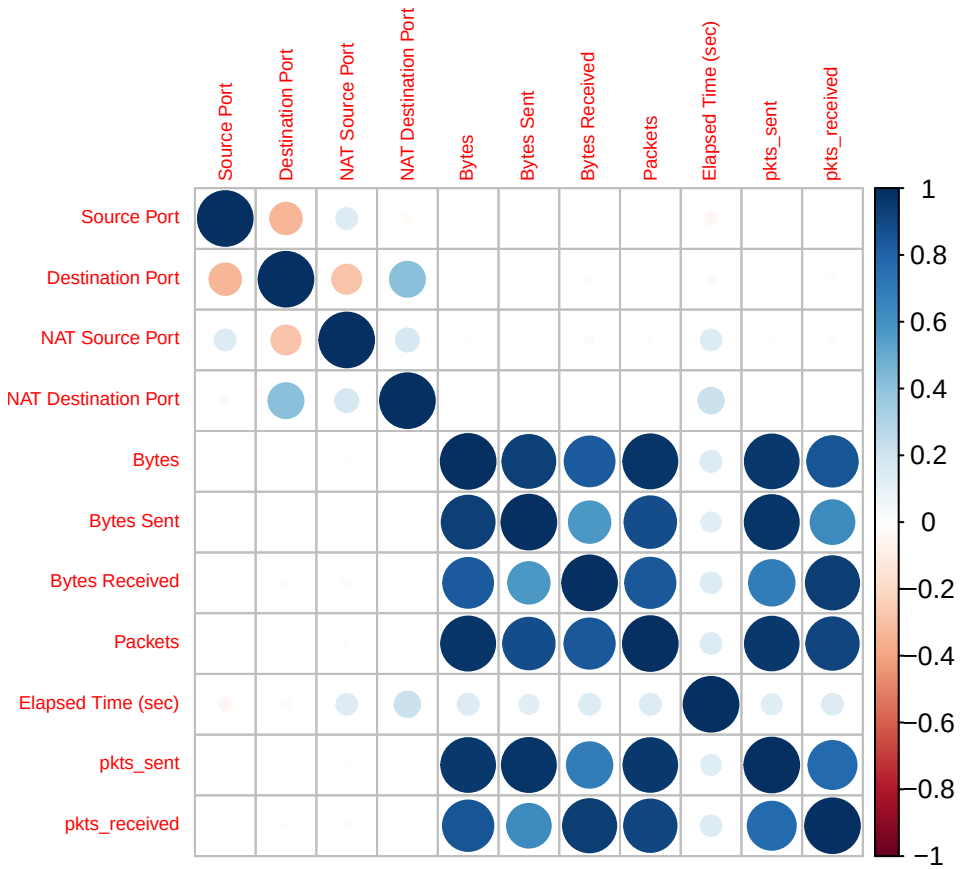
```
##
## -- Column specification --------------------------------------------------------
## cols(
##   `Source Port` = col_double(),
##   `Destination Port` = col_double(),
##   `NAT Source Port` = col_double(),
##   `NAT Destination Port` = col_double(),
##   Action = col_character(),
##   Bytes = col_double(),
##   `Bytes Sent` = col_double(),
##   `Bytes Received` = col_double(),
##   Packets = col_double(),
##   `Elapsed Time (sec)` = col_double(),
##   pkts_sent = col_double(),
##   pkts_received = col_double()
## )
```

```
summary(log2)
```

```
##   Source Port    Destination Port NAT Source Port NAT Destination Port
##  Min.   :    0   Min.   :    0    Min.   :    0   Min.   :    0
##  1st Qu.:49183   1st Qu.:   80    1st Qu.:    0   1st Qu.:    0
##  Median :53776   Median :  445    Median : 8820   Median :   53
##  Mean   :49392   Mean   :10577    Mean   :19283   Mean   : 2671
##  3rd Qu.:58638   3rd Qu.:15000    3rd Qu.:38366   3rd Qu.:  443
##  Max.   :65534   Max.   :65535    Max.   :65535   Max.   :65535
##     Action              Bytes             Bytes Sent        Bytes Received
```

```
##  Length:65532       Min.   :6.000e+01   Min.   :       60   Min.   :         0
##  Class :character   1st Qu.:6.600e+01   1st Qu.:       66   1st Qu.:         0
##  Mode  :character   Median :1.680e+02   Median :       90   Median :        79
##                     Mean   :9.712e+04   Mean   :    22386   Mean   :     74738
##                     3rd Qu.:7.520e+02   3rd Qu.:      210   3rd Qu.:       449
##                     Max.   :1.269e+09   Max.   :948477220   Max.   :320881795
##     Packets          Elapsed Time (sec)   pkts_sent         pkts_received
##  Min.   :       1.0   Min.   :    0.00   Min.   :     1.0   Min.   :     0.0
##  1st Qu.:       1.0   1st Qu.:    0.00   1st Qu.:     1.0   1st Qu.:     0.0
##  Median :       2.0   Median :   15.00   Median :     1.0   Median :     1.0
##  Mean   :     102.9   Mean   :   65.83   Mean   :    41.4   Mean   :    61.5
##  3rd Qu.:       6.0   3rd Qu.:   30.00   3rd Qu.:     3.0   3rd Qu.:     2.0
##  Max.   :1036116.0   Max.   :10824.00   Max.   :747520.0   Max.   :327208.0
```

```
FirewallData <- log2[-(1),-(5) ]
#cor(FirewallData$`Source Port`,FirewallData$`Destination Port`)
#install.packages("ggplot2")
#library (ggplot2)
#ggplot(data=log2)
corrplot(cor(FirewallData),tl.cex = 0.6)
```




```
cor(FirewallData$`Bytes Received`, FirewallData$Bytes)
```

```
## [1] 0.830225
```

```
# The number of bytes received is dependent on the action classification/type
# The NAT Source and Destination Port can predict where bytes are received by
# by the recipient
```

```
# Based on the corrplot, the bytes received has a moderately strong positive
# correlation of 0.830225 with the transmitting message
install.packages("dplyr")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
## (as 'lib' is unspecified)
```

```
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
```

```
Actions <- pull(log2,Action)
install.packages("rpart")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
## (as 'lib' is unspecified)
```

```
install.packages("rpart.plot")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
## (as 'lib' is unspecified)
```

```
install.packages("rattle")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
## (as 'lib' is unspecified)
```

```
install.packages("RColorBrewer")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
## (as 'lib' is unspecified)
```

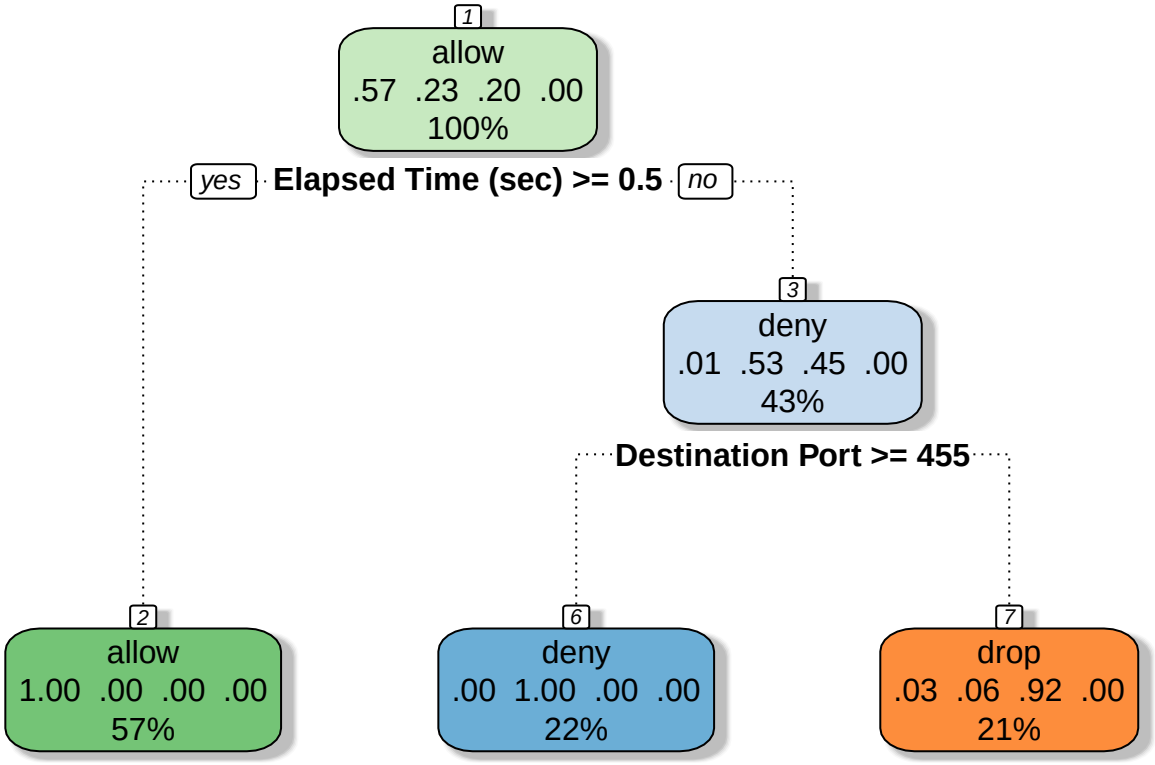
```
library(rpart)
library(rattle)
```

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

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

```
## Rattle: A free graphical interface for data science with R.
## Version 5.4.0 Copyright (c) 2006-2020 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.
```

```
library(rpart.plot)
library(RColorBrewer)
decisionTreeBinary <- rpart(Action ~ . , data = log2, cp=0.1)
fancyRpartPlot(decisionTreeBinary)
```




```
install.packages("tidyverse") # a set of data science tools including dplyr, tidyr and stringr
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
## (as 'lib' is unspecified)
```

```
install.packages("skimr") # a package to facilitate data summaries
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
## (as 'lib' is unspecified)
```

```
install.packages("Hmisc") # a package for data analysis
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
## (as 'lib' is unspecified)
```

```
# Load Libraries
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v tidyr   1.1.2     v stringr 1.4.0
## v purrr   0.3.4     v forcats 0.5.0
```

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

```
library(skimr)
library(Hmisc)
```

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

```
## Loading required package: survival
```

```
## Loading required package: Formula
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:dplyr':
##
##     src, summarize
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
library(ggplot2)
skim(Actions)
```

Table 1: Data summary

Name	Actions
Number of rows	65532
Number of columns	1
Column type frequency:	
character	1
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
data	0	1	4	10	0	4	0

```
dropped<- filter(log2, log2$Action=="drop")
#install.packages("stringr")
#library(stringr)
#filter(Actions,"drop")
#filter(log2,log2$`Bytes Received`== 0)
```