Module Two Exercise

msurdek

10/18/2021

## load packages

library(readr)  
library(dplyr)

## read claims file

data <- read\_csv("/home/msurdek/Documents/School/DAT 610 Optimization & Risk Assessment/Assignments/Module Two Exercise/DAT 610 Auto Accident Personal Injury Claims-1.csv")

## remove empty rows and print dataframe

data <- data[!is.na(data$Claim\_Number),]  
  
data

## # A tibble: 502 × 45  
## Claim\_Number Policy\_ID CLAIM\_AMOUNT PAID\_AMOUNT CLAIM\_SUSPICION\_SCORE IND\_01  
## <dbl> <dbl> <chr> <chr> <dbl> <dbl>  
## 1 5001463 364697 $13,463 $3,646 3 1  
## 2 5004844 426960 $1,246 $594 3 1  
## 3 5005493 426313 $19,883 $15,138 3 1  
## 4 5007366 351603 $16,348 $28 3 1  
## 5 5011314 423014 $2,477 $786 2 5  
## 6 5016984 419258 $37,365 $35,228 3 1  
## 7 5021876 415367 $18,926 $4,328 3 1  
## 8 5023456 365027 $12,990 $0 3 2  
## 9 5024273 346972 $29,493 $0 3 2  
## 10 5029392 351192 $5,255 $1,816 2 2  
## # … with 492 more rows, and 39 more variables: IND\_02 <dbl>, IND\_03 <dbl>,  
## # IND\_04 <dbl>, IND\_05 <dbl>, IND\_06 <dbl>, IND\_07 <dbl>, IND\_08 <dbl>,  
## # IND\_09 <dbl>, IND\_10 <dbl>, IND\_11 <dbl>, IND\_12 <dbl>, IND\_13 <dbl>,  
## # IND\_14 <dbl>, IND\_15 <dbl>, IND\_16 <dbl>, IND\_17 <dbl>, IND\_18 <dbl>,  
## # IND\_19 <dbl>, IND\_20 <dbl>, RIDIT\_01 <dbl>, RIDIT\_02 <dbl>, RIDIT\_03 <dbl>,  
## # RIDIT\_04 <dbl>, RIDIT\_05 <dbl>, RIDIT\_06 <dbl>, RIDIT\_07 <dbl>,  
## # RIDIT\_08 <dbl>, RIDIT\_09 <dbl>, RIDIT\_10 <dbl>, RIDIT\_11 <dbl>, …