Capstone Project - Week1

Amol Gote

11/7/2020

set.seed(123)  
library(tidyverse)

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

## v ggplot2 3.3.2 v purrr 0.3.4  
## v tibble 3.0.1 v dplyr 1.0.0  
## v tidyr 1.1.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

## -- Conflicts -------------------------------------------------------------------------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(caret)

## Loading required package: lattice

##   
## Attaching package: 'caret'

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

library(corrplot)

## corrplot 0.84 loaded

fundamentals\_ds <- read.csv("./data/Fundamentals\_DS.csv", na.strings=c(""," "))  
nrow(fundamentals\_ds)

## [1] 17416

names(fundamentals\_ds)[names(fundamentals\_ds) == "ï..gvkey"] <- "gvkey"  
fundamentals\_ds\_filter <- fundamentals\_ds %>%  
 filter(datafmt == 'STD')   
nrow(fundamentals\_ds\_filter)

## [1] 9078

fundamentals\_ds\_filter\_1 <- fundamentals\_ds\_filter[ lapply( fundamentals\_ds\_filter, function(x) sum(is.na(x)) / length(x) ) < 0.1 ]  
ncol(fundamentals\_ds\_filter\_1)

## [1] 269

write.csv(fundamentals\_ds\_filter\_1, file = "data/fundamentals\_ds\_filter\_1.csv", row.names=FALSE)

fundamentals\_ds\_filter\_1 <- fundamentals\_ds\_filter\_1[ lapply( fundamentals\_ds\_filter\_1, function(x) sum(is.na(x)) / length(x) ) < 0.1 ]  
fundamentals\_ds\_filter\_1 <- subset(fundamentals\_ds\_filter\_1, select = -c(datadate,indfmt,curncd,consol,popsrc,conm,curcd,apdedate,fdate,add1,addzip,busdesc,  
 city,conml,weburl,phone,loc,final,fyr,acchg,fic,  
 xido,xidoc,naicsh,sich,au,auop,auopic,fyrc,ggroup,gind,gsector,gsubind,priusa))  
write.csv(fundamentals\_ds\_filter\_1, file = "data/fundamentals\_ds\_filter\_2.csv", row.names=FALSE, na="")

nzv\_ds <- nearZeroVar(fundamentals\_ds\_filter\_1, saveMetrics = TRUE)  
nzv\_ds <- nzv\_ds[nzv\_ds[,"nzv"] > 0, ]  
nzv\_ds

## freqRatio percentUnique zeroVar nzv  
## datafmt 0.00000 0.01101564 TRUE TRUE  
## ajex 66.32000 1.49812734 FALSE TRUE  
## ajp 68.54545 1.46508041 FALSE TRUE  
## currtr 244.14286 1.31086142 FALSE TRUE  
## pddur 753.08333 0.12117206 FALSE TRUE  
## scf 0.00000 0.01101564 TRUE TRUE  
## upd 55.38509 0.02203128 FALSE TRUE  
## acdo 2115.25000 2.90812954 FALSE TRUE  
## aldo 2159.25000 1.17867372 FALSE TRUE  
## aociother 1630.20000 4.71469487 FALSE TRUE  
## aocisecgl 0.00000 0.01101564 TRUE TRUE  
## ciother 747.90909 4.26305354 FALSE TRUE  
## cstke 1366.00000 5.12227363 FALSE TRUE  
## dcom 4315.50000 0.88125138 FALSE TRUE  
## dcvsr 1322.50000 6.59836968 FALSE TRUE  
## dcvsub 1679.80000 1.81758096 FALSE TRUE  
## dcvt 1295.00000 8.03040317 FALSE TRUE  
## dd 882.00000 5.39766468 FALSE TRUE  
## diladj 905.11111 3.58008372 FALSE TRUE  
## ds 479.58824 3.34875523 FALSE TRUE  
## dudd 2782.66667 2.62172285 FALSE TRUE  
## dvp 678.25000 5.28750826 FALSE TRUE  
## dvpa 2115.00000 1.15664243 FALSE TRUE  
## esopct 4325.00000 0.73804803 FALSE TRUE  
## esopnr 4375.50000 0.08812514 FALSE TRUE  
## esopr 0.00000 0.01101564 TRUE TRUE  
## esopt 4375.50000 0.08812514 FALSE TRUE  
## intc 256.60714 7.42454285 FALSE TRUE  
## itcb 0.00000 0.01101564 TRUE TRUE  
## lifr 1574.60000 4.56047588 FALSE TRUE  
## mib 1201.57143 3.50297422 FALSE TRUE  
## pncad 36.83684 1.37695528 FALSE TRUE  
## pncaeps 36.81053 1.39898656 FALSE TRUE  
## pnrsho 224.11429 4.56047588 FALSE TRUE  
## prca 513.40000 5.82727473 FALSE TRUE  
## prcad 60.77519 0.58382904 FALSE TRUE  
## prcaeps 59.81679 0.59484468 FALSE TRUE  
## prsho 741.09091 4.14188147 FALSE TRUE  
## pstk 96.53846 8.51509143 FALSE TRUE  
## pstkc 190.07317 5.97047808 FALSE TRUE  
## pstkl 161.02174 8.97774840 FALSE TRUE  
## pstkn 100.98750 3.12844239 FALSE TRUE  
## pstkr 1629.80000 5.84930601 FALSE TRUE  
## pstkrv 161.02174 9.10993611 FALSE TRUE  
## rdip 2863.33333 1.41000220 FALSE TRUE  
## rdipa 2849.33333 1.52015863 FALSE TRUE  
## rdipd 438.21053 0.55078211 FALSE TRUE  
## rdipeps 462.55556 0.53976647 FALSE TRUE  
## seqo 1576.60000 7.83212161 FALSE TRUE  
## tstkp 2149.50000 0.02203128 FALSE TRUE  
## txo 2639.00000 2.97422340 FALSE TRUE  
## txw 2805.00000 1.81758096 FALSE TRUE  
## xi 8728.00000 0.12117206 FALSE TRUE  
## xintopt 8207.00000 0.02203128 FALSE TRUE  
## xoptd 0.00000 0.01101564 TRUE TRUE  
## xopteps 0.00000 0.01101564 TRUE TRUE  
## dvpsp\_c 112.85455 7.98634060 FALSE TRUE  
## dvpsx\_c 114.74074 7.85415290 FALSE TRUE  
## adjex\_c 62.96639 1.38797092 FALSE TRUE  
## dvpsp\_f 119.65385 7.77704340 FALSE TRUE  
## dvpsx\_f 110.98214 7.62282441 FALSE TRUE  
## adjex\_f 62.34167 1.41000220 FALSE TRUE  
## rank 0.00000 0.01101564 TRUE TRUE  
## dpact\_fn 31.85081 0.05507821 FALSE TRUE  
## rdipa\_fn 1711.20000 0.02203128 FALSE TRUE  
## rdipd\_fn 133.77419 0.02203128 FALSE TRUE  
## rdipeps\_fn 101.13415 0.02203128 FALSE TRUE  
## stkco\_fn 0.00000 0.01101564 TRUE TRUE

nzv\_ds\_cols <- nearZeroVar(fundamentals\_ds\_filter\_1)  
fundamentals\_ds\_filter\_1 <- fundamentals\_ds\_filter\_1[, -nzv\_ds\_cols]  
write.csv(fundamentals\_ds\_filter\_1, file = "data/fundamentals\_ds\_filter\_3.csv", row.names=FALSE, na="")

ncol(fundamentals\_ds\_filter\_1)

## [1] 167

fundamentals\_ds\_filter\_1 <- subset(fundamentals\_ds\_filter\_1, select = -c(acctstd,ismod,src,acodo,acox,  
 aox,capxv,ceql,cibegni,cicurr,cidergl,  
 cimii,cipen,cisecgl,citotal,cshfd,cshpri,dclo,dcpstk,  
 dltis,dlto,dltr,do,dp,dpvieb,drc,  
 drlt,dv,dvc,epsfx,epspx,exre,  
 fopox,ibadj,ibc,ibcom,ibmii,  
 invch,ivaco,ivaeq,ivao,lco,  
 lcox,lcoxdr,lct,loxdr,mibn,  
 mibt,mii,msa,niadj,np,oprepsx,pnca,ppent,ppeveb,  
 recco,rectr,sale,spced,spceeps,  
 tstkc,txbco,txbcof,txdb,  
 txdba,txdbca,txdc,txdi,  
 txditc,txndb,xopr,exchg,costat,  
 ceoso,cfoso,idbflag,naics,sic,stko))  
write.csv(fundamentals\_ds\_filter\_1, file = "data/fundamentals\_ds\_filter\_4.csv", row.names=FALSE, na="")

ncol(fundamentals\_ds\_filter\_1)

## [1] 87

#fundamentals\_ds\_filter\_1

fundamentals\_restmt\_ds\_filter <- fundamentals\_ds %>%  
 filter(datafmt == 'SUMM\_STD')  
std\_cols <- colnames(fundamentals\_ds\_filter\_1)  
fundamentals\_restmt\_ds\_filter <- subset(fundamentals\_restmt\_ds\_filter, select = c(std\_cols))  
fundamentals\_restmt\_ds\_filter <- fundamentals\_restmt\_ds\_filter[ lapply( fundamentals\_restmt\_ds\_filter, function(x) sum(is.na(x)) / length(x) ) < 0.1 ]  
summary(fundamentals\_restmt\_ds\_filter$ci)

## Length Class Mode   
## 0 NULL NULL

sample\_restmt\_ds\_filter <- fundamentals\_restmt\_ds\_filter %>%  
 filter(gvkey == 1076)  
sample\_ds\_filter <- fundamentals\_ds\_filter\_1 %>%  
 filter(gvkey == 1076)  
#nrow(sample\_restmt\_ds\_filter)  
#nrow(sample\_ds\_filter)  
#head(sample\_restmt\_ds\_filter)  
#head(sample\_ds\_filter)  
  
fundamentals\_ds\_filter\_1$restmt\_at <- 0  
fundamentals\_ds\_filter\_1$restmt\_at\_mag <- 0.0  
  
fundamentals\_ds\_filter\_1$restmt\_capx <- 0  
fundamentals\_ds\_filter\_1$restmt\_capx\_mag <- 0.0  
  
fundamentals\_ds\_filter\_1$restmt\_cogs <- 0  
fundamentals\_ds\_filter\_1$restmt\_cogs\_mag <- 0.0  
  
fundamentals\_ds\_filter\_1$restmt\_dltt <- 0  
fundamentals\_ds\_filter\_1$restmt\_dltt\_mag <- 0.0  
  
fundamentals\_ds\_filter\_1$restmt\_epsfi <- 0  
fundamentals\_ds\_filter\_1$restmt\_epsfi\_mag <- 0.0  
  
fundamentals\_ds\_filter\_1$restmt\_epspi <- 0  
fundamentals\_ds\_filter\_1$restmt\_epspi\_mag <- 0.0  
  
fundamentals\_ds\_filter\_1$restmt\_ib <- 0  
fundamentals\_ds\_filter\_1$restmt\_ib\_mag <- 0.0  
  
fundamentals\_ds\_filter\_1$restmt\_ni <- 0  
fundamentals\_ds\_filter\_1$restmt\_ni\_mag <- 0.0  
  
fundamentals\_ds\_filter\_1$restmt\_nopi <- 0  
fundamentals\_ds\_filter\_1$restmt\_nopi\_mag <- 0.0  
  
fundamentals\_ds\_filter\_1$restmt\_pi <- 0  
fundamentals\_ds\_filter\_1$restmt\_pi\_mag <- 0.0  
  
fundamentals\_ds\_filter\_1$restmt\_reuna <- 0  
fundamentals\_ds\_filter\_1$restmt\_reuna\_mag <- 0.0  
  
fundamentals\_ds\_filter\_1$restmt\_seq <- 0  
fundamentals\_ds\_filter\_1$restmt\_seq\_mag <- 0.0  
  
fundamentals\_ds\_filter\_1$restmt\_teq <- 0  
fundamentals\_ds\_filter\_1$restmt\_teq\_mag <- 0.0  
  
fundamentals\_ds\_filter\_1$restmt\_txt <- 0  
fundamentals\_ds\_filter\_1$restmt\_txt\_mag <- 0.0  
  
fundamentals\_ds\_filter\_1$restmt\_wcap <- 0  
fundamentals\_ds\_filter\_1$restmt\_wcap\_mag <- 0.0  
  
fundamentals\_ds\_filter\_1$restmt\_ci <- 0  
fundamentals\_ds\_filter\_1$restmt\_ci\_mag <- 0.0  
  
for (row in 1:nrow(sample\_restmt\_ds\_filter)){  
 restmt\_item\_gvkey <- as.integer(sample\_restmt\_ds\_filter[row, "gvkey"])  
 restmt\_item\_fyear <- sample\_restmt\_ds\_filter[row, "fyear"]  
 restmt\_item\_at <- sample\_restmt\_ds\_filter[row, "at"]  
 restmt\_item\_capx <- sample\_restmt\_ds\_filter[row, "capx"]  
 restmt\_item\_cogs <- sample\_restmt\_ds\_filter[row, "cogs"]  
 restmt\_item\_dltt <- sample\_restmt\_ds\_filter[row, "dltt"]  
 restmt\_item\_epsfi <- sample\_restmt\_ds\_filter[row, "epsfi"]  
 restmt\_item\_epspi <- sample\_restmt\_ds\_filter[row, "epspi"]  
   
 restmt\_item\_ib <- sample\_restmt\_ds\_filter[row, "ib"]  
 restmt\_item\_ni <- sample\_restmt\_ds\_filter[row, "ni"]  
 restmt\_item\_nopi <- sample\_restmt\_ds\_filter[row, "nopi"]  
 restmt\_item\_pi <- sample\_restmt\_ds\_filter[row, "pi"]  
 restmt\_item\_reuna <- sample\_restmt\_ds\_filter[row, "reuna"]  
 restmt\_item\_seq <- sample\_restmt\_ds\_filter[row, "seq"]  
 restmt\_item\_teq <- sample\_restmt\_ds\_filter[row, "teq"]  
 restmt\_item\_txt <- sample\_restmt\_ds\_filter[row, "txt"]  
 restmt\_item\_wcap <- sample\_restmt\_ds\_filter[row, "wcap"]  
  
 row\_count <- as.integer(nrow(subset(fundamentals\_ds\_filter\_1, gvkey == restmt\_item\_gvkey & fyear == restmt\_item\_fyear)))  
   
 if (row\_count > 0){  
 fundamental\_stmt\_row <- fundamentals\_ds\_filter\_1 %>%  
 filter(gvkey == restmt\_item\_gvkey & fyear == restmt\_item\_fyear)  
  
 stmt\_item\_gvkey <- fundamental\_stmt\_row["gvkey"]  
 stmt\_item\_fyear <- fundamental\_stmt\_row["fyear"]  
 stmt\_item\_at <- fundamental\_stmt\_row["at"]  
 stmt\_item\_capx <- fundamental\_stmt\_row["capx"]  
 stmt\_item\_cogs <- fundamental\_stmt\_row["cogs"]  
 stmt\_item\_dltt <- fundamental\_stmt\_row["dltt"]  
 stmt\_item\_epsfi <- fundamental\_stmt\_row["epsfi"]  
 stmt\_item\_epspi <- fundamental\_stmt\_row["epspi"]  
 stmt\_item\_ib <- fundamental\_stmt\_row["ib"]  
 stmt\_item\_ni <- fundamental\_stmt\_row["ni"]  
 stmt\_item\_nopi <- fundamental\_stmt\_row["nopi"]  
 stmt\_item\_pi <- fundamental\_stmt\_row["pi"]  
 stmt\_item\_reuna <- fundamental\_stmt\_row["reuna"]  
 stmt\_item\_seq <- fundamental\_stmt\_row["seq"]  
 stmt\_item\_teq <- fundamental\_stmt\_row["teq"]  
 stmt\_item\_txt <- fundamental\_stmt\_row["txt"]  
 stmt\_item\_wcap <- fundamental\_stmt\_row["wcap"]  
  
 if (!is.na(restmt\_item\_at) & !is.na(stmt\_item\_at) & restmt\_item\_at != stmt\_item\_at){  
 fundamentals\_ds\_filter\_1$restmt\_at[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- 1  
 magnitude <- ((restmt\_item\_at - stmt\_item\_at)/stmt\_item\_at) \* 100.0  
 magnitude <- as.double(round(magnitude, digits = 3))  
 fundamentals\_ds\_filter\_1$restmt\_at\_mag[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- magnitude  
 }  
  
 if (!is.na(restmt\_item\_capx) & !is.na(stmt\_item\_capx) & restmt\_item\_capx != stmt\_item\_capx){  
 fundamentals\_ds\_filter\_1$restmt\_capx[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- 1  
 if (stmt\_item\_capx == 0.0){  
 magnitude <- 100.00  
 }  
 else{  
 magnitude <- ((restmt\_item\_capx - stmt\_item\_capx)/stmt\_item\_capx) \* 100.0  
 }  
 magnitude <- as.double(round(magnitude, digits = 3))  
 fundamentals\_ds\_filter\_1$restmt\_capx\_mag[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- magnitude  
 }  
  
 if (!is.na(restmt\_item\_cogs) & !is.na(stmt\_item\_cogs) & restmt\_item\_cogs != stmt\_item\_cogs){  
 fundamentals\_ds\_filter\_1$restmt\_cogs[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- 1  
 if (stmt\_item\_cogs == 0.0){  
 magnitude <- 100.00  
 }  
 else{  
 magnitude <- ((restmt\_item\_cogs - stmt\_item\_cogs)/stmt\_item\_cogs) \* 100.0  
 }  
 magnitude <- as.double(round(magnitude, digits = 3))  
 fundamentals\_ds\_filter\_1$restmt\_cogs\_mag[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- magnitude  
 }  
  
 if (!is.na(restmt\_item\_dltt) & !is.na(stmt\_item\_dltt) & restmt\_item\_dltt != stmt\_item\_dltt){  
 fundamentals\_ds\_filter\_1$restmt\_dltt[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- 1  
 if (stmt\_item\_dltt == 0.0){  
 magnitude <- 100.00  
 }  
 else{  
 magnitude <- ((restmt\_item\_dltt - stmt\_item\_dltt)/stmt\_item\_dltt) \* 100.0  
 }  
 magnitude <- as.double(round(magnitude, digits = 3))  
 fundamentals\_ds\_filter\_1$restmt\_dltt\_mag[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- magnitude  
 }  
  
 if (!is.na(restmt\_item\_epsfi) & !is.na(stmt\_item\_epsfi) & restmt\_item\_epsfi != stmt\_item\_epsfi){  
 fundamentals\_ds\_filter\_1$restmt\_epsfi[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- 1  
 if (stmt\_item\_epsfi == 0.0){  
 magnitude <- 100.00  
 }  
 else{  
 magnitude <- ((restmt\_item\_epsfi - stmt\_item\_epsfi)/stmt\_item\_epsfi) \* 100.0  
 }  
 magnitude <- as.double(round(magnitude, digits = 3))  
 fundamentals\_ds\_filter\_1$restmt\_epsfi\_mag[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- magnitude  
 }  
  
 if (!is.na(restmt\_item\_epspi) & !is.na(stmt\_item\_epspi) & restmt\_item\_epspi != stmt\_item\_epspi){  
 fundamentals\_ds\_filter\_1$restmt\_epspi[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- 1  
 if (stmt\_item\_epspi == 0.0){  
 magnitude <- 100.00  
 }  
 else{  
 magnitude <- ((restmt\_item\_epspi - stmt\_item\_epspi)/stmt\_item\_epspi) \* 100.0  
 }  
 magnitude <- as.double(round(magnitude, digits = 3))  
 fundamentals\_ds\_filter\_1$restmt\_epspi\_mag[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- magnitude  
 }  
  
 if (!is.na(restmt\_item\_ib) & !is.na(stmt\_item\_ib) & restmt\_item\_ib != stmt\_item\_ib){  
 fundamentals\_ds\_filter\_1$restmt\_ib[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- 1  
 magnitude <- ((restmt\_item\_ib - stmt\_item\_ib)/stmt\_item\_ib) \* 100.0  
 magnitude <- as.double(round(magnitude, digits = 3))  
 fundamentals\_ds\_filter\_1$restmt\_ib\_mag[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- magnitude  
 }  
  
 if (!is.na(restmt\_item\_ni) & !is.na(stmt\_item\_ni) & restmt\_item\_ni != stmt\_item\_ni){  
 fundamentals\_ds\_filter\_1$restmt\_ni[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- 1  
 magnitude <- ((restmt\_item\_ni - stmt\_item\_ni)/stmt\_item\_ni) \* 100.0  
 magnitude <- as.double(round(magnitude, digits = 3))  
 fundamentals\_ds\_filter\_1$restmt\_ni\_mag[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- magnitude  
 }  
  
 if (!is.na(restmt\_item\_nopi) & !is.na(stmt\_item\_nopi) & restmt\_item\_nopi != stmt\_item\_nopi){  
 fundamentals\_ds\_filter\_1$restmt\_nopi[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- 1  
 if (stmt\_item\_nopi == 0.0){  
 magnitude <- 100.00  
 }  
 else{  
 magnitude <- ((restmt\_item\_nopi - stmt\_item\_nopi)/stmt\_item\_nopi) \* 100.0  
 }  
 magnitude <- as.double(round(magnitude, digits = 3))  
 fundamentals\_ds\_filter\_1$restmt\_nopi\_mag[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- magnitude  
 }  
  
 if (!is.na(restmt\_item\_pi) & !is.na(stmt\_item\_pi) & restmt\_item\_pi != stmt\_item\_pi){  
 fundamentals\_ds\_filter\_1$restmt\_pi[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- 1  
 magnitude <- ((restmt\_item\_pi - stmt\_item\_pi)/stmt\_item\_pi) \* 100.0  
 magnitude <- as.double(round(magnitude, digits = 3))  
 fundamentals\_ds\_filter\_1$restmt\_pi\_mag[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- magnitude  
 }  
  
 if (!is.na(restmt\_item\_reuna) & !is.na(stmt\_item\_reuna) & restmt\_item\_reuna != stmt\_item\_reuna){  
 fundamentals\_ds\_filter\_1$restmt\_reuna[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- 1  
 magnitude <- ((restmt\_item\_reuna - stmt\_item\_reuna)/stmt\_item\_reuna) \* 100.0  
 magnitude <- as.double(round(magnitude, digits = 3))  
 fundamentals\_ds\_filter\_1$restmt\_reuna\_mag[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- magnitude  
 }  
  
 if (!is.na(restmt\_item\_seq) & !is.na(stmt\_item\_seq) & restmt\_item\_seq != stmt\_item\_seq){  
 fundamentals\_ds\_filter\_1$restmt\_seq[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- 1  
 magnitude <- ((restmt\_item\_seq - stmt\_item\_seq)/stmt\_item\_seq) \* 100.0  
 magnitude <- as.double(round(magnitude, digits = 3))  
 fundamentals\_ds\_filter\_1$restmt\_seq\_mag[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- magnitude  
 }  
  
 if (!is.na(restmt\_item\_teq) & !is.na(stmt\_item\_teq) & restmt\_item\_teq != stmt\_item\_teq){  
 fundamentals\_ds\_filter\_1$restmt\_teq[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- 1  
 magnitude <- ((restmt\_item\_teq - stmt\_item\_teq)/stmt\_item\_teq) \* 100.0  
 magnitude <- as.double(round(magnitude, digits = 3))  
 fundamentals\_ds\_filter\_1$restmt\_teq\_mag[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- magnitude  
 }  
  
 if (!is.na(restmt\_item\_txt) & !is.na(stmt\_item\_txt) & restmt\_item\_txt != stmt\_item\_txt){  
 fundamentals\_ds\_filter\_1$restmt\_txt[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- 1  
 if (stmt\_item\_txt == 0.0){  
 magnitude <- 100.00  
 }  
 else{  
 magnitude <- ((restmt\_item\_txt - stmt\_item\_txt)/stmt\_item\_txt) \* 100.0  
 }  
 magnitude <- as.double(round(magnitude, digits = 3))  
 fundamentals\_ds\_filter\_1$restmt\_txt\_mag[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- magnitude  
 }  
  
 if (!is.na(restmt\_item\_wcap) & !is.na(stmt\_item\_wcap) & restmt\_item\_wcap != stmt\_item\_wcap){  
 fundamentals\_ds\_filter\_1$restmt\_wcap[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- 1  
 magnitude <- ((restmt\_item\_wcap - stmt\_item\_wcap)/stmt\_item\_wcap) \* 100.0  
 magnitude <- as.double(round(magnitude, digits = 3))  
 fundamentals\_ds\_filter\_1$restmt\_wcap\_mag[fundamentals\_ds\_filter\_1$gvkey == restmt\_item\_gvkey & fundamentals\_ds\_filter\_1$fyear == restmt\_item\_fyear] <- magnitude  
 }  
   
 }  
}  
#head(fundamentals\_ds\_filter\_1)

nrow(fundamentals\_ds\_filter\_1)

## [1] 9078

# Removing aco == NA as all those rows do not have any data  
fundamentals\_ds\_filter\_2 <- subset(fundamentals\_ds\_filter\_1, !is.na(aco))  
nrow(fundamentals\_ds\_filter\_2)

## [1] 8763

fundamentals\_ds\_filter\_2[is.na(fundamentals\_ds\_filter\_2)] <- 0  
#nrow(fundamentals\_ds\_filter\_2)  
#head(fundamentals\_ds\_filter\_2)

colnames(fundamentals\_ds\_filter\_2)

## [1] "gvkey" "fyear" "tic"   
## [4] "aco" "acominc" "act"   
## [7] "ao" "aocidergl" "aocipen"   
## [10] "aodo" "aoloch" "ap"   
## [13] "aqc" "at" "bkvlps"   
## [16] "caps" "capx" "ceq"   
## [19] "ceqt" "ch" "che"   
## [22] "chech" "ci" "cogs"   
## [25] "cshi" "csho" "cstk"   
## [28] "cstkcv" "dd1" "dilavx"   
## [31] "dlc" "dltt" "dm"   
## [34] "dn" "dpact" "dpc"   
## [37] "dvt" "ebit" "ebitda"   
## [40] "epsfi" "epspi" "esub"   
## [43] "fiao" "fincf" "fopo"   
## [46] "gdwl" "gp" "ib"   
## [49] "icapt" "intan" "intano"   
## [52] "invt" "ivch" "ivncf"   
## [55] "ivst" "lo" "lse"   
## [58] "lt" "ni" "nopi"   
## [61] "nopio" "oancf" "oiadp"   
## [64] "oibdp" "opeps" "pi"   
## [67] "ppegt" "re" "reajo"   
## [70] "rect" "recta" "reuna"   
## [73] "revt" "seq" "siv"   
## [76] "spce" "spi" "sppiv"   
## [79] "sstk" "teq" "tstk"   
## [82] "tstkn" "txp" "txr"   
## [85] "txt" "wcap" "xint"   
## [88] "restmt\_at" "restmt\_at\_mag" "restmt\_capx"   
## [91] "restmt\_capx\_mag" "restmt\_cogs" "restmt\_cogs\_mag"   
## [94] "restmt\_dltt" "restmt\_dltt\_mag" "restmt\_epsfi"   
## [97] "restmt\_epsfi\_mag" "restmt\_epspi" "restmt\_epspi\_mag"  
## [100] "restmt\_ib" "restmt\_ib\_mag" "restmt\_ni"   
## [103] "restmt\_ni\_mag" "restmt\_nopi" "restmt\_nopi\_mag"   
## [106] "restmt\_pi" "restmt\_pi\_mag" "restmt\_reuna"   
## [109] "restmt\_reuna\_mag" "restmt\_seq" "restmt\_seq\_mag"   
## [112] "restmt\_teq" "restmt\_teq\_mag" "restmt\_txt"   
## [115] "restmt\_txt\_mag" "restmt\_wcap" "restmt\_wcap\_mag"   
## [118] "restmt\_ci" "restmt\_ci\_mag"

final\_ds\_initial <- fundamentals\_ds\_filter\_2 %>%  
 group\_by(gvkey,tic) %>%  
 summarize(  
 aco = mean(aco),  
 acominc = mean(acominc),  
 act = mean(act),  
 ao = mean(ao),  
 aocidergl = mean(aocidergl),  
 aocipen = mean(aocipen),  
 aodo = mean(aodo),  
 aoloch = mean(aoloch),  
 ap = mean(ap),  
 aqc = mean(aqc),  
 at = mean(at),  
 bkvlps = mean(bkvlps),  
 caps = mean(caps),  
 capx = mean(capx),  
 ceq = mean(ceq),  
 ceqt = mean(ceqt),  
 ch = mean(ch),  
 che = mean(che),  
 chech = mean(chech),  
 ci = mean(ci),  
 cogs = mean(cogs),  
 cshi = mean(cshi),  
 csho = mean(csho),  
 cstk = mean(cstk),  
 cstkcv = mean(cstkcv),  
 dd1 = mean(dd1),  
 dilavx = mean(dilavx),  
 dlc = mean(dlc),  
 dltt = mean(dltt),  
 dm = mean(dm),  
 dn = mean(dn),  
 dpact = mean(dpact),  
 dpc = mean(dpc),  
 dvt = mean(dvt),  
 ebit = mean(ebit),  
 ebitda = mean(ebitda),  
 epsfi = mean(epsfi),  
 epspi = mean(epspi),  
 esub = mean(esub),  
 fiao = mean(fiao),  
 fincf = mean(fincf),  
 fopo = mean(fopo),  
 gdwl = mean(gdwl),  
 gp = mean(gp),  
 ib = mean(ib),  
 icapt = mean(icapt),  
 intan = mean(intan),  
 intano = mean(intano),  
 invt = mean(invt),  
 ivch = mean(ivch),  
 ivncf = mean(ivncf),  
 ivst = mean(ivst),  
 lo = mean(lo),  
 lse = mean(lse),  
 lt = mean(lt),  
 ni = mean(ni),  
 nopi = mean(nopi),  
 nopio = mean(nopio),  
 oancf = mean(oancf),  
 oiadp = mean(oiadp),  
 oibdp = mean(oibdp),  
 opeps = mean(opeps),  
 pi = mean(pi),  
 ppegt = mean(ppegt),  
 re = mean(re),  
 reajo = mean(reajo),  
 rect = mean(rect),  
 recta = mean(recta),  
 reuna = mean(reuna),  
 revt = mean(revt),  
 seq = mean( seq ),  
 siv = mean( siv ),  
 spce = mean(spce),  
 spi = mean(spi),  
 sppiv = mean(sppiv),  
 sstk = mean(sstk),  
 teq = mean(teq),  
 tstk = mean(tstk),  
 tstkn = mean(tstkn),  
 txp = mean(txp),  
 txr = mean(txr),  
 txt = mean(txt),  
 wcap = mean(wcap),  
 xint = mean(xint),  
 restmt\_at = mean(restmt\_at),  
 restmt\_at\_mag = mean(restmt\_at\_mag),  
 restmt\_capx = mean(restmt\_capx),  
 restmt\_capx\_mag = mean(restmt\_capx\_mag),  
 restmt\_cogs = mean(restmt\_cogs),  
 restmt\_cogs\_mag = mean(restmt\_cogs\_mag),  
 restmt\_dltt = mean(restmt\_dltt),  
 restmt\_dltt\_mag = mean(restmt\_dltt\_mag),  
 restmt\_epsfi = mean(restmt\_epsfi),  
 restmt\_epsfi\_mag = mean(restmt\_epsfi\_mag),  
 restmt\_epspi = mean(restmt\_epspi),  
 restmt\_epspi\_mag = mean(restmt\_epspi\_mag),  
 restmt\_ib = mean(restmt\_ib),  
 restmt\_ib\_mag = mean(restmt\_ib\_mag),  
 restmt\_ni = mean(restmt\_ni),  
 restmt\_ni\_mag = mean(restmt\_ni\_mag),  
 restmt\_nopi = mean(restmt\_nopi),  
 restmt\_nopi\_mag = mean(restmt\_nopi\_mag),  
 restmt\_pi = mean(restmt\_pi),  
 restmt\_pi\_mag = mean(restmt\_pi\_mag),  
 restmt\_reuna = mean(restmt\_reuna),  
 restmt\_reuna\_mag = mean(restmt\_reuna\_mag),  
 restmt\_seq = mean(restmt\_seq),  
 restmt\_seq\_mag = mean(restmt\_seq\_mag),  
 restmt\_teq = mean(restmt\_teq),  
 restmt\_teq\_mag = mean(restmt\_teq\_mag),  
 restmt\_txt = mean(restmt\_txt),  
 restmt\_txt\_mag = mean(restmt\_txt\_mag),  
 restmt\_wcap = mean(restmt\_wcap),  
 restmt\_wcap\_mag = mean(restmt\_wcap\_mag),  
 )

## `summarise()` regrouping output by 'gvkey' (override with `.groups` argument)

head(final\_ds\_initial)

## # A tibble: 6 x 116  
## # Groups: gvkey [6]  
## gvkey tic aco acominc act ao aocidergl aocipen aodo  
## <int> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1076 AAN 0. 2.47e-1 0. 8.21e+1 0 0. 8.21e+1  
## 2 1078 ABT 4.63e+3 -2.39e+3 2.42e+4 2.18e+3 76.8 -2.46e3 2.18e+3  
## 3 1177 AET 1.68e+3 -1.07e+3 7.71e+3 5.64e+3 0 -1.64e3 5.64e+3  
## 4 1234 ATRI 2.86e+0 0. 7.07e+1 7.66e-1 0 0. 7.66e-1  
## 5 1239 ACV 4.76e+1 -7.03e+1 8.04e+2 8.47e+1 -3.29 -5.31e0 8.33e+1  
## 6 1259 ACEL 1.81e-1 0. 8.24e-1 1.06e-1 0 0. 4.00e-3  
## # ... with 107 more variables: aoloch <dbl>, ap <dbl>, aqc <dbl>, at <dbl>,  
## # bkvlps <dbl>, caps <dbl>, capx <dbl>, ceq <dbl>, ceqt <dbl>, ch <dbl>,  
## # che <dbl>, chech <dbl>, ci <dbl>, cogs <dbl>, cshi <dbl>, csho <dbl>,  
## # cstk <dbl>, cstkcv <dbl>, dd1 <dbl>, dilavx <dbl>, dlc <dbl>, dltt <dbl>,  
## # dm <dbl>, dn <dbl>, dpact <dbl>, dpc <dbl>, dvt <dbl>, ebit <dbl>,  
## # ebitda <dbl>, epsfi <dbl>, epspi <dbl>, esub <dbl>, fiao <dbl>,  
## # fincf <dbl>, fopo <dbl>, gdwl <dbl>, gp <dbl>, ib <dbl>, icapt <dbl>,  
## # intan <dbl>, intano <dbl>, invt <dbl>, ivch <dbl>, ivncf <dbl>, ivst <dbl>,  
## # lo <dbl>, lse <dbl>, lt <dbl>, ni <dbl>, nopi <dbl>, nopio <dbl>,  
## # oancf <dbl>, oiadp <dbl>, oibdp <dbl>, opeps <dbl>, pi <dbl>, ppegt <dbl>,  
## # re <dbl>, reajo <dbl>, rect <dbl>, recta <dbl>, reuna <dbl>, revt <dbl>,  
## # seq <dbl>, siv <dbl>, spce <dbl>, spi <dbl>, sppiv <dbl>, sstk <dbl>,  
## # teq <dbl>, tstk <dbl>, tstkn <dbl>, txp <dbl>, txr <dbl>, txt <dbl>,  
## # wcap <dbl>, xint <dbl>, restmt\_at <dbl>, restmt\_at\_mag <dbl>,  
## # restmt\_capx <dbl>, restmt\_capx\_mag <dbl>, restmt\_cogs <dbl>,  
## # restmt\_cogs\_mag <dbl>, restmt\_dltt <dbl>, restmt\_dltt\_mag <dbl>,  
## # restmt\_epsfi <dbl>, restmt\_epsfi\_mag <dbl>, restmt\_epspi <dbl>,  
## # restmt\_epspi\_mag <dbl>, restmt\_ib <dbl>, restmt\_ib\_mag <dbl>,  
## # restmt\_ni <dbl>, restmt\_ni\_mag <dbl>, restmt\_nopi <dbl>,  
## # restmt\_nopi\_mag <dbl>, restmt\_pi <dbl>, restmt\_pi\_mag <dbl>,  
## # restmt\_reuna <dbl>, restmt\_reuna\_mag <dbl>, restmt\_seq <dbl>, ...

final\_ds\_initial\_1 <- final\_ds\_initial   
  
for (row in 1:nrow(final\_ds\_initial\_1)){  
 row\_item\_gvkey <- as.integer(final\_ds\_initial\_1[row, "gvkey"])  
   
 restmt\_at <- final\_ds\_initial\_1[row, "restmt\_at"]  
 restmt\_at\_mag <- final\_ds\_initial\_1[row, "restmt\_at\_mag"]  
 if (restmt\_at >= 0.5){  
 restmt\_at <- 1  
 restmt\_at\_mag <- as.double(restmt\_at\_mag)  
 }  
 else{  
 restmt\_at <- 0  
 restmt\_at\_mag <- 0.0  
 }  
 final\_ds\_initial\_1$restmt\_at[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_at  
 final\_ds\_initial\_1$restmt\_at\_mag[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_at\_mag  
   
 restmt\_capx <- final\_ds\_initial\_1[row, "restmt\_capx"]  
 restmt\_capx\_mag <- final\_ds\_initial\_1[row, "restmt\_capx\_mag"]  
 if (restmt\_capx >= 0.5){  
 restmt\_capx <- 1  
 restmt\_capx\_mag <- as.double(restmt\_capx\_mag)  
 }  
 else{  
 restmt\_capx <- 0  
 restmt\_capx\_mag <- 0.0  
 }  
 final\_ds\_initial\_1$restmt\_capx[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_capx  
 final\_ds\_initial\_1$restmt\_capx\_mag[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_capx\_mag  
   
 restmt\_cogs <- final\_ds\_initial\_1[row, "restmt\_cogs"]  
 restmt\_cogs\_mag <- final\_ds\_initial\_1[row, "restmt\_cogs\_mag"]  
 if (restmt\_cogs >= 0.5){  
 restmt\_cogs <- 1  
 restmt\_cogs\_mag <- as.double(restmt\_cogs\_mag)  
 }  
 else{  
 restmt\_cogs <- 0  
 restmt\_cogs\_mag <- 0.0  
 }  
 final\_ds\_initial\_1$restmt\_cogs[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_cogs  
 final\_ds\_initial\_1$restmt\_cogs\_mag[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_cogs\_mag  
   
 restmt\_dltt <- final\_ds\_initial\_1[row, "restmt\_dltt"]  
 restmt\_dltt\_mag <- final\_ds\_initial\_1[row, "restmt\_dltt\_mag"]  
 if (restmt\_dltt >= 0.5){  
 restmt\_dltt <- 1  
 restmt\_dltt\_mag <- as.double(restmt\_dltt\_mag)  
 }  
 else{  
 restmt\_dltt <- 0  
 restmt\_dltt\_mag <- 0.0  
 }  
 final\_ds\_initial\_1$restmt\_dltt[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_dltt  
 final\_ds\_initial\_1$restmt\_dltt\_mag[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_dltt\_mag  
   
   
 restmt\_epsfi <- final\_ds\_initial\_1[row, "restmt\_epsfi"]  
 restmt\_epsfi\_mag <- final\_ds\_initial\_1[row, "restmt\_epsfi\_mag"]  
 if (restmt\_epsfi >= 0.5){  
 restmt\_epsfi <- 1  
 restmt\_epsfi\_mag <- as.double(restmt\_epsfi\_mag)  
 }  
 else{  
 restmt\_epsfi <- 0  
 restmt\_epsfi\_mag <- 0.0  
 }  
 final\_ds\_initial\_1$restmt\_epsfi[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_epsfi  
 final\_ds\_initial\_1$restmt\_epsfi\_mag[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_epsfi\_mag  
   
   
 restmt\_epspi <- final\_ds\_initial\_1[row, "restmt\_epspi"]  
 restmt\_epspi\_mag <- final\_ds\_initial\_1[row, "restmt\_epspi\_mag"]  
 if (restmt\_epspi >= 0.5){  
 restmt\_epspi <- 1  
 restmt\_epspi\_mag <- as.double(restmt\_epspi\_mag)  
 }  
 else{  
 restmt\_epspi <- 0  
 restmt\_epspi\_mag <- 0.0  
 }  
 final\_ds\_initial\_1$restmt\_epspi[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_epspi  
 final\_ds\_initial\_1$restmt\_epspi\_mag[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_epspi\_mag  
   
 restmt\_ib <- final\_ds\_initial\_1[row, "restmt\_ib"]  
 restmt\_ib\_mag <- final\_ds\_initial\_1[row, "restmt\_ib\_mag"]  
 if (restmt\_ib >= 0.5){  
 restmt\_ib <- 1  
 restmt\_ib\_mag <- as.double(restmt\_ib\_mag)  
 }  
 else{  
 restmt\_ib <- 0  
 restmt\_ib\_mag <- 0.0  
 }  
 final\_ds\_initial\_1$restmt\_ib[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_ib  
 final\_ds\_initial\_1$restmt\_ib\_mag[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_ib\_mag  
   
 restmt\_ni <- final\_ds\_initial\_1[row, "restmt\_ni"]  
 restmt\_ni\_mag <- final\_ds\_initial\_1[row, "restmt\_ni\_mag"]  
 if (restmt\_ni >= 0.5){  
 restmt\_ni <- 1  
 restmt\_ni\_mag <- as.double(restmt\_ni\_mag)  
 }  
 else{  
 restmt\_ni <- 0  
 restmt\_ni\_mag <- 0.0  
 }  
 final\_ds\_initial\_1$restmt\_ni[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_ni  
 final\_ds\_initial\_1$restmt\_ni\_mag[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_ni\_mag  
   
 restmt\_nopi <- final\_ds\_initial\_1[row, "restmt\_nopi"]  
 restmt\_nopi\_mag <- final\_ds\_initial\_1[row, "restmt\_nopi\_mag"]  
 if (restmt\_nopi >= 0.5){  
 restmt\_nopi <- 1  
 restmt\_nopi\_mag <- as.double(restmt\_nopi\_mag)  
 }  
 else{  
 restmt\_nopi <- 0  
 restmt\_nopi\_mag <- 0.0  
 }  
 final\_ds\_initial\_1$restmt\_nopi[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_nopi  
 final\_ds\_initial\_1$restmt\_nopi\_mag[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_nopi\_mag  
   
 restmt\_pi <- final\_ds\_initial\_1[row, "restmt\_pi"]  
 restmt\_pi\_mag <- final\_ds\_initial\_1[row, "restmt\_pi\_mag"]  
 if (restmt\_pi >= 0.5){  
 restmt\_pi <- 1  
 restmt\_pi\_mag <- as.double(restmt\_pi\_mag)  
 }  
 else{  
 restmt\_pi <- 0  
 restmt\_pi\_mag <- 0.0  
 }  
 final\_ds\_initial\_1$restmt\_pi[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_pi  
 final\_ds\_initial\_1$restmt\_pi\_mag[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_pi\_mag  
   
 restmt\_reuna <- final\_ds\_initial\_1[row, "restmt\_reuna"]  
 restmt\_reuna\_mag <- final\_ds\_initial\_1[row, "restmt\_reuna\_mag"]  
 if (restmt\_reuna >= 0.5){  
 restmt\_reuna <- 1  
 restmt\_reuna\_mag <- as.double(restmt\_reuna\_mag)  
 }  
 else{  
 restmt\_reuna <- 0  
 restmt\_reuna\_mag <- 0.0  
 }  
 final\_ds\_initial\_1$restmt\_reuna[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_reuna  
 final\_ds\_initial\_1$restmt\_reuna\_mag[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_reuna\_mag  
   
 restmt\_seq <- final\_ds\_initial\_1[row, "restmt\_seq"]  
 restmt\_seq\_mag <- final\_ds\_initial\_1[row, "restmt\_seq\_mag"]  
 if (restmt\_seq >= 0.5){  
 restmt\_seq <- 1  
 restmt\_seq\_mag <- as.double(restmt\_seq\_mag)  
 }  
 else{  
 restmt\_seq <- 0  
 restmt\_seq\_mag <- 0.0  
 }  
 final\_ds\_initial\_1$restmt\_seq[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_seq  
 final\_ds\_initial\_1$restmt\_seq\_mag[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_seq\_mag  
   
 restmt\_teq <- final\_ds\_initial\_1[row, "restmt\_teq"]  
 restmt\_teq\_mag <- final\_ds\_initial\_1[row, "restmt\_teq\_mag"]  
 if (restmt\_teq >= 0.5){  
 restmt\_teq <- 1  
 restmt\_teq\_mag <- as.double(restmt\_teq\_mag)  
 }  
 else{  
 restmt\_teq <- 0  
 restmt\_teq\_mag <- 0.0  
 }  
 final\_ds\_initial\_1$restmt\_teq[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_teq  
 final\_ds\_initial\_1$restmt\_teq\_mag[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_teq\_mag  
   
 restmt\_txt <- final\_ds\_initial\_1[row, "restmt\_txt"]  
 restmt\_txt\_mag <- final\_ds\_initial\_1[row, "restmt\_txt\_mag"]  
 if (restmt\_txt >= 0.5){  
 restmt\_txt <- 1  
 restmt\_txt\_mag <- as.double(restmt\_txt\_mag)  
 }  
 else{  
 restmt\_txt <- 0  
 restmt\_txt\_mag <- 0.0  
 }  
 final\_ds\_initial\_1$restmt\_txt[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_txt  
 final\_ds\_initial\_1$restmt\_txt\_mag[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_txt\_mag  
   
 restmt\_wcap <- final\_ds\_initial\_1[row, "restmt\_wcap"]  
 restmt\_wcap\_mag <- final\_ds\_initial\_1[row, "restmt\_wcap\_mag"]  
 if (restmt\_wcap >= 0.5){  
 restmt\_wcap <- 1  
 restmt\_wcap\_mag <- as.double(restmt\_wcap\_mag)  
 }  
 else{  
 restmt\_wcap <- 0  
 restmt\_wcap\_mag <- 0.0  
 }  
 final\_ds\_initial\_1$restmt\_wcap[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_wcap  
 final\_ds\_initial\_1$restmt\_wcap\_mag[final\_ds\_initial\_1$gvkey == row\_item\_gvkey] <- restmt\_wcap\_mag  
   
}  
  
restmt\_var\_ds <- subset(final\_ds\_initial\_1, select = c(gvkey,  
 restmt\_at,restmt\_at\_mag,  
 restmt\_capx,restmt\_capx\_mag,  
 restmt\_cogs, restmt\_cogs\_mag,  
 restmt\_dltt, restmt\_dltt\_mag,  
 restmt\_epsfi, restmt\_epsfi\_mag,  
 restmt\_epspi, restmt\_epspi\_mag,  
 restmt\_ib, restmt\_ib\_mag,  
 restmt\_ni, restmt\_ni\_mag,  
 restmt\_nopi, restmt\_nopi\_mag,  
 restmt\_pi, restmt\_pi\_mag,  
 restmt\_reuna, restmt\_reuna\_mag,  
 restmt\_seq, restmt\_seq\_mag,  
 restmt\_teq, restmt\_teq\_mag,  
 restmt\_txt, restmt\_txt\_mag,  
 restmt\_wcap, restmt\_wcap\_mag  
 ))  
  
head(restmt\_var\_ds)

## # A tibble: 6 x 31  
## # Groups: gvkey [6]  
## gvkey restmt\_at restmt\_at\_mag restmt\_capx restmt\_capx\_mag restmt\_cogs  
## <int> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1076 1 -0.067 0 0 1  
## 2 1078 0 0 0 0 0  
## 3 1177 0 0 0 0 0  
## 4 1234 0 0 0 0 0  
## 5 1239 0 0 0 0 0  
## 6 1259 0 0 0 0 0  
## # ... with 25 more variables: restmt\_cogs\_mag <dbl>, restmt\_dltt <dbl>,  
## # restmt\_dltt\_mag <dbl>, restmt\_epsfi <dbl>, restmt\_epsfi\_mag <dbl>,  
## # restmt\_epspi <dbl>, restmt\_epspi\_mag <dbl>, restmt\_ib <dbl>,  
## # restmt\_ib\_mag <dbl>, restmt\_ni <dbl>, restmt\_ni\_mag <dbl>,  
## # restmt\_nopi <dbl>, restmt\_nopi\_mag <dbl>, restmt\_pi <dbl>,  
## # restmt\_pi\_mag <dbl>, restmt\_reuna <dbl>, restmt\_reuna\_mag <dbl>,  
## # restmt\_seq <dbl>, restmt\_seq\_mag <dbl>, restmt\_teq <dbl>,  
## # restmt\_teq\_mag <dbl>, restmt\_txt <dbl>, restmt\_txt\_mag <dbl>,  
## # restmt\_wcap <dbl>, restmt\_wcap\_mag <dbl>

final\_ds\_initial\_2 <- final\_ds\_initial\_1  
#summary(final\_ds\_initial\_2)  
write.csv(final\_ds\_initial\_2, file = "data/final\_ds\_initial\_2.csv", row.names=FALSE, na="")

cor\_matrix\_ds <- subset(final\_ds\_initial\_2, select = -c(gvkey,tic, aodo,seq,ivch,nopio,spce,reuna,dilavx,ebitda,csho,epsfi,   
 ib,pi,  
 oiadp,oibdp,gdwl,esub,  
 restmt\_seq,restmt\_seq\_mag,  
 restmt\_ib,restmt\_ib\_mag,  
 restmt\_epsfi,restmt\_epsfi\_mag, restmt\_cogs\_mag))  
cor\_matrix <- cor(cor\_matrix\_ds)  
cor\_matrix %>%  
 as.data.frame() %>%  
 mutate(var1 = rownames(.)) %>%  
 gather(var2, value, -var1) %>%  
 arrange(desc(value)) %>%  
 group\_by(value) %>%  
 filter(row\_number()==1)

## # A tibble: 2,415 x 3  
## # Groups: value [2,415]  
## var1 var2 value  
## <chr> <chr> <dbl>  
## 1 aco aco 1   
## 2 ni ci 0.990  
## 3 lt at 0.981  
## 4 icapt at 0.979  
## 5 revt cogs 0.979  
## 6 ppegt dpact 0.972  
## 7 teq icapt 0.966  
## 8 teq ceq 0.953  
## 9 oancf ebit 0.943  
## 10 ppegt capx 0.939  
## # ... with 2,405 more rows

#corrplot(cor\_matrix, method = "ellipse")  
  
#ncol(cor\_matrix\_ds)

fundamentals\_final\_ds <- subset(final\_ds\_initial\_2, select = -c(aodo,seq,ivch,nopio,spce,reuna,dilavx,ebitda,csho,epsfi,   
 ib,pi,  
 oiadp,oibdp,gdwl,esub,  
 restmt\_seq,restmt\_seq\_mag,  
 restmt\_ib,restmt\_ib\_mag,  
 restmt\_epsfi,restmt\_epsfi\_mag, restmt\_cogs\_mag))

stocks\_init\_ds <- read.csv("./data/Stocks\_DS.csv", na.strings=c(""," "))  
nrow(stocks\_init\_ds)

## [1] 4187047

names(stocks\_init\_ds)[names(stocks\_init\_ds) == "ï..gvkey"] <- "gvkey"  
stocks\_init\_limited\_cols <- subset(stocks\_init\_ds, select = c(gvkey,cshtrd,prccd,prchd,prcld,prcod,trfd))  
stocks\_init\_limited\_cols <- stocks\_init\_limited\_cols[!is.na(stocks\_init\_limited\_cols$cshtrd)&!is.na(stocks\_init\_limited\_cols$prccd)  
 &!is.na(stocks\_init\_limited\_cols$prchd)&!is.na(stocks\_init\_limited\_cols$prcld)  
 &!is.na(stocks\_init\_limited\_cols$trfd),]  
stocks\_init\_limited\_cols$prcod[is.na(stocks\_init\_limited\_cols$prcod)] <- (stocks\_init\_limited\_cols$prchd + stocks\_init\_limited\_cols$prcld)/2  
  
stocks\_grouped\_data <- stocks\_init\_limited\_cols %>%  
 group\_by(gvkey) %>%  
 summarize(  
 cshtrd\_m = mean(cshtrd),  
 prccd\_m = mean(prccd),  
 prchd\_m = mean(prchd),  
 prcld\_m = mean(prcld),  
 prcod\_m = mean(prcod),  
 trfd\_m = mean(trfd)  
 )

## `summarise()` ungrouping output (override with `.groups` argument)

fundamental\_stocks\_data <- fundamentals\_final\_ds %>%  
 inner\_join(stocks\_grouped\_data, by = 'gvkey')  
#summary(fundamental\_stocks\_data)

cor\_matrix\_ds <- subset(fundamental\_stocks\_data, select = -c(gvkey,tic))  
cor\_matrix <- cor(cor\_matrix\_ds)  
cor\_matrix %>%  
 as.data.frame() %>%  
 mutate(var1 = rownames(.)) %>%  
 gather(var2, value, -var1) %>%  
 arrange(desc(value)) %>%  
 group\_by(value) %>%  
 filter(row\_number()==1)

## # A tibble: 2,849 x 3  
## # Groups: value [2,849]  
## var1 var2 value  
## <chr> <chr> <dbl>  
## 1 aco aco 1   
## 2 lse at 1.00   
## 3 prcld\_m prccd\_m 0.999  
## 4 prchd\_m prccd\_m 0.998  
## 5 prcld\_m prchd\_m 0.995  
## 6 opeps epspi 0.990  
## 7 ni ci 0.990  
## 8 lt at 0.981  
## 9 icapt at 0.979  
## 10 revt cogs 0.979  
## # ... with 2,839 more rows

fundamental\_stocks\_data\_final <- subset(fundamental\_stocks\_data, select = -c(prccd\_m))  
#summary(fundamental\_stocks\_data\_final)  
nrow(fundamental\_stocks\_data\_final)

## [1] 2459

securities\_init\_ds <- read.csv("./data/Securities\_DS.csv", na.strings=c(""," "))  
names(securities\_init\_ds)[names(securities\_init\_ds) == "ï..gvkey"] <- "gvkey"  
securities\_init\_ds\_1 <- subset(securities\_init\_ds, select = -c(iid,isalrt,primiss,ajexm,  
 spgim,spiim,spmim,cheqvm,curcddvm,dvpsxm,  
 sphcusip,sphiid, sphmid,sphname,sphsec,sphtic,sphvg,sph100,  
 cyear,mkvalincl,exchg,tpci,city,  
 conml,costat,ggroup,gind, gsector,gsubind,loc,naics,sic,state, curcdm,   
 navm,adrrm,rawpm,rawxm,cshoq,csfsm,  
 datadate,tic,conm,cmth  
 ))  
  
  
#summary(securities\_init\_ds\_1)

fund\_stock\_securities\_ds <- fundamental\_stocks\_data\_final   
securities\_init\_ds\_2 <- securities\_init\_ds\_1 %>%  
 filter(!is.na(trfm) & !is.na(trt1m)) %>%  
 group\_by(gvkey) %>%  
 summarise(  
 trfm\_m = mean(trfm)  
 )  
  
fund\_stock\_securities\_ds$trfm\_m <- NA  
for (row in 1:nrow(fund\_stock\_securities\_ds)){  
 row\_item\_gvkey <- as.integer(fund\_stock\_securities\_ds[row, "gvkey"])  
 specific\_security <- securities\_init\_ds\_2 %>%  
 filter(gvkey == row\_item\_gvkey)   
 if (nrow(specific\_security) > 0){  
 security\_row <- head(specific\_security, 1)  
 trfm\_m <- as.numeric(security\_row$trfm\_m)   
 fund\_stock\_securities\_ds$trfm\_m[fund\_stock\_securities\_ds$gvkey == row\_item\_gvkey] <- trfm\_m  
 }  
}  
  
  
  
securities\_init\_ds\_3 <- securities\_init\_ds\_1 %>%  
 filter(!is.na(dvrate)) %>%  
 group\_by(gvkey) %>%  
 summarise(  
 dvrate\_m = mean(dvrate)  
 )  
  
fund\_stock\_securities\_ds$dvrate\_m <- NA  
for (row in 1:nrow(fund\_stock\_securities\_ds)){  
 row\_item\_gvkey <- as.integer(fund\_stock\_securities\_ds[row, "gvkey"])  
 specific\_security <- securities\_init\_ds\_3 %>%  
 filter(gvkey == row\_item\_gvkey)   
 if (nrow(specific\_security) > 0){  
 security\_row <- head(specific\_security, 1)  
 dvrate\_m <- as.numeric(security\_row$dvrate\_m)   
 fund\_stock\_securities\_ds$dvrate\_m[fund\_stock\_securities\_ds$gvkey == row\_item\_gvkey] <- dvrate\_m  
 }  
}  
  
  
#summary(fund\_stock\_securities\_ds)

ratings\_init\_ds <- read.csv("./data/Ratings\_DS.csv", na.strings=c("", " "))  
names(ratings\_init\_ds)[names(ratings\_init\_ds) == "ï..gvkey"] <- "gvkey"  
ratings\_init\_ds$datadate <- as.Date(ratings\_init\_ds$datadate, "%m/%d/%Y")  
ratings\_init\_ds$splticrm = factor(ratings\_init\_ds$splticrm, levels=c(levels(ratings\_init\_ds$splticrm), "NR"))  
ratings\_init\_ds$splticrm[is.na(ratings\_init\_ds$splticrm)] = "NR"  
  
ratings\_init\_ds$splticrm\_num\_value <- 0  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "AAA"] <- 100  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "AA"] <- 90  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "AA-"] <- 85  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "A+"] <- 80  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "A"] <- 75  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "A-"] <- 70  
  
  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "BBB+"] <- 65  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "BBB"] <- 60  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "BBB-"] <- 55  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "BB+"] <- 50  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "BB"] <- 45  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "BB-"] <- 40  
  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "B+"] <- 35  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "B"] <- 30  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "B-"] <- 25  
  
  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "CCC+"] <- 20  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "CCC"] <- 19  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "CCC-"] <- 18  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "CC"] <- 17  
  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "D"] <- 10  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "SD"] <- 10  
ratings\_init\_ds$splticrm\_num\_value[ratings\_init\_ds$splticrm == "NR"] <- 0  
ratings\_init\_ds$splticrm\_num\_value <- factor(ratings\_init\_ds$splticrm\_num\_value)  
#levels(ratings\_init\_ds$splticrm)  
#str(ratings\_init\_ds)

fund\_stock\_securities\_rating\_ds <- fund\_stock\_securities\_ds #%>%  
 #filter(gvkey == 1078)  
fund\_stock\_securities\_rating\_ds$sp\_rating <- "NOTRATED"   
rated\_companies <- ratings\_init\_ds %>%  
 filter(splticrm != "NR")  
  
for (row in 1:nrow(fund\_stock\_securities\_rating\_ds)){  
 row\_item\_gvkey <- as.integer(fund\_stock\_securities\_rating\_ds[row, "gvkey"])  
  
 specific\_rating <- rated\_companies %>%  
 filter(gvkey == row\_item\_gvkey) %>%  
 arrange(datadate)  
 if (nrow(specific\_rating) > 0){  
 first\_row <- head(specific\_rating, 1)  
 last\_row <- tail(specific\_rating, 1)  
 start\_value <- as.integer(first\_row$splticrm\_num\_value)  
 end\_value <- as.integer(last\_row$splticrm\_num\_value)  
 if (start\_value == end\_value){  
 fund\_stock\_securities\_rating\_ds$sp\_rating[fund\_stock\_securities\_rating\_ds$gvkey == row\_item\_gvkey] <- "NoCHANGE"  
 }else if (start\_value < end\_value){  
 fund\_stock\_securities\_rating\_ds$sp\_rating[fund\_stock\_securities\_rating\_ds$gvkey == row\_item\_gvkey] <- "INCREASED"  
 }else if (start\_value > end\_value){  
 fund\_stock\_securities\_rating\_ds$sp\_rating[fund\_stock\_securities\_rating\_ds$gvkey == row\_item\_gvkey] <- "DECREASED"  
 }  
 }  
}  
fund\_stock\_securities\_rating\_ds$sp\_rating <- factor(fund\_stock\_securities\_rating\_ds$sp\_rating)  
#summary(fund\_stock\_securities\_rating\_ds)

sca\_fillings\_ds <- read.csv("./data/SCA\_Filings\_and\_Settlements.csv", na.strings=c(""," "))  
sca\_fillings\_ds$SettlementAmount = gsub("\\$", "", sca\_fillings\_ds$SettlementAmount)  
sca\_fillings\_ds$SettlementAmount = as.numeric(gsub("\\,", "", sca\_fillings\_ds$SettlementAmount))  
#summary(sca\_fillings\_ds)

fund\_stock\_securities\_rating\_ds$litigated <- 0  
fund\_stock\_securities\_rating\_ds$litigation\_settlement <- NA  
fund\_stock\_securities\_rating\_ds\_1 <- fund\_stock\_securities\_rating\_ds   
  
for (row in 1:nrow(fund\_stock\_securities\_rating\_ds\_1)){  
 row\_item\_tic <- lapply(fund\_stock\_securities\_rating\_ds\_1[row, "tic"], as.character)  
 row\_item\_gvkey <- as.integer(fund\_stock\_securities\_rating\_ds[row, "gvkey"])  
 specific\_sca\_filings <- sca\_fillings\_ds %>%  
 filter(Ticker == row\_item\_tic)   
 if (nrow(specific\_sca\_filings) > 0){  
 fund\_stock\_securities\_rating\_ds$litigated[fund\_stock\_securities\_rating\_ds$gvkey == row\_item\_gvkey] <- 1  
 specific\_sca\_filings\_max <- specific\_sca\_filings %>%  
 filter(!is.na(SettlementAmount)) %>%  
 arrange(SettlementAmount)  
   
 if (nrow(specific\_sca\_filings\_max) > 0){  
 last\_row <- tail(specific\_sca\_filings\_max, 1)  
 settlement\_amount <- as.numeric(last\_row$SettlementAmount)  
 fund\_stock\_securities\_rating\_ds$litigation\_settlement[fund\_stock\_securities\_rating\_ds$gvkey == row\_item\_gvkey] <- settlement\_amount  
 }  
 }  
}  
fund\_stock\_securities\_rating\_ds$litigated <- as.factor(fund\_stock\_securities\_rating\_ds$litigated)  
#summary(fund\_stock\_securities\_rating\_ds)

#colnames(fund\_stock\_securities\_rating\_ds)

fund\_stock\_securities\_rating\_ds\_final <- subset(fund\_stock\_securities\_rating\_ds, select = -c(aocidergl, aocipen, ceqt, cstkcv, dd1,  
 dpc,icapt, intan, intano, ivncf, ivst,  
 lo, lse, opeps, reajo, recta,  
 spi, tstkn, txp, txr,  
 restmt\_capx, restmt\_capx\_mag, restmt\_epspi, restmt\_epspi\_mag,  
 restmt\_ni, restmt\_ni\_mag, restmt\_pi, restmt\_pi\_mag,   
 restmt\_reuna, restmt\_reuna\_mag, restmt\_teq, restmt\_teq\_mag,   
 restmt\_txt, restmt\_txt\_mag, restmt\_wcap, restmt\_wcap\_mag  
  
 ))  
#summary(fund\_stock\_securities\_rating\_ds\_final)

cor\_matrix\_ds <- subset(fund\_stock\_securities\_rating\_ds\_final, select = -c(gvkey,tic, sp\_rating, litigated))  
cor\_matrix <- cor(cor\_matrix\_ds)  
cor\_matrix %>%  
 as.data.frame() %>%  
 mutate(var1 = rownames(.)) %>%  
 gather(var2, value, -var1) %>%  
 arrange(desc(value)) %>%  
 group\_by(value) %>%  
 filter(row\_number() == 1)

## # A tibble: 1,539 x 3  
## # Groups: value [1,539]  
## var1 var2 value  
## <chr> <chr> <dbl>  
## 1 aco aco 1   
## 2 prcld\_m prchd\_m 0.995  
## 3 ni ci 0.990  
## 4 lt at 0.981  
## 5 revt cogs 0.979  
## 6 ppegt dpact 0.972  
## 7 teq ceq 0.953  
## 8 oancf ebit 0.943  
## 9 ppegt capx 0.939  
## 10 teq at 0.937  
## # ... with 1,529 more rows

nrow(fund\_stock\_securities\_rating\_ds\_final)

## [1] 2459

summary(fund\_stock\_securities\_rating\_ds\_final)

## gvkey tic aco acominc   
## Min. : 1076 0160A : 1 Min. : 0.000 Min. :-19306.573   
## 1st Qu.: 18060 0161A : 1 1st Qu.: 0.433 1st Qu.: -0.584   
## Median : 63516 0170A : 1 Median : 3.498 Median : 0.000   
## Mean : 88413 0173A : 1 Mean : 121.016 Mean : -77.565   
## 3rd Qu.:165053 3ADMA : 1 3rd Qu.: 30.126 3rd Qu.: 0.018   
## Max. :314866 3AEGA : 1 Max. :13859.200 Max. : 3495.339   
## (Other):2453   
## act ao aoloch ap   
## Min. : 0.00 Min. : 0.000 Min. :-5998.250 Min. : 0.00   
## 1st Qu.: 9.57 1st Qu.: 0.162 1st Qu.: -0.940 1st Qu.: 1.13   
## Median : 69.46 Median : 2.664 Median : 0.008 Median : 5.87   
## Mean : 1225.47 Mean : 180.174 Mean : -0.697 Mean : 330.39   
## 3rd Qu.: 365.18 3rd Qu.: 34.485 3rd Qu.: 1.899 3rd Qu.: 52.43   
## Max. :146290.60 Max. :30582.257 Max. : 4914.600 Max. :35222.20   
##   
## aqc at bkvlps caps   
## Min. :-111.000 Min. : 0.0 Min. : -82001.3 Min. : -701.48   
## 1st Qu.: 0.000 1st Qu.: 20.0 1st Qu.: 0.0 1st Qu.: 18.62   
## Median : 0.000 Median : 150.4 Median : 2.2 Median : 94.24   
## Mean : 60.962 Mean : 3718.2 Mean : 1170.1 Mean : 744.56   
## 3rd Qu.: 6.071 3rd Qu.: 1086.5 3rd Qu.: 8.8 3rd Qu.: 347.06   
## Max. :5559.018 Max. :367278.8 Max. :1216746.9 Max. :155311.75   
##   
## capx ceq ch che   
## Min. : -3.325 Min. : -9528.50 Min. : 0.000 Min. : 0.00   
## 1st Qu.: 0.210 1st Qu.: 2.05 1st Qu.: 3.285 1st Qu.: 3.76   
## Median : 3.176 Median : 55.51 Median : 21.818 Median : 29.49   
## Mean : 133.683 Mean : 1252.32 Mean : 301.370 Mean : 423.87   
## 3rd Qu.: 32.301 3rd Qu.: 381.85 3rd Qu.: 103.515 3rd Qu.: 134.90   
## Max. :20279.200 Max. :126661.20 Max. :26204.345 Max. :39531.20   
##   
## chech ci cogs cshi   
## Min. :-1606.500 Min. :-4911.696 Min. : 0.0 Min. : 0.00   
## 1st Qu.: -0.792 1st Qu.: -8.385 1st Qu.: 6.5 1st Qu.: 17.47   
## Median : 0.448 Median : 0.000 Median : 64.6 Median : 39.38   
## Mean : 17.731 Mean : 206.727 Mean : 2227.1 Mean : 148.83   
## 3rd Qu.: 8.657 3rd Qu.: 40.114 3rd Qu.: 587.1 3rd Qu.: 91.28   
## Max. : 5879.489 Max. :16929.062 Max. :325065.8 Max. :8946.25   
##   
## cstk dlc dltt dm   
## Min. : 0.000 Min. : -56.00 Min. : 0.00 Min. : 0.00   
## 1st Qu.: 0.023 1st Qu.: 0.05 1st Qu.: 0.00 1st Qu.: 0.00   
## Median : 0.160 Median : 1.82 Median : 4.25 Median : 0.27   
## Mean : 82.690 Mean : 235.69 Mean : 879.30 Mean : 182.41   
## 3rd Qu.: 1.539 3rd Qu.: 17.19 3rd Qu.: 201.80 3rd Qu.: 22.42   
## Max. :10591.097 Max. :70169.80 Max. :77456.60 Max. :26721.95   
##   
## dn dpact dvt ebit   
## Min. : 0.0 Min. : 0.00 Min. : -12.481 Min. : -425.200   
## 1st Qu.: 0.0 1st Qu.: 1.39 1st Qu.: 0.000 1st Qu.: -4.154   
## Median : 0.0 Median : 15.70 Median : 0.000 Median : 3.639   
## Mean : 522.0 Mean : 747.43 Mean : 79.456 Mean : 334.014   
## 3rd Qu.: 8.3 3rd Qu.: 171.69 3rd Qu.: 2.812 3rd Qu.: 90.028   
## Max. :62177.0 Max. :120589.20 Max. :6572.535 Max. :24345.400   
##   
## epspi fiao fincf   
## Min. :-154.662 Min. :-5158.161 Min. :-15688.750   
## 1st Qu.: -0.339 1st Qu.: -2.531 1st Qu.: -15.535   
## Median : 0.000 Median : -0.048 Median : 0.473   
## Mean : 3.330 Mean : -18.559 Mean : -114.928   
## 3rd Qu.: 1.052 3rd Qu.: 0.000 3rd Qu.: 13.290   
## Max. :6710.943 Max. : 3025.988 Max. : 10025.261   
##   
## fopo gp invt lt   
## Min. :-8039.000 Min. : -240.82 Min. : 0.00 Min. : 0.00   
## 1st Qu.: 0.556 1st Qu.: 0.58 1st Qu.: 0.00 1st Qu.: 8.59   
## Median : 3.704 Median : 43.04 Median : 3.39 Median : 55.16   
## Mean : 47.902 Mean : 1127.46 Mean : 335.93 Mean : 2351.46   
## 3rd Qu.: 16.667 3rd Qu.: 380.34 3rd Qu.: 69.92 3rd Qu.: 576.27   
## Max. :13384.000 Max. :117445.60 Max. :39770.60 Max. :267504.27   
##   
## ni nopi oancf   
## Min. :-4398.397 Min. :-2536.272 Min. : -339.304   
## 1st Qu.: -8.385 1st Qu.: -0.019 1st Qu.: -2.463   
## Median : -0.025 Median : 0.103 Median : 4.985   
## Mean : 206.043 Mean : 22.924 Mean : 337.897   
## 3rd Qu.: 37.759 3rd Qu.: 1.496 3rd Qu.: 92.826   
## Max. :17546.884 Max. :15531.549 Max. :26182.400   
##   
## ppegt re rect revt   
## Min. : 0.00 Min. :-90811.25 Min. : 0.00 Min. : 0.0   
## 1st Qu.: 2.56 1st Qu.: -90.48 1st Qu.: 0.56 1st Qu.: 6.6   
## Median : 34.10 Median : -7.82 Median : 9.62 Median : 112.5   
## Mean : 1554.96 Mean : 755.66 Mean : 443.71 Mean : 3354.6   
## 3rd Qu.: 368.78 3rd Qu.: 105.04 3rd Qu.: 86.19 3rd Qu.: 1011.1   
## Max. :194731.60 Max. :129492.20 Max. :81411.50 Max. :442511.4   
##   
## siv sppiv sstk   
## Min. : 0.00 Min. :-6191.874 Min. : -3.6837   
## 1st Qu.: 0.00 1st Qu.: -0.019 1st Qu.: 0.1295   
## Median : 0.00 Median : 0.000 Median : 3.0452   
## Mean : 211.85 Mean : -8.672 Mean : 29.0287   
## 3rd Qu.: 0.47 3rd Qu.: 0.012 3rd Qu.: 20.3688   
## Max. :130595.75 Max. : 393.600 Max. :3092.3460   
##   
## teq tstk txt wcap   
## Min. : -8234.00 Min. : 0.00 Min. :-8035.500 Min. :-8236.80   
## 1st Qu.: 4.74 1st Qu.: 0.00 1st Qu.: 0.000 1st Qu.: 0.09   
## Median : 62.22 Median : 0.00 Median : 0.129 Median : 24.95   
## Mean : 1353.70 Mean : 347.21 Mean : 71.163 Mean : 297.94   
## 3rd Qu.: 408.72 3rd Qu.: 2.56 3rd Qu.: 14.244 3rd Qu.: 138.60   
## Max. :133340.00 Max. :67539.25 Max. : 7749.600 Max. :31798.00   
##   
## xint restmt\_at restmt\_at\_mag   
## Min. : 0.000 Min. :0.0000000 Min. :-6.700e-02   
## 1st Qu.: 0.080 1st Qu.:0.0000000 1st Qu.: 0.000e+00   
## Median : 1.099 Median :0.0000000 Median : 0.000e+00   
## Mean : 50.602 Mean :0.0004067 Mean :-2.725e-05   
## 3rd Qu.: 14.705 3rd Qu.:0.0000000 3rd Qu.: 0.000e+00   
## Max. :4525.000 Max. :1.0000000 Max. : 0.000e+00   
##   
## restmt\_cogs restmt\_dltt restmt\_dltt\_mag   
## Min. :0.0000000 Min. :0.0000000 Min. : 0.000000   
## 1st Qu.:0.0000000 1st Qu.:0.0000000 1st Qu.: 0.000000   
## Median :0.0000000 Median :0.0000000 Median : 0.000000   
## Mean :0.0004067 Mean :0.0004067 Mean : 0.007324   
## 3rd Qu.:0.0000000 3rd Qu.:0.0000000 3rd Qu.: 0.000000   
## Max. :1.0000000 Max. :1.0000000 Max. :18.010500   
##   
## restmt\_nopi restmt\_nopi\_mag cshtrd\_m   
## Min. :0.0000000 Min. :-194.41200 Min. : 0   
## 1st Qu.:0.0000000 1st Qu.: 0.00000 1st Qu.: 33569   
## Median :0.0000000 Median : 0.00000 Median : 163147   
## Mean :0.0004067 Mean : -0.07906 Mean : 821980   
## 3rd Qu.:0.0000000 3rd Qu.: 0.00000 3rd Qu.: 613083   
## Max. :1.0000000 Max. : 0.00000 Max. :66661743   
##   
## prchd\_m prcld\_m prcod\_m trfd\_m   
## Min. : 0.0002 Min. : 0.0002 Min. : 0.0043 Min. : 1.000   
## 1st Qu.: 1.8324 1st Qu.: 1.6958 1st Qu.: 3.7487 1st Qu.: 1.030   
## Median : 8.3980 Median : 7.9390 Median : 11.1434 Median : 1.153   
## Mean : 20.9993 Mean : 20.1028 Mean : 25.4425 Mean : 2.479   
## 3rd Qu.: 25.2388 3rd Qu.: 24.3799 3rd Qu.: 27.4553 3rd Qu.: 1.511   
## Max. :2250.8331 Max. :2183.9480 Max. :2217.8208 Max. :597.912   
##   
## trfm\_m dvrate\_m sp\_rating litigated  
## Min. : 1.000 Min. :0.0050 DECREASED: 104 0:2034   
## 1st Qu.: 1.000 1st Qu.:0.2956 INCREASED: 219 1: 425   
## Median : 1.000 Median :0.5933 NoCHANGE : 239   
## Mean : 1.582 Mean :0.8036 NOTRATED :1897   
## 3rd Qu.: 1.077 3rd Qu.:1.0000   
## Max. :120.892 Max. :9.5391   
## NA's :18 NA's :1866   
## litigation\_settlement  
## Min. : 1000000   
## 1st Qu.: 3650000   
## Median : 8000000   
## Mean : 30566316   
## 3rd Qu.: 16000000   
## Max. :925500000   
## NA's :2364