library(tidyverse)

library(dplyr)

library(ggplot2)

library(lubridate)

library(randomForest)

library(plyr)

library(readr)

library(janitor)

library(purrr)

#set working directory (where your files are)

setwd("~/Documents/Regals\_Project")

# 1. Import sector data

#2020 sector file is added to this folder

nsesectors <- read.csv(file = 'NSE\_data\_stock\_market\_sectors\_2020.csv')

View(nsesectors)

# 2. - Import NSE Data

#Folder with NSE Data is added to this folder

NSEfolder <- "NSE\_Data"

nsefiles <- list.files(path = NSEfolder, pattern = "\*.csv", full.names=TRUE)

nsefilesv2 <- nsefiles %>%

map(read\_csv) %>%

reduce(rbind)

View(nsefilesv2)

write\_csv(nsefilesv2,"nsefilesv2.csv")

#Change Date format

nsefilesv2$Date2 = format(as.Date(nsefilesv2$DATE, "%d-%b-%y"), "%d-%m-%Y")

nsefilesv2$Date3 <- lubridate::dmy(nsefilesv2$Date2)

View(nsefilesv2)

#Step 2.1 - Join Sector Data to NSE Stock Price Data

nsefilesv3 <- merge(x=nsefilesv2, y=nsesectors, by = c("CODE" ), all = TRUE)

View(nsefilesv3)

#Omit Null Values

nsefilesv4 <- na.omit(nsefilesv3)

View(nsefilesv4)

#Write file to working Directory

write\_csv(nsefilesv4,"nsefilesv4.csv")

#Clean Up column names if necessary

#names(nsefilesv4) <- gsub(" ", "\_", names(nsefilesv4))

#subset to include only the company(s) of interest e.g. Kakuzi

nsefilesv5 <- subset(nsefilesv4, subset = nsefilesv4$CODE == "KUKZ" )

View(nsefilesv5)

#subset to include only relevant dates

nsefilesv6 <- subset(nsefilesv5, subset = nsefilesv5$Date3 >= "2019-01-01" )

View(nsefilesv6)

nsefilesv6$month = format(nsefilesv6$Date3,"%m")

nsefilesv6$year = format(nsefilesv6$Date3,"%Y")

nsefilesv6$monthyear = paste(nsefilesv6$year,nsefilesv6$month )

#plot some data

qplot(nsefilesv6$Previous, geom = "histogram")

nsefilesv7 <- ggplot(nsefilesv6,aes(monthyear ,y=Previous, group=1)) + geom\_jitter() + geom\_smooth(model = "lm")

print (d2)