#retail analysis with walmart data

#importing the data frames

walmart<- read.csv('Walmart\_Store\_sales.csv'stringsAsFactorsx=FALSE)

View(walmart)

#no of rows and columns

dim(walmart)

#str of the var

str(walmart)

install.packages('lubridate')

library(lubridate)

library(dplyr)

walmart <- walmart%>%mutate(new\_dt=dmy(Date))

str(walmart)

summary(walmart)

#aggregate the weekely sales by store

walmart %>% group\_by(walmart) %>% summarise(aggr\_sales = sum(Weekly\_Sales))

library(dplyr)

walmart %>% group\_by(Store) %>% summarise(aggr\_sales = sum(Weekly\_Sales))

install.packages('fansi')

library(fansi)

walmart %>% group\_by(Store) %>% summarise(aggr\_sales = sum(Weekly\_Sales))

install.packages('fansi',dependencies=TRUE)

library(fansi)

walmart %>% group\_by(Store) %>% summarise(agge\_sales = sum(Weekly\_Sales))

library(dplyr)

walmart %>% group\_by(Store) %>% summarise(agge\_sales = sum(Weekly\_Sales))

install.packages('utf8',dependencies=TRUE)

library(utf8)

library(dplyr)

walmart %>% group\_by(Store) %>% summarise(agge\_sales = sum(Weekly\_Sales))

df <- walmart %>% group\_by(Store) %>% summarise(avg\_sales =mean(Weekly\_Sales),sd\_sales=sd(Weekly\_Sales))

df

df%>%mutate(cov=(avg\_Sales/sd\_Sales)\*100)

#quarterly growth rate

#date

install.packages('lubridate')

library(lubridate)

#create quarter, year, semester variables

walmart <- walmart%>%mutate(yr=year(new\_dt),mon=month(new\_dt),qtr=quarter(new\_dt),sem=semester(new\_dt))

View(walmart)

walmart <- walmart %>% mutate(yr\_qtr=paste(yr,"Q",qtr,sep=""))

walmart

df2 <- walmart%>%group\_by(Store,yr\_qtr)%>% summarise(qtr\_Sales=sum(Weekly\_Sales))

df2

View(df2)

df3 <- df2 %>% mutate(qoq\_rate=(qtr\_Sales/lag(qtr\_Sales,4)-1)\*100)

df3

View(df3)

df4 <- df3 %>%filter(yr\_qtr=="2012Q3")%>%arrange(desc(qoq\_rate))

df4

View(df4)

#mean sales on holiday

walmart%>%filter(Holiday\_Flag==0)%>%summarise(avg\_Sales=mean(Weekly\_Sales))

walmart%>%filter(Holiday\_Flag==1,Weekly\_Sales>1041256)

#aggregate the sales monthly and semester wise

fact\_dt=as.factor(walmart$new\_dt)

levels(fact\_dt)

uniq\_date <- data.frame(new\_dt1=levels(as.factor(walmart$new\_dt)))

View(uniq\_date)

uniq\_date$new\_dt=ymd(uniq\_date$new\_dt1)

uniq\_date$r\_1=row\_number(uniq\_date)

View(uniq\_date)

#merge this with the original df

walmart\_new <- inner\_join(walmart,uniq\_date,by='new\_dt')

View(walmart\_new)

walmart$Holiday\_Flag=as.factor(walmart$Holiday\_Flag)

library('ggplot2')

ggplot(data=walmart,aes(x=CPI,y=Weekly\_Sales))+geom\_point(aes(color=Holiday\_Flag))

walmart\_new <- walmart\_new%>%mutate(days=day(new\_dt))

#create a lm model

lin\_model <- lm(data=walmart\_new,Weekly\_Sales~Holiday\_Flag+Unemployment+CPI)

summary(lin\_model)