

```
### Uploading and Reading the Dataset
```

```
walmart<- read.csv("Walmart_Store_sales.csv",header = TRUE)
View(walmart)
str(walmart)
head(walmart)
class(walmart)
```

```
### Descriptive Statistics
```

```
summary(walmart)
```

```
### Checking NA values
```

```
colSums(is.na(walmart))
```

```
## No null values in the dataset
```

```
### Loading all the needed libraries
```

```
library("dplyr")
library("lubridate")
library("zoo")
```

```
#Data Visualization
```

```
library("grid")
library("vcd")
library("ggplot2")
library("plotly")
```

```
## Converting Date column into Date format
```

```
walmart$Date = as.Date(walmart$Date,format="%d-%m-%Y")
```

```
### Q1- which store has max sales?
```

```
Store_sales <-aggregate(Weekly_Sales~Store, data = walmart, sum)
Store_sales
which.max(Store_sales$Weekly_Sales)
```

```
## Store 20 has max sale , Ammount-301397792
```

```
## Q2- Which store has maximum standard deviation
```

```
Store_sales$sales_mean <- aggregate(Weekly_Sales~Store,data= walmart, mean)$Weekly_Sales
Store_sales$sales_sd <- aggregate(Weekly_Sales~Store, data=walmart,sd)$Weekly_Sales
str(Store_sales)
arrange(Store_sales, desc(sales_sd))
```

```
### Store 14 has highest standard deviatiation: 317569.95
```

```
## Q3- Which store/s has good quarterly growth rate in Q3'2012?
# creating copy of Walmart
```

```
walmart2= walmart
```

```
walmart2$month_year <-substr(walmart2$Date,1,7)
Q3_2012 <- filter(walmart2,month_year=="2012-07"| month_year=="2012-08"|month_year=="2012-09")
Q2_2012<- filter(walmart2,month_year=="2012-04"|month_year=="2012-05"|month_year=="2012-06")
```

```

Q3_2012_sales <- summarise(group_by(Q3_2012,Store),sum(Weekly_Sales))

Q2_2012_sales<- summarise(group_by(Q2_2012,Store),sum(Weekly_Sales))

Q3_2012_Growthrate = merge ( Q2_2012_sales , Q3_2012_sales , by = 'Store')
Q3_2012_Growthrate = mutate(Q3_2012_Growthrate, Growth_Rate =
((Q3_2012_sales$`sum(Weekly_Sales)` - Q2_2012_sales$'sum(Weekly_Sales)')*100) /
Q2_2012_sales$'sum(Weekly_Sales)')
gr = arrange(Q3_2012_Growthrate, desc(Growth_Rate))
View(gr)

### Store 15 has highest growth rate in Q3 2012: 13.3307760

###Find out holidays which have higher sales than the mean sales in non-holiday season for
all stores together

SuperBowl = as.Date(c("2010-02-12","2011-02-11","2012-02-10","2013-02-08"))
LabourDay = as.Date(c("2010-09-10", "2011-09-09", "2012-09-07", "2013-09-06"))
Thanksgiving = as.Date(c("2010-11-26", "2011-11-25", "2012-11-23", "2013-11-29"))
Christmas = as.Date(c("2010-12-31", "2011-12-30", "2012-12-28", "2013-12-27"))

Walmart_Holiday = walmart[1:3]
Walmart_Holiday$hflag = ifelse(Walmart_Holiday$Date %in% SuperBowl, "SB",
ifelse(Walmart_Holiday$Date %in% LabourDay, "LD", ifelse(Walmart_Holiday$Date %in%
Thanksgiving, "TG", ifelse(Walmart_Holiday$Date %in% Christmas, "CH","None"))))
aggregate(Weekly_Sales~hflag,data=Walmart_Holiday, mean)

### Thanks giving have highest sales than mean. Mean sales in non-holiday season for all
stores together is 1041256.4 and except Christmas all holidays have higher sales than
average sale in non-holiday sale.

### For Store 1 - Build prediction models to forecast demand

library(dplyr)

semester_store1 = select(filter(walmart, Store==1),-1)
View(semester_store1)
str(semester_store1)
head(semester_store1)

## Linear Model

Walmart_lm = lm(Weekly_Sales ~ Holiday_Flag + Temperature + Fuel_Price+ CPI + Unemployment
, semester_store1)
summary(Walmart_lm)

walmart_lm1 = lm(Weekly_Sales ~ Holiday_Flag + Temperature ++ CPI , semester_store1)
summary(walmart_lm1)

walmart_lm3 = lm(Weekly_Sales ~ Temperature + CPI , semester_store1)
summary(walmart_lm3)

### Change dates into days by creating new variable

Data2= walmart
Data2$Weekdays = weekdays(Data2$Date)
View(Data2)

```