# author - Anupama Rajaram

# Date - Dec 15, 2016

# Description - Code for pricing analysis. Analyze Facebook stock price and create linear regression model.

# ====================================

# Setup the environment

# ====================================

# Start by cleaning up memory of current R session:

rm(list=ls(all=TRUE))

# suppress scientific notation e.g.: 2.4e5 instead 2,400,367

options(scipen = 999)

# decimal places - only 4

options(digits = 4)

# load required library packages:

library(data.table) # for fread()

library(sqldf)

library(plotly)

library('ggplot2') # plotting data

library(quantmod) # get stock prices; useful stock analysis functions

library(xts) # working with extensible time series

library(tidyverse) # ggplot2, purrr, dplyr, tidyr, readr, tibble

library(stringr) # working with strings

# ====================================

# load original dataset

# ====================================

fbkdf = data.frame(fread("fb\_2004\_to\_2016.csv", stringsAsFactors = FALSE))

sp500df = data.frame(fread("sp500\_2004\_to\_2016.csv", stringsAsFactors = FALSE))

sp5k = sp500df[,c(1:2)]

colnames(sp5k) = c("Date", "SP500\_price")

# merge sp500 and fbkdf to benchmark against S&P500.

fbknew = merge(fbkdf, sp5k, by = "Date", all.x = TRUE)

rm(sp500df, sp5k)

# ====================================

# Create derived variables

# ====================================

# calculate daily volatility

fbknew$prc\_change = fbknew$Close - fbknew$Open

# date calculation

fbknew$Date2 = as.Date(fbknew$Date, "%Y-%m-%d")

fbknew$mthyr = paste(as.numeric(format(fbknew$Date2, "%m")), "-",

as.numeric(format(fbknew$Date2, "%Y")), sep = "")

fbknew$mth = as.numeric(format(fbknew$Date2, "%m"))

fbknew$year = as.numeric(format(fbknew$Date2, "%Y"))

# calculate volatility from prev day:

# first order by date

fbknew <- fbknew[order(fbknew$Date2),]

# add a factor for volatility to show change from previous day

fbknew$volt\_chg = 0

for( i in 2:nrow(fbknew)){

fbknew[i, "volt\_chg"] = fbknew[i, "Open"] - fbknew[i-1, "Open"]

}

# ====================================

# Some visualizations

# ====================================

fbkmthdata = sqldf("select avg(Open) as 'Open',

avg(High) as 'High', avg(Low) as 'Low' ,

avg(Close) as 'Close' , avg(Volume) as 'Volume',

mthyr, year, mth

from fbknew

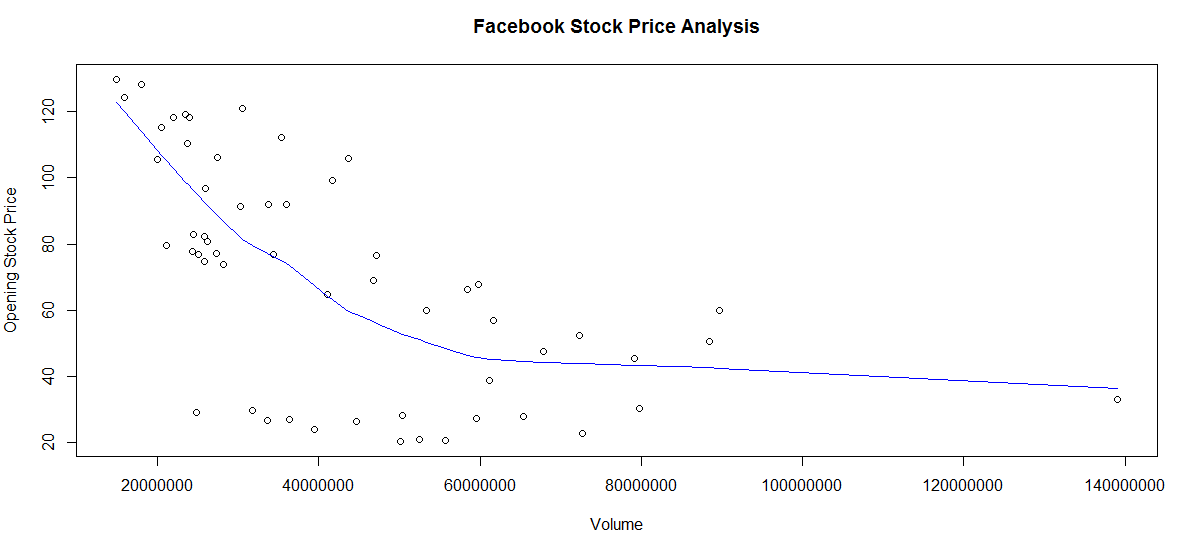
group by mthyr")

# graph for Facebook Stock Price by Volume

plot(fbkmthdata$Volume, fbkmthdata$Open, xlab = "Volume",

ylab = "Opening Stock Price", main = "Facebook Stock Price Analysis")

lines(lowess(fbkmthdata$Volume, fbkmthdata$Open), col="blue")



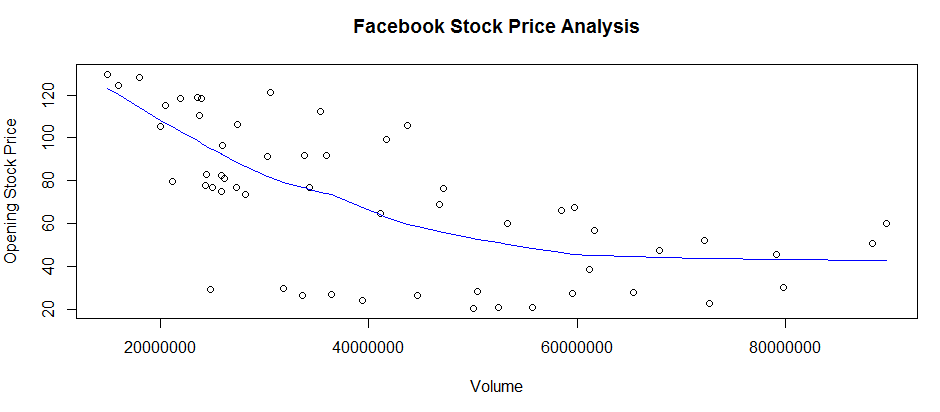
data1 = subset(fbkmthdata, Volume <= 120000000)

# graph for Facebook Stock Price by Volume

plot(data1$Volume, data1$Open, xlab = "Volume",

ylab = "Opening Stock Price", main = "Facebook Stock Price Analysis")

lines(lowess(data1$Volume, data1$Open), col="blue")



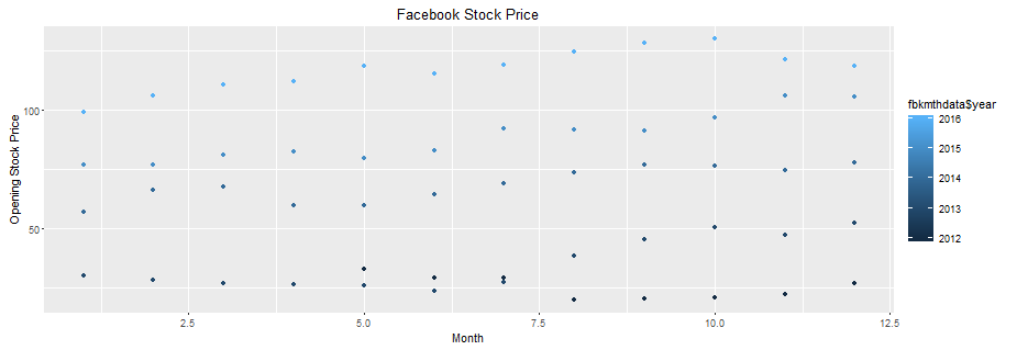
# graph for Stock Price variation by Month and Year

ggplot(fbkmthdata, aes(fbkmthdata$mth, fbkmthdata$Open,

color = fbkmthdata$year))+geom\_point() +

labs(title = "Facebook Stock Price") + xlab("Month") +

ylab( "Opening Stock Price")



# ====================================

# Linear Regression Model

# ====================================

lmodelfb = lm(Open ~ High + Volume + SP500\_price + prc\_change +

mthyr + mth + year + volt\_chg,

data = fbknew)

summary(lmodelfb)

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Output of Linear model

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Call:

lm(formula = Open ~ High + Volume + SP500\_price + prc\_change +

mthyr + mth + year + volt\_chg, data = fbknew)

Residuals:

Min 1Q Median 3Q Max

-2.3790 -0.1743 0.0147 0.2144 1.7985

Coefficients: (2 not defined because of singularities)

Estimate Std. Error t value

(Intercept) -1.66945551409 0.66624643356 -2.51

High 0.97801703854 0.00550464762 177.67

Volume -0.00000000688 0.00000000044 -15.63

SP500\_price 0.00156703762 0.00048151430 3.25

prc\_change -0.47698242483 0.00999426772 -47.73

mthyr1-2014 -0.41313035410 0.20665323521 -2.00

mthyr1-2015 -0.42629218241 0.30730741600 -1.39

mthyr10-2016 0.52566901750 0.52229829064 1.01

mthyr11-2012 -0.05208958149 0.12501702602 -0.42

mthyr2-2016 -0.55308551478 0.40548046199 -1.36

mthyr4-2014 -0.52614571008 0.22220063739 -2.37

mthyr5-2016 0.39859112992 0.46591768800 0.86

mthyr6-2012 -0.21858972318 0.14196103908 -1.54

mthyr7-2016 0.15957335364 0.47633629248 0.34

mthyr8-2014 -0.13206437001 0.28132131150 -0.47

mthyr9-2014 -0.36189047760 0.29836592189 -1.21

mth NA NA NA

year NA NA NA

volt\_chg 0.04619227682 0.00769325998 6.00

Pr(>|t|)

(Intercept) 0.0124 \*

High < 0.0000000000000002 \*\*\*

Volume < 0.0000000000000002 \*\*\*

SP500\_price 0.0012 \*\*

prc\_change < 0.0000000000000002 \*\*\*

mthyr1-2014 0.0458 \*

mthyr11-2013 0.0033 \*\*

mthyr3-2013 0.0151 \*

mthyr4-2013 0.0026 \*\*

mthyr4-2014 0.0181 \*

mthyr5-2013 0.0089 \*\*

mthyr5-2014 0.0179 \*

mthyr6-2013 0.0069 \*\*

mthyr7-2013 0.0466 \*

mthyr8-2013 0.0994 .

volt\_chg 0.0000000026 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.38 on 1093 degrees of freedom

Multiple R-squared: 1, Adjusted R-squared: 1

F-statistic: 1.54e+05 on 60 and 1093 DF, p-value: <0.0000000000000002

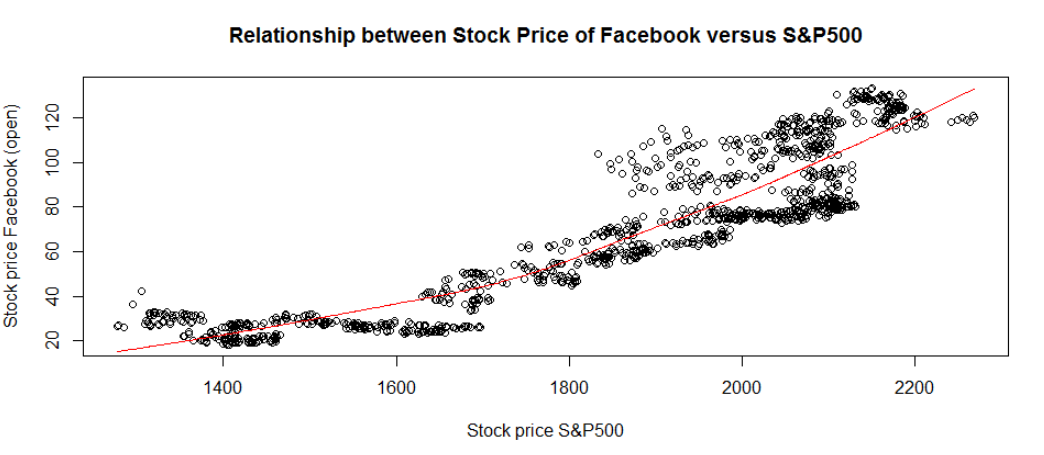
# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

plot(fbknew$SP500\_price, fbknew$Open,

xlab = "Stock price S&P500",

ylab = "Stock price Facebook (open)",

main = "Relationship between Stock Price of Facebook versus S&P500") lines(lowess(fbknew$SP500\_price, fbknew$Open), col="red")

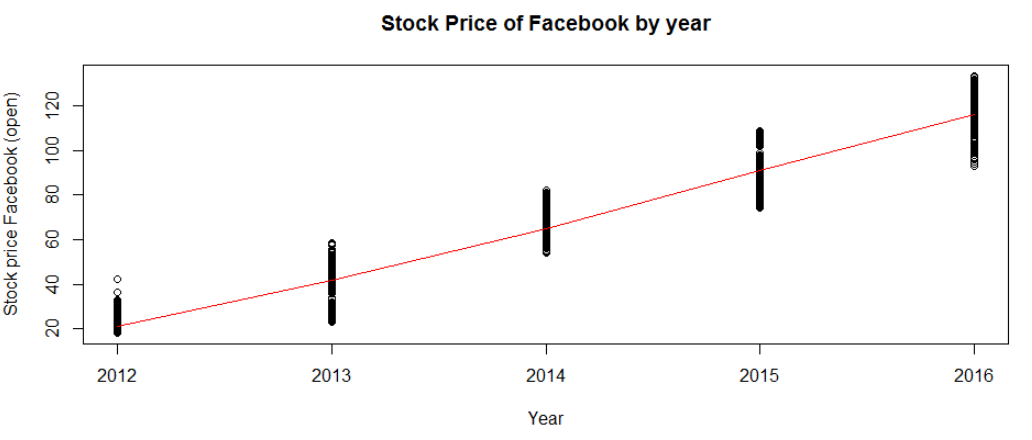


plot(fbknew$year, fbknew$Open,

xlab = "Year",

ylab = "Stock price Facebook (open)",

main = "Stock Price of Facebook by year")

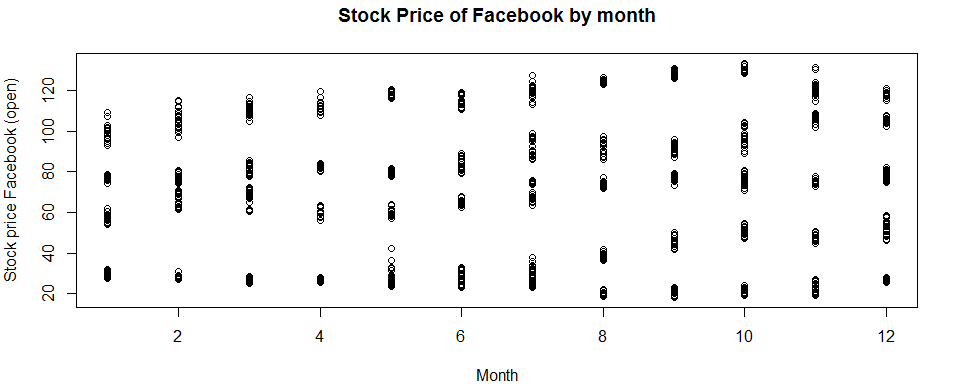
lines(lowess(fbknew$year, fbknew$Open), col="red")

plot(fbknew$mth, fbknew$Open,

xlab = "Month",

ylab = "Stock price Facebook (open)",

main = "Stock Price of Facebook by month")

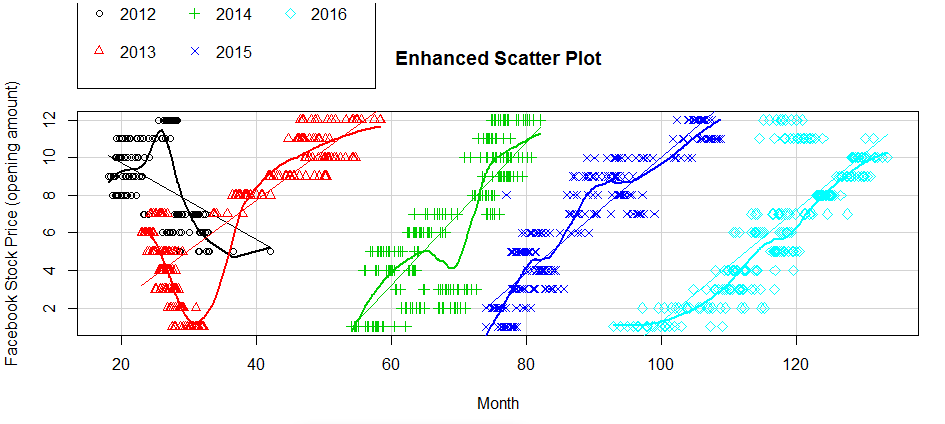


library(car)

scatterplot(fbknew$mth ~ fbknew$Open | fbknew$year,

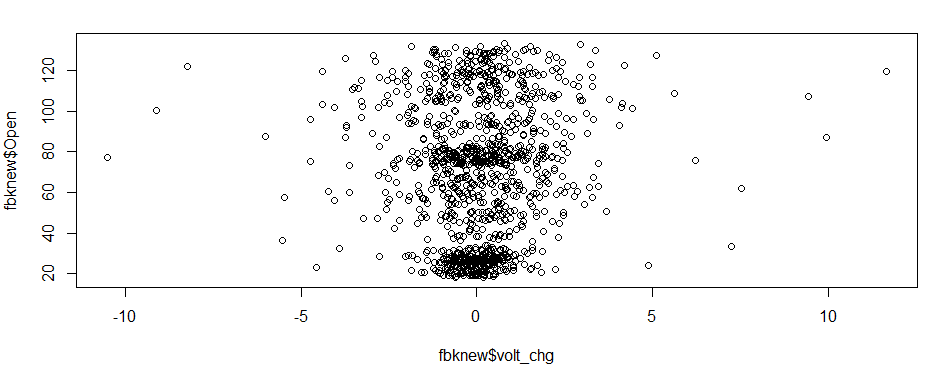
xlab="Month", ylab="Facebook Stock Price (opening amount)",

main="Enhanced Scatter Plot")



plot(fbknew$volt\_chg, fbknew$Open)

lines(lowess(fbknew$volt\_chg, fbknew$Open), col="red")



plot(data1$High, data1$Open, xlab = "Daily High",

ylab = "Opening Stock Price", main = "Facebook Stock Price Analysis")

lines(lowess(data1$High, data1$Open), col="blue")

