Does Investor attention impacts Stock Price and Volume

(A Study on one of the most valued company - Amazon)

Practicum 01

Surendra Kumar Dattatrey

(71710015) | CBA-08 | 27-April-2017

Table of Contents

[1. Abstract 3](#_Toc481069258)

[2. Business problem 3](#_Toc481069259)

[3. Introduction 3](#_Toc481069260)

[3.1 Background 3](#_Toc481069261)

[3.2 Purpose / Motivation 3](#_Toc481069262)

[3.3 Audience 4](#_Toc481069263)

[4. Significance of the study/Learning objective 4](#_Toc481069264)

[5. Research questions and hypothesis 4](#_Toc481069265)

[6. Methodology and Tools 4](#_Toc481069266)

[7. Data 4](#_Toc481069267)

[7.1 Data collection 4](#_Toc481069268)

[7.2 Data cleaning 5](#_Toc481069269)

[8. Analysis and Results 5](#_Toc481069270)

[9. Conclusion and recommendations 15](#_Toc481069272)

[10. Limitations 15](#_Toc481069273)

[11 References 15](#_Toc481069274)

[12. Appendix 1 16](#_Toc481069275)

[13. Appendix 2 17](#_Toc481069276)

# 1. Abstract

I have carried out a study by which I tried to figure out/check, if stock price & volume of any stock have any impact due to the investor attention. For this study, I took public data from Google trends (Week wise Popularity score) and the price – volumes data from Yahoo finance. I applied the univariate and bivariate analysis to check if there are any trends in these two datasets.

I considered only on one company data “**Amazon**” for sake of simplicity of this study in given time. By added tools we can definitely extend the study to include many other dimensions and make the findings more robust and can do many more analysis like, can do correlation study between multiple stocks etc. and their impact. Please refer section Conclusion and Recommendation for more detail.

This particular study assumes that an investor attention via web searches would impact the retail investors who have more sophisticated tools with them for investment into stock markets.

# 2. Business problem

By this study, I tried to find out the relationship between the equity stock and investor sentiments. Basically I want to see variation/changes on volume or Stock price, if investor had an attention on electronic platform such as web. (More precisely to check investor’s attention on electronic platform had any impact.)

# 3. Introduction

## 3.1 Background

I have carried out a study by which I tried to figure out/check, if stock price & volume of any stock have any impact due to the investor attention. For this purpose I collected data from secondary data sources.

## 3.2 Purpose / Motivation

By this study, I will get an understanding of changes in stock prices and volume when users discussed about any organization/company actively over internet. Also, with this analysis we can understand that what was the impact when users talks about any company, does it have any impact on the stock pricing (price goes down or up) or not.

Basically I am trying to find out an answer of below question:

*If people talks about any stocks over internet, does that had any impact on prices or volume. If yes then what?*

## 3.3 Audience

This study would be useful for the investors, Hedge funds, Investment banks etc. for decision making while buying these stocks or any contracts based on these stocks.

To represent this, I took **Amazon** (NYSE: AMZ) as the primary area for the study.

# 4. Significance of the study/Learning objective

In this study, I will get an understanding of changes in stock prices and volume when users discussed about any organization/company actively over internet.

# 5. Research questions and hypothesis

With this study I am trying to research the impact of Investor attention on stock price and volume. I would like figure out the answers of following questions:

1. Does Stock volume and Investor attention had any relation between, if yes what is that?

2. Does Stock price and Investor attention had any relation between, if yes what is that?

3. Does stock price changes when investor is more attentive?

# 6. Methodology and Tools

This study uses following methods to check relationship between users attention to Amazon, its stock price and stock volume over last 5 years.

1. I gathered weekly data (Open Price, Closing Price, Adj. Closed Price, Volume etc.) for the last 5 years.

2. I also gathered weekly data for the previous 5 years to get the trends from Google trends for US location only. For searching the Google Trends data, I used **“Amazon stock price”** keyword.

3. Univariate analysis for the numerical data like – Popularity score, Close Price and volumes.

4. Bivariate analysis for the combinations of Popularity score, Closing price and volume numerical data.

# 7. Data

## 7.1 Data collection

For data collection, I relied on secondary source of data which is available on Google tends and Yahoo Finance to do this study.

From Google Trends, I extract “Weekly Popularity Score” and from Yahoo Finance, I extract weekly “Stock price data”.

## 7.2 Data cleaning

Above 2 datasets are from the two different sources and merged into one .csv file for analysis purpose. I also received it as cleaned data from both sources, so one phase of data science; data cleaning was not at all required on both data sets.

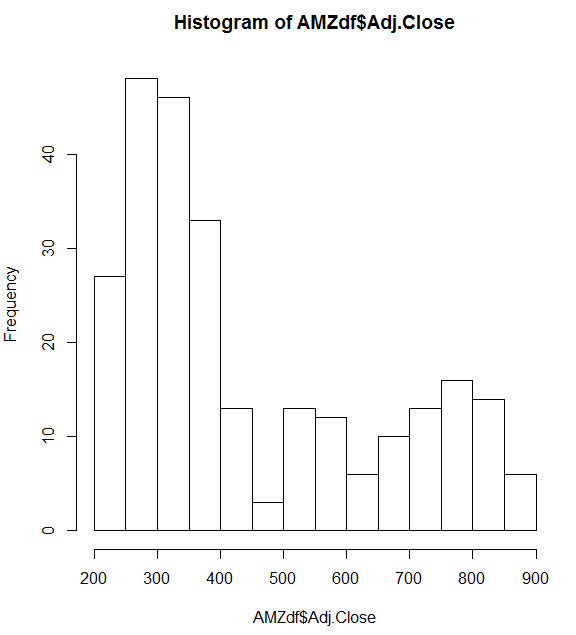
# 8. Analysis and Results

To get some sense from the variables, it’s important for us to understand these variables used in this study *viz*. Popularity Score, Adj. Close and volume.

**Univariate Analysis**

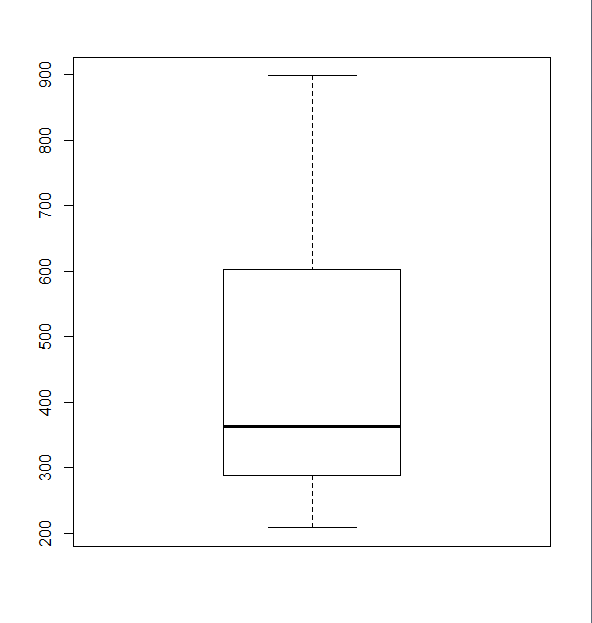
Adjusted Closing Price (Weekly Price):

As I worked on Adj. Closing Price, I also plotted another histogram (Figure 1) and boxplot (Figure 2) for “Adjusted close Price”.



**Figure1. Histogram for Adj. Closed Price**

As I can see the histogram is right skewed, this means that mean is far away from median on the right side, and this could be due some outliers. We can figure out the reason of skewedness creating a boxplot for this variable.



**Figure2. Boxplot for Adj. Closed Price**

As I can see from above box plot, Quartile 3 has more and higher values than other quartiles and due to that histogram is rightly skewed.

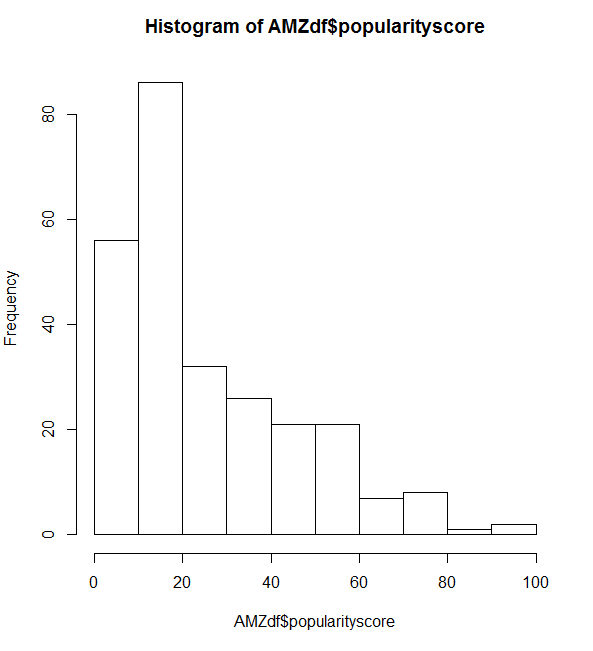
We also can interpret from this:

1. Mean of the Adj. Closed Price > Median of the Adj. Closed Price

2. Quartile 3 had more and higher value than Quartile 1.

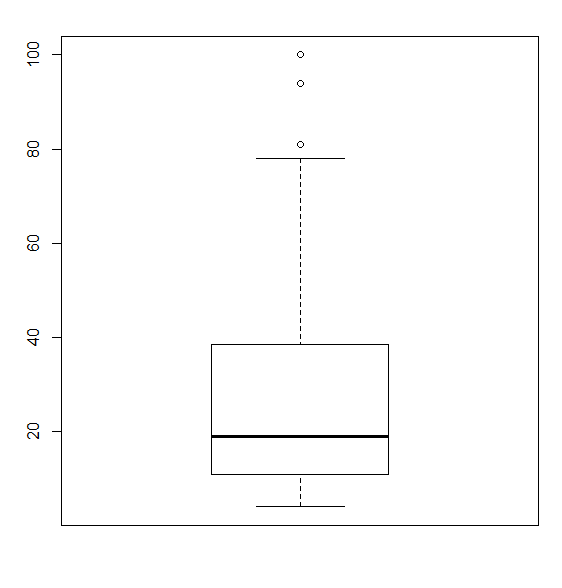
Popularity Score:

A simple histogram populated for the popularity score of Amazon for the past 5 years data and gets the below figure (figure 3). The data is right skewed with some outliers on the right end side of this distribution. Please check appendix 1 for R code and some statistics around this Popularity score data.



**Figure3. Histogram for Popularity Score**

As I can see the histogram is right skewed, this means that mean is far away from median on the right side, and this could be due some outliers. Cause of right skewedness, we can see in boxplot of this variable.



**Figure4. Box Plot for Popularity Score**

Above is box plot (Figure 4) for popularity score which shows a few outliers from score 80 to 100.

As I can see from above drawn box plot, Quartile 3 has more and higher values than other quartiles and due to that histogram is rightly skewed.

We also can interpret from this:

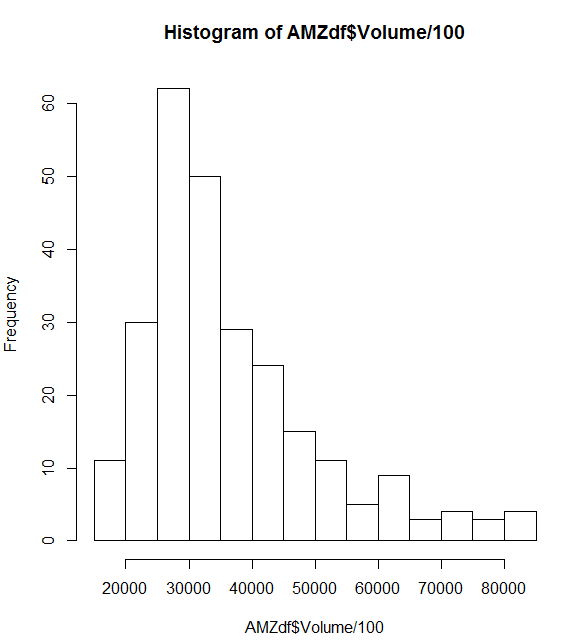
1. Mean of the Popularity score > Median of the Popularity score

2. Quartile 3 had more and higher value than Quartile 1.

Volume:

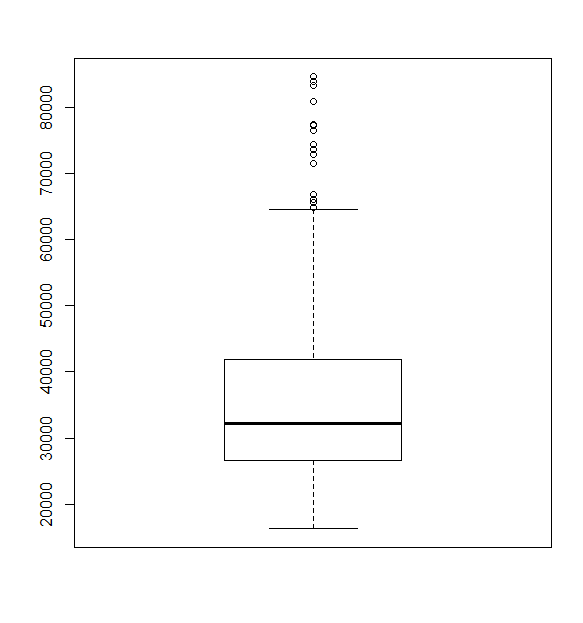
As I plotted Histogram and Box plot for above numerical fields (Stock Volume), I also plotted Histogram and boxplot for volumes as well.

For fitment purpose and better representation, I reduced scale of the X-axis by hundreds.



**Figure5. Histogram of Trade Volume**

As I can see this particular histogram is right skewed, this means that mean is more on the right side than median, and this could be due some outliers.



**Figure6. Boxplot for Volume**

As I can see from above box plot, Quartile 3 has large value followed by so much outliers and due to that histogram is rightly skewed.

As I can see from above drawn box plot, Quartile 3 has more and higher values than other quartiles and due to that histogram is rightly skewed.

We also can interpret from this:

1. Mean of Volume > Median of Volume

2. Quartile 3 had more and higher value than Quartile 1.

All above figures along with the details in Appendix 1 gives a fair idea of the distribution of the 3 variables. This is Univariate analysis which was performed on 3 parameters.

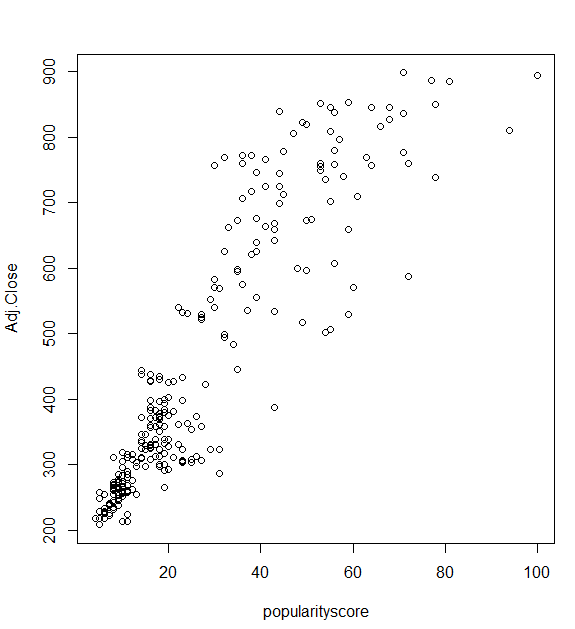
**Bivariate Analysis:**

Next we took the combination of 3 variables. By using 2 variables, we will perform bivariate analysis.

1. Popularity score – Adj. Close price and

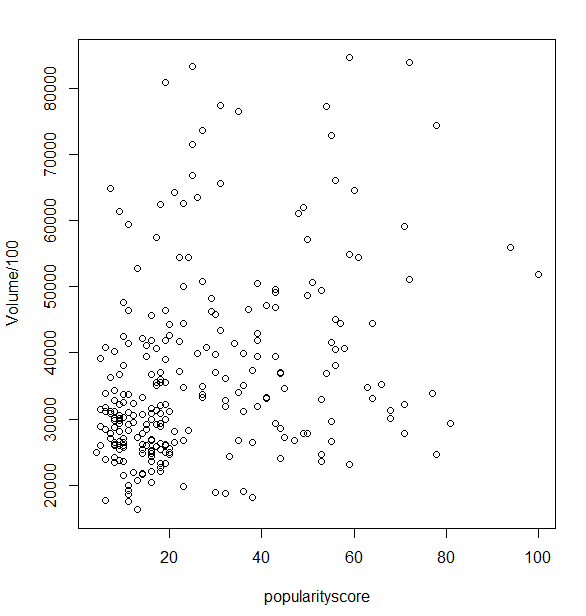
2. Popularity score – Stock volume

These analyses are to check, if there are any relations between these variables and we can check the correlation between these variables. Here we are primarily focusing that if Popularity Score had any impact on Stock Price etc.



**Figure7. Scatter Plot – Popularity Score vs Adj. Close Price**

Aha!!! While looking at the scatter plot plotted between Adj. Close prices and popularity score from google trends, looks like there is high correlation between these two variables. “**This means that when popularity score increases Close price also increase**”, so respective audience can take decision on the basis of this.



**Figure8. Scatterplot – Stock Volume vs Popularity Score**

After checking the relationship between volume and Popularity score, there seems to be some impact of the popularity score on the volumes of the stock traded during the corresponding week for Amazon, as we can see in above scatter plot. (However it’s not that highly correlated as compare to Closed Price and Popularity score.)

The correlation metric found in R corroborates the above findings (see Appendix 2 for R code and metrics).

# 9. Conclusion and recommendations

I tried to see the impact of the web search on the price and volume of Amazon with last 5 years of data, where data was segregated on weekly basis (weekly price and weekly trends).

By using various statistical analyses, I can conclude that there is an impact of the investor attention (measured here via trends score) on the volumes traded on the stock exchange for Amazon and the trends has significant impact on prices.

This was really simple study while keeping certain quantitative variables in mind. There could be so many other factors that can affect the pricing and volume of a stock. Also, this study cannot be used by prospective investor for the very naïve reason, as this study covers only few quantitative aspects of the correlation between prices and Trends. Apart from that there could be multiple factors that could affect the stock pricing and volume.

# 10. Limitations

This study only provides insight on the relations between investor attention and Price, and volume of Amazon. This is just first reference while looking at the Price-volume analysis.

This study is based on below facts/points:

1. This study covers only one company (Amazon) and doesn’t generalize the relationship with stocks of the same groups/category. (like – impact of Walmart stock to Amazon)
2. This study doesn’t classify much on volume, i.e. what is the proportion of Sell side and Buy side. It’s a generic volume number.
3. This study doesn’t check/covers any news/events and their impact on price movement.
4. Also, I performed this study over weekly data, so doesn’t cover daily changes information which would be more precise.

# 11. References

* <https://trends.google.co.in/trends/explore?q=Amazon%20stock%20price>
* <https://in.finance.yahoo.com/quote/AMZN/history?p=AMZN>

# 12. Appendix 1

getwd()

setwd("C:\\Users\\A022032\\Desktop\\Docs\\ISB\\Practicum")

install.packages('e1071')

library('e1071')

AMZdf <- read.csv('table.csv', stringsAsFactors = FALSE)

head(AMZdf)

str(AMZdf)

colnames(AMZdf) c("TrendingWeek","popularityscore","Open","High","Low","Close","Volume","Adj.Close")

# Univariate analyses for Adjusted Close Price weekly

mean(AMZdf$Adj.Close)

[1] 447.7382

median(AMZdf$Adj.Close)

[1] 362.745

hist(AMZdf$Adj.Close)

boxplot(AMZdf$Adj.Close)

# Univariate analyses for PopularityScore

mean(AMZdf$popularityscore)

[1] 26.79615

median(AMZdf$popularityscore)

[1] 19

hist(AMZdf$popularityscore)

boxplot(AMZdf$popularityscore)

# Univariate analyses for Volume weekly

mean(AMZdf$Volume)

[1] 3658953

median(AMZdf$Volume)

[1] 3225200

hist(AMZdf$Volume/100)

boxplot(AMZdf$Volume/100)

# 13. Appendix 2

#Bivariate analyses between Adj. Closing price and Popularity Score

plot( Adj.Close ~ popularityscore , data=AMZdf)

#Correlation matices between Adj. Closing price and Popularity Score

cor(AMZdf$Adj.Close , AMZdf$popularityscore)

[1] 0.9008006

#Bivariate analyses between Volume/Popularity Score

plot(Volume/100 ~ popularityscore , data=AMZdf)

#Correlation matices between Volume/Popularity Score

cor(AMZdf$popularityscore,AMZdf$Volume)

[1] 0.3410846