PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE17CS203

B.Tech, Sem III
Session : Aug-Dec, 2018

UE17CS203 – INTRODUCTION TO DATA SCIENCE

REPORT

EXPLORATORY ANALYSIS ON GAFA STOCK PRICES

DATA SET LINK:	https://www.kaggle.com/stexo92/gafa-stock-prices
TEAM MEMBERS	**NAME : Richa** **SRN : PES1201700688** **EMAIL ID : richa5june99@gmail.com** **CONTACT NO. : 7022385963**
	NAME : Kirthika Gurumurthy **SRN : PES1201700230** **EMAIL ID : kirthikagurumurthy@gmail.com** **CONTACT NO. : 7019318232**

ABSTRACT

Data Science is one of the most flourishing domains of Computer Science.
This assignment focuses on choosing a dataset of interest and drawing insights from it through visualization.
The dataset we chose focuses on the Big Four of the stock market right now, i.e, GAFA which stands for Google, Amazon, Apple and Facebook.
According to the famous theories, it is assumed that the past behaviour of a stock price is rich in information concerning its future behaviour. The patterns of the past prices will give some information about the future trends .The more interesting and more unique alerts, however, make use of detailed statistical analysis to watch the market for useful patterns, and to decide which patterns are interesting enough to report.
We analysed the dataset using exploratory data analysis which mainly focuses on making inferences from a dataset through visual aids like graphs of various kinds like line chart, bar graph, etc.

DATA SET

The dataset chosen provides information about the GAFA (Google,Apple,Facebook,Amazon) stock prices in the time period of 1981 (for Apple) to 2018 (all).The source of the given information is Yahoo Finance and the datatset is chosen from Kaggle , the world's largest community of data scientists and machine learners .

This is a big dataset consisting of 19620 rows of data and 8 columns .The columns included are:

- **Stock**-The stock for which the data points have been recorded
- **Date** -The particular day for which the open, high, low, close , adj close, and volume of the stocks is recorded
- **Open**- The opening price or the price at which a security first trades upon the opening of an exchange on a trading day
- **High**- high is the highest price at which a stock trades over the course of a trading day
- **Low**- low is the lowest price at which a stock trades over the course of a trading day
- **Close**- The close is the end of a trading session in the financial markets when the markets close.
- **Adj Close**- An adjusted closing price is a stock's closing price on any given day of trading that has been amended to include any distributions and corporate actions that occurred at any time before the next day's open.
- **Volume**- The number of shares or contracts traded in a security or an entire market during a given period of time (in this case the time period is the particular trading day).

INTRODUCTION

This year has been an all time high for the fantastic 4 of the tech industry: Google, Amazon, Facebook and Apple (GAFA). Facebook grew bigger than China with 1,55Bn users while Amazon became the most valued retailer in the world, leaving Walmart behind. Apple and Google, the two most valuable brands on the planet, are pursuing their ascension, preempting markets one after the other.

The economic shift we're witnessing gets even clearer when we look at GAFA's aggregate performance over the chosen time period (i.e between the years of 2012 to 2018).

We chose this dataset as we are interested in the financial aspect of data science ,especially in the analysis of stocks as it plays a huge role in determining the economic shift in these companies.

We started off our analysis by first limiting the range of years that we are working on (2012-2018) so as to get a common time period for all four companies to get a basis for comparison.We then split the main file into four separate csv files for each of the companies to simplify working with the data.We plotted several kinds of graphs to analyse the data :

- line graphs to observe the basic variation in each of the attributes high,low,open,close,adjacent close, of the stocks in each of these four companies and compare them.
- Box plots for the adjusted close and volume to observe their distribution for each of the companies Google,Amazon,Facebook and Apple.
- Bar graphs of the adjusted close and volume against the year to observe the yearly variation.
- Plot of adjusted close and volume against the year
- Plot of adjusted close against close
- Grouped bar graphs of the adjusted close against the year and the month for each of the big four.
- A 100 day moving average and a 20 day moving average of the adjusted close
- A candlestick graph is a new graph we explored in the mpl_finance module which plots the open, high, low and close in a single bar
- Heatmap for the correlation values of the Adj Close of each company

We are working more with adjusted close rather than the close as **adjusted closing price** as it has been amended to include any distributions and corporate actions that occurred at any time before the next day's open .

Some of the questions we could ask/answer in our analysis are:

- What is the variation in the open, close, high, low, adj close and volume of stocks of Google,Amazon ,Facebook and Apple of the years?

- How is the volume and the adjusted close of the stocks of the Google, Amazon, Facebook and Apple distributed?
- How does the adjusted close vary with the year and the month for each of the big four and possible reasons for the variation in the same?
- How does the adjusted close differ from the close?
- Why is there a sudden spike in Amazon's open, high, low, close and adj close?
- Is there any trend or seasonality observed for these specs and what is the reason for the same?
- What is the strength and maturity or stage of the current trend?
- Are the company's revenues growing?
- Is the company actually making a profit?
- Can the company beat its competitors in the future?
- Will the company's stock be a good investment?

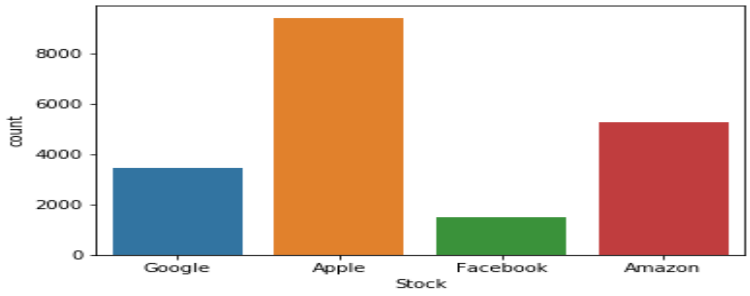
PROCESSING (DATA CLEANING)

The dataset chosen is already sufficiently clean and did not have any unfinished, unreliable, inaccurate or non-relevant parts.

However to make working with each of the columns better we split the main file into 4 separate csv files individually for Google, Apple, Facebook and Amazon. We then made the start date equal for each stock, i.e, we restricted the time period the analysis would be carried out for to bring a common basis for comparison .The number of values of all the different stocks are now equal.

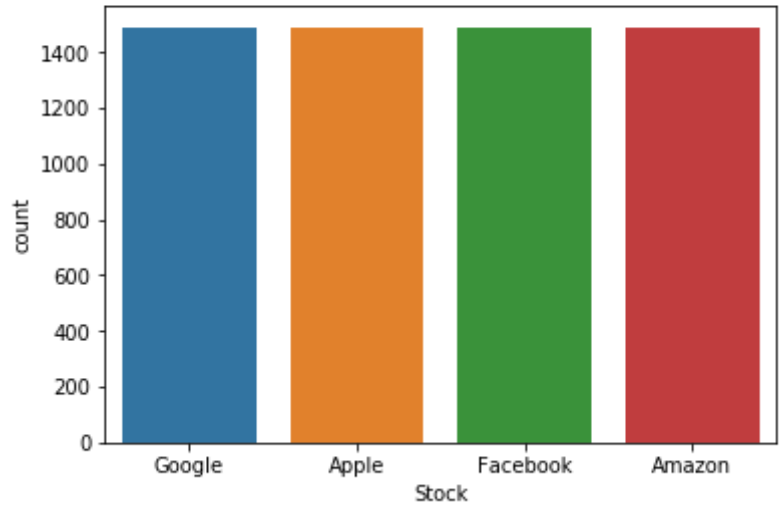
Before setting the start date to be equal

```
Apple       9420
Amazon      5268
Google      3442
Facebook    1490
Name: Stock, dtype: int64
```



After setting the start date to be equal

```
Apple       1490
Facebook    1490
Amazon      1490
Google      1490
Name: Stock, dtype: int64
```

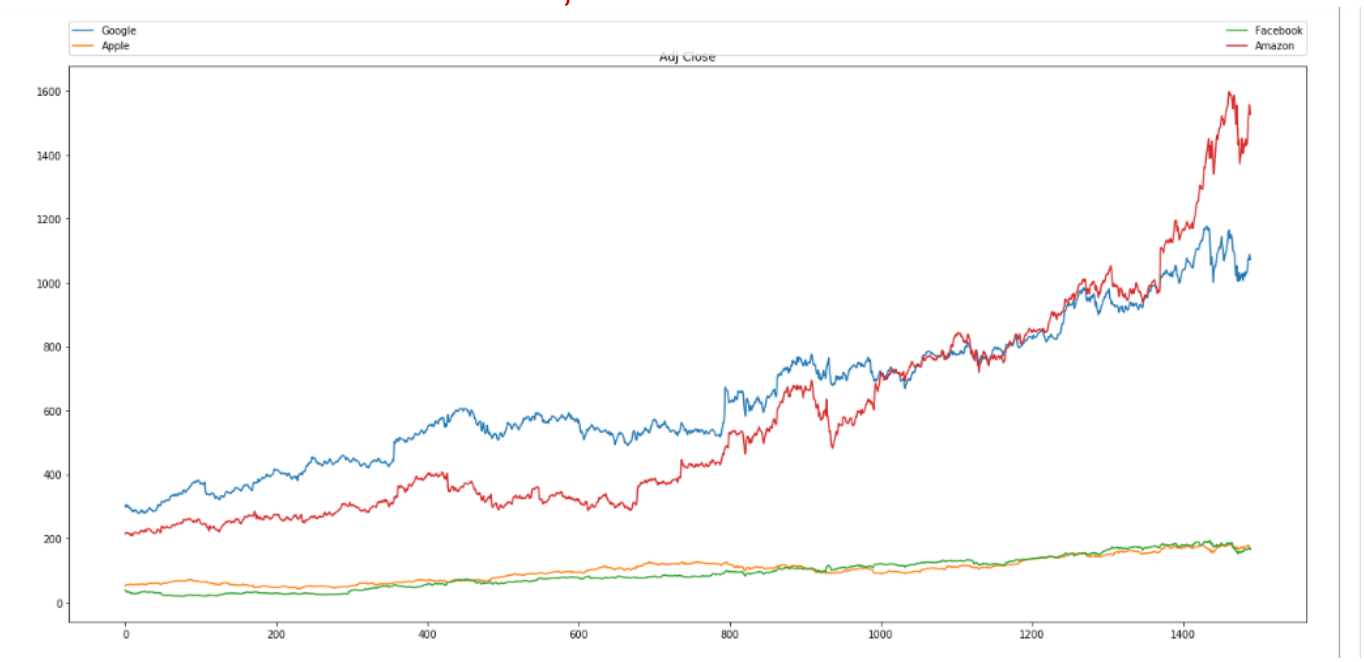


We then converted the date column to date time type to make splitting the date into days,months and years easier and dropped the unnecessary columns created during the splitting of the main files into four separate csv files.

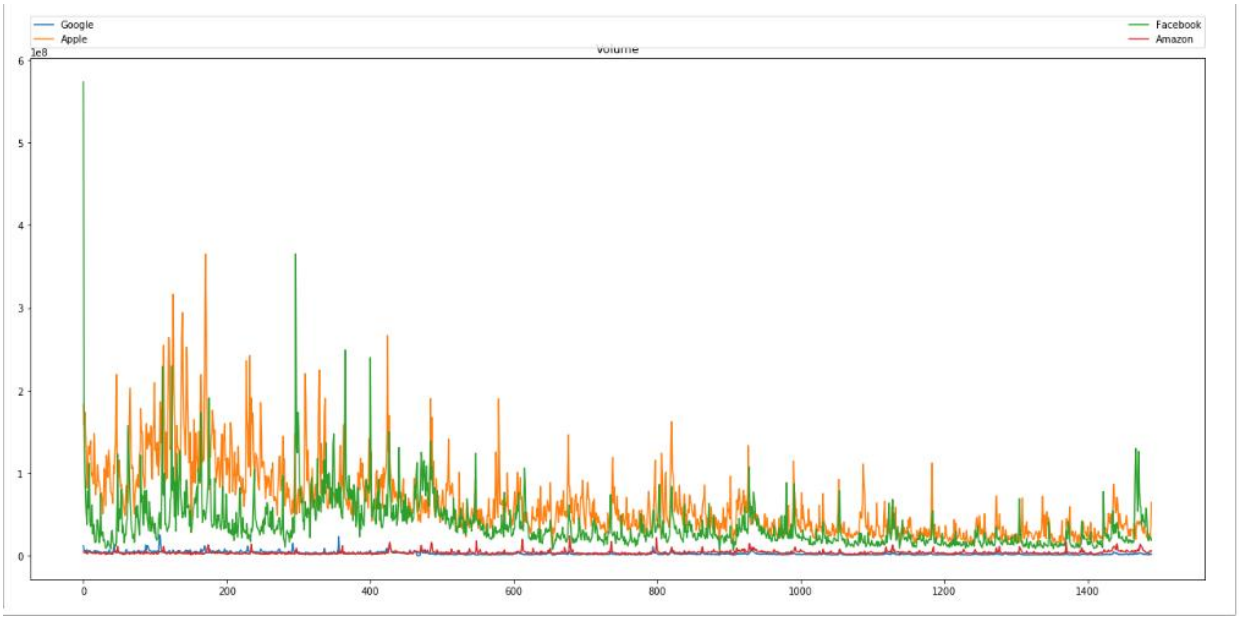
EXPLORATORY ANALYSIS

We started out by plotting line charts for each stock for each column individually to observe the pattern it follows. It can be observed that there is a sudden drastic peak in each of the specs for Amazon (the adj close graph called the price chart and the volume graph has been illustrated). This is mainly due to the result of the company's massive investment spending. Growth amazingly remains unaffected by scale as observed in the volume line graph. Although Amazon keeps getting bigger, its growth isn't coming down all that much.
Initially the stock adjusted close values for Google started out higher that Amazon but this changed as with Amazon's sudden drastic growth in the same.
However the volume parameter of the stocks is highly irregular with no definite pattern as such.
Meanwhile Facebook and Apple show a gradual increasing trend.

Adjusted close



Volume

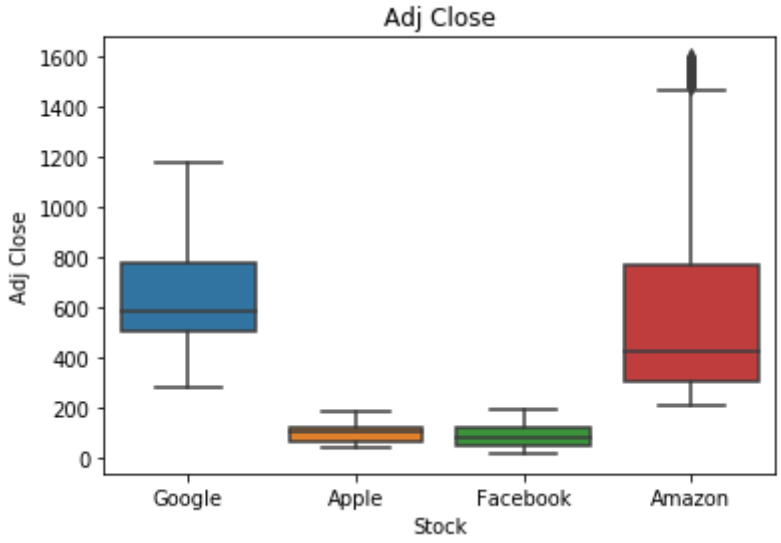


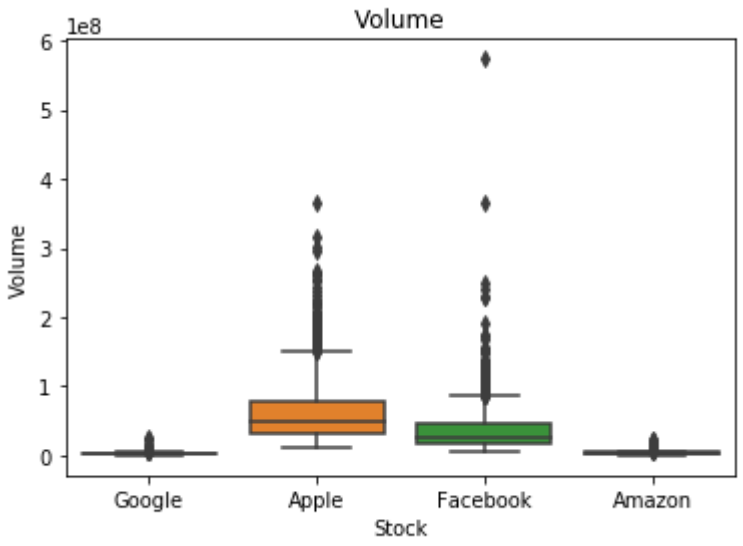
Then we plotted box plots for each of the stock to show us the skewness in data along with the spread of values marking min, max, median, upper quartile and lower quartile for each stock (for adjusted close and volume)

In the adjusted close box plot we can see that the boxes for Facebook and Apple are quite small indicating that the data points,i.e the adjusted close prices hover consistently around the center values while Google and Amazon have more variable adjusted close prices.

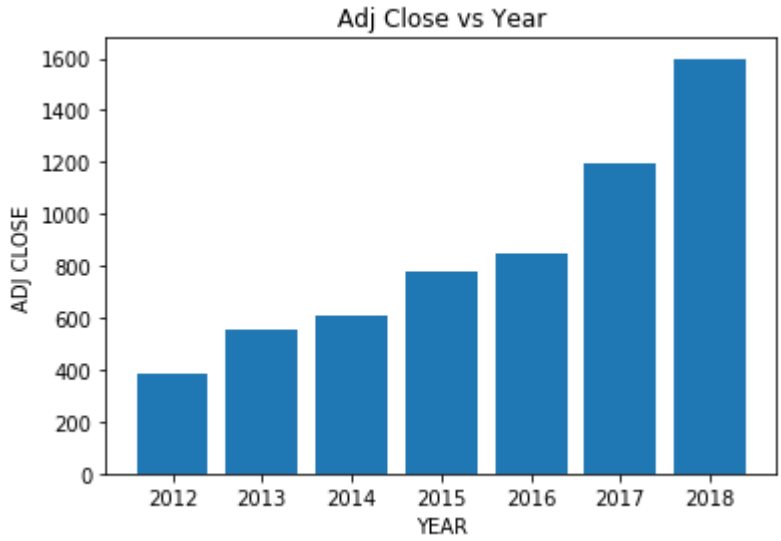
The maximum adjusted close value for Amazon seems to be the highest followed by Google,Facebook and Apple respectively .However the median adjusted close for Amazon is lower than that of Google.The box plots for Google,Facebook and Amazon are right skewed while that of Apple is left indicating that the adjusted close prices for Apple in the recent times are closer together than in the initial time period and the median adjusted close price is higher than the mean and vice versa for the other three stocks.

In the volume box plot we see that the boxes for Google and Amazon are quite small indicating that the volume hovers consistently around the center while Facebook and Apple have more variability in the same. Apple has the maximum volume value (without including outliers). Outliers are observed in this box plot however removing them would not be correct as they only indicate that on a particular day the volume of stocks sold were very high, this is a day wise analysis.

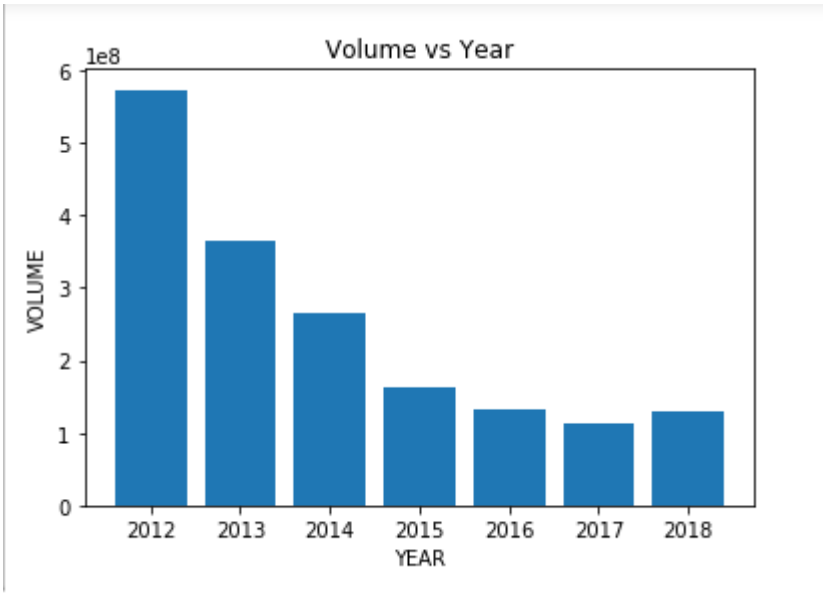




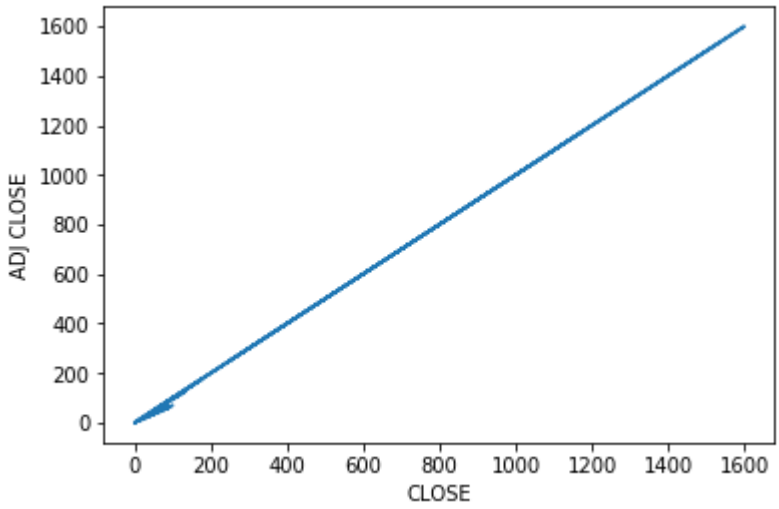
The next thing we did was to plot a cumulative bar graph for "Adj Close" for each year .This shows that the stock prices increase every year.



Then we plotted the cumulative bar graph for "Volume" indicating that as the stock prices increase every year the volume of stocks bought decreases.

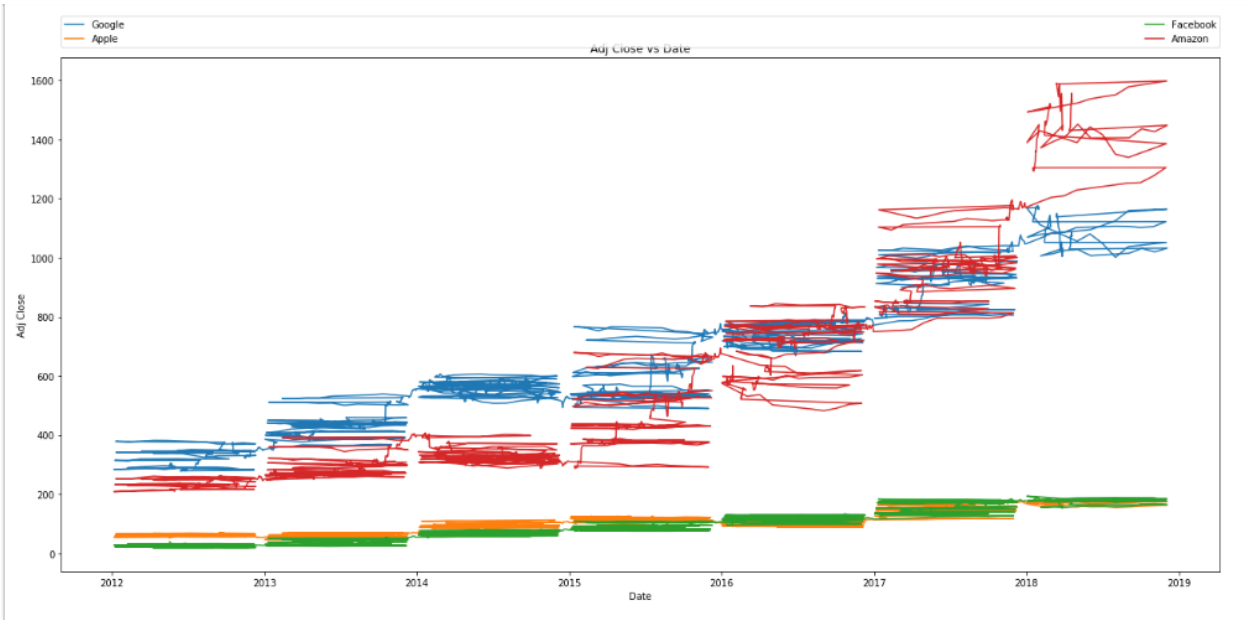


The graph below is plotted between adjusted close and Close .It is a straight line indicating that factors such as dividends, stock splits and new stock offerings did not affect the close values much .

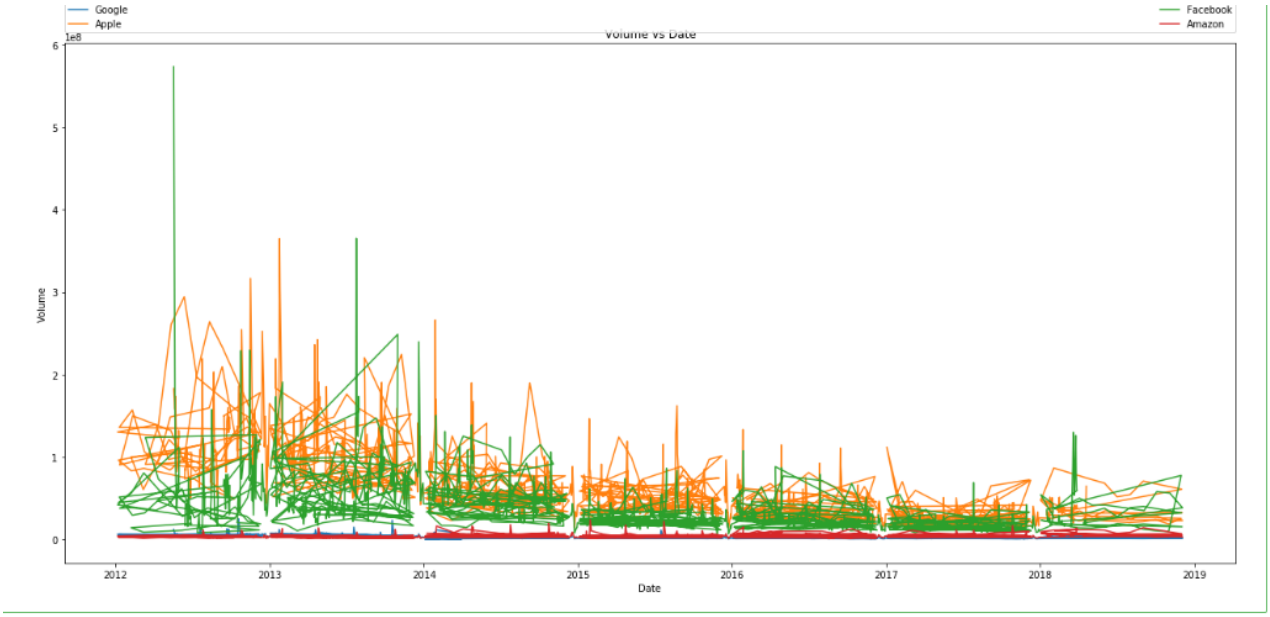


The graphs below helps us analyse the values of the "Adj Close" and "Volume" with the "Year" column (the dataset has new columns named “Year”, “Month” and “Day” containing the year, month and year values derived from the “Date” column) and shows us how the stock prices fluctuated along with the volume bought. Between 2017 and 2018 the adjusted close values for Amazon overtook Google. Google showed the maximum volume of stocks value in 2012.

Adj close vs Date



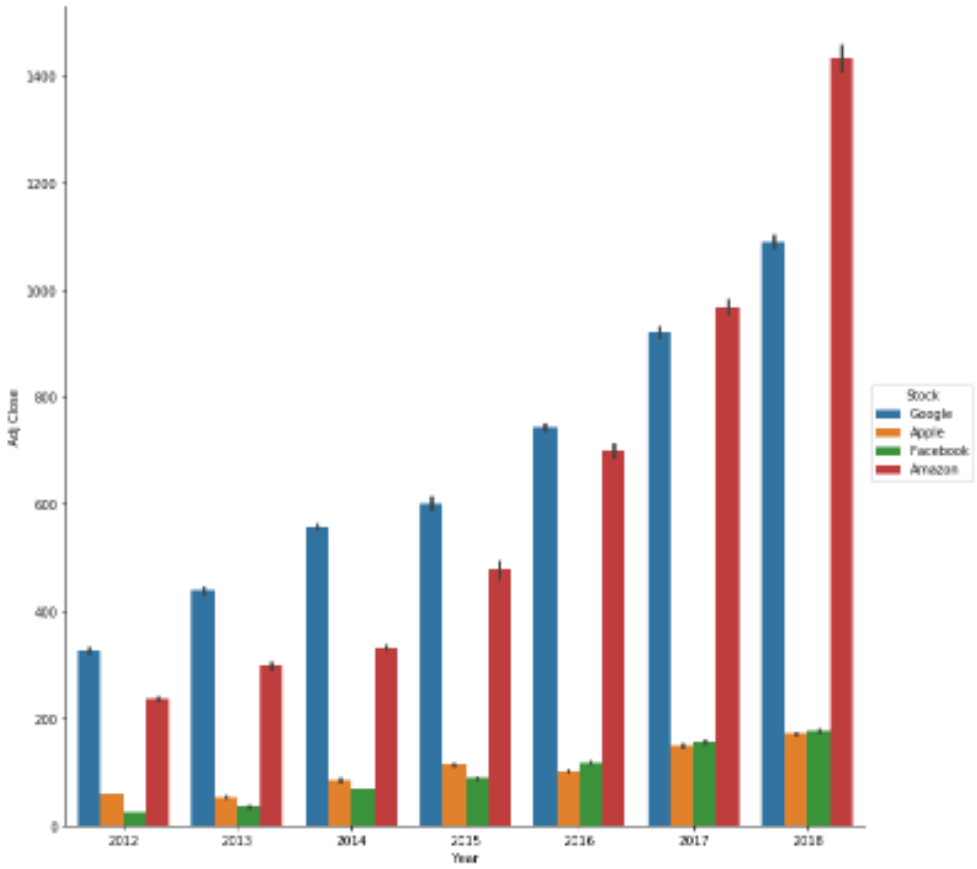
Volume vs Date



Then we plotted a group bar graph: "Adj Close" against the "Year" for each stock to compare the values of "Adj Close" for each stock. It can be observed that Amazon overtook

Google in the Adjusted close prices in 2017. This can be explained by Amazon establishing its Alexa-powered devices as the prevailing standard for non-phone voice assistants and continuing to grow its Amazon Web Services Business. The values for Apple and Facebook are very close together.

Adjusted close vs Year

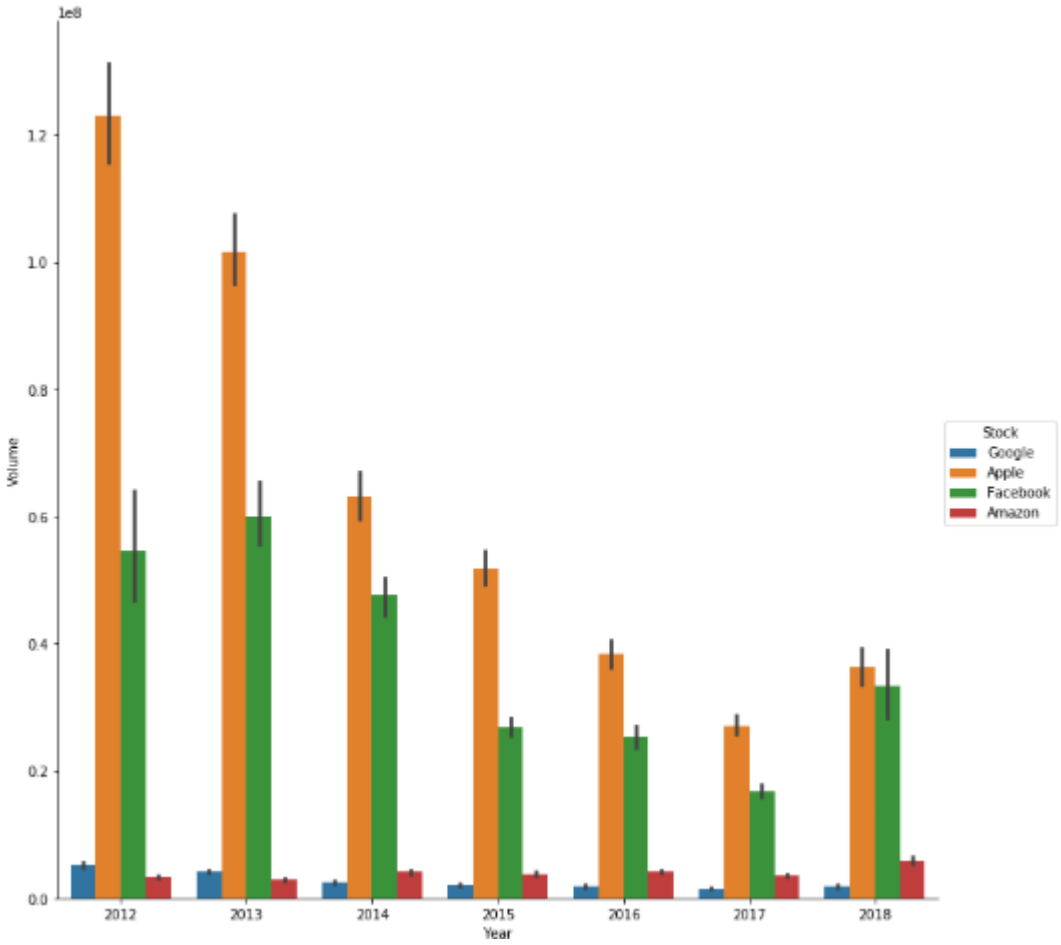


The below grouped bar graph plots the "Volume" against the "Year" for each stock in the same graph to compare the values of "Volume" brought for each stock. It is clearly observed that the trading volume for Apple remains the highest through out the time period but is constantly decreasing from 2012 to 2017 with a slight increase in 2018.The volume change for all the stocks is irregular .

Trading volume can help identify momentum in a stock and confirm a trend. If trading volume increases, prices generally move in the same direction.The change in Google and Amazon is relatively more stable, there is no drastic change in the number of shares .The upward trend is observed the most for Facebook from the year 2017 to 2018 indicating the gain in momentum which will probably continue.

As far as Amazon is considered a large number of stocks have been bought in the year 2018 as compared to Google and this will also most certainly continue considering the growth that Amazon is showing in the recent times.

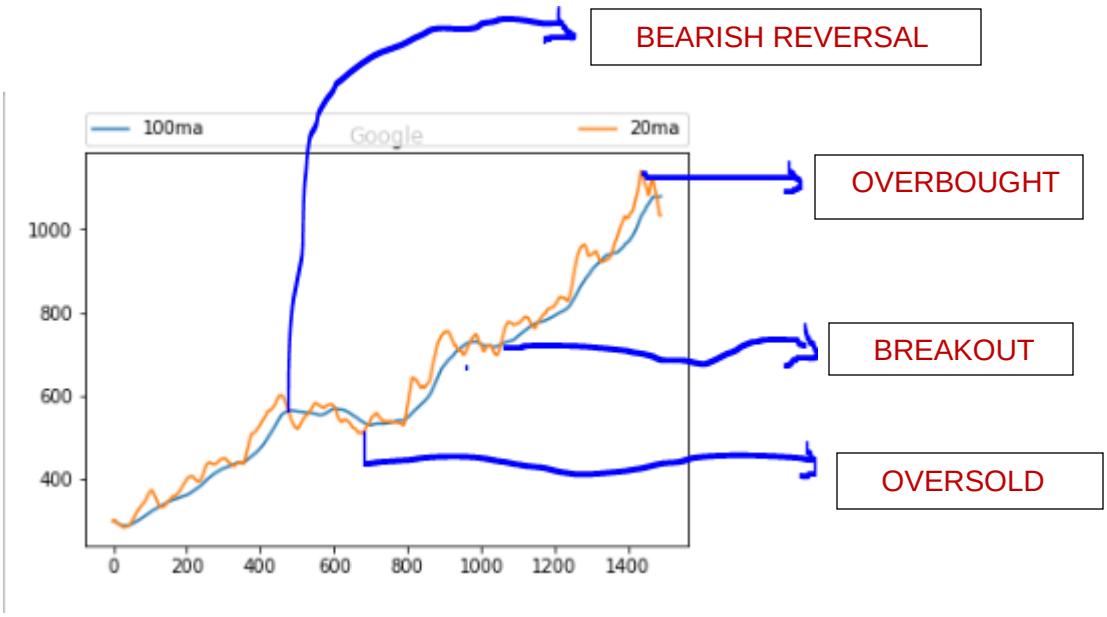
Volume vs Year

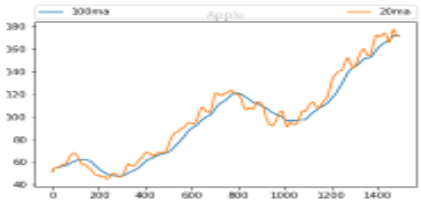


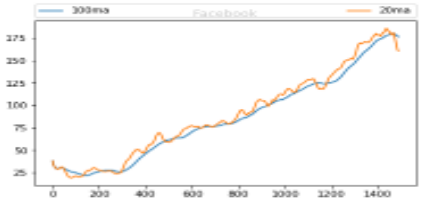
Next we calculated the 100 days and 20 days moving average for each stock. Moving Average (MA) reduces the noise and smoothens out the visual representation of the market and helps us detect patterns that might otherwise go undetected.
Plotting 100MA and 20MA against each other also helps us figure out when the stocks were overbought and oversold.
Short-term moving averages crossing below long-term moving averages is often the sign of a bearish reversal, while a short-term moving average crossover above a long-term moving average could precede a breakout higher.
A Bearish Reversal occurs when today's high is higher than its previous day high and the current price / today's close is lower than its previous day close which can be seen in the graphs indicated below. This can also be seen through the candlestick graph that has been shown later.

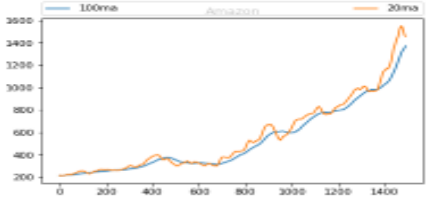
A breakout is most commonly understood as a situation in which the price of an asset breaks through a level of resistance, increasing to an unprecedented price.

We can also identify overbought and oversold patterns .Securities can become overbought and remain overbought in a strong uptrend. Similarly, securities can become oversold and remain oversold in a strong downtrend. In a strong uptrend, prices often move above the upper envelope and continue above this line. In fact, the upper envelope will rise as price continues above the upper envelope. This may seem technically overbought, but it is a sign of strength to remain overbought. The reverse is true for oversold. Overbought and oversold readings are best used when the trend flattens.

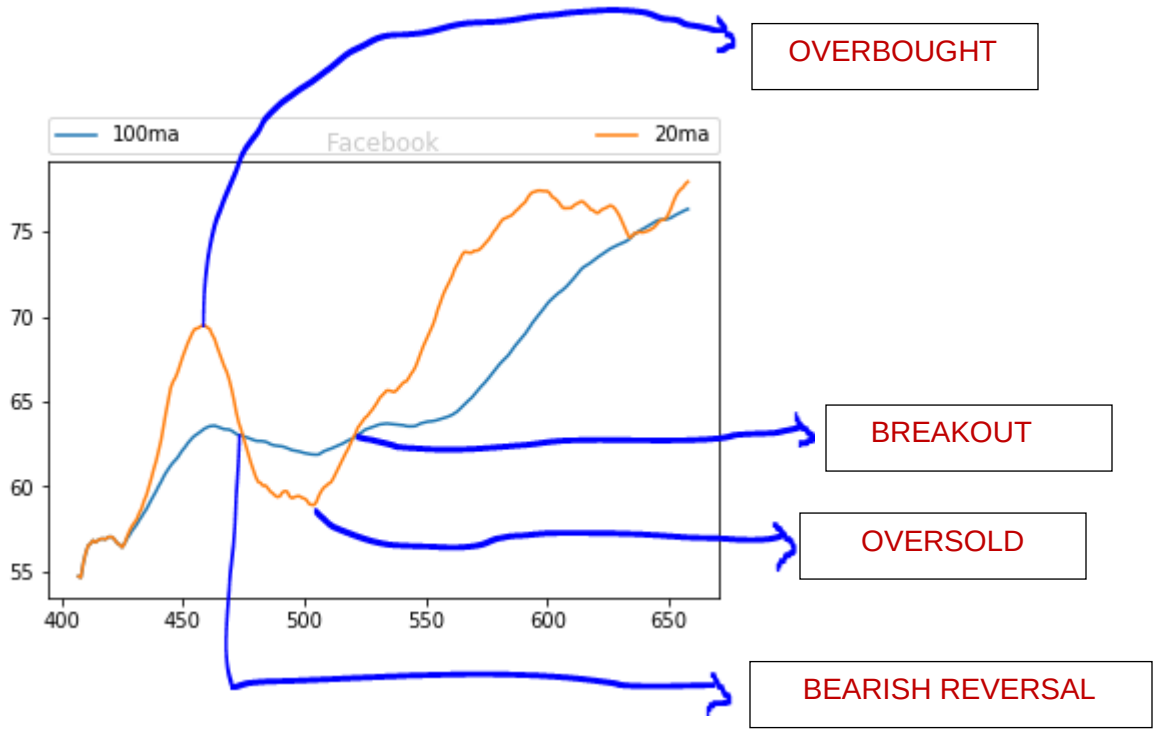




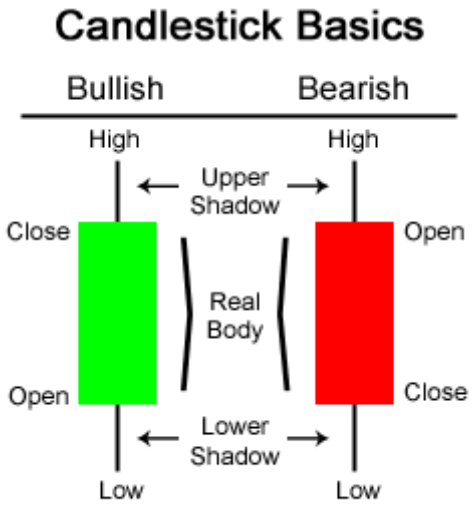




20ma and 100ma plotted for the year 2014 for Facebook



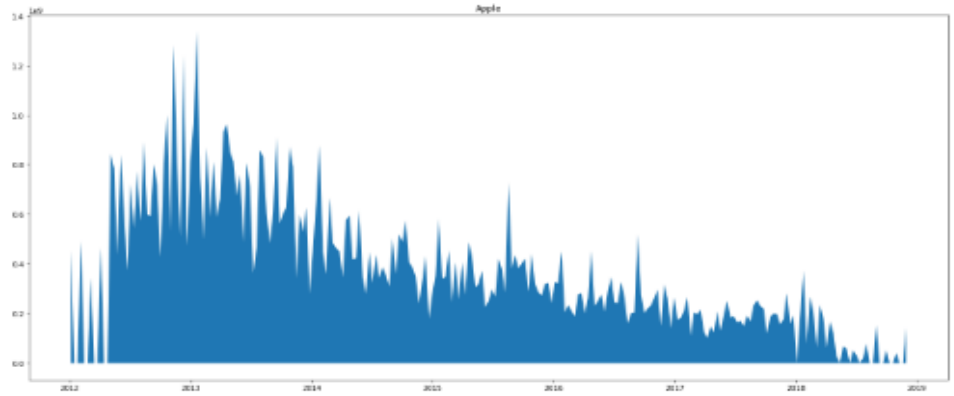
Below graph called candlestick_ohlc (which is a part of the module mpl_finance) for 2015.
A candlestick basically contains the ohlc (open, high, low and close) values on one single bar. The red bar implies that the stock closed at a lower price than it opened indicates selling pressure. The green bar implies that the stock closed at a higher price than the open indicates buying pressure.



Many traders prefer candlestick analysis as it helps them visualize the open, high, low and close of a particular time period (in this case we have taken the average of 10 days for each column and put it in a single dataset for each company) in a single bar. The relationship between the open and close is very important information which can be easily deciphered from a candlestick. Long red candlesticks show strong selling pressure (which after a long decline can lead to panic and capitulation) whereas the long green candlesticks show strong buying pressure (which after a long decline can lead excessive bullishness).

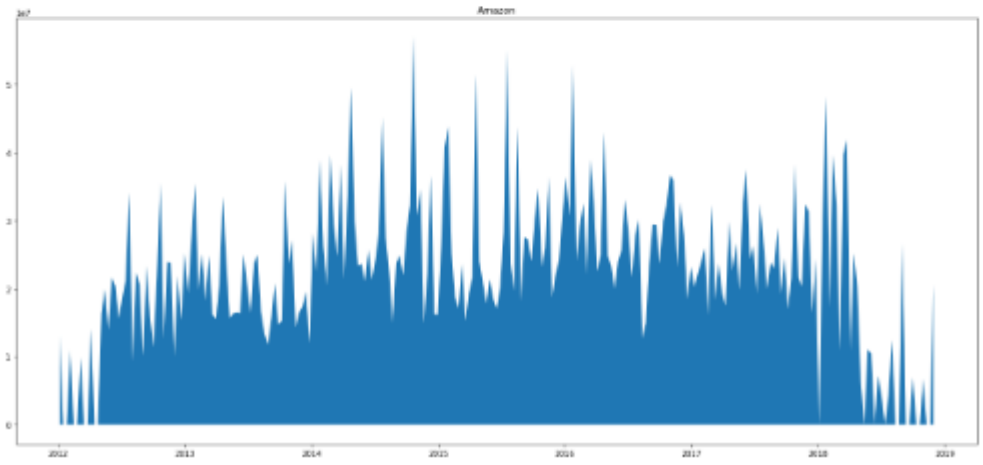
The graphs shown below contain candlesticks for 10 days average for the year 2015 for each of the 4 companies. This graph helped us analyse that Apple had a lot of buying pressure in 2015 whereas Google, Facebook and Amazon had a lot of selling pressure.

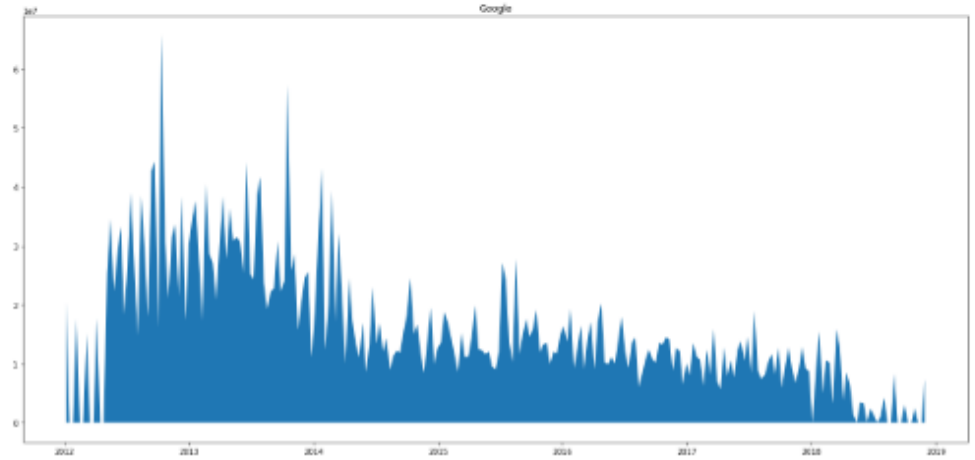
Also plotted is the sum of volume graph which is filled between with color to show the volume trend for each stock over the time period of 2012 to 2018.



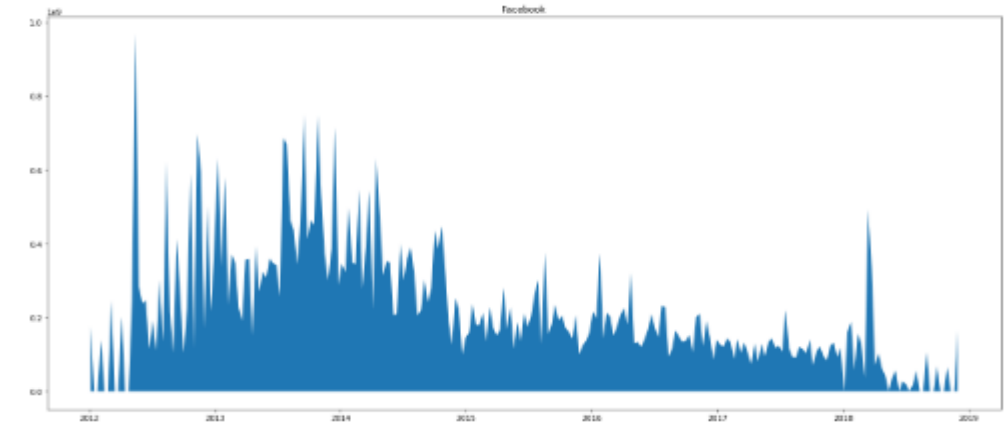
APPLE

AMAZON



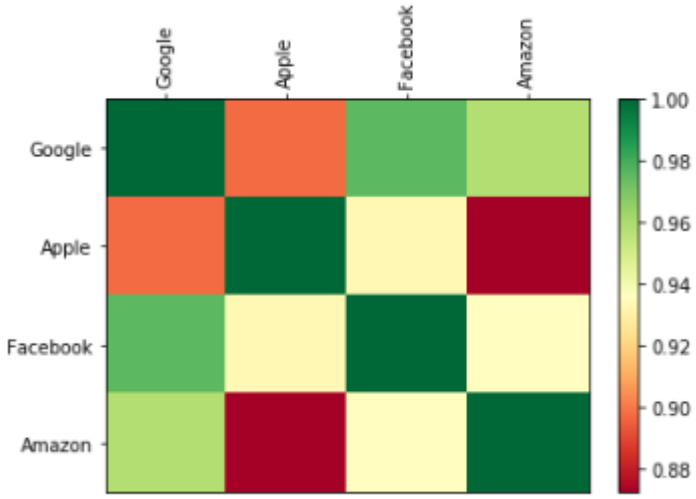


GOOGLE



FACEBOOK

Below is a heatmap from the correlation between all the columns that contain the individual 'Adj Close' value for each stock. 1 represents maximum correlation. Facebook and Google have a fairly strong correlation while amazon and apple have a very weak correlation.
The map has been plotted with the help if the correlation values of the Adj Close values of all the four companies.



CONCLUSION

It can be concluded that seeing the overall trends investing in Amazon would have the best forecasted medium to long term gains, followed closely by Google.
From this exercise we learned a lot about stocks and their technical analysis and how to use moving averages, candlestick plots etc .We learned how to observe the patterns of past prices to give us information about future trends and through careful analysis develop an understanding of these patterns.
The purpose of this report was particularly to understand the general behaviour of GAFA stock prices over 2012 to 2018 and sudden fluctuations involved along with possible reasons though different descriptive statistics.