Exploratory Data Analysis on Stock market Data

Stock data can be obtained from **Yahoo! Finance, Google Finance**, or a number of other sources. A line chart has been performed on The **Dow Jones Industrial Average**.The chart is given below.

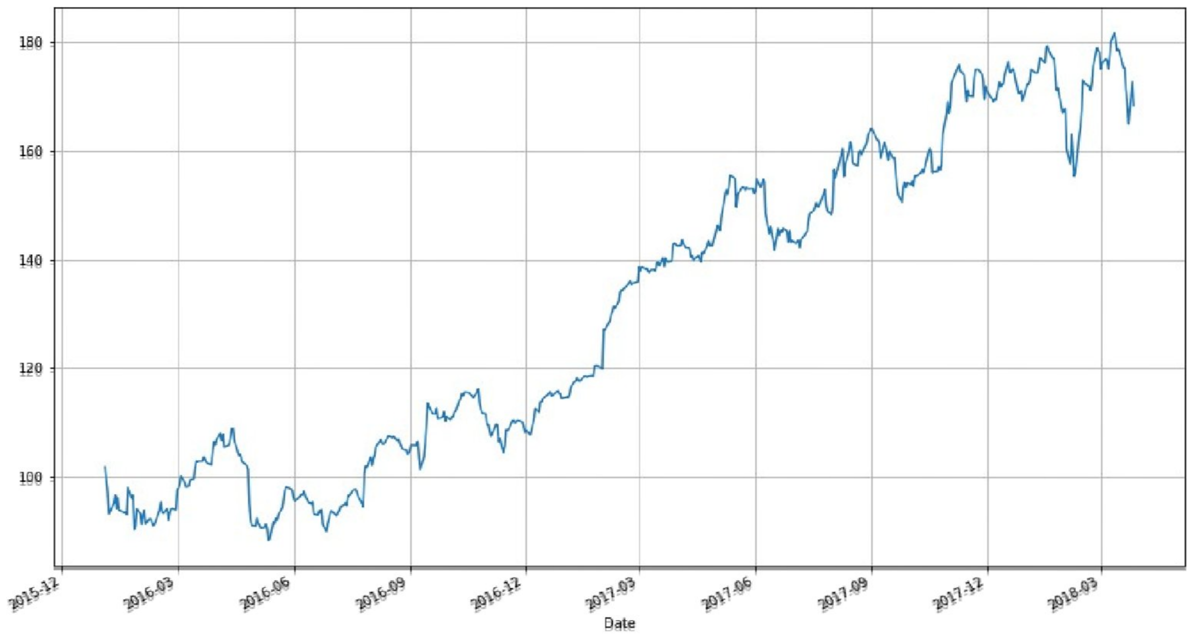

Fig. 1

Line chart has been prepared on Dow Jones Industrial average Index of Stock Market Data starting from **2015 December till 2018 March**. According to the graph the trends have moved steeper upwards according to time.

Uptrend lines act as support and indicate that net-demand (demand less supply) is increasing even as the price rises. A rising price combined with increasing demand is very bullish, and shows a strong determination on the part of the buyers. As long as prices remain above the trend line, the uptrend is considered solid and intact. A break below the uptrend line indicates that net-demand has weakened and a change in trend could be imminent.

With a boxplot-whiskers chart, a black box indicates a day where the closing price was higher than the open (a gain), while a red box indicates a day where the open was higher than the close (a loss). The wicks indicate the high and the low, and the body the open and close (hue is used to determine which end of the body is the open and which the close). The median of the **Dow Jones Industrial average** is shown to increase over time.

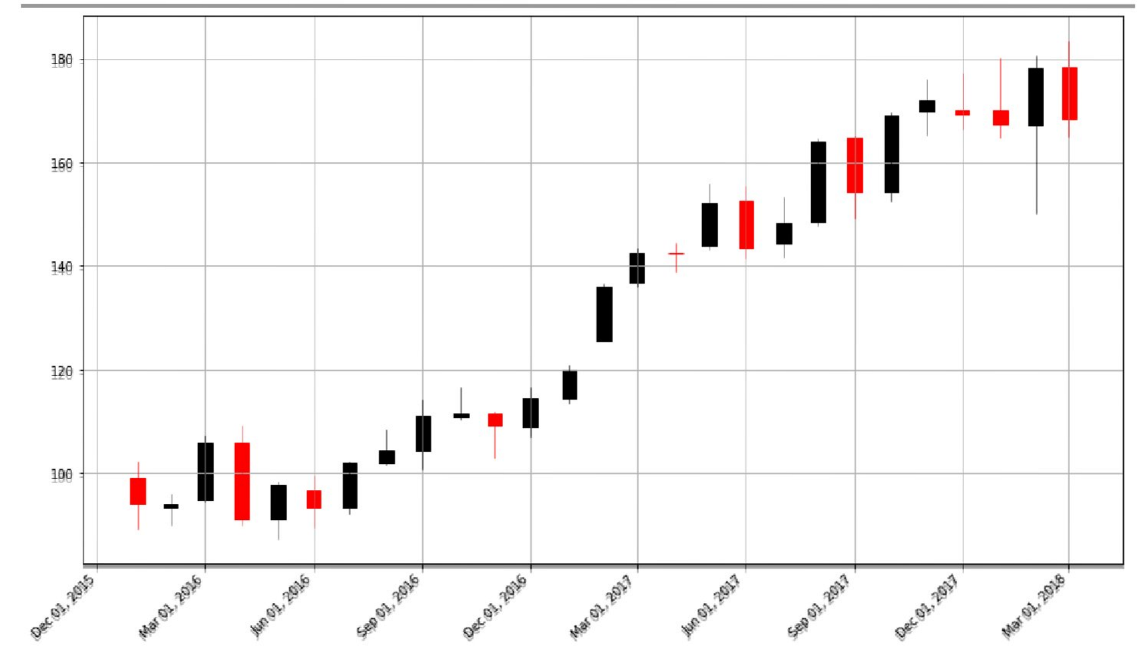

Fig. 2

A plot of multiple financial instruments together is performed. As we may want to compare stocks,compare them to the market, or look at other securities such as **exchange-traded funds (ETFs).** Later, we will also want to see how to plot a financial instrument against some indicator like a moving average.

While absolute price is important, when trading, we are more concerned about the relative change of an asset rather than its absolute price. Google's stocks are much more expensive than Apple's or Microsoft's, and this difference makes Apple's and Microsoft's stocks appear much less volatile than they truly are (that is, their price appears to not deviate much).(**Source Google Finances**)

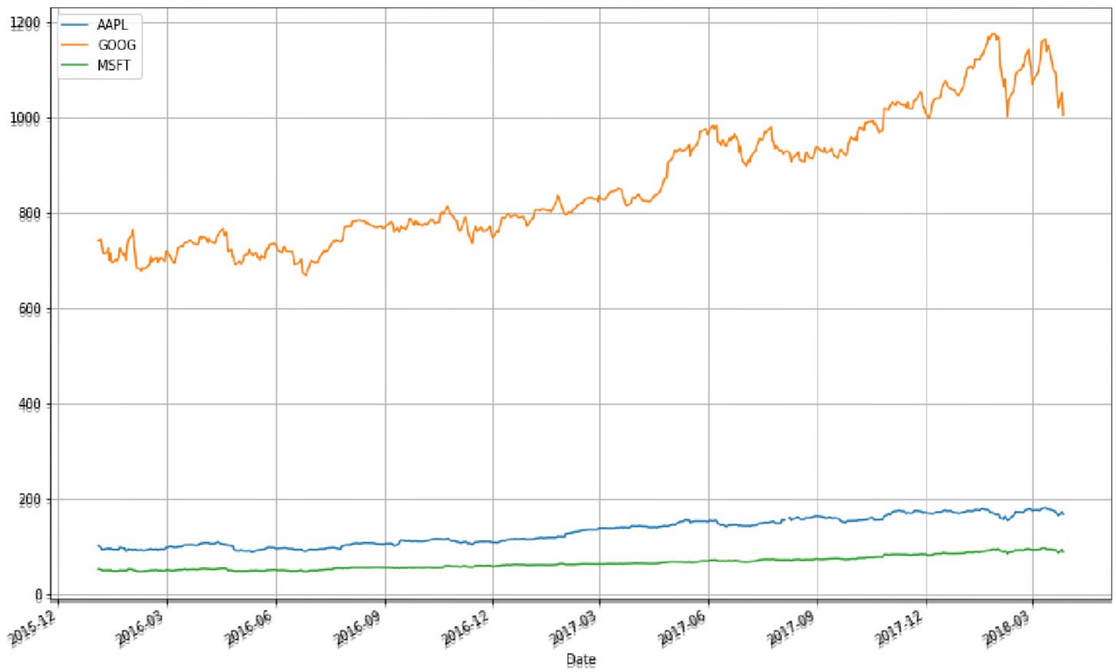

Fig. 3

One solution would be to use two different scales when plotting the data; one scale will be used by Apple and Microsoft stocks, and the other by Google.

AAPL shows a shift from that of GOOG and MSFT and it has a fluctuating slopeline than in GOOG and MSFT.(Time of Interest: 2015 March-2018 March)

A "better" solution, though, would be to plot the information we actually want: the stock's returns. This involves transforming the data into something more useful for our purposes. There are multiple transformations are applied.

One transformation would be to consider the stock's return since the beginning of the period of interest. In other words, we plot:

$$\text{return}_{t,0} = \frac{\text{price}_t}{\text{price}_0}$$

This will require transforming the data in the stocks object, using a **lambda function**, which allows the data to pass a small function defined quickly as a parameter to another function or method.

Fig. 4

This is a much more useful plot. We can now see how profitable each stock was since the beginning of the period. Furthermore, we see that these stocks are highly correlated; they generally move in the same direction, a fact that was difficult to see in the other charts.

Alternatively, we could plot the change of each stock per day. One way to do so would be to plot the percentage increase of a stock when comparing day to day , with the formula:

$$\text{growth}_t = \frac{\text{price}_{t+1} - \text{price}_t}{\text{price}_t}$$

But change could be thought of differently as:

$$\text{increase}_t = \frac{\text{price}_t - \text{price}_{t-1}}{\text{price}_t}$$

These formulas are not the same and can lead to differing conclusions, but there is another way to model the growth of a stock: with log differences.

$$\text{change}_t = \log(\text{price}_t) - \log(\text{price}_{t-1})$$

Fig. 5

The advantage of using log differences is that this difference can be interpreted as the percentage change in a stock but does not depend on the denominator of a fraction. Additionally, log differences have a desirable property: the sum of the log differences can be interpreted as the total change (as a percentage) over the period summed (which is not a property of the other formulations; they will overestimate growth). Log differences also more cleanly correspond to how stock prices are modeled in continuous time.

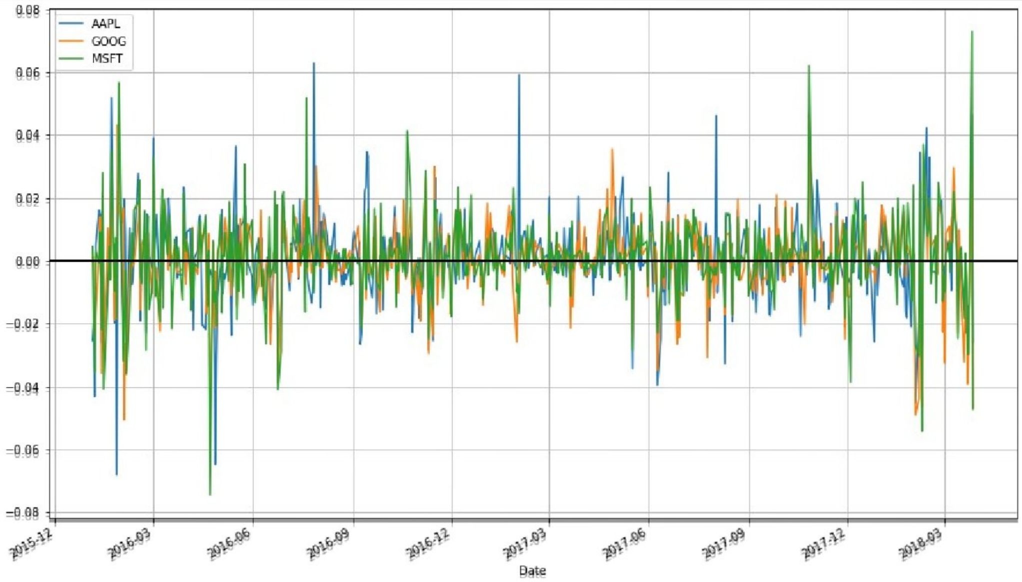

Fig. 6

Looking at returns since the beginning of the period make the overall trend of the securities in question much more apparent. Changes between days, though, are what more advanced methods actually consider when modelling the behavior of a stock. so they should not be ignored.

We often want to compare the performance of stocks to the performance of the overall market. SPY, which is the ticker symbol for the SPDR S&P 500 exchange-traded mutual fund (ETF), is a fund that attempts only to imitate the composition of the S&P 500 stock index, and thus represents the value in “the market.”,Data used here collected from Yahoo! Finance.

Below I get data for SPY and compare its performance to the performance of the stocks.

AAPL
GOOG
MSFT
SPY
1.8
1.6
1.4
1.2
1.0
2015-12
2016-03
2016-06
2016-09
2016-12
2017-03
2017-06
2017-09
2017-12
2018-03
Date

Fig. 7

Applying Moving Averages

A q**-day moving average** is, for a series x_t and a point in time t, the average of the past q days: that is, if MA_t^q denotes a moving average process, then:

$$MA_t^q = \frac{1}{q} \sum_{i=0}^{q-1} x_{t-i}$$

Moving averages smooth a series and helps identify trends. The larger q is, the less responsive a moving average process is to short-term fluctuations in the series x_t. The idea is that moving average processes help identify trends from "noise". **Fast** moving averages have smaller q and more closely follow the stock, while **slow** moving averages have larger q, resulting in them responding less to the fluctuations of the stock and being more stable.

Fig. 8

We can notice how late the rolling average begins. It cannot be computed until 20 days have passed. This limitation becomes more severe for longer moving averages. Because I would like to be able to compute 200-day moving averages, I'm going to extend out how much AAPL data we have. That said, we will still largely focus on 2016.

Fig. 9

We will notice that a moving average is much smoother than the actual stock data. Additionally, it’s a stubborn indicator; a stock needs to be above or below the moving average line in order for the line to change direction. Thus, crossing a moving average signals a possible change in trend, and should draw attention.

Traders are usually interested in multiple moving averages, such as the 20-day, 50-day, and 200-day moving averages. It’s easy to examine multiple moving averages at once.

Fig. 10

The 20-day moving average is the most sensitive to local changes, and the 200-day moving average the least. Here, the 200-day moving average indicates an overall **bearish** trend: the stock is trending downward over time. The 20-day moving average is at times bearish and at other times **bullish**, where a positive swing is expected. You can also see that the crossing of moving average lines indicate changes in trend. These crossings are what we can use as **trading signals**, or indications that a financial security is changind direction and a profitable trade might be made.