[[1]](#footnote-1)

S&P500 Analysis

*Noel Hynes, Student, Letterkenny Institute of Technology*

# Introduction

T

he Standard and Poors 500 (S&P500) is a market index of 500 of the largest companies listed on the stock exchanges in the United States. These companies cover ~ 80% of the American equity market by capitalization.

In this project, we will analyze S&P500 data between February 2013 and February 2017. We will look for trends in the data and try to discover some key aspects of the dataset over this period.

For more in depth analysis, we have merged together our original data file with a second file, which provides additional information for each stock included year it was founded and industry it is involved with

# Key Data Links

## GitHub Repository for Project

* https://github.com/hynesenberg/UCDPA\_NoelHynes2

## Source of all\_stocks\_5yrs file

* https://www.kaggle.com/camnugent/sandp500?select=all\_stocks\_5yr.csv

## Source of Symbol Details

* https://en.wikipedia.org/wiki/List\_of\_S%26P\_500\_companies

# Coding

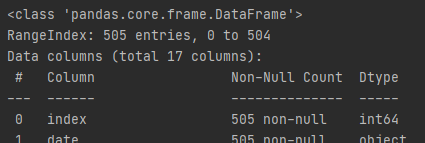
I have included detailed descriptions of each line of code in the python notebook.

# Main Insights

From our working through the datasets, we have established a number of key insights.

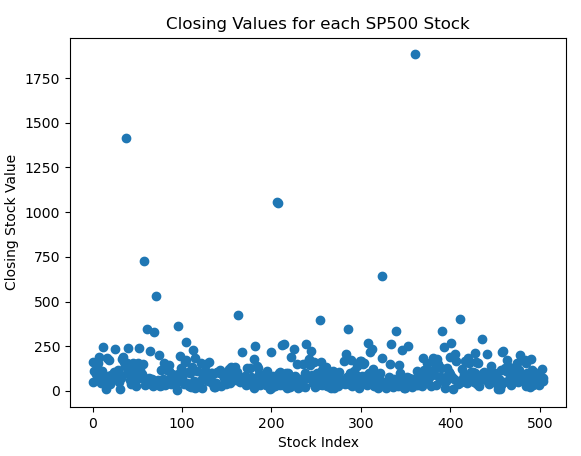
## Number of Stocks

From using the .info() python code, we can see that there are 505 entries for our closing SP500 values. This seems interesting, given I would expect 500 stocks on the S&P500 index. However, upon further analysis and research we can see that 5 companies on the S&P500 have 2 share classes.



## Closing Values

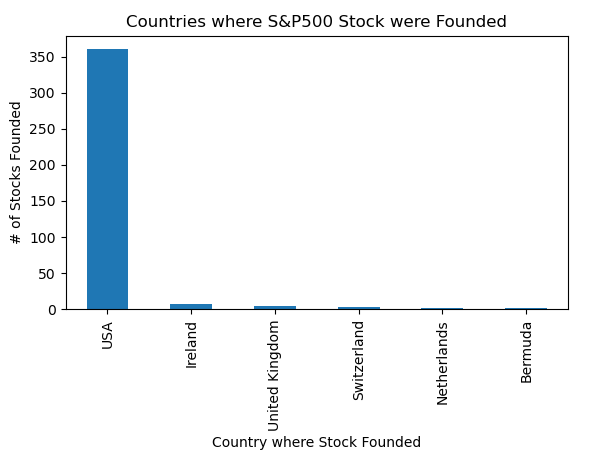
From the first chart we plot we can see the wide range of prices in the market index. We can see prices ranges upto almost 2000 dollars, but with the majority under 250 dollars market price.



## Country Founded

The bar chart below clearly shows that while a small number of S&P 500 companies were not founded in the US, the vast majority of them were founded in USA.

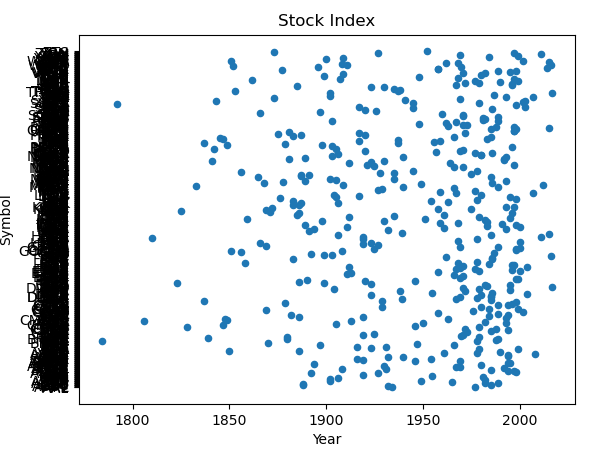
Note that the chart does not quite add up to 505, as some companies were renamed over the years and hence their additional information is not available for this chart as we dropped NaN entries to our table as a result of the merge.



## Founded Dates

The below chart shows the years in which each stock was founded. What we can see here is that there is a large spread over time, with companies being founded as far back as the 1800’s and right up to present day.

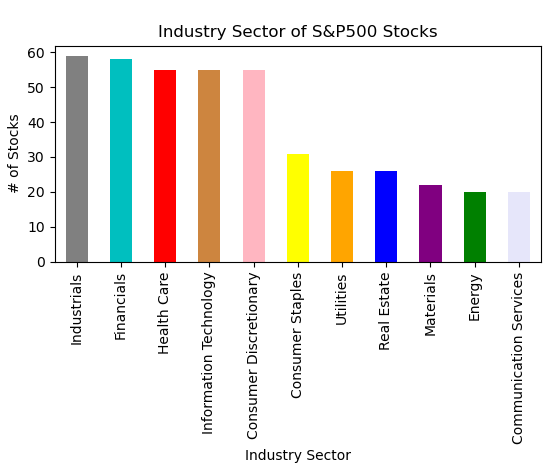
This shows that the S&P500 index is a mix of all ranges of different companies over time.



## Industries

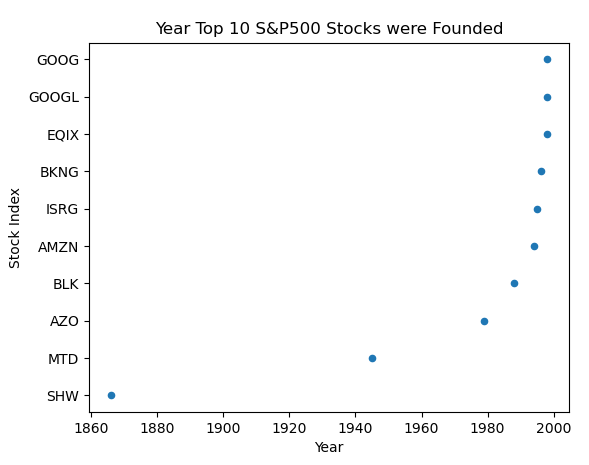
The graph below gives us an idea of the various industries which feature in the SP500 market index. The most popular sectors are industrials, finance and healthcare, but with a large mix of industries overall.

This shows us a well-diversified portfolio of assets.



## Top 10 Companies Founded

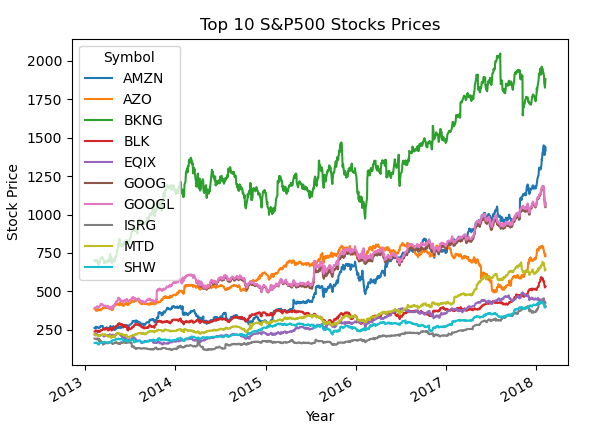
We can see from the next chart that of the top 10 stocks listed on the S&P500, 8 of which have been established since 1980. This shows that although some rival companies were established over 100 years ago, others have overtaken these to become more valuable now.



## Trends

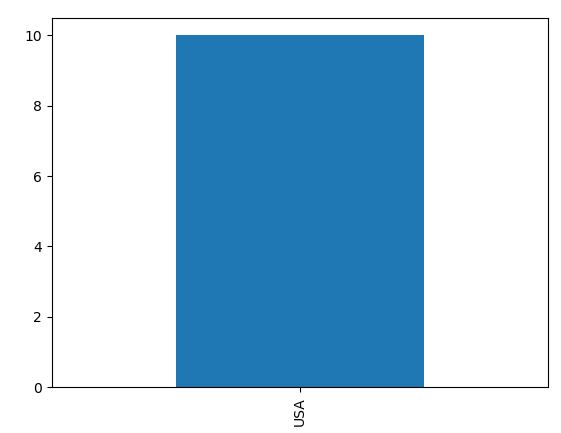
We can see the trend of all of the top 10 prices stocks in the S&P500 index are all showing strong growth trends.

Amazon(Amzn), despite only being founded in 1994, is growing rapidly and looks to be catching upto the top priced S&P500 stock, Booking Holdings Inc. (BKNG).



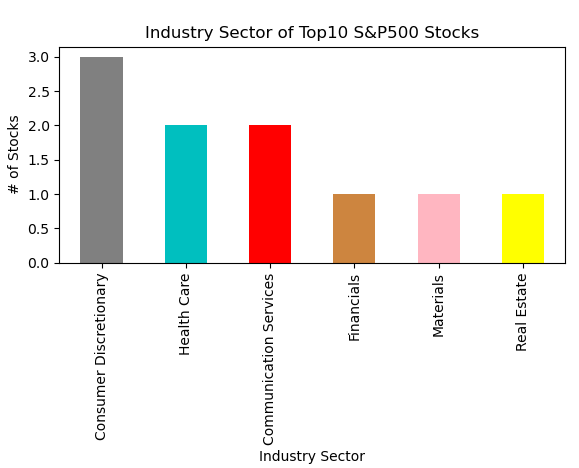
## USA Founded Companies

I was curious to see where all the top companies were founded. Unsurprisingly,



## Top 10 S&P500 Stock Industries

Finally, I looked at the sectors that the top stocks were based in.



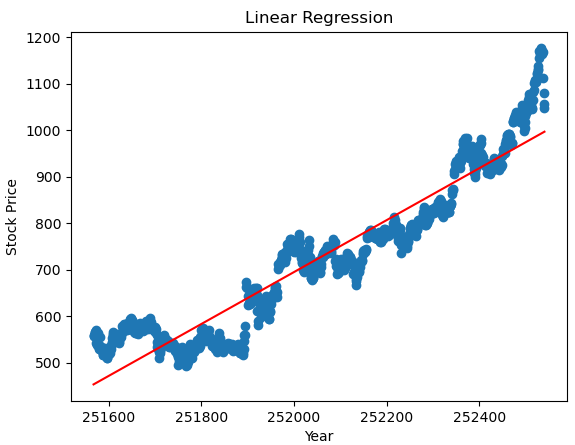
What I find interesting here I that no traditional “I.T.” companies feature in the top 10 stocks. While this is surprising, some of the top performing stocks, such as Google and Amazon, are almost “I.T.” companies but working in non-traditional areas such as sales and Internet of Things.

## Linear Regression

I have carried out a linear regression of the GOOG stock to track the rise of the stock and which can be sued to predict the trend of the stock price going forward.

This fitline/prediction line is a good fit to the data, as the stock has consistently risen over time. A more volatile stock would not be suited to a linear regression.

The R^2 value of the Fitline is 0.8897488, which is a good fit. (1 would be an exact match and 0 would be not a good match)



# Discussion & Conclusion

To conclude, we have analyzed the S&P 500 index from a range of different angles and ascertained a lot of high level information from the data that would not have been able to be viewed from excel due to the large volume of the data.

While USA founded companies dominated the list, this could largely be due to the fact that the S&P500 companies and stocks traded on American stock exchanges. This analysis is skewed as such, and so next step would be to compare with market indices from Asia and other stock exchanges around the world to see how these compare.

I would also have added a polynomial regression as a next step to check predictions for the stock prices at future dates. These values could be compared to values achieved since 2017 to gauge the success of a practice commonly used by traders, “charting”.

Python is a very efficient program and can deal with large quantities of data easily, with a range of readymade libraries to make such analysis user friendly.

1. Noel Hynes, Current Student with the University College Dublin

   (e-mail: noelhynes46@hotmail.com) [↑](#footnote-ref-1)