S&P 500 Data Analysis Project Report

Business Task

The business task is to answer the following questions (suggested by Maven Data Agency (the provider of the dataset)) and also to provide insight into any other interesting aspects of the data that are learned during the analysis which could be used to help improve a business.

1. Which date in the sample saw the largest overall trading volume? On that date, which two stocks were traded most?
2. On which day of the week does volume tend to be highest? Lowest?
3. On which date did Amazon (AMZN) see the most volatility, measured by the difference between the high and low price?
4. If you could go back in time and invest in one stock from 2nd January 2014 – 29th December 2017, which would you choose? What % gain would you realize?

Data Source

The data has been sourced from Maven Data Analytics Agency (mavenanalytics.io), who offer a range of datasets for free, to download and analyse.

Data Organisation

The source data is organised into the following fields:

- Symbol (company ticker symbol)
- Date (date of trading in question. Format: YYYY-MM-DD.)
- Open (share price at start of the trading day. Units: US Dollars.)
- High (highest share price during the trading day. Units: US Dollars.)
- Low (lowest share price during the trading day. Units: US Dollars.)
- Close (share price at end of the trading day. Units: US Dollars.)
- Volume (volume of stock bought/sold during the trading day. Units: Shares.)

Bias / Credibility of the Data

I was unable to verify the dataset against alternative sources online; in a work environment I would investigate this issue further, and query the reliability of this dataset, however for the purpose of this exercise I consider the data reliable enough to use. The data is considered comprehensive as it covers a broad range of time, and no data has been omitted during capture, although a small number of records will be dropped during cleaning. During cleaning, it also became clear that there are thousands of records where the low price is higher than the opening price, which stood out to me as odd, however upon further research, this is to be expected on occasion, and simply means that the price rose *immediately* after the market opened and didn't fall back to the low of the opening price; so these records will not be dropped.

Licence

The data is in the public domain, so can be used freely for any purpose.

Privacy

The data is public, so there are no privacy concerns for this project.

Security

As the data is already in the public domain, there are no security concerns for this project, however the data will be stored securely on my local machine, and a backup kept on Google Drive.

Integrity
The data is of suitable integrity for the purposes of this exercise. Records with null / NaN values, and those where the "high" is lower than the "low" will be dropped during cleaning, as I question the integrity and reliability of these records.

Timeframe
The data covers the period from 2nd January 2014 – 29th December 2017. No other timeframe has been considered as part of this analysis.

Code
Complete Python code for the cleaning and analysis can be found in a PDF alongside this report.

Process

Tools
I have completed cleaning, analysis and visualisations using Python (running in Anaconda Jupyter Notebook), including packages Numpy, Pandas, MatPlotLib and Seaborn.

Cleaning
I have explored the data carefully to perform a thorough cleaning. The prep, cleaning, checking and due diligence process entailed:

- Import the data and make a copy for safety.
- Verify that the correct number of fields and records have been imported.
- Rename "symbol" column to "company", for user friendliness.
- Check for any duplicate rows – there were none.
- Check for duplicate rows considering just date and company - the two fields which make each record unique. There were none.
- Change the data-type of the date column to *datetime*, in the format (YYYY-MM-DD).
- Drop the one record where the high value is lower than the low value, as this record is considered not reliable.
- Drop all records containing 1 or more null / NaN value, (these are blank on the source CSV), to avoid skewing the analysis with potentially unreliable data.
- Check for any records showing 0 volume of trading – there are now none, as they have been dropped as part of the previous step.
- Create a day of the week column (using *.dt.day_name()* method)
- Create a percentage day change column ((close – open) / open * 100)
- Create an absolute day change column (close – open)
- Create a percentage volatility column ((high – low) / low * 100)
- Create an absolute volatility column (high – low)
- Reorder the columns to a more sensible order.
- Reorder the rows, sort by date firstly, then by company, then reset the index.
- Eyeball the cleaned data.

FYI – the following have also been done prior to cleaning:

- Import Numpy, Pandas, and MatPlotLib libraries.
- Set the max number of rows to display at any-one-time to 50, for easier reading.
- Suppress scientific notation, for easier reading.

- Within Pandas, round all numbers to two decimal places, for easier reading.

Analyse & Share

Summary of Analysis Findings

Average stock growth

Between 2nd January 2014 – 29th December 2017, the average S&P 500 opening stock price increased from 72.07 to 107.61, providing a 49.3% total ROI, on average across the index. This journey is shown in the below line chart. When you exclude stocks that joined the market after 2nd January 2014, this average increases to 53.3%. This could be due to the newer stocks having less time to provide a reliable long-term ROI.

Volumes of Trading

The highest volume of trading in a single day took place on the 24th August 2015, with volume traded of 4,607,945,196 shares. On this day, the two companies traded the most were Bank of America (BAC) and Apple (AAPL), with 214,649,482, and 162,206,292 shares traded respectively. Tables of these two rankings are shown below.

date	volume
2015-08-24	4607945196
2016-06-24	4367393052
2015-12-18	4124454411
2016-01-20	4087629753
2016-11-10	4060601612
...	...
2014-12-26	894908944
2015-11-27	791154818
2014-12-24	750895627
2015-12-24	736263173
2017-11-24	728261080

	company	date	volume
201260	BAC	2015-08-24	214649482
201204	AAPL	2015-08-24	162206292

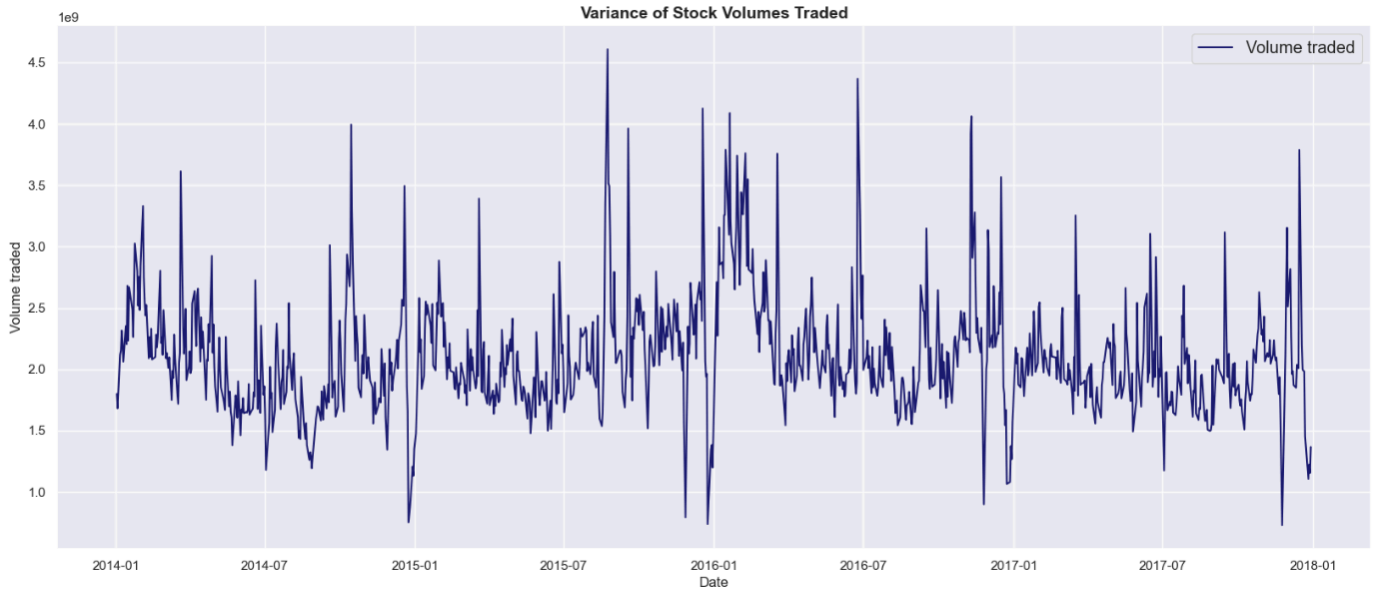
The analysis also showed that these two companies (Bank of America (BAC) and Apple (AAPL)), were also the two most traded companies across the whole time period, with a total of 89,988,444,028 and 45,485,758,169 shares traded respectively. At the opposite end of the spectrum, Aptiv PLC (APTV) and Brighthouse Financial Inc (BHF), were the lowest and second lowest traded companies by volume respectively, with just 122,667,236 and 36,306,643 shares traded respectively over the entire time period. This can be attributed to these two companies only joining the S&P 500 in December and July 2017 respectively. A table of this ranking is shown below:

company	volume
BAC	89988444028
AAPL	45485758169
GE	41734050117
AMD	33522535638
F	33144701045
...	...
HII	335472634
AZO	323425210
MTD	173123302
BHF	122667236
APTV	36306643

The lowest volume of trading in a day was 24th November 2017, with 728,261,080 shares traded. Perhaps unsurprisingly, the lowest 11 days of trading volume were all around either the Christmas or Thanksgiving periods. The lowest volume trading days are shown on the below table.

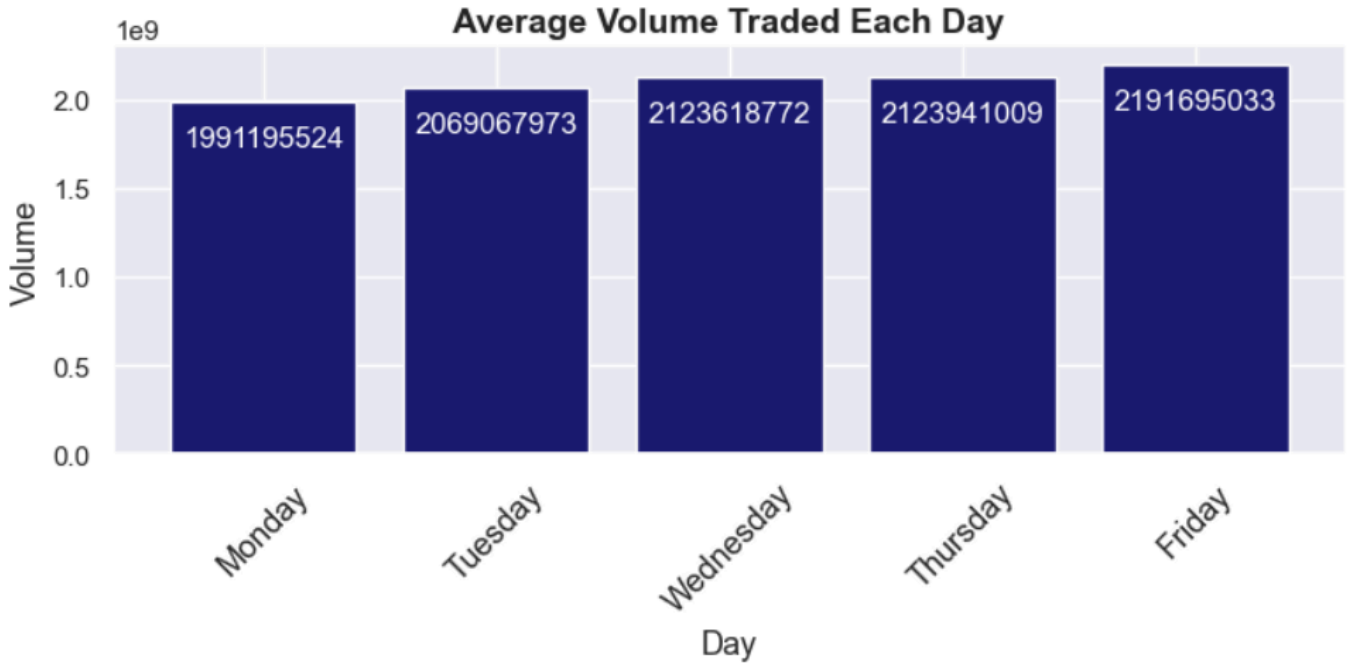
date	volume
2017-11-24	728261080
2015-12-24	736263173
2014-12-24	750895627
2015-11-27	791154818
2014-12-26	894908944
2016-11-25	898051561
2016-12-23	1063636294
2016-12-27	1080274822
2017-12-26	1104102103
2014-12-30	1130517899
2017-12-28	1152466365
2017-07-03	1173653906

The below line chart shows the huge and rapid variation from day to day in volumes of shares traded.



Volume by Day of the Week

The later the day of the week, the higher the average volume of shares traded: Fridays averaged 2,191,695,033 shares traded, Thursdays average 2,123,941,009, Wednesdays closely follow with an average of 2,123,618,772, Tuesdays average 2,069,067,974, and Mondays average just 1,991,195,525 shares traded. The below bar chart illustrates this.



% Day change

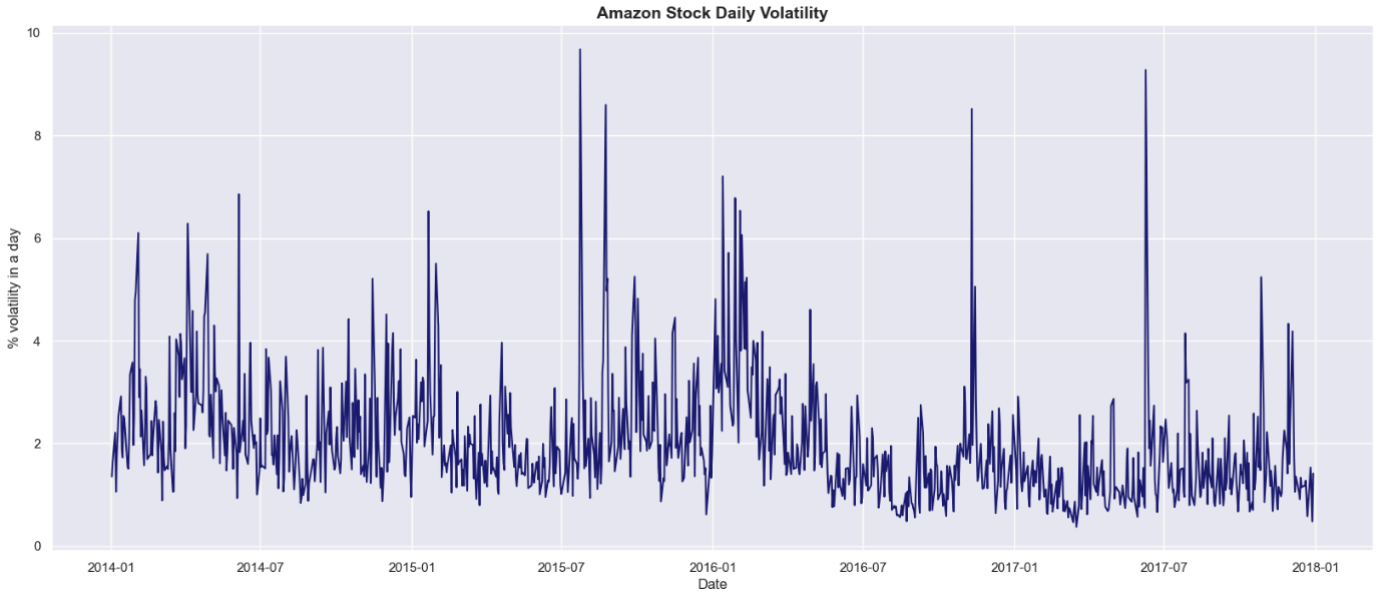
The highest percentage increase in share price in a day, (ie between open and close of trading), was Church & Dwight Co Inc (CHD) on the 19th May 2014, where the stock opened at 18.77 and closed at 33.90. An increase of 80.58%. Meanwhile the biggest loss in a day (from open to close) was Alliant Energy Corp (LNT) on the 19th May 2016, with an open of 35.19 and a close of 17.87, a drop of 49.22%. Further details are shown on the below table.

	company	date	day	open	close	abs_day_change	pc_day_change	low	high	abs_volatility	pc_volatility	volume
45548	CHD	2014-05-19	Monday	18.77	33.90	15.13	80.58	33.70	33.95	0.26	0.76	1078888
250551	WMB	2016-01-14	Thursday	13.42	18.29	4.87	36.29	13.24	18.44	5.20	39.27	42552128
266038	CHK	2016-03-02	Wednesday	2.62	3.40	0.78	29.77	2.60	3.75	1.15	44.23	76288029
266534	CHK	2016-03-03	Thursday	3.37	4.27	0.90	26.71	3.32	4.72	1.40	42.17	138475442
283843	AMD	2016-04-22	Friday	3.19	3.99	0.80	25.08	3.18	3.99	0.81	25.47	143265305
...	...	...	...	...	...	...	...	...	...	...	...	...
220270	PWR	2015-10-16	Friday	23.06	18.74	-4.32	-18.73	18.51	23.06	4.55	24.58	24411177
248764	FCX	2016-01-11	Monday	5.40	4.31	-1.09	-20.19	4.23	5.42	1.19	28.13	117668158
258102	CHK	2016-02-08	Monday	2.56	2.04	-0.52	-20.31	1.50	2.59	1.09	72.67	121984560
258487	WMB	2016-02-08	Monday	14.93	11.16	-3.77	-25.25	10.22	15.00	4.78	46.77	62368018
293534	LNT	2016-05-19	Thursday	35.19	17.87	-17.32	-49.22	35.09	35.78	0.70	2.00	1231722

Amazon Stock Volatility

The date where Amazon saw the highest percent volatility in a day ((high – low) / low * 100), was 24th July 2015 with a volatility of 9.68%, while the lowest percent volatility Amazon saw was 17th March 2017, with only 0.38% volatility across the day. Maven suggested the best method to calculate this metric would simply be high minus low, however, I would suggest the above percentage method is better as that takes into account the natural inflation of the stock over the years. The below table shows the top and bottom 5 days of percentage volatility for Amazon stock, and the below line chart shows how this metric varies over time.

	company	date	day	open	close	abs_day_change	pc_day_change	low	high	abs_volatility	pc_volatility	volume
190907	AMZN	2015-07-24	Friday	578.99	529.42	-49.57	-8.56	529.35	580.57	51.22	9.68	21909381
426013	AMZN	2017-06-09	Friday	1012.50	978.31	-34.19	-3.38	927.00	1012.99	85.99	9.28	7647692
201239	AMZN	2015-08-24	Monday	463.58	463.37	-0.21	-0.05	451.00	489.76	38.76	8.59	10097601
354017	AMZN	2016-11-10	Thursday	778.81	742.38	-36.43	-4.68	717.70	778.83	61.13	8.52	12746994
249616	AMZN	2016-01-13	Wednesday	620.88	581.81	-39.07	-6.29	579.16	620.88	41.72	7.20	7655239
...	...	...	...	...	...	...	...	...	...	...	...	...
330113	AMZN	2016-09-02	Friday	774.11	772.44	-1.67	-0.22	771.70	776.00	4.30	0.56	2181792
326129	AMZN	2016-08-23	Tuesday	763.31	762.45	-0.86	-0.11	761.00	764.70	3.70	0.49	1524131
496488	AMZN	2017-12-28	Thursday	1189.00	1186.10	-2.90	-0.24	1184.38	1190.10	5.72	0.48	1841676
394966	AMZN	2017-03-13	Monday	851.77	854.59	2.82	0.33	851.71	855.69	3.98	0.47	1909672
396966	AMZN	2017-03-17	Friday	853.49	852.31	-1.18	-0.14	850.64	853.83	3.19	0.38	3384403



I have calculated Amazon's absolute volatility via the requested method anyway. The highest day of absolute volatility for Amazon was 9th June 2017, with a price volatility of 85.99 dollars. This was a percentage volatility of 9.28% which is less than the 9.68% on the 24th July 2015, mentioned above. The day of lowest absolute volatility for Amazon was 19th August 2014, with 2.80 dollars of volatility (0.84%), which also does not beat the low percentage of 0.38% on the 17th March 2017. Top and bottom 5 days of absolute volatility for Amazon are shown on the below table.

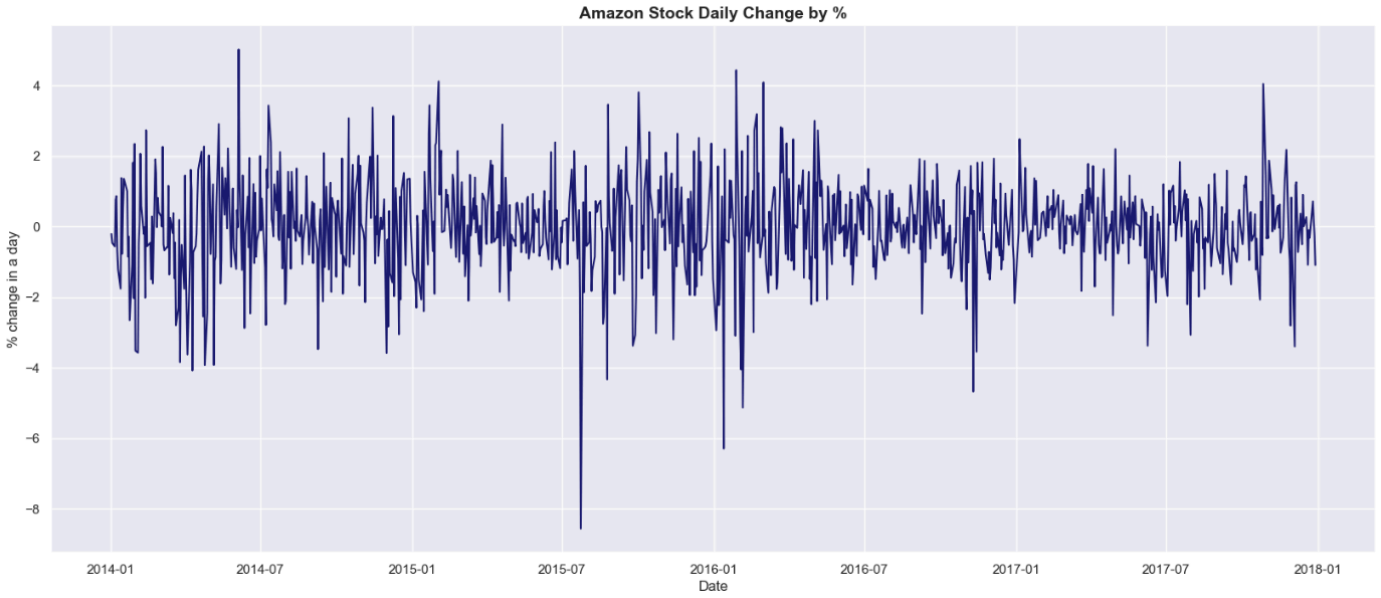
	company	date	day	open	close	abs_day_change	pc_day_change	low	high	abs_volatility	pc_volatility	volume
426013	AMZN	2017-06-09	Friday	1012.50	978.31	-34.19	-3.38	927.00	1012.99	85.99	9.28	7647692
354017	AMZN	2016-11-10	Thursday	778.81	742.38	-36.43	-4.68	717.70	778.83	61.13	8.52	12746994
475304	AMZN	2017-10-27	Friday	1058.14	1100.95	42.81	4.05	1050.55	1105.58	55.03	5.24	16565021
190907	AMZN	2015-07-24	Friday	578.99	529.42	-49.57	-8.56	529.35	580.57	51.22	9.68	21909381
486392	AMZN	2017-11-29	Wednesday	1194.80	1161.27	-33.53	-2.81	1145.19	1194.80	49.61	4.33	9257512
...	...	...	...	...	...	...	...	...	...	...	...	...
149299	AMZN	2015-03-24	Tuesday	373.99	374.09	0.10	0.03	372.27	375.24	2.97	0.80	2228214
121946	AMZN	2014-12-31	Wednesday	311.55	310.35	-1.20	-0.39	310.01	312.98	2.97	0.96	2057766
110722	AMZN	2014-11-26	Wednesday	333.78	333.57	-0.21	-0.06	331.75	334.65	2.90	0.87	1985949
50346	AMZN	2014-06-03	Tuesday	305.75	307.19	1.44	0.47	305.07	307.92	2.85	0.93	2379273
76587	AMZN	2014-08-19	Tuesday	334.87	335.13	0.26	0.08	333.01	335.81	2.80	0.84	1714120

<u>Amazon Day Gains and Losses (percent) ((close – open) / open * 100)</u>

The 5th June 2014 saw a 5.02% gain for Amazon stock, the largest percent gain (from open price to close price) in the dataset, while the biggest % loss in a day for Amazon was on the 24th July 2015, with a loss of 8.56%. Top and bottom 5 of this metric are shown in the below table.

	company	date	day	open	close	abs_day_change	pc_day_change	low	high	abs_volatility	pc_volatility	volume
51316	AMZN	2014-06-05	Thursday	308.10	323.57	15.47	5.02	306.90	327.94	21.04	6.86	7803760
254576	AMZN	2016-01-28	Thursday	608.37	635.35	26.98	4.43	597.55	638.06	40.51	6.78	14015171
132214	AMZN	2015-02-02	Monday	350.05	364.47	14.42	4.12	350.01	365.00	14.99	4.28	10231914
265488	AMZN	2016-03-01	Tuesday	556.29	579.04	22.75	4.09	556.00	579.25	23.25	4.18	5038452
475304	AMZN	2017-10-27	Friday	1058.14	1100.95	42.81	4.05	1050.55	1105.58	55.03	5.24	16565021
...	...	...	...	...	...	...	...	...	...	...	...	...
201731	AMZN	2015-08-25	Tuesday	487.49	466.37	-21.12	-4.33	466.25	489.44	23.19	4.97	5679329
354017	AMZN	2016-11-10	Thursday	778.81	742.38	-36.43	-4.68	717.70	778.83	61.13	8.52	12746994
257552	AMZN	2016-02-05	Friday	529.28	502.13	-27.15	-5.13	499.19	529.45	30.26	6.06	9708929
249616	AMZN	2016-01-13	Wednesday	620.88	581.81	-39.07	-6.29	579.16	620.88	41.72	7.20	7655239
190907	AMZN	2015-07-24	Friday	578.99	529.42	-49.57	-8.56	529.35	580.57	51.22	9.68	21909381

The below line chart shows the change in percentage day gain/loss, from day to day, this doesn't show any particular correlation over time, however it does show the huge variation in this metric from day to day, and it also shows that there have been bigger losses than gains, however naturally there will have been more days of gains than losses, as the stock substantially increased in value over the four years. This supports the theory of long-term investing being a more reliable way to make returns, than short term investing.



<u>Stocks added during the time period.</u>

The analysis showed that there are 22 stocks that joined the S&P 500 between 2nd Jan 2014 and 29th Dec 2017. These were APTV, BHF, BHGE, CFG, CSRA, DWDP, DXC, EVHC, FTV, GOOG, HLT, HPE, HPQ, INFO, KHC, NAVI, PYPL, QRVO, SYF, UA, WLTW, and WRK.

<u>If investing with Hindsight</u>

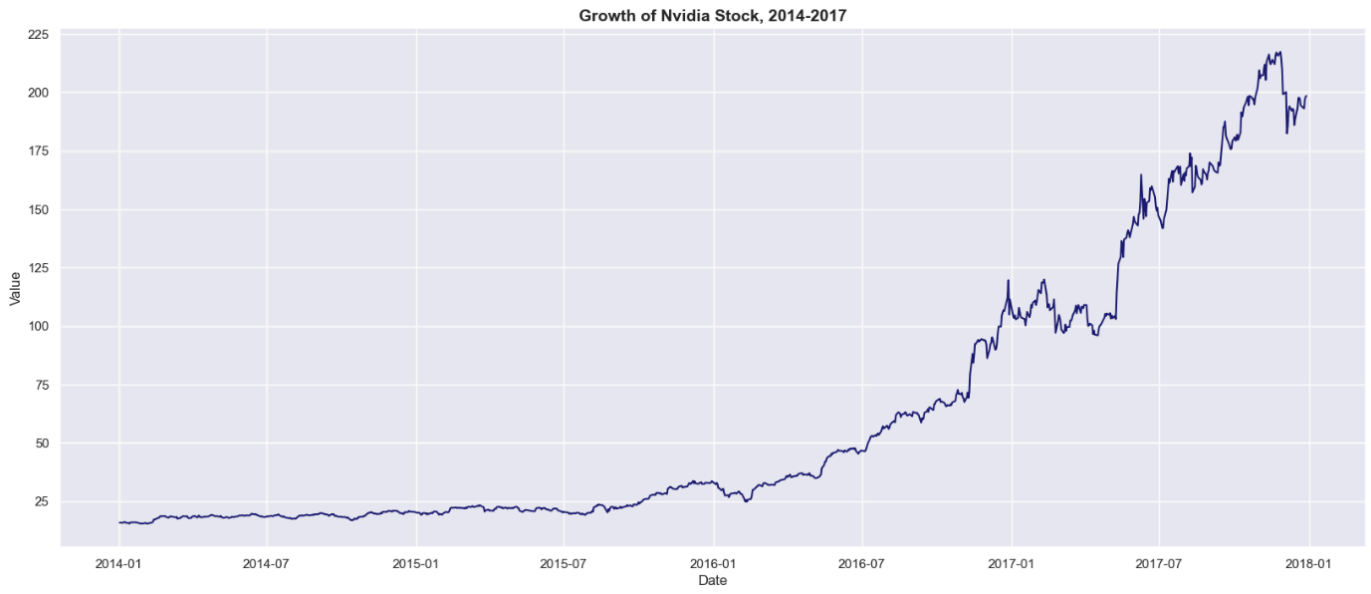
If I could go back in time to 2nd Jan 2014 and make an investment to be sold on the 29th Dec 2017, I would invest in Nvidia (NVDA) stock. Nvidia saw astronomical total growth of 1115.45% over the four years. This is commonly attributed to the increasing popularity of crypto-currency, which is best mined using graphics cards, which Nvidia specialises in manufacturing. No other company has come close to achieving this level of growth during this time period, the 2nd highest percent stock growth was Broadcom (AVGO) which achieved a still-huge growth of 386.09%. The worst

company to have invested in would be Chesapeake Energy (CHK), which lost 85.37% of its value over the four years. Further details of this metric are shown on the below table.

	company	date_x	open	date_y	close	price_difference	price_difference_pc
330	NVDA	2014-01-02	15.92	2017-12-29	193.50	177.58	1115.45
52	AVGO	2014-01-02	52.85	2017-12-29	256.90	204.05	386.09
146	EA	2014-01-02	22.90	2017-12-29	105.06	82.16	358.78
26	ALGN	2014-01-02	57.06	2017-12-29	222.19	165.13	289.40
318	NFLX	2014-01-02	52.40	2017-12-29	191.96	139.56	266.33
...	...	...	...	...	...	...	...
285	MAT	2014-01-02	47.57	2017-12-29	15.38	-32.19	-67.67
134	DISCK	2014-01-02	83.22	2017-12-29	21.17	-62.05	-74.56
133	DISCA	2014-01-02	90.21	2017-12-29	22.38	-67.83	-75.19
385	RRC	2014-01-02	83.13	2017-12-29	17.06	-66.07	-79.48
91	CHK	2014-01-02	27.07	2017-12-29	3.96	-23.11	-85.37

<u>Nvidia Growth</u>

The below line chart shows the exponential rate at which Nvidia's stock grew, the growth only took off in a major way in mid-2016, and the majority of the gains came in the final 18 months of the time period this analysis considers.

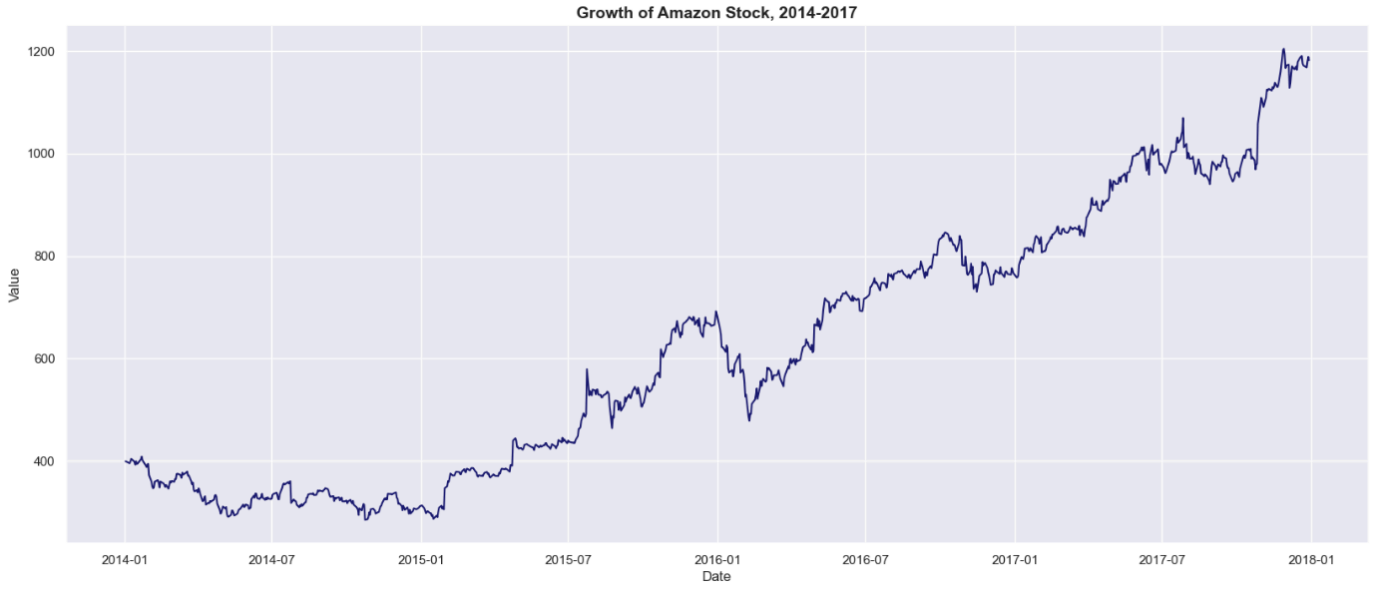


In terms of absolute price difference, Amazon (AMZN) saw the biggest stock price increase, of $770.67, however this only amounted to 193.25%, which while still incredible, is far less than Nvidia's growth. The top and bottom 5 companies in the ranking of this metric are shown on the below table.

	company	date_x	open	date_y	close	price_difference	price_difference_pc
38	AMZN	2014-01-02	398.80	2017-12-29	1169.47	770.67	193.25
344	PCLN	2014-01-02	1159.97	2017-12-29	1737.74	577.77	49.81
197	GOOGL	2014-01-02	558.29	2017-12-29	1053.40	495.11	88.68
310	MTD	2014-01-02	241.09	2017-12-29	619.52	378.43	156.97
156	EQIX	2014-01-02	177.63	2017-12-29	453.22	275.59	155.15
...	...	...	...	...	...	...	...
363	PRGO	2014-01-02	152.45	2017-12-29	87.16	-65.29	-42.83
385	RRC	2014-01-02	83.13	2017-12-29	17.06	-66.07	-79.48
133	DISCA	2014-01-02	90.21	2017-12-29	22.38	-67.83	-75.19
380	RL	2014-01-02	175.44	2017-12-29	103.69	-71.75	-40.90
101	CMG	2014-01-02	530.00	2017-12-29	289.03	-240.97	-45.47

Amazon Growth

In contrast, while Amazon's growth has been bumpy in the short term, over the longer-term Amazon has grown much more consistently than Nvidia. This can be seen from the below chart of Amazon's growth.



Act

Recommendations

It is no secret that predicting the future of the stock market with 100% reliability is impossible. That said, there is still much to learn from the data, which will enable us to better inform future decisions.

This analysis cannot recommend particular stocks to invest in, but it can anticipate average long term growth if invested across the index in similar market conditions. This information can be used for a range of financial planning purposes, such as calculating expected returns over a period of time when invested into an S&P 500 index tracker fund. This analysis has shown that the average stock on the S&P 500 between 2014 - 2017 grew by 49.3%. That's an average growth of 12.33% per year. This information could be used to calculate the approximate growth of a pension investment before a given retirement date, painting a picture of what an investor's retirement will look like. It is important to note that investments should be made for the long term to maximise the chance of predictable returns. Short term investments are far less predictable. It's also important to note that anything can happen on the stock market, crashes can and do happen, and investors can get back less than they invested.

The analysis within this report surrounding average daily trading volume could help a trading platform schedule workers such as Technical Support Engineers onto shifts. Based on this analysis I would recommend workers are scheduled on shifts evenly across the week, as while there is a small change in average volume depending on the day of the week, there are many anomalies frequently occurring within the data, so it would not be wise to schedule fewer workers on shift on a Monday, say, just because it's the day with the lowest average volume traded.

Summary of answers to the specific business questions:

1. Which date in the sample saw the largest overall trading volume? On that date, which two stocks were traded most?
   24th August 2015 with 4,607,945,196 shares traded.
   Bank of America (BAC) and Apple (AAPL), with 214,649,482, and 162,206,292 shares traded respectively.
2. On which day of the week does volume tend to be highest? Lowest?
   Highest: Fridays, average 2,191,695,033 shares traded.
   Lowest: Mondays, average 1,991,195,525 shares traded.
3. On which date did Amazon (AMZN) see the most volatility, measured by the difference between the high and low price?
   Highest *percent* volatility: 24th July 2015: volatility of 9.68% / $51.22.
   Highest *absolute* volatility: 9th June 2017: volatility of 9.28% / $85.99.
4. If you could go back in time and invest in one stock from 2nd January 2014 – 29th December 2017, which would you choose? What % gain would you realize?
   Nvidia: 1115.45% increase ($177.58 increase per share.)
   (Highest absolute price growth: Amazon (AMZN): 193.25%, increase of $770.67 per share.)