DATA ANALYTICS BOOTCAMP

EDA PROJECT 1

WRITTEN

REPORT

Analyzing the impact of COVID-19 on stock market returns and taking a deep-dive into the healthcare sector.

PREPARED BY GROUP 4
ERJOLA, GAYATRI & VINCENT

JULY 2023

Table of Contents

Objective

COVID-19 was one of the most impactful events of the century, with long-lasting effects on the economy and capital markets. In this analysis, we wanted to find out how COVID affected the stock market, and specifically see how the healthcare industry was impacted. Many believed that the healthcare sector flourished in a post-covid world, and we wanted test if the healthcare sector is truly a good investment in times of healthcare crises.

In order to tackle our objective, we broke down our analysis to the following questions:

#	Questions
1.	Which healthcare stocks performed best before and during COVID-19?
2.	Are healthcare sector returns significantly different before and during/after COVID-19?
3.	Are market returns significantly different before and during/after COVID-19?
4.	Is the impact on the healthcare sector significantly different that the market?
5.	Did the healthcare sector outperform the market during/after COVID-19?
6.	How does market capitalization impact returns in the healthcare sector?
7.	Looking at 2 ETFs, SPY and IYH as benchmarks, how do prices and returns differ?

Methodology

This section will describe the data collection methods, our ideation, terminology and scope of the analysis.

Terminology

We defined 'Pre-Covid' as the 3 year period starting 1st of January, 2017 to 31st of December, 2019.
We defined 'Post-Covid' as the 3 year period starting 1st of January, 2020 to 31st of December, 2022.
The market, in our report, refers to North American, publicly traded stock market.
The healthcare sector, in our report, refers to North American, publicly traded companies under the GICS sector 'Healthcare'.
Return in our analysis is defined as (change in stock price/initial stock price)*100. It is in percentage. This will be our measure for stock performance.

Data Sources and Collection Methods

In order to get a list of healthcare companies, we used stockmarketbma.com, and downloaded 10 CSVs with list of companies from the following Sub-Sectors:

1. Biotechnology
2. Drug Manufacturing - General
3. Drug Manufacturing - Special & Generic
4. Diagnostic Research
5. Health Information Service
6. Health Care Plans
7. Medical Care Facilities
8. Medical Devices
9. Medical Distributions
10. Medical Instruments & Supplies

We combined, and cleaned the data to only include companies under the Healthcare GICS Sector.

In order to get a list of companies that represent the market, we used the list of companies that make up the Standard & Poor 500. We sourced the list from wikipedia.com and downloaded it as a CSV file. We believe that this list is a good sample to represent the market as it represents approximately 80% of the total value of the U.S. stock market and includes companies across all sectors. It is frequently used as a benchmark for the market as a whole.

In addition, to plot price and returns over time, we picked 2 ETFs to represent the market and healthcare sector.

For the market, we are using the SPDR S&P 500 ETF Trust with ticker symbol SPY. This is an ETF that tracks the S&P 500 and we believed it was the obvious choice.

For the healthcare sector, we are using the iShares U.S. Healthcare ETF with the ticker symbol IYH. This is an ETF that tracks the performance of the healthcare sector of the U.S. equity market, by including a variety of subsectors including Biotechnology, Pharmaceuticals, Medical Devices etc.

In order to source stock information such as Price, Trading Volume, Market Capitalization etc. we used a pre-existing python package called yahoo_fin. This package parses through yahoo finance websites and pulls the data for analysis. We build data frames on Python by using the lists we found above and retrieving stock information on each ticker symbol.

In order to further clean our data, we made sure to remove companies that were not publicly traded during our analysis period, i.e. newer companies that may have had their IPOs after January 1st, 2017 and we removed companies that went out of business during our analysis period. We ended up with 818 healthcare stocks, and 498 stocks to represent the market.

This process can be reviewed in the folder named 'Resources' in our repository.

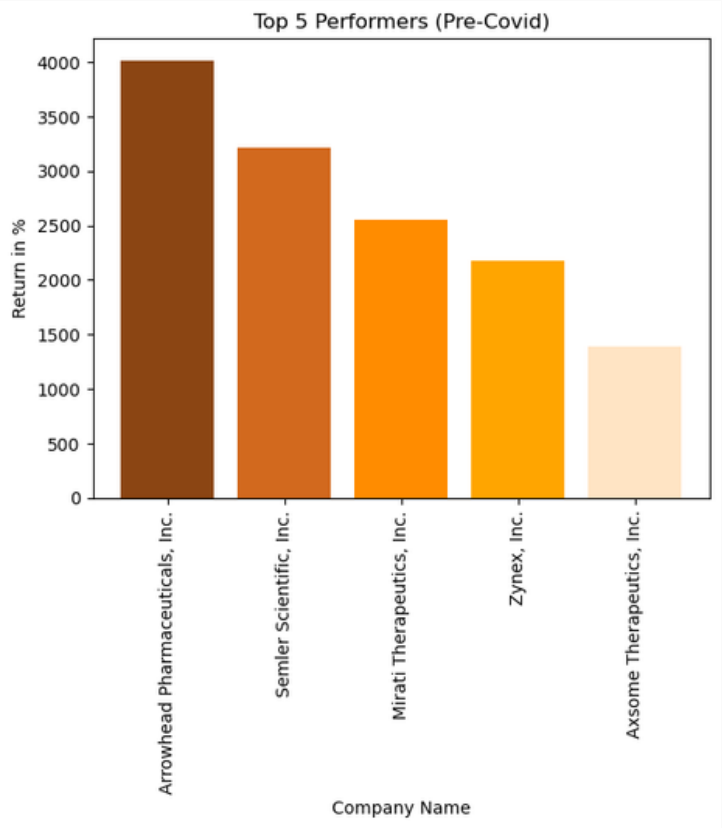
Question 1

Which healthcare stocks performed best before and during COVID-19?

We wanted to see which companies had the highest returns in the pre-covid and post-covid period in the healthcare sector. The results are as follows.

Top Performers - Pre Covid

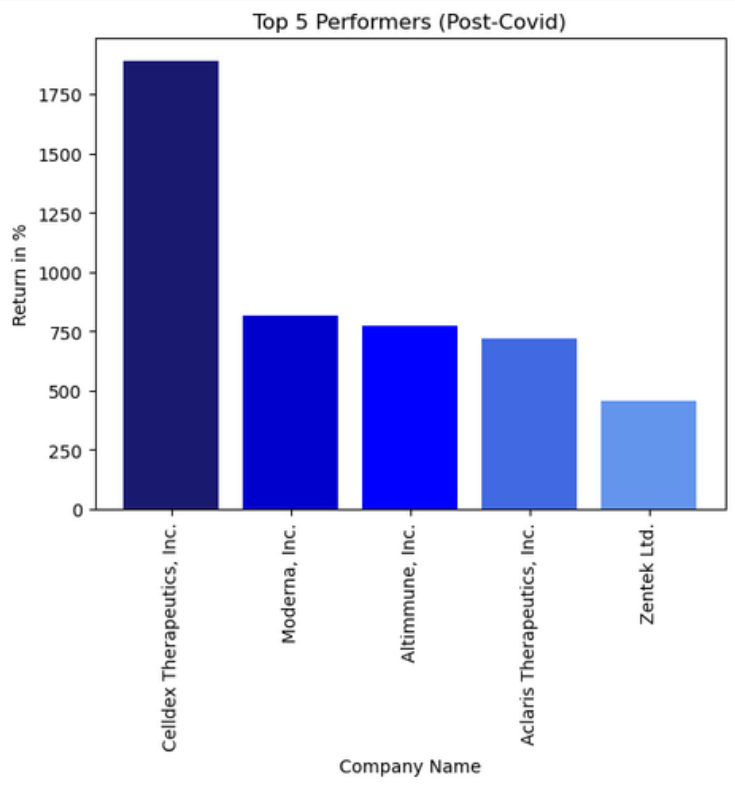
	Company Name	Symbol	GICS sector	Return from 2017-2019 (Pre-Covid)	Market Cap Type	Market Cap	Average Volume
40	Arrowhead Pharmaceuticals, Inc.	ARWR	Health Care	4,015.38%	Mid cap	$2,564,784,660.0	1,425,556
726	Semler Scientific, Inc.	SMLR	Health Care	3,211.03%	Micro cap	$175,992,505.0	37,265
46	Mirati Therapeutics, Inc.	MRTX	Health Care	2,548.04%	Mid cap	$2,250,126,669.0	985,251
711	Zynex, Inc.	ZYXI	Health Care	2,174.28%	Small cap	$443,137,573.0	382,678
36	Axsome Therapeutics, Inc.	AXSM	Health Care	1,383.29%	Mid cap	$2,698,458,227.0	1,033,635



The top performers in the pre-covid period are from all different market cap types. Arrowhead Pharmaceuticals Inc. is the top performer with a return of ~4,015%, followed by Semler Scientific Inc. with a return of ~3,211%. Arrowhead Pharmaceuticals is a Biotechnology company and Semler Scientific is a Medical Devices company.

Top Performers - Pre Covid

	Company Name	Symbol	GICS sector	Return from 2020-2022 (Post-Covid)	Market Cap Type	Market Cap	Average Volume
57	Celldex Therapeutics, Inc.	CLDX	Health Care	1,889.73%	Small cap	$1,656,972,334	591,048
541	Moderna, Inc.	MRNA	Health Care	817.83%	Large cap	$1,793,425,188	3,453,513
191	Altimmune, Inc.	ALT	Health Care	770.37%	Micro cap	$206,478,428	1,312,393
114	Aclaris Therapeutics, Inc.	ACRS	Health Care	720.31%	Small cap	$543,547,657	926,279
805	Zentek Ltd.	ZTEK	Health Care	453.57%	Micro cap	$150,296,313	35,060



The top performers in the post-covid period are from all different market cap types. Celldex Therapeutics Inc. is the top performer with a return of ~1,890%, followed by Moderna Inc. with a return of ~818%. Celldex Therapeutics and Moderna are both Biotechnology companies. We know Moderna is one of the major producers of the COVID-19 vaccine, which is likely responsible for their high returns.

Question 2

Are healthcare sector returns significantly different before and during/after COVID-19?

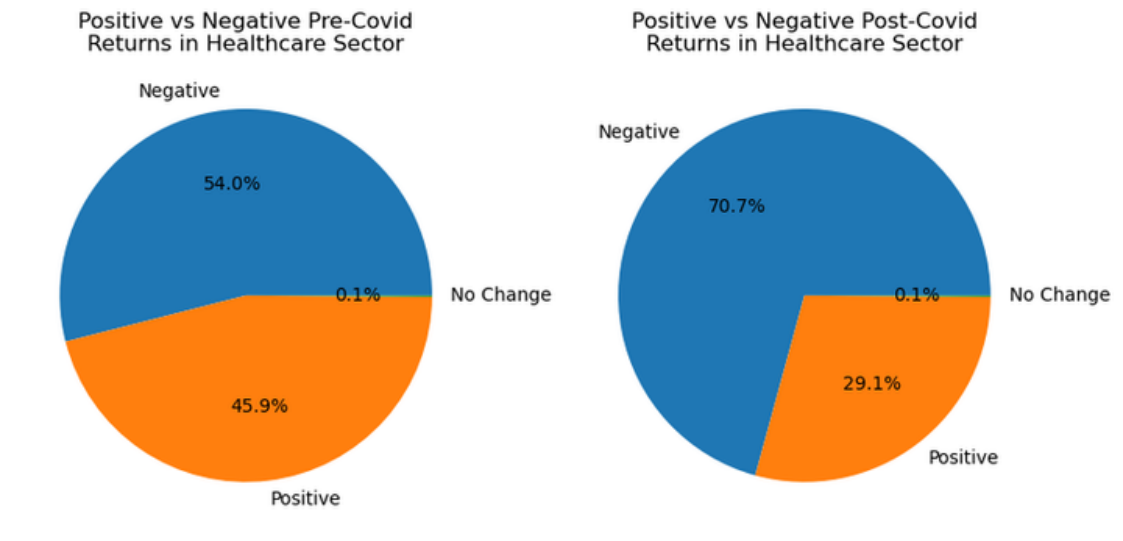
We started by looking the summary statistics to get a birds-eye view of whether the returns are different before and after covid. Then, a two-sample t-test was conducted to measure if there is a statistically significant difference between 'pre-covid' healthcare sector returns and 'post-covid' healthcare sector returns.
We first did the analysis with the full dataset, and then we removed outliers and did the process again. Note that the returns are reported in percentages.

Healthcare Sector - Summary Statistics

	Summary Statistics for Returns of Healthcare Stocks, Pre-Covid
Mean	45.78
Median	-7.39
Standard Deviation	267.67
Variance	71648.54

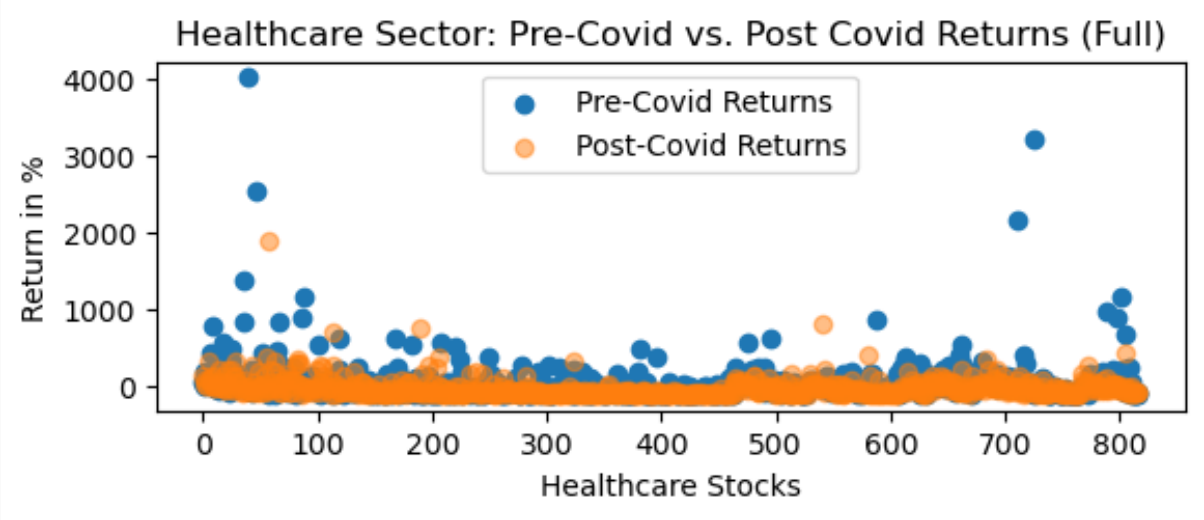
	Summary Statistics for Returns of Healthcare Stocks, Post-Covid
Mean	-16.37
Median	-53.11
Standard Deviation	121.20
Variance	14688.63

Right of the bat, we can see that the average return for the healthcare sector, pre-covid, was 45.78%, which is much higher than the average return in the post-covid period, -16.37%. When looking at the summary statistics, we would say that there is a difference in returns, and that Healthcare stocks performed better before COVID-19. We can see that there is more variation among returns in the pre-covid period, with a standard deviation of 267.67% versus just 121.20%. We can see that in both cases, the median return is lower than the mean return, which means the data is likely positively skewed.

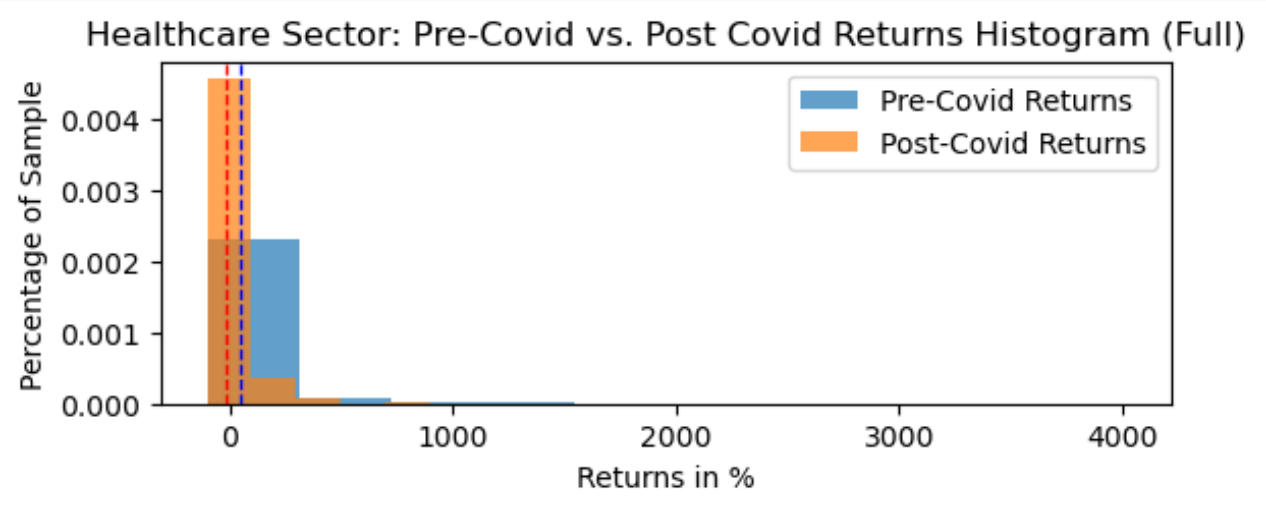


In this pie-chart it can be observed that in the pre-covid period, only 54% of companies experienced negative returns, but in the post-covid period, this number goes up to 70.7%. It would, once again, appear that COVID had a negative impact on returns.

Full Dataset - Plots



Pictured above is a scatter plot of healthcare sector returns. It can be observed that there is more variation in pre-covid returns and it also can be observed that there are higher data points for pre-covid returns. This implies that healthcare stocks performed better in the pre-covid period, and confirms our findings in the summary statistics.



In this histogram plot, it can be observed that there is more variation in pre-covid returns and it also can be observed that the average pre-covid return (marked by the blue line) is higher than the average post-covid return (marked by the red line). We can also see in this graph that many of the post-covid data points hover around 0. Another observation is that the histogram is highly positively skewed. The histogram confirms our findings from the summary statistics.

Full Dataset - Hypothesis Testing

In the graphs, its clear to see that healthcare sector performance is better in the pre-covid period, than the post-covid period. Now we need to test if this difference is statistically significant or just a coincidence. The hypotheses being tested are as follows:

Null Hypothesis:
μ(returns of healthcare pre-covid) = μ(returns of healthcare post covid)

Alternative Hypothesis:
μ(returns of healthcare pre-covid) ≠ μ(returns of healthcare post covid)

Where μ represents the sample mean. We ran a two-sample t-test, performing Welch's t-test, as the variances of the two samples are not equal and it produced the following results:

Ttest_indResult(statistic=6.118763711063338, pvalue=1.2961677176188562e-09)

The p-value is ~0.00 which is significant at every level, therefore we can reject the null hypothesis. The results imply that there is a significant difference in pre-covid and post-covid healthcare returns and that COVID-19 had a negative impact on the returns of healthcare sector.

Outlier Analysis

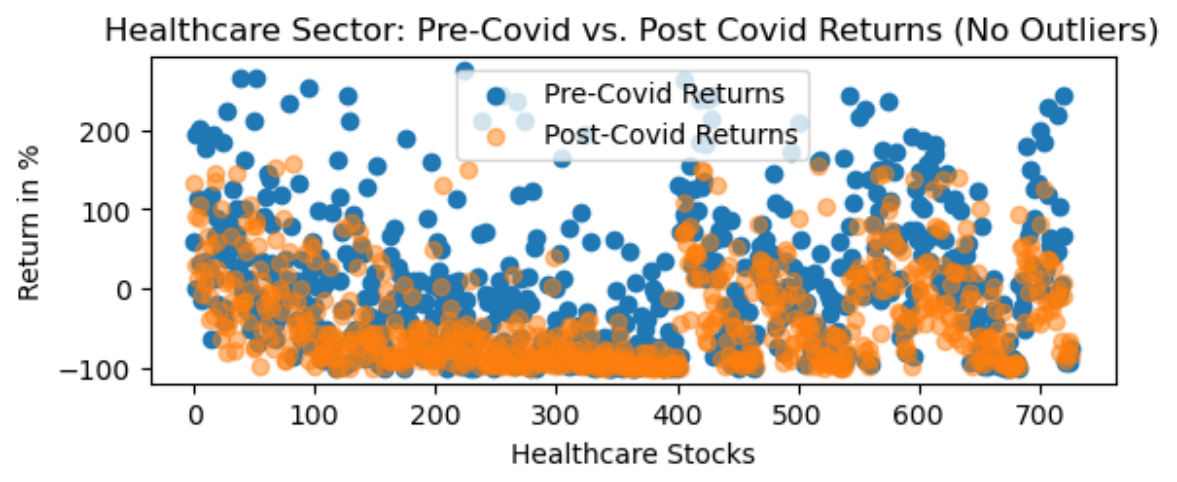
We wanted to remove the outliers in the data and reconduct the tests. We started by looking at summative statistics about quartiles for pre-covid and post-covid data. These were the results:

The lower quartile of pre-covid returns is: -69.69 %
The upper quartile of pre-covid returns is: 67.27 %
The interquartile range of pre-covid returns is: 136.96 %
The the median of pre-covid returns is: -7.39 %
The lower bound of pre-covid returns is: -275.13 %
The upper bound of pre-covid returns is: 272.71 %

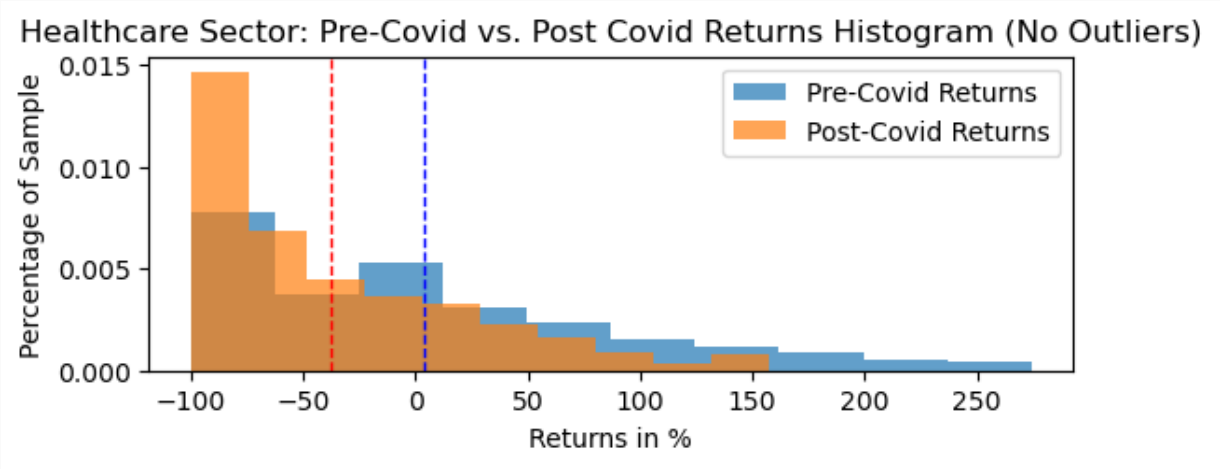
The lower quartile of post-covid returns is: -84.63 %
The upper quartile of post-covid returns is: 13.02 %
The interquartile range of post-covid returns is: 97.65 %
The the median of post-covid returns is: -53.11 %
The lower bound of post-covid returns is: -231.11 %
The upper bound of post-covid returns is: 159.5 %

We then removed the outliers from the data set. There were 90 outliers, leaving us with a new dataset of 727 companies.

Dataset Without Outliers - Plots



Pictured above is scatter plot of healthcare returns with no outliers. It can be observed that there is more variation in pre-covid returns and it also can be observed that there are higher data points for pre-covid returns. This implies that healthcare stocks performed better in the pre-covid period.



Pictured above is histogram plot of healthcare returns not including outliers. It can be observed that there is more variation in pre-covid returns and it also can be observed that the average pre-covid return (marked by the blue line) is higher than the average post-covid return (marked by the red line). It can be observed that most healthcare companies experienced negative returns in the post-covid period. Another observation is that the histogram is positively skewed.

Dataset Without Outliers - Hypothesis Testing

In the graphs, its clear to see that healthcare sector performance is better in the pre-covid period, than the post-covid period, even without outliers. Now we need to test if this difference is statistically significant or just a coincidence. The hypotheses being tested are as follows:

Null Hypothesis:
μ(returns of healthcare pre-covid) = μ(returns of healthcare post covid)

Alternative Hypothesis:
μ(returns of healthcare pre-covid) ≠ μ(returns of healthcare post covid)

Where μ represents the sample mean. We ran a two-sample t-test, performing Welch's t-test, as the variances of the two samples are not equal and it produced the following results:

Ttest_indResult(statistic=10.426609499942277, pvalue=1.7037579766095215e-24)

The new results yield an even smaller p-value, which is ~0.00 and is significant at every level, therefore we can reject the null hypothesis. The results imply that there is a significant difference in pre-covid and post-covid healthcare returns and that COVID-19 had a negative impact on the returns of healthcare sector.

Question 3

Are market returns significantly different before and during/after COVID-19?

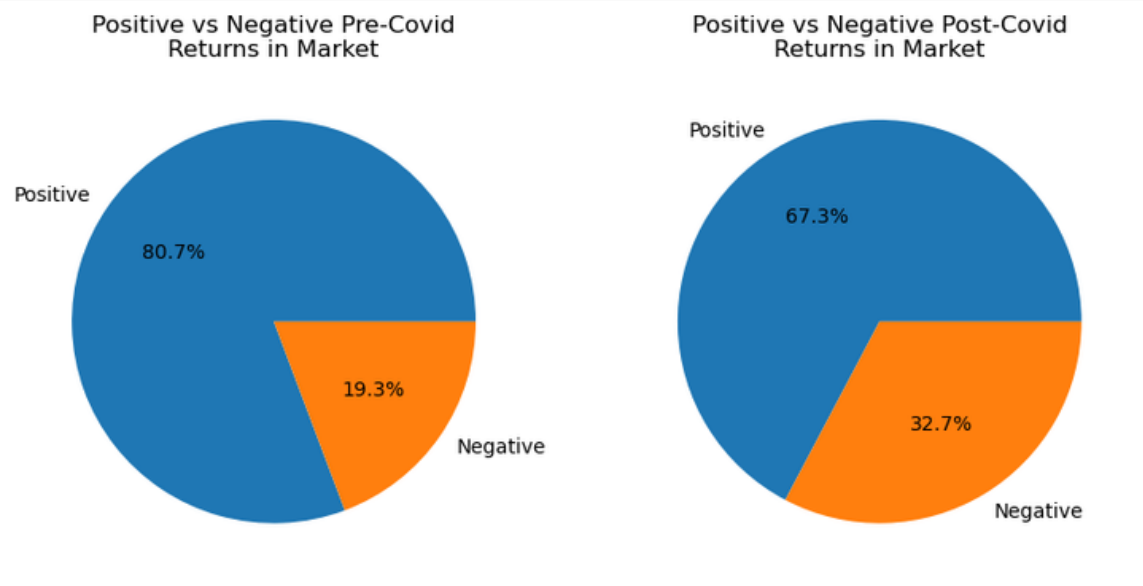
We started by looking the summary statistics to get a birds-eye view of whether the returns are different before and after covid. Then, , we conducted a two-sample t-test to measure if there is a statistically significant difference between 'pre-covid' market returns and 'post-covid' market returns. We first did the analysis with the full dataset, and then we removed outliers and did the process again. Note that the returns are reported in percentages.

Market - Summary Statistics

	Summary Statistics for Returns of Market, Pre-Covid
Mean	59.35
Median	42.98
Standard Deviation	127.02
Variance	16133.59

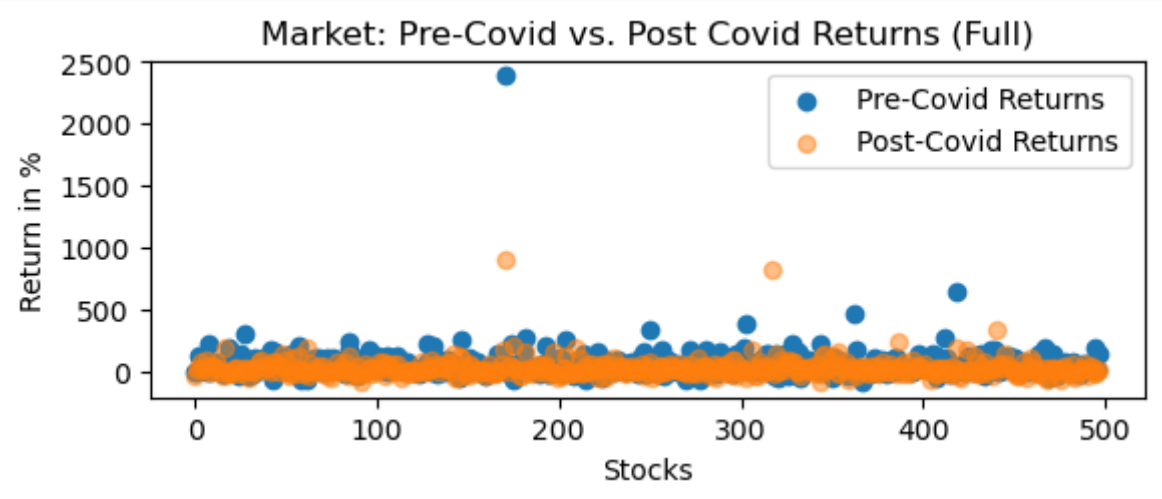
	Summary Statistics for Returns of Market, Post-Covid
Mean	27.95
Median	18.74
Standard Deviation	72.20
Variance	5212.14

Right of the bat, we can see that the average return for the market, pre-covid, was 59.35%, which is much higher than the average return in the post-covid period, 27.95%. When looking at the summary statistics, we would say that there is a difference in returns, and that the market performed better before COVID-19. We can see that there is more variation among returns in the pre-covid period, with a standard deviation of 127.02% versus just 72.20%. We can see that in both cases, the median return is lower than the mean return, which means the data is likely positively skewed.

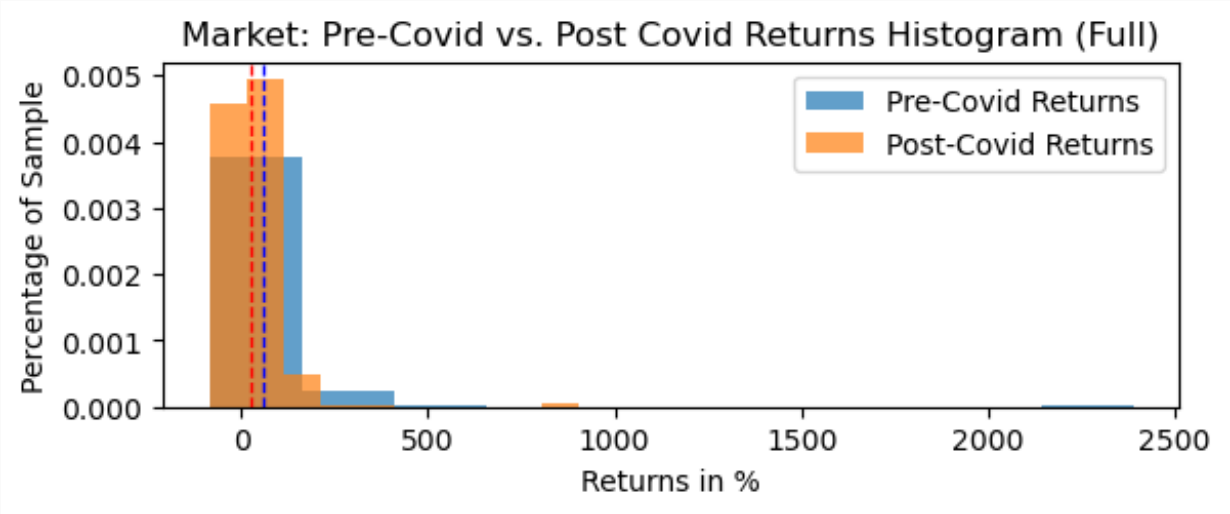


In this pie-chart it can be observed that in the pre-covid period, only 19.3% of companies experienced negative returns, but in the post-covid period, this number goes up to 32.7%. It would, once again, appear that COVID had a negative impact on returns.

Full Data Set - Plots



Pictured above is scatter plot of market returns. It can be observed that there is more variation in pre-covid returns and it also can be observed that there are higher data points for pre-covid returns. This implies that in general, stocks performed better in the pre-covid period, and confirms our findings in the summary statistics.



Pictured above is histogram plot of market returns. It can be observed that there is more variation in pre-covid returns and it also can be observed that the average pre-covid return (marked by the blue line) is higher than the average post-covid return (marked by the red line). We can also see in this graph that many of the post-covid data points hover around 0. Another observation is that the histogram is positively skewed. The histogram confirms our findings from the summary statistics.

Full Dataset - Hypothesis Testing

In the graphs, its clear to see that the market performance is better in the pre-covid period, than the post-covid period. Now we need to test if this difference is statistically significant or just a coincidence. The hypotheses being tested are as follows:

Null Hypothesis:
μ(returns of market pre-covid) = μ(returns of market post covid)

Alternative Hypothesis:
μ(returns of market pre-covid) ≠ μ(returns of market post covid)

Where μ represents the sample mean. We ran a two-sample t-test, performing Welch's t-test, as the variances of the two samples are not equal and it produced the following results:

Ttest_indResult(statistic=4.792199799245037, pvalue=1.9716443542725045e-06)

The p-value is ~0.00 which is significant at every level, therefore we can reject the null hypothesis. The results imply that there is a significant difference in pre-covid and post-covid market returns and that COVID-19 had a negative impact on the returns of the market as a whole.

Outlier Analysis

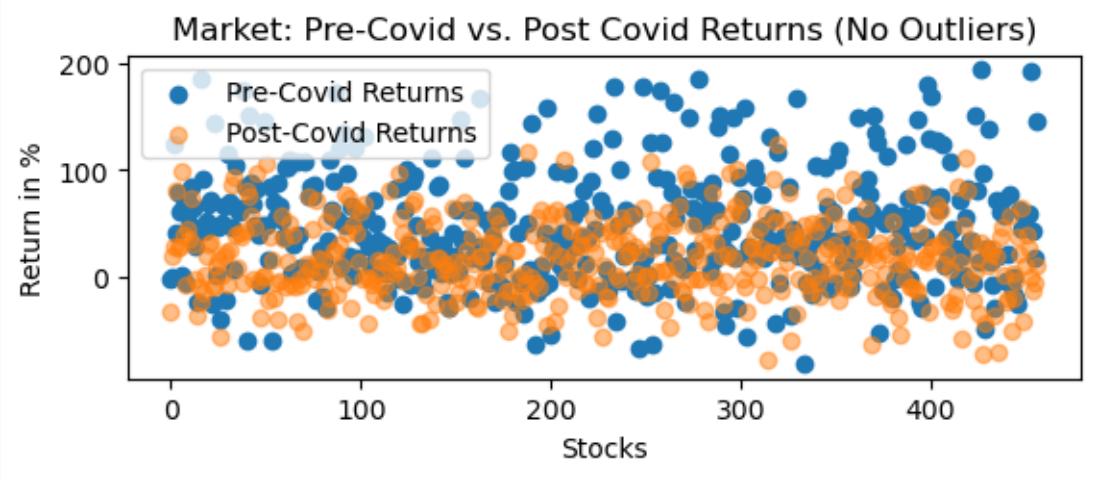
We wanted to remove the outliers in the data and reconduct the tests. We started by looking at summative statistics about quartiles for pre-covid and post-covid data. These were the results:

The lower quartile of pre-covid returns is: 8.51 %
The upper quartile of pre-covid returns is: 83.85 %
The interquartile range of pre-covid returns is: 75.34 %
The the median of pre-covid returns is: 42.98 %
The lower bound of pre-covid returns is: -104.5 %
The upper bound of pre-covid returns is: 196.86 %

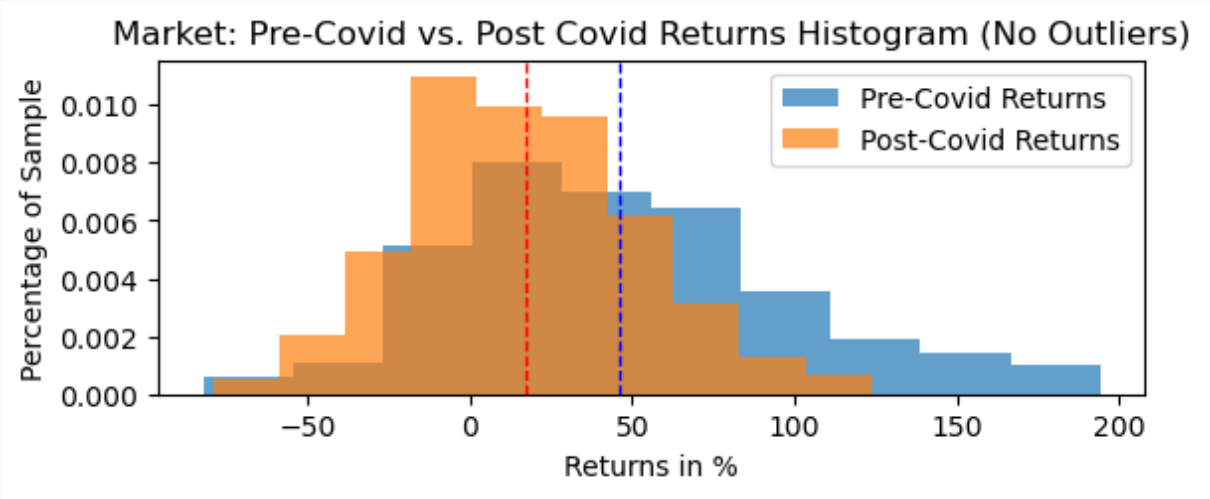
The lower quartile of post-covid returns is: -6.12 %
The upper quartile of post-covid returns is: 45.93 %
The interquartile range of post-covid returns is: 52.05 %
The the median of post-covid returns is: 18.74 %
The lower bound of post-covid returns is: -84.18 %
The upper bound of post-covid returns is: 124.0 %

We then removed the outliers from the data set. There were 41 outliers, leaving us with a new dataset of 457 companies.

Dataset Without Outliers - Plots



Pictured above is scatter plot of market returns with no outliers. It can be observed that there is more variation in pre-covid returns and it also can be observed that there are higher data points for pre-covid returns. This implies that the market performed better in the pre-covid period.



Pictured above is histogram plot of market returns not including outliers. It can be observed that there is more variation in pre-covid returns and it also can be observed that the average pre-covid return (marked by the blue line) is higher than the average post-covid return (marked by the red line). Another observation is that the histogram is positively skewed.

Dataset Without Outliers - Hypothesis Testing

In the graphs, its clear to see that market, in general performed better in the pre-covid period, than the post-covid period, even without outliers. Now we need to test if this difference is statistically significant or just a coincidence. The hypotheses being tested are as follows:

Null Hypothesis:
μ(returns of market pre-covid) = μ(returns of market post covid)

Alternative Hypothesis:
μ(returns of market pre-covid) ≠ μ(returns of market post covid)

Where μ represents the sample mean. We ran a two-sample t-test, performing Welch's t-test, as the variances of the two samples are not equal and it produced the following results:

Ttest_indResult(statistic=9.774880979588726, pvalue=2.1189031494824718e-21)

The new results yield an even smaller p-value, which is ~0.00 and is significant at every level, therefore we can reject the null hypothesis. The results imply that there is a significant difference in pre-covid and post-covid returns and that COVID-19 had a negative impact on the returns of the market as a whole.

Question 4

Is the impact on the healthcare sector significantly different than the market?

In our previous analyses we saw that COVID-19 had a negative impact on both the market as a whole and the healthcare sector as well. In this question, we want to understand, if the impact of COVID on the healthcare sector is different than on the market as a whole. In other words, did healthcare stocks underperform the same way all other sectors did, or did healthcare stocks get impacted differently by COVID. We will be using Difference-in-Differences (or diff-in-diff) testing to evaluate this.

In order to do this statistical test, we first need to find the difference in returns in 'post-covid' and 'pre-covid' periods for both the market and the healthcare sector. This is simply done as:

Post-Covid Return - Pre-Covid Return

Note that since the returns are in percentages, the difference in returns will be in percentage points.

Difference in Returns - Summary Statistics

	Difference in Returns
mean	-31.41
median	-25.21
var	12497.78
std	111.79

Difference in Returns: Market

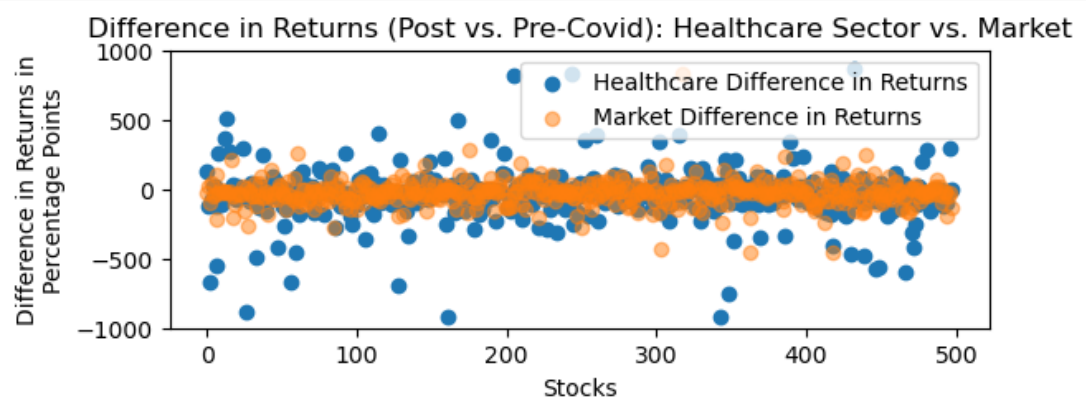
	Difference in Returns
mean	-62.15
median	-27.58
var	85342.85
std	292.13

Difference in Returns: Healthcare sector

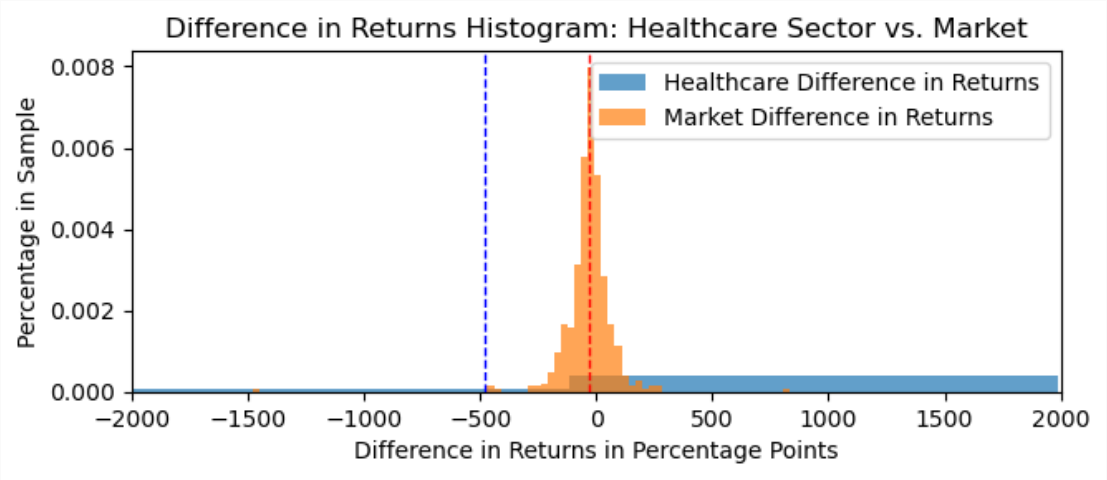
Pictured above are the summary statistics for the calculated difference in returns for market and healthcare sector. From just looking at the above tables, it can be seen that on average the drop in performance is much higher in the healthcare sector, at ~ -62 percentage points, versus just ~ -31 percentage points. It would appear that COVID impacted the healthcare sector far more than the market. We should note that there is also more variation in the differences in returns in the healthcare sector.

Difference in Returns - Plots

In order to compare the market and the healthcare sector, we first wanted to ensure that both sample sizes are the same. We used random sampling to generate a sample of 498 healthcare stocks, to compare to 498 stocks that represent the market.



Pictured above is a scatter plot of differences in returns for healthcare and for the market. It can be observed that there is more variation in healthcare differences, which makes sense as the standard deviation for healthcare is much higher than the market.



Pictured above is a histogram of differences in returns for healthcare and for the market. It can be observed that there is more variation in healthcare differences, with the histogram being extremely spread out for healthcare, and forming a bell shape for the market. The average difference as we saw in the tables above is lower for healthcare, compared to the market.

Diff-in-Diff - Hypothesis Testing

In the graphs above, it's clear to see that the difference in returns is much greater amongst healthcare companies relative to the market. Now we need to test if this difference is statistically significant or just a coincidence. The hypotheses being tested are as follows:

Null Hypothesis:

μ(returns of healthcare post covid) - μ(returns of healthcare pre covid)

=

μ(returns of market post covid) - μ(returns of market pre covid)

Alternative Hypothesis:

μ(returns of healthcare post covid) - μ(returns of healthcare pre covid)

≠

μ(returns of market post covid) - μ(returns of market pre covid)

Where μ represents the sample mean. We ran a two-sample t-test, performing Welch's t-test, as the variances of the two samples are not equal and it produced the following results:

Ttest_indResult(statistic=-1.0773971070380404, pvalue=0.2818254083998124)

The p value is ~0.28 which is not significant at any standard level, therefore we fail to reject the null. This implies that the impact of COVID-19 on healthcare sector returns is not significantly different than the impact on the rest of the market. How can that be, given that the mean difference in return for healthcare is higher than the market? This is explained by the high variation / standard deviation in healthcare returns compared to the market.

Question 5

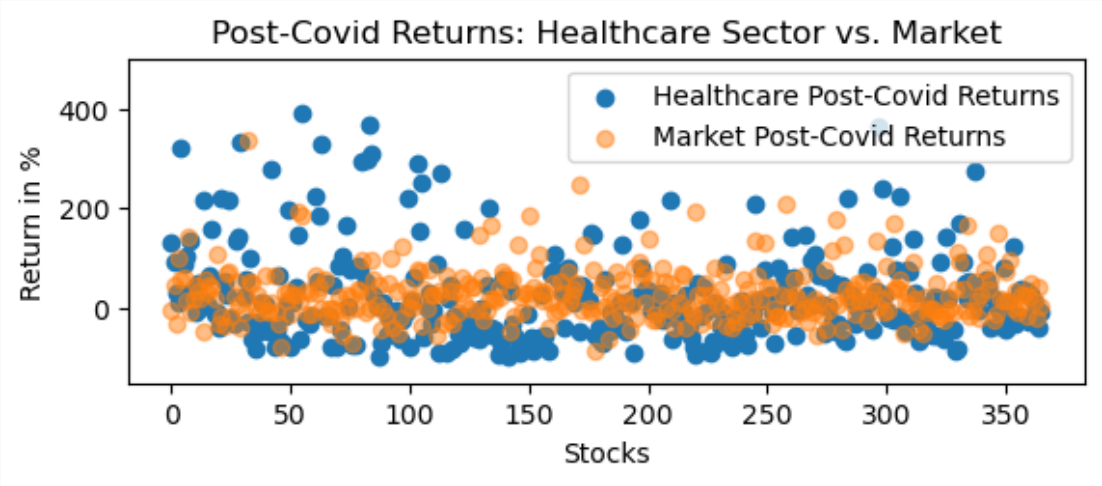
Did the healthcare sector outperform the market during/after COVID-19?

In our previous analyses we saw that COVID-19 had a negative impact on both the market as a whole and the healthcare sector as well. In this question, we want to understand if the healthcare sector outperformed the market in the post-covid period.

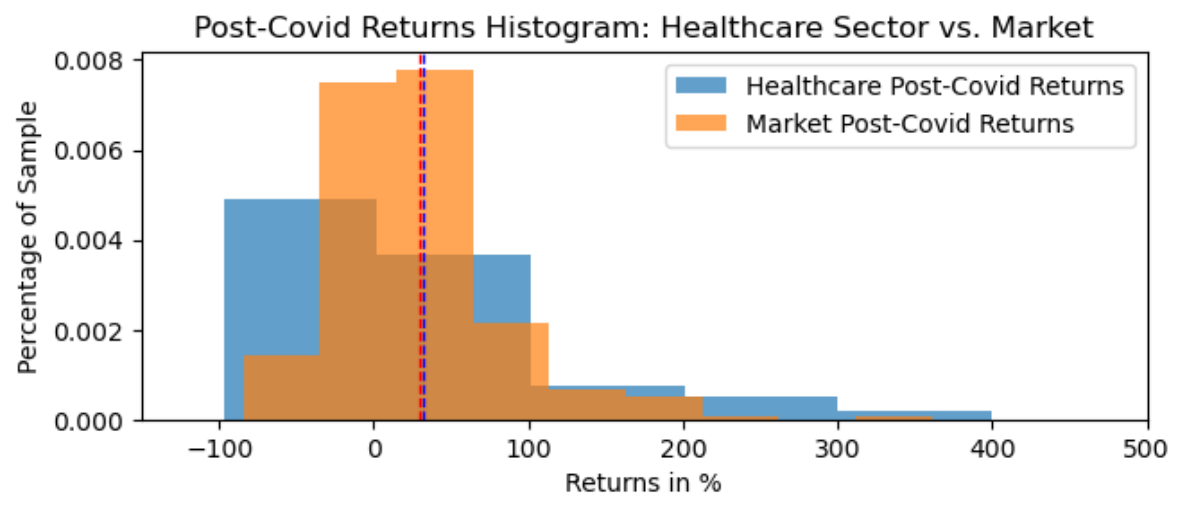
In order to do this we first remove micro-cap stocks from the healthcare dataset. This is because the market dataset does not have micro-cap stocks - companies in the S&P 500 much meet a certain market cap criteria, and because micro-cap stocks tend to skew data. So we removed the micro-cap stocks to make the 2 samples more comparable.

Now we are left with 366 healthcare stocks, so we randomly sample 366 stocks from our market list to create even samples.

Post Covid Returns - Plot



Pictured above is a scatter plot of returns for healthcare sector and the market in the post-covid period. It can be observed that there is much more variation amongst the healthcare returns.



Pictured above is a histogram of post-covid returns for healthcare and for the market. We can see here that the mean return for healthcare (represented by the blue line) is slightly higher than the mean return for the market (represented by the red line). We can see that there is less variation in market returns, with more of a bell shaped histogram, as opposed to the positively skewed healthcare returns.

Post-Covid Returns - Hypothesis Testing

In the graphs above, we saw that the average return for healthcare is slightly higher than the average return for the market. We need to find out if healthcare outperformed the market or if this is just a coincidence from the sample. Thus we use a two-sample, t-test, and test the following hypotheses:

Null Hypothesis:
μ(returns of healthcare post covid) = μ(returns of market post covid)

Alternative Hypothesis:
μ(returns of healthcare post covid) ≠ μ(returns of market post covid)

Where μ represents the sample mean. We ran a two-sample t-test, performing Welch's t-test, as the variances of the two samples are not equal and it produced the following results:

Ttest_indResult(statistic=0.3436044031274581, pvalue=0.7312812319452231)

The p value is ~0.73 which is not significant at any standard level, therefore we fail to reject the null. This implies that there was no significant difference between the performance of healthcare stocks and the market, meaning that healthcare stocks did not outperform the market.

An important thing to note here is that while the average return for healthcare sector and the market were almost similar, there is quite a difference in standard deviation or variation of the returns. In finance, the standard deviation represents risk, and in this case investing only in healthcare stocks is riskier than in the market as a whole.

Optimal investing required maximizing returns while minimizing risk. So between investing in the market, and just the healthcare sector, the optimal choice is the market, as you get the same return for less risk.

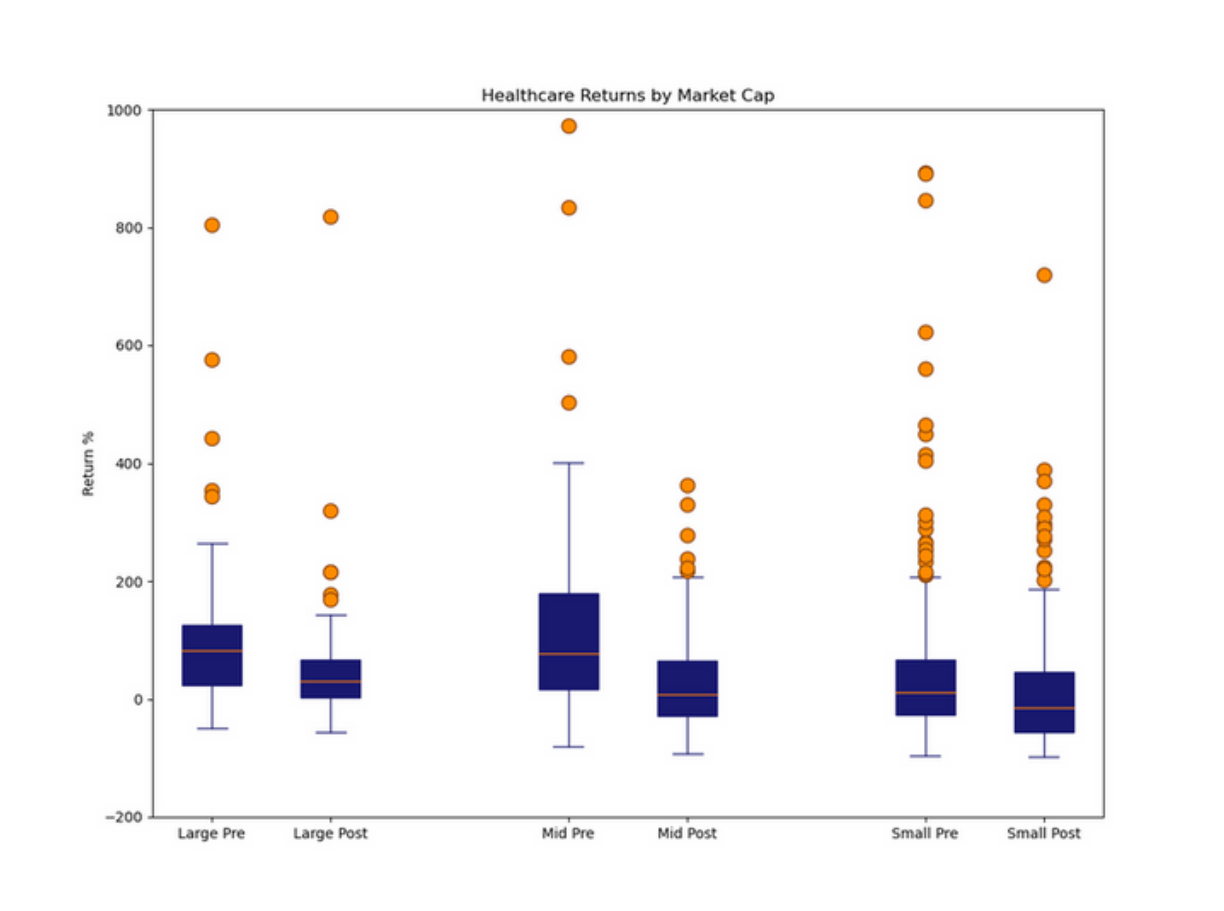
The market is a combination of various sectors, and is therefore a well diversified portfolio. As you diversify your investments, risk or standard deviation of returns go down. Healthcare sector is not well diversified, and thus it was seen in the graphs that there was lots of variation or risk.

Question 6

What impact did market capitalization have on returns?

In our previous analyses we saw that COVID-19 had a negative impact on both the market as a whole and the healthcare sector as well. In this question, we want to analyze how market capitalization played a role in returns.
In this analysis we will be looking at large-cap, mid-cap and small-cap stocks, and comparing the summary statistics pre-covid and post-covid results.

Healthcare Sector - Box Plot



Healthcare Sector - Summary Statistics by Cap Size

Summary Statistics for Returns of Large Cap Healthcare Stocks, Pre-Covid	
Mean	100.76
Median	81.47
Standard Deviation	129.14
Variance	16675.96

Summary Statistics for Returns of Large Cap Healthcare Stocks, Post-Covid	
Mean	50.23
Median	31.06
Standard Deviation	106.96
Variance	11439.45

Summary Statistics for Returns of Mid Cap Healthcare Stocks, Pre-Covid	
Mean	191.41
Median	76.71
Standard Deviation	502.70
Variance	252708.87

Summary Statistics for Returns of Mid Cap Healthcare Stocks, Post-Covid	
Mean	31.91
Median	7.30
Standard Deviation	92.79
Variance	8609.28

Summary Statistics for Returns of Small Cap Healthcare Stocks, Pre-Covid	
Mean	71.88
Median	10.44
Standard Deviation	239.05
Variance	57146.57

Summary Statistics for Returns of Small Cap Healthcare Stocks, Post-Covid	
Mean	25.77
Median	-15.43
Standard Deviation	178.40
Variance	31828.04

Healthcare Sector - Observations

1. In pre-covid times, mid cap healthcare stocks had the highest mean returns.

2. In post-covid times, large cap healthcare stocks had the highest mean returns.

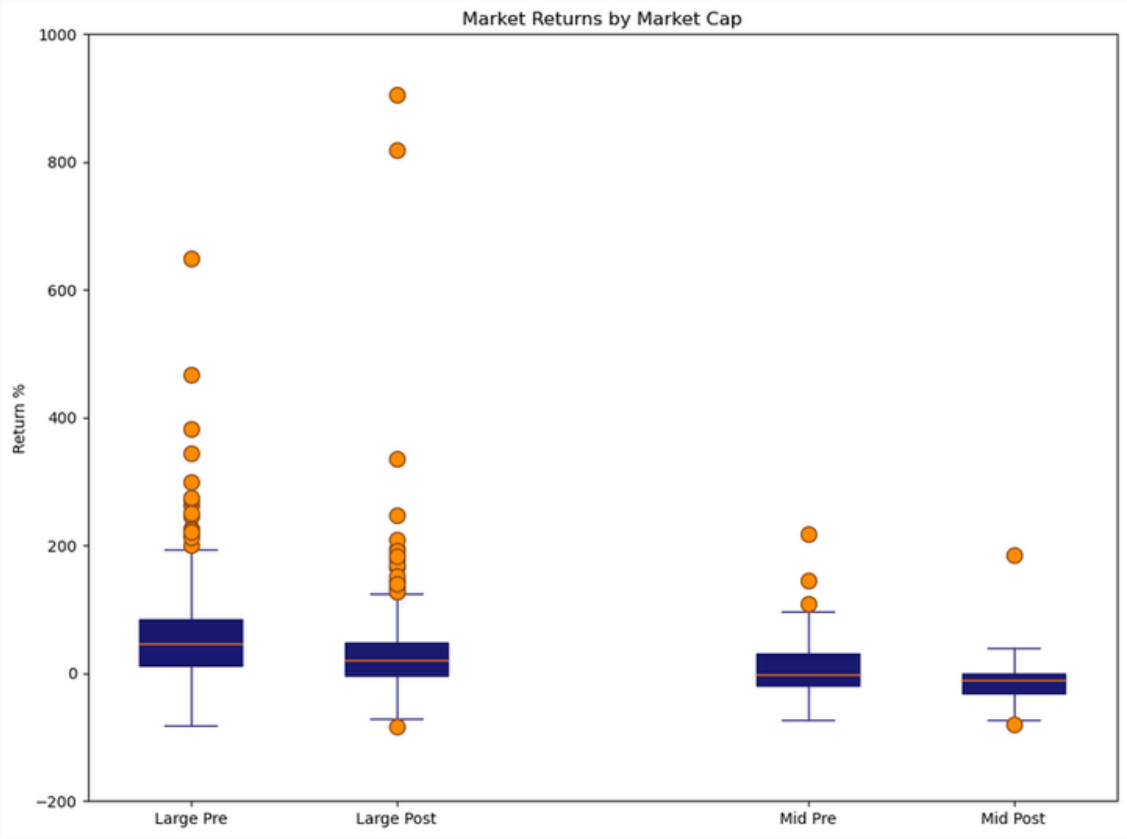
3. In pre-covid times, mid cap healthcare stocks also had the highest standard deviation.

4. In post-covid times, small cap healthcare stocks had the highest standard deviation.

5. There are many outlier values across all cap sizes, with all outliers being above the upper-bounds. This makes sense as stock price returns can only be as low as -100% but can be infinitely high.

6. Mid cap healthcare stocks appear to be impacted the most by COVID-19, just by looking at the box plots.

Market - Box Plot



Market - Summary Statistics by Cap Size

Summary Statistics for Returns of Large Cap Stocks, Pre-Covid	
Mean	62.34
Median	46.81
Standard Deviation	129.89
Variance	16870.29

Summary Statistics for Returns of Large Cap Stocks, Post-Covid	
Mean	30.48
Median	21.04
Standard Deviation	73.05
Variance	5335.96

Summary Statistics for Returns of Mid Cap Stocks, Pre-Covid	
Mean	14.38
Median	-2.79
Standard Deviation	59.47
Variance	3537.06

Summary Statistics for Returns of Mid Cap Stocks, Post-Covid	
Mean	-10.14
Median	-10.41
Standard Deviation	45.14
Variance	2037.69

Market - Observations

Note that there are no small cap stocks in our market sample since we used the S&P 500 as a benchmark for the market. S&P 500 has a minimum market cap requirement of $8.2 Billion.

1. Large cap stocks outperformed mid cap stocks in both pre and post covid periods.

2. Mid cap stocks had lower standard deviation in both periods.

3. There are many outlier values across all cap sizes, with most outliers being above the upper-bounds.

Question 7

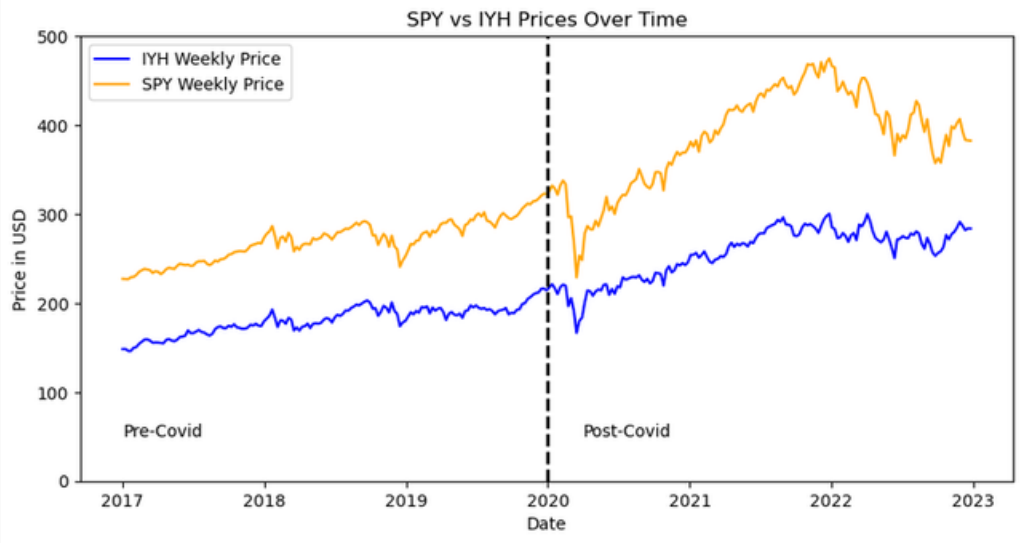
Looking at 2 ETFs, SPY and IYH as benchmarks, how do prices and returns differ?

If we wanted to see the time-series trend of price and returns of healthcare sector and the market, it would be difficult to plot all the stocks on one graph and observe an aggregate trend. In order to overcome this problem, we are going to use two ETFs as benchmarks for the healthcare sector and the market as a whole.

An ETF or exchange-traded fund is a basket of securities that are traded on an exchange like a singular stock. Securities in an ETF are usually chosen with a specific criteria in mind. ETFs can be used to get exposure to multiple stocks without having to individually buy or sell each stock. ETFs are also used to track specific markets or sectors - which is what we will be doing in this analysis.

As mentioned earlier, we are using the SPY ETF as a benchmark for the market and the IYH ETF as a benchmark for the healthcare sector.

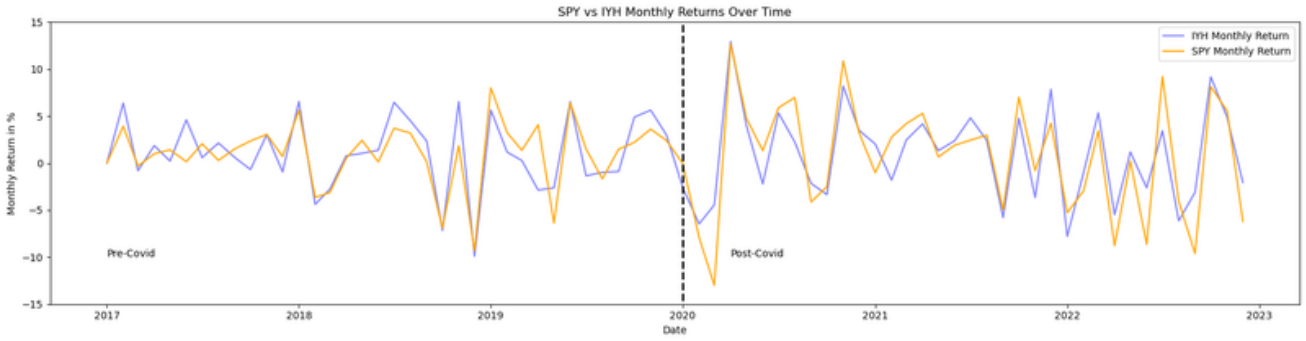
SPY vs. IYH Prices Over Time



Pictured above are two line graphs that represent the prices of each ETF over the time period of our analysis. The black line down the middle separates the graph into pre and post covid periods. We used close prices from January 2017 to December 2022. We chose weekly intervals to smooth out the trendline. It can be observed that the prices move fairly parallel to each other.

We can see that both prices follow a very similar pattern, with a dip at beginning of the post-covid period. This dip, as we know, is the drop in prices we saw on March of 2020 when the pandemic officially started in North America. We can see that since then, both the market and the healthcare sector has recovered, and continues on in a generally upward trend. Around the beginning of 2022, we can that SPY prices increase at a higher rate than IYH, but then adjusts back to be parallel with IYH prices.

SPY vs. IYH Returns Over Time



Pictured above are two line graphs that represent the monthly returns of each ETF over the time period of our analysis. Note that the returns are in percentages. The black line down the middle separates the graph into pre and post covid periods. We used close prices from January 2017 to December 2022. We chose monthly intervals to smooth out the trendline. It can be observed that the returns are quite similar in movement. Like in the previous graph, we can see the dip in the beginning of the pandemic.

The price and return graphs confirm our 2 prior findings. We concluded that the impact of COVID on the market is no different from the impact of COVID on the healthcare sector, which can be seen clearly in the above 2 graphs. The movement of prices and returns are quite similar. We also concluded that the healthcare sector did not outperform the market, and found that the returns are quite similar. This also can be seen in the graph above, where SPY and IYH return lines are quite similar too.

Conclusion, Recommendations & Limitations

Conclusion

In our analysis we looked at the top performers in the healthcare sector, and analyzed the impact of COVID-19 on healthcare sector stocks and the market as a whole.
We found that COVID-19 had a negative impact on the market as a whole and on the healthcare sector as well. We also found that there was no difference in the impact of COVID-19 on the healthcare sector compared to the rest of the market.
Our findings showed that healthcare sector did not outperform the market after COVID-19, contrary to what many people believed.
Cap size is impactful on stock returns, with different cap size categories experiencing different returns and risk (standard deviation).
We used SPY and IYH as benchmarks for the market and healthcare sector, and found that both prices and returns move quite similarly.

Recommendation

In uncertain times, markets get reactive, and there will always be winners and losers. While there were healthcare companies that experienced high returns because of COVID-19, eg. Moderna from our top performers list, there were also healthcare companies that experience negative returns. Healthcare crises, like the global pandemic, do not guarantee that the healthcare sector will outperform the market. It is best to maintain a well diversified portfolio to exposure to all sectors to reduce risk, or standard deviation.

Limitations

One limitation of our analysis is that we used the list of S&P 500 stocks to emulate the market. While it is a fairly diverse list of companies, it does not include small or micro cap stocks, as there is a minimum market cap requirement of $8.2 Billion. A sample of the market including small companies too would have been a better ground for comparison.

Thank you!