

SI 618 PROJECT 1 REPORT

Correlation of Stock Market and Cryptocurrency Trading Volumes after Covid19

Hareeshwar Karthikeyan

hkarthik@umich.edu

October 27, 2021

Motivation:

Cryptocurrency price movements have historically been claimed to be uncorrelated to the stock (or equities) market price fluctuations and cryptocurrencies are often seen as a hedging investment. However, following the market crash at the onset of the recent Covid19 pandemic, both the crypto and stock markets have seen an epic period of growth^{[1][2]}. As operations around the world move online, they have been increasingly investing in technology and since cryptocurrencies are also essentially elements from the blockchain technology, a stronger correlation has been observed between the prices of major cryptocurrency Bitcoin and the stock market indices^[3]. And following the recovery, when there was a stock market dump at the time of inflation, the prices of Bitcoin had a downtrend as well^[4].

This project looks into the trading trends considering the volume in the stocks/crypto trades. The overall price movement of the stock indices and cryptocurrencies have been in a similar bullish trend in the last one year following the market crash at the onset of the pandemic. The objective of this project is to look into how similar the trade volumes associated with this price movement are. Crypto currencies are known for their extreme volatility and poor liquidity and are not the favourites of the conservative investor. A similar trend in the trading volumes of two different assets means that the strength of the price fluctuations in those assets are also similar. If this similarity is found with the equity and crypto markets, it would be a stronger indication that the crypto market is starting to reflect the stock market.

Data sources:

Dataset 1 – Cryptocurrency Historical Prices Dataset from Kaggle^[5]

- This dataset has the historical price information of some of the top crypto currencies by market capitalization. It contains the Date of observation, Opening price on the given day, Highest price on the given day, Lowest price on the given day, Closing price on the given day, Volume of transactions on the given day and the Market capitalization in USD.
- Most studies involving cryptocurrencies involve only bitcoin as it takes almost half of the cryptocurrency market cap. But recently, a few other crypto like BNB, LINK, ETH, ADA and DOGE have gained a lot of market traction, some even overtaking Bitcoin to have the highest traded volume on Binance Exchange a few months back. Hence for this project, along with Bitcoin, we consider the top 10 currencies by their market cap because in combination, they constitute more than 75% of the total cryptocurrency market (info according to coinmarketcap).
- This dataset contains information for each currency in the csv format.

Dataset 2 – Stock Market Data (NASDAQ, NYSE and S&P500) from Kaggle^[6]

- This dataset contains the Date, Volume, High, Low, and Closing Price (for NASDAQ, S&P500, and NYSE listed companies).

- To consider only the major companies in the market, pick the companies pertaining to some top 100- company market indices. The following equity indices are considered.
 - i. NASDAQ 100
 - ii. S&P100
 - iii. NASDAQ 100 Technology Sector
 - iv. NYSE Arca Tech 100
- NYSEUS 100 index was also initially proposed for consideration in the project proposal earlier. However, during analysis it was noted that NYSEUS100 closely follows the NASDAQ100. Hence this index was not considered to avoid any redundancy.
- This dataset contains stocks information in both csv and json formats. The json files, however, are missing for a lot of stocks. Hence the csv files were considered and whatever stocks data was present in this was used for each index considered in this project.

Time Period Considered in the Data

- Three year time period from 01 Jul 2018 to 01 Jul 2021
- The first reported case of Covid 19 was in 31 Dec 2019^[7] and hence the above time period is selected which spans one and half years before and after the onset of Covid19
- Crypto markets are open every day and hence there is data present for all 1097 days in this time period in the dataset
- Stock markets, with its holidays, have data for 756 trading days in this time period.

Data Manipulation Methods:

Getting the required data:

- For the stock indices, the list of the stocks in each index had to be obtained and they had to be extracted from the dataset which contained several hundred stocks traded in a each stock exchange. This involved gathering the list of ticker symbols (since the data had files named for each stock by their ticker symbol) of the stocks for each index. The following resources were used for this purpose
 - Nasdaq100 list^[8]
 - S&P 100 list^[9]
 - Nasdaq100 Top Tech Companies list^[10]
 - NYSE Arca Tech 100 list^[11]
- Once the ticker symbols were obtained, the required files for each index were taken and the ticker symbol information was added to each of them. Then they were all merged together to form the index information. Following this, the data was filtered to contain only the desired three years' data.
- For creating the crypto index, all the data files in the dataset were considered before selecting the top ten by market cap value and merging them to form the index data file. Following this, the data was filtered to contain only the desired three years' data.

Calculations, conversions and processing:

- The project aims to look into the changes in the trading volume and the price changes. For estimating money flow into the assets of the index on the daily basis, a high volume of trades do not necessarily mean that more money went into the trades as their unit/share prices could be smaller. Hence, to get an appropriate estimate of the money flow into the trades, consider,

$$\text{Traded Monetary Volume} = \text{Mean}(\text{Open Price}, \text{Close Price}) * \text{Volume}$$
- The values for each stock/crypto's daily record were calculated as follows
 - Avg price = Average of open and close price
 - Total traded monetary volume(TMV) = Volume * Avg price

- $\%price\ change = 100 * (current\ close\ price - previous\ record's\ close\ price) / previous\ record's\ close\ price$
- $\%tmv\ change = 100 * (current\ tmv - previous\ record's\ tmv) / previous\ record's\ tmv$
- All the values in the data were present as strings. Hence they had to be converted to float data type before the calculations were done.
- Once the values were computed, each record in the index file, the final processing done on them was to group by the dates to contain the average price change percentage and average traded monetary volume(TMV) percentage of all the stock/crypto in the index for every date.
- The crypto and stocks index files had the date values in different formats. These were later converted to the same 'DD-MM-YYYY' format before joining them together.

Handling missing/incomplete data, Joining datasets and handling incorrect data:

- For the NYSE Arca Tech index, only two files were found in the dataset. Hence it was ignored as analyzing it wouldn't yield any real generalized results. Thus, final combined index data files considered are,
 - Crypto - Top 10 by Marketcap
 - Nasdaq100
 - S&P 100
 - Nasdaq100Tech
- In finding the Price and TMV changes with respect to the previous day, the previous day record would not be available for the first data record. This was handled as follows.
 - In the stocks index data, the price change for the first record of each stock was set as the difference between the day's open and close price.
 - In the crypto index data, the price change for the first record of each crypto was simply set as zero.
 - For the volume change, the value was set as zero for the first record of each stock/crypto in the index files.
 - A zero was selected as it wouldn't cause too much effect on the distribution of these values since the change percentage values lied in both the negative and positive regions and a single zero point wouldn't cause a considerable skewness to either side.
- The datasets with the computed values were inner joined on the date column using spark sql.
- The stock market had only 756 dates in the time period (since weekends and holidays have no stock market trading data) while the crypto dataset had entries for all the 1097 days. This was taken care of by using the inner join while joining the stock indexes with the crypto index. When it came to the column entries for each row, the dataset was full and complete without any missing values.
- After this was done, boxplots were used to have a look at the distribution of the values and the extremely large outliers -possibly incorrect change percentage values which were absurdly large and over 10000 percent- were removed. The change values within a few hundred percentages also showed up as outlier points on the boxplot but these weren't removed as these represent the days of strong market trends.

The first few records of the final dataset are as follows.

```
# Order of columns
# 'DATE', 'C_PCP', 'C_TMVCP', 'N_PCP', 'N_TMVCP', 'SP_PCP', 'SP_TMVCP', 'NT_PCP', 'NT_TMVCP'
```

	_c0	_c1	_c2	_c3	_c4	_c5	_c6	_c7
_c8								
02-07-2018	5.121125072620959	9.193665598572519	-1.4197329157191518		0.0	-6.92142945367664		0.0 -1.3828243249559322
03-07-2018	-1.7263876931070241	3.88579704445608	-0.7503701677520215	-34.69465041770667	-0.38754218086889713	-31.618496889332604	-1.4233110883018818	-32.83890509077797
05-07-2018	-0.7984669906309387	5.925269446030441	1.2241598859767095	70.59531435527903	0.9503111182226929	58.35068850659227	1.615999192251064	77.31397164561653
06-07-2018	-0.6435487855691865	-12.91041562423435	1.745983843588868	25.99901405601902	1.0062455628507367	27.340063807157787	1.5456087189055587	-6.8313403065421054
09-07-2018	-1.453540445159325	-5.414889467143977	0.7354070539389318	13.00388006337085	0.8477661177601469	15.089087671577163	0.7502420195560021	8.630149015982997

only showing top 5 rows

Workflow of the source code:

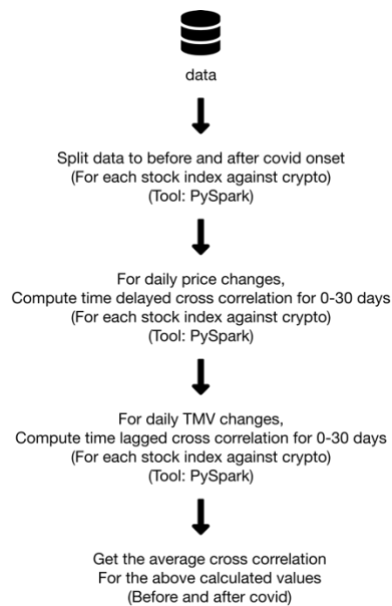
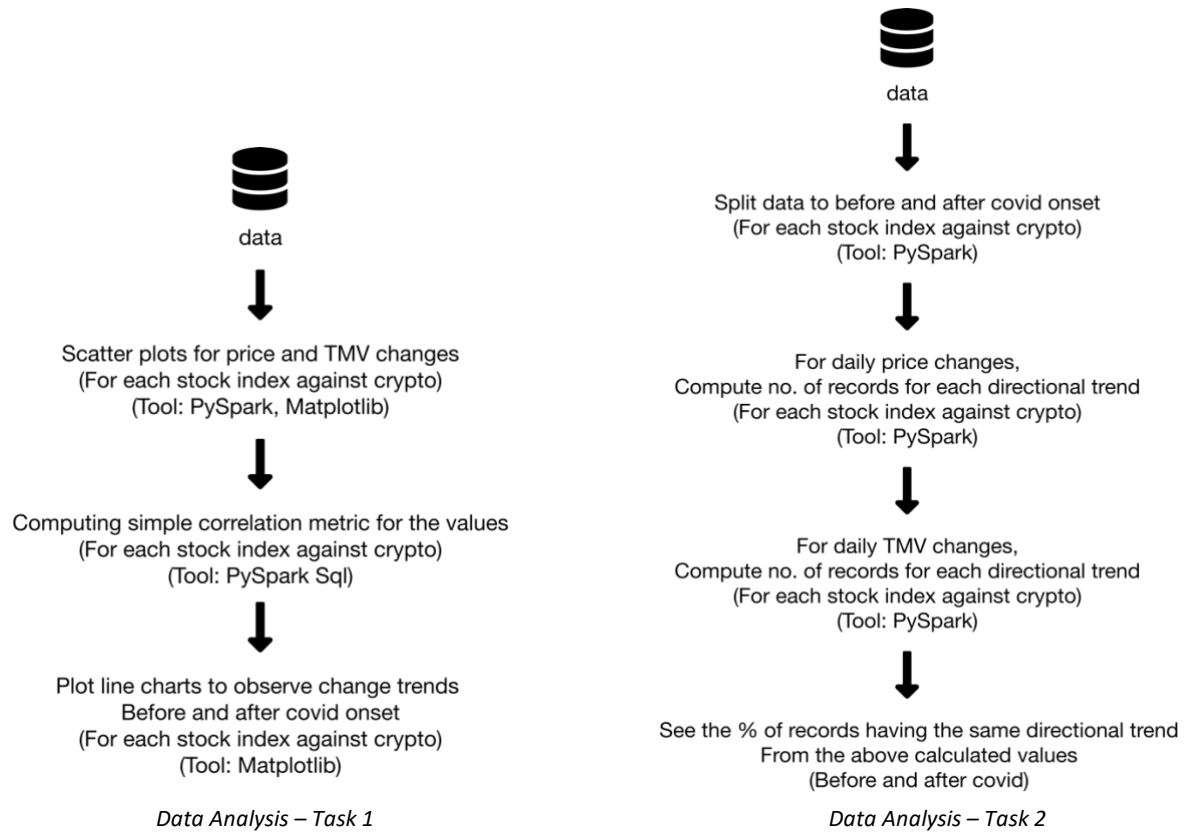
The data manipulation was done with the code split into three different files.

1. Data Manipulation – Stocks.py
2. Data Manipulation – Crypto.py
3. Data Manipulation – Combining Datasets.py



The data analysis tasks were each done in a separate file as follows.

1. Task1.py
2. Task2.py
3. Task3.py



Note: The 'price change' denotes the average daily price change percentage for all the stocks/crypto in a certain index and the 'tmv change' denotes the average daily tmv change percentage for all the stocks/crypto in a certain index. 'Directional trend' denotes either positive or negative change percentage.

Analysis and Visualization:

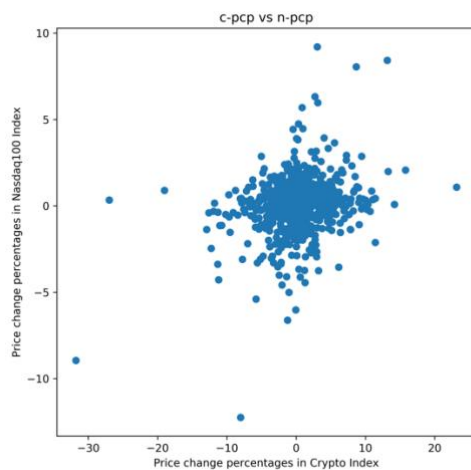
Task 1:

How do the overall comparisons of price and volume changes across the markets look like?

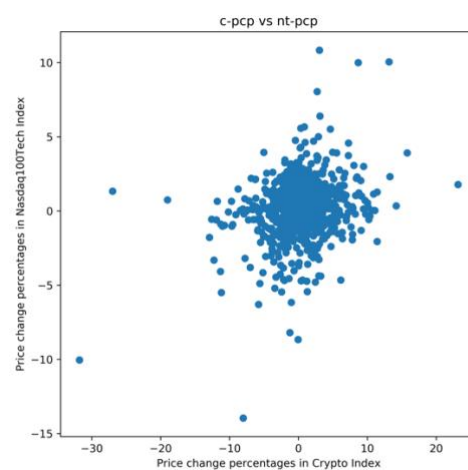
Scatter plots:

First, the data file was loaded as a spark data frame and all the string values were converted to float formats. Then, a scatter plot was plotted for the price change percentage values and the tmv change percentage values. The variable names used are as follows

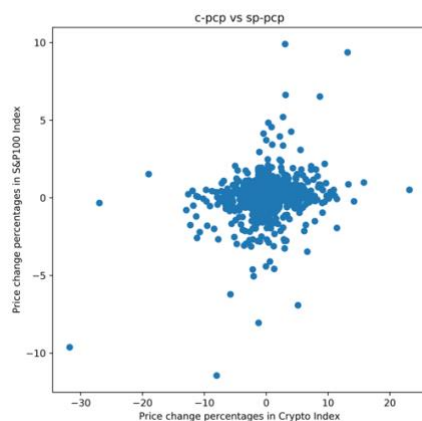
- c-pcp : Crypto – Average price change percentage
- c-tmvcp : Crypto – Average TMV change percentage
- n-pcp : Nasdaq100 – Average price change percentage
- n-tmvcp : Nasdaq100 – Average TMV change percentage
- sp-pcp : S&P100 – Average price change percentage
- sp-tmvcp : S&P100 – Average TMV change percentage
- nt-pcp : Nasdaq100Tech – Average price change percentage
- nt-tmvcp : Nasdaq100Tech – Average TMV change percentage



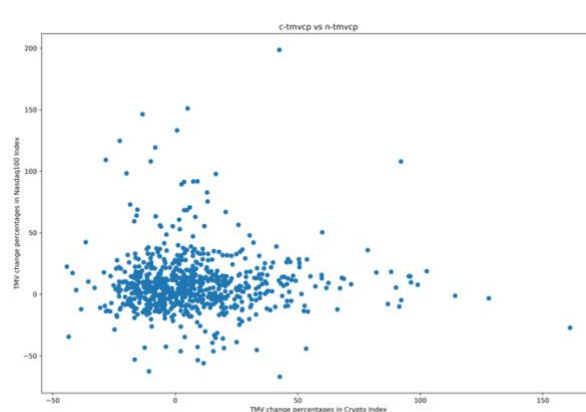
X-axis :c-pcp, Y-axis: n-pcp



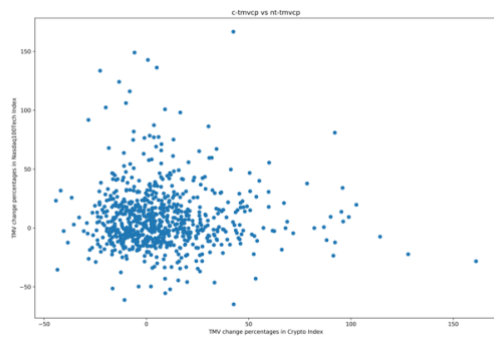
X-axis :c-pcp, Y-axis: nt-pcp



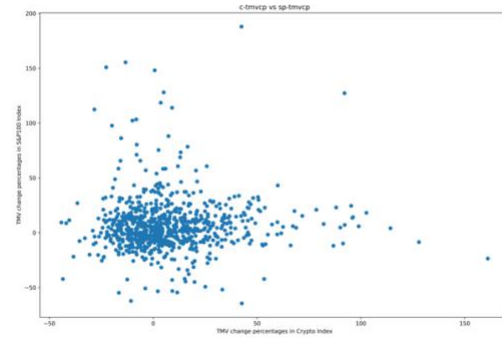
X-axis :c-pcp, Y-axis: sp-pcp



X-axis :c-tmvcp, Y-axis: n-tmvcp



X-axis :c-tmvcp, Y-axis: nt-tmvcp



X-axis :c-tmvcp, Y-axis: sp-tmvpcp

Simple correlation:

Following this, to get a simple static overall estimate, the correlation between the values for each index were taken. *Note that this value does not take into account any temporal value and it doesn't reveal any time related trends if one time series signal has an effect on the other.* Also, as seen below, these values reveal only a mild to no correlation between the indices for the price and tmv change percentages.

```
Correlation between daily average price change percentages
Crypto index and nasdaq100
+-----+
| CORRELATION|
+-----+
|0.2558040114840452|
+-----+

Crypto index and sp100
+-----+
| CORRELATION|
+-----+
|0.22833453790167552|
+-----+

Crypto index and nasdaqTech100
+-----+
| CORRELATION|
+-----+
|0.2543237456125306|
+-----+

Correlation between daily average traded monetary volume change percentages
Crypto index and nasdaq100
+-----+
| CORRELATION|
+-----+
|-0.00409818166758...|
+-----+

Crypto index and sp100
+-----+
| CORRELATION|
+-----+
|0.020775244928730157|
+-----+

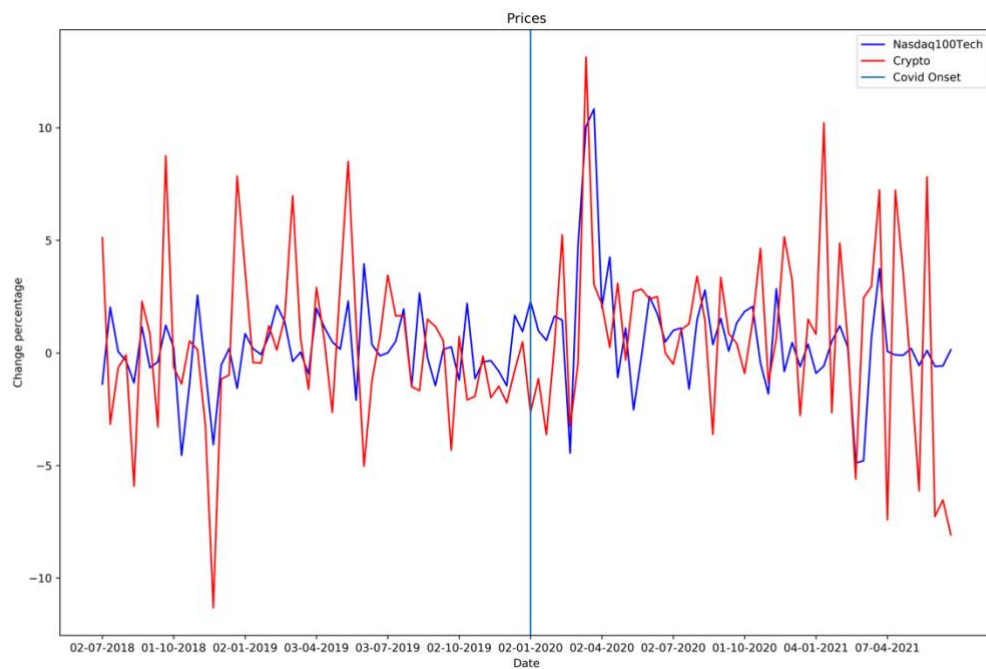
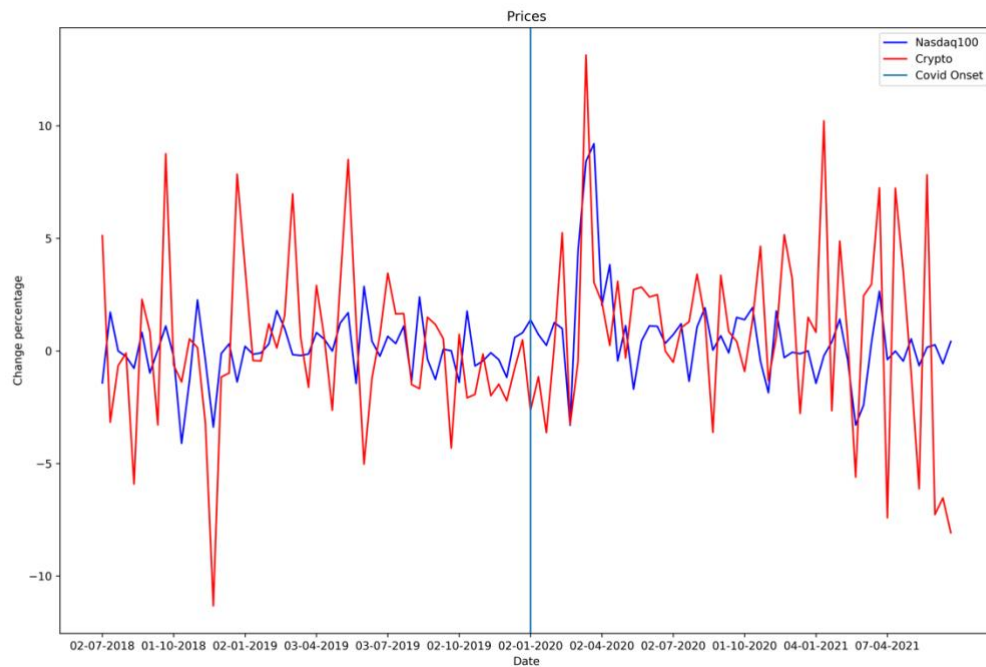
Crypto index and nasdaqTech100
+-----+
| CORRELATION|
+-----+
|-0.02783042218111...|
+-----+
```

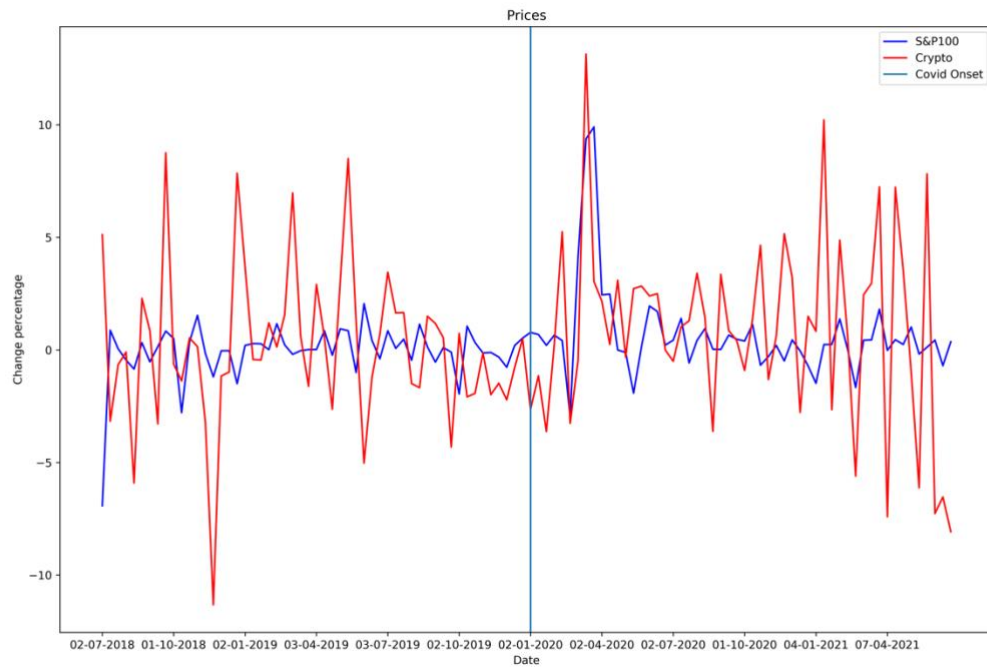
Line chart visualization:

After this, to get a visual outlook of the overall comparison, line charts were plotted to see the progress of the trends over time. The vertical line in the middle represents the central point in the data which is the covid onset time. Initially, in the project proposal, bar charts were proposed to study the traded monetary volume. However, it was seen that the total market cap in the stock markets is

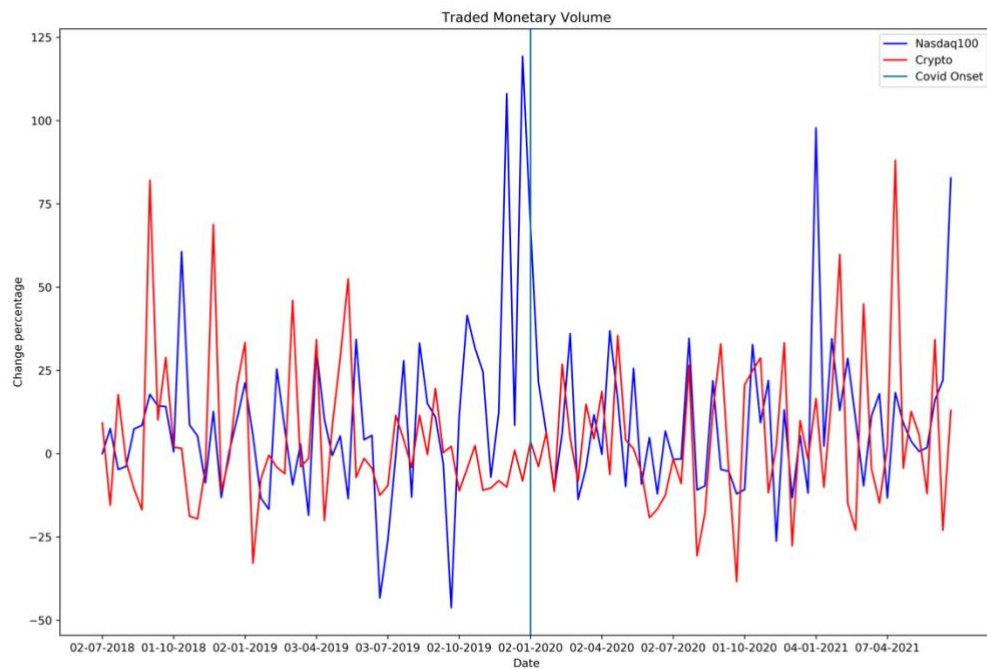
way more higher than the cryptos. Hence the crypto tmv would only be a small fraction of it. To overcome this disparity in the scale, the percentage of change values are considered and line plots are more lucid to see the progress of this value over time rather than bar charts.

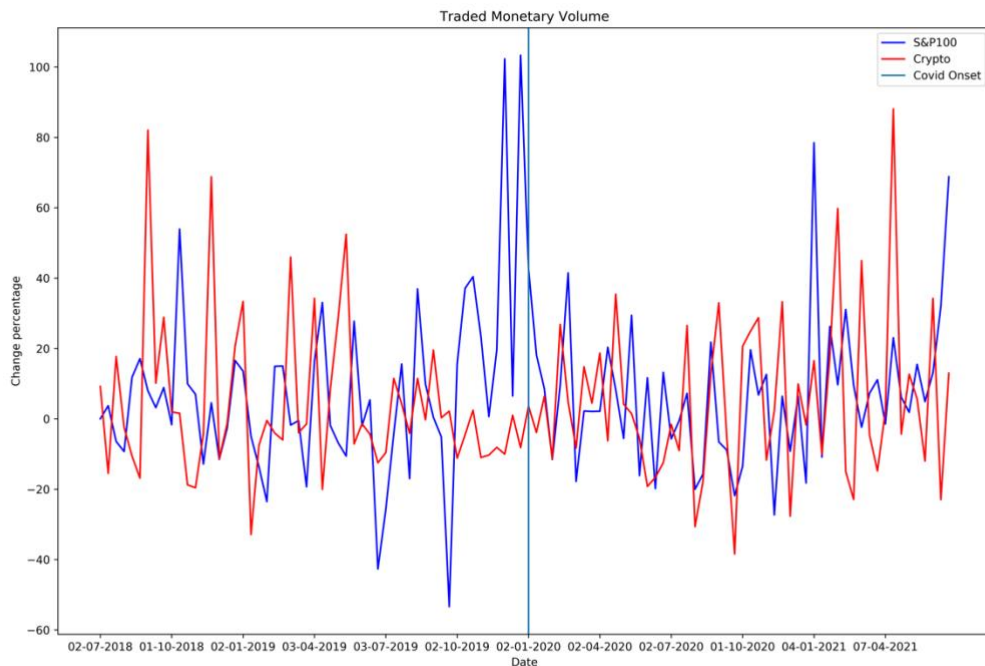
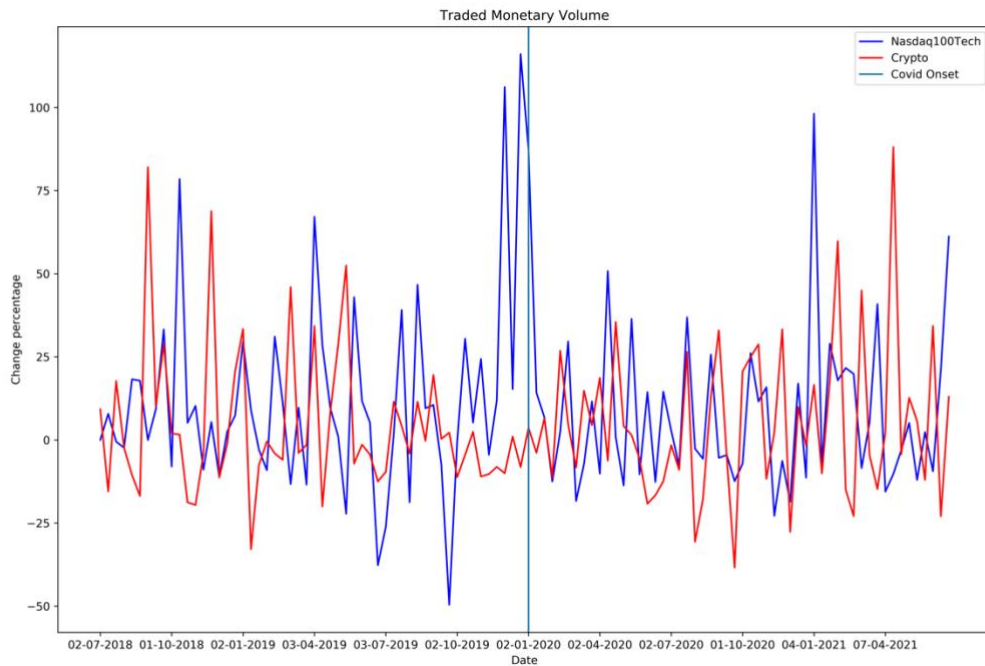
Price change percentages for crypto index against the stock indices





Traded Monetary Volume Change Percentages for crypto index against the stock indices





TMV has strong fluctuations near the period of covid onset in the stock markets but nothing similar is seen in the crypto market. But, it is followed by a drop in the magnitude of the swings in stocks whereas the magnitude of the swings seem to rise at this point in the cryptos

For the direction, no clear correlation can be noticed from this. But we can notice that the swings are in a similar direction in many cases, especially in the after-covid period.

Task 2:

How similar is the directional nature (positive and negative) of the changes in the price and traded volume in the markets before and after covid hit?

After reading the data onto a spark data frame, for both price and tmv change values, the comparisons of the direction were done comparing each index with the crypto index, after splitting the data for periods before and after covid onset. The number of values that had changes in the same direction and otherwise were computed and presented as a table. From these values, the ratio of the values having the same directional nature of changes was obtained.

In before-covid period:

```
Comparing daily avg price change percentage across the indices
Crypto vs Nasdaq100
      nasdaq100_positive  nasdaq100_negative
crypto_positive         112                72
crypto_negative         90                103
Percentage of records having the same direction of price change: 57.02917771883289
Crypto vs S&P100
      s&p100_positive  s&p100_negative
crypto_positive      111                73
crypto_negative      100                93
Percentage of records having the same direction of price change: 54.11140583554376
Crypto vs Nasdaq100Tech
      nasdaq100Tech_positive  nasdaq100Tech_negative
crypto_positive             111                73
crypto_negative             97                96
Percentage of records having the same direction of price change: 54.90716180371353
```

```
Comparing daily avg traded monetary volume change percentage across the indices
Crypto vs Nasdaq100
      nasdaq100_positive  nasdaq100_negative
crypto_positive         146                65
crypto_negative         110                55
Percentage of records having the same direction of tmv change: 53.45744680851064
Crypto vs S&P100
      s&p100_positive  s&p100_negative
crypto_positive      132                79
crypto_negative      94                71
Percentage of records having the same direction of tmv change: 53.98936170212766
Crypto vs Nasdaq100Tech
      nasdaq100Tech_positive  nasdaq100Tech_negative
crypto_positive             125                86
crypto_negative             94                71
Percentage of records having the same direction of tmv change: 52.12765957446808
```

In after-covid period:

```
Comparing daily avg price change percentage across the indices
Crypto vs Nasdaq100
      nasdaq100_positive  nasdaq100_negative
crypto_positive         140                81
crypto_negative         80                76
Percentage of records having the same direction of price change: 57.294429708222815
```

```

Crypto vs S&P100
      s&p100_positive  s&p100_negative
crypto_positive      131          90
crypto_negative       80          76
Percentage of records having the same direction of price change: 54.90716180371353
Crypto vs Nasdaq100Tech
      nasdaq100Tech_positive  nasdaq100Tech_negative
crypto_positive              139          82
crypto_negative              78          78
Percentage of records having the same direction of price change: 57.55968169761273

```

```

Comparing daily avg traded monetary volume change percentage across the indices
Crypto vs Nasdaq100
      nasdaq100_positive  nasdaq100_negative
crypto_positive          137          81
crypto_negative           93          66
Percentage of records having the same direction of tmv change: 53.84615384615385
Crypto vs S&P100
      s&p100_positive  s&p100_negative
crypto_positive          141          77
crypto_negative           86          73
Percentage of records having the same direction of tmv change: 56.763925729442974
Crypto vs Nasdaq100Tech
      nasdaq100Tech_positive  nasdaq100Tech_negative
crypto_positive              117          101
crypto_negative              89          70
Percentage of records having the same direction of tmv change: 49.60212201591512

```

No large changes after covid can be noticed. There are small increases in some and small decreases in some. The ratio of records in a similar direction of change are close to 50 percent in both scenarios for both price and tmv change percentages.

Task 3:

Is there any correlation in the time series data to see if the crypto trends reflect the stock trends in price and volume changes?

This task required computing the cross correlation to see if one time series value is actually related with another over time. The correlation of each stock index to the crypto index had to be studied to see if there are any trend reflections. Hence, for every stock index value, the value of the crypto index for different offsets (0 to 30) were taken. That is, for offset 1, we would take tomorrow's value from the crypto index and for offset 2, the one from the day after tomorrow and so on.

Once these thirty values were obtained for a price change or tmv change value comparison between the stock index and the crypto index, correlations were computed between the stock index value and each of these offset values. This shows if the current value in the stock market is correlated to any future values in the crypto markets. Finally, the average of these cross correlations was taken to get their general idea.

```
--BEFORE COVID--  
AVERAGE CROSS CORRELATIONS (0 to 30 day time lag)  
  
For Avg Daily Price Change Percentage  
Nadaq100 vs Crypto : 0.10301176301910034  
S&P100 vs Crypto : 0.03886703418526572  
Nadaq100Tech vs Crypto : 0.10811318930723292  
  
For Avg Daily TMV Change Percentage  
Nadaq100 vs Crypto : -0.054234646011012806  
S&P100 vs Crypto : -0.03607717065023008  
Nadaq100Tech vs Crypto : -0.07585510348936104
```

```
--AFTER COVID--  
AVERAGE CROSS CORRELATIONS (0 to 30 day time lag)  
  
For Avg Daily Price Change Percentage  
Nadaq100 vs Crypto : 0.32496428480914186  
S&P100 vs Crypto : 0.30090143539740305  
Nadaq100Tech vs Crypto : 0.32127257026528894  
  
For Avg Daily TMV Change Percentage  
Nadaq100 vs Crypto : 0.04416860015625667  
S&P100 vs Crypto : 0.07765885806094824  
Nadaq100Tech vs Crypto : 0.020390698683324285
```

The average cross correlation values are not high in either time periods. But they have tripled for the price change percentage in the after-covid period. For the tmv change percentage, the negative correlations have become positive. These are significant changes in the correlation from one time period to another and could explain the stronger overlap in the line chart observed in the after-covid period.

However, these correlation values are not numerically significant enough and it would be dangerous to make any strong conclusions. The crypto market hasn't actually reflected the stock market trading patterns as seen by this small correlation values which are less than even the 0.5 mark. The crypto market still has to stabilize a lot in terms of both price and monetary volume traded daily to start reflecting the stock market.

Challenges encountered:

- Information about the indices was not available in the stock market dataset. Hence, the ticker symbols of the stocks for the indices had to be manually collected and then a python script had to be used to pick the relevant files for each index.

- For the indices, some files were not present in the stock market data folder pertaining to that particular exchange. For one of the indices, only two stocks' data was available and hence that index had to be dropped from analysis(NYSE ArcaTech100). For the other indices, not all 100 were available (for the Nasdaq100 and the S&P100). Only the available files had to be considered (72 for Nasdaq100 and 68 for S&P100).
- For two cryptocurrencies in the top 10 by marketcap(Polkadot and Uniswap), entries were not present for all dates. The values were simply dropped and not used in calculating average values of the top 10 currencies so as to not disturb the average of the other currency values for those dates.
- On the whole, the greatest challenge was the data cleaning the initial data and curating the final data.csv file for analysis. Almost 50 percent of the code written in this project goes towards the Data Manipulation stage prior to Data Analysis. As a future development on this project, collecting data for every required stock/cryptocurrency using APIs from the stock and cryptocurrency exchanges could produce more relevant and significant results while reducing the effort for data cleaning and curation.

References:

1. <https://www.cnbc.com/2021/03/12/a-stunning-fall-and-a-recovery-how-the-stock-market-has-evolved-one-year-since-covid-hit.html>
2. <https://theconversation.com/bitcoin-why-the-price-has-exploded-and-where-it-goes-from-here-152765>
3. <https://www.etftrends.com/tactical-allocation-channel/bitcoins-correlation-to-markets-hits-a-record-in-2020/>
4. <https://decrypt.co/63468/is-bitcoin-price-correlated-to-stock-market>
5. <https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory>
6. <https://www.kaggle.com/paultimothymooney/stock-market-data>
7. [https://www.who.int/bangladesh/emergencies/coronavirus-disease-\(covid-19\)-update#:~:text=On%20this%20website%20you%20can,on%2031%20December%202019](https://www.who.int/bangladesh/emergencies/coronavirus-disease-(covid-19)-update#:~:text=On%20this%20website%20you%20can,on%2031%20December%202019)
8. <https://www.nasdaq.com/market-activity/quotes/nasdaq-ndx-index>
9. <https://www.barchart.com/stocks/indices/sp/sp100>
10. <https://finance.yahoo.com/quote/%5Endxt/components/>
11. <https://finance.yahoo.com/quote/%5EXCI/components/>