Crypto Market Analysis

Jyoti Kumari

Dec 07 2020

Overview: The objective of this project is to explore the Crypto currency market (i.e., Data downloaded from Coinmarketcap, Yahoo.com and Nomics.com) and draw potential correlations with S&P500 index (Yahoo.com data). The data will be cleaned up, grouped by common variables **and** further analysis/investigation will be done, as noted below.

Overall Question: Our project will focus on providing an analysis to the following questions:

Main Question: Is Bitcoin (BTC) dependent on external factors such as stock market, news, world events?

Sub-Questions:

1. *What is BTC dominance when compared to Altcoins / What drives Crypto Market Cap?*
2. *How is BTC related to S&P500 Index?*
3. *How are Altcoins and BTC related?*

Datasets (Source Files): We have considered 6 datasets in this project. The source of all these datasets are: https://coinmarketcap.com/ , https://nomics.com/ , https://finance.yahoo.com

1. **Preliminary Analysis:**

    a. **Info about Datasets:** This dataset contains the snapshot of data of top 500 crypto currencies as of Dec 5th 2020. Size of the data: 500 Rows and 30 Columns

    b. **Columns used from the above dataset:** The following variables are selected for our analysis

    - Id (Cryptocurrency Ticker Symbol)
    - Market Cap (Market Cap of that individual coin, Int format)
    - 1d.volume (Daily volume of that individual coin, Int format)

2. **BTC Price/Market cap:**

    a. **Info about Datasets:** This dataset contains daily prices (in USD) and volumes of BTC from 2013-04-28 to 2020-11-24. There are 2768 rows and 7 columns in this dataset. The variables in the dataset are: Date, Open, High, Low, Close, Volume, Market Cap

b. **Columns used from the above dataset:** The following variables are selected for our analysis:

- Date: Date of transaction in YYYY-MM-DD format
- Close: Adjusted closing price on the day
- Volume: Total number of Bitcoins transacted on the day
- Market Cap: Total market capitalization of Bitcoin during the day

3. **S&P 500:**

a. **Info about Datasets:** This dataset contains daily prices (in USD) and volumes of index SPX from 11/7/12 to 11/6/20. There are 2015 rows and 7 columns in this dataset. They are: Date, Open, High, Low, Adj.Close, Volume.

b. **Columns used from the above dataset:**

- Date: Date of transaction in DD-MM-YY format
- Close: Closing price of the index on the day
- Volume: Total number of assets traded under SPX on the day

4. **News Articles:**

a. **Info about Datasets:** This dataset contains the sentiment for Bitcoin and the market in general in a particular month and year. This dataset contains 100+ data points. This dataset was manually compiled by searching for cryptocurrency, financial and any other news that could have impacted the stock and crypto market. In the future, we would like to design a web crawler that will search and download all news articles for cryptocurrencies, derive sentiment using "sentiment analysis" and create the dataset.

b. **Columns used from the above dataset:**

- Sentiment: A High(H), Medium (M) or Low (L) sentiment echoed by the news article (Interpreted Sentiments).
- Month: Month of news article (in MM format)
- Action: Summary of news article

5. **Google Trends:**

a. **Info about Datasets:** Downloaded from trends.google.com, this dataset contains trends for Bitcoin and contains 260 data points across 2 columns, both of which were used in our analysis.

b. **Columns used from the above dataset**

- Week: Month and Year of data collection
- Bitcoin: Number of trending searches for Bitcoin

6. **Alternate Coins (Ethereum, ChainLink and PolkaDot)**

a. **Info about Datasets:** There are 3 datasets - one for each of the alternate coins (Ethereum, ChainLink and PolkaDot). The variables in each of these datasets are: Date, Open, High, Low, Close, Volume, Market Cap

- **Ethereum** dataset contains 1937 entries and 7 columns representing prices from 7-Aug-15 to 24-Nov-20.
- **ChainLink** dataset contains 1162 entries and 7 columns representing prices from 20-Sep-17 to 24-Nov-20.
- **PolkaDot** dataset contains 97 entries and 7 columns representing prices from 20-Aug-20 to 24-Nov-20.

Note: There is comparatively less data for other datasets such as Ethereum, ChainLink and PolkaDot as these are comparatively new crypto currencies. For e.g. PolkaDot was introduced in Aug-2020.

b. **Columns used from the above dataset:** The following variables are selected for our analysis for each of the alternate coins mentioned here:

- **Date**: Date of transaction in DD-Mon-YY format
- **Market Cap**: Total market capitalization of the cryptocurrency during the day

Data Exploration/Clean-up:

Before starting our analysis, we perform a sanity check on the data and the variables to ensure its quality and integrity. The sanity checks include examining the data using describe(), value_counts(), len(), range(), head(), tail() functions. We also write all datasets that were cleansed and reformatted to csv files for further inspection. Checks were limited to data and variables that we will be using in our analysis.

Preliminary Analysis: The following steps are undertaken to join the data sets from Nomics.com. All the data sets are of same size. Hence, Pd. Concat() function is used to join the datasets and convert to panadas. The columns were listed from back to front, hence a function called “reordered” was created to reorder the columns. Final master data set (Finaldf2) was created. The required variables (‘id’,’market_cap’,’1d volume’) were selected. 1d Volume column was renamed. The values greater than 0 were filtered to make sure that no erroneous entries were taken into consideration. **Market Cap Percent**: This is calculated column. This was

calculated using Lambda function – dividing the individual market cap by the sum of market cap and multiplying by 100. Format function was created to format the market_cap_percent column. **Market_cap_$ and daily_volume_$** columns were created so that the dataset can be read easily. These are text columns. The market cap was converted to billions and $ sign was added in the front so that the overall text is in the format – "$354.1B". Format function was created and apply method was used so that the format function can be applicable to the panda series. Once the new series was created, pd. Concat function was used to append the new series to the existing dataframe and columns were renamed using rename() method. New Helper column was created which indicates market cap and percent marketcap together in the format – '$354.68B (61.78%)'. String concatenation was done here and pd. Concat function was used to join the series to the main data frame. Market_cap_7 was the final dataframe that was created for plotting

Initial Data Set:

Out[56]:

	id	currency	symbol	name	logo_url	status	price	price_date	price_timestamp	circulating_supply	...	high	hi
0	BTC	BTC	BTC	Bitcoin	https://s3.us-east-2.amazonaws.com/nomics-api/...	active	19139.348870	2020-12-02T00:00:00Z	2020-12-02T21:00:00Z	18568362	...	19688.581651	
1	ETH	ETH	ETH	Ethereum	https://s3.us-east-2.amazonaws.com/nomics-api/...	active	596.459215	2020-12-02T00:00:00Z	2020-12-02T21:00:00Z	113078940	...	1361.744656	
2	XRP	XRP	XRP	Ripple	https://s3.us-east-2.amazonaws.com/nomics-api/...	active	0.618646	2020-12-02T00:00:00Z	2020-12-02T21:00:00Z	45334265890	...	2.770437	
3	USDT	USDT	USDT	Tether	https://s3.us-east-2.amazonaws.com/nomics-api/...	active	1.004891	2020-12-02T00:00:00Z	2020-12-02T21:00:00Z	19410938050	...	1.092650	
4	LTC	LTC	LTC	Litecoin	https://s3.us-east-2.amazonaws.com/nomics-api/...	active	57.311481	2020-12-02T00:00:00Z	2020-12-02T21:00:00Z	65669014	...	366.453150	

Final Dataset:

```
market_cap7.head()
```

	id	market_cap	daily_volume	market_cap_percent	market_cap_$	daily_volume_$	mcap_percent
0	BTC	354676428896	3.012829e+10	61.78	$354.68B	$30.13B	$354.68B (61.78%)
1	ETH	67800079809	1.438233e+10	11.81	$67.80B	$14.38B	$67.80B (11.81%)
2	XRP	28045895552	9.442518e+09	4.89	$28.05B	$9.44B	$28.05B (4.89%)
3	USDT	19505575729	5.453582e+10	3.40	$19.51B	$54.54B	$19.51B (3.4%)
4	LTC	5762550266	5.414855e+09	1.00	$5.76B	$5.41B	$5.76B (1.0%)

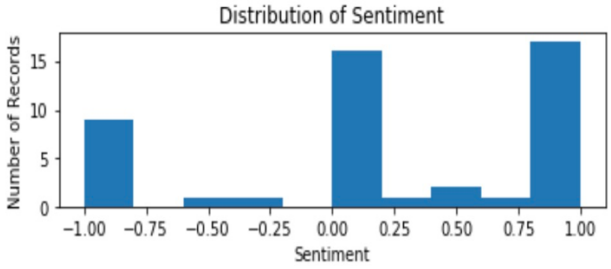
BTC Price Actions: Validating the data

Below code indicates data inspection for BTC Price Action analysis:

```
BTC_df2 = pd.read_csv("sources/Historical_Data_for_Bitcoin.csv")

print("Dataset - first 5 records")
print(BTC_df2.head())
print("Dataset - last 5 records")
print(BTC_df2.tail())
print("Dataset - shape")
print(BTC_df2.shape)
print("Dataset - describe")
print(BTC_df2.describe())
```

These are the initial results of the above checks:

```
Dataset - first 5 records
         Date      Open      High       Low     Close        Volume  \
0  2020-11-24  18365.01  19348.27  18128.66  19107.46  5.146957e+10
1  2020-11-23  18370.02  18711.43  18000.80  18364.12  4.274111e+10
2  2020-11-22  18642.23  18688.97  17671.38  18370.00  4.128043e+10
3  2020-11-21  18621.32  18936.62  18444.36  18642.23  3.965021e+10
4  2020-11-20  17817.08  18773.23  17765.80  18621.31  3.699287e+10

     Market Cap
0  3.545044e+11
1  3.406981e+11
2  3.407902e+11
3  3.458192e+11
4  3.454119e+11
Dataset - last 5 records
            Date    Open    High     Low   Close  Volume    Market Cap
2763  2013-05-02  116.38  125.60   92.28  105.21     0.0  1.168517e+09
2764  2013-05-01  139.00  139.89  107.72  116.99     0.0  1.298955e+09
2765  2013-04-30  144.00  146.93  134.05  139.00     0.0  1.542813e+09
2766  2013-04-29  134.44  147.49  134.00  144.54     0.0  1.603769e+09
2767  2013-04-28  135.30  135.98  132.10  134.21     0.0  1.488567e+09
Dataset - shape
(2768, 7)
Dataset - describe
              Open          High           Low         Close        Volume  \
count   2768.000000   2768.000000   2768.000000   2768.000000  2.768000e+03
mean    3883.310206   3981.097016   3781.481073   3890.090387  7.175396e+09
std     4265.241595   4383.524555   4137.924908   4274.506549  1.186830e+10
min       68.500000     74.560000     65.530000     68.430000  0.000000e+00
25%      416.480000    421.127500    409.322500    416.427500  2.619468e+07
50%     1034.285000   1054.150000   1001.110000   1034.345000  2.107975e+08
75%     7399.777500   7538.407500   7239.102500   7402.305000  9.261024e+09
max    19475.800000  20089.000000  18974.100000  19497.400000  7.415677e+10

         Market Cap
count  2.768000e+03
mean   6.786945e+10
std    7.665694e+10
min    7.784112e+08
25%    5.794512e+09
50%    1.631594e+10
75%    1.307320e+11
max    3.545044e+11
```



It is highly unlikely that we will find news articles every month dating back to 2012 when Bitcoin was introduced. There are also instances when multiple news articles were identified in a certain month and year. Sentiment was converted to a numeric value and aggregated by month and year for analysis with other datasets. Sentiment values for month and year where no news articles were available, was set to neutral (Medium) using the fillna() function. The above graph (refer to Distribution of Sentiment) confirms that sentiment value was recomputed and is contained within the boundaries set for high and low sentiment. Bitcoin data was also aggregated by month and year and the monthly average closing price and volumes were computed for further analysis with other datasets.

Major Alts: For data exploration and analysis and cleanup purposes, followings steps have been performed:

1. Performed Data loading by converting data files downloaded from https://coinmarketcap.com/ to dataframes resulting in 4 different dataframes corresponding to each of the major cryptocurrencies(Bitcoin, Ethereum, ChainLink, and PolkaDot)
2. Added suffix to existing variables with the currency name to create unique columns for each of the cryptocurrencies. For e.g. Market Cap column has been suffixed to Market Cap_btc.

3. Merged all 4 dataframes to a single dataframe using outer merge that resulted in a dataset where all empty cells' values were replaced with 'nan'. Since all 3 altcoins - Ethereum, ChainLink, and PolkaDot came after bitcoin, there were a lot of empty cells for a particular date range.
4. The Single merged data frame was then normalized by comparing the existing market cap value with the global maximum and minimum, in order to show correlation between different altcoins realistically.

Why Normalization?

Normalization helps get rid of a number of anomalies that can make analysis of the data more complicated. It transforms multi scaled data to the same scale that help in visualizing data more efficiently and realistically. In our case, as per the below snapshot, since there is a large variance in the value range for market cap for all the cryptocurrencies, normalization will help transform these variances to a common scale. We will get to see the difference visually in the following section on Charts/Graphs/Plots.

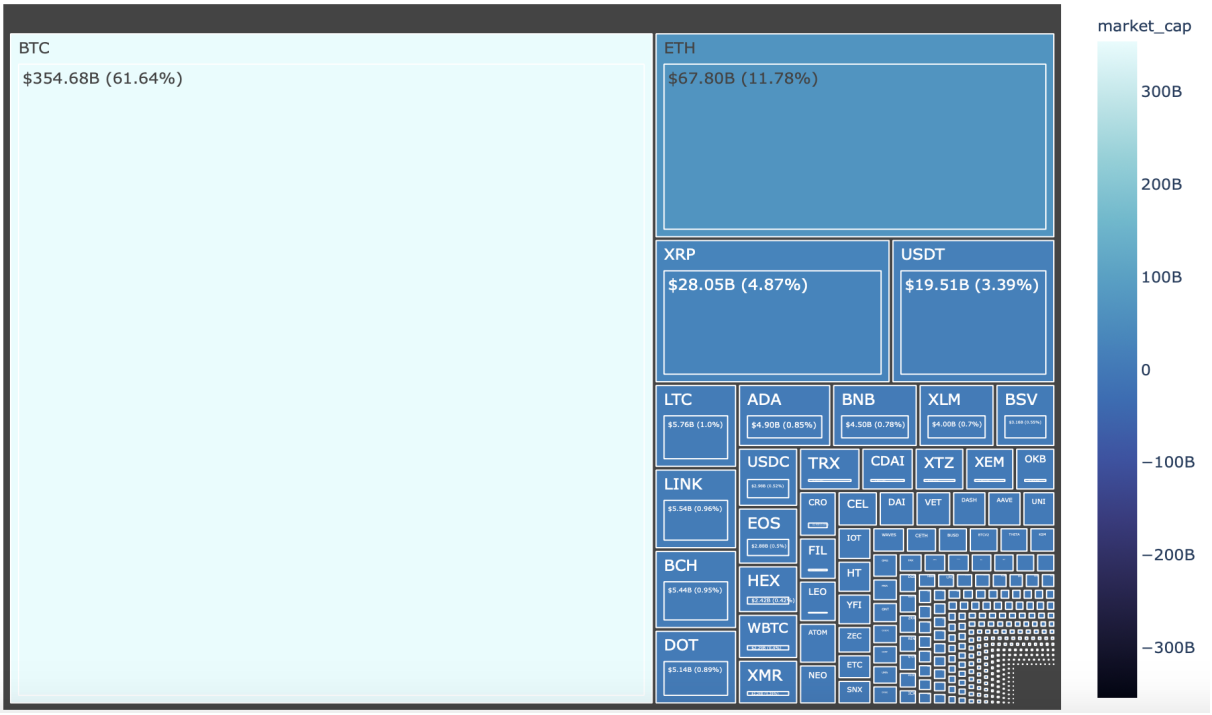
	MCap_btc	MCap_eth	MCap_clk	MCap_pldot
max	354504361032	135,400,735,922.00	6,684,562,685.00	5,234,800,947.00
min	778411179	32,213,626.00	45,895,150.00	0.00

Data Analysis:

Preliminary Analysis: Please refer to the below tables:

Overall Market Structure:

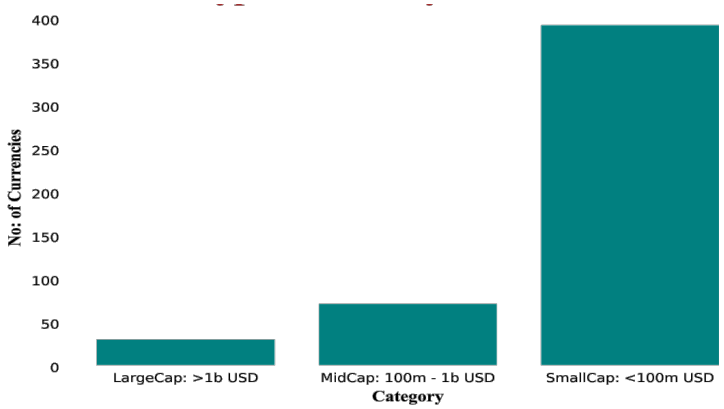
BTC Dominance – 61.64%



This indicates the overview of the market structure of crypto currency market. This treemap was prepared using plotly library. From the chart we can see that BTC has 61.64% marketshare with total marketcap of $354.6Billion as of Dec 5th 2020. Ethereum, XRP (Ripple), LTC (Litecoin), LINK (ChainLink), PolkaDot (DOT) are some of the altcoin majors.

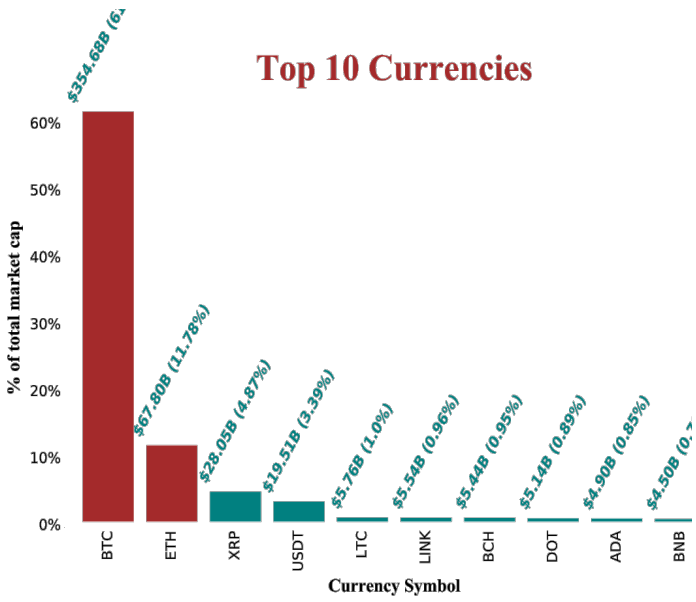
Market structure based on Categories

Few (32) Large Caps (> 1billion USD) dominating the market



In order to understand the structure better, the data was divided into 3 subsets based on the market cap. They are: LargeCap (32 out of 500): > 1 Billion USD Marketcap, MidCap (73 out of 500): 100m – 1Billion USD and Small Cap (394 out of 500): < 100Million USD. We can clearly see from the below two charts that: 1. Large Caps dominate the marketcap and they are fewer (32) in number. 2. BTC and Ethereum contribute to more than 70% of the marketshare 3. BTC is the dominant crypto with 61.64% marketshare. ***Hence, BTC price action drives the overall crypto market cap.***

Top 10 Crypto Currencies:



We wanted to further explore the top 10 Large Caps so that we will understand, the market structure within the large caps. We have used matplotlib to generate the bar graph. We used custom formatting for color coding as well as showing text labels so that the viewers can easily understand the data. The top two crypto currencies, i.e., BTC and ETH are highlighted in Red, and the rest of the top 10 majors are highlighted in teal color.

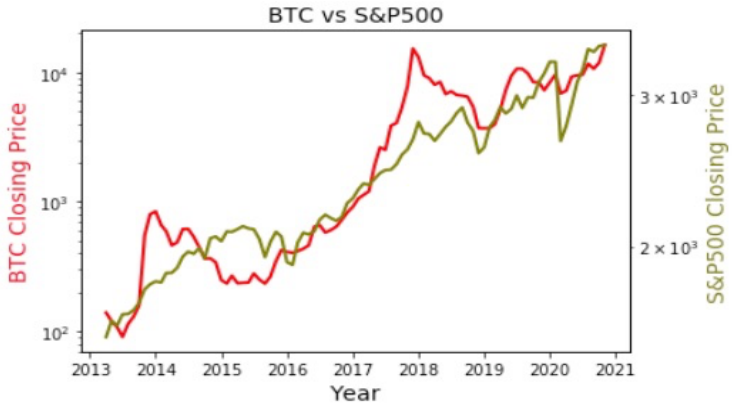
This chart again shows that BTC dominates the crypto market.

This answers one of our initial questions: "*What is BTC dominance when compared to Altcoins / What drives Crypto Market Cap?*

Further analysis is focused on BTC Price Action and Altcoin price action correlated with BTC

Detailed Analysis of BTC Price Action: Please refer to the below tables:

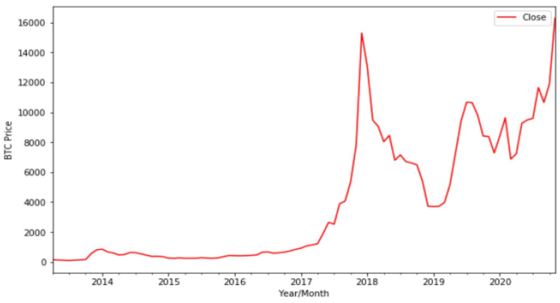
BTC vs. S&P 500:



We find that BTC prices over the years are highly correlated to SPX. We used a logarithmic scale to show BTC and SPX prices while showing the two prices in 2 separate y axes in a single chart. We can also see that around Mar2020, stock market crash, BTC has also fallen considerably. This indicates even though traditionally (2013-2016) BTC is considered independent of SPX, that during world events such as Pandemic (early 2020), BTC price action followed S&P 500

BTC prices outperformed S&P500 in late 2013 and in mid 2017 primarily because of very strong sentiment (i.e., public interest driven by price action) and regulations (recognizing BTC as a tradable currency) that favored BTC during this time.

BTC vs. Volume:

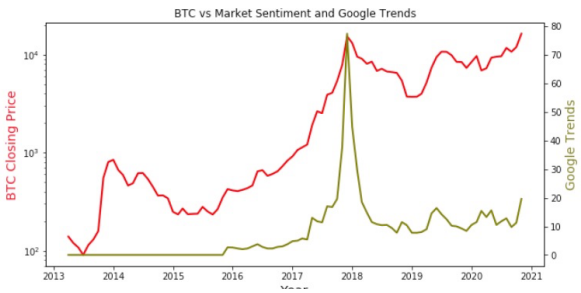


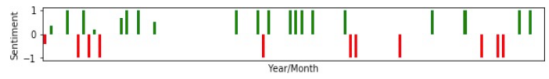
We find that an increase in BTC price is always correlated with an increase in volume. We see a significant rise in price and volume since mid-2017 which can be correlated to the 2017 Bull run prices. The volumes have sustained since then due to aggressive buying and selling of this asset.
BTC price more than halved in early 2014 with clear decrease in volumes.

Note: Volumes were not available for most of 2013 and we used bfill and ffill to update missing values.

BTC has skyrocketed in late 2020 after the election. Volumes have almost tripled after the price dump due to COVID.

BTC Price vs. Sentiment: BTC price is compared against major events such as US Elections timing, Pandemic, Stock Market Crash, Exchange Hack etc., The below analysis illustrates it in more detail:





cont…

We can see that after the last two US Elections in 2016 and 2020, the price of BTC has increased. This behaviour aligns with S&P500 behaviour where the investors find more stability after a clear election outcome. Negative news such as Exchange hack (Mt.Gox Exchange Hack) resulted in 2014 BTC price drop by 20%.

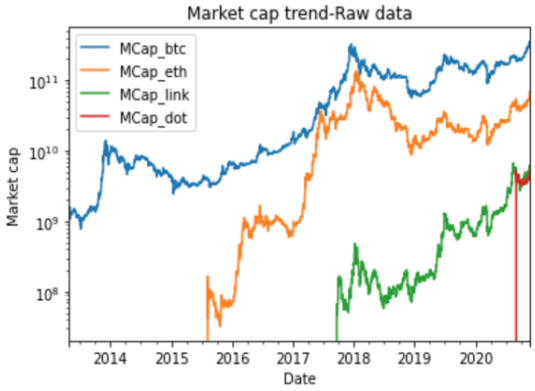
We find a strong correlation with positive news and sentiment with an increase in BTC price. A negative news article such as regulation or delisting from an exchange shows a related downward price pressure on BTC. We can also see a spike in Google Trends at the time of Dec 2017 Bull run, hence showing increased public interest that could have potentially lead to "Fear of Missing Out" (FOMO) situation and could also have contributed to price bubble.

External factors such as pandemic or elections also impacted BTC prices as we can see from the drop in Mar-2020 and the all-time high in Nov/Dec 2020.

This answers our Key Question and sub-question: Is BTC dependent on external factors such as stock market, news, world events? How is BTC related to S&P 500 Index?

Further analysis is based on Altcoins and their correlation with BTC

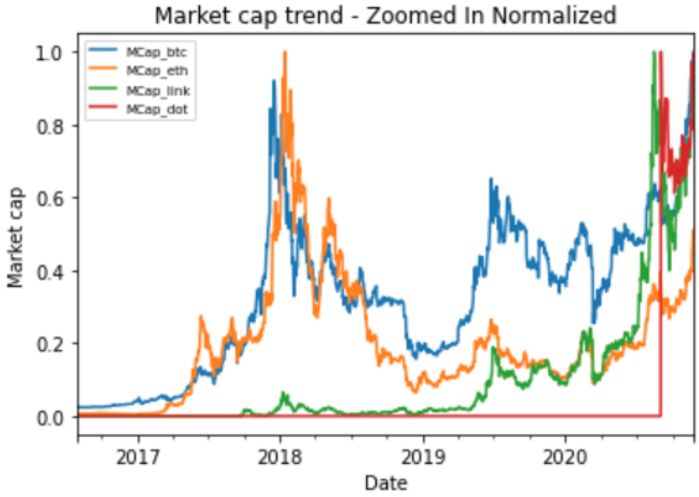
BTC vs. Selected Major Alts (Raw Data): This chart is based on the data directly derived from coinmarket.com. This chart shows the trend of Bitcoin(BTC) and other major Alts such as Ethereum(ETH), ChainLink(LINK) and PolkaDot(DOT).



*From this chart, **it appears that there is strong correlation between BTC and major Alts**. This also shows up that a new currency always sees a sharp rise in the beginning as we see for "eth" in Q3 2015, "link" in Q3 2017 and "dot" in Q3 2020.*

Since there is a huge gap in the start date as well as in the market cap for each of the cryptocurrencies, we decided to further normalize the data and the zoom in from mid-2016 to get better insights

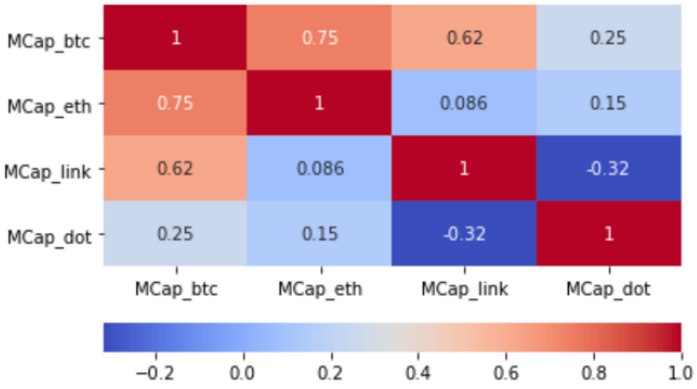
BTC vs. Selected Major Alts (Zoomed Version):

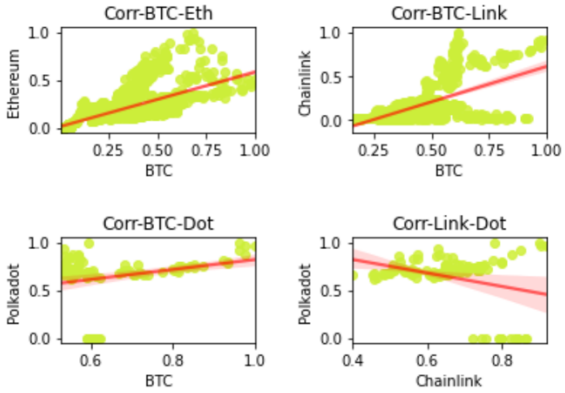


*Typical trend is that all alt coins go along with BTC. They go up when BTC goes up and they go down when BTC goes down. However, the **recovery rate of BTC is faster as compared to alt coins as seen for btc and eth between Q2 2018 to 2020.***

The other interpretations include:
1. BTC recovered faster than Ethereum in terms of its marketcap between 2017 and 2020.
2. LINK outperformed BTC and Ethereum by recovering in 2019 and reaching new All Time High in mid-2019 and ever since that time, it has been on a parabolic run.

Market Cap Correlation – BTC vs. Selected Major Alts:





Moving further, in order to closely analyze correlation between all major crypto currencies, we created heatmaps and regression models to show correlation between BTC and Selected Major Alts.

- ***BTC and Ethereum are strongly correlated***
- *BTC and Chainlink are moderately correlated (ChainLink outperformed BTC since mid-2019 – refer to earlier charts)*
- *BTC and PolkaDot are weakly correlated (Time period is too short, i.e., 3 months to estimate correlation)*
- ***ChainLink and PolkaDot correlation is negatively moderate. Given that DOT is a new coin (as indicated in the previous charts), the launch effect could have played a role in drawing ChainLink investors to it.***

This answers one of our questions: How are Altcoins and BTC related?

Assumptions:

- Data in our datasets were accurate representation of the truth
- Our analysis/interpretations were limited to the data that we have collected. The actual scenarios might be dependent on various other factors in addition to those mentioned in the report
- The sentiments were subjective and were based on our preliminary interpretation based on the identified news article only. The actual sentiments could vary depending on various other factors.
- The data collected is based on the availability on portals such as coinmarketcap.com, yahoo.com and Nomics.com. Given that the crypto currency market is a new field, complete data is not readily available.

Key Insights:

1. BTC is the dominant cryptocurrency that drives the crypto market at large **with 61.8% market cap dominance**

2. BTC is heavily influenced by key factors such as **1. Elections 2. Pandemic 3. Exchange Hacks, Shutdowns, Bankruptcies (Black Swan events) etc.**
3. There is a strong correlation between Major Alts such as Ethereum, ChainLink, PolkaDot and BTC. However, in the last 3 months, the recovery of BTC is much higher when compared to other Altcoins

References

https://nomics.com/

https://coinmarketcap.com/

https://finance.yahoo.com/

https://trends.google.com/trends/explore?q=Bitcoin&geo=US

https://en.wikipedia.org/wiki/Bitcoin