Code First Girls Nanodegree
Data Science Module
August 2021

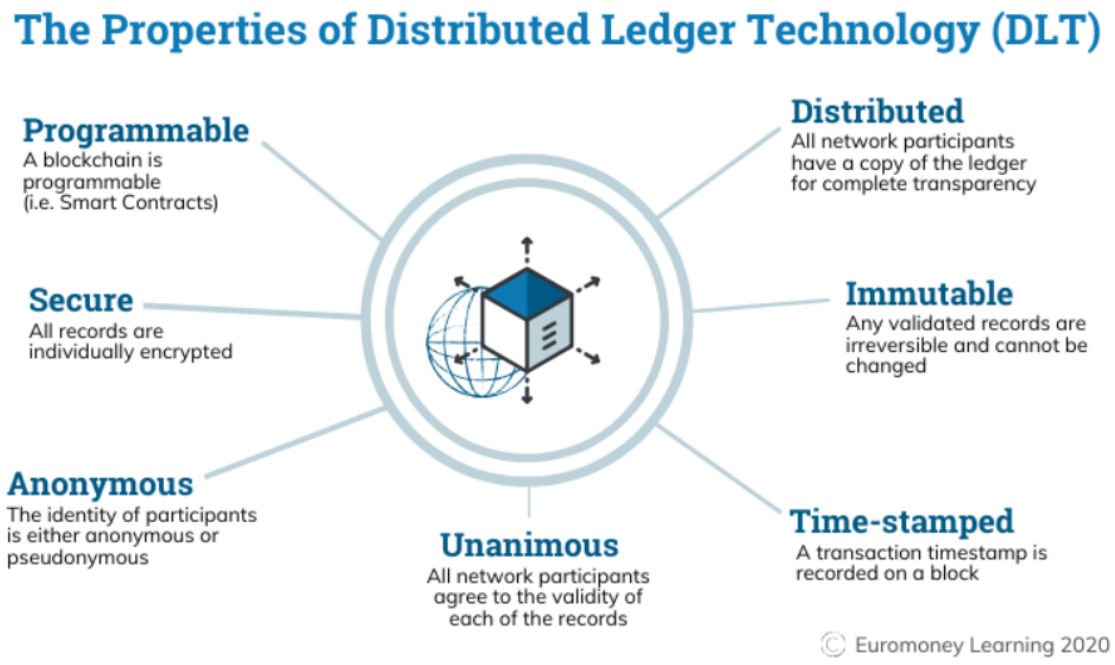
Team Members: Mhai Kunchorn, Noor Syed, Margaret Onakoya

Cryptocurrency and Social Media: Can Twitter Affect the Market?

Cryptocurrency and blockchain technology have been buzzwords in the media and finance industry since Bitcoin came on the scene in 2009. It has garnered fintech fans and critics alike and quickly became a trending hot topic on social media. Since its initial rise, cryptocurrency has had both short and long term fluctuations in the market, which has grown in number of cryptocoins and in complexity over time. With the introduction of new coins and a decentralized, deregulated platform, it has attracted the interest of those both within and outside of the finance industry, including some prominent names and public figures. Using data sourced from LunarCrush API and from research on cryptocurrency, our team has taken a look at the effect social media has played in the cryptocurrency market in terms of several cryptocoin prices and volatility, and analyzed the correlation between the market and the media.

Key Definitions:

- Cryptocurrency: A digital or virtual asset which is secured by cryptography. Online payment using cryptocurrency uses a virtual "token" which gets marked on a ledger entry. The ledger is on a decentralized network across a high number of computers, which makes it very difficult to counterfeit or reuse.
- Blockchain: A digital ledger of transactions that is distributed across the network of computer systems that are on the blockchain. Each time a new transaction occurs on that blockchain, the record is added to each participant's ledger. Each transaction is recorded using a unique and immutable cryptographic signature called a hash.
- Distributed Ledger Technology (DLT): A decentralized database managed across a network of computers by multiple participants. Blocks in the chain contain a number of transactions, and every time a new transaction occurs on the blockchain, a record of that transaction is added to every participant's ledger. Blockchain is a type of DLT.



Project Roadmap

Week 1: Planning

- Discussing topics of interest and possible objectives

Week 2: Planning

- Discussing potential APIs and focusing question we want to answer, delegating work and areas of focus

Week 3: Data Cleaning and Exploration

- Working on code building and exploration of data

Week 4: Data Visualization and Compilation

- Integrating code together and reviewing findings for our overall conclusion

Questions: to invest or not to invest?

Although cryptocurrency is referred to as a "currency", or form of money, it is currently viewed by the US government as a financial asset or property. Given this, as well as its relatively new use in the market and many uncertainties around transacting in cryptocurrency now and into the future, there are inherent risks associated with investing in cryptocurrency. Our team has set out to dive deeper into the question around social media – does what people post online have an effect on the value of cryptocurrency? To what degree is that effect? Given our analysis, would we recommend crytpocurrency as a stable investment, from the perspective of volatility, given public sentiment?

Objectives

- Research cryptocurrency and common cryptocoins
- Pull data around market movements and social media buzz around cryptocoins
- Analyze data to review any correlation between social media volume and market movements
- Draw conclusions on cryptocurrency as an investment

Topic Determination and Background

After cryptocurrency was suggested as an idea, the team researched various APIs which could provide helpful data. From this search of crypocurrency APIs, we located LunarCrush, which tracks price movement as well as social media data and statistics. The social media data drove us to our questions – how correlated are the two, and how does that vary across coins? Each team member then pulled data for their unique cryptocoin and together we worked through the various ways to pull relevant data as it was embedded within the database information and not immediately available. From there we worked on our graph types and data comparisons for consistency in approach and collective views across all coins.

Code First Girls Nanodegree
Data Science Module
August 2021

Key Definitions for API Data

Per LunarCrush Documentation
LunarCrush API

- Social Score: sum of followers, retweets, likes, reddit karma, etc of social posts collected
- Social Volume: number of social posts
- Tweets: number of tweets collected
- Volatility: degree of variation of a trading price series over time as measured by the standard deviation of logarithmic returns
- Close/High/Low: end of day, high, or low price of the day
- Correlation Rank: a score based on how the assets social metrics correlate with price and volume
- Market Dominance: a coin's market capitalization and divided by the total market capitalization of the cryptocurrency market
- Social Dominance: a coin's social volume divided by the cryptocurrency market's total social volume

Steps Specifications and Approach

Through group meetings and check ins, our team have each tackled a unique cryptocoin asset to analyze the price movements, volatility, and social media correlation while collaborating on ideas, code share and suggestions, and data exploration techniques.

- In depth review of LunarCrush documentation to pull relevant and timely data
- Flexible approach with consistent data points analyzed (such as social score and social volume) as well as room to explore other supporting data points for specific assets (such as social and market dominance)
- Outside research of cryptocurrency and tweets which helped provide context for our analysis and specific market movements
- Data was primarily sourced from LunarCrush API as it covered most of the key data points we wanted to review, but we also researched to understand cryptocurrency market and definitions and ensure our analysis was supported and understood, such as our correlation analysis with regard for market news and events

SWOT Analysis

Strengths • Pandas, Matplotlib • Data cleaning • Data visualization	Our team felt fairly confident in using pandas and matplotlib to perform data cleaning, data exploration and data visualization effectively. We were able to both independently carry out these tasks and when we encountered errors were able to communicate and resolve it together. Another strength of our team work and project was efficiency in communication – we were always able to update and share findings via WhatsApp and Slack while not having the time to meet over video call as regularly as we intended.
Weaknesses • Subject matter – cryptocurrency as a new subject • Numpy, Seaborn • Consistent format applied when building code and visualizations	However we tried to incorporate the agile values in our working process but daily SCRUM was not possible so we resorted to more contact via instant messaging and all made our face time over Sundays. The weakness of project was due to the lack of knowledge on the subject matter of cryptocurrency as we had no prior knowledge of the industry resulting in conflict of interest and a team member leaving to join another team where she felt more confident in offering her skills for a more familiar business case. Despite the lack of knowledge the rest of our team pursued cryptocurrency as we felt there was an opportunity to learn about this subject. Finally, we did not have a standardized approach to the format when building the code, which created more work for the integration step.
Opportunities • Practice reading and implementing documentation • Deeper understanding of Python • Organization and time management • Learning about cryptocurrency and the market	Other opportunities we identified were reading and understanding documentation, time management of the project which we did not do well as our planning stage and deliberating over topics and questions took majority of the time away from other areas such as testing. We recognised in order to do well moving forward we would need to be more conscious of time spent on each stage of the project.
Threats • Time allocation – collaboration and number of team members for task delegation • Amount of work coding and presentation to complete in short timeframe • Consolidation of learning Jupyter Notebook and libraries and to implement learning meaningfully in short amount of time	The main threat to our project was due to the lack of time and restricted availability of each member for collaboration and delegation of tasks. There was also a lot of work to get done in a short amount of time and very little time to consolidate the learning of Jupyter notebook and libraries to use it meaningfully in a short space of time. Finally, if we had to do this project differently we would reduce the work load considering the amount of time we had to work on the project and using the time we had more effectively and consciously.

Implementation and Execution

Team member roles:

- Each member selected a cryptocoin asset they were interested in analyzing
  - Mhai researched Bitcoin (BTC) as the flagship cryptocurrency which was created in 2009
  - Noor researched Ethereum (ETH) which has risen in prominence as the cryptocurrency asset with the second highest market capitalization
  - Margaret researched Stellar (XLM) to view what role social media plays on a less popular cryptocurrency
  - Kristen researched Dogecoin (DOGE) which was created in 2013 based off a popular Shiba Inu dog meme
- Kristen assumed role as project manager to help drive the team forward and stay on track with the project requirements and approach, deadlines, and data integration and testing of code and PDF
- Margaret is driving Powerpoint presentation

Tools and Libraries:

- Pandas
- Matplotlib
- Numpy
- Json
- Flask
- Requests
- Urllib
- Seaborn
- Plotly
- Cufflinks
- Math
- Pprint

Agile Framework

Our team used agile values to navigate our project as it allowed us the agility in our working process. The planning stage consumed majority of our project time, we were able to assign a project manager within our team to oversee progress and make decisions to move the project forward respective of the team feedback. Having chosen finance as our domain, we discussed the different topics we were interested in regarding finance and came to the conclusion of pursuing the cryptocurrency idea. We had then devised a plan to look at different API's and came across LunarCrush API which would be ultimately be where we pull the data for our project.
At this stage we spoke about arranging regular meet ups to discuss and work on the project, however this proved to be more challenging than planned as everyone had a full time job as well as having to prioritise the Nanodegree and personal well-being. We realised that the daily SCRUM was not possible and decided that Sundays were likely the best time to meet as that would be a time that worked with everyone and began creating group chats on Slack to share links and progress, a WhatsApp group for instant communication and setting up a repository on GitHub where we would upload our work.

Coming to the 'big question' was another challenge at the planning stage, there were also conflict of interests as not all members of the team felt like there was a strong business case for cryptocurrency; this is because we were also not familiar with the Cryptocurrency market and the way it worked. However, this posed a good opportunity to do research and work on a new field as data science can benefit numerous industries. In the workplace for those who may go on to work for firms which leverage data and business insights there, will be many opportunities where clients come from industries which they may not be familiar. Hence doing research and speaking to people within that field will help to realise and form questions to perform data analysis. We had then created a Miro board to put our ideas down which can be found here: Miro Board. You can also see evidence of us considering taking our project in a different direction when we were uncertain about the cryptocurrency idea. The question we came to was whether there is a correlation between cryptocurrency assets and their social media engagement, and what does this correlation tells us about whether it is a stable asset to invest in or not?

Phase 1: Requirements

- With our question formed, this serves as our key objective and determines what our end project will look like
- Primary data source selected as LunarCrush API
- Key libraries discussed including Pandas, NumPy, Matplotlib, and Seaborn for visualization analysis
- Work delegated across team with each team member analyzing a specific asset
- Key data points identified for further review which also allowed flexibility per asset, such as social score and market dominance

Phase 2: Design

- Design focused on data analysis framework and ensuring appropriate time spent on each step, including selecting useful data, importing and extracting valid data, cleaning, exploring, and visualizing the data for meaningful analysis
- This phase could have been improved and more time spent to improve consistency of design for coding format to assist with integration phase

Phase 3: Development and Coding

- Building code per the data analysis framework
- Collaboration amongst team to help overcome obstacles and review what the data looks like across the different assets

Phase 4: Integration and Testing

- Compiling source code into one cohesive project
- Updating dataframes to clear and unique naming conventions, consolidating repeat or similar code, ensuring all necessary code is included in order to run each piece of code
- Accounting for order of code execution – certain graphs and information must be pulled in a specific order based on the format of the dataframe
- Testing to ensure each line of code runs successfully and shows appropriate data, with all steps of the framework captured

Phase 5: Implementation and Deployment

- Presentation of project to CFG class to determine our objective and findings are clear and understandable outside our group
- Submission of source code to instructors for review of project execution as the deliverable to final stakeholder

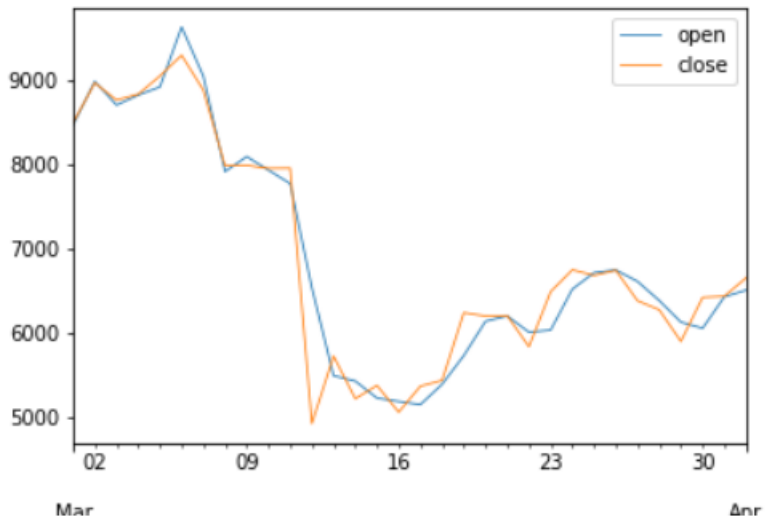
Phase 6: Review

- Review takeaways and lessons learned from our project
- Consider what we would have done differently or what we would like to explore more if we had more time

Key Findings

Analysis Summary – Bitcoin

From our observation, there is certainly a correlation between social media sentiments and Bitcoin price fluctuations, from our data analysis on specific days where we saw a dramatic percentage decrease as well as the spike in the closing price. However, these investigations were performed on outliers, if we are to look at the overview on the percentage change chart and compare it with the social volume chart, we can see that Bitcoin's percentage change is much more stable over the 2 year period despite the social volume having an upward trend over the 2 year period. To be able to create a forecast and more accurate prediction if the volatility of this investment, we would need more than 2 years worth of historical data as the past two years we have faced unprecedented events such as the pandemic which we are still navigating through. Without more data, the prediction does not wholly reflect the true volatility of the cryptocurrency market and specifically for Bitcoin.

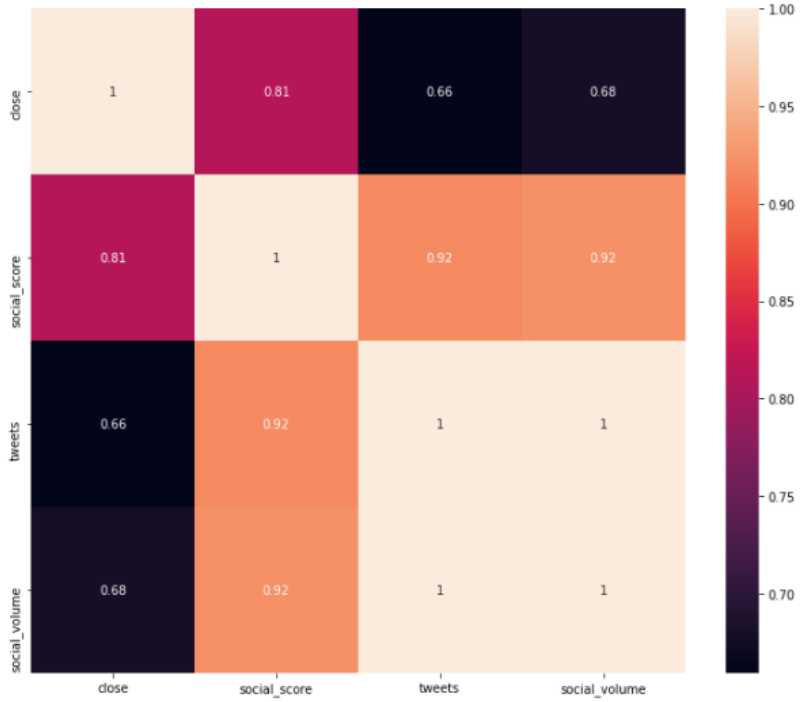


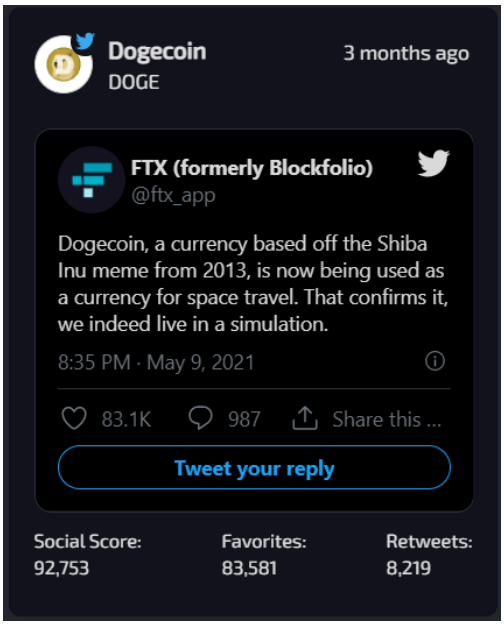
On this chart we can see that on around the 9th March 2020 there was a gradual decrease in price and ultimately on 12th March 2020 BTC opened at $6541 and closes at $4929, a 37% decrease.

In 2021 there was a noticeable outlier between March - May and the discoveries on the following particular days are as follows:
8th March 2021 – Tesla buys $1.5billion worth of bitcoin (43,000 bitcoin) driving up the price as it gave investors confidence that one of the world's most innovative companies trusted in the future of the currency as an investment. Open: $39329 Close: $46095 Percentage Change: +18%
22nd April 2021 – Capital Gains Tax increase announcement by Joe Biden sent retail investors selling their assets in order to avoid having to pay taxes on it and presumably moving into other off-shore investments. The price opened at $54189 and closed at $52017, a 4% decrease.
12 May 2021 – Elon Musk tweeted that bitcoin would no longer be accepted by Tesla as a cryptocurrency payment until it uses more sustainable mining practices. BTC on this day opened at $55168 and dropped 11% and closed the day at $50088. This could be due to lack of confidence from a huge investor.

Analysis Summary – Dogecoin

Dogecoin was created in 2013 by software engineers Billy Marcus and Jackson Palmer, originally as a joke to "poke fun at Bitcoin". In January 2021 the price skyrocketed due to social media activity on Reddit. The same WallStreetBets thread which explained the concept of short selling to Reddit users and used this community to buy and drive up the value of GameStop stock, also called attention to Dogecoin, which is based on a meme. After this Reddit thread skyrocketed the cryptocoin price from obscurity, social media has continued to have an ongoing effect on the price value and fluctuation. At its peak, Dogecoin rose by nearly 100% from where its price started at the beginning of January. It has attracted high profile investors such as Elon Musk and Mark Cuban and our analysis shows that spikes in volatility and price frequently coincide in terms of movement.





Our heatmap also shows a significant and material correlation of 0.81 between close price and social score, and research into the coin itself shows that this correlation cannot be ruled out as a coincidence. This is supported by the asset's initial rise due to social media platform and community focus, and ongoing news and social media volume and content which shows persons of influence directly commenting on their investment status, support of, or opinion of the asset, during times which coincide with high volatility and price fluctuations and spikes.

Analysis Summary – Ethereum

Based on the data analysis conducted for Ethereum, it was found that social media buzz did not have an adverse impact on the open and close price or volatility of the coin. Instead, it was rather due to upcoming software or user developments for Ethereum or knock-on effects from news or conference developments which affected the coins popularity.

Analysis Summary – Stellar

I wanted to analyse the Stellar coin because it is not as popular as Bitcoin or Ethereum. I wanted to see whether the lack of popularity and visibility within social media would effect the open and close price or its market dominance. Overall, because Stellar social dominance is so little, it has no effect on its market dominance and even its Volatility. However on Jan 2021 Stellar received some popularity on Reddit, but for negative reasons. This is demonstrated by the sharp spike of social volume and volatility. Also during this time there was a sharp plummet in Stellar's market dominance.

Conclusion

Our findings show that there is a moderate to high correlation between what is happening in the market and the volume and content of engagement on social media. Some cryptocoin assets, such as Dogecoin, were highly correlated between price and social score. Based on volatility and correlation data, along with sourced tweets and news regarding the coin and social media, it is a fairly volatile asset whose price can be driven from social media directly. For this type of asset, investment can be seen as more risky as it is fairly dependent on the current public sentiment.

For other cryptocurrency assets, this is not as high a concern. For more seasoned assets such as Bitcoin and Ethereum, although social media and price are correlated, our in depth analysis shows that this is more directly driven by the overall market environment and sentiment, such as new capital gains tax regulation, high profile investors, and companies that accept the coin as a form of payment or not. However, when there is a material or unexpected declaration on social media regarding these assets, which quickly reaches many more users than just those who are directly connected to the financial market, this can cause a large spike even in these more stable assets.

Overall, cryptocurrency and blockchain are relatively new and unreglulated assets which can vary even among the different cryptocoins themselves, and should be considered with volatility, market risk and public sentiment risk in mind. Our findings show that although asset prices can be fickle, it is a fast growing market driven by interesting technology and new cryptocoins continue to be introduced and take a stronghold on the market. Whether the cryptocoin is introduced as a joke, or as the hot new way to invest in going to space, both the financial markets and the wider public are willing to invest to see where cryptocurrency goes next.

Sources

- https://lunarcrush.com/developers/docs#
- https://www.investopedia.com/terms/c/cryptocurrency.asp
- https://www.euromoney.com/learning/blockchain-explained/what-is-blockchain
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6107969/
- https://www.forbes.com/sites/sarahhansen/2021/04/22/biden-will-seek-to-nearly-double-capital-gains-tax-for-wealthy-according-to-report/
- https://python.plainenglish.io/how-to-predict-stock-volatility-with-python-46ae341ce804
- https://www.forbes.com/sites/billybambrough/2021/05/19/crypto-price-crash-bitcoin-drops-under-40000-as-ethereum-binances-bnb-cardano-and-dogecoin-collapse/?sh=4577fa6635c5
- https://www.cnbc.com/2021/07/21/elon-musk-jack-dorsey-speak-about-cryptocurrency-at-b-word-conference.html
- https://www.forbes.com/advisor/investing/what-is-dogecoin/