

# Analysis of Twitter's Influence on Bitcoin and Ethereum

Sanya Dod

**Abstract—** In the recent years, cryptocurrency has gained a lot of interest among the investors and academicians because of its increasing market capitalization. Bitcoin and Ethereum are one of the biggest digital currencies by market cap, yet their value doesn't depend on institutional cash guidance or backed by sovereign gold, rather on individual insights and speculations. Similarly social media's growth has been exponential – 100% increase in number of users in mere five years. Hence, examining the relationship between social media references and web search is critical for establishing or forecasting the value of cryptocurrency. In this research, twitter trends are utilized to gauge the price movement of digital currencies and establish if tweets can be utilized to impact the buying choice. Regression and correlation methodology are applied to assess the statistical significance and magnitude of strength of the relationship between closing price of Bitcoin and Ethereum with respect to volume of tweets.

**Index Terms—** Bitcoin, Ethereum, Regression, Social Media, Twitter

## I. INTRODUCTION

The innovation in technology has prompted various advances in financial domain, be it efficiencies in lending, AI based financial modelling, or new forms of digital currencies. In digital currencies, cryptographic currencies have sprung up as encryption techniques are used to control the generation of units of currency and verify the transfer of funds, working autonomously of a national bank [2]. Cryptocurrencies are globally spreading and are acquiring an immense measure of consideration from media, governments and financial organizations.

The most prominent and oldest cryptocurrency is Bitcoin, which was presented in 2008 by group of persons or individual passing by the name of Satoshi Nakamoto [3]. Bitcoin has been introduced as a decentralized, borderless technique for funds transfer. At present, there are more than 400 exchanges on which Bitcoin can be traded from any part of the world. Moreover, well establish exchanges, like NASDAQ, have started showing bitcoin prices thus giving prominence to the currency.

Ethereum paper was published in 2013 by Vitalik Buterin and was launched in 2015. It is similar to Bitcoin in terms of leveraging Blockchain technology and being decentralized. Its main use is being promoted as enabling Smart contracts.

These two biggest cryptocurrencies, estimated by market capitalization, have a joint market worth of 1087 billion dollars, by July 2021 [1]. Bitcoin alone made up almost \$783

billion of this worth. Given the huge worth of these monetary forms, a few groups consider them to use as genuine monetary forms, while others see them as investment openings. This interest for currencies has resulted in huge swings in the prices of the two currencies throughout brief timeframes. During 2017, the worth of a solitary Bitcoin expanded 2000% going from \$863 on January 9, 2017 to a high of \$17,550 on December 11, 2017 [1]. Similarly Ethereum's price fluctuated from USD 736 on Jan 1, 2021 to 3986 on 11th May 2021, a variation of about 400% in 4.5 months.

In this paper, an association between tweets and different crypto currencies (BTC and ETH) are analyzed along with a prediction model which assess the statistical significance of model prediction factor.

## II. LITERATURE REVIEW

Cryptocurrency is a decentralized virtual money, thanks to block chain innovation, which employs cryptography for security. The most notable cryptocurrency is Bitcoin. In contrast to fiat, monetary standards controlled by customary national banks which determine paper currency in circulation, Bitcoin's inventory is restricted as there are just 21 million Bitcoins in circulation. The worth of Bitcoin isn't backed by gold or different products, but it depends entirely on a worth as determined by supply and demand.

Various research studies done on assessing Bitcoin's reasonable worth have come to conclusion that Bitcoin's price is fundamentally evaluated through hypothesis /speculation. In their works, Fantazzini et al. explored models for determining basic worth through strategies for market estimation and looked into various econometric models which utilized cross-sectional, univariate, and multivariate time series models [4]. They infer that the main drivers of bitcoin price dynamics are still mainly of speculative nature, followed by traditional supply and demand related factors. Further global macro financials variables play no role.

One reason for Bitcoin and Ethereum expanding prevalence is web-based media. In the recent times, social media's participation into everyday lives has expanded at an astounding rate. Social media platforms gained 490 million users in 2020 (that's a 13.2% increase, year over year; the previous year's growth was only 7.2%). Social platforms gained 1.3 million users every day, or 15.5 new users every second in 2020. 4.2 billion people use social media around the world (i.e., 53.6% of the planet's total population) 90% of people with access to the internet use social media [5]. One type of correspondence is called microblogging through which clients can, in short messages, share their opinions and

sentiments. One main platform that has been utilized in different studies to mine information is Twitter. Twitter has 353 million users, and 192 million daily active users [5]. Further, the average US user spends 32.7 minutes per day on Twitter and has experienced 36% growth in site-traffic. Other than, the number of clients, Twitter also contains clients that differ from famous people, organization agents, and as Twitter is utilized around the world, it is feasible to get information in various dialects [6] and use Twitter as a corpus for sentiment analysis and opinion mining. All these components make twitter a rich source of information for data analysis.

III. METHODOLOGY AND DATA ANALYSIS

A. Figures and Tables

Different types of statistical analysis were done to examine the past trends between tweets and the price of both cryptocurrencies BTC and ETH.

Publicly available historical pricing dataset [7] from Kaggle is used for this research. This pricing dataset has the historical price information of some of the top crypto currencies by market capitalization.

- Date : date of observation
- Open : Opening price on the given day
- High : Highest price on the given day
- Low : Lowest price on the given day
- Close : Closing price on the given day
- Volume : Volume of transactions on the given day
- Market Cap : Market capitalization in USD

Pricing dataset has data from 29-04-2013 to 27-02-2021 for Bitcoin and from 08-08-2015 to 27-02-2021 for Ethereum.

Historical tweets can be collected using Twitter APIs but these take lot of time to gather historical datasets given the throttling on twitter APIs . In this research, datasets gathered by Pathak, Ajeet Ram [8] for Bitcoin, Ethereum and Facebook are used. They have labeled tweets by finding keywords ‘bitcoin’ and ‘ethereum’ appearing in tweets. In the datasets, it was observed that many bitcoin tweets are also referred by #btc and ethereal by #eth also. So reclassification of originals tweets is done to include tweets mentioning ‘#btc’ or ‘bitcoin’ as Bitcoin Tweets and ‘#eth’ or ‘ethereum’ as Ethereum Tweets. Following are the results of data augmentation:

	<u>Bitcoin Tweets</u>	<u>Ethereum Tweets</u>
<u>Raw Tweets</u>	4,936,184	3,809,617
<u>Reclassified Tweets</u>	6,315,432	3,869,130

Tweets datasets has data from 31-07-2015 to 08-10-2018.

The Reclassified Tweets data set is merged with Pricing dataset, and subsequent research was done on merged dataset.

Merged dataset contains daily date of closing price and corresponding number of tweets.

Firstly, graphical representation of cryptocurrency prices and tweets was done to visually understand the trends. Then, followed by the rate of change in price analysis through graphical representation. After which a correlation analysis is performed between tweets and price for each type of crypto currency to analyze the correlation strength. Lastly, a regression analysis is performed to analyze the statistical significance of variables.

For re-classification and merging of datasets, programs are developed using Python and MySql, and for correlation and regression analysis Microsoft Excel software is used.

B. Data Analysis And Results

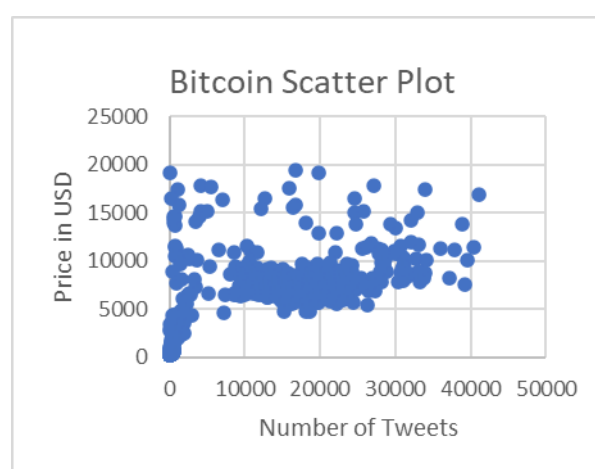


Figure 1: Bitcoin Scatter Plot

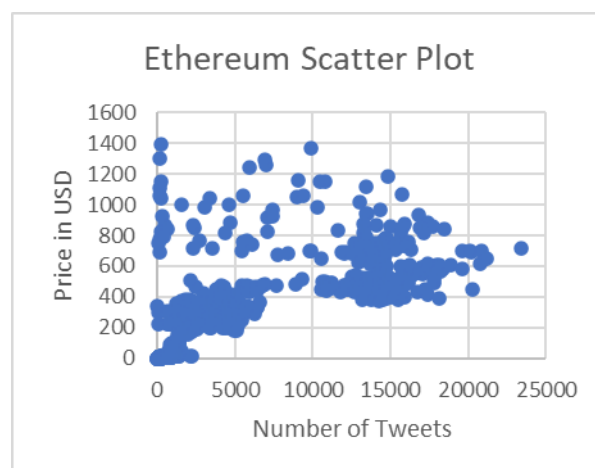
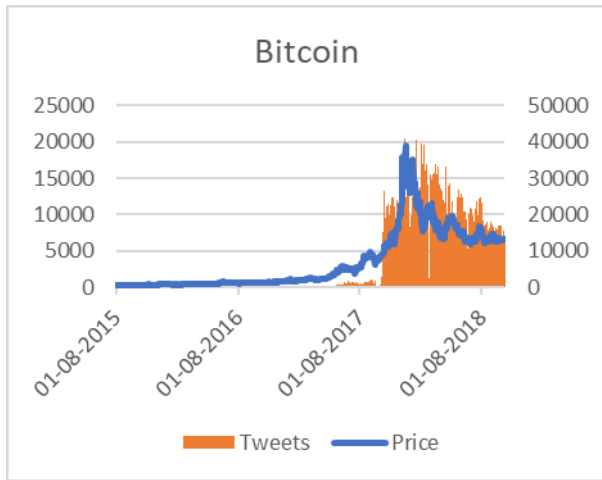
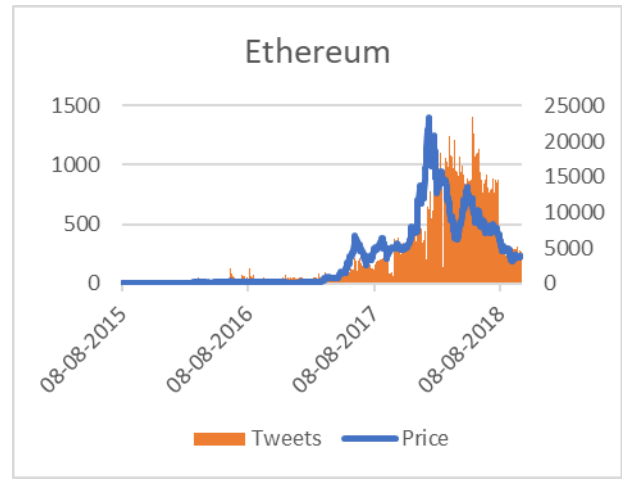


Figure 2: Ethereum Scatter Plot



*Figure 3: Bitcoin Price vs Tweets*



*Figure 3: Ethereum Price vs Tweets*

C. Correlation And Regression Analysis

*Table 01: Correlation between closing price and Respective Currency tweets*

	<b><u>Bitcoin</u></b>	<b><u>Ethereum</u></b>
<b>Pearson Correlation</b>	<b>0.764</b>	<b>0.741</b>
<b>Sig (2-tailed)</b>	<b>0.000</b>	<b>0.000</b>
<b>N (Number of Observations)</b>	<b>1090</b>	<b>1108</b>

*Table 02: Model Summary of Regression Analysis*

	<b><u>Bitcoin</u></b>	<b><u>Ethereum</u></b>
<b>Model</b>		
<b>R</b>	<b>0.764</b>	<b>0.741</b>
<b>R-Square</b>	<b>0.583691</b>	<b>0.54909</b>
<b>Adjusted R-Square</b>	<b>0.5833</b>	<b>0.5486</b>
<b>St.Error of Estimates</b>	<b>2563.7</b>	<b>187.27</b>

Table 03: ANOVA - Bitcoin price & currency tweets

	Regression	Residual	Total
Sum of Squares	10026365989.89	7151147982.	17177513972.5
df	1	1088	1089
Mean Square	10026365989.	6572746.308	
F	1525.445		
Sig	.000 <sup>b</sup>		

Table 04: ANOVA - Ethereum price & currency tweets

	Regression	Residual	Total
Sum of Squares	47237153	38790601.027	86027754
df	1	1106	1107
Mean Square	47237153	35072.876	
F	1346.8		
Sig	.000 <sup>b</sup>		

IV. DISCUSSION

Figure 1 shows scatterplot of Bitcoin Price and number of tweets. Looking at the plot, it’s evident that there exists a correlation between price and number of tweets of Bitcoin. Similar correlation is evident for Ethereum as shown in Figure 2.

To further investigate the correlation, price and tweet counts are plotted in the same graph, shown graphically in Figure 3. Up until the mid of 2017 number of tweets and the closing price of Bitcoin remained stagnant. However, towards the end of the year 2017 the average number of tweets started to increase, as well as the closing price of Bitcoin. An increasing trend can be seen towards the last month of the year 2017. Hence, indicating that there might be a strong correlation between the closing price and tweets

which can be verified through correlation analyses. After the year 2017, number of tweets started to decrease and with that closing price also decreased. Therefore, a clear parallel trend of increase and decrease of the average number of tweets and closing price as shown in figure 1.

Figure 4 shows the trend in closing price on Ethereum and number of tweets. Here as well, there is a strong correlation of increase in price with respect to tweets from early 2017 to early 2018. Towards the later part of 2018, price of Ethereum started to drop and so does the number of tweets.

Moving on to the correlation analysis in table 01. The correlation value between the closing price and the number of tweets for BTC is 0.764, indicating a strong positive correlation strength. Hence, if the number of tweets (positive) increases the price of BTC will also increase. Moreover, the analyses also revealed that correlation between the closing price and tweets is statistically significant because the p-value

(sig value) is 0.000. Being less than alpha value 0.05 the correlation proves to be statistically significant.

Tables 02 and 03, on the other hand, shows the regression results of BTC and tweets in which BTC closing price was dependent variable and number of tweets as independent variable. Model summary (table 02) shows that 58.3% variation in BTC closing price is due to a variable number of tweets. The rest of the variation is due to other factors not accounted for in this model. On the other hand, ANOVA table 03 shows that the variable number of tweets is a statistically significant variable in predicting the closing price of BTC, because the p-value is 0.000. Hence, the regression model proves to be significant.

Analyzing the correlation coefficient of the closing price of ETH and the number of tweets shown in table 01. The correlation coefficient of 0.741 indicates a strong positive relationship strength being statistically significant at a 0.05% level of alpha. Table 02 and table 04 again show the regression result of the closing price of ETH and the number of tweets. Model summary (table 02) shows that 54.9% variation in ETH closing price is due to the variable number of tweets related to ETH. On the other hand, regression results in table 04 confirm that in the case of ETH just like BTC variable number of tweets is a statistically significant variable in predicting the price of ETH. Hence, both the models prove to be significant

## V. CONCLUSION AND FUTURE

From the analysis of the data, it is evident that web-based media references and web information bases like tweets have an impact on cryptocurrency price changes. Regarding the dependency of price of cryptocurrency with respect to tweets, it can be reasoned that the Twitter estimations concerning cryptocurrency value will, in general be, positive. However, there are many other factors which also influence the price of the cryptocurrency like bank guidelines, political risk and regulations.

Hence, further intricate analysis can be explored with additional factors taken into consideration for building the statistical model which can lead to an effective prediction of the crypto currency.

## REFERENCES

- [1] <https://bitinfocharts.com/>
- [2] Schuettel, P. (2017). The concise Fintech compendium. Fribourg: School of Management Fribourg/Switzerland
- [3] Nakamoto, S., & Bitcoin, A. (2008). A peer-to-peer electronic cash system. Bitcoin.–URL: <https://bitcoin.org/bitcoin.pdf>,
- [4] Fantazzini, D., Nigmatullin, E., Sukhanovskaya, V., & Ivliev, S. (2016). Everything you always wanted to know about bitcoin modelling but were afraid to ask. Applied Econometrics
- [5] Cooper, P. (2021). 140+ Social Media Statistics that Matter to Marketers in 2021. [online] Hootsuite Social Media Management. Available at: <https://blog.hootsuite.com/social-media-statistics-for-social-media-managers/>. (Accessed on 05/20/2021)
- [6] Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010, pp. 1320-1326)
- [7] SRK (2021), Cryptocurrency Historical Prices :[online] <https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory>
- [8] Pathak, Ajeet Ram (2019), "Social media data #bitcoin #Ethereum #facebook ", Mendeley Data, V1, doi: 10.17632/chx9mddydb.1