THE IMPACT OF PUBLIC SENTIMENT ON CRYPTO RESILIENCE DURING COVID-19 PANDEMIC

University of Twente - Faculty of Behavioral Science
International Busine Administration

Caţalab, Augustin Ionut : s2269287 Student
Dr. Marisetty Vijaya: 1st Supervisor
Dr. Machado Marcos: 2nd Supervisor

Table of Contents

1. Introduction 1
   1.1.Research objective 1
   1.2.Academic and practical relevance 1
   1.3.Research question: 2
2. Literature review 2
   2.1.Cryptocurrency and crypto-resilience 2
   2.2. Sentiment Analysis and crypto market research 2
   2.3. Relevant models guiding sentiment analysis 3
   2.4. Relationship between global crisis and financial performance 3
   2.5. Concluding literature 4
   2.6. Limitations 4
3. Methodology 4
   3.1. Data collection 5
   3.2.Sentiment model selection 5
   3.3.Regression and correlation analysis 5
   3.4. Variable selection 5
4. Analysis 6
   4.1. Sentiment analysis 6
   4.2. Correlation and regression analysis 7
5. Conclusion and discussion 8
6. References 8
7.Appendix 9
   1. Volatility analysis 10
   2.Return analysis 11

1. Introduction

If in its early stages of development, Cryptocurrency was more of a taboo subject, today is a topic discussed on a daily basis by every mass media outlet. Although the concept is known to humanity ever since the 1980s through the book of cryptographer David Chaum titled "Blind Signatures of untraceable payments". He was also the man to invent the first cryptocurrency called E-cash, and one of the people at that time thinking progressively at the payment system and how it can address the personal privacy and the nature and extent of criminal use of payment"(Kvilhaug S. 2022). From there, a few decades later and a few other failed attempts, Bitcoin appeared as a rejuvenated version of all its failed predecessors such as eCash, B-money, Bit Gold, and Hashcash.

The exponential growth of cryptocurrencies has attracted a lot of attention in recent years and offered the opportunity of all sizes investors to participate in the market. (Caporale, G. M 2018).

The global pandemic spurred interest and investment in cryptocurrencies such as Bitcoin and Ethereum, as well as the growing use of blockchain technology in areas such as supply chain management and digital identity verification (G. Iyer 2021). In an article for McKinsey describing the market performance at the beginning and during the pandemic, it was concluded that the pandemic has acted as an accelerator to pre-existing trends in the capital market such as the rise of digitalization (C. Bradley 2021). Similarly, R. Van Hoek 2020 for Harvard Business School, suggests that the COVID-19 pandemic has presented an opportunity for blockchain technology to gain traction and be more widely adopted.

Research studies have shown that behavioral factors, such as herding theory, prospect theory, and heuristic theory, significantly affect investors' investment decisions in the cryptocurrency market (Almansour, B.-2020). Herding behavior, where investors tend to follow the actions of others, has been observed and is commonly known as herd behavior (Wang, Z et al.2022). This behavior can lead to market inefficiencies and the formation of speculative bubbles (Haykir, O., Yagli, I. 2022). The presence of speculative bubbles underscores the importance of understanding investor sentiment and emotions as they can have a profound impact on cryptocurrency prices and volatility (Haykir, O., Yagli, I. 2022).

In research on the impact of online public opinion regarding the nuclear water crisis in Japan, (Hong, W. (2023) observed an inverse correlation between positive sentiment and stock returns and an increase in volatility. The study concluded that the positive changes in public sentiment translated into a decline in stock returns in the current period and an increase in return in the lag period.

Several studies have explored the relationship between investor sentiment and market dynamics during crisis periods, providing insights into why investors act on sentiment analysis hints during these times. One of the reasons is market timing, which plays a crucial role in identifying trends and firm performance during pre and post-financial crisis periods (Zainudin, Z et al 2019).

Secondly, although it may not provide a safe haven against inflation in times of crisis, as Conlon, T. (2020) argues, cryptocurrency such as bitcoin, can, however, provide “a substantial increase in portfolio downside risk” and, as H.Chaudhari-2020 argues that while using text mining of tweets and articles to determine the relationship between this and price, sentiment does not have an immediate effect on market price but rather can be used to predict with high accuracy the direction.

The opinion of whether sentiment analysis is a good instrument to measure price changes for both crypto and financial markets is ambivalent.

The above-mentioned are especially important as the survey conducted by NORC at the University of Chicago with a sample of over 1000 people, found that the average crypto trader is a male under 40 and 55% do not have a college degree(NORC- 2021). The survey also reveals that the investment decision is based on the information given by trading platforms (26%) and/or social media(25%) and only 2% reach out to professional advice (NORC- 2021). Concluding that the average investor can be easily influenced.

Further on, it can be argued that, based on the given evidence of a relation between sentiment analysis and market fluctuation, as well as the push towards digitalization and the spurred interest in cryptocurrency investment, it is worth researching whether a change in sentiment occurred during the Covid-19 pandemic and whether a positive or negative sentiment correlates with a change in cryptocurrency prices.

1.1.Research objective

Although some consistent study has been carried out on sentiment and the financial market regarding sentiment and market price, I found few current studies that address the historical accuracy of sentiment polarity and crypto market movement.

The objective of this study is to investigate the change in sentiment during the Covid-19 pandemic and examine the correlation between positive/ negative sentiment and cryptocurrency price fluctuations.

Specifically, the study aims to examine public sentiment through open source tweets, before 2020 (pre-crisis) and after 2020 (during the crisis) and determine whether a change in sentiment occurred given the technological push on digitalization and blockchain as the article suggested, (G. Iyer 2021, C. Bradley 2021, R. Van Hoek 2020).

By analyzing sentiment data and cryptocurrency price movements, the research seeks to provide insights into the impact of sentiment on the cryptocurrency market and contribute to our understanding of the relationship between sentiment and price volatility during a crisis event.

1.2.Academic and practical relevance

This research aims to provide valuable insights for investors, researchers, and individuals interested in cryptocurrencies, with a focus on long-term financial investment. By analyzing the risks and opportunities associated with cryptocurrencies, this thesis can help the general public make better decisions about adding cryptocurrencies to their investment portfolios

The research can represent a source of information about the population's reaction to a global crisis and can offer a suggestion in regard to how the crypto market may react in the future. This aspect can have an important impact on an investor's portfolio.

Additionally, this research contributes to the existing literature on cryptocurrencies, which is still relatively limited compared to studies on the financial market. Overall, the findings of this research can increase knowledge and understanding of cryptocurrencies, potentially leading to more profitable investments and better utilization of opportunities.

1.3.Research question:

Is there a correlation between positive or negative sentiment polarity and cryptocurrency price fluctuations?

The Null Hypothesis H0 = There is no significant statistical connection between the sentiment polarity (negative/positive) and cryptocurrency price fluctuation during the Covid-19 Pandemic.

The Alternative Hypothesis H1= There is a significant connection between the sentiment polarity (negative/positive) and cryptocurrency price fluctuation during the Covid-19 Pandemic.

2. Literature review

In this part, the study aims to describe the theoretical concepts approach in the research and provide context and established foundation for the research study. This part will describe the following concepts: cryptocurrency and crypto resilience, relevant work with sentiment analysis and the crypto market, relevant theoretical frameworks guiding sentiment analysis, and the connection between global crisis and financial performance.

2.1.Cryptocurrency and crypto-resilience

As was mentioned in the introduction David Chaum was one of the pioneers of this market, he envisioned a world where third-party involvement in currency transaction to support security will not be necessary, This is because cryptocurrency is a form of digital currency that distinguish itself from fiat money by facilitating per-to-per transaction, throughout the use of blockchain technology, which ensures transparency security and immutability of transaction.

Another influential figure in the "crypto universe" is Satoshi Nakamoto who through his paper "Bitcoin: Peer-to-peer electronic cash system" proposes Bitcoin as a solution to the double spending problem and avoiding a third party for electronic cash transactions (Nakamoto, S. 2008).

Numerous studies looked at what are the diverse factors contributing or hindering the crypto market resilience. One particular study looked into the association of attention to the socio-political conflict between Russia and Ukraine (R.Khalfaoui-2023), measured through Google Trends in 2022. Using quantile cross-spectral analysis it was observed that War attention affects all currencies in the short run and as a response investors seek liquidity.

Early studies on factors influencing the choice of cryptocurrencies to mine and/or use, sheds light on the significance of popularity variables in the cryptocurrency landscape. With a focus on user opinions and preferences, the study reveals that among community strength, currency value, and ease of use/mine, popularity, logo, and name play a key role in selecting a cryptocurrency. (A.A Shehhi 2014).

Studies examining mining models emphasize the importance of self-defense algorithms to enhance the resilience and robustness of cryptocurrency systems(Zhou, C.et al.-2022) against terrorist attacks. And other studies emphasize the network system's characteristic to withstand attacks and carry on transactions providing high resilience and security(D. Segura-2018).

Concluding, the resilience of cryptocurrency is dependent on a multitude of factors such as popularity, defense mechanism, P2P network system characteristics, socio-political events, social-media attention, and arguably the potential to provide diversification and reduce liquidity risk (Ghabri, Y. et al. -2020 & H.Chaudhari-2020).

2.2. Sentiment Analysis and crypto market research

Given the constraints of human cognition and capacity, consistency in processing large amounts of information is a considerable challenge. As such, there is a need for automated systems that can mine and summarize opinions, as these can overcome subjective biases and mental limitations through the use of objective analysis (Liu, B. 2011).

According to G. Vinodhini (2012), sentiment analysis, or opinion mining, is a natural language processing technique used to monitor the general public's attitudes toward a specific topic or product. It involves creating a system that can gather and analyze opinions expressed in various forms such as blog posts, comments, reviews, and tweets.

Sentiment analysis plays a significant role in understanding market behavior and decision-making in various domains and the benefits of using it are manifold.

Predicting customer behavior through the use of sentiment analysis it was observed that its application can provide a comprehensive tool to analyze customer information and feedback for smart digital marketing applications.(Kyaw, N et al. 2023).

Timing the market, which plays a crucial role in identifying trends and firm performance during pre and post-financial crisis periods (Zainudin, Z et al 2019).

Conclusive results have been brought to light in previous studies which have investigated the relationship between sentiment analysis and cryptocurrency values, using machine learning (ML) models and deep learning (DL). These studies aim to determine whether public sentiment, as measured through social media data, can provide insights into or predict cryptocurrency price fluctuations. The results indicate that there is a degree of explanatory power in sentiment analysis when correlated with price volatility.

One such study conducted by D'Amato (2022) focused on analyzing cryptocurrency using sentiment analysis techniques. The research employed ML and DL models to examine the relationship between public sentiment and cryptocurrency values. The findings suggested that sentiment analysis can offer some explanation for the price volatility observed in the cryptocurrency market but it does not offer a strong correlation.

In a similar vein, S. Rhaman (2018) conducted a research paper exploring the association between user sentiment and Bitcoin price fluctuations. The study employed sentiment analysis techniques and statistical methods to analyze user sentiment toward Bitcoin. The research discovered that positive user sentiment was often followed by price increases in Bitcoin, while negative sentiment was associated with price declines. The research further emphasized the usefulness of sentiment analysis as a tool for understanding the factors driving Bitcoin price movements and predicting future prices.

Another study investigating the impact of online public opinion regarding the nuclear crisis in Japan, (Hong, W. (2023), concluded a loss in stock return and an increase in volatility for the short term. even if the sentiment analysis online showed a positive score as dominant

Although a very powerful tool, it's not without any criticism. In a survey on sentiment analysis, while acknowledging the benefits, the author points out the challenges which can hinder the analysis performance. (Hussein D.G.-2016) Some of these factors are negation, the domain which is being analyzed, and the current

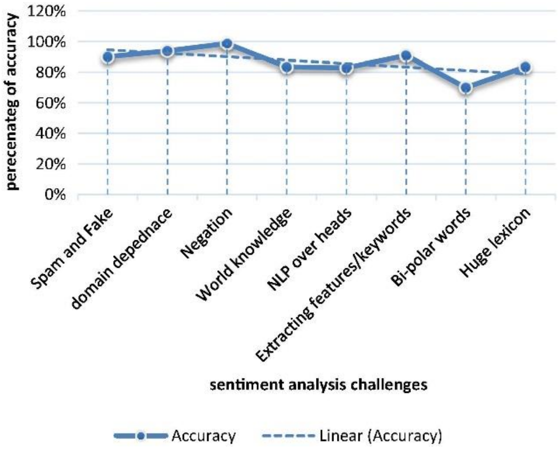

Figure 1Sentiment Analysis challenges & accuracy

knowledge of the subject. Additionally Raheman (2022) in his natural language model comparison mentioned additional obstacles such as Idioms and Non-text data(audio, video, images).

Overall, these studies highlight the potential of sentiment analysis in uncovering patterns and relationships between user sentiment and its interaction with a certain subject such as cryptocurrency, financial market, and customer feedback. By leveraging sentiment analysis techniques, researchers gain insights into the impact of public sentiment on cryptocurrency market dynamics and can potentially predict and assess future or historical price movements.

2.3. Relevant models guiding sentiment analysis

When dealing with a large amount of data, which needs to be analyzed and compounded into meaningful and interpretable results, Natural Language Processing (NLP)is widely used and has developed so fast that we have better ways to deal with raw texts than in recent years. A great deal of scientific papers uses this processing framework in collaboration with pre-trained datasets offered as a coding program library.

Raheman (2022) looks into the applicability of various natural language processing models for sentiment analysis of social media in the context of financial market prediction, with a focus on the cryptocurrency domain. The study aims to investigate the correlation between different sentiment metrics and the price fluctuations observed in Bitcoin. To achieve this, the paper explores multiple approaches for calculating sentiment metrics from textual data, revealing that many of these methods are not particularly accurate for the prediction task at hand. Through a comprehensive evaluation, the researchers identify a single model (Aigents) that outperforms more than 20 other publicly available models in terms of its predictive capabilities. The study analysis also showed a relative similarity in most models, such as Python libraries that provides a simplified and intuitive interface for performing various natural language processes (NLP), including text blob, google NLP, and alfin models.

D'Amato (2022) uses both the machine learning (ML) model technique and the deep learning model to examine the public sentiment effect on cryptocurrency. The ML model is trained to learn patterns and futures from labeled datasets then each text is associated with a sentiment label or also called polarity(positive, negative, or neutral). The deep learning model on the other hand, as in S. Rhaman's study, excels at capturing local patterns and is often employed for sentiment analysis at the sentence or phrase level.

2.4. Relationship between global crisis and financial performance

In his paper, "Could the global financial crisis improve the Performance of the G7 stock markets?" J.P Vieitoaand investigates whether the global financial crisis of 2008 had a positive impact on the performance of the stock markets of the G7 countries (Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States). The study uses data from 1990 to 2016 and employs various econometrics and techniques such as CAPAM, and MV analysis (mean-variance) to analyze the impact of the crisis on the markets.

The results suggest that the crisis had a significant positive impact on the performance of the stock markets in the G7 countries. This finding is attributed to several factors, including the implementation of policy measures by central banks and governments to mitigate the effects of the crisis, such as quantitative easing and fiscal stimulus. (Vieito, J. P 2016)

Overall, the study suggests that the global financial crisis had a positive impact on the performance of the G7 stock markets, although the long-term effects of the crisis on the economies of these countries are still uncertain.

In a McKinsey article titled “The impact of COVID-19 on capital markets, one year in” C. Bradly and P. Stumpner provide insights into how the pandemic has impacted various sectors of the capital markets, such as equity, debt, and alternative investments.

The article also discusses how market participants have responded to the pandemic and overall concludes that the pandemic has acted as an accelerator to pre-existing trends in the capital market such as the rise of digitalization and the increasing focus on sustainability, and has highlighted the importance of resilience in the face of unexpected shocks. The author also suggests that firms should prioritize building resilience into their business models and prepare for similar upcoming scenarios(C. Bradley P. Stumpner – 2021).

This may suggest that the Covid-19 crisis started on December 2019 (World Health Care Organization Archive) and could also have been an important factor in today’s cryptocurrency performance.

2.5. Concluding literature

Based on the literature reviewed, several key insights can be drawn. Firstly, cryptocurrency and crypto-resilience are areas of interest, with pioneers like David Chaum and Satoshi Nakamoto contributing to the development of digital currencies and blockchain technology. Factors such as popularity, defense mechanisms, P2P network characteristics, socio-political events, social media attention, potential diversification, and reduced liquidity risk contribute to the resilience of cryptocurrencies.

Sentiment analysis plays a crucial role in understanding market behavior and decision-making across various domains, including cryptocurrency. It involves the automated mining and analysis of public opinions expressed through various channels such as blogs, comments, reviews, and tweets. Sentiment analysis has been found to provide valuable insights into customer behavior, market trends, and firm performance, and its application in the cryptocurrency domain shows potential for predicting price fluctuations.

Different models and techniques have been employed in sentiment analysis, including machine learning (ML) and deep learning (DL) models. These models enable the analysis of sentiment patterns and can operate at both the document and sentence levels.

The relationship between global crises in some cases shows a positive outcome, such as the 2008 crisis had a positive impact on the performance of G7 stock markets due to various factors, including policy measures implemented by central banks and governments. The COVID-19 pandemic has acted as an accelerator for pre-existing trends in capital markets and highlighted the importance of resilience in the face of unexpected shocks.

However, it is important to note that sentiment analysis has its limitations, including challenges related to negation, domain-specific analysis, and the current knowledge of the subject. Despite these limitations, sentiment analysis remains a valuable tool for understanding the impact of public sentiment on cryptocurrency market dynamics and predicting price movements.

Overall, the literature supports the notion that sentiment analysis, when combined with appropriate models and techniques, can provide insights into the relationship between sentiment and cryptocurrency values, contributing to a better understanding of market behavior and decision-making in the financial domain.

2.6. Limitations

As was mentioned in the literature sentiment analysis is susceptible to errors caused by text which contains ironic statements, non-text data, and negation.

The data-gathering process proved to be much more work-intensive than expected and the accuracy of data is as well susceptible to some degree of error. As of begging of 2023 Twitter announce changes to its API developer platform in regards to price and accessibility which determined to pursue data collection through third-party data providers.

3. Methodology

This methodology section is dedicated to explaining the proposed analysis method for the cryptocurrency correlation between the sentiment of the public.

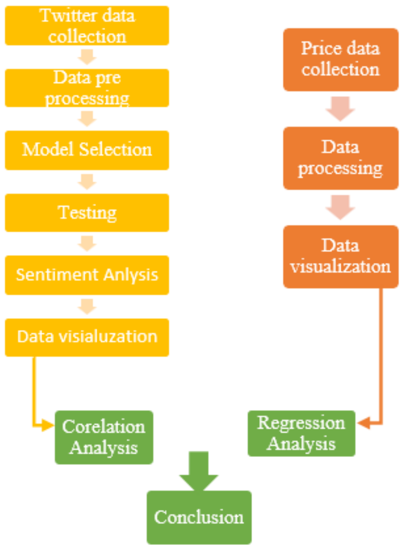

Figure 2 Analysis Action Plan

The research will be conducted in four stages. Firstly we have the literature review where we evaluate the theoretical research already conducted on cryptocurrency and how the sentiment analysis can be used to fit our research.

Secondly, we have the planning and data gathering phase where we gather tweeter data sets for the periods before and during the corona pandemic, acquire the necessary software to process the data(Python, R-studio), and select a model of data processing for sentiment analysis to fit the research.

Thirdly we have the analysis phase where we conduct the research as it was previously planned and gather the results. Finally based on our analysis we formulate a conclusion

3.1. Data collection

Tweets are collected by using the open-source website kaggle.com, which contains datasets libraries on tweets regarding cryptocurrencies.[1]

The way data was collected is by looking up the name of the cryptocurrency and/or the hashtag associated with it. Since the analysis focuses to explain the overall cryptocurrency market, the search word setting is expanded to encompass as many cryptocurrencies as possible. Examples of search words: btc, crypto,1inch,eth, bitcoin, bitcoin news, ada, cardano, crypto, cryptocurrency.

Sentiment will be plotted against an cryptocurrency index, namely the CCI30 index which is designed to measure the overall performance of the crypto market on a short and long-term basis. The index tracks 30 largest cryptocurrencies by market capitalization while excluding stablecoins. The index contains daily data starting with the year 2015 and it is within our range and scope of the research.

The motivation for selecting this dataset is fueled by the historical data availability and ease of access, other data set indexes were considered but they offered a limited about of historical price data.

3.2.Sentiment model selection

In light of Raheman (2022) research on sentiment analysis models for text compiling and interpretation, the decision was made to select textBlob built on Natural Language Processing which provides relatively good accuracy and it is one of the most accessible reasons for which it is widely used as well.

The text analysis comes as a library that needs to be added and run in a Python coding program.

3.3.Regression and correlation analysis

After the sentiment analysis is completed and data has been sorted out correlation and regression analysis is proposed as a means to identify the interaction between variables given by sentiment score and crypto index ccI30.

Correlation is that when two variables are correlated, changes in one variable tend to correspond with changes in the other variable in a specific direction. Knowing this relationship is helpful because it allows us to predict the value of one variable based on the value of the other variable. (J. Frost 2019) Through this method, we intend to confirm the correlation between the sentiment of the population and the price movement of the cryptocurrency.

However, correlation analysis only describes the strength and the direction of the relationship, we need regression analysis to define the relation itself.

There are multiple regression methods but we are interested in the linear regression model because we are looking to understand the degree of influence on cryptocurrency prices caused by the sentiment of tweets between the 2 periods assessed.

The linear regression equation which will be used for a multi-linear regression model with 3 independent variables (X) and one dependent variable (Y) is:

$Y = \beta0 + \beta1X + \beta2X\ 2+\ldots+ \beta nXn + \varepsilon$

where:

Y is the predicted value of the dependent variable

$\beta0$ is the intercept (the value of Y when X = 0)

$\beta1$ is the slope (the change in Y for every one-unit increase in X)

X is the independent variable

ε is the error term (the difference between the predicted value of Y and the actual value of Y)

3.4. Variable selection

Table 1. Variable table

Variable	Role	Obtained by
Volatility	Dependent variable	Excel calculation on logarithmic return on a sample period
Return	Dependent v.	(current price/previous price-1)
Positive score	Independent v.	Given by the sentiment analysis
Negative score	Independent v.	Given by the sentiment analysis
Volume	Control v.	Given by the index cci30

[1]The website is a Google subsidiary and an online community for data scientists and engineers. The platform is dedicated to offering users free access to datasets, to solve data science challenges. (https://www.kaggle.com/datasets/leoth9/crypto-tweets & https://www.kaggle.com/datasets/ilariamazzoli/3-million-tweets-cryptocurrencies-btc-eth-bnb)

4. Analysis

As described in the chapter previous chapter, tweeter data sets were downloaded from Kaggle.com. To ensure the accuracy of the analysis a lot of data entries are required to provide a reliable answer, therefore, after collecting the data, a number of 279.043 tweets were filtered and collected for the period of 2020 to 2022. For 2017-2019, 547.351 tweets were collected concluding a total of 826.394 internet tweets.

Since sentiment will be used to determine the impact on the cryptocurrency market, the crypto index CCI30 which was designed to measure the overall performance of the market. The index provided a total of 2097 data entries from January 1st, 2017 till December 31st, 2022. Within the CCI30 index data set the following variable was provided: open price, high and low of the day, and the closing price and volume.

Because the number of tweets is far bigger than the data provided by the ccI30 index an average of the negative and positive polarity was determined and separated into two variables, positive sentiment score and negative sentiment score, both indicating the average positive and negative score for the corresponding date of the index price.

Further on, the return rate was obtained by dividing the current closing price by the previous day's closing price and subtracting 1 (current day/previous day-1).

Daily volatility is calculated as follows:

$$\sqrt{\frac{1}{2}h_{i,t} - l_{i,t})^2 - (2log2 - 1)c_{i,t}^2}$$

where ci,t = log(closei,t) − log(openi,t), li,t = log(lowi,t) − log(openi,t) and ht = log(highi,t) − log(open i,t).

(*Haykir, O., Yagli, I. 2022)*

4.1. Sentiment analysis

Sentiment analysis was performed using Python coding language and the training model used was NLTK (Natural Language Toolkit) library which is commonly used in numerous sentiment analysis research.

Before running the analysis a series of text cleanings were performed on the tweets to exclude interfering with the analysis (@, #, images, pictures, links).

Table 2 Sentiment Score Descriptive Statistics

Mean	0.25
Standard Error	0.00
Median	0.25
Mode	0.00
Standard Deviation	0.44
Sample Variance	0.19
Kurtosis	-0.52
Skewness	-0.32
Range	2.00
Minimum	-1.00
Maximum	1.00
Sum	203823.97
Count	826394.00

Further on, the program simply picks up the text content of each tweet, passes it through the machine learning model, and attributes a negative, positive, and neutral score, which is stored in the same Excel file. Descriptive statistics provide an overview of the compounded sentiment score of the overall dataset.

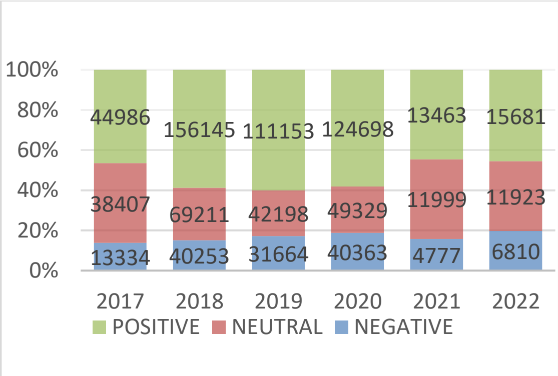

Figure 3 Sentiment Polarity over time

Based on the sentiment score results a few observations are made.

As can be seen in Figure 3. , the number of tweets after the first year of the pandemic started to lower exponentially, and the population sentiment takes a much more neutral attitude toward cryptocurrency.

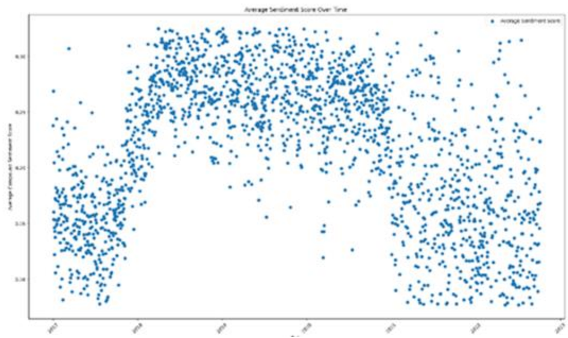

Figure 4 Sentiment over time 2017-2022

Additionally, while plotting the overall results of the sentiment analysis data points it can be observed in Figure 4. that during the last two years of the pandemic (2021-2022) population’s sentiment was very dispersed while from 2017 through 2019 it can be observed a positive trend.

4.2. Correlation and regression analysis

Volatility analysis

The data for the period pre-corona (2017-2020) in regards to volatility includes information on 1074 observations. The correlation matrix 1 (Matrix 1->Appendix) indicates a negative correlation between volume and volatility (-0.233) and a positive correlation between positive score and volatility (0.044). Multicollinearity statistics show that the variables volume, negative score, and positive score have tolerance values of 0.684, 0.919, and 0.665, respectively, and variance inflation factor (VIF) values of 1.463, 1.088, and 1.504, respectively.

The regression analysis for the variable volatility shows that the model has a coefficient of determination (R^2) of 0.067, indicating a low level of fit. The adjusted R^2 is 0.064. Meaning that our model is able to explain only 7.% of the volatility using a Negative score, Positive score, and Volume variable.

Table 3 Goodness of fit statistics (Volatility): Regression (pre-Corona)

Observations	1074
Sum of weights	1074
DF	1070
R^2	0.067
Adjusted R^2	0.064
MSE	0.000
RMSE	0.017

The standardized coefficients (Matrix 2 -> Appendix)show that volume has a negative standardized coefficient of -0.307, indicating a negative impact on volatility. The negative score has a small and insignificant effect (0.015), while the positive score has a positive effect (0.139) on volatility

Additionally, both volume and positive score have a p-value lower than alpha(0.05) which indicates a significant statistical relation between them and volatility.

Data during the Covid-19 crisis (2020-2022) in regards to volatility includes information on 1001 observations. The correlation matrix (Matrix 3)shows a positive correlation of 0.294 between volume and volatility, an insignificant positive correlation between volume and negative score, and a low negative correlation of - 0.194 between positive score and volatility. Multicollinearity statistics indicate that all variables have high tolerance values and low variance inflation factor (VIF) values, suggesting low multicollinearity.

The regression analysis for the volatility variable reveals a coefficient of determination (R^2) of 0.090, indicating a relatively weak fit. The adjusted R^2 is 0.088.

Table 4 Goodness of fit statistics (Volatility): Regression (during Corona)

Observations	1001
Sum of weights	1001
DF	997
R^2	0.090
Adjusted R^2	0.088
MSE	0.000
RMSE	0.018

The standardized coefficients(Matrix 4) show that volume has a positive standardized coefficient of 0.284, indicating a positive impact on volatility. The negative score has a small and insignificant effect (-0.009), while the positive score has a negative effect (-0.064) on volatility.

Similarly, both volume and positive score have a p-value lower than alpha(0.05) which indicates a significant statistical relation between them and volatility.

Return analysis

Looking at the analysis on return for the period before Corona, observing 1094 date entries. The correlation matrix (Matrix 5)shows a low and insignificant level of correlation between sentiment polarity and return and a weak negative correlation of -0.052 between volume and return. Multicollinearity statistics indicate that all variables have high tolerance values and low variance inflation factor (VIF) values, suggesting low multicollinearity.

The regression analysis for the return variable indicates a coefficient of determination (R^2) of 0.007, suggesting a very weak fit. The adjusted R^2 is 0.005.

Table 5 Goodness of fit statistics (Return):Regression (pre-Corona)

Observations	1094
Sum of weights	1094
DF	1090
R^2	0.007
Adjusted R^2	0.005
MSE	0.002
RMSE	0.047

The standardized coefficients (Matrix 6) reveal that volume has a small negative standardized coefficient (-0.005), indicating a weak impact on return. The negative score has a small positive effect (0.014), while the positive score has a stronger negative effect (-0.078) on return. However, these effects are not statistically significant.

During Corona crisis, data include 1002 observations. The correlation matrix (Matrix 7) shows a similar insignificant level of correlation just as before the pandemic of return and sentiment polarity, and a weak negative correlation of -0.065 between volume and return. Multicollinearity statistics indicate that all variables have high tolerance values and low variance inflation factor (VIF) values, suggesting low multicollinearity.

The regression analysis for the return variable indicates a coefficient of determination (R^2) of 0.008, suggesting a very weak fit. The adjusted R^2 is 0.005.

Table 6 Goodness of fit statistics (Return):Regression (during Corona)

Observations	1002
Sum of weights	1002
DF	998
R^2	0.008
Adjusted R^2	0.005
MSE	0.002
RMSE	0.045

The standardized coefficients(Matrix 8) reveal that volume has a small negative standardized coefficient (-0.063), indicating a weak impact on return. Both the negative score and positive score have small positive effects (0.054 and 0.055, respectively) on return, although the statistical significance of these effects is marginal.

5. Conclusion and discussion

Overall, the analysis of the data supports our alternative hypothesis that there is a relationship between sentiment polarity (negative-positive) and the variables of interest (volatility and return) in the corona crisis periods.

In the pre-corona period, the analysis reveals a positive impact of the positive score on volatility. The coefficient of determination (R^2) suggests a low level of fit, indicating that the variables included in the model explain only 6.7% of the volatility. Similarly, the analysis of return shows a weak impact of positive score on return, with the model explaining only 0.7% of the return.

During the corona crisis, the analysis of volatility shows a positive impact on the volume and a negative impact of the positive score on volatility. The coefficient of determination (R^2) suggests a relatively weak fit, with the variables explaining 9% of the volatility. The on-return rate indicates a weak impact of volume and a positive score on return, with the model explaining only 0.8% of the return.

In both periods, volume and positive scores have p-values lower than the significance level of 0.05, indicating a statistically significant relationship with volatility. However, the effects of the variables on return are not statistically significant.

Overall, it is worth mentioning that by looking at **Figure.4,** sentiment toward cryptocurrency changed throughout the periods, and it became more dispersed. However, the results suggest that positive score has a low influence on volatility, and a weak impact on returns having a very low explanatory power for crypto advancement during Corona mentioned in G. Iyer 2021, C. Bradley 2021, R. Van Hoek 2020.

It is important to note that the models used in this analysis explain only a small portion of the variability in volatility and return, indicating that other factors not included in the analysis may also play a significant role. Although the analysis aligns with H.Chaudhari-2020 which says that sentiment does not have an immediate strong effect on the market but rather is an indicator of the direction.

Although a relationship between a positive sentiment score and the volatility of the market was confirmed, given the small p-value, it does not explain alone the price fluctuation in the market.

Further research and analysis are needed to gain a deeper understanding of the factors influencing volatility and return in the cryptocurrency market.

6. References

1. *A. A. Shehhi, M. Oudah and Z. Aung, "Investigating factors behind choosing a cryptocurrency," 2014 IEEE International Conference on Industrial Engineering and Engineering Management, Selangor, Malaysia, 2014, pp. 1443-1447, doi: 10.1109/IEEM.2014.7058877.(goot for later)*
2. *Harshal Chaudhari, Martin Crane, 2020,Cross-correlation dynamics and community structures of cryptocurrencies,Journal of Computational Science,Volume 44,https://doi.org/10.1016/j.jocs.2020.101130.*
3. *Conlon, T., & McGee, R. (2020). Safe haven or risky hazard? Bitcoin during the Covid-19 bear market. Finance Research Letters, 35, 101607. https://doi.org/10.1016/j.frl.2020.101607*
4. *Teti, E., Dallocchio, M., & Aniasi, A. (2019). The relationship between twitter and stock prices. Evidence from the US technology industry. Technological Forecasting and Social Change, 149, 119747. https://doi.org/10.1016/j.techfore.2019.119747*
5. *Remko Van Hoek and Mary Lacity, April 27, 2020 - How the Pandemic Is Pushing Blockchain Forward –“ https://hbr.org/2020/04/how-the-pandemic-is-pushing-blockchain-forward*
6. *Sayim, M., My, N. (2022). An Analysis Of the U.S. Individual Investor Sentiment*

Influence On Cryptocurrency Returns And Volatility. Ann. Finan. Econ., 02(18). https://doi.org/10.1142/s2010495222420015

7. *Almansour, B. (2020). Cryptocurrency Market: Behavioral Finance Perspective. JAFEB, 12(7), 159-168. https://doi.org/10.13106/jafeb.2020.vol7.no12.159*
8. *Wang, Z., Huang, Z., He, R., Feng, Y. (2022). Cryptocurrency and The Herd Behavior.. https://doi.org/10.3233/atde220018*
9. *Haykir, O., Yagli, I. (2022). Speculative Bubbles and Herding In Cryptocurrencies. Financ Innov, 1(8). https://doi.org/10.1186/s40854-022-00383-0*
10. *NORC- University of Chicago 2021 - https://www.norc.org/research/library/more-than-one-in-ten-americans-surveyed-invest-in-cryptocurrenci.html*
11. *Zainudin, Z., Zaki, Z., Hadi, A., Hussain, H., Kantakji, M. (2019). Investor Sentiment and Firm Financial Performance Of Malaysian Ipo Firms: Pre And Post Financial Crisis. IJFR, 5(10), 450. https://doi.org/10.5430/ijfr.v10n5p450*
12. *Hong, W. (2023). Impact Of Online Public Opinion Regarding the Japanese Nuclear Wastewater Incident On Stock Market Based On The Sor Model. MBE, 5(20), 9305-9326. https://doi.org/10.3934/mbe.2023408*
13. *R.Khalfaoui & R., Gozgor, G., & Goodell, J. W. (2023, March). Impact of Russia-Ukraine war attention on cryptocurrency: Evidence from quantile dependence analysis.*
14. *Zhou, C., Xing, L., Guo, J., Liu, Q. (2022). Bitcoin Selfish Mining Modeling and Dependability Analysis. Int J Math, Eng, Manag Sci, 1(7), 16-27. https://doi.org/10.33889/ijmems.2022.7.1.002*
15. *Delgado-Segura, S., Pérez-Solà, C., Herrera-Joancomartí, J., Navarro-Arribas, G., Borrell, J. (2018). Cryptocurrency Networks: a New P2p Paradigm. Mobile Information Systems, (2018), 1-16. https://doi.org/10.1155/2018/2159082*
16. *Ghabri, Y., Guesmi, K., Zantour, A. (2021). Bitcoin and Liquidity Risk Diversification. Finance Research Letters, (40), 101679. https://doi.org/10.1016/j.frl.2020.101679*
17. *Kyaw, N., Tepsongkroh, P., Thongkamkaew, C., Sasha, F. (2023). Business Intelligent Framework Using Sentiment Analysis For Smart Digital Marketing In the E-commerce Era. Asia Social Issues, 3(16), e252965. https://doi.org/10.48048/asi.2023.252965*
18. *A. Raheman 2022 -Social Media Sentiment Analysis for Cryptocurrency Market Prediction" https://doi.org/10.48550/arXiv.2204.10185*
19. *Jim Frost 2019 Regression Analysis an Intuitive guide for using and Interpreting Linear Models page 14 & 27*
20. *https://cci30.com/ (crypto index dataset)*
21. *Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Accessed on March 2023.and retrieved from https://bitcoin.org/bitcoin.pdf*
22. *Alghamdi, S., Alqethami, S., Alsubait, T., & Alhakami, H. (2022). Cryptocurrency Price Prediction using Forecasting and Sentiment Analysis. International Journal of Advanced Computer Science and Applications, 13(10). https://doi.org/10.14569/ijacsa.2022.01310105*
23. *D'Amato, V., Piscopo, G., & Levantesi, S. (2022). Deep learning in predicting cryptocurrency volatility. Physica D: Nonlinear Phenomena, 596, 127158. https://doi.org/10.1016/j.physa.2022.127158*
24. *S. Rahman, J. N. Hemel, S. J. A. Anta, and H. Al Muhee, "Sentiment analysis using r: an approach to correlate bitcoin price fluctuations with change in user sentiments," Ph.D. dissertation, BRAC University, 2018.*
25. *Liu, B. (2011). Web Data Mining. Springer EBooks. https://doi.org/10.1007/978-3-642-19460-3*
26. *G. Vinodhini & RM. Chandrasekaran (2012) Sentiment analysis and Opinion mining: A Survey - International Journal of Advanced Research in Computer Science and Software Engineering Available online at: www.ijarcsse.com*
27. *Hussein, D. G. (2016). A survey on sentiment analysis challenges. Journal of King Saud University: Engineering Sciences, 30(4), 330–338. https://doi.org/10.1016/j.jksues.2016.04.002*
28. *Haykir, O., Yagli, I. (2022). Speculative Bubbles and Herding In Cryptocurrencies. Financ Innov, 1(8). https://doi.org/10.1186/s40854-022-00383-0*

7.Appendix

1. Volatility analysis

Matrix 1 Correlation :pre-corona

	Volume	Negative score	Positive score	Volatility
Volume	**1**	-0.221	0.558	-0.233
Negative score	-0.221	**1**	-0.273	0.044
Positive score	0.558	-0.273	**1**	-0.037
Volatility	-0.233	0.044	-0.037	**1**

Matrix 2 Standardized (Volatility):pre-corona

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)	p-values signification codes
Volume	-0.307	0.036	-8.589	**<0.0001**	-0.377	-0.237	***
Negative score	0.015	0.031	0.471	0.637	-0.046	0.075	***
Positive score	0.139	0.036	3.827	**0.000**	0.068	0.210	°

Matrix 3 Correlation: during Corona

	Volume	Negative score	Positive score	Volatility
Volume	**1**	0.144	-0.186	0.294
Negative score	0.144	**1**	-0.303	0.051
Positive score	-0.186	-0.303	**1**	-0.114
Volatility	0.294	0.051	-0.114	**1**

Matrix 4 Standardized coefficients (Volatility): during Corona

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)	p-values signification codes
Volume	0.284	0.031	9.193	**<0.0001**	0.223	0.344	***
Negative score	-0.009	0.032	-0.295	0.768	-0.072	0.053	***
Positive score	-0.064	0.032	-1.985	**0.047**	-0.127	-0.001	°

2.Return analysis

Matrix 5 Correlation: pre-Corona

	Volume	Negative score	Positive score	return
Volume	**1**	-0.213	0.564	-0.052
Negative score	-0.213	**1**	-0.264	0.036
Positive score	0.564	-0.264	**1**	-0.085
return	-0.052	0.036	-0.085	**1**

Matrix 6 Standardized coefficients (Return):pre-corona

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)	p-values signification codes
Volume	-0.005	0.037	-0.145	0.884	-0.077	0.067	*
Negative score	0.014	0.031	0.460	0.646	-0.047	0.076	°
Positive score	-0.078	0.037	-2.098	**0.036**	-0.151	-0.005	°

Matrix 7 Correlation: during Corona

	Volume	Negative score	Positive score	return
Volume	**1**	0.145	-0.186	-0.065
Negative score	0.145	**1**	-0.303	0.028
Positive score	-0.186	-0.303	**1**	0.051
return	-0.065	0.028	0.051	**1**

Matrix 8 Standardized coefficients (Return): during Corona

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)	p-values signification codes
Volume	-0.063	0.032	-1.953	0.051	-0.126	0.000	°
Negative score	0.054	0.033	1.631	0.103	-0.011	0.119	.
Positive score	0.055	0.033	1.653	0.099	-0.010	0.121	°