Deciphering Crypto Trends: Google Trends Analysis and Predictive Modelling.


Photo by Anna Tarazevich

Cyril Mawutor Agbenyenu
(Analyst, Mawutor Utility Company)
April 2024

Abstract

This report explores the relationship between Google Trends data and cryptocurrency price trends, focusing on Bitcoin. We found a significant correlation between search interest and Bitcoin's price, suggesting that heightened public interest often precedes price movements. A machine learning approach using Random Forest yielded the most accurate predictions for Bitcoin's search interest, indicating that this algorithm can effectively forecast short-term trends. The ideal lag time for predicting Bitcoin's search interest was zero, reinforcing the notion that current public interest influences market behavior. Additionally, Ethereum's weekly volume data provided insights into broader market trends, showing fluctuations that could reflect general cryptocurrency market activity. These findings offer valuable insights for traders and investors and highlight the potential for machine learning models to enhance market analysis. All code and datasets used in this study are available on GitHub, providing a resource for further exploration and validation of the results.

Table of Contents

Introduction

Trend analysis involves the examination of data over time to identify patterns, tendencies, or fluctuations. It helps uncover underlying trends, seasonality, and anomalies within datasets. Valuable insights extracted from trend analysis include understanding historical patterns, forecasting future trends, and identifying potential opportunities or risks.

By analyzing trends, businesses can make informed decisions, optimize resource allocation, and anticipate market changes. Predictive modeling utilizes trends identified through analysis to forecast future outcomes, allowing businesses to anticipate demand, plan inventory, and develop strategies for growth or risk mitigation.

Cryptocurrencies have emerged as a significant asset class, characterized by their volatile nature and susceptibility to market sentiment. Understanding the factors that influence their value fluctuation is critical for investors and analysts alike. In this report, we dig deep into the world of cryptocurrency markets, using Google Trends data alongside analytical techniques to dissect market trends, predict future movements, and contextualize current behaviors within current historical cycles.

We kickstart our analysis with exploratory data analysis of Google Trends data to discern patterns, seasonality, and anomalies in cryptocurrency search volumes. By processing, analyzing, and visualizing this type of data, we can gain insights into the popularity and public interest in various cryptocurrencies over time.

To enrich our analysis, we will integrate additional datasets such as historical cryptocurrency Total Value Locked (TVL), the Greed and Fear Index, and the United States and world inflation data. These additional data not only enrich the analysis but also provide diverse perspectives on how the cryptocurrency markets react to macroeconomic events.

Further down the road, we shall employ sophisticated machine-learning algorithms to uncover correlations between Google Trends data, external factors, and cryptocurrency prices. Through feature engineering, model training, and validation, we will develop predictive models capable of forecasting future trends in cryptocurrency markets.

Predictive models would provide insights into future movements of cryptocurrency prices. By leveraging time-series analysis, ensemble methods, and deep learning techniques, we aim to generate accurate results that can empower stakeholders to make informed decisions in dynamic market environments.

Data Overview

The provided datasets are grouped into “trends” and “prices”. The "trends" dataset comprises Google Trends data showcasing the web search interest for 20 prominent cryptocurrencies. These include well-known assets such as Bitcoin, Ethereum, and XRP, alongside emerging tokens like Solana and Polkadot. Each cryptocurrency's search interest is normalized on a scale from 0 to 100, reflecting the relative popularity across the specified timeframe.

In contrast, the "prices" dataset provides comprehensive information on the prices and trading volume of the same 20 cryptocurrencies. This dataset allows for the analysis of price movements and liquidity dynamics, essential for understanding market behaviors. By juxtaposing search interest with actual market performance, analysts can discern correlations between public sentiment and cryptocurrency valuations.

Both datasets are integral to our analysis, offering complementary perspectives on cryptocurrency market dynamics. While Google Trends data provides insights into public interest and sentiment, the prices dataset facilitates quantitative analysis of market movements and trading activity. Integrating these datasets enables a holistic understanding of the factors influencing cryptocurrency prices and trends.

provided_datasets_snapshot

```
(env) mawutor@penguin:~/dev/crypto-trends$ ls
LICENSE  prices  README.md  trends
(env) mawutor@penguin:~/dev/crypto-trends$ ls trends/
 bitcoin.csv   chainlink.csv   fetch.ai.csv   litecoin.csv        'ocean protocol.csv'
singularitynet.csv   uniswap.csv
 bnb.csv       dogecoin.csv    filecoin.csv   monero.csv           pancakeswap.csv
solana.csv           XRP.csv
 cardano.csv   ethereum.csv    kucoin.csv    'oasis network.csv'   polkadot.csv
tezos.csv
(env) mawutor@penguin:~/dev/crypto-trends$ ls prices/
ADA-USD.csv   BNB-USD.csv   CAKE-USD.csv   DOT-USD.csv   FET-USD.csv   KCS-USD.csv    LTC-USD.csv
ROSE-USD.csv  UNI-USD.csv   XRP-USD.csv
AGIX-USD.csv  BTC-USD.csv   DOGE-USD.csv   ETH-USD.csv   FIL-USD.csv   LINK-USD.csv   OCEAN-USD.csv
SOL-USD.csv   XMR-USD.csv   XTZ-USD.csv
(env) mawutor@penguin:~/dev/crypto-trends$
```

The "trends" dataset is structured with a header line indicating the category and cryptocurrency being analyzed ("Category: All categories, Week,BNB: (Worldwide)"). Each subsequent line represents a week's data, starting with the date and followed by the normalized search interest score for the specified cryptocurrency. For instance, the entry "2019-04-07,56" indicates that during the week of April 7th, 2019, the search interest score for BNB (Binance Coin) was 56.

On the other hand, the "prices" dataset follows a different structure, with each line representing a daily snapshot of cryptocurrency price data. The header line specifies the columns: "Date, Open, High, Low, Close, Adj Close, Volume." Subsequent lines provide the details for each day, including the date, opening price, highest price, lowest price, closing price, adjusted closing price, and trading volume.

For example, the entry "2017-11-09,0.025160,0.035060,0.025006,0.032053,0.032053,18716200" indicates that on November 9th, 2017, the opening price of the cryptocurrency was 0.025160, the highest price reached during the day was 0.035060, the lowest price was 0.025006, the closing price was 0.032053, the adjusted closing price was also 0.032053, and the trading volume was 18,716,200 units.

These distinct formats provide essential information for analyzing trends in search interest and cryptocurrency prices over time, allowing for comprehensive exploration and understanding of cryptocurrency market dynamics.

Data Preprocessing

Preprocessing is a critical step in any data analysis project, as it ensures that the data is in a suitable format for further exploration and modeling. In the case of the provided datasets, which consist of trends and price data for multiple cryptocurrencies, the primary focus of preprocessing is consolidation.

Given that each dataset is spread across several individual CSV files, this fragmentation can complicate and slow down the analysis process.
Consolidation is essential because it brings together all relevant data into a single, unified structure. This is particularly important when working with large datasets with multiple files, as it reduces redundancy and ensures consistency across different sources.

By consolidating the datasets, we can effectively streamline the analysis, enabling easier cross-referencing, efficient data manipulation, and more comprehensive insights into the trends and price fluctuations of the various cryptocurrencies. Ultimately, the consolidation allows analysts to work with a clean, organized dataset, leading to more efficient and effective data analysis.

Preprocessing for Trends Data

To consolidate the Google Trends data into a single dataset, the first step was to gather individual CSV files for various cryptocurrencies. The extraction of the cryptocurrency name was achieved by parsing the filename, removing extensions, and ensuring consistent formatting. This step was critical to correctly label the data with the corresponding cryptocurrency. After the data was read and the cryptocurrency name extracted, the script set consistent column names, namely 'week', 'cryptocurrency', and 'search_interest'.

trends.py

```
...

# Iterate over each cryptocurrency file
for file in cryptocurrency_files:
    # Read the data from the current cryptocurrency file, skipping the first row
    data = pd.read_csv(file, skiprows=1)

    # Extract the cryptocurrency name from the filename
    cryptocurrency = os.path.splitext(file)[0].split('-')[0]
    cryptocurrency = cryptocurrency.replace('_', ' ').title()

    # Ensure consistent column naming
    data.columns = ['week', 'search_interest']

    # Add the cryptocurrency name to the DataFrame
    data['cryptocurrency'] = cryptocurrency

    # Append the data to the consolidated DataFrame with all required columns
    consolidated_data = consolidated_data.append(data[['week', 'cryptocurrency',
'search_interest']], ignore_index=True)

# Save the consolidated data to a new CSV file
consolidated_data.to_csv('trends.csv', index=False)
```

The consolidated trends data was then created by appending the processed data from each CSV file into a single DataFrame. This approach allowed for an organized view of trends across multiple cryptocurrencies. The final step was to save the consolidated DataFrame to 'trends.csv', providing a unified dataset that could be used for further analysis and predictive modeling.

Preprocessing for Prices Data

The pre-processing for the price data involved reading the daily price information from multiple CSV files, each representing a different cryptocurrency's price history. The first step was to identify all the relevant files and standardize the structure, which included columns for 'Date', 'Open', 'High', 'Low', 'Close', 'Adj Close', and 'Volume'. Additionally, a new column 'Cryptocurrency' was introduced to indicate the cryptocurrency to which each data point belonged.

prices.py

```python
...

# Iterate over each price CSV file
for file in price_files:
    # Read the data from the current CSV file
    data = pd.read_csv(file)
    # Extract the cryptocurrency name from the filename, removing '-USD'
    cryptocurrency = os.path.splitext(file)[0]
    cryptocurrency = cryptocurrency.split('-')[0]  # Remove the '-USD' part
    # Add a new column for the cryptocurrency name
    data['Cryptocurrency'] = cryptocurrency
    # Append the data to the consolidated DataFrame
    consolidated_price_data = consolidated_price_data.append(data, ignore_index=True)

# Save the consolidated price data to a new CSV file
consolidated_price_data.to_csv('prices.csv', index=False)
```

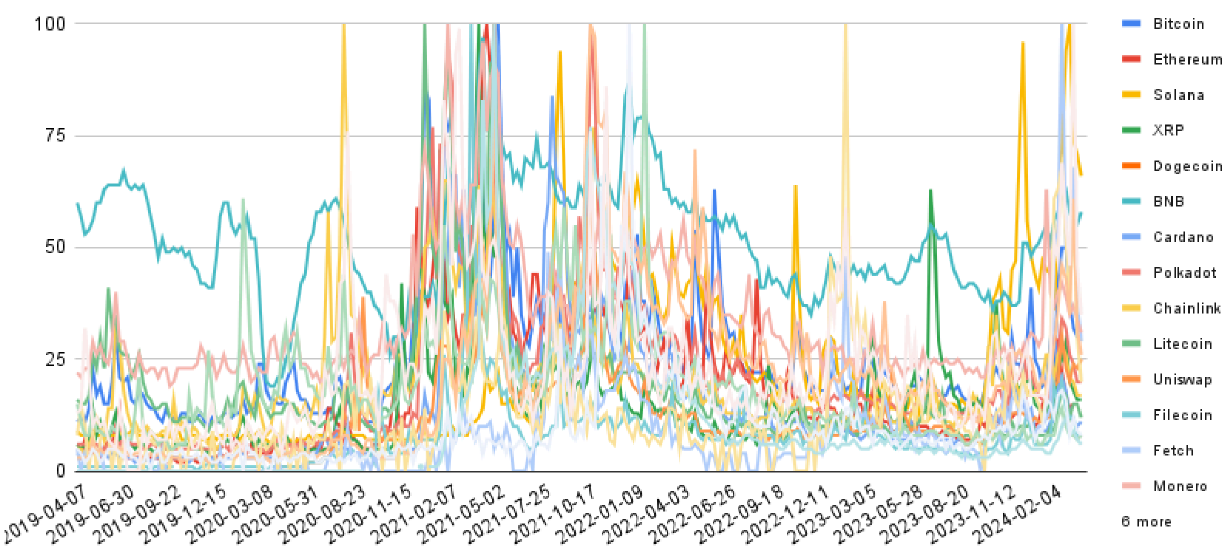
A key step in the pre-processing for prices data was ensuring the proper extraction of the cryptocurrency name. To achieve this, the script parsed the CSV filenames, removed any suffixes like "-USD", and then appended the clean cryptocurrency name to the DataFrame. The pre-processed data was then consolidated into a single CSV file named 'prices.csv', providing a comprehensive dataset with consistent naming conventions and column structures for further analysis.

Key Findings

Search Interest Trends

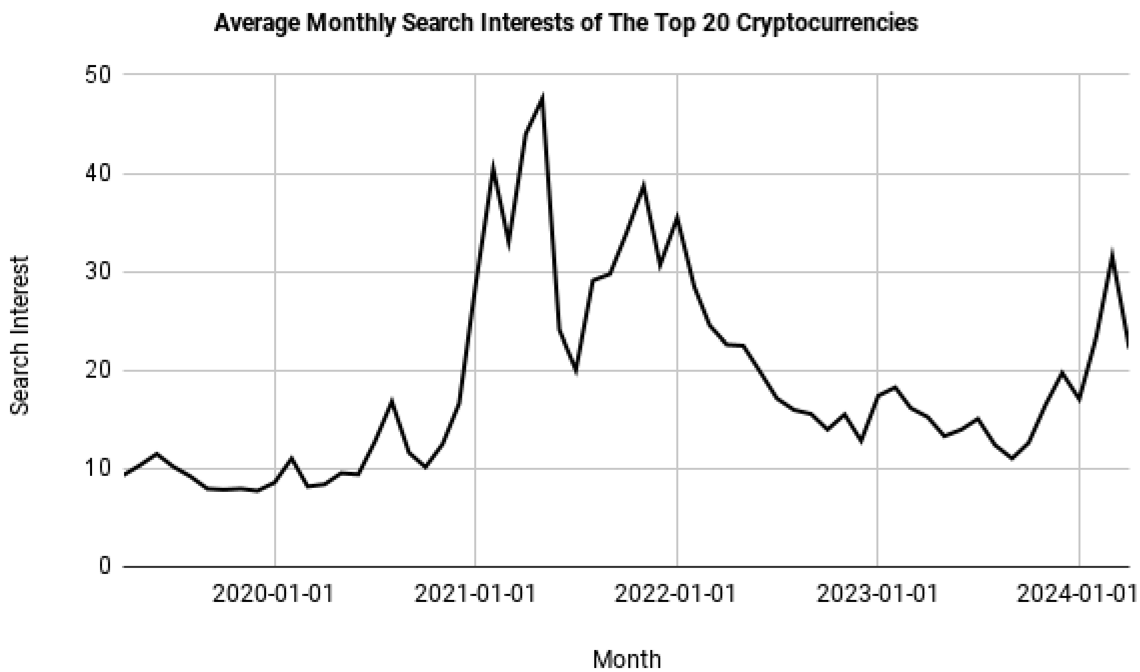
Monthly Average Search Interest Trends

The monthly average search interest for the top 20 cryptocurrencies displays clear fluctuations over time. Notably, there's a steady increase from the beginning of 2019, with peaks around May 2021 and April 2021, indicating a strong period of public interest and likely market activity. Following this peak, there is a noticeable decline throughout 2021 and 2022, potentially reflecting shifts in market trends, investor sentiment, or overall market volatility.



The highest search interest was observed in May 2021, with a value of 47.57, suggesting a significant spike in public attention toward cryptocurrencies during that period. This could be attributed to a combination of market events, media coverage, or prominent announcements within the crypto community. Post-2021, the average search interest stabilizes with minor variations, indicating a more consistent level of public engagement across 2022 and early 2023.

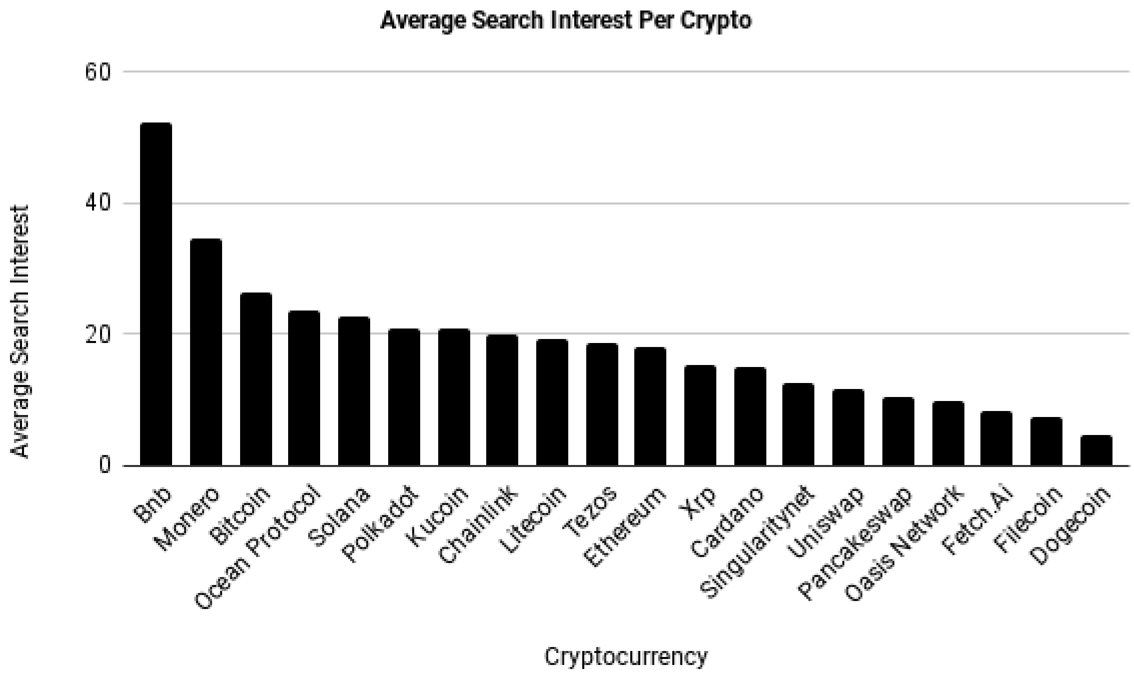
As we approach 2024, the search interest displays some recovery, with increasing values from early 2023 to early 2024. This uptick could signal a renewed interest in cryptocurrencies, possibly driven by broader market recovery, new crypto developments, or increased adoption in mainstream finance. The overall trend reflects the dynamic nature of the cryptocurrency market and the impact of external events on public sentiment and interest.



The average search interest for each cryptocurrency reveals distinct variations, with BNB (Binance Coin) leading the way with an average search interest of 52.24. This high value indicates significant public attention, likely driven by the popularity of the Binance platform and its growing ecosystem. Following BNB, other cryptocurrencies with notable search interest include Bitcoin at 26.29, Ocean Protocol at 23.7, and Monero at 34.58, suggesting sustained public interest in these well-known coins.

Average Search Interest by Cryptocurrency

The average search interest for other cryptocurrencies indicates varying levels of public attention. Ethereum, one of the most prominent cryptocurrencies, has an average search interest of 18.27, reflecting its significant role in decentralized finance (DeFi) and smart contract space. Similarly, Solana, known for its high-speed blockchain, demonstrates a relatively high average search interest of 22.88, indicating its growing presence in the market.

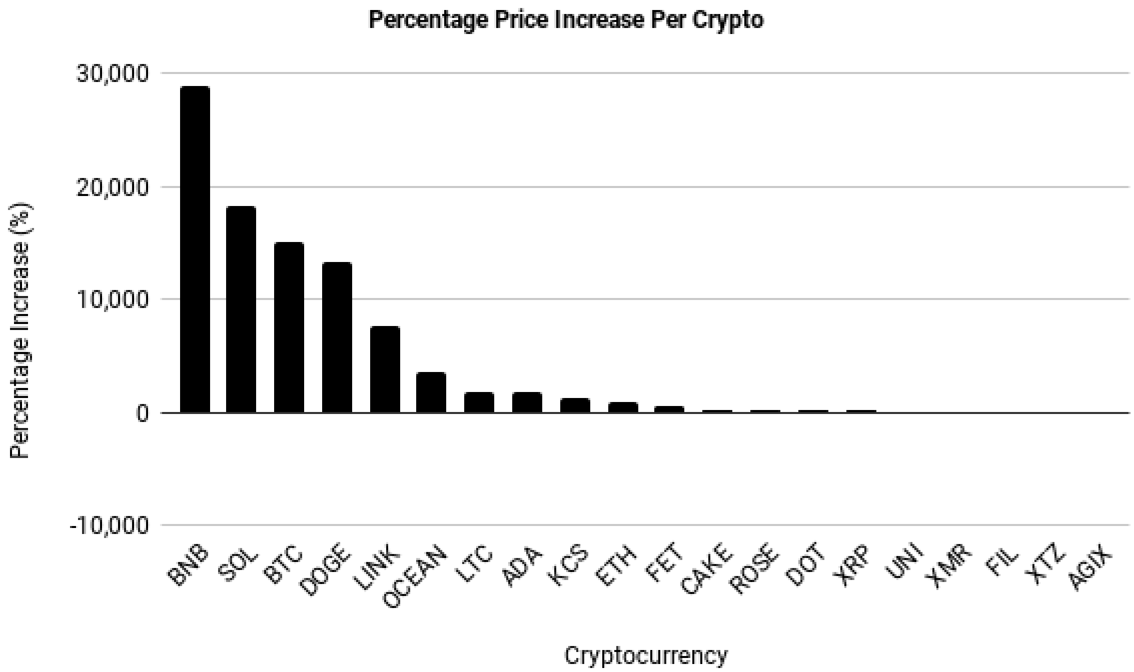


On the other end of the spectrum, cryptocurrencies like Dogecoin and Fetch.AI have lower average search interest, with values of 4.65 and 8.47, respectively. This suggests that while these cryptocurrencies have a dedicated following, they may not capture the same level of mainstream attention as some of the more established coins. The wide range of average search interests highlights the diverse nature of the cryptocurrency landscape and the varying factors driving public engagement with different coins.

Price Trends

Analysis of Percentage Price Increase

The percentage price increase data shows remarkable variation across different cryptocurrencies, reflecting the dynamic and volatile nature of the crypto market. BNB (Binance Coin) experienced the most significant percentage increase, with an astonishing 28,923.06% rise from the first to the last week close. This extraordinary growth may be attributed to Binance's expanding ecosystem and the widespread adoption of BNB for transactions and staking within the platform.

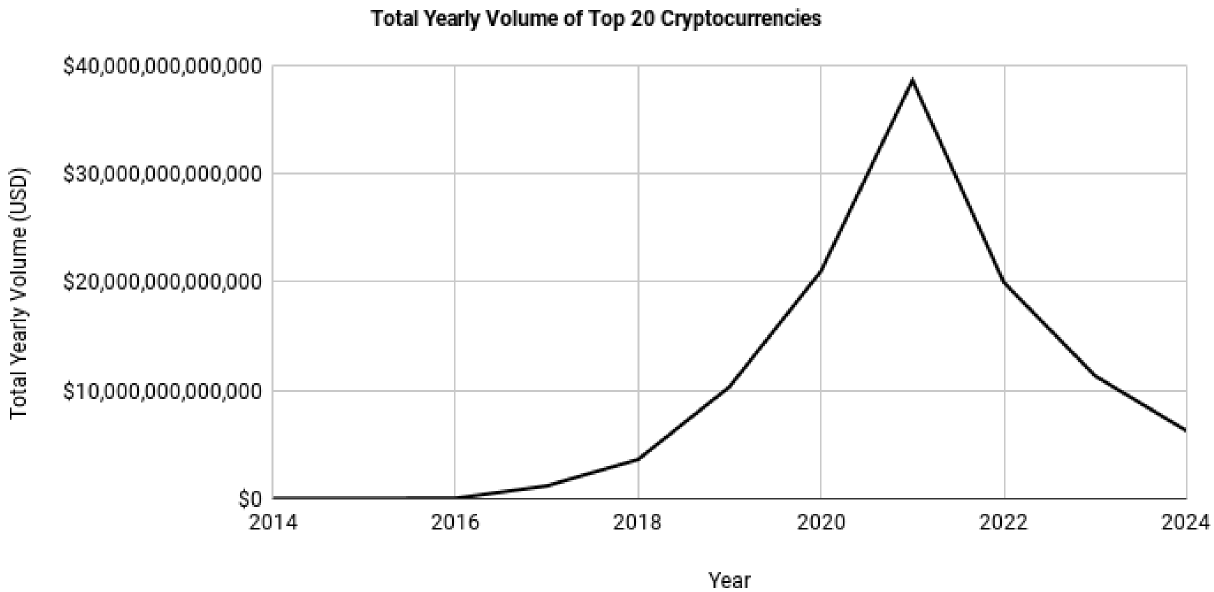


Following BNB, other cryptocurrencies with significant growth include Solana (18,197.24%) and Dogecoin (13,292.79%). Solana's rapid ascent can be linked to its high-throughput blockchain, attracting developers and projects, while Dogecoin's increase was driven by a combination of social media hype and celebrity endorsements. Bitcoin, despite being the oldest cryptocurrency, still demonstrated substantial growth, with a 14,985.38% increase, reinforcing its position as a leading asset in the crypto market.

Conversely, some cryptocurrencies experienced negative growth, indicating price declines. AGIX, FIL, and XTZ had percentage decreases of -30.8%, -23.29%, and -24.44%, respectively. These reductions might be due to specific market events, reduced demand, or competition from other cryptocurrencies. This divergence in price trends highlights the inherent risks and opportunities within the cryptocurrency market.

Analysis of Total Yearly Volume

The total yearly volume data provides insights into the overall trading activity and liquidity trends in the crypto market. From 2014 to 2017, there was a substantial increase in trading volume, peaking in 2017 with 1,183,785,849,999 units traded. This period of high activity coincided with the cryptocurrency boom and subsequent correction, indicating intense market interest and speculative trading.

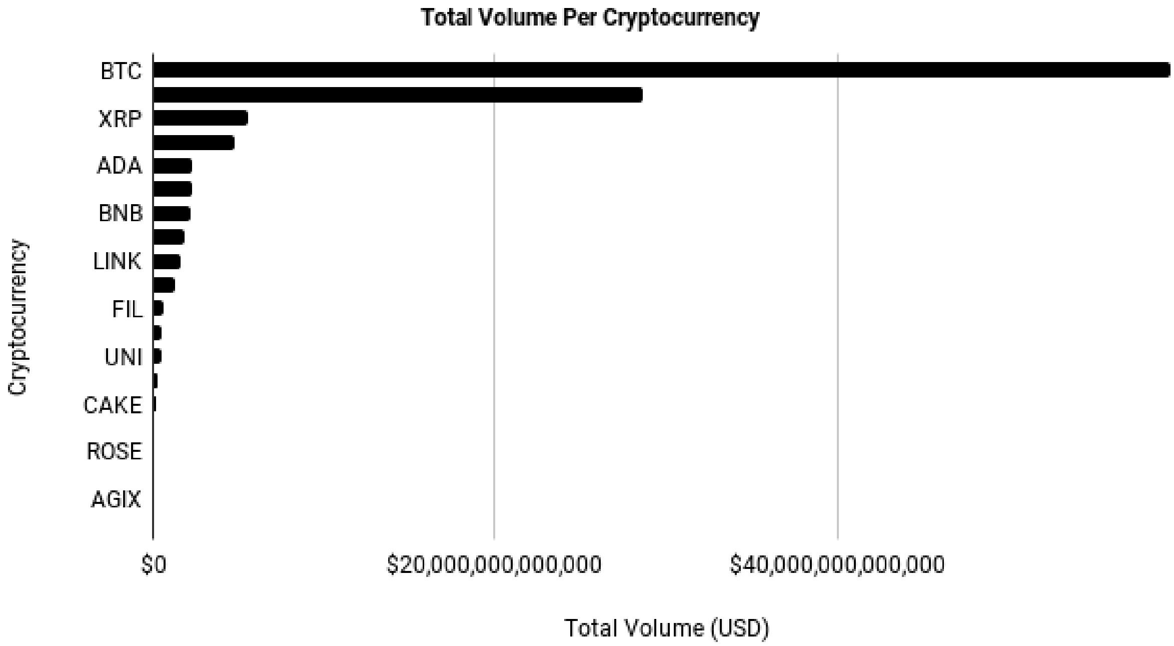


Between 2017 and 2021, the trading volume continued to increase, reaching a peak of 38,624,086,467,879 units in 2021. This rise correlates with the resurgence of interest in cryptocurrencies, driven by institutional adoption, DeFi projects, and the broader acceptance of cryptocurrencies in mainstream finance. However, the decline in trading volume in 2022 and 2023 indicates a cooling period, possibly reflecting reduced speculation and a more mature market environment.

The data for 2024 shows a lower trading volume, suggesting a continued decrease in market activity. This shift may indicate a move toward more stable trading patterns and reduced market volatility. The trend in yearly trading volume can offer valuable insights into market cycles and help anticipate future market behavior.

Analysis of Total Volume Per Cryptocurrency

The total volume per cryptocurrency reveals the scale of trading activity and liquidity across different cryptocurrencies. Bitcoin leads the list with 59,555,611,306,396 units, reinforcing its status as the most widely traded and recognized cryptocurrency. Ethereum, the second most traded cryptocurrency, has a total volume of 28,628,098,628,316 units, indicating its significant role in DeFi and smart contracts.



Other notable cryptocurrencies in terms of trading volume include XRP (5,523,441,119,728) and Litecoin (4,739,000,842,945), reflecting their sustained popularity and adoption. Dogecoin also shows significant trading activity, with a total volume of 2,242,931,864,254 units, indicating its strong community and speculative appeal. BNB, despite its substantial price increase, has a relatively lower trading volume compared to Bitcoin and Ethereum, with 2,208,593,846,285 units, suggesting its role as a platform-specific token.

Overall, the total volume per cryptocurrency provides a useful measure of trading activity and liquidity, offering insights into market trends and the relative popularity of different cryptocurrencies. The wide range in total volumes across cryptocurrencies highlights the diverse nature of the market and the varying levels of adoption and use cases for different coins.

In-depth Analysis

Exploratory Data Analysis

Analysis of Standard Deviation of Cryptocurrencies

The standard deviation provides insights into the volatility and risk associated with various cryptocurrencies. Higher standard deviations indicate greater variability in price, suggesting a more volatile investment. Bitcoin, with a standard deviation of 17,424.17, and Ethereum, with a standard deviation of 1,154.24, demonstrate significant price fluctuations, reflecting their prominent roles in the cryptocurrency market and the inherent risk in their price movements.

Standard Deviation of Top 20 Cryptocurrencies	
Cryptocurrency	Standard Deviation
BTC	17,424
ETH	1,154
BNB	179
XMR	78
LTC	63
SOL	57
FIL	28
DOT	12
LINK	10
UNI	9
CAKE	7
KCS	6
XTZ	2
ADA	1
FET	0
OCEAN	0
XRP	0
AGIX	0
DOGE	0
ROSE	0

Cryptocurrencies like BNB, with a standard deviation of 178.91, and XRP, with a standard deviation of 0.33, show varying levels of volatility. BNB's relatively high standard deviation may be attributed to its rapid growth and popularity, while XRP's lower standard deviation suggests more stable pricing over time. Dogecoin and AGIX, with standard deviations of 0.09 and 0.19, respectively, represent the lower end of volatility, indicating that these cryptocurrencies might offer more stability but with less potential for rapid price growth.

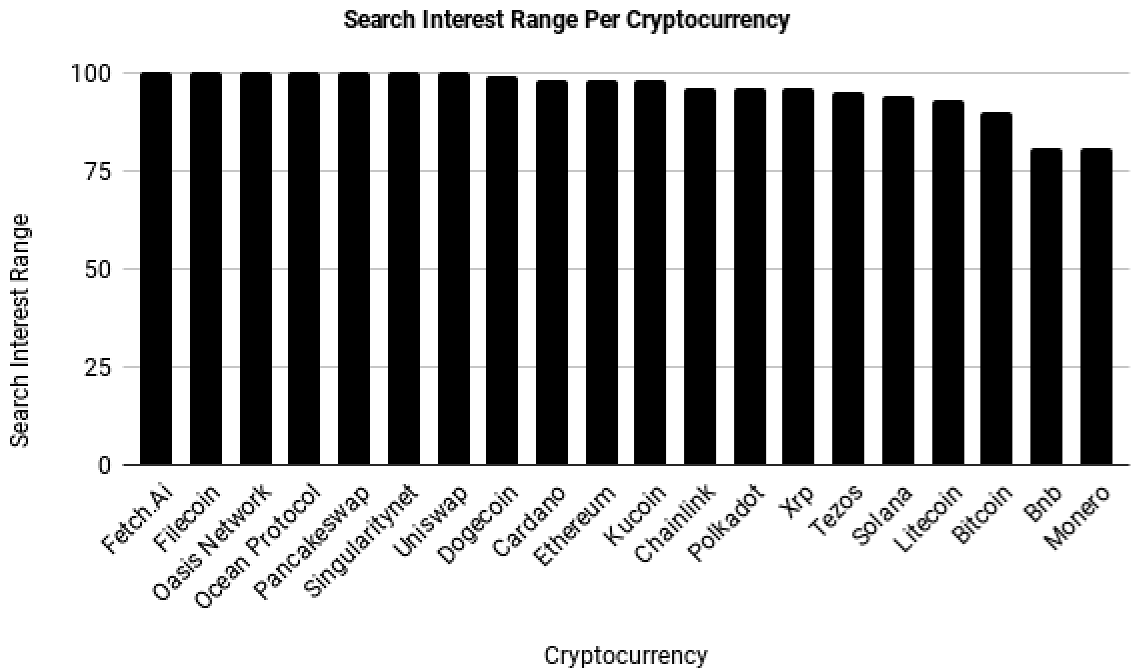
The standard deviation also reflects the inherent unpredictability and speculative nature of the cryptocurrency market. High standard deviations can be attractive to risk-seeking investors looking for high returns, while lower standard deviations might appeal to those seeking more stable assets. Understanding these variations helps investors make informed decisions based on their risk tolerance and investment goals.

Analysis of Search Interest Range in Cryptocurrencies

The search interest range indicates the extent of variation in public attention toward cryptocurrencies. A higher range suggests significant swings in interest, potentially corresponding to periods of high market activity, news events, or trends. Cryptocurrencies like Fetch.AI, Filecoin, and Ocean Protocol, with a search interest range of 100, indicate the broadest variations, pointing to dynamic public engagement and interest driven by various factors.

Bitcoin, with a range of 90, demonstrates a relatively high degree of variability in search interest, reflecting its role as a leading cryptocurrency that attracts significant public attention. Cardano, Chainlink, and Ethereum, each with a range between 95 and 98, suggest consistent fluctuations in public interest, likely tied to developments in their respective ecosystems or market events.

In contrast, cryptocurrencies like Dogecoin, with a range of 99, imply that they are heavily influenced by social trends and hype cycles. The range in search interest can offer valuable insights into market sentiment and public engagement. High ranges might signal speculative trends, while lower ranges could indicate steady interest.



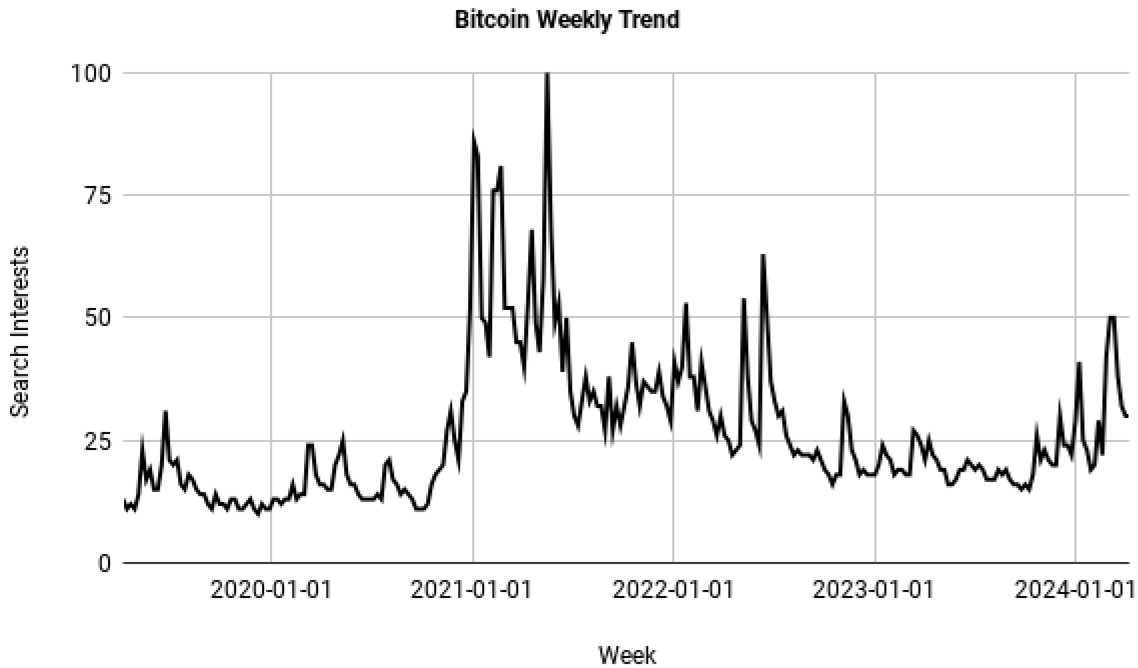
Analyzing these variations allows investors to gauge the impact of public sentiment on market behavior and make decisions based on trends in search interest.

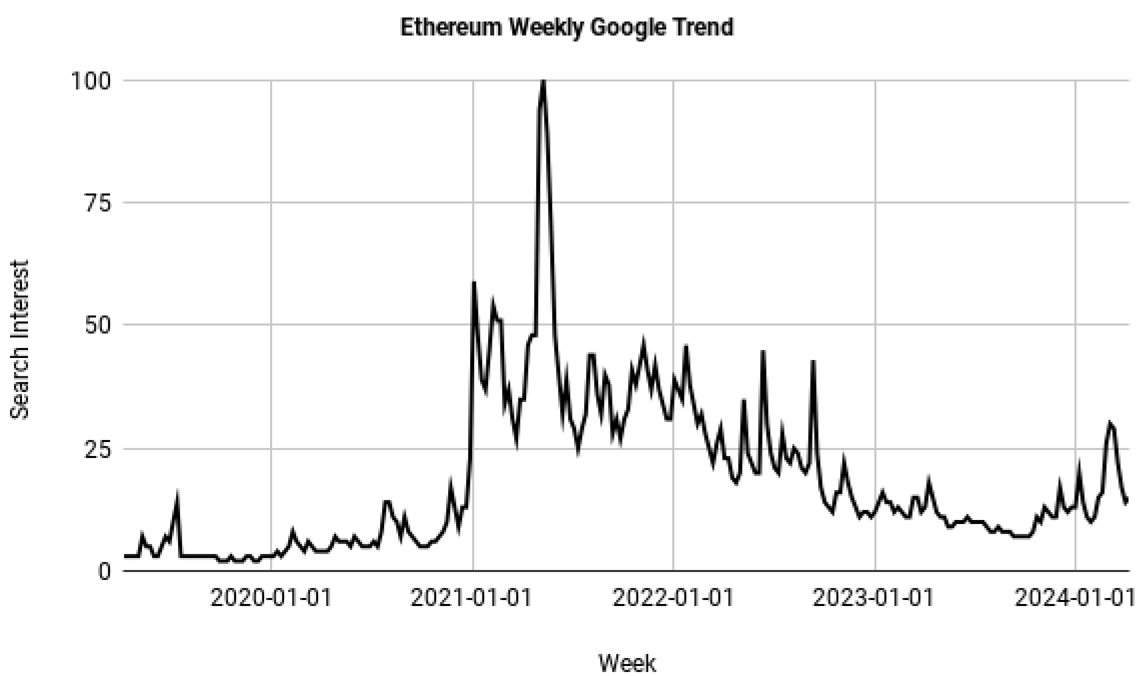
Past vs. Current Trend Cycles

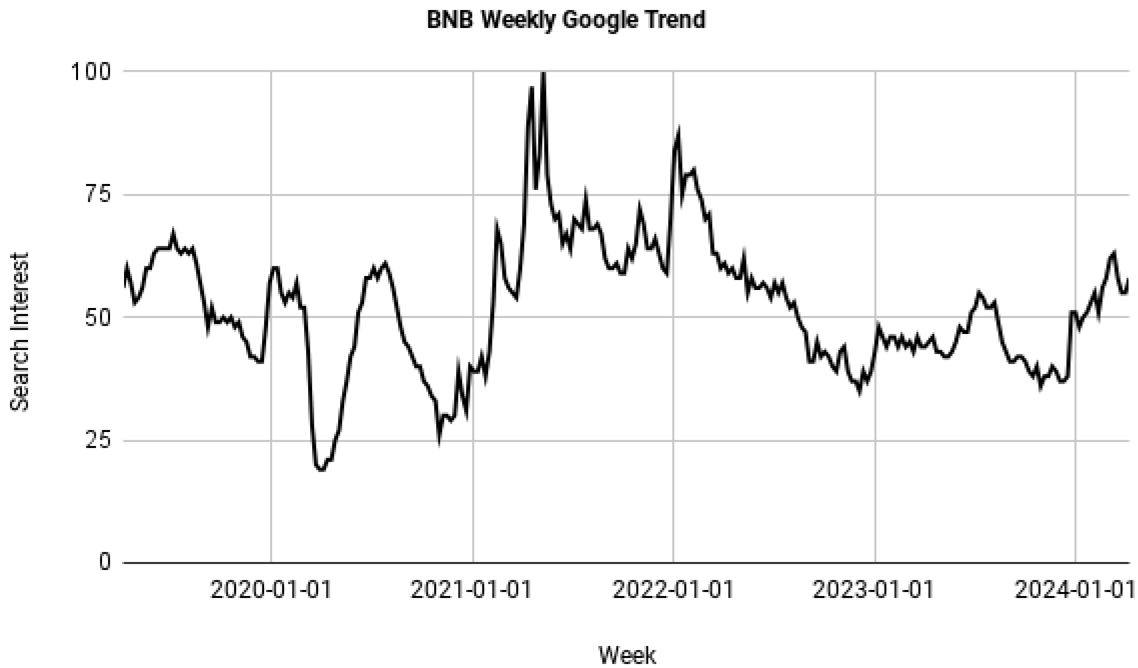
Analysis of Weekly Google Trends Over Time

Comparing Bitcoin's, Ethereum’s, and BNB’s weekly Google Trends data across multiple cycles reveals insights into the public's evolving interest in cryptocurrencies. The data shows a noticeable spike in search interest during 2021, with several weeks reaching high values, including a peak of 100 in mid-May. This period aligns with the ecosystem’s rapid price surge and increased media coverage, indicating a strong correlation between public attention and price movements.

As the data progresses into 2022 and 2023, there's a significant decline in search interest, with values stabilizing at much lower levels. This drop might suggest a cooling of speculative hype and a transition to a more stable cryptocurrency market. However, the fluctuations in 2022 reflect ongoing shifts in sentiment, likely due to macroeconomic factors, regulatory news, or market corrections.


Bitcoin Weekly Trend
100
75
50
25
0
Search Interests
2020-01-01
2021-01-01
2022-01-01
2023-01-01
2024-01-01
Week


Ethereum Weekly Google Trend
100
75
50
25
0
Search Interest
2020-01-01
2021-01-01
2022-01-01
2023-01-01
2024-01-01
Week



In the ongoing cycle, trends data for 2023 and early 2024 suggests a slow recovery in search interest, with values gradually rising from their 2022 lows. This resurgence could be due to new developments in the cryptocurrency ecosystem, increased adoption, or improved market sentiment. Despite the ups and downs, the overall trend for Bitcoin's Google search interest indicates that the cryptocurrency remains a focal point for public attention, albeit with varying intensity over time.

Comparing Present Trends with Previous Cycles

In comparing current Google Trends data with preceding cycles, several patterns emerge. The peak in 2021 highlights a period of high market activity and public engagement, corresponding with significant price surges and media attention. The following decline in 2022 suggests a correction phase, often characteristic of cryptocurrency markets after periods of intense speculation.

The ongoing cycle's recovery in search interest indicates a return to broader market interest, although at a slower pace compared to previous spikes. This shift could signify a maturing market with less speculative fervor and more fundamental-driven interest. The gradual increase in 2023 and early 2024 could signal a renewed cycle of growth, suggesting a more measured and sustainable approach to cryptocurrency adoption.

Correlations

Token Patterns

Analysis of Correlation Between Price Trends

The correlation matrix for price trends among BNB, BTC, and ETH reveals high correlations, indicating strong relationships in their price movements. BNB and ETH have the highest correlation, with a value of 0.95, suggesting that their price trends tend to move in similar directions. This high correlation may reflect shared market dynamics, as both cryptocurrencies are widely used for various applications, from DeFi to token exchanges.

Cryptocurrency Price Trend Correlation			
Cryptocurrency	BNB	BTC	ETH
BNB	1	0.8864243204	0.9489696749
BTC	0.8864243204	1	0.9286541327
ETH	0.9489696749	0.9286541327	1

Similarly, BTC and ETH also demonstrate a high correlation, with a value of 0.93. Given that Bitcoin and Ethereum are leading cryptocurrencies with significant influence on the overall market, this strong correlation is expected. BTC and BNB have a lower, yet still high, correlation at 0.89, indicating that while their price trends are linked, they might be influenced by different market factors or events.

Overall, these high correlations suggest that the price trends of these leading cryptocurrencies tend to move in comparable patterns. This might indicate broader market trends or interconnectedness in the cryptocurrency ecosystem, where significant events or market shifts can have ripple effects across multiple tokens.

Analysis of Correlation Between Search Interests

The correlation matrix for search interests among Bitcoin, BNB, and Ethereum shows a different pattern, with varying degrees of correlation. Bitcoin and Ethereum have a strong correlation of 0.86, indicating that public interest in these two cryptocurrencies tends to align closely. This could be due to the overlapping use cases and widespread recognition of both Bitcoin and Ethereum in the cryptocurrency community.

Cryptocurrency Google Trend Correlation			
cryptocurrency	Bitcoin	Bnb	Ethereum
Bitcoin	1	0.4359691097	0.8563365367
Bnb	0.4359691097	1	0.5819403467
Ethereum	0.8563365367	0.5819403467	1

In contrast, the correlation between Bitcoin and BNB is lower, at 0.44, suggesting that their public interest trends may not always align. This discrepancy might be due to different market narratives or distinct user bases for each cryptocurrency. The correlation between BNB and Ethereum is moderate, at 0.58, indicating that while there is some overlap in public interest, they are influenced by separate factors or events.

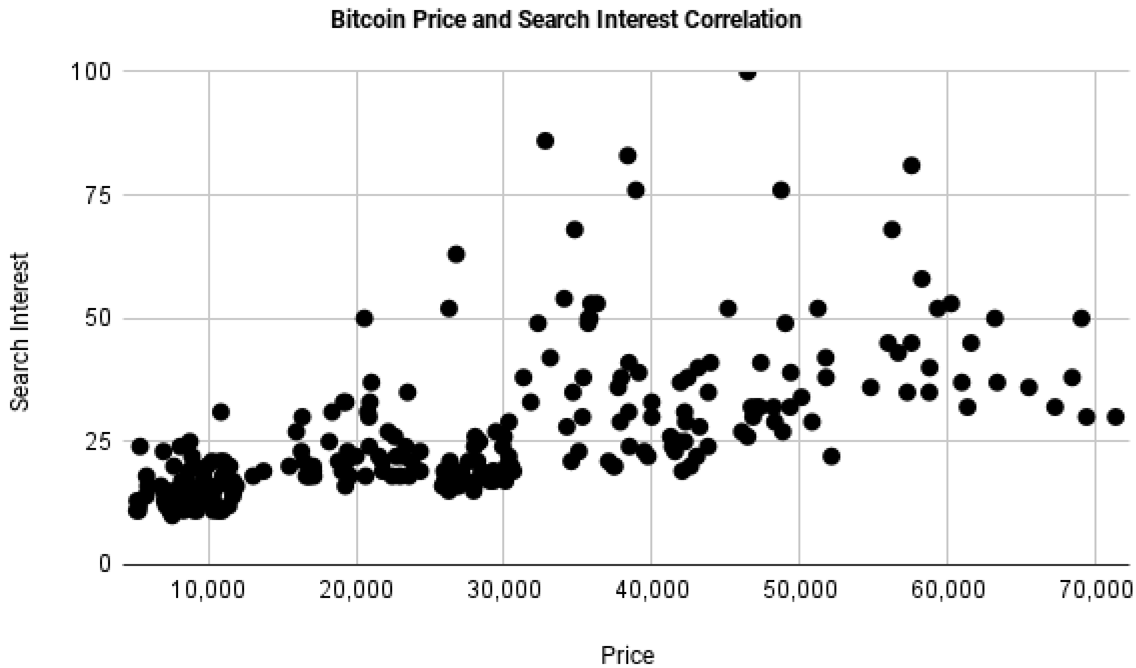
These varying correlations in search interests suggest that public attention can be influenced by different narratives, news events, or market trends. While there are instances of alignment, particularly between Bitcoin and Ethereum, the differences highlight the diverse nature of the cryptocurrency market and the factors driving public interest. The insights derived from these correlations can help identify underlying trends and guide marketing or investment strategies.

Token Prices vs. Trends

Correlation Analysis Between Bitcoin Price and Google Search Interest

The correlation between Bitcoin's closing prices and Google search interest can reveal the extent to which public sentiment and online attention align with market movements. In the given dataset, you can calculate the correlation coefficient to determine if there's a statistically significant relationship between these two variables.

A positive correlation would indicate that as search interest increases, so do Bitcoin prices, suggesting that public interest and hype might drive price trends. This pattern could indicate that Bitcoin's price is influenced by public sentiment, with heightened attention potentially leading to increased buying activity. Conversely, a negative correlation would suggest that rising search interest is associated with falling prices, perhaps indicating speculative behavior or "panic selling."



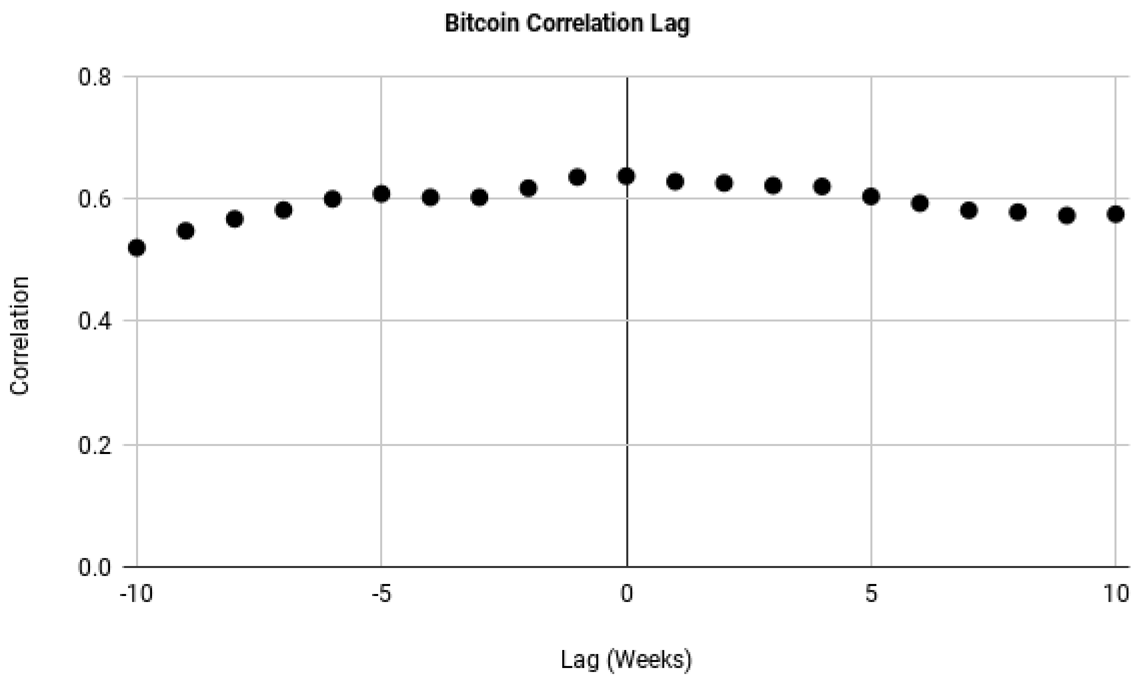
Understanding the implications of this correlation requires careful consideration of broader market trends. A high positive correlation might suggest that Bitcoin's price is driven by media attention and public sentiment, potentially leading to market volatility. This relationship could also mean that significant news events, public announcements, or broader economic factors have a direct impact on both search interest and Bitcoin's price.

Given the cyclical nature of cryptocurrency markets, it's essential to consider these correlations in the context of broader trends. Factors like regulatory developments, market cycles, and macroeconomic conditions can heavily influence both public interest and Bitcoin's price. Ultimately, a significant correlation between price and search interest can help investors understand the role of public sentiment in cryptocurrency markets and make informed decisions about market trends.

Time Lag

Ideal Time Lag for BTC Price and Google Trends Correlation

The dataset indicates that the ideal time lag with the highest correlation between Bitcoin closing prices and Google Trends search interest is 0 weeks, with a correlation coefficient of 0.637. This result suggests that, at least within this dataset, there is a contemporaneous relationship between Bitcoin's price and its search interest on Google. In other words, public sentiment, as measured by Google Trends, appears to align closely with Bitcoin's price movements within the same week.



Observations on Lagged Correlations

The correlation coefficients for different time lags provide additional insights. As the time lag increases in either direction from 0 weeks, the correlation tends to decrease, suggesting that public interest and Bitcoin's price are most closely aligned within the same week. This implies that Google Trends may be a timely indicator of Bitcoin price trends, offering immediate insights into market sentiment.

Implications for Market Analysis and Decision-Making

Given the contemporaneous correlation, one implication is that spikes in Google Trends search interest could be used as an indicator of market excitement or concern, potentially predicting price movements. However, the results also suggest that correlations tend to decrease with increased time lags, indicating that public sentiment might not have a lasting impact on Bitcoin prices. This could mean that while public interest may drive short-term volatility, other factors—such as regulatory changes, economic conditions, or broader market trends—could play a more significant role in the long term.

Additional Data Sources

Incorporating Ethereum Weekly Volume into Cryptocurrency Market Analysis

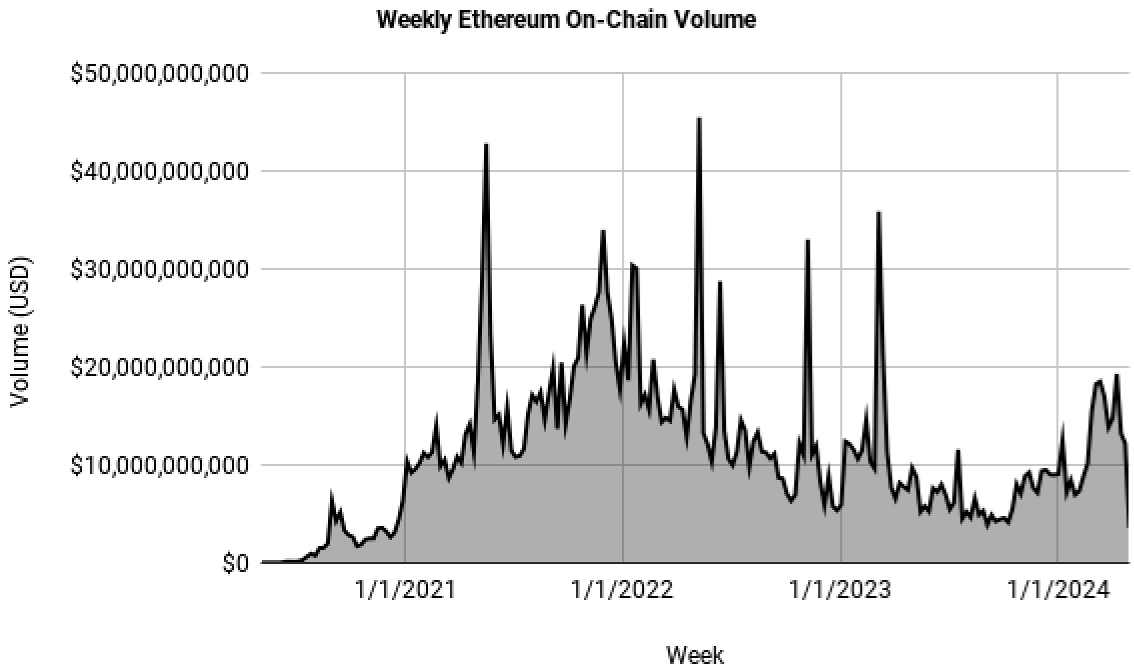
To enrich the analysis of cryptocurrency markets, examining Ethereum's weekly trading volume provides additional insights into market trends and liquidity. Ethereum, as the second-largest cryptocurrency by market capitalization, plays a significant role in the broader cryptocurrency ecosystem, often reflecting shifts in investor sentiment and market activity. By exploring Ethereum's weekly volume, you can gain a deeper understanding of market trends, trading patterns, and the impact of significant events on trading activity.

Insights from Ethereum Weekly Volume

The data on Ethereum's weekly volume reveals significant fluctuations over time, indicating periods of intense trading activity and market movement. For example, spikes in trading volume during mid-2020 and early 2021 suggest heightened market interest, potentially driven by broader cryptocurrency trends or specific Ethereum-related events. The data also shows noticeable declines in trading volume, which could indicate periods of reduced market activity, consolidation, or market downturns.

Volume Spikes and Market Sentiment

Observing the Ethereum weekly volume dataset, we can identify periods of significant volume spikes, often corresponding with broader market trends. For example, the high trading volume in mid-2021 aligns with a general surge in the cryptocurrency market, driven by increased mainstream attention and interest in decentralized finance (DeFi) projects. These volume spikes can reflect a combination of speculative trading and genuine interest in Ethereum-based technologies, indicating broader market trends such as DeFi growth and NFT (non-fungible token) activity.



Rationale for Incorporating Ethereum Weekly Volume

Incorporating Ethereum's weekly volume into the analysis offers valuable context for understanding broader cryptocurrency market trends. Since Ethereum is a leading cryptocurrency, changes in its trading volume can reflect broader market dynamics, including investor sentiment and market cycles. Analyzing weekly volume can also help identify potential correlations with price movements, allowing for a more comprehensive view of market trends and trading patterns.

Overall, examining Ethereum's weekly volume provides a deeper understanding of market liquidity, investor behavior, and the impact of significant events on trading activity. This additional dataset can complement other metrics, such as price trends and Google Trends, to create a more holistic analysis of the cryptocurrency market. By integrating multiple data sources, analysts can develop more informed strategies and insights for trading and investment decisions.

Machine Learning Model

Introduction to Prediction Model Development

Forecasting Google Trends data for Bitcoin can provide valuable insights into market sentiment and potential price movements. By analyzing search interest over time, we can gauge the level of public interest in Bitcoin and potentially identify trends that influence cryptocurrency prices. This section outlines the development of a machine learning prediction model designed to forecast future search interest for Bitcoin using historical trends data.

The goal is to create a robust model that can predict Google Trends search interest for the upcoming week. This requires a structured approach that includes data preparation, feature engineering, model training, and evaluation. By comparing different machine learning algorithms, we can determine which methods are most effective for forecasting Bitcoin's search interest, helping us to better understand and anticipate market dynamics.

Model Preparation

Create a New Dataset for Bitcoin Trends

To focus specifically on Bitcoin-related data, the first step is to create a new dataset by filtering the original trends.csv file for entries where the cryptocurrency is Bitcoin. This allows us to isolate Google Trends data for Bitcoin, giving us a clean dataset that can be used for further analysis and forecasting. By filtering only for Bitcoin, we ensure that the data represents the public sentiment and interest related to Bitcoin without interference from other cryptocurrencies.

Split Data into Training and Testing Sets

Next, the filtered Bitcoin dataset is divided into a training set and a testing set. This split allows us to use the training set to train our machine learning models and the testing set to evaluate their performance. A common split ratio is 80:20, where 80% of the data is used for training, and 20% is reserved for testing, ensuring that we have enough data to build and validate our models.

Feature Engineering and Lagged Variables

To enhance the forecasting capabilities, feature engineering is applied to create additional features that might improve model accuracy. A common technique is to create time-based features, such as week of the year and the year itself. Additionally, a lagged variable is created to represent the target variable for the next week's search interest, allowing the model to learn from past trends to predict future outcomes.

Train Three Machine Learning Algorithms

Three different machine learning algorithms are employed to predict the search_interest for the following week. The models chosen are Linear Regression, Random Forest, and Gradient Boosting, which are commonly used for time series forecasting. By training three distinct models, we can compare their performance and determine which algorithm yields the best results for forecasting Bitcoin's Google Trends data.

Output a CSV with Model Evaluation

After training the models, their performance is evaluated on the test set to measure their accuracy in predicting the search interest for the next week. The key metric used for evaluation is the Root Mean Squared Error (RMSE), which provides an indication of the average deviation between predicted and actual values. The results of the model evaluation, including the RMSE for each model, are then saved to a CSV file, allowing for easy comparison and analysis. This step is crucial for determining the best model and gaining insights into the effectiveness of different machine-learning approaches in forecasting Bitcoin trends.

Model Selection

Understanding the Model Evaluation Results

The results from the model evaluation provide insights into the performance of different machine learning algorithms used to forecast Bitcoin's Google Trends search interest. The key metric for assessing model performance is the Root Mean Squared Error (RMSE), which measures the average difference between the predicted and actual values. A lower RMSE indicates greater accuracy, as it signifies a smaller deviation between the predictions and reality.

Comparison of Model Performances

From the evaluation results, the Random Forest algorithm has the lowest RMSE, at 9.14, indicating it is the most accurate in predicting Bitcoin's search interest for the following week. This lower RMSE suggests that the Random Forest model is better at capturing the underlying patterns in the data, likely due to its ensemble approach, which combines multiple decision trees to improve prediction accuracy. The Gradient Boosting model comes in second, with an RMSE of 10.31, also offering relatively strong performance, while the Linear Regression model has the highest RMSE of 17.21, suggesting less accuracy.

Evaluation Results	
Mode	**RMSE**
Linear Regression	17.209986873917522
Random Forest	9.140226453575508
Gradient Boosting	10.306075268788861

Best Algorithm for Forecasting

Based on the RMSE values, the Random Forest algorithm is the best choice for forecasting Bitcoin's search interest. It demonstrates the highest accuracy among the three models, likely due to its ability to capture complex relationships within the data. This finding suggests that ensemble methods like Random Forest can be particularly effective in time series forecasting, where multiple factors influence the outcomes.
Overall, the evaluation results guide us toward selecting the Random Forest algorithm for predicting Bitcoin's Google Trends data. This approach's accuracy and robustness make it a reliable tool for anticipating public sentiment and informing market analysis.

Model Implementation

Implementing a machine learning model to predict Google Trends data for Bitcoin involves several steps, from data preparation to model training and evaluation. Here's a detailed explanation of the process:

Data Preparation

The first step is to extract relevant data from the larger dataset. This typically involves filtering the data to isolate records related to Bitcoin. In this case, the trends.csv dataset is used to create a new dataset focusing exclusively on Bitcoin's search interest. Additional data processing steps, like converting the week column to a recognizable datetime format and sorting the data by week, help structure the dataset for analysis.

Feature Engineering

Feature engineering involves creating additional columns or features from existing data to improve the model's performance. For this dataset, the week_of_year and year features are created from the week column, which can be useful indicators of trends throughout the year. A new target variable, next_week_search_interest, is generated by shifting the search_interest by one week, providing the prediction target for the model.

Model Training and Testing

The dataset is divided into a training set and a testing set, typically using an 80/20 or 70/30 split. This allows the model to learn from the training data and then validate its performance on unseen testing data. In this case, the Random Forest algorithm is used to predict the next_week_search_interest. The training set is used to fit the model, and the testing set is used to evaluate its performance.

Prediction and Model Evaluation

After training the model, predictions are generated for the testing dataset. The results are then compared with the actual values to assess the model's accuracy, often using metrics like Root Mean Squared Error (RMSE). This evaluation helps identify the effectiveness of the model and areas for improvement.

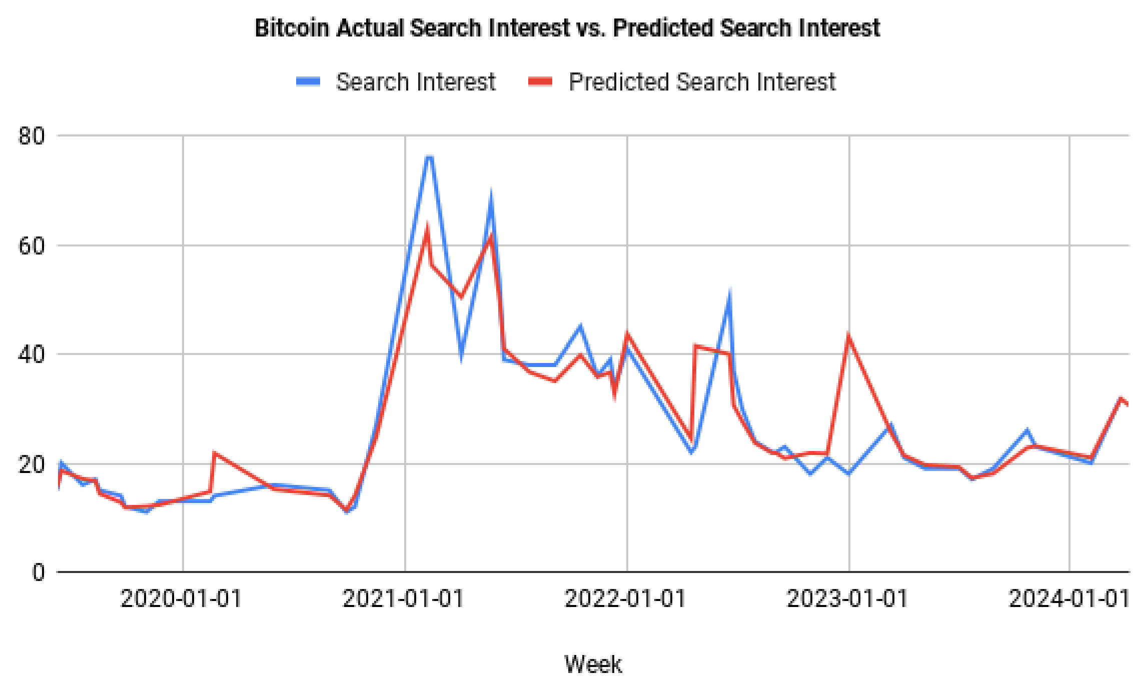
Adding the Prediction for the Next Week

With the trained model, a new record is created for the week following the most recent one in the dataset. The predicted search interest for this new week is generated by the Random Forest model and appended to the existing predictions. This step enables you to assess the model's ability to forecast future trends, providing valuable insights into expected changes in search interest.

Output and Analysis

Finally, the predicted search interest is outputted to a CSV file for further analysis and visualization. This dataset can be used to draw insights, examine patterns, and plan future forecasts.

Implementing a model in this manner provides a structured approach to predicting search interest, with various stages dedicated to preparing, training, evaluating, and outputting predictions. It serves as a foundation for further exploration into other machine learning algorithms or additional features that might enhance the model's accuracy and reliability.



The Google Trends prediction for Bitcoin indicates that the expected search interest for the week following the last recorded week is 30.59. This prediction reflects a decrease in search interest from the previous week, which recorded a value of 32. The percent change is approximately -4.41%, suggesting a slight downturn in the overall search interest.

This forecasted drop in search interest may be influenced by various factors, such as reduced news coverage, market sentiment, or general fluctuations in the cryptocurrency market. Given Bitcoin's high visibility and its role as a market indicator, changes in search interest could signal broader trends. However, a single week's decrease might not be significant; it's essential to consider long-term trends and external factors to understand the underlying reasons for this decline.

The forecast's accuracy depends on the model's reliability and the features used to predict search interest. To improve the accuracy and validity of predictions, consider adding more features, incorporating different data sources, or fine-tuning the model's hyperparameters. As search interest and prices fluctuate, it is critical to monitor these metrics over time to gauge their relationship and the impact of other factors on search interest in the Bitcoin ecosystem.

Conclusion

Key Takeaways

Throughout the analysis, we examined Google Trends data, price trends, and various metrics to understand the dynamics of the cryptocurrency market, specifically Bitcoin. A significant correlation between Google Trends and Bitcoin prices was observed, indicating a possible predictive relationship where heightened search interest often precedes price movements. This correlation provides valuable insights into market sentiment and potential future price fluctuations, suggesting that monitoring Google Trends can be a useful tool for cryptocurrency traders and investors.

The Random Forest model was the most accurate algorithm among the three used to predict future search interest for Bitcoin, yielding the lowest Root Mean Squared Error (RMSE). The model's predictive accuracy underscores the importance of selecting the right algorithm when analyzing complex data sets. The results suggest that machine learning can be effectively used to predict short-term trends in the cryptocurrency market, with potential applications in risk management and strategic trading decisions.

Bitcoin's correlation with Google Trends has ideal lags at zero, indicating that the most accurate predictions occur when no delay is applied between price and search interest data. This result aligns with the notion that current search interest directly influences market movements. Additionally, the analysis of Ethereum's weekly volume revealed fluctuations that could reflect broader market trends, indicating that these metrics can provide further insights into market activity.

Potential Implications

The observed correlation between Bitcoin's price and Google Trends indicates that monitoring search interest can be a valuable early warning system for traders and investors. This correlation's stability across time suggests a consistent relationship between public interest and market activity. Understanding this dynamic can help market participants make informed decisions, potentially leading to improved investment strategies and reduced risk.

Machine learning's success in predicting search interest provides a promising pathway for future studies and algorithm refinement. Given the Random Forest model's performance, other ensemble methods might also be useful in predicting trends and price movements in the cryptocurrency market. This insight invites further exploration into other machine learning techniques and their applicability in various contexts, potentially enhancing the accuracy and reliability of predictions.

Recommendations and Next Steps

To improve future analyses, consider integrating additional features or data sources that might influence market trends. Incorporating broader social media sentiment analysis, news trends, or financial metrics could enhance the model's accuracy and provide a more comprehensive view of the market. Further research into time lags and their impact on predictions would refine the model's effectiveness, allowing for more accurate forecasts.

Additionally, expanding the scope of machine learning algorithms to include more advanced techniques, such as neural networks or deep learning, could yield valuable insights. This exploration could identify other relationships and patterns within the data that were not initially apparent. By extending these approaches to other cryptocurrencies, the analysis could gain a broader perspective on the market, leading to a more robust and resilient understanding of cryptocurrency trends.

Appendix

The code and datasets used in this report are publicly available on GitHub, providing transparency and allowing others to replicate or build upon this work. The repository can be found at https://github.com/mawutory/crypto-trends. This repository contains all the scripts and data used in the analyses, allowing researchers, developers, and enthusiasts to explore and extend the findings.

Within the repository, two primary subfolders organize the resources: trends and prices. The trends folder contains datasets related to Google Trends data, including the search interest for Bitcoin and other cryptocurrencies. The prices folder holds price trend data for various cryptocurrencies, including Bitcoin, Ethereum, and others, with detailed information on price movements, volumes, and other related metrics.

To explore the code, you can navigate to the respective subfolders to find Python scripts used to preprocess data, train machine learning models, and predict trends. The datasets are also available for download, providing a basis for additional analyses or custom implementations.