Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi
Pamukkale University Journal of Engineering Sciences

A flexible and reusable framework for Bitcoin blockchain and price data

Bitcoin blok zinciri ve fiyat verileri için esnek ve yeniden kullanılabilir bir çerçeve

Vasfi Tataroglu[1], Gurhan Gunduz[1], Sezai Tokat[1], Hasan Bulut[2]*

[1]Department of Computer Engineering, Faculty of Engineering, Pamukkale University, Denizli, Türkiye.
vtataroglu@pau.edu.tr, ggunduz@pau.edu.tr, stokat@pau.edu.tr
[2]Department of Computer Engineering, Faculty of Engineering, Ege University, Izmir, Türkiye.
hasan.bulut@ege.edu.tr

Received/Geliş Tarihi: 01.02.2025 Revision/Düzeltme Tarihi: 14.07.2025 doi: 10.5505/pajes.2025.26511
Accepted/Kabul Tarihi: 18.07.2025 Research Article/Araştırma Makalesi

Abstract

Driven by the potential for high returns and financial innovation, the cryptocurrency market is seeing a rapid increase in the number of investors and market value. Unlike other financial assets, cryptocurrencies have a transparent and accessible data structure that makes it possible to examine the relationships between price movements and on-chain metrics. This growing interest, as well as the transparent and accessible structure, is leading to a parallel increase in academic studies focused on analyzing the price dynamics of Bitcoin, the world's first widely accepted cryptocurrency, and the underlying factors by examining Bitcoin blockchain data and developing predictive models. The current literature generally relies on data collected at limited, fixed time intervals on a daily basis. However, the daily and fixed data sets of the cryptocurrency market in the literature become outdated over time and are not sufficient to analyze instantaneous price movements. This poses a significant limitation when analyzing cryptocurrency markets, which are highly volatile and require a focus on rapid price fluctuations and immediate market changes. For detailed analysis, investors and researchers increasingly require data at a higher frequency than daily. In this study, a framework is designed and developed that allows users to obtain Bitcoin blocks and price data for any desired time period and interval. In this way, a comprehensive, up-to-date and flexible framework is offered that can serve various research areas, such as machine learning-based predictions, testing algorithmic trading strategies, investigating correlations between the blockchain and price fluctuations, performing market analysis or detecting anomalies. In addition, various frequency aggregations were defined using the framework developed in this study. In particular, a 10-minute time interval was analyzed based on the average time it takes to generate a Bitcoin block. In this way, a direct and consistent relationship between block information and price movements could be established.

Keywords: Digital Assets, Cryptocurrency Market, Bitcoin, Blockchain, On-Chain Analysis

Öz

Yüksek getiri potansiyeli ve finansal inovasyondan kaynaklanan ilgi, kripto para piyasasının yatırımcı sayısında ve piyasa değerinde hızlı bir artışa neden olmaktadır. Diğer finansal varlıklardan farklı olarak, kripto paralar fiyat hareketleri ile zincir içi (on-chain) metrikler arasındaki ilişkileri incelemeyi mümkün kılan şeffaf ve erişilebilir bir veri yapısına sahiptir. Bu artan ilgi ve şeffaf, erişilebilir yapı, Bitcoin blok zinciri verilerini inceleyen ve fiyat dinamiklerini analiz edip tahmin modelleri geliştiren akademik çalışmalarda da paralel bir büyümeye yol açmaktadır. Mevcut literatürde veriler genellikle günlük bazda ve sabit zaman aralıklarında toplanmakta, ancak kripto para piyasasındaki günlük ve sabit veri setleri zamanla güncelliğini yitirmekte ve anlık fiyat hareketlerini analiz etmek için yeterli olmamaktadır. Bu durum, yüksek volatiliteye sahip kripto para piyasalarında hızlı fiyat dalgalanmaları ve ani piyasa değişikliklerine odaklanmayı zorunlu kılan analizler için önemli bir sınırlamadır. Daha detaylı analiz yapabilmek için, yatırımcılar ve araştırmacılar giderek günlük düzeyin üzerinde bir veri sıklığına ihtiyaç duymaktadır. Bu çalışmada, kullanıcıların istedikleri herhangi bir zaman aralığı ve periyotta Bitcoin blok ve fiyat verilerine ulaşmasını sağlayan bir çerçeve tasarlanmış ve geliştirilmiştir. Bu sayede makine öğrenmesi tabanlı tahminler, algoritmik alım-satım stratejilerinin test edilmesi, blok zinciri ile fiyat dalgalanmaları arasındaki ilişkilerin incelenmesi, piyasa analizleri veya anormallik tespiti gibi çeşitli araştırma alanlarına hizmet edebilecek kapsamlı, güncel ve esnek bir çerçeve sunulmaktadır. Ayrıca bu çalışmada geliştirilen çerçeveyle birlikte farklı frekanslarda veri toplulaştırmaları da tanımlanmıştır. Özellikle bir Bitcoin bloğunun ortalama üretim süresi olan 10 dakikalık zaman aralığı temel alınarak yapılan analiz, blok bilgilerinin fiyat hareketleriyle doğrudan ve tutarlı bir şekilde ilişkilendirilmesine olanak sağlamaktadır.

Anahtar kelimeler: Dijital Varlıklar, Kripto Para Piyasası, Bitcoin, Blok Zinciri, Zincir Üzeri Analiz

1 Introduction

Digital assets represent a broad category of intangible assets that are created, transferred and stored electronically, usually using blockchain or distributed ledger technology[1]. Blockchain is a specialized technology that uses interconnected blocks of data protected by cryptographic hash functions to record information. It is designed as a distributed and decentralized digital ledger whose purpose is to record all transactions in sequential order [2]. A continuously growing sequential list of data records called blocks is stored in this structure. The name blockchain is derived from this property [3].

Cryptocurrencies, a subset of digital assets, are a significant example of the transformative potential of intangible assets in the modern economy. Digital currencies such as Bitcoin and Ethereum use blockchain technology to enable decentralized, secure and transparent transactions without the need for intermediaries such as banks. Cryptocurrencies have a similar value to traditional assets but offer unique advantages due to their digital structure, such as low transaction costs, unlimited

*Corresponding author/Yazışılan Yazar

transfers and programmable functions through smart contracts[4]. While traditional financial systems rely on banks, central banks and other intermediary institutions to manage money transfers, this novel structure reduces reliance on intermediaries such as banks and enables peer-to-peer transfers across geographical boundaries, lowering transaction costs for asset transfers and transforming global economic governance [5].

The most important feature of digital cryptocurrencies is that it is almost impossible to change transactions once they have been recorded in the blockchain. Each participant retains a copy of the entire transaction history, and new transactions can only be made with the consent of the majority of participants in the network. This consensus mechanism preserves the reliability and integrity of the system.

With its decentralized structure, transparency and global investor infrastructure, Bitcoin is one of the most popular digital cryptocurrencies of recent years. Blocks, which are the basic building blocks of Bitcoin, contain transactions that have been authorized within a certain period of time, and these blocks are cryptographically chained together. Each block, with the exception of the first genesis block, contains a hash value as a summary of the previous block, making retroactive changes mathematically complex. This chain structure forms the basis for the security and immutability of Bitcoin.

Each new block, which is generated approximately every 10 minutes, ensures the security of the network and the controlled multiplication of the Bitcoin supply. This duration is dynamically adjusted by a parameter called mining difficulty [6]. Over the years, the mining difficulty has increased and the number of Bitcoins waiting to be mined has decreased. The hash rate, on the other hand, is a measure of the computing power of the miners. As the mining difficulty increases, the hash rate improves accordingly, with the average block production time remaining roughly constant [7]. This mechanism allows Bitcoin to have a predictable and inflation-proof supply program [8].

Block transactions are records that transfer the Bitcoin assets belonging to a user from one address to another and represent the fundamental property of blocks. Each transaction is secured by digital signatures. These signatures mathematically prove that the person making the transaction is the true owner of the Bitcoin digital asset. Transactions are signed with private keys and can be verified with public keys that are visible to everyone [9]. This cryptographic infrastructure allows Bitcoin to function as a secure and verifiable system for transferring digital assets.

With the price of Bitcoin at over one hundred thousand dollars, global interest in cryptocurrency markets has risen again, which has refocused the attention of investors and researchers seeking to understand market dynamics. This historic high has prompted not only individual investors, but also financial institutions and academic circles to re-evaluate Bitcoin. In particular, the approval of exchange-traded Bitcoin spot funds and the growing interest of major financial institutions in crypto-assets show that this market is increasingly being integrated into the traditional financial system.

The Bitcoin price has momentary fluctuations, higher volatility compared to traditional markets and unique dynamics due to its blockchain-based structure, making price prediction an important and challenging area of research. On the one hand, this volatility offers high potential returns, but on the other hand it also poses a major challenge for risk management. While price movements in traditional financial assets can often be explained by economic indicators, company performance or macroeconomic events, changes in the price of Bitcoin are subject to much more complex factors. These factors include non-traditional variables such as network activity, mining difficulties, halving events, regulatory news, technological developments and social media sentiment.

Price forecasting refers to the process of predicting future price levels and plays a crucial role in strategy development, risk management, algorithmic trading applications and better understanding of market trends. In modern financial markets, especially for high-frequency transactions, manual analysis and decision-making processes are proving insufficient. Machine learning and artificial intelligence-based forecasting models are therefore becoming increasingly important. These models go beyond the traditional indicators of technical analysis and incorporate alternative data sources such as transaction data on the blockchain, network metrics and signals from social media.

With the entry of institutional investors into the market, the need for more sophisticated risk management and portfolio optimization tools has increased. Price forecasting models not only predict future prices, but also perform important functions such as identifying potential risk factors, calculating optimal position sizes and dynamically adjusting stop-loss levels. These models also provide valuable insights for recognizing market manipulation, identifying arbitrage opportunities and managing liquidity.

However, the unique challenges of cryptocurrency markets make the direct application of traditional financial models difficult. In these round-the-clock markets, global events can have an immediate impact on prices, and a social media post or protocol update can lead to unexpected price movements. Therefore, a successful price prediction model must integrate both traditional financial indicators and the unique dynamics of the cryptocurrency ecosystem, while adapting quickly to changing market conditions.

Price candlestick charts are powerful analytical tools commonly used to study Bitcoin price movements . These charts visually represents the open, close, high and low prices that occur within a given time interval. This visual representation method, which was developed by Japanese rice traders in the 18th century, is now an indispensable part of the financial markets. Each candlestick combines four fundamental price points in a single chart and provides quick and comprehensive information about the direction and intensity of price movements.

Candlesticks on a daily, hourly or minutely basis allow both short and long-term analysis in this environment where market trends can change rapidly. Each time frame offers a different perspective: daily candlesticks are used for general trend analysis, while hourly candlesticks are effective for capturing medium-term price movements. 1-minute-based candlesticks are crucial for investors who trade short-term. These different time scales allow investors to choose the analytical perspective that best suits their trading strategies and risk tolerance. However, the ever faster reaction of the market as well as the need of investors for more detailed data in minute or even second intervals encourages researchers to increase the frequency of data. There are several important factors behind this trend. First, the widespread use of algorithmic trading systems requires the development of programs capable of making buy and sell decisions within milliseconds. In addition, the integration of global markets and the immediate impact of social media mean that price movements are happening much faster.

High frequency data reveals the microstructure of the market more clearly and makes it possible to model price movements with greater precision. Thanks to this data, it is possible to go beyond the traditional indicators of technical analysis and analyze factors such as momentary shifts in buying and selling pressure, liquidity dynamics and market depth more effectively. For example, even within a one-minute time frame, multiple buying and selling waves can occur, presenting opportunities and risks that cannot be identified by examining daily or hourly data alone.

This high-resolution data analysis provides valuable input, particularly for machine learning models. Deep learning algorithms can recognize complex patterns in high-frequency data, enabling more accurate modeling of price movements. In addition, this data can be used in important areas such as the detection of market manipulation, the identification of arbitrage opportunities and the optimization of risk management strategies.

When examining the existing literature, most studies focus on periods prior to 2017 and often use a daily interval for data frequency, which is insufficient to capture the new dynamics of the market. Moreover, the shared data in the data repositories are outdated over time. This shortcoming clearly shows the need for updatable data sets with flexible resolution.

In this study, we designed and developed a novel framework with a comprehensive 8-step data processing workflow that provides a foundation for the application development process and enables users to obtain Bitcoin blocks and price data in the desired time periods and frequencies. In addition, this framework was used to systematically collect, cleanse and transform Bitcoin block and price data over a large period from 2017 to 2024. In the resulting structure, the raw data is transferred into a regular and usable database format using tools such as Python, Shell Script, Pandas and SQLite. Then, to synchronize with the blockchain data, high-frequency (e.g. 1-minute interval) Bitcoin price data is parametrically combined for the intervals intended for analysis, resulting in a consistent data set that is aligned with the average block time. In this way, it is possible to transform, update and query data over different time periods and at different time levels. Furthermore, with the provided Python code, Flask-based APIs and data query methods, researchers can keep the dataset up to date and quickly integrate it into their own analysis environments, all in the spirit of automation and reusability.

The rest of the paper is organized as follows: Chapter 2 deals with the related work. In chapter 3, we provide some background information that contributes to the understanding of the issues addressed in this study. The proposed framework is explained in Chapter 4. Chapter 5 deals with the case study in which our framework is used. We conclude the paper in Chapter 6.

2 Related work

This study presents a comprehensive, up-to-date and flexible framework that can serve various research areas, such as machine learning-based predictions, testing algorithmic trading strategies, investigating correlations between the blockchain and price fluctuations, market analysis or anomaly detection. Studies in the literature in this area have used a wide range of methods, —from statistical econometric models, machine learning and deep learning-based methods to complex hybrid models that incorporate on-chain metrics. The studies in the literature can generally be categorized along three main axes: traditional statistical and econometric approaches, machine learning and deep learning-based models, and advanced methods that integrate on-chain data, social media sentiment analysis, and high-frequency datasets. These studies will be discussed in the context of the time periods and frequencies they use.

Much of the early research aims to capture Bitcoin price dynamics through time series analysis and econometric models. For example, in [10], the ARIMA model was compared with autoregressive and moving average models using daily data from 2013 to 2017, demonstrating that the ARIMA model produced more consistent results for Bitcoin price prediction. The ARIMA model was applied to 5-minute interval data in [11].

A comparison of LSTM and RNN models using daily data showed that while LSTM offers higher accuracy, its training time is significantly longer [12]. In addition, a neuro-fuzzy controller called PATSOS was developed, successfully predicting the daily price movement of Bitcoin [13]. Using hourly data, a study focused on price prediction by employing a bidirectional LSTM combined with sentiment analysis of Twitter data [14].

In recent years, the use of new data sources in the prediction of cryptocurrency prices has come to the fore. The direct integration of blockchain data into price prediction models has helped to close gaps in this area. For example, through "k-chainlet" analysis of Bitcoin transaction graphs, it has been shown that certain topological structures offer insights into price dynamics [15]. Volatility and order book data, in turn, were combined with temporal mixture models to predict Bitcoin price changes with high accuracy [16].

A general review of the studies in the literature reveals that the majority use daily course data. A major shortcoming in the current literature is that most studies focus on the period prior to 2017 and on daily data. This has led to gaps in understanding the real-time and high-resolution dynamics of highly volatile cryptocurrency markets. Moreover, the integrated study of on-chain metrics with price movements has only been addressed by a limited number of studies. In this context, the framework we present provides an up-to-date and flexible infrastructure by converting data into block intervals at desired time intervals and frequencies, enabling the analysis of price movements at both macro and micro levels. This alleviates the lack of data in the literature and provides an easy-to-use environment for statistical econometric models, machine learning and deep learning based models, and on-chain metrics analysis. High-frequency and up-to-date datasets can better reflect real-time market conditions, improving model performance and enabling new research questions.

In the studies found in the literature, the data was generally obtained by using data repositories that share historical Bitcoin data. For example, a Kaggle dataset provides data up to 2021 separately in Excel format for time periods of 1 minute, 1 hour and 1 day [17]. However, it is not possible to work with this dataset with, for example, 5- or 10-minute intervals. Researchers working in this area can conduct studies with the appropriate intervals if such data is available in the APIs provided by the cryptocurrency exchanges, or otherwise they have to create them themselves for each time period and interval by using their knowledge of data conversion processes.

Our proposed framework eliminates this problem for researchers by providing an infrastructure that facilitates access to data by extending it from a 1-minute interval to any desired interval.

3 Background

3.1 Fundamental bitcoin concepts

Bitcoin was launched in 2008 by an individual or group under the pseudonym Satoshi Nakamoto and is a decentralized, peer-to-peer digital currency. This digital asset is based on blockchain technology, which enables a direct transfer of value between users without relying on a central authority. The blockchain provides a distributed ledger structure that ensures the trust and integrity of data without the need for third parties to act as intermediaries.

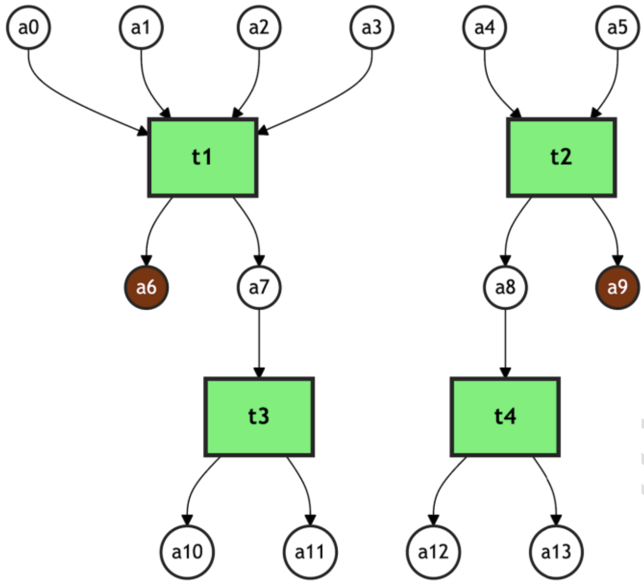

Figure 1: Transaction-address representation in the Bitcoin network

Each "block" contains the transactions that have been authorized within a certain period of time. These blocks are cryptographically interlinked and form a data record structure that is very difficult and costly to change retrospectively. The relationship between transactions and addresses in the Bitcoin network is shown schematically in Figure 1. In Figure 1, the addresses are shown as circles and the transactions as rectangles. The edges show coin transfers. Here there has been no transaction movement in the coin at addresses a6 and a9, so this coin was never spent.

As is illustrated in Figure 1, within a block that flows from left to right in time, the relationship between transactions and addresses is represented as a coin transfer. Although the average block production time on the Bitcoin network is fixed at around 10 minutes, this duration can fluctuate slightly due to changes in mining difficulty and network hash power. To make time measurement more consistent within the network, the Bitcoin protocol typically uses the timestamps of the last 11 blocks to create a metric called "median time" This measurement helps to more reliably reflect the chronological order of blocks.

3.2 Price movements and candlestick charts

One of the fundamental tools used in interpreting price movements in financial markets is the candlestick chart. Each candlestick corresponds to a certain time interval (e.g., 1 minute, 5 minutes, 1 hour, or 1 day) and contains four key price points from that period: the opening (open), closing (close), highest (high), and lowest (low). Candlestick charts help investors and analysts understand the market's direction, volatility, trends, and possible turning points. Especially in cryptocurrency markets, candlestick charts on time frames ranging from hourly down to the minute provide the opportunity for shorter-term and more detailed analyses.

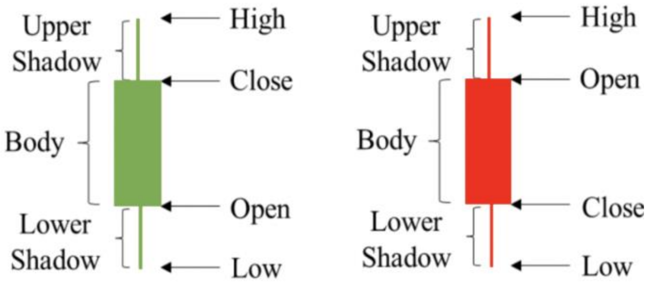

Figure 2: Candlestick chart representation [18]

As is illustrated in Figure 2, the region between the opening and closing price is called the "body" in candlestick charts and is usually shown in green (bullish) or red (bearish). The thin lines extending from the top and bottom of the body are called "shadows" or "wicks" and indicate the highest and lowest prices observed during that time interval. This visual structure not only provides valuable insight into price movements, but also into the balance of power between buyers and sellers in the market. For example, long shadows indicate high volatility, while short shadows indicate that the price has stabilized around a certain level.

3.3 The importance of high-frequency data

Compared to traditional financial markets, cryptocurrency markets are much faster, more volatile and open around the clock. This means that market conditions can change in real time. Using daily data can be insufficient to capture rapid market movements. High frequency data, collected at shorter intervals (minutes or even seconds), allows price fluctuations to be analyzed at a micro level. This in turn allows investors to develop algorithmic trading strategies, understand the microstructure of the market, detect anomalies or gain deeper insights for volatility modeling. High frequency data is also crucial for studying the relationship between blockchain data and prices, as correlations between on-chain metrics such as block creation times, transaction intensity and difficulty adjustments with price movements can become clearer on shorter time scales.

4 Proposed framework

In this section, the proposed framework is described in detail so that it can be easily used by researchers. The raw data coming from different sources are transformed into a current, consistent and integrated dataset for the desired time period and interval, suitable for the analyzes intended by the researchers. For this purpose, an 8-step detailed framework has been developed. Thus, researchers using the proposed framework can easily repeat, extend or customize the data preparation process for different datasets. In the first stage, the raw data is collected from web sources in the appropriate file formats and transferred into a regular database structure. Then, to synchronize with the blockchain data, high-frequency bitcoin price data is collected down to the minute and converted into parameter-based intervals by timestamping.

The processes of systematically collecting, cleaning, transforming and transferring Bitcoin block and price data into a database structure over a broad period from 2017 to 2024 are discussed in detail.

4.1 Data sources, downloading, and formats

4.1.1 Block data

Researchers wishing to access Bitcoin block data would normally need to set up a full node by downloading and synchronizing the entire blockchain. Setting up a full Bitcoin node involves significant technical requirements, as it requires downloading and verifying the entire extensive blockchain data. In this context, a powerful processor, sufficient RAM, fast storage units and a stable internet connection with high bandwidth are of the utmost importance. In addition, the system must run continuously over a long period of time to complete the initial synchronization process, which requires a significant amount of time and effort. In addition, cyber security measures, regular software updates and legal restrictions in certain regions make this process even more complex. Considering the potential additional costs and restrictions from ISPs due to high data transfer volumes, running a full node seems both cumbersome and costly.

To simplify this challenging and costly process, the proposed study uses online resources that regularly provide raw block data in tab-separated values (TSV) format. the .tsv.gz files for each day or for specific time intervals are efficiently retrieved using Python scripts and multithreaded download methods and then decompressed into the raw .tsv format. In this way, the data is structured so that each line corresponds to a block. This allows researchers to access block information without having to download gigabytes of data or set up an entire node.

The block data contains numerous columns that reflect the structure of a Bitcoin block, its economic activity and details of the mining process. These can be listed as hash, time, median_time, size, stripped_size, weight, version, version_hex, version_bits, Merkle root, nonce, bits, difficulty, transaction_count, witness_count, input_count, output_count, input_total, output_total, fee_total, reward and guessed_miner.

4.1.2 Price data

You can obtain historical Bitcoin price data in minute intervals as .csv table data sets from cryptocurrency exchanges or from platforms such as Kaggle. These data sets usually contain the timestamp (date), price information (opening, high, low and closing price), the symbol (e.g. BTCUSD) and the trading volume (volume_btc, volume_usd). In this way, price movements can be analyzed over a wide period of time with a resolution in minute intervals. When we examine the datasets here, we find that they contain price data up to the time they were published. For this reason, the source code (i.e. the framework) to obtain the most recent data is not shared with researchers who wish to use the dataset. Therefore, if researchers wish to re-obtain and review the data themselves, they will not be able to do so.

The price data includes a 1-minute-based timestamp (date), a symbol (e.g. BTCUSD), open/close/high/low prices (open, high, low, close) and trading volume (volume_btc, volume_usd). You can use these columns to analyze the trends in price movements, volatility and volume changes in detail.

4.2 Proposed framework and workflow

In this study, data collection and preprocessing processes are carried out in 8 clearly defined steps:

- Step-1: Downloading Block Data

  The block data is retrieved from the Internet in .tsv.gz format using Python and multithreaded downloader scripts.

- Step-2: Unpacking the Data

  The downloaded .tsv.gz files are unpacked into the raw .tsv format and are thus ready for use. The first lines of the latest .tsv files are checked to ensure that the correct data has been received.

- Step-3: Transferring from TSV to SQLite

  The block data is transferred to an SQLite database engine and the table structure is defined.

- Step-4: Creating an API (Block Data)

  An API based on Flask enables access to block data within certain time intervals.

- Step-5: Transferring Price Data

  1-minute-based Bitcoin price data in .csv format is transferred to the SQLite database engine. Queries are performed over specific date ranges to test the accuracy and integrity of the price data. Additionally, in cases of API limitations or when the primary exchange data provider is inaccessible, the system automatically retrieves price data from alternative exchange sources for the corresponding timestamps, ensuring continuous and reliable data coverage.

- Step-6: API Access to Price Data

  The Flask API allows users to retrieve price data in JSON format for a specific time period.

- Step-7: Time Series Conversion (1 min → n min)

  The data obtained in 1-minute intervals in the previous step is aggregated into n-minute intervals as required. The 1-minute-based price data is converted into the desired n-minute bars and synchronized with the block time. This allows the time scale of the data set to be adjusted depending on the scope and size of the analysis to be performed.

- Step-8: API Access to n-Minute Data

  You can access the n-minute price data via the API.

This structural approach makes all steps from the acquisition of data at the source to its final form before analysis transparent and repeatable. One of the original contributions of the developed system is that the basic data structure is based on price bars at 1-minute intervals, which allows dynamic aggregation into any n-minute interval specified by the user. This flexible time scale transformation mechanism, enabled by the transformation process described in Step 7, allows the raw 1-minute interval dataset to be systematically and flexibly adapted to different time resolutions (e.g., 60 minutes, 120 minutes, 240 minutes, etc.). This approach gives researchers the opportunity to analyze the same data set at different time intervals, allowing for a multidimensional investigation of the dynamics in the cryptocurrency markets.

A key advantage of this methodological innovation is that the time scale transformation mechanism required to analyze high frequency trading data and blockchain metrics simultaneously is available. As a result, researchers can comparatively study market behavior in current time periods with desired time intervals by choosing temporal resolutions suitable for specific research questions and optimizing the size of the dataset according to analytical requirements, thereby improving the effectiveness of multi-scale analysis on cryptocurrency markets. This method increases both the efficiency of the analytical studies and the relevance of the results obtained to the time frame under investigation.

The framework developed is suitable for regularly updating the data set, adding new data sources and performing analyzes at different time periods and intervals. This process, automated with Python scripts and shell commands, offers researchers speed, flexibility and the elimination of repetitive work. In addition, the Flask-based APIs that have been created facilitate remote access to the data, sharing and reuse in other research projects. This framework can also be extended and customized for other cryptocurrencies such as Ethereum. The dataset can be enriched not only with price and block data, but also by integrating other metrics such as network difficulty, hash rate and on-chain transaction counts. Despite occasional interruptions in data sources, API limitations or website access issues, the steps to populate missing data and the automation processes ensure that the quality of the dataset is maintained. Alternative sources or new metrics can be added to the framework to expand the scope of the dataset.

The source code, detailed documentation, and instructions for using the proposed framework are publicly available on GitHub at https://github.com/vtataroglu/PhDWork Researchers can easily access and integrate the provided Python scripts into their own analysis environments.

5 Case study

In this section, data sets with different time intervals n are generated using the proposed framework. In Figure 3, n=10 was chosen to demonstrate a one-week period of Bitcoin prices. A short time interval with high data density is suitable for microstructure analysis of price movements and short-term volatility. In Figure 4, however, n=60 was chosen. Medium length bars preserve some short-term details while providing sufficient granularity for trend detection and technical analysis. In Figures 5 and 6, n=240 and n=1440 are chosen respectively. With a longer time interval, the short-term fluctuations decrease and the medium-term trends become clearer. In Figure 6, with a resolution of only eight days with eight bars, detailed technical analysis and pattern recognition become a greater challenge.

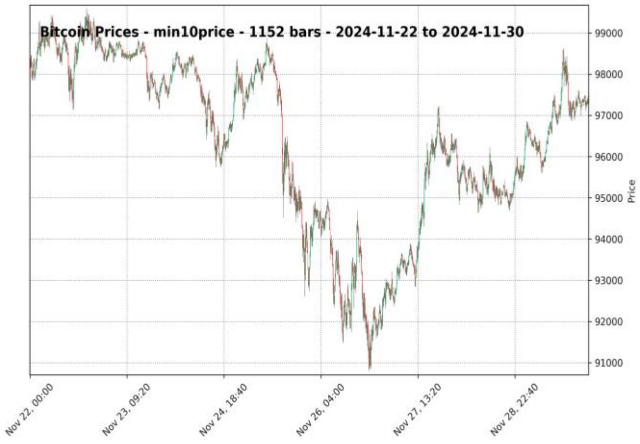

Figure 3: Representation with 10-minute price bars for 2024/11/22 – 2024/11/30

Figure 4: Representation with 60-minute price bars for 2024/11/22 – 2024/11/30

Figure 5: Representation with 240-minute price bars for 2024/11/22 – 2024/11/30

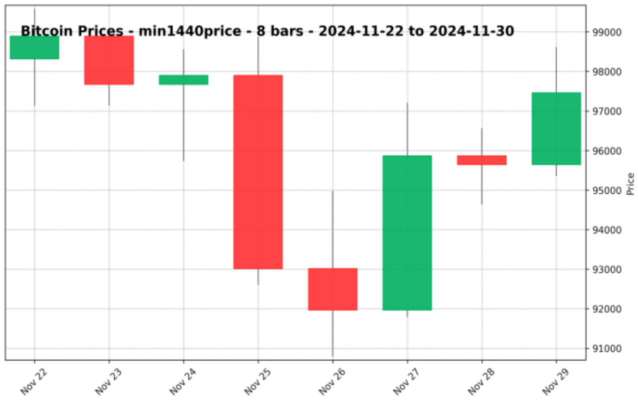

Figure 6: Representation with 1440-minute (daily) price bars for 2024/11/22 – 2024/11/30

For example, if you use daily (1440-minute) bars in Figure 6, the eight period in question is represented by only eight data points, which significantly limits detailed technical analysis, pattern recognition and identification of candlestick chart formations. In contrast, looking at the same time frame with 10-minute bars in Figure 3 provides a much larger number of data points, allowing for a more detailed analysis of the microstructure of price movements, momentary volatility changes and short-term trends. Similarly, Figures 4 and 5 provide a higher density of data than daily data through 60-minute and 240-minute bars respectively, allowing for more comprehensive execution of applications such as technical analysis, image processing and formation detection. This comparative approach highlights the critical impact of the choice of time period and interval scale on data granularity and the quality of potential analytical insights. While short time intervals enable in-depth analysis with high-frequency data, the decreasing data density of longer time frames offers a trend-oriented but more limited perspective in terms of detail.

Table 1: Basic statistical values

Metric	Value
Total Blocks Examined	382536
Average Time	0:09:44
Median Time	0:06:46
Maximum Time	2:19:14

The timestamps of Bitcoin blocks can have slight deviations, as they depend on the system clocks of the various miners in a decentralized network. When a new block is added, the protocol checks whether its timestamp is greater than the average timestamp of the previous 11 blocks and whether it is within two hours of the network-adjusted time. This sometimes results in a block appearing to have been created “before” the previous block (i.e. a negative time difference), even though it is still valid according to the Bitcoin rules. For example, a miner’s local clock could be slightly behind or ahead, or there could be short delays in transmitting the block to the network. Since these negative intervals are usually only a few seconds, they do not jeopardize the integrity or security of the chain. Bitcoin’s fundamental security is based on proof-of-work and the consensus of nodes following the chain with the highest cumulative work, meaning that minor timestamp discrepancies have minimal impact on the system.

Table 2: Top 3 longest block time differences (example)

Rank	Block1	Block2	Difference
1	689300	689301	2:19:14
2	679785	679786	2:02:08
3	824717	824718	2:01:43

Table 1 and Table 2 are based on data obtained by analyzing blocks from block 446045 to block 828580, corresponding to the years 2017-2024. As can be seen in Table 1, the average block time is 9 minutes and 44 seconds, so we can consider 10-minute time intervals as a suitable basis for the analysis. In this way, both the usual block production rate and potential anomalies can be clearly examined, allowing an assessment of the consistency of delay between blocks across the system. Table 2 shows the longest time differences between three blocks.

In this study, by choosing *n* = 10 and converting 1-minute price data into 10-minute time intervals, it is possible to create a data structure that better reflects the Bitcoin block production cycle. Since Bitcoin blocks are created approximately every 10 minutes, reducing price movements to these time windows allows on-chain metrics and price dynamics to be analyzed simultaneously. This approach captures the highest, lowest, open and close price in each 10-minute period so that the price series can be summarized to match the rhythm of block production. This makes it possible to examine events in the blockchain (e.g. successful mining of a new block, adjustments to mining difficulty or changes in transaction volume) and price fluctuations in a more direct, consistent and comparable way. This contributes to a deeper understanding of market trends, volatility patterns and the relationship between blockchain metrics and price.

6 Conclusion

This study presents a framework for obtaining current Bitcoin price data for specific dates and time periods, allowing researchers to gain a deeper understanding of cryptocurrency markets. Processes that normally require tedious tasks such as setting up a full Bitcoin node and downloading the entire blockchain from scratch have been greatly simplified by obtaining raw data from online sources, allowing researchers to quickly access information about blocks, transactions, mining and prices.

One of the main contributions of the study is the development and documentation of an 8-step framework. The proposed framework automates the entire process — from downloading the block data, unpacking the compressed files, transferring the data to a SQLite database, adding missing data points, and converting price data at 1-minute intervals to bars of a desired interval. This allows researchers to update the dataset as often as they like, easily switch between different time intervals or add new metrics to the dataset.

The resulting dataset is not only used for price forecasting, but is also a valuable resource for testing algorithmic trading strategies, performing detailed market analysis, modeling volatility, investigating correlations between on-chain metrics and prices, detecting anomalies, and many other areas. In addition, the dataset and the API services developed allow researchers to access up-to-date high-frequency data without having to set up a full node on their own computer.

To summarize, this study fills the need for current and high-resolution data in the literature, contributing to a deeper investigation of the relationships between Bitcoin's blockchain data and price movements. The high-frequency, up-to-date, and flexible dataset offered here will provide a valuable foundation for many different application areas, including machine learning-based price prediction models, testing algorithmic trading strategies, investigating correlations between blockchain events and price fluctuations, analyzing market microstructure, and anomaly detection. This will provide a robust infrastructure for both large-scale historical analysis and near real-time predictive studies. This infrastructure will contribute to a wide range of next-generation analysis, modeling and policy development, ranging from academic research to industrial applications.

In the future, applying similar approaches to Ethereum or other cryptocurrencies will enable the exploration of relationships between on-chain data and prices across a broader spectrum. By integrating additional metrics from different data sources (such as hash rate, difficulty, transaction fees, number of active on-chain addresses), the dataset can be further enriched. As cryptocurrency markets become increasingly complex, it may consequently become possible to develop more robust economic models, more accurate predictions and more effective investment strategies.

7 Author contribution statement

In the scope of this study, Author-1 was actively involved in the conception and development of the idea, the design of the study, and contributed significantly to the spelling, editing, and content verification of the article. Additionally, Author-1 participated in the literature review, data preparation, and the analysis and interpretation of the results. Author-2 contributed to the evaluation of the results, conducted a thorough literature review, and assisted in the preparation and examination of the data. Author-3 played a key role in the literature review, the analysis of the results, and the critical review of the article's content for accuracy and coherence. Author-4 was instrumental in shaping the research idea, contributing to the study design, and ensuring the article was thoroughly reviewed and checked for content quality and clarity. All authors reviewed and approved the final manuscript.

8 Ethics committee approval and conflict of interest statement

"This study does not require ethics committee approval". "The authors declare no conflicts of interest related to this work".

9 References

[1] Glas T. Digital Assets. In: Asset Pricing and Investment Styles in Digital Assets. Advanced Studies in Diginomics and Digitalization, Cham, Switzerland, Springer, 2022. (https://doi.org/10.1007/978-3-030-95695-0_5)

[2] Van der Laan J. Understanding Blockchain. Editors: Artzt M, Richter T. International Handbook of Blockchain Law: A Guide to Navigating Legal and Regulatory Challenges of Blockchain Technology and Crypto Assets, Holland, Wolters Kluwer Press, 2024.

[3] Ducas E, Wilner A. "The security and financial implications of blockchain technologies: Regulating emerging technologies in Canada". International Journal: Canada's Journal of Global Policy Analysis, 72(4), 538–562, 2017. (https://doi.org/10.1177/0020702017741909)

[4] Narayanan A, Bonneau J, Felten E, Miller A, Goldfeder S. Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction. Princeton, USA, Princeton University Press, 2016.

[5] Gao Q, Zhang J. Blockchain and the Transformation of Global Economic Governance. In: Artificial Intelligence Governance and the Blockchain Revolution. Artificial Intelligence and the Rule of Law, Singapore, Springer, 2024. (https://doi.org/10.1007/978-981-99-9211-9_4)

[6] Ghimire S, Selvaraj H. "A survey on Bitcoin cryptocurrency and its mining". 2018 26th International Conference on Systems Engineering (ICSEng), Sydney, NSW, Australia, 1–6, 2018. (https://doi.org/10.1109/ICSENG.2018.8638208)

[7] Adjei F. "Determinants of Bitcoin expected returns". Journal of Finance and Economics, 7(1), 42-47 , 2019.

[8] Arslanian H. Bitcoin. In: The Book of Crypto, Cham, Switzerland, Palgrave Macmillan, 2022. (https://doi.org/10.1007/978-3-030-97951-5_2)

[9] Li J, Zhang C, Zhang J, Shao Y. "Research on blockchain transaction privacy protection methods based on deep learning". Future Internet, 16(4), 113, 2024.

[10] Roy S, Nanjiba S, Chakrabarty A. "Bitcoin price forecasting using time series analysis". 21st IEEE International Conference of Computer and Information Technology (ICCIT), 1–5, 2018.

[11] Ibrahim A, Kashef R, Corrigan L. "Predicting market movement direction for bitcoin: a comparison of time series modeling methods". Computers & Electrical Engineering, 89, 106905, 2021.

[12] McNally S, Roche J, Caton S. "Predicting the price of bitcoin using machine learning". 26th IEEE Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), 339–343, 2018.

[13] Atsalakis GS, Atsalaki IG, Pasiouras F, Zopounidis C. "Bitcoin price forecasting with neuro-fuzzy techniques". European Journal of Operational Research, 276(2), 770–780, 2019.

[14] Patel R, Chauhan J, Tiwari NK, et al. "A deep learning framework for hourly Bitcoin price prediction using bi-LSTM and sentiment analysis of Twitter data". SN Computer Science, 5, 767, 2024. (https://doi.org/10.1007/s42979-024-03112-9)

[15] Akcora CG, Dey AK, Gel YR, Kantarcioglu M. Forecasting Bitcoin Price with Graph Chainlets. Editors: Phung D, Tseng V, Webb G, Ho B, Ganji M, Rashidi L. Advances in Knowledge Discovery and Data Mining. PAKDD 2018 (Lecture Notes in Computer Science, 10939), 765–776, Cham, Switzerland, Springer, 2018. (https://doi.org/10.1007/978-3-319-93040-4_60)

[16] Guo T, Antulov-Fantulin N. "Predicting short-term Bitcoin price fluctuations from buy and sell orders". arXiv:1802.04065, 2018.

[17] Kaggle. "BTCinUSD dataset". https://www.kaggle.com/ datasets/prasoonkottarathil/btcinusd (18.12.2024).

[18] Chen X, Hu W, Xue L. "Stock price prediction using candlestick patterns and sparrow search algorithm". Electronics, 13(4), 771, 2024. (https://doi.org/10.3390/electronics13040771)