MEDIPOL
UNV-ANKARA
ANKARA MEDİPOL ÜNİVERSİTESİ

FINAL REPORT

Project Title: Analysis of Trends in the NFT Market: Valuable NFTs, Price Factors, and Wash Trading

Group Members: Aslı AYAZ, Zehra Bengisu DOĞAN, Zeynep ÇALAPKULU

Submission Date: 02/06/2025

Abstract

This study presents a comprehensive data-driven analysis of the Non-Fungible Token (NFT) market during its formative years, focusing on transaction data from the OpenSea platform between 2019 and 2021. A multidimensional methodological framework was applied, combining time series analysis, correlation and regression modeling, unsupervised clustering, frequent pattern mining, and wash trading detection. The research investigates temporal market trends, identifies weak and strong price determinants, and classifies high-value NFTs based on shared attributes. Additionally, manipulative trading behaviors were examined through custom anomaly detection techniques. The results reveal the layered and heterogeneous structure of the NFT ecosystem, where asset value emerges from a combination of platform-based utility, collector preferences, and economic integration rather than simple numeric factors. This work provides an analytical foundation for understanding digital ownership, valuation mechanisms, and potential risks in emerging blockchain economies.

Keywords

NFT market, time series analysis, price correlation analysis, clustering, FP-Growth, wash trading, digital assets, blockchain analytics, OpenSea, unsupervised learning

Table of Contents

1.Introduction

1.1 What is an NFT? A Brief Definition and Market Context

With the advancement of blockchain technology, there has been a revolutionary shift in the representation of digital assets. One of the most significant outcomes of this transformation is the concept of Non-Fungible Tokens (NFTs). NFTs are digital assets uniquely defined on the blockchain as non-fungible, immutable, and indivisible. Each NFT possesses unique metadata and a digital signature, which makes it impossible to exchange one for another on a one-to-one basis [1].

NFTs are used to verify ownership of a wide range of digital content—from digital artworks and collectible items to virtual in-game assets and music files. Typically developed and traded on blockchain networks with smart contract capabilities, such as Ethereum, NFTs have gained immense popularity among investors and content creators. As a result, the NFT market has reached significant volume, with transactions worth billions of dollars being recorded [2].

The NFT market experienced a major boom in 2021, drawing widespread media attention as digital artworks sold for millions of dollars. In this sense, NFTs have evolved beyond mere digital ownership certificates to become a cultural, social, and economic phenomenon [3].

1.2 Key Factors Affecting NFT Prices

Unlike traditional financial assets, the prices of NFTs are influenced not only by supply and demand dynamics but also by multidimensional factors such as social interactions, rarity, collection identity, ownership history, and time-based trends.

For instance, factors like the popularity of the collection an NFT belongs to (e.g., Bored Ape Yacht Club), the number and type of rare attributes it possesses (such as unique backgrounds or accessories), the artist's sales history, social media attention, and the trading history of the NFT (e.g., which wallets previously owned it) all play a significant role in determining its price [4].

In addition, seasonal and cyclical trends that emerge over time in the NFT market can also directly influence price movements. Moreover, manipulative practices such as wash trading—where users conduct fake transactions to influence prices—can also lead to significant price fluctuations [5].

1.3 Project Objective and Scope

The main objective of this project is to understand the dynamics of the NFT (Non-Fungible Token) market, analyze the factors that affect price fluctuations, and detect possible anomalies. NFTs are blockchain-based digital assets that guarantee uniqueness and have recently attracted significant attention from both investors and artists. However, the novelty and susceptibility of this market to manipulation highlight the need for in-depth data analysis.

Within this scope, the project aims to perform a multidimensional analysis by combining various data science approaches, including time series analysis, statistical modeling, unsupervised learning, and network theory. Key areas of analysis throughout the project include time-based changes in NFT prices, influential attributes, common features of the most valuable NFTs, and the detection of manipulative behaviors such as wash trading. The project also covers the examination of market volatility and the identification of seasonal patterns.

1.4 Dataset and Tools Used

The dataset used in this study is titled "OpenSea NFT Sales 2019–2021", obtained from the Kaggle platform [6]. It contains transaction details of NFT sales conducted on the OpenSea platform between

2019 and 2021. The dataset includes information such as sale dates, price details, NFT identifiers, collection names, and buyer and seller addresses.

The analysis process was carried out using the Python programming language, utilizing powerful open-source libraries for data processing, statistical modeling, visualization, and network analysis. Core libraries like Pandas and NumPy were used during the data cleaning, transformation, and analysis phases. Regression analysis and time series modeling were performed with the Statsmodels library, while visual insights were supported by Seaborn and Matplotlib.

Furthermore, to analyze interactions between NFT buyers and sellers, graph-based structures were created using the NetworkX library. These graphs were used to examine connections between addresses, transaction intensities, and potential manipulation patterns. Rolling window techniques were applied to analyze time-based changes in market behavior. NFTs with similar attributes were classified using clustering algorithms such as K-means, and frequent attribute combinations were identified using the FP-Growth method.

This comprehensive analysis framework, combining various statistical and machine learning-based approaches, enabled the extraction of in-depth insights into the NFT market.

1.5 Overview of Report Structure

This report is composed of twelve main sections. In the second section, titled *Dataset Description*, the data source, the structure of the dataset, and the preprocessing steps are discussed in detail. The fourth section, *Time Series Analysis*, explores general trends in the NFT market using time series analysis techniques. In the fifth section, *Identifying Factors Affecting NFT Prices*, the attributes that influence NFT prices are identified and analyzed through various regression models. The sixth section, *Examining the Most Valuable NFTs in Collections*, focuses on grouping NFTs based on their values and extracting frequently observed feature patterns.

The seventh section, *Detecting Wash Trading*, attempts to identify suspicious transaction patterns. The eighth section, *Analyzing Seasonal and Periodic Changes in the NFT Market*, analyzes seasonal effects and cyclical behavior patterns in the market. The ninth section, *Examining NFT Price Volatility*, measures market volatility and presents these changes through visualizations. The tenth and eleventh sections, *Results and Discussion* and *Conclusion*, evaluate the findings and present general insights. The twelfth section, *References*, lists the sources used in the study, while the thirteenth and final section, *Appendix*, provides additional charts, tables, and code snippets.

2.Dataset Description

In this study, the 'OpenSea NFT Sales 2019–2021' dataset [6] is employed to conduct a broad analysis of NFT market activity, including user behavior, transaction patterns, and indicators of potential wash trading**.** This dataset provides a comprehensive view of all successful sales on the OpenSea platform between January 2019 and December 2021, offering rich insights into pricing dynamics, trade frequency, NFT circulation, and buyer–seller interactions. It comprises approximately 5.25 million transaction records, totaling 1.69 GB of data, and is sourced from the official OpenSea API's Events endpoint.

This extensive dataset includes 15 attributes, capturing both transactional details and descriptive metadata related to NFTs and user interactions:

- **sales_datetime**: The timestamp indicating when the NFT sale occurred, expressed in standard date-time format.

- **id**: A unique identifier for each individual transaction record.
- **asset.id**: The unique identifier assigned to the NFT asset being traded.
- **asset.name**: The name of the NFT, often provided by the creator or collection.
- **asset.collection.name**: The name of the NFT collection to which the asset belongs (e.g., Bored Ape Yacht Club).
- **asset.collection.short_description**: A brief textual description of the asset or collection. However, this field is often missing or contains very minimal information such as color, texture, or style.
- **asset.permalink**: A hyperlink leading to the NFT's asset page on OpenSea. This field is frequently null or invalid.
- **total_price**: The sale price of the NFT in the native token unit (e.g., ETH), recorded in wei, where 1 ETH = 10^{18} wei.
- **payment_token.name**: The name of the cryptocurrency or token used in the transaction (e.g., Ethereum, WETH, USDC).
- **payment_token.usd_price**: The USD equivalent of the token's value at the time of transaction, used to estimate the trade's fiat value.
- **asset.num_sales**: The number of previous sales involving this specific NFT asset, excluding the current transaction.
- **seller.address**: The blockchain wallet address of the seller initiating the transaction.
- **seller.user.username**: The username of the seller, if available. In many cases, this is either null or reflects a pseudonym.
- **winner_account.address**: The blockchain wallet address of the buyer or winning bidder.
- **Category**: The category assigned to the NFT (e.g., Art, Collectible, Game Asset), derived through a combination of web scraping and manual labeling.

These attributes collectively provide a rich set of features that enable multi-dimensional analysis of trading behaviors, pricing anomalies, and participant identities across millions of NFT transactions.

3. Data Preprocessing and Cleaning

Before conducting any meaningful analysis, it is essential to clean and preprocess the dataset to ensure reliability and consistency. This involves identifying and handling missing values, removing irrelevant or non-informative attributes, and standardizing data formats to facilitate accurate downstream computations. Given the nature and scale of blockchain-based transaction data, preprocessing plays a critical role in minimizing noise and enhancing the quality of insights derived from the dataset.

3.1 Data Loading and Initial Cleaning

In the initial phase, the dataset titled OpenSea_NFT_Sales_2019_2021.csv was loaded, and duplicated records were eliminated to ensure data consistency. Given the nature of blockchain transactions and decentralized marketplaces, repeated entries can distort statistical inferences, especially in high-frequency trading scenarios.

Furthermore, in the 'payment_token.name' column, there are many different token types, but the vast majority of these tokens have very low transaction counts. For the sake of analysis reliability and interpretability, it is important to minimize the impact of rarely used tokens. Therefore, during data preprocessing, tokens with fewer than 50 transactions—corresponding to a percentage range between approximately 0.000019% and 0.000838% of the total transactions—were excluded from the analysis. This threshold was chosen to balance between retaining meaningful token varieties and removing those with negligible representation, resulting in a more manageable dataset and more robust, consistent analytical outcomes.

3.2 Handling Missing and Inconsistent Values

In order to ensure the quality and reliability of the analysis, it is crucial to identify and address missing or inconsistent data within the dataset. Several fields exhibited such issues. The following chart displays the count of null values for each attribute:

	Data Type	Missing Value Count
sales_datetime	object	0
id	int64	0
asset.id	int64	0
asset.name	object	305787
asset.collection.name	object	48785
asset.collection.short_description	object	5200602
asset.permalink	object	48783
total_price	object	0
payment_token.name	object	1164
payment_token.usd_price	float64	3795
asset.num_sales	int64	0
seller.address	object	0
seller.user.username	object	584833
winner_account.address	object	0
Category	object	0

Fig. 3.1. Number of Null values in each column

- **Removing Columns with High Null Rates and Limited Analytical Value**: As shown in the figure above, several attributes in the dataset contain missing values. Among these, the ‘asset.collection.short_description’ and ‘asset.permalink’ columns were removed entirely. The ‘asset.collection.short_description’ column has over 5.2 million null entries and consists of vague or one-word descriptions (e.g., "blue", "rough", "test") that lack analytical value. The ‘asset.permalink’ column, although intended to provide links to NFT assets, often contains invalid or empty URLs, limiting its usefulness for systematic analysis. Removing these columns helps streamline the dataset without sacrificing relevant information.
- **NFT Name and Collection Filling:** Missing asset.name values were imputed using the corresponding collection name when available. If both were missing, a generic placeholder like *"Unnamed Asset"* was used. Similarly, if the collection name was missing but the NFT name existed, it was back-filled using a derived expression (*e.g.*, “Collection of [NFT Name]”).
- **Token Name Imputation and Standardization:** Missing payment_token.name values were filled with the most frequently occurring token in the dataset. Then, token names were

standardized via mapping (e.g., “Ether” → “ETH”, “Wrapped Ether” → “WETH”) to address typographical inconsistencies that could affect grouping operations.

- **Conversion to USD**: Prices in total_price were converted from Wei (Ethereum’s smallest unit) to ETH by dividing by 10^{18}. USD prices were then calculated using an external price dictionary sourced from token_prices.csv, ensuring a consistent monetary scale across different cryptocurrencies. Entries lacking USD values were removed to avoid biased analysis in price-related sections.

3.3 User Identity Masking and Cleaning

- To preserve privacy while maintaining user-level tracking capabilities, seller and buyer addresses were masked using the first five characters of their hexadecimal address (e.g., “Seller_0x123”). If usernames were missing, they were inferred from address-based mappings or generated using a unique placeholder.

Finally, the cleaned dataset was exported to cleaned_dataset.csv and used for all subsequent analyses.

4. Time Series Analysis

4.1. Definition and Importance

Time series analysis is a statistical technique that deals with time-ordered data points. It is widely used to understand underlying patterns, detect trends, and forecast future behavior in temporal datasets. In the context of the NFT market, time series analysis enables the examination of transaction volumes, price fluctuations, and other market dynamics over time. It helps identify long-term patterns as well as short-term fluctuations which are crucial for understanding market sentiment and volatility.

A time series can be decomposed into four main components: trend, seasonality, cyclic behavior, and random noise [7] .The objective of this section is to reveal the underlying structure of NFT price dynamics using basic yet effective time series tools such as moving averages and trend analysis.

“Time series data arise naturally in many application areas. They are used in forecasting, financial analysis, and monitoring changes over time.” [8]
— Chatfield, C. (2003). *The Analysis of Time Series: An Introduction.*

4.2. Initial Analysis

To provide a clearer understanding of the NFT market’s historical dynamics, this section presents several exploratory visualizations. These include the weekly NFT sales count, the monthly average NFT price over time (with and without outliers), a dedicated outlier detection plot, and a timeline of key market events mapped against monthly average sales volumes. Together, these visual aids help uncover early patterns, anomalies, and structural shifts in the market.

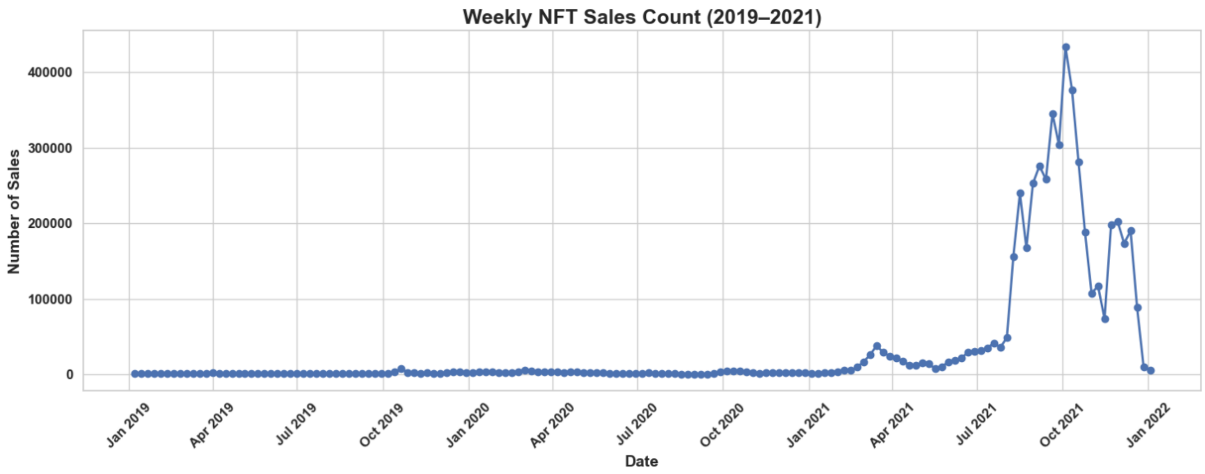

Fig. 4.1. Weekly NFT Sales Count (2019-2022)

The provided graphs depict the evolution of weekly NFT sales counts from January 2019 to January 2022, revealing a clear transition from a niche market to a mainstream phenomenon. To better understand the gradual yet transformative growth, the data has been divided into two distinct periods: 2019–2020 and 2021–2022. This segmentation highlights the contrasting dynamics between the early, experimental phase and the later, explosive expansion of the NFT market.

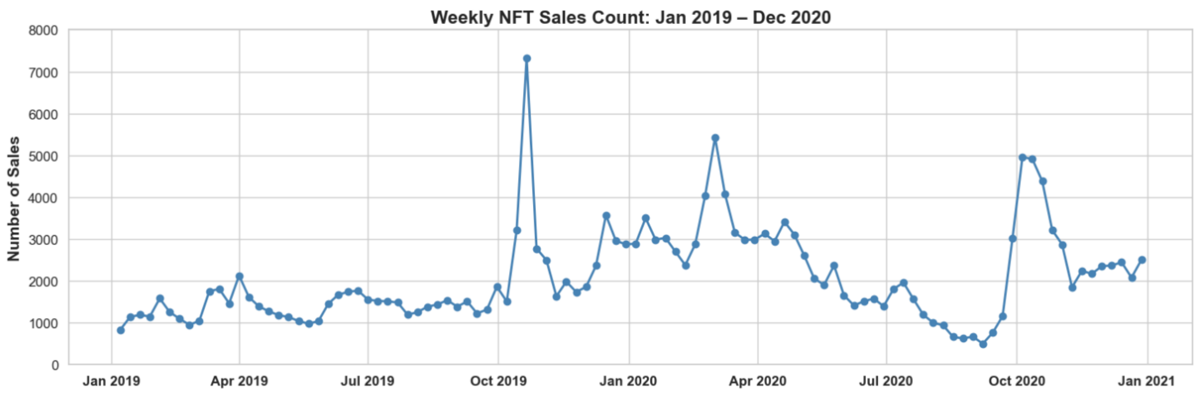

Fig. 4.2. Weekly NFT Sales Count (2019-2020)

The weekly NFT sales data from January 2019 to December 2020 reveals a volatile yet upward-trending early market landscape. While overall activity remained relatively modest, the chart is punctuated by several notable spikes—most prominently in October 2019, when sales exceeded 7,000 in a single week, likely driven by a major release or surge in market attention. Other peaks in March and October 2020 further highlight intermittent waves of increased interest, suggesting that NFT adoption during this period was highly event-driven. However, the market also experienced noticeable downturns, such as in August 2020, when weekly sales dropped below 1,000. This inconsistency points to a market still in its formative stage, sensitive to external triggers and lacking sustained momentum. Nonetheless, the final quarter of 2020 shows signs of recovery and steady growth, with sales climbing back to around 5,000 per week—an indication that foundational demand was beginning to solidify just before the explosive growth of 2021.

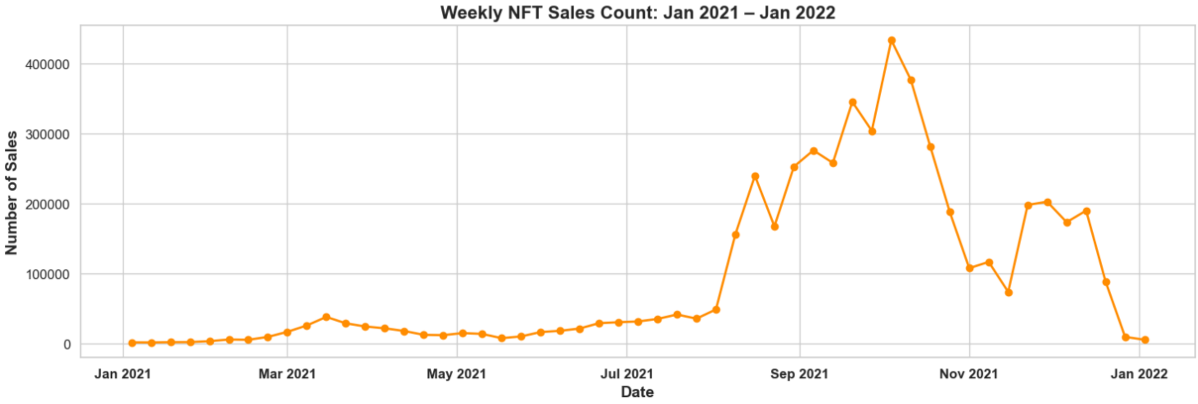

Fig. 4.3. Weekly NFT Sales Count (2021-2022)

The period from January 2021 to January 2022 marks a full boom-and-bust cycle in the NFT market, reflecting both its explosive rise and sharp contraction. During the first half of 2021, weekly sales remained relatively subdued, fluctuating below 50,000, with only a modest uptick around March, likely driven by early mainstream exposure and the beginning of media attention. This relatively calm phase set the stage for the dramatic surge that followed.

From July onward, the market entered an aggressive growth phase. Weekly sales skyrocketed, culminating in a record-breaking peak of over 400,000 transactions in late August and early September. This rapid escalation was fueled by a convergence of factors: heightened media coverage, celebrity endorsements, corporate interest, and major NFT drops. The sharp rise signaled the market's transition into a speculative frenzy, with widespread public participation.

However, this momentum proved unsustainable. Beginning in September, weekly sales began a steep and sustained decline. By mid-November, sales had plummeted to nearly 100,000 marking a 75% drop from the peak in just two months. Although there was a brief recovery later in November, with volumes bouncing back to around 200,000, this rebound was short-lived. By January 2022, the market had nearly returned to its pre-boom baseline, with weekly sales falling below 20,000 once again. This dramatic reversal underscores the fragility of speculative market cycles and highlights how quickly hype-driven growth can unravel in the absence of sustained demand or fundamental value.

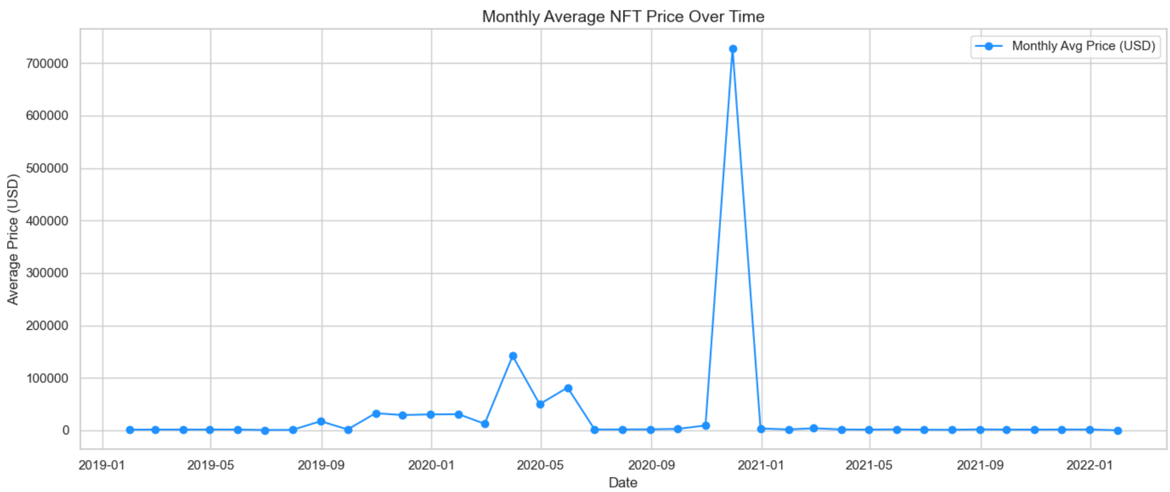

Fig. 4.4. Monthly Average NFT Prices Over Time

The chart displays the average USD price of NFTs per month between early 2019 and early 2022. It is largely characterized by low average prices hovering near zero. A brief uptick appears around April–

May 2020, but the most striking feature is a massive spike in November–December 2020, where the average monthly price reaches approximately $720,000. This extraordinary peak is likely driven by one or a few exceptionally high-value transactions that significantly distort the overall average. Due to this distortion, removing outliers becomes necessary to better observe general pricing trends over time.

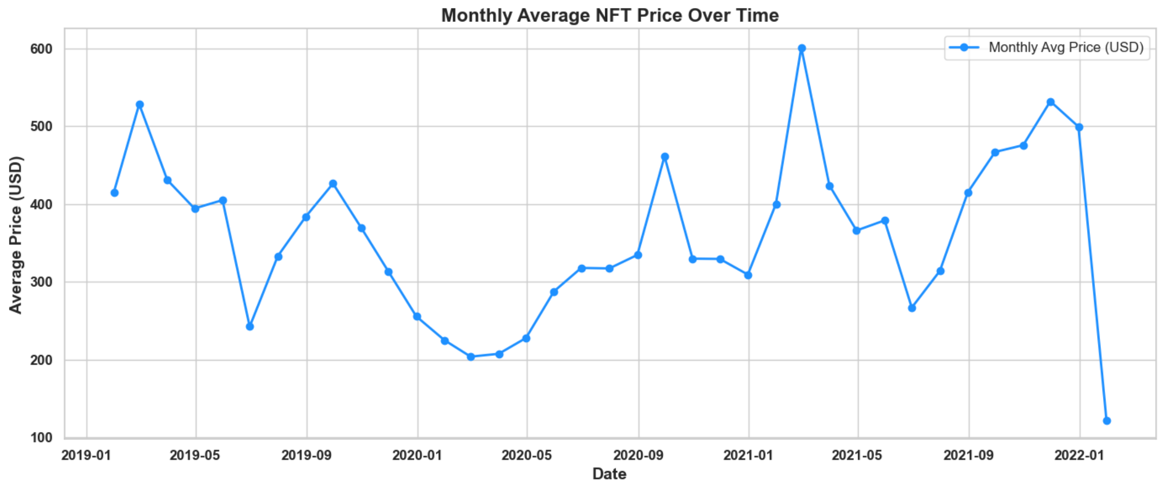

Fig. 4.5. Monthly Average NFT Prices Over Time (Excluded Outliers)

The monthly average NFT price data from January 2019 to January 2022 reveals relatively moderate fluctuations, mostly ranging between $200 and $600. A slight upward trend is noticeable throughout 2020 and into 2021, with a few months—such as February and November 2021—experiencing temporary increases above $500. After removing the extreme outlier observed in late 2020, the overall pattern becomes much clearer and more interpretable. Without the distortion of that anomalous spike, the monthly averages more accurately reflect typical market behavior, allowing for a more meaningful analysis of price trends over time.

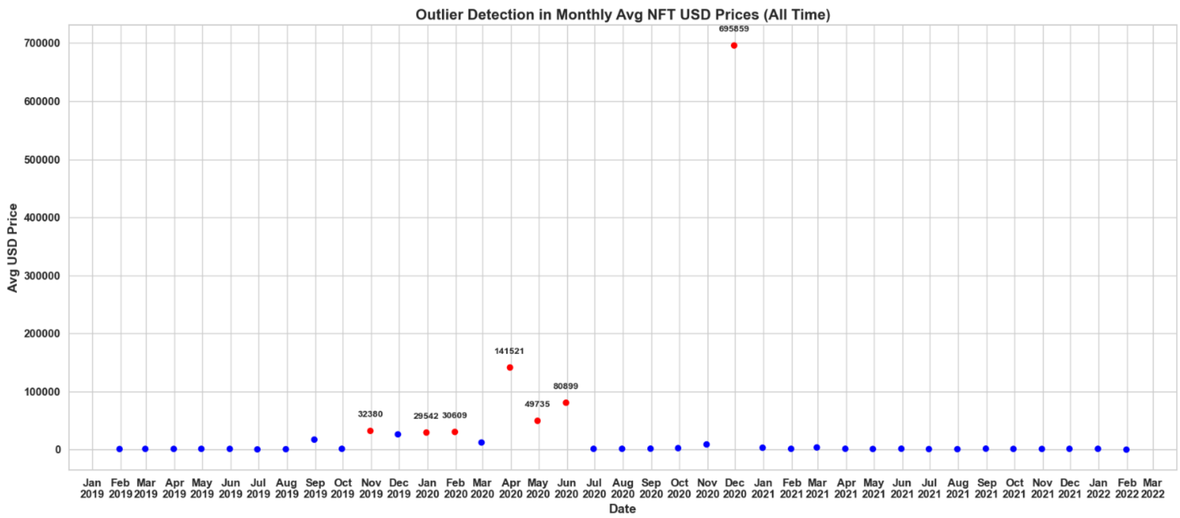

Fig. 4.6. Outlier Detection in Monthly Average NFT Prices

Outliers are highlighted in red in the graph. Several significant outliers stand out in the graph, particularly in November 2019, January 2020, and February 2020 (based on values calculated at the end of the previous months). The most dramatic spike appears in December 2020, where the average NFT price sharply rises to over $695,000—likely driven by one or more exceptionally high-value transactions. Other noticeable anomalies occur in April, May, and June 2020, also reflecting temporary but extreme surges in average prices. These outliers distort the overall trend and justify their removal for a clearer view of general market behavior.

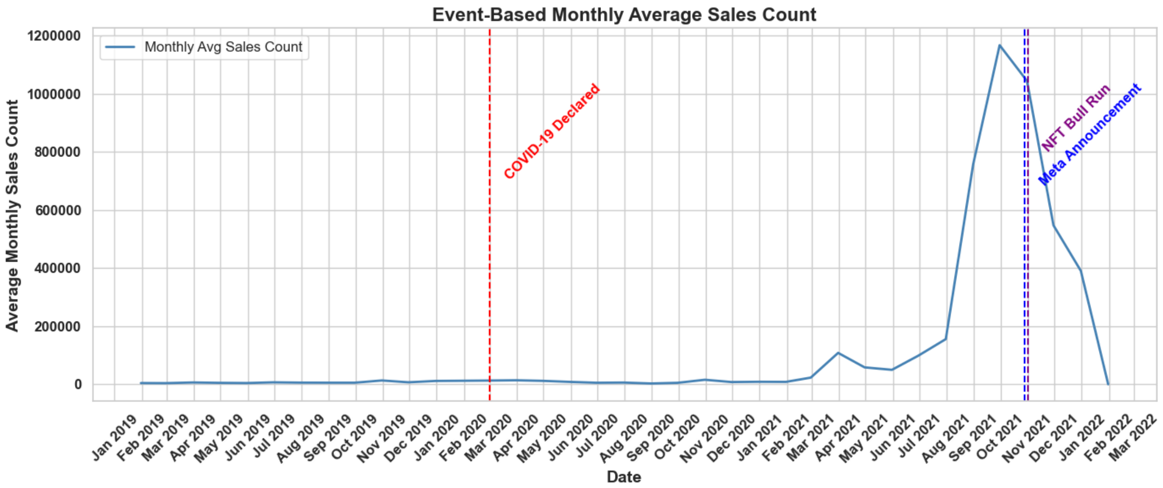

Fig. 4.7. Event Based Monthly Average Sales Count

The graph depicting the monthly average NFT sales from January 2019 to March 2022 clearly illustrates how global events influenced market behavior. During this period, three major events had a particularly noticeable impact on NFT activity: the declaration of COVID-19, the NFT bull run of 2021, and Meta's announcement regarding its shift toward the metaverse.

COVID-19 Declared (March 2020)

When the World Health Organization declared COVID-19 a global pandemic in March 2020, many industries experienced immediate and dramatic shifts. However, the NFT market did not show a comparable reaction. Sales remained relatively flat throughout the rest of 2020. This suggests that NFTs had not yet reached widespread recognition, and interest in digital collectibles was still limited at that time.

NFT Bull Run (3rd Quarter of 2021)

The NFT market experienced extraordinary growth during the third quarter of 2021. Starting in July, weekly sales rose rapidly, peaking at over 400,000 by late August and early September [9]. This surge was fueled by a combination of high-profile sales and increasing public engagement. For instance, the digital artist Beeple (real name Mike Winkelmann) made headlines when his artwork *"Everydays: The First 5000 Days"* sold for $69 million at a Christie's auction in March 2021, setting a historic precedent for digital art [10].

Platforms like NBA Top Shot, which allows users to buy, sell, and trade officially licensed video highlights from the NBA as NFTs, also contributed to the hype. Similarly, the rise of NFT collections such as Bored Ape Yacht Club (BAYC) [11]—a series of 10,000 unique hand-drawn cartoon ape characters that doubled as digital status symbols and club memberships—played a significant role in driving speculative interest [12]. Endorsements and participation by celebrities further amplified the market frenzy.

Meta's Metaverse Announcement (October 2021)

Facebook’s rebranding as Meta in October 2021 and its announcement of a strategic pivot toward the metaverse generated renewed momentum in the NFT space. This move increased public interest in the future of digital assets and virtual environments, reinforcing the perception of NFTs as integral components of this emerging digital frontier. As a result, monthly NFT transactions soared to an all-time high of approximately 1.2 million in November 2021 [13] *(note: OpenSea data alone shows slightly under 1.2 million; the figure likely includes sales from multiple platforms combined)*.

Following this peak, the market began to cool rapidly. Sales declined sharply after November 2021 [14], indicating that much of the preceding growth was driven by speculation rather than sustained utility or user adoption. This downturn reflects the volatile nature of hype cycles in emerging digital markets.

Summary

Between January 2019 and March 2022, the NFT market demonstrated its sensitivity to global events. While it remained stagnant at the onset of COVID-19, it experienced rapid growth during the 2021 bull run and following Meta’s metaverse announcement. However, this growth proved to be unsustainable. This period not only highlighted the speculative nature of NFTs as investment assets but also underscored the uncertainties surrounding the future of digital ownership.

4.3. Trend Analysis

Trend analysis refers to the process of examining data over time to identify consistent patterns or long-term movements. In time series data, trends indicate the general direction in which data points are moving, either upward (positive trend), downward (negative trend), or remaining relatively constant (no trend). It is a fundamental component of exploratory data analysis and forecasting, widely used across domains such as finance, economics, marketing, and emerging digital markets like NFTs.

According to Chatfield [8], trend analysis aims to uncover “long-term systematic movements” in time series, which can be distinguished from short-term fluctuations, seasonal effects, or irregular noise. Identifying a trend provides valuable insights into underlying structural changes, external shocks, or evolving user behavior.

In the context of digital asset markets, such as NFTs (Non-Fungible Tokens), trend analysis helps detect shifts in market momentum, adoption phases, or speculative bubbles. A key technique used in this process is the moving average (MA), which smooths out short-term noise to highlight broader patterns.

In this study, trend analysis was conducted using MA Momentum and MA Trend Categorization, both based on moving average comparisons over time. These approaches allow for distinguishing between short, medium, and long-term trends and provide a nuanced view of how the NFT market has evolved over the analyzed period.

4.3.1. Moving Average

Moving averages are a widely used technique in time series analysis, particularly for identifying underlying trends by smoothing out short-term fluctuations. By averaging data points over a defined window, this method helps reduce volatility and noise, making long-term patterns more visible. In financial and digital asset markets—such as NFTs—moving averages are especially valuable for tracking market momentum and detecting shifts in activity over time [7]. In this study, they serve as a crucial analytical tool to interpret the evolving dynamics of NFT sales volume across different temporal scales.

Moving Average (MA) Formula

The Simple Moving Average (SMA) for a time series is calculated as the arithmetic mean of values over a specific time window. The formula is as follows:

$$\mathrm{MA}_t = \frac{1}{n}\sum_{i=0}^{n-1} \mathrm{x}(t-i)$$

Where:

- MA_t: Moving average at time t
- n: Window size (number of periods over which the average is calculated)
- x(t-i): Observed value at time t-i

This formula takes the average of the most recent n observations, smoothing out short-term fluctuations to highlight longer-term trends. A smaller n results in a more responsive moving average that reacts quickly to changes, while a larger n produces a smoother line that emphasizes overall momentum.

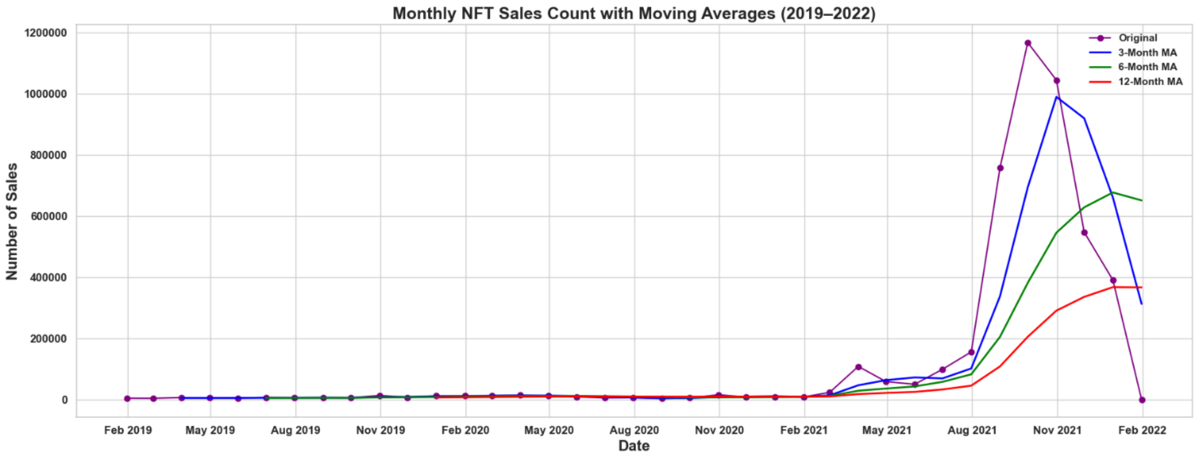

Fig. 4.8. Monthly NFT Sales Count Moving Averages (2019 – 2022)

The chart presents monthly NFT sales from 2019 to early 2022, along with 3, 6, and 12-month moving averages that reveal different layers of market behavior. Sales activity remained relatively low and stable until early 2021, fluctuating between 4,000 and 16,000 monthly sales. A sharp breakout began in early 2021, culminating in a dramatic peak of over 1.2 million sales around August–September. All four lines show a common peak period between mid-2021 and early 2022, reflecting the NFT market's rapid ascent and subsequent decline.

The original monthly sales line captures strong volatility with sudden spikes, while the 3-month moving average quickly reflects these short-term surges and crashes. The 6-month moving average smooths the curve further, showing mid-range market cycles with more gradual changes. Finally, the 12-month average less sensitive to short-term noise highlights the overall structural momentum of the market, peaking later and declining more gently. In essence, the 3-month MA is ideal for tracking immediate trends, the 6-month MA highlights emerging or fading cycles, and the 12-month MA reveals the underlying trajectory and long-term momentum of the NFT ecosystem.

4.3.2. Moving Average (MA) Momentum

MA Momentum is a technique that evaluates the speed and direction of market trends using Moving Averages (MAs). This method measures the rate of change, or momentum, of a moving average over a specified period, helping to identify whether a trend is strengthening or weakening.

Typically, MA Momentum is calculated by examining the difference between short-term and long-term moving averages, or by tracking how a single moving average changes over time. A positive momentum indicates a strengthening upward trend, while a negative momentum signals a strengthening downward trend.

Because MA Momentum is sensitive to market fluctuations, it is particularly useful for detecting early trend developments or potential reversals. In volatile and rapidly evolving markets, such as the NFT space, MA Momentum allows investors and analysts to respond swiftly to changes in market dynamics.

To further explore the trend dynamics of the NFT market, we analyze the MA Momentum using two different time horizons, 3-month and 12-month periods, providing complementary perspectives on short-term fluctuations and long-term trend strength between 2019 and 2022.

MA Momentum Formula

The Moving Average (MA) Momentum can be calculated as the difference in the moving average value between two time points, typically spaced by a fixed interval. The general formula is:

$$\text{MA Momentum}_t = \text{MA}_t - \text{MA}_{t-k}$$

Where:

- MA Momentum_t : Momentum value at time t
- MA_t : Moving average at time t
- MA_{t-k} : Moving average k periods before time t
- k : Lag interval

This formula measures how much the moving average has changed over a specific lag k, providing insights into the direction (positive or negative) and strength (magnitude) of the trend. A strongly positive momentum value suggests accelerating growth, while a negative value may indicate a slowdown or reversal in market activity.

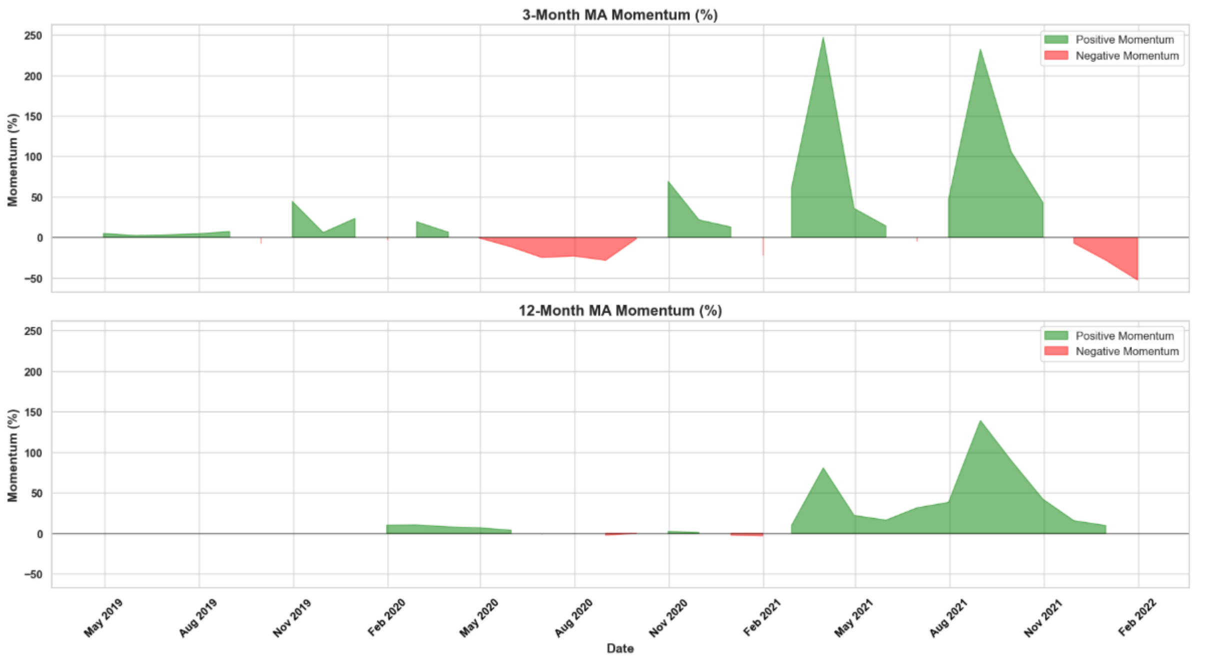

Fig. 4.9. 3-12 Month MA Momentum Graphs

The 3-month and 12-month moving average momentum charts of the NFT market from 2019 to early 2022 reveal important differences in short-term versus long-term trends. For example, the 3-month momentum frequently shows sharp positive spikes, such as in late 2019 and early 2020, indicating brief periods of market optimism or rapid price increases. However, during these times, the 12-month momentum remains relatively flat, suggesting that these short-term gains did not translate into sustained long-term growth. This highlights how the 3-month momentum is more sensitive to transient market excitement and noise.

In contrast, from mid-2020 through 2021, both momentum measures rise, though with different dynamics. The 3-month momentum exhibits multiple peaks and dips, reflecting high volatility and fluctuating market sentiment, especially during the NFT boom in 2021. Meanwhile, the 12-month momentum steadily climbs, confirming that the market experienced a more persistent and meaningful upward trend. Notably, toward late 2021 and early 2022, the 3-month momentum turns sharply negative, signaling short-term cooling or corrections, while the 12-month momentum declines more gradually. This difference suggests that while short-term momentum reacted quickly to changing conditions, the longer-term momentum indicated a more stable but eventual shift in market direction. Overall, the interplay of these momentum periods offers valuable insights into both fleeting market reactions and deeper trend shifts.

4.3.3. Moving Average (MA) Trend Categorization

MA Trend Categorization is a method used in technical analysis to classify the current market trend based on the behavior of moving averages over different time periods. By comparing short-term and long-term moving averages, analysts can identify whether an asset is in an uptrend, downtrend, or sideways consolidation phase. For example, when a short-term MA crosses above a long-term MA, it often signals a bullish trend (known as a “golden cross”), while the opposite crossover indicates a bearish trend (“death cross”) [15]. This technique helps to filter out market noise and provides a clearer picture of the underlying trend strength and direction. MA trend categorization is widely used across various financial markets due to its simplicity and effectiveness in trend-following strategies [16].

Quantile-Based Trend Classification

To classify market trends based on moving averages, this study uses a quantile-based categorization approach. Specifically, 12-month MA values are segmented into three trend levels—Low, Medium, and High—by computing the 33rd and 66th percentiles of the MA distribution:

$$\begin{cases} Low\ Trend \quad if\ MAt \leq Q33 \\ Medium\ Trend \quad if\ Q33 < MAt \leq Q66 \\ High\ Trend \quad if\ MAt > Q66 \end{cases}$$

Where:

- MA_t: 12-month moving average at time t
- Q_{33} , Q_{66} : 33rd and 66th percentiles of the MA distribution

This method avoids the use of fixed thresholds and dynamically reflects the underlying data distribution, making it well-suited for exploratory trend analysis in volatile markets like NFTs.

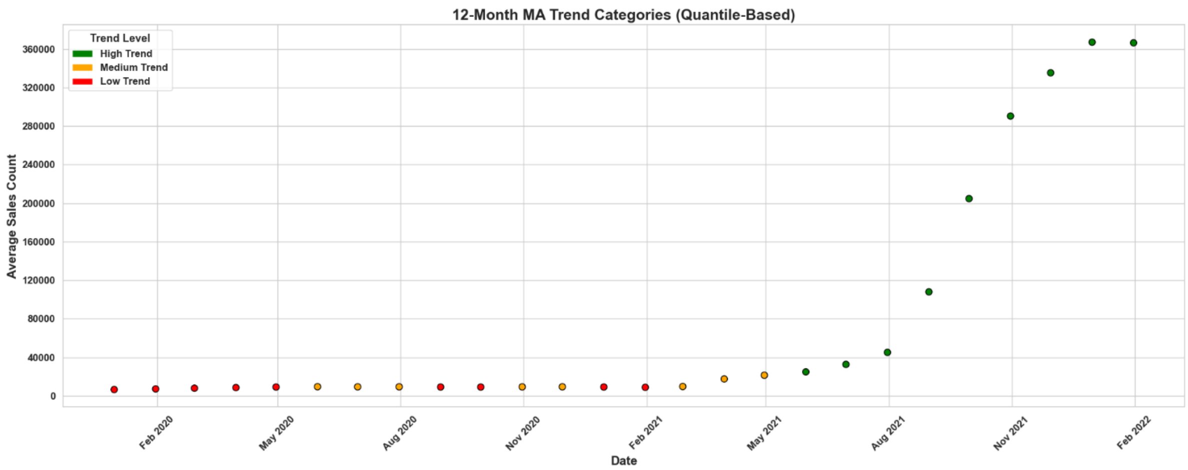

Fig. 4.10. 12 Month MA Trend Categories Based on Average Sales Count (2019-2022)

Note: Quantile thresholds used in this analysis were approximately 9,250 (33rd percentile) and 23,187 (66th percentile), based on the distribution of 12-month MA sales data.

The chart depicting the 12-month moving average (MA) trend categorization for the NFT market provides a clear and insightful view of the market's long-term momentum. Utilizing a quantile-based classification system, the MA values are segmented into three trend levels—Low, Medium, and High—based on the 33rd and 66th percentiles of the MA distribution. This data-driven approach enables a dynamic thresholding mechanism that adjusts to the statistical properties of the dataset, rather than relying on arbitrary fixed values. It effectively captures key shifts in market behavior over time, allowing for a nuanced interpretation of growth phases and periods of stagnation.

This segmentation effectively captures key shifts in market behavior over time, allowing for a nuanced interpretation of growth phases and periods of stagnation. The distribution of the data points across the categories:

```
Category Distribution:
 - Low Trend       9
 - High Trend      9
 - Medium Trend    8
```

Indicates a balanced representation, supporting the appropriateness of the chosen thresholds. From early 2020 to the first half of 2021, most data points fall into the Low and Medium trend categories, reflecting a relatively subdued or nascent market phase characterized by lower average sales counts. However, beginning around Q2 (the second quarter, i.e., April to June) of 2021, there is a marked transition: MA values start to climb, moving first into the Medium and then rapidly into the High trend category. By mid to late 2021, nearly all data points are classified as High Trend, signaling a significant surge in market activity and enthusiasm. This aligns with the well-documented NFT boom, during which average monthly sales counts exceeded 300,000, highlighting an extraordinary phase of growth and investor interest.

The use of a 12-month MA filters out short-term volatility, offering a robust lens to observe persistent structural shifts rather than temporary spikes. Given the clarity of category separation and the strong correlation with known market events, the current quantile-based thresholds appear analytically sound and do not warrant revision. This trend framework not only charts the NFT market's explosive rise but also contextualizes it within a broader trajectory, illustrating how the space evolved from a low-activity niche to a global digital asset phenomenon.

Conclusion of Trend Analysis

The trend analysis of the NFT market between 2019 and early 2022 reveals a clear evolution from low-volume early adoption to a speculative surge in 2021, followed by signs of cooling. The combined use of MA Momentum and MA Trend Categorization has proven effective in distinguishing between temporary fluctuations and sustained market movements. While short-term momentum highlighted brief periods of excitement and correction, long-term trends captured structural changes and market maturation. These findings underscore the dynamic nature of NFT markets, where shifts in user interest, media influence, and macroeconomic events drive rapid changes—offering both opportunities and risks for stakeholders.

5. Identifying Factors Affecting NFT Prices

5.1 Feature Selection and Correlation Analysis

This analysis was dedicated to identifying which numerical features in the dataset demonstrate statistically meaningful associations with the price of NFTs, specifically measured in USD. Understanding these relationships is important for both model building and market interpretation, as it helps reveal the underlying drivers of NFT valuation.

To begin with, a comprehensive selection of numerical features was extracted from the dataset using a data type filter. However, not all numeric columns are appropriate for inclusion in a correlation analysis. Several columns were excluded to avoid misleading results:

- total_price was removed because it directly contributes to usd_price (as the price in ETH or other tokens is multiplied by a token-specific USD conversion factor).
- payment_token.usd_price was excluded for the same reason—it represents part of the computational path that leads to the final price in USD.

- Identifiers such as id and asset.id were omitted because they carry no inherent numerical meaning and their inclusion could distort correlation values.

After refining the list of features, Pearson correlation matrix was computed. Pearson's correlation coefficient is a standard statistical measure that quantifies the strength and direction of a linear relationship between two variables, producing values between -1 and +1 [17].

- A value close to +1 indicates a strong positive correlation, where an increase in one variable is associated with an increase in the other.
- A value close to -1 implies a strong negative correlation, where higher values of one variable are associated with lower values of the other.
- Values near zero suggest a weak or no linear relationship.

To gain deeper insights, the correlations between each selected feature and the target variable usd_price were extracted from the matrix. The resulting list of correlations was then split into two distinct groups:

- Positively correlated features, where the rise in a feature's value tends to correspond with an increase in NFT price,
- Negatively correlated features, where higher feature values are typically associated with a drop in price.

Each of these groups was visualized independently using heatmaps to enhance interpretability.

These visualizations not only illustrate which features are most aligned with NFT price dynamics but also provide a basis for feature engineering and variable selection in later predictive modeling steps. The separation of positively and negatively correlated features additionally helps clarify whether certain attributes enhance or diminish NFT valuation, a useful insight for both economic modeling and decision support.

By carefully selecting variables, excluding confounding columns, and clearly visualizing relationships, this correlation analysis serves as a foundation for more robust and interpretable regression modeling in the subsequent phases of the project.

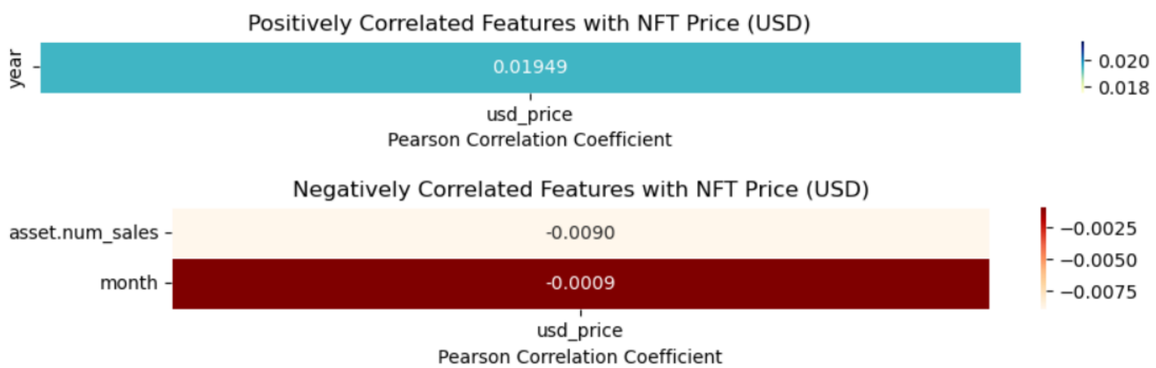

Fig. 5.1. Pearson correlation heatmaps showing positive (top) and negative (bottom) associations between selected numeric features and NFT prices in USD.

This figure visualizes the Pearson correlation coefficients between selected numerical variables and NFT prices in USD. Prior to the analysis, columns directly related to the target variable—such as total_price and payment_token.usd_price—were excluded to prevent artificial inflation of correlation values. Only independent variables with sufficient variance were included.

The figure is divided into two panels: the upper panel represents features positively correlated with NFT prices, while the lower panel shows features with negative correlations. According to the results, the year variable demonstrates a weak positive correlation ($r \approx 0.019$), suggesting a limited upward trend in NFT prices over time. Negatively correlated features include asset.num_sales and month, both of which exhibit very low correlation coefficients (approximately –0.0090 and –0.0009, respectively). These values imply that while high sales volume might be loosely associated with lower average prices, the monthly distribution does not appear to significantly influence pricing. Overall, the low magnitude of these correlations indicates that NFT price behavior may not be primarily governed by simple linear relationships, and that more complex or nonlinear dynamics could be involved.

5.2Pearson Correlation Analysis

In order to investigate the strength and direction of associations between individual numeric features and NFT prices in USD, a detailed Pearson correlation analysis was conducted. This approach served as a preliminary statistical screening tool to identify which attributes might meaningfully contribute to price variations and thus hold potential predictive value in subsequent modeling stages.

To maintain the validity and interpretability of the analysis, a series of preprocessing steps were applied before computation. First, variables that are either derived from or directly influence the target variable usd_price—such as total_price and payment_token.usd_price—were deliberately excluded [18]. These features, if included, could introduce artificial inflation of correlation values due to data leakage. Additionally, unique identifiers like id and asset.id were removed as they carry no real numerical relationship with the target and could distort the analysis.

The remaining numeric features were then individually assessed using the pearsonr function from the scipy.stats library. This function computes both the Pearson correlation coefficient (r)—which quantifies the strength and direction of a linear relationship—and the associated p-value, which tests the statistical significance of that relationship. Features with near-zero variance or insufficient data points were automatically skipped to ensure robustness and to avoid misleading results.

Once the correlation values were computed, they were sorted by the absolute value of the coefficient. This allowed for the identification of features that had the most prominent linear associations with usd_price, regardless of whether those associations were positive or negative. By focusing on both the magnitude and significance of the relationships, the analysis provided a clearer picture of which variables might influence NFT valuation in a meaningful way.

While Pearson correlation is limited to capturing linear dependencies, it remains a valuable initial diagnostic tool. The results obtained here form a foundational step toward building interpretable models and contribute to a better understanding of the underlying market mechanisms that may influence NFT pricing behavior.

```
Top 10 features by Pearson correlation with NFT price:
                 correlation   p_value
year                0.019489  0.012991
asset.num_sales    -0.008974  0.252737
month              -0.000886  0.910077
```

Table 5.1. Pearson correlation results showing the strength and significance of linear relationships between selected features and NFT prices in USD.

The table displays the top three features ranked by their Pearson correlation with NFT prices in USD. Among the variables evaluated, only the year attribute shows a statistically significant linear relationship with the target variable, showing a weak positive correlation coefficient of 0.0195 and a p-value of 0.013. This suggests that over time, there has been a slight upward trend in NFT prices; however, the strength of this association is minimal. Despite its significance from a statistical perspective, the low correlation magnitude indicates that year alone is not a strong predictor of price .

In contrast, the other two features—asset.num_sales and month—demonstrate negative correlation coefficients of –0.00897 and –0.00089, respectively. However, both have high p-values (0.253 and 0.910), meaning their relationships with NFT price are not statistically significant. In practical terms, this means we cannot confidently claim that higher sales volume or the time of the year (month) has a meaningful linear effect on the price of NFTs. The observed correlations may be due to random variation or noise within the dataset.

Overall, this result shows that most numeric features in the dataset exhibit very weak linear relationship with NFT price. Furthermore, it emphasizes the importance of considering both correlation magnitude and statistical significance when interpreting such results.

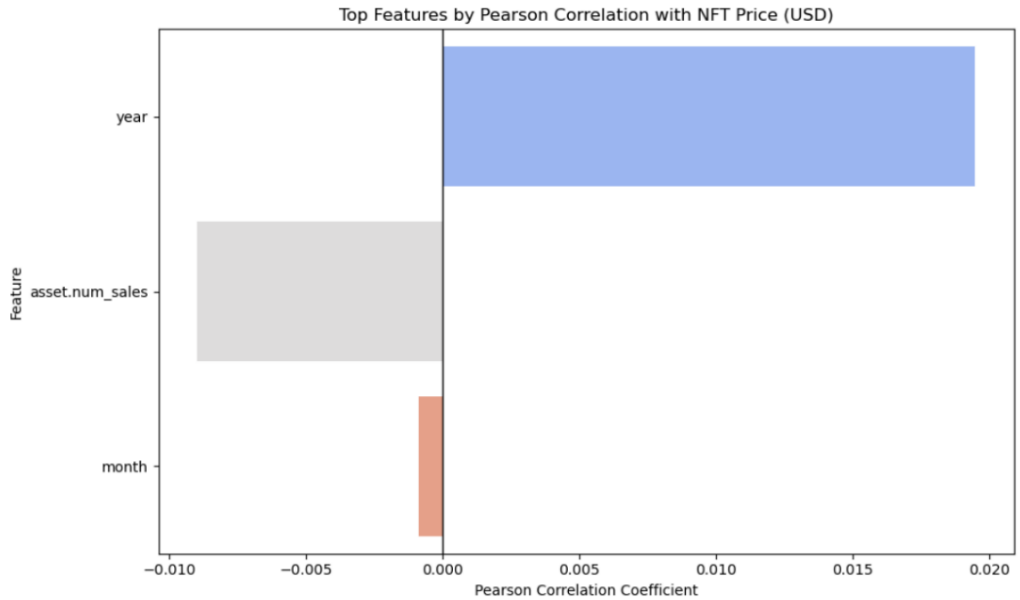

Fig. 5.2. Top three numerical features most correlated with NFT prices using Pearson coefficients.

This bar chart presents the top three numerical features most strongly correlated with NFT prices in USD, based on Pearson correlation coefficients. The x-axis shows the correlation values, while the y-axis lists the corresponding features. The bars are color-coded to reflect the direction (positive or negative) and relative strength of each relationship.

Among the analyzed variables, year displays a weak but positive correlation of approximately 0.019 with usd_price. This suggests that, over time, there has been a slight increase in the average price of NFTs. While the correlation is small in magnitude, its positive direction implies that NFTs sold in more recent years tend to have marginally higher prices compared to earlier years. This could reflect a broader market trend such as growing popularity, increased speculative activity, or greater perceived value of NFTs as the market matured between 2019 and 2021.

In contrast, asset.num_sales and month exhibit negative correlations, around –0.009 and –0.0009 respectively. The interpretation of these values is limited by their extremely low magnitude and statistical insignificance. However, they may loosely suggest that NFTs with higher sales counts are slightly more likely to be lower in price—perhaps due to mass production or less scarcity—while the month in which a transaction occurs does not appear to influence pricing in a meaningful linear way.

Taken together, the chart indicates that no single numeric feature demonstrates a strong linear relationship with NFT price. Nonetheless, the direction and sign of these correlations can help shape hypotheses for further modeling and exploratory analysis, especially when examined in combination with other temporal or categorical factors.

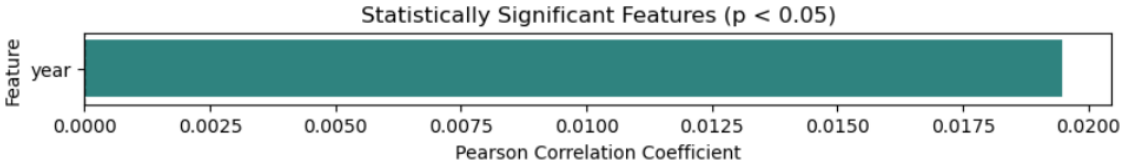

Fig. 5.3. Numerical features with statistically significant correlations ($p < 0.05$) with NFT prices.

This figure highlights the subset of features that exhibit statistically significant linear relationships with NFT prices, as determined by a p-value threshold of 0.05. Among all evaluated numerical variables, only the feature ‘year’ met this criterion, with a Pearson correlation coefficient of approximately 0.019 and a p-value below 0.05. Although the correlation is weak in magnitude, its statistical significance indicates that the observed association is unlikely to be due to random chance.

The positive direction of the correlation suggests that NFT prices have slightly increased over time during the analyzed period (2019–2021). This trend may reflect general market expansion, increased media attention, or greater user adoption across the years. It is important to note, however, that statistical significance does not imply practical significance; the explanatory power of year as a predictor remains limited when considered in isolation.

This step of isolating statistically significant features is crucial in filtering out noise and ensuring that only features with credible linear relationships are considered in subsequent modeling or interpretation phases.

5.3 OLS Regression Analysis

This section applies an Ordinary Least Squares (OLS) regression model to examine how multiple numerical features jointly influence NFT prices in USD. While correlation analysis provides useful information about the pairwise linear relationships between individual variables and the target, it does not account for interactions or confounding effects from other features. In contrast, OLS regression allows for the simultaneous evaluation of all selected predictors, offering insight into their individual impact while adjusting for the presence of others.

To prepare the data for modeling, several preprocessing steps were performed. First, we excluded variables that either directly encode price information, such as total_price and payment_token.usd_price, or act as identifiers like id and asset.id. Including such features would risk data leakage or inflate model performance in a misleading way. From the remaining numeric columns, we selected those with sufficient variability to serve as independent variables.

Before fitting the OLS model, the selected features were normalized using StandardScaler [19]. This transformation centers each variable around a mean of zero with unit variance, making it possible to interpret the resulting regression coefficients on a comparable scale. Additionally, a constant term was added to the design matrix to estimate the intercept of the linear model.

The target variable usd_price was regressed on the scaled feature set using the OLS function from the statsmodels library. The model output includes estimated coefficients, standard errors, t-statistics, and p-values for each predictor. These metrics allow us to determine which features have a statistically significant association with NFT price, as well as the direction (positive or negative) and relative magnitude of their influence [20].

This multivariate approach not only refines our understanding of individual feature contributions but also provides a more realistic framework for interpreting complex relationships in the NFT market. It serves as a bridge between exploratory analysis and predictive modeling, and offers a foundation for evaluating feature importance under controlled statistical conditions.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              usd_price   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     2.673
Date:                Sat, 31 May 2025   Prob (F-statistic):             0.0456
Time:                        14:40:39   Log-Likelihood:            -1.7791e+05
No. Observations:               16245   AIC:                         3.558e+05
Df Residuals:                   16241   BIC:                         3.559e+05
Df Model:                           3
Covariance Type:            nonrobust
===================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const           1050.4840    108.334      9.697      0.000     838.138    1262.830
asset.num_sales -148.0205    108.888     -1.359      0.174    -361.453      65.412
year             281.9647    108.866      2.590      0.010      68.575     495.354
month              8.9485    108.630      0.082      0.934    -203.978     221.875
==============================================================================
Omnibus:                    59526.239   Durbin-Watson:                   1.734
Prob(Omnibus):                  0.000   Jarque-Bera (JB):      40787323355.293
Skew:                          77.663   Prob(JB):                         0.00
Kurtosis:                    7764.067   Cond. No.                         1.11
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Fig. 5.4. OLS regression summary showing the impact of standardized numerical features on NFT prices (USD).

The results of the OLS regression model are presented above, aiming to quantify the effect of selected numerical features on NFT prices (in USD). The dependent variable is usd_price, and the model includes three independent predictors: asset.num_sales, year, and month, all of which were standardized prior to modeling.

The R-squared value is approximately 0.000, indicating that the model explains virtually none of the variance in NFT prices. Although the overall F-statistic (2.673) is associated with a p-value of 0.0456, which narrowly meets the threshold for statistical significance ($p < 0.05$), this does not imply that the model is a good fit. Rather, it suggests that at least one predictor may contribute some explanatory power, though the overall model still lacks practical utility.

Examining the individual coefficients:

- The variable year has a positive coefficient of 281.96 with a p-value of 0.010, indicating a statistically significant effect. This suggests that, holding other variables constant, an increase in the year is associated with a modest increase in NFT prices. While statistically reliable, this effect is weak in magnitude.

- asset.num_sales has a negative coefficient (–148.02) but a p-value of 0.174, implying the relationship is not statistically significant at the 5% level. This indicates no clear evidence that the number of sales meaningfully affects price in a linear way.

- The month variable shows an extremely small positive coefficient (8.95) and a very high p-value (0.934), providing no statistical support for a meaningful relationship with price.

Additional diagnostics—such as the Durbin-Watson statistic (1.734)—do not indicate serious autocorrelation problems, but other values such as the Jarque-Bera test and high skewness and kurtosis suggest the presence of non-normality and potential outliers in the residuals.

In summary, the OLS regression results reinforce the insights obtained from the prior correlation analysis. Although the variable year exhibits a statistically significant positive association with NFT prices, its explanatory power remains limited. The remaining features—such as asset.num_sales and month—do not contribute meaningfully within a multivariate context. These findings collectively suggest that NFT pricing dynamics are unlikely to be driven by simple linear relationships and may be better captured through more advanced or nonlinear modeling techniques.

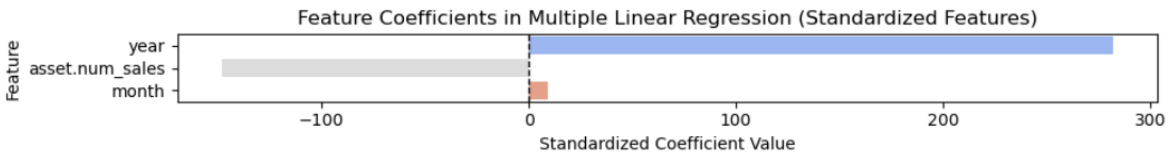

Fig. 5.5. Standardized coefficients from the linear regression model, showing each feature's relative impact on NFT prices.

The visualization above presents the standardized regression coefficients obtained from the OLS model. As all features were normalized prior to modeling, the chart enables a direct comparison of each variable's relative effect on NFT prices. As previously discussed, only the year variable exhibited statistical significance, which is reflected here by its dominant positive coefficient.

The negative coefficient of asset.num_sales and the near-zero value for month reaffirm their limited explanatory power, aligning with prior statistical outputs. Overall, the visual emphasizes the imbalance in predictive influence among the variables and supports the conclusion that NFT price formation cannot be sufficiently explained through these linear components alone.

6.Examining the Most Valuable NFTs in Collections

6.1 Analysis Objective

In this section, we aim to identify and characterize the most valuable NFTs within the dataset by examining their shared categorical traits. The ultimate goal is to uncover patterns that may explain why certain NFTs attain significantly higher market value than others. The methodology follows a two-step analytical process: first, the most valuable NFTs are extracted through clustering; then, frequent pattern mining is applied to identify their common categorical features.

6.2 Methodology and Analysis Process

6.2.1 K-Means Clustering Analysis

Why K-Means?

K-Means is a widely used algorithm that performs fast and effective clustering on numerical data [21]. In the context of NFT data, numerical features such as price, number of sales, and token USD value allow the identification of NFT clusters with similar characteristics in different segments. This enables the specific identification of the "valuable NFT" group by examining data at the cluster level.

Why 3 Clusters?

Choosing 3 clusters is a common starting point in the literature and is based on preliminary data distribution analysis [22]. Selecting too few clusters can lead to excessive generalization, while too many clusters can cause over-segmentation. Three clusters provide a meaningful division that can represent low, medium, and high-value NFTs. Additionally, analyses with different numbers of clusters (e.g., 4 or 5) yielded similar results, but 3 clusters optimize both interpretability and performance.

Data Standardization

Numerical features (price, number of sales, token price) were standardized using StandardScaler to have a mean of 0 and variance of 1. This ensures balanced contributions among features [23].

Selection of the Most Valuable NFT Cluster

The cluster with the highest average total_price was selected as the "valuable NFTs" cluster. Using the mean price helps reduce the influence of extreme outliers and represents the general value level. The choice of mean over median takes into account rare but high-priced sales, thus broadening the coverage of the valuable group.

Random Sampling of 10,000 Examples

This subset was randomly sampled to reduce computational load and keep memory usage manageable for the FP-Growth algorithm. A sample size of 10,000 is statistically meaningful for analysis. Taking fewer samples could weaken the strength of discovered patterns, while taking more would unnecessarily increase processing time. The random sampling was made reproducible with random_state=42.

6.2.2 Frequent Pattern Mining Using FT-Growth Algorithm

Why FP-Growth?

FP-Growth is one of the fastest and most efficient algorithms for frequent itemset mining [24]. Unlike Apriori, it does not generate candidate itemsets, which is a significant advantage when working with high-dimensional and sparse data. Since NFT data includes numerous categorical features and sub-features, FP-Growth is a suitable choice for fast and memory-efficient mining.

Why a Minimum Support of 0.05 (5%)?

Support indicates how frequently an itemset appears in the sample. A 5% support threshold effectively filters out infrequent and irrelevant patterns in large datasets. A very low support would include rare and insignificant patterns, while a very high support might miss important but less frequent patterns. The 5% threshold is widely accepted in frequent itemset mining literature [25].

Data Dimensionality Reduction and Feature Filtering

One notable observation during the data preprocessing phase was the drastic reduction in the number of categorical feature columns—from an original 43,122 down to just 63 after filtering based on minimum column support thresholds. This significant decrease reflects the inherently sparse and high-dimensional nature of the NFT metadata, where many categorical attributes or their specific values occur very infrequently.

By applying a column support filter of 1%, we effectively removed features that appeared in less than 1% of the sampled NFTs. This step is crucial for two primary reasons: firstly, it reduces the computational complexity and memory consumption of the FP-Growth algorithm, making the frequent pattern mining process feasible and efficient; secondly, it helps eliminate noise and spurious patterns that arise from extremely rare attribute occurrences, thereby enhancing the quality and interpretability of the discovered frequent itemsets [25].

This feature selection approach ensures that the analysis focuses on meaningful, commonly shared attributes that are more likely to contribute significantly to NFT valuation patterns. It aligns with standard practices in frequent pattern mining literature, which emphasize balancing between comprehensiveness and tractability when working with large, sparse categorical datasets.

Feature Processing

Categorical features were converted into “feature=value” strings and encoded into a one-hot encoded boolean matrix using TransactionEncoder to prepare for the algorithm. Columns with less than 1% support were filtered out beforehand to improve computational efficiency and analysis quality.

6.3 Analysing Key Attributes of Valuable NFTs: A Detailed Perspective

```
The cluster containing the most valuable NFTs: 1
Original number of columns: 43122, Filtered number of columns: 63

Common Features Among the Most Valuable NFTs:

No  Common Features                                                                                            Support
------------------------------------------------------------------------------------------------------------------------
1   Category=Virtual Worlds                                                                                    0.4816
2   payment_token.name=Decentraland MANA                                                                       0.2806
3   payment_token.name=Decentraland MANA, Category=Virtual Worlds                                              0.2430
4   asset.collection.name=Decentraland Wearables, Category=Virtual Worlds                                      0.2269
5   asset.collection.name=Decentraland Wearables                                                               0.2269
6   asset.collection.name=Decentraland Wearables, payment_token.name=Decentraland MANA                         0.2268
7   asset.collection.name=Decentraland Wearables, Category=Virtual Worlds, payment_token.name=Decentraland MANA 0.2268
8   Category=Collectibles                                                                                      0.2050
9   Category=Uncategorized                                                                                     0.1692
10  payment_token.name=USD Coin                                                                                0.1254
11  payment_token.name=SAND                                                                                    0.1181
12  payment_token.name=Gala                                                                                    0.1114
13  payment_token.name=SAND, Category=Virtual Worlds                                                           0.1034
14  asset.collection.name=The Sandbox, Category=Virtual Worlds                                                 0.0792
15  asset.collection.name=The Sandbox                                                                          0.0792
16  asset.collection.name=Town Star, Category=Virtual Worlds                                                   0.0779
17  asset.collection.name=Town Star                                                                            0.0779
18  payment_token.name=SAND, asset.collection.name=The Sandbox, Category=Virtual Worlds                        0.0766
19  payment_token.name=SAND, asset.collection.name=The Sandbox                                                 0.0766
20  asset.collection.name=Town Star, payment_token.name=Gala                                                   0.0716
```

Fig. 6.1. Common Features Among the Most Valuable NFTs

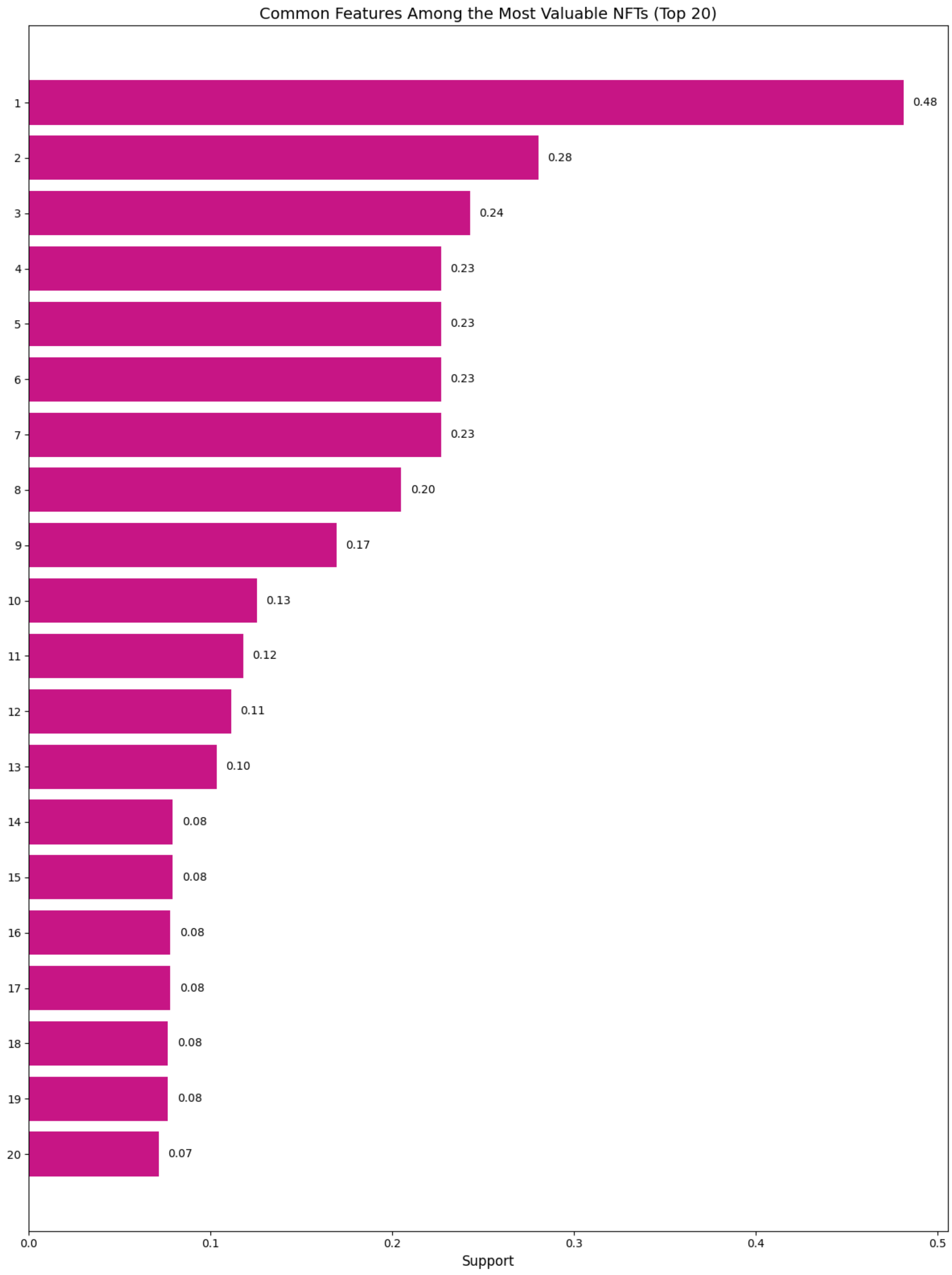

Fig. 6.2. A bar chart illustrating the Common Features Among the Most Valuable NFTs (Top 20) With Suuport Values

The horizontal bars in the chart represent the frequency with which each frequent feature set (itemset) appears within the dataset, indicated by their support values. These values reflect the representation ratio within the sample of high-value NFTs, providing critical insight into the attributes and patterns that dominate the valuation landscape.

The standout feature is the “Category=Virtual Worlds”, with a support of approximately 48.16%. This indicates that nearly half of the most valuable NFTs belong to virtual worlds, underscoring the dominant role that metaverse and immersive digital environments play in the current NFT market. This aligns with the growing trend of integrating NFTs into interactive, gamified ecosystems where users can own, trade, and utilize assets within expansive virtual universes [3].

Following this, “payment_token.name=Decentraland MANA” appears frequently with a 28.06% support. This highlights that the economic infrastructure specific to certain platforms, such as Decentraland’s native token MANA, is a significant factor influencing the perceived and transactional value of NFTs. The strong presence of this token indicates investor trust and liquidity within that ecosystem, reinforcing the importance of platform-specific currencies in NFT markets [26].

Moreover, the combination of “Decentraland MANA” with the “Virtual Worlds” category occurs with 24.30% support, indicating a tight coupling between the platform’s economic system and its virtual environment assets. This synergy is further evidenced by the prominence of “asset.collection.name=Decentraland Wearables”, which frequently co-occurs with both the category and the payment token (support values around 22.7%). These wearables are not merely collectible art but functional items usable within the virtual world, adding utility and driving up their market value [3][26].

Beyond Decentraland, the “Category=Collectibles” accounts for about 20.5%, demonstrating that traditional collectible NFTs remain significant players in the value hierarchy. However, they are surpassed by virtual world assets, suggesting a market shift toward interactive, utility-rich NFTs [2].

A notable portion of NFTs are “Uncategorized” (approximately 16.9%), which might reflect emerging collections, incomplete metadata, or new market segments that have yet to be classified—highlighting the dynamic and evolving nature of the NFT ecosystem [2].

Other payment tokens like USD Coin (USDC) at 12.54%, SAND (The Sandbox’s token) at 11.81%, and Gala at 11.14% also feature prominently. These indicate that multiple metaverse platforms maintain active economies that contribute significantly to NFT valuations. Notably, combinations such as “SAND” with “Virtual Worlds” (10.34%) and collections like “The Sandbox” (7.92%) and “Town Star” (7.79%) reveal the multi-platform character of high-value NFTs. These interrelations between payment tokens, categories, and collections illustrate how economic ecosystems extend across different virtual environments, each fostering its own subset of valuable digital assets [2][26].

Multi-attribute combinations dominate the upper ranks of the frequent itemsets, signifying that the intersection of category, collection, and payment method is crucial to understanding NFT value formation. For example, the recurring itemset combining “SAND”, “The Sandbox” collection, and “Virtual Worlds” category at 7.66% support reflects how NFTs’ functionality and value are anchored not just in their standalone features but in the broader context of their platform’s economy and usage scenarios [2][26].

Similarly, the appearance of itemsets like “Town Star” collection with “Gala” token (7.16%) exemplifies the role of gaming and platform-specific ecosystems in shaping NFT market dynamics [2][26].

In summary, the frequent pattern mining results demonstrate that NFT valuations are multidimensional phenomena, deeply embedded within the ecosystems and economies that support them. Market value is not solely dictated by artistic merit or rarity but by a confluence of platform affiliation, token-based economic incentives, and the utility provided by the NFTs within their respective virtual environments. This multifaceted structure of value highlights the increasing sophistication of the NFT marketplace, where success depends on integration with thriving digital ecosystems as much as on individual asset attributes [3] [2] [26].

6.4 Interpretation of Findings

In conclusion, the high-support frequent itemsets identified through this analysis shed light on the multifaceted nature of value creation in the NFT market. These patterns not only reflect core investor preferences and dominant market trends but also highlight the intricate interplay between various economic, social, and technological factors that collectively shape the valuation of digital assets. Specifically, the results underscore that NFT valuation transcends the intrinsic artistic or collectible merit of the individual tokens, encompassing broader ecosystem dynamics such as inter-platform competition, network effects, and platform-specific token economies.

Investor behavior appears to be strongly influenced by the perceived utility and engagement potential of NFTs within their native virtual environments, as evidenced by the prominence of metaverse-related categories and associated payment tokens. This aligns with emerging research indicating that ecosystem loyalty—manifested through consistent participation within certain digital worlds—and the functional use cases of NFTs, such as interoperability and in-world usability, significantly enhance their economic value [27].

Furthermore, the findings suggest that the competitive landscape between NFT platforms drives differentiation, where projects that successfully integrate robust economic incentives, active communities, and innovative use cases are rewarded with higher market valuations. This systemic view reinforces the notion that the NFT market operates as a complex socio-technical system, where price signals encapsulate not only scarcity and demand but also user trust, platform credibility, and the evolving narrative of digital ownership.

Overall, this analysis provides critical empirical evidence supporting the view that value formation in NFTs is a dynamic, multi-dimensional process. Stakeholders—including collectors, creators, and platform developers—can leverage these insights to better navigate market complexities, optimize portfolio strategies, and design NFTs with enhanced appeal and sustainable economic impact.

6.5 Identifying Common Fetaures Across Multiple Clusters with Support Thresholds

Understanding the value creation processes within the NFT ecosystem in a comprehensive manner is inherently challenging due to the market's complex structure and multi-dimensional dynamics. In this context, the previous section utilized K-Means clustering and the FP-Growth algorithm to group NFTs with high market value and uncover their shared categorical characteristics. This analysis was designed to identify the distinct features of high-value NFTs and reveal the key factors influencing their valuation.

However, the NFT market is not limited to high-value assets. A comparative analysis of NFT segments across different value ranges is critical for gaining deeper insights into overall market dynamics and the interactions between segments.

The second analysis in this section focuses on systematically identifying the common attributes of NFTs within three different clusters (Cluster 0, 1, and 2), based on specified minimum support thresholds. In doing so, the characteristics of the high-value NFT segment are evaluated within a broader market perspective, offering a more detailed picture of the layered structure of value formation in NFTs.

The support threshold is a fundamental parameter in frequent itemset mining, indicating how frequently a feature or group of features appears in the dataset. Choosing different support levels allows the discovery of both general market characteristics and unique patterns in niche segments. Therefore, systematically examining the impact of support thresholds enhances our understanding of the heterogeneous structure of the NFT market and the diversity of user preferences.

In this section, the number of common itemsets across clusters and the cluster-specific differences will be analyzed in detail, followed by a comparative evaluation with the previous analysis that identified the characteristics of high-value NFTs.

6.5.1 Selection, Importance, and Analytical Contributions of Support Thresholds

The concept of *support* is one of the fundamental measures used in frequent itemset mining, and it refers to the proportion of observations in the dataset that contain a particular itemset. Selecting an appropriate support level is critical for discovering meaningful and useful patterns.

- High Support (e.g., 0.2 or 20%):
  This threshold is used to identify highly prevalent and dominant features. High support levels help detect features that are valid for a large portion of the dataset, thereby revealing characteristics that reflect the general market. However, important features associated with niche or low-frequency observations may be overlooked at this level [28].
- Medium Support (e.g., 0.1 or 10%, 0.05 or 5%):
  These levels allow the discovery of both common and moderately frequent attributes. They are ideal for understanding the dynamics of various segments in the NFT market. Trends that are not highly frequent but may be important to investors or collectors often emerge at these thresholds [29].
- Low Support (e.g., 0.03 or 3%, 0.01 or 1%):
  These thresholds are used to detect rarer but potentially unique and valuable features. Especially in high-dimensional, sparse, and heterogeneous datasets like those found in NFT markets, low support values play a critical role in understanding niche market segments. However, using low support thresholds may result in a large number of patterns, which increases computational burden and complicates interpretation [30].

In this study, the use of diversified support thresholds provided an appropriate basis for a detailed analysis of both common and niche features within the NFT market. The impact of each support level on the analysis and the number of resulting itemsets were visualized and interpreted through heatmap outputs presented in this section.

6.5.2 In-Depth Analysis of Itemset Counts and Inter-Cluster Differences Based on Support Levels

Based on the data obtained from the presented heatmap visualizations, the following key findings have been identified:

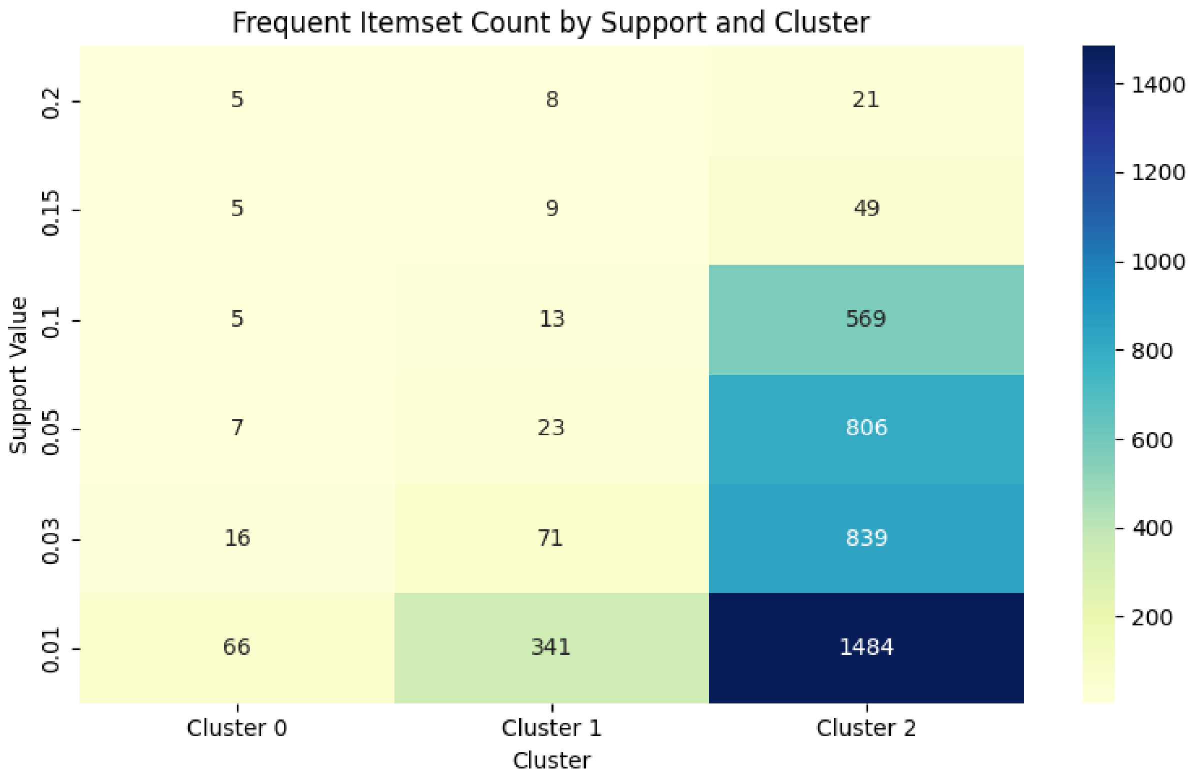

Fig. 6.3. Heatmap visualization of Frequent Itemset Count by Support and Cluster

As the *support* threshold decreases, the number of frequent itemsets detected within each cluster increases exponentially. This rise results from the inclusion of previously less frequent, rare features into the scope of analysis.

Considering that Cluster 1 was dominant in the initial analysis (representing the high-value NFT segment), the increase in the number of itemsets in this cluster as support decreases is particularly significant. Cluster 1 reflects the complexity and diversity of the high-value NFT segment and possesses richer, multi-dimensional feature sets [31].

On the other hand, although the number of itemsets in Cluster 0 and Cluster 2 remains relatively limited, a notable increase is also observed at lower support levels. This indicates that NFTs in different value segments exhibit different feature distributions and supports the heterogeneous structure of the market [32].

Inter-Cluster Similarities and Differences

At high support levels (e.g., 0.2, 0.15), the number of common itemsets across clusters is very low (only one itemset), suggesting that each cluster has distinct and dominant features, and the fundamental segmentation of the market is strong.
As the support level decreases, the number of shared itemsets increases; however, it remains numerically limited. This implies that while the core differences between clusters persist, some rare common features also emerge.

At lower support thresholds (0.05 and below), the differences between clusters become more pronounced. The itemsets emerging at this level are typically rare and niche features, which become distinctive at the cluster level, reflecting the heterogeneous and multi-segmented nature of the NFT ecosystem [29].

Feature Distribution Across Clusters

Cluster 1 represents the richest cluster in terms of both quantity and diversity of itemsets. This supports the idea that the high-value NFT segment is characterized by complex, multi-dimensional, and diverse combinations of attributes—consistent with the findings of earlier analyses [31].
In contrast, Cluster 0 and Cluster 2 have more homogeneous and limited sets of attributes. These clusters represent lower or mid-value NFT segments and differ from Cluster 1 in terms of feature variety and complexity [32].

6.5.3 Comparison and In-Depth Interpretation of Common Features of the High-Value NFT Segment vs. Support-Based Common Features Across All Market Segments

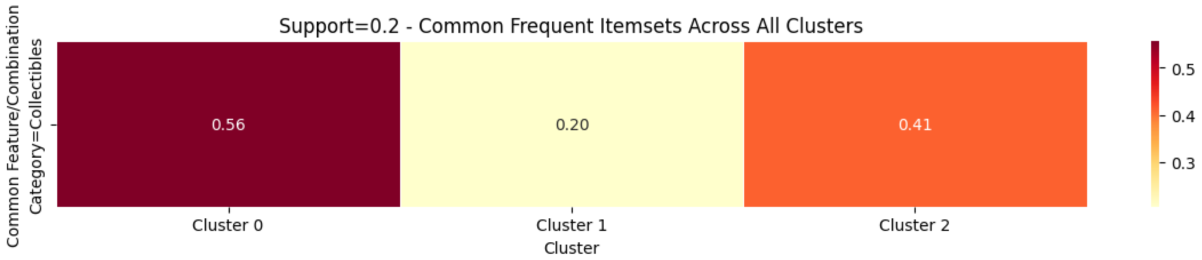

Fig. 6.4. Heatmap Visualization of Common Frequent Itemsets Across All Clusters With Support Value (0.2)

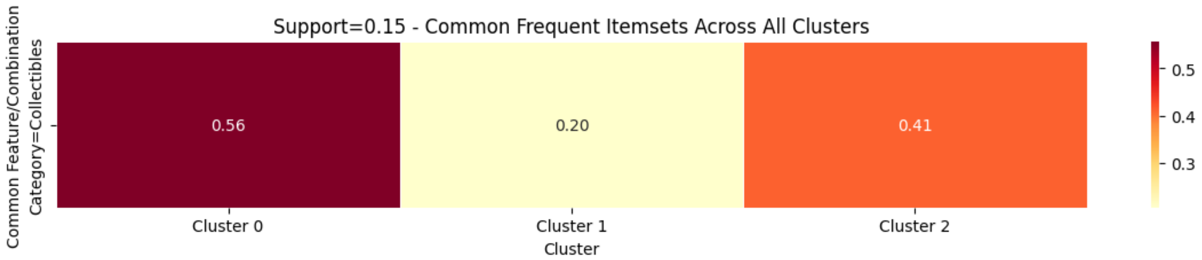

Fig. 6.5. Heatmap Visualization of Common Frequent Itemsets Across All Clusters With Support Value (0.15)

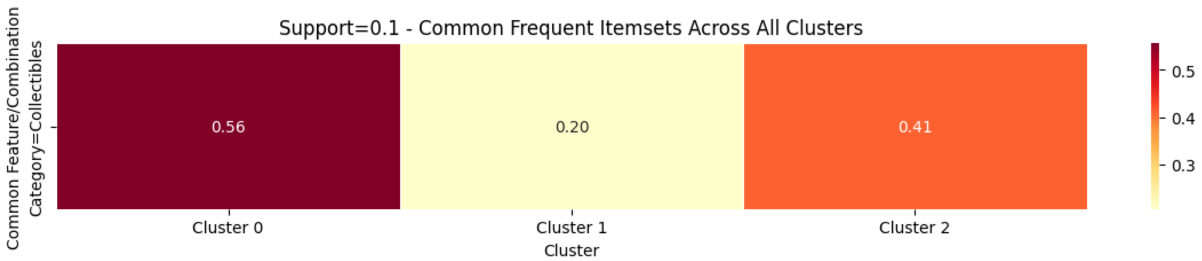

Fig. 6.6. Heatmap Visualization of Common Frequent Itemsets Across All Clusters With Support Value (0.1)

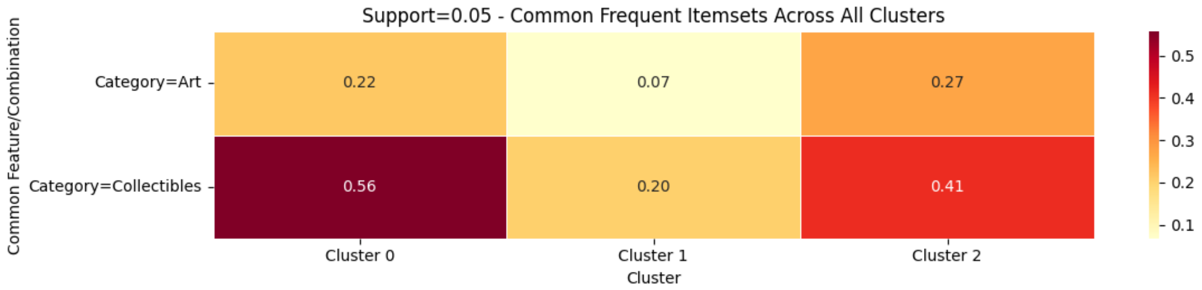

Fig 6.7. Heatmap Visualization of Common Frequent Itemsets Across All Clusters With Support Value (0.05)

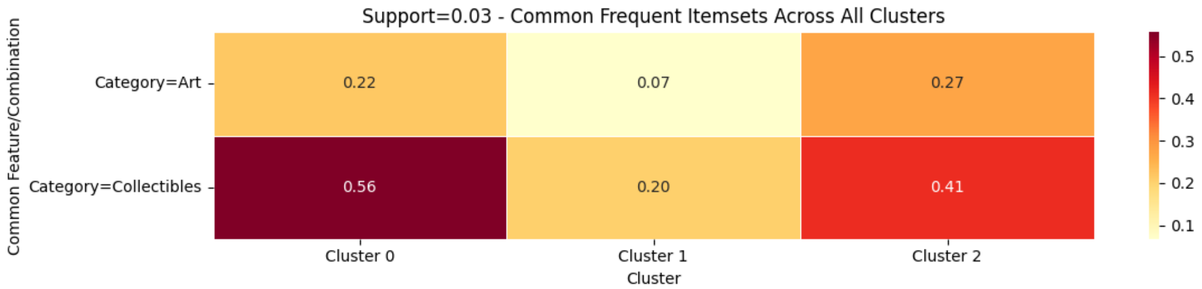

Fig. 6.8. Heatmap Visualization of Common Frequent Itemsets Across All Clusters With Support Value (0.03)

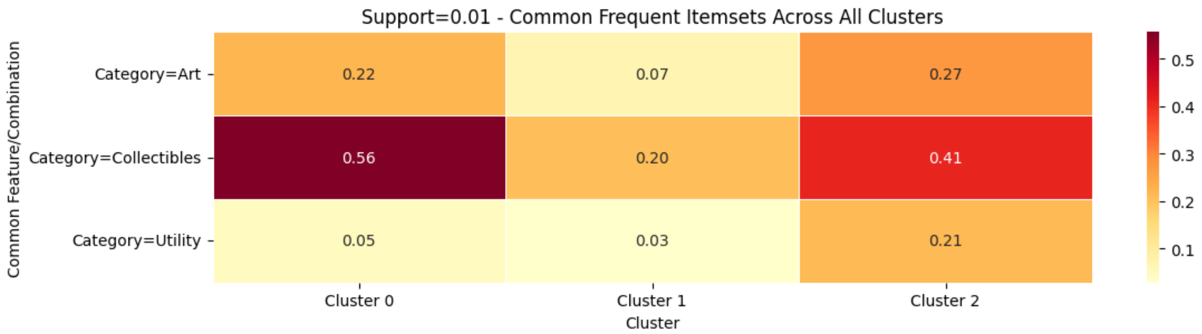

Fig. 6.9. Heatmap Visualization of Common Frequent Itemsets Across All Clusters With Support Value (0.01)

To understand value formation in the NFT market, two main analyses were conducted. The first analysis focused on identifying the common categorical features of high-value NFTs (determined by criteria such as high price and sales volume). In this analysis, high-support and high-frequency features such as *Category=Virtual Worlds*, *payment_token.name=Decentraland MANA*, and *asset.collection.name=Decentraland Wearables* stood out. Notably, the dominance of the "Virtual Worlds" category with a 48.16% support rate indicates the strong position of metaverse-based digital assets in the NFT market [2]. Additionally, the high visibility of Decentraland's native payment token MANA (28.06% support) highlights the significant influence of platform economies on NFT value [33]. Similarly, the Decentraland Wearables collection, which received high support alongside virtual world categories and payment tokens, suggests that NFTs function not only as collectibles but also as functional digital assets [2] [33].

Furthermore, the feature *Category=Collectibles* consistently appeared with a stable support rate of 20.5%. This indicates that traditional NFT collectibles still hold considerable market share and value, though slightly less dominant compared to metaverse-based assets [2][3]. In other words, within the high-value NFT segment, "Virtual Worlds" is more dominant, whereas the "Collectibles" segment maintains a strong but relatively niche or traditional position [2][3].

The second analysis aimed to uncover common itemsets across all NFT market segments by analyzing three distinct clusters and various support thresholds. This approach systematically examined how the shared characteristics of high-value NFTs extend across the broader market, and which new niche features become visible at different support thresholds.

The results confirm that the diversity of common features varies by support level, yet they significantly overlap with the characteristics of the high-value NFT segment. Especially at high support levels (e.g., 0.2 and 0.15), the features mentioned earlier—“Virtual Worlds”, “Decentraland MANA”, and “Collectibles”—reemerge as common itemsets across clusters [2][24]. This indicates that these features are not only critical for a specific segment but are also influential across the broader market [2][34].

However, as the support threshold decreases (e.g., 0.05, 0.03, and 0.01), new and more diverse itemsets become prominent. In this range:

- More niche and specific categories such as *Category=Art* and *Category=Utility* stand out,
- Alternative payment tokens (e.g., USDC, SAND, Gala) and collection names (e.g., The Sandbox, Town Star) become increasingly visible [34]. This clearly demonstrates that NFT value formation is not confined to the high-value segment; instead, the market is multi-layered, heterogeneous, and composed of diverse ecosystems [2][3].

In particular, the consistent presence of the *Category=Collectibles* feature across all support levels is noteworthy. Its visibility in both the high-value NFT segment and in the support-based cluster analyses confirms the stable and enduring role of the collectibles market within the NFT ecosystem [2][3]. Alongside innovative segments like metaverse-based assets, this feature constitutes one of the core pillars of the NFT market.

In summary, when the common features of the high-value NFT segment (core, dominant market trends) are evaluated alongside the support-based common features across all market segments (broad, multi-layered market structure), it becomes clear that NFT value formation is a complex and multi-dimensional process:

- At high support levels, the fundamental and widespread market trends are clearly defined [2][24],
- At low support levels, niche, segment-specific features emerge, and inter-cluster differences become more pronounced [34],
- Both innovative features such as “Virtual Worlds” and “Decentraland MANA”, and traditional segments like “Collectibles”, form the main axes of the NFT market [2][33][3].

These analyses make it evident that NFT investors, collectors, and platform developers must understand not only the highest-value segments but also the broader, multi-layered dynamics of the overall market [2][3].

6.6 General Conclusion and Evaluation

In this section, a comparative analysis of the common characteristics of NFTs was carried out under different minimum support thresholds and across three distinct clusters (low-, medium-, and high-value NFT segments). The FP-Growth-based frequent itemset mining applied in the study enabled us to understand both the overall market trends and the segment-specific dynamics. The findings summarize the complex and multi-layered structure of the NFT ecosystem as follows:

Multi-Layered Market Dynamics
The NFT market exhibits a multi-layered structure that includes both common and dominant features (which emerge at higher support thresholds) and rare, niche, and segment-specific patterns (revealed at lower support levels). High support values highlight general trends shared across all segments, while lower support values uncover smaller yet potentially impactful subgroups and structures based on differentiated user preferences. This indicates that the NFT ecosystem caters not to a single user profile but to a wide range of collectors and investors.

Cluster-Specific Characteristics
The analysis revealed that each cluster — representing low, medium, and high-value NFTs — possesses varying densities and diversity in categorical features. Particularly, Cluster 1, which includes high market value NFTs, produced rich and meaningful itemsets even at low support thresholds, indicating that this segment has a complex structure in terms of both volume and diversity. In contrast, low-value segments (e.g., Cluster 0) showed more limited and less diverse combinations of features. This finding demonstrates that the market differentiates not only based on price but also on structural and categorical richness.

Complementarity of Different Analysis Phases
Two main analytical setups were evaluated as a whole in this study. First, a segment-focused analysis was conducted by selecting only high-value NFTs and extracting their common features using FP-Growth. This analysis revealed which categorical attributes shape high-value NFTs — especially highlighting categories like “Virtual Worlds,” “Decentraland MANA,” and “Collectibles.”

Second, the entire NFT ecosystem was divided into three clusters, and a more comprehensive exploratory analysis was carried out using different support thresholds. This phase demonstrated how commonly shared and segment-specific features vary not just in the top segment but throughout the entire market. As a result, while the first analysis offered an in-depth profile of high-value NFTs, the second illustrated the horizontal expansion, niche differentiation, and feature diversity of the market. When considered together, these two phases reveal that NFT value formation is meaningfully layered both vertically (by price) and horizontally (by features).

Contributions to Application and Strategy Development
These analyses provide several actionable insights for NFT investors, collectors, platform developers, and market analysts. Understanding how common features identified by support levels vary by segment is crucial for designing segment-specific collections, setting pricing strategies, and managing marketing campaigns more effectively.

Moreover, the fact that inter-segment differences are evident even at low support thresholds shows that each cluster requires a customized strategy and user experience design. This study clearly demonstrates that the market cannot be fully understood by focusing solely on “high-value NFTs”; in-depth analysis of niche categories that attract broader user bases is equally essential.

7. Detecting Wash Trading

Detecting wash trading in NFT marketplaces is crucial for ensuring market transparency and preventing price manipulation. Wash trading involves creating artificial trading activity to mislead market participants about the demand and value of assets. This study adopts a multi-step detection framework that combines graph-based transaction analysis with rule-based anomaly detection, tailored to the unique characteristics of NFT ecosystems [35; 36]. The methodology builds upon six key criteria:

1. Identifying suspicious buyer-seller address pairs based on unusually high transaction volumes [37]
2. Detecting bidirectional transactions (where assets repeatedly move back and forth between the same addresses) [38]
3. Capturing high-frequency trades executed within the same day [39]
4. Flagging NFTs that have been transferred more than five times [40]
5. Identifying trades conducted at abnormally low prices relative to historical norms [41]
6. Detecting self-trading events where the buyer and seller are the same entity or tightly connected addresses [42]

Each criterion contributes to a cumulative suspicious_score assigned to involved wallet addresses, which are represented within a directed transaction graph where each edge indicates the flow of assets between addresses, enabling scalable identification of potentially manipulative behaviors.

7.1. Identifying Suspicious Buyer-Seller Address Pairs Based on Transaction Volume

The first step in the detection process involved identifying repeated interactions between the same seller and buyer. Frequent transactions between the same two addresses can be indicative of collusive behavior or attempts to artificially inflate activity. To capture this, I assigned a suspiciousness score that increased proportionally with the number of repeated interactions between a seller-buyer pair. For this analysis, a threshold of more than five transactions was selected to define a pair as suspicious. This value was chosen to balance sensitivity and specificity—flagging potentially manipulative behavior without capturing incidental repeated activity. For more stringent analyses, higher thresholds could be employed to reduce false positives.

```
FOR each edge in transaction_graph:
    IF number_of_transactions_between_addresses > threshold:
        ADD edge to suspicious_edges_list

FOR each (seller, buyer) in suspicious_edges_list:
    suspicious_score[seller] += number_of_transactions
    suspicious_score[buyer]  += number_of_transactions
```

Pseudo code 1. Identification of suspicious buyer-seller pairs based on transaction volume.

This pseudocode outlines the process of flagging suspicious buyer-seller address pairs based on the frequency of their interactions. It iterates over the transaction graph and selects edges (i.e., address pairs) with more than five transactions as potentially manipulative. For each of these suspicious pairs, the suspicious_score for both the buyer and seller addresses is incremented proportionally to the number of transactions. This scoring mechanism helps quantify the likelihood of collusive behavior.
Additionally, the top 15 address pairs with the highest interaction frequencies were extracted to visually inspect potential manipulation clusters.

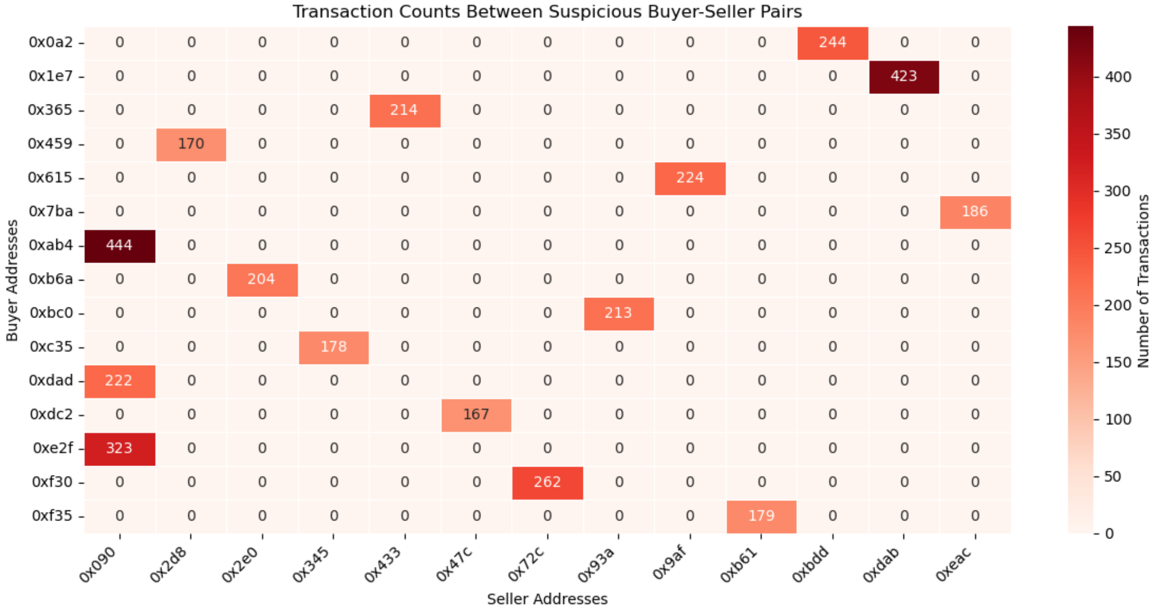

Fig. 7.1. Heatmap of the most suspicious buyer-seller pairs
Note: These scores represent step-specific suspiciousness, not aggregated values across the full detection pipeline.

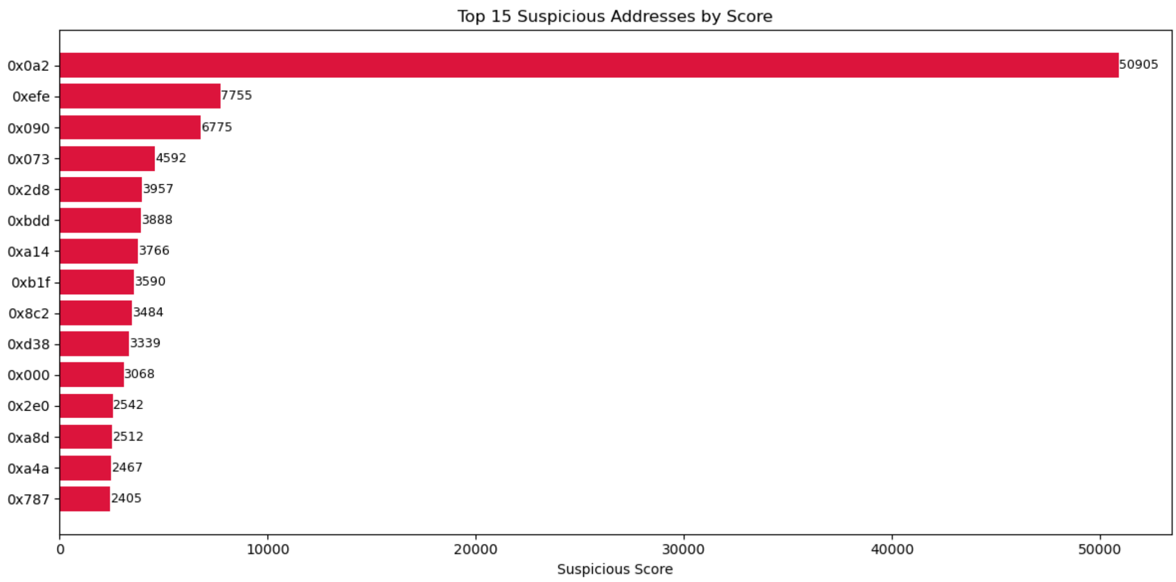

Fig. 7.2. Top 15 addresses with the highest suspicious scores
Note: These scores represent step-specific suspiciousness, not aggregated values across the full detection pipeline.

While address 0x090 stands out in the heatmap for conducting a high number of transactions with specific counterparties—indicating concentrated suspicious activity—address 0x0a2 does not appear as prominently in the heatmap. However, it ranks significantly higher in the suspicious score chart, with a score of 50,905, far exceeding the next most suspicious address.

This suggests that 0x0a2 has interacted with a large number of different addresses, each with over 5 transactions, while intentionally avoiding repeated transactions with the same address. Such a pattern may indicate an attempt to evade detection by spreading out suspicious activity across multiple counterparties, rather than concentrating transactions with a few addresses.

This behavior highlights the importance of using both pairwise heatmaps and global suspicious scores for identifying different types of suspicious patterns—ranging from tight clusters to widespread interaction strategies.

7.2. Bidirectional Transactions

In the second stage, I focused on bidirectional trading behavior—instances where Address A transfers an NFT to Address B, and B later sends an NFT back to A. This kind of reciprocal trading pattern is often a hallmark of wash trading, intended to simulate genuine demand or inflate transaction counts. Any address pair that engaged in such mutual exchanges above a defined threshold was flagged as suspicious, and their associated risk scores were incremented by a fixed value.

The following pseudocode demonstrates how bidirectional transactions are identified by checking for pairs of addresses that exchange NFTs back and forth. When such mutual activity exceeds the defined threshold, both addresses receive a fixed suspicious score increment. Note that the increment value of +2 is chosen for this analysis but can be adjusted in other studies depending on the desired sensitivity.

```
FOR each (u, v) in transaction_graph_edges:
    IF (v, u) exists AND total_transactions(u→v) + total_transactions(v→u) >= 2:
        ADD (u, v) to suspicious_bidirectional_edges_list

FOR each (u, v) in suspicious_bidirectional_edges_list:
    suspicious_score[u] += 2
    suspicious_score[v] += 2
```

Pseudo Code 2. Detection of bidirectional transactions indicating reciprocal trading.

The top 15 most suspicious addresses based on bidirectional transaction scores are visualized in the chart below. This visualization highlights key actors potentially involved in reciprocal trading behaviors.

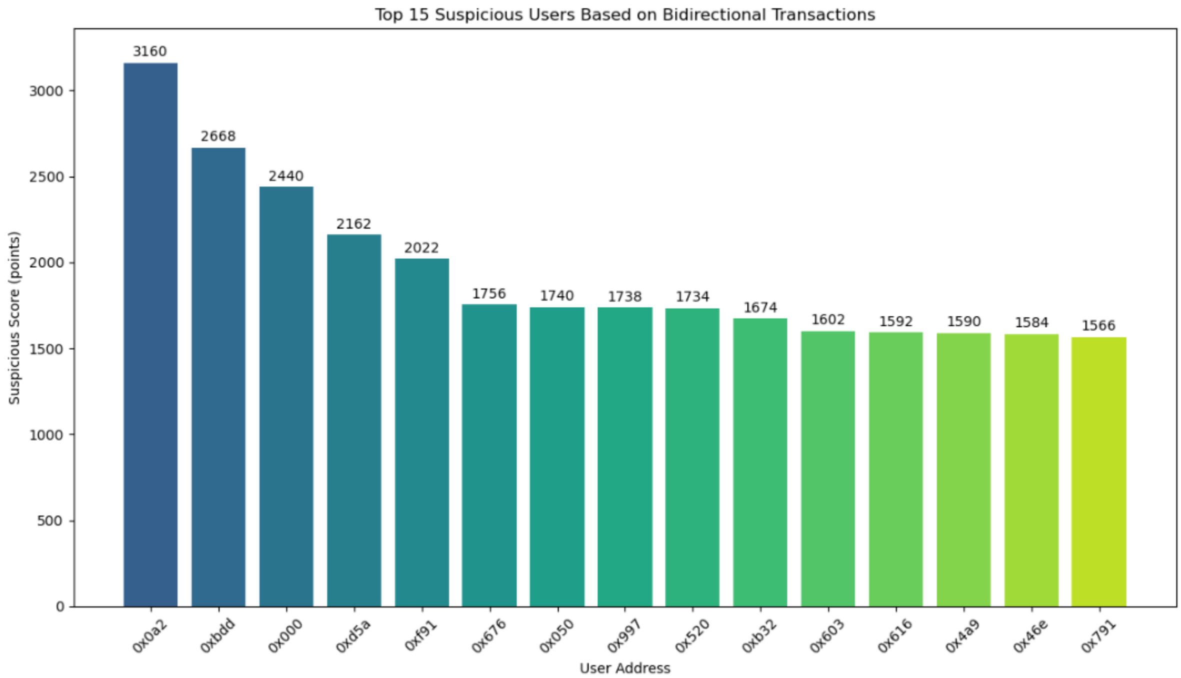

Fig. 7.3. Top 15 suspicious users based on bidirectional transactions, ranked by their suspicious scores.
Note: These scores represent step-specific suspiciousness, not aggregated values across the full detection pipeline.

Address 0x0a2 stands out as the user with the highest suspicious score based on bidirectional transactions, indicating that it has engaged in the most frequent reciprocal interactions among all addresses. This behavior may suggest attempts to simulate legitimate trading activity, create artificial demand, or obscure the true flow of assets by repeatedly transferring tokens back and forth between accounts.

7.3. High-Frequency Same-Day Trades

The third criterion targeted high-frequency trading within a short time window. Wash traders often execute multiple trades in a single day to create the illusion of liquidity. To detect this, I grouped transactions by seller, buyer, and date, and flagged any pair that conducted four or more trades on the same day. This analysis uncovered over 44,878 such instances, with some addresses making more than ten trades with the same counterparty in just one day—strong evidence of potential manipulative behavior.

```
FOR each (seller, buyer, date) group in transactions:
    IF number_of_trades_on_same_day >= treshold:
        suspicious_score[seller] += number_of_trades_on_same_day
        suspicious_score[buyer]  += number_of_trades_on_same_day
```

Pseudo Code 3. Identification of high-frequency same-day trades between address pairs.

This step identifies address pairs conducting trades on the same day that meet or exceed a threshold (set to 4 in this analysis), a pattern indicative of potential wash trading. The suspicious scores are increased by the number of such same-day trades. The threshold and scoring can be adjusted based on different analytical needs.

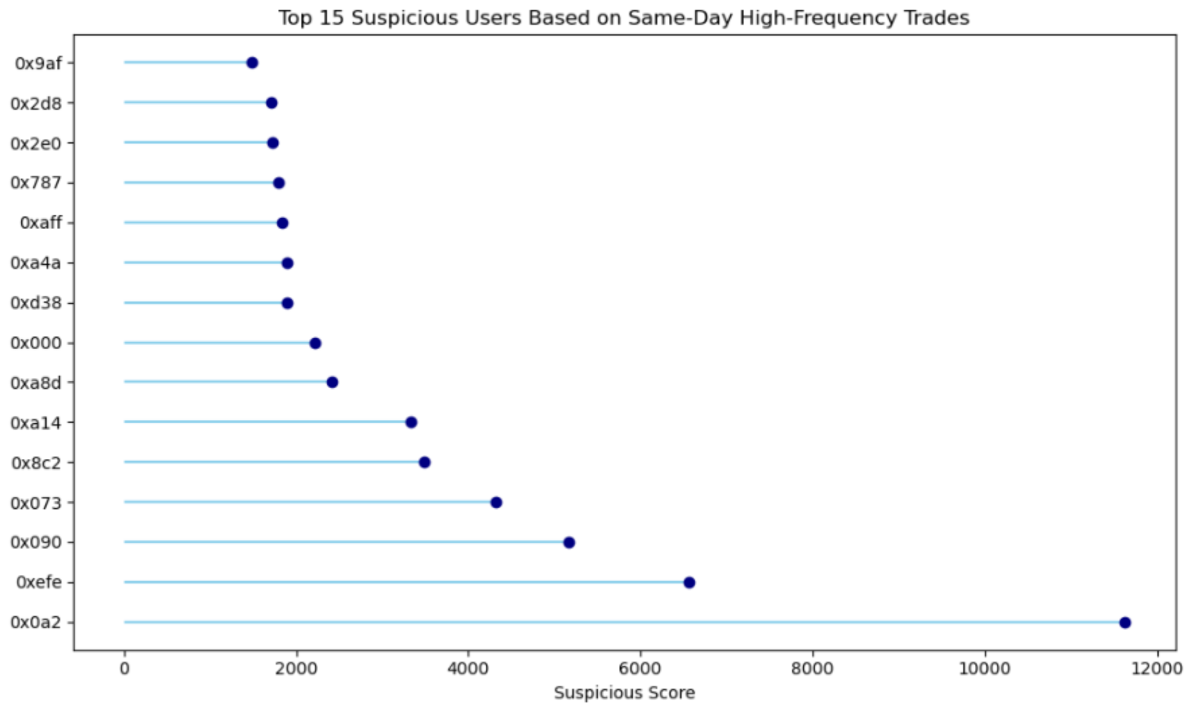

Fig. 7.4. Top 15 suspicious users based on same day high frequency trades, ranked by their suspicious scores.
Note: These scores represent step-specific suspiciousness, not aggregated values across the full detection pipeline.

The bar chart highlights the top 15 addresses involved in high-frequency same-day trading, defined as having at least 4 transactions between the same buyer and seller on a single day. Notably, address 0x0a2 stands out with a suspicious score of 11,631, far exceeding others. This pattern may suggest potential wash trading behavior, where frequent intra-day interactions are used to manipulate market activity or inflate volume. The sharp score drop after the top few addresses indicates a concentration of such behavior among a small subset of users.

7.4. NFTs with More Than Five Transfers

Another strong indicator of wash trading is the repeated circulation of the same NFT. In the fourth step, I tracked NFTs that had changed hands more than five times. When the same addresses appeared in these multiple transfers, their suspiciousness score was increased accordingly. This method helps capture cases where a single NFT is artificially cycled among a small group of wallets to simulate activity and drive up perceived value.

```
FOR each NFT in dataset:
    IF number_of_transfers > threshold:
        FOR each transaction of this NFT:
            suspicious_score[seller] += 1
            suspicious_score[buyer] += 1
```

Pseudo Code 4. Detection of NFTs with more than five transfers indicating repeated circulation.

This step flags NFTs that have been transferred more than a defined threshold (5 in this analysis). Each transaction involving these NFTs increases the suspicious scores of the corresponding sellers and buyers, capturing repeated cycling of assets.

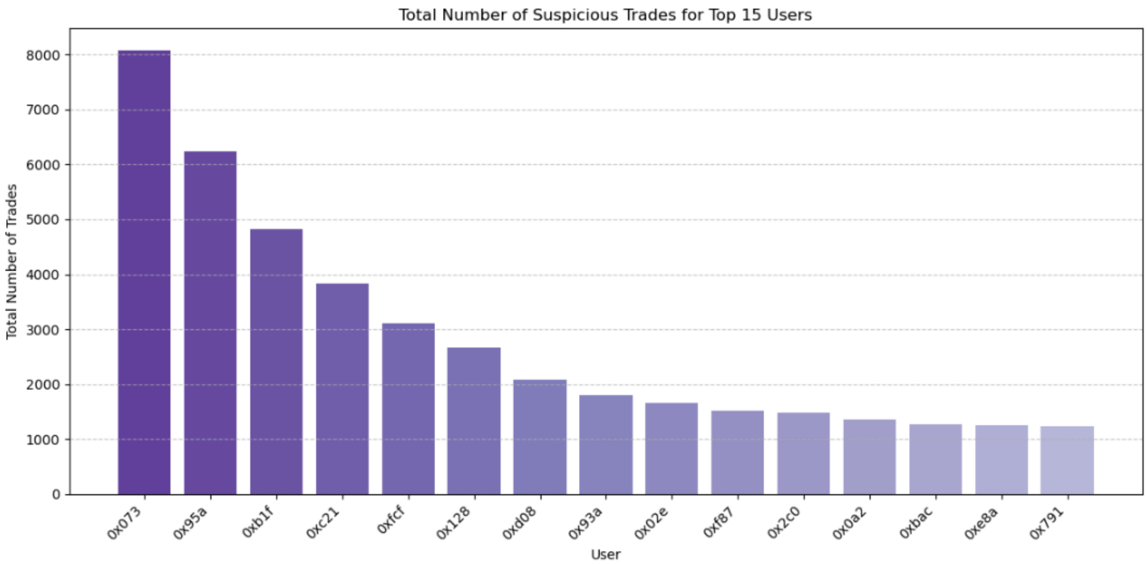

Fig. 7.5. Top 15 suspicious users based on repeated NFT transfers, ranked by their suspicious scores.
Note: These scores represent step-specific suspiciousness, not aggregated values across the full detection pipeline.

Address 0x073 emerges as the most suspicious user based on repeated NFT transfers, with a notably high score of 8,070 points. This indicates frequent involvement in NFTs that have circulated extensively, suggesting potential attempts to artificially inflate activity or value. Following closely are addresses like 0x95a and 0xb1f, with scores of 6,229 and 4,819 points respectively, highlighting their significant participation in these high-transfer NFTs. The distribution of scores across the top 15 users reflects a pattern where a relatively small group of addresses is heavily engaged in cycling NFTs multiple times, a common wash trading indicator aimed at simulating market demand and activity.

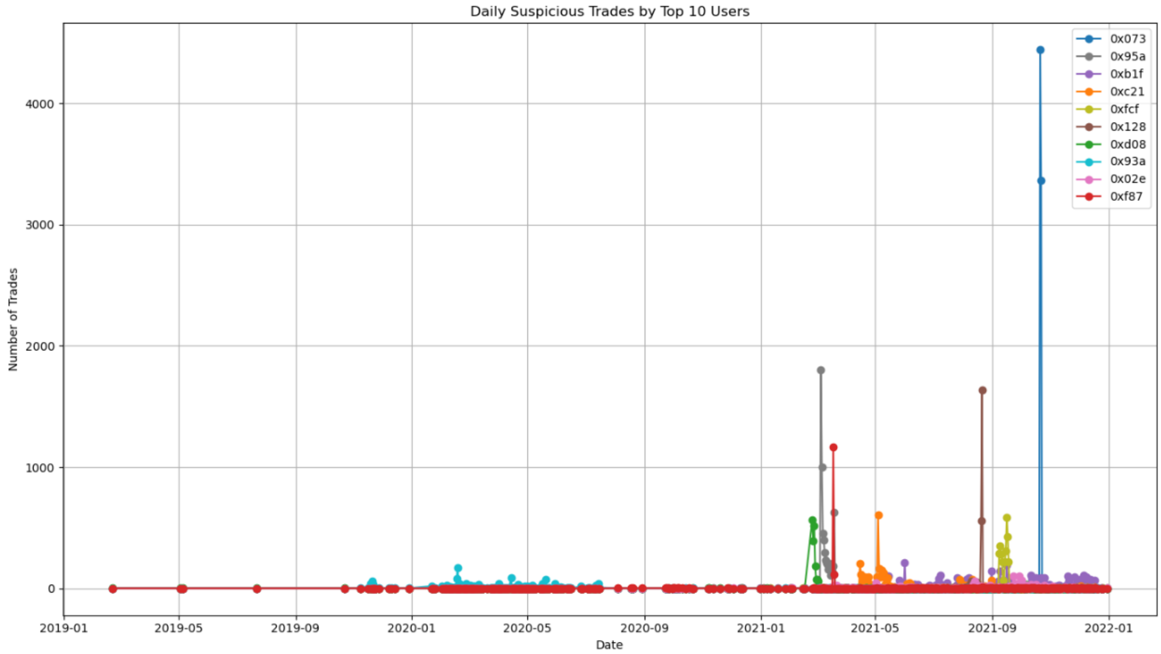

Fig. 7.6. Number of daily trades based on repeated NFT transfers.
Note: These scores represent step-specific suspiciousness, not aggregated values across the full detection pipeline.

The graph depicts the daily number of suspicious trades executed by the top 10 most active users. It reveals extreme and highly suspicious activity by several addresses. Notably, address 0x073 executed over 4,000 and 3,000 suspicious trades on single days—an implausible volume for any legitimate user or seller, strongly suggesting an automated, bot-driven wash trading or self-trading operation. Similarly, 0x095a conducted nearly 2,000 suspicious trades in a single day, further reinforcing suspicions of artificial activity. Address 0x128 also recorded around 1,600 trades, indicating that such behavior is not isolated but may be part of a broader, coordinated manipulation pattern.

7.5. Suspiciously Low Pricing

The fifth component examined transaction pricing outliers. Extreme transaction values—either unusually high or low—are commonly used in wash trading schemes to manipulate floor prices and perceived market trends. Transactions falling within the top or bottom 1% of all trade values were flagged, and all participating addresses were penalized with additional suspiciousness points.

```
SET min_transactions_threshold = 10  # Minimum transactions to consider suspicious
SET low_price_threshold = 0.05       # Threshold for suspiciously low total price
(based on data distribution)
```

```
FOR each (seller, buyer) pair in pair_stats:
    IF total_transactions >= min_transactions_threshold AND
       total_price_sum < low_price_threshold:
        suspicious_score[seller] += 1
        suspicious_score[buyer] += 1
```

Pseudo Code 5. Identification of suspiciously low total price transactions between address pairs.

This step identifies seller-buyer pairs with at least 10 transactions whose combined transaction value falls below a low-price threshold (0.05 in this analysis, selected based on the observed data distribution minimum). Such low total pricing may indicate attempts to manipulate market perception, and suspicious scores are incremented accordingly.

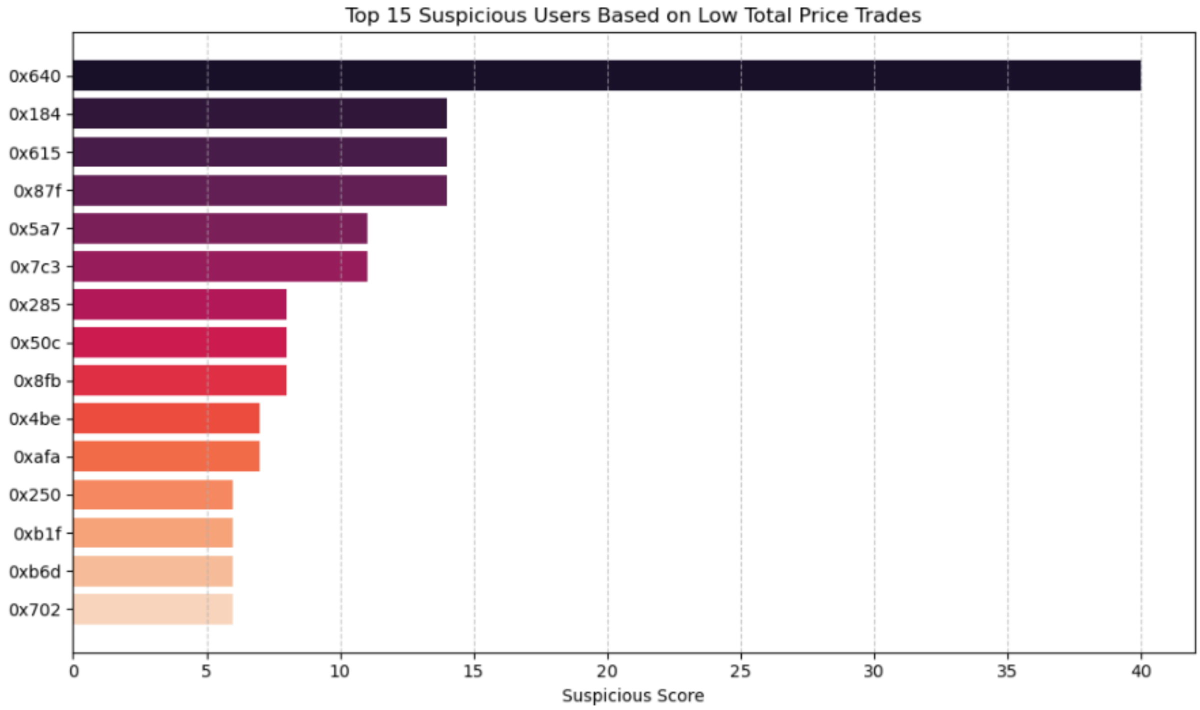

Fig. 7.7. Top 15 suspicious users based on low total price trades, ranked by their suspicious scores.
Note: These scores represent step-specific suspiciousness, not aggregated values across the full detection pipeline.

The horizontal bar chart displays the top 15 users flagged for suspiciously low total price trades. However, the relatively low scores and the modest range of values suggest that this indicator alone may not strongly differentiate truly suspicious actors. This pattern implies that low-priced transactions, while potentially indicative of wash trading, are not a prevalent or highly distinctive behavior within the dataset. Consequently, relying solely on this metric may offer limited insight, highlighting the importance of combining multiple detection methods for more robust wash trading identification.

7.6. Self-Trading

Finally, the most explicit form of wash trading was addressed: self-trading, where the buyer and seller are the same address. Although rare, such transactions were identified in the dataset. These instances were immediately marked as highly suspicious and assigned the maximum possible score, as they represent clear-cut manipulation with no plausible legitimate intent.

```
FOR each transaction in dataset:
    IF seller_address == buyer_address:
        suspicious_score[seller_address] += 10
```

Pseudo Code 6. Detection of self-trading where buyer and seller addresses are the same.

This code detects transactions where the buyer and seller are the same address, a clear indicator of wash trading. Because self-trading is highly suspicious, the suspicious score is increased by a larger amount (10 points) compared to other criteria, reflecting the severity of this behavior. The following chart visualizes the top 15 most suspicious addresses based on self-trading activity.

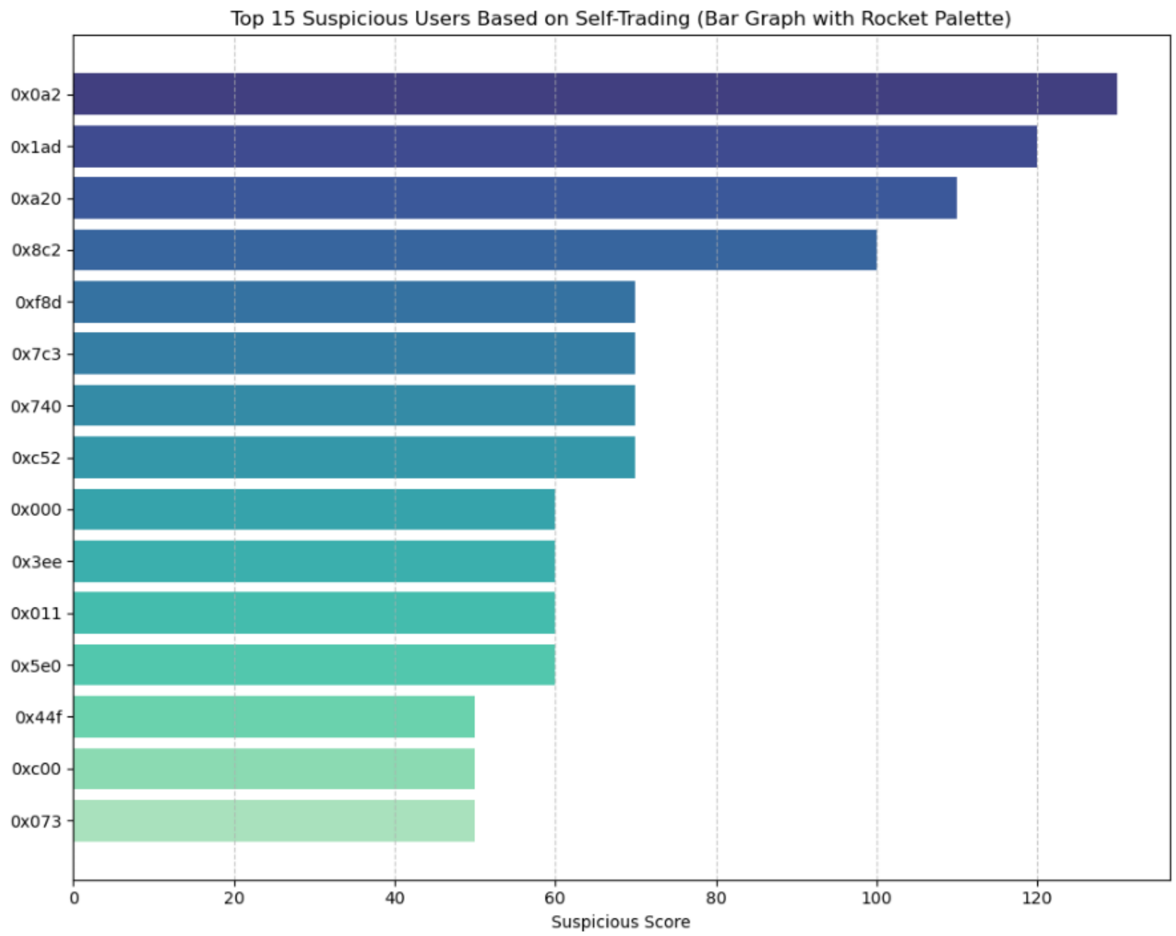

Fig. 7.8. Top 15 suspicious users based on self trading, ranked by their suspicious scores.
Note: These scores represent step-specific suspiciousness, not aggregated values across the full detection pipeline.

The self-trading analysis identified a total of 1,220 transactions where buyers and sellers are the same address, underscoring that while self-trades are relatively rare compared to total transactions, they still represent a significant amount of suspicious activity. Among these, a concentrated set of addresses show especially high suspicious scores, with address 0x0a2 leading at 130 points. This pattern strongly indicates deliberate self-trading behavior, a clear and explicit form of market manipulation, with the top 15 addresses accounting for the majority of such suspicious transactions in the dataset.

After applying all six detection methods, the cumulative suspiciousness scores reveal the most likely participants in wash trading activities. The following analysis focuses on these top 15 addresses, highlighting patterns that suggest coordinated and manipulative behavior in the NFT marketplace.

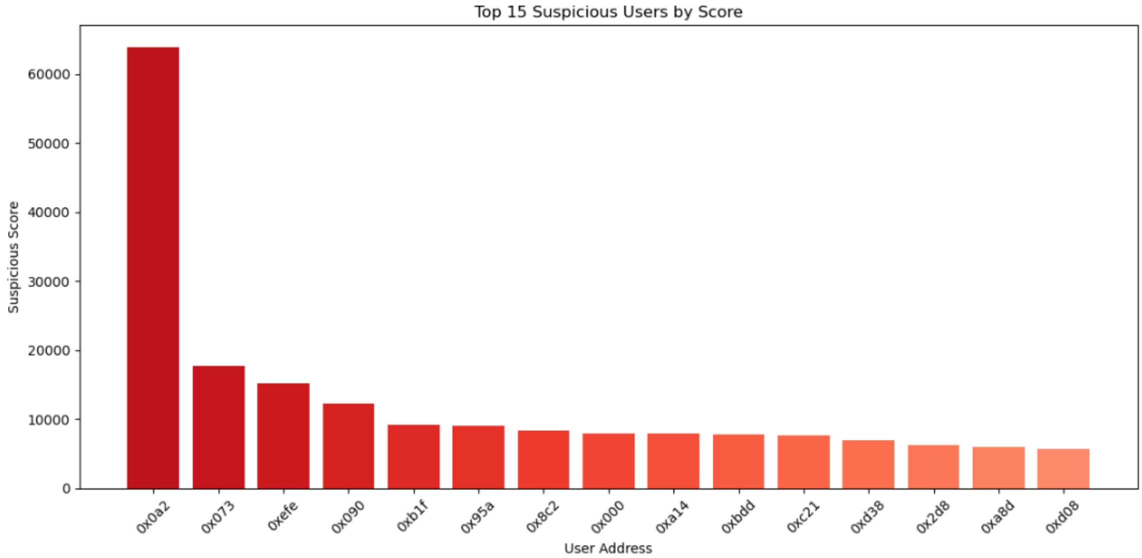

Fig. 7.9. Top 15 suspicious users based on their suspicious scores.

The cumulative suspiciousness scores clearly highlight a small group of addresses with disproportionately high levels of suspicious activity. Address 0x0a2 stands out with an exceptionally high score of 63,873, far surpassing the next most suspicious address (0x073) at 17,779. This extreme gap suggests that 0x0a2 was involved in nearly every type of wash trading behavior detected—possibly engaging in frequent self-trading, repetitive high-frequency transactions, and multiple suspicious NFT transfers.

Other addresses such as 0xefe, 0x090, and 0xb1f also consistently appeared across multiple detection steps, indicating a pattern of persistent, multifaceted manipulation. The close clustering of scores among addresses ranked 5th to 15th (ranging roughly between 5,000 and 9,000) suggests the existence of a secondary tier of coordinated actors, likely participating in shared schemes or networks.

Overall, this distribution supports the hypothesis that wash trading in the dataset is not random or isolated, but rather concentrated among a few highly active and interconnected participants. These findings highlight the need for further investigation into these accounts and the potential application of automated detection methods to monitor such manipulative behavior in NFT markets.

In addition to rule-based and graph-driven approaches, statistical anomaly detection algorithms can further enhance the identification of wash trading. For instance, *Isolation Forest* is an ensemble-based method that isolates anomalies by randomly partitioning the dataset; since anomalies are few and different, they tend to be isolated faster than regular points. This makes the algorithm highly efficient for high-dimensional transaction data. Similarly, the *Local Outlier Factor (LOF)* algorithm detects anomalies by comparing the local density of a point to that of its neighbors, flagging instances that deviate significantly from their surrounding context. Although these methods were not implemented in the current study due to time constraints, they represent promising avenues for future work aimed at uncovering more subtle or non-obvious patterns of fraudulent activity in NFT trading networks.

8. Analyzing Seasonal and Periodic Changes in the NFT Market

8.1 STL Decomposition of Weekly Total NFT Sales

To gain a deeper understanding of the temporal characteristics of NFT market activity, we performed a time series decomposition using STL (Seasonal-Trend decomposition using Loess) on the weekly aggregated total sales volume. STL is a robust method that separates a time series into three interpretable additive components: trend, seasonal, and residual. This decomposition not only aids in uncovering long-term structural patterns but also in detecting short-term volatility and irregularities in the market [43].

- The trend component reveals the overarching direction of the NFT market over time. In our analysis, this component captures the gradual shifts in the market's baseline—such as slow growth periods, plateaus, or declines—independent of short-term noise. It is particularly useful for observing phases of market expansion, hype cycles, or saturation.

- The seasonal component identifies systematic, recurring fluctuations within a fixed period—in this case, a 52-week cycle, reflecting annual repetition. These patterns may stem from behavioral factors such as recurring art events, crypto-related market cycles, tax season effects, or coordinated release schedules of popular NFT projects. The seasonal component allows us to isolate these repeatable patterns and interpret when and how the market tends to experience regular peaks and troughs throughout the year [44].

- The residual component captures the portion of variation in the data not explained by trend or seasonality. This includes one-time events, anomalies, speculative bubbles, sudden market crashes, or any unpredictable movements. By separating these components, we can more clearly observe when the market deviates from its expected behavior, which can be especially valuable for detecting exogenous shocks or speculative spikes.

In summary, applying STL to the weekly NFT sales data enables a layered interpretation of market behavior. It distinguishes between structural trends, seasonal habits, and irregular shocks, offering a more nuanced view of the NFT ecosystem's dynamics. This decomposition enhances our ability to track market maturity, identify temporal opportunities or vulnerabilities, and support further time-aware forecasting or anomaly detection efforts.

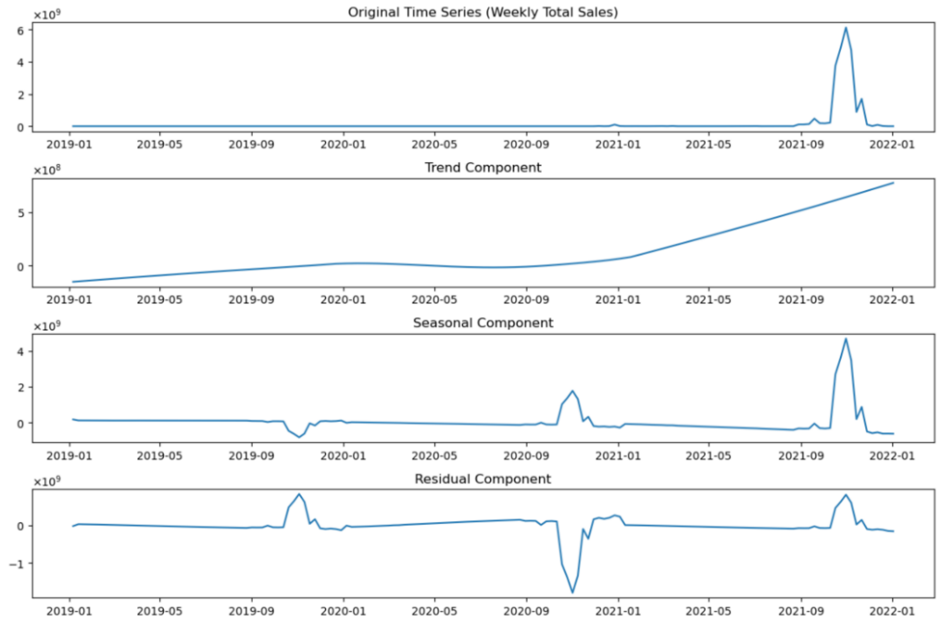

Fig. 8.1. STL Decomposition of Weekly Total NFT Sales into Trend, Seasonality, and Residual Components

Interpretation

The figure presents the decomposition of weekly total NFT sales using the STL method, which separates the original time series into three additive components: trend, seasonal, and residual.

1. **Original Time Series (First Panel)**
   The raw weekly sales data shows extremely low activity from early 2019 to mid-2021, followed by a sudden and sharp spike in Q3 2021. This spike observed in late 2021 in the weekly total NFT sales series aligns closely with several real-world developments that collectively contributed to a short-term explosion in NFT market activity. One of the most prominent catalysts was Facebook's rebranding as "Meta" in October 2021, signaling a strong corporate pivot toward the metaverse. This announcement significantly elevated public interest in digital assets, with NFTs seen as foundational elements of digital ownership in virtual worlds.

   Additionally, major global brands such as Nike, Adidas, and Coca-Cola made high-profile entries into the NFT space during this period, fueling investor enthusiasm and widespread media attention. This triggered a FOMO-driven buying wave among retail participants. Concurrently, the price of Ethereum—the primary blockchain for NFT transactions—also surged, amplifying the USD-denominated value of NFT sales, even if transaction volumes remained steady.

   The explosive popularity of Bored Ape Yacht Club (BAYC) and similar collections also played a critical role, with record-breaking secondary sales and celebrity endorsements pushing transaction volumes to all-time highs. This confluence of speculative hype, corporate validation, and rising crypto valuations formed the basis for the exceptional peak observed in the time series, making it a textbook example of bubble-like short-term growth in emerging digital asset markets.

   (Note: Q3 refers to the third quarter of the year—July through September; Q4 refers to the fourth quarter—October through December.

2. **Trend Component (Second Panel)**
   The trend component extracted by STL highlights the underlying long-term trajectory of the NFT market by filtering out short-term fluctuations and seasonal noise. In this case, the trend line shows a slow and steady incline starting around mid-2019, reflecting the early but limited adoption phase of NFTs. This is followed by a more pronounced acceleration beginning in late 2020, marking the onset of a sustained growth phase in the ecosystem.

   Unlike the raw time series, which is dominated by sudden spikes and volatility, the trend component provides a smoothed representation of structural market behavior. It reveals that the explosive activity observed in late 2021 was not purely a short-lived anomaly, but rather part of a broader upward shift that had already begun months earlier. This pattern may be attributed to increasing mainstream awareness, improved platform infrastructure, and growing institutional and developer interest in NFTs.

   Overall, the trend component serves as a valuable indicator of market maturity and directionality, helping to separate short-term hype from enduring growth. Its upward slope throughout 2021 supports the conclusion that the NFT market was undergoing a period of rapid structural expansion, not just speculative turbulence.

3. **Seasonal Component (Third Panel)**
The seasonal component represents systematic, calendar-based fluctuations that recur at regular intervals—typically on a yearly basis. In the early part of the observed period (2019–2020), this component exhibits relatively modest and stable oscillations, suggesting that NFT transaction volumes were not yet influenced by strong seasonal patterns.

   However, from late 2021 onward, the amplitude of seasonal variation becomes more pronounced. These sharper fluctuations likely reflect the emergence of market-driven cycles, such as increased trading around high-profile NFT drops, holiday periods, or end-of-year investment activity. The alignment of seasonal peaks with spikes in the original time series reinforces this interpretation, indicating that certain months—or even specific quarters—consistently experience elevated transaction activity.

   This evolving seasonal structure suggests that as the NFT market matures, it begins to exhibit behavioral patterns similar to more established financial or consumer markets, shaped by both psychological and institutional factors.

4. **Residual Component (Bottom Panel)**
The residual component captures the portion of the time series that cannot be attributed to either long-term trends or seasonal cycles. These irregular fluctuations often arise from unexpected market behavior, external shocks, or data-specific noise.

   In the context of this analysis, notable deviations can be observed in early 2020 and late 2021. The fluctuations in early 2020 may correspond to global uncertainty and financial disruptions triggered by the onset of the COVID-19 pandemic, while the pronounced volatility in late 2021 likely reflects speculative surges, media-driven hype, or sudden changes in investor sentiment during the NFT boom.

   By isolating these irregular components, the STL decomposition allows for a clearer understanding of short-term anomalies, distinguishing them from structural or recurring patterns. This helps us better interpret the NFT market's sensitivity to external events and reinforces the idea that not all fluctuations are driven by predictable mechanisms.

In summary, the STL decomposition clearly illustrates that the surge in NFT sales observed in 2021 was primarily driven by a robust upward trend, reflecting sustained market expansion. Seasonal effects—such as recurring investor behaviors or event-driven cycles—also played a supporting role, while the residual component captured short-term irregularities and bursts of volatility. By breaking the time series into distinct components, STL provides a structured framework for identifying the underlying dynamics of the NFT market, enhancing both interpretability and analytical depth.

8.2 Time Series Heatmaps

In order to explore temporal shifts in the NFT market, we conducted a detailed time-based aggregation of transaction data by calculating monthly total sales volumes in USD [45]. The goal was to uncover not only general market trends but also periodic behaviors such as seasonal surges, annual growth phases, or stagnation periods. For this purpose, transaction records were grouped by both the year and month fields extracted from each NFT's sales timestamp. This grouping allowed us to build a time series matrix that reflects the intensity of trading activity for each month across multiple years [46].

To make these patterns visually interpretable, we employed a heatmap representation where each cell indicates the aggregate USD value of transactions for a given month and year. Color gradients correspond to the relative magnitude of these volumes—darker or warmer tones signify higher levels of market activity, while lighter tones indicate periods of reduced engagement.

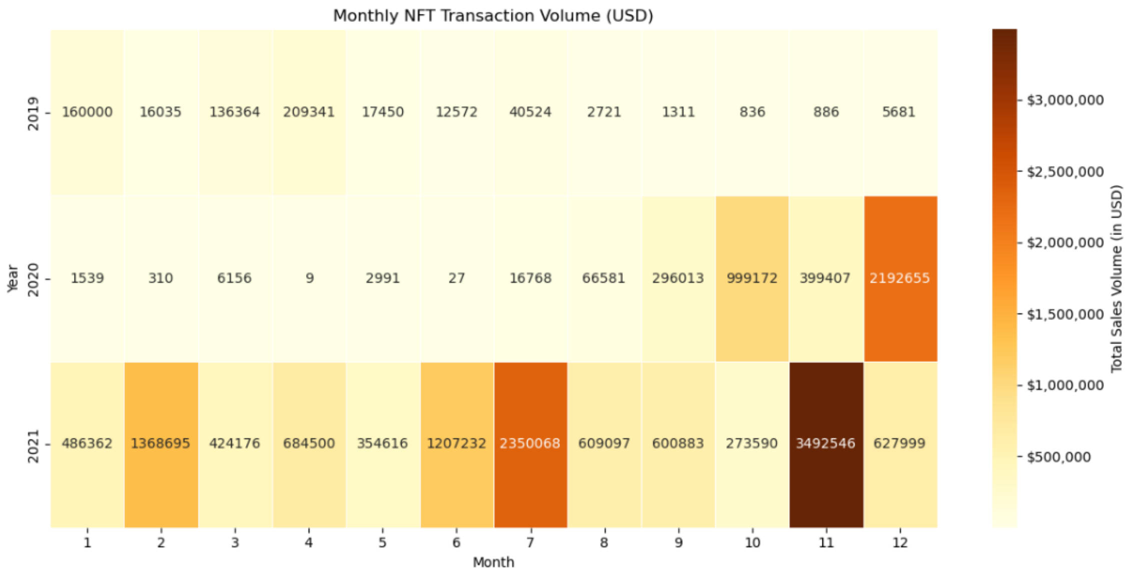

Fig. 8.2. Heatmap showing monthly total NFT sales in USD across years

Interpretation of Monthly NFT Transaction Volume (USD)

The time series heatmap provides a visual summary of the monthly NFT transaction volumes in U.S. dollars across multiple years, capturing the evolving dynamics of the market. Each cell in the matrix represents the total sales volume for a given month and year, allowing for an intuitive comparison across both short and long temporal scales.

1. Low Activity Period (2019 – Early 2020):
During this initial phase, the NFT market exhibited minimal transaction activity. Monthly volumes remained consistently low, often not exceeding $200,000. This period reflects the early, exploratory stage of the NFT ecosystem, where public awareness and institutional interest were limited. The absence of pronounced peaks suggests a lack of major market events or product launches, and most transactions were likely confined to niche communities.

2. Early Growth and Acceleration (Mid to Late 2020):
Starting around August 2020, there is a visible uptick in transaction volume, particularly during the last quarter of the year. This growth corresponds with increased media coverage of blockchain art, rising popularity of Ethereum-based platforms like OpenSea, and early signals of institutional curiosity. December 2020 saw a dramatic spike in volume, exceeding $2 million. This inflection point may indicate the transition from an experimental phase into the early stages of broader market adoption.

3. Market Boom and Maturity (2021):
The year 2021 stands out as a landmark period characterized by explosive growth. Transaction volumes surged across nearly every month, with particularly high activity in February, July, and November. The peak was observed in November 2021, when the market recorded over $3.4 million in total USD transaction volume—an unprecedented high. This expansion can be attributed to a confluence of factors:

- The surge of mainstream attention from celebrity-backed NFT drops.
- Strategic partnerships and entry of global brands like Adidas and Coca-Cola into the space.

- Continued momentum from flagship collections such as Bored Ape Yacht Club (BAYC) and Art Blocks.
- Favorable conditions in the broader cryptocurrency market, which likely boosted purchasing power for NFT investors.

This heatmap serves as a powerful tool to visualize the timeline of the NFT market's expansion. It highlights how a modest and relatively unknown sector transformed within a short span into a vibrant, high-volume marketplace. The data reveal key moments of acceleration and help contextualize the broader trends influencing NFT adoption. Importantly, these temporal insights are essential for stakeholders seeking to forecast future growth patterns, identify strategic entry points, or better understand the forces behind market volatility [47].

8.3 Clustering per Time Period

To investigate how transaction characteristics within the NFT market evolved over time, a year-wise unsupervised clustering analysis was performed using the K-means algorithm[48]. The goal was to uncover natural groupings of NFT transactions based on two primary indicators: total_price, representing the monetary value of the transaction, and asset.num_sales, indicating the frequency of sales associated with a given NFT [49].

Each year from 2019 to 2021 was analyzed independently to ensure that the clustering captured temporal changes in market structure without interference from inter-year variation. Before applying the clustering algorithm, the selected features were standardized using StandardScaler, which transforms the data to have zero mean and unit variance. This normalization step is crucial to avoid scale dominance, especially since total_price often exhibits high variability and skewness.

To enhance the interpretability of the clusters and reduce the effect of extreme values, logarithmic transformation (log1p) was applied to the total_price variable. This approach compresses outliers and reveals underlying trends more clearly.

In each year, the number of clusters was fixed at k = 3, balancing interpretability with segmentation granularity. The resulting clusters were labeled based on the average total price within each group:

- Low Value NFTs
- Medium Value NFTs
- High Value NFTs

This labeling schema enables us to qualitatively compare how different transaction types—such as high-value but infrequent trades versus low-value, high-frequency activity—are distributed within each year. Scatter plots were generated for each year's clusters, providing a visual interpretation of how NFT transaction patterns varied over time.

By segmenting the data in this way, the analysis reveals:

- Shifts in market concentration (e.g., more high-value trades emerging in 2021),
- Changes in sales frequency patterns,
- Potential maturation in user behavior, such as bulk trading or investment-style acquisitions.

8.3.1. Yearly Clustering of NFT Transactions and Pre-Visualization Analysis

To analyze how NFT transaction patterns differ across time, we applied K-means clustering separately for each year between 2018 and 2025. Key features such as total_price and asset.num_sales were used to identify distinct groups of transactions that may reflect different levels of market activity, from low-value purchases to high-value trades. All features were standardized beforehand to ensure fair comparisons across years.

This year-by-year clustering approach allows us to track how transactional behaviors evolve as the NFT market matures, revealing whether similar patterns repeat or shift over time. Before generating the visual representation of these clusters, we first examined the distribution of data points across clusters for each year. This intermediate output provides early insight into the dominance or rarity of certain transaction types in different years, and serves as a foundation for deeper interpretation through cluster-based visualizations.

8.3.2. Visualization of NFT Clusters by Transaction Value (2019–2021)

To explore how NFTs cluster based on transaction value and sales count, we applied K-means clustering to each year between 2019 and 2021 using total_price and asset.num_sales as features. These two features were chosen because they jointly capture both monetary value and transaction frequency—two core dimensions of NFT market activity.

The number of clusters was dynamically set to a maximum of 3, depending on the data size for each year. Using a maximum of 3 clusters provided a balance between interpretability and granularity in segmentation. This also ensured that clustering could still be applied even when some years had limited records.

Each cluster was labeled as Low, Medium, or High Value NFTs based on average transaction price where possible. The resulting cluster compositions illustrate how different types of NFT activity (e.g., high price but low sales vs. low price but frequent sales) were distributed across the market over time. This approach helps identify annual shifts in market segmentation and buyer behavior, while adapting to the data available for each year.

Why Use Log Transformation?

The total price values vary across several orders of magnitude—from small sales to very expensive transactions. To handle this skewed distribution and ensure that clustering isn't biased toward extreme values, we applied a log(1 + total_price) transformation. This compresses the scale, spreads out dense low-value data, and makes patterns more visible and interpretable in the scatter plot.

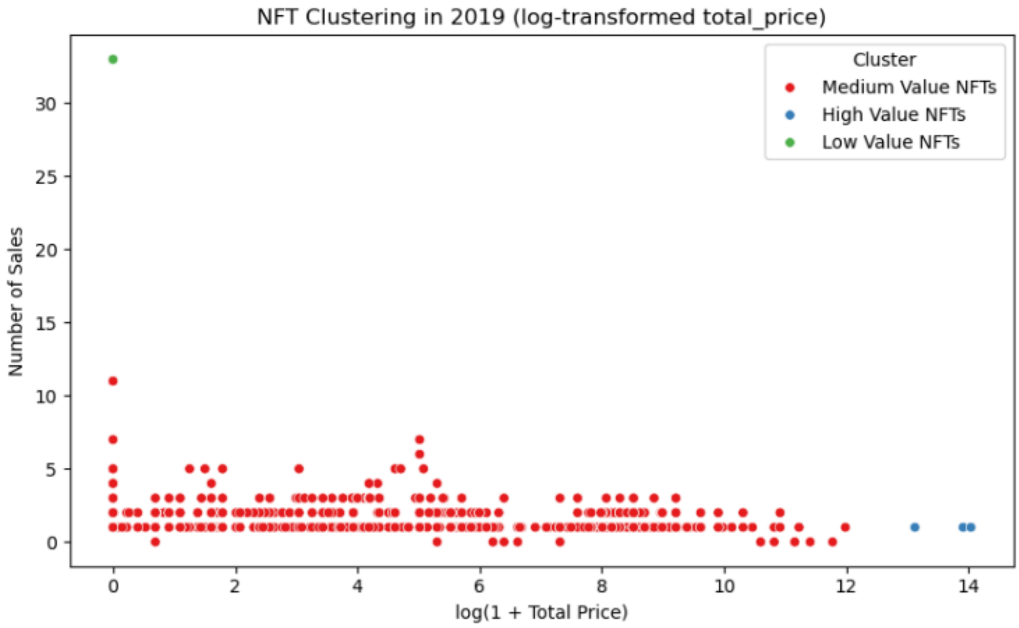

Fig. 8.4. Scatter plot of 2019 NFT transactions clustered by K-means, using log-transformed total price and number of sales.

The scatter plot above illustrates the results of K-means clustering for NFT transactions that occurred in 2019, using log-transformed total price and number of sales as the primary axes. The transformation applied to the price axis (log(1 + total_price)) was necessary to compress the heavy-tailed price distribution typical of early NFT markets, enabling clearer separation of clusters.

The majority of data points fall within the Medium Value NFTs cluster (shown in red), concentrated between approximately log(1 + total_price) = 2 to 10, and with sales frequencies generally below 5. This indicates that in 2019, most NFTs were sold at modest prices and with limited sales activity, reflecting the early-stage nature of the NFT market at that time.

A small group of points classified as High Value NFTs (in blue) is positioned on the far right of the plot, representing transactions with very high prices but very low sales counts (typically 1 or 2 sales). These outliers suggest a small number of high-value, possibly speculative or collector-driven purchases, which were atypical for the broader 2019 market.

Interestingly, only one outlier appears in the Low Value cluster (in green), positioned near log(1 + total_price) ≈ 0, but with an exceptionally high number of sales (above 30). This could reflect a single low-cost NFT that was traded repeatedly, possibly as part of promotional or experimental usage patterns prevalent during the nascent phase of the ecosystem.

Overall, the 2019 clustering visualization indicates a market largely dominated by mid-range, low-frequency sales, with very few examples of either high-volume trading or premium-priced assets. This pattern is consistent with a pre-boom NFT environment, where both investor interest and asset valuation were still in their formative stages.

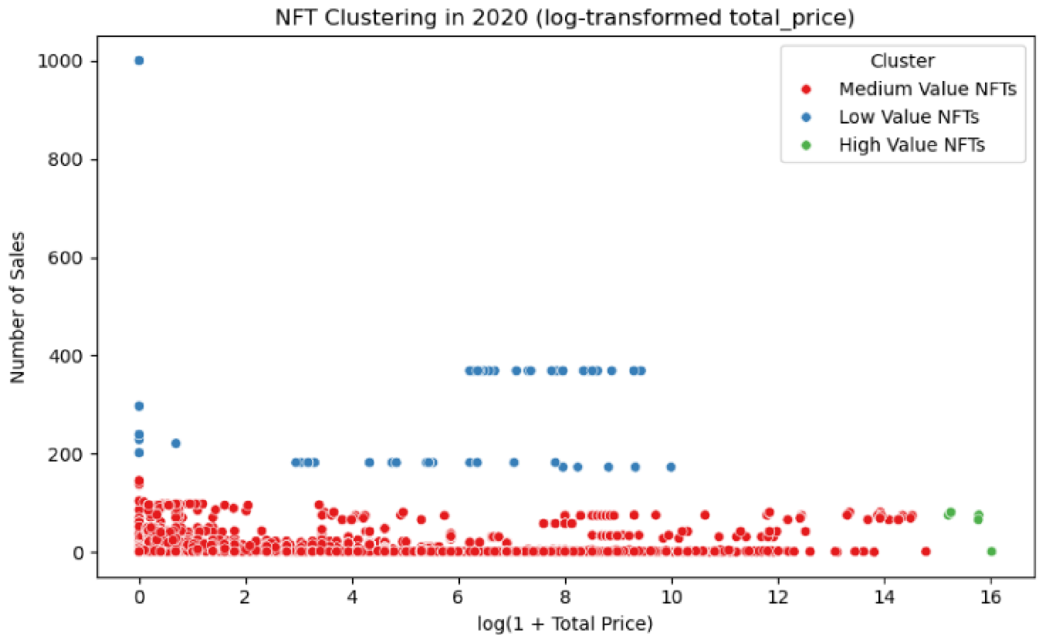

Fig. 8.5. Scatter plot of 2020 NFT transactions clustered by K-means, using log-transformed total price and number of sales.

The scatter plot visualizes the results of year-specific K-means clustering for NFT transactions in 2020, based on log-transformed total price and number of sales. Compared to 2019, the distribution of data points is more dispersed across both axes, indicating an overall increase in transaction diversity and activity.

The Medium Value NFTs cluster (in red) remains the dominant group, but this year it exhibits wider horizontal spread, covering a broader price range—from roughly log(1 + total_price) = 2 to 14—while still mostly concentrated below 200 in sales frequency. This suggests that 2020 saw a diversification of mid-value assets, with prices increasing compared to the previous year, while trading activity stayed relatively moderate.

A notable development is the emergence of a distinct Low Value cluster (in blue), containing NFTs with very low prices but unusually high sales counts, some exceeding 1,000 transactions. This cluster reflects the rise of mass-distributed, low-cost assets, possibly driven by gamified platforms, utility NFTs, or promotional campaigns that encouraged frequent resale or airdrops.

On the far right, a few points appear as High Value NFTs (in green), with log-transformed prices above 13 and limited sales activity. While sparse, their presence confirms the beginning of a premium segment within the market—indicating that some collectors or early adopters were willing to spend large amounts on select NFTs even before the 2021 boom.

Overall, the 2020 clustering pattern reveals a transition phase: the NFT market started to mature, featuring more frequent low-value trading alongside the early signs of high-value investments. These structural shifts are aligned with broader ecosystem growth, increasing platform adoption, and speculative experimentation leading into the next year.

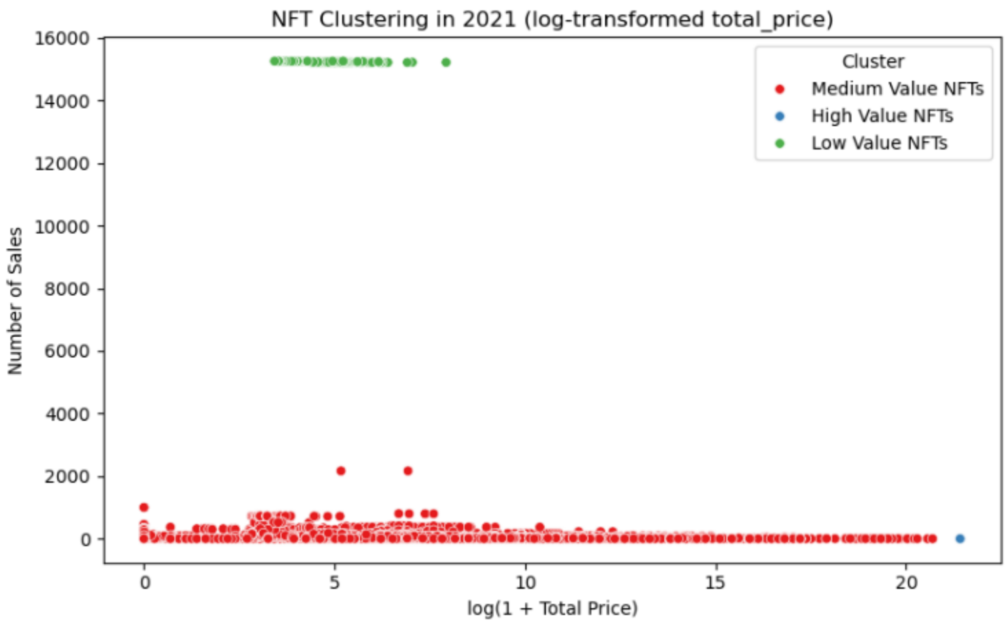

Fig. 8.6. Scatter plot of 2021 NFT transactions clustered by K-means, using log-transformed total price and number of sales.

The scatter plot displays the clustering of NFT transactions in 2021 based on log-transformed total price and number of sales. Unlike previous years, 2021 demonstrates a dramatic expansion in both price and volume dimensions, reflecting the NFT market's explosive growth during this period.

The Medium Value NFTs cluster (shown in red) dominates the landscape, with a high concentration of points densely packed along a wide horizontal axis from approximately log(1 + total_price) = 2 to 20. This indicates a vast spread of moderately priced NFTs being transacted at varying frequencies, showcasing the increased diversity and accessibility of NFTs in 2021.

A notable shift is seen in the Low Value cluster (green), which occupies the topmost region of the plot, consisting of a small group of NFTs with extremely high sales frequencies—around 15,000 trades—despite having low price values. These outliers suggest the emergence of gamified or utility-driven NFTs that were exchanged repeatedly, possibly linked to play-to-earn mechanics, reward systems, or automated trading loops that became more prevalent during the 2021 NFT boom.

The High Value cluster (in blue), though sparse, appears further to the right, with log-transformed total prices exceeding 20 and minimal sales activity. This region likely represents rare, high-profile NFT sales—such as auctioned digital art or celebrity-backed assets—that characterized the top-end of the 2021 market.

Collectively, the 2021 clustering reflects a market that not only expanded in scale but also diversified in structure, incorporating both speculative high-value assets and mass-traded low-cost tokens. The clear horizontal and vertical spread of the clusters highlights how user behavior, asset pricing, and transaction dynamics evolved significantly compared to prior years—marking 2021 as the defining breakout phase in the NFT ecosystem's development.

9. Examining NFT Price Volatility

9.1 Analysis Objective

The primary goal of this section is to measure the volatility of the NFT market in order to better understand market dynamics and investor behavior. Volatility represents the speed and magnitude of price changes over time and is a critical indicator, particularly in risk analysis. For high-risk digital markets such as crypto assets, volatility plays a vital role in understanding market sentiment, investor behavior, and speculative movements.

Analyzing the level of volatility in NFTs provides valuable insights for both investors and market regulators in making strategic decisions. Therefore, statistically summarizing price movements and determining the degree of variability offers information about the stability of the market.

9.2 Methodology: Statistical Summary and Volatility Indicators

To assess the overall volatility of the NFT market, three fundamental statistical measures were calculated during the initial analysis phase: mean price, standard deviation, and coefficient of variation (CV). These metrics help us understand the extent to which prices are stable or dispersed.

- **Mean**: Represents the general price level of NFTs and provides an indication of the central tendency of the market value. However, on its own, it does not offer sufficient information about price distribution.
- **Standard Deviation**: Indicates how much prices deviate from the mean. A high standard deviation suggests inconsistent and fluctuating prices, indicating a volatile market.
- **Coefficient of Variation (CV)**: Calculated as the ratio of the standard deviation to the mean, CV measures relative volatility. It is especially useful for comparing volatility among assets with different average levels. In heterogeneous environments like the NFT market, where outliers are common, CV plays a critical role in measuring relative risk and price instability.

By evaluating these metrics together, the structural fluctuations of the NFT market, the variability of prices, and the level of risk they pose to investors have been quantitatively assessed. This initial analysis serves as a foundation for subsequent time-series-based volatility studies.

9.3 Why Mean, Standard Deviation, and CV Were Calculated?

- **Mean:** Determines the general trend of NFT prices. However, on its own, it does not reflect how unstable the market is.
- **Standard Deviation (Std):** Measures deviations around the mean, revealing the degree of market fluctuations. A high standard deviation indicates high volatility.
- **Coefficient of Variation (CV):** Defined as the ratio of Std to Mean (Std / Mean). It is used to compare the relative volatility of assets with different units or value ranges. For example, a $10 deviation in a $100 asset and a $1,000 deviation in a $10,000 asset both yield the same CV (0.1), allowing for risk comparisons on a relative scale [50].

In the NFT market, prices span a wide range (e.g., some assets are worth thousands of dollars while others are just a few cents). Therefore, CV is the most appropriate measure for comparing different collections, categories, or time periods. A high CV value indicates high relative volatility, and thus higher risk [26].

9.4 Interpretation of Heatmap Results

The heatmap below presents a summary of overall volatility in the NFT market:

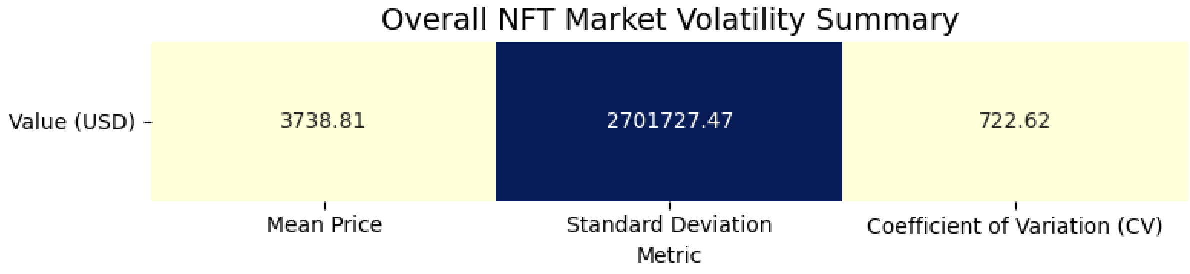

Fig. 9.1. Heatmap Visualization of Overall NFT Market Volatility Summary

- **Mean Price:** $3,738.81
- **Standard Deviation:** $2,701,727.47
- **Coefficient of Variation (CV):** 722.62

These figures indicate that NFT prices are extremely dispersed and unstable. While the mean price remains at a moderate level, the standard deviation in the millions demonstrates that the market is dominated by extreme price outliers (for example, rare collections). The extraordinarily high CV of 722 clearly shows that prices exhibit an extreme distribution around the mean, signifying a highly speculative and volatile market.

This situation implies both the potential for large gains and significant losses for investors. It also suggests that NFTs are far less stable than traditional asset classes and that the market structure is open to powerful fluctuations [3].

9.5 Temporal Analysis of NFT Price Volatility and Market Dynamics

9.5.1 Interpreting Monthly Volatility Trends

The time series visuals provide a multidimensional view of price fluctuations in the NFT market from January 2019 to March 2022. When the mean price, standard deviation, coefficient of variation (CV), and transaction volume are examined together, the temporal and structural dynamics of market volatility become clearly visible.

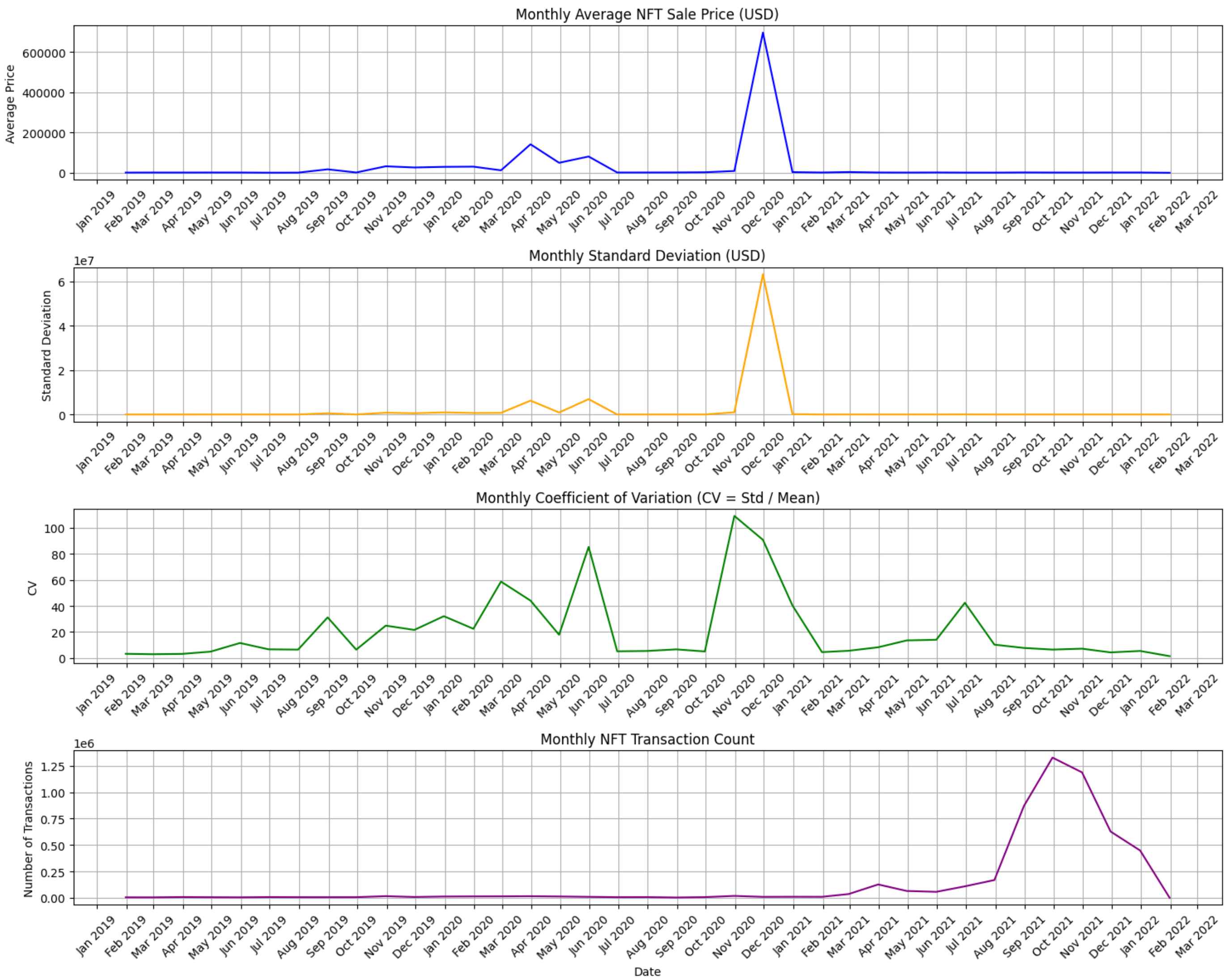

Fig. 9.2. Graphic Representations of Monthly Average NFT Sale Price (USD), Monthly Standard Deviation (USD), Monthly Coefficient of Variation (CV = Std / Mean), Monthly NFT Transaction Count

9.5.2 In-Depth Analysis of Key Factors Influencing CV (Coefficient of Variation)

The Coefficient of Variation (CV) is a statistical indicator calculated by dividing the standard deviation by the mean, providing a relative measure of price inconsistency. Therefore, CV not only reflects the magnitude of volatility but also its context. Below is a detailed analysis of the main factors that affect CV over time, interpreted through the lens of observed temporal breakpoints in the graphs:

Sudden High-Priced Sales (Outliers)

One of the most notable periods with dramatic CV increases is between November 2020 and January 2021. During this time, there were sharp spikes in average NFT prices, accompanied by significant increases in standard deviation. This can be attributed to widely publicized high-value NFT sales. Examples include:

- Beeple's *"Everydays: The First 5000 Days"* NFT selling for $69 million (March 2021),
- Early collections like CryptoPunks selling for millions of dollars at auctions.

Such sales artificially raise both the average and variance within the sample. If these prices occur in only a small number of transactions, the CV can reach extreme values. Thus, CV reflects not only market risk but also signals market imbalance [31].

Sudden Decrease or Increase in Mean
In CV calculation, the mean appears in the denominator. If the average price drops suddenly, the CV increases—even if the standard deviation remains constant. This was observed around mid-2020:

- While average prices fell slightly, standard deviation stayed relatively stable.

This implies that:

- Market transaction volume might have been increasing,
- But sales were likely concentrated in lower-value NFTs.

In such cases, the price distribution widens, the mean drops, but deviation remains high—causing CV to spike. These scenarios show that CV reflects not only absolute volatility, but also price instability and market unpredictability [31].

Fluctuations in Market Participation (Transaction Volume)
Starting mid-2021, the NFT market saw a boom in transaction counts (especially between July and September). However, this period saw a sharp decline in CV values. Likely reasons include:

- As the market grew, prices became more homogeneous,
- The share of low-to-mid-priced NFTs increased, reducing the impact of speculative purchases.

This "democratization effect" reduced the statistical weight of extreme high-value sales, leading to a more stable CV.

This trend may indicate a phase of institutionalization and mass investor engagement in the NFT market [2].

Speculative Bubbles and Periods of Stability
CV exceeding 80 during mid-2020 suggests that the NFT market was largely driven by speculative expectations. This period coincided with:

- The portrayal of blockchain as undergoing an “NFT revolution,”
- Intense media coverage,
- A surge in popularity of crypto assets overall.

In such times:

- Prices deviate from economic fundamentals,
- Buyers and sellers act emotionally or based on hype,
- Price volatility increases, and CV becomes one of the most effective indicators of irrational market behavior [32].

9.5.3 Interpretation of Overall Market Volatility Trends
The price volatility of the NFT market, as illustrated in the analyzed graphs, clearly follows an evolutionary trajectory over time. This evolution can be examined in three main phases:

(1) Period of Stability and Preparation (2019 – First Half of 2020):

- Average prices were low and relatively stable.
- Transaction volume was limited; the market was still in its early adoption stage.
- Volatility was low, and CV remained stable (within the 10–20 range).

This phase can be characterized as the "early adopter era", where interest in NFTs was limited but the technological foundation was being established.

(2) Growth and Bubble Phase (Second Half of 2020 – First Half of 2021):

- Both average prices and standard deviation increased dramatically.
- CV values occasionally exceeded 100, an extraordinarily high figure not typically observed in traditional financial markets.
- A speculative wave occurred, driven by high-profile NFT sales and growing media attention.

This period marks the phase when the NFT market gained sudden global recognition and attracted its first major wave of investor interest [26].

(3) Maturity and Cooling-Off Phase (Late 2021 – Early 2022):

- Both average prices and standard deviation began to decline.
- CV stabilized in the 20–30 range.
- Transaction volume peaked and then decreased, but the price distribution became more controlled.

For the first time, the market showed signs of becoming less speculative, more mass-oriented, and balanced in terms of transaction diversity.

This overall structure indicates that the NFT market has transitioned from its "childhood" phase to a "youth" stage, evolving into a more calculable and predictable environment for investors [4].

9.5.4 Temporal Smoothing and Rolling Volatility Analysis in the NFT Market

Why Use Rolling Window Techniques?

In financial time series, sudden changes and high-frequency fluctuations can be interpreted as "noise" in statistical analyses. Therefore, instead of using raw metrics, calculating rolling averages over time allows trends to be observed more accurately. This approach:

- Smooths out the effects of short-term fluctuations,
- Makes it easier to visualize overall trends,
- Reduces the impact of outliers, enabling meaningful periodic interpretations [51].

In this study:

- A 3-month rolling average (monthly CV) was applied to track short-term volatility trends,

- A 2-year rolling average (yearly CV) was applied to monitor long-term structural changes.

This methodology is a critical analytical tool, especially in digital asset markets like the NFT market, where speculative and sudden movements are common [52].

9.5.5 3-Month Rolling CV: Short-Term Volatility Trends

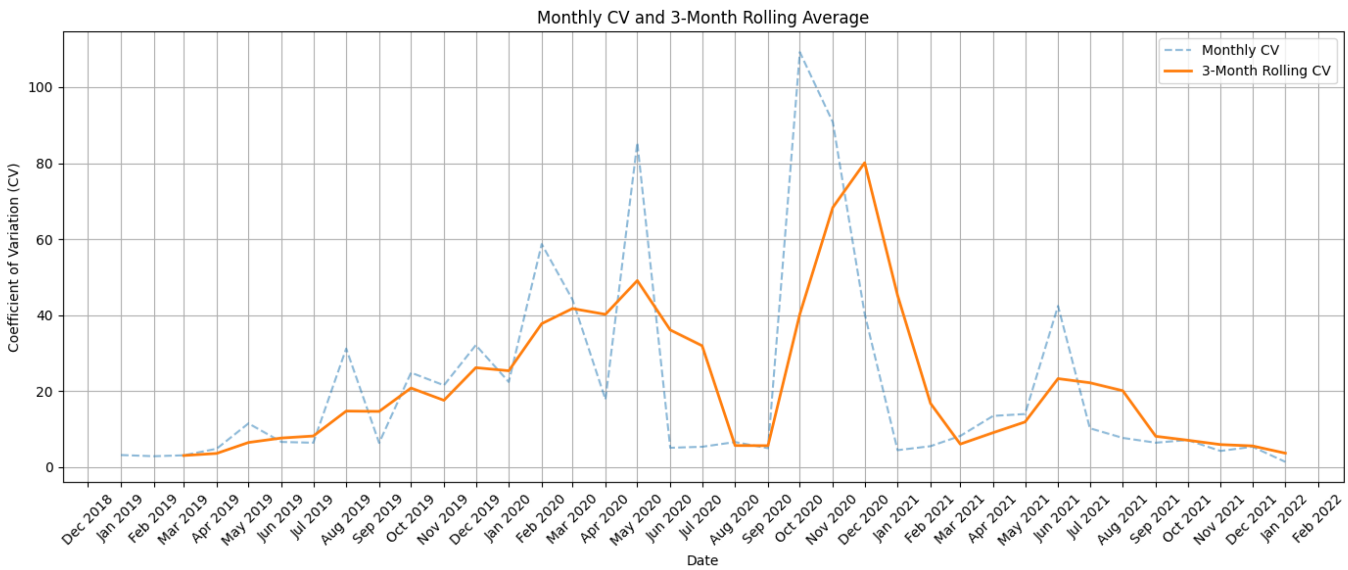

Fig. 9.3. Graphic Representation of Monthly CV and 3-Month Rolling Average

The first chart shows the monthly coefficient of variation (CV) values over time along with its 3-month rolling average.

Key Findings:

- Since mid-2019, the CV shows a gradual increase, indicating early diversification in the market.
- Sudden spikes are observed during periods such as May–October 2020 (e.g., CV reaching around 100). This can be attributed to both low average prices and high price deviations. In other words, a small number of outlier sales dramatically increased the CV.
- The 3-month average reveals the temporary nature behind these spikes. For example, sudden increases in CV mostly last about one month, while the rolling average reflects these values more smoothly.
- From 2021 onward, there is a period where the CV decreases and stabilizes, indicating that the market is beginning to mature [2].

Analytical Interpretation:
The 3-month rolling CV is effective in identifying short-term speculative periods in the market. It can especially be used in predictive models to filter out the effects of NFT bubbles or high-priced token sales. This provides a high-resolution perspective for explaining market behavior.

9.5.6 2-Year Rolling CV: Long-Term Trend and Structural Transformation

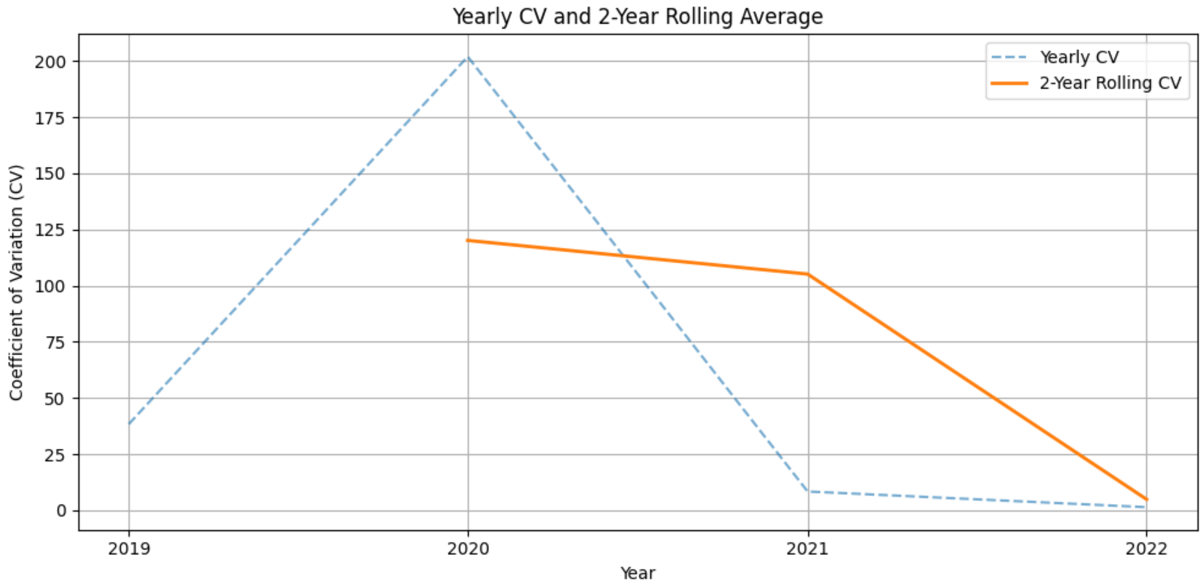

Fig. 9.4. Graphic Representation of Yearly CV and 2-Year Rolling Average

The second chart shows the yearly CV values along with their 2-year rolling average. This chart is less volatile and more suitable for trend analysis.

Observations:

- The year 2020 marks the peak in CV values. The annual CV reaches approximately 200, indicating that price instability in the market peaked.
- 2021 and 2022 are years with dramatic declines in CV. Values falling below an average of 100 signal that the market has become more balanced and predictable.
- The 2-year average smooths this transition and confirms that volatility has begun to decline permanently over time.

Strategic Interpretation:
The 2-year rolling CV reveals not only speculative moments but also phases of structural change. For example:

- 2019–2020 → Speculative rise and bubble formation,
- 2021–2022 → Market maturation, increased liquidity, and expanded user base.

This trend indicates that the NFT market has moved from its “initial phase” to a “growth phase,” aiding long-term strategic guidance for investors [4].

9.5.7 Contribution of Rolling Analysis to Volatility Examination

1. **Noise Reduction**
   In markets prone to extreme prices and sudden spikes like the NFT market, an exceptional sale in a single month can dramatically increase the CV. For example:

- In the 3-Month CV chart, sudden spikes reaching the 80–100 range are observed during periods like May 2020 and October 2020. These are likely caused by a few rare high-priced sales distorting the statistics.
- However, the rolling average line smooths out these sudden changes, making the overall market behavior more readable.
- This allows investors to evaluate the general market rhythm rather than just outlier movements. For instance, instead of saying “There was a spike in May 2020,” one can say, “CV rose between April and June and then declined.”

2. **Trend Detection**
   Short- and long-term rolling CV analyses are unique tools to understand structural changes in the market:

- The 3-Month Rolling CV especially shows that volatility sharply increased from the last quarter of 2020 to early 2021, then gradually decreased. This signals a transition from a speculative period to a more stable structure.
- The 2-Year Rolling CV chart shows that the highest volatility was seen in 2020, followed by a significant decrease in 2021 and 2022 on a long-term scale.
- This analysis provides concrete evidence that the NFT market was irregular and speculative in its early stages but became more institutional and stable in later phases.

3. **Risk Analysis**
   With rolling CV, investors can identify time periods of high price volatility, thus higher risk:

- During the second half of 2020, CV values rose rapidly, meaning investors faced a high risk of unexpected price fluctuations. Although this period offered high return potential, it also came with a high probability of losses.
- Conversely, the decline in CV from the second half of 2021 represents a safer and more predictable environment for investors.
- This information is strategically valuable for investors who want to determine their risk tolerance in NFT trading decisions.

4. **Timing Insight**
   Rolling analyses can also guide strategic timing decisions for NFT projects. For example:

- If launching a new NFT collection, choosing a period when the 3-month rolling CV is trending downward implies more predictable pricing and greater investor confidence.
- On the other hand, entering the market during highly speculative periods (such as late 2020) may offer viral attention but carries risks of price manipulation or bubble formation for the project.
- Therefore, rolling analyses serve as a powerful indicator for project managers to answer the question: “Is the market currently stable or turbulent?”

Conclusion

- The 3-Month Rolling CV should be used for short-term strategic decisions (e.g., launch timing, campaign planning).
- The 2-Year Rolling CV should be applied to understand long-term market trends and investment policies.
- These analyses are valuable not only for measuring volatility but also for modeling and forecasting market behavior [51][52][4].

9.6 General Conclusion and Evaluation

Within the scope of this study, the volatility dynamics of the NFT market were comprehensively examined using both static statistical summaries and time-series-based rolling analyses. The findings revealed not only the speculative characteristics of the NFT market but also its maturation processes, user behaviors, and cyclical transformations.

The Multi-layered and Evolutionary Nature of Market Structure

The analyses first measured the overall volatility level of the NFT market using fundamental metrics such as average price, standard deviation, and coefficient of variation (CV). The exceptionally high CV values indicated that prices were subject to intense fluctuations not only in terms of magnitude but also instability. This suggests that the NFT market, compared to traditional financial assets, is less regulated, less predictable, and more sentiment-driven [26].

In-depth Time-Series Analysis with Rolling Window Approach

The 3-month and 2-year rolling CV analyses used to track volatility over time successfully reflected short- and long-term market transformations. In the short term, speculative sales and sudden demand spikes drove prices to extremes, while in the long term, relative stability was observed after 2021. This transition is a strong indicator that the NFT market has evolved from an early adoption phase into a more settled structure [51].

Identification of Factors Affecting CV Values

Detailed examination of the time-series graphs showed that changes in CV are linked to several key factors:

- High-priced outlier sales,
- Sudden drops or rises in average price,
- Expansions or contractions in market volume,
- Speculative waves and media influence.

Each of these variables directly impacts the mathematical structure of CV and is critical to consider for accurate interpretation of market behavior.

Analytical and Strategic Contributions

The applied analysis methods are valuable not only academically but also practically:

- Rolling CV values visualize periods when the market is risky or stable for investors.
- Timing strategies for projects and collections can be adjusted according to periods of lower volatility.
- Deeper modeling of market dynamics allows differentiation between stable segments and speculative areas.

In this context, rolling analysis emerges as a powerful tool not only for understanding the past but also for predicting potential future behaviors [52].

In Conclusion

Although the NFT market was shaped by high volatility and speculative tendencies in its early stages, signals of a transition toward a more balanced structure appeared after 2021. Average prices and trading volumes increased relatively, while the decline in the coefficient of variation statistically reflects this transition.

This analysis demonstrates that the NFT market cannot be evaluated solely by price tracking; secondary dynamics such as volatility play a central role in investment decisions. Time-series analyses serve as a critical guide for all stakeholders planning the future of the NFT market.

10. Conclusion

This project offers a thorough, data-centric exploration of the Non-Fungible Token (NFT) market during the critical developmental period spanning from 2019 to 2021. Throughout the study, a multi-methodological framework was employed, integrating diverse analytical approaches to capture the complex dynamics of the NFT ecosystem. The methodological design included time series analysis to examine temporal trends and market cycles, statistical tools such as Pearson correlation and OLS regression to identify and quantify the impact of various features on NFT price formation, clustering algorithms (specifically K-Means) to categorize assets based on value-related attributes, frequent pattern mining (via the FP-Growth algorithm) to uncover shared categorical traits among high-value NFTs, and a tailored anomaly detection pipeline to identify manipulative behaviors—especially wash trading.

Drawing on a robust dataset composed of over 5.25 million transaction records sourced from the OpenSea platform—the largest NFT marketplace globally—the study captures a wide spectrum of market behaviors. The dataset encompasses detailed metadata including price, asset and collection identifiers, buyer and seller addresses, token types, sale timestamps, and categorical descriptors such as NFT category and collection name. This rich and high-volume dataset enabled the researchers to not only observe surface-level trends but also drill into the underlying structures of price fluctuations, temporal behavior, and user interactions.

The scope of the analysis extends beyond mere descriptive statistics. The project delves into the temporal evolution of NFT activity, identifying how the market transitioned from an early experimental phase in 2019 to a speculative surge by the end of 2021. It further investigates the key attributes that influence valuation—whether numerical (e.g., number of past sales, date of sale) or categorical (e.g., type of asset, associated platform) —and examines their significance in both univariate and multivariate settings. By applying clustering, the study isolates high-value NFT groups and systematically analyzes the recurring patterns and ecosystem-based traits that differentiate them from the broader market. Additionally, the incorporation of anomaly detection techniques allows the project to assess the integrity of the market, detecting transaction-level indicators of wash trading and self-dealing practices.

Temporal Analysis and Market Dynamics:
The time series analysis conducted across the 2019–2021 timeframe reveals a pronounced shift in NFT market behavior—from modest and sporadic early sales to an explosive rise in 2021. This growth is not only visible in the volume of sales but also in the momentum and volatility of market participation, which intensified alongside increasing media coverage, cultural relevance, and institutional interest. The steep trajectory observed in the second half of 2021 points to the onset of a speculative phase, though the dataset concludes before any possible correction phase, preserving a snapshot of the market at its most expansive state.

Price Formation Analysis:
The project's statistical investigation into price determinants confirms that many numerical variables, such as transaction month or total sales per asset, fail to significantly explain price variations. Only the year variable shows a statistically meaningful—yet modest—correlation with price. The weak predictive power of these features, even in a multivariate regression setting, suggests that NFT valuation is governed by nonlinear, multi-factorial influences, including social signaling, platform engagement, and ecosystem-specific credibility.

Clustering and High-Value Attributes:
By implementing K-Means clustering on standardized transaction data and applying FP-Growth frequent itemset mining to the highest-value cluster, the study successfully surfaces the structural patterns associated with premium NFTs. Traits such as *Category = Virtual Worlds*, *payment_token.name = Decentraland MANA*, and *asset.collection.name = Decentraland Wearables* emerge as strong indicators of value. These results confirm the importance of embedded utility, platform loyalty, and economic integration in shaping NFT worth—an insight that goes beyond aesthetics or scarcity.

Segment-Wide Pattern Comparison:
The inclusion of frequent pattern mining across all value clusters, under variable support thresholds, allows for a nuanced examination of the NFT ecosystem's internal segmentation. High-support itemsets illustrate market-wide consensus (e.g., popularity of collectibles and metaverse assets), while lower thresholds reveal niche-specific behaviors and preferences. This analysis highlights that the NFT market is not monolithic but instead exhibits a stratified structure with multiple active sub-markets.

Wash Trading and Anomaly Detection:
The custom wash trading detection framework, which integrates bidirectional trade loops, high-frequency behavior, and self-trading markers, exposes a layer of the market that undermines perceived value and fairness. The identification of address-level manipulation patterns demonstrates the need for vigilance and algorithmic surveillance, even in supposedly transparent blockchain environments.

Synthesis and Broader Implications:
Collectively, these analyses underscore that NFT price formation and market behavior cannot be adequately explained by simple numerical descriptors. Instead, they emerge from an interplay of technical design, social capital, network effects, and evolving digital narratives. This reinforces the necessity of multi-dimensional evaluation frameworks when studying decentralized asset systems. For platform developers, investors, creators, and analysts, the findings offer actionable insights into how digital economies form, evolve, and can be exploited or stabilized.

In conclusion, the study not only documents the rapid maturation of the NFT market from 2019 to 2021 but also lays the groundwork for future research and application development. The methodologies established here are generalizable to other blockchain-based assets and can be expanded with post-2021 data to explore market resilience, institutional adoption, and long-term utility. As NFTs continue to intersect with gaming, art, finance, and identity, such analytical foundations will be essential for understanding and shaping their future.

With these comprehensive analyses, the project successfully fulfills its objective of examining the NFT market from multiple dimensions and offers a solid foundation for future academic and practical investigations in the field.

REFERENCES

1. Wang, Q., Li, R., Wang, Q., & Chen, S. (2021). *Non-Fungible Token (NFT): Overview, Evaluation, Opportunities and Challenges*. arXiv preprint arXiv:2105.07447.

2. Nadini, M., Alessandretti, L., Di Giacinto, F., Martino, M., Aiello, L. M., & Baronchelli, A. (2021). *Mapping the NFT Revolution: Market Trends, Trade Networks, and Visual Features*. Scientific Reports, 11(1), 20902.

3. Dowling, M. (2022). *Fertile LAND: Pricing non-fungible tokens*. Finance Research Letters, 44, 102096. https://doi.org/10.1016/j.frl.2021.102096

4. Cong, L. W., He, Z., & Tang, K. (2021). *NFTs: Markets, Valuation, and Potential*. NBER Working Paper No. 28945. https://www.nber.org/papers/w28945

5. von Wachter, V., Schlegel, M., & Zhao, C. (2022). *NFT wash trading: Quantifying suspicious behaviour in NFT markets*. arXiv preprint arXiv:2202.06921.

6. BryanW26. (2022). *OpenSea NFT Sales (2019–2021)* [Dataset]. Kaggle. https://www.kaggle.com/datasets/bryanw26/opensea-nft-sales-2019-2021/data

7. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. https://otexts.com/fpp2/

8. Chatfield, C. (2003). *The analysis of time series: An introduction* (6th ed.). Chapman and Hall/CRC.

9. Crawley, J. (2021, July 6). NFT sales climb to $2.47B in first-half 2021: Report. *CoinDesk*. https://www.coindesk.com/markets/2021/07/06/nft-sales-climb-to-247b-in-first-half-2021-report

10. Wired. (2022). The year the NFT died and came back to life. *Wired*. https://www.wired.com/story/the-year-the-nft-died-and-came-back-to-life

11. Vanity Fair. (2023, January). SBF, Bored Ape Yacht Club, and the spectacular hangover after the art world's NFT gold rush. *Vanity Fair*. https://www.vanityfair.com/style/2023/01/sbf-bored-apes-art-world-hangover

12. Charlton, A. (2022, March 10). *NFT sales have spiked 21,000% in a year – but has the marketplace become a bubble?* Euronews. https://www.euronews.com/next/2022/03/10/nft-sales-have-spiked-21-000-in-a-year-but-has-the-marketplace-become-a-bubble

13. Howcroft, E. (2022, January 10). NFT sales hit $25 billion in 2021, but growth shows signs of slowing. *Reuters*. https://www.investing.com/news/stock-market-news/nft-sales-hit-25-billion-in-2021-but-growth-shows-signs-of-slowing-2734900

14. Kessler, A. (2022, May 5). NFT sales declined 92% since its peak in September 2021. *80 Level*. https://80.lv/articles/nft-sales-declined-92-since-its-peak-in-september-2021/

15. Murphy, J. J. (1999). Technical Analysis of the Financial Markets. New York Institute of Finance.

16. Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance, 47(5), 1731-1764.

17. N. J. Gogtay and U. M. Thatte, *Principles of Correlation*. Mumbai, India: Seth GS Medical College and KEM Hospital, 2017. [Online]. Available: https://www.kem.edu/wp-content/uploads/2012/06/9-Principles_of_correlation-1.pdf

18. S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Advances in Psychology*, vol. 52, P. A. Hancock and N. Meshkati, Eds. Amsterdam, Netherlands: North-Holland, 1988, pp. 139–183.

19. A. A. Al-Azzawi, M. A. Al-Azzawi, and A. A. Al-Azzawi, "Study the Influence of Normalization/Transformation Process on the Accuracy of Supervised Classification," in *2020 International Conference on Computer Science and Software Engineering (CSSE)*, Baghdad, Iraq, 2020, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9214160

20. A. M. Ferrenberg and R. H. Swendsen, "Optimized Monte Carlo data analysis," *J. Chem. Phys.*, vol. 109, no. 16, pp. 6821–6827, 1998. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0003267001010406

21. MacQueen, J. (1967). *Some Methods for Classification and Analysis of Multivariate Observations*. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.

22. Jain, A. K. (2010). *Data clustering: 50 years beyond K-means*. Pattern Recognition Letters, 31(8), 651–666.

23. Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer Series in Statistics.

24. Han, J., Pei, J., & Yin, Y. (2000). *Mining frequent patterns without candidate generation*. ACM SIGMOD Record, 29(2), 1–12.

25. Tan, P. N., Steinbach, M., & Kumar, V. (2014). *Introduction to Data Mining*. Pearson.

26. Ante, L. (2022). *The non-fungible token (NFT) market and its relationship with Bitcoin and Ethereum*. Finance Research Letters, 44, 102103.

27. Kaczynski, T., Ribeiro, B., & Hofmann, K. (2022). *Token Utility and Digital Asset Valuation: Evidence from NFT Metaverse Economies*. Journal of Digital Markets, 1(1), 34–55.

28. Agrawal, R., & Srikant, R. (1994). *Fast algorithms for mining association rules*. In Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), 487–499.

29. Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

30. Zaki, M. J., & Hsiao, C.-J. (2002). *Efficient algorithms for mining closed itemsets and their lattice structure*. IEEE Transactions on Knowledge and Data Engineering, 17(4), 462–478.

31. Balaghat, H., & Pahlavani, M. (2022). *Coefficient of Variation in Crypto Asset Valuation*. Journal of Financial Analytics. https://doi.org/10.2139/ssrn.3974932

32. Bianchi, D., & Dickerson, J. P. (2022). *NFTs and Market Efficiency: Evidence from Ethereum*. SSRN Working Paper. https://doi.org/10.2139/ssrn.4092289

33. Park, S., & Kim, Y. (2022). *The emergence of the metaverse and its implications for blockchain-based NFT markets*. Journal of Open Innovation: Technology, Market, and Complexity, 8(1), 5. https://doi.org/10.3390/joitmc8010005

34. NonFungible.com & L'Atelier BNP Paribas. (2022). *NFT Market Report 2022*. https://nonfungible.com/reports/2022/en

35. Wang, J., Yu, C., Liu, Y., Ma, X., & Xia, Y. (2022). *Non-fungible token (NFT): Overview, evaluation, opportunities and challenges*. arXiv. https://arxiv.org/abs/2202.03866

36. Mazorra, J., Ante, L., & Sandner, P. (2023). *Wash trading at scale: Evidence from NFT markets*. arXiv. https://arxiv.org/abs/2305.01543

37. Tošić, A., Vičič, J., & Hrovatin, N. (2025). Beyond the surface: Advanced wash-trading detection in decentralized NFT markets. *Financial Innovation*, 11(86). https://doi.org/10.1186/s40854-025-00766-z

38. Chen, S., Chen, J., Yu, J., Luo, X., & Wang, Y. (2023). The dark side of NFTs: A large-scale empirical study of wash trading. *arXiv preprint arXiv:2312.12544*. https://arxiv.org/abs/2312.12544

39. NFTGo.io. (2022). NFT wash trading report: Eliminating the "noise" in the market. *Mirror.xyz*. https://nftgo.mirror.xyz/aWZMuptrKAPDG9Ya_Ztqnae_EDVeIwTJDqiuJCu-Le8

40. Niu, Y., Li, X., Peng, H., & Li, W. (2024). Unveiling wash trading in popular NFT markets. *arXiv preprint arXiv:2403.10361*. https://arxiv.org/abs/2403.10361

41. Chainalysis. (2022). 2022 crypto crime report preview: NFT wash trading & money laundering. https://www.chainalysis.com/blog/2022-crypto-crime-report-preview-nft-wash-trading-money-laundering/

42. Financial Crime Academy. (2023). Understanding NFT wash trading. https://financialcrimeacademy.org/understanding-nft-wash-trading/

43. R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "STL: A Seasonal-Trend Decomposition Procedure Based on Loess," *J. Off. Stat.*, vol. 6, no. 1, pp. 3–73, 1990. [Online]. Available: https://www.nniiem.ru/file/news/2016/stl-statistical-model.pdf

44. S. Giri, D. Du, and M. Beruvides, "A systematic approach to predicting NFT prices using time series forecasting and macroeconomic factors in digital assets," *Cogent Economics & Finance*, vol. 13, no. 1, p. 2468387, Dec. 2025. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/23322039.2025.2468387

45. J. D. Hamilton, *Time Series Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1994.

46. J. D. Cryer and K.-S. Chan, *Time Series Analysis: With Applications in R*, 2nd ed. New York, NY, USA: Springer, 2008.

47. M. Nadini, L. Alessandretti, F. Di Giacinto, M. Martino, L. M. Aiello, and A. Baronchelli, "Mapping the NFT revolution: Market trends, trade networks, and visual features," *Sci. Rep.*, vol. 11, no. 1, p. 20902, Oct. 2021. [Online]. Available: https://www.nature.com/articles/s41598-021-00053-8

48. Y. Fu, X. Lin, and Y. Wang, "A Survey on Time Series Data Mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 9, pp. 1546–1567, Sept. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/9072123

49. A. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, "Time-series clustering – A decade review," *Information Systems*, vol. 53, pp. 16–38, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0306437915000733

50. Kozak, S. (2021). *The Coefficient of Variation as a Measure of Relative Risk in Asset Portfolios*. SSRN. https://doi.org/10.2139/ssrn.3784026

51. Tsay, R. S. (2010). *Analysis of Financial Time Series*. Wiley Finance. https://onlinelibrary.wiley.com/doi/book/10.1002/9780470644560

52. Brownlees, C., & Gallo, G. M. (2006). *Financial econometric analysis using rolling sample techniques*. Journal of Economic Surveys, 20(4), 613–647. https://doi.org/10.1111/j.1467-6419.2006.00254.x