DATA SCIENCE IN SCIENCE
2023, VOL. 2, NO. 1, 2151950
https://doi.org/10.1080/26941899.2022.2151950

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Non-Fungible Token Transactions: Data and Challenges

Jason B. Cho[a], Sven Serneels[b,c], and David S. Matteson[a]

[a]Department of Statistics and Data Science, Cornell University, Ithaca, New York, USA; [b]Gallop Data, Inc., Denver, Colorado, USA; [c]Department of Mathematics, University of Antwerp, Antwerp, Belgium

ABSTRACT

Non-fungible tokens (NFT) have recently emerged as a novel blockchain-hosted financial asset class that has attracted major transaction volumes. However, preprocessing and analysis of NFT transaction data, which investors often rely on for their investment decisions, pose several challenges not commonly encountered in traditional financial data. These challenges arise mainly due to the non-fungible nature of NFTs as well as the intrinsic characteristics of the blockchain, the primary data source for NFT transactions. Using data consisting of the transaction history of eight highly valued NFT collections, a selection of such challenges is illustrated. These include price differentiation by token traits, the possible existence of lateral swaps and wash trades in the transaction history, and finally, severe price volatility. This paper provides an overall summary of the challenges associated with data analytics on NFT transaction data and lay a foundation for future research on the topic.

ARTICLE HISTORY
Received 16 August 2022
Revised 18 November 2022
Accepted 21 November 2022

KEYWORDS
Non-fungible tokens; NFT; financial transactions; wash trading; volatility; rarity; data analytics

1. Introduction

Blockchain technology has attracted a considerable amount of attention over the past decade as a decentralized ledger. Blockchain technology was initially launched in 2009 as the *Bitcoin Network* based on a whitepaper with the sole purpose of being the immutable public ledger containing records of transfers of digital currency (Nakamoto 2008). Bitcoin's success as a cryptocurrency sparked a wave of development in blockchain technologies in the wake of which many new blockchains were launched. Some of these concerned refining the Bitcoin approach to digital currencies. However, it was soon recognized that blockchain technology could have utility much broader than digital currency. The Ethereum network was established to fill that void: it was designed to be a blockchain that enabled to settle digital contracts and run decentralized applications (Buterin 2014). Several other networks have followed Ethereum's footsteps since.

Ethereum allows its users to run applications on the blockchain, and it gave rise to the idea of a digital, or *smart*, contract. A smart contract is a self-executing contract that consists of lines of code, which is stored on the blockchain. A simple "if-else" statement, such as "If party A pays X amount in Ethereum to party B, party A owns the NFT token X," would be an example of a smart contract. In a traditional setting, there is always a risk of party B not fulfilling the contract, even after party A pays party B the promised amount. No such risk exists for a smart contract. The contract executes automatically as soon as the condition set forth is satisfied. Smart contracts may be applied to various fields. One such area that stands out is a contract that transfers a digital right of ownership to a unique asset, which can be both digital and physical. On the blockchain, such an asset is assigned a unique token identifier, whence the name non-fungible token (NFT). Again, the concept of non-fungible tokens (NFTs) can be applied to many real-world applications, ranging from supply chains to digital sports cards and tokens with utility in games, just to name a few. These unique identifiers may be generated for any digital and physical asset and may be published on the blockchain such as Ethereum and Solana. Such a process is also known as *minting*. NFTs have become most hyped in the realm of digital art, which has received significant mainstream media attention over the past four years. The hype around NFTs has attracted individual and institutional investors, which has caused certain tokens and even entire collections to be highly valued.

Owing to the sheer amount of money that circulates in the NFT markets, high-valued NFTs are widely seen as a novel financial asset class. A broad set of financial services have emerged around NFTs that span lending protocols, hedge and index funds, NFT fractionalization, buy now pay later services, and many more. This phenomenon is referred to as the *financialization of NFT*. In addition, there has been a significant increase in studies exploring the financial aspects of NFT markets in recent years (Bao and Roubaud (2021)). Studies such as Kong and Lin (2022), Mazur (2021), Borri et al. (2022) and Ko et al. (2022) analyze risk-return of NFT in relation to other asset classes, such as bond,

CONTACT Sven Serneels svenserneels@protonmail.com Gallop Data, Inc., Denver, CO, USA

equity, and commodities. All four studies conclude that the NFT market is extremely volatile and behaves differently from traditional asset classes.

A few studies also explore various potential predictors for NFT price. Dowling (2022a) and Ante (2022) explore the correlation between the price of NFT collections and the underlying cryptocurrency such as Bitcoin and Ether. Both studies conclude that cryptocurrency price is weakly correlated to the price of an NFT. The effect of media on NFT price is also explored by Umar et al. (2022) and Kapoor et al. (2022). Particularly, Umar et al. (2022) analyzes how media coverage of COVID-19 affects the price and liquidity in the NFT market. Kapoor et al. (2022) indicates some features from social media, such as the number of likes of posts about a particular NFT collection, to be significant factors in explaining the price of an NFT token.

Success of financial services and research on the NFT markets cannot take place without accurate data. The study by Oh et al. (2022) claims that experienced investors outperform inexperienced investors by 10 percentage points per trade, suggesting a high degree of informational inefficiency in the NFT market. Data on NFT transactions, however, come with a set of challenges not shared with traditional financial data. Such challenges include price differentiation by token traits, the possible existence of lateral swaps and wash trades in the transaction history, and finally, severe price volatility. These challenges are both the result of the process that generates the data and of the nature of the asset. While studies on NFT often use such data, challenges and difficulties in analyzing NFT collections data are rarely explored. This paper will highlight these unique challenges by illustrating them based on the transaction history for a set of eight commonly traded, high valued collections of NFTs and will suggest how to preprocess the raw data so that they become amenable to statistical and financial analysis.

The paper is organized as follows: in Section 2, the data themselves are introduced, along with a discussion on specific characteristics of NFT transaction data and some summary statistics. Section 3 introduces a selection of challenges associated with NFT transaction data and some suggestions to preprocess such data appropriately. Finally, Section 4 provides an outlook into the emerging field of analytics for said data.

2. The Data

NFTs have diverse fields of application in the real world. However, the most financialized segments of the broader NFT market are digital art and gaming. Specific aspects of the tokens and their application need to be considered to analyze data for each segment. To narrow down the scope of the paper, the data described here all resort to the digital art segment of *profile picture collections (PFP)*. These collections are designed in a similar way: an artist creates a base image, to which attributes are added. By varying these attributes in a way that each realized combination occurs only once, a *collection* of unique images is created, each of which follow the base pattern, yet are clearly distinguishable. Ownership for each of these images is then *minted*, which refers to the process of creating an NFT on the blockchain of choice. An illustration of how varying traits lead to unique images is presented in Figure 1 for the collection *Bored Ape Yacht Club*.

2.1. Sources and Structure

All blockchain transactions are publicly available on the blockchain itself and are executed based on one or more *smart contracts*. Information such as the transacted price of an NFT token, involved parties of the transaction, and the timing of the transaction, may be acquired by decoding those smart contracts. However, the latter can be a daunting task. Information on smart contract execution is stored in bytecode, which will only be converted to legible and interpretable information if the corresponding contract is correctly decoded. Moreover, transactions often involve several layers of smart contract execution, for instance, separate smart contracts that represent an aggregator of exchanges (e.g. GemSwap), an individual exchange protocol (e.g. OpenSea), the NFT collection (e.g. Bored Ape Yacht Club) and a contract that defines the currency the token is transacted in (e.g. wrapped Ether). Therefore, while anyone can query any transactions that occurred on the blockchain, creating a reliable end-to-end data pipeline from the blockchain to a database ready to be used for analysis is a significant technical challenge, which is explored in detail in Section 3.2. Because of these technical challenges, several companies have recently spun up data services that provide structured, decoded transaction data for NFTs. One such company is Gallop[1], whose set of APIs is the source for the data in this paper. As of this writing, Gallop provides data of 156 Ethereum and 295 Solana NFT collections.

The data selected for this paper correspond to six PFP collections on the Ethereum blockchain (Bored Ape Yacht Club, Bored Ape Kennel Club, Mutant Ape Yacht Club, Cryptopunks, Meebits, and Autoglyphs) and two on the Solana blockchain (Degenerate Ape Academy and Aurory), all of which are highly valued at the time of writing. An overview of the collections selected, is presented in Table 1. The data contain transaction histories from the collection launch up to March 31, 2022. Data were compiled from the Gallop API responses into a structured table and were analyzed using R (R Core Team (2021)). For the analysis, the following data fields were used: the transacted price of each token in U.S. dollars; timestamp of the transaction; wallet address of the seller, denoting the entity owning the token prior to the transaction; wallet address of the buyer, denoting the entity owning the token after the transaction; and lastly visual traits such as background color, body type, or accessories of the token within the collection. Accompanying this manuscript, the tabular data for these eight collections have been made publicly available (Serneels et al. 2022).

2.2. Exploratory Data Analysis

The average price of a collection, as well as its trend within a given time span, can vary greatly among collections. Price

Figure 1. Sample of six *Bored Ape Yacht Club* tokens with blue beam, cyborg, 3 D (top) and bored, holographic, X-eyes (bottom) eye traits, among other trait differences.

Table 1. Associated blockchain, project launch date, token counts, transaction summary, floor price, and trading volume of the eight profile picture token collections.

		Aurory	Autoglyphs	Bored Ape Kennel Club	Bored Ape Yacht Club
Currency		Solana	Ethereum	Ethereum	Ethereum
Project launch		August 2021	April 2019	June 2021	April 2021
Token count		10,000	512	10,000	10,000
Transactions					
	Daily average	68.78	0.45	70.54	79.85
	Total	14,582	490	20,176	26,751
Floor price					
	Daily minimum	0.005478	0.005300	0.000010	0.000050
	Daily average	21.29	51.81	1.80	5.72
	Daily maximum	35.01	449.00	8.40	102.69
Total trading volume					
	Daily average	3032.63	17.95	303.64	1654.62
	Total	642,917.79	19,496.99	86,841.95	554,296.23
		Cryptopunks	Degenerate Ape Academy	Meebits	Mutant Ape Yacht Club
Currency		Ethereum	Solana	Ethereum	Ethereum
Project launch		June 2017	May 2021	August 2021	August 2021
Token count		9999	10,000	20,000	19,414
Transactions					
	Daily average	10.86	85.60	77.01	139.87
	Total	18,923	19,516	25,566	30,073
Floor price					
	Daily minimum	0.000050	0.052039	0.000900	0.000025
	Daily average	15.59	42.60	2.26	0.76
	Daily maximum	217.00	88.01	6.15	15.01
Total trading volume					
	Daily average	428.46	5222.89	9131.63	1511.11
	Total	746,379.75	1,190,818.21	3,031,702.79	324,889.44

Floor prices and trading volume are listed in their respective cryptocurrency (ETH or SOL).

histories are plotted for the eight collections on a logarithmic scale in Figure 2. The average prices of Aurory, Autoglphys, Bored Ape Kennel Club, Bored Ape Yacht Club, Cryptopunks, Degenerate Ape Academy, Meebits, and Mutant Ape Yacht club were 3582.05, 97,096.10, 12,832.39, 66,391.96, 113,939.13, 7465.30, 341,651.02, and 35,896.81 USD, respectively. Only Autoglyphs and Cryptopunks show a steady increase in price. Bored Ape Yacht Club had a steady increase during the first few months after the collection launch, but became approximately stationary around August 2021. However, the Aurory, Bored Ape Kennel Club and Mutant Ape Yacht Club collections were stationary during the given period, with the exception of a few outliers. Moreover, it is interesting to note that seemingly bimodal price distributions exist for several collections, most notoriously so for Meebits, an effect that will be discussed in more detail in Sections 3.2 and 3.3.

In addition to the price, the trading volume of a financial asset is another technical indicator that traders look to determine liquidity when making trading decisions. Similar to the high trading volume of a stock after the Initial Public Offering (IPO), collections such as Bored Ape Kennel Club, Bored Ape

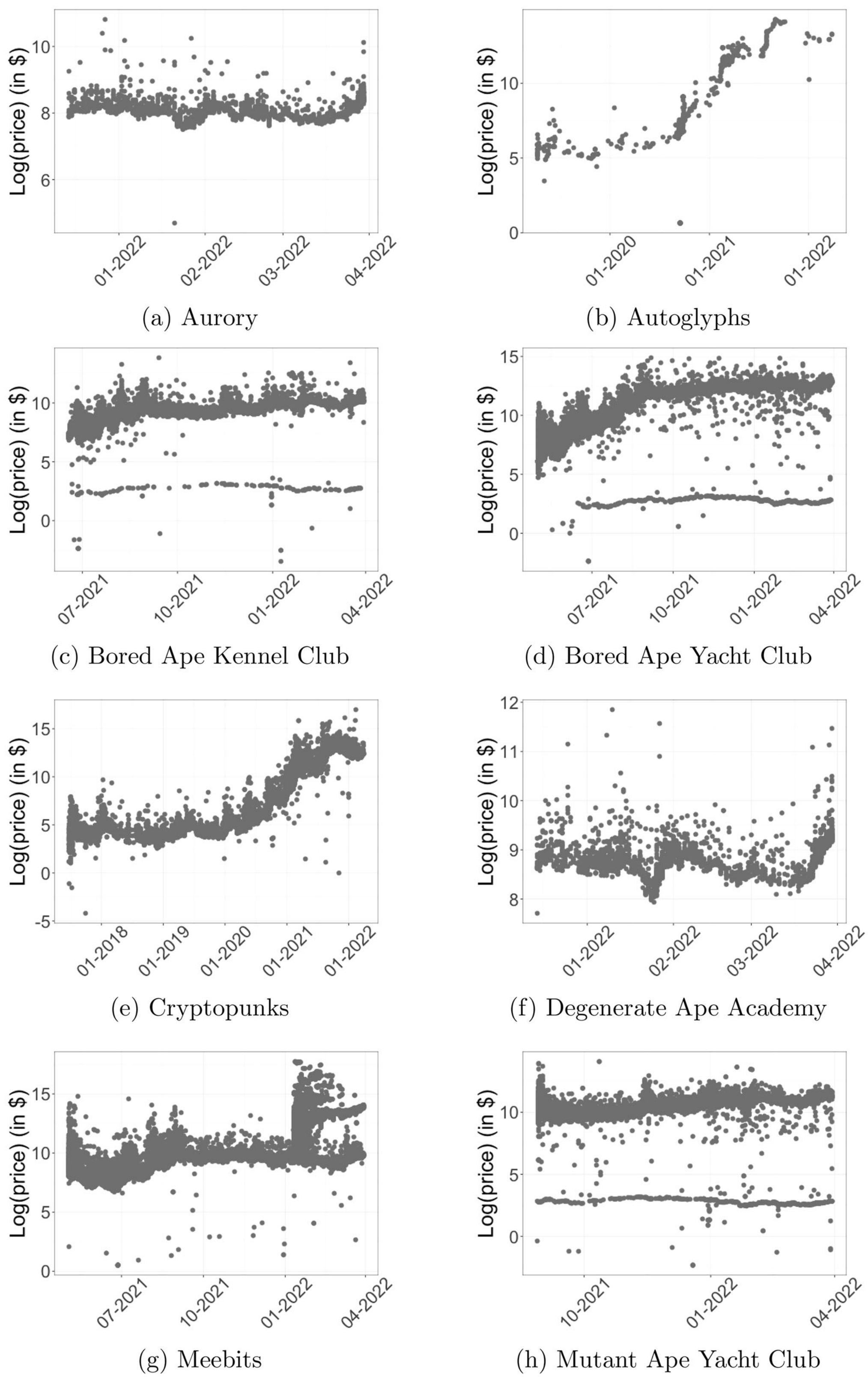


Figure 2. Log(price) of tokens from eight profile picture collections (PFP) in USD from the project launch to March 31, 2022.

Yacht Club, Meebits, and Mutant Ape Yacht Club hit their highest trading volumes right after the project launch as shown in Figure 3. When compared to traditional financial instruments, the trading volumes for all eight collections were low. While some collections reached a thousand to few thousand daily transactions, the median daily transactions of Aurory, Autoglphys, Bored Ape Kennel Club, Bored Ape Yacht Club, Cryptopunks, Degenerate Ape Academy, Meebits, and Mutant Ape Yacht club were 16, 1, 36, 33, 5, 13, 33 and 80 respectively. Trading volume also had a high day-to-day variation with the standard deviation of 15.99, 3.80, 104.63, 167.80, 28.60, 25.44, 123.02, and 392.53 respectively.

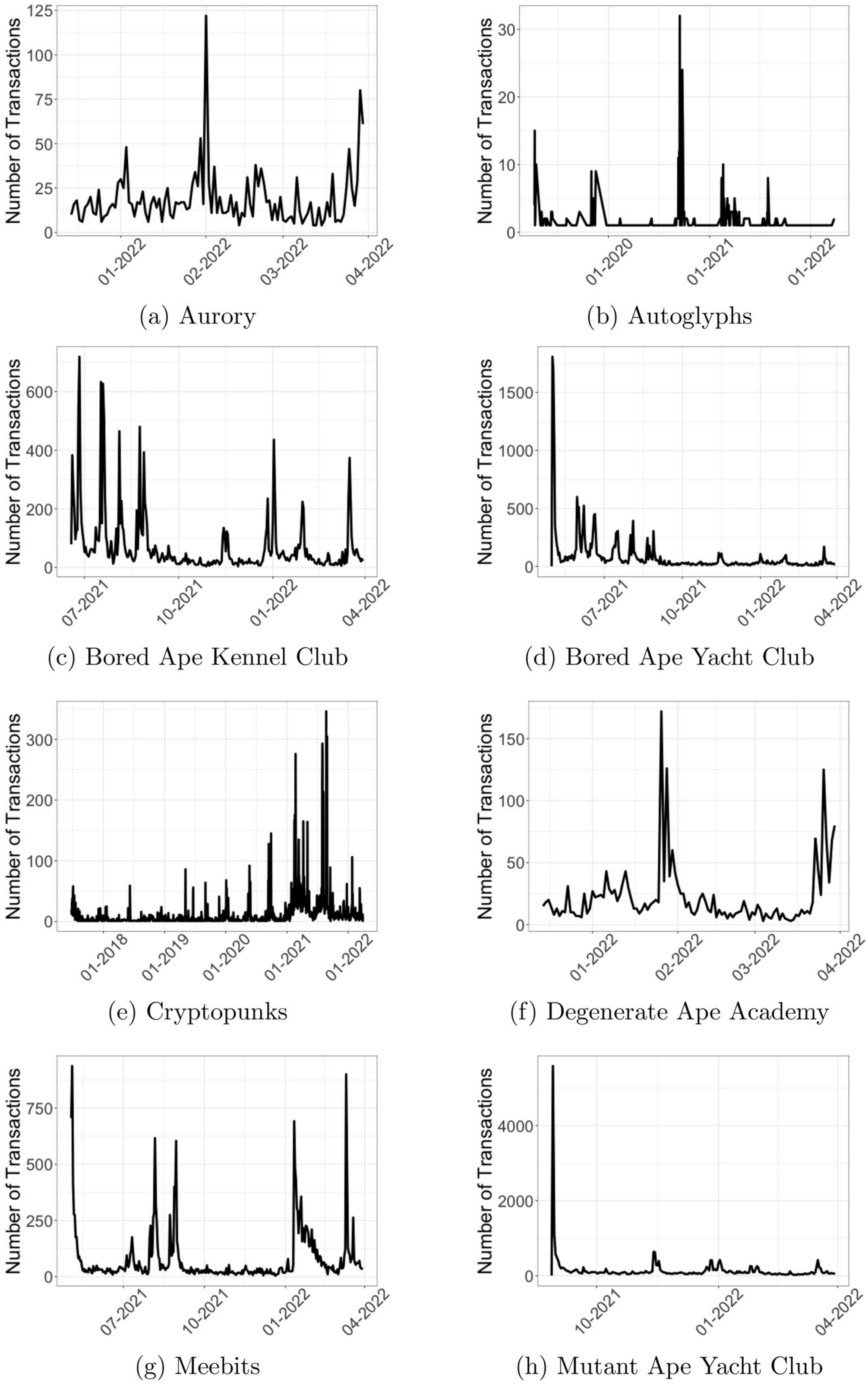
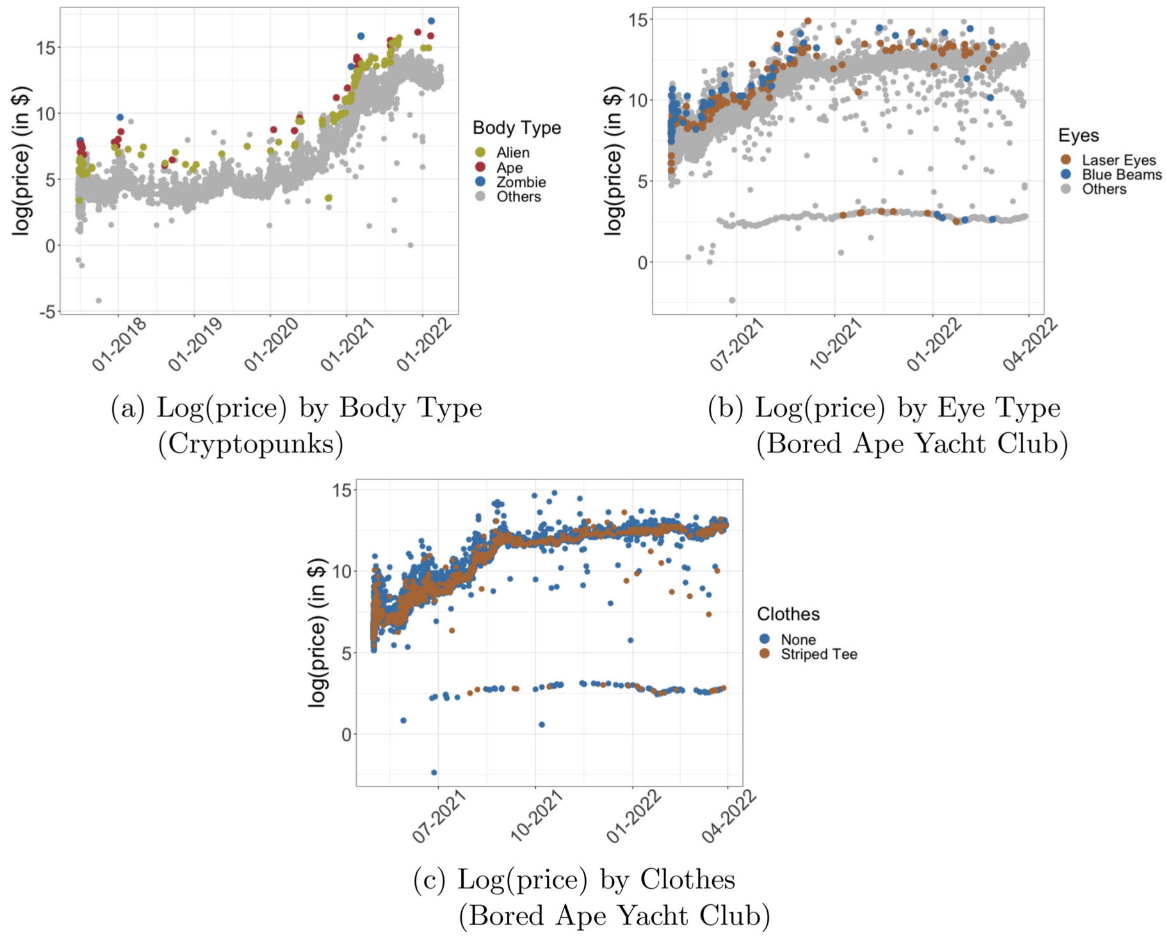


Figure 3. Number of transactions of eight profile picture collections (PFP) in USD from the project launch to March 31, 2022.

3. Challenges and Limitations

3.1. Traits

One of the main challenges to analyzing the price of NFTs at the collection level is the wide price range among tokens, as shown for two collections in Figure 2. Such differences can partially be explained by the design of each token. Tokens are generated according to a base pattern that is then complemented by giving them certain traits. Most collections are designed in such a way that a certain



(a) Log(price) by Body Type (Cryptopunks)

(b) Log(price) by Eye Type (Bored Ape Yacht Club)

(c) Log(price) by Clothes (Bored Ape Yacht Club)

Figure 4. Log(price) of tokens from Bored Ape Yacht Club and Cryptopunks by various traits.

(combination of) traits is much rarer than other combinations. For instance, the Cryptopunks collection consists of 10,000 tokens, distributed across six body types. However, there are only 6 Aliens, 14 Apes, and 24 Zombies, while the remainder of the tokens has either Male or Female body type. The price of a token with Alien, Ape, and Zombie body types was generally considerably higher than the ones with Male and Female body types (Figure 4(a)). Similarly, for the Bored Ape Yacht Club (BAYC) by YugaLabs, only 118 of 10,000 tokens have laser eyes or blue beams radiating from their eyes, and these are traded at prices noticeably higher than tokens with others Eyes traits (Figure 4(b)).

However, not all design traits appear to contribute to the price of a token. For example, in the BAYC collection, the most common Cloth trait is "none": 1886 of the 10,000 tokens have this trait. The next common trait is "Striped Tee," which 412 tokens have. While the tokens with "None" trait are about 4.5 times more common than the ones with "Striped Tee" trait, no noticeable differences are found in their prices, as is illustrated in Figure 4(c).

The relationship between the price of a token and its rarity is explored by calculating the rarity score and analyzing the linear relationship between the score and the average price of the token. The rarity score of a token was calculated by averaging the rarity of a token over multiple traits (hair, background, accessories, etc.) a token is designed to have, where rarity is simply the fraction of the number of tokens

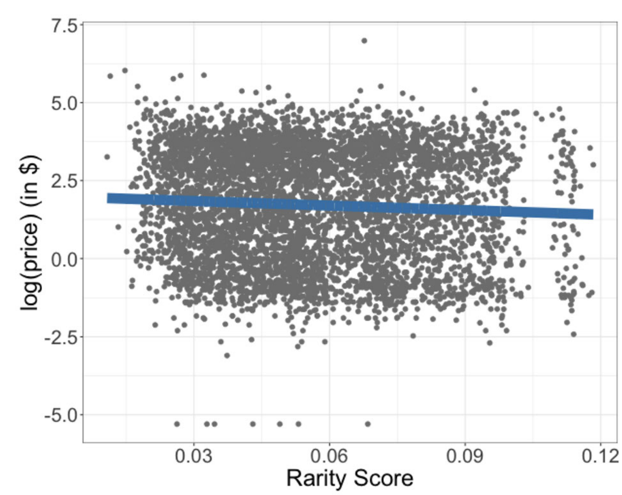


Figure 5. Log(price) vs rarity score of a token (Bored Ape Yacht Club), with the least squares regression line (blue).

with the trait to the total number of tokens in the collection. A negative correlation between rarity and the log average price of a token was detected (Figure 5), suggesting that rarer tokens are indeed traded at higher prices than more common tokens. The slope parameter, $\hat{\beta}$, is -2.7 and significantly different from 0 with the p-value of around 0.005. It means that for every 0.05 increase in the rarity score, the average price of the token decreases by about 4 percent … However, the R^2 of only around 0.001 suggests that the correlation is weak and that there are other meaningful factors that contribute to the price of a token. Here we also note that average rarity is only one way to compute token rarity[2] and other measures may lead to more significant correlations. For instance, Kong and Lin (2022) arrive at similar findings using a more sophisticated hedonic regression model for Cryptopunks transactions from the project launch through May 2021.

3.2. Lateral Swaps

One challenge to decoding NFT transactions on Ethereum is the presence of multiple smart contracts associated with a single transaction. In the simplest case, only one smart contract is required between two parties to trade an NFT. An example would be a peer-to-peer Cryptopunk token transaction. In this simple case, one may find the transacted price of a token by observing the `value` field in the transaction. Since there was only one smart contract involved in this transaction, the value contained in that field reflects the transacted price of the most recent transaction, denominated in the blockchain's primary currency (in this case, ETH). In transactions that involve more than one smart contract, the transaction price has to be retrieved by decoding further layers of smart contract executions, which occurred prior to the execution of the most recent smart contract. In fact, many transactions have a value field equal to zero at the top level but contain non-zero transaction prices in the token contract execution.

All of the above is accounted for in Gallop's data aggregation, yet some atypical transactions may still be found in decoded transaction data. For instance, when users use a smart contract to swap tokens laterally between wallet addresses, they may still have to pay the execution price of the swap contract, which will appear on the blockchain as the value. This results in a set of transactions recorded at a price of 0.005 ETH, which is considerably lower than the fair value token transaction price, as strikes the eye for collections such as Bored Ape Kennel Club, Bored Ape Yacht Club, and Mutant Ape Yacht Club, plotted in Figure 6. As these transaction records are not indicative of the fair value of the token, they should be excluded when performing token or collection valuation analysis[3].

3.3. Wash Trading

Wash trading is a process in which a market participant sells and buys the same token back multiple times within a short period to deceive other market participants about its price and liquidity. If such malicious market participants are successful, they can completely distort transaction price statistics. A good illustration of that effect is shown in the transaction price history for Meebits (Figure 6(d)). Meebits' transaction history follows an approximately smooth and continuous trend through January 2022, but then a vast amount of transactions occur at a wide range of prices deviating from that trend. Unusual transactions of the Meebits

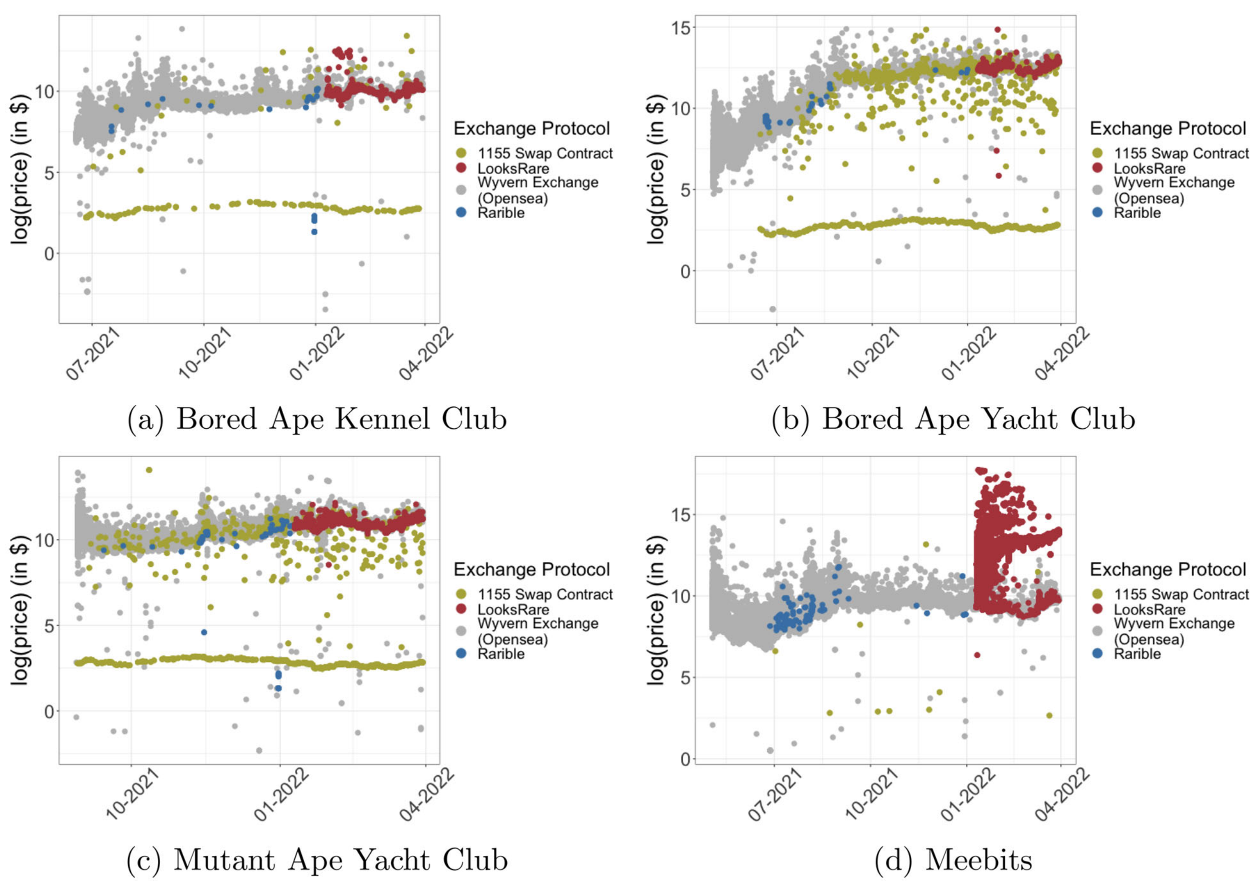


Figure 6. Log(price) of tokens from Bored Ape Kennel Club, Bored Ape Yacht Club, Mutant Ape Yacht Club, and Meebits by Exchange Type.

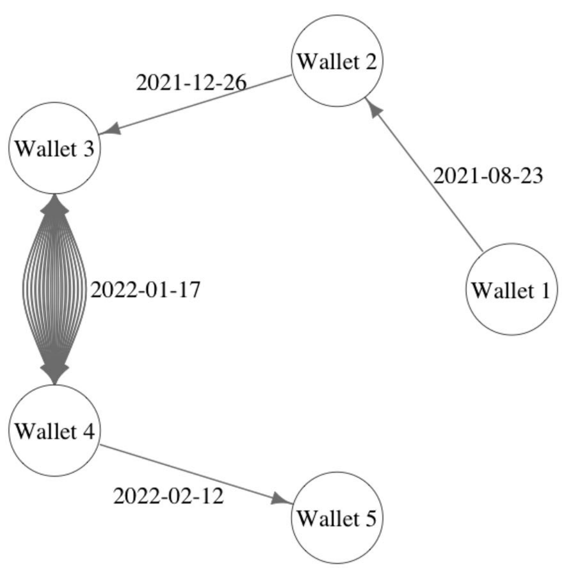
tokens garnered media attention and a few articles have been published on this phenomenon (Howcroft 2022). Many suspect that most, if not all, of the transactions are at significantly higher prices compared to the trend to be wash trades. As can be seen from Figure 6, which groups transactions by exchange protocol, almost all transactions at elevated prices have been executed using the *LooksRare* protocol.

LooksRare[4] launched in January 2022 as a new decentralized marketplace, aiming to take on the competition with the (then and now) dominant marketplace OpenSea[5]. In order to grow market share, the LooksRare protocol rewards users by paying them back a fraction of the transacted amount in their native LOOKS token. However, the LooksRare protocol thereby inherently encourages wash trading. By using LooksRare, wash traders reap a double benefit: they not only succeed at artificially raising average sales prices and volumes for their tokens but they also get rewarded in LOOKS tokens for doing so. Moreover, they can then decide to stake the LOOKS tokens, for which they will reap a second round of rewards in the form of Wrapped Ether (WETH) tokens. An example of how trading a single token in this way can lead to profits of approximately 3.5 million USD, is described in Coffman (2022).

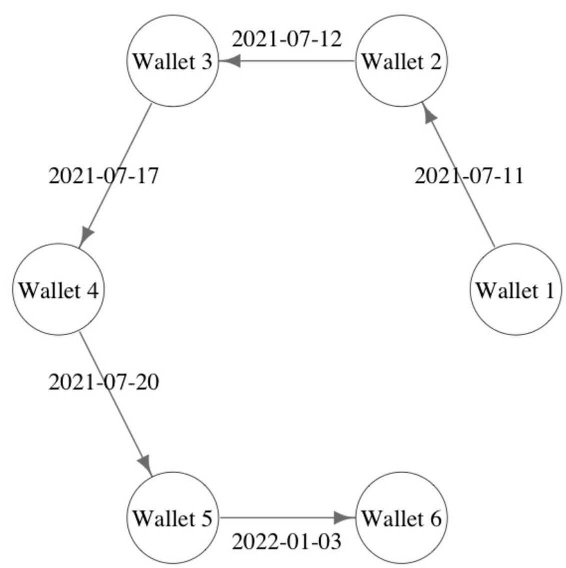
Wash trading can clearly distort market prices by creating a false sense of demand in the market. It thus needs to be accurately identified before analyzing the data. To propose a detailed algorithm for that purpose would lead us beyond the scope of this Data Note. Instead, we will provide a hint as to how either wallets or collections can be flagged for suspicious wash trading activity.

At first, a wash trade is by definition a trade executed between different wallets owned by the same person or entity. On the blockchain, wallets are anonymous and in theory, users can create new wallet addresses each time they trade. However, at least for non-automated traders or less sophisticated wash trading algorithms, creating a new wallet address each time is a cumbersome step. Therefore, one can expect that in such cases, the token traded would *return* to the originating wallet address at a given point in time. In several cases, the latter occurs a striking amount of times, as illustrated in Figure 7(a). Herein, the trade graph is plotted for two example tokens. On the one hand, Bored Ape Kennel Club #282 has unique wallets involved in all of its five transactions (Figure 7(b)), whereas Meebit #17021 was transacted 32 times on January 17, 2022, between the same pair of wallets, notably *via* LooksRare, before eventually being transacted on Opensea into a different wallet on February 12, 2022.

By the anonymous nature of NFT wallets, one can not detect wash trades with absolute certainty. Even though transactions involving BAKC #282 involve unique wallet addresses, it is possible that all these wallets are owned by the same party. Nevertheless, the 32 transactions between a pair of wallets involving Meebit #17021 are most likely the result of wash trading activity. Many other examples of repeated transfers between wallet pairs can be found in the Meebits transaction history. The latter would lead us to believe that a lot of the wash trading activity has been executed with a low degree of automation. Two holistic approaches investigating wash trading for the entire blockchain arrive at similar conclusions. At first, Tariq and Sifat (2022) corroborate the existence of wash trades on the Ethereum and Wax blockchains by performing hypothesis tests on the entirety of NFT transactions. They test for adherence to Benford's law, existence of clusters in pricing



(a) Transactions between 2021-07-11 and 2022-02-12 (Meebit #17021)



(b) Transactions between 2021-07-11 and 2022-01-03 (Bored Ape Kennel Club #282)

Figure 7. Transactions history of Bored Ape Kennel Club #282 and Meebits #17021. Each node represents the parties involved with the transactions. The arrow and the date represent the direction and the timing of the sale respectively.

and test if transaction prices fit Pareto—Levy's law, all of which significantly point to wash trades being present. Secondly, von Wachter et al. (2022) store the entirety of Ethereum transactions that involve NFTs as a directed graph and then extract all closed loops from it, which leads them to conclude that approximately two percent of all NFT transactions are suspect to being wash trades and that the vast majority of these only involve up to a handful of wallet addresses. To identify such closed-loop trades, they employ Johnson (1975)'s exhaustive algorithm to find elementary circuits in a directed graph. Recently, Serneels (2022) has proposed three strategies to flag suspicious wash trading activity in the NFT markets, one of which stools upon the idea presented above. The results therein corroborate the validity of the conclusions arrived at here.

While von Wachter et al. (2022) suspect wash trading to be present in most NFT collections, our analysis suggests that some collections are prone to a higher degree of wash trading activity than others. Figure 8 shows the ratio of the number of unique wallets to the total number of transactions within the data history. A low unique wallet-to-transaction ratio would indicate a large number of transactions occurring between only a few wallet addresses, which is indicative of suspected wash trading. Meebits stands out as a

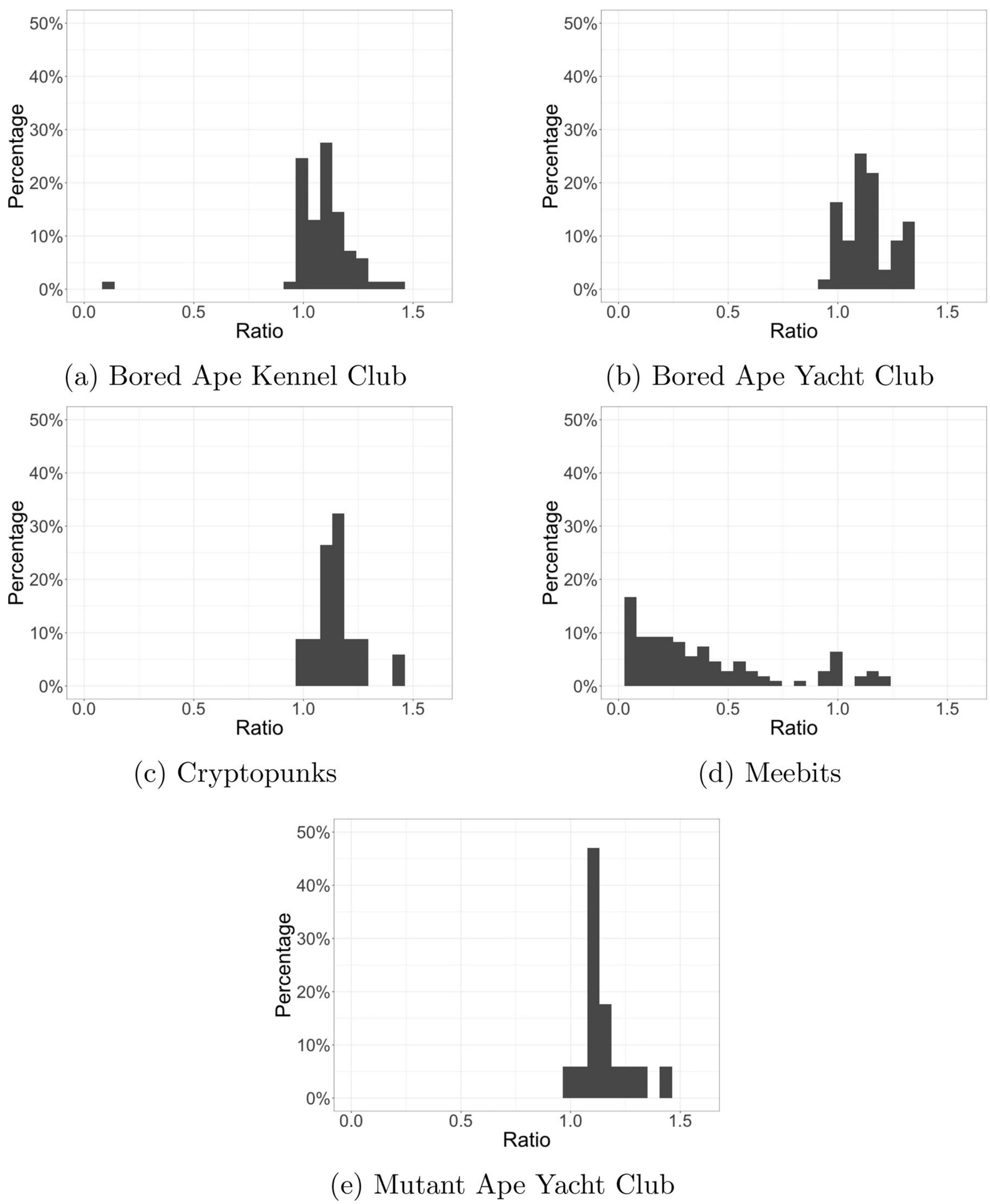
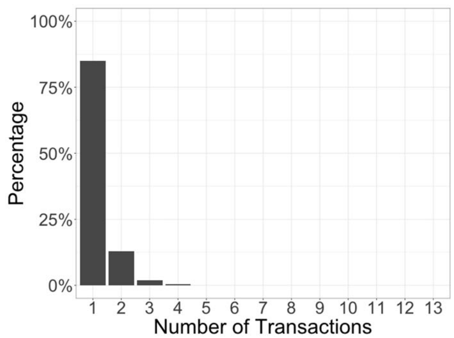


Figure 8. Distribution of the wallet to transactions ratio.

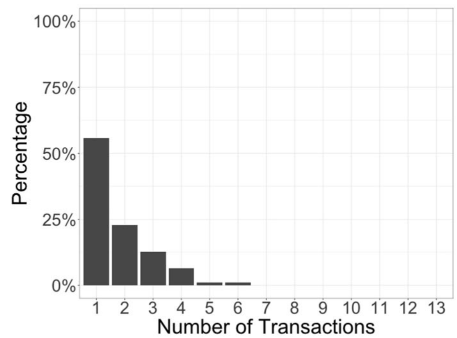
collection with a comparably low number of unique wallets to transaction ratio (Figure 8(d)). Despite LooksRare's effort to discourage wash trading by raising transaction fees and limiting the amount of LOOKS tokens to be rewarded in a day, the data suggest that many continue to exploit the system to this day.

3.4. Market Volatility

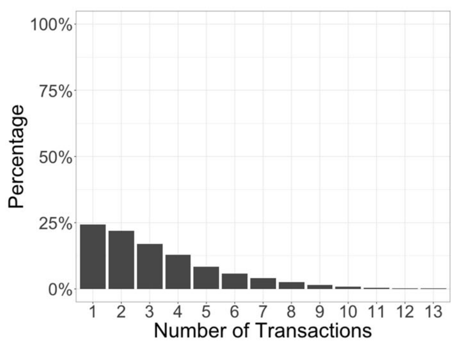
Another challenge associated with the analysis of NFT collections is its extreme volatility, as has been pointed out by Dowling (2022b), Borri et al. (2022), Kong and Lin (2022) and Mazur (2021). Unlike cryptocurrencies or equity markets



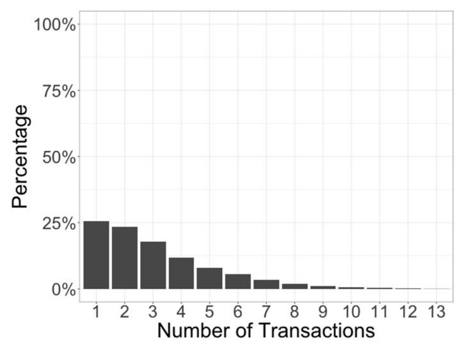
(a) Aurory



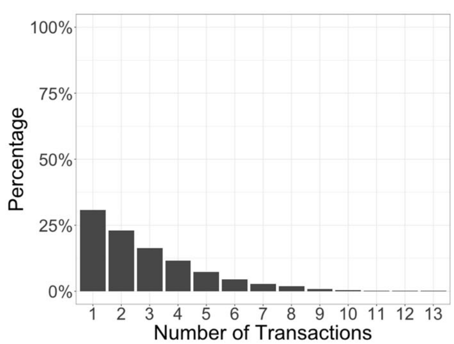
(b) Autoglyphs



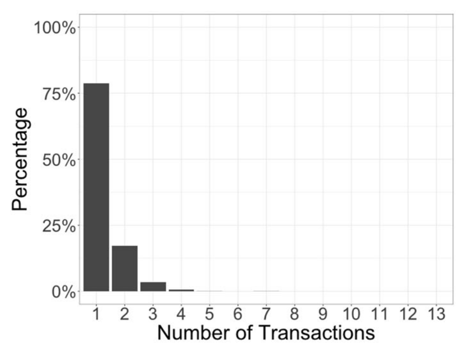
(c) Bored Ape Kennel Club



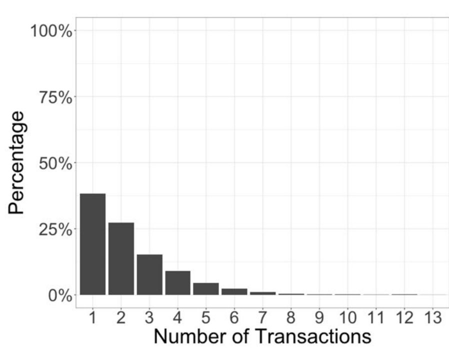
(d) Bored Ape Yacht Club



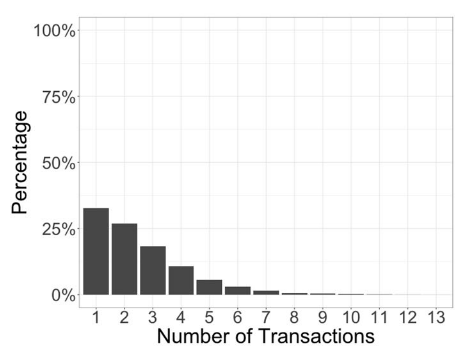
(e) Cryptopunks



(f) Degenerate Ape Academy



(g) Meebits



(h) Mutant Ape Yacht Club

Figure 9. Percentage of the number of transactions per token from the project launch to March 31st, 2022.

where a large number of historical data points exist at a regular interval, the NFT market is highly illiquid with sales occurring sporadically over time as seen in Figure 3. This may be one driver that causes the transacted prices to be extremely volatile, as suggested by Kapoor et al. (2022). Moreover, a large proportion of tokens are transacted only once after the collection launch. Figure 9 shows the distribution of the number of transactions of tokens that were traded at least once during the studied period. It also suggests that only a small fraction of tokens are traded more than three times.

Indeed, the volatility of all eight studied collections during the studied period was extremely high (Figure 10). The

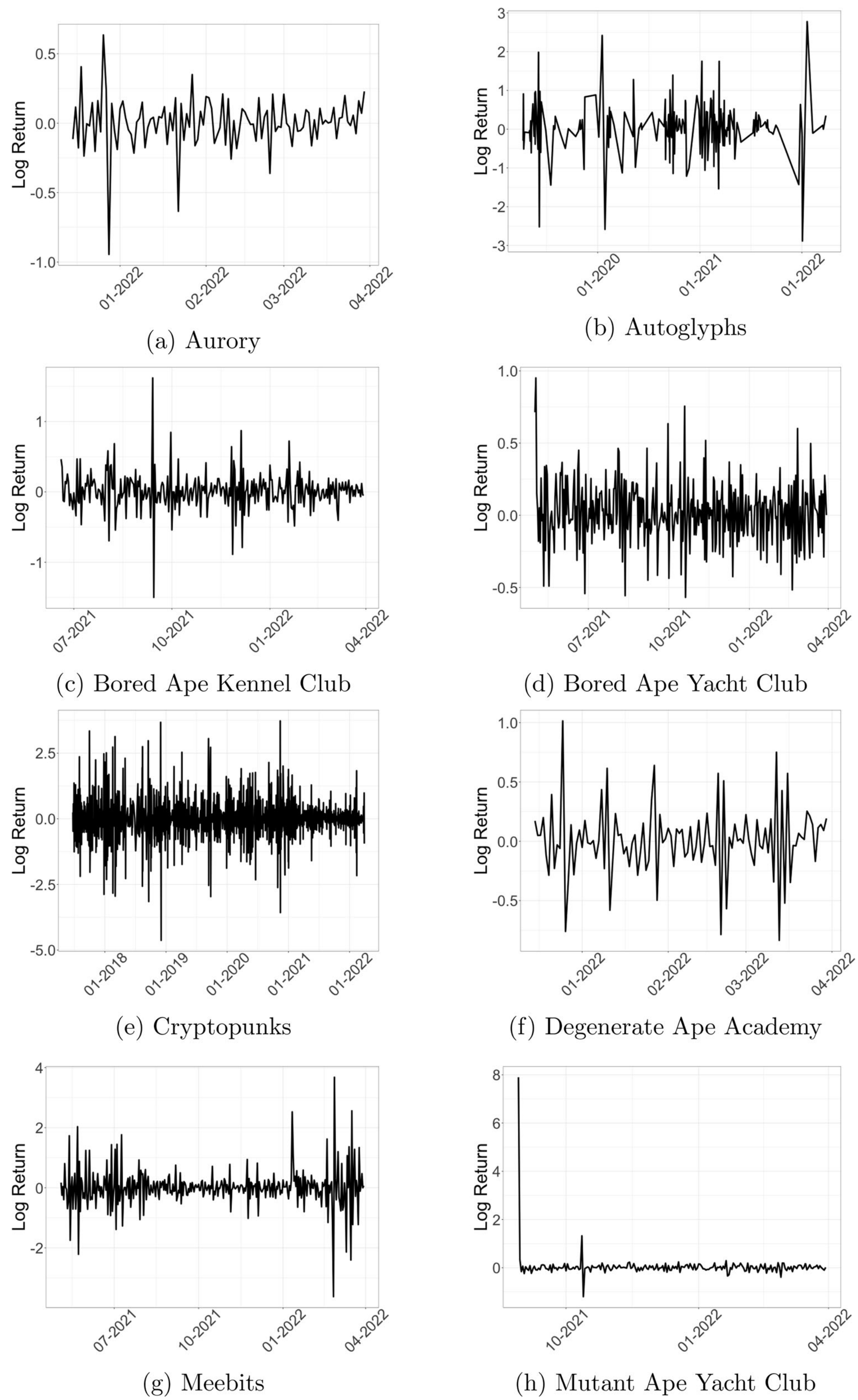


Figure 10. Average daily log return per collection.

Table 2. 12 Month realized volatility of eight profile picture collections.

	Aurory	Autoglphs	Bored Ape Kennel Club	Bored Ape Yacht Club
12-Month realized volatility	291%	1084%	420%	340%
	Cryptopunks	Degenerate Ape Academy	Meebits	Mutant Ape Yacht Club
12-Month realized volatility	1140%	464%	1013%	896%

average daily log returns at the collection level were rather stationary except for Mutant Ape Yacht Club. Average returns for Mutant Ape Yacht club were especially volatile for the first few days but became stationary with relatively constant variance throughout the studied period. 12-month realized volatility, the standard deviation of the daily log return times the square root of the number of trading days of 252, was computed for the eight collections based on the daily average transacted price of tokens (Table 2). Cryptopunks had the highest realized volatility of 1140% and Aurory had the lowest with 291%. For comparison, the annualized volatility of S&P 500 was 14.95% as of March 31st of 2022. The 52-week high of volatility for crude oil, a commodity that is known to be volatile, in the past year was around 102%, according to CBOE Crude Oil Volatility Index (OVX). However, in order to address the issue of illiquidity, transaction prices are being averaged or aggregated to a weekly or monthly level, thus not properly addressing the underlying volatility. For example, in Kapoor et al. (2022), which studies the relationship between the social media mentions of a given NFT token price, the token price is defined as the average selling price of the asset over all its historic sales. In Dowling (2022b), which studies the pricing of NFT of land in a virtual world, data are aggregated at the weekly level to be used for the analysis. In both cases, the average price of a token over an extended time horizon may not accurately represent the value of a token as it fails to capture the underlying high volatility.

4. Conclusions and Outlook

The emergence of NFTs initially spurred interest from blockchain enthusiasts and individual collectors. In recent years, though, as a growing number of NFT collections have become highly valued, the NFT ecosystem has *financialized* and NFTs can now be regarded as a novel financial asset class. Financial decisions rely on data and analysis thereof. This paper has described characteristics common to transaction data for many NFT collections and has illustrated that such data entail unique analytical challenges. Owing to these and other challenges, several authors have identified NFTs as an emerging field of research. Beyond doubt, the existing body of academic literature is still limited and many research questions regarding NFTs remain unanswered. Baals et al. (2022) recognize this fact and propose a research agenda forward with focus on the financial economics for NFTs. While research on financial economics for NFTs will encompass advanced data analytics, future research NFT transactions may offer a set of opportunities beyond that field as well, particularly in data science.

In fact, each of the challenges described in Section 3 may lead to novel data science methods and/or applications being proposed. For instance, to account for the high price differentiation according to certain, but not all, traits, one can imagine constructing a machine learning model that uses information from both transactions and token metadata, possibly even from the images themselves, to obtain accurate token specific price predictions. Section 3.3 hinted at ways to detect wash trading at the transaction and collection level. These ideas could be used to construct a robust detection algorithm. Moreover, NFT data that involve wash trades can be a fertile ground for classification and anomaly detection models, such as robust statistics. At the height of the Meebits wash trade wave caused by LooksRare's reward system, more than half of transactions were suspect wash trades. A fraction of over 50% of anomalies generally poses challenges to be detected by robust statistics as the latter will break down, which provides an interesting avenue for research. Finally, the volatility challenges highlighted in Section 3.4 require deeper investigation, should one want to apply downstream financial analytics, such as risk/reward quantification or NFT portfolio optimization.

As presented, NFT data provide a fertile ground for future data science research with its own unique challenges: price differentiation by token traits, the existence of lateral swaps and wash trades, and severe price volatility. We hope that the material presented here can be a seed for further developments in the field and that the data shared here prove to be useful in bringing about such developments.

Notes

1. https://www.higallop.com.
2. The company https://rarity.tools specializes in token rarity and offers two versions of highly customizable rarity models, the details of which would range beyond the scope of this paper.
3. The public Gallop API will filter these anomalous transactions out, yet they were retained in the data described here as they represent a challenge common in NFT transaction data.
4. https://looksrare.org/
5. https://opensea.io/

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This article was funded by the NSF (2114143, 1940276, 1940124 and 1934985).

ORCID

Jason B. Cho http://orcid.org/0000-0002-4801-5514
Sven Serneels http://orcid.org/0000-0003-1642-189X
David S. Matteson http://orcid.org/0000-0002-2674-0387

References

Ante L. 2022. The non-fungible token (NFT) market and its relationship with bitcoin and Ethereum. FinTech. 1(3):216–224. https://www.mdpi.com/2674-1032/1/3/17.

Baals LJ, Lucey B, Long C, Vigne S. 2022. Towards a research agenda on the financial economics of NFT's. https://ssrn.com/abstract=4070710.

Bao H, Roubaud D. 2021. Recent development in fintech: non-fungible token. FinTech. 1(1):44–46. https://www.mdpi.com/2674-1032/1/1/3.

Borri N, Liu Y, Tsyvinski A. 2022. The economics of non-fungible tokens. http://dx.doi.org/10.2139/ssrn.4052045

Buterin V. 2014. Ethereum: a next-generation smart contract and decentralized application platform. Ethereum.org. https://ethereum.org/en/whitepaper/.

Coffman C. 2022. Wash trading: who, what, why, and what should we do about it? blog.cryptoslam.io. https://blog.cryptoslam.io/wash-trading-who-what-why-and-what-should-we-do-about-it/.

Dowling M. 2022a. Is non-fungible token pricing driven by cryptocurrencies? Financ Res Lett. 44:102097. https://www.sciencedirect.com/science/article/pii/S1544612321001781.

Dowling M. 2022b. Fertile land: pricing non-fungible tokens. Financ Res Lett. 44:102096. https://www.sciencedirect.com/science/article/pii/S154461232100177X.

Howcroft E. 2022. Unreal demand? Irregular sales worth billions fire up wild NFT market. *Reuters*, February.

Johnson DB. 1975. Finding all the elementary circuits of a directed graph. SIAM J Comput. 4(1):77–84.

Kapoor A, Guhathakurta D, Mathur M, Yadav R, Gupta M, Kumaraguru P. 2022. Tweetboost: Influence of social media on NFT valuation. https://arxiv.org/abs/2201.08373.

Ko H, Son B, Lee Y, Jang H, Lee J. 2022. The economic value of NFT: evidence from a portfolio analysis using mean-variance framework. Financ Res Lett. 47:102784. https://www.sciencedirect.com/science/article/pii/S1544612322000976.

Kong DR, Lin T-C. 2022. Alternative investments in the fintech era: the risk and return of non-fungible token (NFT). https://ssrn.com/abstract=3914085.

Mazur M. 2021. Non-fungible tokens (NFT). The analysis of risk and return. http://dx.doi.org/10.2139/ssrn.3953535

Nakamoto S. 2008. Bitcoin: a peer-to-peer electronic cash system. Bitcoin.org: Bitcoin.pdf. https://bitcoin.org/bitcoin.pdf.

Oh S, Rosen S, Zhang AL. 2022. Investor experience matters: evidence from generative art collections on the blockchain. http://dx.doi.org/10.2139/ssrn.4042901

R Core Team. 2021. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. https://www.R-project.org/.

Serneels S, Cho JB, Matteson DS. 2022. Data containing transaction history and visual traits of eight highly valued non-fungible token (NFT) collections. Cornell Univ eCommons. https://ecommons.cornell.edu/handle/1813/111404. 10.7298/6jcj-9461

Serneels S. 2022. Detecting wash trading for nonfungible tokens. Financ Res Lett. https://www.sciencedirect.com/science/article/pii/S1544612322005517.

Tariq SA, Sifat I. 2022. Suspicious trading in nonfungible tokens (NFTs): evidence from wash trading. https://ssrn.com/abstract=4097642.

Umar Z, Abrar A, Zaremba A, Teplova T, Vo XV. 2022. The return and volatility connectedness of NFT segments and media coverage: fresh evidence based on news about the covid-19 pandemic. Financ Res Lett. 49:103031. https://www.sciencedirect.com/science/article/pii/S1544612322002690.

von Wachter V, Jensen JR, Regner F, Ross O. 2022. NFT wash trading: quantifying suspicious behaviour in NFT markets. Financial Cryptography and Data Security. FC 2022 International Workshops. https://ssrn.com/abstract=4037143.