Computer Science

Capstone Report - Spring 2022

Quantifying NFT Communities

Erol Bickici

supervised by
Olivier Marin

Preface

The inspiration behind my research was to contribute to a growing field that has great future potential, yet has a maniac-like current climate. I have personally always dismissed the idea of NFTs (non-fungible assets) as being speculative tools for a few lucky winners to pride themselves on, but over time I realized how there are real-world, useful applications of NFTs such as in real estate, verification, and loan programs. My overreaching motivation for my research is to serve as a first step to show that, at the very least, the current climate of NFTs is not based on pure randomness but is influenced by the various communities supporting them. I aim to encourage and target readers situated in a crossroad who, on one hand, may be hopeful for the long-term future of NFTs while, on the other hand, believe that the current landscape is far too random and baseless to ever grow.

Acknowledgements

First and foremost, I would like to thank Professor Olivier Marin for serving as a guide, advisor, and a critical source of knowledge throughout my entire research. I would also like to thank Professors Prométhée Spathis and Xianbin Gu for also pushing my research to the best it can be whether it was the research direction or technical suggestions. Additionally, I would like to thank Professor Sukjoon Lee and Yegor Mostovoy for assistance in making sense of the statistical angle of my research. Finally, I would like to thank everyone who I annoyed with NFT questions and clarifications to help me better understand this topic that I initially was extremely skeptical to address but ultimately found a newfound respect for.

Abstract

As NFTs (non-fungible tokens) continue to explode in popularity, there is a rise in false, misleading, and/or scam-ridden collections with few tools to gauge the legitimacy of these collections and to value growing NFT projects. The NFT world is seemingly dominated by speculation and a mania to get in on this new trend. I argue that this compromises the greater potential that NFTs could provide in the long-term. My approach aims to examine what I believe to be the backbone of NFT collections, the community behind it, and to analyze its relationship to popular collections to justify valuations. My findings suggest that more expensive NFT collections have a more distributed ownership structure based on their Gini coefficients, although their correlation is weaker than expected with a few standout outliers at the higher end. Additionally, there appears to be a certain Ethereum price range for NFT collections (between 0.05 to 0.07) that exhibits a sudden uptick in correlation compared to that of other price ranges, suggesting external price-influencing factors.

Keywords

NFTs, Collections, Ownership, Centralization, Valuations, Community

Contents

1 Introduction

NFTs, or non-fungible tokens, are tokens that represent a unique asset or item stored on a blockchain [1]. Each NFT has unique characteristics, hence the non-fungible aspect, that make each individual NFT not tradeable on an exact 1:1 basis, unlike cash where one $20 bill is perfectly substitutable for another $20 bill. While NFT characteristics have become increasingly more complex and diverse, the vast majority are attached to content that can include a digital image, such as a JPEG or PNG. All this content is hosted distributedly on the blockchain which make it immutable. It also allows each NFT to be tradeable by cryptocurrency, such as ETH (Ethereum), attaching NFTs to the hype of cryptocurrency in recent years.

NFTs have been around since 2014, but only saw wide-scale popularity and usage in early 2021 [2]. A key factor of this growth was the investment opportunity presented by buying NFTs and holding them in the hopes they would appreciate in value, which resulted in total NFT sales in 2021 of $24.9 billion from $94.9 million in the year prior [3]. Some NFT collections, or a group of individual NFTs with identical structure but differing attributes (most commonly between 100 - 10,000 total items), have had a few, such as the wildly popular Bored Ape Yacht Club (BAYC), having their 10,000 items trading with a floor price (or minimum price) of 108 ETH, or just over $234,279 each (priced at the time of writing where 1 ETH =$2,169.25) [4].

A key driver to the huge success of many NFT collections is the community that owns and supports the collection. The BAYC, for instance, has organized in-person Yacht parties and free cryptocurrency distributed to its NFT owners while U.S. politician Andrew Yang's GoldenDAO, another NFT collection, promotes AAPI (Asian Americans and Pacific Islanders) solidarity and empowerment by in-person meets, discussions, and business initiatives - exclusively for their members [5][6]. Each of these collections starts from a small group of founding members who can begin initiatives such as the ones mentioned, but sometimes are overcome by selfishness.

The emergence of this rapidly growing, entirely new, and unique industry with millions in transactions everyday unfortunately has not gone without bad actors. Most NFT collections are compelling purchases due to their utility, brand, or financial uplook. However, some individuals and groups are capitalizing on the popularity boom for the sole purpose of garnering short-term profit at the expense of early, enthusiastic backers. One NFT collection named FrostiesNFT came under fire earlier in early 2022 for pulling a "rug-pull", or vigorously promoting a collection to only quickly take profits and run, which resulted in a scam loss of $1 million as well as money-laundering charges pressed by the Department of Justice against its original creators [7]. Collections like Frosties are not isolated instances as OpenSea, the largest marketplace for buying and selling NFTs, has recently exclaimed how "80% of the NFTs created for free on the platform are plagiarized works, fake collections, and spam" [8].

Catalyzing the effect of these inherently destructive or malicious collections is a lack of tools to combat them. The NFT market has yet to mature enough to allow for tools that can vet the validity and value of growing collections, from both a consumer and business perspective. Tools such as Nansen and Trait Sniper, amongst others, largely focus on identifying investment opportunities amongst NFTs by viewing insight such as trading volume, media coverage, or a specified NFTs' content attributes, such as an image's hair color [9][10]. These tools are certainly useful for seeing and capitalizing on NFT trends while also analyzing one's current portfolio, but fail to really conceptualize on how the community that backs these collections relates to its own valuation.

This failure to analyze NFT communities is surprising due to the importance that many place on the communities for future growth. The demand for the most expensive collections, in particular, have been frequently cited as being extremely popular, exclusive clubs that represent "people finding community through a shared experience" [11]. The NFT mania has gathered waves of newcomers to the fields of NFTs with the hopes to join a collection with a strong community backing, but instead now are facing a roadblock.

Taking a step back, the larger problem here is the negative image that NFTs carry to the public. Current sentiment over NFTs is mostly along the lines of "it's just an image" or "it's just based on pure speculation and nothing else", which is synonymous of all NFTs from those with pictures of literal ducks to also those that can be used for real social good, such as hosting user identity, land deeds, and even combating child labor - all solutions to issues currently facing developing nations [12][13]. My motivation stems from this false, narrow narrative that many have over NFTs, where I aim to find the sound in the noise by examining what I believe to be the root cause of NFT growth, and specifically valuation: the community.

The question I wish to address is if and how NFT communities contribute to the growth of their corresponding collections. My research will focus on examining a wide-variety of collections and their communities from popular (popular being defined as recent trading prices) collections to those that have been generally accepted as scams. Specifically, I wish to examine community characteristics such as the NFT collection's ownership distribution.

2 Related Work

Current research into NFTs is rather limited due to its young history, putting my research into a unique situation to be built off a rather wide diversity of existing studies. While there are few studies that focus on the ownership of NFTs, there are still pieces of existing research elsewhere that individually are limiting in my hypothesis' scope, but each still having aspects that contribute to my research when concatenated together.

2.1 Valuation

Before analyzing a community's relationship with NFTs, it is first important to dissect the core concepts of what influences NFT valuation. One study conducted by Olav Velthuis sets the stage for the influencing variables by examining real-world artwork and explains how restricting valuation purely on price leaves out other potential characteristics that can influence valuation, such as a social, cultural, or community value attached [14]. While Velthuis does suggest a community influence on pricing, his research excludes NFTs while also suffering from loose connections and explanations of these external effects and how they can influence valuation, which is where my research will focus on.

Velthuis's findings are also supported by research by Pavel Kireyev, but with an NFT scope. Kireyev's methodology was to use structural models built around bidding and asking prices for NFTs, ultimately pointing to a psychological factor that influenced valuations (such as having more bids for an item naturally makes it more appealing to non-NFT holders) [15]. Kireyev's research, although narrower in focus to NFTs specifically, is unfortunately too broad in scope as it views NFTs at the market level and not considering individual collections and the existing community - choosing instead to focus on the auction side.

Fazli et al. somewhat address this void, by having their research focus more on finding valuation amongst the NFTs themselves and in relation to similar 'siblings' [16]. They accomplished this by utilizing deep learning and computer vision to extract the contents of NFTs and then used cosine distance as a similarity metric amongst images. They were able to estimate the selling prices of NFTs where 80% of all historical sales fell within 1 Ether range of its 5 nearest neighbors by drawing up relations amongst a variety of NFT projects. Now while this research demonstrates correlations amongst the NFTs themselves, there are no real metrics on the communities that back them or a quantifiable value attached, but there are findings that similar collections have similar communities.

In terms of quantifiable variables in NFT communities, Lucey et al. suggest that there may be challenges to approaching a methodology to gauge communities [17]. Their research draws correlations between NFT hype and the cryptocurrency market, but they discover that this is not consistently true as NFTs do not follow typical "asset pricing detection models" that can be used to draw correlations with its backing community. Now while there has yet to be any research that directly demonstrates the connection between these unexplained bubbles and NFT communities, we can reasonably assume from Pavel Kireyev's study that the price increases are likely due to media propagation and certain key NFT-related events, such as with Christie's Cyberpunks auction [15].

2.2 Community

Specifically focusing on NFT communities, Casale-Brunet et al. analyzed the community makeup and financial history of a select few NFT projects while also being the first to really emphasize and embrace community and its influence on NFTs [18]. Their study examined 8 large NFT projects to topologically map out the activity of each collection. Surprisingly, their research showed stark differences amongst the collections when it came to commonality (being involved in more than 1 NFT community simultaneously) of wallet holders. They measured this using a reciprocity coefficient that "measures the proportion of mutual connections on a direct graph" which showed a very low exchange between wallets after minting (the initial NFT release). The reciprocity is shown below:

Collection	Reciprocity	Transitivity	Assortativity
HashMasks	0.049	0.019	-0.015
Art Blocks Curated	0.040	0.039	0.313
CryptoPunks	0.044	0.058	0.474
Bored Ape Yacht Club	0.058	0.012	0.026
Acclimated Moon Cats	0.021	0.032	0.616
CryptoVoxels	0.048	0.012	0.020
Decentraland	0.071	0.037	0.109
Meebits	0.064	0.012	0.057
ALL (merged)	0.054	0.023	0.045

Table 1: Network graph connectivity and clustering statistics for each NFT collection in terms of terms of reciprocity, transitivity, and assortativity,. The merged graph considers all collections together [18].

A limit with Casale-Brunet et al.'s research, however, was the very small, handpicked sample size they analyzed. The NFT collections they selected, such as Cryptopunks and Bored Ape Yacht Club are well-known for high community engagement being two of the highest traded NFT collections, which limits the researchers' findings on only these prestigious collections selected because they were the "most significant ones in terms of both market capitalization and popularity, across categories such as profile pictures, curated digital art and metaverse " [18]. My research aims to focus on a larger range of collections where my results would ideally represent the whole NFT space that takes into account smaller, less active NFT communities as well.

For the methodology of detecting these communities, we can turn our attention to Papadopoulos et al. with their research into community detection methods, particularly around social networks [19]. Their research discusses a variety of definitions and attributes of social graphs with the 'weighted membership' being the most related to NFT collections due to some wallets owning more of a collection than others. The study expands by exclaiming the difference in popular detection methods based on the graph type such as with cohesive subgraph directories (where community satisfaction is specified) or vertex clustering (with distance between nodes being calculated). While these methods are certainly useful for the first step of detecting and calculating a community, there is no reference to NFTs in their research which are affected by hype cycles and other unique events and therefore should be treated differently than social communities, particularly because of the financial incentive of NFTs.

Cyclical and external events may have an influence on an NFT's community, and hence its price, however, other research by Baytaş et al. approaches NFT collection influences from a wide variety of angles including that of the collection owners, developers, original creators, investors, startups, and even corporations [20]. This angle was one of the first mentions of describing the stakeholders involved in NFT projects, but with this wide range of angles comes a lack of depth for each one. Focusing on the community, Baytaş and his colleagues do touch on the importance of the current owning community to influence the utility that comes from projects, thus correlating the community with the growth, or potential future growth, of the collection.

2.3 Ownership

Let us now shift our attention to the direct makeup of NFT communities and how its ownership distribution can correlate with valuation. Chang et al. conducted research into network DEA models and constructed approaches for three different ownership methods: centralized, distributed, and hybrid [21]. While their work relating directly to DEA models does not correlate with my own research, a very interesting finding of their research is the clear effect of the ownership distribution of their systems based on efficiency scores generated by their DEA models. This point will be carried throughout my own research as I aim to really see how negative (or not) the ownership structures of NFTs are to its underlying valuation.

Expanding on the ownership distribution, one study by Vasan et al. examined these ownership structures in relationship with its NFT peers [22]. Their research focused heavily on artists/creators on the Foundation network (another NFT marketplace) where users are able invite others onto the platform to promote their own NFT artwork. Vasan et al. leveraged these inherently built clusters to find relationships amongst its users. They formed these clusters by linking together invitational links to find correlations of higher average sales per user with the highest earning clusters. Amongst these users, they also found that individual sales prices tended to be relatively stable at an individual level between predictable ranges. Additionally, they found that amongst the bidders of the 'higher' tier artists, there were only a small number of collectors that actively partook in the bids, rather than a vibrant, large community. These findings demonstrated consistency amongst the NFT community with artists and collectors but stopped short of comparing clusters, or communities, with one another besides the references of total revenue. Combining this research with Chen's study that examined the overall benefits of a decentralization network, or in our case a decentralized community, we can suggest that common actors owning multiple similar collections as Vasan et al. pointed to, may be a negative influence on NFT collection valuation [23].

3 Solution

My key hypothesis is that more expensive NFT collections would have larger, more active communities compared to less-expensive ones. I aim to analyze the ownership distribution of NFT collections to examine if the higher-end collections would have a more equal distribution of its assets compared to lower-to-middle tier collections (tiered by ETH price).

In order to test this hypothesis, I would first need to collect NFT data that contains the owners' wallets, their assets, and the recent price. To collect this data, I relied on a 3rd party solution that found and parsed the NFT data I was looking for. Here, I used Moralis as the data provider which did the heavy-lifting of interacting directly with the blockchain on my behalf, while running the workload in Google Cloud's Cloud Functions in a NodeJS environment. In my workload, I filtered through only the necessary attributes for my research and computed my calculations, finally storing the results in a NoSQL Firestore database hosted on the cloud.

For my research, I decided to keep the idea of 'popular' collections simple and straightforward so as to not over-complicate my studies and keep a limit on controversy in my methods leading up to my results. For this reason, I decided to settle on a 'popular' NFT collection being defined by the average price of the most recent trades (past 500 if possible) on a given collection. This method differs from a study by Pinto-Gutiérrez, Gaitán, and Velasquez which suggests that in order to gauge for popularity, a collection's Google search popularity can be examined, but this unfortunately has a bias against hugely popular collections or a recent media discussion on a given collection[24]. For my research, I wished to gauge popularity based on 'skin in the game' popularity rather than merely general queries.

For calculating the average, I use three different methods: standard mean, trimmed mean, and 2SD (standard-deviation) cutoff mean. For trimmed mean, I calculated the mean of only the middle 60% of values while for 2SD mean, I calculated the mean of values that appear only within 2SD of the median. The use of these methods is due to the outliers in NFT trades that would skew the data drastically one way, such as in the case where some NFTs were gifted for free or accidentally sold for extremely low prices, such as in the case of one Bored Ape accidentally sold for 1/100th of its intended price due to a "lapse of concentration" [25].

Armed with a relative standard to compare the 'popularity' of collections, I will then be dividing these collections into differing tiers based on their prices. The goal with this division of tiers is to establish a way to easily compare collections from lower tiers with those of higher tiers. For example, how do collections within the 0.00 to 0.02 ETH price range compare to that of the tier that ranges between 0.10 to 0.15. These ranges for each tier are set to accurately split up the collections into relatively equal chunks where each of the 10 tiers consists of around 10% of the total collections, providing for a distributed dataset to view and compare my analysis.

For my analysis, I will be calculating and studying different metrics that will gauge the ownership distribution of each NFT collection.

The first and most important metric is the Gini coefficient, a measure of inequality by viewing distribution of ownership, commonly used in economics for measuring economic inequality in a populace with a measure of 0 being perfectly equal ownership and 1 being perfectly unequal [26]. To compute the Gini coefficient, the first step is to get the given population and find their ownership of an asset. In my research, this step involves getting the wallets and how many NFT assets of a given collection each wallet owns, then to graph the Lorenz curve and calculate the area difference between the curve and a x = y linear line, demonstrated below:

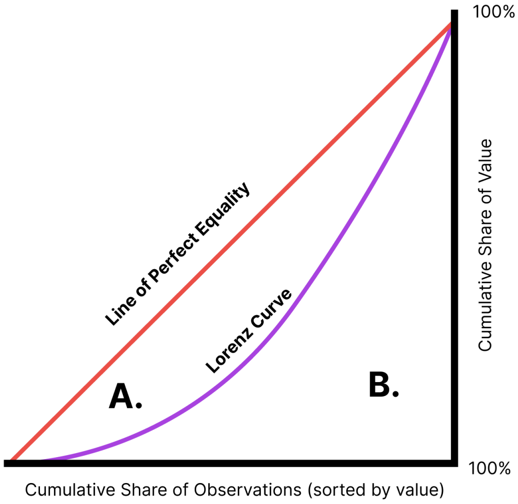

Figure 1: The Lorenz Curve and Gini Coefficient where: Gini = A/(A+B)

Another metric I aim to use is Pearson's coefficient to map correlations between the popularity of collections, once again defined by recent price history, and to ownership attributes such as Gini. For correlation, Pearson's coefficient is ideal for gauging the relationship between two continuous, random variables by determining if a linear correlation exists and also suggesting the strength of the correlation [27]. The formula is given below:

$$P_{x,y} = \frac{\sum x \cdot y - \frac{\sum x \cdot \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \cdot \left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

Other metrics I use includes the average wallet size, unique owners to total assets ratio, and total unique owners amongst others.

4 Results and Discussion

My results span a total of 1,858 collections available on the Ethereum blockchain with a total standard mean ETH price of 0.656, or approximately $1,423.03 (at the time of writing), and an average Gini of 0.4570.

4.1 Distribution

My first finding is that the Gini coefficient generally does decrease with higher NFT valuations, indicating a more distributed ownership amongst its assets, unlike its lower-end peers, although this correlation was not as apparent as my initial hypothesis would have suggested. This is likely due to small collections having fewer owners coupled with its relatively cheap cost that allows for these minority owners to easily become majority owners. Higher-tier collections initially set themselves up with a high-price barrier that limits any single owner from owning multiple items very difficult due to cash constraints (i.e., not everyone can afford to purchase 100 NFT assets worth $1,000 each).

Collection ETH Range (in ETH)	Total # Collections	% Of Total Collection	Average Price (ETH)	Trimmed Average Price (middle 80%) (ETH)	Average Gini	Average Wallet Size (# total assets)
0 - 0.02	189	10%	0.0125	0.0108	0.4722	89.0256
0.02 - 0.035	189	10%	0.0277	0.0273	0.4805	82.6314
0.035 - 0.05	180	9%	0.0425	0.0422	0.4950	78.9236
0.05 - 0.07	191	10%	0.0594	0.0598	0.4646	79.9958
0.07 - 0.1	217	11%	0.0851	0.0839	0.4792	78.3392
0.1 - 0.15	182	10%	0.1226	0.1230	0.4759	78.1122
0.15 - 0.25	210	11%	0.1968	0.1933	0.4423	72.6717
0.25 - 0.5	187	10%	0.3575	0.3533	0.4173	68.1281
0.5 - 1.5	172	9%	0.8813	0.8903	0.4291	66.5970
1.5 - 50	167	9%	6.4096	6.1892	0.3862	66.0779

Table 2: NFT collections data at various pricing tiers with calculated Gini. Gini and wallet size generally decreasing with more expensive NFT collections.

My findings with overall wallet sizes also support this claim that higher tier collections limit the total assets that their owners can purchase with another linear decrease in wallet sizes relative to their collection sizes. Here there is a clear decrease in the average wallet size based on the collection price. Cheaper collections naturally make sense for owners to own more since it's cheaper to own while also opening the door for investment speculation. We see this speculation in other industries such as with penny stocks too where a common investor sentiment is that at least one of these must be a winner, therefore I will purchase as many as I can, suggesting similar speculation to NFTs that exist in other financial areas such as in Penny Stocks [28].

Returning to the Gini coefficient, one surprising find is that in a few of the most popular, expensive collections, there are clear deviations in their Gini's from their relative peers. Cryptopunks, for example, is an NFT collection considered in the 1.5-50 ETH tier, with an average Gini coefficient of 0.3862, yet Cryptopunks has a Gini of 0.6089, a +0.3028 or a 57.66% difference compared to its peers in the same price-range. My initial theory here was that NFT collections such as Cryptopunks have had their Gini inflated by outside media coverage. Being one of the most well-known collections, I hypothesized that this has led to more individuals desiring to hold on for the sake of being a part of this hugely desirable collection and refusing to sell, concentrating the ownership in a few owners. This argument, however, can be debunked by examining collections like the BAYC with a Gini of 0.3286 that closely resembles its peers' Ginis in the 1.5 to 50 ETH tier, while also generally being accepted as being as desirable as Cryptopunks, if not more so, suggesting an alternative external reason [29].

4.2 Correlation

With a seemingly linear decrease in Gini coefficient and price, my results in the Pearson's coefficient did not support this claim as I hypothesized. Comparing all three means (standard mean, 20% trimmed mean, and 2SD mean), the correlation seemed to be close to 0 for all 10 tiers of collections. Interestingly, however, there is a sudden uptick in the correlation between the price of an NFT collection and its Gini at the ETH price-range tier of 0.05 to 0.07. The results are displayed below:

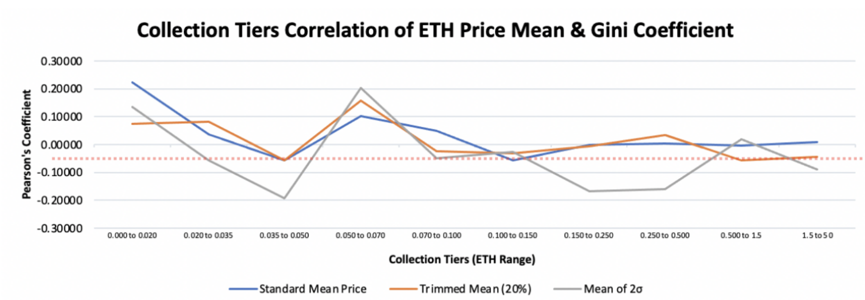

Figure 2: Three mean prices (standard, trimmed, and SD) with corresponding correlation at various price points. Notice the spike in correlation at price tier of 0.05-0.07.

This sudden uptick reaches a correlation around 0.2 which, according to scientific consensus on the Pearson's coefficient, represents a fairly weak correlation, although existing [27]. Calculating for significance, or the likelihood that this single uptick is precise, represents a p-value of 0.1585. This insignificance is likely due to the limit on data at this specific tier where it accounts for 191 collections which is not enough to confidently say that its Pearson's coefficient is accurate, or statistically significant.

The reason for this uptick is certainly puzzling and shocking as it appears for all 3 means, skyrocketing its correlation from close to 0 for standard mean and trimmed mean while being -0.2 for 2SD mean. The lack of p-value significance comes with a lack of certainty to this bump which compels further research to be done to confirm or deny this anomaly.

	Mean Price (ETH Range 0.05-0.07)
Mean Price (ETH Range 0.05–0.07)	1.0000
Gini	0.1024
p-value	0.1585

Table 3: Shows calculated Gini and p-value for NFT collections with average mean price between 0.05-0.07 (in ETH), or the tier with the 'bump' in correlation. The p-value of 0.1585 indicates the correlation as statistically insignificant.

Taking a step back and examining the correlation between price and Gini in its entirety, my findings suggest that a weak relationship does exist generally with NFT valuations in collections. Here, the correlation is -0.1146, -0.12387, and -0.73567 for standard mean, trimmed mean, and 2SD mean, respectively, with a p-value indicating significance of 0.0000, suggesting it as significant and accurate.

	Mean Price (ETH Range 0.00-30)
Mean Price (ETH Range 0.00–30)	1.0000
Gini	-0.0973*
p-value	0.0000

Table 4: Shows calculated Gini and p-value for NFT collections with average mean price between 0.00-30 (in ETH). Note, the star indicates significance.

Collection ETH Range (ETH)	Total # Collections	% Of Total Collections	Average Price (ETH)	Average Gini	Correlation (Average Price & Gini)
0.05-10	1287	68%	0.5897	0.4388	-0.1146

Table 5: Shows middle 70% of NFT collections with calculated Pearson's correlation along price and Gini.

5 Discussion

One early issue I faced was deciding on how to obtain the data. Interestingly, and almost ironically, blockchain data revolving around NFTs was not as easily accessible as originally anticipated. The most obvious first approach was to utilize the API of the current market leader in NFTs: OpenSea [30]. An issue I faced, however, was that you must request an API key from them directly and specify how the API would be used. While seemingly straightforward, I unfortunately did not get a response back until over a month later, so I had to switch services. The provider I eventually used, Moralis, was also not without its issues [31].

The biggest issue I faced throughout my research was the spam-prevention that Moralis utilized. API call abuse is not unique to this one provider, but it still limits the data processing I am able to accomplish. In order to limit call deny requests and still fetch the data effectively, I had to throttle the requests with a few random seconds delay between each call. Without the delay limitation, my research could have possibly incorporated more NFT collections data while also providing me with greater flexibility for adjustments if needed, such as if I missed a certain attribute or wanted to adjust how far back I viewed the collection trades.

Another adjustment I made during my research was to switch from using Google Cloud's Compute Engine to Cloud Functions. I began planning to use Compute Engine since my research initially wanted to account for other factors besides just NFT ownership such as social media influence and more technical financial analysis, such as floor price and volatility. To focus up my research, I ultimately shifted my analysis on exclusively NFT ownership which subsequently made me shift away from my initially planned architecture and tools. Google's Compute Engine is more suited for a large range of loosely connected functionality, or simply more complex workloads while Google's Cloud Functions is suited for single, simple functionality calls, which was more suited for my updated, focused analysis.

6 Conclusion

In terms of communities and their NFT collections, there seems to be findings of how its owners are correlated with their price. However, this correlation is weaker than originally expected. On one hand, the ownership distribution, governed by the Gini coefficient, seems to be decreasing with higher-priced NFT collections. On the other hand, the Pearson's correlation shows only a weak correlation, with a sudden hike at the ETH price range between 0.05 to 0.07, suggesting external factors that potentially influenced the result. I suggest that these external factors, such as media coverage or social buzz, also play a role in influencing some notable collections as well, such as Cryptopunks whose Gini coefficient is detected to be 57.66% higher than its relatively priced peers. While my research does not definitively declare any sort of explanation with absolute certainty for these outliers, I do believe that my findings can suggest other areas for research expansion.

The first suggestion I have for future research is to get more in-depth, complete popularity measurements. My simplistic approach and definition of finding the average price based on recent history is likely to have skewed my results with an unfair bias against collections with a lower number of total NFT items. For example, a collection with fewer items would likely have its 500 recent trades happen over a long stretch of time with a wide-vary in its price due to Ethereum's price volatility. A possibly fairer analysis would be to not only consider when the trades occurred but also consider additional, external factors to define a popular collection, such as social media chatter like how Fazli et al. briefly discussed in their research [16].

I would also encourage a more efficient approach to address outliers, such as in the example of Cryptopunks. I would suggest an analysis of media chatter or any other metrics that could possibly give weight to investors desiring a collection for the sake of brand. My research worked extensively with the direct ownership of these NFT collections, but it would certainly be interesting to view the influence that 'NFT mania' and brand reputation has on a NFT collections price.

A final suggestion I have for more research is to shine a brighter light onto NFT collections that have been generally accepted as scams, such as FrostiesNFT. With the young history of NFTs, my research unfortunately did not have the luxury of working with a large amount of historical NFT data to accurately portray how different scams such as FrostiesNFT differ from their relatively priced peers. For FrostiesNFT specifically, my results show that its Gini was 0.6069 compared to its relatively priced peers' Gini of 0.4722. This one datapoint, however, is not large enough to draw any conclusion, but suggests future potential studies. The harsh truth is that there will more rug-pulls, money-grabs, and other forms of NFT scams that will continue to happen in the future and my hope is that other research will be able to analyze these collections in aggregate to find similarities that may be used to help consumers and businesses identify these scams before it is too late - as with more data comes more insight.

References

[1] "Non-fungible tokens (NFT)," https://ethereum.org/en/nft/, ethereum.org.

[2] A. Hamilton, "The beginning of nfts - A brief history of NFT art," https://www.zenofineart.com/blogs/news/the-beginning-of-nfts-a-brief-history-of-nft-art, Mar 2022, zeno Fine Art.

[3] Person and E. Howcroft, "NFT sales hit $25 billion in 2021, but growth shows signs of slowing," https://www.reuters.com/markets/europe/nft-sales-hit-25-billion-2021-growth-shows-signs-slowing-2022-01-10/, Jan 2022.

[4] "Bored Ape Yacht Club - collection," https://opensea.io/collection/boredapeyachtclub, openSea.

[5] M. Graham, "Meet apecoin: Bored ape yacht club's new crypto token," https://boardroom.tv/apecoin-bored-ape-yacht-club-token/, Mar 2022, boardroom.

[6] "Goldendao NFT," https://www.goldendao.xyz/, goldenDao NFT.

[7] M. Williamshailing, "Frosties NFT rug pull team arrested by DOJ," https://www.business2community.com/nft-news/frosties-nft-rug-pull-team-arrested-by-doj-02461717, Mar 2022, business 2 Community.

[8] Pearson, Jordan, "More than 80% of nfts created for free on OpenSea are fraud or spam, Company says," https://www.vice.com/en/article/wxdzb5/more-than-80-of-nfts-created-for-free-on-opensea-are-fraud-or-spam-company-says, Jan 2022, .VICE.

[9] "Introducing the NFT leaderboards," https://www.nansen.ai/post/introducing-the-nft-leaderboards, nansen.

[10] "Trait Sniper," https://www.traitsniper.com/features.

[11] L. Dvoskin, "Why community is the secret to NFT Success," https://www.rollingstone.com/culture-council/articles/community-secret-nft-success-1283244/, Jan 2022, rolling Stone.

[12] A. Popescu, "Non-fungible tokens (nft)-innovation beyond the craze," in *5th International Conference on Innovation in Business, Economics and Marketing Research*, 2021.

[13] K. B. Wilson, A. Karg, and H. Ghaderi, "Prospecting non-fungible tokens in the digital economy: Stakeholders and ecosystem, risk and opportunity," *Business Horizons*, 2021.

[14] O. Velthuis, "An interpretive approach to meanings of prices," *The Review of Austrian Economics*, vol. 17, no. 4, pp. 371–386, 2004.

[15] P. Kireyev, "Nft marketplace design and market intelligence," *SSRN Electronic Journal*, pp. 1–33, 2022.

[16] M. Fazli, A. Owfi, and M. R. Taesiri, "Under the skin of foundation nft auctions," *arXiv preprint arXiv:2109.12321*, 2021.

[17] B. M. Lucey, Y. Wang, and S. A. Vigne, "Bubbles all the way down? detecting and date-stamping bubble behaviour in defi and nft markets," *Detecting and Date-Stamping Bubble Behaviour in Defi and Nft Markets*.

[18] S. Casale-Brunet, P. Ribeca, P. Doyle, and M. Mattavelli, "Networks of ethereum non-fungible tokens: A graph-based analysis of the erc-721 ecosystem," in *2021 IEEE International Conference on Blockchain (Blockchain)*. IEEE, 2021, pp. 188–195.

[19] S. Papadopoulos, Y. Kompatsiaris, A. Vakali, and P. Spyridonos, “Community detection in social media,” *Data mining and knowledge discovery*, vol. 24, no. 3, pp. 515–554, 2012.

[20] M. A. Baytaş, A. Cappellaro, and Y. Fernaeus, “Stakeholders and value in the nft ecosystem: Towards a multi-disciplinary understanding of the nft phenomenon,” in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 2022, pp. 1–8.

[21] T.-S. Chang, K. Tone, and Q. Wei, “Ownership-specified network dea models,” *Annals of Operations Research*, vol. 214, no. 1, pp. 73–98, 2014.

[22] K. Vasan, M. Janosov, and A.-L. Barabási, “Quantifying nft-driven networks in crypto art,” *Scientific reports*, vol. 12, no. 1, pp. 1–11, 2022.

[23] Y. Chen and C. Bellavitis, “Blockchain disruption and decentralized finance: The rise of decentralized business models,” *Journal of Business Venturing Insights*, vol. 13, p. e00151, 2020.

[24] C. Pinto-Gutiérrez, S. Gaitán, D. Jaramillo, and S. Velasquez, “The nft hype: What draws attention to non-fungible tokens?” *Mathematics*, vol. 10, no. 3, p. 335, 2022.

[25] Woodyatt, Amy, “’bored ape’ nft worth $284,495 accidentally sells for just $2,844,” https://www.cnn.com/style/article/nft-accidentally-bored-ape-sold-intl-scli/index.html, Dec 2011, . CNN.

[26] Lamb, Evelyn, “Ask gini: How to measure inequality,” https://www.scientificamerican.com/article/ask-gini/, Nov 2012, scientific American.

[27] Nickolas, Steven, “What do correlation coefficients positive, negative, and zero mean?” https://www.investopedia.com/ask/answers/032515/what-does-it-mean-if-correlation-coefficient-positive-negative-or-zero.asp, Dec 2011, investopedia.

[28] A. Bhattacharyya and A. Chandra, “The cross-section of expected returns on penny stocks: Are low-hanging fruits not-so sweet?” *Available at SSRN 2386949*, 2016.

[29] S. Bhowmick, “Cryptopunks: Why NFT collectors love these punky little characters,” https://www.outlookindia.com/website/story/business-news-cryptopunk-i-why-nft-collectors-love-these-punky-little-characters/397061, Oct 2021, outlookindia.

[30] “API overview,” https://docs.opensea.io/reference/api-overview, openSea Developer Documentation.

[31] “What is Moralis?” https://docs.moralis.io/introduction, moralis.