

An analysis of 7,020,950 NFT transactions on the Ethereum blockchain

The Moonstream team
info@moonstream.to

October 21, 2021

Abstract

We present the Ethereum NFTs dataset, a representation of the activity on the Ethereum non-fungible token (NFT) market between April 1, 2021 and September 25, 2021, constructed purely from on-chain data.

This dataset consists of all 7,020,950 token mints and transfers across 727,102 accounts between block 12,150,245 and block 13,296,011.

We analyze this data with a view towards answering the following questions:

1. Who owns NFTs on the Ethereum blockchain?
2. Are NFTs for a small number of wealthy elite or are they for the masses?
3. How should one measure the utility of a non-fungible token?

The distribution of the number of NFTs owned by Ethereum addresses is Zipfian in nature. This indicates that the NFT market is indeed an open market, free for anyone to participate in and with low barriers to entry.

Most NFTs have few owners relative to the number of tokens in their token supply.

The probability distribution that a contract induces on owners of its tokens reflects the utility of that contract. Based on this observation, we propose an entropy-based measure of utility for NFT contracts – their *ownership entropy*.

1 Introduction

A non-fungible token, or NFT, is a unique and non-interchangeable unit of data stored on a digital ledger (blockchain)[1]. It is non-fungible in the sense that it immutably and exclusively represents a data asset. This means that it can be used to represent virtually anything – ideas, physical objects, intellectual property, etc. It will never change what it represents and the represented object or concept admits no *other* representation on the blockchain.

The global market for NFTs has seen a massive boom between June and October of 2021. Visual artists and other content creators are digitizing their

creations as NFTs to distribute their work to patrons. Game producers are tokenizing assets in computer and mobile games as NFTs to create shared worlds with other content creators.

Conventional digital representations of physical works can be replicated arbitrarily and indefinitely. For example, if you own a digital copy of a book, you could in principle make arbitrarily many clones of that copy and distribute it to anyone who asked for it.

In contrast, NFTs naturally reflect the scarcity of the objects they represent. This essential scarcity makes NFTs a perfect tool to globally conserve value when transferring ideas and assets from one digital reality to another. NFTs allow people to create common representations of scarce resources across multiple realities.

The recent boom in the NFT market [2] has led to an increased variance in utility across NFTs. Similarly, there is a growing number of first-time NFT buyers. This paper analyzes these variances and derives statistics that can be used to classify NFTs and NFT purchasers.

2 The Ethereum NFTs Dataset

The majority of recent NFT action has been centered around the Ethereum blockchain. This made Ethereum the natural starting point for a series of analyses, of which this is paper represents the first.

This section introduces the Ethereum NFTs dataset [3]. It describes the structure of NFT data on the Ethereum blockchain, elaborates on our methods of data collection, and articulates the structure of the dataset.

2.1 Contracts, tokens, events, and ERC721

On the Ethereum blockchain, non-fungible tokens are created using Ethereum smart contracts [4]. The most famous examples of Ethereum NFT contracts are CryptoPunks [5] and CryptoKitties [6].

Ethereum smart contracts expose public methods which may be executed by participants on the Ethereum blockchain. When a smart contract method is executed, it can change the state of the Ethereum blockchain by changing the state of addresses associated with the smart contract and also by emitting events which log the activity of that smart contract.

An NFT contract typically represents a plurality of non-fungible tokens. NFT contracts normally represent a thematically consistent set of tokens. This is reflected by the language of NFT marketplaces like OpenSea [7] and Nifty Gateway [8], which call them *collections*.

Anyone wishing to create a non-fungible token on the Ethereum blockchain is free to implement their tokens in any manner whatsoever subject to the requirement of non-fungibility. The most *common* implementation follows the Ethereum ERC721 protocol [9].

There is a growing ecosystem of secondary marketplaces such as OpenSea *opensea* that are immediately accessible to NFTs which follow the ERC721 standard. Tokens which do not follow the standard, such as CryptoPunks [5], have to create ERC721 wrappers in order to participate in these marketplaces. This, combined with the existence of high quality, battle-tested implementations of the ERC721 standard, means that ERC721-compliant tokens account for the overwhelming majority of Ethereum NFTs.

ERC721 contracts emit a *Transfer* event into global Ethereum state whenever a token is created on that contract or an existing token on that contract is transferred from one address to another. This event has the following structure:

Listing 1: ERC721 Transfer event

```
event Transfer(address indexed _from, address indexed _to, uint256 indexed
_tokenId);
```

We built the Ethereum NFTs dataset by scanning all blocks between block number 12,150,245 and 13,296,011. Our scan of these 1,145,767 blocks yielded transfer activity for 7,020,950 tokens from 9,292 NFT contracts across 727,102 addresses. These mints and transfers form the core of the dataset.

The Ethereum NFTs dataset is built purely from on-chain Ethereum data, collected by Moonstream [10].

The Ethereum NFTs dataset is a single SQLite database consisting of two core relations – *mints* and *transfers*. The *mints* relation enumerates all NFT mints (creations) that took place in the window of time for which we collected data. The *transfers* relation enumerates all transfers of NFTs from one owner to another which took place in the same window of time.

Both relations have the same schema. This is the creation statement for the *mints* table.

Listing 2: *mints* relation

```
CREATE TABLE mints
(
    event_id TEXT NOT NULL UNIQUE ON CONFLICT FAIL,
    transaction_hash TEXT,
    block_number INTEGER,
    nft_address TEXT REFERENCES nfts(address),
    token_id TEXT,
    from_address TEXT,
    to_address TEXT,
    transaction_value INTEGER,
    timestamp INTEGER
);
```

The *from_address* and *to_address* fields denote the address of the original owner and the address of the new owner for each transfer. They do *not* denote

the *from* and *to* parameters of the actual Ethereum transaction that initiated the mint or transfer.¹

The dataset also contains other relations, derived from these core relations, but included in the dataset for ease of analysis. These are:

1. *nfts* – available metadata about the NFT contracts represented in the dataset
2. *current_market_values* – the current (estimated) market value of each NFT in WEI
3. *current_owners* – the current owner of each NFT
4. *transfer_statistics_by_address* – number of transfers in and out of every address that was involved in an NFT transfer in the window of time the dataset represents

Throughout this paper, we take the world *current* to mean “as of block 13,296,011”.

2.2 Caveats

The Ethereum NFTs dataset is constructed purely from events on the Ethereum blockchain. It does not include any data from Layer 2 networks like Polygon. Nor does it include any data from centralized APIs like the OpenSea API. It does not account for events or data from any non-ERC721 smart contracts associated with these platforms on the Ethereum blockchain.

This means that two parties could exchange a positive amount of funds for a transfer off-chain and conduct the transfer on-chain and we would not be able to distinguish the transfer from a gift.

It is also possible for a single transaction to involve multiple NFT transfers. [11] is an example of such a transaction involving several NFT layers on top of a Loot token [12].

If a transaction involves multiple NFT transfers and has a non-zero value, it is difficult to understand whether that value is related to the transfers and, if so, how it distributes over the transfers.

For that reason, in this first version of our dataset, the valuation numbers should only be treated as a rough estimate of the actual value for each NFT.

2.3 Access

The complete Ethereum NFTs dataset is available on Kaggle under a Creative Commons license [3].

¹The *to* parameter would be the address of the NFT contract, which we list under *nft_address*. We plan to include the transaction *from* information in a future iteration of the dataset.

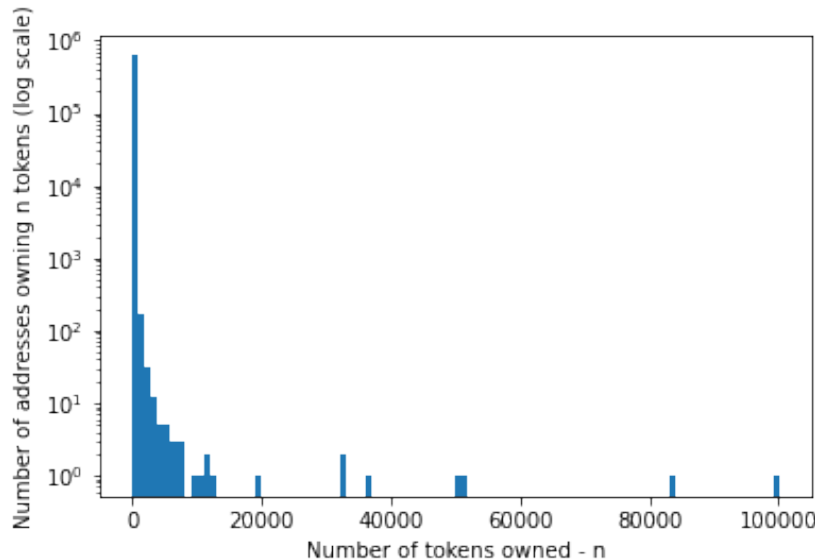


Figure 1: Addresses by number of tokens owned

3 Who is buying NFTs?

It is difficult to understand how much of the hype surrounding NFTs is manufactured and how much of it reflects the situation on the market.

Is there a small number of people who each carry significant NFT holdings? Are they driving the hype and carrying the market with them?

Or is the market home to many different people, each of whom own a relatively small number of NFTs?

For each $n > 0$, let A_n denote the number of addresses that assumed ownership of exactly n NFTs between block 12,150,245 (April 1, 2021) and block 13,296,011 (September 25, 2021).

Figure 1 plots A_n against n on a logarithmic scale.

Of course, the NFT owners in the full dataset includes the addresses of smart contracts which act as exchanges and clearinghouses for NFTs and work with thousands and even tens of thousands of NFTs at a time. It also includes the addresses of bots which may not be implemented as smart contracts but which automatically submit transactions based on their triggering logic.

Figure 2 considers only those addresses which assumed ownership of at most 1500 NFTs between block 12,150,245 (April 1, 2021) and block 13,296,011 (September 25, 2021).

Even this graph is better viewed on a logarithmic scale, as in Figure 3.

These statistics suggest one additional hypothesis - that the distribution of the number of NFTs per owner follows a Zipf distribution.

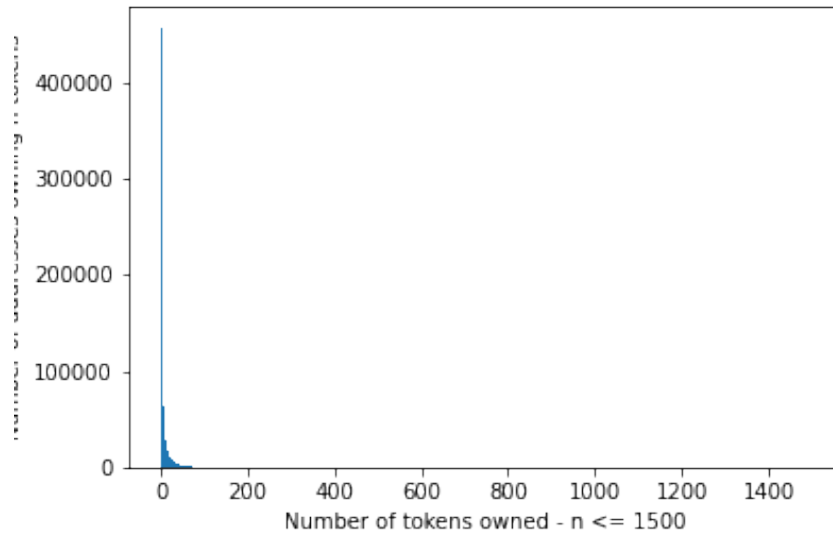


Figure 2: Addresses with 1500 tokens or fewer by number of tokens owned

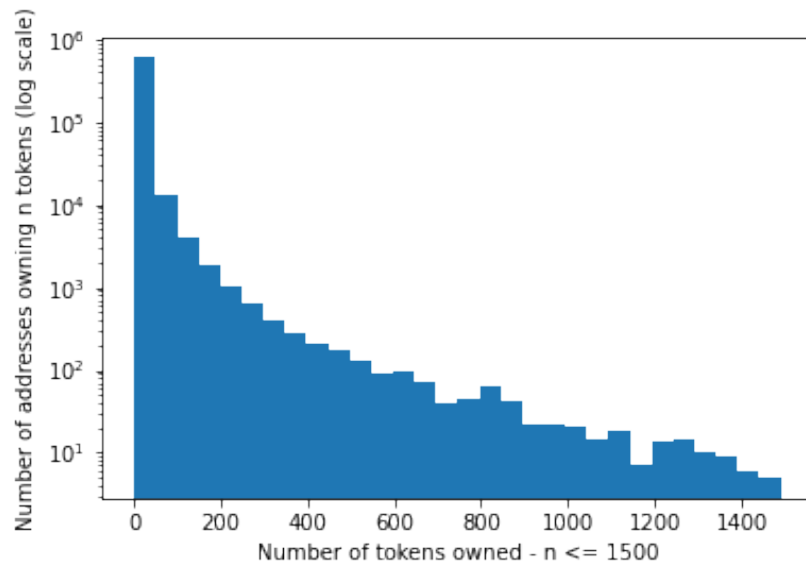


Figure 3: Addresses with 1500 tokens or fewer by number of tokens owned (log scale)

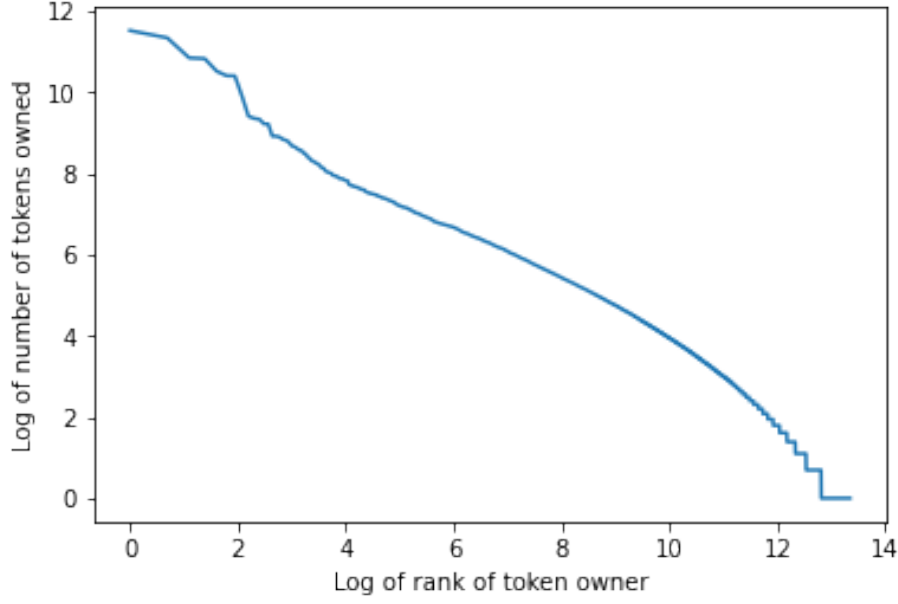


Figure 4: Log of number of tokens owned vs. log of number of addresses owning that number of tokens

NFTs owned n	Number of addresses	Proportion of addresses	Total tokens owned by addresses	Proportion of tokens owned by addresses
$n \geq 1$	625,354	1	7,020,950	1
$1 \leq n \leq 1000$	625,107	0.9996	6,112,780	0.8707
$1 \leq n \leq 100$	615,658	0.9845	4,036,089	0.5749
$1 \leq n \leq 10$	520,834	0.8329	1,335,177	0.1902
$1 \leq n \leq 5$	456,399	0.7298	842,892	0.1201
$1 \leq n \leq 2$	348,948	0.558	438,090	0.0624
$n = 1$	259,806	0.4155	259,806	0.037

Table 1: Number of addresses by NFTs owned

This hypothesis is clearly supported [13] by Figure 4, which plots the log-log graph of each number n of tokens that an address could own relative to the number addresses owning that number of tokens.

Table 1 quantifies this relationship further. It shows, for example, that 83.29% of the addresses which assumed ownership of an NFT between block 12,150,245 (April 1, 2021) and block 13,296,011 (September 25, 2021) did so for only a handful of tokens n , where $1 \leq n \leq 10$.

It is possible that there is a small number of people or organizations who are creating a distinct wallet for each NFT they purchase, but doing so at a scale that would our analysis would be technologically and operationally complex enough, and expensive enough, that it is virtually impossible.

What this data shows us is that the Ethereum NFT market is open in the sense the vast majority of its participants are small-time purchasers who likely make their purchases manually. There are few barriers to entry for those who wish to participate in this market.

There is also a great inequality in the Ethereum NFT market in the sense that the top 16.71% of NFT owners control 80.98% of the NFTs. This latter statistic does require a little more nuance in its interpretation, however, as many of those owners are marketplaces and clearinghouses like OpenSea [7], Nifty Gateway [8], and other platforms of the same ilk. We plan to expand on this analysis in a future report.

4 The utility of an NFT

People buy NFTs for different reasons. Some buyers may purchase an NFT to support their favorite artists or communities. Others may prefer to purchase NFTs that bring them extrinsic utility. A good example of this kind of utility is the Ethereum Name Service (ENS) [14], which allows anyone to create a human-friendly name (such as `vitalik.eth`) associated with the Ethereum addresses (and more). The associations are represented as NFTs on the ENS registry contract, and many services (e.g. Coinbase, Metamask, etc.) support resolution of ENS names as part of transfers and other blockchain operations.

The Ethereum blockchain is simultaneously home to NFTs like the Ethereum Name Service and governance NFTs for various decentralized protocols, and NFTs that only have subjective value like those containing links to digital art and other forms of media.

How do we distinguish between NFTs that represent intrinsic, subjective value and those which have clear extrinsic utility?

This is a question that the blockchain community and blockchain regulators have been concerned with for several years. Until now, there has been no objective, quantitative measure of the level of extrinsic utility of a token.

We propose that the extrinsic utility of a token should be measured at the level of its NFT collection (smart contract). For tokens that do represent extrinsic utility for a large pool of users, other tokens deployed as part of the same contract are must also represent similar utility. Market forces demand this, as exaggerated scarcity diminishes utility [15].

Let us consider a few different possible statistics that could act as measures of extrinsic utility for NFT contracts. We will start by considering statistics that perform poorly as measures of utility and address their faults to arrive at good candidates.

4.1 Maximum token value

We could attempt to use the maximum value of a token in an NFT contract as a measure of its utility. However, this statistic does not capture distributional information about other tokens in the same contract. We would not be able to use this statistic to understand if all the tokens in the same contract seemed to have similar utility. This makes it a poor statistic for the purposes of measuring external utility.

4.2 Distribution of value over the tokens in a collection

This statistic has an advantage over the maximum token value in that it encodes information about all the tokens in a collection.

It suffers from two problems:

1. It is not a scalar statistic. We would need to calculate several moments of the token value distribution over the collection in order to capture all the information it contains, and this could make it awkward to work with.
2. It requires us to estimate the value of the tokens in a contract. The estimation of value from on-chain data is difficult because people are not required to exchange monetary value on the blockchain. It would be a simple matter for two parties to exchange money off-chain and then exchange their NFTs on chain.

This second problem is a practically insurmountable obstacle to the use of any statistic based on the distribution of values over the tokens in an NFT contract.

4.3 Distribution of number of transfers over the tokens in a collection

This statistic has an advantage over the previous statistics in that it doesn't require us to estimate the value of NFTs in a collection.

Like the previous candidate, this too is not a scalar statistic. It would require a great deal of care in analysis in interpretation to use this distribution of transfers to draw conclusions about the utility of an NFT contract from this distribution.

A more serious concern is that both the value of tokens and the number of times they are transferred is dependent on the particular form of their extrinsic utility. One can imagine use cases in which tokens derive utility through being transferred or through being volatile in value, and other use cases in which tokens derive utility through being held or through being stable in value.

Because the *form* of utility could have such a drastic effect on this candidate and the previous one, neither is an ideal candidate for a measure of utility. Our measure of utility should be independent of the form of the utility. We cannot predict how people will derive utility from their NFTs in the future, but we would like to be aware of when they start deriving it.

4.4 Distribution of ownership over the tokens in a collection

Now we narrow in on an invariant. We discussed why it is sensible to associate extrinsic utility with an NFT contract - a full collection of NFTs - rather than with the individual tokens. Because extrinsic utility applies to large populations

of users rather than to a small number of individuals, and excessive scarcity diminishes utility.

This means that, if an NFT collection has extrinsic utility, then it should have many distinct owners relative to its number of tokens.

Suppose that a few parties strike out to purchase most of the tokens. Then the tokens would gain monetary value, and would become good vehicles for investment. But this represents a gain in intrinsic utility and a reduction in extrinsic utility. So we see that the dynamics whereby extrinsic utility is traded off for intrinsic utility correspond to an increased concentration of ownership among a few addresses as compared to a dispersion of ownership across many addresses.

The level of dispersion of ownership across the tokens in an NFT contract is invariant to the particular form that the external utility of the tokens takes. If the form involves many transfers, for example, it still doesn't significantly affect the dispersion at any single point in time.

This notion of dispersion of ownership is an invariant of the NFT contract under different forms of extrinsic utility which nonetheless captures how attractive NFTs in that contract are to the general Ethereum community.

The notion of information theoretic entropy formalizes this concept of dispersion. We propose a statistic called *ownership entropy* as a measure of the external utility of the tokens in an NFT contract.

4.5 Ownership entropy

Let π be a probability distribution on the sample space $1, 2, \dots, n$ for $n \geq 1$. Denote $\pi = (\pi_1, \pi_2, \dots, \pi_n)$, where π_j is the probability associated with the event j .

Then the entropy of π is defined as:

$$H(\pi) = \sum_{j=1}^n -\pi_j \log(\pi_j).$$

(Here, \log represents the logarithm for base 2, although the constant is not so important.)

The entropy $H(\pi)$ is maximized for the distribution π which assigns equal probability to all its outcomes. In fact, from Jensen's inequality,

$$H(\pi_1, \dots, \pi_n) \leq \log(n),$$

with the maximum achieved if and only if $\pi_1 = \pi_2 = \dots = \pi_n = \frac{1}{n}$.

$H(\pi)$ is an information theoretic measurement of how well distributed the probability mass of π is over its sample set, and is maximized when the probability mass is evenly distributed. The units of entropy are bits (as in binary digits).

This makes it a natural candidate to measure the dispersion of ownership over the tokens of an NFT contract.

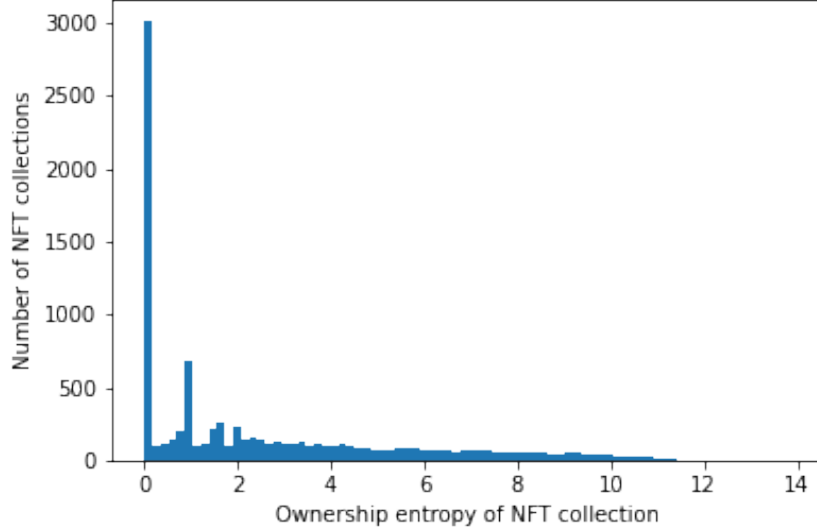


Figure 5: NFT contracts by ownership entropy

For an NFT contract C , let T denote the set of tokens (represented by their token IDs) present in C . For each token $t \in T$, let A_t denote the address owning that token. It is possible for A_t to be any Ethereum address, including the 0 address.

We can think of C as a probability distribution over its tokens whereby we select each token t with probability $\frac{1}{|T|}$. This induces a probability distribution π_C on the set \mathcal{A} of all Ethereum addresses whereby, for any address $A \in \mathcal{A}$, the probability of π_C selecting that address is:

$$\pi_{C,A} = \frac{|t \in T : A_t = A|}{|T|}.$$

We define the ownership entropy of C to be the entropy $H(\pi_C)$ of this probability distribution that C induces on \mathcal{A} .

Figure 5 plots the number of contracts on the Ethereum blockchain that saw mint or transfer activity between block number 12,150,245 and 13,296,011, by ownership entropy.

The tokens with highest ownership entropy are those that have clear utility, such as the Ethereum Name Service [14] and the Uniswap v3 Position token [16]. The tokens at the bottom are NFT releases of art collections which have failed to gain any traction.

Table 2 shows a sample of token collections at varying levels of ownership entropy, with a link to each project as well as the number of tokens which were active in that collection between between block 12,150,245 (April 1, 2021) and

NFT Collection	Ownership entropy	Level	Tokens traded	Owners registered
Ethereum Name Service	13.864019	High utility	145,303	64,717
Rarible Governance Token	13.831032	High utility	86,375	42,644
Uniswap v3 Position NFT	13.742724	High utility	132,087	38,790
The Enigma Project	6.011324	Highly speculative	381	108
The RTFKT Capsule Space Drip	6.010405	Highly speculative	112	78
Dommies	6.009679	Highly speculative	1,034	236
World Cup Token	2.004886	Stalled	11	5
SHADYCON (associated with Eminem)	2.002227	Stalled	101	23

Table 2: A sample of tokens at varying levels of ownership spread

block 13,296,011 (September 25, 2021), and the number of distinct addresses which currently² hold ownership of these tokens.

The cases of Ethereum Name Service and the Uniswap position NFT are particularly interesting because holders realize the value of those NFTs in very different ways - an address is much more likely to hold onto an ENS token and much more likely to trade their liquidity position on Uniswap v3. Despite the differences between these two contracts in the form of their utility, they both rise to the top when we consider their ownership entropies.

We also see that, at the lower ranges, the ownership entropy serves to measure adoption for tokens which have no extrinsic utility (like the RTFKT Capsule Space Drop and Dommies versus World Cup Token).

This data also highlights the importance of considering ownership entropy as a time series, and of tracking the differences in ownership entropy over time to reflect changing market perceptions of NFTs. There is a big difference between World Cup Token, which launched years ago and SHADYCON which only launched recently. We can only draw a true comparison between these tokens after both have been allowed time to achieve stable dynamics on the blockchain.

5 Conclusions

Our analysis paints a picture of the Ethereum NFT market as an open and free market which exhibits the same kinds of wealth disparities as conventional markets.

It also provides early validation for the viability of ownership entropy as a means of quantifying the extrinsic utility of NFT contracts.

In future versions of this report, we plan to:

1. Conduct analysis of the openness of the Ethereum market over time. Rather than only considering data in a single window of time (in the case of this report, April 1, 2021 to September 25, 2021), we will consider the time series of the same statistics generated at frequent intervals from 2016 until the time of publication of the report.
2. Expand the analysis of ownership entropy into an analysis of ownership information gain - the change in ownership entropy over time.

²As of block 13,296,011.

3. Enrich our dataset and our analyses with information about the addresses which funded NFT transfers.
4. Enrich our dataset and our analyses with side information about NFT valuations from centralized sources (like the OpenSea API).
5. Provide further validation for ownership entropy as a measure of extrinsic utility, not only of ERC721 tokens, but also of ERC20 tokens.

6 Collaboration

The calculations presented here, as well as a more elaborate analysis, are available as a Kaggle notebook at:

<https://www.kaggle.com/simiotic/ethereum-nft-analysis>

The Moonstream platform, which we used to gather this data, is free software released under the Apache 2.0 License [17]:

<https://github.com/bugout-dev/moonstream>

The Ethereum NFTs dataset is available on Kaggle under a Creative Commons license (CC-BY-4.0) [18]:

<https://www.kaggle.com/simiotic/ethereum-nfts>

You can reach the Moonstream team at any time on Discord:

<https://discord.gg/K56VNUQGvA>

References

- [1] Wikipedia.org. Non-fungible token.
- [2] Elizabeth Howcroft. NFT sales surge as speculators pile in, sceptics see bubble. Reuters. August 25, 2021.
- [3] Sophia Arakelyan, Andrey Dolgolev, Neeraj Kashyap, Nana Landau, Daria Navoloshnikova, Tim Pechersky, Yhtyyar Sahatov, Sergei Sumarokov. Ethereum NFTs. Kaggle, 2021. (doi:10.34740/KAGGLE/DSV/2698517)
- [4] Ethereum.org. Introduction to Smart Contracts.
- [5] CryptoPunks. <https://www.larvalabs.com/cryptopunks>
- [6] CryptoKitties. <https://www.cryptokitties.co>
- [7] OpenSea. <https://opensea.io>
- [8] Nifty Gateway. <https://niftygateway.com>

- [9] William Entriken, Dieter Shirley, Jacob Evans, Nastassia Sachs. EIP-721: A Non-Fungible Token Standard. January 24, 2018.
- [10] Moonstream. <https://moonstream.to>
- [11] Transaction `0xa578aac6db19c464f69492747fa147985006281a57d77e46316fe09fb406deb2`. Etherscan.io
- [12] Loot Project. <https://www.lootproject.com/>
- [13] David M W Powers. Applications and explanations of Zipf’s law. Joint conference on new methods in language processing and computational natural language learning, 1998. Association for Computing Linguistics. pp. 151-160.
- [14] Ethereum Name Service. <https://ens.domains>
- [15] Karl Menger. On the Origin of Money. The Economic Journal Vol. 2, No. 6, June 1892.
- [16] The Uniswap Team. Introducing Uniswap V3
- [17] Apache License, Version 2.0
- [18] Creative Commons Attribution 4.0 International (CC BY 4.0).