Analysis of NFT markets and traders network

Aaditya S Goel
PES1UG20CS001,Dept. of CSE
PES University
Bengaluru, India

Swathi Rupali N.V
PES1UG20CS600,Dept. of CSE
PES University
Bengaluru, India

Aditya Anil Rokhade
PES1UG20CS018,Dept. of CSE
PES University
Bengaluru, India

Ankur D
PES1UG20CS537,Dept. of CSE
PES University
Bengaluru, India

***Abstract*— NFT or Non-fungible Tokens are blockchain linked digital assets which are tradeable in decentralised markets. NFT stands for assets which are unique and cannot be cloned. Implying that no two NFTs are the same. NFTs have made headlines both by exclusivity when a digital-art was sold for 69.3 million USD by people and by critics claiming that it is just digital and inanimate to be valued. In this paper we study and analyse how NFT market is different from traditionally operating stocks shares market. Estimate how much NFT is actually decentralised by means of distribution and volume of sales. The obscurity and seemingly high technological barrier of entry into the NFT market seems to keep traditional economists and financial experts from studying the structure and evolution of this market.Realise if we could use NFT as an asset class like gold statistically**

I. Introduction

NFTs are now being used to commodify digital objects in different contexts, such as art, gaming, and sports collectibles. Originally NFTs were part of the Ethereum blockchain but increasingly more blockchains have implemented their own versions of NFTs. The profitability of NFTs has motivated celebrities to create their own NFTs, with collectibles of NBA and famous football players getting sold for hundreds of thousands dollars. The first popular example of NFTs is CryptoKitties, a collection of artistic images representing virtual cats that are used in a game on Ethereum that allows players to purchase, collect, breed, and sell them on Ethereum5 . In July 2020, the NFT market started to grow and attracted a huge attention in March 2021, when the artist known as Beeple sold an NFT of his work for $69.3 million at Christie's 8 . The purchase resulted in the third-highest auction price achieved for a living artist, after Jef Koons and David Hockney.

Research on NFTs is still limited, and focuses mostly on technical aspects, such as copyright regulations, components, protocols, standards, and desired properties; new blockchain-based protocols to trace physical goods. Emperical studies that aim at characterizing properties of the market have focused on a limited number of NFT collections, such as, CryptoKitties, Cryptopunks, and Axie, or on a single NFT market, such as, Decentraland or SuperRare. These analyses revealed that the digital abundance of NFTs in digital games has led to a substantial decrease of their value .Further, it was shown that NFTs valued by experts are more successful, and that, based on 16,000 NFTs sold in the SuperRare market, the structure of the the NFT co-ownership network is highly centralized, and small-world-like.

Its 'digital smart contract' and 'mutually not interchangeable' characteristics are what makes it different from other crypto assets.

The lockdown measures taken in the pandemic significantly boosted digital engagement and further stimulated the growth of NFT. In addition to gaining widespread attention from investors and enthusiasts, academic research around NFT also gradually emerged. On one hand, NFT is still in the early stages of development, with both great potential and uncertainty; this also means that there is not enough material or data for research to support a large number of studies. On the other hand, it is also because NFT is a cross-cutting area of research. It requires scholars to have broader knowledge accumulation and understanding in multiple areas.

The aim of this study is to reflect on the latest trends, development, growth and profitability of NFTs market on the basis of associated variables. Since the NFTs market bloomed in a way that very few people anticipated since the beginning of the last year, the study also aims to foresee the scope of NFTs in the near future. The aim of our project is to have a comparitative study for NFT's before covid and during covid.

II. Literature Reviews

The work of Matthieu Nadini et al. [**?**] serves as our primary reference. Given the relative infancy of the NFT market, and the seemingly high technological barrier of entry, its overall structure and evolution has not been analyzed as intricately as traditional markets.This paper intends to do the same. It starts off with statistical properties of the market such as the total volume of NFTs traded over a period, their distribution in terms of collections, the frequency of trades of certain collections, the distribution of prices etc. This is followed up with an analysis of the network of traders in NFT markets in the form of graphs. This in particular can hold much importance in understanding how traders in NFT markets behave differently from traders in say equity markets. The writers of the paper follow this up with a model to predict NFT prices and use that to extract good predictors.

Since mid-2020, the NFT market has seen an increase in the role of investors, traders, and artists, to provide a wider view of NFT marketplace.The author has visualized it by analyzing the total profit /loss incurred by the 5 major sectors. The 5 sectors include Art, Collectibles, Gaming, Metaverse and Utilities. It is seen that the difference between total profits and the total losses is very large, making NFTs market a volatile yet profitable investment portfolio. It is also found that when it comes to NFT sales, Ethereum is still by far the most important blockchain. According to their sales transaction analysis of top NFT's traded in Ethereum, they have proved a negative correlation between number of transactions and sales. NFTs are safeguarded by complex codes that are impossible to alter, making them a more secure way to buy and sell digital properties. Every coin, however, has two sides. Unregulated NFTs and the craze surrounding them could cause volatility and they aren't readily exchangeable for cash, so liquidity is an issue.

Christian Pinto-Gutiérrez proves through their results, how the remarkable increase in prices of major cryptocurrencies can explain the hype around NFT's. Using vector autoregressive models(VAR) they show that how Bitcoin returns significantly predict next week's NFT's growth in popularity, measured by google search queries. Wavelet coherence analysis also suggests that bitcoin and Ethereum returns are significant drivers of the attention received by NFT's. It is also found that the excitement around cryptocurrencies induced by record-high prices in 2021 could explain the NFT growth in popularity during the same period. Thus, we should extend our understanding of the effects of the leading cryptocurrencies on new blockchain developments, such as the NFT market.

Non-Fungible Tokens (NFT). The Analysis of Risk and Return.This paper revolves around the risk and reward analysis of investment into NFT projects spread across NFT gaming, NFT decentralized finance, NFT music and media. It starts with exploration of liquidity mining, which translates to creating NFT liquidity by depositing and generating revenue with time as the floor price increases. It then explores data gathered about NFT projects such as Theta, PancakeSwap Decentraland from Binance. On analysis, we figure that the average first day returns is 130% excluding outliers contrasted to the IPO-day returns of 35% Further it dwells upon risk analysis in two levels, for raw and adjusted returns. The first being sharpe ratio-which explains returns contrasted with the votality. Surprisingly, the sharpe ratios analysed for the NFT projects would equal stock-market sharpe ratio of 0.32. Secondly, build risk-adjusted models on comparison with bitcoin, yielding 60% returns. This validates that investments into NFT would perform better than BTC over a long term(¿5y).

III. EDA

Before we start visualizing our data and try to make inferences, we do some data cleaning to lose some unimportant attributes and to make the data easier to work with. First, we drop all the columns that hold images, i.e. 'image url1', 'image url2', 'image url3'. We do not believe that these fields are critical to our analysis. In the paper by Matthieu Nadini et al, they vectorize the images and use the distance between them as one of the features for their linear regression predictor model. We don't intend to replicate that segment of their work and hence choose to drop them.

After further exploration, we notice that 'Seller username' and 'Buyer username' have a lot of missing values(4.97 million and 5.17 million respectively out of 6 million data points). At first glance, this seems worrisome and feels like we could potentially be losing out on a lot of useful information. But it must be noted that there are no missing values in 'Seller address' and 'Buyer address', both of which are encoded hex values that store the digital addresses of buyers and sellers. These are used to create trader and NFT networks which too are used as features for the Regression model. We believe this is important information and thus choose to retain them.

The next important step we take in our analysis is the splitting of the data into two parts, pre-COVID19 and post-COVID. The pandemic served as a catalyst in the growth and evolution of the NFT market and thus its effect cannot be ignored. As we will show through our inferences, the changes NFTs underwent during the pandemic define how they are viewed by the general public today.

A. EDA inferences

In our dataset we see that Atomic is the most popular platform for NFT exchange by a mile. OpenSea comes in at second and Cryptokitties comes in at third place. But as we understand, our dataset seems to reveal that our data is slightly dated and doesn't give the full picture about the current state of affairs in the NFT market.

Therefore, as mentioned in the previous section, we partition the data into two segments. The first one stretches from 23rd November 2017 to 25th July 2020 and contains exactly 2 million transactions. The choice of 25th July was not specific, but we wished to choose a date during which most of the world had just begun the first lockdown stint, forcing people to move to digital media of communication, entertainment. This is about the time when NFTs were being noticed.

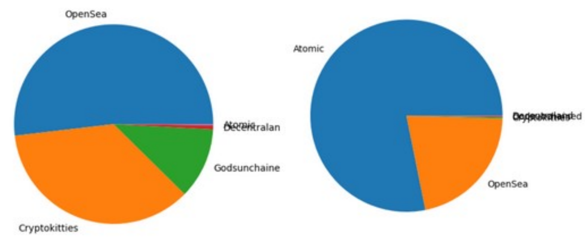

Fig. 1. Market share of platform

The second split contains transactions between 25th July 2020 to 27th April 2021. This split has over 4 million transactions. This number is staggering and puts into picture the magnitude at which the NFT market grew during this time. Most of the advanced economies of the world were under lockdown throughout this 10 month period and fueled the rapid expansion of NFTs. The fact that twice as many transactions happened during this 10 month period, compared to the previous 35 months stands as testimony to the rise of NFTs as digital assets.

1) Market: We start off by analyzing the most popular platforms for buyers and sellers to conduct their NFT transactions. Before COVID19 the most popular market place was OpenSea with over 50% share followed by Cryptokitties with around a 32% share, then followed by Godsunchained, Decentraland and Atomic. But after COVID19 the largest marketplace is Atomic by an overwhelming margin(over 75%) followed by OpenSea whose market share reduced to less than 25%. But owing to the large increase in the number of transactions overall, it still has a sizeable number of transactions taking place on its platform. The biggest surprise however is Cryptokitties. Cryptokitties went from being one of the first and biggest NFT projects to an insignificant speck amongst its contemporaries. Cryptokitties was the project that propelled the rise of the now ubiquitous Ethereum blockchain. Experts when talking about the importance of

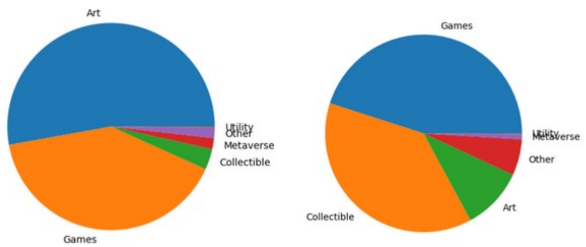

Fig. 2. Most popular categories

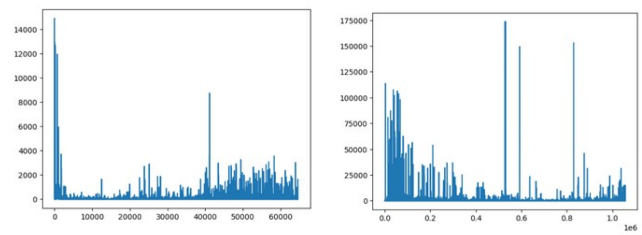

Fig. 3. Prices of different categories

Cryptokitties state that before it became popular, blockchains were associated primarily only with cryptocurrencies. Cryptokitties played a crucial role in showcasing what was possible with NFTs.

2) Category: After analyzing the market we now move on to analyzing the different categories of NFTs. Pre-COVID NFTs pertaining to Art and Games dominate all transactions. Utilitarian NFTs refer to those NFTs that function for purposes like tickets and so on. Post pandemic Games and Collectibles are the most popular categories of NFTs. The sudden rise of collectibles is important to us because it is representative of what NFTs represent to us in general. They are seen as digital assets much like Bitcoins valued for their exclusivity which drives their prices sky-high. But they differ from cryptocurrencies in that all bitcoins are the same in essence and value, but each NFTs is non-duplicable, unique on the blockchain and in its value. Although NFTs from the same collection are seemingly priced similarly, their value lies solely in the minds of buyers. This is a key differentiator between NFTs and other potential digital assets.

3) Prices: A descriptive summary of the prices before and after COVID reveals interesting results. Before the pandemic, the average price of the transactions was \$34.22 with a standard deviation of \$566 and a maximum of \$214k. After the pandemic, the average price shoots up to \$201.41, a staggering standard deviation of \$6,689.60 and a maximum of \$7.5 million. The difference before and after the pandemic are stratospheric to say the least. Further, Art and Collectible categories of NFTs have similar measures of central tendency while the NFTs associated with games are valued much lower. Thus although NFTs in games have persisted through the NFT revolution, they don't seem to have caught up with other categories in terms of monetary value.

IV. Price Prediction

Now after analysis, we will now attempt to predict the price of the NFT transactions as accurately as possible. An important part of the predictive process is to first identify what features will be good predictors. The reduction in the number of inputs not only reduces the computational cost of the model but also helps improve performance. We are also choosing to predict prices only for the Godsunchained collection as it makes little sense to try and predict the prices of an aggregated set of different NFTs. In the primary paper of reference, for the prediction of stock prices, the researchers use elaborate feature engineering to obtain 11 features. The model they employ for all three sets is multi-variate linear regression. The 11 features are segregated into three different sets.

The third one is important. The paper notes that the previous day's sale price is a strong indicator of the present day's price and thus uses an average of the previous week's prices. This is an approach we found interesting.

Instead of using an average of the prices we use the prices themselves as features into our Linear Regression model. We believe the nuances of price variation over multiple days and transactions is better explained this way than by averaging them where we'd lose these details.

Another feature we notice they don't include is the 'crypto_price' which shows the price of the NFT in ETH cryptocurrency. This is a very interesting property. Cryptocurrencies are digital assets themselves, and through their rise and fall they can potentially capture more features about the NFTs than regular currencies. This is further reinforced by our correlation test that showed that there was little similarity between the price of the

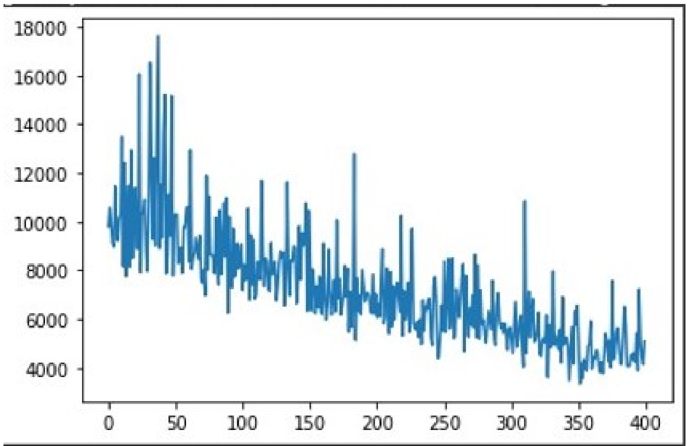

Fig. 4. ANN Model Training Loss

dollar and the price of ETH. With all these features at our disposal we conduct Multi-variate Linear Regression on the given data with the price_usd as the predicted value. We obtain an R_square score of 0.59 and an RMSE score of 106. This is not a great score by most standards, but given the relative simplicity of our features compared to the ones mentioned in the paper, our fit is relatively on-par. The initial paper achieves around an R_square score of 0.77 With more complex feature engineering, we believe that our approach might deliver better performance than the existing model. But given time and resource constraints we are not able to showcase this at the moment. We however intend to work on them soon.

One point we do notice in the paper is that Linear regression is the only model used. This gives them fairly decent results but we believe there's a lot of scope for improvement with the use of more power models, statistical and even possible Deep Learning models. For example, we used the popular ARIMAX model with the same features on the same test set and obtained an RMSE score of 92.92. Our ANN model did not perform well. The training accuracy converged to a low value but testing accuracy remained fairly high. We are trying to better this, and possibly explore the use of LSTMs.

V. Contributions and Peer review Feedback

A. Individual Contributions

Study conception: Ankur Datta ,Aaditya Rokhade and Aaditya Goel

Design: Swathi Rupali and Aaditya Goel

Preprocessing and EDA: Ankur Datta, Aaditya Goel

Model Development: Aaditya Rokhade, Swathi Rupali

All authors reviewed the results and approved the final version of the manuscript.

B. Peer-review Feedback

We were advised to use a pretrained model to predict our results. But we chose to try ones that we had learnt in class and focus on feature selection similar to the paper. But we wish to explore complex models soon.

VI. Conclusion

This project has been a lot more elaborate than we initially imagined. The study of their process has exposed us to very interesting feature engineering methods involving the use of graphs centrality measures and image vectors as features themselves for the predictive model. The paper does a great job at this aspect but seems to restrain from exploring different predictive models to the same depth. We believe that with a little time and effort to replicate their features, we can definitely improve upon the complexity of the predictive model to obtain significantly better results,

References

[1] Matthieu Nadini , Laura Alessandretti, Flavio Di Giacinto , Mauro Martino, Luca Maria Aiello Andrea Baronchelli 2021. Mapping the NFT revolution: market trends, trade networks, and visual features

[2] Mieszko Mazu, 2021. Non-Fungible Tokens (NFT). The Analysis of Risk and Return.

[3] Smriti Agrawal, Mohneet Sandhu, 2021. A Comprehensive Study on the Evolution of the NFT Market Future Prospects

[4] Christian Pinto-Gutiérrez, Sandra Gaitán , Diego Jaramillo and Simón Velasquez, 2022. The NFT Hype: What Draws Attention to Non-Fungible Tokens?