A PRICING OF DIGITAL COLLECTIBLES: AN EMPIRICAL EVIDENCE OF NFT VALUATION

by

Ivan Volovyk

A thesis submitted in partial fulfillment of
the requirements for the degree of

MA in Economic Analysis.

Kyiv School of Economics

2022

Thesis Supervisor: ______________________ Professor Olesia Verchenko

Approved by ____
Head of the KSE Defense Committee, Professor

____

____

____

Date _______________________________

Kyiv School of Economics

Abstract

PRICING OF DIGITAL COLLECTIBLES:
AN EMPIRICAL EVIDENCE OF NFT
VALUATION

by Ivan Volovyk

Thesis Supervisor: Professor Olesia Verchenko

In the year 2021, the total value locked in NFTs is growing to reach 60% of the global art market and 25% of the Ukrainian GDP. Even though by the end of this research all crypto dropped by 70%, I still believe that we have too few works conducted on this unique type of asset, which incorporates collectible, gaming, and financial mechanics. I used several hedonic models to estimate the determinants which drive prices inside the single collection, which will help artists to create more successful artworks in the future and users, to price their collectibles better, thus bringing liquidity into the market. I've found that the rarity trait is the big driver behind collectible pricing. Also, the cryptocurrency price influence NFT significantly, together with the token's previous price. Moreover, we have found some differences between primary and secondary markets, and the negative effect of the whitelisting procedure. This will help collection creators design better launches in the future.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I want to express my gratitude to all Research Workshop professors and my thesis advisor, prof. Olesia Verchenko. She had enough patience to describe to me repeatedly what it means to do research and what is expected from me.

I would also like to thank my classmates: Mykyta Horovoi for pinpointing some mistakes in my model and helping to structure my thoughts in a proper direction, and Inna Omelchenko, for extensive help in formatting this document and explaining to me the thesis formatting guidelines.

My special appreciation to my closest friends and family – for their encouragement and care, especially to my grandmother Alla – for motivational words like "stop complaining and do it" or "research won't do it by itself."

LIST OF ABBREVIATIONS

NFT. Non-fungible token. Digital ownership of virtual assets or collectibles.

EVM. Ethereum virtual machine. The main blockchain on which digital assets and financial instruments are located.

De Fi. Decentralized finance. The financial instruments are decentralized and created on the blockchain. They are permissionless and traded 24/7.

Trait. A characteristic that the subset of algorithmically generated NFT from the set of all collections has. For example, specific type of pants or hat color.

Chapter 1

INTRODUCTION

The total value of transactions in the global art market amounted to 65.1 billion U.S. dollars in 2021. Ukraine's GDP is about 200 billion for the same year. But the NFT art market total value of transactions is estimated to reach 40 bullions. Despite all this, there are less than 10 researchers conducted on the NFT market at the time of the start of this study. Despite a lot of prejudice related to cryptocurrency in the academic field, it's our obligation as economists to study the most influential social and financial phenomena. And undoubtedly, the biggest of them in 2021 was the NFT boom.

During the preparatory phase of this study, I concluded, that the two biggest barriers to entering the field for scholars were the understanding of the market and the complication in data gathering. The crypto is taught at the beginning, and it's hard to get all parts of it quickly. Despite all this, the benefits of the NFT market are what it is the simplest part of crypto so far. To do the research in this field, you need to know some of the basic concepts, and then you could jump right in to test various interesting hypotheses, from buyers' behavior to finance or some social patterns.

But the data gathering procedure is even harder. Even though, after creating services like Alchemy, or the Graph we don't need to set up a blockchain node to filter raw transactions and work with them, it still requires some programming experience to work with various APIs. Moreover, the data available to the public by some well-known resources like Kaggle are about the most popular parts of crypto, but it will be hard to find all trades for particularly NFT or some DeFi-related info. I used 4 services to gather data for this research, and 3 programming languages at least. For some data parts, like whitelisting, I couldn't find the data on the internet at all. So, I ended up reading the smart contract code and filtering the blockchain for the transaction on interest.

So, I believe one of the biggest contributions of this research is the dataset and data gathering techniques that I distribute to my fellow researchers[1]. Also, investigating the price determinants of the NFT collection, I ended up testing several hypotheses from different fields. For example, testing for the relationship between traits rarity and token prices contributed to the research of a well-known “Snob effect”. I tested different art indexes as proxy for traits, which is a separate area of interest in art-related economic studies. Also, I derived some important implications about primary and secondary markets and whitelisting, which could potentially help to improve the quality of NFT listings in the future. In the end, I tested for the relations between cryptocurrency and the NFT market and found a big correlation that contradicts to some researchers' assumptions about NFT market independence from the general crypto market. This, I believe, will help fellow scientists to gain interest in this market which is unfairly undervalued in terms of researchers for such total value locked.

NFT stands for a non-fungible token, tradable ownership of the virtual asset. This, new markets for digital assets emerged to some prominence in early 2021 and grew to about $550m of lifetime total traded volume. But the first NFT was created much earlier, back in 2014. Over $200m of that trade happened in the month of March alone. This trade growth was matched by large growth in public discussion and traditional media coverage of NFTs (Dowling 2021). This was associated with various high-profile NFT sales and the launch of many new projects.

---

[1] Repository with the datasets and data mining techniques which were used in the research: https://github.com/ivanvolov/thesis-code

For example, NFT by Beeple, the first tweet ever was sold for 2.9 million, the meme animation Nyan Cat was auctioned for about $0.6 million, and the band Kings of Leon sold their music rights as NFTs for the equivalent of $2 million (Ante 2021). Because an NFT is a right of ownership, not the object itself, it could be any type of digital asset. The most common types are picture and video based. They could be used to sell digital collectibles, artworks, and digitalized characters from sports and other games. Also, NFTs are perfectly suited to be a representative of the objects in virtual worlds. For instance, Decentraland land NFTs or Axie Infinity gaming characters.

NFTs have clearly become a ground-shifting paradigm that was produced by highly innovative approaches. For the first time since the creation of the Internet, artists can monetize digital content and do it without relying on the various counterparties or legislations. Gamers can now be not only users but owners of the digital world. We could also reproduce rare collectibles from the physical world digitally. For example, an artist could draw a painting, take a photo of it, wrap the photo into NFT and burn the physical artwork itself to make the token unique. After that, this painting could be sold in any part of the world instantly, and without constraints. Also, we could not only reproduce but produce digital entities online. And this production process will be relied on capital and labor, bringing the same value to the object, as it was if it is produced physically.

Various NFT projects have developed on the Ethereum blockchain, which at this point still has a strong dependency and relationship with cryptocurrency markets (Dowling 2021). It happens because NFTs are often traded against Ethereum's native cryptocurrency Ether, and Ether is also used to perform transactions, which is essentially required to operate your digital property. However, their pricing is seeming to be distinct from cryptocurrency pricing in terms of volatility transmission (Dowling 2021). Also, other EVM compatible blockchains, like Solana, Near, and Tezos become the better place for buying and selling digital assets. It happened due to low transaction fees thus making them more accessible for a general audience. Also, they have become the prevalent choice for gaming NFTs in consequence of quick transactions speed, which is greatly demanded in this field.

An NFT lifecycle starts with registering ownership of a digital asset on a blockchain, which is called minting. There are various types of minting mechanisms, and creators are choosing them according to their proposes. If you are selling artwork, you could just put it on the auction, choose its type, and wait for the better price to win. Gaming NFTs have their own special minting procedure, which is performed with the game mechanics in the mind. For example, Crypto Kitties is a game with the main mechanic of collecting digital cats, characterized by a set of discrete visual attributes, such as their fur, pattern, eyes, and color. Gamers can bread and trade these cats. By breeding cats, participants create additional cats with attributes that depend on the attributes of the parents and a random component. So, you need to buy some cats first, and after that, you could start minting additional cats by breading. Axie Infinity added a battling mechanic to this setup, so now you could get the characters of your opponent's by battling them.

In this research, we will concentrate on the algorithmically generated NFTs, which use slightly different minting mechanics. There are a lot of varieties, but the common pattern is the following: you need to buy a token, and then you could reveal its properties. So before revealing the procedure all participants will be uncertain what type of NFT they have bought. The properties are assigned according to a random process, so there is no mechanism of influencing the result of revealing otherwise from buying more NFTs. Also, the properties have some predetermined distribution, so one property could be rarer than the other thus influencing the rarity of the token itself. This approach has some minor variations, but it could lead to unpredicted results thus highly influencing the result of the whole campaign (Hasu, Anish Agnihotri 2021). For example, the minting procedure could be set up off-chain or on-chain, which would affect the customer's trust. You could reveal the NFT after buying or set some mechanism that will wait till the last collectible is sold. This will influence the anxiousness of people and could reduce their willingness to pay. Lastly, the scarcity of preminted NFT's could create a natural auction, which could be won by monopolists in capital or technology, in cost of pushing general users back. This could prevent the community of holders from forming, which could lead to mispricing of the asset in the long run.

An example of an algorithmically generated NFT, which forms part of the data analyzed in this study, is the Meebits created by Larva Labs. Meebits are one of the largest markets in NFTs at the time of writing, with an approximately $18.1k average price. The Larva Labs brand is well-known in the De Fi space. Firstly, they are known for the CryptoPunks, the most famous NFT collection created. CryptoPunks was launched in June 2017 and has not gathered a lot of attention. But in 2021, the collection showed its real potential. A few CryptoPunks were sold for more than $7 million each. So, as expected, the Meebits project was very well received by the community. Also, the CryptoPunks collectors got these Meebits for free. It increased the social media hype among top NFT influencers and spiked the initial price. A total of 11,000 Meebits went to existing Larva Labs collectors, and the remaining 9,000 were sold on auction. Despite the high initial price of 2,5 ETH (about $7,500), all Meebits sold within 8 hours (Modesta Masoit 2020).

Due to their pure collectible nature, Meebits NFT collection appears to be the best environment to study price determinants of art like tokens. Also, the second word in the NFT culture is "rarity", which means that consumers potentially should value token rarity more than other related characteristics. The rarity in this market is determined by token parameters, which have some predetermined distribution, we could compare two tokens and decide, which of them is rarer than another. So, we could contribute not only to the art research field but to the field of examining the impact of rarity using this unique market. Also, it has some benefits for researchers, because all transactions are presented in the blockchain, so we could study the whole dataset which is impossible to do with physical collectibles. Moreover, digital assets can't be loosed or damaged, which means their rarity is not diminished in time compared to rare coins or paintings. The reverse causality problem will also not affect the estimation, because the mintage amount is fixed and cannot be changed. And more importantly, we won't have a correlation between rarity and quality, because collectibles don't have external usage except merely trading and collecting.

The rest of the thesis is structured as follows. Chapter 2 presents an overview of the relevant academic literature on estimating the determinants of prices in art markets. An overview of the basic theoretical framework used in the thesis is given in Chapter 3. In Chapter 4, a description of the data used is provided. Chapter 5 provides obtained empirical results. Chapter 6 concludes the thesis with summarizing the discussion.

C h a p t e r   2

LITERATURE REVIEW

This paper mainly contributes to three strands of literature. Firstly, it could contribute to the literature on NFT and cryptocurrency. Some studies are focusing on treating NFT as an investment opportunity. For instance, (Kong and Lin 2021) provided a comprehensive analysis that NFTs serve as a novel investment instrument and investigated their return. Some researchers focus their attention on the topics like cointegration between markets (Ante 2021) or NFT phenomena themselves. Also, there are findings of volatility transmission between markets (Dowling 2021). For example, NFT pricing is quite distinct from cryptocurrency pricing in terms of volatility transmission. Also, there is low spillover between NFT markets. However, in cryptocurrencies and stock markets we observed a high spillover effect.

Others are focused on NFT auctions and intricacies in behaviors of biers and sellers (Casale-Brunet, Ribeca, Doyle and Mattavelli 2021) or (Fazli, Ali, Taesiri and Reza 2021). For example, this study is concentrated on the Crypto Kitties market (Kireyev and Lin 2021) and emphasizes some problems in the valuation of different tokens for auction participants. Also, the study highlights some difficulties in using hedonic regression for NFT valuation.

The Crypto Kitties series generated about 30 million USD in transactions since late 2017 and allows users to own and trade digital cats. This collection was also one of the first applications of NFTs and has several transactions of individual items in excess of 100,000 USD. Crypto Kitties have different mechanics compared to other collections studied in the field. Every token consists of sets of different attributes which could be random or inherited from the cat parents Fig. 1.

Classical literature on pricing and valuation of traditional collectibles often relies on linear hedonic price regressions to uncover the valuation of the price for every attribute. But in this paper, the author decided to develop a structural model of buyer behavior that accounts for the descending auction selling mechanism. This was done due to various factors connected with the NFT market nature.

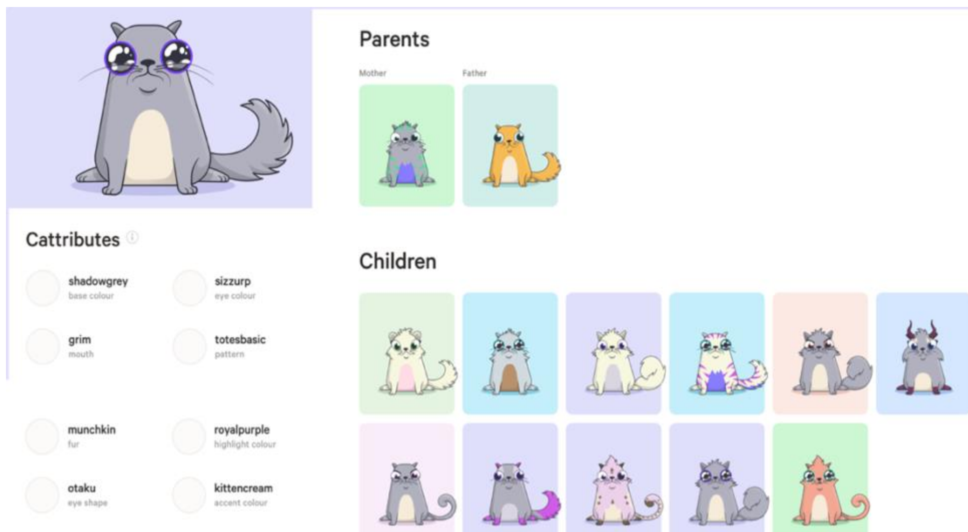

Figure 1. Example of Crypto Kitty with attributes

The mechanism states that users who acquire a digital cat may sell it by posting a descending auction on the marketplace. He also needs to specify a starting price, ending price, and duration. Also, the existence of failed auctions was added to the model and corrects for idiosyncrasies in seller pricing decisions. This results in a sale, and sellers pricing sub-optimally.

Mispricing and selection biases were also identified due to only using data on successful sales and setting the very wide intervals of closing and starting prices by sellers. As the result, the evidence that buyers value digital collectibles much like physical collectibles, but sellers may be uncertain of the value of their items was found. Also, this collection is a gaming NFT collection, which means that the cats have economic and investment usage and can't be treated as purely collectible items. So, here I could contribute to studying a price determinant of algorithmically generated NFT which has only esthetic and collectible usages.

As we could see, researchers now are mainly focusing on the market aggregates market rather than on some collections. It happens due to the novelty of the concept and the hardship of data acquisition. But there is a little amount of work dedicated to some particular token. I believe that looking at the collection on the micro-level could bring some valuable insights. Thus, this work could contribute by examining the Meebits collection and its community behavior. To the best of my knowledge, it's the first work conducted on the Meebits collection.

Secondly, we could contribute to the research that tried to measure the market value of rarity. The NFT market has very distinct features of valuing rare items, so this makes it a perfect candidate for analysis. Before, addressing this issue, scholars tried to look at rare coins (Koford and Tschoegl 1998) in which they measure the market value of rarity by examining identical rare coins with different mint sizes. They have studied data on prices and original mintages of the Morgan Dollar and Liberty Head with Coronet Eagle to test if the rarity could influence the price. In this example, rarity means pure scarcity. Researchers concluded that price correlates positively with rarity and elasticity of price with respect to original mintages sizes from -0.28 to -0.84. They also have not found evidence of premium for 'super rarity' in the dataset. Moreover, the demand for a complete set has not also been considered the source of rarity.

But they were faced with different obstacles. Firstly, the physical collectibles market is not liquid and transparent enough. They used a dataset of around two hundred transactions which was not enough to illustrate the whole market participants' behaviors. Also, this type of physical collectible gets lost over time, it is unsure how many of them were not lost after a certain period. This could bring some uncertainty about the true rarity of a set of coins.

Another research was conducted to estimate the demand for rarity using the data for unique audio recordings (Cameron and Sonnabend 2020). They studied price transactions of rare audio recordings sold on the online marketplace Discogs. This removes the issues of transparency to a certain extent because all transactions are digitally recorded and cannot be missed. However, this approach is still doesn't account for transactions that happened behind the scenes, and not recorded in the marketplace records.

The new approach to these studies begins with the invention of the NFT and its rapid growth. This technology allows for conducting research in the most transparent way. You could see all transactions together with their prices on the blockchain. Also, this information could be obtained by any market participant which means that this is the game of complete information. The mintage size and other parameters are predetermined, so could not be changed after the NFT listing. Also, tokens could not be lost, and their characteristics are not diminished with time. All these properties combined together lead to solving all of the problems which were encountered by the research above. This makes the NFT market the best fit for estimating the elasticity of rarity.

Among the first scholars who decided to use this new opportunity was (Lee 2021). He introduced a new market called NBA Top Shot. It is an officially licensed digital collectible of the National Basketball Association (NBA) video highlights. NBA is the premier men's professional basketball league in the world. This product is an NFT version of traditional basketball cards. Each product is called a moment and features an NBA highlight in video format. For example, the video of a Ja Morant moment features an NBA player Ja Morant's dunk play in an NBA game held on December 11, 2019. There are only 299 originals at this moment. But here the same products are offered with a different level of rarity in this market. There are only 5 moments with a circulation of 25, 49, 250, 299, and 1000 which share the same player Ja Morant's video highlight. Thus, these groups of moments provide the same utility to users and are only different in rarity. As the result, the elasticity of price concerning circulation (a rarity) is -0.74 on average. Also, less popular products are more elastic than more popular products.

But these studies could be improved by going further and introducing the new market of algorithmically generated NFTs and the Meebits collection This market is interested in the way, that the tokens don't have a particular rarity trait, which makes it more similar to the art market. But the traits of the item have a specific distribution, so buyers intuitively could understand which tokens are rarer. On the contrary, the Crypto Kitties could be considered the gaming NFT, and rarity is highly correlated with the utility which could be obtained by breading more unique or older cats. This means that the problem which appears in the first research on rare coins is still in place.

The NBA top shot could also be described as a purely collectible market. Moreover, the minting mechanism is very similar to rare coins studied above, so we are selling identical products but with different issuing sizes. It means that the same product could be minted in the future but with different mintage sizes. It returns us to the second problem. Therefore, the object of this research would be the Meebits collection which doesn't have a gaming characteristic, and all tokens were minted and revealed simultaneously. Every item in the collection has a set of properties with certain values, and the values are distributed randomly due to their rarity. This makes this market unique conjunction of art and collectibles, and we could study the price impact of the new meaning of "rarity".

The third stack of literature is the literature on the price determinants in the traditional art market. There is the various study in the field, for example (Rengers and Velthuis' 2002) in the German art galleries. Also, scholars have addressed different topics like differences in determinants depending on selling prices group in (Scorcu and Zanola 2011) or some complex size price relationship in Asian art markets (Nahm 2010). Also, (Renneboog and Spaenjers 2013) analyzed the art returns from 1957 and studied the investment opportunities of paintings. But the NFT market, especially an algorithmically generated part of it could be treated as a new form of art which have its obstacles in evaluation.

So, I believe my work was among the first to estimate the impact of different attributes on price together with studying some art indexes which is emerging in the field. The above studies suggest the importance of estimating the price determinants in the NFT markets. In the following chapter, we will discuss a methodology for estimating using the Meebits transaction dataset.

C h a p t e r   3

METHODOLOGY

As we know, the hedonic OLS regression is commonly used in the analysis of the various art markets to determine the relationship between a set of characteristics of collectibles and their corresponding hammer prices (Rengers and Velthuis' 2002). This technique is quite established and reliable and received various modifications created for some particular cases. For example, scholars used quintile models to study the price determinants of Picasso paintings (Scorcu and Zanola 2011). It was necessary because the traditional approach relies upon the mean of the conditional distribution of the dependent variable. But various studies suggested that some characteristics are expected to be valued differently across a given distribution of selling price, which was later confirmed for the Picasso art market.

Others used some nonlinear models with a two-step nonparametric kernel estimation technique (Nahm 2010) in the Korean art market to estimate the relationship between size and price. It was applicable, due to the complex nonlinear relationship in preferences between the size of the painting and its price. Although, the literature on pricing artworks is always loosely related to the calculation of the art price index (Ginsburg 2006), which is why the hedonic model become extremely popular. So, in the end, this hedonic regression framework ends up being modified for the particular art market it's been used in and its distinct characteristics.

Expectedly, the main direction of the modification for the NFT art market, or the market of digital collectibles becomes the direction of rarity. The distinct feature of digital collectibles, it's that every collection and token in it has a particular rarity trait, which could be estimated explicitly or implicitly. That is why the methodology is very similar to the one used in the physical world for examining the dependence between rarity and prices.

For example, this model Eq. 1. from (Koford and Tschoegl 1998) for a physical world which is the logarithmic transformation of the model derived theoretically using some psychophysical evidence and assuming constant proportional erosion of coins over time is very similar to the model Eq. 2. for the digital collectibles of the NBA Top Shot market (Lee 2021).

$$\ln(Price_i) = \alpha + \beta_1 \ln(Mintage_i) + \beta_2 Age_i + \varepsilon_i \tag{3.1}$$

Here $Price_i$ is the price of the ith coin, defined by year and place of issue. $Mintage_i$ is the quantity of the ith coin that the government minted in the year and place. Also, Age_i is the number of years between ith mintage and the last mintage in the series.

And here, researchers used fixed effects regressions circulation as reverse to rarity:

$$\ln\left(y_{jgt}\right) = \beta_1 \ln\left(x_{ig}\right) + \alpha_g + \tau_t + \varepsilon_{jgt} \tag{3.2}$$

where β_1 is the coefficient on the log of circulation and g is a group number. The j variable is a moment, depending on g. The t variable is the time of day and y_{jgt} - price for moment j in group g at time t. x_{ig} is the circulation for moment j in group g. But in both studies, the token has the explicit rarity trait, which could be clearly added to the model.

Other work conducted on digital gaming collectibles, used machine learning methods together with hedonic regression, to build the structural model and account for mispricing and selection biases. It was proven, that sometimes a hedonic regression is not enough to clearly estimated the price determinants, due to market specifics. The following model was used to estimate the buyers' preferences:

$$\ln(P_{jt}) = \sum_k \beta_k X_{kjt} + \alpha W_j + f(t) + \gamma Z_{jt} + \varepsilon_{jt} \qquad (3.3)$$

where P denotes a sale price of jth token at time t, W is a set of continuous attributes of the item with associated coefficients α. X_{kjt} - an indicator for the presence of discrete attribute k in item j at time t with β as the associated coefficient. They also included time-specific fixed effects $f(t)$, Z_{jt} is a set of control variables with associated coefficient γ, and ε_{jt} is an error term. They also came up with the conclusion, that some characteristics (generation ID in this case) could be used as a proxy for rarity.

In the Meebits market, which we are examining in this study, we have around 500 distinct attributes and about 6 continuous attributes. Also, this model could suffer from multicollinearity, which is common for similar studies from the physical world. For example, (Renneboog and Spaenjers 2013) have used a lot of various action artist levels, painting levels, and auction level characteristics, but compensated for this with the big dataset. In my dataset, I have a lot of different tokens, but most of them were traded 1-3 times, so it would be better to find the proxy for asset-level distinct characteristics to improve our model. Also, this collection is quite new, so I can't use some sophisticated techniques due to a lack of data. So, we decided to estimate the classic hedonic regression Eq. 4. And do several robustness checks. We will test two models with different district characteristics. The first will be where the color and type for the specific trait are added separately, and in the second model, we will make intersects out of color and types and see, if this produces better results. This will be equivalent to creating a discrete trait for every different piece of clouts and accessories. This will help to better interpret the result because the change in trait type will not change the corresponding subset of available trait colors.

$$\ln(P_{jt}) = \sum_k \beta_k X_{kj} + \alpha W_j + f(t) + \varepsilon_{jt} \qquad (3.4)$$

Here P denotes a sale price of jth token at time t, W is set continuous attributes of the item with associated coefficients α. We will use the previous times and prices of the token, together with the cryptocurrency rate and some network-related characteristics. X_{kjt} - an indicator for the presence of discrete attribute k in item j with β as the associated coefficient. We also included time-specific fixed effects $f(t)$, we will test for a month and week-related effects, as suggested in the traditional financial markets, and ε_{jt} is an error term.

Then, we will use three different ratings, as a proxy for all discrete characteristics, and add them to the model. The first model will have all of them together with descript attributes, the second will have tree ratings alone. The model from 1 to 3 will have only one of the corresponding ratings as a proxy for rarity. Also, these ratings are greater, the rarer is the token (the rarer traits it has). This will help us to test the hypothesis about the dependence between price and rarity. I have a hypothesis that the rarity dependence on prices is not linear, so we will add a quadratic term for rating to the model and test it.

Separately, for the model without rating and with discrete variables, we will obtain the coefficients for every attribute. When we will remain only economically and statistically significant and regress these attributes' α on the corresponding attributes rarity score parameter. The score parameters are a value calculated in the first rating by the following formula

$$S_j = \frac{20000}{N_j} \tag{3.5}$$

where S_j is a score value for the attribute j, and N_j is the amount of time this attribute accrued in the dataset. By conducting this procedure on all remaining coefficients and separately, on groups of the coefficients divided by trait type we will get the alternative estimation of the dependents between the trait's rarity and its contribution to the price.

Another hypothesis is to estimate the dependence between NFT price and cryptocurrency price. Because theory suggests that we don't have a correlation between crypto price and NFT market aggregates, we have found some dependence between eth price and token price. Also, the effect of the primary and secondary markets together with the "whitelisted" effect was measured. We divided the dataset between primary and secondary markets and used the best model from the previous stage to estimate two models. When we will compare coefficients and intersection between their confidence intervals to test the hypothesis that there is a significant difference between primary and secondary markets.

To test the difference between the behaviors of users who received whitelisted tokens for free and other participants, we added an indicator for whitelisted sellers. Also, as we have a various type of whitelisted sellers (for example someone received 1 token for free, and at the same time another seller could receive 100 tokens) we added the "amount whitelisted" as a separate variable to estimate how this effect changes depending on the number of free tokens.

Chapter 4

DATA

For the purpose of this research, I decided to use the data on the Meebits NFT collection. This collection was issued by a group of artists from Larva Labs and consisted of 20000 unique algorithmically generated NFT. This mean, that every one of them is unique by its appetence and traits (parameters which are used to generate and describe specific NFT), thus could be viewed as a collectible object or a unique piece of art. As far as I know, this is the first research conducted on this class of NFT. Previously, the NBA Top shot and Cryptokitties collections were examined.

The first one could be also viewed as a pure collectible token, but the generation algorithm is completely different which brings another economic meaning to this asset. They are minted in packs that consist of n (mintage size) identical tokens (master token). And every master token is a digital illustration of the specific basketball event, and player (moment). Also, several bunches could be minted using one master token but have different mintage sizes, so the rarity of the token in the bunch could be determined by the mintage size and some moment-specific characteristics. This is very similar to the collectible card games with physical tokens which is quite popular in the US. Moreover, the supply of tokens is changing over time, by issuing another butch and adding more recent master tokens by the issuing firm. Thus, they are closer to the rare coins, examined in the work of (Koford and Tschoegl 1998). then to the pictures and art objects.

The other collection which has received the researchers' attention in the field is Cryptokitties. This is gaming NFT, which is used to play in the game with the same name. It's also a play-to-earn game, where you could earn real money to buy breading and fighting using your Cryptocat. So, on one hand, this token has not only strong entertaining usage but also could serve as a barrier to entering the game, which makes it similar to the game subscription or license in the traditional gaming market. On the other hand, it also serves as an investment asset or the means of production, because you could produce other cats by breading or conquering tokens of your enemies. The supply of this token is also increasing, over time. Thus, it's also not purely an art collection, with several economic meanings and approaches to the valuation.

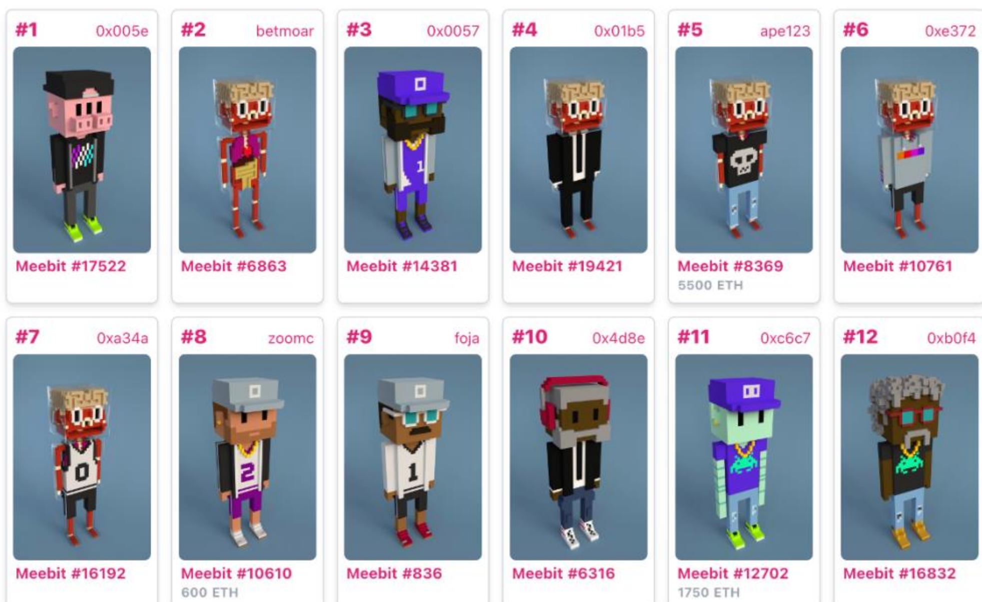

Figure 2. Example of Meebits appearance controlled by traits

But the algorithmically generated NFT is a unique class of assets that have a strong economic characteristic of art, and thus, could be studied using similar methods to the studies of price determinants in the art markets (Nahm 2010). Also, they could be used as a digital avatar or the representation in the metaverse, but we won't cover this usage in the research. The supply of this token is constant, and the collection is released only once. So, the holders could not feel the diminishing value of their tokens due to the new issuance. However, the same artist or group or artist could produce new collections, which is also the case for Larva Labs. They had already issued 2 collections, the Autoglyphs and Crypto Punks, before appearing in the Meebits, which have brought them fame and reputation on the market. But it is also the case for the alive traditional artists, they also could write a new painting. And every new painting could potentially increase the value of the existing works, by bringing him more attention and thus, increasing prices. At the same they, the new collection could in general increase the price of the existing collections in the digital market. So, the algorithmically generated NFTs should theoretically incorporate these features.

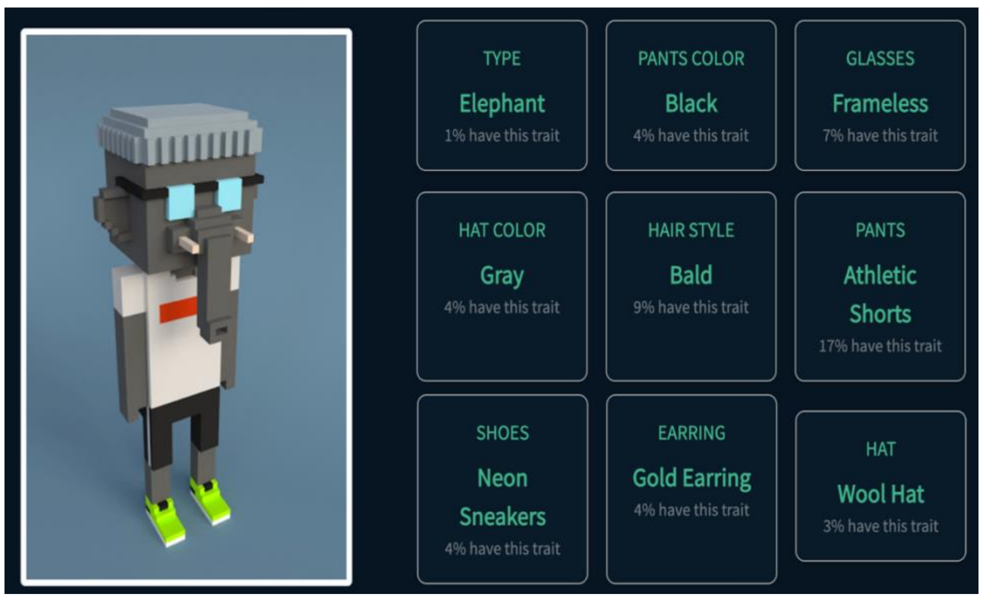

Figure 3. Meebit with its traits and corresponding rarity

The tokens of this asset class could be described by traits. Every token could randomly have some traits from the closed set of possible traits, which is randomly assigned to them at the time of issuing the collection. The traits have full control of the token appearance Fig. 2. Some collection even concentrates only on the traits themselves and leave the appearance interpretation to the community. But Meebits have the traits interpretation built in their algorithm, so anybody could generate a specific Meebit picture, providing the set of traits. Also, some traits' values are more unique than others buy design Fig. 3. So, the community should value their tokens based not only on their appearance but the overall rarity of the traits or some specific parts. In the results, the most expensive Meebit currently sold has the unique trait of "Dissected body type", which only exists in 5 out of 20000 tokens.

We will use the dataset of the transaction of buying and selling tokens on the OpenSea auction platform. The vast majority of trades were conducted using this auction platform, so we could safely approximate the dataset using only this source of the transaction. Out of 20000 NFTs of the Meebit collection, only 7099 unique tokens were sold using marketplaces. This account for roughly 35% of the collection. This is not much, but it's a common issue with liquidity for these markets. Also, not every owner decided to sell their tokens, and want to leave them as their virtual identity.

Table 1. The distribution of sales for one token

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	1.000	2.000	1.927	2.000	10.000

1	2	3	4	5	6	7	8	9	10
3380	2054	920	453	174	71	31	9	1	2

I was using the OpenSea API and NodeJs to get the record of about 13000 transactions conducted on this marketplace. I was also interested only in Buy and Sell events and ignored all the bids and asks events of the auction. Separate research could be conducted to analyze the auction level date and evaluate the structure of the auction. Also, (Kireyev and Lin 2021) conducted a study on the auction level in Cryptokitties and found out what sellers tend to set up very wide price ranges, so these events could bring distortion to the price determinants estimation.

Moreover, the dataset which was used in the study contains about 150 000 records, which is larger compared to the current market. This happens due to the gaming nature of the Cryptokitties collection, and it is much older than the one examined in the study. So, we will wait for the market to mature and leave the auction analyses for future scholars. To illustrate the problem with liquidity, out of all transactions, we have the following distribution of the sales of one unique NFT in the collection Table 1. So, every token was sold mostly 1,2, or 3 times. Only some particular tokens were sold more than 7 times. This suggests that we could treat this date as a plane or cross-section dataset.

Every trait is a categorical variable, which could have a value from 0 to N, where N is specific for every trait. The dataset has the following traits group (21 in total):

- Body type: every token has a body type, and there are 8 of them. The values are (from least rare to most rare): *Human, Pig, Elephant, Robot, Skeleton, Visitor, Dissected*. Here *Human* body trait is common for 18881 tokens in the collection, and only 5 NFT have a *Dissected* body type. The correct values for the trait rarity were also scraped from the OpenSea and could be viewed in Appendix A.

- Pants type. Every Meebit wears pants. We have only one NFT without pants (№6863), but he was not sold and is not present in the So, we could assume that all tokens have pants. This trait has 10 distinct values and includes skirts and jeans.

- Pants color. Most treats have a similar trait which represent its color. For example, *Cargo Pants* could have 5 different color combinations: Blue Camo, Camo, Dark Grey, Dark Red, and Denim (from most to least rare).

There are 20 different pants color schemes, so in total, we have 10*20~200 possible combinations of pants with its color. Likely, specific pants type is only associated with a small subset of pants schemes, which reduces the overall number of combinations. Also, the trait *No Color* is present, and two types of pans have this trait (*Ripper Jeans, Suite Pants*). Moreover, this pants type has no more color variation, so we could exclude the *No Pants* trait from the sample because it has no useful information. Thus, if we have *Ripper Jeans*, it automatically means that it has no color. This approach also helps to address the multicollinearity issues which is huge in this dataset. Also, *Argyle* color, is only present in *Athletic Shorts*, so could be also excluded from the model.

- Shoes are similar to pants. Every remained token wear shoe, and we have 23 model types. The Shoe Color trait represents the color of shoes, and we have 9 of them. The *No Color* trait is also excluded due to similar considerations as above.

- We could conduct the same procedure for the following trait groups, excluding *No Color* and colors which are unique for some trait values. As the result we have:

    - Shirt - 34 types of shirts and 16 types of *Shirt Color*.
    - Hair Style – 22 types of hairstyles including bold, which have no color. Hair Color – 12 colors.
    - Beard – 7 types. Also, this trait together with others that will follow next is not mandatory, so could either be present or not present in the NFT. Beard Color has a variation of 4 colors.
    - Hat – 7 types. Hat Color – 10 colors.
    - Glasses – 5 types. Glasses Color – 3 colors.

    - Overshirt – 5 types. Overshirt Color – 15 colors. This trait is special because it correlates to Shirt. It happens because if the Shirt type is Hoodie, the Meebit can't wear an overshirt.

- Also, we have 3 trait groups without color. The first one is Tattoo. Due to the uniqueness of every tattoo, I decided to reduce this trait to binary dummy, which is true, of the token has tattoo mad false, if not. Also, Meebit could have one of the three types of a necklace (*Gold Chain, Gold Necklace*). And one of 3 types of Earing (*Gold Earring*, *Gold Earrings, Gold Hoops*).

- The last remaining trait category is the Jersey number. This trait only exists if Meebit wears a type of Jersey collection Shirt: Classic Jersey, Basketball Jersey, Snoutz Jersey" Jersey. It could have value from 0 to 10.

As the result, we have a lot of categorical variables which are transformed into around 500 dummies. Also, I will use two variations of explanatory variables in the model and determine, which will have better descriptive power. The first one is the translation of all categorical variables into dummies. And the second one is the intersection between corresponding dummies of type and color. Although, the first set will have fewer variables, thus, could isolate the effect more accurately, it will be harder to interpret, because the change in Pants type will lead to changes in the color, which is acceptable for Pants value, thus, misleading the "ceterus paribus" condition.

Additionally, we could get the auction level date from the OpenSea API. We have "price in eth", which represents the amount of a particular cryptocurrency that was paid for the order. Also, we have the "cryptocurrency type" variable, so translated the price into the dollar value, using the data on ETH/USD pair from CoinMarketCap. Similar research in the field included the eth price and cryptocurrency rates separately into the model, but I think it's overcomplication because market participants usually translate their investments into per dollar terms. Also, we know the date of all sales Fig. 4.

We are concentrated on sales, conducted from the beginning of May, the issuing of the collection till the end of November. This is all data available so far, and it incorporates the most active trading before the market went into stagnation at the end of February 2022. So, we have 7 months of observation. The black line represents the number of sales conducted during the day. The gray line represents the number of sales conducted during the day if the price of the sale was the all-time largest. The yellow line is the same as the gray, but for the secondary market. As we could see, we have an increase in transactions at the start, then we have a sudden drop.

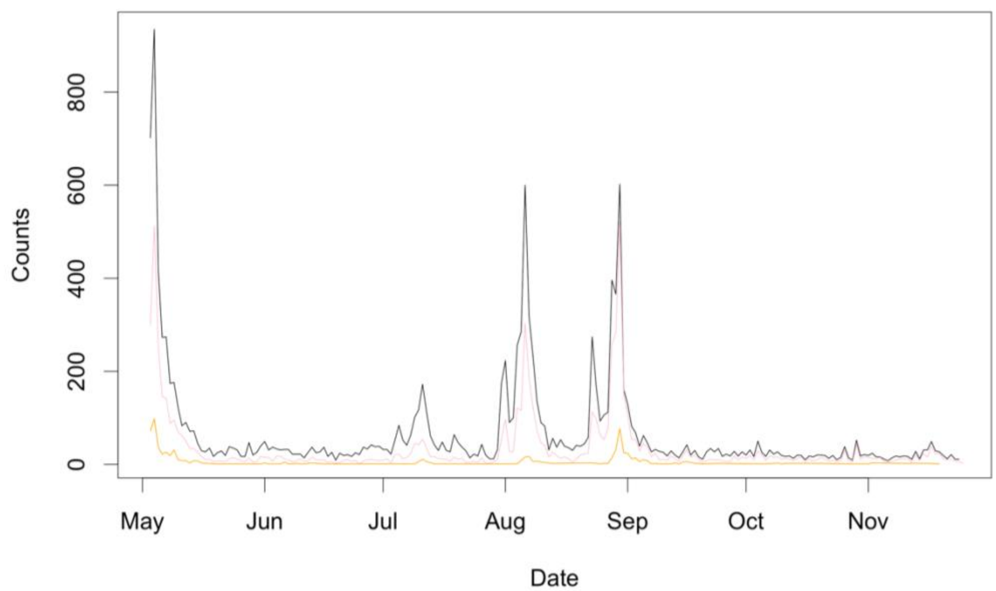

Figure 4. The time distribution of sales

The increase happens due to the first sales on the primary market, but then, after all, NFTs were minted, and people started to sell them off to get profit, which caused a death loop. After that, we have 3 main spikes in the trading activity.

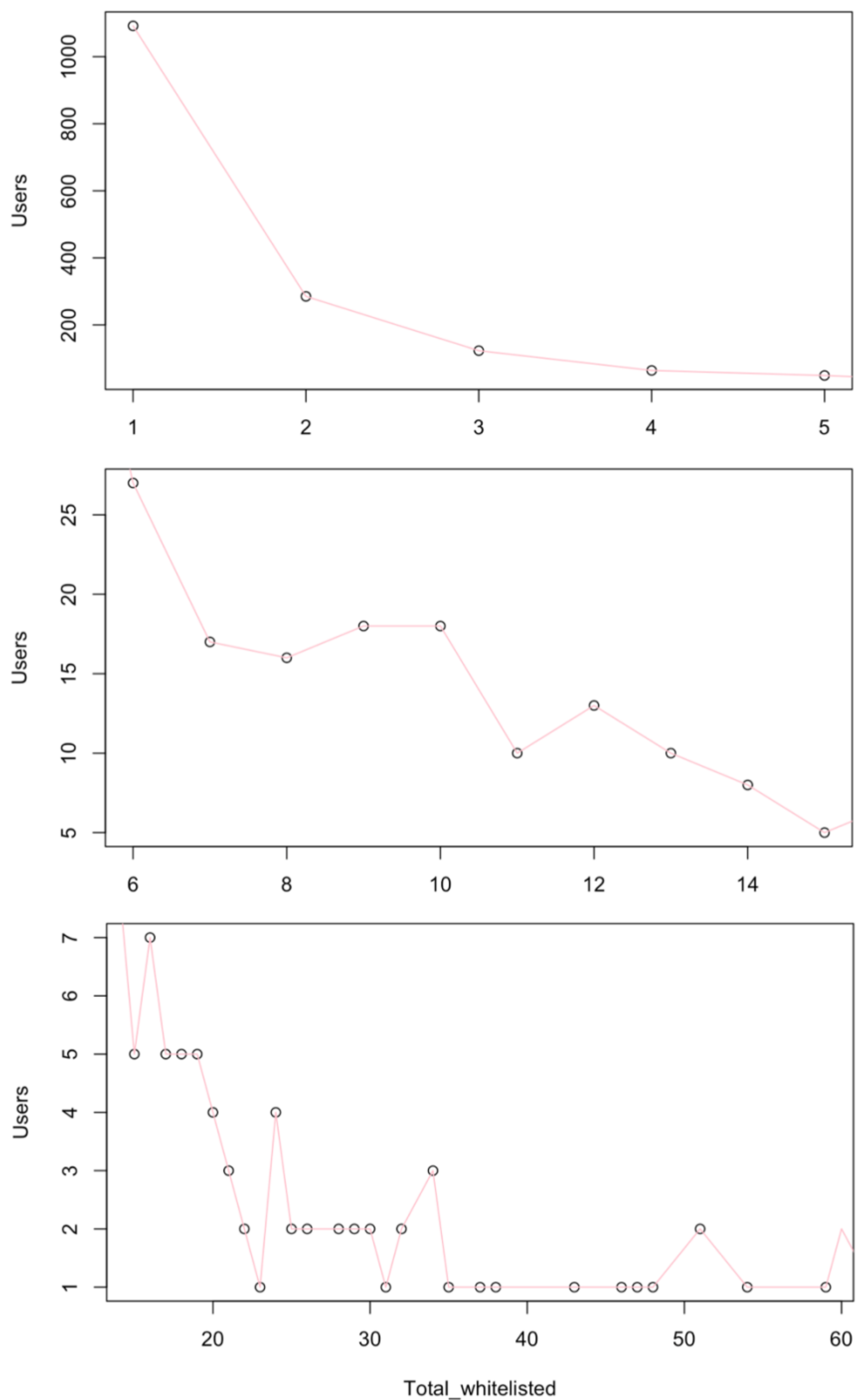

Figure 5. The number of free tokens whitelisted by one user

The gray and yellow lines depict the main patterns of the black line. Also, we could see that the yellow line is mostly near 0, except for the spikes. This means, that the all-time largest sales are not present on the secondary market. Overall, the data looks quite stationary, except for several spikes. This means, that we could try to incorporate time-specific daily or weekly effects in our model to account for them.

Also, I add the explanatory variable, which accounts for the primary and secondary markets. This division is justified, because all collection is revealed simultaneously, and people don't have references based on which they should value their tokens. Thus, the first sale of every token is marked as the primary market, and the next is marked as secondary. We predict what people will have different behavior patterns and price determinant in both markets, due to incompleteness of information.

Also, the previous price, for the already traded token is added as a separate variable to the dataset. It makes sense because people would in general look at the price of the previous sale before purchasing the token again. Moreover, the time after the previous sale is calculated and added to the sample, because it's suggested by the literature on the topic (Nadini, M., Alessandretti, L. 2021). Also, theory suggests, that the increase and decrease in cryptocurrency itself could change the mood of the market, so we incorporated the variable which represents the percentage increase/decrease in ETH/USD price from the previous sale.

$$\text{Score} = \sum_{t=0}^{trait\ Count} \frac{20000}{\beta_t} \tag{4.1}$$

The community is using special indexes to determine the token rarity, and thus, try to estimate the proper prices. I used 3 types of indexes. The first one is a rarity value calculated using the LooksRare methodology

where β_t is the number of times when trait t occurs in other NFTs in a dataset. This index accounts for all traits and their rarity but omits the combination of indexes or specific community preferences for particular traits. The second one is rarity ranking. It was created by the RarityTool platform and was web scraped from their website using R. It ranks all tokens from 1 to 20000 and potentially incorporates weights and some intersection of traits.

This rank was built on another rarity score index, calculated by the same tool, and then sorted descendingly. By this procedure, the distances between close-ranked NFTs should be distorted. So, theoretically, the rarity score should be the best proxy got the token individual characteristic and represents its rarity simultaneously.

As I mention above, Larva Labs has issued 2 collections before releasing the Meebits. They decided to distribute it in the following way: 11000 to the token holders of the existing collection, and 9000 to the public for sale. This means, that 55% of people received the token for free (excluding transaction price), so this affected the prices on the primary market. Also, some users have several different tokens, so they potentially received more than one free Meebit. As the result we have two variables, "whitelisted" and "whitelisted count", which represent the drop parameters for every seller in the data frame.

To get this information we have used the Ether Scan, to find the time range, in which people were allowed to mint free tokens. The range is between from May 3 to 8 (between blocks 12393673 and 12358264). Then I parsed the Ethereum blockchain using Alchemy API and get the set of transactions for this period. After that, I used Infura to check every transaction for success status, and filter by the “Mint with Punk O Glyph” method. As the result, I obtained the data frame which represents every minter of free Meebit Fig. 6. As we could see, most people get about 4-5 free tokens, but somebody managed to get around 200.

Also, we could use the auction-level information about buyers and sellers to construct a networking graph. It could be insightful to study the NFT transfers in this form. As a result, we have a network with very low density, so we decided not to use it in this research. However, we have used this information to calculate the indegree and outdegree distribution of all nodes and add it to the dataset by buyer. It would be a good proxy of how many deals of each type (buying - accumulating or selling - brokering) an actor makes and could improve our model. We used buyer’s nodes here because, in the end, they are the ones to decide, what the price will be, due to the auction nature. This actor-level data could bring some insides, about what actor characteristic helps to determine the price.

Collecting this dataset was very tedious work, which require both knowledge of the crypto market, programming, and blockchain-related techniques from the individual. I assume, that the barrier to entering data mining the blockchain data, is one of the main reasons for the little research in the field. Thus, I contribute to the literature and future research just by gathering and cleaning this dataset which could potentially bring more scholars into the topic.

C h a p t e r   5

ESTIMATION RESULTS

In this chapter, I present the estimation results of my model. First, we will start by choosing between two specifications. The first one has 210 discrete variables, and the second one has 586 discrete attributes.

Table 2. Estimation results

	Dependent variable:	
	log(price_in_usd)	
	(1)	(2)
Close	0.001^{***}	0.001^{***}
	(0.00001)	(0.00001)
winner_flow	0.022^{***}	0.031^{***}
	(0.002)	(0.003)
winner_t_sold	-0.007^{***}	-0.009^{***}
	(0.0003)	(0.0004)
winner_t_bought	0.004^{***}	0.006^{***}
	(0.0003)	(0.0004)
prev_price	0.003^{***}	0.003^{***}
	(0.001)	(0.001)
cur_increased	-0.00000	0.00000
	(0.00000)	(0.00000)
time_from_prev	-0.00000	0.00000
	(0.00000)	(0.00000)
whitelisted	0.009	0.012
	(0.008)	(0.010)
whitelisted_c	-0.0004^{***}	-0.0004^{**}
	(0.0001)	(0.0002)
trait_Type_2	1.318^{***}	1.310^{***}
	(0.031)	(0.035)
trait_Type_3	2.249^{***}	2.148^{***}
	(0.037)	(0.042)
trait_Type_4	3.546^{***}	3.464^{***}
	(0.065)	(0.076)
trait_Type_5	3.651^{***}	3.600^{***}
	(0.076)	(0.088)
trait_Type_6	5.022^{***}	4.895^{***}
	(0.121)	(0.141)
weekday_Fri	-0.172***	-0.167***
	(0.011)	(0.013)
weekday_Sat	-0.102***	-0.102***
	(0.012)	(0.013)
weekday_Sun	-0.137***	-0.141***
	(0.011)	(0.013)
weekday_Thu	-0.226***	-0.227***
	(0.012)	(0.014)
weekday_Tue	0.031***	0.033***
	(0.011)	(0.013)
weekday_Wed	-0.149***	-0.139***
Constant	7.240***	6.903***
	(0.097)	(0.052)
Adjusted R^2	0.827	0.772
F Statistic	286.999^{***}	75.568^{***}
Note:	$^{*}p<0.1$ $^{**}p<0.05$ $^{***}p<0.01$	

The second is an intercept and only consists of existing attributes which are present in the sample. We could see part of the estimation results in Table 2, and the whole report could be found in Appendix B. I decided to show only continuous variables and common discrete values.

Table 3. Seven regressions, estimation results

	Dependent variable:						
	log(price_in_usd)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Close	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.00001)	(0.00001)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
score_r	0.00004**		0.461***	0.873***			
	(0.00002)		(0.024)	(0.024)			
Rating	-0.00001***		-0.034***		-0.047***		
	(0.00000)		(0.001)		(0.001)		
rarityS	0.001***		1.262***			2.124***	
	(0.00004)		(0.060)			(0.062)	
win_f	0.022***	0.022***	0.066***	0.077***	0.071***	0.080***	0.087***
	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
win_ts	-0.007***	-0.007***	-0.010***	-0.011***	-0.010***	-0.012***	-0.012***
	(0.0003)	(0.0003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
win_tb	0.004***	0.004***	0.006***	0.007***	0.006***	0.008***	0.008***
	(0.0003)	(0.0003)	(0.0005)	(0.001)	(0.0005)	(0.001)	(0.001)
prev_p	0.003***	0.003***	0.008***	0.006***	0.008***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)
white	0.008	0.009	0.058***	0.071***	0.062***	0.065***	0.074***
	(0.008)	(0.008)	(0.013)	(0.014)	(0.013)	(0.014)	(0.015)
white_c	-0.0004***	-0.0004***	-0.001**	-0.001***	-0.0004*	-0.0004*	-0.0004*
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Fri	-0.175***	-0.172***	-0.167***	-0.154***	-0.163***	-0.167***	-0.156***
	(0.011)	(0.011)	(0.018)	(0.019)	(0.018)	(0.019)	(0.020)
Sat	-0.102***	-0.102***	-0.102***	-0.096***	-0.103***	-0.098***	-0.097***
	(0.011)	(0.012)	(0.018)	(0.019)	(0.019)	(0.020)	(0.020)
Sun	-0.138***	-0.137***	-0.135***	-0.125***	-0.135***	-0.133***	-0.128***
	(0.011)	(0.011)	(0.018)	(0.019)	(0.018)	(0.019)	(0.020)

Table 3 - Continued

	Dependent variable:						
	log(price_in_usd)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sun	-0.138***	-0.137***	-0.135***	-0.125***	-0.135***	-0.133***	-0.128***
	(0.011)	(0.011)	(0.018)	(0.019)	(0.018)	(0.019)	(0.020)
Thu	-0.225***	-0.226***	-0.242***	-0.247***	-0.245***	-0.254***	-0.256***
	(0.012)	(0.012)	(0.019)	(0.021)	(0.020)	(0.021)	(0.022)
Tue	0.033***	0.031***	0.032*	0.031*	0.038**	0.031*	0.035*
	(0.011)	(0.011)	(0.017)	(0.018)	(0.017)	(0.018)	(0.019)
Wed	-0.148***	-0.149***	-0.148***	-0.142***	-0.146***	-0.138***	-0.135***
	(0.012)	(0.012)	(0.018)	(0.020)	(0.019)	(0.020)	(0.021)
Con	7.401***	7.240***	7.148***	6.768***	7.438***	6.848***	6.947***
	(0.098)	(0.097)	(0.062)	(0.065)	(0.063)	(0.065)	(0.068)
Adj. R^2	0.831	0.827	0.564	0.496	0.536	0.491	0.446
F Stat.	292.8	286.9	737	613.1	718.7	600.2	526.17
Note:						*p<0.1 **p<0.05 ***p<0.01	

As we could see, the model with intercepts has lower R2 and looks over complicated, due to the extra 100 variables which are taken into account. Also, other coefficients of interest have a similar value and overlap in their confidence interval, so we will proceed with smaller specifications for the next models.

Let's compare 7 mod els Table 3, the first two have all vectors of discrete attributes plus 3 ratings, and the second has only the vector. The third has only three ratings and from the fourth to six, we have only one rating correspondingly. In the last regression, we have neither attributes nor their proxy. As we could see, the ratings are doing a good job of incorporating all individual characteristics. Also, adding them to the first model is not improving it at all. The coefficient of interest is not changing much between the models.

Also, the last model is the worst, so the ratings definitely have some descriptive power. The last conclusion is that by combining the rating together in the third model, we have some improvements in comparison to the individual ratings.

Table 4. Regression with quadratic terms

	Dependent variable:				
	log(price_in_usd)				
	(1)	(2)	(3)	(4)	(5)
I(score_rating2)		-0.045***	-0.099***		
		(0.003)	(0.002)		
I(Rating2)		0.003***		0.006***	
		(0.0002)		(0.0002)	
I(rarityScore2)		-0.481***			-0.916***
		(0.023)			(0.016)
Close	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.00001)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
score_rating	0.00004**	0.996***	2.195***		
	(0.00002)	(0.047)	(0.038)		
Rating	-0.00001***	-0.066***		-0.165***	
	(0.00000)	(0.004)		(0.004)	
rarityScore	0.001***	5.091***			9.180***
	(0.00004)	(0.210)			(0.136)
winner_flow	0.022***	0.050***	0.062***	0.064***	0.054***
	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)
winner_total_sold	-0.007***	-0.009***	-0.009***	-0.010***	-0.009***
	(0.0003)	(0.0005)	(0.001)	(0.001)	(0.001)
winner_total_bought	0.004***	0.005***	0.006***	0.006***	0.006***
	(0.0003)	(0.0004)	(0.0005)	(0.0005)	(0.0005)
prev_price	0.003***	0.008***	0.007***	0.008***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
whitelisted	0.008	0.053***	0.058***	0.061***	0.055***
	(0.008)	(0.012)	(0.013)	(0.013)	(0.013)
whitelisted_count	-0.0004***	-0.0004**	-0.0003	-0.0004**	-0.0005**
	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Constant	7.401***	6.808***	6.498***	7.854***	6.517***
	(0.098)	(0.065)	(0.061)	(0.062)	(0.059)
Adjusted R^2	0.831	0.613	0.558	0.572	0.588
F Statistic	292.802***	803.861***	752.031***	793.925***	850.365***
Note:			*p < 0.1	**p < 0.05	***p < 0.01

Next, let's add a quadratic term to the ratings and decide, which one is better Table 4. Also, we divided all rating values by 1000 in order to simplify the interpretation.

We could see that all models improved a lot, and the best one is with the rarityScore rating. Even though, combining all ratings together brings a slightly higher R2, we could see that the other bettas are the same, so I will go with the last rating in order to simplify the model. Also, we could not reject the hypothesis, that price depends on rarity nonlinearly, because the coefficients near the rarityScore are statistically and economically significant. As we could see, the effect of rarity is diminishing. The price of the token is also dependent on the previous price heavily.

Moreover, the more buyer sold tokens, the less will be the price, and the more bought the more he is willing to pay. This is illustrated by the coefficients on the winner_flow variable. Also, we could see that the "Close" variable which represents this day's Ethereum close price is a good predictor of the token price. Also, this variable is stationary, so we could use it in our model.

So, if the price of eth is changed by 1 dollar, we could expect the change of token price by about 0.1 percentage point. It could look like not an economically significant term, but the asset price is really volatile and could change by 100 dollars a day easily, so we would have about 10 percentage points change.

Considering the day of the week effect, we could see that the effect is only positive on Friday, then is slowly diminishing to Monday, and then have a strong decrease to its bottom on Wednesday. It's quite similar to the effect found in some traditional stock markets. Moreover, the month effect is following the general crypto trend during the months of research, so it's become even more evident that the NFT market is highly correlated with crypto.

Table 5. Two markets regression

	Dependent variable:	
	log(price_in_usd)	
	(1)	(2)
I(rarityScore2)	-0.907^{***}	-7.985^{***}
	(0.020)	(0.894)
Close	0.001^{***}	0.0005^{***}
	(0.00002)	(0.00002)
rarityScore	9.105^{***}	7.872^{***}
	(0.173)	(0.320)
winner_flow	0.051^{***}	0.038^{***}
	(0.004)	(0.007)
winner_total_sold	-0.009^{***}	-0.008^{***}
	(0.001)	(0.001)
winner_total_bought	0.004^{***}	0.005^{***}
	(0.001)	(0.001)
prev_price		0.400^{***}
		(0.008)
whitelisted	-0.034^{**}	0.137^{***}
	(0.015)	(0.023)
whitelisted_count	-0.0004	0.009^{***}
	(0.0002)	(0.002)
month_06	-0.222^{***}	-0.182^{***}
	(0.039)	(0.035)
month_07	0.414^{***}	0.255^{***}
	(0.040)	(0.033)
month_08	0.934^{***}	0.754^{***}
	(0.020)	(0.019)
month_09	1.051^{***}	0.701^{***}
	(0.038)	(0.023)
month_10	0.621^{***}	0.431^{***}
	(0.051)	(0.026)
month_11	0.318^{***}	0.184^{***}
	(0.062)	(0.034)
weekday_Friday	-0.217^{***}	-0.077^{***}
	(0.025)	(0.020)
weekday_Saturday	-0.154^{***}	-0.035^{*}
	(0.027)	(0.019)
weekday_Sunday	-0.184^{***}	-0.075^{***}
	(0.026)	(0.019)
weekday_Thursday	-0.303^{***}	-0.125^{***}
	(0.027)	(0.022)
weekday_Tuesday	0.055^{**}	-0.010
	(0.023)	(0.020)
weekday_Wednesday	-0.180^{***}	-0.071^{***}
	(0.026)	(0.021)
Constant	6.541^{***}	3.527^{***}
	(0.085)	(0.091)
R^2	0.570	0.722
Adjusted R^2	0.568	0.722
F Statistic	468.125^{***}	812.646^{***}
Note:		$^{*}p<0.1^{**}p<0.05^{***}p<0.01$

Now we could test the hypothesis about the difference between the two markets. We run two similar models on primary and secondary markets Table 5. As we could see, the coefficients are similar, and we could not reject the null hypothesis that the markets are different, because only 54% of coefficients are not overlapping on their confidence interval.

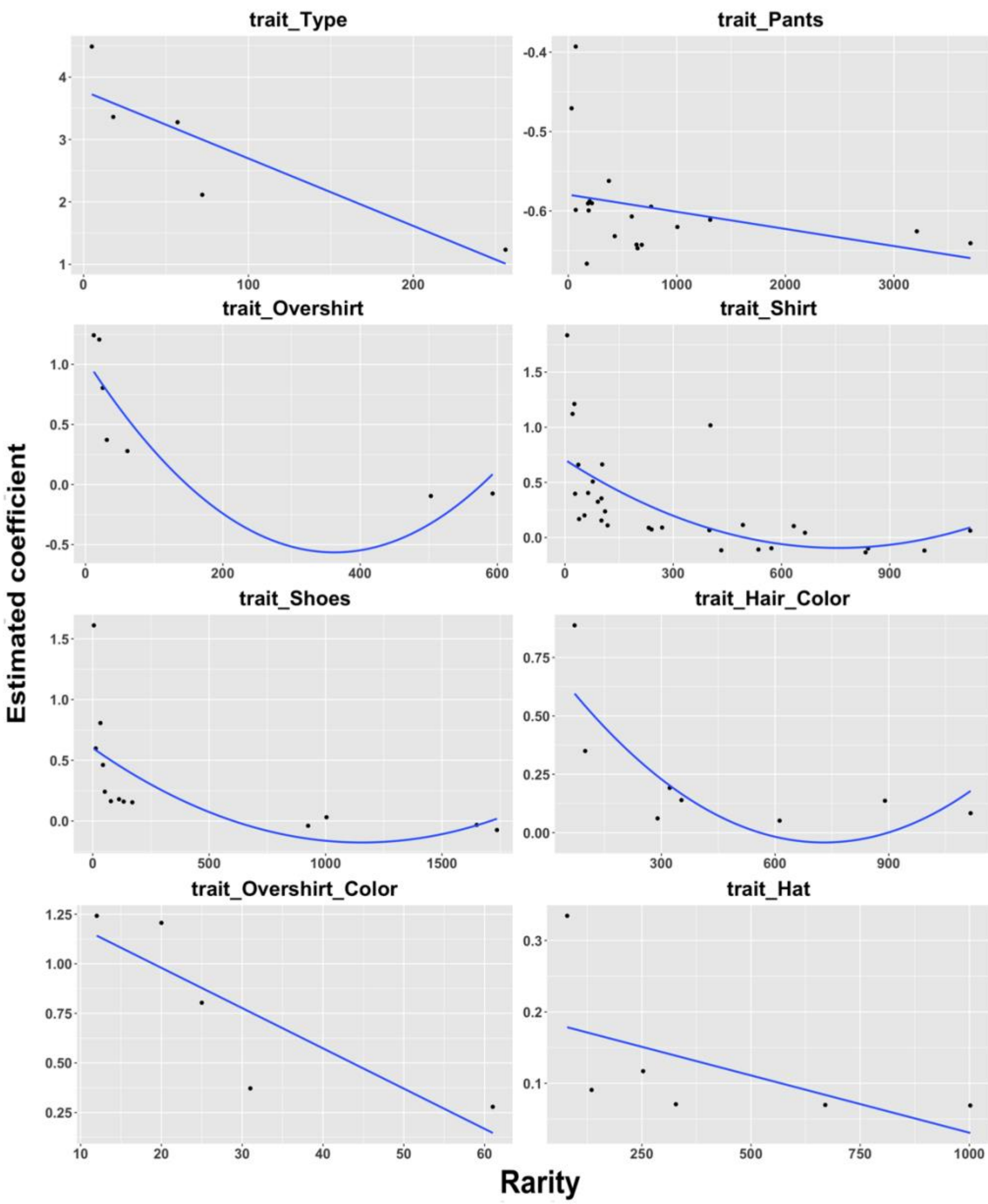

Figure 6. The rarity impact on prices for various trait groups

But we could see, that for some determinants the difference between the primary and secondary markets is significant. For example, in the day of the week or month effect. In general, the coefficient for the time effect in primary markets is higher than for secondary. We could make an assumption that it happens because the market is not mature enough, so could be influenced by external factors. Also, the whitelisting effect differs. For the primary market, the whitelisted NFTs recived less price, and the more free tokens the seller have, the less the price. But for the secondary market, the whitelisted tokens are pricier when other, minted with ETH.

Returning to the main price determinants, we already know that the rarity of tokens influences prices nonlinearly. But we could estimate this relationship for some particular trait groups. I extracted the betas from the model, grouped them by the trait group, and fitted the line. We have the betas on the y axis and score, which represent the rarity of the trait on the x axis. The smaller the score, the more rare the trait. So, in the end, we see, that the impact of a rarity on prices is nonlinear.

In the previous model, we estimated using ratting as a rarity aggregate, whet return on prices is diminishing. But in some traits group, we see that it's not always the case. In the example of Hat, we see that the influence of all tait types except the rarest one is quite constant. But the rarest Hat has 3 times higher an impact than others. I conclude that It's evidence of premium for "super rarity".

Chapter 6

CONCLUSIONS

The main purpose of the research was to attract the attention of other scholars to this emerging field with a tremendous total value locked. As was emphasized above, most of the works conducted so far were concentrated on some macro aggregates of the field. Thus, I decided to concentrate on the micro-level, the level of individual collection, and study it thoroughly. A lot of time was dedicated to the data gathering, in order to acquire the most complete dataset which will contain token level, user-level, and market-level variables.

We have studied several hypotheses and found various price determinants in this market. First of all, unlike other studies suggest, NFT and the crypto market are heavily linked together. We have found evidence for about a 0.1 percentage point increase in token prices in response to the increase in crypto prices by 1 dollar. Which is a lot considering the volatility of crypto. This could be the topic of future research, on how to use different NFT collections as a hedge against crypto market volatility. But the main determinant of the algorithmically generated token price was its rarity.

Although this market has a quite unique approach to rarity, it's still possible to measure. We have found, using both the discrete variables approach and using an aggregate, that the token price heavily depends on rarity. We have also found evidence for super rarity. In a particular attributes group (for example "Hat") the price depends constantly on the hat type, but for the rarest hat, the coefficient is increased by 3 times. Also, I calculated and estimated several privately created indexes for tokens rarity, and found out, what they are a good proxy for token individual characteristic. This hypotesisi tested unsing this variables suggested the deminishing return in rarity, thich is also goes according to the literature on the physicall colectibles. As the result, we could conclude that buyers value digital collectibles much like physical collectibles. The same behaviour was found in the reserch on Kripto Kitties, so it could be the feature of market as a whole.

This makes the market a perfect place for studying the “Snob effect” or the people's preferences towards rarity. As was mentioned above, the previous researchers struggled due to the issues connected with data gathering and intransparent in the physical world, and digital collectibles could solve them. Also, we studied the differences between primary and secondary markets, and found out, that even though we can’t reject the hypothesis about their similarity, we still could highlight some distinct characteristics. First of all, the primary market is more vulnerable to the various time effects. Secondly, the dependents between the price of the token and its previous price are quite significant.

Also, the whitelisting mechanics is quite crucial to price formation. On the primary market, the whitelisted tokens tend to be sold at 4 percentage points cheaper than the non whitelisted. This is happening due to people, who have received them for free dumping the market in order to gain more profit. This effect is increasing with the number of tokens "dropped" to the person. But at the same time, the users from the whitelisted group tend to pay for tokens by 14 percentage points more than their colleagues. This will be very helpful for future collection creators because the most vulnerable part of the NFT project is its launch (Hasu and Agnihotri 2021) but the main instrument of marketing in the field is still dropping. Thus, as we fund, the collection that whitelists too many tokens tend not only to oversupply but also triggers some behavior patterns of their audience which could have some potential problems in the future.

Finally, we incorporated some network techniques and studied the dependence between price and the buyer's inbounds and outbounds edges. As a result, the more buyer sells, the more it tends to underpay. And from the other hand, the more buyers buy, the more he tends to increase the price. This relationship is not very solid and should be studied further using some behavior or game-theoretic model, but it's still interesting to consider.

The hypotheses tested in this paper are barely scrunching the surface, but I believe will show, how many important behavioral and social phenomena could be studied using "overpriced pictures on the Internet, which could be right-clicked and downloaded buy any time".

WORKS CITED

Ante, Lennart, Non-fungible Token (NFT) Markets on the Ethereum Blockchain: Temporal Development, Cointegration and Interrelations (August 13, 2021).

Cameron, S., and Sonnabend, H. 2020. "Pricing the Groove: hedonic equation estimates for rare vinyl records." *Applied Economics* 52:50, 5516-5530.

Casale-Brunet, S., Ribeca, P., Doyle, P., Mattavelli, M., Networks of Ethereum Non-Fungible Tokens: A graph-based analysis of the ERC-721 ecosystem (October 24, 2021).

Dowling, M. 2021. "Fertile land: pricing non-fungible tokens." Finance Res. Lett. forthcoming, https://doi.org/10.1016/j.frl.2021.102096.

Ginsburgh V.A., Mei, J., and Moses, M. (2006), The computation of price indices, in Ginsburgh, V.A., Throsby, D. (eds.), *Handbook of the Art and Culture*, North-Holland.

Hasu and Agnihotri, A., "A Guide to Designing Effective NFT Launches," Paradigm, October 13, 2021, https://www.paradigm.xyz/2021/10/a-guide-to-designing-effective-nft-launches/

Kireyev, Pavel and Lin, Ruiqi, Infinite but Rare: Valuation and Pricing in Marketplaces for Blockchain-Based Nonfungible Tokens (October 20, 2021). INSEAD Working Paper No. 2021/60/MKT.

Koford, K., and Tschoegl, A. 1998. "The market value of rarity," *Journal of Economic Behavior & Organization* 34, 445-457.

Kong, De-Rong and Lin, Tse-Chun, Alternative Investments in the Fintech Era: The Risk and Return of Non-fungible Token (NFT) (August 30, 2021).

Lee, Yeonjoon, Measuring the Impact of Rarity on Price: Evidence from NBA Top Shot (August 16, 2021).

Modesta Masoit, Dappradar, https://dappradar.com/blog/how-to-value-the-meebits-nft-collection (accessed Oct. 13, 2021).

Mohammad Amin, F., Ali O., Mohammad, Taesiri Mohammad R., Under the skin of foundation NFT auctions (September 28, 2021)

Nahm, Joonwoo. "Price Determinants and Genre Effects in the Korean Art Market: A Partial Linear Analysis of Size Effect." *Journal of Cultural Economics* 34, no. 4 (2010): 281–97. http://www.jstor.org/stable/41811061.

Rengers, M. and Velthuis, O. (2002). "Determinants of prices for contemporary art in dutch galleries," 1992–1998. *Journal of Cultural Economics*, 26(1):1–28.

Renneboog, L., and Spaenjers, C. (2012) "Buying Beauty: On Prices and Returns in the Art Market." *Management Science*, Vol. 59, No. 1, 2013.

Scorcu, A., and Zanola, R. (2010), "The 'Right' Price for Art Collectibles. A Quantile Hedonic Regression Investigation of Picasso Paintings", Working Paper series, *Rimini Centre for Economic Analysis*.

APPENDIX

THE NFT INDIVIDUAL TRAITS RARITY

Table 6. Individual trait rarity

Trait name	Occurrence (out of 20000)
Body type	
Human	18881
Elephant	256
Skeleton	57
Pig	711
Robot	72
Visitor	18
Dissected	5
Shoes Color	
Black	2992
Gray	2764
White	2198
Yellow	1004
Magenta	835
Red	833
Green	747
Purple	893
Pants	
Skirt	2384
Leggings	2532
Regular Pants	3293
Ripped Jeans	1948
Trackpants	677
Cargo Pants	2868
Short Leggings	2418
Athletic Shorts	3379
Suit Pants	500

Trait name	Occurrence (out of 20000)
Shoes	
Skater	1701
High Tops	925
Sneakers	1649
Canvas	3808
Neon Sneakers	799
Urban Boots	1106
LL 86	33
Workboots	1590
Basketball	1667
Classic	1736
Running	639
Sandals	1066
Slides	1764
High Boots	684
LL Baby Blue	225
LL Orange	133
LL Moonboots	44
LL Retro	78
LL Tall	113
LL RGB	52
LL High Tops	170
LL Alien	13
No Shoes	5

Table A.1 - Continued

Trait name	Occurrence (out of 20000)	Trait name	Occurrence (out of 20000)
Shirt		Flamingo Tee	37
Diagonal Tee	753	Snoutz Jersey	64
Oversized Hoodie	1123	Punk Tee	26
Halter Top	467	CGA Shirt	101
Hoodie Up	412	Glyph Shirt	6
Logo Tee	996	Pants Color	
Ghost Tee	833	Camo	3210
Classic Jersey	536	Dark Gray	3704
Tee	1123	Luxe	585
Invader Tee	1792	Denim	3783
Windbreaker	891	Leopard Print	374
Hoodie	2107	Blue Camo	1005
Meepet Tee	488	Green	182
Suit Jacket	493	Posh	428
Suit	1212	Dark Red	1307
Jersey	400	Gray	638
Tube Top	572	Yellow	171
Heart Hoodie	886	Black	763
Bare Chest	403	White	629
Tie-dyed Tee	121	Purple	218
Basketball Jersey	433	Magenta	188
Skull Tee	1972	Red	197
Long-sleeved	840	Green Plaid	69
Lines	232	Argyle	31
Snoutz Tee	269	Red Plaid	69
Hawaiian	91	Beard Color	
Stylized Hoodie	101	Dark	3341
Snoutz Hoodie	39	Blond	450
Snoutz Skull Tee	103	Silver	72
Heart Tee	77	Brown	453

Table A.1 - Continued

Trait name	Occurrence (out of 20000)
Beard	
Full	1449
Biker Mustache	129
Stubble	1585
Big	369
Mustache	628
Muttonchops	156
Medical Mask	113
Shirt Color	
Black	3116
Camo	665
Gray	2732
Red	749
Purple	815
Yellow	844
White	1970
Green	634
Posh	28
Green Plaid	285
Magenta	813
Blue Camo	118
Red Plaid	240
Leopard Print	21
Argyle	111
Luxe	54
Overshirt	
Collar Shirt	1337
Jean Jacket	316
Athletic Jacket	503
Trenchcoat	593

Trait name	Occurrence (out of 20000)
Leather Jacket	461
Overshirt Color	
Argyle	20
Gray	239
White	180
Magenta	188
Black	227
Red	165
Green	118
Yellow	194
Camo	127
Green Plaid	61
Posh	12
Purple	187
Red Plaid	66
Luxe	25
Blue Camo	31
Glasses	
Aviators	1057
Round Glasses	1299
Sunglasses	1083
Elvis	1374
Nerdy	966
Frameless	1339
Specs	592
3D	120
Glasses Color	
White	1220
Charcoal	1232
Dark Red	405

Table A.1 – Continued

Trait name	Occurrence (out of 20000)
Hair Style	
Very Long	731
Pulled Back	758
Half-shaved	39
Simple	2570
Long	773
Messy	664
Wild	1400
Buzzcut	3277
Fade	1407
One Side	360
Ponytail	789
Spiky	1345
Bald	1752
Bob	389
Straight	774
Bun	235
Curly	732
Big Bangs	760
Mohawk	643
High Flat Top	404
Fiery Mohawk	41
Pigtails	157
Necklace	
Gold Necklace	1577
Gold Chain	2161
Earring	
Gold Hoops	401
Gold Earrings	782
Gold Earring	859

Trait name	Occurrence (out of 20000)
Hair Color	
Blonde	612
Auburn	322
Dark	11709
Light Blue	353
Brown	1795
Dyed Red	740
Blond	1116
Silver	290
Bleached	890
Purple Dye	210
Blue	71
Rainbow	99
Hat	
Cap	1002
Headphones	1285
Wool Hat	670
Bandana	166
Trucker Cap	328
Backwards Cap	664
Brimmed	135
Snoutz Cap	79
Hat Color	
Purple	275
Black	920
Red	236
Camo	229
Gray	863
Yellow	298
White	745
Magenta	253

Table A.1 – Continued

Trait name	Occurrence (out of 20000)
Jersey Number	
0	140
1	129
2	133
3	168
4	146
5	145
6	142
7	133
8	148
9	149

Trait name	Occurrence (out of 20000)
Hat Color	
Green	209