DESCRIPTIVE AND PREDICTIVE ANALYSIS OF NFT MARKET

A Thesis

by

Onur Can Çabuk

Submitted to the
Graduate School of Sciences and Engineering
In Partial Fulfillment of the Requirements for
the Degree of

Master's Degree

in the
Department of Data Science

Özyeğin University
Jan 2023

DESCRIPTIVE AND PREDICTIVE ANALYSIS OF NFT MARKET

Approved by:

Asst. Prof. Erinç Albey, Advisor
Dept. of Industrial Eng.
Özyeğin University

Asst. Prof. Mehmet Önal
Dept. of Industrial Eng.
Özyeğin University

Prof. Mehmet Güray Güler
Dept. of Industrial Eng.
Yıldız Technical University

Date Approved: 29 Dec 2022

To my family and friends who supported me throughout my journey.

ABSTRACT

Non-fungible tokens (NFTs) are digital assets on a blockchain that have unique identification codes and metadata that make them distinguishable from one another. NFTs can represent a wide range of digital assets, including game cards, artwork, and even real estate. Due to these characteristics, NFTs have gained a tremendous interest from people around the world, leading to huge returns on investment in the NFT market. However, there are only a few studies on the market in the literature.

This paper examines various aspects of the NFT market to shed light on its dynamics and wallet behaviors. First, a descriptive analysis of the market is performed to show its overall trend. The transactional behaviors of wallets are then analyzed, and a segmentation is made to gain a general understanding of the user portfolio. The buyers of a specific NFT collection (Bored Ape Yacht Club) are then studied by comparing them to the overall market, revealing differences in transactional tendencies and macro indicators. Finally, machine learning models are developed to predict the transactional behaviors of wallets.

Our analysis has revealed that the growth of the NFT market is largely driven by new entrants to the market, but lately there has been a significant decrease in the number of new wallets entering the market. We have also found that the majority of wallets in the market have only one transaction and hold only one token, suggesting that these are users who are experimenting with the market. When we look at the Bored Ape Yacht Club sample, however, we see that these users are highly engaged with the market, with high trading frequencies and a diverse portfolio. Finally, our predictive models show that the transactional behaviors of wallets can be predicted, which opens up opportunities for optimization in various areas.

ÖZETÇE

Nitelikli-Fikri Tapular (NFT'ler), benzersiz tanımlama kodlarına ve onları birbirinden ayıran meta verilere sahip, bir blok zincirindeki kripto varlıklardır. NFT'ler, bir oyun kartını, sanat eserini ve hatta fiziksel bir emlağı temsil edebilir. Bu özellikleriyle, NFT'ler kısa sürede büyük ilgi görmüş ve yüzlerce kat yatırım getirisi sağlamıştır. Ancak, literatürde NFT pazarı ile ilgili yalnızca bir kaç araştırma mevcuttur.

Bu çalışma, NFT pazarını farklı yönlerden inceleyerek pazar dinamiklerine ve cüzdan davranışlarına ışık tutmaktadır. İlk olarak, piyasanın genel eğilimini göstermek için piyasanın tanımlayıcı bir analizi yapılmıştır. Ardından, cüzdanların işlemsel davranışları analiz edilmiş ve portföy hakkında genel bir anlayışa sahip olmak için bir segmentasyon yapılmıştır. Sonrasında, belirli bir NFT koleksiyonunun (Bored Ape Yacht Club) alıcıları, genel pazarla karşılaştırılarak analiz edilmiş ve makro göstergelerin yanı sıra işlem eğilimlerindeki farklılıklar göz önüne serilmiştir. Son olarak, cüzdanların işlemsel davranışlarını tahmin etmek için makine öğrenimi modelleri geliştirilmiştir.

Sonuç olarak, pazarın işlemsel büyümesinin büyük ölçüde pazara girişlerle desteklendiğini ve son dönemde pazara giriş yapan cüzdan sayısında önemli bir düşüş olduğunu keşfedilmiştir. Ek olarak, piyasadaki cüzdanların çoğunun sadece bir işlem yaptığı görülmüştür bu da bunların piyasada keşif yapan kullanıcılar olduğunu göstermektedir. Ancak Bored Ape Yacht Club örneklemi incelendiğinde, bu kullanıcıların yüksek alım satım sıklığına ve geniş bir portföye sahip olan, piyasa ile yakından ilgilenen kullanıcılar oldukları görülmüştür. Son olarak, tahmine dayalı modeller, cüzdanların işlemsel davranışlarının tahmin edilebileceğini kanıtlamaktadır ve bu da pazarlama kampanyalarının vb. optimizasyonu için bir alan yaratmaktadır.

ACKNOWLEDGEMENTS

I would like to begin by expressing my gratitude to my advisor, Erinç Albey, for his support and guidance throughout my education in data science. His expertise and experience in the field were invaluable in helping me conduct my research. It has been an honor to study under his guidance. I am also grateful to my family, particularly my mother Hatice Çabuk, my father Bahri Çabuk, and my sister Zeynep Çabuk, for their love and support. I would also like to thank my friends for their support throughout my life. Lastly, I would like to thank those who are no longer with us for their contributions to who I am today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Motivation and Contributions

Over the last decade, cryptocurrencies have gained significant interest and have established a multi-billion dollar market. In 2021, the crypto market entered a bull market, which led to a sudden surge of interest in non-fungible tokens (NFTs) from around the world. This resulted in huge returns on investment for many people. In January 2021, the total market value of the NFT market was around $ 70 million, and by January 2022, it had reached an all-time high of $35 billion [1]. The Ethereum blockchain dominates the NFT market, but there are other blockchain networks like Flow, Polygon, and WAX that also support the creation and trading of NFTs [2]. Despite the huge market value, there is a lack of research on the NFT market, with most existing studies focusing on the underlying NFT protocols.

The Ethereum blockchain data is publicly available for anyone who wants to take it, however, the data structure of it is pretty complex. The input data of a transaction is encoded in hexadecimal format and in order to make sense of the input data one should first properly decode the data. However, this decoding process is only possible through using the ABI(Application Binary Interface) decoder of the contract that is subject to the transaction. Given that there are thousands of different NFT contracts in the chain, it is challenging to decode all types of transactions. Also, the decoded data of contracts have different data structures which make it even harder to obtain the data in a structured way.

Along with having a general understanding of the crypto world, the above-mentioned traits of the blockchain data create an entry barrier in terms of analyzing the NFT market, which justifies the mere number of studies in the area despite the multi-billion dollar market value. In our study, first, we will obtain the blockchain data in an efficient way as described in Chapter 4. We will then use this data to conduct both descriptive and predictive analyses of the NFT market to better understand the market dynamics.

In the descriptive part of our analysis, we will begin by examining the overall growth of the NFT market to understand it's drivers. We will then focus on wallets, which can be thought of as individual customers, and analyze their transactional behavior to better understand their characteristics and segment them based on their transactional patterns. Finally, we will focus on a specific NFT project called Bored Ape Yacht Club, and compare the user portfolio for this project to the overall market, using both macro indicators and transactional behaviors. This will provide valuable insights into the market and the behavior of experienced users.

The predictive part of our analysis will focus on wallet analytics and will be divided into two main categories: market-level metrics and project-level metrics. We will propose a set of metrics that can be used by actors in the NFT market (such as NFT marketplaces, NFT creators, and crypto lenders) to predict the future transactional behavior of wallets. These metrics will provide valuable insights into the market, allowing these actors to make more informed decisions. We will then use machine learning models to predict these metrics and present the results of our analysis.

1.2 Organization of the Thesis

The rest of the thesis is organized as follows: Chapter 2 provides a literature review of the relevant topics, including blockchain technology, NFTs, and the methodologies used in the study (XGBoost and resampling methods). Chapter 3 describes the problem in detail, and Chapter 4 explains the data architecture and the process for obtaining the data. Chapter 5 presents the results of the descriptive analyses of the NFT market, and Chapter 6 describes the predictive modeling strategy and presents the results of the predictive analysis. Finally, Chapter 7 offers comments and discussion on the results and potential future work.

CHAPTER II

LITERATURE REVIEW

2.1 Blockchain

In the recent years, cryptocurrencies, especially Bitcoin, became widely adopted all over the world, and the Blockchain technology is the core foundation of crpytocurrencies. Blockchain is defined as a public ledger in which all the transactions are stored in blocks of chain and the chain grows every time when a block is appended to it [3]. Blockchain has four key characteristics; decentralization, persistency, anonymity and auditability. These core features allow Blockchain to save cost and increase efficiency [4].

- Decentralization

  Decentralization feature allows a transaction to be conducted between two peers(P2P) thus removing the central agency (i.e. banks) which can significantly reduce server costs.

- Persistency

  Every transaction should be confirmed and recorded in distributed blocks over the network, so it is nearly impossible to hack or cheat the system.

- Anonymity

  Users can use the network via a generated address with no personal or private information. And it is also possible for a person to create multiple addresses to avoid identity exposure.

- Auditability

Given that each transaction is validated and recorded with a datetime, anyone can easily verify the record and trace the previous records by accessing any of the nodes in the network.

Although Blockchain is well known because of cryptocurrencies, it's implication areas are not limited to cryptocurrencies. It also has other uses cases including digital assets, online payments [5], smart contracts [6], public services [7], internet of things [8], reputation systems [9] and security services [10].

Even though the Blockchain technology has a great potential to become the backbone of the future internet systems, it has had its unique challenges. Firstly, scaling is a huge problem for Blockchain, block size of bitcoin network is limited to 1 megabyte and a new block is mined in the network in about 10 minutes [11]. Despite the fact that Ethereum network is faster, it is still not fast enough to deal with high frequency trading. Secondly, consesus algorithms like proof of stake (PoS) and proof of work (PoW) are facing serious challenges. PoW networks are inefficient in terms of electricity consumption [12] which is a big issue in the modern era as the scientists urge to give up on fossil fuel for the sake of environment. Ethereum's new PoS network requires validator nodes to stake 32 ethereums and given that ETH's price is well above thousand dollars, many of the users who is interested in becoming a validator wouldn't be able to invest that much money. This issue is similar to the challenges users had with proof-of-work, where only loaded users had better chances to increase their mining success rate. Also, users staking higher amounts are the ones who get better chances of getting chosen to become validators and earn rewards [13] which could mean that the rich would get richer. Therefore, these challenges should be addressed to make Blockchain more reliable in the future.

2.2 NFT

NFT(Non Fungible Token)'s are cryptographic assets in blockchain, derived from smart contracts. NFT tokens have unique identification codes and metada that distinguish them from each other. However crpytocurrency tokens are indistinguishable and equivalent which makes them a suitable medium for financial transactions. In contrast, NFT's are unique which cannot be exchanged like-for-like, making it convenient for identifying someone or something [14]. To give an example, by using NFTs, a creator can easily prove the existence and ownership of a digital asset in the form of videos, images, arts [15], event tickets etc. Additionally, the creator of NFT can also earn commissions every time a successful trade happens. Because of this, NFT concept draw huge attention from the art world, especially digital creators which led to investors to have an interest in the market. With overall crpyto market entering the bull market in late 2020, NFT market has also seen extraordinary growth, becoming the most popular Fintech application and crypto asset in 2021 [16]. According to Nansen, a crypto analytics company, the market cap of NFT is around 11.3 billion dollars [17]. While some NFTs are traded on low value some are really expensive, i.e. an artist known as Beeple sold his own artwork as an NFT for 69.3 million dollarson March 2021 [18]. There are also some NFT projects traded at high value, CryptoPunks [19] CryptoKitties [20], Bored Ape Yatch Club [21]. However the NFT market is still immature and volatile. With the crypto crash that happened in 2022 when The Federal Reserve, the central bank of the United States, started to hike the interest rates to fight with inflation, NFT market experienced a massive drop, as did Bitcoin, the flagship of cryptocurrencies.

2.3 Predictive Methods: XGBoost

Tree boosting algorithms are the algorithms that ensemble the outputs of the weak decision tree learners to obtain a superior learner. The performance of the superior learner is dependent on two main things; the data and the weak learner. Insufficient data and noise in the data can cause the ensemble learner to perform poorly in terms of predictive capability. However, tree boosting algorithms can produce reasonably good results on a wide range of problems with sufficient and well-constructed datasets [22]. XGBoost (Extreme Gradient Boosting) is an open-source tree boosting machine learning algorithm that can be used in both regression and classification problems. Kaggle, a website that holds competitions for machine learning problems, published that out of 25 competitions hosted in 2015, XGBoost was the winning algorithm in 17 [23] therefore it's success is well recognized in the machine learning community. In addition to it's predictive performance, it's short execution speed, thanks to parallel processing using multiple CPUs, [24] makes it a viable candidate for any problem that have a tabular dataset. All these features, make XGBoost a proper benchmark for machine learning problems.

2.4 Sampling Methods

Imbalanced data in Machine Learning refers to an unequal distribution of classes within a dataset [25]. This issue happens mostly in classification tasks in which the labels in a given dataset are not uniformly distributed. Learning algorithms that do not take class imbalance into account frequently become overwhelmed by the majority class and ignore the minority class [26]. For instance, in a dataset where the majority class constitutes 99% of all the data, a learning algorithm that maximizes the accuracy rate might choose to categorize all samples as the majority class in order to obtain a high accuracy rate of 99%. However, in this case, all minority class samples will be incorrectly classified. Therefore, in a setting where there is a high level

of class imbalance, imbalance must be carefully managed to build a good classifier that could classify the minority class correctly [27]. Cost-sensitive learning, another significant problem in machine learning, is closely related to class imbalance as well. Misclassifying an instance of a minority class is typically more serious than misclassifying an instance of a majority class. Consider approving a fraudulent credit card application, it is more costly to approve a fraudulent application to decline a credible one. There are two main methods to solve this problem; adding records to the minority class by resampling the minority class or deleting records from the majority class. Undersampling is the process of decreasing the amount of majority target instances. Some common undersampling methods are random sampling, tomeks' links [28], and cluster centroids [29]. Oversampling can be performed by increasing the amount of minority class instances or samples by producing new instances or repeating some instances. Some common oversampling methods are random oversampling, SMOTE [30] and Adasyn [31].

CHAPTER III

DETAILED PROBLEM DESCRIPTION

The NFT market is relatively unexplored, there are few studies conducted on NFTs and fewer on the NFT market. This is partly due to the fact that blockchain data is different from traditional transactional data and must be decoded to be understood by humans. Hence some background research is needed to have an understanding of the data. This need for background research creates an entry barrier for those who want to analyze the market, therefore the dynamics of the market are still largely unknown. In our study, we will focus on three main aspects of the NFT market.

- Firstly, we will explore the NFT market, and understand the growth that has happened there by analyzing the major indicators of the market.

- Secondly, we will study the user portfolio, and examine transactional behaviors and tendencies of the wallets. We will then perform a segmentation on the customer portfolio by using the findings of our analysis.

- Lastly, we will focus on just one NFT project, Bored Ape Yacht Club(BAYC). We will examine the overall BAYC sample by looking at macro indicators and comparing it to the overall NFT market. Moreover, we will analyze the wallets of this sample to understand the transactional tendencies of the BAYC buyers by comparing them to the overall market. Finally, we will go into details of wallet analytics to explore the opportunities there. We will propose set of metrics that would help actors in the field, such as NFT marketplaces, creators, and lenders, to take action by either assessing the risk of the customers or the probability of buying a token and build predictive models to help solve these problems.

3.1 NFT Market

Even though the history of the NFT dates back to 2015, the market did not experience remarkable growth until 2021. The market started to flourish in early 2021 and experienced exceptional growth around mid-2021, and had consistent growth until the recent past. However, there are no studies regarding the NFT market to understand the growth and wallet behaviors. The transactional growth of the NFT Market is driven by two main factors. The first one is the wallets trading in the market. Wallets are like bank accounts that are required for a user to make transactions therefore wallets can be treated as customers or users. The second factor is the NFT projects that are traded in the market. A project is a collection of unique NFT tokens which are governed by the same contract address. In our study, we will first examine the growth by looking into transactions, wallets, and projects from different aspects.

3.2 Wallet Analysis

In the next part of our analysis, we will examine the behavior of wallets in the NFT market. As mentioned before, wallets can be considered as customers, so we will approach them from that perspective and study their transactional habits. We will explore different aspects of individual wallets to gain insights into their characteristics and identify potential opportunities to segment them based on their behavior.

3.3 Collection Analysis

After completing the analysis of the wallets, we will turn our attention to Bored Ape Yacht Club, one of the major projects in the NFT market, if not the biggest. The reason why we chose the Bored Ape Yacht Club is that it is one of the pioneer NFT projects in the market and its price is over $100,000, which indicates that the buyers of this collection are heavily interested in the NFT market. First, we will compare the general trends of the sample to the NFT market by using macro indicators.

Then, we will examine the data using exploratory data analysis techniques to understand the transactional behaviors of individual wallets and compare them to the overall market to gain an understanding of how experienced users behave. We will then propose a set of metrics that can help actors in the NFT field, such as NFT marketplaces, NFT creators, and crypto lenders, make informed decisions. The proposed metrics can help them assess the risk of a wallet or the probability of buying a token, which would allow them to take the appropriate actions. Finally, we will build machine learning models to predict these metrics, which would provide solutions to the challenges faced by these actors in the market.

CHAPTER IV

DATA ARCHITECTURE

Blockchain data is publicly available for anyone who wants to take it, however, as stated earlier, one of the main reasons why the NFT market is not being analyzed in the literature is that there is some background research needed to understand the overall blockchain data structure. Also, the raw data needs to be decoded first from hex decimal format to human readable values. Therefore, even though there are brilliant scientists all over the world, most of them are not able to analyze the data because of a lack of understanding of the data structure. At the start of the project, We've also tried to fetch the data from the publicly available Ethereum blockchain itself, however, the returned data from the API is hard to decompose. For a transaction in a blockchain, underlying data can only be decoded using the contract of the token, which creates difficulty and lack of organization in decoding the data. After researching possible ways to acquire blockchain data we've decided o use the Flipside's [32] API which offers access to already decoded blockchain data. However, the data returned from the API is raw and should be transformed in a meaningful way to have a dataset that is ready for predictive modeling. Apart from fetching transactional data and sales data using the Flipside API, we've also acquired different datasets from the internet by scraping web pages.

4.1 Architecture

We have various data sources that compose the final dataset. Therefore, customized data loading processes are needed to tackle each unique problem while acquiring the data. Mainly, Python [33] programming language is used while fetching data. Acquired data is then stored in a local MySQL [33] instance in the raw format. All the python scripts in the project are hosted in the GitHub [34] version controller.

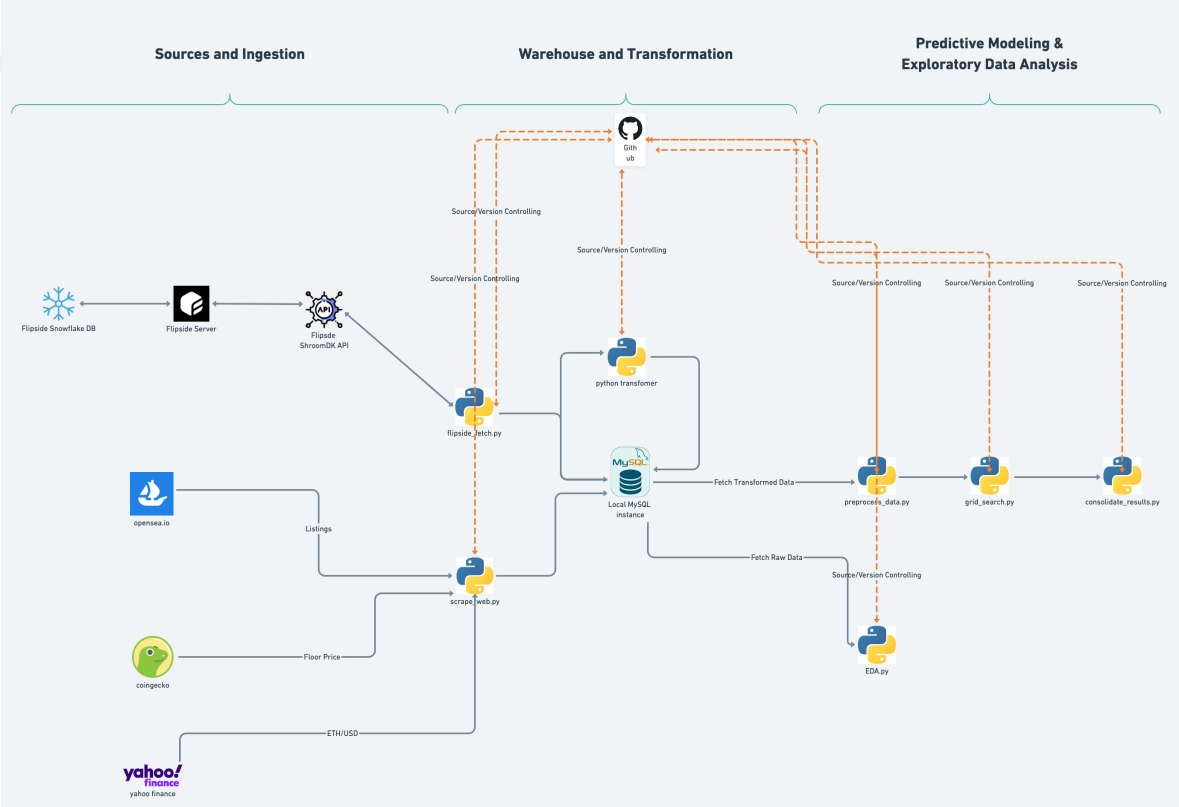

Figure 1: Data Architecture

After acquiring raw data, the data is transformed through a python transformer to obtain a meaningful dataset. Finally, raw and transformed data is fetched from the MySQL instance to perform exploratory data analysis and build predictive models.

4.2 Data Sources

As can be seen from the Data Architecture figure 1, we have various data sources for the final dataset. The main data source of the project is the Flipside API and the remaining sources are used to enrich the data.

4.2.1 Flipside

As mentioned earlier, Flipside is the main data source of the project. Flipside provides access to one of the most comprehensive blockchain data sets in all of crypto, covering dozens of blockchains, NFT marketplaces, and thousands of protocols. Flipside decrypts the blockchain data to transform it into a human-readable format and stores it in a database. It also offers an API service to query its database. The API requires an API key which can only be acquired through an NFT. During the data collection, we fetched two different tables from this data source using the API.

4.2.1.1 NFT Transactions Table

The first sourced table is the NFT Transactions table, it contains information regarding the NFT transactions, including but not limited to; transaction hash, event index, from wallet address of transaction, to wallet address of transaction, block number, event time, token id, contract address of the token, etc.

4.2.1.2 NFT Sales Table

Second sourced table is the NFT sales table, it contains data regarding the NFT sales that happened in the main NFT marketplaces; opensea [35], blur [36], looksrare [37], rarible [38], x2y2 [39], sudoswap [40], and nftx [41]. The data has information including but limited to; buyer wallet address, seller wallet address, nft contract address, nft tokenid, transaction hash, event timestamp, currency of the payment, contract address of the currency, token price in terms of the transaction currency, gas fee, platform fee, creator comission fee and the USD equivalents of these fees.

4.2.2 OpenSea

Open sea is world's first and largest digital marketplace for crypto collectibles and non-fungible tokens. One can buy, sell, and discover exclusive digital items. OpenSea has a listing feature for sellers, one can list their tokens for sale and get bids from the potential buyers. We collected token listings data of the Bored Ape Yatch Club (BAYC) collection from OpenSea by scraping the website. However we were not able to scrape the bidding data in a timely manner, which would help predictive models in terms of performance.

4.2.3 CoinGecko

CoinGecko provides a fundamental analysis of the digital currency market. In addition to tracking price, volume, and market capitalization, CoinGecko tracks community growth, open source code development, major events, and on-chain metrics. We've used CoinGecko to gather floor price of the BAYC collection. Data is collected by scraping the website. There are much more information in CoinGecko that could be used as features while building the predictive models, however we were not able to scrape them in a timely manner.

4.2.4 Yahoo Finance

Yahoo! Finance is a media property that is part of the Yahoo! network. It provides financial news, data and commentary including stock quotes, press releases, financial reports, and original content. It also offers some online tools for personal finance management. We've used Yahoo! Finance to collect the ETH/USD parity to used in our predictive models. Data is scraped through a python script.

4.3 Fetching Data

4.3.1 Flipside

We used Flipside's API to fetch the data, API access requires a key that can only be acquired through minting an NFT. However fetching the Flipside data is tricky, reason behind this is that API has a rate limit for requests, and also there is a limit on the number of rows that can be returned from the API. We have curated a fetching process that first checks the number of rows that would be returned by a query, and if this number is greater than the allowed return row limit, the script breaks the query into pieces by filtering the query using where clause and checks again until the returned rows for all the queries are within limits. As there is a rate limit on the API and sometimes the API throws an error, there are try, except and finally clauses on the python script to ensure that all the data is fetched without a problem and nothing goes unnoticed. After developing the script that is required to fetch the data, we have also created a process to store the data on the specified table in the local MySQL instance. After testing the whole process that fetches the data and inserts it into the database, we've collected the data for NFT transactions and NFT sales up until October 1st, 2022.

4.3.2 Scraping Websites

Given that there are multiple websites that need to be scraped. First we've created functions that would be required to scrape the website, then we've created a unique process for each website that is customized to fetch the data the in an efficient way. Python programming language is used to develop the script, and libraries used are including but are not limited to; Selenium, Requests, and BeatufilSoup. Finally, scraped data is then stored on the specified table in the local MySQL instance.

4.4 Transforming Data

Given that the data fetched from different sources are in a raw format, a transformation step is required to make the data ready for modeling. There are multiple steps of transformation for each data source. Transformation steps can be summarized as;

- First, each data set is cleaned and de-duplicated to ensure the data integrity along the process
- Secondly, there are sanity checks in process to ensure that the data is correct and meaningful.
- Thirdly, all the data sources joined to create a meaningful data set
- Finally, there are processes to create a historical data by following a snapshotting logic.

All the transformation steps are written and maintained in a python script.

CHAPTER V

DESCRIPTIVE ANALYSIS

5.1 NFT Market

As stated earlier, the NFT market experienced huge growth in 2021. We will explore what happened on the market by using exploratory data analysis techniques. Data up to October 1st, 2022 is covered in the study.

An NFT transaction has two main components; wallets making the transaction(sender and receiver) and an NFT token. Therefore, we will analyze the data by using three mediums; transactions, wallets, and NFT Projects. Firstly, we will analyze all of the NFT market without filtering anything and check the macro indicators there to understand the overall view of the market. Then we will filter the data based on projects to include only the NFT projects which are recognized by the market and analyze it to draw a more meaningful picture of the market.

5.1.1 Overall Market

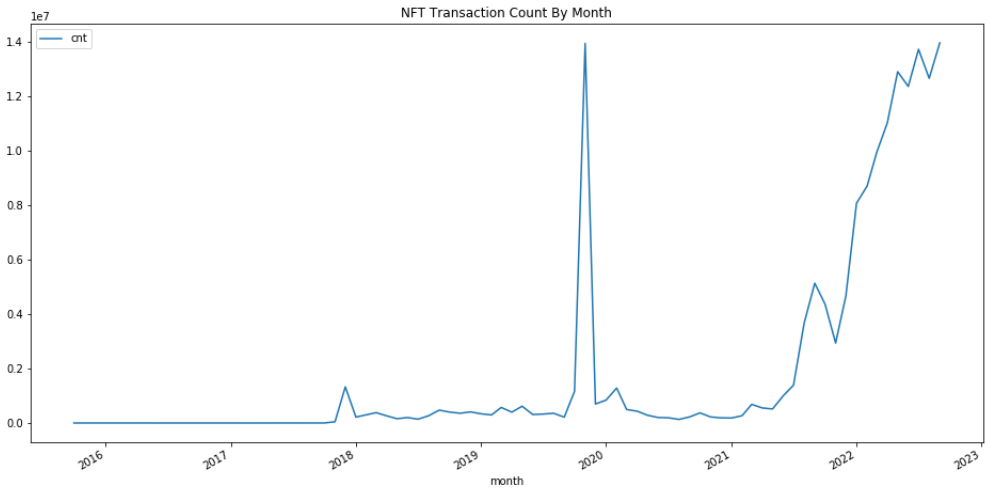

Figure 2: Monthly Transaction Counts of The NFT Market

As can be seen from the Figure 2, the NFT market has experienced tremendous growth in 2021 in terms of transaction count, even though the growth rate has decreased in the second half of 2022, the growth continues. There was a total of 182.726 transactions on January 2021, and the monthly transaction count jumped up to 8.071.166 on January 2022. From January 2022 to September 2022, the monthly transaction count nearly doubled to 14 Million. The reason behind the spike in November 2019 is the release of an NFT project named Gods Unchained in high quantities.

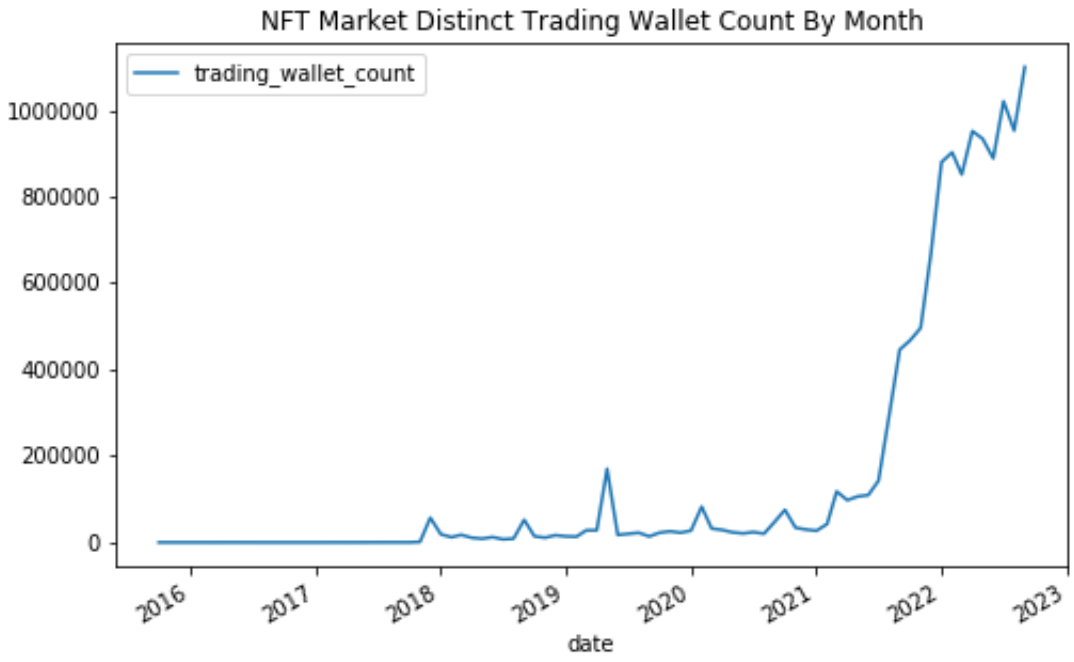

Figure 3: NFT Market Distinct Trading Wallet Count By Month

The same trend in Figure 2 is also present in Figure 3, which shows the monthly counts of the distinct wallets that are trading in the market. The number of wallets trading in the market was less than 100.000 in January 2021, and on September 2022, it was up to 1.000.000 which means a tenfold increase.

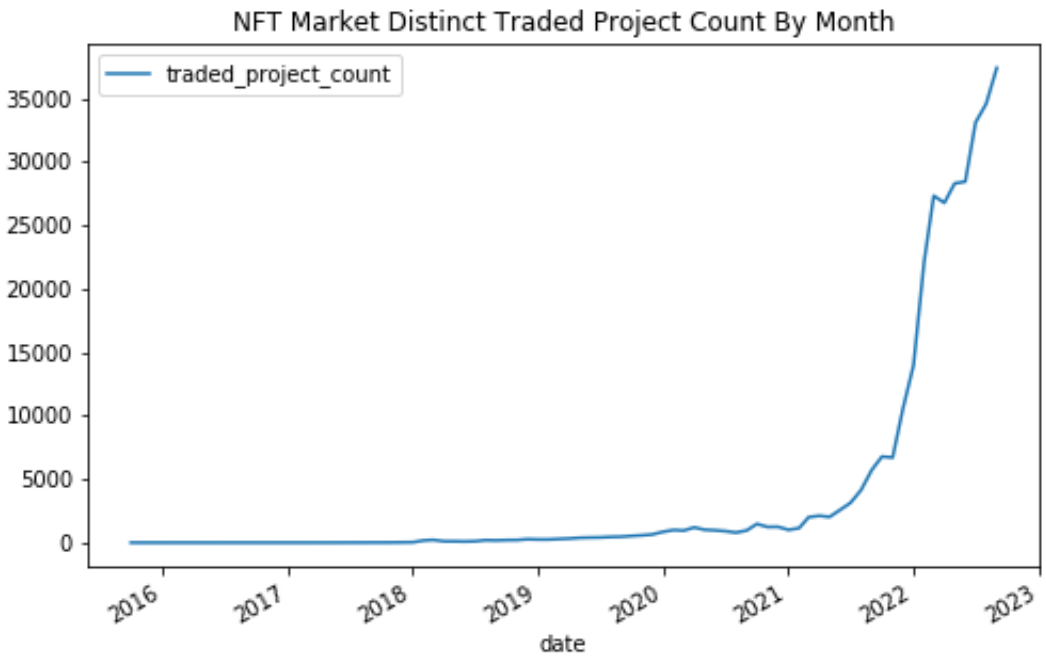

Figure 4: NFT Market Distinct Traded NFT Project Count By Month

The same trend in Figure 3 is also present in Figure 4, which shows the monthly counts of the distinct NFT projects that are traded in the market. However, the growth rate doesn't seem to be decreasing in 2022-H2 in Figure 4. This could mean that users that are still trading in the market are exploring opportunities with different NFT projects.

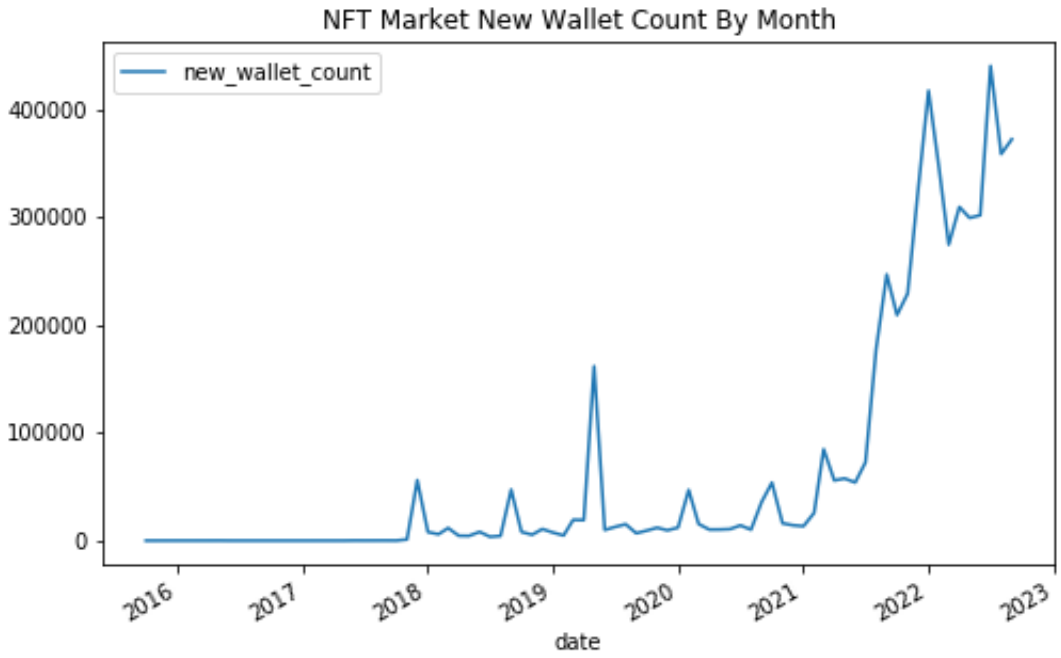

Figure 5: Number of Wallets Entering the NFT Market By Month

The same trend is followed in 2021 in Figure 5, which shows the number of wallets entering the market, it can be argued that the number of wallets entering the market is still growing in 2022, if we exclude the peak happened in January, 2022. We would also like to underline that the data that is present in Figure 3 includes the data in Figure 5, which means that nearly 40% of the wallets trading in the market are the ones that are entering the market.

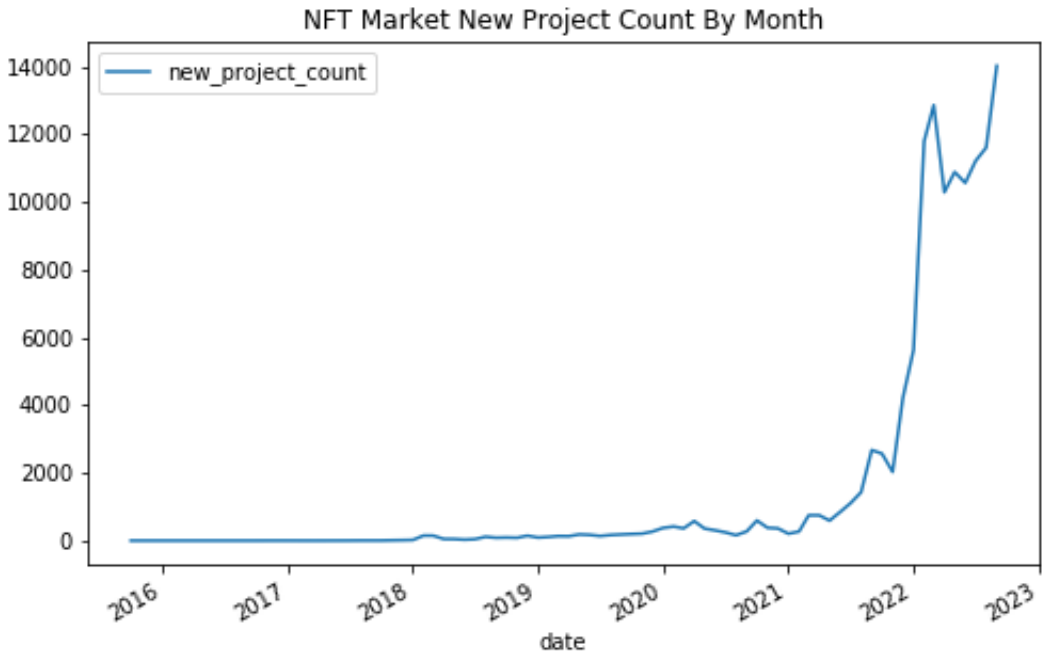

Figure 6: Number of NFT Projects Entering the NFT Market By Month

The same trend in Figure 5 is also present in Figure 6, which shows the number of NFT projects entering the market. There are around 12.000 new projects released each month and there seems to be growth happening in the last 3 months.

5.1.2 Market - Filtered on Relevant Projects

The NFT market seemed to be growing in 2022 in the previous section. However, these numbers could be misleading as there are many NFT projects with a low value. Therefore we should not take the projects that have a low value into account in order to perform a meaningful analysis. To eliminate these projects from the analysis we've used the following filters;

- A project should have at least 100 sales on the marketplaces.
- Mean price of these sales should be more than 10 USD.
- Median price of these sales should be more than 10 USD.

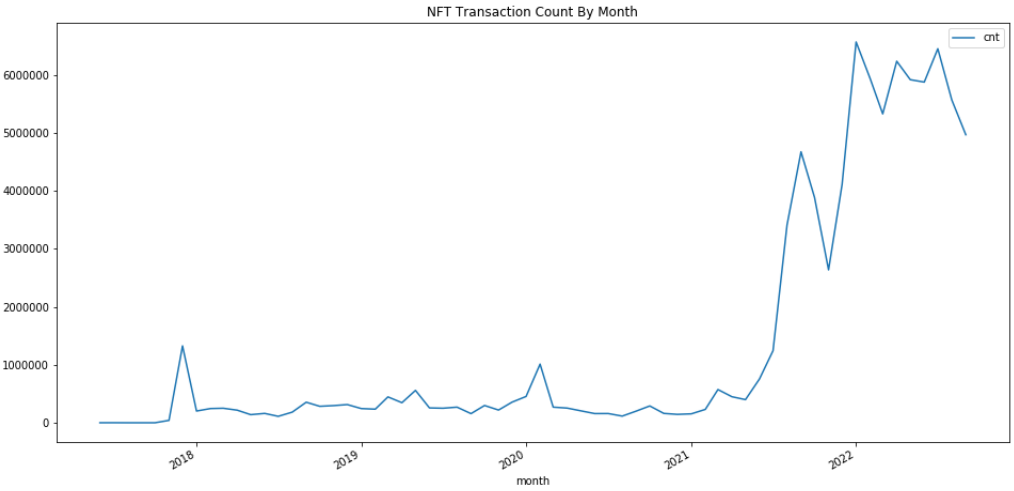

Figure 7: Monthly Transaction Counts of The Filtered NFT Market

As can be seen from Figure 7, the market experienced enormous growth in 2021, however, this growth stopped in 2022 and the transaction count stabilized. When compared with Figure 2 which shows the monthly transaction counts in the market without any filtration, it is obvious that the growth happening in 2022 is because of the projects that have a low value. I would also want to underline that approximately 60% of the transactions are coming from NFT projects that have a low value.

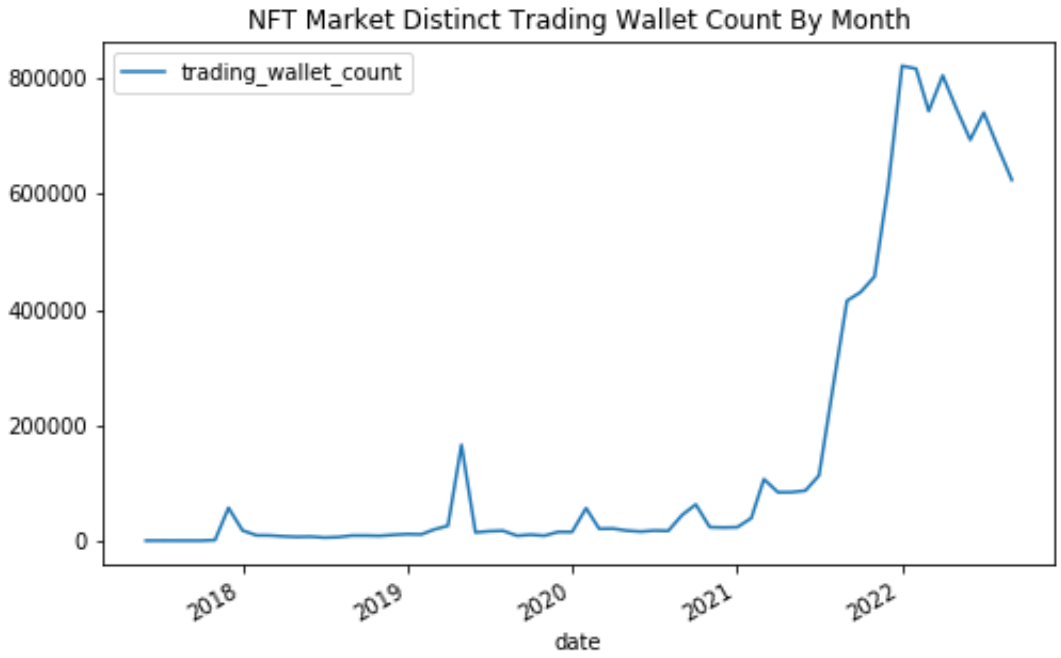

Figure 8: Filtered NFT Market Distinct Trading Wallet Count By Month

As can be seen from Figure 8, which shows the monthly counts of the distinct wallets that are trading in the market, there is a drop in 2022 which draws a different conclusion when compared with Figure 3. When examining the market without filtering any NFT projects, we've seen that the growth continues in 2022 even though the rate of growth is decreasing. However, when we filter out the projects that are not recognized by the market, we can clearly see a drop in the trading wallet counts in 2022.

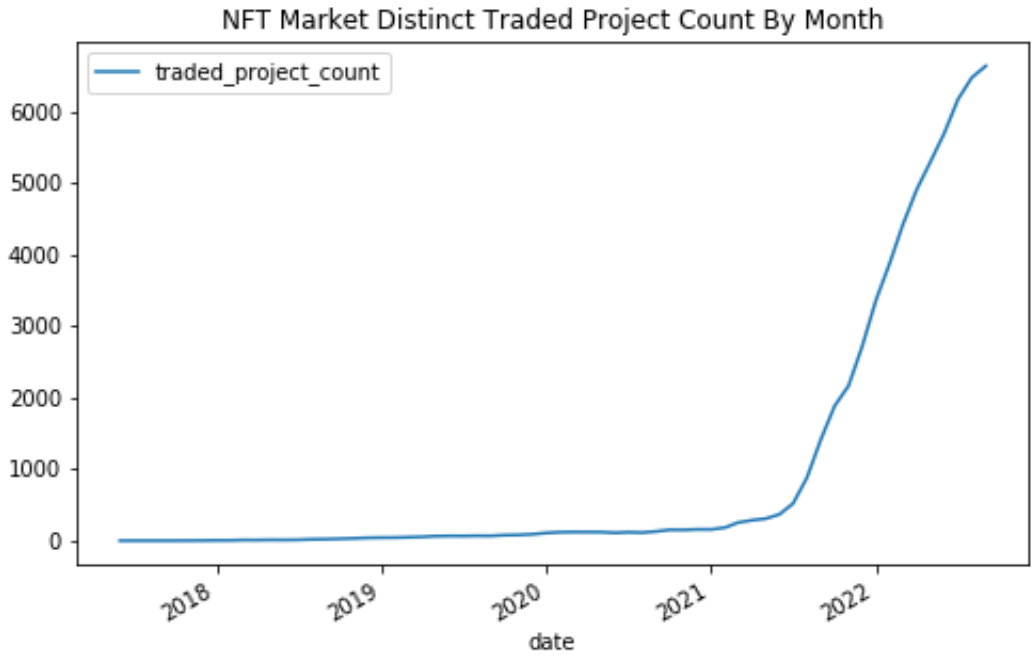

Figure 9: Filtered NFT Market Distinct Traded NFT Project Count By Month

It is obvious from Figure 9, which shows the monthly counts of the distinct NFT projects that are traded in the market, the growth that started in 2021 continues in 2022. If we combine this information with the previous chart, we can conclude that, even though the transaction counts and trading wallets are decreasing in 2022, users in the market are trying out different NFT projects and broadening their portfolios.

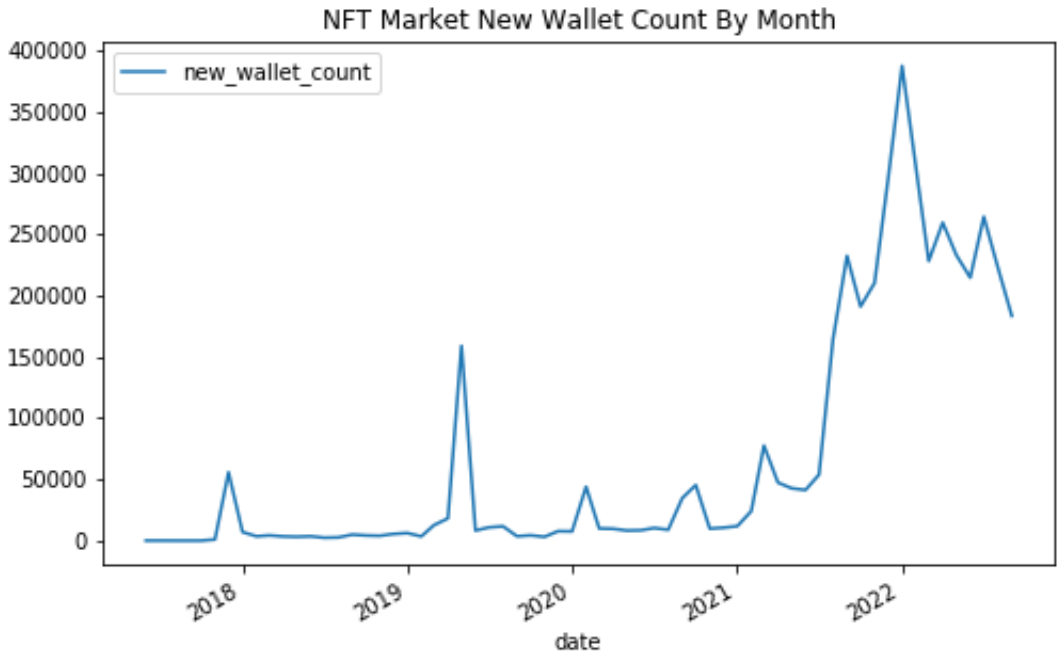

Figure 10: Number of Wallets Entering the Filtered NFT Market By Month

As can be seen from Figure 10, which shows the number of wallets entering the market, there is huge a drop in 2022. The monthly number of wallets entering the market decreased from ≈400.000 to ≈200.000 which means a 50% decrease. This draws a different picture than Figure 5 which shows the trend in the non-filtered market.

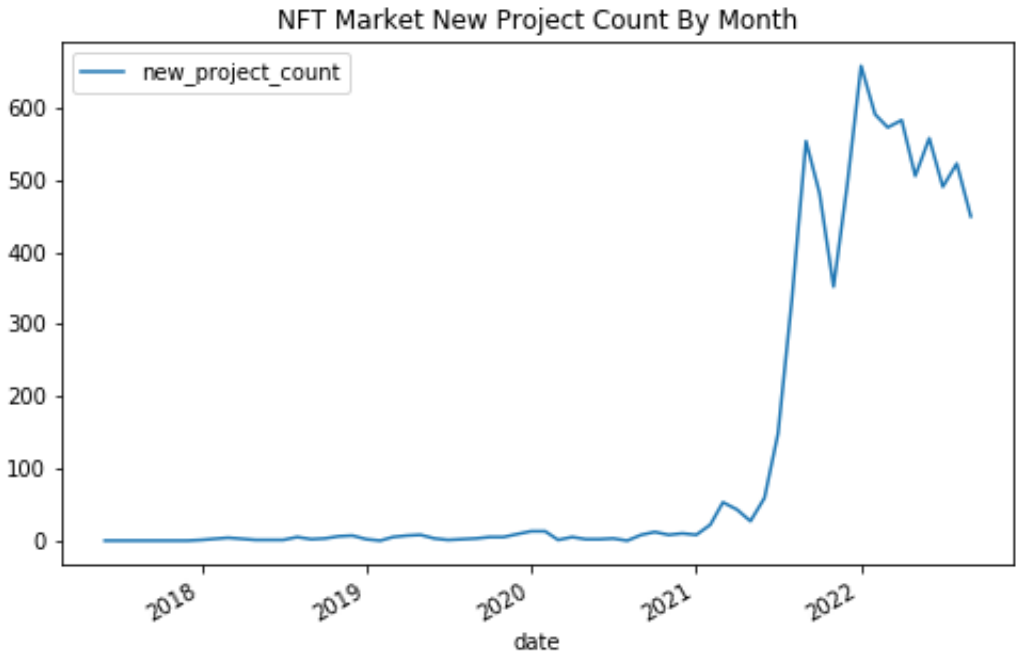

Figure 11: Number of NFT Projects Entering the Filtered NFT Market By Month

As can be seen from Figure 11, the number of new NFT projects is decreasing in 2022 as well as the number of new wallets. However, the magnitude of the decrease is lesser in the new NFT projects.

To conclude, when the data is analyzed without any filtering in place, it is observed that the growth that happened in the NFT market in 2021 continued in 2022 as well, yet, the rate of growth in 2022 was a bit low when compared to 2021. However, when the unrecognized and low-valued projects were filtered out from the data, we observed that the growth that started in 2021 stopped in 2022 and the market stabilized. Even though the growth does not continue that does not necessarily mean that the market is going down. The macro indicators of the market as of the end of September 2022 match the level of mid-late 2021. However, there is a decrease in the latest month in every indicator but the traded project count, which raises a concern regarding the future of the market.

5.2 Wallet Analysis

When we applied a filter on the NFT projects to include only the ones that have been recognized by the market, we have seen that the growth stopped in 2022 and the market stabilized. However, we do not have any information regarding how the wallets in the market are behaving. To analyze the wallets more in detail, we have drawn a random sample of 5000 wallets from the population that has a filter on NFT projects. Firstly, we have checked if the sample follows the same trends in macro indicators as population, and confirmed that it is. Related graphs can be seen from Figures 54,55,56, 57 and 58 in the Appendix section. Then we started analyzing the wallets in terms of transactional behaviors.

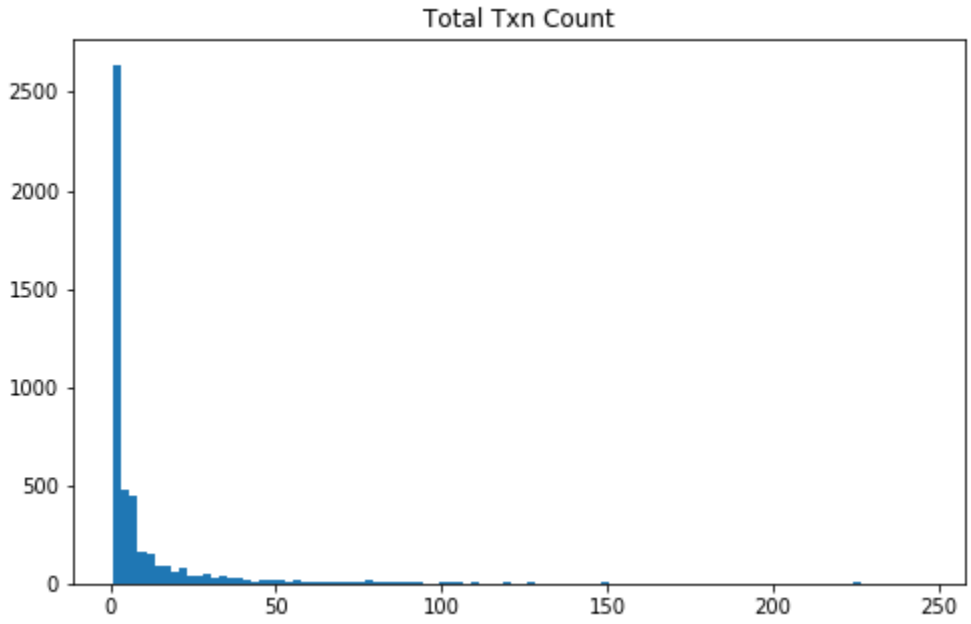

Figure 12: Histogram of Wallets Based on Number Of Transactions

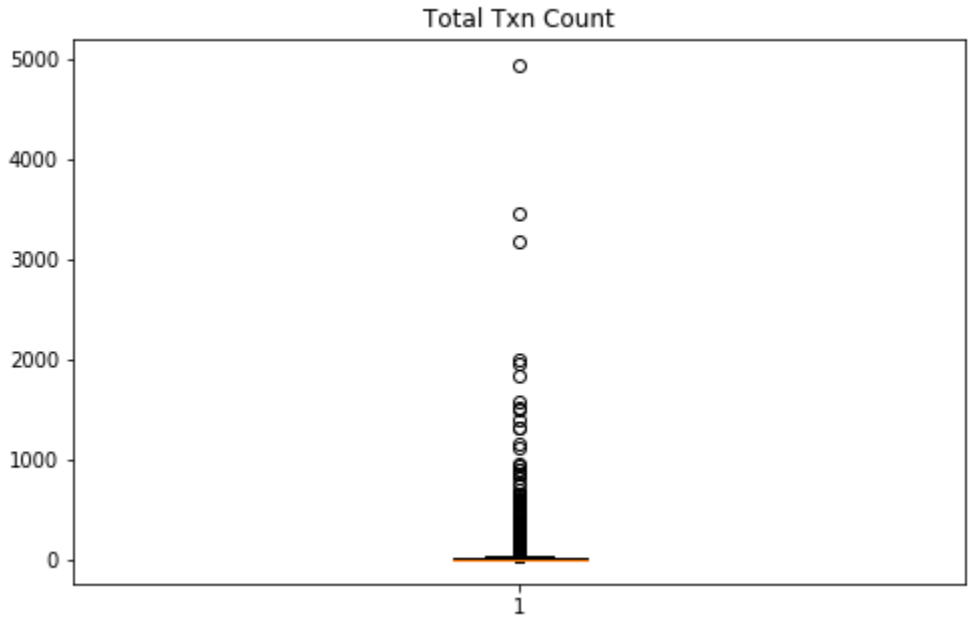

Figure 13: Boxplot of Wallets Based on Number Of Transactions

As can be seen from Figure 12, approximately half of the wallets have just 1 transaction. Also, there are just a few wallets that have more than 50 transactions. Moreover, as can be seen in the Figure 13 some wallets have up to 5000 transactions which can be referred to as outliers.

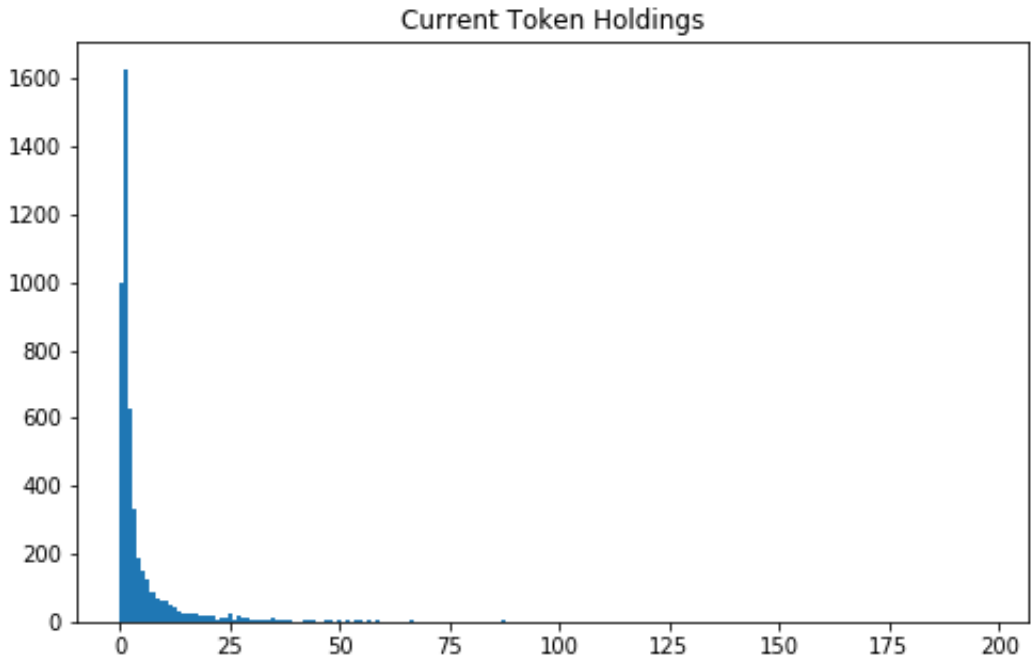

Figure 14: Histogram of Wallets Based on Token Holdings

As can be seen from Figure 14, approximately 20% of the sample have 0 token holdings, which means that they have sold all the tokens they bought. Additionally, 30% of the sample just have 1 token and there are just a few wallets that have more than 25 tokens. There are also outliers in terms of token holdings as it was in the transaction counts. Related boxplot can be seen from Figure 59 in the Appendix section.

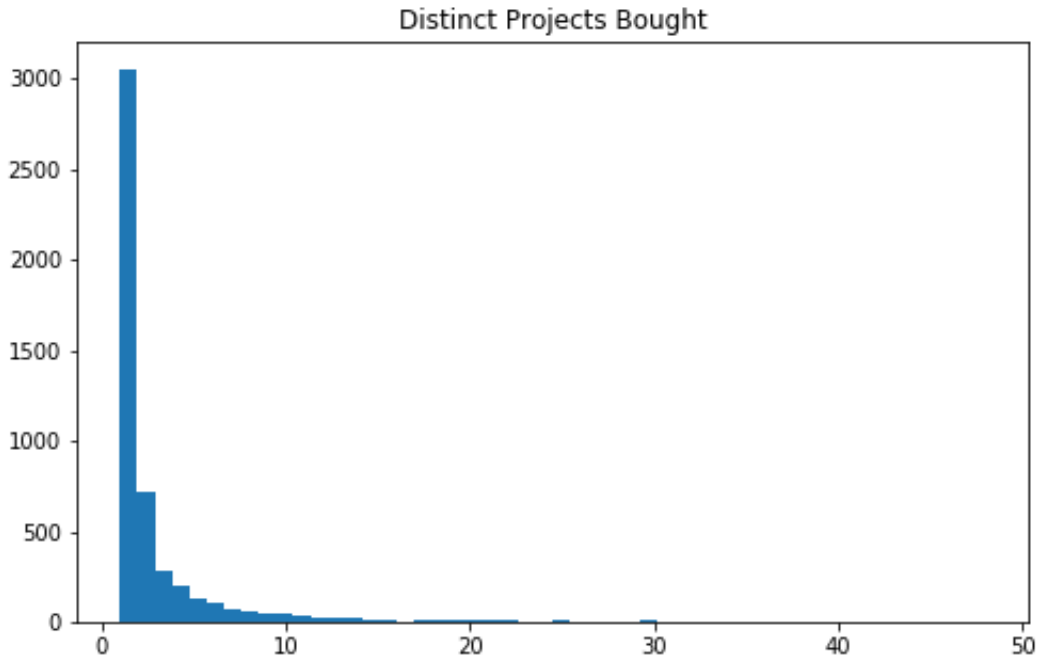

Figure 15: Histogram of Wallets Based on Number of Distinct Projects Bought

As can be seen from Figure 15, approximately 60% of the sample bought tokens from just one NFT project. This means that most of the sample tend to buy from just one project and are not interested in other projects. There are also outliers in terms of the number of projects bought. Related boxplot can be seen from Figure 60 in the Appendix section.

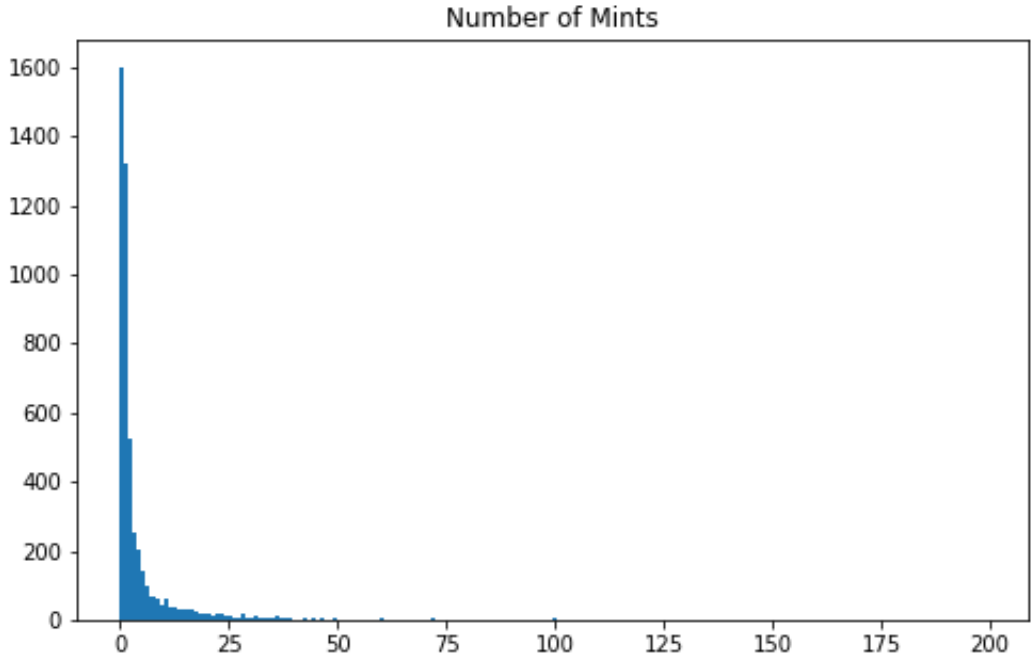

Figure 16: Histogram of Wallets Based on Number of Mints

Minting an NFT is, in more simple terms, uniquely publishing an NFT token on a blockchain to make it officially a commodity that can be bought and sold. As can be seen from the Figure 16, approximately 32% of the sample do not have any mint transactions and $\approx$ 25% of the sample just had one minting event. Again, there are just a few wallets that have more than 25 mints. There are also outliers in terms of the number of mints. Related boxplot can be seen from Figure 61 in the Appendix section.

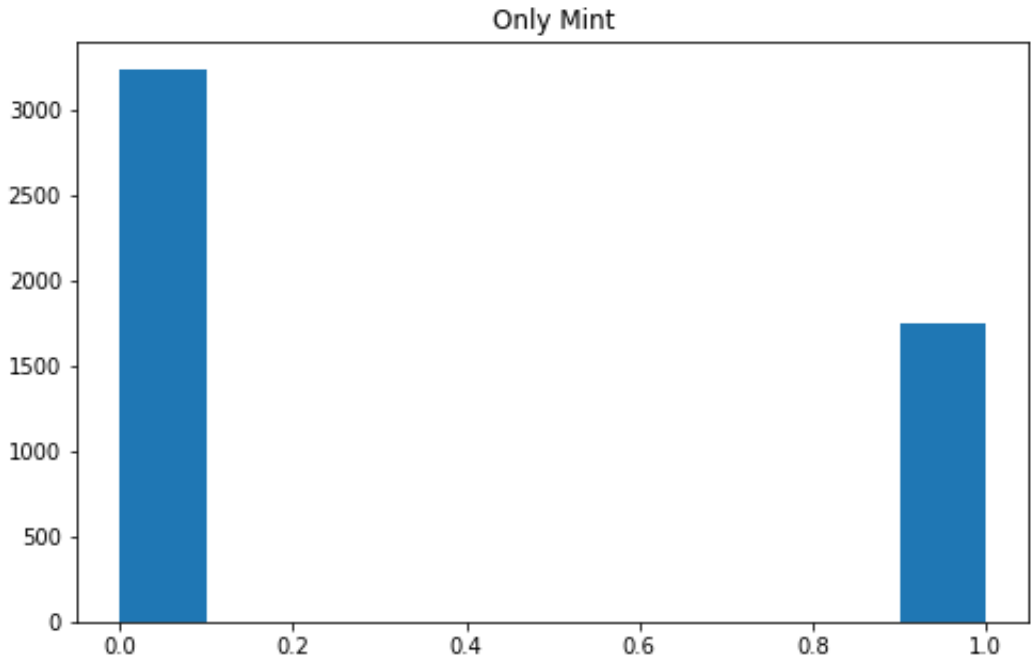

Figure 17: Distribution of the wallets in terms of having only mint transactions

As it can be seen from the Figure 17, $\approx 35\%$ of the sample only have minting transactions which can be segmented as only-minters.

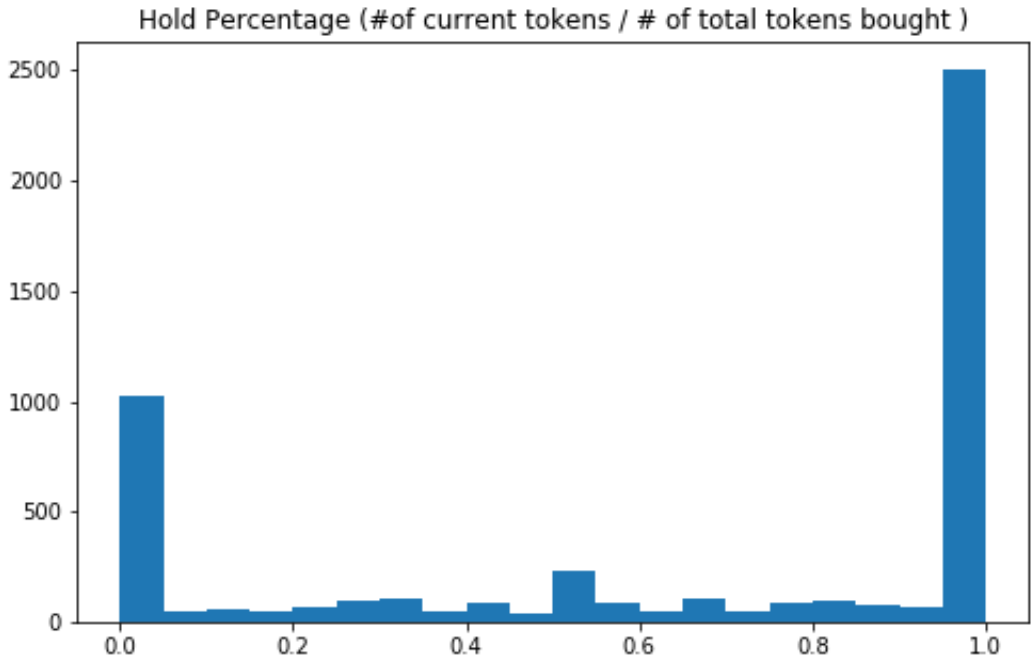

Figure 18: Histogram of Wallets Based on Holding Percentage

Holding percentage can be formulated as following;

holding percentage = number of current token holdings / total number of tokens bought

As it can be seen from the Figure 18, approximately 50% of the sample did not sell any of the tokens they bought, on the other hand, 20% of the wallets sold all of the tokens they bought. Remaining of the sample is distributed between (0-1) without any obvious skew. Additionally, the holding percentage distribution of only minters is similar to the distribution of not-only minters. However, there are fewer wallets distributed between (0-1) in the only minters as can be seen from Figure 62 in the Appendix section.

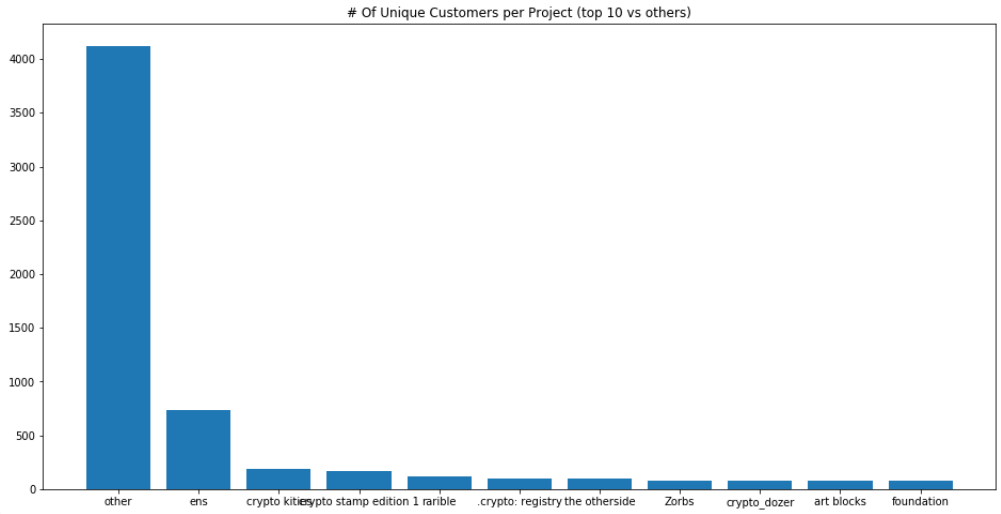

Figure 19: Histogram of NFT Projects Based on Number of Unique Buyers

As can be seen from Figure 19, wallets of this sample are not interested in the same NFT projects. The most common project bought from the wallets is the ENS, which is bought by around 15% of the sample. Therefore we can conclude that the sample is not concentrated in any of the NFT projects.

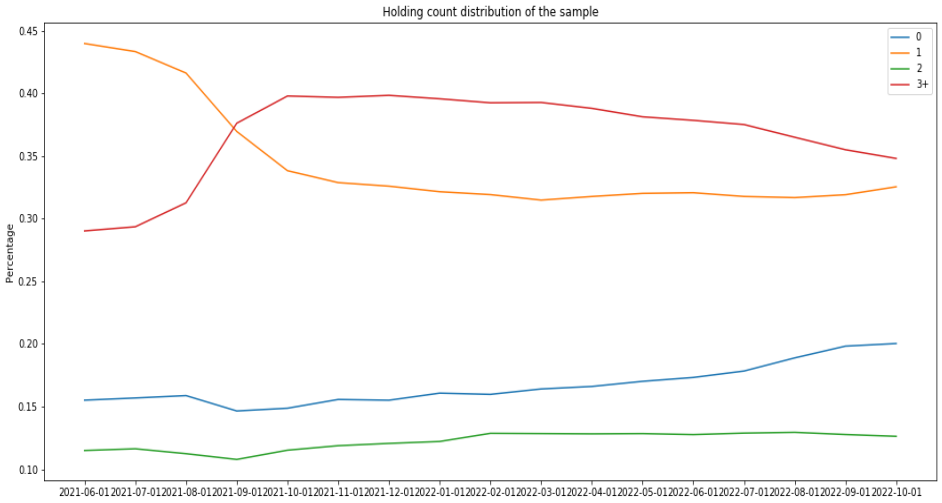

Figure 20: Distribution of the sample in terms of token holdings over time

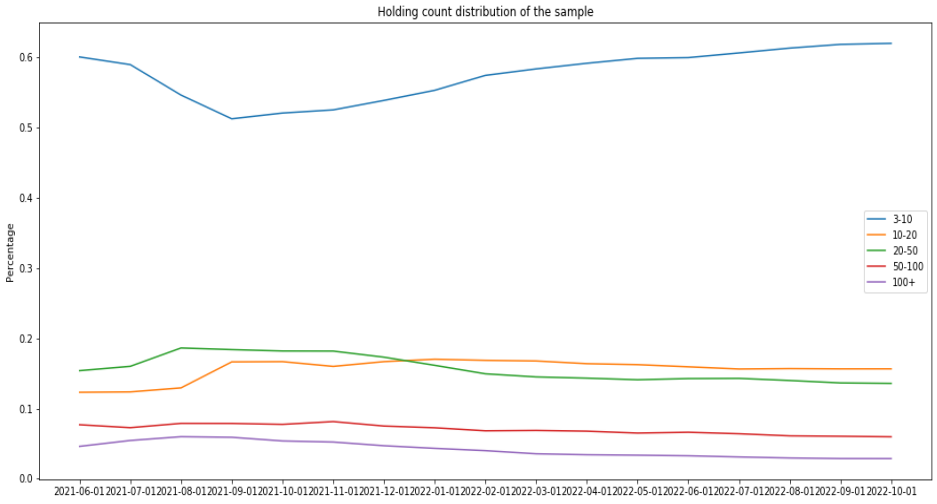

Figure 21: Distribution of the sample in terms of 3+ token holdings over time

Figure 20 summarizes how the token-holding tendencies of the sample changed over time. As can be observed, the percentage of the wallets that have more than or equal to 3 tokens was around 30% on June 1st, 2021 and the percentage of the wallets that have just 1 token was about 45%. Over time, the percentage of wallets that have $\geq$ 3 tokens increased and stabilized at around 36%, also the percentage of the wallets that have just 1 token decreased and stabilized at around 33%. Additionally, the percentage of the wallets that have 0 tokens increased to 20% which can be signaling that wallets are exiting the market.

Figure 21 shows how the $\geq$ 3 token holders are distributed over time. 60% of the $\geq$ 3 token holders have $>$ 3 and $\leq$ 10 token holdings. Around 30% of the $\geq$ 3 token holders have 10 to 50 tokens and the remaining 10% have more than 50 tokens and this distribution does not drastically change over time. Non-normalized stacked-bar plots of this data can be seen from Figures 63 and 64 in the Appendix section.

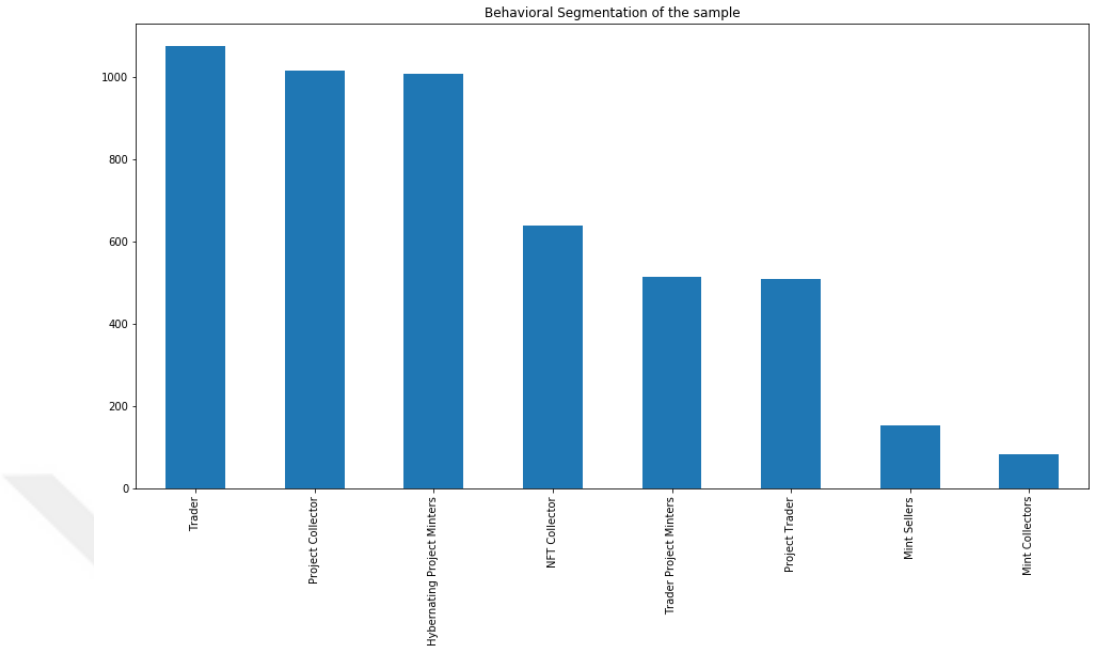

Figure 22: Behavioral Segmentation of The Sample

Using the characteristics of the wallets mentioned above, we've created 8 different segments;

- Trader
- Project Collector
- Hybernating Project Minter
- NFT Collector
- Trader Project Minter
- Project Trader
- Mint seller
- Mint collector

The definition of the segments can be seen in Figure 65 in the Appendix section.

As can be seen from Figure 22, the three most frequent segments are; Traders, Project

Collectors, and Hybernating Project Minters. The trader segment represents the group that buys different NFT projects and trades them. The Project Collector segment represents the group that buys from just one project and holds it. Finally, the Hybernating Project Minter segment represents the group that mints just 1 project and holds it. There are three segments that have medium frequencies; NFT Collectors, Trader Project Minters, and Project Traders. The NFT Collector segment represents the group that buys tokens from different projects and holds them. The trader Project Minters segment represents the group that mints from just 1 project and trades it. Project Traders represent the group that buys from just one project and sells it. Finally, there are two minor segments; Mint Sellers and Mint Collectors. Mint Sellers are the ones that sell the tokens they minted and Mint Collectors are the ones that hold the minted tokens. It should be noted that the wallets in the segments that are reffered to as 'Mint', only acquire NFT tokens by minting it.

5.3 Focusing on a Specific Collection: Bored Ape Yatch Club

As can be seen from the previous Wallet Analysis section, 50% of the wallets in the sample have just 1 transaction and 60% of the sample buys from just one NFT project. This can be interpreted as most of the wallets belonging to experimental and inexperienced users. In order to analyze how experienced and well-funded users behave in the market, we've decided to analyze a sample that buys at least 1 Bored Ape Yacht Club(BAYC) Token. As mentioned earlier, BAYC is one of the most popular and expensive tokens, so the people who are buying BAYC are the users who are deeply interested in the market. To analyze these users in detail, we've drawn a random sample of 5000 wallets from the wallets that bought at least 1 BAYC token.

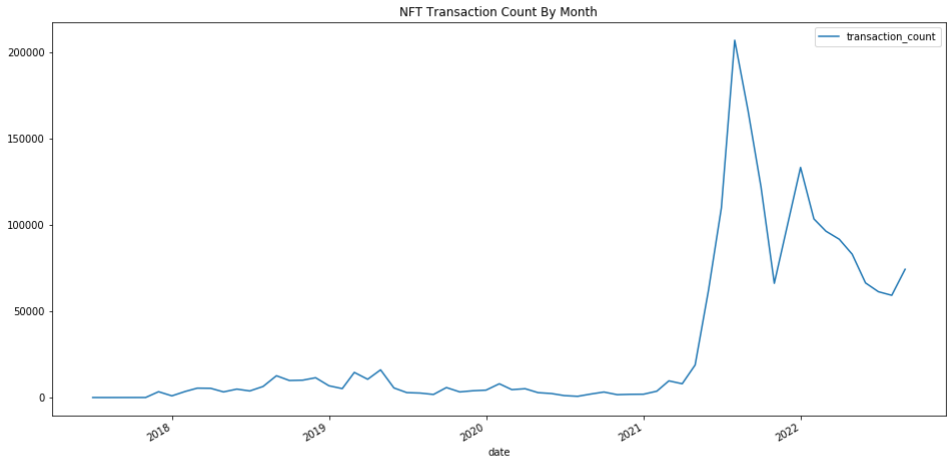

Figure 23: Monthly Transaction Counts, BAYC Sample

As can be seen from Figure 23, the trend explored in the market is present in the BAYC sample as well. There is approximately a 50% decrease in the monthly transaction counts. However, we would like to underline that the monthly transaction counts of the BAYC sample are significantly greater than the previous sample. While the previous sample which represents the whole NFT market has a maximum of 12.000 transactions per month, the BAYC sample has 200.000 which proves that the BAYC sample is quite interested in the market. Other graphs regarding the macro indicators which also prove that the BAYC sample is heavily interested in the market can be seen from Figures 66, 67, 68 and 69 in the Appendix section.

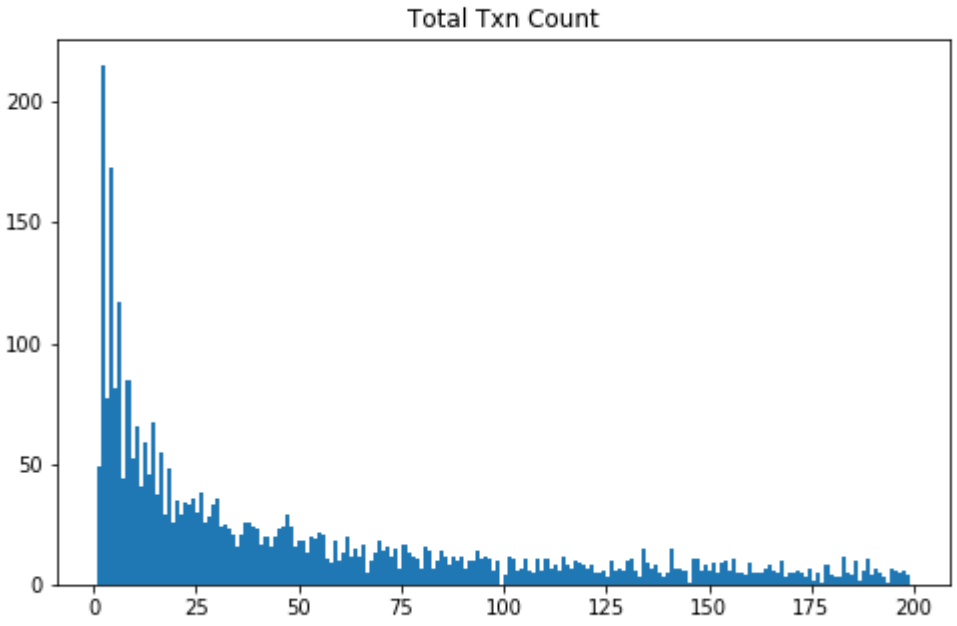

Figure 24: Histogram of Wallets Based on Number Of Transactions, BAYC Sample

As can be seen from Figure 24, wallets of the BAYC sample have much more transactions when compared to the overall sample. There are also outliers in the data in terms of transaction counts, as some wallets have approximately 50000 transactions. Related boxplot which shows the outilers can be seen in Figure 70 in the Appendix section.

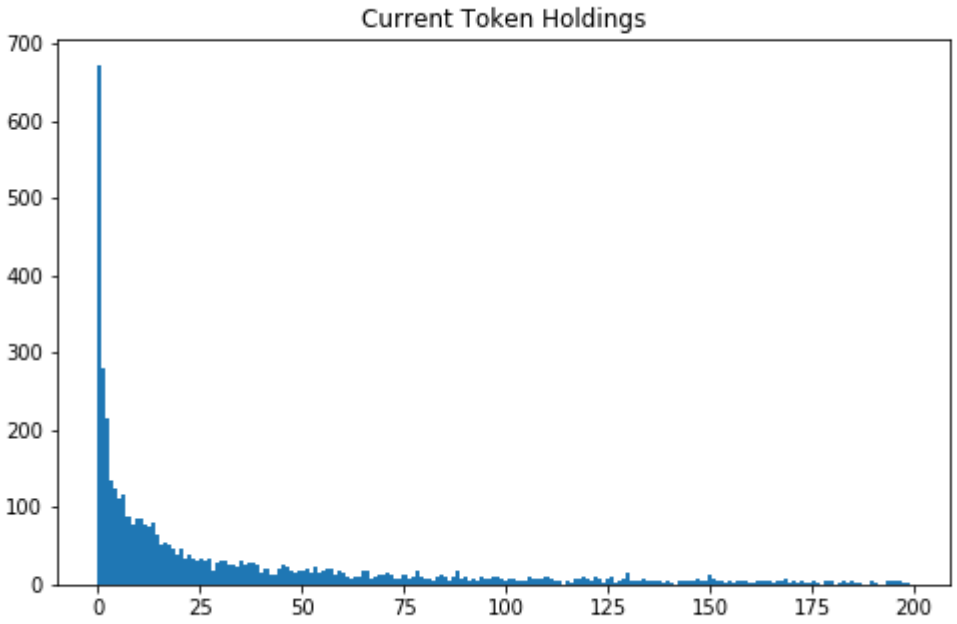

Figure 25: Histogram of Wallets Based on Number Of Token Holdings, BAYC Sample

As expected, wallets of the BAYC sample have much more token holdings when compared to the overall market. Again there are also outliers in the data, which can be seen from Figure 71 in the Appendix section.

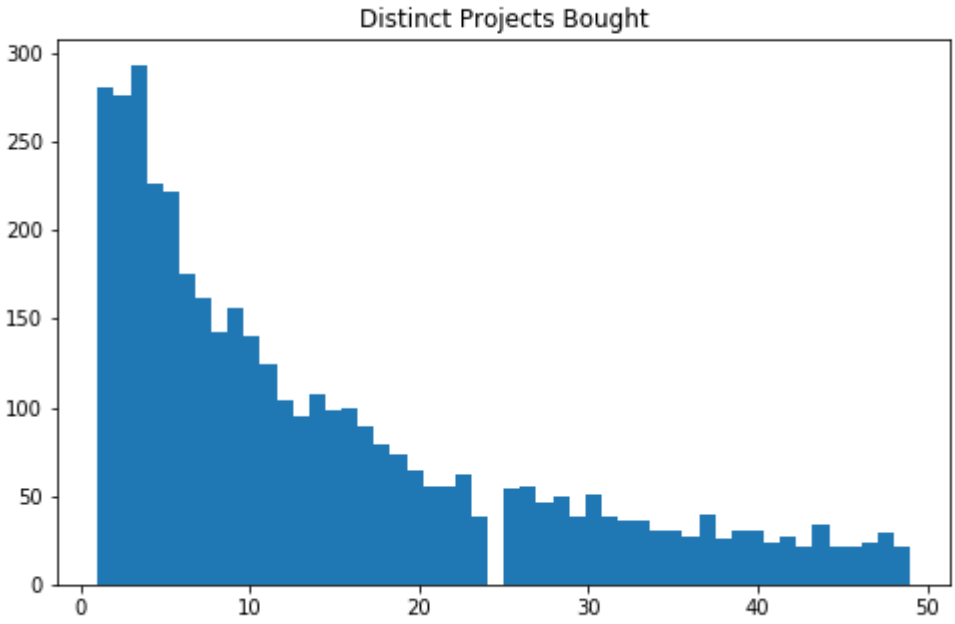

Figure 26: Histogram of Wallets Based on Number of Distinct Projects Bought, BAYC Sample

As stated earlier, BAYC buyers are heavily interested in the market, and as can be seen from Figure 26, wallets of the BAYC sample buy tokens from various different projects which can prove this interest.

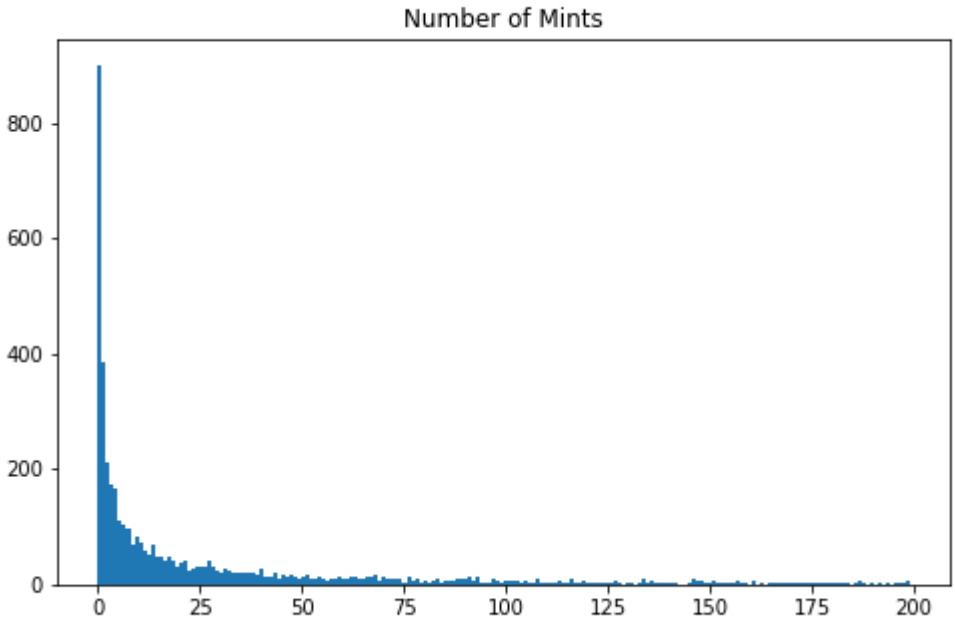

Figure 27: Histogram of Wallets Based on Number of Mints, BAYC Sample

Even though there is just one only-mint wallet in the BAYC sample, the minting count of the sample is much higher when compared to the overall sample.

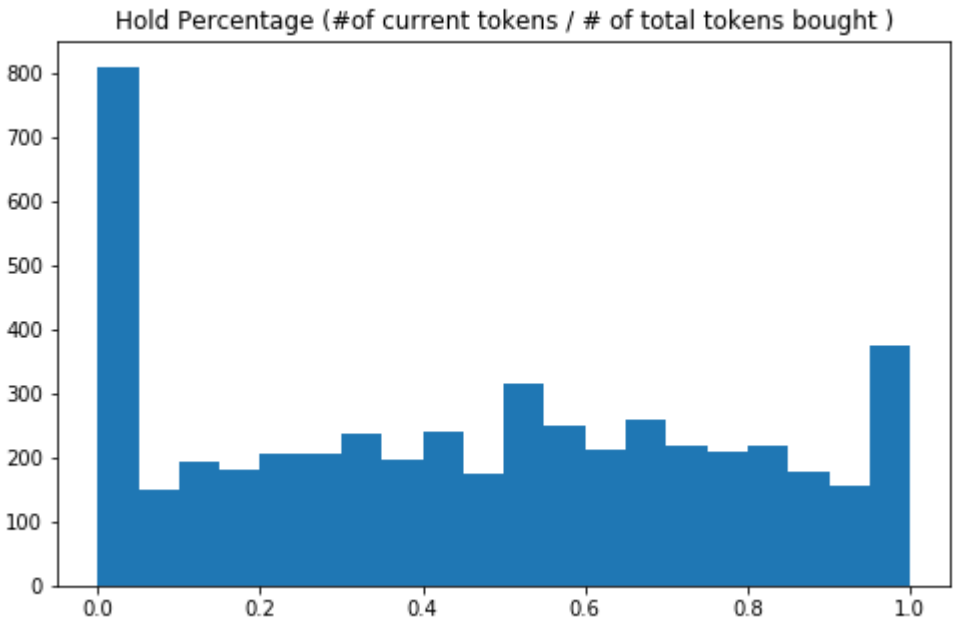

Figure 28: Histogram of Wallets Based on Holding Percentage, BAYC Sample

As can be seen from Figure 28 Holding Percentage of the sample skewed in 0 and 1. However, the number of wallets having 0 or 1 holding percentage is significantly lower than the overall market. Thus, there are much more wallets between (0-1) which means that wallets of this sample tend to trade tokens rather than holding every token or selling all of the tokens.

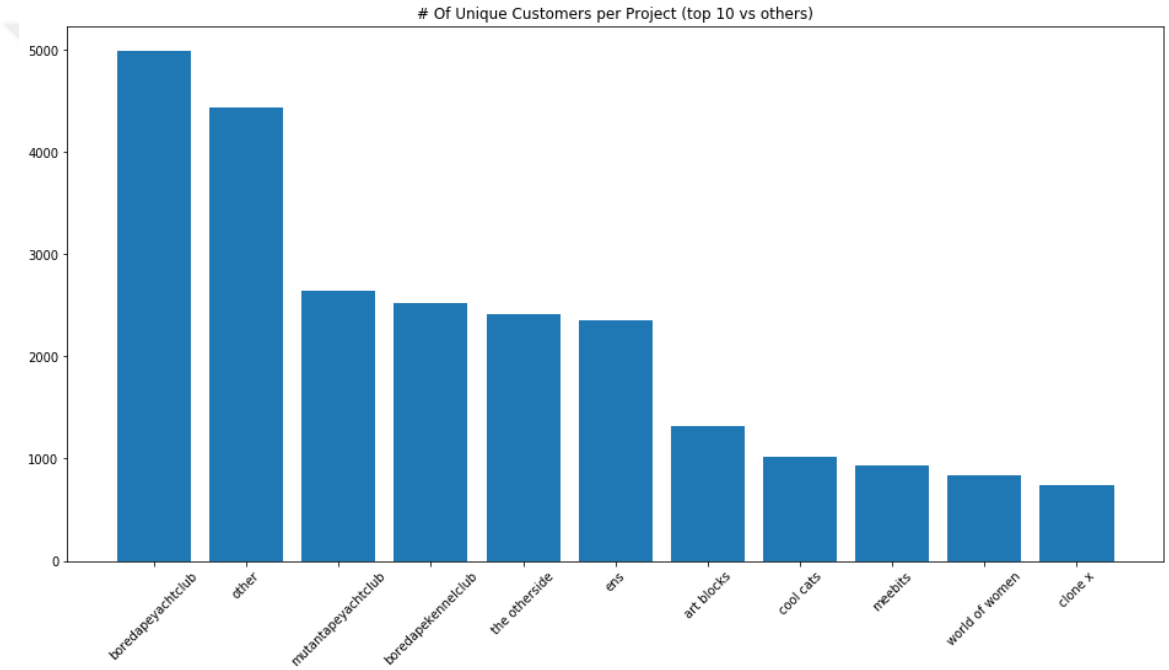

Figure 29: Histogram of NFT Projects Based on Number of Unique Buyers, BAYC Sample

As can be seen from Figure 29, wallets of this sample are interested in the same NFT projects. Approximately 50% of the sample bought tokens from Mutant APE Yacht Club, Bored APE Kennel Club, The Otherside and, ENS. However, this was not the case for the overall sample as the most common project was only bought by 15% of the sample.

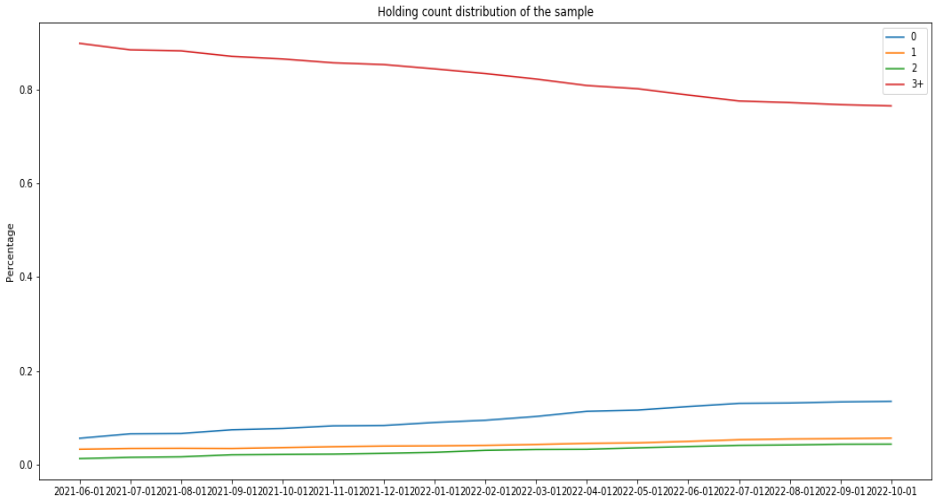

Figure 30: Distribution of the sample in terms of token holdings over time, BAYC Sample

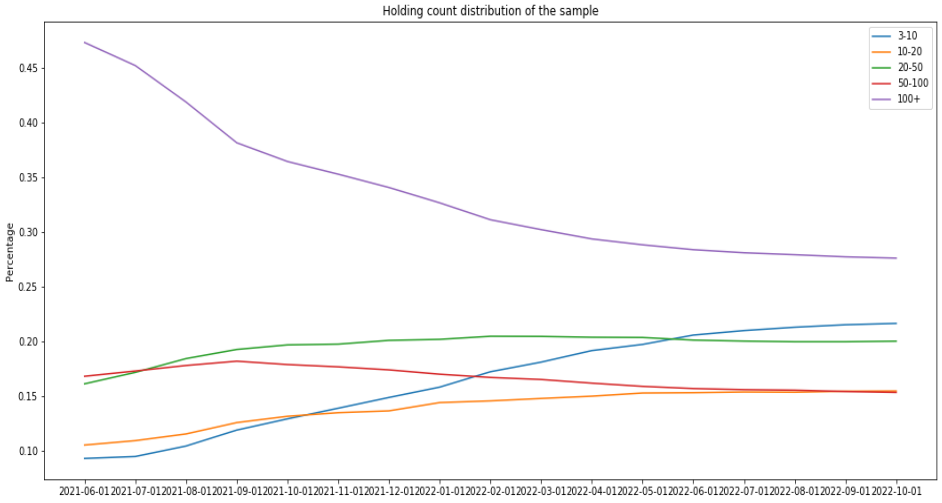

Figure 31: Distribution of the sample in terms of 3+ token holdings over time, BAYC Sample

Figure 30 summarizes how the token-holding tendencies of the sample changed over time. As can be observed, the percentage of the wallets that have more than or equal

to 3 tokens was around 85% on June 1st, 2021 and the percentage of the wallets that have just 1 token was about 3%. Over time, the percentage of wallets that have ≥ 3 tokens decreased and stabilized at around 80%, also the percentage of the wallets that have 0 tokens increased and stabilized at around 8%. Additionally, the percentage of the wallets that have 1 or 2 tokens was steady over time.

Figure 21 shows how the ≥ 3 token holders are distributed over time. 45% of the ≥ 3 token holders had >100 token holdings and this percentage decreased and stabilized at around 30%. In the meantime, the percentage of > 3 and ≤ 10 token holders increased from 8% to 22%. Both of the figures show a different trend from the previous sample as the wallets in BAYC sample tend to hold more tokens.

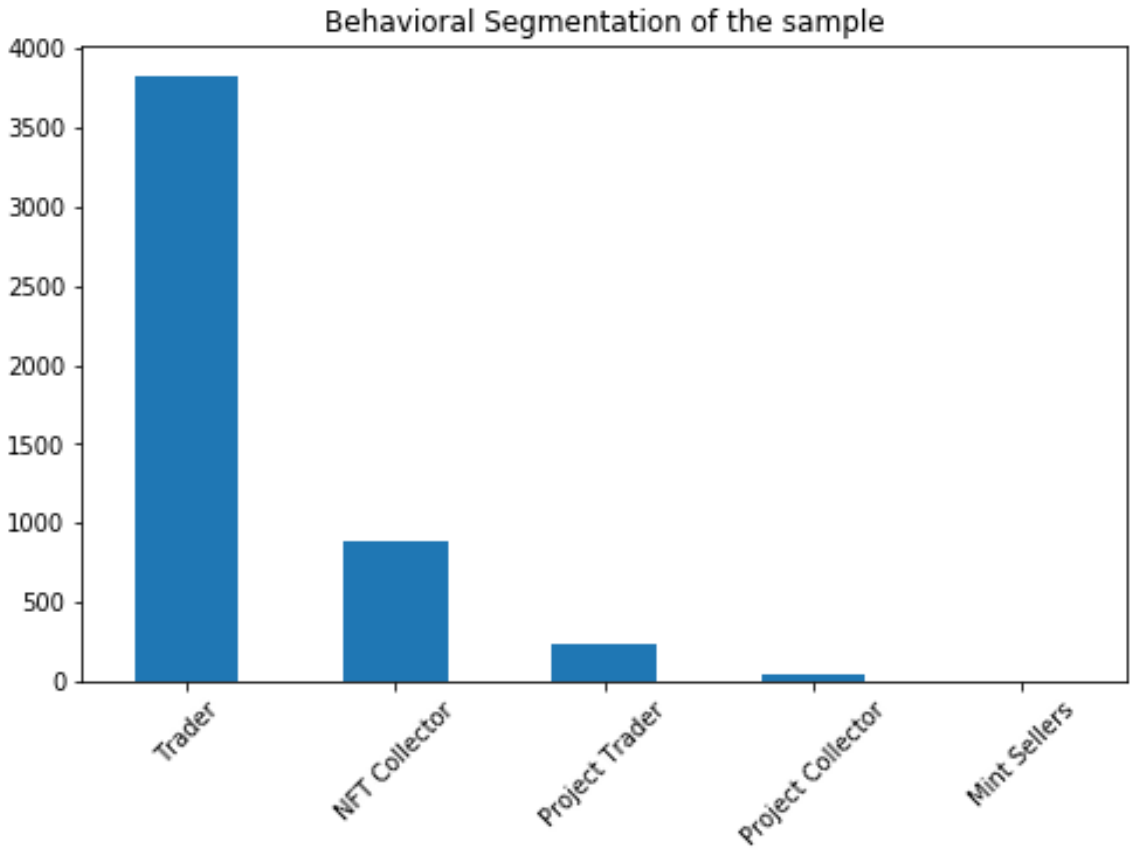

Figure 32: Behavioral Segmentation of The Sample, BAYC Sample

As can be seen from Figure 32, when we apply the same segmentation to the BAYC

sample, it is obvious that most of the wallets belong to the Trader segment. This was expected as we've seen that there are much more wallets having holding percentages between (0-1). There are also around 800 NFT Collectors and 200 Project Traders.

To sum up, the BAYC sample has different dynamics and interests when compared to the overall sample. BAYC sample is quite interested in the market and trades at high frequencies. Additionally, the BAYC sample tries out different NFT projects as well. Given that the sample has much more data in terms of transactions and quite interested in the market, we've chosen this sample to try our predictive models.

CHAPTER VI

WALLET ANALYTICS: PREDICTIVE ANALYSIS

6.1 Overview

As mentioned earlier, we've chosen the BAYC sample to try out predictive models. The reason behind this is that this sample trades at high frequencies and volumes therefore information that would be gathered from the sample would be much more valuable. Additionally, because of high frequency trading this sample has much more data which would help predictive models in learning.

Two main categories of predictive metrics were explored in this study; market-related metrics and collection-related metrics. Market-related metrics are the metrics that are in the market level which means they are inclusive of all the NFT projects, and collection-related metrics are the metrics in the collection level. We've chosen metrics that could help the actors in the market such as NFT Marketplaces, NFT creators, Crypto Lenders, etc., by providing predictive visibility on what the wallets are going to do.

6.2 Market Related Metrics

6.2.1 Is Trading

The first predictive model is built upon is trading target. Target definition is, in simple terms, like the following; looking at the status of Wallet X at time T, is Wallet X going to have a transaction in the following 7 days (T, T+7 days) or not? For example, this information would help NFT marketplaces to optimize their marketing activities by not investing in users that won't make any transactions anyway. When you think of it, this information could be extended and used in other areas as well, a

marketplace can estimate the next period's revenue in terms of transaction commission and even further the expected revenue from a customer thus the lifetime value, etc. As mentioned earlier in the Data Architecture chapter, data is preprocessed to obtain a data set that is ready for modeling.

Snapshot_Date	Wallet_ID	Feature_1	Feature_2	Feature_X	Is_Trading
2022-06-01	X	1	1	1	0
2022-06-01	Y	1	0	1	1

Figure 33: Data Structure Example

Figure 33 contains a fictional example of the data structure. The raw data is snapshotted to acquire the status of the wallets at the snapshot date, and the target is calculated by taking the data between the snapshot date and snapshot date + 7 days into account. For all of the market related metrics, the data is filtered to include only the wallets that have a token at the time of the snapshot.

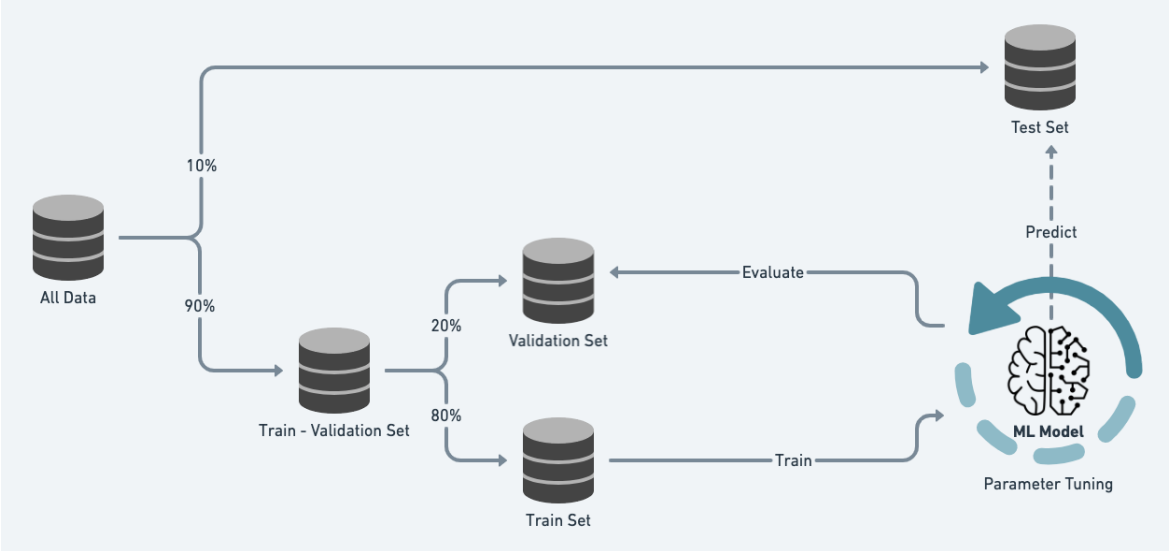

Figure 34: Modeling Pipeline

Figure 34 explains the structure of the modeling pipeline. The data was first randomly split into 10% Test and 90% Train-Validation sets. Then the Train-Validation

set is split into 80% Train and 20% Validation set. The machine learning model is trained on the Train set and evaluated on the Validation set. By using the Validation set while evaluating the performance of the model, we prevent the model from over-fitting on the train set. There is a grid search in place in the model building to find the best sampling method and correlation threshold to eliminate correlated features. Basically what happens is this; for each sampling method (over sampling, under sampling, no sampling) and correlation threshold for feature elimination(no elimination, 0.9, 0.8, 0.7) pair, there is a hyper parameter tuning in place. Then we pick the best model among all the hyper tuned models. Hyper-parameters of the ML model were tuned by using a python library named HyperOpt, and the AUC-ROC score is used as the model evaluation metric.

Figure 35: Is Trading Target Distribution Across Datasets

Figure 36: Is Trading Target Distribution Across Datasets

Figure 35 and Figure 36 summarize the datasets in terms of size and target distribution. Is Trading was one of the balanced targets unlike the collection related metrics which will be covered in the next sections.

We've used the XGBoost classifier to predict the outcomes, the reason behind this is, as explained in the literature review, the classifier's well-known success in the latest competitions. Hyper-parameters of the XGBoost model are tuned based on AUC performance on the Validation set. After having the tuned model trained, three different prediction thresholds are explored, again using the Validation set. The first threshold is determined using Youden Threshold, the second one determined using the False Positive Rate and True Positive Rate. And the third threshold is determined using the F1 score. Because of the fact that is trading is well balanced, these thresholds don't change the results too much. However, as it will be explored in the next sections, it has a great impact on imbalanced datasets. Deciding on which threshold to use actually depends on the intentions of the modeler, for example, one might want to capture all the True Positives by sacrificing on Precision which means having lower

thresholds, etc.

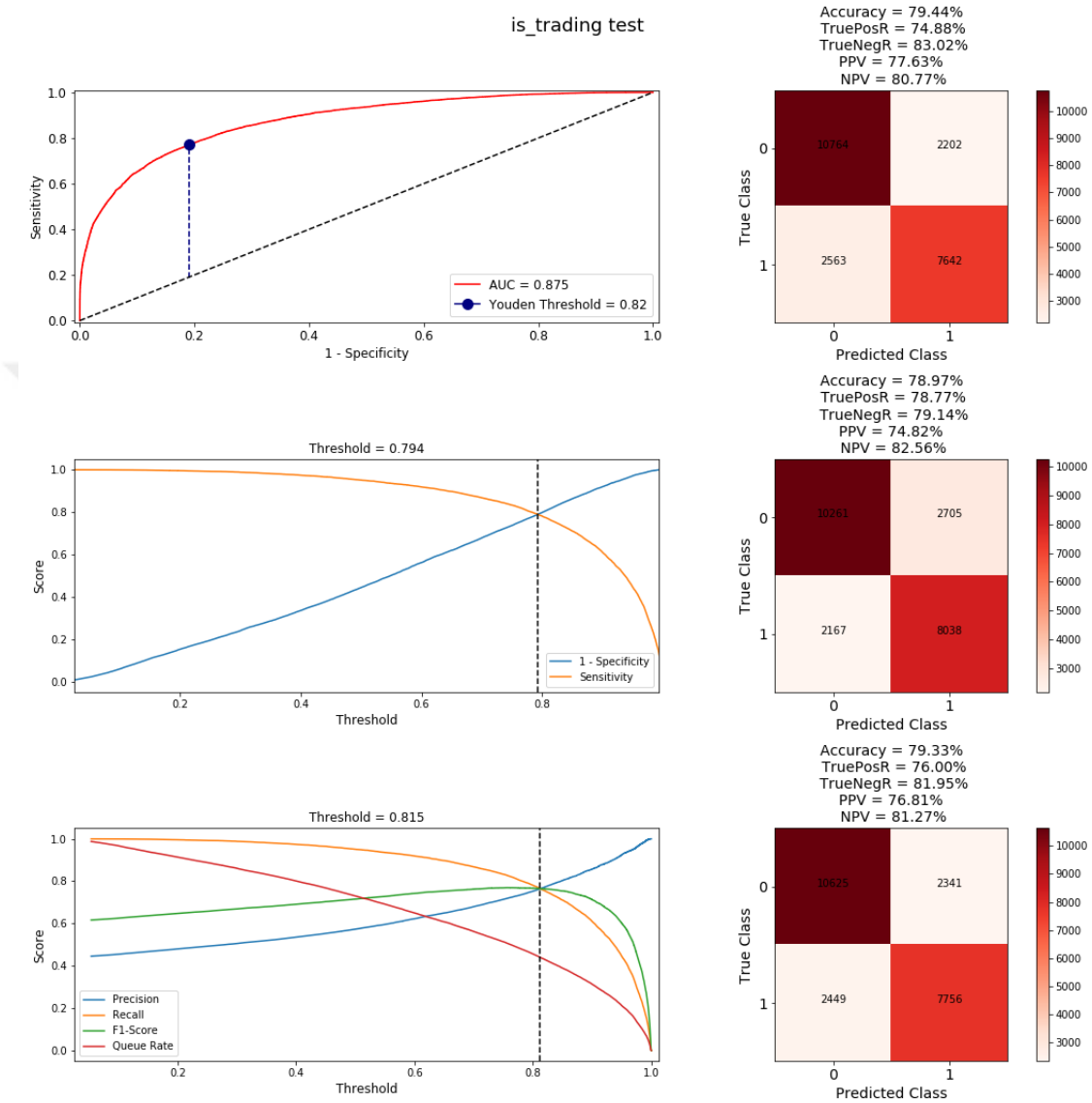

Figure 37: Is Trading Results on Test Set

Figure 37 summarizes the results of the Test set with different thresholds decided using the Validation set. Model achieved an AUC of 0.875 on the Test Set. For this specific problem, we've chosen the F1 threshold which is 0.815, mostly because, both of the miss-classifications have the same cost to us. We've achieved 79.33% Accuracy with a Positive Predictive Value (PPV) of 76.81% and a Negative Predictive Value (NPV) of 81.27%.

6.2.2 Is Buying a New Project

The second predictive model is built upon Is Buying a New Project metric. Target definition is basically like the following; looking at the token holdings of Wallet X at time T, is Wallet X going to buy a new token that is not in his/her portfolio in the following 7 days (T, T+7 days) or not? Again, this metric can be useful for many use cases, for example, this can be a good metric to check the wallet's health in terms of risk for crypto lenders, this can also be used by marketplaces to concentrate their marketing efforts on those users and it can be extended even more by using apriori analyzes. Furthermore, collection creators can also target these users to sell their tokens. Modeling pipeline and data structure is same as Is Trading and can be seen from Figure 33 and Figure 34. Again, the data is filtered to include only the wallets that have a token at the time of the snapshot.

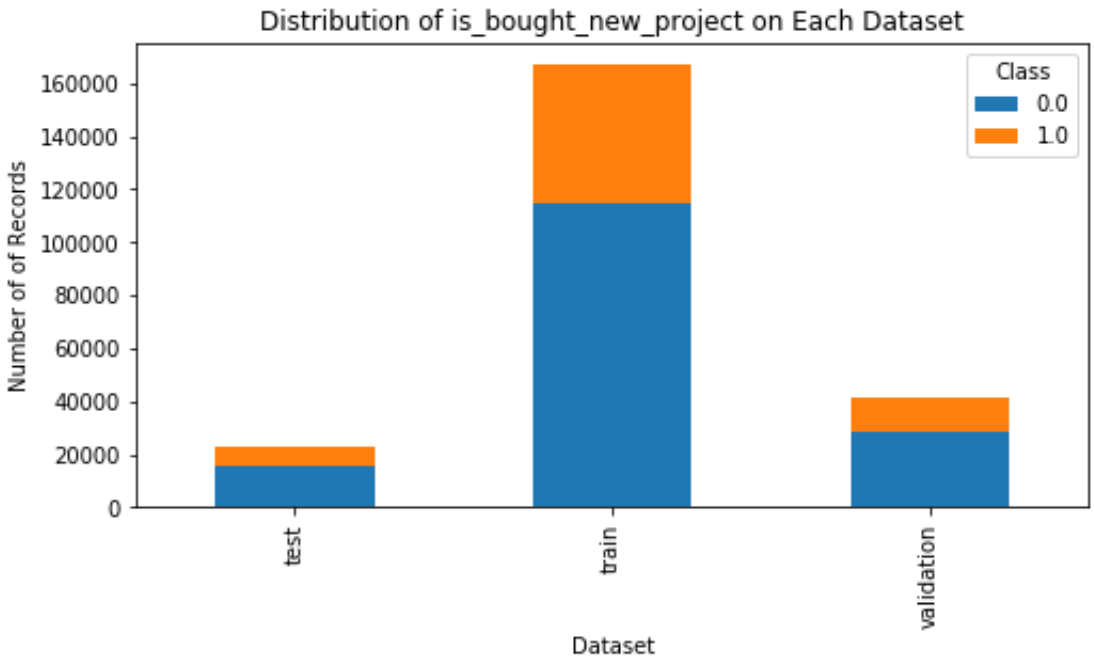

Figure 38: Is Buying a New Project Target Distribution Across Datasets

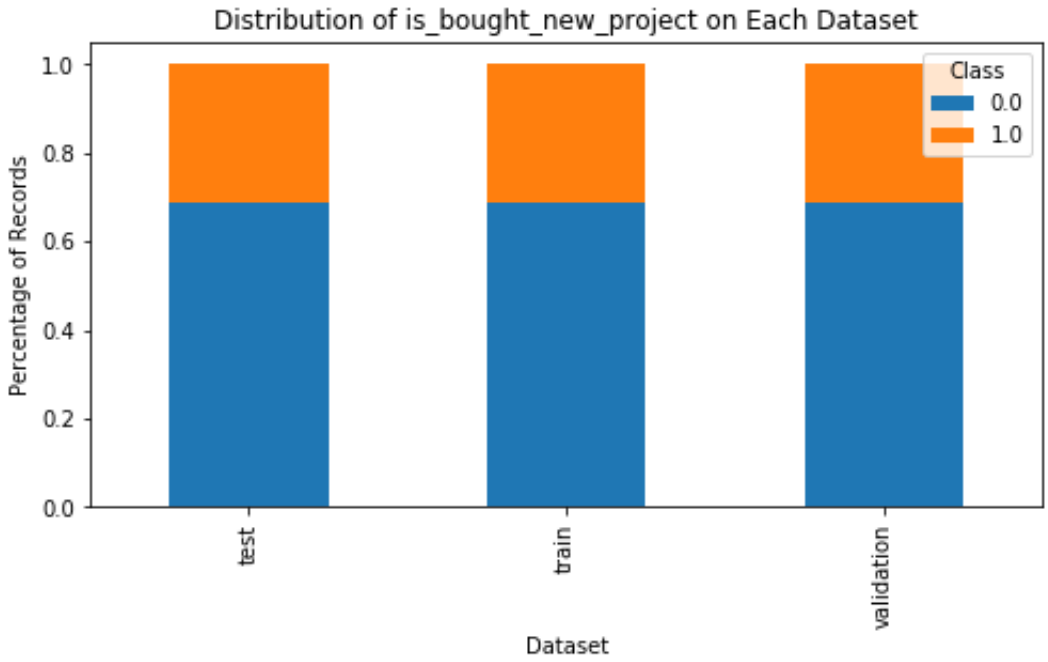

Figure 39: Is Buying a New Project Target Distribution Across Datasets

The target distribution and the size of the dataset for the metric can be seen from Figure 38 and Figure 39. The majority class constitutes approximately 2/3 of the data.

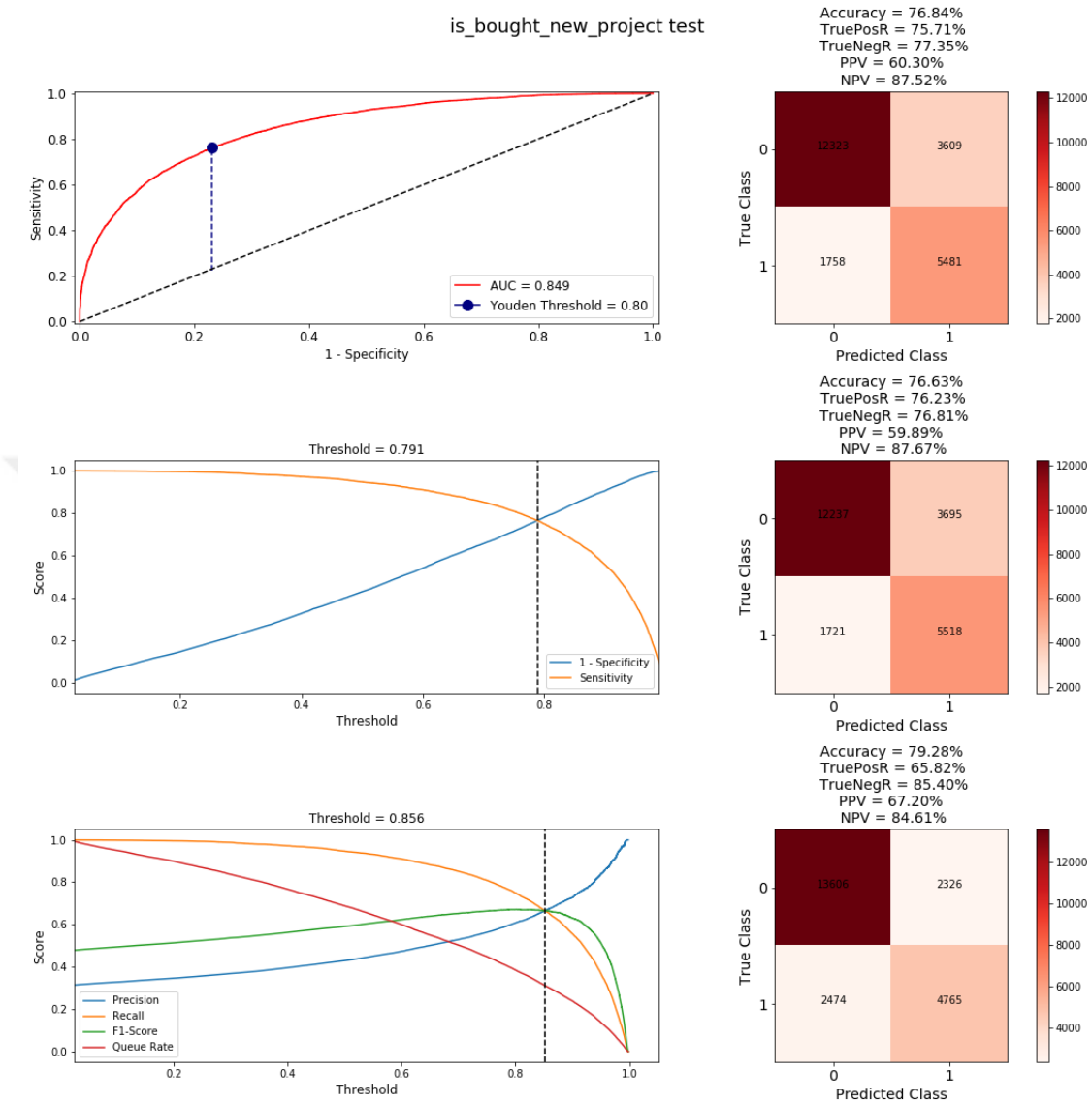

Figure 40: Is Buying a New Project Results on Test Set

As can be seen from Figure 40, we've achieved an AUC of 0.849 on the Test set. The effect of the prediction threshold not as obvious as it was for Is Exiting a Project, however, it is still a tool that can be used to maximize the model performance in terms of the costs associated with the different types of mis-classifications. As you can see, if we pick the 0.856 threshold, which maximizes the F1 score, we would have 67% Positive Predictive Value, 65% True Positive Rate, 85% True Negative Rate, and 84% Negative Predictive Value.

6.2.3 Is Exiting a Project

The third predictive model, also the last one for the market-related metrics, is built upon Is Exiting a Project metric. Target definition is basically like the following; looking at the token holdings of Wallet X at time T, is Wallet X going to sell all of the tokens for a particular project in the following 7 days (T, T+7 days) or not? Again, this metric can be useful for many use cases, for example, there are NFT and Crypto lending businesses growing, and this information regarding wallets could be a good signal to estimate the risk of a wallet that is borrowing. This information could also be used to bid on the NFTs of the wallets that are going to exit from a project in order to obtain the NFTs at lower prices given that the Wallet is probably going to sell. Modeling pipeline and data structure is same as Is Trading and can be seen from Figure 33 and Figure 34. Again, the data is filtered to include only the wallets that have a token at the time of the snapshot.

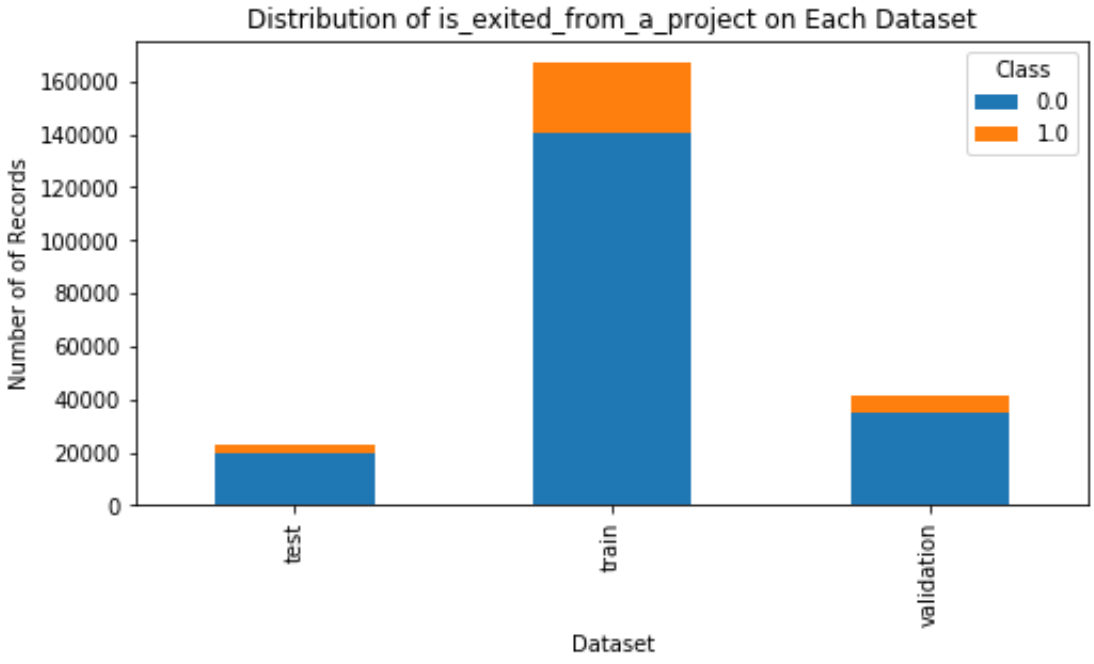

Figure 41: Is Exiting a Project Target Distribution Across Datasets

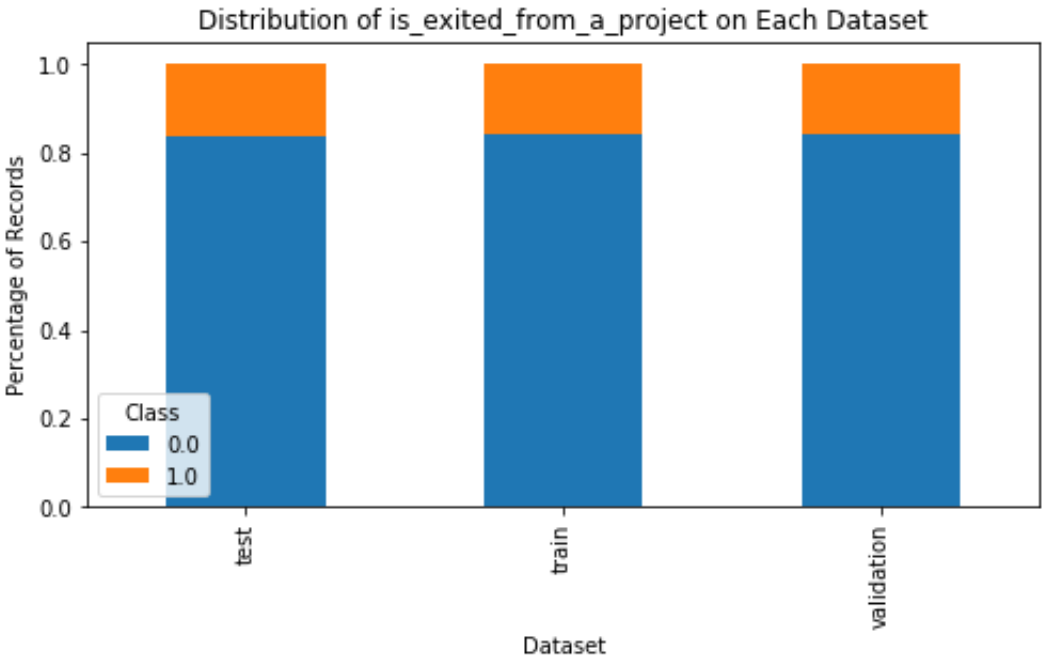

Figure 42: Is Exiting a Project Target Distribution Across Datasets

The target distribution and the size of the dataset for the metric can be seen from Figure 41 and Figure 42. It is obviously an imbalanced data as the majority class constitutes approximately 82% of the data.

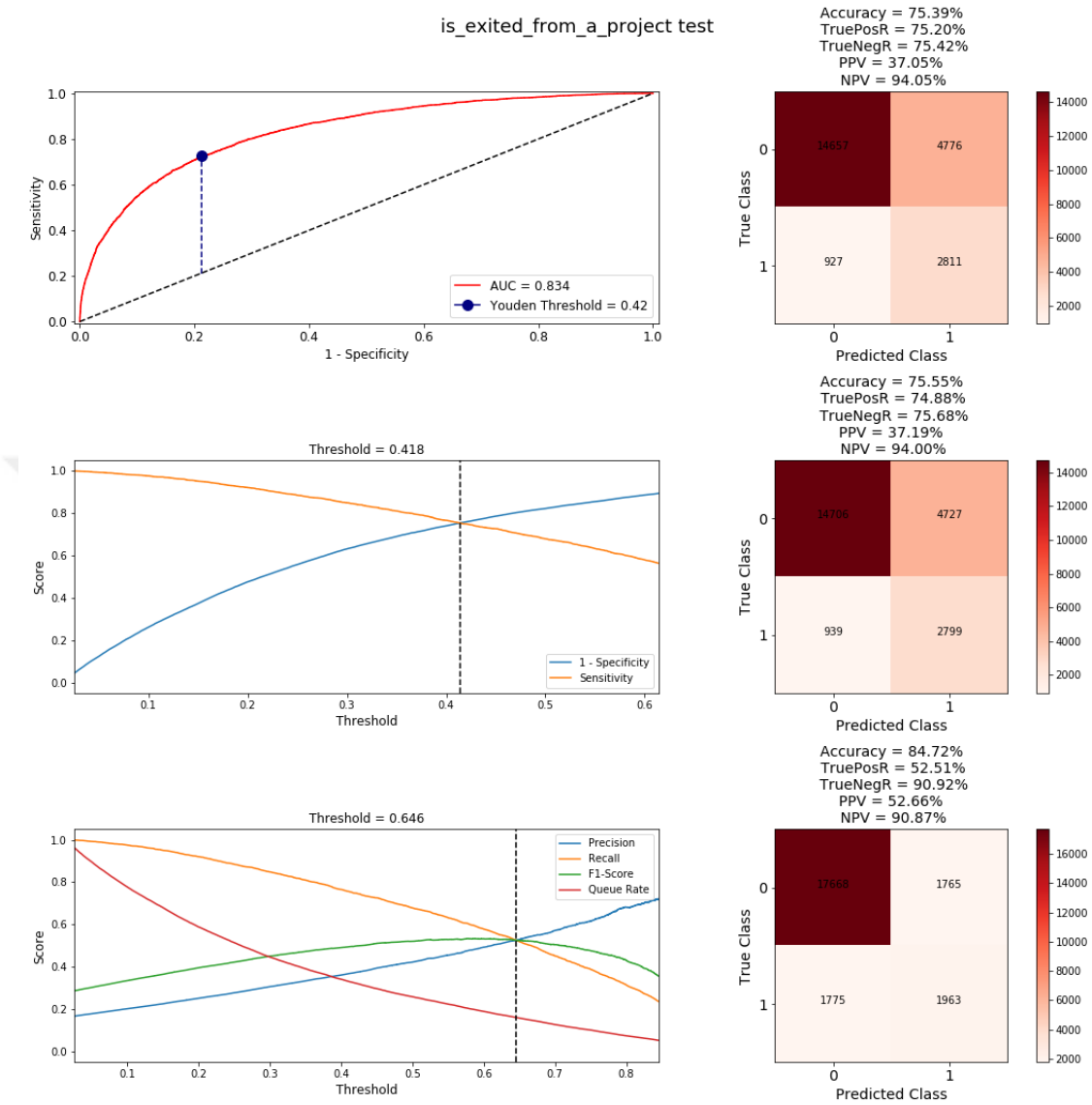

Figure 43: Is Exiting a Project Results on Test Set

We've achieved an AUC of 0.834 on the Test set. As mentioned earlier, prediction thresholds can have a great impact on the confusion matrix for the imbalanced problem. As you can see, if we pick the 0.646 threshold, which maximizes the F1 score, we would have 52% Positive Predictive Value, 52% True Positive Rate, 90% True Negative Rate, and 90% Negative Predictive Value. However, as stated earlier, one might want to capture all the True Positives with the cost of Positive Predictive Value by lowering the prediction threshold, thus overshooting. For example, if this metric

was a true indicator of the risk of the customer, lenders might choose to overshoot in order to capture all the wallets that are going to exit from a project.

6.3 Collection Related Metrics

6.3.1 Is Trading BAYC

The first collection related metric is Is Trading BAYC and it can be described as; looking at the status of Wallet X at time T, is Wallet X going to have a BAYC transaction in the following 7 days (T, T+7 days) or not? It is simply Is Trading, however, this time we are only interested in transactions of one particular project, the BAYC. The thing with the Collection Related Metrics is that, these metrics are harder to predict as there are less data on them, mostly because for all of the collection related metrics, the data is filtered to include only the wallets that have a token of that specific project at the time of the snapshot. Additionally, the feature space we had initially was not sufficient to predict such metrics, which is why we decided to fetch BAYC listings from OpenSea, BAYC floor price from CoinGecko, and ETH/USD data from Yahoo! Finance. The predictive modelling process is the same as Is Trading, therefore we wont repeat the process over here. Only difference is that, on the market related metrics data was filtered to include the wallets that have at least 1 token holding, however on the collection related metrics data is filtered to include the wallets that have at least 1 token holding for the particular project.

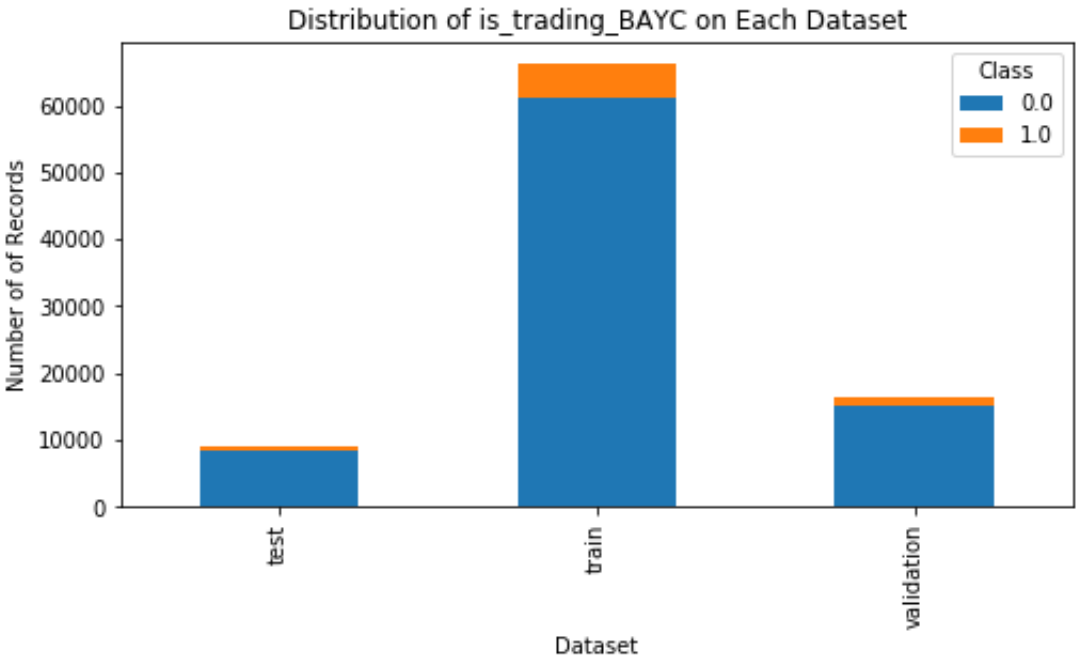

Figure 44: Is Trading BAYC Target Distribution Across Datasets

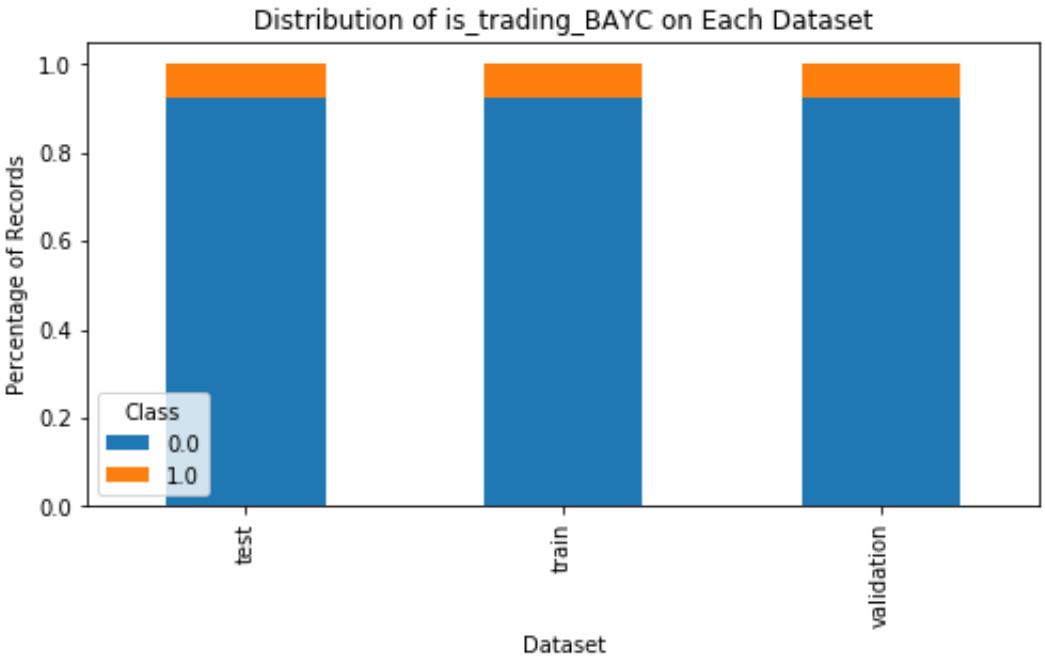

Figure 45: Is Trading BAYC Target Distribution Across Datasets

As can be seen from Figure 44, number of observations in the dataset is particularly lower when compared to market related metrics. Additionally, as can be seen from

Figure 45, the data is highly imbalanced as there are only 8% positive classes.

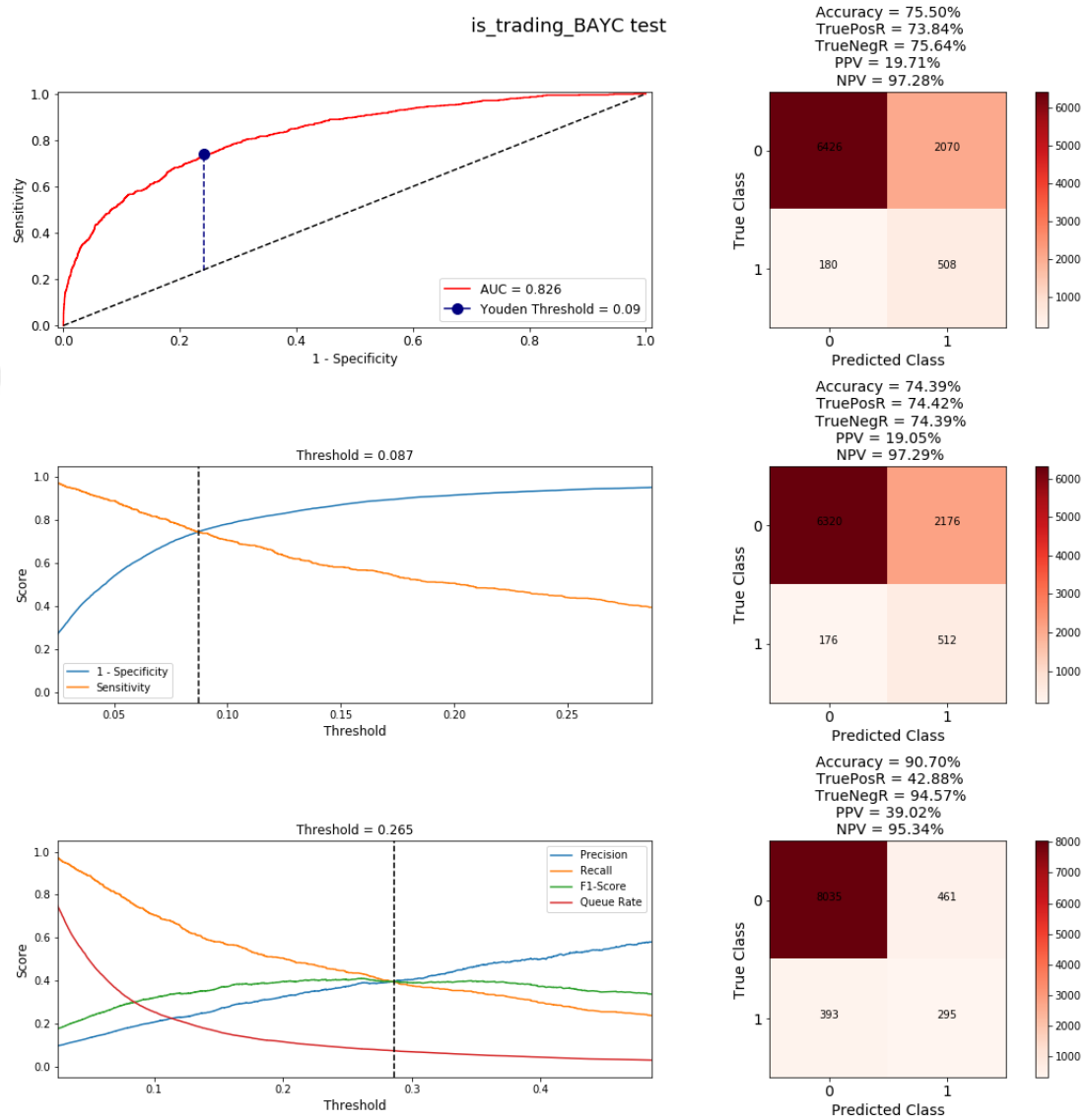

Figure 46: Is Trading BAYC Results on Test Set

As can be seen from Figure 46, we've achieved an AUC of 0.826 on the Test set. As mentioned earlier, prediction thresholds can have a great impact on the confusion matrix for the imbalanced problem. As you can see, if we pick the 0.265 threshold, which maximizes the F1 score, we would have a 39% Positive Predictive Value, 42% True Positive Rate, 94% True Negative Rate, and 95% Negative Predictive Value. However, if we were to choose the threshold 0.087, we would have a 19% Positive

Predictive Value, 74% True Positive Rate, 74% True Negative Rate, and 97% Negative Predictive Value.

If we were to create a marketing campaign for an NFT marketplace and want to target only wallets that are likely to trade a BAYC token in the next 7 days, we would choose a prediction threshold of 0.087. Without a predictive model, we would need to target all 9,184 wallets in the test set. However, with the help of the predictive model, we could target only 2,688 wallets, resulting in a 70% decrease in the cost of the campaign. This demonstrates the power and importance of using predictive models to optimize the processes in the NFT space.

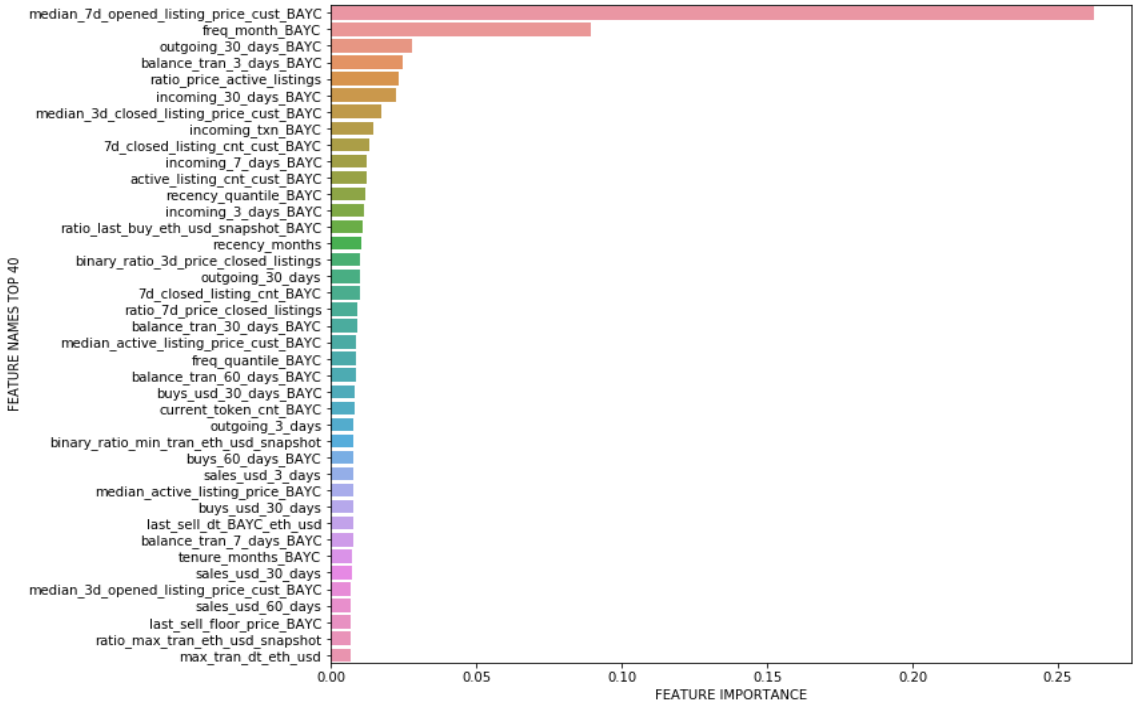

Figure 47: Is Trading BAYC Feature Importances

Feature importances of the model can be seen in Figure 47. The most important feature is median_7d_opened_listing_price_cust_BAYC, which means the median price of the BAYC listings that are listed by the wallet in the last 7 days. The second important feature is freq_month_BAYC which means the average monthly BAYC

transaction count of the wallet since the first BAYC buy. The third feature is outgoing_30_days_BAYC which means the number of BAYC tokens sent from the wallet in the last 30 days. The fourth feature is balance_tran_3days_BAYC which means the balance of the wallet in terms of BAYC tokens in the last 3 days. The fifth feature is ratio_price_active_listings which is the ratio of the median price of the active BAYC listings of the wallet to the median listing price of all BAYC listings. As expected, feature importances suggest that the listing features are important signals in terms of NFT trading.

6.3.2 Is Buying BAYC

The second collection related metric is Is Buying BAYC and it can be described as; looking at the status of Wallet X at time T, is Wallet X going to buy a BAYC token in the following 7 days (T, T+7 days) or not? It is simply Is Trading BAYC, however, this time we are only interested in buying transactions.

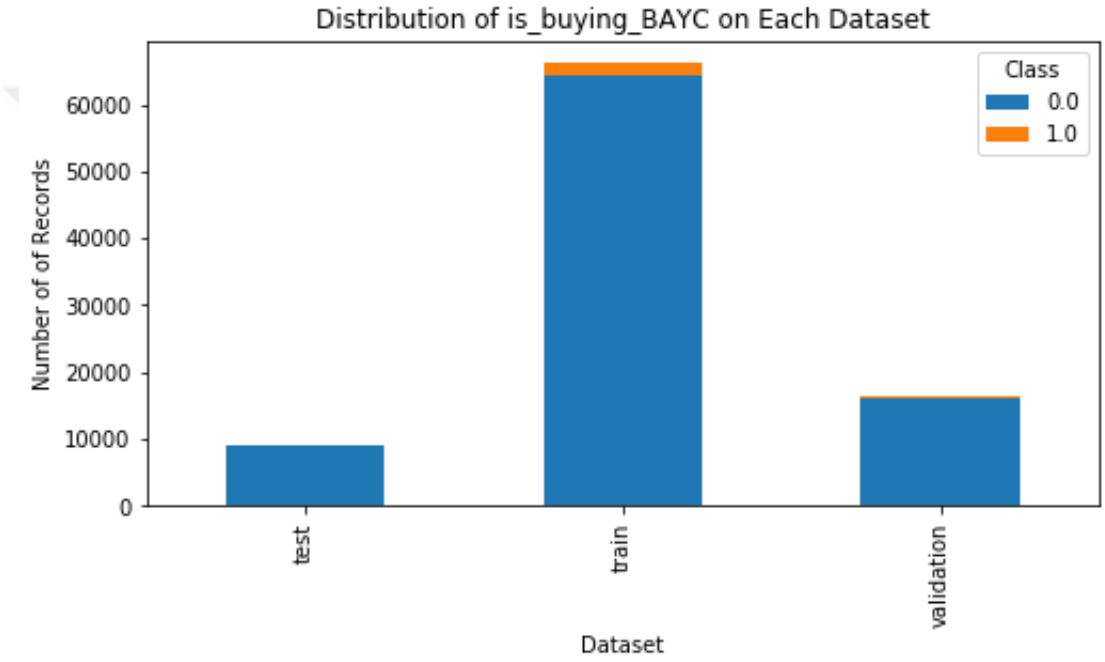

Figure 48: Is Buying BAYC Target Distribution Across Datasets

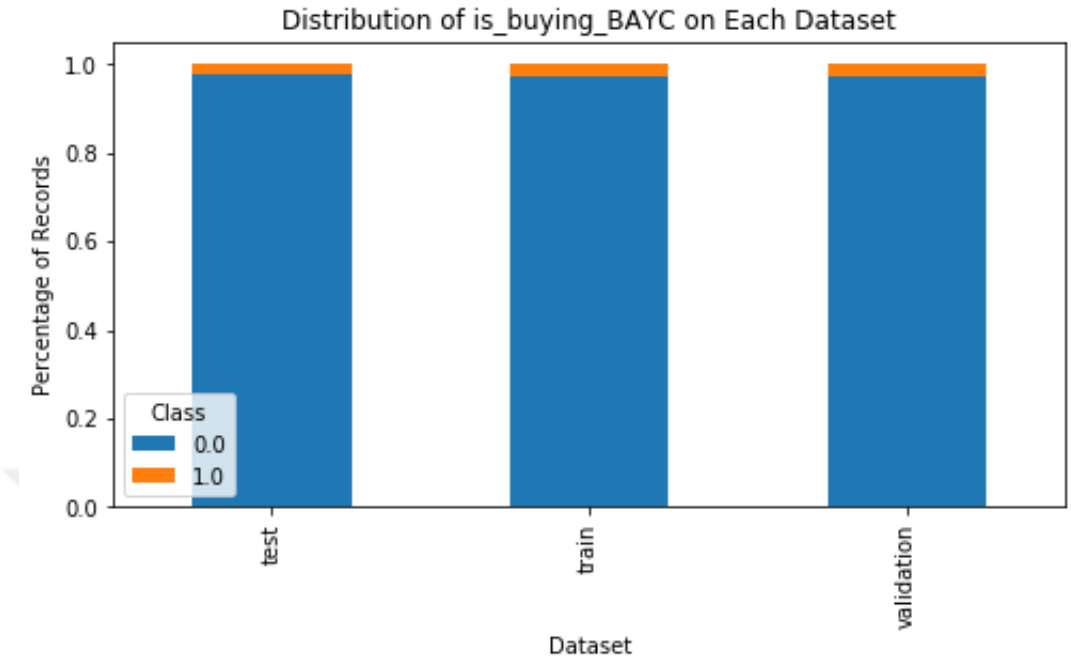

Figure 49: Is Buying BAYC Target Distribution Across Datasets

As can be seen from Figures 48 and 49, this problem is a highly imbalanced classification problem. The main reasons for the imbalance are; we are filtering the data to include only the wallets that have a BAYC token at the snapshot date, and people do not frequently buy multiple BAYC tokens because of the high prices.

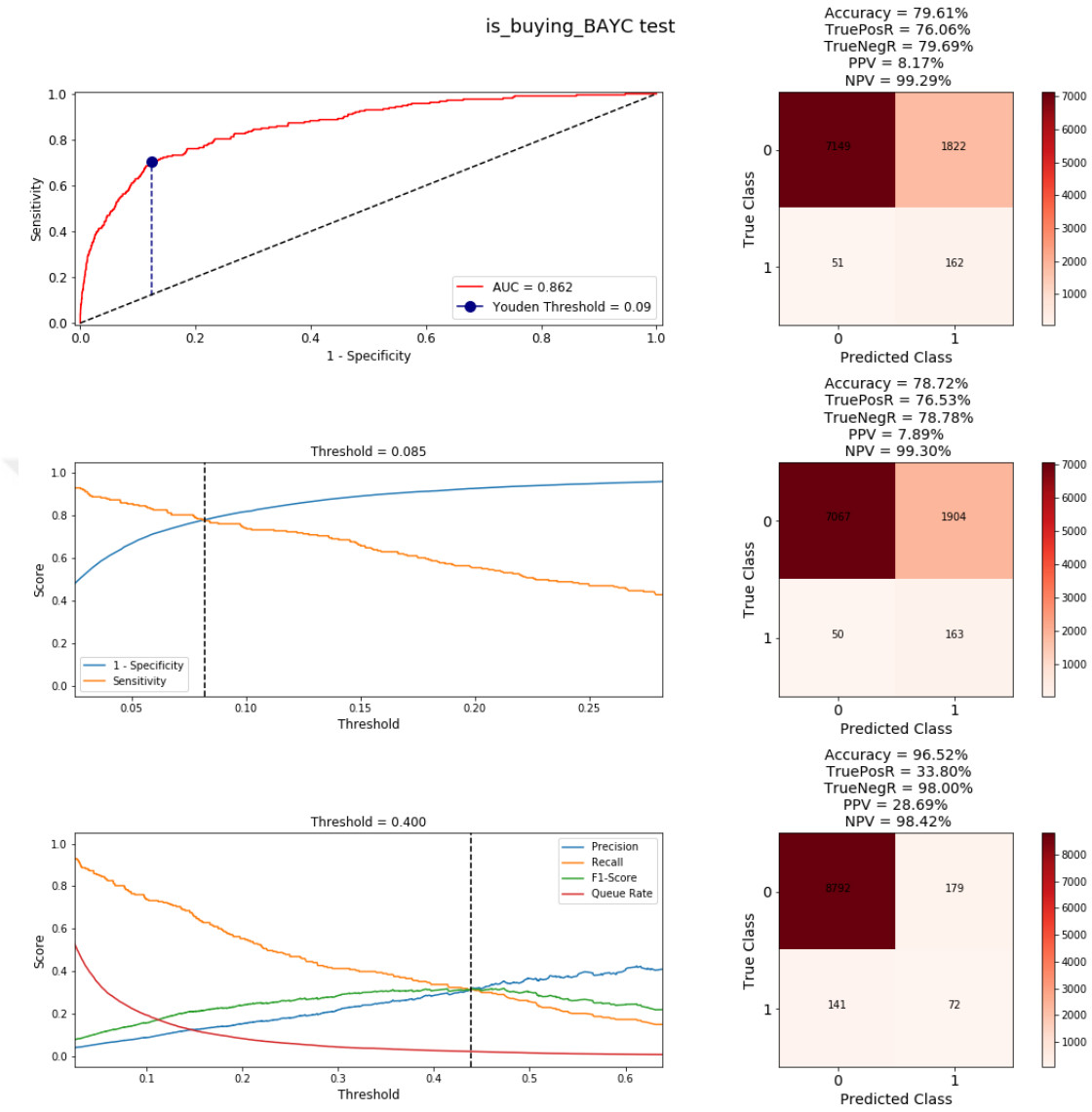

Figure 50: Is Buying BAYC Results on Test Set

As can be seen from the Figure 53, we've achieved an AUC of 0.862 in the Test set, and with a sacrifice on Positive Predictive Value and True Negative Rate we could achieve a True Positive Rate of 76% which means that we can correctly capture 76% of the positive cases.

6.3.3 Is Selling BAYC

The third and final collection related metric is Is Selling BAYC and it can be described as; looking at the status of Wallet X at time T, is Wallet X going to sell a BAYC token

in the following 7 days (T, T+7 days) or not? It is simply Is Trading BAYC, however, this time we are only interested in selling transactions. Again, this information could be valuable in the case of bidding on the listings of the sellers, as they are probably going to sale their tokens with a reasonable price.

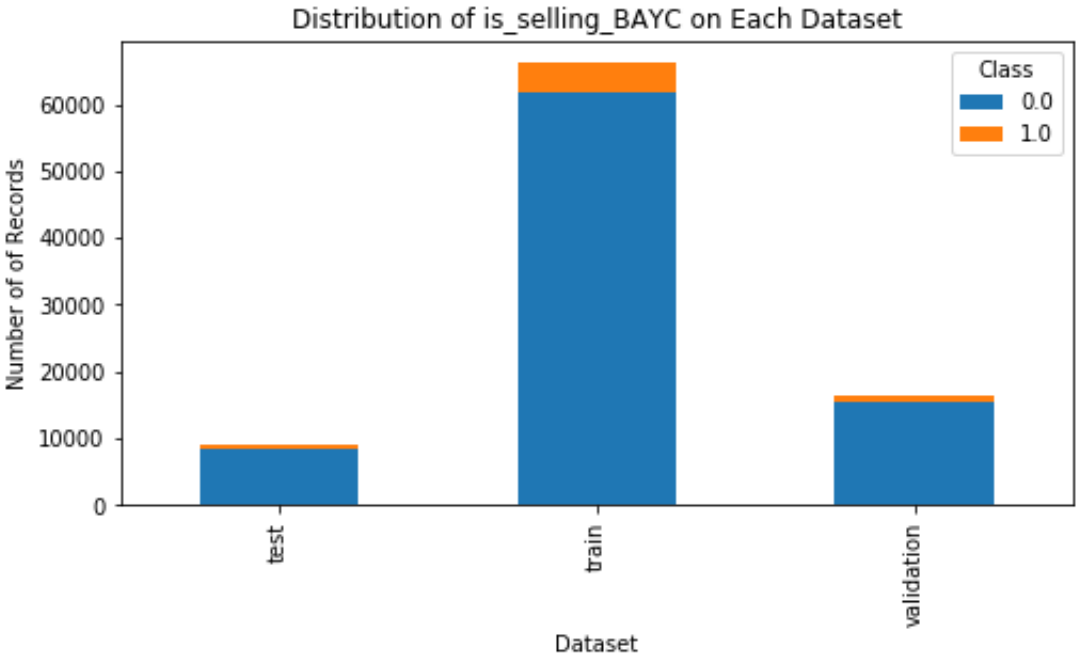

Figure 51: Is selling BAYC Target Distribution Across Datasets

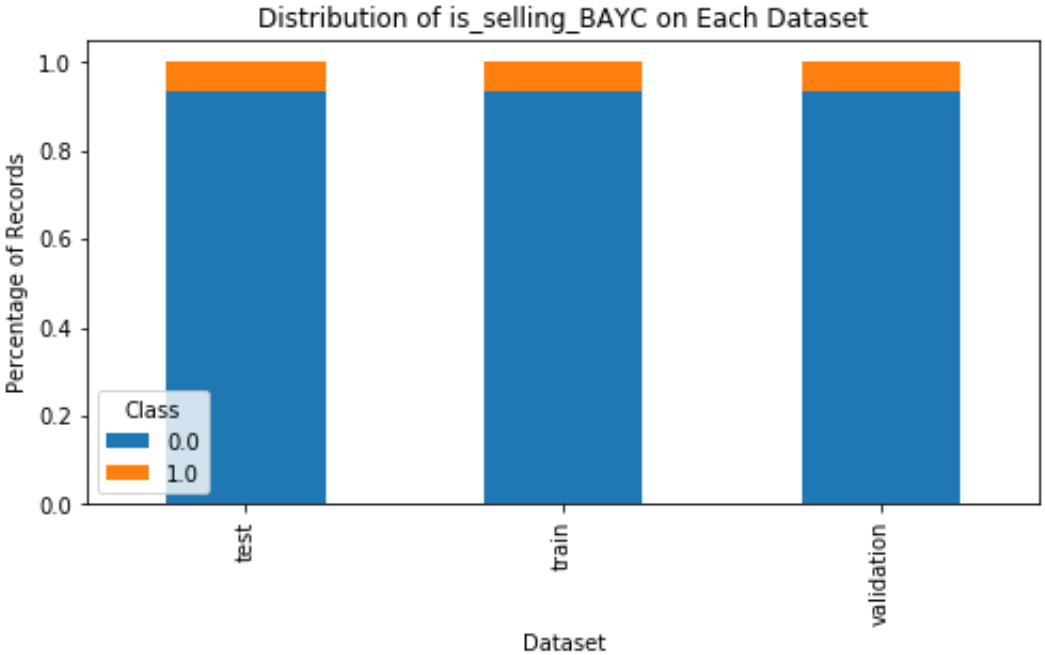

Figure 52: Is selling BAYC Target Distribution Across Datasets

As can be seen from Figures 51 and 52 this is another imbalanced problem and positive class constitutes only 7% of the data.

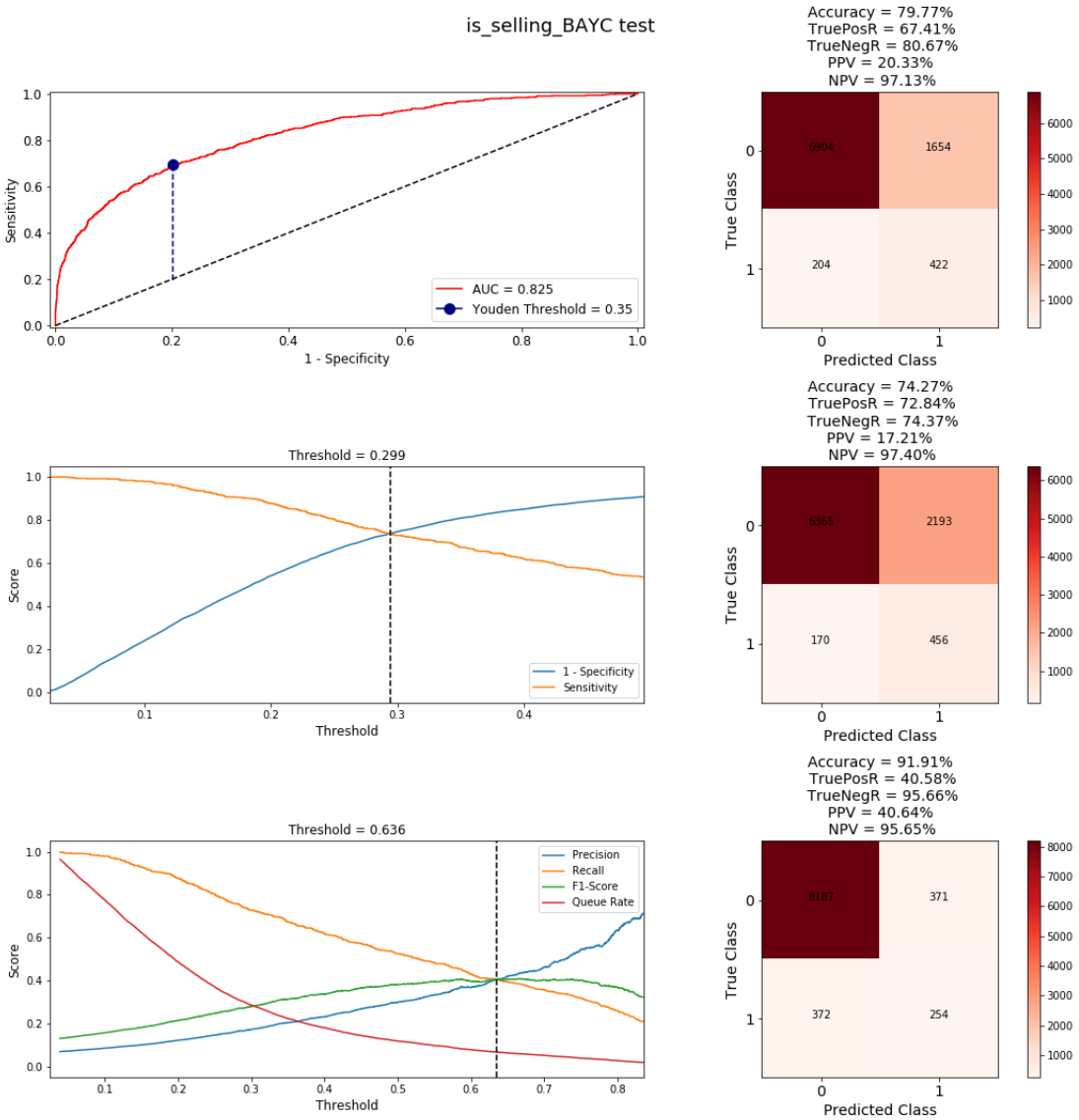

Figure 53: Is selling BAYC Results on Test Set

We could achieve 0.825 AUC score on the Test set, and with a sacrifice on True Negative Rate and Positive Predictive Value, we could correctly capture 72% of the positive classes.

CHAPTER VII

RESULTS AND CONCLUSIONS

In this thesis, we have performed a comprehensive analysis of the NFT market from various perspectives to gain insight into its dynamics. The study is divided into two main parts: a descriptive analysis of the market and a predictive analysis of wallet behavior. The goal of this research is to provide a better understanding of the market.

In the descriptive part of our analysis, we have explored the market using three different approaches allowing us to gain a deeper understanding of the market and its characteristics and providing valuable insights into its trends and dynamics.

- Descriptive Analysis of the Market

  In this part of the analysis, we have studied the major indicators of the market and found that the growth that started in 2021 stopped in 2022, and the number of new wallets entering the market decreased significantly. This might suggest that the market has reached a mature phase. However, it is not possible to conclude as there are concerns about a global recession, and markets around the world are also experiencing a decrease.

- Descriptive Analysis of Wallets

  When the transactional behaviors of the wallets are analyzed, we found that most of them have only one transaction and hold only one token, indicating that most of the users are inexperienced users that are exploring the market. To gain a better understanding of the wallet portfolio, we performed a segmentation of the user base based on their transactional behaviors. This allowed us to identify different segments with distinct characteristics and behaviors, providing valuable insights into the market.

- Descriptive Analysis of Wallets Focusing on Specific Collection

  We've also analyzed the buyers of the BAYC project, one of the most recognized and valuable projects in the NFT space, and seen that these users are heavily interested in the market and they tend to buy tokens from various NFT projects. We also discovered that BAYC buyers are active traders and have a strong interest in similar NFT projects. This suggests that they are experienced users who have a good understanding of the market.

On the predictive side of the analysis, we have explored the potential of wallet analytics in the NFT space and built predictive models for metrics that can help actors in the market, such as NFT marketplaces, NFT creators, and crypto lenders. These models provide predictive visibility on wallet behavior, allowing these actors to make more informed decisions. The predictive analysis is divided into two main parts: market-level metrics and project-level metrics. These two approaches provide different perspectives on the market and help us understand the trends and dynamics at different levels of granularity.

- Market Level Metrics

  On market-level metrics, we've proposed three different metrics;

  "Is Trading", "Is Buying a New Project" and "Is Exiting a Project". These metrics are on the market level which means they are inclusive of all NFT projects and have the potential to help actors in the NFT space to optimize their processes. We've used tree-based machine learning models to predict the behaviors of the wallets and achieved AUC scores of 0.875, 0.849, and 0.834 respectively.

- Project Level Metrics

  On project-level metrics, we've proposed three different metrics;

  "Is Trading BAYC", "Is Buying BAYC" and "Is Selling BAYC". These metrics are specific to a collection and also have the potential to help NFT marketplaces, NFT creators, and Crypto Lenders to optimize their processes. We've used tree-based machine learning models to predict the behaviors of the wallets and achieved AUC scores of 0.826, 0.862, and 0.825 respectively.

In summary, we have performed a detailed analysis of the NFT market, with a focus on wallet analytics, which has not been studied in the literature before, and provided valuable insights into its trends and dynamics. We have proposed some meaningful metrics that can help actors in the NFT space by providing visibility on the transactional behavior of wallets, and we have built machine-learning models to predict these metrics.

Our future work will be based on improving the performance of the predictive models by incorporating token-level features into the dataset. Currently, our models are token agnostic, which means they do not consider the unique characteristics of each token, such as rarity or fair value. Additionally, we will expand the feature space of the models by adding bidding data on token listings, which will enhance the predictive capabilities of machine learning models and allow us to make more accurate predictions.

REFERENCES

[1] NFTGO, "Global NFT Market Overview." https://nftgo.io/analytics/market-overview. Accessed 15 December, 2022.

[2] CoinTelegraph, "Blockchains vie for NFT market, but Ethereum still dominates — Report." https://cointelegraph.com/news/blockchains-vie-for-nft-market-but-ethereum-still-dominates-report. Accessed 15 December, 2022.

[3] Z. Zheng, S. Xie, H.-N. Dai, X. Chen, and H. Wang, "Blockchain Challenges and Opportunities: A Survey," *International Journal of Web and Grid Services*, vol. 14, no. 4, pp. 352–375, 2018.

[4] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends," in *IEEE International Congress on Big Data (BigData Congress)*, pp. 557–564, 2017.

[5] G. W. Peters and E. Panayi, *Understanding Modern Banking Ledgers Through Blockchain Technologies: Future of Transaction Processing and Smart Contracts on the Internet of Money*, pp. 239–278. Cham: Springer International Publishing, 2016.

[6] A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papamanthou, "Hawk: The Blockchain Model of Cryptography and Privacy-Preserving Smart Contracts," in *IEEE Symposium on Security and Privacy (SP)*, pp. 839–858, 2016.

[7] B. W. Akins, J. L. Chapman, and J. M. Gordon, "A Whole New world: Income Tax Considerations of The Bitcoin Economy," *Pitt. Tax Rev.*, vol. 12, p. 25, 2014.

[8] Y. Zhang and J. Wen, "An IoT Electric Business Model Based on The Protocol of Bitcoin," in *18th International Conference on Intelligence in Next Generation Networks*, pp. 184–191, IEEE, 2015.

[9] M. Sharples and J. Domingue, "The Blockchain and Kudos: A Distributed System For Educational Record, Reputation and Reward," in *European Conference on Technology Enhanced Learning*, pp. 490–496, Springer, 2016.

[10] C. Noyes, "Bitav: Fast Anti-malware by Distributed Blockchain Consensus and Feedforward Scanning," *arXiv preprint arXiv:1601.01405*, 2016.

[11] A. Gervais, G. O. Karame, K. Wüst, V. Glykantzis, H. Ritzdorf, and S. Capkun, "On the Security and Performance of Proof of Work Blockchains," in *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, (New York, NY, USA), p. 3–16, Association for Computing Machinery, 2016.

[12] F. Saleh, "Blockchain without Waste: Proof-of-Stake," *The Review of Financial Studies*, vol. 34, pp. 1156–1190, 07 2020.

[13] M. Andoni, V. Robu, D. Flynn, S. Abram, D. Geach, D. Jenkins, P. McCallum, and A. Peacock, "Blockchain Technology in The Energy Sector: A Systematic Review of Challenges and Opportunities," *Renewable and Sustainable Energy Reviews*, vol. 100, pp. 143–174, 2019.

[14] Q. Wang, R. Li, Q. Wang, and S. Chen, "Non-Fungible Token (NFT): Overview, Evaluation, Opportunities and Challenges," *CoRR*, vol. abs/2105.07447, 2021.

[15] M. Franceschet, G. Colavizza, T. Smith, B. Finucane, M. L. Ostachowski, S. Scalet, J. Perkins, J. Morgan, and S. Hernández, "Crypto Art: A Decentralized View," *Leonardo*, vol. 54, pp. 402–405, 08 2021.

[16] H. Bao and D. Roubaud, "Non-Fungible Token: A Systematic Review and Research Agenda," *Journal of Risk and Financial Management*, vol. 15, no. 5, 2022.

[17] Nansen, "NFT Market Statistics and Trends." https://www.nansen.ai/guides/nft-statistics-2022. Accessed 15 December, 2022.

[18] Christies, "Beeple NFT Sale." https://www.christies.com/about-us/press-archive/details?PressReleaseID=9970. Accessed 15 December, 2022.

[19] Larvalabs, "CryptoPunk Website." https://www.larvalabs.com/cryptopunks. Accessed 15 December, 2022.

[20] CryptoKitties, "CryptoKitties Website." https://www.cryptokitties.co/. Accessed 15 December, 2022.

[21] B. A. Y. Club, "BAYC Website." https://boredapeyachtclub.com/. Accessed 15 December, 2022.

[22] Y. Freund, R. Schapire, and N. Abe, "A Short Introduction to Boosting," *Journal-Japanese Society For Artificial Intelligence*, vol. 14, no. 771-780, p. 1612, 1999.

[23] T. Chen and C. Guestrin, "Xgboost: A Scalable Tree Boosting System," in *Proceedings of The 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

[24] S. Ramraj, N. Uzir, R. Sunil, and S. Banerjee, "Experimenting XGBoost algorithm for prediction and classification of different datasets," *International Journal of Control Theory and Applications*, vol. 9, no. 40, 2016.

[25] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results," in *11th International Conference on Information and Communication Systems (ICICS)*, pp. 243–248, 2020.

[26] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Special Issue on Learning From Imbalanced Data Sets," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 1–6, 2004.

[27] X.-Y. Liu, J. Wu, and Z.-H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 2, pp. 539–550, 2009.

[28] I. Tomek, "A Generalization of The k-NN Rule," *IEEE Transactions on Systems, Man, and Cybernetics*, no. 2, pp. 121–126, 1976.

[29] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python Toolbox to Tackle The Curse of Imbalanced Datasets in Machine Learning," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 559–563, 2017.

[30] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[31] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive Synthetic Sampling Approach For Imbalanced Learning," in *IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, pp. 1322–1328, IEEE, 2008.

[32] Flipside, "Flipside Website." https://flipsidecrypto.xyz/. Accessed 15 December, 2022.

[33] Python, "Python Website." https://www.python.org/. Accessed 15 December, 2022.

[34] GitHub, "GitHub Website." https://github.com/. Accessed 15 December, 2022.

[35] Opensea, "Opensea Marketplace." https://opensea.io/. Accessed 15 December, 2022.

[36] Blur, "Blur Marketplace." https://blur.io/. Accessed 15 December, 2022.

[37] Looksrare, "Looksrare Marketplace." https://looksrare.org/. Accessed 15 December, 2022.

[38] Rarible, "Rarible Marketplace." https://rarible.com/. Accessed 15 December, 2022.

[39] x2y2, "x2y2 Marketplace." https://x2y2.io/. Accessed 15 December, 2022.

[40] sudoswap, "SudoSwap Marketplace." https://sudoswap.xyz/. Accessed 15 December, 2022.

[41] Nftx, "NFTX Marketplace." https://nftx.io/. Accessed 15 December, 2022.

APPENDIX A

APPENDIX

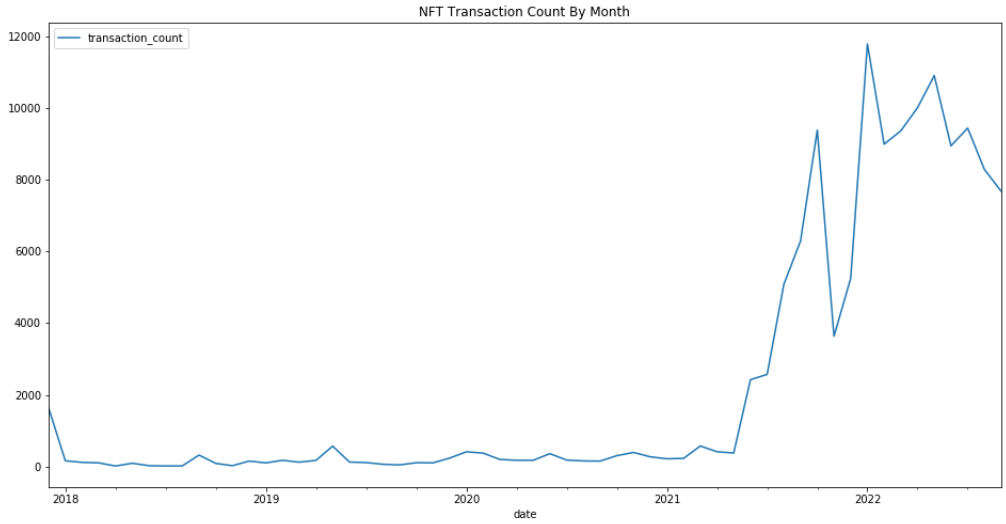

Figure 54: Monthly Transaction Counts of the Sample

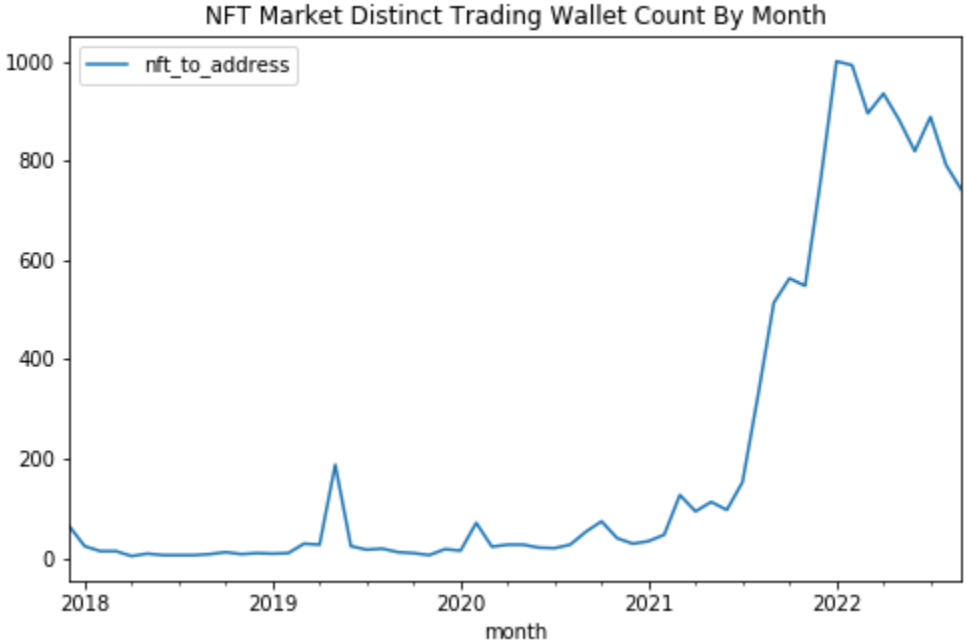

Figure 55: Monthly Trading Wallet Counts of the Sample

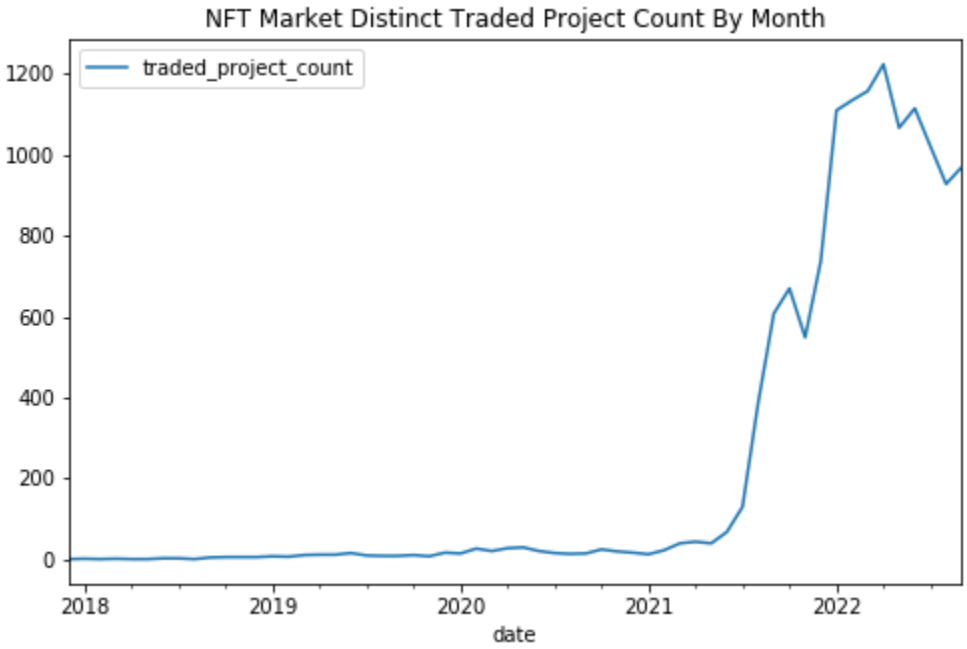

Figure 56: Monthly Traded Project Counts of the Sample

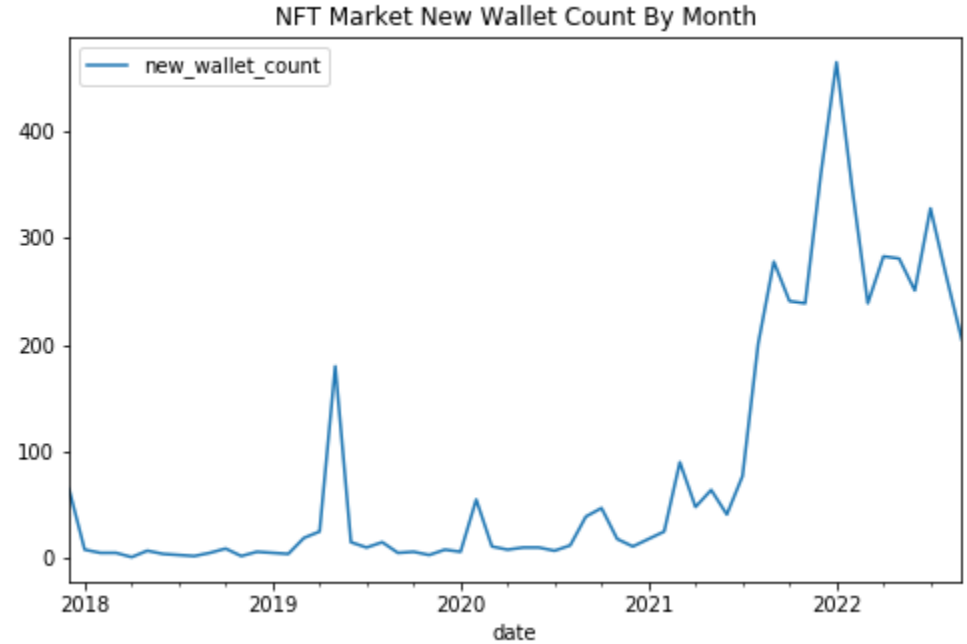

Figure 57: Monthly New Wallet Counts of the Sample

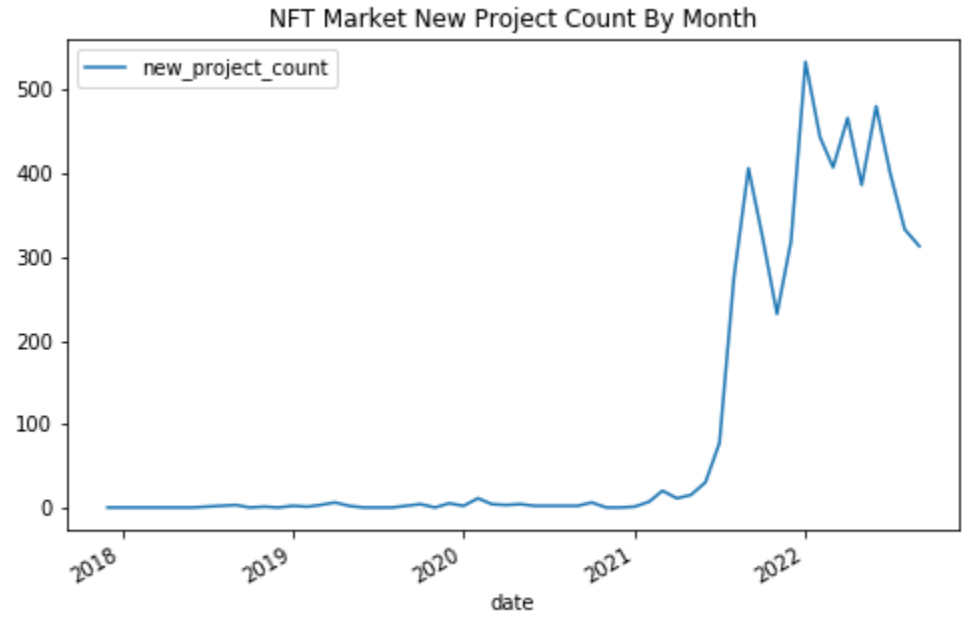

Figure 58: Monthly New Project Counts of the Sample

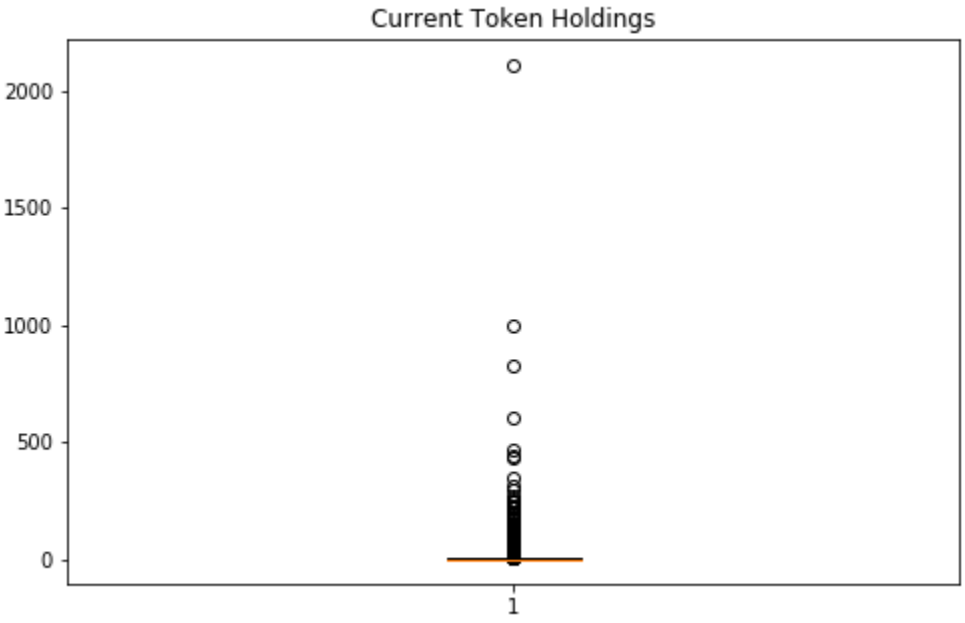

Figure 59: Boxplot of Wallets Based on Token Holdings

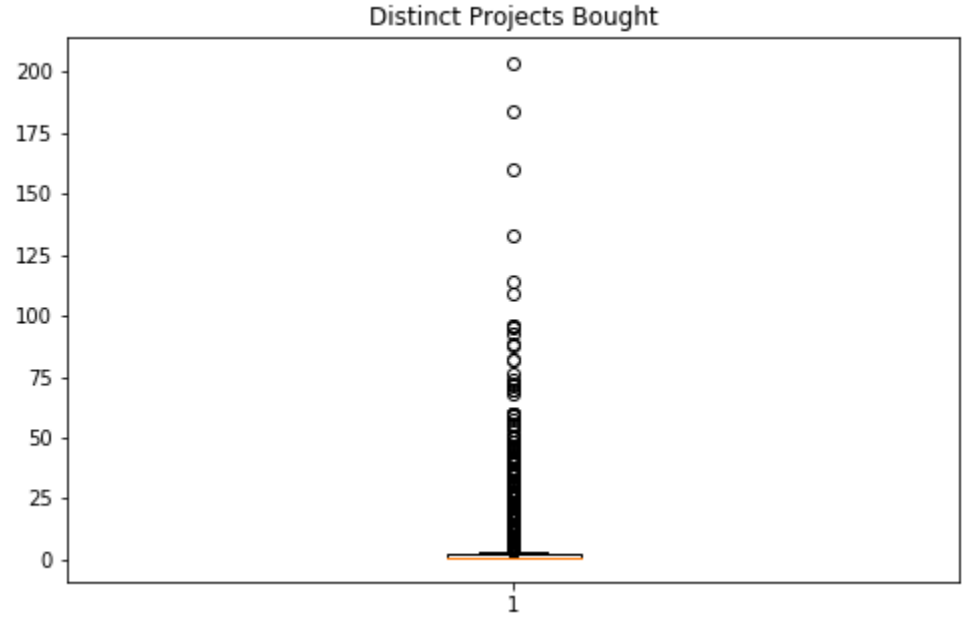

Figure 60: Boxplot of Wallets Based on Number of Distinct Projects Bought

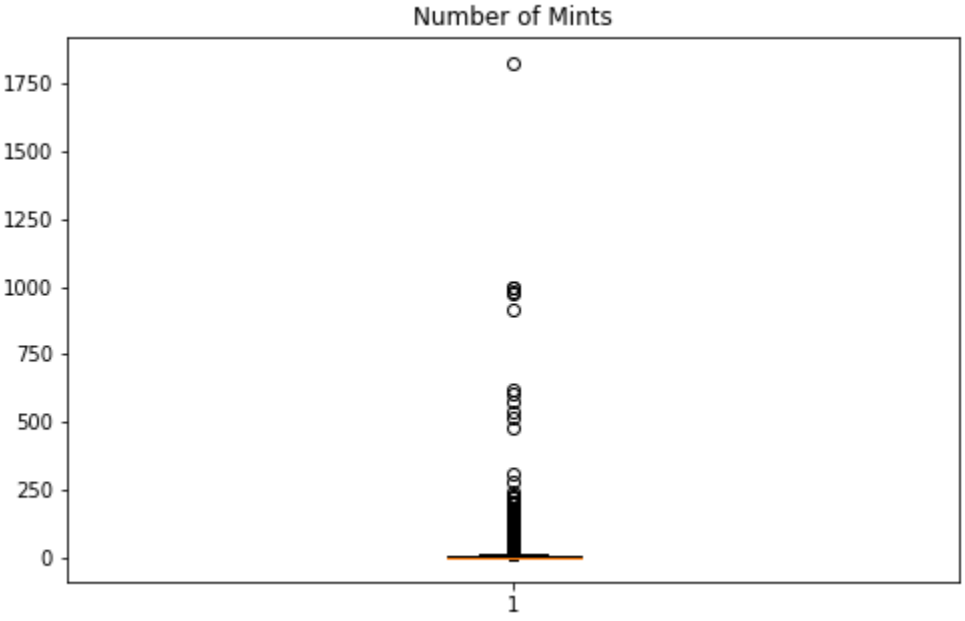

Figure 61: Boxplot of Wallets Based on Number of Mints

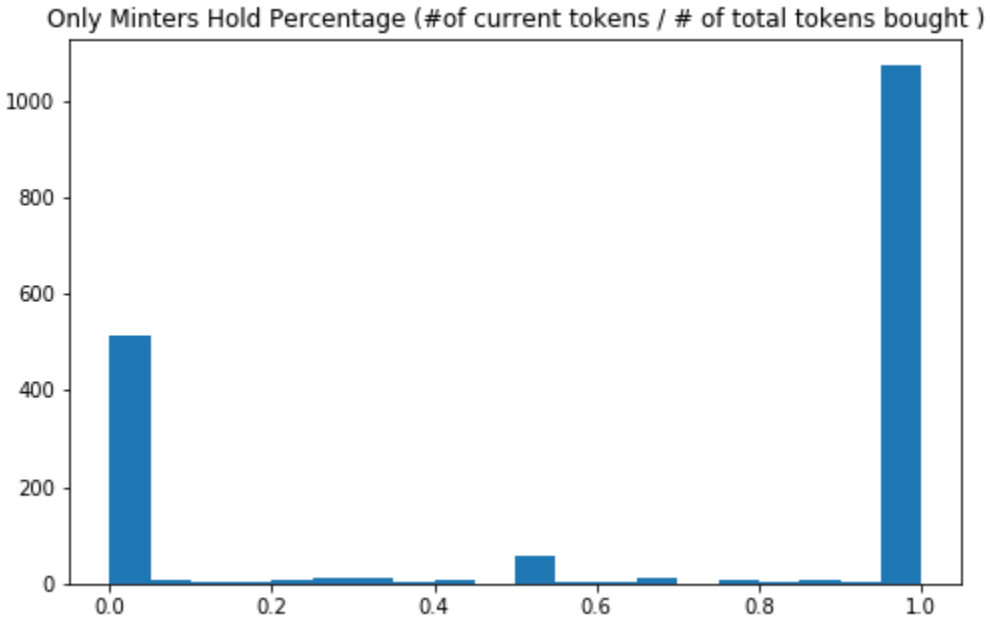

Figure 62: Histogram of Only Mint Wallets Based Token Holding Percentage

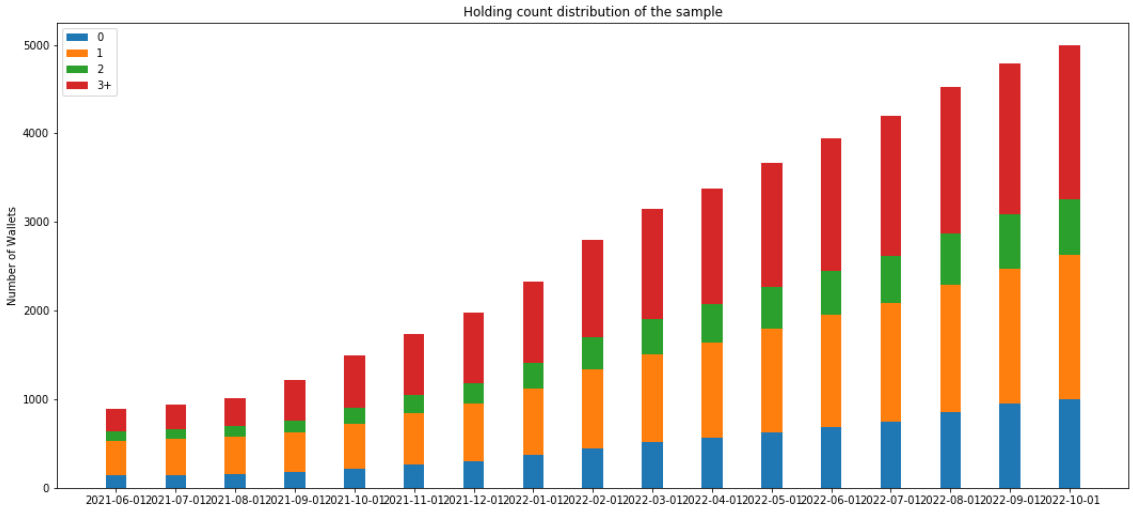

Figure 63: Distribution of the sample in terms of token holdings over time - Barplot

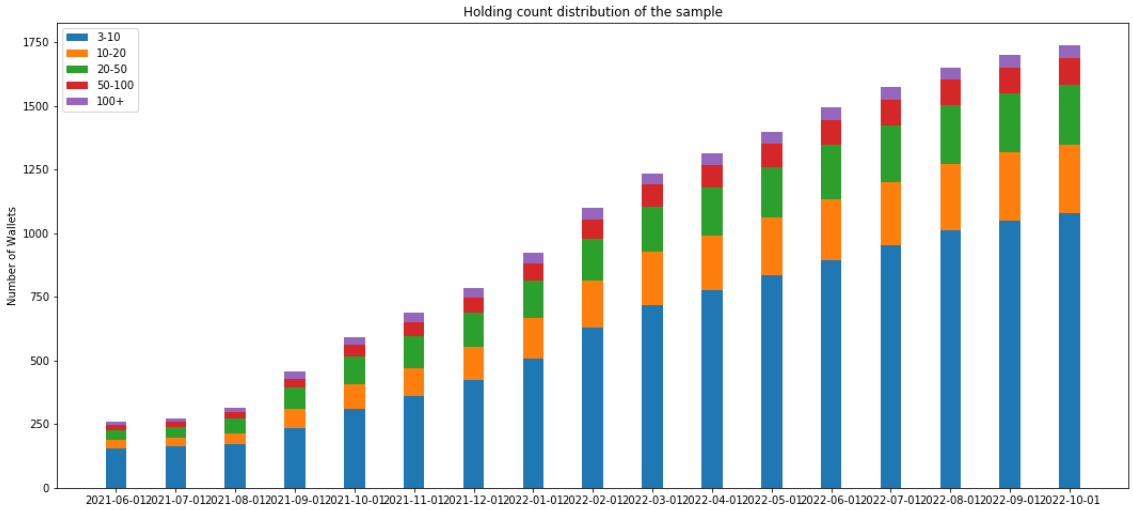

Figure 64: Distribution of the sample in terms of 3+ token holdings over time - Barplot

```
df['Segment']=df.apply(lambda row:
                       'Trader' if row['only_mint']==0 and row['distinct_projects_bought']>1 and row['hold_percentage']<0.8 else
                       'NFT Collector' if row['only_mint']==0 and row['distinct_projects_bought']>1 and row['hold_percentage']>=0.8 else
                       'Project Collector' if row['only_mint']==0 and row['distinct_projects_bought']==1 and row['hold_percentage']>=0.8 else
                       'Project Trader' if row['only_mint']==0 and row['distinct_projects_bought']==1 and row['hold_percentage']<0.8 else
                       'Hybernating Project Minters' if row['only_mint']==1 and row['distinct_projects_bought']==1 and row['hold_percentage']>=0.8 else
                       'Trader Project Minters' if row['only_mint']==1 and row['distinct_projects_bought']==1 and row['hold_percentage']<0.8 else
                       'Mint Sellers' if row['only_mint']==1 and row['distinct_projects_bought']>1 and row['hold_percentage']<0.8 else
                       'Mint Collectors' if row['only_mint']==1 and row['distinct_projects_bought']>1 and row['hold_percentage']>=0.8 else
                       'Other',axis=1)
```

Figure 65: Definition of segments

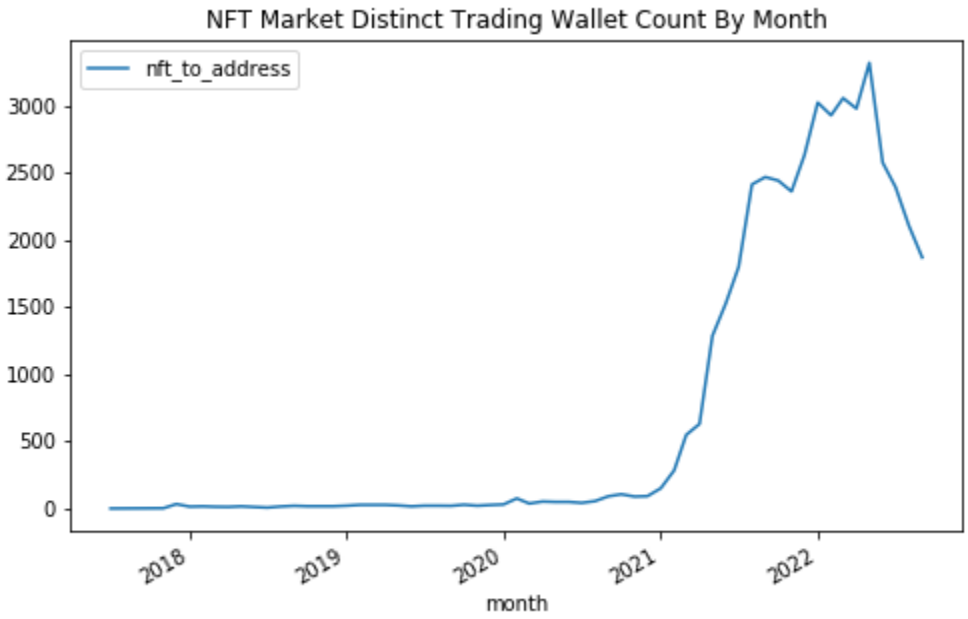

Figure 66: Monthly Trading Wallet Count of the BAYC Sample

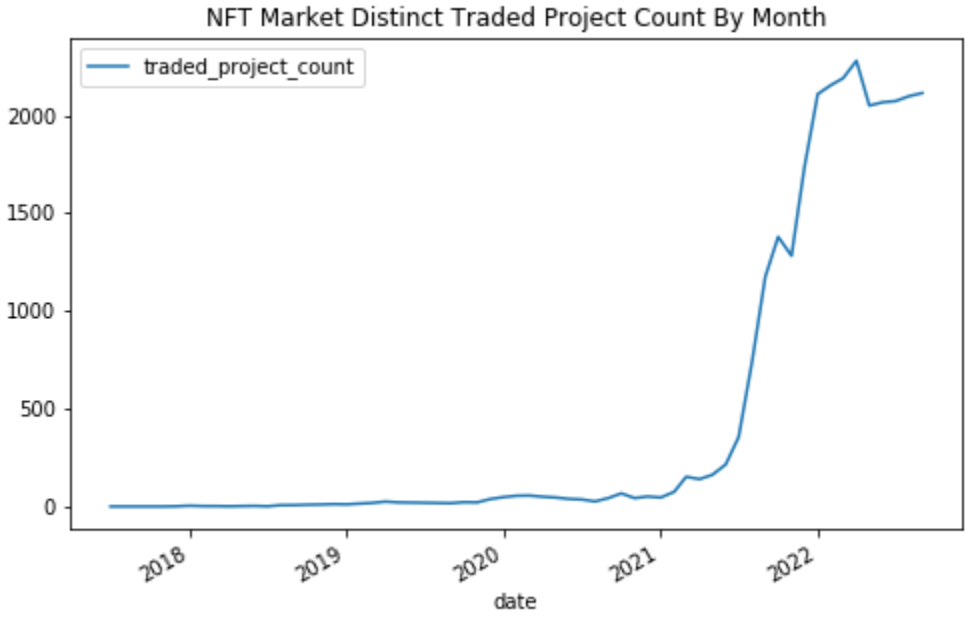

Figure 67: Monthly Traded Project Count of the BAYC Sample

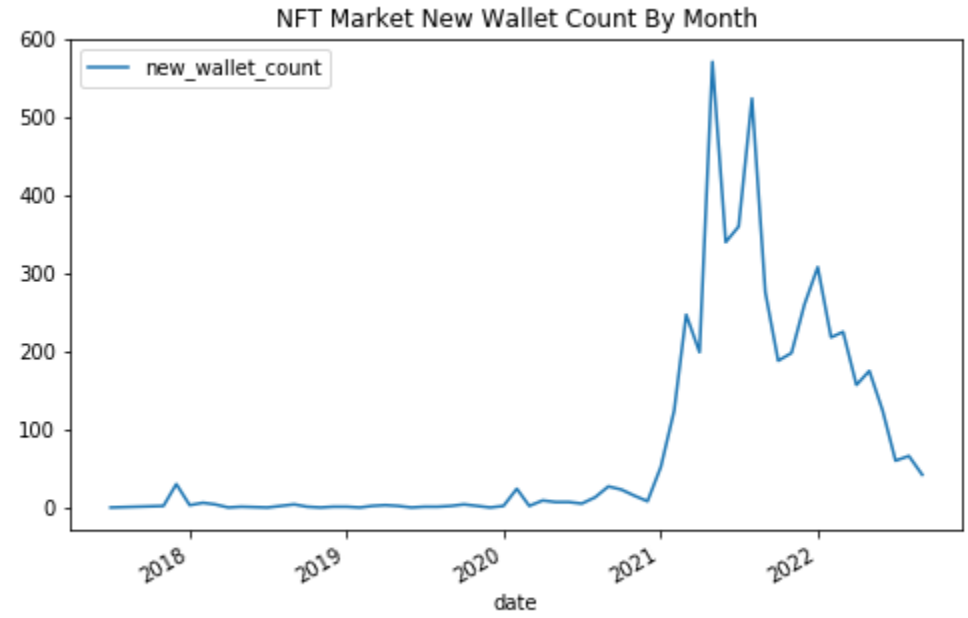

Figure 68: Monthly New Wallet Count of the BAYC Sample

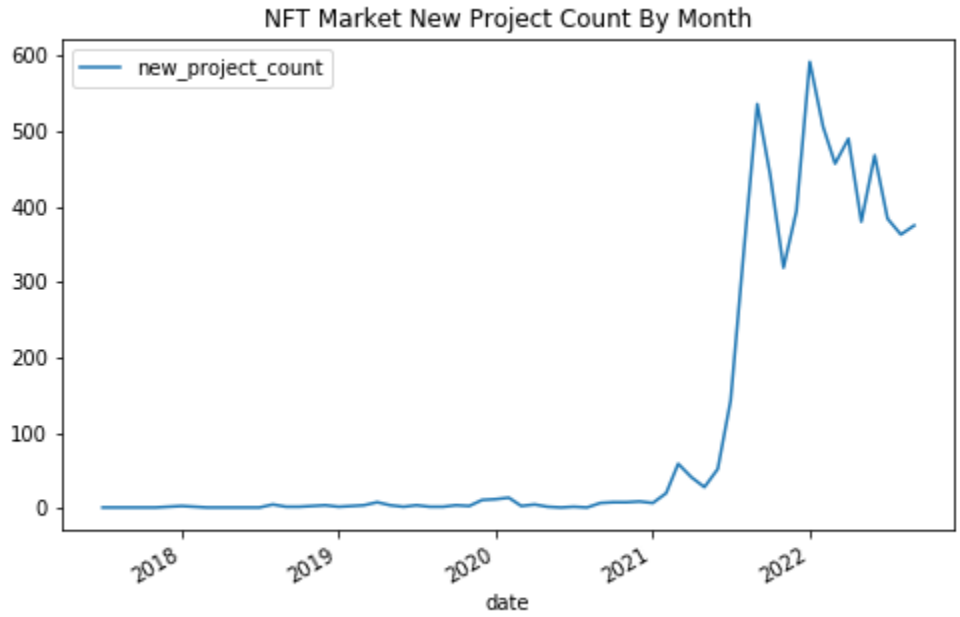

Figure 69: Monthly New Project Count of the BAYC Sample

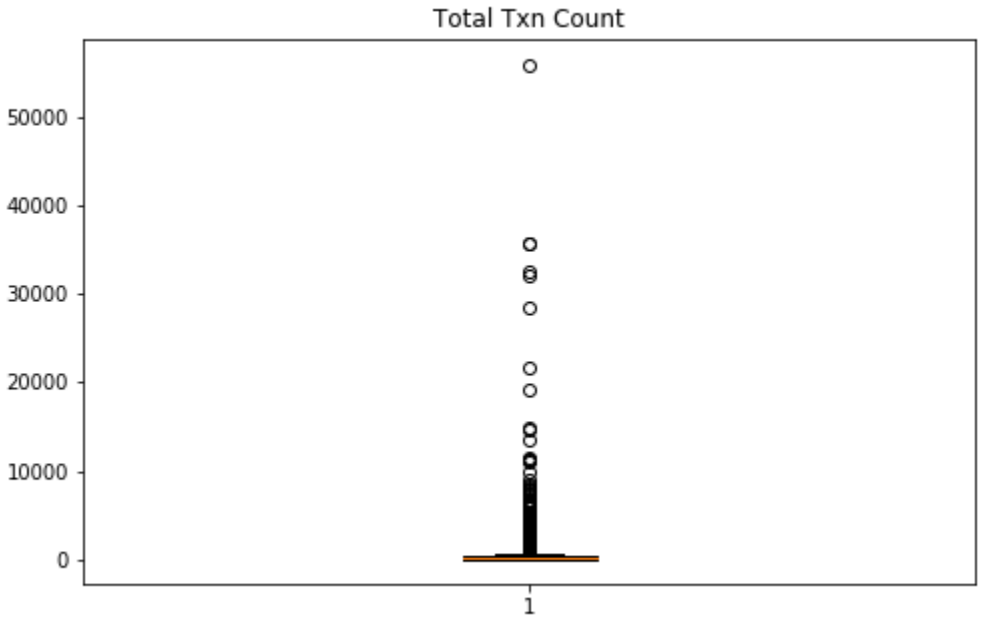

Figure 70: Boxplot of Wallets Based on Transaction Count, BAYC Sample

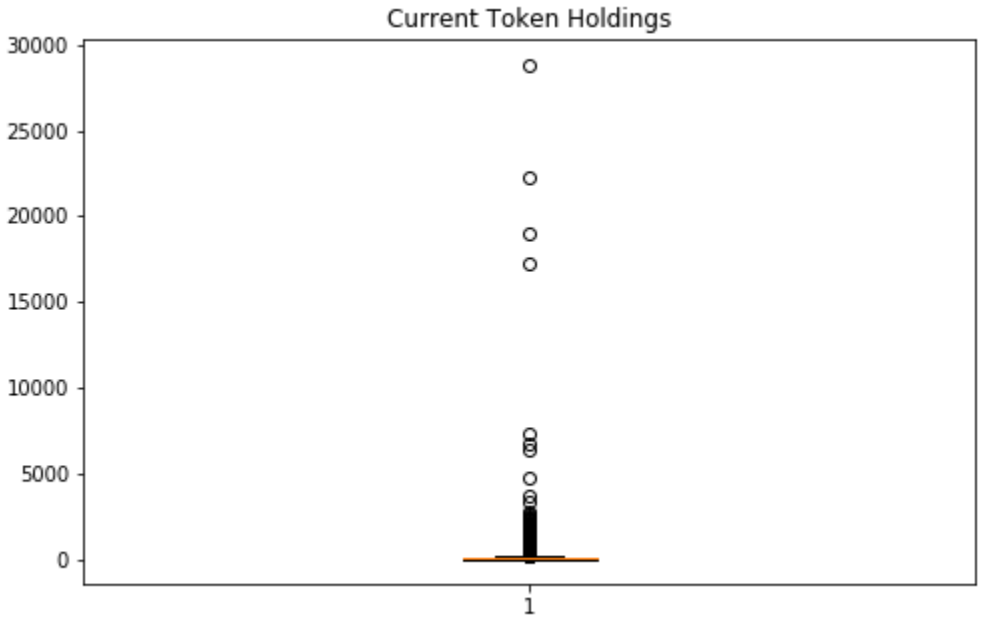

Figure 71: Boxplot of Wallets Based Token Holdings, BAYC Sample

VITA

Onur Can Çabuk received his bachelor's degree in Management Information Systems from Boğaziçi University in 2019. Since his graduation, he has been working in different industries in data scientist and data analyst roles. He joined the Master of Science program in Data Science at Özyegin University in 2020. Currently, Onur Can is working as Data Scientist at Zeplin. His research focuses on machine learning, deep learning, and customer analytics.