

Project Study: NFT Marketplace and the Bored Ape Yacht Club Collection

Yuxuan Zhao yz4397, Yifeng Wang yw3933, Donglin Gu dg3294,
Gehui Liang gl2800, Shifeng Zhang sz3104

May 8, 2023

1 Introduction

In this project study, we aim to explore the NFT marketplace, with a particular focus on the Bored Ape Yacht Club (BAYC) collection. The BAYC is a popular NFT collection of 10,000 unique digital apes, which have become highly sought after by collectors and investors. Our study will examine the 2000 random BAYC collection and its impact on the NFT marketplace, as well as the broader trends in the NFT market.

2 Research Questions

Our research will seek to answer the following questions:

- What is the current market value of the Bored Ape Yacht Club collection, and how does it fluctuate recently?
- What are the key players of attributes in the Bored Ape Yacht Club collection, and how have they influenced the market?
- What is the relationship between the collection's rarity and its sales price?
- Is there a separation between the NFTs based on their attributes?
- What are the popular attributes in the collection, and how does it affect the price of the NFT?

3 Methodology

To answer our research questions, we conducted a literature review of existing research on the Bored Ape Yacht Club collection and the NFT marketplace. We also analyzed data from NFT marketplaces such as OpenSea tools to better understand the current state of the market.

Since we could not acquire a valid API key for OpenSea API, we have decided to use NFT port instead, which is a free API, but with less information and useful functions.

We implemented the 'Retrieve Contract NFTs' function as a primary method to retrieve all useful information like metadata, last sale price, and rarity related to the BAYC collection. An example code is provided below.

```
import requests

url = "https://api.nftport.xyz/v0/nfts/0xbc4ca0eda7647a8ab7c2061c2\
e118a18a936f13d?chain=ethereum&page_number=1&page_size=50\
&include=metadata&refresh_metadata=\
false"

headers = {"accept": "application/json"}

response = requests.get(url, headers=headers)

print(response.text)
```

To effectively analyze the 2000 records retrieved from the NFT port API about the Bored Ape Yacht Club (BAYC), it is important to preprocess the data to ensure that it is in a clean and usable format. This preprocessing process includes flattening the JSON text into columns, which involves transforming complex nested data structures into a tabular format that is easier to work with. For example, the metadata field in the JSON record may contain nested attributes, such as "earings: Silver". To flatten this data, we create a new column with the name "metadata earrings" and a value of "silver". Any unknown attributes are represented as None values in the corresponding column.

Aims to find the trends of rarity and price of BAYC NFTs, two encoding techniques were applied for the categorical variables: one-hot encoding and ordinal encoding; and three machine learning models were used for the prediction tasks: XGBoost, Random Forest, and Linear Regression.

4 Results

4.1 Relationship between Daily Volume and Daily Sales: Insights and Market Trends

We extracted transaction data from OpenSea, a leading NFT marketplace, including daily sales, daily average price, and daily volume. Our objective was to identify trends and patterns in the BAYC market based on these properties. Through our visualizations, we investigated the relationship between daily volume and daily sales for BAYC NFTs. Daily volume refers to the total monetary

value of all BAYC NFT transactions on a given day, while daily sales represent the total number of transactions. We generated a scatter plot to analyze the correlation between these two variables, which provided insights into market dynamics and demand for BAYC NFTs. By understanding the correlation between daily volume and daily sales, collectors, artists, and investors can better comprehend the BAYC market trends and make more informed decisions regarding their involvement in this rapidly expanding digital art space.

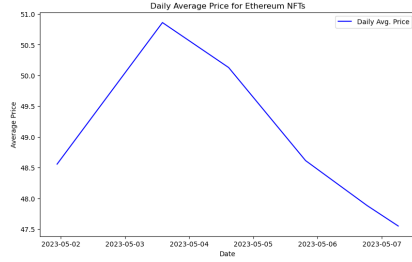


Figure 1: Average Price Fluctuation

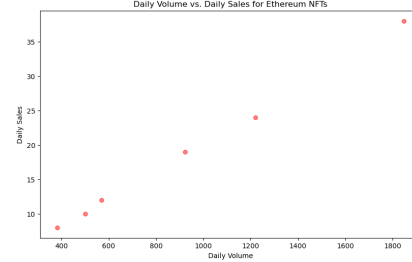


Figure 2: Daily Sales Fluctuation

4.2 Clustering

UMAP (Uniform Manifold Approximation and Projection) is a powerful dimensionality reduction technique commonly used for visualizing high-dimensional data in lower dimensions. Here is a figure of Visualizing UMAP

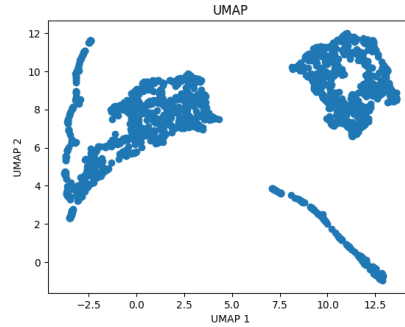


Figure 3: Cluster

Records about the collection can be separated into three clusters from a higher dimension. Meaning there are higher dimensional separations between the data points, at the same time, there are similarities within the cluster.

4.3 Key Players of Attributes in BAYC Collection

- Background: from the Figure 1, we can know that grey is the least popular background in the market.
- Clothes: from the Figure 2, we can know that Striped tee is the most popular clothes in the market, and Black suit is the least popular clothes in the market.
- Earring: from the Figure 3, we can know that Silver Hoop and Silver Stud are the most popular two earrings in the market, and Diamond Stud is the least popular clothes in the market.
- Eyes: from the Figure 4, we can know that Bored eyes are the most popular eyes in the market.
- Fur: from the Figure 5, we can know that Brown, Black and Dark Brown are top 3 popular color for fur in the market, and solid gold is the least popular clothes in the market.
- Hat: from the Figure 6, we can know that Beanie is the most popular hat in the market, and Laurel Wreath is the least popular hat in the market.
- Mouth: from the Figure 7, we can know that Bored and Bored Unshaven are the most popular Mouths in the market.
- In summary, our analysis of the Bored Ape Yacht Club NFT collection reveals distinct preferences among collectors for specific attributes. The least popular background color is grey, while Striped tee emerges as the most popular clothing option, and Black suit is the least favored. In terms of earrings, Silver Hoop and Silver Stud are the most sought-after, with Diamond Stud being the least popular. Bored eyes stand out as the preferred eye style. For fur colors, Brown, Black, and Dark Brown lead in popularity, while Solid Gold is the least favored. Among hats, the Beanie is the top choice, and Laurel Wreath ranks as the least popular. Lastly, Bored and Bored Unshaven are the most desirable mouth attributes. This analysis underscores the importance of understanding collector preferences when designing, marketing, and trading NFTs within the digital art and collectibles space.

4.4 Relationship between the Rarity and its Average Sales Price

To investigate the relationship between the collection’s rarity and its average sales price, we first group the BAYC collection into seven key attributes. We

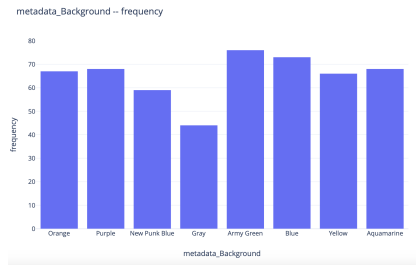


Figure 4: Background

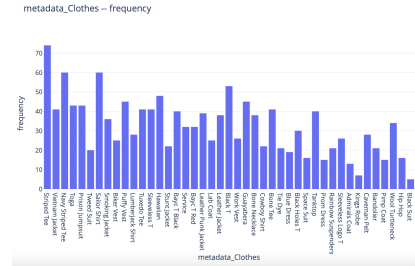


Figure 5: Clothes

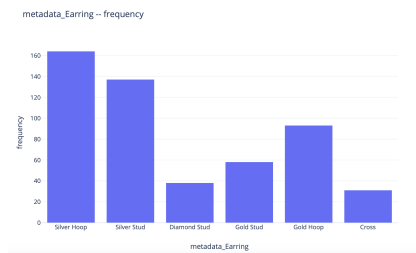


Figure 6: Earring

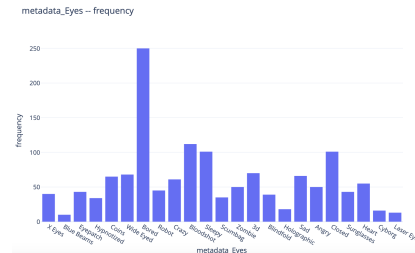


Figure 7: Eyes

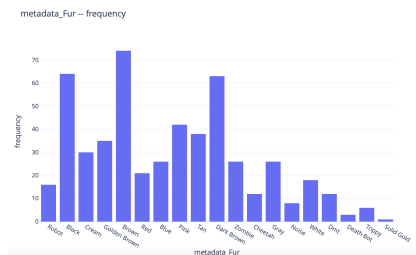


Figure 8: Fur

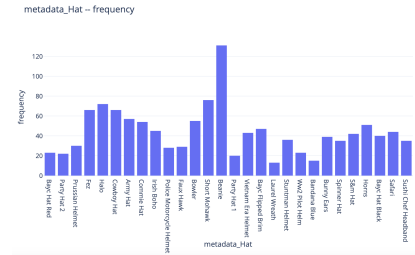


Figure 9: Hat

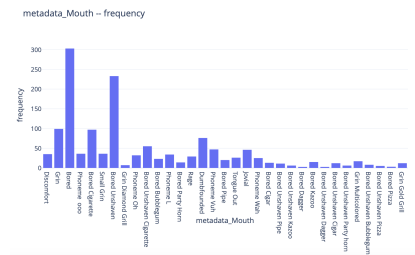


Figure 10: Mouth

encode each attribute for data analysis and use Python to obtain scores as an indicator to evaluate its rarity. Then we calculate each category's average rarity score and average ETH price, and visualize the data by creating three-dimensional plots as follows:

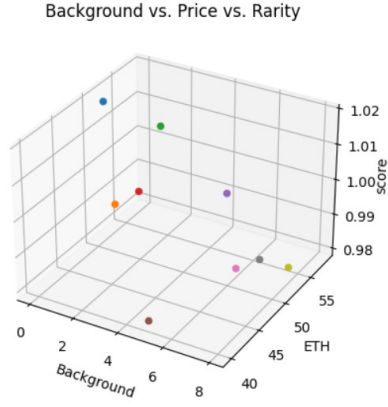


Figure 11: Background vs. Price vs. Rarity

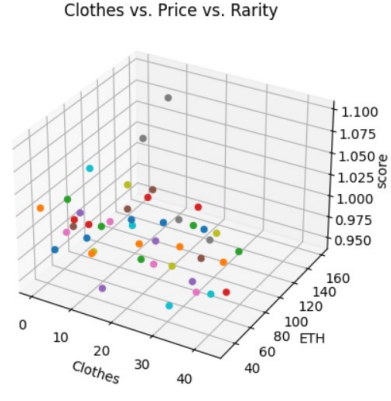


Figure 12: Clothes vs. Price vs. Rarity

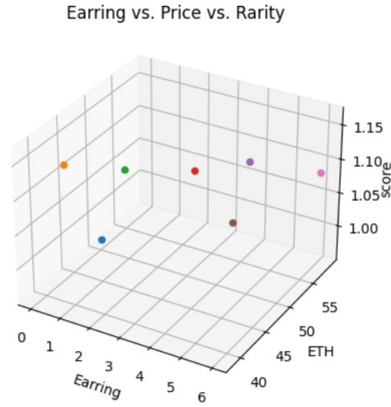


Figure 13: Earring vs. Price vs. Rarity

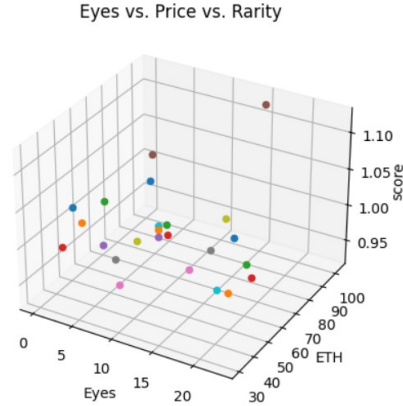


Figure 14: Eyes vs. Price vs. Rarity

- Background: Figure 8 shows the score of background is concentrated around 1.0, while The sales prices are mostly distributed in the range of 50-60 ETH.
- Clothes: Figure 9 shows The six most expensive NFTs in the Clothes category all have a rarity score of above 1.0.
- Earring: Figure 10 shows the score of highest price for Earring group is

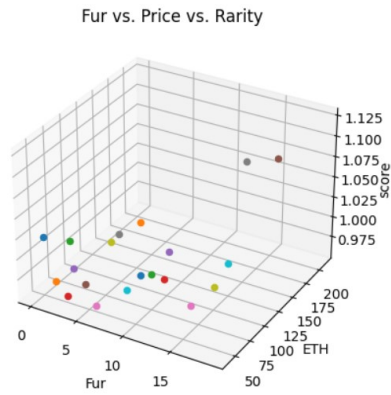


Figure 15: Fur vs. Price vs. Rarity

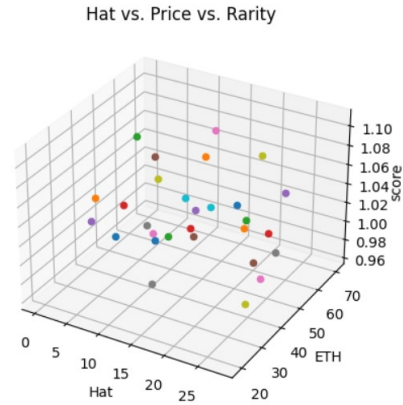


Figure 16: Hat vs. Price vs. Rarity

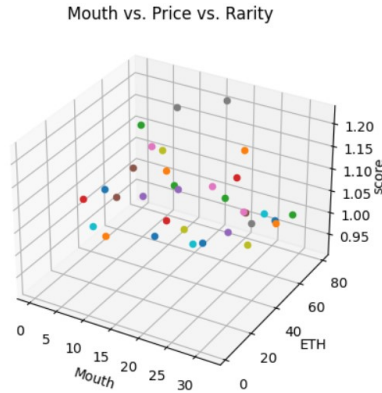


Figure 17: Mouth vs. Price vs. Rarity

concentrated around 1.1.

- Eyes: Figure 11 shows that Laser Eyes has the highest price in the market, and its rarity score is around 1.10.
- Fur: Figure 12 shows that 'Solid Gold' and 'Trippy' have exceptionally high selling prices, with a score of 1.15 and 1.2, respectively.
- Hat: Figure 13 shows that 'Party Hat 2' has the highest prices, with a score of 1.06.
- Mouth: Figure 14 shows that 'Bored Party Horn' and 'Bored Unshaven Pizza' have the highest prices, with a score of 1.15 and 1.2, respectively.

In conclusion, for the most data of collection, the NFTs with a rarity score of around 1.10 – 1.20 has a relatively higher sales price in the market.

4.5 Rarity and Sales Price Trends exploration

To find the trends of the Rarity and Sales Price of BAYC collection, we used 6 columns of metadata: Background, Fur, Clothes, Mouth, Eyes, Hat as features to make regression models. We used two encoding techniques for the categorical variables and three machine learning models for the regression separately. The results are as follows

Encoding	Model	MSE Train	R^2 Train	MSE Test	R^2 Test
Ordinal	XGBoost	0.0118	0.0775	2.3431e-05	0.9983
Ordinal	Random Forest	0.0110	0.1432	0.0016	0.8815
Ordinal	Linear Regression	0.0105	0.1803	0.0115	0.1420
One-Hot	XGBoost	0.0102	0.0496	3.5995e-07	0.99997
One-Hot	Random Forest	0.0095	0.1098	0.0014	0.8982
One-Hot	Linear Regression	0.0090	0.1565	0.0084	0.4028

Table 1: Comparison of encoding techniques and models for BAYC rarity prediction

The highest R-squared values for the rarity prediction models using one-hot encoding and ordinal encoding were 0.1565 and 0.1803, respectively. This suggests that the models can explain a small portion of the variance in rarity. However, the R-squared values indicate that a significant portion of the variance in rarity remains unexplained by the models. True labels and the predicted labels generated by each model are as follows:

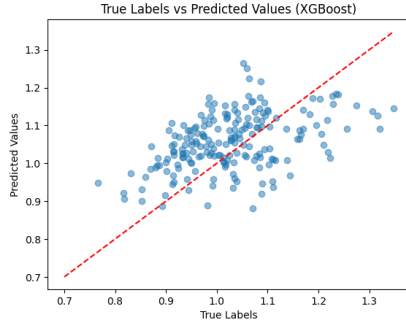


Figure 18: Xgboost onehot

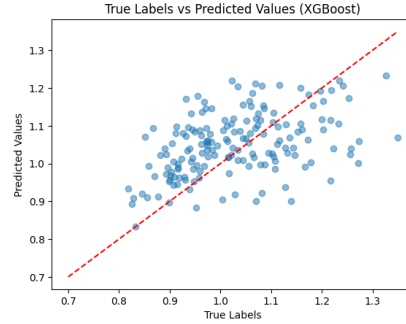


Figure 19: Xgboost ordinal

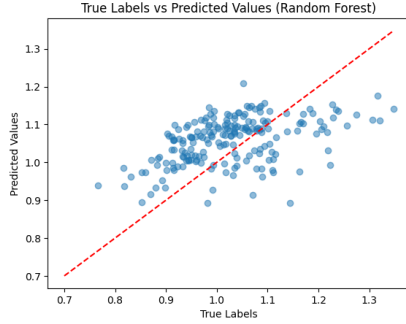


Figure 20: Random Forest onehot

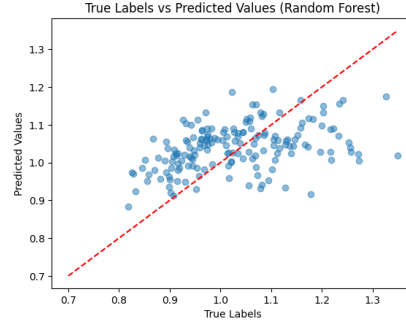


Figure 21: Random Forest ordinal

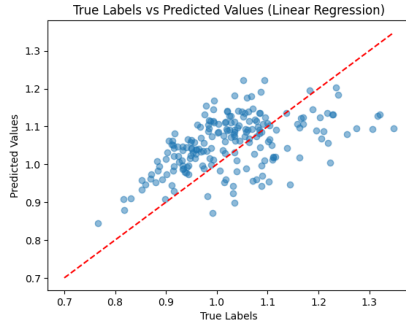


Figure 22: Linear Regression one-hot

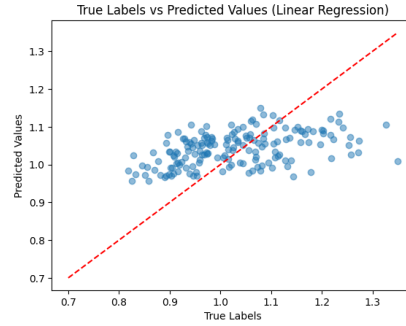


Figure 23: Linear Regression ordinal

The R-squared values for the price prediction models were found to be negative. A negative R-squared value indicates that the models are not able to explain the variance in price, and in fact, perform worse than simply using the mean value of the target variable as the prediction. After second thoughts, we think this maybe caused by the imbalance of the NFTs' prices, big proportion of them are close to 0, so it maybe be difficult to be predicted.

To better build the models in the future, we need to retrieve more data and maybe we can use a time series model to make predictions of the trends of prices and rarity scores of BAYC collection.

5 Conclusion

In this project study, we examined the NFT marketplace, with a particular focus on the Bored Ape Yacht Club collection. By conducting a thorough analysis of the collection and its impact on the market, we provided valuable insights into the trends and challenges facing the NFT marketplace. We believe that

our study will be of interest to collectors, investors, and developers in the NFT community, as well as anyone interested in the future of digital assets.