Luca Ordronneau

NFT Price Analysis

Ocean Data Challenge

Discord :
lucanew#3793

GitHub :
lucaordronneau

Version of
December 3, 2023

Introduction

Non-fungible tokens (NFTs) have experienced a surge in popularity, leading to a rapidly expanding market. Accurate forecasts of their value are crucial for investment decisions and assessing pricing risks associated with NFT financial products. This technical report focuses on developing data analytics reports and a machine learning model to anticipate the floor price of NFTs, specifically targeting the Bored Ape Yacht Club and Azuki collections.

Utilizing a comprehensive dataset from Transpose data, we conduct a series of data analyses to understand the factors influencing NFT prices. Our investigation includes examining daily transaction trends, identifying NFT clusters based on their attributes, and exploring correlations between NFT characteristics and prices. Additionally, we assess the relationship between transaction volume, floor price, and the price of ETH.

Based on these analyses, we develop a machine learning model capable of determining the current floor price of the rarest NFT in a chosen collection. This report presents our findings and insights, detailing our approach to data analysis and the rationale behind our choices for constructing the prediction model. Our work aims to contribute valuable information to the growing field of NFT price analysis and support decision-making in the ever-evolving NFT market.

Data Analysis

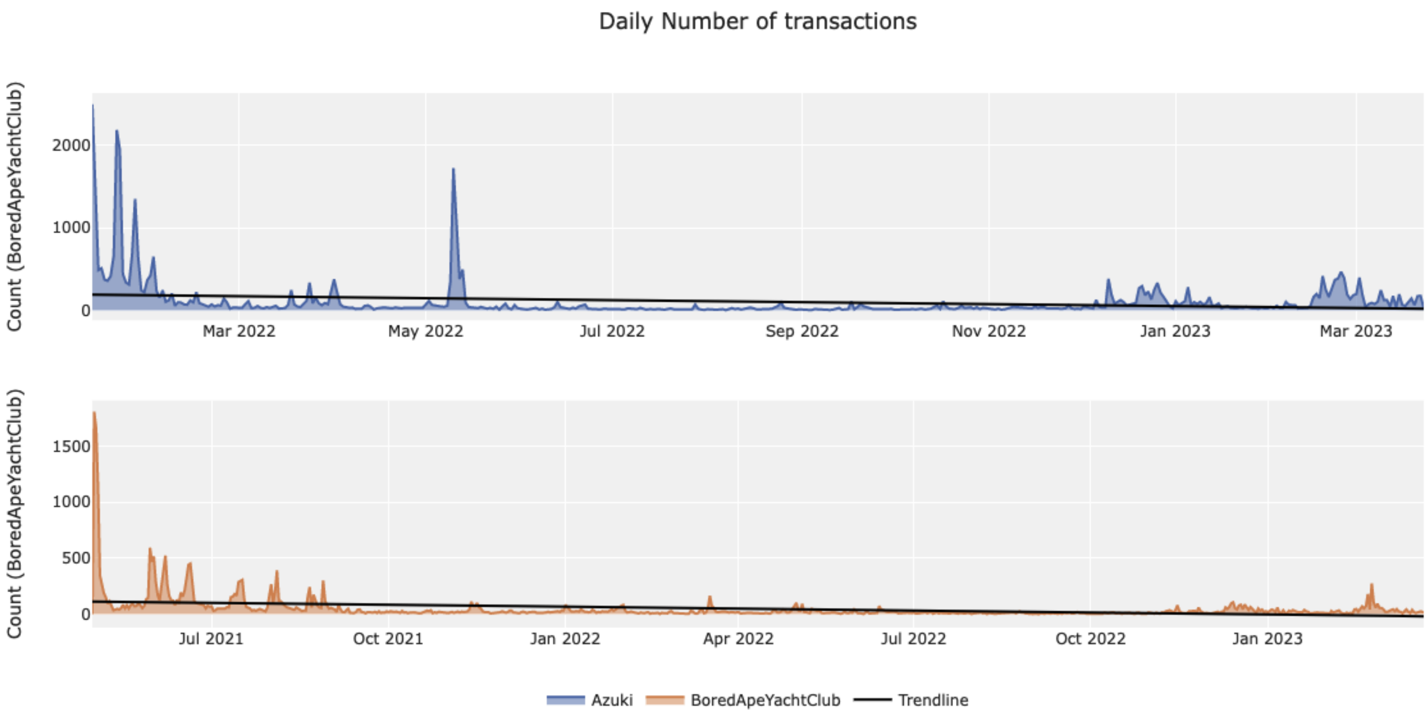
Daily transactions analysis

In this analysis, we investigate how the number of daily transactions for two NFT collections, Bored Ape Yacht Club and Azuki, has changed over time. The provided dataset contains transaction records, which have been aggregated by day and collection. A linear regression trendline has been fitted to each collection's transaction data.

The results show that both collections experienced a large number of transactions at the beginning, likely due to initial hype and low prices. The peak in transactions for both collections reached around 2000. Over time, the number of daily transactions has decreased, as shown by the downward trendlines.

Statistically, the Bored Ape Yacht Club collection has an average of 48 transactions per day, with a median of 19 transactions. The Azuki collection sees a higher transaction volume, with an average of 104 transactions per day and a median of 33 transactions.

Several factors could contribute to the decrease in transactions over time, such as market saturation, high prices, changing market environment, the hype cycle, rarity and uniqueness of NFTs within the collection, and evolving market dynamics.

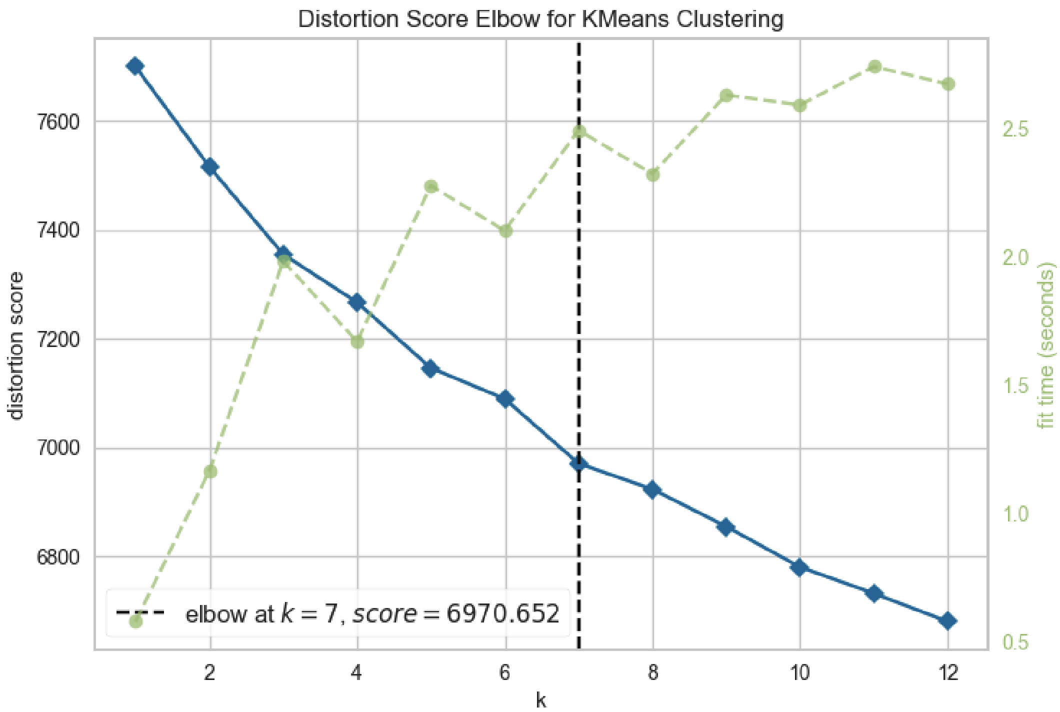


The graphical representation of the daily transaction data illustrates the initial hype in both projects, followed by a gradual decline in the number of daily transactions. This information is valuable for understanding the popularity and trading activity of these NFT collections over time.

Clustering NFTs by Attributes & Characteristics

In this technical report, we aim to identify any clusters or groups of NFTs within the Bored Ape Yacht Club and Azuki collections based on their attributes or characteristics. The approach taken for this analysis involves the use of text tokenization, TF-IDF vectorization, and the KMeans clustering algorithm.

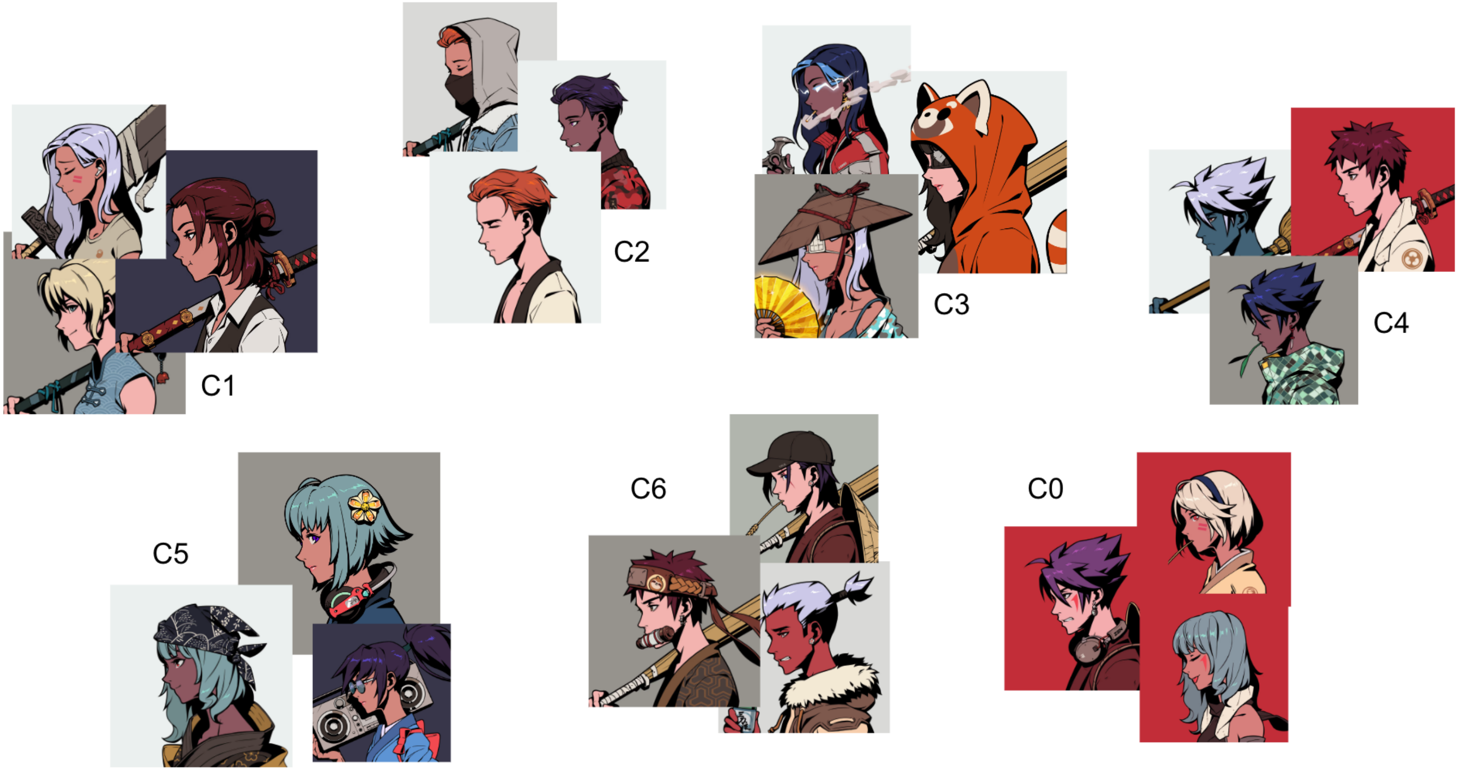
The method begins by selecting the NFTs belonging to the specified collection and extracting the unique flattened properties. These properties are parsed, and the attributes and values are combined to create a single string. The text is tokenized and preprocessed to remove any non-alphabetic tokens and convert them to lowercase. A document-term matrix is created using the TF-IDF vectorizer, which is then fed into the KMeans clustering algorithm. The Elbow Method is used to determine the optimal number of clusters for each collection.

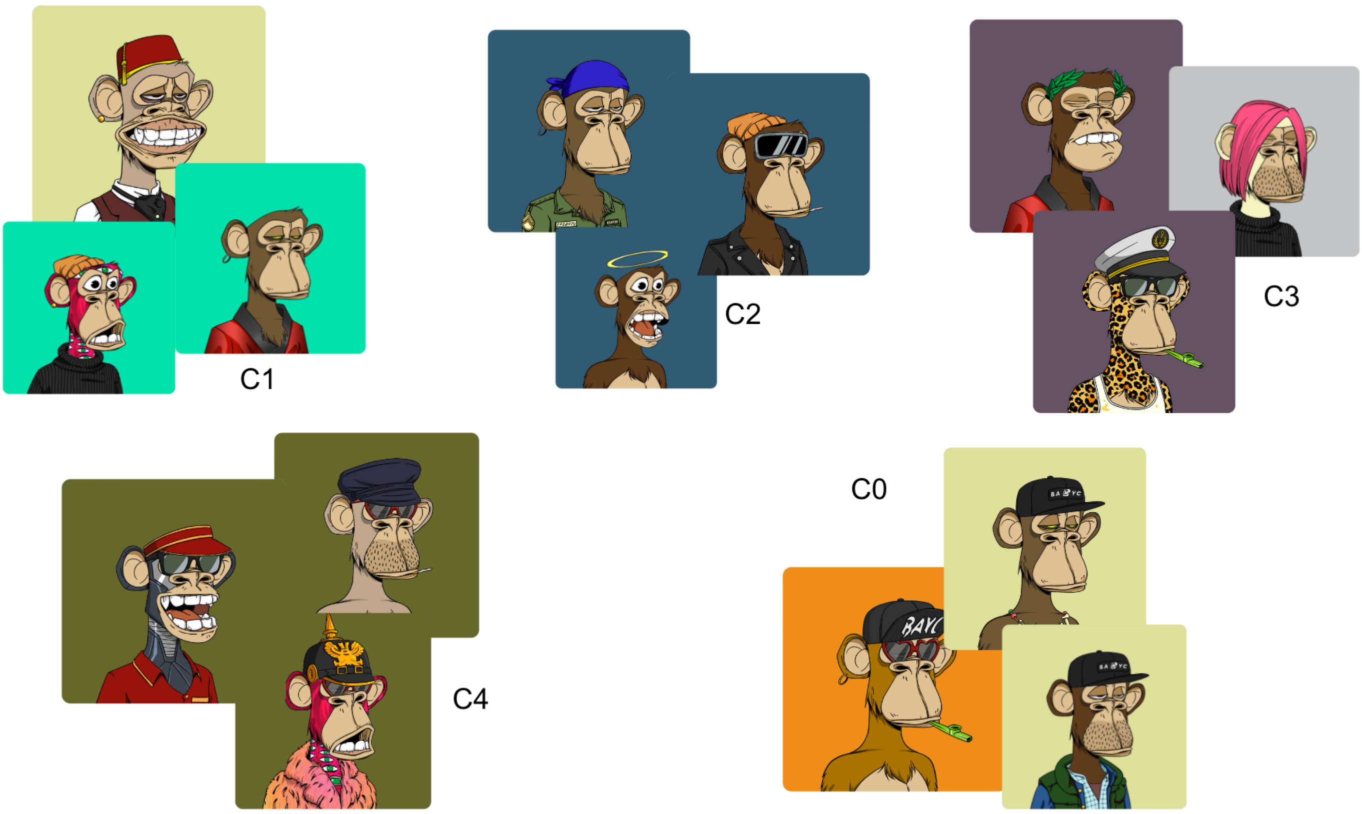


For the Azuki collection, the analysis revealed 7 clusters, each characterized by distinct attributes. For example, Cluster 1's most frequent unique attributes are related to hair, such as Maroon Bun, Maroon Half Bun, and Blue Half Bun. Similarly, the other clusters can be characterized by attributes such as face, eyes, mouth, and clothing.

In the Bored Ape Yacht Club collection, 5 clusters were identified. Cluster 0, for instance, is characterized by attributes like the Bayc Flipped Brim hat, Bayc Hat Black, and Bayc T Black clothes. The other clusters are distinguished by attributes such as earrings, backgrounds, and clothing items.

In summary, this analysis has identified distinct clusters within both the Azuki and Bored Ape Yacht Club NFT collections based on their attributes and characteristics. The key features of these groups include attributes related to hair, face, eyes, mouth, clothing, hats, earrings, and backgrounds.


C1
C2
C3
C4
C5
C6
C0


C1
C2
C3
C4
C0

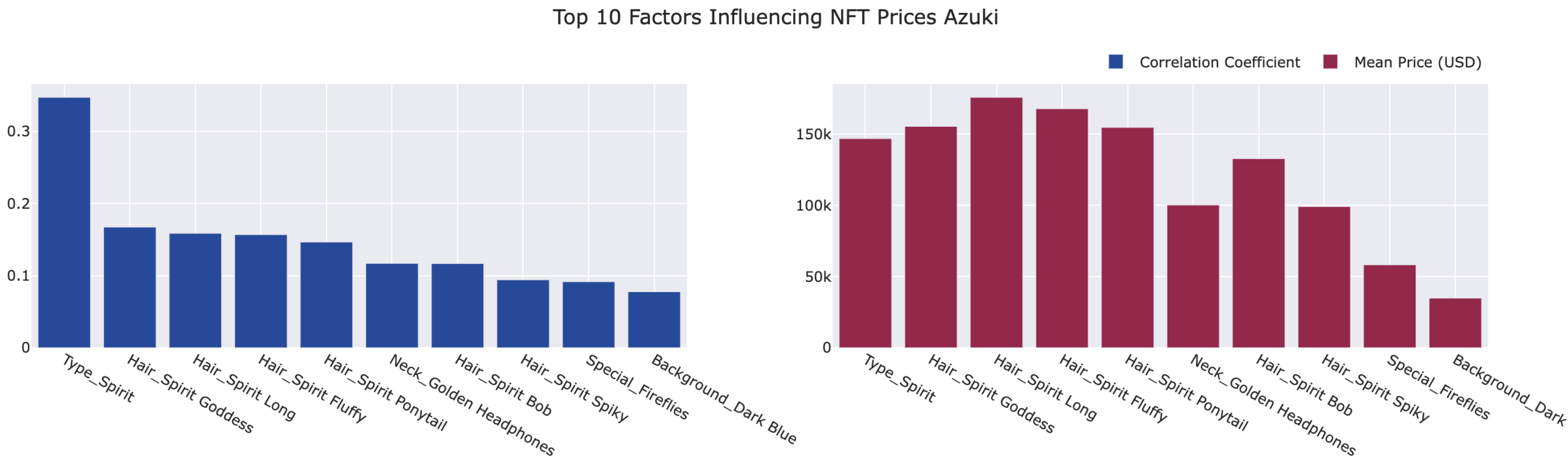
Correlations between NFT Characteristics and Prices

For this question, we aim to identify any correlations between the characteristics of NFTs within the Azuki and Bored Ape Yacht Club collections and their prices. The approach taken for this analysis involves the creation of dummy variables for each attribute-value pair, followed by the computation of correlation coefficients between these dummy variables and the usd_price column. The top 10 most important factors influencing the prices are then determined based on these correlation coefficients.

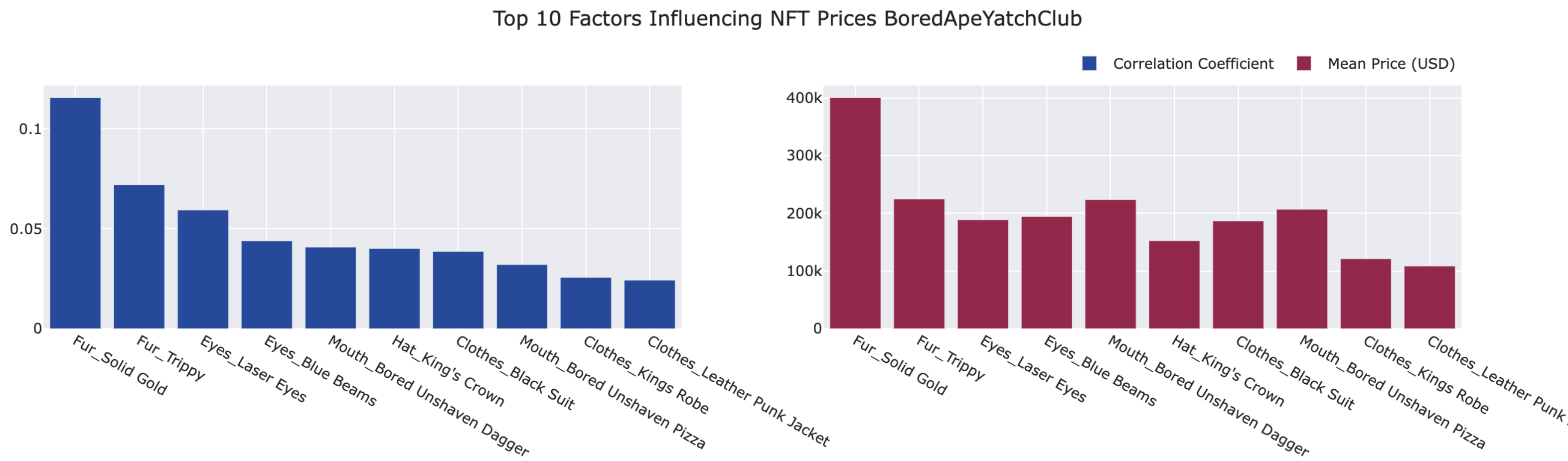
It is essential to understand that the correlation coefficients might not be very strong, as the relationship between NFT attributes and prices can be quite complex and may not be well-captured by linear correlations. A high positive correlation indicates that NFTs with a certain attribute are generally more valuable, while a high negative correlation indicates that NFTs with that attribute are generally less valuable. A correlation close to 0 suggests that the attribute has little or no impact on the price of the NFT.

For the Azuki collection, the analysis reveals that the most influential factors are related to the "Type" and "Hair" attributes, with "Type: Spirit" having the highest positive correlation (0.347) and a mean price of $146,833.24 when this attribute is present. On the other hand, "Type: Human" has a strong negative correlation (-0.131), indicating that NFTs with this attribute have a lower mean price of $23,957.67.

In the Bored Ape Yacht Club collection, the most influential factors are related



to "Fur," "Eyes," and "Mouth" attributes. The "Fur: Solid Gold" attribute has the highest positive correlation (0.116) and a mean price of $400,467.09 when this attribute is present. Other attributes, such as "Eyes: Laser Eyes" and "Mouth: Bored Unshaven Dagger," also exhibit positive correlations and higher mean prices.



Correlation between Transaction Volume and Floor Price

For this question, we aim to determine the correlation between the number of transactions in a collection and its floor price for the Azuki and Bored Ape Yacht Club collections. To achieve this, we first calculate the floor price for each collection by finding the lowest usd_price within each collection for a given period (e.g., weekly). Next, we aggregate the number of transactions by collection and time period using the timestamp column. Finally, we compute the correlation coefficient between the aggregated number of transactions and the floor price.

Our analysis reveals the following results:

Weekly Correlation between the number of transactions and the floor price for Azuki collection: **-0.256**

Weekly Correlation between the number of transactions and the floor price for Bored Ape Yacht Club collection: **-0.281**

These negative correlation coefficients of around -0.25 to -0.28 indicate a weak negative relationship between the number of weekly transactions and the floor price for both collections. In other words, when the number of weekly transactions increases, the floor price tends to decrease slightly, and vice versa.

However, it is important to note that the correlation coefficients suggest a relatively weak relationship between the two variables. Therefore, strong conclusions should not be drawn from this observation alone. Other factors might influence the floor price and the number of weekly transactions, and it is essential to take these into account when analyzing the NFT market.

Correlation Between Collection's Transaction Count and ETH Price

We aim to determine the correlation between the number of transactions in a collection and the price of ETH (price of the market in USD) for the Azuki and Bored Ape Yacht Club collections. To achieve this, we first aggregate the number of transactions by collection and time period (e.g., weekly) using the timestamp column. We then compute the mean usd_price and mean eth_price for each week. Next, we calculate the ETH market price by dividing the mean usd_price by the mean eth_price. Finally, we compute the correlation coefficient between the aggregated number of transactions and the ETH market price.

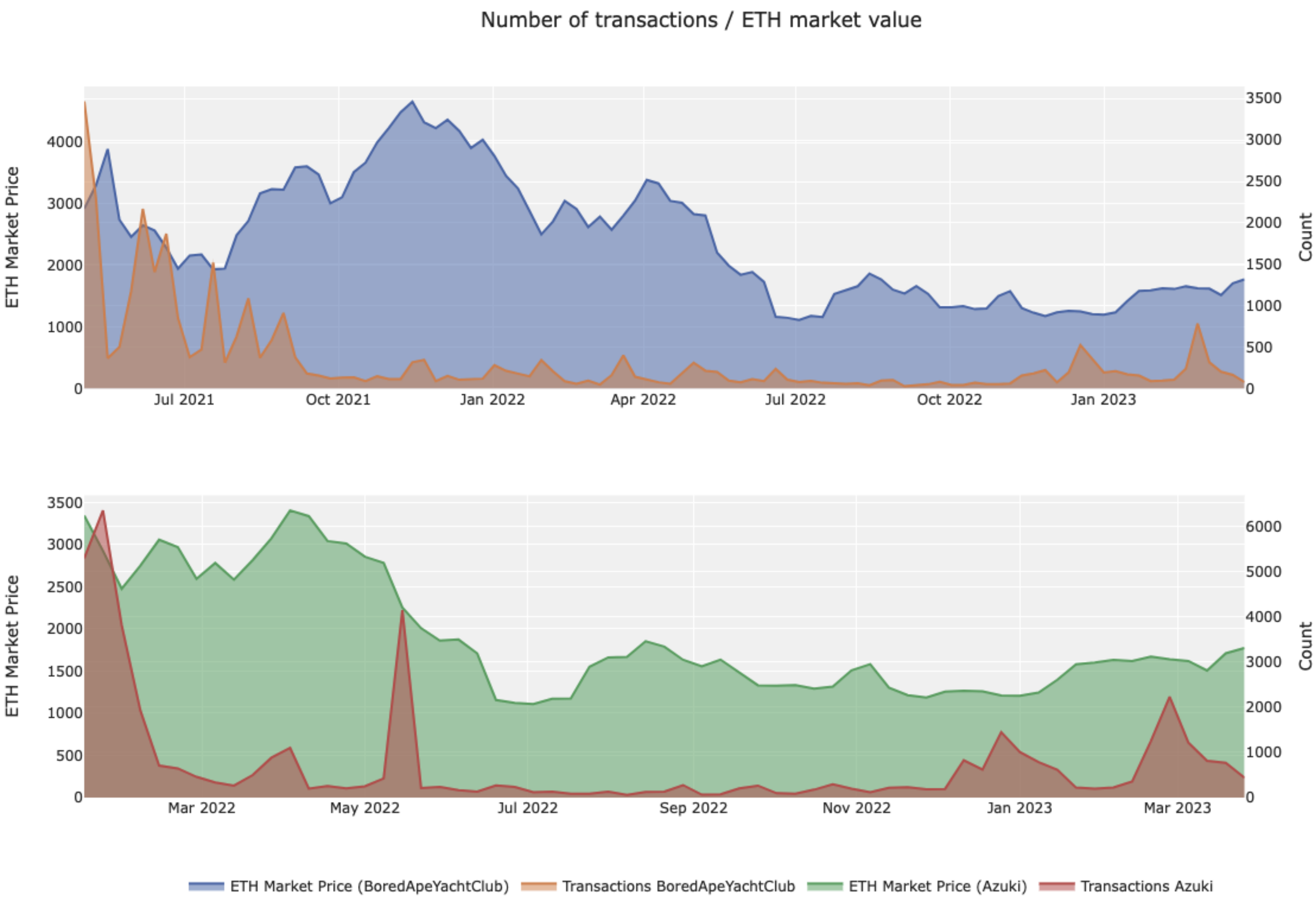
Our analysis reveals the following results:

Weekly Correlation between the number of transactions and ETH price for Azuki collection: **0.375**

Weekly Correlation between the number of transactions and ETH price for Bored Ape Yacht Club collection: **0.113**

These correlation coefficients indicate that there is a weak positive relationship between the number of weekly transactions and the price of ETH for both collections. In other words, when the price of ETH tends to increase slightly the number of weekly transactions increases. However, the relationship is stronger for the Azuki collection than for the Bored Ape Yacht Club collection.

It is essential to note that the correlation coefficients suggest a relatively weak relationship between the two variables. Therefore, strong conclusions should not be drawn from this observation alone. Other factors might influence the number of weekly transactions, and it is important to take these into account when analyzing the NFT market.



Prediction Model

In our analysis, we aimed to develop a machine learning model that can predict the current floor price of the rarest NFT in a given collection. In order to realise this model, we added qome fetures such rarity score, number of transactions, volume in USD, maximum and mean USD sale price, exchange name, and cyclical time encoding.

We began by calculating the floor price for each NFT in the collection by identifying the minimum usd_price for each unique NFT (token_id) on a daily basis.

Next, we computed the rarity score and other important features based on the attributes we identified through our analysis.

The rarity score is calculated by looking at the distribution of attributes across all NFTs in a collection and assigning a score to each attribute based on how common or rare it is. The rarity score for a particular NFT is then calculated as the sum of scores for all of its attributes.

For example, consider a collection of NFTs where each NFT has two attributes: color and shape. Suppose that there are 10 NFTs in the collection, and the distribution of colors and shapes is as follows:

- 3 NFTs are blue circles
- 2 NFTs are green circles
- 1 NFT is a red circle
- 2 NFTs are blue squares
- 1 NFT is a green square
- 1 NFT is a red square

To calculate the rarity score for a blue circle NFT in this collection, we would assign a score of 0.3 to the color blue (since it appears in 3 out of 10 NFTs) and a score of 0.5 to the shape circle (since it appears in 5 out of 10 NFTs). The rarity score for this NFT would be $0.3 + 0.5 = 0.8$.

By comparing the rarity scores of different NFTs in a collection, we can get a sense of which NFTs are the most unique or rare.

We then created a new DataFrame containing the token_id, rarity score, number of transactions, volume in USD, maximum usd sale, mean USD sale, correlation scores for exchange name, and cyclical time encoding (sinus and cosinus) for month, week, day of the week, and hour. In the development of our machine learning model, we took care to ensure that the selected features did not leak any information about the floor price, thus preventing the model from exploiting unintended information and enabling it to generate accurate and reliable predictions based solely on relevant and appropriate factors.

For our machine learning model, we chose to use the XGBoost Regressor, a powerful and widely-used algorithm for regression tasks. We used cross-validation to tune the model's parameters and ensure its robustness, utilizing the findings from our cluster analysis. To further refine our model, we employed Weight and Biases for hyperparameter tuning.

The rarest NFT for the Azuki collection is identified as #9605.

Finally, after training our model on the entire dataset with the tuned parameters, we used it to predict the floor price of the rarest NFTs in the Azuki collection. By comparing the predicted floor price with the actual market price, we can determine that the NFT is overpriced.

Conclusion

In conclusion, our technical report demonstrates the potential of data analytics and machine learning models in predicting the floor price of rare NFTs within a specific collection. By examining various factors, such as rarity score, number of transactions, volume in USD, maximum and mean USD sale price, correlations of attributes, and cyclical time encoding, we have developed a comprehensive understanding of the Bored Ape Yacht Club and Azuki collections.

Our machine learning model, utilizing the XGBoost Regressor, was able to predict the floor price of the rarest NFT in each collection. The rarest NFT for the Azuki collection is identified as #9605. Our model predicted a floor price of $250,780 for this NFT, while its actual market price is $1,424,801. Given this significant difference, the NFT appears to be overpriced.

We predicted the price of the NFT as we were 2022-03-30.

It is crucial to note that while our model offers valuable insights, it may not capture the complex dynamics and various factors affecting NFT prices. Therefore, caution should be exercised when drawing conclusions or making investment decisions based solely on these predictions. Ultimately, our work contributes to the growing field of NFT price analysis and supports decision-making in the ever-evolving NFT market.