Ranking Features

Ranking 1: Total Earnings - This feature is based on the total amount earned by a wallet holder on purchasing NFTs. The more positive the earnings for that wallet holder , the more valuable their membership should be, as they are more invested in the ecosystem.

Ranking 2: Average amount spent - This feature is based on the average amount spent by a wallet holder on purchasing NFTs. The more money a wallet holder has spent on NFTs, the more valuable their membership should be, as they are more invested in the ecosystem.

Ranking 3: Number of transactions - This feature is based on the total number of transactions made by a wallet holder. It assumes that a wallet holder who has made a larger number of transactions is more actively engaged in the ecosystem and therefore more valuable.

Ranking 4: Number of unique currencies - This feature is based on the number of unique currencies a wallet holder has used to purchase NFTs. A wallet holder who has used a larger number of currencies is more globally connected and has a more diverse portfolio.

Ranking 5: Average Duration - This feature is based on the average duration of time that a wallet holder has held onto their NFTs. A wallet holder who has held onto their NFTs for a longer period of time is more committed to the ecosystem and is therefore more valuable.

Ranking 6: Number of NFT contract addresses - This feature is based on the number of unique NFT contract addresses a wallet holder has interacted with. A wallet holder who has interacted with a larger number of NFT contract addresses has a more diverse NFT portfolio and is therefore more valuable.

Bucketing rules

The scoring system is designed to rate wallet holders based on several key features, including their average amount spent, profit and loss, number of transactions, and more. We are using binning to assign membership structures to wallet holders based on their ranking in each of the above mentioned seven features.

The goal of binning is to transform continuous numeric features into categorical features, which can be more easily interpreted and analyzed. This process involves grouping values into a set number of bins based on their range. For each feature, we first calculate the range of values that the column contains. Based on the range, we create bins with custom labels using the pd.cut() function in Pandas.

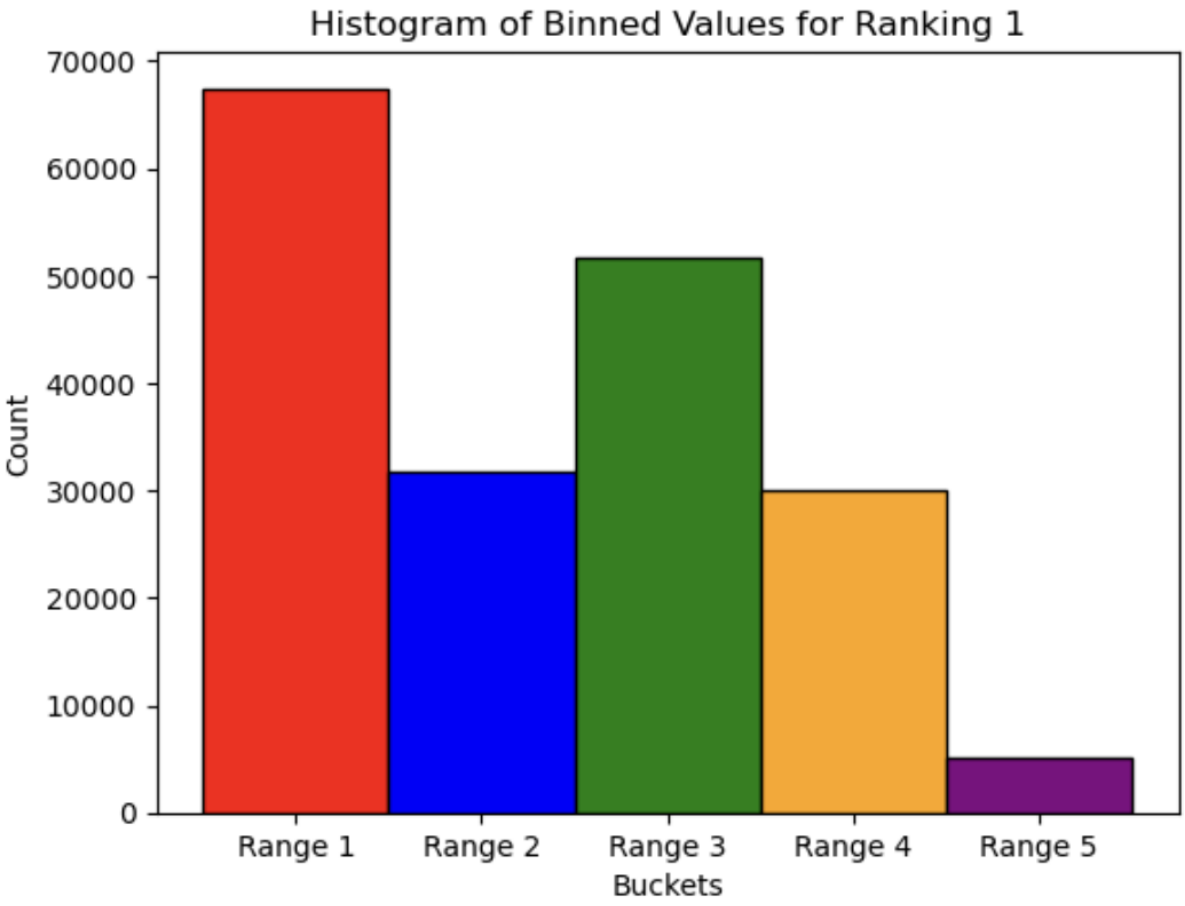
To apply binning to the ranking features, we use the Bucketizer function, which takes a column of continuous numeric values and outputs a column of bucket indices. Bucketizer takes two inputs: the input column to be binned and an array of bin boundaries. The bin boundaries are a list of values that divide the range of the input column into a set of intervals or buckets.

For each ranking feature, we first calculate the range of values for that feature and then divide the range into a set number of bins. The bin boundaries are then used as input to the Bucketizer function, which assigns each value in the input column to a bin based on its corresponding bin boundary. The resulting output is a new column of categorical values, each representing the bin in which the corresponding numeric value falls.

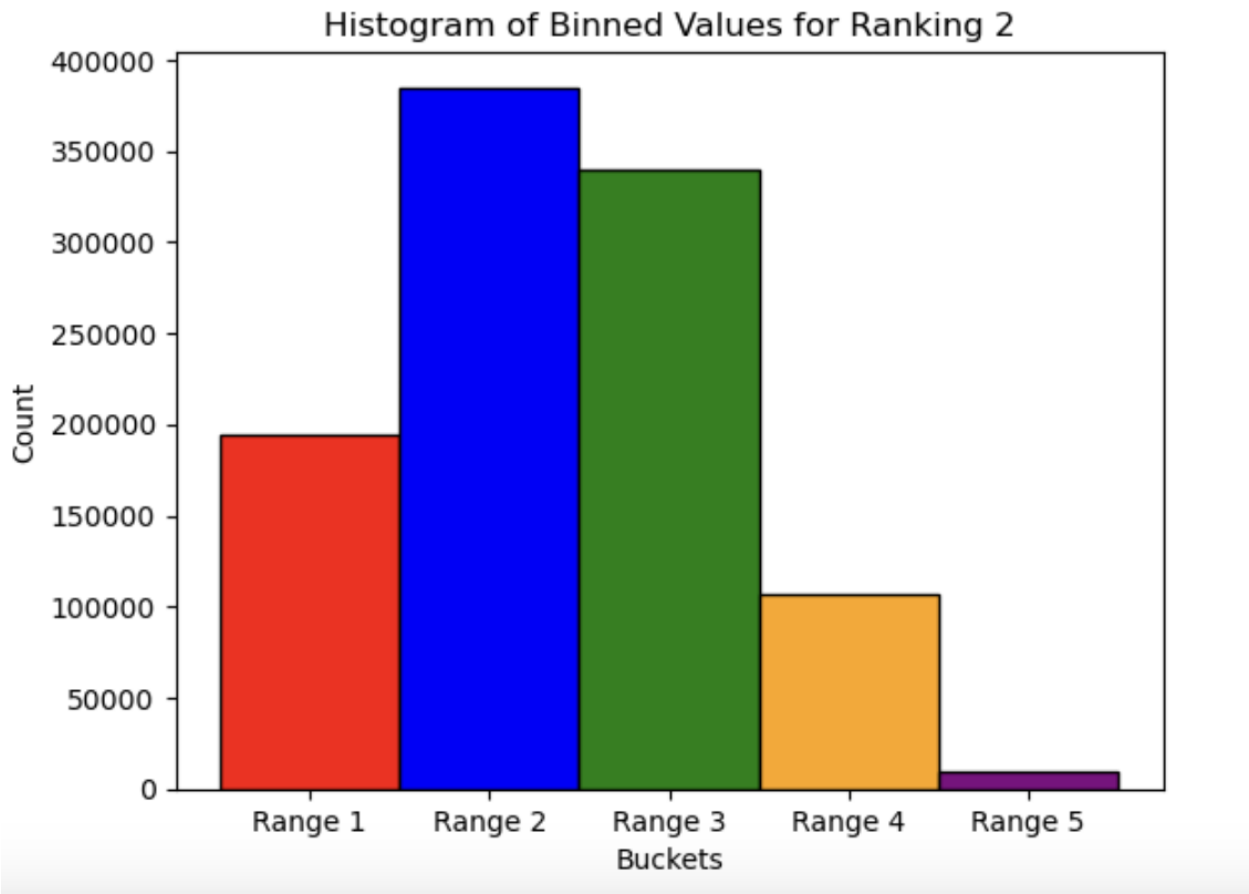
Once binning is applied to all the ranking features, the resulting columns are joined to create a final membership structure for each wallet holder. This structure assigns a membership level to each wallet holder based on their bin in each ranking feature. The membership level can be used to define rewards, benefits, or other incentives for wallet holders who have demonstrated more engagement, investment, and commitment to the NFT ecosystem.

After creating bins for each feature, we use the Bucketizer function in PySpark to assign bin numbers to each feature. Finally, we join the binned features to create the membership structure for each wallet holder based on their ranking in each feature.

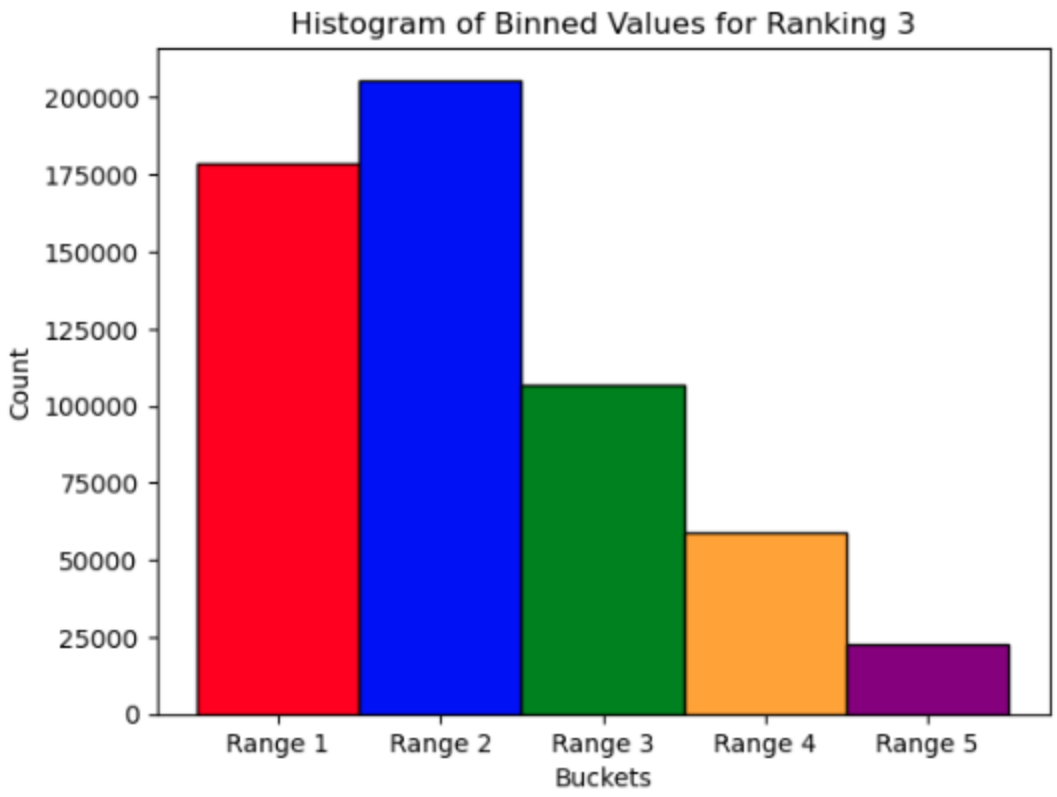
The buckets have been defined for each ranking as follows:
Ranking 1 - Total Earnings
[-16018783.0, -50, 50, 1000, 10000, 60688892.0]



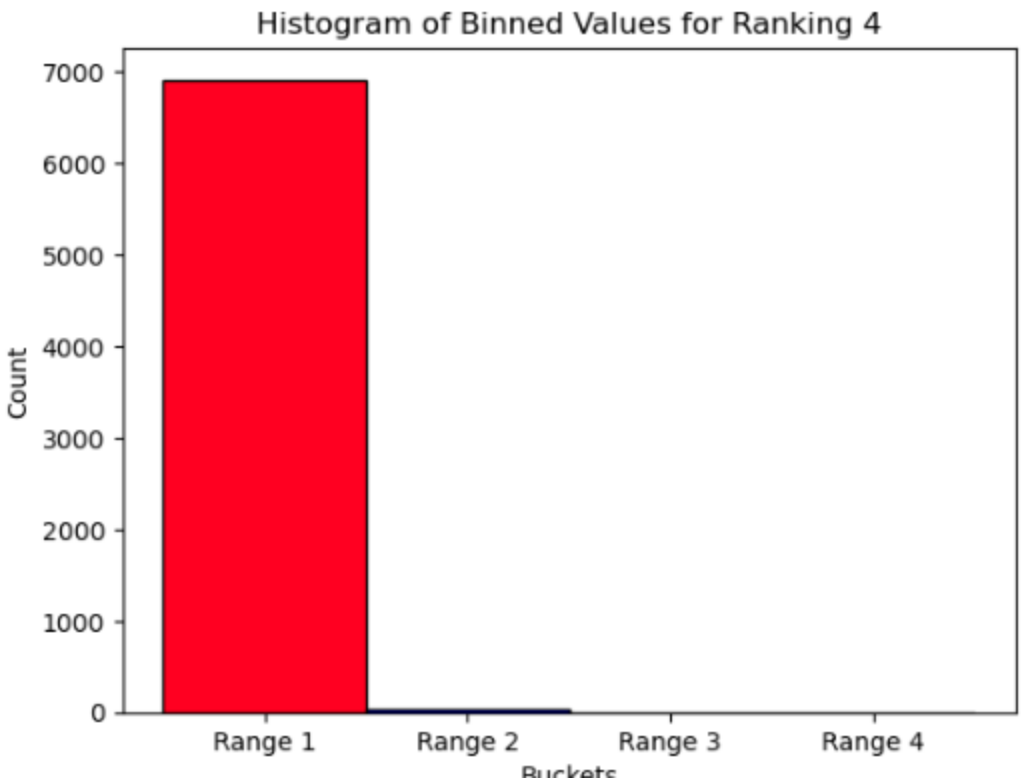
Ranking 2 - Amount Spent
[0.0, 100, 1000, 10000, 100000,100139800]



Ranking 3 - number of transactions
[1, 2, 5, 10, 20,116]



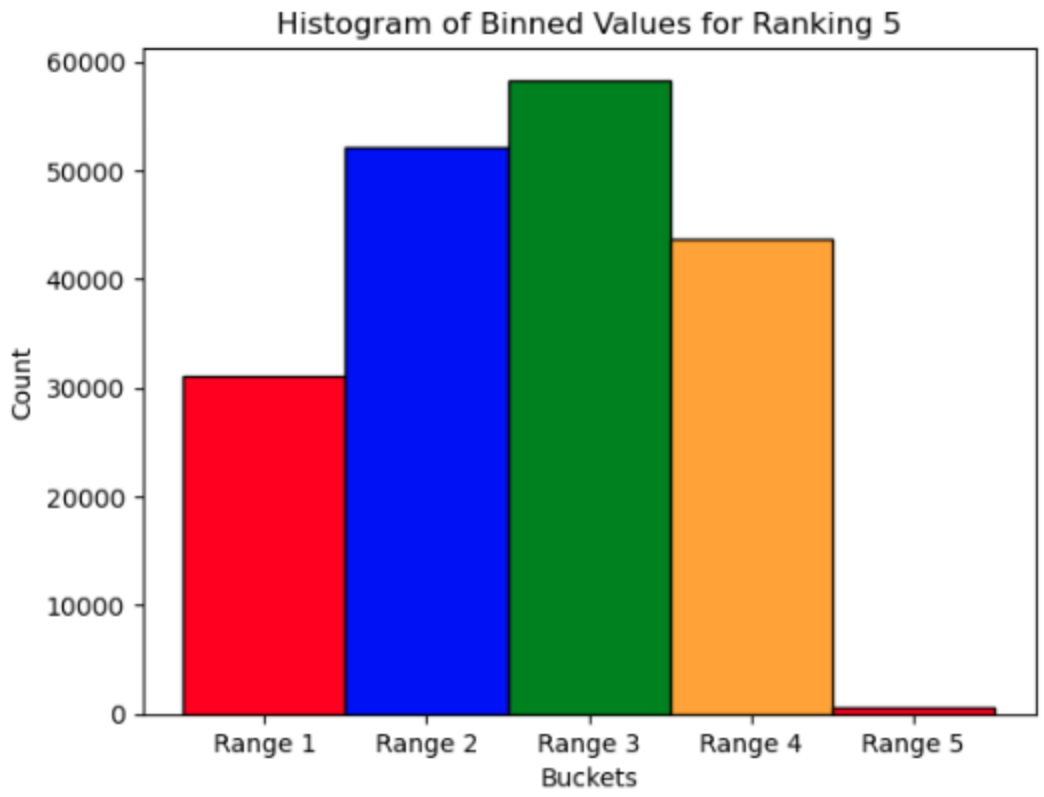
Diversity Ranking 4 - Number of unique currencies done in transactions
[1, 2, 3, 4, 6]
Min : 1
Max : 4, hence 4 ranges only



Ranking 5 - Average Holding Duration (in seconds)
Max value : 9.1290008E7 seconds

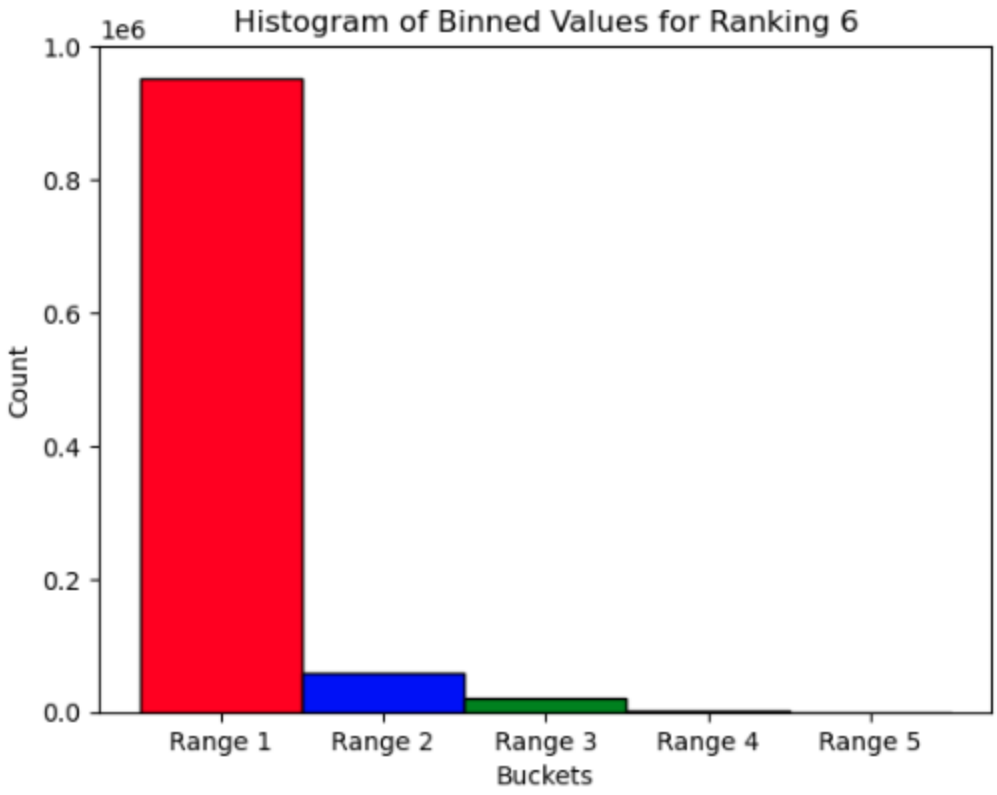
Buckets divided in a day, in a week, in a month, in a year and more than 1 year(2.89 years based on max value)

[0, 86400,604800,2628000, 31540000, 91290008.0]



Ranking 6 - Number of NFT Contract Addresses

[0, 10, 20, 50, 100, 116.0]


Histogram of Binned Values for Ranking 6
1e6
1.0
0.8
0.6
0.4
0.2
0.0
Count
Range 1
Range 2
Range 3
Range 4
Range 5
Buckets

Tiering Criteria

With the given weights:

weight_1 = 10
weight_2 = 10
weight_3 = 5
weight_4 = 10
weight_5 = 10
weight_6 = 10

Each weight number corresponds to the weight of the respective ranking rule.

To determine the values in the "tier" column, the approximate 20th and 80th percentiles of the "Score" column are first computed using the approxQuantile method. These percentiles divide the distribution of scores into three tiers:

- Tier 1: Scores below or equal to the 20th percentile.
- Tier 2: Scores above the 20th percentile and below or equal to the 80th percentile.
- Tier 3: Scores above the 80th percentile.

If a customer's score is below or equal to the 20th percentile, the "tier" column is set to "tier1". If the score is above the 20th percentile but below or equal to the 80th percentile, the "tier" column is set to "tier2". Otherwise, if the score is above the 80th percentile, the "tier" column is set to "tier3".