Binanace Accounts Analysis:

Project Link:

Github: [abhashgoyal/Binance-Account-Ranking-Algorithm]

Linkedin: [Abhash Goyal | LinkedIn]

Aim :

Dataset Information: Historical trade data from various Binance accounts over **90 days,** containing:

Port_ID : Unique identifiers for accounts.
Trade_History: Historical trades with details like timestamp, asset, side (BUY/SELL), price, and more.

Objective: Analyze the dataset to calculate financial metrics for each account, rank them, and provide a top 20 list.

Metrics to Calculate:
ROI (Return on Investment)
PnL (Profit and Loss)
Sharpe Ratio
MDD (Maximum Drawdown)
Win Rate
Win Positions
Total Positions

Findings: Data Exploration

Total Columns: 15:
{Port_ID, time, symbol, side, price, fee, feeAsset, quantity, quantityAsset, realizedProfit, realizedProfitAsset, baseAsset, qty, positionSide, activeBuy}
Total Unique Users - 149
Total Transactions made by users : 211277
Time - **Data of 60 days were given (UNIX TIME)**
Total Stocks Traded : 153

Missing Transactions: ONE Trade History was empty:

The dataset initially consisted of 150 users, with two columns: **Port_id** and **Trade History**. Upon analysis, it was discovered that one record in the "Trade History" column contained a missing value. This record was removed to ensure data integrity. After cleaning, the final dataset includes 149 complete records for both columns, resulting in a total of 149 valid entries.

The Trade History column contains detailed information about user trades in a JSON-like structured format. Each record documents multiple trades, providing a comprehensive view of trading activities. Below are the key aspects observed in the data:
Key Metrics Captured:

- **Time**: Trade execution timestamps.
- **Symbol**: Assets traded (e.g., SOLUSDT).
- **Trade Direction**: All trades are marked as BUY.
- **Position Side**: All trades are LONG.
- **Realized Profit**: No profit realized in any of the trades (0.0 USDT).
- **Fees**: Fees incurred are deducted in USDT.
- **Active Buy**: Some trades were marked as activeBuy: True, indicating active orders.
- **Profitability**: No realized profits recorded, indicating possible unrealized outcomes.
- **Trade Volume**: Significant trading activity with high transaction volumes.

Assumptions:

1. **Realized profit** represents the total profit/loss for each trade - It includes all trading fees and commissions - Negative values indicate losses after fees - Zero values may indicate unrealized positions or break-even trades

Methodology

1. Data Analysis:
   Explored the data and mentioned the findings.
2. EDA- Exploratory Data Analysis
   - Converted the raw data into Data Frames, and a more readable format.
   - UNIX Time to Date Time Format.
3. Feature Engineering:

- Calculated each metrics from the listed metrics.

Algorithm Used:

1. Z-Score Test:
   The Z-Score Test is a statistical method used to measure how many standard deviations a data point is from the mean of a dataset. It helps identify whether a value is significantly different from the expected norm.
   $$Z = \frac{X - \mu}{\sigma}$$
   where,
   X: The raw score or data point whose Z-score is being calculated.
   μ: The mean (average) of the population or dataset.
   σ: The standard deviation of the population or dataset.

2. MCDA(Topsis Score)
   TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) is a multi-criteria decision analysis (MCDA) method used to rank and evaluate alternatives by comparing their proximity to an ideal solution.

3. Machine Learning Approach (Ensemble Learning)
   Ensemble learning is a machine learning approach that combines the predictions of multiple models (base learners) to produce a more accurate and robust result than any individual model.

   a. Base Model LightBG
      LightGBM is a highly efficient and scalable framework for gradient boosting, designed to optimize performance on large datasets and high-dimensional data.

   b. XGBOOST
      XGBoost is an open-source, scalable, and efficient implementation of gradient boosting, optimized for speed and performance. It is widely regarded as one of the most powerful machine learning algorithms due to its high predictive accuracy and ability to handle various data types and tasks.

   c. Random Forest

Random Forest is an ensemble learning method primarily used for classification and regression tasks. It builds multiple decision trees during training and merges their predictions to improve accuracy and control overfitting. Random Forest is one of the most popular and robust machine learning algorithms due to its ability to handle large datasets, its versatility, and its relatively simple implementation.

After Implementation of the algorithms I have the following findings:

	Port_ID	z_score	topsis_score	z_score_rank	topsis_rank
8	3826087012661391104	2.540907	0.520940	1.0	2.0
2	3768170840939476993	1.535170	0.449858	2.0	4.0
75	3999240873283311617	1.326136	0.491707	3.0	3.0
96	4020204877254599680	1.317961	0.572640	4.0	1.0
62	3986814617275053313	1.207571	0.381569	5.0	6.0
47	3956048468100538880	1.060482	0.338650	6.0	17.0
143	4039129759104249600	0.963624	0.356644	7.0	9.0
16	3891020560590657281	0.951884	0.339048	8.0	16.0
133	4035430878731345664	0.802350	0.338201	9.0	18.0
14	3886752488982104320	0.756363	0.329665	10.0	24.0
112	4029749871687083265	0.752262	0.348420	11.0	14.0
17	3907081197088384000	0.745922	0.361155	12.0	8.0
58	3977234346014419201	0.735990	0.355183	13.0	10.0
144	4039279455324236544	0.733673	0.324058	14.0	31.0
107	4028701921959171840	0.718550	0.362562	15.0	7.0
100	4022641794255717633	0.672999	0.326564	16.0	26.0
39	3944658614777849089	0.670193	0.331279	17.0	22.0
36	3943533600390906881	0.663843	0.320108	18.0	38.0
109	4029422834086627072	0.640813	0.332824	19.0	21.0
99	4022565861939831809	0.639179	0.347417	20.0	15.0

Z score and MCDA:

Ranking Comparison Metrics:
Average Rank Difference: 12.79
Rank Correlation: 0.924

Matching Portfolios in Top 20:

Rank	Percentage (%)
Top 5	80.0%
Top 10	60.0%
Top 20	70.0%

Conclusion: Z score and Topsis Score give a good accuracy of 70% in portfolio matching

	Port_ID	Ensemble_Score	Rank
5	3768170840939476993	0.891498	1.0
11	3826087012661391104	0.810318	19.0
19	3891020560590657281	0.872043	6.0
33	3936410995029308417	0.847611	15.0
42	3944658614777849089	0.863353	11.0
50	3956048468100538880	0.874198	3.0
61	3977234346014419201	0.868943	7.0
64	3983074113875692800	0.861356	12.0
84	4004713168329653760	0.848869	14.0
98	4019895412775450368	0.823466	18.0
105	4023697433751327232	0.863722	9.0
106	4023697881318718465	0.847050	16.0
112	4029422834086627072	0.872193	5.0
115	4029749871687083265	0.863532	10.0
118	4030555430101054209	0.810164	20.0
136	4035430878731345664	0.872965	4.0
146	4039129759104249600	0.891344	2.0
147	4039279455324236544	0.859715	13.0
149	4040843843196854529	0.833715	17.0
151	4041860229502600193	0.866647	8.0

ENSEMBLE VS Z SCORE VS MCDA(TOPSISI SCORE):

Overlap Analysis:
Portfolios common between Z-Score and TOPSIS: 14
Portfolios common between Z-Score and Ensemble: 11
Portfolios common between TOPSIS and Ensemble: 11
Portfolios common across all three methods: 8

Unique Portfolios Analysis:
Portfolios unique to Z-Score: 3
Portfolios unique to TOPSIS: 3
Portfolios unique to Ensemble: 6

Major Outcome

1. Top Performers Across Methods

Upon reviewing the top performers across all three methods, the following observations were made:

- **Portfolio ID 3768170840939476993**:
  - **Ranked 2nd in Z-Score**: This portfolio shows a strong performance in terms of standardized metrics.
  - **Ranked 4th in TOPSIS**: The portfolio performs well when considering multiple criteria simultaneously.
  - **Ranked 2nd in Ensemble Score**: Combining both Z-Score and TOPSIS ranking methods, this portfolio retains a top position.
- **Portfolio ID 3826087012661391104**:
  - **Ranked 1st in Z-Score**: This portfolio leads in terms of standardized performance metrics.
  - **Ranked 2nd in TOPSIS**: It performs consistently well when evaluating multiple factors.
  - **Ranked 20th in Ensemble Score**: This portfolio's position changes considerably when combined with the other two methods, indicating the influence of different ranking criteria.

2. Ranking Variations

- **Significant Rank Differences**: A few portfolios demonstrate substantial differences in their rankings across the three methods. This highlights that each method might prioritize different performance factors, leading to variability in rankings.
- **Ensemble Methods Role**: The Ensemble method appears to balance out extreme rankings from individual methods. While some portfolios may perform extremely well in one method and poorly in another, Ensemble provides a more balanced and stable ranking.
- **Agreement Between Z-Score and TOPSIS**: Both Z-Score and TOPSIS rankings exhibit some level of agreement, particularly in the higher-performing portfolios. However, there are also significant differences in ranking orders, suggesting that these methods prioritize different aspects of portfolio performance.

3. Method Characteristics

Each of the ranking methods has distinct characteristics that affect their ranking outcomes:

- **Z-Score**: The Z-Score method focuses on standardized performance metrics, ensuring that portfolios are compared based on their relative performance. It highlights portfolios that stand out in terms of overall performance when compared to the mean.
- **TOPSIS**: The TOPSIS method takes into account multiple criteria simultaneously, considering both the best and worst possible performance. This multi-criteria approach provides a more comprehensive ranking but may lead to different outcomes than methods that focus on a single performance measure.
- **Ensemble Method**: The Ensemble method combines both the Z-Score and TOPSIS rankings, leveraging the strengths of both methods. This provides a more balanced view, mitigating the extremes of individual methods and offering a holistic assessment of portfolio performance.

4. Portfolio Distribution

- **Overlapping Top Portfolios**: There is a notable overlap in the top 20 portfolios across the methods. Several portfolios consistently rank high across all three methods, suggesting their strong performance in various areas.
- **Variable Rankings for Some Portfolios**: Some portfolios show considerable fluctuation in their rankings depending on the method used. This variability

highlights the impact of different ranking criteria and the need for a comprehensive approach to performance evaluation.

Future Scope:

1. Developing an automated robust model for finding the top performers.
2. Developing a Front-End for showing the user their graphs.
3. Using more advanced models for testing and experimenting.

Conclusion:

This comparative analysis of Z-Score, TOPSIS, and Ensemble rankings demonstrates the value of employing multiple ranking techniques to assess portfolio performance. While each method provides useful insights, using them together in an Ensemble approach yields a more balanced and reliable assessment. The variations observed across individual methods underline the importance of considering multiple perspectives when evaluating portfolio performance.