Report for Analysis of Historical Trade Data

1. Introduction:

This report addresses the study of historical transaction data from many Binance accounts over a 90-day period. The primary goal was to calculate critical financial data for each account, rate them, and identify the top 20 accounts based on performance.

2. Data Exploration and Cleaning

- **Data Loading**: The dataset was loaded from the CSV file TRADES_CopyTr_90D_ROI.csv.
- **Initial Inspection:** We examined the data's structure and content, identifying columns like Port_IDs and Trade_History.
- **Handling Missing data:** To preserve the analysis's integrity, rows with missing data were eliminated.
- **Trade History Transformation:** The Trade_History column, which was originally represented as json, was turned into lists to facilitate manipulation.

3. Feature Engineering:

- **Trade Data Transformation:** Each trade record was retrieved and connected with the appropriate Port_ID. This produced a new DataFrame (df_trades) with all trade information.
- **Position Type Creation:** A new column, Position_Type, was added to categorize trades according to their side and position type.
- **Trade_value:** It was calculated as the product of quantity and price.

4. Metrics Calculation:

The key performance measures were calculated as follows:

- **PnL (Profit and Loss):** The overall profit/loss for each account was determined by adding the realizedProfit.
- **Win Positions:** The number of profitable transactions was calculated.
- **Total Positions:** The total number of deals was tallied.
- **ROI (Return on Investment):** It is calculated as the total PnL divided by the absolute sum of PnL across all accounts and expressed as a percentage.
- **Win Rate:** The win rate was calculated by dividing the number of winning positions by the total number of positions, and it is expressed as a percentage.
- **Return Calculation:** Each trade's return was computed by dividing the realized profit by quantity.

5. Sharpe Ratio and Maximum Drawdown:

- **Sharpe Ratio:** The Sharpe Ratio was calculated to determine each account's risk-adjusted return, using a risk-free rate of 0.01.

$$\text{Sharpe Ratio} = \frac{\text{mean return} - \text{risk-free rate}}{\text{standard deviation of returns}}$$

- **Maximum Drawdown (MDD):** This metric was developed to calculate the maximum observable loss from a peak to a trough in an account's returns.

6. Ranking Algorithm:

- A scoring system was created to rank accounts using a weighted total of metrics:

$$\text{Score} = (0.5 \times \text{PnL}) + (0.3 \times \text{Win Positions}) + (0.2 \times \text{ROI})$$

- Accounts were ranked based on their scores, with the top 20 accounts chosen for reporting.

7. Deliverables:

- The entire analytic code was stored as a Jupyter Notebook/Python script.
- Calculated Metrics CSV: A CSV file containing the calculated metrics was generated (Final_account.csv).
- Top 20 Accounts List: The top 20 accounts were identified based on their rank and performance indicators.

8. Outcomes:

The analysis produced the following top 20 accounts based on the Rank system:

	Rank	Port_ID	Score	pnl	Win_Positions	Win_rate	ROI	Sharpe_Ratio	Total_Positions
96	1.0	4020204877254599680	36599.321467	71998.855953	1993	32.942149	9.967452	-0.445459	6050
75	2.0	3999240873283311617	21998.215633	42574.473679	2366	52.321981	5.893969	-0.217712	4522
98	3.0	4021669203289716224	13372.797511	26427.331592	528	31.391201	3.658574	-0.600381	1682
50	4.0	3960874214179953664	10154.177423	19567.471286	1233	51.568381	2.708902	-0.332736	2391
17	5.0	3907081197088384000	9770.497509	18015.997370	2540	61.397148	2.494117	0.290320	4137
62	6.0	3986814617275053313	9003.183289	16337.461881	2780	78.221722	2.261742	0.391549	3554
48	7.0	3956076827719377409	9001.170998	16790.012238	2019	51.557712	2.324393	-0.088706	3916
107	8.0	4028701921959171840	8913.088043	17601.401398	373	54.294032	2.436721	-0.825975	687
5	9.0	3788465932399412480	7391.569777	13960.966457	1369	44.607364	1.932743	-0.280950	3069
99	10.0	4022565861939831809	7210.781638	14197.577076	372	52.765957	1.965499	-1.187391	705
63	11.0	3987739404272887297	6646.948275	12464.606315	1381	22.826446	1.725588	-0.267964	6050
27	12.0	3931992636670880512	5916.769905	10374.965290	2430	40.165289	1.436300	-0.154440	6050
84	13.0	4008711265867865600	5699.466031	11298.706390	166	42.131980	1.564182	-1.327305	394
30	14.0	3936410995029308417	5342.852748	10681.514000	6	30.000000	1.478738	0.298094	20
32	15.0	3939318616482048768	4923.365214	9775.389110	118	37.460317	1.353295	-0.236557	315
73	16.0	3998572645139652353	4830.558582	9603.585360	95	35.714286	1.329511	-0.670886	266
51	17.0	3962024093501274881	4758.604928	9314.494060	337	26.472899	1.289490	-1.168020	1273
125	18.0	4031493134338259457	4746.500624	9082.698288	683	36.156697	1.257400	-2.066688	1889
37	19.0	3944088772635000577	4623.155907	7581.491983	2774	45.752928	1.049574	-0.023377	6063
93	20.0	4017263283405469185	4181.080938	7141.366419	2034	33.619835	0.988644	-0.702526	6050

9. Conclusion:

The top-performing accounts were effectively identified using calculated financial parameters. The findings emphasize the need of utilizing a complete scoring system to evaluate account performance, which might be useful in future investment decisions.