Data Analysis Report

By Refat Ahmed

September 28, 2025

Dataset Overview

- **Transactions analyzed:** 59,317

- **Unique logins (users):** 600

- **Columns present:** 14

  - `login, ticket, symbol, type, open_time, close_time, open_price, close_price, stop loss, take profit, pips, reason, volume, profit`

Data Quality

The dataset demonstrates the following qualities:

- No missing values were detected.

- No duplicate records were found.

- The dataset maintains a consistent structure across all entries.

.

Exploratory Data Analysis (EDA)

Profit Distribution

- **Mean profit per trade:** ≈ 22.6

- **Median profit:** $2.0 \rightarrow$ distribution is right-skewed.

- **Minimum profit:** $-12{,}250$, **Maximum profit:** $19{,}061$

- **Standard deviation:** $687 \rightarrow$ indicates wide variation in profitability.

Trade Characteristics

- A total of 63 different symbols were traded, with **XAUUSD** being the most frequent.

- Both trade types are represented: **Buy** and **Sell**.

	ticket	open_price	close_price	stop loss	take profit	pips	reason	volume	profit
count	5.931700e+04	59317.000000	59317.000000	59317.000000	5.931700e+04	5.931700e+04	59317.000000	59317.000000	59317.000000
mean	4.820009e+07	9247.848230	9249.311204	6119.568535	4.895936e+03	1.115243e+02	7.236711	176.008463	22.606170
std	2.781208e+07	20700.786785	20702.285285	17961.190194	1.736828e+04	3.093160e+04	6.774457	1768.198799	687.227407
min	7.631230e+05	0.308400	0.325050	0.000000	0.000000e+00	-1.465680e+06	0.000000	1.000000	-12250.000000
25%	2.785692e+07	70.730000	70.940000	0.000000	0.000000e+00	-1.890000e+02	1.000000	20.000000	-102.000000
50%	5.709664e+07	2645.190000	2645.790000	1.248030	9.317200e-01	7.200000e+00	4.000000	51.000000	2.000000
75%	7.241683e+07	2755.480000	2755.220000	2667.400000	2.648650e+03	2.430000e+02	16.000000	150.000000	81.860000
max	8.051714e+07	109111.000000	108304.000000	442230.000000	1.085760e+06	1.560000e+06	17.000000	100000.000000	19061.100000

8 rows × 27 columns

Figure 1: Profit distribution across all trades.

Profitability Analysis

Overall Summary

- Total profit of Top 10 Most Profitable Logins: 291,508.07

- Total profit of Top 10 Least Profitable Logins: -114,355.67

Top 10 Most Profitable Logins

Login Encoded	Login Name	Profit
396	13378390	53891.98
511	55009560	28475.44
50	13088202	27848.61
146	13205503	27049.34
40	13070589	27023.68
498	55008451	27021.14
147	13205506	26494.85
370	13361147	25136.16
0	11173702	24301.54
519	55010677	24265.33

Table 1: Top 10 most profitable logins.

Top 10 Least Profitable Logins

Login Encoded	Login Name	Profit
61	13103928	-14778.82
329	13333728	-13868.00
539	55011482	-12215.00
23	13018096	-12194.31
193	13251499	-11405.24
462	13410127	-10571.86
507	55009211	-10102.92
226	13276691	-10010.77
34	13054222	-9635.14
76	13131614	-9573.61

Table 2: Top 10 least profitable logins.
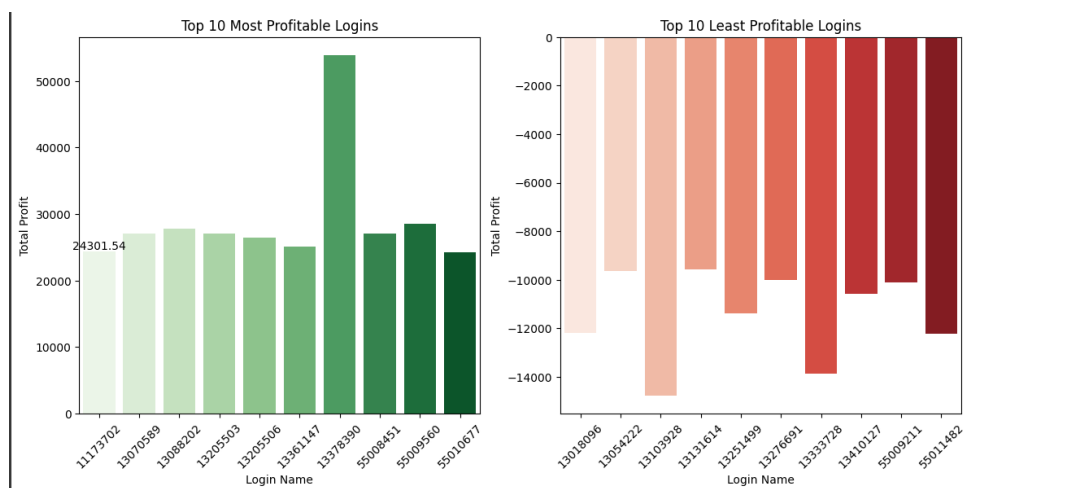
Figure 2: Comparison of top 10 most and least profitable logins.

Cumulative Profit Over Trade Sequence

The top 5 logins ranked by cumulative profit are:

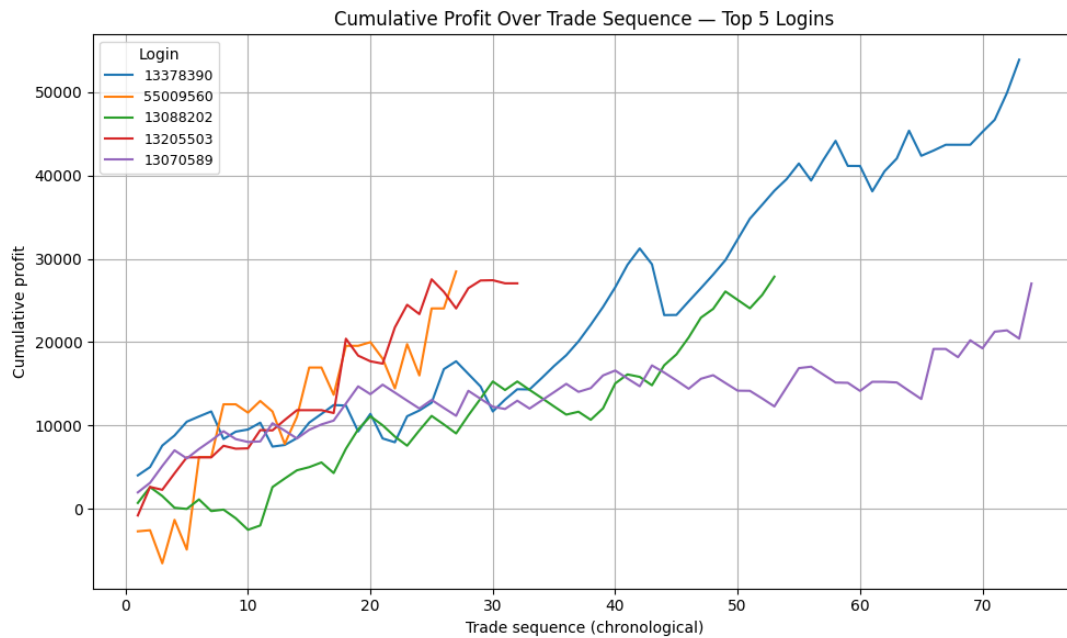$$\{13378390,\ 55009560,\ 13088202,\ 13205503,\ 13070589\}$$

Figure 3: Cumulative profit over trade sequence for the top 5 most profitable logins.

The least 5 logins ranked by cumulative profit are:

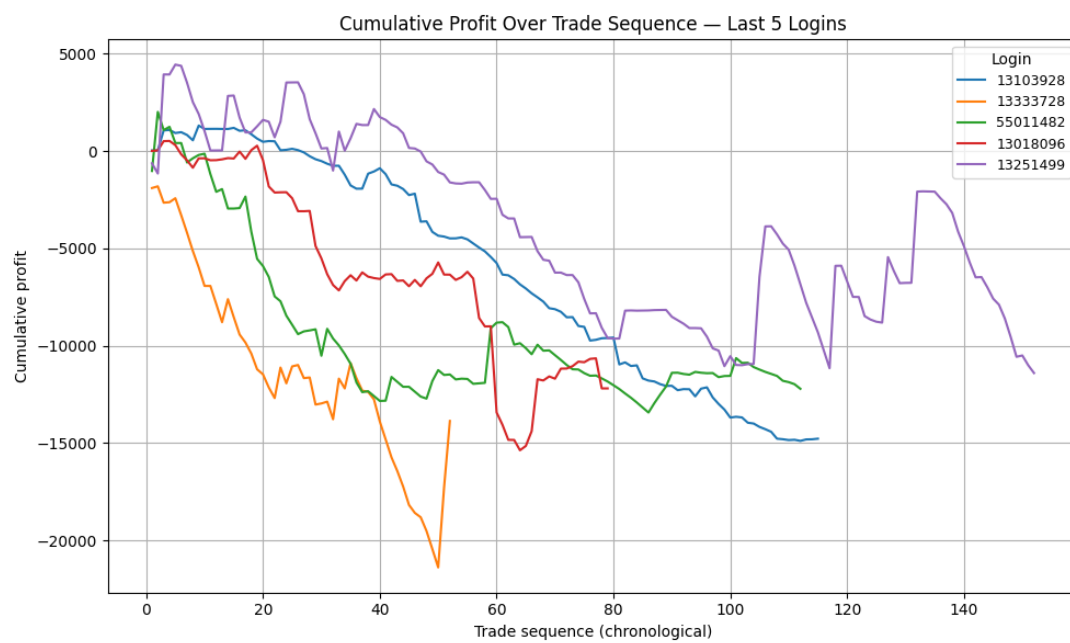$$\{13103928,\ 13333728,\ 55011482,\ 13018096,\ 13251499\}$$

Figure 4: Cumulative profit over trade sequence for the 5 least profitable logins.

Correlation Analysis

Pearson (linear) and Spearman (rank-based) correlations between `profit` and selected features. Table 3 summarizes the key results from Pearson (linear) and Spearman (rank-based) correlations between `profit` and numeric features.

Feature	Pearson r	Spearman ρ	p-value (Spearman)
pips	0.138	0.874	$< 10^{-249}$
reason	0.079	0.140	$< 10^{-256}$
type_encoded	-0.021	-0.028	4.42×10^{-11}
open_time_hour	-0.016	-0.085	1.17×10^{-28}
stop loss	0.012	-0.078	3.21×10^{-22}
volume	0.012	-0.040	1.73×10^{-6}
open_price	-0.009	0.016	2.42×10^{-2}
close_price	-0.009	0.016	2.41×10^{-2}
take profit	0.010	0.056	1.46×10^{-1}

Table 3: Pearson and Spearman correlations between `profit` and selected features.
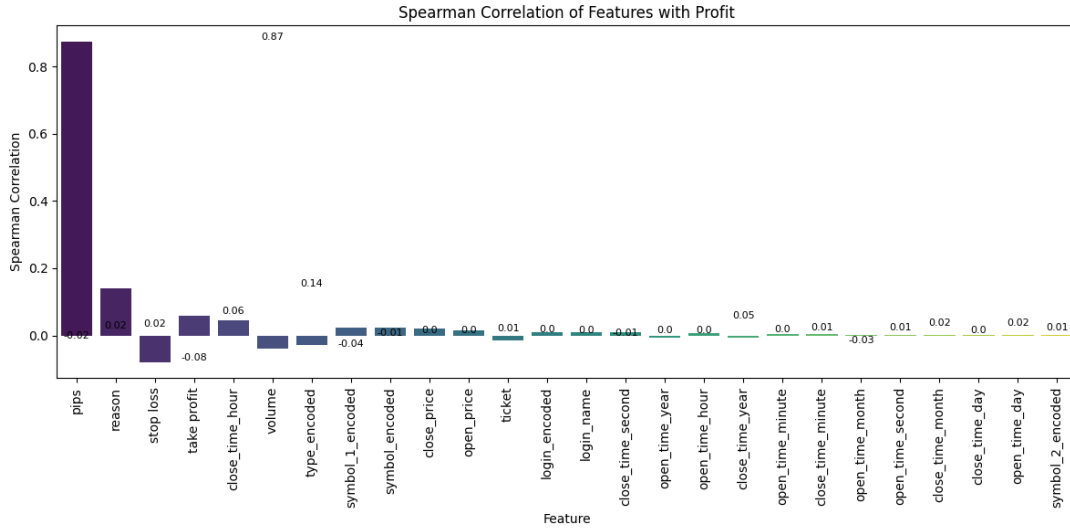
Most numeric features show negligible linear ($|r| < 0.1$) and monotonic ($|\rho| < 0.1$) associations with profit, suggesting limited direct predictive power. The only notable exception is `pips`, where Pearson's $r = 0.138$ suggests a weak linear link, but Spearman's $\rho = 0.874$ indicates a very strong monotonic relationship. This implies that while profit does not increase linearly with pips, their ranks are closely aligned, possibly due to the influence of trade volume or instrument-specific scaling. Other features such as `reason` and `stop loss` show weak associations, and their statistical significance is driven mainly by the large sample size rather than meaningful effect sizes.

Further analysis confirms that the relationship between pips and profit is primarily monotonic rather than linear. Spearman correlation and the scatter plot indicate that higher-magnitude pips generally correspond to higher-magnitude profits or losses,
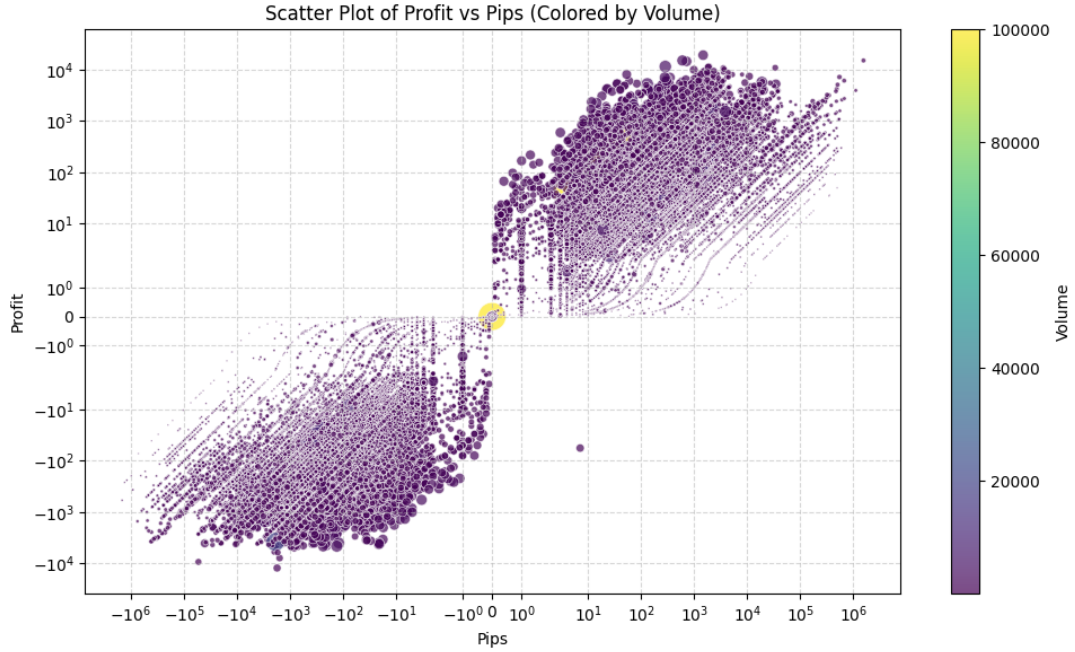
Figure 5: Spearman and Pearson Correlation.

especially for trades with larger volumes. This suggests that modeling approaches accounting for non-linear patterns or rank-based relationships would better capture the influence of pips on profit.

ANOVA Test and Kruskal-Wallis Test for Categorical vs Profit Relations

To understand the relationship between categorical features and profits, we applied both parametric and non-parametric tests. For normally distributed groups, a t-test was used, while the Kruskal-Wallis test was applied to non-normal distributions or groups with unequal variances. The results are summarized in Table 4. The symbol_1 and symbol_2 indicates the currency pair individually and symbol_encoded indicates the currency pair.

Table 4: Statistical tests between categorical features and profit.

Feature	Test	Statistic	p-value	Significance
login_encoded	Kruskal-Wallis	3981.05	0.0	Significant
symbol_encoded	Kruskal-Wallis	274.997	1.31e-28	Significant
symbol_1_encoded	Kruskal-Wallis	142.037	1.75e-15	Significant
symbol_2_encoded	Kruskal-Wallis	83.390	2.17e-09	Significant
type_encoded	t-test	5.163	2.44e-07	Significant

These results indicate that all tested categorical features have a statistically significant relationship with profit. Non-parametric Kruskal-Wallis tests confirm strong differences across groups for login and symbol features, while the t-test shows a significant effect of type on profit.

Conclusion

In conclusion, this analysis highlights the diverse profitability outcomes across users and trading behaviors within the dataset. While the overall distribution of profit is highly skewed and marked by extreme outliers, the investigation into top-performing and least-profitable logins underscores significant disparities in trading success. The strong monotonic relationship between pips and profit suggests that while linear associations are limited, rank-based measures capture meaningful patterns that can inform future predictive modeling. Furthermore, categorical factors such as login identity, trade symbols, and trade type show statistically significant impacts on profitability, reinforcing the role of trader-specific strategies and market instruments in shaping outcomes.

Overall, the findings emphasize the importance of considering both non-linear dynamics and categorical distinctions in evaluating trading performance. These insights can serve as a foundation for refining trading strategies, enhancing risk management, and guiding the development of more robust analytical models in subsequent research.