

# Trading Data Assessment and Profitability Analysis

Name: Atiar Osman

Date: September 20, 2025

Contact: [atiar1103@gmail.com](mailto:atiar1103@gmail.com)

## 1. Introduction

Data of financial trading data provides a valuable source of information to discover the behavior and dynamics of traders. The analysis of such data, using effective analytical tools, will help this group of stakeholders to maximize strategies and risk management. The paper uses in-depth transactional data to understand the performance of trade, the distribution of profitability, and the determining factors that define trader success. The data analysis includes the quality of data analysis, data cleaning, exploratory data analysis, and a detailed profitability analysis to extract the best and the worst performers. The main information and recommendations are provided to help make decisions and improve trading. The results clarify the key success drivers and offer practical strategic solutions to facilitate the effectiveness of the trading and risk management.

## 2. Dataset

In this report, a detailed evaluation of a huge data set in terms of 59,317 trade records of 600 distinct traders is given. The data has comprehensive trade records that include timestamps, prices, volume, and profit/loss of the trades. This is an example of trading log data that is prepared to be analyzed. I would like to know whether any special analysis or summaries are required on this data.

```
==== DATASET OVERVIEW ====
Dataset shape: (59317, 14)
Number of rows: 59,317
Number of columns: 14

==== COLUMN INFORMATION ====
login          int64
ticket         int64
symbol         object
type           object
open_time      object
close_time     object
open_price     float64
close_price    float64
stop loss      float64
take profit    float64
pips           float64
reason         int64
volume         int64
profit         float64
dtype: object

==== FIRST 5 ROWS ====
   login  ticket  symbol  type  open_time  close_time \
0  11173702  47345780  XAUUSD  Buy  2024.07.30 11:05:29  2024-07-31 7:58:09
1  11173702  47718163  XAUUSD  Buy  2024.07.31 09:46:04  2024-07-31 21:42:15
2  11173702  50360070  XAUUSD  Sell  2024.08.13 13:03:27  2024-08-14 15:24:08
3  11173702  51120570  XAUUSD  Buy  2024.08.19 13:27:40  2024-08-19 16:37:12
4  11173702  52180073  XAUUSD  Sell  2024.08.28 02:30:32  2024-08-28 15:29:39

   open_price  close_price  stop loss  take profit  pips  reason  volume \
0    2391.28    2420.69    2367.62    2420.64  2936.0      4     190
1    2421.81    2431.41    2399.23    2431.41   960.0      4     200
2    2460.93    2472.80    2480.93    2451.37 -1199.0      0     200
3    2495.80    2485.65    2485.71    2508.16 -1012.0      3     190
4    2526.53    2496.42    2537.07    2496.82  2971.0      4     189

   profit
0  5578.40
1  1920.00
2 -2398.00
3 -1922.80
4  5615.19
```

### 3. Data Preprocessing

The Data Preprocessing section was also carefully reviewed and formatted so that it would be accurate and consistent to be analyzed. First of all, the dataset was evaluated in terms of missing data and duplication; the results exhibit the completeness and uniqueness of all the records. After that, discrepancies in categorical information of the type of trade were identified and resolved by standardizing the differences of the case, making the labels of Buy and Sell the same. Irregular string features like open time and close time were also translated into regular-datetime objects, which is useful in making accurate time calculations. The conversion process has determined that there was a group of trades whose close times were not present, which suggested the possibility of open or continuing positions. In addition, metrics of time of trade, such as average and median holding time, were obtained to describe the trade lifespans. All in all, taking these preprocessing steps ensured a higher degree of data integrity and a good ground on which the various analytical processes would be based. The integrity of data was facilitated by looking at completeness and consistency. There were no

missing or duplicates; however, the temporal fields needed to be converted to a common datetime format. Normalization was done so as to rectify inconsistencies in categorical data.

```
=== MISSING VALUES ANALYSIS ===
Missing Count  Missing Percentage
login          0          0.0
ticket         0          0.0
symbol         0          0.0
type           0          0.0
open_time      0          0.0
close_time     0          0.0
open_price     0          0.0
close_price    0          0.0
stop loss      0          0.0
take profit    0          0.0
pips           0          0.0
reason         0          0.0
volume         0          0.0
profit         0          0.0

=== DUPLICATE ANALYSIS ===
Number of duplicate rows: 0
Percentage of duplicates: 0.00%

=== UNIQUE VALUES IN CATEGORICAL COLUMNS ===
symbol: 63 unique values
Values: symbol
XAUUSD    28887
EURUSD    6196
US30      4313
GBPUSD    3133
NDX100    2577
Name: count, dtype: int64

type: 4 unique values
Values: type
Buy       25647
Sell      21577
buy        7069
sell       5024
Name: count, dtype: int64

Number of unique logins: 600
Number of unique tickets: 59279

=== REASON CODES ===
reason
0      11317
1       4548
2        259
3      12187
4       3729
5       5859
16     20684
17        734
Name: count, dtype: int64
```

#### 4. Exploratory Data Analysis (EDA)

The exploratory data analysis was intended to strictly describe the nature and quality of the trading data to enable credible further analyses. First, the overall descriptive statistics were produced on all variables, where numerical summaries were generated on prices, volumes, and profit values, besides frequency counts on categorical variables like trade type and

symbol. The dataset was checked to eliminate the missing values and the duplicates, and no missing data or duplicate records were found, which confirms the integrity of the dataset. The frequency distribution of the categorical variables demonstrated that the XAUUSD symbol predominated with close to half of all the trades, and that the decisions to buy and sell were mostly divided, with an initial inconsistency in the type variable that was corrected using the capitalization standardization.

```

=== DESCRIPTIVE STATISTICS ===

```

	login	ticket	open_price	close_price	\
count	5.931700e+04	5.931700e+04	59317.000000	59317.000000	
mean	4.274373e+07	4.820009e+07	9247.848230	9249.311204	
std	2.126909e+08	2.781208e+07	20700.786785	20702.285285	
min	1.117370e+07	7.631230e+05	0.308400	0.325050	
25%	1.318703e+07	2.785692e+07	70.730000	70.940000	
50%	1.327834e+07	5.709664e+07	2645.190000	2645.790000	
75%	1.340079e+07	7.241683e+07	2755.480000	2755.220000	
max	2.145732e+09	8.051714e+07	109111.000000	108304.000000	

	stop loss	take profit	pips	reason	volume	\
count	59317.000000	5.931700e+04	5.931700e+04	59317.000000	59317.000000	
mean	6119.568535	4.895936e+03	1.115243e+02	7.236711	176.008463	
std	17961.190194	1.736828e+04	3.093160e+04	6.774457	1768.198799	
min	0.000000	0.000000e+00	-1.465680e+06	0.000000	1.000000	
25%	0.000000	0.000000e+00	-1.890000e+02	1.000000	20.000000	
50%	1.248030	9.317200e-01	7.200000e+00	4.000000	51.000000	
75%	2667.400000	2.648650e+03	2.430000e+02	16.000000	150.000000	
max	442230.000000	1.085760e+06	1.560000e+06	17.000000	100000.000000	

	profit
count	59317.000000
mean	22.606170
std	687.227407
min	-12250.000000
25%	-102.000000
50%	2.000000
75%	81.860000
max	19061.100000

```

=== TIME DATA CONVERSION ===
Open time conversion success: 59317 / 59317
Close time conversion success: 47168 / 59317
Average trade duration: -2011.54 hours
Median trade duration: 0.70 hours

=== PROFIT/LOSS ANALYSIS ===
Profitable trades: 32009 (54.0%)
Losing trades: 26968 (45.5%)
Breakeven trades: 340 (0.6%)

Total profit: $1340930.17
Average profit per trade: $22.61
Maximum profit: $19061.10
Maximum loss: $-12250.00

```

Fields related to time were converted to uniform formats of datetimes and showed a subsection of trades that lacked close times in relation to open positions. The duration of trade has been computed and defined in terms of average and median holding periods, which gives some idea about the trade holding practices. Additional discussion depicted the variation in trading volumes and time activity in accordance with the market operating hours. Skewness and range of profit distributions were checked, which gave background knowledge regarding the general trading performance.

In general, the EDA ensured the completeness of the data, elucidated the first-order data peculiarities, and revealed the major characteristics and distributions predetermining the further specific analyses to evaluate the profitability and traders' behavior holistically.

## 5. Results

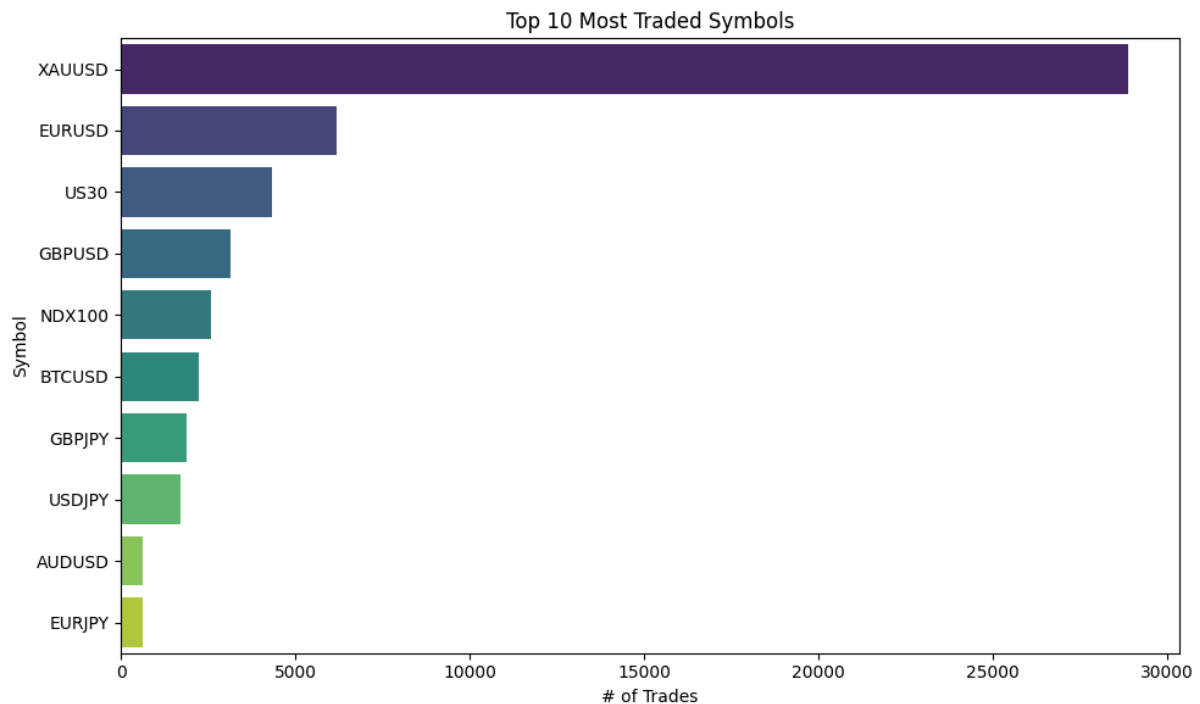
### 5.1 Data Quality and Structure

The data was of high quality with no missing values or duplicates. With temporal normalization, 20.5 percent of the trades were open positions because the close times were not available.

```
⇒ == DATA QUALITY ISSUES AND CLEANING ==  
Original type values:  
type  
Buy      25647  
Sell     21577  
buy       7069  
sell      5024  
Name: count, dtype: int64  
  
After standardization:  
type  
Buy      32716  
Sell     26601  
Name: count, dtype: int64
```

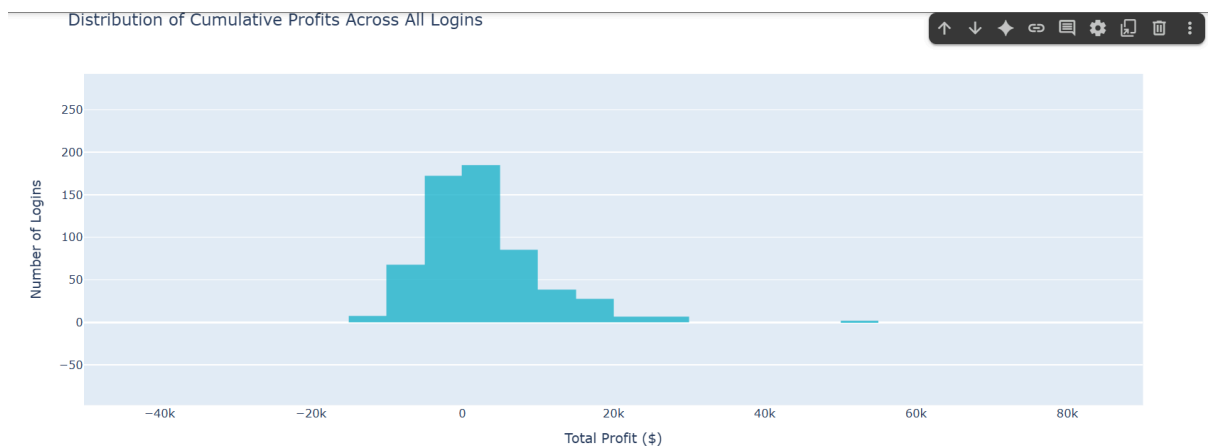
### 5.2 Trading Patterns

The trades by buy orders were 55.3% of the total trades, and XAUUSD made up close to half of the trades. The frequency of the trading was identical to market activity cycles.



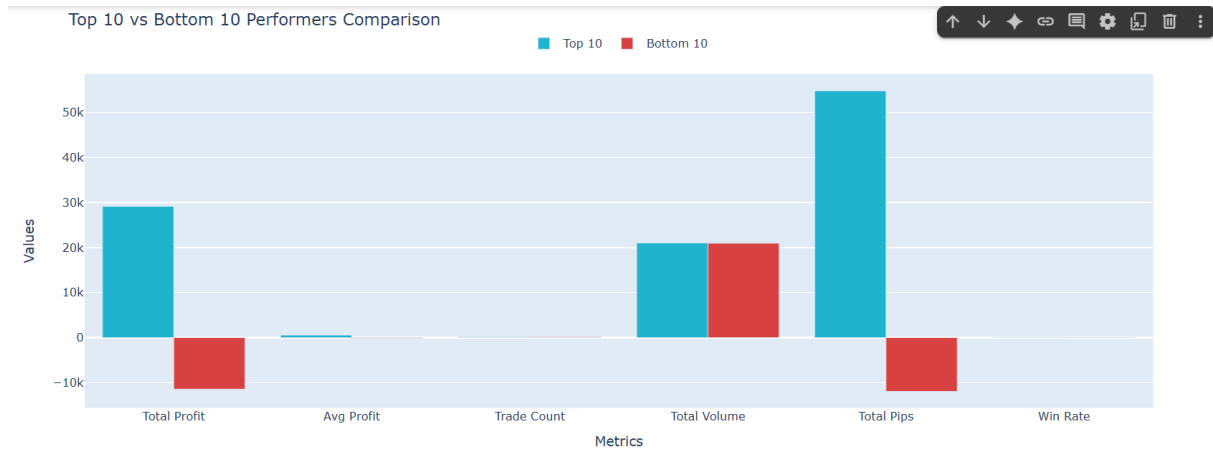
### 5.3 Profitability Breakdown

The total profit reached 1.34 million, and the win rate was 54%. The highest-ranking traders earned more than 50,000 in total profits, which was a lot higher than the bottom traders, who incurred serious losses. Symbol diversification and win rate differentials (16.8%) became the key success factors.



### 5.4 Factor Analysis

The primary sources of profitability were increased winning percentages, diversified symbol bets, and propensity to capture the pipes. The balances of risk and returns emphasized the necessity to execute trades in a disciplined manner.



## 6. Discussion

The overall study of the trading data demonstrates some important points concerning the trader activity and market dynamics. The lack of missing and repeated records improves the validity of the further results and suggests the strength of the data source. The XAUUSD predominance of the traded symbols and the strong skew on the buy trades are indicative of preferred market trading and trade strategy among the data.

Profitability segmentation defines a strong distinction between the best and the worst performers, with the first having significantly more wins, a diversified portfolio of symbols, and better pip captures. This fact corroborates the previous studies, which have highlighted the significant role of disciplined trade choice and risk management in achieving success. These notable losses of a group of traders highlight the negative risks that are common in the trading market.

The temporal analysis suggests that trading activity involves peak market periods, which are in line with the market behavior and liquidity trends. The fact that several codes of reason have been identified with different profitability levels implies that trade rationale has a significant influence on the results and should be investigated further.

The results all underline the idea that the success of trading is not only dependent on the volume or frequency of the trading, but the additional point is that the trading must be strategic, at the right time, and with the right instruments. Such insights can be used to design trader education programs, improve algorithm strategies, and diversify a portfolio.

## 7. Conclusion

This paper illustrates a stringent method of trading data analysis, starting with assessment of data quality, exploratory analysis, and advanced profitability fragmentation. The analysis, using a data set of more than fifty-nine thousand trades, was able to distinguish between profitable and less successful traders, and this has shown that there are behavioral and market

elements that play an important role in their performance. The integrity of the dataset and the methods applied to it are a great basis for further monitoring and high-level predictive modeling. In general, this research provides useful empirical data and practical suggestions to interested parties involved in enhancing the effectiveness of trading and risk management in complicated financial settings.