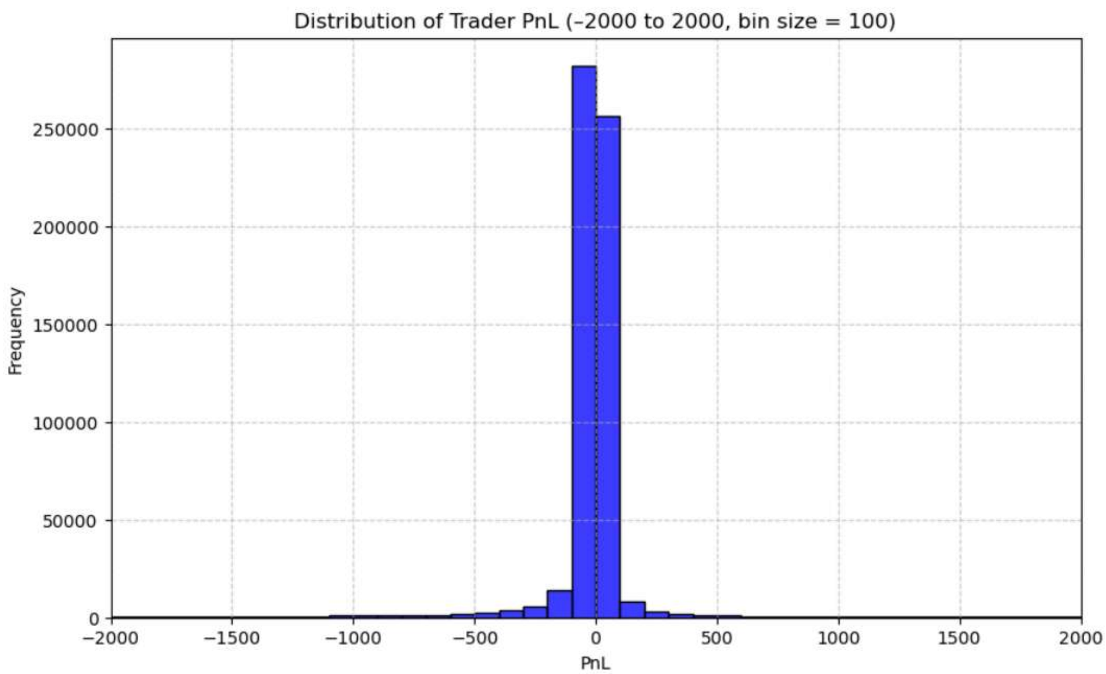
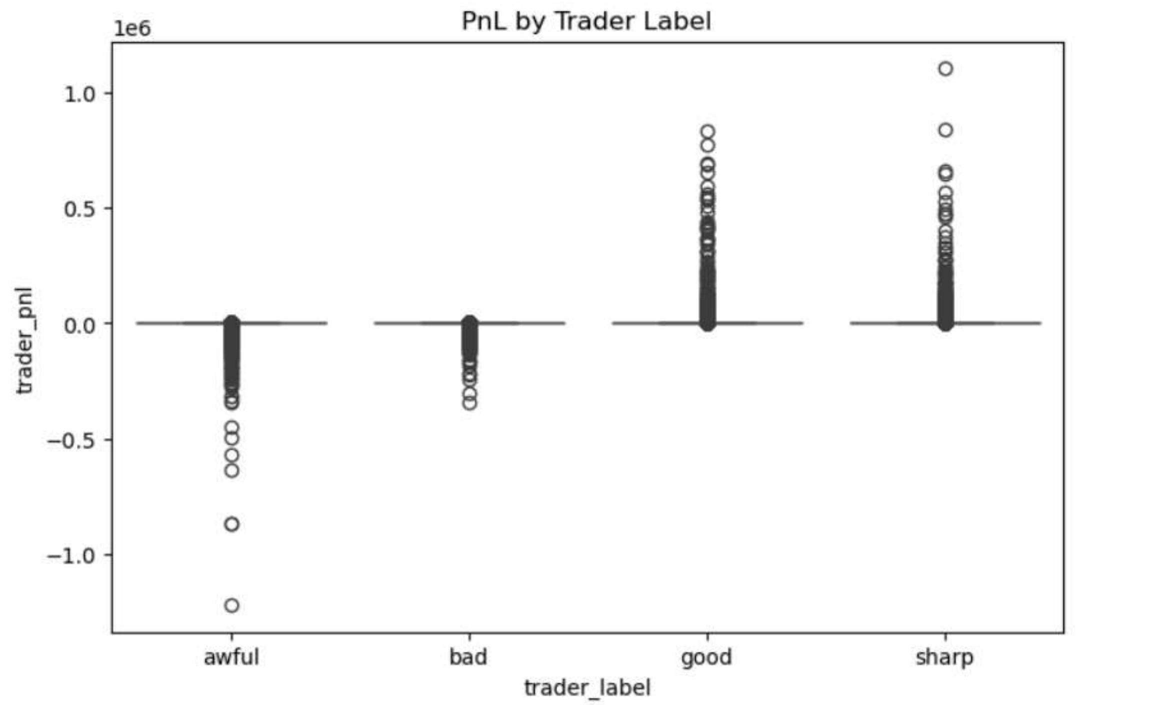
This dataset consists of over 600,000 records with 41 features describing trader performance, behavior, and topical involvement. It includes both numeric variables (PnL, Volume, etc) and categorical labels (trader classification: awful, bad, good, sharp).

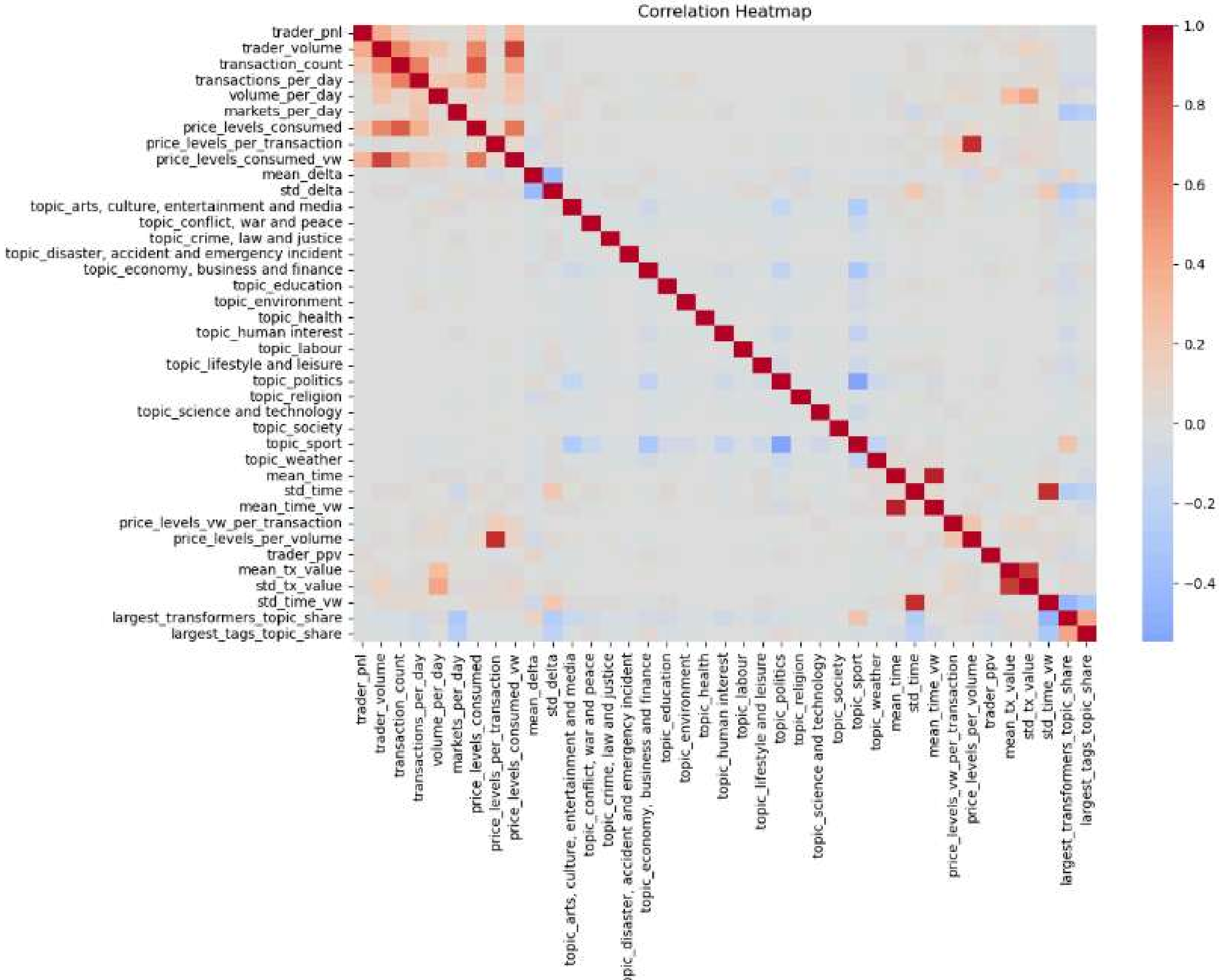
The distribution of trader profit and loss (PnL) is highly skewed, as visualized in the histogram. Most traders cluster around small PnL values, there are some users who go past this mark both ways. This suggests that while most traders exhibit modest results, a small subset achieves extreme outcomes, which is consistent with the thought I had going into this EDA.

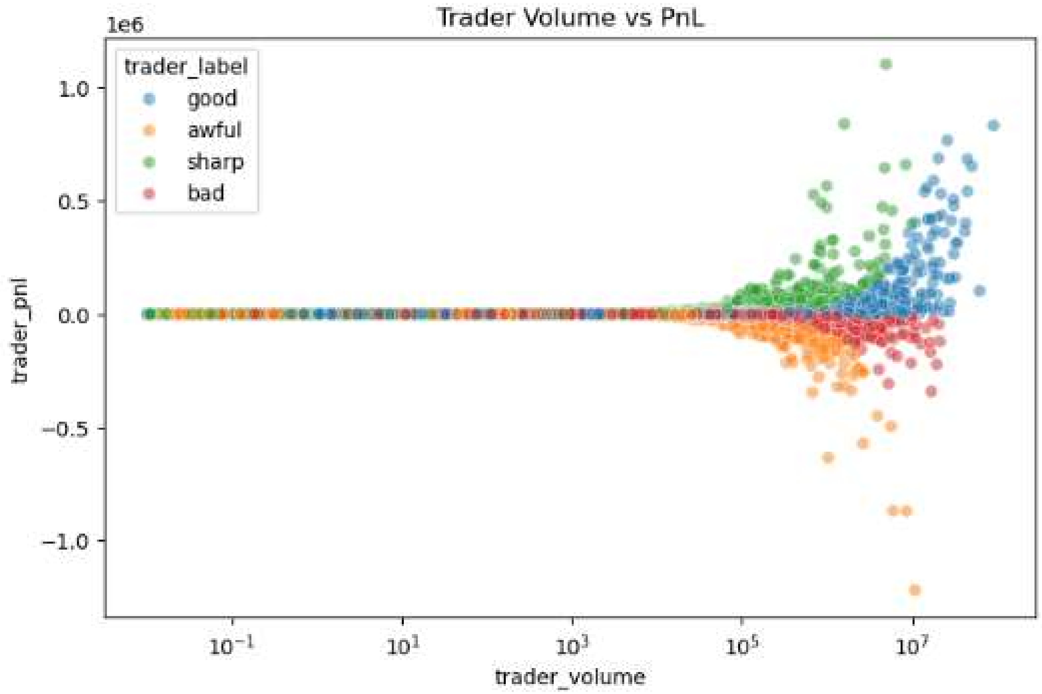


The boxplot of PnL across trader labels reinforces this. “Sharp” traders consistently outperform with average PnL above +550, while “Awful” traders average nearly -476. “Good” traders show steady profitability, and “Bad” traders small losses. The largest winner had a PnL of 1.1 million, while the largest loser had a PnL of -1.1 million. These are both extreme outliers compared to the averages I explained above.



Correlation analysis revealed that trader volume has the strongest positive correlation with PnL (0.41). This indicates that more active traders tend to achieve better outcomes. This was not what I expected going into it, and I was very surprised by this revelation. Related features such as order book depth consumption (price_levels_consumed_vw) also correlate positively, reinforcing that market engagement depth supports profitability. Transaction count and transactions per day also correlate positively with PnL, though more weakly. In contrast, topic variables such as “economy”, “politics” and “sports” showed very small correlations, which implies that the topics themselves don’t really matter. The scatter plot of volume versus PnL highlighted this further. High volume traders are disproportionately represented among profitable outcomes, especially within the “sharp” label group. Conversely, low volume traders dominate the “awful” and “bad” categories. This emphasizes that consistent, large-scale activity,rather than occasional trades, is central to long term profitability.


Correlation Heatmap
trader_pnl
trader_volume
transaction_count
transactions_per_day
volume_per_day
markets_per_day
price_levels_consumed
price_levels_per_transaction
price_levels_consumed_vw
mean_delta
std_delta
topic_arts, culture, entertainment and media
topic_conflict, war and peace
topic_crime, law and justice
topic_disaster, accident and emergency incident
topic_economy, business and finance
topic_education
topic_environment
topic_health
topic_human interest
topic_labour
topic_lifestyle and leisure
topic_politics
topic_religion
topic_science and technology
topic_society
topic_sport
topic_weather
mean_time
std_time
mean_time_vw
price_levels_vw_per_transaction
price_levels_per_volume
trader_ppv
mean_tx_value
std_tx_value
std_time_vw
largest_transformers_topic_share
largest_tags_topic_share
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4



In summary, the exploratory data analysis shows that factors such as volume and book depth are the most important drivers of profitability. The dataset could be used for machine learning applications, particularly classification (predicting trader label) or regression (forecasting PnL). Topic features, however, contribute little explanatory power, and could likely be excluded or down weighted in modeling. Overall, the data strongly supports insights into how trader activity translates into profitability