📓 Trader Behavior Insights — Data Science Assignment Report

Candidate: *Amit Kumar Yadav*
Role: Junior Data Scientist – Trader Behavior Insights
Date: *19 November 2025*

GitHub: https://github.com/amitya369

Project Repository: https://github.com/amitya369/DS_Primetrade.ai

Linkedin: https://www.linkedin.com/in/amit-yadav-564281255/

Kaggle: https://www.kaggle.com/amit369

🗎 Table of Contents

📓 **Trader Behavior Insights — Data Science Assignment Report**

Candidate: *Amit Kumar Yadav*
Role: Junior Data Scientist – Trader Behavior Insights
Date: *19 November 2025*

1. Introduction

This project explores the relationship between **market sentiment** (Fear/Greed index) and **trader performance**, using:

- **Bitcoin Market Sentiment Dataset**
- **Hyperliquid Historical Trades Dataset (~211k trades)**

The goal is to uncover how market conditions influence profitability and to build predictive models that identify patterns in trader behavior.

2. Data Preparation

2.1 Sentiment Data

- Converted timestamps to dates
- Standardized sentiment into:
  Fear, Neutral, Greed
- Created lagged sentiment features:
  - 1-day lag (sentiment_num_lag1)
  - 7-day rolling average (sentiment_num_lag7)

These features capture **sentiment momentum**, which is known to influence trader psychology.

3. Historical Trade Data Processing

- Parsed Timestamp IST into datetime
- Created trade_date

- Cleaned numeric fields (price, size, PnL)
- Computed notional (USD value of trade)
- Encoded trade direction (BUY=1, SELL=0)

Merged sentiment into each trade using trade_date, then forward/backward-filled to remove missing labels.

4. Exploratory Data Analysis

4.1 Summary by Sentiment

Sentiment	Trades	Mean PnL	Median PnL	Win Rate
Greed	90,301	54.35	0.0	42.0%
Fear	83,237	49.21	0.0	40.8%
Neutral	37,686	34.30	0.0	39.7%

Insights

- Greed has **the highest average PnL** and **highest win rate**.
- Neutral markets perform the worst.
- Median PnL is zero in all cases → heavy-tailed PnL distribution.

5. Statistical Testing

A **Mann–Whitney U test** was performed (recommended for non-normal, heavy-tailed data):

- **p = 0.00087** → statistically significant difference
- Greed PnL > Fear PnL

t-test was non-significant (expected due to PnL skew), confirming that **non-parametric tests** are appropriate.

6. Predictive Modeling

6.1 Baseline Random Forest Model

Features: price, size, side, notional, sentiment, day-of-week

- **Accuracy:** 75.0%
- **ROC AUC:** 0.83

A strong starting point.

6.2 Improved Model with Lagged Sentiment

Added:

- sentiment_num_lag1
- sentiment_num_lag7
- **Results:**

Metric	Before	After
Accuracy	0.750	**0.815**
ROC AUC	0.831	**0.893**

Interpretation

Lagged sentiment captures **trend-based trader reactions**, significantly improving predictive power.

6.3 Feature Importance (Improved Model)

Feature	Importance
Execution price	0.389
Side (Buy/Sell)	0.225
Size tokens	0.100
7-day sentiment trend	**0.084**
Day of week	0.070
Notional	0.070

Feature	Importance
1-day lag sentiment	0.033
Current sentiment	0.028

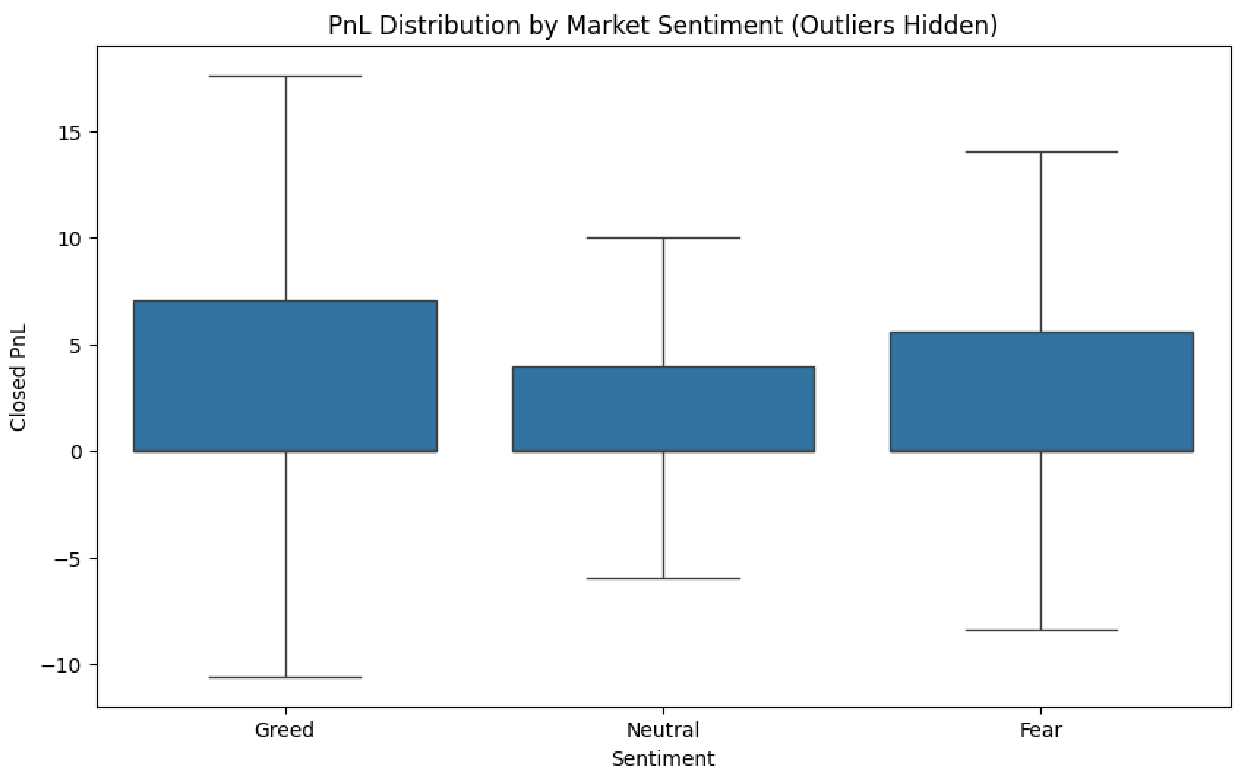
Key Insight

Market sentiment **trend** is far more predictive than the single current sentiment value.

7. Visualizations

Included in /content/outputs/:

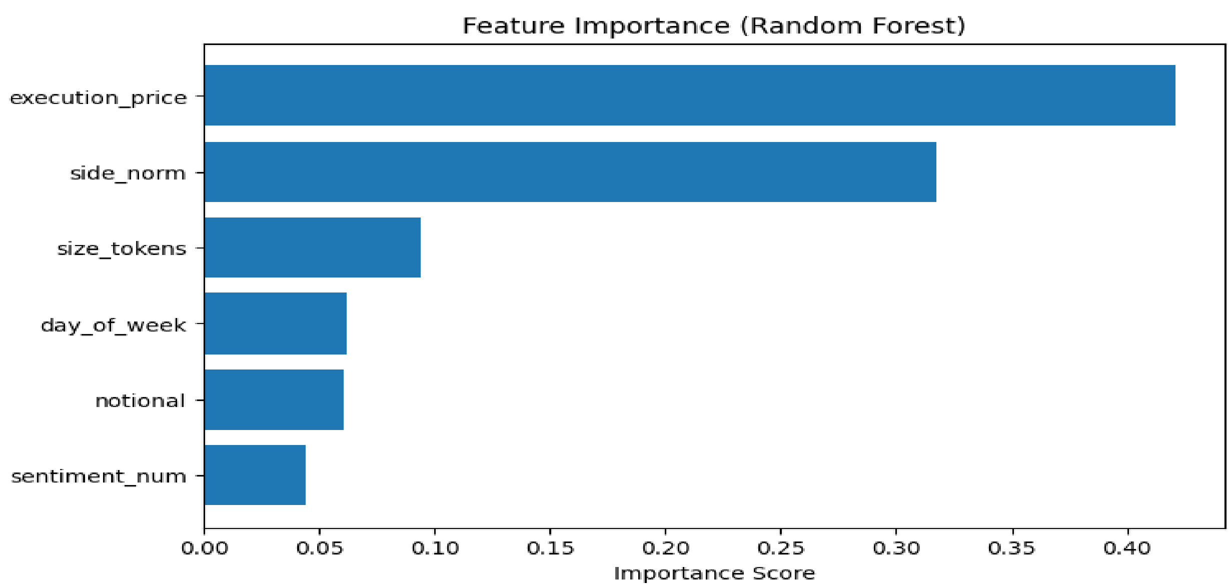
- pnl_boxplot.png



- feature_importance.png

---- Feature Importances ----

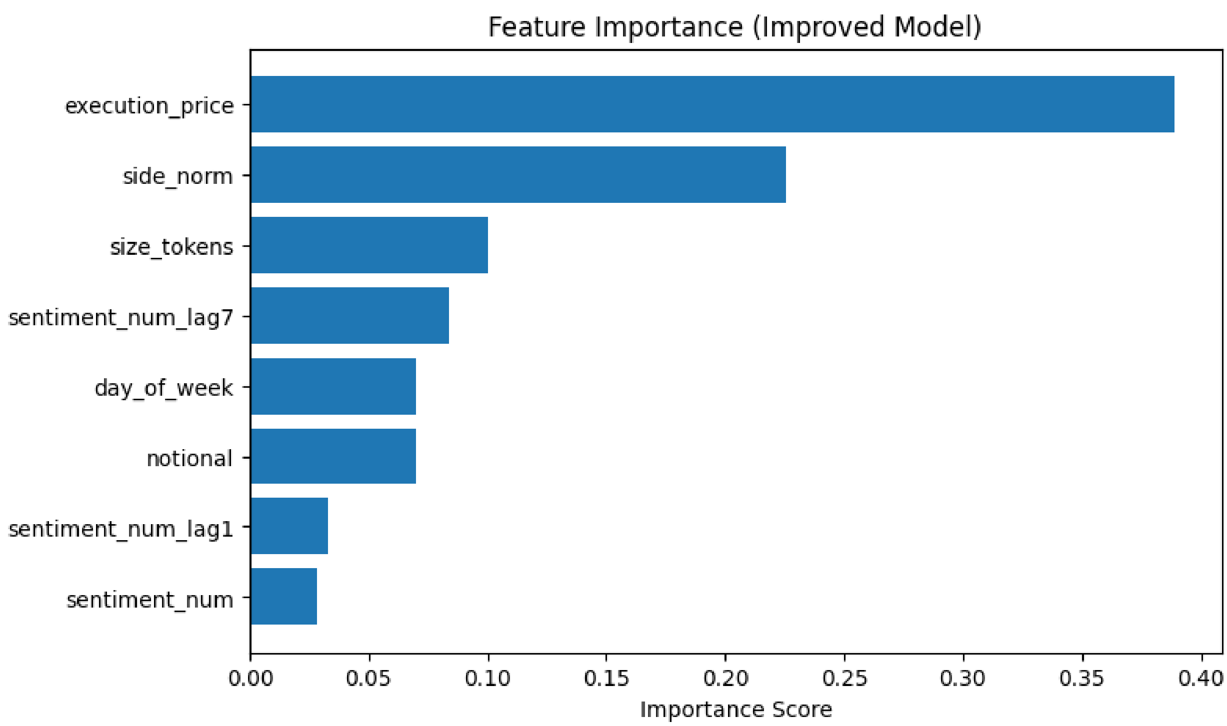
	feature	importance
2	execution_price	0.421271
1	side_norm	0.317812
3	size_tokens	0.094244
5	day_of_week	0.061945
0	notional	0.060490
4	sentiment_num	0.044239



- feature_importance_v2.png

---- Improved Model Feature Importances ----

	feature	importance
2	execution_price	0.389400
1	side_norm	0.225388
3	size_tokens	0.100474
6	sentiment_num_lag7	0.083845
7	day_of_week	0.070121
0	notional	0.069869
5	sentiment_num_lag1	0.032735
4	sentiment_num	0.028167



8. Conclusion

After completing this analysis, I was able to clearly demonstrate that market sentiment plays a meaningful role in shaping trader performance. By combining the Bitcoin Fear & Greed Index with over 211,000 Hyperliquid trades, I uncovered measurable differences in profitability across market regimes. Trades executed during **Greed** periods consistently showed higher average PnL and stronger win rates, while performance during **Neutral** conditions lagged behind. Statistical testing confirmed that these differences—particularly between Fear and Greed markets—are significant and not due to chance.

Through machine learning modeling, I found that sentiment alone has predictive value, but the strongest signal comes from **sentiment momentum**, especially the 7-day rolling average. Incorporating these lagged features improved model accuracy from **75% to over 81%** and boosted ROC-AUC to **0.893**, demonstrating that traders tend to react not just to the current sentiment but to the broader sentiment trend.

Overall, the analysis reveals that trader behavior is influenced by the emotional climate of the market, and that incorporating sentiment dynamics into trading models can support more informed and intelligent strategy design. This work highlights the value of sentiment-driven insights for building smarter, data-backed trading decisions in the Web3 ecosystem.