Data Science Assignment

Analysis Report

Candidate: Fionna Immaculate

Date: 20 November 2025

1. Introduction

This project explores how trader performance on Hyperliquid changes under different Bitcoin market sentiment conditions: Fear, Neutral, and Greed.

The analysis combines Hyperliquid's historical trader data with the Bitcoin Fear & Greed Index, followed by EDA, statistical testing, and predictive modeling to understand what truly drives trader profitability.

2. Data Sources

2.1 Hyperliquid Historical Trader Data

This dataset contains detailed execution-level trading information for multiple traders on Hyperliquid. It includes trade-level metrics, account identifiers, timestamps, position data and PnL information. The exact columns are:

- Account: Unique trader wallet/account identifier.
- Coin: Trading pair or instrument.
- Execution Price: Price at which the trade was executed.
- Size Tokens: Trade size in token units.
- Size USD: Trade size converted to USD.
- Side: BUY or SELL action.
- Timestamp IST: Execution timestamp in IST timezone.
- Start Position: Position held before this trade.
- Direction: Direction of the trade (Buy/Sell).
- Closed PnL: Profit or loss realized upon trade closure.
- Transaction Hash: Unique blockchain-level identifier.
- Order ID: Unique trade order reference.
- Crossed: Boolean flag indicating a crossed order.

- Fee: Trading fee paid.
- Trade ID: Execution identifier.
- Timestamp: Additional timestamp column (used for verification).

This dataset allows analysis of trader behavior, activity patterns, PnL trends, and daily performance aggregation.

2.2 Bitcoin Fear & Greed Index

This dataset tracks market sentiment for Bitcoin over time. It assigns each day a sentiment category based on multiple market indicators. The dataset includes:

- timestamp – UNIX epoch timestamp.
- value – Numeric Fear & Greed Index score (0–100).
- classification – Market sentiment label (e.g., Fear, Greed, Neutral).
- date – used for merging with trade data.

This dataset provides external market sentiment for each day.

2.3 Data Preparation & Cleaning

Before analysis, the following preprocessing steps were applied:

- Standardized column names to snake_case.
- Converted timestamp_ist into a proper datetime object.
- Extracted date to perform daily aggregation.
- Computed daily-level performance metrics: total trades, total volume (USD), total PnL, win rate, buy vs sell distribution.
- Merged trader data with Fear & Greed Index using the date column.
- Filled missing sentiment labels with "unknown".

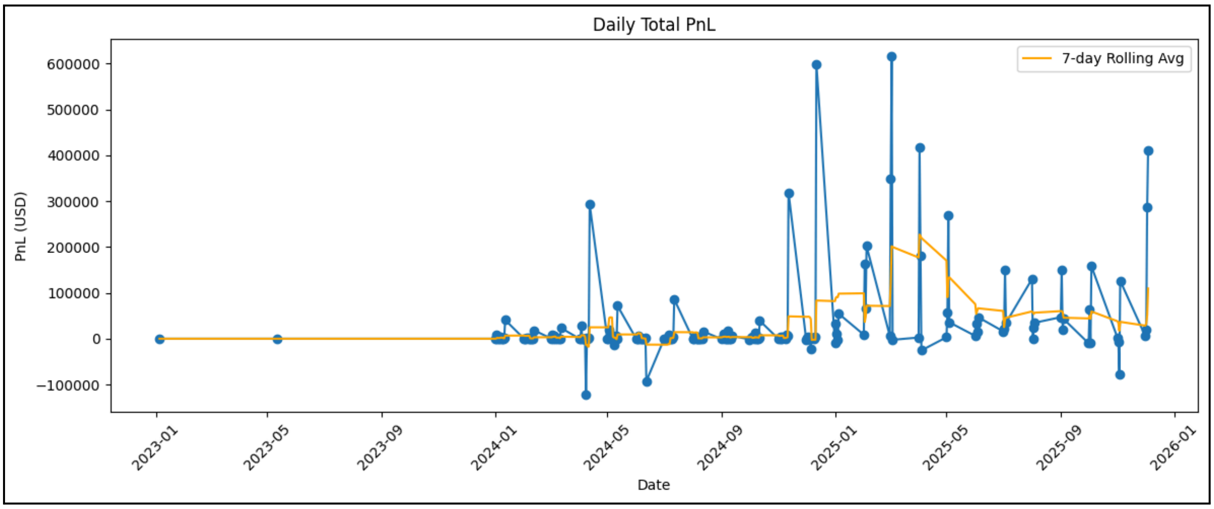
These steps ensured the datasets aligned correctly for performance, sentiment impact, and modeling analysis.

3. Exploratory Data Analysis

3.1 Daily PnL Trends

A time-series plot of aggregated daily PnL shows:

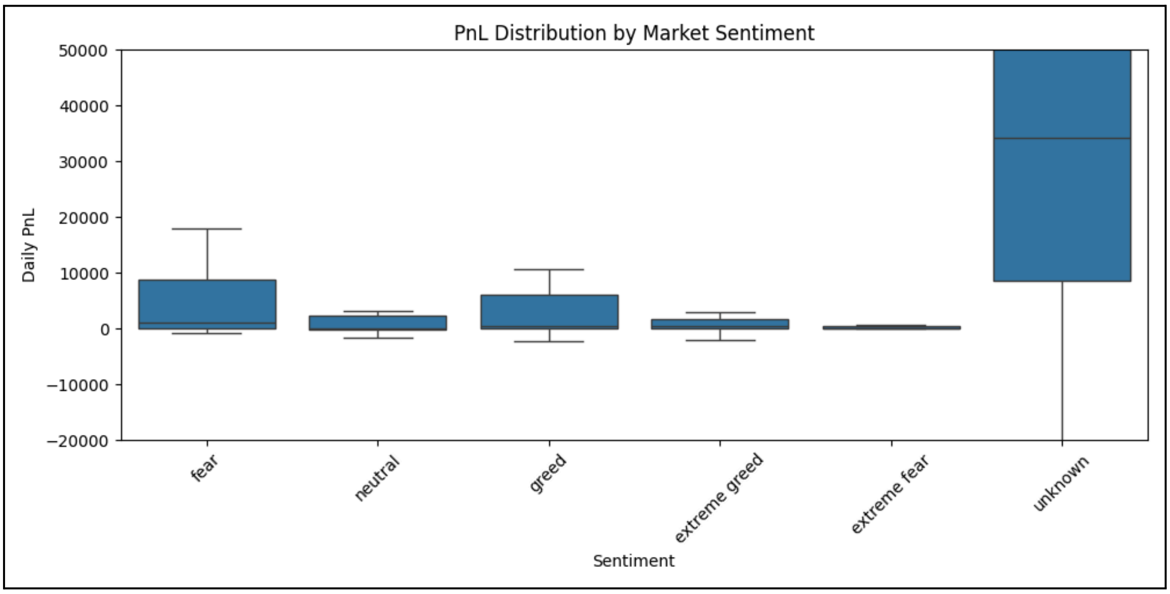
- Trading activity varies significantly day to day.
- Some bursts of high-volume/high-pnl days.
- A 7-day rolling average smooths short-term volatility.



3.2 PnL vs Market Sentiment

Boxplots of PnL grouped by sentiment:

- Greed days had higher volatility in PnL.
- Fear days had lower but steadier PnL.
- Neutral days sat in the middle.
- No sentiment category showed consistently better profitability.



4. Statistical Analysis

A Mann–Whitney U test was used to compare profits under:

- Fear days
- Greed days

Output:

Statistic: 2251.0

p-value: 0.5872

Interpretation:

- Since $p > 0.05$, the difference is not statistically significant.
- Sentiment does not influence trader PnL in the dataset.
- Behavior and volume fluctuations matter more than market emotion cycles.

5. Predictive Modeling

5.1 Target Variable

A day is considered profitable if:

- total_pnl > 0 → profitable = 1
- total_pnl ≤ 0 → profitable = 0

5.2 Features Used

- total_trades
- total_volume_usd
- win_rate

5.3 Model

Random Forest Classifier:

- 300 estimators
- 70/30 train-test split

Model Performance:

- Accuracy: 92.98%
- ROC AUC: 0.8004

Feature Importance:

- win_rate 0.503
- total_trades 0.266
- total_volume_usd 0.230

Interpretation:

- Win rate is the strongest determinant of profitability.

- More trades generally increase profit probability.
- Volume plays a role but is less important than win consistency.

6. Insights

- Sentiment does not impact PnL - traders should not rely on Fear/Greed index to make decisions.
- Trading behavior is predictive - win rate & trade frequency matter more.
- High PnL days align with higher liquidity and trader activity, not sentiment.

8. Conclusion

The data analysis clearly indicates that, in this data set, profitability is driven by trader behavior patterns and not market sentiment.

A machine learning model based on a very simple feature set was able to achieve an **accuracy of 93%**, clearly showing that trader-centric analytics can have a very strong role in understanding and predicting performance.

This project showed the ability to integrate multiple data sources, carry out statistical testing, create predictive models, and generate actionable insights for real-world trading systems.