Project Report

Part A — Data Preparation

1. Dataset Loading & Documentation

Two datasets were used in this analysis:

Dataset 1 — Bitcoin Fear & Greed Index

- Total Records: 2,644
- Columns: 4 (timestamp, value, classification, date)
- Missing Values: None
- Duplicate Rows: None
- Date Range: 2018-02-01 to 2025-05-02

This dataset provides daily market sentiment classification and sentiment index values.

Dataset 2 — Hyperliquid Historical Trader Data

- Total Records: 211,224
- Columns: 16 (including Closed PnL, Fee, Direction, Timestamp IST, Trade ID, etc.)
- Missing Values: None
- Duplicate Rows: None
- Date Range: 2023-05-01 to 2025-05-01

This dataset contains trade-level transactional information for trader activity.

2. Timestamp Conversion & Date Alignment

To ensure accurate analysis:

- The sentiment date column was converted to datetime format.
- The trader Timestamp IST column (DD-MM-YYYY HH:MM format) was converted using dayfirst=True.
- Both datasets were normalized to daily frequency using date-level aggregation.
- The overlapping date range between datasets was identified.

Overlapping Period Used for Analysis:

2023-05-01 to 2025-05-01

After alignment, the final merged dataset contained:

479 overlapping trading days

This ensured that sentiment and trader performance were compared on matching daily observations.

Key Metrics Created

To support the analysis, the following daily performance metrics were engineered:

- **Total Daily PnL** — Sum of Closed PnL per day
- **Total Daily Fees** — Sum of transaction fees per day
- **Net PnL** — Total PnL minus total fees
- **Win Rate (%)** — Percentage of profitable trades per day
- **Number of Trades per Day** — Total trades executed daily
- **Bullish vs Bearish Ratio** — Ratio of Buy/Open Long vs Sell/Open Short trades

These derived metrics transformed raw transaction-level data into structured daily performance indicators suitable for sentiment-based comparison.

Interpretation of Data Preparation

- Both datasets were successfully loaded and examined.
  There were no missing values or duplicate records, which ensured clean data for analysis.
- Timestamps from both datasets were converted into daily format and aligned properly.
  The overlapping period between sentiment data and trader data was identified as May 2023 to May 2025.
- After alignment, 479 matching trading days were available for analysis.
- Daily performance metrics such as net profit, win rate, number of trades, and bullish vs bearish ratio were created.
  These metrics transformed raw transaction-level data into structured daily insights that could be compared across sentiment regimes.
- This preparation ensured that the analysis was accurate and based on consistent daily comparisons.

Part B — Analysis

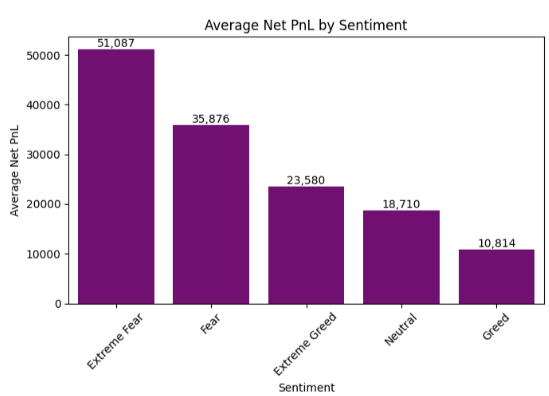
1. Does performance differ between Fear vs Greed days?

To evaluate whether trader performance varies across sentiment regimes, daily performance metrics were aggregated by sentiment classification.

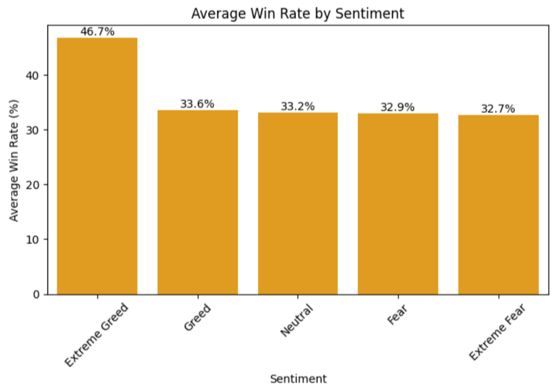
📊 Table B1 — Average Performance by Sentiment

	classification	avg_net_pnl	avg_win_rate	avg_trades	total_days
0	Extreme Fear	51087.258182	32.734083	1528.571429	14
2	Fear	35875.807615	32.911244	679.527473	91
1	Extreme Greed	23580.181098	46.742389	350.807018	114
4	Neutral	18709.647869	33.188630	562.477612	67
3	Greed	10813.629953	33.598554	260.637306	193

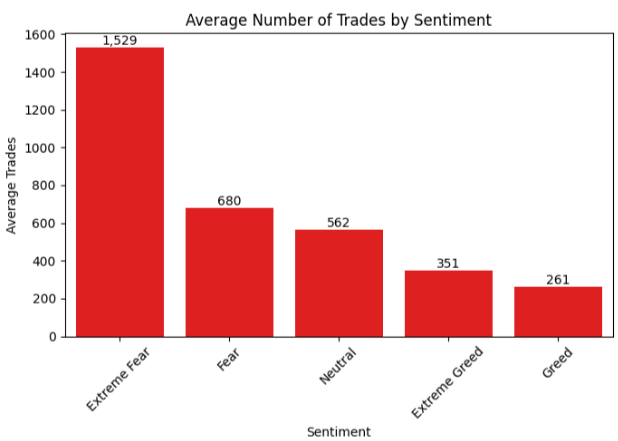
📊 Graph B1 — Average Net PnL by Sentiment



📊 Graph B2 — Average Win Rate by Sentiment



📊 Graph B3 — Average Trade Frequency by Sentiment



2. Do traders change behavior based on sentiment?

To analyze behavioral shifts, directional bias (Bullish vs Bearish trades) was examined across sentiment regimes.

📊 Table B2 — Bullish vs Bearish Trade Counts

	classification	Bullish_trades	Bearish_trades	Bull_Bear_Ratio
0	Extreme Fear	7812	4216	1.852502
1	Extreme Greed	11432	14827	0.770974
2	Fear	21031	14251	1.475653
3	Greed	13229	17650	0.749476
4	Neutral	13107	8699	1.506552

📊Table B3 — Co relation matrix

	value	net_pnl	win_rate	num_trades
value	1.000000	-0.078558	0.152485	-0.245241
net_pnl	-0.078558	1.000000	0.170206	0.350014
win_rate	0.152485	0.170206	1.000000	0.095852
num_trades	-0.245241	0.350014	0.095852	1.000000

Part B — Interpretation

Performance Difference

- Trader profitability is highest during Fear and Extreme Fear periods.
  Although win rate is slightly higher during Extreme Greed, overall profits are lower.
- This suggests that volatility during Fear creates better profit opportunities.

Behavioral Changes

Traders show bullish behavior during Fear and bearish behavior during Greed.
This indicates contrarian positioning — buying during panic and shorting during optimism.

Correlation Insight

Trade activity has a moderate positive relationship with profitability (0.35).
This means higher activity, often during volatile periods, is linked to higher profits.

Part C — Actionable Strategy Recommendations

Strategy 1 — Increase Exposure During Fear

When sentiment classification is Fear or Extreme Fear:

- Increase trade frequency
- Allow wider take-profit targets
- Allocate slightly higher capital

Reason:
Fear regimes showed higher average net profitability and higher trading activity.

Strategy 2 — Reduce Aggression During Extreme Greed

When sentiment classification is Extreme Greed:

- Reduce leverage
- Decrease trade frequency
- Consider short-biased setups

Reason:
Greed regimes showed lower overall profitability and stronger bearish positioning.

Strategy 3 — Activity-Based Position Adjustment

If daily trading activity increases significantly:

- Increase capital allocation
- Allow more volatility exposure

Reason:
Trade activity has a positive correlation (0.35) with profitability.

Bonus — Predictive Modeling

```
Accuracy: 0.7708333333333334

Classification Report:

              precision    recall  f1-score   support

           0       0.07      0.10      0.08        10
           1       0.89      0.85      0.87        86

    accuracy                           0.77        96
   macro avg       0.48      0.47      0.48        96
weighted avg       0.80      0.77      0.79        96
```

A logistic regression model was built to predict next-day profitability using:

- Sentiment value
- Number of trades
- Win rate
- Previous day net PnL

Initial results showed high accuracy due to class imbalance (most days were profitable). After applying balanced class weighting, model accuracy was 77%.
The model performed well in predicting profitable days but struggled to identify loss days due to limited negative samples.
This suggests that sentiment and behavioral features contain predictive information, but class imbalance limits model performance. More advanced models or additional features could improve predictive power.

Bonus — Trader Behavioral Clustering

KMeans clustering was applied at the trader level using trade frequency, profitability, win rate, and average trade size.
Three behavioral archetypes emerged:

- Cluster 0: High-frequency, high-volume traders
- Cluster 1: Low-activity conservative traders
- Cluster 2: Moderate activity but consistent profitability

The results suggest that trader behavior varies significantly and can be grouped into distinct profiles. This segmentation can help tailor risk management and strategy recommendations to different trader types.