Data Science Report – Web3 Trading: Sentiment vs Trader Behavior

Author: Banoth Anand

Date: October 23, 2025

1 Executive Summary

Objective: Measure how trader behavior (PnL, win-rate, volume, fees) aligns or diverges from Bitcoin sentiment (Fear vs Greed).

Approach: Clean sentiment and trade data, engineer features, aggregate daily KPIs, and join with daily sentiment. Visualize relationships and regime effects.

Headlines (fill after running):

- Win-rate tends to be [higher/lower/flat] during [Greed/Neutral/Fear].
- Daily PnL shows [positive/negative/weak] association with the Fear & Greed index.
- Volume is [pro/contra]-cyclical relative to sentiment levels.
- Fees exhibit [expected drag/no clear link] to daily PnL.

2 Data

Sentiment (from `notebook_1.ipynb`): `csv_files/cleaned_sentiment_daily.csv`

- Columns: `date`, `sentiment_value` (0–100), `sentiment_class` (Extreme Fear → Extreme Greed)
- Derived: 7D/30D rolling stats, weekly summaries

Trades (from `notebook_2.ipynb`): `csv_files/cleaned_trades.csv`

- Standardized columns include: `account`, `symbol`, `execution_price`, `size_tokens`, `size_usd`, `side`, `event`, `closed_pnl`, `fee`, `fee_rate`, `trade_value`, `time`, `date`, `is_win`
- Handling: chunked read, numeric coercions, IST→UTC time normalization

3 Methodology

Cleaning & Feature Engineering:

- Sentiment: unify to daily `date`; compute rolling means/volatility and weekly aggregates.
- Trades: compute `trade_value`, `fee_rate = fee / trade_value`; derive `is_win` from `closed_pnl`.

Aggregation:

- By `date`: `trades`, `win_rate`, `pnl_sum`, `pnl_mean`, `volume_tokens` (or USD), `fees_sum`, `avg_fee_rate`.
- By `date, symbol`: same metrics for coin-level analysis.

Join: Merge daily KPIs with sentiment on `date` → `csv_files/kpi_with_sentiment.csv`.

Visualization (saved to `outputs/`):

- `fg_daily_with_30d_mean.png`
- `fg_regime_counts.png`
- `fg_weekly_range_mean.png`
- `pnl_vs_sentiment.png`
- `winrate_by_sentiment_regime.png`
- `fees_vs_pnl.png`
- `volume_vs_sentiment.png`

4 Results (fill after execution)

- **Regime Distribution:** [counts/percentages from `sentiment_regime_counts.csv`].
- **PnL vs Sentiment:** [direction/strength and notable periods].
- **Win-rate by Regime:** [which regimes stand out and by how much].
- **Fees vs PnL:** [relationship observation].
- **Volume vs Sentiment:** [relationship observation].
- **Symbol-level Highlights:** [from `kpi_by_symbol_date.csv` if relevant].

5 Insights & Recommendations

- **Regime-aware risk:** Adjust sizing based on `sentiment_value` or `sentiment_class` thresholds if results show consistent edges.
- **Fee discipline:** Avoid days or symbols where fee rate historically erodes expectancy.
- **Monitoring:** Track rolling sentiment changes; consider transitions (e.g., Fear→Neutral) as potential signals.

6 Limitations & Next Steps

- Missing/limited leverage fields in sample; proxy signals used instead.
- Potential selection bias in account/symbol coverage.
- Next steps: add time-of-day/weekday features; run walk-forward validation for regime-conditioned strategies.

7 Code Link

- Colab Notebook Link