Data Science Report: Sentiment-Aware Trading Behavior Analysis

1. Introduction

This report explores the relationship between trader behavior and market sentiment—specifically, how human traders react under states of fear and greed in the Bitcoin market. The investigation uses two real-world datasets: the Fear & Greed Index (representing market sentiment) and historical trader execution logs from Hyperliquid. Understanding how sentiment influences decision-making in financial markets is critical for designing more effective trading strategies, risk management systems, and even automated agents. The purpose of this analysis is to identify correlations or divergences between sentiment classification and actual trading behavior such as volume, profitability, and risk.

2. Dataset Description

The project is based on two datasets:

1. Fear & Greed Index (Sentiment Data):
   - Contains date and classification fields, where classification is either Fear or Greed.
   - This label represents the daily dominant market emotion.

2. Hyperliquid Trader Data (Execution Logs):
   - Fields include timestamp, account, symbol, side, execution price, size, closed_pnl, etc.
   - Represents raw trade-level data for thousands of unique executions.
   - Important numeric fields used: size (volume), closed_pnl (profitability), and timestamp.

Both datasets were cleaned, column names normalized, and merged on a common date field to allow comparative analysis.

3. Methodology

The analysis pipeline follows these steps:

1. Preprocessing:
   - All column headers were stripped of whitespace, converted to lowercase, and formatted with underscores.
   - Date columns were parsed using pandas' to_datetime and reduced to date only.

2. Merging:
   - The trader dataset and sentiment index were merged on date, using a left join to retain all trading events.

3. Exploratory Data Analysis (EDA):
   - A series of plots and statistical summaries were generated:
     - Trade count by sentiment
     - Average trade volume by sentiment
     - Profitability distribution via boxplots
     - Descriptive statistics for volume and profit grouped by sentiment

4. Handling Missing Data:
   - Many rows had missing closed_pnl or size values. Instead of excluding these, fallback statistics (.describe()) were printed if visualizations failed due to NaNs.

5. Visualization:
   - Plots were created using Seaborn and Matplotlib and saved as .png files in an output directory.

4. Key Findings

a) Trade Activity by Sentiment

- During Greed periods, the total number of trades tended to increase compared to Fear periods.
- This could reflect increased trader confidence and a tendency to take more positions when sentiment is positive.

b) Average Trade Volume

- Greed days generally showed higher average trade volume (size) per trade.
- This suggests not only are traders more active in greedy markets, but they are also more willing to commit higher volume.

c) Profitability Distribution

- The closed_pnl distribution during Greed was more positively skewed, with higher median and fewer extreme losses.
- Fear periods exhibited higher variance and more loss-heavy outliers.

d) Fallback Stats Support Visuals

- In cases where plotting was not possible (e.g., empty or missing values), the .describe() summaries confirmed trends observed visually.

5. Limitations

While the trends are suggestive, some limitations are noted:

- Missing values in critical fields like closed_pnl and size reduced the number of usable rows.
- No data on leverage was available, despite it being mentioned in the problem statement.

- The sentiment dataset is a single-label classification per day, which might not capture intra-day changes in market mood.

6. Conclusion

This exploratory analysis demonstrates that trader behavior is indeed influenced by market sentiment:
- Traders are more active, confident, and profitable during Greed conditions.
- Fear states lead to lower volumes and wider PnL variability, indicating more cautious and erratic behavior.

These insights can be used in developing sentiment-aware risk models, or as features in machine learning-based trading agents. Future work may expand on this by incorporating intra-day sentiment or extending the analysis to additional instruments beyond Bitcoin.

7. Recommendations

- Data Enrichment: Add real-time sentiment feeds and intraday metrics for finer resolution.
- Modeling: Apply clustering or predictive ML models to detect behavior anomalies.
- Backtesting: Use sentiment conditions as filters in strategy backtesting to improve expected returns.