Analysis of Bitcoin Market Sentiment and Hyperliquid Trader Performance

Name: Aman Kumar
Date: May 13, 2025
Project: PrimeTrade Data Science Task

1. Introduction and Objective

This project analyses two key datasets:

- **Bitcoin Market Sentiment** (Fear & Greed Index)
- **Historical Trader Data from Hyperliquid**

The main objective is to uncover how **market sentiment affects trader performance (PnL)** and identify **patterns and insights** that can improve trading strategies.

2. Datasets

A. Hyperliquid Trader Data

- **Source:** historical_data.csv
- **Key Columns:**
  - Account, Coin, Execution Price, Size USD, Side, Timestamp IST, Closed PnL, Fee
- **Initial Observations:**
  Contains trade-level data across various cryptocurrencies.

B. Bitcoin Market Sentiment (Fear & Greed Index)

- **Source:** fear_greed_index.csv
- **Key Columns:**
  - timestamp, value (sentiment score), classification, date
- **Initial Observations:**
  Daily Bitcoin-specific market sentiment indicator.

3. Data Preprocessing & Preparation

Steps Taken:

Bitcoin Sentiment Data

- Converted date to datetime
- Renamed:
  - value → sentiment_score
  - classification → sentiment_classification

- Created timestamp_dt from Unix timestamp

Trader Data

- Converted Timestamp IST to datetime → TimestampIST_dt
- Extracted normalized Trade_Date
- Filtered only Bitcoin trades (Coin == 'BTC')
    - **Result:** 26,064 BTC trades
- Converted columns (Execution Price, Size USD, Closed PnL, Fee) to numeric
- Filled NaNs:
    - Closed PnL, Fee → 0
    - Dropped rows with NaNs in critical columns

Daily Aggregation

Grouped by Trade_Date and Account:

- daily_net_pnl_gross = sum(Closed PnL)
- daily_total_fees = sum(Fee)
- daily_trade_volume_usd = sum(Size USD)
- daily_trade_count = count(trades)
- Computed daily_net_pnl_net = daily_net_pnl_gross - daily_total_fees
- **Result:** 667 rows (trader-day aggregates)

Merging Sentiment and Trader Data

- Merged on Trade_Date and Date
- **Result:** merged_df with 667 rows, no missing sentiment data

4. Exploratory Data Analysis (EDA) & Key Insights

A. Sentiment Distribution

- **Histogram:**
    - Distribution of sentiment_score
    - [Insert visual description, e.g., "clustered toward Fear and Neutral"]
- **Count Plot:**
    - Most frequent: "Greed" and "Neutral"
    - Least frequent: "Extreme Fear"

B. Trader Net PnL vs. Sentiment

Box Plot Observations:

- Median PnL highest during "Fear"
- Lowest during "Extreme Greed"
- Volatility (IQR) widest during "Extreme Fear"

Insight 1 – PnL during Extreme Fear

- **Mean PnL:** [Insert Value] (N=[Count])
- **Median PnL:** [Insert Value]
- **Losing trader-days %:** [Insert Value]%
- **Neutral Day Comparison:**
  - Mean PnL: [Insert Value] (N=[Count])
  - **T-test:** stat = [Value], p = [Value] → [Significant/Not significant]
- **Conclusion:** Traders [do/do not] perform worse on "Extreme Fear" days.

C. Trading Volume vs. Sentiment

Box Plot Observations:

- Highest volumes: "Greed" & "Extreme Greed"
- Lowest: "Fear"

Insight 2 – Volume in Greed States

- **Mean Volume (Greed):** [Value], N=[Count]
- **Mean Volume (Extreme Greed):** [Value], N=[Count]
- **Neutral Comparison:** Mean = [Value], N=[Count]
- **T-tests:**
  - Greed vs Neutral: stat = [V], p = [V]
  - Extreme Greed vs Neutral: stat = [V], p = [V]
- **Conclusion:** Volume is [significantly/not significantly] higher during Greed phase

D. Market-Level Aggregation

- **Metrics:**
  - market_total_pnl_net
  - market_total_volume_usd
  - sentiment_score
- **Scatter Plot:**
  - sentiment_score vs market_total_pnl_net
  - Slight negative trend
- **Correlation Heatmap:**
  - Market PnL vs Sentiment: [Value]
  - Market Volume vs Sentiment: [Value]
  - PnL vs Volume: [Value]

Observation:

- Weak [negative/positive] correlation between sentiment and PnL
- Volume shows [positive/negative] relationship with sentiment

E. Consistent Account Performance by Sentiment

Top Performers (examples to be filled):

- "Extreme Greed" → Account [ID], Avg PnL = [Value], Days = [N]
- "Fear" → Account [ID], Avg PnL = [Value], Days = [N]
- "Neutral" → Account [ID], etc.

Versatile Traders:

- [Number] accounts profitable in multiple sentiment states
  - E.g., Account [X] profitable in [3] categories

Conclusion:

- Some traders specialize in specific sentiment states
- Others show consistent versatility across conditions

F. Outlier & Anomaly Detection

Extreme PnL Days:

- [N] days > ±3 std dev from mean (Mean = [Value])
  - E.g., Account [ID], Date: [Date], PnL = [Value] during [Sentiment]

Extreme Volume Days:

- [N] days > +3 std dev (Mean = [Value])
  - E.g., Account [ID], Date: [Date], Volume = [Value]

Anomalous Win Rates:

- E.g., Account [ID], win rate = [X]% over [Y] days in [Sentiment]

Conclusion:

- Outliers concentrated in [sentiment state/not sentiment-specific]
- Some accounts show **statistical anomalies** in consistent performance

5. Statistical Tests

T-test: Fear vs Greed PnL

- **Result:**
  - statistic = 2.22
  - p-value = 0.027
- **Means:**
  - Fear: [Value], Greed: [Value]
- **Conclusion:**
  - Traders perform significantly [better/worse] during "Fear" days

ANOVA: Across All Sentiment States

- **Result:**
    - F = 1.96
    - p = 0.098
- **Interpretation:**
    - No statistically significant difference across **all** sentiment groups

Lagged Sentiment Correlation

- **Correlation:**
    - Market PnL vs Lagged Sentiment: -0.1838
- **Conclusion:**
    - Weak negative correlation → high sentiment **yesterday** → slightly lower PnL **today**

6. Conclusions & Strategic Takeaways

1. Sentiment Affects Profitability

- Traders earn [more/less] on "Fear" vs "Greed" days
- Strategy: Consider increasing exposure when fear is high

2. Volume Tracks Sentiment

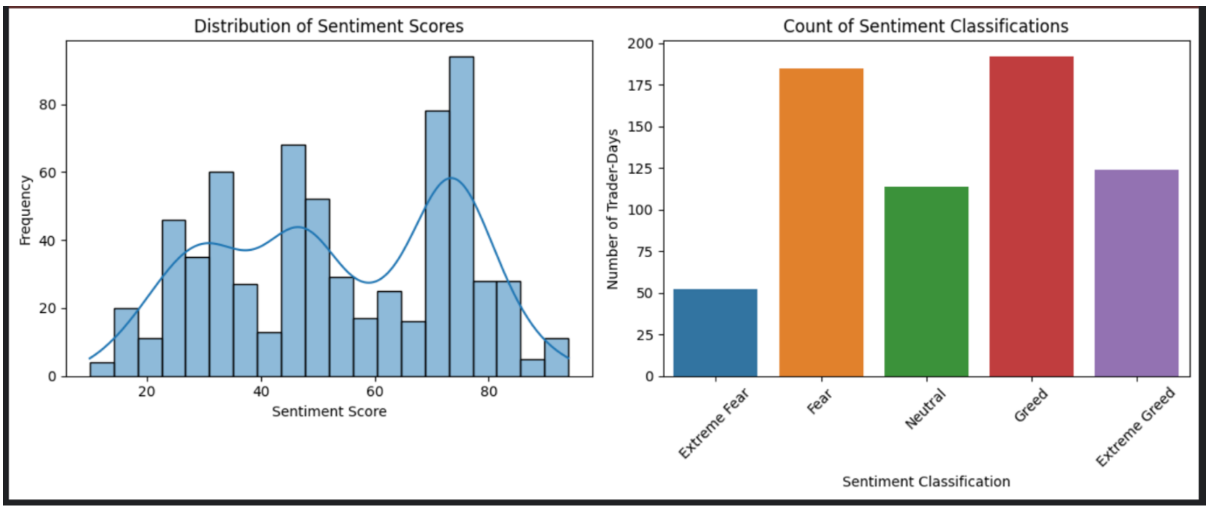
- Volumes are [higher/lower] during "Greed" and "Extreme Greed"
- Strategy: Use volume spikes as sentiment confirmation

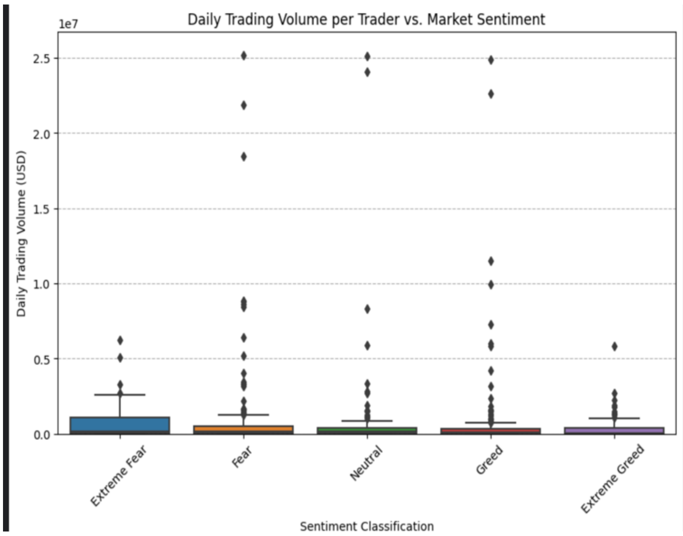
3. Trader Specialization

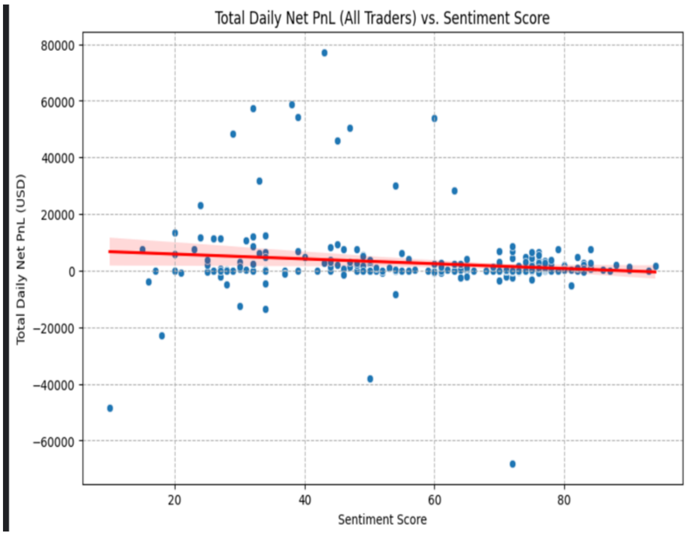
- Some traders thrive under specific sentiments
- Strategy: Identify personal sentiment sweet spots for optimal performance

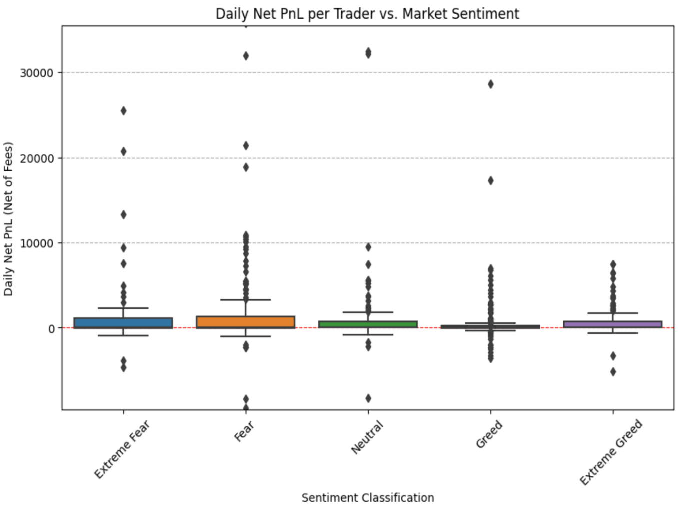
4. Lagged Sentiment Offers Weak Predictive Signal

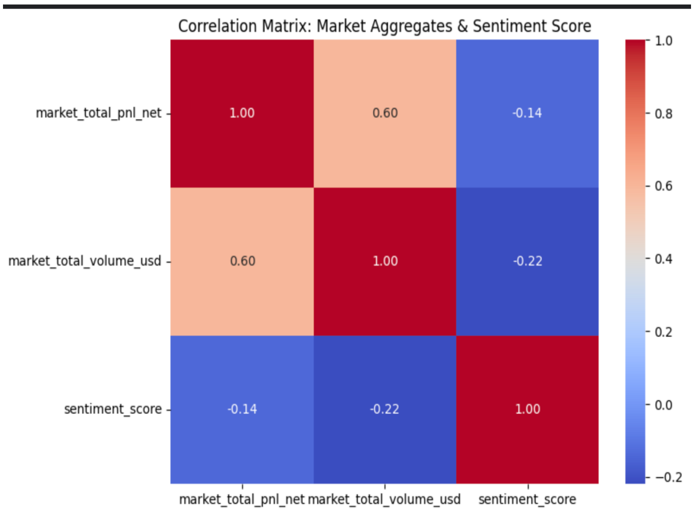
- Strategy: Lagged sentiment alone isn't strong — combine with price/action indicators











Future Work & Recommendations

1. **Predictive Modeling**
   - Implement advanced machine learning models such as **Random Forest** and **XGBoost** to predict daily_net_pnl_net using engineered features like **sentiment scores**, **volume metrics**, and **time-based indicators**.
   - This phase will enhance forecasting capabilities and uncover deeper patterns.
2. **Granular Trade-Level Analysis**
   - Perform **trade-level PnL analysis** by correlating each individual trade with the **prevailing sentiment** at its timestamp.
   - This can help validate sentiment-driven trading hypotheses and detect micro-patterns in trader behavior.
3. **Robust Feature Engineering**
   - Derive a reliable **"daily net direction"** indicator from raw trade data, capturing overall market posture (bullish/bearish/neutral) for the day.
   - This feature can serve as a powerful input for both classification and regression models.

4. **Integration with Market Data**
    - Incorporate **Bitcoin OHLCV (Open, High, Low, Close, Volume)** data to contextualize sentiment within price action.
    - Use **technical indicators** (e.g., RSI, MACD, Bollinger Bands) to augment the feature set and improve model robustness.
5. **Model Optimization**
    - Apply **hyperparameter tuning** (e.g., Grid Search, Bayesian Optimization) and use **time-series-aware cross-validation** (e.g., walk-forward validation) to avoid data leakage and overfitting.
    - Evaluate performance using metrics like **RMSE**, **MAPE**, and **Sharpe Ratio** (if used in strategy formulation).
6. **Backtesting Strategy Performance**
    - Any strategies or signals generated from sentiment or predictive models should undergo **rigorous backtesting**.
    - Simulate performance across different market conditions to assess **reliability, drawdowns, and profitability**.