PROJECT REPORT
Analyzing Trading Behaviour vs. Market Sentiment

Name: N Hari Sai Vignesh
E-mail: nhs.vignesh@gmail.com

1. Objective

To analyze how trading behaviour (profitability, risk, volume) aligns or diverges from overall market sentiment (Fear & Greed Index) and to identify hidden trends or signals that could influence smarter trading strategies.

Executive Findings: In addition to successfully uncovering a significant "hidden trend," the analysis refuted a widely held belief.

- There was **no statistically significant** correlation between trading **profitability** (PnL) and market sentiment. Both concurrent (same-day) and lagged (previous-day) sentiment are affected by this.
- In trading volume, we found a **statistically significant predictive signal**. The overall **trading volume** of today is predictably impacted by the sentiment of yesterday.
- Only the **extremes** are responsible for this effect. Trading volume is substantially **higher** on the day after **Extreme Fear** and significantly **lower** on the day after **Extreme Greed.**

2. Methodology

The analysis was conducted in a 5-step "drill-down" process using the given two datasets.

1. **Data Preparation:** The merged_daily_data.csv we used was already clean. To create it, two separate datasets were prepared:
   - **Historical Data:** This raw data (likely at a per-trade level) would be cleaned by converting timestamps to dates. It would then be aggregated by date to calculate daily totals (total_volume_usd, total_pnl_usd) and statistics (avg_pnl_usd, pnl_volatility).
   - **Sentiment Data:** This daily-level data would be cleaned by checking for any missing value or classification entries and ensuring the dates were in a compatible format.

   These two cleaned, daily-level datasets were then merged on the date column to create the single file we analyzed.

2. **Exploratory Data Analysis (EDA):** First, we used time-series plots, boxplots, and heatmaps to visualize the combined data. This was done in order to visually identify any possible (yet unproven) trends or correlations between trading metrics such as volume and PnL and sentiment classifications.

3. **Test 1: Concurrent ANOVA:** To determine whether today's sentiment has a significant impact on today's trading activity, we performed an ANOVA statistical test. The findings demonstrated that there is no statistical correlation between PnL, volume, or volatility, indicating that the data does not support a same-day strategy.

4. **Test 2: Lagged ANOVA:** In order to check for a delayed effect, we then developed a new "lagged" feature for yesterday's sentiment (classification_lag_1). According to the ANOVA test we then conducted, sentiment from yesterday has a statistically significant correlation with today's total_volume_usd, but it still does not predict PnL.

5. **Test 3: Tukey's HSD Post-Hoc Test:** We performed a Tukey's HSD post-hoc test to determine the precise source of the volume signal. The exact relationship was identified by this test: trading volume is significantly lower on days after Extreme Greed and significantly higher on days after Extreme Fear.

3. Key Findings

Finding 1: Sentiment Fails to Predict Profitability

A primary goal was to see if sentiment could predict PnL. Our analysis proves it cannot.

- **Concurrent Test:** The ANOVA test for *today's* sentiment vs. *today's* total_pnl_usd yielded a p-value of 0.2978. This is not significant.
- **Lagged Test:** The test for *yesterday's* sentiment vs. *today's* total_pnl_usd yielded a p-value of 0.4204. This is also not significant.

Based on this data, a trader cannot reliably predict profit or loss by looking at the current or previous day's market sentiment.

Finding 2: A Significant "Hidden Signal" in Volume

The analysis of trading *volume* (activity) was far more successful.

- **Concurrent Test:** The test for *today's* sentiment vs. *today's* total_volume_usd was not significant (p-value: 0.0985).
- **Lagged Test:** The test for *yesterday's* sentiment vs. *today's* total_volume_usd was STATISTICALLY SIGNIFICANT (p-value: **0.0173**).

We proved that a 1-day lag exists. Yesterday's sentiment is a genuine predictor of today's market participation.

Finding 3: The Signal is Driven *Only* by Extremes

The Tukey's HSD test isolated the precise nature of the signal.

- The test compared all possible pairs of the 5 sentiment classifications.
- The only pair to return a significant result (reject = True) was Extreme Fear vs. Extreme Greed (p-value: 0.0217).

- **The Effect:** On average, a day following Extreme Greed had $3.77 million less in trading volume than a day following Extreme Fear.

4. Actionable Insights

This analysis provides two clear, data-driven insights to inform trading strategies:

1. Do not use the Fear & Greed Index as a direct signal for profitability. Strategies such as "buy on fear, sell on greed" are not supported by this data and do not show any statistical edge.
2. The real value of the index is as a 1-day-ahead predictor of market *activity*.
    - After an Extreme Fear day, expect high volume and high liquidity, which could be a good time to place big orders since the high activity can reduce slippage.
    - After an Extreme Greed day, expect low volume and low liquidity. This signals market hesitation. On these days, one might consider reducing trade sizes or using algorithms designed for lower-liquidity environments.

GitHub:
https://github.com/ComradeV7/ds_N-Hari-Sai-Vignesh

Google Drive:
https://drive.google.com/drive/folders/11Q2v6WW7Pd4f5idD7Kvq60s8FcyVja-e?usp=drive_link