Project Report

1. Project Overview

This project analyzes the relationship between trading performance and market sentiment. Using historical trading data alongside a sentiment indicator (Fear & Greed Index), the objective is to explore whether sentiment trends have a measurable impact on trader performance and come out with good trading strategies. The notebook combines data exploration, visualization, and analysis to uncover potential insights for trading strategies.

2. Data Sources

1. **Historical Trading Data (historical_data.csv)**
    - Contains trader-level information including timestamps, positions, and realized profit and loss (PnL).
    - Used to compute daily and cumulative PnL trends.
2. **Sentiment Data (fear_greed_index.csv)**
    - Provides market sentiment values based on the Fear & Greed Index.
    - Used to measure investor psychology, ranging from extreme fear to extreme greed.
    - Aggregated at the daily level for comparison with trading outcomes.

3. Methodology

- **Data Loading & Cleaning**
    - Imported datasets using pandas.
    - Verified structure with .head() and summary functions.
    - Ensured date fields were properly converted to datetime.
- **Aggregation & Transformation**
    - Trading data was grouped by date to calculate **daily PnL** and **cumulative PnL**.
    - Sentiment data was grouped by date to calculate **daily average sentiment** and rolling averages for smoothing trends.

- **Visualization** (Used matplotlib and seaborn to plot):
  - Daily and cumulative PnL trends.
  - Sentiment index over time.
  - Comparisons between sentiment and trader performance.

Analysis & Insights

1. Correlation between Trader PnL and Market Sentiment

- The correlation coefficient between **daily PnL** and **sentiment value** is **-0.0826**.
- This is very close to zero, indicating **no significant linear relationship** between daily trading profits and the sentiment index.
- In other words, sentiment alone does not directly predict daily PnL.

2. Cumulative PnL Over Time

- The cumulative PnL shows a **sharp exponential rise** starting late 2024 into 2025.
- This suggests that trading performance was relatively flat for a long time but then captured a strong upward trend, possibly due to a change in strategy, favorable market conditions, or concentration in a few high-performing assets.

3. Market Sentiment Trends

- Sentiment is highly **cyclical**, fluctuating between fear (low values) and greed (high values).
- There is no steady upward or downward trend, which matches real-world investor psychology cycles.
- Comparing this to PnL suggests that **PnL gains were not strongly tied to sentiment shifts**, reinforcing the weak correlation result.

4. Distribution of PnL by Side (BUY vs SELL)

- Both **BUY** and **SELL** trades exhibit a wide spread of outcomes, including large positive and negative PnLs.
- The distributions are fairly similar, suggesting **trading success was not biased towards one side**.
- Extreme outliers (both losses and gains) exist, which could drive volatility in performance.

5. Performance Metrics

- **Total PnL:** ~10.3 million, showing strong absolute performance.
- **Average Daily PnL:** ~49, which is relatively small compared to the cumulative gains, suggesting many small trades plus occasional large winners.
- **Win Rate:** 41% - less than half of trades are profitable, but this is common in trading when **risk–reward is asymmetric**.
- **Max Drawdown:** -419k, indicating exposure to significant downside risk.
- **Sharpe Ratio:** 0.30 - below the standard benchmark of 1.0 for "good risk-adjusted returns." This means returns were positive but highly volatile.

6. Top 10 Coins by Total PnL

- PnL is concentrated in a few coins: @107, HYPE, SOL, ETH, and BTC.
- This indicates **profit concentration**: a majority of gains were driven by a handful of assets rather than evenly distributed.
- Suggests that asset selection (or luck in timing these assets) played a huge role in performance.

7. PnL by Day of Week

- Highest PnL occurs on **Monday and Tuesday**, with Friday also being strong.
- Lowest PnL occurs on **Sunday**, possibly due to lower liquidity or weaker market moves on weekends.

- This suggests that **weekday trading was more profitable than weekend trading**, a useful timing insight.

8. PnL by Hour of Day

- Strong gains around **11 AM and 12 PM**, as well as **evening hours (7 PM – 9 PM)**.
- Lower PnL during early morning and late night.
- Suggests there are **optimal trading windows** in the day when volatility and opportunity are higher.

9. Trader Performance vs Market Sentiment (Dual Axis)

- The dual-axis chart shows **PnL rising sharply** in early 2025, while sentiment was volatile and did not track PnL closely.
- This confirms again: **PnL performance was strategy/market-driven rather than sentiment-driven.**
- However, certain spikes in PnL do align with sharp sentiment moves, indicating **opportunistic gains in extreme sentiment phases**.

Final Conclusion

- **Sentiment and PnL are not strongly correlated** - traders did not systematically profit more in “fear” or “greed” conditions.
- **Performance was highly concentrated** in a few coins and specific time windows (weekday mornings and evenings).
- Despite a relatively low win rate (41%) and high volatility (low Sharpe ratio), the strategy delivered **large absolute profits** due to outsized gains on certain trades.
- Risk management remains a challenge, as shown by significant drawdowns.
- Future improvements could focus on:
    - Reducing downside volatility (risk-adjusted performance).
    - Diversifying gains across more assets.
    - Testing predictive power of **extreme sentiment levels** rather than average daily sentiment.

PREDICTION MODEL RESULTS

Accuracy = **71%** (68 correct out of 96).

Performance on Class 1 (Majority class)

- **Precision (0.90):** When the model predicts class 1, it's correct 90% of the time.
- **Recall (0.75):** It identifies 75% of actual class 1 cases.
- **F1-score (0.82):** Balanced performance on class 1.

Performance on Class 0 (Minority class)

- **Precision (0.16):** When the model predicts class 0, it's only correct 16% of the time (poor).
- **Recall (0.36):** Out of 11 true class 0 samples, it only catches 4.
- **F1-score (0.22):** Very low, showing weak performance on minority class.

The CatBoost model is **strong for the profit detection** but **weak for the loss detection** due to imbalance. It's useful if your priority is detecting profits correctly, but risk management is an issue for this model.