Final Report: Analyzing Trader Performance vs. Market Sentiment

Author: Rana Talukdar **Date:** August 10, 2025

---

1. Executive Summary

This project aimed to uncover the relationship between trader performance and market sentiment using historical trade data and the Fear & Greed Index. While initial analysis suggested that trading during periods of "Extreme Greed" was highly profitable, a deeper investigation using advanced machine learning models.

The true driver of profitability was not market sentiment, which had less than 1% predictive power. Instead, the success of the trading strategy was almost entirely dependent on a **hidden pattern**: executing a **Sell action** on a **single, specific asset (@107)**.

A predictive model was successfully built with **95% accuracy** to identify winning trades. Based on its analysis, a clear, data-driven strategic framework was developed to capitalize on this core profitable action while avoiding identified loss-making scenarios.

---

2. Project Objective

The primary goal was to explore the relationship between trader performance and market sentiment, uncover hidden patterns, and deliver insights that can drive smarter trading strategies.

---

3. Data Preparation & Cleaning

The first step was to prepare the data for analysis. This involved several key changes to the datasets.

¥ **Data Loading:** Two separate CSV files were used: one containing historical trader data and one containing the daily Fear & Greed sentiment index.

¥ **Data Merging:** To analyze both datasets together, they were merged into a single table. The common link between them was the **date**. We matched the sentiment data for a specific date to all the trades that occurred on that same date.

¥ **Feature Engineering:** To build a predictive model, we needed a clear target. We created a new column called **Win**. This column was given a value of **1** if a trade's Closed PnL (Profit and Loss) was greater than zero (a win), and **0** otherwise (a loss or breakeven). This simple change turned our complex data into a straightforward question for a model to answer: "Will this trade be a win or a loss?"

---

4. Exploratory Data Analysis (EDA)

With the data prepared, we performed an initial exploration to look for obvious trends.

- **Initial Hypothesis:** We created a "Fear vs. Greed" summary table. This table showed a surprising pattern: trades made during **'Extreme Greed'** had a much higher average profit and win rate than trades made during 'Extreme Fear'. This led to our initial hypothesis that 'Extreme Greed' was a highly profitable condition to trade in.

- **Deeper Look with Boxplots:** A boxplot visualization of the PnL for each sentiment category revealed that while the *average* profit was high during 'Greed', the *median* profit (the middle-of-the-road trade) was close to zero. This was our first clue that a few very large winning trades were skewing the results, and the relationship was more complex than it seemed.

---

5. Predictive Modeling: Finding the Truth

The goal of this phase was to build a "brain" that could learn the patterns leading to a winning trade. We tested several types of machine learning models to see which one could best understand the data.

Why We Chose XGBoost (And Why Others Performed Poorly)

We trained a variety of models, and the results were very clear. The models fell into two distinct camps:

- **The Losers: Distance-Based Models (like K-Nearest Neighbors & SVM)**
  - **How they work:** These models make decisions based on the "distance" between data points.
  - **Why they failed:** They are extremely sensitive to the scale of the data. Our Size USD column had numbers in the thousands, while other features were small numbers (like 1s and 0s). These models were tricked into thinking the Size USD was thousands of times more important than anything else, simply because the number was bigger. This completely broke their ability to find real patterns, resulting in terrible performance.
- **The Winners: Tree-Based Models (like Random Forest & XGBoost)**
  - **How they work:** These models work like a series of "if/else" questions. For example, "IF the coin is @107 AND the action is Sell, THEN predict a win."
  - **Why they succeeded:** They are not sensitive to the scale of the data. More importantly, they are experts at finding complex, **non-linear relationships** and **feature interactions**—exactly the kind of hidden patterns we were looking for.

XGBoost was chosen as the champion because it is an advanced version of a tree-based model. It builds its trees sequentially, where each new tree learns from the mistakes of the previous one. This "boosting" process often makes it the most accurate and powerful model for complex, tabular data like ours. We then **fine-tuned** XGBoost to optimize its settings, achieving a final **F1-Score of 0.95**, meaning it was highly accurate.

---

6. Uncovering the Hidden Pattern

After building our highly accurate XGBoost model, we performed the most critical step: we asked the model what it learned. We analyzed its **feature importances** to see which factors it relied on most when making its predictions. As the significance of the columns was not given hence the column **values** was taken as the measurement of the market sentiment with respect to **classification**.

The result was the **non-obvious conclusion** of this project:

- Market sentiment (value) had a negligible importance of **less than 1%**.
- The model's predictions were overwhelmingly driven by the trade **Direction** (e.g., 'Open Long', 'Sell'), which accounted for over **70%** of the predictive power.

This debunked our initial hypothesis. The relationship between sentiment and profit was a distraction; the real story was hidden in the trade actions themselves.

---

7. Strategic Insights & Recommendations

By analyzing the trades the model found most important, we isolated the precise source of profitability.

- **The "Golden Rule":** The data overwhelmingly supports one core strategy: **The highest probability of success lies in executing a Sell order on the @107 asset when the market sentiment is in 'Extreme Greed'.** This single, specific pattern was the engine of profitability in the dataset.
- **The "Red Flags":** The data also clearly showed that trading during **'Extreme Fear'** was a consistent loss-maker and should be avoided.

---

8. Conclusion

This project successfully moved from a broad and misleading initial hypothesis to a specific, data-driven, and actionable strategic insight. It demonstrated that by using the right analytical tools, it's possible to look past obvious but weak signals (like market sentiment) and uncover the hidden patterns that truly drive performance. The final deliverable is not just an analysis but a clear framework for making smarter trading decisions based on historical evidence.

Appendix: Note on Missing Data

It is important to note that the leverage column, while mentioned as an example in the assignment description, was **not present** in the provided historical data file. Consequently, all analyses related to leverage could not be performed. This report represents the most complete analysis possible with the available data.