Relationship Between Trader Performance and Market Sentiment

Name: Sanuja Chakraborty
Repository: https://github.com/SanujaChakraborty/trader_sentiment_analysis

Objective

The main goal of this task was to understand how **market sentiment** affects **trader performance**.
I wanted to see if traders behave differently — or perform better or worse — when the overall market mood shifts between **"Fear"** and **"Greed."**

In simple terms, I explored whether traders tend to take riskier positions or achieve higher profits when sentiment is optimistic versus cautious.

Data Overview

I worked with two main datasets:

- **Trades data (df_trades)** — includes trade-level details like entry/exit price, position size, closed profit/loss (PnL), and whether the trade was a buy or sell.
- **Sentiment data (df_sent)** — contains daily or weekly "fear/greed" indicators, or text-based sentiment classifications.

To make analysis smoother, I also created a **daily aggregated dataset** where I summarized mean and total PnL per day, grouped by sentiment.

Some preprocessing steps included:

- Converting timestamps into proper date formats.
- Handling missing columns (leverage, size, symbol) with safe defaults.
- Creating a fallback sentiment score when raw sentiment labels were missing (to keep the model consistent).

Data Exploration

Once the data was cleaned and aligned by date, I explored how profit and sentiment interacted.

One of the key visuals I created was a **boxplot of Mean Closed PnL vs Sentiment**, saved as
outputs/box_mean_closedpnl_by_sentiment.png.

This helped me visually compare trading outcomes across different emotional phases of the market.

Insights from the visuals:

- “Greedy” sentiment days showed slightly higher mean PnL compared to “Fear” periods.
- However, the spread (volatility) was larger, meaning traders took bigger risks.
- There wasn’t a perfect correlation — other factors like leverage and timing clearly matter too.

Modeling & Results

I built a simple **logistic regression model** to predict whether a trade would be profitable (closed_pnl > 0).

Features used:

- Leverage
- Trade size
- Side (Buy/Sell, encoded as 1/0)
- Sentiment score

After scaling features and running a 5-fold cross-validation, I measured the model’s performance using **ROC-AUC**.

Result:

```
Logistic AUC CV (5-fold): [0.48352632 0.79617614 0.56611678 0.5521482
0.73320312]

Mean AUC: 0.6262341141542769
```

Observations & Key Takeaways

Findings:

- Traders tend to take slightly riskier or larger positions when sentiment is positive ("Greedy").
- Profitability shows a weak but noticeable correlation with sentiment.
- Pure sentiment alone isn't enough to predict success — market context and strategy matter more.

Limitations:

- Missing or incomplete sentiment data required creating synthetic scores.
- The dataset was relatively small (about 188 records), limiting statistical confidence.

Next Steps:

- Use a larger dataset with real-time sentiment metrics.
- Try non-linear models like Random Forest or XGBoost to capture complex relationships.
- Introduce time-based features (e.g., lagged sentiment or rolling averages).

Conclusion

Working on this project helped me understand how quantitative trading data and sentiment analysis can be combined to uncover behavioral patterns.

Even with limited data, I saw that market emotion subtly influences trading behavior — and that combining psychology and numbers can give traders an extra edge.

This assignment strengthened my understanding of:

- Data cleaning and alignment between two datasets,
- Exploratory visualization for financial behavior, and
- Basic predictive modeling using logistic regression.

Overall, it was a practical blend of data science + trading psychology, and I really enjoyed bringing both together in this analysis.