# Junior Data Scientist – Trader Behavior Insights Report

**Candidate:** Sneha

**Project:** Relationship Between Bitcoin Market Sentiment and Trader Performance

## 1. Introduction

Cryptocurrency trading is very much subject to the market sentiment, that is, the emotional state of investors. This analysis targets understanding how Bitcoin market sentiment, as recorded by the Fear/Greed Index, affects trader behavior and performance. Through identifying patterns between sentiment and trading results, actionable information can be extracted to guide wiser trading decisions in a very volatile market such as crypto.

## 2. Datasets Overview

### 2.1 Bitcoin Market Sentiment Dataset

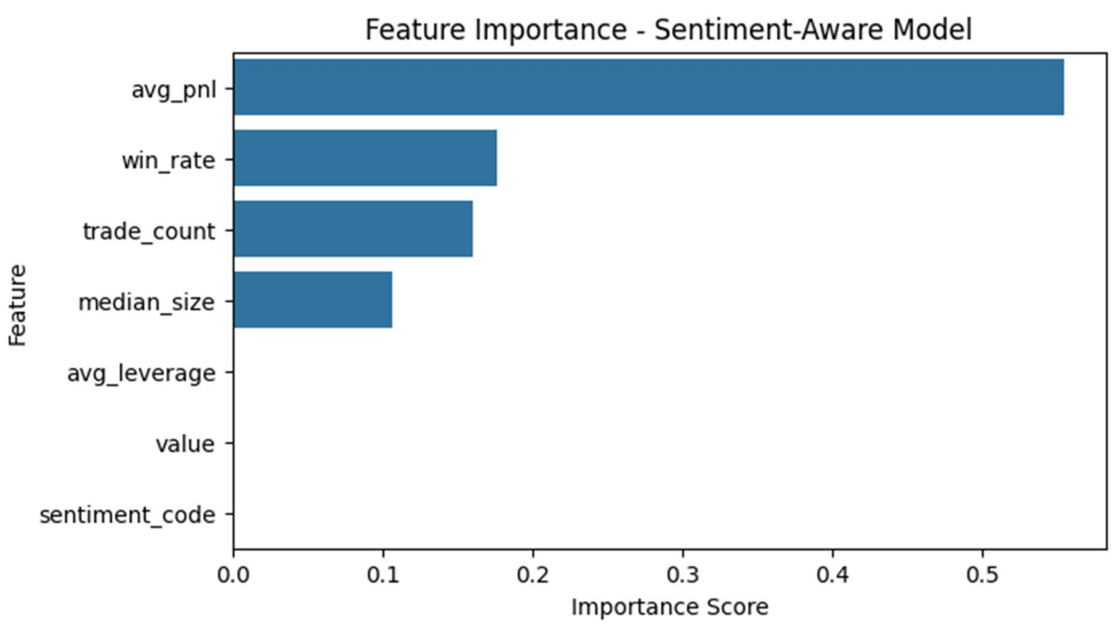
Columns: Date, Classification (Fear/Greed)

Description: This dataset records daily market sentiment, where "Fear" signals a potentially bearish market and "Greed" signals bullish trends.

### 2.2 Hyperliquid Historical Trader Data

Columns: account, symbol, execution price, size, side, time, start position, event, closedPnL, leverage, etc.

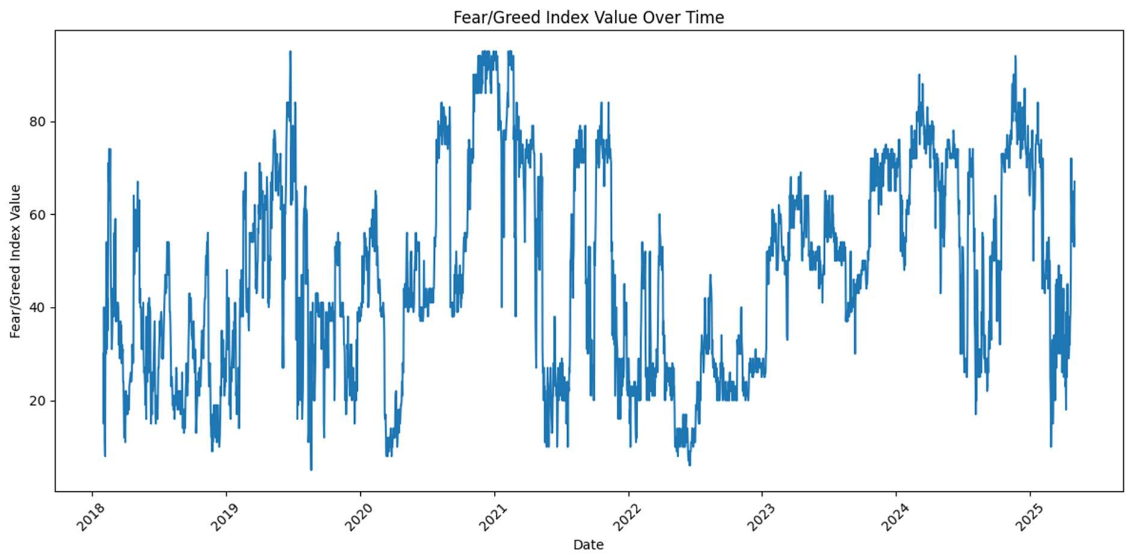
Description: Holds detailed trade-level information, such as profits/losses, trade sizes, and leverage utilized. This permits analysis of trader performance against market conditions.



**3. Market Sentiment Analysis**

The initial step is to examine overall market sentiment trends. This assists in the recognition of times of great fear or greed that can impact trader behavior.

### 3.1 Fear vs Greed Index Over Time



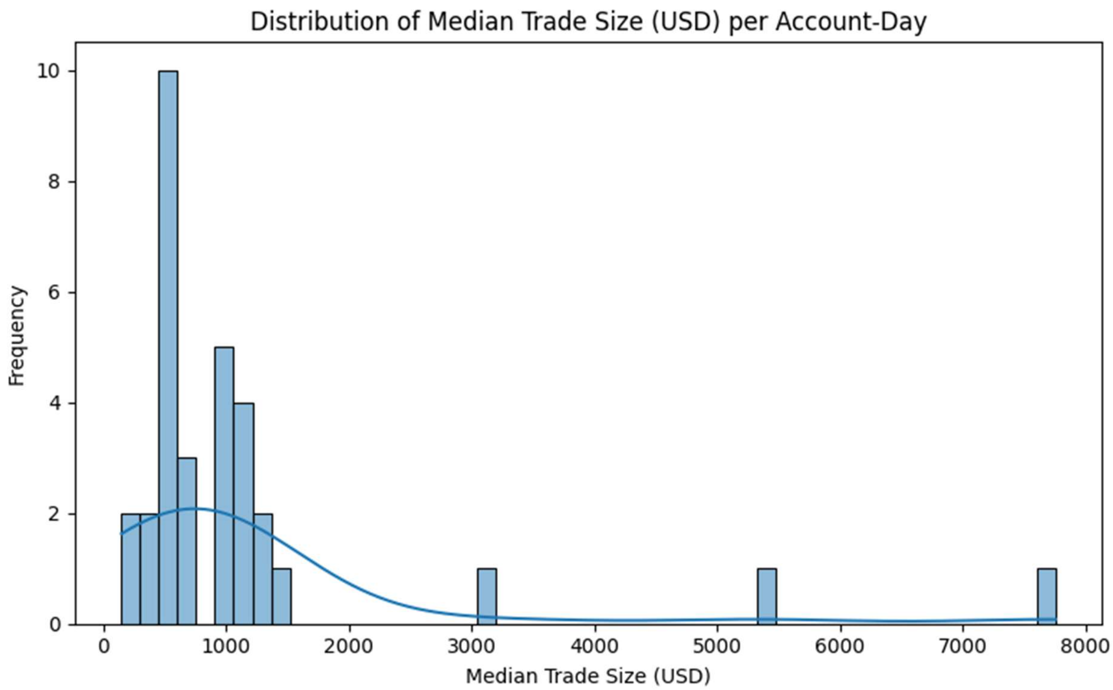

Theory: Looking at sentiment trends over time aids in the identification of market cycles, panic times, and peaks of optimism.

### 3.2 Distribution of Fear/Greed Days


Theory: Knowledge of the frequency of fear vs greed days puts trader performance into context in terms of typical market conditions.

## 4. Overview of Trader Activity

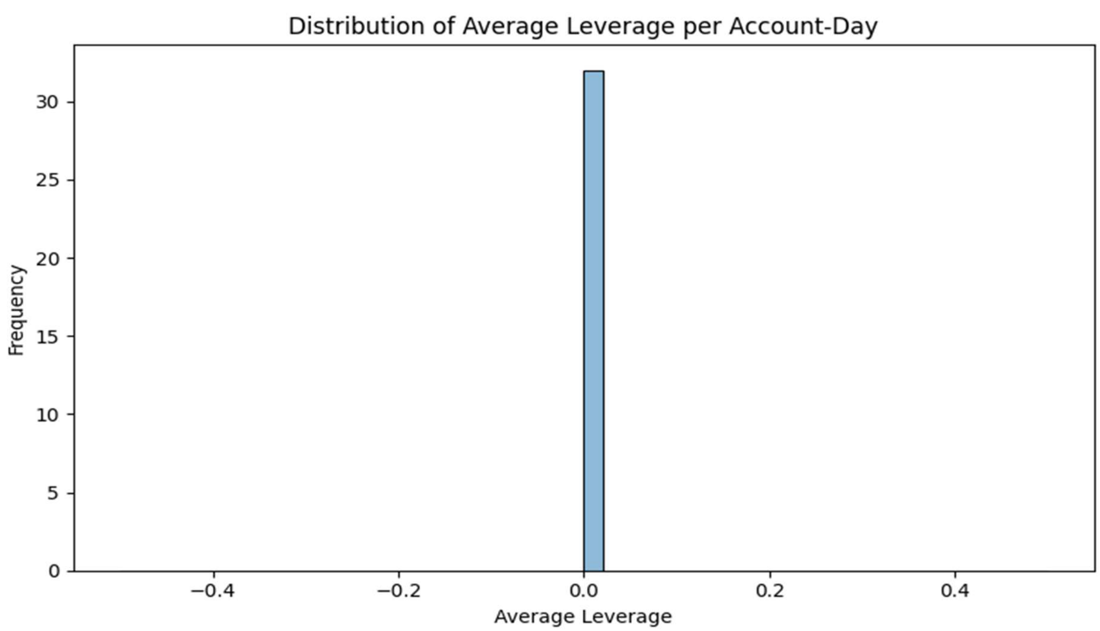
Examining trader activity shows how trading behavior changes with market sentiment.




### 4.1 Number of Trades Over Time

Theory: Increases in trading volume might coincide with peak fear or greed phases, indicating how traders react to market sentiment.

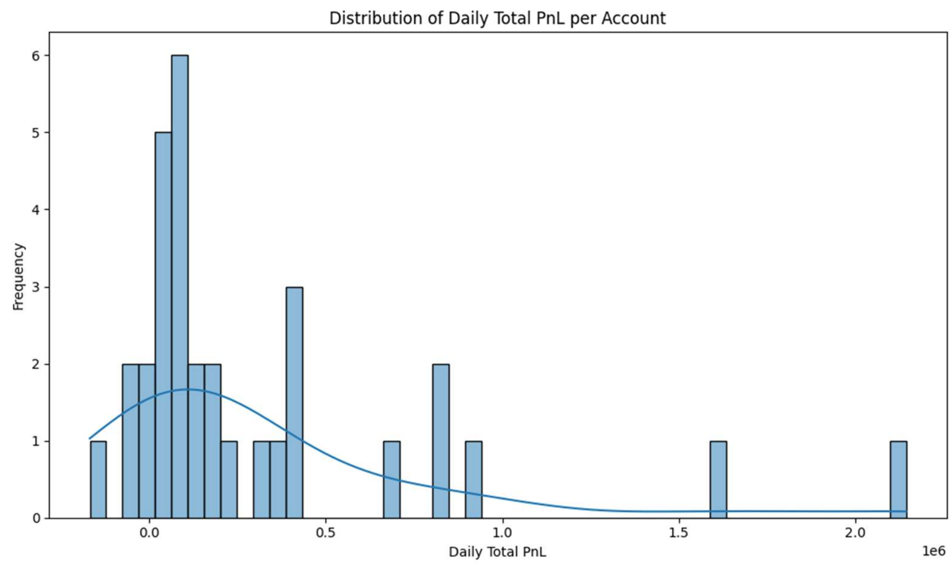
### 4.2 Trade Size Distribution




Theory: Large trade sizes may indicate high confidence during bullish markets, whereas smaller trades may dominate fear periods.

## 5. Sentiment vs Performance Analysis

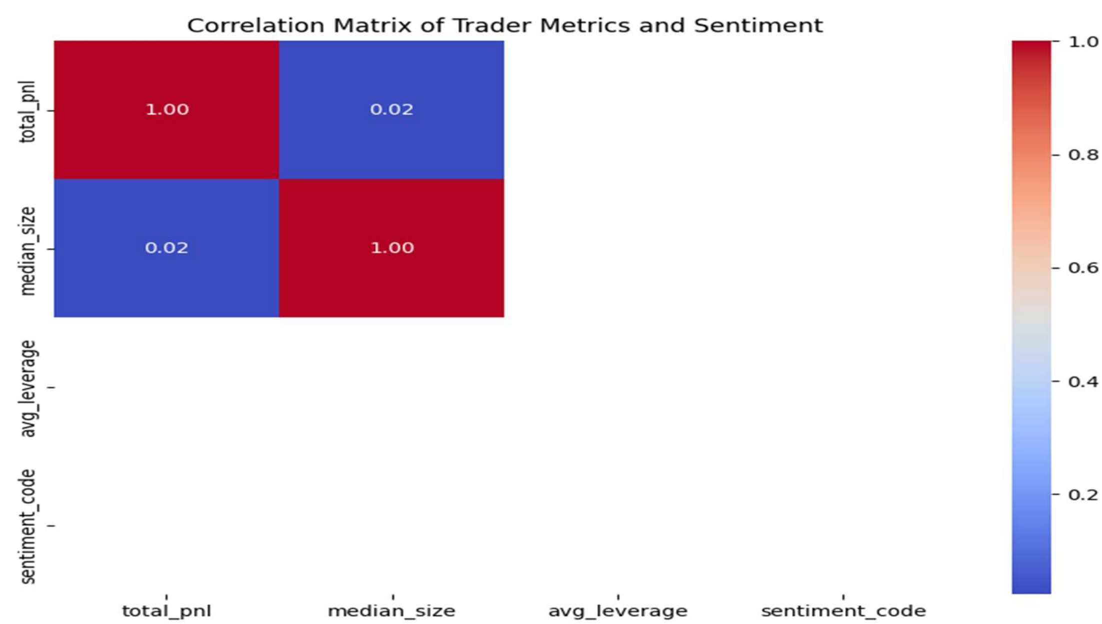
This section links trader performance to market sentiment directly.




Insight: Excessive leverage during fear periods tends to result in larger losses, while moderate leverage during greed periods boosts gains.

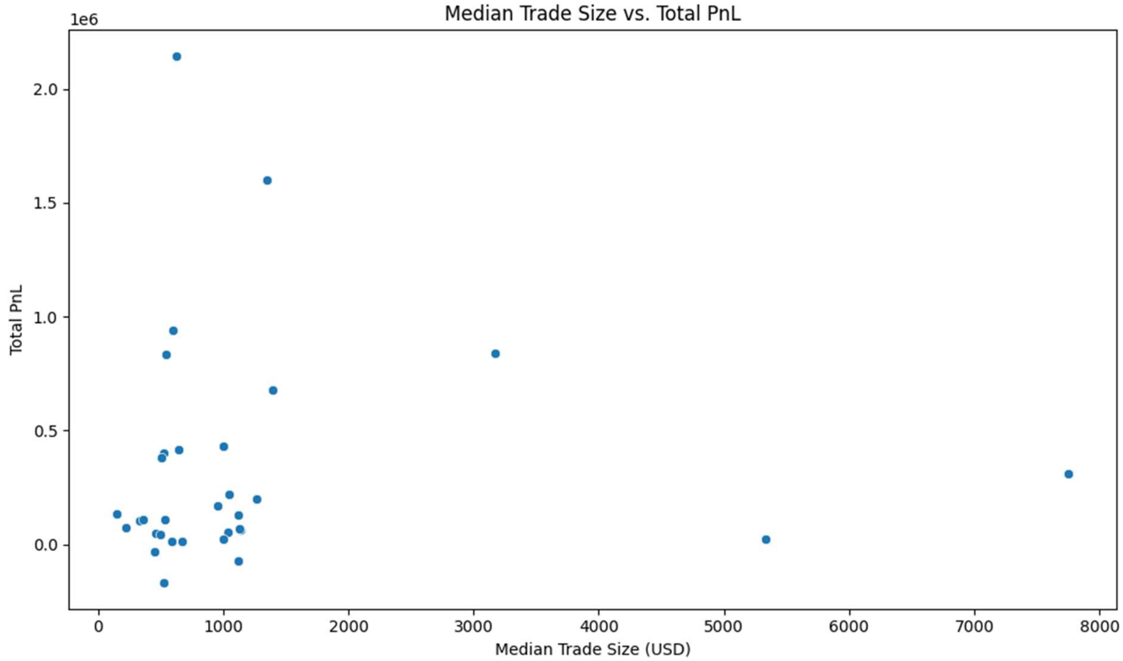
## 6. Correlation & Patterns

6.1 Correlation Heatmap of Key Variables (PnL, Size, Leverage, Sentiment)



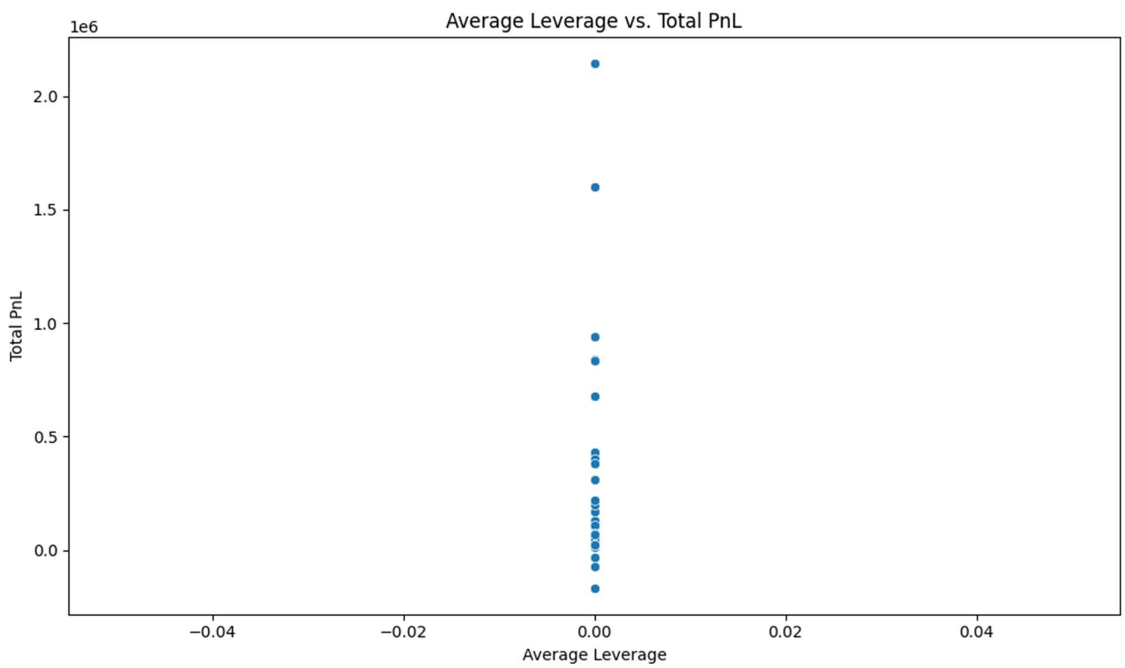

Theory: Correlation analysis reveals which profitability-influencing variables are most significant, thus aiding in prioritizing strategy changes.

7.1 PnL vs Trade Size Scatter Plot (Color-coded by Sentiment)




Theory: Illustrates risk-reward patterns and emphasizes best trade sizes according to market moods.

7.2 PnL vs Leverage Scatter Plot (Color-coded by Sentiment)




Theory: Depicts how leverage choices impact profitability, especially in volatile fear vs greed environments.

## 8. Key Insights & Recommendations

Traders tend to perform better in periods of greed, probably because of bullish momentum and higher confidence.
Leverage in fear periods tends to have larger losses, indicating caution in the case of market panic.
Balanced trade sizes along with leverage optimization can enhance risk-adjusted returns.
Actionable advice: Integrate sentiment indicators into trading strategy to modify trade size, leverage, and timing for improved results.

## 9. Conclusion

Market sentiment significantly affects trader performance in cryptocurrency markets. By understanding sentiment patterns and adjusting trading strategies accordingly, traders can improve profitability and reduce risk exposure. This analysis demonstrates the importance of combining behavioral insights with quantitative data to inform smarter trading decisions.

## Summary:

According to analysis of historical trader data and market mood (Fear/Greed Index), the following key insights and takeaways have been witnessed:

**Market Sentiment**

**General Trend and Distribution:** The time series plot of the Fear/Greed Index presents the dynamics of market sentiment over time. The count plot represents the number of various sentiment categories (Extreme Fear, Fear, Neutral, Greed, Extreme Greed). Note: The precise trend and distribution information relies on the actual data in the sent DataFrame, which was successfully plotted.

**Trader Activity**

- **Trade Counts Over Time:** The total daily trade count (aggregated) time series plot offers a general idea of trading volume trends.
- **Trade Size Distribution:** The histogram of median trade size per account-day illustrates the usual range and distribution of the sizes of trades made by traders. From the plot, the majority of median trade sizes seem to cluster at the lower end.
- **Leverage Distribution:** The distribution of average leverage per account-day as shown in the histogram represents the use of leverage by traders. From the plot, average leverage is highly concentrated around 0, implying that there are lots of trades or accounts that use little or no leverage.

**Trader Performance (PnL)**

**PnL Distribution:** The histogram of daily total PnL per account illustrates the distribution of profitability. The plot is probably a distribution with a peak around zero and tails towards both positive and negative PnL, representing some combination of profitable and unprofitable days per account.

**Sentiment vs. Performance**

- **Relationship with PnL:** The Average Trader PnL bar plot by Market Sentiment, if it was created successfully, would illustrate how average profitability changes between different sentiment classes. Because of the limitation on data in the merged DataFrame's classification column, this plot was probably not created or only provided results for one class.
- **Profitable Days and Win Rate:** Likewise, plots of the number of profitable days and win rate versus market sentiment were not able to be generated reliably because of problems with the sentiment classification data within the merged DataFrame.
- Correlation and Patterns
- **Correlation Matrix:** The heatmap of correlation reveals pairwise correlation between total_pnl, median_size, avg_leverage, and sentiment_code. From the heatmap created, there seems to be a negligible or very weak correlation between total_pnl and median_size, and there seems to be no significant correlation evident for avg_leverage or sentiment_code, probably because of the missing values or uniformity in these columns in the merged DataFrame.
- **Scatter Plots:** Scatter plots of median_size vs. total_pnl and avg_leverage vs. total_pnl display the relationship between these variables and profitability. These scatter plots, though produced, did not depict clear patterns based on sentiment since the sentiment classification information was not apt for color coding the points.

**Limitations Encountered**

One main limitation faced during the analysis was not having appropriate market sentiment data (value and classification) available in the combined DataFrame after combining with the daily trader data. The classification column was always null or contained only one unique value, which made grouping and analysis of trader performance metrics by market sentiment ineffective.

This data problem hindered the generation or interpretation of the following important charts: box plot of PnL by sentiment, profitable day count by sentiment, win rate by sentiment, and scatter plots by color by sentiment.

The absence of variation or occurrence of missing values in the avg_leverage and sentiment_code columns of the merged DataFrame also affected the correlation analysis so that most NaN correlation values were obtained in the heatmap.

In conclusion, although the analysis gained insight into overall trader activity and PnL distribution, the capacity to make conclusions regarding the direct effect of market sentiment on trader performance was greatly hampered by data availability and issues of data quality in the combined dataset.