
Data Science Analysis Report: Web3 Trading Behavior vs. Market Sentiment

I. Problem Statement and Methodology (Techniques Used)

The objective of the assignment was to analyze how collective **trader behavior** (profitability, risk, volume, and leverage/aggressiveness) aligns with or diverges from **market sentiment** (Fear vs. Greed), and to identify statistical trends that could inform smarter trading strategies.

The methodology employed a robust data science pipeline, utilizing several key techniques:

A. Data Engineering and Feature Creation

1. Data Cleaning and Preprocessing:

- **Dynamic Column Mapping:** The script was designed to dynamically identify and clean key columns like PnL and Volume.
- **Timestamp Normalization:** The Timestamp_IST column was converted to datetime objects, and a daily Date was extracted, which was crucial for merging with the daily sentiment data.

2. Robust Leverage/Aggressiveness Proxy:

- Since a direct leverage column might be inconsistent or missing, a more robust feature, **Calculated_Position_Ratio**, was engineered. This proxy measures trade size (Size_USD) relative to the collateral/position value ($\text{Start_Position} * \text{Execution_Price}$) on a trade-by-trade basis.

3. Daily Aggregation:

- The trade-level data was aggregated to a daily frequency (daily_metrics) to align with the daily sentiment data. Key metrics calculated include:
 - **Profitability:** Total_PnL, Average_PnL.
 - **Risk:** PnL_Volatility (Standard Deviation of PnL).
 - **Behavioral:** Total_Volume_USD, Max_Position_Ratio, Total_Trades, and Daily_Win_Rate.

B. Statistical and Machine Learning Models

- T-Test (Statistical Validation):**
 - A **Welch’s t-test** (assuming unequal variances) was used to determine if the difference in Total_PnL between Fear days and Greed days was **statistically significant** (i.e., not due to random chance).
 - Random Forest Classifier (Signal Identification):**
 - This supervised learning model was trained to predict the Sentiment_Group (Fear or Greed) based *only* on the daily trader behavioral metrics (Total_PnL, Volume, Leverage, etc.).
 - The output, **Feature Importances**, is used to identify **which trader behavior metric is the strongest signal** for the underlying market sentiment.
 - KMeans Clustering (Regime Discovery):**
 - This unsupervised learning model was applied to four key behavioral features to group trading days into **three distinct profiles** or "regimes" (e.g., High-Risk/High-Reward, Low-Activity).
 - The **Cluster Centers** provide a quantifiable average profile for each identified trading regime.
 - ARIMA (Time Series):**
 - An ARIMA (Autoregressive Integrated Moving Average) model was fitted to the Total_PnL time series to assess basic time-series dependencies and potential predictability.
 - Risk-Adjusted Reward (RAR):**
 - The ratio of **Average PnL / PnL Volatility** was calculated for both sentiment groups. This metric serves as a proxy for the **Sharpe Ratio**, measuring the average return generated per unit of risk taken.
-

II. Detailed Insights and Explanations

The analysis used **412 matched trading days** where both trader activity and clear Fear/Greed sentiment data were available. The results clearly show that collective trader behavior differs significantly between the two regimes.

A. Comparative Analysis: Fear vs. Greed

The average daily metrics show a clear pattern of behavioral divergence:

Metric (Avg. per	Fear Regime	Greed Regime	Behavioral
------------------	-------------	--------------	------------

Day)			Implication
Average PnL (USD)	\$32.23	\$45.85	Greed days are more profitable .
PnL Volatility (Risk)	\$382.02	\$259.27	Risk is lower on Greed days, contradicting common assumptions.
Daily Win Rate	0.33	0.38	Traders have a higher win rate during Greed.
Total Volume (USD)	\$5.69 Million	\$1.35 Million	Trading activity (volume) is 4.2 times higher during Fear.
Max Position Ratio	6.25 Million	5.10 Million	Fear days, on average, saw slightly higher peak aggressiveness (leverage proxy).
Risk-Adjusted Reward (RAR)	0.084	0.177	Greed periods offer 2.1 times better return per unit of risk.

Key Takeaways:

- Greed is Superior Performance:** The data overwhelmingly suggests that **Greed periods are associated with statistically superior risk-adjusted performance** (higher PnL, higher Win Rate, and surprisingly, lower Volatility).
- Volume Skew:** The highest trading volume occurs during **Fear** days, suggesting that low sentiment drives high transactional activity (likely due to forced selling/capitulation or aggressive accumulation by a subset of traders).

B. Behavioral Signals (Random Forest)

The Random Forest model identified which trader behavior metric is the most predictive of the market sentiment:

Feature	Importance Score	Insight
Total_Volume_USD	29.9%	The total money transacted is the strongest signal differentiating a Fear day from a Greed day.
Max_Position_Ratio	19.6%	The maximum trade aggressiveness is the second most important signal.
Lag_PnL	14.8%	Yesterday's PnL has moderate predictive power for today's sentiment group.

Signal Implication: The strong importance of Total_Volume_USD confirms that **market volume is the primary behavioral consequence of market sentiment**, followed closely by the degree of leverage/risk-taking.

C. Trading Regimes (KMeans Clustering)

KMeans identified three distinct day types, revealing "hidden trends" in trading profiles:

Cluster	Total PnL	PnL Volatility	Total Volume	Implication
---------	-----------	----------------	--------------	-------------

	(USD)		(USD)	
Cluster 1	\$185,670	\$1,608	\$15.34 Million	"Alpha Regime" : Days of extreme profitability, but accompanied by extremely high risk and volume.
Cluster 2	\$2,554	\$75	\$1.83 Million	"Extreme Leverage Regime" : Days marked by the highest calculated Max Position Ratio (1.5 Billion), but with low PnL and low risk. This suggests aggressive traders are often involved in non-profitable, low-volatility days.
Cluster 0	\$6,066	\$165	\$1.22 Million	"Low-Activity Regime" : Balanced, moderate PnL and volume.

Trend Discovery: The highest absolute profitability (Cluster 1) is not defined by leverage alone, but by a combination of **high PnL, high risk, and massive volume**, representing a highly active and volatile market day.

III. Final Summarized Insights and Strategic Recommendation

Statistical Validation

The difference in **Total Daily PnL** between Fear and Greed days was found to be **statistically significant** ($p=0.0246$), confirming that the observed profitability difference is not random.

Strategic Recommendation

The analysis suggests that the market's collective behavior tends toward efficiency during Greed. Strategies should consider the following:

The data suggests that traders, as a collective, tend to reduce volume but use **higher position ratios** (aggressiveness) during **Greed** periods, which... is associated with a more favorable risk-adjusted return than Fear periods. Strategies should focus on **managing volatility during Greed rather than avoiding the market entirely**.

While the largest trading days (by volume) occur during Fear, the most efficient and profitable periods (by RAR) occur during **Greed**. A smart strategy should seek to **exploit the favorable risk-adjusted environment of Greed** while recognizing that Fear days are prone to extreme, volume-driven movements (capitulation or accumulation) that are more difficult to navigate profitably.