Methodology

- **Data Preparation**: Loaded and preprocessed Fear/Greed index and historical trader data. Converted timestamps and handled missing values.
- **Metrics Calculation**: Computed daily PnL, win rate, trade size, leverage, trade count, and long/short ratios per account.
- **Trader Segmentation**: Segmented traders by leverage (High/Low) and trading frequency (Frequent/Infrequent) based on median values.
- **Trader Clustering**: Applied K-means clustering and PCA for dimensionality reduction to identify behavioral archetypes.
- **Predictive Model**: Developed a Logistic Regression model to predict next-day profitability using sentiment and behavioral features.

Insights

- **Profitability by Sentiment**: Traders showed higher average daily PnL during 'Fear' and 'Extreme Fear' conditions, despite sometimes lower win rates. This suggests contrarian success or higher impact trades during uncertainty.
- **Leverage Usage**: Extremely high average leverage values were noted in 'Greed' and 'Fear', warranting further investigation into outliers.
- **Long/Short Bias**: A pervasive long-only trading behavior was observed across all segments, with very few short trades.
- **Segmented Performance**: Low-leverage, frequent traders performed better in 'Fear' and 'Extreme Fear' markets.
- **Clustering Limitations**: K-means resulted in highly imbalanced clusters, indicating potential issues with cluster number or feature representation.
- **Predictive Model Weakness**: The Logistic Regression model achieved 64.76% accuracy but failed to predict non-profitable days effectively (0.05 recall for class 0), indicating a class imbalance problem.

Strategy Recommendations

- **Improve Predictive Model**: Address class imbalance (e.g., SMOTE, class weights), explore new features (e.g., volatility), and consider advanced models (e.g., XGBoost).
- **Refine Trader Clustering**: Re-evaluate the optimal number of clusters and conduct detailed profiling of more balanced clusters.
- **Investigate Anomalies**: Deep dive into the extremely high leverage values and the predominant long-only ratio to understand underlying causes (e.g., platform specifics, trader psychology, data interpretation).