DS Report: Analysis of Trader Performance and Market Sentiment and Inspection:

1.Initial Data Loading:

- I loaded the Historical Trader Data (df) and confirmed it has 211,224 entries and no missing values initially . Key columns include 'Closed PnL', 'Size USD', 'Side', 'Timestamp IST', 'Execution Price', etc.
- I loaded the Bitcoin Market Sentiment Dataset (sentiment_df) and confirmed it has 2,644 entries and no missing values initially . Key columns are 'date', 'value', and 'classification'.

2.Data Cleaning - Converting Timestamps:

- I converted the 'Timestamp IST' column in df and the 'date' column in sentiment_df to datetime objects . This was crucial for merging and time-based analysis.

3. Merging Datasets:

- I merged df and sentiment_df into merged_df based on the date . The resulting merged_df contains trade-level data aligned with the daily market sentiment.

4.Handling Missing Sentiment after Merge:

- Checking merged_df for missing values revealed 6 missing values in the 'value' and 'classification' columns . These were trades that occurred on dates not present in the sentiment dataset.
- I imputed these missing sentiment values: 'classification' was set to 'Neutral', and 'value' was set to 50 . A final check confirmed no remaining missing values in these columns.

5.Exploratory Data Analysis (EDA) - Key Findings:

- Summary Statistics: Displayed summary statistics for merged_df , giving us an overview of the numerical features like average 'Closed PnL' (~48.75), typical 'Size USD' (median ~597), etc.
- Average PnL by Sentiment Classification: Calculated the average 'Closed PnL' for each sentiment category :
- Extreme Fear: 34.54
- Extreme Greed: 67.89
- Fear: 54.29
- Greed: 42.74
- Neutral: 35.43

- **Insight: This showed that 'Extreme Greed' had the highest average PnL.**
- Trading Volume by Sentiment: Calculated total 'Size USD' by sentiment .

- **BUY vs. SELL Proportions in Extreme Periods**: Filtered data for Extreme Fear and Extreme Greed and analyzed BUY/SELL proportions:
  - Extreme Fear: ~51.1% BUY, ~48.9% SELL (slight buy bias)
  - Extreme Greed: ~44.9% BUY, ~55.1% SELL (pronounced sell bias)
  - **Insight: Indicated some contrarian behavior.**
- PnL by Side in Extreme Periods: Calculated average 'Closed PnL' for BUY vs. SELL during extreme periods :
- Extreme Fear: BUY (34.11), SELL (34.98) - Similar
- Extreme Greed: BUY (10.50), SELL (**114.58**) - Significant difference
- **Insight: Selling into Extreme Greed was associated with the highest average profitability.**
- Time Series Plot & Correlation: Visualized daily average PnL and sentiment over time and calculated their linear correlation (-0.01) .
  - **Insight: The weak correlation suggests no strong linear daily relationship, but the visual and categorical analysis points to patterns at sentiment extremes.**

6.Predictive Modeling - Process and Results:

- Feature Engineering: Created time-based features (hour, day of week, month) and one-hot encoded 'Side' and 'classification' (cell 3016d241). Crucially, added lagged sentiment value and lagged trading features (Execution Price, Size, Position, Fee) . Handled resulting missing values by dropping rows .
- Data Split: Defined features (X) and target (y) :split data into training/testing sets.
- Model Training: Trained a Random Forest Regressor model (original, updated with lagged sentiment, final with all lagged features) .
- Model Evaluation: Evaluated models using MAE, MSE, RMSE, and R^2 on test sets.
  - Original Model (basic features): $R^2 = 0.08$
  - Updated Model (with lagged sentiment): $R^2 = 0.43$ (Significant improvement)
  - Final Model (with all lagged features): $R^2 = 0.26$ (Better than original, but slightly lower than updated)
  - Insight: Lagged sentiment is a strong predictor; adding more lagged trading features slightly reduced R^2 in this model setup but the final model is still much better than the original.
- Feature Importances: Analyzed feature importances for the final model :
  - Top features include: 'Size USD' (0.166), 'Fee' (0.127), 'Start Position' (0.119), 'Size Tokens' (0.118).
  - Important lagged features: lagged_start_position (0.046), lagged_sentiment_value (0.041), lagged_size_usd (0.035), lagged_size_tokens (0.034), lagged_execution_price (~0.030).

- Insight: Confirmed importance of trade characteristics, lagged trading data, and lagged sentiment.

5. Key Findings and Insights for Smarter Trading Strategies

Based on the comprehensive analysis of the historical trader data and the Bitcoin market sentiment, several key findings and signals emerge that could potentially inform smarter trading strategies:

- Extreme Greed Represents a Historically Profitable Selling Opportunity:
    - Analysis of average 'Closed PnL' by sentiment classification revealed that trades executed during periods of Extreme Greed sentiment exhibited the highest average profitability (average PnL of 67.89).
    - Furthermore, a deeper look into trading behavior during Extreme Greed showed a pronounced tendency towards SELL trades (~55.1% of trades in this category were SELLs).
    - Crucially, when examining the average PnL specifically within Extreme Greed, SELL trades had a significantly higher average PnL (~114.58) compared to BUY trades (~10.50) during the same sentiment periods.
    - Implication: This strong association between Extreme Greed sentiment, selling behavior, and high average profitability suggests a potential contrarian strategy: identifying periods of excessive market optimism (Extreme Greed) and considering taking SELL positions to potentially capitalize on subsequent price corrections.
- Lagged Sentiment as a Predictive Signal:
    - While the overall linear correlation between daily average PnL and current daily sentiment value was weak (-0.01), our predictive modeling highlighted the significance of recent sentiment history.
    - Including the sentiment value from the previous trade (lagged_sentiment_value) as a feature in our Random Forest model dramatically improved its ability to explain the variance in 'Closed PnL' (increasing the R^2 from 0.08 to 0.43).
    - The feature importance analysis also ranked lagged_sentiment_value as one of the more important predictors.
    - Implication: This suggests that considering the sentiment of the market just prior to a trade can be a more impactful signal for predicting its potential outcome than solely relying on the current, simultaneous sentiment reading.
- Historical Trading Context is Highly Relevant:
    - Beyond sentiment, the characteristics of the trade itself ('Size USD', 'Fee', 'Start Position', 'Execution Price') were consistently identified as the most important features for predicting PnL across our models.

    - Moreover, lagged versions of these trading metrics (lagged_start_position, lagged_size_usd, lagged_size_tokens, lagged_execution_price, lagged_fee) also showed considerable importance in the final model's predictions.
    - Implication: This underscores that a comprehensive understanding of recent trading activity and market conditions leading up to a trade is vital. Incorporating features that capture the state of the trader's position, recent trade sizes, prices, and costs can significantly enhance the ability to predict a trade's profitability.
- Predicting Exact Outcomes Remains Complex:
    - Despite engineering numerous features and using a powerful model, the final model's R^2 of 0.26 indicates that a substantial portion (74%) of the variance in 'Closed PnL' in the test set is not explained by the features used.
    - Implication: While the identified signals and predictive model can provide valuable data-driven insights and potentially increase the likelihood of profitable trades on average, they do not guarantee profitability for every trade. Trading outcomes are influenced by many real-time and unpredictable factors not captured in this dataset. Therefore, robust risk management strategies are essential alongside any strategy based on these findings.