Project Report

Project Setup and Environment Check:

- We started by checking the Python version (sys.version) to ensure we were in a compatible environment. (Although a request for Python 3.9 was made, we continued with the available 3.12 runtime).
- We identified and installed necessary Python libraries for data analysis and market analysis, including pandas, numpy, matplotlib, seaborn, scipy, scikit-learn, statsmodels, and yfinance.
- We set up the project directory structure, creating a root folder (ds_<yourname>), a folder for CSV files (csv_files), and an outputs folder (outputs).

Data Loading and Initial Inspection

- We uploaded the historical_data.csv (trades) and fear_greed_index.csv (sentiment) files from your local machine into the csv_files directory in Colab.
- We loaded these CSV files into pandas DataFrames named trades and sent respectively.
- We performed initial checks on the DataFrames by displaying their heads (.head()), information (.info()), and descriptive statistics (.describe()) to understand their structure, data types, and basic characteristics.

Data Cleaning and Preprocessing

- We addressed the timestamp columns in both dataframes. We converted the numerical timestamp column in the sent DataFrame from Unix seconds to datetime objects (pd.to_datetime(sent['timestamp'], unit='s')).
- Similarly, we converted the numerical Timestamp column in the trades DataFrame, correctly identifying that it was in milliseconds (pd.to_datetime(trades['Timestamp'], unit='ms')).
- We extracted the date part from the trades['Timestamp'] column and created a new Date column, ensuring it was also in datetime format for merging.
- We converted the sent['date'] column to datetime format to ensure compatibility for merging.

Data Merging and Aggregation

- We merged the trades and sent dataframes based on their date columns to create a combined dataframe (merged_df) containing both trade-level data and daily sentiment information.
- We then aggregated the trade-level data in merged_df to create a daily summary (daily_trades_summary), calculating the total trading volume (total_size_usd), total closed profit/loss (total_closed_pnl), and the number of trades (number_of_trades) for each day.

- We combined the daily_trades_summary with the daily sentiment data from the original sent DataFrame to create a final daily_analysis_df where each row represents a day with aggregated trading metrics and corresponding sentiment data.

Feature Engineering and Normalization

- We normalized the sentiment classifications in daily_analysis_df, mapping "Extreme Fear" to "Fear" and "Extreme Greed" to "Greed" to simplify analysis into three main categories: 'Fear', 'Neutral', and 'Greed'.
- We created a lagged sentiment value feature (sentiment_value_lag_1) by shifting the 'value' column by one day to potentially explore if previous day's sentiment impacts current day's trading.

Analysis of Relationship between Sentiment and Trading Behavior

- We calculated the correlation matrix between the numerical sentiment value and the daily aggregated trading metrics (total_size_usd, total_closed_pnl, number_of_trades) to quantify linear relationships.
- We visualized these relationships using scatter plots to provide a visual representation of the correlations.
- We analyzed the average trading metrics grouped by the normalized sentiment classification to see how behavior differs across Fear, Neutral, and Greed states.
- We used a boxplot to visually compare the distribution of Total Closed PnL across the normalized sentiment classifications.
- We performed a Mann-Whitney U statistical test to determine if there was a statistically significant difference in Total Closed PnL between the 'Fear' and 'Greed' sentiment classifications.

Visualization of Daily Trends

- We created time series plots for the daily total size USD, total closed PnL, and number of trades to visualize the trends of these metrics over time.

Saving Processed Data and Code

- We saved the final daily_analysis_df as daily_summary.csv in the csv_files and outputs directories.
- We discussed methods for saving the entire notebook's code into a Python file for reproducibility.

Summary of Findings:

- Sentiment and Trading Activity: Our analysis consistently showed an inverse relationship between market sentiment value (Greed higher, Fear lower) and trading activity (volume and number of trades). Lower sentiment (Fear) was associated with higher trading activity, while higher sentiment (Greed) was associated with lower activity.

- Sentiment and Profitability: The relationship between sentiment and profitability (Total Closed PnL) was less clear. While the average total closed PnL was highest during "Fear" periods in our sample, a Mann-Whitney U statistical test did not find a significant difference in profitability between "Fear" and "Greed" classifications at the 0.05 significance level. This suggests that based on this analysis, we cannot definitively conclude that one sentiment state leads to statistically different profitability compared to the other.
- Statistical Significance: It's crucial to note that while descriptive statistics might show differences in averages, statistical tests help determine if these differences are likely real or due to random chance. In the case of profitability between Fear and Greed, the test suggested the difference was not statistically significant.

Insights

- Sentiment as an Activity Predictor: Market sentiment appears to be a reasonable indicator of overall trading volume and the frequency of trades. Periods of fear may present more dynamic trading environments.
- Profitability is Complex: The link between simple daily sentiment classification and overall daily profitability is not straightforward and doesn't appear statistically significant based on this analysis. This suggests that profitability is likely influenced by many other factors beyond just the daily sentiment level, or that the relationship is more nuanced (e.g., lagged effects, specific market conditions, individual trader skill).
- Opportunity in Fear? While not statistically confirmed, the observation of higher average PnL during fear periods is an interesting point for further qualitative investigation. It could hint that while many might panic sell during fear, those who trade effectively during such times (perhaps buying opportunities) can achieve higher returns on average.