Data Science Report: Trader Behavior vs. Market Sentiment

Introduction

The objective of this assignment was to explore and analyze the relationship between trader behavior and market sentiment. The goal was to use two key datasets—a Bitcoin Market Sentiment dataset and Historical Trader Data from Hyperliquid —to identify trends that could influence smarter trading strategies.

Data Validation and Primary Finding

The initial and most critical phase of the analysis was data validation. During this step, a significant temporal discrepancy was discovered between the two datasets. An examination of their date ranges revealed that they do not overlap:

- Historical Trader Data: Covers the period from March 1, 2024, to March 7, 2024.
- Fear & Greed Sentiment Data: Covers the period from February 1, 2018, to October 26, 2022.

This lack of a common time period makes it impossible to merge the datasets and directly analyze how daily market sentiment affects trading behavior. This data validation issue is the primary and most important finding of this project.

Analysis of Individual Datasets

While the core objective was unachievable, an exploratory data analysis (EDA) was performed on each dataset individually to extract any available insights.

- Trader Data Insights (March 2024): The one-week of trading data showed that buy-side trades were more frequent than sell-side trades, indicating a generally bullish stance among traders during that specific period. The distribution of profit and loss was characterized by numerous small gains and losses, with a few significant outliers.
- Sentiment Data Insights (2018-2022): The historical sentiment data revealed the volatile nature of the cryptocurrency market, with slightly more days classified as 'Fear' than 'Greed'. The data clearly illustrated major market cycles, including extended periods of fear during market downturns and sustained greed during bull runs.

Conclusion and Professional Recommendation

The primary goal of correlating daily trading activity with market sentiment could not be met due to the data mismatch. However, the ability to identify and diagnose such data quality issues is a fundamental skill for a data scientist.

In a professional setting, the correct next step would be to report this discrepancy to the project manager or data engineering team. The recommendation would be to acquire datasets that cover the same time period to enable the intended analysis. This project underscores the importance of rigorous data validation before proceeding to the modeling or analysis phase.