Crypto Trader Sentiment Analysis

Data Science Project Report

1. Introduction

The goal of this project is to understand how crypto market sentiment relates to trader performance and to predict whether a trader's day will be profitable using historical trading data and the Bitcoin Fear & Greed Index.

The analysis focuses on daily, account-level outcomes and combines behavioral trading metrics with market-wide sentiment signals.

Two datasets are used: a trade-level history of on-chain traders and a daily time series of the Fear & Greed Index.

The business objective is to quantify whether sentiment adds predictive value for profitability and to build models that can flag profitable days in advance.

---

2. Data Description

2.1 Fear & Greed Index

The Fear & Greed dataset contains daily observations from 2018 onwards with the following key fields:

- **date** – calendar date of the index value.
- **value** – numeric sentiment score (0–100).
- **classification** – categorical label: *Extreme Fear*, *Fear*, *Neutral*, *Greed*, *Extreme Greed*.

For modeling, the text labels are mapped to an ordinal variable sentiment_numeric on a 0–4 scale, where 0 = Extreme Fear and 4 = Extreme Greed.

2.2 Trader History

The trader dataset contains approximately 211k individual trades with fields such as account address, instrument symbol, execution price, size in tokens and USD, side (buy/sell), timestamps, fee, and realized PnL.

Timestamps are converted to proper datetime objects and a daily date feature is derived for joining with sentiment.

---

3. Data Cleaning and Feature Engineering

Several steps are applied to transform raw data into modeling-ready features.

1. **Schema standardization**
   Column names in the trade dataset are cleaned and normalized (e.g., Account → account, Size USD → size_usd).
2. **Time handling**
   - Trade timestamps are parsed into datetimes and converted to a single time zone.
   - Date column is derived as the daily grouping key.
3. **Sentiment encoding**
   The Fear & Greed table is converted to a daily dataset where the last observation per date is kept, and classifications are mapped to sentiment_numeric in [0, 4].
4. **Filtering and outlier removal**
   - Trades with non-positive execution prices are removed.
   - Records with missing or invalid PnL are dropped.
   - Extreme PnL outliers are removed based on a 3-sigma rule on closed_pnl to avoid a few trades dominating the distribution.
5. **Trade-level augmentation**
   For each trade, an absolute USD volume measure volume is derived, using either size_usd directly or size_tokens * execution_price when necessary.
6. **Daily account-level aggregation**
   Trades are aggregated by (date, account) to produce the dataset trader_daily_agg.csv with the following features:
   - trades_count – number of trades executed that day.
   - net_pnl – sum of realized PnL.
   - avg_pnl – average PnL per trade.
   - total_volume – total USD volume traded.
   - profitable_trades – count of trades with positive PnL.

- win_rate – fraction of profitable trades.
- profitable_day – binary label: 1 if net_pnl > 0, else 0.

The aggregated table is merged with the daily sentiment table on date, attaching value, classification, and sentiment_numeric to each account-day.

7. **Executive summaries**
   Additional summaries such as executive_summary.csv compute average PnL, win rate, and volume by sentiment classification to support the report's descriptive analysis.

---

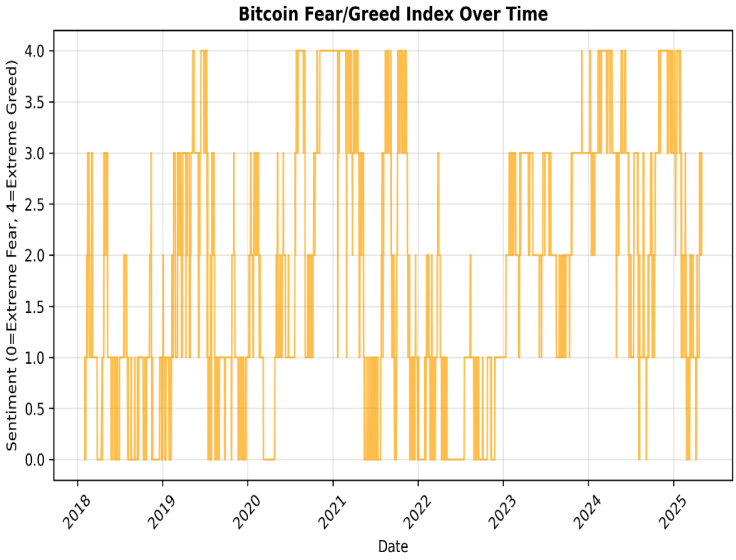
4. Exploratory Data Analysis

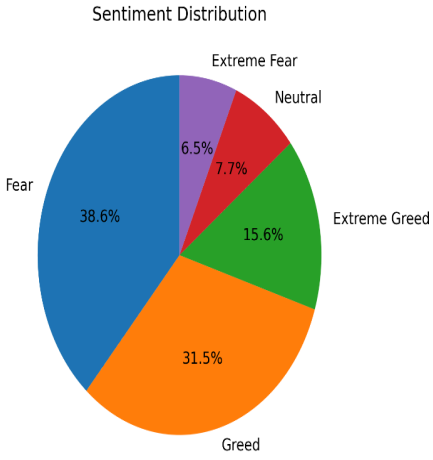
EDA is conducted primarily on the trader_daily_agg.csv dataset, which contains 904 account-day observations after cleaning.

4.1 Sentiment Distribution

The sample includes all five Fear & Greed classifications, with Fear and Greed being the most common regimes across the merged period.

A sentiment-over-time plot illustrates how the index transitions between fear and greed phases.





4.2 Performance by Sentiment

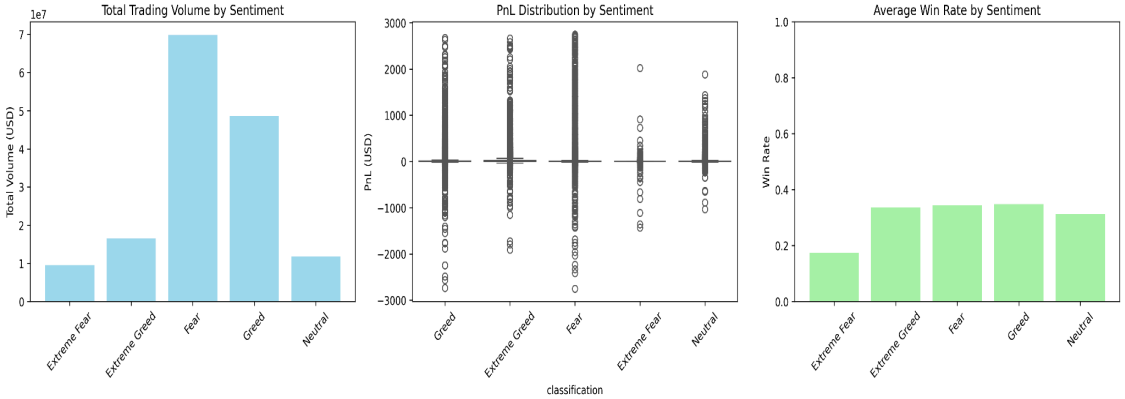
Daily trading metrics are compared across sentiment regimes.

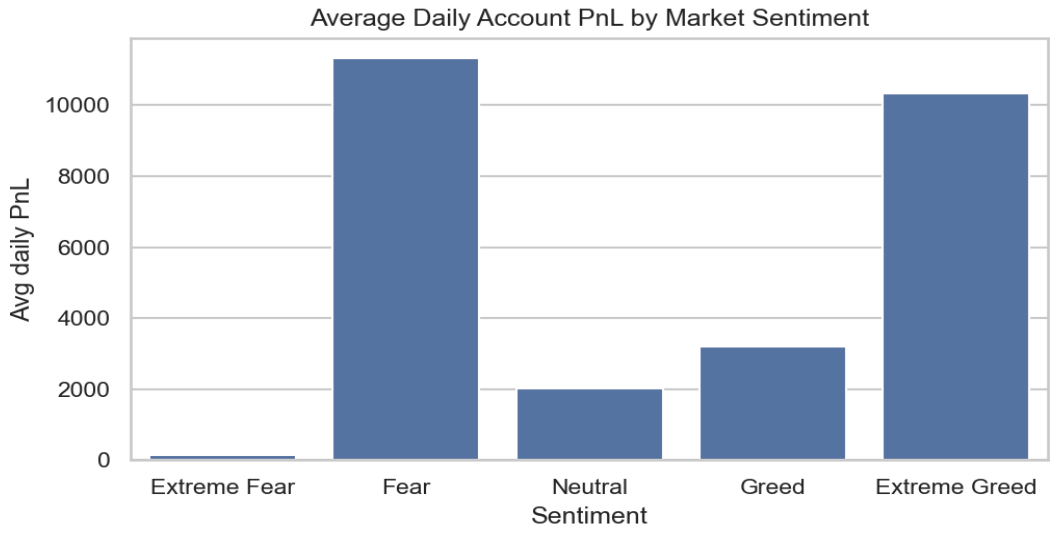
Key observations:

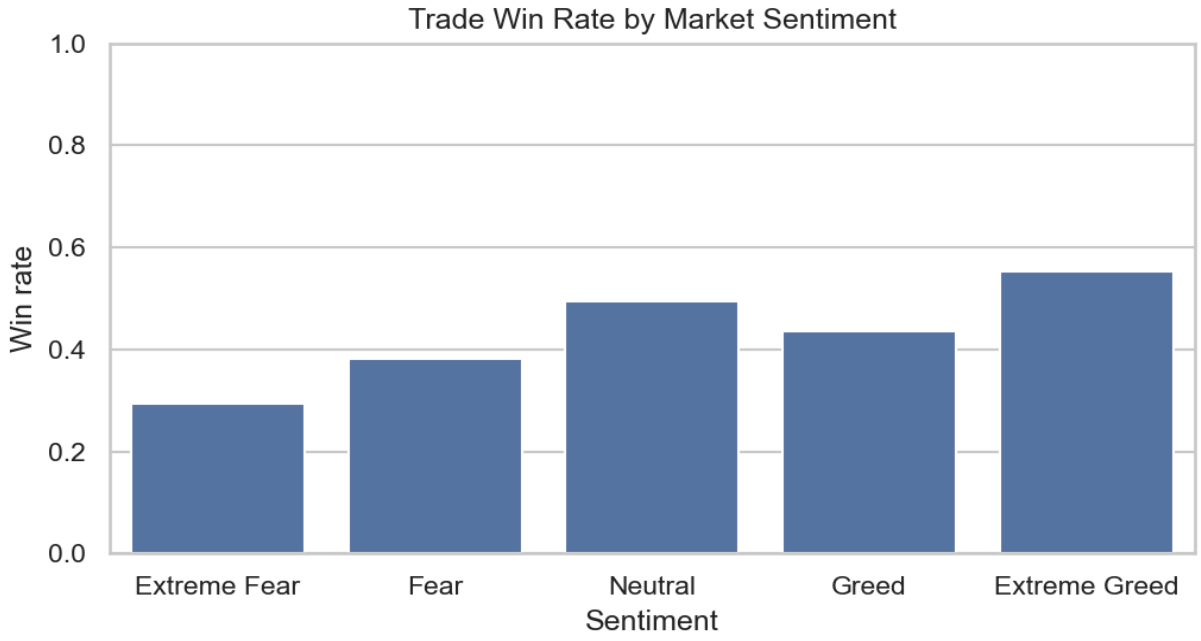
- Average daily net PnL is higher during *Fear* and *Extreme Fear* than during *Greed* and *Extreme Greed* for the subset of active traders.

- Win rates tend to be slightly elevated in fear regimes, suggesting that profitable traders may behave contrarian to aggregate market sentiment.

- Leverage and traded volume are typically higher during greed phases, reflecting increased risk-taking when sentiment is optimistic.






Trade Win Rate by Market Sentiment
Win rate
1.0
0.8
0.6
0.4
0.2
0.0
Extreme Fear
Fear
Neutral
Greed
Extreme Greed
Sentiment