## Crypto Derivatives Analysis

**Objective -**
**Analyse how trading behaviour (profitability, risk, volume, leverage) aligns or diverges from overall market sentiment (fear vs greed). Identify hidden trends or signals that could influence smarter trading strategies.**

```
Trading
│
├──── 1. Spot / Cash Market
│      ├──── Stocks: Buy/Sell actual company shares
│      │   (e.g., buy 100 Reliance shares → you own them)
│      └──── Crypto: Buy/Sell actual tokens
│         (e.g., buy 1 BTC → you own BTC in wallet)
│
├──── 2. Margin Trading (Leverage using borrowed funds)
│      ├──── Stocks: Borrow from broker to buy more shares
│      │   (e.g., ₹1 lakh own + ₹4 lakh borrowed → buy ₹5 lakh shares)
│      └──── Crypto: Borrow BTC/USDT to trade bigger
│         (e.g., 10x leverage → $100 becomes $1,000 position)
│
├──── 3. Derivatives
│      │
│      ├──── A. Futures
│      │      ├──── Stocks: Futures contracts (expiry weekly/monthly)
│      │      │   (e.g., NIFTY Futures, Reliance Futures)
│      │      └──── Crypto: Perpetual Futures (no expiry, funding fee mechanism)
│      │         (e.g., BTCUSDT Perpetual on Binance, Bybit, dYdX)
│      │
│      └──── B. Options
│           ├──── Stocks: Call & Put options on shares/indices
│           │   (e.g., NIFTY 22,000 Call Option)
│           └──── Crypto: Call & Put options on BTC/ETH
│              (e.g., Buy BTC Call Option at $60k strike)
│
└──── 4. Advanced / Other Forms
     ├──── Stocks: ETFs, Mutual Funds, Arbitrage
     └──── Crypto: Staking, Yield Farming, Liquidity Pools
```

**Our datasets are of derivatives (perpetual futures trading data)**

## Key Features of Perpetual Futures (Our Dataset)

1. **Leverage -> not present In out dataset**
    - **You trade with borrowed money (2x, 5x, 10x, even 100x).**
    - **Example: With $100 and 10x leverage → $1,000 position.**
    - **Risk: If price moves against you slightly → liquidation.**
2. **Direction (Long vs Short)**
    - **Long (BUY): Bet price will go up.**
    - **Short (SELL): Bet price will go down.**
3. **PnL (Profit & Loss)**
    - **Mark-to-market every second.**
    - **If you long BTC at $60k and price goes to $61k → profit.**
    - **If it goes to $59k → loss.**
4. **Funding Rate (Cross/Isolated Margin)**
    - **Ensures perpetual contracts stay close to spot market.**
    - **Example:**
        - **If everyone is long (greedy), longs pay shorts every 8 hours.**
        - **If everyone is short (fearful), shorts pay longs.**
5. **Order Matching**

- **Every trade needs two sides: a buyer and a seller.**
- **The order matching engine pairs your order with the best available opposite order.**

**Example:**

- **You place a BUY 1 BTC @ $60,000.**
- **There's a seller offering SELL 1 BTC @ $60,000.**
- **The engine matches → trade executes.**

**If no seller exists at your price, your order just sits in the order book until someone comes along.**

# Datasets overview -

1. Trade -> what user did (buy, sell, price_size, PnL)
2. Sentiment -> What the market mood was (Fear/Greed index)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 211224 entries, 0 to 211223
Data columns (total 16 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   Account           211224 non-null  object
 1   Coin              211224 non-null  object
 2   Execution Price   211224 non-null  float64
 3   Size Tokens       211224 non-null  float64
 4   Size USD          211224 non-null  float64
 5   Side              211224 non-null  object
 6   Timestamp IST     211224 non-null  object
 7   Start Position    211224 non-null  float64
 8   Direction         211224 non-null  object
 9   Closed PnL        211224 non-null  float64
 10  Transaction Hash  211224 non-null  object
 11  Order ID          211224 non-null  int64
 12  Crossed           211224 non-null  bool
 13  Fee               211224 non-null  float64
 14  Trade ID          211224 non-null  float64
 15  Timestamp         211224 non-null  float64
dtypes: bool(1), float64(8), int64(1), object(6)
memory usage: 24.4+ MB
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2644 entries, 0 to 2643
Data columns (total 4 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   timestamp       2644 non-null   int64
 1   value           2644 non-null   int64
 2   classification  2644 non-null   object
 3   date            2644 non-null   object
dtypes: int64(2), object(2)
memory usage: 82.8+ KB
```

# Dataset 1: Historical Trades (Users' Transactions)

**Shape: (211224, 16)**
**Columns:**

**1. Account**

- **What it is: Wallet or user ID executing the trade.**
- **Influence: Different accounts = different trading styles (scalper, investor, whale).**
- **Why it matters: You can group by accounts to see whether some users behave more rationally in Fear vs Greed.**
- **Example:**
    - **Account A always buys in Greed → often loses.**
    - **Account B buys in Fear → gains more.**

**2. Coin**

- What it is: The asset traded (e.g., BTC, ETH, SOL).
- Influence: Different coins react differently to sentiment.
- Why it matters: Altcoins usually exaggerate Fear/Greed moves compared to BTC.
- Example:
  - BTC PnL may be stable in Fear days,
  - But DOGE trades may swing wildly.

Each coin has different volatility, liquidity, and correlation with sentiment.

Liquidity => How easily you can buy or sell an asset without moving its price too much.

High liquidity → Lots of buyers & sellers, deep order books.

- Example: Bitcoin (BTC) and Ethereum (ETH).
- You can trade millions of dollars, and price barely moves.

Low liquidity → Few buyers & sellers, shallow order books.

- Example: a small altcoin.
- Even a $50k order can push the price up or down significantly.

Volatility => How much the price moves (fluctuates) in a given period of time.

High volatility → price changes fast, big swings
(e.g., Bitcoin jumping from $60k → $65k → $62k in a day).
Low volatility → stable price (e.g., a stablecoin hovering around $1).

2.1 Sentiment Sensitivity

- BTC (Bitcoin) usually drives the market. Fear/Greed in BTC often spills over into all other coins.
- Altcoins (smaller market caps) tend to overreact to sentiment.
  - Fear → Bigger crashes.
  - Greed → Bigger pumps.

Example:

- On a Fear day (Index = 20):

  - **BTC may drop 2%.**
  - **DOGE may drop 10%.**

**2.2 Liquidity Differences**

- **BTC & ETH are highly liquid → trades execute closer to execution price.**
- **Small-cap coins → higher slippage, more volatility.**

**Impact: A trader buying $50,000 worth of BTC might see little slippage. But $50,000 in a tiny altcoin can move the market itself.**

**2.3 Risk Profile by Coin**

- **BTC/ETH = “safer” trades.**
- **Meme coins/altcoins = “high risk, high reward.”**
- **Traders’ coin choices during Fear vs Greed can reveal risk appetite.**

**Example:**

- **In Greed (Index > 70): More trades in altcoins (people chasing higher returns).**
- **In Fear (Index < 30): Trades consolidate into BTC/ETH (people retreating to safety).**

**2.4 PnL Behavior by Coin**

- **Profitability depends not just on trade direction but also on which coin was traded.**
- **Hypothesis:**
  - **Trades in altcoins during Greed = more losses (FOMO buying at tops).**
  - **Trades in BTC during Fear = more gains (buying dips).**

**Example:**

- **Greed = User buys SHIB after hype → loses when correction happens.**
- **Fear = User buys BTC at $35k → gains when it recovers to $40k.**

**2.5 Correlation with Sentiment Index**

- **You can measure whether a coin is more/less correlated with the Fear & Greed Index.**
- **BTC is usually highly correlated (since index is BTC-heavy).**

- Some altcoins may not align perfectly (their moves depend on hype, news, etc.).

3. Execution Price

- What it is: Price at which trade was executed.
- Influence: Impacted by market volatility and liquidity.
- Why it matters: Compare execution price vs sentiment — were trades made during price spikes (Greed) or dips (Fear)?
- Example:
    - Buy BTC at $60,000 (Greed),
    - Sell BTC at $40,000 (Fear).

4. Size Tokens / Size USD

- What it is: Trade size in tokens and equivalent in USD.
- Influence: Larger positions = higher risk. Users might scale down exposure in Fear.
- Why it matters: Lets you see risk-taking behavior across Fear vs Greed.
- Example:
    - User trades $100k on Greed days, but only $5k on Fear days.

5. Side

- What it is: Buy or Sell.
- Influence: Driven by sentiment. Fear usually = more sells; Greed = more buys.
- Why it matters: Simple way to test: Do users buy more in Greed and sell more in Fear?
- Example:
    - 80% of trades are “Buy” when Greed > 70.

6. Start Position / Direction

- What it is: Indicates whether it was a long or short position.
- Influence:
    - Long = bullish outlook (more common in Greed).

- Short = bearish outlook (more common in Fear).

- Why it matters: Directly shows whether users follow or oppose market sentiment.
- Example:
  - Greed day: 90% of users go Long.
  - Fear day: more Shorts open.

7. Closed PnL (Profit & Loss)

- What it is: Profit or loss from trade when closed.
- Influence: Execution timing, leverage, fees, and sentiment-driven market moves.
- Why it matters: The ultimate performance metric — do users profit more on Fear days or Greed days?
- Example:
  - Average PnL = +$200 on Fear days.
  - Average PnL = –$50 on Greed days.

8. Fee

- What it is: Trading fees deducted.
- Influence: Higher for leveraged or high-frequency trades.
- Why it matters: Eats into profitability. Important to test if traders still profit after fees.
- Example:
  - Gross profit = $100, Fee = $20 → Net = $80.

9. Timestamp / Timestamp IST

- What it is: Date & time of the trade.
- Influence: Needed to merge with Fear & Greed data (by day).
- Why it matters: This is the key bridge column that links trades to daily sentiment.
- Example:

- 02-12-2024 22:50 → matches to Fear Index value for 2024-12-02.

**10. Transaction Hash, Order ID, Trade ID, Crossed**

- **What they are:** Technical/logging details for trade execution.
- **Influence:** Not important for sentiment analysis.
- **Why it matters:** Mostly irrelevant here — you can ignore these.

## Dataset 2: Fear & Greed Index

**Shape:** (2644, 4)
**Columns:**

**1. timestamp**

- **What it is:** UNIX timestamp for the sentiment reading.
- **Why it matters:** Used to merge with trade timestamps.

**2. date**

- **What it is:** Human-readable date.
- **Why it matters:** Easier to align with Timestamp IST from trade data.

**3. value**

- **What it is:** The sentiment score (0–100).
- **Influence:** Market-wide mood (fear or greed).
- **Why it matters:** The core column to compare against trade behavior.
- **Example:**
    - Value = 15 (Extreme Fear) → most users likely selling.
    - Value = 80 (Extreme Greed) → most users likely buying.

**4. classification**

- **What it is:** Categorical label ("Fear", "Greed", etc.).
- **Influence:** Direct mapping of value.

- Why it matters: Easier grouping than raw scores.
- Example:
  - classification = Fear → Value = 30.
  - classification = Greed → Value = 70.

## Phase 1: Data Preparation & Exploration

### Step 1: Data Quality Assessment & Preprocessing

- Data integrity ensured by cleaning and standardizing date formats across trading and Fear & Greed Index datasets, allowing for reliable analysis and visualization.
- Missing values were identified and handled through imputation or exclusion to preserve analytical quality.
- Merged all datasets on matching daily dates, establishing a consistent basis for cross-domain analysis.
- Created derived metrics:
  - Daily trading volume per coin
  - Mean execution price per coin
  - Sentiment regime labels (e.g., Extreme Fear, Neutral, Extreme Greed)

### Step 2: Exploratory Data Analysis

- Trading volume patterns:
  - Explored aggregated trading volumes by coin and day, noting temporal spikes or drops and differences in liquidity across assets.
- Fear & Greed Index distribution:
  - Plotted daily sentiment index values, highlighting fluctuations and extreme sentiment periods over time.
  - Generated bar charts for sentiment classification frequencies, emphasizing dominant and rare regimes.

## Additional Analytical Context

### 1. Percentage Change and Rate of Change (ROC)

$$ROC = \frac{(Current\ Value - Previous\ Value)}{Previous\ Value},$$

Percentage Change (pct_change) shows how much a value (like price or volume) has increased or decreased compared to its previous value, expressed as a percentage. It helps track how much something grows or shrinks over time.

Rate of Change (ROC) measures the speed or momentum of that change over a set period, indicating how quickly prices or volumes are rising or falling. It helps identify strengthening or weakening trends.

2. Genetic Algorithm

- Usage:
    - Implemented for optimizing trade selection or strategy parameters automatically based on fitness criteria (e.g., maximizing PnL, minimizing drawdown).
- Contribution:
    - Enabled robust search through potential strategy configurations, enhancing overall trading system effectiveness.

3. Hypothesis Tests Performed

Here is an expanded section for your report, detailing the additional statistical tests and measures performed in your analysis:

1. Skewness and Kurtosis

- Purpose:
    - Assessed the shape and nature of the price and volume distributions across coins and sentiment regimes.
- Details:
    - Skewness measured asymmetry in each metric’s distribution.
    - Kurtosis quantified the presence of outliers and extremity in distribution tails.

2. Mann-Whitney U Test

- Purpose:

- Compared median values of execution price and volume between two sentiment regimes (e.g., Extreme Fear vs Neutral), especially when normality assumptions did not hold.

- Details:

  - Used to determine if distributions came from different populations.

3. Kruskal-Wallis Test and Dunn’s Post Hoc Test

- Purpose:

  - Assessed mean or median differences among more than two sentiment regimes (Extreme Fear, Neutral, Extreme Greed).

- Details:

  - Kruskal-Wallis: Non-parametric test for comparing multiple independent groups.

  - Dunn’s test: Identified which regime pairs had significant differences after finding an overall effect.

4. Chi-Squared Contingency Test (chi2_contingency)

- Purpose:

  - Examined the relationship between categorical variables, such as sentiment regime and trade outcome (win/loss).

- Details:

  - Tested if the distribution of trade outcomes was independent of, or related to, sentiment classification.

5. Sharpe Ratio

- Purpose:

  - Calculated risk-adjusted profitability of trades for each sentiment regime.

- Details:

  - Sharpe ratio = Mean return / Standard deviation of return

    Standard deviation of returnMean return within each regime.

  - Higher ratios indicated better risk-adjusted performance during specific sentiment periods.

## Phase 3: Behavioural Analysis

### Profitability vs Sentiment Analysis

- **PnL Performance:**
    - **Grouped trades by sentiment regime, analyzed average and median profits during Extreme Fear, Neutral, and Extreme Greed periods.**
- **Success Rate:**
    - **Calculated proportion of profitable trades within each regime, identifying sentiment conditions with higher win rates.**
- **Risk-adjusted Returns:**
    - **Used Sharpe ratio and similar metrics by regime to capture risk/reward trade-offs, highlighting regimes with superior risk-adjusted performance.**