Anomaly Detection in Option Chain Data — Summary Report

Problem Understanding

The objective of the task was to detect **statistical or structural anomalies** in a 1-second resampled option chain dataset of the same underlying and expiry, recorded over two days. These anomalies may include:

- Sudden price movements,
- Implied volatility (IV) spikes,
- Unusual volume activity across instruments.

The data included multiple fields per option instrument (e.g., LTP, IV, traded volume) along with global fields like the underlying price. The dataset was multi-indexed and provided in `.parquet` format.

Anomaly Detection Logic

Based on the engineered features, the following rule-based logic was used:

- **IV Spike**: Detected when the z-score of IV > 3 (i.e., IV exceeds 3 standard deviations from its moving average).
- **Volume Spike**: Triggered when the rolling average of volume change exceeds 2.
- **LTP Jump**: Flagged when the absolute LTP difference exceeds 3× the standard deviation.

Only events occurring after the first few seconds of data (10-sec warmup) were considered valid.

Outcome

The script identifies anomalous `(timestamp, instrument, reason)` tuples and saves them to `results.csv`. Had return the reason why we considered them as anomaly in the last column.

- `IV spike`
- `Volume spike`
- `LTP jump`

This code generalized over various days and similar datasets, from which we can fetch a dataframe.