

Stock Data Quality and Coverage Summary

The dataset comprises twenty stocks with trading data spanning from 4 January 2021 to 22 December 2023, covering 1,082 trading days for nearly all tickers, with two extending to 25 December 2023. Data issues identified during processing included missing values, formatting inconsistencies, and outliers. Missing values were handled using forward-fill and back-fill techniques, resulting in a complete dataset.

Data cleaning steps included removing an unintended space from the rsi_14 field name and converting the volume column from float to integer. No duplicate records were found. Outliers were treated by capping values at the 5th and 95th percentiles. The dataset is now prepared for further analysis or modelling.

Missing Values Summary

Indicator	Missing (%)
sma_20	1.999742
macd	2.006193
rsi_14	2.128758
bb_upper	0.129016
bb_lower	0.129016
bb_width	2.135208
volume_ratio	1.999742
momentum_10	1.290156
momentum_20	2.580312
volatility_20	0.258031