EDA Insight Summary

Stock Price and Trend Analysis

The dataset reveals distinct price trends across selected stocks. STK009 (CapitalBank) shows a steady uptrend, STK017 (PowerGen) exhibits a clear downtrend, while STK001 (TechCorp) moves sideways with little directional change. Overall, most stocks move in line with market-level trends, with returns generally centred near zero but occasionally showing large deviations.

Average daily return: Highest for STK009 (0.14)
Typical (median) return: Highest for STK017 (0.17)
Volatility: STK017 experiences the greatest price fluctuations (std 2.97)

Sector Comparison

Performance varies by sector:

Average return: Finance leads, followed by Consumer and Healthcare; Technology lags.
Volatility: Energy stocks are the most volatile, highlighting greater risk.

Sectors differ in both median returns and dispersion, with Energy, Finance, and Technology showing higher variability.

Volume Analysis

Examining high-volume days:

On the busiest 10% of trading days, the average return was -0.68%, suggesting that volume spikes often coincide with sell-offs or negative market sentiment.
Correlation between volume and returns is generally weak: a few stocks (e.g., STK008, STK004) show slight positive correlation, while others (e.g., STK001, STK019) show weak negative correlation.

Overall, large-volume days are associated with significant price movements, but the relationship with next-day returns is generally modest, except around specific events.

Technical Indicators

RSI oscillates within typical bounds (30–70), indicating overbought or oversold conditions.
MACD crossovers align well with trend reversals in the sample stocks.

Correlation Analysis

Price-related columns are strongly positively correlated (e.g., open, high, low, close, adjusted close). Technical indicators such as MACD, moving averages, and Bollinger Bands show moderate correlations, indicating overlapping information but also some independent signals.

Feature Selection and Modelling

The following features were selected for trend prediction:
`momentum_10, momentum_20, sma_20, sma_200, macd, macd_signal, bb_width, atr_14, volume_sma_20, volume_ratio, trend_label`

Training set class distribution: Uptrend 35%, Downtrend 34%, Sideways 31%
Testing set class distribution: Uptrend 37%, Sideways 34%, Downtrend 29%

The classes are reasonably balanced, so oversampling techniques like SMOTE were deemed unnecessary.