1. Introduction: Context and Dataset Overview

The NASDAQ-100 index comprises the 100 largest non-financial companies listed on NASDAQ. Our dataset contains daily observations from January 2010 through September 2021 for every constituent: Open, High, Low, Close, Adjusted Close, Volume, and ticker “Name.” Because these firms span technology, retail, biotechnology, and more, the dataset captures broad market behavior over boom and bust cycles (including the 2020 COVID-19 shock).

In “Business Analysis Informed by Machine Learning,” we will:
• Articulate concrete business questions that leverage time-series and cross-sectional patterns.
• Select and apply machine learning techniques that address those questions.
• Translate model outcomes into actionable strategies for investors, portfolio managers, or corporate decision-makers.

2. Identifying Key Business Questions

When working with historical stock data, organizations typically seek to answer questions in three broad areas:

Financial Forecasting (Price Prediction & Risk Assessment)

Primary Question 1: Can we predict short-term stock price movements (e.g., next-day or next-week Adjusted Close) for high-volume, large-cap technology stocks in the NASDAQ-100?

Primary Question 2: Which stocks exhibit the highest volatility or tail-risk, and when should a portfolio rebalance to minimize drawdowns?

Market Trend Analysis (Sector & Cross-Sectional Patterns)

Primary Question 3: Do clusters of stocks (e.g., by sector or correlation behavior) move together, and can we identify early sector rotation signals?

Primary Question 4: How does trading volume spike relative to price changes, and can anomalies in volume forecast upcoming momentum or reversals?

Operational Efficiency & Automated Trading Insights

Primary Question 5: Which technical indicators (e.g., moving averages, momentum oscillators) combined with machine learning feature selection most reliably identify profitable entry/exit points?

Primary Question 6: Are there persistent patterns—such as mean reversion across certain

tickers—where algorithmic trading strategies could exploit short-term inefficiencies?

These questions combine forecasting (time-series regression/classification), clustering (unsupervised learning), and anomaly/momentum detection. Each yields different business insights: price predictions inform trading decisions, clustering reveals diversification opportunities, and volume-based anomalies suggest tactical rebalancing.