Objectives

This technical documentation aims to detail the derivation and preprocessing steps undertaken on the data before commencing the analysis.

Data Collection

Identify Data Sources

The initial step involves identifying market sectors before proceeding to select stocks. Within each sector, we select the top 5 companies for comparative analysis.

Health Care	1. Eli Lilly and Co. (LLY) 2. Merck & Co. Inc. (MRK) 3. Pfizer Inc. (PFE) 4. Abbott Laboratories (ABT) 5. Amgen Inc. (AMGN)
Materials	1. Air Products and Chemicals, Inc. (APD) 2. BHP Group Ltd (BHP) 3. Freeport-McMoRan Inc. (FCX) 4. Southern Copper Corporation (SCCO) 5. The Sherwin-Williams Company (SHW)
Energy	1. Brookfield Renewable (BEP) 2. Chevron Corp (CVX) 3. Exxon Mobil Corp (XOM) 4. NextEra Energy (NEE) 5. Shell PLC (SHEL)
Consumer Discretionary	1. Amazon.com Inc. (AMZN) 2. LVMH Moet Hennessy Louis Vuitton SE (LVMHF) 3. Tesla Inc. (TSLA) 4. The Home Depot Inc. (HD) 5. Toyota Motor Corp (TM)
Consumer Staples	1. Coca-Cola Co (KO) 2. Costco Wholesale Corp (COST) 3. Nestle SA (NSRGY) 4. Procter & Gamble Co (PG) 5. Walmart Inc (WMT)
Industrials	1. Caterpillar Inc. (CAT) 2. General Electric (GE) 3. Honeywell International Inc. (HON) 4. Union Pacific Corp. (UNP) 5. United Parcel Service Inc. (UPS)
Utilities	1. American Electric Power Company, Inc. (AEP)

	2. Duke Energy Corporation (DUK) 3. NextEra Energy, Inc. (NEE) 4. Sempra (SRE) 5. The Southern Company (SO)
Financials	1. Berkshire Hathaway Inc. (BRK.B) 2. Industrial and Commercial Bank of China (IDCBY) 3. JPMorgan Chase & Co. (JPM) 4. Mastercard Inc. (MA) 5. Visa Inc. (V)
Information Technology	1. Alphabet Inc Class A (GOOGL) 2. Apple Inc (AAPL) 3. Meta Platforms Inc (META) 4. Microsoft Corp (MSFT) 5. NVIDIA Corp (NVDA)
Communication Services	1. Comcast Corporation (CMCSA) 2. Netflix (NFLX) 3. T-Mobile US (TMUS) 4. Verizon Communications Inc. (VZ) 5. Walt Disney (DIS)
Real Estate	1. American Tower Corporation (AMT) 2. Equinix (EQIX) 3. Prologis (PLD) 4. Public Storage (PSA) 5. Welltower Inc. (WELL)

Stock Prices

Stock prices are sourced from the NASDAQ website, gathering a decade's worth of data for analysis. Stock prices are sourced from the NASDAQ website due to its status as a reputable and widely recognized stock exchange platform. NASDAQ offers reliable and up-to-date stock market data, ensuring accuracy and accessibility, making it an ideal source for comprehensive stock price information crucial for analysis and decision-making.

The selection of a 10-year timeframe is justified by its comprehensive span, offering a robust historical perspective for more accurate and insightful analysis.

Refer to aggregate_stock_prices.py for data cleaning.

Financial Indicators

Financial Indicators data are gathered from the SEC website. The utilization of data from the SEC website to gather Financial Indicators is driven by the site's credibility and regulatory oversight, ensuring the accuracy and integrity of financial disclosures submitted by companies. This choice guarantees reliable and standardized financial data, pivotal for in-depth analysis.

Refer to get_financial_data.ipynb for retrieving of financial indicators from SEC
Refer to aggregate_financial_data.ipynb for data cleaning.

Data Storage

Data storage is facilitated by MySQL. In handling stock data, a relational database like MySQL is preferred over non-relational databases due to its ability to establish structured relationships between different stock-related data points. This relational structure enables efficient querying, analysis, and management of interconnected stock data, ensuring integrity and consistency in financial datasets.

Additionally, non-relational databases are well-suited for storing dynamic data or for applications managing diverse data types. In scenarios involving algorithmic trading, where frequent updates occur, and diverse data like hourly price updates and unstructured news data are handled, the flexibility of non-relational databases becomes advantageous.

Refer to script - stocks.sql for creating databases and tables
Refer to script - insert_to_mysql.py for inserting values