**Chapter 1: Stock Selection Using Alpha Vantage API and Filtering Criteria**

**Introduction:**

In this chapter, we will outline the initial step in our pairs trading strategy: selecting a universe of stocks from a specific sector or industry, identifying stocks within that sector or industry, and applying filtering criteria based on market capitalization, liquidity, and trading volume. Pairs trading is a market-neutral trading strategy that involves identifying and trading correlated stocks to profit from temporary discrepancies in their price relationship. By focusing on a specific sector or industry, we can identify stocks with a higher potential for correlation, which is crucial for the success of a pairs trading strategy. We have chosen the S&P 500 index as our stock universe, as it represents a diverse range of large-cap companies across various sectors, providing a comprehensive selection of stocks to choose from for our strategy.

1. **Data Collection:**

We start by collecting the constituents of the S&P 500 index from a publicly available source, such as Wikipedia.



**Snapshot of the components of the SP500 in Wikipedia table.**

We use the Pandas library to read the HTML table containing the list of S&P 500 companies and their corresponding sectors (GICS Sectors) directly into a DataFrame. This dataset provides us with the necessary information to filter stocks based on a specific sector of interest.

****

**All sectors of the components of the SP500 in Wikipedia table.**

1. **Selecting a Sector:**

In our analysis, we have focused on the following sectors: **Energy**, **Real Estate**, **Communication Services**, **Information Technology**, **Health Care**, **Industrials**, and **Financials**. By working with these sectors, we aim to cover a diverse range of companies and industries within the S&P 500 index.

1. **Fetching Stock Data with Alpha Vantage API:**

We use the *Alpha Vantage API* to fetch stock data, including the daily-adjusted time series and company overview for each stock in the selected sector. The daily-adjusted time series provides information about stock prices, dividends, and stock splits, while the company overview provides data on shares outstanding and other financial metrics. We utilize the **TimeSeries** and **FundamentalData** endpoints of the Alpha Vantage API to fetch this information.

1. **Defining Filtering Criteria:**

To select stocks based on specific criteria, we define the following filtering criteria in our code:

* Market Capitalization: To select stocks based on market capitalization, we set a minimum and maximum market capitalization range to focus on stocks with a market capitalization between 2 million USD and 10 billion USD. This range is chosen to target mid-cap companies, which typically exhibit a balance between growth potential and stability compared to their small-cap and large-cap counterparts. Mid-cap stocks can provide investors with an opportunity for higher returns than large-cap stocks while maintaining a lower risk profile than small-cap stocks. By focusing on this market capitalization range, we aim to identify stocks that have the potential for strong performance while maintaining a manageable level of risk for our pairs trading strategy.
* Liquidity Ratio: We define a minimum liquidity ratio of 0.1% to ensure that stocks have a sufficient amount of liquidity for trading. The liquidity ratio is calculated as the average daily trading volume divided by the shares outstanding. This ratio serves as an important measure of the ease with which a stock can be bought or sold in the market without significantly affecting its price. A higher liquidity ratio indicates that the stock can be traded more easily, reducing transaction costs and minimizing the potential impact of trading on the stock's price.

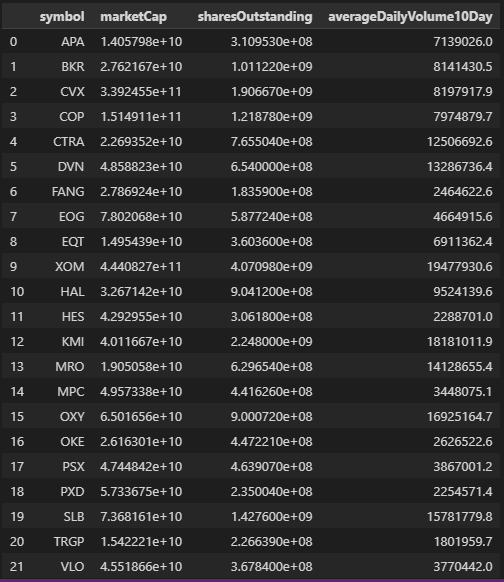
By setting a minimum liquidity ratio of 0.1%, we aim to select stocks that exhibit a reasonable level of trading activity relative to their outstanding shares, ensuring that our pairs trading strategy can be executed efficiently. This criterion helps to avoid stocks with limited trading activity or low liquidity, which may be subject to higher price volatility and transaction costs, thus increasing the risk and reducing the potential profitability of our strategy.

This minimum threshold of 0.1% strikes a balance between maintaining liquidity while also allowing for a wider range of stocks to be included in our analysis, potentially increasing the opportunities for successful pairs trading.

* Average Daily Trading Volume: We set a minimum average daily trading volume of 500,000 shares to ensure that the selected stocks have adequate trading activity. Average daily trading volume is an important metric to consider, as it reflects the number of shares that are traded on average during a given trading day. Stocks with higher trading volumes typically have higher liquidity, which allows for easier execution of trades and lower transaction costs.

By setting a minimum average daily trading volume of 500,000 shares, we aim to filter out stocks with low trading activity that might be difficult to trade in practice. This threshold helps to ensure that our pairs trading strategy can be implemented efficiently and with minimal impact on the stock's price due to our trades. By focusing on stocks with adequate trading activity, we also reduce the risk of facing price manipulation or "thin" markets, which could adversely affect the performance of our pairs trading strategy.

1. **Filtering Stocks:** We loop through the stocks in the selected sector and fetch their data using the Alpha Vantage API. We calculate the market capitalization, liquidity ratio, and average daily trading volume for each stock, and then apply our filtering criteria to select the stocks that meet these criteria. The selected stocks are then stored in a DataFrame and can be saved as a CSV file for further analysis or use in trading algorithms.



**Example of the components of the SP500’s energy sector.**

**Conclusion:** In this chapter, we have outlined the process of selecting stocks from a specific sector using the Alpha Vantage API and applying filtering criteria based on market capitalization, liquidity ratio, and average daily trading volume. This approach allows investors and traders to focus on a specific sector and select stocks that meet their desired criteria, which can be useful in building a targeted investment portfolio or developing sector-focused trading strategies such as our pairs trading strategy.

**Chapter 2: Gathering Historical Price Data for Selected Stocks**

**Introduction:**

In this chapter, we discuss the process of collecting historical price data for the selected stocks identified in the previous step. We have chosen a specific set of stocks based on their market capitalization, liquidity, and trading volume. This data collection is crucial for the development of our pairs trading strategy, as it provides us with the necessary information to identify potential pairs with a high degree of statistical similarity.

To begin the data collection process, we first set the time period for our historical price data. We have chosen a start date of January 1, 2015, and an end date of March 26, 2023. This time frame is selected for several reasons:

1. **Sufficient data for analysis**: By choosing a time frame of over eight years, we ensure that we have enough data points to perform data mining for our pairs trading strategy. This lengthy time frame allows us to capture multiple market cycles and various macroeconomic conditions, thereby providing a robust dataset for analysis.
2. **Recent macroeconomic conditions:** By including data up to the present day, we capture the impact of recent macroeconomic events and market trends on our selected stocks. This information is valuable in understanding how these events might affect the performance of our pairs trading strategy and assessing the potential risks involved. Some of the key macroeconomic events and trends that may have influenced our selected stocks during this period include:
   1. *Monetary policy:* Central banks around the world have implemented various monetary policy measures, such as interest rate changes and quantitative easing, in response to changing economic conditions. These actions can significantly impact the performance of stocks, as they influence borrowing costs, liquidity, and overall market sentiment.
   2. Geopolitical events: Over the past several years, there have been several significant geopolitical events, such as trade wars, Brexit, and international conflicts. These events can create uncertainty and volatility in the financial markets, which can, in turn, affect the performance of individual stocks and sectors.
   3. *Technological innovations:* Rapid advancements in technology, particularly in areas such as artificial intelligence, renewable energy, and digital communication, have had a profound impact on various industries. Companies that are well positioned to capitalize on these innovations may experience significant growth, while those that fail to adapt may struggle.
   4. *Regulatory changes:* Changes in regulations and government policies can have a direct impact on the performance of specific industries and stocks. For example, changes in environmental regulations can affect the energy sector, while changes in healthcare policy can impact the healthcare industry.
   5. *Economic cycles:* Our dataset captures various stages of the economic cycle, including periods of expansion, contraction, and recovery. Understanding how stocks perform during different phases of the economic cycle can help inform our pairs trading strategy and improve its risk-adjusted returns.

By considering these recent macroeconomic events and trends, we can gain a better understanding of the potential risks and opportunities associated with our pairs trading strategy. This knowledge can help us make more informed decisions as we develop and refine our trading algorithm.

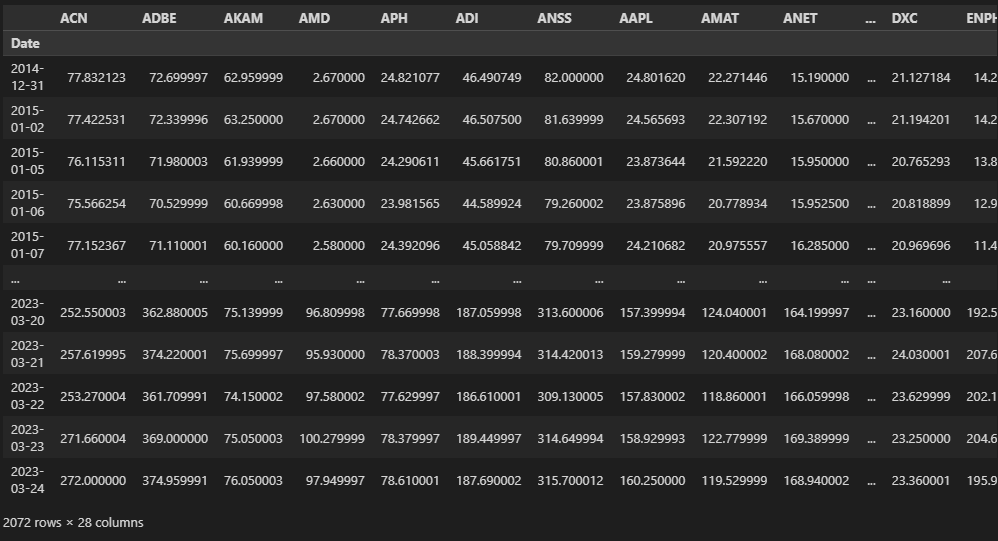
1. **Daily frequency**: We have chosen to collect daily price data for our analysis. Daily data strikes a balance between providing enough granularity to capture short-term price movements and reducing the impact of intraday noise on our analysis. Additionally, daily data is more readily available and accessible compared to higher-frequency data (e.g., intraday or tick data), making it a practical choice for our analysis.

After setting the time period for data collection, we proceed to gather the historical price data for each of our selected stocks. We use the **yfinance** library to fetch this data, focusing specifically on the adjusted closing prices.

We choose to focus on the adjusted closing price in our pairs trading strategy for several key reasons. The adjusted close accounts for corporate actions such as dividends and stock splits, providing a consistent and comparable price series. This enables accurate performance measurement and reduces the impact of temporary price fluctuations. Furthermore, using the adjusted close simplifies data processing, as it allows for direct comparison across different time periods without additional adjustments. By utilizing the adjusted closing price, we can better identify long-term trends and relationships between stocks, which is crucial to the success of our pairs trading strategy.

To ensure the quality of our dataset, we check for missing data points and fill any gaps using linear interpolation. This method estimates the missing values by assuming a linear relationship between the adjacent data points, providing a reasonable approximation for our analysis.

Once we have collected and cleaned the historical price data for each stock, we store this data in a single DataFrame, with each stock's adjusted closing prices in a separate column. This consolidated dataset is then saved to a CSV file for further analysis and use in our pairs trading strategy.



**Example of the Adjusted Close data of the financial sector.**

**Conclusion:**

In summary, collecting historical price data for our selected stocks is a critical step in the development of our pairs trading strategy. By choosing an appropriate time frame and frequency for our data, we ensure that our analysis captures the relevant market dynamics and trends, providing a solid foundation for our trading strategy.