**Project Proposal**

**Introduction**

With investing becoming increasingly popular, especially among retail investors, developing tools that assist in making informed decisions is more valuable than ever. This project aims to explore patterns and behaviors in the stock market using a data-driven approach that integrates multiple sources.

By collecting and combining data from TradingView, Yahoo Finance, and Wikipedia, I aim to investigate the short-term performance of companies in the S&P 500 index. The data includes company-level fundamentals (from TradingView), daily historical price behavior (from Yahoo Finance), and sector classifications (from Wikipedia). These datasets will be used to analyze relationships between volume, volatility, price movement, and sector-based performance.

The primary goal of this project is to identify which stocks may potentially be buy, sell, or hold opportunities and to evaluate which sectors are currently performing well across the S&P 500. This analysis can benefit investors and analysts looking to better understand how price changes relate to broader market behavior.

**Data**

***Dataset 1:* Trading View S&P500 Components:** <https://www.tradingview.com/symbols/SPX/components/>

* Scraped using Selenium
* Currently includes the following fields for the first 100 companies in the S&P 500 index:
  + Ticker (symbol), company name, stock price, price % change, and volume

***Dataset 2:* Yahoo Finance API:**

* Retrieved from the “yfinance” python package
* For each ticker in Dataset 1 (100 total), I downloaded 1-day historical data:
  + Open, high, low, and close (OHLC)

***Additional Dataset to be Used:* Wikipedia S&P 500 List with Sectors**

<https://en.wikipedia.org/wiki/List_of_S%26P_500_companies>

* Will be used to scrape GICS Sector and GICS Sub-Industry classification for each ticker

**Integration**

I cleaned and merged the TradingView and Yahoo Finance datasets using the ticker symbol as a common identifier. The sector and sub-industry information from Wikipedia will also be merged using the same field.

The final dataset will include stock performance metrics, sector information, and a signal column to classify each stock as buy, sell or hold based on price behavior.

**Data Dictionary**

|  |  |  |
| --- | --- | --- |
| **Field** | **Type** | **Description** |
| Ticker | Text | Stock symbol used for merging and analysis |
| Company Name | Text | Full name of the company |
| Stock Price | Numeric | Latest available price from TradingView |
| Price % Change | Numeric | Stock price percentage change (with +/- and % sign) |
| Price % Change (Numeric) | Text | Cleaned stock price % change as float |
| Date | Date | |  | | --- | | Trading day date | |  | |
| Open | Numeric | Opening price on that day |
| High | Numeric | Highest price on that day |
| Low | Numeric | Lowest price on that day |
| Close | Numeric | Closing price on that day |
| Volume | Numeric | Latest stock volume (converted to number) |
| GICS Sector | Text | Sector classification from Wikipedia |
| GICS Sub-Industry | Text | GICS sub-industry classification from Wikipedia |
| Signal | Text | Classification of stock as Buy, Sell, or Hold |

**Purposed Analysis**

**1. Which S&P 500 stocks are likely a good Buy, Sell, or Hold based on short-term price behavior?**  
I will use a combination of percentage change, opening and closing prices, and the daily highs and lows to develop a simple logic for generating a trading signal for each stock.

**2. Which industry sectors are performing the best and worst in terms of average return and price movement?**  
Using the GICS sector classification, I will group stocks and calculate average returns, volume, and volatility metrics to evaluate sector performance.

**3. Which S&P 500 stocks exhibit the highest and lowest volatility based on daily price range?**  
Volatility will be estimated using the daily high-low range for each stock. Stocks with the widest ranges will be highlighted.

**Conclusion**

To conclude, the final analysis will include a clean and integrated dataset, a set of trading signals, and supporting visualizations. These insights will help identify strong performers and assist investors in making more informed trading decisions**.**