STOCK DASHBOARD

Interactive Dashboard: Visualizing Stock Insights of the Top 15 US and French Companies by Market Cap

20/12/2024

Fares Boudelaa

CONTENTS

1 INTRODUCTION

The stock market, a marketplace where buyers and sellers trade publicly held companies, is often difficult to understand and usually unpredictable. For the average person, the abundance of financial data can be overwhelming, while differing regulations across countries further complicate comparisons between markets. This project aims to bridge these gaps by providing visualizations that showcase the behaviors of the 15 largest American and French companies, both collectively and individually. Significant effort was dedicated to preprocessing the data manually using Pandas and encoding visuals with D3 to ensure clarity and accessibility.

2 DATASET

2.1 Overview

The dataset used for this project was provided by Alpha Vantage. Stocks represent a time series, i.e., a timestamp and a price. Due to their complex behavior, it is often useful to record additional information to extract more details. Both the French (post 2004, data before is hard to find) and American markets operate only during weekdays, meaning that trading securities is not possible during weekends or off hours. Thus, a stock usually has four key features:

- timestamp: the specific time point (e.g., hour, week, month, minute).
- open: the price of the security at the start of a given time period.
- close: the price of the security at the end of the time period (market closure).
- high: the peak price of the security during the time period.
- low: the lowest price of the security during the time period.

A stock also has volume, which represents the number of times a share is traded during a specific time period. It is important to note that the number of shares of a company can change for various reasons. For example, during a stock split (where one share becomes two shares), every share owned by an investor is doubled. For this reason, we cannot rely solely on the closing price as our main metric.

To address this limitation, we introduce the adjusted close, which adjusts the closing price using historical data. For this project, we use monthly adjusted stock prices as defined by Alpha Vantage:

> "This API returns monthly adjusted time series (last trading day of each month, monthly open, monthly high, monthly low, monthly close, monthly adjusted close, monthly volume, and monthly dividend) of the equity specified, covering 20+ years of historical data."

Additional data includes historical unemployment figures and interest rates for both countries as macroeconomic indicators. Furthermore, the dataset includes the number of shares outstanding and the percentage of revenue streams from different regions globally: EMAE, Japan and China, North America, and the rest of the world. Often grouped in this manner, this data was extracted from company websites. Although it is fairly inaccurate as companies are not required to release such information.

2.2 Data Preprocessing

Using Python and the `pandas` library, we first preprocess the data. Given the availability of the adjusted close price, we can compute other adjusted features using appropriate formulas.

To enhance the dataset, we derive additional features such as the **moving average** and **lagged values**. Furthermore, we compute more insightful features, including:

1. **Market Capitalization (Market Cap):**

$$\text{Market Cap}_t = \text{close}_t \times \text{number of shares}$$

   Market Cap provides the total market value of a company's outstanding shares, offering a sense of its size and market presence.

2. **Return:**

$$\text{Return}_t = \frac{\text{close}_t - \text{close}_{t-1}}{\text{close}_{t-1}}$$

   Returns calculate the percentage change in a stock's price, serving as an indicator of its profitability.

These features are more robust as they reduce the impact of seasonality, cyclical market behavior, and noise, thereby providing a more accurate perception of a company's performance.

For broader market-level insights, we employ **indices**:

- An **index** is a statistical measure designed to track the performance of a specific set of assets. There are multiple ways to compute an index depending on the context but in our case, we can just use a simple index with an arbitrary value $\alpha = 1000$

$$\text{Base Value} = \alpha \times \frac{\sum_i MC_i(0) \times \text{close}_i(0)}{\sum_i MC_i(0)}$$

$$\text{Index}(t) = \text{Base Value} \times \frac{\sum_i MC_i(t) \times \text{close}_i(t)}{\sum_i MC_i(t)}$$

Another valuable metric for market-wide analysis is the **Pearson correlation coefficient**, defined as:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

The correlation coefficient evaluates the linear relationship between two stocks:

- A value of -1 indicates an inverse linear relationship (when one stock's price rises, the other falls).
- A value of 0 indicates no linear relationship.
- A value of 1 suggests a direct linear relationship.

In this project, since all stocks rise in price, we restrain the domain of correlation values to $[0, 1]$ for simplicity.

3 VISUALIZATION

The visualization of this data is done in two parts: an individual stock dashboard and a dual dashboard for market comparison. The objective of this design is to have a page focused on the company and another for the entire marketplace.

3.1 Individual Stock Dashboard

3.1.1 • candlestick

The dashboard is comprised of the financial candlestick representation. Unlike the usual candlestick representation that uses statistics (confidence intervals, percentiles, medians, etc.), the financial candlestick provides detailed insights. As seen in figure 1, it uses the all four important features of the candlestick to provide detailed information about the price movement. It can also provide quick insight as seen by the colour, where green it indicates growth and red indicates decrease.

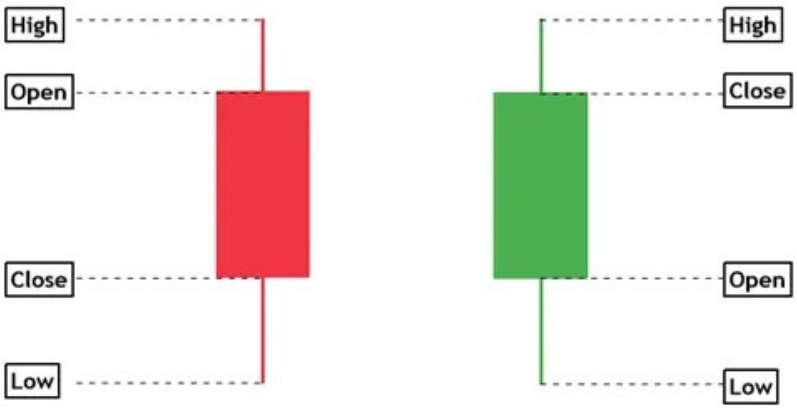

Figure 1: Candlestick

3.1.2 • moving averages

A simple plot of the moving average of the close price shows the long-term performance of the company, as explained earlier. We also add two plots representing the moving averages of the low and high prices, in red and green, respectively. These plots highlight the uncertainty of the stock's price over a specified rolling window, providing insight into the variability and risk associated with the company's performance during that period.

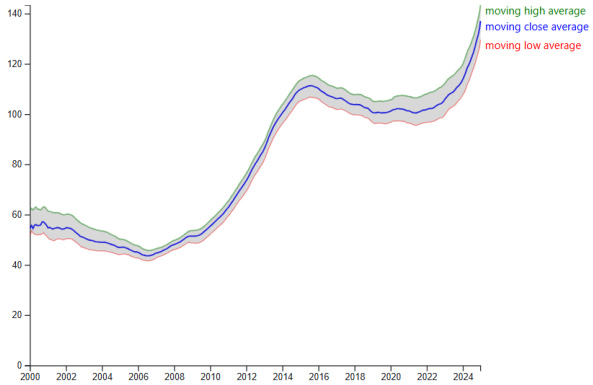

Figure 2: Candlestick

3.1.3 • Histogram of volume per quarter

We also include a histogram of the volume categorized by quarters: Q1 being January, February, and March; Q2 being April, May, and June, and so on. Categorizing volume by quarter provides insight into the trading habits of the market. For example, historically well-performing companies tend to have an equal distribution of volume throughout the year, whereas companies experiencing parabolic price movements display very uneven volume.

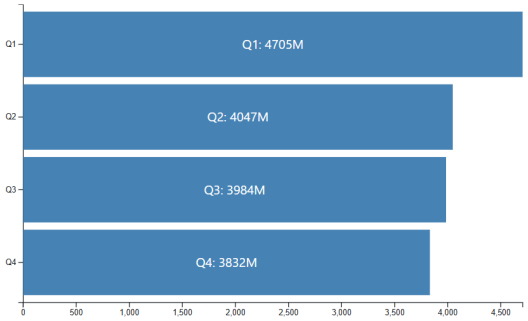

Figure 3: Moving Averages

3.1.4 • Lag plot

Stock prices and returns are usually modelled as autoregressive, meaning the current value depends on previous lagged features. We can visualize this relationship using a scatter plot. This scatter plot includes buttons that allow us to select a lag of 1, 2, or 3 months. In some cases, companies have straightforward plots showing a clear linear relationship due to their slow and steady growth rate. For an investor, this represents a safer investment choice. Other companies exhibit more complex lag plots, revealing different growth periods.

Moreover, a lag plot also indicates the monthly speed of growth, characterized by the slope's angle—a steeper slope indicates faster growth, and vice versa.

Finally, the color of the points represents the time lag: red for one month, orange for two months, and yellow for three months. This intuitive use of color helps the user visually grasp time as an additional dimension. Moreover, the progressively weaker linear relationship between the points highlights the diminishing correlation between the lag features and the close price as the lag increases. A simple yet rich plot.

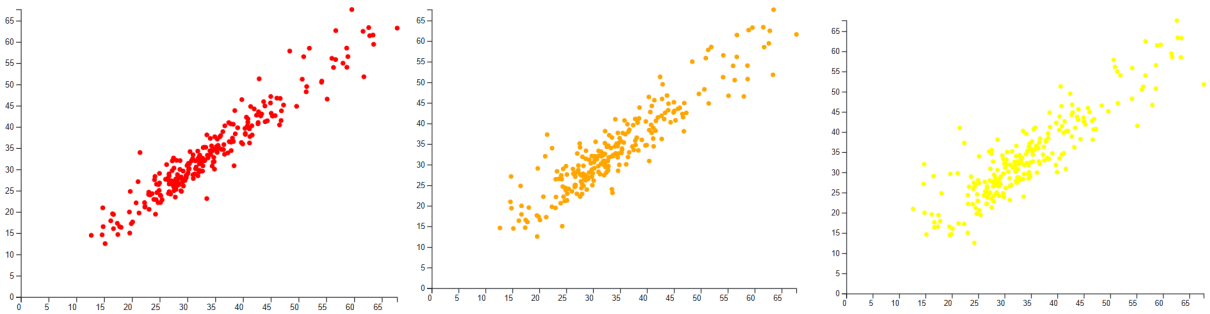

Figure 4: Lag plots

3.1.5 • scatter plot

Finally, we include another scatter plot showing the return of a stock plotted against the change in an economic indicator. This scatter plot attempts to display any relationship, or lack thereof, between the company and the country's economic health. For example, American tech giants are relatively unaffected by unemployment, whereas big banks show a nonlinear relationship between interest rates and returns.

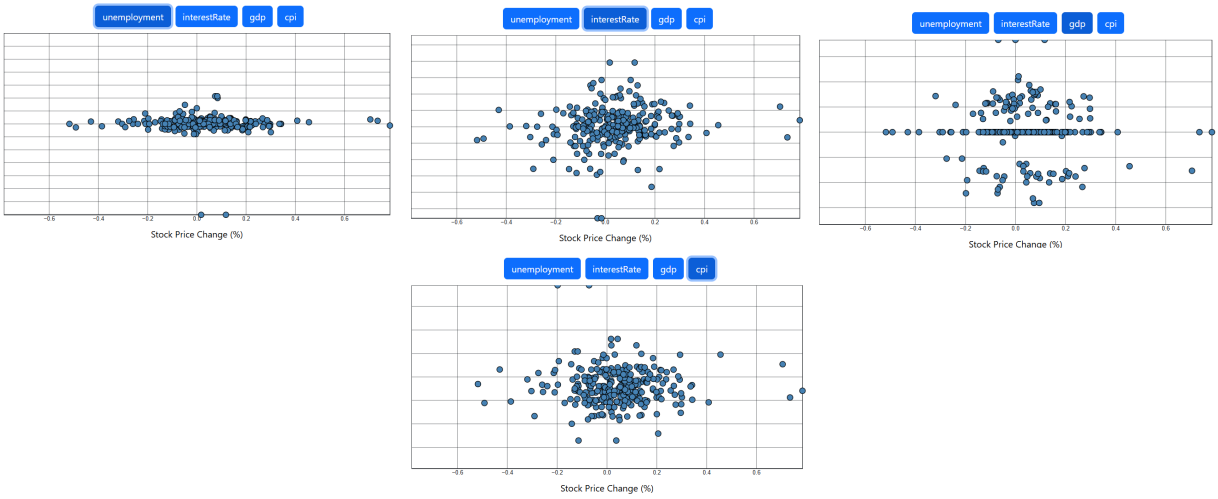

Figure 5: Lag plots

3.2 Dual Dashboard for Market Comparison

3.2.1 • Index Plot

The index plot visualizes the calculated indices of the French and American markets. This is arguably the most important plot of the entire project. It effectively demonstrates the difference between the French and American markets. Starting from the same arbitrary value, the U.S. index grows much more rapidly, while the French index stagnates.

The plot also highlights price stability. While the French index does not grow significantly, it remains largely unaffected by world events. The opposite is true for the U.S. index, which is more volatile and reflects the influence of global market dynamics.

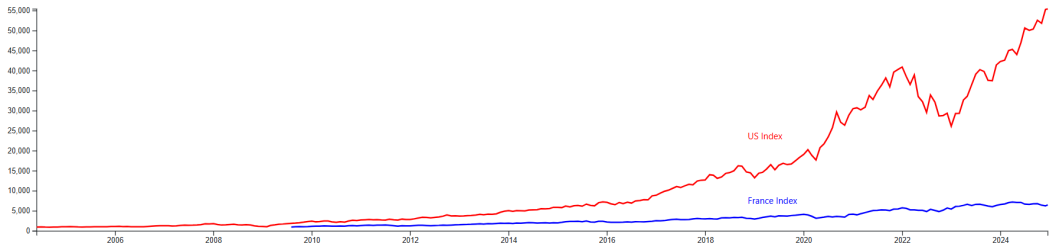

Figure 6: Index Plot

3.2.2 • Correlogram heatmap

The Correlogram heatmap highlights linear relationships between different stocks. It is an effective visualization channel for representing stock correlations due to its ability to condense a large amount of numerical information into an intuitive visual format. Each cell represents the Pearson correlation between the two corresponding stocks, with the intensity of the colour displaying the strength of the relation.

This representation provides a quick and effortless way to visualize the homogeneity of the market's historic movement. It also provides clues as to what companies have stocks that move together. For example, we see that Axa, BNP and Credit Agricole have strong relationships, same with Kering and LVMH, which would make sense since they are in the same markets.

Finally, this also allows for a quick comparison between French and us markets, where US stocks movements are more heterogeneous unlike the French stocks.

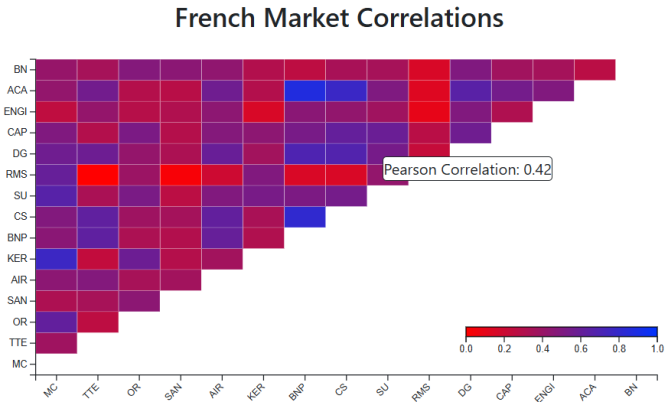

Figure 7: Correlogram heatmap

3.2.3 • Closing Price Matrix

An overview at the end of the page displays plots of the 30 stocks' closing prices. Each market has a series of similar graphs using the same scale and axes, allowing them to be easily compared. This is a simple yet effective visualization of the relative performance and trends of multiple stocks over time, enabling quick identification of outliers, patterns, or market-wide movements.

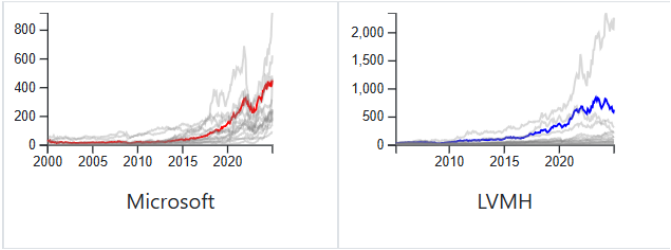

Figure 8: Closing Price Matrix

3.2.4 • Price Change Heatmap

This color-strip displays the monthly percentage change of the closing price. By using color to encode price changes, it simplifies complex data and avoids clutter in multi-stock comparisons. It is an effective way to observe changes in stock prices across many stocks without displaying all their candlesticks. For market-wide insight, this visualization helps users identify or discover time periods where all stocks move in unison, such as the 2008 financial crisis, where the strips have strong red hues.

The color-strip also helps users identify clusters of stocks with similar behaviors. For example, Netflix, Meta, and Tesla share overlapping periods of growth.

Another valuable insight that can be gained is the age of the companies. The largest French companies have been public for nearly 20 years (Engie resulted from a merger of two public companies), while almost half of the American companies are younger than 20 years.

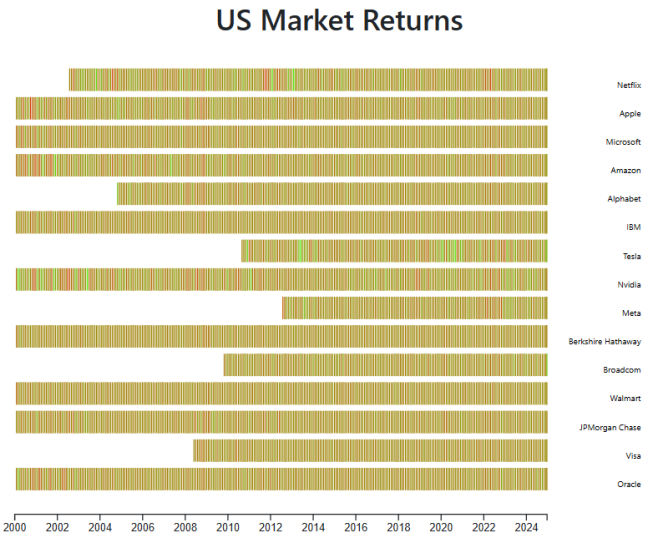

Figure 9: Price Change Heatmap

4
POTENTIAL IMPROVEMENTS AND CONCLUSIONS

4.1 Potential Improvements

There is much more to explore in this project. One particular visual encoding that failed to due many issues is a map encoding for each stock not to mention the lack of available data for french companies.

The project couldve also used prettier color choices and better overall design with more animation to captivate the users.

An encoding channel that was underutilized in the project but retained for its intuitive appeal was the bubble chart. This visualization highlights the stark size difference between French and American companies. For instance, American giants like NVIDIA dominate so significantly that even the smallest among the top American companies can comfortably rank within the top 20 largest in the French market. While the bubble chart effectively conveys these size disparities at a glance, it falls short in providing deeper insights or uncovering nuanced trends within the data. Also there is an issue with adding labels to the bubbles.

4.2 Conclusion

In conclusion, the objective was to demonstrate the different behaviours of the 15 biggest American and French companies and to inform the layman about individual companies. While access to more comprehensive datasets could enhance the visualization further, the current data provides a strong foundation for user engagement and curiosity.