Navigating the Financial Tides: A Year in Stock Market Analysis

by

Pious Khemka - 862467916
Vishesh Shukla - 862464003

MGT 286A: Capstone in Business Analytics
Mr. Hani A M Aldirawi
March 12, 2024

1.Introduction:

This report presents a comprehensive academic inquiry into the financial performance of four prominent companies throughout the year 2023. Employing rigorous statistical techniques and robust data analysis, our study endeavors to address pivotal questions crucial for informed investment decisions. Through meticulous examination, we seek to unravel the intricacies of stock market dynamics, shedding light on key metrics such as stock price evolution, daily return averages, moving averages, inter-stock correlations, and risk assessment.

Furthermore, we aim to extend the boundaries of knowledge by exploring advanced predictive methodologies, particularly focusing on the application of Long Short-Term Memory (LSTM) models to forecast future stock behaviors. As a practical demonstration, we present a predictive model tailored to anticipate the closing price of JPM's stock.

This report is structured to provide a comprehensive overview of our methodology, findings, and implications. It serves as a scholarly contribution to finance, offering insights that can inform academic discourse and guide practical investment strategies. Through our rigorous analysis, we contribute to a deeper understanding of financial markets and empower stakeholders with actionable insights.

2. Research Questions:

Through rigorous statistical analysis and data-driven methodologies, this study seeks to address the following key inquiries regarding the financial performance of four finance companies over the entirety of 2023:

- What are the patterns and trends in the change of stock prices over time for the selected companies?
- What is the average daily return of the stocks, and how does it vary across the companies studied?
- How do moving averages reflect and elucidate the performance trends of the various stocks?
- What is the correlation between different stocks, and how does it impact diversification strategies?
- How can the risk associated with investing in a particular stock be quantified and assessed?
- Can future stock behavior be predicted, and to what extent can Long Short-Term Memory (LSTM) models assist in forecasting the closing price of a stock, using JPM as a case study?

These research questions serve as the guiding framework for our comprehensive analysis, aiming to provide nuanced insights into the financial dynamics of the selected companies and inform investment decision-making processes.

3. Dataset Description

The dataset utilized in this study comprises one year of stock price data for four prominent financial firms, sourced directly from Yahoo Finance through Python's Yahoo Finance library. The dataset has been meticulously collected and cleaned, ensuring its reliability and accuracy for analysis purposes.

Data Source: Yahoo Finance library
Timeframe: One year
Companies Included: Four financial firms.

Data Attributes:

- **Company Name:** Name of the financial firm.
- **Opening Price:** The price at which the stock opened for trading on a given day.
- **Highest Price:** The highest price at which the stock traded during the day.
- **Lowest Price:** The lowest price at which the stock traded during the day.
- **Closing Price:** The price at which the stock closed for trading on a given day.
- **Adjusted Closing Price:** The closing price adjusted for any corporate actions or other factors that may affect comparability.
- **Volume of Trade:** The total number of shares traded during the day.

The dataset encompasses a range of vital information essential for analyzing stock market dynamics, including daily price fluctuations, trading volumes, and adjusted closing prices. This rich dataset will serve as the foundation for our comprehensive analysis, enabling us to explore various aspects of the financial performance of the selected companies over one year.

3. Analysis Methodology

The methodology used in this study is structured to provide a comprehensive understanding of the financial performance of four selected financial firms over one year. Leveraging Python's Yahoo Finance library for data acquisition, the collected dataset was inherently clean, minimizing the need for extensive preprocessing. Our analysis approach comprises the following key steps:

- **Data Acquisition:** Utilizing Python's Yahoo Finance library, one year of historical stock price data for the selected financial firms was directly collected from Yahoo Finance. This dataset encompasses essential attributes such as opening price, highest price, lowest price, closing price, adjusted closing price, and volume of trade.
- **Descriptive Analysis:** The initial phase of analysis involved conducting descriptive statistics to gain insights into the central tendencies, dispersion, and distributions of the key variables. This included calculating measures such as mean, median, standard deviation, and quartiles for each attribute to understand the overall characteristics of the dataset.
- **Exploratory Data Analysis (EDA):** Following descriptive analysis, exploratory data analysis was performed to uncover patterns, trends, and relationships within the data. This involved visualizing the stock price movements over time through time series plots, exploring correlations between different stocks, and identifying any anomalies or outliers that may warrant further investigation.
- **Quantitative Analysis:** To quantify the relationships and trends observed in the exploratory phase, quantitative analysis techniques were applied. This included calculating average daily returns, moving averages, and correlation coefficients between the stock prices of the selected financial firms. Additionally, measures of risk, such as volatility and beta, were computed to assess the level of risk associated with investing in each stock.

- **Predictive Modeling:** As a culminating aspect of the analysis, predictive modeling was employed to forecast future stock behaviors. Long Short-Term Memory (LSTM) models, known for their effectiveness in handling sequential data, were utilized to predict the closing price of a stock. JPM was selected as a case study for this predictive modeling exercise.
- **Validation and Interpretation:** Finally, the results of the analysis were validated and interpreted to derive meaningful insights and implications. The findings were critically evaluated within the context of financial theory and market dynamics, providing valuable guidance for investment decision-making and future research endeavors.

By adhering to this systematic analysis methodology, we aim to provide a robust and insightful examination of the financial performance of the selected companies, offering actionable insights for investors and researchers alike.

Descriptive Statistics:

Following is the Descriptive Statistics of the dataset. The data include 250 records for each company we have selected for our analysis.

Descriptive Statistics:

	Open	High	Low	Close	Adj Close	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	138.073550	139.450020	136.833420	138.205210	135.177616	
std	122.976737	124.235936	121.893996	123.128558	120.184712	
min	25.320000	25.650000	24.959999	25.170000	24.796028	
25%	36.375000	36.960000	36.235001	36.582500	35.345244	
50%	88.019999	88.225000	86.514999	87.710001	85.799608	
75%	200.862503	202.584999	199.562500	200.202503	198.323601	
max	385.570007	387.760010	383.630005	386.410004	383.690948	

	Volume
count	1.000000e+03
mean	2.014586e+07
std	2.170095e+07
min	4.601000e+05
25%	5.641825e+06
50%	1.290320e+07
75%	3.006100e+07
max	2.184033e+08

Figure 1: Descriptive Statistics

Figure 1 provides insights into the behavior of the four stocks. Overall, the mean opening price is $138.07 with a standard deviation of $122.98, indicating significant variability in the opening prices. The range between the minimum and maximum prices is wide, from $25.32 to $385.57, highlighting diverse market performances. Similarly, the mean volume of 20.15 million shares with a standard deviation of 21.70 million suggests varying levels of trading activity across the stocks. The quartiles indicate the distribution of prices and volumes within the dataset, with 25% of observations falling below the 25th percentile and 75% below the 75th percentile. These statistics provide a comprehensive overview of stock behavior, helping investors assess risk and potential returns.

```
Information about the Data:
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 1000 entries, 2023-01-03 to 2023-12-29
Data columns (total 7 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Open          1000 non-null   float64
 1   High          1000 non-null   float64
 2   Low           1000 non-null   float64
 3   Close         1000 non-null   float64
 4   Adj Close     1000 non-null   float64
 5   Volume        1000 non-null   int64
 6   company_name  1000 non-null   object
dtypes: float64(5), int64(1), object(1)
memory usage: 62.5+ KB
None
```

Figure 2: More information on the data

Figure 2 provides more context about the dataset. It says that the data is a time series dataset with 1000 entries, spanning from January 3, 2023, to December 29, 2023. Each entry includes information on the opening, high, low, and closing prices of the stocks, as well as adjusted close prices and trading volume. The data types for the numeric columns are consistent (float64 for prices and int64 for volume), ensuring proper handling of numerical operations. Additionally, there's a categorical column "company_name", which indicates the name of each stock. This structured data format, combined with the time series nature, facilitates further analysis and visualization, enabling insights into the trends and behaviors of the stocks over the specified time.

Change in stock price over time:

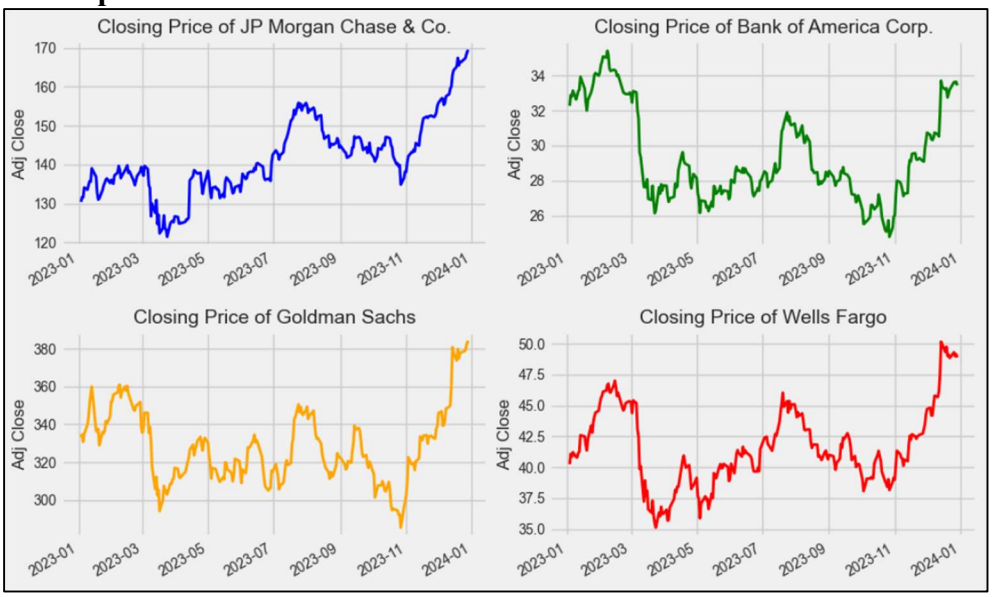

Figure 3: Change in stock price.

Figure 3 is a time series graph showing adjusted closing price of four stocks: J.P. Morgan Chase & Co., Bank of America Corp., Goldman Sachs, and Wells Fargo, over a period ranging from January 2023 to January 2024. The fact that the all stocks show similar dips and highs

around the same time suggests they might be responding to broader market trends or industry-specific events.

Here is one additional thing to consider:

- Not all movements will be identical: Even though the stocks move together to an extent, as they belong to the same sector, they won't necessarily have identical price movements. Each company has its own financial health, prospects, and investor sentiment that can influence its stock price.

For all stocks it can be noticed that the lowest closing price of all time in the year 2023 was recorded in the month of March. Potential Contributing Factors:

- Global Economic Conditions: The year 2023 might have witnessed global economic concerns like inflation, rising interest rates, or geopolitical tensions. These factors can negatively affect stock market performance, potentially leading to widespread price drops.
- Industry-Specific Events: The financial sector, which the four companies belong to, might have been impacted by specific events or regulations in 2023, causing a decline in their stock prices.
- Company-Specific News: Each company might have faced negative news or internal challenges in March 2023, further contributing to the price decrease observed in the graph.

These are just some guesses, and pinpointing the exact reason would require looking at financial news and market data from March 2023.

Volume of Sales:

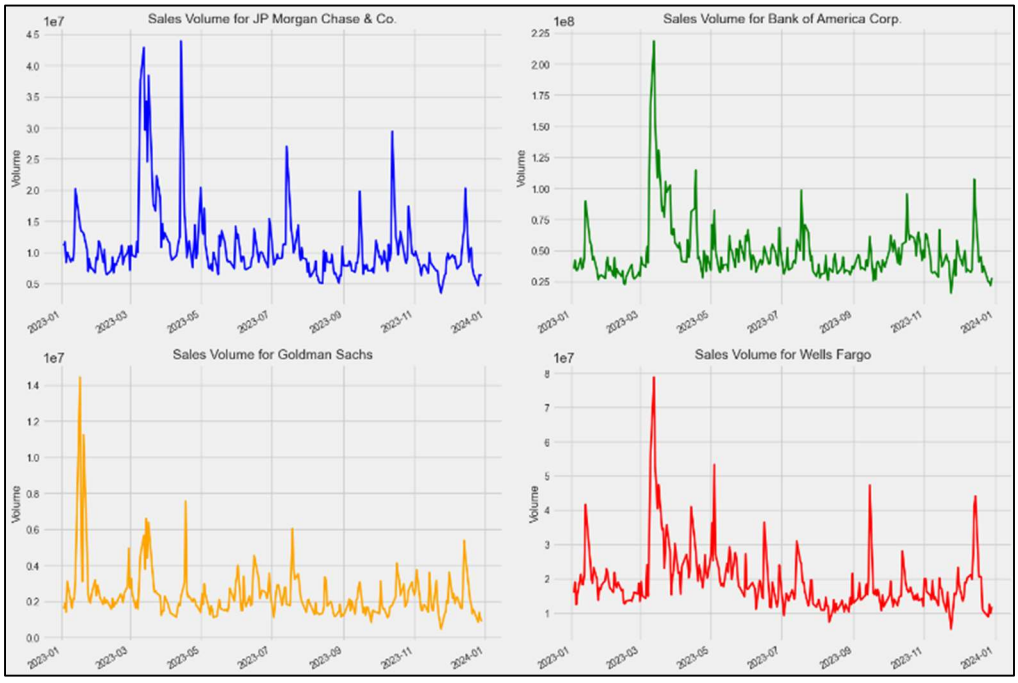

Figure 4: Volume of Sales

In Figure 4, overall, the graph suggests that JPM has the highest and most stable sales volume among the four companies. BAC maintains a steady volume, while GS and WFC exhibit more fluctuations. For all stocks it can be noticed that there was a spike in the sales volume in

the month of March. The reasons could be the same as those driving stock price changes. Any news can influence market sentiment. As a result, traders and investors will engage in many transactions consistent with their plan.

Moving Averages:

A moving average (MA) is a technical analysis tool used to smooth out price fluctuations and identify trends in a stock's price. It essentially calculates the average closing price of a stock over a chosen period.

We have calculated moving averages for several timeframes: 10 days, 20 days, 50 days, 100 days, and 200 days. Each moving average will create a line on your stock chart, and by looking at these lines, you can get clues about the price trend and potential support or resistance levels. Here's a breakdown of what each timeframe might indicate:

- Short-term (10-day and 20-day): These moving averages are more sensitive to recent price changes and can be helpful for identifying short-term trends and potential trading opportunities. However, they can also be quite volatile and generate a lot of noise.
- Mid-term (50-day): This timeframe smoothes out some of the short-term fluctuations but is still responsive to price changes over the past few months. It can be useful for gauging the stock's overall momentum.
- Long-term (100-day and 200-day): These moving averages are less sensitive to short-term price movements and can help identify long-term trends. They can also indicate support and resistance levels, as the price may tend to bounce off these lines.

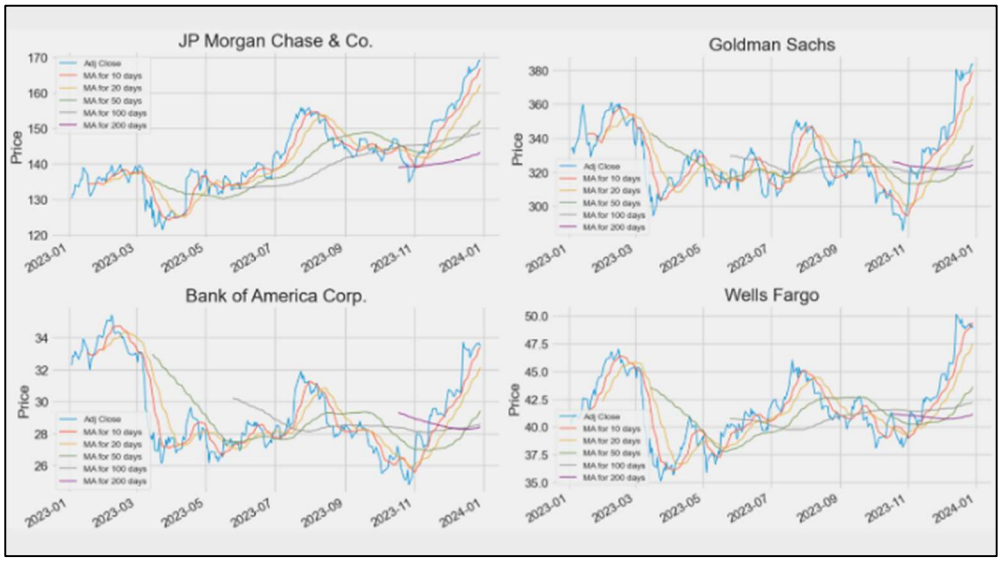

Figure 5: Moving Averages

Figure 5 shows the moving average of all 4 companies. Moving Average – MA of JP Morgan Chase & Co suggest a upward trend, Goman Sachs and Wells Fargo suggest a flat trend with a deviations in between. Whereas Bank of America Corp. suggests a downward trend at the start of January 2024.

Averages Daily Return:

Daily return is calculated by subtracting the opening price from the closing price. If you are calculating for a per-share gain, you simply multiply the result by your share amount.

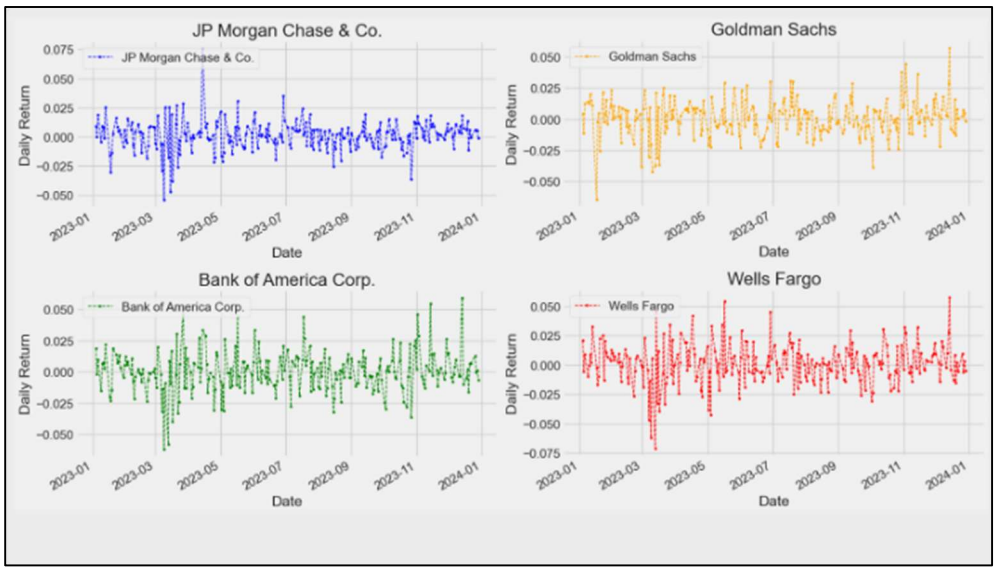

Figure 6: Average Dailly Return

Figure 6 suggests that JP Morgan Chase & Co., Goldman Sachs, and Bank of America Corp. generally had positive daily returns, while Wells Fargo generally had negative daily returns. There could be several reasons why JP Morgan Chase & Co., Goldman Sachs, and Bank of America outperformed Wells Fargo in terms of daily returns during the period shown. Here are some possibilities:

- Company Specific Factors: Business Performance, exposure to Specific Events or recent Scandals.
- Market Sector Performance: Differing Industry Performance.

Daily Return Histogram:

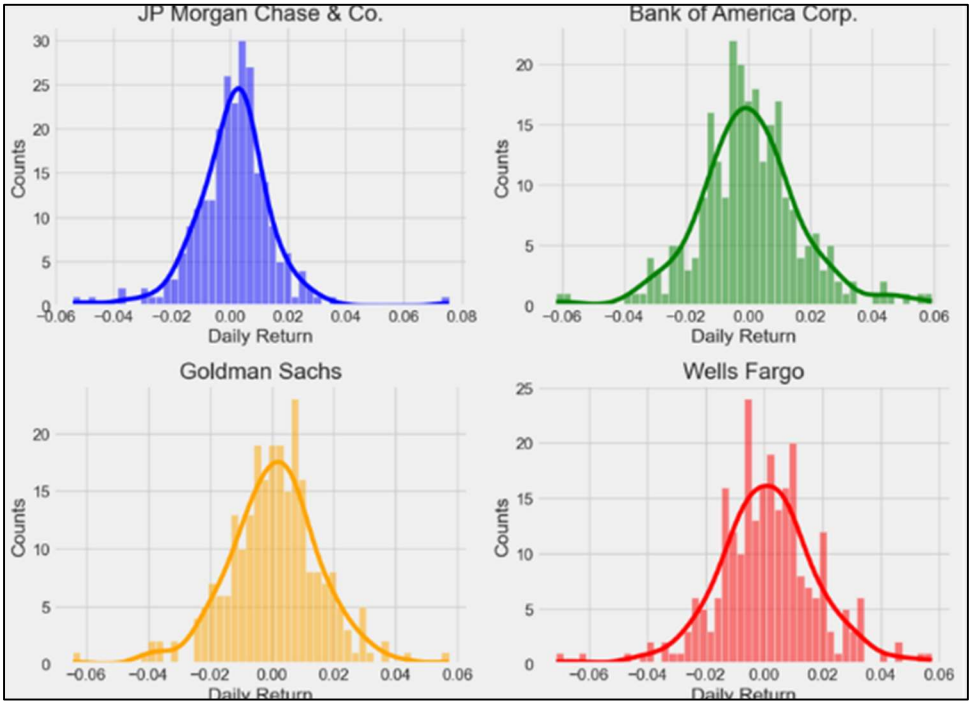

Figure 7: Daily Returns Histogram

In Figure 7 the x-axis shows the daily return, which is the change in price from the previous day divided by the previous day's price. A positive number indicates a price increase, while a negative number indicates a price decrease. The y-axis shows the number of days (frequency) that the daily return fell within a specific range. For example, in the JP Morgan Chase histogram, the tallest bar around 0.02 indicates that there were more days where the daily return was between 0.02 and 0.04 than any other range.

The histograms for all four companies appear to be centered around zero, which suggests that there were more days with small positive or negative returns than days with large positive or negative returns. The histogram for Wells Fargo appears to have a wider spread compared to the other three, which suggests that Wells Fargo's daily returns were more volatile. There were more days with larger positive and negative returns for Wells Fargo compared to JPM, GS, and BAC. The histogram for JPM appears to have the least spread, which suggests that JPM's daily returns were the least volatile during this period.

Correlation of Daily Returns of these Stocks:

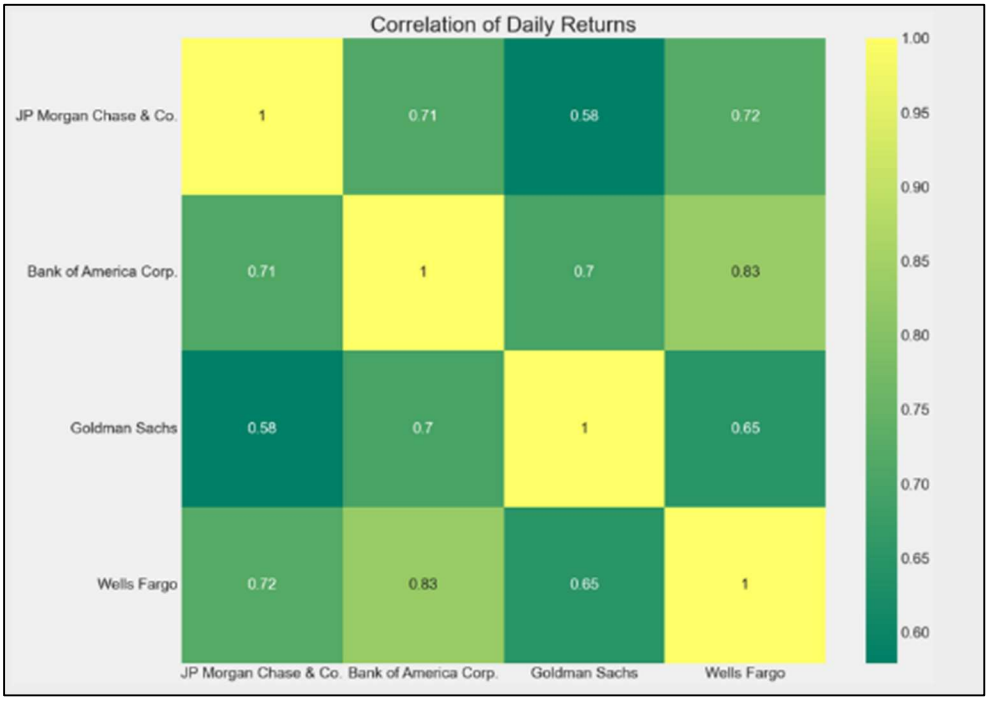

Figure 8: Correlation.

In Figure 8 the correlation coefficient between BAC and WFC is 0.83, which indicates a Strong positive correlation.

The correlation coefficient between JPM and the other three companies (BAC, GS, WFC) is around 0.7, which indicates a strong positive correlation. This means that the daily returns of JPM tended to move in the same direction as the daily returns of the other three companies. The correlation coefficient between BAC and the other three companies (JPM, GS, WFC) is also around 0.7, indicating a strong positive correlation. Like JPM, the daily returns of BAC also tended to move in the same direction as the other three companies.

The correlation coefficient between GS and WFC is around 0.56, which indicates a moderate positive correlation. There is a positive relationship, but it's not quite as strong as the correlations between JPM/BAC and the other companies.

The correlation coefficient between WFC and JPM/BAC is around 0.7, which is a strong positive correlation. However, it's important to note that the correlation is positive, but WFC's returns were generally negative according to the previous graph you sent. This means that while WFC's returns moved in the same direction as JPM and BAC, they moved in the opposite direction in terms of price change (because WFC's returns were negative).

Possible Reasons for the Correlations:

- Market Movements: The stock market can move up or down as a whole, and this can affect the daily returns of all stocks, including JPM, BAC, GS, and WFC.

- Industry Factors: The financial sector can also experience periods of strong performance or weakness, and this can affect the daily returns of all banks (JPM, BAC, WFC).

Risk Vs Return Analysis:

Daily return for a company on a specific day = ((Closing price today - Closing price yesterday) / Closing price yesterday) * 100

The means of the daily returns for each company serve as a measure of expected return. It represents the average daily gain or loss one might anticipate from the stock.

The standard deviation of the daily returns for each company is used as a measure of risk. It indicates how much the actual returns typically deviate from the average, showcasing the volatility of the stock.

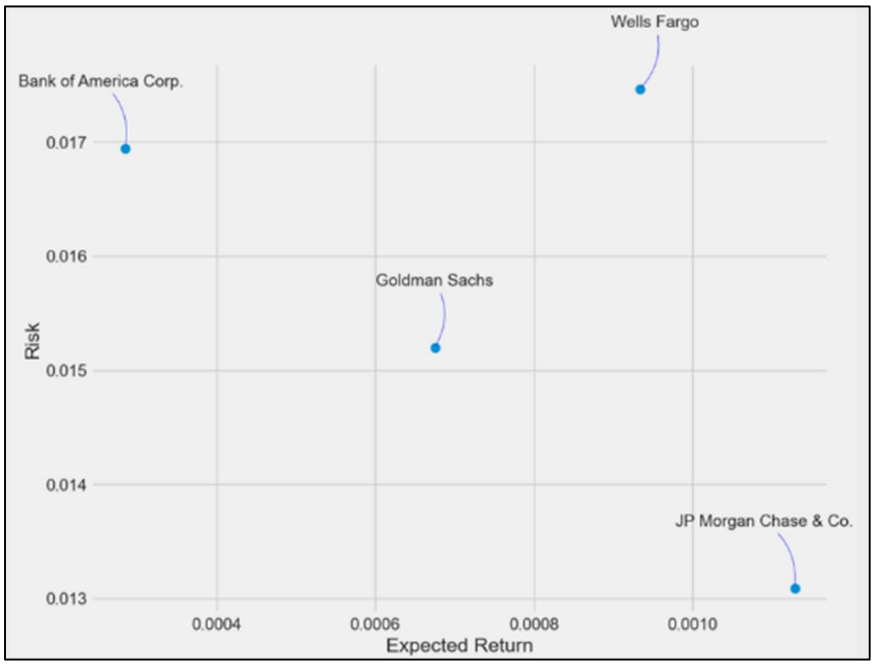

Figure 9 Risk vs Return.

The risk vs expected return plot in Figure 9 shows the expected return plot on the y-axis and risk (measured by standard deviation) on the x-axis for four stocks.

The plot confirms the general principle that investors expect higher returns for taking on greater risk. Stocks like WFC and JPM (located in the right side of the graph) have a higher expected return compared to BAC and WFC (located in the left side of the graph). JPM has the lowest risk followed by GS. Whereas WFC and BAC has high risk as per the data.

Investors seeking higher returns need to consider their risk tolerance. The plot helps visualize this trade-off. For example, if an investor is risk-averse, they might prefer JPM or GS with lower expected returns but also lower risk. Conversely, an investor with a higher risk

tolerance might be drawn to WFC for the potential of higher returns despite the increased risk.

The fact that all four stocks appear relatively close together suggests that they might not offer significant diversification benefits when included together in a portfolio. Since their risk and return profiles seem similar, they might move in tandem to a certain degree.

Stock prediction using LSTM:

LSTM stands for Long Short-Term Memory network. It's a type of recurrent neural network (RNN) architecture specifically designed to handle sequential data, like stock prices.

Regular RNNs struggle with long-term dependencies in sequential data. They can learn short-term patterns well, but as the time gap between relevant data points increases, the RNN may not be able to capture the connection. Imagine predicting tomorrow's stock price. A standard RNN might consider yesterday's closing price but struggle to remember information from weeks or months ago that could also be influential.

LSTMs introduce memory cells that can remember values for longer periods. These cells have gates that control the flow of information:

- Input gate: Decides what new information to store in the cell.
- Forget gate: Decides what information to forget from the cell's previous state.
- Output gate: Decides what information from the cell to use for the current prediction.

By selectively remembering and forgetting information, LSTMs can learn complex relationships between data points spread out over time.

Benefits of LSTMs for Stock Price Prediction:

- Capturing Long-Term Trends: LSTMs can consider historical price movements, economic indicators, and news events from extended periods, potentially leading to more accurate predictions.
- Identifying Patterns: LSTMs can identify complex patterns in historical data that might not be obvious with simpler models. These patterns could be related to seasonal trends, market cycles, or reactions to specific events.

Figure 10: Stock prediction using LSTM.

Figure 10, a graph showing the actual closing price of JPMorgan Chase & Co. (JPM) compared to the predicted prices generated by a Long Short-Term Memory (LSTM) model. The LSTM model seems to capture the general upward trend in JPM's closing for December 2023. The LSTM model was trained using 90% of the records and testing using the remaining 10%.

Model accuracy:

On average, the squared difference between the predicted and actual closing prices was 6.9447 (MSE). Squaring the error magnifies larger differences. Since the typical closing price range is $160-$170, an error of, for example, $5 would result in a squared error of 25. Therefore, an MSE of 6.9447 suggests that on average, the model's predictions deviated from the actual prices by more than $2 (square root of 6.9447 is approximately 2.6).

The average absolute difference between the predicted and actual closing prices was $2.0955 (MAE). This metric directly reflects the average magnitude of the errors. Considering the typical price range of $160-$170, an error of $2.0955 implies that the model's predictions were generally within $2 of the actual closing prices, on average.

6. Conclusion:

In conclusion, the comprehensive analysis conducted on the financial performance of four prominent companies throughout 2023 has illuminated crucial insights into stock market dynamics and investment strategies. Through meticulous examination of price trends, trading volume dynamics, moving averages, and correlations of daily returns, this study has provided investors with valuable tools to navigate the complexities of the financial landscape. Notably, the analysis revealed patterns indicating potential responses to broader market trends or industry-specific events, underscoring the importance of staying attuned to macroeconomic factors. Moreover, the application of advanced predictive techniques, such as Long Short-Term Memory (LSTM) models, showcased promising capabilities in forecasting future stock behaviors, offering investors an additional tool for decision-making and risk management.

Furthermore, the risk vs. return analysis underscored the fundamental trade-off between expected returns and risk, emphasizing the need for investors to carefully balance their risk tolerance with investment objectives. While strong positive correlations between certain stocks were observed, suggesting potential benefits of diversification strategies, investors must remain vigilant of broader market movements that may influence individual stock performances. Overall, this study contributes to a deeper understanding of financial markets, empowering stakeholders with actionable insights to make informed investment decisions and navigate the dynamic landscape of finance effectively.

6. Project member contribution:

Member 1	**Member 2**
Pious Khemka	Vishesh Shukla
Contribution	**Contribution**
• Project background research. • Coding for data analysis in Python. • Drafting interpretations from the analysis. • Report drafting. (Overall workload distribution 50%)	• Assisted in coding for data analysis in Python. • Project presentation preparation. • Drafting interpretations from the analysis. • Report drafting. (Overall workload distribution 50%)