Final Report: Stock Analysis and Visualization

Team 5: Kaichen Bian, Shuhong Cai, Longtao Chen, Siying Chen, Zeyu He

1. Overview

This report provides an in-depth analysis of four prominent companies—Microsoft (MSFT), Walmart (WMT), Johnson & Johnson (JNJ), and Bank of America (BAC)—across various dimensions, including financial performance, risk assessment, and market sensitivity. By leveraging data from Bloomberg, Yahoo Finance, and the Fama-French database, the analysis employs advanced data cleaning techniques and robust financial modeling to offer actionable insights for investors. The database design adheres to rigorous normalization principles, ensuring data integrity and facilitating accurate calculations, while visualizations enhance understanding of key financial metrics and risk profiles.

Key insights reveal that Microsoft outperforms the others in growth and profitability, fueled by its dominance in the tech sector, while Walmart and Johnson & Johnson offer stability for risk-averse investors. Bank of America presents higher volatility, making it more suitable for aggressive investment strategies. Comprehensive risk measures, including Value at Risk (VaR) and Expected Shortfall (ES), highlight significant differences in risk exposure among the selected stocks. Additional analyses, such as the Fama-French 5-Factor Model and Sharpe Ratios, provide deeper insights into stock-specific dynamics and guide tailored investment recommendations.

2. Stocks Picking

We choose four different large companies in different industries:

Walmart (WMT): Walmart is a multinational retail corporation operating a chain of hypermarkets, discount department stores, and grocery stores, known for its low prices and extensive global presence.

Microsoft (MSFT): Microsoft is a leading technology company renowned for its software products, including Windows, Office, and Azure cloud services, as well as its hardware like the Xbox gaming console.

Johnson & Johnson (JNJ): Johnson & Johnson is a global healthcare leader specializing in pharmaceuticals, medical devices, and consumer health products like Band-Aid and Tylenol.

Bank of America (BAC): Bank of America is one of the largest financial institutions in the world, providing banking, investment, and wealth management services to individuals and businesses.

Before conducting any data-driven analyses, we performed a preliminary assessment based on the companies' fundamentals and recent developments.

From a long-term growth and innovation standpoint, Microsoft (MSFT) appears to hold greater promise, given its diversified advancements in cloud computing, artificial intelligence, enterprise software, and gaming ecosystems, as well as its strong financial performance.In comparison, Walmart demonstrates steady growth and defensive stability in the retail sector, though its potential growth trajectory may lag behind that of leading technology firms. Johnson & Johnson exhibits defensive characteristics as a healthcare entity but faces uncertainties stemming from ongoing legal challenges. Bank of America stands to benefit from stable economic conditions and rising interest rates; however, its growth path is relatively even-paced and lacks the robust innovation drivers seen in high-tech sectors.

Thus, considering growth potential and innovation momentum, Microsoft (MSFT) emerges as the company with potentially the strongest long-term outlook among the four.

3. Value Proposition

Our value proposition is designed to empower three primary customer segments in the stock market: investment firms and business managers seeking advanced analytical tools for strategic decision-making, portfolio managers and financial analysts requiring sophisticated solutions to process and interpret complex datasets, and individual investors desiring instant, actionable insights to optimize their trading strategies.

We enhance automation and accuracy with stored procedures that ensure consistent and error-free calculations, providing a reliable foundation for decision-making. For example, investment firms can visualize financial performance trends through dynamic dashboards, enabling them to compare sector growth or risk profiles at a glance. Our solutions also improve decision-making by enabling users to analyze real-time data, historical trends, and predictive models. Portfolio managers can, for instance, access heatmaps of portfolio diversification or charts displaying asset performance correlations, making complex data intuitive and actionable.

Furthermore, we enhance efficiency by removing emotional bias from trading decisions. Using data-driven visualizations, individual investors can monitor live stock price movements, set automated trade triggers based on candlestick patterns, and review visual backtests of strategies, ensuring they capitalize on market opportunities swiftly. By combining robust data analytics with clear, interactive visualizations, we aim to transform raw data into meaningful insights that guide investors toward more confident and precise decisions.

4. Data Preparation

4.1 Initial Data Quality

The dataset for this analysis was sourced from three different providers: Bloomberg, Yahoo Finance, and the Fama French website. Bloomberg supplied the return data for four major stocks: Bank of America (BAC), Johnson & Johnson (JNJ), Microsoft (MSFT), and Walmart (WMT). These return values provide insight into the stocks' performance and volatility over time. Yahoo Finance contributed financial statement data for the same four stocks, including critical financial metrics such as Revenue, Net Income, Earnings Per Share (EPS), Total Assets, and Total Liabilities, which are essential for evaluating each company’s financial health. Lastly, the Fama French website provided key stock market factors such as Mkt-RF, SMB, HML, RMW, CMA, and RF. These factors help to explain the broader market influences on stock returns.

Upon initial review, the dataset presented several quality issues. Missing values were found in critical metrics like Return, Volatility, and various financial metrics, particularly Revenue and Net Income. Outliers were also detected in the financial statement data, especially in the Revenue and Net Income columns, which could potentially distort the analysis. Additionally, there were inconsistencies in data types, with numeric columns such as Return and Financial Ratios mistakenly formatted as text, making calculations and analysis problematic.

4.2 Data Cleaning Process

The data cleaning process involved several key steps:

4.2.1 Removing missing values

The first step in the cleaning process was addressing missing values, which appeared in several critical columns. Missing values were primarily found in the Return and Volatility metrics from Bloomberg and financial data such as Revenue and Net Income from Yahoo Finance. We resolved this issue by removing rows with missing critical data points, ensuring that the dataset remained complete and reliable for analysis.

4.2.2 Handling outliers

Outliers were particularly prevalent in financial metrics such as Revenue and Net Income, as well as in stock return data. To prevent skewed results, we applied the Interquartile Range (IQR) method to identify and remove extreme values. This ensured that the dataset remained representative of typical stock performance and financial metrics, without being overly influenced by anomalies.

4.2.3 Normalizing data types

Finally, data types across the dataset were inconsistent, with some numeric fields being treated as text. This posed a problem for performing accurate financial calculations and analyses. To address this, we normalized all relevant columns—such as Return, Revenue, and EPS—to numeric formats, ensuring consistent data types throughout the dataset.

```python
# FS
# Step 1: Removing missing values
def remove_missing_values(df):
    return df.dropna()

# Step 2: Handling outliers (using IQR method)
def remove_outliers(df, column):
    df.loc[:, column] = pd.to_numeric(df[column], errors='coerce')
    Q1 = df[column].quantile(0.25)
    Q3 = df[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    return df[(df[column] >= lower_bound) & (df[column] <= upper_bound)]

# Step 3: Normalizing data types
def normalize_data_types(df, columns):
    for column in columns:
        df.loc[:, column] = pd.to_numeric(df[column], errors='coerce')
    return df

# Step 1: Clean Financial Statements
cleaned_financial_statements = remove_missing_values(financial_statements_df)

# Step 2: Handle outliers (example for 'Revenue (B)')
cleaned_financial_statements_no_outliers = remove_outliers(cleaned_financial_statements, 'Revenue (B)')

# Step 3: Normalize data types for financial metrics
columns_to_normalize = ['Revenue (B)', 'Net Income (B)', 'Earnings Per Share (EPS)', 'Total Assets (B)', 'Total Liabilities (B)']
cleaned_financial_statements_normalized = normalize_data_types(cleaned_financial_statements_no_outliers, columns_to_normalize)

# Step 4: Handling zero values for 'Capital Expenditures (CapEx) (B)'
def handle_zero_values(df, column):
    # Replace zero values with NaN
    df.loc[df[column] == 0, column] = pd.NA
    # Forward-fill or backward-fill missing values
    df[column] = df[column].fillna(method='ffill').fillna(method='bfill')
    return df
```

Figure 1: Financial Statement

```python
# Fama
# Step 1: Removing missing values
def remove_missing_values(df):
    return df.dropna()

# Step 2: Handling outliers (using IQR method)
def remove_outliers(df, columns):
    for column in columns:
        df.loc[:, column] = pd.to_numeric(df[column], errors='coerce')
        Q1 = df[column].quantile(0.25)
        Q3 = df[column].quantile(0.75)
        IQR = Q3 - Q1
        lower_bound = Q1 - 1.5 * IQR
        upper_bound = Q3 + 1.5 * IQR
        df = df[(df[column] >= lower_bound) & (df[column] <= upper_bound)]
    return df

# Step 3: Normalizing data types
def normalize_data_types(df, columns):
    for column in columns:
        df.loc[:, column] = pd.to_numeric(df[column], errors='coerce')
    return df

# Apply all steps in one process
columns_to_process = ['Mkt-RF', 'SMB', 'HML', 'RMW', 'CMA', 'RF']

# Clean Fama Factors data
cleaned_fama_factors = remove_missing_values(fama_factors_df)
cleaned_fama_factors_no_outliers = remove_outliers(cleaned_fama_factors, columns_to_process)
cleaned_fama_factors_normalized = normalize_data_types(cleaned_fama_factors_no_outliers, columns_to_process)

# Export the cleaned data to Desktop as Excel
output_path = '/Users/longtaochen/Desktop/Cleaned_Fama_Factors.xlsx'  # Replace with your correct path

# Write the cleaned dataset to Excel
cleaned_fama_factors_normalized.to_excel(output_path, sheet_name='Cleaned_Fama_Factors', index=False)

print(f"Fama Factors data has been exported to {output_path}")
```

Figure 2: Fama French

```python
# Stock
# Step 1: Removing missing values
def remove_missing_values(df):
    return df.dropna()

# Step 2: Handling outliers (using IQR method)
def remove_outliers(df, columns):
    for column in columns:
        df.loc[:, column] = pd.to_numeric(df[column], errors='coerce')
        Q1 = df[column].quantile(0.25)
        Q3 = df[column].quantile(0.75)
        IQR = Q3 - Q1
        lower_bound = Q1 - 1.5 * IQR
        upper_bound = Q3 + 1.5 * IQR
        df = df[(df[column] >= lower_bound) & (df[column] <= upper_bound)]
    return df

# Step 3: Normalizing data types
def normalize_data_types(df, columns):
    for column in columns:
        df.loc[:, column] = pd.to_numeric(df[column], errors='coerce')
    return df

# Clean and process each stock sheet using a single set of functions
cleaned_stocks = {}
columns_to_process = ['Close', 'Return']  # Columns to process in the stock data

for stock, df in stock_data.items():
    # Apply all steps for each stock sheet
    cleaned_df = remove_missing_values(df)
    cleaned_df_no_outliers = remove_outliers(cleaned_df, columns_to_process)
    cleaned_df_normalized = normalize_data_types(cleaned_df_no_outliers, columns_to_process)
    cleaned_stocks[stock] = cleaned_df_normalized

# Concatenate cleaned stock data for export
cleaned_stocks_df = pd.concat(cleaned_stocks)
```

Figure 3: Stocks

During the data cleaning process, we gained several insights. For instance, Microsoft (MSFT) and Walmart (WMT) exhibited higher volatility in their returns compared to Bank of America (BAC) and Johnson & Johnson (JNJ). Moreover, Walmart showed substantial growth in Revenue, but its Net Income data contained several outliers, highlighting potential financial fluctuations that required careful attention.

After completing the data cleaning process, the dataset's overall quality improved significantly. All missing values were addressed, and outliers were removed using the IQR method, resulting in a more stable and reliable dataset. The data types were normalized, ensuring that calculations could be performed accurately without errors. With these corrections, the final dataset was deemed sufficient for further financial analysis.

5. Databasen Design

The database design for our stock market analytics project effectively supports complex financial data analysis while maintaining data integrity and optimization.

The EER diagram showcases a logical and comprehensive data model that effectively captures the relationships between different entities in the stock market domain. Key entities include sector_index, sectors, stocks, financial_statement_v2, and various return-related tables (daily, monthly, and yearly return_and_volatility). The relationships between these entities are clearly defined with appropriate relationships.

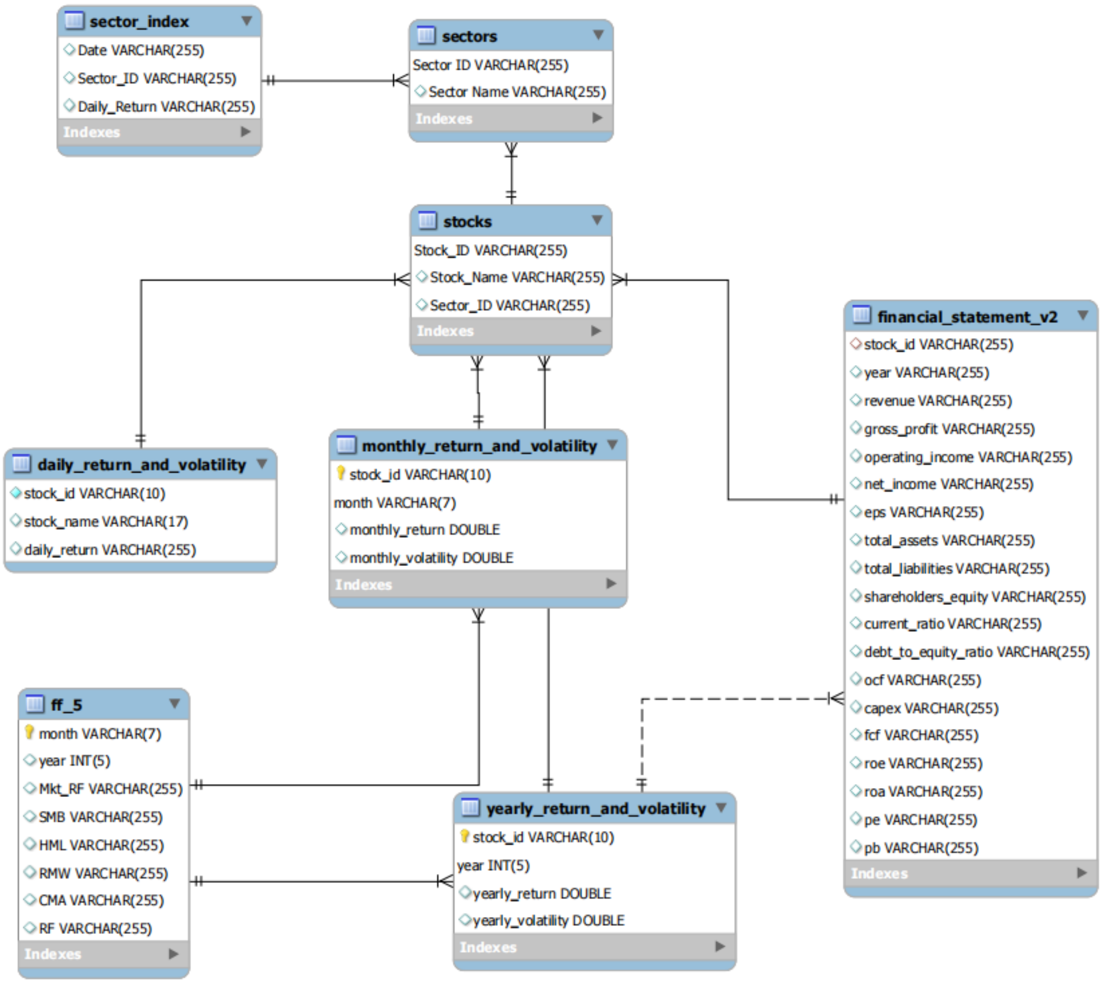

Figure 4: EER Diagram

The implementation strictly adheres to the three normalization forms (1NF, 2NF, and 3NF), ensuring data integrity and minimizing redundancy. First normalization form is achieved by maintaining atomic values across all tables, as evidenced in the stocks table with attributes like stock_id and stock_name. Second normalization form is implemented through proper primary key definitions, with composite keys (stock_id, date) appropriately used for time-series data tables. Third normalization form is maintained by eliminating transitive dependencies, as demonstrated by the separation of sector information into its own table with Sector_id as a foreign key in the stocks table.

Technical competence is further demonstrated through the sophisticated SQL implementations shown in the figure 6-9. The code exhibits proper use of CREATE TABLE, ALTER TABLE, and PRIMARY KEY constraints. The implementation includes thoughtful data cleaning procedures, such as handling NULL dates, and shows mastery of SQL syntax through complex queries involving joins, aggregations, and window functions, as evidenced in the excess return and Sharpe ratio calculations.

```
CREATE TABLE bus211a.bac AS
SELECT
    stock_id,
    date,
    'Bank of America' AS stock_Name,
    `Return` AS `daily_return`
FROM bus211a.bac_origin;
```

Figure 5: BAC Daily Return Table Setup

```
DELETE FROM bac WHERE date IS NULL;
```

Figure 6: Null Data Cleaning

```
ALTER TABLE bac ADD PRIMARY KEY (stock_id, date);
ALTER TABLE jnj ADD PRIMARY KEY (stock_id, date);
ALTER TABLE msft ADD PRIMARY KEY (stock_id, date);
ALTER TABLE wmt ADD PRIMARY KEY (stock_id, date);
ALTER TABLE monthly_return_and_volatility ADD PRIMARY KEY (stock_id, month);
ALTER TABLE yearly_return_and_volatility ADD PRIMARY KEY (stock_id, year);
```

Figure 7: Example of Adding Primary Keys

```
ALTER TABLE daily_return_and_volatility
ADD CONSTRAINT fk_stock_id FOREIGN KEY (stock_id) REFERENCES stocks(stock_id);
ALTER TABLE stocks
ADD CONSTRAINT fk_stock_id_daily FOREIGN KEY (stock_id) REFERENCES stocks(stock_id);
```

Figure 8: Example of Adding Foreign Keys

The design's effectiveness is validated through practical examples, such as the calculation of monthly return and volatility in figure 6, which successfully leverage the database structure to perform complex financial analyses. This demonstrates that the design not only meets theoretical database principles but also serves its intended analytical purposes effectively.

```
-- BAC
CREATE TABLE monthly_return_and_volatility AS
SELECT
    stock_id,
    DATE_FORMAT(date, '%Y-%m') AS month,
    EXP(SUM(LOG(1 + IFNULL(daily_return, 0)))) - 1 AS
monthly_return,
    STDDEV(IFNULL(daily_return, 0)) * SQRT(21) AS
monthly_volatility
FROM
    bus211a.bac
GROUP BY
    stock_id, month;
```

Figure 9: Monthly Return and Volatility Calculation

In conclusion, the database design exhibits professional-grade implementation of database concepts, with clear attention to both theoretical correctness and practical utility in the context of stock market analysis.

6. Visualizations

6.1 Financial Data Analysis

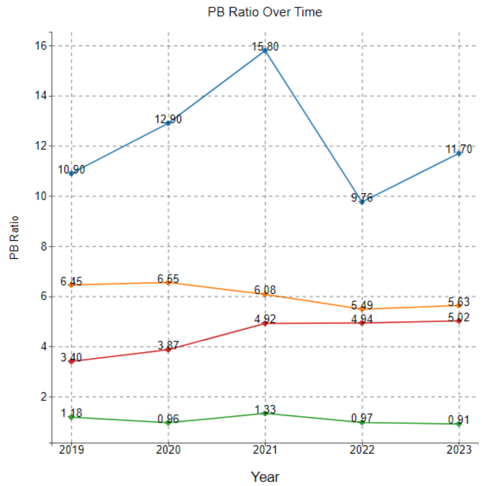

Figure 10: RB Ratio Over Time

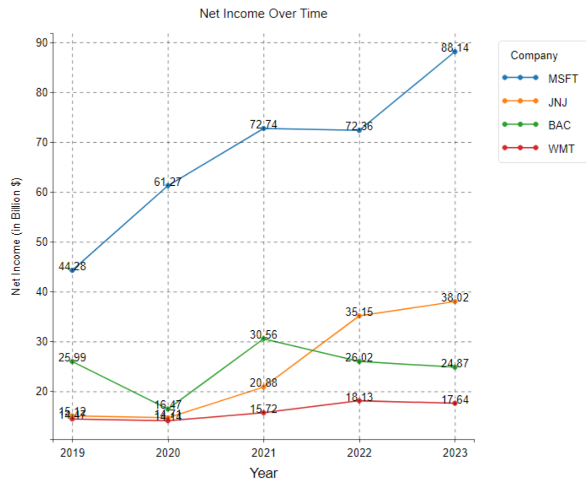

Figure 11: Net Income Over Time

To provide a preliminary assessment of the four selected stocks, we began by examining five-year financial indicators, such as net income, gross profit, ROA, ROE, PE, and PB ratios. For initial visualization, we focused on two metrics. From the net income data, it is clear that Microsoft (MSFT) significantly outperforms the others, reflecting its substantially stronger profitability. Meanwhile, the Price-to-Book (PB) ratio—a measure comparing a company's market value to its book value—can offer contrasting insights. A lower PB ratio may suggest that a company is undervalued or operates in a capital-intensive industry where asset values are high, but it may also indicate financial challenges. Conversely, a higher PB ratio may reflect investor optimism and is common for companies holding substantial intangible assets, such as brand value or advanced technologies, but it may also indicate the stock has been overvalued.

However, directly comparing these metrics across different industries can be misleading, as each sector maintains its own set of typical financial benchmarks. Therefore, to achieve more accurate evaluations and recommendations, more sophisticated analysis is needed, employing advanced financial modeling techniques and in-depth comparative analysis.

6.2 Cumulative Returns Analysis

The cumulative returns of Bank of America (BAC), Microsoft (MSFT), Walmart (WMT), and Johnson & Johnson (JNJ) were analyzed alongside the risk-free rate (RF) for the period from 2019 to 2023. Among the four stocks, MSFT demonstrated the highest growth, achieving a cumulative return of approximately 300% by the end of 2023. This exceptional performance highlights Microsoft's dominance in the tech sector, fueled by its advancements in cloud computing and digital transformation. In contrast, JNJ exhibited steady but modest growth, with cumulative returns reaching about 50%, reflecting its stability and appeal to risk-averse investors. WMT maintained consistent growth, achieving around 70% in cumulative returns, while BAC showed greater fluctuations, ultimately reaching approximately 60% by the end of the period.

The RF curve remained flat, with cumulative returns close to 0, serving as a baseline for comparison. The chart illustrates the substantial performance premium of stocks over the risk-free rate. MSFT's steep upward trajectory highlights its superior return potential, while JNJ's slower but steady incline underscores its reliability. The visualization also reveals that WMT and BAC, while displaying moderate growth, were more susceptible to market dynamics. These trends, supported by key data points, can offer a comprehensive understanding of the diverse risk-return profiles of the selected stocks, aiding investors in aligning their portfolios with their risk tolerance and investment goals.

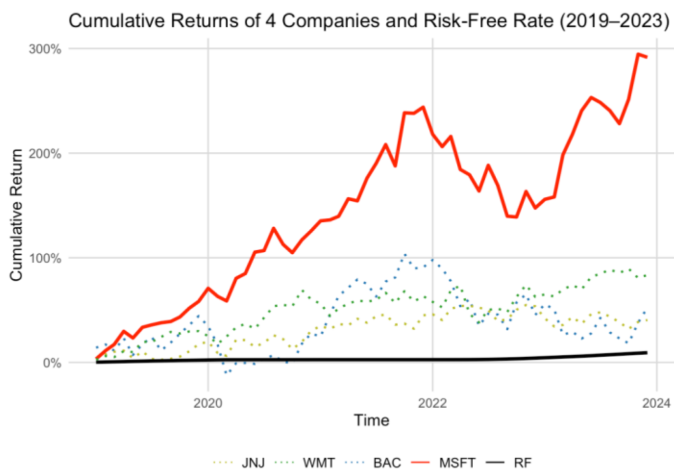

Figure 12: Cumulative Returns of 4 Companies and Risk-free Rate (2019-2023)

6.3 Risk Measures: Value at Risk (VaR) and Expected Shortfall (ES)

The Value at Risk (VaR) analysis using daily returns provides insights into the potential losses for different companies under normal market conditions at 95% and 99% confidence levels. The histograms for BAC, JNJ, MSFT, and WMT illustrate the distribution of daily returns with clear visual indicators of the VaR thresholds. For example, BAC exhibits higher risk levels, with a VaR 99% value of -0.0561, signifying larger potential losses compared to other stocks. Similarly, JNJ's risk is relatively lower, with a VaR 99% of -0.0363. These results are valuable for comparing risk profiles across stocks, aiding in investment decisions.

The graphs also categorize return data based on their proximity to the VaR thresholds, highlighting instances where daily returns fall outside these confidence intervals. The visualizations enable investors to better understand the risk characteristics of each stock, emphasizing how often extreme negative returns occur. Combined with annotations and marked thresholds, the analysis effectively conveys critical risk metrics for practical application in portfolio management and financial decision-making.

```
##
## --- BAC ---
## VaR 95% (Matched Return): -0.032
## VaR 99% (Matched Return): -0.0561
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_bin()`).
```

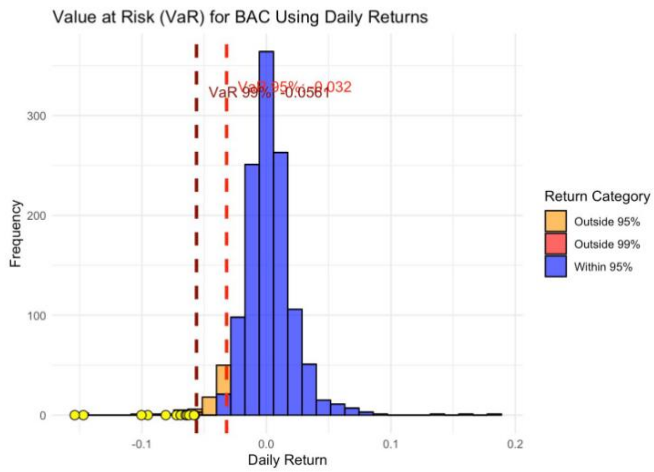

Figure 13: VaR for BAC

```
##
## --- JNJ ---
## VaR 95% (Matched Return): -0.0166
## VaR 99% (Matched Return): -0.0363
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_bin()`).
```

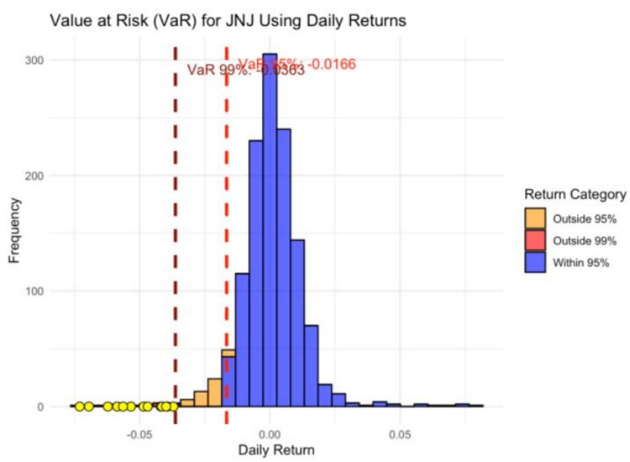

Figure 14: VaR for JNJ

```
##
## --- MSFT ---
## VaR 95% (Matched Return): -0.0291
## VaR 99% (Matched Return): -0.0454
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_bin()`).
```

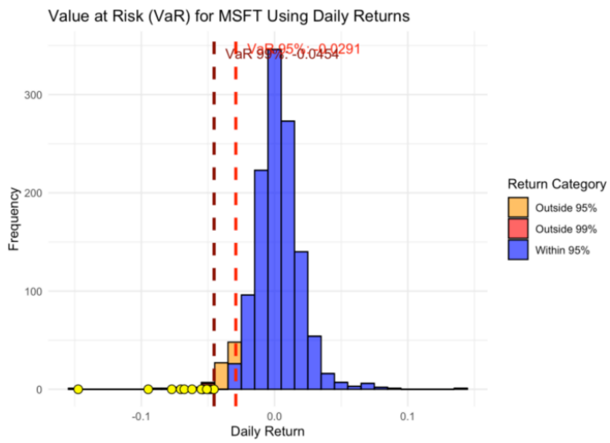

Figure 15: VaR for MSFT

```
##
## --- WMT ---
## VaR 95% (Matched Return): -0.0188
## VaR 99% (Matched Return): -0.0325
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_bin()`).
```

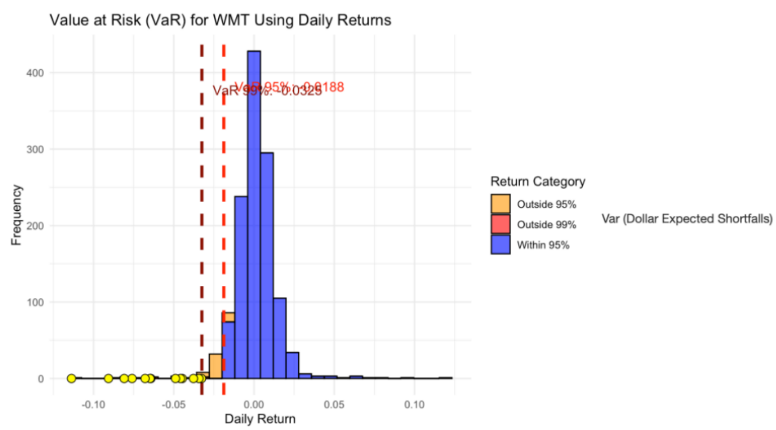

Figure 16: VaR for WMT

The dollar-based Value at Risk (VaR) and Expected Shortfall (ES) analyses for a $10,000 investment provide deeper insights into the financial risk exposure of the selected stocks. At the 95% confidence level, BAC shows the highest dollar VaR at $319.83, while JNJ records the lowest at $166.24, reinforcing JNJ's stability. At the more stringent 99% confidence level, BAC’s dollar VaR rises to $561.46, confirming its higher susceptibility to extreme losses. These results align with the risk profiles observed earlier, with BAC standing out as the riskiest among the stocks analyzed.

The dollar ES, which reflects average losses exceeding the VaR threshold, further highlights the contrast in risk exposure. At the 95% confidence level, BAC's dollar ES reaches $494.59, significantly higher than JNJ’s $291.51, while at the 99% level, BAC’s dollar ES soars to $837.05, compared to $513.75 for JNJ and $617.64 for WMT. The bar plots effectively illustrate these variations, emphasizing BAC's higher exposure to extreme losses and reaffirming JNJ and WMT as lower-risk options. This analysis provides critical insights for investors seeking to balance risk and return when allocating capital.

```
##
## --- BAC ---
## VaR 95% (Potential Loss): $ 319.83
## VaR 99% (Potential Loss): $ 561.46
## ES 95% (Average Loss Beyond VaR): $ 494.59
## ES 99% (Average Loss Beyond VaR): $ 837.05
##
## --- JNJ ---
## VaR 95% (Potential Loss): $ 166.24
## VaR 99% (Potential Loss): $ 362.56
## ES 95% (Average Loss Beyond VaR): $ 291.51
## ES 99% (Average Loss Beyond VaR): $ 513.75
##
## --- MSFT ---
## VaR 95% (Potential Loss): $ 290.83
## VaR 99% (Potential Loss): $ 454.02
## ES 95% (Average Loss Beyond VaR): $ 424.81
## ES 99% (Average Loss Beyond VaR): $ 674
##
## --- WMT ---
## VaR 95% (Potential Loss): $ 188.41
## VaR 99% (Potential Loss): $ 325.23
## ES 95% (Average Loss Beyond VaR): $ 309.86
## ES 99% (Average Loss Beyond VaR): $ 617.64
```

Figure 17: VaR and ES Risk Metrics for 4 Stocks

Figure 18: Dollar ES for 95% Confidence

Figure 19: Dollar ES for 99% Confidence

6.4 Market Sensitivity and Alpha

The regression analysis quantifies the beta of each stock, measuring its sensitivity to market movements (Mkt-RF). Bank of America (BAC) exhibits the highest beta at 1.38, indicating it is highly sensitive to market fluctuations and behaves aggressively in response to market trends. In contrast, Walmart (WMT) has the lowest beta at 0.44, reflecting its defensive nature and limited sensitivity to broader market movements. Johnson & Johnson (JNJ) and Microsoft (MSFT) show intermediate betas of 0.49 and 0.82, respectively, indicating moderate sensitivity to market changes.

In terms of alpha, which represents the stock's performance beyond market-driven returns, Microsoft (MSFT) stands out with a significant positive alpha of 0.0137, highlighting its consistent outperformance relative to its risk exposure. On the other hand, JNJ and WMT show less compelling alpha values, indicating their returns are largely aligned with market or external factors. BAC's negative alpha suggests underperformance relative to its risk, reinforcing its aggressive yet inconsistent behavior. These findings provide valuable insights into the stocks' market behavior, aiding investors in understanding risk-adjusted returns and selecting investments based on their sensitivity and alpha characteristics.

```
##
## --- Regression Results for Stock: BAC ---
##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.192782 -0.040665  0.000118  0.041577  0.138530
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.006135   0.007741  -0.792    0.431
## X            1.376238   0.137269  10.026 2.82e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05862 on 58 degrees of freedom
## Multiple R-squared:  0.6341, Adjusted R-squared:  0.6278
## F-statistic: 100.5 on 1 and 58 DF,  p-value: 2.82e-14
```

Figure 20: Regression Results for BAC

```
## --- Regression Results for Stock: MSFT ---
##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.083869 -0.033906 -0.004688  0.027679  0.117881
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.013701   0.005838   2.347   0.0224 *
## X           0.824570   0.103535   7.964 7.03e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04421 on 58 degrees of freedom
## Multiple R-squared:  0.5224, Adjusted R-squared:  0.5141
## F-statistic: 63.43 on 1 and 58 DF,  p-value: 7.033e-11
```

Figure 21: Regression Results for MSFT

```
## --- Regression Results for Stock: JNJ ---
##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.110777 -0.029752  0.002154  0.029014  0.082125
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.000467   0.005510  -0.085    0.933
## X            0.494727   0.097707   5.063 4.46e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04173 on 58 degrees of freedom
## Multiple R-squared:  0.3065, Adjusted R-squared:  0.2946
## F-statistic: 25.64 on 1 and 58 DF,  p-value: 4.464e-06
```

Figure 22: Regression Results for JNJ

```
## --- Regression Results for Stock: WMT ---
##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.15972 -0.02324  0.00197  0.02729  0.11303
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.004753   0.005988   0.794 0.430591
## X           0.442539   0.106185   4.168 0.000104 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04535 on 58 degrees of freedom
## Multiple R-squared:  0.2305, Adjusted R-squared:  0.2172
## F-statistic: 17.37 on 1 and 58 DF,  p-value: 0.0001038
```

Figure 23: Regression Results for WMT

The excess return analysis compares Bank of America (BAC) and Walmart (WMT) against market excess returns (Mkt−RF) from 2019 to 2023. For BAC, its excess returns exhibit higher volatility compared to market excess returns, with substantial spikes and dips, reflecting BAC's sensitivity to market fluctuations. Notably, BAC's excess return shows occasional outperformance over the market, though it also records pronounced underperformance during negative market trends. This pattern reinforces BAC's aggressive nature and significant exposure to broader market changes.

Conversely, Walmart (WMT) displays a more stable excess return pattern relative to the market, underscoring its defensive characteristics. The WMT excess returns generally align closely with market movements but demonstrate fewer extremes in deviation. This consistent performance highlights Walmart's lower risk profile and its attractiveness to risk-averse investors. Overall, the visual comparison of excess returns across both stocks effectively illustrates their respective market sensitivities, aligning with the beta and alpha findings.

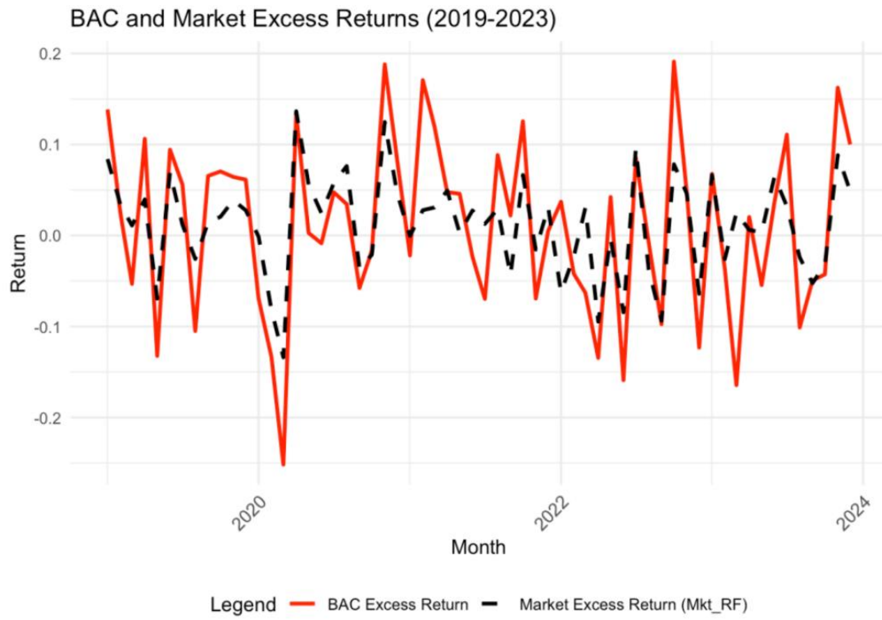

Figure 24: BAC and Market Excess Returns (2019-2023)

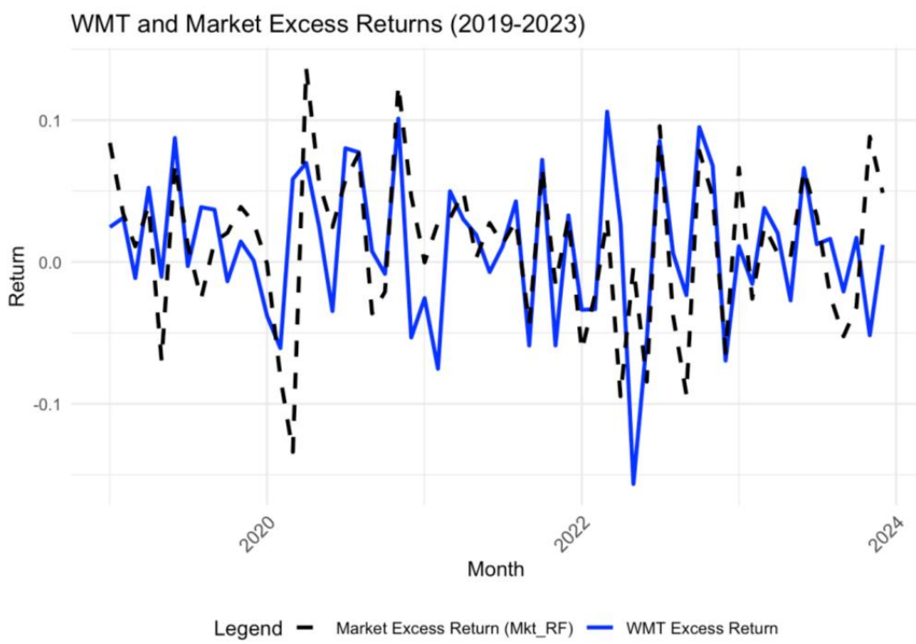

Figure 25: WMT and Market Excess Returns (2019-2023)

6.5 Deeper Factor Analysis: Insights from the Fama-French 5-Factor Model

The Fama-French 5-Factor Model introduces a multifaceted perspective on returns by incorporating additional factors beyond market returns, including size (SMB), value (HML), profitability (RMW), and investment (CMA). In applying this model to Walmart (WMT), the results highlighted the utility of extending beyond the traditional market model. With an adjusted R^2 of 32.76%, the inclusion of additional factors demonstrated a moderate improvement in explaining WMT's returns compared to the simpler beta-based approach.

Among the factors, HML (value) and CMA (conservatism) emerged as significant, indicating that WMT's performance is partially driven by its value orientation and conservative investment strategy. However, the other factors, such as SMB (size) and RMW (profitability), were found to be statistically insignificant. This suggests that the broader factor-based model captures some of WMT's style-driven returns, yet a significant portion of variability remains unexplained, pointing to potential model limitations.

6.6 Recommendations for Enhanced Modeling:

To address these gaps, future iterations of the analysis could benefit from simplifying the model by excluding insignificant factors like SMB and RMW, reducing the risk of overfitting. Additionally, integrating macroeconomic indicators such as interest rates and inflation could provide context for broader economic influences on WMT's performance. Sector-specific metrics like consumer demand and supply chain metrics may also help capture unique dynamics affecting retail. Lastly, the exploration of time-varying coefficients could account for evolving market conditions and refine the model's predictive power. These steps aim to align the model more closely with WMT’s unique characteristics and the external environment.

```
## Call:
## lm(formula = excess_return ~ Mkt_RF + SMB + HML + RMW + CMA,
##     data = wmt_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.14802 -0.01553 -0.00024  0.02894  0.06257
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0008123  0.0057387   0.142    0.888
## Mkt_RF       0.5619282  0.1139994   4.929 8.23e-06 ***
## SMB         -0.2533236  0.2485259  -1.019    0.313
## HML         -0.3867582  0.1847434  -2.093    0.041 *
## RMW          0.2551931  0.2641824   0.966    0.338
## CMA          0.5293996  0.2628644   2.014    0.049 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04203 on 54 degrees of freedom
## Multiple R-squared:  0.3845, Adjusted R-squared:  0.3276
## F-statistic: 6.748 on 5 and 54 DF,  p-value: 5.961e-05
```

Figure 26: Fama-French Regression Results for WMT

6.7 Sharpe Ratio Analysis and Key Comparisons

The Sharpe Ratio provides a valuable metric for comparing the risk-adjusted returns of different stocks. Among the analyzed companies, MSFT stands out with the highest Sharpe Ratio, signifying its superior ability to deliver returns relative to its risk. WMT and JNJ exhibit moderate Sharpe Ratios, making them appealing to risk-averse investors seeking stable performance. Complementing this analysis, other metrics provide further insights: MSFT has a significant positive alpha, indicating consistent outperformance, while BAC's high beta reflects its high market sensitivity. In contrast, WMT and JNJ, with lower betas, demonstrate defensive characteristics, ideal for cautious investment strategies.

6.8 Recommendations for Investors：

Based on the comprehensive comparison, MSFT emerges as the optimal choice for investors aiming for high returns coupled with strong risk-adjusted performance, supported by its highest cumulative return, positive alpha, and favorable news sentiment. On the other hand, WMT is recommended for risk-averse investors, given its defensive attributes, moderate Sharpe Ratio, and relatively low VaR and ES values, signaling lower exposure to extreme losses. JNJ offers similar defensive traits but lags in cumulative returns, making WMT the more balanced option for conservative investors.

Comprehensive Comparison Table

Metric	BAC	MSFT	WMT	JNJ
News Sentiment	Positive	Positive	Neutral	Neutral
Financial Statements	Moderate	Strong	Moderate	Strong
Sharpe Ratio	0.4644	1.2525	0.7355	0.4508
Sharpe Ratio Level	Low	**High**	Moderate	Low
Beta	1.38	0.82	0.44	0.49
Alpha	Insignificant	**Significant**	Insignificant	Insignificant
VaR (99%)	$561.46	$454.02	$325.23	$362.56
ES (99%)	$837.05	$674.00	$617.64	$513.75
Cumulative Return	High	**Highest**	Moderate	Low

6.9 Visualizations Conclusion

In summary, MSFT demonstrates the best overall investment potential, offering a strong balance of performance, risk-adjusted returns, and significant alpha, making it ideal for growth-oriented investors. For those prioritizing stability and defensive characteristics, WMT stands out with its moderate Sharpe Ratio and alignment with style factors. However, it is important to acknowledge the limitations of this analysis, as historical data may not accurately predict future performance, and the current models do not fully account for the variability in returns for all stocks. To enhance future analysis, incorporating macroeconomic and sector-specific variables, along with dynamic modeling approaches, could provide a more comprehensive understanding of market behavior and evolving investment opportunities.

7. Conclusion

Our detailed analysis of four major companies, Microsoft (MSFT), Walmart (WMT), Johnson & Johnson (JNJ), and Bank of America (BAC), evaluates their investment potential based on financial performance, risk metrics, and market sensitivity. The analysis employed rigorous data preparation, addressing missing values, outliers, and data type inconsistencies to ensure reliable results. Key financial metrics, including net income, ROA, and PE ratios, were analyzed alongside advanced risk measures like Value at Risk (VaR) and Expected Shortfall (ES). Visualization of cumulative returns over five years revealed Microsoft's exceptional growth (300%) compared to the steady but moderate growth of Walmart and Johnson & Johnson, while Bank of America displayed high volatility.

Additional analyses, such as regression for beta and alpha values and the Fama-French 5-Factor Model, provide insights into market sensitivity and stock-specific drivers. Microsoft emerged as the top choice for growth-oriented investors, supported by its strong risk-adjusted returns and positive alpha. Walmart and Johnson & Johnson appealed to risk-averse investors with their stability and lower exposure to extreme losses, while Bank of America attracted aggressive investors due to its high beta and potential for significant gains. To enhance future analyses, integrating macroeconomic indicators, sector-specific variables, and dynamic modeling approaches is recommended to better capture evolving market conditions and investment opportunities.