Portfolio Diversification using Efficient Frontier

Bachelor Of Technology in

Computer Science And Engineering

By

Nikhil Sri Ram Pulluri (AP22110011164)

Hemanth Kumar Polisetti (AP22110011354)

Under the Guidance of

, **Dr. Tapas Kumar Mishra**

SRM University–AP

Neerukonda, Mangalagiri, Guntur Andhra Pradesh – 522240

April - 2025

Certificate

Date: 29-04-2025

This is to certify that the work in this Project, entitled "**Portfolio Diversification using Efficient Frontier"** has been carried out by **Nikhil Sri Ram Pulluri, and Hemanth Kumar Polisetti** under my supervision. The work is genuine, original, and suitable for submission to the SRM University—AP for the award of a Bachelor of Technology in **the School of Engineering and Sciences**.

Supervisor

Dr. Tapas Kumar Mishra

Assistant Professor,

Computer Science and Engineering.

Acknowledgements

Firstly, I would like to extend our heartfelt gratitude to the head of the organization, Vice Chancellor **Prof Manoj Kumar Arora**, and Dean , whose leadership and support were instrumental in the successful completion of this project. Their guidance and encouragement were invaluable at every step.

Secondly, I would like to thank my faculty mentor, **Dr. Tapas Kumar Mishra** sir for their insights, effective management, and expertise, which greatly contributed to making this software a success. Their motivation and support were crucial in boosting our morale and navigating through challenges.

Lastly, I would like to thank my team members who provided effective management, knowledge, and skills throughout the development process. Their planning and support helped us overcome all hurdles.

Nikhil Sri Ram Pulluri (AP22110011164)

Hemath Kumar Polisetti (AP22110011354)

Date:- 29-04-2025

Abstract

This project investigates strategies for portfolio diversification with the goal of managing financial risk and optimizing investment returns. The study encompasses several key steps: collecting historical stock price data for major companies, including Apple, Tesla, Nvidia, Netflix, and AMD, along with the S&P 500 index, using the Alpha Vantage API; preprocessing this data to handle missing values and prepare it for analysis; performing exploratory data analysis (EDA) through statistical calculations and data visualizations to discover market trends and relationships; assessing the volatility of stocks by calculating standard deviations of daily returns; and exploring optimal portfolio allocations using the Efficient Frontier approach [1][2][4]. Technologies employed include Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn. By integrating these elements, the project aims to leverage both classical Sharpe Ratio maximization [1], modern machine learning techniques for clustering and diversification [2], technical indicators such as Bollinger Bands [3], and empirical portfolio optimization methods [4] to provide insights that investors can use to construct well-diversified portfolios.

Table of Contents

List of Figures

List of Formulae

Introduction

This project addresses the critical need for effective portfolio diversification strategies in today's complex financial markets, where investors face heightened volatility and uncertainty [1][2]. This project will test different portfolios on the analysis for risk management [1][2]. The goal is to analyze historical stock data and provide investors with a robust, data-driven framework for building resilient portfolios that optimize returns while minimizing risk exposure [1][2].

Building upon the foundation laid by Modern Portfolio Theory (MPT) and Sharpe Ratio maximization [1], the project leverages data-driven insights obtained through a combination of technical analysis, statistical modeling, and machine learning techniques [2]. Historical stock price data is sourced using the Alpha Vantage API for a set of major companies, including Apple (AAPL), Tesla (TSLA), Nvidia (NVDA), Netflix (NFLX), and AMD, along with the S&P 500 index (SPY) [2].

To ensure robust and reliable analysis, the collected data undergoes rigorous preprocessing, including the handling of missing values and extraction of adjusted closing prices [2]. This is followed by comprehensive exploratory data analysis (EDA) to identify patterns, trends, and outliers in stock prices [2]. Furthermore, the project explores risk assessment and portfolio optimizations that leverage the Efficient Frontier for asset allocation [1][2].

In addition, technical indicators such as Bollinger Bands are incorporated to enhance market timing strategies [3]. By analyzing price volatility relative to moving averages, Bollinger Bands provide insight into potential overbought or oversold conditions [3]. This indicator plays a pivotal role in the project's decision-making framework, helping to determine whether to hold, sell, or buy new stocks based on observed price behavior [3].

The integration of machine learning is mentioned as a key extension, allowing for a more refined assessment of non-linear relationships and improved predictive capabilities [2]. By providing insights into individual stock volatility, correlations, and optimal diversification strategies, this project aims to empower investors to make informed decisions that align with their risk tolerance and investment objectives [1][2]. The tools and technologies used to achieve the goals are

Scikit-learn (for machine learning models), Pandas, NumPy, Matplotlib, and Seaborn (for visualization) [2].

As recent studies suggest the need to leverage alternative methodologies alongside classical approaches[4], this project will analyze market dynamics and demonstrate AI-driven portfolio optimization to build resilient and personalized investment solutions [2].

Technologies Used

- Pandas, NumPy
- Matplotlib, Seaborn (visualization)
- Scikit-learn (machine learning models)

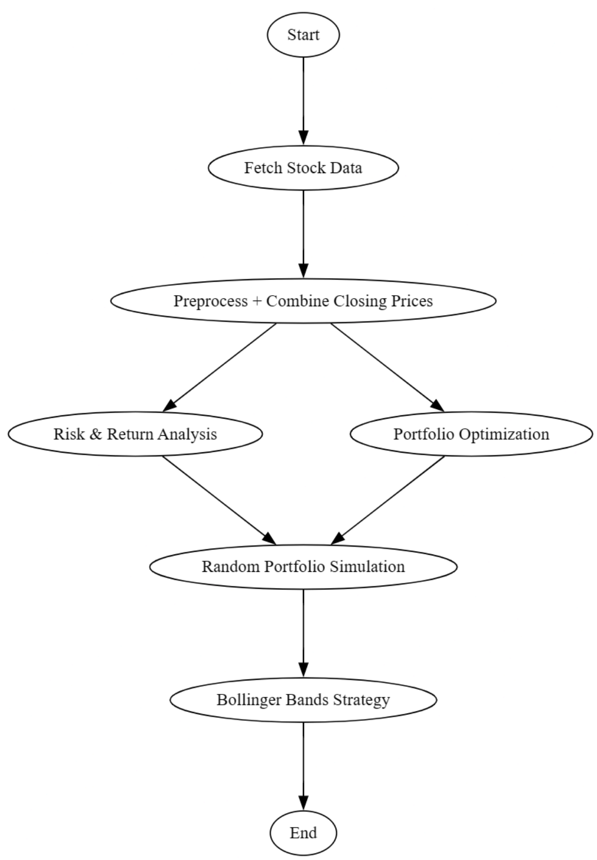

Figure - 1 | *Flow Chart of the Approach*

Literature Review

Recent developments in portfolio optimization and financial data analysis reflect a growing emphasis on integrating classical financial theory with advanced data-driven and algorithmic methods. The foundational principles of Modern Portfolio Theory (MPT), particularly mean-variance optimization and the maximization of the Sharpe ratio, remain central to the construction of efficient portfolios. Maller et al. (2010) provide a rigorous statistical framework for understanding the distribution and estimation of the maximum Sharpe ratio, highlighting both the practical challenges and theoretical implications of relying on sample-based estimates in real-world portfolio selection [1]. Their work underscores the importance of considering estimation error and the potential divergence between the theoretically optimal and empirically achievable Sharpe ratios, especially when the number of assets or the dimensionality of the problem increases.

Despite the enduring relevance of classical optimization, recent research has increasingly explored the limitations of relying solely on traditional volatility and covariance-based diversification metrics. Kuzheliev et al. (2025) demonstrate that while classical approaches such as Markowitz optimization and Sharpe ratio maximization often yield superior risk-adjusted returns, the integration of machine learning techniques—specifically, cluster analysis—can provide valuable complementary insights [2]. By grouping assets based on multidimensional features including returns, volatility, correlations, skewness, and technical indicators, clustering analysis uncovers hidden patterns and asset groupings that may be overlooked by standard approaches. Their findings suggest that clustering enhances the understanding of asset diversity and supports the development of alternative diversification metrics, such as cluster-based diversification ratios, which capture variance both within and between asset groups. However, their empirical results also indicate that clustering should be viewed as a complement rather than a replacement for classical methods, as the latter remain more effective for direct risk-return optimization.

Technical analysis tools, such as Bollinger Bands, have also gained prominence in both academic and applied finance for their ability to identify market conditions and inform trading decisions. Lento et al. (2007) investigate the informational content of Bollinger Bands, finding that these indicators can provide statistically significant signals for market timing and risk management [3]. The incorporation of such technical indicators into portfolio construction frameworks allows for more dynamic and responsive investment strategies, particularly in volatile or rapidly changing market environments.

Empirical studies have further examined the practical aspects of portfolio construction, including the impact of weighting schemes and optimization constraints. Dye and Groth (2000) compare value-weighted and simple optimization approaches, providing evidence on the trade-offs between simplicity, empirical performance, and the risk of overfitting in portfolio allocation [4]. Their findings reinforce the notion that while sophisticated optimization techniques can offer theoretical advantages, practical considerations such as estimation error, transaction costs, and model robustness must be carefully managed.

In summary, the literature demonstrates a clear trend toward hybridizing classical financial models with modern data-driven and machine learning approaches. While mean-variance optimization and Sharpe ratio maximization continue to serve as foundational tools for portfolio selection and risk management [1], the adoption of clustering analysis and technical indicators enriches the analytical toolkit available to investors [2][3]. Empirical research on weighting methods and real-world constraints further grounds these advances in practical portfolio management [4]. This convergence of theory, empirical analysis, and machine learning reflects the evolving complexity of financial markets and the ongoing search for more robust, diversified, and adaptive investment strategies.

Purpose

The purpose of this project is to analyze stock market data, assess investment risks, and optimize portfolio diversification to maximize returns while minimizing risk. By leveraging historical stock price data, statistical analysis, and financial modeling, the project aims to provide investors with data-driven insights for better decision-making.

The following are the significant purposes of the work:

- **Data-Driven Investment Insights**: Collect and process stock market data to evaluate historical trends and price movements [2].

- **Risk Assessment**: Measure stock volatility and analyze correlations to understand market fluctuations and asset dependencies [2].

- **Portfolio Optimization**: Use financial models, including the Efficient Frontier and Sharpe Ratio maximization, to identify the best asset allocation strategies [1].

- **Diversification Strategy**: Reduce investment risk by recommending an optimal mix of stocks that balance risk and return, incorporating both classical and cluster-based diversification metrics [1][2].

- **Visualizing Market Trends**: Generate correlation heatmaps, return distributions, and time-series plots to help investors understand stock performance and relationships [2].

- **Incorporation of Technical Indicators**: Integrate technical analysis tools such as Bollinger Bands to enhance the timing of investment decisions and further support risk management [3].

- **Empirical Validation of Optimization Methods**: Compare value-weighted and optimized portfolios to assess practical trade-offs between simplicity, empirical performance, and risk of overfitting [4].

By achieving these objectives, this project serves as a comprehensive tool for investors to construct a well-diversified and risk-adjusted portfolio, enhancing financial stability and long-term profitability.

Scope

- Retail Investors – Individuals seeking diversified, risk-optimized portfolios.
- Financial Analysts & Traders – Professionals analyzing stock trends and risk factors.
- Investment Firms – Organizations optimizing portfolios for risk-adjusted returns.
- Students & Researchers – Those studying quantitative finance and market behavior.
- FinTech Startups – Companies building AI-powered investment tools.

Our Contributions

In this project, we combine several well-known financial analysis methods into a unified framework for investments. Although the financial research community is familiar with individual concepts such as the Bollinger Bands, Efficient Frontier, Beta calculation, and Sharpe Ratio, our work stands out for the way we integrate, adapt, and use them in tandem for useful portfolio decision-making.

First, in addition to the S&P 500 index, we gather real-time market data for five significant technological stocks: Tesla, Apple, Nvidia, AMD, and Netflix. We carry out a thorough statistical study that includes evaluation of the Sharpe Ratio for risk-adjusted performance comparison, thorough correlation studies between price levels and returns, and standard risk measurements (volatility, returns distribution).

Next, we use random portfolio simulation based on Dirichlet distributions and the Efficient Frontier approach to optimize portfolio creation. Under realistic, useful limitations, this dual technique enables the discovery of portfolios with the highest Sharpe ratio and those with the lowest volatility.

Our project's distinctive feature is the integration of a dynamic Bollinger Bands investment recommendation system, in which the bands' width is modified according to the user's anticipated holding duration (short-, medium-, or long-term). Standard academic or industry models usually lack this time-horizon-based personalisation of investment strategy.

Overall, by combining horizon-based strategy changes, thorough end-to-end portfolio analysis, and real-world usability into a single workflow, our work not only duplicates but also significantly expands upon current approaches. This gives investors a more sensible, risk-adjusted, and knowledgeable method of managing their portfolios.

Results and Analysis

date	Open	High	Low	Close	Volume
2025-03-26 00:00:00	220.77	224.1	220.08	223.75	34493583.0
2025-03-24 00:00:00	221.0	221.48	218.58	220.73	44299483.0
2025-03-21 00:00:00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2025-03-20 00:00:00	213.99	217.4899	212.22	214.1	48862947.0
2025-03-19 00:00:00	214.22	218.76	213.75	215.24	54385391.0
2025-03-18 00:00:00	214.16	215.15	211.49	212.69	42432426.0
2025-03-17 00:00:00	213.31	215.22	209.97	214.0	48073426.0
2025-03-14 00:00:00	211.25	213.95	209.58	213.49	60107582.0
2025-03-13 00:00:00	215.95	216.8394	208.42	209.68	61368330.0
2025-03-12 00:00:00	220.14	221.75	214.91	216.98	62547467.0
2025-03-11 00:00:00	223.805	225.8399	217.45	220.84	76137410.0
2025-03-10 00:00:00	235.54	236.16	224.22	227.48	71451281.0
2025-03-07 00:00:00	235.105	241.37	234.76	239.07	46273565.0
2025-03-06 00:00:00	234.435	237.86	233.1581	235.33	45170419.0
2025-03-05 00:00:00	235.42	236.55	229.23	235.74	47227643.0
2025-03-04 00:00:00	237.705	240.07	234.68	235.93	53798062.0
2025-03-03 00:00:00	241.79	244.0272	236.112	238.03	47183985.0
2025-02-28 00:00:00	236.95	242.09	230.2	241.84	56833360.0

Figure - 2 | *Fetching of Stock Data of Apple, Nvidia, Tesla, Netflix and AMD from Alpha Vantage API*

```
print(NFLX.isnull().sum(),AMD.isnull().sum(), GSPC.isnull().sum(), sep='\n\n')
```

```
Open      0
High      0
Low       0
Close     0
Volume    0
dtype: int64

Open      0
High      0
Low       0
Close     0
Volume    0
dtype: int64

Open      0
High      0
Low       0
Close     0
Volume    0
dtype: int64
```

Figure - 3 | *Checking for any null values in the dataset*

Verifying that the data is clean and reliable. The data we considered contain no null values. The presence of no missing values in each stock dataset is a positive finding, indicating that we have complete and reliable data for analysis. This allows for accurate assessments of various aspects of the stocks, simplifies data preparation, and ensures more robust and trustworthy studies.

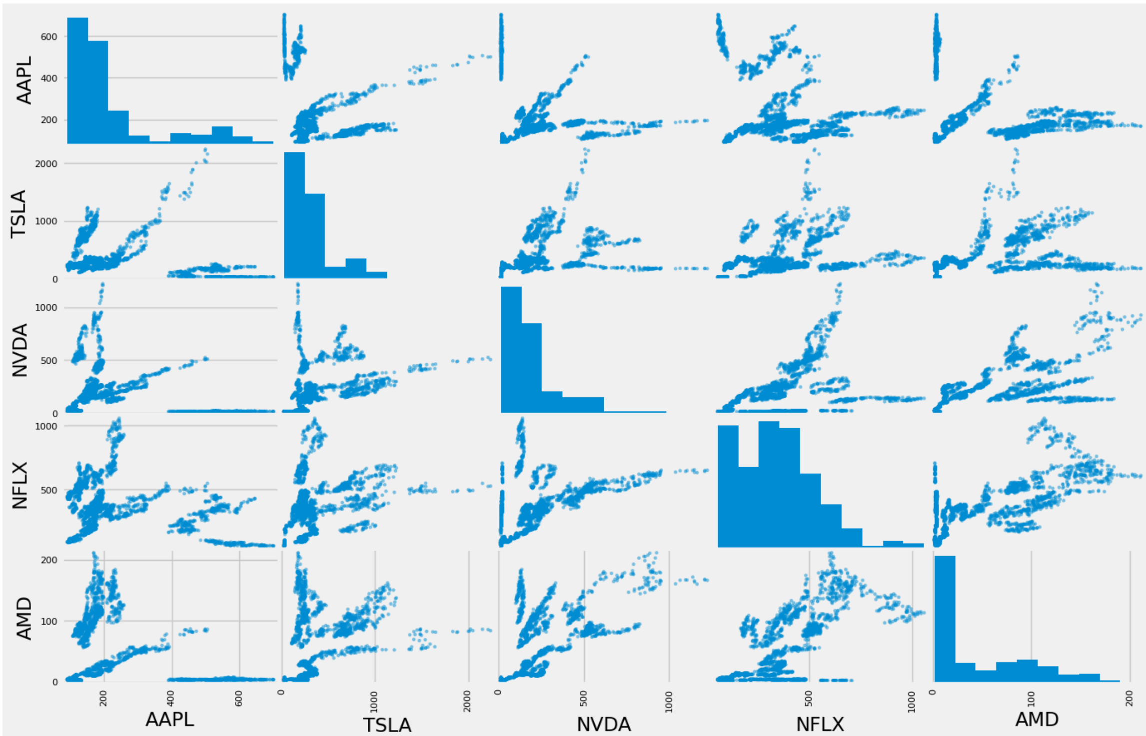

Figure - 4 | *Binning of Stock Data to understand the data distribution*

By dividing stock price values into bins, you get an idea of how frequently prices fall within specific intervals. This helps identify price trends, such as whether stock prices are concentrated within certain ranges or spread out.

Stock Correlation HeatMap

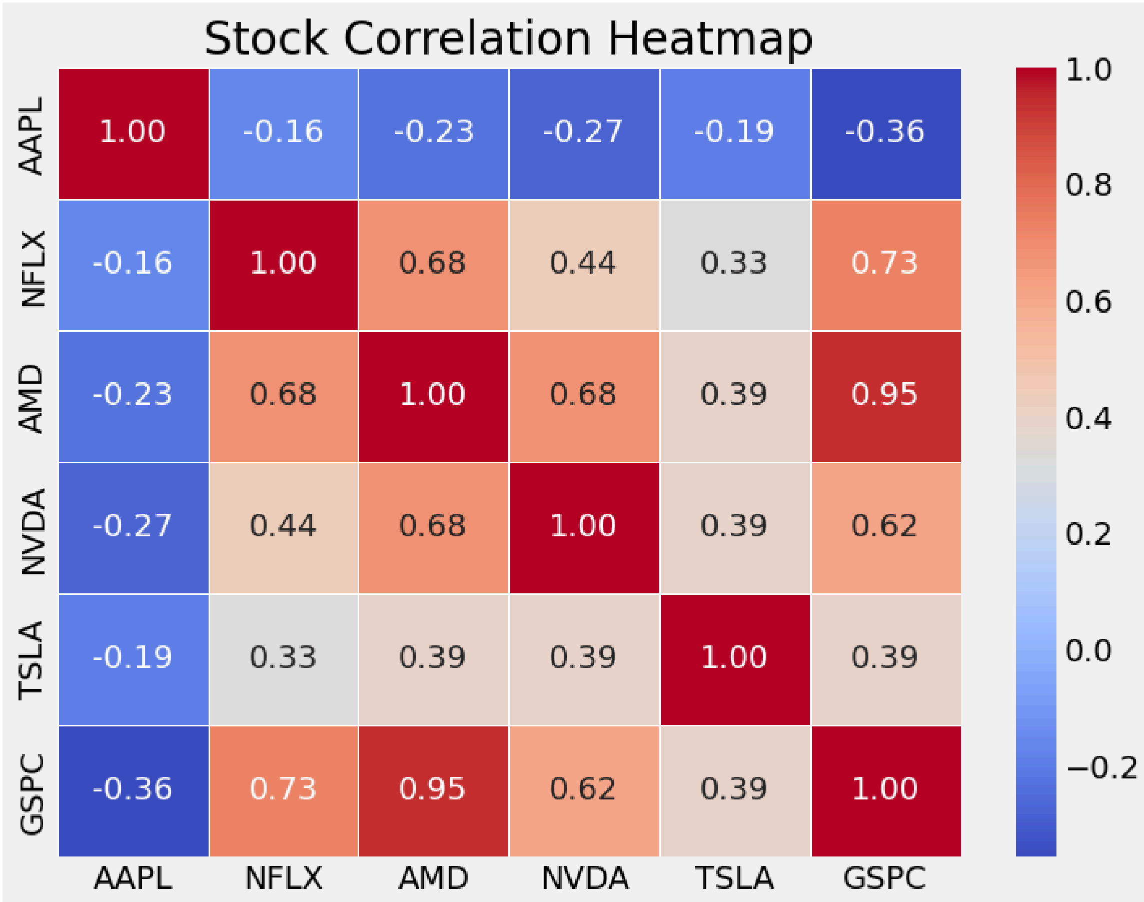

Figure - 5 | *Stock Correlation Heatmap*

This approach computed the relation between the stock closing prices we have considered and the values range from -1 to 1.

Volatility

The formula for Simple Return is a basic calculation that measures the percentage change in the value of an investment over a given period.

$$\beta_i = \frac{Var(R_m)}{Cov(R_i, R_m)} \text{ ....} \textbf{(1)}$$

$\beta_i : Beta\ Security\ of\ asset\ i$

$Cov(R_i, R_m)$: $Covariance\ between\ the\ return\ of\ asset\ i\ (R_i)\ and\ the\ return\ of\ the\ market\ (R_m).$

$Var(R_m)$: $Variance\ of\ the\ return\ of\ the\ market\ (R_m)$

$$R_t = \frac{(P_t - P_{t-1})}{P_{t-1}} \text{ ....} \textbf{(2)}$$

R_t: $Return\ at\ time\ t\ -\ the\ percentage\ change\ in\ the\ asset's\ price\ from\ time\ (t-1)\ to\ t.$

P_t: $Price\ at\ time\ t\ -\ the\ current\ price\ of\ the\ asset.$

P_{t-1}: $Price\ at\ time\ (t-1)\ -\ the\ previous\ price\ of\ the\ asset$

date	AAPL	NFLX	AMD	NVDA	TSLA	GSPC	R_AAPL	R_NFLX	R_AMD	R_NVDA	R_TSLA	R_GSPC
2012-05-01	582.13	81.360	7.59	13.23	33.78	140.74	NaN	NaN	NaN	NaN	NaN	NaN
2012-05-02	585.98	82.230	7.63	12.85	33.94	140.32	0.006614	0.010693	0.005270	-0.028723	0.004737	-0.002984
2012-05-03	581.82	75.970	7.41	12.63	32.46	139.25	-0.007099	-0.076128	-0.028834	-0.017121	-0.043606	-0.007625
2012-05-04	565.25	73.145	7.18	12.26	31.83	137.00	-0.028480	-0.037186	-0.031039	-0.029295	-0.019409	-0.016158
2012-05-07	569.48	73.450	7.18	12.47	32.47	137.10	0.007483	0.004170	0.000000	0.017129	0.020107	0.000730

Figure - 6 | *Daily volatility (in %) of each stock*

The descriptive statistics table provides valuable insights into the risk levels of each stock, as represented by their standard deviations. Among the stocks analyzed, Apple (AAPL) exhibits the highest level of risk with a standard deviation of 0.3447, indicating relatively significant price fluctuations and potential volatility. Following AMD, Tesla (TSLA) shows a lower but still substantial level of risk.

Comparatively, the risk associated with the S&P 500 market is relatively lower, as indicated by its standard deviation of 0.0105. This suggests that, on average, individual stocks such as Apple and NVIDIA carry higher risks than the overall market represented by the S&P 500. These findings highlight the importance of considering risk factors when making investment decisions. Investors may choose to allocate their portfolios based on their risk appetite, taking into account the higher volatility of specific stocks like Apple and NVIDIA compared to the broader market.

Daily Stock Closes – Watchlist

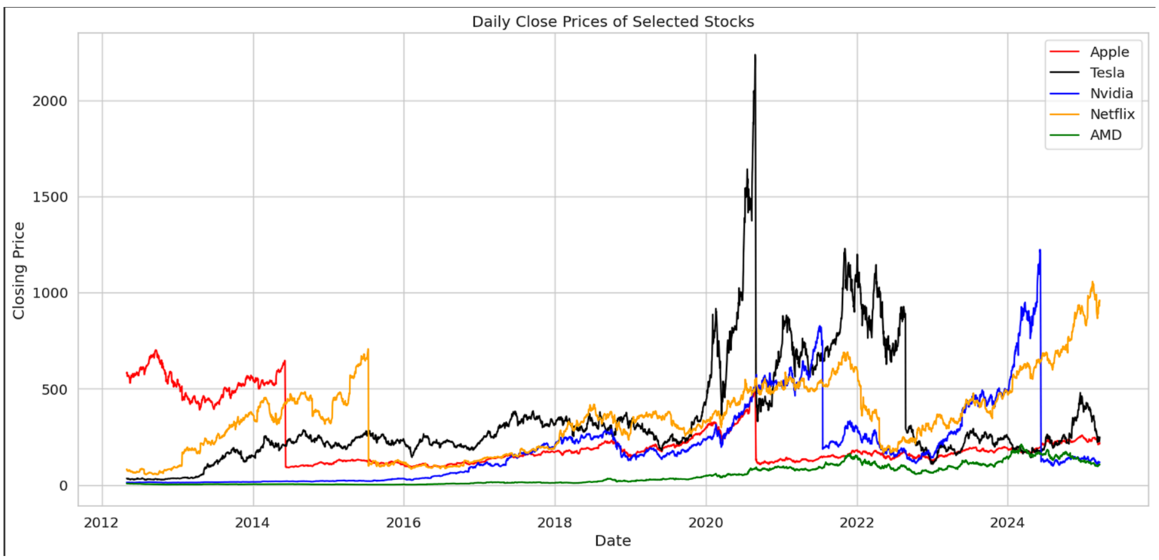

Figure - 7 | *Daily close Prices for Selected Stocks*

Daily Stock Closes – S&P 500 Focus

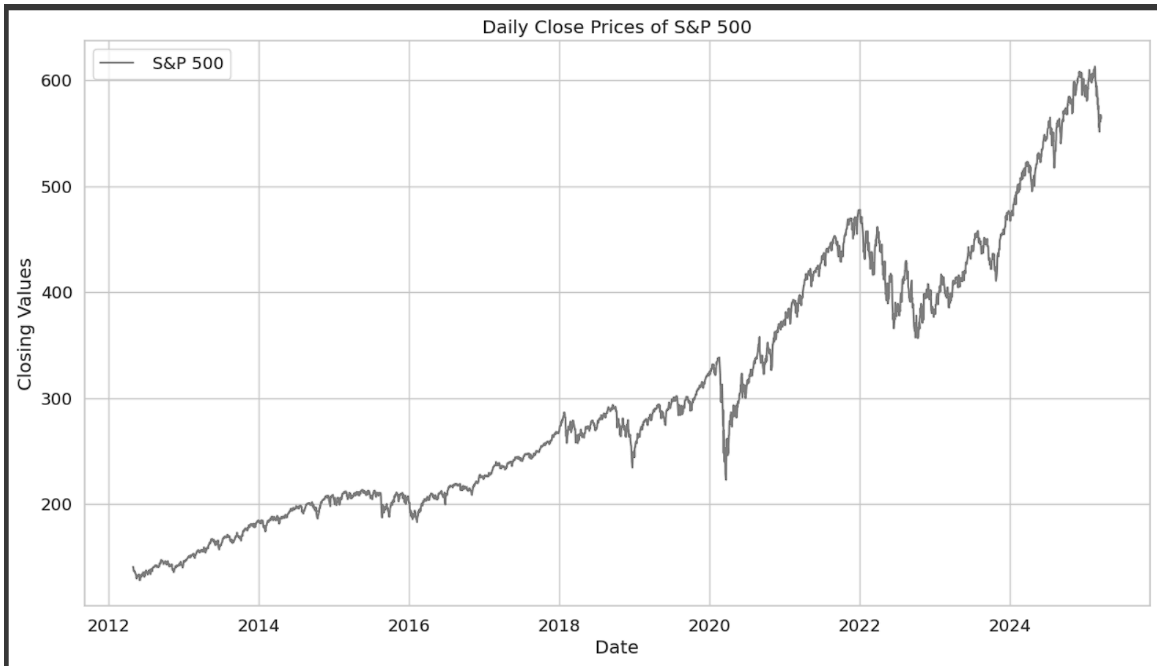

Figure - 8 | *Daily Closing Prices of S&P 500*

Daily Returns of the stock (Time Series Plot and Distribution Plot)

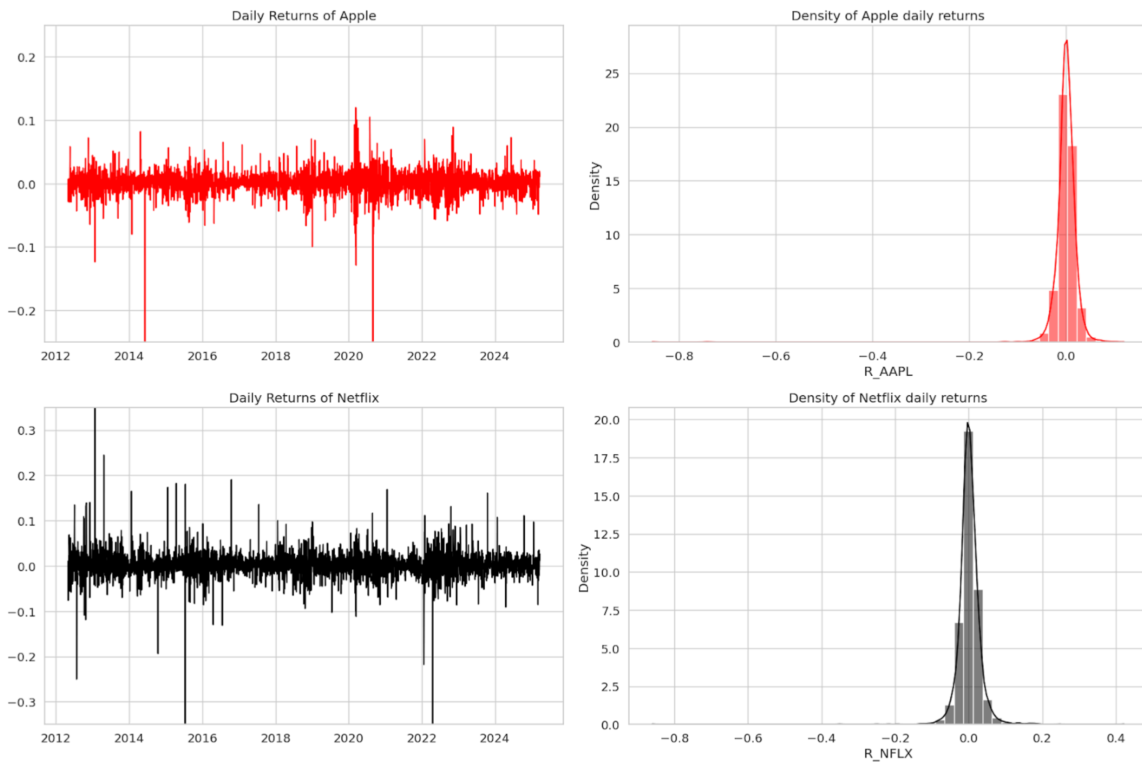

Figure - 9 | *Apple & Netflix*

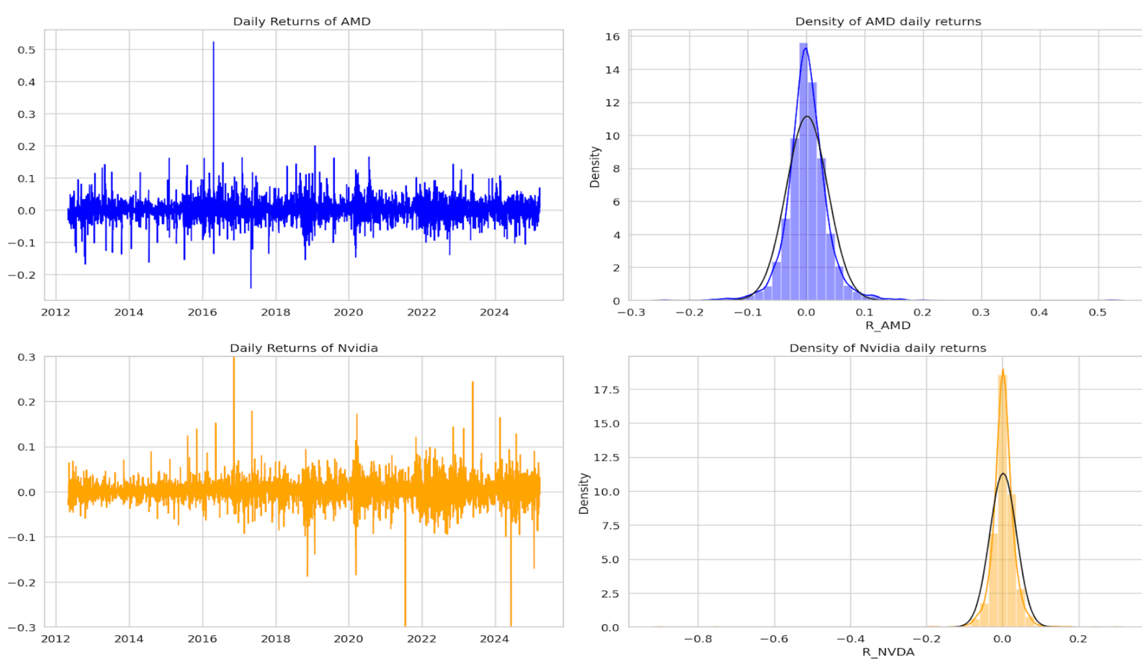

Figure - 10 | *AMD & Nvidia*

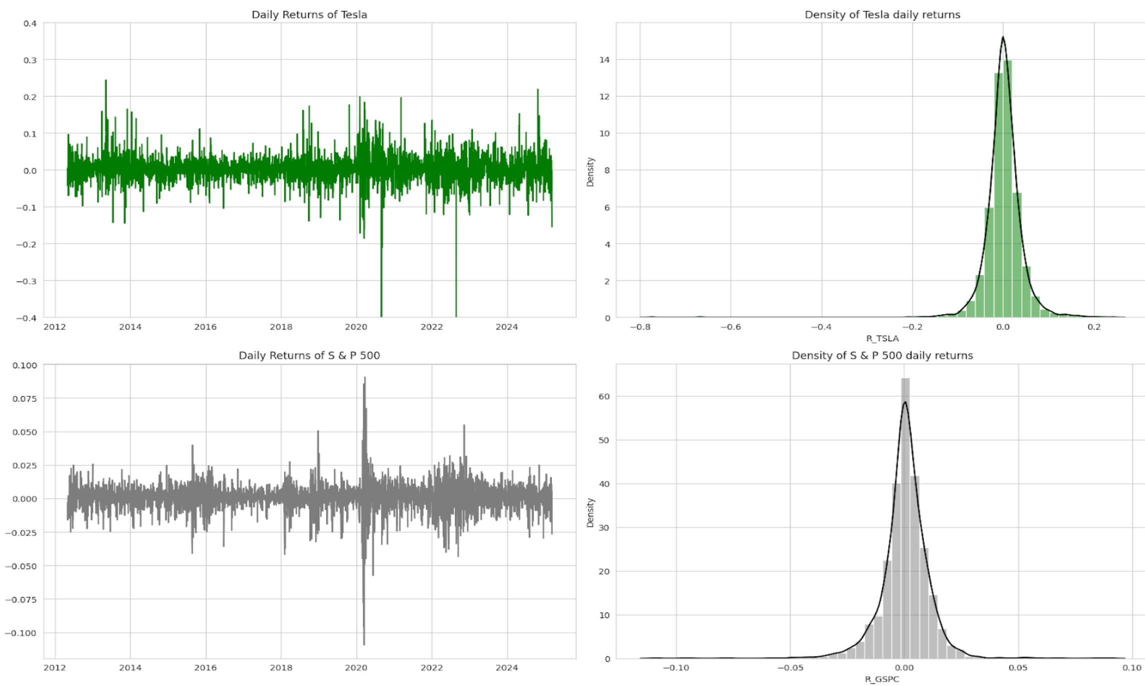

Figure - 11 **|** *Tesla and S&P 500*

The return graphs indicate periods of high volatility in Microsoft, Apple, Apple, and the S&P 500, particularly at the beginning of 2020 due to the COVID-19 crisis. During this time, the stock returns experienced significant fluctuations and increased volatility, reflecting the pandemic's market uncertainty and economic impact.

Apart from these volatile periods, the density of returns for each company appears to be symmetric around the origin, suggesting a balanced distribution of positive and negative returns. Additionally, the returns of each company exhibit a pattern that closely follows a normal distribution. This implies on average, the returns of these stocks tend to cluster near the mean value, with fewer extreme outliers.

Understanding the volatility and distribution of returns is crucial for investors as it provides insights into the potential risks and rewards associated with investing in these companies. It enables investors to evaluate historical performance and make informed decisions based on risk appetite and return expectations.

Correlation Analysis Between Tech Stocks and the S&P 500

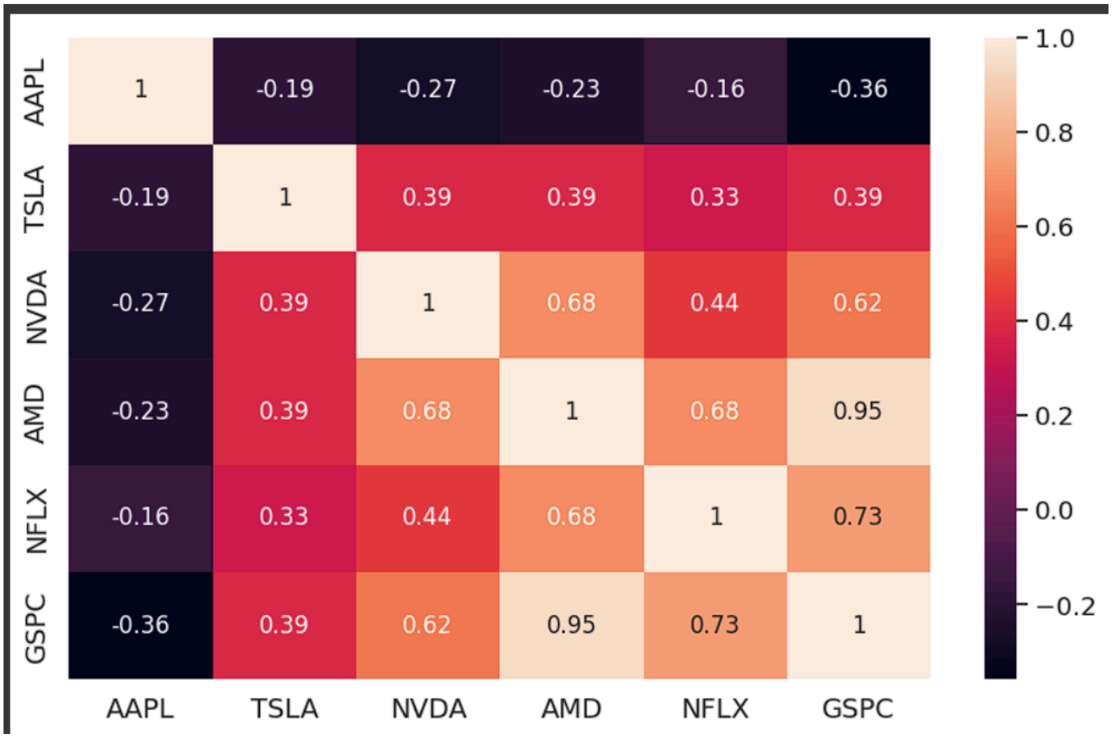

Figure - 12 | *Tech Giants vs S&P 500 – Correlation*

Correlation of Daily Returns: Apple, Nvidia, AMD, Netflix, Tesla & S&P 500

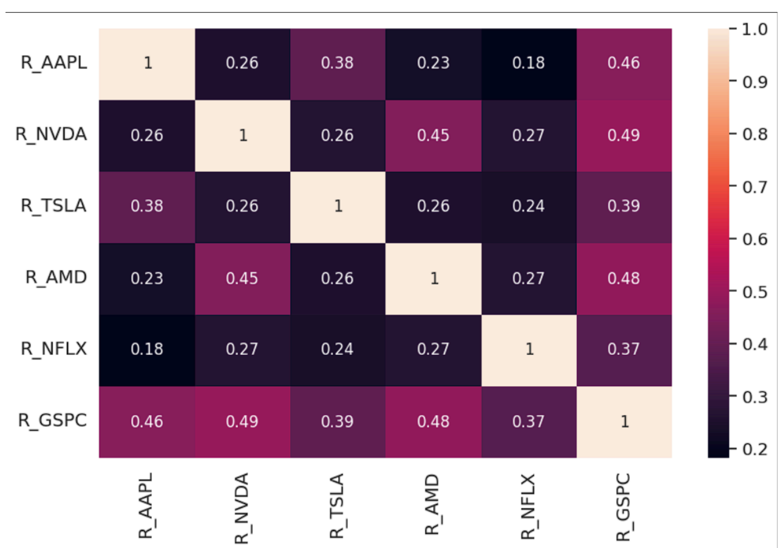

Figure - 13 | *Daily Return Correlation: Tech Stocks vs. S&P 500*

Portfolio optimization using Efficient Frontier

$$E[R_p] = \Sigma\ w_i \mu_i \text{ ....} \textbf{(3)}$$

$E[Rp]$: *Expected return of the portfolio*

w_i: *Weight of asset i in the portfolio*

μ_i: *Expected return of asset i*

$$\sigma_p = \sqrt{wT\ \Sigma w} \text{ ....} \textbf{(4)}$$

σ_p: *Standard deviation of the portfolio returns*

w: *Weight vector*

w^T: *Transpose of the weight vector*

Σ: *Covariance matrix of asset returns*

$$SharpeRatio = \frac{(E[R_p] - R_f)}{\sigma p} \text{ ....} \textbf{(5)}$$

$E[R_p]$: *Expected return of the portfolio*

R_f: *Risk − free rate*

σ_p: *Standard deviation (Risk) of the portfolio*

$$max_w = \frac{w^T \mu - R_f}{\sqrt{w^T\ \Sigma w}} \text{ ....} \textbf{(6)}$$

w: *Weight vector*

$w^T\mu$: *Expected portfolio return*

R_f: *Risk − free rate*

$w^T \Sigma w$: $Portfolio\ variance$

$\sqrt{w^T \Sigma w}$: $Portfolio\ standard\ deviation(Risk)$

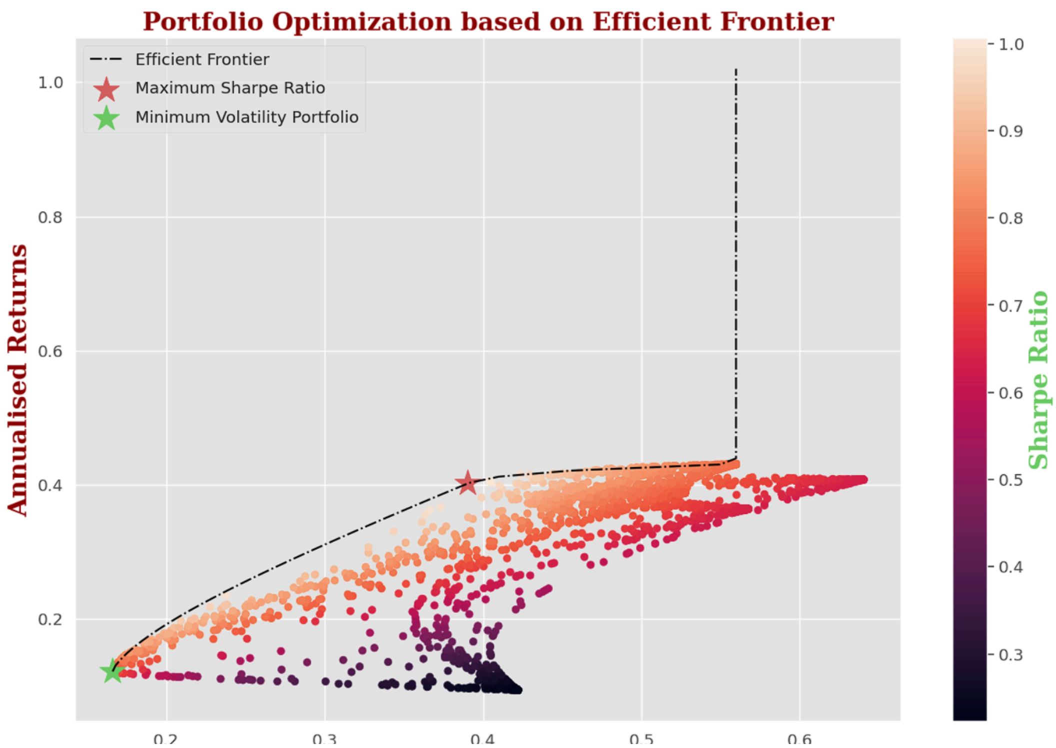

Figure - 14 **|** *Portfolio Optimization based on Efficient Frontier*

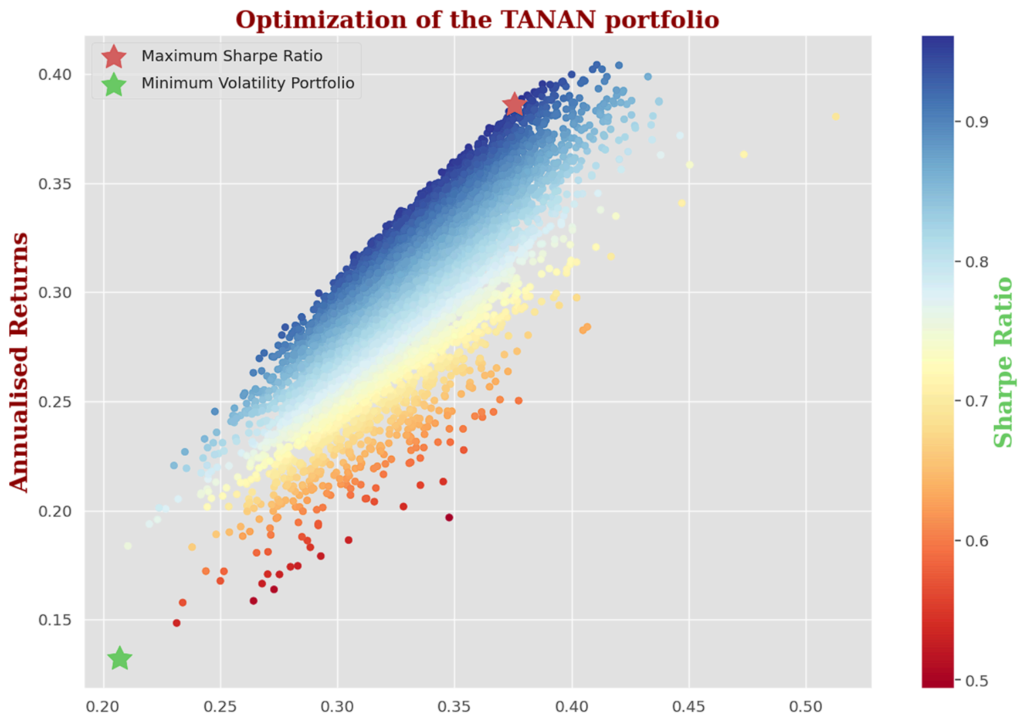

Figure - 15 **|** *Optimization of the TANAN portfolio*

Bollinger bands Strategy for Market Entry/Exit for given holding period

$$\text{Upper Band} = \mu_t + k \cdot \sigma_t \quad \text{..............................(7)}$$

$$\text{Lower Band} = \mu_t - k \cdot \sigma_t \quad \text{..............................(8)}$$

μ_t: *Moving average at time t*

σ_t: *Standard deviation*

k: *Multiplier*

Tesla:

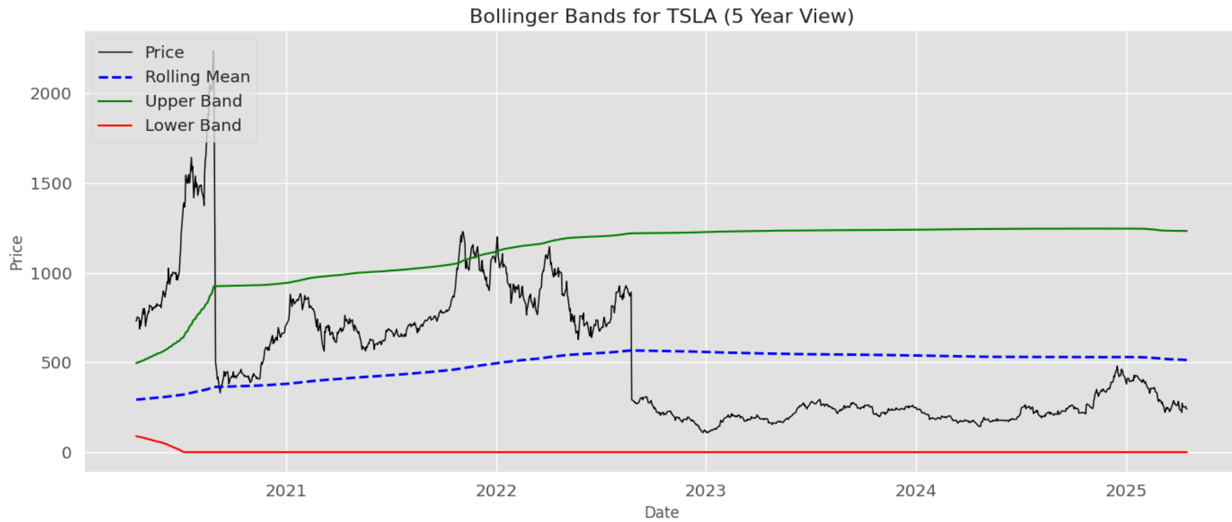

Figure - 16 | *Bollinger Bands for TSLA for 5 years*

Apple:

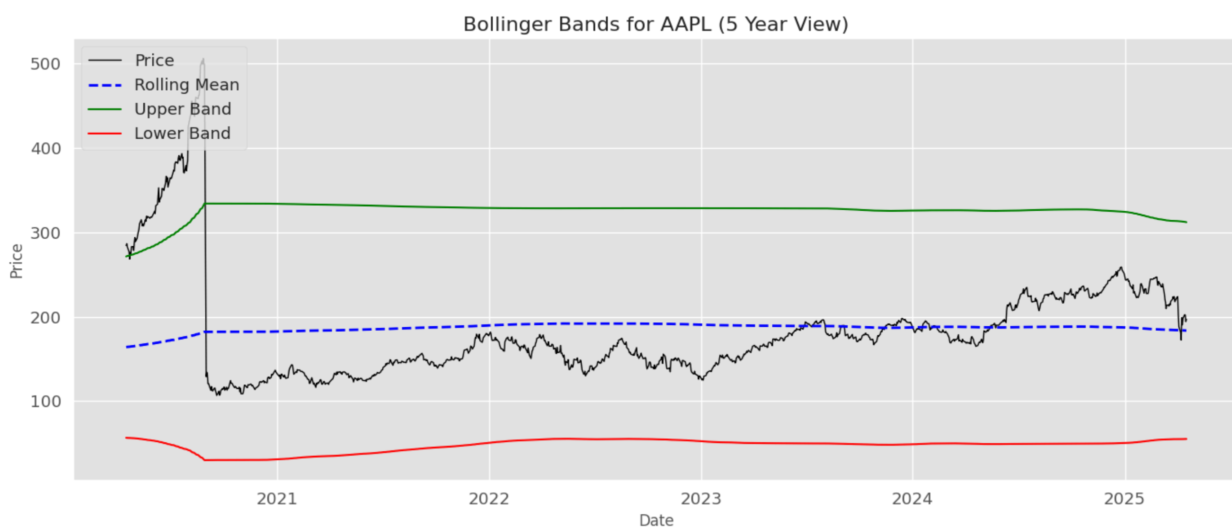

Figure - 17 | *Bollinger Bands for AAPL for 5 years*

Nvidia:

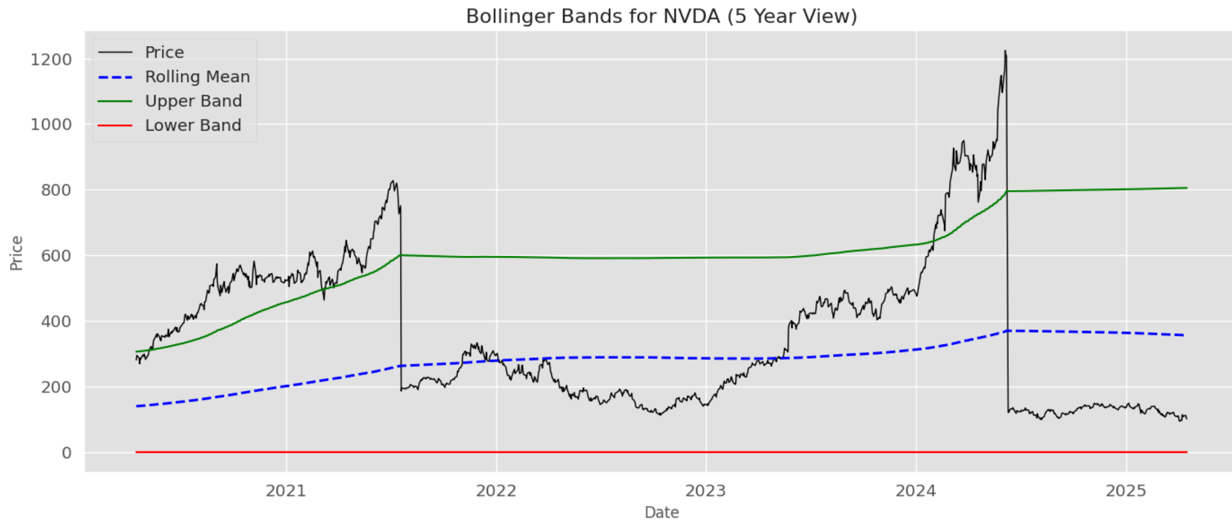

Figure - 18 **|** *Bollinger Bands for NVDA for 5 years*

AMD:

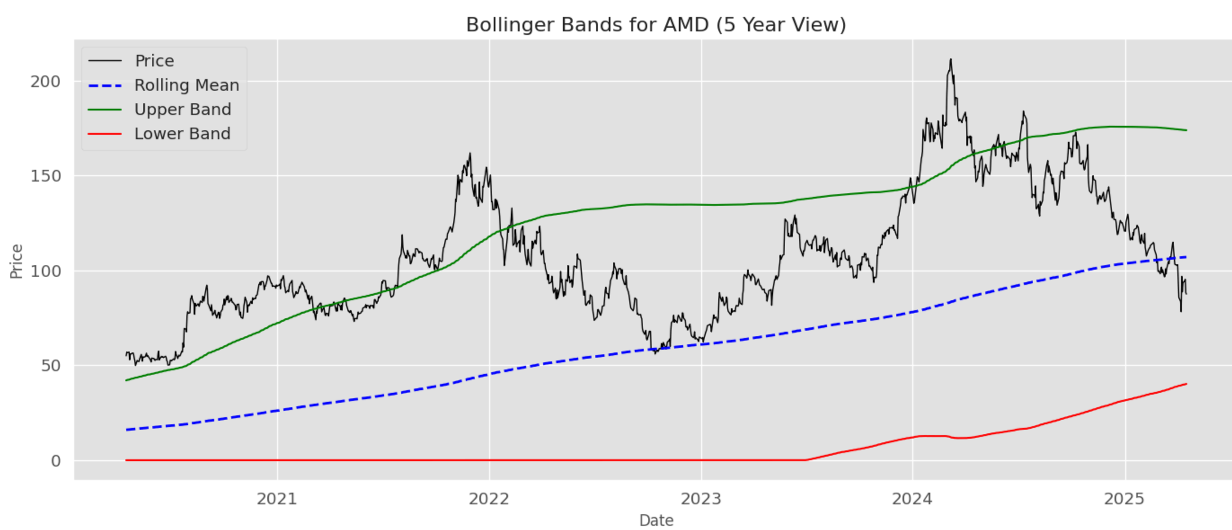

Figure - 19 **|** *Bollinger Bands for AMD for 5 years*

Netflix:

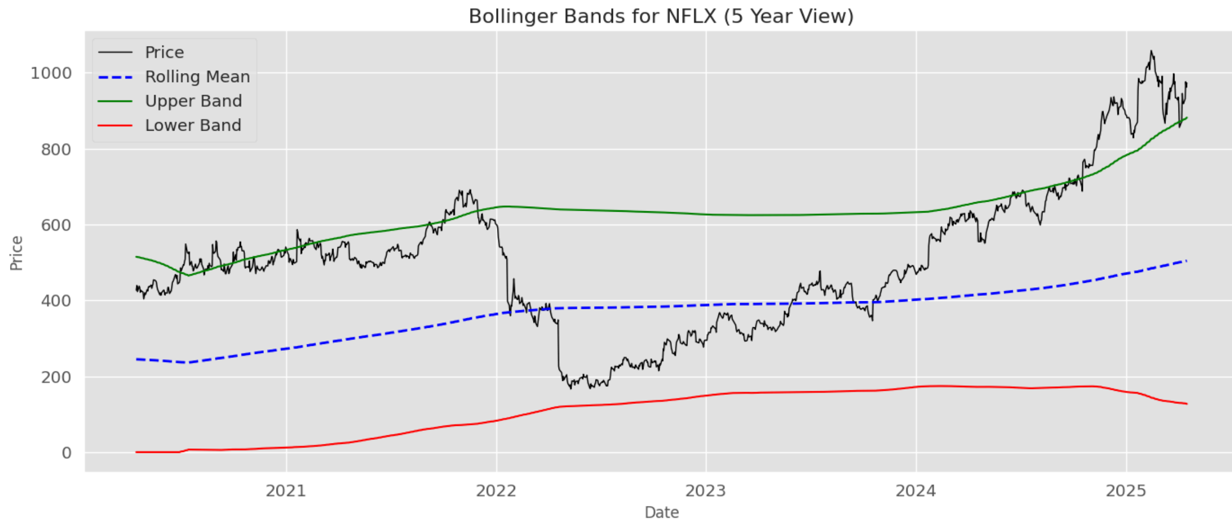

Figure - 20 **|** *Bollinger Bands for NFLX for 5 years*

GSPC:

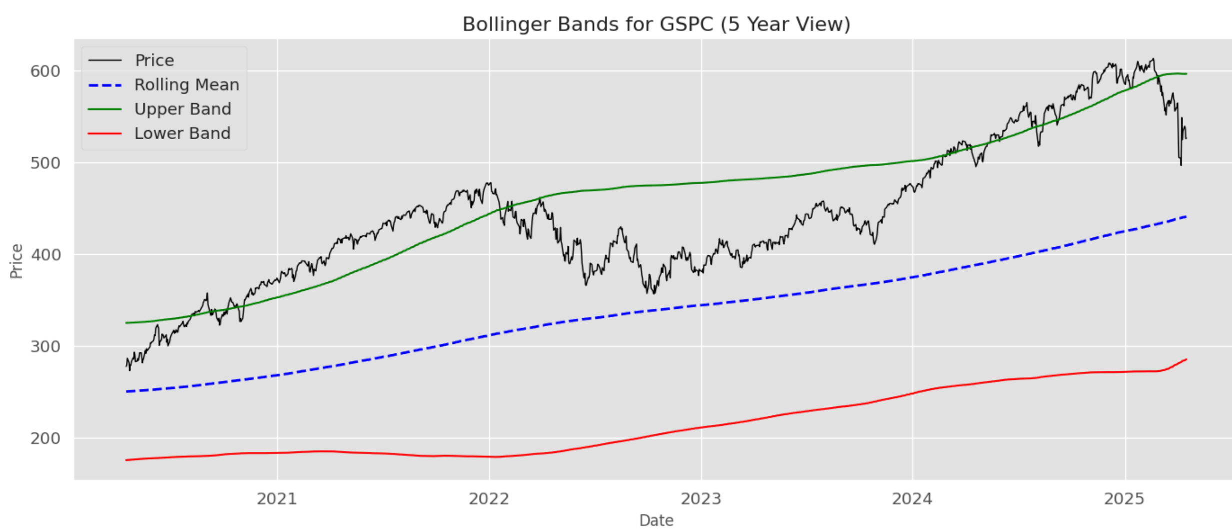

Figure - 21 **|** *Bollinger Bands for GSPC for 5 years*

Stock signal Analysis using bollinger bands

Enter your expected holding period in years: **5**

Strategy Selected: Medium-term strategy (standard bands)

Using num_std_dev = 2.0 for a 5 year(s) holding period.

--- TSLA Analysis ---

Latest Price : 241.37

Bollinger Upper Band: 1232.36

Bollinger Lower Band: 0.00

Suggested Action : Hold

--- AAPL Analysis ---

Latest Price : 196.98

Bollinger Upper Band: 312.06

Bollinger Lower Band: 55.24

Suggested Action : Hold

--- NVDA Analysis ---

Latest Price : 101.49

Bollinger Upper Band: 805.37

Bollinger Lower Band: 0.00

Suggested Action   : Hold

--- AMD Analysis ---

Latest Price   : 87.50

Bollinger Upper Band: 173.82

Bollinger Lower Band: 40.21

Suggested Action   : Hold

--- NFLX Analysis ---

Latest Price   : 973.03

Bollinger Upper Band: 881.74

Bollinger Lower Band: 127.26

Suggested Action   : Sell

--- GSPC Analysis ---

Latest Price   : 526.41

Bollinger Upper Band: 596.24

Bollinger Lower Band: 285.47

Suggested Action   : Hold

Conclusion

This project provided a thorough analysis of key technology stocks and the S&P 500 index over a multi-year period. By examining price trends, return distributions, and correlations, we identified strong relationships among tech stocks and varying risk profiles.

Using portfolio optimization techniques, including the Efficient Frontier and Sharpe Ratio maximization, we derived optimal asset allocations that balance expected returns against risk. Beta calculations offered insights into each stock's market sensitivity, aiding risk assessment.

The application of Bollinger Bands generated actionable trading signals based on price volatility, supporting informed buy, sell, or hold decisions aligned with investment horizons.

In summary, integrating statistical analysis, portfolio theory, and technical indicators enables more informed, data-driven investment decisions. Future work could extend this framework by incorporating additional risk measures and expanding asset classes to enhance portfolio diversification and trading strategies.

- **Investment Insights:**
    - Tech stocks offer higher growth potential but come with increased volatility and risk.
    - Diversification across multiple assets reduces overall portfolio risk.
    - Portfolio optimization techniques help balance risk-return trade-offs effectively.
- **Trading Strategy:**
    - Bollinger Bands provide actionable entry/exit signals based on price volatility.

References

1. Maller, R. A., Durand, R. B., & Jafarpour, H. (2010). Optimal portfolio choice using the maximum sharpe ratio. *The Journal of Risk, 12*(4), 49-73. Retrieved from https://www.proquest.com/scholarly-journals/optimal-portfolio-choice-using-maximum-sharpe/docview/817784573/se-2

2. Kuzheliev, Mykhailo & Zherlitsyn, Dmytro & Rekunenko, Ihor & Nechyporenko, Alina & Stabias, Sergii. (2025). Expanding portfolio diversification through cluster analysis beyond traditional volatility. Investment Management and Financial Innovations. 22. 147-159. 10.21511/imfi.22(1).2025.12.

3. Lento, C., Gradojevic, N., & Wright, C. S. (2007). Investment information content in Bollinger Bands? *Applied Financial Economics Letters*, *3*(4), 263–267. https://doi.org/10.1080/17446540701206576

4. Dye, R.T. and Groth, J.C. (2000), "Value weighting and simple optimization of portfolios: an empirical examination", Managerial Finance, Vol. 26 No. 6, pp. 23-35. https://doi.org/10.1108/03074350010766729
5. https://pypi.org/project/yfinance/

6. https://www.investopedia.com/terms/c/closingprice.asp

7. https://www.alphavantage.co/