Benchmark of Portfolio optimization techniques on Nasdaq dataset

Project for
Big Data in Business, Economics and Society
(a.y. 2024-2025)

Babak Khalilvandian
Cristian Longoni
Sergio Verga

Introduction

In this project, different techniques of portfolio management will be compared:

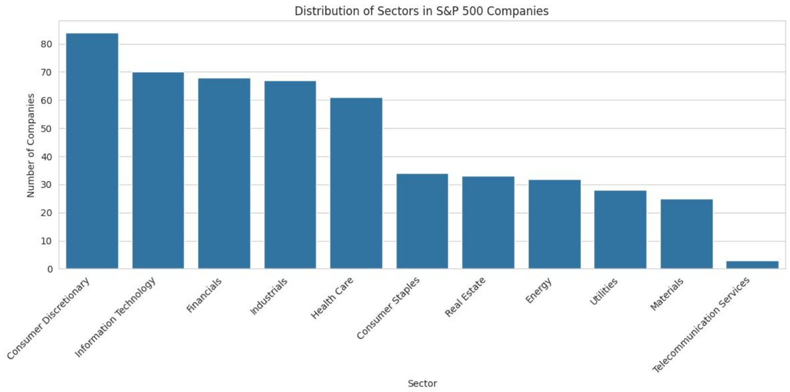
- Constant weights portfolio
- Markowitz portfolio
- LSTM based portfolio

First day of the month for prediction of **closing price** (with dividends correction)

Dataset preprocessing & exploration

15 Stocks were randomly chosen to be experimented by our strategies. The choices were from S&P 500 Stocks and selected from 3 Top Sectors considering number of companies:

- Consumer Discretionary
- Information Technology
- Financials



Dataset preprocessing & exploration

- Data for the selected symbols were gathered from Nasdaq nasdaq.com/market-activity/indexes/screener
- A range of 10 years were selected.
- Start Date: **2015 - 06 - 13**
- End Date: **2025 - 06 - 12**
- Added a **No Risk** option with constant price of 1

	Symbol	Name	Sector
0	CMG	Chipotle Mexican Grill	Consumer Discretionary
1	MHK	Mohawk Industries	Consumer Discretionary
2	GRMN	Garmin Ltd.	Consumer Discretionary
3	LEG	Leggett & Platt	Consumer Discretionary
4	ORLY	O'Reilly Automotive	Consumer Discretionary
5	BAC	Bank of America Corp	Financials
6	MS	Morgan Stanley	Financials
7	XL	XL Capital	Financials
8	CB	Chubb Limited	Financials
9	AXP	American Express Co	Financials
10	IBM	International Business Machines	Information Technology
11	EBAY	eBay Inc.	Information Technology
12	EA	Electronic Arts	Information Technology
13	AVGO	Broadcom	Information Technology
14	IT	Gartner Inc	Information Technology

Dataset preprocessing & exploration

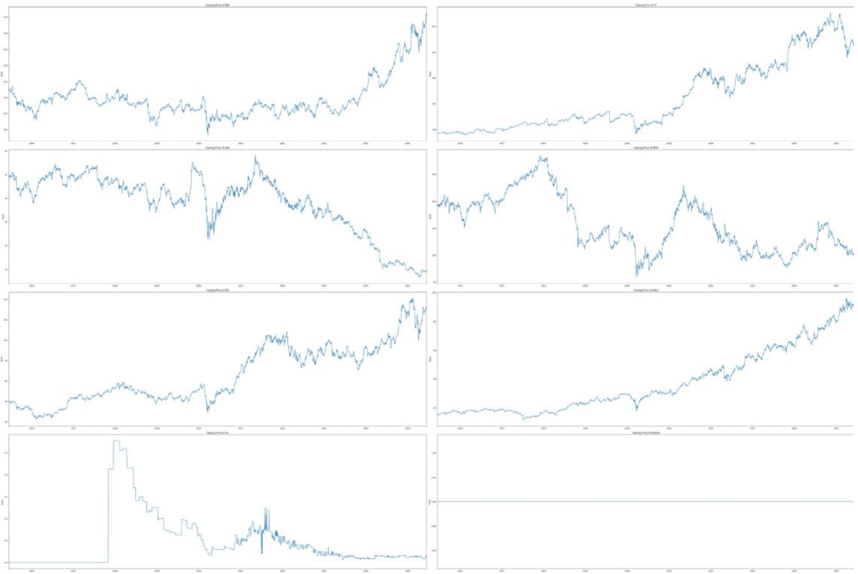
Dataset Includes:

- 2609 rows of Data
- Daily prices in business days
- Close: Closing price of the symbol
- Open: Starting Price of the day
- High & Low: Highest and lowest price recorded in the day
- Volume: Share traded of the symbol in the day

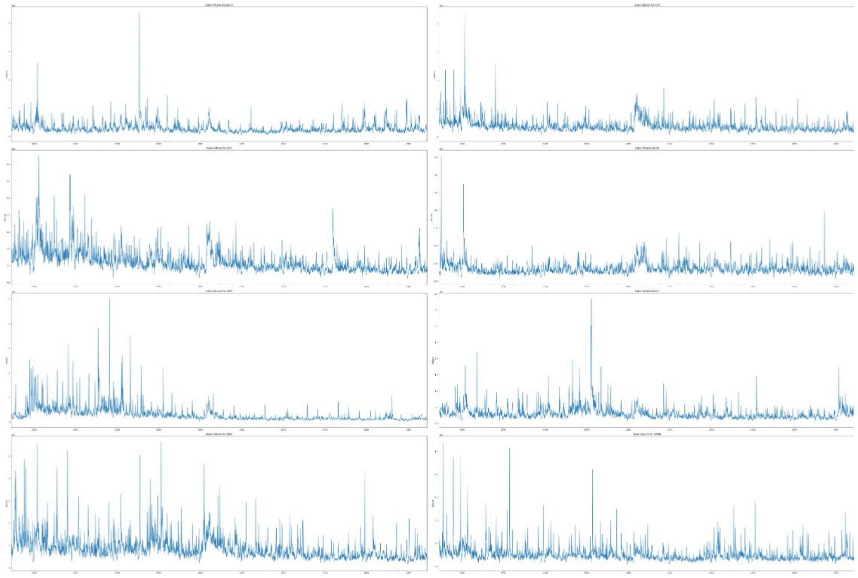
```
Stats for the company EA
```

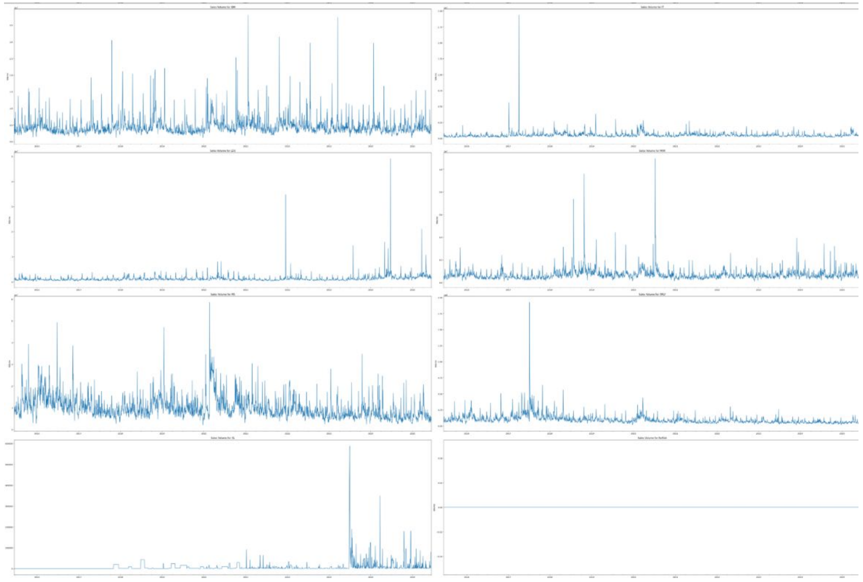
	close	volume	open	high	low
count	2609.00	2609.00	2609.00	2609.00	2609.00
mean	114.89	3022023.53	114.88	116.16	113.56
std	25.14	2147819.13	25.16	25.31	24.99
min	55.50	583933.00	54.39	56.57	53.01
25%	94.47	1852324.00	94.44	95.87	92.98
50%	121.50	2498714.00	121.43	123.00	120.21
75%	135.11	3514937.00	134.94	136.59	133.54
max	167.97	38356770.00	168.46	168.50	166.93

Daily closing price

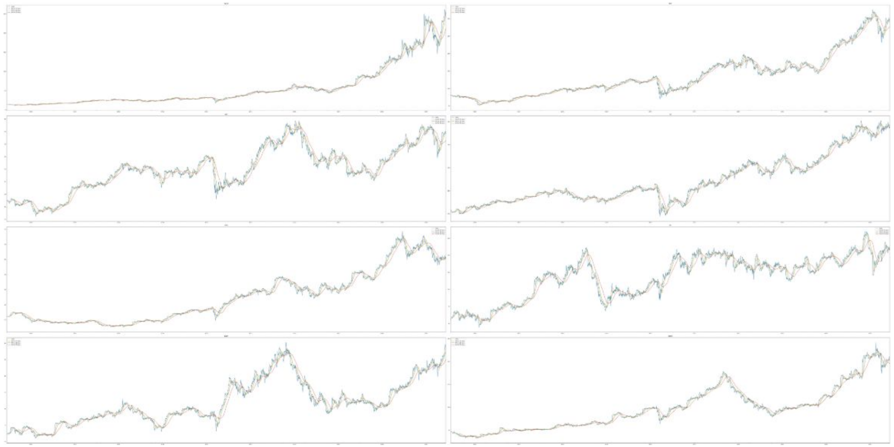


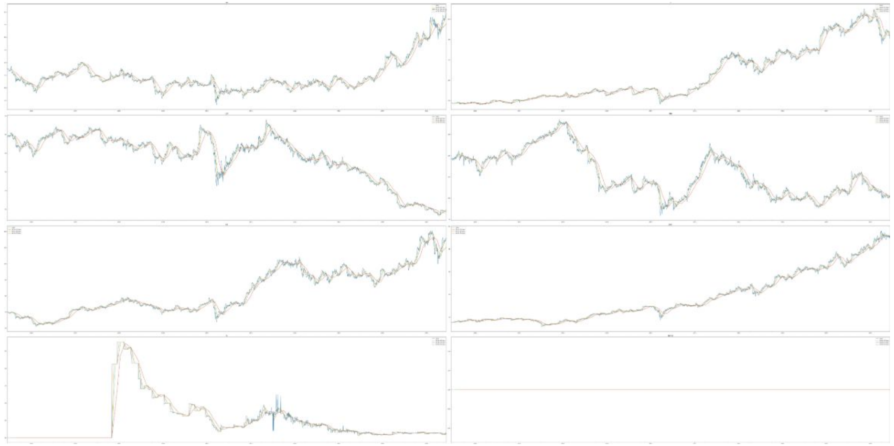
Daily volume





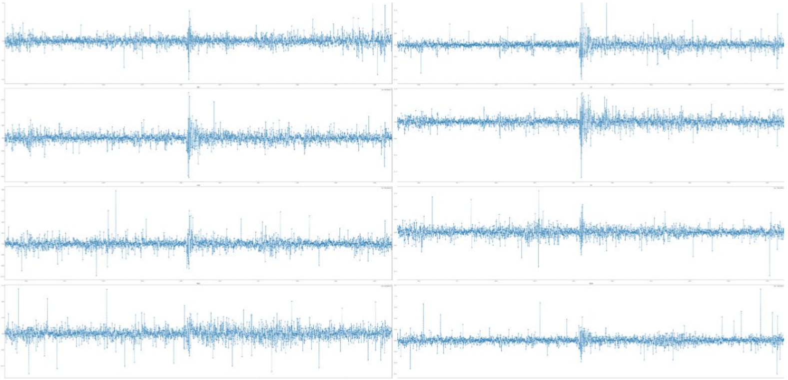
10, 20 and 50 days moving average on closing price

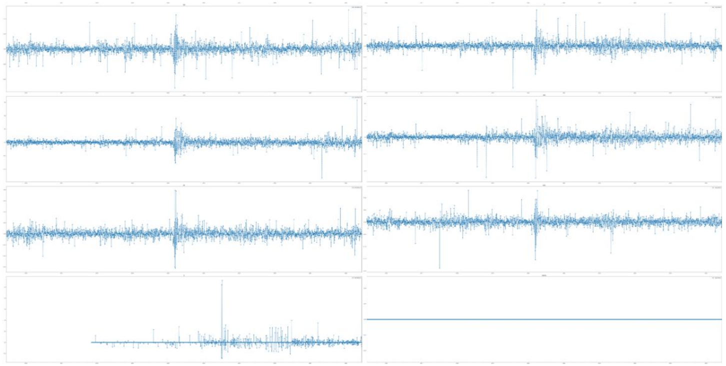


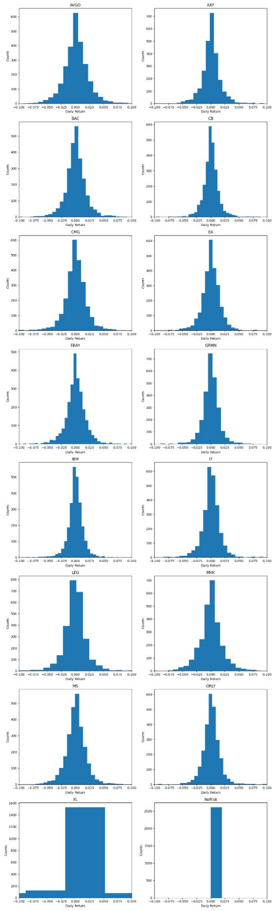


Daily return

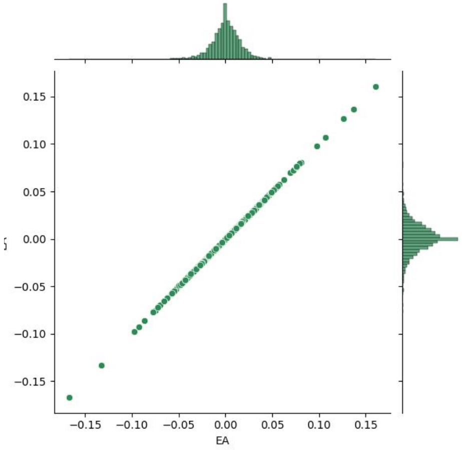
Exploratory data analysis is performed in the search of meaningful patterns

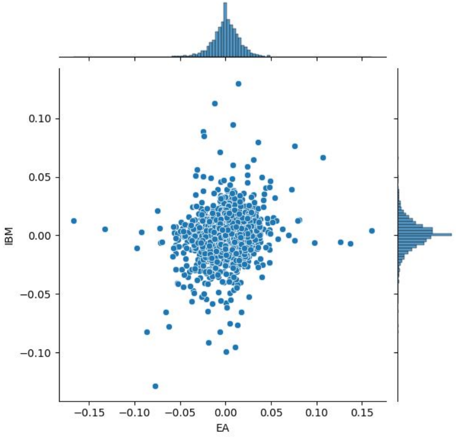






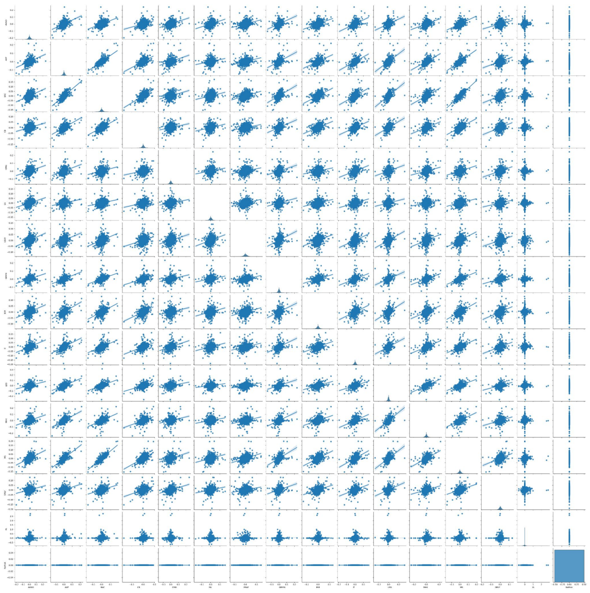
Scatterplots

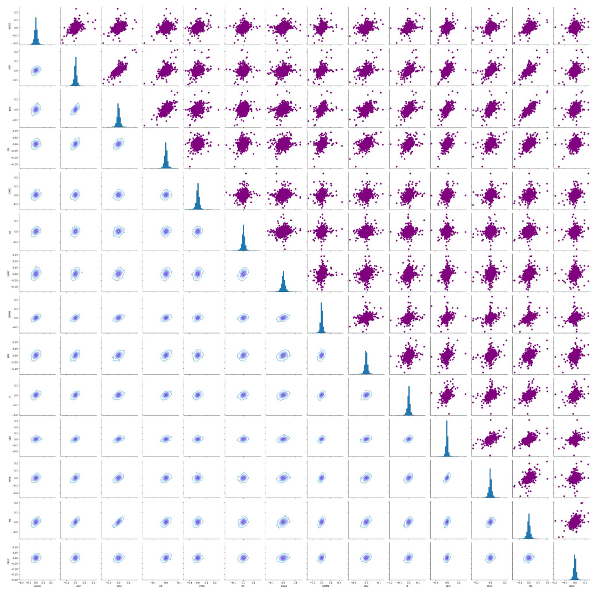


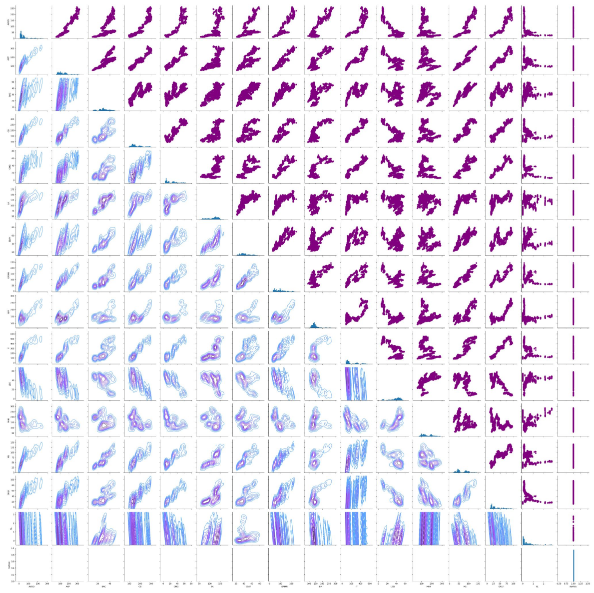


Scatterplots are investigated to look for comparing daily returns of different stocks. In this example, EA and IBM

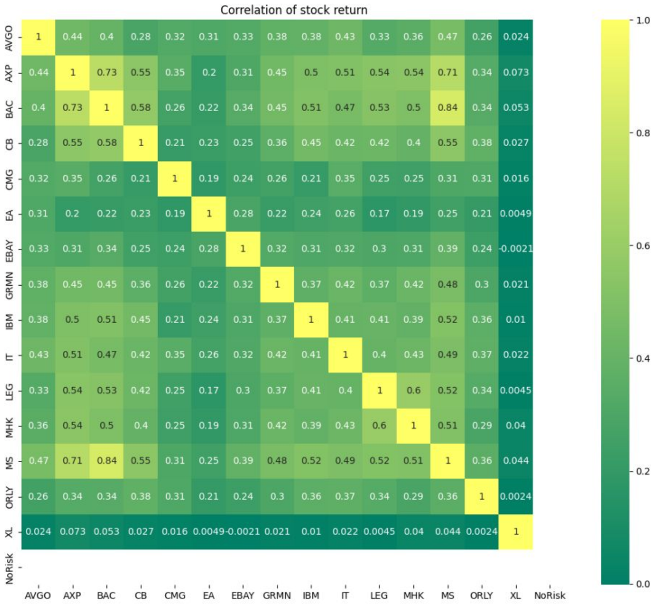
Scatterplots (iterating for all combinations)

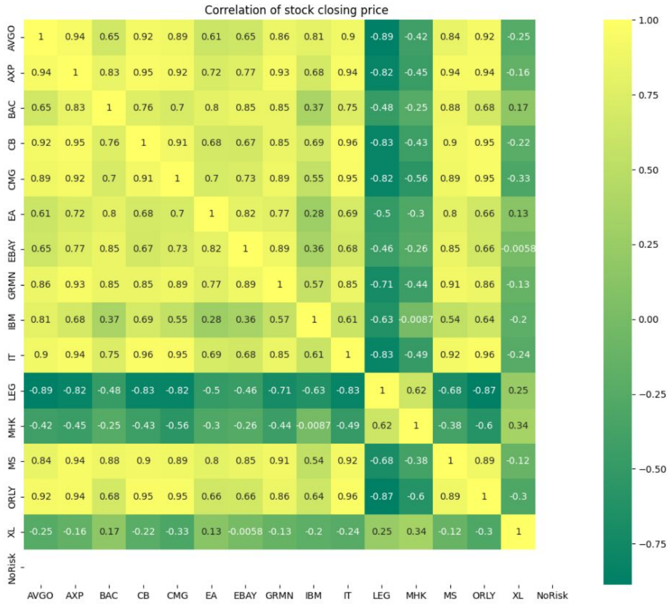




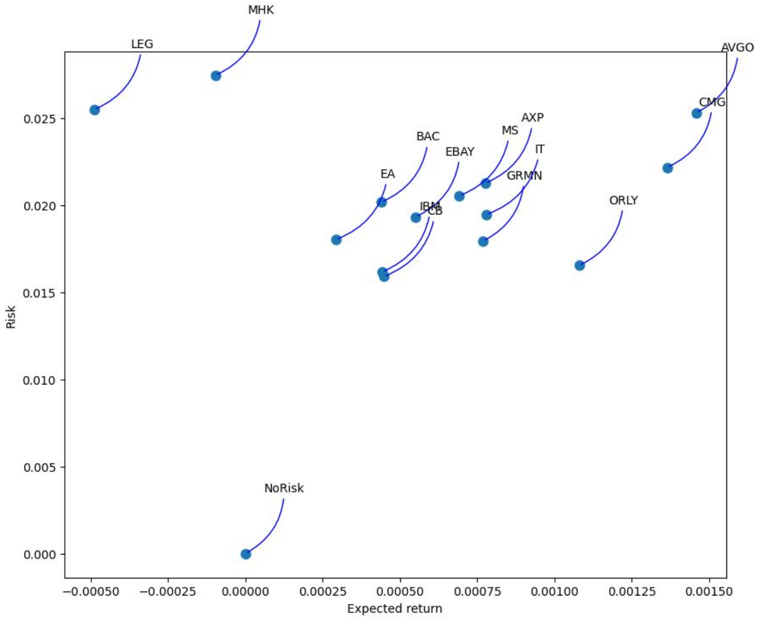


Correlation analysis (return & closing price)





Risk vs Expected Return



Constant Weights Portfolio as a Baseline

As a Baseline, all options are allocated the same budget for investment and backtraded.

Constant Weights Portfolio

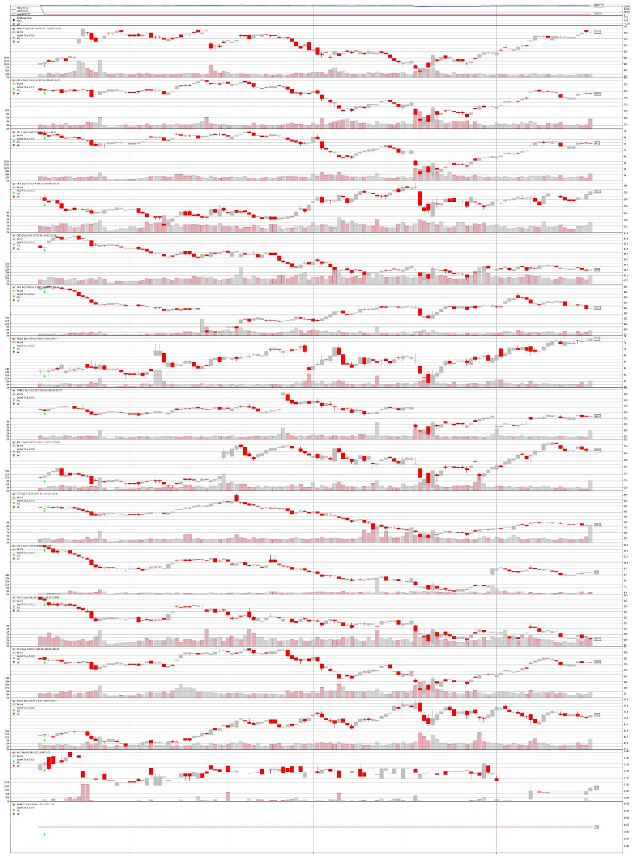
Simply, all symbols were assigned equal weights therefore equal amount of investment.

	Symbol	Weight	Position_Size
0	AVGO	0.062	6200.0
1	AXP	0.062	6200.0
2	BAC	0.062	6200.0
3	CB	0.062	6200.0
4	CMG	0.062	6200.0
5	EA	0.062	6200.0
6	EBAY	0.062	6200.0
7	GRMN	0.062	6200.0
8	IBM	0.062	6200.0
9	IT	0.062	6200.0
10	LEG	0.062	6200.0
11	MHK	0.062	6200.0
12	MS	0.062	6200.0
13	ORLY	0.062	6200.0
14	XL	0.062	6200.0
15	NoRisk	0.062	6200.0

Constant Weights Results

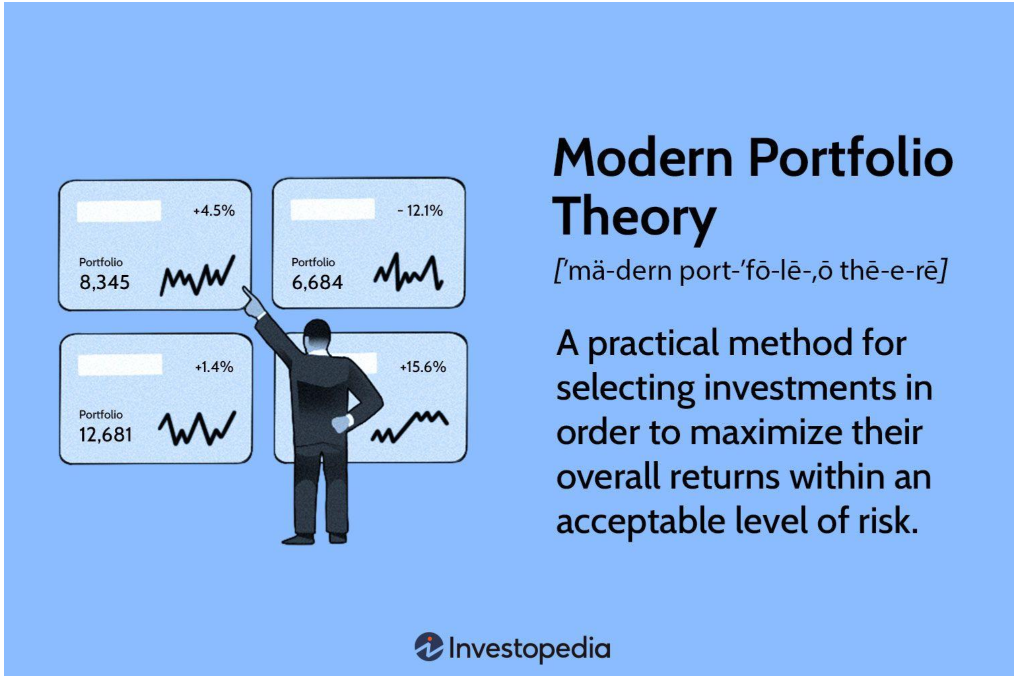
This is a baseline of a portfolio selection and any good strategy must be able to statistically outperform it.

- `Buy and Hold since 2024-12-01`
- `Initial Value: 100000`
- `Final Value: 96057.77`
- `Total Gain: -3942.23`
- `Percent Gain: -3.94 %`



Portfolio Optimization with Markowitz

In this section, Portfolio Optimization is performed in order to build the optimal portfolio in terms of Expected Returns and Expected Volatility



Markowitz part outline

solved the **Minimum Variance Portfolio** problem:

$$\min_{w} \sqrt{w^\top \Sigma w} \quad \text{subject to} \quad \sum w = 1,\ 0 \leq w_i \leq 1$$

- Group each stock based on sector (Consumer Discretionary, Financials, Information Technology) and impose to invest equally (33,33%) for each sectors
- We consider only closing prices
- Calculates daily returns for each stock, dictionary with sector names mapped to return matrices (e.g. { 'Finance': df_of_daily_returns, ... })

For each sector:

- Applies **minimum variance optimization** using the **Ledoit-Wolf shrinkage estimator,** compute a **robust covariance matrix**
- Optimizes weights such that the portfolio volatility is minimized under the constraint ∑weights = 1.

- Each sector's portfolio is then scaled to contribute **equally** to the global portfolio.

The final portfolio is a weighted sum of each optimized sector portfolio.

Ledoit-Wolf shrinkage estimator

Useful when:

- You have **many assets**
- And **not enough historical data** (short time series)

It improves the **stability and accuracy** of the covariance matrix, which is critical for portfolio optimization (like Markowitz), where this matrix determines portfolio risk and diversification.

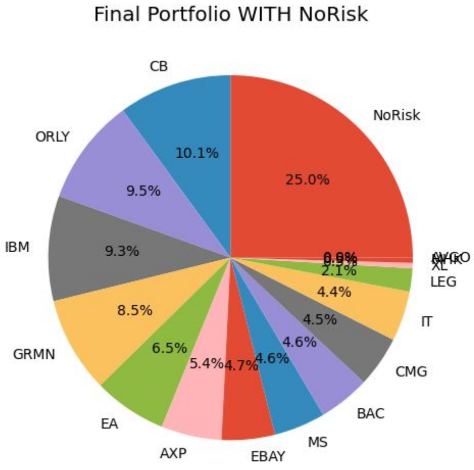
$$\Sigma_{\mathrm{LW}} = \lambda\Sigma_{\mathrm{target}} + (1 - \lambda)\Sigma_{\mathrm{sample}}$$

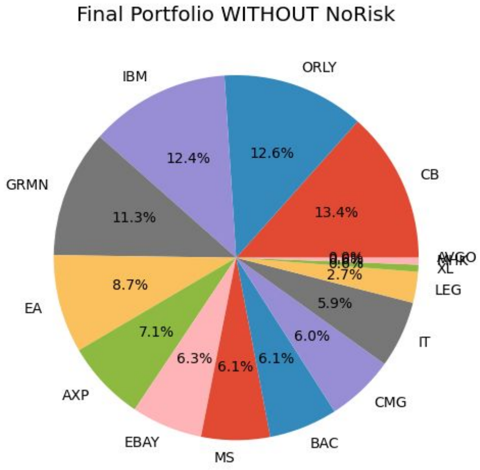
Where:

- Σ_{sample}: empirical covariance matrix (calculated from historical data)
- Σ_{target}: structured, well-conditioned matrix (e.g. a diagonal matrix or scaled identity)
- λ: **shrinkage intensity**, a value between 0 and 1

Final Portfolios Visualizations

Ticker	Weight
NoRisk	2.500000e-01
CB	1.005295e-01
ORLY	9.467176e-02
IBM	9.321222e-02
GRMN	8.489468e-02
EA	6.519476e-02
AXP	5.360300e-02
EBAY	4.722305e-02
MS	4.563379e-02
BAC	4.552913e-02
CMG	4.527778e-02
IT	4.436997e-02
LEG	2.056986e-02
XL	4.704607e-03
MHK	4.585931e-03
AVGO	1.694066e-19





Ticker	Weight
CB	1.340393e-01
ORLY	1.262290e-01
IBM	1.242830e-01
GRMN	1.131929e-01
EA	8.692634e-02
AXP	7.147067e-02
EBAY	6.296407e-02
MS	6.084505e-02
BAC	6.070550e-02
CMG	6.037037e-02
IT	5.915997e-02
LEG	2.742648e-02
XL	6.272809e-03
MHK	6.114575e-03
AVGO	2.258755e-19

Introducing a NoRisk asset reduces overall portfolio risk by allocating part of the capital to a non-volatile asset.

Efficient Frontier

Computes:

- Mean returns and covariance matrix (Ledoit-Wolf) on training + validation
- Efficient frontier: for 100 target returns, find min volatility portfolio
- Tangency portfolio: maximizes the Sharpe ratio directly using SLSQP

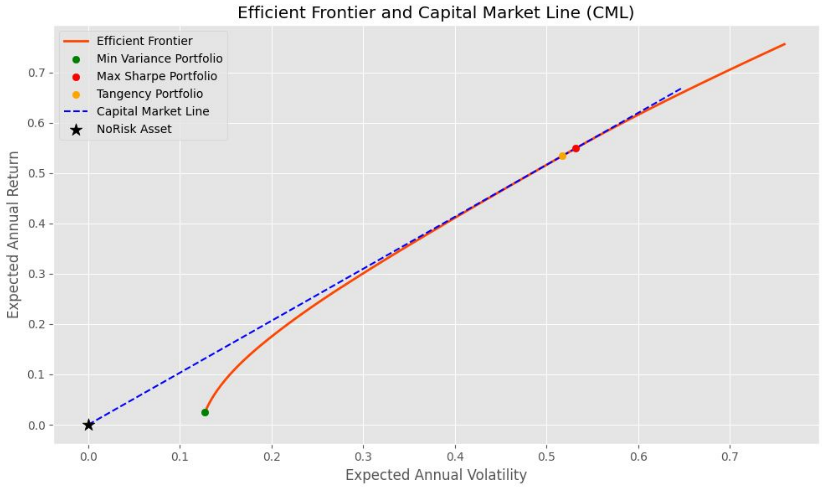
rf = 0.0 for the NoRisk asset since has 0% daily return, CML starts from (0, 0)

The Tangency Portfolio (yellow) maximizes the Sharpe ratio when a risk-free asset is available.

The Max Sharpe Portfolio (red) and the Tangency Portfolio are nearly overlapping → indicating that the risk-free asset has little impact on optimal weights (but not zero).

Maximized the **Sharpe Ratio** directly:

$$\max_{w} \frac{w^\top \mu - r_f}{\sqrt{w^\top \Sigma w}} \quad \text{subject to} \quad \sum w = 1$$



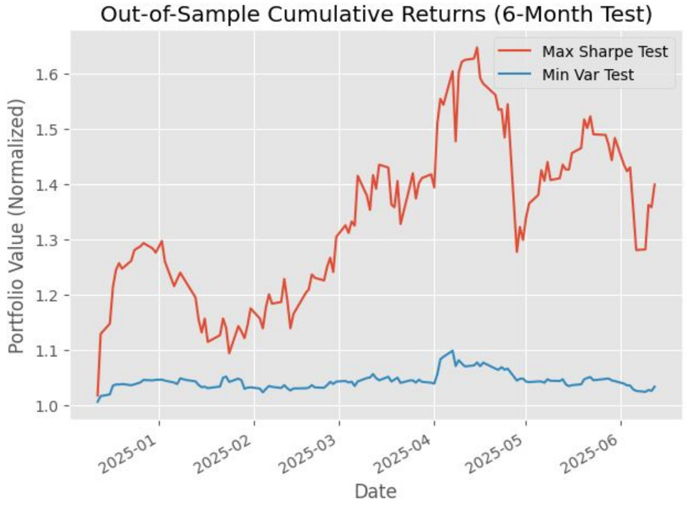

```
Max Sharpe Portfolio Return:     0.5493
Max Sharpe Portfolio Volatility:0.5320
Max Sharpe Ratio:                0.0650
```

Dataset split:

- **Test** (last 6 months)
- **Validation** (previous 6 months)
- **Training** (all previous months)

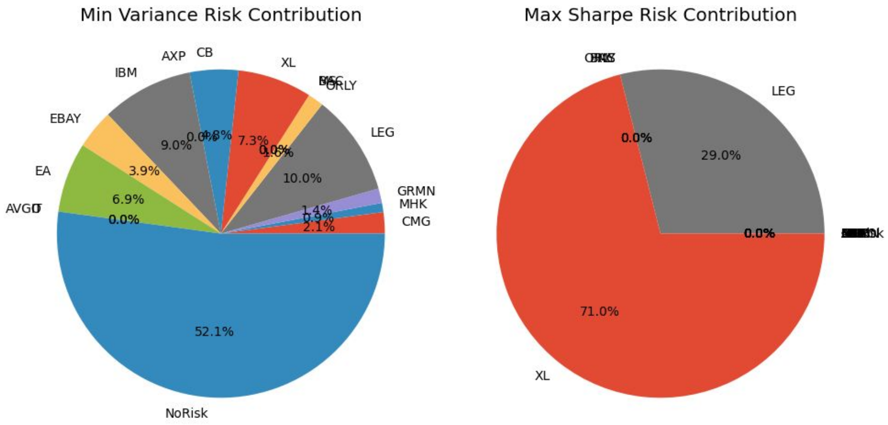
Mininum Variance and Max Sharpe Portfolios

```
=== Out-of-Sample Performance (Last 6 Months) ===

Max Sharpe Portfolio:
  Return:     0.8285
  Volatility: 0.5360
  Sharpe:     1.5458
  Sortino:    0.1273
  Calmar:     3.6845
  CVaR 5%:    0.0721

Min Variance Portfolio:
  Return:     0.0707
  Volatility: 0.1016
  Sharpe:     0.6955
  Sortino:    0.0630
  Calmar:     1.0406
  CVaR 5%:    0.0139
```



Conclusion: Max Sharpe delivers much higher returns but with high volatility. Min Variance is more stable, but with low profitability.

Diversification metrics




```
Diversification Metrics:
 - Min Variance: Entropy = 1.6251, Herfindahl = 0.3333
 - Max Sharpe:   Entropy = 0.5761, Herfindahl = 0.6124

--- Risk Contributions (as % of total volatility) ---
Min Variance Portfolio:
NoRisk: 52.15%

Max Sharpe Portfolio:
LEG: 28.98%
XL: 71.02%
```

```
--- Minimum Variance Portfolio Weights ---
CMG: 0.0278
MHK: 0.0077
GRMN: 0.0196
LEG: 0.0740
ORLY: 0.0225
XL: 0.0197
CB: 0.0578
IBM: 0.0940
EBAY: 0.0451
EA: 0.0794
NoRisk: 0.5525
```

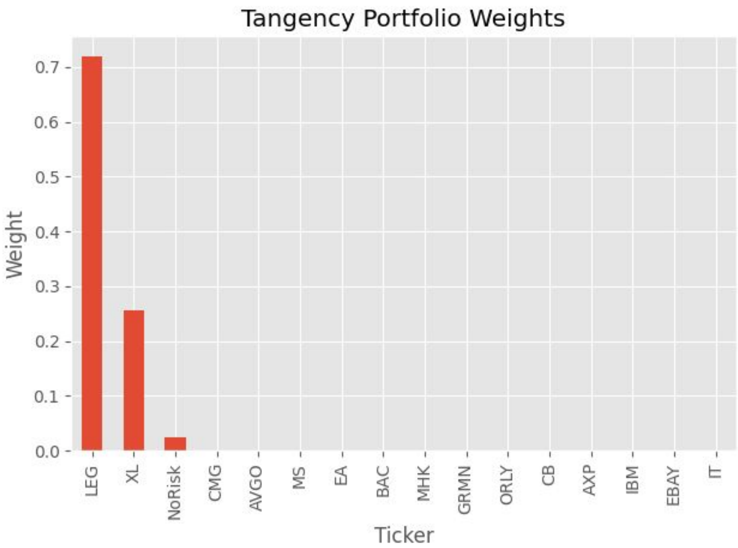
```
--- Maximum Sharpe Ratio Portfolio Weights ---
LEG: 0.7371
XL: 0.2629
```

The Minimum Variance Portfolio concentrates risk mainly on NoRisk (52%) and IBM.

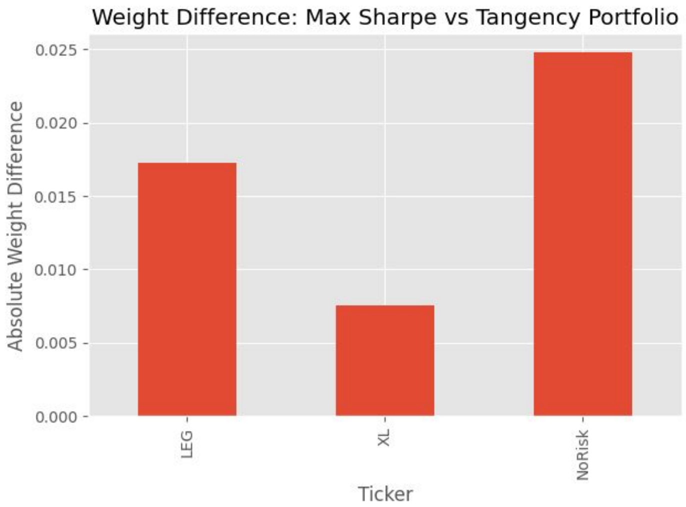
The Max Sharpe Portfolio is extremely concentrated: XL and LEG contribute 100% of the portfolio risk.

Conclusion: Minimizing variance promotes diversification; maximizing Sharpe may reduce it significantly.

Tangency Portfolio




```
--- Tangency Portfolio (True Max Sharpe) ---
Expected Return: 0.5343
Volatility:      0.5175
Sharpe Ratio:    1.0325
Weights (NoRisk included):
  LEG: 0.7198
  XL: 0.2554
  NoRisk: 0.0248
```




```
Difference between Max Sharpe and Tangency weights:
L1 norm: 0.049641
L2 norm: 0.031157
```

The Max Sharpe (from grid) and Tangency (from optimization) portfolios are theoretically equivalent

Differences are very small (< 2.5%) and mostly on low-weight assets

Tangency portfolio slightly allocates to NoRisk to reduce marginal volatility.

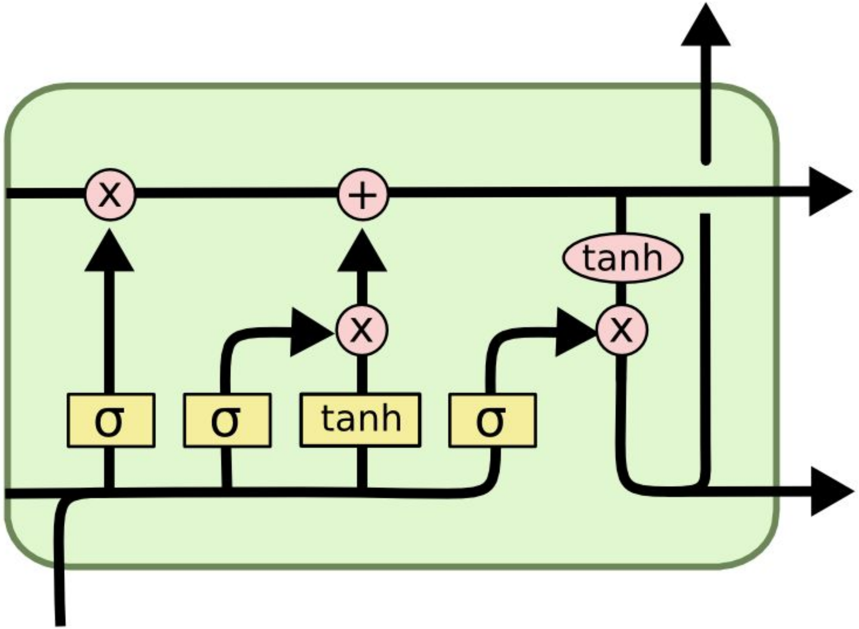
Conclusion

Findings:

- The Markowitz-optimized portfolio delivers consistent theoretical results.
- Adding a NoRisk asset **improves the risk-return profile** and stability.
- There's a **trade-off between diversification and Sharpe maximization**.
- The **out-of-sample analysis** confirms Max Sharpe's dominance **only if higher risk is acceptable**.
- The Tangency Portfolio emerges as a **theoretically stronger choice** in presence of a risk-free asset.

Long-Short-Term-Memory neural networks

In this section, LSTMs are exploited to optimize stock closing prices prediction and maximize the expected return



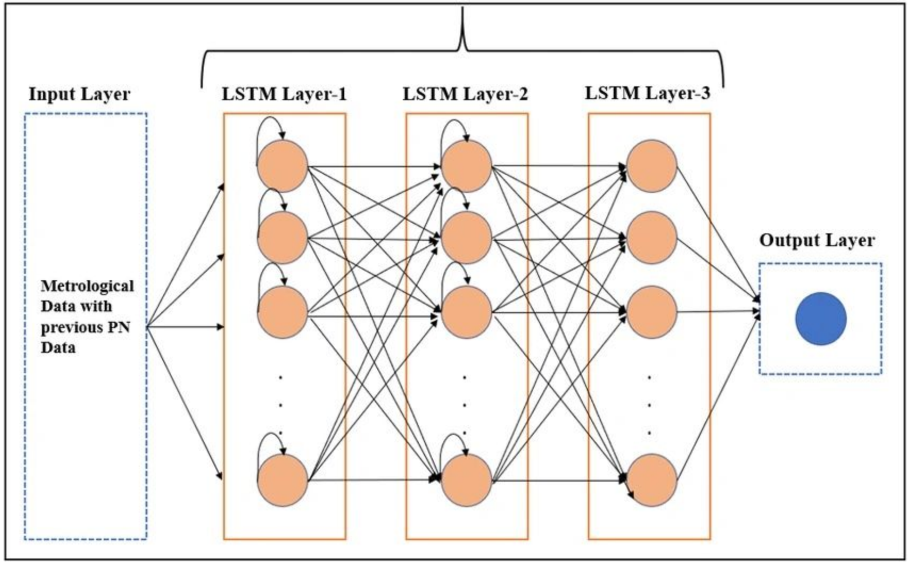
LSTM part outline

Model: Long Short-Term Memory (LSTM) Neural Network

- Designed for sequential data like time series
- Handling long-term dependencies in stock prices

Architecture:

- 2 stacked **LSTM** layers
- **Dropout** regularization to prevent overfitting
- **Dense output layer** for closing price prediction



LSTM Training and Inference

Dataset split:

- **Test** (last 6 months)
- **Validation** (previous 6 months)
- **Training** (all previous months)

Hyperparameter optimization to select the **optimal model**

The **optimal model** is then re-trained on **training+validation**

Test set predictions with each optimized model

Hyperparameter	Search space
LSTM 1	[32, 64, 96, 128]
Dropout 1	[0.1, 0.2, 0.3, 0.4, 0.5]
LSTM 2	[32, 64, 96, 128]
Dropout 2	[0.1, 0.2, 0.3, 0.4, 0.5]
Optimizer	[Adam, RMSprop]
Sequence length	Fixed at 30 timesteps
Batch size	Fixed at 64

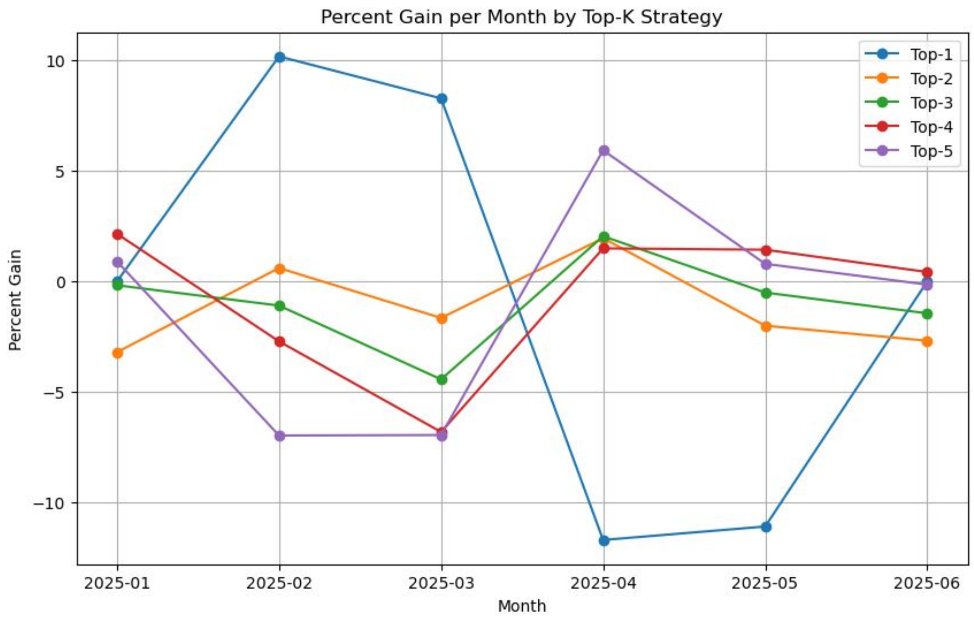
Optimizing the portfolio with LSTMs

Top-k strategy

- Tracking k in **[1,5]**
- Monitoring k in **[1,16]**

Maximizing the profit according to LSTM predictions

Evaluation using backtrading to exploit the **LSTM predictions** (for each strategy)



LSTM-based portfolio management improvements

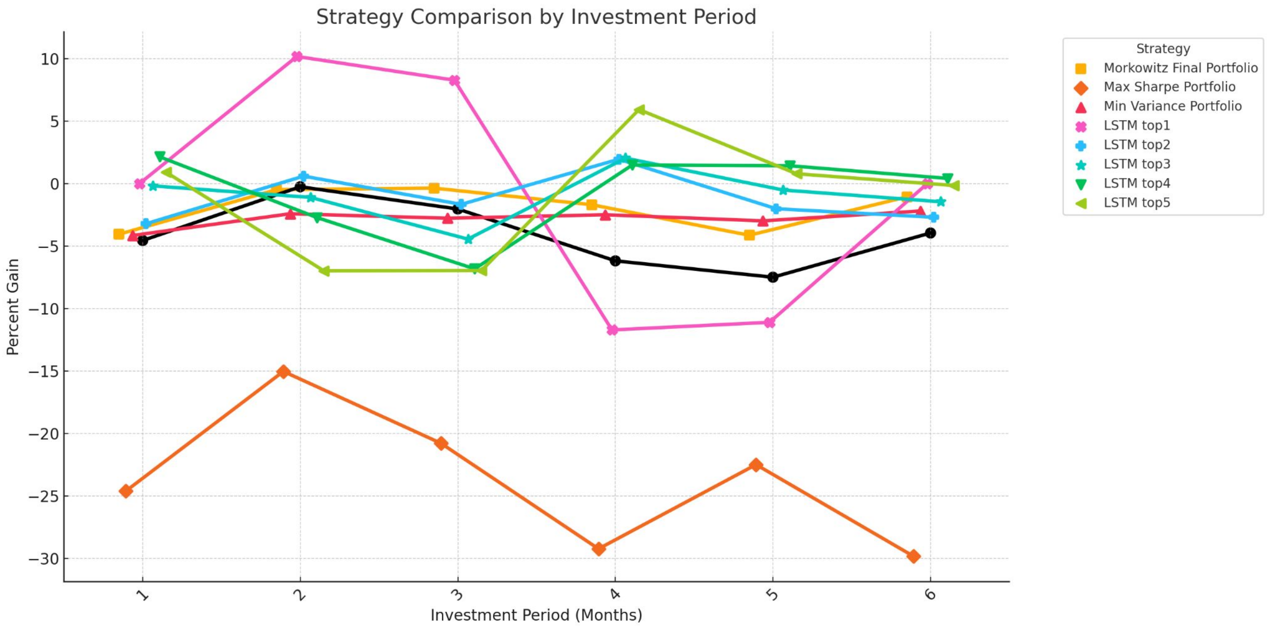
Further experiments may include the following:

1. Accounting for **Seasonality**
2. **Retrain every month** for the next month prediction
3. Analyze predictions with **daily granularity**
4. For top-k portfolios (with k>1), elaborate **more advanced weight subdivision**

Results & conclusions

In this section, the different techniques about portfolio management on Nasdaq are benchmarked

Results



- **LSTM Top-K strategies** showed higher adaptability;
  - **Top-3, Top-4 and Top-5** consistently outperformed Top-1. In the shortest horizon **Top-1** reaches maximum profitability(+10% in month 2) but with high instability. (-1.44%, 0.43%, -0.14%)
  - Top-3/Top-4/Top-5 offer optimal trade-off between return and risk
- **Markowitz optimization** achieved balanced risk-adjusted returns but underperformed in dynamic conditions.
  - **Max Sharpe** delivers much higher returns but with high volatility (-29.82 %)
  - **Min Variance** is more stable, but with low profitability (-2.18 %)
- **Constant-weight portfolios** remained stable but lacked responsiveness to market changes (-3.94 %)