A Comparative Study of Euclidean Distance k-Means and DTW-Based Hierarchical Clustering for Stock Grouping in the S&P 500

Giulia Talà, Nicolò dal Monte, Gloria Desideri, Lorenzo Pirozzi, Giacomo Maggiore and Simone Zani

BlackswanQuants Student Organization

ARTICLE INFO

Keywords:
Portfolio Optimization
Dynamic Time Warping (DTW)
Euclidean Distance
S&P 500
Time Series Analysis
Financial Markets
Machine Learning Techniques
Quantitative Finance
Clustering
Unsupervised Machine Learning

ABSTRACT

This paper proposes a data-driven framework for constructing a capital-efficient proxy portfolio that replicates the performance of the SP 500 index using a reduced number of constituent assets. Traditional index replication strategies, such as full replication or market capitalization weighting, often pose limitations in terms of transaction costs, scalability, and practical implementation for individual or institutional investors with constrained resources.

We introduce a methodology that combines time-series clustering with constrained optimization to address these limitations. First, we apply unsupervised learning techniques to group stocks based on similarity in their historical price trajectories. We benchmark KMeans clustering with Euclidean distance against hierarchical clustering using Dynamic Time Warping (DTW), identifying the latter as the most effective in capturing temporal dependencies and aligning stock behaviors.

A representative subset of stocks is then selected from each cluster to form the candidate portfolio. We formulate a constrained linear optimization problem, inspired by Markowitz's mean-variance framework, to determine the asset weights that minimize the tracking error with respect to the SP 500 index.

Empirical results from historical backtests show that the proposed method effectively replicates the index with high accuracy, delivering superior performance while substantially reducing portfolio size. This approach presents a more interpretable, scalable, and computationally efficient alternative to conventional index investment strategies.

1. Introduction

In the evolving landscape of financial markets, the construction of efficient and diversified portfolios remains a central objective for both institutional asset managers and retail investors. Traditional investment strategies typically rely on direct stock selection, mutual funds, or passive index tracking via exchange-traded funds (ETFs). While effective, these methods often entail limited flexibility, high capital requirements, or insufficient customization.

This study proposes a novel methodology for index replication, with a specific focus on the SP 500. The objective is to develop a reduced and more accessible portfolio that approximates the index's behavior while potentially enhancing its risk-adjusted returns. To this end, we leverage historical price data to identify structural similarities across constituent equities through unsupervised learning techniques.

The analysis begins with a comparative evaluation of clustering algorithms for financial time series, ultimately selecting hierarchical clustering with Dynamic Time Warping (DTW) as the most suitable method for capturing nuanced temporal dependencies and alignment-invariant relationships in stock behavior.

Following the clustering phase, representative assets are selected from each cluster to ensure coverage of the underlying dynamics present in the index. Subsequently, a constrained mean-variance optimization problem is formulated within the classical Markowitz framework, where the objective is to minimize portfolio variance for a given expected return, subject to constraints designed to approximate the risk-return characteristics of the SP 500 index. Constraints are introduced to reflect practical investment considerations, including budget normalization, non-negativity, and upper bounds on individual weights, thereby ensuring implementability and diversification.

The resulting portfolio is subjected to a backtesting procedure to evaluate its historical performance relative to the full index. Results indicate that the proposed method is capable of delivering close index tracking while utilizing a substantially smaller set of assets, thereby enhancing capital efficiency and reducing transaction complexity.

This research demonstrates that the integration of time-series clustering and constrained portfolio optimization provides a transparent and effective framework for constructing interpretable, low-dimensional approximations of broad market indices.

2. Literature Review

Several studies in financial time series analysis have examined the effectiveness of clustering methods based on Euclidean distance and Dynamic Time Warping (DTW). Puspita and Zulkarnain (2020) demonstrated that DTW-based clustering significantly outperforms traditional Euclidean distance methods in grouping stock data, particularly when evaluated using metrics such as the Silhouette Score and computational efficiency. Similarly, Aqsari, Prastyo and Puteri Rahayu (2022) confirmed the superiority of DTW over Euclidean-based approaches in capturing temporal dynamics within financial data. Additionally, Chang, Lin, Koc, Chou and Huang (2016) applied Affinity Propagation clustering to the S&P 500 stocks, demonstrating that selecting representative stocks from these clusters can lead to more stable and diversified investment portfolios.

Collectively, these studies underscore the importance of selecting appropriate distance metrics and clustering techniques when analyzing financial time series, particularly for tasks related to stock grouping and portfolio construction.

2.1. Overall Hypothesis

Hypothesis 1. *Using a 10-year daily dataset, constrained by computational resources, is hypothesized to diminish clustering robustness. Expanding the sample size would likely capture a more comprehensive range of time-series dynamics, improving accuracy and stability.*

Hypothesis 2. *Ensuring stationarity, normalization, and continuity in financial time series is considered essential to prevent distortions in distance calculations. Techniques such as z-score normalization and differencing are expected to enhance clustering quality and interpretability.*

Hypothesis 3. *Clustering performance is hypothesized to depend on the compatibility between the chosen distance measure and the algorithm. For instance, Dynamic Time Warping (DTW) is more effective when combined with hierarchical or k-medoids methods, rather than centroid-based algorithms like k-means.*

Hypothesis 4. *Hierarchical methods are hypothesized to be less sensitive to changes in the distance metric than partition-based approaches. DTW is expected to yield more temporally aligned groupings than Euclidean distance, albeit at increased computational cost.*

3. Methodology

Source Code The authors developed all code and implementations for this study. While external libraries (e.g., SciPy, scikit-learn) are used, the core methods, algorithms, and preprocessing were fully customized for applying an Clustering methods to the data. The code is accessible at this link.

The clustering pipeline implemented in `clustering_class` follows five stages: data validation, preprocessing, similarity computation, clustering, and performance evaluation.

Integrity of input time series is verified and any detected anomalies or missing-data patterns result in an error to prevent unreliable analysis.

Preprocessing transforms each series by computing percentage changes item

$$X'_t = \frac{X_t - X_{t-1}}{X_{t-1}},$$

where X'_t is the daily percentage return at time t,
X_t is the adjusted closing price at time t.

and backward-filling NaN values; percentual change normalization is applied when using both Euclidean-based KMeans and DTW-based methods.

Two similarity measures are supported: Euclidean distance

$$d_{\mathrm{E}}(X, Y) = \sqrt{\sum_{t=1}^{n} (X_t - Y_t)^2},$$

where X_t and Y_t are the values at time t in sequences X and Y, respectively,

and Dynamic Time Warping (DTW), which aligns sequences by warping their time axes. A cost matrix

$$D_{i,j} = (X_i - Y_j)^2$$

is constructed, then cumulative cost is computed via

$$C(i,j) = D_{i,j} + \min\{C(i-1,j), C(i,j-1), C(i-1,j-1)\},$$

with boundary condition $C(1,1) = D_{1,1}$. The optimal warping path satisfies boundary, monotonicity, and step-size constraints, and its cost defines the DTW distance $\sqrt{C(n,m)}$. A Sakoe–Chiba band of radius r restricts alignments to $|i - j| \leq r$, reducing computation and preventing pathological warps.

The module FastDTW approximates DTW in linear time by performing recursive low-resolution alignments and refining at finer levels, using a default window radius of 5% of the series length. The resulting pairwise distance matrix is cached using a hash of metric parameters and saved to disk (e.g., `fastdtw_distance_matrix_radius5.npy`) for reuse.

Hierarchical clustering uses complete linkage

$$d_{\mathrm{linkage}}(C_i, C_j) = \max_{x \in C_i, y \in C_j} d(x, y),$$

where C_i and C_j are the clusters, and $d(x, y)$ is the distance between points x and y from the respective clusters,

with the number of clusters K set manually and clusters extracted by cutting the dendrogram. KMeans clustering minimizes

$$\sum_{k=1}^{K} \sum_{x \in C_k} \|x - \mu_k\|^2$$

where μ_k is the centroid of cluster C_k, using only the Euclidean metric. Both methods accept user-defined parameters for cluster count and linkage type.

Cluster quality is evaluated via silhouette scores

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}},$$

where $a(i)$ and $b(i)$ are the mean intra-cluster and nearest-inter-cluster distances, respectively.

Economic metrics such as total return, average return, and variance are aggregated by cluster to assess financial relevance. Optimal cluster count is chosen using the elbow method by plotting

$$\mathrm{SSE}(K) = \sum_{k=1}^{K} \sum_{x \in C_k} \|x - \mu_k\|^2$$

versus K, with the elbow indicating diminishing variance reduction, and dendrogram inspection. Final review includes plots of series grouped by cluster under both Euclidean and DTW metrics for comprehensive analysis.

4. Portfolio Construction and Evaluation Framework

The quality of the dynamic–time–warping (DTW) clusters is assessed through the construction of a *representative* portfolio composed of 50 equities.

Because clustering algorithms are inherently unsupervised and therefore do not provide an intrinsic measure of explanatory power, the capacity of a cluster-based portfolio to replicate the S&P 500 constitutes an external validation test. The hypothesis formulated in this study suggests that if the clustering procedure succeeds in capturing the essential temporal dynamics of index components, a portfolio that balances representatives across clusters should display price behaviour comparable to that of the benchmark.

4.1. Selection of Representative Assets

Let the set S of S&P 500 constituents be partitioned into K disjoint clusters $C_1, \dots, C_K$ obtained with a DTW-based similarity metric. For cluster C_k containing N_k equities, the standardised price trajectories are denoted $\mathcal{T}k = X_i^{(k)}i = 1^{N_k}$ with $X_i^{(k)} \in \mathbb{R}^T$. The empirical cluster centroid is

$$\bar{X}^{(k)} = \frac{1}{N_k} \sum_{i=1}^{N_k} X_i^{(k)}$$

A measure of representativeness is defined as the Euclidean distance between an individual trajectory and the centroid,

$$d_i^{(k)} = \|X_i^{(k)} - \bar{X}^{(k)}\|_2, \qquad i = 1, \dots, N_k$$

Assets in C_k are ranked by ascending $d_i^{(k)}$, and the $n = \lfloor 50/K \rfloor$ equities with the smallest distances, where $\lfloor 50/K \rfloor$ denotes the integer part of $50/K$, are retained. This procedure ensures that the most central and archetypal members of each cluster are included in the candidate universe.

4.2. Background on Mean–Variance Portfolio Theory

Mean–variance optimisation, introduced by Markowitz (1952), frames portfolio selection as a quadratic optimisation problem in which the variance of portfolio return is minimised for a predetermined level of expected return. Under the assumptions of frictionless markets and jointly elliptically distributed returns, the feasible set of portfolios in $(\sigma_p, , \mathbb{E}[R_p])$ space is a convex parabola. The upper branch of this parabola constitutes the *efficient frontier*, whose elements dominate all other feasible portfolios with identical risk or return.

Several concepts are central to the Markowitz framework:

- **Global Minimum-Variance Portfolio (GMVP).** The left-most point of the frontier, obtained by solving the optimisation problem without an expected-return constraint.
- **Tangency Portfolio.** When a risk-free asset with rate r_f is available, the Sharpe-ratio-maximising portfolio is located at the tangency between the frontier and the capital market line (CML); capital is then allocated along the CML rather than the frontier.
- **Long-Only Frontier.** Imposing non-negative weights restricts the feasible set to a sub-region of the full frontier. The present study adopts this restriction to reflect common implementation constraints.

4.3. Optimisation Problem

Denote by $\mu \in \mathbb{R}^n$ the vector of expected excess returns of the $n = 50$ selected equities and by $\Sigma \in \mathbb{R}^{n\times n}$ their covariance matrix. Portfolio weights are represented by $w \in \mathbb{R}^n$. The optimisation implemented in this study seeks the minimum-variance portfolio that matches the annualised return r^* of the S&P500 over the calibration window, under a full-investment and long-only constraint:

$$\begin{aligned}
&\min_{w \in \mathbb{R}^n} \; w^\top \Sigma\, w \\
&\text{s.t.} \quad \mu^\top w = r^*, \\
&\qquad\;\; \mathbf{1}^\top w = 1, \\
&\qquad\;\; w \geq 0
\end{aligned}$$

The problem is convex; sequential quadratic programming (SQP) therefore converges to the global optimum. Expected returns are estimated by the mean historical return, and the covariance matrix is computed from daily log-returns, both annualised with a factor of 252 trading days.

4.4. Evaluation Procedure

We evaluate the portfolio optimization strategy using two distinct modes:

In-sample (Lookback): In this mode, the portfolio weights are calibrated using a specific historical data window. The performance, measured by the realized cumulative return of the optimized portfolio, is then evaluated over the *same* historical window and benchmarked against the index trajectory during that period. This assesses the model's fit to the calibration data.

Out-of-sample (Lookforward): Here, the portfolio weights are calibrated on a historical window, identical to the in-sample case. However, these calibrated weights are subsequently applied to a *forward* window of equal length, immediately following the calibration period. This protocol rigorously evaluates the out-of-sample robustness of both the clustering procedure and the optimization routine by testing their predictive power on unseen data.

Daily portfolio returns are computed as $R_{p,t} = w^\top r_t$, where w represents the vector of portfolio weights (calibrated according to the evaluation mode) and r_t is the vector of asset returns observed on day-t.

The performance of the resulting portfolio return series is assessed using several standard metrics:

- Cumulative Growth: The total growth achieved over the evaluation window.
- Annualized Return: The geometric average annual rate of return.
- Annualized Volatility (σ_p): The annualized standard deviation of the portfolio returns, measuring risk.
- Sharpe Ratio (SR): A measure of risk-adjusted return, calculated as:

$$\mathrm{SR} = \frac{\mathbb{E}[R_p] - r_f}{\sigma_p}$$

where $\mathbb{E}[R_p]$ is the expected portfolio return, σ_p is the portfolio volatility, and the risk-free rate r_f is assumed to be zero for this study.

The following image, **Figure 1** describes visually the workflow, which will be described in detail in the following section.

Empirical results, comparing the performance under both evaluation modes, are presented graphically and in tabular form in Section 6.5.

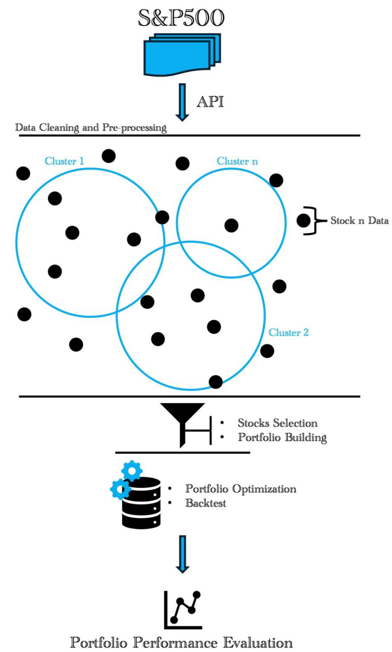

Figure 1: Workflow pipeline.

5. Data Preprocessing and Analysis

5.1. Data Retrieval

Historical stock price data were obtained via the Yahoo Finance API using the `yfinance` library. Daily data for the S&P 500 index (ticker `ĜSPC`) covering the past ten years were downloaded. The DataFrame index was reset, and the `Date` column was converted to `datetime` format. Initial inspection was performed with standard Pandas commands, and the processed DataFrame was serialized to a pickle file for subsequent analysis.

5.2. Data Cleaning

Missing values were quantified by computing the count of null entries per column. Stocks with more than 30% missing observations (31 tickers) were removed to preserve dataset integrity. Remaining missing values were handled using forward- and backward-filling to maintain continuity in the time series. Both the raw and cleaned datasets were saved as pickle files to facilitate efficient reuse in later stages.

5.3. Exploratory Data Analysis

To gain initial insights into the data and prepare for DTW clustering, we performed an exploratory data analysis on the cleaned dataset. We calculated key descriptive statistics for the daily returns of each stock to understand their basic statistical properties.

The daily returns were computed as the percentage change in adjusted closing prices. Based on our analysis using a custom `StatisticalAnalysis` module, the mean daily average return across all stocks is approximately 0.0646%, indicating small positive returns on average. The low standard deviation (approximately 0.0394%) of these returns suggests that most stocks exhibit similar daily return patterns.

In addition, 205 of the S&P 500 stocks display positive skewness in their daily returns, indicating a higher probability of experiencing large positive returns.

All stocks exhibit positive kurtosis, implying that extreme events (both positive and negative) are more likely than predicted by a normal distribution.

Interpretation: The modestly positive average return (0.0646%) aligns with expectations for large-cap equities in mature markets, reflecting long-term growth trends. The relatively low standard deviation (0.0394%) across the cross-section suggests homogeneous short-term volatility, which can be interpreted as evidence of systemic co-movement and market integration—factors critical in justifying the use of time-series similarity measures for clustering.

The presence of positive skewness may reflect the equity premium associated with growth-oriented firms or sectors, and has implications for risk-seeking portfolio construction strategies.

Moreover, the universally positive kurtosis suggests fat-tailed behavior, where extreme return events occur more frequently than under Gaussian assumptions. From a financial standpoint, this feature is consistent with the empirically observed heavy tails in equity markets, underscoring the need for robust distance metrics, such as DTW, that accommodate temporal and distributional irregularities when performing time-series clustering.

6. Implementation

Following the methodology presented in Section 3, all project components were combined to derive the final clustering outputs.

First, the pickled dataset was loaded and supplied to an instance of the `TimeSeriesClustering` class. Percentage change was chosen as the representation method, as it standardizes each series across varying price levels and ensures that clustering is driven by relative dynamics rather than absolute scale differences.

6.1. DTW Hierarchical Clustering

The code implements hierarchical clustering on financial time series data using the FastDTW (Fast Dynamic Time Warping) distance, which is a computationally efficient approximation of the standard DTW metric. This distance measure is particularly suitable for financial applications, as it accounts for potential temporal misalignments in time series representing asset price dynamics, while offering significant performance improvements over exact DTW.

As we can see in **Figure 2**, the number of clusters is manually specified as **five**, following **visual inspection of the hierarchical clustering dendrogram**, which exhibits a clear separation into five principal branches at a dissimilarity threshold of approximately 60. This choice reflects a model selection approach commonly adopted in empirical asset clustering, where visual diagnostics and domain-specific considerations complement algorithmic criteria.

The pipeline first checks for the existence of a previously computed FastDTW distance matrix on disk. If found, it is loaded to avoid redundant computation—a critical efficiency in cross-sectional studies involving high-dimensional financial time series. Otherwise, the matrix is computed using **complete linkage** hierarchical clustering with **FastDTW** as the dissimilarity metric, then cached both in memory and on disk for subsequent reuse. A unique cache key ensures consistency and modular reuse across clustering experiments.

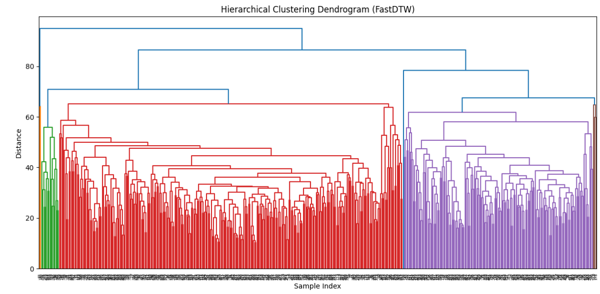

Figure 2: Dendogram Visualization for Hierarchical Clustering: Identifying the optimal number of clusters.

Hierarchical clustering is then executed with the specified configuration. The resulting cluster labels and a **silhouette score** of **0.29** are reported. While the silhouette coefficient is a standard metric for evaluating clustering compactness and separation, its interpretability is limited under **non-Euclidean metrics** such as DTW, where the lack of the triangle inequality and centroid definition weakens its theoretical grounding.

To address this, the **Cophenetic Correlation Coefficient (CPCC)** is computed, yielding **0.3412**. CPCC quantifies the correlation between the original pairwise distances and those implied by the dendrogram structure, offering a more suitable diagnostic for hierarchical models. The moderate CPCC value indicates that the dendrogram preserves some, but not all, of the underlying distance structure, which may be expected given the complexity and potential noise in financial time series data.

Overall, this procedure facilitates the unsupervised discovery of asset groupings with analogous temporal behavior under elastic alignment, supporting interpretable structural segmentation in financial markets where time dynamics are nonlinearly distorted and traditional distance metrics fail to capture meaningful similarity.

6.2. Euclidean K-Means Clustering

The selection of the optimal number of clusters in KMeans clustering is a crucial step in uncovering meaningful patterns within financial time series data. In this study, the elbow method is utilized to determine the appropriate cluster count. Specifically, the elbow curve in **Figure 3** is generated for a range of cluster values (from 1 to 15), with the optimal choice being set at **three** clusters. This decision is supported by the elbow method's visualization, which identifies a significant change in the rate of decrease of the within-cluster sum of squares, signaling diminishing returns beyond three clusters.

Following the elbow method's indication, the KMeans clustering algorithm is executed with `n_clusters=3`, employing `k-means++` initialization to prevent poor cluster centroid initialization and `n_init=10` to ensure that the algorithm converges to a stable solution. This configuration is widely used in empirical asset clustering studies, as it mitigates issues of suboptimal initialization, which can lead to inaccurate

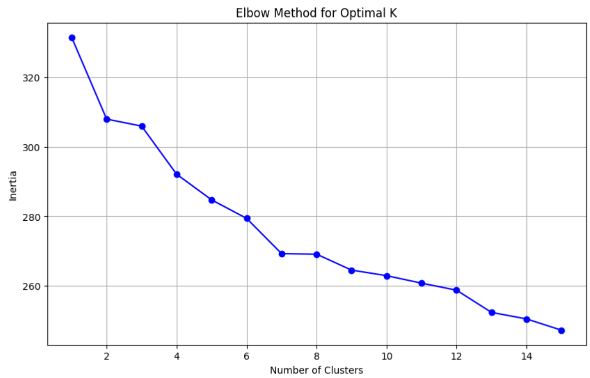

Figure 3: Elbow Method for KMeans Clustering: Identifying the optimal number of clusters.

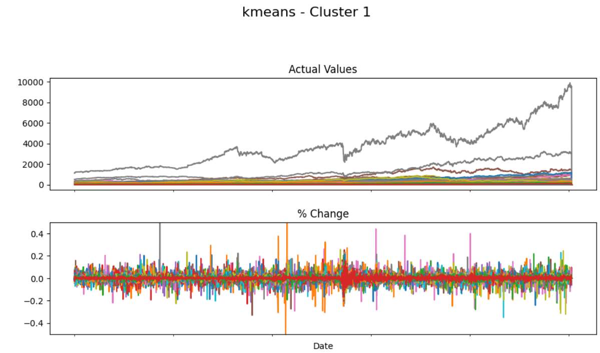

Figure 4: Clustering Results, Cluster 1.

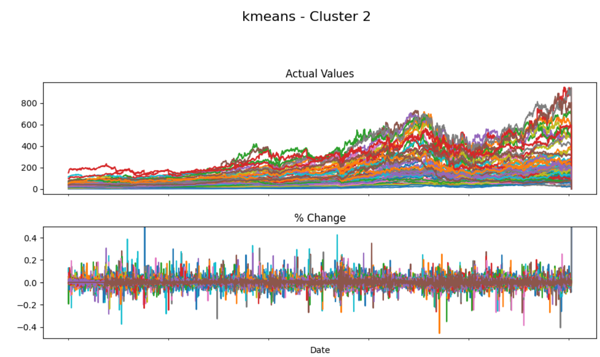

Figure 5: Clustering Results, Cluster 2.

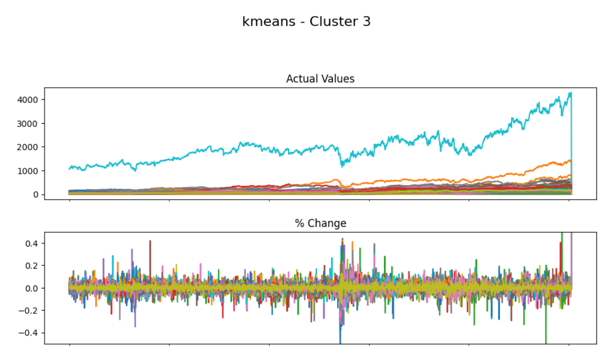

Figure 6: Clustering Results, Cluster 3.

clustering of financial instruments with complex and noisy time.

The results of the clustering are examined through the silhouette score, which quantifies how similar each data point is to its own cluster relative to other clusters. In this case, the silhouette score is **0.0869**, a relatively low value. In the context of quantitative finance, such a score suggests that the financial assets in the dataset are not clearly separable into well-defined clusters. A silhouette score closer to 1 would indicate a high degree of separation between clusters, while a score near 0 implies that the clusters may be overlapping or poorly defined. The score of 0.0869 thus reflects the complexity and possible noise inherent in the data, making it challenging to form distinct asset groupings based solely on the features considered by the clustering algorithm.

While the clustering structure reveals three groups of assets, the moderate silhouette score indicates that further refinement is needed. Potential avenues for improvement include the incorporation of additional features, such as market factors or risk metrics.

The KMeans clustering results serve as a preliminary and comparative step in the analysis, offering insights into how assets may behave in relation to each other under the chosen metric.

6.3. Clustering Results

The **Euclidean K-Means clustering** method groups assets based on their risk-return profiles. In **Figure 4** Cluster 1 (244 assets) shows moderate returns with low volatility, suggesting stability. In **Figure 5** Cluster 2 (66 assets) has higher returns and volatility, indicating a more aggressive, high-risk, high-reward nature. In **Figure 6** Cluster 3 (159 assets) represents conservative assets with lower returns and moderate volatility.

The **DTW hierarchical clustering** method captures temporal relationships. In **Figure 7** Cluster 1 (443 assets) is stable with moderate returns and low volatility, similar to Cluster 1 in the Euclidean analysis. **Figure 8** Cluster 2 (2 assets) reveals outliers with high returns and volatility, reflecting high-risk, high-reward assets. **Figure 9** Cluster 3 (23 assets) shows strong returns with higher risk, while in**Figure 10** Cluster 4 (1 asset) is an extreme outlier with exceptional performance. This method highlights more dynamic asset behavior, identifying both stable and exceptional outliers.

Table 1 summarizes the indicators computed for each clustering method and for every cluster.

6.4. Post-Clustering Statistical Analysis

To better interpret the structure and significance of the clusters produced by the hierarchical clustering with *Dynamic Time Warping* (DTW)—which constitutes the central

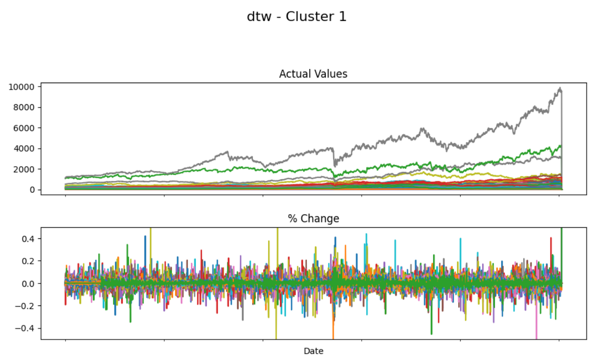

Figure 7: Clustering Results, Cluster 1.

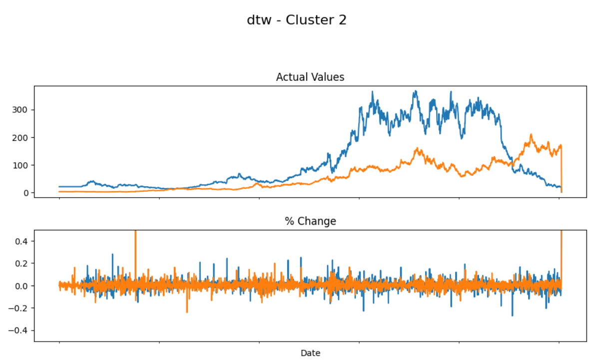

Figure 8: Clustering Results, Cluster 2.

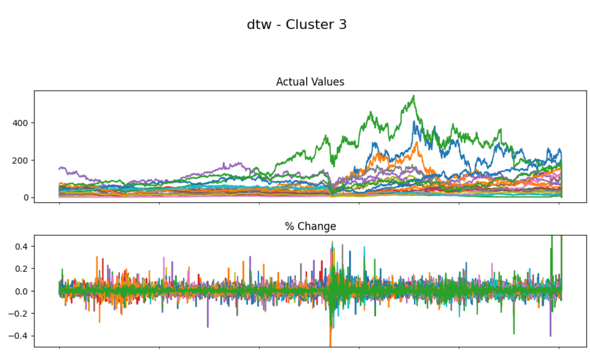

Figure 9: Clustering Results, Cluster 3.

clustering approach in this study—a comprehensive post-clustering statistical analysis was conducted. This examination serves to validate and visualize the internal coherence and external separation of the resulting groups of assets.

6.4.1. Heatmap matrix of Clusters

For each relevant DTW-derived cluster, a heatmap of the DTW distance matrix was constructed to summarize pairwise dissimilarities visually. Cohesive clusters typically appear as low-distance blocks along the diagonal, while off-diagonal structures offer insight into inter-cluster relationships.

Figure 11 shows a heatmap of pairwise distances, revealing that most stocks in the index appear highly similar.

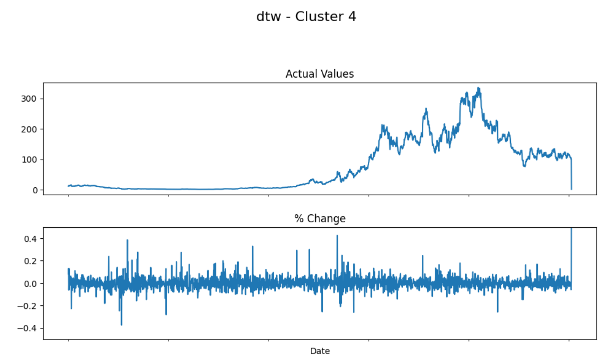

Figure 10: Clustering Results, Cluster 4.

Cluster	Size	Mean Return	Volatility	Total Return
Cluster Analysis for Euclidean Hierarchical Clustering				
1	244	0.000600	0.012682	2.844667
2	66	0.001066	0.018040	13.014381
3	159	0.000574	0.015923	1.740039
Cluster Analysis for `FastDTW` Hierarchical Clustering				
1	443	0.000636	0.014592	3.096884
2	2	0.001559	0.022968	31.510126
3	23	0.000906	0.023953	16.787884
4	1	0.002197	NaN	8.652964

Table 1
Cluster Analysis for Euclidean and FastDTW Hierarchical Clustering

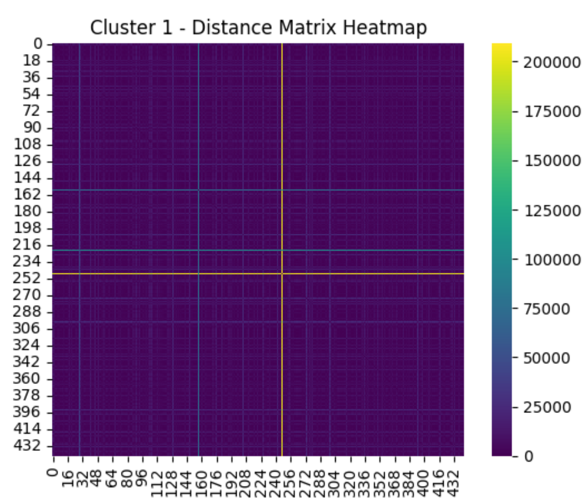

Figure 11: Clustering Results, Cluster 2.

This observation may stem from several factors: the time range of the data, the fact that many stocks in the index belong to similar sectors, the use of a market-cap-weighted index, and the specific constraint band used in the DTW calculation (e.g., Sakoe–Chiba radius). These elements can reduce variability across distance measures, leading to an overall appearance of closeness.

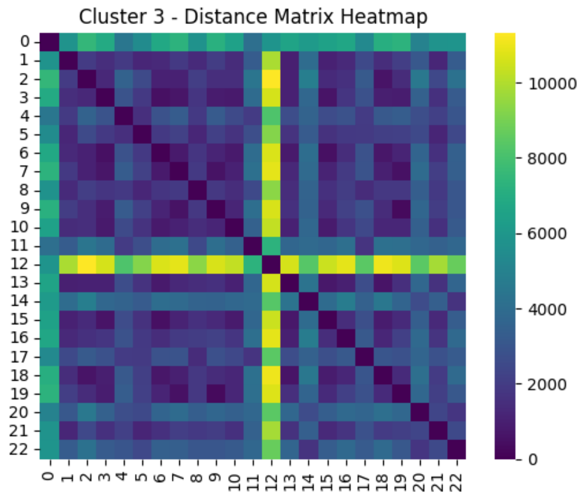

Figure 12: Clustering Results, Cluster 2.

6.4.2. Cluster 3, 23 tickers

Figure 12, on the other hand, includes fewer stocks, providing a clearer distinction between distance values. This allows for more visible clustering patterns and highlights dissimilarities that may be obscured in larger, more homogeneous datasets.

6.4.3. T-SNE Of Overall Clusters

The following figures depict the low-dimensional embeddings of financial time series derived from two distinct clustering methodologies—Dynamic Time Warping (DTW) and K-Means—via t-Distributed Stochastic Neighbor Embedding (t-SNE). This dimensionality reduction technique enables the visualization of high-dimensional temporal behavior of assets in a two-dimensional space while preserving the local structure of the data.

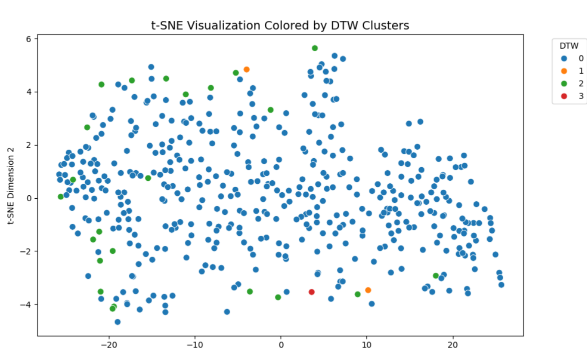

Figure 13: Clustering Results, T-SNE DTW Clsuster.

In **Figure 13**, each point corresponds to a time series representing an individual asset, embedded from its original functional representation into two dimensions via t-SNE. Coloring is performed according to DTW-based clustering, where DTW (Dynamic Time Warping) measures similarity between time series by allowing for elastic shifts in the time dimension, making it particularly suitable for capturing latent temporal alignments.

The predominance of a single cluster (Cluster 0, in blue) suggests that under the DTW distance metric, the majority of financial time series exhibit similar temporal dynamics. This can be attributed to DTW's flexibility in aligning time series with local temporal distortions, leading to tighter grouping and fewer distinguishable macro patterns. Minority clusters (Clusters 1–3) denote more idiosyncratic or anomalous temporal behaviors that deviate under warping constraints. The sparse and dispersed nature of these minority groups highlights DTW's sensitivity to temporal dissimilarities, even when statistical properties are globally similar.

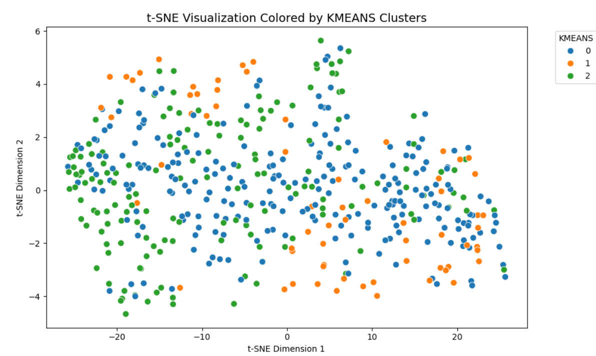

Figure 14: Clustering Results, T-SNE KMEANS Clusters

In **Figure 14**, constrasting to the DTW-based visualization, the K-Means clustering yields a more even and well-distributed segmentation into three main groups (clusters 0–2). This reflects K-Means' reliance on Euclidean distances in a fixed feature space, where separation is driven more by overall shape and volatility levels than by temporal alignment. The visual separability of clusters in this projection suggests that K-Means effectively identifies structural or behavioral classes within the asset universe, such as high-volatility versus low-volatility equities or bullish versus bearish patterns.

These t-SNE projections illuminate fundamental distinctions between time series clustering methodologie.. DTW, by design, aligns series based on latent temporal warping, grouping assets that follow similar temporal rhythms irrespective of phase shifts. This makes it particularly advantageous for capturing regime-switching behavior or sectoral synchronization with lag. Conversely, K-Means excels in segmenting series with distinct statistical properties, such as differing variances or linear trends, often corresponding to structural differences in asset fundamentals or macroeconomic exposure.

The divergence in cluster distributions between **Figures 12** and **Figures 13** underscores the methodological importance of the underlying similarity metric. From a portfolio management perspective, this has profound implications: DTW clustering may enhance strategy design for temporal regime tracking or relative value strategies, while K-Means clustering provides a basis for volatility targeting or diversification through statistical decorrelation.

The use of t-SNE here is not merely illustrative but analytical. By projecting high-dimensional, temporally rich data into an interpretable space, it reveals non-trivial latent structures that would otherwise be obfuscated. Moreover, the visual coherence (or lack thereof) of clusters validates the internal consistency of each method, offering an empirical lens for assessing cluster robustness.

6.5. Performance Evaluation

Finally, as outlined in Sections 4.3 and 4.4, the optimization procedure was employed to determine the top *n* representative stocks along with their corresponding weights.

For the look-back evaluation, the relevant weights were as follows: JNJ (Johnson & Johnson) with a weight of 0.29948, PEP (PepsiCo) with 0.17838, WM (Waste Management) with 0.18649, JKHY (Jack Henry & Associates) with 0.08904, UPS (United Parcel Service) with 0.04044, CME (CME Group) with 0.03995, AIZ (Assurant) with 0.03723, TRGP (Targa Resources) with 0.0, and NVDA (NVIDIA) with 0.0. For the look-forward evaluation, the corresponding relevant weights were identical.

The identical weights arise from their calculation based on the same historical data. The application of these weights to both in-sample and out-of-sample periods provides evidence that, to a certain extent, the optimal stock allocations exhibit consistency across past and future market conditions. This suggests that the optimized weights possess a certain level of robustness, remaining relevant not only in the historical context but also in their predictive capacity for future performance.

The resulting portfolio was then evaluated against the benchmark using both retrospective and prospective temporal analysis. The outcomes are presented below.

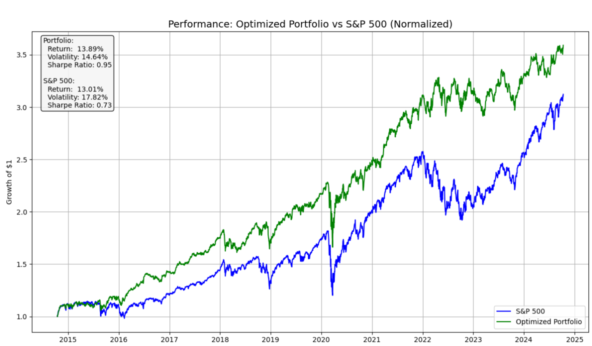

Figure 15: Optimization result, look-backward mode.

Figure 15 illustrates the comparative performance in a look-backward (in-sample) evaluation over the period from 2015 to the beginning of 2025. The plot displays the cumulative growth of an initial $1 investment, normalized at the start date. The optimized portfolio (green line) consistently outperforms the S&P 500 index (blue line) throughout this extensive backtesting period. Performance metrics corroborate this visual evidence: the optimized portfolio achieved an annualized return of 13.89% with a significantly lower annualized volatility of 14.64%, yielding a strong Sharpe ratio of 0.95. In contrast, the S&P 500 benchmark generated a comparable return of 13.01% but with substantially higher volatility (17.82%) and a consequently lower Sharpe ratio of 0.73. These in-sample results indicate the efficacy of the optimization procedure in constructing a portfolio with superior historical risk-adjusted returns.

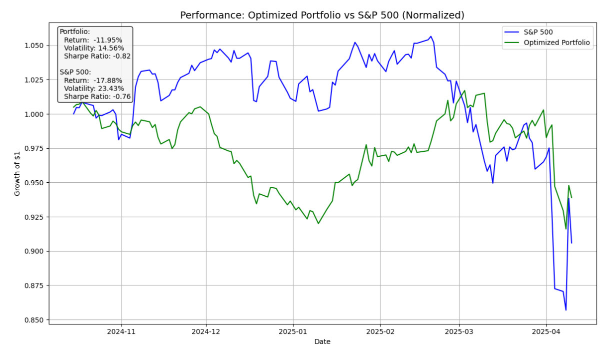

Figure 16: Optimization result, look-forward mode.

Figure 16 presents the performance evaluation in a look-forward (out-of-sample) mode, assessing the optimized portfolio's behavior on data from November 2024 to April 2025. This period, subsequent to the data potentially used for optimization, was characterized by challenging market conditions. Both the optimized portfolio and the benchmark experienced negative returns. However, the optimized portfolio demonstrated notable relative outperformance, incurring a smaller loss (-11.95%) compared to the benchmark (-17.88%). Crucially, the portfolio also exhibited significantly lower volatility (14.56%) than the S&P 500 (23.43%) during this downturn. While both strategies yielded negative Sharpe ratios (-0.82 for the portfolio vs. -0.76 for the benchmark), the portfolio's ability to mitigate losses more effectively and operate with lower risk highlights valuable defensive characteristics of the strategy.

6.6. Limitations and Practical Considerations

In both the look-backward and look-forward evaluations, the optimized portfolio demonstrates superior relative performance compared to the benchmark. This overall outperformance, particularly the enhanced risk-adjusted returns in the longer backtest and the capital preservation during the downturn, can be attributed to the integration of several methodologically rigorous components. Specifically, the clustering of assets via Dynamic Time Warping (DTW) allowed for the identification of latent temporal structures in asset price trajectories, capturing non-linear alignments that are often missed by traditional distance metrics. This enabled the selection of representative assets that retain the essential dynamics of broader market segments. These representatives were then used as the input universe for a constrained mean-variance optimization problem, with the expected returns and covariance matrix estimated from historical data and the target return anchored to that of the S&P 500.

However, despite the promising empirical results, several limitations must be acknowledged when interpreting these findings. A primary concern relates to the practical implementation and the stationarity assumptions inherent in the framework. Even in the lookback configuration, the use of static clusters and optimization parameters derived from historical data does not fully account for dynamic changes in market structure or evolving cross-asset relationships, which may lead to model drift over time.

Furthermore, while the look-forward test provides out-of-sample validation, its realism is constrained by the fact that the optimal cluster structures and model parameters identified from past data may not hold perfectly in future, unseen market regimes. Relying on historically derived inputs inevitably introduces challenges related to non-stationarity.

Moreover, the DTW-based clustering, while powerful in capturing shape similarities, does not inherently consider the full spectrum of stochastic properties relevant to financial time series, such as volatility clustering or tail dependence, which could materially affect portfolio risk assessment.

Finally, the reliance on historical average returns as proxies for expected returns, and the assumption of a stationary covariance structure within the mean-variance optimization, introduce further well-known sources of model risk and estimation error. Markowitz optimization is particularly sensitive to estimation errors in both the expected returns and the covariance matrix, often resulting in portfolios that are unstable or overly concentrated. Additionally, it assumes that investors have quadratic utility or that returns are normally distributed—assumptions that often do not hold in real-world financial markets.

Therefore, while the approach demonstrates theoretical and empirical merit under the controlled conditions of this study, its direct application to real-world portfolio management should be undertaken with caution.

7. Conclusion

This study explored the application of Dynamic Time Warping (DTW) for clustering time series data of S&P 500 constituent stocks, highlighting its effectiveness in capturing temporal similarities that are not aligned in a traditional sense. DTW offers a significant advantage over conventional distance measures by accommodating shifts and distortions in time series data, which is crucial in the context of financial markets where such misalignments are common.

In comparison with K-means clustering, which underperformed in this study, DTW demonstrated superior performance in identifying meaningful clusters of stocks with similar return patterns. The inability of K-means to handle the temporal distortions inherent in financial data resulted in suboptimal clusterings that failed to reveal deeper market insights. On the other hand, the DTW-based clustering method was able to more accurately group stocks with similar temporal behavior, thus providing a more refined and meaningful segmentation of the market.

Furthermore, through a comprehensive optimization process and benchmark comparisons with traditional clustering methods, this study empirically validated that DTW-based clustering outperforms K-means in terms of both cluster coherence and predictive ability. The optimized portfolios derived from DTW clusters were shown to offer superior risk-return profiles compared to those based on K-means, demonstrating the practical advantages of using DTW for financial time series analysis.

The results of this study underscore the importance of selecting appropriate distance measures for time series clustering in financial applications. DTW's ability to capture the dynamic nature of financial time series offers significant potential for enhancing portfolio management, risk assessment, and asset allocation strategies.

Future research could build on these findings by incorporating additional financial and macroeconomic factors to further refine the clustering process. Additionally, hybrid methods that combine DTW with machine learning algorithms or other optimization techniques could provide even more robust models for market prediction and trend analysis.

In conclusion, this study contributes to the literature on time series clustering in finance, demonstrating that DTW-based clustering outperforms K-means and providing a foundation for its further application in quantitative finance.

8. Acknowledgements

This work would not have been possible without the support of the Quants at BlackSwan Quants, a student club that brings together enthusiasts of Quantitative Finance. The club fosters an environment where individuals can collaborate on increasingly challenging projects, applying the skills learned throughout their academic journey while building a network of ambitious people always eager to learn new things and meet new people.

Riccardo Polo is gratefully acknowledged for his valuable support during the initial phase of the project, and is extended best wishes for his future studies.

References

Aqsari, H.W., Prastyo, D.D., Puteri Rahayu, S., 2022. Clustering stock prices of financial sector using k-means clustering with dynamic time warping, in: Proceedings of the 2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), IEEE. pp. 503–507. doi:10.1109/ICITISEE57756.2022.10057714.

Chang, C.C., Lin, Z.T., Koc, W.W., Chou, C., Huang, S.H., 2016. Affinity propagation clustering for intelligent portfolio diversification and investment risk reduction, in: Proceedings of the 2016 International Conference on Computational and Business Data (CCBD), IEEE. pp. 145–150. doi:10.1109/CCBD.2016.037.

Markowitz, H.M., 1952. Portfolio selection. Journal of Finance 7, 77–91. doi:10.2307/2975974.

Puspita, P.E., Zulkarnain, Z., 2020. A practical evaluation of dynamic time warping in financial time series clustering, in: 2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS), IEEE. pp. 61–68. doi:10.1109/ICACSIS51025.2020.9263123.