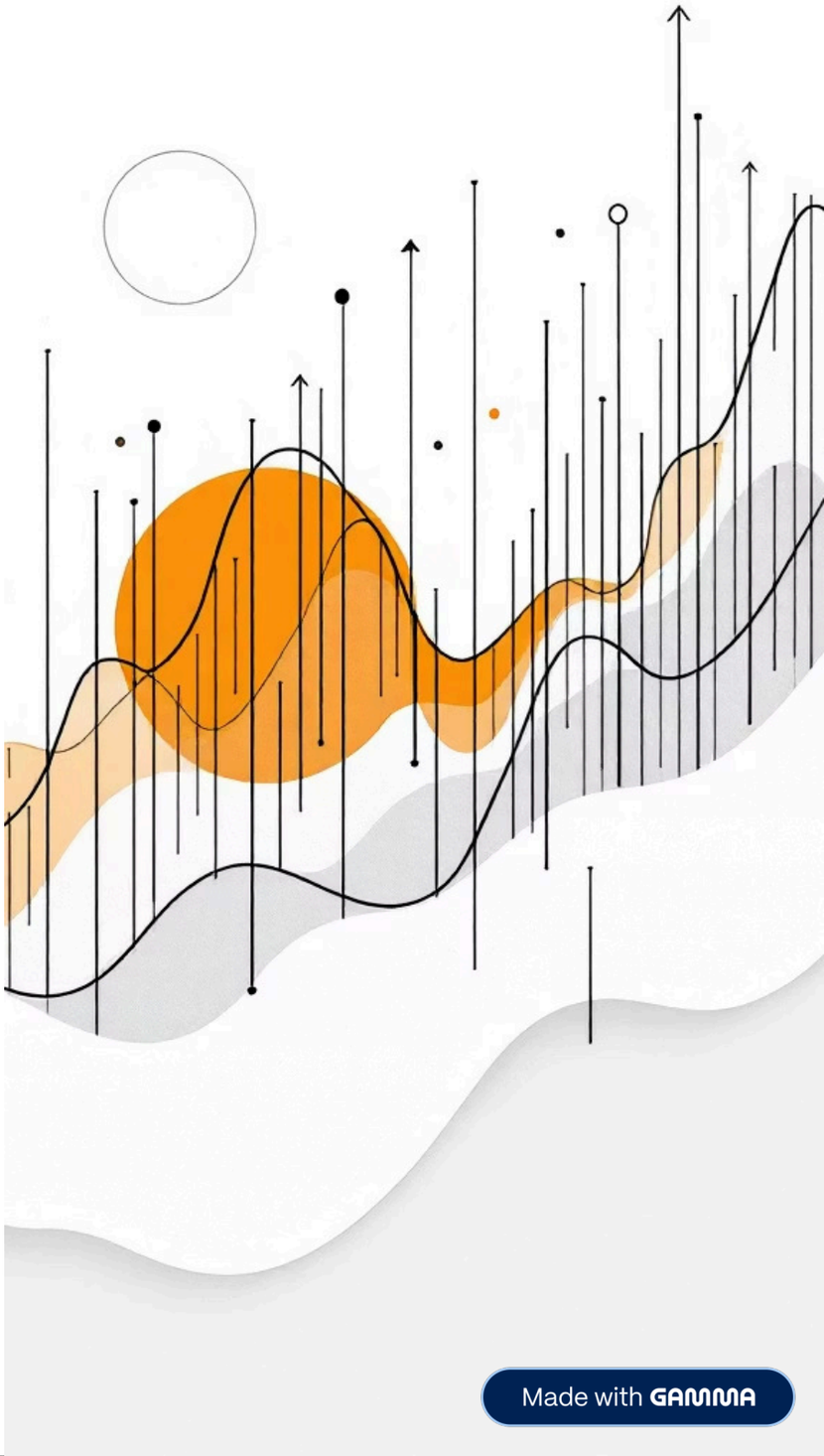
Portfolio Analysis with PCA and K-Means

This presentation details the analysis and clustering of portfolios using three years of real-world data from 50 stocks obtained via Yahoo Finance. We employed Principal Component Analysis (PCA) and K-Means Clustering to uncover hidden connections between securities and identify groups (clusters) with similar risk and return characteristics.

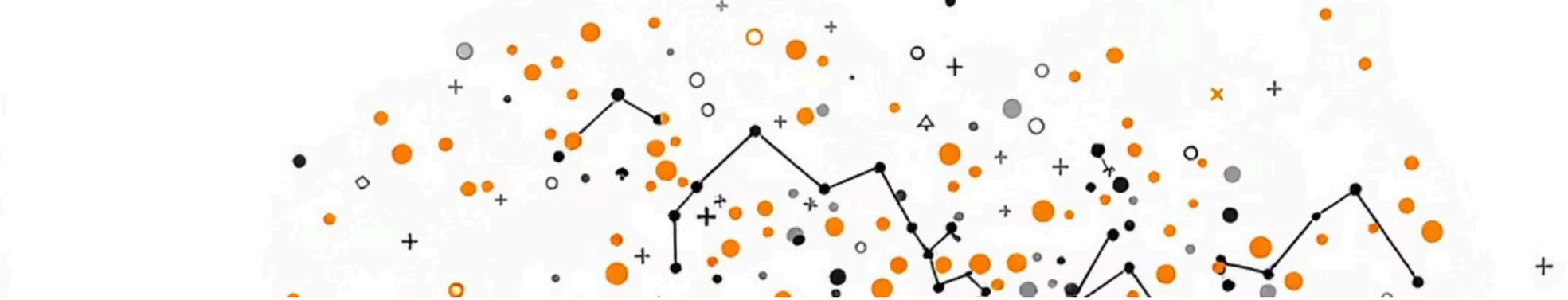
Finally, equal-weighted portfolios were assembled for each cluster, and their key performance and risk indicators were evaluated.



Data and Pre-processing

Our analysis utilised daily adjusted close prices and trading volumes for 50 large and mid-cap stocks across various sectors, including technology, finance, energy, and consumer goods. Three features were calculated for each stock: annual volatility, average volume, and average market cap, all using rolling windows of 1, 3, 5, 10, and 30 days.

All variables underwent standardisation to ensure a mean of 0 and a variance of 1. This crucial step guarantees that variables with different units of measurement have equal weighting in the PCA analysis.



Principal Component Analysis (PCA)

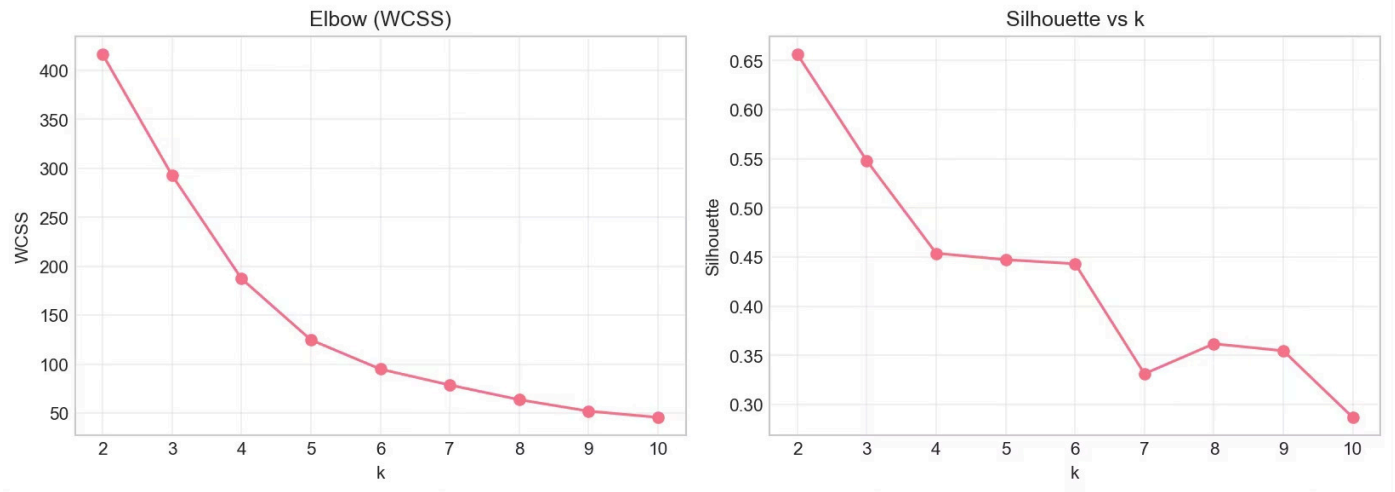
Principal Component Analysis (PCA) was applied to simplify the dataset. The primary goal was to identify principal components that explain over 90% of the total variance, ensuring essential information remained intact.

PC	Variance_Explained	Cumulative_Variance	Eigenvalue
PC1	0.5848365063793055	0.5848365063793055	8.95157917927508
PC2	0.1962749724208807	0.7811114788001863	3.004208761544091
PC3	0.09735029631808391	0.8784617751182702	1.4900555558890387
PC4	0.06603526318427795	0.9444970383025482	1.0107438242491518

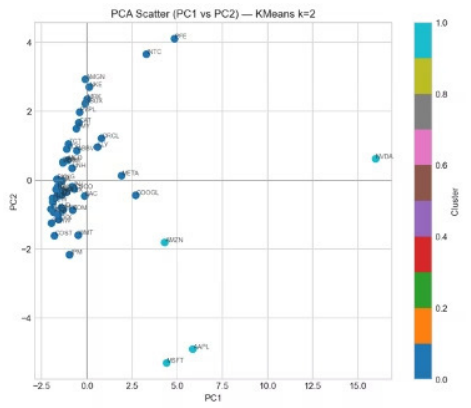
The diagram above illustrates the arrangement of stocks in the space of the first two principal components (PC1 and PC2). Stocks with similar characteristics tend to cluster together, indicating potential similarities in their behaviour regarding volatility, volume, and market capitalisation.

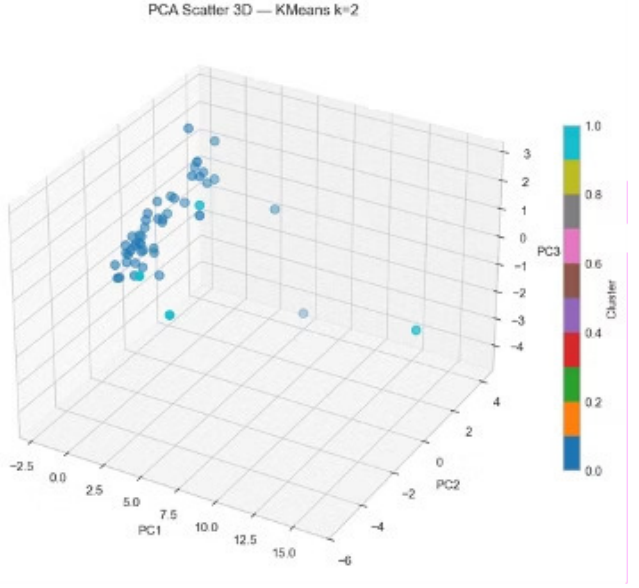
K-Means Clustering

K-Means was applied to the PCA-transformed data. The Elbow method, which examines the change in WCSS (Within Cluster Sum of Squares) relative to 'k', and the Silhouette index, measuring cluster cohesion, were used to determine the optimal number of clusters (k).



The Elbow diagram suggests an optimal 'k' of around 2–3 groups, indicating two main categories of stocks in the dataset.





The 2D and 3D scatter plots display the clusters in the principal component space. The clusters show moderate separation, with Cluster 0 comprising higher market capitalisation and lower volatility stocks, and Cluster 1 containing higher-risk stocks with potentially greater returns.

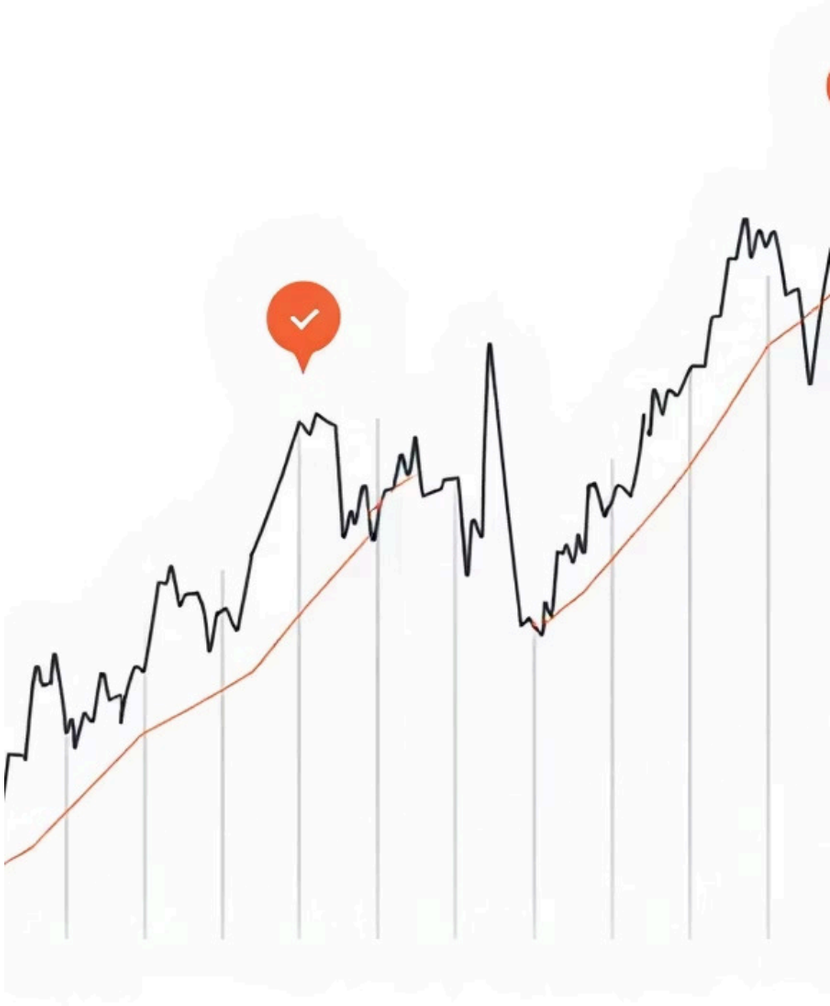
0	46
1	4

Table 1: Cluster Composition

Equal-Weighted Portfolios per Cluster

An equal-weighted portfolio was created for each cluster, ensuring each stock contributed equally. This approach eliminates market capitalisation bias, allowing for a fair comparison between groups. The equity curves below illustrate the value evolution of these portfolios.

The portfolios exhibit distinct behaviours: Cluster 0's portfolio shows a more stable upward trend, while Cluster 1's demonstrates higher volatility but also a greater potential for outperformance.

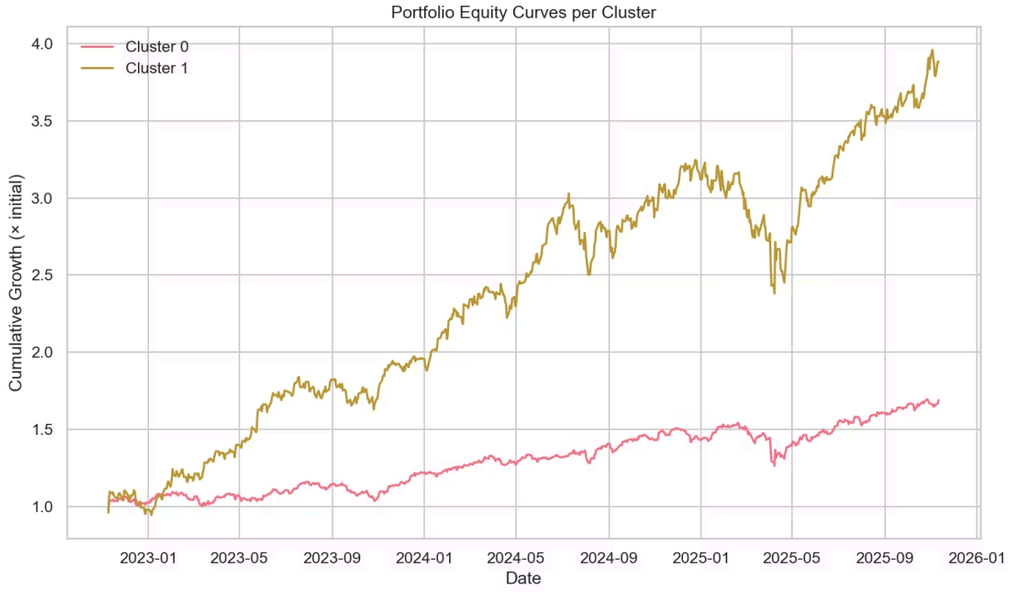


Performance Results

The key performance indicators for each portfolio are presented in the table below.

Cluster	Cumulative Return	Volatility Annualized	Sharpe Ratio	Max Drawdown	Alpha	Beta
0.0	0.6897	0.1432	1.1079	-0.1798	0.0113	0.8461
1.0	2.8823	0.2688	1.68	-0.2663	0.1737	1.5056

Cluster 1's portfolio achieved a higher cumulative return but also exhibited increased volatility. Conversely, Cluster 0's portfolio showed lower risk and a higher Sharpe ratio, indicating better risk-adjusted performance.



Conclusions

The combined use of PCA and K-Means proved highly effective in deciphering stock interactions and grouping them based on shared risk and return characteristics. PCA successfully reduced dataset dimensions while preserving critical information, and K-Means identified groups with distinct investment profiles.

The equal-weighted portfolios highlighted the differentiation between these groups, offering valuable insights for risk management and strategic capital allocation. This methodology provides a robust framework for understanding and optimising portfolio structures.