# Portfolio Clustering Framework

## Hierarchical Clustering for Asset Classification and Portfolio Diversification

**Report Generated:** November 30, 2025 at 14:59
**Analysis Period:** 2022-12-01 to 2025-11-30
**Number of Assets:** 12
**Number of Clusters:** 7

*Educational Open-Source Framework*
*Free Data Sources: Yahoo Finance*

## Executive Summary

This report presents a comprehensive analysis of portfolio diversification using hierarchical clustering techniques. The framework analyzed 12 assets over a 3-year period, identifying 7 distinct asset clusters based on correlation patterns.

**Key Findings:**
• The clustering algorithm successfully identified natural groupings in the asset universe
• Portfolio diversification ratio of 1.86 indicates strong diversification benefits
• Equal cluster weighting strategy ensures balanced exposure across asset classes
• Effective number of independent assets: 9.48

**Methodology:**
• Correlation-based distance metric: $d = \sqrt{(2(1-\rho))}$
• Ward's hierarchical clustering for minimum variance
• Automatic cluster determination via elbow method
• Equal allocation across clusters for maximum diversification

## Asset Universe

| Asset | Cluster | Annual Return | Annual Vol | Sharpe | Weight |
|---|---|---|---|---|---|
| AAPL | C2 | 21.72% | 25.99% | 0.84 | 4.76% |
| AGG | C1 | 4.14% | 6.07% | 0.68 | 7.14% |
| EEM | C3 | 13.32% | 16.36% | 0.81 | 4.76% |
| EFA | C3 | 14.19% | 14.50% | 0.98 | 4.76% |
| GLD | C5 | 28.15% | 16.21% | 1.74 | 14.29% |
| GOOGL | C2 | 39.00% | 30.42% | 1.28 | 4.76% |
| JNJ | C7 | 7.98% | 17.05% | 0.47 | 14.29% |
| JPM | C4 | 30.44% | 22.85% | 1.33 | 14.29% |
| MSFT | C2 | 22.92% | 23.35% | 0.98 | 4.76% |
| TLT | C1 | -1.55% | 15.43% | -0.10 | 7.14% |
| USO | C6 | 0.29% | 29.90% | 0.01 | 14.29% |
| VNQ | C3 | 5.51% | 18.07% | 0.30 | 4.76% |

## Cluster Analysis

The hierarchical clustering algorithm identified 7 distinct clusters based on correlation patterns in asset returns. Each cluster represents a group of assets with similar behavior patterns, providing natural diversification boundaries.

**Cluster Composition:**

**Cluster 1** (2 assets, 14.3% portfolio weight):
Assets: AGG, TLT

**Cluster 2** (3 assets, 14.3% portfolio weight):
Assets: AAPL, GOOGL, MSFT

**Cluster 3** (3 assets, 14.3% portfolio weight):
Assets: EEM, EFA, VNQ

**Cluster 4** (1 assets, 14.3% portfolio weight):
Assets: JPM

**Cluster 5** (1 assets, 14.3% portfolio weight):
Assets: GLD

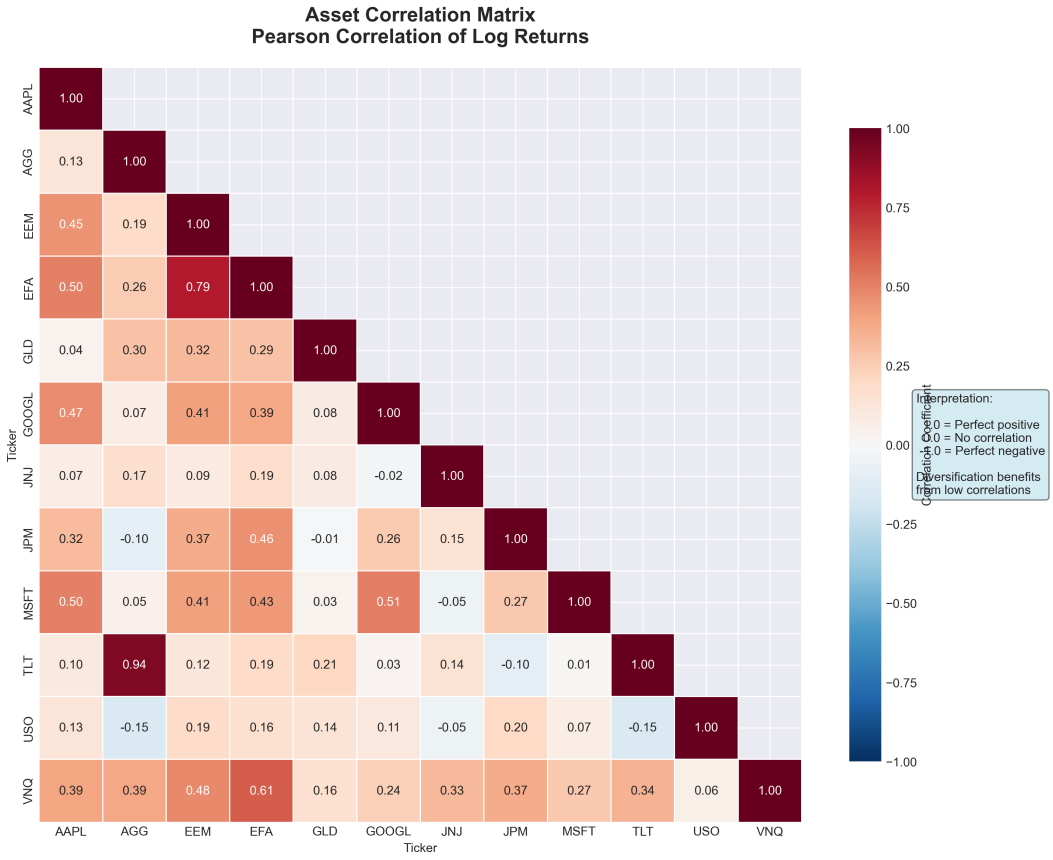
**Cluster 6** (1 assets, 14.3% portfolio weight):
Assets: USO

**Cluster 7** (1 assets, 14.3% portfolio weight):
Assets: JNJ

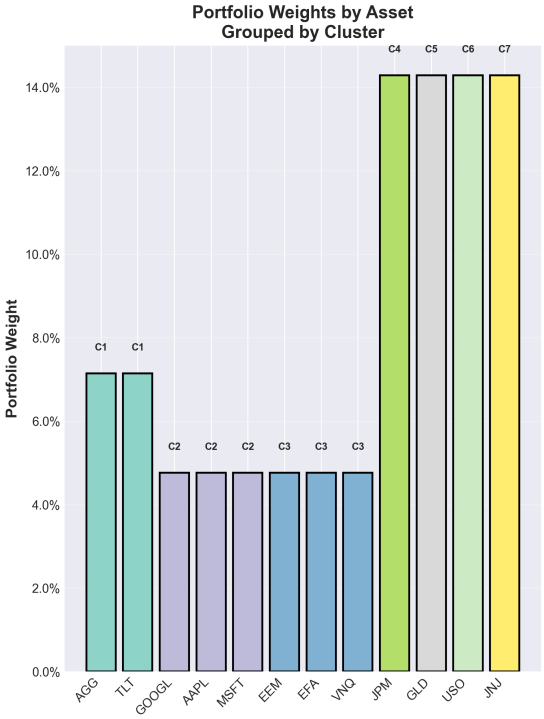
# Visualizations

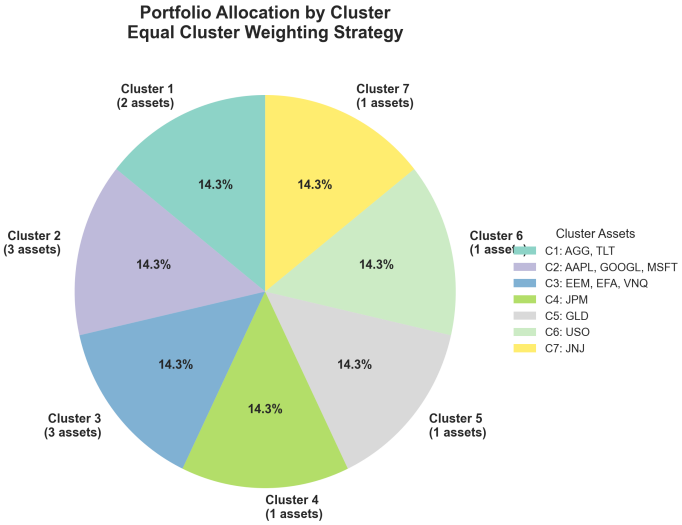
## Hierarchical Clustering Dendrogram

## Correlation Matrix

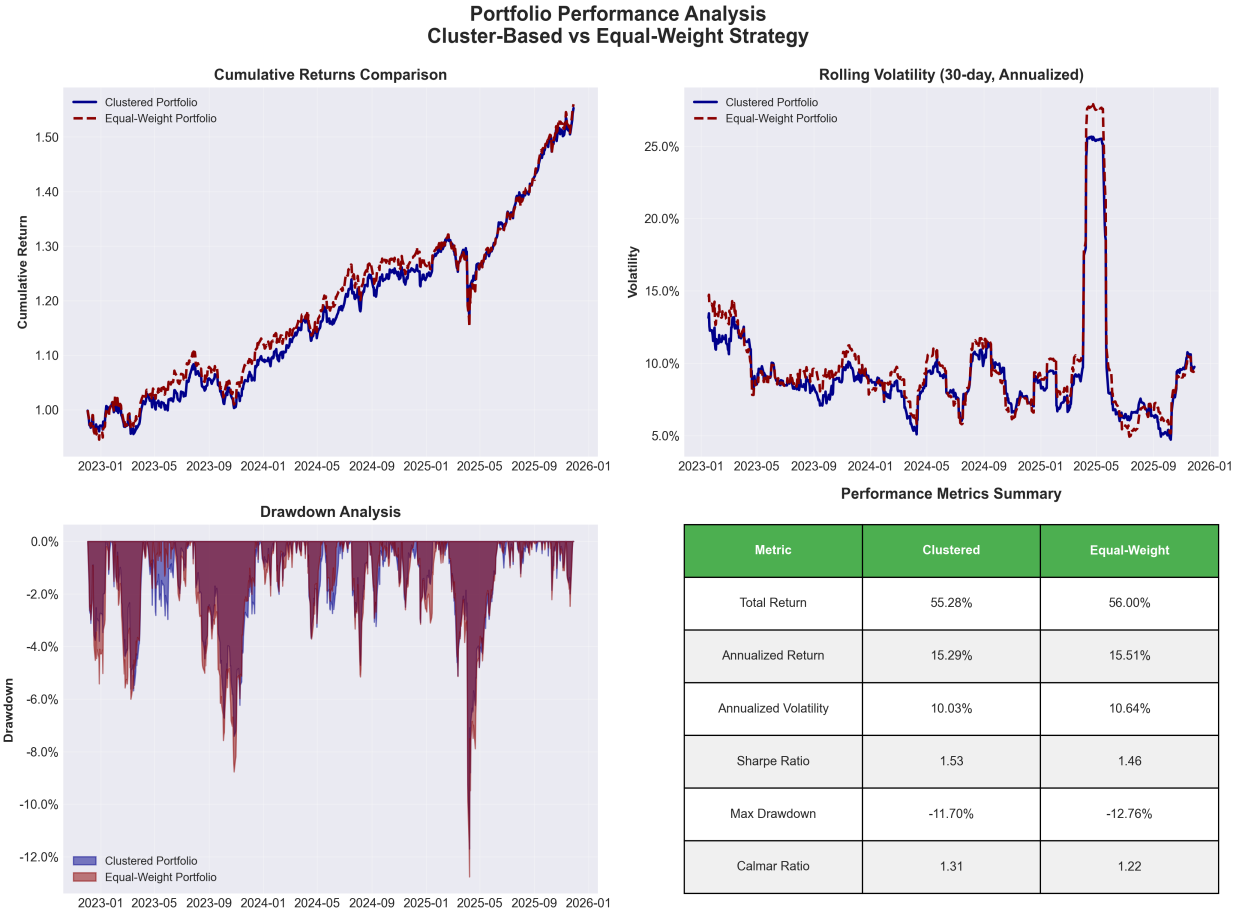


## Portfolio Composition



# Performance Analysis




**Performance Metrics Summary**

| Metric | Clustered | Equal-Weight |
|---|---|---|
| Total Return | 55.28% | 56.00% |
| Annualized Return | 15.29% | 15.51% |
| Annualized Volatility | 10.03% | 10.64% |
| Sharpe Ratio | 1.53 | 1.46 |
| Max Drawdown | -11.70% | -12.76% |
| Calmar Ratio | 1.31 | 1.22 |

## Methodology

**1. Data Collection:**
Historical price data retrieved from Yahoo Finance, a free and reliable data source. Adjusted close prices account for stock splits and dividends.

**2. Return Calculation:**
Logarithmic returns computed for time-additivity and improved statistical properties: $r_t = \ln(P_t / P_{t-1})$

**3. Correlation Analysis:**
Pearson correlation matrix calculated to measure linear relationships between asset returns.

**4. Distance Metric:**
Correlation converted to distance using: $d = \sqrt{(2(1-\rho))}$, ensuring proper metric properties for hierarchical clustering.

**5. Hierarchical Clustering:**
Ward's method applied to minimize within-cluster variance. Optimal cluster count determined via elbow method analyzing merge distances.

**6. Portfolio Optimization:**
Equal allocation across clusters ensures diversification. Within each cluster, assets receive equal weights, implementing a naive diversification strategy that maximizes cluster-level diversification benefits.

## Conclusion and Educational Notes

This framework demonstrates the practical application of hierarchical clustering for portfolio construction and asset classification. The methodology provides several educational insights:

**Key Learnings:**
• Correlation-based clustering reveals natural asset class groupings
• Diversification benefits arise from low inter-cluster correlations
• Equal cluster weighting provides systematic diversification
• Hierarchical methods offer interpretable, transparent classification

**Practical Applications:**
• Asset allocation across diverse investment universes
• Risk management through systematic diversification
• Portfolio rebalancing based on changing correlations
• Educational tool for understanding portfolio theory

**Limitations and Considerations:**
• Historical correlations may not predict future relationships
• Equal weighting ignores expected returns and risk preferences
• Clustering stability should be monitored over time
• Transaction costs and constraints not considered

**Open Source and Educational Use:**
This framework is designed for educational purposes using only free, publicly available data sources. The code is transparent, well-documented, and extensible for further research and learning.