K-Means Clustering and Markowitz Portfolio Optimization of Nifty 200 Stocks

1. Introduction

This project applies k-means clustering and Markowitz portfolio optimization to analyse Nifty 200 stocks from 2021-01-13 to 2024-02-13.. The objective is to group stocks based on fundamental financial metrics, construct optimal portfolios for each cluster, and evaluate their performance against the Nifty 200 index. By focusing on intrinsic financial characteristics rather than market-driven metrics, the study provides actionable insights for investors with varying risk profiles.

The analysis evaluates Reward-to-risk ratio**,** Sharpe Ratio, Treynor Ratio, beta, computes Value at Risk (VaR**)**, and interprets the risk-return trade-offs for each cluster portfolio. Python libraries such as Pandas, NumPy, Scikit-learn, and Matplotlib were used for data preprocessing, clustering, and performance evaluation.

2. Data Collection

Data Sources

1. **Screener Website**: Extracted fundamental financial ratios for Nifty 200 stocks.
2. **NSE Website**: Obtained ticker symbols corresponding to the stocks.
3. **Yahoo Finance API**: Used for historical stock price data.
4. **Upstox API**: Retrieved Nifty 200 index data.
5. **FRED API**: Provided the risk-free rate (10-year Indian bond yield).

Preprocessing

- A structured CSV file was created containing stock tickers and their respective financial ratios.
- Automated removal of stocks with failed data downloads from Yahoo Finance and restructuring of data frame, ensured data completeness.
- Standardized financial features to ensure equal influence during clustering.
- Stocks were sorted alphabetically, by their tickers, to align with optimal weights during portfolio construction and dates of cluster returns and nifty 200 returns were aligned for beta calculation and cumulative returns.
- Time frame variables (start and end dates) were defined at the beginning of the code for consistency and automation.

3. Selection of Factors

Financial Metrics

Eight fundamental factors were selected based on their ability to capture profitability, efficiency, growth potential, income generation, and financial stability:

1. **ROE % (Return on Equity)**: Measures profitability relative to shareholder equity.
2. **ROCE % (Return on Capital Employed)**: Evaluates efficiency in capital utilization.
3. **OPM % (Operating Profit Margin)**: Indicates operational efficiency.
4. **Free Cash Flow (Rs.Cr.)**: Reflects cash generation capabilities beyond accounting profits.
5. **Piotroski Score**: Composite measure of financial health.
6. **EPS Diluted Growth (%)**: Captures earnings growth potential.
7. **Dividend Yield (%)**: Represents income generation for shareholders.
8. **Debt/Equity Ratio**: Measures leverage and financial risk.

Justification

These metrics provide a comprehensive view of a company's financial health:

- Profitability ratios (**ROE**, **ROCE**) highlight management efficiency in generating returns.
- Operational metrics (**OPM**) reveal how well companies manage costs.
- Free Cash Flow indicates liquidity strength, while Piotroski Score measures overall financial stability.
- Growth metrics (**EPS growth**) capture expansion potential, while Dividend Yield reflects shareholder income.
- Debt/Equity ratio assesses leverage and associated risks.

4. Clustering Analysis

Methodology

1. Standardized the selected features to ensure equal influence during clustering.
2. Applied k-means clustering to group stocks into clusters based on their financial characteristics.
3. Determined the optimal number of clusters using the elbow method, which identified six clusters as optimal.

4. Also used silhouette score to confirm the optimal number of clusters and chosen the method with the highest no. of clusters, for diverse view.

Cluster Characteristics & Interpretation:

Cluster 0 (26 stocks, dominated by Bandhan Bank)

- High ROE (23.6%) and ROCE (21.9%) indicate efficient capital use.
- Moderate OPM (28.8%) suggests decent profitability but not as high as some other clusters.
- Exceptionally high Free Cash Flow (₹23,646 Cr) signals strong cash generation.
- Healthy Piotroski Score (6.57) indicates financial strength.
- EPS growth (22.4%) is solid, with moderate dividend yield (2.4%).
- Debt-to-equity ratio (2.05) is relatively high.

This cluster likely represents a mix of high-growth and high-cash-generating firms, possibly a mix of financial and non-financial companies. The dominance of Bandhan Bank suggests that within this cluster Bandhan Bank might have the best balance of risk and return.

Cluster 1 (141 stocks, no single stock dominates completely)

- Moderate ROE (21.6%) and ROCE (20.75%) suggest good profitability.
- OPM (25.3%) is slightly lower than Cluster 0.
- Free Cash Flow is positive but significantly lower (₹1,961 Cr).
- Piotroski Score (6.40) is strong, indicating good financial health.
- EPS growth (22.68%) is high, but Dividend Yield (0.97%) is lower.
- Low Debt-to-Equity ratio (1.20) indicates lower financial leverage.

This is likely a broadly diversified cluster of stable, well-performing stocks with moderate growth potential and strong financial health. Since it has 141 stocks, it may contain a mix of sectors, including both growth and value stocks.

Cluster 2 (4 stocks, dominated by Bajaj Finance)

- Moderate ROE (20.2%) and very low ROCE (10.88%) indicate that capital efficiency is low relative to other clusters.
- Exceptionally high OPM (76.02%) suggests the presence of firms with strong pricing power.
- Very large negative Free Cash Flow (₹-74,911 Cr) suggests aggressive expansion.
- Moderate Piotroski Score (5.25).

- EPS Growth (26.7%) is one of the highest, but Dividend Yield (1.98%) is moderate.
- High Debt-to-Equity ratio (5.81) suggests heavy leverage.

Likely a cluster of high-growth, highly leveraged companies, possibly in financial services or capital-intensive sectors. The dominance of Bajaj Finance clearly indicates it might have the best balance of risk and return.

Cluster 3 (1 stock, ICICI Bank)

- ROE (18.8%) and ROCE (7.6%) are relatively lower than other clusters.
- Moderate OPM (30.3%).
- Free Cash Flow (₹1,53,675 Cr) is extremely high.
- Piotroski Score (6.0) is healthy.
- EPS Growth (57.02%) is the highest among clusters.
- Very high Dividend Yield (8.0%).
- Debt-to-Equity ratio (6.45) is the highest, indicating a heavily leveraged business.

ICICI Bank alone forms this cluster, meaning its financial characteristics drive the cluster. The high leverage, strong cash flow, and high dividend yield suggest a financial institution with strong profitability and risk-adjusted returns.

Cluster 4 (7 stocks, dominated by M&M Finance)

- ROE (11.45%) and ROCE (11.46%) are the lowest among clusters, suggesting weaker capital efficiency.
- OPM (40.72%) is decent.
- Large negative Free Cash Flow (₹-26,847 Cr) suggests capital-intensive businesses.
- Piotroski Score (3.85) is the lowest, indicating weaker financial strength.
- EPS Growth (7.18%) is the lowest, meaning slower growth.
- Dividend Yield (0.97%) is moderate.
- Debt-to-Equity ratio (3.49) is high.

This cluster seems to consist of highly leveraged, slow-growing companies that might be in capital-intensive industries such as infrastructure, utilities, or finance. The dominance of M&M Finance indicates it might have the best balance of risk and return.

Cluster 5 (19 stocks, no single dominant stock)

- ROE (13.67%) and ROCE (8.86%) are on the lower side.
- OPM (52.27%) is relatively high.

- Negative Free Cash Flow (₹-7,954 Cr) suggests capital-intensive operations.
- Piotroski Score (5.10) is moderate.
- EPS Growth (19.94%) is decent, and Dividend Yield (1.11%) is moderate.
- Debt-to-Equity ratio (4.58) is high.

This cluster represents moderate-growth, capital-intensive firms that have decent profitability but require significant capital investments. The presence of multiple stocks suggests diversification across different sectors.

Final Insights:

- Clusters 0 and 1 contain a mix of high-quality growth and value stocks, with strong financials.
- Cluster 2 contains highly leveraged high-growth stocks, such as Bajaj Finance.
- Cluster 3 is purely ICICI Bank, emphasizing strong cash flow and high dividends.
- Clusters 4 and 5 contain more capital-intensive, lower-growth firms, possibly infrastructure, NBFCs, or utilities.

5. Performance Metrics

Methodology

Markowitz portfolio optimization was applied separately to each cluster to derive optimal weight allocations maximizing the Sharpe ratio.

Insights (according to Reward to Risk,Sharpe,Treynor Ratios and Beta**):**

1. **Best Risk-Adjusted Performer – Cluster 1**

- Sharpe Ratio (0.73) & Reward-to-Risk Ratio (1.16) are the highest.
- Low Beta (0.71) indicates lower volatility, making it an efficient risk-adjusted performer.
- Interpretation: A well-balanced, low-risk portfolio with strong return per unit of risk.

2. **Strong Performer – Cluster 3 (ICICI Bank Dominates)**

- Sharpe Ratio (0.67) and Reward-to-Risk Ratio (0.95) are high.
- Moderate Beta (1.10) means it carries somewhat higher risk but good return efficiency.
- Interpretation: A high-growth, high-profitability stock with good risk-adjusted returns, though exposed to stock-specific risks.

3. **Moderate Performer – Cluster 4 (M&M Finance Dominates)**

- Sharpe Ratio (0.39) and Reward-to-Risk Ratio (0.57) are moderate.
- High Beta (1.49) suggests higher market volatility.
- Interpretation: While moderately efficient, this cluster is highly sensitive to market movements, making it riskier than the benchmark.

4. **Weak Performer – Cluster 2 (Bajaj Finance Dominates)**

- Low Sharpe Ratio (0.26) & Reward-to-Risk Ratio (0.48) indicate inefficient risk-adjusted returns.
- High Beta (1.34) means high volatility but without strong compensation in returns.
- Interpretation: The cluster exhibits high risk with low return efficiency, making it a weak performer in risk-adjusted terms.

5. **Poor Performer – Cluster 5 (19 Stocks, High Debt/Equity)**

- Very Low Sharpe Ratio (0.02) and Reward-to-Risk Ratio (0.33).
- Moderate Beta (1.05) suggests exposure to market fluctuations.
- Interpretation: Over-leveraged and inefficient in generating returns for the risk taken, leading to poor overall performance.

6. **Worst Performer – Cluster 0 (Bandhan Bank Dominates)**

- Negative Sharpe Ratio (-0.57) & Reward-to-Risk Ratio (-0.38) indicate severe underperformance.
- High Beta (1.22) suggests it took excessive risk but failed to generate positive returns.
- Interpretation: Worst in risk-adjusted terms, indicating excessive volatility without reward.

6. VaR Computation

Methodology

Value at Risk (VaR) was calculated at a confidence level of 99% using historical simulation methods

Cluster-wise Risk Analysis

Cluster	VaR (99% Confidence Level)	Risk Level	Risk Interpretation
Cluster 1	**-0.0257**	**Low Risk**	Best-performing cluster with diversified exposure and low downside risk.
Cluster 5	**-0.0349**	**Moderate-Low Risk**	19-stock portfolio with manageable leverage, leading to lower downside risk.
Cluster 3	**-0.0408**	**Moderate Risk**	Single-stock cluster (ICICI Bank) with stable fundamentals but some downside risk.
Cluster 4	**-0.0584**	**Moderate-High Risk**	7-stock cluster (M&M Finance-dominated) with exposure to financial stocks, making it relatively risky.
Cluster 2	**-0.0555**	**Moderate-High Risk**	4-stock cluster (Bajaj Finance dominates), high operating margins but negative free cash flow increased risk.
Cluster 0	**-0.0726**	**Highest Risk**	Bandhan Bank dominates; weak fundamentals and high volatility lead to maximum downside risk.

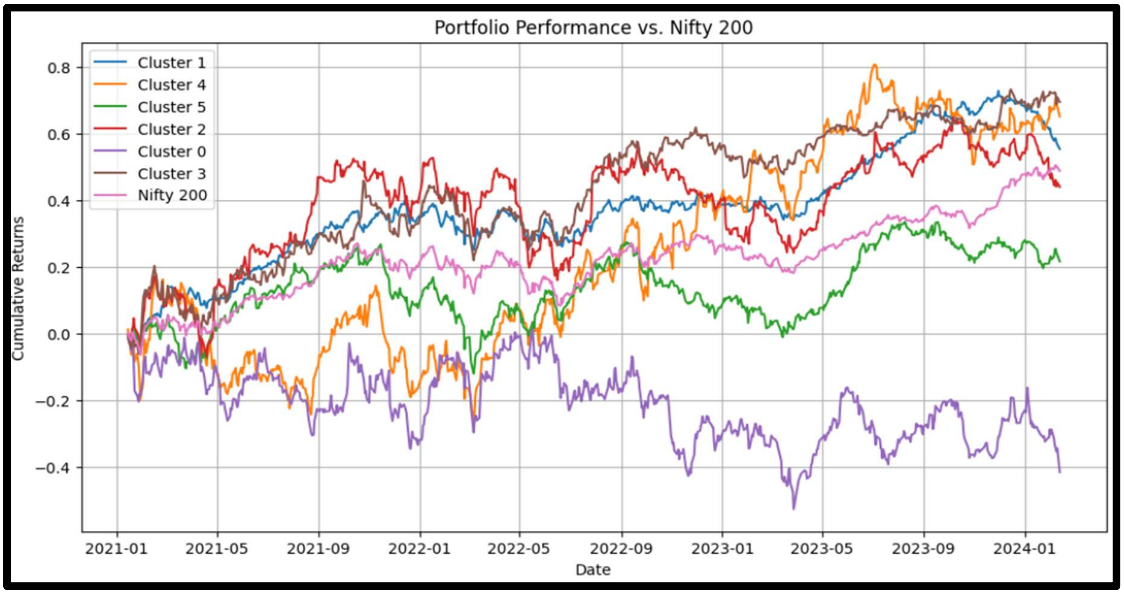
Risk Implications for Portfolio Allocation:

- Investors seeking low-risk, stable returns should favour Cluster 1.
- Moderate-risk seekers can consider Cluster 3 and Cluster 5.
- High-risk, high-reward portfolios may look at Clusters 2 and 4.
- Cluster 0 should be avoided or allocated cautiously due to extreme risk exposure.

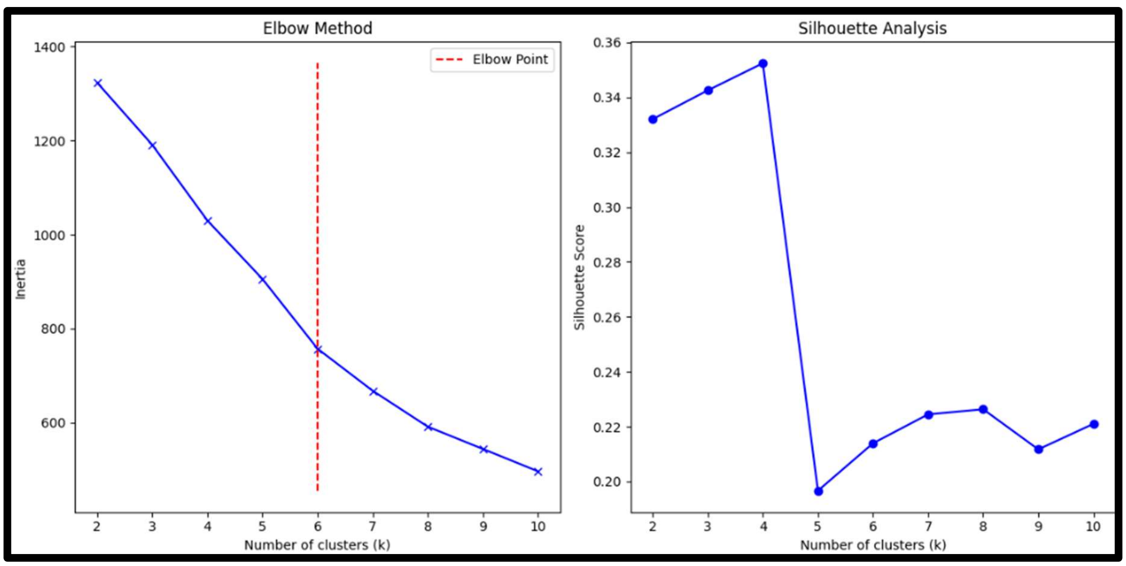
7.Cumulative Return Insights & Risk Implications

1. **Best Cumulative Performers**
   - Cluster 3 (69.46%) and Cluster 4 (65.12%) outperformed the NIFTY 200 index (48.75%).
   - Cluster 3's single-stock exposure (ICICI Bank) makes it riskier but highly profitable.
   - Cluster 4's portfolio (M&M Finance and similar stocks) has high volatility but strong returns.
2. **Moderate to Strong Performers**
   - Cluster 1 (55.36%) surpassed NIFTY 200, indicating well-balanced portfolio performance.
   - Cluster 2 (43.96%) is slightly below NIFTY 200 but shows resilience despite weak risk-adjusted returns.
3. **Underperformers**
   - Cluster 5 (21.61%) struggled with poor financial characteristics, despite not losing money.
   - Cluster 0 (-41.67%) suffered major losses, showing extreme risk exposure and inefficiency.

APPENDIX



Portfolio Performance vs. Nifty 200
Cluster 1
Cluster 4
Cluster 5
Cluster 2
Cluster 0
Cluster 3
Nifty 200
Cumulative Returns
Date



Elbow Method
Elbow Point
Inertia
Number of clusters (k)
Silhouette Analysis
Silhouette Score
Number of clusters (k)

1. Financial Indicators (Cluster-wise Averages)

Cluster	ROE %	ROCE %	OPM %	Free Cash Flow (Rs. Cr.)	Piotroski Score	EPS Growth	Dividend Yield %	Debt/Equity
0	23.62	21.94	28.82	23,646.37	6.58	22.41	2.41	2.05
1	21.64	20.75	25.33	1,961.99	6.40	22.68	0.97	1.20
2	20.22	10.89	76.03	-74,911.23	5.25	26.71	1.98	5.81
3	18.80	7.60	30.37	153,675.82	6.00	57.02	0.80	6.45
4	11.46	11.46	40.72	-26,847.66	3.86	7.19	0.98	3.49
5	13.67	8.87	52.27	-7,954.92	5.11	19.94	1.11	4.58

2. Performance Metrics

Cluster	Reward to Risk Ratio	Sharpe Ratio	Treynor Ratio	Beta
0	-0.3822	-0.5697	-0.1685	1.2208
1	1.1626	0.7332	0.1617	0.7148
2	0.4812	0.2574	0.0582	1.3386
3	0.9534	0.6728	0.1474	1.1010
4	0.5659	0.3882	0.0995	1.4871
5	0.3286	0.0176	0.0037	1.0476

3. Value at Risk (VaR) – 99% Confidence Level

Cluster	VaR (99% Confidence Level)
0	-0.0726
1	-0.0257
2	-0.0555
3	-0.0408
4	-0.0584
5	-0.0349

4.Cumulative returns of optimised Cluster Portfolios

Cluster	Final Cumulative Return (%)
0	-41.67%
1	55.36%
2	43.96%
3	69.46%
4	65.12%
5	21.61%
Nifty 200	48.75%