Trade&Ahead – Stock Grouping

Unsupervised Learning

16 December 2023

Contents / Agenda

- Executive Summary
- Business Problem Overview and Solution Approach
- EDA Results
- Data Preprocessing
- K-Means Clustering
- Hierarchical Clustering
- Appendix

Executive Summary

- Optimizing Diversification: Securities categorized into four clusters, ranging from very aggressive to mildly aggressive, facilitate diversified investments to maximize earnings in varying market conditions.

- Dynamic Analysis for Market Volatility: Acknowledging stock market volatility, dynamic clustering and continuous analysis of stock movement across clusters are essential for making accurate predictions as more data is added daily.

- Strategic Client Recommendations: Tailoring investment recommendations based on client financial goals, risk tolerance, and behaviors, Trade&Ahead can utilize clusters as potential portfolios or delve deeper into financial statement analysis to identify stocks that deviate from cluster profiles for more strategic decision-making.

Business Problem Overview and Solution Approach

Context: Trade&Ahead is looking to increase their risk management when investing in stocks. A diversified approach not only enhances returns but also mitigates risks by minimizing potential losses during market downturns. Utilizing cluster analysis aids in identifying stocks with similar characteristics and minimal correlation, streamlining the stock selection process and fortifying the portfolio against vulnerability to losses.

Objective: Analyze stock price and financial indicator data for select companies on the New York Stock Exchange, employing clustering techniques to group stocks based on provided attributes and deliver insightful findings on the distinctive characteristics of each group.

Solution Approach: Compare results from K-means Clustering and Hierachical Cluster to identify similarity in clusters, then determine key distinctive characteristics.

EDA Results

- *Exploratory Data Analysis reveals that the distribution of Current_Price and Estimated_Shares_Outstanding across securities in all sectors is right-skewed, featuring several positive outliers.*
- *The Health Care and Financial sectors stand out with notably high positive Price_Change in the last 13 weeks, presenting favorable opportunities for investors.*
- *The Information Technology and Financial sectors exhibit some of the highest Cash_Ratios, making them more favorable compared to other sectors.*
- *In contrast, the Real Estate sector showcases minimal variation in both Price_Change and Cash_Ratio across its securities, positioning it as a safer investment choice for risk-averse investors.*
- *On the other hand, the Energy sector displays significant variance in Price_Change across its securities, indicating higher volatility and risk. However, this sector features securities with elevated P/E_Ratios, suggesting investors are willing to invest more in a single share relative to its earnings, making it an intriguing but riskier investment choice.*

Link to Appendix slide on data background check

Data Preprocessing

- Duplicate value check
  - No duplicate values
- Missing value treatment
  - No missing values
- Outlier check
  - Outliers were seen as appropriate reflections of the market and volatility and were not treated. See appendix for visualization.
- Data preparation for modeling
  - Data was scaled using StandardScalar()

***Note**: You can use more than one slide if needed*

K-Means Cluster Summaries

Cluster 0 - Large Market Capitalization / Dow Jones Industrial Average:
- 11 stocks, predominantly from Financials, Health Care, IT, and Consumer Discretionary sectors.
- Characteristics include low volatility, significant cash outflows, highest net incomes, and a large number of shares outstanding.

- **Cluster 1 - "Cash is King":**
- 13 stocks, mainly from Healthcare and IT sectors.
- Exhibits moderate volatility, profitability, and notable cash ratios with cash inflows.

- **Cluster 2 - S&P 500 / Diversification:**
- 280 stocks (84% of the dataset) across all sectors.
- Features low P/E ratios and outliers on negative P/B ratios, representing a diversified portfolio.

- **Cluster 3 - "Ride the Energy Rollercoaster" portfolio / Growth mindset:**
- 29 stocks, predominantly from the Energy sector.
- Characterized by low stock prices, high ROE, high beta, volatility, and mostly negative net incomes.

- **Cluster 4 - High Earnings for a High Price:**
- 7 stocks, mostly from Health Care and Consumer Discretionary sectors.
- Notable for highest stock prices, favorable cash ratios, and the most favorable P/B ratios, with highest earnings-per-share.

Hierarchical Clustering Summary

- Optimal Number of clusters using Hierarchical Clustering
- Cluster Profiling

***Note**: You can use more than one slide if needed*

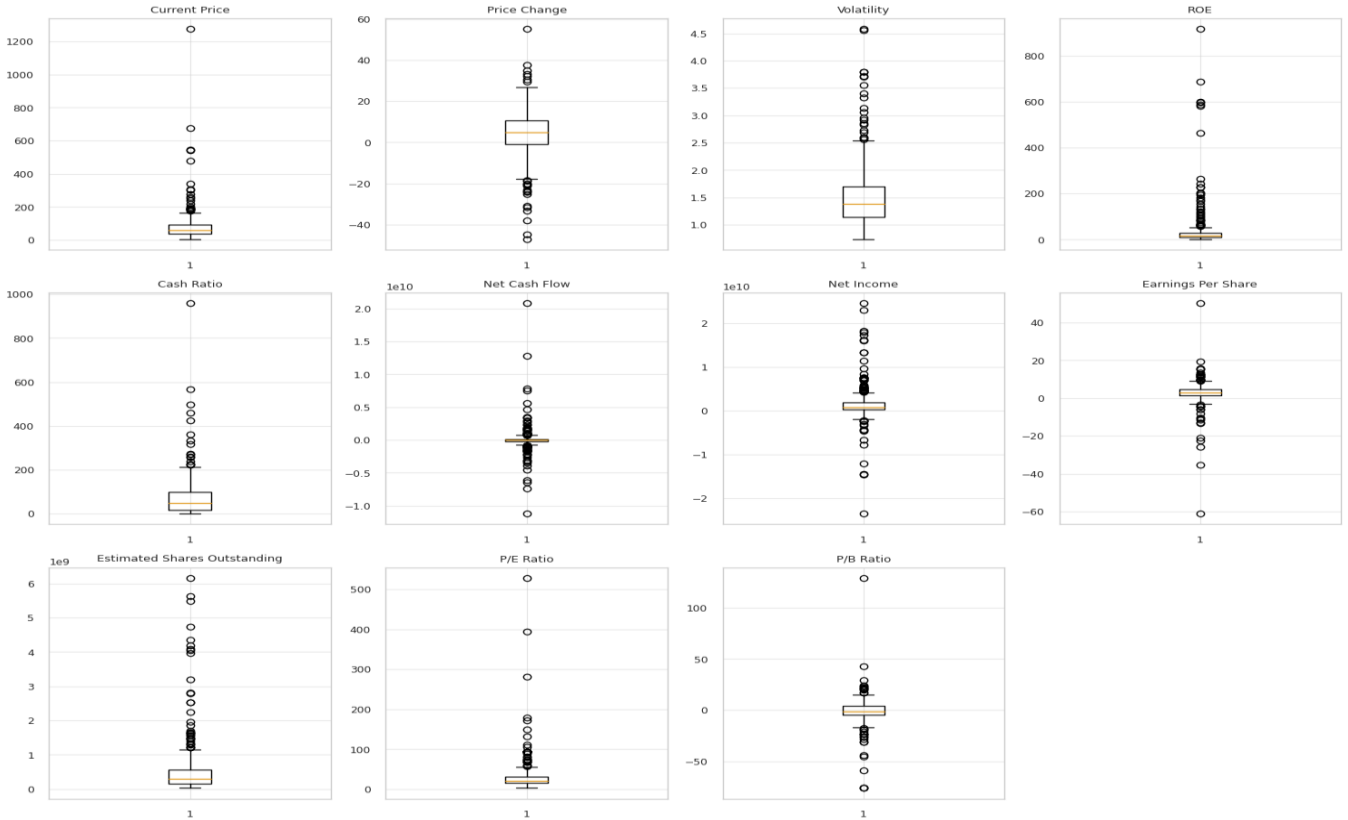
Link to Appendix slide on Hierarchical Clustering

APPENDIX

Data Background and Contents

- Ticker Symbol: An abbreviation used to uniquely identify publicly traded shares of a particular stock on Ticker Symbol: An abbreviation used to uniquely identify publicly traded shares of a particular stock on a particular stock market
- Company: Name of the company
- GICS Sector: The specific economic sector assigned to a company by the Global Industry Classification Standard (GICS) that best defines its business operations
- GICS Sub Industry: The specific sub-industry group assigned to a company by the Global Industry Classification Standard (GICS) that best defines its business operations
- Current Price: Current stock price in dollars
- Price Change: Percentage change in the stock price in 13 weeks
- Volatility: Standard deviation of the stock price over the past 13 weeks
- ROE: A measure of financial performance calculated by dividing net income by shareholders' equity (shareholders' equity is equal to a company's assets minus its debt)
- Cash Ratio: The ratio of a company's total reserves of cash and cash equivalents to its total current liabilities
- Net Cash Flow: The difference between a company's cash inflows and outflows (in dollars)
- Net Income: Revenues minus expenses, interest, and taxes (in dollars)
- Earnings Per Share: Company's net profit divided by the number of common shares it has outstanding (in dollars)
- Estimated Shares Outstanding: Company's stock currently held by all its shareholders
- P/E Ratio: Ratio of the company's current stock price to the earnings per share
- P/B Ratio: Ratio of the company's stock price per share by its book value per share (book value of a company is the net difference between that company's total assets and total liabilities)a particular stock market

Outliers Visualization



K-Means Clustering Technique

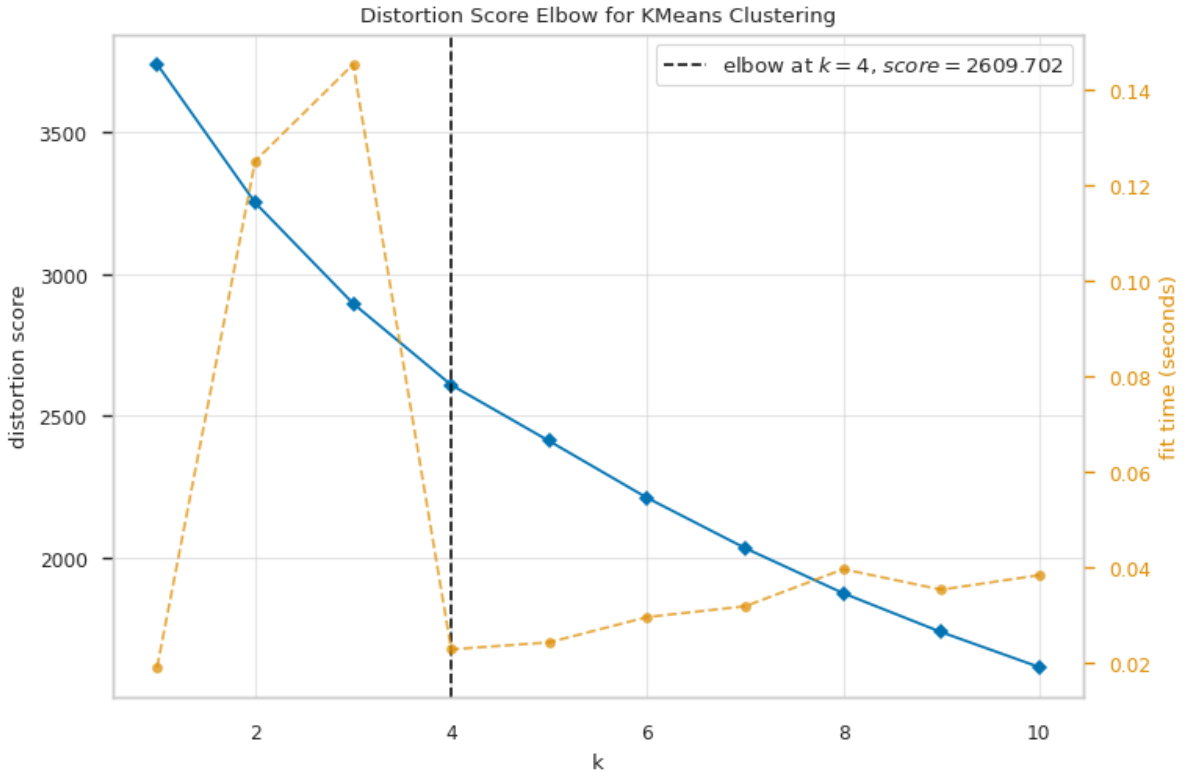
- Please update regarding application of K-Means Clustering
- Observations using Elbow Curve along with visuals
- Observations from Silhouette scores for different number of clusters

***Note**: You can use more than one slide if needed*

K-Means Clustering Summary

- Model – KMeans
- Visualizer – KElbow

- Distortion Elbow appears at k=4 with a distortion score of 2609.702

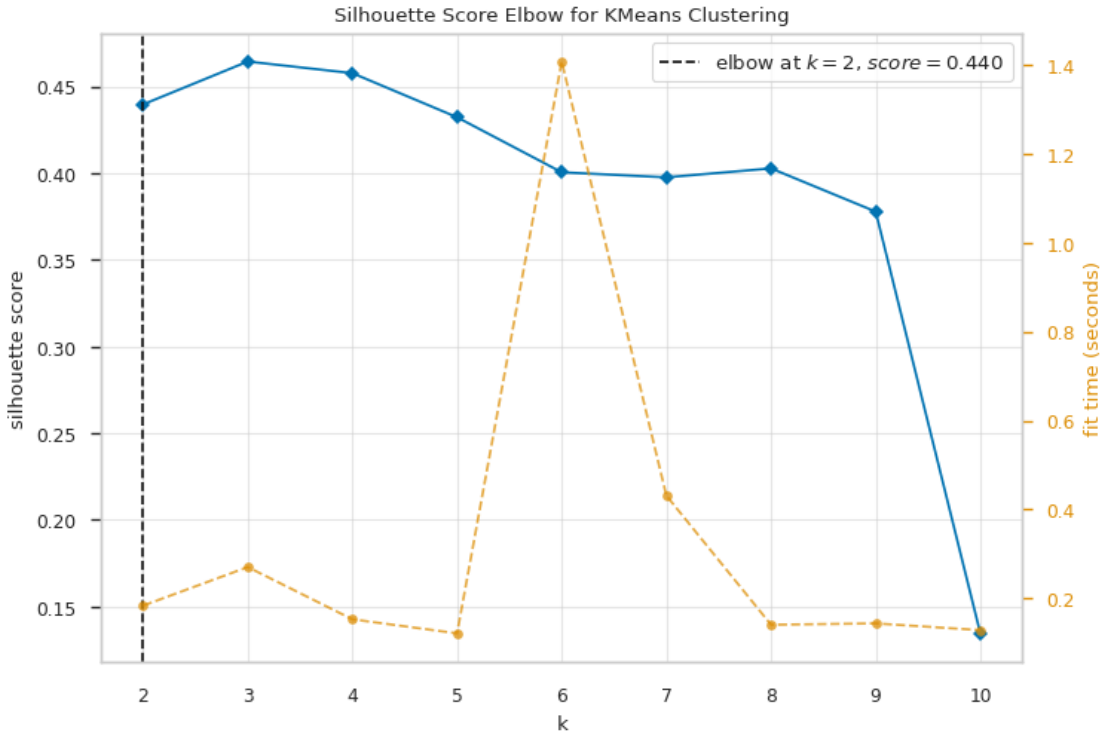


Link to Appendix slide on K-Means Clustering

K-Means Clustering Summary

- Model – KMeans
- Visualizer – KElbow
- Metric - Silhouette

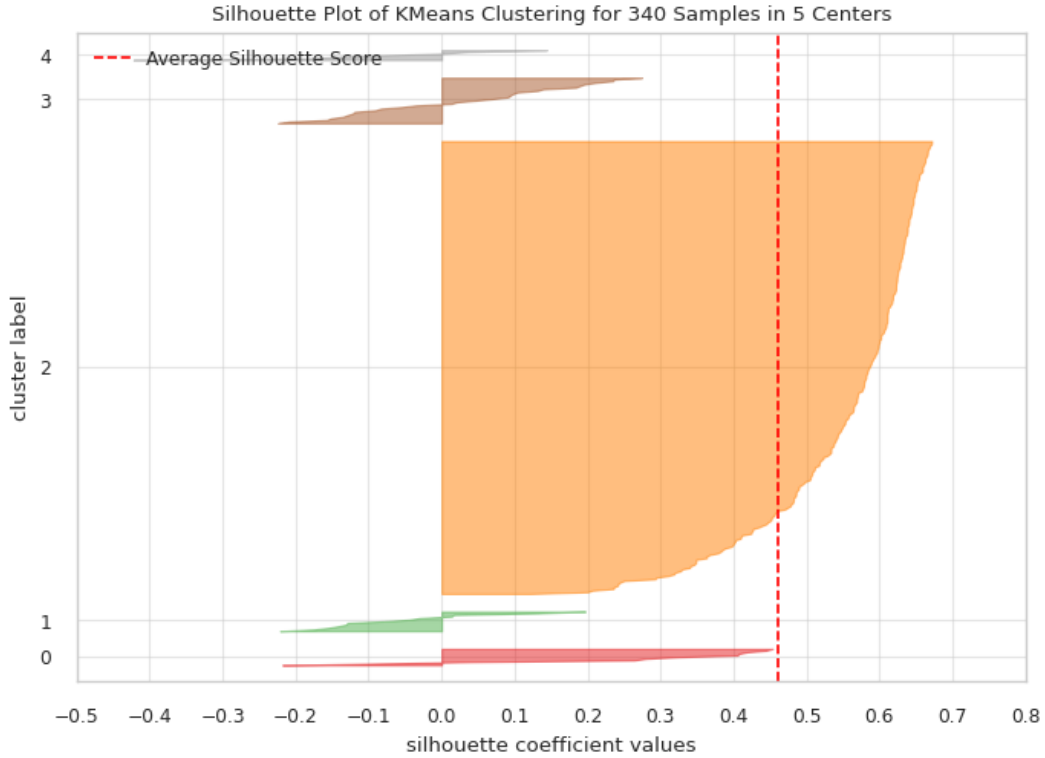
- However the Silhouette Score optimizes at a K-value of 2.



Link to Appendix slide on K-Means Clustering

K-Means Clustering Summary

- Model – KMeans
- Visualizer – Silhouette



Link to Appendix slide on K-Means Clustering

KMeans_clusters	Current Price	Price Change	Volatility	ROE	Cash Ratio	Net Cash Flow	Net Income	Earnings Per Share	Estimated Shares Outstanding	P/E Ratio	P/B Ratio	Count
0	72.399112	5.066225	1.388319	34.620939	53.000000	-14046223.826715	1482212389.891697	3.621029	438533835.667184	23.843656	-3.358948	277
1	50.517273	5.747586	1.130399	31.090909	75.909091	-1072272727.272727	14833090909.090910	4.154545	4298826628.727273	14.803577	-4.552119	11
2	38.099260	-15.370329	2.910500	107.074074	50.037037	-159428481.481481	-3887457740.740741	-9.473704	480398572.845926	90.619220	1.342067	27
3	234.170932	13.400685	1.729989	25.600000	277.640000	1554926560.000000	1572611680.000000	6.045200	578316318.948800	74.960824	14.402452	25

Hierarchical Clustering Technique

- Please update regarding application of Hierarchical Clustering
- Observations using different linkage methods
- Dendrograms for linkage methods used and their observations
- Observations from Cophenetic correlation for different combinations of distance and metrics

***Note**: You can use more than one slide if needed*

K-Means vs Hierarchical Clustering

- Comparison of clusters obtained from K-Means and Hierarchical Clustering on various parameters

***Note**: You can use more than one slide if needed*

Great Learning

Happy Learning !