Trade and Ahead Project

Unsupervised Learning Project

Contents

Business Problem Overview and Solution Approach

The Problem

- The stock market is a strong long-term investment option, offering wealth creation, inflation protection, tax benefits, and the power of compounding. Starting early can significantly grow retirement savings and help achieve financial goals.
- A diversified portfolio maximizes returns while reducing risk, cushioning losses during downturns. However, analyzing multiple stocks across numerous financial metrics to find the right picks can be tedious.
- Cluster analysis helps identify stocks with similar traits and low correlation, enabling investors to diversify across segments and reduce portfolio risk.
- Trade&Ahead is a financial consultancy firm who provide their customers with personalized investment strategies.
- Trade&Ahead has provided data comprising stock price and some financial indicators for a few companies listed under the New York Stock Exchange.
- Trade&Ahead seeks data analysis to group stocks by attributes and provide insights into each cluster's characteristics.

Solution Approach

- The approach is to draw insight from the data provided, perform exploratory data analysis, make important observations, and perform cluster analysis that can help group stocks by attributes. This will enable investors to analyze stocks across segments and reduce portfolio vulnerability to risks.

Data Background and Content

The data contains the different attributes of stocks. The data has 340 rows and 15 columns. The columns types are integer, object and float (4 int, 4 object and 7 float). The data has the following columns:

- 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)

There are no duplicates in the data.

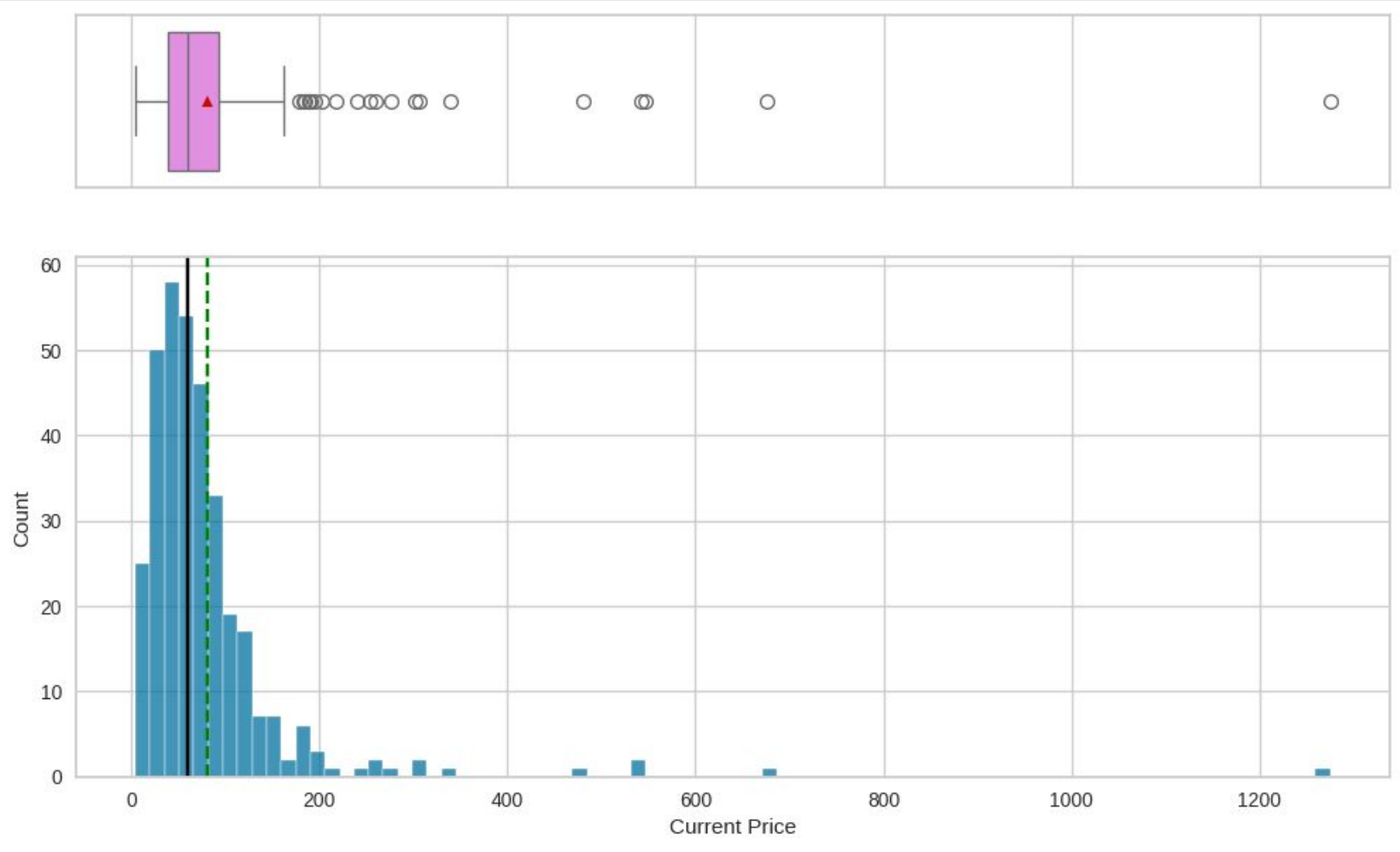
There are no missing values in the dataset

Statistical Summary of Dataset

- There are 11 GICS sectors and 104 sub-industries, these reflect broad market representation.
- The minimum current stock price is 4.5 dollars while the maximum is approx 1275 dollars. This is a significantly wide range.
- The average percentage change in the stock price in 13 weeks stands at 4.078194. Price change varies widely, ranging from -47.13 to 55.05.
- The minimum volatility is 0.733163 while the maximum is 4.580042. This is a significantly wide range.
- ROE and Cash Ratio have a wide range of values. ROE varies from 1.0 to 917 while Cash Ratio varies from 0 to 958.
- The standard deviation values for Net Cash Flow, Net Income, P/E Ratio, and P/B Ratio are substantially higher than the third quartile (75%), indicating the presence of potential outliers.
-

EDA Results

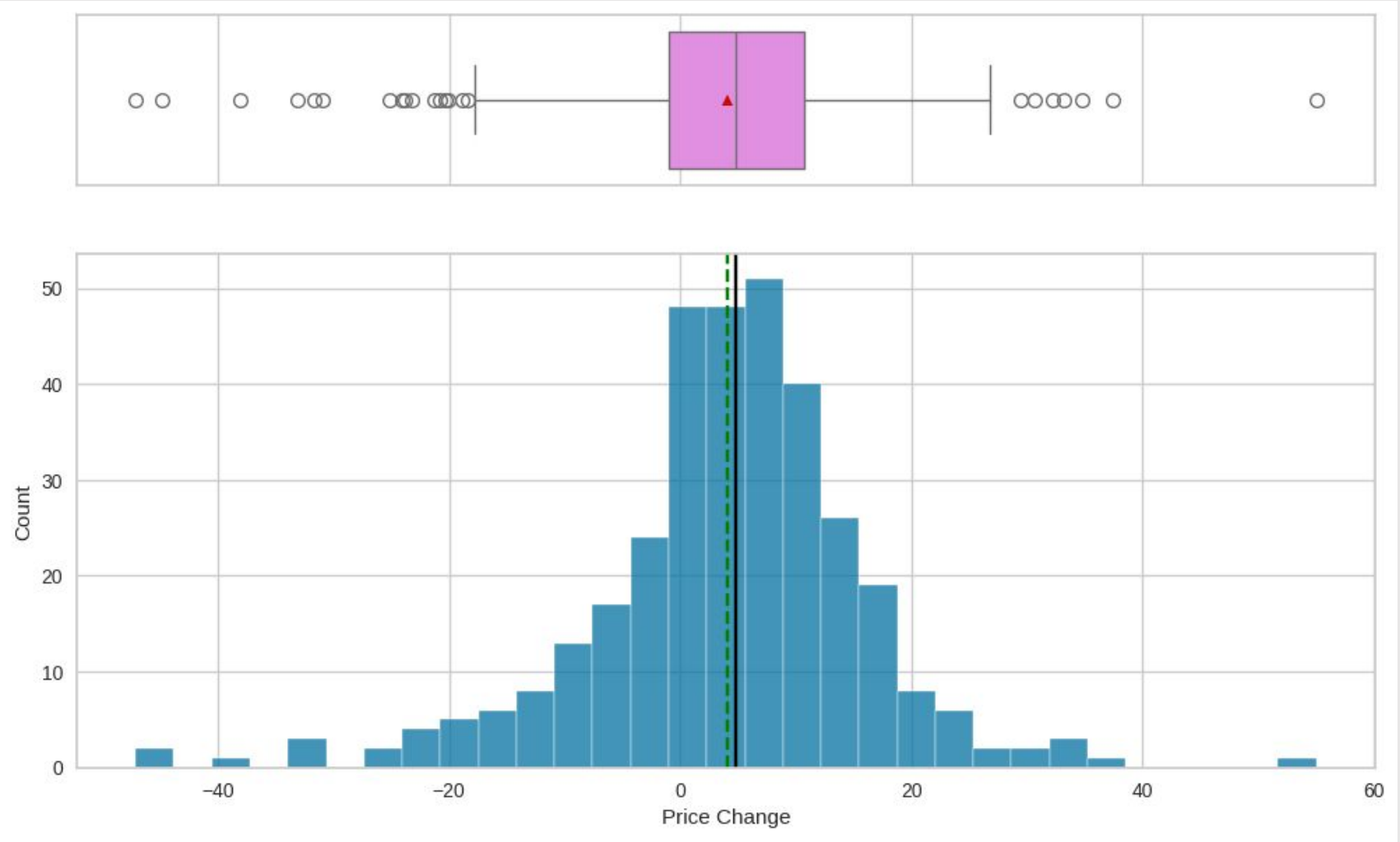
Univariate analysis: Observation on Current Price



- Current Stock price exhibits a right-skewed distribution, with some prices found at the lower end and a long tail corresponding to higher values.
- This shows the presence of a few stocks with exceptionally high prices relative to the majority.
- Unsurprisingly, there is no stock that is less than 0 dollars.

EDA Results

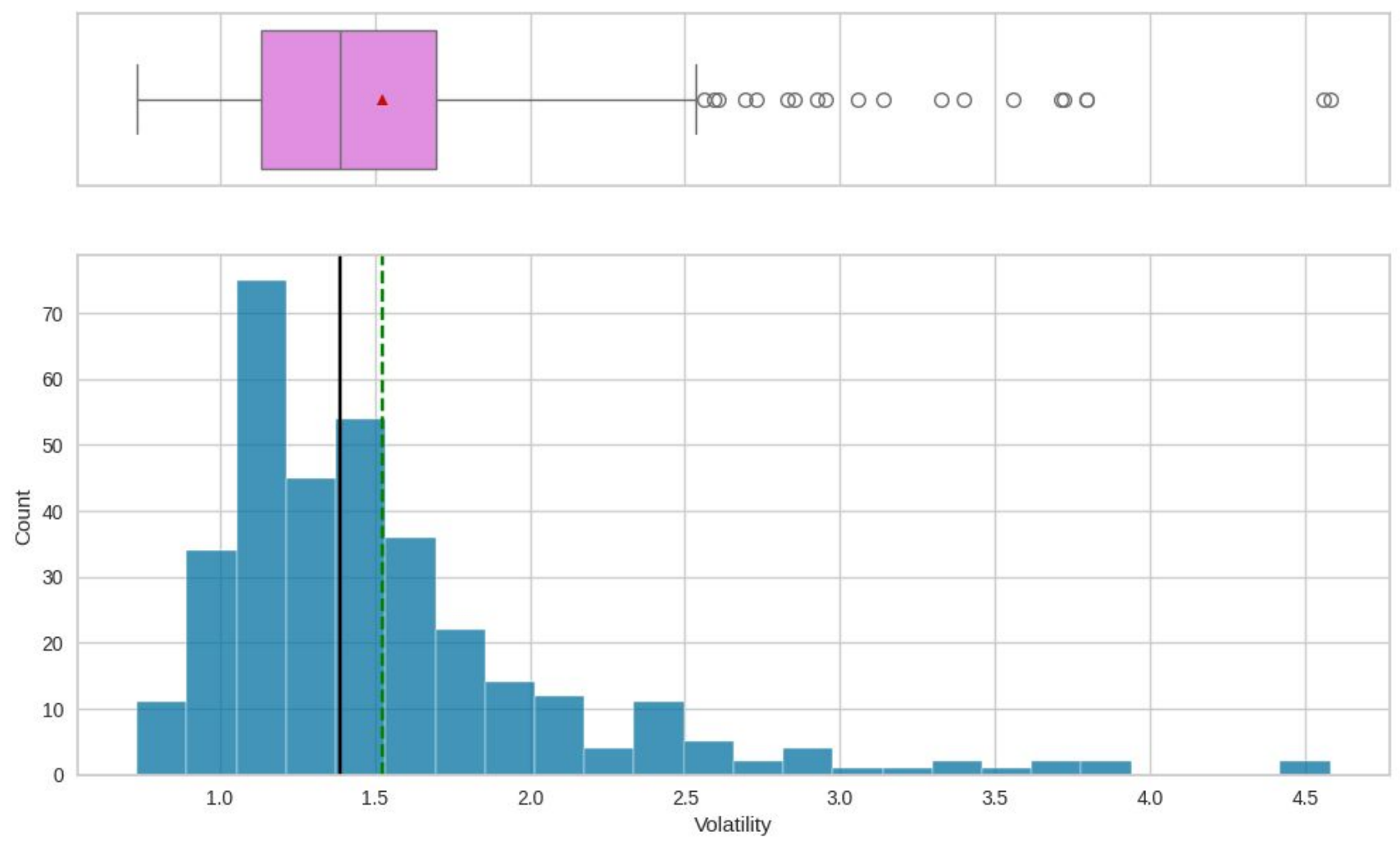
Univariate analysis: Observation on Price Change



- The distribution of Price Change is almost centered around zero, with long tails extending in both the positive and negative directions.
- The distribution is almost uniform

EDA Results

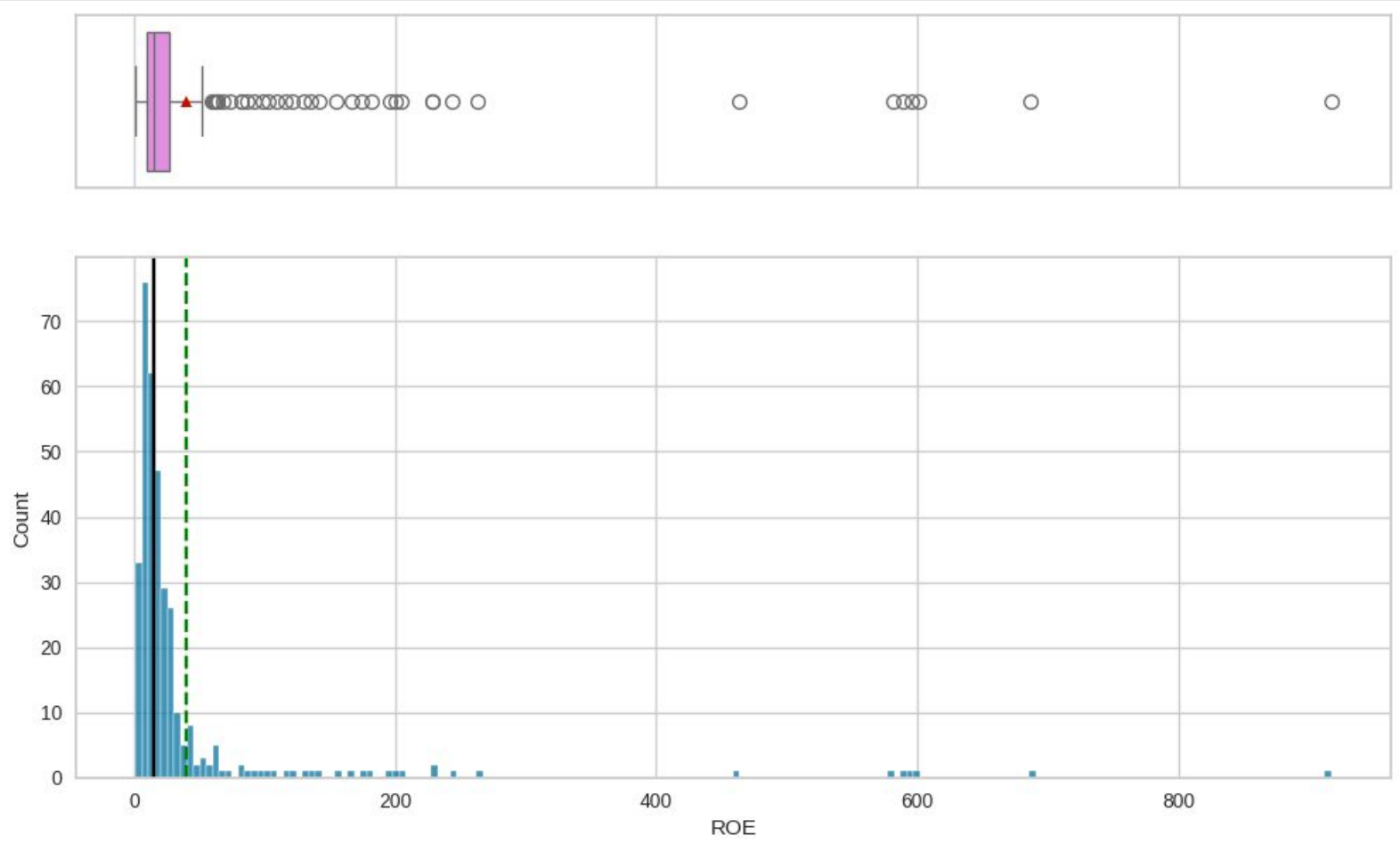
Univariate analysis: Observation on Volatility



- Volatility exhibits a rightly skewed distribution.
- This is not a Gaussian distribution.
- The boxplots shows the presence of outliers.
- The highest standard deviation of stock price over a period of 13 weeks is 45.

EDA Results

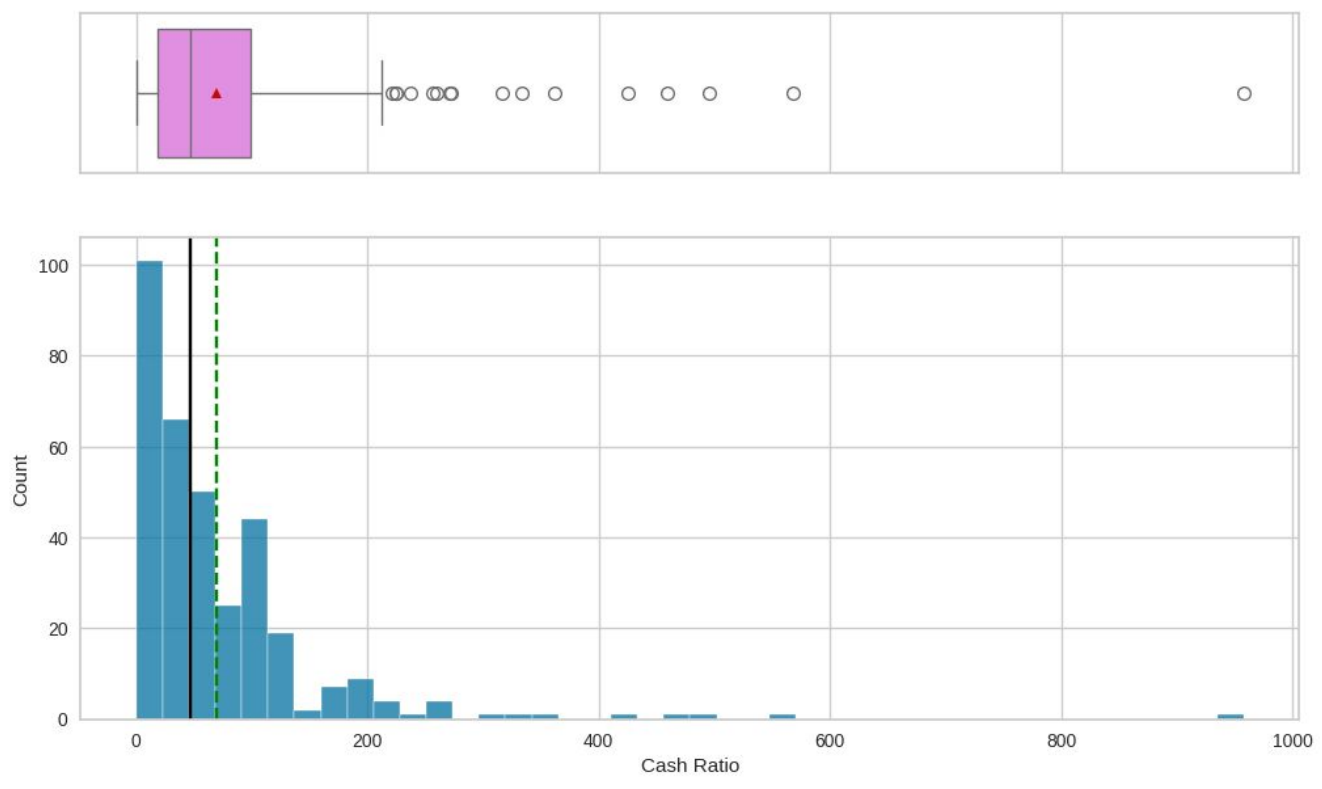
Univariate analysis: Observation on ROE



- ROE is a measure of financial performance.
- ROE exhibits a right-skewed distribution, with some ROEs found at the lower end and a long tail corresponding to higher ROE values.
- This shows that a few stocks have exceptionally good financial performance relative to the majority.
- Unsurprisingly, ROE values are all positive (greater than zero).

EDA Results

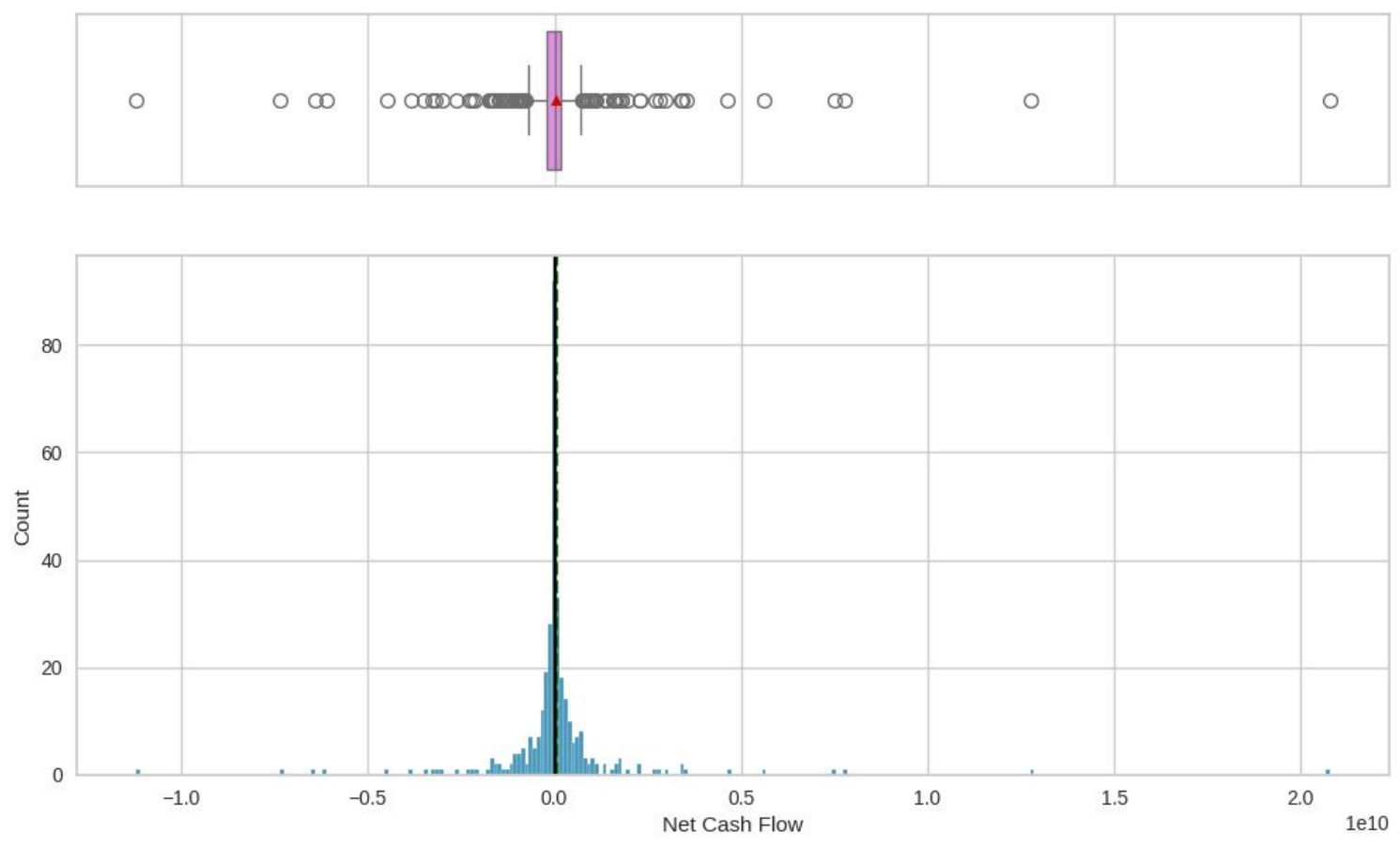
Univariate analysis: Observation on Cash Ratio



- The Median is less than the Mean of the distribution of Cash Ratio.
- There are outliers as shown by the Boxplot.
- The distribution of Cash Ratio is skewed to the right.
- There are no negative values for Cash ratio

EDA Results

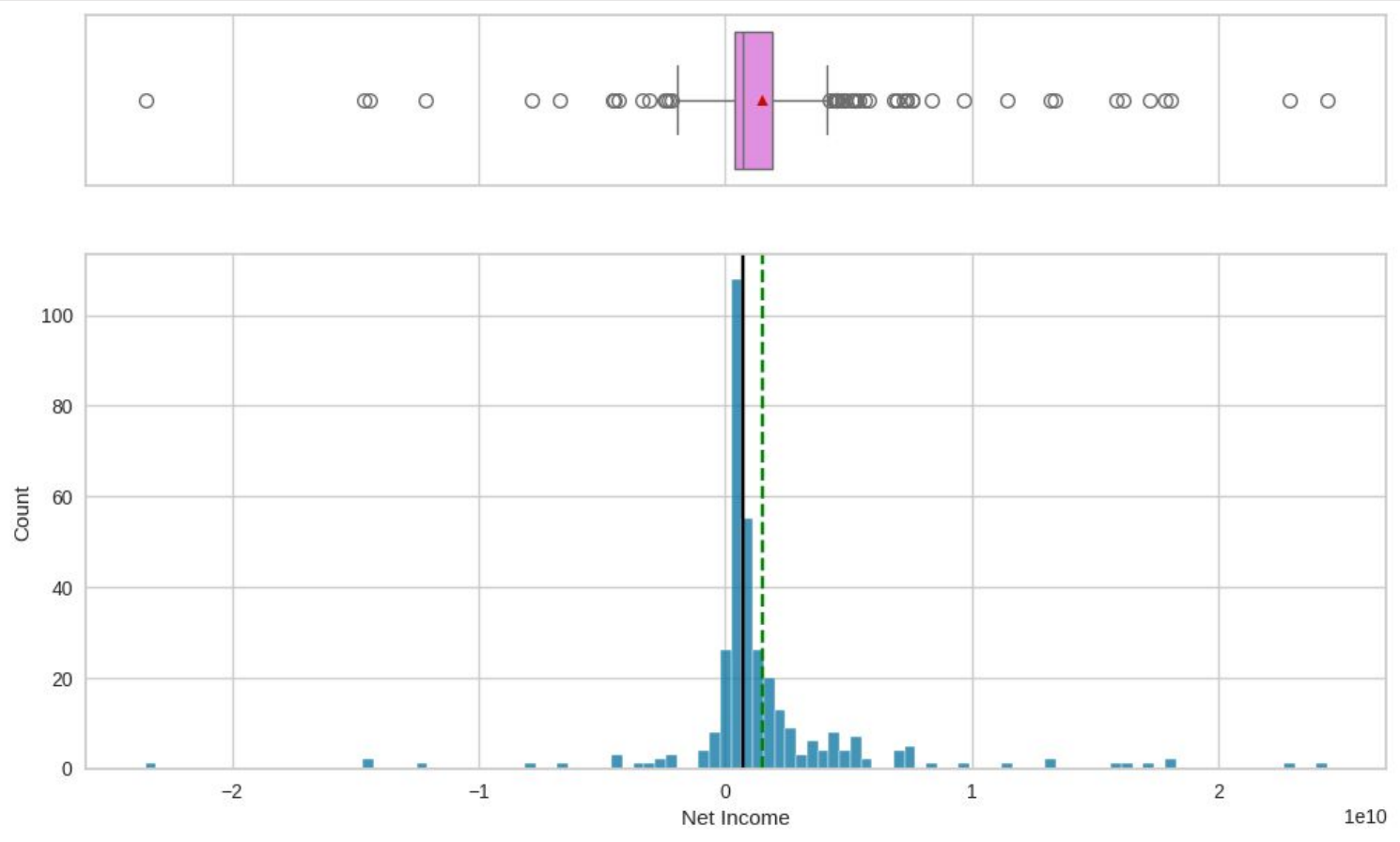
Univariate analysis: Observation on Net Cash Flow



- Here is a barplot for the Net Cash Flow (difference between cash inflow and outflow).
- There are outliers in both positive and negative directions.
- The distribution seems to have a longer tail in the positive direction.

EDA Results

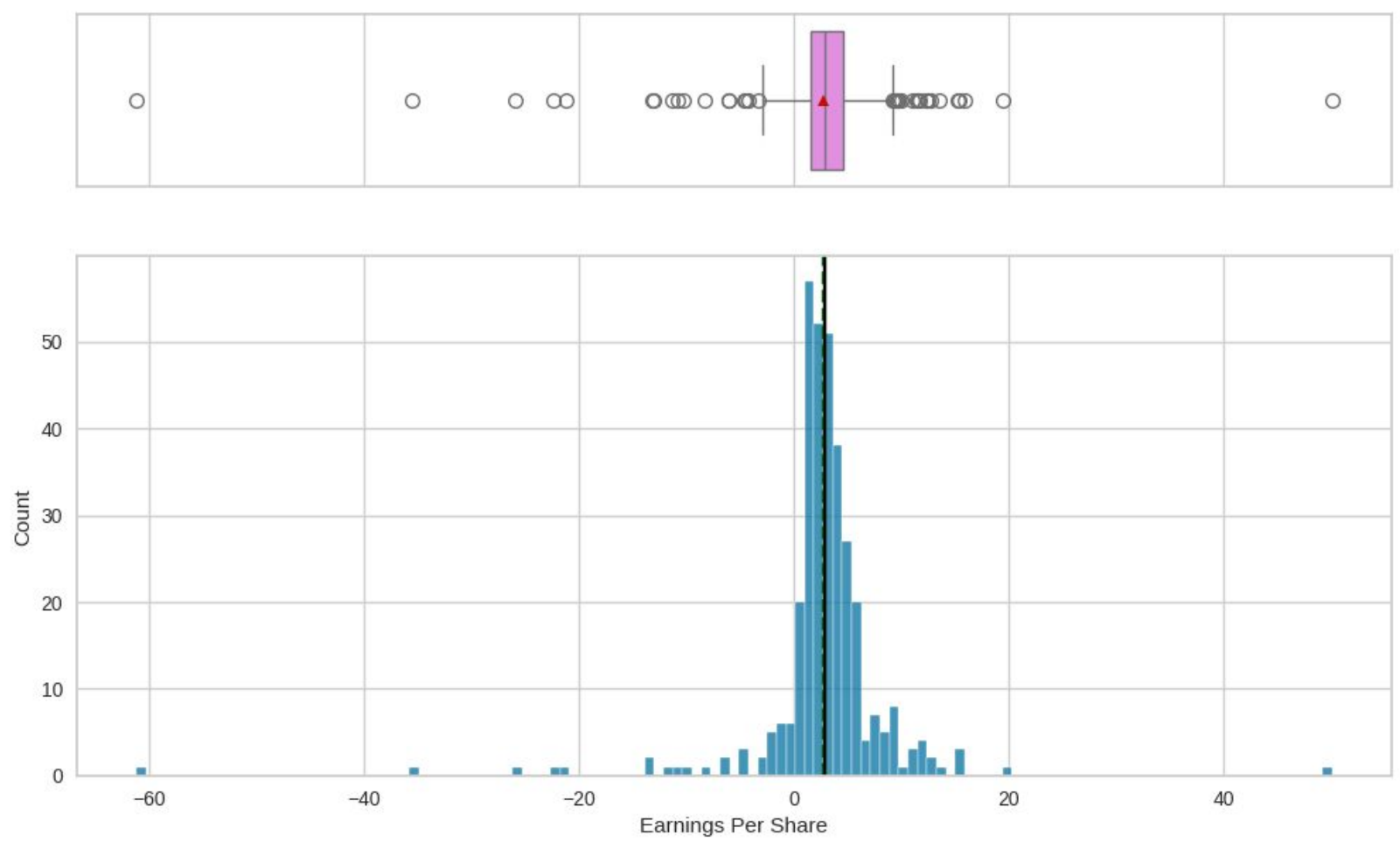
Univariate analysis: Observation on Net Income



- The distribution extends in both positive and negative directions.
- This shows that some companies have negative net income and are failing while some have positive net income and are doing well.
- The median is less than the mean, the distribution seems to be rightly skewed.

EDA Results

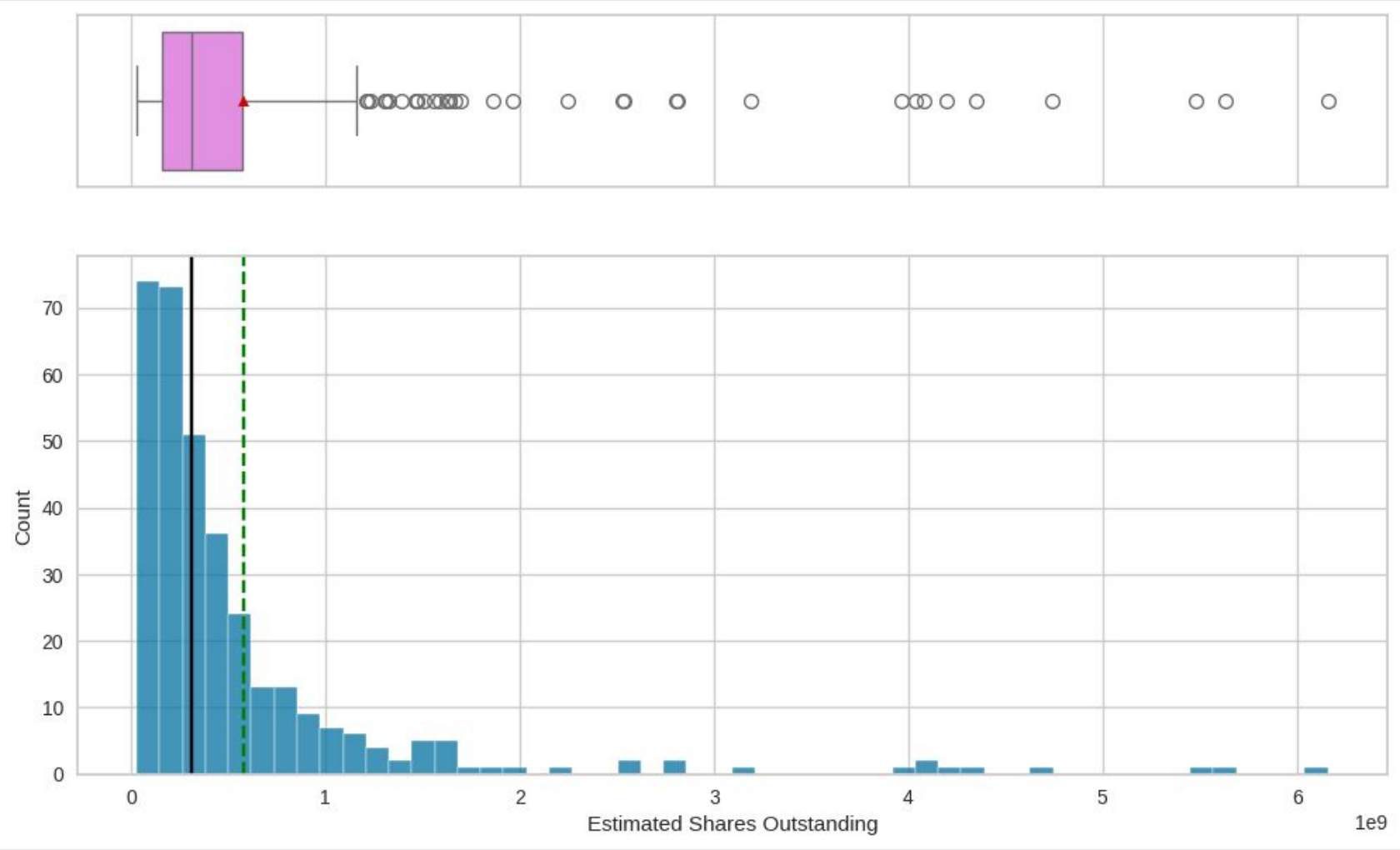
Univariate analysis: Observation on Earnings Per Share



- The distribution extends in both positive and negative directions.
- This shows that some companies have negative Earnings Per Share while some are positive.
- The median is greater than the mean, the distribution appears to be skewed to the left.

EDA Results

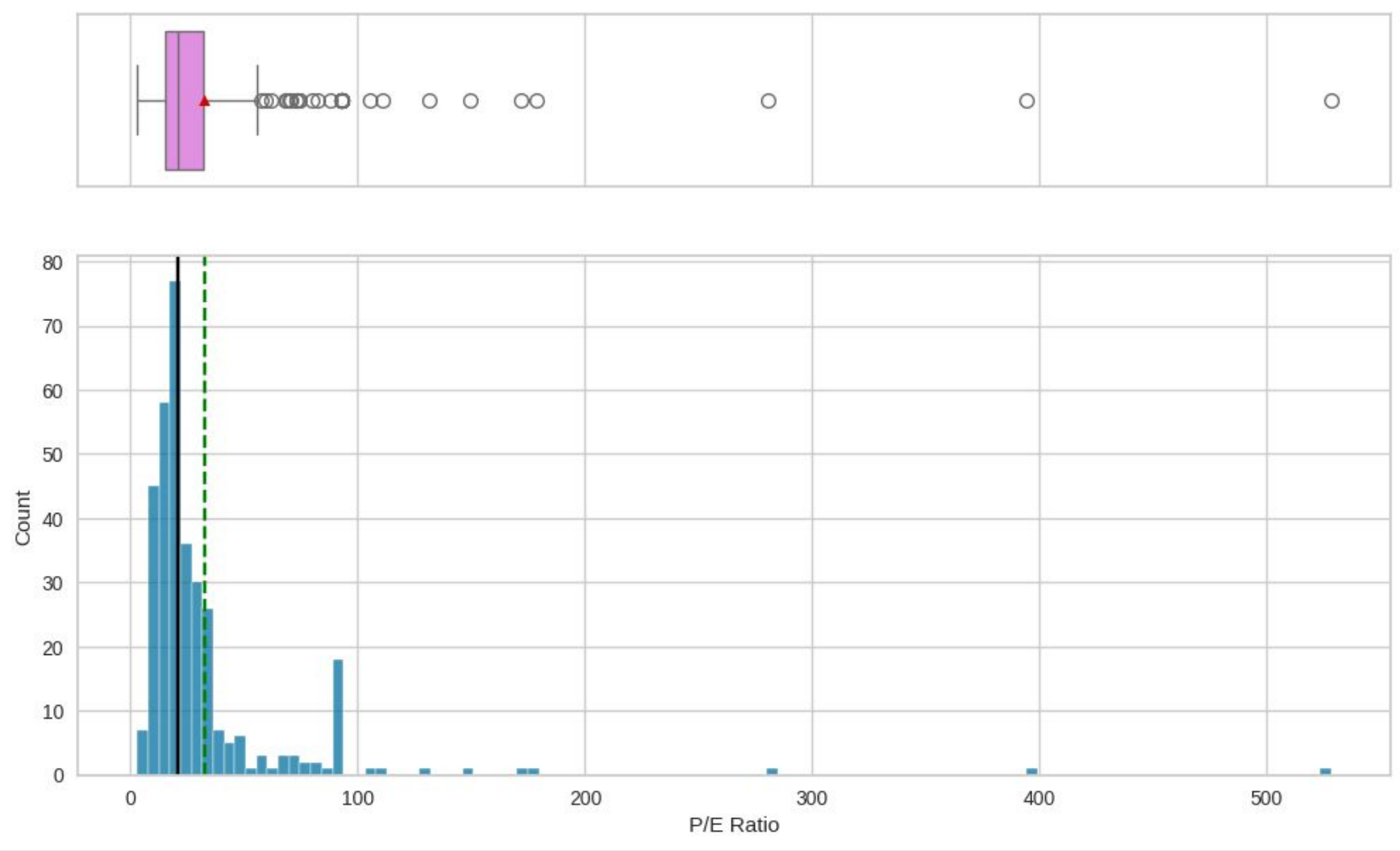
Univariate analysis: Observation on Estimated Shares Outstanding



- The distribution extends only in the positive direction.
- Estimated Shares Outstanding has values strictly greater than zero.
- The median is less than the mean, the distribution is obviously rightly skewed.
- There are outliers.
- Some companies have significantly higher Estimated Shares Outstanding.

EDA Results

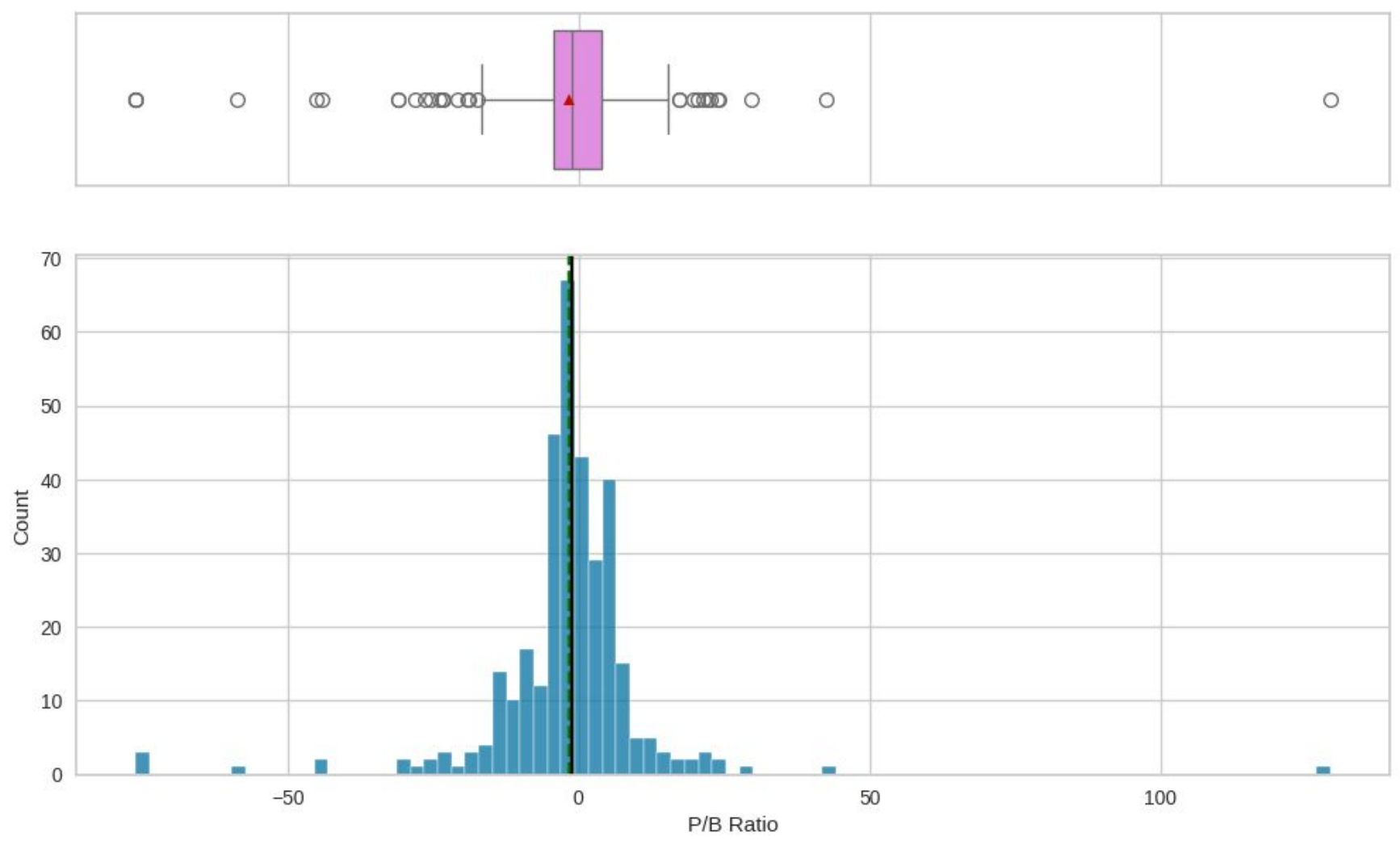
Univariate analysis: Observation on P/E Ratio



- The distribution extends only in the positive direction.
- P/E Ratio has values strictly greater than zero.
- The median is less than the mean, the distribution is obviously rightly skewed.
- There are outliers.
- Some companies have significantly higher P/E Ratio.

EDA Results

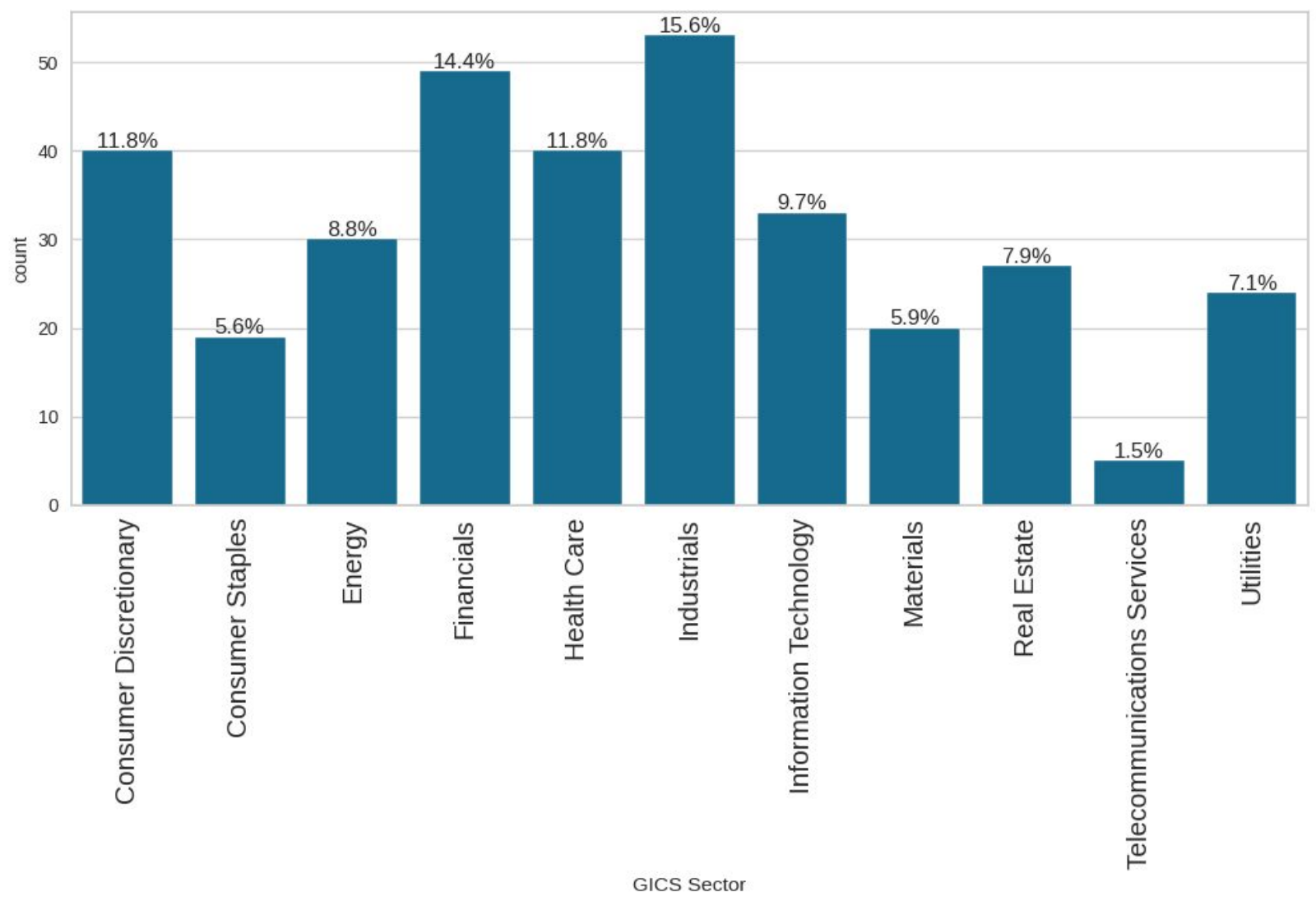
Univariate analysis: Observation on P/B Ratio



- The distribution extends in both positive and negative directions.
- There are outliers.
- Certain companies exhibit considerably higher P/B ratios, while others show substantially lower values.

EDA Results

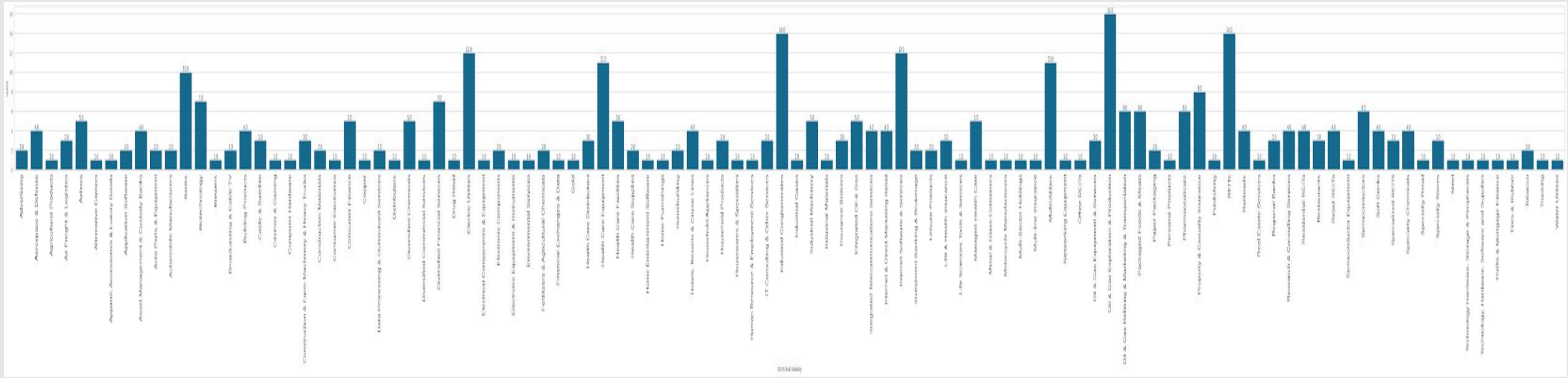
Univariate analysis: Observation on GICS Sector



- The GICS Sector denotes a company's assigned industry based on its business operations.
- Industrials lead with 15.6% of companies, followed by Financials (14.4%), Health Care and Consumer Discretionary (11.8% each), Information Technology (9.7%), Energy (8.8%), Real Estate (7.9%), Utilities (7.1%), Materials (5.9%), Consumer Staples (5.6%), and Telecom Services (1.5%).

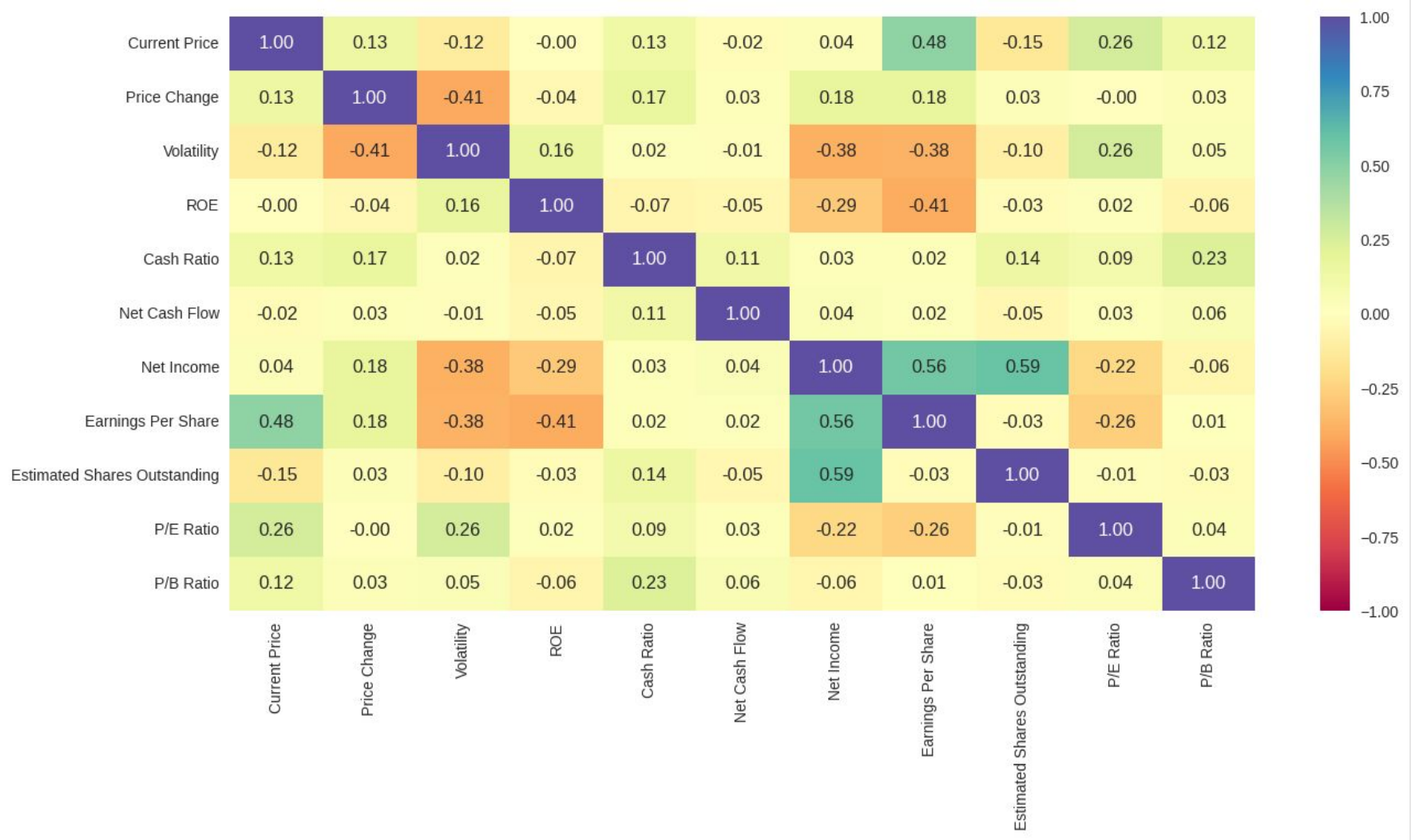
EDA Results

Univariate analysis: Observation on GICS Sub Industry



- Among all the sub industry, Oil and Gas Exploration and Production has the highest count (of 16) as the sub industry group assigned to a company by GICS that best describes its business operations.

EDA Results
Bivariate analysis



- No pair of variables exhibits a strong correlation.
- There is moderate correlation between the following variables:

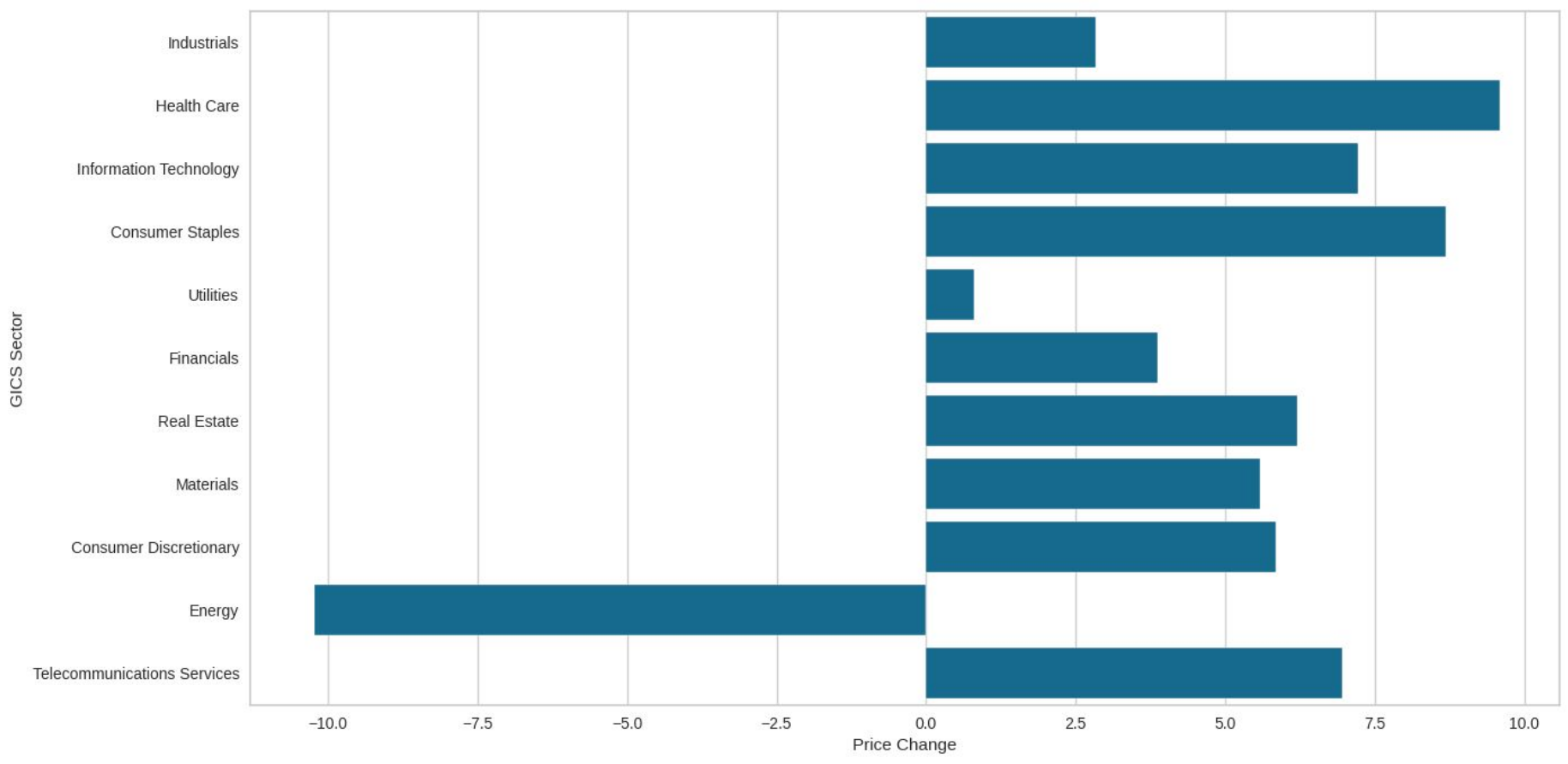
a. Negative correlation between: Volatility and Price Change; Earnings Per Share and ROE

b. Positive correlation between Earnings Per Share and Current Price; Earnings Per Share and Net Income; Estimated Shares Outstanding and Net Income

EDA Results
Bivariate analysis: GICS Sector vs Price Change

Let us check the stocks of which economic sector have seen the maximum price increase on average.

EDA Results
Bivariate analysis: GICS Sector vs Price Change



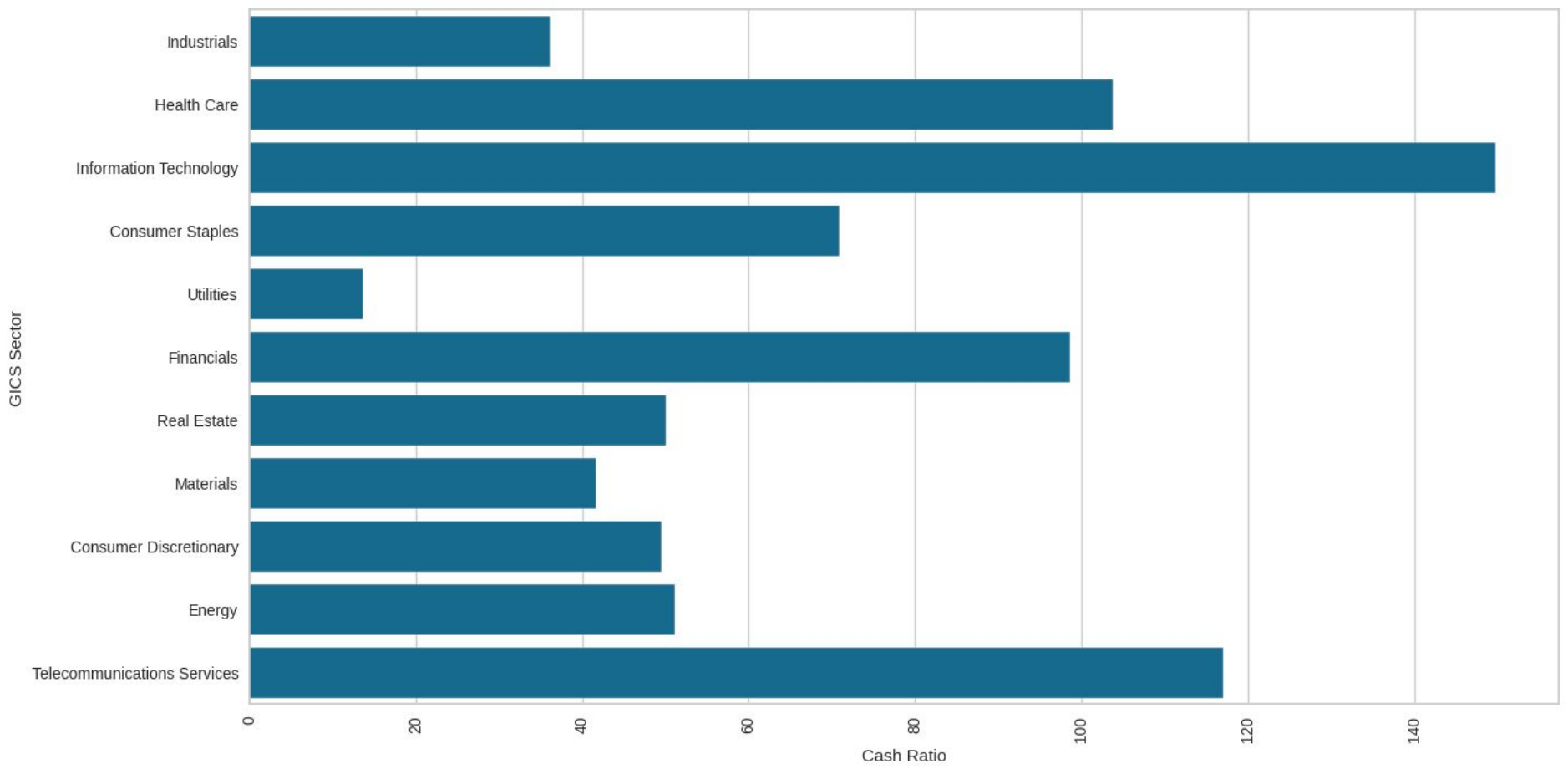
- Stocks of Health Care Sector have seen the maximum price increase on average, followed by Consumer Staples. The least is Energy with negative price change

EDA Results
Bivariate analysis: GICS Sector vs Cash Ratio

Cash ratio provides a measure of a company's ability to cover its short-term obligations using only cash and cash equivalents.

Let us see how the average cash ratio varies across economic sectors.

EDA Results
Bivariate analysis: GICS Sector vs Cash Ratio



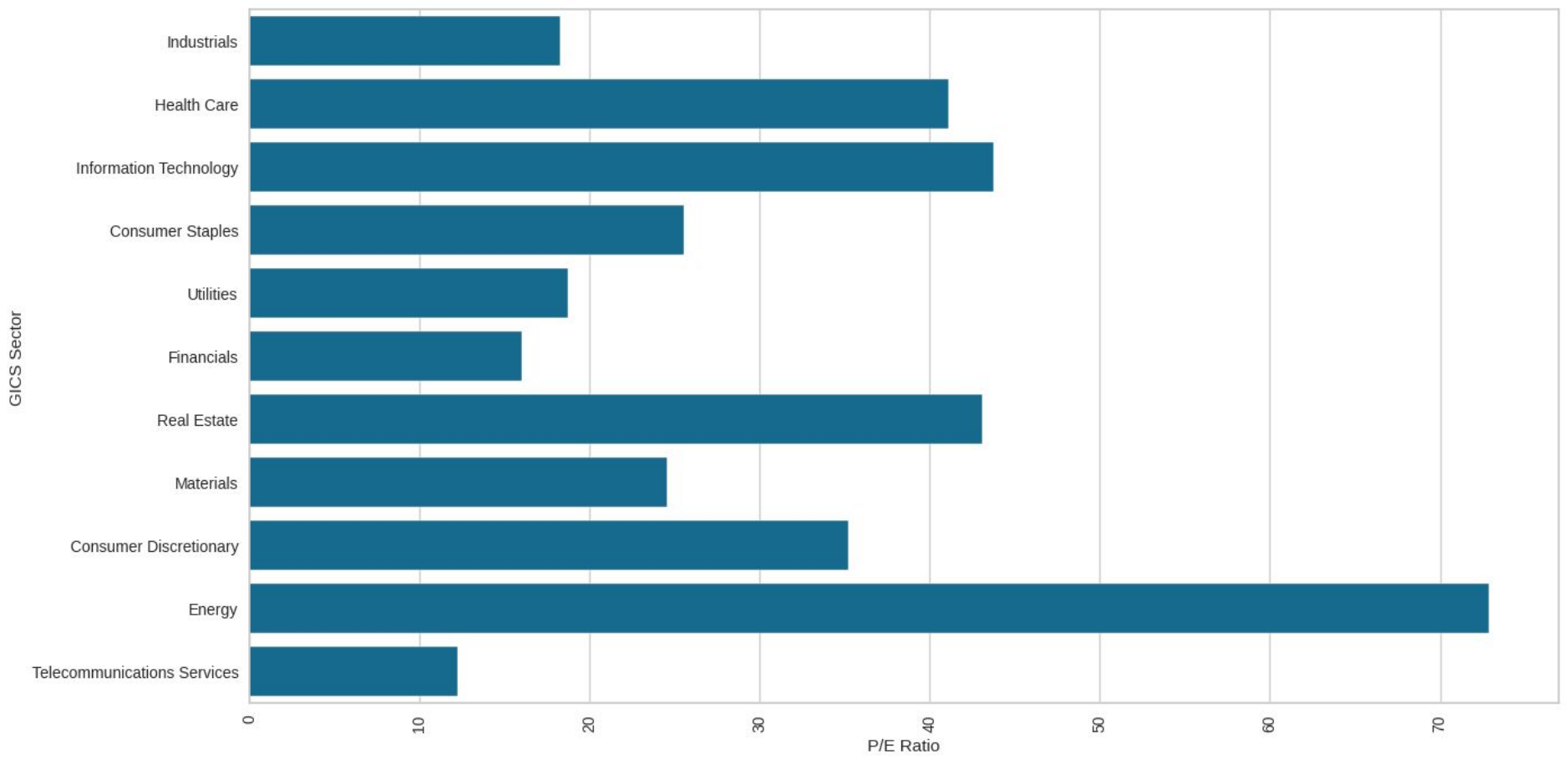
- Information Technology has the highest Cash Ratio, followed by Telecommunications Services.
- The GICS Sector with the least Cash Ratio is Utilities.

EDA Results
Bivariate analysis: GICS Sector vs P/E Ratio

P/E ratios can help determine the relative value of a company's shares as they signify the amount of money an investor is willing to invest in a single share of a company per dollar of its earnings.

Let us see how the P/E ratio varies, on average, across economic sectors.

EDA Results
Bivariate analysis: GICS Sector vs P/E Ratio



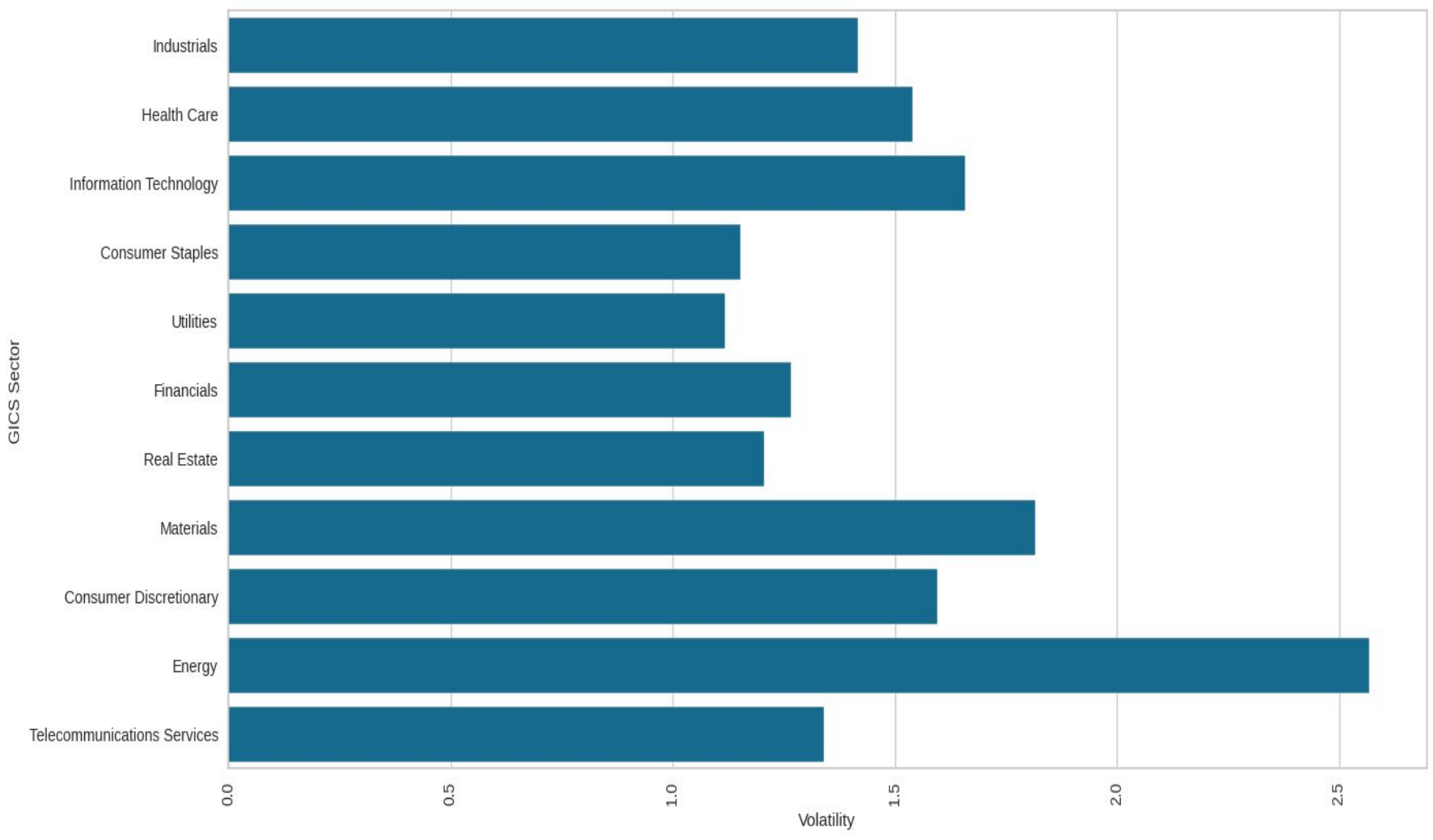
- Among all the GICS Sectors, Energy shows the highest P/E Ratio, followed by Information Technology and Real Estate, while Telecommunications Services records the lowest.

EDA Results
Bivariate analysis: GICS Sector vs Volatility

Volatility accounts for the fluctuation in the stock price. A stock with high volatility will witness sharper price changes, making it a riskier investment.

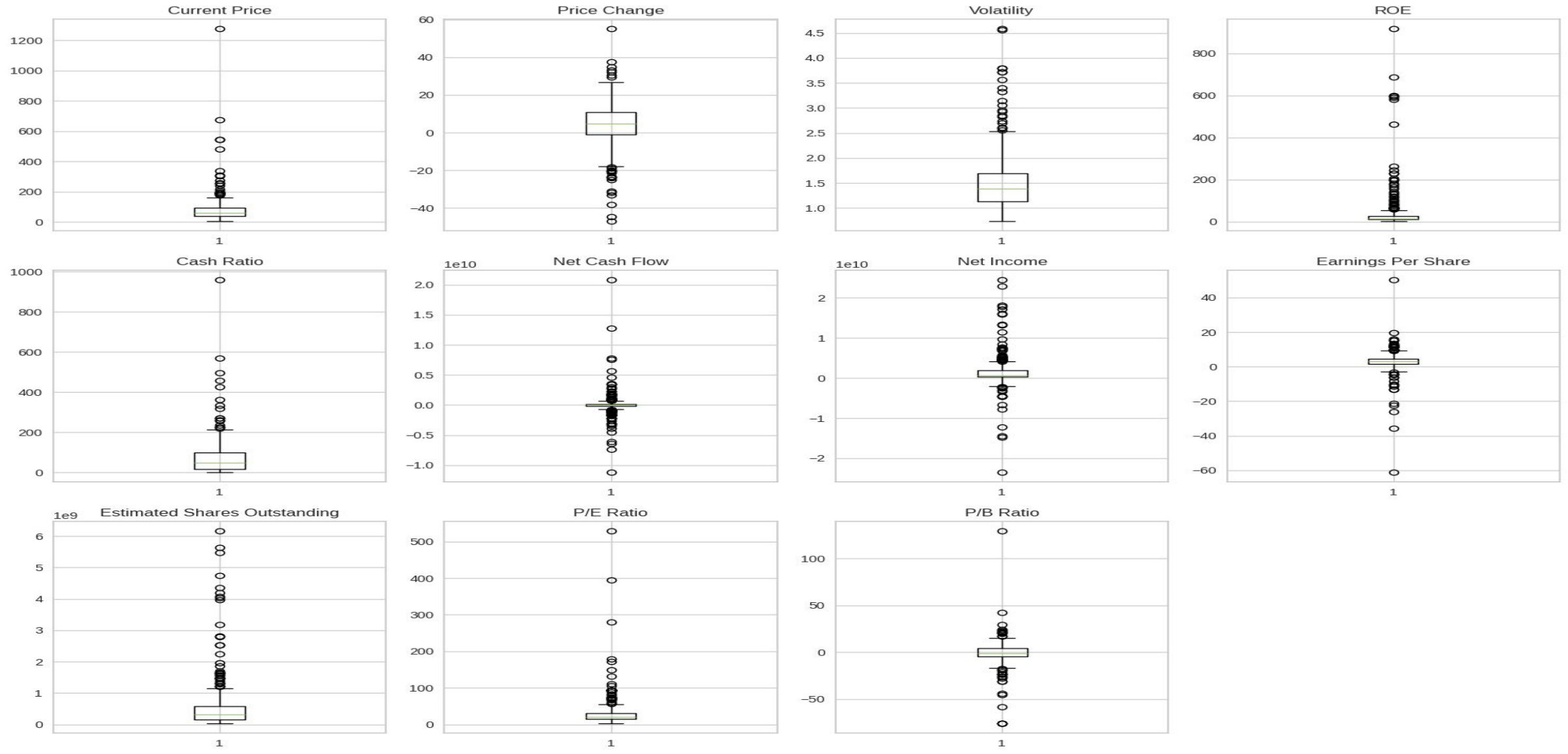
Let us see how volatility varies, on average, across economic sectors.

EDA Results
Bivariate analysis: GICS Sector vs Volatility



- Energy has the highest volatility, making it the most risky sector to invest stocks. This is probably due to its sensitivity to oil and gas price fluctuations, geopolitical events, and regulatory changes.
- Utilities have the lowest volatility probably because they provide essential services with steady demand. Their revenues are relatively stable and less affected by economic cycles or market shocks. This makes utility stocks more predictable and less risky for investors.

Data Preprocessing: Outlier check



Data Preprocessing: Outlier check

- There are eleven boxplots in the image on the previous slide.

- All the boxplots show the presence of outliers.
- Several companies exhibit exceptionally high stock prices, while others have very low or negative net incomes, as well as some with extremely high net incomes.
- Similar patterns of extreme values are observed across the other variables.
- We will not treat the outliers, as these values are typical for stock-related attributes

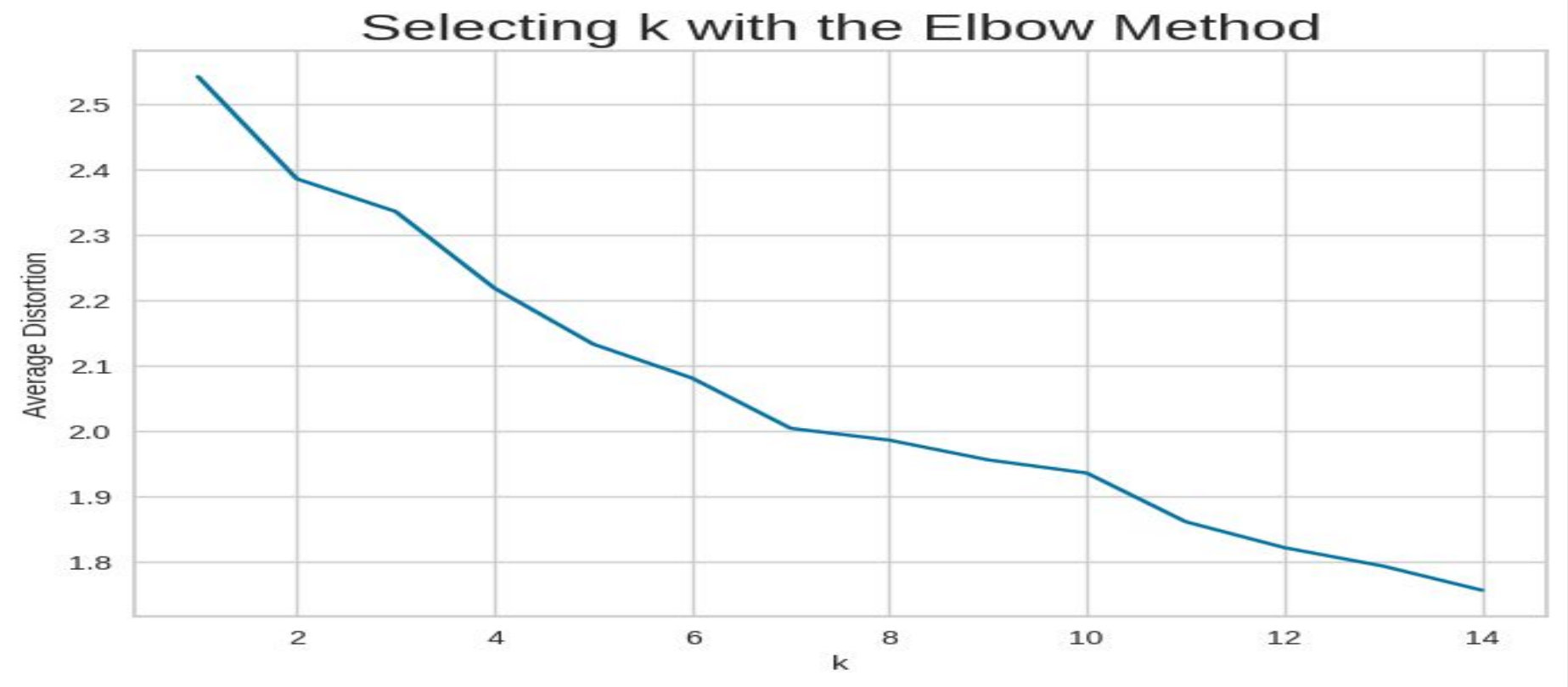
Data Preprocessing: Scaling

- We have decided to scale all our numerical columns via the z-score normalization, meaning that each column will now have mean of zero and standard deviation of 1.

- Scaling ensures all features contribute equally to clustering, preventing variables with larger ranges from dominating the results.

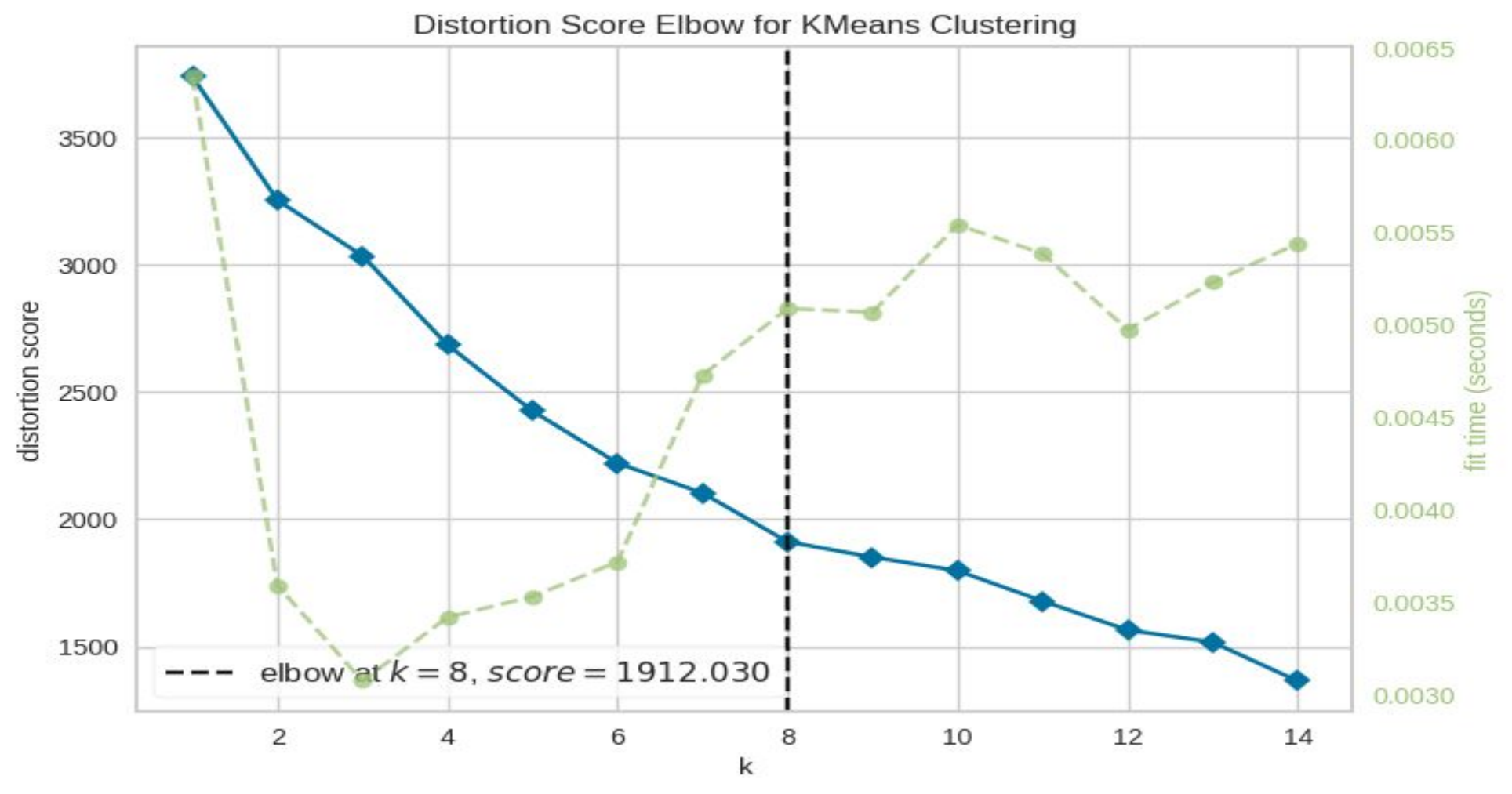
We have created a dataframe of the scaled data.

K-means Clustering: Elbow Plot



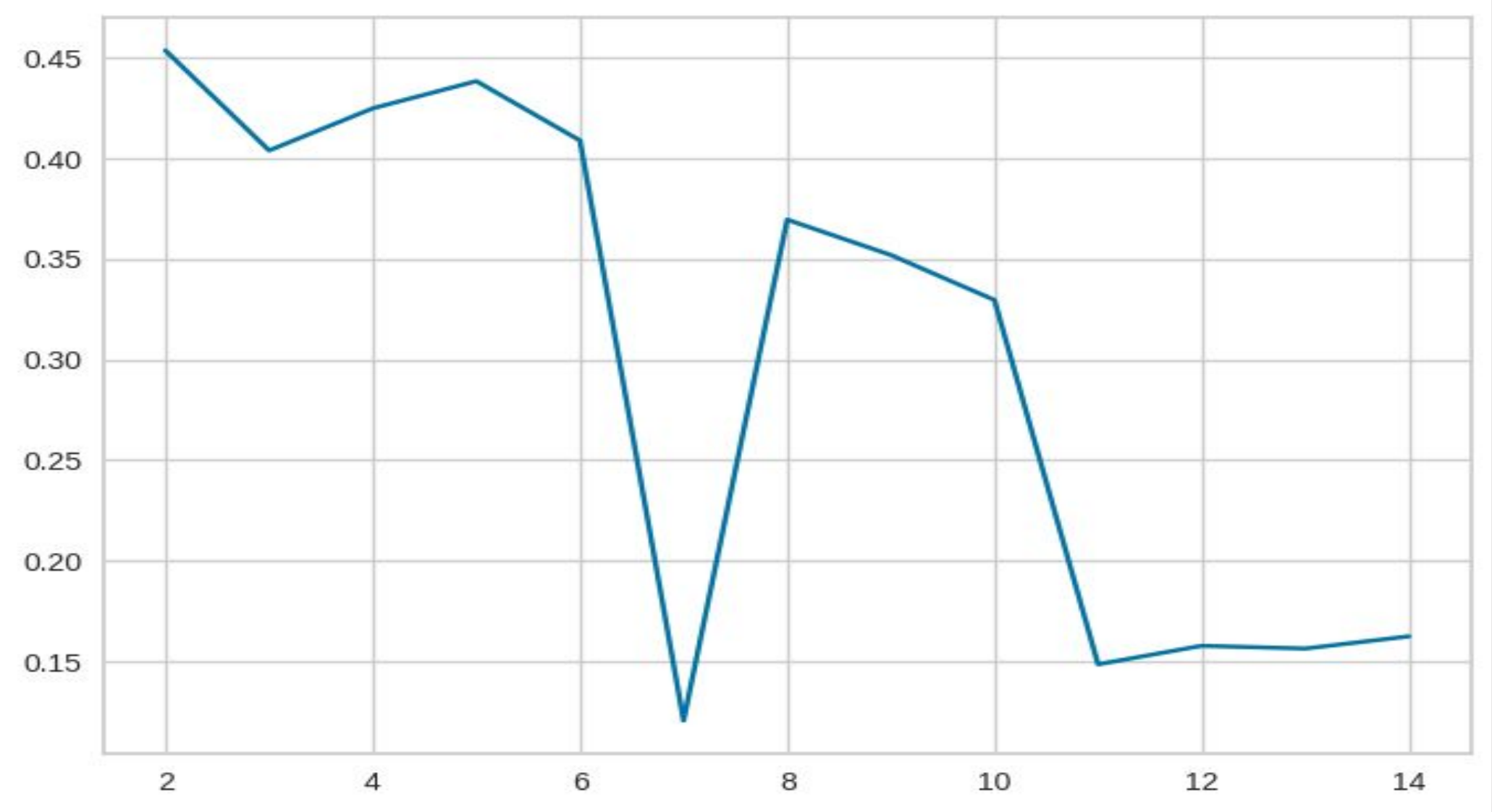
- There are a number of points on the curve where the curve sharply decreases
- So, it is not clear what k should be.

K-means Clustering: Elbow Plot



- Elbow at k=8 has the score of 1912.030
- We will use the Silhouette method to determine k.

K-means Clustering: Silhouette Method

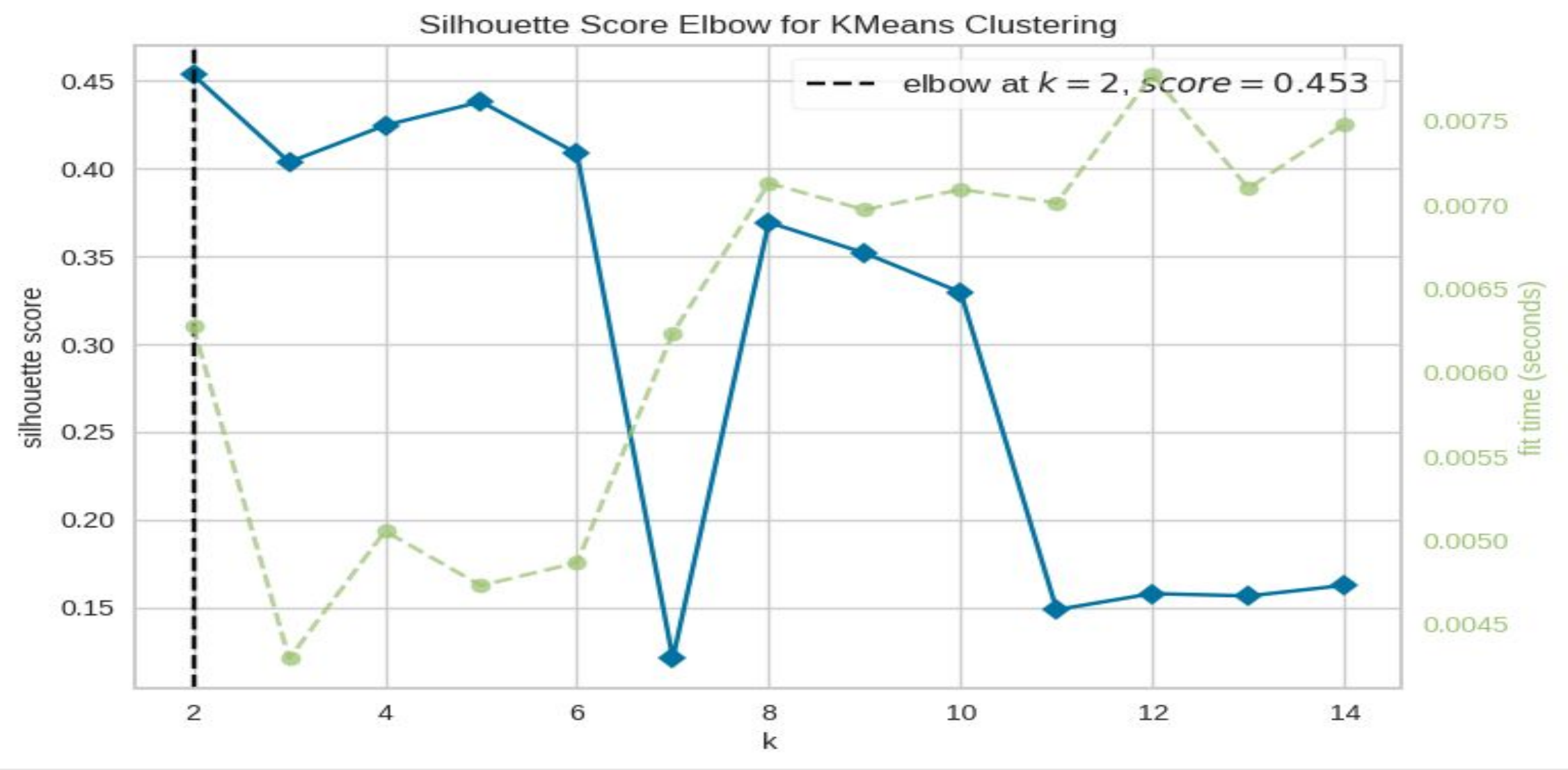


K-means Clustering: Silhouette Method

- The Silhouette score measures how well data points are grouped within their assigned clusters versus how close they are to neighboring clusters.
- A Silhouette score close to +1 shows that the point is well clustered, that is, tight within its own cluster and far from others.
- A Silhouette score that is close to 0 shows that the point is on the boundary between two clusters (not clearly belonging to one).
- A Silhouette score that is close to -1 shows that the point is poorly clustered, likely assigned to the wrong cluster.
- The chosen value of k is where the average Silhouette score is maximized.

Based on the explanations above, n_clusters=2 corresponds to the highest Silhouette score of approx 0.45, so it appears that k=2 is a good choice.

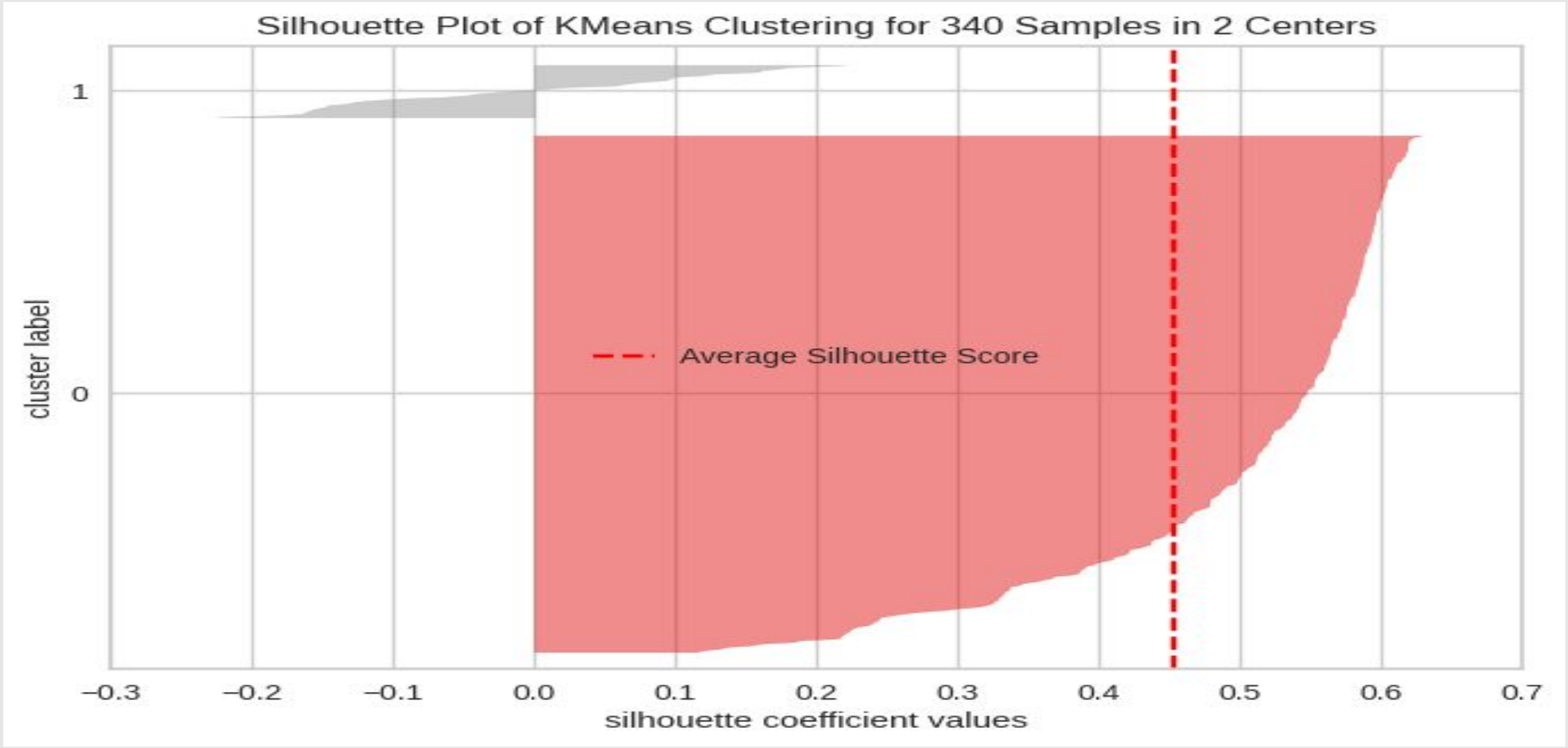
K-means Clustering: Silhouette Method



K-means Clustering: Silhouette Method

- Looking at the image on the previous slide, the choice of k=2 seems to produce the highest Silhouette score.

K-means Clustering: Silhouette Method



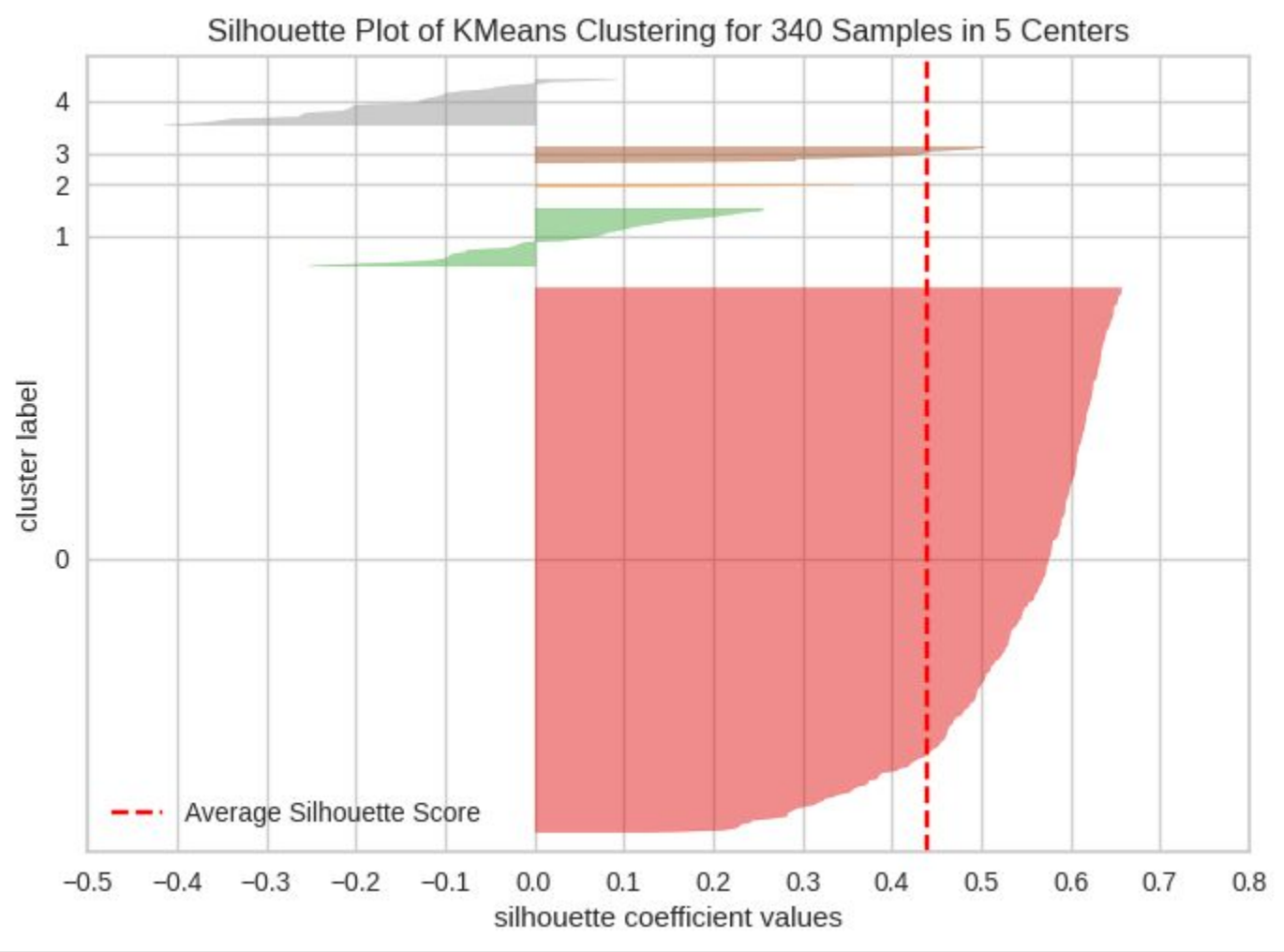
K-means Clustering: Silhouette Method

- Using the SilhouetteVisualizer, we can compare the average Silhouette line across different values of k.
- The choice of k=2 seems to produce the highest Silhouette coefficient values.

- On the next four slides, we show graphs for n_clusters = 5, 6, 8, and 10. In all cases, the Silhouette coefficient is lower than for k = 2, same is true for other values of k, they all produce lower Silhouette coefficient.

- So, we stick to k=2.
- With this in mind, let us create our final model.

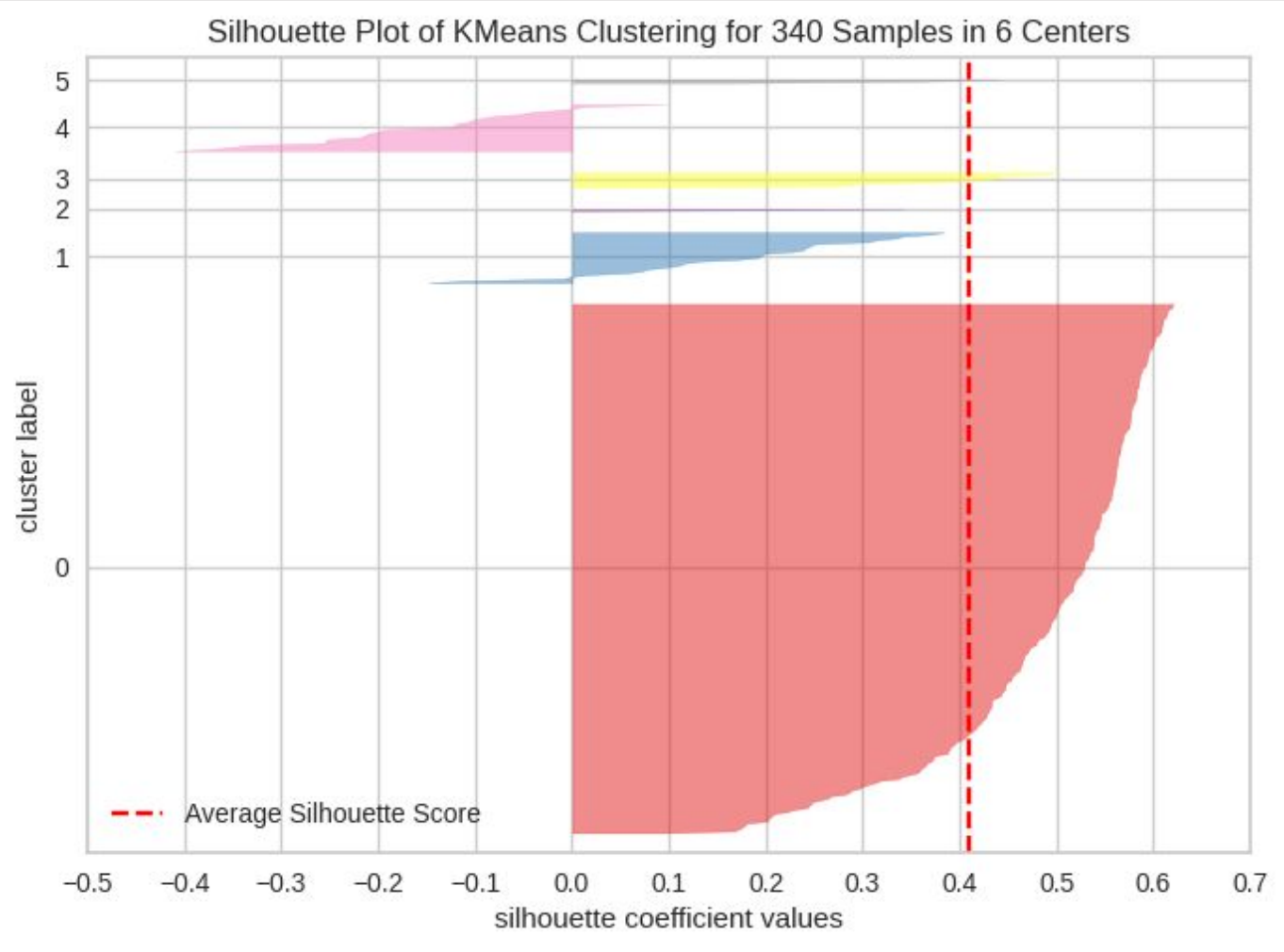
K-means Clustering: Silhouette Method



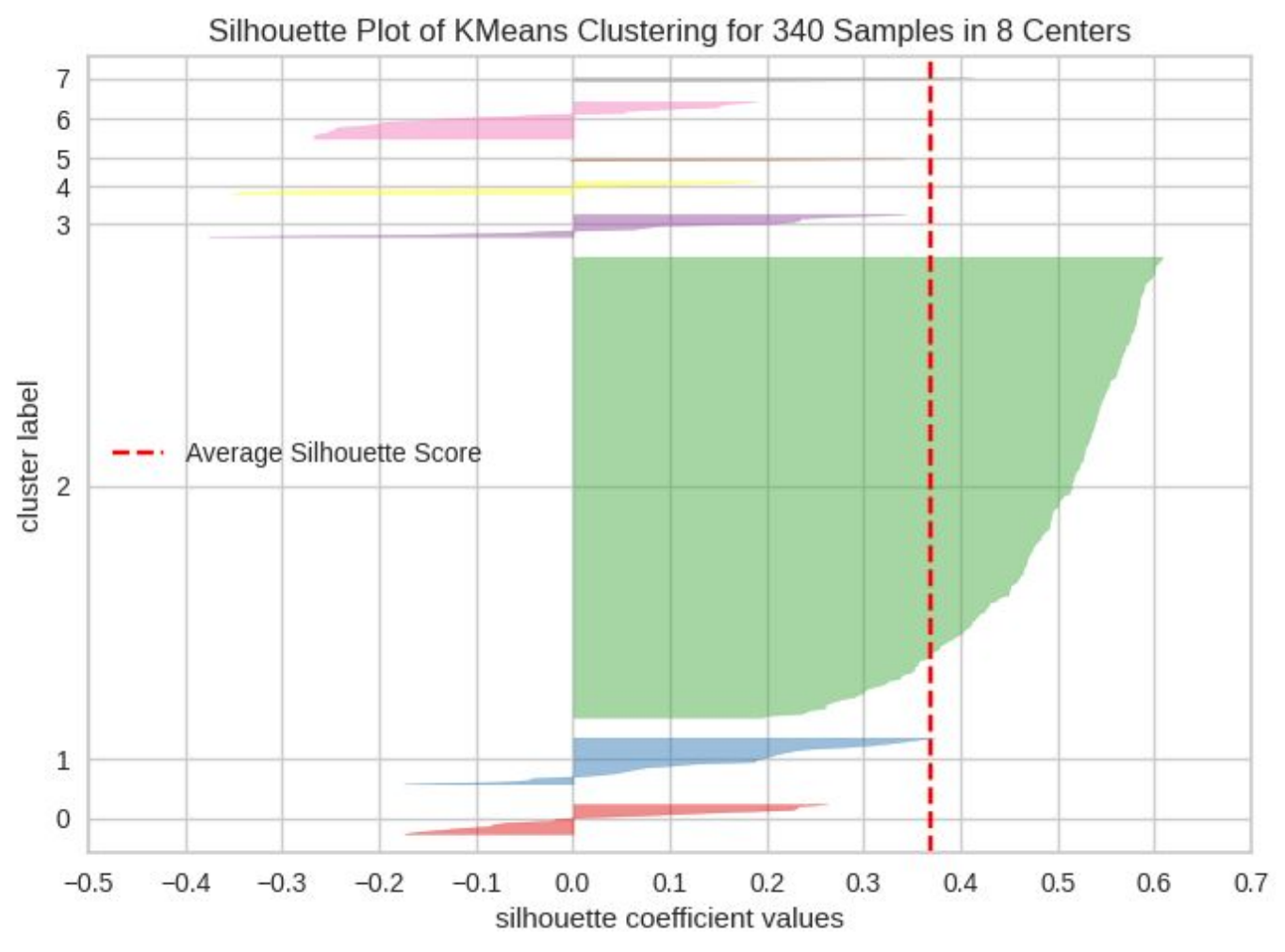
- This is for n_clusters=5

K-means Clustering: Silhouette Method

- This is for n_clusters=6



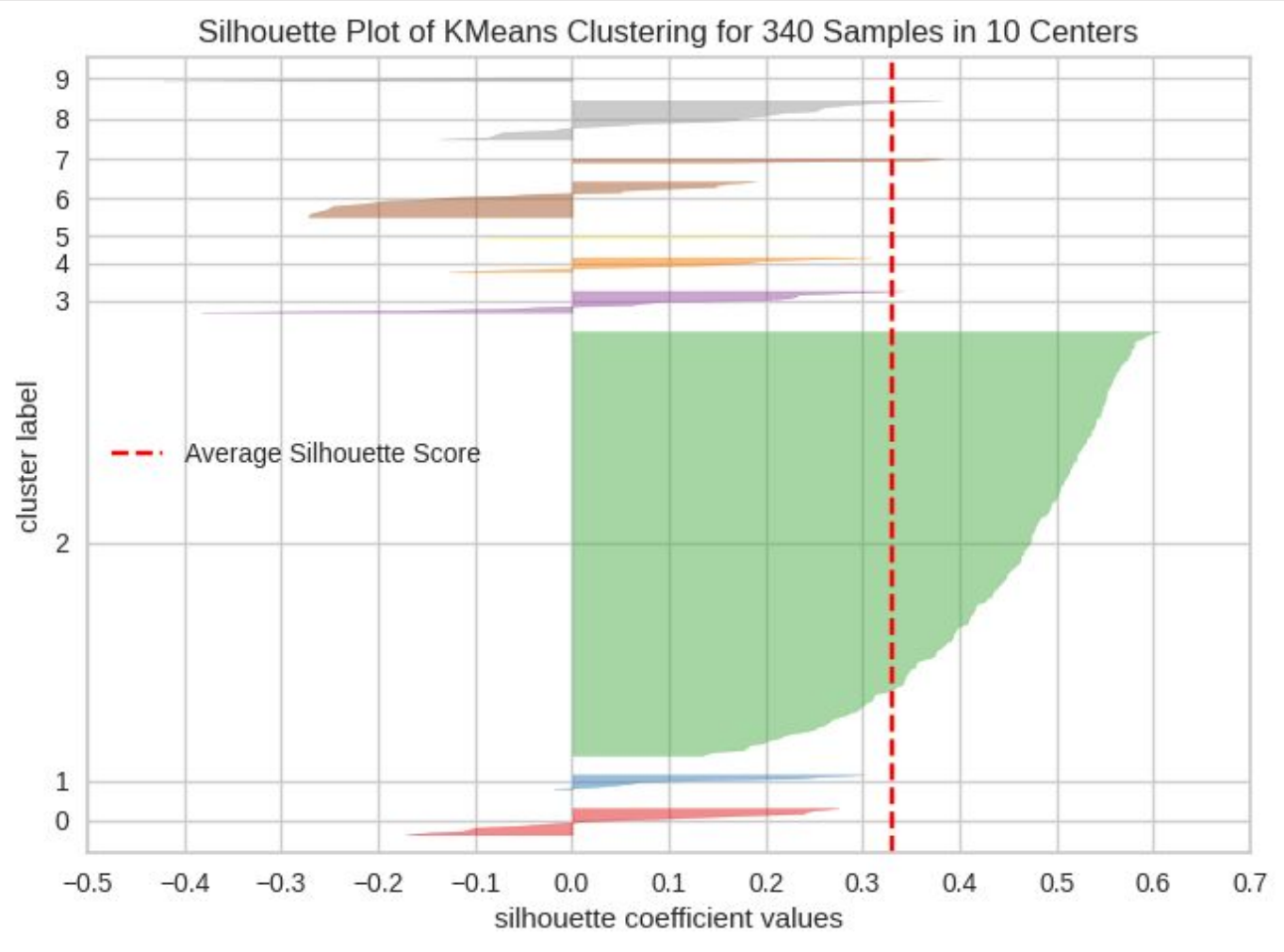
K-means Clustering: Silhouette Method



- This is for n_clusters=8

K-means Clustering: Silhouette Method

- This is for n_clusters=10



K-means Clustering: Creating Final Model

KMeans(n_clusters=2, random_state=1)

- KMeans algorithm will partition our data into two clusters, with consistent repeatable result.

K-means Clustering: Cluster Profiling

There are two clusters 0 and 1, and we will profile them on the next slide

K-means Clustering: Cluster Profiling

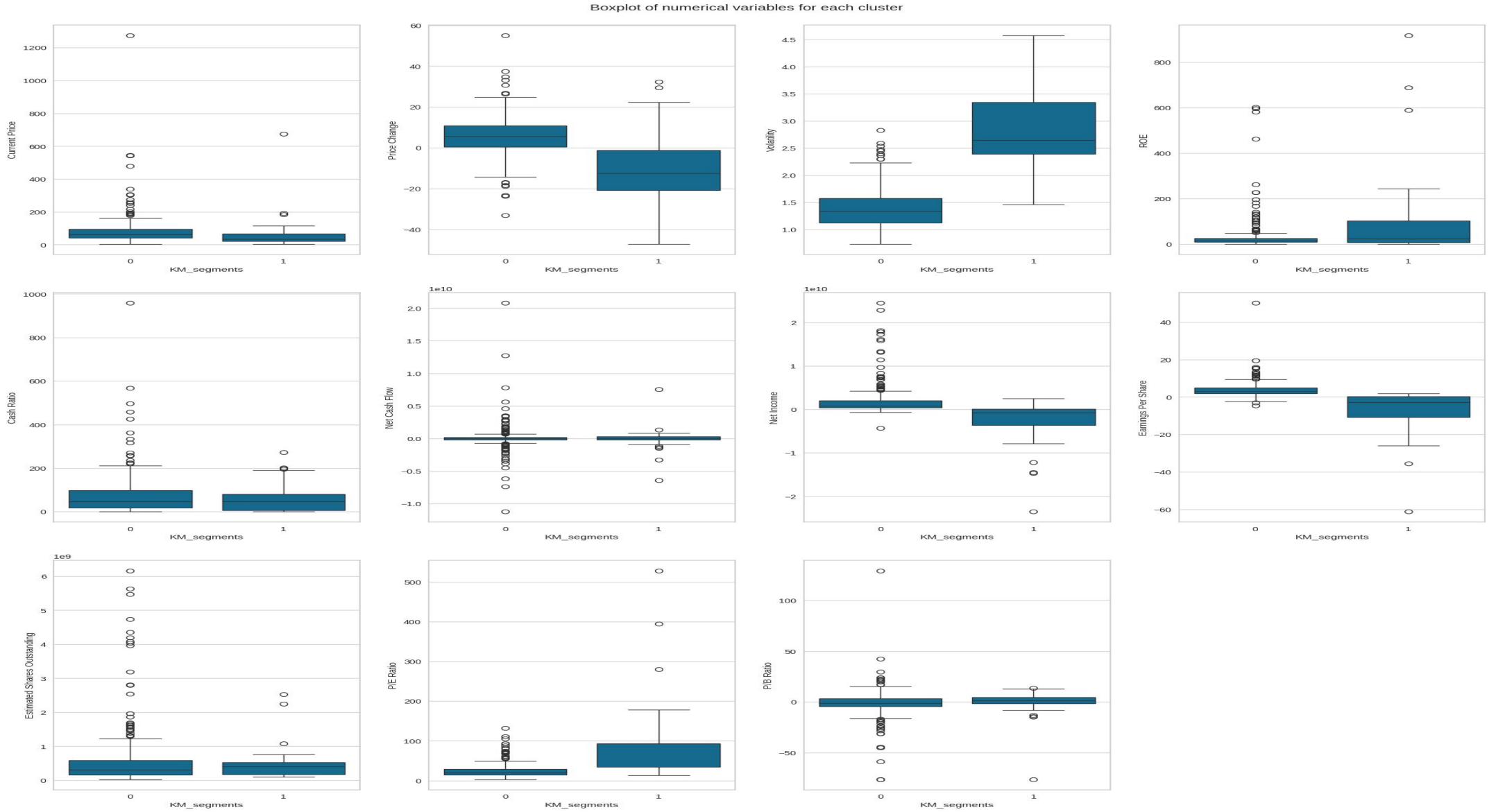
Feature	Cluster 0 (308 count)	Cluster 1 (32 count)
Current Price	82.11	68.84
Price Change	+5.57	-10.29
Volatility	1.39 (low)	2.79 (high)
ROE	31.78	114.81
Cash Ratio	70.56	64.81
Net Cash Flow	+70 million	-84 million
Net Income	+1.99 billion	-3.24 billion
Earnings Per Share (EPS)	3.89	-7.94
Estimated Shares Outstanding	584 million	503 million
P/E Ratio	24.67	109.07
P/B Ratio	-1.78	-1.14
Price Trend / Overall Comment	Positive, stable, profitable	Decreasing, volatile, loss-making

K-means Clustering: Cluster Profiling

KM_segments	GICS Sector	Security
0	Consumer Discretionary	37
	Consumer Staples	19
	Energy	8
	Financials	49
	Health Care	39
	Industrials	52
	Information Technology	29
	Materials	19
	Real Estate	27
	Telecommunications Services	5
	Utilities	24
1	Consumer Discretionary	3
	Energy	22
	Health Care	1
	Industrials	1
	Information Technology	4
	Materials	1

The table above shows how many companies in each GICS sector belong to each KMeans cluster.

K-means Clustering: Cluster Profiling- Boxplots



K-means Clustering: Cluster Profiling- Boxplots

There are 22 boxplots in the image on the previous slide (11 for each cluster).

All the boxplots show the presence of outliers, except the boxplot of Volatility for Cluster 1.

K-means Clustering: Cluster Profiling

Companies in **Cluster 0** have:

- higher average price change
- higher current price
- higher cash ratio
- Lower volatility and ROE
- Wider range of cash flow
- Higher net income and earnings per share
- Higher estimated shares outstanding
- Lower P/E ratio
- Wider range of P/B ratio

when compared to companies in **Cluster 1**

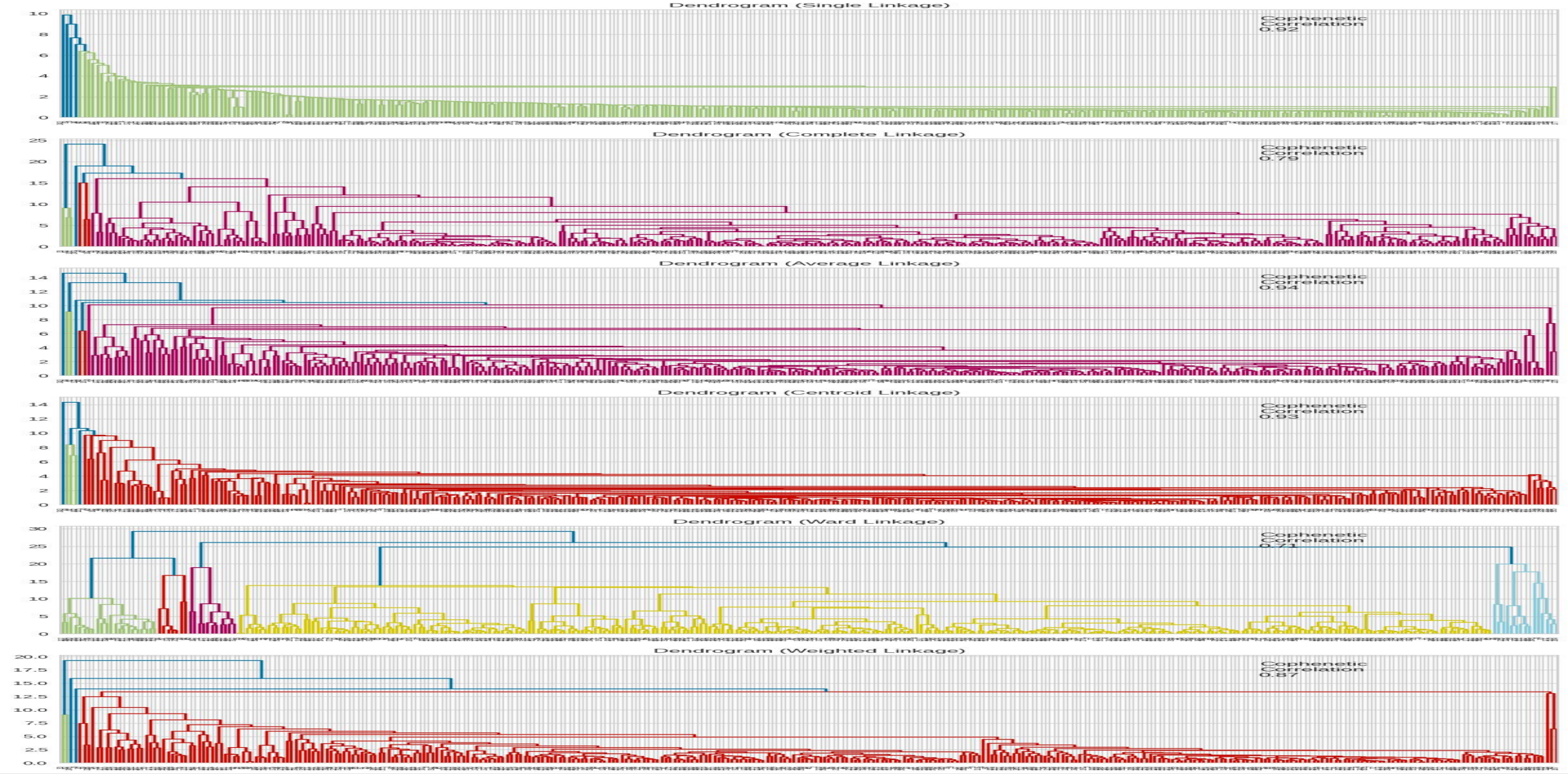
Hierarchical Clustering: Cophenetic Correlation

Highest cophenetic correlation is 0.9422540609560814, which is obtained with Euclidean distance and average linkage.

So, let us explore different linkage methods with Euclidean distance only.

Highest cophenetic correlation is 0.9422540609560814, which is obtained with average linkage.

Hierarchical Clustering: Checking Dendograms



Hierarchical Clustering: Checking Dendrograms

- The average linkage has the highest Cophenetic correlation of 0.94
- The appropriate number of cluster for average linkage appears to be 7
- However, the dendrogram for the average linkage appears to be unbalanced.
- Similar observation holds true for the Centroid and Single linkages with high Cophenetic correlation of 0.93 and 0.92 respectively.
- We need to use a linkage with symmetrical clusters
- The ward linkage with a Cophenetic correlation of 0.71 appears to have balanced clusters at n_clusters=5.
- So, we will use the ward linkage.

Hierarchical Clustering: Creating Model Using sklearn

AgglomerativeClustering

AgglomerativeClustering(n_clusters=5)

Hierarchical Clustering: Cluster Profiling

- There are 5 clusters, namely Clusters 0, 1, 2, 3 and 4. We will compare these clusters.
- Cluster 0 has 15 companies, the highest current stock price, highest price change, cash ratio, earnings per share, P/E and P/B ratio.
- Cluster 1 contains 7 companies with the highest ROE, lowest Cash Ratio, lowest negative Net Cash flow, Net Income, Earnings Per Share and P/B ratio.
- Cluster 2 contains 11 companies with the highest Net Cash Flow and Net Income and Estimated Shares Outstanding.
- Clusters 3 is the most populated cluster with 285 companies, it has a negative P/B Ratio and a low Volatility.
- Cluster 4 has the highest volatility, lowest price change, negative net cash flow, net income and earnings per share.

Hierarchical Clustering: Cluster Profiling

Cluster 0 contains companies associated with the following GICS Sector:

- Consumer Discretionary
- Consumer Staples
- Health Care
- Information Technology
- Real Estate
- Telecommunications Services

Cluster 1 contains companies in the following GICS Sectors

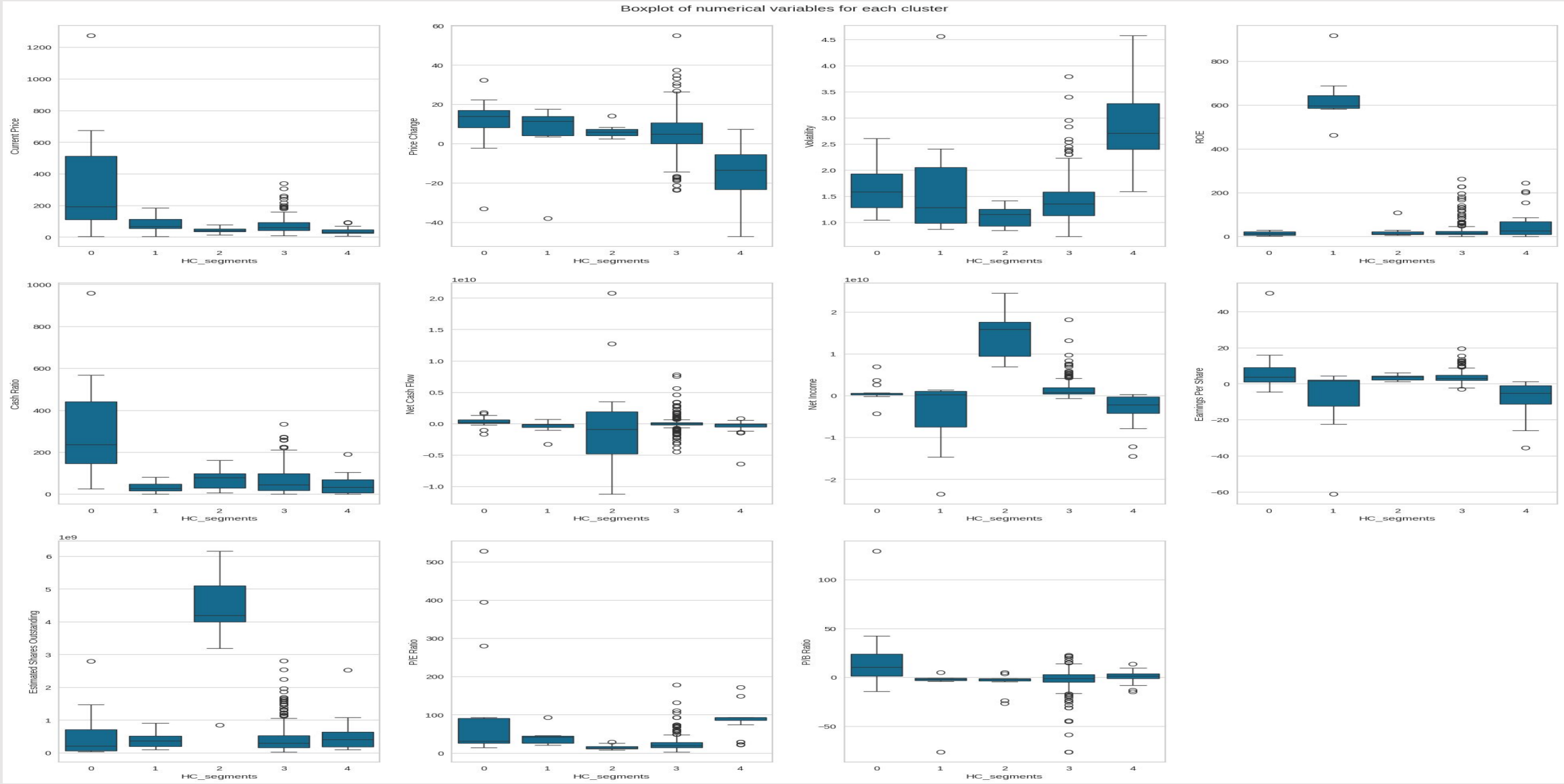
- Consumer Discretionary
- Consumer Staples
- Energy
- Financials
- Industrials

Cluster 2 contains companies in
Consumer Discretionary
Consumer Staples
Energy
Financials
Health Care
Information Technology
Telecommunications Services

Cluster 3 contains companies in
Consumer Discretionary,
Consumer Staples
Energy
Financials
Health Care
Industrials
Information Technology
Materials
Real Estate
Telecommunications Services

Cluster 4 contains companies in
Energy
Information Technology
Materials.

Hierarchical Clustering: Cluster Profiling- Boxplots



Hierarchical Clustering: Cluster-Boxplots

Cluster 0 contains 15 companies with:

- Outliers across nearly all features, except ROE and Volatility.
- Most of the stocks with the highest current prices, the highest median stock price, and a wider price range.
- The highest median price change.
- Most stocks with the highest cash ratio and strong earnings per share (EPS).
- The best P/B ratios and extreme outliers in P/E ratios.

Hierarchical Clustering: Cluster-Boxplots

Cluster 1 includes 7 companies characterized by:

- High stock price changes with a relatively narrow range, and the second-highest median price change.
- A broad range of stock volatility, with the second-highest median volatility.
- Exceptionally high Return on Equity (the highest among all clusters).
- Low levels of Cash Ratio, Net Cash Flow, Estimated Shares Outstanding, P/E ratio, and P/B ratio.
- The lowest Net Income values.
- Predominantly negative Net Income and Earnings Per Share.

Hierarchical Clustering: Cluster-Boxplots

Cluster 2 includes 11 companies characterized by:

- Very low current stock prices and price changes, both within a narrow range.
- The lowest volatility, along with generally small ROE values.
- The lowest Net Cash Flow overall, though with some extreme positive outliers.
- The highest Net Income.
- Relatively low Earnings Per Share, with a narrow range.
- The highest Estimated Shares Outstanding.
- Low P/E ratio and P/B ratio.
- Predominantly negative Net Cash Flow.

Hierarchical Clustering: Cluster-Boxplots

Cluster 3 includes 285 companies characterized by:

- Numerous positive and negative outliers in Net Cash Flow.
- Mostly positive Net Income, with several outliers; the same pattern is observed for Earnings Per Share, Estimated Shares Outstanding, and the P/E ratio.
- The second-lowest median P/E ratio.
- A median P/B ratio near zero, with several outliers.
- Predominantly negative P/B ratio.

Hierarchical Clustering: Cluster-Boxplots

Cluster 4 includes 22 companies characterized by:

- Among the lowest stock cash prices, with one outlier.
- Mostly negative stock price changes, with a negative median.
- Extremely high volatility, indicating companies in this cluster may be risky investments.
- Relatively high ROE.
- Relatively low Cash Ratio, with one outlier.
- Predominantly negative Net Cash Flow, with a relatively small range.
- Predominantly negative Net Income and Earnings Per Share.
- High P/E ratio, with a relatively narrow range.

K-means vs Hierarchical Clustering

You compare several things, like:

- **Which clustering technique took less time for execution?**
- **Which clustering technique gave you more distinct clusters, or are they the same?**
- **How many observations are there in the similar clusters of both algorithms?**
- **How many clusters are obtained as the appropriate number of clusters from both algorithms?**
- **You can also mention any differences or similarities you obtained in the cluster profiles from both the clustering techniques.**

K-means vs Hierarchical Clustering

- The K-means clustering technique was significantly faster to execute than the Hierarchical/Dendrogram clustering technique.
- Compared to K-means, hierarchical clustering resulted in clusters that were more distinct and separable.
- Cluster 0 from K-means is similar to Cluster 0 from Hierarchical in the sense that they both have the highest current price, price change, cash ratio and earnings per share
- ROE is highest in Cluster 1 (K-means) and Cluster 1 (Hierarchical).
- P/B ratio and P/E ratio are highest in Cluster 1 (K-means) and Cluster 0 (Hierarchical).
- Cluster 1 (K-means) has the higher volatility while Cluster 4 (Hierarchical) has the highest volatility.
- K-means clustering resulted in 2 clusters while Hierarchical clustering yielded 5 clusters.
- **Hierarchical clustering appears to be a more suitable approach for analyzing the problem; however, the use of Ward linkage resulted in a lower Cophenetic correlation, indicating a trade-off in cluster accuracy.**

Actionable Insights and Key Take away

- We will examine each cluster obtained from Hierarchical clustering.

Actionable Insights and Key Take away

Cluster 0 (Insights)

Outliers across nearly all features (except ROE and Volatility):

- These companies show unusual or extreme values in many financial metrics, meaning some may represent unique opportunities, but they also carry higher uncertainty.

Highest current stock prices, highest median price, and wider price range:

- These are generally well-established or in-demand companies. Their high prices suggest strong market confidence, but the wide range also means valuations vary significantly across companies in this group.

Greatest median price change:

- Investors can expect both high-growth potential and significant risk. Some companies may deliver exceptional returns, while others could underperform.

High cash ratios and strong EPS:

- Companies here are financially healthy, holding solid liquidity and generating strong earnings relative to shares. This signals operational strength and lower short-term risk.

Best P/B ratios and extreme P/E outliers:

- Many firms appear fairly valued relative to their book value, which is attractive to investors. However, extreme P/E outliers indicate that some companies may be overvalued or undervalued, requiring careful selection.

Actionable Insights and Key Take away

Cluster 0 (Key Take Away)

- Cluster 0 contains financially strong companies with high prices and earnings power, but also wide variability and some extreme outliers.
- They may offer excellent opportunities for growth-oriented investors, but stock-picking is crucial to avoid overpriced or risky outliers.

Actionable Insights and Key Take away

Cluster 1 (Insights)

High stock price changes within a narrow range, with the second-highest median change:

- These companies tend to see significant movement in stock prices, suggesting active trading and opportunities for gains. The relatively narrow range means this behavior is somewhat consistent across the group.

Broad range of volatility, with the second-highest median:

- Investors face substantial risk here. Some companies may be extremely volatile, while others are more stable, making careful selection essential.

Exceptionally high Return on Equity (ROE):

- On paper, these firms are very efficient at generating returns from shareholder capital. However, extremely high ROE can sometimes be misleading if driven by negative equity or unusual financial structures.

Low Cash Ratio, Net Cash Flow, Shares Outstanding, P/E, and P/B ratios:

- These companies may be undercapitalized and less liquid, with limited cash reserves. The low valuation ratios suggest they could be undervalued by the market or simply lack investor confidence.

Lowest Net Income values:

- Profitability is a major concern here. Despite strong ROE, these companies are struggling to generate bottom-line earnings.

Predominantly negative Net Income and EPS:

- Persistent losses indicate poor earnings performance, which makes these stocks risky long-term investments.

Actionable Insights and Key Take away

Cluster 1 (Key Take Away)

- Cluster 1 represents high-risk, speculative companies.
- The strong stock price movements and very high ROE might attract short-term traders, but negative earnings, weak cash positions, and high volatility make them unsuitable for conservative or income-focused investors.
- They may offer short-term trading opportunities but carry significant downside risk.

Actionable Insights and Key Take away

Cluster 2 (Insights)

Very low stock prices and price changes, within a narrow range:

- These are low-priced, relatively stable stocks. They are unlikely to deliver explosive short-term gains but may appeal to cautious investors seeking predictability.

Lowest volatility and small ROE values:

- The stability is reinforced by low volatility, but the modest ROE suggests these firms are not particularly efficient at generating shareholder returns.

Lowest Net Cash Flow overall, with some positive outliers:

- Most companies here struggle to generate cash from operations, raising concerns about liquidity. However, the few outliers with strong positive cash flow could be hidden gems worth closer analysis.

Highest Net Income:

- Despite weak cash flows, these firms report strong accounting profits. This could be due to non-cash items (like accruals) or differences in earnings quality, so investors should be cautious.

Relatively low EPS, with a narrow range:

- Profits per share are modest, partly because these companies also have a very high number of shares outstanding.

Highest Estimated Shares Outstanding:

- A large share base dilutes earnings, limiting EPS growth potential. It also suggests many companies may rely on issuing equity to raise capital.

Low P/E and P/B ratios:

- The market values these firms cheaply relative to their earnings and book value, indicating they may be undervalued or that investors are skeptical of their ability to sustain profits.

Predominantly negative Net Cash Flow:

- This reinforces liquidity concerns, as strong reported earnings are not translating into actual cash.

Actionable Insights and Key Take away

Cluster 2 (Key Take Away)

- Cluster 2 represents low-priced, low-volatility companies with strong reported Net Income but weak cash generation and diluted earnings.
- They may look undervalued based on P/E and P/B ratios, but cash flow issues raise red flags.
- Investors should be cautious, focusing on whether profits are sustainable and backed by real cash flows.

Actionable Insights and Key Take away

Cluster 3 (Insights)

Numerous positive and negative outliers in Net Cash Flow:

- Cash generation is highly inconsistent. Some firms may have exceptionally strong liquidity, while others are cash-strapped, making this a mixed group with uneven financial stability.

Mostly positive Net Income, with several outliers (also in EPS, shares outstanding, and P/E):

- Overall, companies in this cluster are profitable, but the presence of significant outliers means earnings quality and shareholder returns vary widely. Some firms could be high performers, while others may face structural issues.

Second-lowest median P/E ratio:

- These companies tend to be relatively cheap compared to earnings, which may make them attractive for value investors. However, extremely low P/E ratios can also signal market skepticism about future growth.

Median P/B ratio near zero, with several outliers:

- This suggests the market values many of these companies close to (or below) their book value. Outliers indicate that some are highly overvalued or undervalued, requiring careful stock selection.

Predominantly negative P/B ratio:

- A negative P/B typically reflects companies with negative equity, which can be a sign of financial distress despite profitability on paper. This introduces risk.

Actionable Insights and Key Take away

Cluster 3 (Key Take Away)

- Cluster 2 is a mixed bag: many companies are profitable and appear undervalued based on P/E ratio, but weak balance sheets (negative equity leading to negative P/B ratio) and inconsistent cash flow increase the risk profile.
- For investors, it offers potential value opportunities but requires thorough due diligence to avoid financially unstable firms.

Actionable Insights and Key Take away

Cluster 4 (Insights)

Among the lowest stock prices, with one outlier:

- These are low-priced stocks, often considered “cheap,” but low price alone doesn’t necessarily mean good value.

Mostly negative price changes, with a negative median:

- The overall trend is downward, suggesting weak market sentiment or underperformance.

Extremely high volatility:

- These stocks are highly unpredictable, making them risky. Price swings can create opportunities for short-term traders, but they carry significant downside for long-term investors.

Relatively high ROE:

- On the surface, these firms seem efficient at generating returns from shareholder equity. However, given their weak earnings, high ROE might be artificially inflated by small or negative equity bases.

Relatively low Cash Ratio, with one outlier:

- Most companies have limited liquidity, which raises concerns about their ability to cover short-term obligations.

Predominantly negative Net Cash Flow (with a small range):

- Cash generation is consistently poor, and there’s little variation across the group: a strong signal of weak operational health.

Predominantly negative Net Income and EPS:

- These firms are generally unprofitable, reducing their long-term investment appeal.

High P/E ratio, with a narrow range:

- Despite losses and weak fundamentals, the market still assigns high valuations relative to earnings. This suggests potential overvaluation or investor speculation.

Actionable Insights and Key Take away

Cluster 4 (Key Take Away)

- Cluster 3 represents high-risk, speculative stocks.
- They are low-priced but volatile, with declining prices, negative cash flows, and poor earnings.
- While high ROE and high P/E ratios may catch attention, they likely reflect structural distortions rather than real strength.
- These companies might appeal to aggressive traders but are unsuitable for conservative or long-term investors.

Overall Insights and Key Take away

Trade&Ahead should first evaluate clients' financial objectives, risk appetite, and investment preferences, and subsequently suggest a cluster of stocks as a potential portfolio tailored to these needs.

We provide insights and recommendations based on the following yardsticks:

- Potential for Growth
- Stability vs Volatility
- Value of Investment
- Other comments

Overall Insights and Key Take away

Potential for Growth

- Investors may consider allocating resources to Cluster 1 and Cluster 4 for growth-oriented opportunities.
- Cluster 1 is characterized by substantial stock price fluctuations and high ROE, while Cluster 4 contains highly volatile stocks that could generate short-term gains.
- Caution is advised, however, due to negative earnings, weak cash flows, and elevated P/E ratios.

Overall Insights and Key Take away

Stability vs Volatility

- A diversified investment approach across clusters is recommended to manage risk and volatility.
- Cluster 2 offers stable earnings with low volatility, while selective exposure to Cluster 1 or Cluster 4 can provide higher growth potential.
- This strategy supports a more balanced and resilient portfolio.

Overall Insights and Key Take away

Value of Investment

- Investments in Cluster 2 and Cluster 3 may be suitable for value-focused strategies.
- Cluster 2 exhibits strong net income alongside low P/E and P/B ratios, whereas Cluster 3 demonstrates predominantly positive earnings and favorable P/E ratios.
- Thorough due diligence is recommended to verify that these companies are not experiencing underlying cash flow or balance sheet weaknesses.

Overall Insights and Key Take away

Other comments

- Elevated ROE in Clusters 1 and 4 may be misleading due to negative equity or losses.
- Investors should emphasize cash flow stability and earnings quality in conjunction with valuation metrics when making investment decisions.

Overall Insights and Key Take away

Other comments

- For most investors seeking a balance of stability, value, and moderate growth, Cluster 2 is the most favorable, followed by Cluster 3 for those willing to accept moderate risk for potential upside.
- Clusters 1 and 4 are better suited for speculative or short-term trading strategies.