UNSUPERVISED LEARNING PROJECT

GUIDED

LIST OF CONTENT

LIST OF FIGURES

LIST OF TABLES

PROBLEM STATEMENT

❑ CONTEXT

Investing in stocks has long been a proven method for building wealth, fighting inflation and taking advantage of tax benefits. With the power of compound interest, starting early can significantly grow savings for the future goals, such as retirement. To maximize returns and minimize risk, a diversified portfolio is essential. However, analyzing numerous financial metrics for stock selection can be overwhelming. Cluster analysis offers a solution by grouping stocks with similar characteristics and minimizing correlations helping investors spread their risk across various market segments.

Trade & Ahead, a financial consultancy firm, specializes in offering personalized investment strategies to their clients. As a data scientist, I have to analyze the stock price data and financial indicators for several companies listed on the New York Stock Exchange. My role involves examining the data, classifying the stocks into distinct groups based on their attributes and providing insights into the characteristics of each group to support tailored investment recommendations.

❑ DATA DICTIONARY

- **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)

❑ DATA OVERVIEW

- There are **340 rows** and **15 columns** present in the dataset.
- All the columns have complete data, there is no missing values in the dataset.
- Data types:

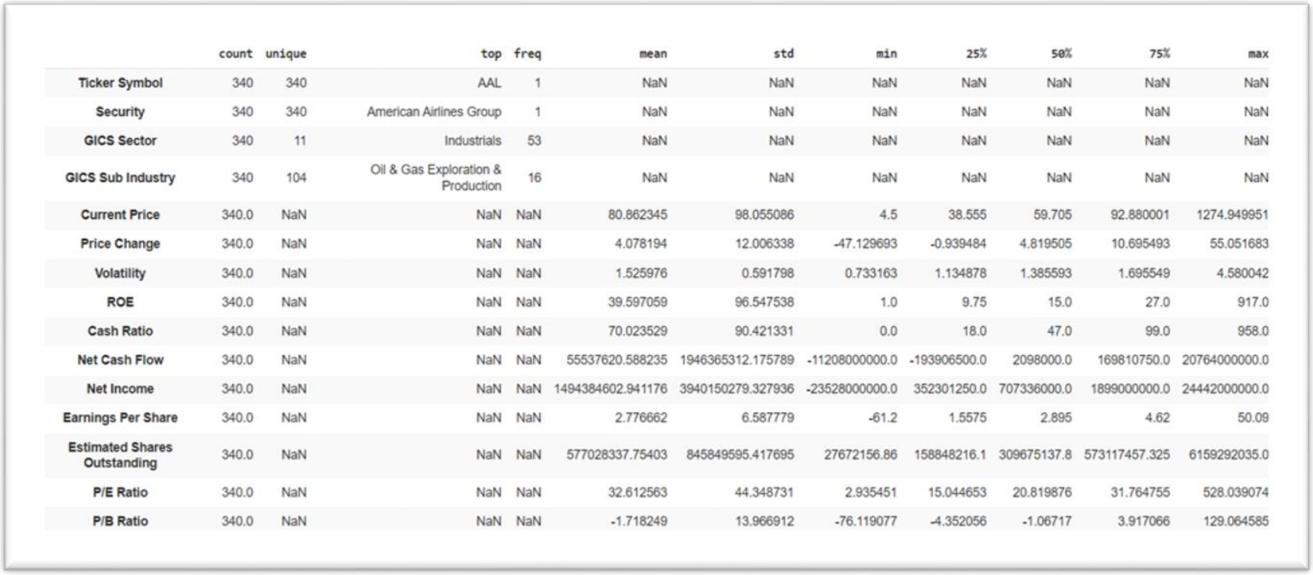
a) Float64: 7

b) Int64: 4

c) Object: 4

- There are no duplicate values in the dataset.

❑ STATISTICAL SUMMARY OF THE DATASET



	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Ticker Symbol	340	340	AAL	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Security	340	340	American Airlines Group	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
GICS Sector	340	11	Industrials	53	NaN	NaN	NaN	NaN	NaN	NaN	NaN
GICS Sub Industry	340	104	Oil & Gas Exploration & Production	16	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Current Price	340.0	NaN	NaN	NaN	80.862345	98.055086	4.5	38.555	59.705	92.880001	1274.949951
Price Change	340.0	NaN	NaN	NaN	4.078194	12.006338	-47.129693	-0.939484	4.819505	10.695493	55.051683
Volatility	340.0	NaN	NaN	NaN	1.525976	0.591798	0.733163	1.134878	1.385593	1.695549	4.580042
ROE	340.0	NaN	NaN	NaN	39.597059	96.547538	1.0	9.75	15.0	27.0	917.0
Cash Ratio	340.0	NaN	NaN	NaN	70.023529	90.421331	0.0	18.0	47.0	99.0	958.0
Net Cash Flow	340.0	NaN	NaN	NaN	55537620.588235	1946365312.175789	-11208000000.0	-193906500.0	2098000.0	169810750.0	20764000000.0
Net Income	340.0	NaN	NaN	NaN	1494384602.941176	3940150279.327936	-23528000000.0	352301250.0	707336000.0	1899000000.0	24442000000.0
Earnings Per Share	340.0	NaN	NaN	NaN	2.776662	6.587779	-61.2	1.5575	2.895	4.62	50.09
Estimated Shares Outstanding	340.0	NaN	NaN	NaN	577028337.75403	845849595.417695	27672156.86	158848216.1	309675137.8	573117457.325	6159292035.0
P/E Ratio	340.0	NaN	NaN	NaN	32.612563	44.348731	2.935451	15.044653	20.819876	31.764755	528.039074
P/B Ratio	340.0	NaN	NaN	NaN	-1.718249	13.966912	-76.119077	-4.352056	-1.06717	3.917066	129.064585

FIGURE 1

➢ Insights based on the figure:

1. Stock Prices:

- The average stock price is $80.86, with significant variability (standard deviation of $98.05)
- The minimum stock price is $4.50, while the maximum is $1274.95, indicating a wide range of stock prices in the dataset.

- The distribution suggests a majority of stocks are priced below $100 but there are few high-priced outliers.

2. Price Change:

- The average price is $4.08, but the high standard deviation of 12.00 indicates substantial variation.
- The minimum price change is -$47.13, while the maximum is $55.05, showing that stocks have both lost and gained values significantly over the time.

3. Volatility:

- The average volatility is 1.53, with a minimum of 0.73 and maximum of 4.58
- This indicates that most stocks have moderate volatility with some more volatile stocks being clear outliers.

4. Return On Equity (ROE):

- The average ROE is 39.60 but with a high standard deviation of 96.55 indicating extreme variation.
- The maximum ROE is 91.7% which suggests that some companies have extremely high profitability relative to their equity.

5. Cash Ratio:

- The average cash ratio is 70.02, but the standard deviation is also high (90.42) with a minimum of 0 and maximum of 958.
- This shows that some companies hold a large portion of their assets in cash, while others hold very little.

6. Net Cash Flow:

- The average net cash flow is around $55.54 million, but there is a significant spread with some companies reporting as much as $20.76 billion and others showing large negative cash flows (up to -$11.21 billion).
- This highlights the wide variety in the company sizes and financial health.

7. Net Income:

- The average net income is $1.149 billion, but there is large variability (standard deviation of $3.94 billion).
- Some companies have significant losses (as much as -$23.53 billion) while the most profitable companies report incomes as high as $24.446.

8. Earnings Per Share (EPS):

- The average EPS is $2.78 with a wide range from -$61.20 to $50.09, showing large differences in profitability per share.

- Negative EPS values suggest companies with losses.

9. Estimated Share Outstanding:

- The average company has about 577 million shares outstanding with some companies as small as 27 million and others with over 6 billion shares.
- This suggests a broad range of company sizes from small firms to large corporations.

10. Price-to-Earnings (P/E) Ratio:

- The average P/E Ratio is 32.61, with a high standard deviation of 44.35
- Some companies have extremely high P/E Ratios (up to 528), suggesting they are highly valued compared to their earnings while others much lower with a minimum of 2.94

11. Price-to-Book (P/B) Ratio:

- The P/B Ratio has an unusual mean of –1.71 likely due to few highly negative values skewing the distribution.
- The range from –76.12 to 129.06 suggests that some stocks are trading well below their book value, while others are trading significantly above.

These statistics highlight the diversity in the financial performance, size and valuation of the companies in the dataset.

The noticeable variation in many metrics such as ROE, net cash flow, P/E ratio indicates that the dataset contains companies across a wide range of industries and financial health.

❑ EDA (EXPLORATORY DATA ANALYSIS)

❖ UNIVARIATE ANALYSIS

♦ **Box plot & Histogram of ‘Current Price’ distribution:**

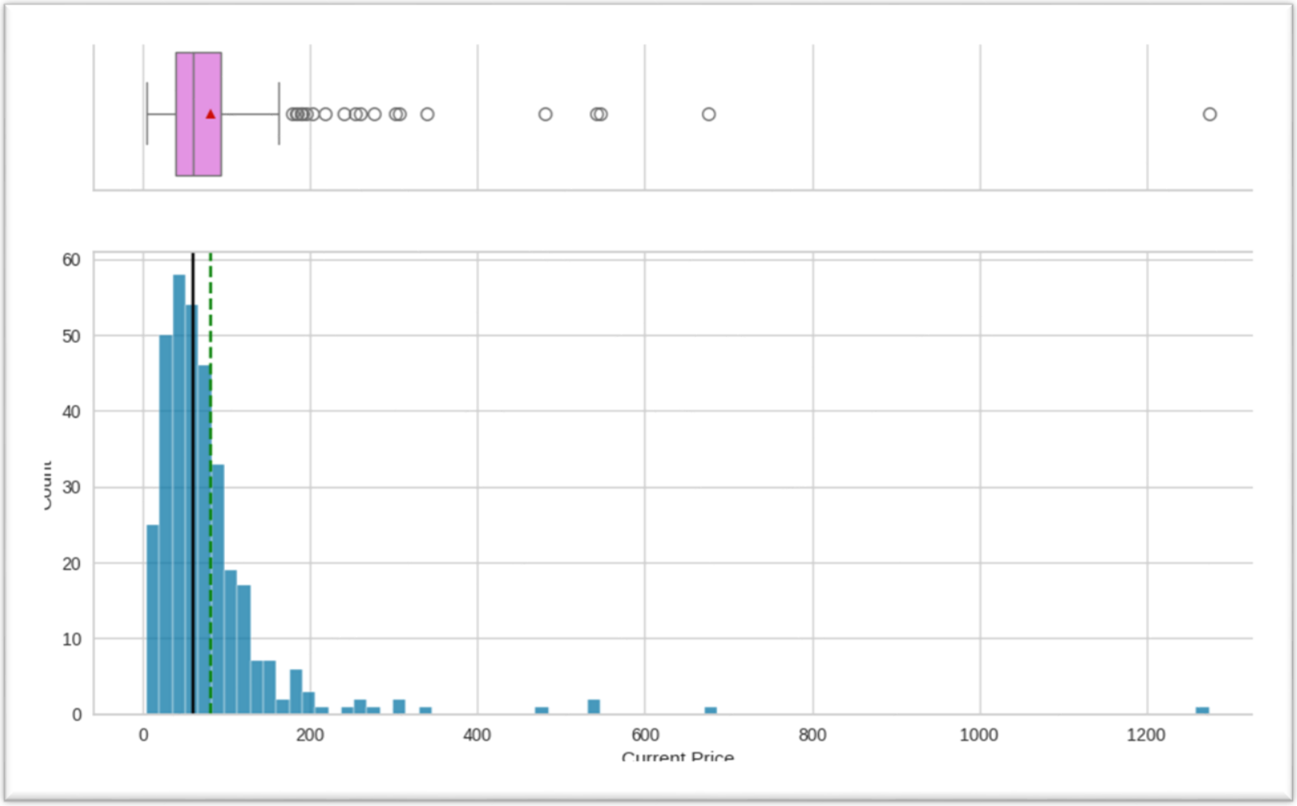


FIGURE 2

➢ Insights on the box plot:

- The box plot shows that the majority of the data points are clustered towards the lower end of the price range, as indicated by the tight IQR.
- The median is relatively close to the lower boundary of the IQR.
- There are several outliers which are significantly higher in value than the majority of the data, with prices reaching beyond 1200.

- The box plot suggests a right skewed distribution as there are more outliers on the high end and the right whisker is longer compared to the left indicating a long tail.

➢ Insights based on the histogram:

- The majority of the data points are concentrated around the lower price values, with the highest frequency occurring between 0 and 100.
- The mode appears to be between 50 and 100 with a count close to 60.
- Similar to boxplot, histogram shows that the distribution is heavily skewed to the right as there are few data points beyond 300.
- The visualization indicates a non-normal distribution with a concentration of lower-priced items and small set of outliers at the higher end.

♦ Box plot & Histogram of ‘Price change’ distribution:

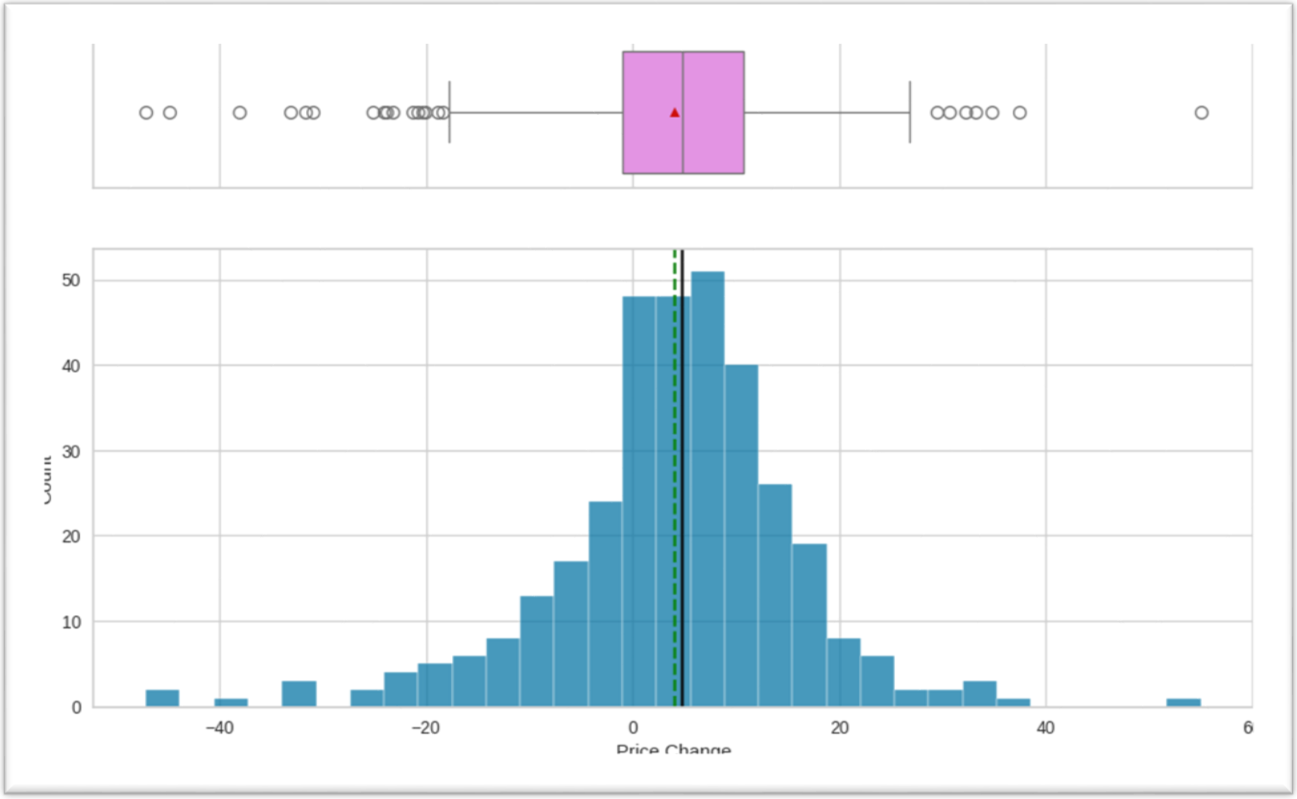


FIGURE 3

➢ Insights on the box plot:

- The median price change is close to zero. The IQR ranges slightly into the positive price change region but is generally symmetric around zero.
- There are outliers on both the sides of the distribution. On negative side, the outliers go as low as –40 and on the positive side they reach approximately 60.

- The whiskers are relatively symmetric suggesting there are extreme values on both the ends, the core data points are distributed fairly evenly around zero.

➢ Insights based on the histogram:

- The histogram shows a roughly symmetric, bell-shaped distribution around a price change of zero. This suggests a near normal distribution with majority of price change clustering around zero.
- The skewness appears to be minimal, with a near equal spread of price changes on both the positive and negative sides.
- The mean price change is close to zero, indicating a balanced distribution with little or no skew.
- The near normal distribution around zero indicates that for most of the items, prices tend to stay relatively stable, with increases and decreases occurring symmetrically.

♦ Box plot & Histogram of ‘Volatility’ distribution:

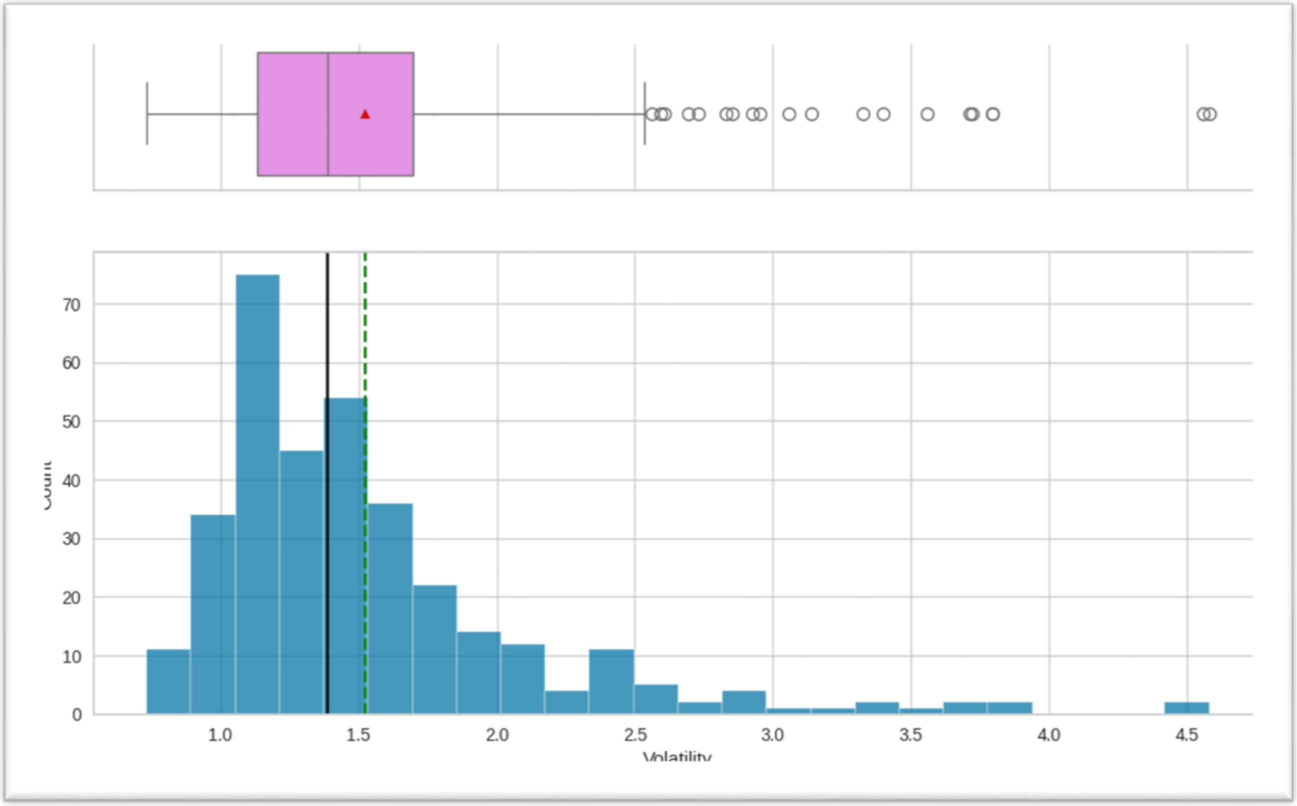


FIGURE 4

➢ Insights on the box plot:

- The box plot suggests that the middle 50% of the data (IQR) is slightly skewed towards lower values of volatility.
- There are numerous outliers present on the higher end. The values range from around 3.0 to 4.5, indicating small subset of highly volatile cases.
- The distribution is right-skewed with a concentration of lower volatility values and a few cases of high volatility.

- **Insights based on the histogram:**

- The histogram reveals that the majority of the data falls between 1.0 and 2.0 in volatility. There is a steep drop-off after 2.0 with a very few data points beyond that range.
- The mode lies between 1.0 and 1.5, with a peak count near 70.
- The histogram also confirms a right-skewed distribution. The bulk of data is in the lower volatility range (1.0 to 2.0) with a long tail extending beyond 2.0
- The mean is higher than the median suggesting the influence of high volatility outliers.
- The plot indicates that the majority of the items are stable with low volatility, but there are a few outliers that shows significant price fluctuations causing a right skewed distribution.

- **Box plot & Histogram of 'Return on Equity (ROE)' distribution:**

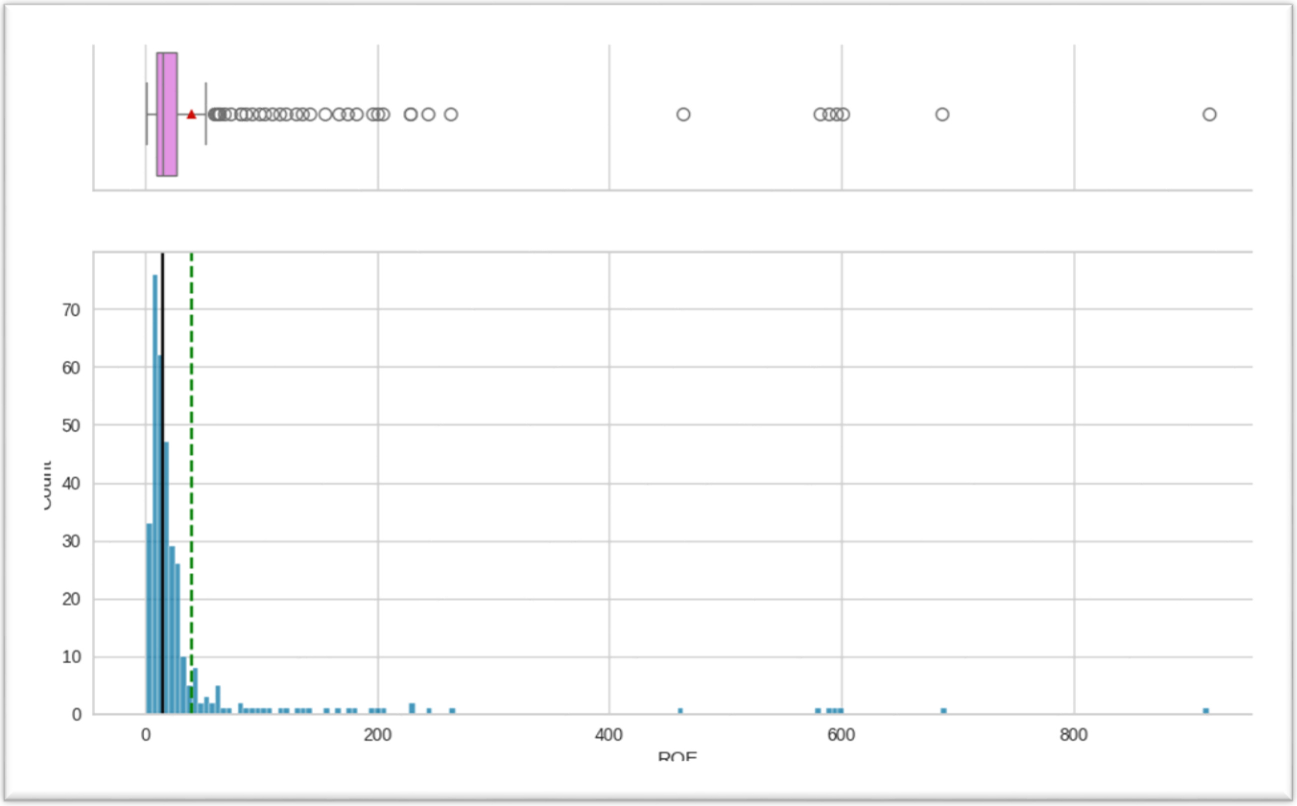


FIGURE 5

➢ Insights on the box plot:

- The majority of the ROE values are tightly clustered near the lower end of the range.
- There are several extreme outliers with values extending far beyond the bulk of the data. These outliers suggests that small number of firms have exceptionally high ROE.
- The mean ROE appears to be higher than the bulk of the data due to the influence of these outliers.

➢ **Insights based on the histogram:**

- The histogram confirms the skewed distribution where most firms have ROE values near zero or slightly positive.
- A long tail in the distribution indicates that few firms have disproportionately high ROE's showing the presence of extreme values.
- The extreme outliers significantly impact the overall statistics like the mean, pulling it away from the more typical ROE values seen in the bulk of data.
- The plot suggests that while most firms may be generating modest returns, a few are performing extra ordinarily well.

♦ **Box plot & Histogram of 'Cash Ratio' distribution:**

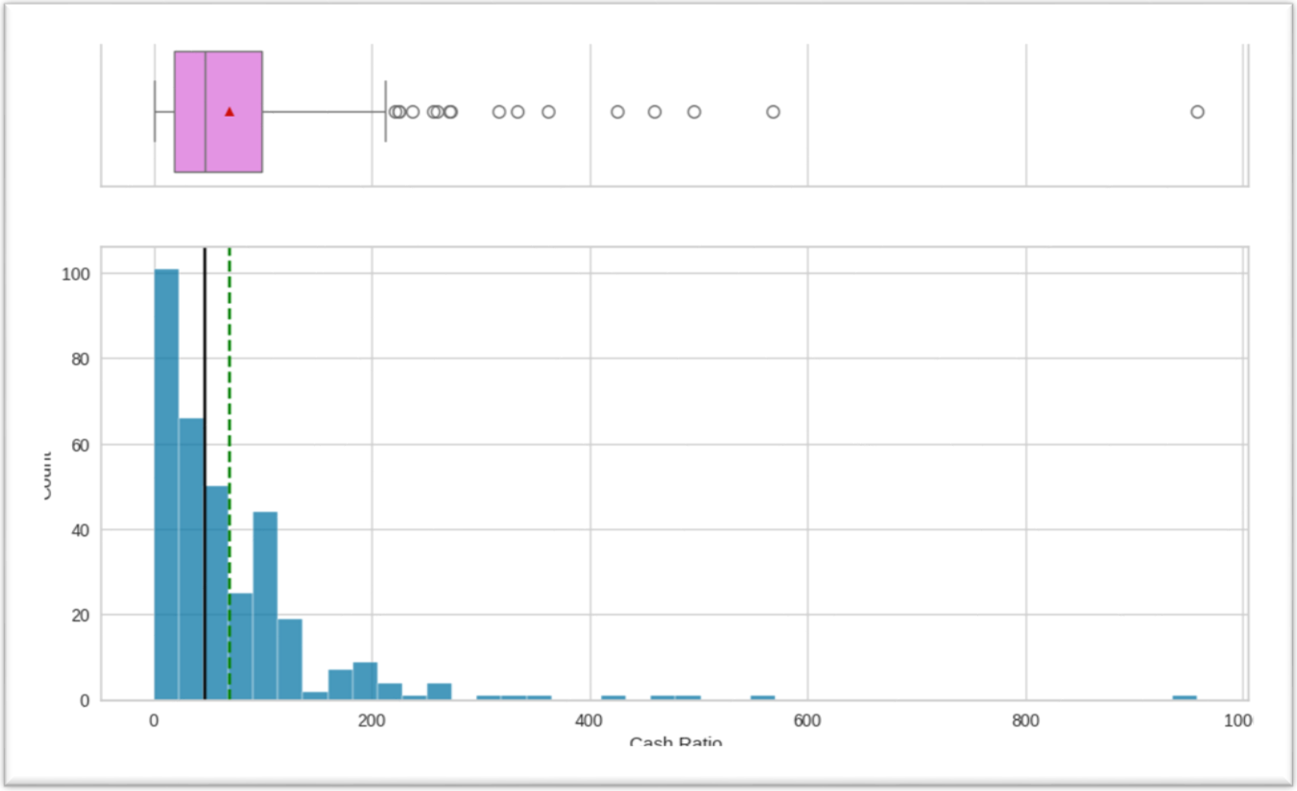


FIGURE 6

➢ Insights on the box plot:

- The majority of the data is concentrated near the lower end of the scale.
- A number of outliers are present, with cash ratio values extending significantly higher than the IQR. These extreme values represent firms that hold disproportionality high amount of cash relative to their liabilities.
- The mean cash ratio is higher than the median due to the presence of the outliers.

➢ Insights based on the histogram:

- The histogram reinforces the right-skewed distribution. The majority of the firms have relatively low cash ratios, indicating that most of them do not hold large amounts of cash in proportion of their liabilities.
- However, there is a small subset of firms with significantly higher cash ratios, pulling up the average.
- The concentration of firms on the left suggests that the mean is skewed by the few high value outliers.
- The presence of high outliers increases the mean, but the median remains relatively low, suggesting that the typical firms have limited liquidity compared to these outliers.

♦ Box plot & Histogram of ‘Net Cash Flow’ distribution:

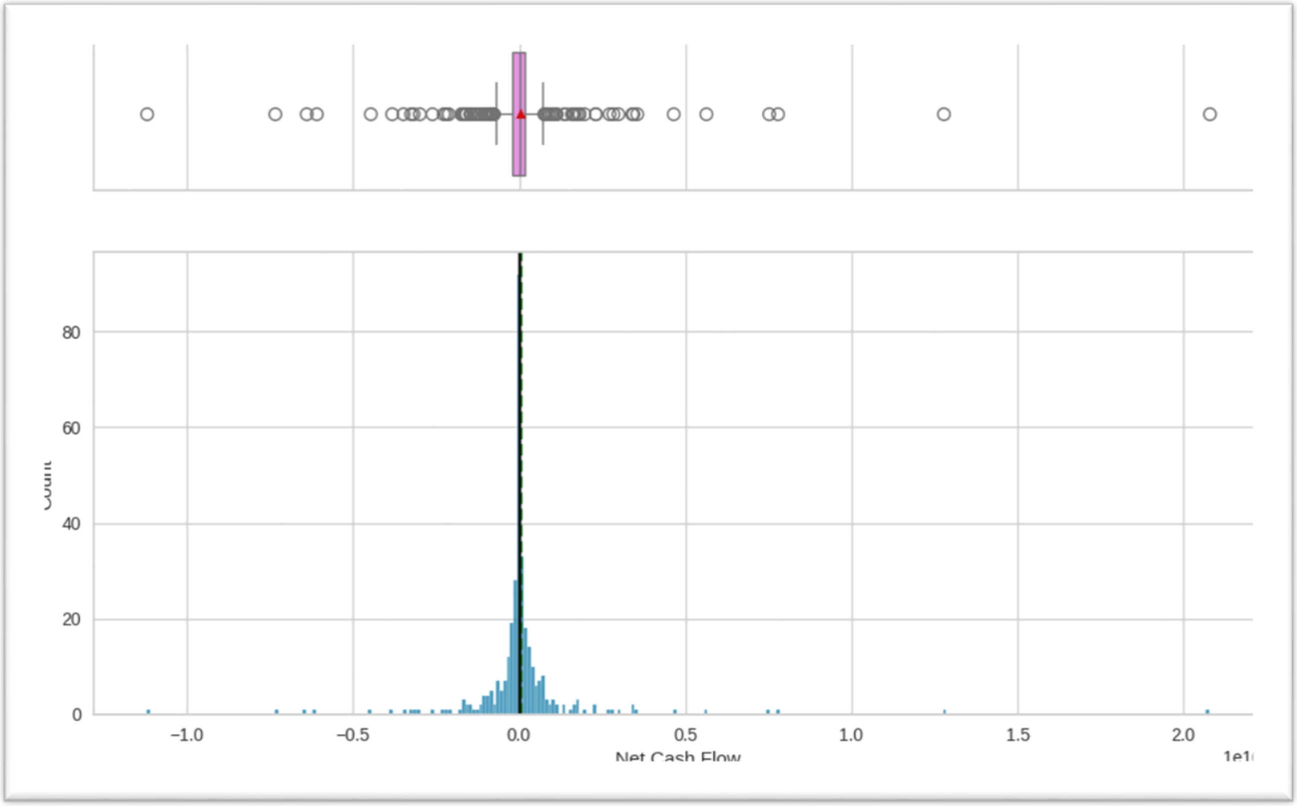


FIGURE 7

➢ Insights on the box plot:

- The distribution of net cash flow is tightly concentrated around the median with a narrow IQR.
- Several outliers are present on both the sides of the distribution showing extreme values of both positive and negative cash flows. These outliers represent companies with either unusually high net inflows or outflows of cash.

- The median appears to be close to the median, suggesting that despite the presence of outliers the overall distribution of net cash flow is relatively symmetric.

➢ Insights based on the histogram:

- The histogram confirms that the majority of the firms have net cash flows values very close to zero, with the bulk of observations tightly packed in the center.
- There are few extreme positive and negative outliers, creating a long tail in both the directions of histogram. These outliers are firms that either experienced significant cash inflows or faced substantial cash outflows during the period measured.
- Despite these outliers, the majority of firms are clustered around zero net cash flow, implying typical companies are not experiencing large cash surpluses or deficits.
- The net cash flow is centered tightly around zero suggesting that most firms have small net changes in their cash position. This could suggest that companies are either balancing inflows and outflows or that their operations are operating near a breakdown point with respect to cash.

♦ Box plot & Histogram of ‘Net Income’ distribution:

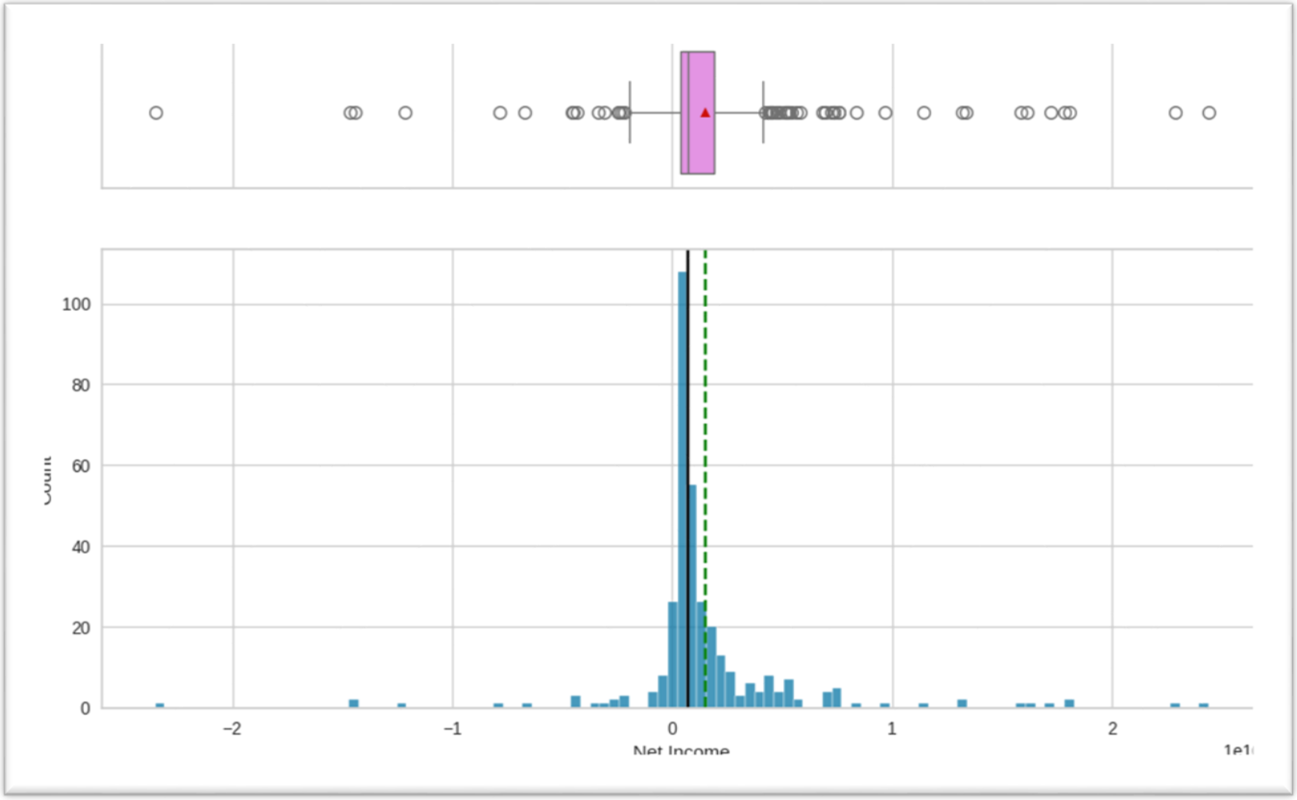


FIGURE 8

➢ Insights on the box plot:

- Most of the net income values are tightly clustered around the median with a small IQR.
- Several outliers on both the positive and negative ends indicate that few companies have either significantly higher or lower net incomes compared to the majority of the firms.

- The mean is slightly skewed to the right, indicating that the outliers on the positive side (companies with large profits) pull the mean away from the median.

➢ Insights based on the histogram:

- The histogram shows a sharp peak around zero, suggesting that most companies have a net income very close to break-even.
- The histogram is right-skewed with a few companies earning much higher profit contributing to the long right tail.
- This indicates that a small number of firms are driving much of the profitability.
- There are few companies with substantial negative net income, contributing to the left tail of the distribution.
- The presence of the outliers, both in positive and negative ranges, suggest that while most companies are operating with marginal profits or losses, there are firms experiencing either strong financial performance or substantial losses.

♦ Box plot & Histogram of 'Earnings Per Share' distribution:

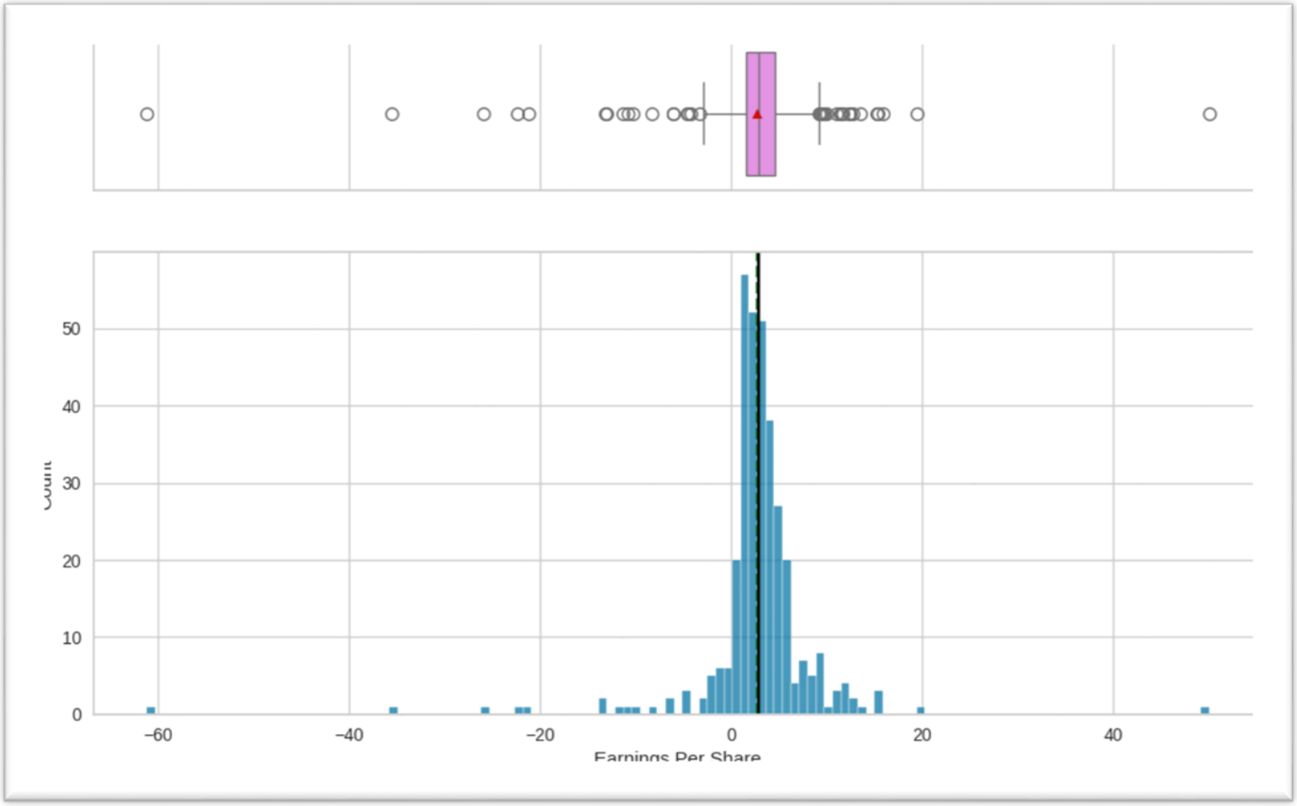


FIGURE 9

➢ Insights on the box plot:

- The IQR is narrow suggesting that most companies have similar Earnings per share (EPS) values.
- The mean is slightly to the right indicating presence of higher positive earnings, even though most of the companies have lower earnings per share values.
- There are numerous outliers present on both the ends with several firms reporting extremely negative or positive earnings per share. These outliers suggests that some

companies are either struggling with losses or are exceptionally profitable.

➢ **Insights based on the histogram:**

- The histogram shows a peak around zero, indicating that the majority of companies have earnings per share values close to break-even.
- The distribution is right-skewed with a long tail extending towards higher positive earnings per share values. This suggests that while most companies are hovering around zero or slightly positive earnings, a small group of firms are achieving significantly higher profits per share.
- The mean and the median are centered near zero confirming that most companies have low earnings per share (EPS) values.
- The plot suggests that while many companies are performing moderately in terms of earnings per share, there are stand out performers that heavily influence the mean, along with a few underperforming firms that drag the distribution towards negative values.

♦ **Box plot & Histogram of 'Estimated Shares Outstanding' distribution:**

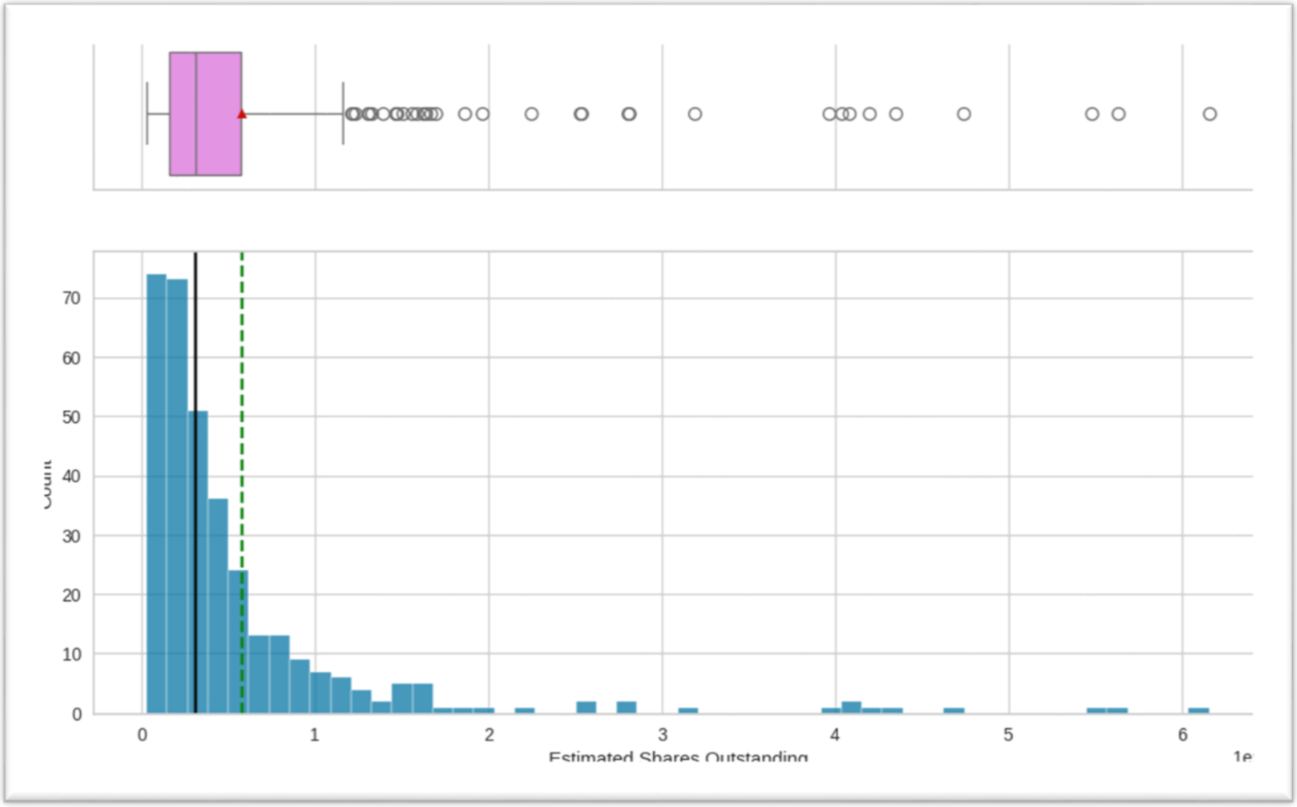


FIGURE 10

➢ Insights on the box plot:

- The box plot suggests that IQR of estimated shares outstanding is relatively small compared to the overall range of the data.
- The mean is near the median indicating a slight skew towards higher values.
- Several firms with outstanding shares exceeding 3 to 6 are considered outliers. This suggests that some large

companies possibly in sectors like tech and finance have issued more shares compared to the majority of others.

➢ Insights based on the histogram:

- The histogram shows a right-skewed distribution with most companies having a low number of shares outstanding, clustered around the lower end of the scale.
- The bulk of companies have shown outstanding in the range of 0 to 1 billion.
- The mean and the median are close to each other which aligns with the slightly skewed distribution but still centers the majority of the data at lower share levels.
- These insights shows that while most companies have moderate number of shares outstanding, a few large firms skew the distribution significantly towards the higher values.

♦ Box plot & Histogram of ‘P/E Ratio’ distribution:

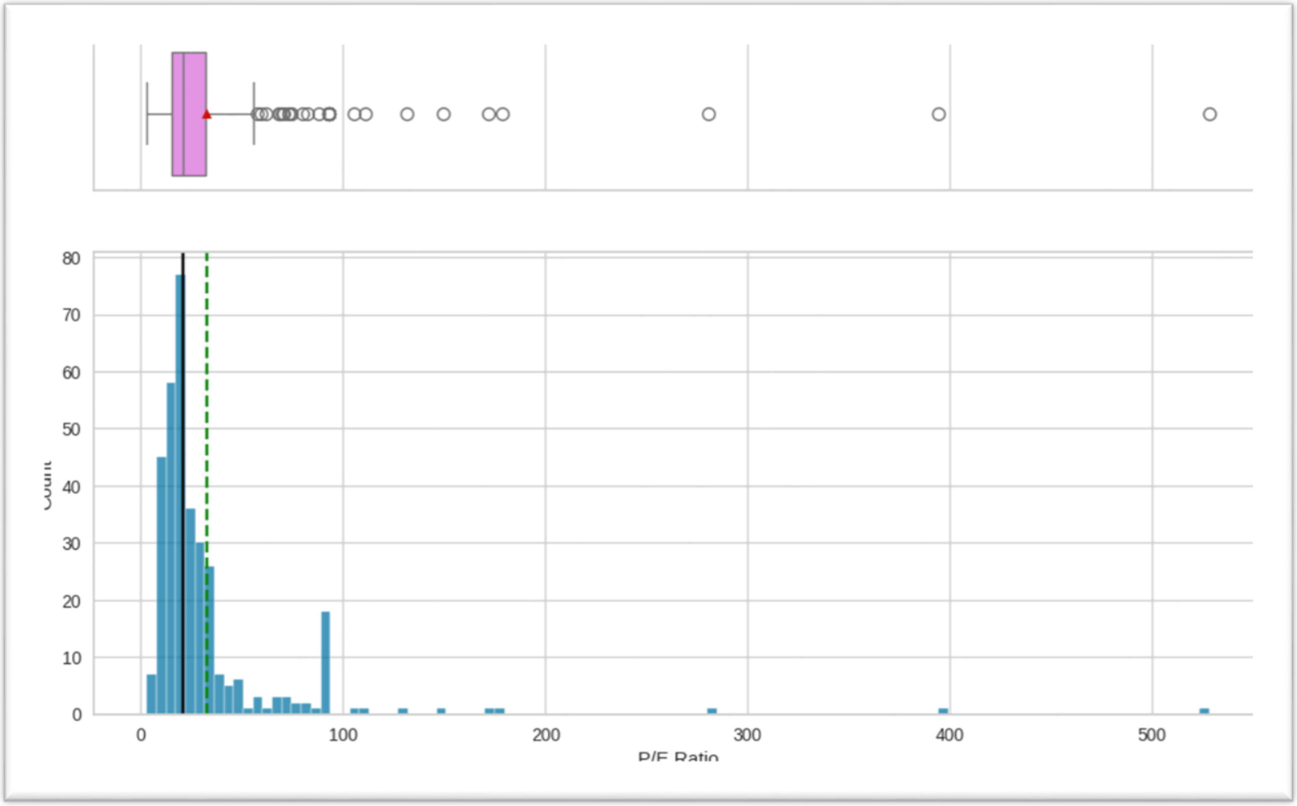


FIGURE 11

➢ Insights on the box plot:

- The P/E Ratio has a long tail with significant outliers.
- The majority of the data is clustered on the lower end.
- The median P/E Ratio is closer to the lower end.
- There are numerous outliers with values stretching up to 500, which is visible in the right side of the plot.

➢ Insights based on the histogram:

- The histogram shows that most of the ‘P/E Ratios’ are concentrated between 0 and 50.
- The distribution is highly right-skewed with a large concentration of values at the lower P/E Ratios and fewer occurrences as the ratio increases.
- There are very few data points with extremely high P/E Ratios (above 100) representing the outliers seen in the box plot.
- The mean appears to be higher than the median, showing positive skewness in the data.
- The distribution is heavily right-skewed, meaning most of the companies have a lower P/E Ratio with some having very high ratios potentially indicating overvaluation.

♦ **Box plot & Histogram of ‘P/B Ratio’ distribution:**

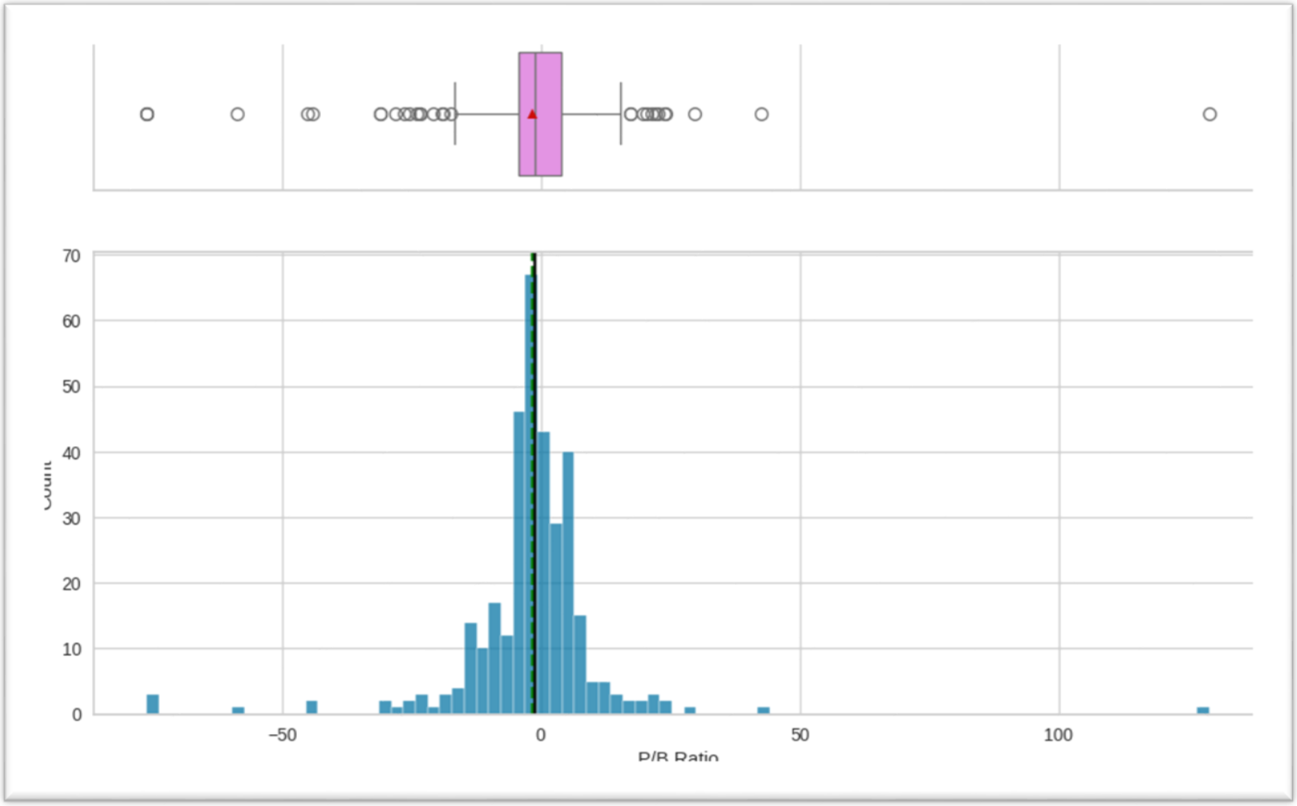


FIGURE 12

➢ Insights on the box plot:

- The IQR is centered around 0, suggesting that most of the stocks have P/B Ratio close to their book value.
- There are number of outliers present, particularly on the far ends.
- The whiskers indicate a wide range of P/B Ratios among the stocks.

- The wide distribution of P/B Ratio implies that different types of companies are included, from value-oriented stocks (low P/B) to growth-oriented ones (high P/B).

- **Insights based on the histogram:**

- The majority of the stocks have a P/B close to 0, as indicated by the central peak.
- There is a sharp spike near P/B ratio= 0, indicating that most stocks are trading near their book value.
- The distribution appears to have long tail on both the sides, showing there are several extreme outliers both positive and negative.
- A significant number of stocks have P/B ratios between –50 and –100, indicating companies whose market values has dropped far below their book value, possibly distressed or highly undervalued.
- On the other hand, few stocks have extremely high P/B ratios (above 50 and reaching 100+) suggesting companies trading at a significant premium to their book value. This represents growth stocks or companies with high investor confidence.
- The visualization suggests that the investment strategies are based on P/B ratios could involve distinguishing between different groups: low P/B vs high P/B.

♦ Bar Plot displaying distribution of stocks across various GICS Sectors:

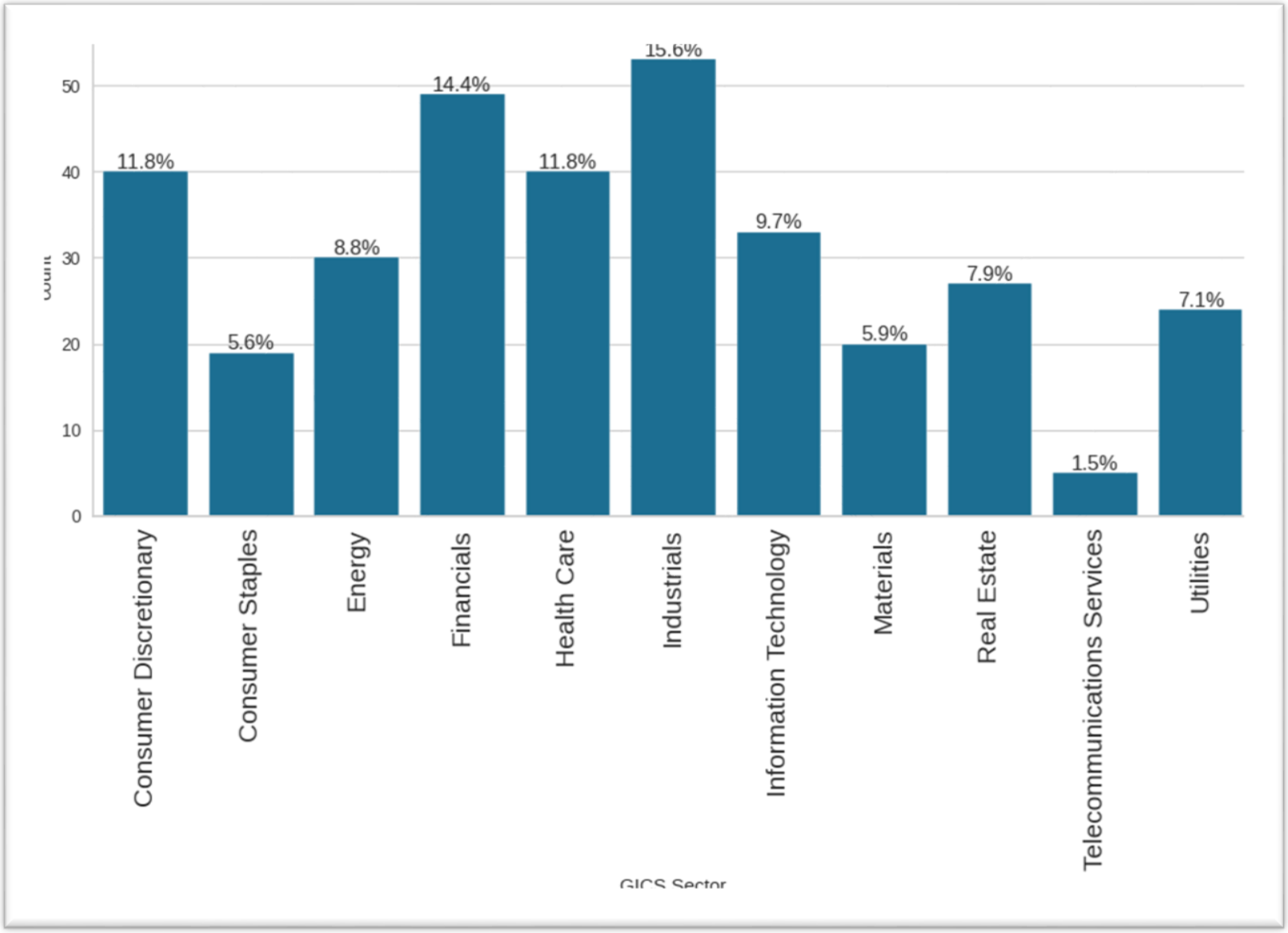


FIGURE 13

➢ Insights on the bar plot:

1. Industrials:

- It has the largest representation among the sectors, indicating a strong presence of companies within industries such as manufacturing, construction, machinery and transportation. This indicates a potential emphasis on industrial growth or investment in capital goods.

2. Financials:

- It is the second largest sector, showing a high concentration of companies related to the banking, insurance and real estate investment trusts (REITs). This reflects the critical role that financial institutions play in the broader economy.

3. Health Care & Consumer Discretionary:

- Both have significant weights.

a) The healthcare sector includes companies in pharmaceuticals, biotechnology and medical devices, indicating strong investor interest in health and life sciences.

b) The Consumer Discretionary sector comprises of companies involved in non-essential goods and services which tend to perform well in periods of economic growth.

4. **Information Technology:**

- It is also a prominent sector, likely representing companies in software, hardware and technology services. This reflects the continuous focus on technological advancement and the sectors growth potential.

5. Energy & Utilities:

- Both are well represented, highlighting companies related to power, gas and oil production. Energy tends to be cyclical, while utilities are considered more stable and defensive providing consistent returns in different market conditions.

6. Real Estate:

- It represents a reasonable portion, indicating a fair presence of companies focused on property development, management or investment. This signifies investor interest in asset backed industries like real estate during times of inflation or economic uncertainty.

7. Material & Consumer Staples:

- Both the sectors have low representation reflecting companies in industries like chemicals, construction, materials and essential goods (e.g., food, beverages and household items). These sectors provide stability during economic downturns but may not be growth focused.

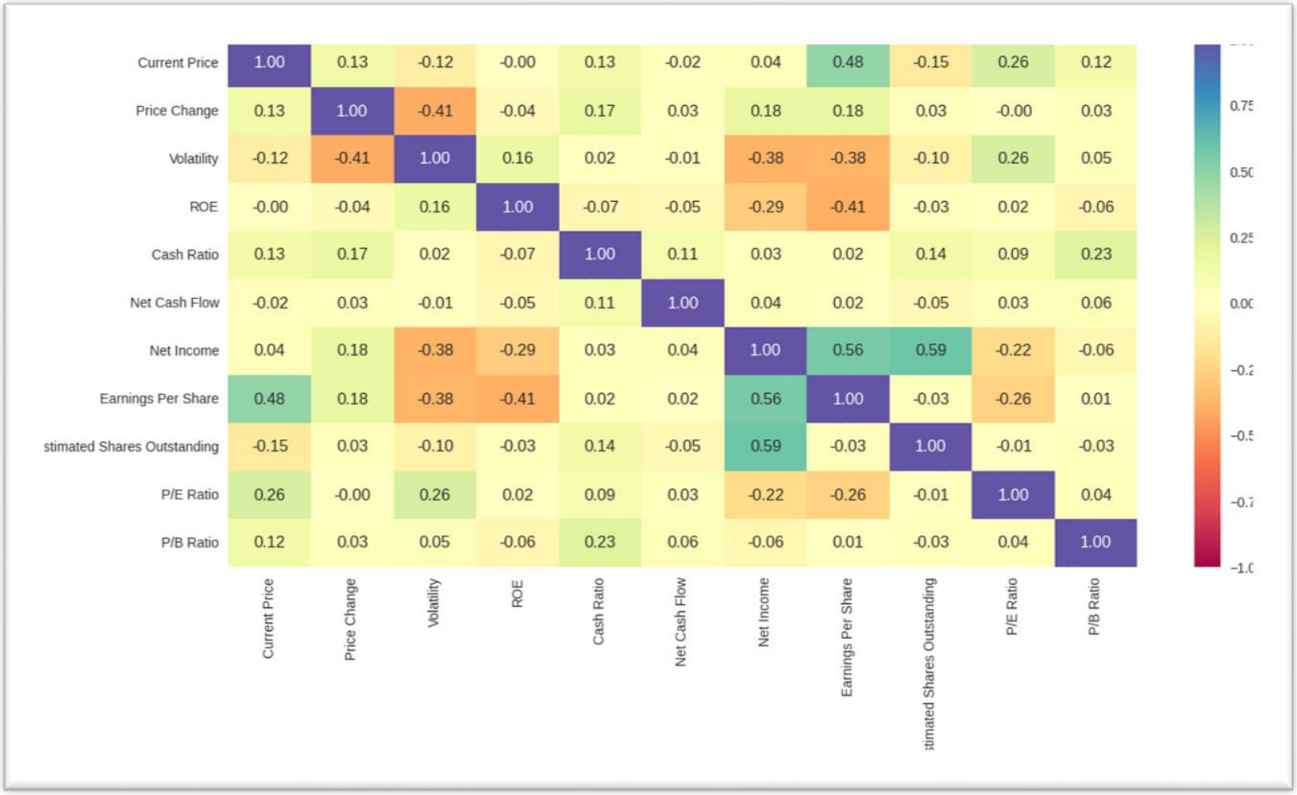
8. Telecommunication Services:

- It has the lowest representation suggesting fewer companies in this sector. Telecommunication services are often seen as defensive but their low share here indicates limited focus or lower volatility compared to other sectors.
- The distribution suggests a balanced mix between growth sectors like technology and discretionary spending and

more stable sectors such as utilities and consumer staples.

❑ BIVARIATE ANALYSIS

♦ Heatmap showing correlation matrix between several financial variables:



	Current Price	Price Change	Volatility	ROE	Cash Ratio	Net Cash Flow	Net Income	Earnings Per Share	stimated Shares Outstanding	P/E Ratio	P/B Ratio
Current Price	1.00	0.13	-0.12	-0.00	0.13	-0.02	0.04	0.48	-0.15	0.26	0.12
Price Change	0.13	1.00	-0.41	-0.04	0.17	0.03	0.18	0.18	0.03	-0.00	0.03
Volatility	-0.12	-0.41	1.00	0.16	0.02	-0.01	-0.38	-0.38	-0.10	0.26	0.05
ROE	-0.00	-0.04	0.16	1.00	-0.07	-0.05	-0.29	-0.41	-0.03	0.02	-0.06
Cash Ratio	0.13	0.17	0.02	-0.07	1.00	0.11	0.03	0.02	0.14	0.09	0.23
Net Cash Flow	-0.02	0.03	-0.01	-0.05	0.11	1.00	0.04	0.02	-0.05	0.03	0.06
Net Income	0.04	0.18	-0.38	-0.29	0.03	0.04	1.00	0.56	0.59	-0.22	-0.06
Earnings Per Share	0.48	0.18	-0.38	-0.41	0.02	0.02	0.56	1.00	-0.03	-0.26	0.01
stimated Shares Outstanding	-0.15	0.03	-0.10	-0.03	0.14	-0.05	0.59	-0.03	1.00	-0.01	-0.03
P/E Ratio	0.26	-0.00	0.26	0.02	0.09	0.03	-0.22	-0.26	-0.01	1.00	0.04
P/B Ratio	0.12	0.03	0.05	-0.06	0.23	0.06	-0.06	0.01	-0.03	0.04	1.00

FIGURE 14

➢ Insights based on the heatmap:

1. **Current Price & Earnings Per Share:**

- There is a moderately positive correlation between current price and EPS. This suggests that companies with higher earnings per share tends to have higher stock prices which is intuitive as earnings are often a driver of stock value.

2. **Volatility:**

- Volatility has a moderate negative correlation with both the net income and EPS. This means that more volatile stocks tend to have lower earnings or more inconsistent earnings, possibly reflecting riskier investments.

3. **Net Income & Earnings Per Share (EPS):**

- There is a strong positive correlation between net income and EPS which makes sense as both are the earnings-related metrics. Higher net income generally leads to higher EPS, as EPS is derived from the net income.

4. **Price/ Earnings (P/E) Ratio & Net Income:**

- Net income has a moderate negative correlation with P/E Ratio. This suggests that the companies with higher net income may have lower P/E ratios, possibly indicating that their stock prices are not overly inflated relative to earnings.

5. **Return On Equity (ROE):**

- Roe shows a negative correlation with net income and Earnings Per Share. This indicates that higher earnings do not necessarily translate into better returns on equity for all the companies, possibly due to high equity bases or inefficient capital allocation.

6. **P/E & P/B Ratios:**

- The P/E & P/B Ratios have a positive correlation though weak. This suggests loose relationship between the market valuation based on earnings and the market valuation based on the book value.

7. **Price Change & Volatility:**

- Price change has a moderate negative correlation with volatility. This indicates that stocks with higher volatility tend to have less favorable or more uncertain price changes, which imply riskier or speculative assets.

8. **Estimated Shares Outstanding (ESO) and Net Income:**

- There is a positive correlation between ESO and net income which suggests that larger companies tend to have higher absolute earnings.
- Understanding these correlations can help in forming groups of stocks based on risk profiles, earning potentials and market valuation which could be useful for personalized investment strategies.

♦ **Bar Plot of Current Price of various stocks:**

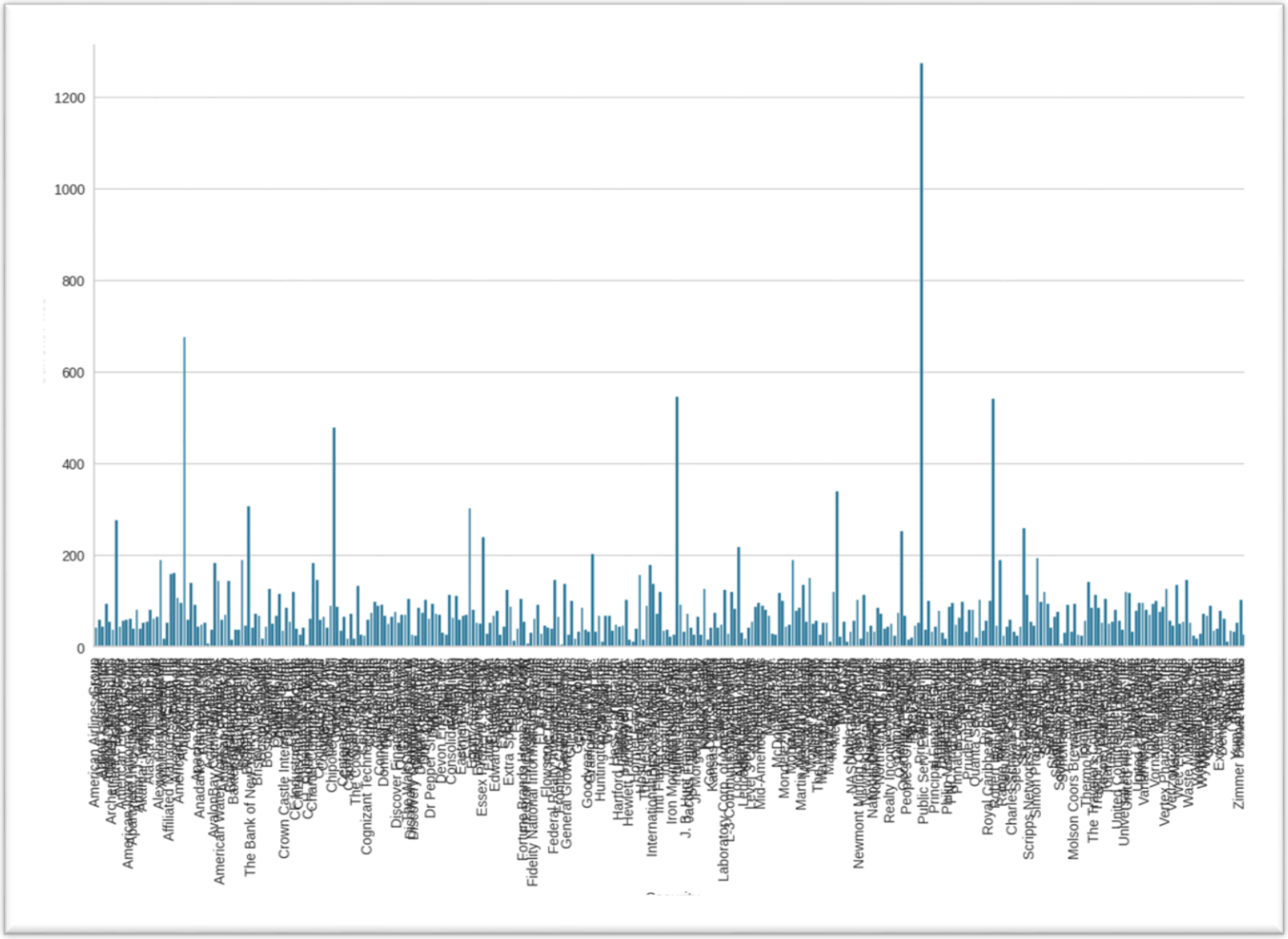


FIGURE 15

➢ Insights based on the Bar Plot:

1. Price Range Variability:

- The majority of the stocks have current prices under $200, indicating that most securities are relatively low priced.
- A few outliers are significantly higher with prices exceeding $600, $800 and even those above $1200. These represents high value stocks or companies in industries like

technology or luxury sectors where stock prices tend to be higher.

2. **Distribution of Stock Prices:**

- There is a significant clustering of stock prices in lower range, suggesting that many companies in the dataset fall within a similar price bracket, representing a specific market segment. (e.g., mid cap or small cap companies).
- The wide gap between the highest and the lowest price stocks suggests that dataset includes companies from different market capitalization, from penny stocks to more mature large cap firms.

3. **Prominent Outliers:**

- A few stocks stand out with very high current prices, well beyond the average range of dataset. These could be major players in their sectors, (like well-known blue-chip stocks) or stocks that have seen massive recent gains.

♦ **Bar Plot of ‘Cash Ratio’ across various GICS Sectors:**

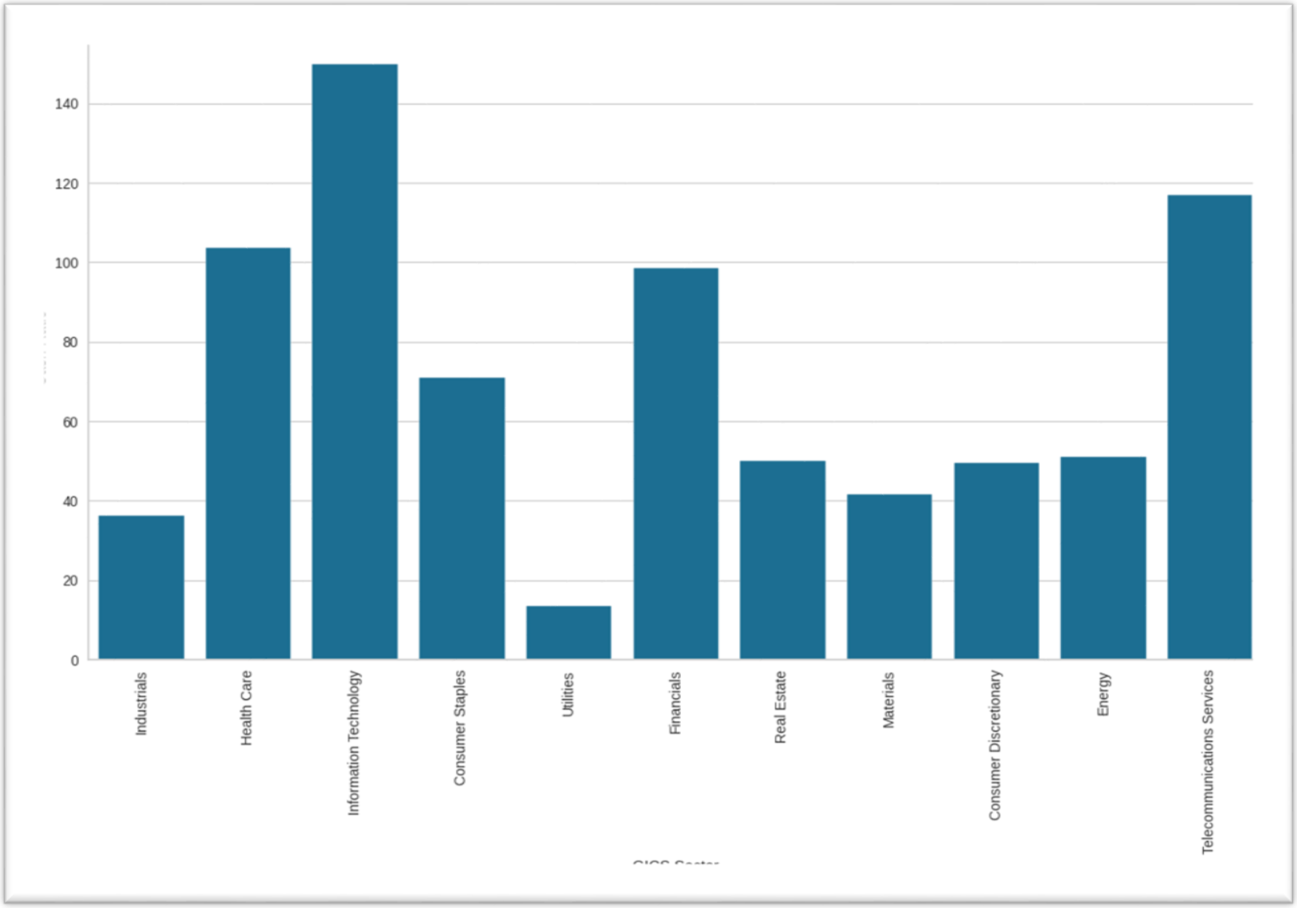


FIGURE 16

➢ Insights based on the Bar Plot:

1. **Information Technology** has the highest cash ratio significantly above 140. This suggests that companies in this sector maintain substantial cash reserves related to their short-term liabilities, possibly to manage risk or capitalize on investment opportunities.
2. **Healthcare & Telecommunication Services** also exhibit relatively higher cash ratios (over 100 and close to 120

respectively) indicating strong liquidity position in these sectors.

3. **Utilities** has the lowest cash ratio, below 20 implying that companies in this sector rely more on other assets or financial mechanisms to cover short term liabilities. This may be due to the capital-intensive nature of utilities, which typically invest in long term infrastructure.
4. **Financials & Consumer Staples** show moderately high cash ratios around 80 – 100 reflecting a cautious approach towards maintaining liquidity.
5. **Industrials, real estates, consumer discretionary, energy & materials** have similar mid-range cash ratios (between 40-60) indicating moderate liquidity in these sectors.
6. This chart represents that these sectors like information technology, healthcare and telecommunication services have a higher focus on maintaining liquid assets, while sectors like utilities, industrial may prioritize other asset types or investment strategies.

- ♦ **Bar Plot representing 'P/E Ratio' for different GICS Sectors:**

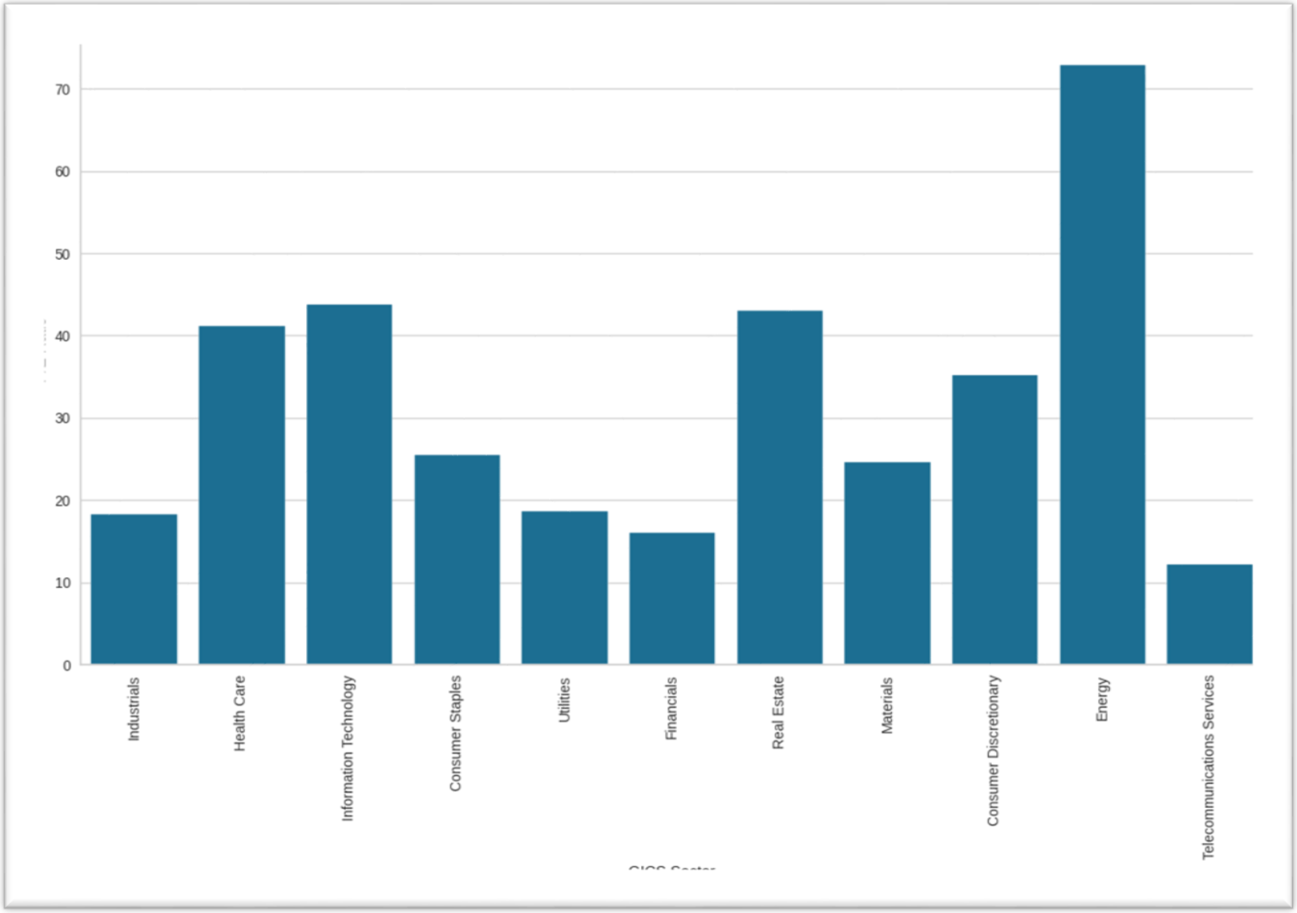


FIGURE 17

➢ Insights based on the Bar Plot:

1. **Energy** stands out with the highest P/E Ratio around 70 indicating strong investor confidence in future earnings growth for this sector. The high ratio also reflects high valuation expectations or an anticipated recovery if the sector is underperforming.
2. **Healthcare and Information Technology** both exhibits high P/E ratios, over 40. This suggests that these sectors

are also viewed as having strong growth potential, possibly due to innovation or demographic trends driving demand.

3. **Real Estate** has a P/E ratio close to 50, which could imply that investors expect the sector to perform well, possibly due to increasing property values or rental income potential.
4. **Consumer Discretionary** has a moderate P/E ratio just above 30, suggesting steady investor confidence in consumer spending and growth prospects.
5. **Utilities and financials** have the lowest P/E ratios both under 20, indicating more modest growth expectations in these sectors. Utilities in particular, tend to have stable but slower growth, while the financial sectors reflect market concerns or steady earnings.
6. **Telecommunication Services** also shows a relatively low P/E ratios which may suggest lower growth prospects or more stable, consistent earnings.
7. **Industrial, Consumer Staples and Materials** fall into the mid-range, with P/E ratios between 20 and 30. These sectors are viewed as having a moderate growth potential balancing between cyclical and defensive characteristics.

Overall, the sectors with the highest P/E ratios- Energy, Healthcare, Information Technology and Real Estate are seen as the high growth or opportunity sectors while utilities and financials represent more stable, lower growth sectors.

- **Bar Plot represents ‘Volatility’ of various GICS Sectors:**

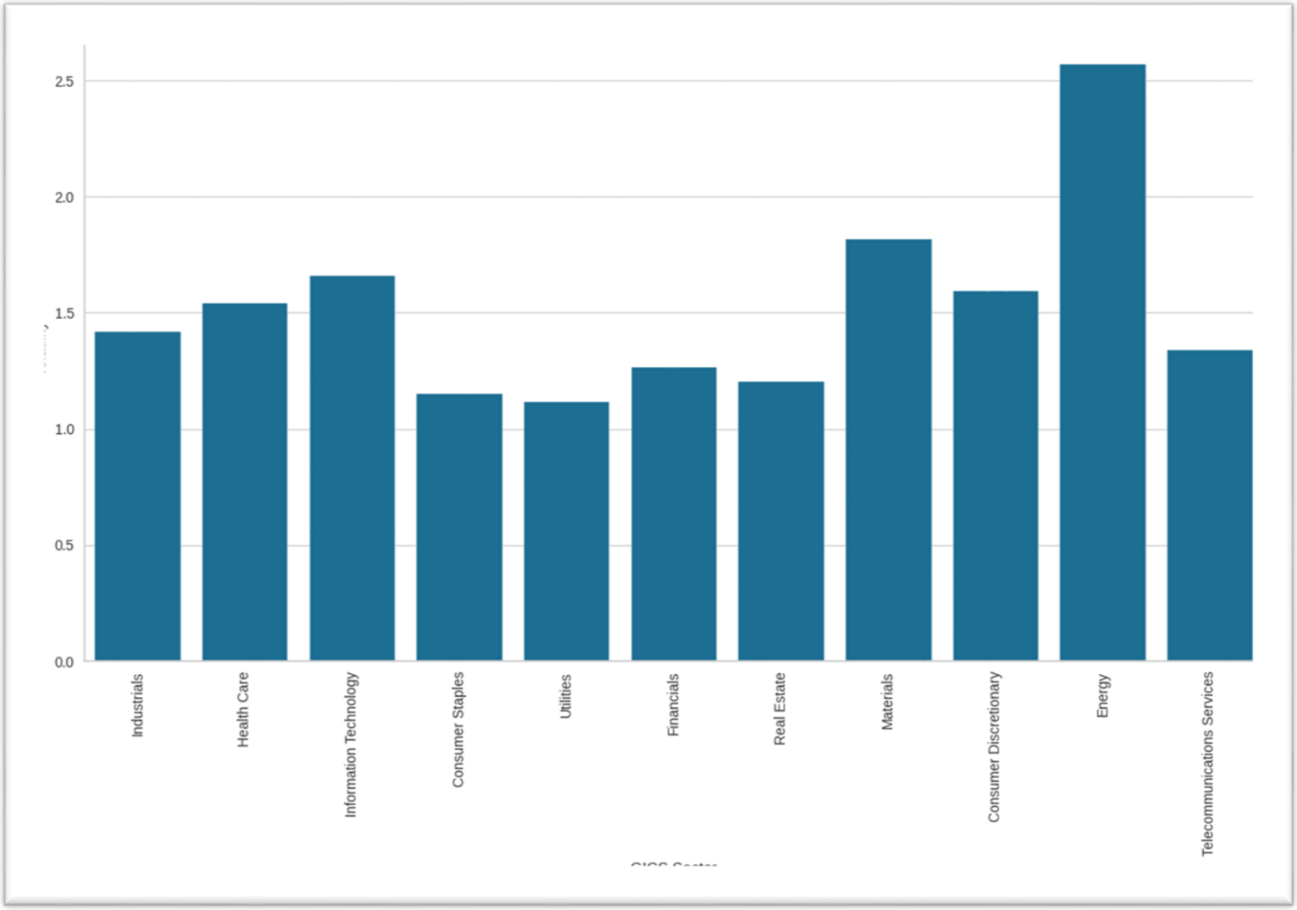


FIGURE 18

➢ Insights based on the Bar Plot:

1. **Energy** has the highest volatility above 2.5. This suggests that stocks in energy sector experience significant price

fluctuations, likely due to external factors like oil prices, geopolitical events and regulatory changes.

2. **Consumer Discretionary & Materials** show relatively high volatility, both around 2.0. These sectors are more sensitive to economic cycles which can cause large swings in their stock prices.
3. **Information Technology, healthcare, financials** exhibit moderate volatility ranging from 1.5 to 1.7. These sectors tend to have strong growth potential but they also face risks related to innovation, regulation, market conditions contributing to their price fluctuations.
4. **Telecommunication Services, consumer staples and utilities** show lower volatility around 1.0. These sectors are typically more stable as they provide essential goods and services, making them less susceptible to dramatic price swings.
5. **Industrial and real estate** falls in the middle with volatility slightly above 1.5. These sectors experience moderate fluctuations, reflecting their mix of growth potential and sensitivity to economic conditions.

Sectors like energy, consumer discretionary and materials have seen as riskier investments due to higher volatility, while telecommunication services, consumer staples and utilities offer more stability with lower volatility.

❑ DATA REPROCESSING

CHECKING OUTLIERS

♦ Box Plot representing various Financial Metrics:

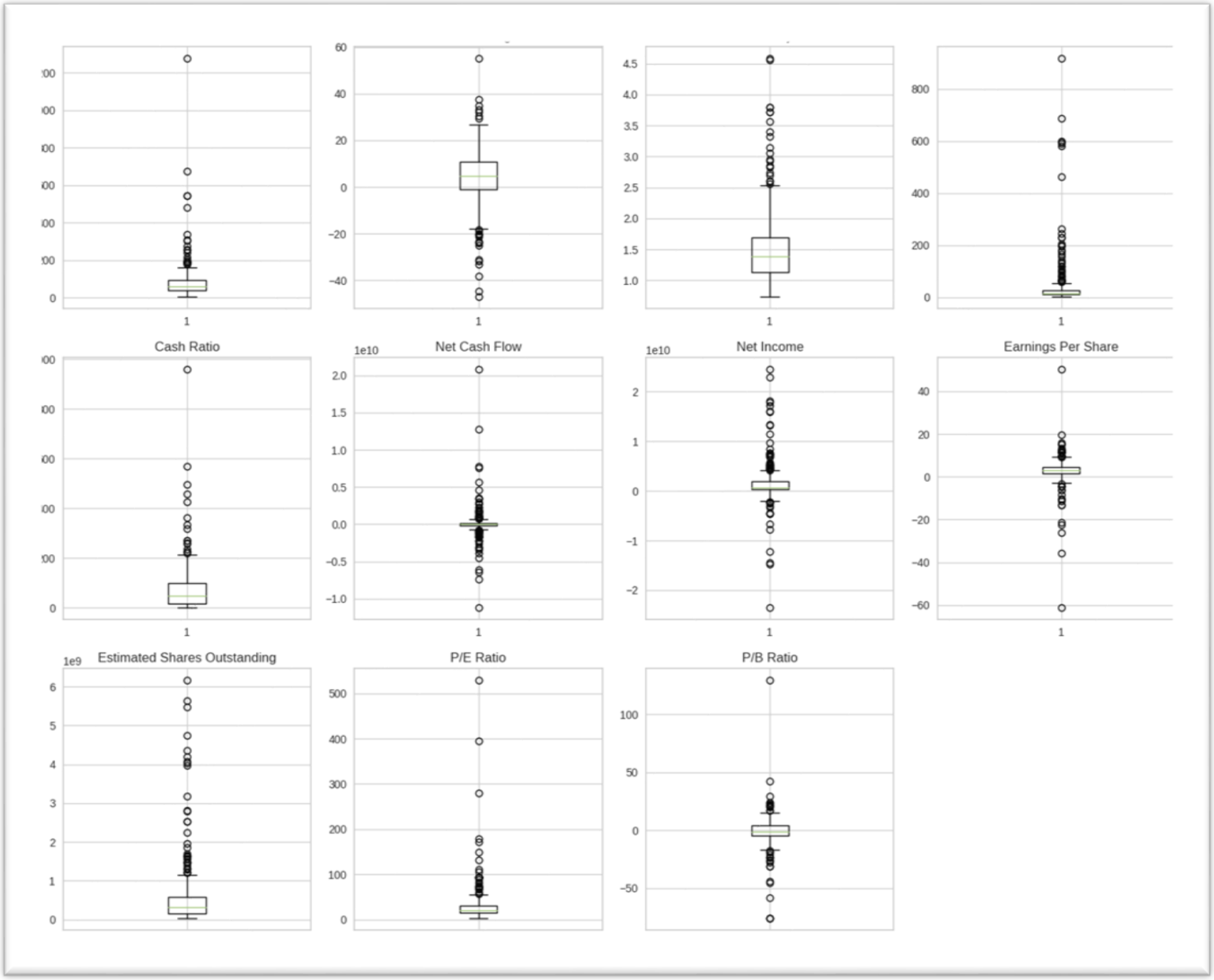


FIGURE 19

➢ Insights based on the Bar Plot:

1. Current Price:

- The median current price is quite low, but there are few extreme outliers with much higher values, indicating a significant variation in stock prices among the companies.

2. Price Change:

- Most of the companies have small price changes, with a concentration near 0. However, a few outliers have both positive and negative large price changes, suggesting some stocks experienced significant movement.

3. Volatility:

- Volatility is generally low, with IQR being close to 1. There are few outliers with high volatility suggesting that most companies have stable prices but some are quite risky.

4. Return On Equity (ROE):

- The median ROE is low, but there are several large outliers both positive and negative indicating that while most companies have moderate returns, some perform exceptionally well or poorly in terms of equity returns.

5. Cash Ratio:

- The median cash ratio is small but there are extreme outliers with very high cash reserves. This implies that while most companies operate with modest liquidity, few have extremely high liquidity positions.

6. Net Cash Flow:

- The median net cash flow is close to 0, with a very narrow IQR. There are both positive and negative outliers, indicating significant cash flow differences among companies.

7. Net Income:

- Net income shows a similar pattern to net cash flows, with most companies clustered around a small median but with a few experiencing very high or low incomes.

8. Earnings Per Share (EPS):

- The majority of the companies have some highly negative or positive EPS values, indicating wide disparity in company profitability.

9. Estimated Shares Outstanding (ESO):

- Most companies have low number of shares outstanding but there are some significant outliers with large amounts of shares, suggesting a range in company sizes or market structures.

10. Price to Earnings (P/E) Ratio:

- Most companies have modest P/E ratio but there are several extreme outliers. This could indicate some companies are highly overvalued or undervalued based on their earnings.

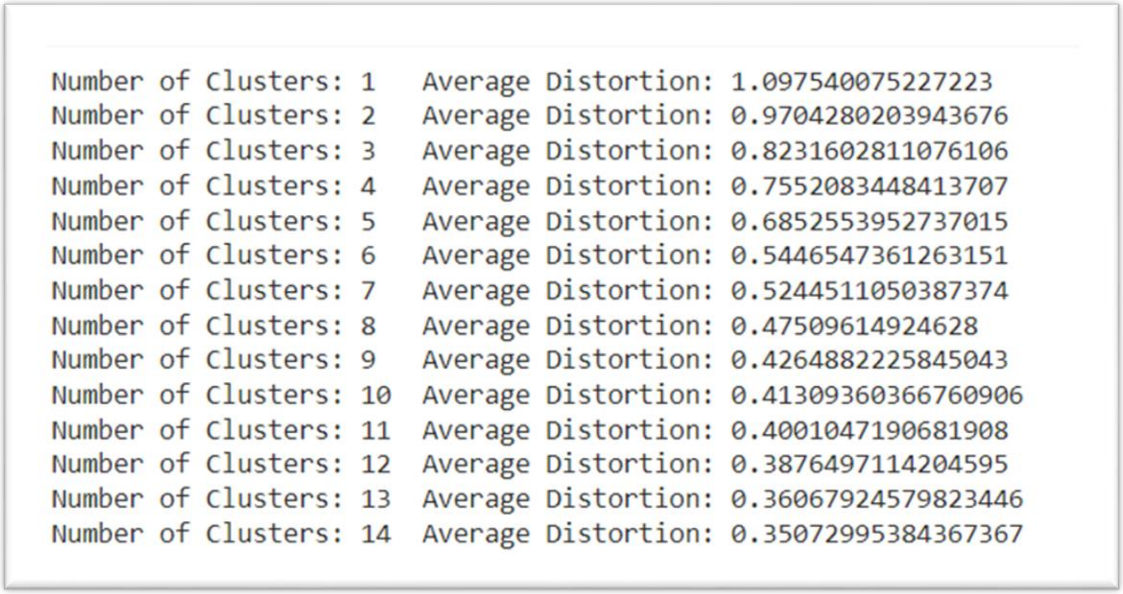
11. Price to Book (P/B) Ratio:

- The P/B ratio shows a similar pattern with a majority of companies clustered around a small value but with both positive and negative outliers, implying variations in the market value relative to the book value.

- ✓ The dataset shows significant presence of outliers across most metrics, indicating substantial variations in financial performance and characteristics across companies.
- ✓ Most metrics are concentrated in a narrow range, with outliers skewing the distribution implying that few companies are influencing the extremes while most remain in a more typical range.
- ✓ The wide variations in the metrics like ROE, P/E ratio and net income suggests significant performance differences across firms, like reflecting industry variations or different financial strategies.

❑ K- MEANS CLUSTERING

♦ CHECKING ELBOW POINT TABLE:



```
Number of Clusters: 1   Average Distortion: 1.097540075227223
Number of Clusters: 2   Average Distortion: 0.9704280203943676
Number of Clusters: 3   Average Distortion: 0.8231602811076106
Number of Clusters: 4   Average Distortion: 0.7552083448413707
Number of Clusters: 5   Average Distortion: 0.6852553952737015
Number of Clusters: 6   Average Distortion: 0.5446547361263151
Number of Clusters: 7   Average Distortion: 0.5244511050387374
Number of Clusters: 8   Average Distortion: 0.47509614924628
Number of Clusters: 9   Average Distortion: 0.4264882225845043
Number of Clusters: 10  Average Distortion: 0.41309360366760906
Number of Clusters: 11  Average Distortion: 0.4001047190681908
Number of Clusters: 12  Average Distortion: 0.3876497114204595
Number of Clusters: 13  Average Distortion: 0.36067924579823446
Number of Clusters: 14  Average Distortion: 0.35072995384367367
```

TABLE 1

- **Plot showing Elbow Method for Selection of Optimal Number of Clusters:**

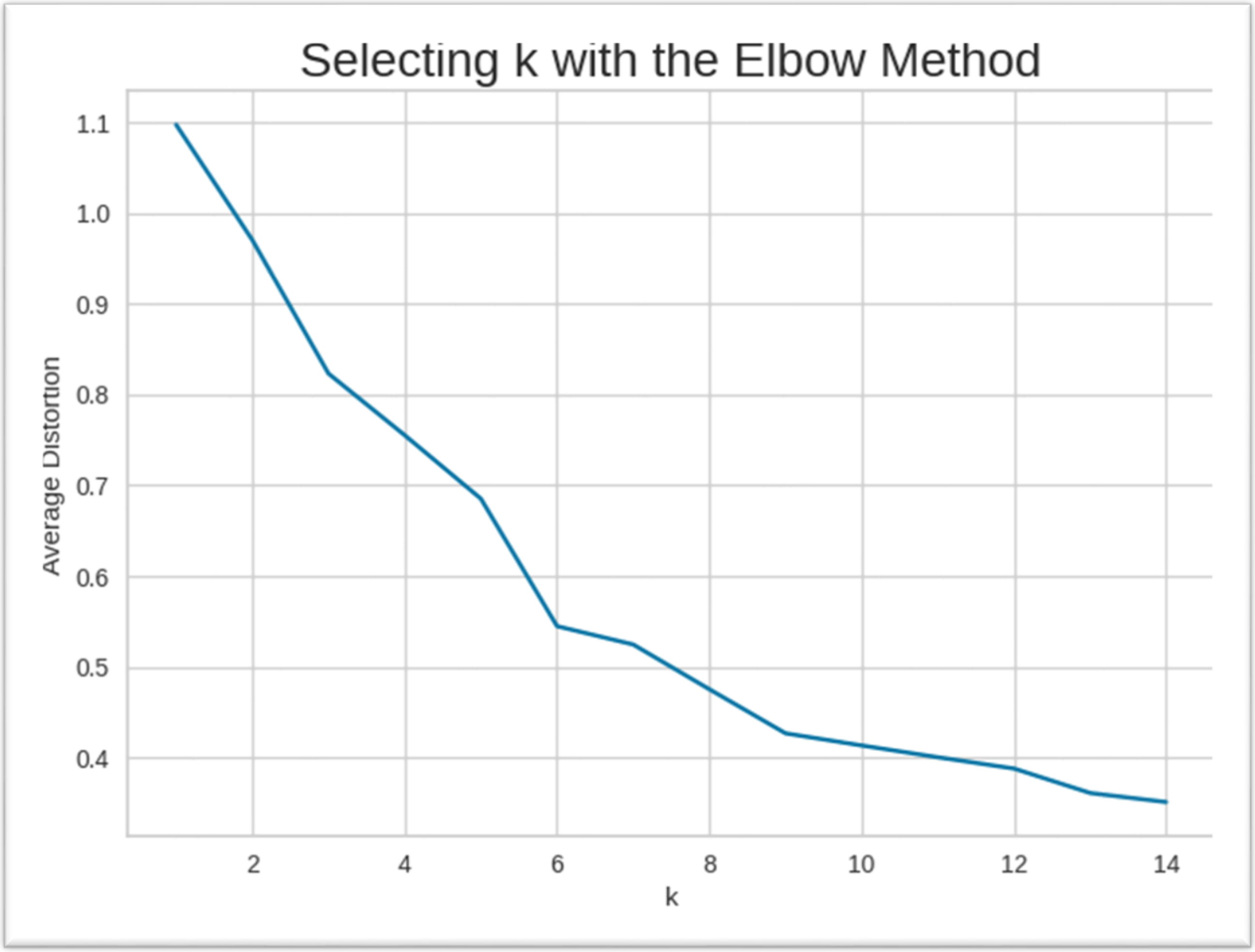


FIGURE 20

➢ Insights based on the Plot:

1. Elbow Point:

- The elbow point is observed around K=6. This is where the rate of reduction in distortion slows down significantly. The optimal number of clusters is often chosen at this point because increasing K beyond this does not provide substantial improvement in clustering performance.

2. Diminishing Returns:

- After K=6, the decrease in distortion becomes much less steep. This indicates that adding more clusters does not significantly improve the model and may lead to overfitting.

3. Interpretation:

- Selecting K=6 appears to balance the complexity (number of clusters) with distortion minimization. Beyond this, the improvement in fitting clusters is marginal.

✓ This plot suggests that the dataset can be well clustered with around 6 clusters as this is where the tradeoff between simplicity and accuracy is optimal.

♦ **Plot representing K- means Distortion Scores and Fit times:**

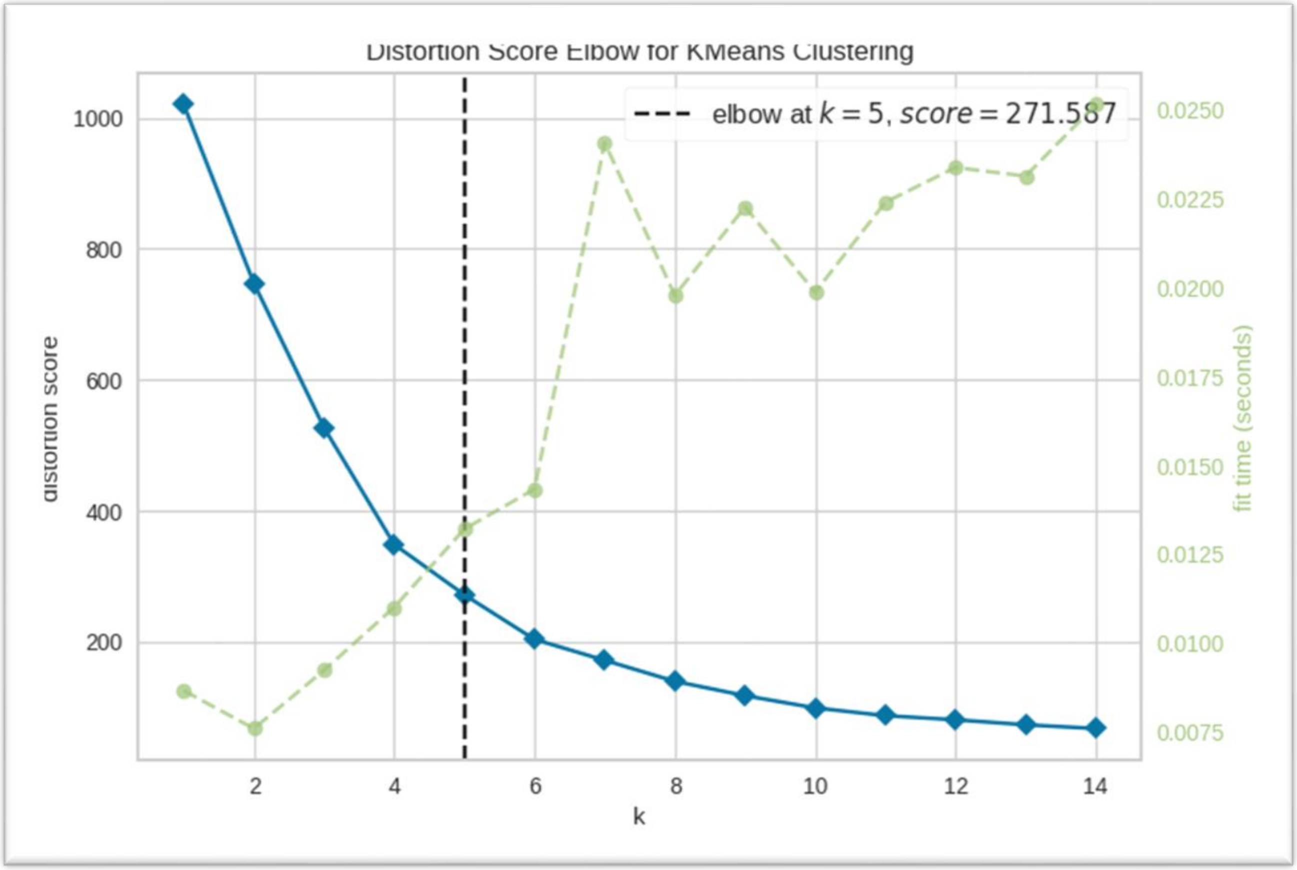


FIGURE 21

➢ Insights based on the Plot:

1. Distortion Score:

- The distortion score decreases as the number of clusters (k) increases, particularly sharply from K=2 to K=6. The curve levels off significantly after K=6 which is consistent with the prior insight that this might be optimal number of clusters. The plot indicates the elbow at K=5, but the distinction between K=5 and K=6 for being the ‘elbow point’ could be a subject for further validation.

2. Fit Time:

- The fit time remains relatively low and stable from K=2 to about K=8 and then begins to increase with larger values of K. This indicates that higher values of K lead to longer computation times, which is expected as the clustering algorithm needs to process more cluster centers.

3. Elbow Point:

- The plot identifies the elbow point at K=5 with a score of 271.57. This suggests that K=5 provides a good trade-off between minimizing distortion and maintaining a reasonable fit time.

4. Selection of K:

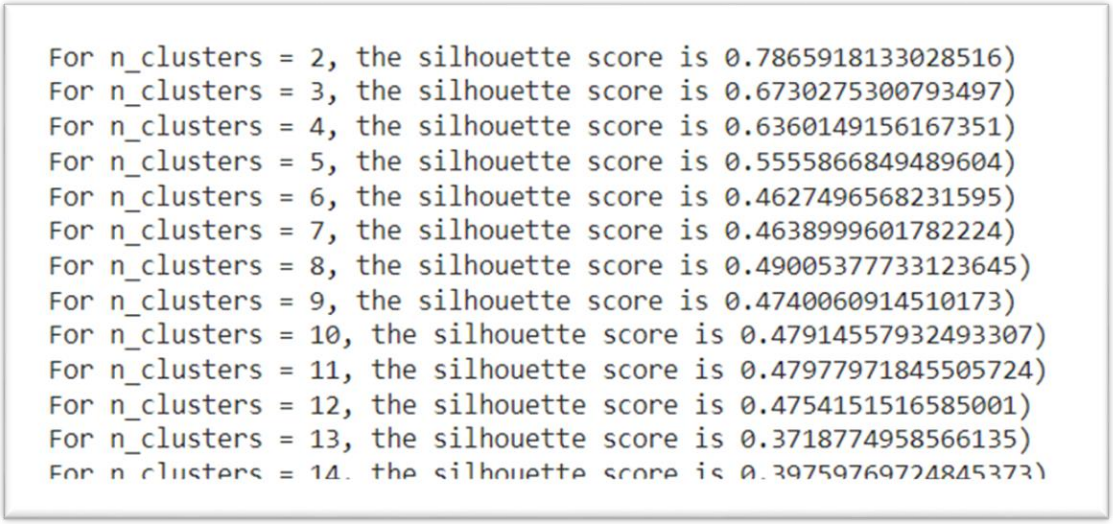
- Although the elbow is identified at K=5, choosing between K=5 and K=6 might depend on specific project needs. If minimizing distortion is highly a priority and a slight increase in computational cost is acceptable, K=6 could be considered. Conversely K=5 could be preferred for slightly faster computations with acceptable distortion.

5. Performance Evaluation:

- The plot also serves as a practical tool for evaluating the performance of the clustering algorithm across different values of K in terms of computational time and effectiveness (via distortion score). Such evaluation is crucial for making informed decisions in cluster analysis, particularly when handling large datasets or requiring efficient processing.

- This method helps in determining the number of clusters that could more efficiently balance computational efficiency and clustering accuracy.

♦ SILHOUETTE SCORES TABLE:



```
For n_clusters = 2, the silhouette score is 0.7865918133028516)
For n_clusters = 3, the silhouette score is 0.6730275300793497)
For n_clusters = 4, the silhouette score is 0.6360149156167351)
For n_clusters = 5, the silhouette score is 0.5555866849489604)
For n_clusters = 6, the silhouette score is 0.4627496568231595)
For n_clusters = 7, the silhouette score is 0.4638999601782224)
For n_clusters = 8, the silhouette score is 0.49005377733123645)
For n_clusters = 9, the silhouette score is 0.4740060914510173)
For n_clusters = 10, the silhouette score is 0.47914557932493307)
For n_clusters = 11, the silhouette score is 0.47977971845505724)
For n_clusters = 12, the silhouette score is 0.475415151658500l)
For n_clusters = 13, the silhouette score is 0.371877495856613 5)
For n_clusters = 14, the silhouette score is 0.39759769724845373)
```

TABLE 2

♦ Plot showing comparison between Silhouette Scores & Fit Times for K means Clustering with different values of K:

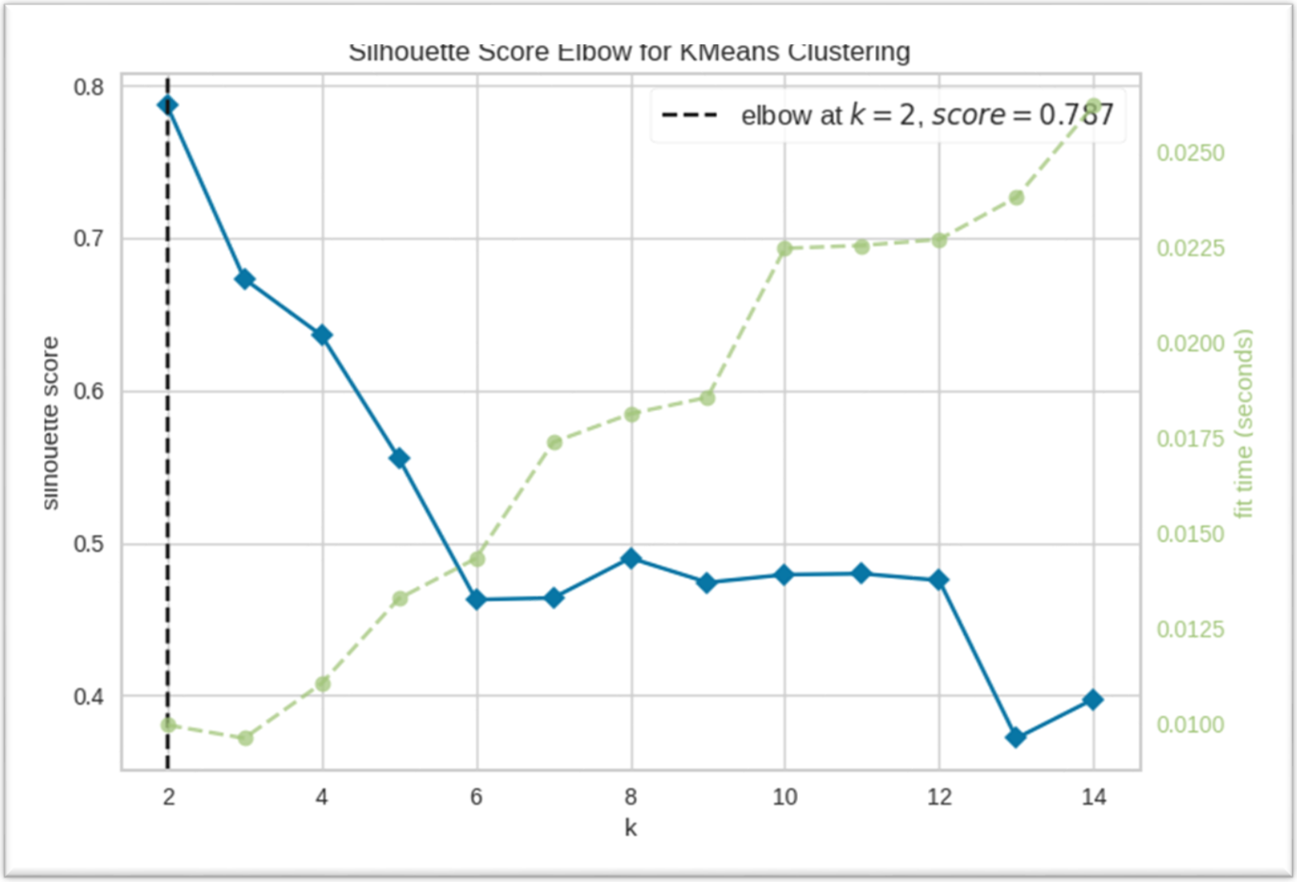


FIGURE 22

➢ Insights based on the Plot:

1. Silhouette Score:

- The silhouette score is a measure of how similar an object is to its own cluster compared to other clusters. It ranges from –1 to 1, with higher values indicating better clustering.
- At K=2, the silhouette score is the highest (around 0.787) suggesting that 2 clusters provide the best separation.

- As K increases, the silhouette score decreases sharply until K=4 and then stabilizes fluctuating between 0.45 and 0.5 indicating that increasing the number of clusters does not improve the clustering quality significantly beyond 2 clusters.

2. Elbow Method:

- The Elbow method is used to identify the optimal number of clusters by looking for a sharp change in the silhouette score. In this case the elbow is clearly at K=2, which is why the plot highlights this value with dashed black line.
- After K=2, the gain in clustering quality (Silhouette scores) diminishes, indicating that2 clusters may be the most appropriate for the data.

3. Fit time:

- The fit time increases gradually as K increases particularly after K=6. This is expected because as the number of clusters increases, the complexity of K-means increases, leading to longer computation times.
- Even though the fit time for K=2 is relatively low, choosing values of K.2 would result in marginal improvements to silhouette score but at the cost of significantly longer fit times.

✓ Optimal number of clusters: Based on the silhouette score and elbow method, K=2 seems to be the best choice.

✓ Trade-off with fit time: Increasing the K results in only minor improvements in clustering quality while significantly increasing the fit time, making K=2 an efficient choice computationally.

♦ **Silhouette Plot showing clustering quality for K-means applied to 340 samples across 3 clusters:**

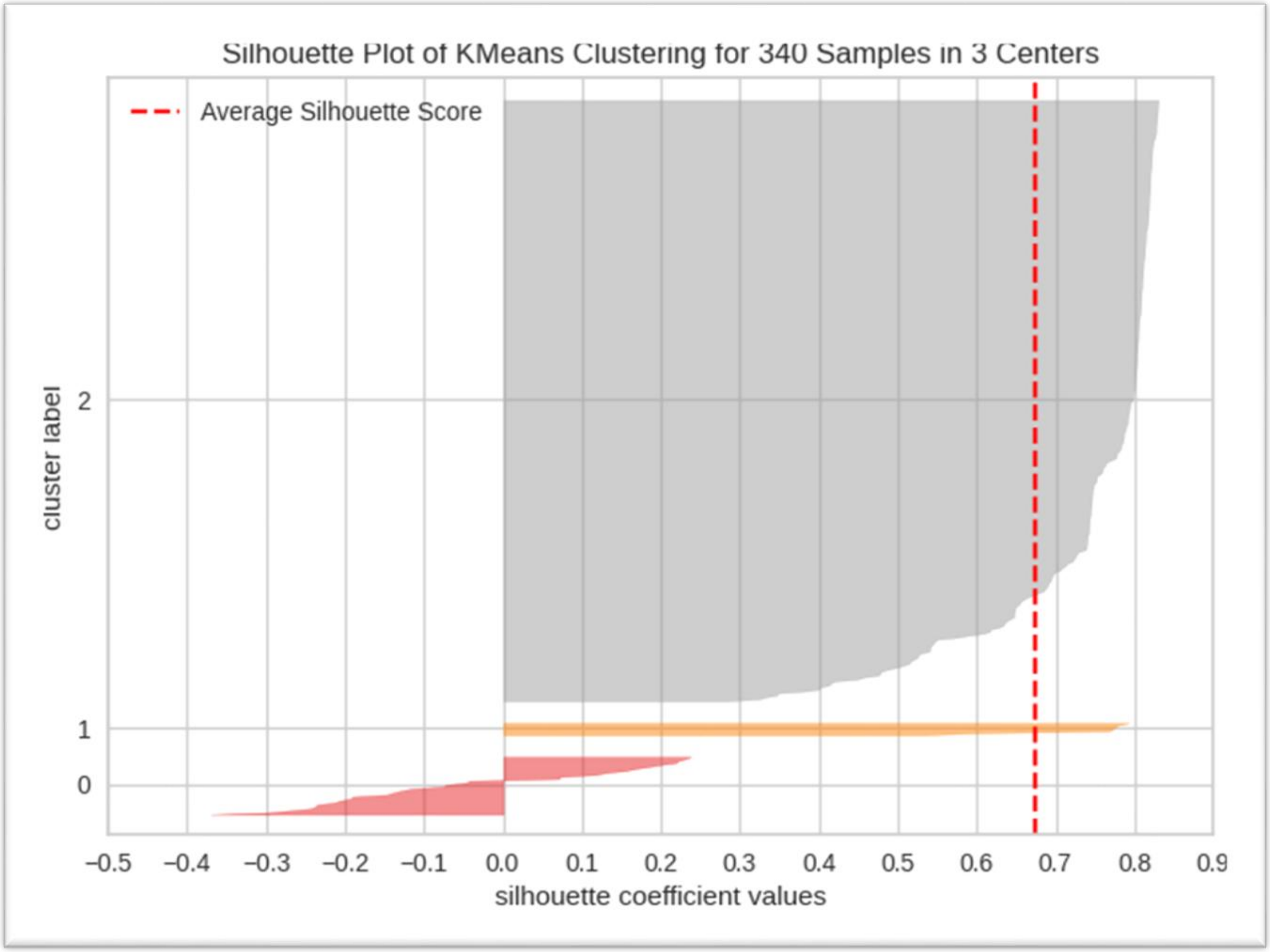


FIGURE 23

➢ Insights based on the Plot:

1. Cluster Sizes:

- The size of the clusters varies significantly. Cluster 2 is the largest, while cluster 0 and cluster 1 are smaller in comparison.

2. Silhouette Coefficient Distribution:

- The Silhouette Coefficient measures how will each sample is clustered. Value close to 1 indicates well-clustered samples, while values close to 0 suggests samples on the cluster boundary and negative values indicate incorrect clustering.
- Cluster 2: The majority of the samples in this cluster have high silhouette score, mostly above 0.6, indicating a well-defined and cohesive cluster.
- Cluster 1: Has moderately positive silhouette score mostly around 0.2 to 0.4. This indicates that the clustering is acceptable, though not as well defined as cluster 2.
- Cluster 0: Contains a significant proportion of samples with negative silhouette score, suggesting that many points in this cluster may be incorrectly assigned or are on the boundary between clusters.

3. Average Silhouette Score:

- The average silhouette score across all the samples is around 0.7

- This high average score suggests that overall clustering is strong, though the issues in cluster 0 reduce the overall clustering quality.

4. **Cluster 0 issues:**

- The negative silhouette values for cluster 0 suggests that some points in this cluster may be better suited to other clusters or that this cluster is not well-separated from others.

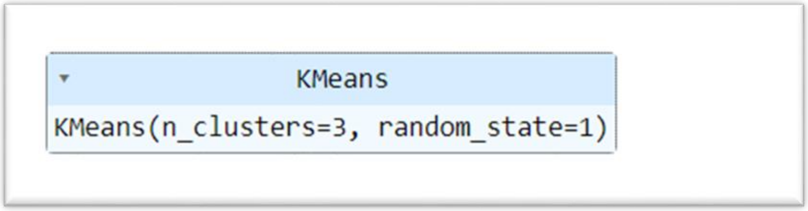
✓ Cluster 2 is well-separated and strongly cohesive, indicating high-quality clustering.

✓ Cluster 1 is reasonably well-clustered but shows some overlap or boundary issues.

✓ Cluster 0 has the most problems, with a number of points poorly assigned, leading to negative silhouette score. This cluster likely needs refinement, possibly by adjusting the number of clusters or revisiting the data's characteristics to better define the separation.

Overall, while the clustering generally performs well (Average score around 0.7), further analysis of cluster 0 could improve the results.

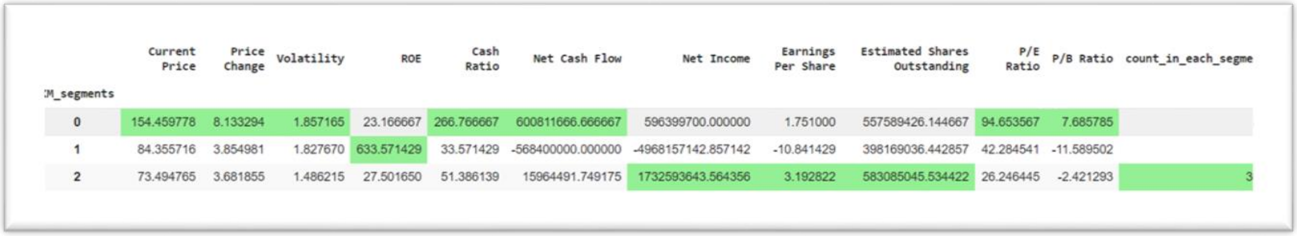
❑ CREATING FINAL MODEL:



```
▾ KMeans
KMeans(n_clusters=3, random_state=1)
```

❑ CLUSTER PROFILING

- **Table represents stock grouping from K-means segmentation based on various metrics:**



	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_in_each_segme
:M_segments												
0	154.459778	8.133294	1.857165	23.166667	266.766667	600811666.666667	596399700.000000	1.751000	557589426.144667	94.653567	7.685785	
1	84.355716	3.854981	1.827670	633.571429	33.571429	-568400000.000000	-4968157142.857142	-10.841429	398169036.442857	42.284541	-11.589502	
2	73.494765	3.681855	1.486215	27.501650	51.386139	15964491.749175	1732593643.564356	3.192822	583085045.534422	26.246445	-2.421293	3

TABLE 3

➢ Insights based on the Table:

1. Segment 0:

- Current Price: The highest at 154.46, indicating that stocks in this group tend to have higher market valuations.

- Price Change: 8.13% is the highest among the segments, reflecting strong recent price growth.
- Volatility: 1.86, suggesting moderate risk, compared to other segments.
- Return On Equity (ROE): 23.17% indicating healthy profitability but not the highest.
- Cash Ratio: 266.77, extremely high liquidity in this group, suggesting these companies have strong cash reserves.
- Net Cash Flow: $600.8 million showing positive operational efficiency.
- Net Income: $596.4 million, indicating profitability.
- Earnings Per Share (EPS): 1.75, representing moderate returns for the shareholders.
- Estimated Shares Outstanding: 557.59 million, quite large but not the highest.
- P/E Ratio: 94.65, very high indicates the stocks are overvalued.
- P/B Ratio: 7.69, the highest suggesting these stocks trade at a high premium over book value.
- Count: 30 stocks, a smaller group but composed of financially stable companies.

2. Segment 1:

- Current Price: 84.36, indicating mid-range valuations.
- Price Change: 3.85% showing moderate growth.

- Volatility: 1.83, similar to other segments.
- Return On Equity (ROE): 633.57%, a very high ROE, suggesting exceptional profitability. However, this may be an outlier.
- Cash Ratio: 33.57, significantly lower than segment 0 but still indicative of decent liquidity.
- Net Cash Flow: -$568.4 million, showing significant outflows which might be a concern.
- Net Income: -$496.8 million, showing negative profitability.
- Earnings Per Share (EPS): -10.84, reflecting poor performance.
- Estimated Shares Outstanding: 398.17 million, the smallest in the table.
- P/E Ratio: 42.28, indicating these stocks are relatively expensive despite negative earnings.
- P/B Ratio: -11.59, reflecting a high discrepancy between market price and book
- Count: Only 7 stocks, suggesting this group could be distressed or speculative.

3. Segment 2:

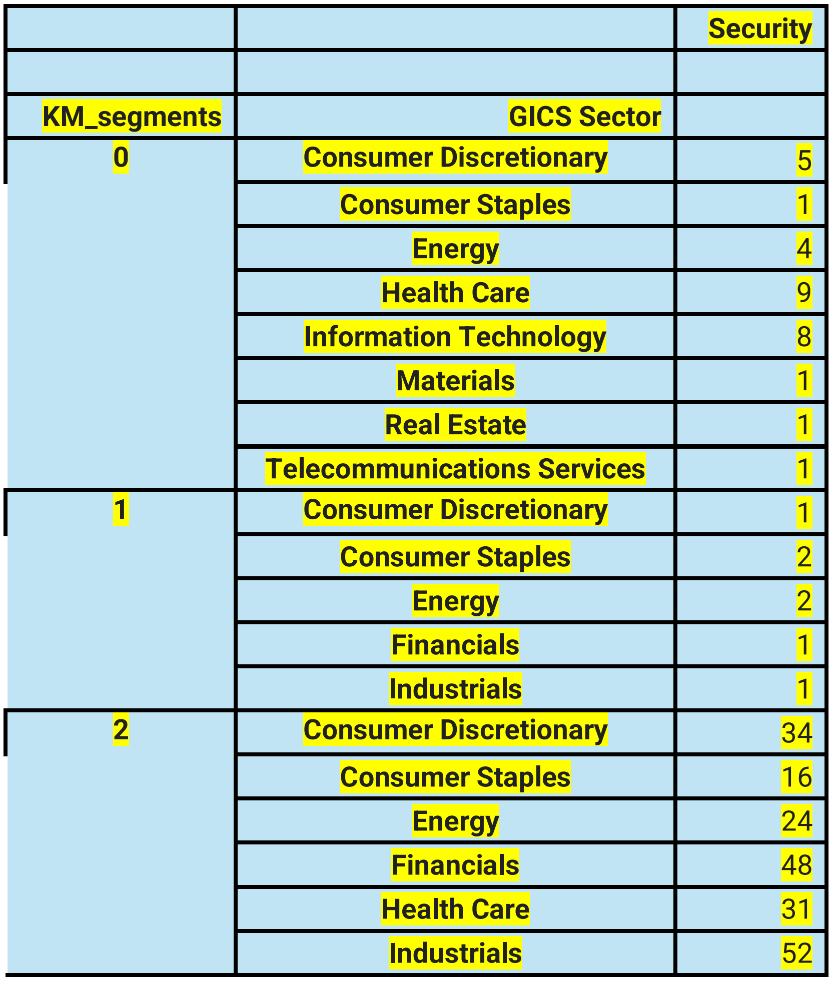
- Current Price: 73.49, the lowest indicating these stocks are more affordable.
- Price Change: 3.68% similar to segment 1, reflecting modest growth.

- Volatility: 1.49, the lowest in the table, indicating that these stocks are less risky.
- Return On Equity (ROE): 27.50% reflecting good profitability though much lower than the segment 1.
- Cash Ratio: 51.39, better liquidity than segment 1, but much lower than segment 0.
- Net Cash Flow: $159.64 million, indicating positive cash flow, though lower than segment 0.
- Net Income: $1.73 billion, the highest showing strong earnings despite lower stock prices.
- Earnings Per Share (EPS): 3.19, the highest EPS showing good returns to the shareholders.
- Estimated Shares Outstanding: 583.09 million, the largest in the table.
- P/E Ratio: 26.25 reasonable and indicating moderate valuations.
- P/B Ratio: -2.42, suggesting these stocks are trading below thin book value, which could mean they are undervalued.
- Count: 303 stocks, the largest group, indicating these are more common or stable stocks.

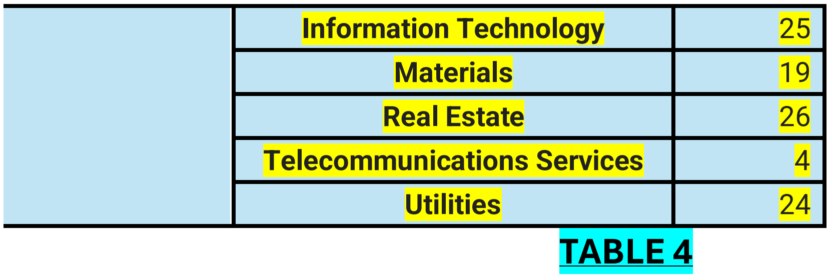
✓ Segment 0 consists of higher priced, high growth and highly liquid companies with high P/E and P/B ratios suggesting they may be overvalued but stable.

- ✓ Segment 1 appears to be highly speculative with negative earnings and cash flows but high ROE. It is a small group, likely composed of distressed companies.
- ✓ Segment 2 represents a large group of lower priced, lower volatility stocks that are profitable with solid cash flows, making them potentially undervalued investments.

- ♦ **Table showing distribution of stock segments based on the GICS sectors classifying each KM_segment:**



		Security
KM_segments	GICS Sector	
0	Consumer Discretionary	5
	Consumer Staples	1
	Energy	4
	Health Care	9
	Information Technology	8
	Materials	1
	Real Estate	1
	Telecommunications Services	1
1	Consumer Discretionary	1
	Consumer Staples	2
	Energy	2
	Financials	1
	Industrials	1
2	Consumer Discretionary	34
	Consumer Staples	16
	Energy	24
	Financials	48
	Health Care	31
	Industrials	52



	Information Technology	25
	Materials	19
	Real Estate	26
	Telecommunications Services	4
	Utilities	24

TABLE 4

➢ Insights based on the Table:

1. Segment 0:

- Sector Concentration: This group is relatively diverse with stocks spread across various sectors, but the largest concentration is in the healthcare (9 stocks) and information technology (8 stocks).
- Sector Highlights:

a) Healthcare and Information Technology dominates, indicating these sectors represent a significant portion of the higher priced, highly liquid stocks in segment 0.

b) There is a notable presence in energy and consumer discretionary which suggests that some growth-oriented sectors are also represented.

c) Sectors like consumer staples, materials, real estate and telecommunication services has minimal representation, suggesting that these sectors are less common in this high value, high liquidity group.

2. Segment 1:

- Small and specialized: This segment is small and focused on only a few sectors, reflecting its speculative and risky nature.

a) Energy and consumer staples have the highest presence indicating that these distressed or high-risk stocks come from sectors related to commodities or essentials.

b) Other sectors include consumer discretionary, financials and industrials but each with only 1 stock showing narrow sectoral diversity in this speculative group.

3. Segment 3:

- Broadly distributed: Segment 2 is the largest and most diverse with representation across all the sectors, indicating that this group contains the most common, stable stocks.
- Sector Highlights:

a) Industrials and financials are heavily represented showing that these sectors are key contributors to the low volatility, mid- priced stocks in this group. These sectors typically feature established companies with consistent revenues.

b) Healthcare and energy also have significant representation, reflecting the importance of these sectors in stable investments.

c) Consumer discretionary and information technology shows that these are still growth oriented and tech driven stocks in this lower risk group, despite generally lower valuations.

d) Utilities and real estate have strong representation further suggesting that this segment is oriented towards stability and value, as these sectors are typically associated with lower volatility and steady returns.

e) Materials and telecommunication services are smaller contributors but still present, providing a well-rounded sector distribution.

- **Box plot showing distribution of various financial metrics for each KM_segments:**

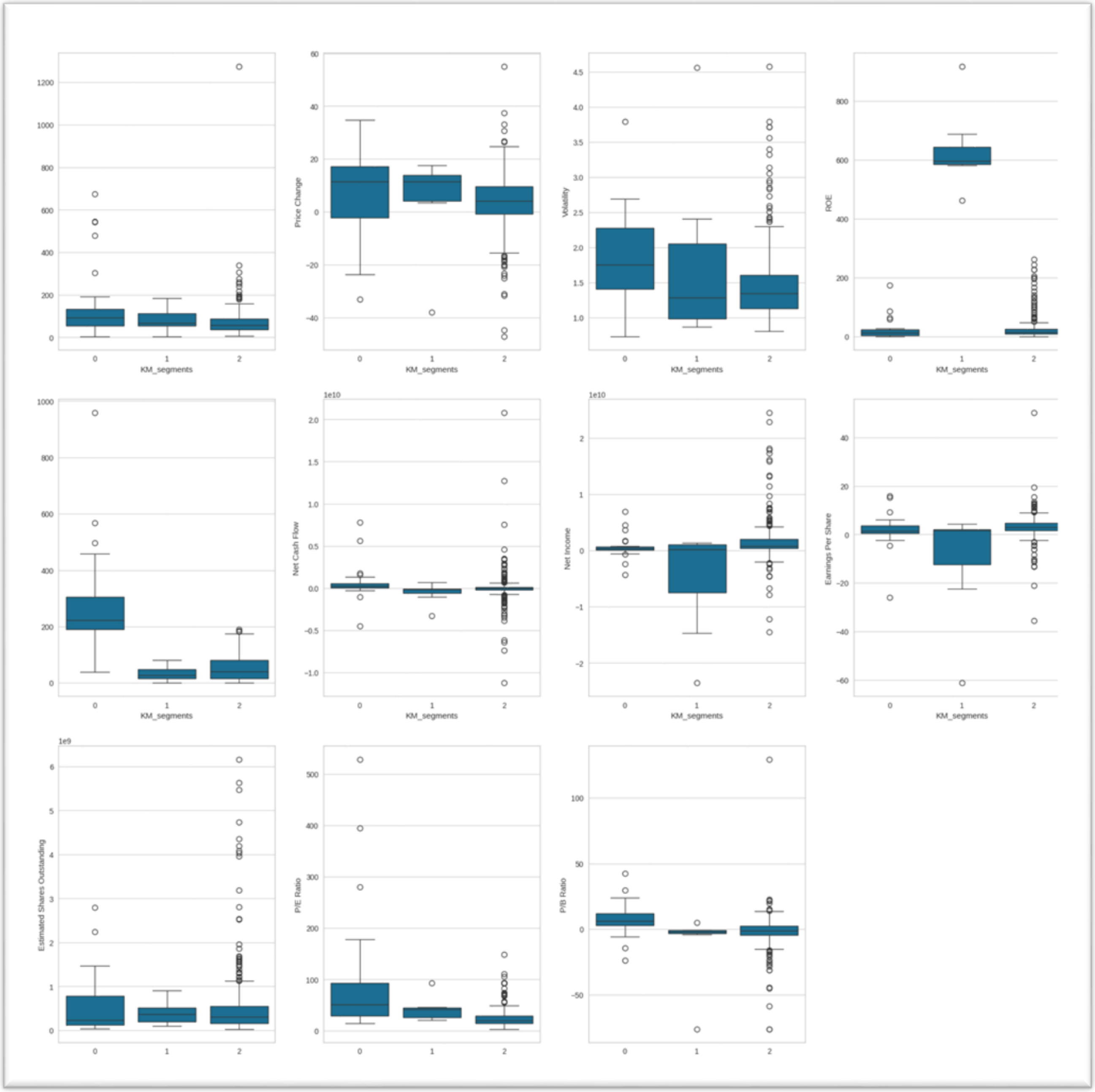


FIGURE 24

➢ Insights based on the Plot:

1. Segment 0:

- Current Price: Median is higher than the other segments but there is a considerable range of stock prices with some extreme outliers.
- Price Change: High price change with moderate variability but there are few negative outliers.
- Volatility: Segment 0 has the second highest volatility though still with a moderate range.
- Return on Equity (ROE): It is widely dispersed with several extreme outliers though the median is relatively high.
- Cash Ratio: Extremely high compared to other segments, indicating significant liquidity for most stocks. A few outliers skew the data.
- Net Cash Flow: Positive overall, though there is a wide variation with some stocks showing negative cash flow.
- Net Income: High positive net income, but with some negative outliers.
- Earnings Per Share (EPS): Mid-level earnings per share showing decent shareholder returns though there are several low outliers.
- Estimated Shares Outstanding (ESO): The share count is moderate, though there are few companies with very high number of outstanding shares.
- P/E Ratio: It has high median showing that these stocks are generally expensive relative to earnings. However, some stocks have very high P/E values, indicating overvaluation.

- P/B Ratio: The median P/B ratio is highest among the segments with substantial positive outliers reflecting that segment 0 stocks are often trade at a significant premium to their book value.

2. Segment 1:

- Current Price: This group has a similar price range to segment 2, but with few stocks at higher prices.
- Price Change: There is a broad range, with some stocks seeing positive change while others experience large losses, indicating high volatility in stock performance.
- Volatility: Similar to segment 0 but with some outliers indicate extreme volatility.
- Return on Equity (ROE): High variability in ROE, including some extreme positive and negative outliers. This indicates a mix of highly profitable companies and some that are struggling.
- Cash Ratio: Low liquidity with most companies having low cash reserves.
- Net Cash Flow: Negative cash flow for the majority of the companies though the range is wide, with a few companies showing very negative outliers.
- Net Income: The majority of the stocks in segment 1 have negative net income, showing weak profitability but with a few stocks performing exceptionally well.

- Earnings Per Share (EPS): Mostly negative suggesting these stocks are underperforming in terms of shareholder returns.
- Estimated Shares Outstanding (ESO): Relatively low, indicating smaller companies with fewer shares.
- P/E Ratio: Similar to segment 0, but with less overall variation. Some stocks are highly overvalued.
- P/B Ratio: Mostly negative or low indicating that these stocks trade at a discount to their book value, which may signal distress.

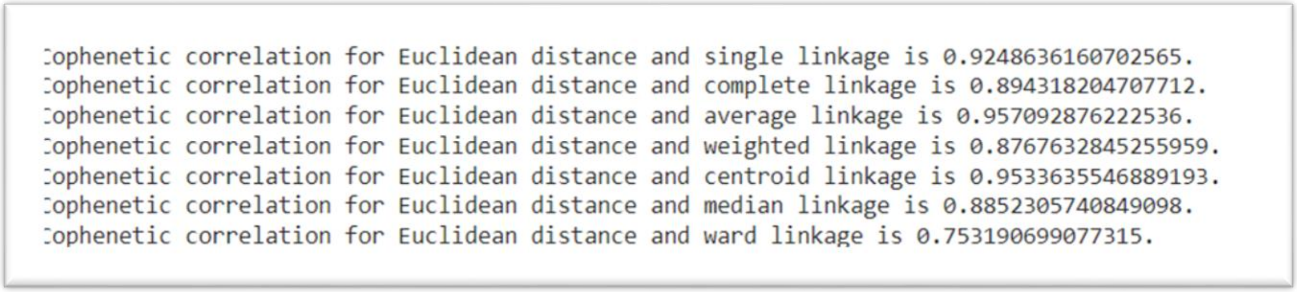
3. Segment 2:

- Current Price: Segment 2 stocks have the lowest current prices overall with fewer extreme outliers.
- Price Change: Moderate price changes with lower overall growth compared to segment 0. There is a smaller range of outliers indicating more consistent performance.
- Volatility: Lowest volatility of 3 segments with most stocks showing stable price movements.
- Return on Equity (ROE): Relatively stable with fewer extreme outliers, suggesting these stocks are more consistent in generating returns.
- Cash Ratio: Moderate liquidity with most stocks having reasonable cash reserves compared to segment 1 but lower than segment 0.

- Net Cash Flow: Positive net cash flow for the majority of the companies though the range indicates that some companies struggle with negative cash flows.
- Net Income: The highest median net income, showing that this segment contains more profitable companies although there are a few negative outliers.
- Earnings Per Share (EPS): Higher median EPS compared to the segment 1, showing better returns for the shareholders.
- Estimated Shares Outstanding (ESO): The widest range of outstanding shares indicating that this segment contains both small and large cap industries.
- P/E Ratio: Lower median compared to segment 0, suggesting more reasonable valuations.
- P/B Ratio: Generally low or negative, implying that segment 2 stocks are undervalued or trading at or below their book value, making them more attractive for the value investors.

❑ HIERARCHICAL CLUSTERING

♦ Computing Cophenetic Correlation TABLE:



```
Cophenetic correlation for Euclidean distance and single linkage is 0.9248636160702565.
Cophenetic correlation for Euclidean distance and complete linkage is 0.894318204707712.
Cophenetic correlation for Euclidean distance and average linkage is 0.957092876222536.
Cophenetic correlation for Euclidean distance and weighted linkage is 0.8767632845255959.
Cophenetic correlation for Euclidean distance and centroid linkage is 0.9533635546889193.
Cophenetic correlation for Euclidean distance and median linkage is 0.8852305740849098.
Cophenetic correlation for Euclidean distance and ward linkage is 0.753190699077315.
```

TABLE 5

➢ Insights based on the Cophenetic Correlation Table:

- **Average linkage (Cophenetic Correlation=0.957):**

This method produces the highest Cophenetic Correlation meaning it most closely represents the true distances in the data. Average linkage calculates the distance between the clusters as the average of all the pair wise distances between the points in the 2 clusters. Given its high score, it seems to balance well between compactness and separation of clusters, making it the most appropriate method for hierarchical clustering in this case.

- **Centroid Linkage (Cophenetic Correlation= 0.953):**

The second-best method which also preserves the structure of the data well. It calculates the distance between clusters based on the distance between their centroids (or mean points).

- **Single Linkage (Cophenetic Correlation= 0.925):**

This method also performs relatively well. Single linkage (or nearest neighbor) tends to form long chain –like clusters, which might not capture dense groupings as well as average linkage but still gives good representation of the distances between data points.

- **Complete Linkage (Cophenetic Correlation= 0.894):**

This method looks at the farthest distance between the points in clusters. While it performs fairly well, it does not preserve the data structure as accurately as the average or centroid linkage.

- **Other Methods:**

a) Weighted Linkage (0.877), median linkage (0.885) and ward linkage (0.753) shows lower cophenetic correlation. Particularly, Ward's method has the lowest score indicating it is not as effective in capturing the true distances between the points. Ward linkage minimizes the variance within clusters, which might result in overly compact clusters at the expense of accurately representing the original distances.

✓ The average linkage method should be preferred for this hierarchical clustering task as it produces the highest Cophenetic Correlation (0.957) indicating that it is best preserves the pairwise distribution and thus the structure of data. This method strikes a balance between minimizing inta-cluster distances and maximizing inter-cluster separation.

❑ CHECKING DENDOGRAMS

♦ DENDROGRAM PLOT:

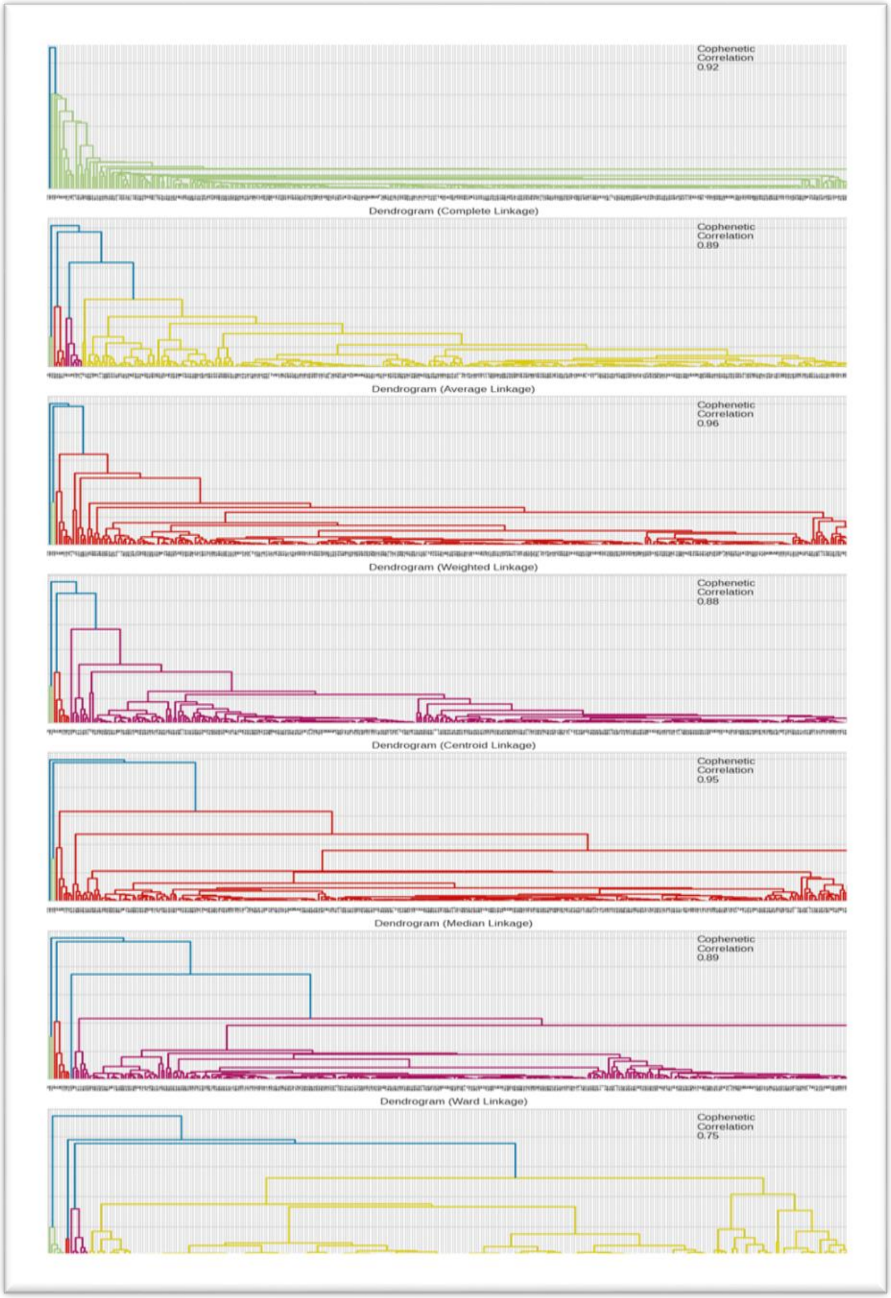


FIGURE 25

➢ Insights based on the Plot on each linkage method:

1. Single Linkage (Cophenetic Correlation= 0.92):

- The clusters are formed based on the minimum distance between points in different clusters.
- The dendrogram shows many small, elongated branches characteristic of ‘chaining’ behavior, where points are gradually merged into clusters often creating long chains of points.
- The relatively high Cophenetic Correlation (0.92) indicates that the structure somewhat reflects the true distances between the data points.

2. Complete Linkage (Cophenetic Correlation= 0.89):

- Clusters are formed based on the maximum distance between the points in different clusters.
- The dendrogram reveals more balanced and compact clusters compared to single linkage, but still some small clusters exist.
- The Cophenetic Correlation is slightly lower than that of the single linkage, showing that complete linkage captures the structure of the data moderately well.

3. Average Linkage (Cophenetic Correlation= 0.96):

- Clusters are merged based on the average pairwise distribution between all the points in different clusters.
- The dendrogram shows well-distributed cluster formations with a balance between compact clusters and smaller branches.

- The highest Cophenetic Correlation (0.96) indicates that average linkage is the most effective method at preserving the true structure of the data in this analysis.

4. **Weighted Linkage (Cophenetic Correlation= 0.88):**

- This method is similar to average linkage but gives weight to the size of the clusters.
- The dendrogram shows some compact clusters but also several small, thin branches.
- The Cophenetic Correlation (0.88) is relatively high but lower than the average linkage, meaning it provides a reasonable but not optimal representation of the data structure.

5. **Centroid Linkage (Cophenetic Correlation=0.95):**

- Clusters are merged based on the distance between their centroids. (Mean points)
- The dendrogram shows a similar characteristic to average linkage with a fairly balanced clusters and branches.
- The Cophenetic Correlation (0.95) is very high suggesting this method is effective at preserving the data structure, though slightly less than average linkage.

6. **Median Linkage (Cophenetic Correlation= 0.89):**

- This method uses the median of the distances between the clusters to merge them.
- The dendrogram shows a mix of large and small clusters but also some narrow branches.

- The Cophenetic Correlation (0.89) indicates a moderately good representation of the data, though not as strong as average or centroid linkage.

7. **Ward's Linkage (Cophenetic Correlation= 0.75):**

- Ward's method minimizes the variance with each cluster when merging.
- The dendrogram reveals large, distinct clusters with less granularity than other methods.
- The lowest Cophenetic Correlation (0.75) suggests that ward linkage is not as effective in preserving the true distances between the data points compared to the other linkage methods.

✓ Average linkage is the most effective method for this dataset, as evidenced by the highest Cophenetic Correlation (0.96), indicating that it best preserves the true distances between points while forming clusters.

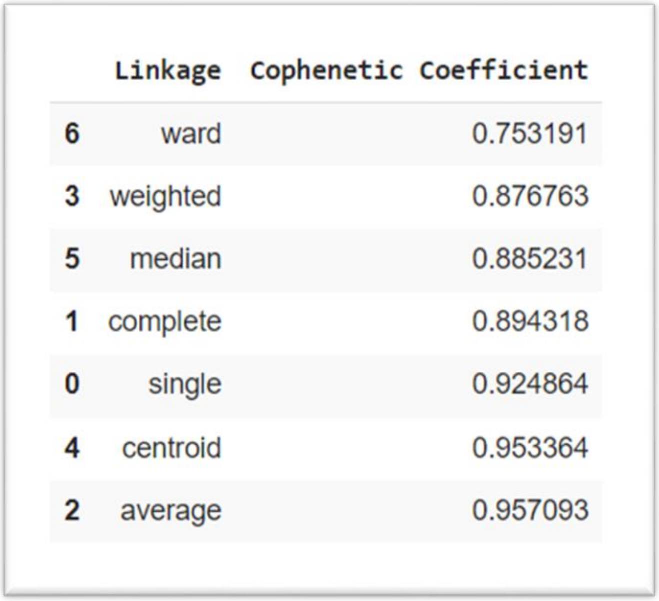
✓ Centroid linkage also performs well, with a Cophenetic Correlation close to average linkage (0.95).

✓ Ward linkage shows the least effective clustering with a low Cophenetic Correlation (0.75), indicating that the clustering forms deviates significantly from the true distances in the data.

✓ Single linkage tends to produce long chains of points, a phenomenon visible in the dendrogram and confirmed by its relatively high Cophenetic Correlation (0.92).

✓ Overall, average and centroid linkage method provides the best balance between capturing the true relationships in the data and forming well distributed clusters.

♦ **Table showing the Cophenetic Coefficient for Various hierarchical clustering linkage methods:**



	Linkage	Cophenetic Coefficient
6	ward	0.753191
3	weighted	0.876763
5	median	0.885231
1	complete	0.894318
0	single	0.924864
4	centroid	0.953364
2	average	0.957093

TABLE 6

➢ Insights based on the table:

1. Average Linkage (Cophenetic Coefficient= 0.957):

- This method produces the highest Cophenetic Coefficient, indicating that it best preserves the pairwise distances and thus the true structure of the data. It suggests that the average linkage should be preferred method for hierarchical clustering in this case.

2. **Centroid Linkage** (**Cophenetic Coefficient= 0.953):**

- This method performs almost as good as average linkage. With a very high Cophenetic Coefficient, centroid linkage is also effective in capturing the true relationships in the data, making it a strong alternative to the average linkage.

3. **Single Linkage (Cophenetic Coefficient= 0.925):**

- Single linkage performs reasonably well with a high Cophenetic Coefficient. However, single linkage tends to create long 'chains' of points which may not always be desirable when the goal is to form compact clusters.

4. **Median Linkage (Cophenetic Coefficient= 0.885):**

- Median linkage provides moderate performance. Though it preserves the data structure better than methods like ward or weighted linkage, it is not as strong as average, centroid or single linkage.

5. **Weighted Linkage (Cophenetic Coefficient= 0.877):**

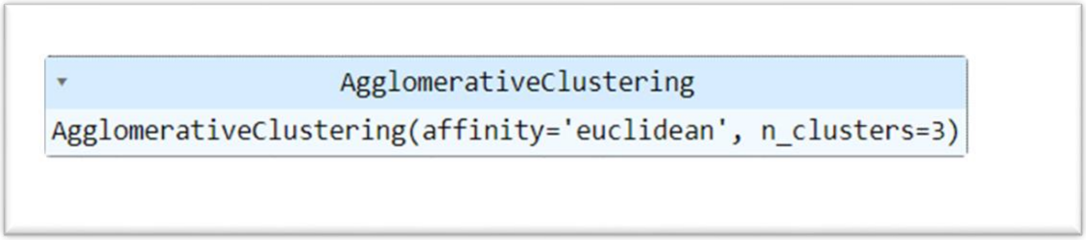
- Weighted linkage performs slightly worse than the median linkage with a Cophenetic Coefficient of 0.877. While it can still capture the data structure, its less effective than other methods.

6. **Ward Linkage:**

- Ward linkage shows the lowest Cophenetic Coefficient indicating it is the least effective method at preserving the pairwise distances between the data points. Ward's method focuses on minimizing the variance within clusters, but this might come at the expense of accurately reflecting the true structure of the data.

- ✓ Average linkage is the best clustering method based on the Cophenetic Coefficient, followed closely by centroid linkage.
- ✓ Single linkage performs well but can result in elongated clusters.
- ✓ Complete linkage is a moderate performer but not as effective as average or centroid linkage.
- ✓ Ward linkage is the least effective in terms of preserving the data structure with lowest Cophenetic Coefficient.

♦ **Creating model using sklearn:**

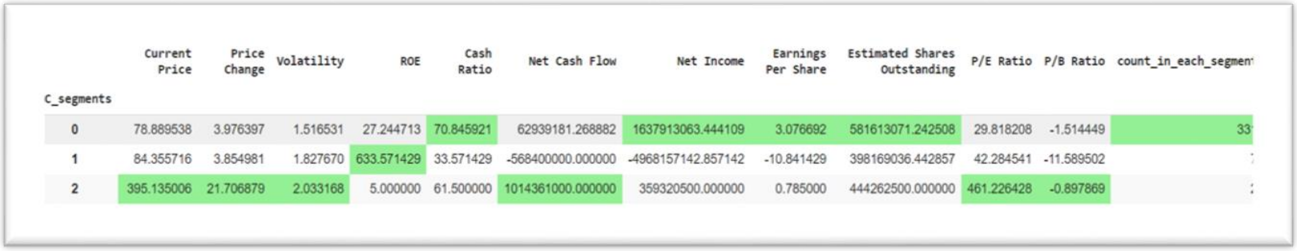


▾ AgglomerativeClustering

```
AgglomerativeClustering(affinity='euclidean', n_clusters=3)
```

❑ CLUSTER PROFILING

- **Table Summarizing financial metrics for different hierarchical clustering segments:**



	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_in_each_segmen
C_segments												
0	78.889538	3.976397	1.516531	27.244713	70.845921	62939181.268882	1637913063.444109	3.076692	581613071.242508	29.818208	-1.514449	33
1	84.355716	3.854981	1.827670	633.571429	33.571429	-568400000.000000	-4968157142.857142	-10.841429	398169036.442857	42.284541	-11.589502	[illegible]
2	395.135006	21.706879	2.033168	5.000000	61.500000	1014361000.000000	359320500.000000	0.785000	444262500.000000	461.226428	-0.897869	[illegible]

TABLE 7

➢ Insights based on the table:

1. Segment 0 (331 stocks):

- Moderate Current Price: $78.89
- Price change is 3.98% and volatility is 1.52, indicating a relative stable performance.
- ROE is 27.24% showing that companies in this segment are providing decent return on equity.
- A cash ratio of 70.85 suggests that these companies are well equipped to cover their short-term liabilities with a strong liquidity position.
- The net cash flow is positive indicating that firms generate adequate cash after operational expenses.
- Net income stands at $1.64 billion, showing profitability.

- Earnings Per Share of 3.08 is stable.
- The P/E Ratio of 29.82 indicates that investors are willing to pay for future earnings growth, but the companies are not overvalued.
- Negative P/B of –1.51 suggests undervaluation compared to the book value or that the companies have large intangible assets or debts.

2. Segment 1 (7 stocks):

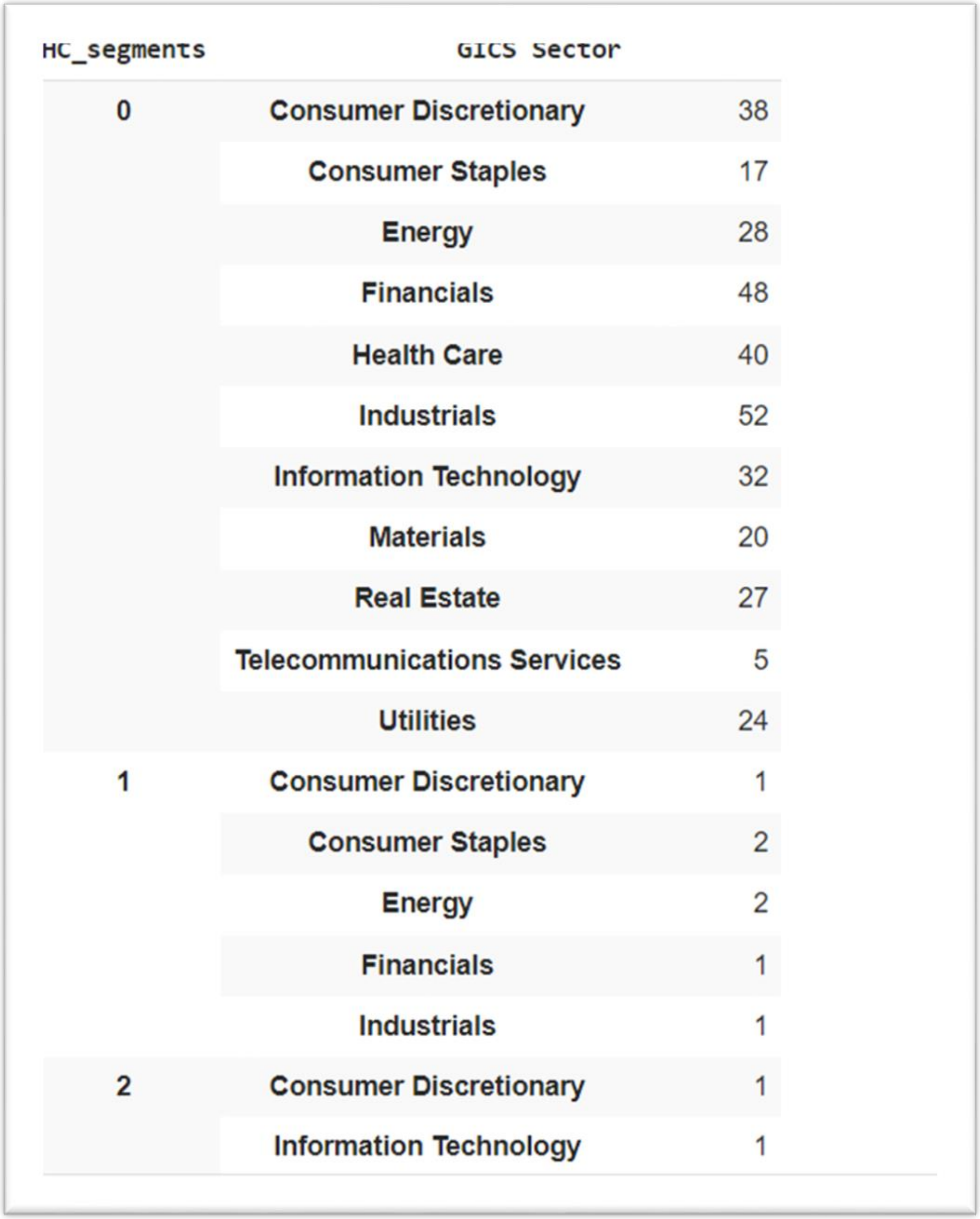
- Stable Current Price: $84.36 similar to segment 0.
- The price change is 3.85% with volatility at 1.83.
- ROE of 633.57% is extremely high, suggesting that these companies are generating massive returns on equity, possibly due to a low equity base or exceptional profitability.
- Cash ratio of 33.57 suggests moderate liquidity, but still acceptable.
- A large negative cash flow and net income point to financial distress or significant losses.
- Earnings per share of –10.84 indicates that these companies are not currently profitable.
- A high P/E ratio of 42.28 could suggest investors expect future improvements in performance, despite the current situation. The P/B ratio of –11.59 implies serious issues with balance sheet health or valuation.

3. Segment 2 (2 stocks):

- $395.14, significantly higher than the other segments indicating these are likely large cap or premium stocks.
- The price change is 21.71% suggesting a strong momentum and recent growth.
- Volatility is 2.03, indicating slightly more risk compared to other segments but still within a manageable range.
- ROE is 5.00 which is significantly lower than other segments, indicating lower returns relative to equity.
- Cash ratio of 61.50 shows strong liquidity though not as high as segment 0.
- These companies generate a substantial positive cash flow of $1.01 billion, indicating excellent operational performance.
- Net income is $359.32 million, indicating profitability but not at the level of segment 0.
- The Earnings Per Share of 0.785 suggests relatively modest earnings for the price.
- P/E ratio of 461.23 indicates that these companies are highly overvalued with expectations of future growth. This suggests speculative investing.
- The P/B ratio of –0.90 suggests an undervaluation compared to the book value or potential balance sheet risks.

- ✓ Segment 0 appears to be the most stable group while segment 1 contains higher risk companies and segment 2 represents a high growth but potentially overvalued segment.

- ♦ **Table showing HC segments and GICS sectors:**



HC_segments	GICS Sector	
0	Consumer Discretionary	38
	Consumer Staples	17
	Energy	28
	Financials	48
	Health Care	40
	Industrials	52
	Information Technology	32
	Materials	20
	Real Estate	27
	Telecommunications Services	5
	Utilities	24
1	Consumer Discretionary	1
	Consumer Staples	2
	Energy	2
	Financials	1
	Industrials	1
2	Consumer Discretionary	1
	Information Technology	1

TABLE 8

➢ Insights based on the table:

1. Segment 0- Diverse and balanced:

- Contains the majority of companies across multiple sectors, indicating a well-diversified segment.
- Significant representation from industrials (52), financials (48), healthcare (40) highlighting the prevalence of these sectors.
- Information technology (32) and energy (28) also have strong representation, reflecting the importance of technology and energy in this segment.
- This segment is broad based, suggesting stability with exposure to various market forces.

2. Segment 1- Niche, low representation:

- Very few companies with only 7 companies total spread thinly across multiple sectors.
- Many consists of consumer staples (2) and energy (2) making it a concentrated, high-risk group with less diversification.

3. Segment 2- Highly concentrated:

- Contains only 2 companies, one each from consumer discretionary and information technology.
- This segment likely consists of high-priced, niche stocks, possibly outliers with unique characteristics.

- ✓ Segment 0 is highly diversified with strong representation across major sectors, suggesting stability and reduced risks.
- ✓ Segment 1 is a smaller, concentrated segment indicating a higher risk profile.
- ✓ Segment 2 is extremely concentrated with just two companies likely representing niche, high growth stocks.

- ♦ **Box plots of numerical variable for each cluster (HC_segments):**

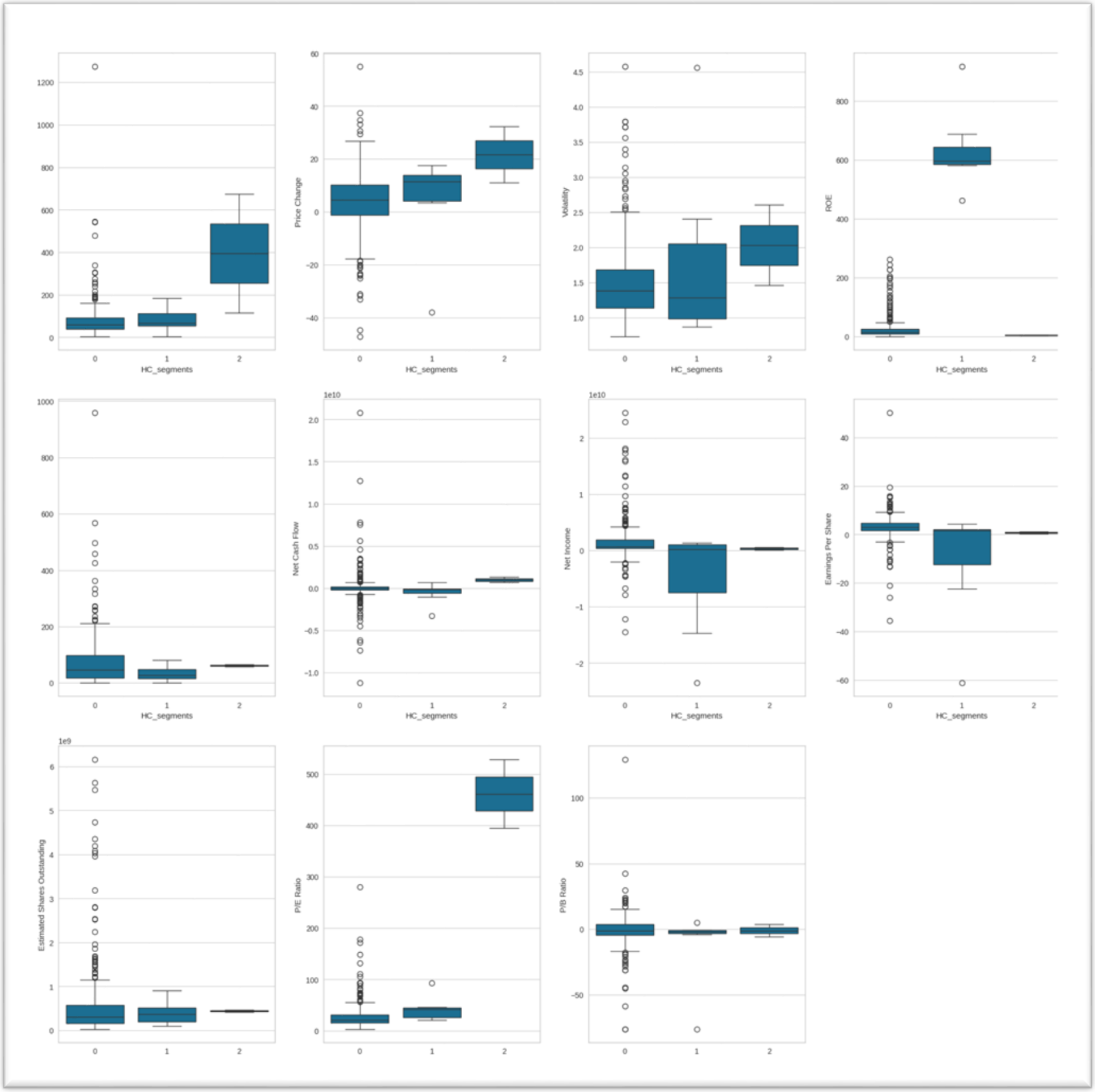


FIGURE 26

➢ Insights based on the boxplots:

1. Current Price:

- Segment 2 has significantly higher stock prices compared to segment 0 and 1, indicating presence of high value stocks.
- Segment 0 and segment 1 shows lower median prices with segment 1 showing variability and a few outliers.

2. Price Change:

- Segment 2 has the largest price increases, indicating strong recent growth.
- Segment 0 shows moderate price changes, while segment 1 has much smaller and consistent price changes.

3. Volatility:

- Segment 2 is the most volatile, suggesting higher risk but also potential for higher returns.
- Segment 0 has moderate volatility while segment 1 shows lower volatility, possibly indicating more stable or defensive stocks.

4. Return on Equity (ROE):

- Segment 1 stands out with an extremely high ROE suggesting exceptional returns for companies in this segment.
- Segment 0 has the moderate ROE, while segment 2 has the relative low ROE, despite its high prices and volatility.

5. Cash ratio:

- Segment 0 shows a relatively high cash ratio indicating stronger liquidity for these companies.

- Segment 1 and 2 have much lower cash ratios with segment 1 being particularly low, possibly indicating financial stress.

6. Net Cash Flow:

- Segment 2 exhibits the highest Net Cash Flow, reflecting strong cash generation capabilities.
- Segment 0 has a moderate net cash flow, while segment 1 has negative cash flow indicating possible cash burns or financial challenges.

7. Net Income:

- Segment 2 has the highest net income, suggesting strong profitability for the companies in this segment.
- Segment 0 has a negative net income, but segment 1 shows negative net income, reinforcing the financial difficulties inferred from other metrics.

8. Earnings Per Share (EPS):

- Segment 0 has the highest median EPS, while segment 1 shows negative earnings.
- Segment 2 has positive but relatively low EPS despite its high net income and stock prices.

9. Estimated Shares Outstanding (ESO):

- Segment 0 and 1 shows more consistency in the number of shares outstanding.
- Segment 2 shows lower median outstanding shares, potentially contributing to its higher stock prices.

10. P/E Ratio:

- Segment 2 has extremely high P/E ratios, indicating either strong growth expectations or overvaluation.
- Segment 0 has moderate P/E ratios while segment 1 is more dispersed but generally lower possibly due to financial instability.

11. P/B Ratio:

- Segment 2 has negative P/B ratio for some companies suggesting either high growth stocks or financial instability.
- Segment 0 has a moderate P/B ratio while segment 1 shows negative values, indicating undervaluation or financial distress.

✓ Segment 2 is composed of high-priced, volatile and highly valued companies with strong net income and growth potential but higher risk.

✓ Segment 0 represents a more balanced group with moderate risk, reasonable ROE and stable earnings.

✓ Segment 1 is characterized by financial challenges with negative earnings, low cash flow and lower growth potential but also the highest ROE.

❑ K-MEANS vs HIERARCHICAL CLUSTERING

❖ QUESTIONS AND ANSWERS

Q1. Which clustering technique took less time for execution?

Ans 1. **K-means execution time: 0.0174 seconds**

Hierarchical clustering execution time: 0.0045 seconds

Hierarchical clustering technique took less execution time.

Q2. Which clustering technique gave you more distinct clusters or are they same?

Ans 2. **K means cluster counts: {0:303, 1:7, 2:30}**

Hierarchical clustering counts {0:331, 1:7, 2:2}

Q3. How many observations are there in the similar clusters of both the algorithms?

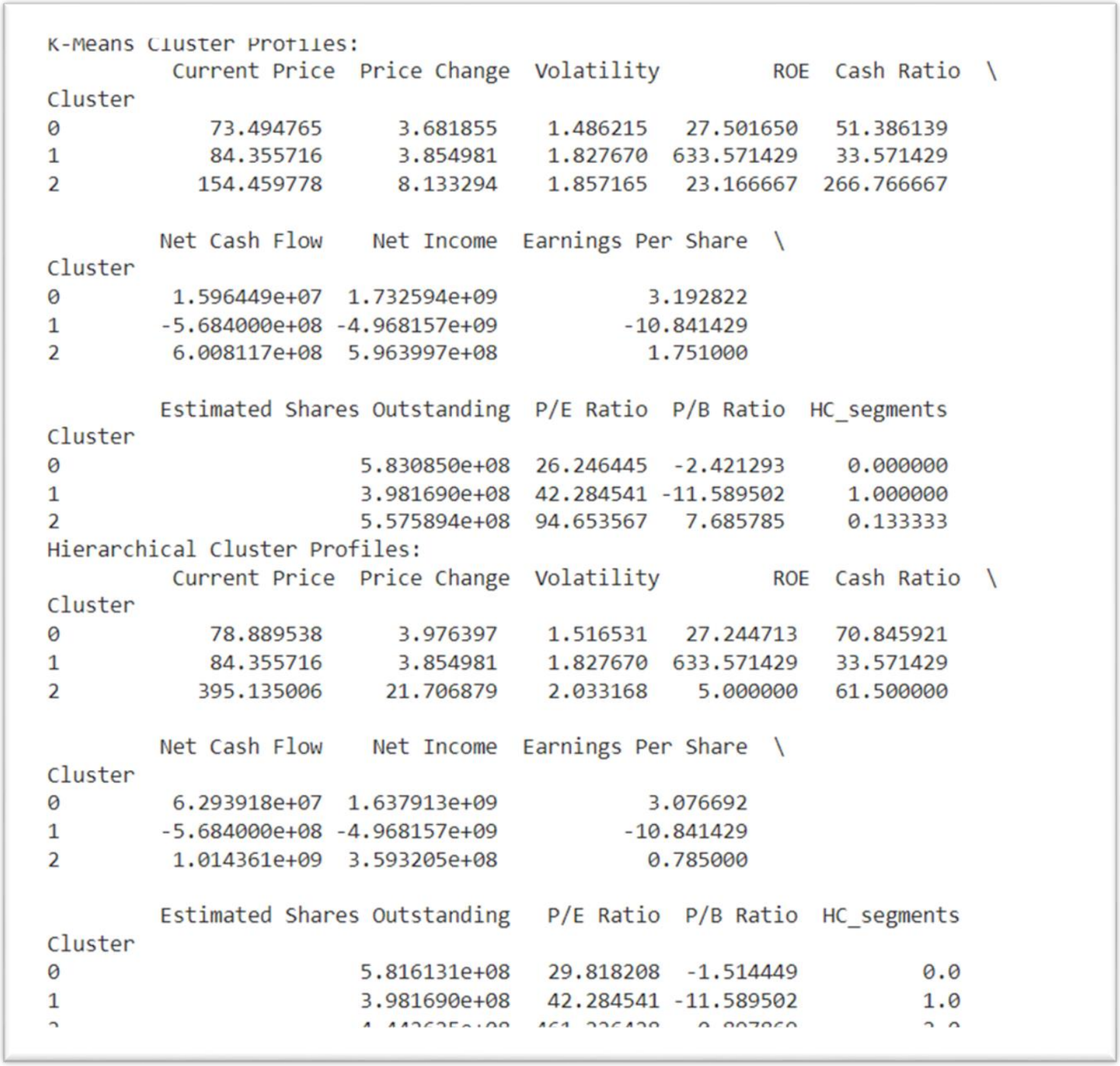
Ans 3. **Adjusted Rand Index: 0.35438**

Q4. How many clusters are obtained as the appropriate number of clusters from both the algorithms?

Ans 4. **K-Means Silhouette Score: 0.673802**

Hierarchical clustering Silhouette Score: 0.7984225

- **Table showing K-means and Hierarchical clustering profiles:**



```
K-Means Cluster Profiles:
          Current Price  Price Change  Volatility         ROE  Cash Ratio  \
Cluster
0             73.494765      3.681855    1.486215   27.501650   51.386139
1             84.355716      3.854981    1.827670  633.571429   33.571429
2            154.459778      8.133294    1.857165   23.166667  266.766667

         Net Cash Flow    Net Income  Earnings Per Share  \
Cluster
0         1.596449e+07  1.732594e+09            3.192822
1        -5.684000e+08 -4.968157e+09          -10.841429
2         6.008117e+08  5.963997e+08            1.751000

         Estimated Shares Outstanding  P/E Ratio  P/B Ratio  HC_segments
Cluster
0                        5.830850e+08  26.246445  -2.421293     0.000000
1                        3.981690e+08  42.284541 -11.589502     1.000000
2                        5.575894e+08  94.653567   7.685785     0.133333
Hierarchical Cluster Profiles:
          Current Price  Price Change  Volatility         ROE  Cash Ratio  \
Cluster
0             78.889538      3.976397    1.516531   27.244713   70.845921
1             84.355716      3.854981    1.827670  633.571429   33.571429
2            395.135006     21.706879    2.033168    5.000000   61.500000

         Net Cash Flow    Net Income  Earnings Per Share  \
Cluster
0         6.293918e+07  1.637913e+09            3.076692
1        -5.684000e+08 -4.968157e+09          -10.841429
2         1.014361e+09  3.593205e+08            0.785000

         Estimated Shares Outstanding   P/E Ratio  P/B Ratio  HC_segments
Cluster
0                        5.816131e+08   29.818208  -1.514449          0.0
1                        3.981690e+08   42.284541 -11.589502          1.0
2                        4.442625e+08  461.226438   0.907869          2.0
```

TABLE 9

➢ Insights based on the table:

1. Price and Price Change:

- K-means cluster 2 has the highest current price at $154.46 and the largest price change of 8.13, indicating high growth potential stocks in this cluster.
- Hierarchical clustering shows an even higher current price of $395.14 and a significant price change of 21.71, signaling premium stocks with large recent price movement.
- Cluster 1 in both the methods has consistent values, with slightly lower current prices around $84 and a smaller price change.

2. Volatility:

- Hierarchical clustering 2 has the highest volatility at 2.03 suggesting that stocks in this cluster are more volatile reflecting a higher risk.
- K means cluster 2 shows a comparable but slightly lower volatility of 1.86.
- Cluster 0 in both the methods has the largest volatile stocks making it more stable but with lower price changes.

3. Return on Equity (ROE):

- Both the methods show cluster 1 with an extremely high ROE of over 600% suggesting an outlier with very high profitability, likely due to a smaller number of high-performance companies.
- Cluster 2 in both the methods has much lower ROE, with hierarchical cluster 2 showing only 5% indicating more expensive but less profitable companies in terms of ROE.

4. Cash Ratio:

- K-means cluster 2 stands out with the highest cash ratio at 266.77 indicating these companies have significant liquidity.
- Hierarchical cluster 0 also shows a healthy cash ratio of 70.85 implying these companies are fairly liquid.
- Cluster 1 in both the methods has the lowest cash ratio, possibly indicating weaker liquidity or financial strain.

5. Net Cash Flow:

- Hierarchical cluster 2 has the highest net cash flow ($1.01 billion) significantly higher than the other clusters, while K-means cluster 2 also shows a substantial net cash flow of $600 million.
- Cluster 1 in both the methods show negative cash flow which could indicate struggling companies.

6. Net Income:

- K-means cluster 0 and hierarchical cluster 0 have the highest net income values around $1.6 billion suggesting strong profitability.
- Cluster 2 in both the methods has the positive but much lower net income, while cluster 1 shows negative income in both the methods reinforcing financial difficulties.

7. Earnings Per Share (EPS):

- K-Means cluster 0 has the highest EPS, at 3.19 followed by hierarchical cluster 0 with 3.07 indicating these clusters contain companies that are profitable for shareholders.

- Cluster 1 in both the methods has negative Earnings per share, indicating losses, while hierarchical cluster 2 has a very low EPS despite high stock prices.

8. **Estimated Shares Outstanding (ESO):**

- Hierarchical cluster 0 and K-means cluster 0 have the highest number of shares outstanding around 580 million, suggesting these clusters may contain larger, more established companies.
- Cluster 1 in both the methods has fewer shares outstanding (around 398 million), possibly indicating smaller companies or those with recent share buy packs.

9. **P/E Ratio:**

- Hierarchical cluster 2 has an extremely high P/E ratio (461.23) implying investors are expecting significant growth or these stocks might be overvalued.
- K-means cluster 2 has a high but more moderate P/E ratio of 94.65, still indicating high growth stocks.
- Cluster 1 i9n both the methods have the second highest P/E ratio while cluster 0 is more conservatively valued.

10. **P/B ratio:**

- Cluster 2 in both the methods has a positive P/B ratio indicating higher valuation relative to book value, whereas cluster 1 has a highly negative P/B ratio, signaling financial stress or undervaluation.

- Cluster 0 shows a slightly negative P/B ratio, but much closer to 0, indicating stable financials relative to book value.

❑ ACTIONABLE INSIGHTS AND RECOMMENDATIONS

♦ <u>ACTIONABLE INSIGHTS:</u>

1. Sector Performance Analysis:

- **Industrials:**

The stocks for Americal Airlines Group (AAL), shows a high price change of 9.99% and a remarkable Return on Equity (ROE) of 135% indicating strong profitability. However, its net cash flow is negative signaling potential liquidity concerns.

  - Focus on managing the cash reserves to ensure operational stability, despite strong earnings performance.

- Healthcare: Both AbbVie and Abbott Laboratories exhibit positive net income and relatively lower P/E ratios (18.80 and 15.28 respectively) indicating the stocks are fairly valued. However, the volatility is moderate.

  - Continue investing in research and development to maintain competitive advantage, while monitoring market sentiment to manage volatility.

- Information Technology: Adobe System has a high P/E ratio of 74.56, suggesting potential overvaluation. Its ROE of 9% is modest compared to the peers.
  - Re-evaluate pricing strategies or seek innovation to justify the high valuation as current profitability is not supporting at the higher price.

2. Cash Flow and Liquidity Management:

- The cash ratio is a key indicator of a company’s liquidity. Companies like Analog Devices (ADI) and Adobe System (ADI) have cash ratios of 272% and 180% which is much higher than needed, possibly signaling an inefficient use of resources.
  - Companies should consider reinvesting excess cash into growth opportunities or returning the capital to shareholders to improve the return on assets.
- Conversely, American Airlines shows a negative net cash flow of -$604 million which combined with its strong earnings indicates high operational outflows.
  - The airlines should implement cost control measures and consider refinancing or restructuring its debt to improve its liquidity position.

3. Valuation Metrics:

- The P/B ratio of AbbVie and Abbott Laboratories are negative which may indicate a discrepancy in market expectations versus the book value.

- These companies should engage in the investor relations activities to better communicate the value of their underlying assets and improve the market confidence.
- Adobe Systems Inc has an exceptionally high P/E ratio (74.56) which may indicate an overvaluation.
- Adobe should explore measures to increase the earnings to justify its market price or face potential price concerns.

4. Profitability:

- Companies like American Airlines and AbbVie demonstrates strong ROE of 135% and 130% respectively. Highlighting high profitability. However, these companies have relatively high volatility which can concern risk-averse investors.
- Implementing hedging strategies or diversifying revenue streams could help in mitigating volatility and ensuring steady returns.
- Analog devices show a low ROE of 14% combined with recent price change of –1.83%
- The company should consider exploring cost-cutting strategies or improving the operational efficiency to enhance the shareholder returns.

5. Earnings and Growth Potential:

- Earnings Per Share (EPS) is generally positive across the dataset with American Airlines leading at $11.39 EPS, followed by AbbVie at $3.15.

- ▪ Companies with strong EPS should reinvest in growth areas or focus on shareholder returns like dividend or share buybacks to capitalize on their earnings potential.
- Adobe System shows a relatively low EPS of 1.26 despite its high market valuation.
- ▪ Adobe should focus on improving the operational efficiency or expanding into high growth areas to enhance its earnings potential and better align with market expectations.

6. Rish and Volatility Management:

- Volatility is particularly high in AbbVie (2.20%) and Analog Devices (1.70%) which suggests the stock prices may fluctuate significantly.
- ▪ These companies should focus on stabilizing their earnings through diversified portfolios or long-term contracts to reduce volatility and provide more predictable returns for the investors.

7. Investment Strategy Recommendations:

- Buy: Stocks like AbbVie and Abbott Laboratories offer balanced growth potential with relatively stable earnings and reasonable valuations.

- Hold: Adobe Systems may be overvalued but could justify its price with continued innovation and expansion into new markets.
- Sell/ Watch: Analog Devices show recent negative price movement and relatively low ROE which may warrant caution unless there is evidence of operational improvements.

The insights aim to guide strategic decision-making by focusing on the performance metrics, risk and profitability while aligning resource with market demands.

♦ **RECOMMENDATIONS:**

1. Optimize Cash Flow Management:

- Excessive Cash flow reserves:

Companies like Analog Devices (ADI) and Adobe Systems (ADI) have high cash ratios (272% and 180% respectively) indicating excess liquidity that could be better utilized. These companies should consider:

a) Reinvesting cash in growth opportunities, R&D or strategic acquisitions to increase the operational efficiency and market share.

b) Implementing share holder friendly initiatives such as stock buybacks or increased dividends payouts to maximize the shareholder value.

- Cash flow deficits:

American Airlines (AAL) is experiencing a negative net cash flow (-$604 million) despite the high earnings. The airline should:

a) Focus on reducing the operational costs, especially in fuel and labor to improve the cash flow.
b) Explore refinancing options or debt restructuring to strengthen liquidity.

2. Address Overvaluation Concerns:

- High P/E Ratios:

Adobe System has a very high P/E Ratio (74.56) suggesting overvaluation. To prevent market correlations Adobe should:

a) Accelerate innovation in key areas such as cloud services or Artificial Intelligence to justify its market valuation.
b) Increase operational efficiency to drive higher earnings growth and better align valuation with financial performance.
c) Monitor competitor movements in the tech space and respond with aggressive strategic initiatives.

3. Improve Asset Utilization and Profitability:

- Low Return on Equity (ROE):

Analog Devices (ADI) has a relatively low ROE of 14% which could indicate inefficient uses of its assets. ADI should:

a) Focus on improving the operational processes, reducing the costs and enhancing product development to increase returns.
b) Consider optimizing its capital structure by leveraging more debt to generate higher returns for shareholders.

4. **Mitigate Volatility to attract Risk-averse Investors:**

- Companies like AbbVie and Analog Devices show relatively high volatility (2.20% and 1.70% respectively) which may deter risk averse investors. These firms should:

a) Engage in hedging strategies to reduce the market risks.
b) Diversify their revenue streams to create more stable, predictable earnings, which will reduce volatility and attract long term investors.
c) Communicate with investors about long-term strategies to reassure them about the company's ability to withstand market fluctuations.

5. Explore new growth opportunities:

- Earnings growth potential:

Several companies like AbbVie and Abbott Laboratories demonstrate solid EPS, indicating strong earnings potential. These companies should:

a) Consider expanding into new geographical markets or product segments particularly in high-growth regions or under penetrated markets to capitalize on their strong financial positions.
b) Increase investment in innovation, especially in healthcare, to maintain a competitive edge and create long term growth opportunities.

6. Strengthen Investor Relations:

- Negative P/B Ratios:

Companies like AbbVie and Abbott Laboratories exhibit negative P/B Ratios, potentially indicating undervaluation by the market or poor perception of the company's assets. To address this, these companies should:

a) Strengthen investor relations by providing clearer communication regarding the company's asset values, growth strategies and market positions.
b) Highlight the intrinsic value of their business in investor reports and public communications to improve the market sentiment and stock valuation.

7. Investor in Sector Specific Initiatives:

- <u>Sector-Specific Insights:</u>

Companies should leverage sector specific trends:

a) Healthcare companies like AbbVie and Abbott laboratories should continue investing in cutting-edge pharmaceutical research and development to stay ahead in competitive market.

b) Information Technology companies like Adobe Systems should focus on expanding the software solutions in high-demand sectors such as Artificial Intelligence, cyber security and cloud computing.

8. Rebalance Investment Portfolios:

- Based on the performance and valuation metrics:
- Buy: AbbVie and Abbott exhibit strong earnings and moderate valuations, making them attractive buys.
- Hold: Adobe system has a strong potential but is currently overvalued. Investors should wait for the earnings to catch up before making any moves.
- Sell/ Watch: Analog Devices shows concerning signs with a negative price change and relatively low ROE. Investors should monitor the company closely for improvements before committing further.

These recommendations focus on leveraging financial strengths, managing risks and optimizing the growth opportunities. They are designed to help each company improve its market position, enhance profitability and attract a more diversified investor base.