Trade&Ahead

Travis Bruns

Table of Contents

- Context
- Objective
- Data Overview
- Data Summary
- Univariate Analysis
- Bivariate Analysis
- K-Means Clustering
- Hierarchical Clustering
- K-Means vs Hierarchical Clustering
- Recommendations

Context

The stock market has consistently proven to be a good place to invest in and save for the future. There are a lot of compelling reasons to invest in stocks. It can help in fighting inflation, create wealth, and also provides some tax benefits. Good steady returns on investments over a long period of time can also grow a lot more than seems possible. Also, thanks to the power of compound interest, the earlier one starts investing, the larger the corpus one can have for retirement. Overall, investing in stocks can help meet life's financial aspirations.

It is important to maintain a diversified portfolio when investing in stocks in order to maximize earnings under any market condition. Having a diversified portfolio tends to yield higher returns and face lower risk by tempering potential losses when the market is down. It is often easy to get lost in a sea of financial metrics to analyze while determining the worth of a stock, and doing the same for a multitude of stocks to identify the right picks for an individual can be a tedious task. By doing a cluster analysis, one can identify stocks that exhibit similar characteristics and ones that exhibit minimum correlation. This will help investors better analyze stocks across different market segments and help protect against risks that could make the portfolio vulnerable to losses.

Objective

Trade&Ahead is a financial consultancy firm who provide their customers with personalized investment strategies. They have hired you as a Data Scientist and provided you with data comprising stock price and some financial indicators for a few companies listed under the New York Stock Exchange. They have assigned you the tasks of analyzing the data, grouping the stocks based on the attributes provided, and sharing insights about the characteristics of each group.

Data Overview

The data provided is of stock prices and some financial indicators like ROE, earnings per share, P/E ratio, etc.

- 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)

Raw Data Summary

- The Source Data includes 340 rows and 15 columns.
- The Data types are 7 float, 4 integers, and 4 objects.
  - Float: Current Price, Price Change, Volatility, Earnings Per Share, Estimated Shares Outstanding, P/E Ratio, and P/B Ratio
  - Integer: ROE, Cash Ratio, Net Cash Flow, and Net Income
  - Object: Ticker Symbol, Security, GICS Sector, and GICS Sub Industry
- The Data does not include any missing values.
- The Data does not include any duplicate values.
- The Data is not missing any values.
- The majority of the stocks are classified as Industrials Financials, Consumer Discretionary, Health Care, Information Technology, or Energy.

Data Summary

- Industrials account for 15.6% of the securities.
- Financials account for 14.4% of the securities.
- Consumer Discretionary stocks account for 11.8% of the securities.
- Healthcare stocks account for 11.8% of the securities.
- Information Technology accounts for 9.7% of the securities.
- Energy accounts for 8.8% of the securities.
- Real Estate accounts for 7.9% of the securities.
- Utilities account for 7.1% of the securities.
- Materials account for 5.9 of the securities.
- consumer Staples account for 5.6% of the securities.
- Telecommunication Services account for 1.5% of the securities.

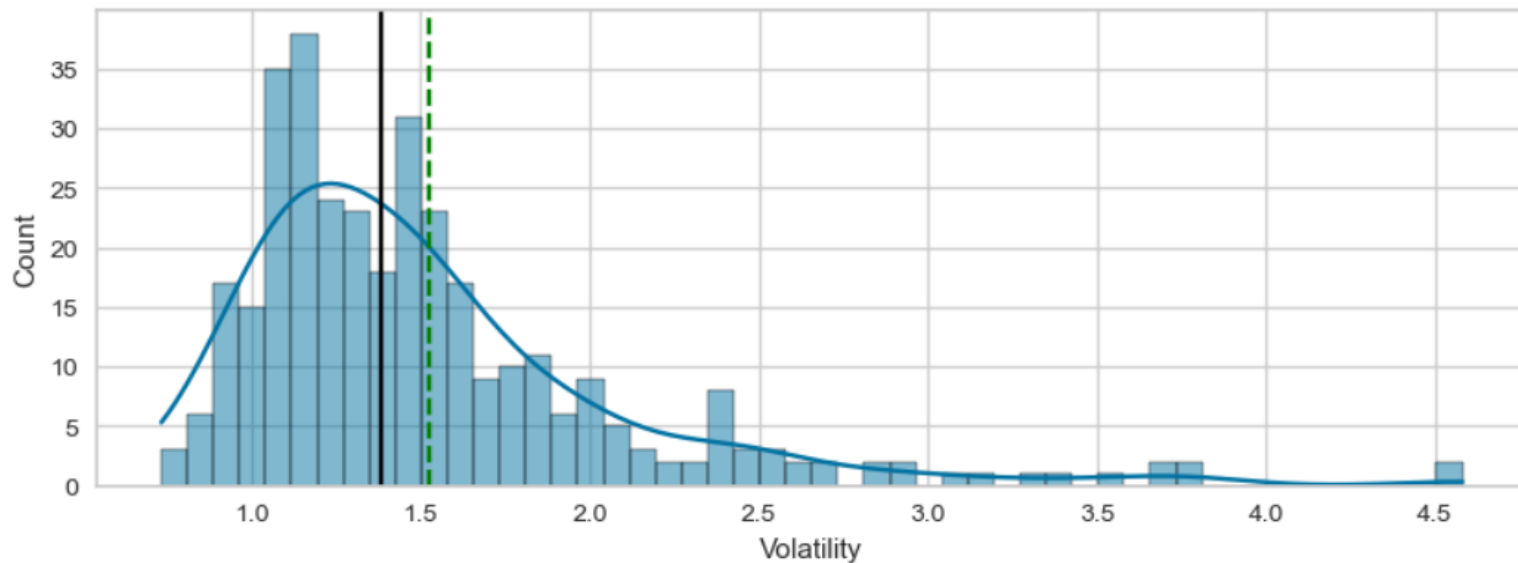# Univariate Data Analysis – Current Price
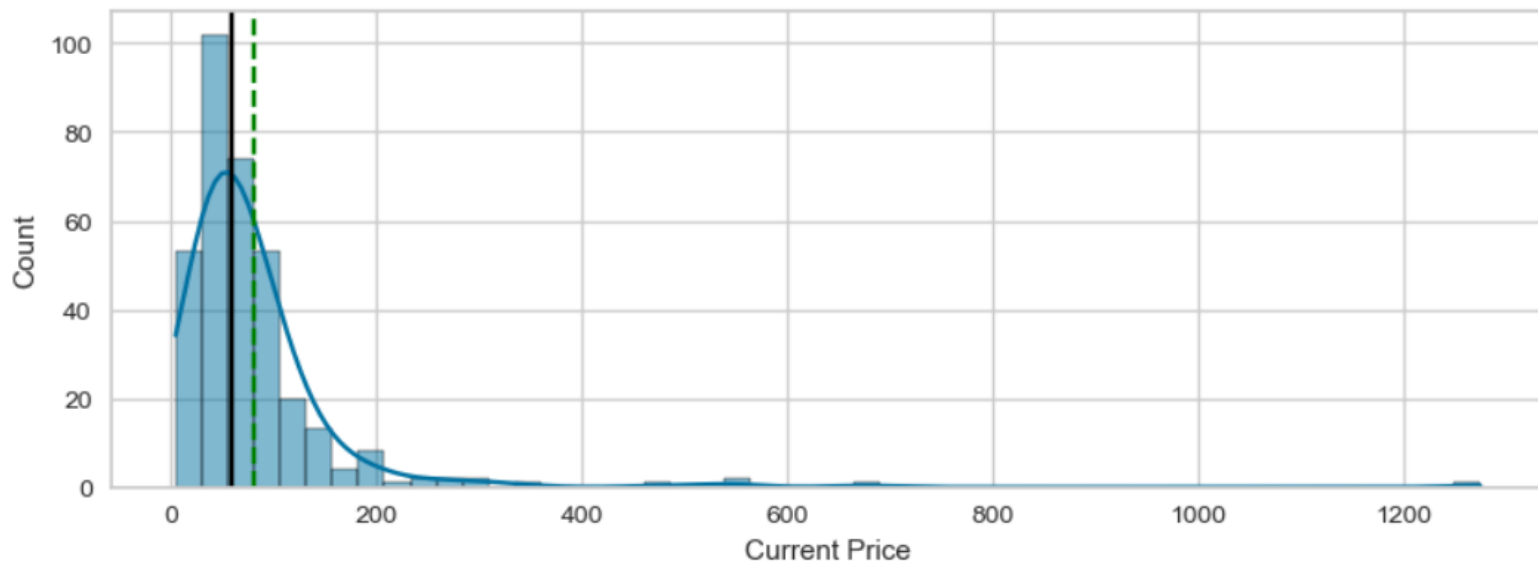
Current Price is right-skewed and has extreme upper outliers.

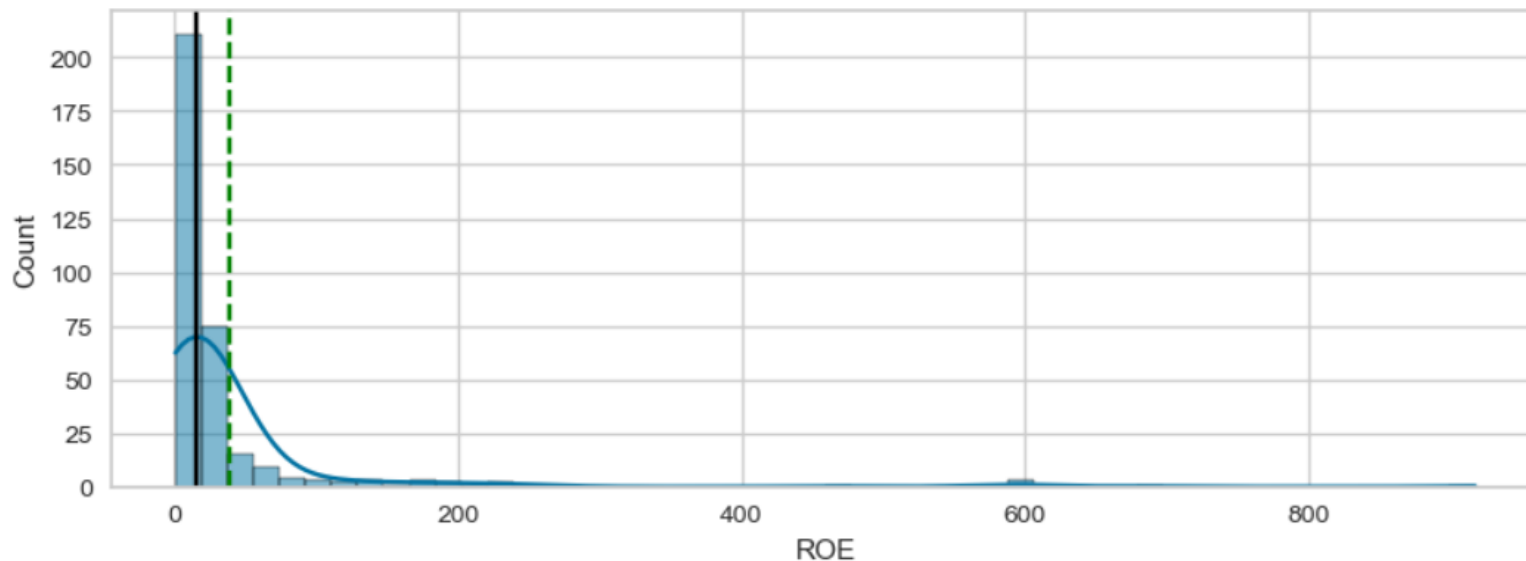# Univariate Data Analysis – Price Change
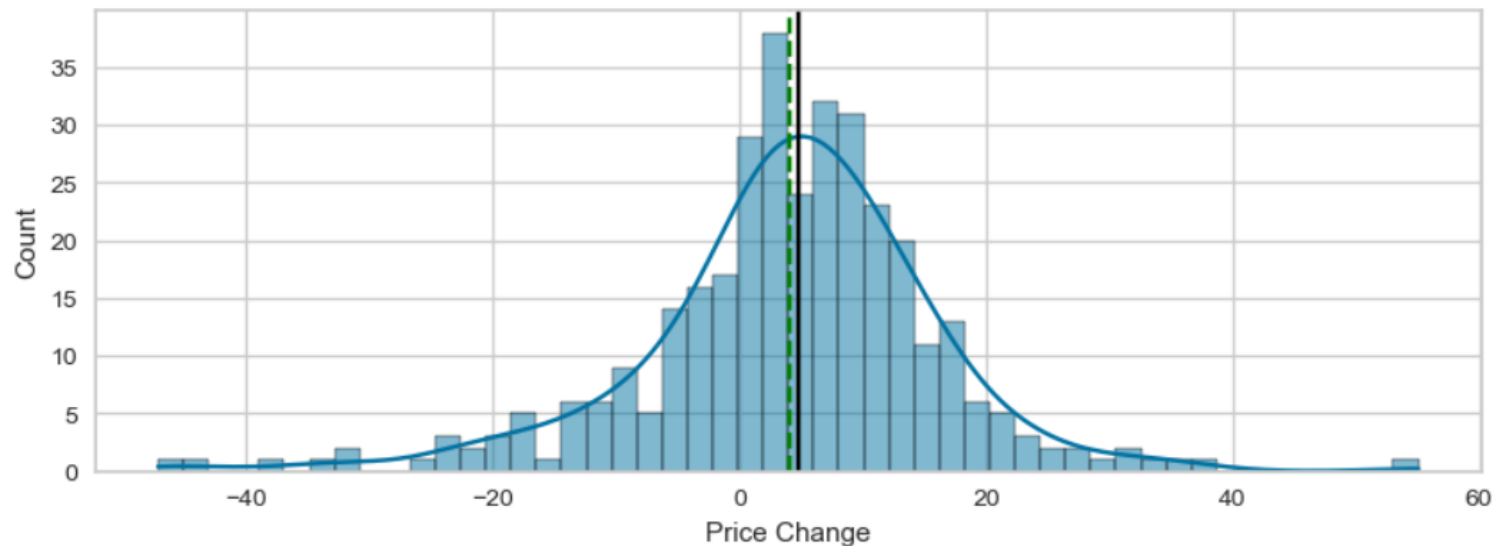
Price Change is normally distributed.
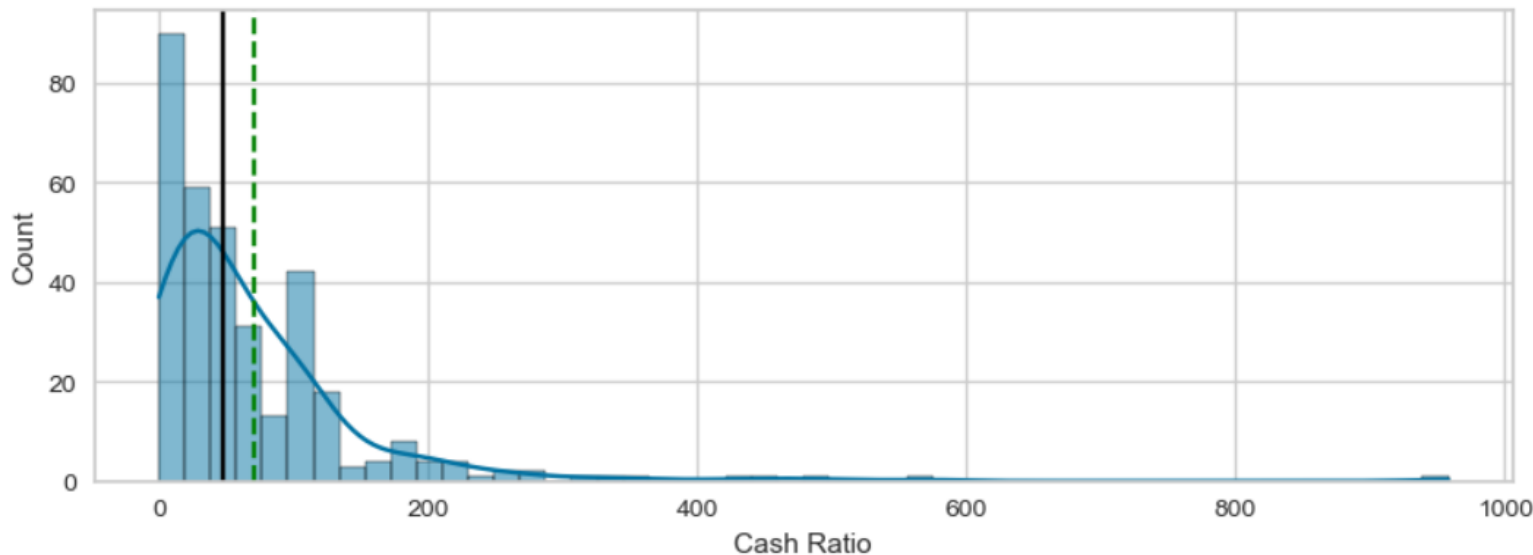
Univariate Data Analysis – Volatility

Volatility is right-skewed and has upper outliers.
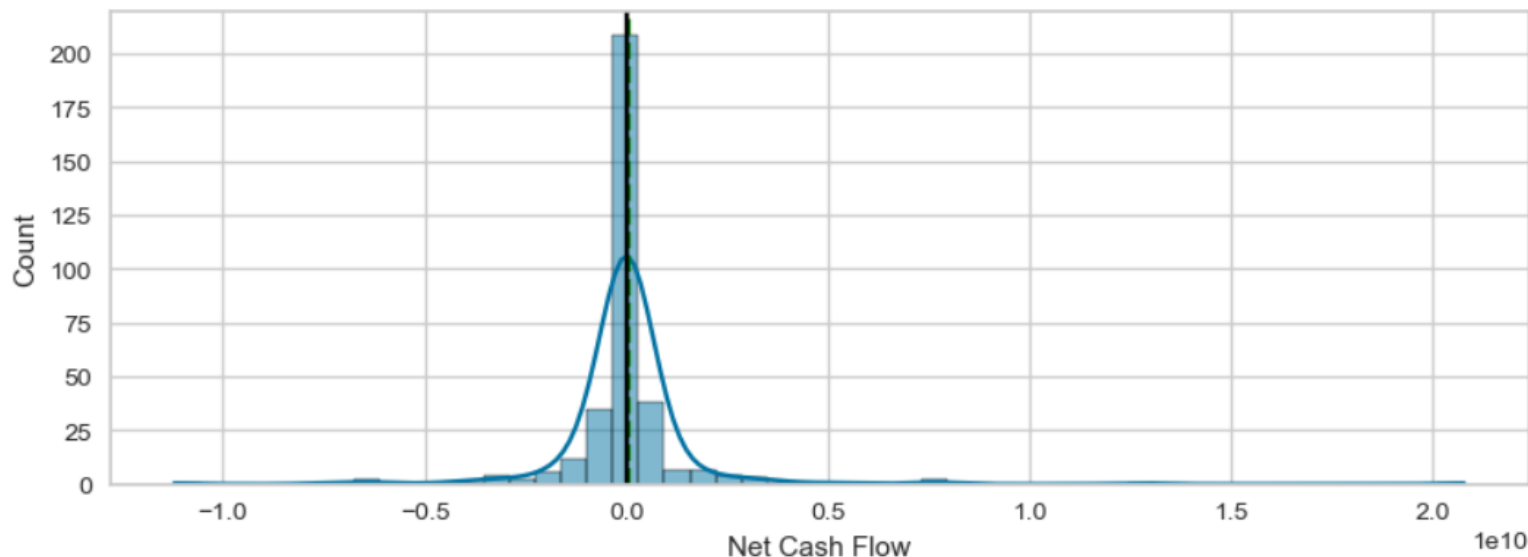
Univariate Data Analysis – ROE

ROE is right-skewed and has upper outliers.

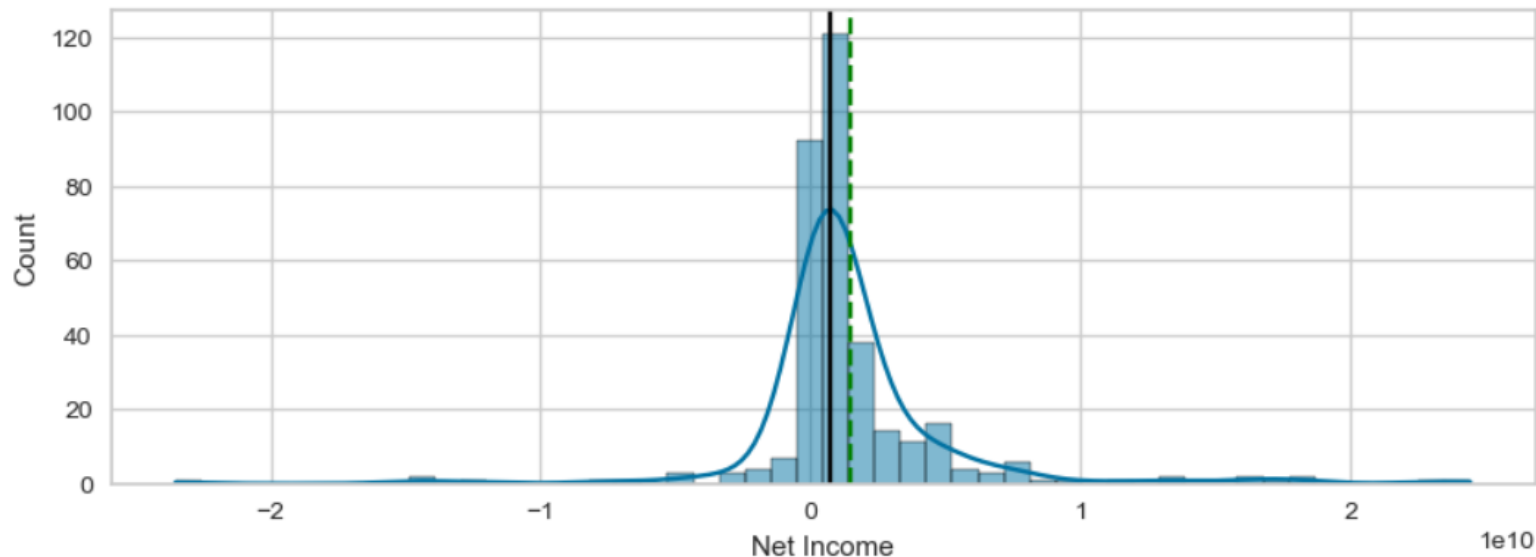# Univariate Data Analysis – Cash Ratio

Cash Ratio is right-skewed and has upper outliers.
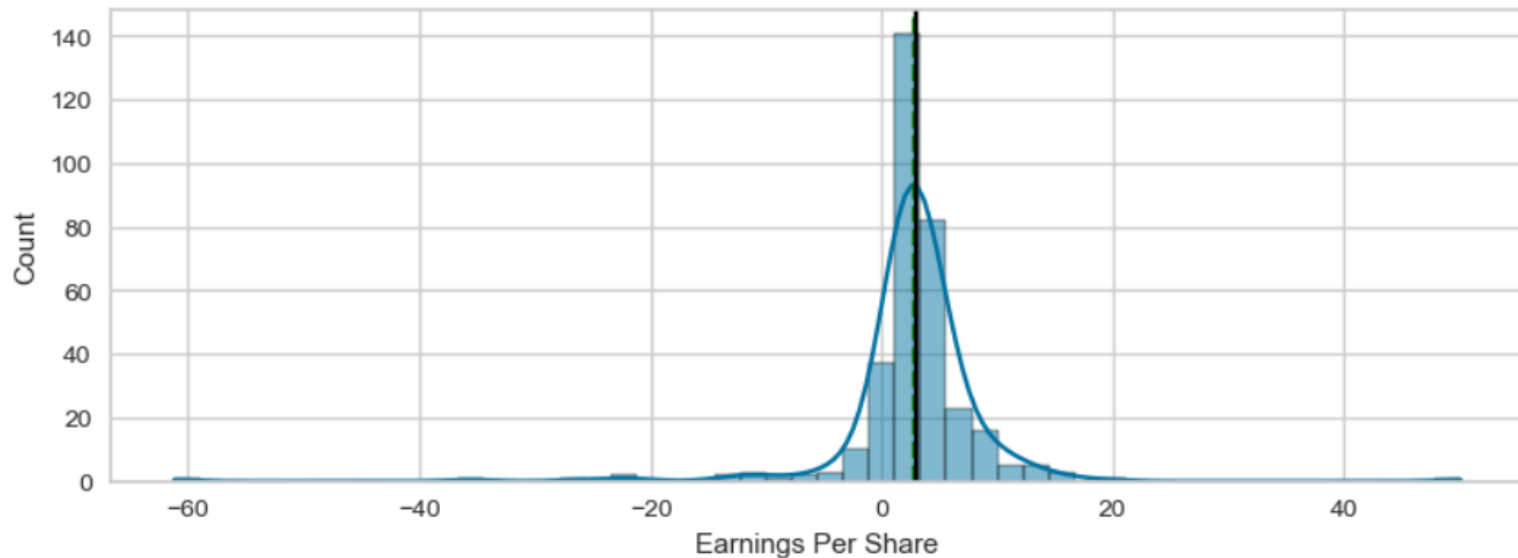
Univariate Data Analysis – Net Cash Flow

Net Cash Flow is normally distributed.

Univariate Data Analysis – Net Income

Net Income is normally distributed.
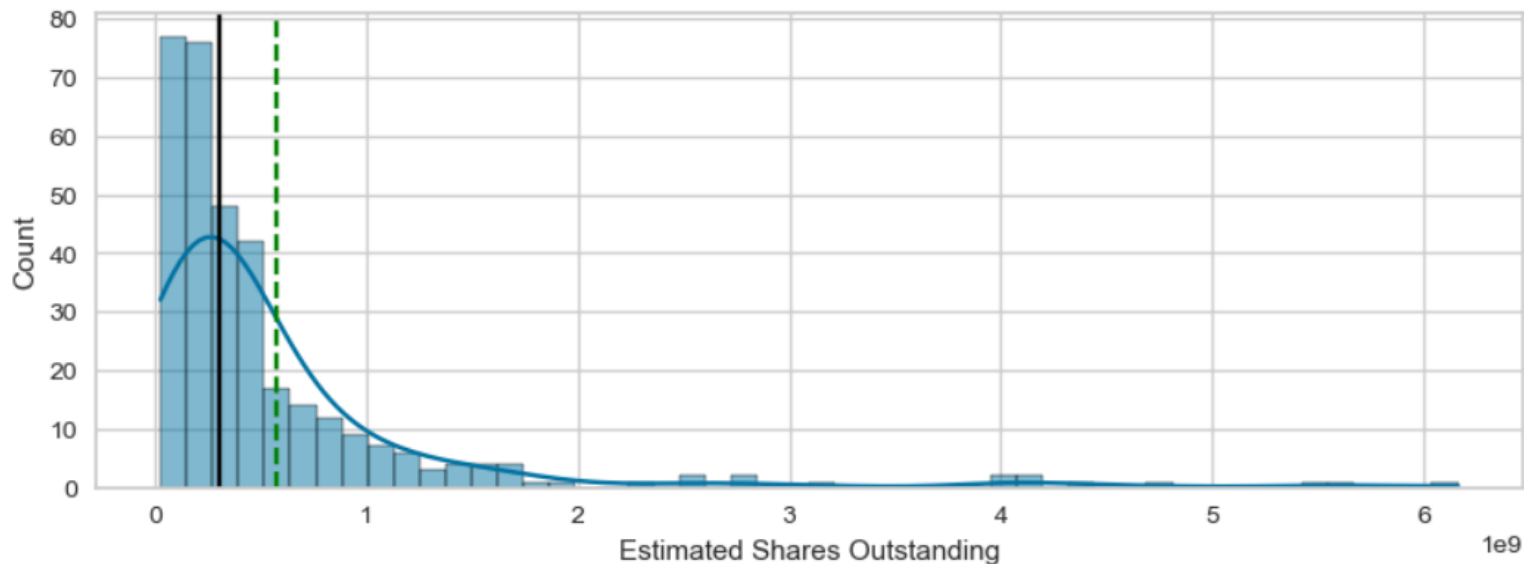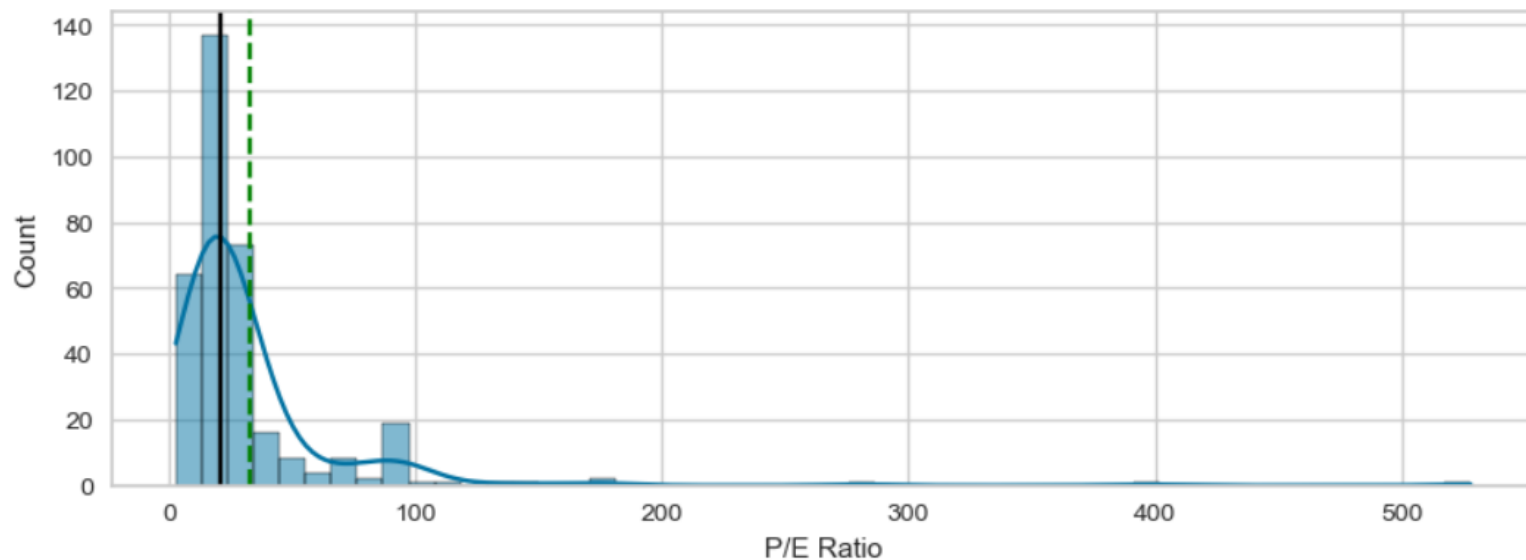
Univariate Data Analysis – Earnings Per Share

Earnings per Share is normally distributed.

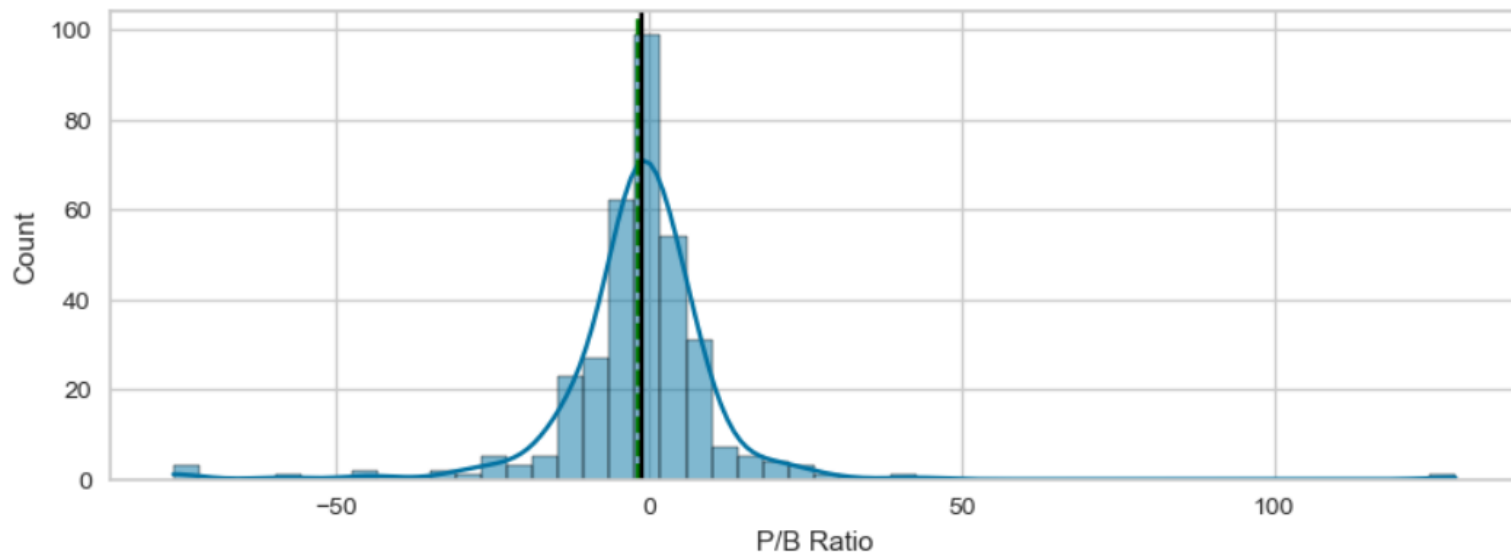# Univariate Data Analysis – Estimated Shares Outstanding

Estimated Shares Outstanding is right-skewed and has upper outliers.

Univariate Data Analysis – P/E Ratio

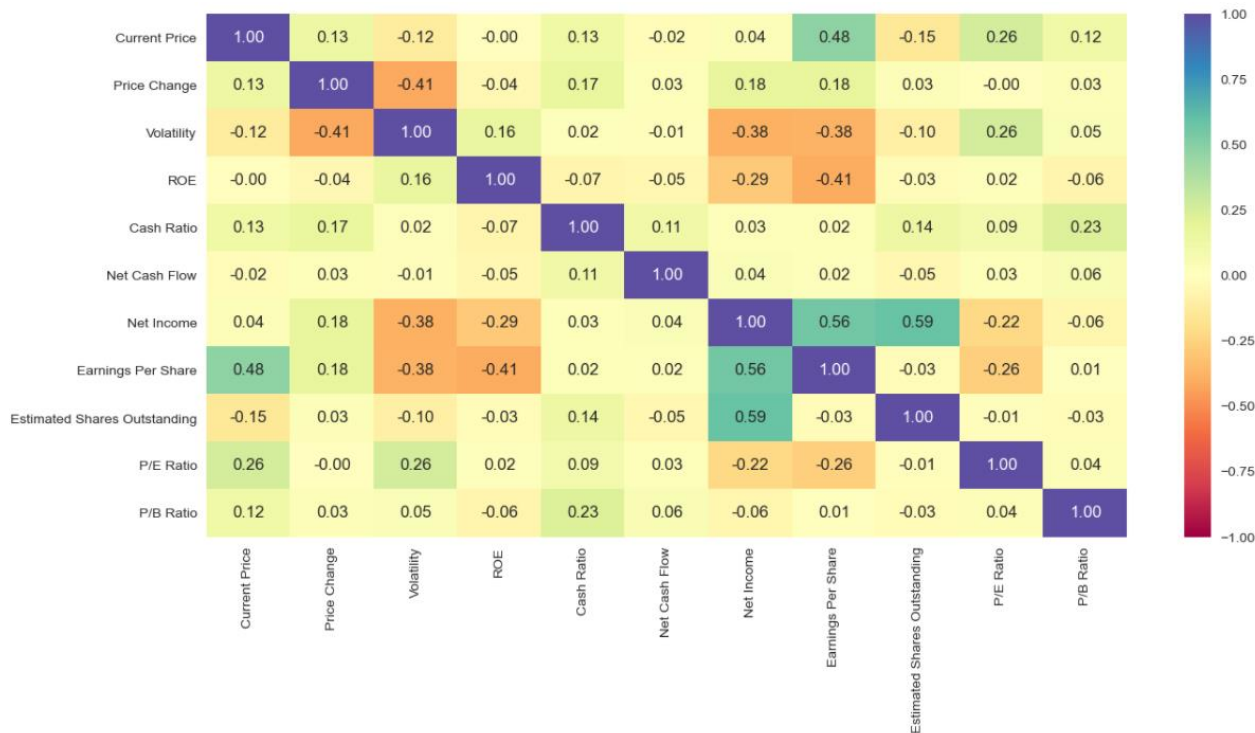

P/E Ratio is right-skewed and has upper outliers.

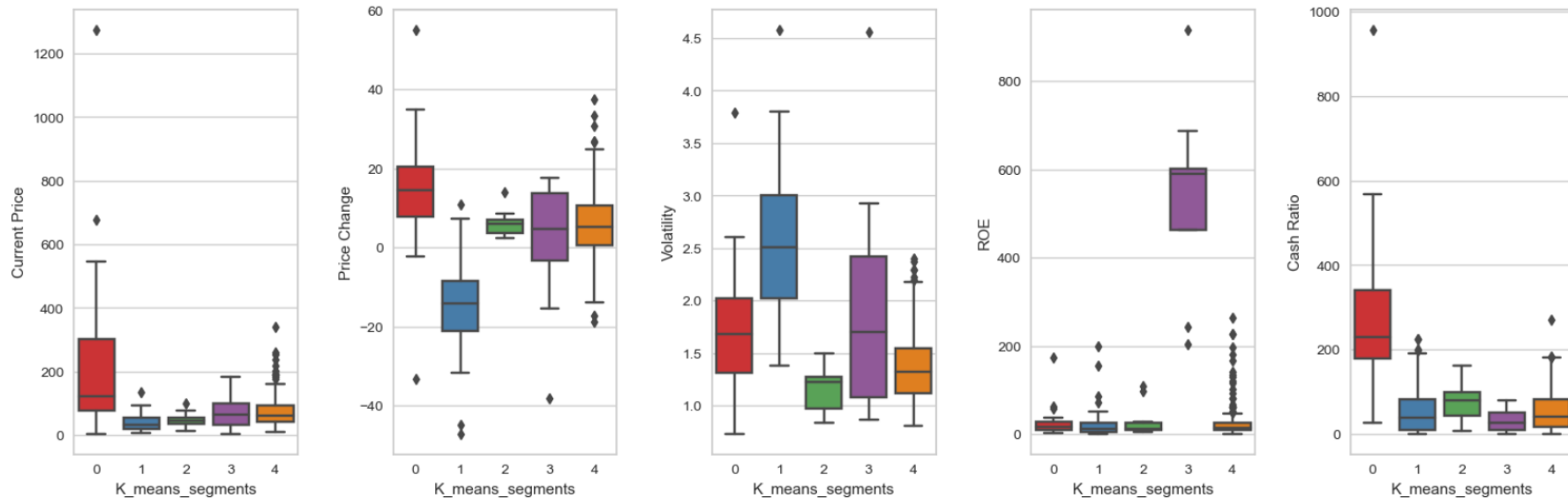# Univariate Data Analysis – P/B Ratio

P/B Ratio is normally distributed.

Bivariate Data Analysis – Integer and Float Data

- Net Income is correlated to Estimated Shares Outstanding and Earnings per Share. It is also somewhat negatively correlated to Volatility and ROE.

- Earnings Per Share is correlated to Current Price and Net Income. It is somewhat negatively correlated to Volatility, ROE, and P/E Ratio.

- P/E Ratio is somewhat correlated to Current Price and Volatility. It is somewhat negatively correlated to Net Income and Earnings Per Share.

- Cash Ratio is somewhat correlated to P/B Ratio.

	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
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
Estimated 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

K-Means Clustering with 5 Clusters - Visuals

K-Means Clustering with 5 Clusters - Insights

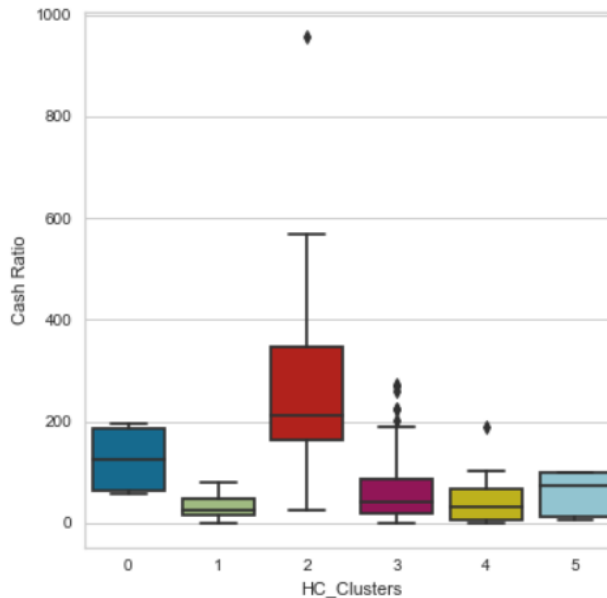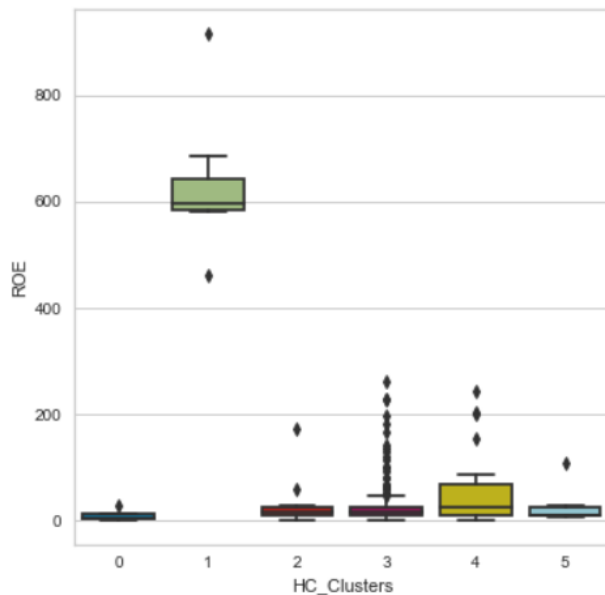| Cluster 0 (red)	Cluster 1 (blue)	Cluster 2 (green)
• 24 Securities • Current price is higher than other clusters with one outlier being over $1200. • Price change is almost all above 0 with about 25% being over 20% increase. • Volatility lies predominantly between 1.25 and 2. • ROE is low and lies near clusters 1, 2, and 4. • Cash ratio is all positive with almost 75% of companies having a cash ratio greater than 200%.	• 35 Securities • Current price is on average lower than the other clusters. • Price change is predominantly negative - the worst performing cluster. • Volatility is on average higher than other clusters with 75% lying above 2.0. • ROE is low and lies near clusters 0, 2, and 4. • Cash ratio has a range between 0 and 200% with the majority of companies being below 100%.	• 13 Securities • Current price for all securities is below $100. • Price change is positive with the cluster having the smallest range of all clusters. • Volatility is the least of all clusters with the range of securities being ~.75 and 1.5. • ROE is low and lies near clusters 0, 1, and 4. • Cash ratio is positive for all securities and ranges ~5% to ~175%.

K-Means Clustering with 5 Clusters - Insights

Cluster 3 (red)	Cluster 4 (blue)
9 Securities - Current price rangest between just above 0 to just under 200. - Price change is from ~-15% to ~25 with over 75% providing positive returns. - Volatility contains the largest midrange with its interior 50% ranging from just above 1 to just below 2.5. - ROE is significantly higher than the other clusters with all companies having ROE greater than 200. - Cash ratio contains the smallest range with companies being between 0 and 100%.	259 Securities - Current price contains the smallest midrange with most companies being between 5 and 175. - Price change is from ~-15% to ~18 with over 50% providing positive returns. - Volatility midrange lies between ~1.1 and ~1.6. - ROE is low and lies near clusters 0, 1, and 2. - Cash ratio has a range between 0 and 200% with the majority of companies being below 100%. It is very similar to cluster 1.

Hierarchal Clustering with 6 Clusters - Visuals
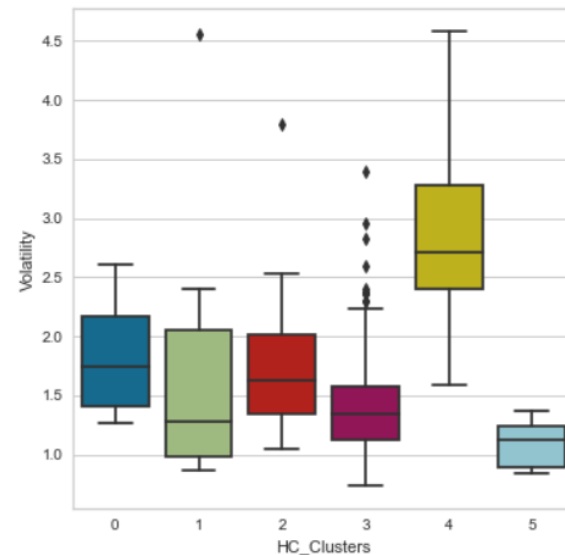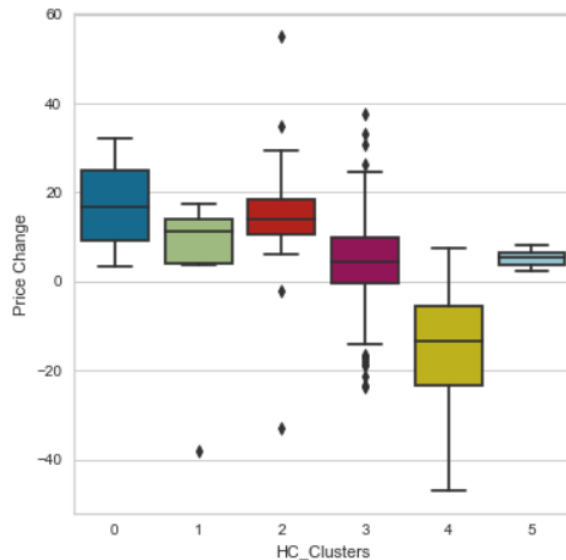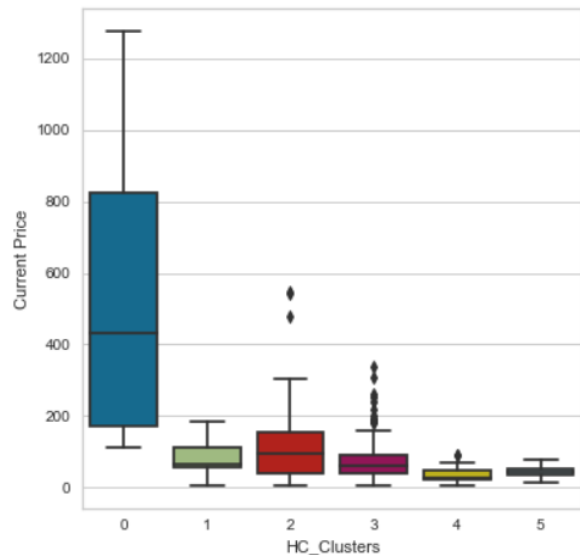
Hierarchal Clustering with 6 Clusters - Visuals

Hierarchal Clustering with 6 Clusters - Insights

Cluster 0 (blue)	Cluster 1 (green)	Cluster 2 (red)
• 4 Securities • Current price midrange is ~190 to 810. It is the portfolio for the priciest stocks. • Price change is positive ranging from ~$3% to 32%. • Volatility is moderate with a midrange between 1.4 and 2.2 • ROE is the lowest of all clusters. • Cash ratio is all positive with a range of ~70 to 200%	• 7 Securities • Current price range is a little over 0 to just under 200. • Price change is consistently positive. • Volatility of the cluster includes the majority of the securities being below 1.5. • ROE is the highest of all clusters with all securities returning over 400%. • Cash ratio on average is lower than all other clusters.	• 23 Securities • Current price contains a portfolio of the second priciest stocks. • Price change is strong with a majority of the underlying securities of the cluster exhibiting double digit returns. • Volatility is moderate with an approximate range of 1 to 2.5. • ROE is low and very similar to clusters 4 and 5. • Cash ratio is strongest for this portfolio with almost 75% of the stocks having cash ratios than the stocks of other portfolios.

Hierarchal Clustering with 6 Clusters - Insights

Cluster 3 (violet)	Cluster 4 (yellow)	Cluster 5 (light blue)
• 275 Securities • Current price contains several outliers driving averages higher. It is the portfolio with the third cheapest stocks. • Price change is a normal distribution with the second largest range with 25% of securities having negative returns. • Volatility for this cluster includes several outliers and contains the second lowest values of all the clusters. • ROE of this portfolio contains a significant amount of outliers with most securities being relatively close to 0%. • Cash ratio contains some outliers but is similar to clusters 4 and 5.	• 22 Securities • Current price is the portfolio with securities with the lowest current prices. • Price change is the worst performing portfolio with over 75% of securities having negative returns. • Volatility for this cluster has the largest range going from a little over 1.5 to above 4.5. • ROE is low and very similar to clusters 2 and 5. • Cash ratio is similar to clusters 3 and 5.	• 9 Securities • Current price is the second lowest portfolio of securities. • Price change is the narrowest of ranges provides a return of approximately 2-10%. • Volatility is the narrowest of ranges and includes all securities being relatively involatile with all securities being belwo 1.5. • ROE is low and very similar to clusters 2 and 4. • Cash ratio is similar to clusters 3 and 4.

K-Means vs Heirarchal Clustering

Trade&Ahead should decide between K-Means Cluster 0 and Hierarchal Cluster 2. These two clusters are properly sized to hedge against risks while small enough to prevent the law of diminishing returns to prevent substantial returns. An ideal portfolio size needed to acheive this is 20-25 securites which both of these clusters achieve at 24 and 23 securities, respectively.

With the current economy preparing for rising interest rates, cash on hand statistics such as Cash Ratio, Net Cash Flow, and Net Income should be considered along with Price Change, Volatility and P/E ratio in determining which portfolio to select.

K-Means Cluster 0	Hierarchal Cluster 2
Cash Ratio: 279.9% Net Cash Flow: $459M Net Income: $1B Price Change: 14.28% Volatility: 1.77 P/E Ratio: 90.1%	Cash Ratio: 125.5% Net Cash Flow: $106M Net Income: $854M Price Change: 17.2% Volatility: 1.84 P/E Ratio: 307.1%

Business Recommendations

Although Hierarchal Cluster 2 provides a better historical return, Trade&Ahead should choose K-Means cluster 0 as its portfolio due to is better cask flow and lower volatility will better prepare Trade&Ahead to weather rising interest rates and inflation while still gaining significant annual returns.