Trade & Ahead
Business Presentation

Contents

The presentation consists of three Four Sections:

1. Background and Business Problem Overview
2. Data Overview
3. Exploratory Data Analysis
4. K-Means Clustering
5. Heirarchical Clustering
6. Business recommendations

Background of the Business and Objective

Background

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

To analyze the data and grouping the stocks based on the attributes provided

Provide insights about the characteristics of each group.

Data Overview

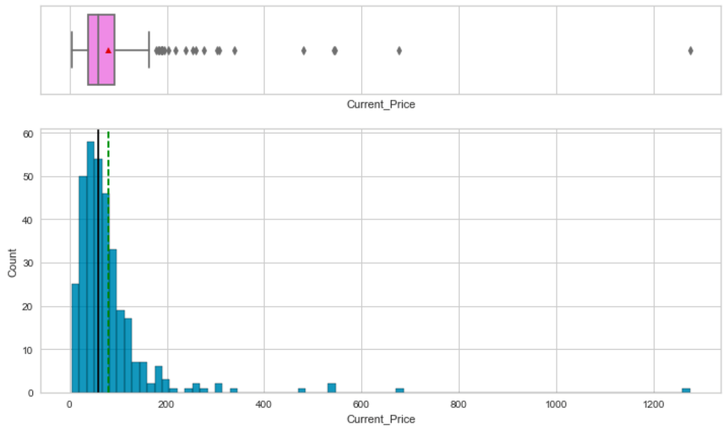
Ticker symbol	An abbreviation used to uniquely identify publicly traded shares of a particular stock on a particular stock market
Company	Name of the company
GICS Sector	The specific economic sector assigned to a company by the Global Industry Classification Standard (GICS) that best defines its business 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)

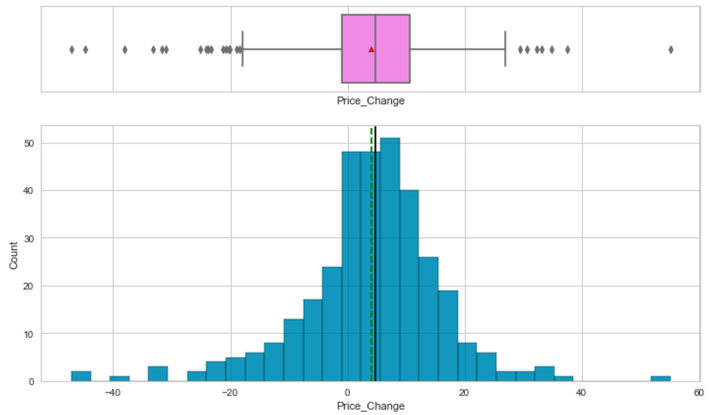
The data provided is of stock prices and some financial indicators like ROE, earnings per share, P/E ratio, etc.

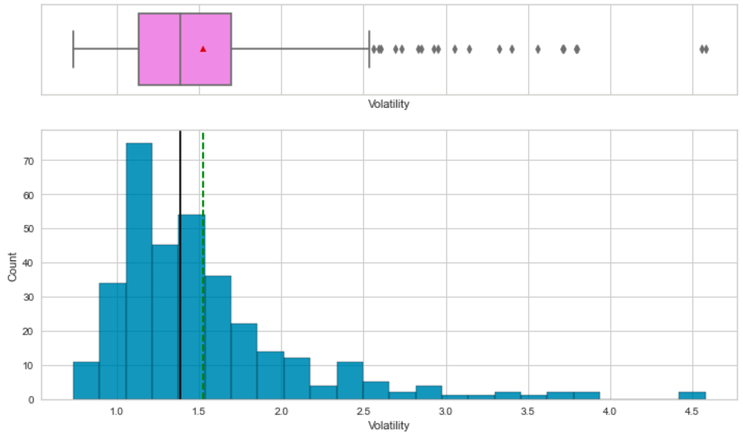
340 observations and 15 columns in the dataset
No missing values in the data
No Duplicate values
Columns: float64, int64, and object datatype

EDA

- Univariate data analysis:





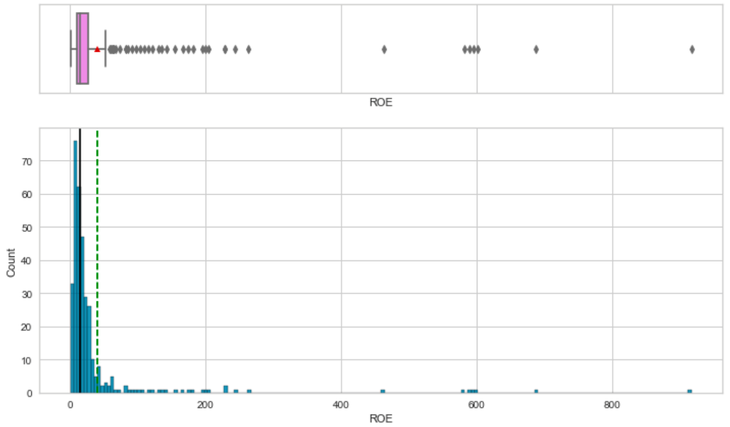


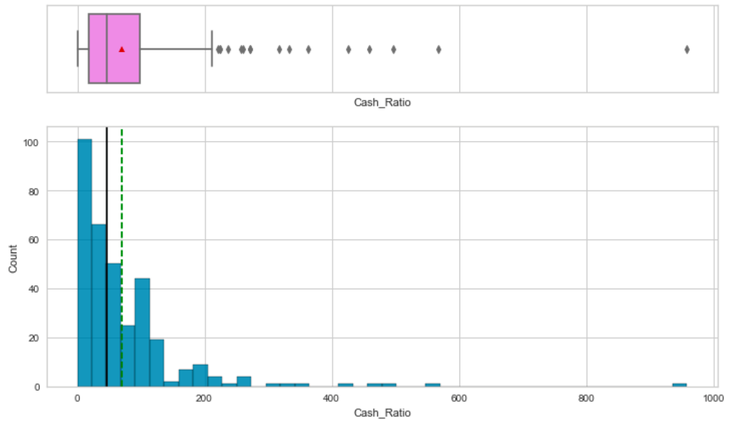
Observations:

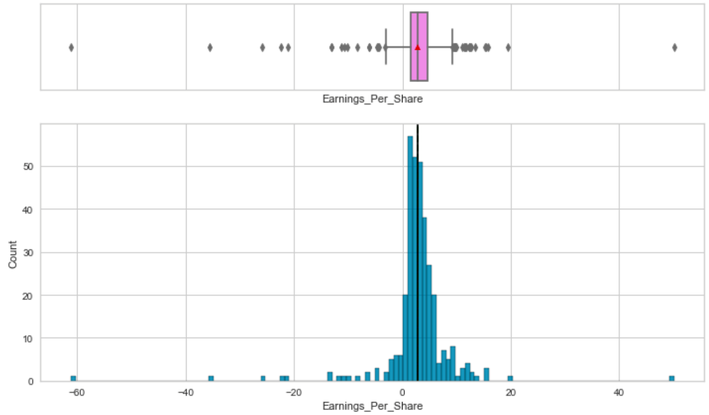
- There is right skewness in Current_Price and Volatility
- Price_change is slightly left skewed
- Outliers are present in the data

EDA

- Univariate data analysis:





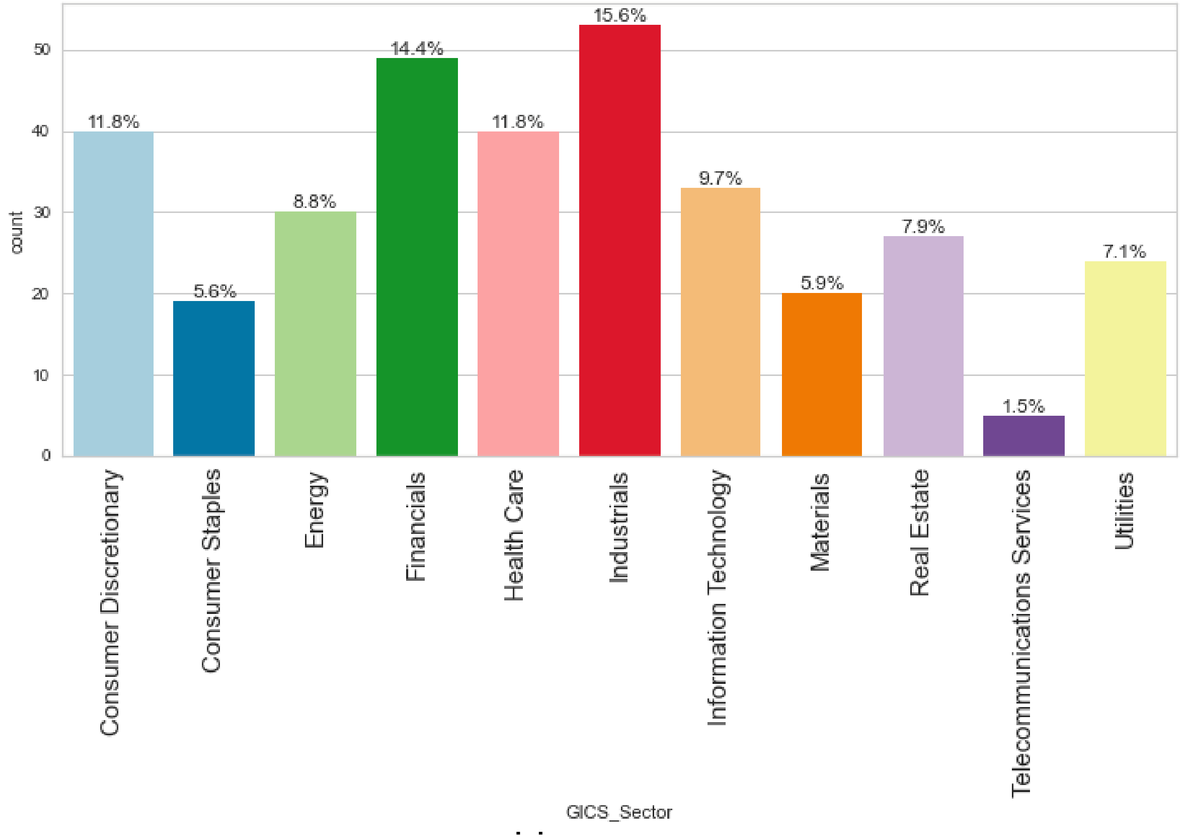


Observations:

- There is right skewness in ROE and Cash_Ratio

EDA

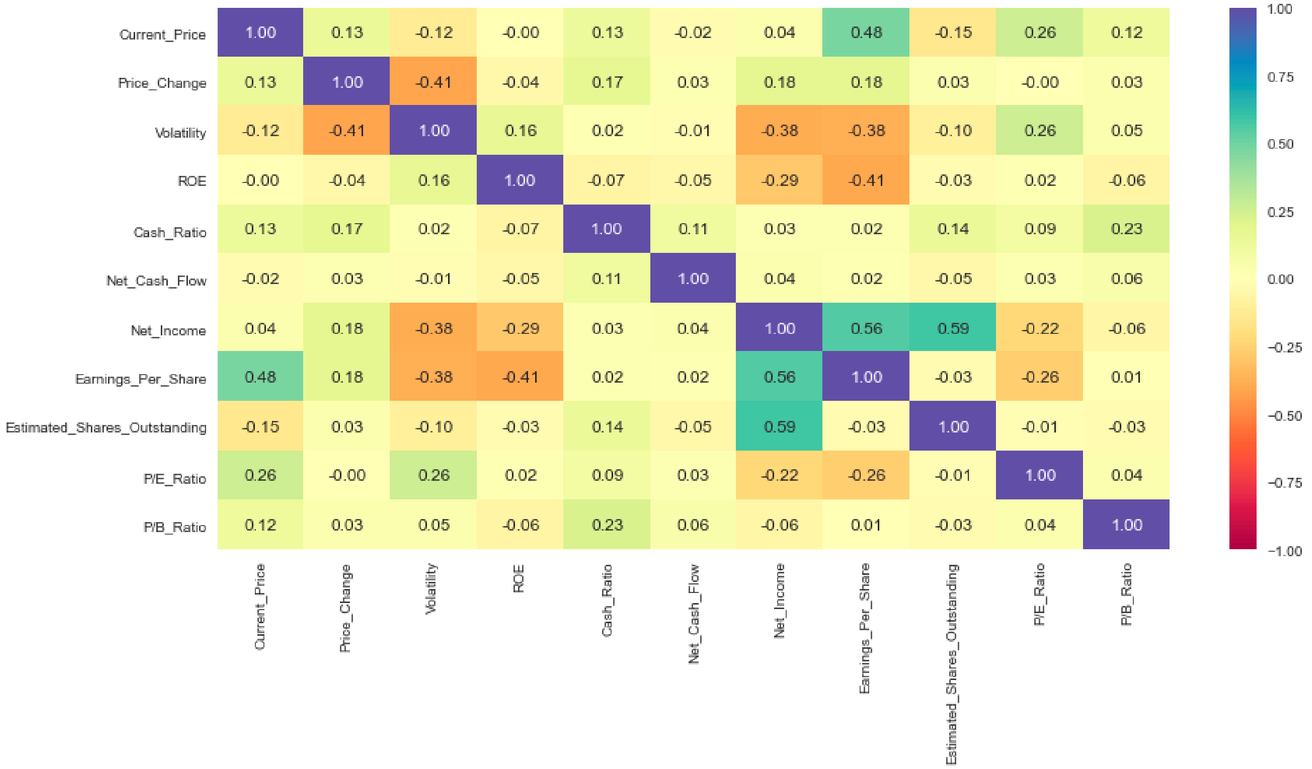
Univariate data analysis



- Most stocks are from Industrial sector followed by Financials.
- Consumer staples and utilities stocks are lowest in data count.

EDA

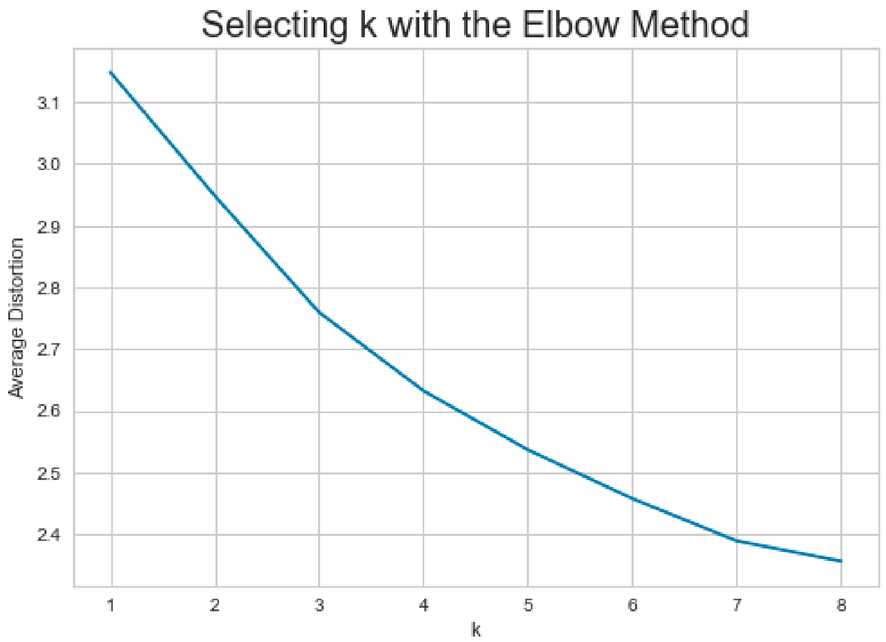
Bivariate data analysis

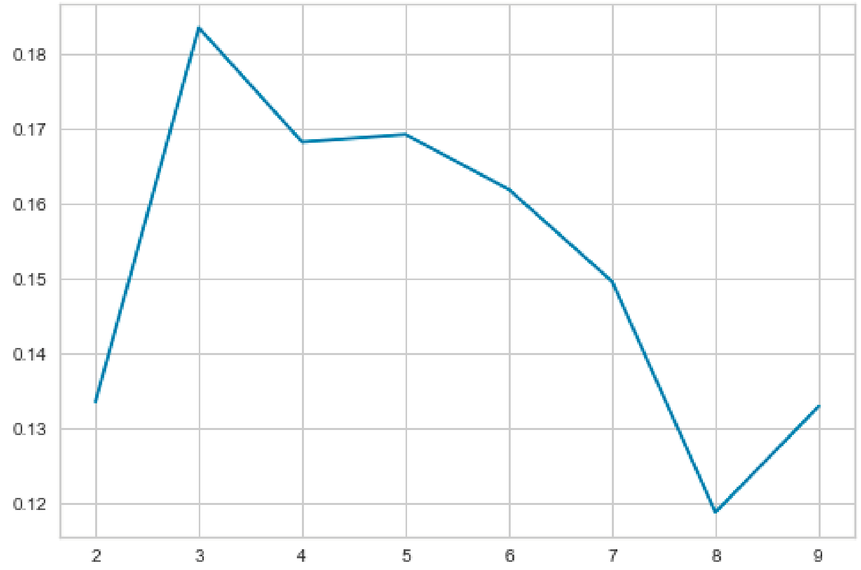


Observations:

- No significant correlation is observed between the variables.

K-Means Clustering





From the silhouette score and the elbow, it seems that 5 is a good value of k

Cluster profile

- **Cluster 0 38**
- **Cluster 1 138**
- **Cluster 2 28**
- **Cluster 3 73**
- **Cluster 4 63**

K-Means Clustering Observations

Cluster 0:
- Stocks current price are high.
- price change for these stocks in cluster 0 are highest.
- Stocks in cluster 0 are moderately volatile.
- ROE for stocks in cluster is moderately high.
- Cash ratio for the stocks in the cluster 0 are highest.

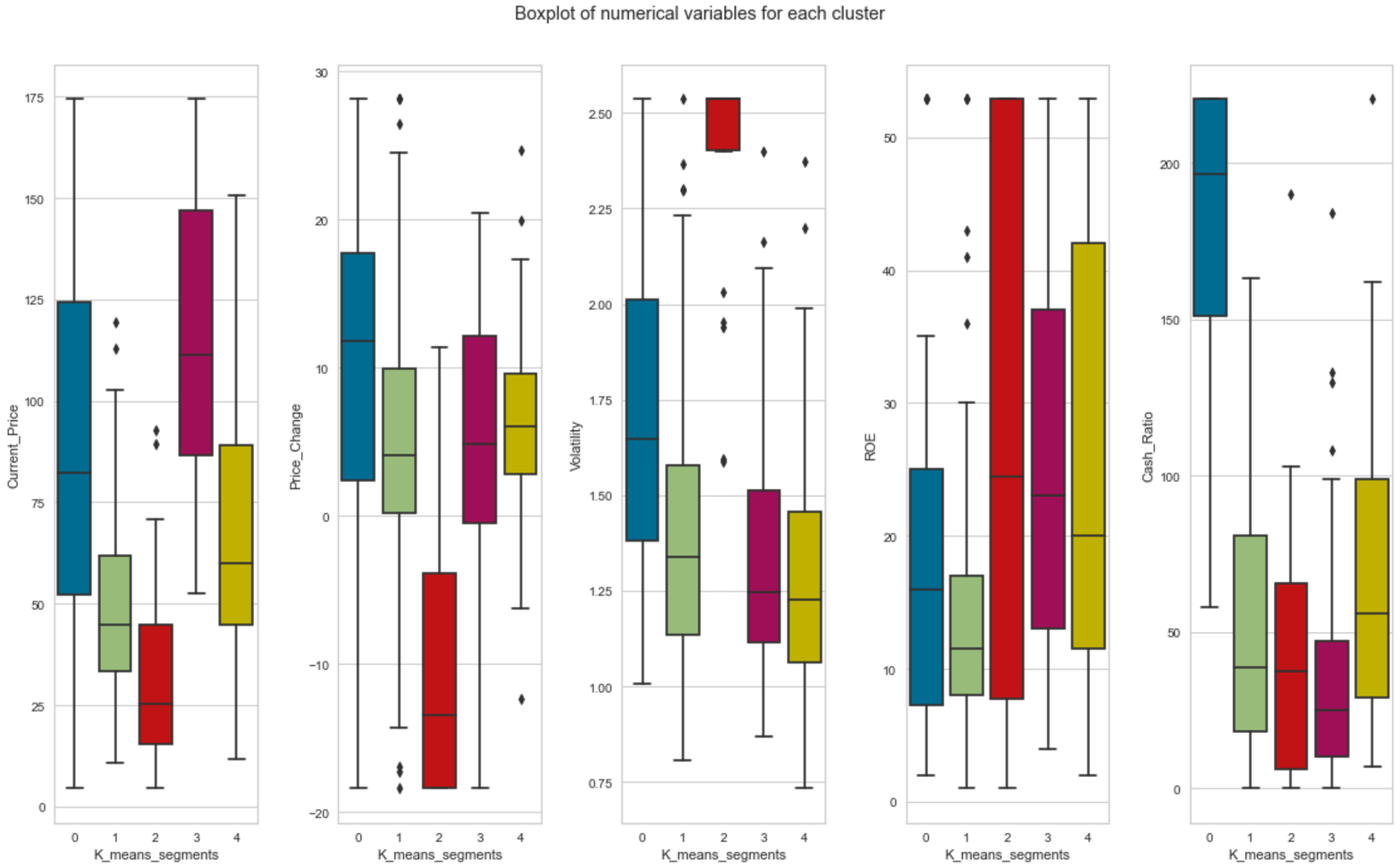
Cluster 1:
- Stocks current price are slightly lower.
- price change of the stocks in this cluster range from negative to positive.
- Stocks in cluster 1 are less volatile than cluster 0.
- ROE for these stocks in this cluster are highest.
- Cash ratio ranges from moderate to high for these stocks.

Cluster 2:
- Stocks current price are lower than any other stocks in any clusters.
- price change of these stocks are negative.
- Stocks in cluster 2 are highly volatile.
- ROE for stocks in cluster is highest.
- Cash ratio for these stocks ranges from low to moderately high.

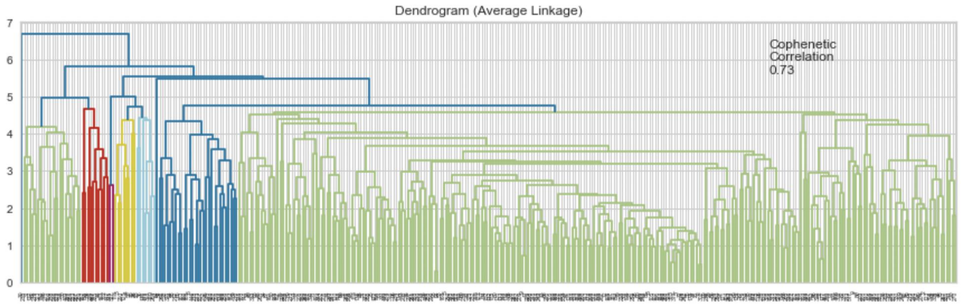
Cluster 3:
- Stocks current price are in highest in the range.
- price change of these stocks are negative to low.
- Stocks in cluster 3 are less volatile than cluster 2.
- ROE for stocks in cluster 3 is positive.
- Cash ratio for these stocks are lowest among all the clusters.

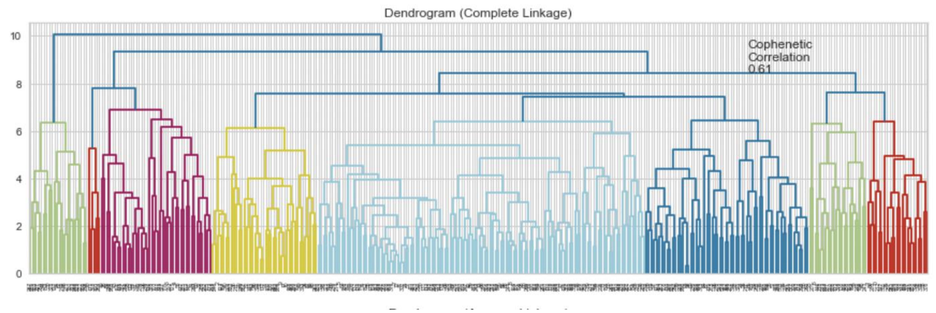
Cluster 4:
- Stocks current price are in the medium range.
- price change of these stocks are positive.
- Stocks in cluster 4 are moderately volatile.
- ROE for stocks in cluster 4 is high.
- Cash ratio for these stocks second highest.

K-Means Clustering



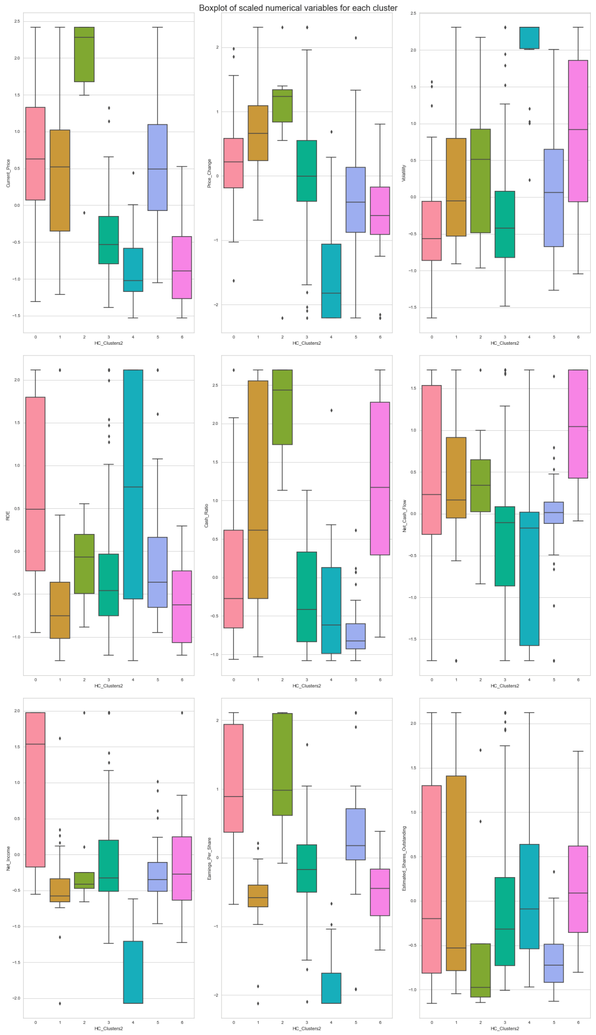
Hierarchical Clustering



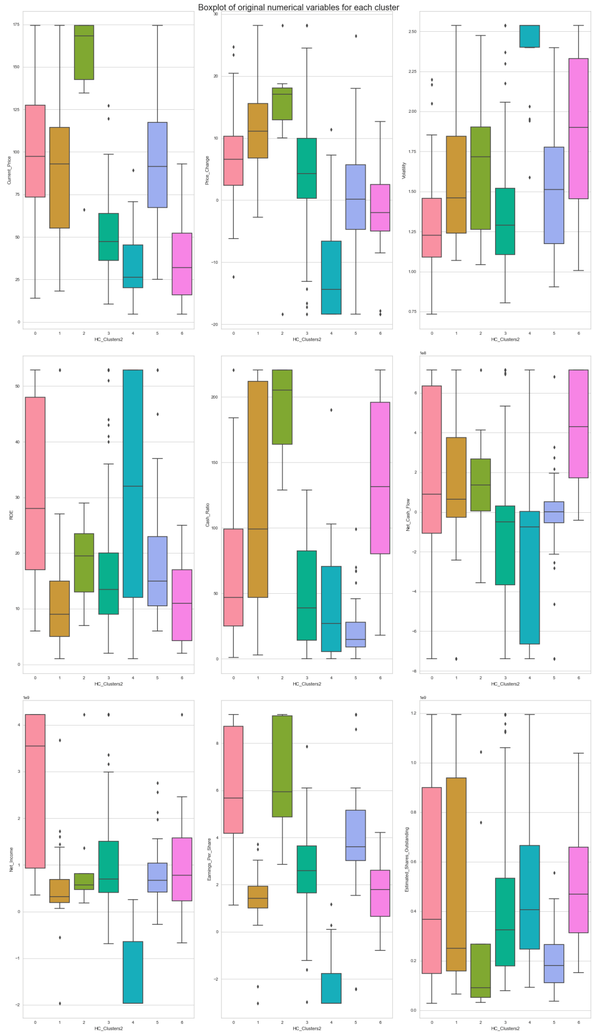


- **Highest cophenetic correlation is 0.7325610568988988, which is obtained with Euclidean distance and average linkage. However, the clusters have less variability**
- **Euclidean distance and complete linkage provides clusters with more variability**
- **From the dendrograms, we deduced that 7 is the appropriate number of clusters for complete linkage**

Heirarchical Clustering



Scaled Data



Original Data

Hierarchical Clustering Observations

Using Agglomerative Clustering clustering technique with Euclidean affinity and complete linkage we observed that.

- - In total 7 cluster created highest distribution of stocks happened in the cluster 3 which are 150 followed by cluster 0 which are 73.
- - Stock prices are highest for the stocks in the cluster 2, cluster 0 and cluster 1.
- - Stock prices are lowest for the stocks in the cluster3 and cluster4.
- - Highest price change for the 13-week duration happened for the stocks in cluster 2.
- - Cluster 4 and cluster 6 had negative price change for the stocks.
- - Stocks in cluster 4 has highest volatility which means they are most risky stocks.
- - Cluster4 has highest ROE observed.
- - Stocks in cluster 2 has highest cash ratio.
- - Stocks in cluster 6 been observed with highest net cash flow.
- - Stocks in cluster 0 has highest net income.
- - Stocks in cluster 2 has highest Earnings per share.
- - Stocks in cluster 4 has highest P/E ratio.

Findings and Recommendations

Based on k-means and hierarchical techniques, we conclude that:

- **Cluster 2,Cluster 0 and Cluster 1 stock prices are on the higher side which are very expensive stocks.**
- **Cluster 2 stocks price change is positive which shows stability within these stocks. The price change is high but provide stable rise over 13 weeks.**
- **Cluster 2 stocks are less volatile than any other cluster which means there is less risk involved for the investors.**
- **ROE for stocks in cluster 2 is low to moderate which shows solid growth among these companies.**
- **Cash ratio for cluster 2 is very high with median value of 220 which provides a measure of companies' ability to cover its short-term obligation.**
- **Cluster 2 Net cash flow is the second best among all the clusters which shows growth among these companies.**
- **Cluster 2 stocks have the highest earnings per share which shows company's profitability and is one of the most popular metrics that analysts refer to when evaluating stocks.**
- **Cluster 2 stocks have a high P/E ratio, which helps in determining the relative value of a company's shares as they signify the amount of money an investor is willing to invest in a single share of a company per dollar of its earnings**
- **Based upon these findings we recommend our customers to add the stocks from cluster 2 into their portfolio for long term growth potential. Cluster 2 has a mixed bag of 8 stocks.**

greatlearning
Power Ahead

Happy Learning !