greatlearning
Power Ahead

# Trade & Ahead Project

# 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.

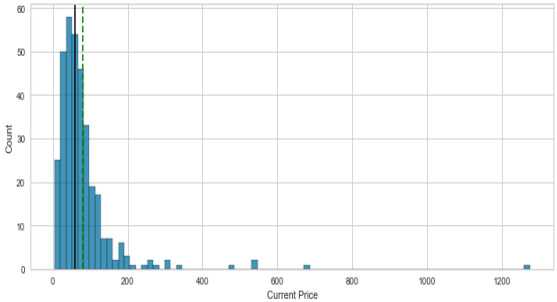
# Objective

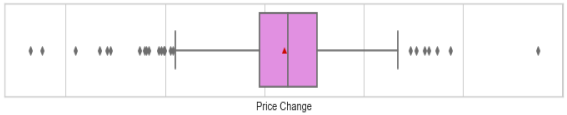
Trade&Ahead is a financial consultancy firm who provide their customers with personalized investment strategies. They have hired you as a Data Scientist and provided you with data comprising stock price and some financial indicators for a few companies listed under the New York Stock Exchange. They have 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.

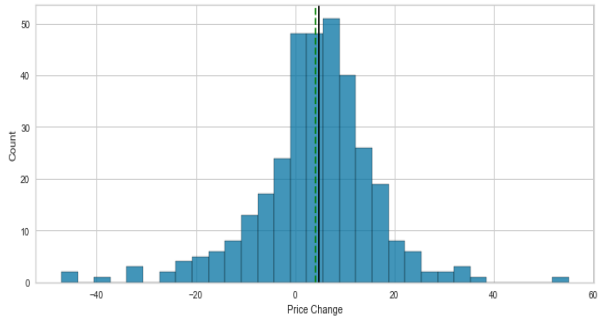
# Data Information

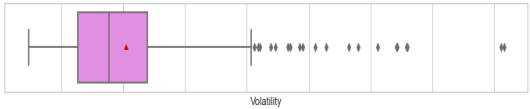
| Variable | Description |
|---|---|
| Ticker Symbol | An abbreviation used to uniquely identify publicly traded shares of a particular stock |
| Company | Name of the company |
| GICS Sector | The specific economic sector assigned to a company by the Global Industry Classification Standard |
| GICS Sub Industry | The specific sub-industry group assigned to a company by the Global Industry Classification Standard |
| 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 |
| 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) |

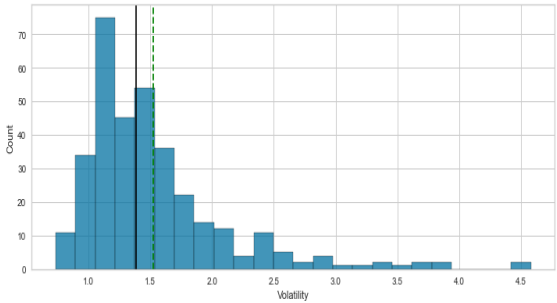
# Exploratory Data Analysis

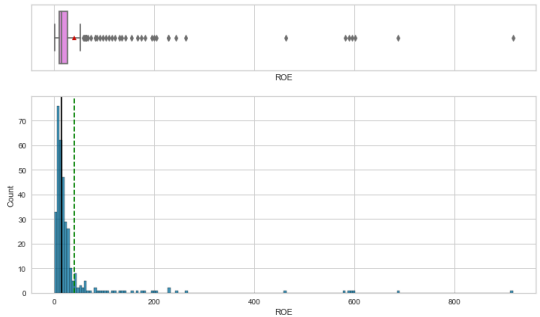



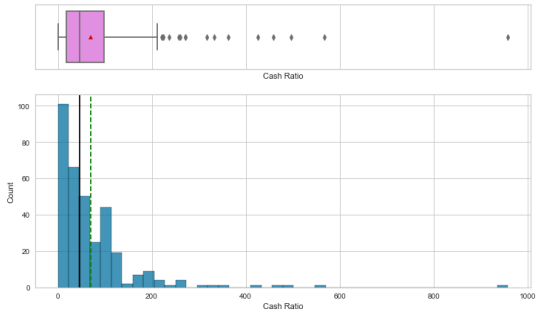



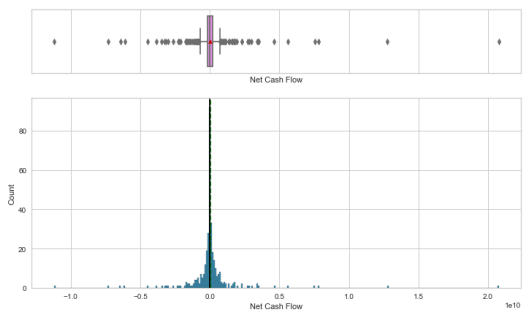



# Exploratory Data Analysis



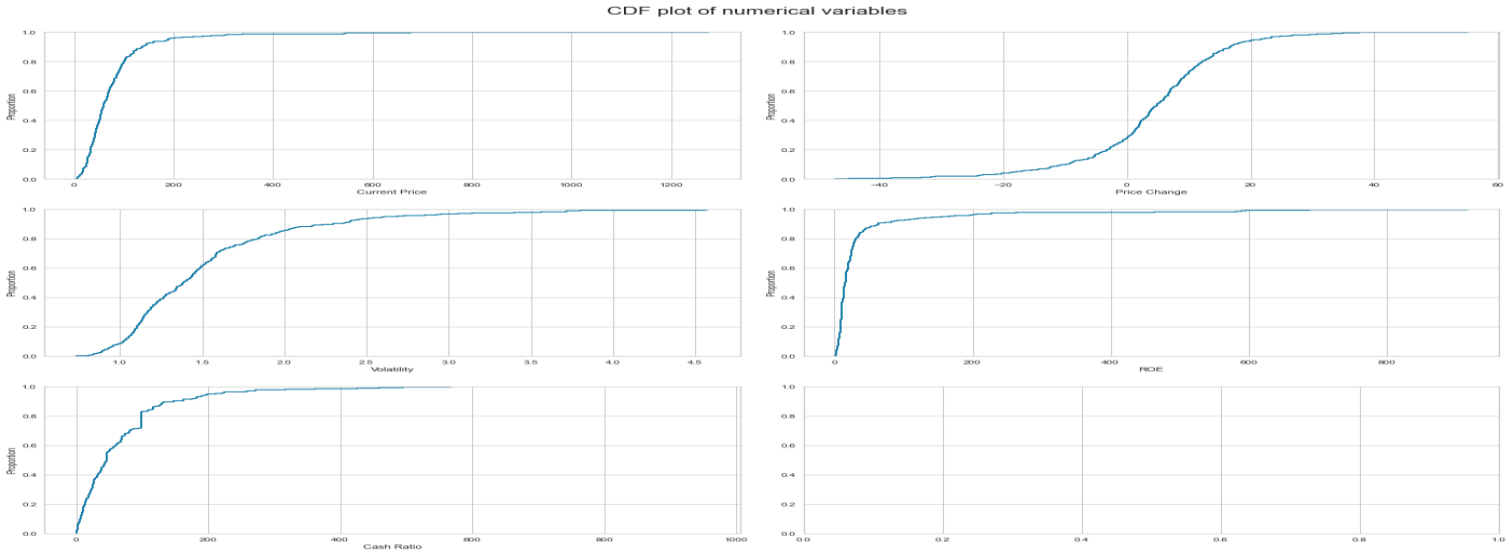




**Observations**;

- Current Price have right-skewed distributions with upper outliers.
- The volatility is rightly skewed with an average of 1.5.

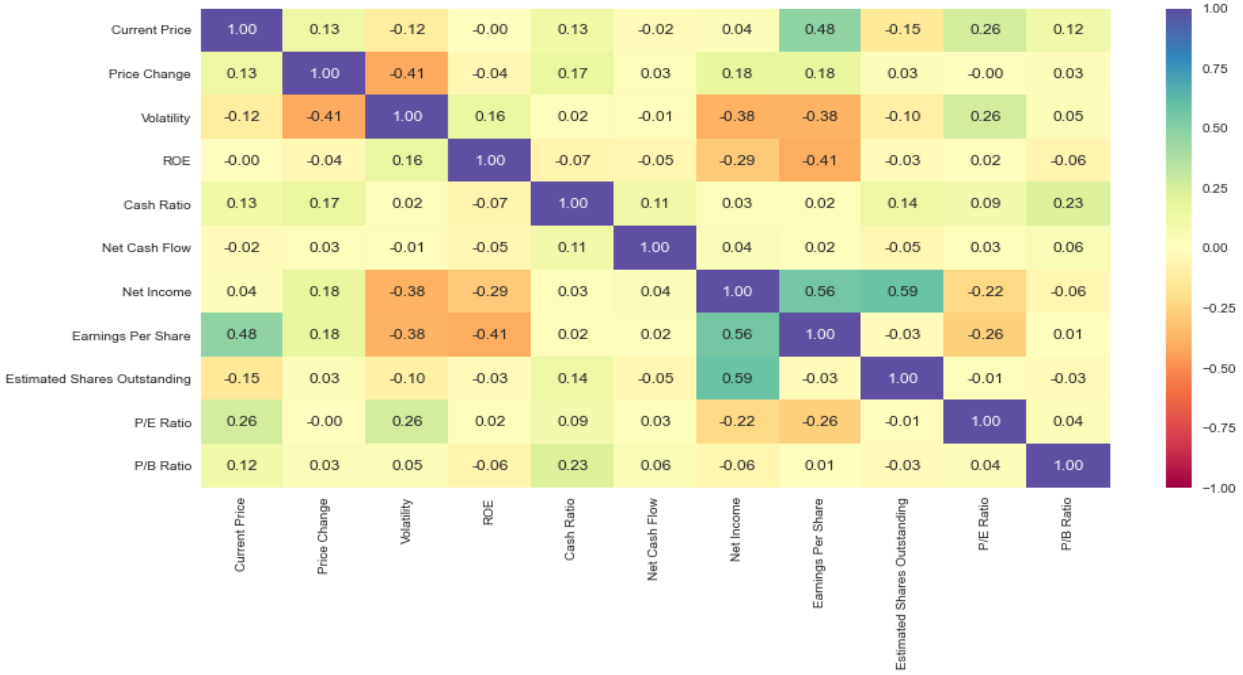
# Exploratory Data Analysis




**Observations**

- 90% of the stocks have current prices less than 200.
- 50% of the stocks have a price change of 5 unit .
- 50% of the stocks have a volatility of 25 leaning to the right.
- 50% off the stocks have a ROE of 400.
- Cash ratio for the stocks ranges from 100-500

# Exploratory Data Analysis

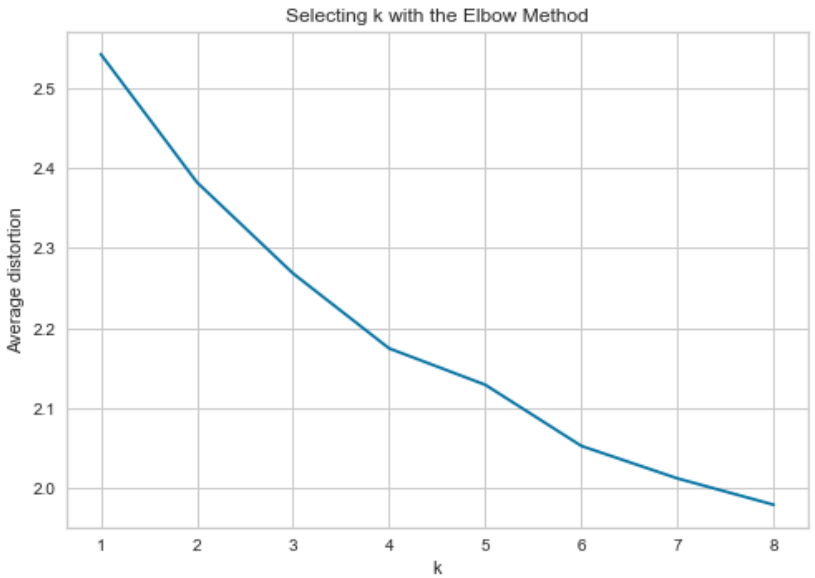



**Observations**;

- The current price and the cash ratio are highly positively correlated, which is obvious.
- Cash Ratio is somewhat negatively correlated with ROE.

.

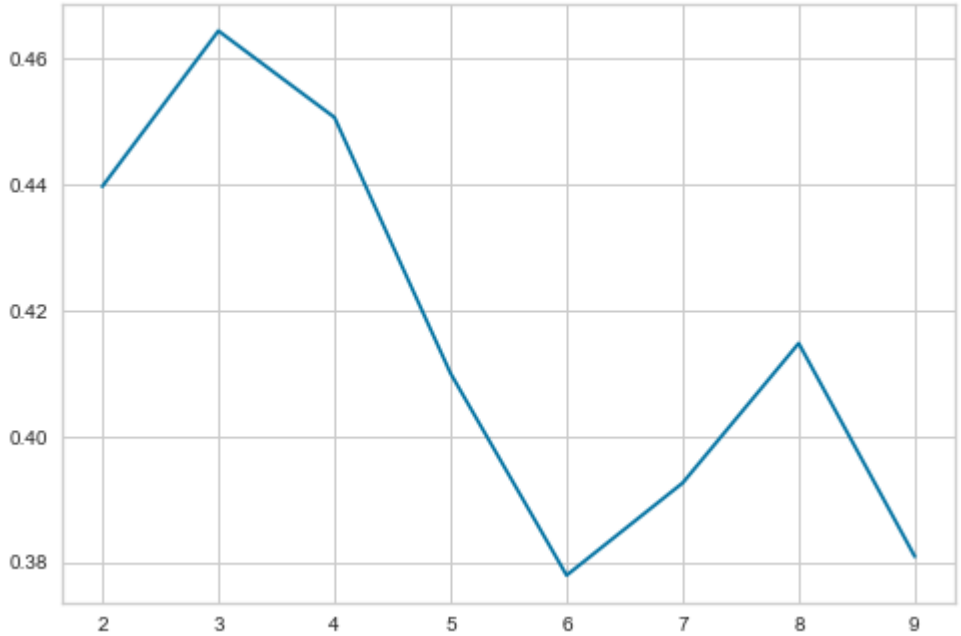
# K MEANS CLUSTERING



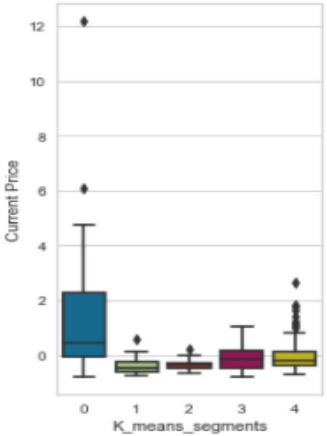

```
Number of Clusters: 1   Average Distortion: 2.5425069919221697
Number of Clusters: 2   Average Distortion: 2.382318498894466
Number of Clusters: 3   Average Distortion: 2.2683105560042285
Number of Clusters: 4   Average Distortion: 2.1745559827866363
Number of Clusters: 5   Average Distortion: 2.129043494736725
Number of Clusters: 6   Average Distortion: 2.052570356366889
Number of Clusters: 7   Average Distortion: 2.0119926937778194
Number of Clusters: 8   Average Distortion: 1.9791030728404386
```

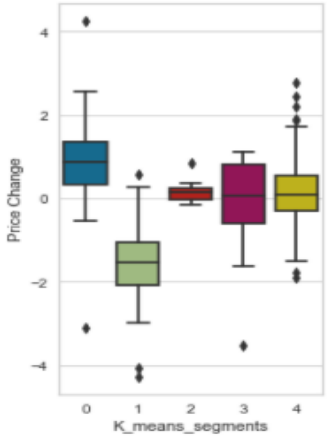
# CLUSTERING

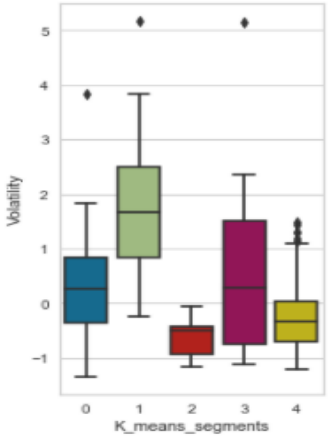
```
For n_clusters = 2, the silhouette score is 0.43969639509980457)
For n_clusters = 3, the silhouette score is 0.4644405674779404)
For n_clusters = 4, the silhouette score is 0.4506868801070228)
For n_clusters = 5, the silhouette score is 0.40999356683171667)
For n_clusters = 6, the silhouette score is 0.3778823699608175)
For n_clusters = 7, the silhouette score is 0.3925655757490979)
For n_clusters = 8, the silhouette score is 0.414772888480993)
For n_clusters = 9, the silhouette score is 0.38086668239902466)
```

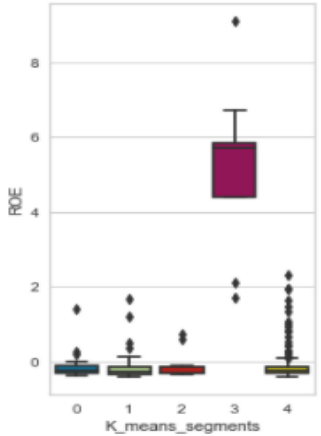


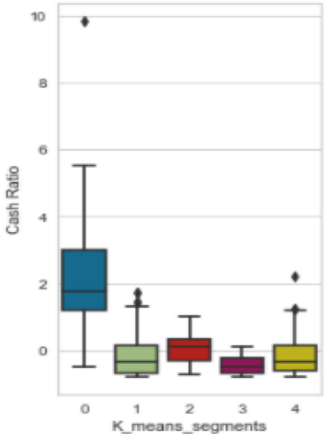
# Boxplot of Scaled Numerical Variable for Each Cluster



# Linkage methods with Euclidean distance and Cophenetic Correlation

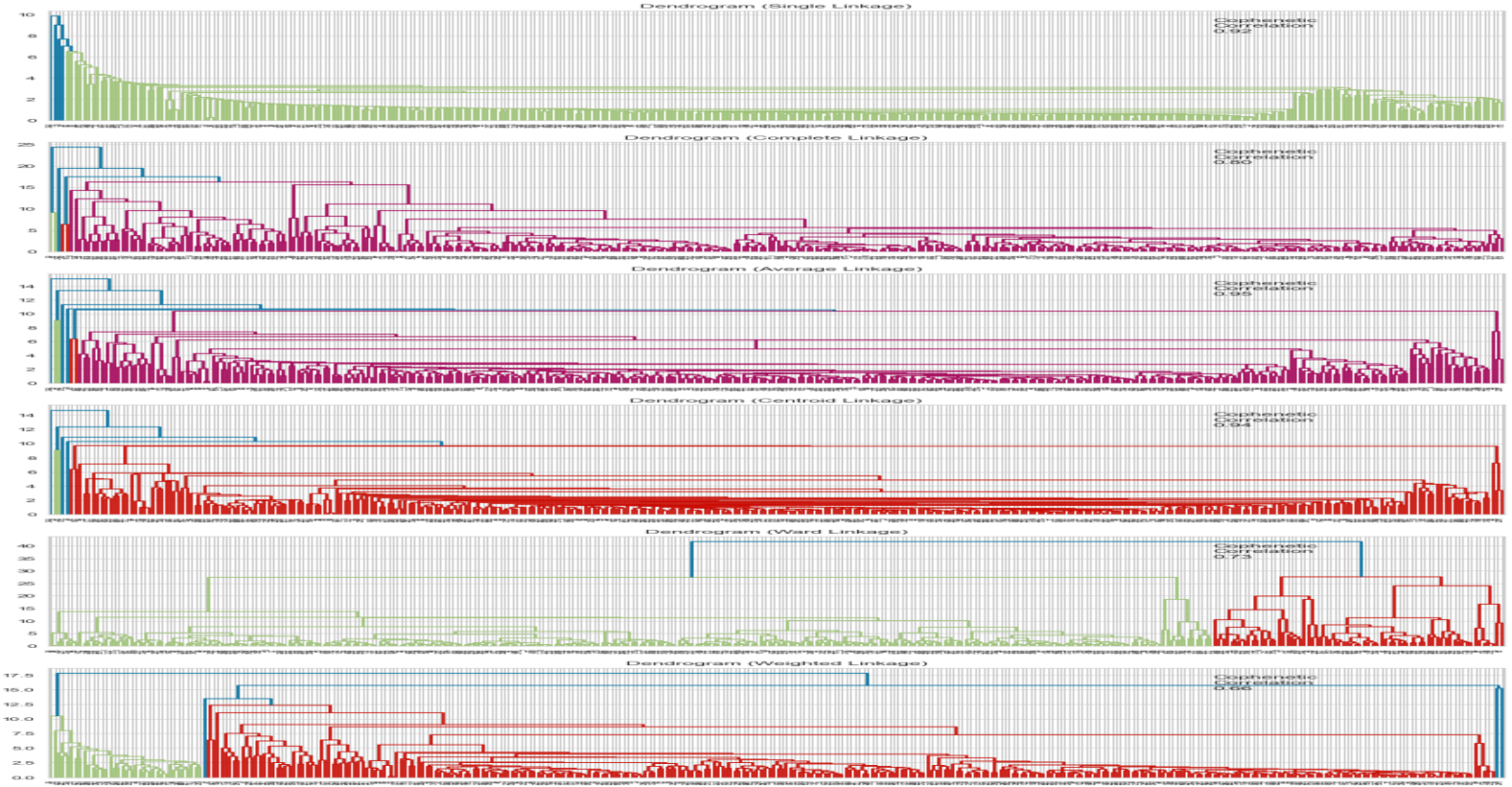
```
Cophenetic correlation for Euclidean distance and single linkage is 0.9245196884893159.
Cophenetic correlation for Euclidean distance and complete linkage is 0.804706102299582.
Cophenetic correlation for Euclidean distance and average linkage is 0.9453100811955032.
Cophenetic correlation for Euclidean distance and weighted linkage is 0.6627654898087273.
Cophenetic correlation for Chebyshev distance and single linkage is 0.9167442245950579.
Cophenetic correlation for Chebyshev distance and complete linkage is 0.8147175389813458.
Cophenetic correlation for Chebyshev distance and average linkage is 0.9374309398424928.
Cophenetic correlation for Chebyshev distance and weighted linkage is 0.909227804034383.
Cophenetic correlation for Mahalanobis distance and single linkage is 0.9349160370798356.
Cophenetic correlation for Mahalanobis distance and complete linkage is 0.8420923890794115.
Cophenetic correlation for Mahalanobis distance and average linkage is 0.9346420885297709.
Cophenetic correlation for Mahalanobis distance and weighted linkage is 0.8761020180095022.
Cophenetic correlation for Cityblock distance and single linkage is 0.9377940959076201.
Cophenetic correlation for Cityblock distance and complete linkage is 0.7819823728539451.
Cophenetic correlation for Cityblock distance and average linkage is 0.9278890656868077.
Cophenetic correlation for Cityblock distance and weighted linkage is 0.6579436145039212.
```

```
Highest cophenetic correlation is 0.9453100811955032, which is obtained with Euclidean distance and average linkage.
```

```
Cophenetic correlation for single linkage is 0.9245196884893159.
Cophenetic correlation for complete linkage is 0.804706102299582.
Cophenetic correlation for average linkage is 0.9453100811955032.
Cophenetic correlation for centroid linkage is 0.9426384108485938.
Cophenetic correlation for ward linkage is 0.7287469098815194.
Cophenetic correlation for weighted linkage is 0.6627654898087273.
```

```
Highest cophenetic correlation is 0.9453100811955032, which is obtained with average linkage.
```

# Visualizing the Decision Tree



# Insight

**Cluster 0:**

- The average current price change of the stocks is 78
- The price change for this particular  stocks is 4.1
- The earnings per share of this stock is 2.9 per share
- The average cash ratio for the stocks is 66.8

**Cluster 1:**

- The average current price change of the stocks is 26
- The Volatility for this particular  stocks is 1.3
- The earnings per share of this stock is 3.3 per share
- The average cash ratio for the stocks is 130

**Cluster 2:**

- The current price of the stock in this cluster is similar  to those in Cluster 1, but the price change is comparatively lower (like negative)
- On average, the Estimated Shares Outstanding is 519,573,983.00 while the P/E Ration is 61

**Cluster 3:**

- The current price in the cluster is very high compared to the rest of the clusters while the price change are really low.
- The volatility for this cluster is like 1.3

**Cluster 4:**

- The Net Income in this cluster is 3,669,000,000.00 while the P/B Ratio and P/E Ration is 78 and 6 respectively
- The cash ration in this cluster is 958 while the Net Cash Flow is 592,000,000.00.

# Recommendations

- Thus, dividing stocks into groups with "similar characteristics" can help in portfolio construction to ensure we choose a collection of stocks with sufficient diversification between them.
- Investors who are more concern with downside risk and diversified portfolio might maximize their ROE
- The modern portfolio theory can be useful to investors trying to construct efficient and diversified portolio

greatlearning
*Power Ahead*

# Happy Learning !