Business Presentation Trade & Ahead

Contents

- Problem Definition
- Key findings and Insights
- Solution
- Potential benefits of implementing solution

Business Problem Overview and Solution Approach

It is important to maintain a diversified portfolio when investing in stocks in order to maximise 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 which 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.

Trade&Ahead a financial consultancy firm provide their customers with personalized investment strategies. They want to know the analysis of the data provided, how the stocks will be grouped based on the attributes provided, and insights about the characteristics of each group.

Business Problem Overview and Solution Approach

We will be majorly focusing on

- Analyzing the data provided
- Grouping the stock based on the attributes provided
- Providing insights on the characteristics of each group based on their cluster

Data Overview

Rows	Columns
340	15

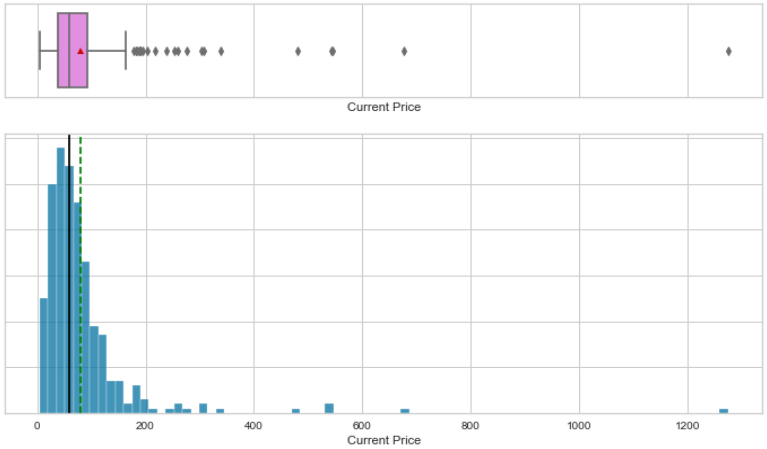
Note:

- There are no missing values in the dataset.
- There are no duplicate entries in the dataset

```
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Ticker Symbol               340 non-null    object
 1   Security                    340 non-null    object
 2   GICS Sector                 340 non-null    object
 3   GICS Sub Industry           340 non-null    object
 4   Current Price               340 non-null    float64
 5   Price Change                340 non-null    float64
 6   Volatility                  340 non-null    float64
 7   ROE                         340 non-null    int64
 8   Cash Ratio                  340 non-null    int64
 9   Net Cash Flow               340 non-null    int64
 10  Net Income                  340 non-null    int64
 11  Earnings Per Share          340 non-null    float64
 12  Estimated Shares Outstanding  340 non-null    float64
 13  P/E Ratio                   340 non-null    float64
 14  P/B Ratio                   340 non-null    float64
dtypes: float64(7), int64(4), object(4)
```

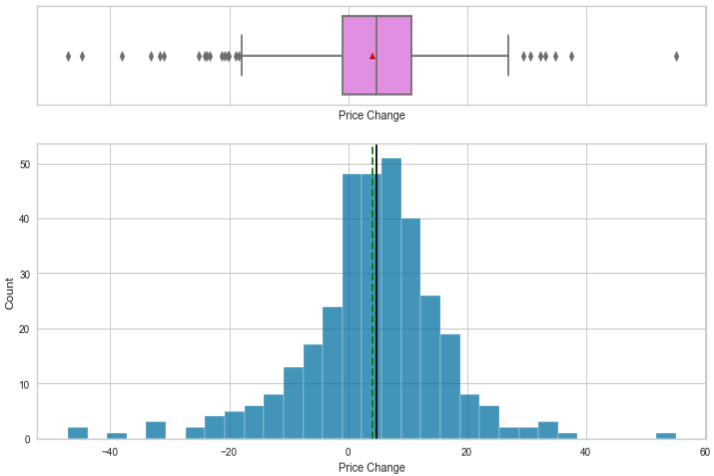
Univarate Analysis: Histogram and Box plots

- Current Price



- Current price is slightly right skewed which means some prices are more than $59
- Mean current prices is around $80

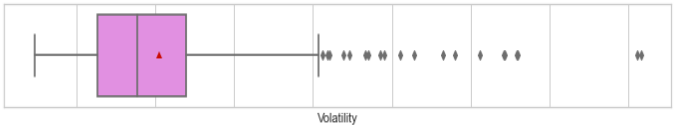
- Price Change

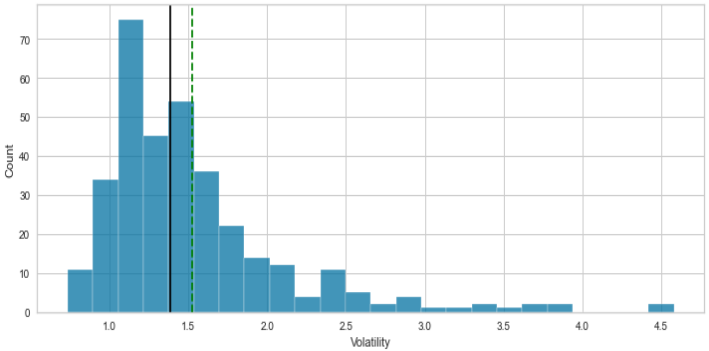


- The distribution appears normally distributed
- With mean and median price change closely aligned to each other
- There are outliers on both the right and left skew

Univarate Analysis: Histogram and Box plots

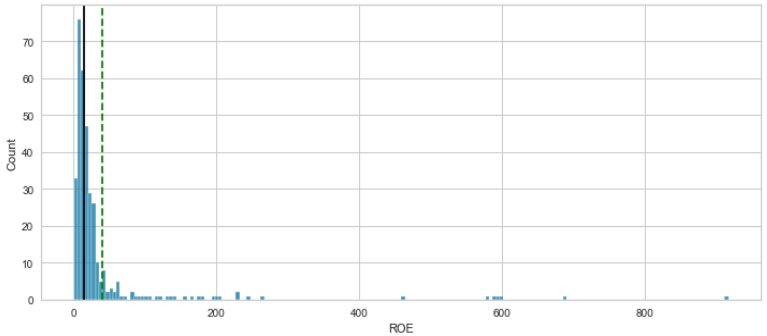
- Volatility





- The distribution is right skewed
- There are some outliers where volatility is more than 2.5
- The mean volatility is about 1.5

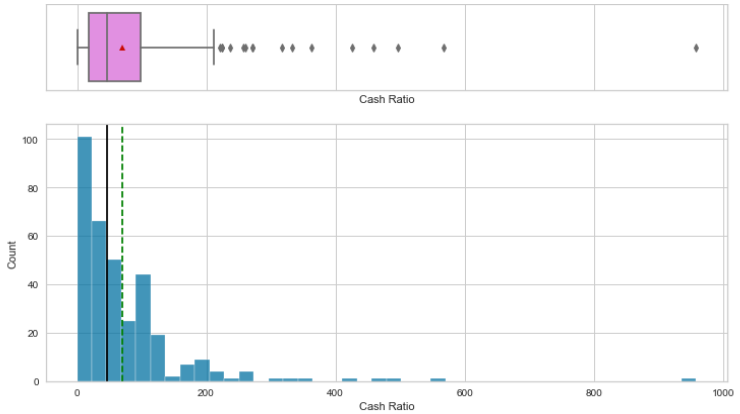
- Return on Equity (ROE)



- There are significant outliers in the data
- The mean Return on Equity of about 39

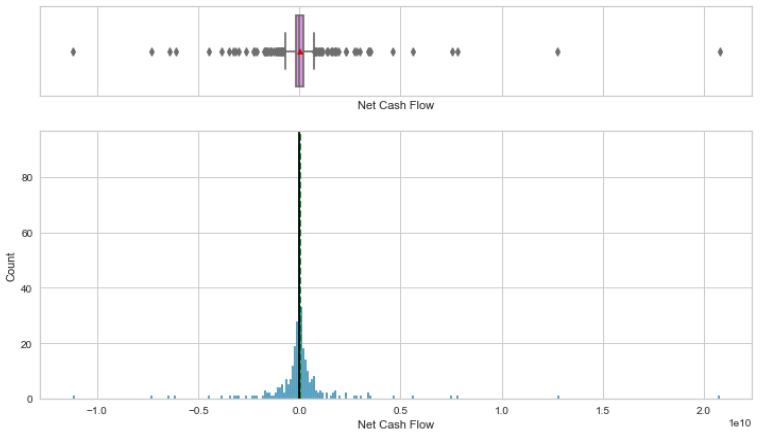
Univarate Analysis: Histogram and Box plots

- Cash Ratio



- The distribution is right skewed
- There are some outliers where cash ratio is more than 70
- The mean cash ratio is about 70

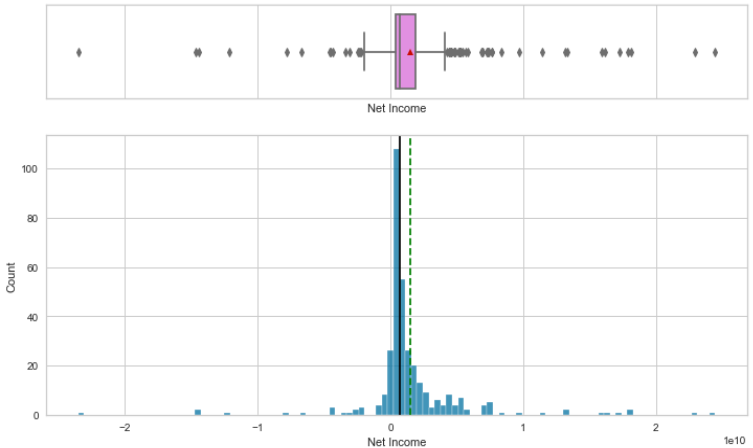
- Net Cash Flow



- The distribution is normally distributed
- The mean and median values in the graph appear close to each other

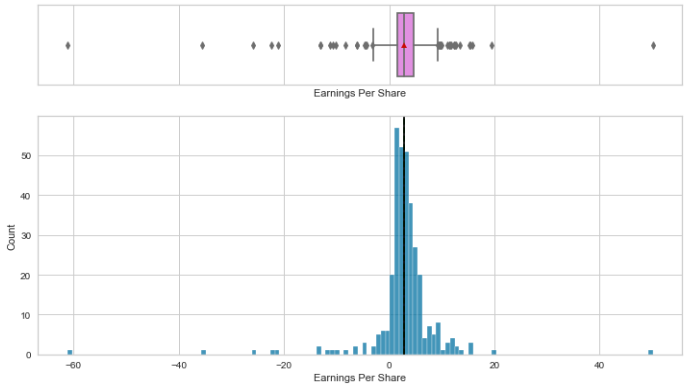
Univarate Analysis: Histogram and Box plots

- Net Income



- The distribution is right skewed with a few outliers
- The mean is greater than the median figure
- There are some outliers with negative net income

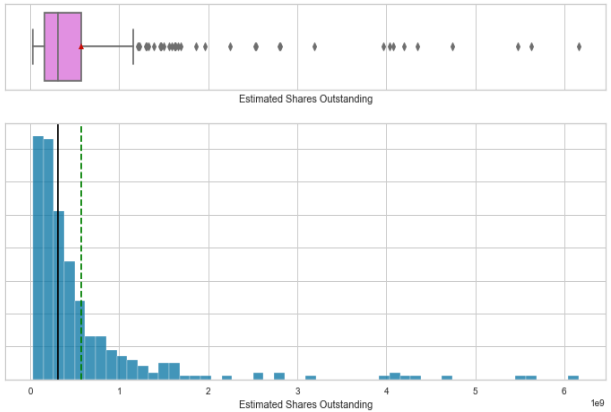
- Earnings Per Share



- The distribution appears normally distributed
- The mean and median figures being fairly close to each other
- There are some observations with negative earnings per share
- There is an extreme value with earning per share less than -40

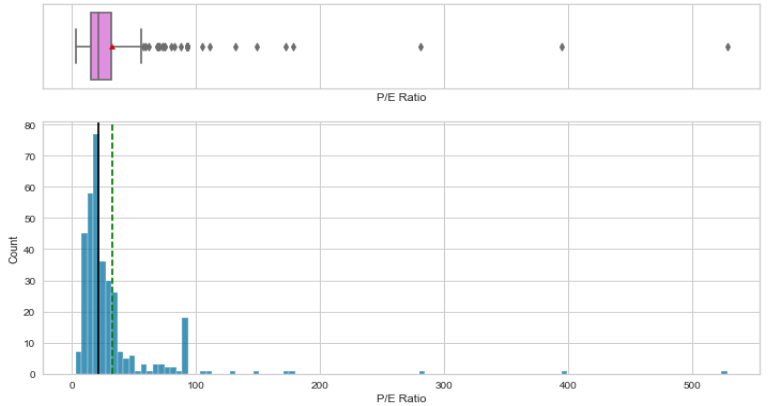
Univarate Analysis: Histogram and Box plots

- Estimated Shares Outstanding



- The distribution is slightly left skewed
- There are a few companies with estimated shares outstanding

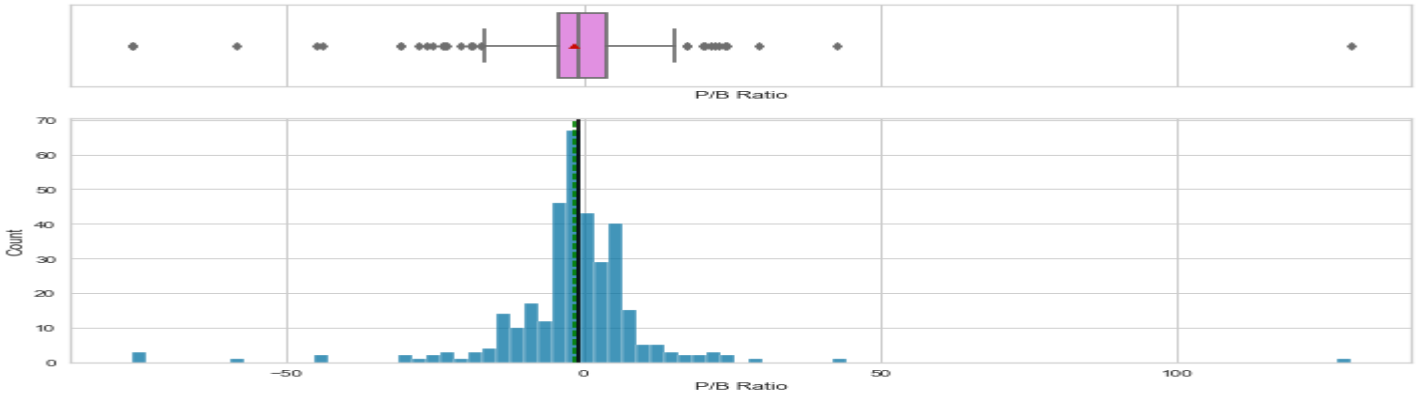
- Price Earning Ration(P/E Ratio)



- The distribution is slightly left skewed
- It has three outliers with high P/E ratio

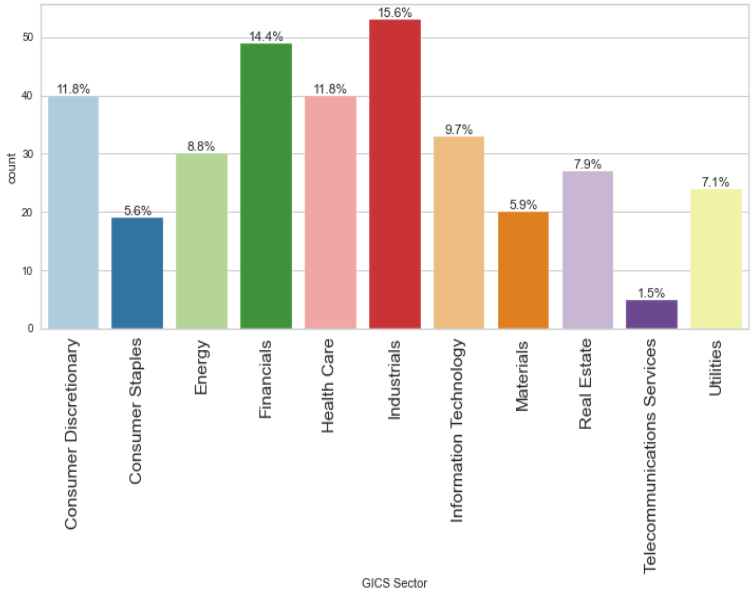
Univarate Analysis: Histogram and Box plots

- Price Book Ratio



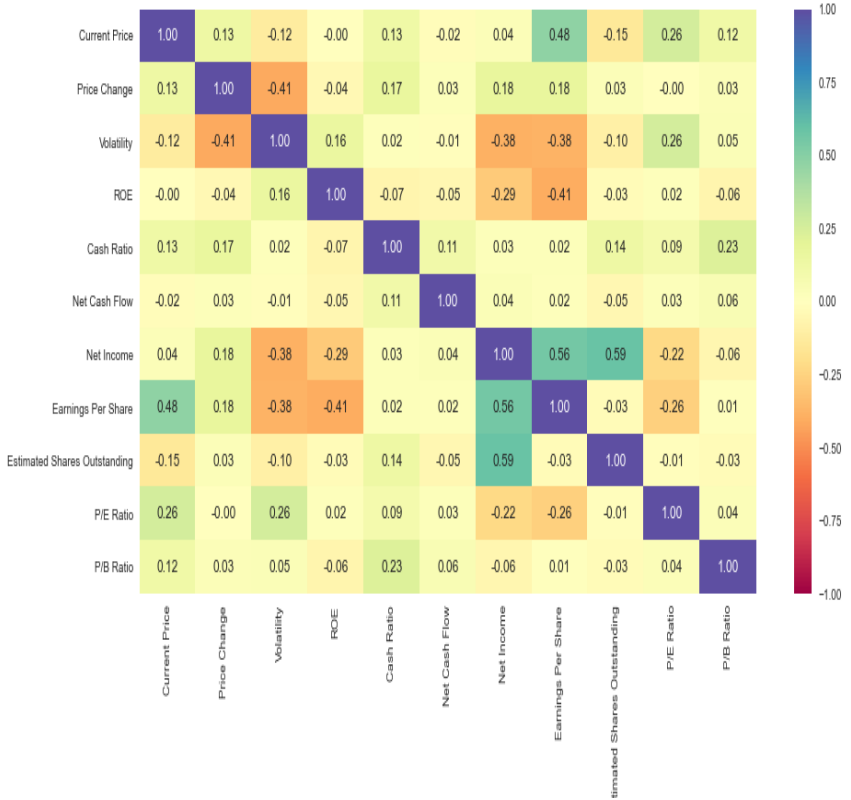
- The distribution appears normally distributed
- There is one outlier with a price to book ratio greater than 100
- There are a few organizations with negative price to book ratio in the distribution

Labelled Barplot for the GICS Sector



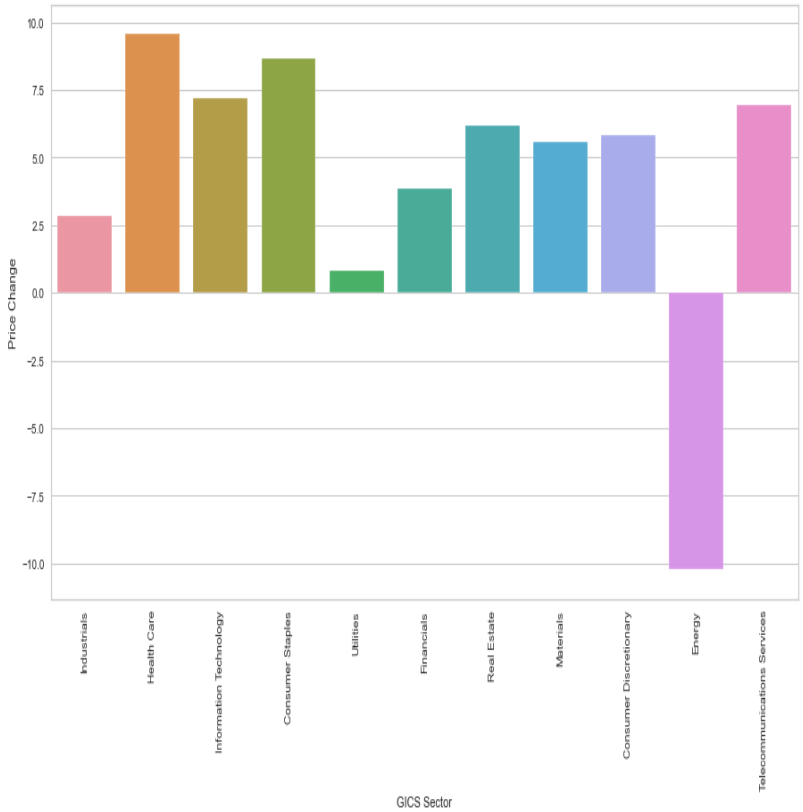
- The industrials have the highest count at 15.6% followed by financials at 14.4%
- The telecommunications services has the lowest count at 1.5% followed by consumer staples at 5.6%

Bivarate Analysis: Heat Map



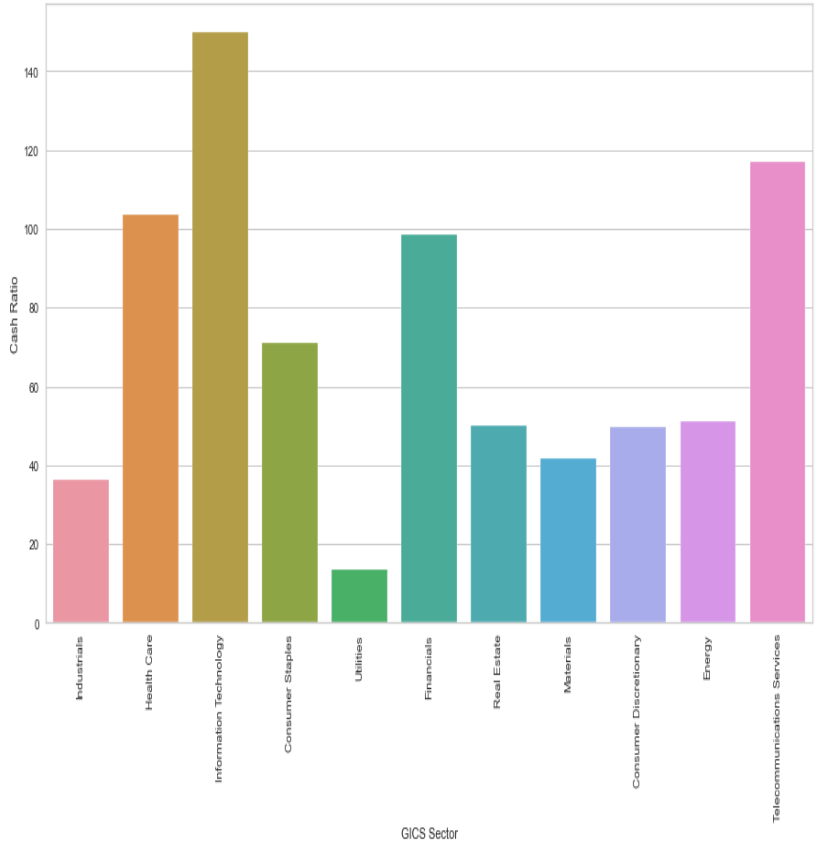
- There is a positive correlation between earnings per share, current price and net income
- Estimated shares outstanding has a positive correlation with net income
- Return on Equity has a slight positive correlation with volatility
- Cash ratio has a slight positive correlation with price to book ratio
- There is a high negative correlation between Earnings per share and Return on Equity

Barplot: Price Increase by GICS Sector



- The highest price change occurs in the Health Care sector followed by Consumer Staples
- The lowest price change occurs in Utilities
- There is a negative price change in the Energy Sector

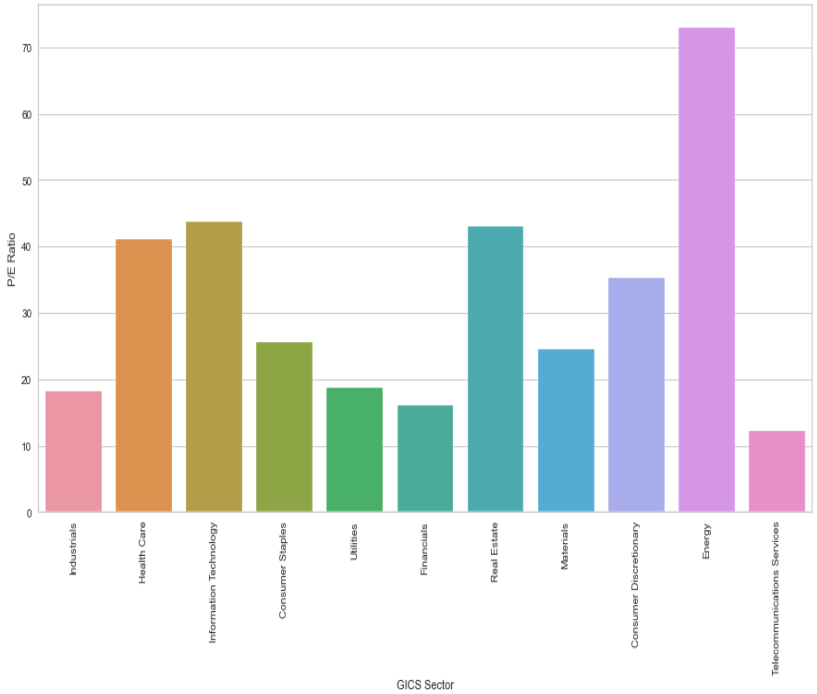
Barplot: Variance of Cash Ratio across Economic sectors



The sector with the highest cash ratio is the Information Technology sector followed by Telecommunications service

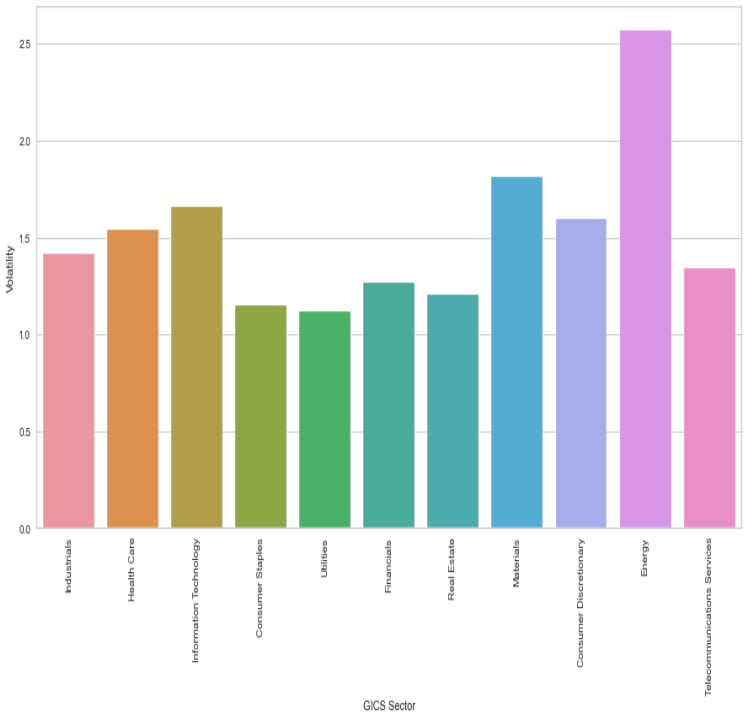
The sector with the lowest cash ratio is the utilities sector followed by Industrials

Barplot: Variance of P/E Ratio across economic sectors



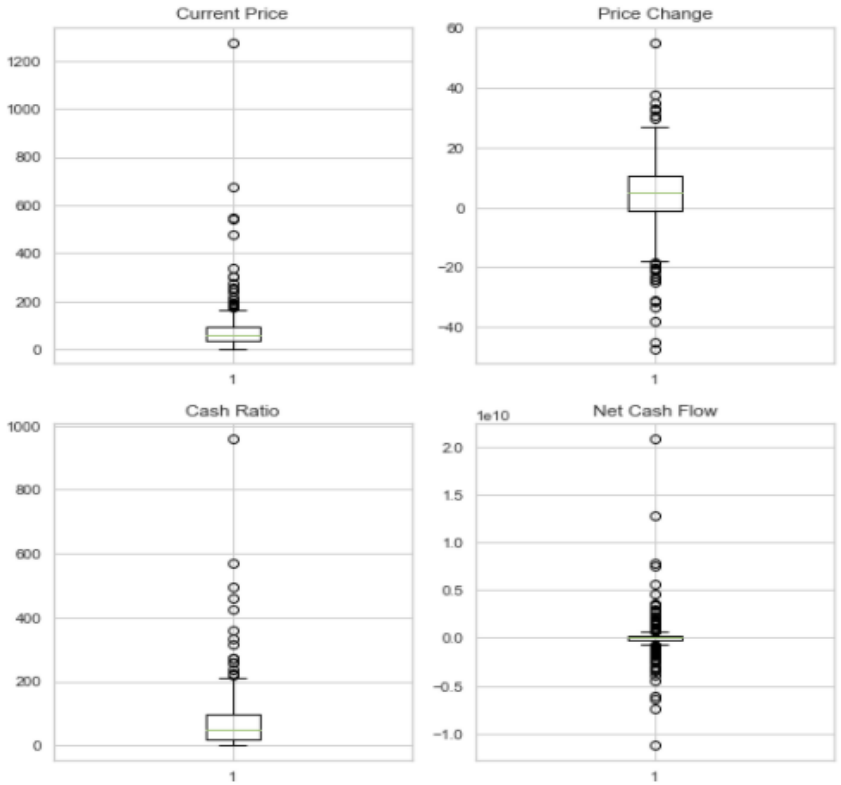
- The sector with the highest P/E ratio is the Energy sector followed by the Information technology sector
- The sector with the lowest P/E ratio is the Telecommunications sector followed by the Financials sector

Bar plot: Volatility across economic sectors



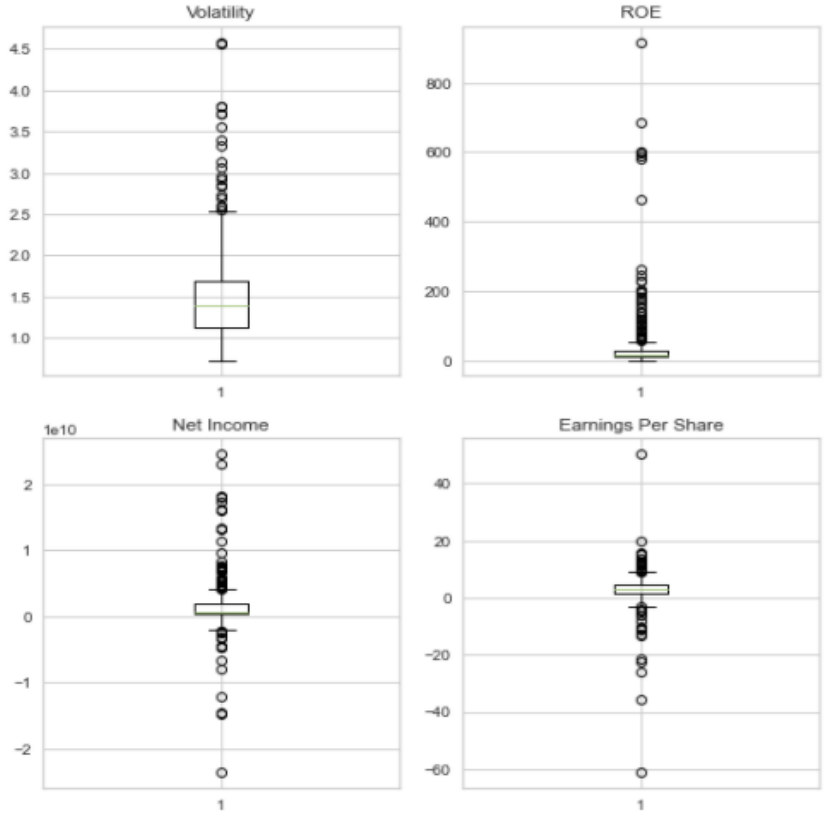
- The sector with the highest volatility is the Energy sector followed by the Materials and Information technology sector respectively
- The sector with the lowest volatility is the Utilities sector followed by the Consumer Staples sector

Boxplots: Outlier Check



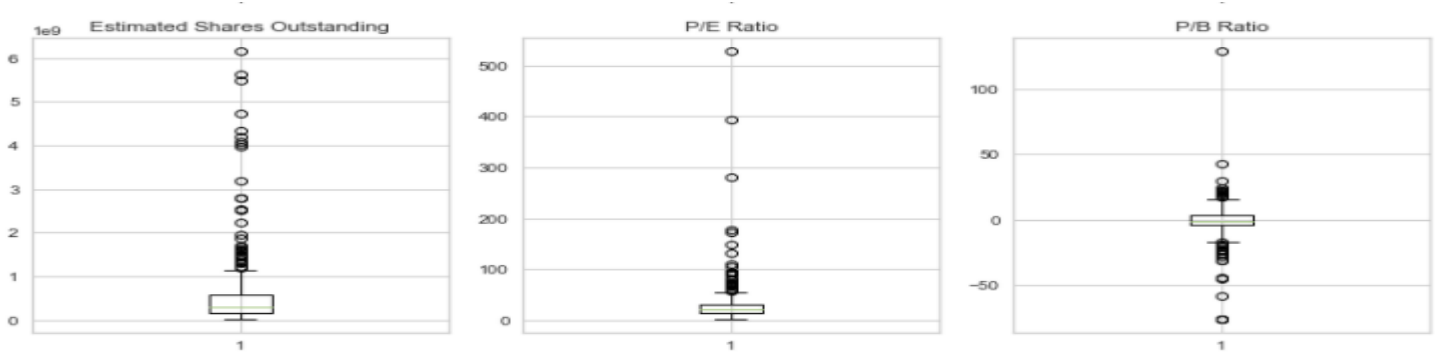
- Current Price
  - There are outliers but all above the upper quartile and upper whisker
- Price Change
  - There are outliers but they are above and below the lower and upper whisker. With some of the outliers having a negative figure
- Cash Ratio
  - The outliers are all above the upper whisker with one outlier having a significant cash ratio
- Net Cash Flow
  - The outliers are all above the upper and lower whisker with some outliers having negative net cash flow

Boxplots: Outlier Check



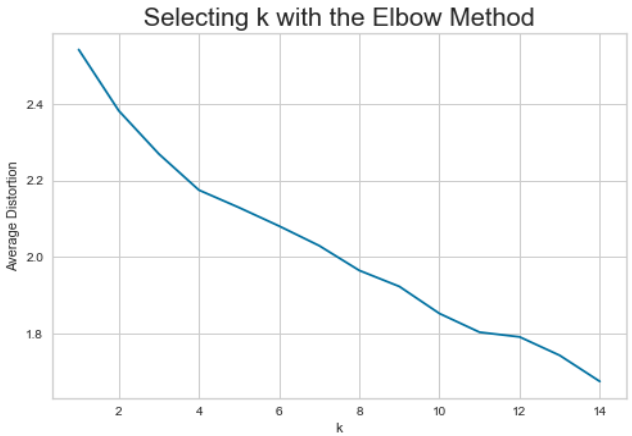
- Volatility
  - The outliers are all above the upper whisker with one particular outlier having a volatility greater than 4.5
- ROE
  - The outliers are all above the upper whisker with one particular outlier having a ROE greater than 800
- Net Income
  - The outliers are both above and below the upper and lower whisker respectively
- Earning Per Share
  - The outliers are both above and below the upper and lower whisker respectively

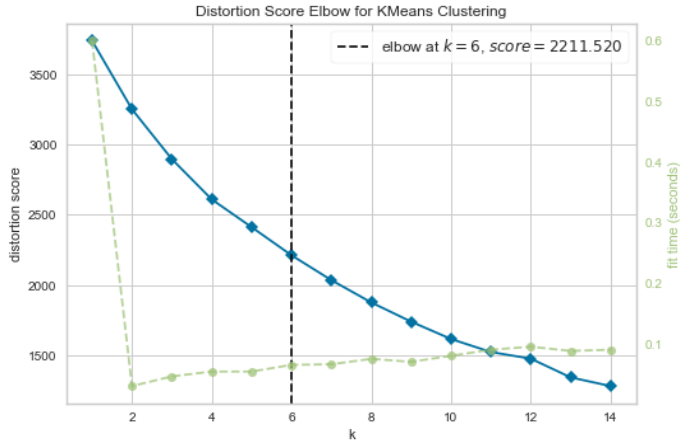
Outlier Check



- Estimated Shares Outstanding
  - The outliers are all above the upper whisker with one particular outlier having a volatility greater than 6
- P/E Ratio
  - The outliers are all above the upper whisker with one particular outlier having a volatility greater than 6
- P/B Ratio
  - The outliers are both above and below the upper and lower whisker respectively

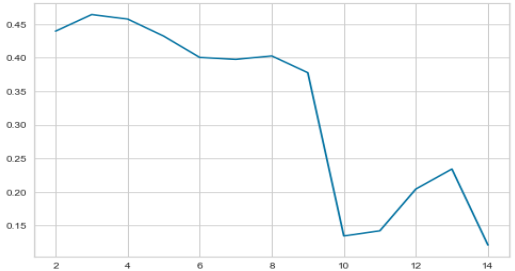
K-means Clustering

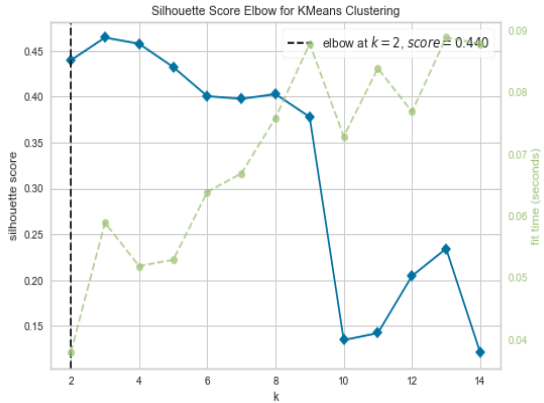


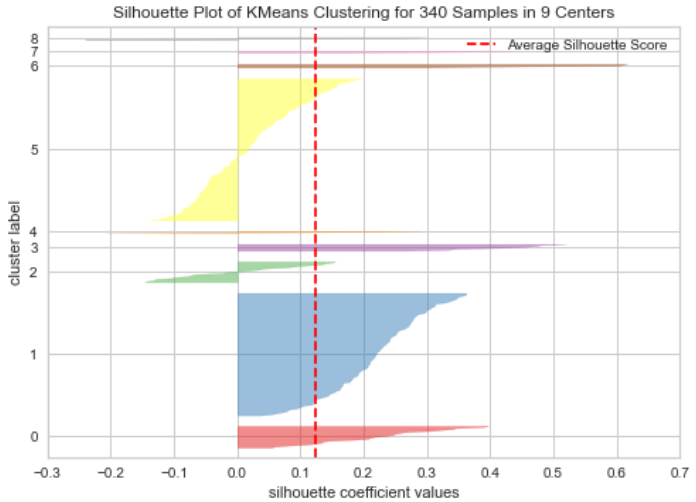


- There are different elbow shapes that can be seen at 4, 6, 10,12
- Based on the distortion score elbow for K-Mean clustering the elbow is at k =6

K means clustering: Silhouette scores







- A lot of the k clusters had a presence of clusters with below average silhouette scores
- K-9 clusters have their clusters greater than the average silhouette score

Cluster Profiles

KM_segments	GICS Sector	Security
0	Energy	19
	Industrials	1
	Information Technology	2
	Materials	1
1	Consumer Discretionary	15
	Consumer Staples	3
	Energy	1
	Financials	25
	Health Care	9
	Industrials	27
	Information Technology	10
	Materials	9
	Real Estate	7
	Telecommunications Services	1
	Utilities	19
2	Consumer Discretionary	5
	Consumer Staples	1
	Energy	1
	Health Care	8
	Information Technology	5
	Real Estate	1
	Telecommunications Services	1
3	Consumer Staples	1
	Energy	1
	Financials	3
	Health Care	1
	Telecommunications Services	2
4	Energy	3
5	Consumer Discretionary	17
	Consumer Staples	12
	Energy	5
	Financials	19
	Health Care	20
	Industrials	24
	Information Technology	14
	Materials	10
	Real Estate	19
	Telecommunications Services	1
	Utilities	5
6	Consumer Discretionary	1
	Consumer Staples	2
	Financials	1
	Industrials	1
7	Consumer Discretionary	1
	Health Care	1
	Information Technology	1
8	Consumer Discretionary	1
	Financials	1
	Health Care	1
	Information Technology	1

Name: Security, dtype: int64

Hierarchical Clustering

```
Cophenetic correlation for Euclidean distance and single linkage is 0.9232271494002922.
Cophenetic correlation for Euclidean distance and complete linkage is 0.7873280186580672.
Cophenetic correlation for Euclidean distance and average linkage is 0.9422540609560814.
Cophenetic correlation for Euclidean distance and weighted linkage is 0.8693784298129404.
Cophenetic correlation for Chebyshev distance and single linkage is 0.9062538164750717.
Cophenetic correlation for Chebyshev distance and complete linkage is 0.598891419111242.
Cophenetic correlation for Chebyshev distance and average linkage is 0.9338265528030499.
Cophenetic correlation for Chebyshev distance and weighted linkage is 0.9127355892367.
Cophenetic correlation for Mahalanobis distance and single linkage is 0.9259195530524591.
Cophenetic correlation for Mahalanobis distance and complete linkage is 0.7925307202850002.
Cophenetic correlation for Mahalanobis distance and average linkage is 0.9247324030159737.
Cophenetic correlation for Mahalanobis distance and weighted linkage is 0.8708317490180428.
Cophenetic correlation for Cityblock distance and single linkage is 0.9334186366528574.
Cophenetic correlation for Cityblock distance and complete linkage is 0.7375328863205818.
Cophenetic correlation for Cityblock distance and average linkage is 0.9302145048594667.
Cophenetic correlation for Cityblock distance and weighted linkage is 0.731045513520281.
```

- The highest cophenetic correlation is 0.9422540609560814 which is obtained with Euclidean distance and average linkage

Cluster Arrangements K-means

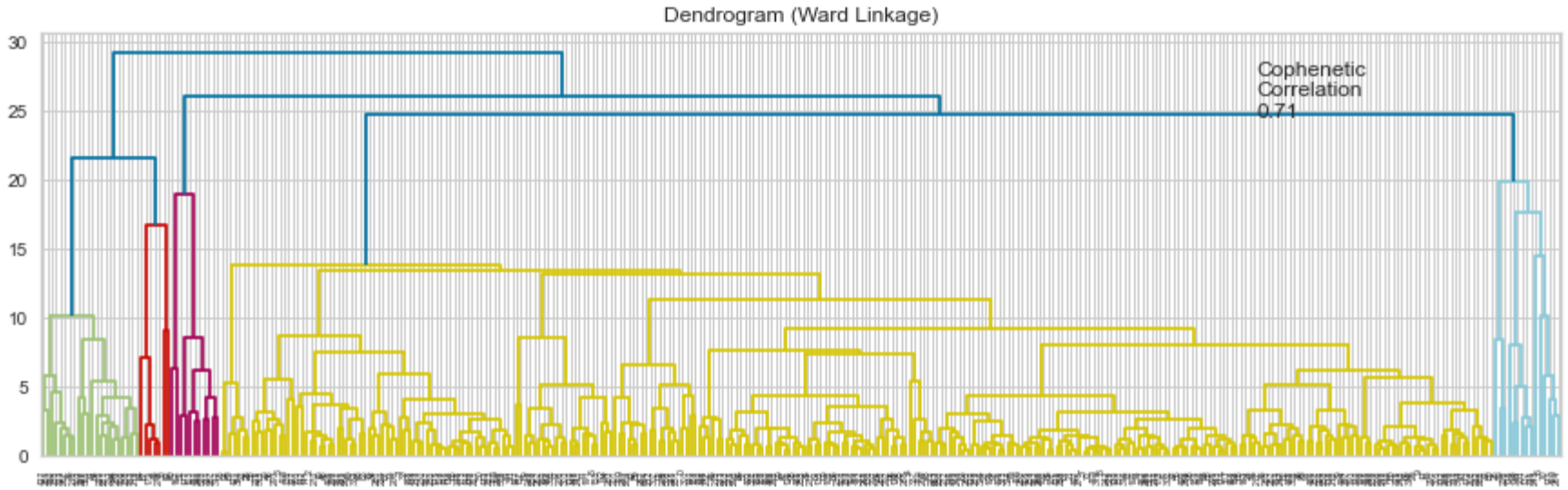
- Cluster 0 appears to be arranged based on:
  - Price Change
  - Net Income
  - Earnings per share
  - Price book ratio
- Cluster 1 appears to be arranged based on:
  - Net Cash Flow
  - Net Income
  - Earnings per share
  - Estimated shares outstanding
- Cluster 2 appears to be arranged based on:
  - Net Cash Flow
  - P/E Ratio
  - P/B Ratio
- Cluster 3 appears to be arranged based on:
  - Cash Ratio
  - Net Cash Flow
  - Earnings per share
  - Estimated Shares Outstanding
  - P/E Ratio
  - P/B Ratio

Different linkage methods with Euclidean distance

```
Cophenetic correlation for single linkage is 0.9232271494002922.
Cophenetic correlation for complete linkage is 0.7873280186580672.
Cophenetic correlation for average linkage is 0.9422540609560814.
Cophenetic correlation for centroid linkage is 0.9314012446828154.
Cophenetic correlation for ward linkage is 0.7101180299865353.
Cophenetic correlation for weighted linkage is 0.8693784298129404.
```

- The highest cophenetic correlation is 0.9422540609560814 which is obtained with average linkage

Dendograms for the different linkage method with Euclidean distance



Cluster Profile

HC_segments	GICS Sector	Security
0	Consumer Discretionary	1
	Consumer Staples	2
	Energy	22
	Financials	1
	Industrials	1
	Information Technology	1
	Materials	1
1	Consumer Discretionary	3
	Consumer Staples	1
	Health Care	5
	Information Technology	4
	Real Estate	1
	Telecommunications Services	1
2	Consumer Discretionary	1
	Consumer Staples	1
	Energy	1
	Financials	4
	Health Care	1
	Information Technology	1
	Telecommunications Services	2
3	Consumer Discretionary	35
	Consumer Staples	15
	Energy	7
	Financials	44
	Health Care	34
	Industrials	52
	Information Technology	27
	Materials	19
	Real Estate	26
	Telecommunications Services	2
	Utilities	24

Name: Security, dtype: int64

Cluster Arrangements

- Cluster 0 appears to be arranged based on:
  - Volatility
  - Return on Equity
  - P/E Ratio
  - Price book ratio
- Cluster 1 appears to be arranged based on:
  - Current Price
  - Price change
  - Volatility
  - Cash ratio
  - Price book ratio
- Cluster 2 appears to be arranged based on:
  - Net Cash Flow
  - Net Income
  - Estimated shares outstanding
  - P/B Ratio
- Cluster 3 appears to be arranged based on:
  - Current Price
  - Price Change
  - Net Cash Flow
  - Net Income

Compare Clusters Obtained from K-means and Hierachical Clustering Technique

K-means

- Cluster 0 appears to be arranged based on:
  - Price Change
  - Net Income
  - Earnings per share
  - Price book ratio
- Cluster 1 appears to be arranged based on:
  - Net Cash Flow
  - Net Income
  - Earnings per share
  - Estimated shares outstanding
- Cluster 2 appears to be arranged based on:
  - Net Cash Flow
  - P/E Ratio
  - P/B Ratio
- Cluster 3 appears to be arranged based on:
  - Cash Ratio
  - Net Cash Flow
  - Earnings per share
  - Estimated Shares Outstanding
  - P/E Ratio
  - P/B Ratio

Hierachical

- Cluster 0 appears to be arranged based on:
  - Volatility
  - Return on Equity
  - P/E Ratio
  - Price book ratio
- Cluster 1 appears to be arranged based on:
  - Current Price
  - Price change
  - Volatility
  - Cash ratio
  - Price book ratio
- Cluster 2 appears to be arranged based on:
  - Net Cash Flow
  - Net Income
  - Estimated shares outstanding
  - P/B Ratio
- Cluster 3 appears to be arranged based on:
  - Current Price
  - Price Change
  - Net Cash Flow
  - Net Income

Compare Clusters Obtained from K-means and Hierachical Clustering Technique

- In k-means clusters net income and net cash flow are reflected in all the clusters
- In hierarchical clustering it is volatility and net income
- The price book ratio is synonymous with cluster 0 in both k-means and hierarchical clustering techniques
- There is no variable synonymous in both k-means and hierarchical clustering techniques in cluster 1
- The net cash flow is synonymous with cluster 1 in both k-means and hierarchical clustering techniques
- The net cash flow is synonymous with cluster 2 in both k-means and hierarchical clustering techniques

Actionable Insights and Recommendations

The actionable insights as noted on the various clusters are as follows:

- Cluster 0 is arranged on earnings more for the investor than for the company. Thus, individuals with a preference for short term/immediate returns on investment would fit into this group
- Cluster 1 is arranged for reinvestment in the company for long term growth. Thus for individuals more interested in the growth of the organization in the long term and reinvestment of their returns on investment in the company this would appear to be an ideal fit. However, there is volatility thus high risk investors would also be a good fit.
- Cluster 2 is arranged for investors with a penchant for reinvestment in the company in the long term. However, without volatility, thus returns might be lower in the long term. For investors with safe to moderate returns this group would be an ideal fit. In addition, investors also interested in non dilution of their original holdings would also be a good fit for this group.
- Cluster 3 is arranged also for investors with long term goals but also with returns on their investments in the short term

Business Recommendations

In structuring a portfolio the following factors should be considered:

- Net Income: This plays a very important role in the cluster arrangement, as individuals definitely want a return on their investment, thus sectors with constant price changes and low volatility would be idle
- Volatility: Volatility could be positive or negative, however given the need for diversification in portfolios there is a need for slight volatility especially in markets that tend to follow the same trends
- Externalities: External factors such as government regulation, environmental impact factors, legal rules must be incorporated into the analysis as they have significant effect on outcomes
- Dividend pay out: Information on dividend paid or accrued for the different sectors should also incorporated as this speaks directly to the risk appetite of investors or potential investors
- Cash ratio: This is also important as it impacts the ability of an organization to meet not just its short term obligations but its long term obligations

Potential Benefits of Implementing Business Recommendation

- Clients: Ability of the clients to sustain and increase any investments made in any of the sectors given the diversity of their recommended portfolios
- Company: Increase in informal advertisements in terms of recommendations made by existing clients which in turn would grown not only the business but their client base
- Profitability: Increase in clients would impact on fees charged, increase in fees would lead to increase in both short and long term working capital and an ability to meet recurrent and long term expenditure
- Diversification: Ability to build funds for clients to invest in, given the ability to cluster, diversification can be incorporated in these funds

greatlearning
Power Ahead

Happy Learning !