

Business Presentation

Contents

Business Problem Overview and Approach

Data Overview

EDA

K-Means Clustering

Hierarchical Clustering

K-Means Vs Hierarchical Clustering

Business Insights and Recommendations

Business Problem Overview and Solution Approach

Core business idea

- Maximize investment returns of customers with personalized investment strategies based on patterns in stock clusters.

Problem to tackle

- Analyzing the financial indicators data and grouping the stocks based on the attributes provided.

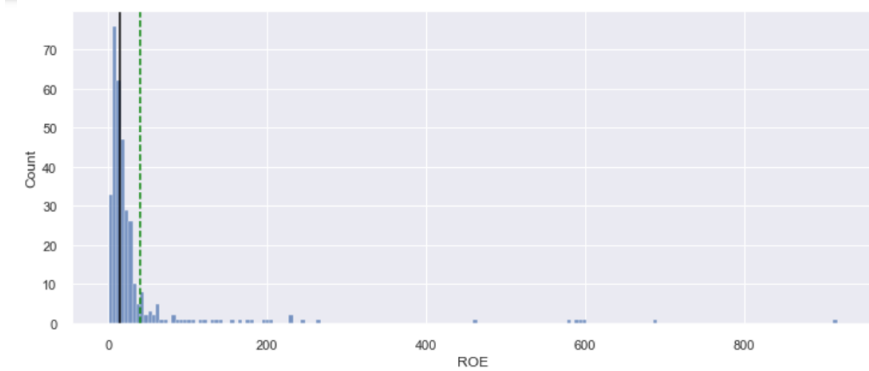
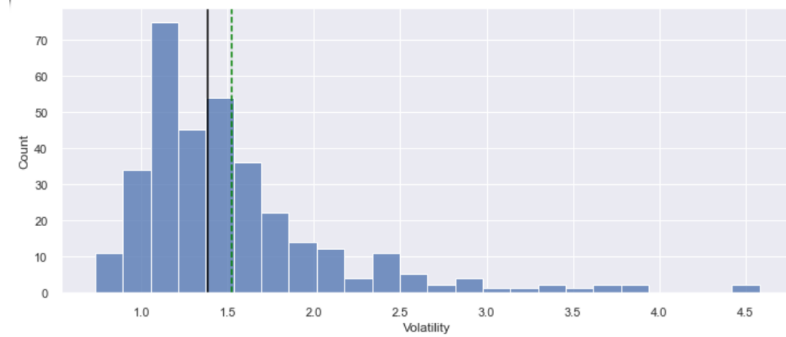
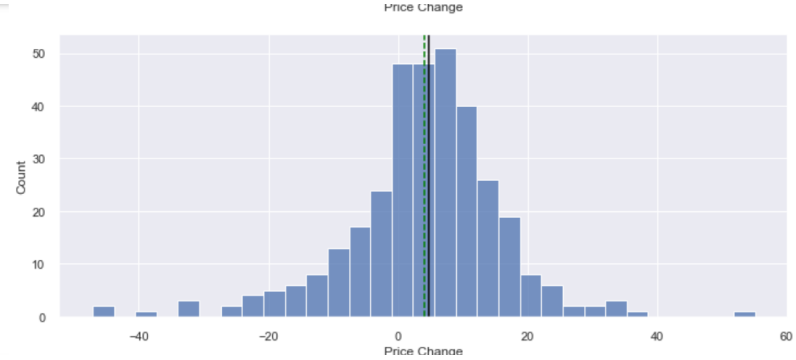
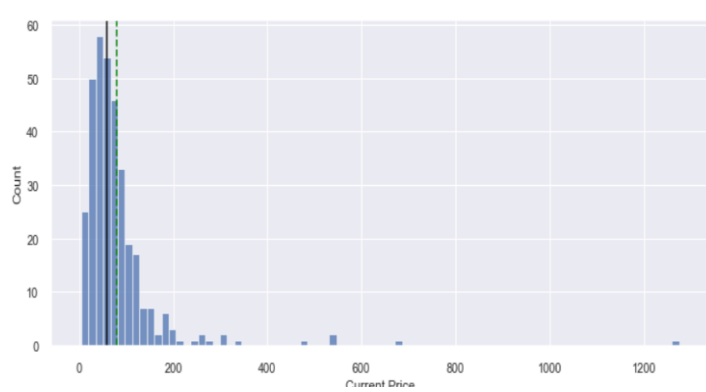
Financial implications

- Based on characteristics of each group of stocks, customers with varying risk tolerance and investment capabilities can be advised to increase their financial returns and inturn improve the revenue of the financial consultancy.

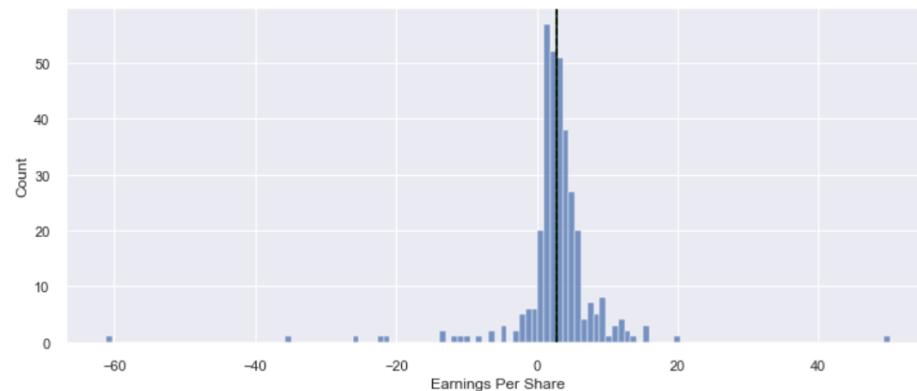
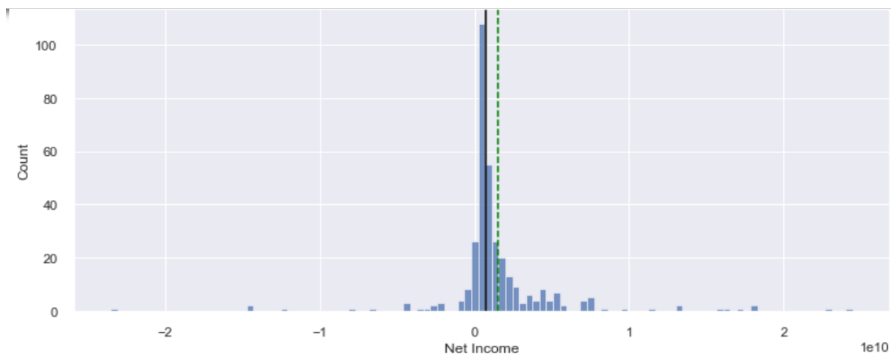
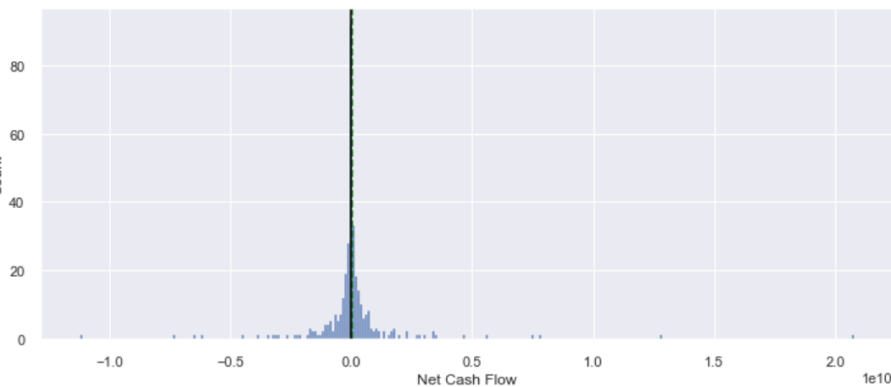
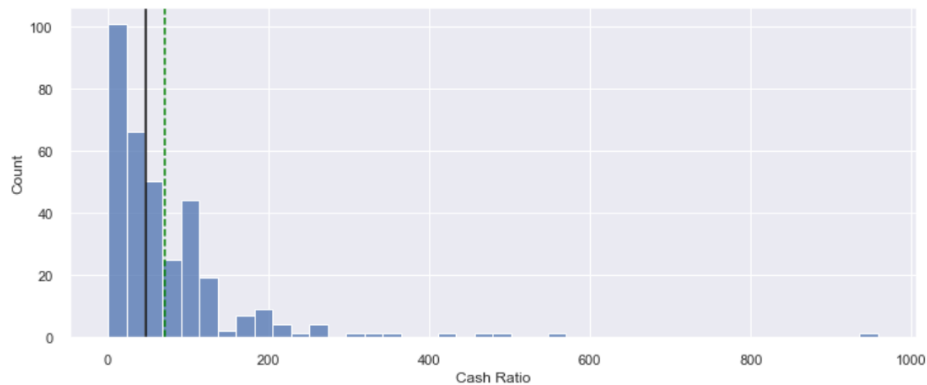
Data Overview

- The data contains information about 340 company and their stock details.
- The characteristics include Security, GICS Sector, GICS Sub Industry, Current Price, Price Change and many more. There are 15 such characteristics.
- Object Variables 'Ticker Symbol', 'Security', 'GICS Sector', 'GICS Sub Industry' have been converted to Categorical Variables.
- There are no duplicate rows.
- There are no missing values.
- There are many outliers.

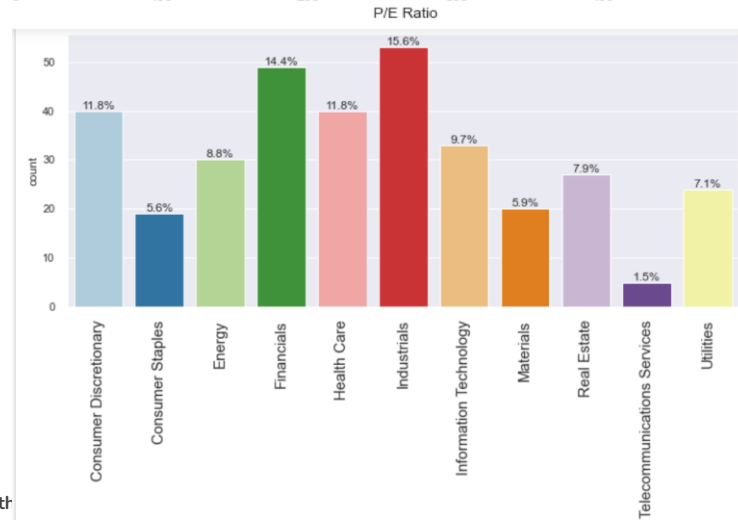
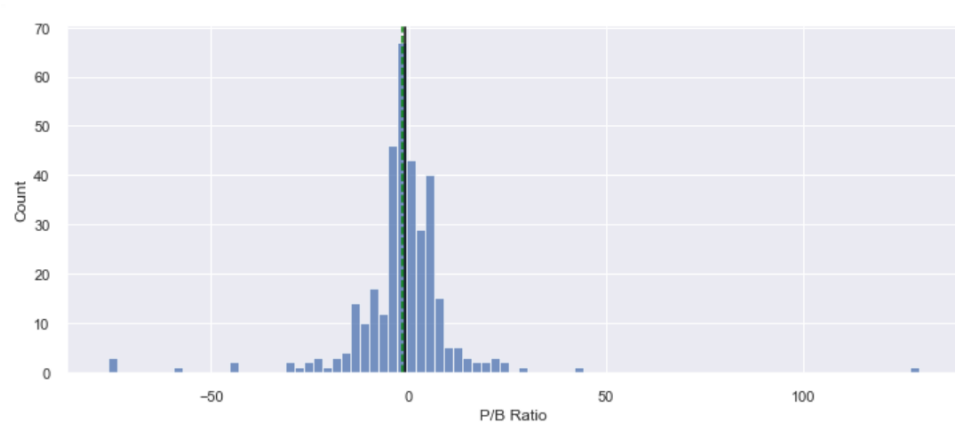
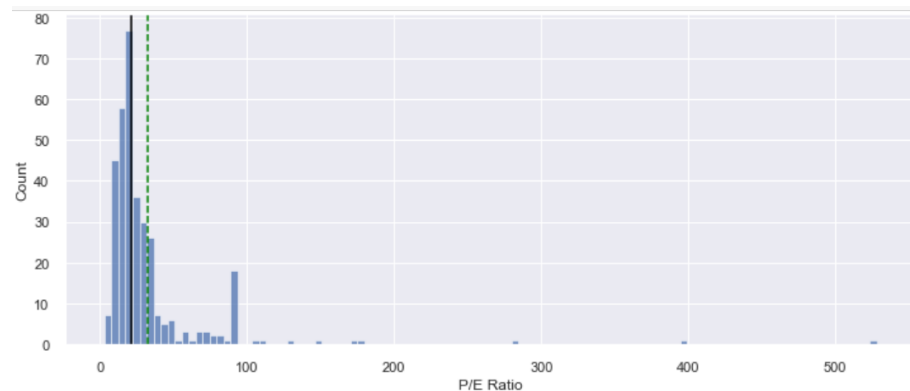
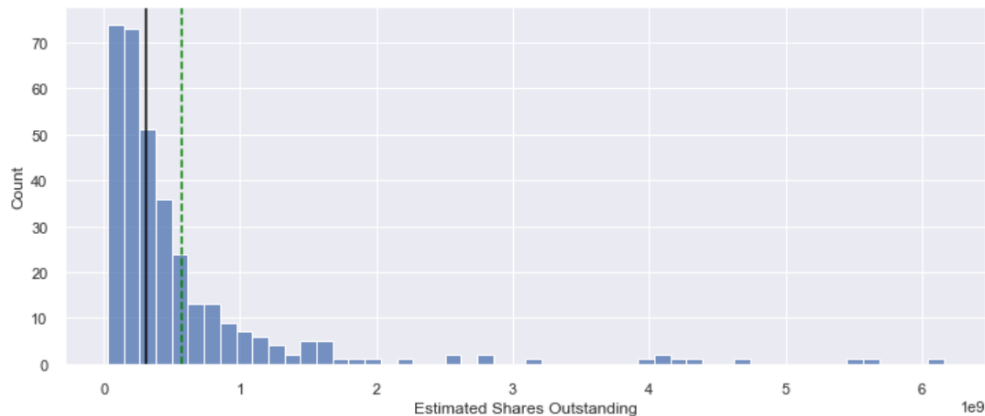
EDA



EDA(Contd.)



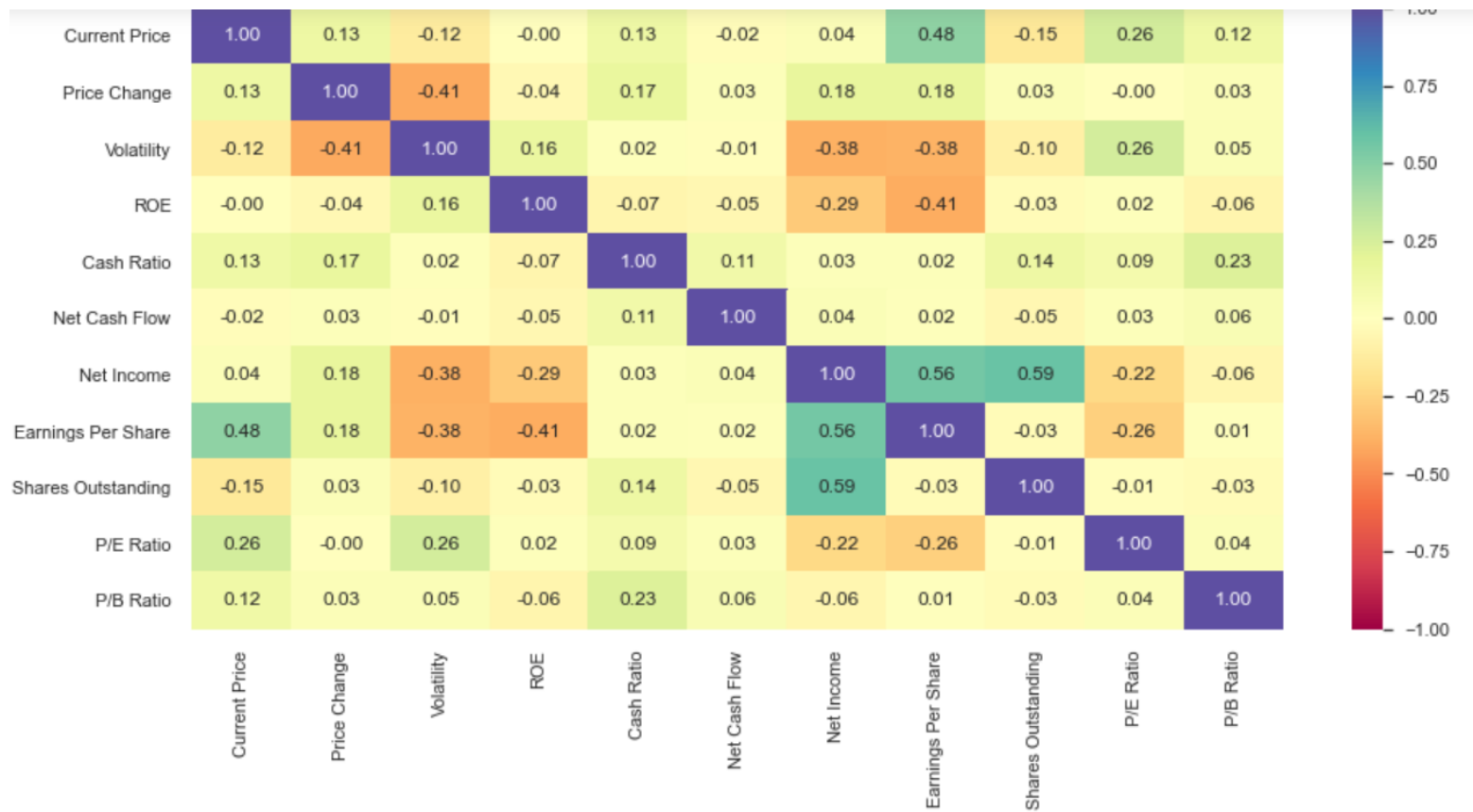
EDA(Contd.)



Observation

- Current Price, Volatility, ROE, Cash Ratio, Estimated Shares Outings and P/E Ratio are positively skewed. There are some outliers.
- We see a bell shape curve for Price change, Net Cash Flow, Net Income, Earnings Per Share and P/B Ratio having outliers at both the end.
- In most of the cases, current Price is less than 100 but, there are some that go till 1300.
- Price change ranges between -45 and ~60.
- Volatility ranges between 0.5 and 1.0.
- Mostly ROE value is less than 50.
- In most of the cases, cash ratio is less than 100 but, it can go upto 900.
- Though, Net Cash flow ranges between -1.5 and 2.3, Mostly, it ranges between -0.25 and 0.25.
- Most of the cases, Earnings per share is less than 5.
- P/E Ratio is always greater than 0. Mostly it is less than 25. Highest being ~525.
- In some cases, P/B ratio is less than 0. A low P/B ratio could also mean the company is earning a very poor.
- 15.6% of the stocks come under Industrials sector followed by Financials.
- Only 1.5% of the stocks come under Telecommunications Services.

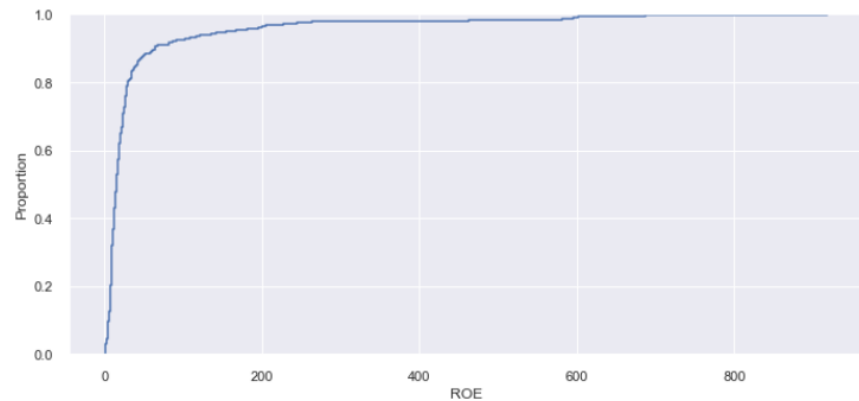
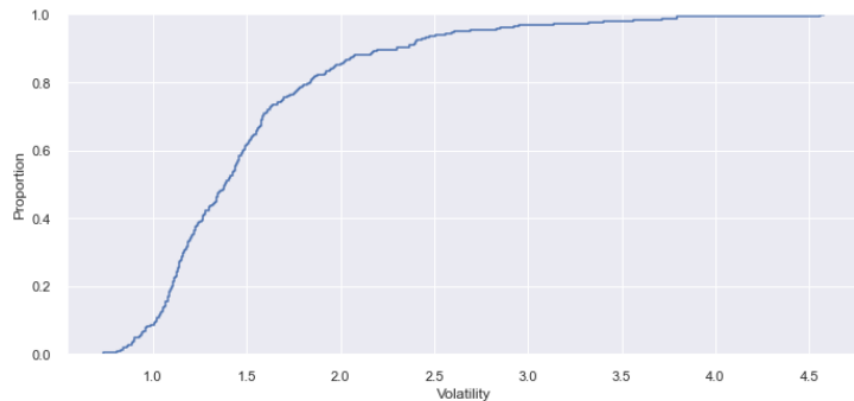
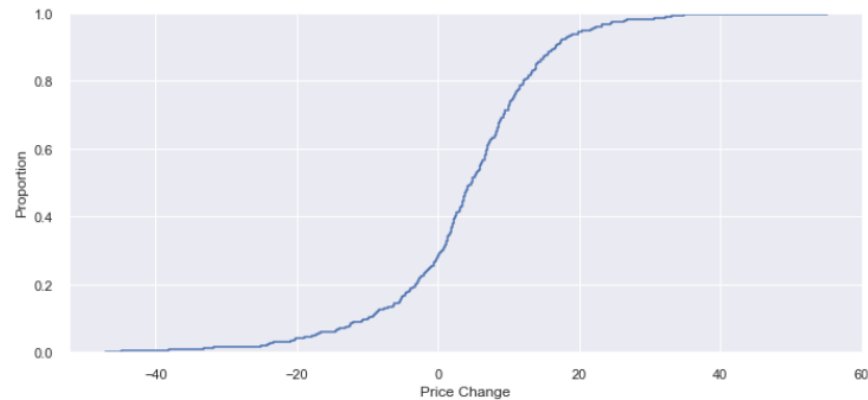
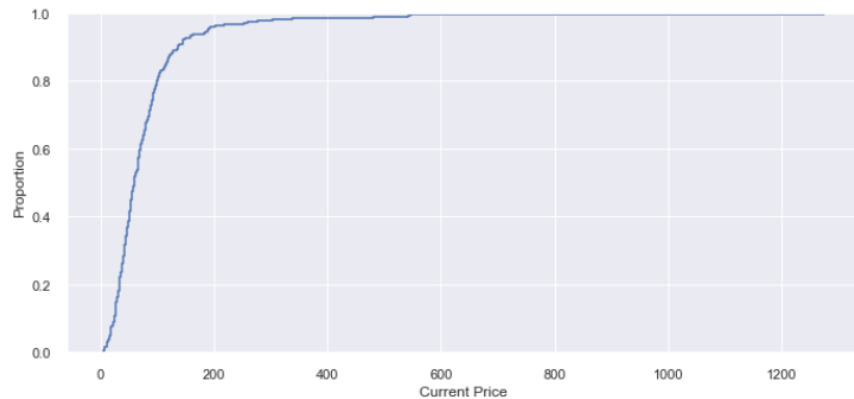
Bivariate Analysis



Observation

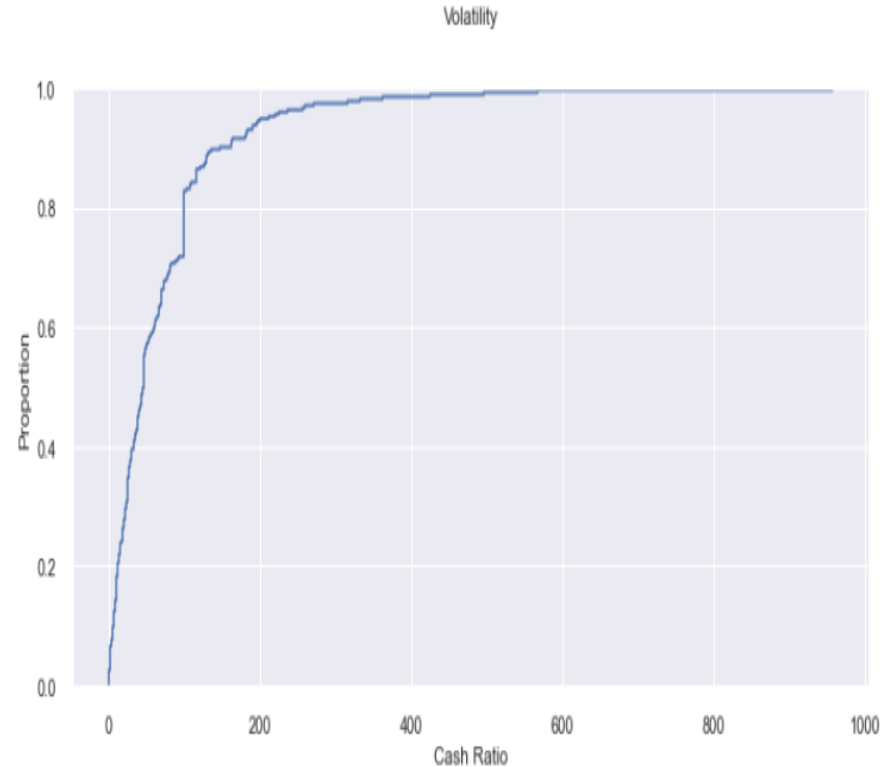
- ***A. The highest correlation is between Estimated Shares Outstanding and Net Income(0.59) followed by Earnings Per Share and Net Income***
- ***B. There is negative correlation between***
 - Volatility and Current Price, Price Change, Net Cash Flow, Net Income, Earnings Per Share and Estimated Shares Outstanding.
 - Current Price and ROE, Net Cash Flow.
 - Price Change and ROE.
 - ROE has negative correlation with all except Volatility and P/E Ratio.
 - Net Cash FLOW is negatively correlated with current price, Volatility and Estimated shares outstanding.
 - Earnings per share and Estimated shares outstanding, P/E Ratio.
 - P/E Ratio is negatively correlated with Price Change, Net Income, Earnings Per Change and Estimated shares outstanding. showing, if one value increases the other decreases.

CDF Plots



CDF Plots(Contd.)

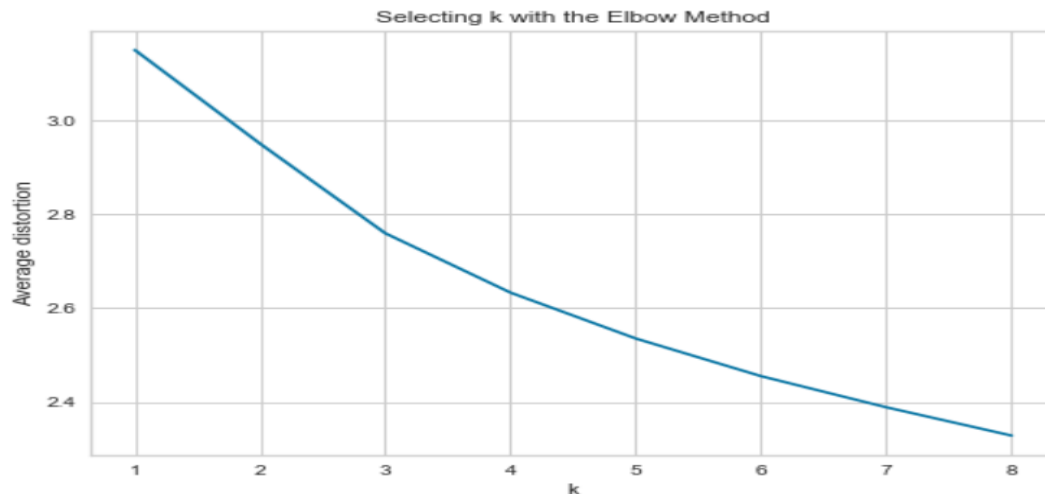
1. 90% of the stocks,current price is less than 120.
2. Price change varies between -40 and 40.
3. 95% of the stocks, the Volatility is less than 2.5
4. 98% of the stocks, ROE value is less than 200.
5. 95% of the stocks have cash ratio less than 200.



K-Means Clustering(Elbow Method)

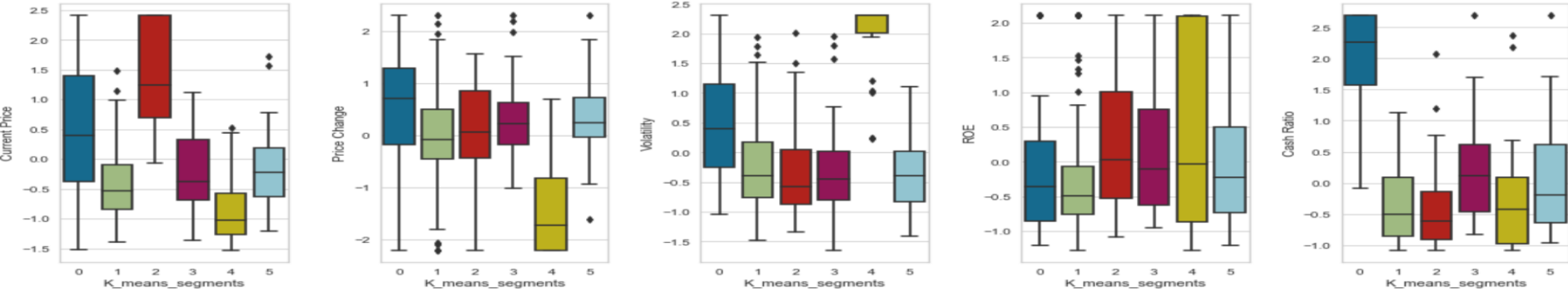
- Appropriate value of K seems to be 3 or 4.

Number of Clusters: 1	Average Distortion: 3.1482665386211908
Number of Clusters: 2	Average Distortion: 2.9491046912239587
Number of Clusters: 3	Average Distortion: 2.7586235347054844
Number of Clusters: 4	Average Distortion: 2.6324515003371367
Number of Clusters: 5	Average Distortion: 2.5349290565854523
Number of Clusters: 6	Average Distortion: 2.454942164539365
Number of Clusters: 7	Average Distortion: 2.3887833542845023
Number of Clusters: 8	Average Distortion: 2.328622793920396

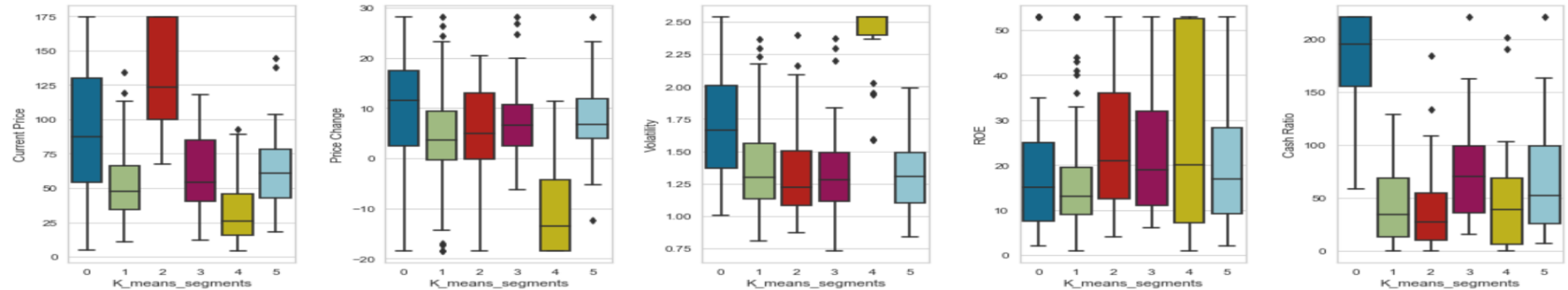


K-Means (Box Plots)

Boxplot of scaled numerical variables for each cluster

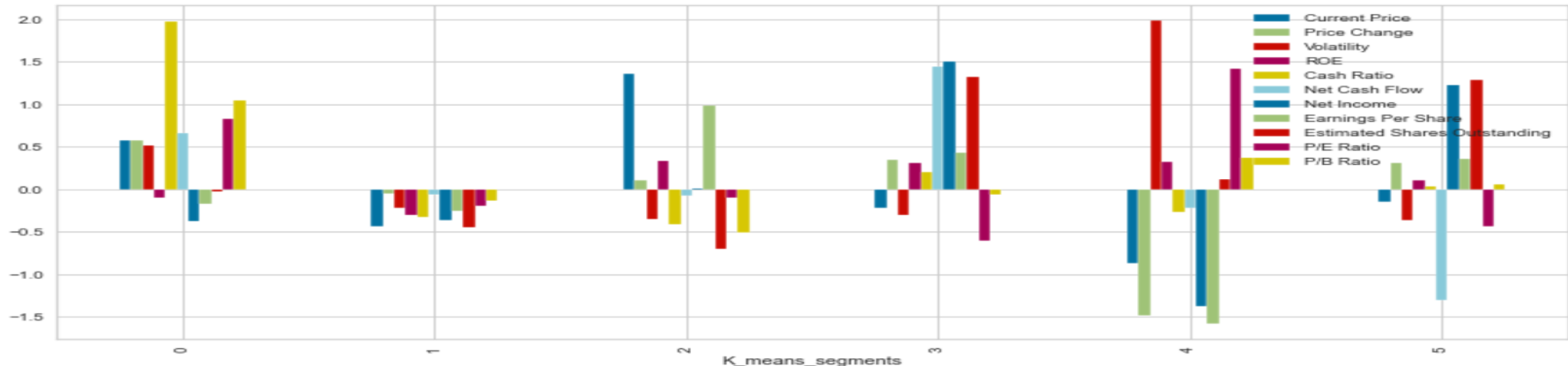


Boxplot of original numerical variables for each cluster



Finding clusters with silhouette coefficient

1. After considering average silhouette score and comparing with the plots, we observe that
2. n_ number of clusters = 6, 7 , 8 are pretty decent as the distribution is nearly similar. Though there is one cluster that has more distribution. They also have negative silhouette scores.
3. number of clusters = 6 have 5 negative silhouette scores while the others have 6 negative silhouette scores.
4. Hence, we will consider number of clusters = 6, as silhouette score is high enough and there is knick at 6 in the elbow curve.



Hierarchical Clustering

1. Cophenetic correlation for Euclidean distance and average linkage is 0.73. This value is the highest value which will be considered.
2. Hence we will consider Average linkage.
3. Number of clusters = 6 appears to be appropriate

From Average Linkage,

1. Cluster 1 has 290 companies, Cluster 2 has 15 companies, cluster 0 has 31 countries, cluster 4 has 1 companies, cluster 3 has 1 company and Cluster 5 has 2.
2. There is much variability here, Lets see how Ward linkage looks as it has more distinct clusters from dendogram.

From Ward Linkage,

1. Cluster 0 has 197 companies.
2. Cluster 4 has 25 companies.
3. Cluster 3 has 23 Companies.
4. Cluster 1 has 52 companies.
5. Cluster 2 has 43 companies.

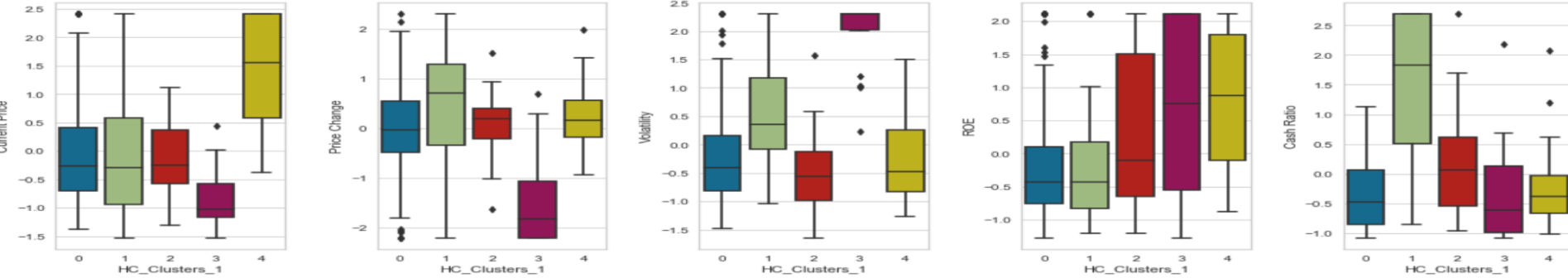
Hierarchical Clustering

	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	K_means_segments	HC_Clusters	coun
1														
0	67.842897	4.165814	1.379430	17.510152	43.710660	-22317788.705584	791304172.588832	3.002418	274869321.770762	23.712726	-3.088535	1.472081	1.005076	
1	71.569176	9.460529	1.745668	17.067308	147.490385	136610663.461538	924025872.596154	2.246563	619951719.447115	32.907674	5.329730	1.423077	1.365385	
2	65.316907	5.686257	1.236535	25.276163	67.127907	59225113.372093	3958289215.116279	4.497558	1012930930.315697	16.923639	-4.105437	3.883721	1.000000	
3	32.947827	-11.234540	2.393770	31.478261	43.347826	-136632809.782609	-1356654576.086957	-2.242609	473814086.308804	50.100914	0.689859	4.000000	0.000000	
4	130.412701	7.182146	1.337093	33.130000	51.440000	-147072915.000000	2570160510.000000	8.128350	332216165.291600	17.183748	3.078650	2.760000	1.000000	

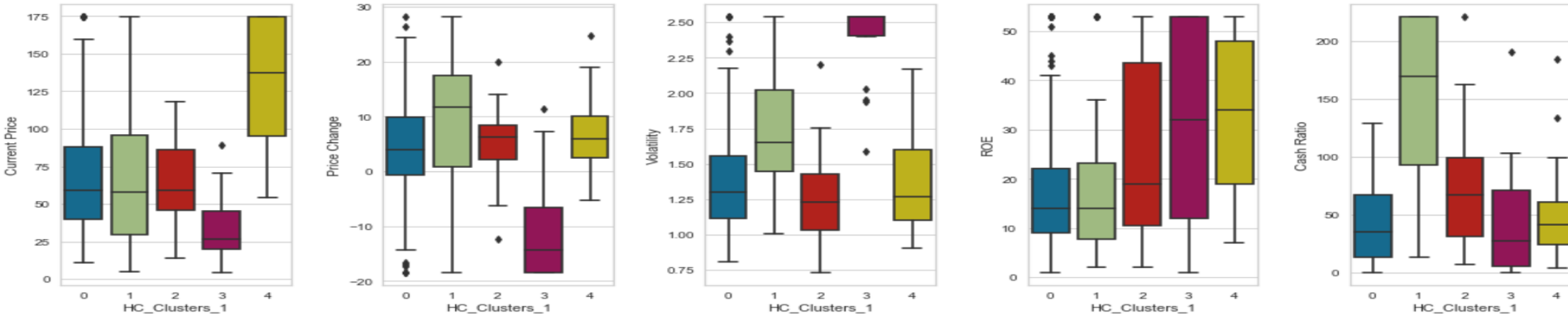
1. Cluster 0 has high number of segments.
2. Cluster 1 has Price change, Cash Ratio, Net Cash Flow and P/B ratio.
3. Cluster 2 has Net Income, Estimated Shares Outstanding.
4. Cluster 3 has high Volatility and P/E Ratio.
5. Cluster 4 has high current Price, ROE and Earnings per share.

Box-plots

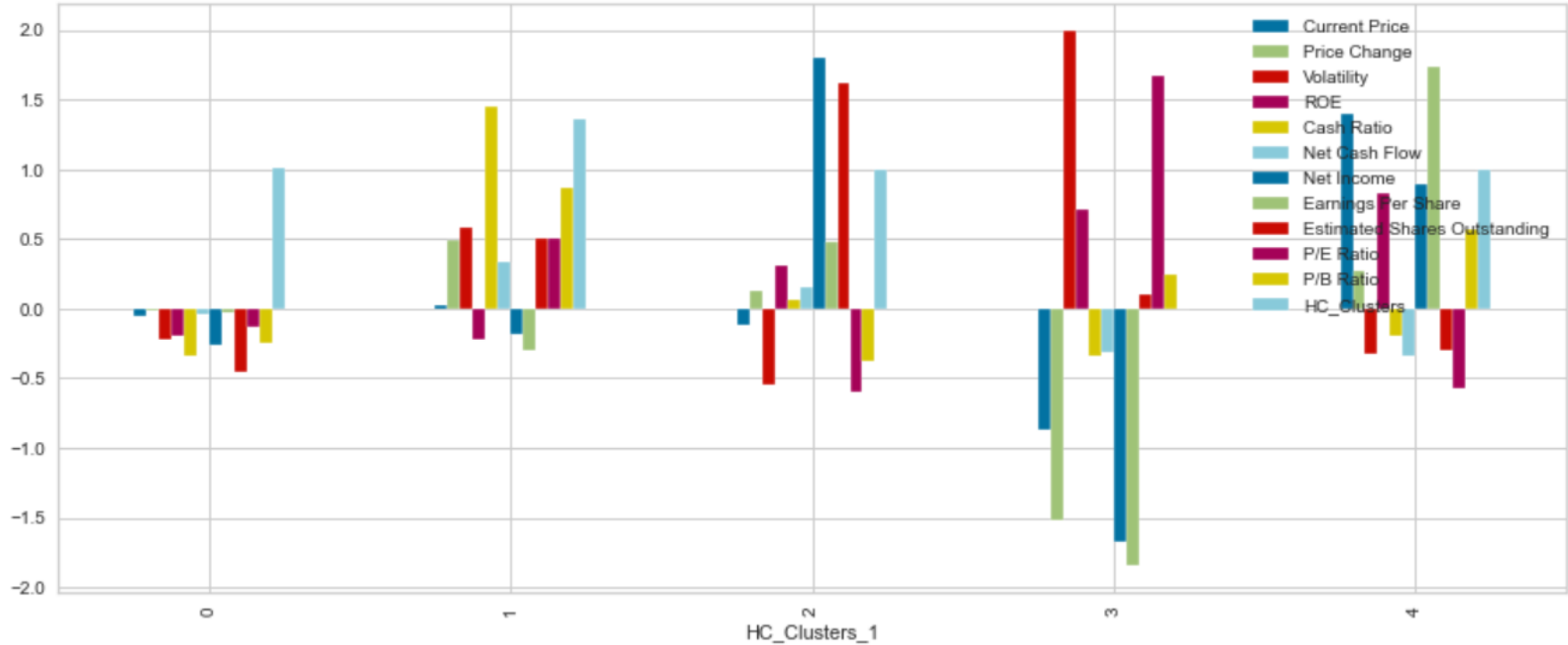
Boxplot of scaled numerical variables for each cluster



Boxplot of original numerical variables for each cluster



Hierarchical Clustering



K-Means Vs Hierarchical Clustering

- Hierarchical clustering took a lot of time to create the dendograms.
 - Execution time for K- Means clustering was way faster.
 - From both algorithms, we see that # optimal clusters is similar.
 - We ended up with 6 clusters through K-Means clustering.
-
- cluster 0 had 35 companies.
 - cluster 1 had 139 companies.
 - cluster 2 had 59 companies.
 - cluster 3 had 35 companies.
 - cluster 4 had 30 companies.
 - cluster 5 had 42 companies.

K-Means Vs Hierarchical Clustering

- In Hierarchical clustering, we decided to end up with 5 clusters.
 - Cluster 0 with high number of segments had 197 companies.
 - Cluster 4 high current Price, ROE and Earnings per share had 25 companies.
 - Cluster 3 with high Volatility and P/E Ratio had 23 Companies.
 - Cluster 1 with high Price change, Cash Ratio, Net Cash Flow and P/B ratio had 52 companies.
 - Cluster 2 with high Net Income, Estimated Shares Outstanding had 43 companies. In both methods we ended with some clusters which have very similar features:
 - All companies with high current price, ROE and EPS ended in one cluster.
 - Companies with high volatility and P/E ratio ended in one cluster.
 - Companies with high net income and Estimated shares outstanding ended in one cluster.

Business Insights and Recommendations

We have found companies with some very similar metrics as follows:

- Companies with high current price, ROE and EPS
- Companies with high volatility and P/E ratio.
- Companies with high net income and Estimated shares outstanding.
- Companies with high price change, cash ration and P/B ratio. Based on the risk tolerance and growth appetite of the customers, they can be directed to invest in either of these clusters.
- Customers looking for quick but risky returns could be routed to the cluster with high volatility.
- Customers looking for stable long term returns could be routed to the cluster with high nt income, estimated shares outstanding as the current price is in the average range.
- Customers who have high investment funds can be directed to invest in clusters that have high current price but also high ROE and EPS.
- Customers with low investment funds but higher risk tolerance can be directed towards the clsuter with high volatility as the current price of this cluster is very low.

greatlearning
Power Ahead

Happy Learning !

