Trade&Ahead

Trade&Ahead – Unsupervised Learning

04/06/2022

Contents / Agenda

- Executive Summary
- Business Problem Overview and Solution Approach
- EDA Results
- Data Preprocessing
- Model Performance Summary
- Appendix

Business Problem Overview and Solution Approach

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 it also provides some taxation 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.

In order to reach this financial goals is important to maintain a diversified portfolio when investing in stocks to maximize earnings under any market condition. 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.

Due to the objectives of Trade&Ahead to provided its customers with a personalized investment strategy they seek a better understating of the current they manage. For this they want to analyzed the different cluster of stocks through the shared attributes provided in the data set.

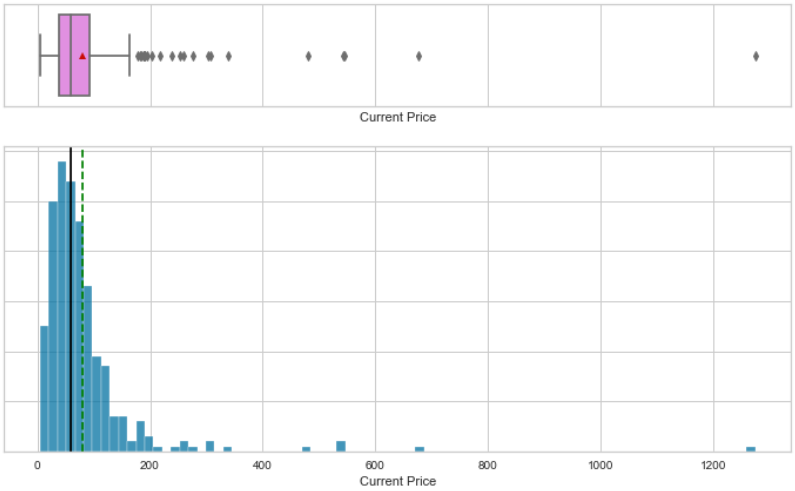
Executive Summary

- Through the development of this project the elaboration of clusters can take different routes through unsupervised learning. The objective of this analysis is to group different kinds of stocks besides their classical portfolio by sector management. In the end, the Hierarchical Cluster through log-transformed variable was chosen.
- The cluster 4 was elaborated by the energy sector which led to the highest ROE among the clusters but, these shares also showed a negative Earning Per Share and negative income.
- Cluster 7 conformed of Real State, Consumer discretionary, Consumer Staples, Financial and Industrial can give a high ROE cluster.
- The cluster 6 shows a cluster of Consumer Discretionary, Energy, Health care and Information Technology can give the highest return per share.
- The clusters 1 and 7 show a positive change in their price change which can lead to portfolio management of high trading stocks in order to obtain a positive result in short-term investment.

EDA Results – Univariate Analysis

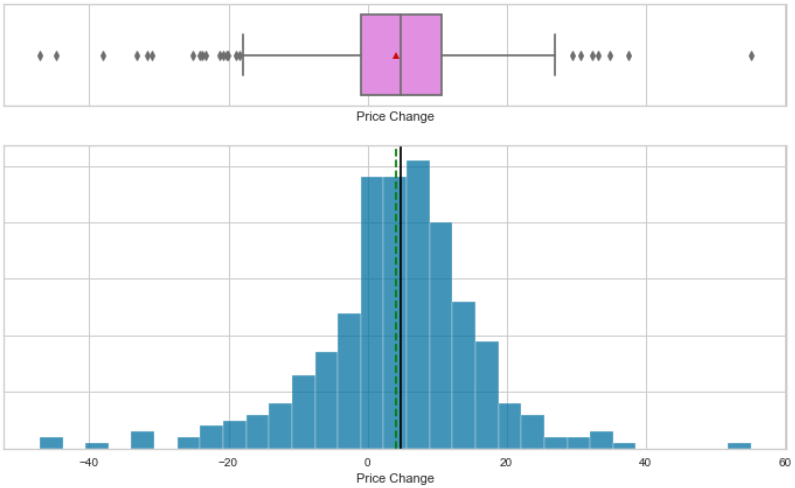
- Current Price

This variable shows the actual price of the stocks in the market(to the day of the study). Due to the spectrum of companies, it's expected to see prices over 200 and under this price. This variable shows a skewness to the right as the higher the value the uncommon it



- Price Change

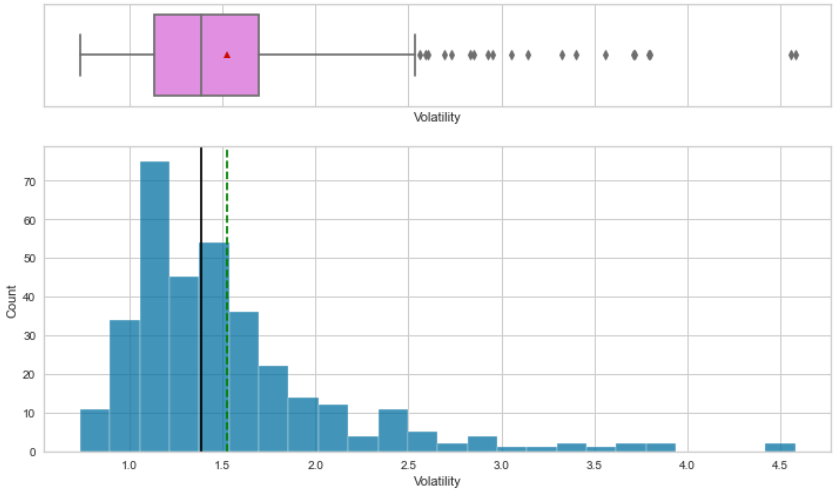
The change in price in stocks is a common situation. This distribution shows a bell shape that goes with a mean of about $ 5 positive in change. The worst change was over - $40 and the best for almost $55.



EDA Results – Univariate Analysis

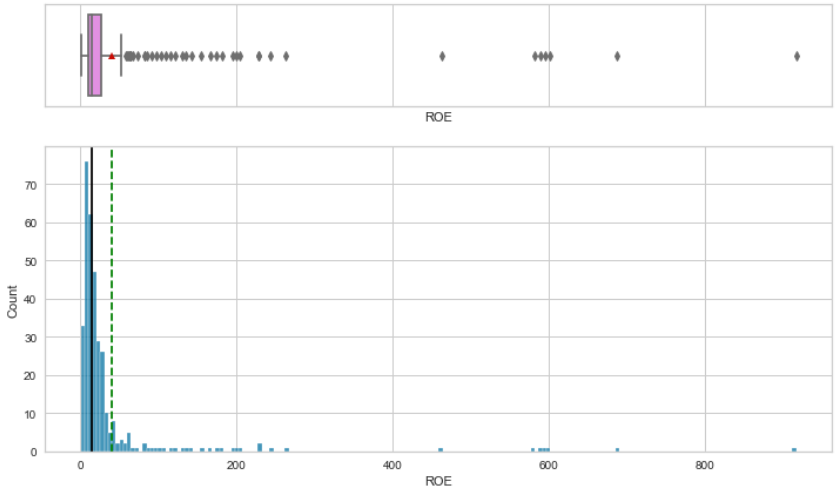
- Volatility

The volatility shows that the vast majority of movement is under 2.0 standard deviations and the mean is almost 1.5 times. This variable shows a skewness to the right.



- ROE

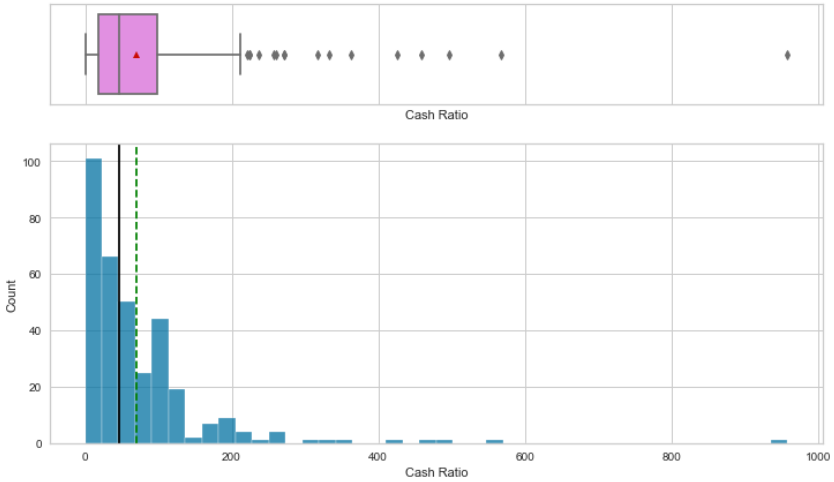
The ROE of the invested companies shows that are positive and generating benefits. The median of this measurement is way beyond the mean. In this case, the outliers are moving the mean.



EDA Results – Univariate Analysis

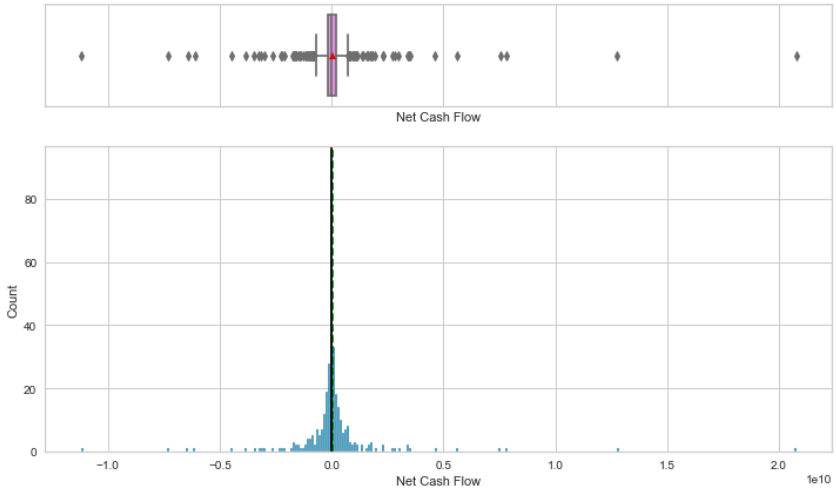
- Cash Ratio

This ratio shows that most of the stocks in the portfolio have a cash ratio under $100 and again one company is making the mean go over the mean. In this variable, the outliers are also moving the mean to the right.



- Net Cash Flow

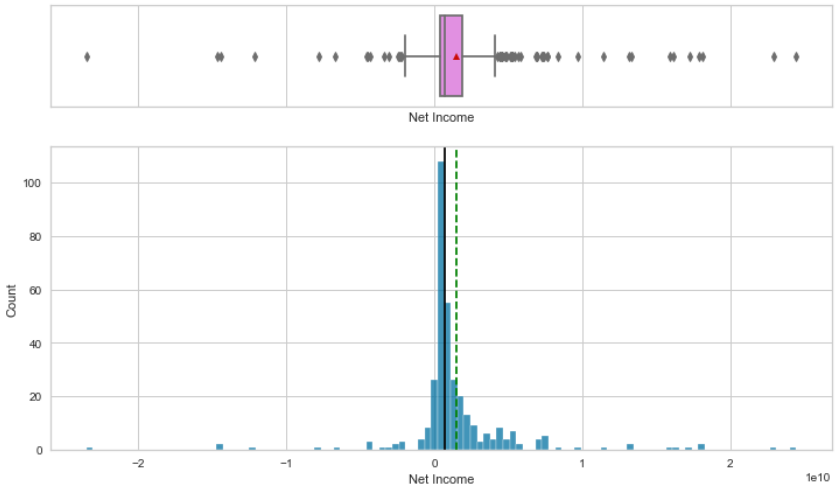
The cash flow shows that the vast majority of companies stay near 0 in the Net Cash Flow, due to the scale of the units. This makes this variable difficult to treat in its current state.



EDA Results – Univariate Analysis

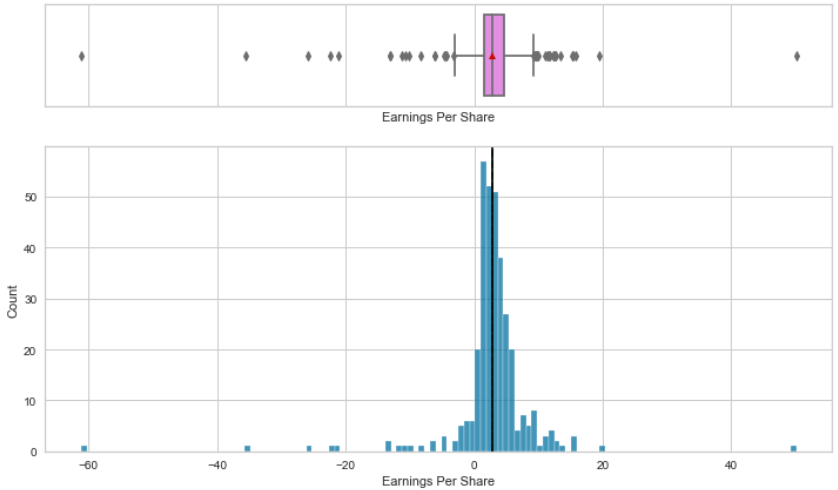
- Net Income

In a similar situation to Net Cash Flow, the highest values affect the interpretation of the variable.



- Earnings Per Share

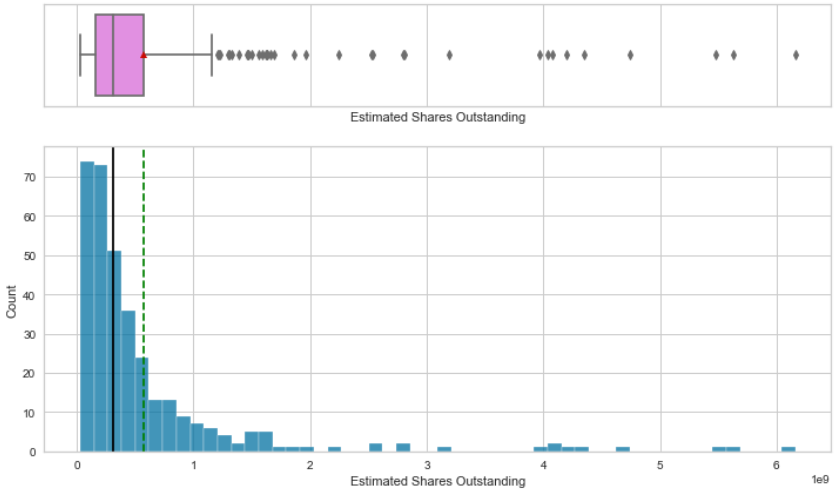
The earnings per share show outliers on both sides as the values concentrate around the mean.



EDA Results – Univariate Analysis

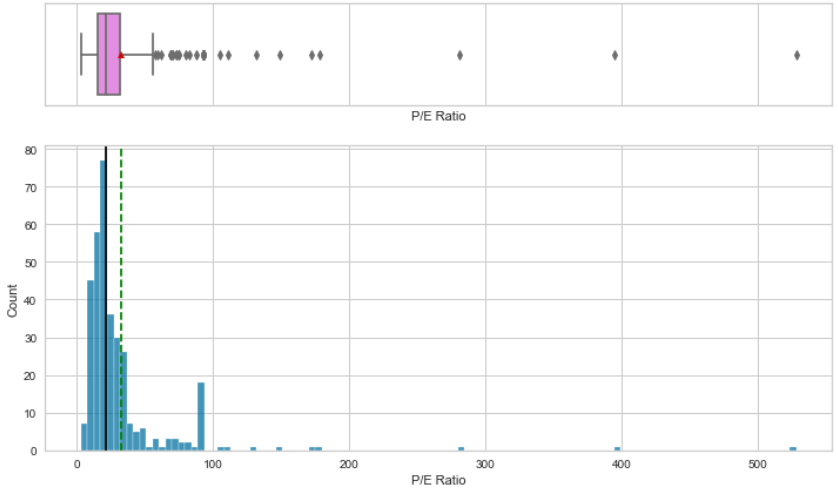
- Estimated Shares Outstanding

The amount of share shows a right-skewed variable as the outliers are over six times the mean values.



- P/E Ratio

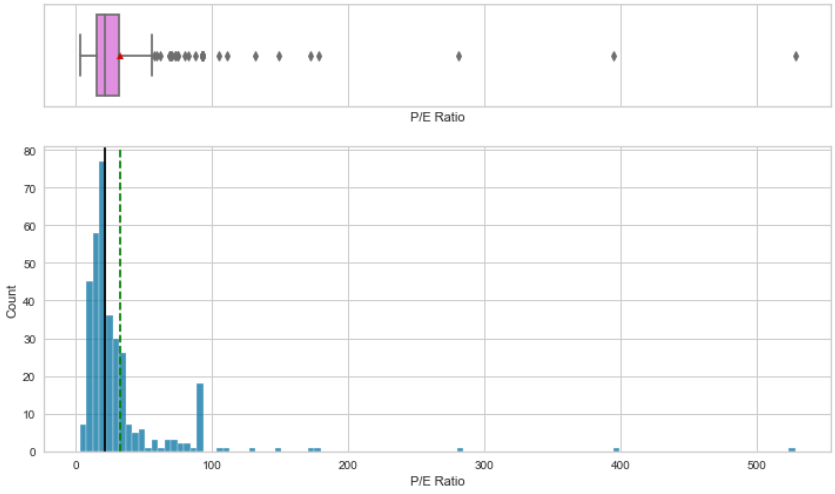
This variable also shows right-skewed variables due to the high number of outliers. The highest values are over 10 times the mean values.



EDA Results – Univariate Analysis

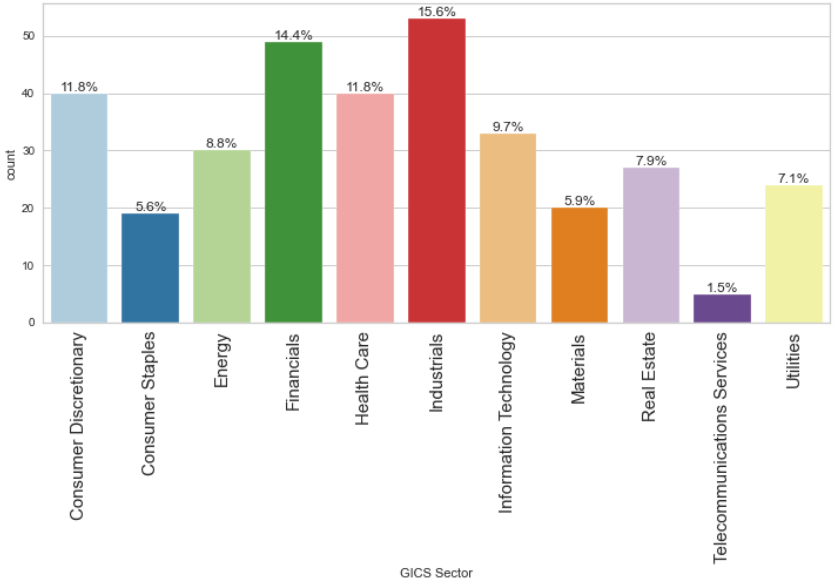
- P/B Ratio

As in the previous cases, the highest values cause a right skewness for this variable.



- GICS Sector

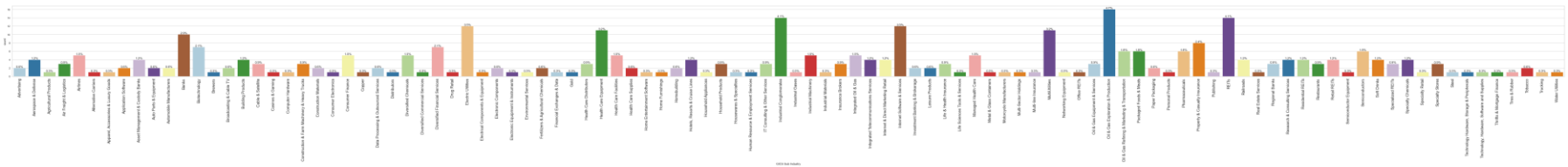
The variables show that the least common type of stock is telecommunication services. Also, this is related to several companies in these services compared to the rest of the industries.



EDA Results – Univariate Analysis

- GICS Sub Industry

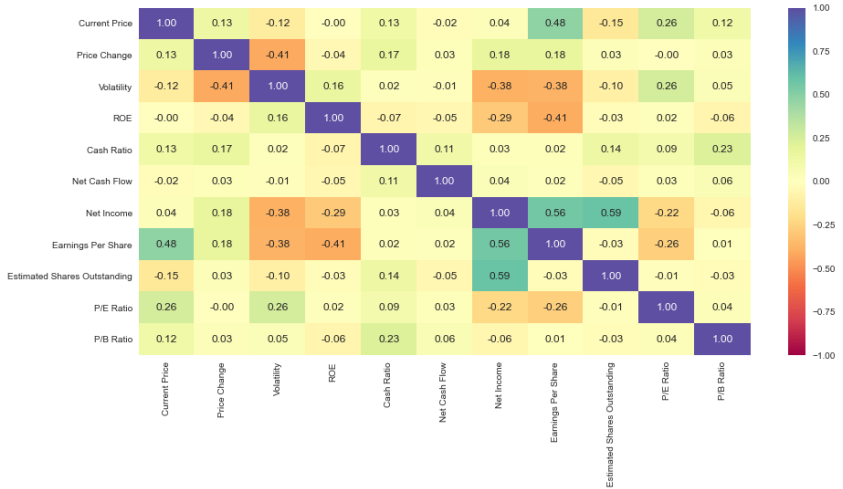
Due to the number of sub-industries, we can highlight how Oil & Gas Exploration & Production is the highest among all sub-industries with 4.7% of the total. Then RETs and Industrial conglomerates have a share of 4.1% each. The vast majority of the shares are under 1% and most of them are between 0.6% and 0.3%.



EDA Results – Bivariate Analysis

- Heatmap

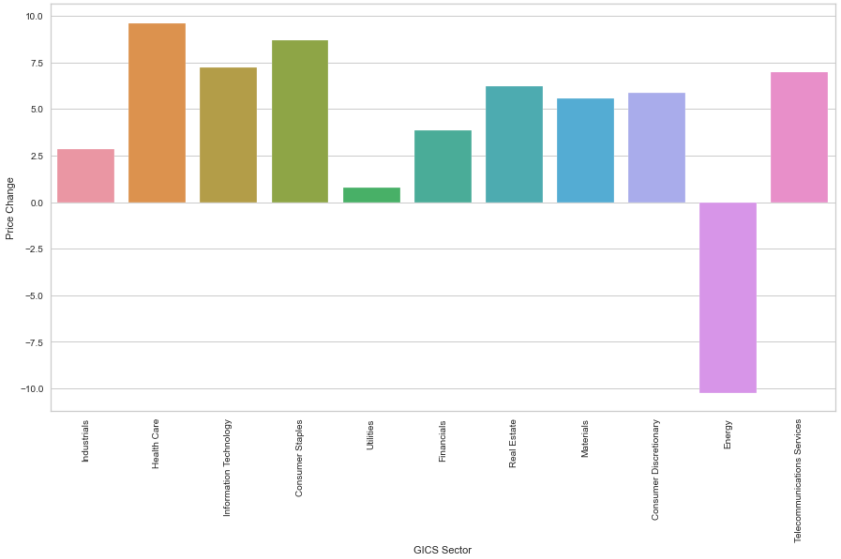
The heatmap shows the relationship between the numeric variables. Volatility and Price change show a negative correlation which is expected between this pair of variables. Then a higher relationship is shown between the Current Price and Earnings Per Share, due to the expected return for the shares with their higher prices. Next, a high correlation is shown between Net Income with Earning per Share and Estimated Shares Outstanding. Finally, ROE and earnings per Share should share a positive ratio as these two variables measures of profitability of a company.



EDA Results – Bivariate Analysis

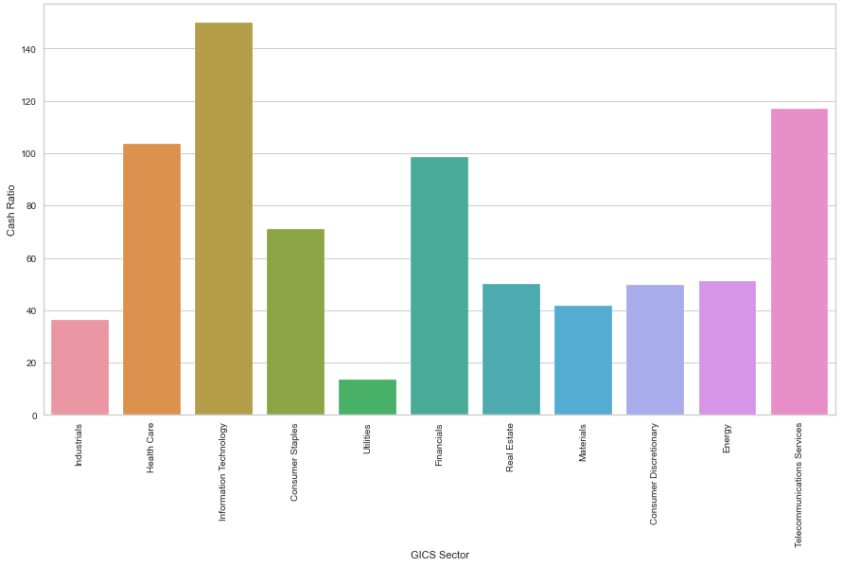
- Price Change vs Economic Sectors

Almost all sectors show a positive price change, except for the energy sector. This sector perhaps is affected by the trend in renewal energy, which can disincentive the usage of traditional energy.



- Cash ratio vs Economic Sectors

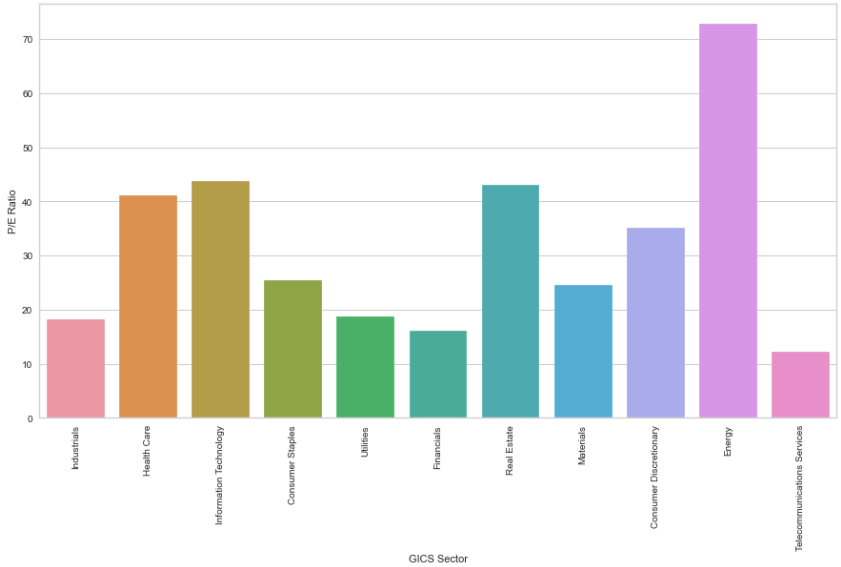
All sectors show an ability to cover the short-term obligations with cash and cash equivalent. The technology sector is the one with the highest cash ratio, perhaps this is due to the investment in R&D.



EDA Results – Bivariate Analysis

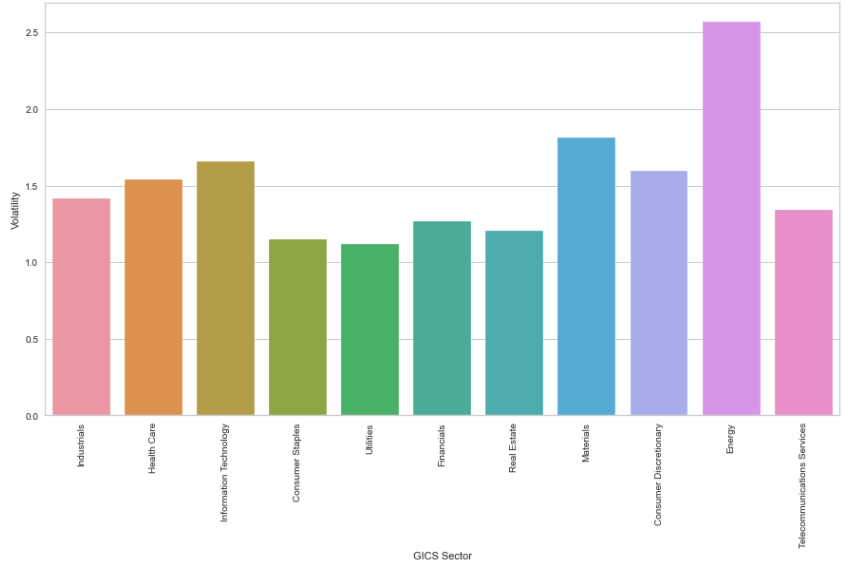
- P/E vs Economic Sectors

The P&E ratio shows that Telecommunications services are perhaps one of the best choices as it brings the highest return over the lowest investment.



- Volatility vs Economic Sectors

The utility sector is the one that shows more stability across all the sectors as it is the one closer to 1. The energy sector also shows to be the one with the highest volatility. Also, this can help to explain it higher P/E ratio.



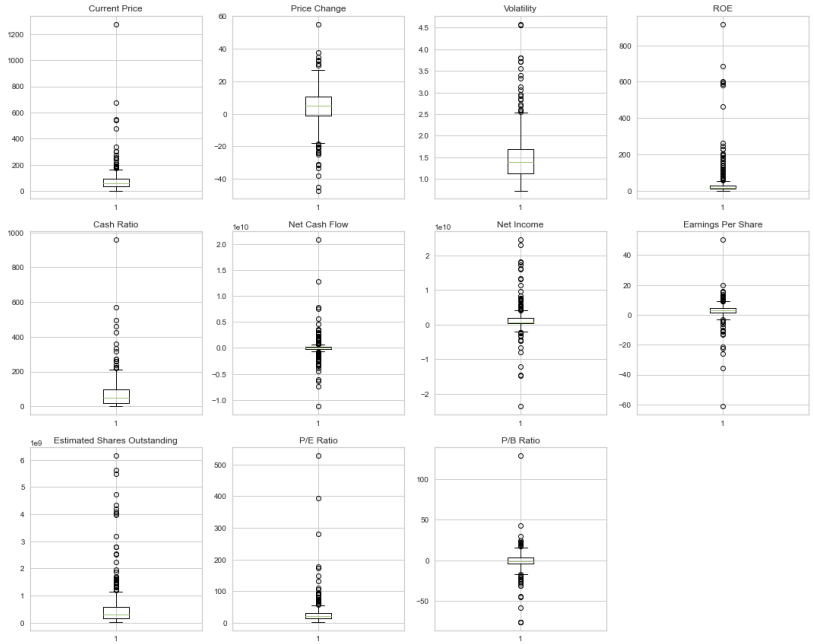
Data Preprocessing

- Duplicate value check: A duplicated Analysis was done through the dataset and it found that none of the records was a duplicate.
- Missing value treatment: A missing value treatment was realized and it found that no value was missing in the data set.

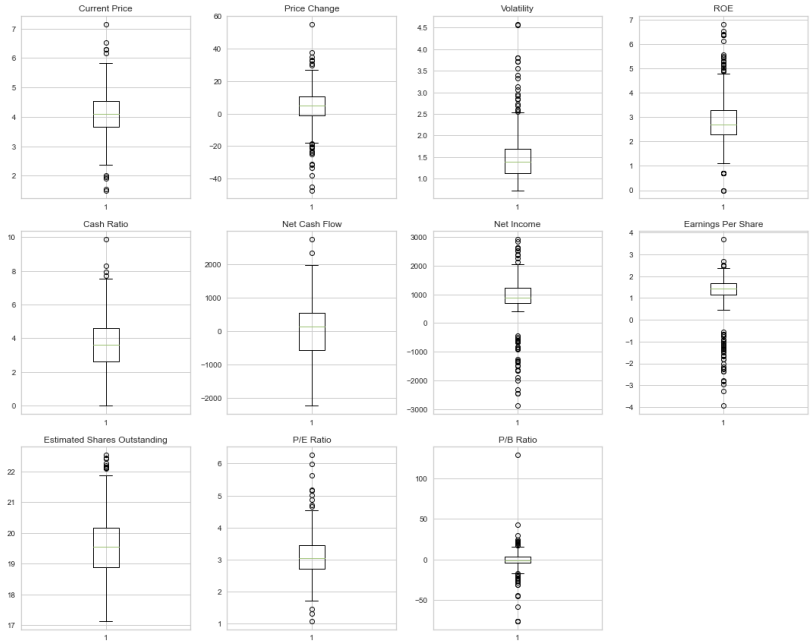
Data Preprocessing

- Outlier check: An outlier check was done in order to validate the data before a transformation.

Original Outliers



Transformation Outliers



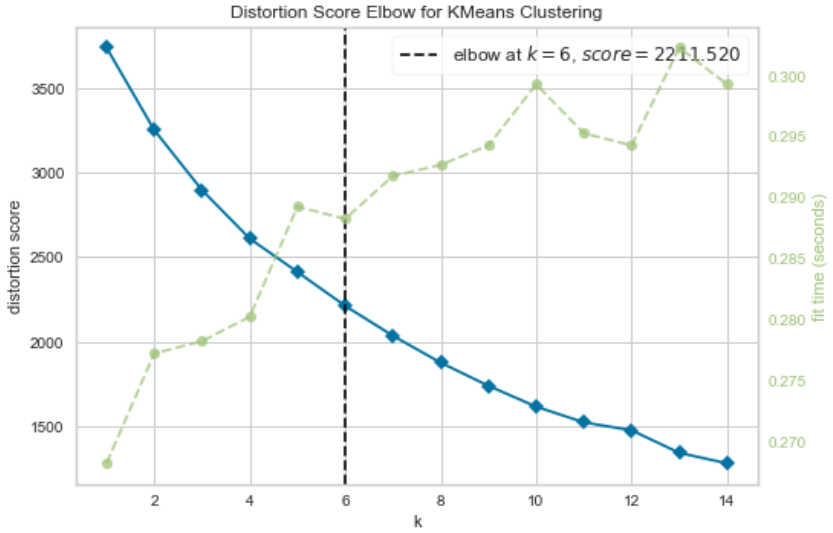
Data Preprocessing

- The previous transformation was done with the following procedure:
- Four variables were transformed with a cubic transformation as the results of the log would result in errors: Cash Ratio, Net Cash Flow and, Net Income.
- The model was evaluated with the log and without the log transformation of the data.
- Data preparation for modelling: To prepare the data for the modelling a Z-Score transformation was done to compare the values in unsupervised learning.

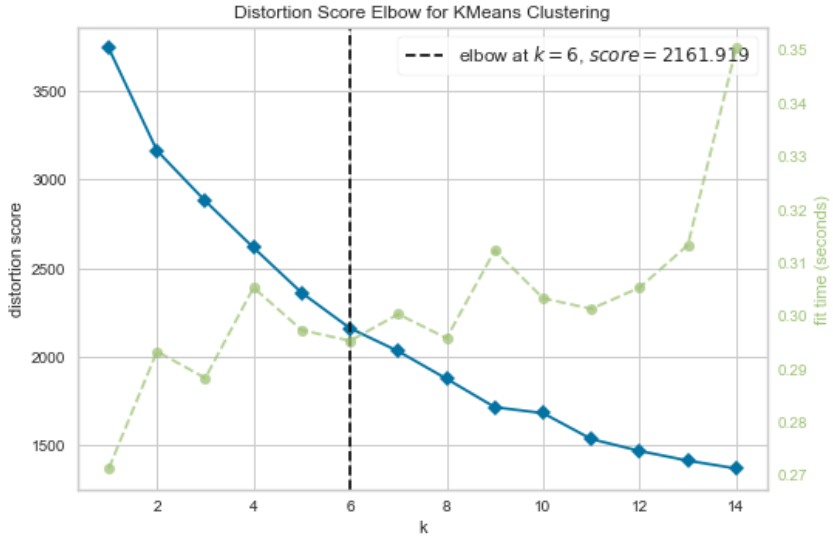
Model Performance Summary – K Means Clustering

- The following parameters were obtained from the models:
  - Without Log transformations: K-Means: 6 clusters
  - With log transformations: K-Means: 6 clusters

Without Log Transformations



With Log Transformations

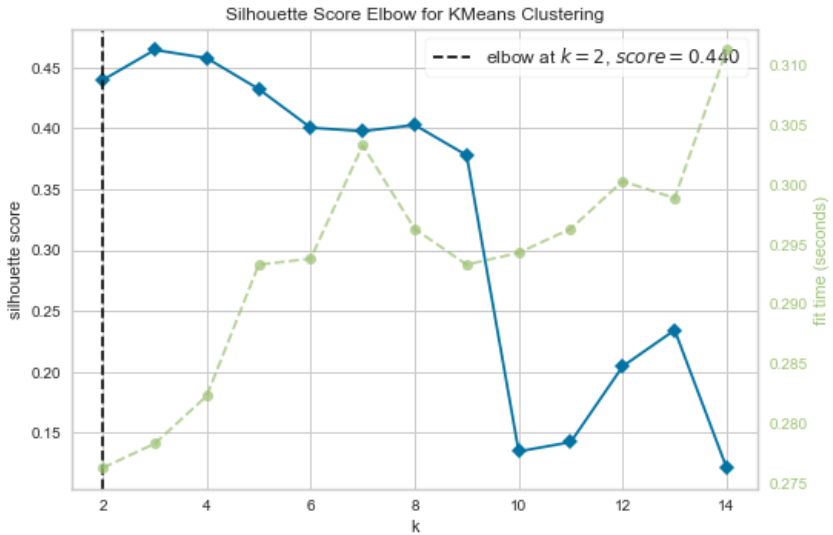


Link to Appendix slide on model assumptions

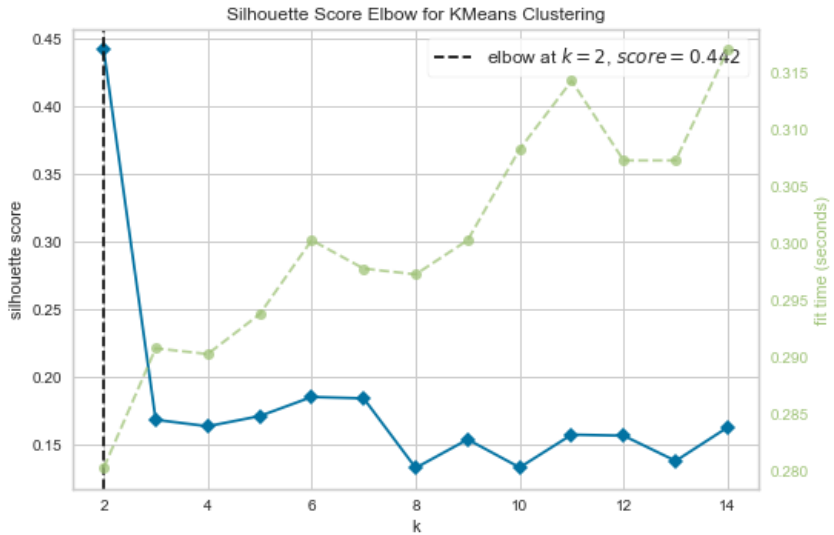
Model Performance Summary – K Means Clustering

- The following Silhouette scores were obtained from the models:

Without Log Transformations



With Log Transformations



Link to Appendix slide on model assumptions

Model Performance Summary – K Means Clustering

- This technique brought 6 clusters for each model. The model with the log transformation brought two large groups against a single large group of the model without the log transformation.

Cluster Number	K Means W/O Log	K Means W Log
0	9	27
1	267	96
2	3	7
3	27	11
4	11	194
5	23	5

- The complete list of cluster profiling is in the appendix in order to see the elements in complete detail.

Link to Appendix slide on model assumptions

Model Performance Summary – Hierarchical Clustering

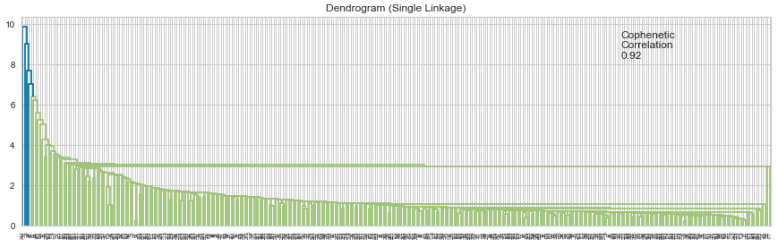
- The following parameters were obtained from the models:
  - Without Log transformations: Hierarchical: 7 clusters.
  - With log transformations: Hierarchical: 8 clusters.

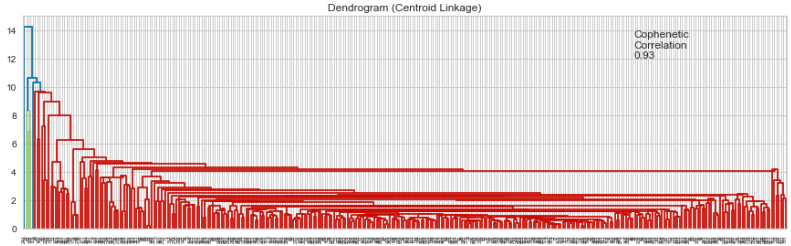
- The cluster technique that took less time was the K-means method. This is because the Hierarchical must do many iterations in each run. The Hierarchical run took over 55.39 seconds to complete and the K-means took over 23 seconds. Also, this technique brought more clusters than the K-means one.
- In both methodologies the distance metrics applied were: Euclidean, Chebyshev, Mahalanobis and, Cityblock. Each of the also applied the following linkage method: single, complete, average, and weighted. In both cases, the Euclidean distance with the average method brought the best Cophenetic Correlation.

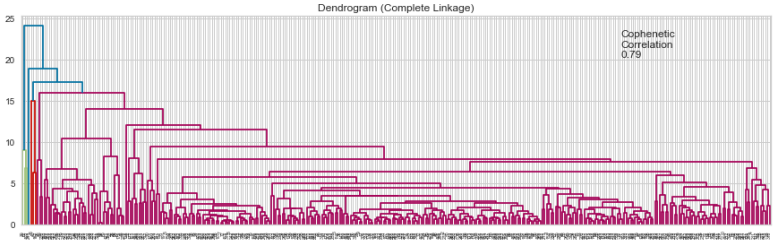
Link to Appendix slide on model assumptions

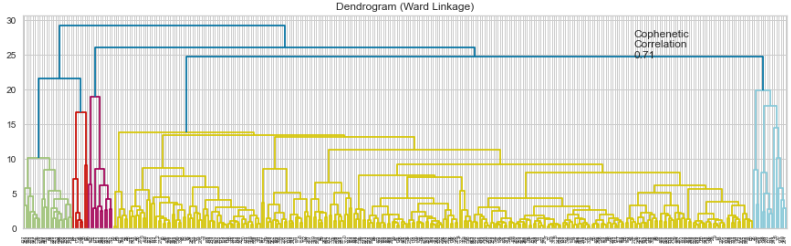
Hierarchical Clustering – Model Without log transformation

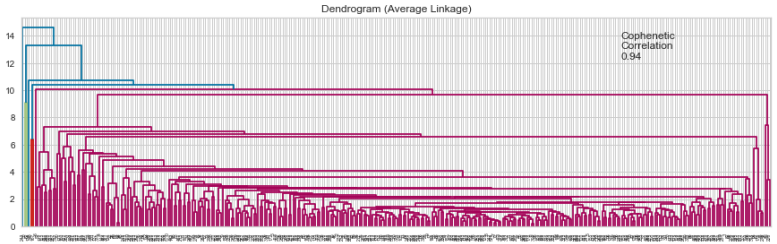
Dendrograms

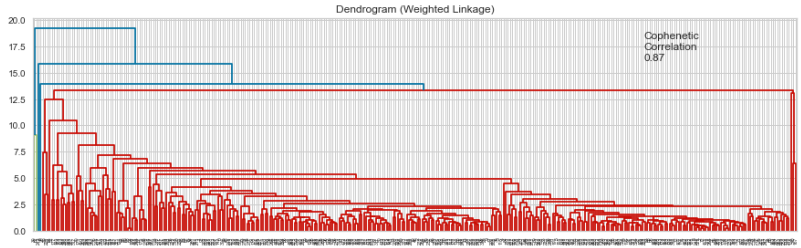






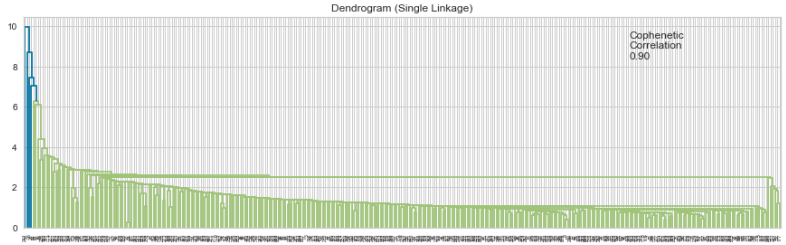


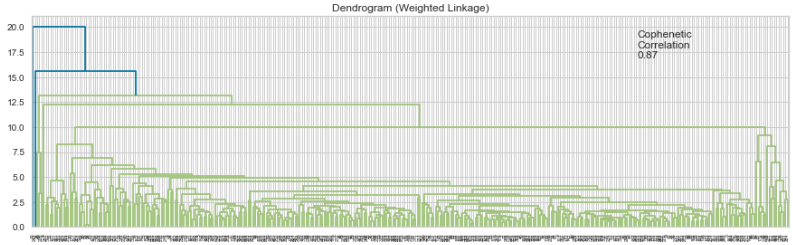


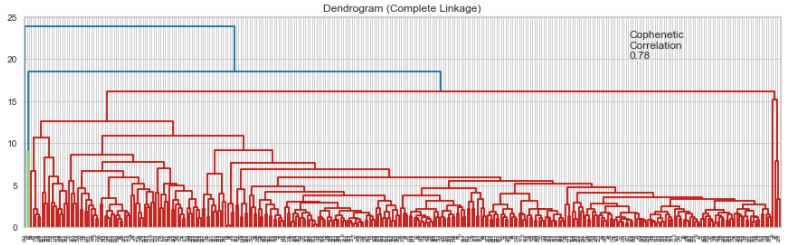


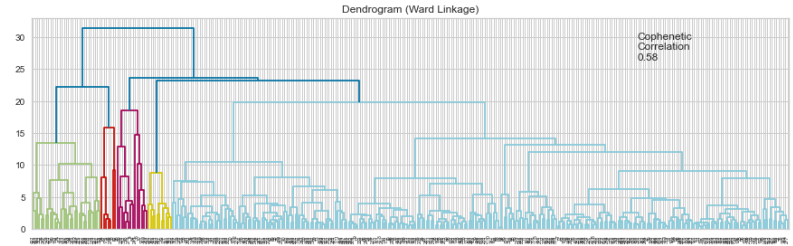
Hierarchical Clustering – Model With log transformation

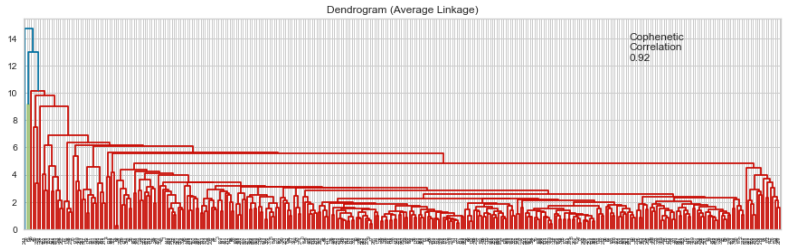
Dendrograms

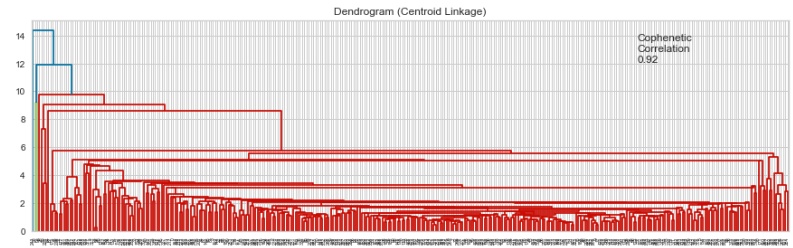












Model Performance Summary – Hierarchical Clustering

- In both cases the cophenetic relation ship brough 0.94 for the model without log transformation and 0.92 for the model with the log transformation. This technique brought a selection of clusters of 7 and 8 as is seen in the table below.

Cluster Number	Hierarchical W/O Log	Hierarchical W/O Log
0	2	6
1	2	31
2	3	198
3	1	11
4	1	2
5	1	79
6	330	8
7	N/A	5

- The complete list of cluster profiling is in the appendix in order to see the elements in complete detail.

Link to Appendix slide on model assumptions

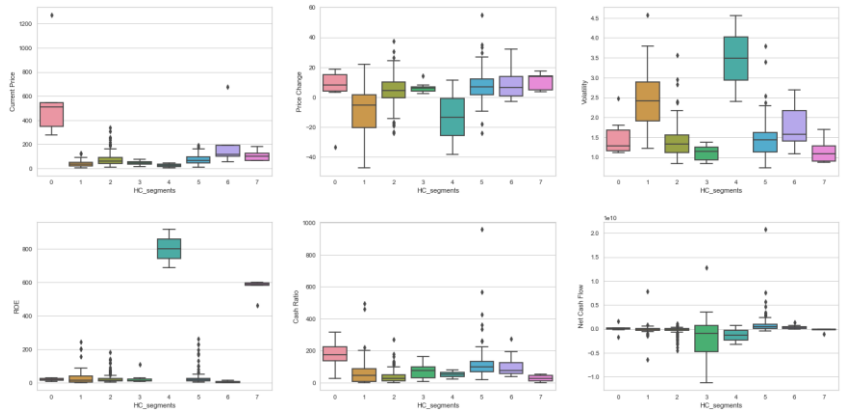
Model Performance Summary – K-means vs Hierarchical Clustering

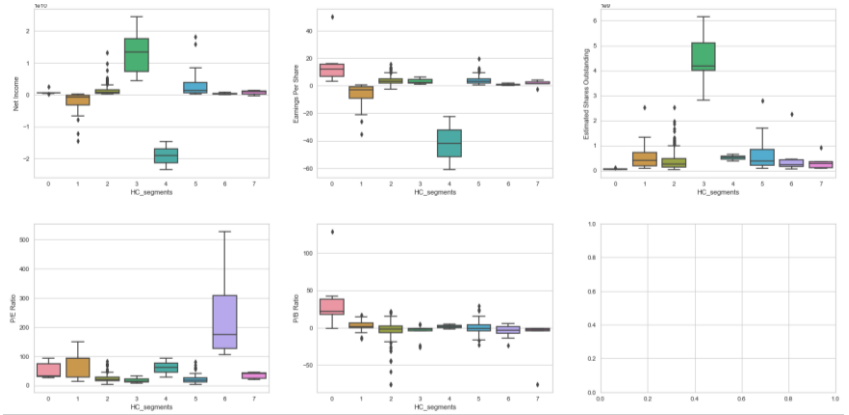
- As the results show the K-means brought 6 clusters and the Hierarchical Clustering brought 7 and 8 clusters. Between these, the selected cluster is the one of Hierarchical Cluster with the log transformations. This is due to the distribution of these clusters and elements corresponding to each element.
- The cluster 4 is composed of 2 energy sector companies. This is a constant in the formation of the clusters as there were at least 1 or 2 companies that made a cluster of the energy sector.

Link to Appendix slide on model assumptions

Model Assumptions – Model Without log transformation

8 Clusters



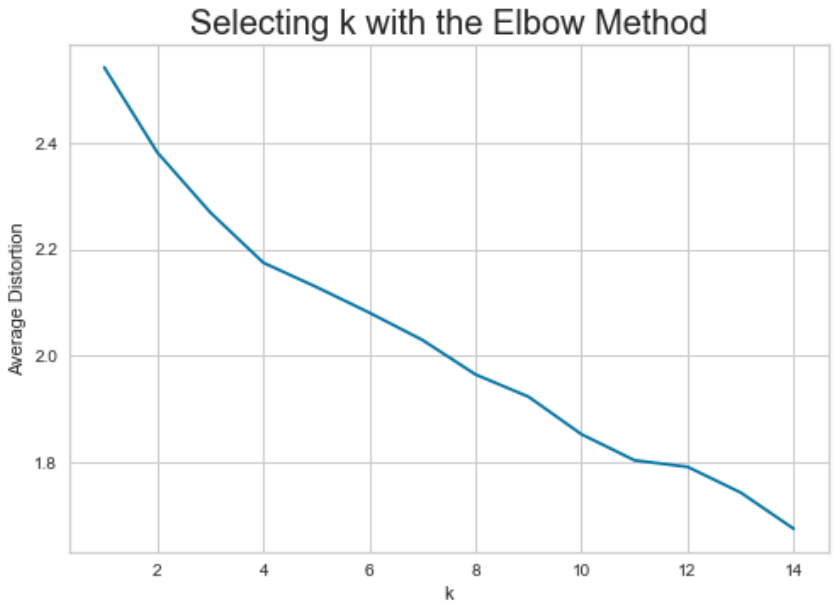


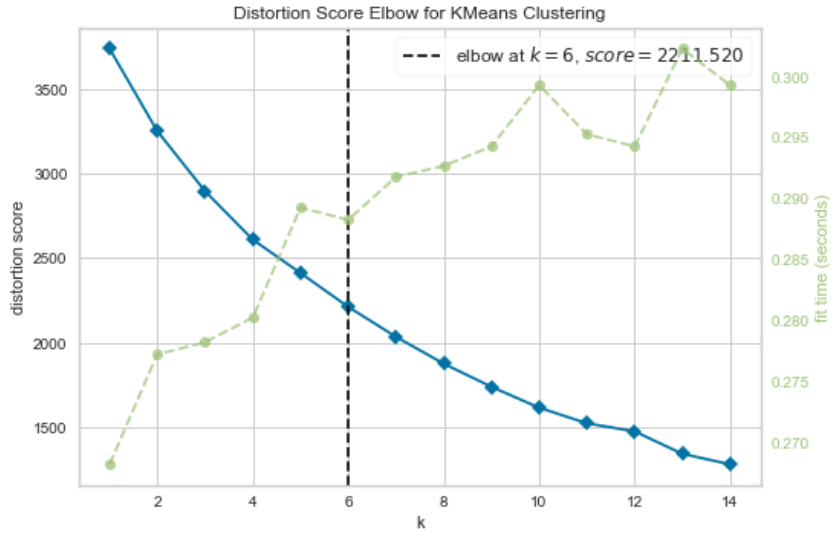
APPENDIX

Data Background and Contents - Dictonary

- **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)

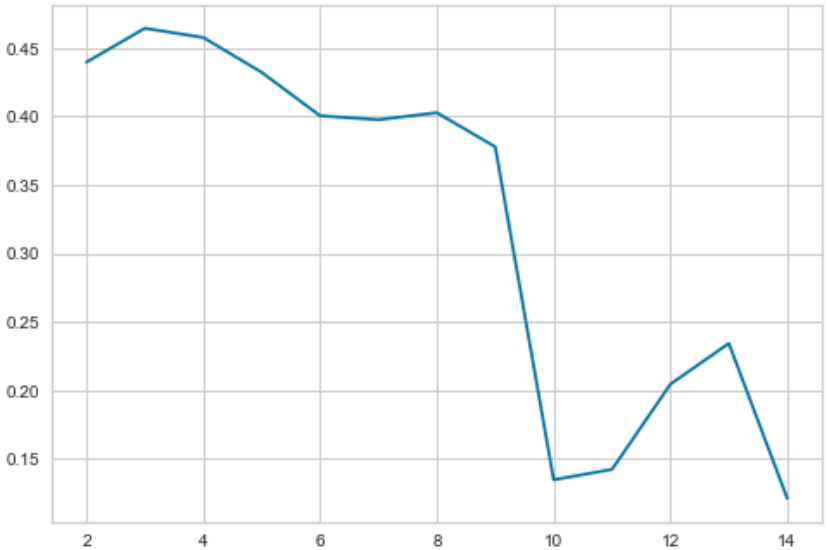
Model Assumptions – Model Without log transformation

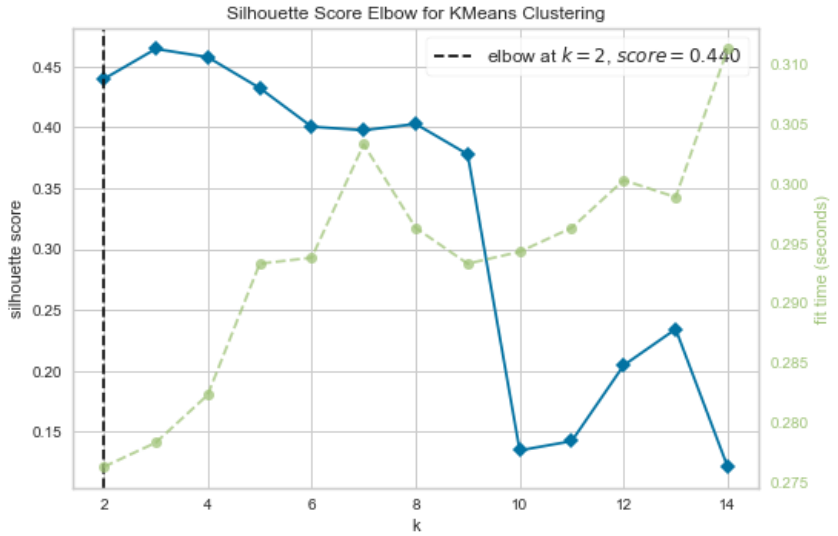




Model Assumptions – Model Without log transformation

Silhouette scores





Model Assumptions – Model Without log transformation

6 Clusters

KM_segments	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	count_in_each_segment
0	527.129995	8.680613	1.499723	24.666667	140.555556	181796333.333333	1059105888.888889	14.333333	118984485.275556	97.539861	26.728638	9
1	72.196386	4.983796	1.368419	35.220974	50.149813	-2768119.850187	1510231314.606742	3.656124	429600138.717041	23.514365	-3.639469	267
2	26.990000	-14.060688	3.296307	603.000000	57.333333	-585000000.000000	-17555666666.666668	-39.726667	481910081.666667	71.528835	1.638633	3
3	37.282919	-14.529500	2.820301	40.666667	47.555556	-133624777.777778	-1904442925.925926	-4.957037	503635899.112592	86.787432	1.378738	27
4	50.517273	5.747586	1.130399	31.090909	75.909091	-1072272727.272727	14833090909.090910	4.154545	4298826628.727273	14.803577	-4.552119	11
5	79.534493	15.175743	1.804140	25.565217	298.347826	1527981782.608696	1576087391.304348	2.008696	786284564.973913	52.669163	6.735099	23

Model Assumptions – Model Without log transformation

6 Clusters – Average Link

KM_segments	GICS Sector	
0	Consumer Discretionary	3
	Health Care	4
	Information Technology	1
	Real Estate	1
1	Consumer Discretionary	33
	Consumer Staples	17
	Energy	5
	Financials	45
	Health Care	28
	Industrials	51
	Information Technology	19
	Materials	17
	Real Estate	26
	Telecommunications Services	2
	Utilities	24
2	Energy	3
3	Energy	20
	Industrials	2
	Information Technology	3
	Materials	2
4	Consumer Discretionary	1
	Consumer Staples	1
	Energy	1
	Financials	3
	Health Care	2
	Information Technology	1
	Telecommunications Services	2
5	Consumer Discretionary	3
	Consumer Staples	1
	Energy	1
	Financials	1
	Health Care	6
	Information Technology	9
	Materials	1
	Telecommunications Services	1

Model Assumptions – Model Without log transformation

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 0.792530720285.

Cophenetic correlation for Mahalanobis distance and average linkage is 0.9247324030159736.

Cophenetic correlation for Mahalanobis distance and weighted linkage is 0.8708317490180427.

Cophenetic correlation for Cityblock distance and single linkage is 0.9334186366528576.

Cophenetic correlation for Cityblock distance and complete linkage is 0.7375328863205819.

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

Model Assumptions – Model Without log transformation

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 0.9314012446828155.

Cophenetic correlation for ward linkage is 0.7101180299865353.

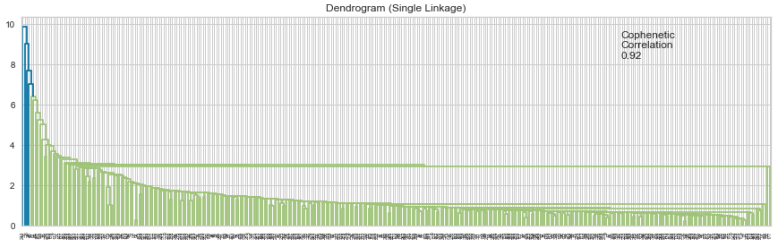
Cophenetic correlation for weighted linkage is 0.8693784298129404.

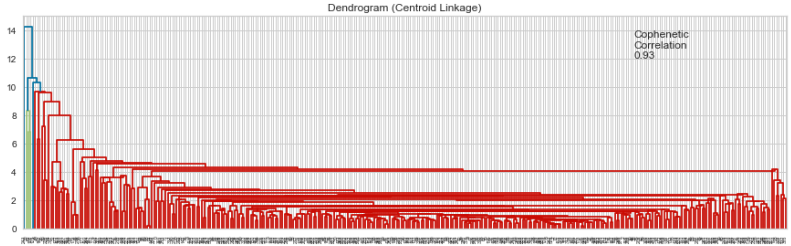
****
****

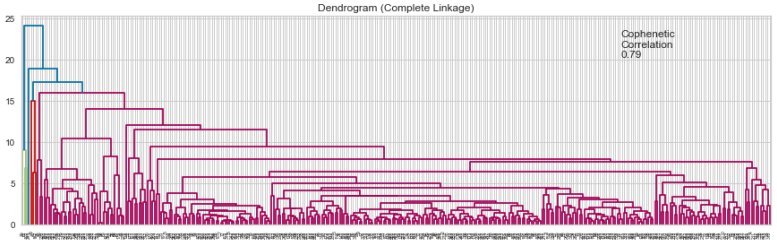
The highest cophenetic correlation is **0.9422540609560814**, which is obtained with **average** linkage.

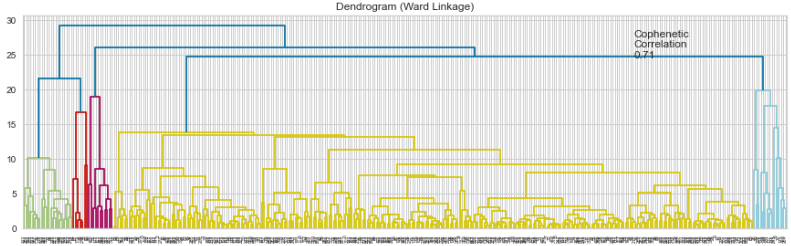
Model Assumptions – Model Without log transformation

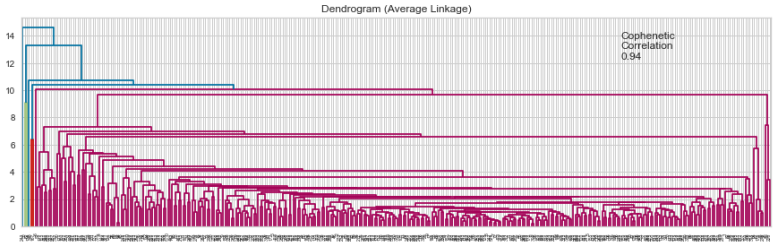
Dendrograms

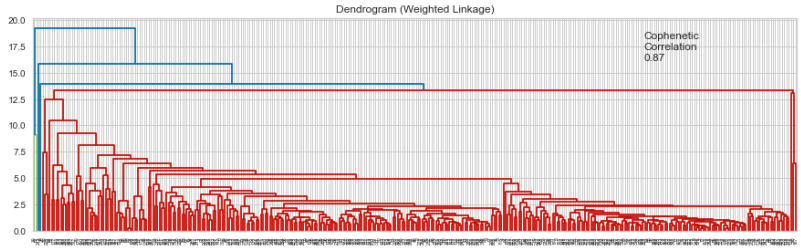












Model Assumptions – Model Without log transformation

7 Clusters – Average Link

KM_segments	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	count_in_each_segment
0	24.485001	-13.351992	3.482611	802.000000	51.000000	-1292500000.000000	-19106500000.000000	-41.815000	519573983.250000	60.748608	1.565141	2
1	25.640000	11.237908	1.322355	12.500000	130.500000	16755500000.000000	13654000000.000000	3.295000	2791829362.100000	13.649696	1.508484	2
2	327.006671	21.917380	2.029752	4.000000	106.000000	698240666.666667	287547000.000000	0.750000	366763235.300000	400.989188	-5.322376	3
3	104.660004	16.224320	1.320606	8.000000	958.000000	592000000.000000	3669000000.000000	1.310000	2800763359.000000	79.893133	5.884467	1
4	1274.949951	3.190527	1.268340	29.000000	184.000000	-1671386000.000000	2551360000.000000	50.090000	50935516.070000	25.453183	-1.052429	1
5	276.570007	6.189286	1.116976	30.000000	25.000000	90885000.000000	596541000.000000	8.910000	66951851.850000	31.040405	129.064585	1
6	75.017416	3.937751	1.513415	35.621212	66.545455	-39846757.575758	1549443100.000000	2.904682	562266326.402576	29.091275	-2.146308	330

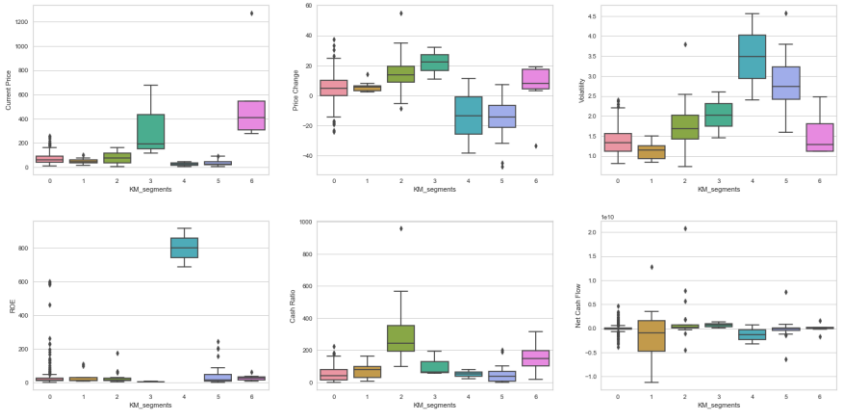
Model Assumptions – Model Without log transformation

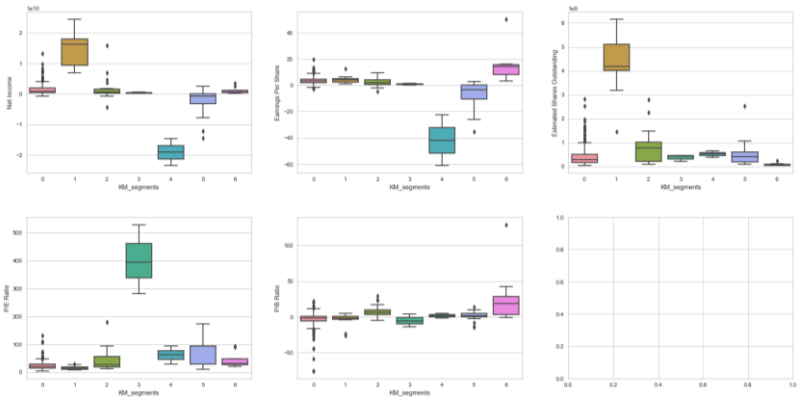
7 Clusters – Average Link

HC_segments	GICS Sector	
0	Energy	2
1	Financials	1
	Information Technology	1
2	Consumer Discretionary	1
	Health Care	1
	Information Technology	1
3	Information Technology	1
4	Consumer Discretionary	1
5	Information Technology	1
6	Consumer Discretionary	38
	Consumer Staples	19
	Energy	28
	Financials	48
	Health Care	39
	Industrials	53
	Information Technology	29
	Materials	20
	Real Estate	27
	Telecommunications Services	5
	Utilities	24

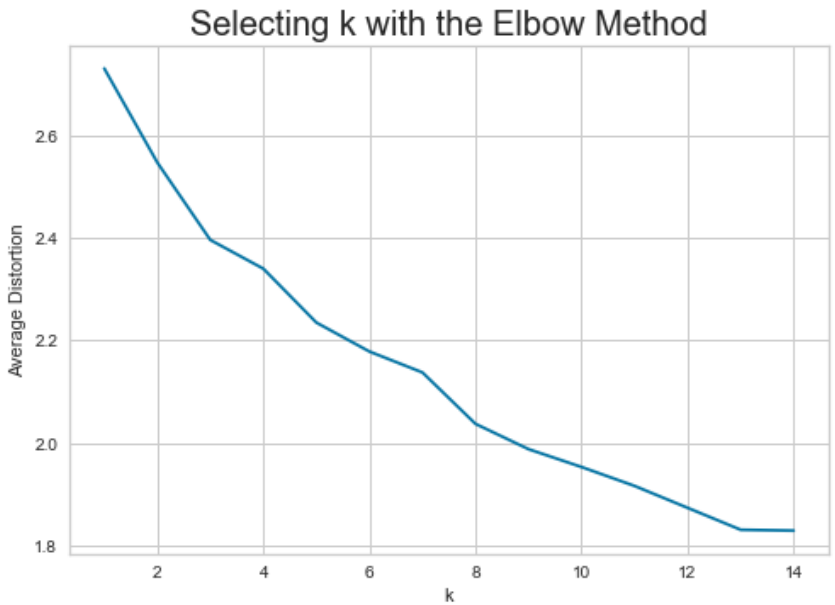
Model Assumptions – Model Without log transformation

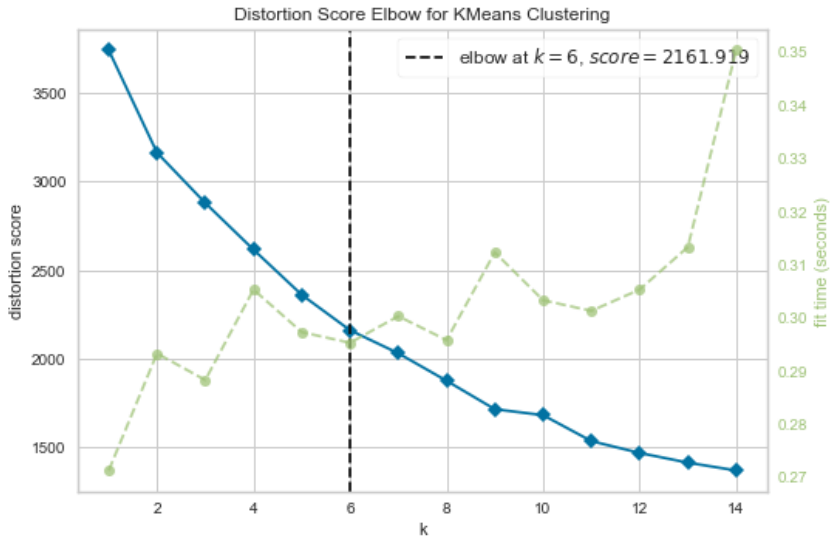
7 Clusters





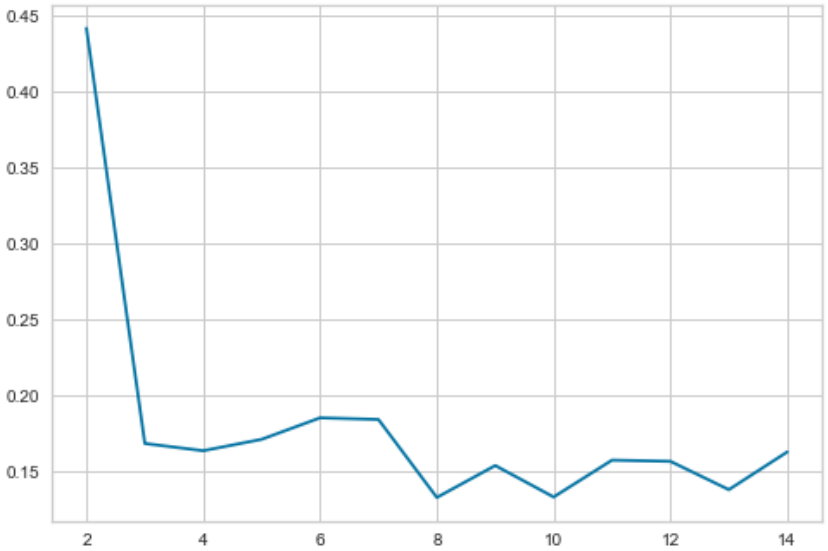
Model Assumptions – Model With log transformation

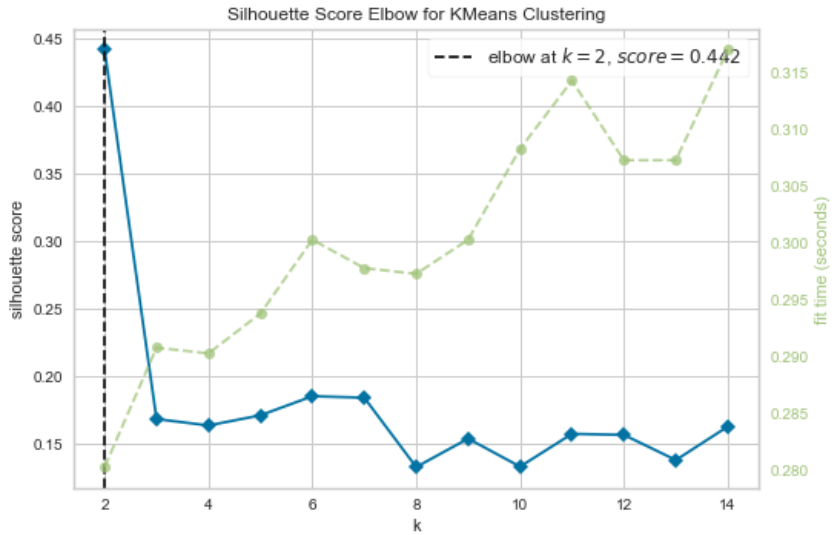




Model Assumptions – Model Without log transformation

Silhouette scores





Model Assumptions – Model With log transformation

6 Clusters

KM_segments	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	count_in_each_segment
0	38.489215	-13.297160	2.763743	48.222222	69.962963	-90620592.592593	-2761534851.851852	-6.655556	527598513.460741	90.395011	2.051778	27
1	81.398681	8.634281	1.500829	29.677083	129.729167	1020183333.333333	2247899468.750000	4.071823	552551395.469062	29.614417	1.613840	96
2	84.355716	3.854981	1.827670	633.571429	33.571429	-568400000.000000	-4968157142.857142	-10.841429	398169036.442857	42.284541	-11.589502	7
3	46.120000	6.142327	1.110799	23.090909	68.818182	-1230545454.545455	13590727272.727272	3.173636	4421329193.909091	17.099041	-5.228699	11
4	72.279305	4.041815	1.376225	23.386598	38.907216	-303392216.494845	1274665664.948454	3.531649	395824232.287165	23.779825	-3.809067	194
5	703.943988	7.611139	1.626217	17.200000	185.000000	-47032600.000000	969563600.000000	17.742000	137567111.218000	141.451007	16.613966	5

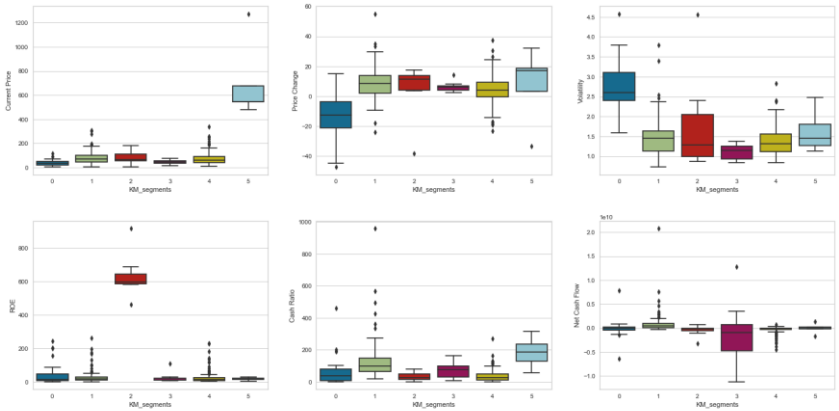
Model Assumptions – Model Without log transformation

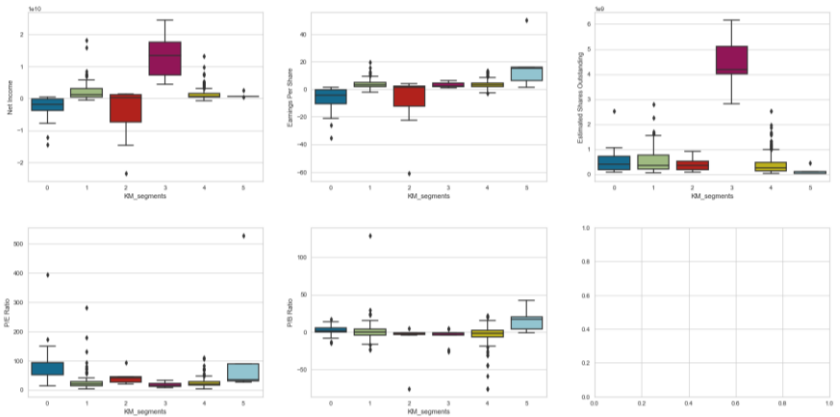
6 Clusters – Average Link

HC_segments	GICS Sector	
0	Energy	21
	Industrials	2
	Information Technology	3
	Materials	1
1	Consumer Discretionary	6
	Consumer Staples	7
	Energy	1
	Financials	24
	Health Care	21
	Industrials	10
	Information Technology	17
	Materials	4
	Real Estate	3
	Telecommunications Services	2
	Utilities	1
2	Consumer Discretionary	1
	Consumer Staples	2
	Energy	2
	Financials	1
	Industrials	1
3	Consumer Discretionary	1
	Consumer Staples	1
	Energy	1
	Financials	3
	Health Care	2
	Information Technology	1
	Telecommunications Services	2
4	Consumer Discretionary	29
	Consumer Staples	9
	Energy	5
	Financials	21
	Health Care	15
	Industrials	40
	Information Technology	12
	Materials	15
	Real Estate	24
	Telecommunications Services	1
	Utilities	23
5	Consumer Discretionary	3
	Health Care	2

Model Assumptions – Model With log transformation

6 Clusters





Model Assumptions – Model With log transformation

Hierarchical Clustering

Cophenetic correlation for Euclidean distance and single linkage is 0.8996292255487788.

Cophenetic correlation for Euclidean distance and complete linkage is 0.776055851803682.

Cophenetic correlation for Euclidean distance and average linkage is 0.9203326877548739.

Cophenetic correlation for Euclidean distance and weighted linkage is 0.87374538079358.

Cophenetic correlation for Chebyshev distance and single linkage is 0.8785023423897879.

Cophenetic correlation for Chebyshev distance and complete linkage is 0.7249969604046684.

Cophenetic correlation for Chebyshev distance and average linkage is 0.9037963551926433.

Cophenetic correlation for Chebyshev distance and weighted linkage is 0.8676212968421244.

Cophenetic correlation for Mahalanobis distance and single linkage is 0.8950606025209297.

Cophenetic correlation for Mahalanobis distance and complete linkage is 0.7848939996697087.

Cophenetic correlation for Mahalanobis distance and average linkage is 0.9058194588808304.

Cophenetic correlation for Mahalanobis distance and weighted linkage is 0.8612919355222921.

Cophenetic correlation for Cityblock distance and single linkage is 0.9117569013032729.

Cophenetic correlation for Cityblock distance and complete linkage is 0.6938389524104923.

Cophenetic correlation for Cityblock distance and average linkage is 0.8896615370534131.

Cophenetic correlation for Cityblock distance and weighted linkage is 0.6600597459291137.

****

The highest cophenetic correlation is **0.9203326877548739**, which is obtained with **Euclidean** distance and **average** linkage.

Model Assumptions – Model With log transformation

Euclidean Distance

Cophenetic correlation for single linkage is 0.8996292255487788.

Cophenetic correlation for complete linkage is 0.776055851803682.

Cophenetic correlation for average linkage is 0.9203326877548739.

Cophenetic correlation for weighted linkage is 0.87374538079358.

Cophenetic correlation for ward linkage is 0.5797121852479153.

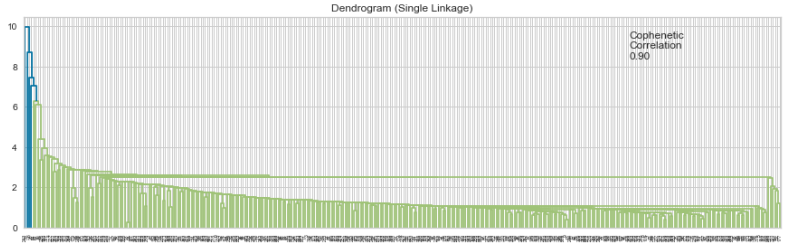
Cophenetic correlation for centroid linkage is 0.9187273316992014.

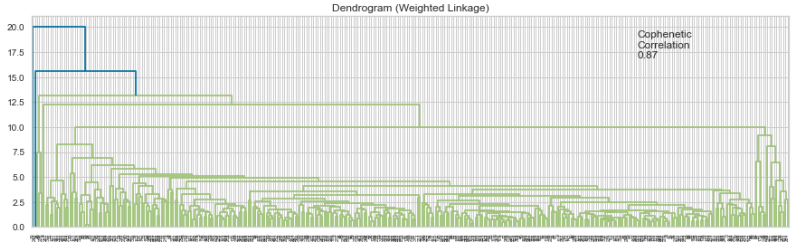
****
****

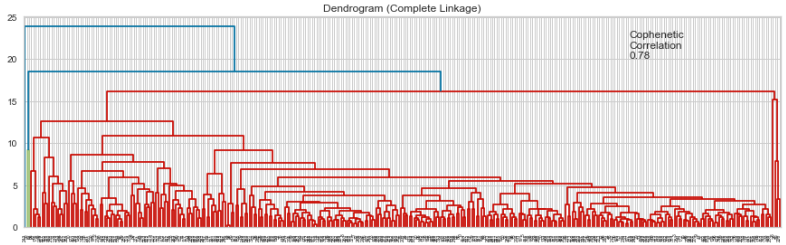
The highest cophenetic correlation is **0.9203326877548739**, which is obtained with **average** linkage.

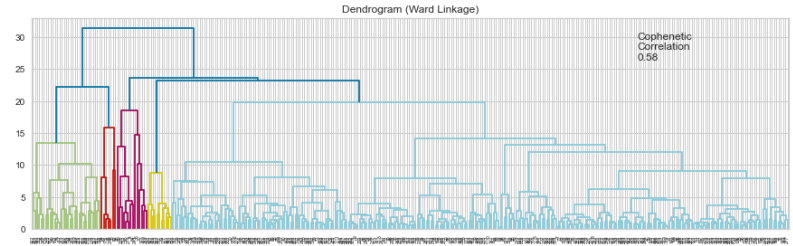
Model Assumptions – Model With log transformation

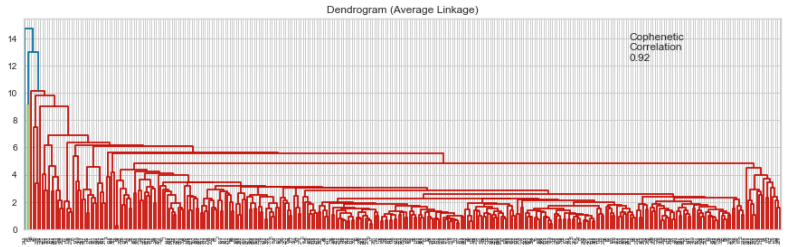
Dendrograms

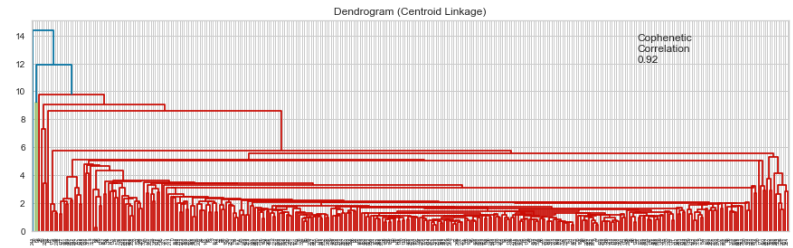












Model Assumptions – Model With log transformation

8 Clusters – Ward Link

HC_segments	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	count_in_each_segment
0	570.466654	3.666088	1.515959	19.833333	176.000000	23440500.000000	839355500.000000	16.598333	57823170.553333	50.550419	38.681119	6
1	39.567283	-8.877878	2.453617	42.387097	85.419355	-107099322.580645	-2517176967.741935	-6.234516	543417231.779032	65.400062	2.627772	31
2	74.000395	4.364337	1.394020	22.065657	40.232323	-272490171.717172	1307582141.414141	3.741111	387537048.586010	23.161291	-3.768200	198
3	46.120000	6.142327	1.110799	23.090909	68.818182	-1230545454.545455	13590727272.727272	3.173636	4421329193.909091	17.099041	-5.228699	11
4	24.485001	-13.351992	3.482611	802.000000	51.000000	-1292500000.000000	-19106500000.000000	-41.815000	519573983.250000	60.748608	1.565141	2
5	70.376118	7.658644	1.497114	37.101266	130.240506	1146219240.506329	2598393050.632911	4.193481	592786124.590506	22.615611	0.345930	79
6	191.776252	9.510192	1.796697	5.125000	107.125000	372345125.000000	273987875.000000	0.837500	501564722.140000	237.614616	-5.041878	8
7	108.304002	10.737770	1.165694	566.200000	26.600000	-278760000.000000	687180000.000000	1.548000	349607057.720000	34.898915	-16.851358	5

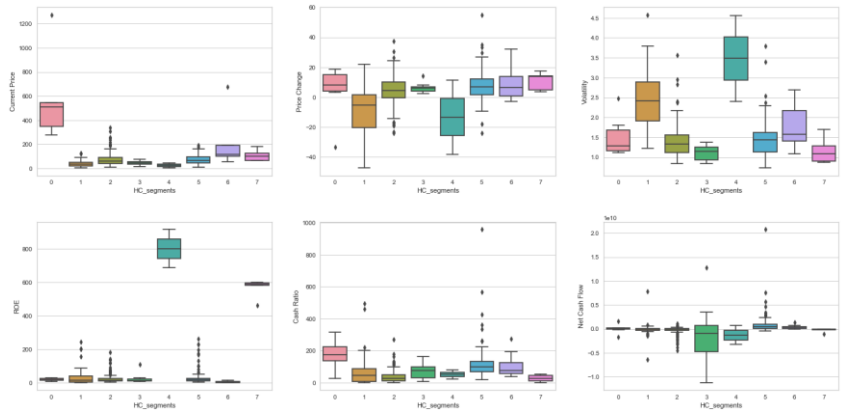
Model Assumptions – Model Without log transformation

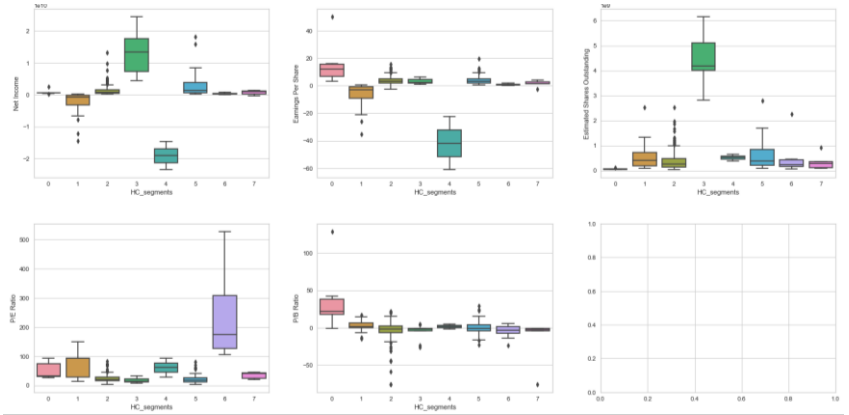
7 Clusters – Average Link

0	Consumer Discretionary	2
	Health Care	2
	Information Technology	1
	Real Estate	1
1	Energy	19
	Health Care	2
	Industrials	3
	Information Technology	2
	Materials	1
	Real Estate	1
	Telecommunications Services	1
	Utilities	2
2	Consumer Discretionary	30
	Consumer Staples	10
	Energy	6
	Financials	21
	Health Care	15
	Industrials	41
	Information Technology	13
	Materials	16
	Real Estate	24
	Telecommunications Services	1
	Utilities	21
3	Consumer Discretionary	1
	Consumer Staples	1
	Energy	1
	Financials	3
	Health Care	2
	Information Technology	1
	Telecommunications Services	2
4	Energy	2
5	Consumer Discretionary	5
	Consumer Staples	6
	Energy	1
	Financials	24
	Health Care	16
	Industrials	8
	Information Technology	14
	Materials	3
	Telecommunications Services	1
	Utilities	1
6	Consumer Discretionary	1
	Energy	1
	Health Care	3
	Information Technology	2
	Real Estate	1
7	Consumer Discretionary	1
	Consumer Staples	2
	Financials	1
	Industrials	1

Model Assumptions – Model Without log transformation

8 Clusters





Great Learning

Happy Learning !