

Trade & Ahead

Data Analysis Observations & Recommendation

John D Noble

February 2022

- I. Business Problem Overview & Solution Approach
- II. Data Description
- III. Exploratory Data Analysis (EDA)
 - I. GICS Sector & Subsector Distribution
 - II. Bivariate Correlation Analysis
- IV. Key Metrics
 - I. Price Change X GICS Sector
 - II. Cash Ratio X GICS Sector
 - III. P/E X GICS Sector
 - IV. Volatility X GICS Sector
 - V. Outlier Check
- V. Scaling
- VI. Modeling Approaches - K-Means & Hierarchical Clustering
- VII. K-means Clustering
- VIII. Hierarchical Clustering
 - I. Hierarchical Clustering – Linkage “Average” to diversify the clusters
- IX. Data Insights
- X. Business Insights and Recommendations
- XI. APPENDIX 1
 - I. K-Means Portfolio 1
 - II. Hierarchical Clustering Portfolio 1
 - III. Hierarchical Clustering Portfolio 2
 - IV. Data Summary

Business Problem Overview & Solution Approach

This analysis is intended to use unsupervised machine learning techniques to classify stocks across different market segments and help protect against risks that could make the portfolio vulnerable to losses.

Current State

- The stock market has consistently proven to be a good place to invest in and save for the future.
- Steady returns on investments through the power of compound interest, the earlier one starts investing, the larger the corpus one can have for retirement.
- Historically, while historical returns do not guarantee future success, the key to managing downside and upside potential is to have a balanced portfolio across a broad basket of stocks.
- Single stock investing carries more inherent risk, and by diversifying within a GICS Sector and/or across non-correlated sectors and investors can construct a balanced portfolio.

What are we trying to solve for?

Can we leverage Unsupervised Machine Learning that can help in cluster like stocks together and shortlist other portfolio that have different risk/return characteristics.

Key Questions We Will Answer!

1. Do we have the proper data?
2. What modeling approaches will yield the best results?
3. Can we find the “right” clusters of stocks and have the clusters significantly differentiated by both risk and return?

Financial Implications

The objective is to build various Unsupervised models (K-means and Hierarchical Clustering) to group stocks into uncorrelated clusters. By doing so the investor can maximize their risk/return profile and make appropriate diversified investments while minimize downside losses due to credit, market, and liquidity risk.

Data Description

Data Description

- New York Stock Exchange data e.g. stock price and some financial indicators

Data Dictionary

- 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)

Key Facts

Number of variables	15
Number of observations	340
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	40 KB
Variable types	
Categorical	4
Numeric	11
Average record size in memory	120.4 B

Data Preparation Made Before Modeling

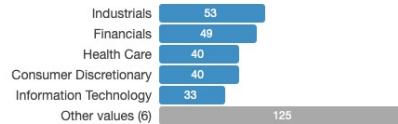
- We will use StandardScaler() to scale all the numerical variables to a Mean = 0 and STD = 1
- While outliers do exist, they are “normal” and will not be treated.

Data Description - EDA

GICS Sector

Categorical

Distinct	11
Distinct (%)	3.2%
Missing	0
Missing (%)	0.0%
Memory size	2.7 KIB



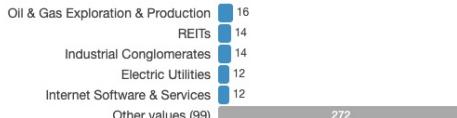
[Toggle details](#)

GICS Sub Industry

Categorical

HIGH CARDINALITY

Distinct	104
Distinct (%)	30.6%
Missing	0
Missing (%)	0.0%
Memory size	2.7 KIB



[Toggle details](#)

Current Price

Real number ($\mathbb{R}_{\geq 0}$)

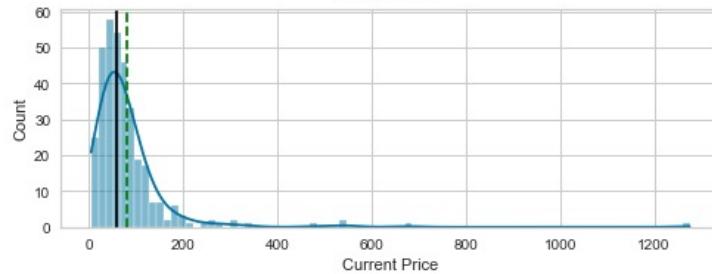
Distinct	336
Distinct (%)	98.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%

Mean	80.86234477
Minimum	4.5
Maximum	1274.949951
Zeros	0
Zeros (%)	0.0%
Memory size	2.7 KIB



[Toggle details](#)

Current Price is right skewed



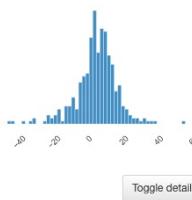
Data Description - EDA

Price Change

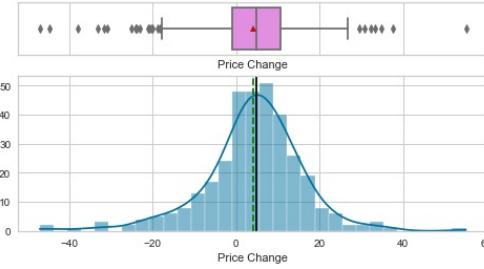
Real number (\mathbb{R})

UNIQUE

Distinct	340	Mean	4.078193812
Distinct (%)	100.0%	Minimum	-47.12969338
Missing	0	Maximum	55.05168339
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	2.7 KIB



Toggle details



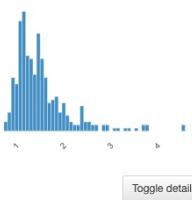
Normal

Volatility

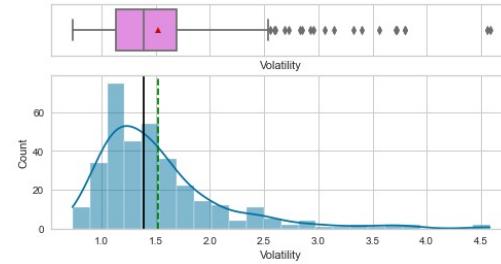
Real number ($\mathbb{R}_{\geq 0}$)

UNIQUE

Distinct	340	Mean	1.525976254
Distinct (%)	100.0%	Minimum	0.733163184
Missing	0	Maximum	4.58004173
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	2.7 KIB



Toggle details



Skew

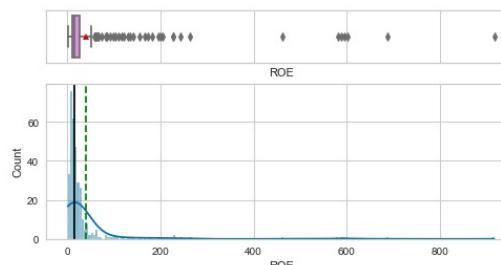
ROE

Real number ($\mathbb{R}_{\geq 0}$)

Distinct	82	Mean	39.59705882
Distinct (%)	24.1%	Minimum	1
Missing	0	Maximum	917
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	2.7 KIB



Toggle details



Skew

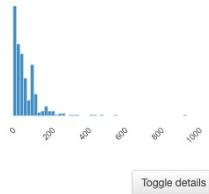
Data Description - EDA

Cash Ratio

Real number ($\mathbb{R}_{\geq 0}$)

ZEROS

Distinct	124	Mean	70.02352941
Distinct (%)	36.5%	Minimum	0
Missing	0	Maximum	958
Missing (%)	0.0%	Zeros	4
Infinite	0	Zeros (%)	1.2%
Infinite (%)	0.0%	Memory size	2.7 KIB



Toggle details



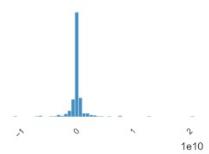
Cash Ratio

Skew

Net Cash Flow

Real number (\mathbb{R})

Distinct	332	Mean	55537620.59
Distinct (%)	97.6%	Minimum	-1.1208e+10
Missing	0	Maximum	2.0764e+10
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	2.7 KIB



Toggle details



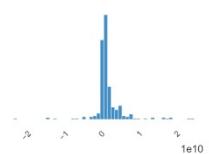
Net Cash Flow

Normal

Net Income

Real number (\mathbb{R})

Distinct	335	Mean	1494384603
Distinct (%)	98.5%	Minimum	-2.3528e+10
Missing	0	Maximum	2.4442e+10
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	2.7 KIB



Toggle details



Net Income

Normal

Data Description - EDA

Earnings Per Share

Real number ($\mathbb{R}_{\geq 0}$)

Distinct	268	Mean	2.776661765
Distinct (%)	78.8%	Minimum	-61.2
Missing	0	Maximum	50.09
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	2.7 kB

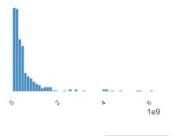


Toggle details

Estimated Shares Outstanding

Real number ($\mathbb{R}_{\geq 0}$)

Distinct	323	Mean	577028337.8
Distinct (%)	95.0%	Minimum	27672156.66
Missing	0	Maximum	6159292035
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	2.7 kB

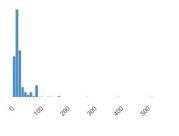


Toggle details

P/E Ratio

Real number ($\mathbb{R}_{\geq 0}$)

Distinct	295	Mean	32.61256332
Distinct (%)	86.8%	Minimum	2.935450768
Missing	0	Maximum	528.0390742
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	2.7 kB

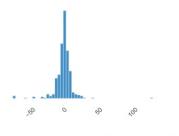


Toggle details

P/B Ratio

Real number (\mathbb{R})

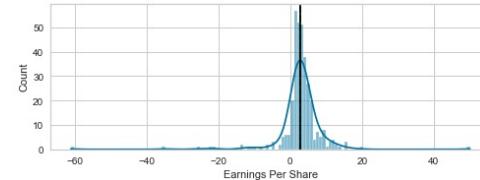
Distinct	299	Mean	-1.718249363
Distinct (%)	87.9%	Minimum	-76.11907749
Missing	0	Maximum	129.0645854
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	2.7 kB



Toggle details



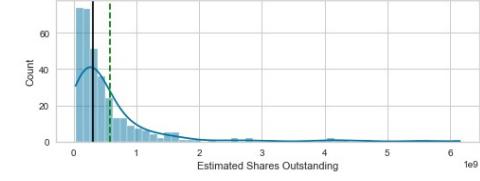
Earnings Per Share



Normal



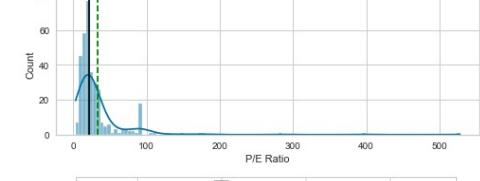
Estimated Shares Outstanding



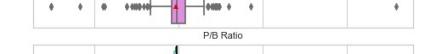
Skew



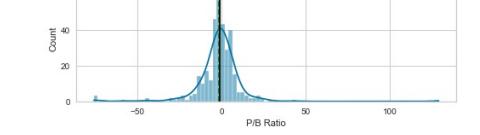
P/E Ratio



Skew



P/B Ratio

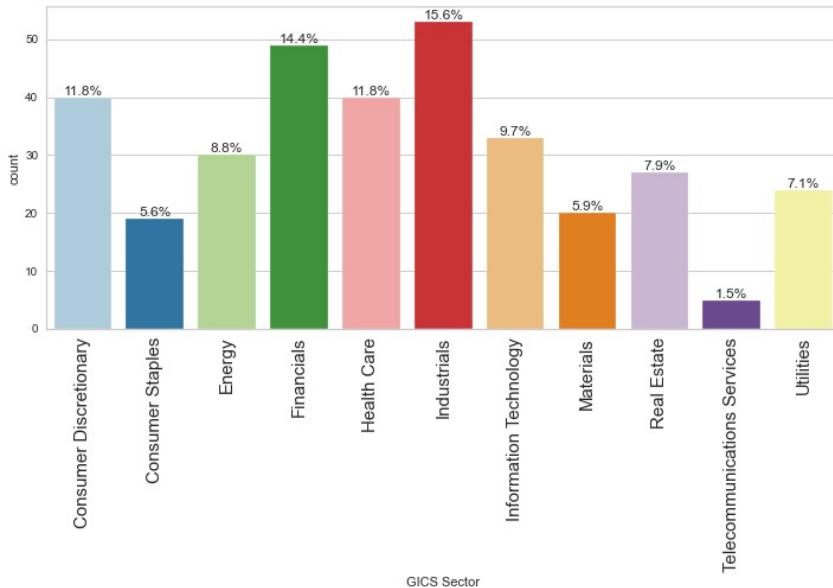


Normal

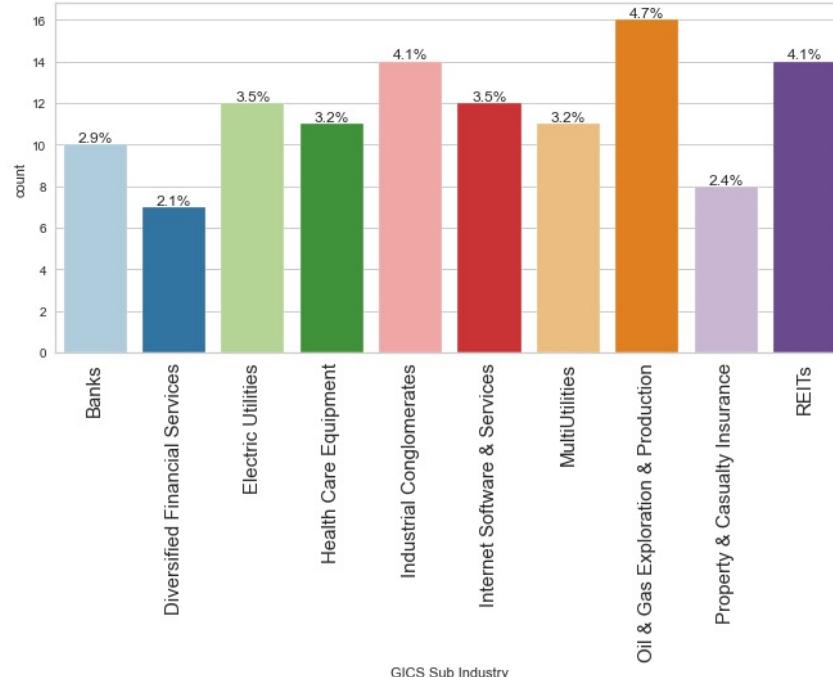
Exploratory Data Analysis (EDA)

Investors have many sector and sub-sector options to choose from when creating an optimal portfolio.

The GICS Sector distribution

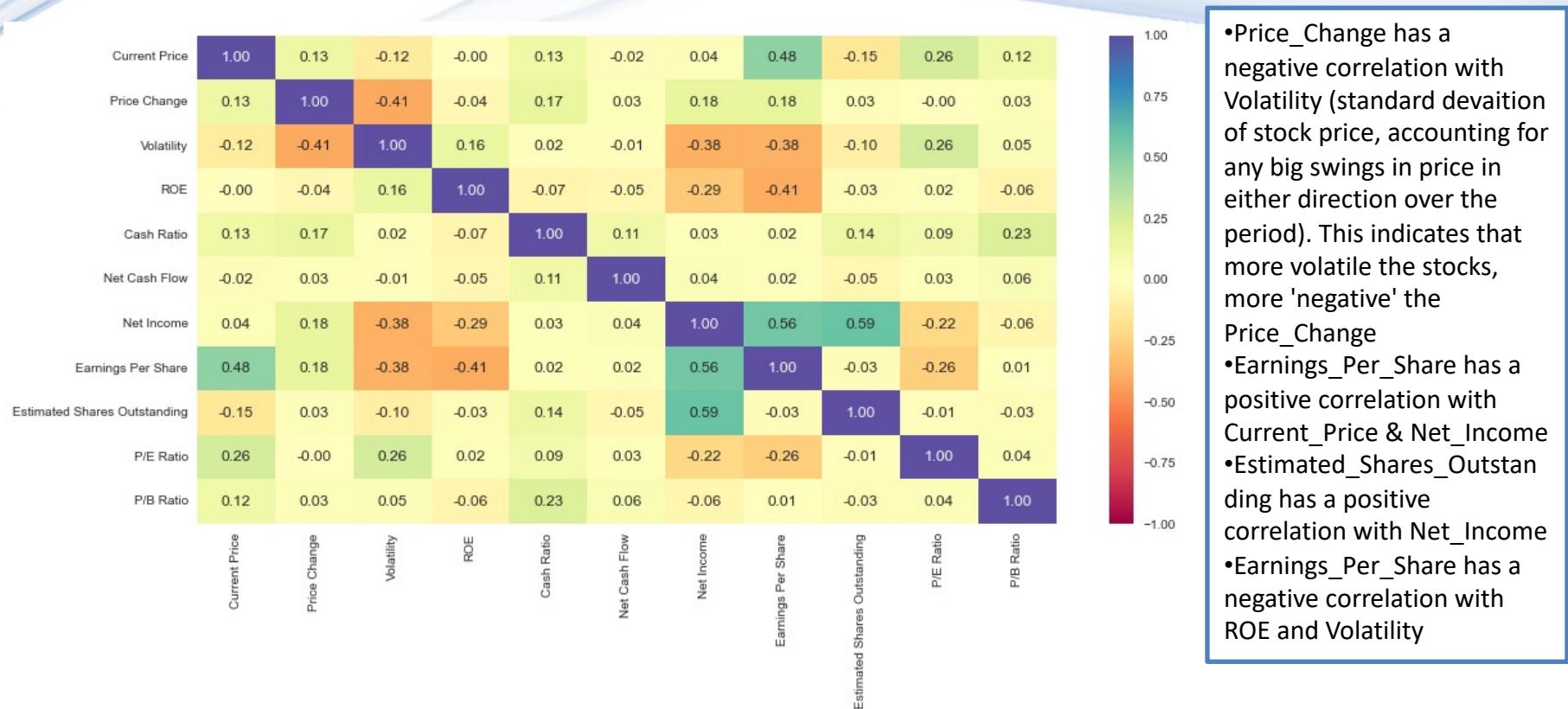


The GICS SubSector distribution



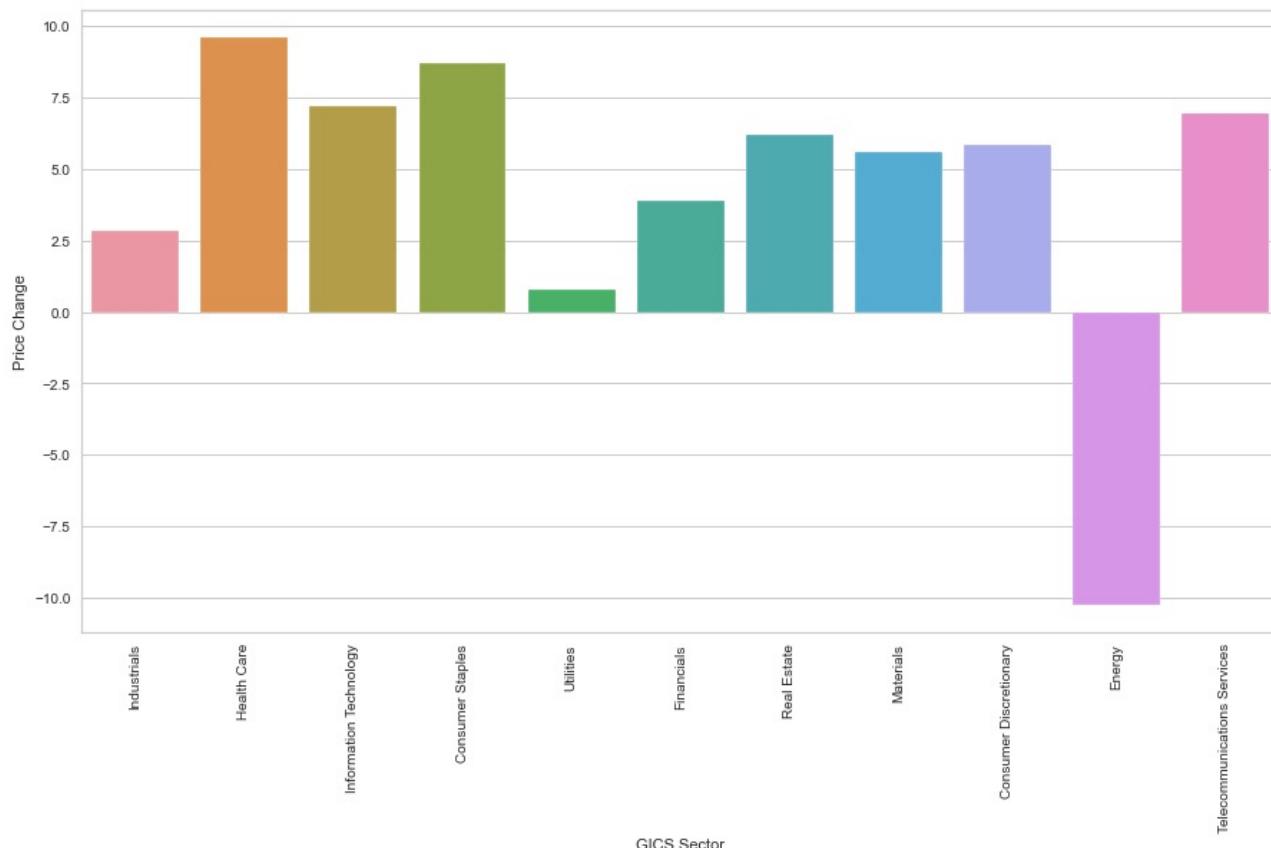
Exploratory Data Analysis (EDA)

Investors have many sector and sub-sector options to choose from when creating an optimal portfolio.



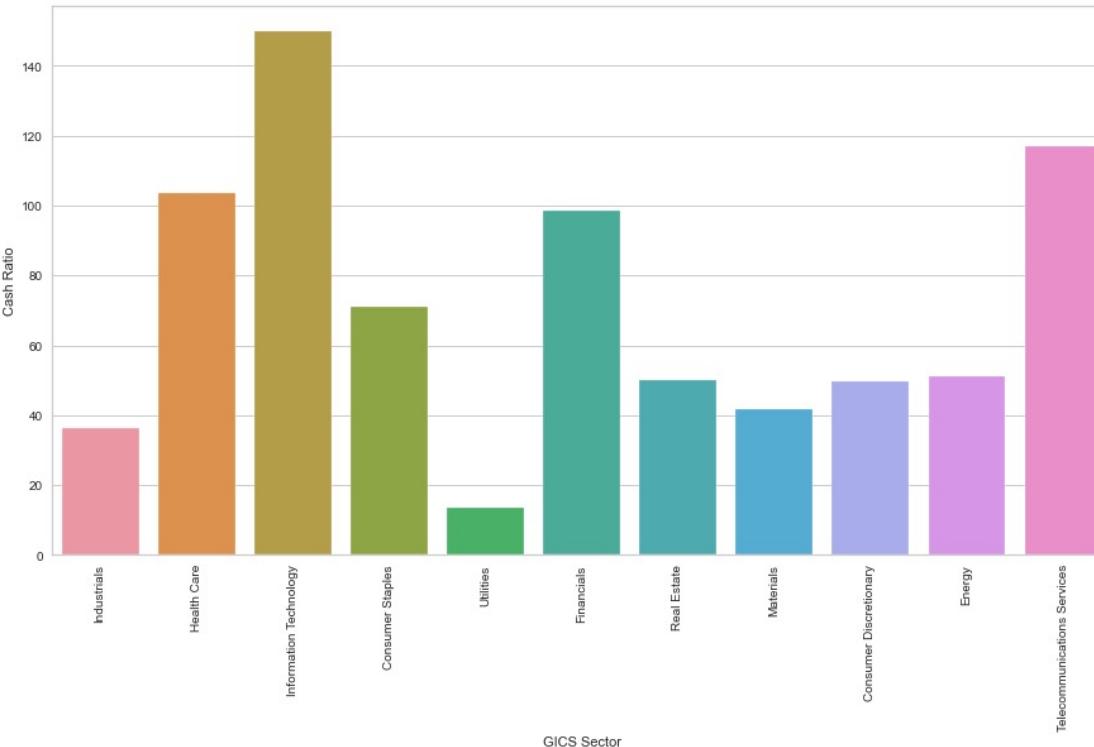
Exploratory Data Analysis (EDA)

Let's check the stocks of which economic sector have seen the maximum price increase on average.



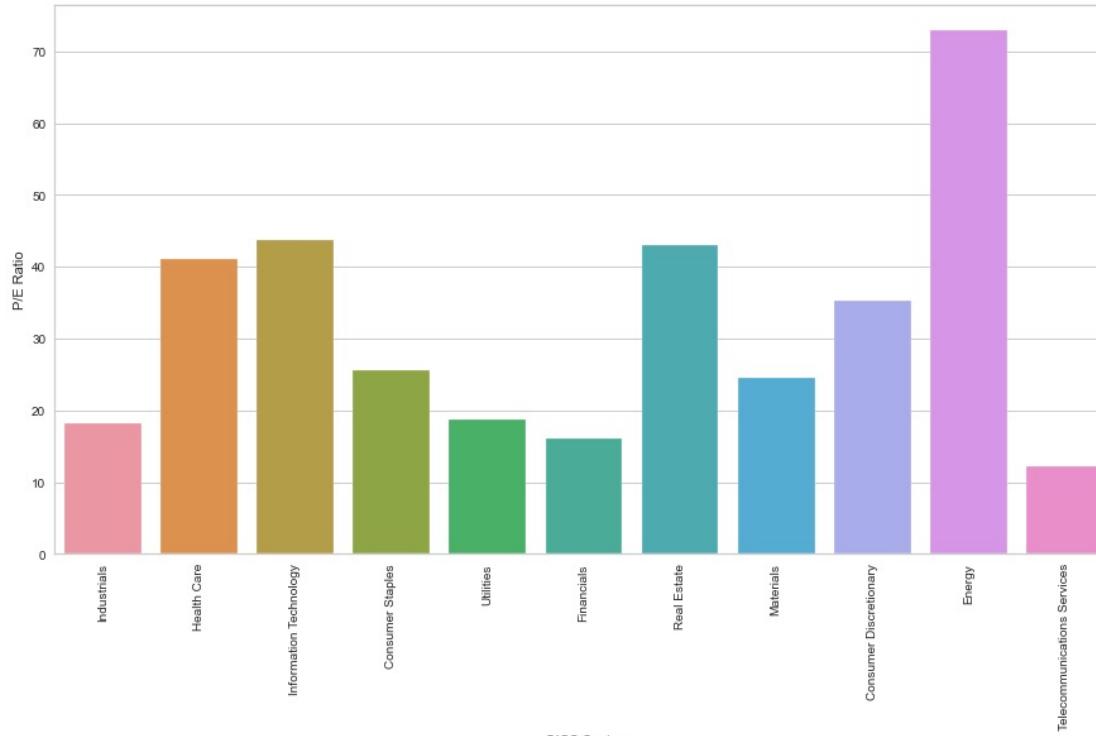
Exploratory Data Analysis (EDA)

Cash ratio provides a measure of a company's ability to cover its short-term obligations using only cash and cash equivalents. Let's see how the average cash ratio varies across economic sectors..



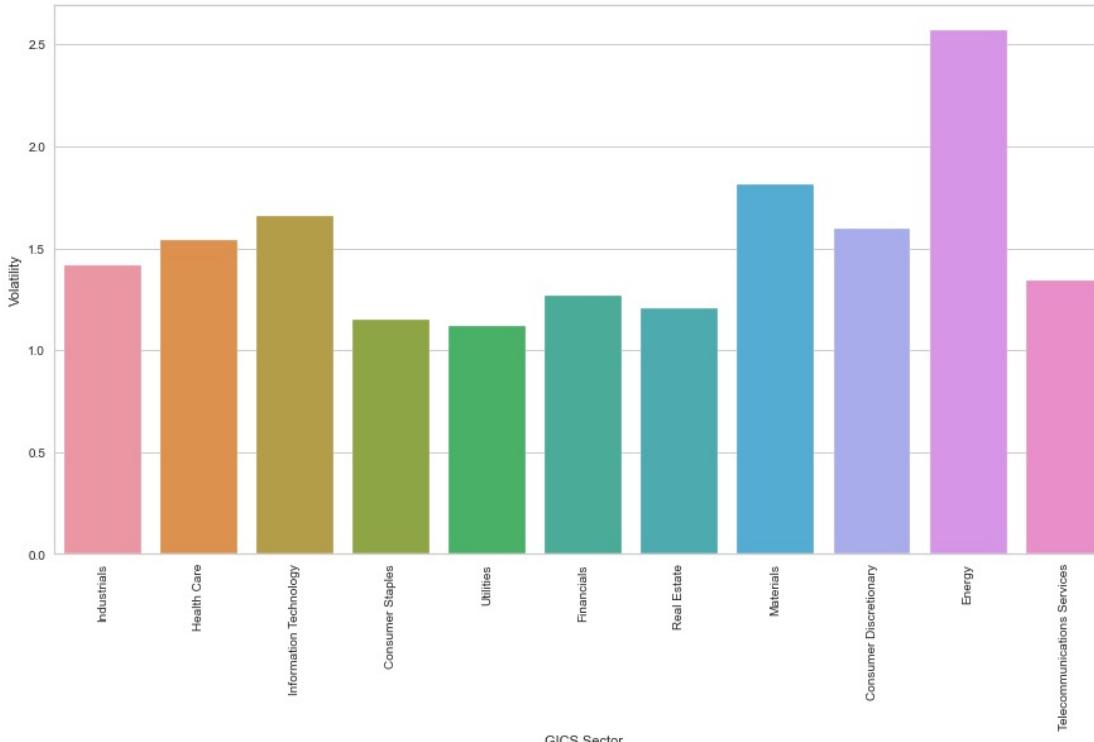
Exploratory Data Analysis (EDA)

P/E ratios can help determine the relative value of a company's shares as they signify the amount of money an investor is willing to invest in a single share of a company per dollar of its earnings. Let's see how the P/E ratio varies, on average, across economic sectors.



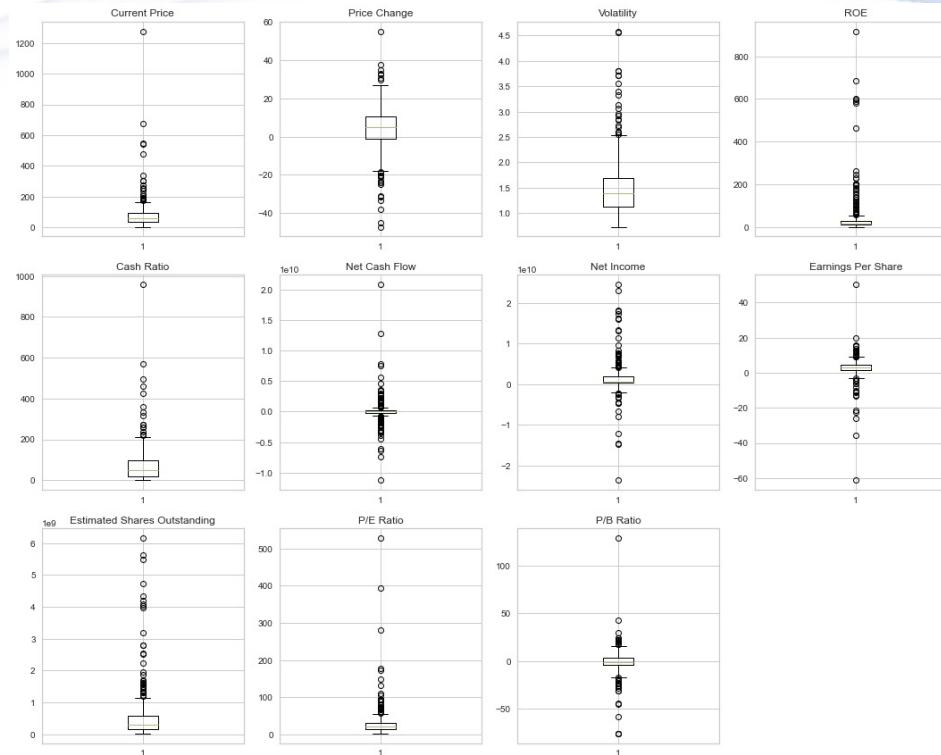
Exploratory Data Analysis (EDA)

Volatility accounts for the fluctuation in the stock price. A stock with high volatility will witness sharper price changes, making it a riskier investment. Let's see how volatility varies, on average, across economic sectors.



Exploratory Data Analysis (EDA)

There are outliers, but this is considered normal for this data, so we will leave the outliers untreated.



Data Preprocessing

Scaling the data set before clustering

```
class sklearn.preprocessing.StandardScaler(*, copy=True, with_mean=True,  
with_std=True)
```

- Standardize features by removing the mean and scaling to unit variance. The standard score of a sample x is calculated as:
$$z = (x - u) / s$$
- where u is the mean of the training samples or zero if `with_mean=False`, and s is the standard deviation of the training samples or one if `with_std=False`.
- Centering and scaling happen independently on each feature by computing the relevant statistics on the samples in the training set. Mean and standard deviation are then stored to be used on later data using [transform](#).
- Standardization of a dataset is a common requirement for many machine learning estimators: they might behave badly if the individual features do not more or less look like standard normally distributed data (e.g. Gaussian with 0 mean and unit variance).

- After feature engineering (scalar transformation), the relationship between the attributes have been maintained. However the attributes are now all on the same scale, with an average of 0, standard deviation of 1
- The data has no missing values, nor duplicate entries. Outliers have been identified, but not treated and they are assumed to be real data points & not anomalies in this context for modeling

Modeling Approaches

We will use both K-Means & Hierarchical Clustering

The [KMeans](#) algorithm clusters data by trying to separate samples in n groups of equal variance, minimizing a criterion known as the *inertia* or within-cluster sum-of-squares (see below). This algorithm requires the number of clusters to be specified. It scales well to large number of samples and has been used across a large range of application areas in many different fields.

- The k-means algorithm divides a set of N samples X into K disjoint clusters , each described by the mean of the samples in the cluster. The means are commonly called the cluster “centroids”; note that they are not, in general, points from X, although they live in the same space.
- The K-means algorithm aims to choose centroids that minimise the **inertia**, or **within-cluster sum-of-squares criterion**:

$$\sum_{i=0}^n \min_{\mu_j \in C} (\|x_i - \mu_j\|^2)$$

Hierarchical clustering is a general family of clustering algorithms that build nested clusters by merging or splitting them successively. This hierarchy of clusters is represented as a tree (or dendrogram). The root of the tree is the unique cluster that gathers all the samples, the leaves being the clusters with only one sample.

The [AgglomerativeClustering](#) object performs a hierarchical clustering using a bottom up approach: each observation starts in its own cluster, and clusters are successively merged together. The linkage criteria determines the metric used for the merge strategy:

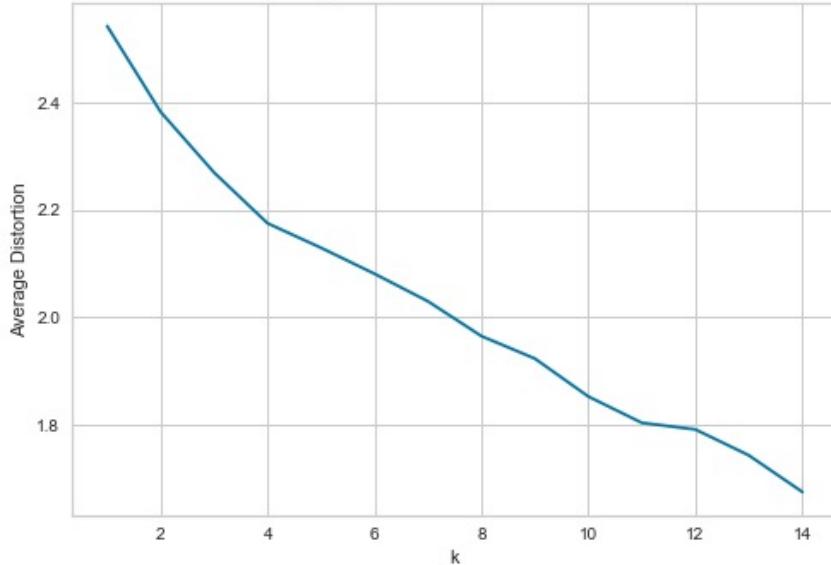
- **Ward** minimizes the sum of squared differences within all clusters. It is a variance-minimizing approach and in this sense is similar to the k-means objective function but tackled with an agglomerative hierarchical approach.
- **Maximum** or **complete linkage** minimizes the maximum distance between observations of pairs of clusters.
- **Average linkage** minimizes the average of the distances between all observations of pairs of clusters.
- **Single linkage** minimizes the distance between the closest observations of pairs of clusters.

[AgglomerativeClustering](#) can also scale to large number of samples when it is used jointly with a connectivity matrix, but is computationally expensive when no connectivity constraints are added between samples: it considers at each step all the possible merges.

K-means Clustering

The appropriate value for k seems to be between 4 and 6. but likely 5 is a good number.

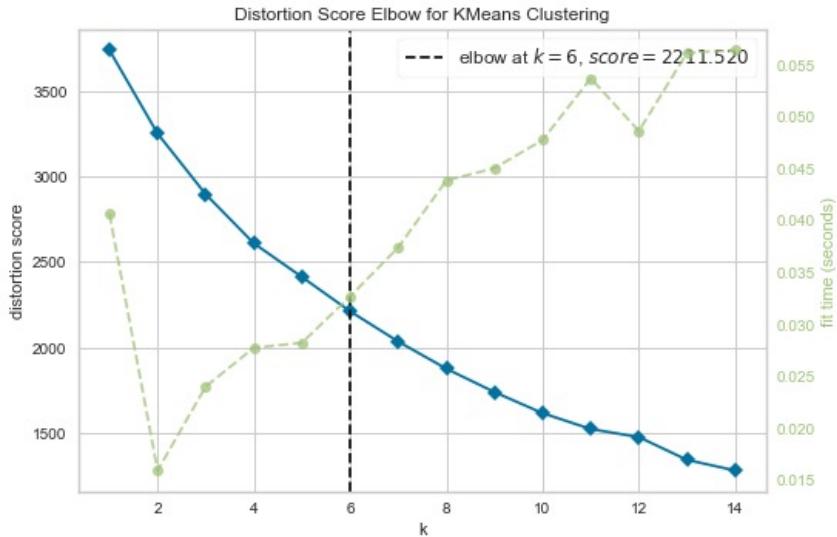
Selecting k with the Elbow Method



Number of Clusters: 1	Average Distortion: 2.5425069919221697
Number of Clusters: 2	Average Distortion: 2.382318498894466
Number of Clusters: 3	Average Distortion: 2.2692367155390745
Number of Clusters: 4	Average Distortion: 2.1745559827866363
Number of Clusters: 5	Average Distortion: 2.128799332840716
Number of Clusters: 6	Average Distortion: 2.080400099226289
Number of Clusters: 7	Average Distortion: 2.0289794220177395
Number of Clusters: 8	Average Distortion: 1.964144163389972
Number of Clusters: 9	Average Distortion: 1.9221492045198068
Number of Clusters: 10	Average Distortion: 1.8513913649973124
Number of Clusters: 11	Average Distortion: 1.8024134734578485
Number of Clusters: 12	Average Distortion: 1.7900931879652673
Number of Clusters: 13	Average Distortion: 1.7417609203336912
Number of Clusters: 14	Average Distortion: 1.673559857259703

K-means Clustering

The appropriate value for k seems to be between 4 and 6, but likely 5 is a good number.

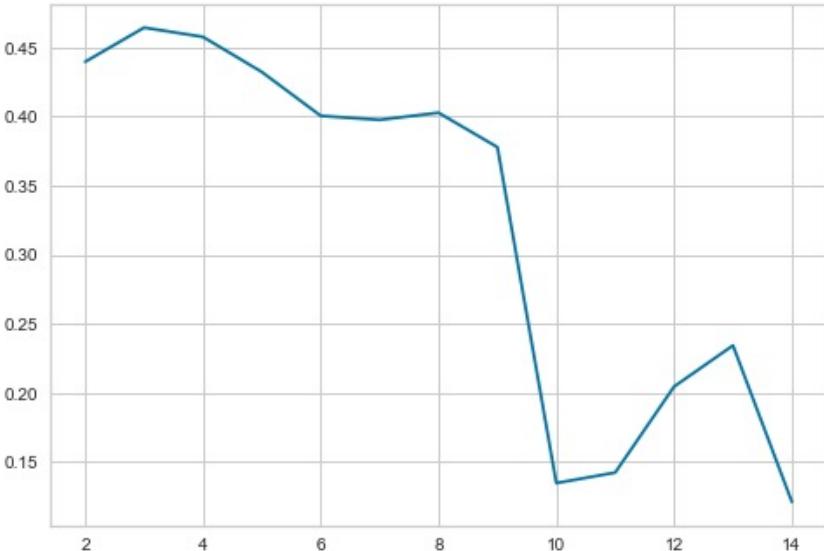


Good diversity across clusters

KM_segments	count_in_each_segment
0	277
1	11
2	27
3	25

K-means Clustering

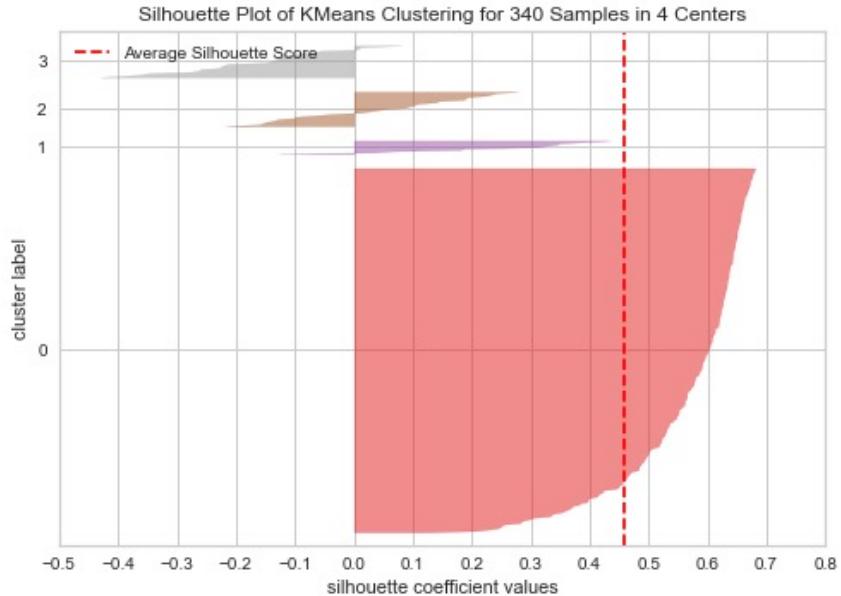
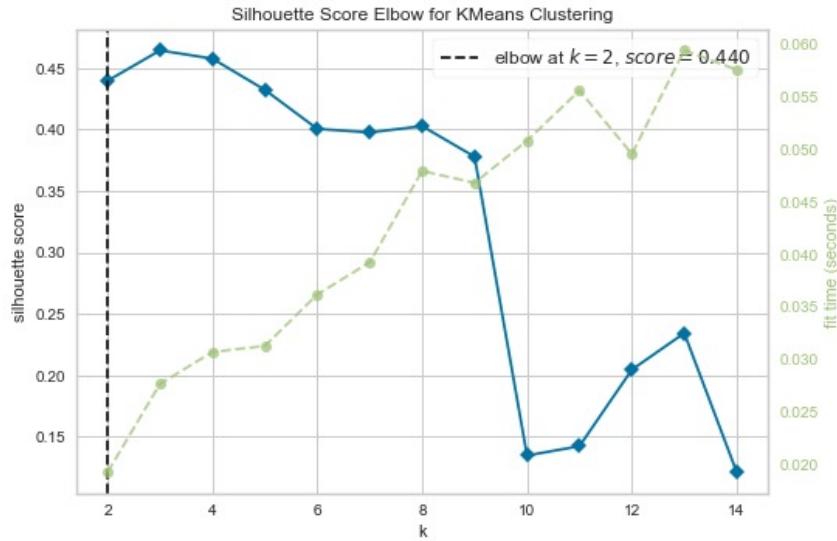
This would suggest that 4 based on the silhouette scores would be a good choice.



```
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 0.4577225970476733)
For n_clusters = 5, the silhouette score is 0.43228336443659804)
For n_clusters = 6, the silhouette score is 0.4005422737213617)
For n_clusters = 7, the silhouette score is 0.3976335364987305)
For n_clusters = 8, the silhouette score is 0.40278401969450467)
For n_clusters = 9, the silhouette score is 0.3778585981433699)
For n_clusters = 10, the silhouette score is 0.13458938329968687)
For n_clusters = 11, the silhouette score is 0.1421832155528444)
For n_clusters = 12, the silhouette score is 0.2044669621527429)
For n_clusters = 13, the silhouette score is 0.23424874810104204)
For n_clusters = 14, the silhouette score is 0.12102526472829901)
```

K-means Clustering

This would suggest that 4 based on the silhouette scores would be a good choice.



K-means - Cluster Profiles¹

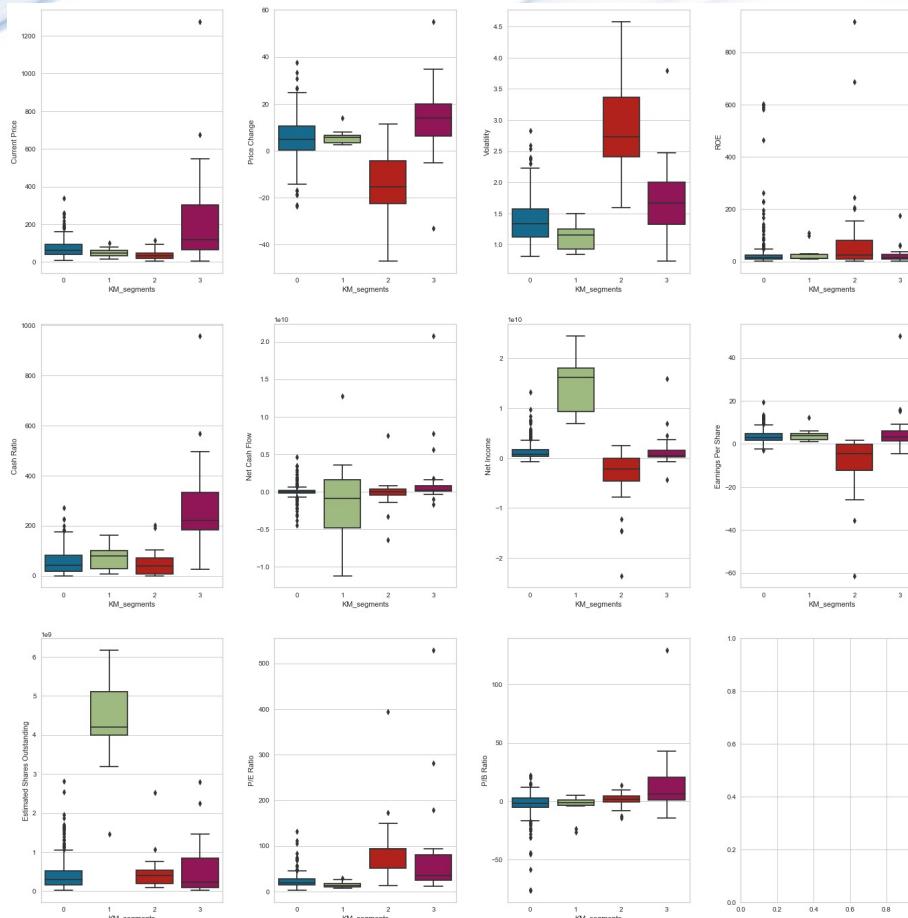
KM_segments	Current Price	Price Change	Volatility	ROE	Cash Ratio	Net Cash Flow	Net Income	Earnings Per Share	Estimated Shares Outstanding	P/E Ratio	P/B Ratio	count_in_each_segment
0.00	72	5	1	35	53	-14046224	1482212390	4	438533836	24	-3	277
1.00	51	6	1	31	76	-1072272727	14833090909	4	4298826629	15	-5	11
2.00	38	-15	3	107	50	-159428481	-3887457741	-9	480398573	91	1	27
3.00	234	13	2	26	278	1554926560	1572611680	6	578316319	75	14	25

KM_segments	GICS Sector	
0	Consumer Discretionary	33
	Consumer Staples	17
	Energy	6
	Financials	45
	Health Care	29
	Industrials	52
	Information Technology	24
	Materials	19
	Real Estate	26
	Telecommunications Services	2
	Utilities	24
1	Consumer Discretionary	1
	Consumer Staples	1
	Energy	1
	Financials	3
	Health Care	2
	Information Technology	1
	Telecommunications Services	2
	Energy	22
	Industrials	1
	Information Technology	3
	Materials	1
3	Consumer Discretionary	6
	Consumer Staples	1
	Energy	1
	Financials	1
	Health Care	9
	Information Technology	5
	Real Estate	1
	Telecommunications Services	1

Name: Security, dtype: int64

- Cluster 3
 - Has about 25 different securities
 - This cluster has securities with very high average Current_Price of \$234
 - As well, cluster has securities with very high Cash_Ratio of 278, high earnings per share of 6.05 and high P/B_Ratio of 14.4
 - Cluster 2
 - Has about 27 different securities
 - This cluster has securities with among the lowest average Current_Price of \$38 and negative average Price_Change of -15.3
 - As well, cluster has securities with among the highest volatility of 2.9, making them more riskier
 - This cluster also has the lowest Cash_Ratio of 50, low Net_Income (i.e., negative) of -3.887458e+09, low Earnings_Per_Share (i.e., negative) of -9.4
 - Cluster 0
 - Has the majority of securities at 277
 - This cluster has securities that lie in between Cluster 0 & Cluster 1 i.e, intermediate Current_Price of \$72, intermediate Price_Change of 5.066 (neither positive & high like cluster 0 nor negative and high like cluster 2), and intermediate volatility
 - Cluster 1
 - Has only 11 different securities
 - This cluster is very similar to Cluster 3 in its behavior but with two striking differences:
 - Has securities with some of the highest average Net_Income of 1.483309e+10 (almost 10 times of those in Cluster 3)
 - Has securities with some of the highest Estimated_Shares_Outstanding of 4.298827e+09 (almost 10 times of those in Cluster 3)
- Clusters 3 and 2 are the safe clusters, with clusters 2 containing more exclusive securities. Clusters 0 and 1 are more riskier securities, former being high performing and later historically speaking low performing

K-means - Cluster Profiles¹



•Cluster 3

- Has about 25 different securities
- This cluster has securities with very high average Current_Price of \$234
- As well, cluster has securities with very high Cash_Ratio of 278, high earnings per share of 6.05 and high P/B_Ratio of 14.4

•Cluster 2

- Has about 27 different securities
- This cluster has securities with among the lowest average Current_Price of \$38 and negative average Price_Change of -15.3
- As well, cluster has securities with among the highest volatility of 2.9, making them more riskier
- This cluster also has the lowest Cash_Ratio of 50, low Net_Income (i.e., negative) of -3.887458e+09, low Earnings_Per_Share (i.e., negative) of -9.4

•Cluster 0

- Has the majority of securities at 277
- This cluster has securities that lie in between Cluster 0 & Cluster 1 i.e., intermediate Current_Price of \$72, intermediate Price_Change of 5.066 (neither positive & high like cluster 0 nor negative and high like cluster 2), and intermediate volatility

•Cluster 1

- Has only 11 different securities
- This cluster is very similar to Cluster 3 in its behavior but with two striking differences:
 - Has securities with some of the highest average Net_Income of 1.483309e+10 (almost 10 times of those in Cluster 3)
 - Has securities with some of the highest Estimated_Shares_Outstanding of 4.298827e+09 (almost 10 times of those in Cluster 3)

Clusters 3 and 2 are the safe clusters, with clusters 2 containing more exclusive securities. Clusters 0 and 1 are more riskier securities, former being high performing and later historically speaking low performing

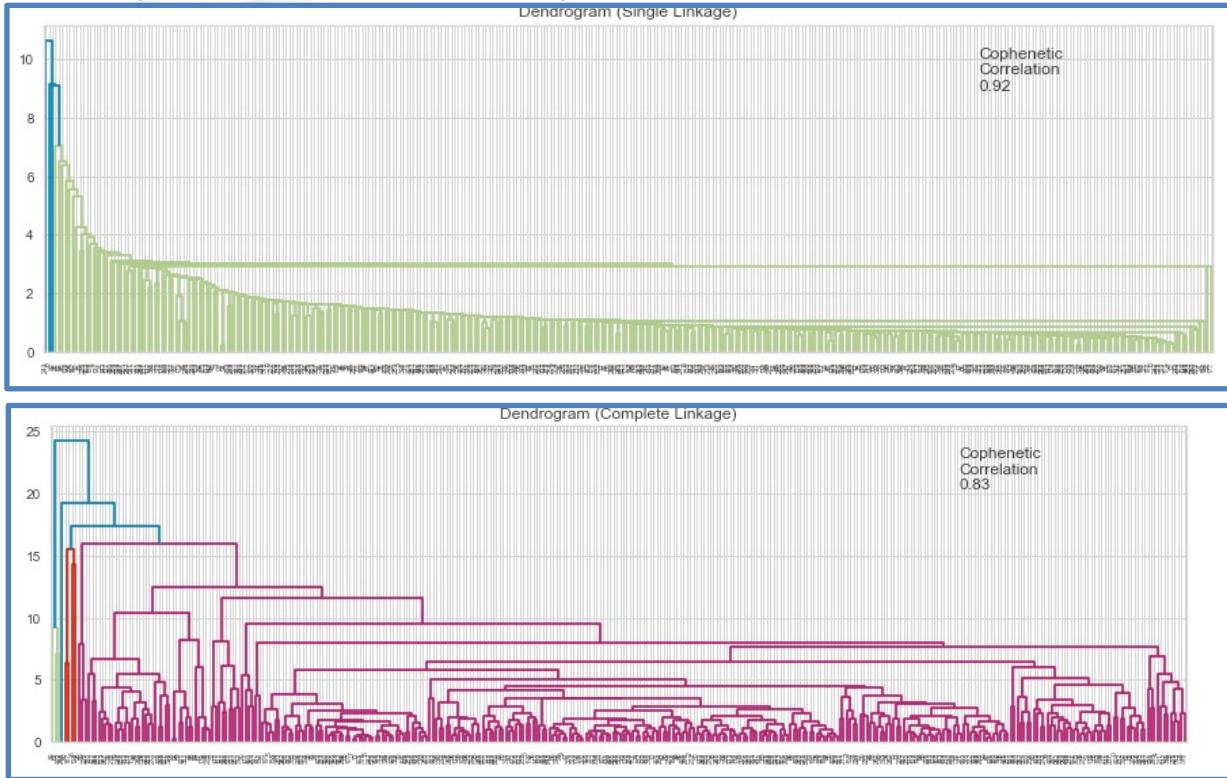
Hierarchical Clustering

```
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.  
*****  
Highest cophenetic correlation is 0.9422540609560814, which is obtained with Euclidean distance and average linkage.
```

```
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.  
*****  
Highest cophenetic correlation is 0.9422540609560814, which is obtained with average linkage.
```

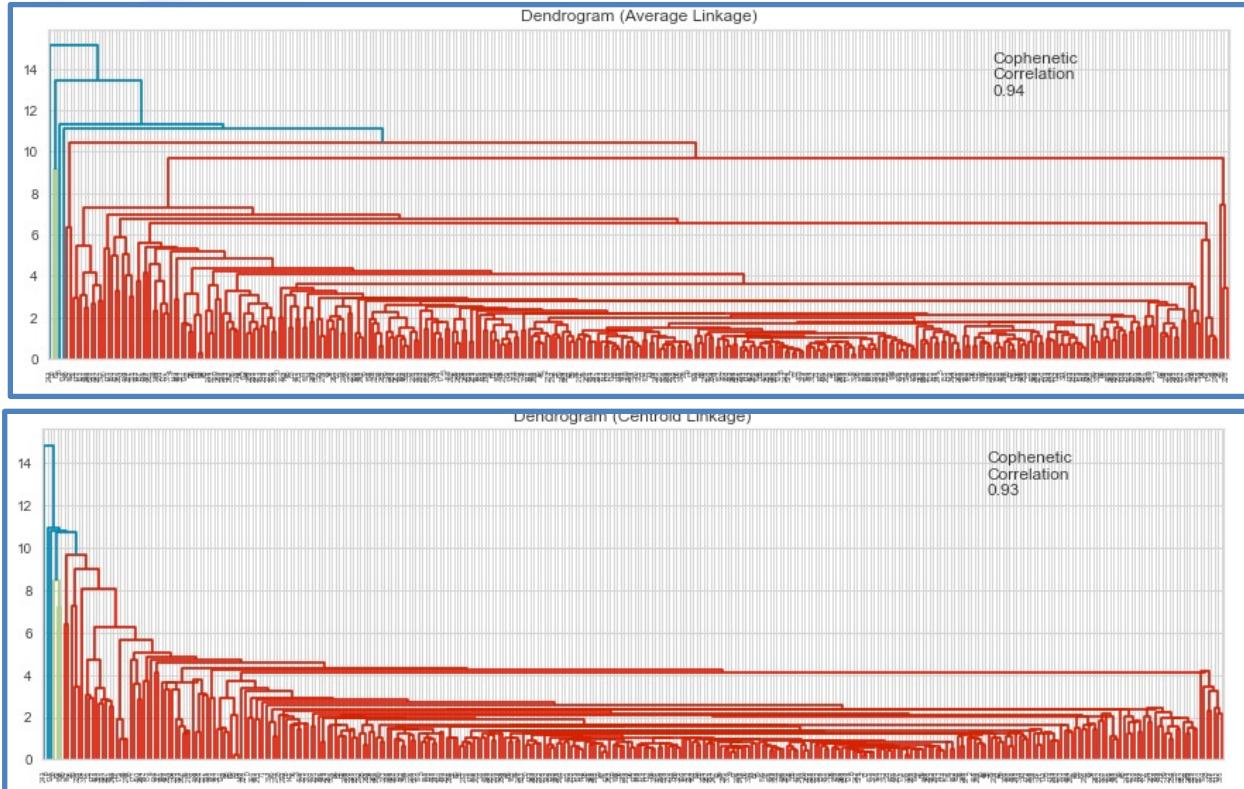
Hierarchical Clustering – Denodgram

Let's view the dendograms for the different linkage methods with Euclidean distance.



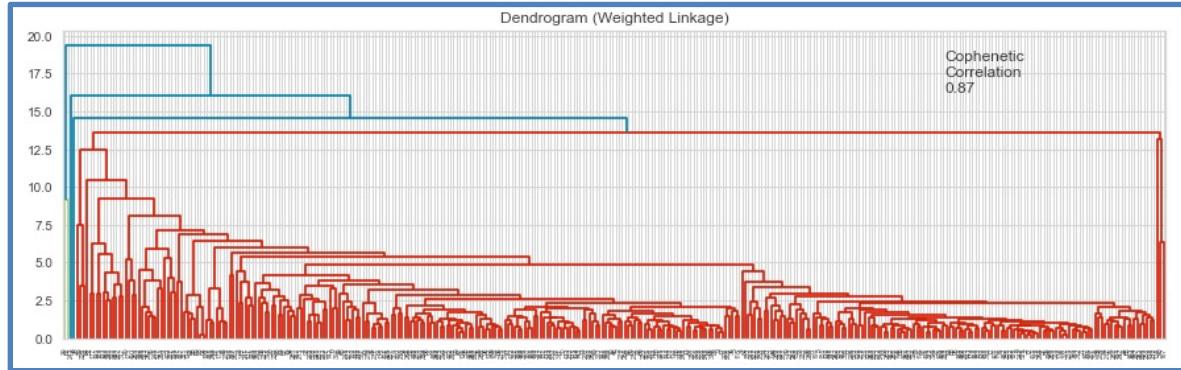
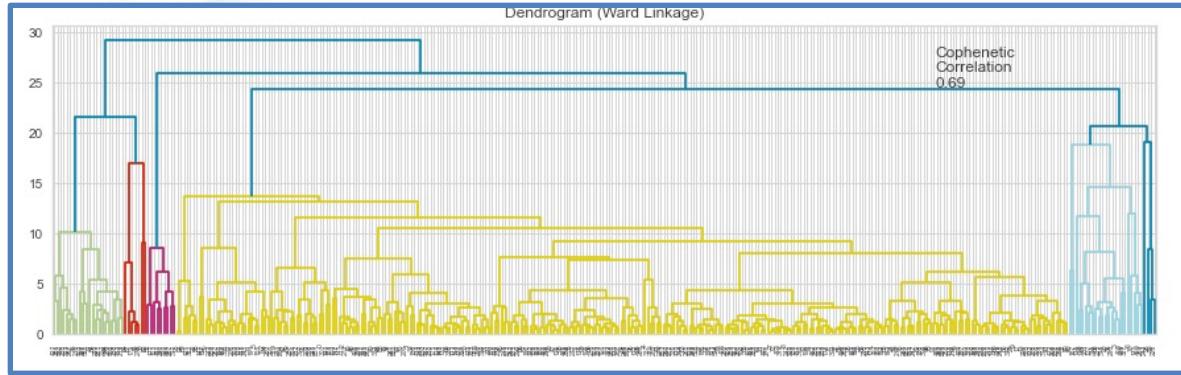
Hierarchical Clustering – Denodgram

Let's view the dendograms for the different linkage methods with Euclidean distance.



Hierarchical Clustering – Denodgram

Let's view the dendograms for the different linkage methods with Euclidean distance.



Hierarchical Clustering – Denodgram

Let's view the dendograms for the different linkage methods with Euclidean distance.

AgglomerativeClustering(linkage='average', n_clusters=6)

Linkage	Cophenetic Coefficient
4 ward	0.710118
1 complete	0.787328
5 weighted	0.869378
0 single	0.923227
3 centroid	0.931401
2 average	0.942254

Hierarchical Clustering – Cluster Profiles²

	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
HC_segments												
0	77.287589	4.09973	1.518066	35.336336	66.900901	-33197321.32	1538074667	2.88527	560505037.3	32.441706	-2.174921	333
1	25.64	11.237908	1.322355	12.5	130.5	16755500000	13654000000	3.295	2791829362	13.649696	1.508484	2
2	24.485001	-13.351992	3.482611	802	51	-1292500000	-19106500000	-41.815	519573983.3	60.748608	1.565141	2
3	104.660004	16.22432	1.320606	8	958	592000000	3669000000	1.31	2800763359	79.893133	5.884467	1
4	1274.949951	3.190527	1.26834	29	184	-1671386000	2551360000	50.09	50935516.07	25.453183	-1.052429	1
5	276.570007	6.189286	1.116976	30	25	90885000	596541000	8.91	66951851.85	31.040405	129.06459	1

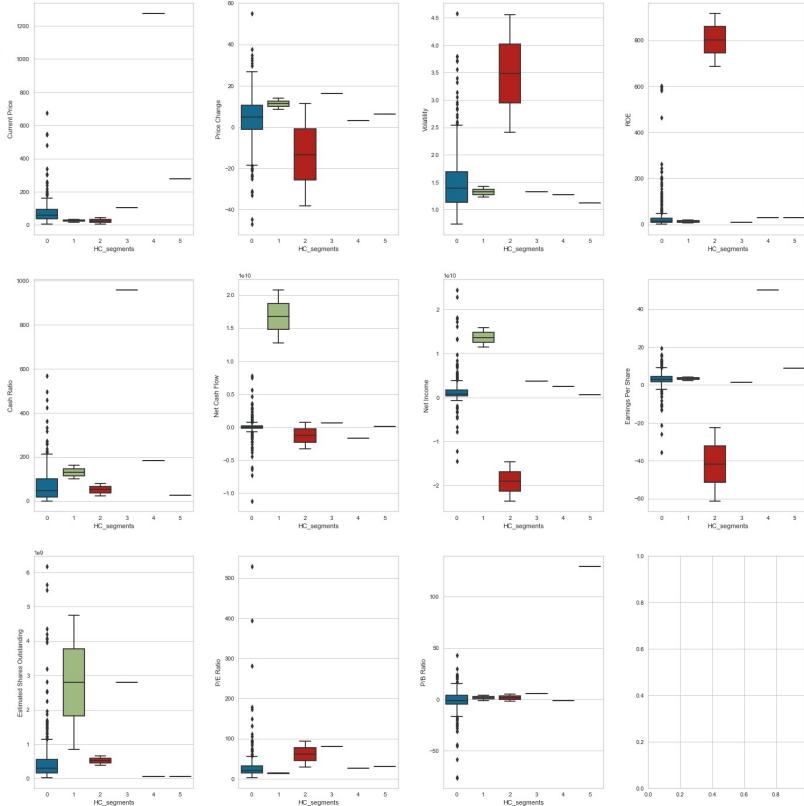
HC_segments	GICS Sector	
0	Consumer Discretionary	39
	Consumer Staples	19
	Energy	28
	Financials	48
	Health Care	40
	Industrials	53
	Information Technology	30
	Materials	20
	Real Estate	27
	Telecommunications Services	5
	Utilities	24
1	Financials	1
	Information Technology	1
2	Energy	2
3	Information Technology	1
4	Consumer Discretionary	1
5	Information Technology	1
Name: Security, dtype: int64		

< No diversity in the clusters 1 – 5.

	count_in_each_segment
HC_segments	
0	333
1	2
2	2
3	1
4	1
5	1

Let's try this again with "Ward" to diversify the clusters!

Hierarchical Clustering – Cluster Profiles²



< No diversity in the clusters 1 – 5.

HC_segments	count_in_each_segment	
	0	333
1	2	2
2	2	2
3	1	1
4	1	1
5	1	1

Let's try this again with "Ward" to diversify the clusters!

Hierarchical Clustering – Cluster Profiles II Now

Using “Ward” to diversify the clusters

	Current Price	Price Change	Volatility	ROE	Cash Ratio	Net Cash Flow	Net Income	Earnings Per	Estimated Shares	P/E Ratio	P/B Ratio	count_in_each_segment
HC_segments												
0.00	74.16	-8.15	2.54	178.50	46.44	-383950875.00	-3233045437.50	-7.80	475111396.80	105.66	-2.46	32.00
1.00	72.43	5.26	1.43	25.51	60.88	196157425.09	1623022236.93	3.65	462816085.05	24.65	-2.62	287.00
2.00	46.67	5.17	1.08	25.00	58.33	-3040666666.67	14848444444.44	3.44	4564959946.22	15.60	-6.35	9.00
3.00	326.00	7.72	1.55	17.00	360.33	186503166.67	1008736416.67	9.54	589434517.80	41.12	25.26	12.00

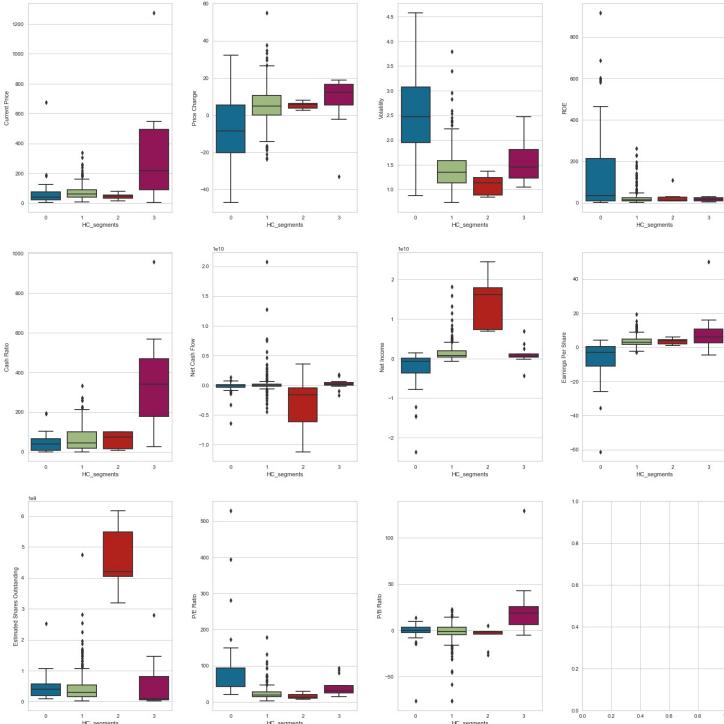
HC_segments	GICS Sector	
0	Consumer Discretionary	2
	Consumer Staples	2
	Energy	22
	Financials	1
	Health Care	1
	Industrials	1
	Information Technology	2
	Materials	1
1	Consumer Discretionary	35
	Consumer Staples	15
	Energy	7
	Financials	45
	Health Care	34
	Industrials	52
	Information Technology	28
	Materials	19
	Real Estate	26
	Telecommunications Services	2
	Utilities	24
2	Consumer Discretionary	1
	Consumer Staples	1
	Energy	1
	Financials	3
	Health Care	1
	Telecommunications Services	2
	Consumer Discretionary	2
	Consumer Staples	1
	Health Care	4
	Information Technology	3
	Real Estate	1
	Telecommunications Services	1
3	Name: Security, dtype: int64	

Good diversity across clusters

	count_in_each_segment
HC_segments	
0	11
1	12
2	22
3	285
4	7
5	3

Hierarchical Clustering – Cluster Profiles II ³ Now

Using “Ward” to diversify the clusters



Good diversity across clusters

HC_segments	count_in_each_segment	
	0	11
1	12	
2	22	
3	285	
4	7	
5	3	

Data Insights

Basic Data Observations⁴

- ❑ No security/ company is repeated twice
- ❑ There are 11 unique GICS_Sector, and Industrials is the most common of the same
- ❑ There are 104 unique GICS_Sub_Industry, and Oil & Gas Exploration & Production is the most common of the same
- ❑ The average Current_Price is \$80.86 while the median Current_Price is \$59.75. This indicates data is right skewed with some high outliers
- ❑ The average Price_Change (% change in price in the last 13 weeks) & average Volatility (Std deviation of price over last 13 weeks) are \$4.078 and 1.5 while the median values are \$4.819 and 1.38
- ❑ Average ROE (return on equity) is 39.59 while the median ROE is only 15. This indicates data is right skewed with some outliers. A higher ROE indicates that a company is able to generate higher profits without needing much capital and is preferred by investors
- ❑ Cash ratio is a measure of the company to pay off short term liabilities with cash & cash equivalents. While a high cash ratio is preferred by investors, a very high cash ratio, may also indicate that cash is not being re invested or is not returning further interest. The average is 70.2 and median is 47
- ❑ The average Net_Cash_Flow is \$55537620 while median Net_Cash_Flow is \$2098000.0 (right skewed, high outliers)
- ❑ The average Net_Income is \$1494384602.941176 while median Net_Income is \$707336000.0 (again right skewed, high outliers). The Net Cash flow is the amount of money that actually comes in and goes out of a business during a period of time. Net income is the profit or loss that a business has after subtracting all expenses from the total revenue. Higher value of both is preferred
- ❑ The average Earnings_Per_Share (net profit/shares outstanding) is 2.78 while median is 2.895 (left skewed)
- ❑ The average Estimated_Shares_Outstanding is 577028337.754029 while median is 309675137.8 (right skewed with outliers)
- ❑ The average P/E , P/B ratio are 32.61 and -1.7 while median values are 20.8 and -1.06 respectively. P/E and P/B are the ratio of company's stock price per share by earnings per share & by (total assets - total liabilities) per share. The average P/B ratio being negative indicates there are some companies in serious financial stress with liabilities exceeding assets (left skewed)

Business Insights and Recommendations

Clustering Profiles

- ❑ Out of 340 securities in our data set, both clustering methods clustered 277, 285* and 333 securities in a similar fashion. The industry segregation into clusters yielded similar results across both algorithms. The Hierarchical Clustering with “Ward” distance yielded a similar result to K-Means.
- ❑ One cluster (25+securities) was identified as very aggressive (& high performing) belonging predominantly to Health Care followed by Consumer Discretionary and Information Technology sectors
- ❑ Another cluster (25+securities) was identified as very aggressive (but historically low performing) belonging predominantly to Energy sector
- ❑ Another cluster (~10securities) was identified as moderately aggressive (& high performing) belonging predominantly to Financials sector
- ❑ Finally, a major cluster (270+securities) was identified as mildly aggressive & safe investment option. This cluster is diversified with securities predominantly belonging to Industrials, followed by Financials, Consumer Discretionary, Real Estate, & Informational Technology sectors

Recommendations

- ❑ Securities were divided into four different groups based on their performance: very aggressive (high and low performing), moderately aggressive (high and low performing), and mildly aggressive (low and high performing).
- ❑ While it is vital to divide equities among a variety of assets that are diversified, it is also important to optimize returns in any market environment.
- ❑ It is crucial to remember, however, that the stock market is notoriously unpredictable, and that historical indicators may not always be indicative of future developments.
- ❑ It is necessary to conduct further research into dynamic clustering (as fresh data is uploaded each day) and movement of stocks across cluster groups as a result of changing market conditions in order to make more accurate forecasts.

APPENDIX 1 - K-Means Portfolio 1

- In cluster 0, the following companies are present: ['American Airlines Group' 'AbbVie' 'Abbott Laboratories' 'Adobe Systems Inc' 'Archer-Daniels-Midland Co' 'Ameren Corp' 'American Electric Power' 'AFLAC Inc' 'American International Group, Inc.' 'Apartment Investment & Mgmt' 'Assurant Inc' 'Arthur J. Gallagher & Co.' 'Akamai Technologies Inc' 'Albemarle Corp' 'Alaska Air Group Inc' 'Allstate Corp' 'Allegion' 'Applied Materials Inc' 'AMETEK Inc' 'Affiliated Managers Group Inc' 'Ameriprise Financial' 'American Tower Corp A' 'AutoNation Inc' 'Anthem Inc.' 'Aon plc' 'Amphenol Corp' 'Arconic Inc' 'Activision Blizzard' 'AvalonBay Communities, Inc.' 'Broadcom' 'American Water Works Company Inc' 'American Express Co' 'Boeing Company' 'Baxter International Inc.' 'BB&T Corporation' 'Bard (C.R.) Inc.' 'The Bank of New York Mellon Corp.' 'Ball Corp' 'Bristol-Myers Squibb' 'Boston Scientific' 'BorgWarner' 'Boston Properties' 'Caterpillar Inc.' 'Chubb Limited' 'CBRE Group' 'Crown Castle International Corp.' 'Carnival Corp.' 'CF Industries Holdings Inc' 'Citizens Financial Group' 'Church & Dwight' 'C. H. Robinson Worldwide' 'Charter Communications' 'CIGNA Corp.' 'Cincinnati Financial' 'Colgate-Palmolive' 'Comerica Inc.' 'CME Group Inc.' 'Cummins Inc.' 'CMS Energy' 'Centene Corporation' 'CenterPoint Energy' 'Capital One Financial' 'The Cooper Companies' 'CSX Corp.' 'CenturyLink Inc' 'Cognizant Technology Solutions' 'Citrix Systems' 'CVS Health' 'Chevron Corp.' 'Dominion Resources' 'Delta Air Lines' 'DuPont (E.I.)' 'Deere & Co.' 'Discover Financial Services' 'Quest Diagnostics' 'Danaher Corp.' 'The Walt Disney Company' 'Discovery Communications-A' 'Discovery Communications-C' 'Delphi Automotive' 'Digital Realty Trust' 'Dun & Bradstreet' 'Dover Corp.' 'Dr Pepper Snapple Group' 'Duke Energy' 'DaVita Inc.' 'eBay Inc.' 'Ecolab Inc.' 'Consolidated Edison' 'Equifax Inc.' 'Edison Int'l' 'Eastman Chemical' 'Equity Residential' 'Eversource Energy' 'Essex Property Trust, Inc.' 'E*Trade' 'Eaton Corporation' 'Entergy Corp.' 'Exelon Corp.' 'Expeditors Int'l' 'Expedia Inc.' 'Extra Space Storage' 'Fastenal Co' 'Fortune Brands Home & Security' 'FirstEnergy Corp' 'Fidelity National Information Services' 'Fiserv Inc' 'FLIR Systems' 'Fluor Corp.' 'Flowserve Corporation' 'FMC Corporation' 'Federal Realty Investment Trust' 'General Dynamics' 'General Growth Properties Inc.' 'Corning Inc.' 'General Motors' 'Genuine Parts' 'Garmin Ltd.' 'Goodyear Tire & Rubber' 'Grainger (W.W.) Inc.' 'Hasbro Inc.' 'Huntington Bancshares' 'HCA Holdings' 'Welltower Inc.' 'HCP Inc.' 'Hartford Financial Svc.Gp.' 'Harley-Davidson' 'Honeywell Int'l Inc.' 'HP Inc.' 'Hormel Foods Corp.' 'Henry Schein' 'Host Hotels & Resorts' 'The Hershey Company' 'Humana Inc.' 'International Business Machines' 'IDEXX Laboratories' 'Intl Flavors & Fragrances' 'International Paper' 'Interpublic Group' 'Iron Mountain Incorporated' 'Illinois Tool Works' 'Invesco Ltd.' 'J. B. Hunt Transport Services' 'Jacobs Engineering Group' 'Juniper Networks' 'Kimco Realty' 'Kimberly-Clark' 'Kansas City Southern' 'Leggett & Platt' 'Lennar Corp.' 'Laboratory Corp. of America Holding' 'LKQ Corporation' 'L-3 Communications Holdings' 'Lilly (Eli) & Co.' 'Lockheed Martin Corp.' 'Alliant Energy Corp' 'Leucadia National Corp.' 'Southwest Airlines' 'Level 3 Communications' 'LyondellBasell' 'Mastercard Inc.' 'Mid-America Apartments' 'Macerich' 'Marriott Int'l.' 'Masco Corp.' 'Mattel Inc.' 'Moody's Corp' 'Mondelez International' 'MetLife Inc.' 'Mohawk Industries' 'Mead Johnson' 'McCormick & Co.' 'Martin Marietta Materials' 'Marsh & McLennan' '3M Company' 'Altria Group Inc' 'The Mosaic Company' 'Marathon Petroleum' 'Merck & Co.' 'M&T Bank Corp.' 'Mettler Toledo' 'Mylan N.V.' 'Navient' 'NASDAQ OMX Group' 'NextEra Energy' 'Newmont Mining Corp. (Hdg. Co.)' 'Nielsen Holdings' 'Norfolk Southern Corp.' 'Northern Trust Corp.' 'Nucor Corp.' 'Newell Brands' 'Realty Income Corporation' 'Omnicon Group' 'O'Reilly Automotive' 'People's United Financial' 'Pitney-Bowes' 'PACCAR Inc.' 'PG&E Corp.' 'Public Serv. Enterprise Inc.' 'PepsiCo Inc.' 'Principal Financial Group' 'Procter & Gamble' 'Progressive Corp.' 'Pulte Homes Inc.' 'Philip Morris International' 'PNC Financial Services' 'Pentair Ltd.' 'Pinnacle West Capital' 'PPG Industries' 'PPL Corp.' 'Prudential Financial' 'Phillips 66' 'Praxair Inc.' 'PayPal' 'Ryder System' 'Royal Caribbean Cruises Ltd' 'Robert Half International' 'Roper Industries' 'Republic Services Inc' 'SCANA Corp.' 'Charles Schwab Corporation' 'Spectra Energy Corp.' 'Sealed Air' 'Sherwin-Williams' 'SL Green Realty' 'Scripps Networks Interactive Inc.' 'Southern Co.' 'Simon Property Group Inc' 'S&P Global, Inc.' 'Stericycle Inc' 'Sempra Energy' 'SunTrust Banks' 'State Street Corp.' 'Skyworks Solutions' 'Synchrony Financial' 'Stryker Corp.' 'Molson Coors Brewing Company' 'Tegna, Inc.' 'Torchmark Corp.' 'Thermo Fisher Scientific' 'The Travelers Companies Inc.' 'Tractor Supply Company' 'Tyson Foods' 'Tesoro Petroleum Co.' 'Total System Services' 'Texas Instruments' 'Under Armour' 'United Continental Holdings' 'UDR Inc' 'Universal Health Services, Inc.' 'United Health Group Inc.' 'Unum Group' 'Union Pacific' 'United Parcel Service' 'United Technologies' 'Varian Medical Systems' 'Valero Energy' 'Vulcan Materials' 'Vornado Realty Trust' 'Verisk Analytics' 'Verisign Inc.' 'Ventas Inc' 'We Energies Group Inc' 'Whirlpool Corp.' 'Waste Management Inc.' 'Western Union Co' 'Weyerhaeuser Corp.' 'Wyndham Worldwide' 'Xcel Energy Inc' 'XL Capital' 'Dentsply Sirona' 'Xerox Corp.' 'Xylem Inc.' 'Yum! Brands Inc' 'Zimmer Biomet Holdings' 'Zions Bancorp' 'Zoetis']
- In cluster 3, the following companies are present: ['Analog Devices, Inc.' 'Alliance Data Systems' 'Alexion Pharmaceuticals' 'Amgen Inc' 'Amazon.com Inc' 'Bank of America Corp' 'BIOGEN IDEC Inc.' 'Celgene Corp.' 'Chipotle Mexican Grill' 'Equinix' 'Edwards Lifesciences' 'Facebook' 'First Solar Inc' 'Frontier Communications' 'Halliburton Co.' 'Intuitive Surgical Inc.' 'McDonald's Corp.' 'Monster Beverage' 'Priceline.com Inc' 'Regeneron' 'TripAdvisor' 'Vertex Pharmaceuticals Inc' 'Waters Corporation' 'Wynn Resorts Ltd' 'Yahoo Inc.]
- In cluster 2, the following companies are present: ['Apache Corporation' 'Anadarko Petroleum Corp' 'Baker Hughes Inc' 'Chesapeake Energy' 'Cabot Oil & Gas' 'Concho Resources' 'Devon Energy Corp.' 'EOG Resources' 'EQT Corporation' 'Freeport-McMoran Cp & Gld' 'Hess Corporation' 'Hewlett Packard Enterprise' 'Kinder Morgan' 'Marathon Oil Corp.' 'Murphy Oil' 'Noble Energy Inc' 'Netflix Inc.' 'Newfield Exploration Co' 'National Oilwell Varco Inc.' 'ONEOK' 'Occidental Petroleum' 'Quanta Services Inc.' 'Range Resources Corp.' 'Southwestern Energy' 'Teradata Corp.' 'Williams Cos.' 'Cimarex Energy']
- In cluster 1, the following companies are present: ['Citigroup Inc.' 'Ford Motor' 'Gilead Sciences' 'Intel Corp.' 'JP Morgan Chase & Co.' 'Coca Cola Company' 'Pfizer Inc.' 'AT&T Inc' 'Verizon Communications' 'Wells Fargo' 'Exxon Mobil Corp.]

APPENDIX 2 - Hierarchical Clustering Portfolio 1

- In cluster 0, the following companies are present: ['American Airlines Group' 'AbbVie' 'Abbott Laboratories' 'Adobe Systems Inc' 'Analog Devices, Inc.' 'Archer-Daniels-Midland Co' 'Ameren Corp' 'American Electric Power' 'AFLAC Inc' 'American International Group, Inc.' 'Apartment Investment & Mgmt' 'Assurant Inc' 'Arthur J. Gallagher & Co.' 'Akamai Technologies Inc' 'Albemarle Corp' 'Alaska Air Group Inc' 'Allstate Corp' 'Allegion' 'Alexion Pharmaceuticals' 'Applied Materials Inc' 'AMETEK Inc' 'Affiliated Managers Group Inc' 'Amgen Inc' 'Ameriprise Financial' 'American Tower Corp A' 'Amazon.com Inc' 'AutoNation Inc' 'Anthem Inc.' 'Aon plc' 'Anadarko Petroleum Corp' 'Amphenol Corp' 'Arconic Inc' 'Activision Blizzard' 'AvalonBay Communities, Inc.' 'Broadcom' 'American Water Works Company Inc' 'American Express Co' 'Boeing Company' 'Baxter International Inc.' 'BB&T Corporation' 'Bard (C.R.) Inc.' 'Baker Hughes Inc' 'BIOGEN IDEC Inc.' 'The Bank of New York Mellon Corp.' 'Ball Corp' 'Bristol-Myers Squibb' 'Boston Scientific' 'BorgWarner' 'Boston Properties' 'Citigroup Inc.' 'Caterpillar Inc.' 'Chubb Limited' 'CBRE Group' 'Crown Castle International Corp.' 'Carnival Corp.' 'Celgene Corp.' 'CF Industries Holdings Inc' 'Citizens Financial Group' 'Church & Dwight' 'C. H. Robinson Worldwide' 'Charter Communications' 'CIGNA Corp.' 'Cincinnati Financial' 'Colgate-Palmolive' 'Comerica Inc.' 'CME Group Inc.' 'Chipotle Mexican Grill' 'Cummins Inc.' 'CMS Energy' 'Centene Corporation' 'CenterPoint Energy' 'Capital One Financial' 'Cabot Oil & Gas' 'The Cooper Companies' 'CSX Corp.' 'CenturyLink Inc' 'Cognizant Technology Solutions' 'Citrix Systems' 'CVS Health' 'Chevron Corp.' 'Concho Resources' 'Dominion Resources' 'Delta Air Lines' 'Du Pont (E.I.)' 'Deere & Co.' 'Discover Financial Services' 'Quest Diagnostics' 'Danaher Corp.' 'The Walt Disney Company' 'Discovery Communications-A' 'Discovery Communications-C' 'Delphi Automotive' 'Digital Realty Trust' 'Dun & Bradstreet' 'Dover Corp.' 'Dr Pepper Snapple Group' 'Duke Energy' 'DaVita Inc.' 'Devon Energy Corp.' 'eBay Inc.' 'Ecolab Inc.' 'Consolidated Edison' 'Equifax Inc.' 'Edison Int'l' 'Eastman Chemical' 'EOG Resources' 'Equinix' 'Equity Residential' 'EQT Corporation' 'Evsource Energy' 'Essex Property Trust, Inc.' 'E*Trade' 'Eaton Corporation' 'Energy Corp.' 'Edwards Lifesciences' 'Exelon Corp.' 'Expeditors Int'l' 'Expedia Inc.' 'Extra Space Storage' 'Ford Motor' 'Fastenal Co' 'Fortune Brands Home & Security' 'Freeport-McMoran Cp & Gld' 'FirstEnergy Corp' 'Fidelity National Information Services' 'Fiserv Inc' 'FLIR Systems' 'Fluor Corp.' 'Flowserve Corporation' 'FMC Corporation' 'Federal Realty Investment Trust' 'First Solar Inc' 'Frontier Communications' 'General Dynamics' 'General Growth Properties Inc.' 'Gilead Sciences' 'Corning Inc.' 'General Motors' 'Genuine Parts' 'Garmin Ltd.' 'Goodyear Tire & Rubber' 'Grainger (W.W.) Inc.' 'Halliburton Co.' 'Hasbro Inc.' 'Huntington Bancshares' 'HCA Holdings' 'Welltower Inc.' 'HCP Inc.' 'Hess Corporation' 'Hartford Financial Svc.Gp.' 'Harley-Davidson' 'Honeywell Int'l Inc.' 'Hewlett Packard Enterprise' 'HP Inc.' 'Hormel Foods Corp.' 'Henry Schein' 'Host Hotels & Resorts' 'The Hershey Company' 'Humana Inc.' 'International Business Machines' 'IDEXX Laboratories' 'Intl Flavors & Fragrances' 'International Paper' 'Interpublic Group' 'Iron Mountain Incorporated' 'Intuitive Surgical Inc.' 'Illinois Tool Works' 'Invesco Ltd.' 'J. B. Hunt Transport Services' 'Jacobs Engineering Group' 'Juniper Networks' 'JPMorgan Chase & Co.' 'Kimco Realty' 'Kimberly-Clark' 'Kinder Morgan' 'Coca Cola Company' 'Kansas City Southern' 'Leggett & Platt' 'Lennar Corp.' 'Laboratory Corp. of America Holding' 'LKQ Corporation' 'L-3 Communications Holdings' 'Lilly (Eli) & Co.' 'Lockheed Martin Corp.' 'Alliant Energy Corp' 'Leucadia National Corp.' 'Southwest Airlines' 'Level 3 Communications' 'LyondellBasell' 'Mastercard Inc.' 'Mid-America Apartments' 'Macerich' 'Marriott Int'l.' 'Masco Corp.' 'Mattel Inc.' 'McDonald's Corp.' 'Moody's Corp' 'Mondelez International' 'MetLife Inc.' 'Mohawk Industries' 'Mead Johnson' 'McCormick & Co.' 'Martin Marietta Materials' 'Marsh & McLennan' '3M Company' 'Monster Beverage' 'Altria Group Inc' 'The Mosaic Company' 'Marathon Petroleum' 'Merck & Co.' 'Marathon Oil Corp.' 'M&T Bank Corp.' 'Mettler Toledo' 'Murphy Oil' 'Mylan N.V.' 'Navient' 'Noble Energy Inc' 'NASDAQ OMX Group' 'NextEra Energy' 'Newmont Mining Corp. (Hldg. Co.)' 'Netflix Inc.' 'Newfield Exploration Co' 'Nielsen Holdings' 'National Oilwell Varco Inc.' 'Norfolk Southern Corp.' 'Northern Trust Corp.' 'Nucor Corp.' 'Newell Brands' 'Realty Income Corporation' 'ONEOK' 'Omnicom Group' 'O'Reilly Automotive' 'Occidental Petroleum' 'People's United Financial' 'Pitney-Bowes' 'PACCAR Inc.' 'PG&E Corp.' 'Public Serv. Enterprise Inc.' 'PepsiCo Inc.' 'Pfizer Inc.' 'Principal Financial Group' 'Procter & Gamble' 'Progressive Corp.' 'Pulte Homes Inc.' 'Philip Morris International' 'PNC Financial Services' 'Pentair Ltd.' 'Pinnacle West Capital' 'PPG Industries' 'PPL Corp.' 'Prudential Financial' 'Phillips 66' 'Quanta Services Inc.' 'Praxair Inc.' 'PayPal' 'Ryder System' 'Royal Caribbean Cruises Ltd' 'Regeneron' 'Robert Half International' 'Roper Industries' 'Range Resources Corp.' 'Republic Services Inc' 'SCANA Corp' 'Charles Schwab Corporation' 'Spectra Energy Corp.' 'Sealed Air' 'Sherwin-Williams' 'SL Green Realty' 'Scripps Networks Interactive Inc.' 'Southern Co.' 'Simon Property Group Inc' 'S&P Global, Inc.' 'Stericycle Inc' 'Sempra Energy' 'SunTrust Banks' 'State Street Corp.' 'Skyworks Solutions' 'Southwestern Energy' 'Synchrony Financial' 'Stryker Corp.' 'AT&T Inc' 'Molson Coors Brewing Company' 'Teradata Corp.' 'Tegna, Inc.' 'Torchmark Corp.' 'Thermo Fisher Scientific' 'TripAdvisor' 'The Travelers Companies Inc.' 'Tractor Supply Company' 'Tyson Foods' 'Tesoro Petroleum Co.' 'Total System Services' 'Texas Instruments' 'Under Armour' 'United Continental Holdings' 'UDR Inc' 'Universal Health Services, Inc.' 'United Health Group Inc.' 'Unum Group' 'Union Pacific' 'United Parcel Service' 'United Technologies' 'Varian Medical Systems' 'Valero Energy' 'Vulcan Materials' 'Vornado Realty Trust' 'Verisk Analytics' 'Verisign Inc.' 'Vertex Pharmaceuticals Inc' 'Ventas Inc' 'Verizon Communications' 'Waters Corporation' 'Wec Energy Group Inc' 'Wells Fargo' 'Whirlpool Corp.' 'Waste Management Inc.' 'Williams Cos.' 'Western Union Co' 'Weyerhaeuser Corp.' 'Wyndham Worldwide' 'Wynn Resorts Ltd' 'Cimarex Energy' 'Xcel Energy Inc' 'XL Capital' 'Exxon Mobil Corp.' 'Dentsply Sirona' 'Xerox Corp.' 'Xylem Inc.' 'Yahoo Inc.' 'Yum! Brands Inc' 'Zimmer Biomet Holdings' 'Zions Bancorp' 'Zoetis'
- In cluster 5, the following companies are present: ['Alliance Data Systems'] In cluster 2, the following companies are present: ['Apache Corporation' 'Chesapeake Energy']
- In cluster 1, the following companies are present: ['Bank of America Corp' 'Intel Corp.]
- In cluster 3, the following companies are present: ['Facebook']
- In cluster 4, the following companies are present: ['Priceline.com Inc']

APPENDIX 3 - Hierarchical Clustering Portfolio 2

- In cluster 3, the following companies are present: ['American Airlines Group' 'AbbVie' 'Abbott Laboratories' 'Adobe Systems Inc' 'Analog Devices, Inc.' 'Archer-Daniels-Midland Co' 'Ameren Corp' 'American Electric Power' 'AFLAC Inc' 'American International Group, Inc.' 'Apartment Investment & Mgmt' 'Assurant Inc' 'Arthur J. Gallagher & Co.' 'Akamai Technologies Inc' 'Albemarle Corp' 'Alaska Air Group Inc' 'Allstate Corp' 'Applied Materials Inc' 'AMETEK Inc' 'Affiliated Managers Group Inc' 'Ameriprise Financial' 'American Tower Corp A' 'AutoNation Inc' 'Anthem Inc.' 'Aon plc' 'Amphenol Corp' 'Arconic Inc' 'Activision Blizzard' 'AvalonBay Communities, Inc.' 'Broadcom' 'American Water Works Company Inc' 'American Express Co' 'Boeing Company' 'Baxter International Inc.' 'BB&T Corporation' 'Bard (C.R.) Inc.' 'BIOGEN IDEC Inc.' 'The Bank of New York Mellon Corp.' 'Ball Corp' 'Bristol-Myers Squibb' 'Boston Scientific' 'BorgWarner' 'Boston Properties' 'Caterpillar Inc.' 'Chubb Limited' 'CBRE Group' 'Crown Castle International Corp.' 'Carnival Corp.' 'Celgene Corp.' 'CF Industries Holdings Inc' 'Citizens Financial Group' 'Church & Dwight' 'C. H. Robinson Worldwide' 'CIGNA Corp.' 'Cincinnati Financial' 'Comerica Inc.' 'CME Group Inc.' 'Cummins Inc.' 'CMS Energy' 'Centene Corporation' 'CenterPoint Energy' 'Capital One Financial' 'The Cooper Companies' 'CSX Corp.' 'CenturyLink Inc' 'Cognizant Technology Solutions' 'Citrix Systems' 'CVS Health' 'Chevron Corp.' 'Dominion Resources' 'Delta Air Lines' 'Du Pont (E.I.)' 'Deere & Co.' 'Discover Financial Services' 'Quest Diagnostics' 'Danaher Corp.' 'The Walt Disney Company' 'Discovery Communications-A' 'Discovery Communications-C' 'Delphi Automotive' 'Digital Realty Trust' 'Dun & Bradstreet' 'Dover Corp.' 'Dr Pepper Snapple Group' 'Duke Energy' 'DaVita Inc.' 'eBay Inc.' 'Ecolab Inc.' 'Consolidated Edison' 'Equifax Inc.' 'Edison Int'l' 'Eastman Chemical' 'Equity Residential' 'EQT Corporation' 'Eversource Energy' 'Essex Property Trust, Inc.' 'E*Trade' 'Eaton Corporation' 'Entergy Corp.' 'Edwards Lifesciences' 'Exelon Corp.' 'Expeditors Int'l' 'Expedia Inc.' 'Extra Space Storage' 'Fastenal Co' 'Fortune Brands Home & Security' 'FirstEnergy Corp' 'Fidelity National Information Services' 'Fiserv Inc' 'FLIR Systems' 'Fluor Corp.' 'Flowserve Corporation' 'FMC Corporation' 'Federal Realty Investment Trust' 'First Solar Inc' 'General Dynamics' 'General Growth Properties Inc.' 'Gilead Sciences' 'Corning Inc.' 'General Motors' 'Genuine Parts' 'Garmin Ltd.' 'Goodyear Tire & Rubber' 'Grainger (W.W.) Inc.' 'Halliburton Co.' 'Hasbro Inc.' 'Huntington Bancshares' 'HCA Holdings' 'Welltower Inc.' 'HHC Inc.' 'Hartford Financial Svc.Gp.' 'Harley-Davidson' 'Honeywell Int'l Inc.' 'Hewlett Packard Enterprise' 'HP Inc.' 'Hormel Foods Corp.' 'Henry Schein' 'Host Hotels & Resorts' 'The Hershey Company' 'Humana Inc.' 'International Business Machines' 'IDEXX Laboratories' 'Int'l Flavors & Fragrances' 'International Paper' 'Interpublic Group' 'Iron Mountain Incorporated' 'Illinois Tool Works' 'Invesco Ltd.' 'J. B. Hunt Transport Services' 'Jacobs Engineering Group' 'Juniper Networks' 'Kimco Realty' 'Kansas City Southern' 'Leggett & Platt' 'Lennar Corp.' 'Laboratory Corp. of America Holding' 'LKQ Corporation' 'L-3 Communications Holdings' 'Lilly (Eli) & Co.' 'Lockheed Martin Corp.' 'Alliant Energy Corp' 'Leucadia National Corp.' 'Southwest Airlines' 'Level 3 Communications' 'LyondellBasell' 'Mastercard Inc.' 'Mid-America Apartments' 'Macerich' 'Marriott Int'l.' 'Masco Corp.' 'Mattel Inc.' 'McDonald's Corp.' 'Moody's Corp' 'Mondelez International' 'MetLife Inc.' 'Mohawk Industries' 'Mead Johnson' 'McCormick & Co.' 'Martin Marietta Materials' 'Marsh & McLennan' '3M Company' 'Atria Group Inc' 'The Mosaic Company' 'Marathon Petroleum' 'Merck & Co.' 'M&T Bank Corp.' 'Mettler Toledo' 'Mylan N.V.' 'Navient' 'NASDAQ OMX Group' 'NextEra Energy' 'Newmont Mining Corp. (Hldg. Co.)' 'Nielsen Holdings' 'Norfolk Southern Corp.' 'Northern Trust Corp.' 'Nucor Corp.' 'Newell Brands' 'Realty Income Corporation' 'Omnicom Group' 'O'Reilly Automotive' 'People's United Financial' 'Pitney-Bowes' 'PACCAR Inc.' 'PG&E Corp.' 'Public Serv. Enterprise Inc.' 'PepsiCo Inc.' 'Principal Financial Group' 'Procter & Gamble' 'Progressive Corp.' 'Pulte Homes Inc.' 'Philip Morris International' 'PNC Financial Services' 'Pentair Ltd.' 'Pinnacle West Capital' 'PPG Industries' 'PPL Corp.' 'Prudential Financial' 'Phillips 66' 'Quanta Services Inc.' 'Praxair Inc.' 'PayPal' 'Ryder System' 'Royal Caribbean Cruises Ltd' 'Robert Half International' 'Roper Industries' 'Republic Services Inc' 'SCANA Corp' 'Charles Schwab Corporation' 'Sealed Air' 'Sherwin-Williams' 'SL Green Realty' 'Scripps Networks Interactive Inc.' 'Southern Co.' 'Simon Property Group Inc' 'Stericycle Inc' 'Sempra Energy' 'SunTrust Banks' 'State Street Corp.' 'Skyworks Solutions' 'Synchroony Financial' 'Stryker Corp.' 'Molson Coors Brewing Company' 'Tegna, Inc.' 'Torchmark Corp.' 'Thermo Fisher Scientific' 'TripAdvisor' 'The Travelers Companies Inc.' 'Tractor Supply Company' 'Tyson Foods' 'Tesoro Petroleum Co.' 'Total System Services' 'Texas Instruments' 'Under Armour' 'United Continental Holdings' 'UDR Inc' 'Universal Health Services, Inc.' 'United Health Group Inc.' 'Unum Group' 'Union Pacific' 'United Parcel Service' 'United Technologies' 'Varian Medical Systems' 'Valero Energy' 'Vulcan Materials' 'Vornado Realty Trust' 'Verisk Analytics' 'Verisign Inc.' 'Vertex Pharmaceuticals Inc' 'Ventas Inc' 'Wec Energy Group Inc' 'Whirlpool Corp.' 'Waste Management Inc.' 'Western Union Co' 'Weyerhaeuser Corp.' 'Wyndham Worldwide' 'Wynn Resorts Ltd' 'Xcel Energy Inc' 'XL Capital' 'Dentsply Sirona' 'Xerox Corp.' 'Xylem Inc.' 'Yum! Brands Inc' 'Zimmer Biomet Holdings' 'Zions Bancorp' 'Zoetis']
- In cluster 1, the following companies are present: ['Alliance Data Systems' 'Amgen Inc' 'Chipotle Mexican Grill' 'Equinix' 'Facebook' 'Frontier Communications' 'Intuitive Surgical Inc.' 'Monster Beverage' 'Priceline.com Inc' 'Regeneron' 'Waters Corporation' 'Yahoo Inc.]
- In cluster 4, the following companies are present: ['Allegion' 'Apache Corporation' 'Chesapeake Energy' 'Charter Communications' 'Colgate-Palmolive' 'Kimberly-Clark' 'S&P Global, Inc.]
- In cluster 5, the following companies are present: ['Alexion Pharmaceuticals' 'Amazon.com Inc' 'Netflix Inc.]
- In cluster 2, the following companies are present: ['Anadarko Petroleum Corp' 'Baker Hughes Inc' 'Cabot Oil & Gas' 'Concho Resources' 'Devon Energy Corp.' 'EOG Resources' 'Freeport-McMoran Cp & Gld' 'Hess Corporation' 'Kinder Morgan' 'Marathon Oil Corp.' 'Murphy Oil' 'Noble Energy Inc' 'Newfield Exploration Co' 'National Oilwell Varco Inc.' 'ONEOK' 'Occidental Petroleum' 'Range Resources Corp.' 'Spectra Energy Corp.' 'Southwestern Energy' 'Teradata Corp.' 'Williams Cos.' 'Cimarex Energy']
- In cluster 6, the following companies are present: ['Bank of America Corp' 'Citigroup Inc.' 'Ford Motor' 'Intel Corp.' 'JPMorgan Chase & Co.' 'Coca Cola Company' 'Pfizer Inc.' 'AT&T Inc' 'Verizon Communications' '37' 'Wells Fargo' 'Exxon Mobil Corp.]

APPENDIX 4 – Data Summary

count	unique	top	freq	mean	std	min	0.25	0.50	0.75	max
Ticker Symbol	340.00	340.00	AAL	1.00	NaN	NaN	NaN	NaN	NaN	NaN
Security	340.00	340.00	American Airlines Group	1.00	NaN	NaN	NaN	NaN	NaN	NaN
GICS Sector	340.00	11.00	Industrials	53.00	NaN	NaN	NaN	NaN	NaN	NaN
GICS Sub Industry	340.00	104.00	Oil & Gas Exploration & Production	16.00	NaN	NaN	NaN	NaN	NaN	NaN
Current Price	340.00	NaN	NaN	NaN	80.86	98.06	4.50	38.56	59.71	92.88
Price Change	340.00	NaN	NaN	NaN	4.08	12.01	-47.13	-0.94	4.82	10.70
Volatility	340.00	NaN	NaN	NaN	1.53	0.59	0.73	1.13	1.39	1.70
ROE	340.00	NaN	NaN	NaN	39.60	96.55	1.00	9.75	15.00	27.00
Cash Ratio	340.00	NaN	NaN	NaN	70.02	90.42	0.00	18.00	47.00	99.00
Net Cash Flow	340.00	NaN	NaN	NaN	55537620.59	1946365312.18	-11208000000.00	-193906500.00	2098000.00	169810750.00
Net Income	340.00	NaN	NaN	NaN	1494384602.94	3940150279.33	-23528000000.00	352301250.00	707336000.00	1899000000.00
Earnings Per Share	340.00	NaN	NaN	NaN	2.78	6.59	-61.20	1.56	2.90	4.62
Estimated Shares Outstanding	340.00	NaN	NaN	NaN	577028337.75	845849595.42	27672156.86	158848216.10	309675137.80	573117457.33
P/E Ratio	340.00	NaN	NaN	NaN	32.61	44.35	2.94	15.04	20.82	31.76
P/B Ratio	340.00	NaN	NaN	NaN	-1.72	13.97	-76.12	-4.35	-1.07	3.92
										129.06