# TRADE & AHEAD PROJECT

BY: NIHAL KALA

# PROBLEM STATEMENT

- There are many financial metrics one must look at to make stock investments that would maximize earnings and lower risk in the stock portfolio.
- With the use of cluster analysis, we can determine which groups of stocks have similar characteristics and minimum correlation.
- Objective: Trade & Ahead, a financial consultancy firm, has asked to analyze stock price and other metrics for various companies and group stocks in clusters based on the attributes and find insights about each of these attributes.
- Some questions to ask:
  - Which characteristic are we going to see stocks grouped most similarly and least similarly?
  - What are the difference between using K-cluster and hierarchical clustering methods?
  - What are the most useful financial indicators one must look when investing in stocks?

# DATA SAMPLE

| | Ticker Symbol | Security | GICS Sector | GICS Sub Industry | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **102** | DVN | Devon Energy Corp. | Energy | Oil & Gas Exploration & Production | 32.000000 | -15.478079 | 2.923698 | 205 | 70 | 830000000 | -14454000000 | -35.55 | 4.065823e+08 | 93.089287 | 1.785616 |
| **125** | FB | Facebook | Information Technology | Internet Software & Services | 104.660004 | 16.224320 | 1.320606 | 8 | 958 | 592000000 | 3669000000 | 1.31 | 2.800763e+09 | 79.893133 | 5.884467 |
| **11** | AIV | Apartment Investment & Mgmt | Real Estate | REITs | 40.029999 | 7.578608 | 1.163334 | 15 | 47 | 21818000 | 248710000 | 1.52 | 1.636250e+08 | 26.335526 | -1.269332 |
| **248** | PG | Procter & Gamble | Consumer Staples | Personal Products | 79.410004 | 10.660538 | 0.806056 | 17 | 129 | 160383000 | 636056000 | 3.28 | 4.913916e+08 | 24.070121 | -2.256747 |
| **238** | OXY | Occidental Petroleum | Energy | Oil & Gas Exploration & Production | 67.610001 | 0.865287 | 1.589520 | 32 | 64 | -588000000 | -7829000000 | -10.23 | 7.652981e+08 | 93.089287 | 3.345102 |
| **336** | YUM | Yum! Brands Inc | Consumer Discretionary | Restaurants | 52.516175 | -8.698917 | 1.478877 | 142 | 27 | 159000000 | 1293000000 | 2.97 | 4.353535e+08 | 17.682214 | -3.838260 |
| **112** | EQT | EQT Corporation | Energy | Oil & Gas Exploration & Production | 52.130001 | -21.253771 | 2.364883 | 2 | 201 | 523803000 | 85171000 | 0.56 | 1.520911e+08 | 93.089287 | 9.567952 |
| **147** | HAL | Halliburton Co. | Energy | Oil & Gas Equipment & Services | 34.040001 | -5.101751 | 1.966062 | 4 | 189 | 7786000000 | -671000000 | -0.79 | 8.493671e+08 | 93.089287 | 17.345857 |
| **89** | DFS | Discover Financial Services | Financials | Consumer Finance | 53.619999 | 3.653584 | 1.159897 | 20 | 99 | 2288000000 | 2297000000 | 5.14 | 4.468872e+08 | 10.431906 | -0.375934 |
| **173** | IVZ | Invesco Ltd. | Financials | Asset Management & Custody Banks | 33.480000 | 7.067477 | 1.580839 | 12 | 67 | 412000000 | 968100000 | 2.26 | 4.283628e+08 | 14.814159 | 4.218620 |

Data Shape: 340 rows, 15 columns

# DATA INFO

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 340 entries, 0 to 339
Data columns (total 15 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Ticker Symbol               340 non-null    object
 1   Security                    340 non-null    object
 2   GICS Sector                 340 non-null    object
 3   GICS Sub Industry           340 non-null    object
 4   Current Price               340 non-null    float64
 5   Price Change                340 non-null    float64
 6   Volatility                  340 non-null    float64
 7   ROE                         340 non-null    int64
 8   Cash Ratio                  340 non-null    int64
 9   Net Cash Flow               340 non-null    int64
 10  Net Income                  340 non-null    int64
 11  Earnings Per Share          340 non-null    float64
 12  Estimated Shares Outstanding  340 non-null  float64
 13  P/E Ratio                   340 non-null    float64
 14  P/B Ratio                   340 non-null    float64
dtypes: float64(7), int64(4), object(4)
memory usage: 40.0+ KB
```

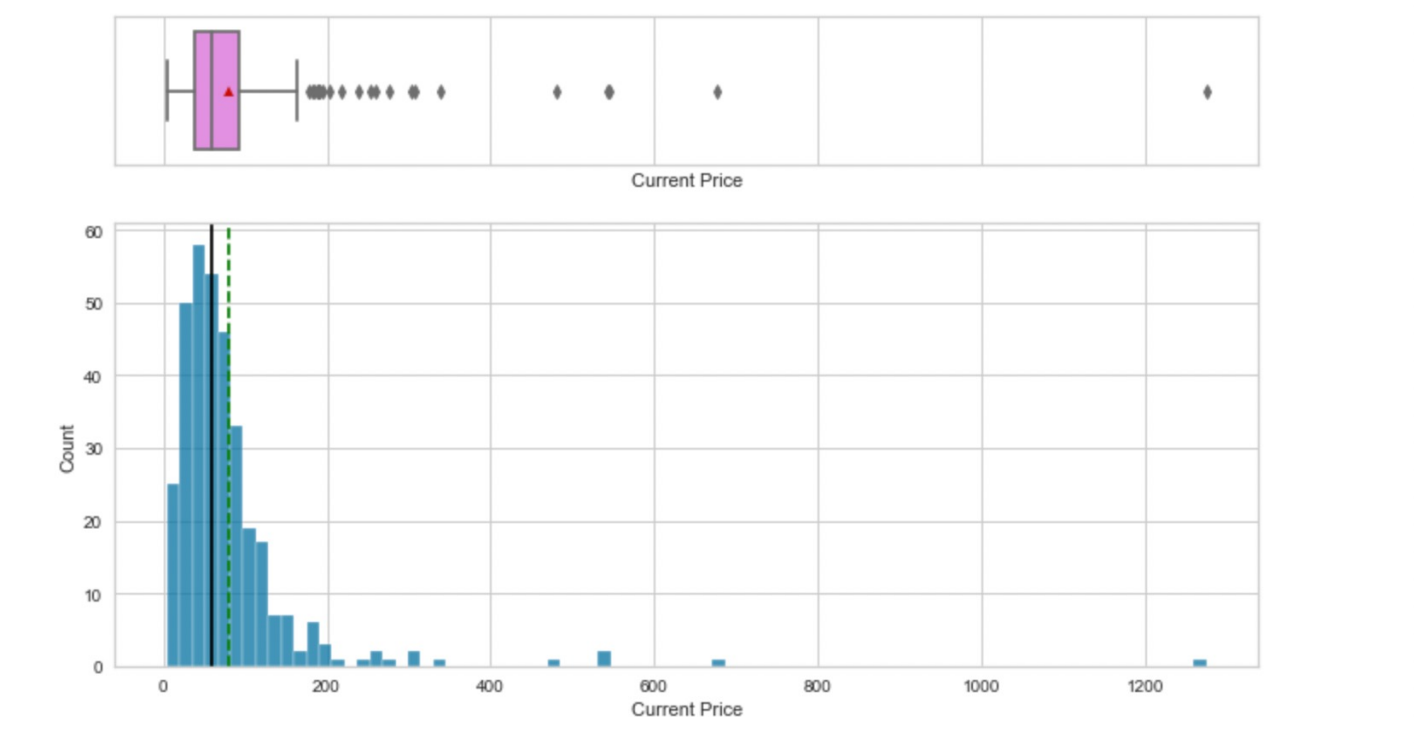
No duplicated or null values.

# DESCRIPTION OF DATA

| | count | unique | top | freq | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Ticker Symbol** | 340 | 340 | D | 1 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| **Security** | 340 | 340 | Nielsen Holdings | 1 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| **GICS Sector** | 340 | 11 | Industrials | 53 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| **GICS Sub Industry** | 340 | 104 | Oil & Gas Exploration & Production | 16 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| **Current Price** | 340.0 | NaN | NaN | NaN | 80.862345 | 98.055086 | 4.5 | 38.555 | 59.705 | 92.880001 | 1274.949951 |
| **Price Change** | 340.0 | NaN | NaN | NaN | 4.078194 | 12.006338 | -47.129693 | -0.939484 | 4.819505 | 10.695493 | 55.051683 |
| **Volatility** | 340.0 | NaN | NaN | NaN | 1.525976 | 0.591798 | 0.733163 | 1.134878 | 1.385593 | 1.695549 | 4.580042 |
| **ROE** | 340.0 | NaN | NaN | NaN | 39.597059 | 96.547538 | 1.0 | 9.75 | 15.0 | 27.0 | 917.0 |
| **Cash Ratio** | 340.0 | NaN | NaN | NaN | 70.023529 | 90.421331 | 0.0 | 18.0 | 47.0 | 99.0 | 958.0 |
| **Net Cash Flow** | 340.0 | NaN | NaN | NaN | 55537620.588235 | 1946365312.175789 | -11208000000.0 | -193906500.0 | 2098000.0 | 169810750.0 | 20764000000.0 |
| **Net Income** | 340.0 | NaN | NaN | NaN | 1494384602.941176 | 3940150279.327937 | -23528000000.0 | 352301250.0 | 707336000.0 | 1899000000.0 | 24442000000.0 |
| **Earnings Per Share** | 340.0 | NaN | NaN | NaN | 2.776662 | 6.587779 | -61.2 | 1.5575 | 2.895 | 4.62 | 50.09 |
| **Estimated Shares Outstanding** | 340.0 | NaN | NaN | NaN | 577028337.754029 | 845849595.417695 | 27672156.86 | 158848216.1 | 309675137.8 | 573117457.325 | 6159292035.0 |
| **P/E Ratio** | 340.0 | NaN | NaN | NaN | 32.612563 | 44.348731 | 2.935451 | 15.044653 | 20.819876 | 31.764755 | 528.039074 |
| **P/B Ratio** | 340.0 | NaN | NaN | NaN | -1.718249 | 13.966912 | -76.119077 | -4.352056 | -1.06717 | 3.917066 | 129.064585 |

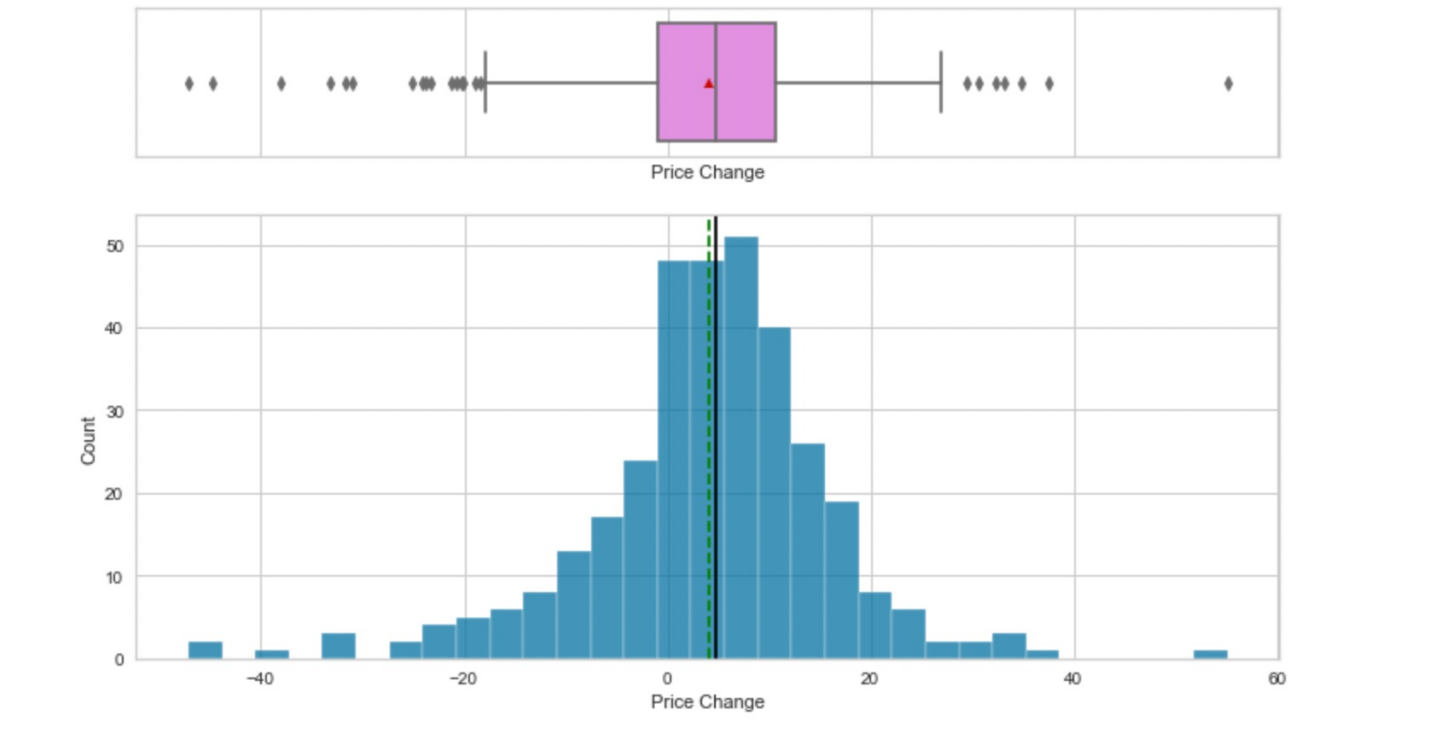
Average Current Price: 80.862
Average Price Change: 4.08
Average Volatility: 1.53
Average ROE: 39.6
Average Cash Ratio: 70
Average EPS: 2.776662
Average P/E Ratio 32.61
Average P/B Ratio: -1.72

# HISTOGRAM-BOX PLOT: CURRENT PRICE



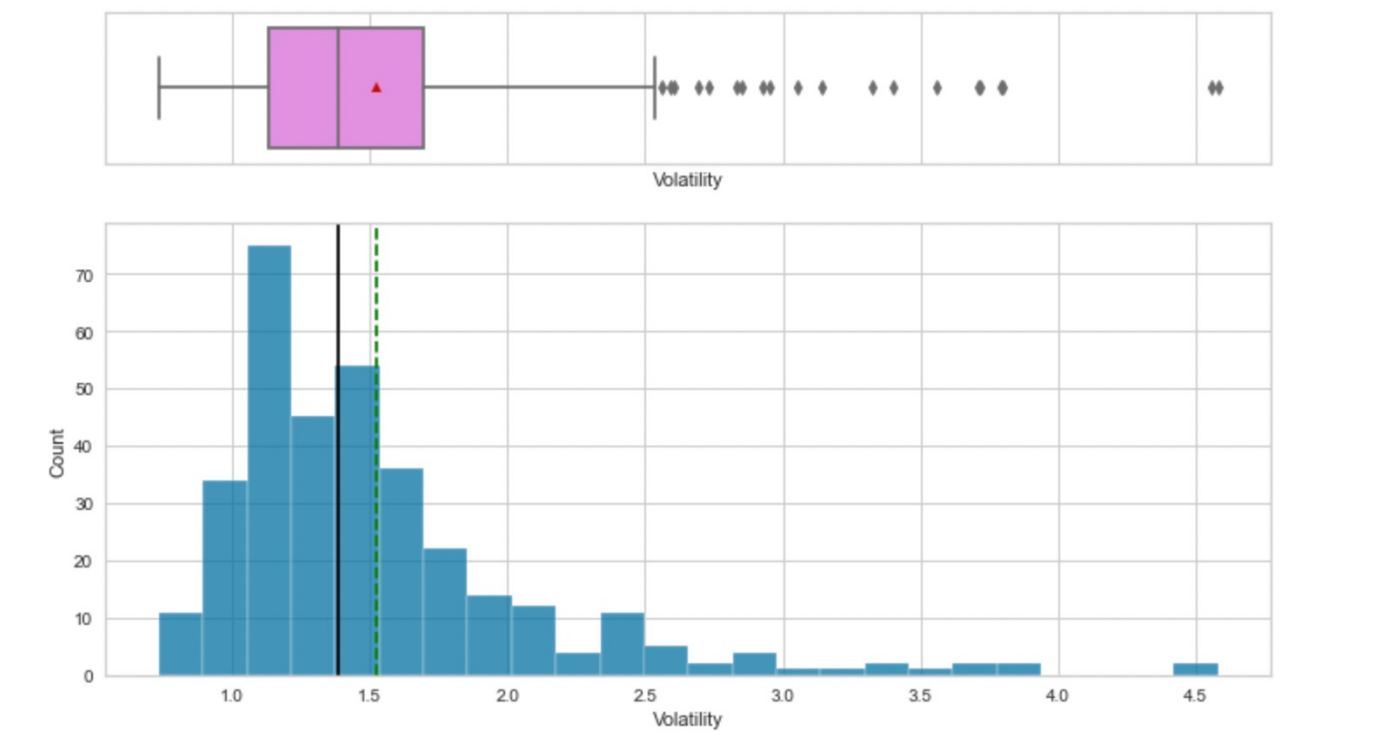

Mean: 80.862345
Range: 4.5-1275
Median: 59.705
IQR: 38.555-92.88
Right-skewed distribution with upper outliers, suggesting stocks with expensive current prices

# HISTOGRAM-BOX PLOT: PRICE CHANGE



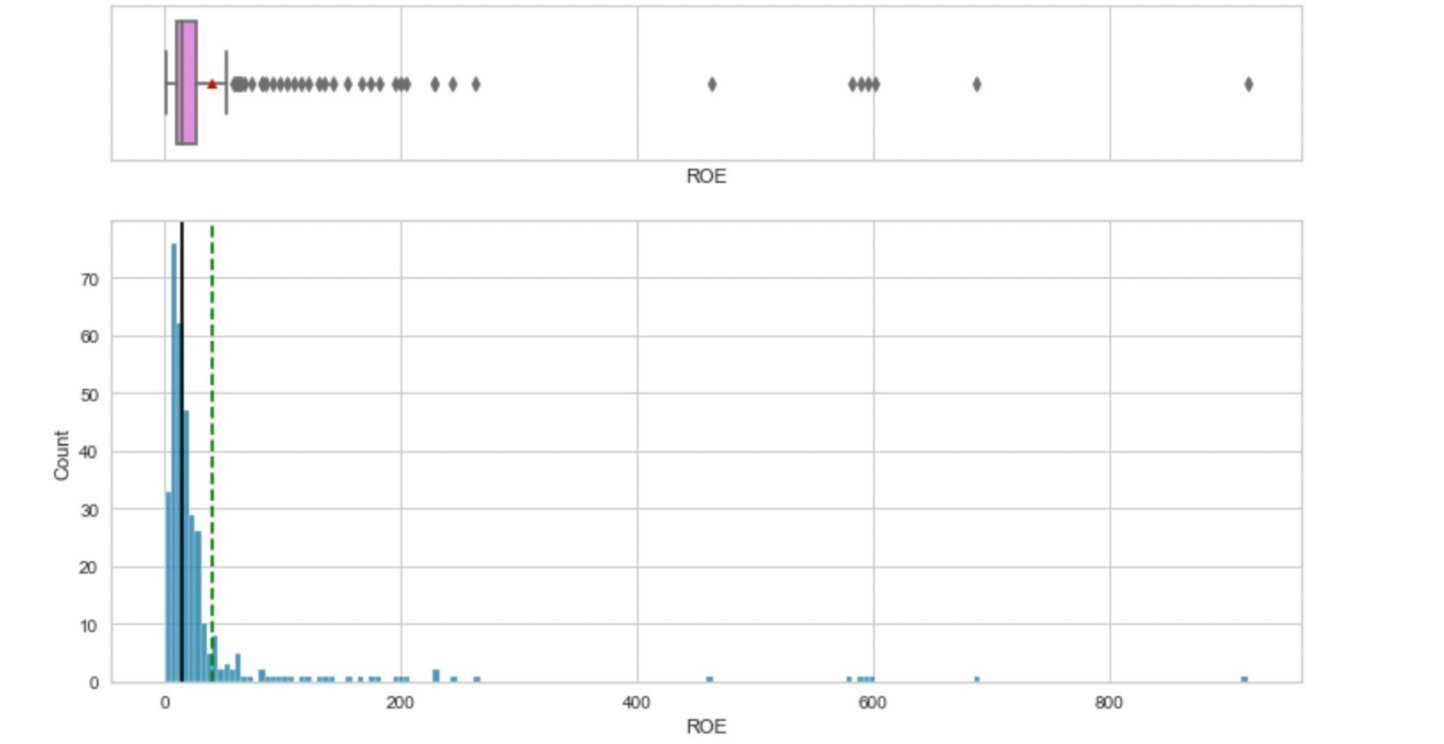

Mean: 4.078194
Range: -47.129693-55.051683
Median: 4.819505
IQR: -0.939484-10.695493
More outliers to the left than the right,
suggesting more negative price changes

# HISTOGRAM-BOX PLOT: VOLATILITY



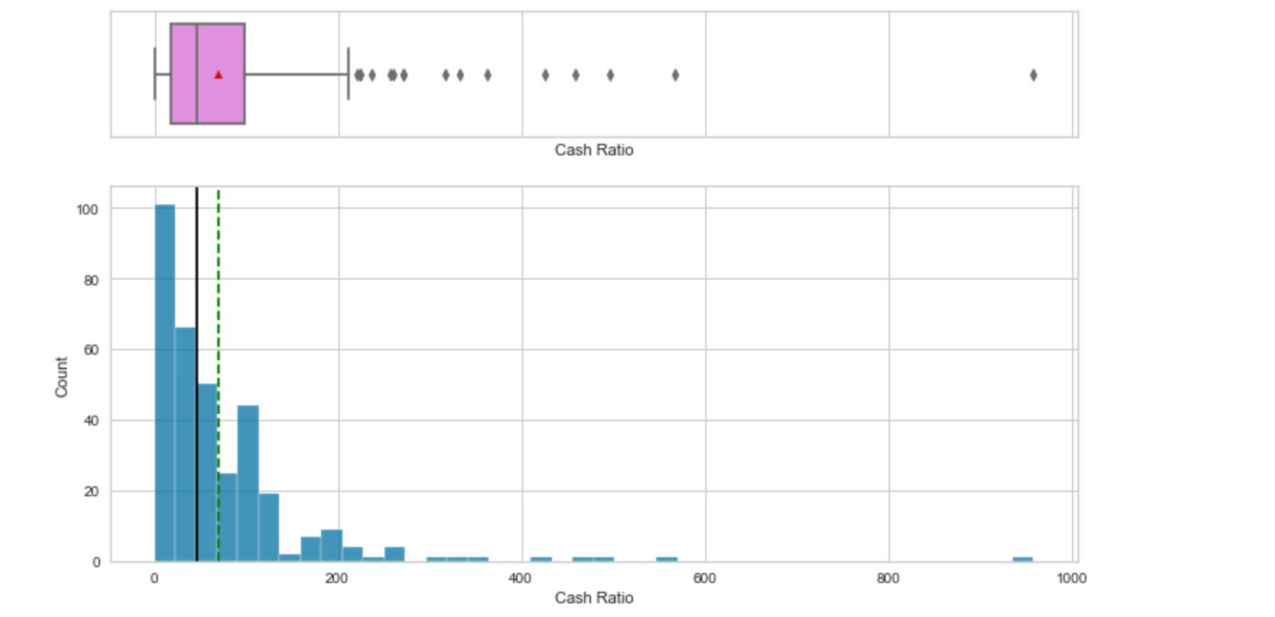

Mean: 1.525976
Range: 0.733163-4.580042
Median: 1.385593
IQR: 1.134878-1.695549
Right-skewed distribution with upper outliers, suggesting stocks with very high volatility

# HISTOGRAM-BOX PLOT: ROE



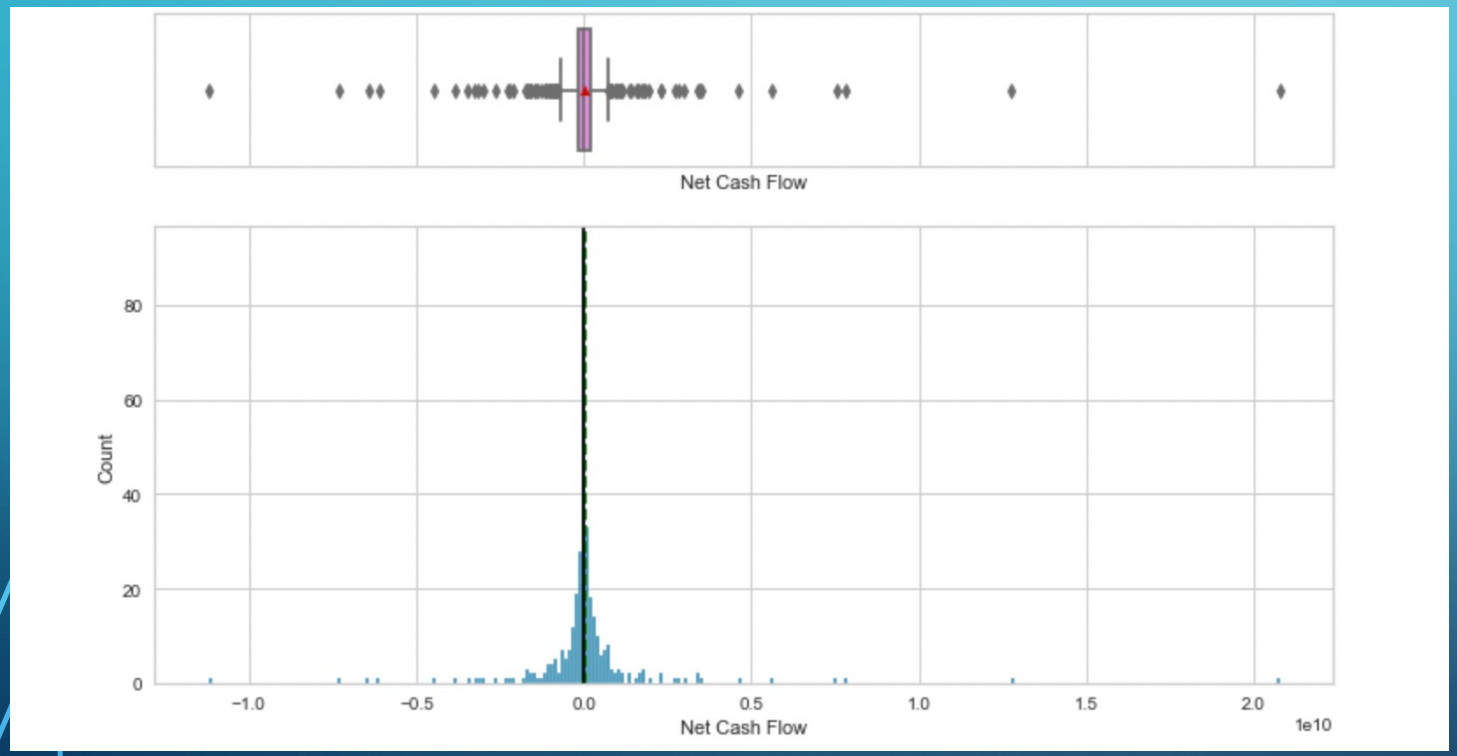

Mean: 39.597059
Range: 1-917
Median: 15
IQR: 9.75-27
Heavily Right-skewed distribution with upper outliers, suggesting stocks with very high return on equity

# HISTOGRAM-BOX PLOT: CASH RATIO



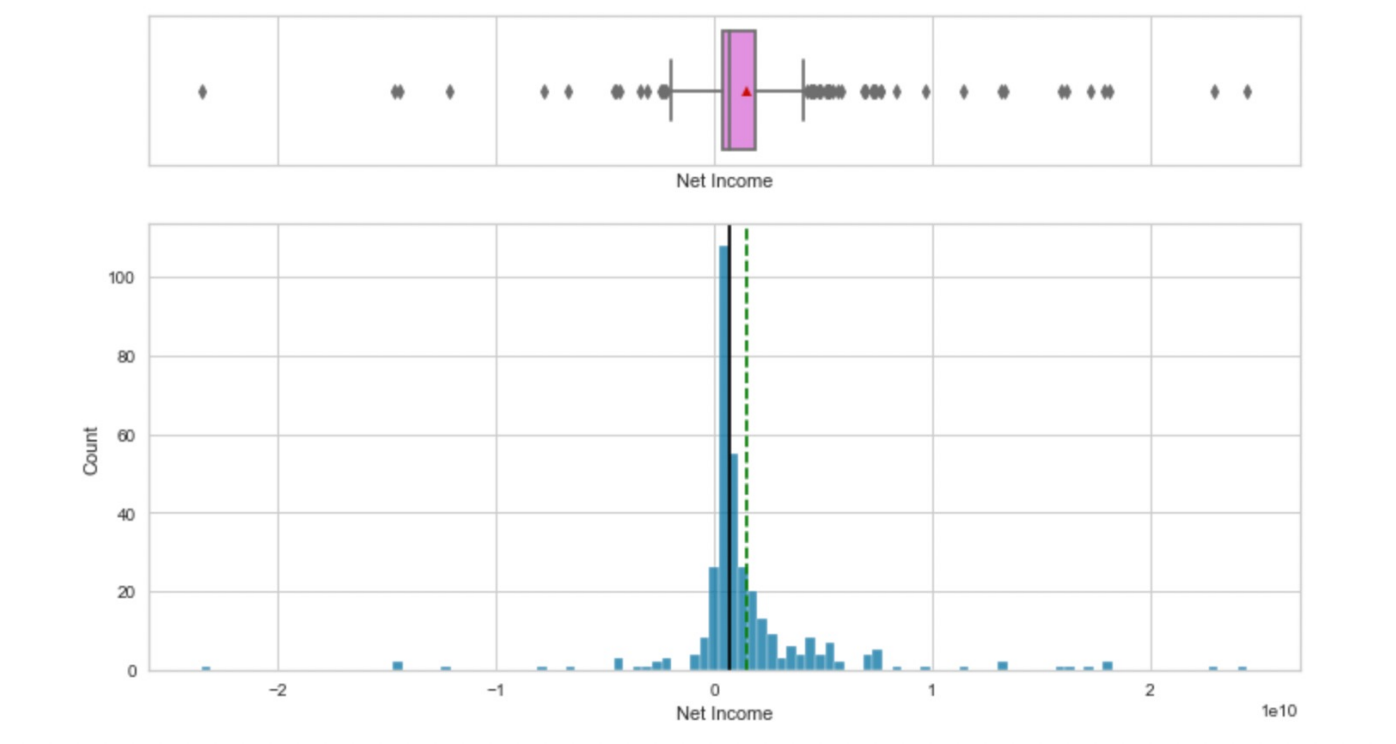

Mean: 70.024
Range: 0-958
Median: 47
IQR: 18-99
Right-skewed distribution with upper outliers, suggesting there are few stocks with high cash ratio

# HISTOGRAM-BOX PLOT: NET CASH FLOW



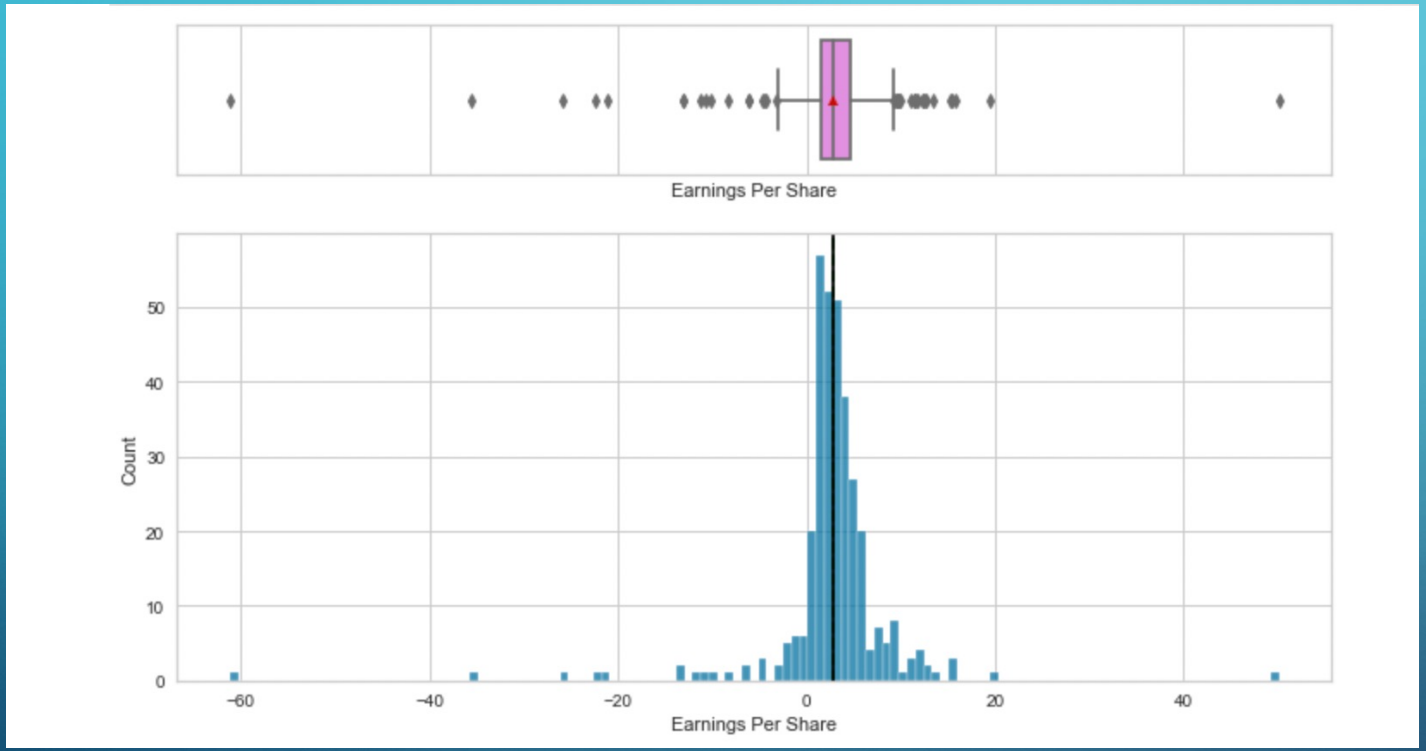

Mean: 55537620
Range: -11208000000-20764000000
Median: 2098000
IQR: -193906500-169810750
Very wide distribution for Net Cash Flow with outliers to the left and right of the interquartile range.

# HISTOGRAM-BOX PLOT: NET INCOME



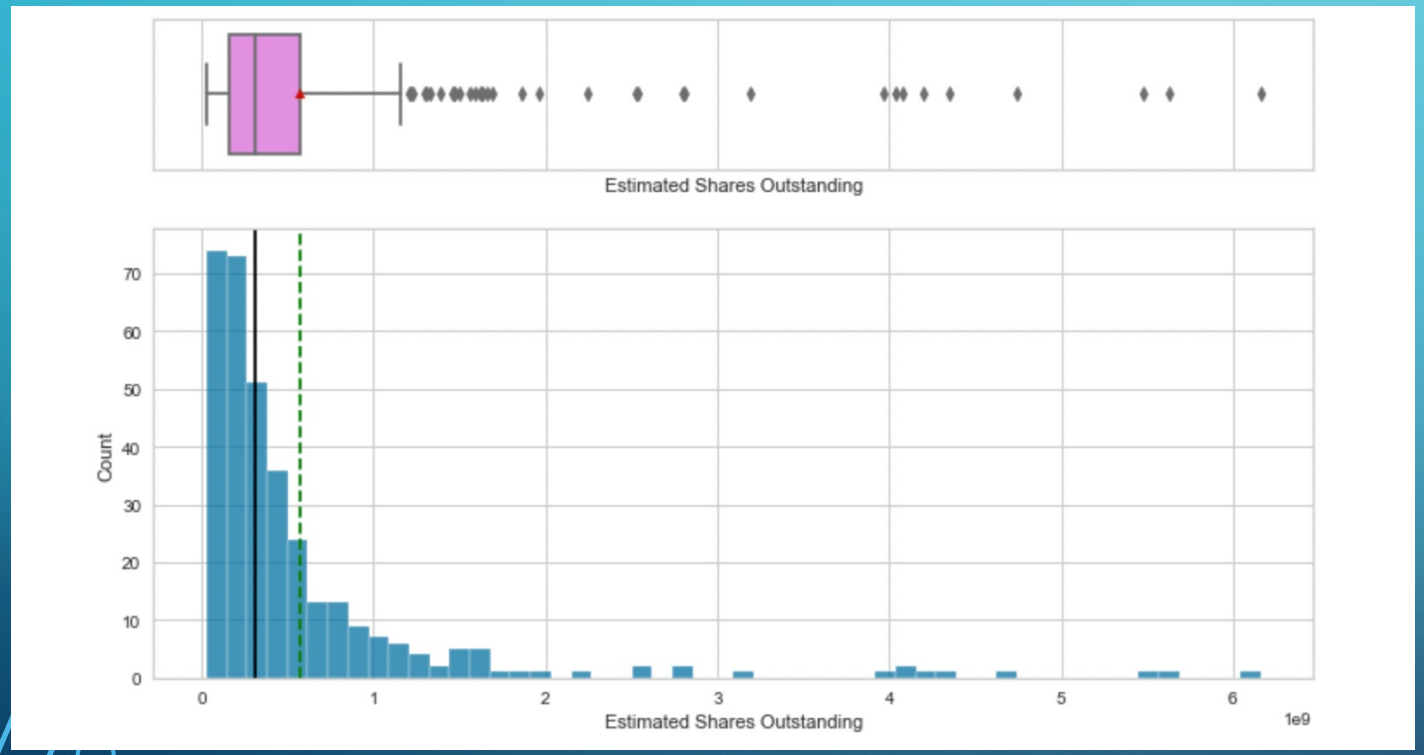

Mean: 1494384602
Range: -23528000000-24442000000
Median: 707336000
IQR: 352301250-1899000000
Very wide distribution for Net Income with outliers to the left and more to right of the interquartile range.

# HISTOGRAM-BOX PLOT: EARNINGS PER SHARE



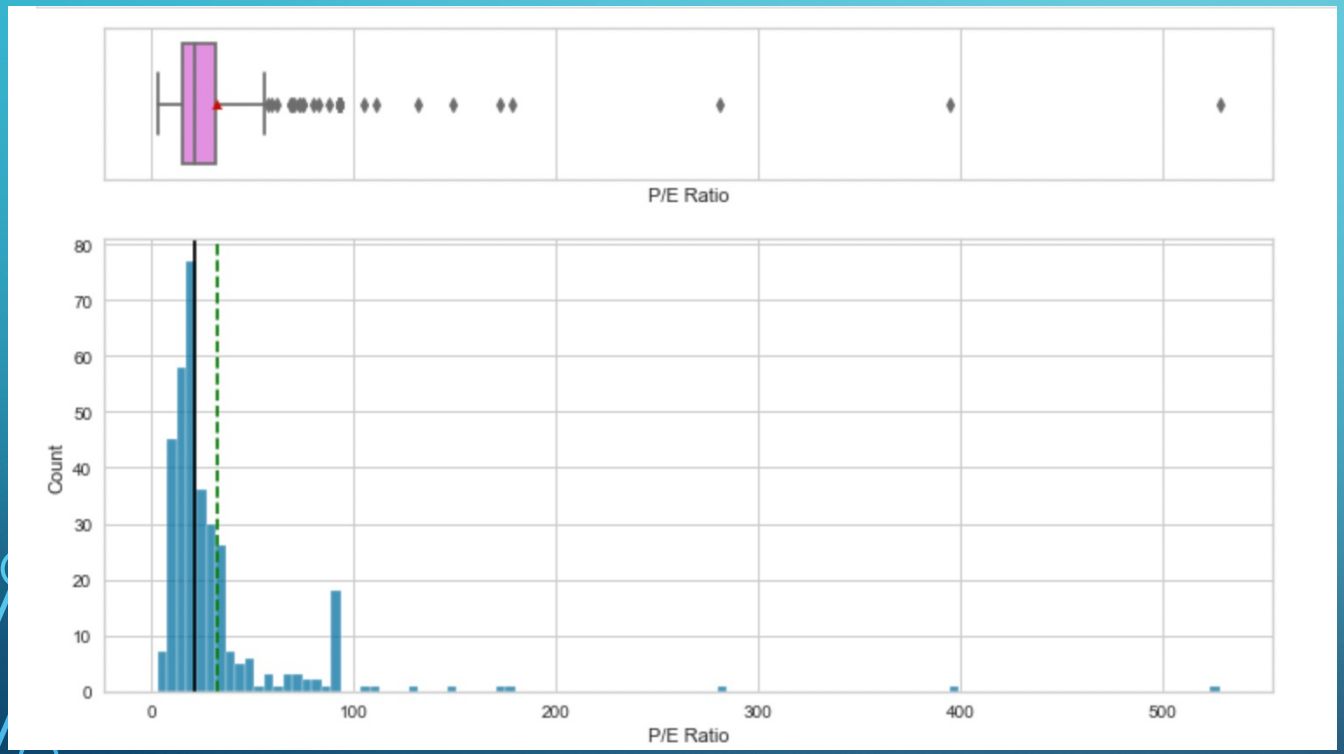

Mean: 2.77662
Range: -61.2-50.09
Median: 2.895
IQR: 1.5575-4.62
Earnings per share mean and median are nearly the same and there outliers to the left and right of the interquartile range.

# HISTOGRAM-BOX PLOT: ESTIMATED SHARES OUTSTANDING



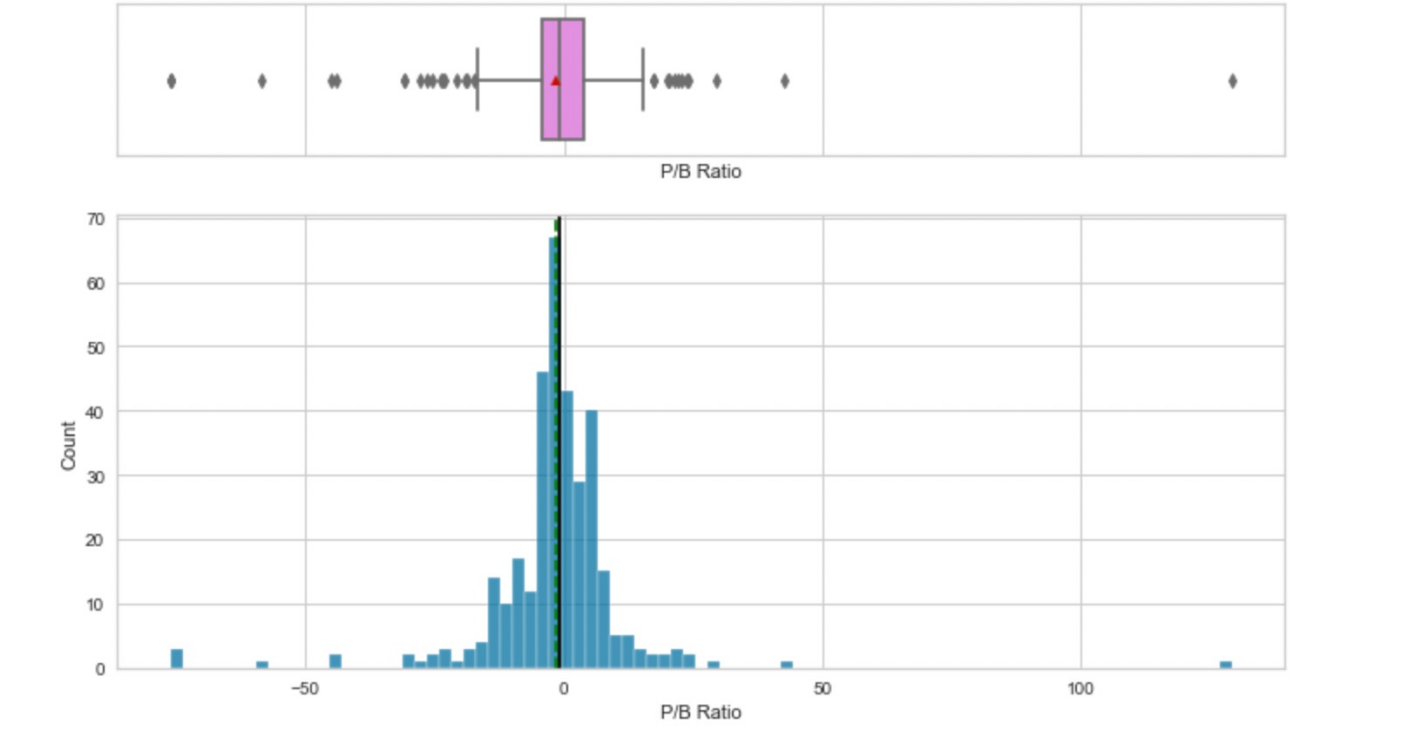

Mean: 577028337
Range: 27672156-6159290235
Median: 309675137
IQR: 158848216-573117457
-Very Right Skewed Distribution suggesting stocks with many estimated shares outstanding.

# HISTOGRAM-BOX PLOT: P/E RATIO



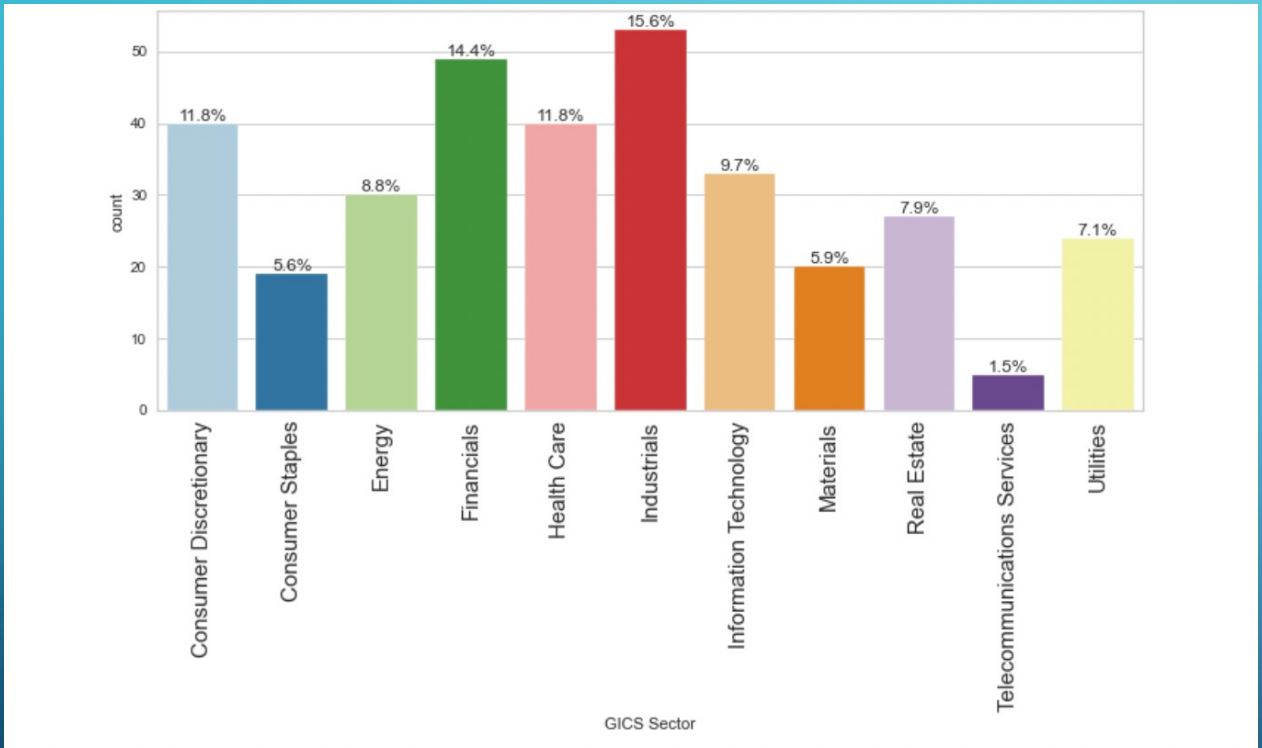

Mean: 32.612563
Range: 2.935451-528.039074
Median: 20.819876
IQR: 15.044653-31.764755
Right-skewed distribution suggesting stocks exist with high P/E Ratio

# HISTOGRAM-BOX PLOT: P/B RATIO



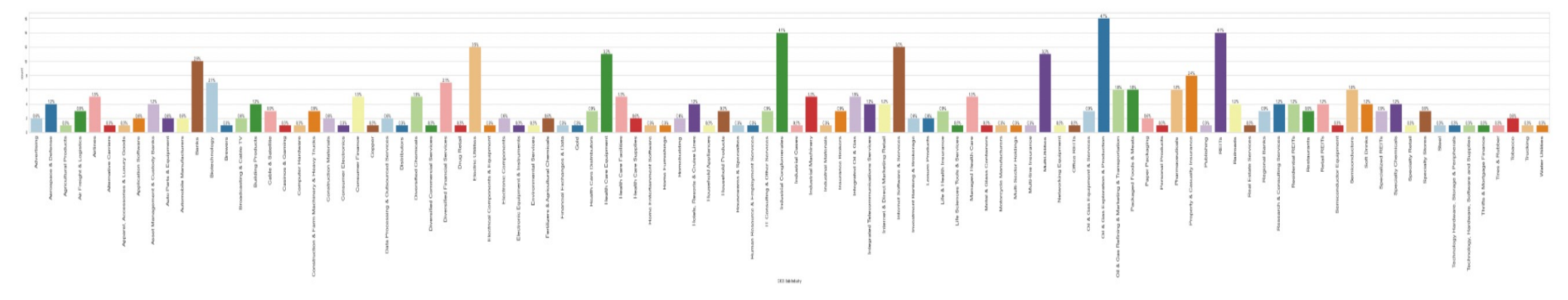

Mean: -1.718249
Range: -76.119077-129.064585
Median: -1.06717
IQR: -4.352056-3.917066
Outliers exist left and right of Interquartile Range

# GIC SECTORS BAR PLOT



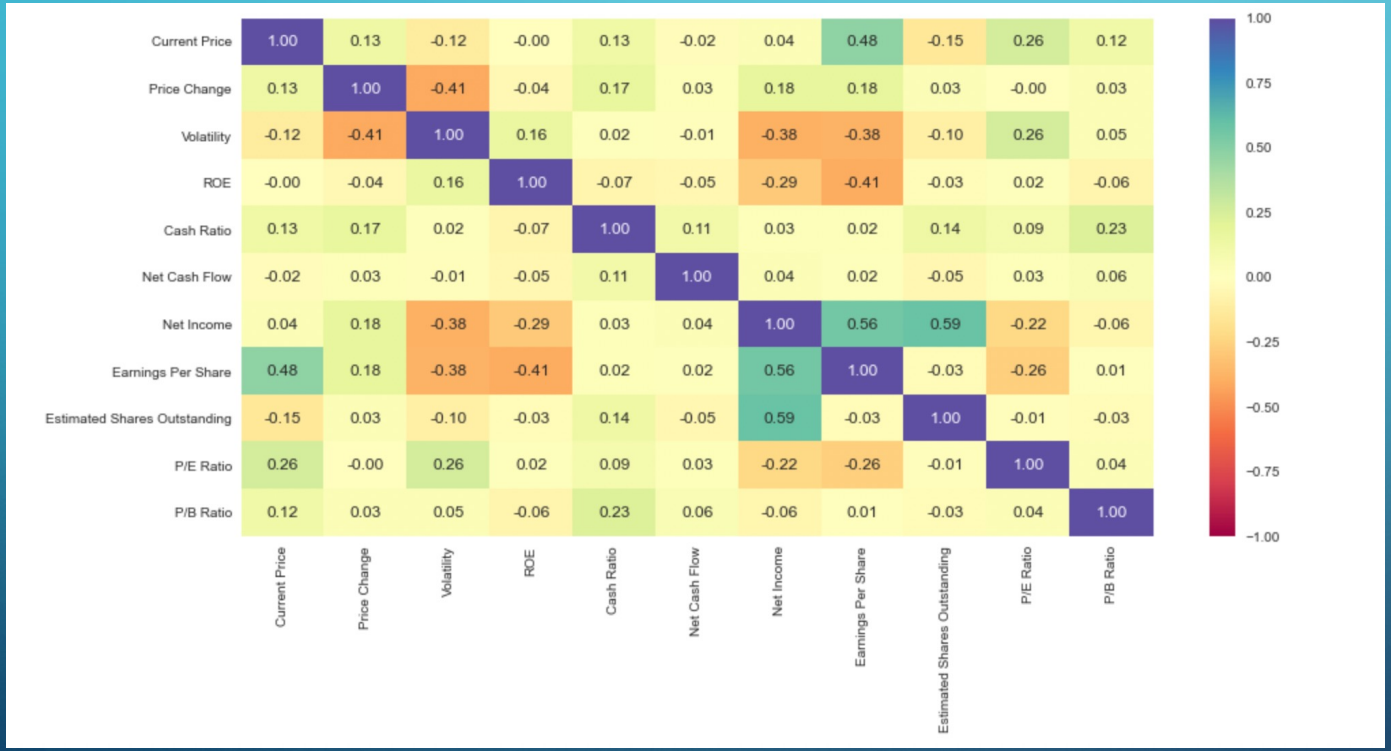

The GICS Sector with the greatest number companies is industrials, while the least number is telecommunications services.

# GIS SUB SECTORS BAR PLOT



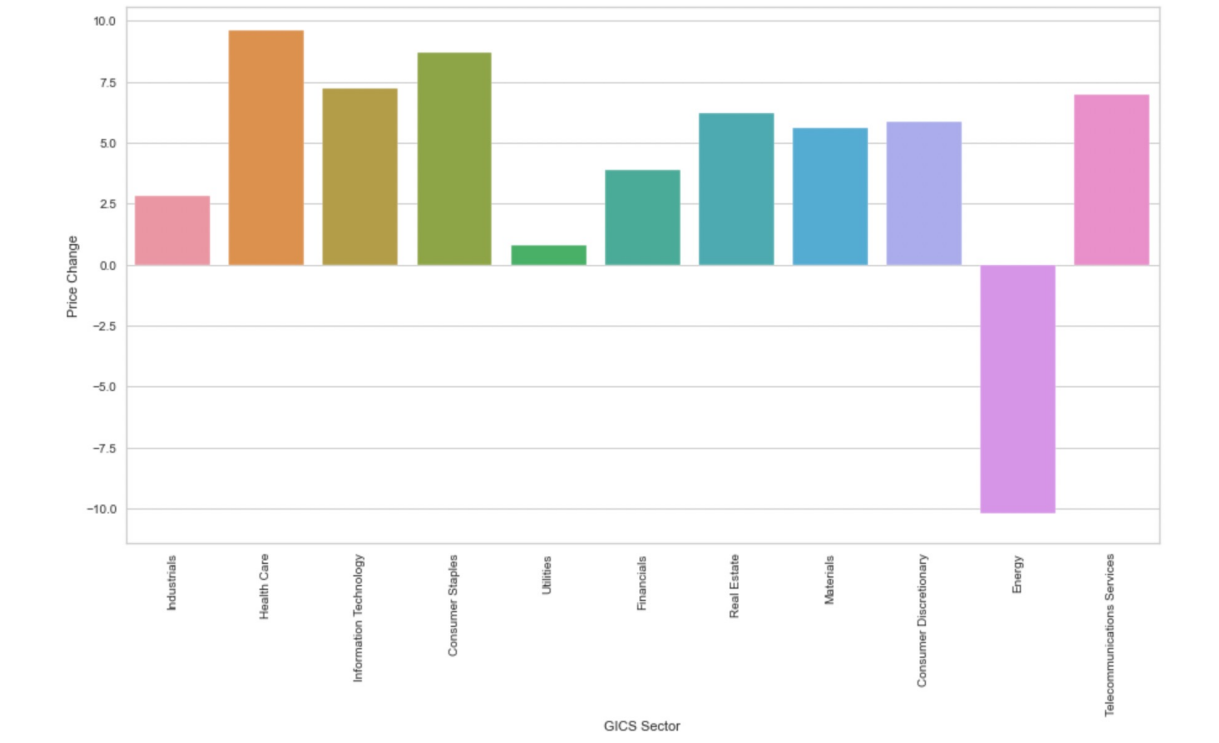
THE GICS SUB SECTORS with the greatest number of companies is oil/gas production, REITS, and industrial conglomerates.

# CORRELATION HEAT MAP



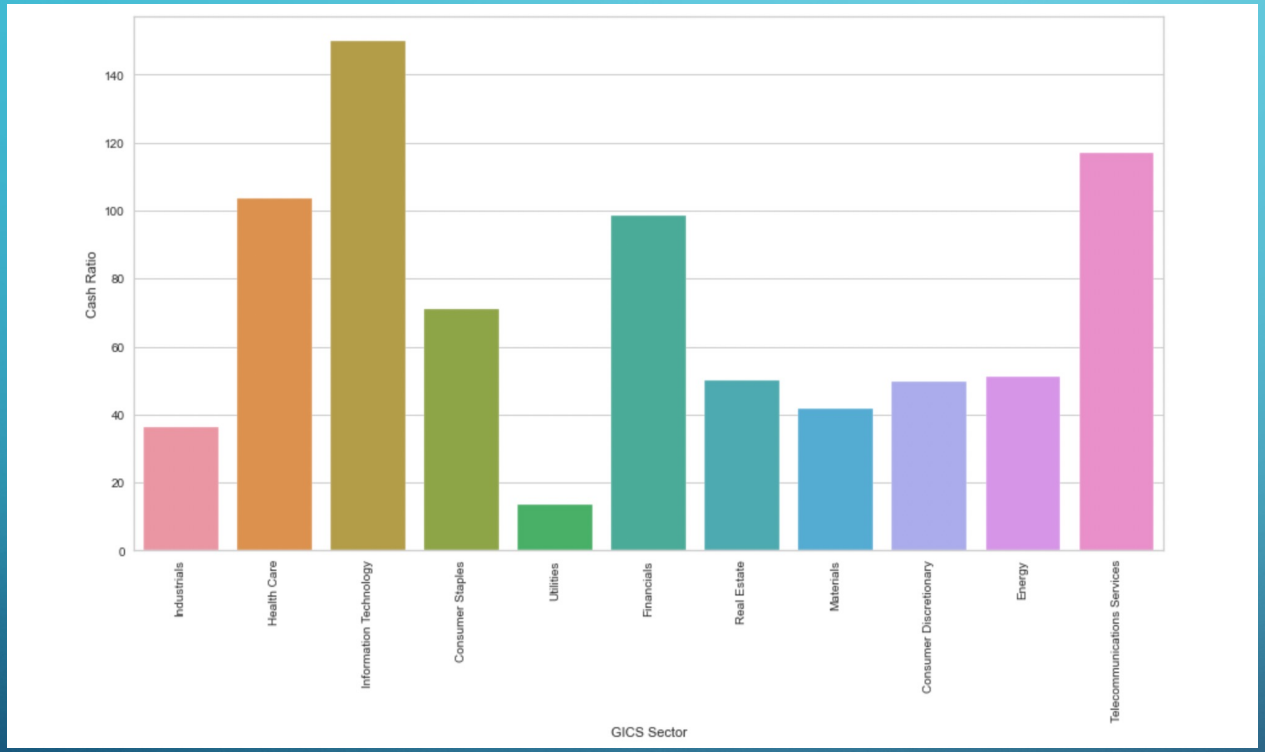

Noticeable negative correlations exist between volatility/price change, return on equity/earnings per share, and earnings per share/volatility. There are slightly strong positive correlations between net income/earnings per share, net income/estimated shares outstanding, and earnings per share/current price.

# GICS SECTOR: PRICE CHANGE BAR GRAPH



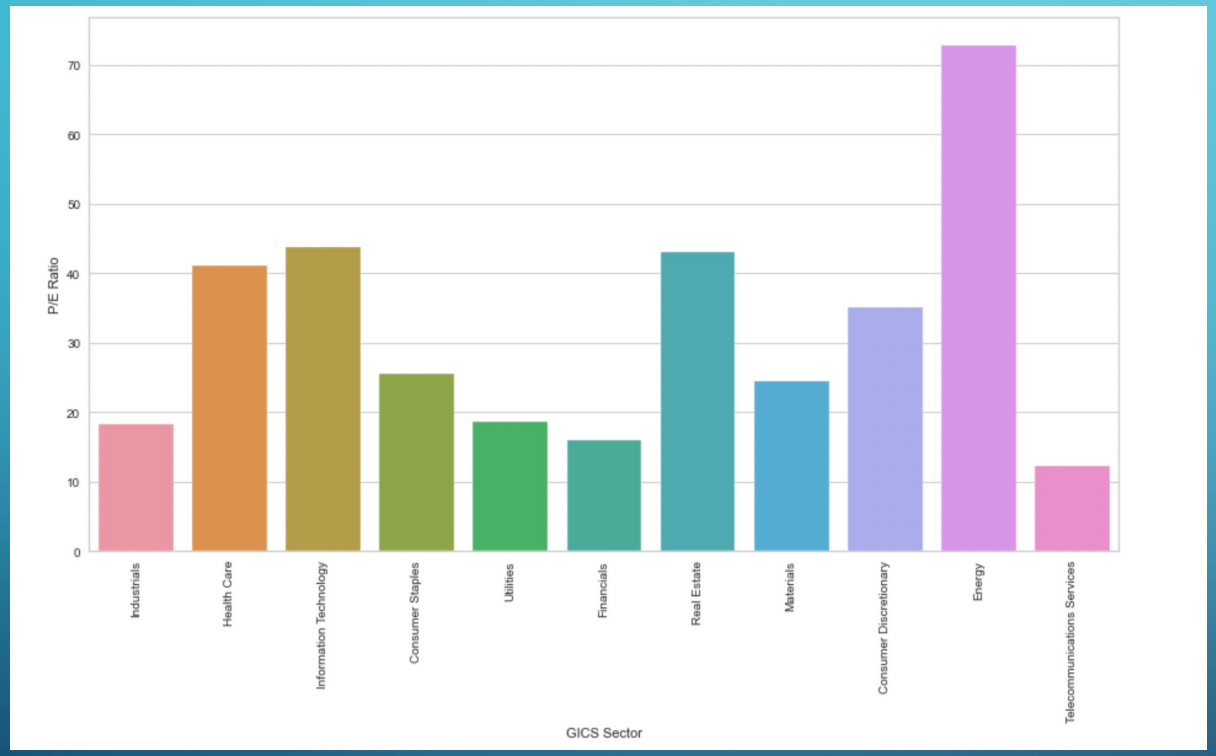

Energy sector has an average negative price change of nearly -10, while health care has the largest positive price change of approximately 9.

# GICS SECTOR: CASH RATIO



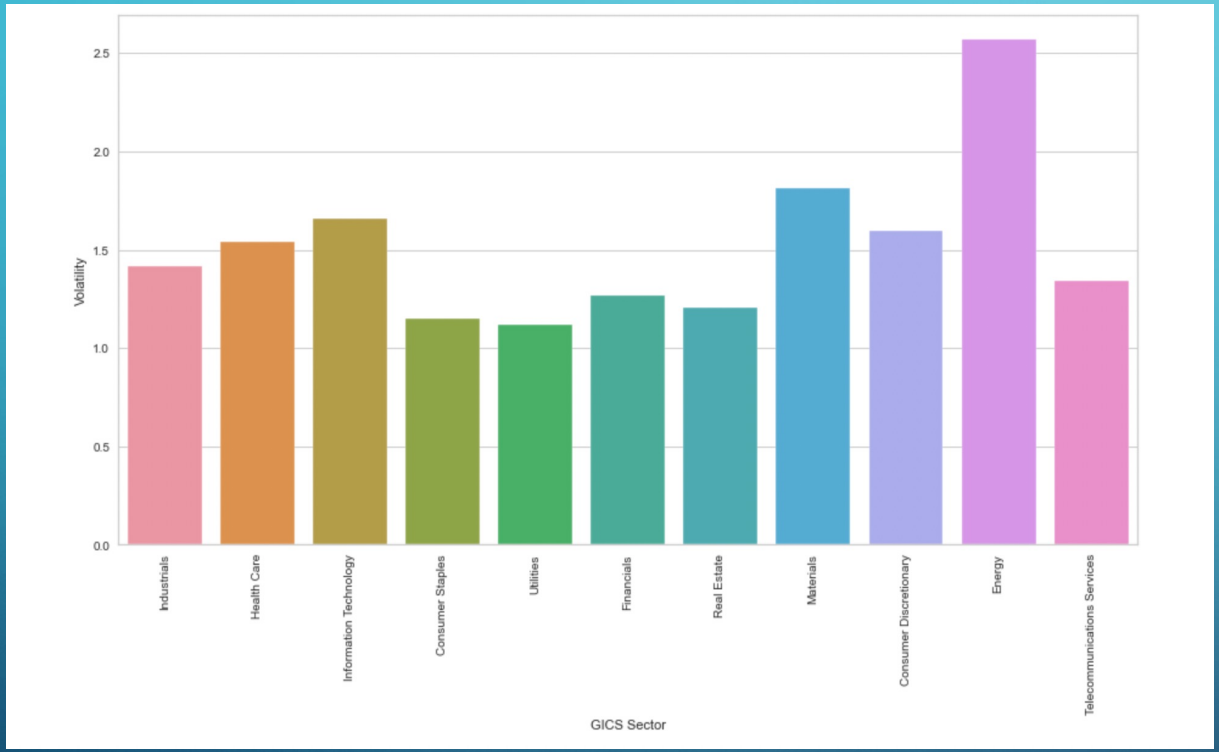

IT has highest cash ratio of over 140, while utilities has lowest cash ratio of about 10.

# GICS SECTOR: P/E RATIO



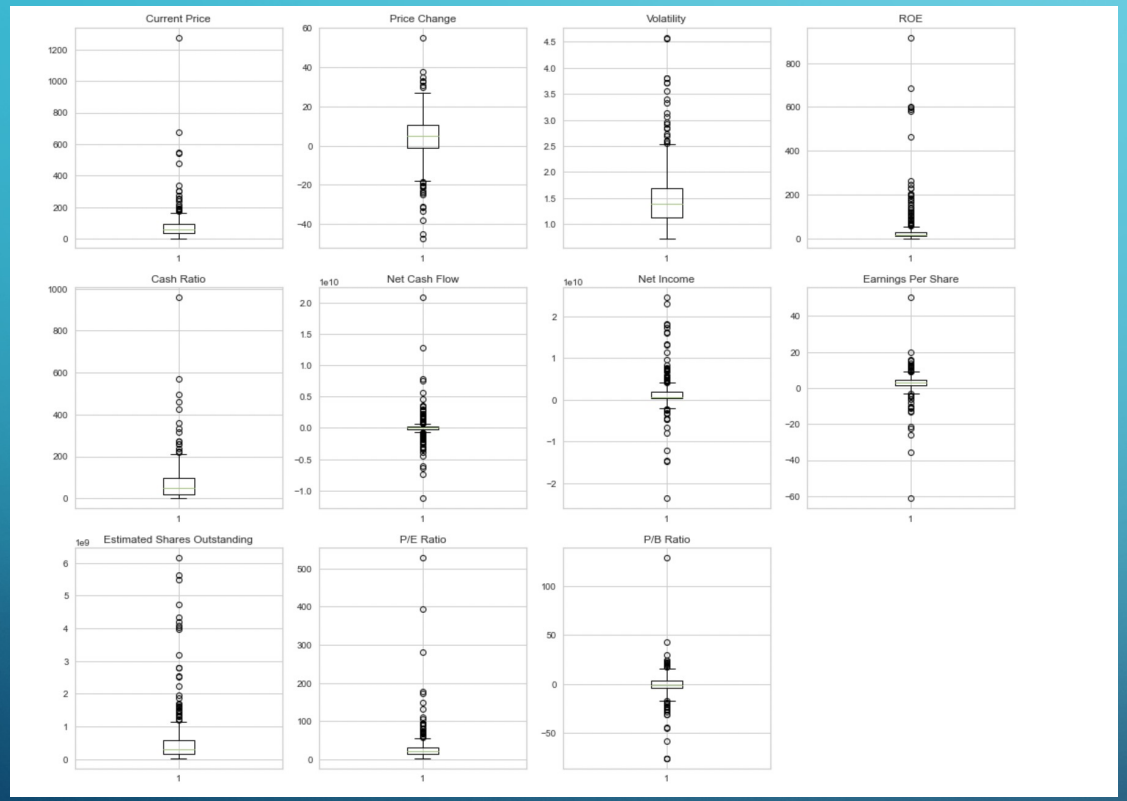

Energy has highest P/E ratio of over 70, while telecommunication services has lowest P/E ratio of a little over 10.

# GICS SECTOR: VOLATILITY




Energy has highest volatility of a little over 2.5 and utilities has lowest with just a little over 1.

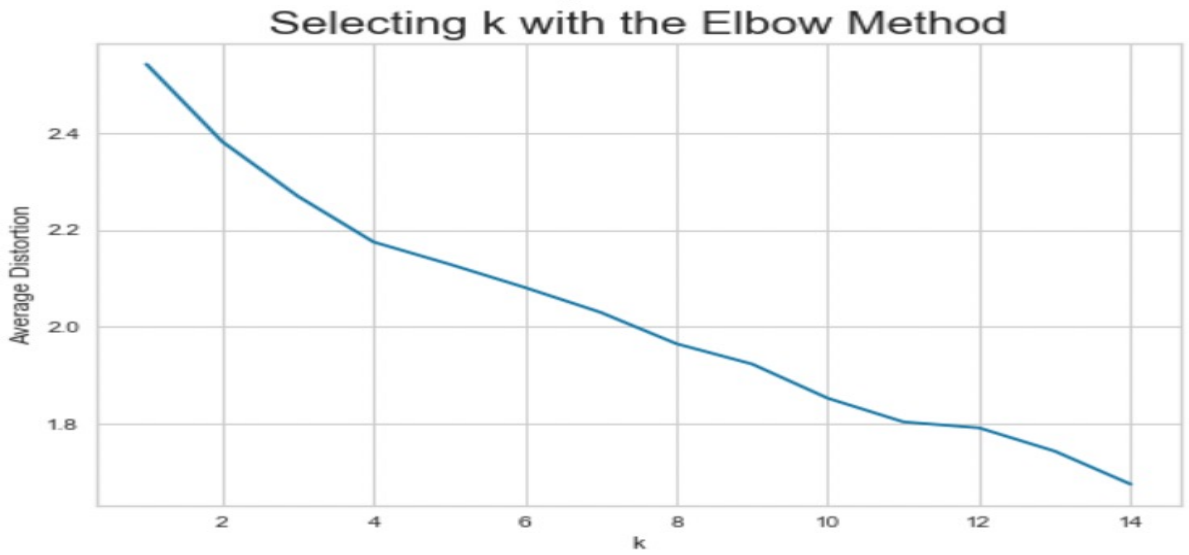
# OUTLIER CHECK



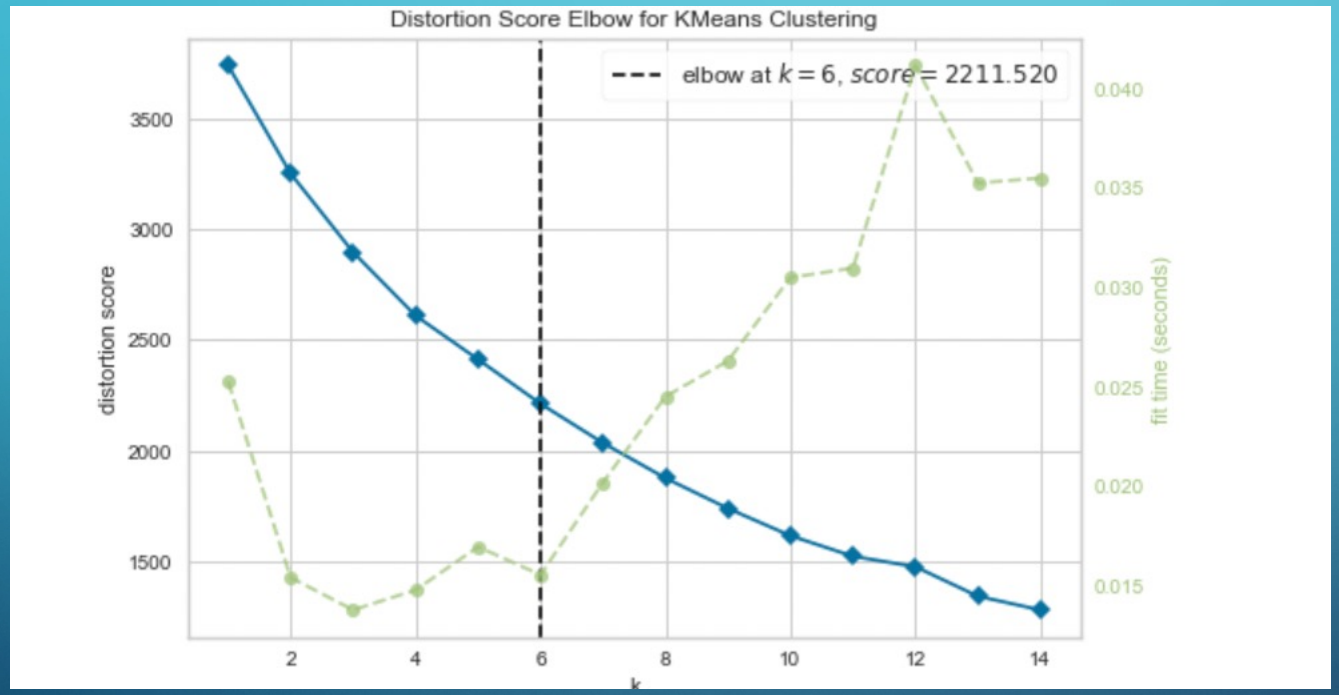

Volatility, Estimated Shares Outstanding, and ROE have many outliers to the right, while net cash flowm earnings per share, and net income have many outliers that are close to the interquartile ranges.

# K-MEANS CLUSTERING

```
Number of Clusters: 1    Average Distortion: 2.5425069919221697
Number of Clusters: 2    Average Distortion: 2.382318498894466
Number of Clusters: 3    Average Distortion: 2.2692367155390745
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
```



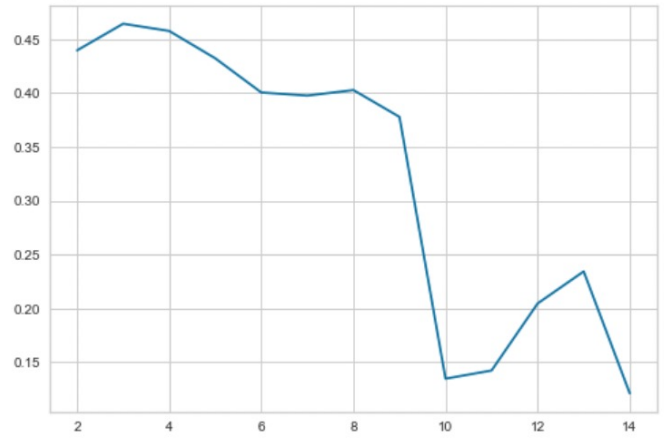
# FINAL MODEL/DISTORTION SCORE K-MEANS CLUSTERING



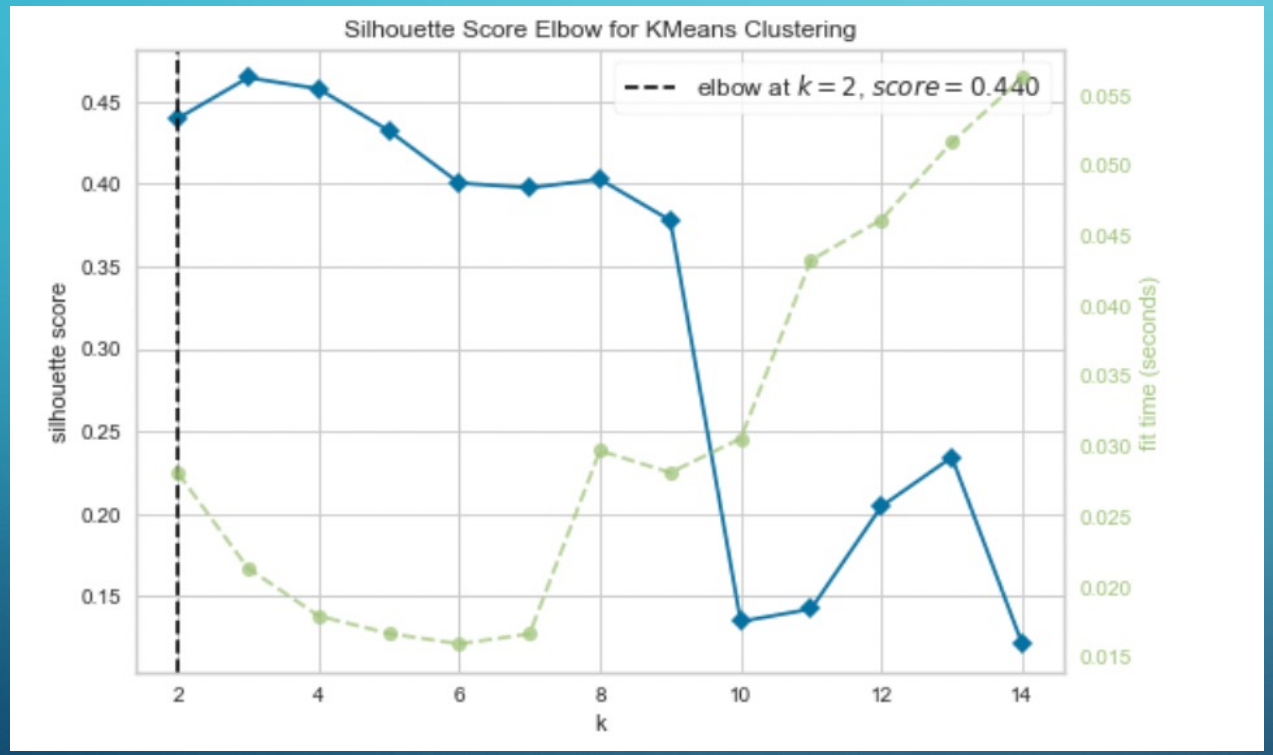

It appears when elbow at k=6, it's appropriate value to use.

# CHECKING SILHOUETTE SCORES

```
For n_clusters = 2, the silhouette score is 0.43969639509980457)
For n_clusters = 3, the silhouette score is 0.4644405674779403)
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 0.40054227372136175)
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)
```

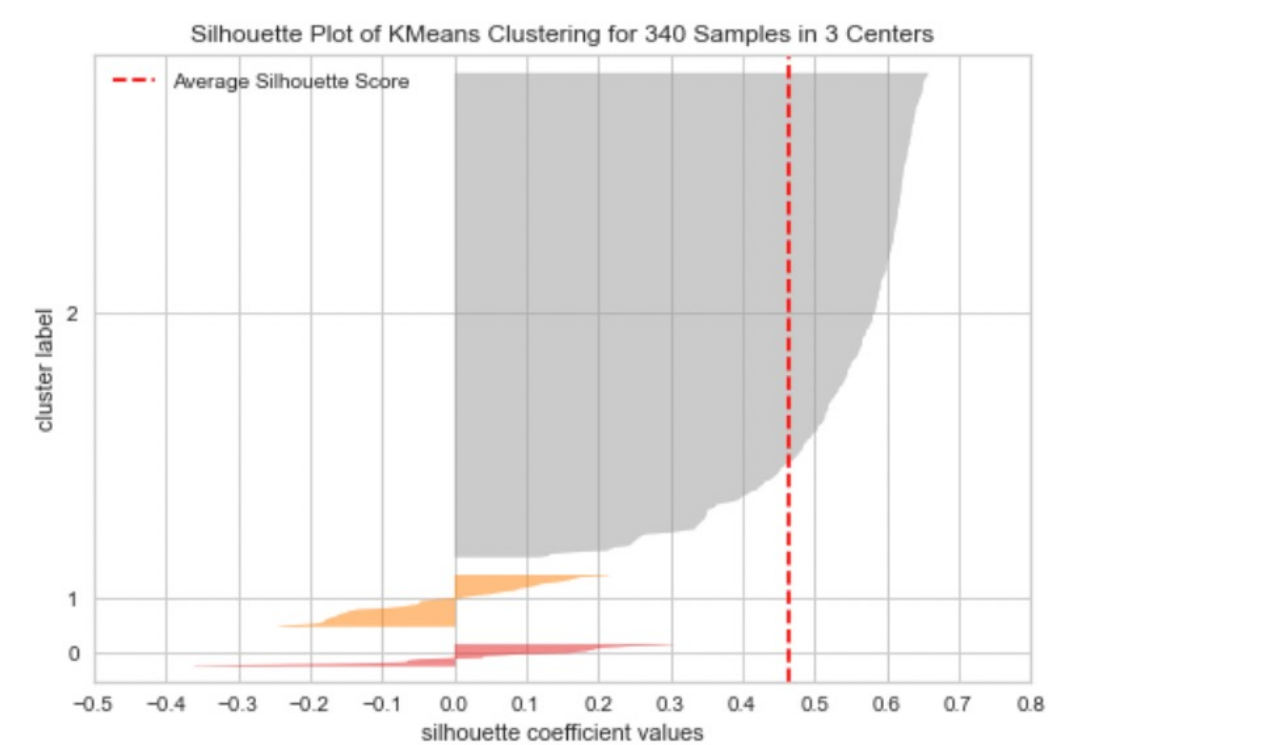


# FINAL MODEL/SILHOUETTE SCORE K-MEANS CLUSTERING




From values of silhouette scores, it seems that 2 or 3 is appropriate value for k.

# FINDING OPTIMAL NUMBER OF CLUSTERS




Choosing 3 because it provides the highest silhouette score.

```
▼ KMeans
KMeans(n_clusters=3, random_state=1)
```

# CLUSTER PROFILING

| 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 | 52.142857 | 6.779993 | 1.175153 | 26.142857 | 140.142857 | 760285714.285714 | 13368785714.285715 | 3.769286 | 3838879870.871428 | 20.654832 | -3.529270 | 14 |
| 1 | 64.183438 | -10.557046 | 2.797776 | 96.531250 | 70.718750 | 159171125.000000 | -3250005968.750000 | -7.886875 | 526459323.057500 | 111.333230 | 1.783445 | 32 |
| 2 | 84.045331 | 5.542488 | 1.404255 | 34.040816 | 66.608844 | 10698350.340136 | 1445333183.673469 | 3.890051 | 427206184.715408 | 24.613743 | -2.013147 | 294 |

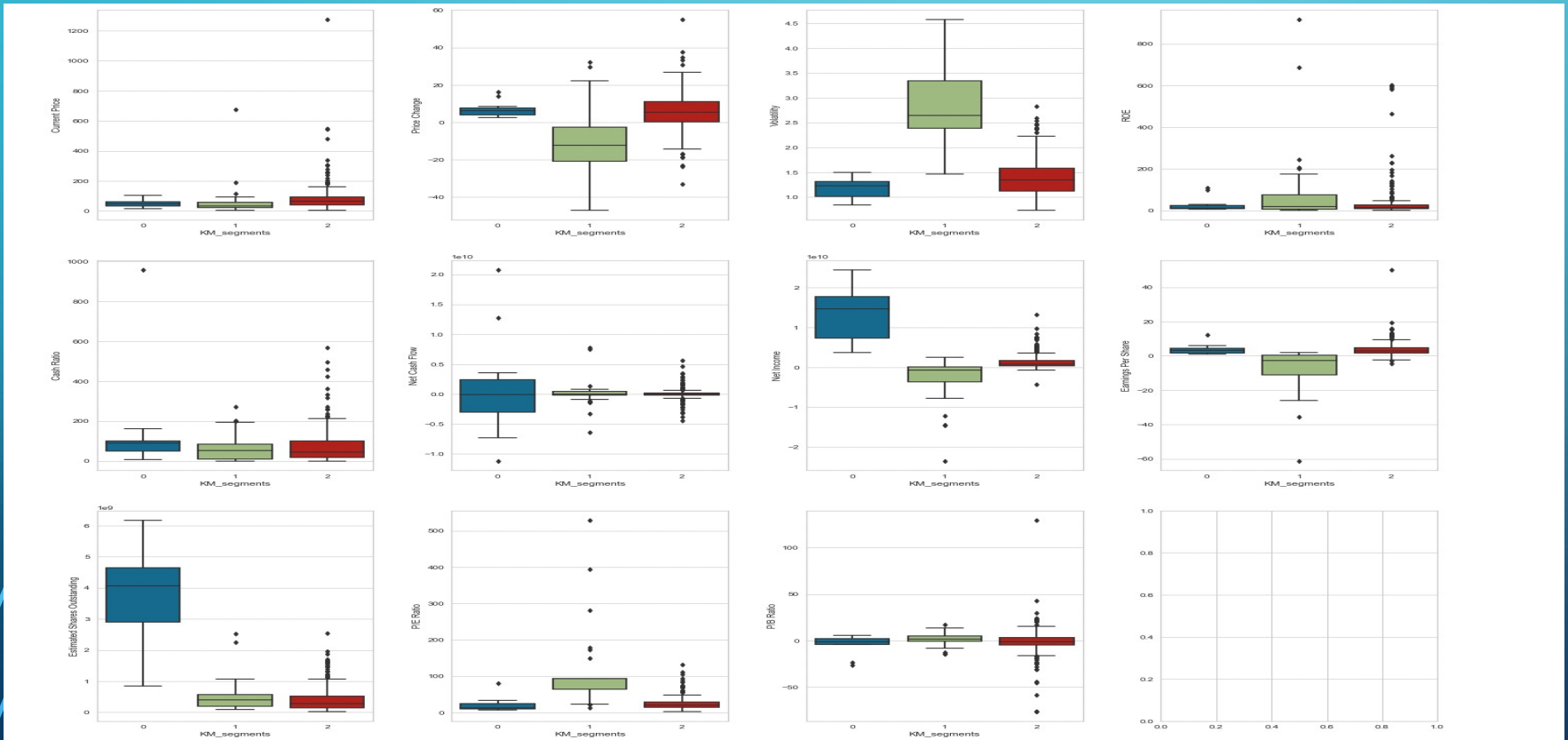
# CLUSTER ORGANIZATION: K-MEANS CLUSTERING

| KM_segments | GICS Sector | |
|---|---|---|
| 0 | Consumer Discretionary | 1 |
| | Consumer Staples | 1 |
| | Energy | 1 |
| | Financials | 4 |
| | Health Care | 3 |
| | Information Technology | 2 |
| | Telecommunications Services | 2 |
| 1 | Consumer Discretionary | 2 |
| | Energy | 23 |
| | Health Care | 1 |
| | Industrials | 1 |
| | Information Technology | 4 |
| | Materials | 1 |
| 2 | Consumer Discretionary | 37 |
| | Consumer Staples | 18 |
| | Energy | 6 |
| | Financials | 45 |
| | Health Care | 36 |
| | Industrials | 52 |
| | Information Technology | 27 |
| | Materials | 19 |
| | Real Estate | 27 |
| | Telecommunications Services | 3 |
| | Utilities | 24 |

Name: Security, dtype: int64

Cluster 0: More Financials/Healthcare, but very small cluster
Cluster 1: Primarily Energy
Cluster 2: Cluster with most companies with lots of Customer Discretionary, Financials, Healthcare, Industrials, etc.

# CLUSTER BOX PLOTS




Observations: As seen with Cluster 1, it has highest volatility, largest negative price change, and highest P/E Ratio, while Cluster 2 has many outliers with large cash ratio, but Cluster 0 has highest net income and estimated shares outstanding.

# HIERARCHICAL CLUSTERING: COPHENETIC CORRELATION

```
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 0.7925307202850004.
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.
```

As one can see, highest cophenetic correlation is 0.942 with Euclidean distance and average linkage.

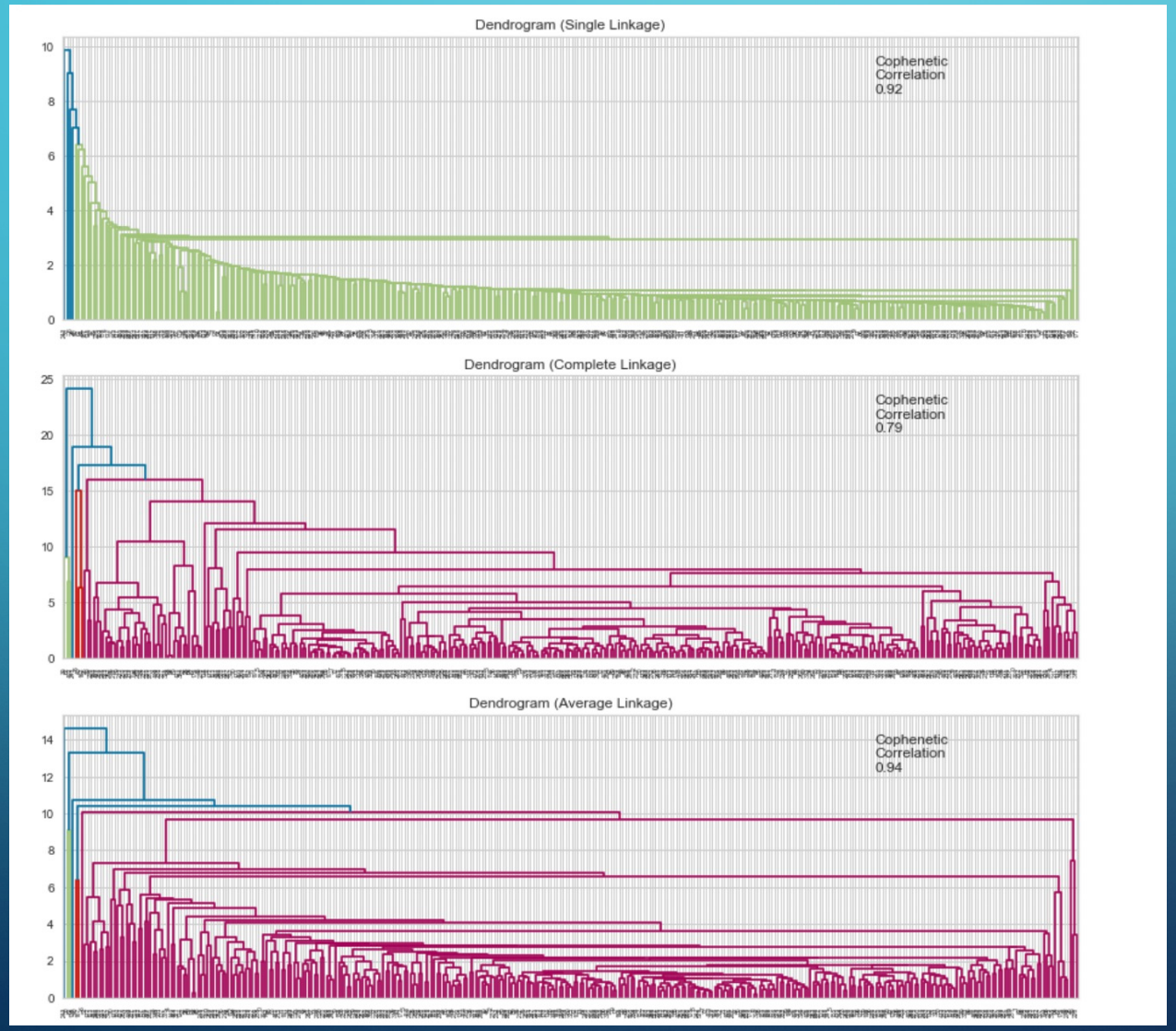
# COPHENETIC CORRELATION BY LINKAGE

```
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.
```

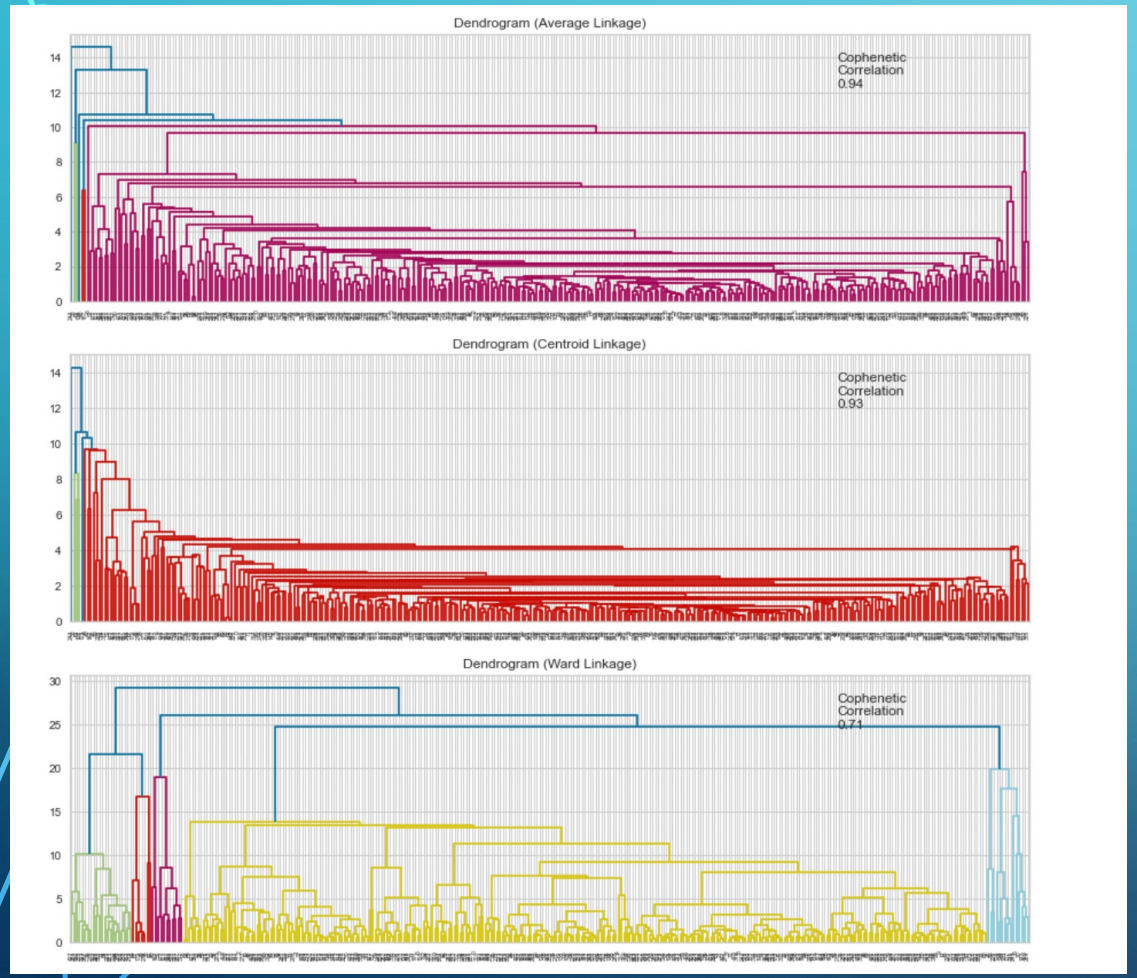
Average Linkage has highest cophenetic correlation and ward linkage has lowest cophenetic correlation.

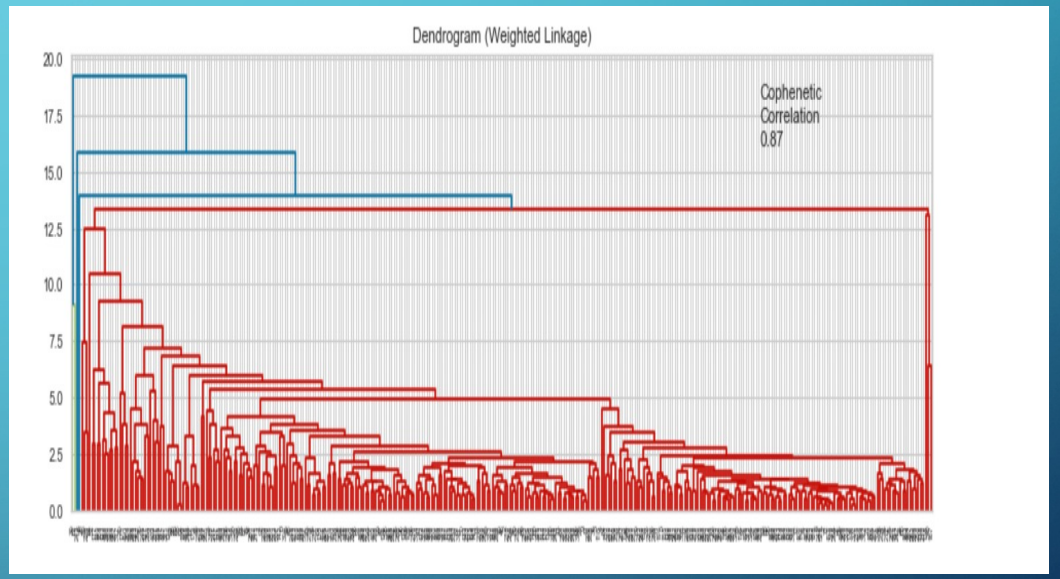
| | Linkage | Cophenetic Coefficient |
|---|---|---|
| **4** | ward | 0.710118 |
| **1** | complete | 0.787328 |
| **5** | weighted | 0.869378 |
| **0** | single | 0.923227 |
| **3** | centroid | 0.931401 |
| **2** | average | 0.942254 |

# DENDROGRAMS



# DENDROGRAMS CONTINUED.




Dendrogram for ward shows distinct and separate clusters, while average has predominant cluster.

# CLUSTER PROFILING

- Using 6 Clusters

```
AgglomerativeClustering
AgglomerativeClustering(n_clusters=6)
```

| 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 | 563.992491 | 17.235667 | 1.839399 | 10.250000 | 125.500000 | 105834000.000000 | 853500250.000000 | 13.085000 | 287806305.492500 | 307.105187 | -4.254889 | 4 |
| 1 | 84.355716 | 3.854981 | 1.827670 | 633.571429 | 33.571429 | -568400000.000000 | -4968157142.857142 | -10.841429 | 398169036.442857 | 42.284541 | -11.589502 | 7 |
| 2 | 152.566666 | 14.908086 | 1.769506 | 24.434783 | 281.913043 | 1747221304.347826 | 1866621956.521739 | 3.802174 | 759756952.867391 | 38.674023 | 16.027369 | 23 |
| 3 | 72.421687 | 4.563230 | 1.403434 | 25.218182 | 55.014545 | 72801872.727273 | 1572467469.090909 | 3.728564 | 445003946.148763 | 24.188244 | -2.966949 | 275 |
| 4 | 36.440455 | -16.073408 | 2.832884 | 57.500000 | 42.409091 | -472834090.909091 | -3161045227.272727 | -8.005000 | 514367806.201818 | 85.555682 | 0.836839 | 22 |
| 5 | 46.672222 | 5.166566 | 1.079367 | 25.000000 | 58.333333 | -3040666666.666667 | 14848444444.444445 | 3.435556 | 4564959946.222222 | 15.596051 | -6.354193 | 9 |

# CLUSTER ORGANIZATION: HIERARCHICAL CLUSTERING THROUGH EUCLIDEAN DISTANCE AND WARD LINKAGE

| HC_segments | GICS Sector | |
|---|---|---|
| 0 | Consumer Discretionary | 2 |
| | Health Care | 1 |
| | Information Technology | 1 |
| 1 | Consumer Discretionary | 1 |
| | Consumer Staples | 2 |
| | Energy | 2 |
| | Financials | 1 |
| | Industrials | 1 |
| 2 | Consumer Discretionary | 3 |
| | Consumer Staples | 1 |
| | Financials | 1 |
| | Health Care | 7 |
| | Information Technology | 8 |
| | Materials | 1 |
| | Real Estate | 1 |
| | Telecommunications Services | 1 |
| 3 | Consumer Discretionary | 33 |
| | Consumer Staples | 15 |
| | Energy | 7 |
| | Financials | 44 |
| | Health Care | 31 |
| | Industrials | 52 |
| | Information Technology | 23 |
| | Materials | 18 |
| | Real Estate | 26 |
| | Telecommunications Services | 2 |
| | Utilities | 24 |
| 4 | Energy | 20 |
| | Information Technology | 1 |
| | Materials | 1 |
| 5 | Consumer Discretionary | 1 |
| | Consumer Staples | 1 |
| | Energy | 1 |
| | Financials | 3 |
| | Health Care | 1 |
| | Telecommunications Services | 2 |

Name: Security, dtype: int64

Choosing ward over average since ward provides separate and distinct clusters with more variability.

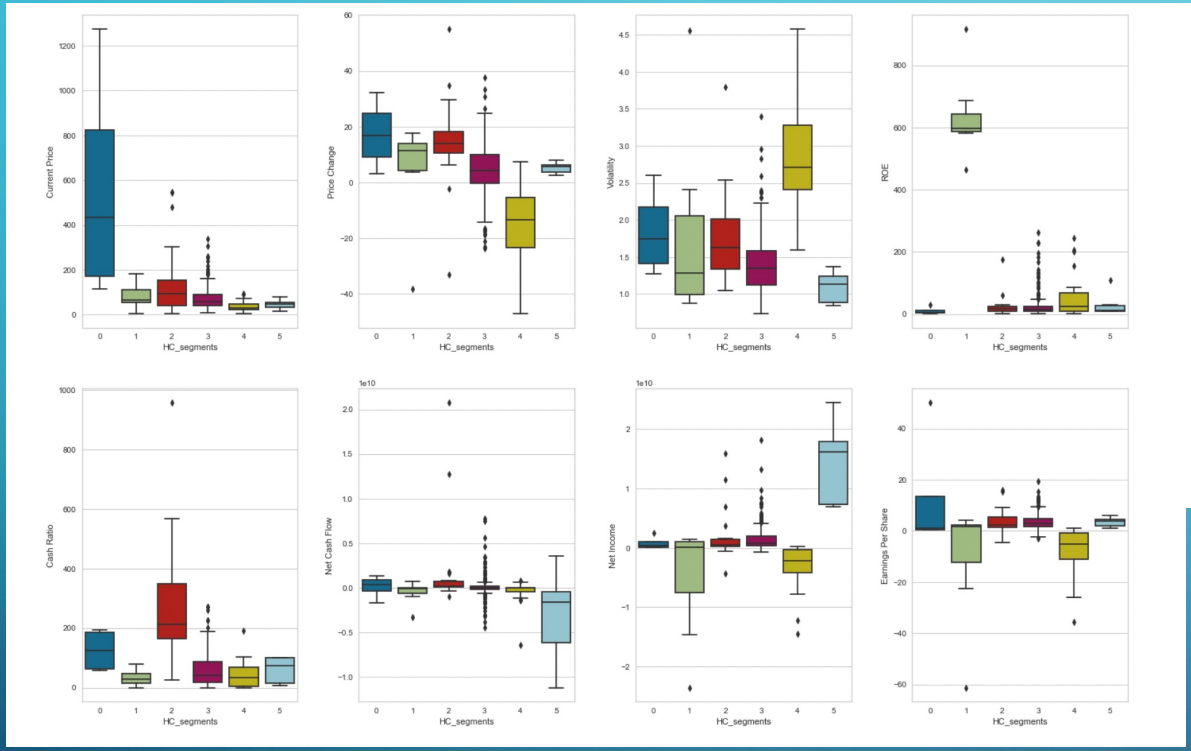
Clusters 0-1 and 5: Very small clusters

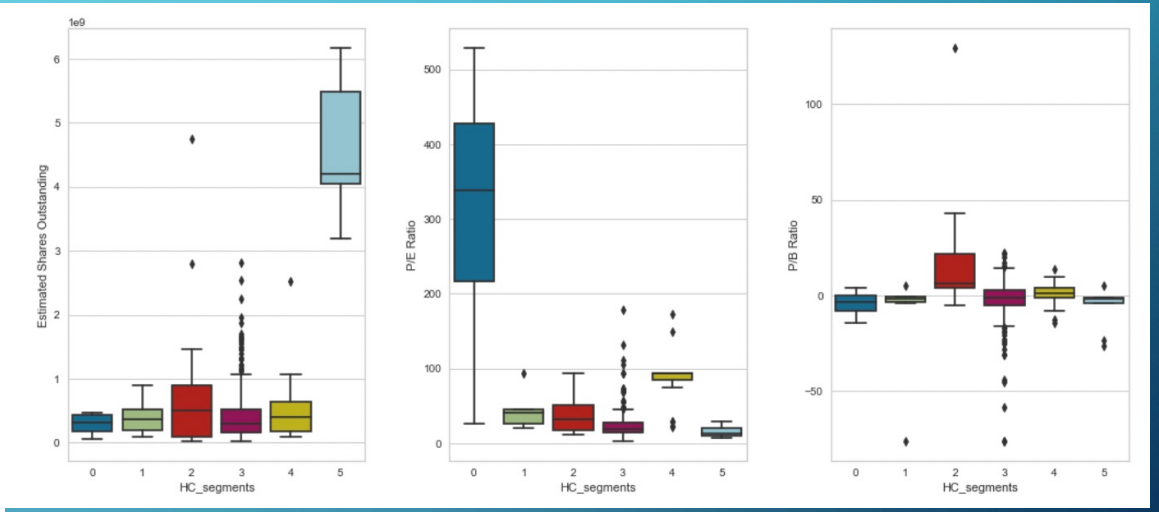
Cluster 2: Information technology companies dominate this cluster, followed closely by healthcare.

Cluster 3: Consumer Discretionary, financials, health care, industrials, information technology, materials, and real estate have a significant amount of companies in this cluster.

Cluster 4: Dominated by energy companies.

# CLUSTER BOX PLOTS




Cluster 0 has largest range of current prices, while Cluster 4 has largest range in negative price change and volatility. Cluster 1 has largest return on equity, while Cluster 2 has largest range in cash ratio. Cluster 3 has many outliers in net cash flow, while Cluster 5 has largest range in net cash flow (primarily negative), net income (primarily positive), and estimated shares outstanding. Cluster 3 has many outliers to the left in P/B ratio, while Cluster 0 has the largest range in P/E Ratio.

# INSIGHTFUL CONCLUSIONS

- Through hierarchical and K-mean clusters, I was able to find out that energy stocks have large negative price fluctuations and high volatility/P/E ratios.
- Clusters that have a significant amount of information technology companies have stocks with high cash ratios.
- K-mean clusters showed healthcare and financial stocks have high net income and estimated shares outstanding, while it wasn't necessarily the same with hierarchical clustering.
- Hierarchical clustering went more in depth with categorizing due to clear and distinct clusters through the Ward linkage.
- Earnings per share are generally less with more volatility in both cluster models, while they're more if net income is high.

# RECOMMENDATIONS

- Try to invest primarily in stocks with less volatility and low P/E ratios.
- Invest in stocks with high cash ratios such as information technology stocks.
- Keep an eye on various financial metrics such as earnings per share and net income as they are positively correlated. Stocks with positive values of these are ones to invest in.