Business Presentation

Contents

In this presentation, we are going to get deep into the stock market and investing in stocks.

It is important to maintain a diversified portfolio when investing in stocks in order to maximize earnings under any market condition. Having a diversified portfolio tends to yield higher returns and face lower risk by tempering potential losses when the market is down. It is often easy to get lost in a sea of financial metrics to analyze while determining the worth of a stock and doing the same for a multitude of stocks to identify the right picks for an individual can be a tedious task. By doing a cluster analysis, one can identify stocks that exhibit similar characteristics and ones that exhibit minimum correlation. This will help investors better analyze stocks across different market segments and help protect against risks that could make the portfolio vulnerable to losses.

Business Problem Overview and Solution Approach

Trade & Ahead is a financial consultancy firm who provide their customers with personalized investment strategies. They have provided us with data comprising stock price and some financial indicators for a few companies listed under the New York Stock Exchange. They have asked us to analyzing the data, grouping the stocks based on the attributes provided, and sharing insights about the characteristics of each group.

Data Overview

- we have 340 rows and 15 columns

- There is no missing value.

- Also, we have outliers in our data

- 4 columns are objects 7 columns are float and others are integer

- There is no duplicated values

| 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)

- industrials and financial are on the top of The specific economic sector
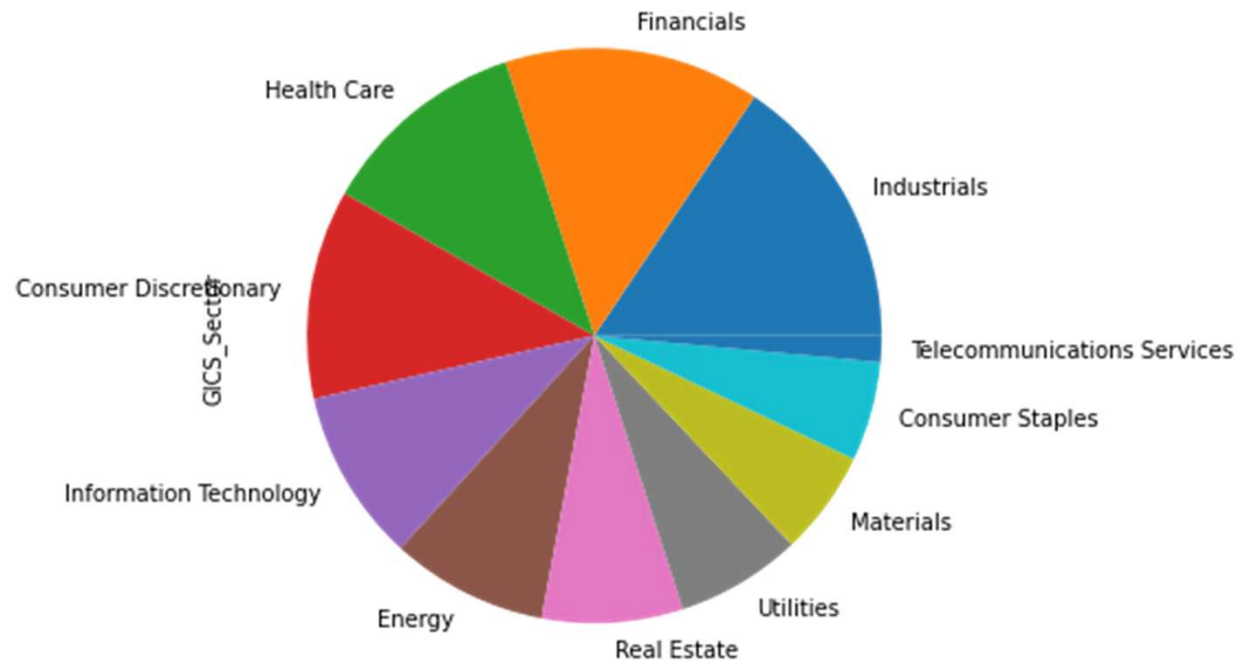
EDA

- What does the distribution of stock prices look like?

- Current stock price has a normal distribution but is highly skewed to the right. its mean is around 90$

Current stock price in dollars
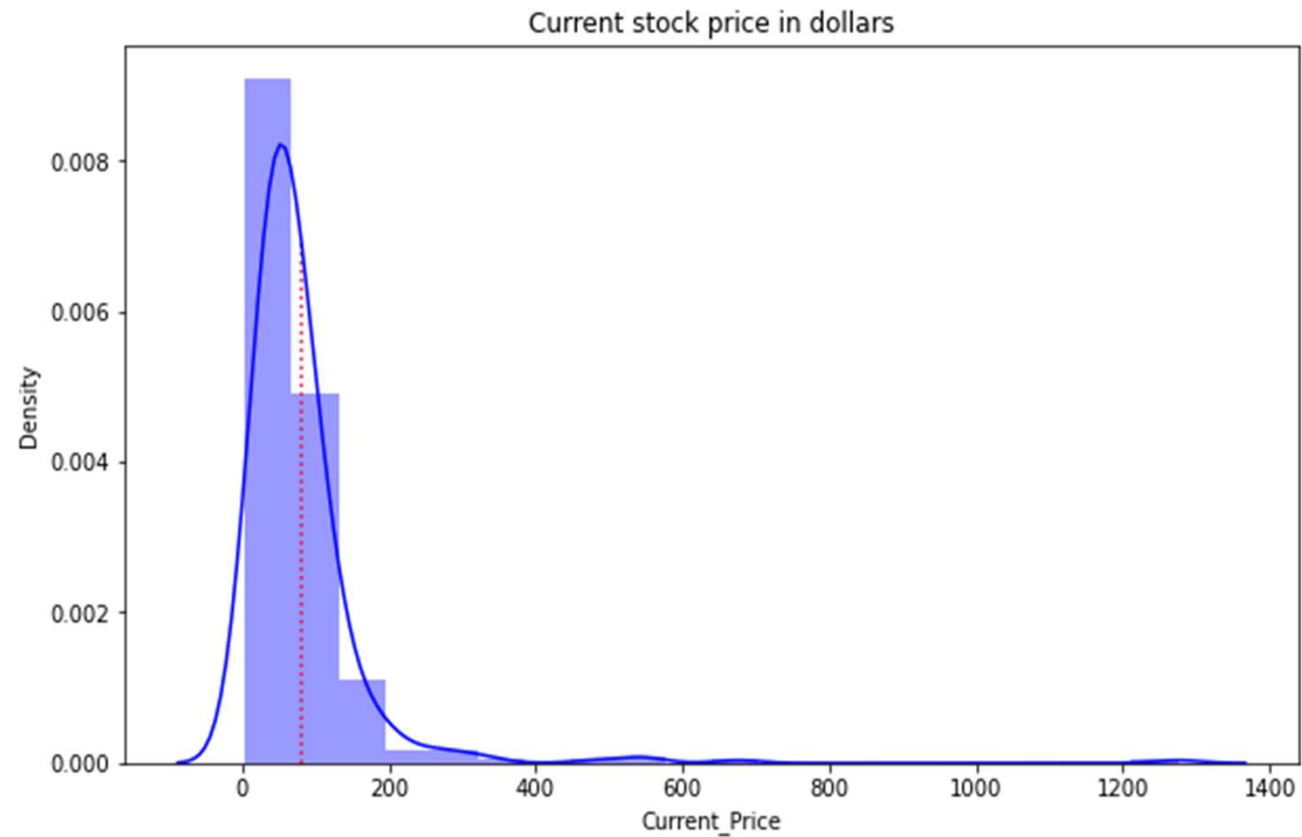
EDA

- the Percentage change in the stock price in 13 weeks has a normal distribution and mean around 5

Percentage change in the stock price in 13 weeks
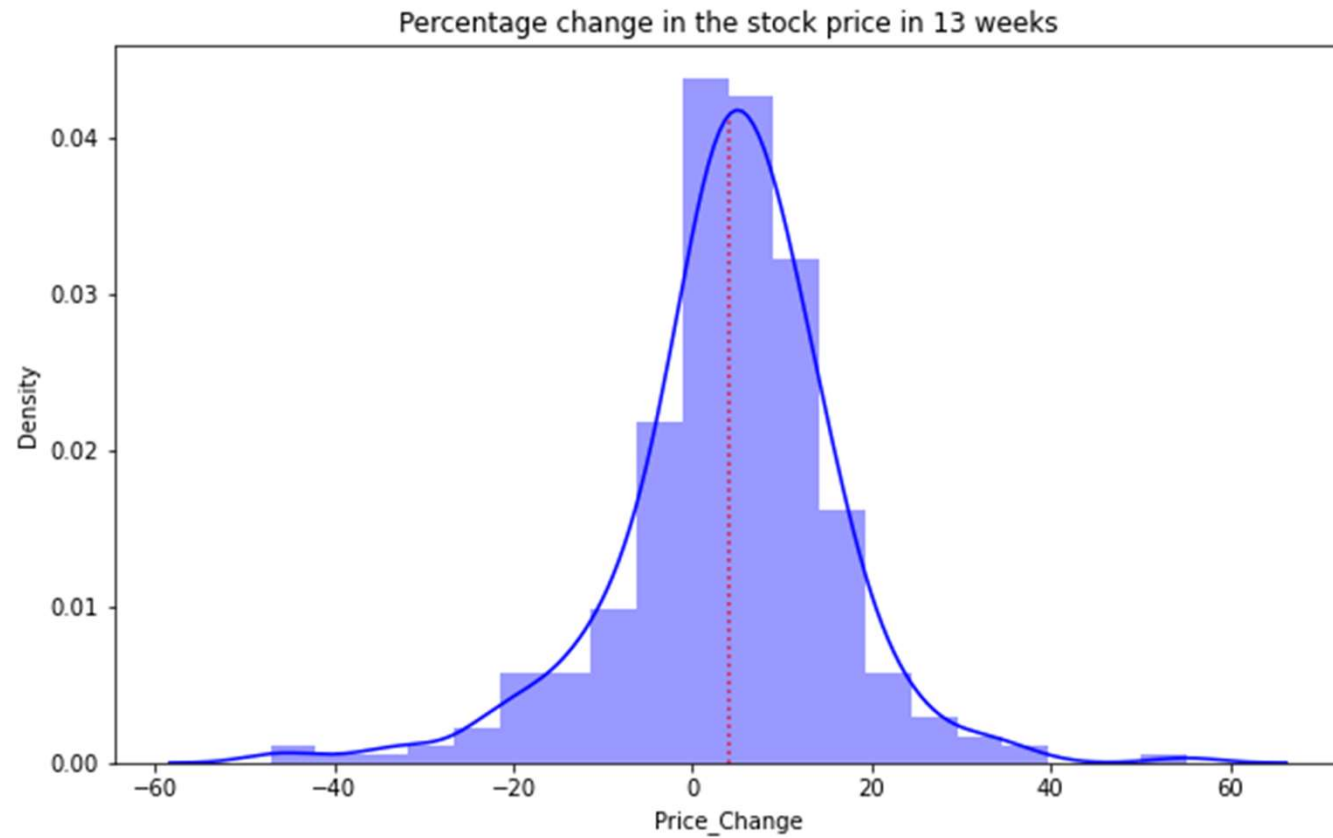
EDA

- Standard deviation of the stock price over the past 13 weeks has a mean of around 1.6 and is skewed to the right.

Standard deviation of the stock price over the past 13 weeks
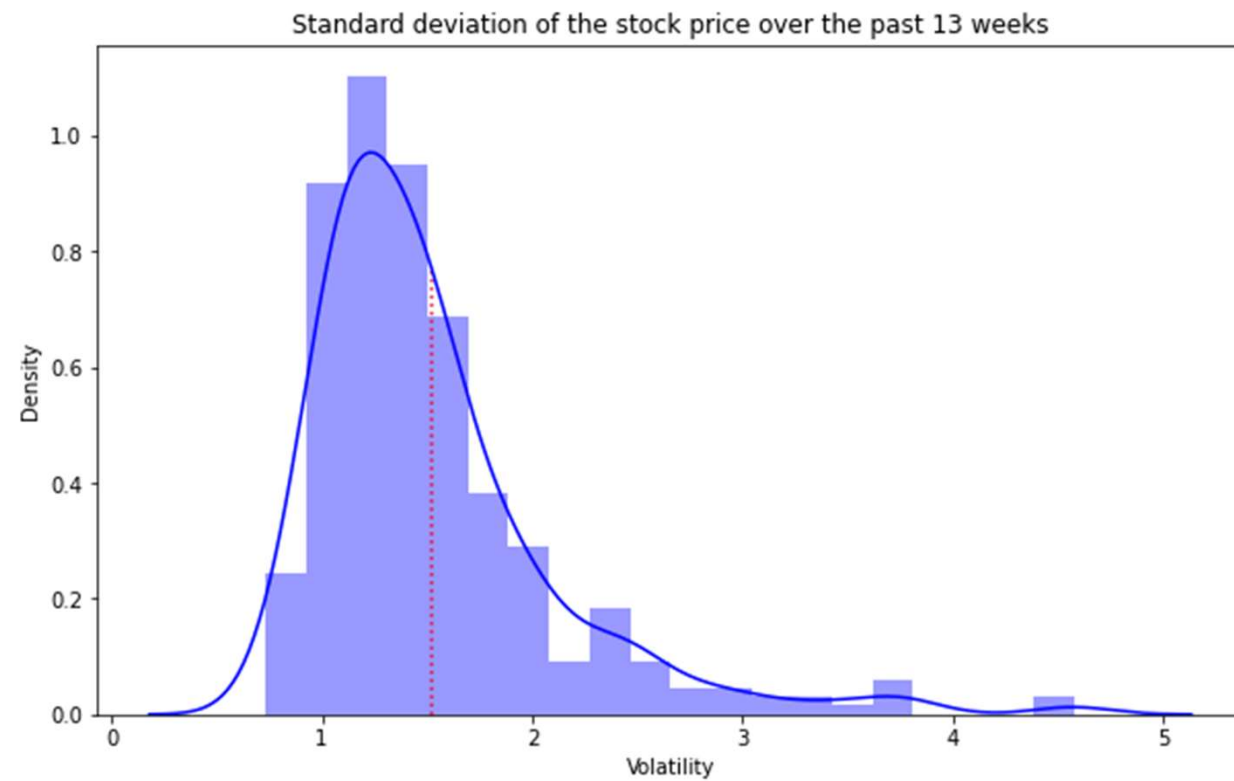
EDA

- How are the different variables correlated with each other?

it seems that some variables have a correlation. for example, between Price_Change and Volatility, there is a significant negative correlation(-0.41)
Also, Net_Income has a strong positive correlation with Estimated_Shares_Outstanding and Earnings_per_Share(0.59 & 0.56)
Earnings_per_Share has positive correlatin with Current_Price (0.48) and also has negative coorelation with REO and Volatility variables(-0.4 & -0.38).
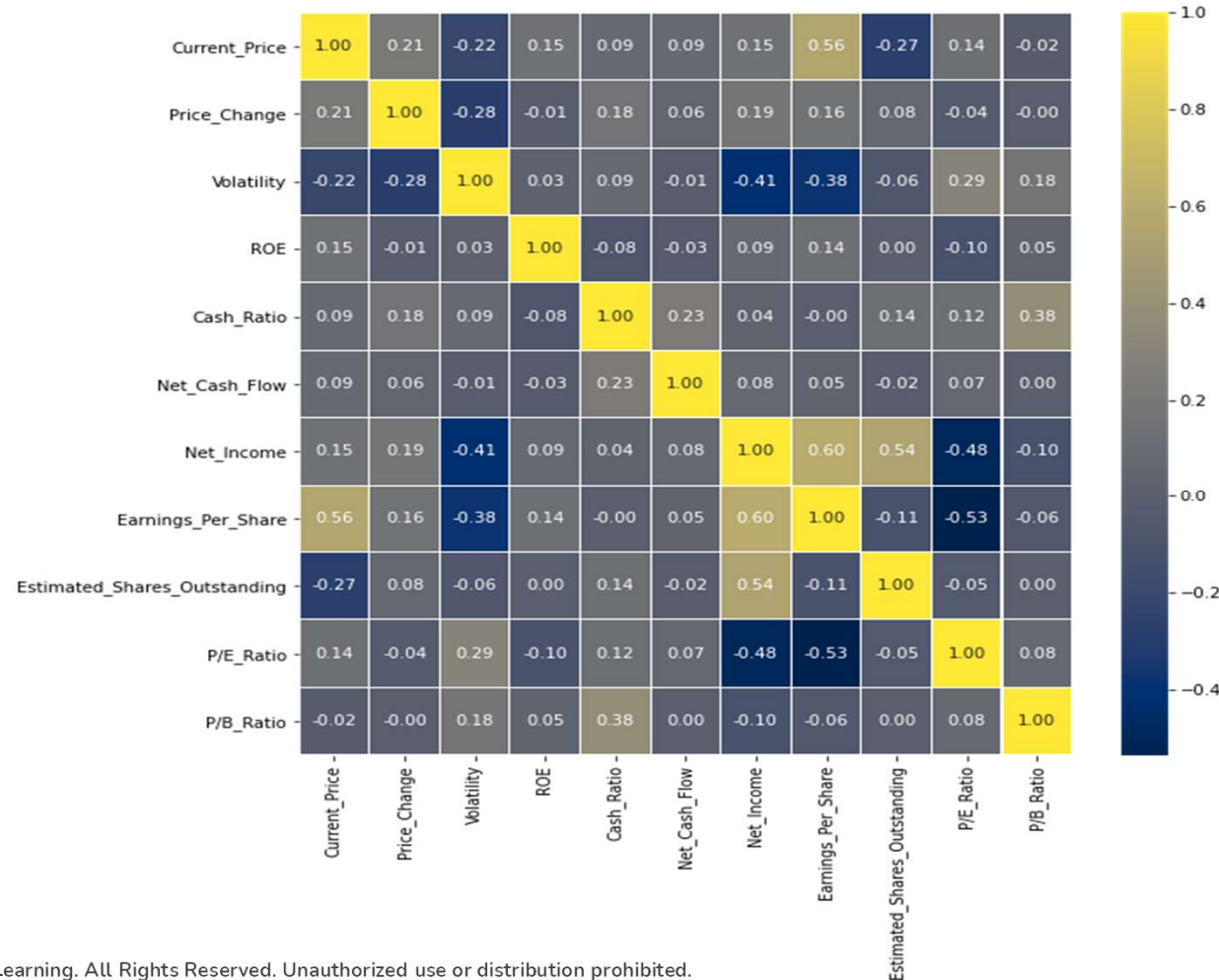
EDA

- The stocks of which economic sector have seen the maximum price increase on average?

- Health Care have seen the maximum price increase on average
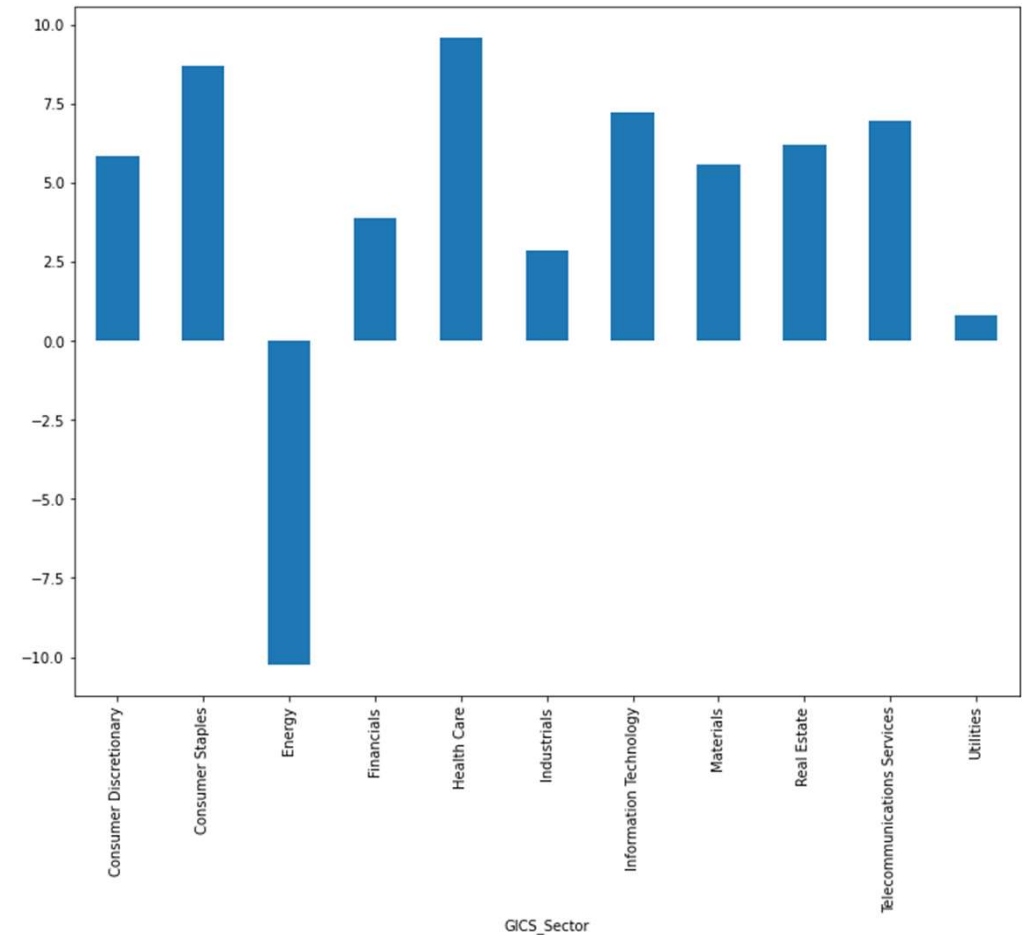
EDA

- Cash ratio provides a measure of a company's ability to cover its short-term obligations using only cash and cash equivalents. How does the average cash ratio vary across economic sectors?

- The average Cash ratio in economic sectors of Information Technology, Telecommunication Services, Healthcare, and Financials are higher than the others; also in Utilities section is the lowest.
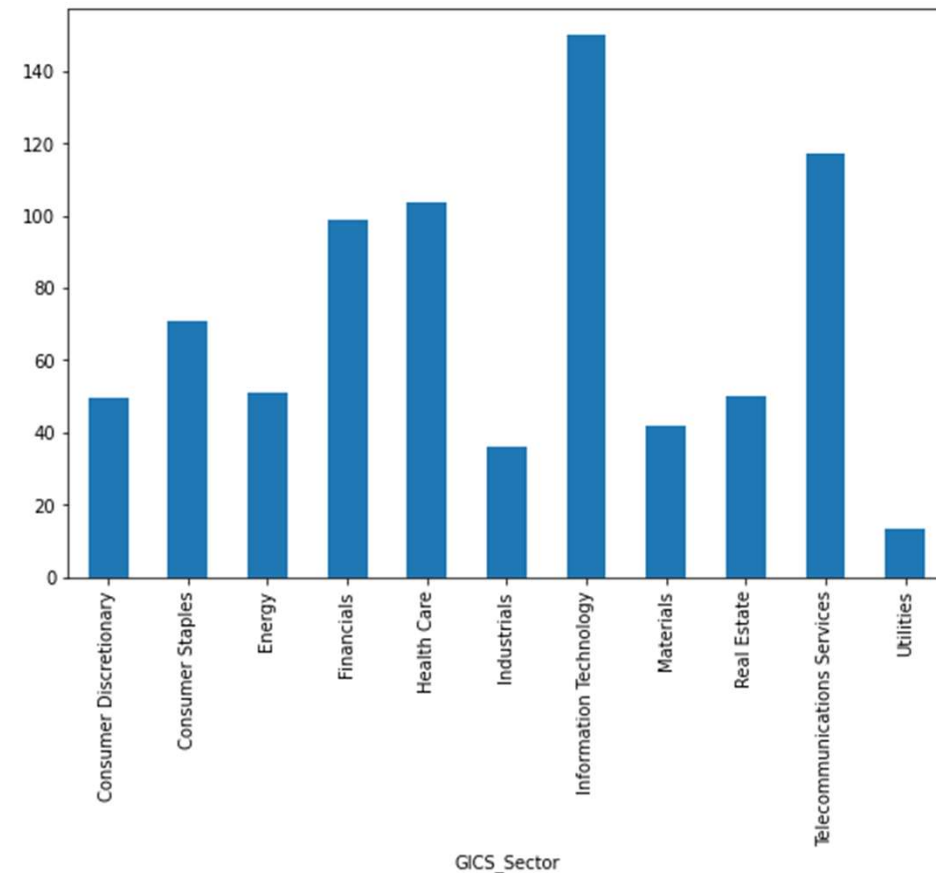
EDA

- the companies with high P/E_Retio have lower Cash_Ratio also Current stock price of Consumer Discretionary and Health care sections is higher than the others.
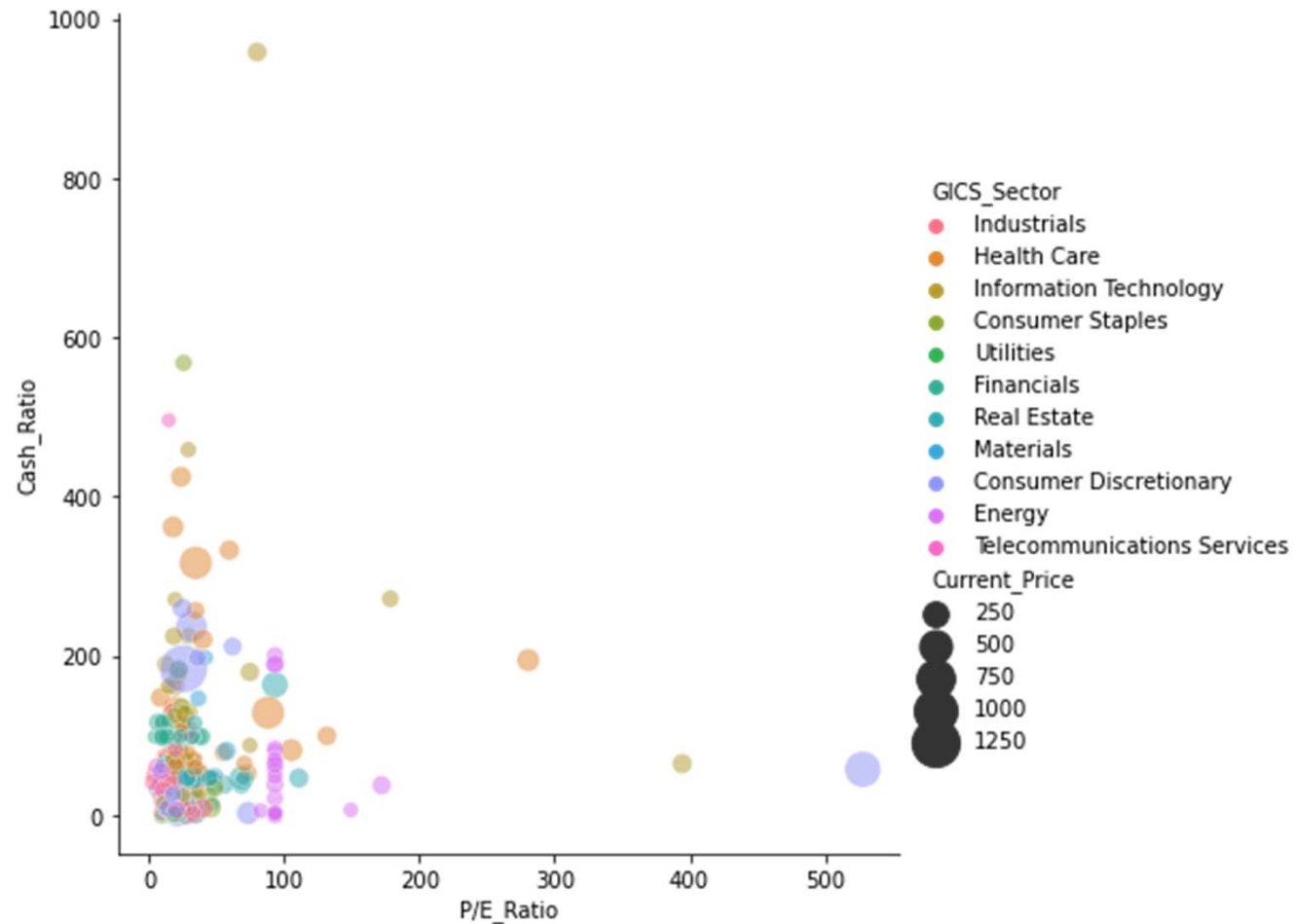
EDA

- 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. How does the P/E ratio vary, on average, across economic sectors?

- the relative value of company's shares in Energy, Information Technology, Real estate, and Healthcare are higher than the others
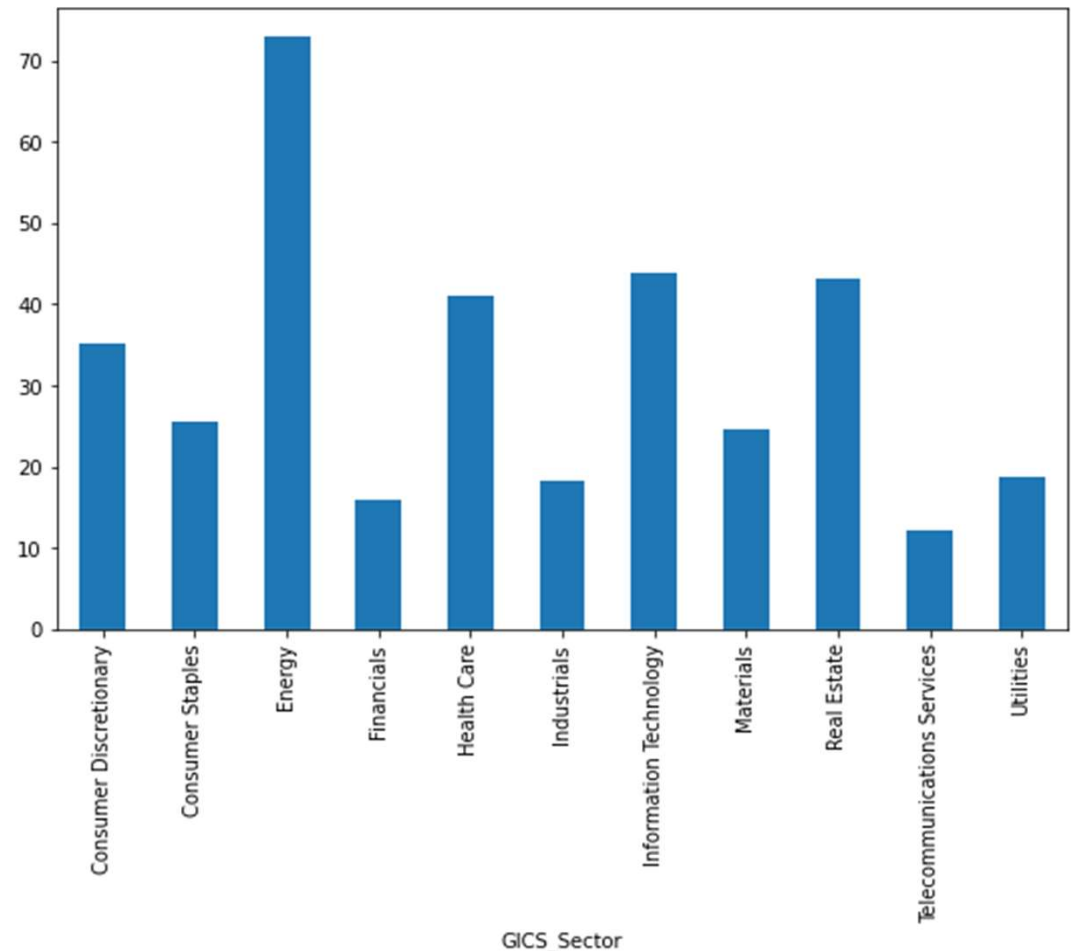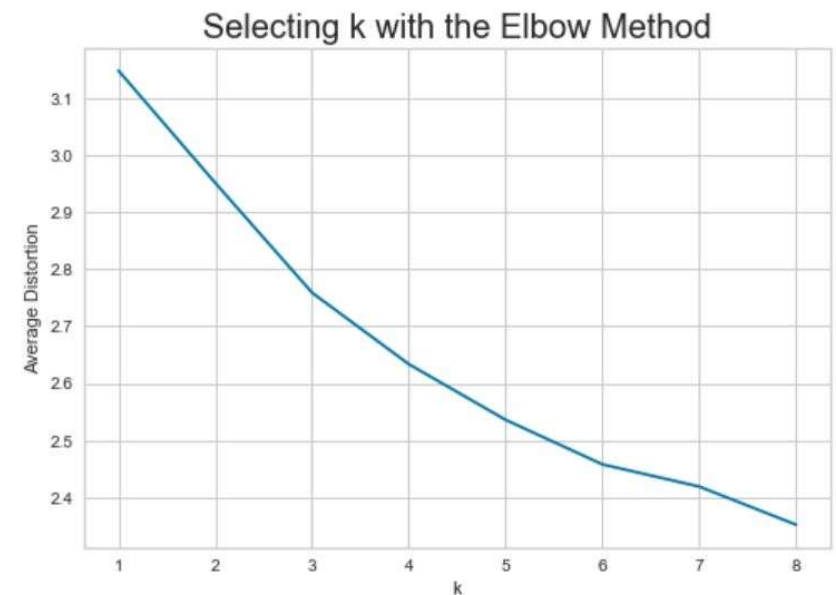
K-means Clustering

1. Apply K-means Clustering and Plot the Elbow curve:

The appropriate value of k from the elbow curve
seems to be 6

```
Number of Clusters: 1    Average Distortion: 3.1482665386211908
Number of Clusters: 2    Average Distortion: 2.95133394846628
Number of Clusters: 3    Average Distortion: 2.75878536682733137
Number of Clusters: 4    Average Distortion: 2.6333554308572875
Number of Clusters: 5    Average Distortion: 2.5357225475110132
Number of Clusters: 6    Average Distortion: 2.45759954836683995
Number of Clusters: 7    Average Distortion: 2.418702726282235
Number of Clusters: 8    Average Distortion: 2.35180072371567484
Text(0.5, 1.0, 'Selecting k with the Elbow Method')
```


Selecting k with the Elbow Method

2. Check Silhouette Scores:

From the silhouette scores, it seems that 3 is a good value of k.

```
For n_clusters = 2, silhouette score is 0.15155312647170657
For n_clusters = 3, silhouette score is 0.183981753237767
For n_clusters = 4, silhouette score is 0.17445259290694196
For n_clusters = 5, silhouette score is 0.14383334964717953
For n_clusters = 6, silhouette score is 0.16229740368621268
For n_clusters = 7, silhouette score is 0.13374980919162344
For n_clusters = 8, silhouette score is 0.127566621998718698
For n_clusters = 9, silhouette score is 0.1280216851592048

[<matplotlib.lines.Line2D at 0x19d84a511c0>]
```
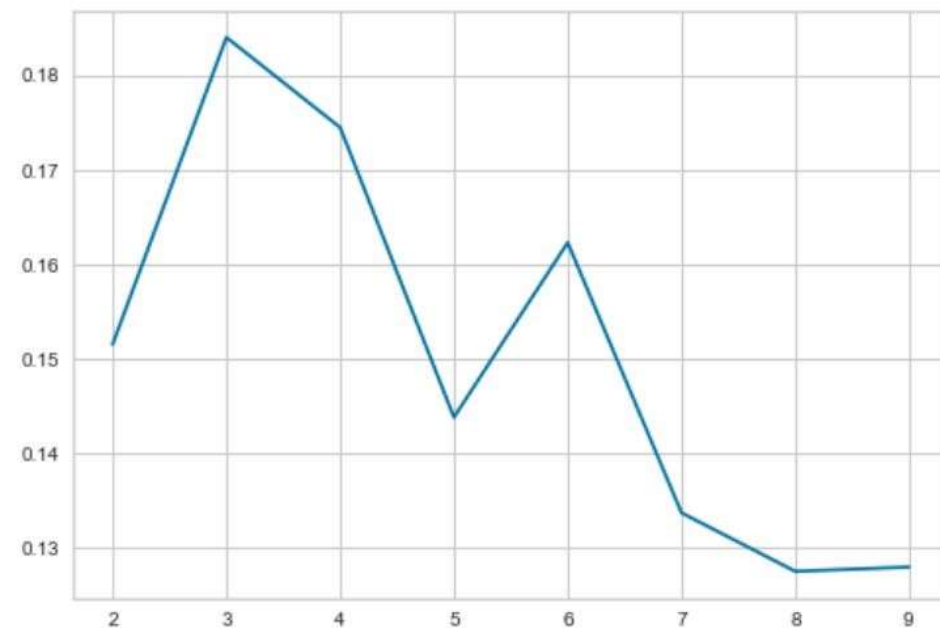
K-means Clustering

display cluster profiles

Cluster 0: There are 82 GICS_Sector in this cluster.
Current stock price is $0.04
Ratio of the company's current stock price to the earnings per share (or P/E ratio) is-0.48.
ROE or a measure of financial performance for companies of this cluster is 0.25.

Cluster 1: There are 215 GICS_Sector in this cluster.
Current stock price is $0.09
Ratio of the company's current stock price to the earnings per share (or P/E ratio) is -0.06
ROE or A measure of financial performance for companies of this cluster is 0.10

Cluster 2: There are 43 GICS_Sector in this cluster.
Current stock price is $-0.61
Ratio of the company's current stock price to the earnings per share (or P/E ratio) is 1.24
ROE or A measure of financial performance for companies of this cluster is 0.18

Cluster 2, the specific economic sector are good places to investing based on cluster profiling done.

| K_means_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	0.047570	0.323258	-0.370048	0.252727	0.213592	0.072363	1.416282	0.550664	1.259448	-0.487327	0.011401	82
1	0.105167	0.096921	-0.215759	-0.132439	-0.178794	-0.017814	-0.305083	0.052293	-0.507625	-0.062404	-0.121426	215
2	-0.616548	-1.101051	1.784468	0.180251	0.486658	-0.048923	-1.175402	-1.311568	0.136388	1.241343	0.585389	43

Hierarchical Clustering

- Looking at the best cophenetic coefficient we get is for "Average" linkage.

- But looking at dendrogram 'ward' shows the good difference between clusters.

- Let's make a dendrogram for the last 25 formed clusters using complete linkage to have a better view since the above dendrograms are very populated

	LinkageMethod	CopheneticCoefficient
0	single	0.693211
1	complete	0.614780
2	average	0.732561
3	ward	0.575331
4	median	0.608879

Hierarchical Clustering

- Let's make a dendrogram for the last 25 formed clusters using complete linkage to have a better view since the above dendrograms are very populated

- I take a maximum distance around 25 to form the different clusters as clearly visible it cuts the tallest vertical lines.
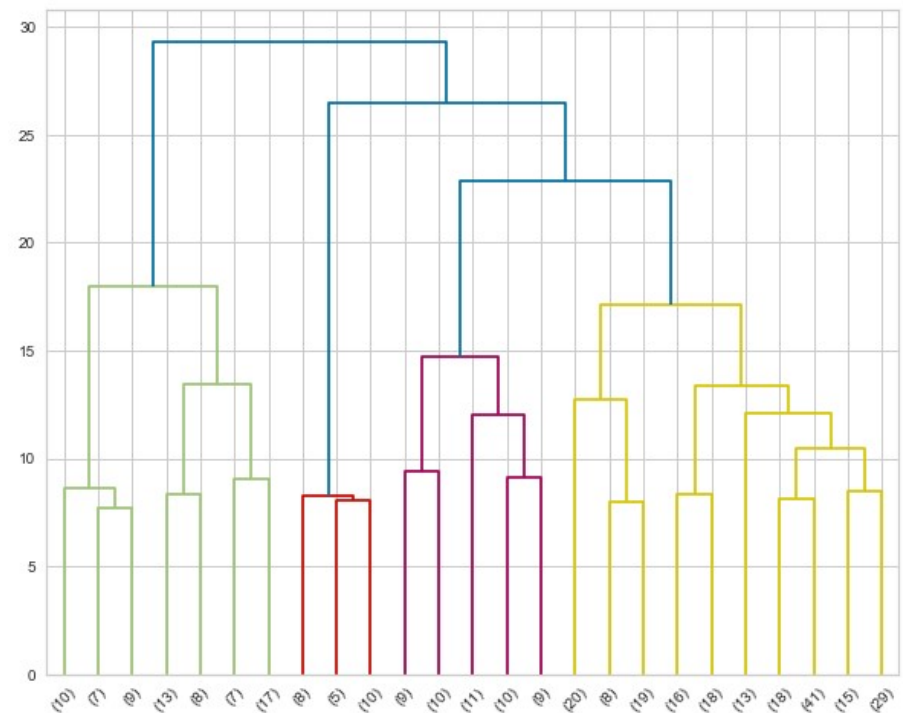
Hierarchical Clustering

display cluster profiles

Cluster 0 There are 246 GICS_Sector in this cluster.

Current stock price is $67.50

Ratio of the company's current stock price to the earnings per share (or P/E ratio) is 25.67.

ROE or A measure of financial performance for companies of this cluster is 17.32.

Cluster 1 There are 71 GICS_Sector in this cluster.

Current stock price is $92.26

Ratio of the company's current stock price to the earnings per share (or P/E ratio) is 17.24

ROE or A measure of financial performance for companies of this cluster is 28.04

Cluster 2 There are 23 GICS_Sector in this cluster.

Current stock price is $32.94

Ratio of the company's current stock price to the earnings per share (or P/E ratio) is 50.10

ROE or A measure of financial performance for companies of this cluster is 31.47

HC_Clusters	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_segments
0	67.499188	5.268043	1.449476	17.321138	63.725610	9136626.524390	804380557.926829	2.792983	343712374.399797	25.671760	-1.495496	246
1	92.264676	6.207562	1.303518	28.044014	67.274648	-9445112.676056	3387592945.422535	5.884894	756266701.881760	17.248779	-0.886916	71
2	32.947827	-11.234540	2.393770	31.478261	43.347826	-136632809.782609	-1356654576.086957	-2.242609	473814086.308804	50.100914	0.689859	23

K-means vs Hierarchical Clustering

- Cluster 2 of K-means appears like Cluster 2 of Hierarchical
- K-means silhouette score is better than Hierarchical (0.19 vs 0.20)
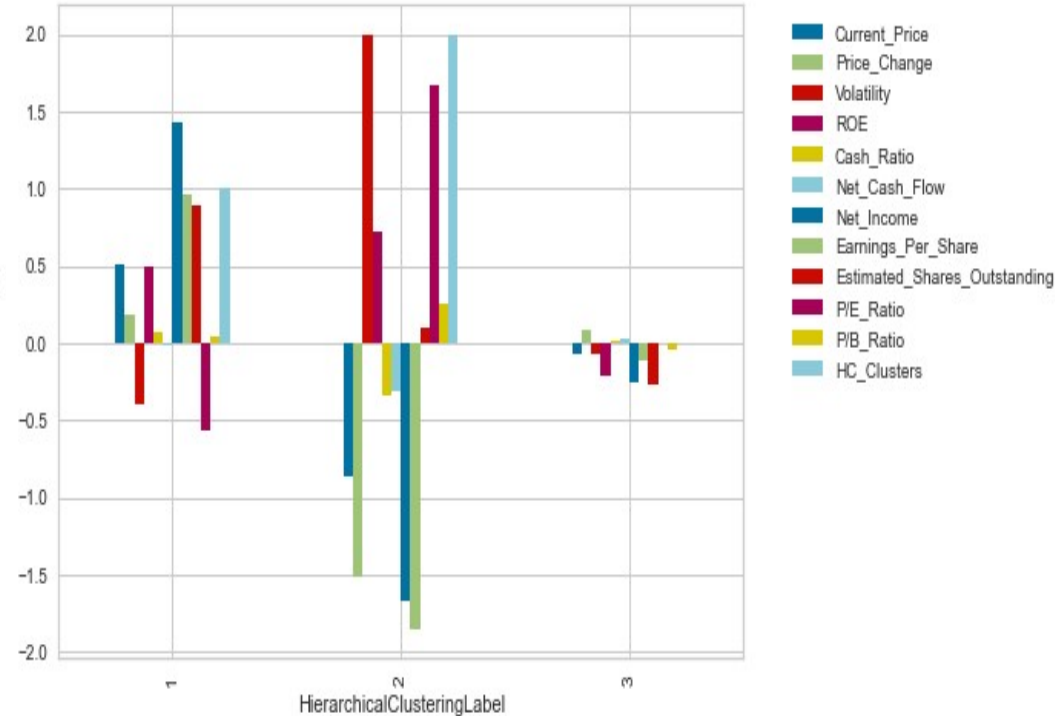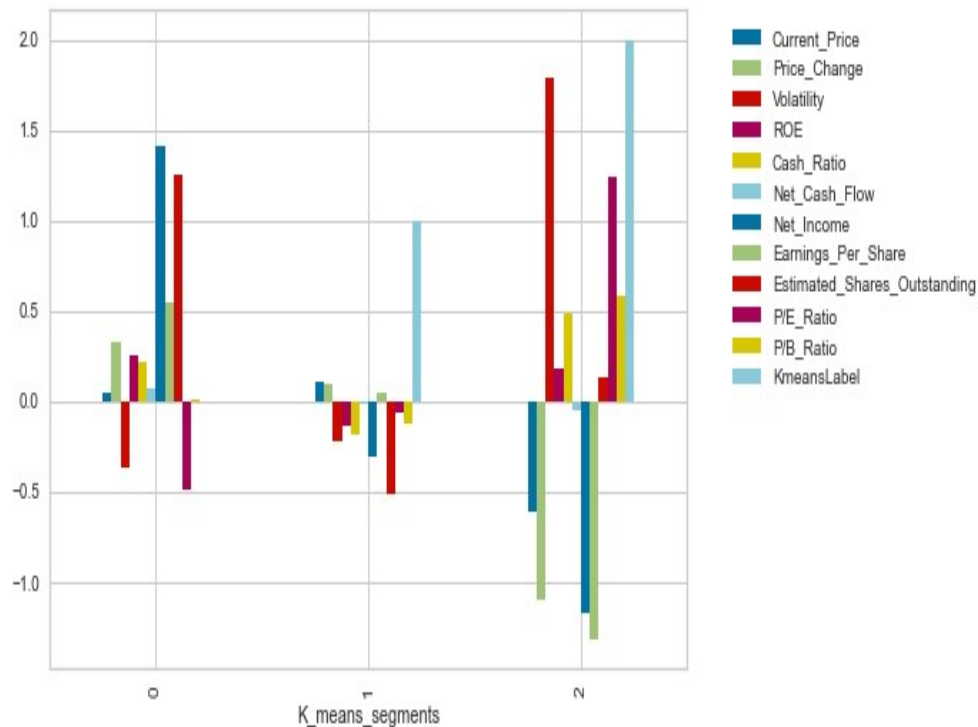
Business Insights and Recommendations

➢ for both clustering k-means and Hierarchical, cluster 2 have better result but, in k-means clustering ratios of (P/E Ratio, P/B Ratio, ROE and Cash Ratio) are better. this group is better for investment.

➢ Financial performance of cluster 3 of Hierarchical and 1 of k-means clustering is very low, These companies need to concentrate more on marketing to get up their financial performance.

➢ Performance of cluster 1 of Hierarchical and 0 of k-means are look like each other ratio of P/E is very low which means Ratio of the company's current stock price to the earnings per share are low. Also, cash ratio is very low in both so, these groups are not suggested.