

Trade&Ahead Project

Contents

- * Business Problem Overview and Solution Approach
- * Objective
- * Data Overview
- * EDA
 - Univariate Analysis
 - Bivariate Analysis
- * K-Means Clustering
- * Hierarchical Clustering
- * Business Insights and Recommendations

Business Problem Overview and Solution Approach

The stock market has consistently proven to be a good place to invest in and save for the future. There are a lot of compelling reasons to invest in stocks. It can help in fighting inflation, create wealth, and also provides some tax benefits. Good steady returns on investments over a long period of time can also grow a lot more than seems possible. Also, thanks to the power of compound interest, the earlier one starts investing, the larger the corpus one can have for retirement. Overall, investing in stocks can help meet life's financial aspirations.

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.

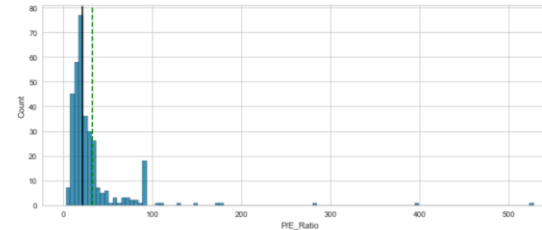
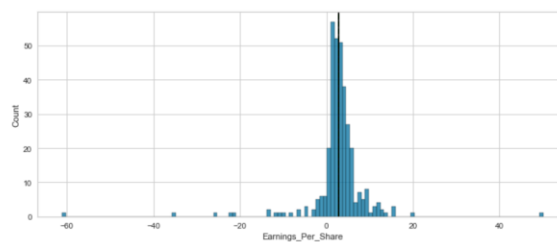
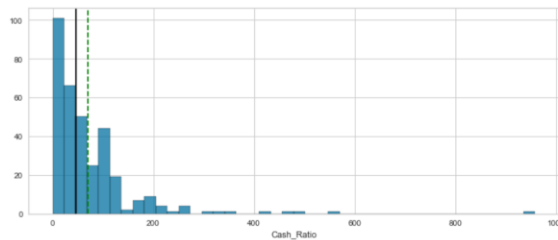
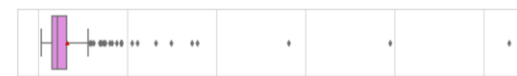
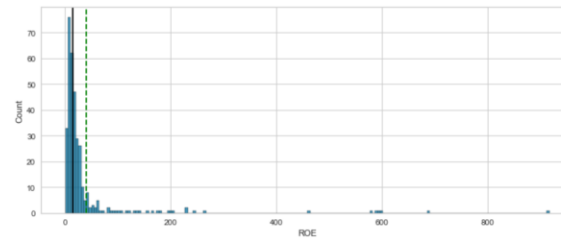
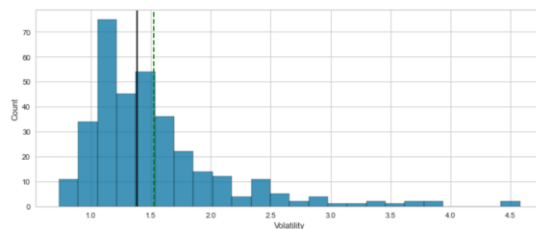
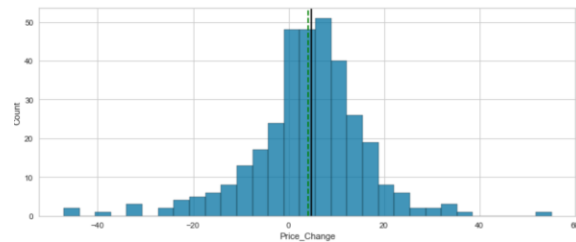
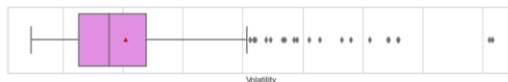
Objective:

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

- 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)

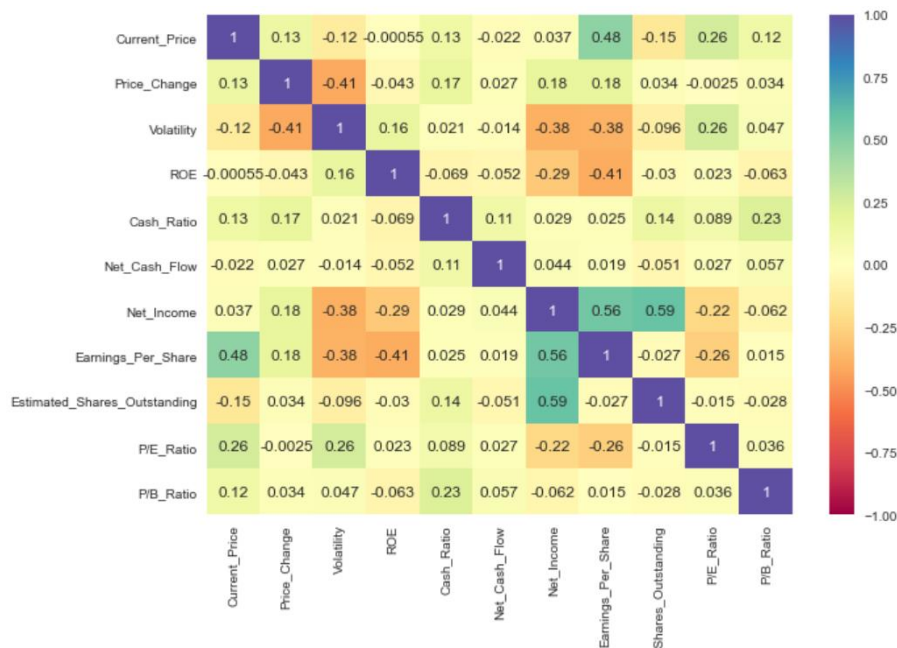
EDA

• UNIVARIATE ANALYSIS

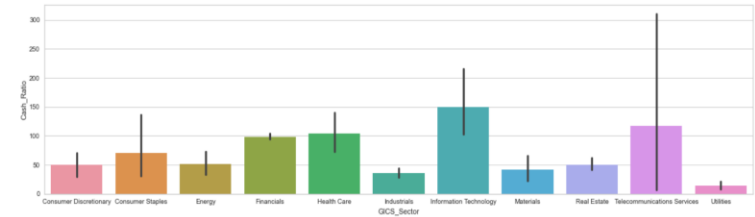
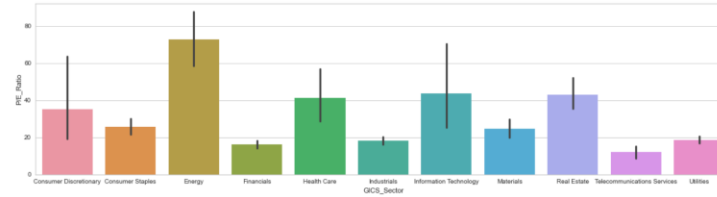
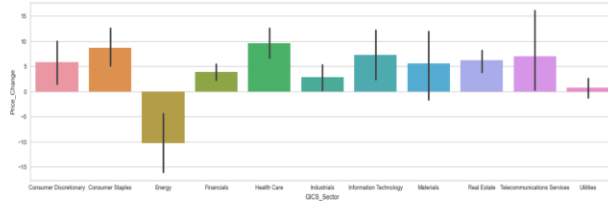


Bivariate Analysis

- There is strong positive correlation for Net_Income with Earnings_Per_Share and Estimated_Shares_Outstanding.
- There are no significant correlation among other variables.



GICS Sector with other Indicators

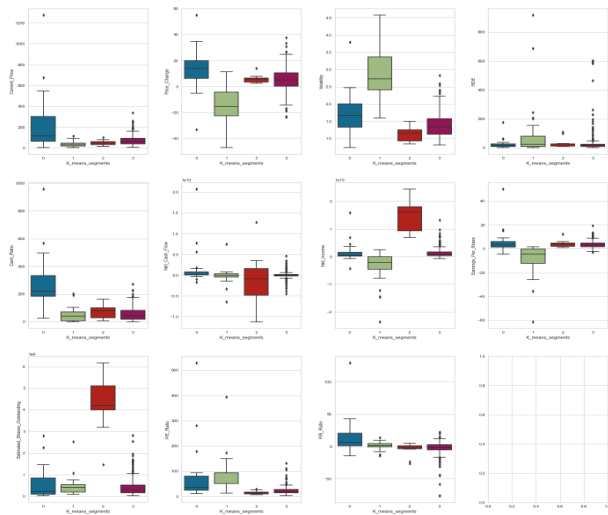


Insights Based on EDA

- Companies are divided into 11 economic sectors known as GICS_Sector.
- Most of the companies come under the sector Industrials with 15.6%.
- Only 1.5% companies come under the sector Telecommunications Services.
- GICS_Sub_Industry has 104 unique values.
- The sub-sector 'Oil and Gas Exploration and Production' has the most count with 4.7%.
- The lowest percent is 0.3% and distributed among many sub-sectors.
- The indicators Price_Change, Net_Cash_Flow, Earnings_Per_Share and P/B_Ratio follow a Normal Distribution,
- All other indicators are skewed towards the right.
- All the indicators are having outliers.
- There is strong positive correlation for Net_Income with Earnings_Per_Share and Estimated_Shares_Outstanding.
- There are no significant correlation among other variables.
- Volatility decreases as Price_Change increases.
- Earnings_Per_Share decreases as ROE increases.
- Earnings_Per_Share increases as Current_Price increases.
- Energy sector has the highest P/E ratio with average of around 75.
- Telecommunications Services has the least P/E ratio with average of around 15.
- Information Technology sector has the highest average of Cash_Ratio around 150.
- Utilities sector has the least average of Cash_Ratio with average of around 15.
- HealthCare Sector has the highest average Price increase with around 10% increase.
- Utilities Sector had least average price increase with 1% increase.
- Energy Sector had price decrease with average of -10%.

K-means Clustering

	Current_Price	Price_Change	Volatility	ROE	Cash_Ratio	Net_Cash_Flow	Net_Income	Earnings_Per_Share	Estimated_Shares_Outstanding	P/E_Ratio	P/B_Ratio	count_in_each_segment
K_means_segments												
0	234.170932	13.400685	1.729989	25.600000	277.640000	1554926560.000000	1572611680.000000	6.045200	578316318.948800	74.960824	14.402452	25
1	38.099260	-15.370329	2.910500	107.074074	50.037037	-159428481.481481	-3887457740.740741	-9.473704	480398572.845926	90.619220	1.342067	27
2	50.517273	5.747586	1.130399	31.090909	75.909091	-1072272727.272727	14833090909.090910	4.154545	4298826628.727273	14.803577	-4.552119	11
3	72.399112	5.066225	1.388319	34.620939	53.000000	-14046223.826715	1482212389.891697	3.621029	438533835.667184	23.843656	-3.358948	277



Insights

4 is taken as appropriate no. of clusters as silhouette score is high enough and there is a knick at 4 in elbow curve.

Cluster 0 :

- Current Price, Price Change, Cash_Ratio, Net_Cash_Flow, Earnings_Per_Share and P/B Ratio of the companies is the highest.
- ROE is the lowest.

Cluster 1 :

- Volatility, ROE and P/E Ratio of the companies is the highest.
- Current Price, Price Change, Cash_Ratio, Net_Income, Earnings_Per_Share are the lowest.

Cluster 2 :

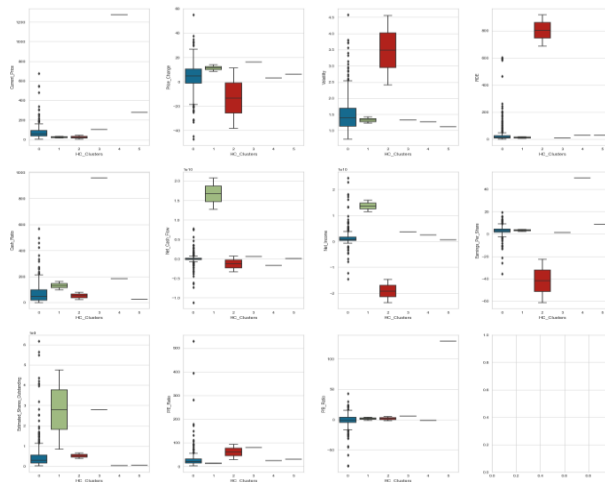
- Net_Income and Estimated_Shares_outstanding is the highest.
- Volatility, Net_Cash_Flow, P/E Ratio and P/B Ratio are the lowest.

Cluster 3:

- The most number of companies are in this cluster.
- There are no high indicators in any area.
- Earnings_Per_Share and Estimated_Shares_outstanding is the lowest.

Hierarchical Clustering

	Current_Price	Price_Change	Volatility	ROE	Cash_Ratio	Net_Cash_Flow	Net_Income	Earnings_Per_Share	Estimated_Shares_Outstanding	P/E_Ratio	P/B_Ratio	K_means_segments	count_in_each_segments
HC_Clusters													
0	77.287589	4.099730	1.518066	35.336336	66.900901	-33197321.321321	1538074666.666667	2.885270	560505037.293543	32.441706	-2.174921	2.630631	333
1	25.640000	11.237908	1.322355	12.500000	130.500000	16755500000.000000	13654000000.000000	3.295000	2791829362.100000	13.649696	1.508484	1.000000	2
2	24.485001	-13.351992	3.482611	802.000000	51.000000	-1292500000.000000	-19106500000.000000	-41.815000	519573983.250000	60.748608	1.565141	1.000000	2
3	104.660004	16.224320	1.320606	8.000000	958.000000	592000000.000000	3669000000.000000	1.310000	2800763359.000000	79.893133	5.884467	0.000000	1
4	1274.949951	3.190527	1.268340	29.000000	184.000000	-1671386000.000000	2551360000.000000	50.090000	50935516.070000	25.453183	-1.052429	0.000000	1
5	276.570007	6.189286	1.116976	30.000000	25.000000	90885000.000000	596541000.000000	8.910000	66951851.850000	31.040405	129.064585	0.000000	1



Insights

Cluster 0 :

- It has highest number of companies 333.
- No Indicators are high
- P/B ratio is lowest.

Cluster 1 :

- Net_Cash_Flow, Net_income is the highest.
- P/E Ratio is the least.

Cluster 2 :

- Volatility and ROE is the highest.
- Current_Price, Price_Change, Net_income, Earnings_Per_Share are the lowest.

Cluster 3:

- Price_Change, Estimated_Shares_outstanding, P/E Ratio and Cash_Ratio are the highest.
- ROE is the lowest.

Cluster 4:

- Current_Price, Earnings_Per_Share are the highest.
- Net_Cash_Flow, Estimated_Shares_outstanding is the lowest.

Cluster 5:

- P/B Ratio is the highest.
- Volatility and Cash Ratio is the lowest.

K-means vs Hierarchical Clustering

- K-Means Clustering had companies more Spread out between different clusters as :25,27,11,277
- In HC it is 333,2,2,1,1,1 which indicates no variability
- In K-Means Cluster 0 has Current Price, Price Change, Cash_Ratio, Net_Cash_Flow, Earnings_Per_Share and P/B Ratio of the companies is the highest and ROE is the lowest. This makes it easier to pick as the Best Cluster.
- In HC some indicators are high on some clusters . There is no definite indication of which cluster is the best.

Business Insights and Recommendations

Use K-Means Clustering Technique.

- **Cluster 0 :**
- Current Price, Price Change, Cash_Ratio, Net_Cash_Flow, Earnings_Per_Share and P/B Ratio of the companies is the highest.
- ROE is the lowest.
- **Cluster 1 :**
- Volatility, ROE and P/E Ratio of the companies is the highest.
- Current Price, Price Change, Cash_Ratio, Net_Income, Earnings_Per_Share are the lowest.
- **Cluster 2 :**
- Net_Income and Estimated_Shares_outstanding is the highest.
- Volatility, Net_Cash_Flow, P/E Ratio and P/B Ratio are the lowest.
- **Cluster 3:**
- The most number of companies are in this cluster.
- There are no high indicators in any area.
- Earnings_Per_Share and Estimated_Shares_outstanding is the lowest.

For getting high results on most of the indicators Cluster 0 companies stocks are the best ones to be invest in.

greatlearning
Power Ahead

Happy Learning !

