




Trade&Ahead Project Business Presentation



Objective

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

Data Information

- **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).

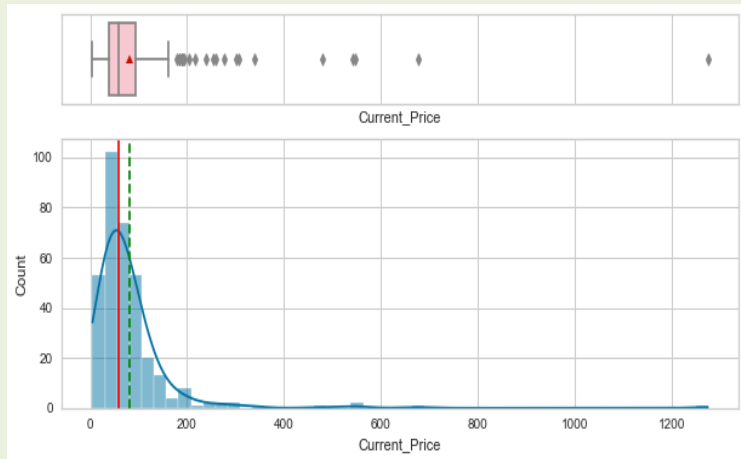
Data Overview

- The data contains information of 340 companies and their financial indicators(15 variables).
- There were no missing values and duplicate values in the dataset.
- The Ticker_Symbol, Company, GICS_Sector, and GICS_Sub_Industry columns have been converted to category data types leading to a reduced memory space.
- The Ticker_Symbol column was dropped from the dataset since all the values in that column were unique.
- There were quite a few outliers in the data, however, they were not treated since they were genuine values.

EDA Results

Univariate Analysis

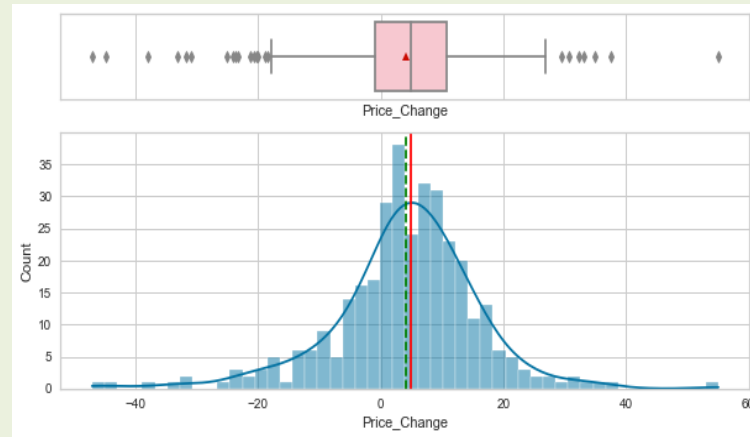
Current_Price



- The distribution of current stock price variable is right-skewed. There are outliers towards the right.

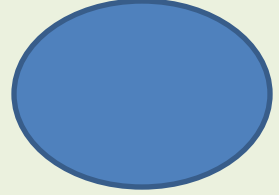
- The median current stock price is ~60 dollars.

Price_Change

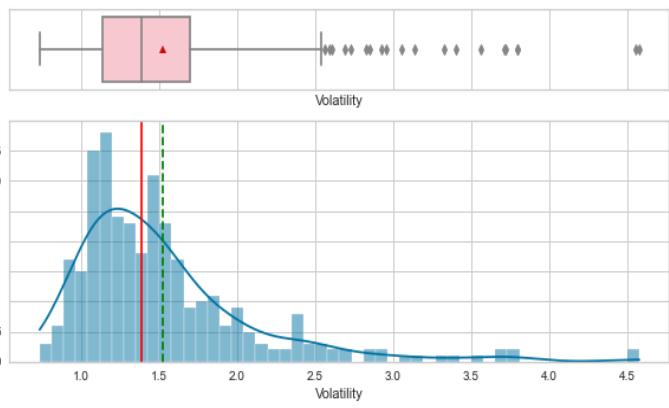


- The distribution of Price_Change variable is close to normal.
- There are outliers on both ends of the boxplot of the Price_Change variable.

Univariate Analysis – Volatility & ROE

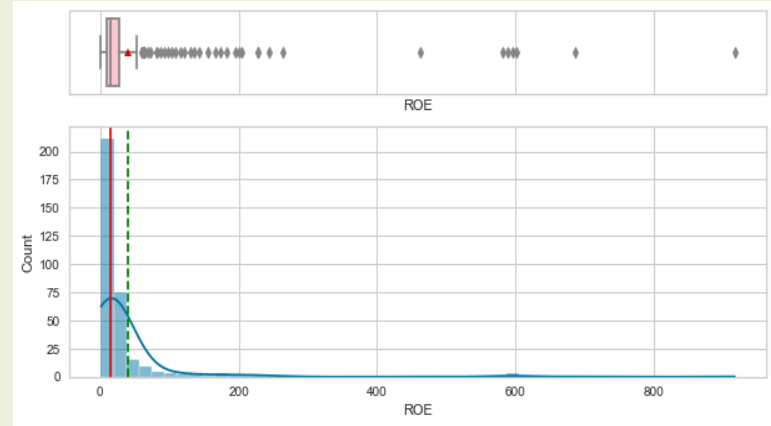


Volatility



- The distribution of Volatility is right-skewed.
- There are outliers towards the right.

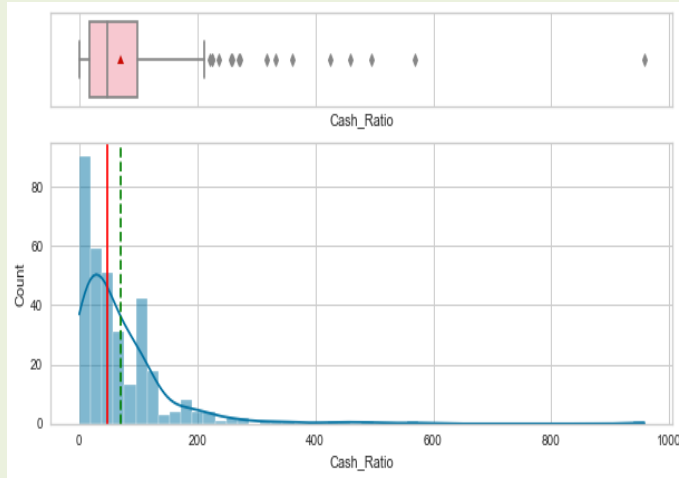
ROE



- The distribution of ROE variable is highly right skewed.
- There are outliers towards the right.

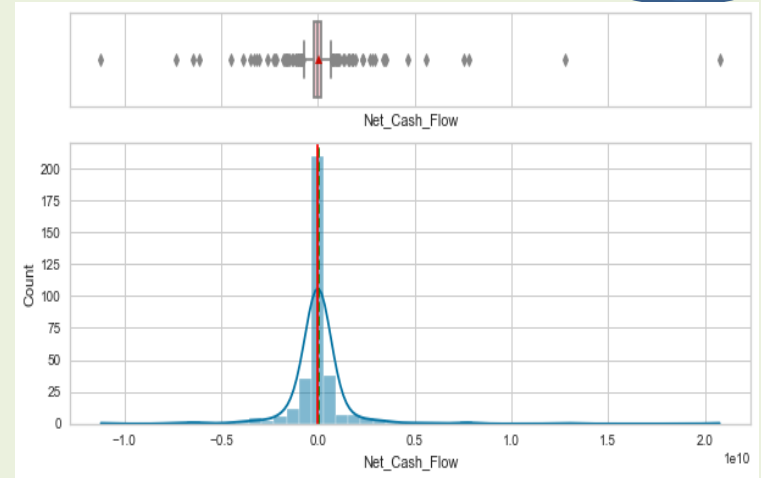
Univariate Analysis – Cash_Ratio & Net_Cash_Flow

Cash_Ratio



- The distribution of Cash_Ratio variable is right skewed.
- There are outliers towards the right.

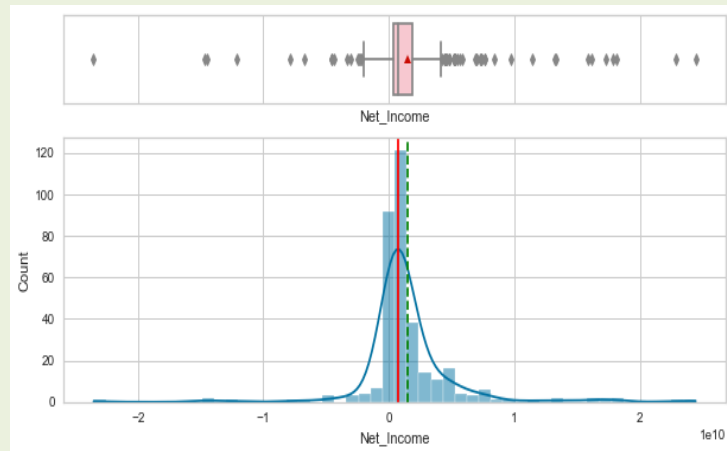
Net_Cash_Flow



- The distribution of Net_Cash_Flow variable is close to normal.
- There are outliers on both ends of the box plot of the Net_Cash_Flow variable.

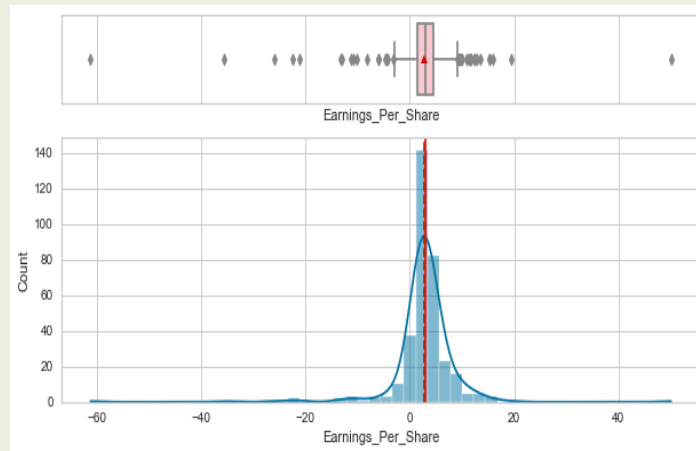
Univariate Analysis – Net_Income & Earnings_Per_Share

Net_Income



- The distribution of Net_Income variable is close to normal.
- There are outliers on both ends of the box plot of the Net_Income variable.

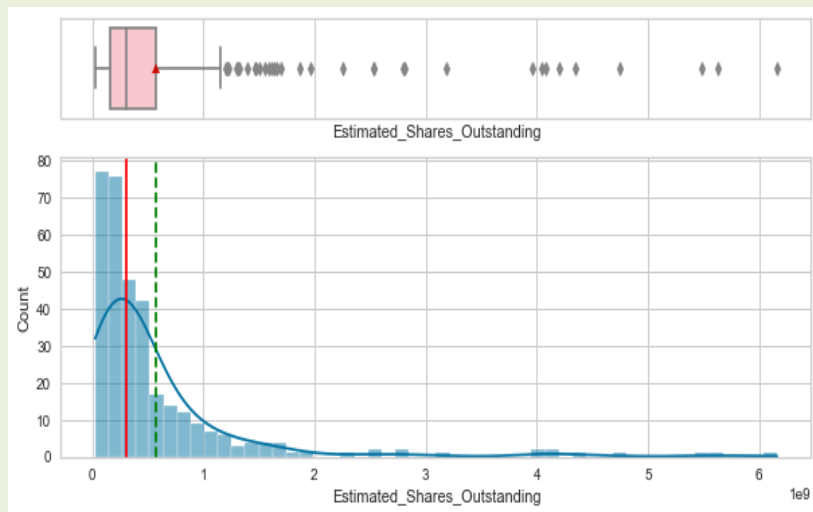
Earnings_Per_Share



- The distribution of Earnings_Per_Share variable is close to normal. The median and the mean values are very close.
- There are outliers on both ends of the boxplot.

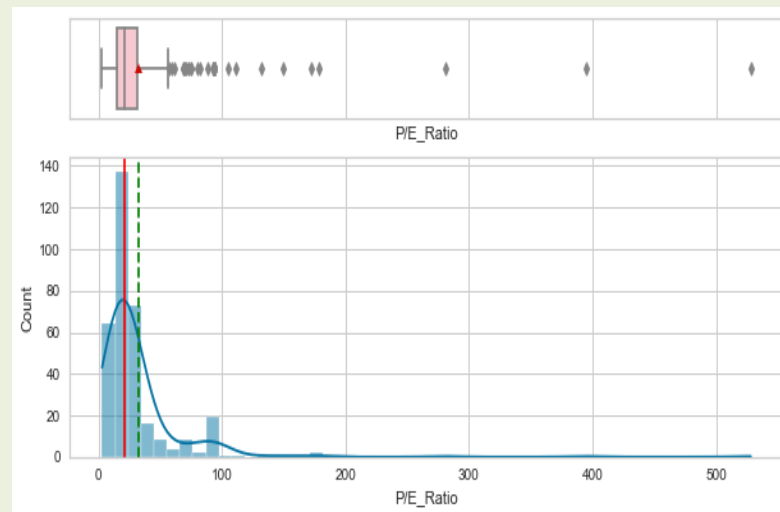
Univariate Analysis – Estimated_Shares_Outstanding & P/E_

Estimated_Shares_Outstanding



- The distribution of Estimated_Shares_Outstanding variable is right-skewed. This indicates that most companies' stock held by all the stakeholders are at the lower end.
- There are outliers towards the right.

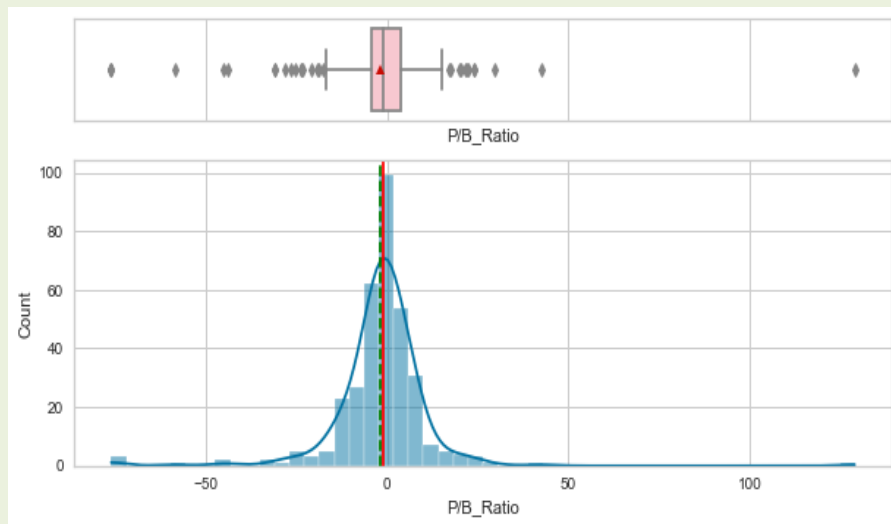
P/E_Ratio



- The distribution of P/E_Ratio variable is right-skewed.
- There are outliers towards the right.

Univariate Analysis – P/B_Ratio

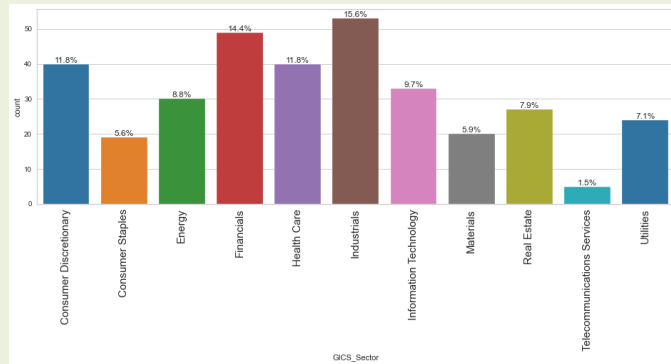
P/B_Ratio



- The distribution of P/B_Ratio variable is close to normal. The median and mean values are very close.
- There are outliers on both ends of the boxplot.

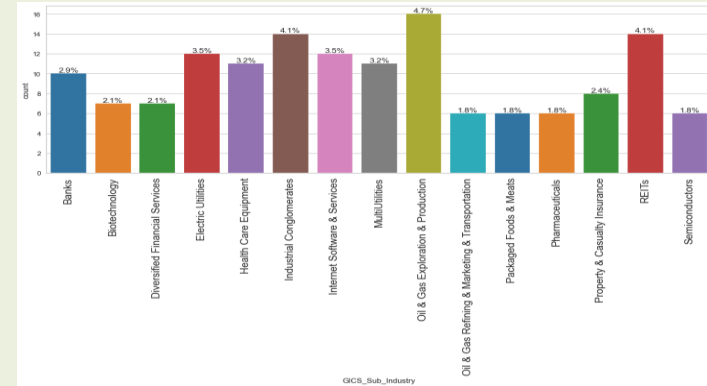
Univariate Analysis – GICS_Sector & GICS_Sub_Industry

GICS_Sector



- There are 11 different economic sectors assigned to companies in this dataset.
- 15.6% of companies in this dataset belong to the industrial economic sector, 14.4% of companies belong to the Financial economic sector, 11.8% of companies belong to the Healthcare economic sector, 11.8% of companies belong to the Consumer Discretionary economic sector, and 9.7% belong to the information technology economic sector.
- Only 1.5% of companies belong to the Telecommunications Services economic sector.

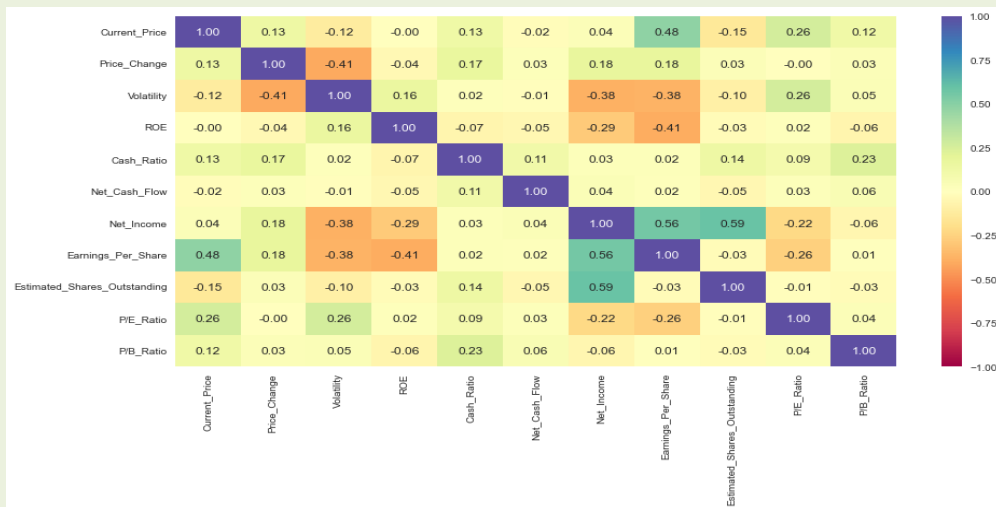
GICS_Sub_Industry



- There are 104 different specific sub-industry groups in this dataset to which the companies are assigned to.
- 4.7% of companies in this dataset are assigned to the sub-industry group-Oil & Gas Exploration & Production, 4.1% of companies are assigned to the sub-industry group-REITs and 4.1% of companies are assigned to the sub-industry group-Industrial Conglomerates.

Bivariate Analysis

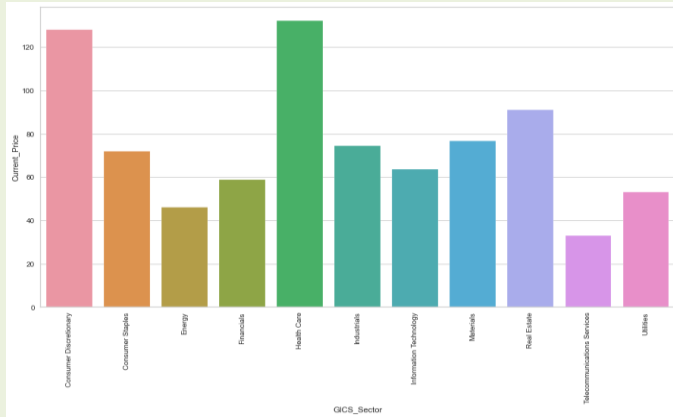
Correlation Amongst Variables



- Net_Income shows an average positive correlation with Earnings_Per_Share and Estimated_Shares_Outstanding.
- Earnings_Per_Share shows an average positive correlation with Current_Price.
- Price_Change shows an average negative correlation with Volatility
- Earnings_Per_Share shows an average negative correlation with ROE.

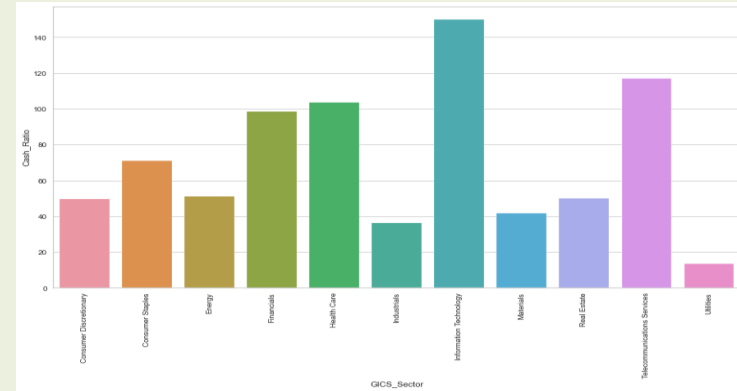
Bivariate Analysis

GICS_Sector vs Current_Price



- The stocks of the Healthcare economic sector have shown the maximum average price followed by the Consumer Discretionary and the Real Estate economic sectors.
- On the otherhand,the stocks of the Telecommunications Service economic sector have shown the lowest average price

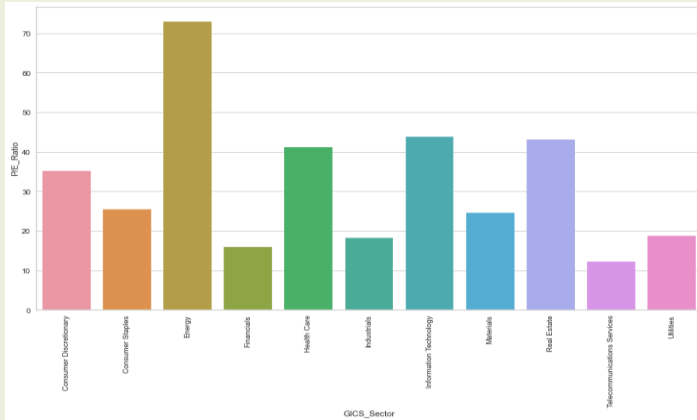
GICS_Sector vs Cash_Ratio



- The economic sector Information technology has the highest average cash ratio followed by the economic sectors Telecommunications services and Healthcare.
- On the otherhand,the economic sector Utilities have the lowest average cash ratio.

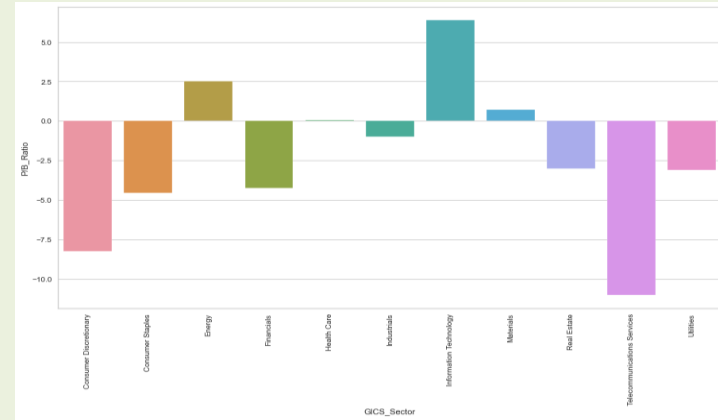
Bivariate Analysis

GICS_Sector vs P/E_Ratio



- The economic sector with the highest average P/E ratio is the Energy sector followed by the Information Technology and Real Estate economic sectors.
- The Telecommunications Services economic sector has the lowest average P/E ratio

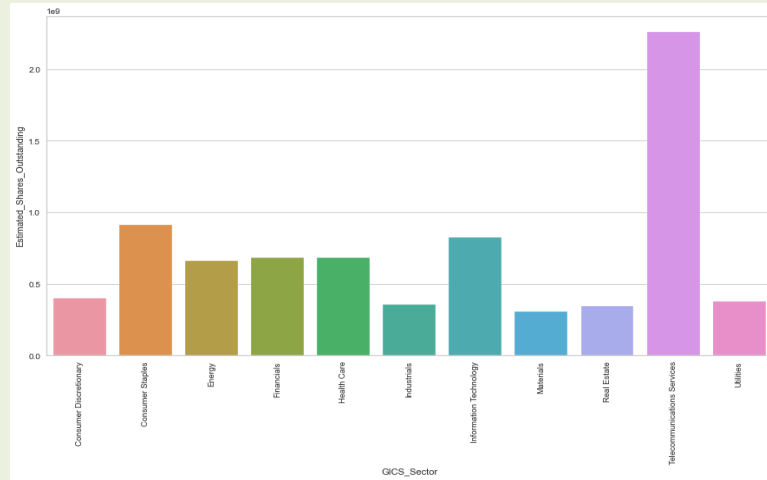
GICS_Sector vs P/B_Ratio



- Information Technology economic sector has the highest average P/B Ratio.
- Telecommunications Services economic sector has the lowest average P/B Ratio.

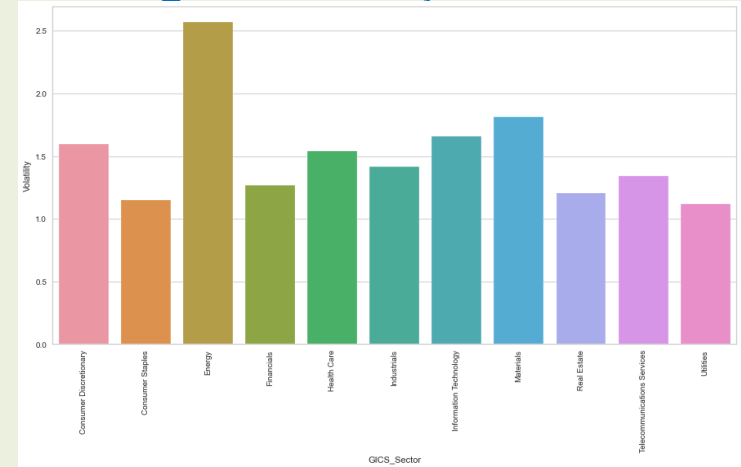
Bivariate Analysis

GICS_Sector vs Estimated_Shares_Outstanding



- The economic sector Telecommunications services have the highest average Estimated Shares Outstanding.

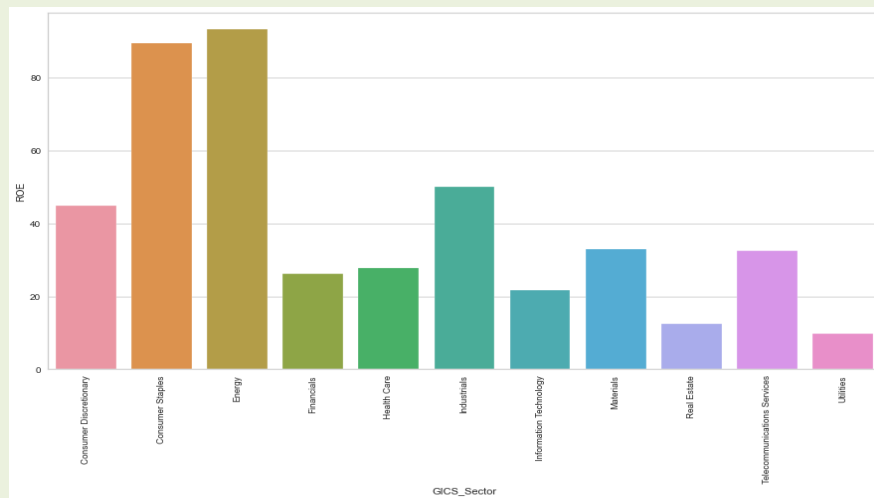
GICS_Sector vs Volatility



- The Energy economic sector has the highest average volatility followed by the sectors Materials and Information Technology. This might imply that these sectors are more susceptible to wild fluctuation in the stock price.
- The Utilities economic sector have the lowest average volatility.

Bivariate Analysis

GICS_Sector vs ROE



- The Energy economic sector has the highest average ROE followed by the Consumer Staples and industrials economic sectors.
- The Utilities economic sector has the lowest average ROE.

K-means Clustering - Cluster Profiling

Kmeans-Segments	Current_Price	Price_Change	Volatility	ROE	Cash_Ratio	Net_Cash_Flow	Net_Income	Earnings_Per_Share	Estimated_Shares_Out	P/E_Ratio	P/B_Ratio	Count_in_each_Segment
0	72.399111	5.066224592	1.388318825	34.62093863	53	-14046223.83	1482212390	3.621028881	438533835.7	23.8436558	-3.358947913	277
1	50.51727264	5.747586465	1.130398854	31.09090909	75.90909091	-1072272727	14833090909	4.154545455	4298826629	14.80357669	-4.552119343	11
2	38.09925963	-15.3703295	2.910500321	107.0740741	50.03703704	-159428481.5	-3887457741	-9.473703704	480398572.8	90.61921995	1.342067035	27
3	234.1709317	13.40068518	1.729988627	25.6	277.64	1554926560	1572611680	6.0452	578316318.9	74.96082356	14.40245164	25

K-means Clustering

Note:

Cluster 0:

- There are **277 companies** in this cluster.
- This cluster have an average of 23.84 P/E Ratio. It has a quite low average P/E Ratio.
- The maximum stock price in this cluster is less 400dollars with a median stock price of ~ 70dollars and an average stock price of ~73dollars .
- Volatility in this cluster ranges from ~0.5 to ~2.8 with an average of ~1.4.
- Companies in this cluster have cash ratio less than 380.The average cash ratio of this cluster is 53.
- **Most companies in this cluster belong to the industrial economic sector followed by the financial economic sector.**

Cluster 1:

- There are **11 companies** in this cluster.
- This cluster have a maximum stock price of less than 180dollars and an average stock price of ~60dollars with the lowest average volatility.
- The average cash ratio of this cluster is ~76.
- **Most companies in this group belong to the financials economic sectors.**

Cluster 2:

- There are **27 companies** in this cluster.
- Companies in this cluster have stock price of less than 200dollars.This cluster has the lowest average stock price (38.1dollars) with the highest volatility(an average of ~3).
- This cluster have the highest average P/E Ratio and low average P/B Ratio(less than 2).
- This cluster have the lowest average cash ratio(of ~50).
- **Most companies in this dataset belongs to the Energy economic sector.**

Cluster 3 :

- There are **25 companies** in this cluster.
- This cluster have an average P/E Ratio of ~75 and the highest average P/B Ratio.
- This cluster have the highest average stock price with an average volatility of ~1.73.
- This cluster have the highest average cash ratio.
- **Most of the companies in this dataset belongs to the healthcare economic sector followed by the Consumer Discretionary and Information Technology economic sectors.**

Hierarchical Clustering - Cluster Profiling

HC_Clusters	Current_Price	Price_Change	Volatility	ROE	Cash_Ratio	Net_Cash_Flow	Net_Income	Earnings_Per_Share	Estimated_Shares_Outst	P/E_Ratio	P/B_Ratio	Count_in_each_Segment
0	29.21	-13.48105574	2.711602667	29	32.5	-56575000	-2339890500	-6.07	404361979.5	93.0892875	1.273529946	24
1	114.379997	14.0350948	1.630258903	15	198	160383000	546733000	2.3	422900000	34.33913174	5.884467157	27
2	46.220001	5.532912271	1.130337336	11	74	-1649000000	16150000000	3.85	4194805195	13.00478493	-1.886881188	9
3	60.545	4.611341626	1.338095096	15	42	2098000	790650000	3.04	291516863.2	19.25631942	-1.360395882	280

Hierarchical Clustering

Note:

Cluster 0:

- There are **24 companies** in this cluster.
- This cluster has the highest average P/E Ratio .
- This cluster have an average P/B Ratio of less than 50.
- This cluster has the lowest average stock price with a high volatility(they have the highest volatility).
- This cluster has the lowest average and median Cash Ratio.
- **Most companies in this cluster belongs to the Energy economic sector.**

Cluster 1:

- There are **27 companies** in this cluster.
- **Most companies in this cluster belongs to the Information Technology economic sector followed by the Health care sector.**
- This cluster have a quite high average P/E Ratio and the highest average P/B Ratio.
- This cluster has the highest average stock price with moderate volatility(average volatility less than 2).
- It has the highest average and median cash ratio.

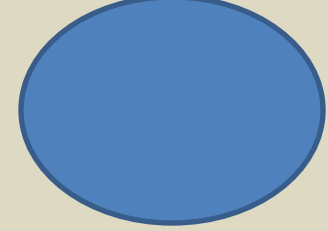
Cluster 2:

- There are **9 companies** in this cluster.
- **A great number of companies in this cluster belong to the Financials economic sector followed by the Telecommunications Services.**
- It has the lowest average P/E Ratio.
- This cluster has a relatively low average stock price with the lowest average volatility.
- Companies in this cluster have a relatively low cash ratio.

Cluster 3 :

- There are **280 companies** in this cluster.
- **Most companies in this cluster belongs to the Industrials economic sector.**
- They have a moderate average stock price with a moderate average volatility.
- They have a low average P/B_Ratio.
- They have a P/E Ratio less than 200
- Companies in this cluster have cash ratio less than 400.

Recommendations:



Based on cluster profile above done by the K-means Clustering Algorithm,

- ❑ **Cluster 2** consists of companies with very low stock price and a high volatility and a very high P/E Ratio and a very low cash ratio.
- ❑ **Cluster 3** consists of companies with a very high stock price and quite low volatility and a very high cash ratio and a quite high P/E Ratio.
- ❑ **Trade&Ahead firm** should handed over the above clusters profiles above with their given financial metrics to financial experts so they can come out with strategic insights for investments in order to help investors have a better view as to where to invest across different market segments and avoid vulnerable risks and losses.