

Trade&Ahead Project Business Presentation

This file is meant for personal use by mavanipurva@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Contents

- Business Problem Overview
- Data Overview
- Exploratory Data Analysis (EDA)
- Clustering and Optimal Clusters
- 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. Investing in stocks 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. Overall, investing in stocks can help meet life's financial aspirations.
- It is important to maintain a diversified portfolio when investing in stocks as it tends to yield higher returns and reduces risk by tempering potential losses when the market is down. Trade&Ahead is a financial consultancy firm that wants to conduct a cluster analysis on their stock data as it can help identify stocks that exhibit similar characteristics and ones that exhibit minimum correlation, and consequently, maximize earnings under any market condition.
- They have provided data comprising stock price and some financial indicators for a few companies listed under the New York Stock Exchange. The task at hand is to analyze the data, group the stocks based on the attributes provided, and share insights about the characteristics of each group.

This file is meant for personal use by mavanipurva@gmail.com only.

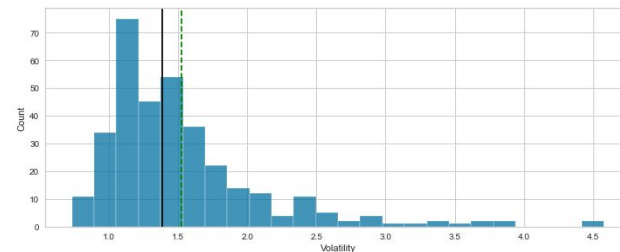
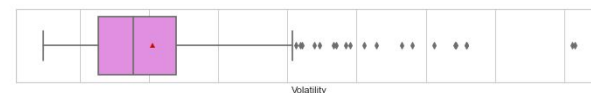
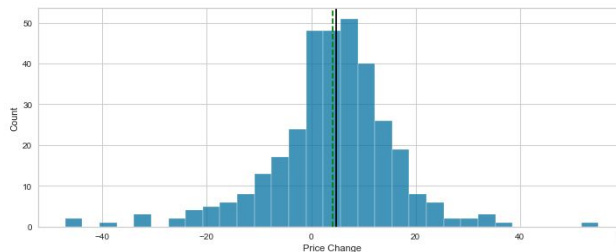
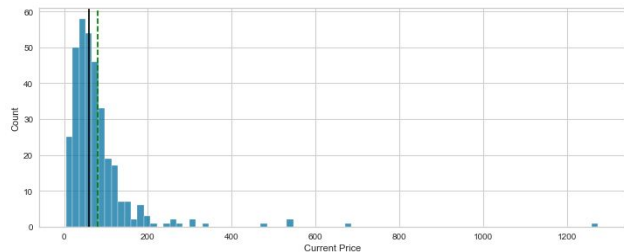
Sharing or publishing the contents in part or full is liable for legal action.
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Data Overview

- The data contains information about 340 companies and their stocks under NYSE.
- The company information includes its financial performance over the past quarter, with attributes like GICS sector and sub-industry, net income, number of outstanding shares, P/E ratio, etc., while the stock information includes attributes like current price, price change, volatility, etc.
- There are no missing or duplicate values in the data.
- The data will be scaled before we use clustering algorithms to segment the players.

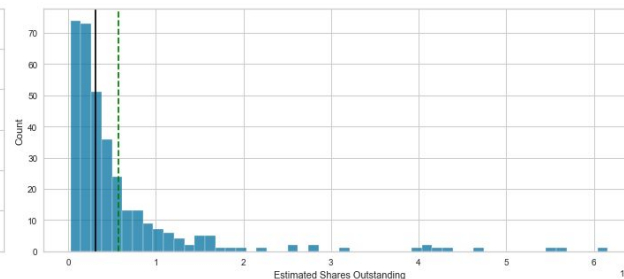
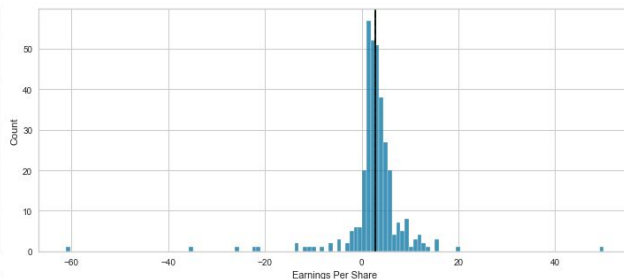
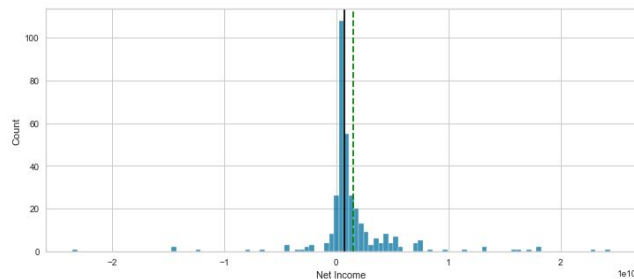
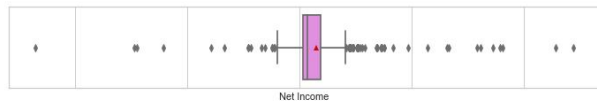
Exploratory Data Analysis (EDA)

- The stock price distribution is heavily right-skewed and very few stocks are price above \$200.
 - There are a few outliers with prices greater than \$750.
- The change in stock price is close to normally distributed.
- The distribution of volatility is slightly right-skewed and few stocks show very high volatility.



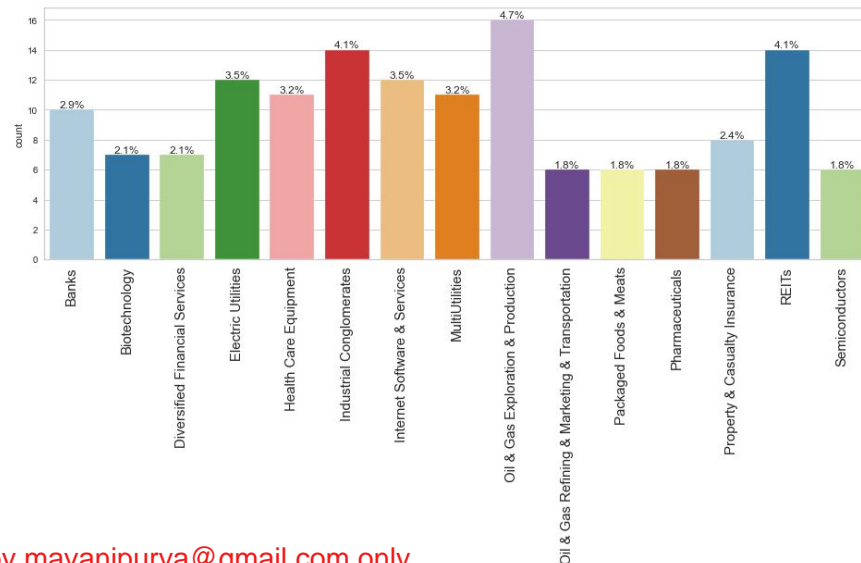
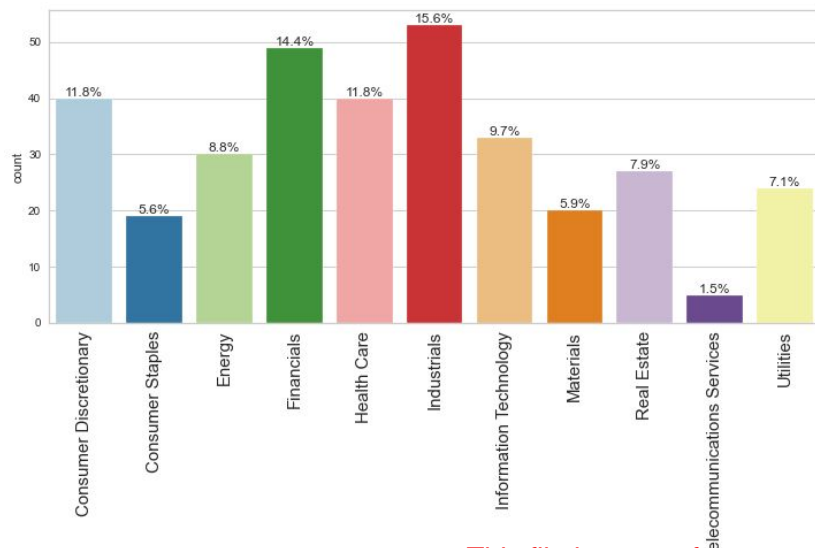
Exploratory Data Analysis (EDA)

- The distribution of net income is close to normal with a lot of outliers.
 - This indicates that some companies made high profits while some incurred heavy losses.
- The distribution of earnings per share is similar to that of net income, which is expected.
- 50% of the companies in the data have less than half a billion outstanding shares, while a few companies have more than 2 billion outstanding shares.



Exploratory Data Analysis (EDA)

- 30% of the companies in the data are from the Financials and Industrials sectors.
- Telecommunication Services has the least weightage of companies in the data.
- Most of the companies in the data belong to the Oil and Gas Exploration and Production sub-industry.



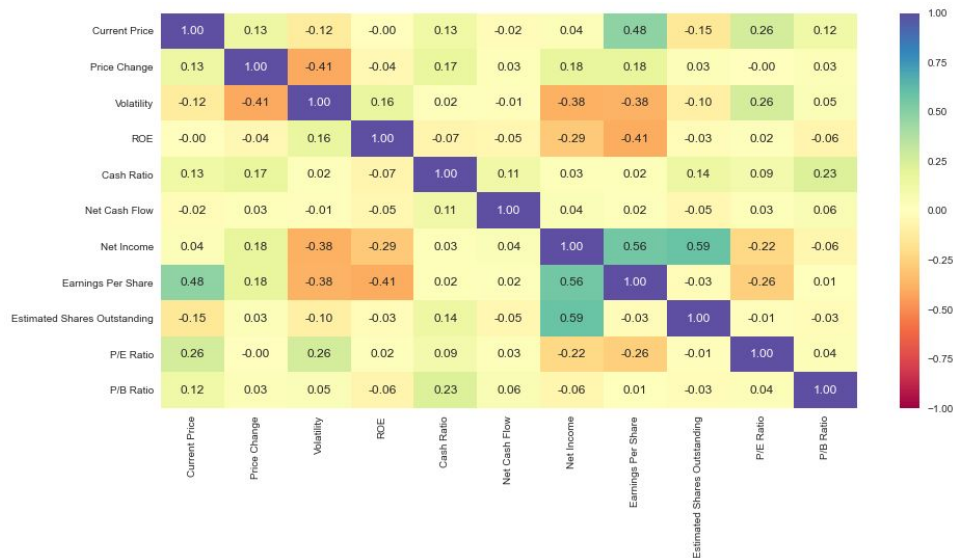
This file is meant for personal use by mavanipurva@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

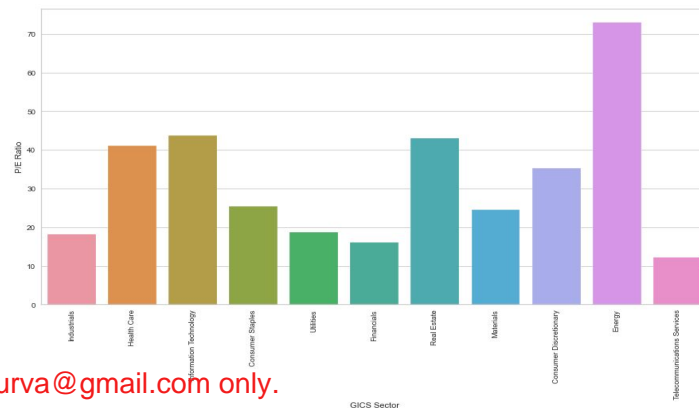
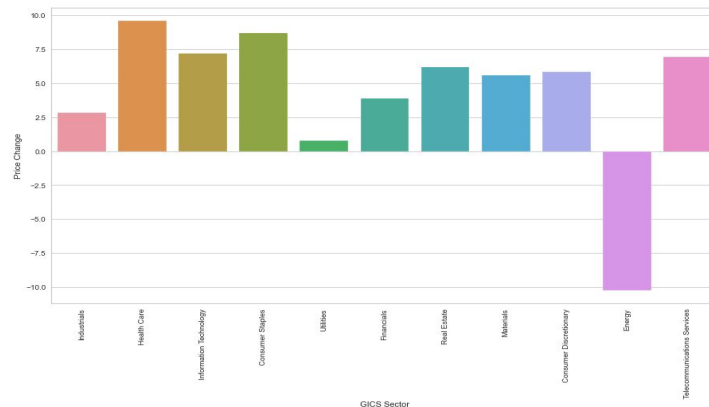
Exploratory Data Analysis (EDA)

- None of the variables show a very high correlation with each other.
- The net income of a company is positively correlated with the earnings per share and the number of outstanding shares.
 - This makes sense as the more money a company makes, the more each shareholder will earn and more people will be willing to invest in the company.
- Percentage change in price and price volatility are negatively correlated.



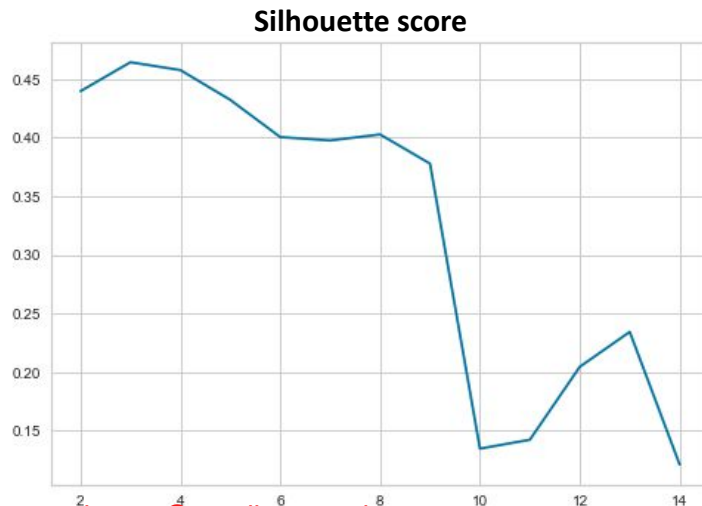
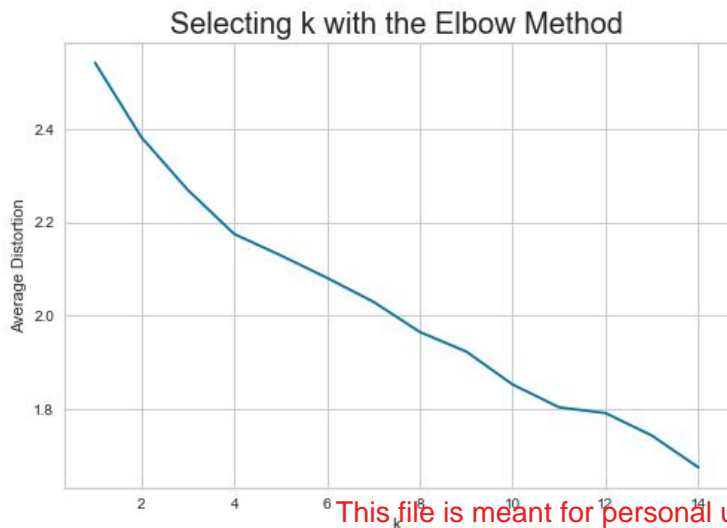
Exploratory Data Analysis (EDA)

- Stocks from the Healthcare sector witnessed the maximum price increase, with an ~9% increase, on average.
- Stocks from the Energy sector were the only ones to witness a price drop on average.
- The P/E ratio is the highest (on average) for the Energy sector companies but they also witness an average price decrease over the past quarter.
 - This might indicate that the stocks of this sector are overvalued.



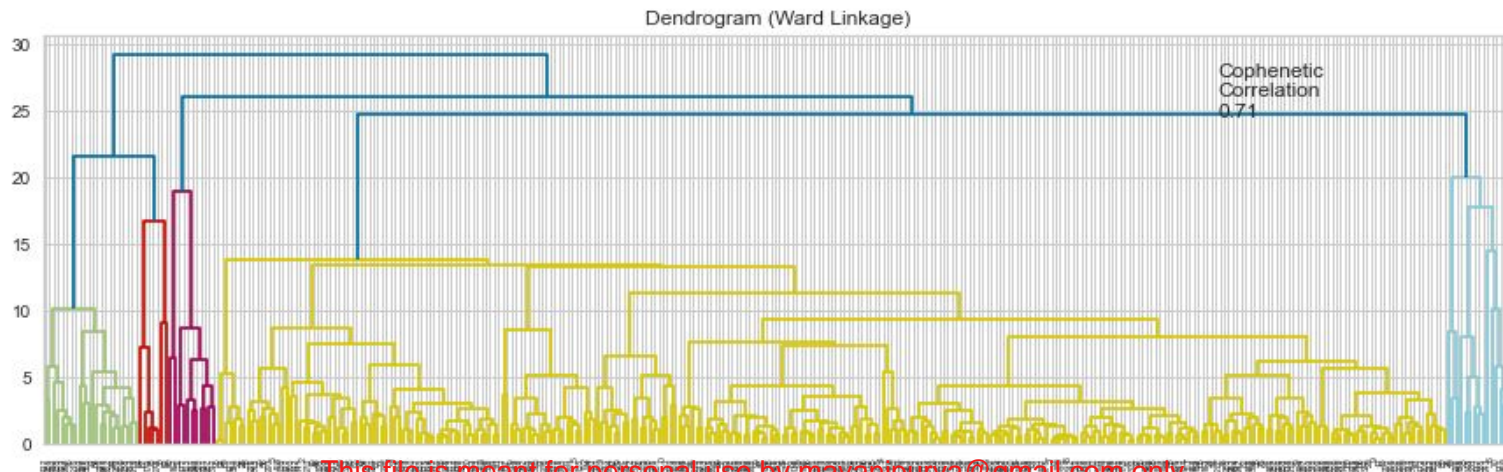
K-means Clustering

- The appropriate value for k is not clearly visible from the elbow plot, but there does seem to be a knick at 4.
- The silhouette score too is high (0.4577) for $k=4$.
- So, we will choose 4 as the value of k .



Hierarchical Clustering

- For hierarchical clustering, we used Euclidean distance and the different linkage methods tested are single, complete, weighted, centroid, average, and Ward.
- The cophenetic correlation is maximum (0.94) with Euclidean distance and average linkage, but the Ward linkage gives more distinct clusters.
- The dendrogram for Ward linkage is shown below.



This file is meant for personal use by mavanipurva@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Business Insights and Recommendations

- Both K-means and Hierarchical clustering yield two similar clusters and two slightly dissimilar clusters. A few companies were swapped between the clusters obtained using both techniques.
- We chose the clusters obtained using Hierarchical clustering and grouped the companies into four groups.
- **Cluster 0** companies are the ones who performed poorly in the previous quarter. Most of the companies in this cluster are from the Energy sector, which as a whole performed poorly in the previous quarter. Investors will have to be careful and dig deeper into the financial statements of these companies for a better analysis.
- **Cluster 1** companies have highly-priced stocks and have shown the most growth in the previous quarter. Large proportions of stocks in this cluster are from Healthcare and IT sectors. The high earnings per share also indicate that these stocks will justify their high price tag with good returns. The high P/E ratio indicates that these stocks are either overvalued or investors expect good growth potential from these companies.

This file is meant for personal use by mavanipurva@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.
Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Business Insights and Recommendations

- **Cluster 2** companies are the ones which have a low price and have had a great previous quarter. They are less riskier investments and are traded in large volumes. The low P/E ratio might also indicate that these stocks are undervalued, but we will have to dig deeper to conclude the same with confidence.
- **Cluster 3** company stocks are moderately priced and are less volatile. They had a good previous quarter and their stocks are traded in moderate volumes. The moderate earnings per share also indicate that these stocks will yield good returns.
- Trade&Ahead should look into more financial (fundamental and technical) indicators to make better predictions of stock price movements and assessment of company valuation.
- They should also conduct cluster analysis separately for each of the economic sectors as it will help them to provide better investment recommendations to their clients.



Happy Learning !

