Trade & Head - Stock Portfolio Analysis

Sathya P

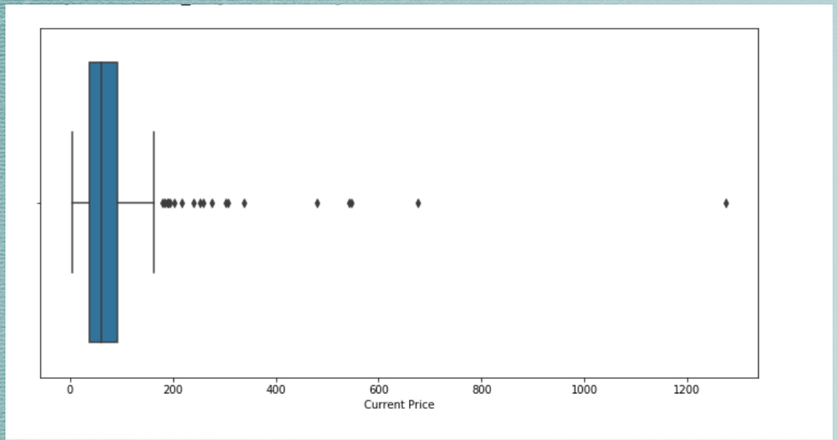
Overview

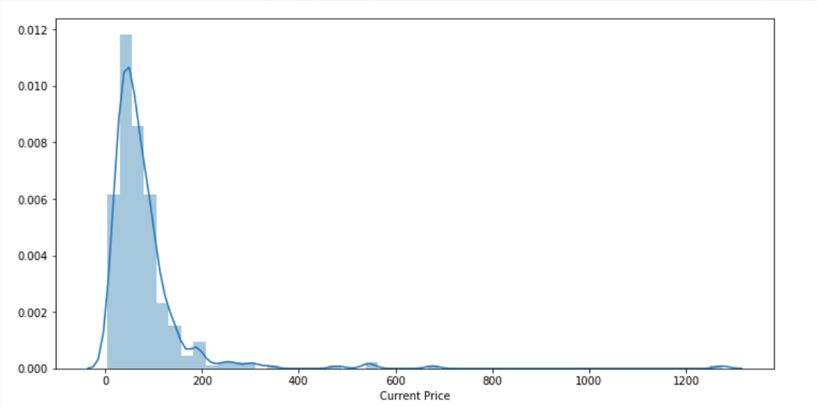
Trade&Ahead is a financial consultancy firm who provide their customers with personalized investment strategies. They have hired you as a Data Scientist and provided you with data comprising stock price and some financial indicators for a few companies listed under the New York Stock Exchange. They have 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 Set

1) Data set has 340 records with 15 attributes
2) No null values in the columns
3) few outliers for the Current Price

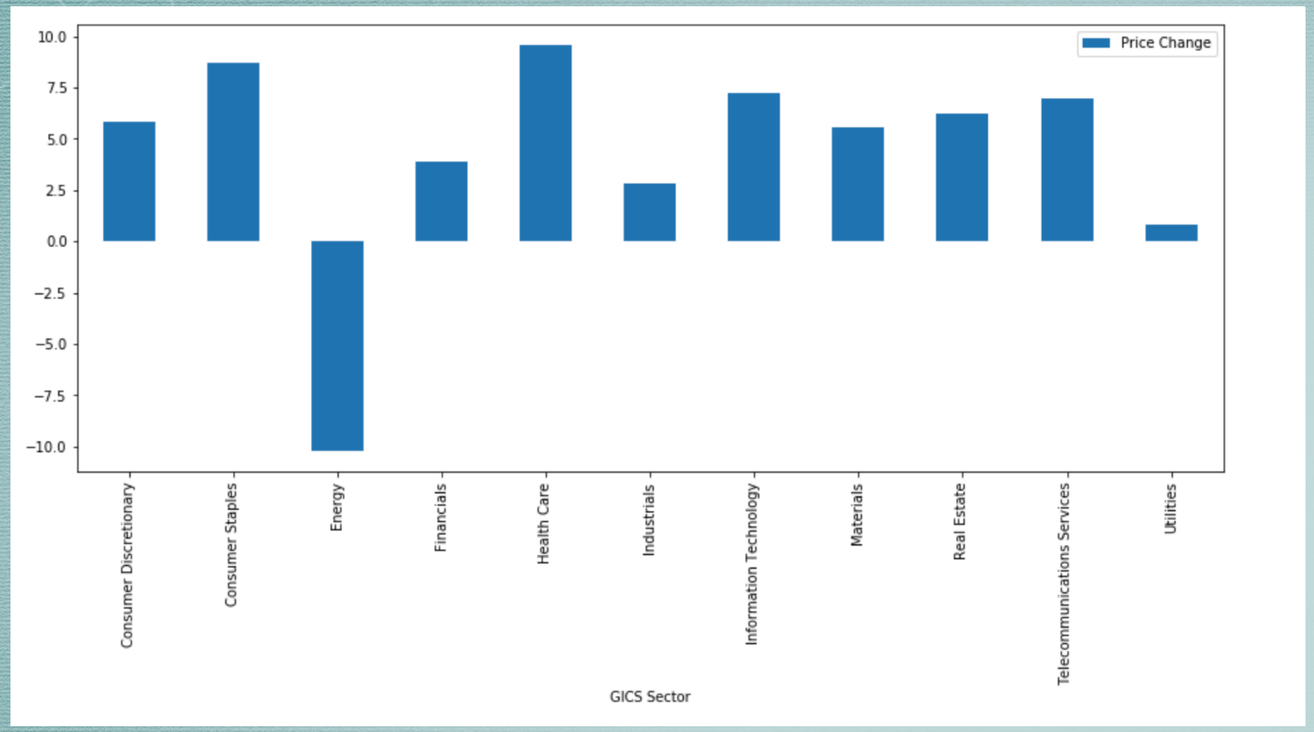
EDA - Current Price Data Distribution





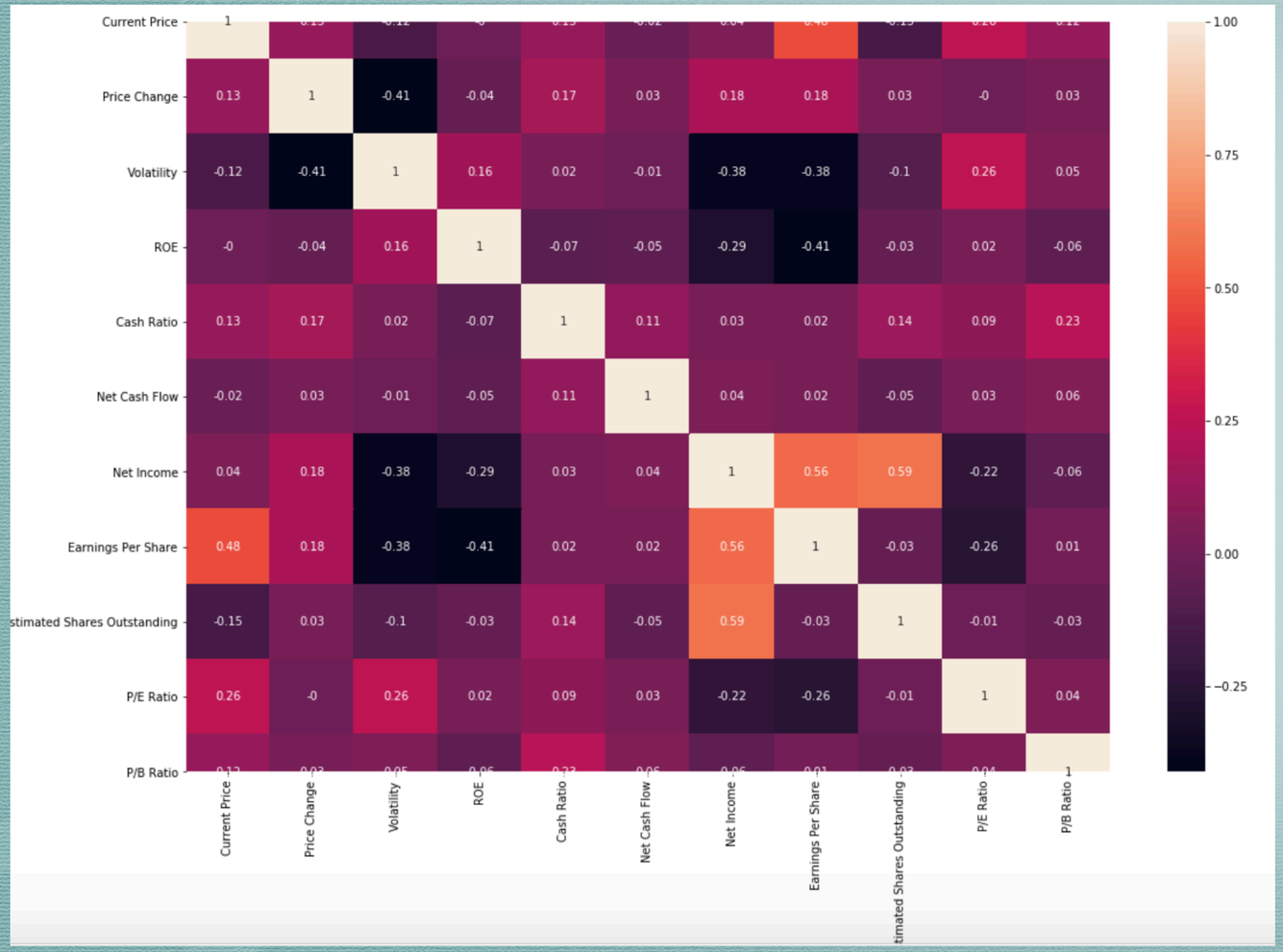
The current price data distribution is slightly right skewed

EDA - Average price Change with Sector wise Price Data Distribution



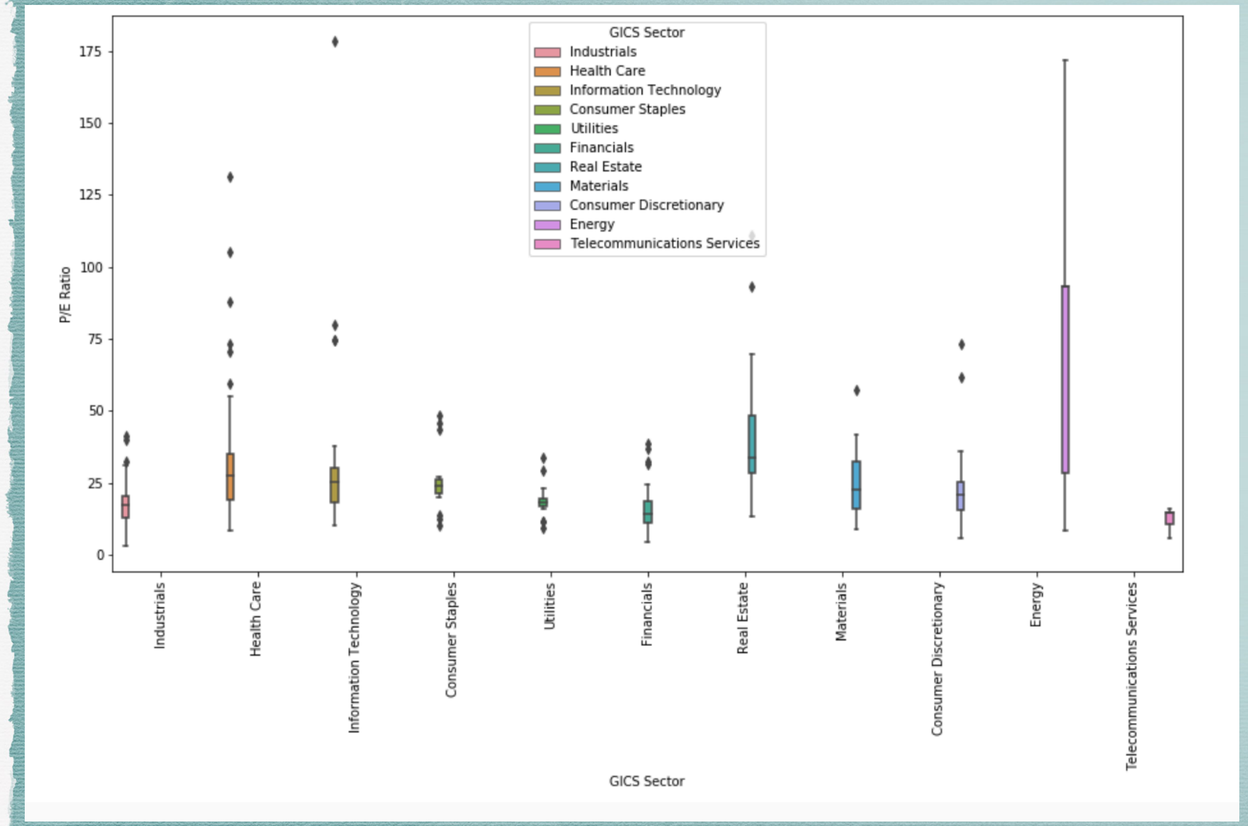
The Health Care has high average Price change
Energy sector is going in the negative for the average Price change
Consumer Supplies is the second highest average price change

Heat Map



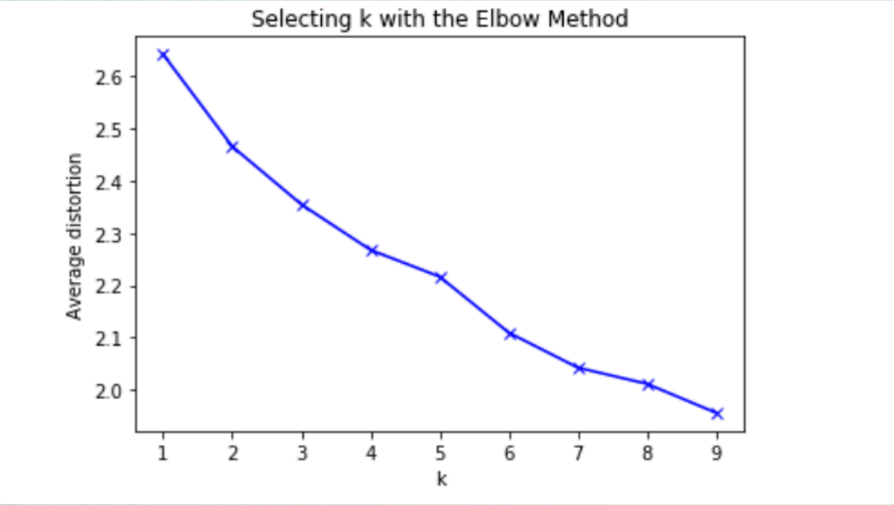
Net Income is one which is having co-relation greater than 0.5 with "Earnings Per Share" and "Estimated Shares Outstanding"

EDA - P/E ratio Average



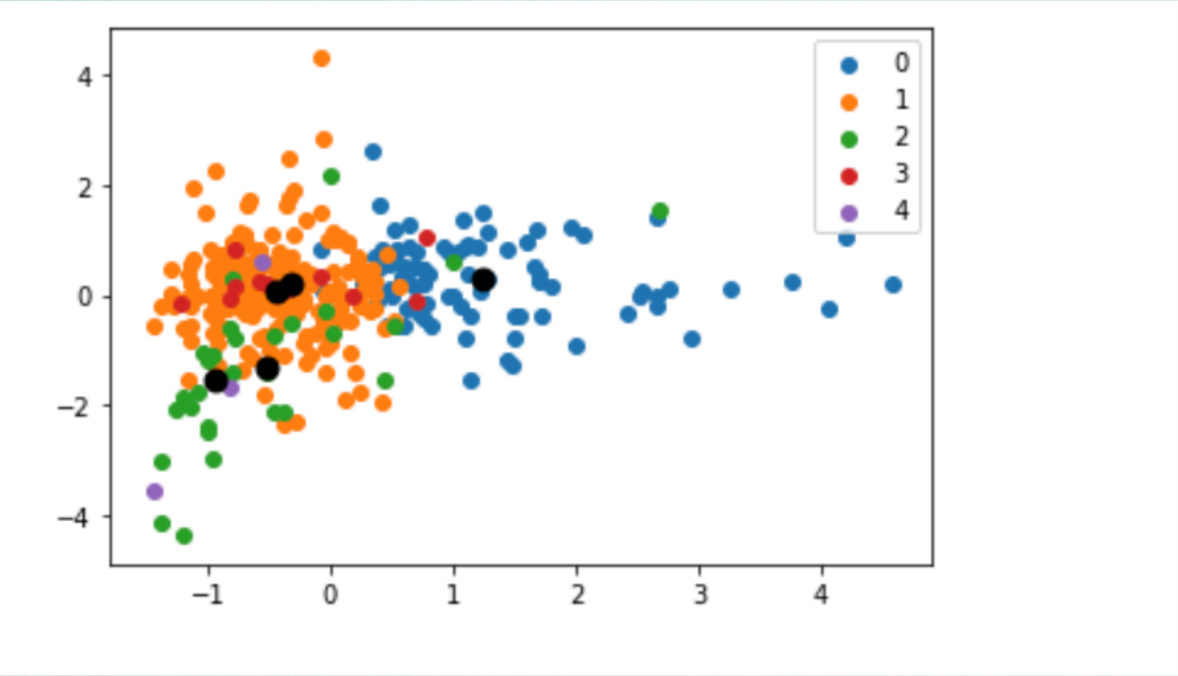
1) Energy has the highest P/E ration in average
2) Second highest is health care sector.

K - Means : Elbow Method



Though the bend is not coming out clearly as there are many bends, let us look at 5 clusters

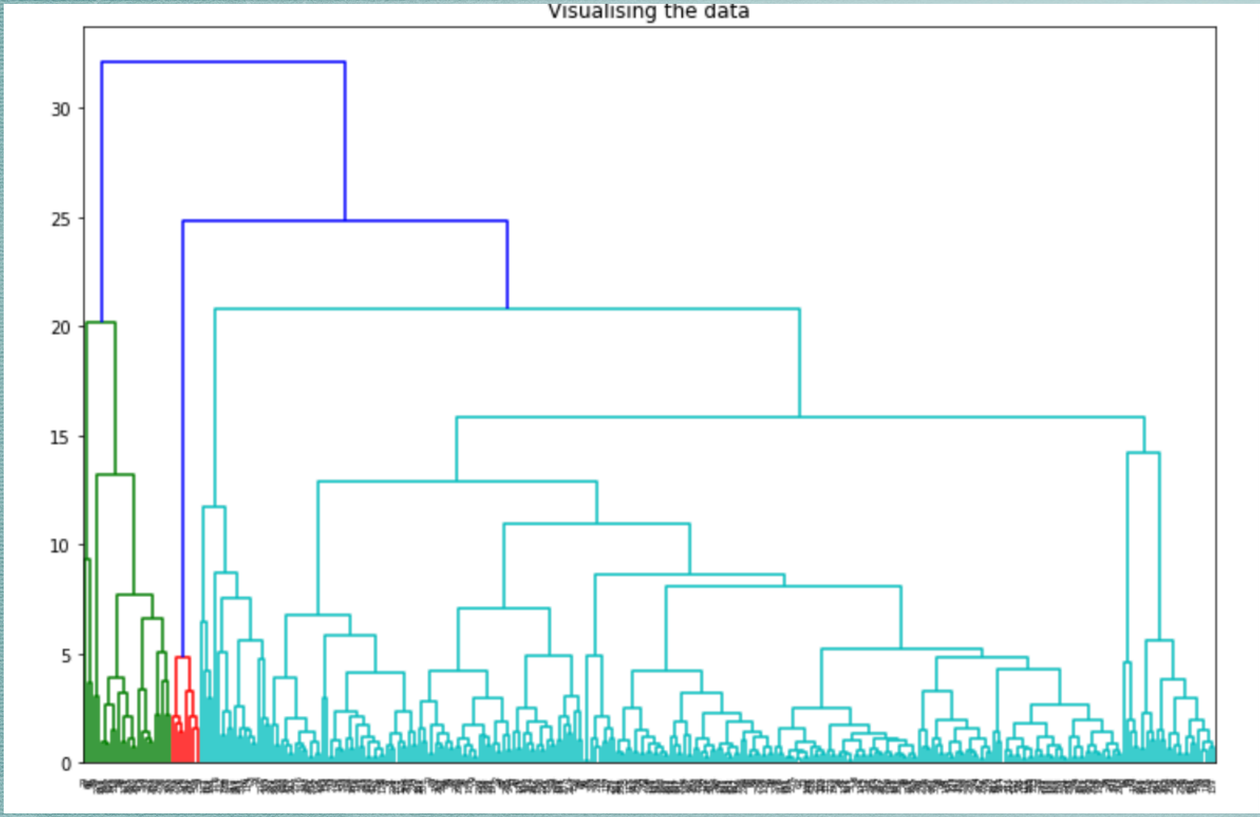
K - Means : Scatter Plot with 5 Clusters



Scatter plot with count of records for each group

1	201
0	87
2	29
3	12
4	3

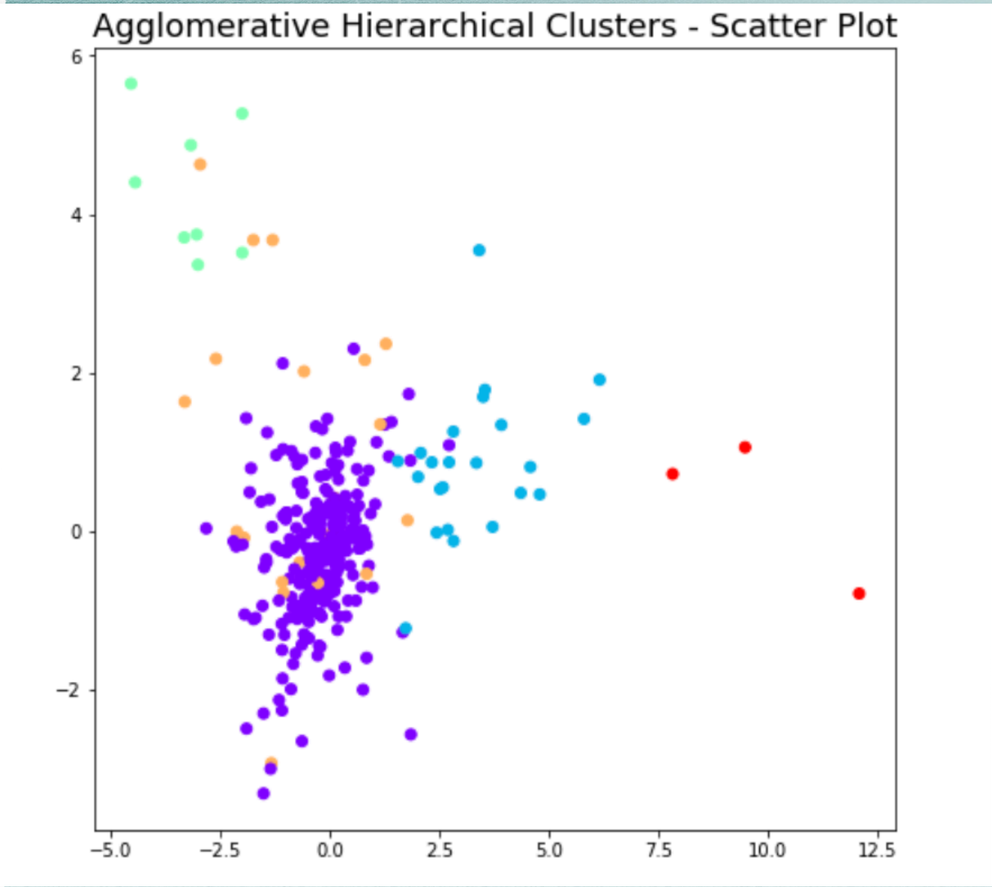
Hiearchical Clustering : PCA & dendrogram



Thru PCA reduced to 5 components /Features

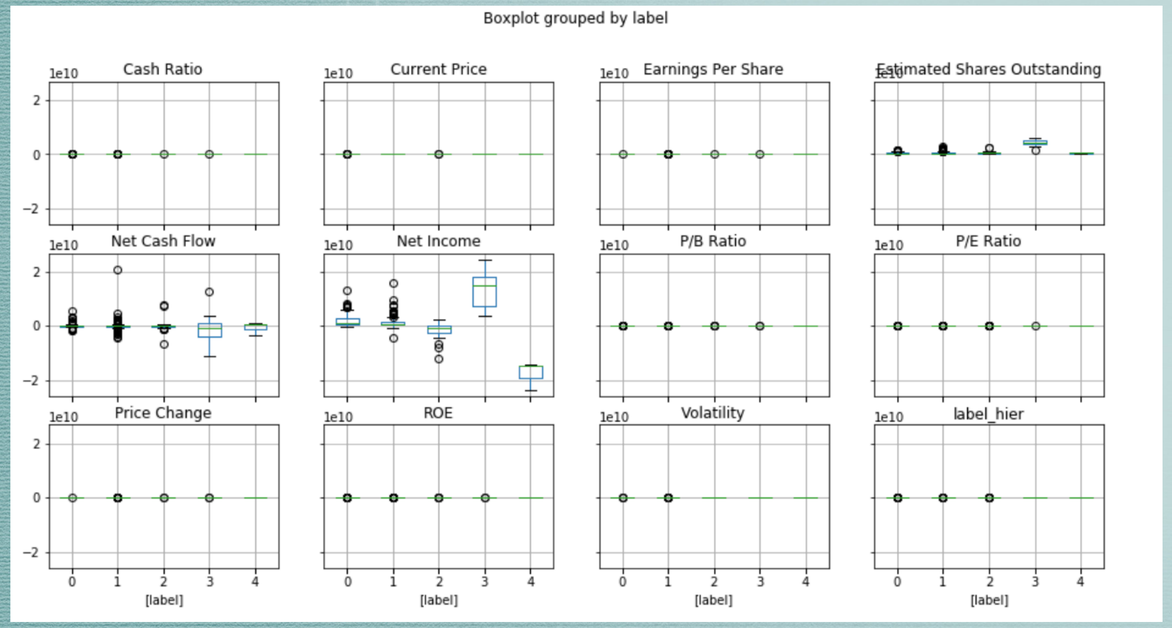
And from the dendrogram we chose 5 clusters between 15-20

Hierarchical Clustering : Scatter Plot with 5 Clusters

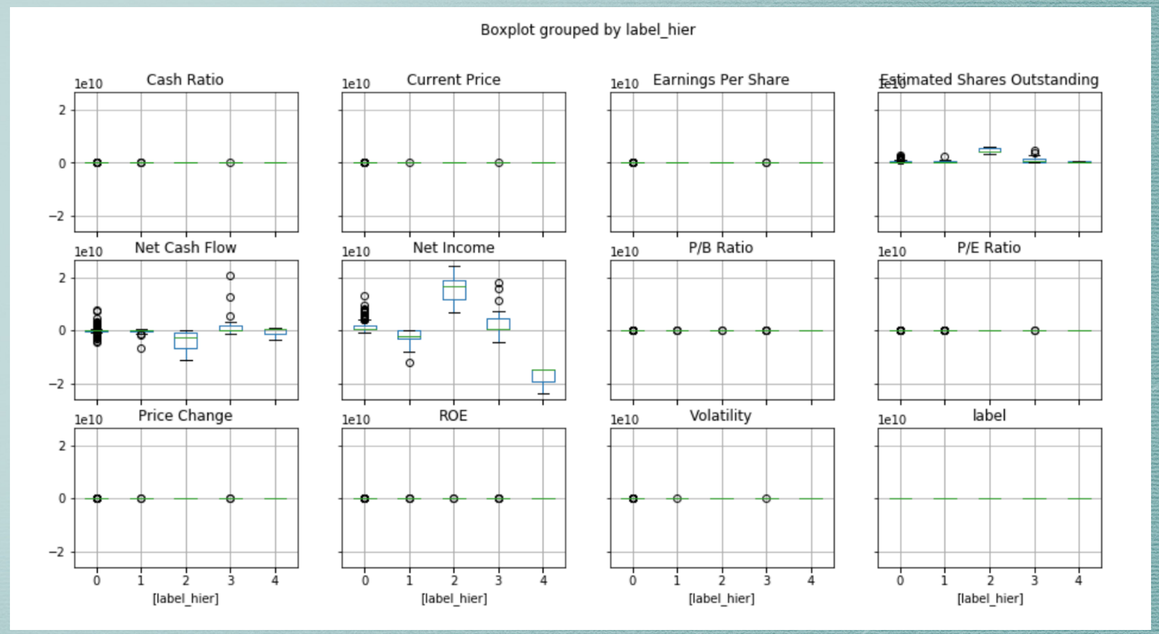


We use agglomerative hierarchical clustering, clustering starts from individual points and clusters are formed upward until one cluster – root cluster remains. It is also called as **bottom-up hierarchical clustering**

K-Means Vs Hierarchical Clustering



K Means Clustering



Hierarchical Clustering

Business Insights

The Hierarchical clustering has more clear clustering than the K - Means
Lable -2 in Hierarchical CLustering are high Net Income records
Label-3 in k-Means are high Net Income records
The clustering in both k-Means& Hierarchical is mainly based on the Net Income .
The second influencing feature for the clustering is Net Cash Flow

Thank You