Trade & Ahead – Stock Market Segmentation

K-Means Clustering/Hierarchal Clustering

Arpan Dinesh

08/03/2023

CONTENTS

- Executive Summary

- Business Problem

- Solution Approach

- Data Overview

- EDA

- Data Pre-Processing

- K-Means Clustering

- Hierarchal Clustering

- K-Means vs Hierarchal Clustering

- Business Recommendations

EXECUTIVE SUMMARY

Trade&Ahead, a renowned financial consultancy firm, endeavors to revolutionize personalized investment strategies by harnessing data-driven insights. The firm's objective is to empower clients navigating the stock market with informed decisions, ensuring alignment with unique financial aspirations.

The proposed solution approach encompasses comprehensive steps for robust cluster analysis. Commencing with exploratory data analysis and diligent pre-processing, the process proceeds to K-means and Hierarchical Clustering, ensuring meticulous model evaluation and comparison. The approach culminates in actionable business recommendations, guiding clients towards profitable choices.

K-means clustering yields strategic recommendations. Cluster 2's robust financials make it an ideal candidate for diversification. Cluster 1, high-risk yet high-reward, appeals to risk-seeking investors. Cluster 0's undervaluation potential suits value investing, while Cluster 3 offers stability and Cluster 3 sustains long-term growth.

Hierarchical clustering augments recommendations. Tailored diversification across clusters caters to individual risk profiles. Cluster 2 beckons for undervalued opportunities, while Cluster 1's growth potential merits attention. Cluster 0 and 3, balanced and stable, harmonize risk and return.

The clustering comparison highlights K-means efficiency for predefined clusters, while Hierarchical Clustering excels in accommodating complex structures. The choice hinges on data characteristics and the firm's insights and portfolio goals.

In synergy, these approaches equip Trade&Ahead to redefine investment paradigms, steering clients towards strategic decisions that converge with their financial dreams.

BUSINESS PROBLEM

Trade&Ahead, a financial consultancy firm, aims to optimize personalized investment strategies for clients in the stock market. The firm possesses a dataset with stock prices and financial indicators of companies listed on the New York Stock Exchange.

The task at hand is to perform cluster analysis on this data, grouping stocks based on similar attributes and minimal correlation. Through this approach, Trade&Ahead seeks to create diversified portfolios that maximize returns and minimize risks effectively.

By providing valuable insights into the characteristics of each group, the firm empowers clients to make well-informed investment decisions, aligning with their financial aspirations.

SOLUTION APPROACH

1. Exploratory Data Analysis: Data Overview, Univariate Analysis.

2. Data Pre-Processing: Missing/Duplicate Value Check & Treatment, Outlier Value Check, Scaling Data.

3. K-Means Clustering: Elbow Curve, Silhouette Scores, Cluster Profiling

4. Hierarchical Clustering: Linkage Methods, Dendrograms, Cophenetic Correlations, Cluster Profiling

5. Performance Evaluation: Model Comparison, Final Model Selection.

6. Business Recommendations: Actionable Insights, Profitable Suggestions.

DATA OVERVIEW

- All features and their definitions are shown in the dictionary below.

- The EDA section will explain each variable in detail.

- Random samples from the dataset and the statistical distribution can be seen in the pictures to the right.

- The data has no missing or duplicate values.

Random Sample of Data

| Ticker Symbol	Security	GICS Sector	GICS Sub Industry	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
DVN	Devon Energy Corp.	Energy	Oil & Gas Exploration & Production	32.000000	-15.478079	2.923698	205	70	830000000	-14454000000	-35.55	4.065823e+08	93.089287	1.785616
FB	Facebook	Information Technology	Internet Software & Services	104.660004	16.224320	1.320606	8	958	592000000	3669000000	1.31	2.800763e+09	79.893133	5.884467
AIV	Apartment Investment & Mgmt	Real Estate	REITs	40.029999	7.578608	1.163334	15	47	21818000	248710000	1.52	1.636250e+08	26.335526	-1.269332
PG	Procter & Gamble	Consumer Staples	Personal Products	79.410004	10.660538	0.806056	17	129	160383000	636056000	3.28	4.913916e+08	24.070121	-2.256747
OXY	Occidental Petroleum	Energy	Oil & Gas Exploration & Production	67.610001	0.865287	1.589520	32	64	-588000000	-7829000000	-10.23	7.652981e+08	93.089287	3.345102
YUM	Yum! Brands Inc	Consumer Discretionary	Restaurants	52.516175	-8.698917	1.478877	142	27	159000000	1293000000	2.97	4.353535e+08	17.682214	-3.838260
EQT	EQT Corporation	Energy	Oil & Gas Exploration & Production	52.130001	-21.253771	2.364883	2	201	523803000	85171000	0.56	1.520911e+08	93.089287	9.567952
HAL	Halliburton Co.	Energy	Oil & Gas Equipment & Services	34.040001	-5.101751	1.966062	4	189	7786000000	-671000000	-0.79	8.493671e+08	93.089287	17.345857
DFS	Discover Financial Services	Financials	Consumer Finance	53.619999	3.653584	1.159897	20	99	2288000000	2297000000	5.14	4.468872e+08	10.431906	-0.375934
IVZ	Invesco Ltd.	Financials	Asset Management & Custody Banks	33.480000	7.067477	1.580839	12	67	412000000	968100000	2.26	4.283628e+08	14.814159	4.218620

Data Dictionary

- 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 is 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)

Statistical Summary of Data

|	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Ticker Symbol	340	340	AAL	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Security	340	340	American Airlines Group	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
GICS Sector	340	11	Industrials	53	NaN	NaN	NaN	NaN	NaN	NaN	NaN
GICS Sub Industry	340	104	Oil & Gas Exploration & Production	16	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Current Price	340.0	NaN	NaN	NaN	80.862345	98.055086	4.5	38.555	59.705	92.880001	1274.949951
Price Change	340.0	NaN	NaN	NaN	4.078194	12.006338	-47.129693	-0.939484	4.819505	10.695493	55.051683
Volatility	340.0	NaN	NaN	NaN	1.525976	0.591798	0.733163	1.134878	1.385593	1.695549	4.580042
ROE	340.0	NaN	NaN	NaN	39.597059	96.547538	1.0	9.75	15.0	27.0	917.0
Cash Ratio	340.0	NaN	NaN	NaN	70.023529	90.421331	0.0	18.0	47.0	99.0	958.0
Net Cash Flow	340.0	NaN	NaN	NaN	55537620.588235	1946365312.175789	-11208000000.0	-193696500.0	2098000.0	169810750.0	20764000000.0
Net Income	340.0	NaN	NaN	NaN	1494384602.941176	3940150279.327936	-23528000000.0	352301250.0	707336000.0	1899000000.0	24442000000.0
Earnings Per Share	340.0	NaN	NaN	NaN	2.776662	6.587779	-61.2	1.5575	2.895	4.62	50.09
Estimated Shares Outstanding	340.0	NaN	NaN	NaN	577028337.75403	845849595.417695	27672156.86	158848216.1	309675137.8	573117457.325	6159292035.0
P/E Ratio	340.0	NaN	NaN	NaN	32.612563	44.348731	2.935451	15.044653	20.819876	31.764755	528.039074
P/B Ratio	340.0	NaN	NaN	NaN	-1.718249	13.966912	-76.119077	-4.352056	-1.06717	3.917066	129.064585

EDA - UNIVARIATE(Current Price & Price Change)

- The current price feature's mean is larger than it's median, and it can be seen that it is right skewed.

- The current price feature has outliers that may have to be treated.

- The price change feature has a relatively normal distribution with similar mean and median values.

- The price change feature has outliers on either whisker, which may have to be treated.
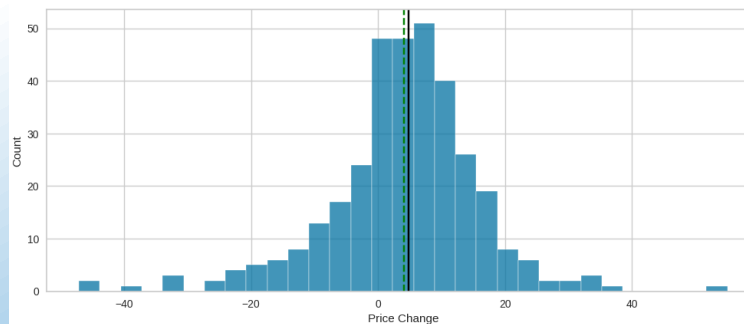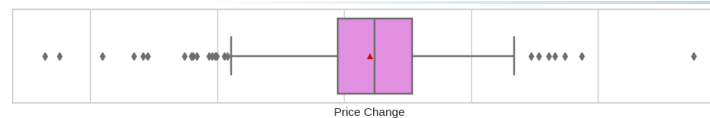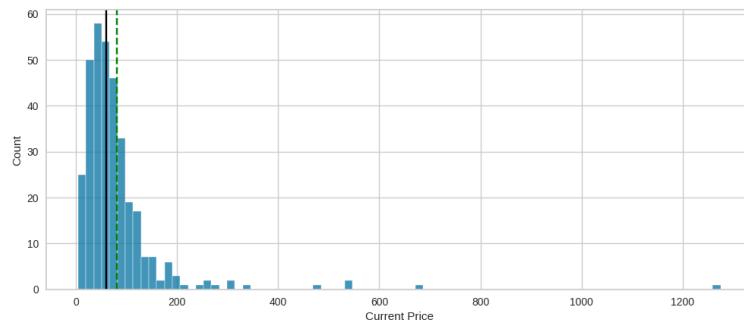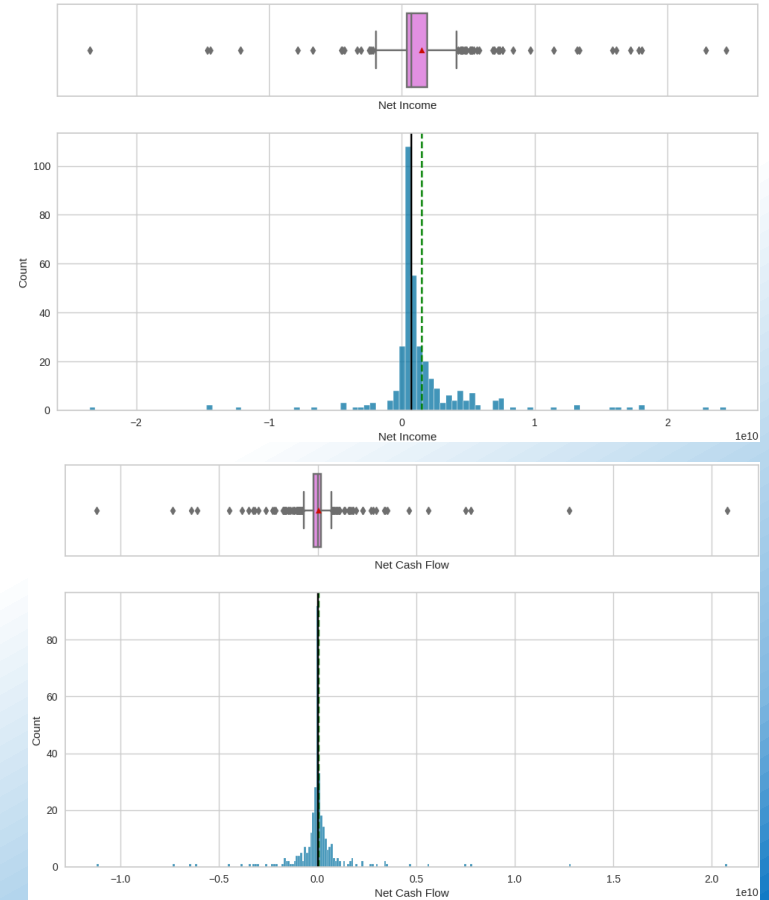
EDA – UNIVARIATE(P/E & P/B Ratios)

- The P/E Ratio's mean is a lot larger than it's median, and it can be seen that it is right skewed.

- The P/E Ratio has outliers that may have to be treated.

- The P/B Ratio has a relatively normal distribution with similar mean and median values.

- The P/B Ratio has outliers on either whisker, which may have to be treated.
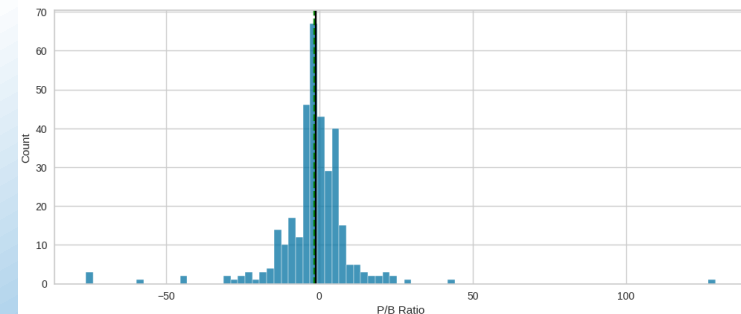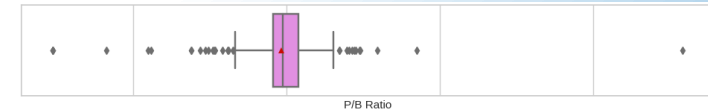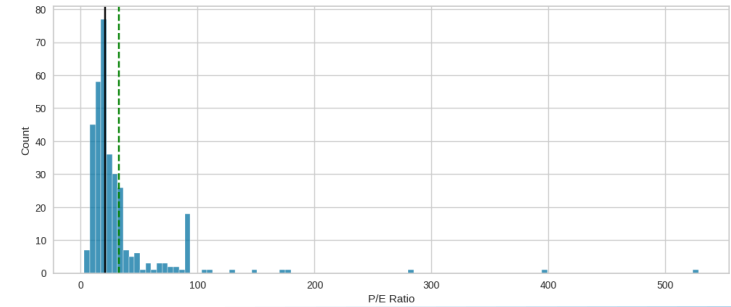
EDA – UNIVARIATE(Earnings/Share & Shares Outstanding)

- The Earnings per Share feature has a relatively normal distribution with similar mean and median values.

- The Earnings per Share feature has outliers on either whisker, which may have to be treated.

- The Estimated Shared Outstanding feature's mean is larger than it's median, and it can be seen that it is right skewed.

- The Estimated Share Outstanding feature has outliers that may have to be treated.
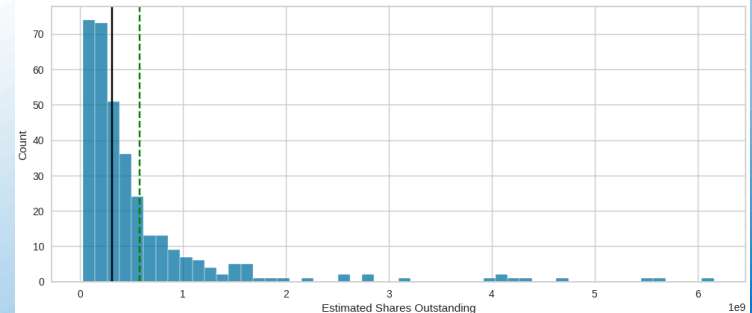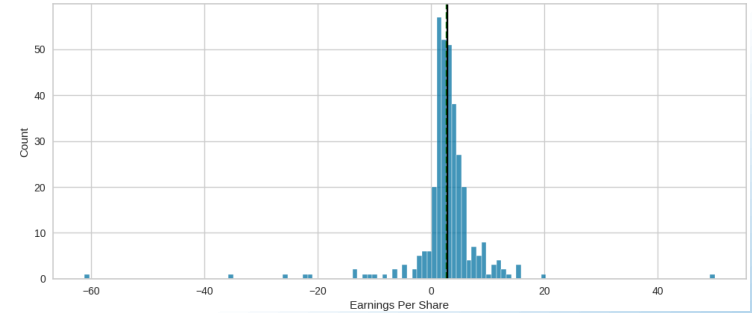
EDA – UNIVARIATE(Net Income & Net Cash Flow)

- The Net Income feature's mean is larger than it's median, and it can be seen that it is slightly right skewed.

- The Net Cash Flow feature has a relatively normal distribution with similar mean and median values.

- The Net Cash Flow and Net Income features have outliers on either whisker, which may have to be treated.

EDA – UNIVARIATE(Cash Ratio, ROE & Volatility)

- The Cash Ratio, ROE and Volatility features all have a right skewed distribution with a larger mean than median.

- All three features contain outliers above the right whisker, which may have to be treated.
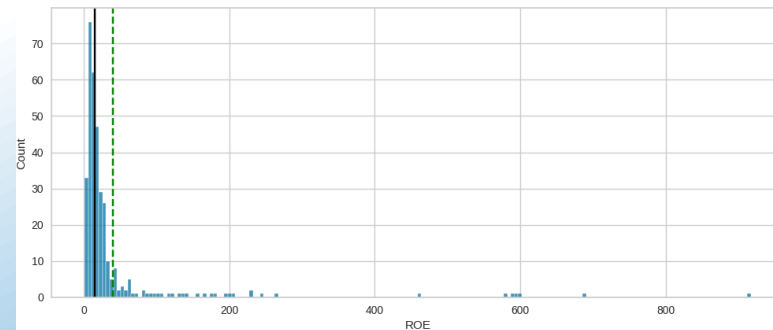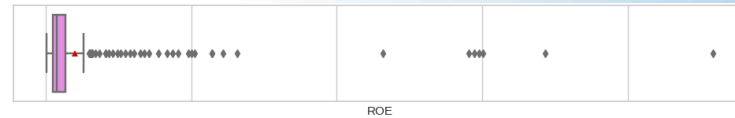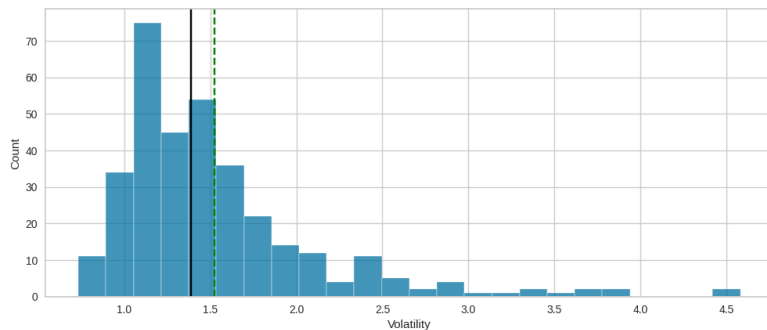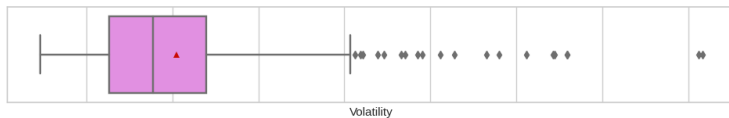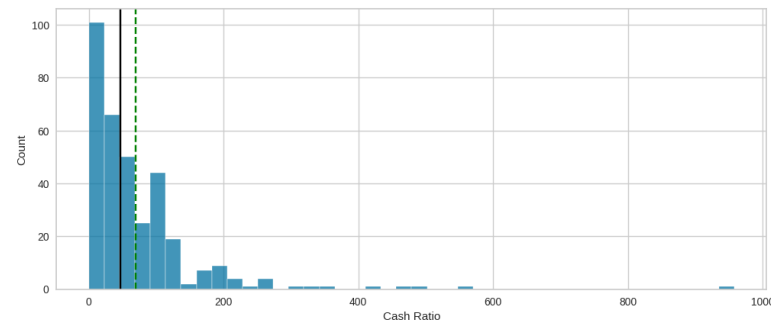
EDA – UNIVARIATE(GICS Sectors)

- 15.6% of the companies are in the Industrials sector.

- Only 1.5% of companies are in the Telecommunications Services sector.

- More than 50% of companies are in the financials, health care, industrials or consumer discretionary sectors.

- Less than 10% of companies are Information Tech companies.
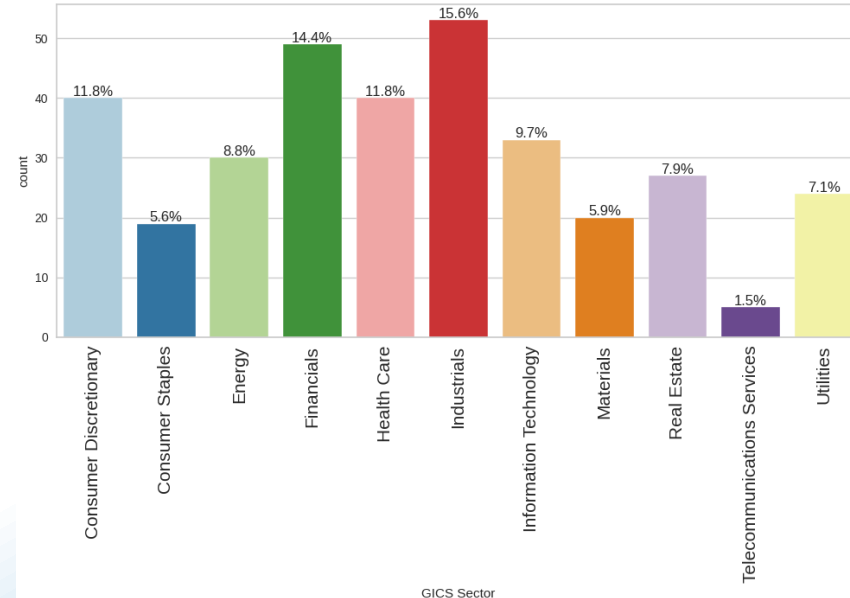
EDA – BIVARIATE(Correlation Matrix)

- Net Income & Estimated Shares Outstanding have the largest positive correlation(0.59), followed by Net Income & Earnings per Share(0.56) which makes sense.

- Current Price & Earnings Per Share have the 3rd highest positive correlation: 0.48.

- The largest negative correlation(-0.41) is between Volatility & Price Change, and between Earnings per Share & ROE.

- Earnings Per Share seems to be influencing, or affected by, other features the most.
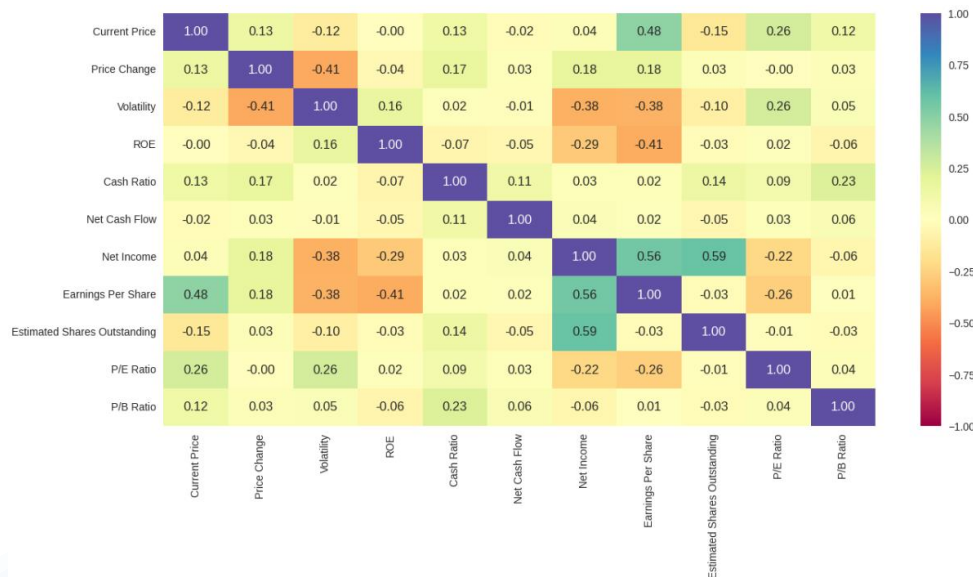
EDA – BIVARIATE(Price Change & Cash Ratio across GICS Sectors)

- The Health Care sector witnesses the highest price change on average: around ~$9.

- The Energy sector is the only one with a negative price change on average: ~-$11.

- The Utilities sector witnesses the least positive price change on average: ~$1.

- Companies in the IT Sector have the highest cash ratio: ~145. Companies in the Utilities Sector have the least cash ratio on average: ~15. All companies have a positive cash ratio.

- The sector with the highest price change(Energy sector ~-$11 on average) has a cash ratio of ~50.

Price Change for each economic sector
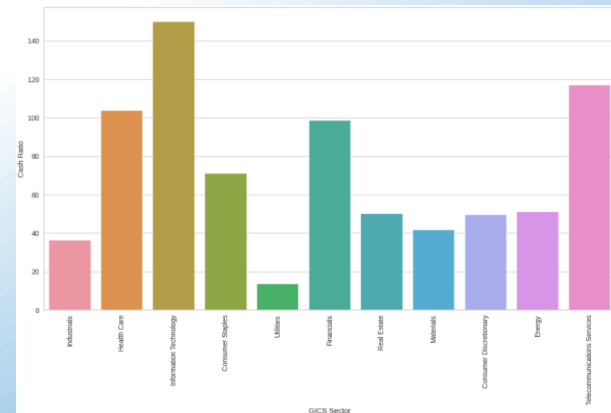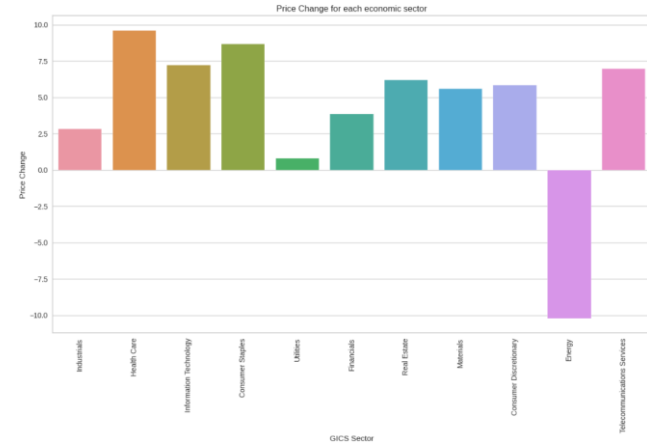
EDA – BIVARIATE(P/E Ratio & Volatility across GICS Sectors)

- Companies in the energy sector have the highest stock price to earnings per share ratio(>70). Note that the energy sector also has the highest price change on average: ~-$11.

- Companies in the Telecommunication Services sector have the lowest stock price to earnings per share ratio: ~16.

- Considering the Energy sector has the largest price change(~-$11), it also has the highest volatility: >2.5

- Companies in the Utilities and Consumer Staples sectors have the lowest volatility: ~1.2 & ~1.1 respectively.
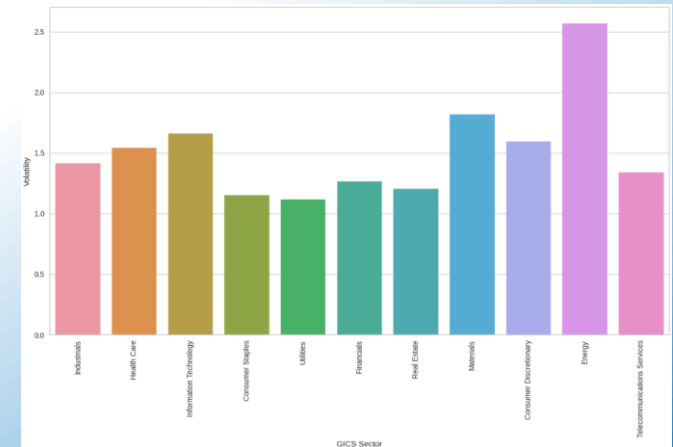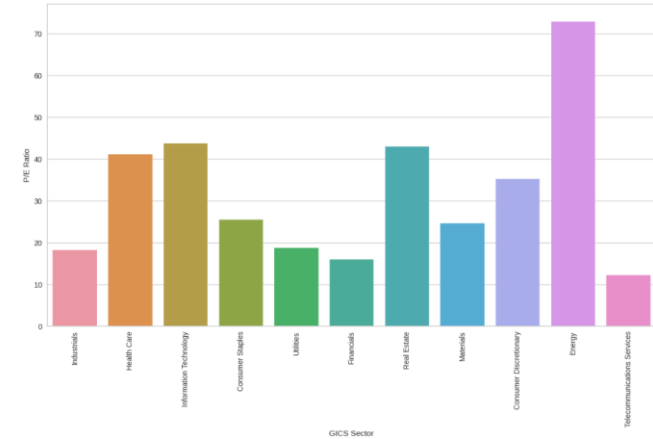
DATA PRE-PROCESSING – OUTLIER VALUES & SCALED DATA

- As seen in the EDA section, all the numerical columns contain outliers. They are all real data and will not be capped or floored.

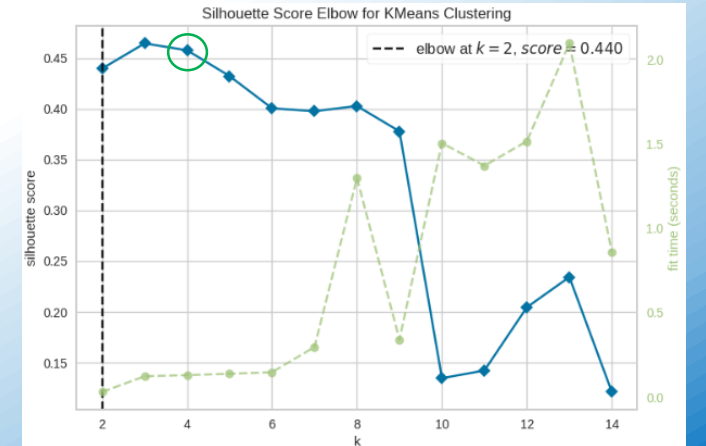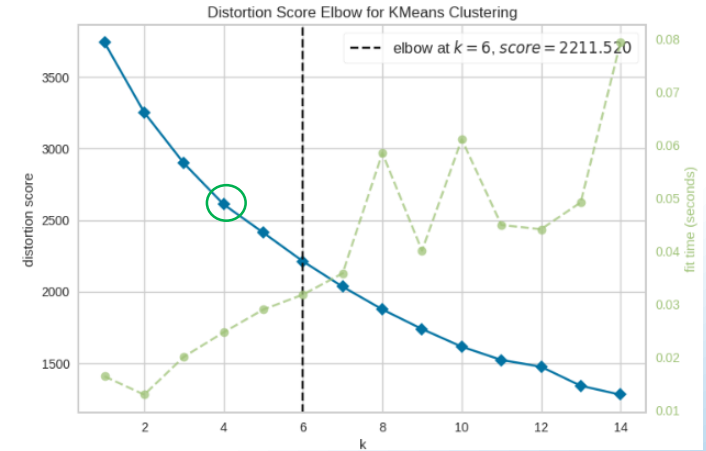- All the numerical features are scaled using the z-score, and now have a mean ~ 0 and standard deviation ~ 1.
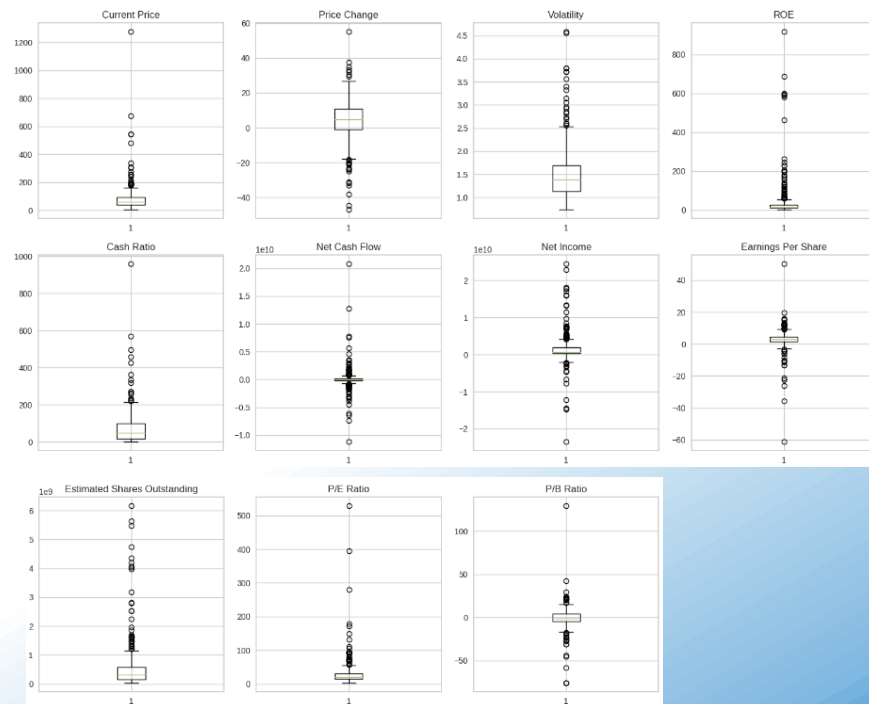
Stat Summary of Scaled Data

|	count	mean	std	min	25%	50%	75%	max
Current Price	340.0	7.836868e-18	1.001474	-0.779918	-0.432101	-0.216088	0.122741	12.195670
Price Change	340.0	-2.873518e-17	1.001474	-4.271357	-0.418535	0.061834	0.551963	4.251806
Volatility	340.0	-7.967483e-17	1.001474	-1.341642	-0.661838	-0.237564	0.286960	5.168258
ROE	340.0	2.873518e-17	1.001474	-0.400362	-0.309599	-0.255142	-0.130667	9.101176
Cash Ratio	340.0	-8.098097e-17	1.001474	-0.775555	-0.576194	-0.255000	0.320933	9.834905
Net Cash Flow	340.0	2.612289e-18	1.001474	-5.795489	-0.128348	-0.027497	0.058798	10.655237
Net Income	340.0	1.044916e-17	1.001474	-6.359977	-0.290285	-0.200045	0.102842	5.832629
Earnings Per Share	340.0	5.747037e-17	1.001474	-9.725729	-0.185337	0.017990	0.280224	7.192572
Estimated Shares Outstanding	340.0	-1.044916e-16	1.001474	-0.650430	-0.495119	-0.316542	-0.004630	6.609320
P/E Ratio	340.0	7.836868e-18	1.001474	-0.670162	-0.396715	-0.266300	-0.019145	11.187619
P/B Ratio	340.0	2.612289e-17	1.001474	-5.334786	-0.188853	0.046685	0.404071	9.377562

K-MEANS CLUSTERING – Elbow Method & Silhouette Score

- On the Elbow Curve Plot, the initial "elbow" shape occurs at k = , shown in the green circle.

- K = 4 has a quick fit time: ~ 0.025 s.

- The Silhouette score at K = 4 is the second highest for all k-values at ~ 0.46 shown in the green circle.

- 4 is the chosen number of clusters for the final model considering the Elbow Curve Plot, Silhouette Scores and the Fit Times.

- The Silhouette plots for K=(4,7) is shown on the next slide for reference.

Distortion Score Elbow for KMeans Clustering

--- elbow at $k = 6$, $score = 2211.520$

Silhouette Score Elbow for KMeans Clustering

--- elbow at $k = 2$, $score = 0.440$

# K-MEANS CLUSTERING - Silhouette Plots for K = 4,5,6,7
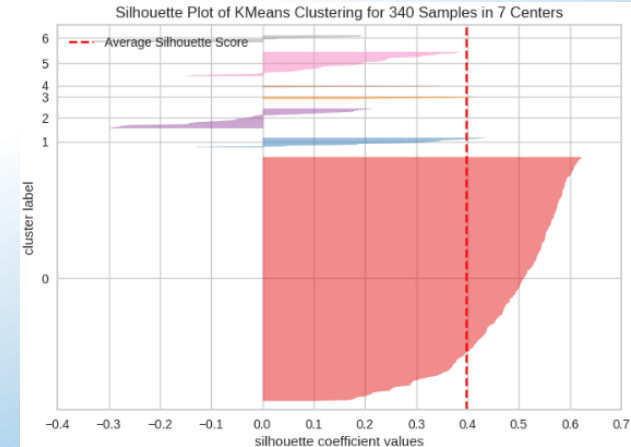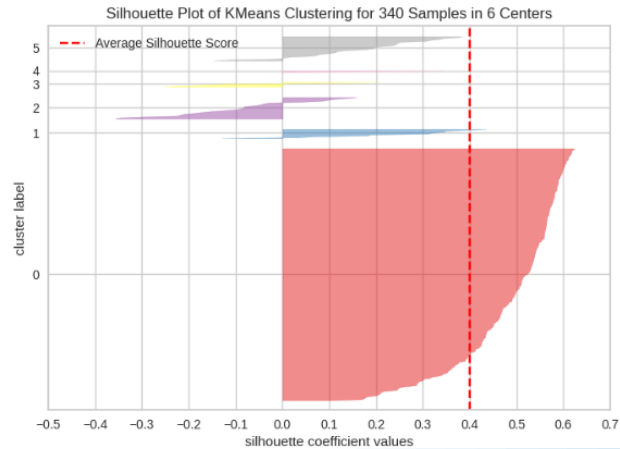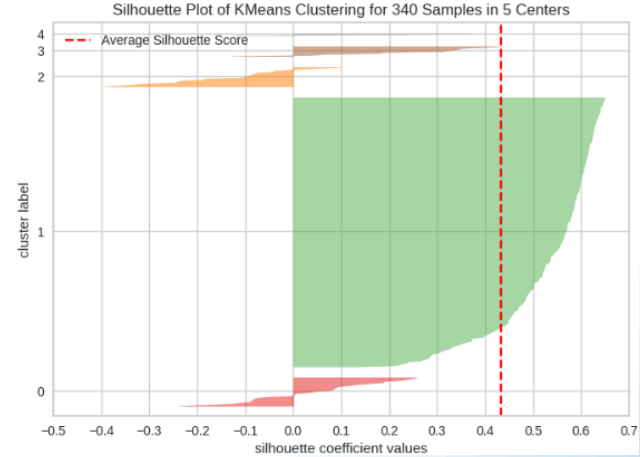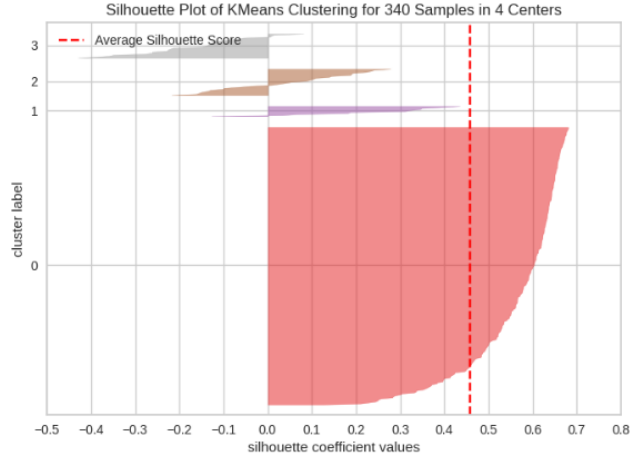
K-MEANS CLUSTERING – Final Model Features

- **Cluster 0**: This cluster contains only 11 companies and has the maximum net income and estimated shares outstanding, suggesting a substantial financial scale and potential influence in the market. Additionally, it boasts the lowest price-to-book (P/B) ratio, indicating potential undervaluation relative to the company's assets.

- **Cluster 1**: Among the clusters, this one stands out with the highest volatility, signifying significant price fluctuations. It also features the highest Return on Equity (ROE), reflecting strong profitability potential. Moreover, this cluster has the highest price-to-earnings (P/E) ratio, potentially indicating an optimistic market sentiment.

- **Cluster 2**: Notably, this cluster demonstrates the highest cash ratio, implying a strong liquidity position. It showcases the highest net cash flow, indicating substantial positive cash generation. Additionally, Cluster 2 has the highest price-to-book (P/B) ratio, current price, and price change, suggesting potentially strong investor confidence and positive market performance.

- **Cluster 3**: This cluster contains almost 80% of the companies(270), and showcases notable features such as a positive average earnings per share (EPS), suggesting a potential for earnings growth. The moderate Return on Equity (ROE) indicates a reasonable level of profitability. Additionally, the moderate price change and current price suggest relatively stable price dynamics in comparison to the other clusters.

Feature values for each cluster

KM_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	50.517273	5.747586	1.130399	31.090909	75.909091	-1072272727.272727	14833090909.090910	4.154545	4298826628.727273	14.803577	-4.552119	11
1	36.253627	-14.482619	2.867901	96.900000	48.533333	-178762300.000000	-3469565300.000000	-8.434000	501463317.368000	85.261572	1.404728	30
2	203.769424	12.981809	1.744529	23.896552	270.517241	1200766344.827586	1396729068.965517	5.032414	609604308.507586	69.869318	13.346229	29
3	73.854019	5.116180	1.369515	35.262963	50.637037	4512725.925926	1512994744.444444	3.723870	430296731.305963	23.486573	-3.567829	270

K-MEANS CLUSTERING - Final Model Box-Plots

Boxplot of numerical variables for each cluster
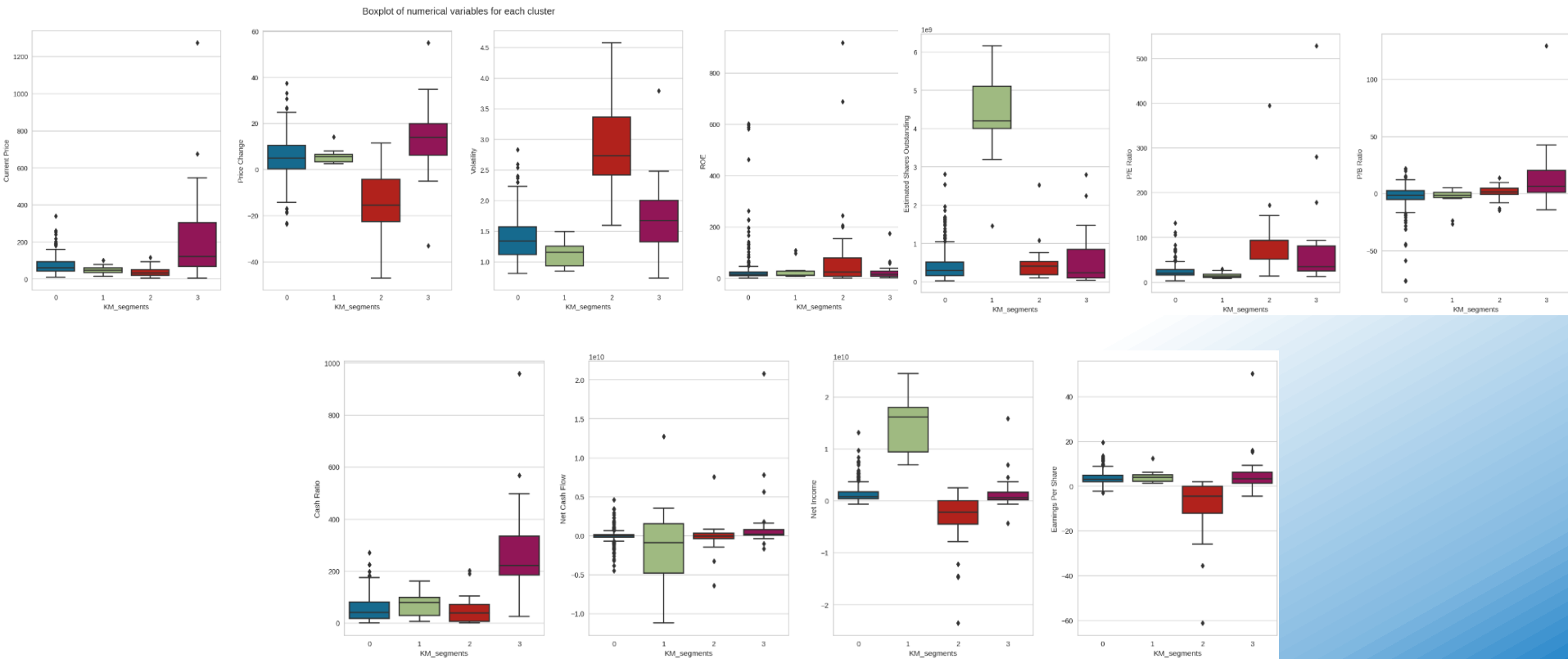
K-MEANS CLUSTERING - BUSINESS RECOMMENDATIONS

- **Diversification Strategy**: Encourage investors to consider companies from Cluster 2 for diversification purposes. The cluster showcases a strong liquidity position, positive net cash flow, and high price-to-book (P/B) ratios. Investors could potentially benefit from these financially robust companies to mitigate risk and enhance portfolio stability.

- **High-Risk, High-Reward Allocation**: For investors seeking potentially high rewards and willing to bear higher risk, Cluster 1 might be appealing. The cluster has companies with the highest volatility and Return on Equity (ROE), indicating strong profitability prospects. However, investors should exercise caution due to the high volatility.

- **Value Investing Opportunities**: Highlight Cluster 0 to investors interested in value investing. This cluster contains companies with potential undervaluation based on their lowest price-to-book (P/B) ratios. Investors could explore these companies for potential growth considering their substantial net income and estimated shares outstanding.

- **Stable Growth Holdings**: Recommend Cluster 3 to investors looking for stable growth. With a majority of companies from various sectors, this cluster displays moderate price change and current prices. These companies might offer consistent growth potential without extreme fluctuations.

- **Long-Term Sustainability**: Suggest investors explore companies in Cluster 3 with high average Return on Equity (ROE) and positive earnings per share (EPS). These metrics reflect strong profitability and potential for sustained growth, making them attractive for long-term investment strategies.

Hierarchical Clustering – Cophenetic Correlations

- Hierarchal Clustering is done using 4 different distance metrics: Euclidian, Chebyshev, Mahalanobis and City Block. Each of these metrics is linked using 4 linkage methods: single, complete, average and weighted.

- The cophenetic correlation is computed for each combination of distance metric & linkage method(Table 1).

- The "average" linkage method has the highest cophenetic correlation = 0.95 on the complete linkage method through the Euclidian distance.

- The City Block distance metric captures outliers well which makes it a good algorithm to use for the data. Better variability in the clusters may be achieved through the Euclidian distance metric, so both metrics will be analyzed with Dendrograms in the next few slides.

Table 1

Distance Metric	Linkage Method	Cophenetic Correlation
Euclidian	Single	0.93
	Complete	0.60
	Average	0.95
	Weighted	0.89
Chebyshev	Single	0.91
	Complete	0.81
	Average	0.93
	Weighted	0.91
Mahalanobis	Single	0.93
	Complete	0.85
	Average	0.94
	Weighted	0.89
City Block	Single	0.93
	Complete	0.77
	Average	0.92
	Weighted	0.88

Hierarchical Clustering – Dendrograms(Euclidian)
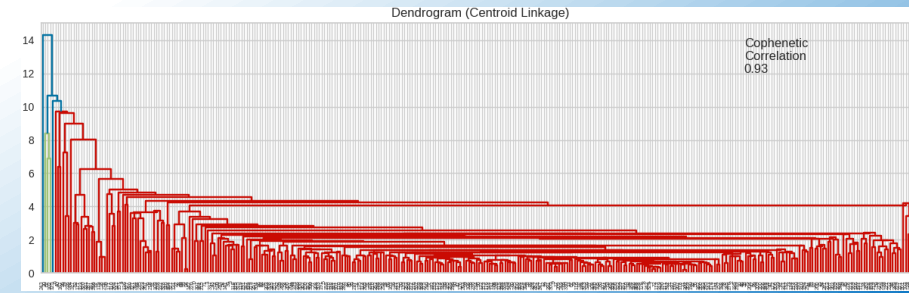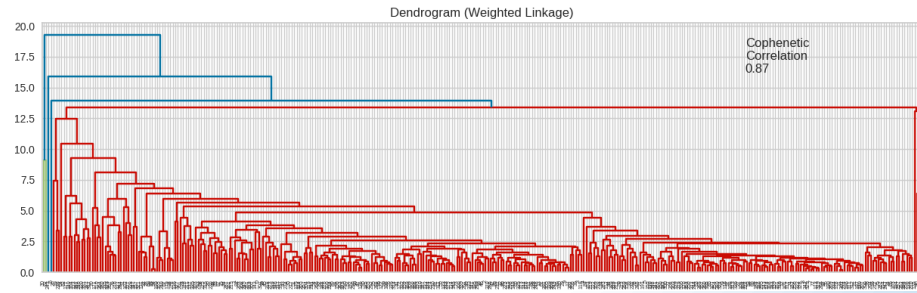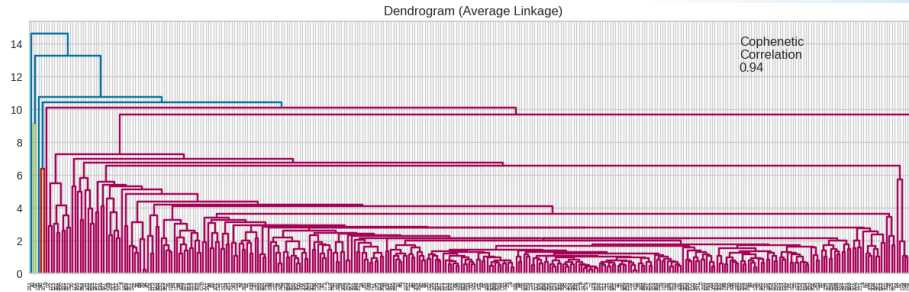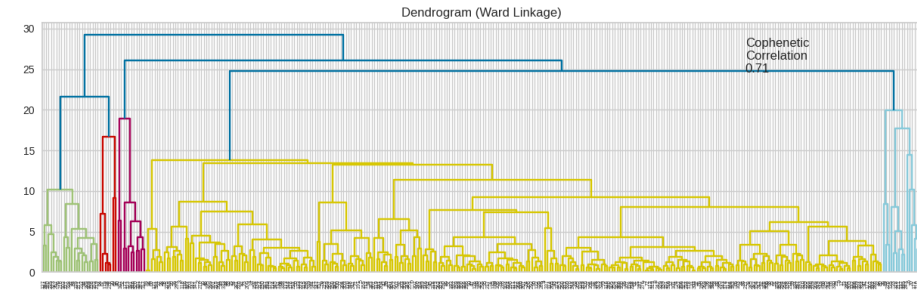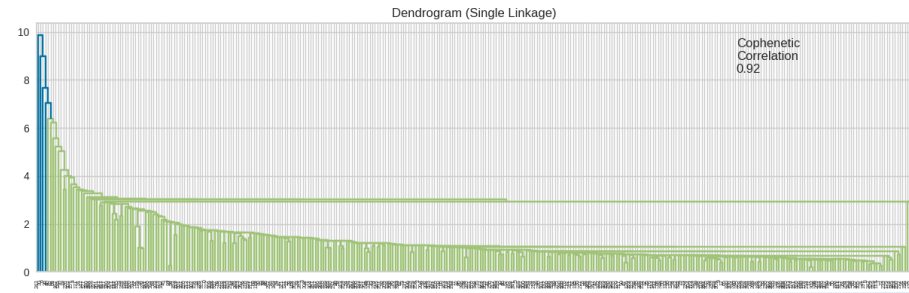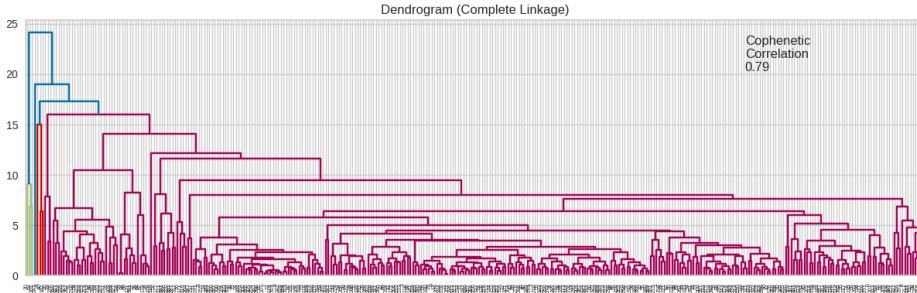
Hierarchical Clustering – Dendrograms(Manhattan)
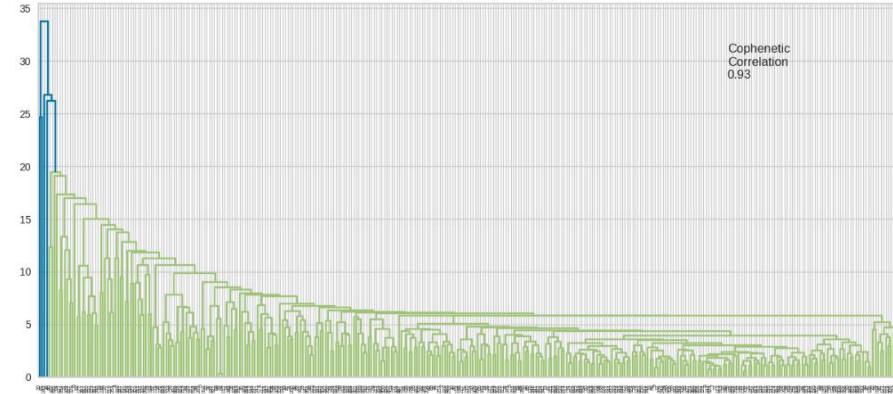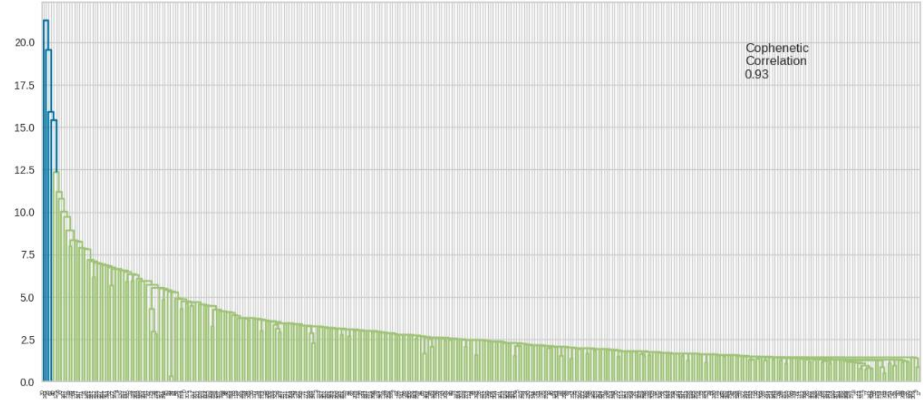
Hierarchical Clustering - Cluster Profiles(Euclidian v/s Manhattan)

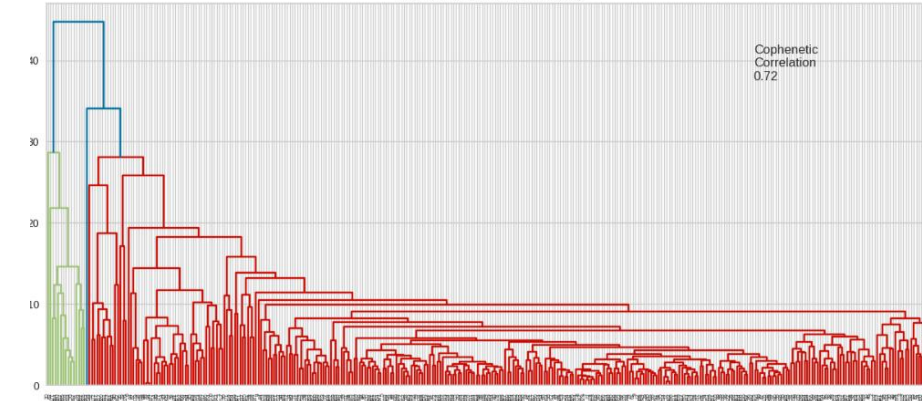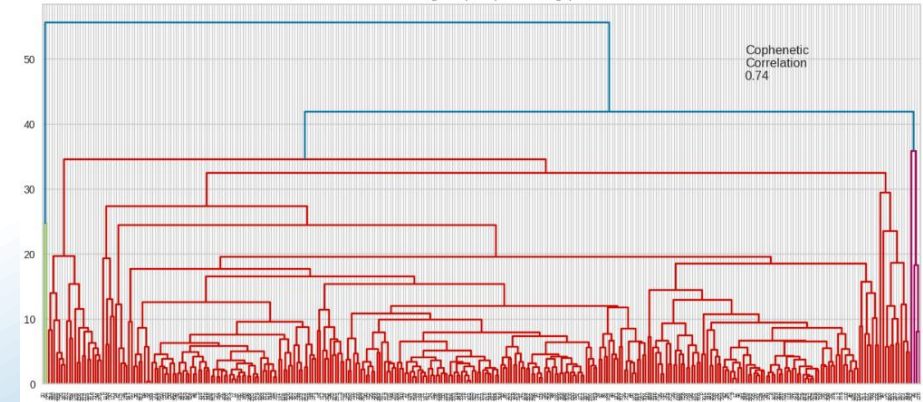| Euclidian Distance Clustering Profiles (5 Clusters)		
Linkage Method(Cophenetic Coeff.)	Cluster*	# of Companies
Ward(0.71) **Final Model**	0	15
	1	7
	2	11
	3	285
	4	22
Complete(0.79)	0	55
	1	2
	2	279
	3	1
	4	3
Single(0.92)	0	336
	1	1
	2	1
	3	1
	4	1
Average(0.94)	0	335
	1	2
	2	1
	3	1
	4	1

Cluster Variability

Manhattan Distance Clustering Profiles(5 Clusters)		
Linkage Method(Cophenetic Coeff.)	Cluster	# of Companies
Complete(0.75)	0	313
	1	22
	2	3
	3	1
	4	1
Single(0.93)	0	336
	1	1
	2	1
	3	1
	4	1
Average(0.93)	0	336
	1	1
	2	1
	3	1
	4	1

No Variability

*5 Clusters seems to be the best choice of clusters based on the Dendrogram. It also gives good variability. Please note that Trade & Ahead could use more or less clusters and test the results.

- The Manhattan distance measure captures the outliers well, but the Euclidian distance measure provides a lot more variability in clustering, which would help investors more in making informed decisions.

Hierarchical Clustering – Final Model Features

- **Cluster 0 (25 companies):** Characterized by moderate price decrease (11.51%), high volatility (2.74), and healthy ROE (51.08%). Despite the volatility, their stable profitability and moderate cash ratio (50.04) offer potential.

- **Cluster 1 (12 companies):** Boasting a high current price (325.99) and positive price change (7.72%), these companies exhibit potential growth. The low P/E (41.12) and P/B (25.26) ratios suggest undervaluation.

- **Cluster 2 (11 companies):** Featuring low current price (42.85) and moderate price change (6.27%), this cluster offers potential value investments. Positive net cash flow counterbalances the negative P/B (-4.92), signaling financial strength.

- **Cluster 3 (285 companies):** Moderate current price (72.76) and price change (5.21%) signify stability. Balanced ROE (25.60%) and P/E (24.72), despite negative P/B (-2.65), indicate diverse investment prospects.

- **Cluster 4 (7 companies):** With remarkably high ROE (633.57%) and negative EPS (-10.84), caution is advised due to translating ROE to per-share earnings. Negative P/B (-11.59) could suggest overvaluation concerns.

Feature values for each cluster

HC_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	71.308401	-11.514513	2.736509	51.080000	50.040000	-332305120.000000	-2747214160.000000	-6.954400	496655257.693600	123.407703	0.097734	25
1	325.996105	7.724708	1.545762	17.000000	360.333333	186503166.666667	1008736416.666667	9.544167	589434517.801667	41.121871	25.258283	12
2	42.848182	6.270446	1.123547	22.727273	71.454545	558636363.636364	14631272727.272728	3.410000	4242572567.290909	15.242169	-4.924615	11
3	72.760400	5.213307	1.427078	25.603509	60.392982	79951512.280702	1538594322.807018	3.655351	446472132.228456	24.722670	-2.647194	285
4	84.355716	3.854981	1.827670	633.571429	33.571429	-568400000.000000	-4968157142.857142	-10.841429	398169036.442857	42.284541	-11.589502	7

Hierarchical Clustering – Final Model Box-Plots

Boxplot of numerical variables for each cluster
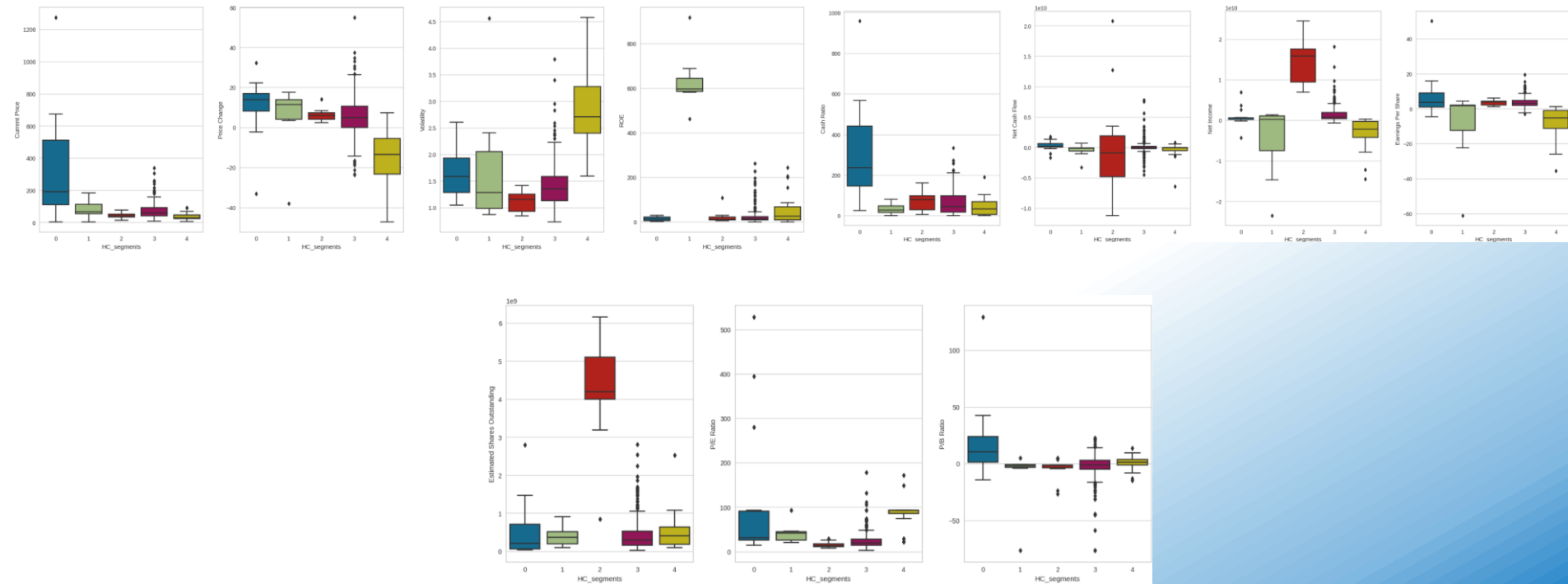
HIERARCHICAL CLUSTERING – BUSINESS RECOMMENDATIONS

- **Customized Diversification Plans**: Tailor investment portfolios for clients by strategically allocating funds across clusters. Combine stable growth from Cluster 3, high-reward potential from Cluster 1, and value opportunities from Cluster 2 to achieve diversification that aligns with client risk profiles.

- **Long-Term Value Hunting**: Advise clients to explore companies in Cluster 2 for potential undervalued opportunities. Focus on firms with strong cash generation and positive net cash flow, utilizing their financial stability to identify long-term value

- **Value-Based Growth Strategy**: Recommend clients consider Cluster 1 for undervalued growth prospects. Companies with high current prices, positive price changes, and low P/E and P/B ratios offer attractive value-based growth potential.

- **Balanced and Stable Portfolios**: Suggest a balanced approach by including Cluster 0's companies with moderate price changes and healthy ROE. Combine them with stable options from Cluster 3 to create portfolios that balance risk and return.

- **High-Growth Caution**: Caution clients interested in Cluster 4 due to exceptionally high ROE but negative EPS. Encourage thorough analysis before investing, focusing on aligning high profitability with sustainable earnings to avoid high-risk scenarios.

K-MEANS v/s HIERARCHICAL CLUSTERING

- K-means clustering took lesser time to execute than Hierarchical clustering.

- The K-means Cluster contained more distinct clusters than the Hierarchical one.

- 4 Clusters are generated in the K-Means method while 5 are generated in the Hierarchal one.

- Majority of companies do fall into the same cluster in both clustering techniques.

Trade & Ahead should consider K-means clustering when seeking a faster execution time and clear, predefined clusters. It's suitable when distinct clusters are anticipated and computational efficiency is crucial. On the other hand, if flexibility in the number of clusters and capturing diverse shapes is paramount, Hierarchical Clustering is recommended. This method allows for more nuanced interpretations, accommodating varying cluster structures. The choice should align with the data's characteristics and the firm's goals for insights and portfolio optimization.

THANK YOU

Arpan Dinesh