

Business Presentation

Trade&Ahead





Contents

1. Overview and Business Problem
2. Solution approach
3. Data
4. Model overview and performance summary
5. Key Findings and insights which can drive business decisions
6. Business recommendations

Overview and Business problem

The stock market has traditionally provided investors with an opportunity to invest in equities (securities) to provide a hedge against inflation and a wealth model. Traditional rates of returns approach 10% per year on average. As a professional consultancy firm, Trade&Ahead is involved in a aggressive and competitive environment to exceed these returns while reducing the risk associated with the investments of their clients.

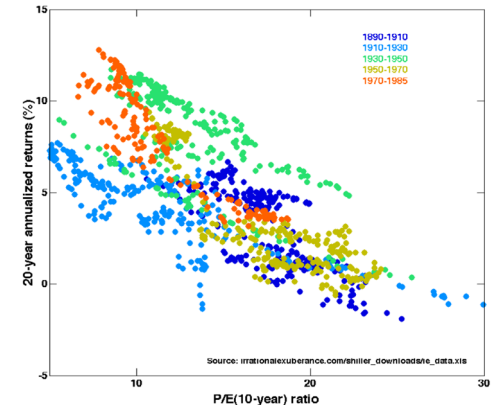
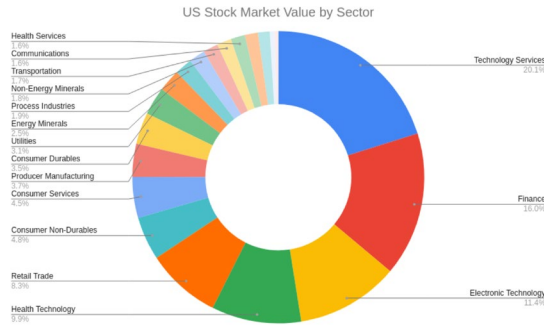
A popular indices – the S & P 500 is generally considered by many investors to be the “market”. The annualized average return is approximately 10.5%. Our objective is to exceed this so our investors will not simply invest in a “no load” fund such as the S & P 500 or the Russell 2000 or another competitor. Thus we need to outpace this investment consistently.

S&P 500 Historical Price Chart



Business Problem- overview and description of the problem

- There are 340 securities we can choose from in order to develop our strategy.
- Individual investors require individual plans , so flexibility is important
- Economic conditions can be volatile , so any plan should also consider not only the potential for gain, but the potential for short term losses.
- Economic seasonality can bring some types of businesses into vogue today, and out tomorrow. Our calculus should consider the peaks and valleys depending on the customers desires and risk tolerance (beta)



Solution - Approach

- As a consultancy to Trade@Ahead, we can employ unsupervised machine learning tools such as K-Means Cluster Analysis and Hierarchical Cluster Analysis to provide insight into the Securities to determine performers, non-performers and trend companies.
- We will utilize both techniques to see which model is optimized for the problem.
- By clustering the securities , we hope to see where there are consistent trends which we can capitalize on. Clustering will also increase our dimensionality and allow us to observe hidden observations that we can hopefully capitalize on.

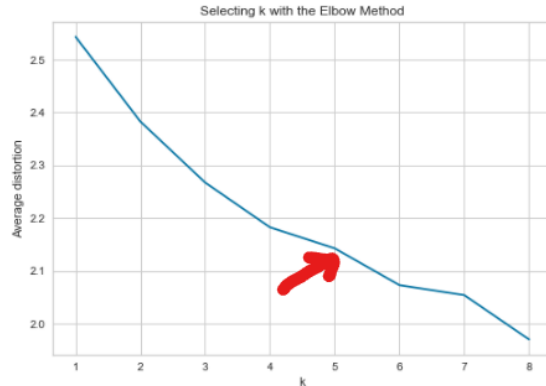
Data Overview

- * 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)

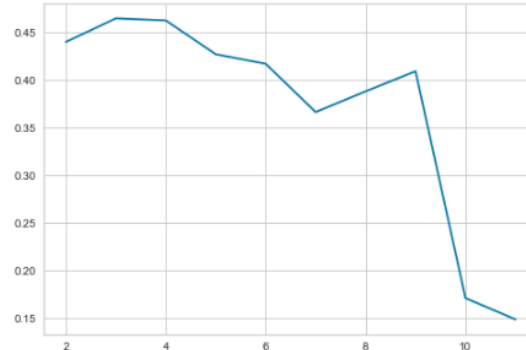
Model Overview and Performance Summary

- K-Means Clustering- selection of the number of clusters. We utilized 2 methods, “the elbow method” and the silhouette score methodology. In our analysis we determined the elbow method was optimal elbow versus the utilization of silhouette. Primarily due to the low silhouette scores comparatively.

Number of Clusters: 1	Average Distortion: 2.5425069919221697
Number of Clusters: 2	Average Distortion: 2.382318498894466
Number of Clusters: 3	Average Distortion: 2.2668604252073914
Number of Clusters: 4	Average Distortion: 2.1822426161442072
Number of Clusters: 5	Average Distortion: 2.1423948645166737
Number of Clusters: 6	Average Distortion: 2.072769863902725
Number of Clusters: 7	Average Distortion: 2.053883804477707
Number of Clusters: 8	Average Distortion: 1.9700587501311286

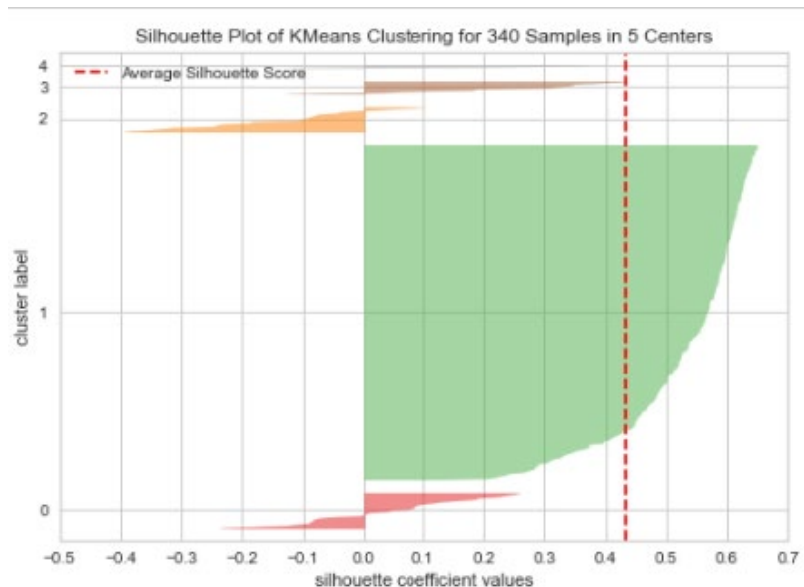


For n_clusters = 2, the silhouette score is 0.43969639509980457)
For n_clusters = 3, the silhouette score is 0.4644405674779404)
For n_clusters = 4, the silhouette score is 0.4622558112748279)
For n_clusters = 5, the silhouette score is 0.4267713416628856)
For n_clusters = 6, the silhouette score is 0.4168471220569999)
For n_clusters = 7, the silhouette score is 0.36616981702277573)
For n_clusters = 8, the silhouette score is 0.38791289243541743)
For n_clusters = 9, the silhouette score is 0.4090171173372503)
For n_clusters = 10, the silhouette score is 0.17112555407721344)
For n_clusters = 11, the silhouette score is 0.1488657653384614)



Model Overview and Performance Summary

- K-Means Clustering(cont.)
- Optimization evaluation we observed that a majority of the observations were within the mean silhouette plotting.



Model Overview and Performance Summary

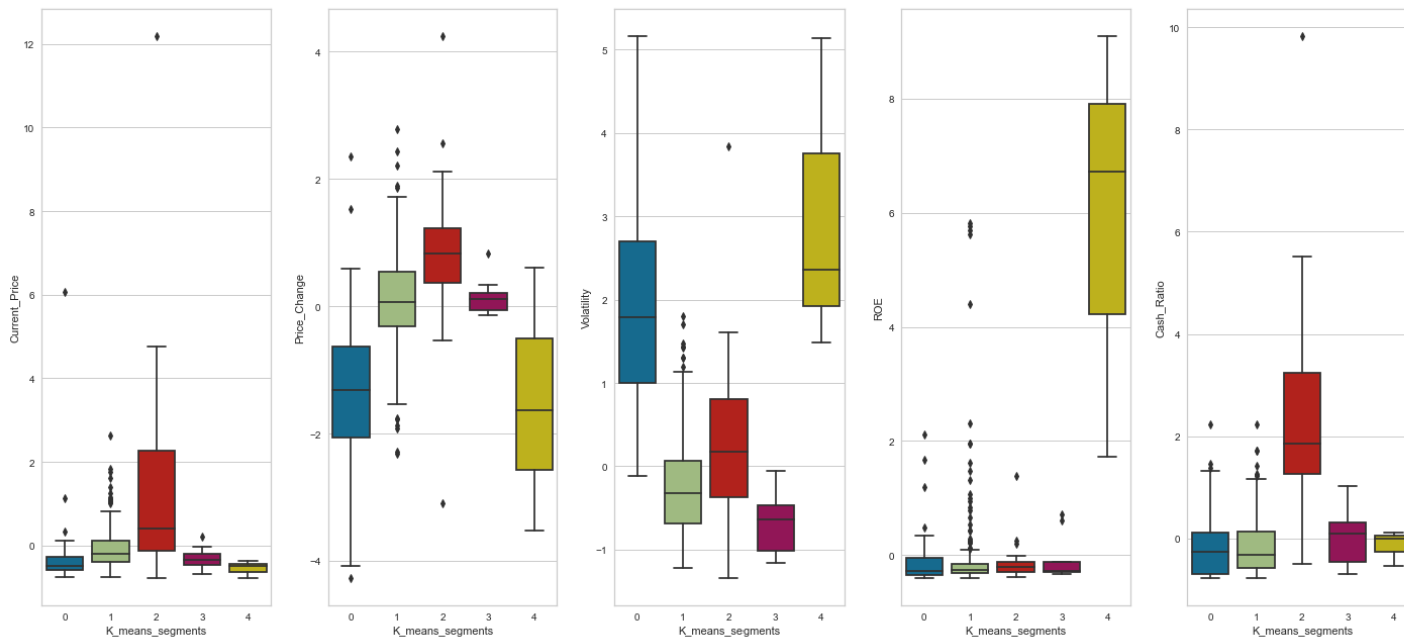
- K-Means(Cont.)
- Breakdown of the 5 clusters and their characteristics.
- Of note – we observe cluster 2 and 3 and 4 as critical components in the analysis

	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_segments
K_means_segments												
0	65.174668	-11.542247	2.690220	37.300000	65.366667	195008366.666667	-1677736033.333333	-4.401667	544473664.718000	113.488924	1.424161	30
1	72.738269	5.179897	1.380738	34.825455	53.138182	-10147287.272727	1488641570.909091	3.636164	437961614.918582	23.680917	-3.395254	275
2	233.251108	13.682869	1.719008	29.333333	296.523810	1398716380.952381	1835686380.952381	7.126190	508721791.962857	37.805996	16.758218	21
3	50.517273	5.747586	1.130399	31.090909	75.909091	-107227272.727272	14833090909.090910	4.154545	4298826628.727273	14.803577	-4.552119	11
4	26.990000	-14.060688	3.296307	603.000000	57.333333	-585000000.000000	-1755566666.666668	-39.726667	481910081.666667	71.528835	1.638633	3

Model Overview and Performance Summary

- K-Means(Cont.)

Boxplot of scaled numerical variables for each cluster



Model Overview and Performance Summary

○ K-Means(Cont.)

Insights

- **Cluster 0:**
 - This cluster has 147 Securities.
 - This cluster's current price has the second highest IQR
 - Ranks 2nd for positive associated price change
 - Ranks 4th for volatility
 - Ranking is 2nd for ROE per IQR, however it should be noted that with the exception of cluster 3, these all tend to be very closely aligned.
 - Ranks 3rd for cash ratio
- **Cluster 1:**
 - Cluster has 124 Securities
 - Ranks 3rd for current price
 - Ranks 4th for price change
 - Ranks 3rd for volatility
 - Ranks along with cluster 0 and others , but has several outliers on the top end that should be investigated
 - Ranks number 2 for cash ratio
- **Cluster 2:**
 - Cluster has 24 Securities
 - 1st in current price
 - 1st in positive price change
 - 2nd in volatility
 - Ranks along with the others in ROE with no real separation
 - Ranks number 2 by IQR for Cash ratio, however ranks 4th for 50% - mixed results
- **Cluster 3:**
 - Cluster has 34 Securities
 - 4th in current price
 - Last in change in current price- with large range
 - 1st in volatility - large range
 - 1st in ROE- large range of IQR
 - Last in cash ratio
- **Cluster 4:**
 - Cluster has 11 Securities
 - Last in current price
 - 3rd in positive price change with a tight range
 - Least volatile of the group
 - In the general group with the others for ROE
 - 2nd in cash ratio by median

Model Overview and Performance Summary

- Hierarchical Clustering- using Cophenetic correlation evaluation of Euclidean, Chebyshev, Mahalanobis and Cityblock (Manhattan) distance to find the highest coefficient to help select the proper technique.
- Euclidean distance and average linkage was the highest cophenetic correlation

Cophenetic correlation for Euclidean distance and single linkage is 0.9285583986172286.
Cophenetic correlation for Euclidean distance and complete linkage is 0.8567551088146715.
Cophenetic correlation for Euclidean distance and average linkage is 0.9416941280429498.
Cophenetic correlation for Euclidean distance and weighted linkage is 0.8914277699936133.
Cophenetic correlation for Chebyshev distance and single linkage is 0.9097107055043151.
Cophenetic correlation for Chebyshev distance and complete linkage is 0.7176448644438254.
Cophenetic correlation for Chebyshev distance and average linkage is 0.9373323239308111.
Cophenetic correlation for Chebyshev distance and weighted linkage is 0.8959147291677868.
Cophenetic correlation for Mahalanobis distance and single linkage is 0.9278960290214097.
Cophenetic correlation for Mahalanobis distance and complete linkage is 0.7905902397868768.
Cophenetic correlation for Mahalanobis distance and average linkage is 0.9365486257072162.
Cophenetic correlation for Mahalanobis distance and weighted linkage is 0.8457994014255156.
Cophenetic correlation for Cityblock distance and single linkage is 0.9335524238799446.
Cophenetic correlation for Cityblock distance and complete linkage is 0.8151784918186171.
Cophenetic correlation for Cityblock distance and average linkage is 0.9185297621803099.
Cophenetic correlation for Cityblock distance and weighted linkage is 0.825845425041732.

Model Overview and Performance Summary

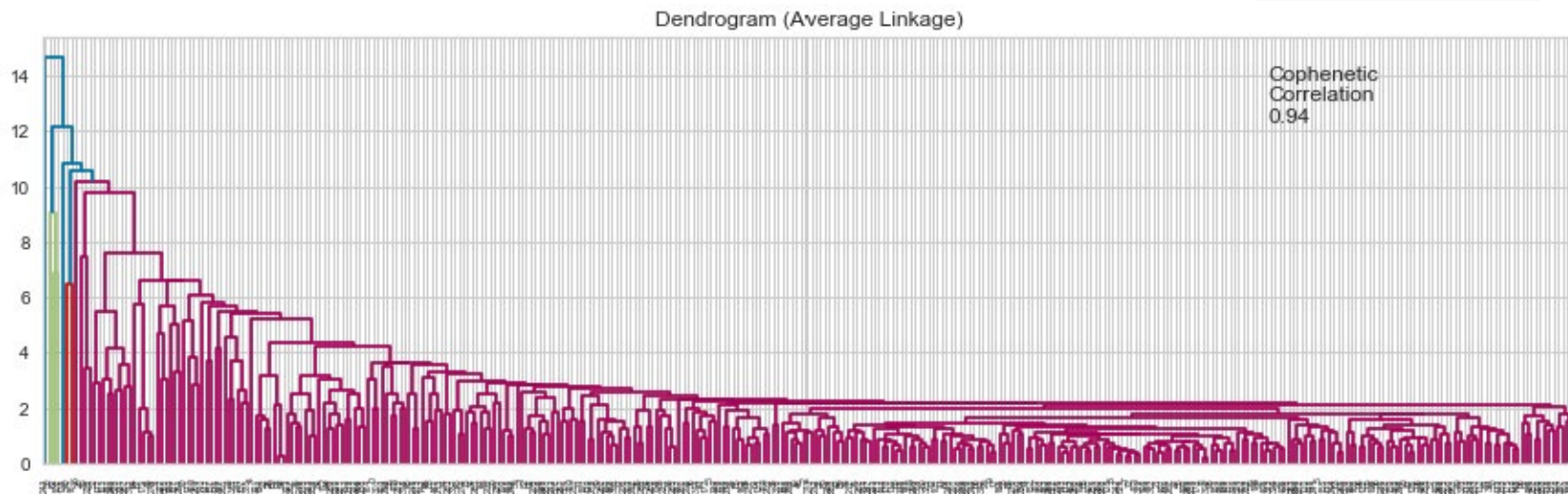
- Hierarchical Clustering (cont.)- determine which linkage of clusters will be optimal.
- Average linkage came in at .9416- which is very good.

```
Cophenetic correlation for single linkage is 0.9285583986172286.  
Cophenetic correlation for complete linkage is 0.8567551088146715.  
Cophenetic correlation for average linkage is 0.9416941280429498.  
Cophenetic correlation for centroid linkage is 0.9389964702246901.  
Cophenetic correlation for ward linkage is 0.7001925759044317.  
Cophenetic correlation for weighted linkage is 0.8914277699936133.  
Wall time: 22.0 ms
```

Model Overview and Performance Summary

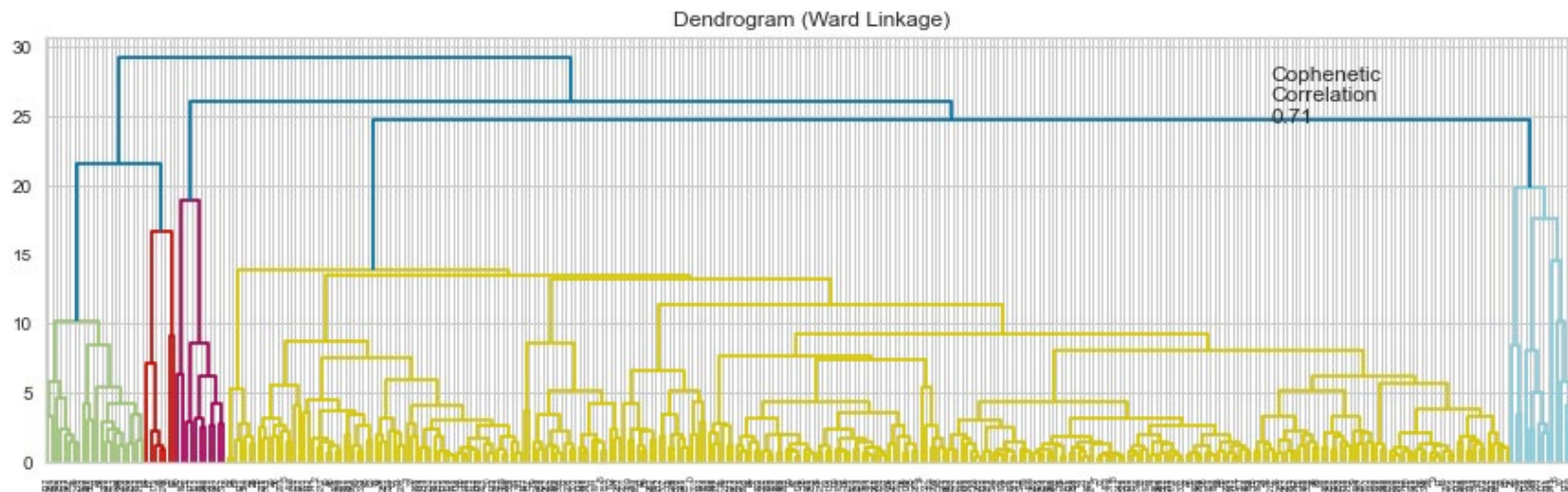
- Hierarchical Clustering (cont.)- Average linkage Dendrogram

	Linkage	Cophenetic Coefficient
0	single	0.923227
1	complete	0.787328
2	average	0.942254
3	centroid	0.931401
4	ward	0.710118
5	weighted	0.869378



Model Overview and Performance Summary

- Hierarchical Clustering (cont.)- advantages due to the obvious visual cluster balance.
- While the cophenetic isn't quite as good, its still good enough to be useful for the analysis and provides balance for the different cluster types which will be helpful to balance the natural outliers.



Model Overview and Performance Summary

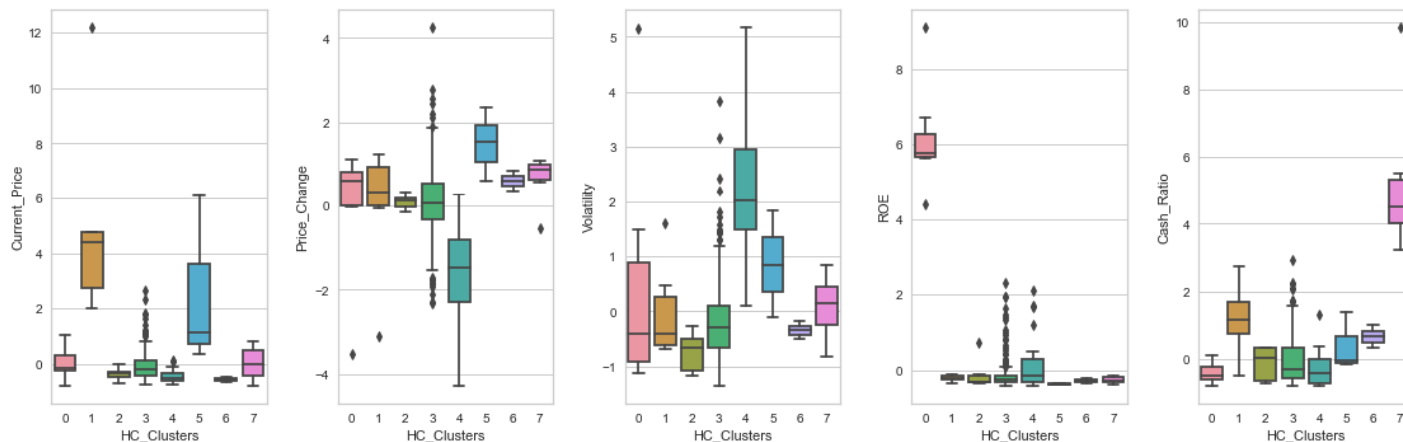
- Hierarchical Clustering- (cont)
- Paying particular attention to clusters 0 , 1 and 2 for the relationships with volatility, earnings per share and return on equity.

	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	84.355716	3.854981	1.827670	633.571429	33.571429	-568400000.000000	-4968157142.857142	-10.841429	398169036.442857	42.284541	-11.589502	1.857143	7
1	570.466654	3.666088	1.515959	19.833333	176.000000	23440500.000000	839355500.000000	16.598333	57823170.553333	50.550419	38.681119	2.000000	6
2	46.672222	5.166566	1.079367	25.000000	58.333333	-304066666.666667	14848444444.444445	3.435556	4564959946.222222	15.596051	-6.354193	3.000000	9
3	72.760400	5.213307	1.427078	25.603509	60.392982	79951512.280702	1538594322.807018	3.655351	446472132.228456	24.722670	-2.647194	1.014035	285
4	36.440455	-16.073408	2.832884	57.500000	42.409091	-472834090.909091	-3161045227.272727	-8.005000	514367806.201818	85.555682	0.836839	0.181818	22
5	327.006671	21.917380	2.029752	4.000000	106.000000	698240666.666667	287547000.000000	0.750000	366763235.300000	400.989188	-5.322376	0.000000	3
6	25.640000	11.237908	1.322355	12.500000	130.500000	1675550000.000000	13654000000.000000	3.295000	2791829362.100000	13.649696	1.508484	2.500000	2
7	81.525556	11.783328	1.575565	14.166667	544.666667	349565833.333333	1178117333.333333	2.490000	1121045865.050000	31.693323	11.835447	2.000000	6

Model Overview and Performance Summary

- Hierarchical Clustering- (cont)
- Pay particular attention to Cluster 4 with the negative price change and high volatility
- Cluster 5 has high positive price change and significantly lower volatility
- Cluster 0 has very high return on equity but relatively low volatility
- Cluster 7 has low volatility, but has a very high cash ratio.

Boxplot of scaled numerical variables for each cluster



Key Findings

Hierarchical clustering was a superior model ; however, it was very helpful to also have K-Means to provide a more complete understanding of the clusters and the relationships that overlapped and provided additional confidence.

- The cluster with highest Current Price is cluster 1 which is comprised of Alliance Data Systems, Chipotle Mexican Grill, Equinix, Intuitive Surgical Inc, Priceline.com Inc, Regeneron
- Price change highest overall positive would come from cluster 5 (Alexion Pharmaceuticals, Amazon.com, Netflix Inc)
- Price change lowest comes from cluster 4 (22 securities)
- Highest Volatility comes from cluster 4 as well, lowest volatility comes from cluster 2
- The best ROE comes from the 0 cluster which comprised of the following securities: Allegion (Industrials-Building Products), Apache Corporation(Energy- Oil and Gas exploration), Chesapeake Energy(Energy- Integrated Oil and Gas) , Charter Communications (Consumer Discretionary-Cable and Satellite), Colgate-Palmolive(Consumer Staples- Household Products), Kimberly-Clark (Consumer Staples- Household Products), S&P Global, Inc (Financials- Diversified Financial Services)
- Cash Ratio- best would come from cluster 7 (Amgen, Facebook, Frontier Communications, Monster Beverage, Waters Corporation, Yahoo Inc)
- Very important- Cluster 0 has a outlier for ROE, which is what I might consider an extremely important variable when considering investing
- Cluster 7 offered the highest Cash Ratio, another important factor because it expresses how liquid a security and it could offer the strength of the security.

Key Findings- cont.

- Both Hierarchical and KMeans provided similar results, generally the more risk the higher the potential reward.
- KMEANS observed in cluster 2 a group of 21 which provided a 7.126 return on investment with a fairly low volatility rating 1.719
- HIERARCHICAL observed 6 in cluster 1 which were all contained within cluster 2 from KMEANS, however it was able to indicate 16.598 return on investment with a 1.515 volatility factor (beta)
- this is extremely important, while cluster 2 is good, to provide a lower beta (volatility) rating along with a 42% improvement in risk is a material development.
- KMEANS was able to isolate 3 stocks- (Apache Corporation, Chesapeake and Devon Energy Corporation which have high volatility rates, but return very high Return on Equity. Devon Energy was also present in cluster 4 for hierarchical clustering. All of these companies are in Energy and 2 of them are in exploration and production.
- HIER was able to observe 9 securities in cluster 2 which overlapped with the KMEANS cluster 3 of 11 securities. This group had a very low volatility (Beta) rating but provided consistent returns. the 9 securities : Citigroup, Ford Mo, Gilead Sciences, Intel, JPMorgan Chase, Coca Cola Company, Pfizer, Verizon Communications, Wells Fargo, Exxon Mobile Corp.
- Highlighted items can provide insight to provide a balanced portfolio to clients based upon their risk tolerance along with their expectations for rates of returns. Obviously - a low risk individual would have a weighted portfolio with more of the cluster 2 Hierarchical observations, whereas those with high risk would overweight their selections with Cluster 1 stocks.

Business Recommendations

Careful analysis of the groups by their volatility and their rates of return can be utilized to provide an equation to fit to the needs of the individual investor. We can develop a model which we can use once we determine through client surveys the risk tolerance, earnings expectations and overall investment strategy to provide a tailored product that is more likely to provide specific performance.

- For example , if a high-risk tolerant individual would like to maximize opportunities and they have a long-term approach to their investment strategy – the investment advisor could use an overweight of Cluster 2 (K-Means) or Cluster 1 (HIERARCHICAL) combined with a healthy diversified mix of lower risk stock from Cluster 5 (HIER) . This will average out the losses that will occur by the more conservative cluster 5 while keeping a good portion for when the significant gains occur from Cluster 1.
- We recommend using a weighted average connected to the risk score (beta) combined with the associated return on equity and price change metrics to provide a useable model that could be adapted and easily changed as necessary for the client.
- High Risk = High Reward
- Groups with low risk but mediocre rewards are very valuable for the foundations of the stock mix.

Business Recommendations- cont.

- This model would need to be run on a periodic basis to ensure that organic clusters associated with securities maintained or changed in order to provide the best modeling for consumers.
- Working with additional machine learning could assist in providing guidance in development of algorithms to enhance the risk vs. reward matrix along with probability as historic data is compiled.



Thank you .

- By - Frederick Duff