Stock market analysis: K-means clustering and creating a stock portfolio

Ayaz Aliyev

Table of Contents

1. Introduction

1.1 Introduction and Aims

A stock market is a marketplace where businesses trade, purchase and sell shares in an effort to increase profits while reducing costs (Göçken et al., 2016; M et al., 2018). According to Nti et al. (2019), financial markets are one of the main improvement-driven tools in economies and their impact on the global economy and economic growth is unprecedented. Consequently, financial markets influence all sectors from education to technology (Shah et al., 2019) and it plays an important role in them as fuel (Pan & Mishra, 2018). Because of all these reasons stock analysis can be considered crucially important.

There is not only one proven way to analyze stocks, because financial variables are not the only factor to take into account. Political and economic factors also have manipulative effects on financial markets (Shah et al., 2019; Bintara & Tanjung, 2019). Although stocks are highly sensitive to external factors (Gandhmal & Kumar, 2019) creating a stock portfolio via clustering is one of the most used method. In this analysis, researchers try to find possible upward or downward trends of stocks to build a portfolio rather than finding the exact price of the stock for next year (Ng & Khor, 2016)

Technology sector investments are highly profitable. Because investing in technology can be a great way to diversify a portfolio and benefit from the potential for long-term growth. Technology companies often have high growth potential, as they are often at the forefront of innovation and can benefit from the rapid changes in the industry (Ortega-Argilés et al., 2014). Moreover, technology companies often have high margins, which can lead to higher returns (Cornelli et al., 2020).

Table 1 demonstrates the aim and objectives of this study:

Aim and objective	
Aim	*To create a profitable stock portfolio between technology sector stocks*
Objective	*Using k-means clustering and financial ratios and proving that they are really essential in creating a profitable stock portfolio*

Table 1.

1.2 Literature review

There are widely accepted 2 theories in stock analysis: fundamental and technical analysis (Herawati & Angger, 2018; Bintara & Tanjung, 2019). Fundamental analysis is based on the analysis of financial statements and it is mainly accepted as a long-term analysis, while technical analysis is short-term and generally, uses only the stock prices. Technical analysis

is a generally used method and it is common in literature while fundamental analysis is a less used method because of its complexity (Bustos & Pomares-Quimbaya, 2020).

The clustering technique is used to find similar and dissimilar data points and k-means clustering is the most used one among them (Sinaga & Yang, 2020). K-means clustering has been used in many kinds of research to classify stocks for their similar features, divide the stock portfolio, reduce the risks and finally, choose the right stocks to invest (Nanda et al., 2010; Momeni et al., 2015; Wu et al., 2022).
However, obviously, k-means is not the only clustering technique to classify stocks and create stock profiles. For instance, Ng and Khor (2016) used EM (Expectation Maximization) technique and Nanda et al. (2010) used Fuzzy C-means and SOM neural network techniques as well as k-means clustering.

2. Methodology

2.1 Data set and R packages

In this report, "200 + indicators of the US stock market between 2014 and 2018" data set has been used (Carbone, 2019). This data set can be considered noisy because it has so many outliers, mistyping, and missing values. Although it has pros too: a wide range of financial ratios have been given calculated for each year, so there is no need to do further calculations. In this report, a financial sheet from 2014 for the technology sector has been analyzed and other years' price variations have been used to see short-term and long-term changes.

These R packages have been utilized in the analysis of stocks: "ggplot2", "dplyr", "gridExtra", "factoextra", "tidyverse", "cluster", "writexl".

"dplyr", "tidyverse" are used to pre-process the data, "ggplot2", "gridExtra" are used to create visualizations to explain data more clearly and "factoextra" and "cluster" are used to do cluster analysis. Moreover, "writexl" package has been used to export R data to Excel for creating the tables which are used in this report.

2.2 Used variables

Financial ratios have been utilised extensively for financial assessments over the years, and it has been demonstrated that they are highly helpful for assessing stocks and companies (Mokhtar et al., 2014). In table 2, some examples of the widely used financial ratios are given.

Literature	Type of Research	Used variables and mentioned variables
Ng and Khor (2016)	**Stock analysis** **Clustering - EM**	Asset Turnover = ATR, Debt Ratio = DR, Return on Equity = ROE, Dividend Yield = DY,Price Earnings ratio = PER, Cash Ratio= CR
Nanda et al. (2010)	**Stock analysis** Clustering - K-means clustering, Self-organizing maps (SOM), Fuzzy C-means	1. Return of stocks - Short-term and long-term. 2. Ratios - Price earning = PE, Price to book value = PBV, Price/cash EPS = PCEPS, EV/EBIDTA = EVE, Market cap/sales = MCS
Agrawal et al. (2013)	**General research** on fundamental and technical analysis	The Price-to-Book Ratio = PBR, Price-to-Earnings Ratio = PER, The PEG Ratio = PEG, Dividend Yield = DY, Debt to Equity Ratio = DER, Returns on Equity = ROE
Mokhtar et al. (2014)	**Fuzzy Delphi approach** to finding best variables in stock analysis	Return on Equity = ROE, Earnings per share = EPS, Operating profit margin = OPM, Net profit margin = NPM, Return on Assets = ROA, Debt to equity ratio = DER
Pok (2017)	**Analysing the financial health** of firms	Liquidity ratio or cash ratio = CR, Interest ratio = IR, Debt ratio = DR, Non-permissible income ratio = NPIR
Momeni et al. (2015)	**Stock market clustering** via K-means	Return on assets = ROA, Earnings Per Share = EPS, Return on equity = ROE, Profit to sales ratio = P/S ratio, Operating profit margins = OPM

Table 2.

Ng and Khor (2016), Nanda et al. (2010) and Momeni et al. (2015) have used financial ratios in the stock analysis while Pok (2017) used CR, IR, DR and NPIR to analyse the financial health of the company. The other two articles, Agrawal et al. (2013) and Mokhtar et al. (2014), mainly focused on the analysis of ratios that which ratios are more effective. According to the search, CR, DR, DY, ROE, ATR and PER are the variables that are mainly used in analyses and they are highly effective in stock analysis. Consequently, these variables are used in this report to analyse the stock and create a stock portfolio.

Table 3 shows the formulas of the used variables in the report:

Formulas of used variables

Cash Ratio	=	Cash/Current liabilities
Return on Equity	=	Net income/ Shareholder's Equity
Dividend Yield	=	Dividend per share/Price per share
Price Earnings ratio	=	Price per share/Earnings per share
Asset Turnover	=	Total sales/Assets
Debt Ratio	=	Total liability /Assets

Table 3.

Cash Ratio – It is a liquidity measure and helps firms to cover short-term debts. The cash ratio helps companies to avoid recession and continue to complete their strategic aims. One recent research shows that US companies tend to increase their cash ratio which caused a decrease in net debts (BATES et al., 2009).

Return on Equity – Return on equity is one of the unique measures to see the profitability of the company. Although some researchers mentioned that higher ROE means lower profitability (Fatila & Syahril, 2022), mainly, it is accepted that firms which have higher profitability and competent management tend to have high ROE (Adawiyah & Setiyawati, 2019).

Dividend Yield – Dividend yield means which percentage of the price of each share is returned by cash to the investor. Although a high dividend yield is attractive, high dividend payments can delude investors because they can cause an increase in debts in the company (Ng & Khor, 2016).

Price Earnings ratio – PER indicates how fast companies develop and usually higher PER is preferred by investors. However sometimes PER is used to find undervalued and overvalued companies and in this situation, investors prefer lower PER (Ng & Khor, 2016).

Asset Turnover - ATR is one of the main ratios to take into account while evaluating stocks. It is generally associated with higher productivity and good use of capital. Moreover, Patin et al. (2020) suggest that in the long term, ATR has a positive effect on stock return.

Debt Ratio – Husna & Satria (2019) mentioned that a high debt ratio means that the company has the capital to invest in its future and it means a higher debt ratio is better. However, according to Ng and Khor (2016), higher debt cause recession in the firms and a lower firm should be chosen to invest.

In this analysis higher ratios are preferred for investment.
Moreover, “Price Variation” (PV) and “sector” variables have been used in this analysis. PV shows the price change (%) for a particular year, so, it is really effective to evaluate stock investments. The “sector” variable has been used to filter data for the technology sector, so, it helps us to focus on technology sector stocks and reach our aim in this report.

2.3 Pre-processing methods and results

2.3.1 Data cleaning

After choosing the mentioned variables in 2.2, the first step was to delete N/A values because they do not create any meaning in data and create noisy data. Secondly, the number of zero values has been deleted because zero values create noisy data, too and make it hard to evaluate the stocks. Shen & Zhu (2019) also used the zero values cleaning method and mentioned that if one variable has zero values over 50% we can accept it as a noisy variable and discard it. In our analysis, instead of removing the whole variable, we removed only the zero values. Additionally, we already mentioned in the variable selection (2.2), we would prefer to invest in higher performed ratios, so it is also healthy for our analysis to remove zero values. In table 3, the percentage of zero values has been indicated for each ratio.

Financial ratios	Percentage of zero values
Cash Ratio	0.25%
Return on Equity	0.10%
Dividend Yield	60%
Price Earnings ratio	36%
Asset Turnover	5.50%
Debt Ratio	22%

Table 4.

After cleaning the zero values, the remained stocks have been filtered for the technology sector only because as we mentioned in this report we focus on technology sector stocks. Reportedly, there were 94 stocks after cleaning N/A and zero values and filtering for the technology sector. In 3. Section (Results and Discussion), these stocks have been mentioned and further results of data cleaning are described.

2.3.2 Outlier detection – Interquartile range

The Boxplot method or Interquartile range is one of the most used methods in analyses (Schwertman et al., 2004; Dovoedo & Chakraborti, 2014). In this outlier detection method we

first, set minimum and maximum ranges and consider the stocks out of these borders as outliers. Table 5 demonstrates the formulas used to find outliers and how to detect non-outliers in these stocks.

Minimum and Maximum range	Formulas
Minimum range	Q1 - (1.5*IQR)
Maximum range	Q3 + (1.5*IQR)
Non-outlier (x as particluar stock's ratio)	x>Minimum range, x<Maximum range

Table 5.

2.3.3 Evaluating the stock prices

As mentioned above, price variation is used to evaluate the stocks. We would look at the price variation of 5 years and PV is given in percentage. So the formula in the table 6 was used to find overall precentage changes.

Formula
Year 1: (1 + 0.x)
Year 2: (1 + 0.x) * (1 + 0.y)
Year 3: (1 + 0.x) * (1 + 0.y) * (1 + 0.z)
Year 4: (1 + 0.x) * (1 + 0.y) * (1 + 0.z) * (1 + 0.a)
Year 5: (1 + 0.x) * (1 + 0.y) * (1 + 0.z) * (1 + 0.a)* (1 + 0.b)

Explanation	
x	2015 PV
y	2016 PV
z	2017 PV
a	2018 PV
b	2019 PV

Table 6.

2.4 Analysis – K-means clustering

K-means clustering is one of the most used methods in stock analysis and it is mainly used for the diversification of stocks (Gandhmal and Kumar, 2019). Consequently, the main aim of applying k-means in stocks in this report is to reduce the risk of buying the same type of

stocks and to create a diverse stock portfolio. This has been a long-known term in finance "Do not put all of your eggs in the same basket", so clustering exactly serves this aim.

Although it is very famous and easy to apply, it has disadvantages too. For example, because of its unsupervised nature, the number of clusters must be known before the execution of the algorithm and it creates some hardships (Sinaga & Yang, 2020). Finding an optimal number of clusters requires further analysis. The most famous analyses are "silhouette", "wss" and "gap stat" and they have been used by various researchers in their research (Nazarov & Baimukhambetov, 2022; Wu et al., 2022). These methods have been used in this report too and further results have been mentioned in the 3. section.

3. Results and discussion

3.1 Pre-processing results

After "Data cleaning" (mentioned in 2. section), we should clean our data from outliers too. Figure 1 shows the boxplot visualization before the outlier detection method has been applied to the data.

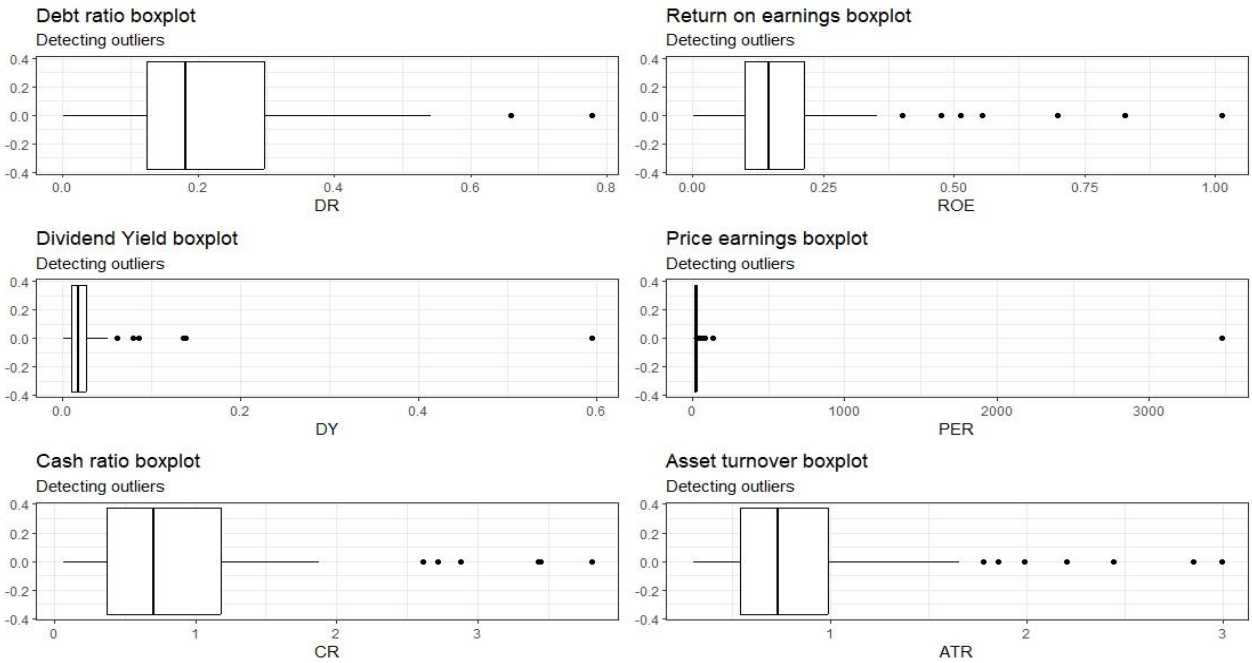

Fig 1.

As can be seen that there are some outliers which make our data noisy. We should remove them to do more precise clustering and further analysis.

Figure 2 indicates the situation of stocks after cleaning the outliers. Now our data is ready for analysing.

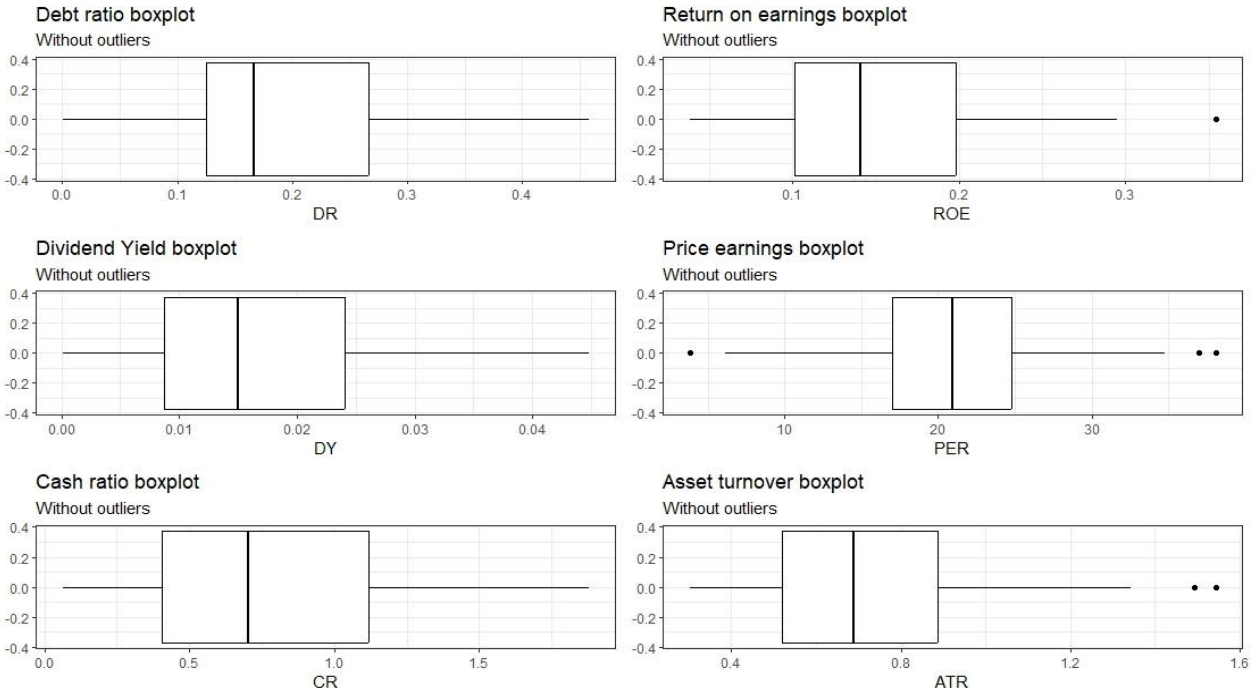

Fig 2.

3.2 Stocks preview before the analysis

After outlier detection, we have 63 non-outlier and 31 outlier stocks. Table 7 indicates the number of stocks that is not outlier while table 8 shows outlier stocks.

Non-outlier stocks											
1	INTC	12	WIT	23	AMAT	34	UMC	45	CSGS	56	WHR
2	AAPL	13	SSNC	24	NVDA	35	EBIX	46	OTEX	57	ENS
3	MSFT	14	AYI	25	TXN	36	MEI	47	AME	58	FLIR
4	HPQ	15	ASML	26	KLAC	37	KBAL	48	HUBB	59	APH
5	ORCL	16	BMI	27	NTAP	38	AMOT	49	LFUS	60	ADI
6	CSCO	17	CW	28	MCHP	39	FORTY	50	DAKT	61	ERIC
7	ATVI	18	MLAB	29	HIMX	40	HURC	51	FICO	62	FIS
8	XRX	19	BELFA	30	INTU	41	DOX	52	MTSC	63	BRKS
9	WDC	20	QADB	31	XLNX	42	VSH	53	EVOL		
10	TSM	21	SAP	32	DBD	43	ASX	54	CTS		
11	LRCX	22	ADTN	33	ABB	44	IAC	55	PLPC		

Table 7.

Outlier stocks					
1	NOK	12	LDOS	23	AVT
2	AVGO	13	BDC	24	SAIC
3	JBL	14	STM	25	EVTC
4	ACN	15	MXIM	26	SIMO

5	IBM	16	NTES	27	ESE
6	GLW	17	CDW	28	MGIC
7	STX	18	SFUN	29	TESS
8	CY	19	MSI	30	JCS
9	SABR	20	WSO	31	OCC
10	TEL	21	JCOM		
11	TDC	22	BLKB		

Table 8.

In our clustering analysis, we use only non-outlier stocks, however, in the stock selection process we accept outliers as another cluster and choose stocks from them too. This method has been used by Ng and Khor (2016) and the authors showed that it can be really effective.

3.3 Analysing results

3.3.1 Forming clusters

As we mentioned in the methodology, for doing clustering analysis, firstly, we should define a number of optimal clusters. To define it, 3 methods have been used: Silhouette, WSS and Gap stat. Figures 3, 4 and 5 demonstrate the optimal number of clusters according to Silhouette, WSS and Gap stat analyses respectively.

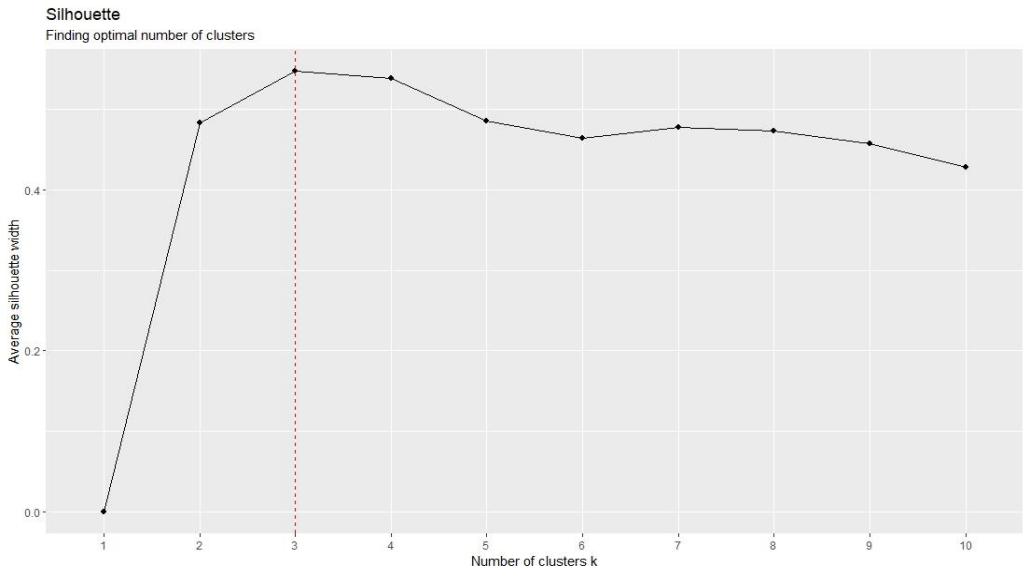

Fig 3.

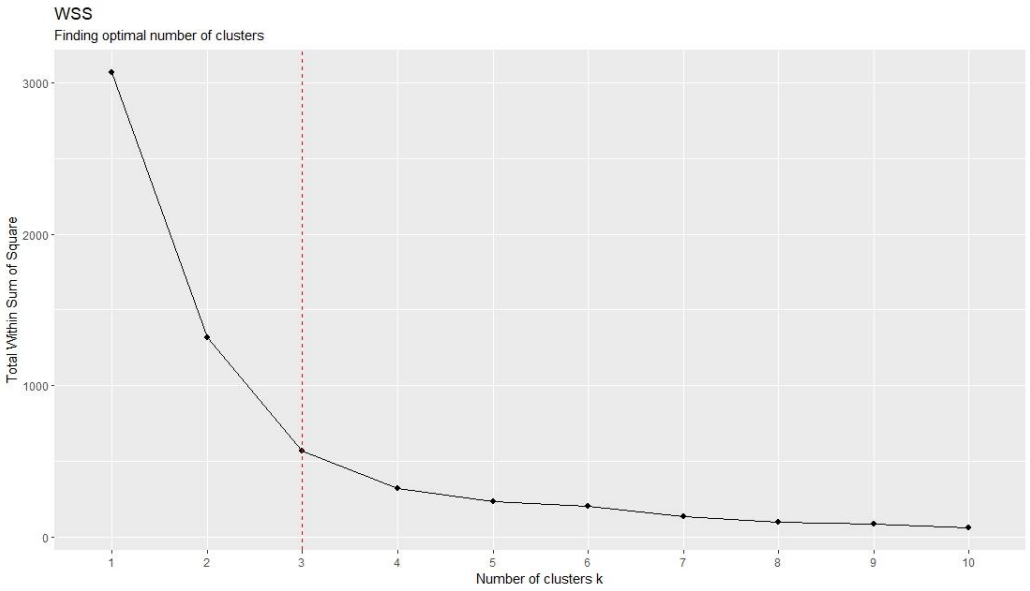

Fig 4.

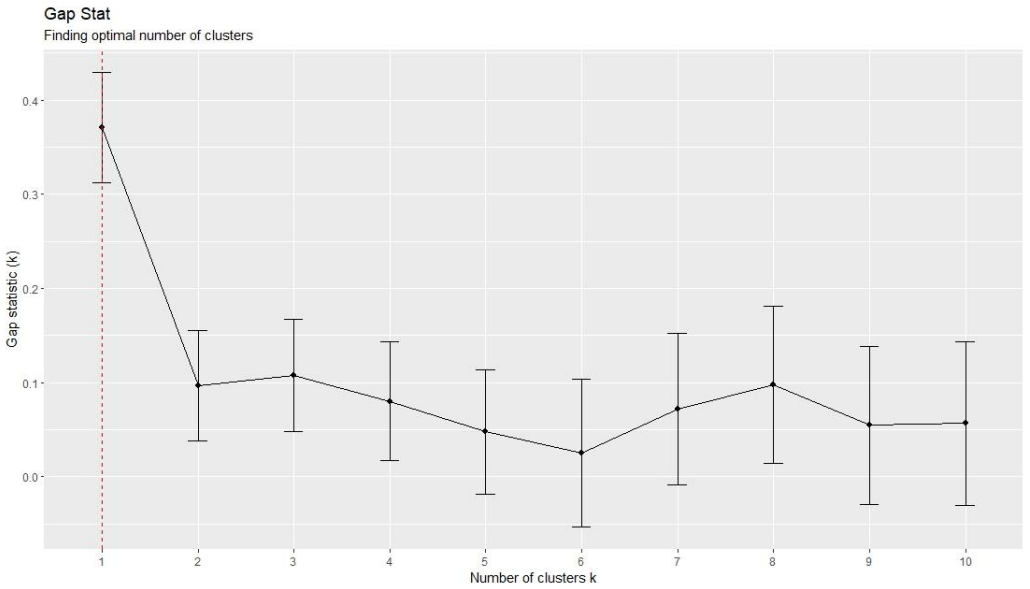

Fig 5.

Analyses show that according to Silhouette and WSS analysis, an optimal number of clusters is 3, however, Gap Stat suggest that an optimal number of cluster is 1. We accept that our optimal number for clusters is 3 because 2 out of 3 analyses suggest this.

After finding an optimal number for clustering we can do a clustering analysis.

After an analysis, we have 22 stocks in cluster 1, 9 stocks in cluster 2 and 32 stocks in cluster 3. Figure 6 indicates the visualisation of our clustering analysis and table 9 demonstrates the average performances of each cluster.

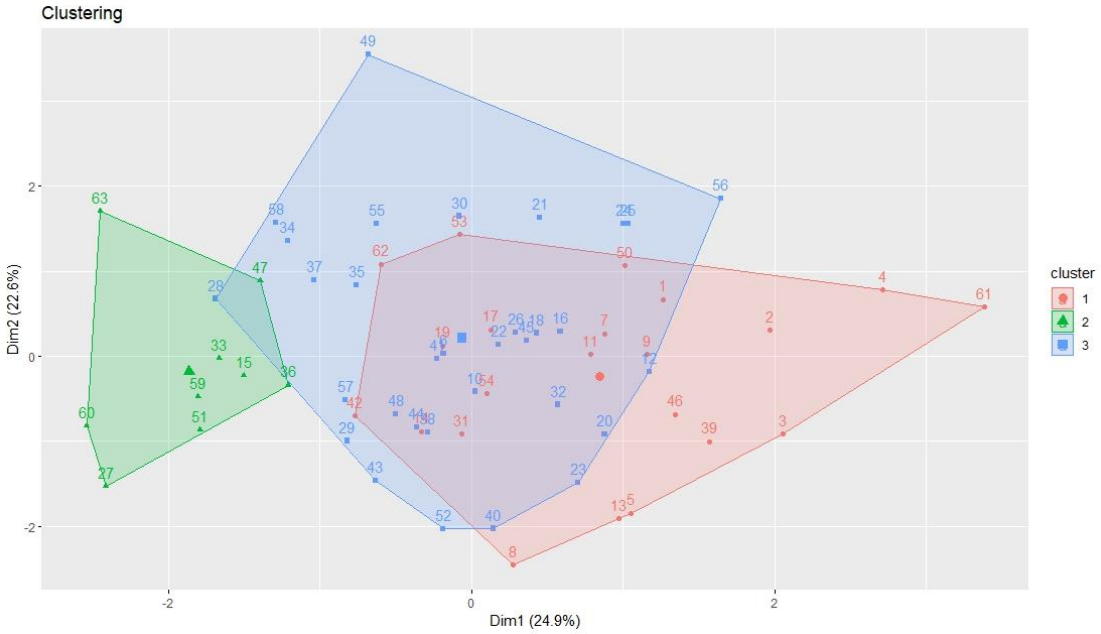

Fig 6.

Cluster Numbers	Debt Ratio	Dividend Yield	Cash Ratio	Price-earnings ratio	Asset turnover	Return on earnings
Cluster 1	0.176686364	0.01545	0.88880564	14.10773182	0.681127079	0.172259091
Cluster 2	0.193077778	0.007477778	0.662848639	33.3157	0.773156837	0.125311111
Cluster 3	0.186396875	0.019403125	0.758985853	22.60234375	0.776215657	0.150021875

Table 9. Red marks show the best performances for each ratio

According to this table, each cluster has the same amount of strengths and weaknesses, if we want to differentiate them specifically, we need to get into more deep details. For sure investors can invest in one of these clusters based on their investment strategies but this kind of detailed analysis is not our research aim, so to reach the aim we need to look at stocks and with choosing the best stocks from each cluster to build our portfolio.

Table 10, 11 and 12 shows the clustered stocks and the best-performed ratios are red-marked from each stock. We would choose 3 best-performed stocks from each cluster. Additionally, table 13 demonstrates the outlier stocks, as we mentioned before we would give a chance to outliers and choose the 2 best-performed stocks from them.

Cluster 1

Name	Debt Ratio	Cash Ratio	Price earnings ratio	Dividend Yield	Return on earnings	Asset turnover
INTC	0.1486	0.159952533	15.5565	0.0242	0.2095	0.607943417
AAPL	0.1522	0.218194427	15.4253	0.0181	0.3542	0.788456644
MSFT	0.1314	1.878553425	15.6767	0.0257	0.2459	0.503718443
HPQ	0.1892	0.346015777	6.1257	0.0374	0.1875	0.548911885
ORCL	0.267	1.234901661	17.343	0.0114	0.2337	0.42402455
CSCO	0.1984	0.339542632	17.28	0.0278	0.1386	0.448672314
ATVI	0.2953	1.786293294	17.6754	0.0099	0.1154	0.301051769
XRX	0.2799	0.232225148	10.6168	0.0274	0.0949	0.45842071
WDC	0.1573	1.249414824	13.5029	0.0135	0.1829	0.976192012
TSM	0.1677	1.594629006	14.4695	0.0006	0.2488	0.510288637
LRCX	0.1671	1.010852079	17.599	0.0027	0.1257	0.576395924
DOX	0.0405	1.186315685	17.3132	0.013	0.1243	0.687260681
VSH	0.1356	1.32492074	17.6875	0.017	0.0644	0.761504891
ASX	0.2978	0.454735681	13.4548	0.0011	0.1507	0.768885163
IAC	0.2537	1.470097966	12.2068	0.0191	0.2083	0.730474749
UMC	0.1276	0.92891645	15.9972	0.001	0.0483	0.450710407
EBIX	0.1923	0.749433984	10.1131	0.0177	0.147	0.337879999
MEI	0.0834	0.977329975	11.419	0.0104	0.2452	1.34283232
KBAL	0.0004	0.53873817	14.8322	0.0153	0.0758	0.753056861
AMOT	0.448	0.364695739	15.5855	0.0042	0.2477	1.494376979
FORTY	0.1399	0.701923227	3.8103	0.0357	0.2489	0.575476876
HURC	0.0138	0.806041645	16.6797	0.0067	0.092	0.938260506

Table 10. Red marks show the best performances for each ratio

Cluster 2						
Name	**Debt Ratio**	**Cash Ratio**	**Price earnings ratio**	**Dividend Yield**	**Return on earnings**	**Asset turnover**
WIT	0.1027	0.83877578	33.6853	0.0001	0.227	0.86455413
SSNC	0.2735	0.5796289	37.019	0.0022	0.0974	0.33883869
AYI	0.1624	1.17428268	30.7297	0.0042	0.1511	1.10482829
ASML	0.0946	0.83751992	32.5063	0.0095	0.1593	0.47986753
BMI	0.2226	0.06216204	28.5337	0.0125	0.1385	1.06920547
CW	0.2807	0.78692571	29.911	0.0074	0.0767	0.65983784
MLAB	0.1692	0.54576603	34.5785	0.0064	0.1399	0.54059818

BELFA	0.3661	0.63330651	34.7826	0.01	0.0384	0.76654061
QADB	0.0659	0.50727018	38.0952	0.015	0.0995	1.13414078

Table 11. Red marks show the best performances for each ratio

Cluster 3						
Name	**Debt Ratio**	**Cash Ratio**	**Price earnings ratio**	**Dividend Yield**	**Return on earnings**	**Asset turnover**
AMAT	0.1478	1.063408	23.9091	0.019	0.1362	0.688629
NVDA	0.1895	1.217971	20.6133	0.0201	0.0987	0.569585
TXN	0.2666	0.451091	20.4866	0.0232	0.2715	0.750921
KLAC	0.1346	0.704392	20.6952	0.0248	0.1588	0.529171
NTAP	0.1075	0.850219	18.7166	0.0171	0.1685	0.68647
MCHP	0.2509	1.388373	24	0.0297	0.1851	0.474777
HIMX	0.1561	0.521844	21.2105	0.0335	0.1409	1.009061
INTU	0.0959	0.597467	25.7767	0.0093	0.2947	0.815805
XLNX	0.3095	0.984148	22.8987	0.0184	0.229	0.472973
DBD	0.2158	0.31728	19.5706	0.0332	0.2152	1.16767
ABB	0.1709	0.349358	18.7168	0.037	0.1594	0.888032
ADI	0.1272	0.802804	24.7363	0.0292	0.1323	0.417624
ERIC	0.0822	0.40653	26.2644	0.0049	0.0802	0.77662
FIS	0.349	0.308212	26.1345	0.0154	0.1036	0.441651
BRKS	0.0107	0.826714	22.3617	0.0324	0.0488	0.620597
SAP	0.2993	0.38815	20.9203	0.0239	0.1682	0.455335
ADTN	0.039	0.617088	26.9136	0.0165	0.0813	0.852866
WHR	0.2173	0.122099	23.3422	0.0148	0.1331	0.993501
ENS	0.1388	0.412972	21.858	0.0072	0.1206	1.065712
FLIR	0.1588	1.716563	22.7535	0.0124	0.1244	0.651533
APH	0.3802	1.271614	23.8097	0.0084	0.2439	0.765184
CSGS	0.2993	0.948152	22.7909	0.0248	0.0996	0.879075
OTEX	0.3383	0.667326	26.3407	0.013	0.1329	0.416622
AME	0.2669	0.403373	22.0209	0.0063	0.1804	0.62638
HUBB	0.1796	1.309896	19.3884	0.0193	0.1688	1.011837
LFUS	0.1822	1.455357	22.777	0.0095	0.1348	0.795643
DAKT	0.0103	0.340802	24.8269	0.0302	0.1093	1.544184
FICO	0.4579	0.307773	19.6786	0.0015	0.2087	0.661735
MTSC	0.1231	0.29755	24.8152	0.0175	0.1627	1.157814
EVOL	0.0003	1.021514	19.5208	0.0448	0.1647	0.671007
CTS	0.1641	1.681728	22.5696	0.009	0.0915	0.884215
PLPC	0.0951	0.535778	22.8577	0.0146	0.0529	1.09667

Table 12. Red marks show the best performances for each ratio

Outliers						
Name	**Debt Ratio**	**Cash Ratio**	**Price earnings ratio**	**Dividend Yield**	**Return on earnings**	**Asset turnover**
NOK	0.1361	1.001235	6.9069	0.0787	0.3147	0.55842
AVGO	0.5251	1.57874	82.2286	0.0131	0.0811	0.40692
JBL	0.1984	0.23148	18.1176	0.0148	0.1076	1.858799
ACN	0.0015	0.603243	17.4828	0.0229	0.5132	1.777684
IBM	0.3472	0.214143	13.4035	0.0265	1.013	0.79127
GLW	0.1085	2.611015	12.5989	0.0174	0.1146	0.323155
STX	0.413	1.100083	12.2146	0.0293	0.5544	1.445849
CY	0.3273	0.377452	132	0.0303	0.0863	0.976074
SABR	0.6596	0.173568	84.4583	0.0089	0.8264	0.56674
TEL	0.1914	0.615327	13.5484	0.1374	0.1977	0.594135
TDC	0.1494	0.838191	18.5085	0.0847	0.215	0.872286
LDOS	0.3203	0.426165	23.3711	0.5945	0.1028	1.382749
BDC	0.5422	1.417106	45.8198	0.0025	0.0922	0.707911
STM	0.2	1.142129	53.3571	0.0502	0.0256	0.822301
MXIM	0.2272	3.442811	27.048	0.0308	0.146	0.55694
NTES	0.0675	3.425894	16.8239	0.004	0.2034	0.385866
CDW	0.5211	0.234673	24.4236	0.0056	0.2615	1.987278
SFUN	0.333	0.905244	11.9968	0.1353	0.4003	0.402973
MSI	0.3262	1.757333	12.6805	0.0194	0.475	0.564233
WSO	0.1694	0.085175	24.7113	0.0187	0.1713	2.202341
JCOM	0.3482	2.721159	23.8462	0.0177	0.1516	0.351296
BLKB	0.2975	0.369222	68.6667	0.0111	0.1522	0.598422
AVT	0.1843	0.186584	11.2177	0.0135	0.1116	2.444272
SAIC	0.3469	0.450355	15.8841	0.0151	0.2997	2.847961
EVTC	0.7789	0.409985	26.3452	0.0181	0.6976	0.408652
SIMO	0.0013	3.808601	17.9167	0.0254	0.1462	0.789348
ESE	0.0473	0.238824	3478	0.0115	0.0007	0.627871
MGIC	0.0149	2.882383	16.5278	0.0361	0.0839	0.732898
TESS	0.0131	0.184732	18.8687	0.0198	0.1415	2.995752
JCS	0.0063	1.057099	45.6522	0.061	0.0228	1.187316
OCC	0.1998	0.11713	46.1	0.0174	0.0221	1.658274

Table 13. Red marks show the best performances for each ratio

3.4 Building a stock portfolio

After looking at the stocks, the best-performed of them have been chosen and decided that they are worth to be invested. Table 14 demonstrates the best-performed stocks. We build our stock portfolio with them. Investors should take into account that we invest in these stocks for the long term and moreover, they can avoid investing in outliers because although chosen outlier stock performances are good enough, they are still having high risk because of their unreliable financial ratio performances.

Chosen stocks from each cluster			
Cluster 1	**Cluster 2**	**Cluster 3**	**Outlier stocks**
MSFT	BELFA	DBD	SABR
ATVI	QADB	DAKT	BDC
AMOT	AYI	APH	

Table 14.

3.4.1 Evaluating our investments - 2015 and trends over the years

After building our stock portfolio we need to check what is our gain or in other words what is the stock return for each stock. We would look at the price variation tables of 5 years (2015-2019) and try to evaluate stock return for the first (2015) and 5th (2019) years. In the tables and graphics, other years (2016-2018) have been given too, for making the whole picture clear and to help understand the price variation fluctuation over the years. This kind of evaluation helps us to understand 2 main points: Whether financial sheet analysis helps us to invest for the short (1 year) and long term (5 years) or not and how price variation changes (accelerate, slow down or cease) over the years.

Table 15 and Figure 7 demonstrate the Cluster 1 stocks:

Price Variation (%)					
Names	**2015**	**2016**	**2017**	**2018**	**2019**
MSFT	21.88	16.51	39.71	20.22	58.26
ATVI	94.27	-3.23	73.87	-27.21	27.35
AMOT	14.07	-13.93	48.62	28.49	9.07

Table 15.

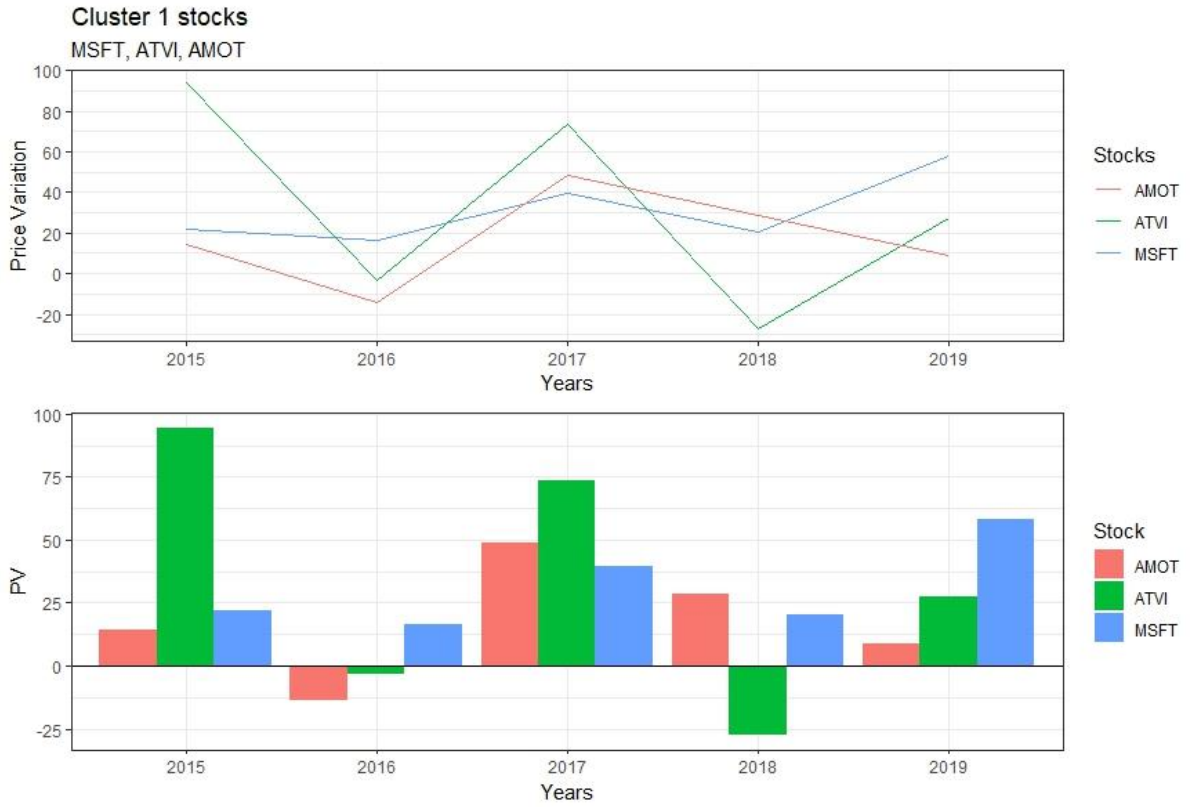

Fig 7.

Cluster 1 stocks' overall performance is great and for the first year, their prices have been increased. Figure 7 indicates the whole fluctuation over the years.

Table 16 and Figure 8 show the Cluster 2 price variation:

Price Variation (%)					
Names	**2015**	**2016**	**2017**	**2018**	**2019**
BELFA	-33.46	67.02	-13.56	-36.05	-0.89
QADB	-0.99	45.57	22.07	-3.84	29.89
AYI	67.64	-0.79	-24.64	-35.83	20.25

Table 16.

Fig 8.

In the comparison with 1st cluster, cluster 2 stocks haven't shown sharp increase rates and their price variation fluctuated. Only "AYI" shares had increased in 2015.

In table 17 and Figure 9, 3rd cluster's stocks have been given:

Price Variation (%)					
Names	**2015**	**2016**	**2017**	**2018**	**2019**
DBD	-8.45	-11.32	-35.18	-85.59	291.11
DAKT	-25.12	36.74	-13.33	-16.72	-14.59
APH	-1.68	33.35	31.36	-6.82	36.99

Table 17.

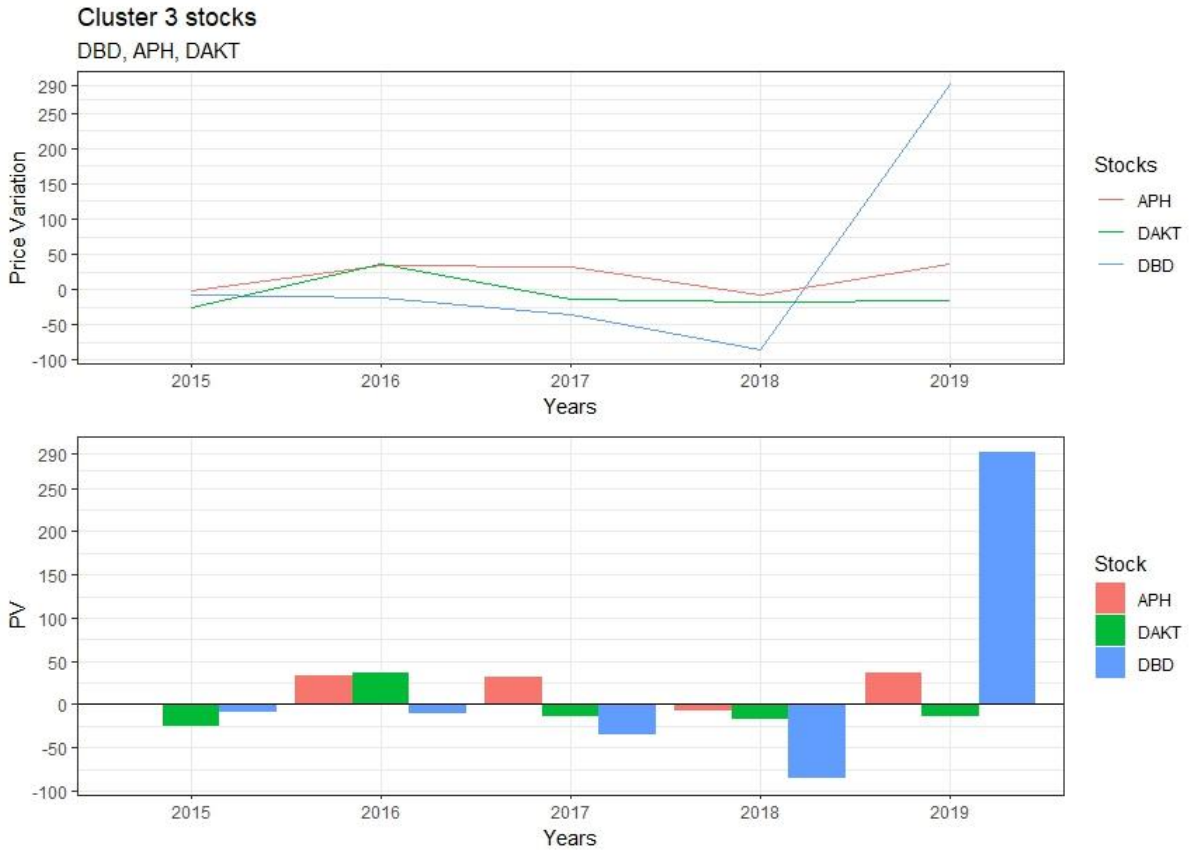

Fig 9.

The worst performance has been noted between the 3rd cluster's stocks, none of the stocks' prices increased during 2015 and the general trend over the years was downward. Although in 2019 we can see a sharp increase in the price of "DBD" it doesn't mean that in 5 years it showed a price increase. We should note that every year's price variation is given in percentage and every year's price change affects the next year, after looking at the price changes over the year we will look at the overall changes in 5 years, too.

Table 18 and Figure 10 shows the Outliers' price variation:

Price Variation (%)					
Names	**2015**	**2016**	**2017**	**2018**	**2019**
SABR	41.19	-7.28	-15.01	7.28	7.76
BDC	-39.15	61.92	0.84	-46.95	31.48

Table 18.

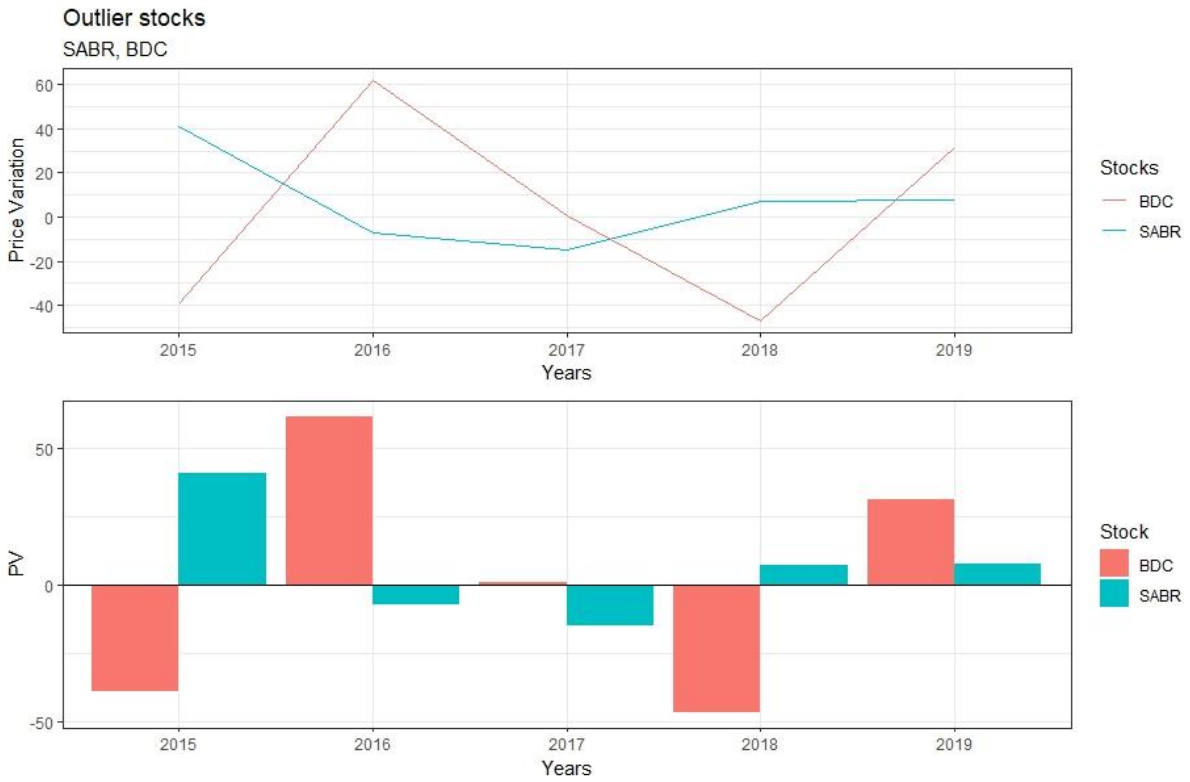

Fig 10.

As we can expect outliers haven't shown a stable trend and sharply fluctuated. In 2015, SABR shares increased while BDC shares decreased.

3.4.2 Evaluating our investments – Comparision of changes in first and last years

As mentioned above price variation has given in % in our data set, so we used the "annual growth rate" formula (Table 6) to calculate the overall changes in 5 years.

Figure 11 shows the price variation of 2015 and overall price variation of 5 years:

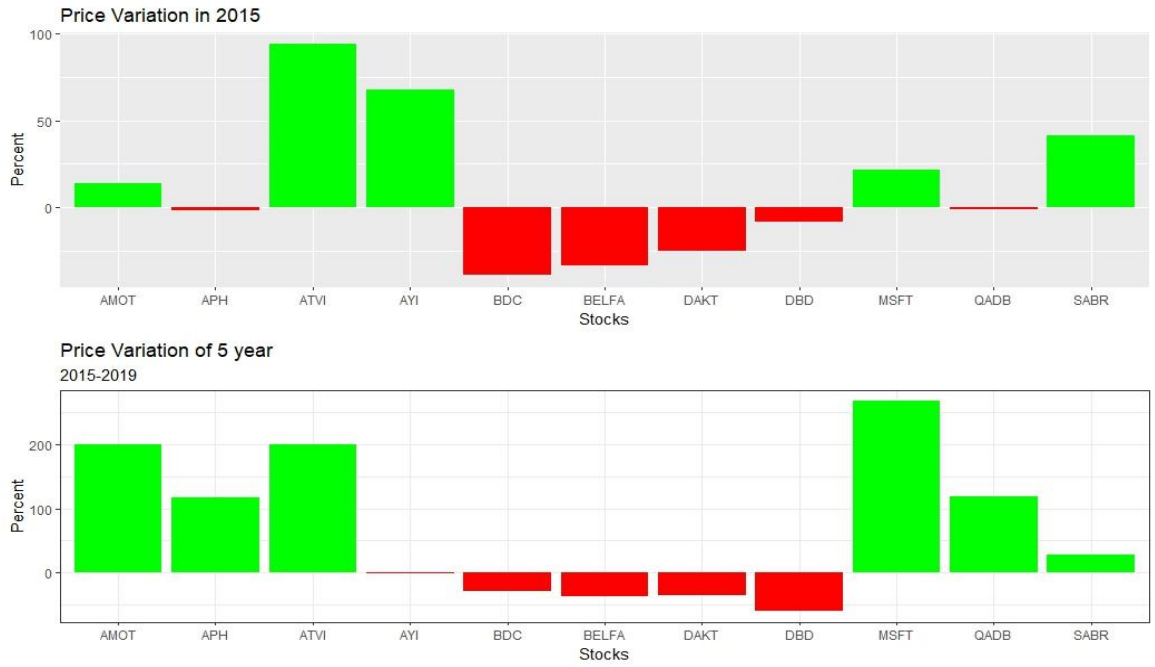

Fig 11.

Although “ATVI”, “AYI” and “SABR” were the best-performed stocks of 2015, after 4 years overall the best-performed stocks are “AMOT”, “ATVI” and “MSFT”. Especially, “MSFT” shares have increased more than 3 times and it’s a record among other stocks. “BDC”, “BELFA”, “DAKT” and “DBD” kept their negative trend over the years. “QADB” and “APH” get the most unexpected results in this analysis, although they have shown negative trends in 2015, after 4 years, their prices have increased overall.

3.5 Discussion

According to our analysis, creating a stock portfolio with the k-means clustering technique is efficient for 1 and 5-year investments and our results are similar to other stock portfolio builded literature (Ng & Khor, 2016: Nanda et al., 2010). Although this report has one difference: Evaluating the 5 years of stock price movement. Moreover, according to our analysis in the long-term (5 years), stock prices tend to show higher price increase rates and even some stocks which have a negative trend in the first year slightly show a positive trend in 5 years.

3.6 Limitations

This analysis has some limitations:

1. Not using stock prices – working with percentages can be deluded and creates some hardships in the interpretation of results.

2. Technology sector-focused analysis – although focusing on one sector helps to do more healthy comparisons between stocks, it also reduces the diversification of stocks. For example, if something bad happened in technology, all the stock prices would go down and investors would lose money. In real-life examples, it is better to diversify the stock portfolio for the sector too.

4. Conclusion

In this report, we have reached our aim and objective which have been mentioned in 1.1. Our aim was to create a profitable stock portfolio in the technology sector and according to our results, we can say that our stock portfolio is profitable and investors should use the clustering technique to diversify stocks and manually choose the best-performed stocks to invest. Our objective was to prove that financial ratios are really important in stock portfolio building and it is proven too. According to our analysis, stocks which have strong financial ratios mainly bring a good stock return in the short and long term.

In conclusion, stock portfolio building is a great way to diversify your investments and increase your chances of earning a return on your investments. It is important to remember that stock portfolio building is not a get-rich-quick scheme and requires careful research and analysis. It is also important to remember that stock portfolio building involves risk and you should always be aware of the risks associated with investing in stocks. With the right strategy and research, stock portfolio building can be a great way to increase your wealth.

5. References

1. M, H., E.A., G., Menon, V. K., & K.P., S. (2018). NSE stock market prediction using deep-learning models. *Procedia Computer Science*, *132*, 1351–1362. https://doi.org/10.1016/j.procs.2018.05.050
2. Göçken, M., Özçalıcı, M., Boru, A., & Dosdoğru, A. T. (2016). Integrating metaheuristics and artificial neural networks for improved stock price prediction. *Expert Systems with Applications*, *44*, 320–331. https://doi.org/10.1016/j.eswa.2015.09.029
3. Pan, L., & Mishra, V. (2018). Stock market development and economic growth: Empirical evidence from China. *Economic Modelling*, *68*, 661–673. https://doi.org/10.1016/j.econmod.2017.07.005
4. Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2019). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, *53*(4), 3007–3057. https://doi.org/10.1007/s10462-019-09754-z
5. Shah, D., Isah, H., & Zulkernine, F. (2019). Stock market analysis: A review and taxonomy of prediction techniques. *International Journal of Financial Studies*, *7*(2), 26. https://doi.org/10.3390/ijfs7020026
6. Bintara, R., & Tanjung, P.R. (2019). Analysis of Fundamental Factors on Stock Return. http://dx.doi.org/10.6007/IJARAFMS/v9-i2/6029

7. Bustos, O., & Pomares-Quimbaya, A. (2020). Stock market movement forecast: A systematic review. *Expert Systems with Applications*, *156*, 113464. https://doi.org/10.1016/j.eswa.2020.113464
8. Gandhmal, D. P., & Kumar, K. (2019). Systematic analysis and review of Stock Market Prediction Techniques. *Computer Science Review*, *34*, 100190. https://doi.org/10.1016/j.cosrev.2019.08.001
9. Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, *8*, 80716–80727. https://doi.org/10.1109/access.2020.2988796

10. Nanda, S. R., Mahanty, B., & Tiwari, M. K. (2010). Clustering Indian stock market data for portfolio management. *Expert Systems with Applications*, *37*(12), 8793–8798. https://doi.org/10.1016/j.eswa.2010.06.026

11. Wu, D., Wang, X., & Wu, S. (2022). Construction of stock portfolios based on K-means clustering of continuous trend features. *Knowledge-Based Systems*, *252*, 109358. https://doi.org/10.1016/j.knosys.2022.109358

12. Momeni, M., Mohseni, M., & Soofi, M. (2015). Clustering Stock Market Companies via K- Means Algorithm. *Kuwait chapter of Arabian Journal of Business & Management Review*, *4*, 1-10.

13. Ng, K.-H., & Khor, K.-C. (2016). STOCKPROF: A stock profiling framework using data mining approaches. *Information Systems and e-Business Management*, *15*(1), 139–158. https://doi.org/10.1007/s10257-016-0313-z

14. Mokhtar, M., Shuib, A., & Mohamad, D. (2014). Identifying the critical financial ratios for stocks evaluation: A fuzzy delphi approach. *AIP Conference Proceedings*. https://doi.org/10.1063/1.4903606

15. Agrawal, J., Chourasia, V.S., & Mittra, A.K. (2013). State-of-the-Art in Stock Prediction Techniques. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Energy, 2*, 1360-1366.

16. Pok, W.C. (2017). Analysis of Syariah quantitative screening norms among Malaysia Syariah-compliant stocks. *Investment management & financial innovations, 9*, 69-80.

17. BATES, T. H. O. M. A. S. W., KAHLE, K. A. T. H. L. E. E. N. M., & STULZ, R. E. N. É. M. (2009). Why do U.S. firms hold so much more cash than they used to? *The Journal of Finance*, *64*(5), 1985–2021. https://doi.org/10.1111/j.1540-6261.2009.01492.x
18. Fatila, M., & Syahril. (2022). THE EFFECT OF RETURN ON ASSEST (ROA) AND RETURN ON EQUITY (ROE) ON COMPANY VALUE. *Jurnal Ekonomi*, *11*(02), 835–841. Retrieved from http://ejournal.seaninstitute.or.id/index.php/Ekonomi/article/view/422
19. Adawiyah, N.R., & Setiyawati, H. (2019). The Effect of Current Ratio, Return on Equity, And Firm Size on Stock Return (Study of Manufacturing Sector Food and Beverage in Indonesia Stock Exchange). *Scholars Bulletin*.
20. Patin, J.-C., Rahman, M., & Mustafa, M. (2020). Impact of total asset turnover ratios on equity returns: Dynamic Panel Data Analyses. *Journal of Accounting, Business and Management (JABM)*, *27*(1), 19. https://doi.org/10.31966/jabminternational.v27i1.559
21. Husna, A., & Satria, I. (2019). Effects of return on asset, debt to asset ratio, current ratio, firm size, and dividend payout ratio on firm value. *International Journal of Economics and Financial Issues*, *9*(5), 50–54. https://doi.org/10.32479/ijefi.8595
22. Schwertman, N. C., Owens, M. A., & Adnan, R. (2004). A simple more general boxplot method for identifying outliers. *Computational Statistics & Data Analysis*, *47*(1), 165–174. https://doi.org/10.1016/j.csda.2003.10.012
23. Dovoedo, Y. H., & Chakraborti, S. (2014). Boxplot-based outlier detection for the location-scale family. *Communications in Statistics - Simulation and Computation*, *44*(6), 1492–1513. https://doi.org/10.1080/03610918.2013.813037
24. Shen, X., & Zhu, Z.-J. (2019). MetFlow: An interactive and integrated workflow for metabolomics data cleaning and differential metabolite discovery. *Bioinformatics*, *35*(16), 2870–2872. https://doi.org/10.1093/bioinformatics/bty1066
25. Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, *8*, 80716–80727. https://doi.org/10.1109/access.2020.2988796
26. Wu, X., Yu, G., Zhang, K., Feng, J., Zhang, J., Sahakian, B. J., & Robbins, T. W. (2022). Symptom-based profiling and multimodal neuroimaging of a large preteenage population identifies distinct obsessive-compulsive disorder–like subtypes with neurocognitive differences. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *7*(11), 1078–1089. https://doi.org/10.1016/j.bpsc.2021.06.011
27. Nazarov, D., & Baimukhambetov, Y. (2022). Clustering of dark patterns in the user interfaces of websites and online trading portals (e-commerce). *Mathematics*, *10*(18), 3219. https://doi.org/10.3390/math10183219
28. Ortega-Argilés, R., Piva, M., & Vivarelli, M. (2014). The productivity impact of R&D investment: Are high-tech sectors still ahead? *Economics of Innovation and New Technology*, *24*(3), 204–222. https://doi.org/10.1080/10438599.2014.918440
29. Cornelli, Giulio and Frost, Jon and Gambacorta, Leonardo and Rau, P. Raghavendra and Wardrop, Robert and Ziegler, Tania. (September 25, 2020). Fintech and Big Tech Credit: A New Database). BIS Working Paper No. 887. https://ssrn.com/abstract=3707437

30. Carbone, N. (2019). 200+ Financial Indicators of US stocks (2014-2018). Retrieved from https://www.kaggle.com/datasets/cnic92/200-financial-indicators-of-us-stocks-20142018.