Investing101

Andre Lee
Singapore Management University
andre.lee.2019@mitb.smu.edu.sg

Boey Yi Heen
Singapore Management University
yiheen.boey.2019@mitb.smu.edu.sg

Ng Wee Kien
Singapore Management University
weekien.ng.2020@mitb.smu.edu.sg

ABSTRACT

Most retail investors rely on traditional financial data websites such as Yahoo Finance to retrieve relevant financial data to aid in their investing decisions. However, these sites' offerings tend to be homogeneous and comprise usual information such as price quotes, market commentary and corporate actions. Investors are left on their own to make sense of these information, without tools to further perform actions such as price forecasting or segmenting stocks into groups. These tools could possibly enhance their investing decisions.

Having identified this gap, we developed Investing101, an interactive visual analytics dashboard to allow investors to not only retrieve basic price information of stocks, but also apply machine learning models such as ARIMA forecasting and hierarchical clustering. The dashboard will also allow users to construct bespoke portfolios and see their historical performances. On top of this, the dashboard also provides users with a tool that readily aggregates and interprets candlestick patterns for the user. All these different tools come together to deliver a product that adds value to investors, on top of what traditional finance data websites can provide.

1. INTRODUCTION

Lorem ipsum dolor sit amet, consectetur adipiscing[1] elit. Pellentesque rhoncus ut tellus eu tristique. Duis pharetra velit vitae viverra elementum. Nulla metus dui, pulvinar id enim at, pellentesque semper ipsum. Phasellus cursus dignissim ipsum, sed congue orci pretium quis. Maecenas rhoncus leo a cursus euismod. Fusce a erat eu ipsum tristique tempus at sed tortor. Aliquam erat volutpat. Donec at pretium lorem. Donec pretium nunc id nunc bibendum convallis. Phasellus quis enim id massa feugiat egestas hendrerit sollicitudin nibh. Sed blandit eros id tellus porta, eget ullamcorper urna posuere. Pellentesque laoreet lacus nibh, a mattis libero viverra sit amet. Sed vitae diam interdum, pharetra neque sit amet, dictum lectus.

Donec massa justo, ultricies quis facilisis sed, tristique nec metus. Vestibulum id condimentum diam. Integer semper augue id porttitor ultrices. Cras vulputate felis eu diam porttitor, ac pulvinar nisi imperdiet. Donec eros felis, imperdiet vel malesuada at, varius et quam. Phasellus facilisis non risus eu placerat. Sed ac mollis lorem.

2. LITERATURE REVIEW

2.1 Technical Indicators

In the literature review conducted, it was found that there have been various attempts by users to provide tools for analysing stock price action. Existing charting or dashboarding tools often focus on only simple charting or technical indicator functionalities. Most open source/ free tools only offer the various functionalities in isolation, and users are unable to have a holistic view uniting technical analysis, portfolio analysis, and the other components to allow for complete decision making when making buying and selling decisions.

In this sub-module, the team will put together various submodules spanning across: Charting, Technical Analysis, Correlation Analysis, Clustering, Forecasting, Portfolio Analysis.

2.2 Forecasting

Time series forecasting is not the same as technical analysis and can be seen as a natural extension/logical next step after conducting technical analysis. The main difference is that time series forecasting gives you an exact forecasted price, while technical analysis only predicts the future movement (up/down) of the price (Berdiell, 2015). The fundamental idea of this method is to seek out patterns in the historical stock prices with a hybrid approach. A hybrid approach combine multiple different models to forecast stock prices. For example, the papers of Markowska-Kaczmar and Dziedzic (2008) and Wang et al (2015) both proposed tackling stock price forecasting with an amalgamation of multiple models instead of just relying on one form of forecasting. This has been shown to result in superior forecasting accuracy and performance as compared to using a dedicated forecasting method. The researches of Dey et al. (2016) has also shown similar results albeit with some overfitting issues and a limited testing scenario.

The modification made in building the sub-module to the existing hybrid techniques is that the application will do a forecast based on 5 different models then present the mean of the 3 best performing ones.

2.3 Clustering

Da Silva, Cunha, and da Costa (2005) [sergio?] carried out hierarchical clustering on 816 stocks listed in North and South America from 1997-1999 on the following variables: return, risk, earnings-price ratio, book value-price ratio, sales-price ratio, sales-number of stocks ratio and dividend yield. Their research showed that clusters with the best risk-return profile from the time window where clustering was conducted continued to be so in a second time window from 2000-2001. However, their research did not comprise visualisations of the cluster characteristics. In addition, their research only showed results for one fixed combination of variables. Lastly, there was no insight as to how the winning clusters performed against a benchmark index, which is potentially of interest.

This research provides opportunities to address the limitations of existing literature as described above. The team looks to incorporate interactivity features and appropriate visualisations into the sub-module as part of improving on existing work.

3. CHARTING

4. TIME SERIES FORECASTING

This sub-module of the dashboard is a tool to carry out time series forecasting on a stock. For this sub-module, forecasting will be carried out on a chosen ticker. The intent for this sub-module is to allow the user to identify future trends of a stock and entry/exit opportunities. The example given will be based on the technology company APPLE (AAPL).

4.1 Data Preparation

The sole data source for this sub-module is the tidyquant package on R.

We use the tidyquant package to pull historical prices of the stocks. The extracted data is then broken down into training and testing sets to build the forecasting model. The dashboard is reactive and will be able to pull any amount of data chosen by the user as long as it is available on Yahoo finance.

4.2 Model building

4.3 Case Study

5. CLUSTERING

The next sub-module of the dashboard is a tool to carry out hierarchical clustering on stocks. For this sub-module, clustering analysis will be carried out on 89 current component stocks of the NASDAQ-100 Index. The Nasdaq-100 Index is a basket of the 100 largest non-financial companies listed on the Nasdaq Exchange in the US. Companies listed in the Nasdaq-100 Index largely belong to the technology sector, and some household names include Apple, Google, Tesla and Facebook. COVID-19 accelerated the rise of the digital economy and drove digital transformation initiatives in companies around the world; it also proved to be a boon for technology industries such as cloud computing, videoconferencing and cybersecurity. There has been huge investor interest in technology stocks following their outperformance in 2020 and the ongoing hype of digitisation. Many rookie investors have also entered the fray, hoping to be involved in these companies' growth. We choose to focus on stocks in the NASDAQ-100 Index for this sub-module to provide users with a tool to aid their decisions in investing in the biggest US-listed technology stocks.

5.1 Clustering variables

For this sub-module, hierarchical clustering will be carried out on 5 variables which have been selected to cover a holistic range of financial indicators of a company. The time horizon for a clustering is one year. The 5 variables are as follows: "1-Year Return,""1-Year Volatility,""Dividend Yield,""Return on Assets,""Total Debt to Total Assets."

5.2 Data Preparation

There are 2 main data sources for this sub-module: 1) Bloomberg Terminal 2) tidyquant package on R

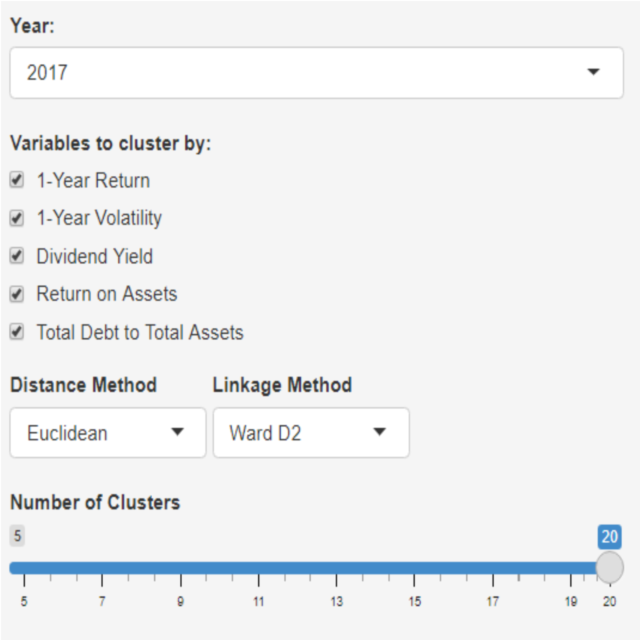
The Bloomberg Terminal is used to pull the historical yearly financial ratios and volatility for the stocks. Bloomberg nicely stores these data points which can be easily pulled into csv format. For each stock in our stock universe, yearly historical values from 2015-2020 are pulled for the variables as mentioned above.

We use tidyquant package solely to pull historical prices of the stocks, which we then convert to yearly returns. The reason we use tidyquant rather than using Bloomberg is because tidyquant package has functions which can help us automatically calculate returns for different periodicities, such as annual or monthly. In this context, we look at yearly returns from 2015-2020 for the stocks in our stock universe.

5.3 Case Study

We demonstrate how to use the Clustering Analysis sub-module of our dashboard by carrying out a round of clustering analysis on a combination of selectable variables.

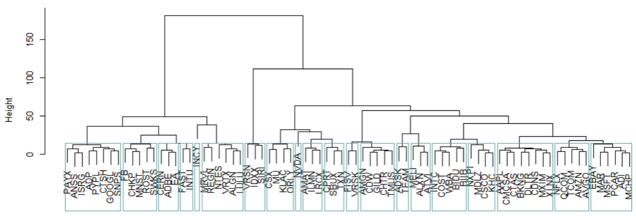
On the side panel of the dashboard, interactivity features are available to allow the user to select the following: "Year," "Variables to cluster by:" "Distance Method," "Linkage Method" and "Number of Clusters."



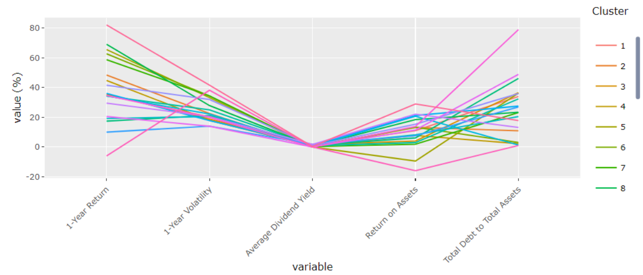
For this case study, we carry out clustering in 2017 on all 5 variables. We select Euclidean to be the distance measure to compute the distance matrix, and Ward D2 as the agglomeration method for the hierarchical clustering algorithm. The resulting dendrogram is cut into 20 clusters.

5.4 Clustering Results

The main panel of the dashboard has 4 main outputs. At the top of the panel is the dendrogram showing how the stocks have been cut into clusters, in this case 20 clusters.



Below the dendrogram is a line graph showing the combination of variable values for each cluster. For each cluster, the value seen on the line graph for each variable is the average value of that variable for all the stocks in the cluster. For example, if Cluster 19 has a value of -6% for 1-Year Return, it means that the average 1-Year Return for all the stocks in Cluster 19 is -6%. There is an interactivity option to isolate the view for a single cluster to strip out the noise from the other clusters.



Next, the user will be able to see the breakdown of each cluster's constituent stocks in a table. For example, we see from the table below for our case study that Cluster 4 consists of the stocks ADP, ANSS, CTSH, GOOGL, ISRG, PAYX, PYPL and SNPS.

	Cluster	Stocks
1	1	AAPL,BKNG,CDNS,CMCSA,CTAS,DLTR,MXIM,XLNX
2	2	ADBE,CERN,EA
3	3	ADI,EBAY,MAR,MCHP,MSFT,PCAR
4	4	ADP,ANSS,CTSH,GOOGL,ISRG,PAYX,PYPL,SNPS
5	5	ADSK,TEAM
6	6	ALGN,LULU,MRVL,NTES,REGN,VRTX
7	7	ALXN,ATVI,MELI
8	8	AMAT,ILMN,LRCX
9	9	AMGN,CDW,CHTR,GILD,TMUS
10	10	AMZN,AVGO,NFLX,QCOM,TCOM

Showing 1 to 10 of 20 entries   Previous 1 2 Next

The final output on the main panel is each individual cluster's return and volatility profile. An additional panel calculating the Returns to Volatility ratio is included, arranged in descending order. This way, users can see instantly which cluster performed the best in terms of risk-adjusted returns for that year. We see that Cluster 4 performed the best in 2017.

	Cluster	Returns to Volatility	1-Year Return	1-Year Volatility
1	4	2.51	44.75	17.84
2	8	2.48	69	27.83
3	2	2.15	48.33	22.45
4	13	2.03	36	17.75
5	5	1.97	65.5	33.16
6	20	1.97	82	41.58
7	3	1.89	35.5	18.78
8	6	1.83	62.67	34.31
9	18	1.74	34.67	19.87
10	1	1.72	34.88	20.22

Showing 1 to 10 of 20 entries   Previous 1 2 Next

5.5 Clustering Results - Backtesting

We then explore a use case as to how to use the clustering results to aid in investing decisions. Following from the finding that Cluster 4 stocks performed the best in 2017, the user may want to use this to explore if outperformance of a cluster continues into the next year 2018. This is an example of momentum investing, a strategy that aims to capitalize on the continuance of existing trends.

Using the sub-module (Weekien), we construct a portfolio consisting of. . . . . . compare against XLK benchmark. . . .outperform!

6. CONCLUSION

Duis nec purus sed neque porttitor tincidunt vitae quis augue. Donec porttitor aliquam ante, nec convallis nisl ornare eu. Morbi ut purus et justo commodo dignissim et nec nisl. Donec imperdiet tellus dolor, vel dignissim risus venenatis eu. Aliquam tempor imperdiet massa, nec fermentum tellus sollicitudin vulputate. Integer posuere porttitor pharetra. Praesent vehicula elementum diam a suscipit. Morbi viverra velit eget placerat pellentesque. Nunc congue augue non nisi ultrices tempor.

7. TABLES

Financial Ratio	Definition
1-Year Return	Return on an investment generated ov percentage of the initial amount of inv
1-Year Volatility	Statistical measure of the rate of fluct over 1 year. Volatility is also known as
Dividend Yield	Shows how much a company pays out to its stock price, expressed as a perce
Return on Assets	A type of return on investment (ROI) profitability of a business in relation to return, the more productive and efficie economic resources
Total Debt to Total Assets	A leverage ratio that indicates the per financed with debt. The higher the ra leverage and financial risk

References

[1] Fenner, M. 2012. One-click science marketing. Nature Materials. 11, 4 (Mar. 2012), 261–263.