DeepAudit - Increment 2



**Thomas Bienasz**

**Xinrui Chen**

**Keerthana Kannan**

**Kurtis Lazenby**

**Ella Salafranca**

**Nathan Soares**

30-09-2021

COMP3850 - Computing Industry Project

# 

| **Revision Table** | |
| --- | --- |
| *Section:* | *Changes Made:* |
| Solutions Architecture | Added a before snapshot of the data and highlighted potential issues that may have occurred if this transformation hadn’t happened. |
| Feature Engineering | Providing a table of the features that will be used for modelling for all ASX listed companies. |

# 

# 

# 

# PROJECT ASSUMPTIONS

Throughout this project, there have been several instances where the team has needed to use their subjective judgement in order to make detailed decisions in the absence of complete information. These assumptions are listed below.

## Project Plan Assumptions

1. All meetings and correspondence between the team and the client take place exclusively via online channels due to the current COVID-19 pandemic.
2. The project will require no funding throughout the duration of the project lifecycle.
3. All resources and plans of action have been approved and recognised by the client.
4. All roles and responsibilities described are consistent with the submitted project plan, though their allocation may differ between each deliverable as discussed amongst the team. This has been organised in a manner that prioritises efficiency in accomplishing the project’s objectives.
5. All team members will abide by the guidelines identified within this project plan document.
6. Failure to identify changes made to the project deliverables and tasks within the determined project timeline may result in major project delays.
7. The project will integrate feedback received from the client during each meeting, which will then be included in the next deliverable submission.
8. The project scope will remain as is, unless stated otherwise. Any changes must first be discussed with the client and receive their approval; revisions must be documented in accordance with the granted approval.
9. All deliverables and output generated will be handed over to the client at project completion.

## Data Assumptions

1. All data is sourced from reliable platforms, or substantiated with appropriate justification if otherwise.
2. All data gathered is valid and accurate, as well as relevant to the project.
3. All data gathered, as well as the project’s files and models will be stored in the team’s GitHub repository for the project.
4. The team will thoroughly screen and select financial statements for incompleteness, and gather financial data from a range of industries to maximise the utility of the model and its application to new data.
5. Data gathered should follow the same naming conventions, if not it needs to be modified before being used to run models with.
6. Financial data for both half yearly and full yearly needs to be uploaded so they can be combined into a separate workbook “Workboooks\_combined”

## Model Assumptions

1. Beneish M Score:
   1. Gross Margin Index: Since the gross margin ratio wasn’t available in the financial statements, net profit margin was used as it was the most closely related substitute.
   2. Sales Growth Index and SG&A Index: The ’Total Revenue excluding interest’ item was used instead of ‘Sales(Revenue)’, as it provides a more complete measure of revenue by including ‘Other revenue’
   3. SG&A Ratio: Instead of the ‘SG&A expense’, the ‘Operating Expense’ item from the P&L sheet was used as these items are synonymous with each other.
   4. TATA Ratio: Instead of ‘Net Income’, the ‘PreTax Profit’ from the P&L sheet was used as these items are synonymous with each other.
   5. DSRI: Due to completeness issues in the datasets, the DSRI ratio may have to swap between the combined & separated datasets depending on the particular company’s reporting structure in their financial statements.
2. Due to the unique financial statements of banks and other financial institutions, they will likely not be included in theBeneish model. This is due to a tendency for critical data to be missing which is necessary for different models to function.
3. Time based indices should generally use the combined dataset (i.e. HY and FY), as this provides a more frequent interpretation of the change in the indices over time. With this being said, the separated HY and FY datasets may be of use for identifying seasonal patterns in the data.

# 

# 

# 

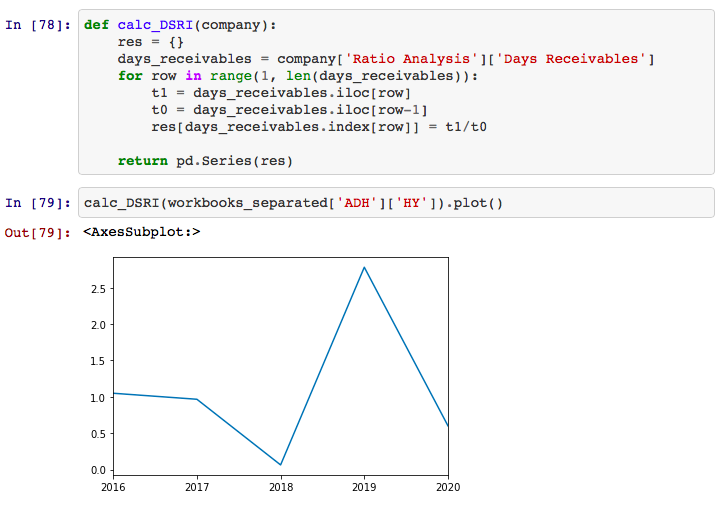
# 

# PROTOTYPE (Increment 1)

In this section we will introduce how our prototype looks and functions. Our team has chosen the Beneish M-score model as our primary solution to detect fraudulent activity in financial statements. In order to calculate the Beneish M-Score, a variety of financial ratios including DSRI, GMI and several others are weighted and summed for each financial period. If the M-Score for a company exceeds a certain threshold, that period is considered to be anomalous. Here we examine the prototype’s components more closely:

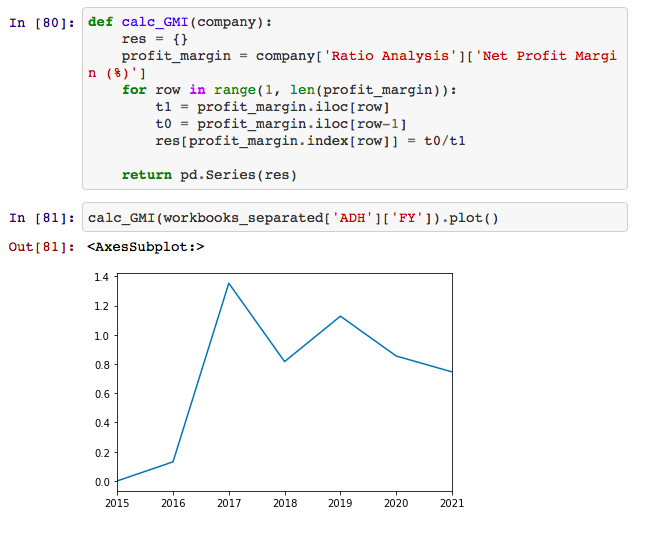
## 

## DSRI (Days Sales in Receivable Index)



In the above screenshot, we calculated the ratio of days sales in receivables for a specific period by the previously known period.

## GMI (Gross Margin Index)

**

The above metric calculates the ratio of the previous period’s net profit margin to the current period. As a result of gross profit data missing from our datasets, we opted to use the ‘Net Profit Margin’ as a proxy for our calculations.

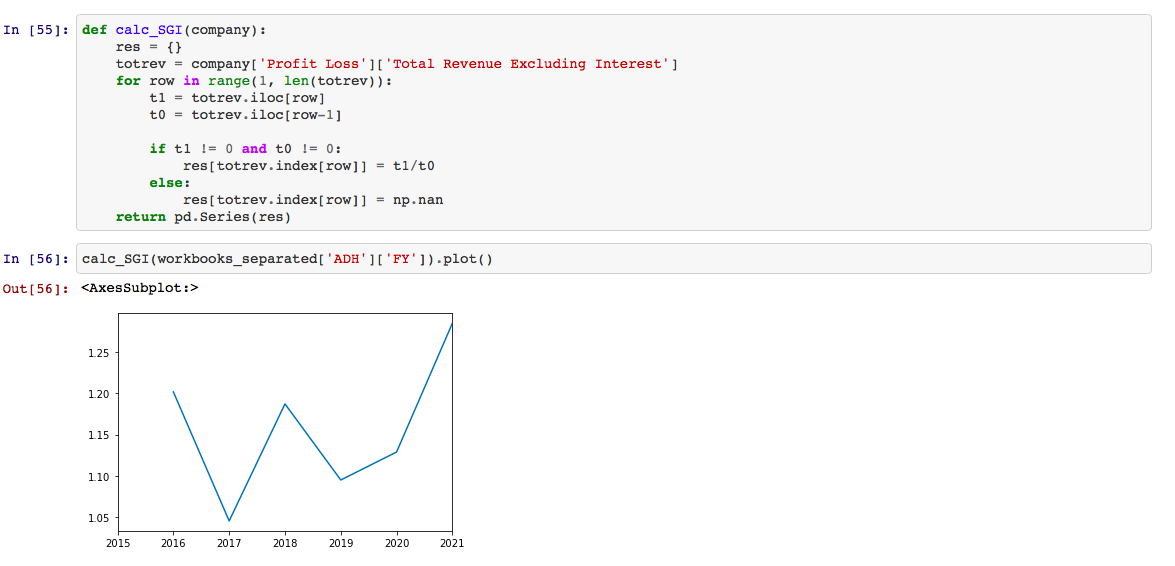
## 

## AQI (Asset Quality Index)

## 

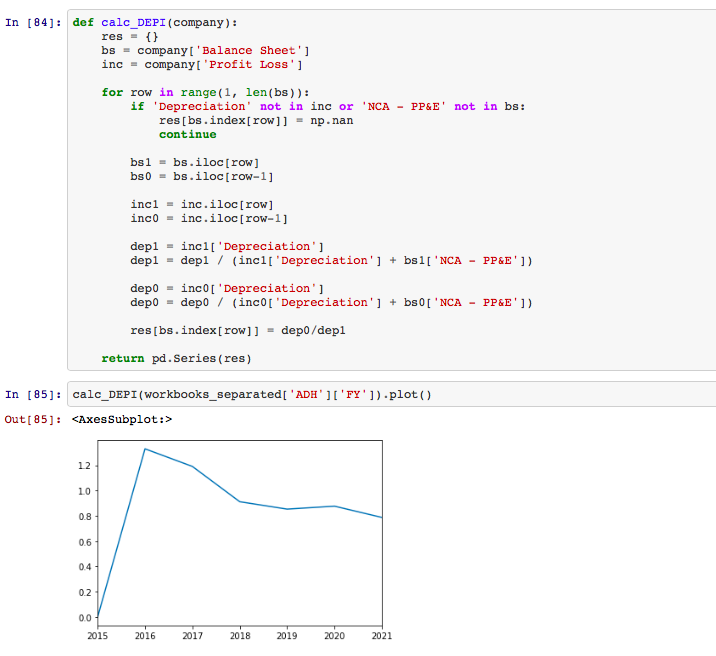
This method was used to measure the quality of a particular company's assets and was calculated by the ratio of non current assets, property and equipment to total assets, drawn from the balance sheet from each company's financial statement

## SGI (Sales Growth Index)



The above calculation measures the growth rate of a particular business’ revenue from sales, drawn from the profit and loss statement. Consistently high growth rates over time may indicate anomalous activity as growth rates tend to normalise closer to the rate of inflation as a firm matures.

## DEPI (Depreciation Index)

**

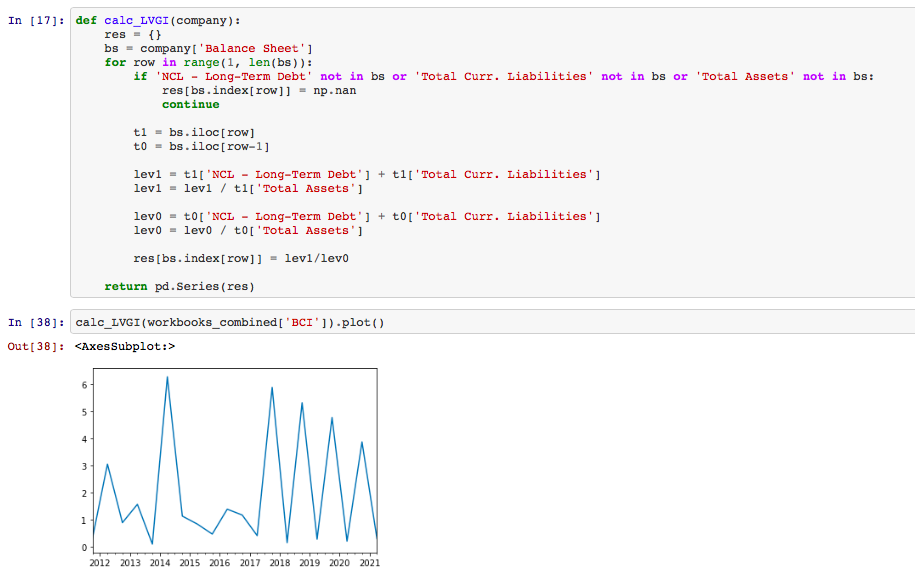
DEPI shows the ratio of the current depreciation rate against last year’s depreciation rate. This index was used to determine if companies' assets are depreciating at a slower or faster rate.

## TATA (Total Accruals to Total Assets)

## 

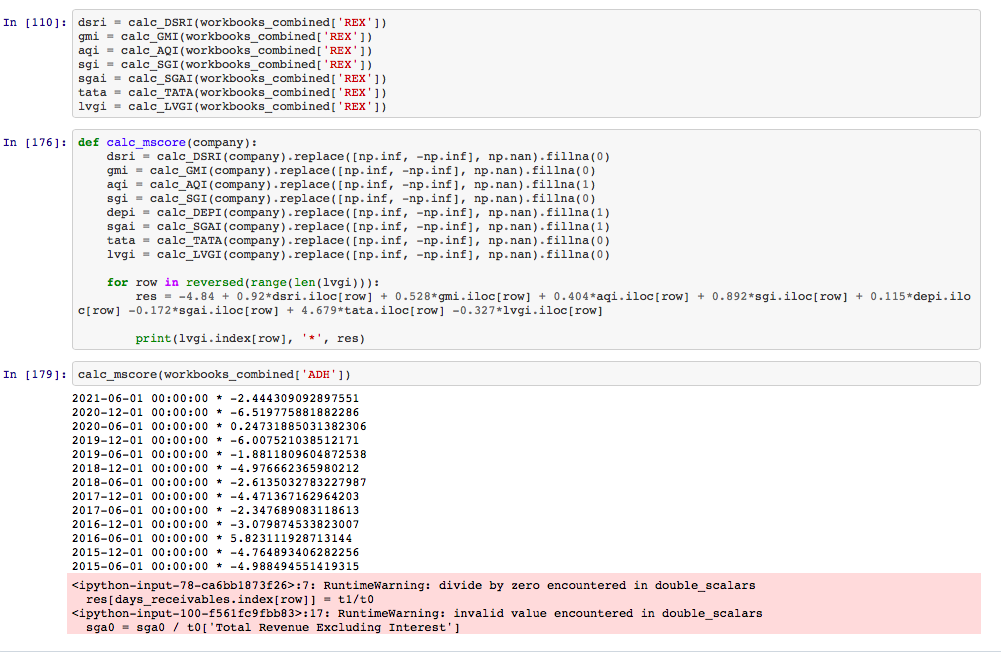
The above ratio is calculated as the change towards working capital accounts apart from cash less depreciation.

## LVGI (Leverage Index)

**

The LVGI is measured as being the total debt relative to total assets. In order to calculate this we used the ‘NCL - Long-Term Debt’, ‘Total Current Liabilities’ and ‘Total Assets’ from each company's balance sheet.

## M-Score Calculation



Finally, we implemented all the above ratio calculations and took their weighted sum in order to find the M-score. The list of dates and numbers corresponds with the M-Score calculated for that period of the company’s financial statements. Where values exceed -1.78, then that period is considered to be anomalous and likely to reflect fraudulent reporting. In the above example, we found the M-Score for Adairs Limited, a retailer of home furnishings. From the results in cell 179, the June reporting period for 2016 and 2020 were identified as anomalous, as the M-Scores exceeded -1.78. This coincides with the period in 2016 in which Adairs was fined by ASIC for failing to disclose that their forecasted figures for the period were likely to be materially lower than the market consensus (LaFrenz, 2017), demonstrating that the model was effective and accurate in identifying anomalous reporting periods.

The semantic meaning behind each ratio is explored in further detail in the Algorithms and Models section of this document.

# 

# 

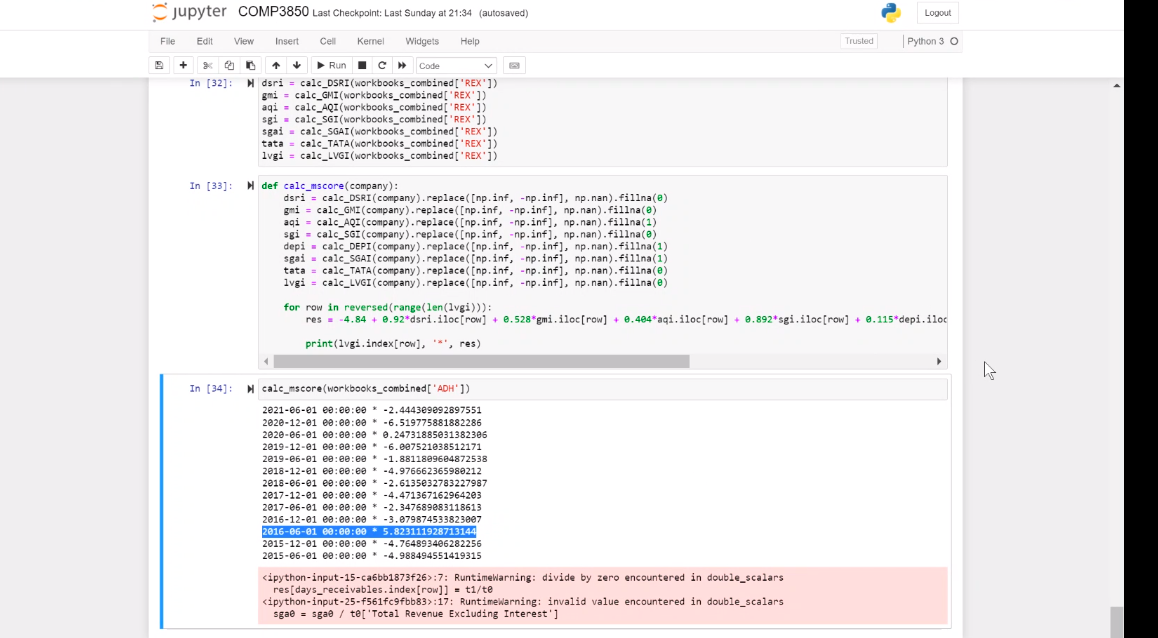
# 

# 

# Sponsor Meeting and Feedback

## Comments made by our sponsor

* To keep in mind that the end user needs to be able to understand the model’s output. In practice, we need to ensure that a normal employee at BT Financial Group can interpret and understand the results to identify whether a period likely contains fraudulent reporting or not.



* For example as above, consider using a heatmap or breakdown of each of the Beneish M-Score inputs to give users a clearer indication of whether fraud occurred and how it was detected.
* Try to make the front-end of the model simple so users can understand how to use it
* Consider that this project and our research will be handed over to BT Financial Group to further develop it. In effect, be clear in explaining what the code is trying to accomplish through effective comments and be clear why particular approaches and models were used and whether they were successful.
* Add short descriptions to plots to better communicate the trends, outputs and ideas derived from the data and model
* Make note of any issues, limitations, defects or loopholes in the model so that expectations of the model’s results are effectively managed, considering that this tool may be used in a legal context to support a case that a firm engaged in fraudulent reporting. We should be confident in the model’s results and ensure they are as accurate as possible.
* Consider methods of testing the accuracy of our model such as testing our model on new datasets. While we may not know whether in reality a particular firm has engaged in fraudulent reporting or not, we can identify potential anomalies and cross-reference this with penalty notices in future to establish the effectiveness of the model.
* We should also test the calculations and design processes of the model beyond testing its rules and logic to gain a higher degree of confidence in its accuracy.

## How will DeepAudit Respond to the Client’s Feedback?

The feedback received from our sponsor was very positive and will be beneficial for our future analysis. He firstly commented on the visualisation for our m-score analysis and mentioned to make the final preview of our results to be more clear for the end user to understand. It should clearly highlight which values come under being ‘manipulated data’ and for which period of time. Hence, the Deep Audit team will look at applying heat maps, scatter plots and any other descriptive graphs that shall make the presentation of the results more simple and readable. These plots will clearly indicate whether a score is either less than or greater than the -1.78 threshold.

Another important comment mentioned had been about adding in more descriptive comments towards our source code. Our team has already started to implement more detailed comments to our code and will continue to apply this methodology. This would make code maintenance a lot easier, detecting future bugs much quicker, as well as provide our user IT team more clarification on what the algorithm computes behind the scenes.

Another suggestion from our sponsor was to include short descriptions under all the plots in order to show multiple trends among the analysis. Deep Audit shall include more detailed descriptions next to our created plots and make sure it provides a great amount of detail for the end user.

Lastly, our sponsor noted to make sure to run multiple tests on a wide variety of financial statements and observe if our algorithms can still detect fraudulent activity without previous knowledge of such actions occurring in the data. Our team will tackle this suggestion once we are confident that our model and algorithms perform the required analysis correctly to our gathered data sets. When satisfied, then we will move on to other financial statements to execute further testing.

# 

# 

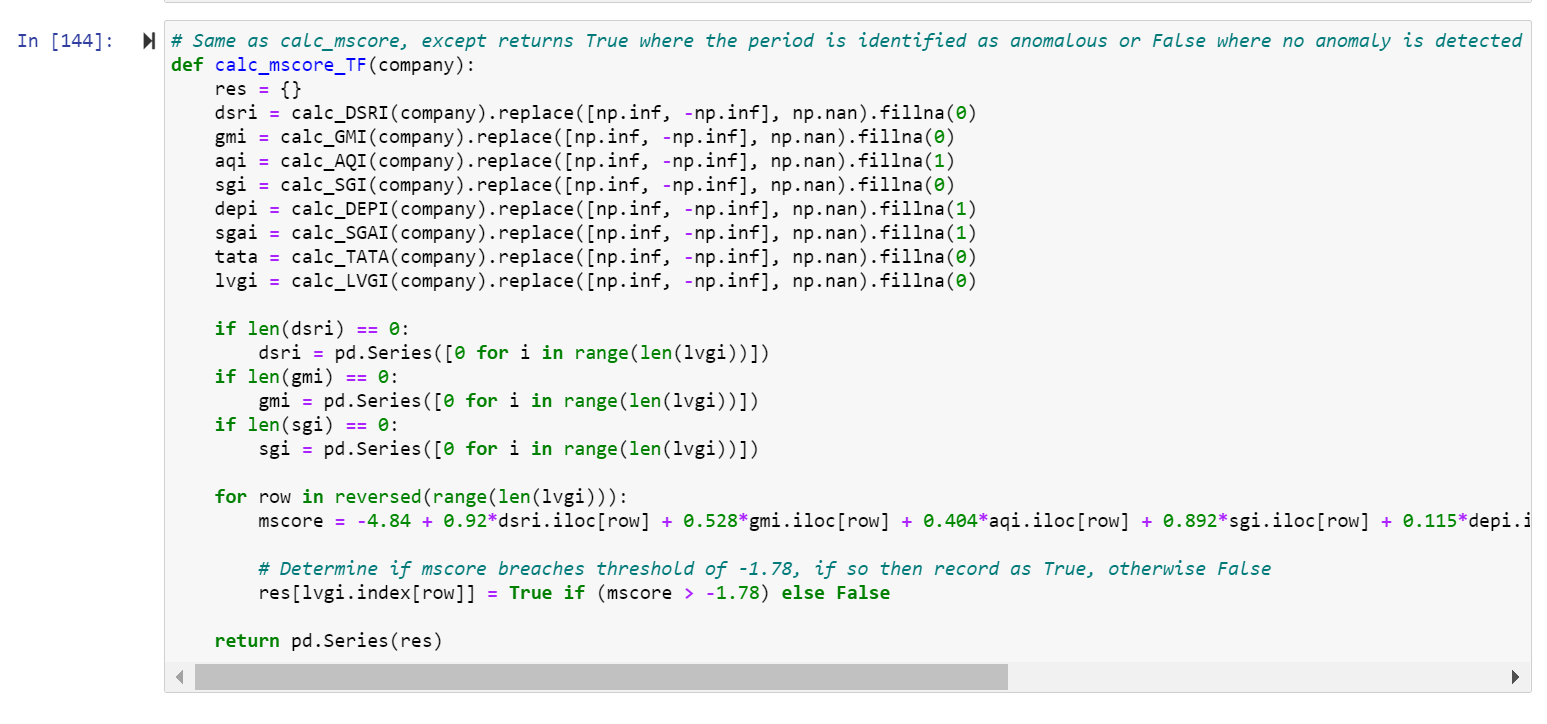
# 

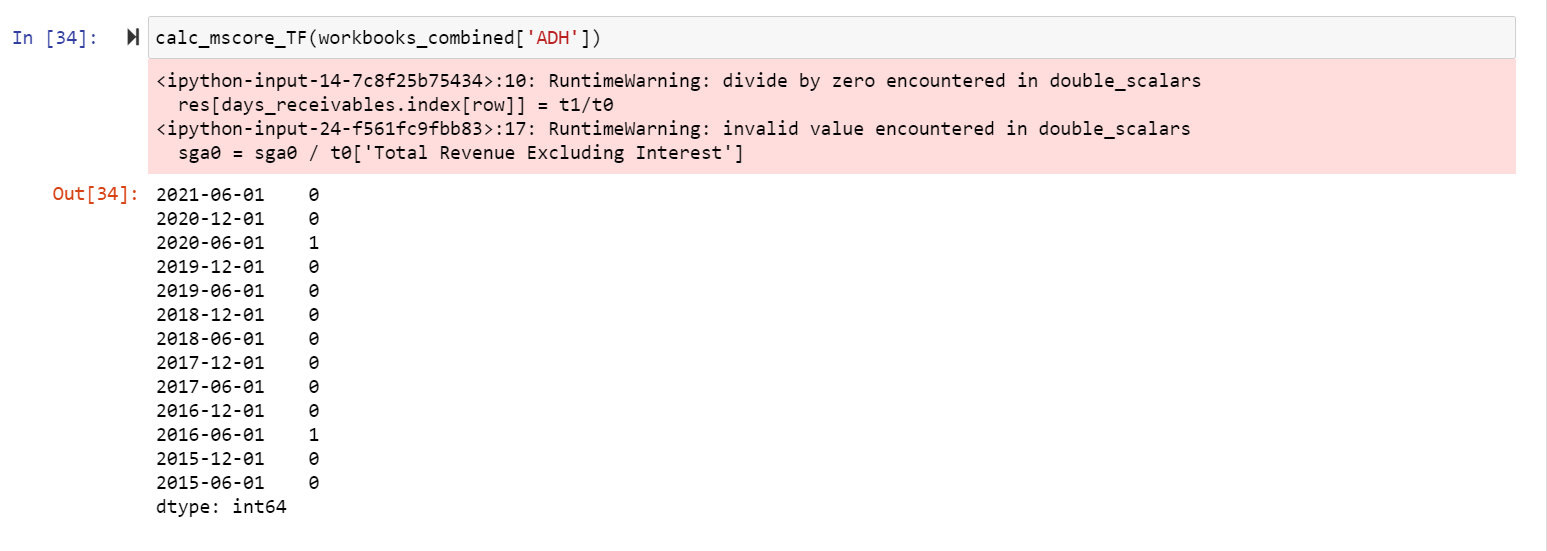
# 

# 

# Prototype (Increment 2)

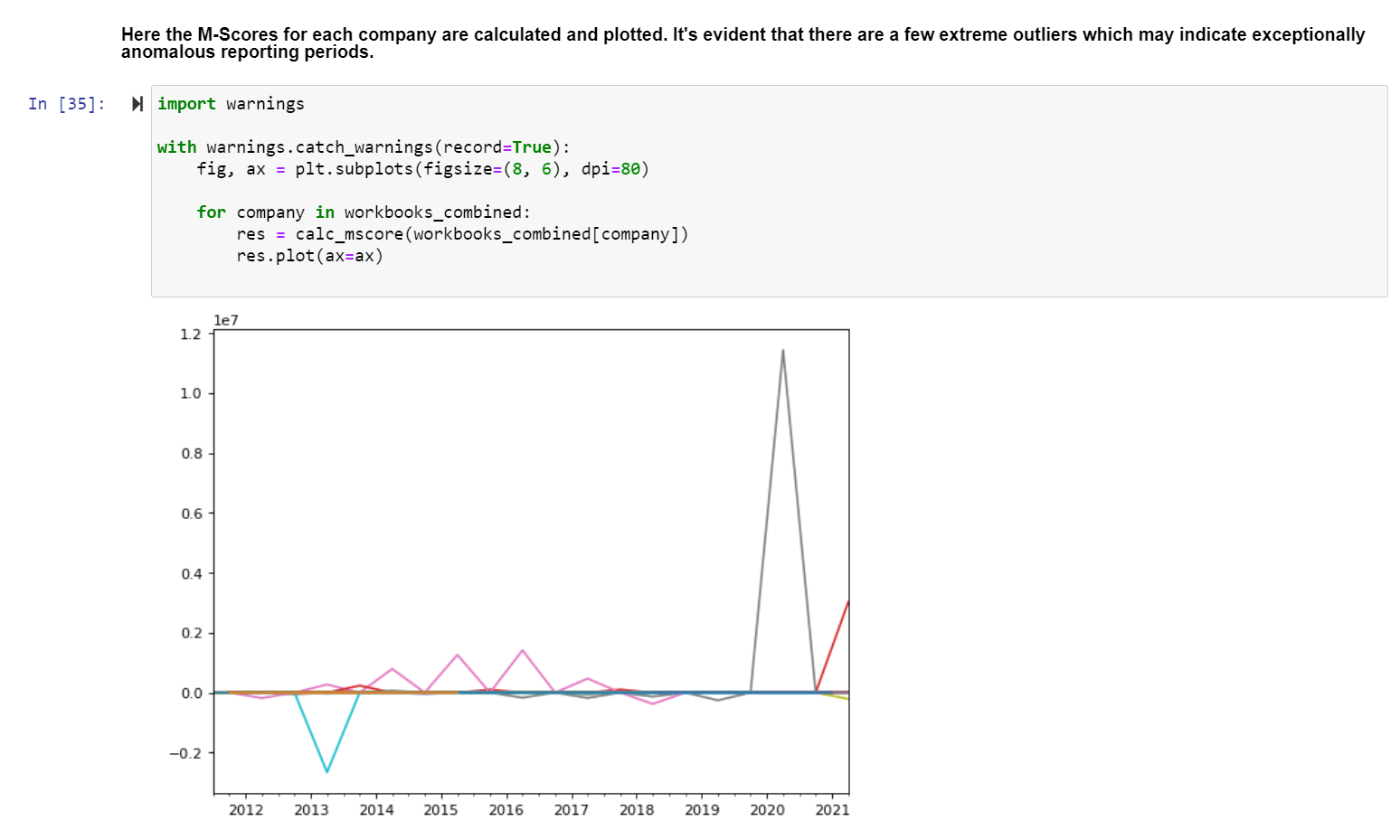
Based on our sponsor’s feedback we decided to implement the additions to our project.





This function aims to easily identify which m-scores are over the threshold of -1.78 which indicates potential manipulation in the financial statement. The 0 represents no manipulation according to the Beneish M-score and 1 represents a high possibility of manipulation in that year.

We also decided to add a visual plot to showcase the great variance amongst all companies’ financial statements, where some have incredibly higher m-scores compared to others.



To elaborate our findings of m-scores for each company, we utilised penalty notices from ASIC to compare with our fraud detection.



# Feature Engineering

## Trends and Patterns

As part of the data preparation process, the team conducted an extensive investigation into the data to identify what data was available, what was missing and the format the data was presented in. Alongside this process and from further analysis, the team inevitably identified several trends and patterns in the financial statement data.

**Half Yearly vs Full Yearly Data vs Combined Data**

We identified that the interim (half-yearly) and full-yearly reports typically each contained data that was exclusive to that reporting period. For example, the datasets for companies such as ADH often only reported their total investing cash flows in the interim reports, which had an impact on the calculation frequency of the TATA ratio for the Beneish M-Score calculation.

It also became quickly evident that the value of time-based indices such as SGAI, SGI and AQI among others would change depending on whether they were calculated based on a rolling 6 month period (i.e. by using interim and full-yearly reports consecutively) or a rolling 12 month period (e.g. by comparing this years half-yearly report to last year’s half-yearly report). This does provide us with the opportunity however to better explore the change in the data across time periods while taking into account the impact of seasonality on the data. For example, a retailing business may consistently report higher half-yearly earnings growth as this accounting period would align with Christmas and New Year, being a time of significantly increased consumer demand as sales across industries.

**Financial Institution Data (i.e. Banks)**

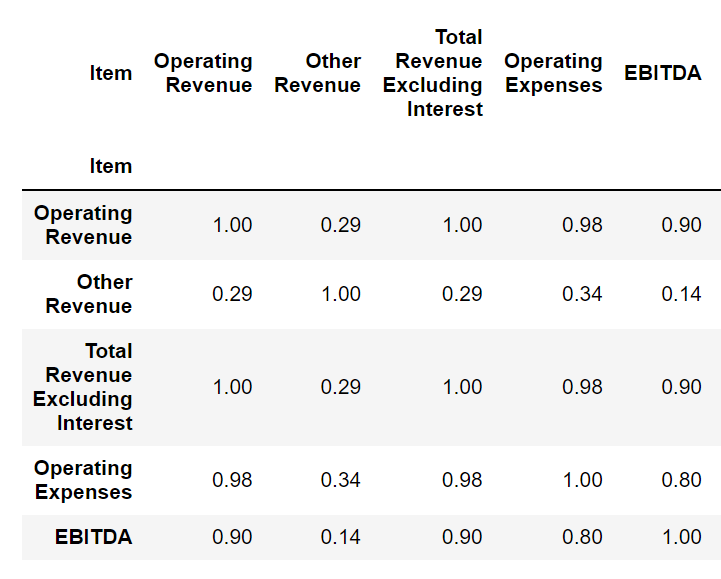
Through our investigation into the different companies and industries our sample data came from, we identified that banks such as ANZ, WBC and CBA had significantly different reporting items compared to retail or service based businesses. The team determined that since financial institutions typically did not generate revenue from sales but instead through earnings on interest, that it would not be possible to correctly calculate the ratios considered in the Beneish M-Score calculation.

This also presents an issue for the cluster analysis if we choose to compare retail businesses and financial institutions in the same model. This is because the features present in retail businesses including sales revenue, capital expenditures and depreciation are missing from the financial statements of the banks. In response, we plan to investigate whether other items in the financial statements can serve as proxies for these missing values.

**Feature Select and Correlation**

One concern the team has regarding the use of a clustering model for identifying anomalous reporting periods is the extremely large number of features present in each company’s financial statements, reaching as many as 100 to 175 for some firms. Because the computational complexity of a clustering model increases exponentially with the number of features, it is infeasible to use all these features in the clustering model, as testing the model would take an exorbitant amount of time. Instead, we must choose a smaller selection from the available features or create our own (as with the Beneish M-Score ratios).

Selecting and generating features is in itself a complex and multifaceted issue, though a typical approach for reducing the number of features is to identify those features which are highly correlated with each other and keep only 1 of these correlated features. This can be visualised with a correlation matrix, as partially shown here for ADH’s income statement:



From this matrix we can identify that Operating Revenue and Operating Expenses are highly correlated with a correlation coefficient of 0.98. In other words, almost all of the movement in ADH’s Operating Expenses can be determined from its Operating Revenue. For this reason, if we were to conduct a cluster analysis on ADH’s income statement data overtime, we can omit the Operating Expenses variable to reduce the total number of features in the model and reduce its computational complexity, allowing us to test the model more rapidly. In order to replicate this procedure across all of the financial statement items for ADH we would have to determine a threshold for the correlation coefficient at which we can consider removing correlated features, such as 90% for example. In such a case for ADH, the EBITDA feature would also be removed as it is captured by the Operating Revenue.

However this is a flawed approach considering that the correlated features within a particular company may differ significantly from other firms - for example, Operating Expenses might only be correlated with Operating Revenue with a coefficient of 0.2 for REX instead of 0.98, and if the Operating Expense variable was to be removed, then potentially significant information would be left out of the model, reducing its accuracy. For this reason, any variable that is removed must have a consistently high correlation coefficient across all firms to prevent useful information from these variables being lost.

## Data Characteristics Table

Below is a data characteristic table, where it displays each individual point of data which is applicable to our modelling system. It specifies the spreadsheet tab and data item name, which will help identify it within the excel spreadsheet. All of the items listed below are the most important variables when calculating the M-Score, which is our main modelling system.

| **Spreadsheet Tab** | **Data Item** |
| --- | --- |
| Ratio Analysis | Day Sales in Receivable Index |
| Ratio Analysis | Gross Margin Index |
| Balance Sheet | Asset Quality Index |
| Profit Loss | Sales Growth Index |
| Profit Loss | Depreciation Index |
| Profit Loss | Sales, General & Administrative Expenses Index |
| Cash Flow | Total Accruals to Total Assets |
| Balance Sheet | Leverage Index |

## Data Characteristics

**Data Description and Data Pipeline**

Data characteristics are an important component of our project, as it allows us to assess whether information is beneficial in certain situations. When manually assessing the data sets, we used a metric to determine which data would fit our modelling system. This metric consisted of accuracy, completeness, reliability, relevance and timeliness.

Once this metric was established, we were able to apply this to the planning of our pipeline. This was especially relevant in stages 1 and 2 as this was when the research and transformation of the data occurred, as it was necessary to ensure that the data we were downloading was as relevant and reliable as possible. In order to make sure that the data we were downloading was relevant we needed to refer back to our modelling system, which required the input of ASX financial reports. Therefore, the relevance of the data was determined by whether the data provided in the sheets were applicable to our models. The reliability is solely related to the source of the data, which was sourced from MorningStar, this source provides large amounts of financial information for all companies listed on the ASX. In terms of ensuring that the data is accurate and complete, this occurred during the transformation and cleaning stage. Any incomplete or inaccurate data was transformed within the Jupyter Notebook, both pre and post transformation a manual inspection of the data was conducted to determine the accuracy and completeness of our data sets. Finally the timeliness of the data was decided in the research phase, as it became apparent that a longer sample of time would yield better results, therefore a 10 year data sample was chosen, both Full-Yearly and Half-Yearly reports.

**Financial Items Naming Scheme**

Because each firm has slightly different reporting standards, the financial item names between companies differ to some extent. The extent of these variations ranges from simply the capitalisation of the words (e.g. Operating revenue vs Operating Revenue), to abbreviations of the same item (e.g. Net Operating Profit after Tax vs NPAT), to having completely different names for the same item (e.g. Sales Revenue vs Operating Revenue). To address this, as part of the data cleaning process we have identified those item names which are duplicates of each other and have synthesised a consistent naming scheme for these item names across all the financial statements.

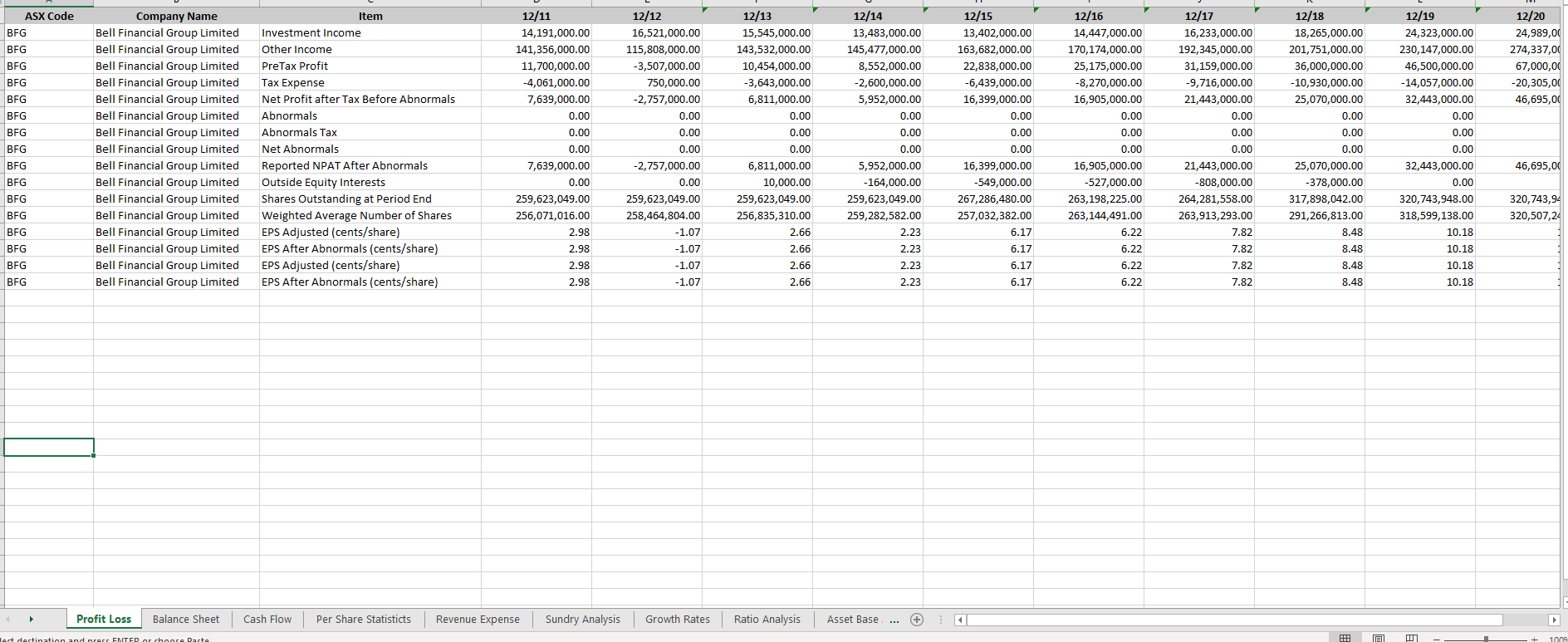
## Feature Names and Definitions

**Summary of Beneish M-Score Features**

A detailed description of the Beneish M-Score features is provided in the Algorithms and Models section of this document, and for the sake of brevity is only summarised here. Fundamentally, the DSRI, GMI, AQI, SGI, DEPI, SGAI, TATA and LVGI indices are all features that have been generated using the available data from each company’s income statement, balance sheet, cash flow statement and ratio analysis sheet where appropriate. These values are all fractional values with no set range, however it should be noted that any time series index, such as AQI, is only calculated when there is data for 2 consecutive periods - where this isn’t the case, the calculated ratio for that particular period is set to 0 or 1 depending on the definitions made by Beneish.

# Solution Architecture

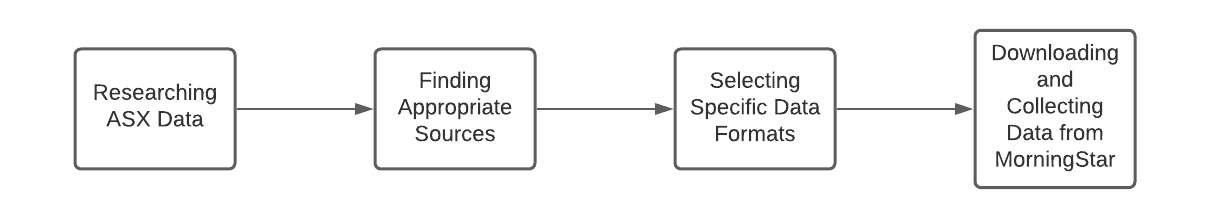
## Before Transformation - Issues that may have occurred



This is an example of a spreadsheet directly downloaded from MorningStar in csv format. As seen in the original data there are duplicate rows, 0 sum results in some of the columns and poorly named rows which do not align with the rest of the data set. However, all of these issues can be solved with our simple transformation process as seen down below. This transforms the data into an easier to read format and is more concise as duplicates and inconsistencies are removed.

## Stage 1 - Researching Data (Week 1-4)

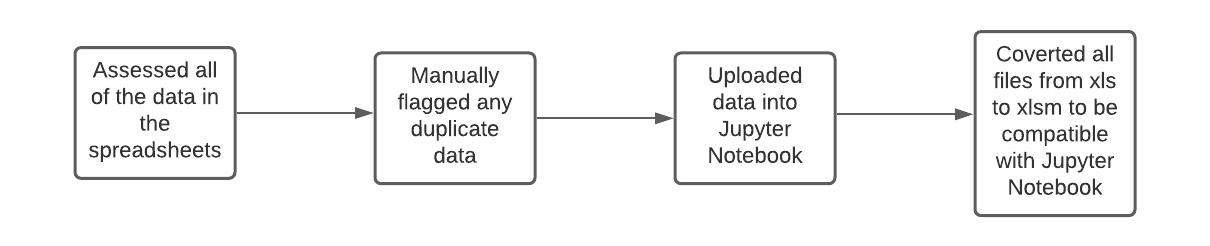
In Stage 1, we conducted in depth research into specific ASX data that we needed to extract in order for our models to detect fraudulent activity. Our resource for this stage was <https://www.morningstar.com.au/>, this was more time consuming than expected because there were a lot of different variations of data available for all ASX listed companies. MorningStar mainly offered spreadsheets in excel format which contained mostly financial data and some pdf files containing company announcements and infringements. We selected financial excel data from MorningStar, both Half-Yearly and Full-Yearly reports over a 10 year time period, as these were the most relevant data sets for our proposed modelling scheme. Once the formats were selected we downloaded spreadsheets and collated them into one folder ready for data transformation.



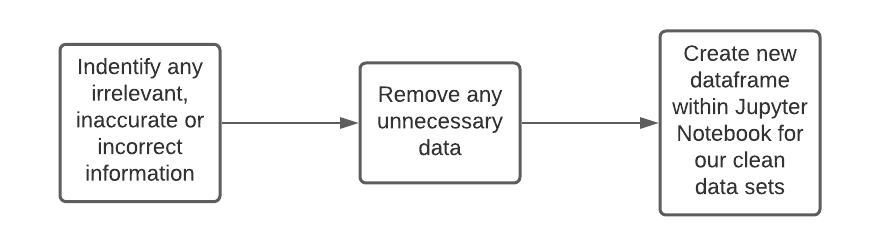
## Stage 2 - Data Transformation (Week 5)

In Stage 2, we manually assessed all the data in the spreadsheets to confirm that we were working with the correct data sets. Before all the data sets were uploaded into our program, we created a naming system for all the files to reduce complications. The system was the ASX ticker followed by either HY (half-yearly) or FY (full-yearly), for example ANZ\_FY. Next we manually flagged and removed any duplicate rows that may make our modelling systems inaccurate.

Once this step was completed we uploaded the entire data set into Jupyter Notebook and created a function that simultaneously converted all the excel files from xls to xlsm, thus allowing us to properly analyse the data and prepare for cleaning. A resource that we referred to in this stage was the project document provided by the client tabulating companies that have been flagged as well as the corresponding ASIC media announcement of the penalties incurred by the company found on <https://asic.gov.au/about-asic/news-centre/find-a-media-release/>.



## Stage 3 - Data Cleaning (Week 6)

In Stage 3, we used Jupyter Notebook to determine whether there was any unnecessary data located in the existing spreadsheets. If there was anything noticed, such as missing or corrupted data, this would be removed. Once the cleaning process was completed new data frames were created in Jupyter Notebook in order to preserve the cleaned data sets. 

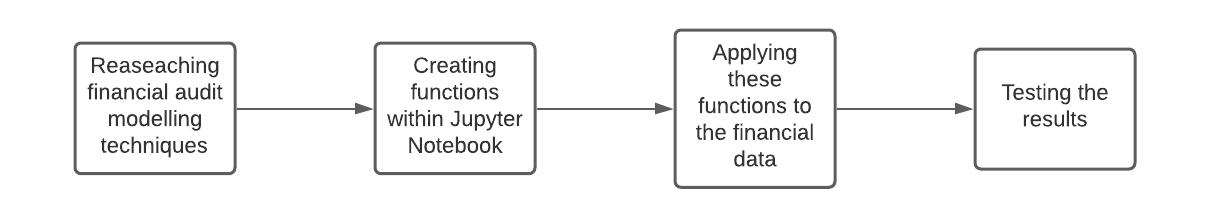
## Stage 4 - Modelling Data (Week 7-8)

In Stage 4, research into different modelling techniques was conducted, allowing us to narrow down a few specific techniques which were best fit for the data sets in our possession. The Beneish M-Score was selected as our main modelling technique, which was found through this resource:

<https://www.researchgate.net/publication/252059255_The_Detection_of_Earnings_Manipulation>.

This allowed us to gain a deeper understanding of how this modelling system works and how to properly apply it to our data sets. The research component was the most time consuming part of this stage as the concepts were hard to grasp at first. Once the techniques were fully understood, we were able to apply our new found knowledge by creating several different functions within Jupyter Notebook, which would allow us to model and calculate the Beneish M-Score.

Once all of these functions were completed, we were able to apply them to our newly created data frames. Finally, the data sets will be run through these functions to determine whether any of the individual data sets display an abnormal Beneish M-Score. If this is the case, the results are then compared against the infringement notices and media release dates, to determine whether there has been any fraudulent reporting within the financial data.



# Algorithms/Models/Methods

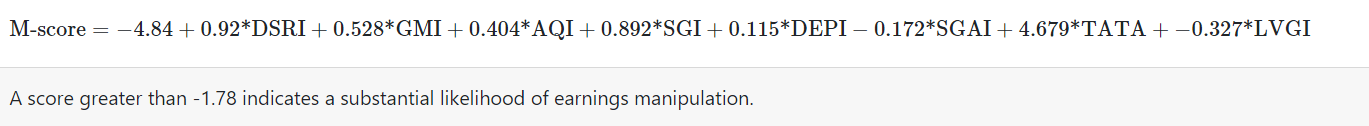
## Beneish M-Score

The statistical model we have chosen to use is able to identify distinct characteristics and provide estimates to detect manipulation in financial statements. An extremely vital part of automating the auditing process of financial statements is being able to detect the company’s manipulation within these statements. Both half yearly and full yearly data for a company was used in the Beneish M-score calculation to detect the likelihood of the company manipulating within its financial statements.

We are basing our initial analysis of the financial statement using Beneish M-Score to be able to identify the specific year(s) a company is likely to have manipulated its data. The calculations were based on raw financial data collected from MorningStar to attain more accurate results. The Beneish M-Score is calculated using very important variables(ratios) which will be discussed further in the document to identify inflation of revenues, deflation of expenses and other characteristics such as the sales growth. (M.Beneish,1999) By using financial data that had been collected from the past 10 years, we were able to pin-point the years and period in which a specific company would have manipulated data in the financial statement; this was later confirmed by checking the ASIC media release.

## 

## Weighted Factors of the Beneish M-Score



**1. Days Sales in Receivable Index (DSRI)**

The Days Sales in Receivables Index is a measure used to calculate the ratio of days sales in receivables for a particular period by the previous period. DSRI can be found within the ‘Ratio Analysis’ worksheet for each company as ‘Days Receivables’. Alternatively to calculate the Days Receivable, the following formula can be used: ((Receivables/Sales) \* 365). The DSRI should generally be well balanced with slight fluctuations for each period but sudden increase is indicative of revenue manipulation.

**2. Gross Margin Index (GMI)**

The Gross Margin Index is a metric that gathers the ratio of the prior period gross margin by current period gross margin. This ratio is usually found in the ‘Ratio Analysis’ worksheet of the financial statements. Due to the missing data, we decided to select the Net Profit Margin instead, as it is a good approximation of the company’s profitability. The main purpose of this index is to flag any deterioration in earnings, to further investigate the company’s motive to inflate profits.

**3. Asset Quality Index (AQI)**

The Asset Quality Index is a ratio calculation based on a company’s assets which include the non-current assets, PP&E and total assets. The Asset Quality Index is calculated using the data from the ‘Balance Sheet’ worksheet of the company’s financial statement. Asset Quality Index (AQI) is calculated using the current period asset quality over the prior. Asset Quality is measured using the following formula: (Total Assets - (Current Assets + PP&E)/ Total Assets. This index could suggest any excessive capitalization of costs within the company for a particular period in the financial statement.

4. **Sales Growth Index (SGI)**

The Sales Growth Index is computed using the company's sales(revenue) for the current period by the sales(revenue) of the previous year. The data needed for this calculation is obtained from the ‘Profit Loss’ sheet from the financial statement for each company. The value selected to be used for the sales(revenue) calculation is ‘Total Revenue Excluding Interest’. Drastic changes in between each period in growing companies could suggest that there is some manipulation trying to meet earnings targets or high capital needs.

**5. Depreciation Index (DEPI)**

The Depreciation Index represents the ratio of depreciation of the current period from the previous period (Prior depreciation Rate/Depreciation rate). To find the depreciation rate, the data for ‘Depreciation’ from the ‘Profit Loss’ sheet and ‘PP&E’ from the ‘Balance Sheet’ is used. To calculate the rate of depreciation, ‘Depreciation’ is divided by the addition of ‘Depreciation’ and ‘PP&E’(Depreciation Rate = Depreciation/ (Depreciation + PP&E)). If the ratio being greater than 1 indicates that the depreciation rate has lowered which could reflect the increase in income due to policy changes or adopting new methods within that period.

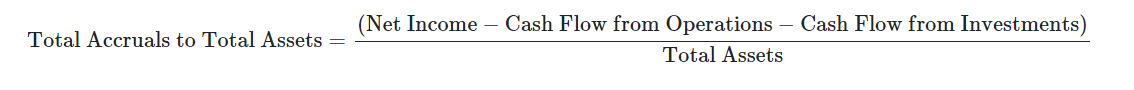
**6.** **Sales, General & Administrative Expenses Index (SGAI)**

The Sales, General & Administrative Expenses Index is measured using the ratio of SGA expenses for the current period over the prior (SG&A Ratio/ Prior SG&A Ratio). The ‘Profit Loss’ sheet from the financial statement of each company was used to select the ‘Operating Expense’ and ‘Total Revenue Excluding Interest’. When the efficiency of SGAI decreases, it could make companies predisposed to manipulating earnings.

**7. Total accruals to total assets (TATA)**

Total Accruals to Total Assets focuses on the cash flow of the company against the total assets. There are a new sheets that are necessary to calculate the TATA, from the ‘Profit Loss’ sheet ‘PreTax Profit’, from ‘Cash Flow’ sheet ‘Net Operating Cashflows’ and ‘Net Investing Cashflows’ and finally ‘Total Assets’ from ‘Balance Sheet’.

The following formula was used to calculate TATA:



Manipulation in financial statements can be shown using TATA as it looks at non-cash earnings and indicates that a higher positive accruals means higher likelihood of earnings manipulation.

**8.** **Leverage Index (LVGI)**

The Leverage Index is a ratio based on the current period Leverage and the prior period Leverage(Leverage/Prior Leverage). To compute the Leverage ‘NCL - Long-Term Debt’, ‘Total Current Liabilities’ and ‘Total Assets’ from the Balance Sheet of each company is required.

The following formula was used to calculate LVGI:

Leverage = (Long term debt + current liabilities) / Total assets

The higher the leverage ratio, the higher the possibility of the company being involved in financial fraud.

## 

## Why Benish M-Score?

According to research on the ‘Likelihood of a Company’s Manipulation of Its Financial Statement: An Empirical Analysis Using Beneish M-Score Model’, it was noted that the Beneish M-Score simply points out the likelihood of companies manipulating their earnings and it cannot detect with 100% accuracy ***(M.Arman &S.Sharmin,2019)***. After further research was conducted, we were able to narrow down to other methods that could potentially be utilised to automate the auditing financial statements to detect fraudulent activity or anomalies. Beneish M-Score was chosen as our first model to implement as it was a mathematical approach that can easily show the chances of manipulation, especially when it is broken into various different financial ratios, allowing individual analysis to get a clearer picture.

# 

# Benford’s Law

Benford’s Law is a powerful analytical tool used for detecting anomalies and involves examining the frequency of the digits in the financial statements. Any accounting or financial data is expected to conform to the Benford distribution and the frequencies can be computed by measuring deviations. Benford’s Law is another possible approach to audit financial statements as it looks at the raw dataset and can be utilised to detect any trends in naturally occurring data that is used for fraud detection.

# 

# Other Models and Algorithms

The Beneish M-Score and Beneford’s Law provide enough information to detect a certain in which a company has manipulated its financial data but it cannot be clearly identified visually.

A simple clustering algorithm that can be used to further improve the auditing process is K Means Clustering. The data is broken down into groups based on similarity with a predefined number of clusters. This is a good way to visually detect any anomalies but unfortunately is only limited to two features being used at the same time rather than the whole financial data.

With the assistance of Isolation Forest, a tree structure based model, the data can be subsampled and makes it possible to identify outliers easily detecting anomalies. Isolation Forest separates each point in a dataset creating tree structures based on randomly selected features and is considered to be the best algorithm for anomaly detection.

# Data Descriptions

## Data Being Used

To automate the auditing of financial statements, we decided to focus on collecting the financial data for companies that were listed in the project document where fraud was detected by ASIC. We used Morningstar in order to gain access to different companies' financial information for the past 11 years. Morningstar allowed us to collect the data in excel format with multiple sheets for the companies listed by ASIC. We were able to separate the data for full year and half year to gather all the information on a certain company’s financial data.

## Data Being Generated

The financial statements for the past few years that were generated on Morningstar were used to do our analysis. It was a dataset that had multiple worksheets, these worksheets include ‘Profit Loss’, ‘Balance Sheet’, ‘Cash flow’,’Per Share Statistics’, ‘Revenue Expense’ , ‘Sundry Analysis’, ‘Growth Rates’, ‘Ratio Analysis’ and ‘Asset Base Analysis’. The worksheets were consistent across companies selected and also between half and full yearly statements.

## Data Being Stored

As a group we decided the best way to store our data is to utilise Github, as we are dealing with large files. It is ideal to use Github as it allows us to update our notebook and other documents quickly and efficiently while being able to get the most up to date file as well. Github allows us to avoid duplicate file issues and also have it on the cloud not taking up local drive storage. We have used separate folders that have the excel files for the companies selected with both the half and full yearly financial data. A new file was created with a combination of both the periods as well, along with other edits to the file type for easy access. The jupyter notebook file has the groups combined project after it has been tested out in other notebooks individually in order to have no mishaps with the final project notebook.

# 

# Model evaluation

## Evaluating Data/Outputs

For this project, we collected data from MorningStar including the balance sheet, income statement and statement of cash flow for the period 2010-2021 for companies for which the data exists for the whole of this duration. The data from MorningStar is compared to the data from different sources to ensure correctness. The data was pre-processed, which included cleaning the financial data, reordering the variables, and removing records with missing values or imputing missing values with zero to ensure data accuracy. Then the data was used to calculate Beneish M-Score; two results were evaluated through half-yearly financial data and full-yearly financial data to check the accuracy of the output. For correctness, financial data was collected from the company who has released penalty notices, executed Beneish M-Score to compare the result with the released penalty notice.

## Model Comparison - Metrics for Regression and Classification

Our proposed approach for comparing the models identified in this report is to use the metrics detailed below as benchmarks for each model, ultimately choosing the model which has the least error while correctly predicting the most incidences of fraudulent reporting.

**Loss Functions and Metrics for Regression**

Loss functions are used to optimize or tune models **(Parmar, 2018)**, whereas metric functions are used to evaluate and select models. Due to the inability to determine regression accuracy, the same metrics are utilized to evaluate performance as well as model error for optimization. In our case, the Beneish M-Score regression model can be evaluated using the following metrics to assess its performance:

**Mean Square Error**: calculates the average of the squares of the errors or deviations, i.e. the difference between the estimated and real value. It assists in assigning greater weights to outliers, therefore decreasing the problem of overfitting. This property is particularly useful considering that our objective is to identify anomalous data points in the dataset that stray far from the mean of the population.

**Mean Absolute Error**: this is the absolute difference between the estimated value and real value. It decreases the weight for outlier errors when compared to the mean squared error. It should be noted however that this is of lesser importance given the large difference in scale of the values reported in financial statements between companies.

**Smooth Absolute Error**: this is the absolute difference between the estimated and actual values for forecasts that are near to the true value, and it is the square of the difference between the estimated and true values for outliers.

### 

## Confusion Matrix

A matrix called the confusion matrix is built for each classification model prediction to show the number of test cases properly and wrongly categorized **(Narkhede, 2018)**. This would be a particularly relevant evaluation technique for the cluster analysis and K-means model approach for fraud detection. It should be noted that the classification of a company exhibiting anomalous reporting in a period is verified only by an ASIC penalty notice or media release that specifically identifies the company involved. In other words, there is no way to verify real positive cases of fraud in the market without an investigation by ASIC to confirm the model’s findings.



Where:

* TN = Number of negative cases correctly classified (I.e., no fraud)
* TP = Number of positive cases correctly classified (I.e., correctly detected fraud)
* FN = Number of positive cases incorrectly classified as negative (i.e., undetected fraud)
* FP = Number of negative cases incorrectly classified as positive (i.e., incorrectly detected fraud)

**Precision**

Precision is the metric used to identify the correctness of classification.

*Precision = TP/(TP+FP)*

This equation is intuitively the ratio of correct positive classifications to total predicted positive classifications. The larger the percentage, the higher the accuracy, implying that the model is better able to properly categorize the positive class.

**Recall**

Recall indicates the number of positive cases correctly classified out of the total number of positive cases.

*Recall = TP/(TP+FN)*

**F1 Score**

Because the F1 score is the harmonic mean of Recall and Precision, it balances out the strengths of each. It's beneficial in situations where both recall and precision are important, such as identifying plane parts that may need to be repaired. Precision will be necessary to save the cost (since plane components are highly costly) and recall will be required to guarantee that the machinery is stable and does not endanger human life.

*F1-Score = 2 \* ((precision \* recall) / (precision + recall))*

# 

# Performance Evaluation Results

The methodology of our testing on the first implementation of the Beneish M-Score model has already been thoroughly explained earlier in this document, though for completeness we will summarise the results of our testing here.

Testing was conducted on a number of firms, including ADH, REX and BCI. For testing ADH, the combined half and full-yearly financial data from ADH was used to calculate Beneish M-Scores for each 6 month period from 2011 to 2020. This was achieved by the periodic calculation of each of the 7 indices, whose weighted sum indicated anomalous reporting had occurred. Particular weight was due to the value of DSRI (Days Sales in Receivable Index), AQI (Asset Quality Index), SGAI (Sales, General & Administrative Expenses Index) ascending in 2016/17 and 2020, as well as the value of DEPI (Depreciation Index) and TATA (Total Accruals to Total Assets) ascending in 2016. The M-Scores for the June reporting period for 2016 and 2020 were identified as anomalous as they exceeded -1.78. This coincides with the period in 2016 for which ADH was issued a penalty notice by ASIC for poor disclosure of their materially significant forecasting errors (LaFrenz, 2017), demonstrating that the model was effective and accurate in identifying anomalous reporting periods.

Also, ground truth is used from the penalty notice issued dates, but it is not an accurate representation of when the fraud has occurred. However, it is the closest data can be found from reliable sources of information. From the predictions below, we've identified that the model successfully detected 15 out of 28 (53.6%) anomalous periods based on the in-scope penalty notices, and missed 13 of them. The model confirmed a considerable proportion of the true negative periods, i.e. the periods which were not anomalous (347 out of 557 or 62.3%). Interestingly, the model flagged 210 periods as anomalous that did not appear in the ground truth dataset derived from the penalty notices. This could be an indicator that there were anomalous events during these periods that were missed by ASIC and could correspond with undetected suspicious activity. Alternatively, this could indicate that the model was too ambitious in detecting anomalies.

|  | **Actual 0** | **Actual 1** |
| --- | --- | --- |
| **Predicted 0** | 347 | 13 |
| **Predicted 1** | 210 | 15 |

*Precision = TP/(TP+FP)* = 15/(15+210) = 0.0667

*Recall = TP/(TP+FN)* = 15/(15+347) = 0.0414

*F1-Score = 2 \* ((precision \* recall) / (precision + recall))* = 0.0510

In the future, more financial data from small firms and big firms who have released penalty notices will be collected to execute Beneish M-Score to compare the result with the released penalty notice to evaluate the percentage of correct test and incorrect test to determine the accuracy of the Beneish M-Score and the difference of performance between big firm and small firms.

References

Arman, Mohammad & Sharmin, Shirin. (2019). Likelihood of a Company's Manipulation of Its Financial Statement: An Empirical Analysis Using Beneish M-Score Model. <https://www.researchgate.net/publication/336577715\_Likelihood\_of\_a\_Company's\_Manipulation\_of\_Its\_Financial\_Statement\_An\_Empirical\_Analysis\_Using\_Beneish\_M-Score\_Model> [Accessed 18 September 2021]

Beneish, Messod. (1999). The Detection of Earnings Manipulation. Financial Analysts Journal - FINANC ANAL J. 55. 24-36. 10.2469/faj.v55.n5.2296. <https://www.scribd.com/doc/33484680/The-Detection-of-Earnings-Manipulation-Messod-D-Beneish> [Accessed 18 September 2021]

Coursera. 2021. Fraud Prediction Models. [online] Available at: <https://fr.coursera.org/lecture/accounting-analytics/fraud-prediction-models-3-5-BbkJj> [Accessed 18 September 2021].

Gmtresearch.com. 2021. Beneish’s M-Score | Accounting Ratio | GMT Research. [online] Available at: <https://www.gmtresearch.com/en/accounting-ratio/beneishs-m-score/> [Accessed 18 September 2021].

LaFrenz, C. Adairs fined by regulator over non-disclosure [online] Available at <.https://www.afr.com/companies/retail/adairs-fined-by-regulator-over-nondisclosure-20171020-gz4pl4> [Accessed 25 September 2021]

Narkhede.(2018). Understanding Confusion Matrix. <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62> [Accessed 20 September 2021]

Parmar.(2018). Common Loss Functions on Machine Learning. <https://towardsdatascience.com/common-loss-functions-in-machine-learning-46af0ffc4d23> [Accessed 20 September 2021]

Petrík, Vladimír. (2016). APPLICATION OF BENEISH M-SCORE ON SELECTED FINANCIAL STATEMENTS.<https://www.researchgate.net/profile/Vladimir-Petrik/publication/311733912\_APPLICATION\_OF\_BENEISH\_M-SCORE\_ON\_SELECTED\_FINANCIAL\_STATEMENTS/links/58581cf508ae3852d2543fd3/APPLICATION-OF-BENEISH-M-SCORE-ON-SELECTED-FINANCIAL-STATEMENTS.pdf>[Accessed 18 September 2021]

A picture containing drawing, clock

Description automatically generated

***COMP3850 Project Deliverable Certificate***

| Name of Deliverable | *Scripts and Model Execution* |
| --- | --- |
| Date Submitted | *21/ 10 / 2021* |
| Project Group Number | *6* |
| Rubric stream being followed for this deliverable (highlight one)  *Note: the feasibility study has the same rubric for all streams.* | *SOFTWARE Rubric*  *GAMES Rubric*  *CYBERSECURITY Rubric*  *DATA SCIENCE Rubric* |

We, the undersigned members of the above Project Group, collectively and individually certify that the above Project Deliverable, as submitted, **is entirely our own work**, other than where explicitly indicated in the deliverable documentation.

| INITIALS | SURNAME | GIVEN NAME | STUDENT NUMBER | SIGNATURE *(IN-PERSON OR DIGITAL)* |
| --- | --- | --- | --- | --- |
| ***KL*** | ***Lazenby*** | ***Kurtis*** | ***45610266*** | ***kl*** |
| ***KK*** | ***Kannan*** | ***Keerthana*** | ***45870969*** | ***kk*** |
| ***NS*** | ***Soares*** | ***Nathan*** | ***45382417*** | ***ns*** |
| ***TB*** | ***Bienasz*** | ***Thomas*** | ***45982201*** | ***tb*** |
| ***ES*** | ***Salafranca*** | ***Ella*** | ***45524742*** | ***es*** |
| ***XC*** | ***Chen*** | ***Xinrui*** | ***44871481*** | ***xc*** |

*© Macquarie University, 2021****List of tasks completed for the deliverable and activities since last deliverable certificate with totals for each individual team member and whole team*** *(copy individual total row for each member and copy pages if more pages needed)*

| *Performed by*  *(Student Names)* | *Duration*  *(hrs)* | *Complexity*  *(L, M, H)* | *Name of task* | *Checked by*  *(Initials)* |
| --- | --- | --- | --- | --- |
| *Kurtis* | *15* | *m* | *Data sources and file locations* | *kl* |
| *Nathan* | *15* | *m* | *Data sources and file locations, introduction* | *ns* |
| *Thomas* | *15* | *m* | *End-User training* | *tb* |
| *Ella* | *15* | *m* | *Installation and configuration* | *es* |
| *Keerthana* | *15* | *m* | *Models* | *kk* |
| *Xinrui* | *15* | *m* | *Support and maintenance* | *xs* |

Scripts & Model Execution



**Thomas Bienasz**

**Xinrui Chen**

**Keerthana Kannan**

**Kurtis Lazenby**

**Ella Salafranca**

**Nathan Soares**

20-10-2021

COMP3850 - Computing Industry Project

# Table of Contents

[**Table of Contents**](#_gci2zv7wka0p) **1**

[**Introduction**](#_ol61dg3ghou5) **2**

[**Data Sources and Format (Kurtis/Nathan)**](#_7li80ppdvr5) **2**

[**File Locations (Kurtis)**](#_3v800iepnx0z) **3**

[**Installation and Configuration (Ella)**](#_h98rfjpd7qal) **3**

[**Model Parameters (Nathan)**](#_rlzm8fh0tyoq) **4**

[**Model Setup (Keerthana)**](#_wypxpdwgg7pr) **5-15**

[**Output (Keerthana)**](#_z9328m26uv2) **15-16**

[**Support & Maintenance (Sherry)**](#_23uqzcux5nis) **16**

# 

# 

# 

# 

# 

# Introduction

This document serves as a guide for users of our financial statement anomaly detection model. It includes specifications on the type of data used, the files involved, software installation and support as well as a detailed model setup guide.

# Data Sources, Formats and File Locations

The data sources for the model can be broadly classified as simply the financial statements of any publicly listed company. This most crucially includes the Income Statement and Balance Sheet of the company.

In order to make the data ingestion process as efficient as possible, we decided to use the financial statements found in a ‘.csv’ format through Morningstar’s *DatAnalysis Premium* service, accessible [here](https://datanalysis-morningstar-com-au.simsrad.net.ocs.mq.edu.au/). This is particularly useful because the structure of the sheets is standardised to a sufficient degree between companies, making the normalisation of the data and its features more streamlined. Morningstar provides financial statements split by interim and full-yearly data, both of which are utilised by our model for different purposes.

Strictly speaking, it is possible to obtain the financial statement data from any source as long as it is formatted in such a way that the model will be able to correctly extract the necessary data from it. Here we describe the steps for importing new data from both Morningstar and an alternate data source:

1. Morningstar Approach:
   1. Download interim and full-yearly statement data for the chosen company from Morningstar
   2. Rename these statements in the format, ‘TICKER\_PERIOD’. E.g., for Commonwealth Bank, ‘CBA\_HY’ for the interim data and ‘CBA\_FY’ for the full-yearly data
   3. Store these files in the /datasets/ folder
   4. Run the ‘Morningstar dataset conversion to Excel file (.xlsx)’ cells in the notebook. The resulting Excel files will be stored in the /datasets\_excel/ folder
2. Generic Approach:
   1. Download interim and full-yearly statement data for the chosen company from an alternative source and rename them in the ‘TICKER\_PERIOD’ naming scheme.
   2. Use the data to create an Excel workbook with sheets called ‘Profit Loss’ and ‘Balance Sheet’ and populate the columns according to the format visible in the existing datasets in the /datasets\_excel/ folder
   3. Save these files in the /datasets\_excel/ folder in .xlsx format

Additionally, the ‘COMP3850 Penalty Dates.xlsx’ file located in the root folder contains a dataset that the team compiled from the penalty notices issued by ASIC with relation to each of our chosen companies. The dataset shows the dates corresponding to each company’s incidents of a breach of ASIC’s regulations based on the ASIC penalty notice. To add to this dataset, add the ticker name for your desired company and list below it the dates on which malicious activity occurred based on an ASIC penalty notice.

# Installation and Configuration

This section details instructions on how to access the model from the user’s PC. The anomaly detection model is currently stored in a GitHub repository and runs on Jupyter Notebook. In order to set up the model, the user will need to set up the following:

1. Access Jupyter Notebook (via [Anaconda](https://www.anaconda.com/products/individual))

* Although there are different methods to accessing Jupyter Notebook, Anaconda will be assumed as the main access point to Jupyter in this manual.
* The program is available on Windows, MacOS and Linux, please refer to the appropriate download links available on the page.
* Once downloaded and set up, you can now launch Jupyter Notebook.

1. Clone GitHub repository to your local device

* Open command prompt and type:
  + git clone [https://github.com/someonealive0/deepaudi](https://github.com/someonealive0/deepaudit)t or;
* If your PC has GitHub Desktop installed, you can clone it by accessing the repository, clicking Code > Open with GitHub Desktop
  + Access it from the same link given above

1. Launch Jupyter Notebook
   1. Open the Anaconda Powershell and navigate to the project folder
   2. Enter the following command into the console to open Jupyter notebook:
      1. 
   3. Click the COMP3850.ipynb file to open the project notebook.

# Model Parameters

The main model we have implemented is the Beneish M-Score model, whose inputs consist of financial ratios derived from the data for each company. The weighting of these inputs are derived by Beneish, and thus for the sake of consistency we have not modified these parameters. However, there is a degree of subjectivity in the calculation of these ratios owing to the different reporting standards used by different companies, resulting in some ratios being impossible to calculate.

By default, in these cases where a ratio is not possible to calculate (such as the Days Sales Receivables Index for a financial institution like ANZ), the value for that ratio is 0 unless specified otherwise in the Beneish M-Score parameter weighting description, referenced in the Scoping Document. However, it is feasible to substitute these ratios with similar available data based on the user’s discretion. One such example is the substitution of the Net Profit Margin for the Gross Profit Margin when calculating the GMI ratio, which was necessary for Adairs Limited as the Gross Profit Margin was not reported.

For users who would like to modify the Beneish M-Score weighting parameters, this could be achieved by training a new logistic regression model while using the same ratios. In order to do this, there would need to be a dataset that contains the indicator of whether a reporting period for a particular company was truly anomalous or fraudulent, which is not publicly available. This may be approximated however using the penalty notices issued by ASIC to identify the periods in which firms engaged in fraudulent behaviour.

# 

# 

# 

# 

# Model Setup and Outputs

### Instruction Guide: How to set up the models (With Outputs)

| 1. Run(select the ) next to import libraries cell to correctly import all relevant libraries for the analysis. |
| --- |
| 1. Read in all relevant financial datasets that need to be used for the model.   Example: |
| 1. Make sure to clean the data by checking for duplicates or misnamed/misspelt names and create a combined workbook with both half and full yearly financial data for the company if necessary.   An example of this can be found under the “Cleaning” heading in the notebook. |
| 1. Various plots of the different features in the data can be visualised to better under the financial data.   An example of this can be found under the “Visualisation” heading in the notebook.  Example: |
| 1. To make comparison plots on various features follow the example given below.   An example of this can be found under the “Visualisation” heading in the notebook. |
| **Note: To run the calculation for the company it must be saved under its asx code in the data\_excel folder.**  **For separated workbooks simply change the asx code for the chosen company marked in red below, wherever it is mentioned.**    **The ASX code for the workbooks\_separated requires the specification of either ‘HY’ or ‘FY’. (Workbooks\_combined does not require this)**    **To access a particular worksheet within that financial data add another section specifying the name of the sheet.** |
| 1. Run the ‘calc\_DSRI’ cell and make sure there are no errors in the output.   Example: |
| 1. Before running the DSRI plot cell make sure to change the workbook and company you want to use.   Change the section circled in red, appropriately. |
| 1. Repeat steps 6 and 7 for the chosen company and period for GMI, AGI, SGI,DEPI, SGAI,TATA,LVGI. Make sure to change the relevant workbook type, ASX code and if necessary HY or FY. |
| These are example outputs for step 8 - Company Adairs Ltd (ADH): |
| 1. To get a list of M-Score for particular periods run the following cells.   Make sure to change the ASX code here.      This cell should produce an output with a list of m-scores. |
| 1. To see if the m-scores are above -1.78 threshold value, run the cells shown below.     Make sure to change the ASX code here. |
| 1. To get a list of all the m-score and a simple plot run the cells shown below.       The outputs should look like this.      This list and plot will continue the m-score for all the companies that are in the data\_excel file. |
| 1. To check the penalties notice collected in the excel file contacting specific dates for financial penalty use the following cells.     Make sure to change the ASX code here. |
| 1. To evaluate the model performance using the following cell with a confusion matrix output. |

# End-User Training

A very important aspect of this project is how DeepAudit’s machine learning algorithm will be delivered to our clients/users. Firstly, they will need to be given adequate training in order to understand how to implement it to their specific data sets and how to operate it correctly to get the most accurate results. There are a fair few steps involved so these calculations can be executed, but our team will ensure that every user will gain the sufficient skills and thorough knowledge which is needed. Initially, these calculations will be performed on our data set that we have created so the users can easily practice their new skills. Repetition is key when it comes to training. The more the users will perform these tasks on our data, the more comfortable they will feel when it comes to carrying out these procedures on real data presented by their company for them to analyse. At the end of the day, our clients will need to make sure they are able to implement any given financial statements into our algorithm and be confident to identify from the results which businesses have conducted data manipulation to then report their findings.

In addition, the training would include briefly explaining the specific calculations being done before getting the final results so our clients gain general knowledge of what is happening in the background. DeepAudit has made it so oru algorithm produces clear outputs such as charts and comments, which will give our clients a more of a visual learning experience as well. By adopting this approach, it makes it a lot easier for the users to visualize some of the trends if they don’t fully understand the numbers. If at any point the client is stuck or has any problems, they will be able to reach out to us through multiple support services.

# Support & Maintenance

Maintenance for this project is mainly focused on "risk management," or ensuring that the client’s mission-critical systems are well maintained and consistently perform at their optimum. Our Assistance and Maintenance Service is designed to provide both preventative and ad hoc reactive support to ensure sustainability, such as fixing bugs, updating the model with new information and improving the user experience through aesthetic changes to the notebook.

Preventive Maintenance Training

Our development team will train the client’s software development team to perform the essential 1st Level Support for any problem solving required inside the company. This will be supported by the appropriate design documentation, a user manual, and an application maintenance guide presented upon handover.

Issue Escalation Procedure

An Issue Log procedure and escalation procedure will be implemented to record, track, and monitor each system issue until it is resolved. This can be raised to the DeepAudit developers or via the client’s internal software development team.

Guaranteed Response Time

Depending on the severity of the issue, our Service Level Agreement guarantees that the client’s issue will be resolved in a timely manner. Once our Issue Log Tracker signs a call from the customer, we will start the investigation. Issues of low urgency, such as visualisation changes, spelling fixes, etc. are given a shorter guaranteed response time (1 week), while issues of higher severity are given considerably longer windows (i.e., from 2-4 weeks).