Business Context and Problem Formulation

In the context of today's finance, credit rating is a critical tool in financial risk management which is concerned by bond investors, corporates and government officers. Credit rating is an evaluation process of a debtor's capability to fulfil its financial obligations, based on several audits and analyses. The practice of credit risk assessment involves mitigating financial losses by understanding the potential borrower's ability to pay back the debts. This process significantly influences corporations, financial institutions' financing decisions and government regulations. Therefore, developing a model that enhances the prediction accuracy of a company's rating is essential and challenging.

This report aims to utilise a historical dataset that consists of 30 financial variables of the firms and respective credit ratings given by Standard and Poors for selecting a predictive model of a firm's ratings that produces the promising prediction accuracy measured in mean absolute error (MAE). Hence, it gives a good explanation of the main factors driving the ratings.

Data Processing

1. Handling missing data

To ensure that there is no distortion during the analysis process, missing values are initially spotted and managed. Missing data is a common occurrence in real datasets, and it spots missing data in 12 variables including Cash, Retained_Earnings, Dividends_per_Share_Pay_Date_Calendar, Interest_and_Related_Expense_Total, Market_Value_Total_Fiscal, Book_Value_Per_Share, Comprehensive_Income_Parent, Employees, Inventories_Total, Earnings_Per_Share_from_Operations, Operating_Activities_Net_Cash_Flow, and Financing_Activities_Net_Cash_Flow. There are a total of 108 observations with missing values which is less than 5 per cent of the sample size. Hence, as the sample size is large enough to afford the data deletion, I choose to simply remove the observations where null entries are in one or several columns, accepting the risk that it might lose some valuable information, which can influence the validity of the analysis. There is another way to handle missing values which is data imputation. However, I did not choose this method due to the concern of data randomness as well as the little information, and dataset understanding available to support imputation.

2. Handling data duplication

The problem of data duplication is also inspected during the processing of raw data, nevertheless, there is no data duplication in the dataset.

3. Reference problem with columns name

When referencing or indexing column names that contain slash '/' or parentheses '()', the problem pops up as it cannot recognise these two symbols as a part of the column name (Appendix 1). Therefore, these columns are renamed for better indexing and referencing in the following sections.

The symbol slash ‘/’ in column names is replaced with the word ‘over’ and the parentheses ‘()’ is removed, for example, the original column name "Sales/Turnover_(Net)" is now retitled as "Sales_over_Turnover_Net".

Exploratory Data Analysis (EDA)

1. Statistical measurements

First, the descriptive statistic is used to get initial basic information i.e., mean, min, max, standard deviation, and quartile range of numerical variables in the dataset. For a target variable, the firm’s ratings are in the ordinal range of 1 to 4 and the majority of observations were gathered in ranking 2 and 3 with 75% of rankings being equal or above 2, 50% being equal or above 3 and 25% being equal or above 3. For the explanatory variables, it is noticeable that most of these variables have large standard deviations relative to the means, signalling the problem of non-normal distribution in data. For example, the variable ‘Market_Value_Total_Fiscal’ has a standard deviation of 42046.9034 which is a large number compared to the mean of 17472.7784. Moreover, comparing the maximum value (615336.4559) of this variable to its mean, there is also an enormous gap between these two numbers implying the potential existence of outliers in data.

2. Skewness, Kurtosis and Outliers

Following the above finding, the skewness and kurtosis are further calculated to understand the presence of outliers in the dataset. There is a potential problem of huge outliers in the dataset since 28 out of 30 explanatory variables (excluding variables 'ID', and 'Year') have a kurtosis of greater than 7 in which -7 to 7 is considered as the appropriate kurtosis range for normally distributed data. For further skewness assessment, three variables 'EBTI_over_total_asset', 'EBTI_over_REV', and 'Financing_Activities_Net_Cash_Flow' appear to be negatively skewed (skewness is smaller than the ideal level for kurtosis which is -2) with large negative outliers on the left side of the distribution. Other predictors (excluding variables 'Net_Cash_Flow', 'ID', and 'Year') are positively skewed with skewness larger than 2. Thus, there is potentially the presence of both sizable negative and positive outliers in the dataset. Therefore, it needs to be cautious about the effect of outliers on the models. Models that are robust to outliers might be better in predicting this dataset.

3. Correlation and Relationship between variables

The correlations between 30 explanatory variables and the target variable “Rating” are calculated. The top 10 highest correlations (based on absolute value) between predictors and response variables are stated in the table below.

Explanatory variables	Correlation
Market_Value_Total_Fiscal	- 0.4207
Gross_Profit_Loss	-0.3850

Common_Equity_Liquidation_Value	-0.3733
Stockholders_Equity_Total	-0.3725
Earnings_Before_Interest	-0.3713
ID	0.3695
Revenue_Total	-0.3661
Sales_over_Turnover_Net	-0.3661
Dividends_per_Share_Pay_Date_Calendar	-0.3652
Operating_Activities_Net_Cash_Flow	-0.3517

The relationships between variables and target variable are also examined through the regression plots. It is apparent that the relationship between some variables including ‘Cash’, ‘Debt_in_Current_Liabilities_Total’, ‘Long_Term_Debt_Total’, ‘Assets_Total’, ‘Dividends_per_Share_Pay_Date_Calendar’ with ‘Rating’ display a certain nonlinear behaviour (Appendix 2). This nonlinear pattern can be better captured by the KNN model than the linear model.

Collinearity between explanatory variables

Two notable variables 'Sales_over_Turnover_Net' and ‘Revenue_Total’ were found perfectly collinear. Thus, having both two variables as a model’s predictors probably would not obtain any new information or improve the predictive accuracy of the model. To reduce the risk of perfect collinearity in future analysis, one of two variables can be dropped at the data processing stage. However, I choose to retain both variables and will remove them manually when modelling to keep the dataset as original.

Moreover, the four variables ‘total_asset/total_liabilities’, ‘EBTI/total_asset’, ‘gross_profit/rev’, and ‘Earnings_Per_Share_from_Operations’ also appear to be nearly perfectly collinear with the variable ‘Total_debt/total_asset’. The table below shows the correlation between four variables with ‘Total_debt/total_asset’. In the linear regression context, the presence of high collinearity can raise challenges in separately understanding the individual effects of covariates on the target response, thus further posing the problem of the model’s interpretability when explaining the driving factors of response.

Explanatory variables	**Correlation with ‘Total_debt/total_asset’**
EBTI_over_total_asset	0.9998

gross_profit_over_rev	0.9997
total_asset_over_total_liabilities	0.9996
Earnings_Per_Share_from_Operations	0.9991

4. Variables are ‘invalid’ for predictive purposes

Two noticeable variables ‘ID’ and ‘Year’ appear to be questionable for its predictive value. The problem can arise from the inexplicable nature of these two variables in explaining the relationship with the credit rating of the firm. The variable ‘Year’ appears to be almost the same for every observation and The features ‘ID’ and ‘Year’ seem to uselessly give information or meaning into how a certain prediction is estimated and understanding the overall connection between predictors and response. Hence, these two variables can not state anything about the new firm’s credit rating from unseen data, consequently, it lacks the power to generalise. The variable ‘ID’ is created following a continuous ordinal sequence which might violate the independence assumption and create non-random patterns in the data, further affecting model validity (if it is used as a predictor in the model).

Methodology - Model and Variable Selections

1. Data split

The original dataset is firstly split randomly into two sets such as a training and test set with the ratio 8:2 and then the training set is split again into a training set and validation set with the ratio 3:1. The training and validation set will be used in the sequential model selection process for tuning the hyperparameter and selecting the ‘optimal’ subset of predictors for models. Whereas, the test set will be used to inform the final prediction accuracy of the selected model at the end, minimising the risk of having the model that might be overfitting on the validation set.

2. The linear regression model

To fit the linear regression model, the variables initially were chosen through a backward stepwise selection. The algorithm starts with fitting all 30 independent variables in the model, then iteratively excluding a predictor that did not give the improvement to the model. In this variable selection exercise, adjusted r-squared is set to be a criterion to select the ‘best’ among a set of models that fit different subsets of variables. Adjusted r-squared will penalise the addition of insignificant variables in the model. A chosen subset of the predictors includes a total of 19 predictors which generates the highest possible adjusted r-squared of 46.5%.

With the goal of finding a simpler linear regression model, additional criteria to select variables are considered. This time, to choose a subset of fewer predictors, I based upon the result of hypothesis testing with the significance level set to be 5%. The null hypothesis is that the coefficient estimate is equal to 0. When fitting the linear regression model with 19 variables, the predictive coefficients of 5 variables ‘Stockholders_Equity_Total’, Book_Value_Per_Share’, ‘Inventories_Total’, ‘Common_Equity_Liquidation_Value’, and ‘Comprehensive_Income_Parent’ appears to be statistically insignificant with the p-value greater than significant level. Hence, these corresponding covariates are removed from the model.

Moreover, from the above EDA section, ‘ID’ was mentioned for its lack of interpretability. Consequently, I also exclude ‘ID’ and get a final subset of 13 predictors for the linear model. Notably, excluding the ‘ID’ variable from the model negatively affects model performance as well as prediction accuracy since the adjusted r-squared reduced from 46.3% to 42.1% and the absolute mean error (MAE) in the training set increased from 0.5846 to 0.6073. It is noteworthy that in terms of predicting firm ratings, to some extent the ‘ID’ variable still has predictive ability regardless of its interpretability and the power to generalise is still the concernedly problems. A final subset of 13 predictors for the linear model includes 'Market_Value_Total_Fiscal', 'Employees', 'Financing_Activities_Net_Cash_Flow', ‘'Earnings_Before_Interest', 'Long_Term_Debt_Total', 'total_asset_over_total_liabilities', 'Debt_in_Current_Liabilities_Total', 'Operating_Activities_Net_Cash_Flow', 'EBTI_over_REV', 'Net_Cash_Flow', 'Earnings_Per_Share_from_Operations', 'Liabilities_Total', 'Total_debt_over_total_asset', and 'Dividends_per_Share_Pay_Date_Calendar'.

I also consider another subset of predictors which is modified from the above group of 13 predictors. Following the finding in the EDA section, I dropped three of the variables that have problems of high collinearity, namely 'total_asset_over_total_liabilities', ‘Employees’ and 'Earnings_Per_Share_from_Operations' from the regression, retaining a subset of 10 variables. This is done to ensure that in later sections keeping all other variables constant, the individual effect of each predictor to target variable ‘Rating’ can be appropriately explained through the linear regression model. Additionally, high collinearity implies that the information provided by these three variables can be redundant in the existence of the variable 'Total_debt_over_total_asset', thus the removal of these three can reduce the model complexity without severely affecting regression fit. Notably, the presence of a high correlation between independent variables (Appendix 3) violates the assumption of no multicollinearity of linear regression, which potentially leads to biased estimates of the regression coefficients. The performance of two linear regression models is shown in the below table.

Measurements	**A subset of 13 predictors**	**A subset of 10 predictors**
Adjusted r-squared	42.1%	40.4%
Training MAE	0.6073	0.6166
Validation MAE	0.6021	0.6141

The performance in the validation set of two linear models are very similar, since the reduction of 2 predictors only resulted in validation MAE to increase by 0.12. This implies the trade-off between bias and variance, since the decrease in numbers of covariate causes the model bias to increase but at the same time induces the model's variance to decrease. In conclusion, I will sacrifice some prediction accuracy for the lowering of model complexity, thus I choose a subset of 10 variables as a final group of predictors for the linear regression model. Noteworthily, 3 of 10 variables ('Long_Term_Debt_Total', 'Debt_in_Current_Liabilities_Total', 'Dividends_per_Share_Pay_Date_Calendar') in the linear model are found to likely have a nonlinear relationship with 'Rating' in EDA section, hence the linearity assumptions of the model might be questionable.

3. The KNN regression model

Building the KNN regression model involves choosing the best set of variables while searching for the appropriate hyperparameter i.e., the number of neighbours. At the start, I re-use the function from tutorial 5 with little modification to select the appropriate numbers of neighbours and obtain the performance in the validation set for each model fitting a different subset of predictors. To choose the subsets of predictors, I write a loop that iteratively adds a new variable to the model, then uses the aforementioned function to choose an appropriate number of neighbours and calculate validation mean absolute error, one-at-a-time. The loop adds in variables in the descending order of its correlation to response and stops when the inclusion of the new variable did not give any additional improvement to the performance of the KNN model in the validation set. Note that the loop only starts to compare the validation MAE and breaks when the new MAE is larger than the current one after the KNN models fit more than 5 predictors. This condition follows the intuition of bias-variance trade-off in which we aim to find a model that can optimise the error in unseen data that is neither too simple nor too complex. Finally, the model fits 9 predictors 'Market_Value_Total_Fiscal', 'Gross_Profit_Loss', 'Common_Equity_Liquidation_Value', 'Stockholders_Equity_Total', 'Earnings_Before_Interest', 'Revenue_Total', 'Dividends_per_Share_Pay_Date_Calendar', 'Operating_Activities_Net_Cash_Flow', and 'Retained_Earnings' with the number of neighbours is 1 was selected.

Additionally, the KNN method is able to capture nonlinear behaviours in the dataset, hence I include 4 additional variables 'Cash', 'Debt_in_Current_Liabilities_Total', Long_Term_Debt_Total', and 'Assets_Total' that were found potentially having a nonlinear relationship with dependent variable 'Rating' in the EDA section. The following table presents the performance of two KNN models.

Measurements	**A subset of 9 predictors**	**A subset of 13 predictors**
Number of neighbours (k)	1	1
Validation MAE	0.3179	0.2832

A final KNN regression model is the one that fits the subset of 13 predictors with 1 neighbour since it performs better in the validation set and produces higher accuracy. However, it is concerned that the small number of neighbours of 1 might lead to the problem of overfitting.

4. The logistic regression model

The third model used to predict the target variable ‘Rating’ is logistic regression. Logistic regression is implemented as the linear model for classification where it models the probability that the response variable y_i is classified as a specific category. In this model family, the target variable ‘Rating’ is treated as a multi-categorical variable with four categories 1 (the group of best rankings), 2, 3, and 4 (the group of worse rankings). Moreover, logistic regression also does not have any significant hyperparameter to fine-tune.

The set of variables for logistic regression was chosen using a Sequential Feature Selector based on classification accuracy. Classified accuracy indicates the portion of a total number of predictions that are accurate predictions. The subset of 5 predictors chosen includes 'Gross_Profit_Loss', 'Retained_Earnings', 'Total_debt_over_total_asset', 'Dividends_per_Share_Pay_Date_Calendar', 'Earnings_Per_Share_from_Operations'. The performance of the model is shown in the table below.

Metrics	**A subset of 5 predictors**
Validation MAE	0.6724
AUC score	0.7692
Training accuracy	0.45
Validation accuracy	0.46

Since logit models act as classifiers, thus to accurately evaluate its performance, a new classification metric AUC score is introduced. AUC score implies the model's capability of distinguishing between classes, and the higher the AUC, the better the model is.

5. A single predictor model

A single predictor 'Market_Value_Total_Fiscal' is selected as it has the highest correlation with the target variable ‘Rating’. Then the variable was fitted in the above three model classes to find the optimal predictive model with one predictor. The validation results of the three models are stated below.

Metrics	**Linear regression**	**KNN regression (k = 35)**	**Logistic regression**

Validation MAE	0.7478	0.5160	1.5356

Ultimately, the KNN model that uses a single predictor 'Market_Value_Total_Fiscal' has the best performance in the validation set with the lowest MAE.

Analysis and Conclusion

1. The final model & Trade-Off Between Prediction Accuracy and Model Interpretability

Among three model families, the KNN regression model with 13 predictors and 1 neighbour performs the best in terms of optimising the accuracy of the predictions which is measured in mean absolute error. The model produces an estimated prediction error of 0.2832 when predicting data in the validation set, which decreases by 0.3334 compared to linear regression and 0.3892 compared to logistic regression. The final accuracy of the KNN model is implied by its performance in a test set that has not been previously used in training, selecting and fine-tuning phrases. Surprisingly, the KNN regression model produces a test MAE of 0.2331 which substantially improves from validation error. Thus, the model does a pretty good job of predicting unseen data and is not misled by overfitting, giving a satisfactory out-of-sample forecast performance.

However, despite the fact that the KNN model provides promising prediction accuracy, it is limited in interpreting the relationship of individual predictors with the response. This denotes the trade-off between prediction accuracy and model interpretability. The KNN model appears to be a flexible approach that can properly capture any nonlinear relationship in the input data, however potentially lead to more complicated but less interpretable estimates. In contrast, linear regression as a more restrictive model can easily explain the relationship between Y and X_i through simpler coefficient estimates. Thus in the goal of inference, the linear regression model will be used to understand the driving factors of a firm's credit ratings even though it has worse performance than the KNN in the test set (model’s test MAE is 0.6204). Moreover, the linear regression model potentially encounters the problem of overfitting since the test MAE increases by 0.0063 compared to the train MAE.

2. Interpretation of driving factors of the target variable.

The estimated predictive model of the firm’s credit ratings using linear regression is:

$$\begin{aligned}
Predicted\,rating\ &= 2.3134\ - 0.000014 \times Market\_Value\_Total\_Fiscal \\
&- 604.7946 \times\ Financing\_Activities\_Net\_Cash\_Flow\ + 0.000068 \times\ Earnings\_Before\_Interest \\
&- 0.000014 \times Long\_Term\_Debt\_Total\ +\ 0.000017 \times Debt\_in\_Current\_Liabilities\_Total \\
&- 604.7946 \times Operating\_Activities\_Net\_Cash\_Flow - 0.3090 \times EBTI\_over\_REV \\
&+ 604.7946 \times Net\_Cash\_Flow\ - 0.000002 \times Liabilities\_Total\ +\ 1.7822 \times Total\_debt\_over\_total\_asset \\
&- 0.1542 \times Dividends\_per\_Share\_Pay\_Date\_Calendar
\end{aligned}$$

The effects of an individual variable on firm’s ratings can be explained as holding all other variables constant:

- As the total fiscal market value of the firm increased by 1 (million) dollars, the firm credit rating decreased by 0.000014 on average.
- As the firm’s net cash flow from financing activities increased by 1 (million) dollars, the firm credit rating decreased by 604. 7946 on average.
- As the firm’s earnings before interest increased by 1 (million) dollars, the combined MPG increased by 0.000068 on average.
- As the firm’s total long-term debt increased by 1 (million) dollars, the firm credit rating decreased by 0.000014 on average.
- As the firm’s total debt in current liabilities increased by 1 (million) dollars, the combined MPG increased by 0.000017 on average.
- As the firm’s net cash flow from operating activities increased by 1 (million) dollars, the firm credit rating decreased by 604. 7946 on average.
- As the firm’s EBIT/EV ratio increased by 1, its credit rating decreased by 0.3090 on average.
- As the firm’s net cash flow increased by 1 (million) dollars, the firm credit rating increased by 604. 7946 on average.
- As the firm’s total liabilities increased by 1 (million) dollars, the firm credit rating decreased by 0.000002 on average.
- As the firm’s D/E ratio increased by 1, the firm credit rating increased by 1.7822 on average
- As the firm’s dividend paid per share increased by 1 dollar, the firm credit rating decreased by 0.1504 on average.

3. Implications of the error functions

The Mean Absolute Error (MAE) is a common objective function that measures the average magnitude of model-predicted errors. However, the interpretation of the target variable ‘Ratings’ might suggest a more appropriate objective function for the prediction model. For instance, the larger the error in the prediction of credit rating the bigger the financial losses that lenders or firms have to bear. In particular, the overprediction of credit rating may imply that the firm can borrow loans with a lower interest rate than it should, which adversely affects the lender or bondholder. Oppositely, the underprediction can result in the firm paying a higher interest rate than it should which could cause financial losses for the firm. Thus, penalising large errors more than small errors in predicting a firm's credit rating is crucial; however, MAE cannot do it.

The more ‘appropriate’ objective function is Root Mean Squared Error (MSE) which penalises large and small errors differently. Since RMSE squares the differences between predictions and true values of the response variable, it imposes a greater penalty for large errors than for small errors. The table below compares the performance of the candidate models using the MAE and the RMSE in the validation set.

Metrics	**Linear regression**	**KNN regression (k=1)**	**Logistic regression**

Validation MAE	0.6166	0.2832	0.6724
Validation RMSE	0.7782	0.6223	0.9961

Both the validation RMSE and MAE suggest that the KNN regression model has the best predictive performance. However, there is a decrease in the difference between the predicted error of KNN models and other models.

4. Limitations

Firstly, it is found that independent variables are reported in different scales, for instance, variables like ‘Market_Value_Total_Fiscal’ are stated in value whereas variables like ‘Total Debt/Total Asset’ and ‘Rating’ are stated in ratio. This can affect the model’s coefficient estimates where the input variables in larger units might overwhelm the smaller value of predictors. Therefore, it suggests the normalisation for larger values of variables in the data processing stage. One method that can be used is min–max normalisation which performs a linear transformation on original data. It is noted that in this analysis, normalisation of variables is lacking, leading to the large gaps between the unit effect of every single predictor on response variable since coefficient estimates are in different scales when interpreting the results of the linear regression model. Thus it is hard to use the linear coefficients to compare the importance of each driving factor of a firm's rating. The application of Mahalanobis distance replaces Euclidean distance in the KNN model can mitigate the effect of different scaled variables.

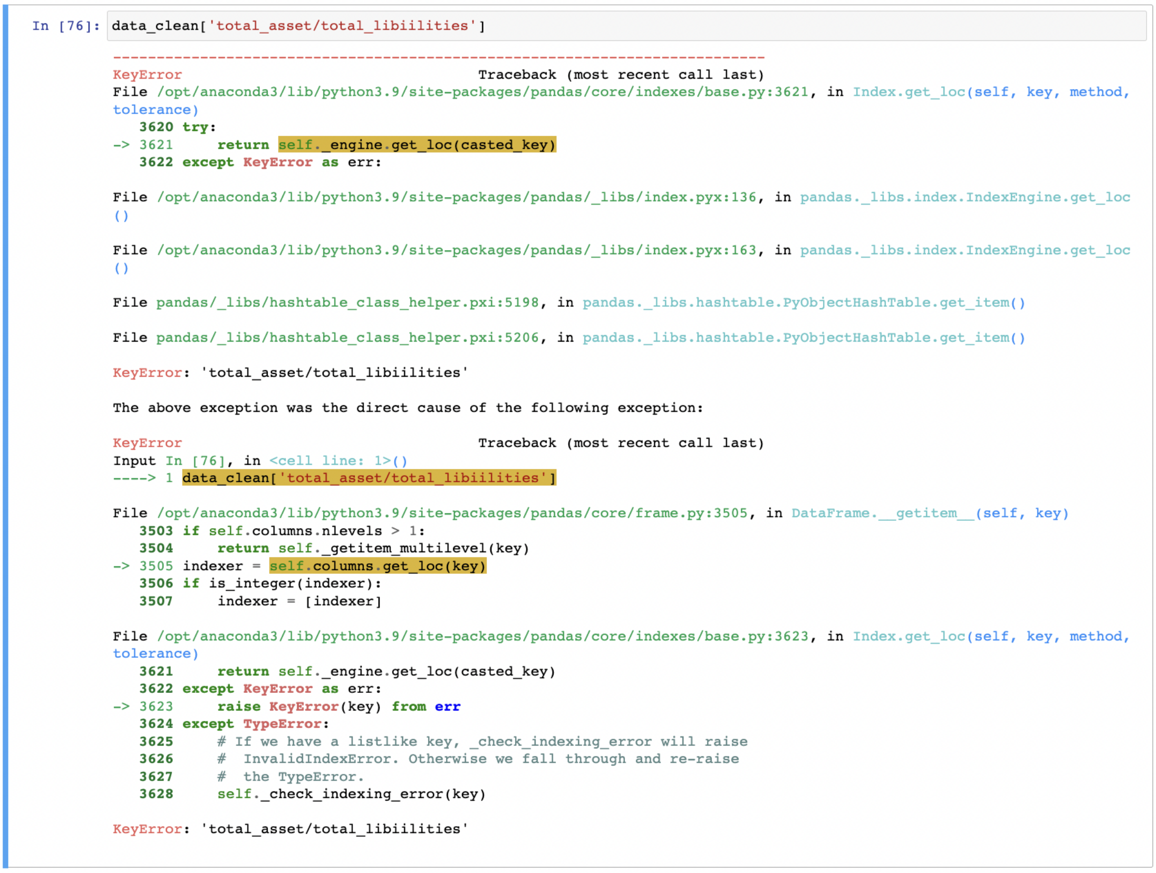
Secondly, the statistical methods can be limited because of their assumption about the distribution and independence of the data. The use of hypothesis testing in choosing variables for linear regression concerns the problem of multicollinearity in the dataset which could affect the accuracy of reference results i.e., p-value.

Thirdly, there is a trade-off between the choice of observations for each set of data in data splitting and the accuracy in error approximation in the model. The ratio chosen for data splitting is based on common choice. However, the increase in observations that are set aside for evaluating the model performance (since in the analysis, I hoard data for both validation and test set) reduces the numbers of train data which could induce systematic bias into the model.

Finally, potential outliers are detected in EDA but have not yet been handled. Outliers in the dataset might potentially bias the model's estimates, increase the variance and decrease the model's accuracy in the analysis since both linear regression and KNN regression are not robust to outliers.

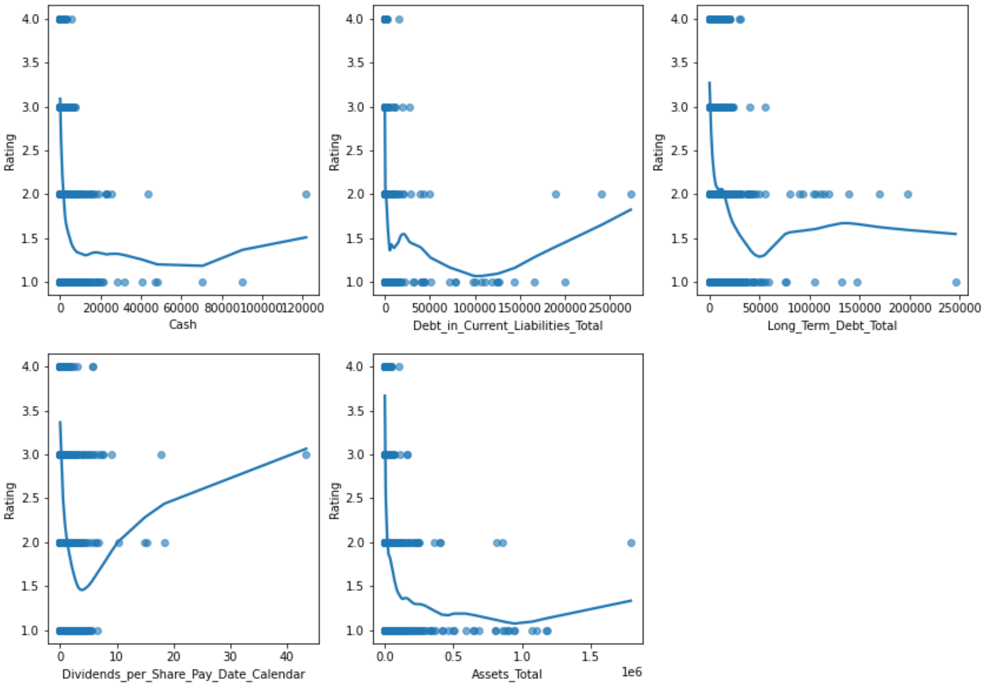
Appendices

1. Appendix 1: The error was generated by the original column names.

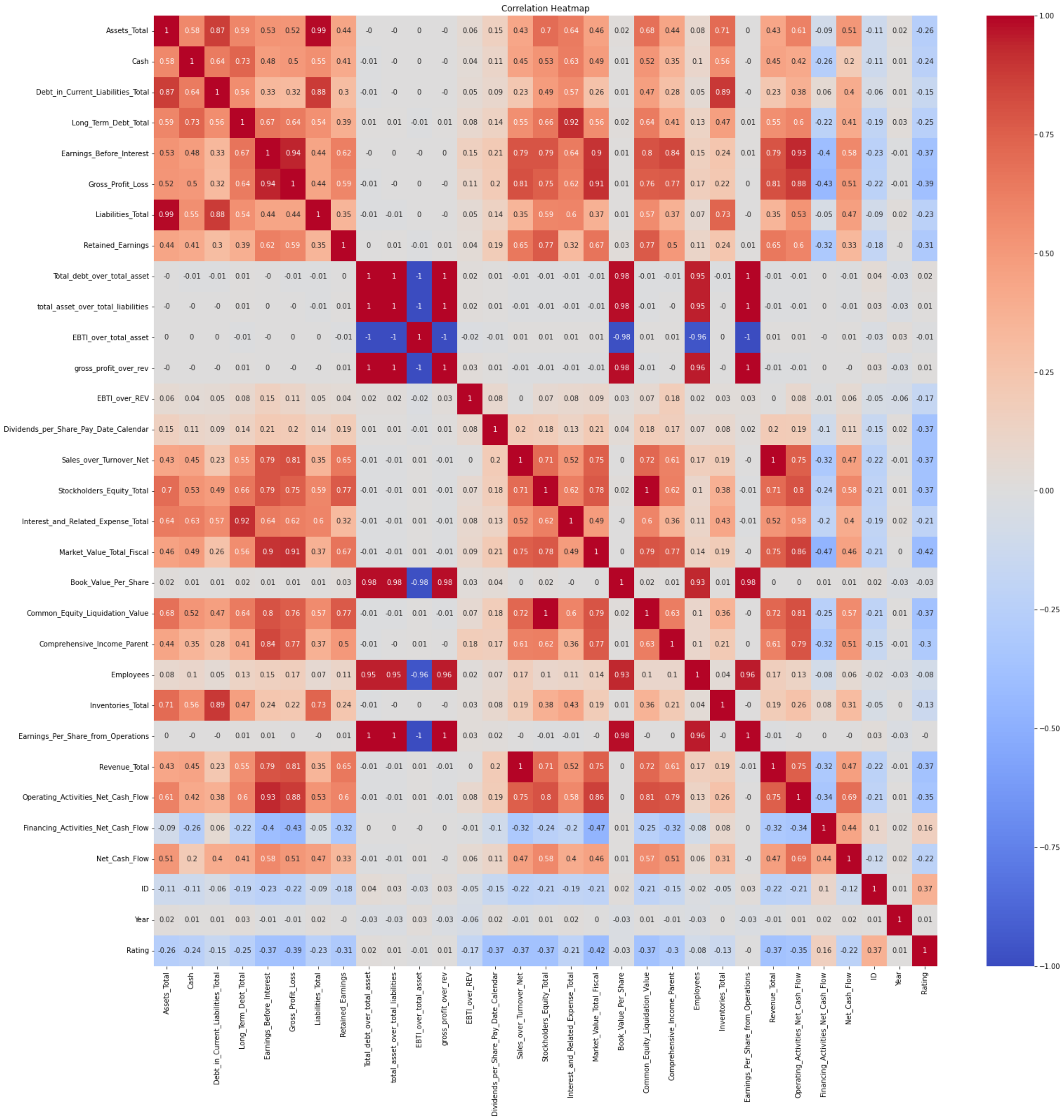


```
In [76]: data_clean['total_asset/total_libiilities']

---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
File /opt/anaconda3/lib/python3.9/site-packages/pandas/core/indexes/base.py:3621, in Index.get_loc(self, key, method,
tolerance)
   3620 try:
-> 3621     return self._engine.get_loc(casted_key)
   3622 except KeyError as err:

File /opt/anaconda3/lib/python3.9/site-packages/pandas/_libs/index.pyx:136, in pandas._libs.index.IndexEngine.get_loc
()

File /opt/anaconda3/lib/python3.9/site-packages/pandas/_libs/index.pyx:163, in pandas._libs.index.IndexEngine.get_loc
()

File pandas/_libs/hashtable_class_helper.pxi:5198, in pandas._libs.hashtable.PyObjectHashTable.get_item()

File pandas/_libs/hashtable_class_helper.pxi:5206, in pandas._libs.hashtable.PyObjectHashTable.get_item()

KeyError: 'total_asset/total_libiilities'

The above exception was the direct cause of the following exception:

KeyError                                  Traceback (most recent call last)
Input In [76], in <cell line: 1>()
----> 1 data_clean['total_asset/total_libiilities']

File /opt/anaconda3/lib/python3.9/site-packages/pandas/core/frame.py:3505, in DataFrame.__getitem__(self, key)
   3503 if self.columns.nlevels > 1:
   3504     return self._getitem_multilevel(key)
-> 3505 indexer = self.columns.get_loc(key)
   3506 if is_integer(indexer):
   3507     indexer = [indexer]

File /opt/anaconda3/lib/python3.9/site-packages/pandas/core/indexes/base.py:3623, in Index.get_loc(self, key, method,
tolerance)
   3621     return self._engine.get_loc(casted_key)
   3622 except KeyError as err:
-> 3623     raise KeyError(key) from err
   3624 except TypeError:
   3625     # If we have a listlike key, _check_indexing_error will raise
   3626     #  InvalidIndexError. Otherwise we fall through and re-raise
   3627     #  the TypeError.
   3628     self._check_indexing_error(key)

KeyError: 'total_asset/total_libiilities'
```

2. Appendix 2: The nonlinear relationship between 5 independent variables and target variable



3. Appendix 3: Correlation Heatmap


Correlation Heatmap
Assets_Total 1 0.58 0.87 0.59 0.53 0.52 0.99 0.44 -0 -0 0 -0 0.06 0.15 0.43 0.7 0.64 0.46 0.02 0.68 0.44 0.08 0.71 0 0.43 0.61 -0.09 0.51 -0.11 0.02 -0.26
Cash 0.58 1 0.64 0.73 0.48 0.5 0.55 0.41 -0.01 -0 0 -0 0.04 0.11 0.45 0.53 0.63 0.49 0.01 0.52 0.35 0.1 0.56 -0 0.45 0.42 -0.26 0.2 -0.11 0.01 -0.24
Debt_in_Current_Liabilities_Total 0.87 0.64 1 0.56 0.33 0.32 0.88 0.3 -0.01 -0 0 -0 0.05 0.09 0.23 0.49 0.57 0.26 0.01 0.47 0.28 0.05 0.89 -0 0.23 0.38 0.06 0.4 -0.06 0.01 -0.15
Long_Term_Debt_Total 0.59 0.73 0.56 1 0.67 0.64 0.54 0.39 0.01 0.01 -0.01 0.01 0.08 0.14 0.55 0.66 0.92 0.56 0.02 0.64 0.41 0.13 0.47 0.01 0.55 0.6 -0.22 0.41 -0.19 0.03 -0.25
Earnings_Before_Interest 0.53 0.48 0.33 0.67 1 0.94 0.44 0.62 -0 0 -0 0 0.15 0.21 0.79 0.79 0.64 0.9 0.01 0.8 0.84 0.15 0.24 0.01 0.79 0.93 -0.4 0.58 -0.23 -0.01 -0.37
Gross_Profit_Loss 0.52 0.5 0.32 0.64 0.94 1 0.44 0.59 -0.01 -0 0 -0 0.11 0.2 0.81 0.75 0.62 0.91 0.01 0.76 0.77 0.17 0.22 0 0.81 0.88 -0.43 0.51 -0.22 -0.01 -0.39
Liabilities_Total 0.99 0.55 0.88 0.54 0.44 0.44 1 0.35 -0.01 -0.01 0 -0 0.05 0.14 0.35 0.59 0.6 0.37 0.01 0.57 0.37 0.07 0.73 -0 0.35 0.53 -0.05 0.47 -0.09 0.02 -0.23
Retained_Earnings 0.44 0.41 0.3 0.39 0.62 0.59 0.35 1 0 0.01 -0.01 0.01 0.04 0.19 0.65 0.77 0.32 0.67 0.03 0.77 0.5 0.11 0.24 0.01 0.65 0.6 -0.32 0.33 -0.18 -0 -0.31
Total_debt_over_total_asset -0 -0.01 -0.01 0.01 -0 -0.01 -0.01 0 1 1 -1 1 0.02 0.01 -0.01 -0.01 -0.01 -0.01 0.98 -0.01 -0.01 0.95 -0.01 1 -0.01 -0.01 0 -0.01 0.04 -0.03 0.02
total_asset_over_total_liabilities -0 -0 -0 0.01 0 -0 -0.01 0.01 1 1 -1 1 0.02 0.01 -0.01 -0.01 -0.01 -0.01 0.98 -0.01 -0 0.95 -0 1 -0.01 -0.01 0 -0.01 0.03 -0.03 0.01
EBTI_over_total_asset 0 0 0 -0.01 -0 0 0 -0.01 -1 -1 1 -1 -0.02 -0.01 0.01 0.01 0.01 0.01 -0.98 0.01 0.01 -0.96 0 -1 0.01 0.01 -0 0.01 -0.03 0.03 -0.01
gross_profit_over_rev -0 -0 -0 0.01 0 -0 -0 0.01 1 1 -1 1 0.03 0.01 -0.01 -0.01 -0.01 -0.01 0.98 -0.01 -0 0.96 -0 1 -0.01 -0.01 0 -0 0.03 -0.03 0.01
EBTI_over_REV 0.06 0.04 0.05 0.08 0.15 0.11 0.05 0.04 0.02 0.02 -0.02 0.03 1 0.08 0 0.07 0.08 0.09 0.03 0.07 0.18 0.02 0.03 0.03 0 0.08 -0.01 0.06 -0.05 -0.06 -0.17
Dividends_per_Share_Pay_Date_Calendar 0.15 0.11 0.09 0.14 0.21 0.2 0.14 0.19 0.01 0.01 -0.01 0.01 0.08 1 0.2 0.18 0.13 0.21 0.04 0.18 0.17 0.07 0.08 0.02 0.2 0.19 -0.1 0.11 -0.15 0.02 -0.37
Sales_over_Turnover_Net 0.43 0.45 0.23 0.55 0.79 0.81 0.35 0.65 -0.01 -0.01 0.01 -0.01 0 0.2 1 0.71 0.52 0.75 0 0.72 0.61 0.17 0.19 -0 1 0.75 -0.32 0.47 -0.22 -0.01 -0.37
Stockholders_Equity_Total 0.7 0.53 0.49 0.66 0.79 0.75 0.59 0.77 -0.01 -0.01 0.01 -0.01 0.07 0.18 0.71 1 0.62 0.78 0.02 1 0.62 0.1 0.38 -0.01 0.71 0.8 -0.24 0.58 -0.21 0.01 -0.37
Interest_and_Related_Expense_Total 0.64 0.63 0.57 0.92 0.64 0.62 0.6 0.32 -0.01 -0.01 0.01 -0.01 0.08 0.13 0.52 0.62 1 0.49 -0 0.6 0.36 0.11 0.43 -0.01 0.52 0.58 -0.2 0.4 -0.19 0.02 -0.21
Market_Value_Total_Fiscal 0.46 0.49 0.26 0.56 0.9 0.91 0.37 0.67 -0.01 -0.01 0.01 -0.01 0.09 0.21 0.75 0.78 0.49 1 0 0.79 0.77 0.14 0.19 -0 0.75 0.86 -0.47 0.46 -0.21 0 -0.42
Book_Value_Per_Share 0.02 0.01 0.01 0.02 0.01 0.01 0.01 0.03 0.98 0.98 -0.98 0.98 0.03 0.04 0 0.02 -0 0 1 0.02 0.01 0.93 0.01 0.98 0 0 0.01 0.01 0.02 -0.03 -0.03
Common_Equity_Liquidation_Value 0.68 0.52 0.47 0.64 0.8 0.76 0.57 0.77 -0.01 -0.01 0.01 -0.01 0.07 0.18 0.72 1 0.6 0.79 0.02 1 0.63 0.1 0.36 -0 0.72 0.81 -0.25 0.57 -0.21 0.01 -0.37
Comprehensive_Income_Parent 0.44 0.35 0.28 0.41 0.84 0.77 0.37 0.5 -0.01 -0 0.01 -0 0.18 0.17 0.61 0.62 0.36 0.77 0.01 0.63 1 0.1 0.21 0 0.61 0.79 -0.32 0.51 -0.15 -0.01 -0.3
Employees 0.08 0.1 0.05 0.13 0.15 0.17 0.07 0.11 0.95 0.95 -0.96 0.96 0.02 0.07 0.17 0.1 0.11 0.14 0.93 0.1 0.1 1 0.04 0.96 0.17 0.13 -0.08 0.06 -0.02 -0.03 -0.08
Inventories_Total 0.71 0.56 0.89 0.47 0.24 0.22 0.73 0.24 -0.01 -0 0 -0 0.03 0.08 0.19 0.38 0.43 0.19 0.01 0.36 0.21 0.04 1 -0 0.19 0.26 0.08 0.31 -0.05 0 -0.13
Earnings_Per_Share_from_Operations 0 -0 -0 0.01 0.01 0 -0 0.01 1 1 -1 1 0.03 0.02 -0 -0.01 -0.01 -0 0.98 -0 0 0.96 -0 1 -0.01 -0 0 -0 0.03 -0.03 -0
Revenue_Total 0.43 0.45 0.23 0.55 0.79 0.81 0.35 0.65 -0.01 -0.01 0.01 -0.01 0 0.2 1 0.71 0.52 0.75 0 0.72 0.61 0.17 0.19 -0.01 1 0.75 -0.32 0.47 -0.22 -0.01 -0.37
Operating_Activities_Net_Cash_Flow 0.61 0.42 0.38 0.6 0.93 0.88 0.53 0.6 -0.01 -0.01 0.01 -0.01 0.08 0.19 0.75 0.8 0.58 0.86 0 0.81 0.79 0.13 0.26 -0 0.75 1 -0.34 0.69 -0.21 0.01 -0.35
Financing_Activities_Net_Cash_Flow -0.09 -0.26 0.06 -0.22 -0.4 -0.43 -0.05 -0.32 0 0 -0 0 -0.01 -0.1 -0.32 -0.24 -0.2 -0.47 0.01 -0.25 -0.32 -0.08 0.08 0 -0.32 -0.34 1 0.44 0.1 0.02 0.16
Net_Cash_Flow 0.51 0.2 0.4 0.41 0.58 0.51 0.47 0.33 -0.01 -0.01 0.01 -0 0.06 0.11 0.47 0.58 0.4 0.46 0.01 0.57 0.51 0.06 0.31 -0 0.47 0.69 0.44 1 -0.12 0.02 -0.22
ID -0.11 -0.11 -0.06 -0.19 -0.23 -0.22 -0.09 -0.18 0.04 0.03 -0.03 0.03 -0.05 -0.15 -0.22 -0.21 -0.19 -0.21 0.02 -0.21 -0.15 -0.02 -0.05 0.03 -0.22 -0.21 0.1 -0.12 1 0.01 0.37
Year 0.02 0.01 0.01 0.03 -0.01 -0.01 0.02 -0 -0.03 -0.03 0.03 -0.03 -0.06 0.02 -0.01 0.01 0.02 0 -0.03 0.01 -0.01 -0.03 0 -0.03 -0.01 0.01 0.02 0.02 0.01 1 0.01
Rating -0.26 -0.24 -0.15 -0.25 -0.37 -0.39 -0.23 -0.31 0.02 0.01 -0.01 0.01 -0.17 -0.37 -0.37 -0.37 -0.21 -0.42 -0.03 -0.37 -0.3 -0.08 -0.13 -0 -0.37 -0.35 0.16 -0.22 0.37 0.01 1
Assets_Total
Cash
Debt_in_Current_Liabilities_Total
Long_Term_Debt_Total
Earnings_Before_Interest
Gross_Profit_Loss
Liabilities_Total
Retained_Earnings
Total_debt_over_total_asset
total_asset_over_total_liabilities
EBTI_over_total_asset
gross_profit_over_rev
EBTI_over_REV
Dividends_per_Share_Pay_Date_Calendar
Sales_over_Turnover_Net
Stockholders_Equity_Total
Interest_and_Related_Expense_Total
Market_Value_Total_Fiscal
Book_Value_Per_Share
Common_Equity_Liquidation_Value
Comprehensive_Income_Parent
Employees
Inventories_Total
Earnings_Per_Share_from_Operations
Revenue_Total
Operating_Activities_Net_Cash_Flow
Financing_Activities_Net_Cash_Flow
Net_Cash_Flow
ID
Year
Rating