11/25/2023

Part A

George Mamvura

Table of Contents

Executive Summary

Businesses or companies can fall prey to default if they are not able to keep up their debt obligations. Defaults will lead to a lower credit rating for the company which in turn reduces its chances of getting credit in the future and may have to pay higher interest on existing debts as well as any new obligations. From an investor's point of view, he would want to invest in a company if it is capable of handling its financial obligations, can grow quickly, and is able to manage the growth scale.

A balance sheet is a financial statement of a company that provides a snapshot of what a company owns, owes, and the amount invested by the shareholders. Thus, it is an important tool that helps evaluate the performance of a business.

Data that is available includes information from the financial statement of the companies for the previous year.

Data Description

The dataset has multiple columns whose description can be found in the Python file as there are a lot of columns.

The data has 2058 rows and 58 columns

Sample of the dataset:

Here is a sample of the dataset showing the first 5 rows for a few columns.

Co_Code	Co_Name	_Operating_Expense_Rate	_Research_and_development_expense_rate	_Cash_flow_rate	_Interest_bearing_debt_interest_rate
16974	Hind.Cables	8.820000e+09	0.000000e+00	0.462045	0.000352
21214	Tata Tele. Mah.	9.380000e+09	4.230000e+09	0.460116	0.000716
14852	ABG Shipyard	3.800000e+09	8.150000e+08	0.449893	0.000496
2439	GTL	6.440000e+09	0.000000e+00	0.462731	0.000592
23505	Bharati Defence	3.680000e+09	0.000000e+00	0.463117	0.000782

Exploratory Data Analysis

The EDA was run using automatic methods and a further deep dive was ran. The EDA displayed in this document is for the variables that were determined as important for the highest performing model. Some of the EDA such as the distribution was shown in the automatic EDA.

Let us check the data

```
 #   Column                                   Non-Null Count  Dtype
---  ------                                   --------------  -----
 0   Co_Code                                  2058 non-null   int64
 1   Co_Name                                  2058 non-null   object
 2   Operating_Expense_Rate                   2058 non-null   float64
 3   Research_and_development_expense_rate    2058 non-null   float64
 4   Cash_flow_rate                           2058 non-null   float64
 5   Interest_bearing_debt_interest_rate      2058 non-null   float64
 6   Tax_rate_A                               2058 non-null   float64
 7   Cash_Flow_Per_Share                      1891 non-null   float64
 8   Per_Share_Net_profit_before_tax_Yuan_    2058 non-null   float64
 9   Realized_Sales_Gross_Profit_Growth_Rate  2058 non-null   float64
 10  Operating_Profit_Growth_Rate             2058 non-null   float64
 11  Continuous_Net_Profit_Growth_Rate        2058 non-null   float64
 12  Total_Asset_Growth_Rate                  2058 non-null   float64
 13  Net_Value_Growth_Rate                    2058 non-null   float64
 14  Total_Asset_Return_Growth_Rate_Ratio     2058 non-null   float64
 15  Cash_Reinvestment_perc                   2058 non-null   float64
 16  Current_Ratio                            2058 non-null   float64
 17  Quick_Ratio                              2058 non-null   float64
 18  Interest_Expense_Ratio                   2058 non-null   float64
 19  Total_debt_to_Total_net_worth            2037 non-null   float64
...
 56  Equity_to_Liability                      2058 non-null   float64
 57  Default                                  2058 non-null   int64
dtypes: float64(53), int64(4), object(1)
```

Q1 Outlier Treatment

Outliers were detected through various statistical methods; we visually identified the outliers and we statistically identified the points that are 1.5 standard deviations from the mean.

To mitigate the impact of the outliers and ensure a more robust analysis, we decided to apply data clipping to the data. We identified the threshold we would clip the values to and thus limited their influence on statistical measures and model outcomes. Upper and lower bounds for each column were identified according to their respective distributions and values beyond these bounds were replaced with the threshold values. This was applied consistently for all 49 columns.

Data Clipping has the following advantages:
1. Preserves the Data Integrity: Clipping allows for retention of valuable information in the dataset while minimising the impact of extreme values

2. Improved model stability: By removing the effect or outlier, the stability and reliability of statistical analyses and machine learning models are enhanced

Q2 Missing Value Treatment

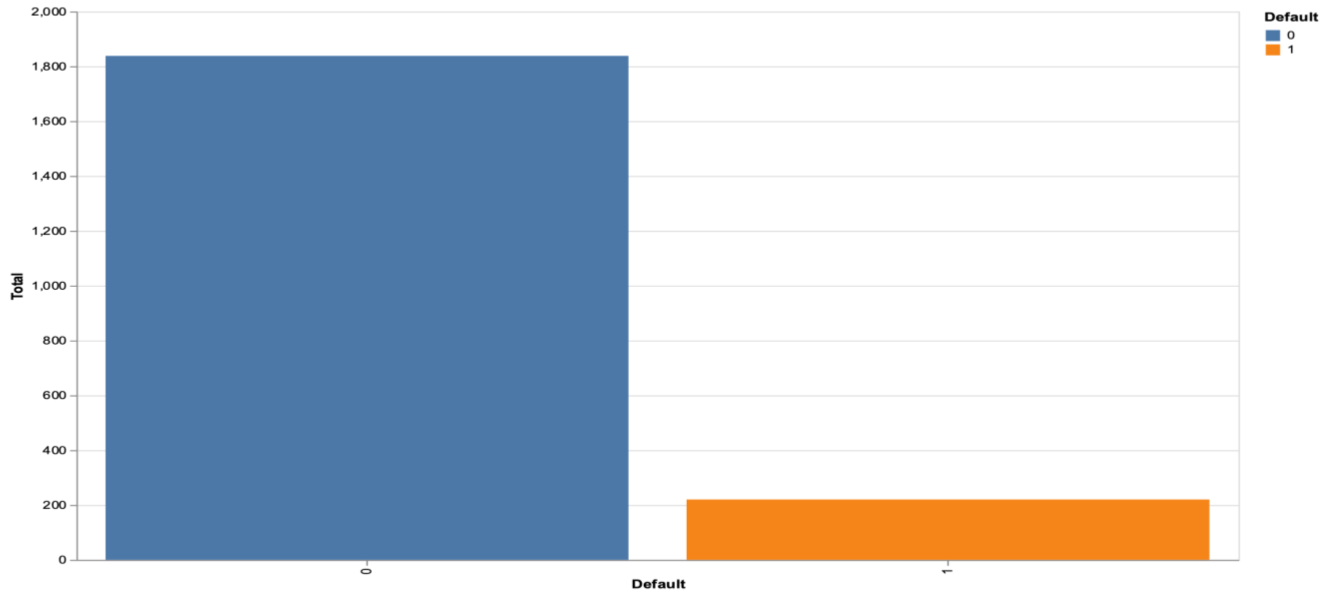
The data contains 298 missing values, and the missing values were replaced using KNN nearest neighbor imputation. KNN imputation involves predicting the missing values based on the values of their nearest neighbours in the dataset.

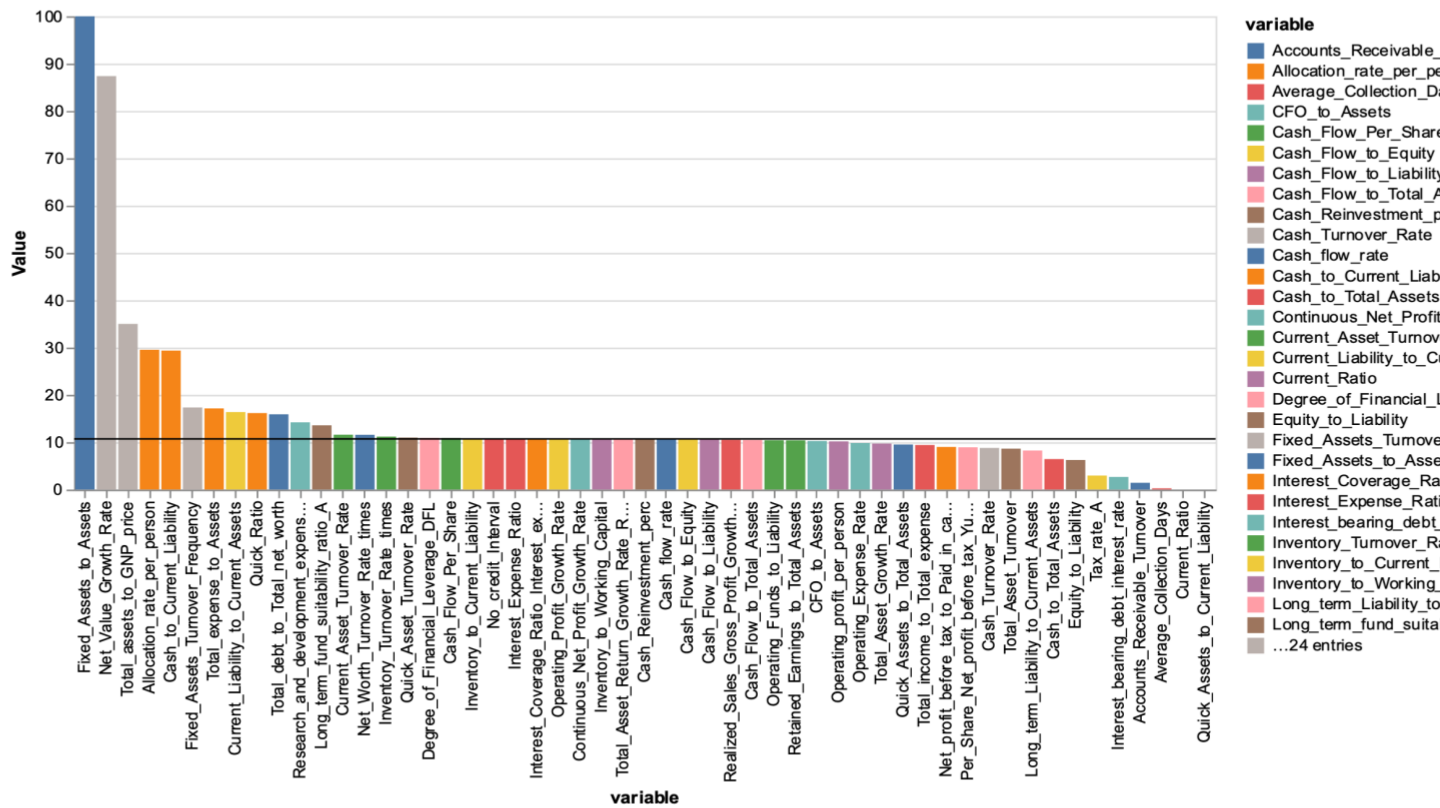
```
There are 298 missing values and you have to interpolate

 The missing columns and the total of missing values are:
 Cash_Flow_Per_Share                     167
Total_debt_to_Total_net_worth            21
Cash_to_Total_Assets                     96
Current_Liability_to_Current_Assets      14
dtype: int64
```

Q3. Univariate, Bivariate analysis

The variables are ordered in order of importance to the model and this showcases their distribution

The distribution of default as shown below shows that the default variable is unbalanced with 10.7% of the individuals in our dataset as defaulters and the remaining 89.3% as non-defaulters.

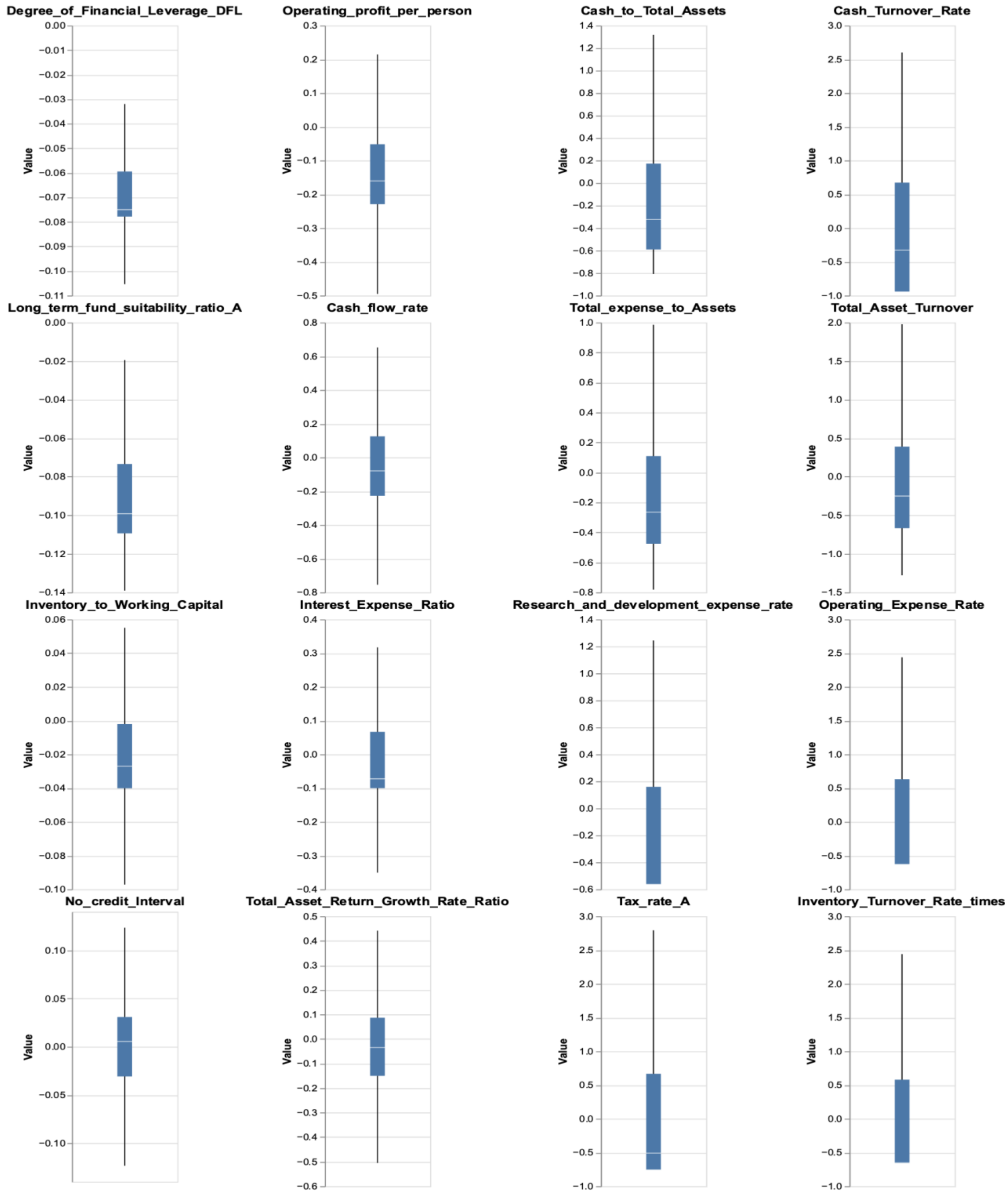


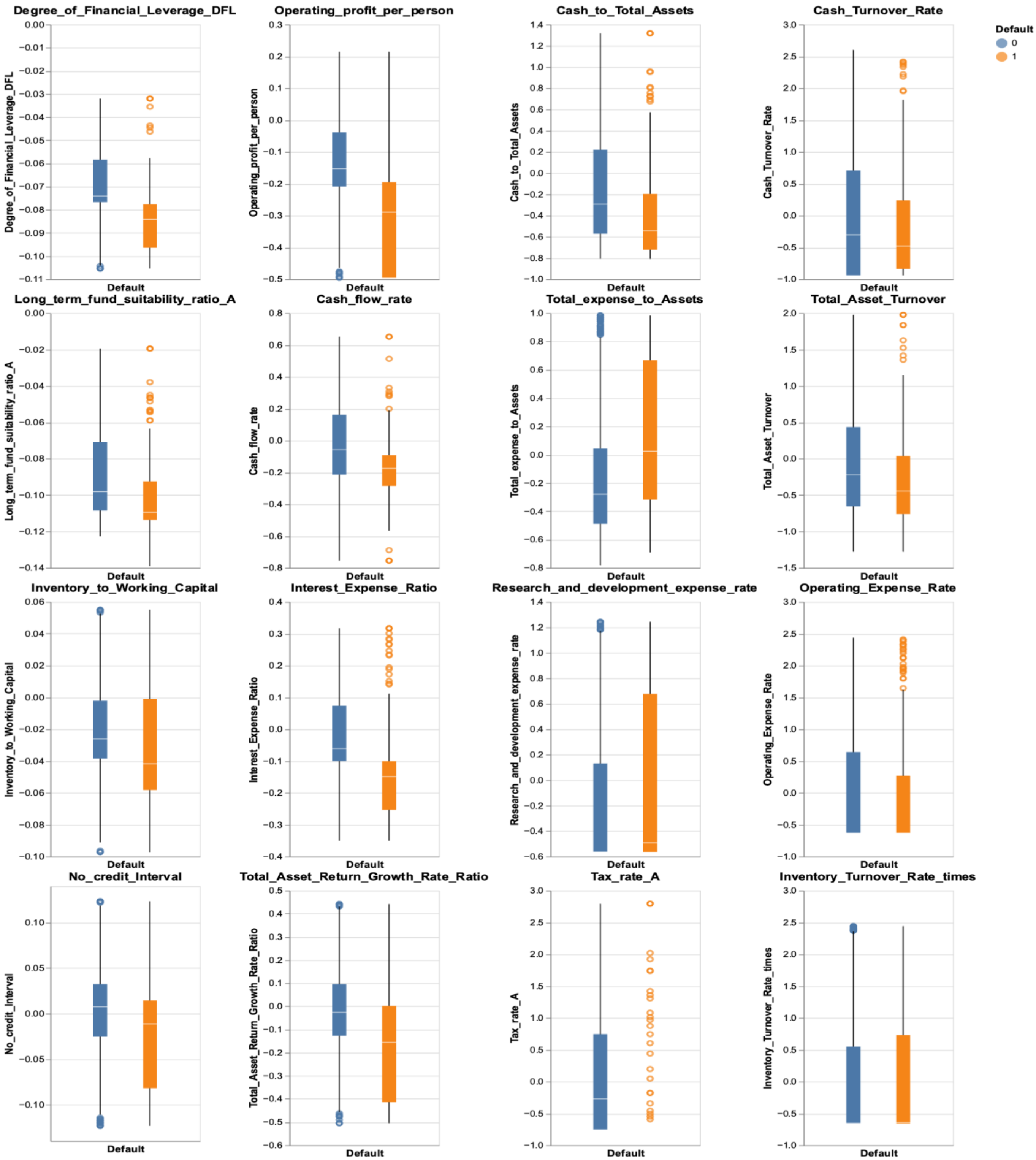


The graph above shows the percentage difference between the two groups, default and non-default in terms of mean value for the variables in the columns. We want to establish an initial hypothesis as to which columns could be important in determining the class of variables. The horizontal black line determines the percentage of defaulters in the overall data set. Deviations from this value would show the variables that could potentially be important in predicting default. Given the analysis I would expect the following variables to help predict Default:

I expect the following variables to help predict Defualt:

- Fixed_Assets_to_Assets (positive relationship), Net_Value_Growth_Rate (positive relationship), Total_assets_to_GNP_price (positive relationship), Cash_to_Current_Liability (positive relationship), Allocation_rate_per_person (positive relationship), Fixed_Assets_Turnover_Frequency (positive relationship), Total_expenses_to_Assets (positive relationship), Current_liability_to_Current_Assets (positive relationship), Tax_rate_A (negative relationship), Current_Ratio (negative relationship), Quick_Assets_to_Current_Liability (negative relationship), Accounts_Receivable_Turnover (negative relationship) and Average_Collection_Days (negative relationship).


Degree_of_Financial_Leverage_DFL
Operating_profit_per_person
Cash_to_Total_Assets
Cash_Turnover_Rate
Long_term_fund_suitability_ratio_A
Cash_flow_rate
Total_expense_to_Assets
Total_Asset_Turnover
Inventory_to_Working_Capital
Interest_Expense_Ratio
Research_and_development_expense_rate
Operating_Expense_Rate
No_credit_Interval
Total_Asset_Return_Growth_Rate_Ratio
Tax_rate_A
Inventory_Turnover_Rate_times
Value


Degree_of_Financial_Leverage_DFL
Operating_profit_per_person
Cash_to_Total_Assets
Cash_Turnover_Rate
Default
0
1
Long_term_fund_suitability_ratio_A
Cash_flow_rate
Total_expense_to_Assets
Total_Asset_Turnover
Inventory_to_Working_Capital
Interest_Expense_Ratio
Research_and_development_expense_rate
Operating_Expense_Rate
No_credit_Interval
Total_Asset_Return_Growth_Rate_Ratio
Tax_rate_A
Inventory_Turnover_Rate_times

We can see that you can see clear separation in terms of the variables and the probability of default with the separation reducing as the importance of the variables in predicting default decreases.

Q4. Train Test Split

We use a 33% split for the test size and 67% for the train size.

Imbalanced Data

We have noticed that there are imbalanced classes in the data and we correct this by using SMOTE which is (Synthetic Minority Oversampling Technique). Smote works by:

- creating synthetic samples from the minor class (no default) instead of creating copies

- randomly choosing one of the k-nearest-neighbours and using it to create a similar, but randomly tweaked, new observations.

Q5. Logistic regression model

Choosing Important variables

We also use step-wise regression to recursively remove variables with the highest p-value above 0.05 until all the variables that are left are significant for the model.

Logit Regression Results

Dep. Variable:	Default	No. Observations:	2462
Model:	Logit	Df Residuals:	2445
Method:	MLE	Df Model:	16
Date:	Tue, 21 Nov 2023	Pseudo R-squ.:	0.3930
Time:	09:29:44	Log-Likelihood:	-1035.9
converged:	True	LL-Null:	-1706.5
Covariance Type:	nonrobust	LLR p-value:	6.644e-276

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.4030	0.363	-12.142	0.000	-5.114	-3.692
No_credit_Interval	-4.7094	0.924	-5.098	0.000	-6.520	-2.899
Cash_flow_rate	-1.8361	0.233	-7.867	0.000	-2.294	-1.379
Interest_Expense_Ratio	1.4599	0.458	3.188	0.001	0.562	2.357
Inventory_to_Working_Capital	-5.1986	1.539	-3.377	0.001	-8.216	-2.181
Cash_to_Total_Assets	-1.2534	0.128	-9.829	0.000	-1.503	-1.003
Inventory_Turnover_Rate_times	-0.2413	0.063	-3.803	0.000	-0.366	-0.117
Research_and_development_expense_rate	0.6436	0.088	7.319	0.000	0.471	0.816
Total_expense_to_Assets	1.2556	0.158	7.946	0.000	0.946	1.565
Operating_profit_per_person	-2.6359	0.431	-6.111	0.000	-3.481	-1.790
Total_Asset_Return_Growth_Rate_Ratio	-1.2528	0.263	-4.758	0.000	-1.769	-0.737
Cash_Turnover_Rate	-0.4431	0.067	-6.600	0.000	-0.575	-0.311
Total_Asset_Turnover	-0.5161	0.105	-4.896	0.000	-0.723	-0.309
Degree_of_Financial_Leverage_DFL	-17.3491	3.994	-4.344	0.000	-25.177	-9.521
Tax_rate_A	-0.5249	0.093	-5.617	0.000	-0.708	-0.342
Long_term_fund_suitability_ratio_A	-16.5455	2.314	-7.152	0.000	-21.080	-12.011
Operating_Expense_Rate	0.2498	0.067	3.711	0.000	0.118	0.382

Cut-off

The optimal cut-off for the model is 0.44 and this improves the quality of predictions for the model.

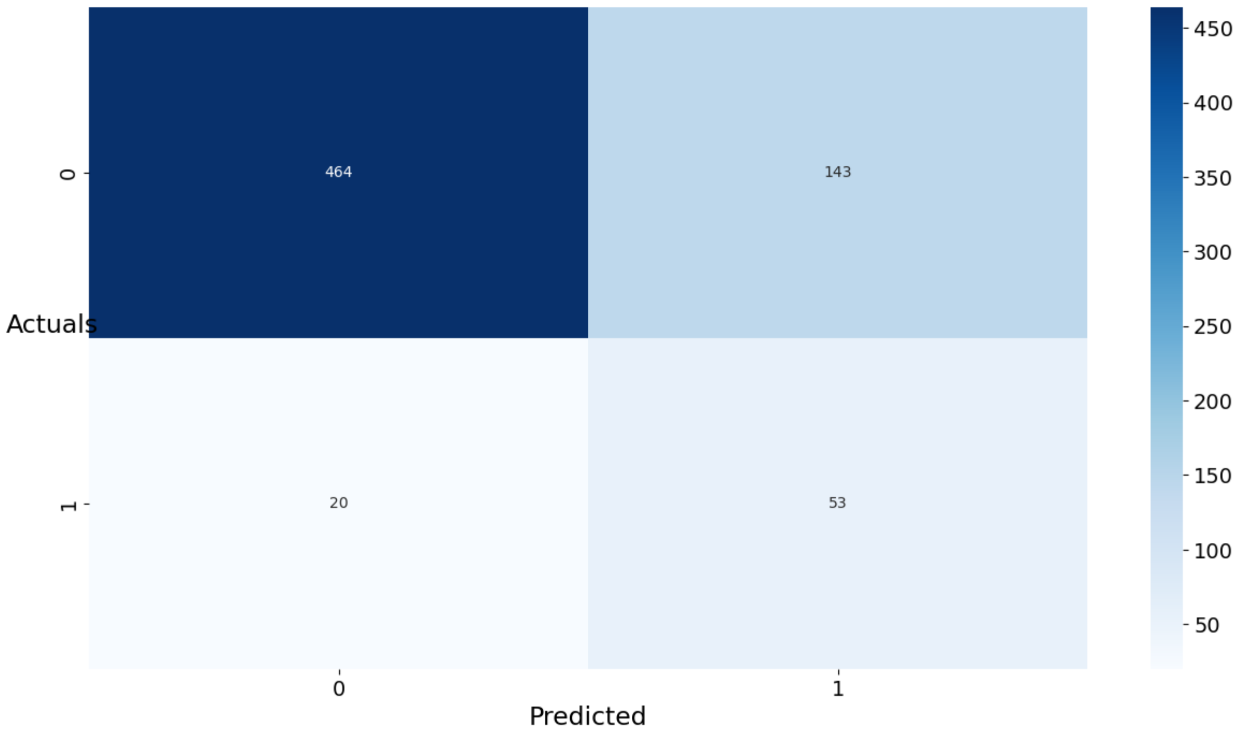
Model-building Approach

Given the collinearity of data and the large number of variables we will use VIF (Variable Inflation Factor) to filter the variables. The primary goal is to identify and mitigate issues arising from high intercorrelation among independent variables, which can impact the model's interpretability and stability. We choose a VIF factor of 4 to filter variables out.

We also filter out the columns with low variance, i.e. less than 0.0001 as they do not hold much information and they have the same value which will most likely not be useful in differentiating between default and non-default.

We also use step-wise regression to recursively remove variables with the highest p-value above 0.05 until all the variables that are left are significant for the model.

Q6. Model performance on Test



	precision	recall	f1-score	support
0	0.959	0.764	0.851	607
1	0.270	0.726	0.394	73
accuracy			0.760	680
macro avg	0.615	0.745	0.622	680
weighted avg	0.885	0.760	0.802	680

The model has a recall score of 0.73

Interpretation of Model variables

For the SMOTE dataset, we find the following observations as significant:

- No_credit_Interval: A more negative coefficient for No_credit_Interval suggests that a longer interval between credit is associated with a lower likelihood of default.

- Total_income_to_Total_expense (-3.5942): A more negative coefficient for Total_income_to_Total_expense indicates that a higher ratio of total income to total expense is associated with a lower likelihood of default.

- Cash_flow_rate: A more negative coefficient for Cash_flow_rate suggests that a higher cash flow rate is associated with a lower likelihood of default.

- Inventory_to_Working_Capitl: A more negative coefficient for Inventory_to_Working_Capital implies that a higher ratio of inventory to working capital is associated with a lower likelihood of default.

- Cash_to_Total_Assets: A more negative coefficient for Cash_to_Total_Assets suggests that a higher ratio of cash to total assets is associated with a lower likelihood of default.

- Research_and_development_expense_rate: A more positive coefficient for Research_and_development_expense_rate indicates that a higher research and development expense rate is associated with a higher likelihood of default.

- Cash_Flow_Per_Share: A more negative coefficient for Cash_Flow_Per_Share suggests that a higher cash flow per share is associated with a lower likelihood of default.

- Total_expense_to_Assets: A more positive coefficient for Total_expense_to_Assets indicates that a higher ratio of total expenses to assets is associated with a higher likelihood of default.

- Operating_profit_per_person: A more positive coefficient for Operating_profit_per_person suggests that a higher operating profit per person is associated with a higher likelihood of default.

- Cash_Turnover_Rate: A more negative coefficient for Cash_Turnover_Rate indicates that a higher cash turnover rate is associated with a lower likelihood of default.

- Total_Asset_Turnover: A more negative coefficient for Total_Asset_Turnover suggests that a higher total asset turnover is associated with a lower likelihood of default.

- Degree_of_Financial_Leverage_DFL: A more negative coefficient for Degree_of_Financial_Leverage_DFL indicates that a higher degree of financial leverage is associated with a lower likelihood of default.

- Long_term_fund_suitability_ratio_A : A more negative coefficient for Long_term_fund_suitability_ratio_A suggests that a higher long-term fund suitability ratio is associated with a lower likelihood of default.

- Operating_Expense_Rate: A more positive coefficient for Operating_Expense_Rate indicates that a higher operating expense rate is associated with a higher likelihood of default.

Q7. Build a Random Forest Model on Train Dataset. Also showcase your model building approach

The construction of a Random Forest model involves an iterative process aimed at optimizing hyperparameters for enhanced predictive performance. The following approach shows the systematic steps undertaken in this model-building process :

1. **Random Hyperparameter Search:**
    - **Objective:** The initial phase involves employing a random hyperparameter search algorithm. This algorithm systematically explores a wide range of hyperparameter combinations to identify an optimal set.
    - **Hyperparameter Exploration:** The algorithm probes various combinations of hyperparameters, including the number of trees (n_estimators), the number of features considered for splitting (max_features), maximum depth of each tree (max_depth), minimum number of samples required to split an internal node (min_samples_split), minimum number of samples required at a leaf node (min_samples_leaf), and the method of selecting samples for training each tree (bootstrap).
    - **Search Range Definition:** The search algorithm explores a predefined search space to ensure a comprehensive exploration of hyperparameter configurations.
2. **Optimal Hyperparameter Search Range:**
    - **Extraction of Optimal Parameters:** Once the random hyperparameter search is complete, the algorithm identifies the hyperparameter combination that yields the best performance metrics.
    - **Refinement of Search Range:** Based on the results obtained, a refined and narrower search range is established. This refined range will be used in subsequent fine-tuning steps.
3. **Fine-Tuning Hyperparameters:**
    - **Grid Search and Cross-Validation:** The refined search range is employed in a more granular optimization process. Grid search and cross-validation techniques are applied to further scrutinize the hyperparameter space and identify the most optimal combination.

- **Evaluation Metrics:** The performance of each hyperparameter configuration is assessed using the evaluation metrics recall.

4. **Model Training and Validation:**

    - **Training the Random Forest Model:** The final step involves training the Random Forest model using the identified optimal hyperparameters on the training dataset.

    - **Validation on Test Data:** The model's performance is then validated on an independent test dataset to ensure generalizability and assess its predictive capabilities on unseen data.

Benefits of the Approach:

- **Optimal Model Performance:** The systematic exploration and fine-tuning of hyperparameters aim to achieve the best possible performance from the Random Forest model.

- **Robust Generalization:** The model is designed to generalize well to unseen data, minimizing overfitting and ensuring reliability in real-world scenarios.

- **Efficient Exploration:** The combination of random search followed by fine-tuning strikes a balance between exploration and exploitation, making the hyperparameter search process efficient and effective.

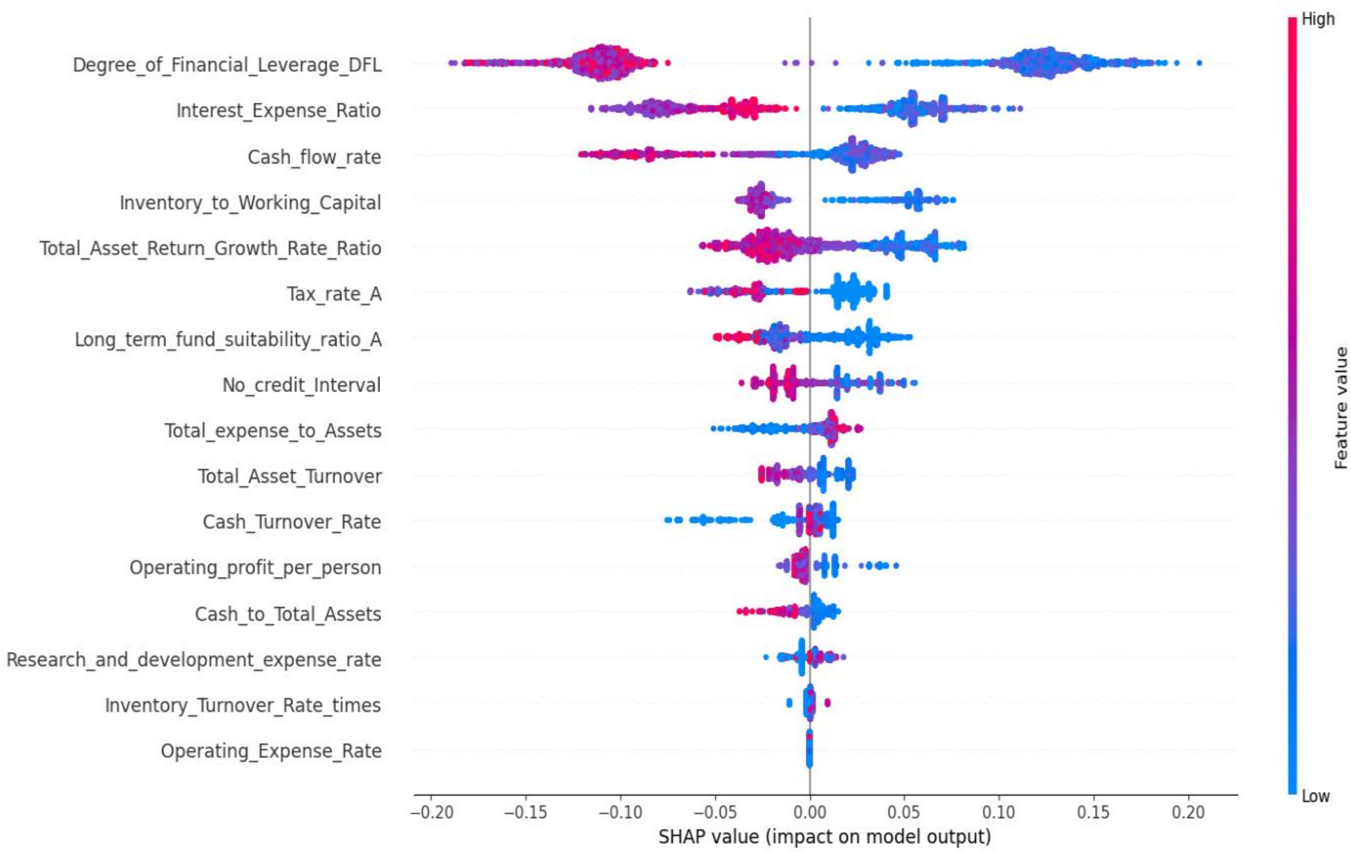
Q8. Validate the Random Forest Model on test Dataset and state the performance metrics. Also state interpretation from the model

	precision	recall	f1-score	support
0	0.97	0.78	0.86	607
1	0.30	0.79	0.44	73
accuracy			0.78	680
macro avg	0.64	0.79	0.65	680
weighted avg	0.90	0.78	0.82	680

The model score is 0.79 recall.

SHAP VALUES and VARIABLE IMPORTANCE

Shap values help us to interpret the impact of variables on the predicted variables.



1. **Long_term_fund_suitability_ratio_A:**
    - Firms with a higher Long-term Fund Suitability Ratio are less likely to default.
2. **Degree_of_Financial_Leverage_DFL:**
    - Firms with a higher Degree of Financial Leverage are less likely to default.
3. **Cash_to_Total_Assets:**
    - Firms with a higher proportion of cash to total assets are less likely to default.
4. **Total_Asset_Return_Growth_Rate_Ratio:**
    - Firms with a higher Total Asset Return Growth Rate Ratio are less likely to default.
5. **Operating_profit_per_person:**
    - Firms with higher operating profit per person are less likely to default.
6. **No_credit_Interval:**
    - Firms with a longer interval for not having credit are less likely to default.
7. **Inventory_to_Working_Capital :**
    - Firms with a lower ratio of inventory to working capital are less likely to default.
8. **Cash_flow_rate:**
    - Firms with a higher cash flow rate are less likely to default.
9. **Total_expense_to_Assets :**
    - Firms with a higher ratio of total expenses to assets are more likely to default.
10. **Interest_Expense_Ratio:**
    - Firms with a higher Interest Expense Ratio are more likely to default.
11. **Tax_rate_A :**
    - Firms with a higher tax rate are more likely to default.
12. **Cash_Turnover_Rate:**

- Firms with a higher cash turnover rate are more likely to default.

13. **Total_Asset_Turnover :**

  - Firms with a higher total asset turnover are more likely to default.

14. **Inventory_Turnover_Rate_times:**

  - Firms with a higher inventory turnover rate are more likely to default.

15. **Operating_Expense_Rate:**

  - Firms with a higher operating expense rate are more likely to default.

16. **Research_and_development_expense_rate :**

  - Firms with a higher research and development expense rate are more likely to default.

Q9. Build a LDA Model on Train Dataset. Also showcase your model building approach

The development of a Linear Discriminant Analysis (LDA) model involves a streamlined process that bypasses a random hyperparameter search and directly engages in a focused grid search with cross-validation. The following delineates the steps in this targeted model-building strategy:

1. **Grid Search with Cross-Validation:**

  - **Objective:** Unlike the random search utilized in some models, the LDA model starts with a grid search strategy that systematically explores a predefined hyperparameter space.

  - **Hyperparameter Grid Definition:** The grid encompasses key hyperparameters relevant to LDA, such as solver options, shrinkage parameters, and priors. Each combination within this grid is exhaustively evaluated during the search.

  - **Cross-Validation:** To robustly assess model performance, cross-validation is employed, typically in the form of k-fold cross-validation. This involves partitioning the dataset into k subsets, training the model on k-1 folds, and validating on the remaining fold. This process is repeated k times, and the average performance is computed.

2. **Optimal Hyperparameter Identification:**

    - **Evaluation Metrics:** The effectiveness of each hyperparameter configuration is gauged using appropriate evaluation metrics, considering the nature of the classification task. Common metrics include accuracy, precision, recall, or F1-score.

    - **Identification of Optimal Hyperparameters:** The combination of hyperparameters that yields the best average performance across the cross-validation folds is selected as the optimal configuration.

3. **Model Training and Validation:**

    - **Training the LDA Model:** With the optimal hyperparameters identified, the LDA model is trained on the entire training dataset.

    - **Validation on Test Data:** The model's performance is then validated on a separate test dataset to ensure its ability to generalize to new, unseen data.

Benefits of the Approach:

- **Precision in Hyperparameter Search:** By directly employing a grid search with cross-validation, the LDA model ensures a meticulous exploration of the hyperparameter space.

- **Efficient Resource Utilization:** The targeted grid search minimizes computational resources while maximizing the effectiveness of the hyperparameter tuning process.

- **Enhanced Generalization:** The model is honed to generalize well to unseen data, prioritizing stability and reliability in practical applications.

Q10. Validate the LDA Model on test Dataset and state the performance metrics. Also state interpretation from the model

The model performs well on the test set and scores 0.73.

	precision	recall	f1-score	support
0	0.96	0.78	0.86	607
1	0.29	0.73	0.41	73
accuracy			0.78	680
macro avg	0.62	0.75	0.64	680
weighted avg	0.89	0.78	0.81	680

Model Intepretation

The magnitude of the weight indicates the strength of the influence, and a high absolute weight suggests a more substantial impact on the linear combination forming the discriminant functions.

Degree_of_Financial_Leverage_DFL:

Higher values indicate a significant negative impact on the output. This suggests that companies with a higher degree of financial leverage are associated with lower likelihood of default.

Long_term_fund_suitability_ratio_A:

Higher values suggest a negative impact on the output. Companies with a lower long-term fund suitability ratio are associated with higher values of the dependent variable.

Inventory_to_Working_Capital:

Higher values indicate a negative impact. Companies with a higher ratio of inventory to working capital are associated with lower likelihood of default.

No_credit_Interval:

Higher values indicate a negative impact. A longer no credit interval is associated with lower likelihood of default.

Operating_profit_per_person:

Higher values suggest a negative impact. Companies with higher operating profit per person are associated with lower likelihood of default.

Cash_flow_rate:

Higher values indicate a negative impact. Companies with a higher cash flow rate are associated with lower likelihood of default.

Interest_Expense_Ratio:

Higher values suggest a positive impact. Companies with a higher interest expense ratio are associated with higher values of the dependent variable.

Total_Asset_Return_Growth_Rate_Ratio:

Higher values indicate a negative impact. Companies with a higher ratio are associated with lower likelihood of default.

Cash_to_Total_Assets:

Higher values indicate a negative impact. Companies with a higher ratio of cash to total assets are associated with lower likelihood of default.

Total_expense_to_Assets:

Higher values suggest a positive impact. Companies with a higher ratio of total expenses to total assets are associated with higher values of the dependent variable.

Research_and_development_expense_rate:

Higher values suggest a positive impact. Companies with a higher research and development expense rate are associated with higher values of the dependent variable.

Tax_rate_A:

Higher values indicate a negative impact. Companies with a higher tax rate are associated with lower likelihood of default.

Cash_Turnover_Rate:

Higher values indicate a negative impact. Companies with a higher cash turnover rate are associated with lower likelihood of default.

Total_Asset_Turnover:

Higher values indicate a negative impact. Companies with a higher total asset turnover are associated with lower likelihood of default.

Operating_Expense_Rate:

Higher values suggest a positive impact. Companies with a higher operating expense rate are associated with higher values of the dependent variable.

Inventory_Turnover_Rate_times:

Higher values indicate a negative impact. Companies with a higher inventory turnover rate are associated with lower likelihood of default.

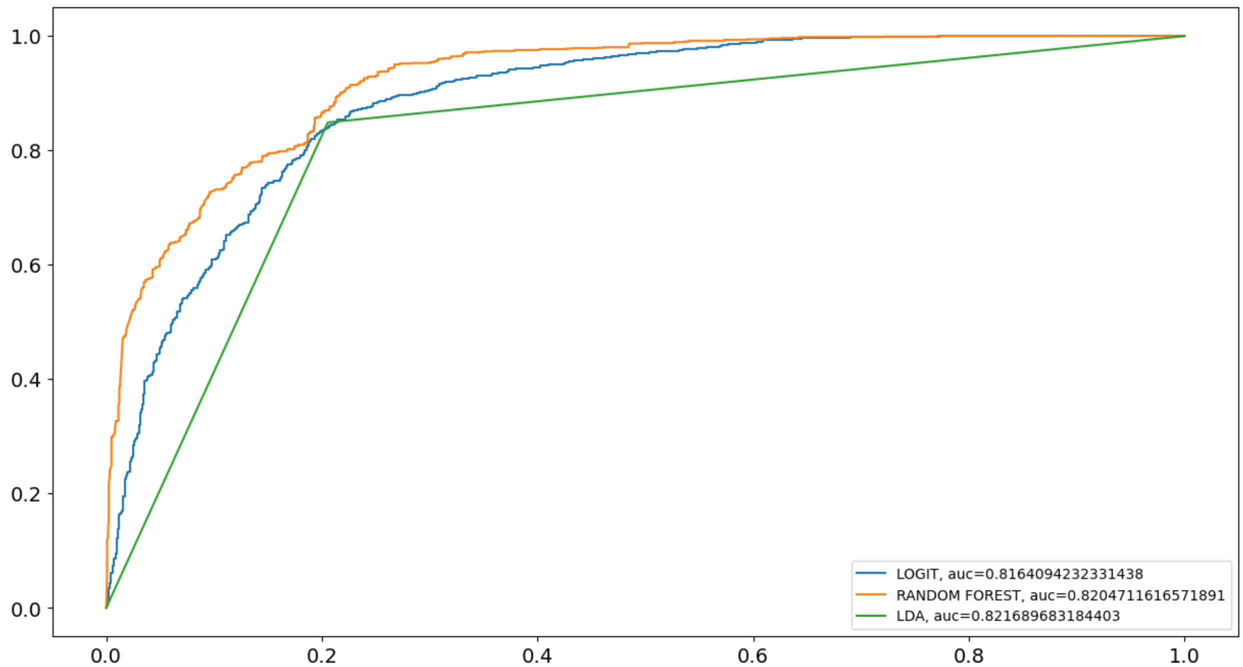
Q11. Compare the performances of Logistic Regression, Random Forest, and LDA models (include ROC curve)

Model	Recall
Logit	0.726
Random Forest	0.79

LDA	0.73

We can see that the Random Forest model is the most performant model with a recall of 0.79

ROC CURVE



Q12. Conclusions and Recommendations

The analysis of the LDA model weights provides valuable insights into the factors influencing the discriminant function and, consequently, the probability of default for the considered dataset. The magnitude and direction of the weights highlight the relative importance of each variable in shaping the discriminant functions. Based on these findings, several observations and recommendations can be made:

Risk Indicators:

1. There are a few risk indicators that group the factors together and they are given below:
   Liquidity Risk Indicators:
    - Cash_to_Total_Assets
    - Cash_flow_rate

- Cash_Turnover_Rate

2. **Operational Efficiency Risk Indicators:**
    - Operating_profit_per_person
    - No_credit_Interval
    - Inventory_Turnover_Rate_times
    - Research_and_development_expense_rate
    - Total_Asset_Turnover
3. **Financial Leverage Risk Indicators:**
    - Degree_of_Financial_Leverage_DFL
    - Long_term_fund_suitability_ratio_A
4. **Cost Management and Profitability Risk Indicators:**
    - Total_expense_to_Assets
    - Operating_Expense_Rate
    - Interest_Expense_Ratio
    - Total_Asset_Return_Growth_Rate_Ratio
5. **Tax and Financial Planning Risk Indicators:**
    - Tax_rate_A

Overall Recommendations to Reduce Default:

- **Improve Liquidity:** Maintain a healthy proportion of cash to total assets.
- **Manage Leverage:** Be cautious with the degree of financial leverage.
- **Effective Cost Management:** Control total expenses relative to assets.
- **Reduce Interest Expenses:** Lower the interest expense ratio.
- **Tax Planning:** Consider strategies to optimize tax rates.

- **Efficient Cash Flow:** Enhance cash flow rates.
- **Operational Efficiency:** Improve operating profit per person.
- **Credit Management:** Extend the interval without credit.
- **Inventory Management:** Keep a lower ratio of inventory to working capital.
- **Strategic Investment:** Monitor and control research and development expenses.
- **Cost of Goods Sold Efficiency:** Optimize inventory turnover rates.
- **Cash Turnover:** Control the cash turnover rate.
- **Asset Efficiency:** Manage total asset turnover effectively.

1. Liquidity Risk Indicators:

- **Analysis:**
    - **Impact on Default Probability:** High liquidity, as indicated by a higher cash-to-total-assets ratio, strong cash flow rates, and efficient cash turnover, is associated with a lower probability of default. Companies with good liquidity are better positioned to meet their financial obligations, reducing the risk of default.
- **Recommendation Impact:**
    - Implementing the recommendations in this category is likely to decrease the probability of default by enhancing the company's ability to handle short-term financial challenges.

2. Operational Efficiency Risk Indicators:

- **Analysis:**
    - **Impact on Default Probability:** Operational efficiency, reflected in factors like operating profit per person, credit management (no_credit_Interval), and inventory turnover, is associated with a lower probability of default. Efficient operations contribute to stable financial performance.
- **Recommendation Impact:**

- Following the recommendations in this category is likely to decrease the probability of default by improving overall operational effectiveness and financial stability.

3. Financial Leverage Risk Indicators:

- **Analysis:**
    - **Impact on Default Probability:** Higher financial leverage (Degree_of_Financial_Leverage_DFL) is associated with an increased probability of default. Managing leverage is crucial for financial stability.
- **Recommendation Impact:**
    - The recommendations here focus on exercising caution with financial leverage, and implementing them can reduce the probability of default by avoiding excessive financial risk.

4. Cost Management and Profitability Risk Indicators:

- **Analysis:**
    - **Impact on Default Probability:** Efficient cost management, controlled total expenses, and positive profitability indicators are associated with a lower probability of default. Conversely, high interest expense ratios increase the probability of default.
- **Recommendation Impact:**
    - Implementing recommendations in this category is likely to decrease the probability of default by improving cost management, reducing expenses, and optimizing profitability.

5. Tax and Financial Planning Risk Indicators:

- **Analysis:**
    - **Impact on Default Probability:** Higher tax rates (Tax_rate_A) are associated with an increased probability of default. Strategic tax planning is crucial for financial health.
- **Recommendation Impact:**

- Strategic tax planning, as recommended, is likely to decrease the probability of default by optimizing the company's tax position.