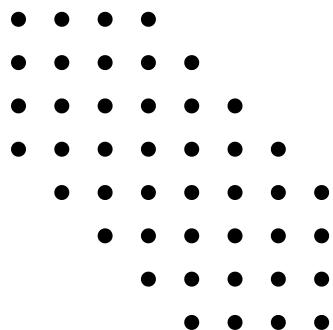


# Credit Risk Analysis Report

**Presented by :**  
Sanjay Rajan J



# TABLE OF CONTENTS

<b>CHAPTER NO</b>	<b>CONTENT</b>	<b>PAGE NO</b>
	<b>LIST OF FIGURES</b>	3
1.	<b>DATA OVERVIEW</b>	5
2.	<b>EXPLORATORY DATA ANALYSIS</b>	13
	2.1 Univariate Analysis	13
	2.2 Bivariate Analysis	17
3.	<b>DATA PREPROCESSING</b>	20
4.	<b>MODEL BUILDING</b>	22
	4.1 Logistic Regression	22
	4.2 Random Forest Classifier	23
5.	<b>MODEL PERFORMANCE IMPROVEMENT</b>	24
	5.1 Logistic Regression	24
	5.2 Random Forest Classifier	26
6.	<b>MODEL PERFORMANCE COMPARISON AND FINAL MODEL SELECTION</b>	28
7.	<b>ACTIONABLE INSIGHTS &amp; RECOMMENDATIONS</b>	30

# LIST OF FIGURES

<b>FIG NO.</b>	<b>NAME</b>	<b>PAGE</b>
1.	Data Info	10
2.	Numerical Statistics 1	11
3.	Numerical Statistics 2	12
4.	Total assets distribution	13
5.	Net_worth distribution	13
6.	Total_income distribution	13
7.	Sales distribution	14
8.	Borrowings distribution	14
9.	Profit_after_tax distribution	14
10.	EPS distribution	15
11.	Cash_Profit distribution	15
12.	Debt_to_equity_ratio_times distribution	15
13.	default distribution	16
14.	Heatmap	17
15.	Boxplots of variables w.r.t default	18
16.	Null value count	20
17.	Null value percentage	21
18.	Model Summary 1	22
19.	Logit Model Perf 1	23
20.	RFC Model Perf 1	23
21.	ROC curve train	24
22.	Logit Model Summary	25
23.	Logit Model Perf 2	26

24.	RFC Model Perf 2	27
25.	Training set Perf	28
26.	Test set Perf	28
27.	Top 15 important features	30

# **1. DATA OVERVIEW**

## **CONTEXT**

In the realm of modern finance, businesses encounter the perpetual challenge of managing debt obligations effectively to maintain a favourable credit standing and foster sustainable growth. Investors keenly scrutinize companies capable of navigating financial complexities while ensuring stability and profitability. A pivotal instrument in this evaluation process is the balance sheet, which provides a comprehensive overview of a company's assets, liabilities, and shareholder equity, offering insights into its financial health and operational efficiency. In this context, leveraging available financial data, particularly from preceding fiscal periods, becomes imperative for informed decision-making and strategic planning.

## **OBJECTIVE**

A group of venture capitalists want to develop a Financial Health Assessment Tool. With the help of the tool, it endeavours to empower businesses and investors with a robust mechanism for evaluating the financial well-being and creditworthiness of companies. By harnessing machine learning techniques, they aim to analyse historical financial statements and extract pertinent insights to facilitate informed decision-making via the tool. Specifically, they foresee facilitating the following with the help of the tool:

1. Debt Management Analysis: Identify patterns and trends in debt management practices to assess the ability of businesses to fulfill financial obligations promptly and efficiently, and identify potential cases of default.
2. Credit Risk Evaluation: Evaluate credit risk exposure by analysing liquidity ratios, debt-to-equity ratios, and other key financial indicators to ascertain the likelihood of default and inform investment decisions.

They have hired you as a data scientist and provided you with the financial metrics of different companies. The task is to analyse the data provided and develop a predictive model leveraging machine learning techniques to identify whether a given company will be tagged as a defaulter in terms of net worth next year. The predictive model will help the organization anticipate potential challenges with the financial performance of the companies and enable proactive risk mitigation strategies.

#### DATA DICTIONARY:

- Net worth Next Year: Net worth of the customer in the next year
- Total assets: Total assets of customer
- Net worth: Net worth of the customer of the present year
- Total income: Total income of the customer
- Change in stock: Difference between the current value of the stock and the value of stock in the last trading day
- Total expenses: Total expenses done by the customer
- Profit after tax: Profit after tax deduction

- PBDITA: Profit before depreciation, income tax, and amortization
- PBT: Profit before tax deduction
- Cash profit: Total Cash profit
- PBDITA as % of total income: PBDITA / Total income
- PBT as % of total income: PBT / Total income
- PAT as % of total income: PAT / Total income
- Cash profit as % of total income: Cash Profit / Total income
- PAT as % of net worth: PAT / Net worth
- Sales: Sales done by the customer
- Income from financial services: Income from financial services
- Other income: Income from other sources
- Total capital: Total capital of the customer
- Reserves and funds: Total reserves and funds of the customer
- Borrowings: Total amount borrowed by the customer
- Current liabilities & provisions: current liabilities of the customer
- Deferred tax liability: Future income tax customer will pay because of the current transaction
- Shareholders funds: Amount of equity in a company which belongs to shareholders
- Cumulative retained profits: Total cumulative profit retained by customer
- Capital employed: Current asset minus current liabilities

- TOL/TNW: Total liabilities of the customer divided by Total net worth
- Total term liabilities / tangible net worth: Short + long term liabilities divided by tangible net worth
- Contingent liabilities / Net worth (%): Contingent liabilities / Net worth
- Contingent liabilities: Liabilities because of uncertain events
- Net fixed assets: The purchase price of all fixed assets
- Investments: Total invested amount
- Current assets: Assets that are expected to be converted to cash within a year
- Net working capital: Difference between the current liabilities and current assets
- Quick ratio (times): Total cash divided by current liabilities
- Current ratio (times): Current assets divided by current liabilities
- Debt to equity ratio (times): Total liabilities divided by its shareholder equity
- Cash to current liabilities (times): Total liquid cash divided by current liabilities
- Cash to average cost of sales per day: Total cash divided by the average cost of the sales
- Creditors turnover: Net credit purchase divided by average trade creditors

- Debtors turnover: Net credit sales divided by average accounts receivable
- Finished goods turnover: Annual sales divided by average inventory
- WIP turnover: The cost of goods sold for a period divided by the average inventory for that period
- Raw material turnover: Cost of goods sold is divided by the average inventory for the same period
- Shares outstanding: Number of issued shares minus the number of shares held in the company
- Equity face value: cost of the equity at the time of issuing
- EPS: Net income divided by the total number of outstanding share
- Adjusted EPS: Adjusted net earnings divided by the weighted average number of common shares outstanding on a diluted basis during the plan year
- Total liabilities: Sum of all types of liabilities
- PE on BSE: Company's current stock price divided by its earnings per share

➤ Shape:

There are 4256 rows and 51 columns in this dataset.

➤ Basic Info:

**Fig.1 Data Info**

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4256 entries, 0 to 4255
Data columns (total 51 columns):
 #   Column           Non-Null Count Dtype  
 --- 
 0   Num              4256 non-null   int64  
 1   Networth_Next_Year 4256 non-null   float64 
 2   Total_assets       4256 non-null   float64 
 3   Net_worth          4256 non-null   float64 
 4   Total_income        4025 non-null   float64 
 5   Change_in_stock    3706 non-null   float64 
 6   Total_expenses     4091 non-null   float64 
 7   Profit_after_tax   4102 non-null   float64 
 8   PBDITA             4102 non-null   float64 
 9   PBT                4102 non-null   float64 
 10  Cash_profit        4102 non-null   float64 
 11  PBDITA_as_perc_of_total_income 4177 non-null   float64 
 12  PBT_as_perc_of_total_income   4177 non-null   float64 
 13  PAT_as_perc_of_total_income 4177 non-null   float64 
 14  Cash_profit_as_perc_of_total_income 4177 non-null   float64 
 15  PAT_as_perc_of_net_worth    4256 non-null   float64 
 16  Sales               3951 non-null   float64 
 17  Income_from_fincial_services 3145 non-null   float64 
 18  Other_income         2700 non-null   float64 
 19  Total_capital        4251 non-null   float64 
 20  Reserves_and_funds   4158 non-null   float64 
 21  Borrowings           3825 non-null   float64 
 22  Current_liabilities_&_provisions 4146 non-null   float64 
 23  Deferred_tax_liability 2887 non-null   float64 
 24  Shareholders_funds   4256 non-null   float64 
 25  Cumulative_retained_profits 4211 non-null   float64 
 26  Capital_employed     4256 non-null   float64 
 27  TOL_to_TNW           4256 non-null   float64 
 28  Total_term_liabilities__to__tangible_net_worth 4256 non-null   float64 
 29  Contingent_liabilities__to__Net_worth_perc   4256 non-null   float64 
 30  Contingent_liabilities 2854 non-null   float64 
 31  Net_fixed_assets     4124 non-null   float64 
 32  Investments          2541 non-null   float64 
 33  Current_assets        4176 non-null   float64 
 34  Net_working_capital   4219 non-null   float64 
 35  Quick_ratio_times     4151 non-null   float64 
 36  Current_ratio_times   4151 non-null   float64 
 37  Debt_to_equity_ratio_times 4256 non-null   float64 
 38  Cash_to_current_liabilities_times 4151 non-null   float64 
 39  Cash_to_average_cost_of_sales_per_day 4156 non-null   float64 
 40  Creditors_turnover    3865 non-null   float64 
 41  Debtors_turnover      3871 non-null   float64 
 42  Finished_goods_turnover 3382 non-null   float64 
 43  WIP_turnover          3492 non-null   float64 
 44  Raw_material_turnover 3828 non-null   float64 
 45  Shares_outstanding    3446 non-null   float64 
 46  Equity_face_value      3446 non-null   float64 
 47  EPS                  4256 non-null   float64 
 48  Adjusted_EPS          4256 non-null   float64 
 49  Total_liabilities     4256 non-null   float64 
 50  PE_on_BSE             1629 non-null   float64 
dtypes: float64(50), int64(1)
memory usage: 1.7 MB
```



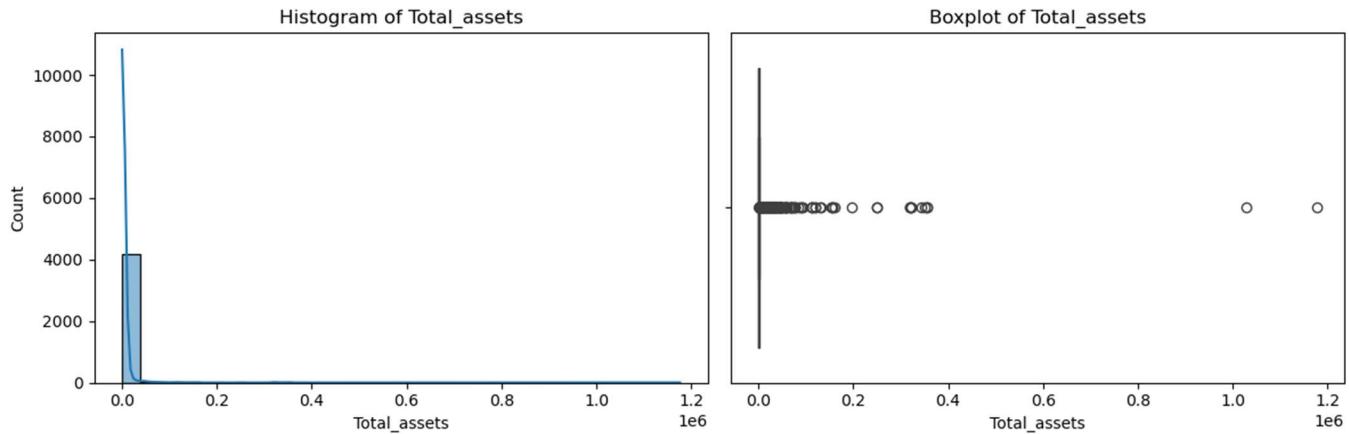
### Fig.3 Numerical Statistics 2

Total_term_liabilities_to_tangible_net_worth	4256.00	1.85	15.88	-325.60	0.05	0.34	1.00	456.00
Contingent_liabilities_to_Net_worth_perc	4256.00	55.71	369.17	0.00	0.00	5.36	31.01	14704.27
Contingent_liabilities	2854.00	948.55	12056.74	0.10	6.00	37.85	195.32	559506.80
Net_fixed_assets	4124.00	1209.49	12502.40	0.00	26.20	93.85	352.82	636604.60
Investments	2541.00	721.87	6793.86	0.00	1.00	8.20	63.80	199978.60
Current_assets	4176.00	1350.36	10155.57	0.10	36.60	148.35	515.00	354815.20
Net_working_capital	4219.00	162.87	3182.03	-63839.00	-1.10	16.70	86.50	85782.80
Quick_ratio_times	4151.00	1.50	9.33	0.00	0.41	0.67	1.03	341.00
Current_ratio_times	4151.00	2.26	12.48	0.00	0.93	1.23	1.72	505.00
Debt_to_equity_ratio_times	4256.00	2.87	15.60	0.00	0.22	0.79	1.75	456.00
Cash_to_current_liabilities_times	4151.00	0.53	4.80	0.00	0.02	0.07	0.19	165.00
Cash_to_average_cost_of_sales_per_day	4156.00	145.16	2521.99	0.00	2.88	8.04	21.97	128040.76
Creditors_turnover	3865.00	16.81	75.67	0.00	3.72	6.17	11.69	2401.00
Debtors_turnover	3871.00	17.93	90.16	0.00	3.81	6.47	11.85	3135.20
Finished_goods_turnover	3382.00	84.37	562.64	-0.09	8.19	17.32	40.01	17947.60
WIP_turnover	3492.00	28.68	169.65	-0.18	5.10	9.86	20.24	5651.40
Raw_material_turnover	3828.00	17.73	343.13	-2.00	3.02	6.41	11.82	21092.00
Shares_outstanding	3446.00	23764909.56	170979041.33	-2147483647.00	1308382.50	4750000.00	10906020.00	4130400545.00
Equity_face_value	3446.00	-1094.83	34101.36	-999998.90	10.00	10.00	10.00	100000.00
EPS	4256.00	-196.22	13061.95	-843181.82	0.00	1.49	10.00	34522.53
Adjusted_EPS	4256.00	-197.53	13061.93	-843181.82	0.00	1.24	7.62	34522.53
Total_liabilities	4256.00	3573.62	30074.44	0.10	91.30	315.50	1120.80	1176509.20
PE_on_BSE	1629.00	55.46	1304.45	-1116.64	2.97	8.69	17.00	51002.74

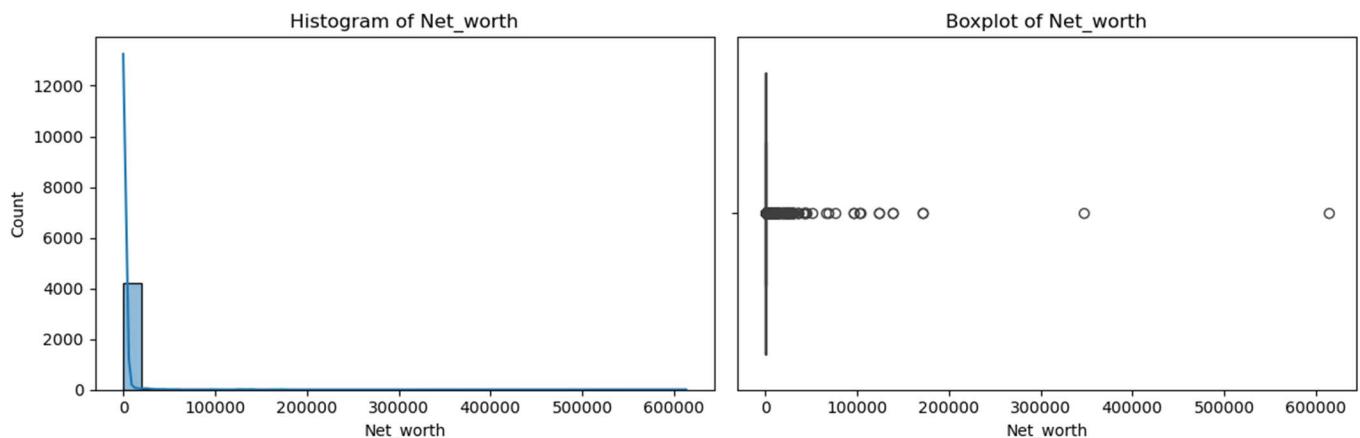
## 2. EXPLORATORY DATA ANALYSIS

### 2.1 UNIVARIATE ANALYSIS:

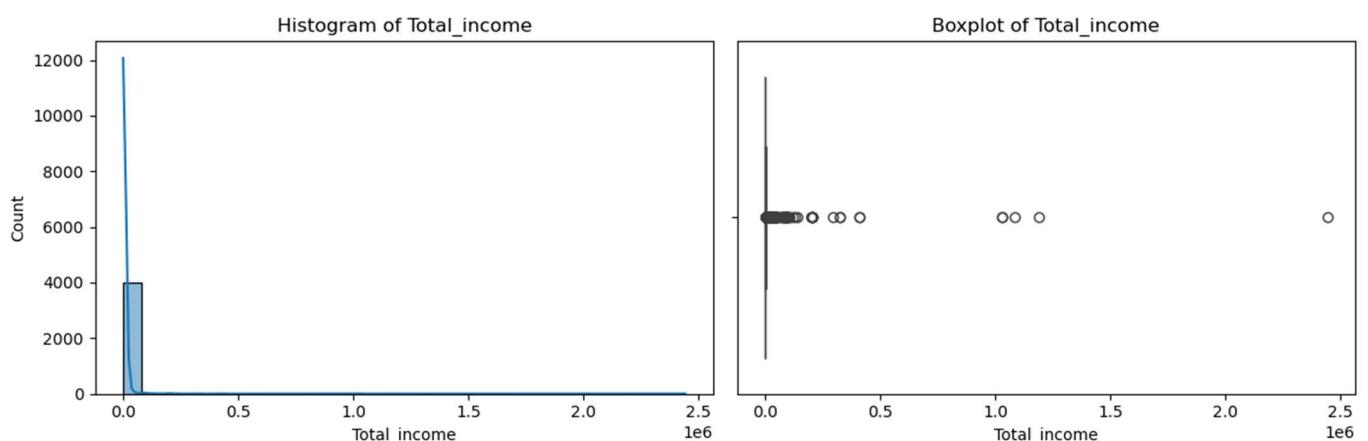
**Fig.4 Total assets distribution**



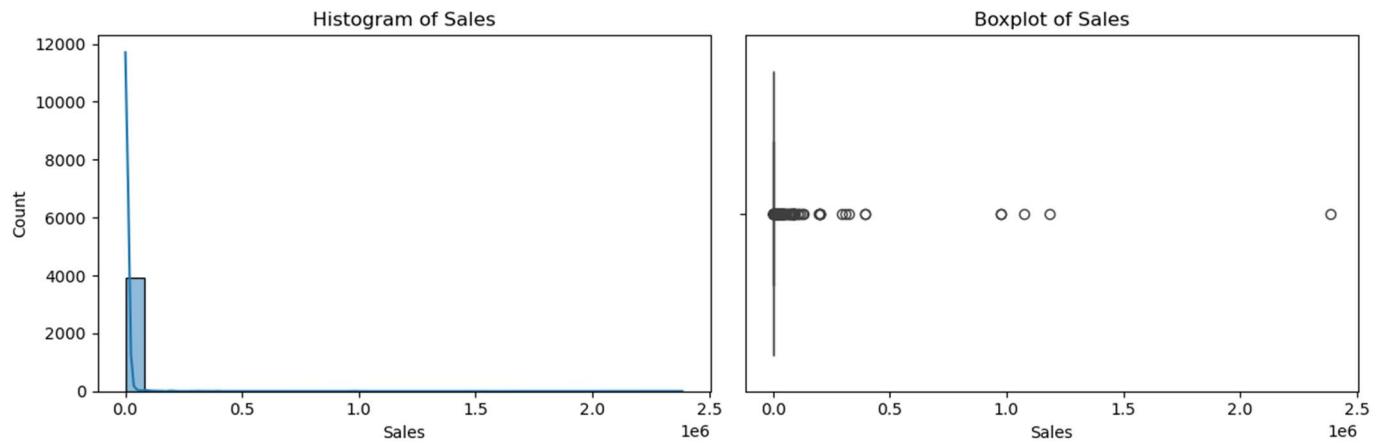
**Fig.5 Net\_worth distribution**



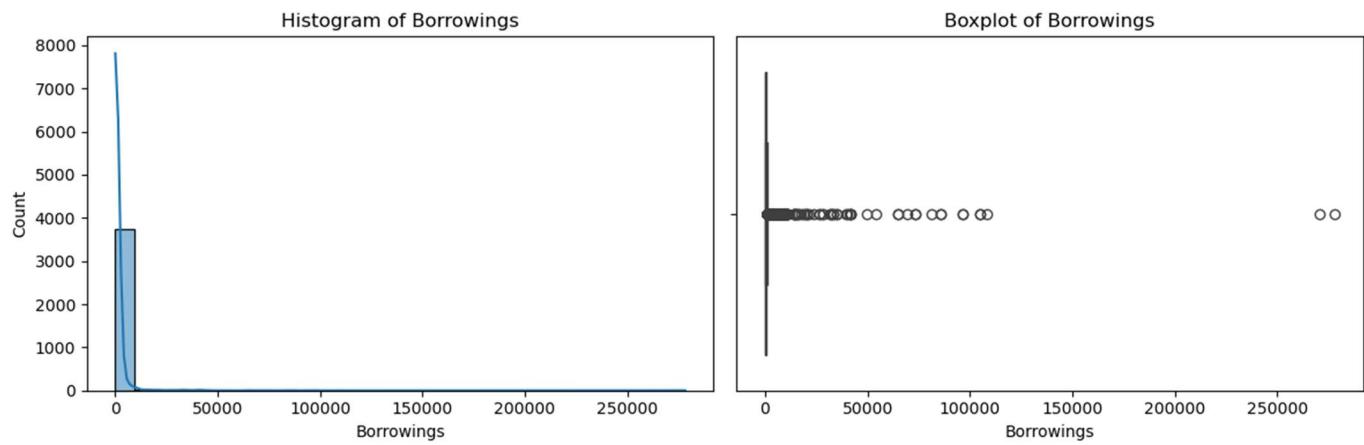
**Fig.6 Total\_income distribution**



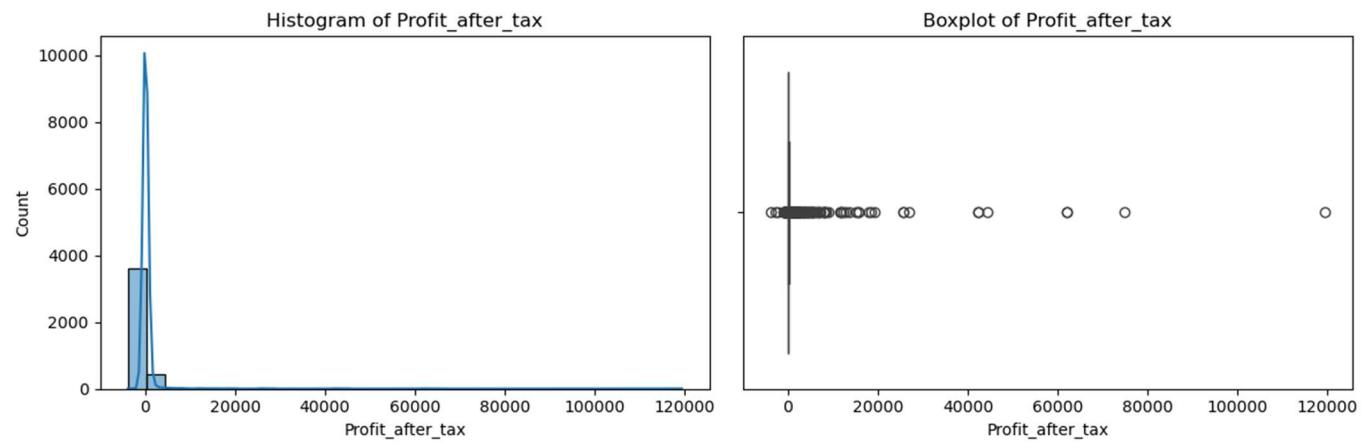
**Fig.7 Sales distribution**



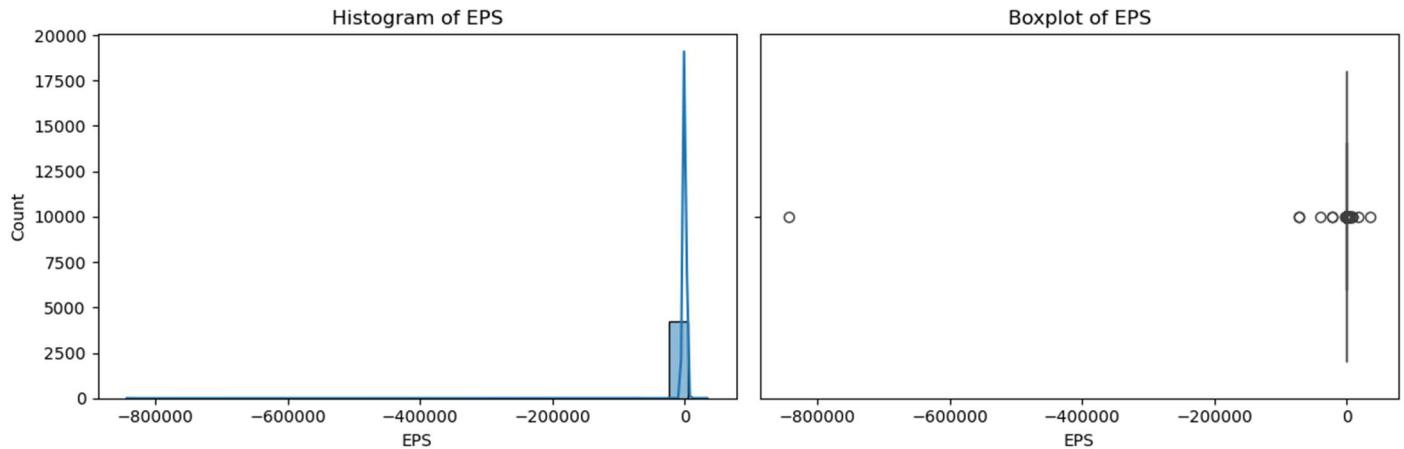
**Fig.8 Borrowings distribution**



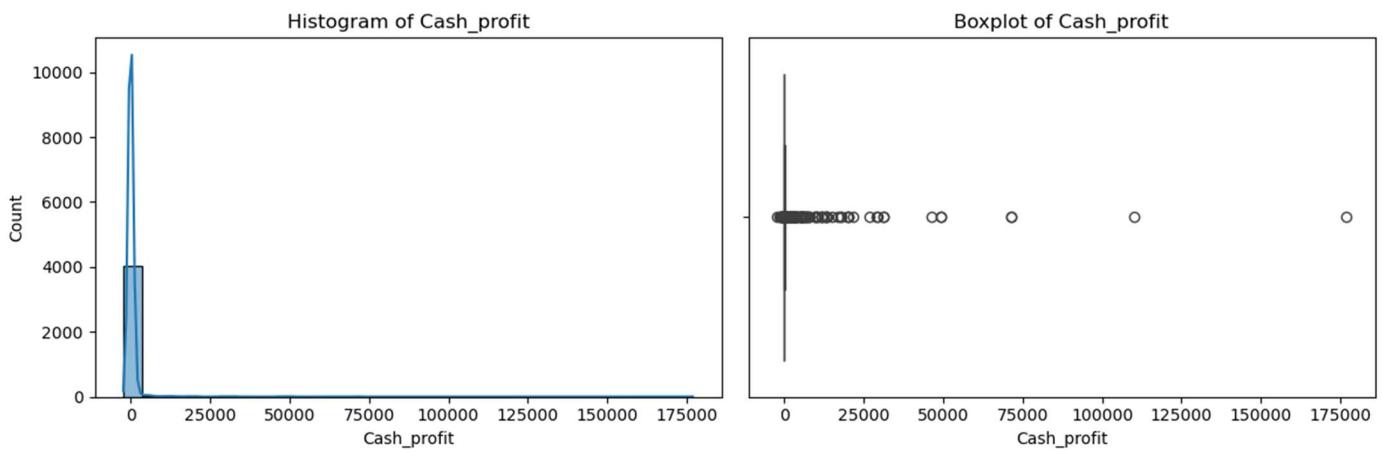
**Fig.9 Profit\_after\_tax distribution**



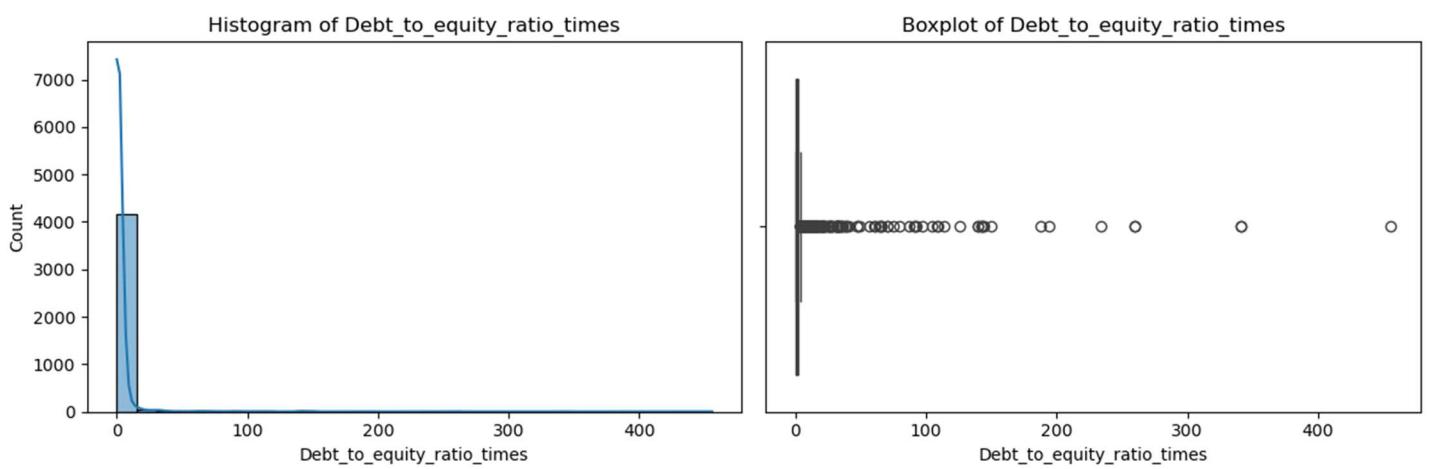
**Fig.10 EPS distribution**



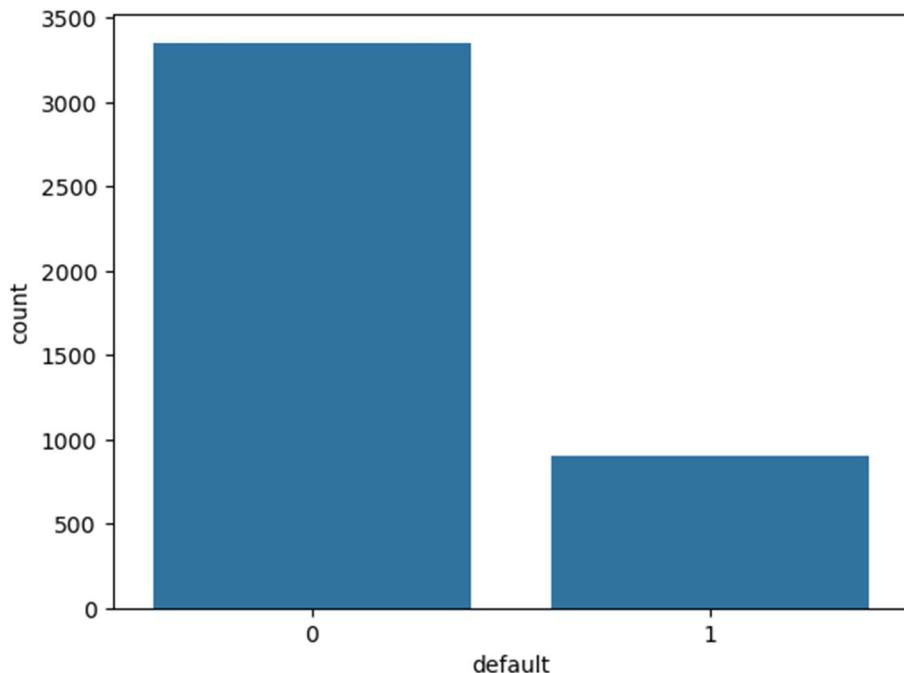
**Fig.11 Cash\_Profit distribution**



**Fig.12 Debt\_to\_equity\_ratio\_times distribution**



**Fig.13 default distribution**

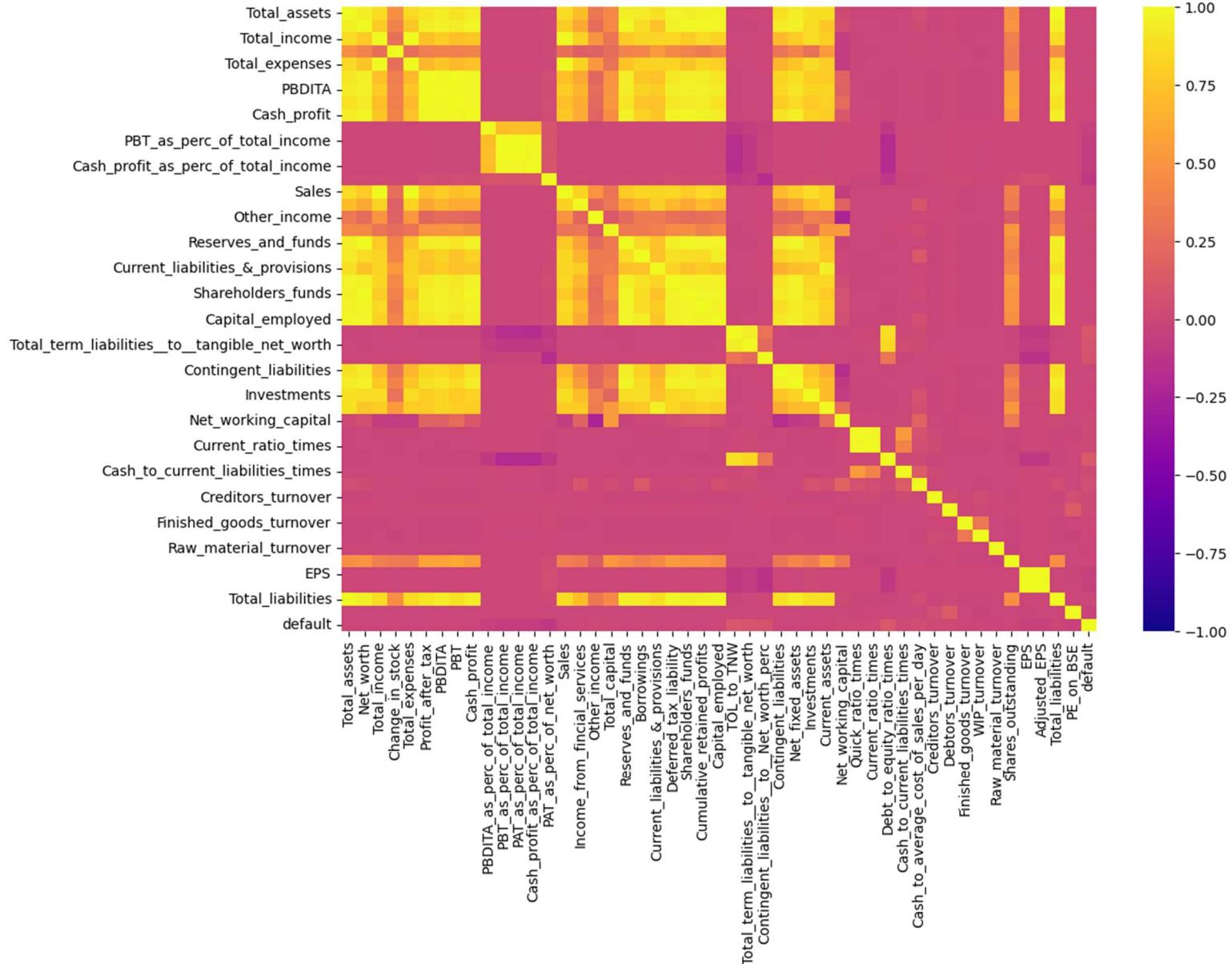


**Inferences:**

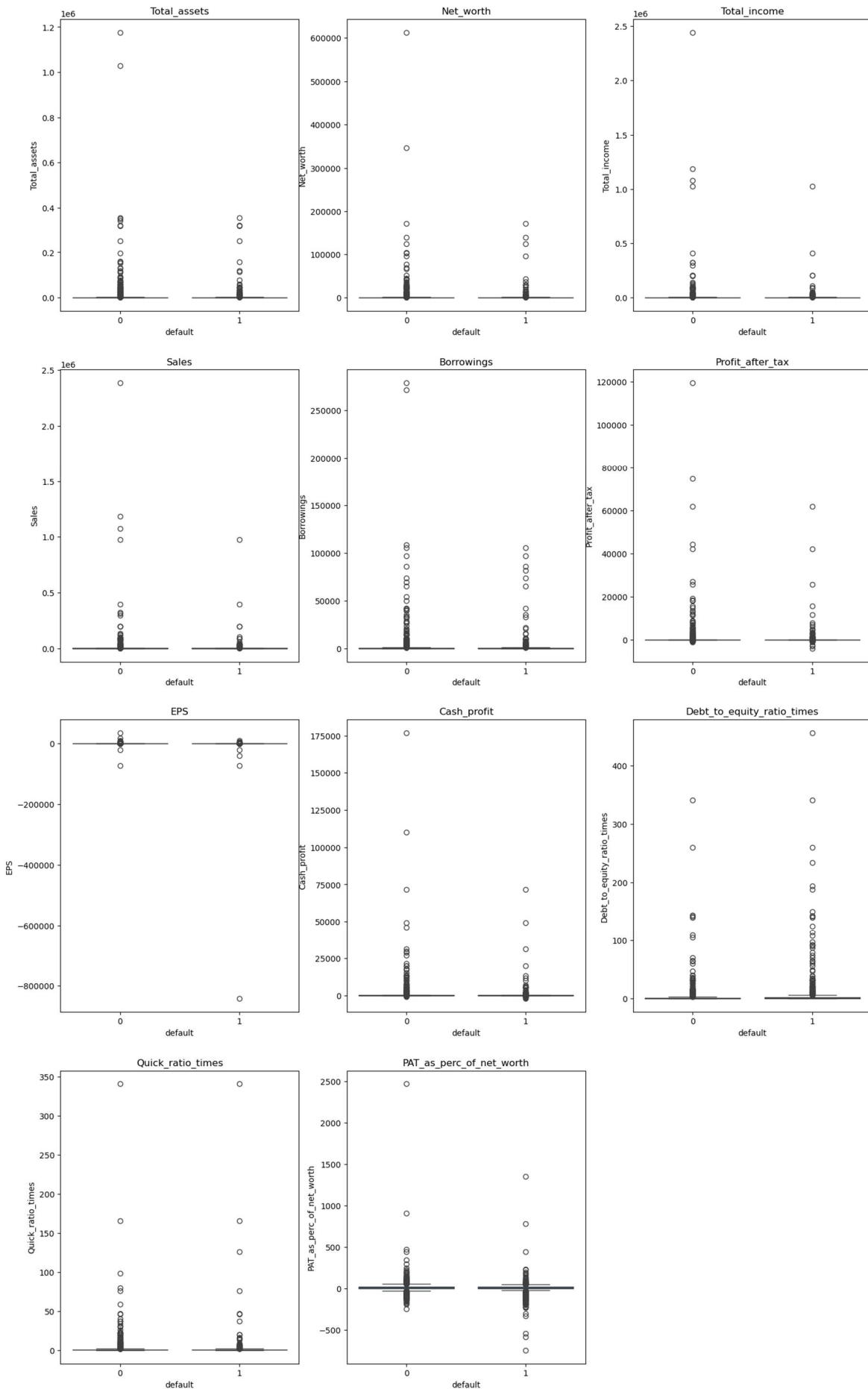
- Class 0 has the majority of the distribution by having 3352 values and Class 1 having 904 values.
- There is a class imbalance in the dataset.
- Majority of the numerical distribution are highly right skewed.
- EPS distribution shows a left skewed distribution.

## 2.2 BIVARIATE ANALYSIS:

**Fig.14 Heatmap**



### Fig.15 Boxplots of variables w.r.t default



## **Inferences:**

- Outliers are present in most of the numerical variables.
- Minimal Outliers are found in cash\_profit, total\_income and profit\_after\_tax.
- In Debt\_to\_equity\_ratio\_times, Higher medians and more outliers are found in defaulters.
- In PAT\_as\_perc\_of\_net\_worth, the plot indicates weaker returns signaling financial inefficiency.

### 3. DATA PREPROCESSING

#### Target Variable Creation:

- The target variable is “default”.
  - Flag 0 : if net\_worth\_next\_year is positive or 0
  - Flag 1 : if net\_worth\_next\_year is negative

#### Missing Value Treatment:

**Fig.16 Null value count**

Total_assets	0
Net_worth	0
Total_income	231
Change_in_stock	550
Total_expenses	165
Profit_after_tax	154
PBDITA	154
PBT	154
Cash_profit	154
PBDITA_as_perc_of_total_income	79
PBT_as_perc_of_total_income	79
PAT_as_perc_of_total_income	79
Cash_profit_as_perc_of_total_income	79
PAT_as_perc_of_net_worth	0
Sales	305
Income_from_fincial_services	1111
Other_income	1556
Total_capital	5
Reserves_and_funds	98
Borrowings	431
Current_liabilities_&_provisions	110
Deferred_tax_liability	1369
Shareholders_funds	0
Cumulative_retained_profits	45
Capital_employed	0
TOL_to_TNW	0
Total_term_liabilities_to_tangible_net_worth	0
Contingent_liabilities_to_Net_worth_perc	0
Contingent_liabilities	1402
Net_fixed_assets	132
Investments	1715
Current_assets	80
Net_working_capital	37
Quick_ratio_times	105
Current_ratio_times	105
Debt_to_equity_ratio_times	0
Cash_to_current_liabilities_times	105
Cash_to_average_cost_of_sales_per_day	100
Creditors_turnover	391
Debtors_turnover	385
Finished_goods_turnover	874
WIP_turnover	764
Raw_material_turnover	428
Shares_outstanding	810
EPS	0
Adjusted_EPS	0
Total_liabilities	0
PE_on_BSE	2627
default	0
	dtype: int64

- Null values are treated using KNN imputer method.
- Columns having more than 30% null values are dropped.

**Fig.17 Null value percentage**

PE_on_BSE	0.67
Investments	0.51
Other_income	0.46
Contingent_liabilities	0.42
Deferred_tax_liability	0.42
Income_from_fincial_services	0.38
Change_in_stock	0.31

### **Outlier Treatment:**

- Outlier values were identified using the IQR method.
- Outlier values are converted into NaN values, which are removed in missing value treatment.
- Treating Outlier values helps in improving data quality and model results.

### **Data Split:**

- Data is split in the ratio of 7:3 for training and testing.

### **Data Scaling:**

- Data is scaled by StandardScaler.

### **Data Resampling:**

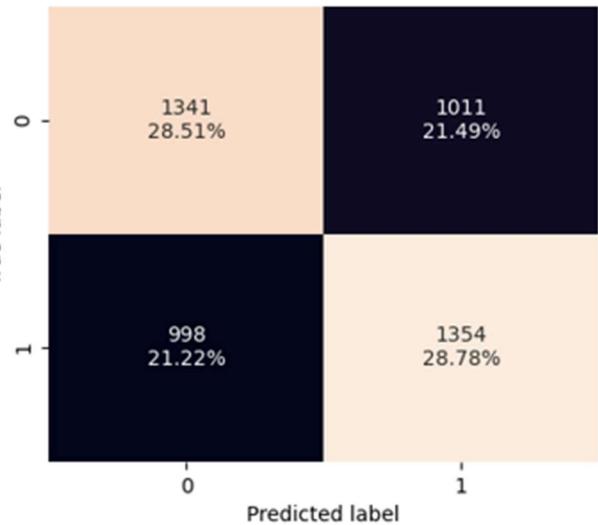
- Class 0 has 3352 samples and Class 1 has 904 samples.
- Data is resampled using Oversampling technique.



**Fig.19 Logit Model Perf 1**

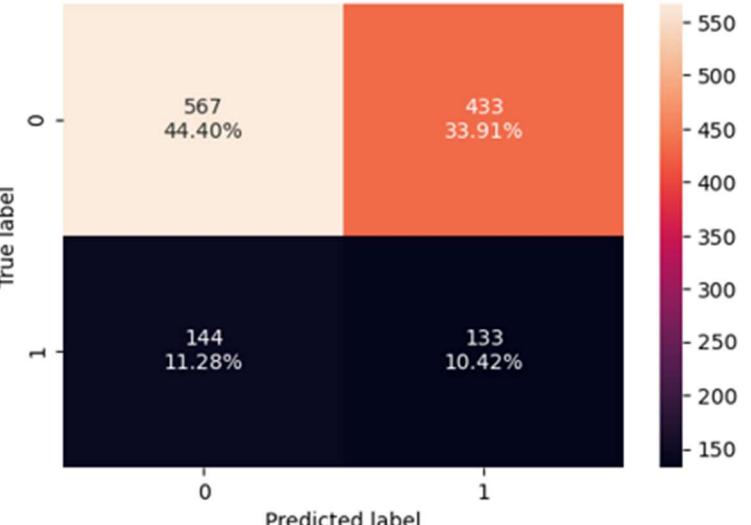
Training Set Performance

	Accuracy	Recall	Precision	F1
0	0.57	0.58	0.57	0.57



Test Set Performance

	Accuracy	Recall	Precision	F1
0	0.55	0.48	0.23	0.32



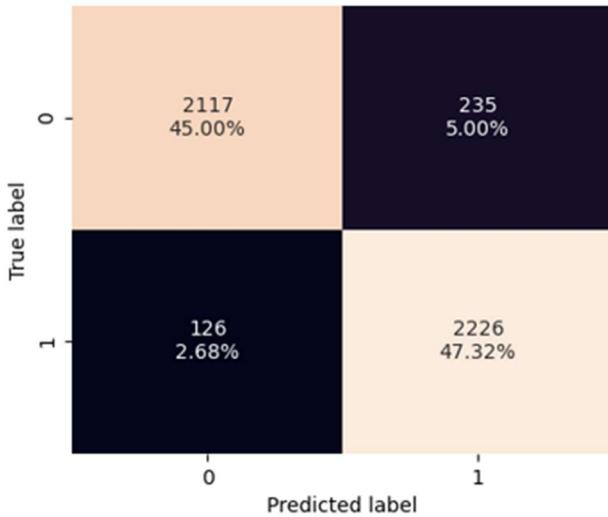
## 4.2 RANDOM FOREST CLASSIFIER

- A simple Random forest classifier model is trained.

**Fig.20 RFC Model Perf 1**

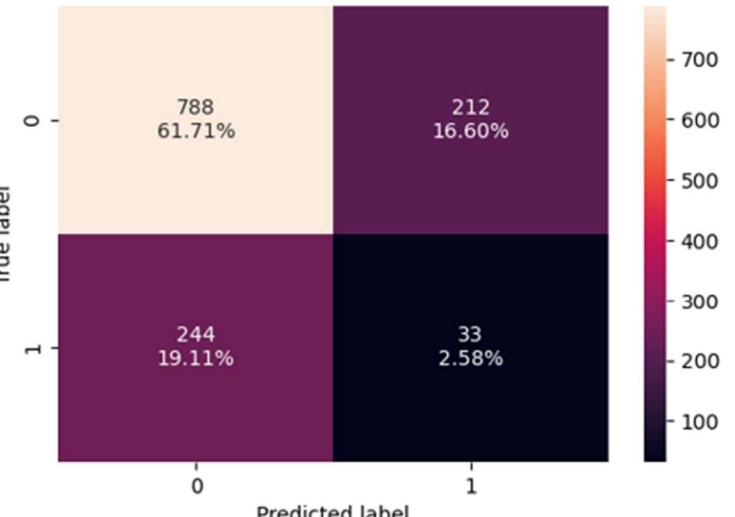
Training Set Performance

	Accuracy	Recall	Precision	F1
0	0.92	0.95	0.90	0.92



Test Set Performance

	Accuracy	Recall	Precision	F1
0	0.64	0.12	0.13	0.13



# 5. MODEL PERFORMANCE IMPROVEMENT

## 5.1 LOGISTIC REGRESSION:

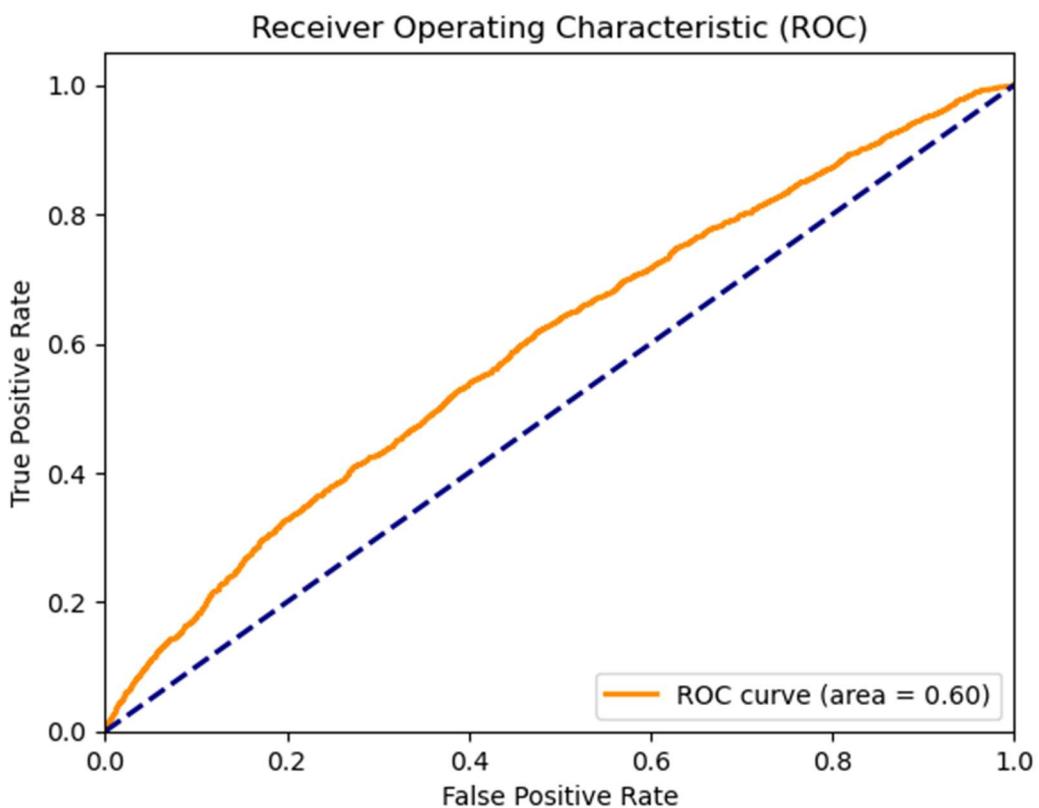
### Treating Multicollinearity:

- There are several columns having VIF value more than 5. Dropping all those columns will lead to data loss. So VIF value threshold is increased to 10.
- Multicollinearity is treated by dropping columns one by one having VIF value more than 10.

### Optimal Threshold value:

- The optimal threshold is obtained from ROC curve.
- Optimal threshold = 0.492

**Fig.21 ROC curve train**





## Model Performance:

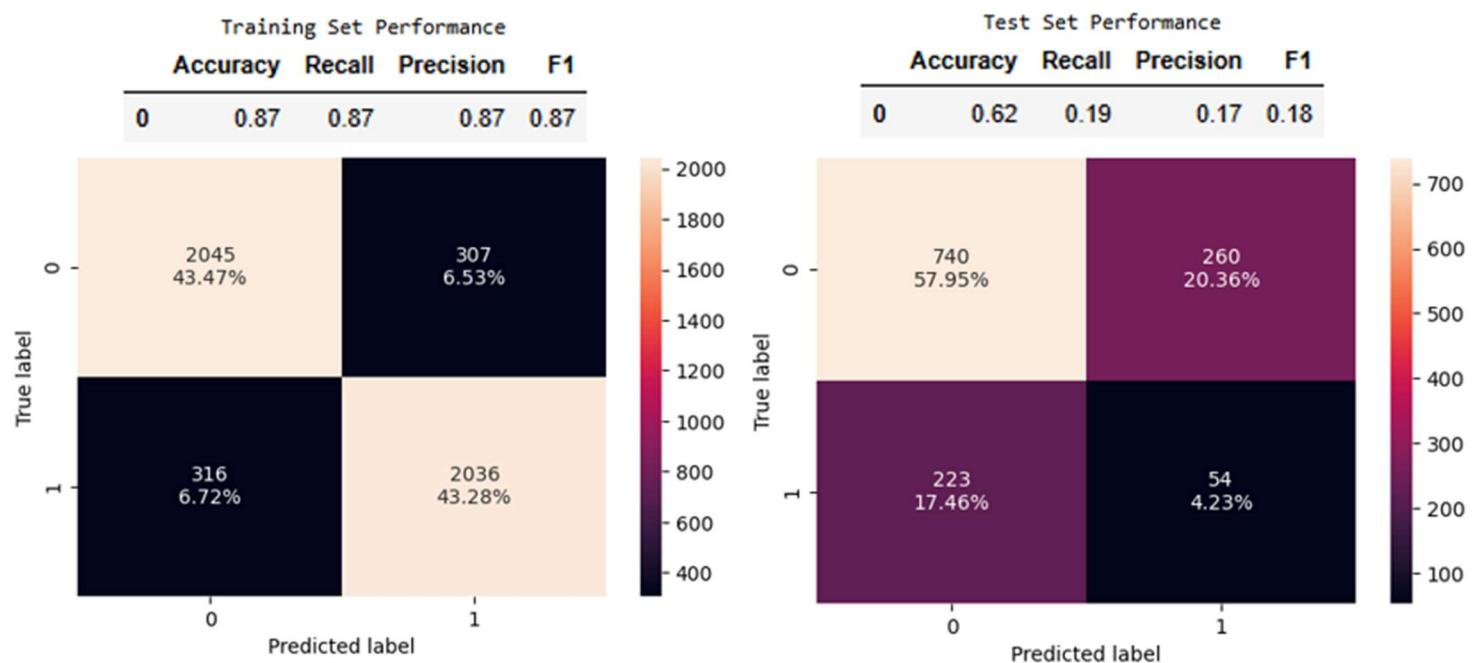
Fig.23 Logit Model Perf 2



## 4.2 RANDOM FOREST CLASSIFIER

- A hyper-parameter tuned Random forest classifier model is used.
- Parameters are selected using RandomizedSearchCV method.
- Parameters: {'n\_estimators': 75,  
'min\_samples\_split': 10,  
'min\_samples\_leaf': 3,  
'max\_depth': 10}

**Fig.24 RFC Model Perf 2**



## 6. MODEL PERFORMANCE COMPARISON AND FINAL MODEL SELECTION

Training Set Performance:

**Fig.25 Training set Perf**

Training performance

	Accuracy	Recall	Precision	F1
Logistic Regression	0.57	0.58	0.57	0.57
Tuned Logistic Regression	0.57	0.62	0.56	0.59
Random Forest	0.92	0.95	0.90	0.92
Tuned Random Forest	0.87	0.87	0.87	0.87

Test Set Performance:

**Fig.26 Test set Perf**

Testing performance:

	Accuracy	Recall	Precision	F1
Logistic Regression	0.55	0.48	0.23	0.32
Tuned Logistic Regression	0.54	0.55	0.25	0.34
Random Forest	0.64	0.12	0.13	0.13
Tuned Random Forest	0.62	0.19	0.17	0.18

## **FINAL MODEL SELECTION:**

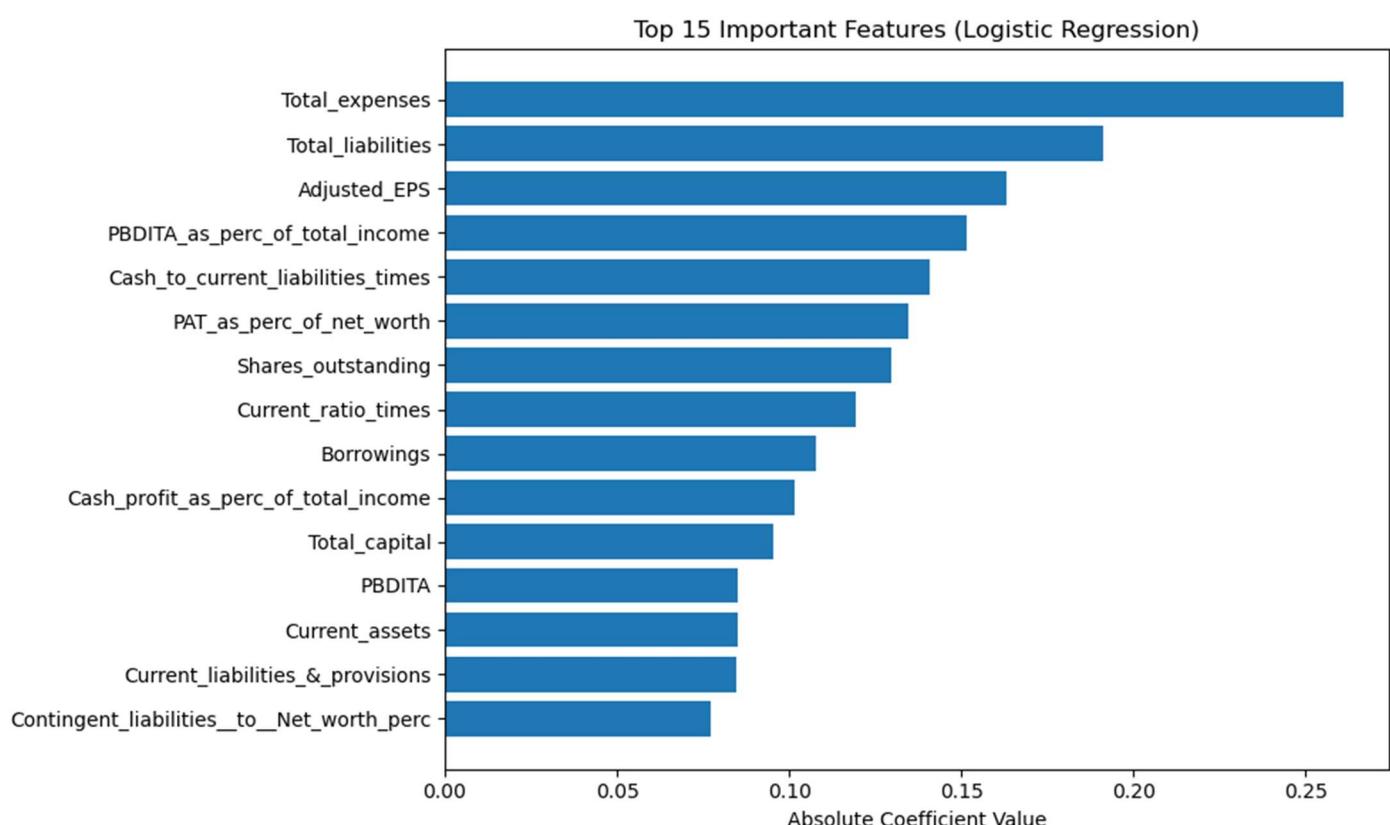
- Even though the Random Forest models performs very well in training set, it still underperforms in test set.
- The recall score of Random forest models are very low in test dataset.
- Logistic Regression Models performs decently in both training and testing dataset.
- Even though, the Random Forest model have the highest accuracy score on test dataset, it underperforms in Recall score which is our choice of metric.
- The best model for predicting the defaulters is Tuned Logistic Regression as it has the highest recall score (0.55) and decent accuracy score.

## 7. ACTIONABLE INSIGHTS & RECOMMENDATIONS

### Inferences:

Important Features:

**Fig.27 Top 15 important features**



- Total\_expenses is the top feature which implies that companies having high expenses are most likely to default.
- Total\_liabilities have a slight positive correlation with default risk which indicates that companies having high debt and liabilities are more vulnerable to default.

- Adjusted EPS (Earnings Per Share) is also a key-indicator which indicates that low EPS companies are most likely to default.

### **Business Recommendations:**

- An early-warning system or approach can be made with the top features which enables to identify clients at risk of default.
- To increase profit margins, concentrate on companies having steady revenue growth and effective cost control.
- Regular maintenance of debt-to-equity ratios will help the company.
- Excessive borrowing should be avoided and limits should be created based on their company performance.
- Clear contracts should be made which defines terms and penalties for defaulters.
- Total Expenses and Total liabilities of a company should be regularly monitored with a real-time company data.
- Companies should be segmented based on their company asset values and loans should be provided with respect to their levels.