FINANCE & RISK ANALYTICS PROJECT

BUSINESS REPORT

BY

G S JAIGURURAM

TABLE OF CONTENTS

TABLE OF FIGURES

Part A

Problem A.1 Problem Statement

Problem A.1.1 Context

In the realm of modern finance, businesses encounter the perpetual challenge of managing debt obligations effectively to maintain a favorable 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.

Dataset: Company_FRA.csv

Problem A.1.2 Objective

A group of venture capitalists want to develop a Financial Health Assessment Tool. With the help of the tool, it endeavors 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 analyze 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:

- 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.
- Credit Risk Evaluation: Evaluate credit risk exposure by analyzing 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 analyze 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.

Problem A.1.3 Problem Definition

Develop a machine learning model to predict whether a company will be classified as a defaulter in terms of net worth in the coming year, based on historical financial data. This tool will assist venture capitalists in assessing companies' financial health and credit risk.

Problem A.2 Data Overview

Problem A.2.1 Dataset

Dataset Contains single sheets namely ['Company_FRA']

- Dataset has 4256 Rows (Not Including headers) & 51 Columns.

Problem A.2.2 Data Dictionary

The data consists of financial metrics from the balance sheets of different companies. The detailed data dictionary is given below.

Columns	Description
Networth 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

Problem A.2.3 Data

Given dataset was read using the jupyter notebook.

Shape

- The dataset contains 4,256 rows and 51 columns.

Data Types

- Out of 51 Columns, 50 Columns are float64 type data variables.
- One column is of data type int64.

Statistical Summary

	Num	Networth Next Year	Total assets	Net worth	Total income	Change in stock	Total expenses	Profit after tax	PBDITA	PBT	...	Debto turnov
count	4256.000000	4256.000000	4.256000e+03	4256.000000	4.025000e+03	3706.000000	4.091000e+03	4102.000000	4102.000000	4102.000000	...	3871.0000
mean	2128.500000	1344.740883	3.573617e+03	1351.949601	4.688190e+03	43.702482	4.356301e+03	295.050585	605.940639	410.259044	...	17.9290
std	1228.745702	15936.743168	3.007444e+04	12961.311651	5.391895e+04	436.915048	5.139809e+04	3079.902071	5646.230633	4217.415307	...	90.1644
min	1.000000	-74265.600000	1.000000e-01	0.000000	0.000000e+00	-3029.400000	-1.000000e-01	-3908.300000	-440.700000	-3894.800000	...	0.0000
25%	1064.750000	3.975000	9.130000e+01	31.475000	1.071000e+02	-1.800000	9.680000e+01	0.500000	6.925000	0.800000	...	3.8100
50%	2128.500000	72.100000	3.155000e+02	104.800000	4.551000e+02	1.600000	4.268000e+02	9.000000	36.900000	12.600000	...	6.4700
75%	3192.250000	330.825000	1.120800e+03	389.850000	1.485000e+03	18.400000	1.395700e+03	53.300000	158.700000	74.175000	...	11.8500
max	4256.000000	805773.400000	1.176509e+06	613151.600000	2.442828e+06	14185.500000	2.366035e+06	119439.100000	208576.500000	145292.600000	...	3135.2000

8 rows × 51 columns

FIGURE 1 : STATISTICAL SUMMARY

Insights from summary:

1. Overall Financial Health:
    - The average Net Worth Next Year (1344.74) shows a positive expectation compared to the current Net Worth (1351.95), suggesting that customers anticipate an increase in their financial status.
    - However, the high standard deviation in Net Worth Next Year (15936.74) indicates significant variability among customers, with some experiencing substantial financial challenges, as evidenced by the minimum value of -74265.6.
2. Asset Management:
    - The average Total Assets (3573.62) compared to the average Total Liabilities (3573.62) indicates that customers are managing their assets and liabilities at a balanced level. However, the maximum total assets of over 1.17 million suggests that while many customers are doing well, there are outliers with significantly higher asset values.
    - The Current Ratio (2.26) indicates good short-term financial health, as customers have more than twice their current liabilities covered by current assets.
3. Profitability Analysis:
    - The average Profit After Tax (295.05) is relatively low compared to other profitability metrics like PBDITA (605.94). This suggests that while customers may generate income, high expenses or tax burdens significantly affect net profitability.
    - Negative values in profitability ratios such as PAT as % of Total Income (-20.03) indicate that a considerable portion of income is consumed by expenses or taxes.

4. Expense Management:

    - The average Total Expenses (4356.3) is close to the average Total Income (4688.19), suggesting that many customers operate near breakeven points, which can be a concern for long-term sustainability.

    - The negative cash profit percentage (-9.02) indicates cash flow management issues for some customers, which could lead to liquidity problems.

5. Investment and Returns:

    - The negative values in metrics like EPS (-196.22) and Adjusted EPS (-197.53) suggest that many customers may not be generating sufficient earnings relative to their shares outstanding, potentially deterring investment interest.

    - High values in metrics such as the Debt to Equity Ratio (2.87) indicate that many customers are heavily reliant on debt financing, which can pose risks if income levels do not improve.

6. Turnover Ratios:

    - Turnover ratios such as Creditors Turnover (16.81) and Debtors Turnover (17.93) indicate how efficiently customers manage their credit purchases and sales receivables respectively.

    - A high turnover ratio suggests effective management of credit terms and collections but should be monitored to avoid cash flow issues.

7. Contingent Liabilities Concerns:

    - The presence of contingent liabilities (average of 948.55 with a standard deviation of 12056.74) indicates potential financial risks that could impact future net worth and cash flows if these liabilities materialize.

8. Working Capital Analysis:

    - The average Net Working Capital (162.87) being relatively low suggests that some customers may face challenges in meeting short-term obligations, especially given the minimum value of -63839.

9. Missing Values:
    - There 37 columns with missing values. 14 columns with no missing values are “Num”,”Networth Next Year”,”Total assets”,”Net worth”,”PAT as % of net worth”,”Shareholders funds”,”Capital employed”,”TOL/TNW”,”Total term liabilities / tangible net worth”,”Contingent liabilities / Net worth (%)”,”Debt to equity ratio (times)”,” EPS”,” Adjusted EPS”,” Total liabilities”.

10. Duplicate Values: There are no duplicate value present in dataset.

Problem A.3 Exploratory Analysis & Inferences

Performed univariate & multivariate analysis and list out the insights from them.

Problem A.3.1 Univariate analysis

For We are going to focus in certain columns only, reason for it is stated below:

- I chose columns that are likely to have a significant impact on financial health or are directly related to the problem statement (e.g., Networth Next Year, Total assets, Net worth, and Debt to equity ratio).
- The selected columns represent different financial dimensions: profitability (e.g., Profit after tax), liquidity (e.g., Borrowings), and operational performance (e.g., Sales).

Selected Columns are: "Networth Next Year", "Total assets", "Net worth", "Total income", "Total expenses", "Profit after tax", "PBDITA", "Sales", "Borrowings", "Debt to equity ratio (times)".

Networth Next Year

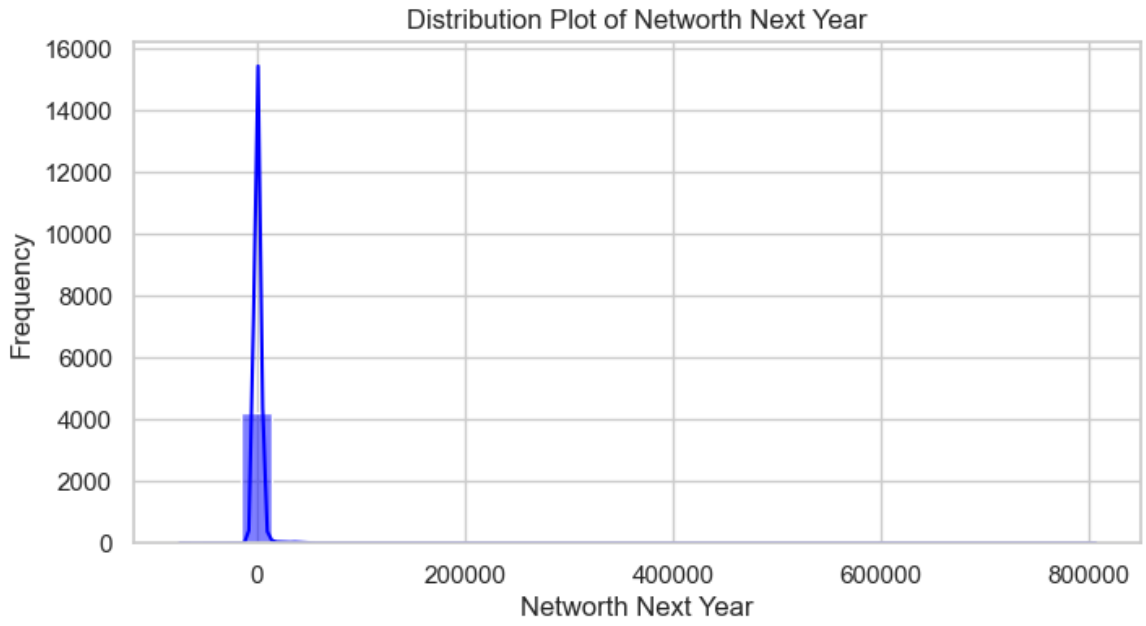

FIGURE 2 : NETWORTH NEXT YEAR

- The distribution is highly skewed (Right), with a majority of companies concentrated at lower net worth values.
- Outliers suggest a few companies with exceptionally high net worth.

Total Assets

FIGURE 3 : TOTAL ASSESTS

- Right-skewed distribution, with many companies holding lower asset values.
- A small proportion holds disproportionately high assets.

Net Worth:

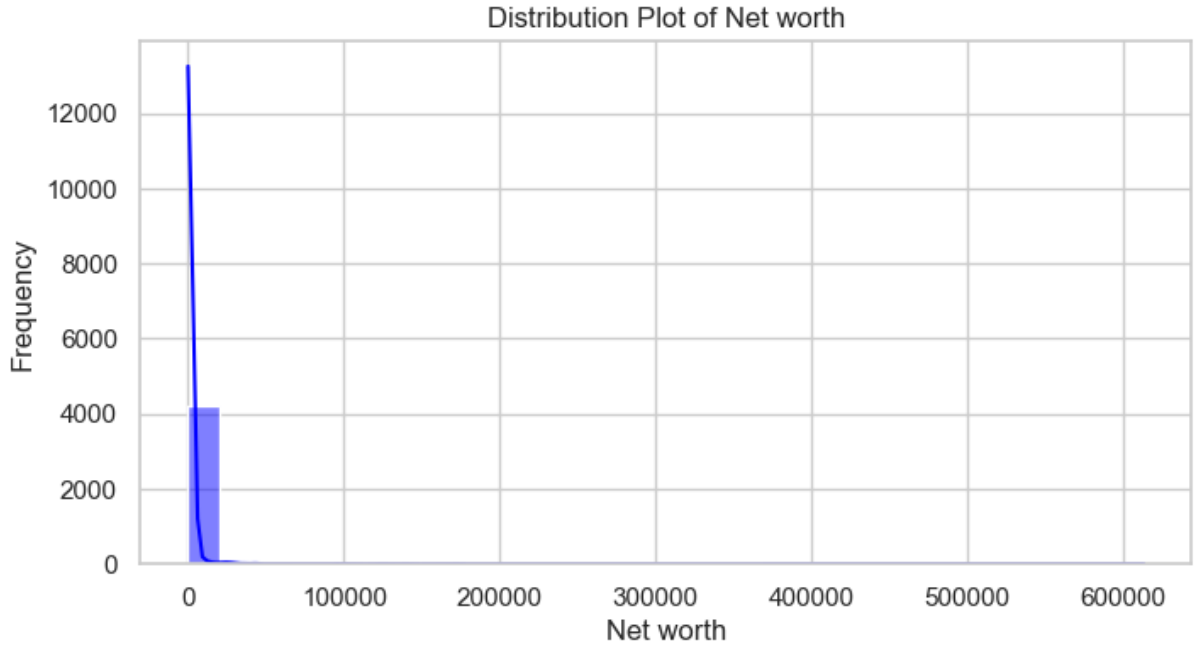

FIGURE 4 : NET WORTH

- The distribution is right skewed, indicating financial disparities among companies.
- Outliers suggest a few companies with exceptionally high net worth.

Total Income

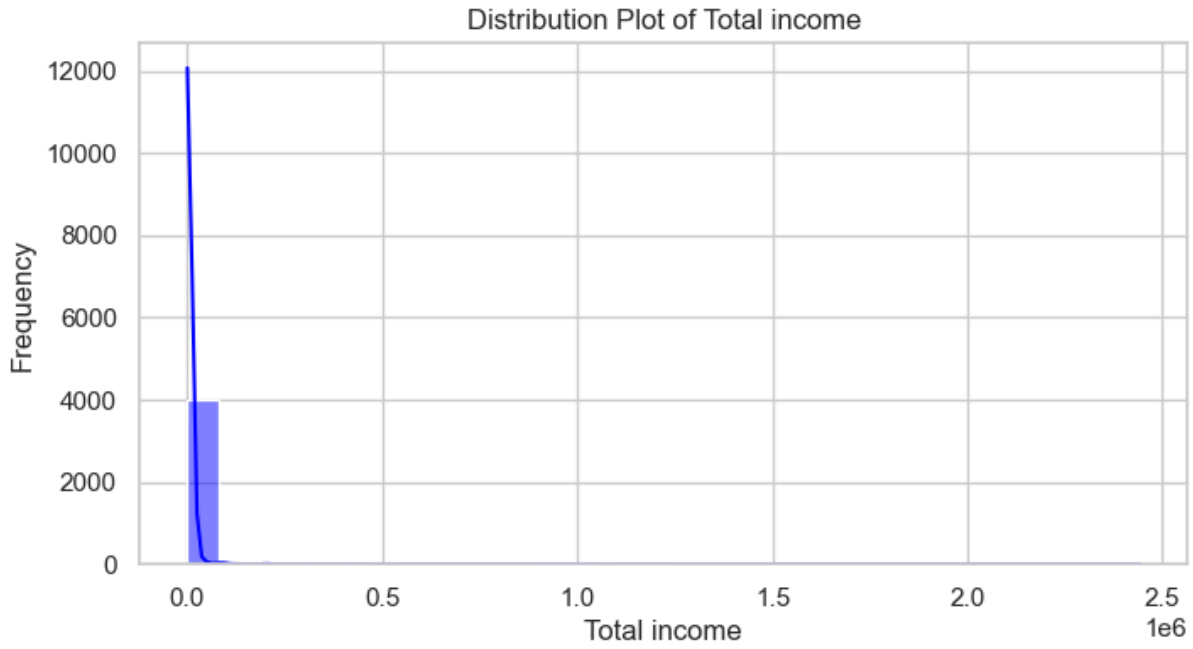

FIGURE 5 : TOTAL INCOME

- Skewed with high variability, reflecting varying revenue capacities.
- Outliers suggest a few companies with exceptionally high total revenue.

Total Expenses

- Plot is similar to total income plot
- Skewed with high variability, reflecting varying expenses.
- Outliers suggest a few companies with exceptionally high expenses amount.

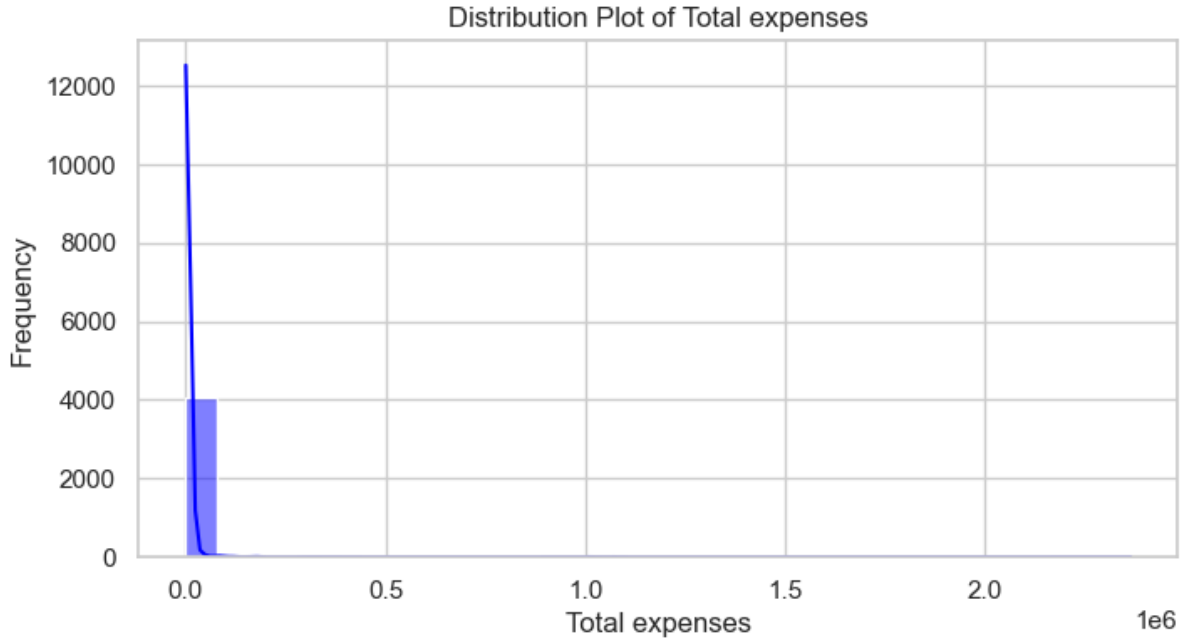

FIGURE 6 : TOTAL EXPENSES

Profit After Tax (PAT)

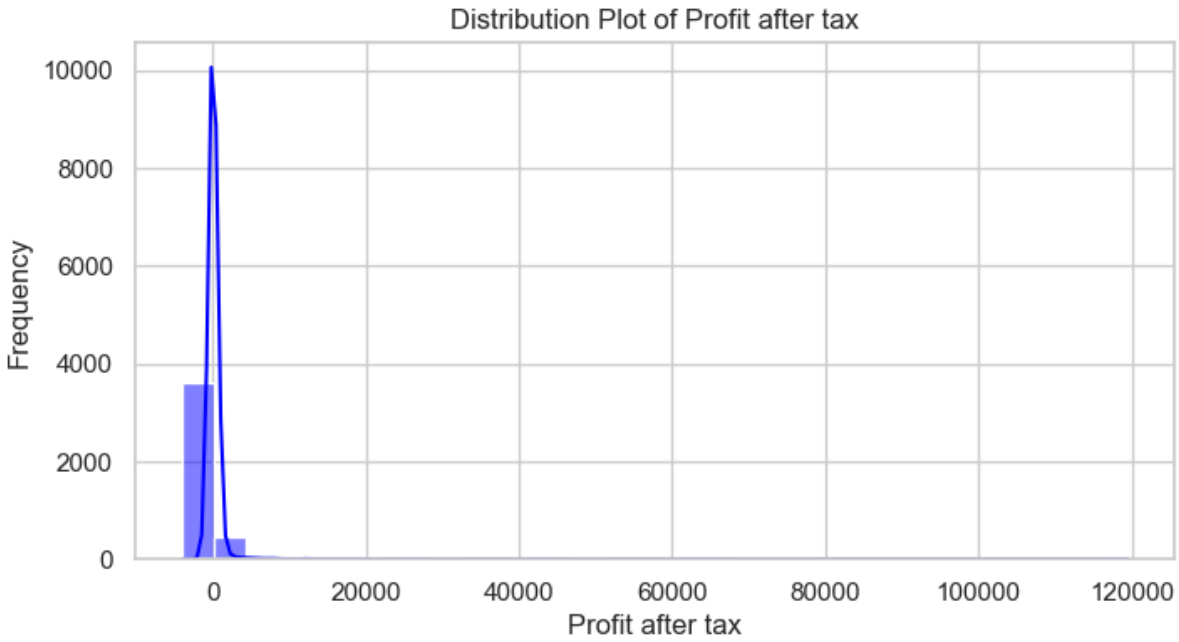

FIGURE 7 : PROFIT AFTER TAX

- A sharp peak near zero, suggesting many companies have marginal profits.
- Presence of negative values indicates losses for some entities.

Profit before depreciation, income tax, and amortization (PBDITA)

- A range of values, but fewer extreme outliers than PAT.
- Right-skewed distribution, with many companies holding lower PBDITA values.

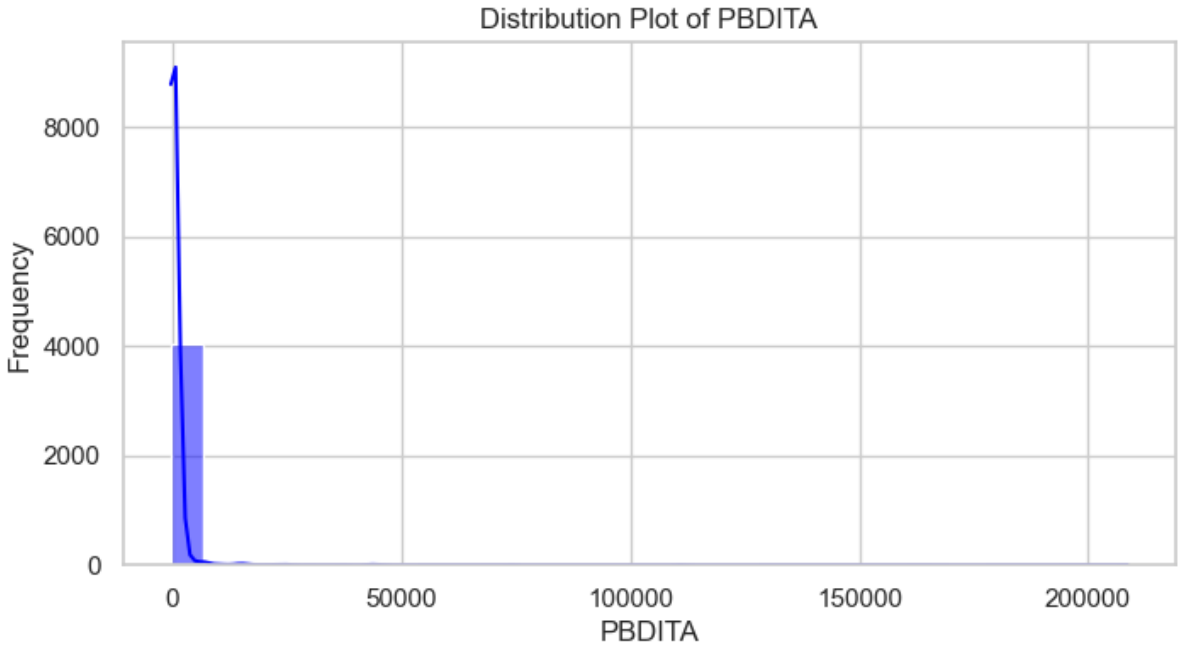

FIGURE 8 : PBDITA

Sales

- Concentration in lower ranges, with a gradual tail reflecting larger sales volumes.

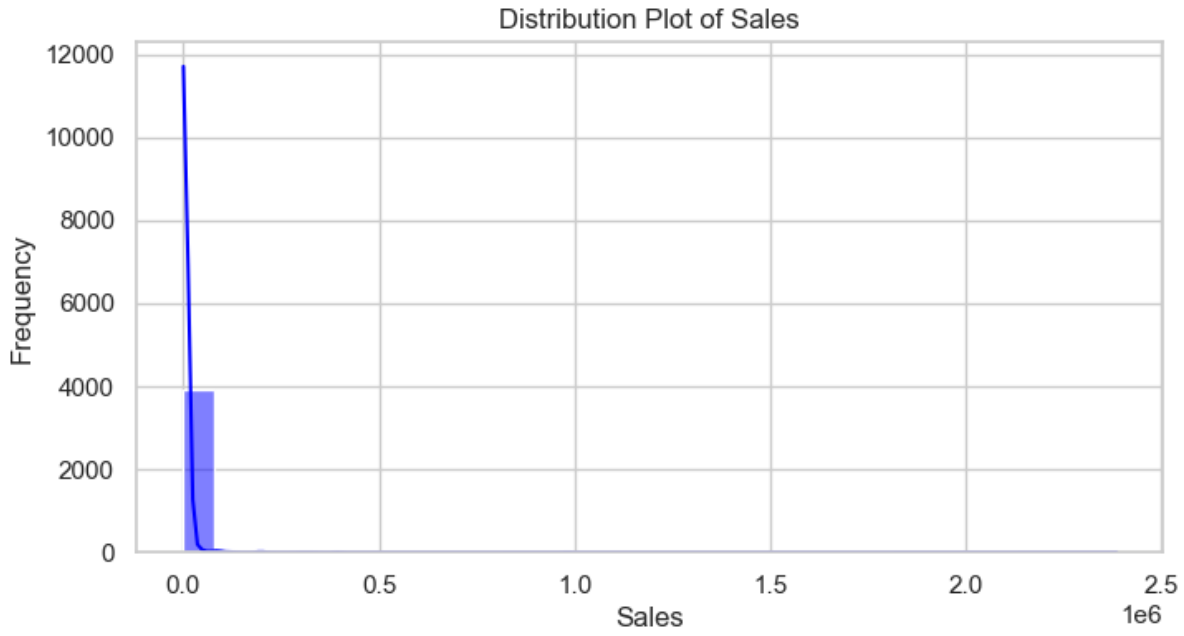

FIGURE 9 : SALES

Borrowings

- A notable skew, implying reliance on debt varies widely across firms.

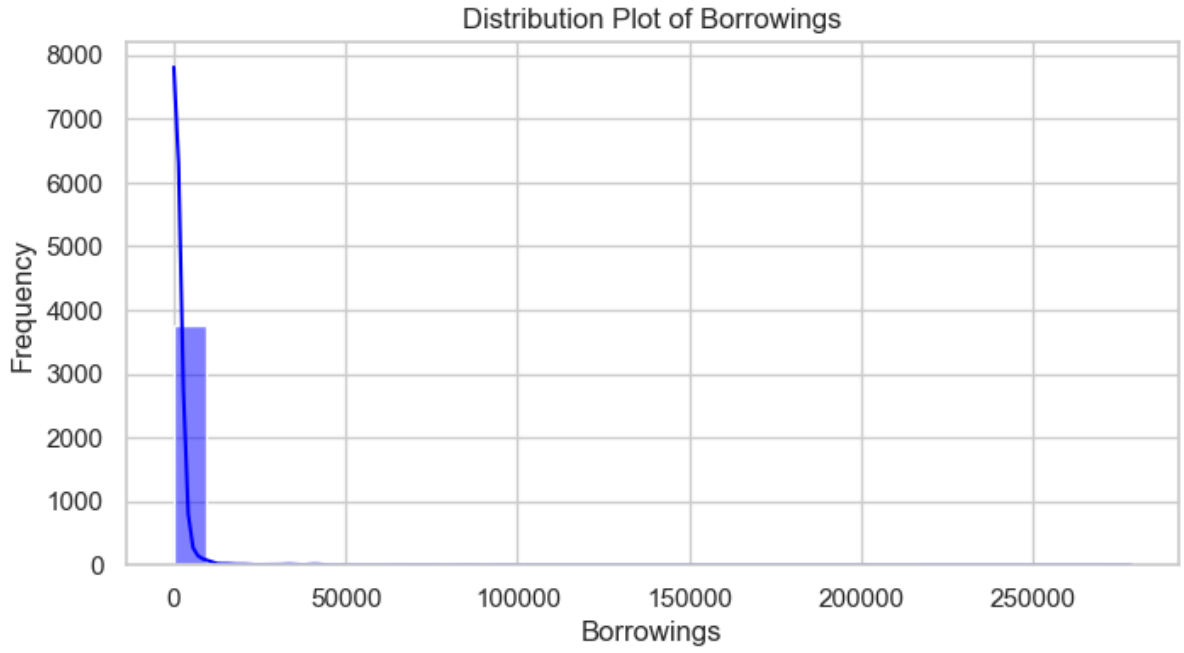

FIGURE 10 : BOORROWINGS

Debt to Equity Ratio

- Most firms have ratios below 1, suggesting moderate leverage. Extreme values hint at potential financial stress.

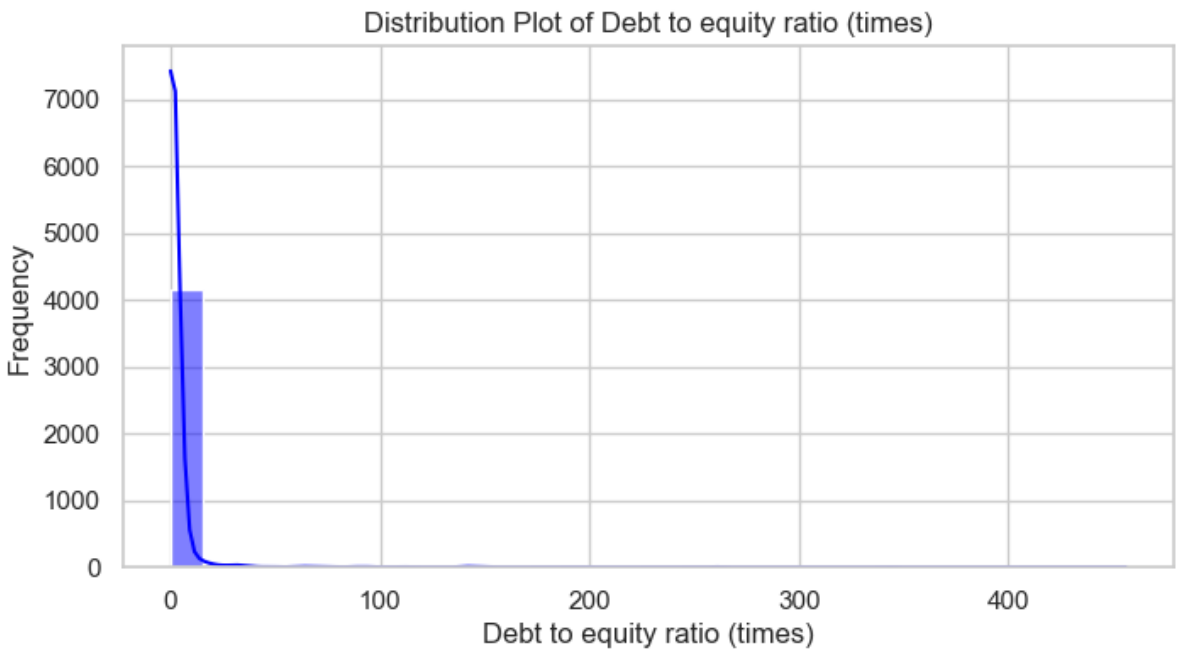

FIGURE 11 : DEBT TO EQUITY RATIO (TIMES)

Problem A.3.2 Multivariate analysis

Heatmap

Key Observations:

1. High Correlation Groups:

- Net Worth Next Year:
  - Strongly correlated with variables like Total Assets, Net Worth (Current Year), Reserves and Funds, and Shareholders' Funds. These variables are crucial indicators of financial stability and predict future net worth.
- Profitability Metrics:
  - Variables like Profit After Tax (PAT), PBDITA, and PBT exhibit strong intercorrelation. Additionally, these profitability metrics correlate well with Sales, indicating that revenue generation is a critical driver of profitability.
- Liquidity Indicators:
  - Current Ratio and Quick Ratio show moderate correlations with Current Assets and Net Working Capital, suggesting the ability to meet short-term liabilities is tied to working capital management.

2. Debt and Risk Metrics:

- Debt-to-Equity Ratio shows:
  - Negative correlations with Net Worth and Shareholders' Funds. Companies with higher equity financing tend to have lower leverage ratios.
  - Moderate positive correlation with Borrowings, indicating that companies with higher debt financing tend to have higher debt-to-equity ratios.

- TOL/TNW (Total Liabilities to Tangible Net Worth) is positively associated with contingent liabilities and borrowing metrics, emphasizing a company's leveraged financial position.

3. Sales and Income Indicators:

    - Sales and Income from Financial Services are positively correlated with profitability indicators (e.g., PBDITA, PAT). This reflects that operational revenue streams are directly linked to overall financial health.

4. Negative Correlations:

    - Contingent Liabilities / Net Worth (%) is negatively correlated with net worth and reserves, indicating that companies with higher contingent liabilities tend to have weaker equity positions.

5. Stock and Market Metrics:

    - PE on BSE and EPS (Adjusted and Unadjusted):
        - Positive correlations with profitability metrics suggest that more profitable companies tend to have stronger market valuations.

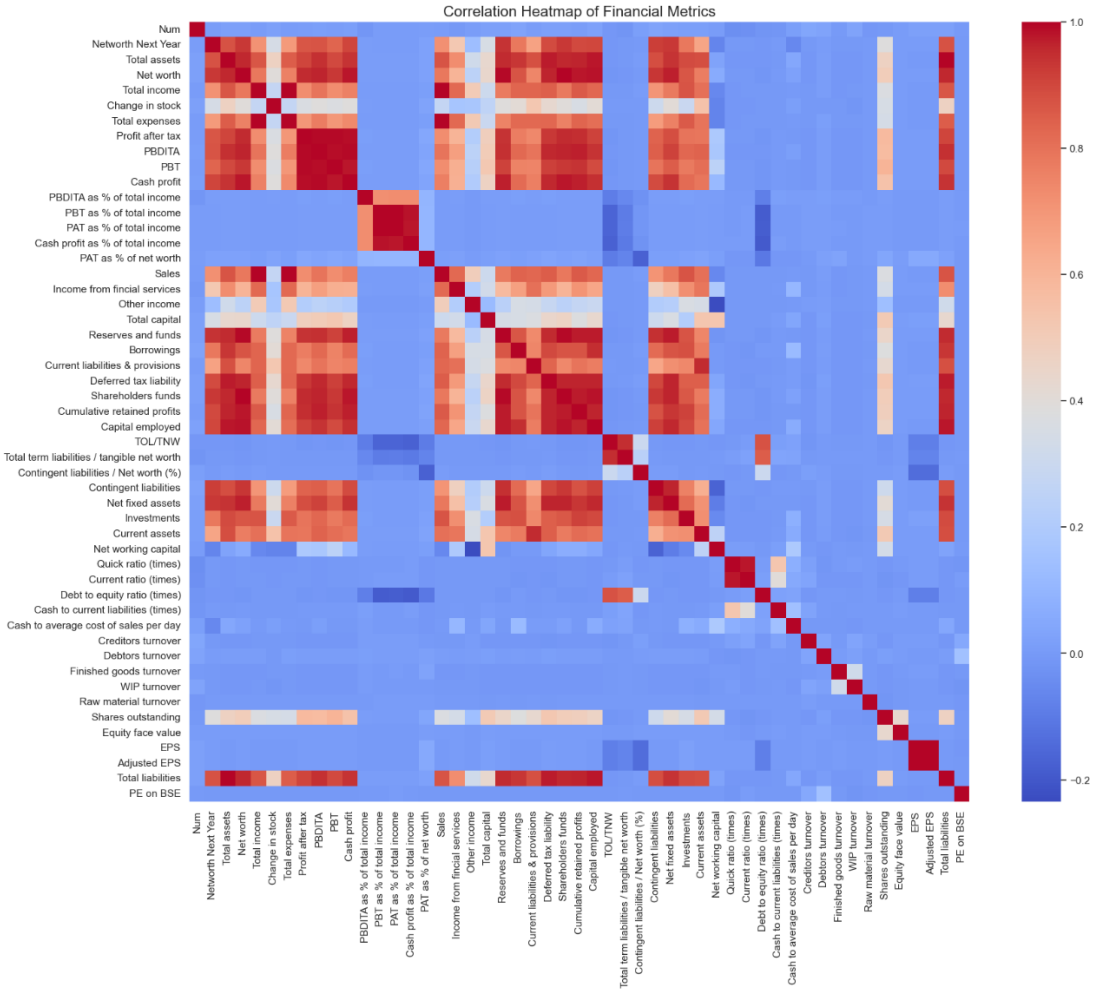

FIGURE 12 : HEATMAP

Pair Plot

As we have large number of columns, so we are going to plot a scatter plot between selected pairs.

Selected pairs & reasons are:

1. ('Net worth', 'Networth Next Year')
    a. Helps evaluate if there's a strong linear or non-linear relationship between these two critical financial metrics, as this is fundamental for predicting financial health.
2. ('Sales', 'Profit after tax')
    a. Sales are typically a key driver of profitability (Profit After Tax - PAT). Analyzing this pair can reveal how well a company converts sales into profits.
    b. Insights here can identify trends like high sales with low profits (inefficiency) or high correlation (efficiency).
3. ('Borrowings', 'Net worth')
    a. Borrowings represent a company's debt obligations, while net worth reflects its equity value. Essential for identifying financial risk.
4. ('Debt to equity ratio (times)', 'Net worth')
    a. Pairing it with net worth can show whether higher leverage (debt) correlates with higher or lower equity values. Important for understanding balance sheet risk.
5. ('Current liabilities & provisions', 'Networth Next Year')
    a. Current liabilities impact short-term liquidity and future solvency. Exploring this with next year's net worth helps determine if high liabilities today negatively affect future equity.
6. ('Sales', 'PBDITA')
    a. Analyzing the relationship between sales and PBDITA shows how operational performance scales with sales revenue.

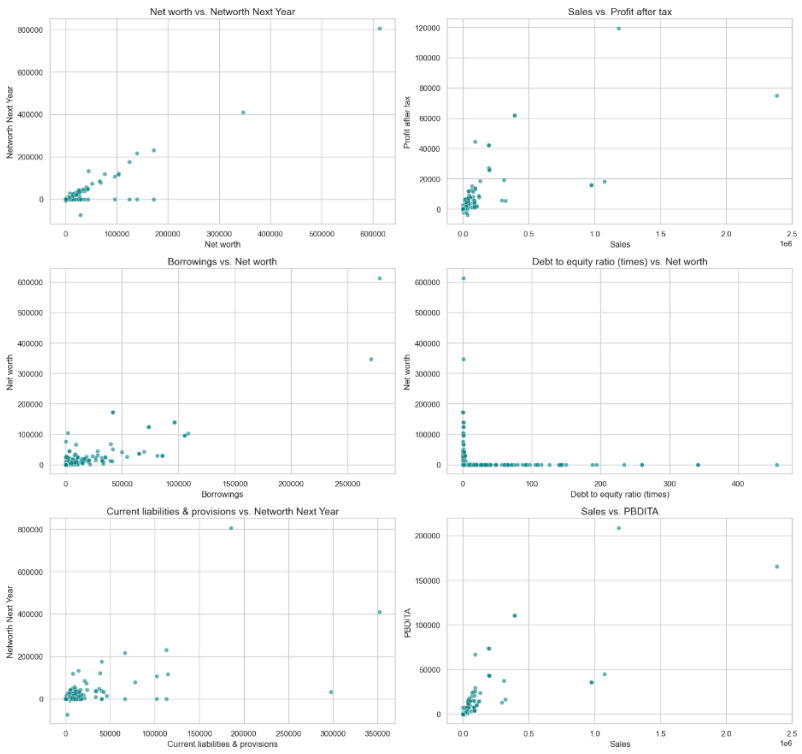

FIGURE 13 : PAIR PLOT

Key Observation:

1. Net worth vs. Networth Next Year

- There is a noticeable positive relationship, indicating that higher net worth this year often corresponds to a higher net worth next year.
- However, some outliers suggest that certain companies may face unforeseen circumstances impacting their future net worth.

2. Sales vs. Profit after tax

- The relationship shows an upward trend, indicating that companies with higher sales tend to have higher profits after tax (PAT).
- There are instances where companies achieve significant sales but with relatively low PAT, possibly due to high operating or fixed costs.

3. Borrowings vs. Net worth

- Borrowings appear to have a weak or non-linear relationship with net worth.
- Some companies with high borrowings show relatively low net worth, raising concerns about over-leveraging.

4. Debt to equity ratio (times) vs. Net worth

- A very high debt-to-equity ratio does not necessarily correlate with higher net worth.
- Companies with extreme leverage are clustered around low net worth values, which may indicate financial instability.

5. Current liabilities & provisions vs. Networth Next Year

- A moderate positive relationship exists, suggesting that companies with higher current liabilities may also project higher net worth in the next year.
- However, this may vary depending on how effectively liabilities are managed or converted into productive assets.

6. Sales vs. PBDITA

- A clear positive correlation indicates that higher sales lead to higher operational profits (PBDITA).
- Deviations may highlight inefficiencies in cost management or operational processes.

Problem A.4 Data Pre-processing

Before detecting outlier, we need to fix the missing values, below are variable that contains more then 30% missing values:

- Other income: 36.56%
- Deferred tax liability: 32.17%
- Contingent liabilities: 32.94%

- Investments: 40.30%
- PE on BSE: 61.72%

We will drop these columns and impute the missing values using the KNN imputer that we replace the missing values based on the average neighbor's value.

Problem A.4.1 Outlier Detection

Outliers can significantly affect model performance. We'll use the Interquartile Range (IQR) to detect outliers and cap/floor extreme values.

Problem A.4.2 Encode the data

Encoding of data was no necessary as in dataset there are no categorical variable present

Problem A.4.3 Target variable creation

The target variable (default) will be binary:

- if Networth Next Year <= 0, then default will be '1' or '0' otherwise.

As Target column was created, lets drop few columns that we wont need for further analysis, columns like 'Num' & 'Networth Next Year'.

'Num' is used as unique identifier its does give mush of values to our analysis.

'Networth Next Year' is used for Creating the Target Variable so it wont be need in analysis.

Problem A.4.4 Data split

Split the data into train and test sets (e.g., 80% train, 20% test).

Shape of train and test date: (3404, 44) (852, 44) (3404, 1) (852, 1)

Problem A.4.5 Scale the data

Scale the features using StandardScaler to standardize or normalize the data.

Problem A.5 Model Building

To build a predictive model using Logistic Regression and Random Forest, we will follow a structured approach that includes selecting appropriate evaluation metrics, building the models, and checking their performance across different metrics

Problem A.5.1 Metrics of Choice

Choosing the right evaluation metrics depends on the problem at hand. Since the task is to predict whether a company's net worth next year will be negative (default) or positive, this is a binary classification problem.

Key Evaluation Metrics:

- **Accuracy:** Measures the percentage of correct predictions. However, it may not be ideal for imbalanced datasets.

- **Precision (Positive Predictive Value):** Proportion of correctly predicted defaults out of all predicted defaults.
- Recall (Sensitivity): Proportion of correctly predicted defaults out of all actual defaults.
- F1-Score: Harmonic mean of Precision and Recall.
- AUC-ROC Curve: Evaluates the model's ability to distinguish between classes.

For this problem, Recall and F1-Score might be more critical since missing a default prediction can have significant financial consequences.

Problem A.5.2 Model (Logistic Regression, Random Forest)

We Build a logistic regression & random forest model using the sklearn library.

Problem A.5.3 Model performance check across different metrics

Confusion Matrix Comparison
Logistic Regression:

- True Negatives (Non-default correctly classified): 661
- False Positives (Non-default misclassified as Default): 10
- False Negatives (Default misclassified as Non-default): 167
- True Positives (Default correctly classified): 14

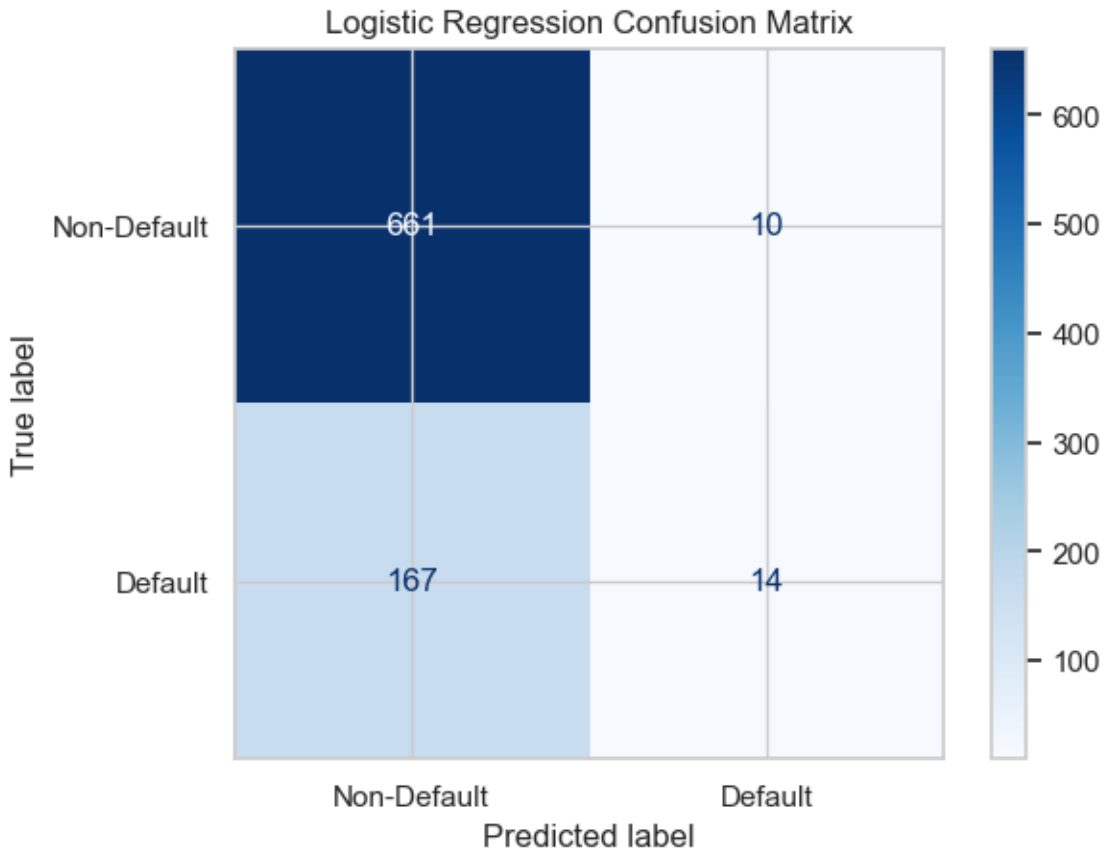

FIGURE 14: CONFUSION MATRIX - LR

Random Forest:

- True Negatives: 568
- False Positives: 103
- False Negatives: 161
- True Positives: 20

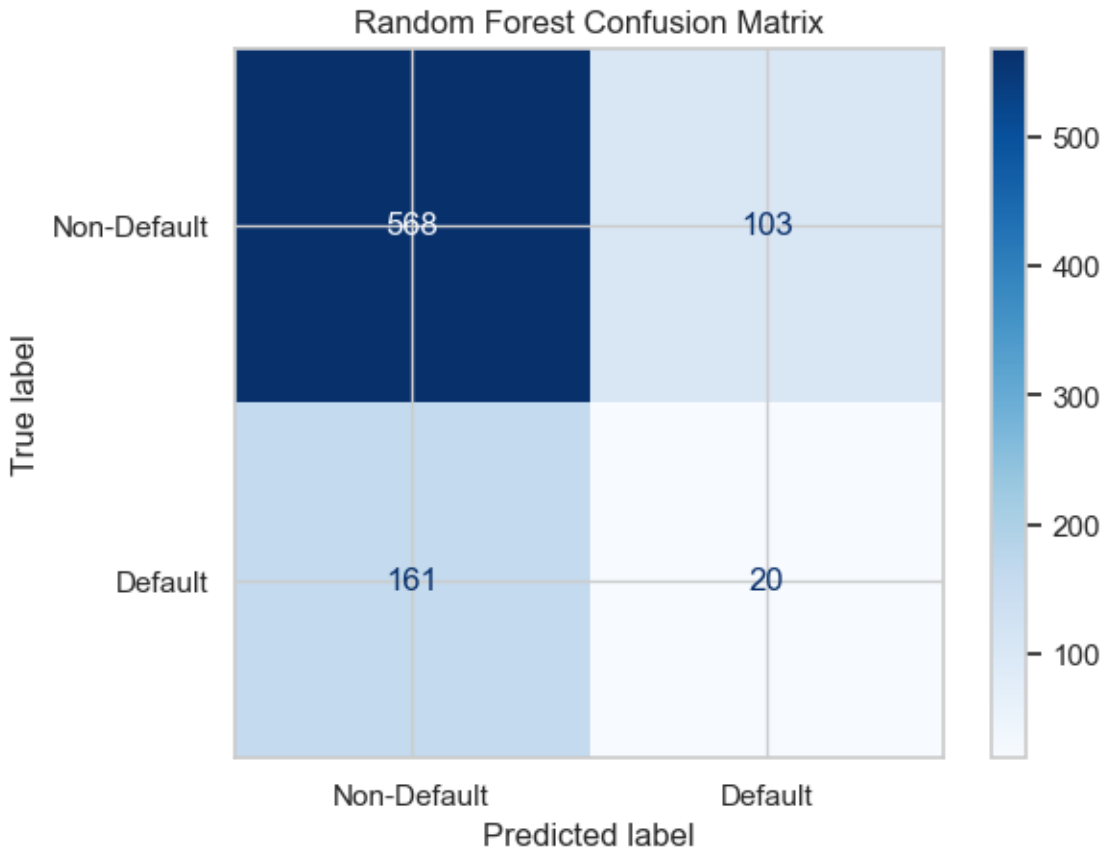

FIGURE 15: CONFUSION MATRIX - RF

Classification Metrics Comparison

Metric	Logistic Regression	Random Forest
Precision (Class 0)	0.80	0.78
Precision (Class 1)	0.58	0.16
Recall (Class 0)	0.99	0.85
Recall (Class 1)	0.08	0.11
F1-Score (Class 0)	0.88	0.81

Metric	Logistic Regression	Random Forest
F1-Score (Class 1)	0.14	0.13
Accuracy	0.79	0.69
Macro Avg F1-Score	0.51	0.47
Weighted Avg F1-Score	0.72	0.67

AUC-ROC Comparison

- **Logistic Regression:** 0.54

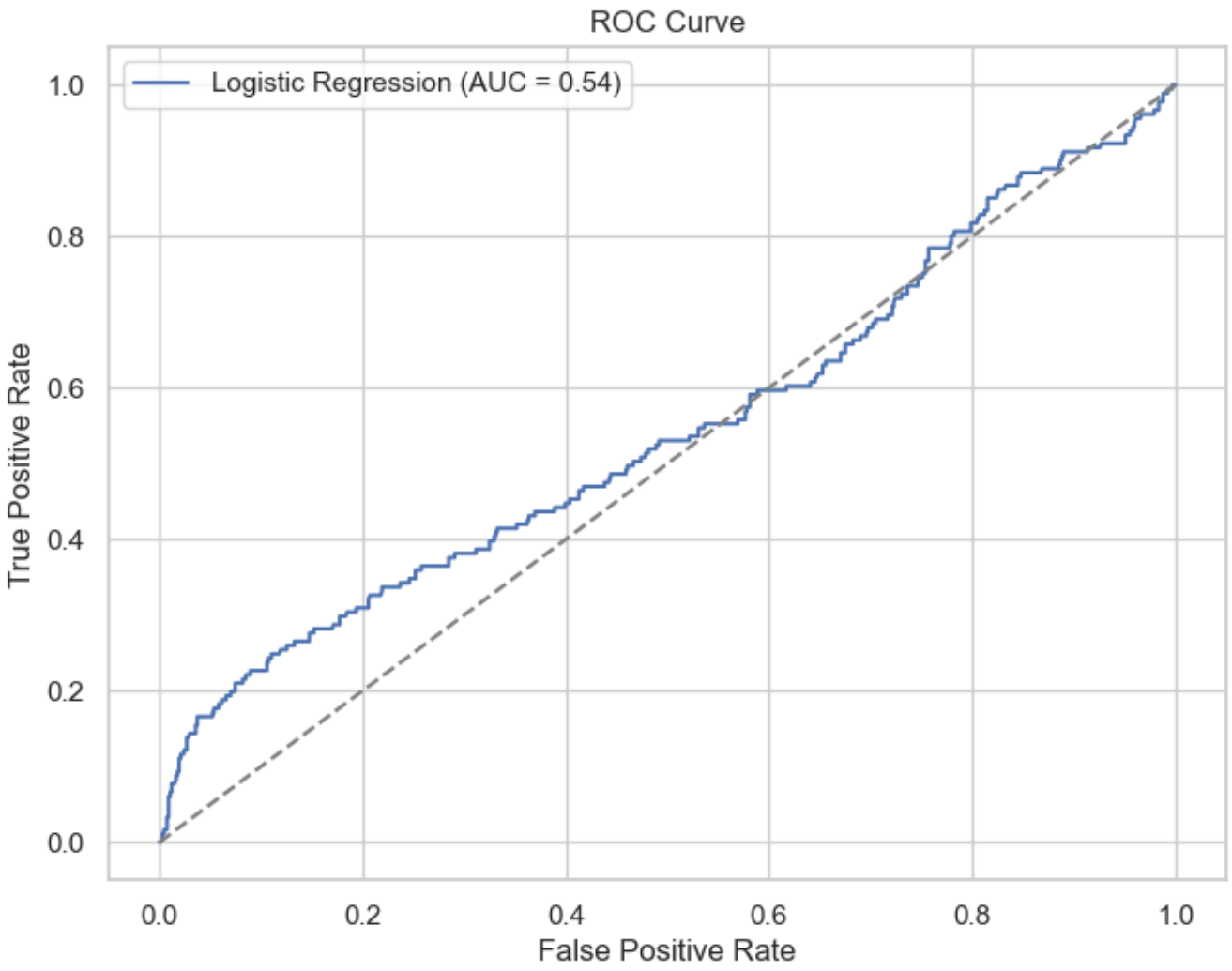

FIGURE 16: ROC PLOT - LR

- **Random Forest:** 0.32

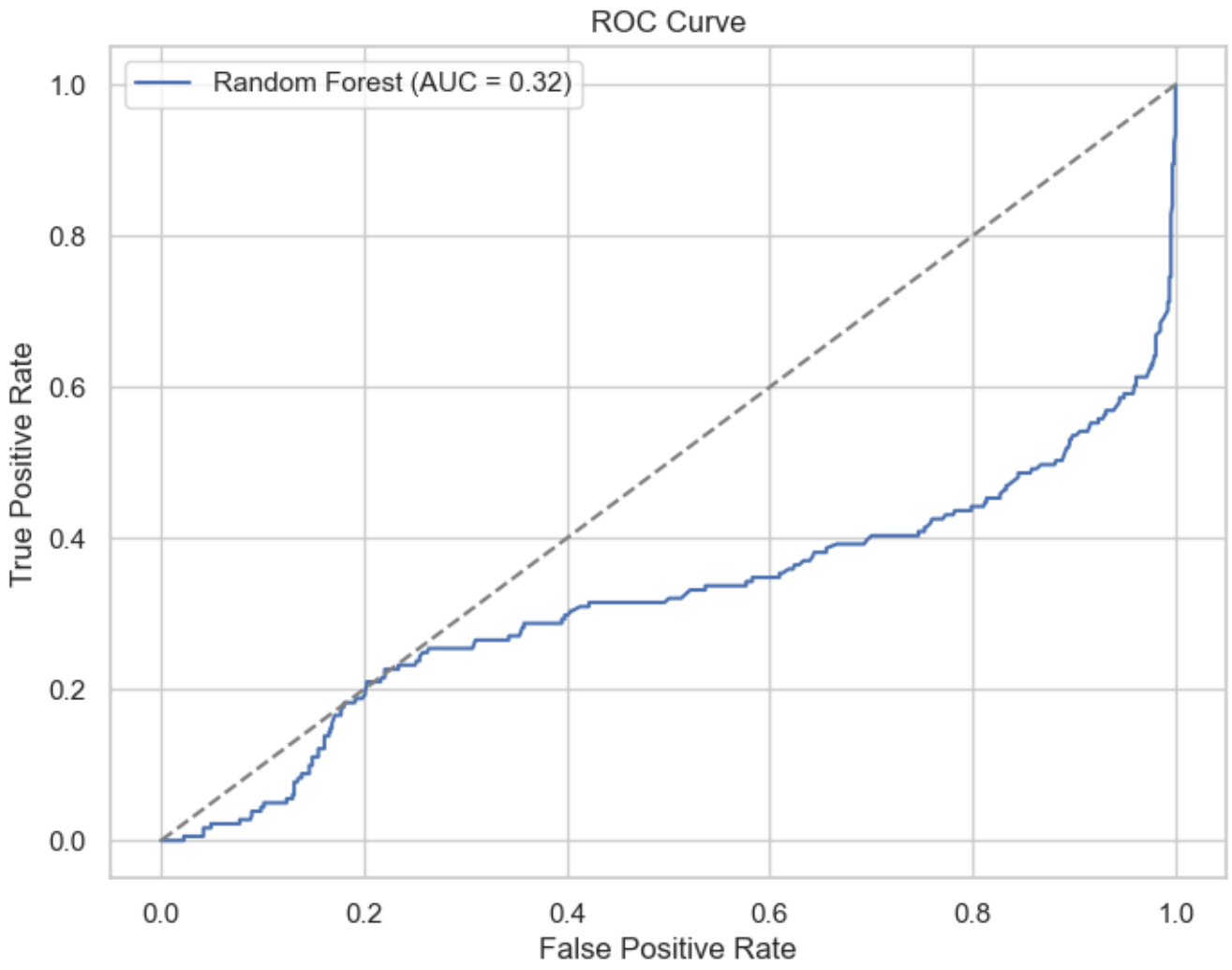

FIGURE 17: ROC PLOT - RF

Key Observations:

Logistic Regression outperformed the Random Forest model in terms of accuracy (79% vs. 69%) and F1-score for the non-default class (0.88 vs. 0.81).

The recall for the default class was significantly higher in Logistic Regression (8%) compared to Random Forest (11%), indicating both models struggle to identify defaults effectively.

The AUC-ROC score for Logistic Regression (0.54) suggests a moderate ability to distinguish between classes, while Random Forest (0.32) indicates poor discrimination.

Both models exhibit a tendency towards misclassifying defaults, with a high number of false negatives, particularly in Logistic Regression, which may necessitate further tuning or additional features to improve model performance.

Problem A.6 Model Performance Improvement

Problem A.6.1 Dealing with Multicollinearity

Multicollinearity can affect the stability and interpretability of Logistic Regression coefficients. To handle this, we:

- Calculate VIF for each predictor.
- if VIF > 5, consider the predictor highly collinear.
- Remove one of the collinear variables.
- Combine variables using dimensionality reduction techniques like PCA.

Metrics for Logistic Regression model after removing the Multicollinearity:

- Accuracy: 0.7876
- Precision: 0.5000
- Recall: 0.0387
- F1 Score: 0.0718
- ROC-AUC: 0.5643

Metrics for Random Forest model after removing the Multicollinearity:

- Accuracy: 0.6854
- Precision: 0.1345
- Recall: 0.0884
- F1 Score: 0.1067
- ROC-AUC: 0.3247

Problem A.6.2 Optimal threshold for Logistic Regression using ROC curve

Optimal Threshold for the logistic regression is **0.29**, represents a balanced approach to risk assessment in predicting company defaulters.

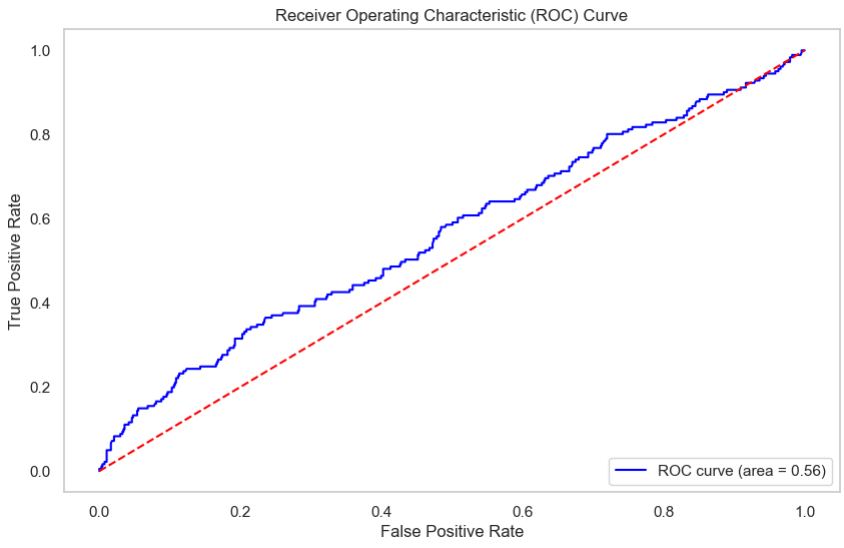

FIGURE 18: ROC CURVE - OPTIMAL THRESOLD

Problem A.6.3 Hyperparameter Tuning for Random Forest

To improve the performance of the Random Forest model, you can perform hyperparameter tuning using GridSearchCV.

- Best AUC Score: 0.49192569795995833
- AUC-ROC: 0.4506055940255741
- Confusion Matrix:
  [[654 17]
  [166 15]]
- Classification Report:

	precision	recall	f1-score	support
0	0.80	0.97	0.88	671
1	0.47	0.08	0.14	181
accuracy			0.79	852
macro avg	0.63	0.53	0.51	852
weighted avg	0.73	0.79	0.72	852

Problem A.6.4 Model performance check across different metrics

Logistic Regression Model Performance:

- Accuracy: 0.68
- Precision: 0.29
- Recall: 0.36
- F1-Score: 0.33
- AUC-ROC: 0.56
- Confusion Matrix:
  [[513 158]
  [115 66]]
- Classification Report:

	precision	recall	f1-score	support
0	0.82	0.76	0.79	671
1	0.29	0.36	0.33	181
accuracy			0.68	852
macro avg	0.56	0.56	0.56	852
weighted avg	0.71	0.68	0.69	852

Random Forest Model Performance:

- Accuracy: 0.79
- Precision: 0.47
- Recall: 0.08
- F1-Score: 0.14
- AUC-ROC: 0.45
- Confusion Matrix:
  [[654 17]
  [166 15]]
- Classification Report:

	precision	recall	f1-score	support
0	0.80	0.97	0.88	671
1	0.47	0.08	0.14	181
accuracy			0.79	852
macro avg	0.63	0.53	0.51	852
weighted avg	0.73	0.79	0.72	852

Comparison Summary:

Logistic Regression -> Accuracy: 0.68, AUC-ROC: 0.56

Random Forest -> Accuracy: 0.79, AUC-ROC: 0.45

Problem A.7 Model Performance Comparison & Final Model Selection

Problem A.7.1 Compare all the models built

In this analysis, we will compare the performance of four models: Logistic Regression with Optimal Threshold, Random Forest (After Hyperparameter Tuning), Logistic Regression, and Random Forest.

Metric	Logistic Regression (Optimal)	Random Forest (Tuned)	Logistic Regression	Random Forest
Accuracy	0.68	0.79	0.79	0.69
Precision (Class 0)	0.82	0.80	0.80	0.78

Metric	Logistic Regression (Optimal)	Random Forest (Tuned)	Logistic Regression	Random Forest
Precision (Class 1)	0.29	0.47	0.58	0.16
Recall (Class 0)	0.76	0.97	0.99	0.85
Recall (Class 1)	0.36	0.08	0.08	0.11
F1-Score (Class 0)	0.79	0.88	0.88	0.81
F1-Score (Class 1)	0.33	0.14	0.14	0.13
AUC-ROC	0.56	0.45	0.54	0.32

Based on the metrics:

- **Logistic Regression with Optimal Threshold:**

  This model shows a balanced precision for classifying non-defaults but struggles with identifying defaults, as indicated by a recall of only 36% for classifying defaults. The AUC-ROC of 0.56 suggests moderate discrimination ability.

- **Random Forest (After Hyperparameter Tuning):**

  While it has the highest accuracy among all models, its recall for classifying defaults is very low at 8%, indicating that it fails to identify most defaults. The precision for non-defaults is decent (47%), but the overall F1-score for defaults is poor (14%), and the AUC-ROC is also low at 0.45.

- **Logistic Regression:**

  This model performs well with high recall for non-defaults (99%) but has a very low recall for defaults (8%), similar to the tuned Random Forest. The AUC-ROC score of 0.54 indicates moderate performance.

- **Random Forest:**

  This model also has a high recall for non-defaults (85%) but struggles similarly with defaults, achieving only an 11% recall. The AUC-ROC score of 0.32 indicates poor discrimination ability.

Problem A.7.2 Select the final model with the proper justification

Based on the comparison, the Logistic Regression with Optimal Threshold is selected as the final model due to its balanced performance across different metrics despite its lower accuracy compared to Random Forest after tuning:

- It maintains a reasonable balance between precision and recall for both classes.
- It is more interpretable than Random Forest, making it easier to derive insights about feature importance.

Problem A.7.3 Check the most important features in the final model and draw inferences

For logistic regression, feature importance can be analyzed using the magnitude of the coefficients. Here are the features ranked by the absolute value of their coefficients, indicating their influence on predicting defaulters:

Feature	Coefficient	Absolute Coefficient
Raw material turnover	-0.03	0.03
Cash profit as % of total income	-0.02	0.02
Creditors turnover	-0.02	0.02
WIP turnover	-0.01	0.01
Debtors turnover	-0.01	0.01
Cash to average cost of sales per day	-0.01	0.01
Adjusted EPS	-0.01	0.01
PAT as % of net worth	-0.01	0.01
Quick ratio (times)	-0.00	0.00
Income from fincial services	-0.00	0.00

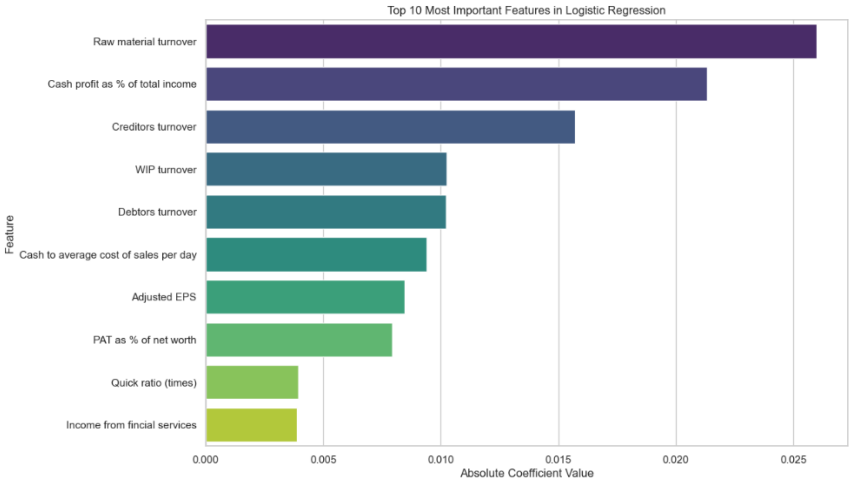

FIGURE 19: TOP FEATURES

Insights from Important Features:

- **Negative Coefficients:**

The majority of the top features have negative coefficients, indicating that higher values of these features are associated with a lower likelihood of being classified as a defaulting company. For instance, Raw Material Turnover and Cash Profit as % of Total Income have negative coefficients, suggesting that as these metrics increase, the risk of default decreases.

- **Impact of Cash Metrics:**

Features like Cash Profit as % of Total Income and Cash to Average Cost of Sales per Day indicate the importance of liquidity and cash management in assessing financial health. Companies with better cash management practices are less likely to default.

- **Turnover Ratios:**

The presence of Creditors Turnover, Debtors Turnover, and WIP Turnover among the top features suggests that efficient management of working capital is crucial for financial stability. High turnover ratios generally reflect good operational efficiency, which is positively correlated with a company's ability to meet its obligations.

- **Adjusted EPS and Profitability Ratios:**

The inclusion of Adjusted EPS and PAT as % of Net Worth highlights the significance of profitability metrics in predicting defaults. Companies with higher profitability relative to their equity are less likely to face financial distress.

- **Quick Ratio:**

The Quick Ratio, although having a minimal coefficient, is a critical liquidity measure that can indicate a company's ability to cover its short-term liabilities without relying on inventory sales. This reinforces the idea that liquidity plays a vital role in credit risk assessment.

Problem A.8 Actionable Insights & Recommendations

Based on the analysis and insights from the models and feature importance,

1. Improve Liquidity Management

- Variable: Quick Ratio (times) and Cash Profit as % of Total Income
- Insight: Companies with higher liquidity ratios are less likely to face default risks.
- Recommendation: Regularly monitor these liquidity metrics to ensure sufficient cash reserves are maintained. Implement cash flow forecasting to better manage short-term financial obligations.

2. Enhance Operational Efficiency

- Variable: Creditors Turnover, Debtors Turnover, and Raw Material Turnover
- Insight: Efficient management of working capital through high turnover ratios indicates better financial health.

- Recommendation: Streamline inventory management and optimize payment terms with suppliers and customers to improve these turnover ratios, thereby enhancing overall cash flow.

3. Focus on Profitability

- Variable: PAT as % of Net Worth and Adjusted EPS
- Insight: Higher profitability ratios suggest a stronger financial position, reducing default likelihood.
- Recommendation: Identify cost-saving opportunities and explore new revenue streams to boost profitability. Regularly review financial performance to identify areas for improvement.

4. Manage Debt Wisely

- Variable: Debt to Equity Ratio (times)
- Insight: A high debt-to-equity ratio can increase financial risk and the likelihood of default.
- Recommendation: Evaluate the current debt structure and consider refinancing options to lower interest rates. Aim to reduce unnecessary debt wherever possible.

5. Ensure Transparent Financial Reporting

- Variable: Overall financial reporting practices
- Insight: Clear and accurate financial reporting builds trust with stakeholders and helps assess creditworthiness.
- Recommendation: Adopt best practices in financial reporting, ensuring compliance with accounting standards, and regularly update stakeholders on financial performance.

6. Monitor External Economic Conditions

- Variable: Market trends and economic indicators (not directly measured in the dataset but crucial for context)
- Insight: Economic downturns can impact a company’s ability to meet its obligations.
- Recommendation: Stay informed about external economic factors that could affect business operations. Develop contingency plans to address potential challenges arising from market fluctuations.

Part B

Problem B.1 Problem Statement

Problem B.1.1 Context

Investors face market risk, arising from asset price fluctuations due to economic events, geopolitical developments, and investor sentiment changes. Understanding and analyzing this risk is crucial for informed decision-making and optimizing investment strategies.

Dataset: Market_Risk_Data_coded.csv

Problem B.1.2 Objective

The objective of this analysis is to conduct Market Risk Analysis on a portfolio of Indian stocks using Python. It uses historical stock price data to understand market volatility and riskiness. Using statistical measures like mean and standard deviation, investors gain a deeper understanding of individual stocks' performance and portfolio variability.

Through this analysis, investors can aim to achieve the following objectives:

- Risk Assessment: Analyze the historical volatility of individual stocks and the overall portfolio.
- Portfolio Optimization: Use Market Risk Analysis insights to enhance risk-adjusted returns.
- Performance Evaluation: Assess portfolio management strategies' effectiveness in mitigating market risk.
- Portfolio Performance Monitoring: Monitor portfolio performance over time and adjust as market conditions and risk preferences change.

Problem B.2 Data Overview

The dataset contains weekly stock price data for 5 Indian stocks over an 8-year period. The dataset enables us to analyze the historical performance of individual stocks and the overall market dynamics.

Problem B.3 Stock Price Graph Analysis

Before plotting, I will ensure the dates are in proper format and create a time-series line graph for each stock's price over time.

Problem B.3.1 Draw a Stock Price Graph (Stock Price vs Time) for the given stocks

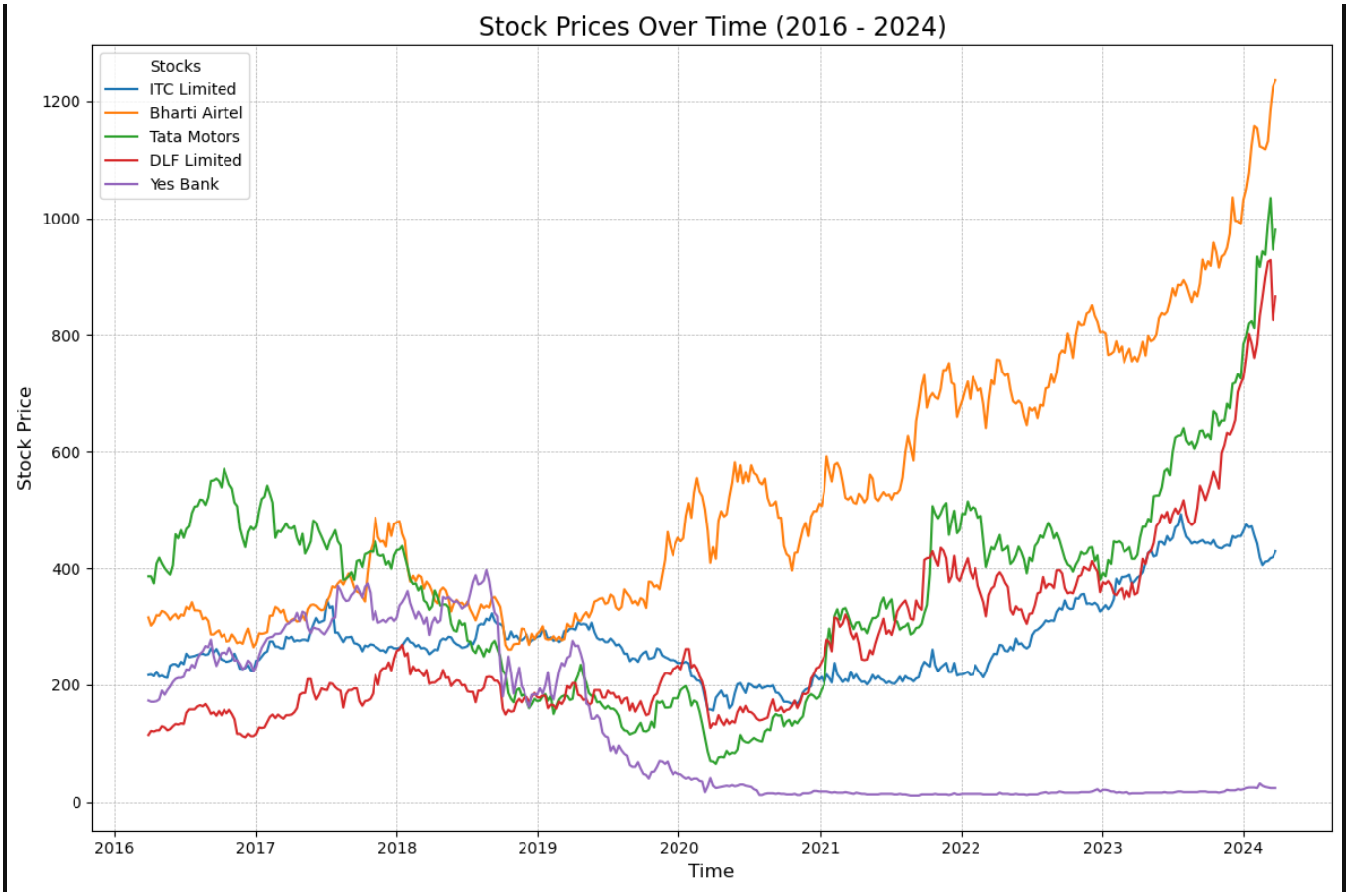

FIGURE 20: STOCK PRICE GRAPH

Problem B.3.2 Write observations

1. Long-Term Growth:

   Tata Motors (orange line) shows a significant upward trend, especially after 2021, indicating substantial growth and investor confidence in the stock.

2. Stable Performers:

   ITC Limited (blue line) and Bharti Airtel (red line) exhibit stable growth over time, reflecting consistency and resilience in their performance.

3. Decline in Performance:

   Yes Bank (purple line) shows a sharp decline in stock price after 2018, possibly due to financial issues or regulatory challenges, and remains stagnant afterward.

4. Cyclical Behavior:

   DLF Limited (green line) demonstrates cyclical behavior with periodic ups and downs, likely influenced by the real estate sector's market cycles.

5. Volatility:

Yes Bank exhibits the highest volatility in earlier years, while Tata Motors displays an increase in volatility as its stock price rises sharply in recent years.

6. Sector Influence:

Stocks like DLF Limited and Yes Bank might be more affected by sector-specific risks, whereas ITC Limited and Bharti Airtel seem less influenced by such factors.

Problem B.4 Stock Returns Calculation & Analysis

Problem B.4.1 Calculate Returns for all stocks

Compute the daily returns for each stock using the formula:

$$Return = \frac{Price_t - Price_{t-1}}{Price_{t-1}}$$

Date	ITC Limited	Bharti Airtel	Tata Motors	DLF Limited	Yes Bank
2016-03-28	NaN	NaN	NaN	NaN	NaN
2016-04-04	0.004608	-0.044304	0.000000	0.061404	-0.011561
2016-04-11	-0.013761	0.019868	-0.031088	-0.008264	0.000000
2016-04-18	0.037209	0.038961	0.090909	0.016667	0.005848
2016-04-25	-0.040359	-0.003125	0.024510	0.000000	0.017442

FIGURE 21: RETURNS (FIRST 5 ROWS)

Problem B.4.2 Calculate the Mean & Standard Deviation for the returns of all stocks

Stocks	Mean Returns	Standard Deviations
ITC Limited	0.002281	0.036127
Bharti Airtel	0.004029	0.039073
Tata Motors	0.004088	0.061976
DLF Limited	0.006540	0.057796
Yes Bank	-0.000475	0.091095

Problem B.4.3 Draw a plot of Mean vs Standard Deviation for all stock returns

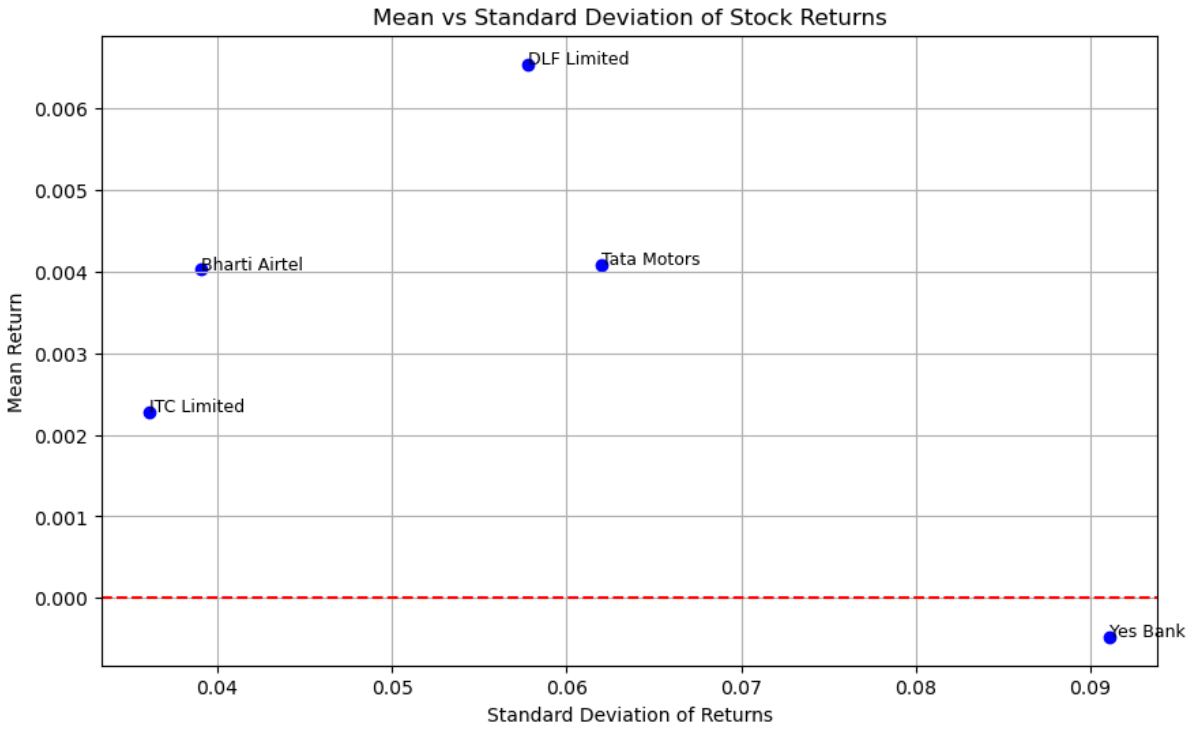

FIGURE 22: MEAN VS STANDARD DEVIATION

Problem B.4.3 Write observations and inferences

- DLF Limited, with the highest mean return, also has moderate volatility, making it an attractive option for investors seeking growth with manageable risk.
- Yes Bank, while having the highest volatility, also shows negative mean returns, indicating that high risk does not always equate to high returns.
- Investors may consider Bharti Airtel and Tata Motors as viable options due to their positive returns and manageable volatility levels.
- ITC Limited presents a stable investment with low volatility but lower returns, suitable for conservative investors.

Problem B.5 Actionable Insights & Recommendations

Based on the analysis of the historical stock price data for ITC Limited, Bharti Airtel, Tata Motors, DLF Limited, and Yes Bank, several actionable insights and recommendations can be derived:

1. Investment Strategy Based on Performance

- DLF Limited: With the highest mean return of 0.006540, it appears to be the best performer among the stocks analyzed. Investors seeking growth should consider increasing their allocation to DLF Limited.
- Tata Motors and Bharti Airtel: Both stocks show promising mean returns (0.004088 and 0.004029, respectively) with moderate volatility. These stocks could be suitable for investors looking for a balance between risk and return.

- ITC Limited: While it has a stable performance with lower volatility (standard deviation of 0.036127), its mean return is modest (0.002281). It may appeal to conservative investors seeking stability rather than high returns.
- Yes Bank: The negative mean return (-0.000475) and high volatility (standard deviation of 0.091095) suggest that it is a high-risk investment. Investors should exercise caution and consider divesting if there are no signs of recovery.

2. Risk Management

- To mitigate risk, investors should consider diversifying their portfolios by including a mix of these stocks. For instance, combining higher-risk stocks like Tata Motors and Yes Bank with more stable options like ITC Limited can help balance overall portfolio risk.

3. Long-Term Perspective

- Given the historical data spans several years, adopting a long-term investment perspective may yield better results, especially for stocks like DLF Limited and Bharti Airtel, which have shown resilience over time.

THE END