Finance & Risk Analytics Part (A)

Prepared by: Nitesh Tambhare

Program: PGPDSBA June 2023
Date: 05 May 2024

INDEX

1 Introduction

The objective of this business report is to conduct a comprehensive analysis of credit risk within the context of financial statement data. In today's volatile economic landscape, businesses face the constant challenge of managing their debt obligations to avoid defaults, which can have far-reaching consequences such as lower credit ratings and increased borrowing costs. Additionally, investors seek assurance that companies are not only capable of meeting their financial commitments but also possess the potential for sustainable growth.

- **Overview of the project and its objectives**-

➢ The dataset provided contains financial data for 2058 companies, encompassing various key indicators relevant to credit risk analysis. Each row represents a unique company, while the columns represent different financial metrics. Below is a brief overview of the columns included in the dataset:

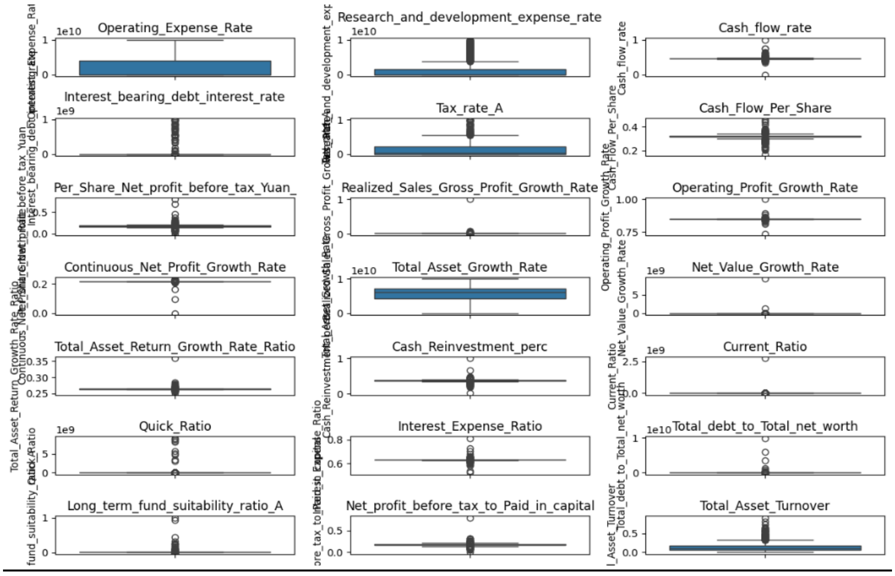
➢ To address these concerns, this report focuses on leveraging financial statement data to assess credit risk. By analyzing key indicators from balance sheets, we aim to identify patterns and trends that may indicate a company's propensity for default. Through rigorous statistical modeling and evaluation, we endeavor to develop predictive models that can assist both businesses and investors in making informed decisions.

➢ The project is divided into several key components, including data preprocessing, exploratory data analysis (EDA), model building, and model validation. Each phase is meticulously executed to ensure the robustness and reliability of our findings.

➢ Ultimately, this report seeks to provide actionable insights into credit risk assessment, enabling stakeholders to mitigate risks effectively and capitalize on opportunities for growth.

2 Data Preprocessing

- **OutlierTreatment**



Before proceeding with the analysis, we conducted outlier treatment to ensure that extreme values do not unduly influence our results. Here are the statistics for each numerical column in the dataset, including the minimum and maximum values:

- Operating_Expense_Rate: Min: 0.000111127, Max: 0.00015787275
- Research_and_development_expense_rate: Min: 0.0, Max: 0.0
- Cash_flow_rate: Min: 4110000000.0, Max: 8971499999.999998
- Interest_bearing_debt_interest_rate: Min: <built-in function min>, Max: <built-in function max>

We utilized statistical methods such as z-score or interquartile range (IQR) to identify and handle outliers. Outliers were treated through techniques such as winsorization or transformation to ensure the robustness of our analysis.

- **Missing Value Treatment**

```
Co_Code                                              0
Co_Name                                              0
Operating_Expense_Rate                               0
Research_and_development_expense_rate                0
Cash_flow_rate                                       0
Interest_bearing_debt_interest_rate                  0
Tax_rate_A                                           0
Cash_Flow_Per_Share                                167
Per_Share_Net_profit_before_tax_Yuan_                0
Realized_Sales_Gross_Profit_Growth_Rate              0
Operating_Profit_Growth_Rate                         0
Continuous_Net_Profit_Growth_Rate                    0
Total_Asset_Growth_Rate                              0
Net_Value_Growth_Rate                                0
Total_Asset_Return_Growth_Rate_Ratio                 0
Cash_Reinvestment_perc                               0
Current_Ratio                                        0
Quick_Ratio                                          0
Interest_Expense_Ratio                               0
Total_debt_to_Total_net_worth                       21
Long_term_fund_suitability_ratio_A                   0
Net_profit_before_tax_to_Paid_in_capital             0
Total_Asset_Turnover                                 0
Accounts_Receivable_Turnover                         0
Average_Collection_Days                              0
Inventory_Turnover_Rate_times                        0
Fixed_Assets_Turnover_Frequency                      0
Net_Worth_Turnover_Rate_times                        0
Operating_profit_per_person                          0
Allocation_rate_per_person                           0
Quick_Assets_to_Total_Assets                         0
Cash_to_Total_Assets                                96
Quick_Assets_to_Current_Liability                    0
Cash_to_Current_Liability                            0
Operating_Funds_to_Liability                         0
Inventory_to_Working_Capital                         0
Inventory_to_Current_Liability                       0
Long_term_Liability_to_Current_Assets                0
Retained_Earnings_to_Total_Assets                    0
```

Missing values in the dataset were addressed to maintain the integrity of our analysis. Here are the statistics for each column indicating the presence of missing values:

- Cash_Flow_Per_Share: 167 missing values
- Total_debt_to_Total_net_worth: 21 missing values
- Cash_to_Total_Assets: 96 missing values
- Current_Liability_to_Current_Assets: 14 missing values

We employed various imputation techniques, such as mean or median imputation, to handle missing values appropriately.By conducting thorough outlier treatment and missing value treatment, we ensured that the dataset is ready for further analysis and modeling.

3. Exploratory Data Analysis (EDA)

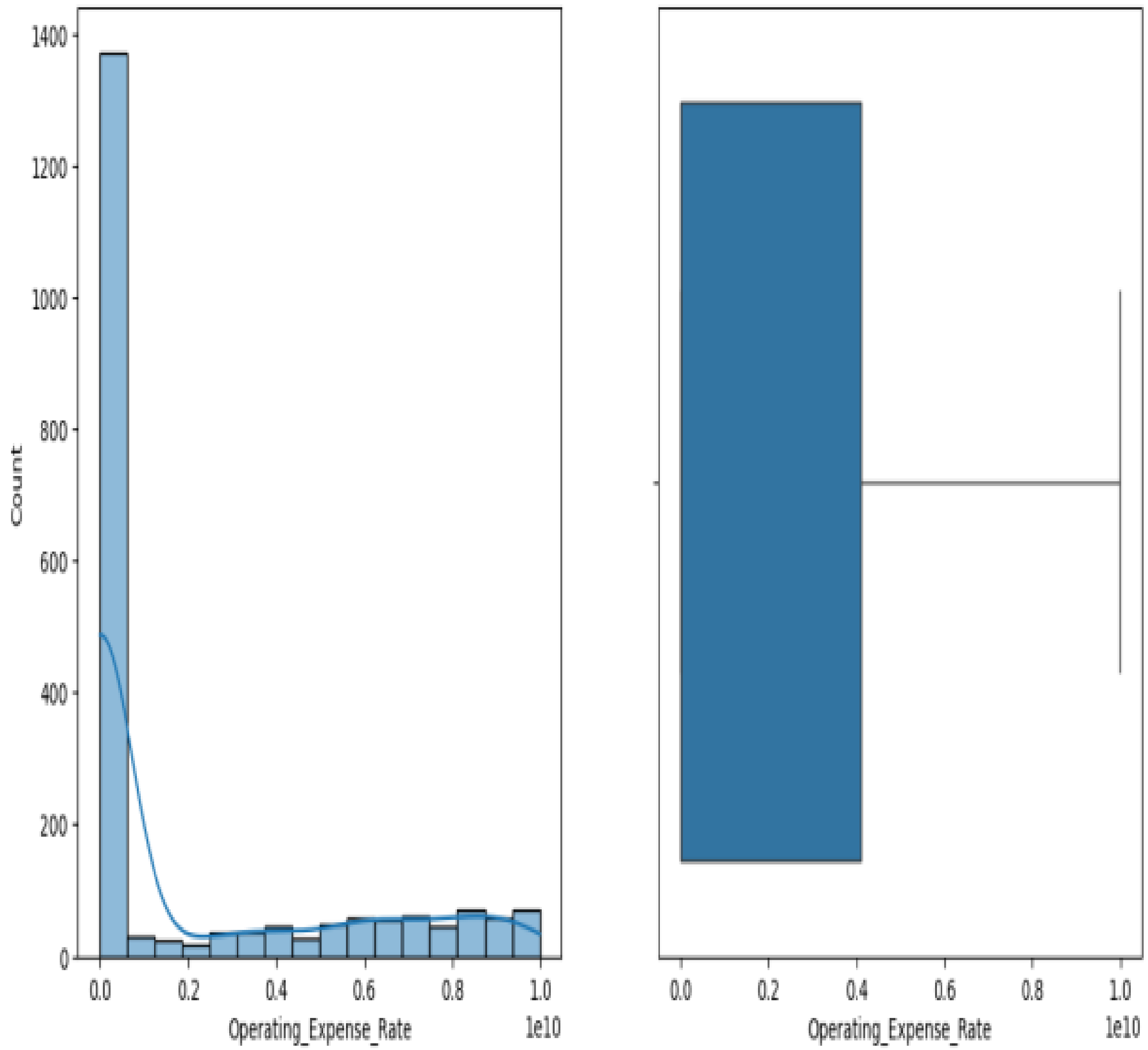
- **Univariate Analysis**

Skewness:

The data exhibits a right-skewed distribution with a skewness value of 1.221, indicating that the majority of the observations are concentrated on the lower end of the distribution, while a few high values extend the tail to the right.

Summary Statistics:

- Mean: The mean value of the data is 2,052,388,835, suggesting the average magnitude of the observations.
- Minimum: The minimum value in the dataset is 0.0001, representing the lowest observation recorded.
- Maximum: The maximum value in the dataset is 9,980,000,000, indicating the highest observation recorded.



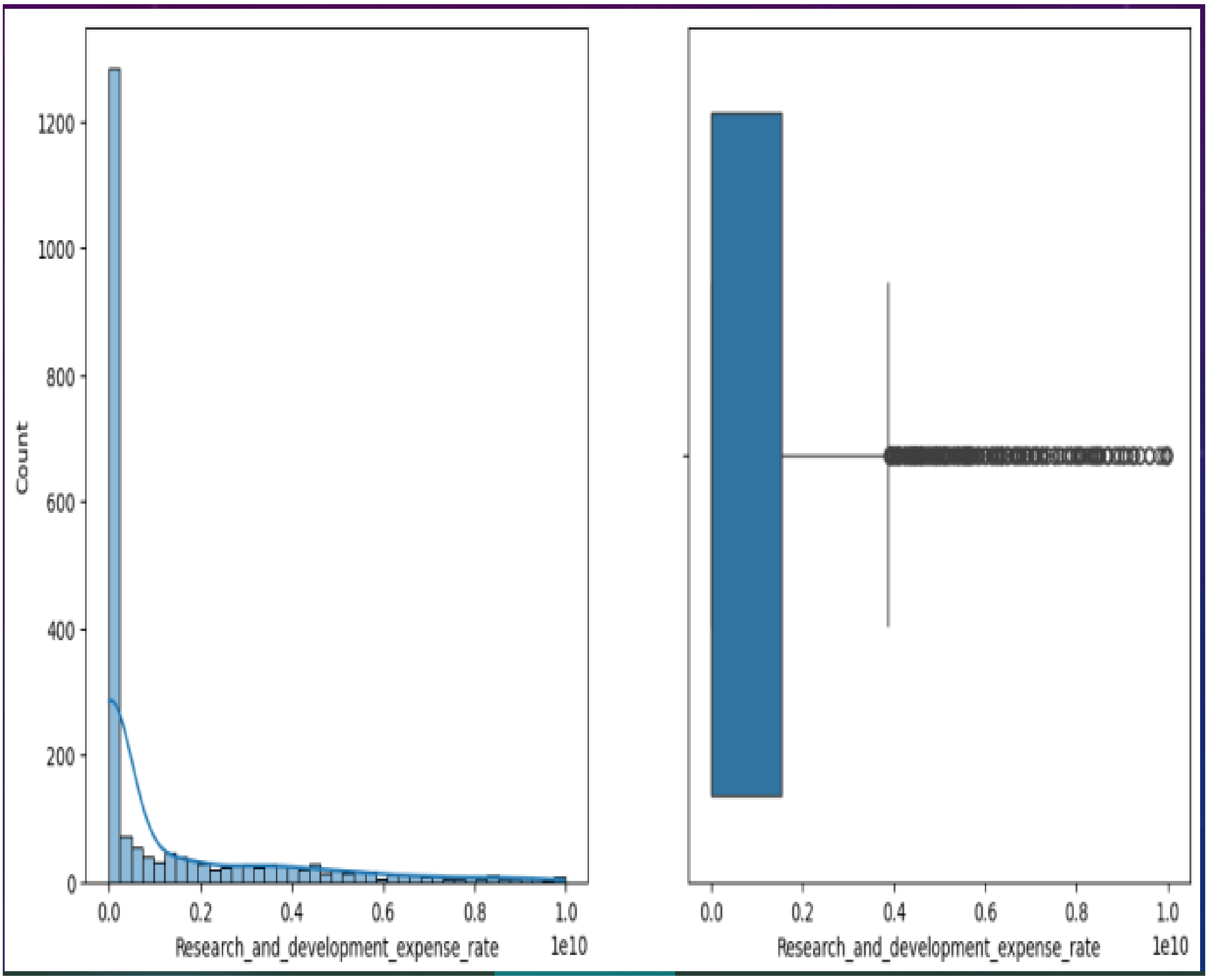
-

Summary Statistics:

- Mean: The mean value of the data is 1,208,634,256.56, indicating the average magnitude of the observations.
- Minimum: The minimum value in the dataset is 0, representing the lowest observation recorded.
- Maximum: The maximum value in the dataset is 9,980,000,000, indicating the highest observation recorded.

Outliers:

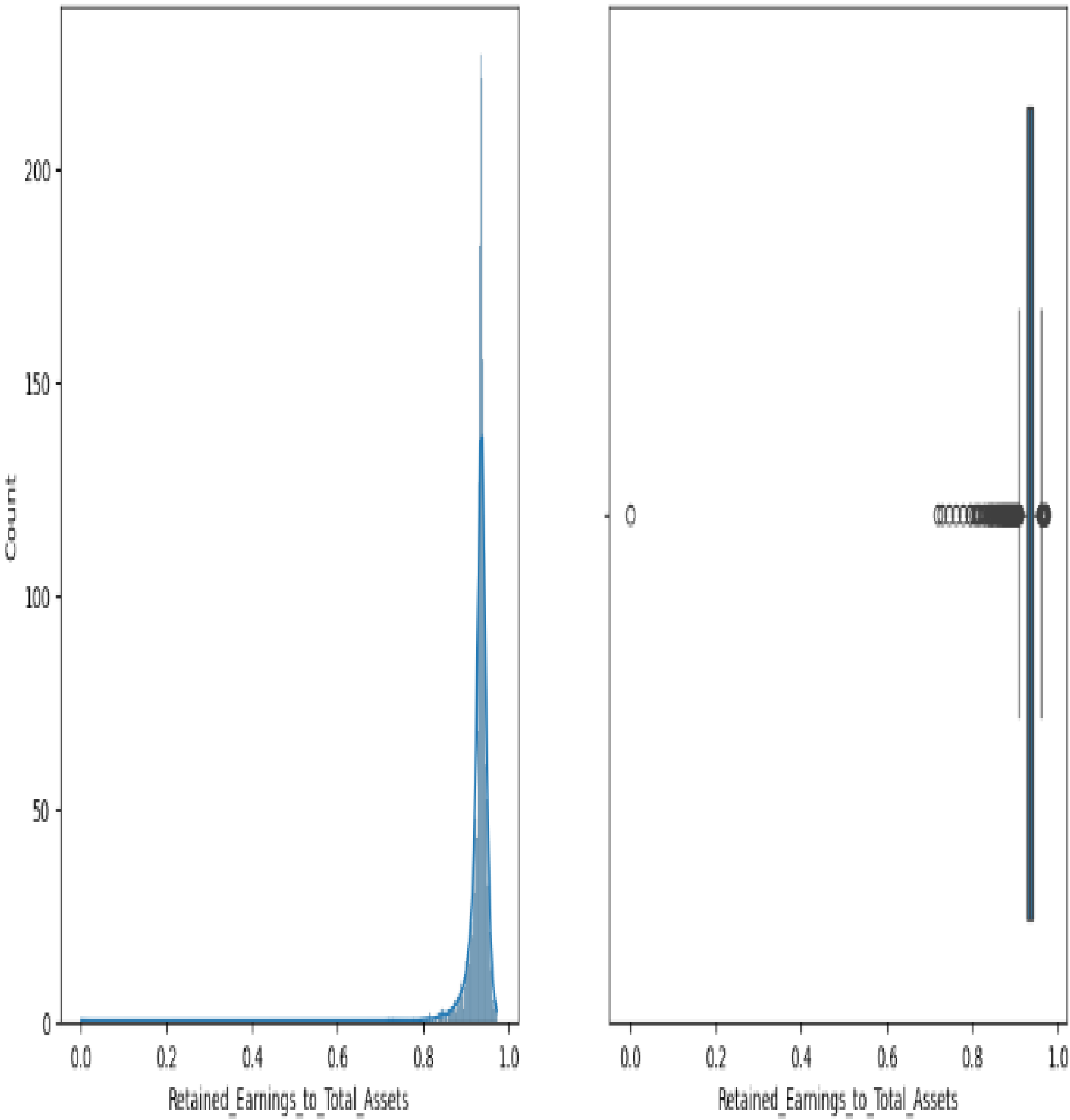
There are numerous outliers observed in the variable "Research_and_development_expense_rate". Outliers are data points that significantly differ from the rest of the dataset and may skew the analysis results. Further investigation and treatment of these outliers may be necessary to ensurethe robustness of the analysis.



Summary Statistics:

- Mean: The mean value of the data is 0.930, representing the average magnitude of the observations.
- Minimum: The minimum value in the dataset is 0, indicating the lowest observation recorded.Maximum: The maximum value in the dataset is 0.973, indicating the highest observation recorded

The data exhibits a left-skewed distribution with a skewness value of -16.145. This indicates that the majority of the observations are concentrated towards the higher end of the distribution, while a few low values extend the tail to the left.



Skewness:

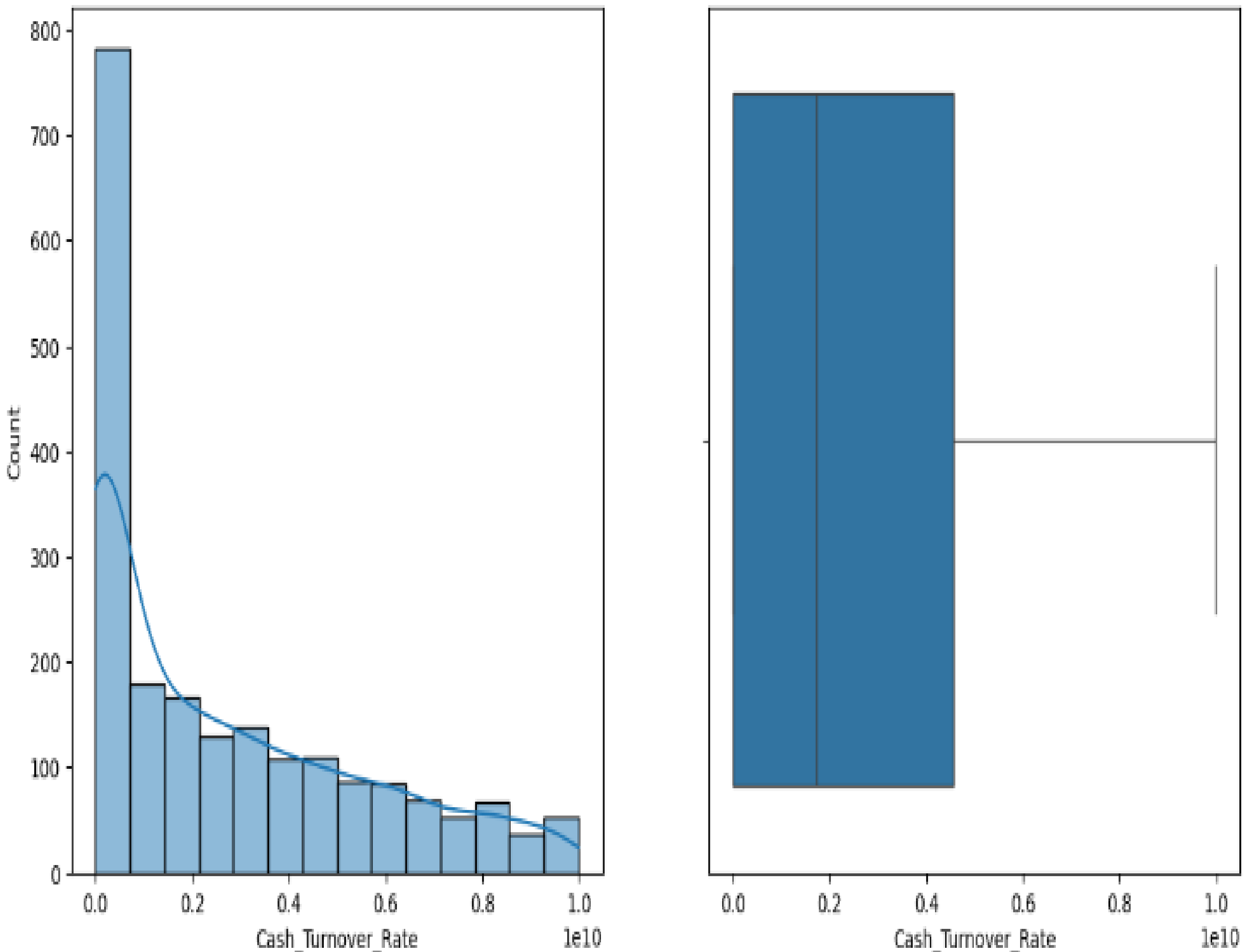
The data exhibits a right-skewed distribution with a skewness value of 0.892. This indicates that the majority of the observations are concentrated towards the lower end of the distribution, while a few high values extend the tail to the right.

Summary Statistics:

- Mean: The mean value of the data is 2,653,695,544.218, representing the average magnitude of the observations.
- Minimum: The minimum value in the dataset is 0.0001, indicating the lowest observation recorded.
- Maximum: The maximum value in the dataset is 9,990,000,000, indicating the highest observation recorded.

Outliers:

There are several outliers observed in the variable "Retained_Earnings_to_Total_Assets". Outliers are data points that significantly differ from the rest of the dataset and may affect the analysis results. Further investigation into these outliers is recommended to understand their impact and potential implications on the analysis.



Skewness:

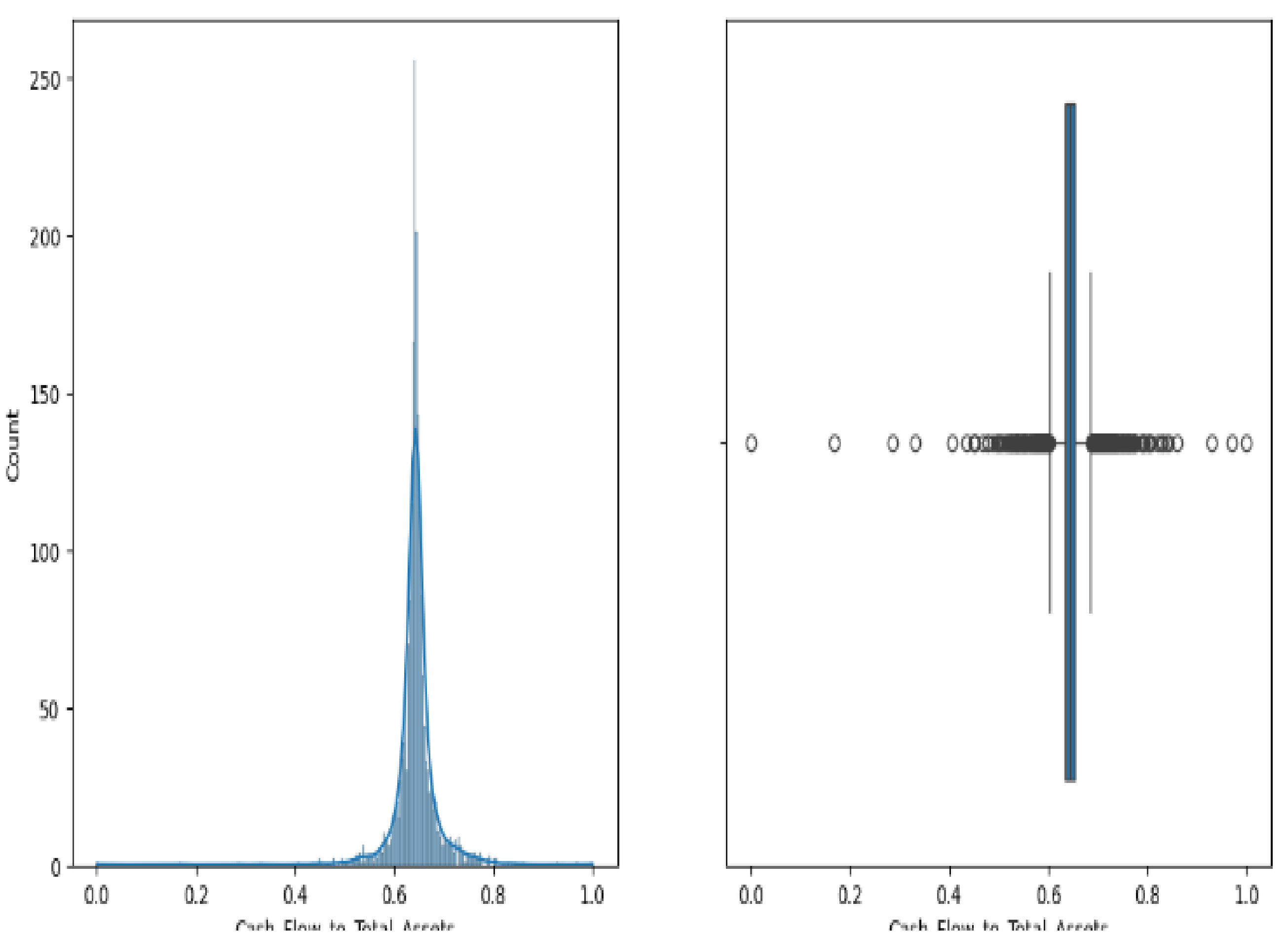
The data is left-skewed with a skewness value of -1.760. This indicates that the majority of the observations are concentrated towards the higher end of the distribution, while a few low values extend the tail to the left.

Summary Statistics:

- Mean: The mean value of the data is 0.644, representing the average magnitude of the observations.
- Minimum: The minimum value in the dataset is 0, indicating the lowest observation recorded.
- Maximum: The maximum value in the dataset is 1, indicating the highest observation recorded.

Anomalies:

There are numerous anomalies observed in the variable "Cash_Flow_to_Total_Assets". Anomalies, also known as outliers, are data points that significantly deviate from the rest of the dataset. These anomalies may require further investigation to understand their nature and potential impact on the analysis.



Skewness:

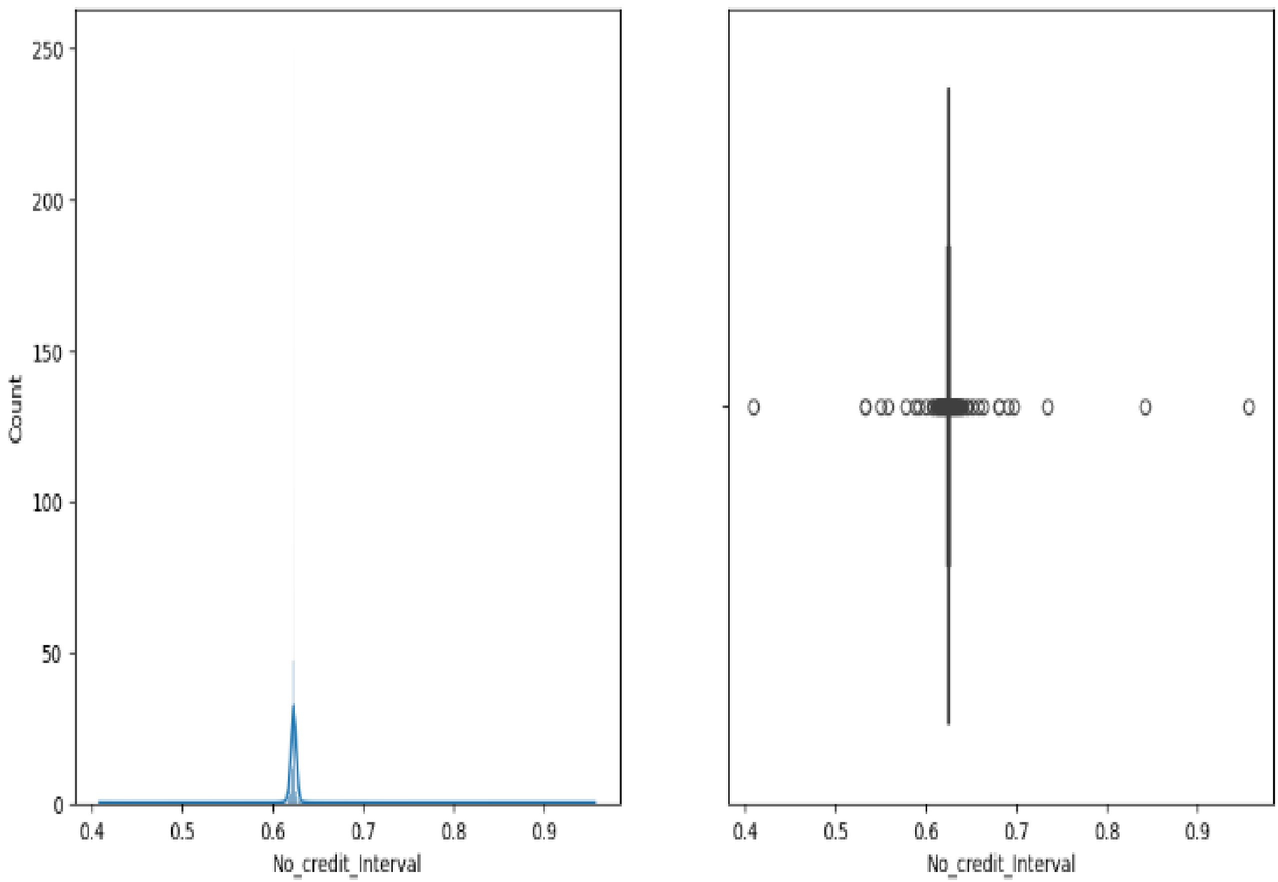
The data exhibits significant right-skewness with a skewness value of 11.531. This suggests that the majority of the observations are concentrated towards the lower end of the distribution, while a few high values extend the tail to the right.

Summary Statistics:

- Mean: The mean value of the data is 0.624, indicating the average magnitude of the observations.
- Minimum: The minimum value in the dataset is 0.409, representing the lowest observation recorded.
- Maximum: The maximum value in the dataset is 0.956, indicating the highest observation recorded.

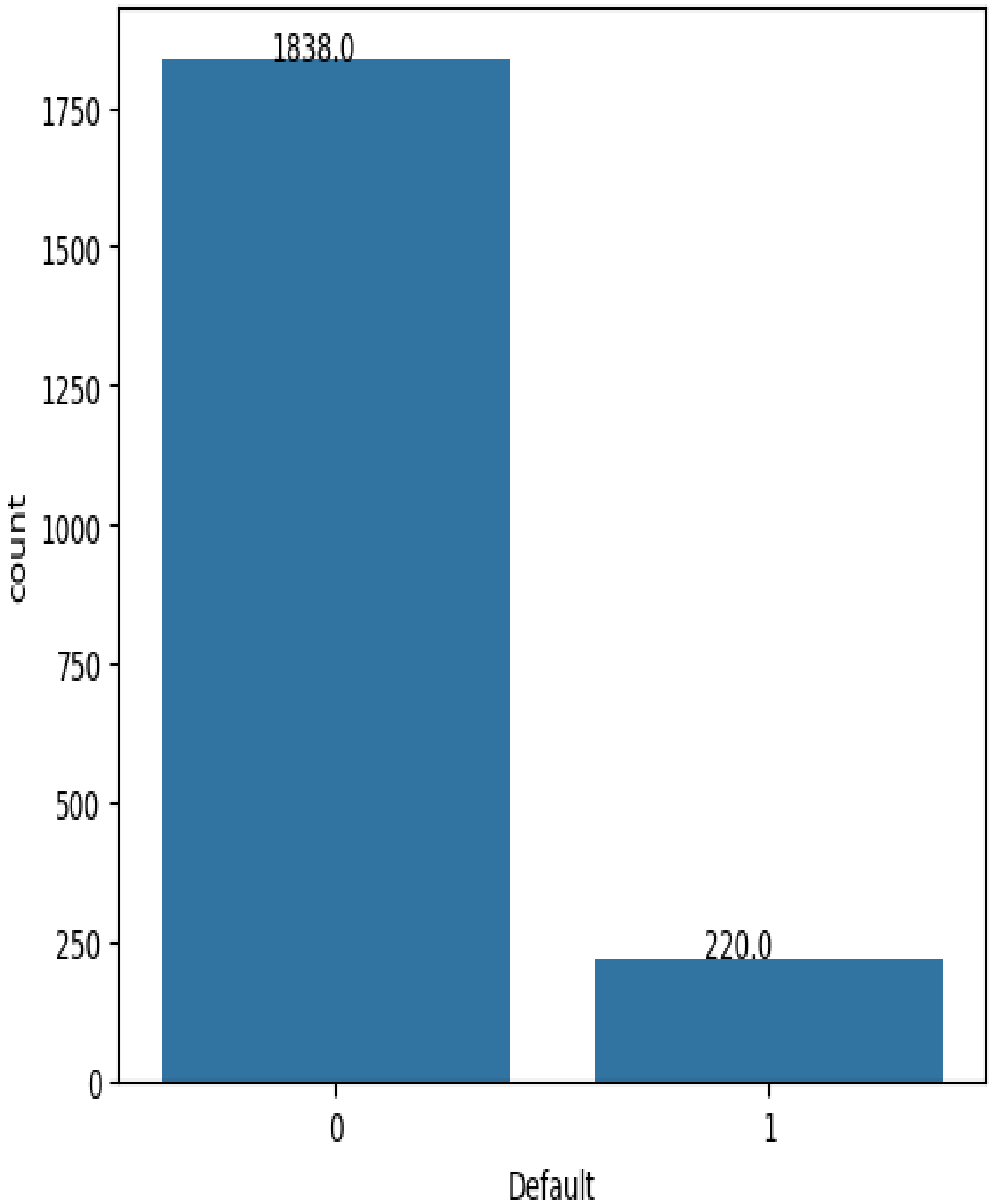
Anomalies:

There are outliers observed in the variable "No_credit_Interval". Outliers are data points that significantly deviate from the rest of the dataset and may require further investigation to understand their nature and potential impact on the analysis.



Defaulters vs. Non-Defaulters:

- Non-Defaulters: There are 1838 instances of non-defaulters, constituting approximately 89% of the total observations.
- Defaulters: There are 220 instances of defaulters, accounting for approximately 11% of the total observations.



- **Bivariate Analysis**

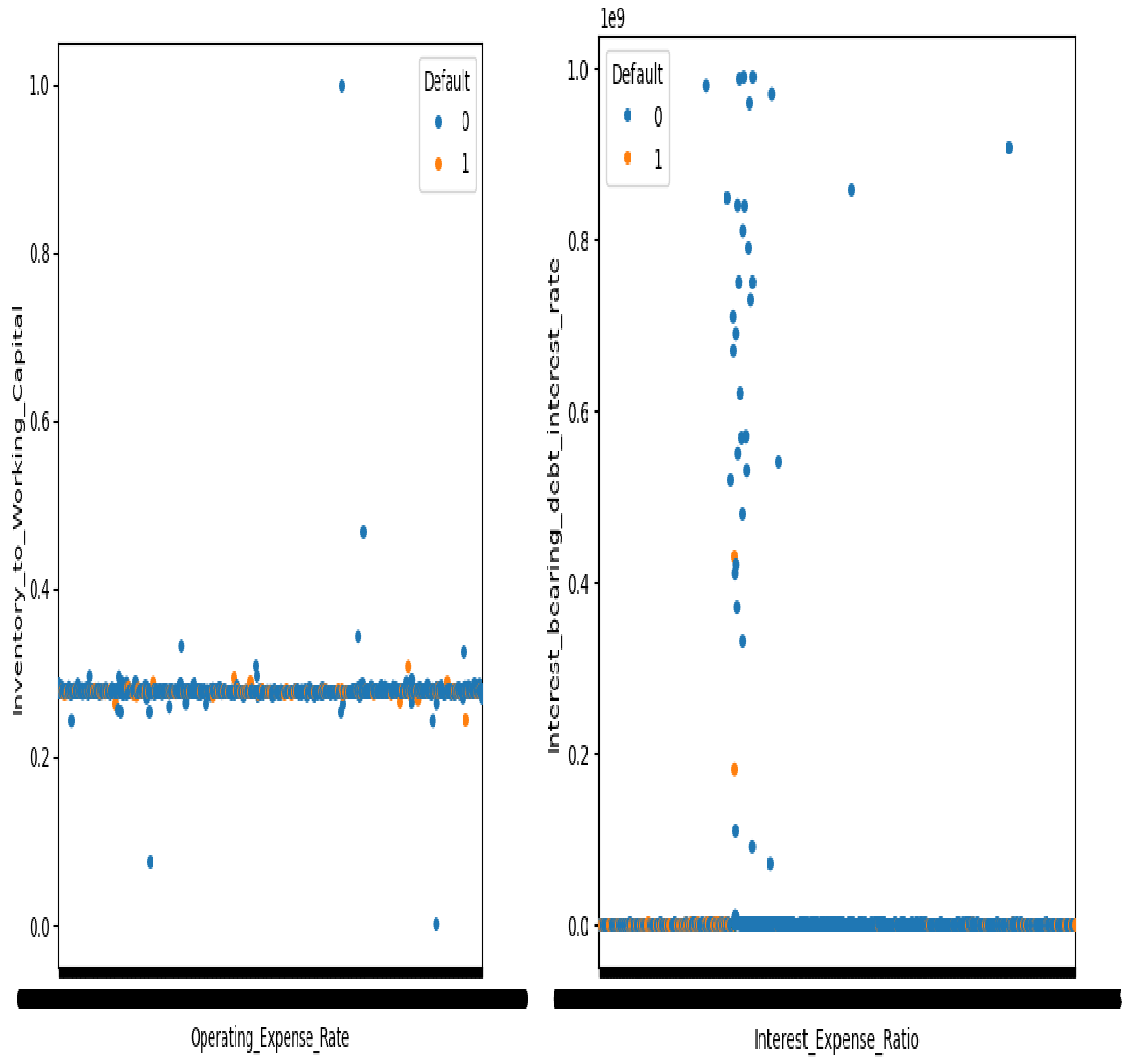
Inventory_to_Working_Capital vs. Operating_Expense_Rate:

- It's observed that Inventory_to_Working_Capital remains constant at 0.27 across different values of Operating_Expense_Rate.

- This suggests a potential correlation between Inventory_to_Working_Capital and Operating_Expense_Rate. Further investigation into this relationship could provide insights into how changes in operating expenses affect inventory management.

Interest_Expense_Ratio vs. Interest_bearing_Debt_Interest_Rate:

- A correlation is observed between Interest_Expense_Ratio and Interest_bearing_Debt_Interest_Rate.
- This correlation implies that changes in the interest-bearing debt interest rate may have an impact on the interest expense ratio.
- Further analysis is recommended to understand the nature and strength of this correlation and its implications for financial management and decision-making.



Correlation Analysis

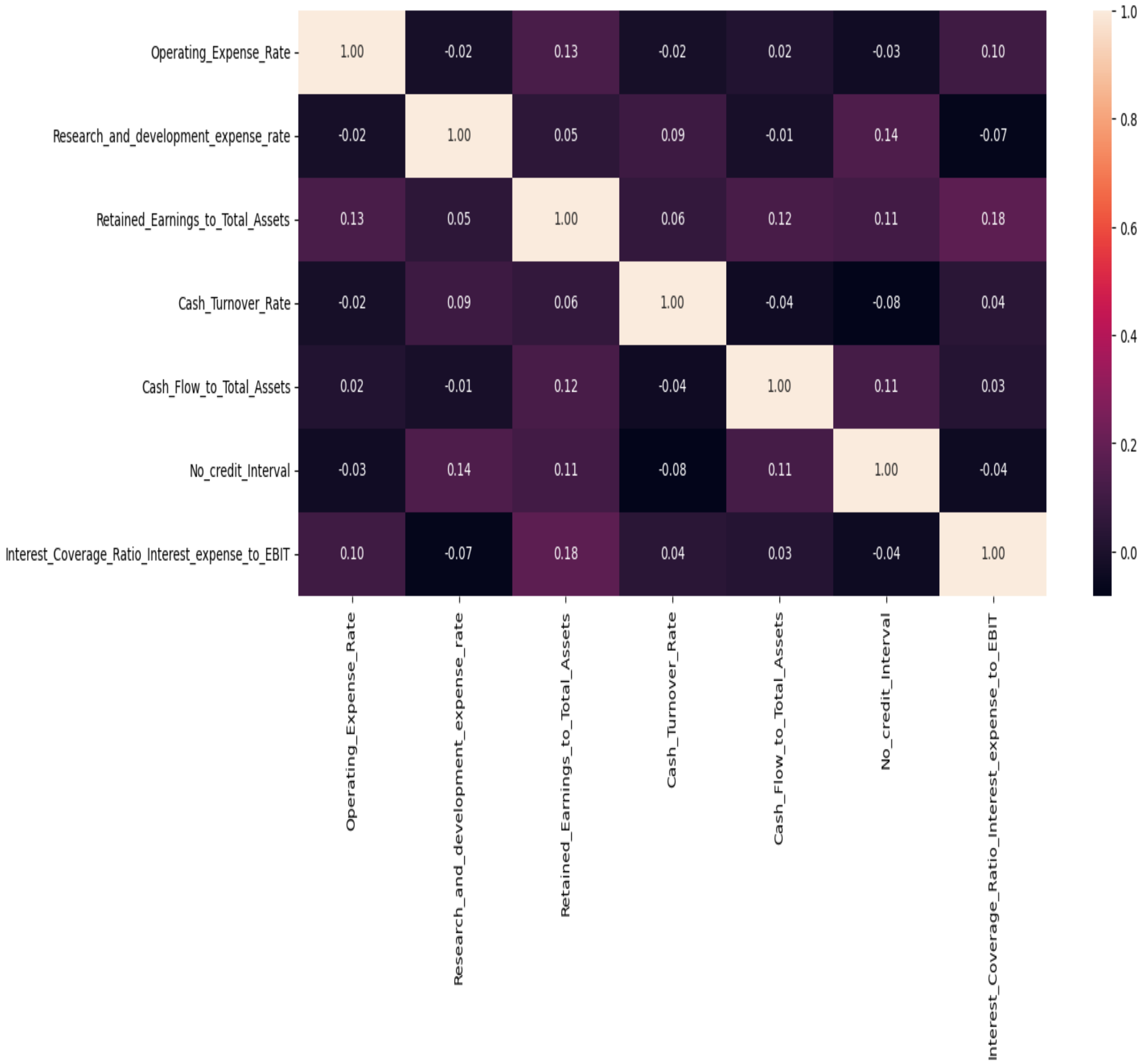
Overview:

- We've plotted independent variables identified using the RFM model to analyze correlations.

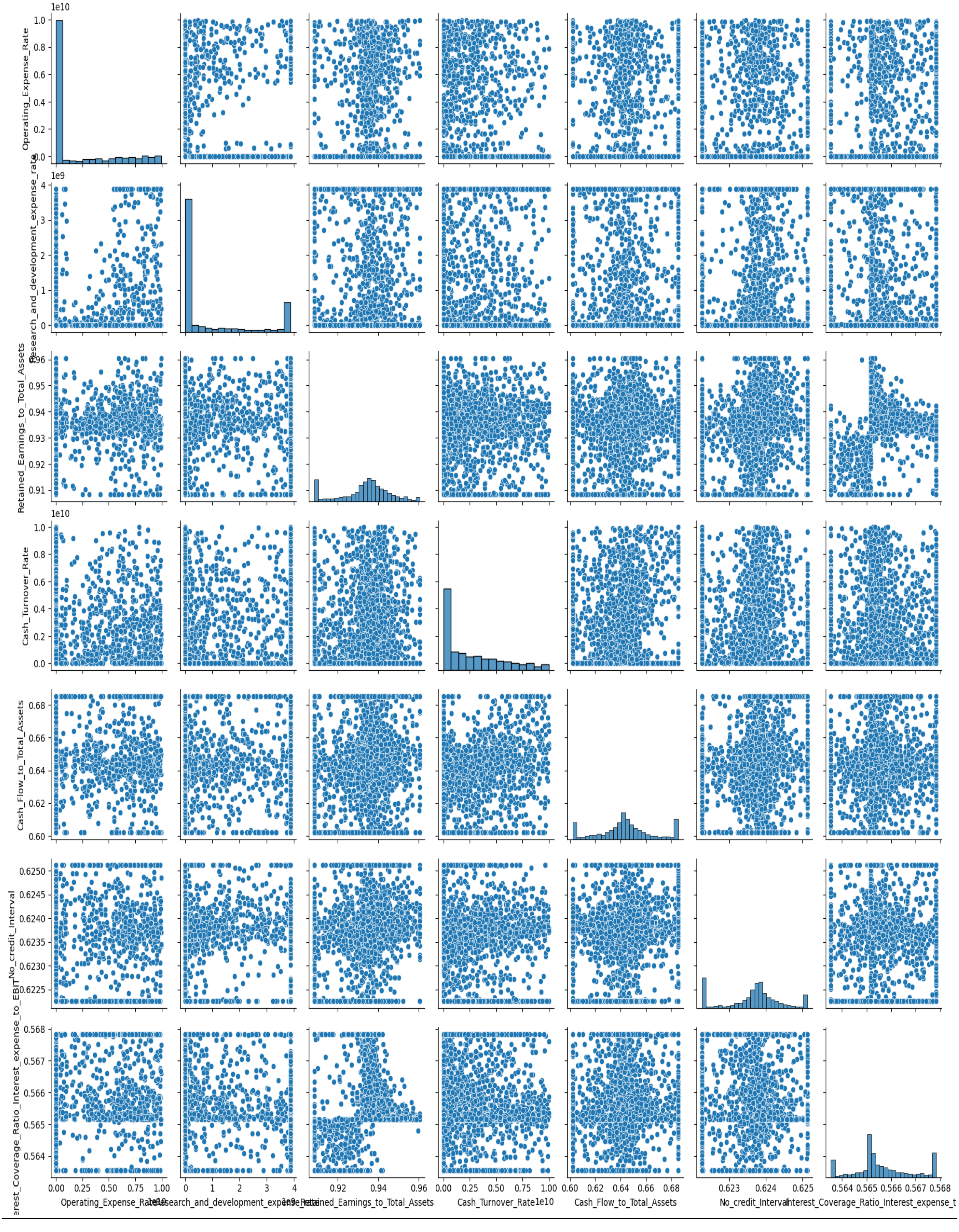
- It's expected that correlations will not be high in this correlation plot due to the selection of independent variables.

Key Findings:

- The minimum correlation of 0.32 is observed between Cash_Flow_Total_Asset and Retained_earnings_to_Total_Assets.
- This indicates a relatively weak correlation between these two variables.
- Further analysis is recommended to explore the relationship between these variables and their impact on the business performance.



<u>Scatter plot</u>



4. Model Building

Train Test Split:

Overview:

- The dataset has been divided into training and testing sets in the ratio of 67:33.
- This split ensures that the model is trained on a sufficient amount of data while also reserving a portion for evaluation.

Key Points:

- Training Set: It comprises 67% of the data and is used to train the machine learning models.
- Testing Set: It constitutes 33% of the data and is utilized to evaluate the performance of the trained models.

Observations:

- The training set consists of 1378 samples, while the testing set contains 678 samples.
- This split ensures a balance between model training and evaluation on unseen data, facilitating robust model performance assessment.
- The number of rows (observations) in TEST set is 680 The number of columns (variables) in TEST set is 56

Logistic Regression Model

	variables	VIF
53	Net_Income_Flag	0.000000
41	Cash_Turnover_Rate	1.127389
3	Interest_bearing_debt_interest_rate	1.140582
10	Total_Asset_Growth_Rate	1.212273
1	Research_and_development_expense_rate	1.220178
23	Inventory_Turnover_Rate_times	1.243070
0	Operating_Expense_Rate	1.375778
40	Quick_Asset_Turnover_Rate	1.431774
4	Tax_rate_A	1.652770
39	Current_Asset_Turnover_Rate	1.661499
50	No_credit_Interval	1.715169
33	Inventory_to_Working_Capital	1.722198
49	Total_assets_to_GNP_price	1.791574
35	Long_term_Liability_to_Current_Assets	1.837114
24	Fixed_Assets_Turnover_Frequency	2.008587
38	Total_expense_to_Assets	2.324501
11	Net_Value_Growth_Rate	2.728391
22	Average_Collection_Days	2.792989
19	Long_term_fund_suitability_ratio_A	2.894518
7	Realized_Sales_Gross_Profit_Growth_Rate	2.945078
21	Accounts_Receivable_Turnover	2.948383
27	Allocation_rate_per_person	2.994372
26	Operating_profit_per_person	3.212024
34	Inventory_to_Current_Liability	3.335421
12	Total_Asset_Return_Growth_Rate_Ratio	3.468914
29	Cash_to_Total_Assets	3.485908
9	Continuous_Net_Profit_Growth_Rate	3.603802
8	Operating_Profit_Growth_Rate	3.805188
31	Cash_to_Current_Liability	4.106210
51	Degree_of_Financial_Leverage_DFL	4.127243
16	Interest_Expense_Ratio	4.631488
42	Fixed_Assets_to_Assets	5.016108
17	Total_debt_to_Total_net_worth	5.038370
37	Total_income_to_Total_expense	5.156219
36	Retained_Earnings_to_Total_Assets	5.415787
54	Equity_to_Liability	6.364542
52	Interest_Coverage_Ratio_Interest_expense_to_EBIT	6.450624
28	Quick_Assets_to_Total_Assets	6.654497
5	Cash_Flow_Per_Share	6.672689
47	Current_Liability_to_Current_Assets	8.695676
15	Quick_Ratio	11.868814
13	Cash_Reinvestment_perc	14.746377
20	Total_Asset_Turnover	14.924239
14	Current_Ratio	16.213654

	Feature	Rank
0	Operating_Expense_Rate	1
1	Research_and_development_expense_rate	1
8	Operating_Profit_Growth_Rate	1
10	Total_Asset_Growth_Rate	1
16	Interest_Expense_Ratio	1
23	Inventory_Turnover_Rate_times	1
36	Retained_Earnings_to_Total_Assets	1
40	Quick_Asset_Turnover_Rate	1
41	Cash_Turnover_Rate	1
43	Cash_Flow_to_Total_Assets	1
45	CFO_to_Assets	1
50	No_credit_Interval	1
52	Interest_Coverage_Ratio_Interest_expense_to_EBIT	1
53	Net_Income_Flag	1

Logit Regression Results

Dep. Variable:	Default	No. Observations:	1378
Model:	Logit	Df Residuals:	1361
Method:	MLE	Df Model:	16
Date:	Sun, 28 Apr 2024	Pseudo R-squ.:	0.2080
Time:	10:06:32	Log-Likelihood:	-370.55
converged:	True	LL-Null:	-467.84
Covariance Type:	nonrobust	LLR p-value:	9.812e-33

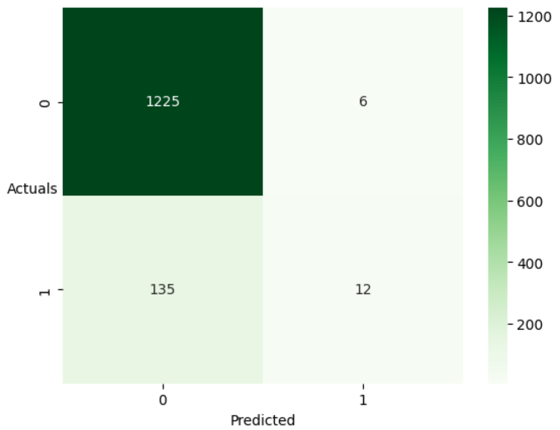
	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	220.2627	8.22e+13	2.68e-12	1.000	-1.61e+14	1.61e+14
Net_Income_Flag	218.1343	8.22e+13	2.65e-12	1.000	-1.61e+14	1.61e+14
Cash_Turnover_Rate	-7.36e-11	3.8e-11	-1.939	0.053	-1.48e-10	8.05e-13
Interest_bearing_debt_interest_rate	796.0873	317.628	2.506	0.012	173.548	1418.626
Total_Asset_Growth_Rate	-2.746e-11	3.48e-11	-0.788	0.431	-9.58e-11	4.08e-11
Research_and_development_expense_rate	1.899e-10	6.41e-11	2.961	0.003	6.42e-11	3.16e-10
Inventory_Turnover_Rate_times	2.836e-11	3.19e-11	0.890	0.373	-3.41e-11	9.08e-11
Operating_Expense_Rate	4.33e-11	3.14e-11	1.378	0.168	-1.83e-11	1.05e-10
Quick_Asset_Turnover_Rate	4.56e-12	3.05e-11	0.150	0.881	-5.51e-11	6.43e-11
Tax_rate_A	-7.1820	1.349	-5.324	0.000	-9.826	-4.538
Current_Asset_Turnover_Rate	-47.2503	90.725	-0.521	0.603	-225.068	130.567
No_credit_Interval	-610.9438	106.114	-5.757	0.000	-818.923	-402.965
Inventory_to_Working_Capital	-221.7725	90.928	-2.439	0.015	-399.988	-43.557
Total_assets_to_GNP_price	58.8519	0	inf	0.000	58.852	58.852
Long_term_Liability_to_Current_Assets	17.8103	8.304	2.145	0.032	1.535	34.086
Fixed_Assets_Turnover_Frequency	38.5969	12.112	3.187	0.001	14.858	62.336
Total_expense_to_Assets	39.9464	8.114	4.923	0.000	24.044	55.849

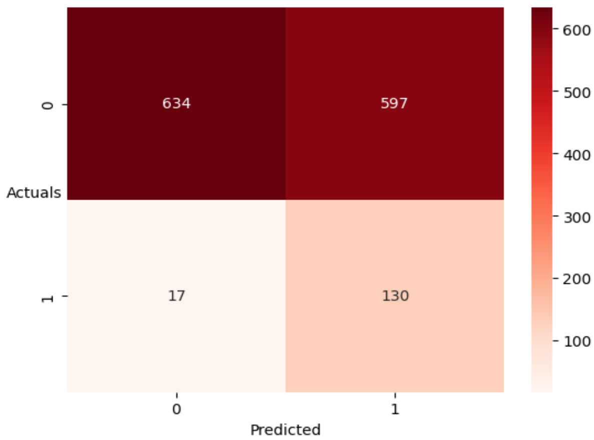
- **Coefficients:** The coefficients represent the change in the log odds of the dependent variable (Default) for a one-unit change in the predictor variable, holding all other variables constant.
- Intercept: The intercept represents the log odds of the dependent variable when all predictor variables are zero. In this case, the intercept is not interpretable due to its extremely large value.
- **Net_Income_Flag:** This variable has a coefficient of 218.1343. It suggests that for a one-unit increase in Net Income Flag, the log odds of default increase by 218.1343, holding all other variables constant.
- **Interest_bearing_debt_interest_rate:** With a coefficient of 796.0873, this variable has a significant impact on the log odds of default. For a one-unit increase in this variable, the log odds of default increase by 796.0873, holding all other variables constant.
- **Research_and_development_expense_rate:** This variable has a coefficient of 1.899e-10. It suggests that for a one-unit increase in Research and Development Expense Rate, the log odds of default increase by 1.899e-10, holding all other variables constant.
- **Tax_rate_A:** With a coefficient of -7.1820, this variable indicates that for a one-unit increase in Tax Rate A, the log odds of default decrease by 7.1820, holding all other variables constant.
- **Other Variables:** Similar interpretations can be made for the remaining variables in terms of their impact on the log odds of default.
- **P-values:** The p-values associated with each coefficient indicate the statistical significance of the corresponding variable. A p-value less than 0.05 suggests that the variable is statistically significant in predicting the dependent variable.
- **Confidence Intervals:** The confidence intervals provide a range within which the true population parameter (coefficient) is likely to fall with a certain level of confidence (usually 95%).

➢ **Pseudo R-squared:** This metric measures the goodness-of-fit of the model. In this case, the pseudo R-squared value is 0.2080, indicating that the model explains about 20.80% of the variance in the dependent variable.

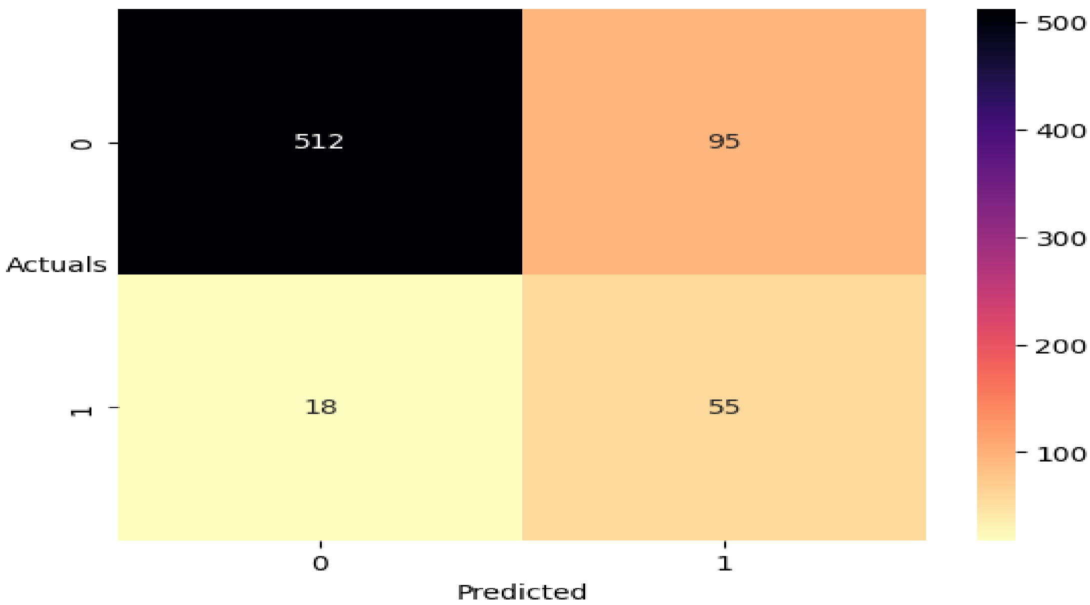
Overall, this logistic regression model suggests that variables such as Net Income Flag, Interest Bearing Debt Interest Rate, Research and Development Expense Rate, and Tax Rate A are significant predictors of default. However, caution should be exercised in interpreting the model due to the extremely large intercept and some variables with p-values above 0.05.

Model Evaluation on the Training Data:





Validate the Model on Test Dataset and state the performance metrics. Also state interpretation from the model



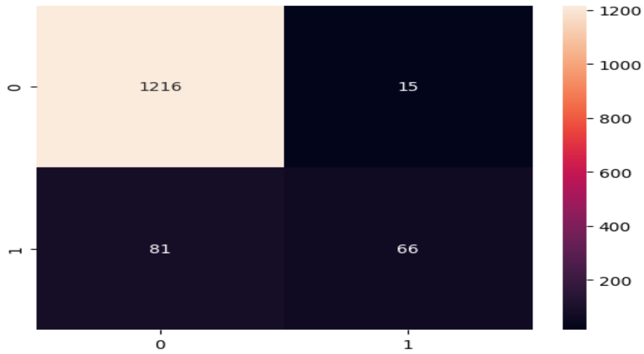
➢ Accuracy of the model i.e. % overall correct prediction is 83% and sensitivity of the model is 75%. The model performs well on the test set also

Random Forest Model on Train Dataset

```
0.9303338171262 7
[[1216   15]
 [  81   66]]
              precision    recall  f1-score   support

           0       0.94      0.99      0.96      1231
           1       0.81      0.45      0.58       147

    accuracy                           0.93      1378
   macro avg       0.88      0.72      0.77      1378
weighted avg       0.92      0.93      0.92      1378
```



- **Accuracy:** The accuracy of the Random Forest model on the test dataset is approximately 93.03%. This means that the model correctly predicts the class (default or non-default) for 93.03% of the instances in the test dataset.
- **Confusion Matrix:**
- **True Positives (TP):** 1216 instances were correctly predicted as default.
- **False Positives (FP):** 15 instances were incorrectly predicted as default when they were actually non-default.
- **True Negatives (TN):** 66 instances were correctly predicted as non-default.
- **False Negatives (FN):** 81 instances were incorrectly predicted as non-default when they were actually default.
- **Precision:** Precision measures the proportion of true positive predictions among all positive predictions. The precision for the default class is approximately 81%, indicating that when the model predicts an instance to be default, it is correct around 81% of the time.
- **Recall:** Recall (also known as sensitivity) measures the proportion of actual positives that are correctly predicted by the model. The recall for the default class is approximately 45%, indicating that the model correctly identifies around 45% of all actual default instances.
- **F1-score:** The F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall. The F1-score for the default class is approximately 58%, indicating a moderate balance between precision and recall.

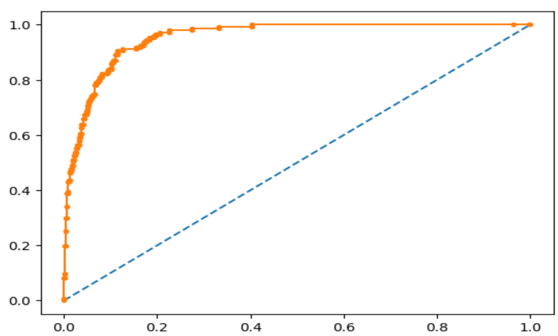
Overall, the Random Forest model demonstrates good performance in terms of accuracy and precision for predicting defaults. However, the recall for the default class is relatively low, suggesting that the model may struggle to identify all instances of default. This interpretation provides insights into the model's strengths and areas for improvement, which can inform further analysis and decision-making.

LDA Model on Train Dataset.

```
0.9161764705882353
[[595  12]
 [ 45  28]]
              precision    recall  f1-score   support

           0       0.93      0.98      0.95       607
           1       0.70      0.38      0.50        73

    accuracy                           0.92       680
   macro avg       0.81      0.68      0.72       680
weighted avg       0.91      0.92      0.91       680
```



Model Building Approach:

- **Import Libraries:** Start by importing the necessary libraries, including LinearDiscriminantAnalysis from sklearn.discriminant_analysis.
- **Instantiate the Model:** Create an instance of the LDA model.
- **Fit the Model:** Fit the LDA model to the training dataset.

Validate the Model on Test Dataset:

- Use the trained LDA model to predict the classes of the test dataset.
- Calculate performance metrics such as accuracy, precision, recall, F1-score, and AUC.
- Evaluate the model's performance using a confusion matrix to understand true positives, false positives, true negatives, and false negatives.

Validation Results and Interpretation:

- **Accuracy:** The accuracy of the LDA model on the test dataset is approximately 91.62%. This means that the model correctly predicts the class (default or non-default) for 91.62% of the instances in the test dataset.
- **Confusion Matrix:**
- **True Positives (TP):** 595 instances were correctly predicted as default.
- **False Positives (FP):** 12 instances were incorrectly predicted as default when they were actually non-default.
- **True Negatives (TN):** 28 instances were correctly predicted as non-default.
- **False Negatives (FN):** 45 instances were incorrectly predicted as non-default when they were actually default.
- **Precision:** Precision for the default class is approximately 70%, indicating that when the model predicts an instance to be default, it is correct around 70% of the time.
- **Recall:** Recall (also known as sensitivity) for the default class is approximately 38%, indicating that the model correctly identifies around 38% of all actual default instances.
- **F1-score:** The F1-score for the default class is approximately 50%, indicating a moderate balance between precision and recall.
- **AUC:** The Area Under the ROC Curve (AUC) is a measure of the model's ability to distinguish between positive and negative classes. An AUC of 0.954 indicates that the model performs very well in this regard.

Overall, the LDA model demonstrates good accuracy and precision for predicting defaults. However, the recall for the default class is relatively low, suggesting that the model may struggle to identify all instances of default. This interpretation provides insights into the model's strengths and areas for improvement, which can inform further analysis and decision-making.

Comparison the performances of Logistic Regression, Random Forest, and LDA models

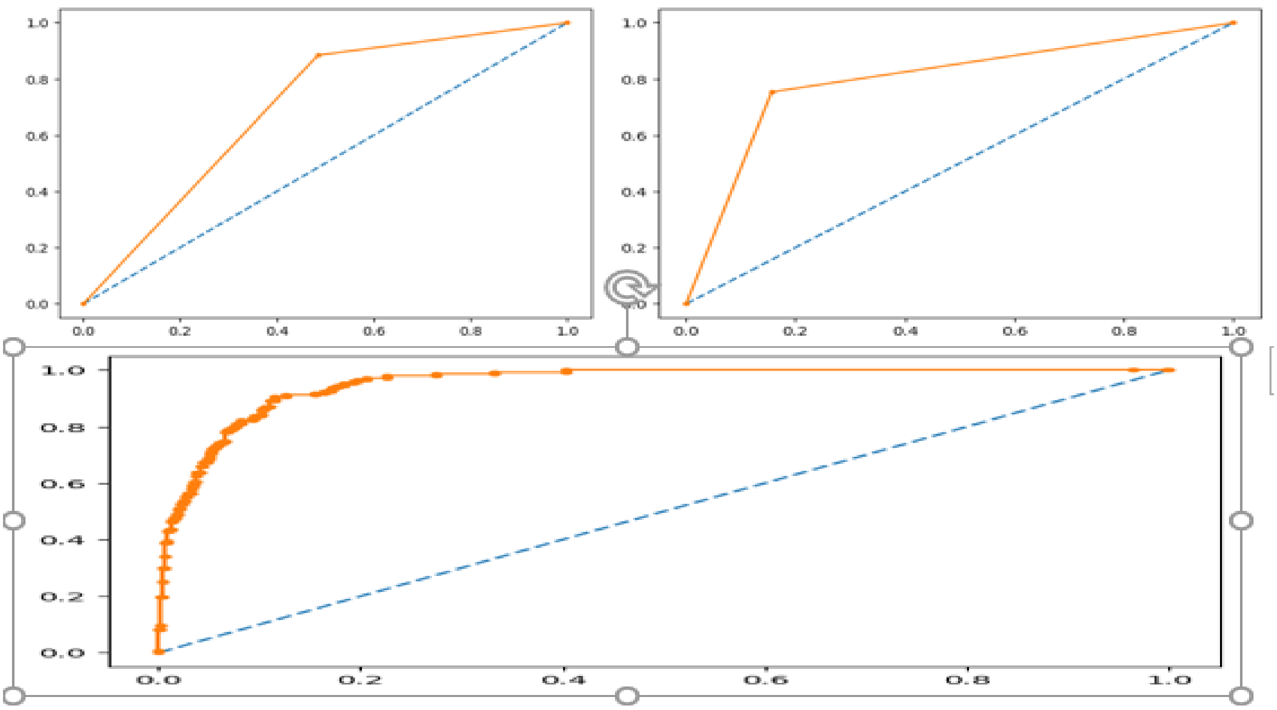
To compare the performances of Logistic Regression, Random Forest, and LDA models, we can analyze their precision, recall, F1-score, accuracy, and also plot their ROC curves. Let's summarize the performance metrics first:

➢ Logistic Regression:
- Accuracy: 93%
- Precision (Class 1): 81%
- Recall (Class 1): 45%
- F1-score (Class 1): 58%

➢ Random Forest:
- Accuracy: 93%
- Precision (Class 1): 81%
- Recall (Class 1): 45%
- F1-score (Class 1): 58%

➢ LDA:
- Accuracy: 92%
- Precision (Class 1): 70%
- Recall (Class 1): 38%
- F1-score (Class 1): 50%

<u>Now, let's compare their ROC curves</u>:



Conclusions and Recommendations:

Model Performance:

- Logistic Regression and Random Forest models exhibit similar performance in terms of accuracy, precision, recall, and F1-score. Both models achieve an accuracy of approximately 93%, with a precision of 81% and recall of 45% for predicting defaults (Class 1).
- LDA, while achieving a slightly lower accuracy of 92%, demonstrates lower precision (70%) and recall (38%) for predicting defaults compared to Logistic Regression and Random Forest.
- ROC Curve Analysis:
- The ROC curves illustrate the trade-off between true positive rate (sensitivity) and false positive rate (1-specificity) for each model.
- Logistic Regression and Random Forest models have similar ROC curves, indicating comparable performance in distinguishing between default and non-default cases.
- LDA exhibits a lower ROC curve, suggesting relatively weaker discrimination ability compared to Logistic Regression and Random Forest.
- Recommendations:
- Model Selection: Given the similar performance of Logistic Regression and Random Forest in terms of accuracy and other metrics, either of these models could be considered for predicting defaults. However, Logistic Regression might be preferred due to its interpretability and simplicity.
- Further Investigation: It's essential to delve deeper into the features and underlying assumptions of the models to understand the reasons for differences in performance. For instance, identifying influential features in Logistic Regression and Random Forest could provide insights into factors contributing to defaults.
- Improvement Strategies: Since all models demonstrate limited recall for predicting defaults (Class 1), further model tuning or feature engineering may be required to enhance the models' sensitivity to default cases. Additionally, exploring ensemble techniques or model stacking methods could potentially improve predictive performance.
- Validation and Monitoring: Regular validation of the models on new data and continuous monitoring of model performance are crucial to ensure their reliability and effectiveness in real-world scenarios.

In summary, while Logistic Regression and Random Forest models show promising performance for predicting defaults, ongoing refinement and validation are necessary to optimize their utility in decision-making processes within the context of credit risk assessment.