# FRA Graded Project

Name : Soumalya Sen
PGP-DSBA (PGPDSBA.O.JULY23.A)
Date: 26/05/2024

# Index

# PART A: Define the problem and perform Exploratory Data Analysis

## 1. Problem definition

- Executive Summary:
  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.

- Introduction:
  A renowned credit rating organization wants 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, the organization aims to analyze historical financial statements and extract pertinent insights to facilitate informed decision-making via the tool. Specifically, the organization foresees 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.

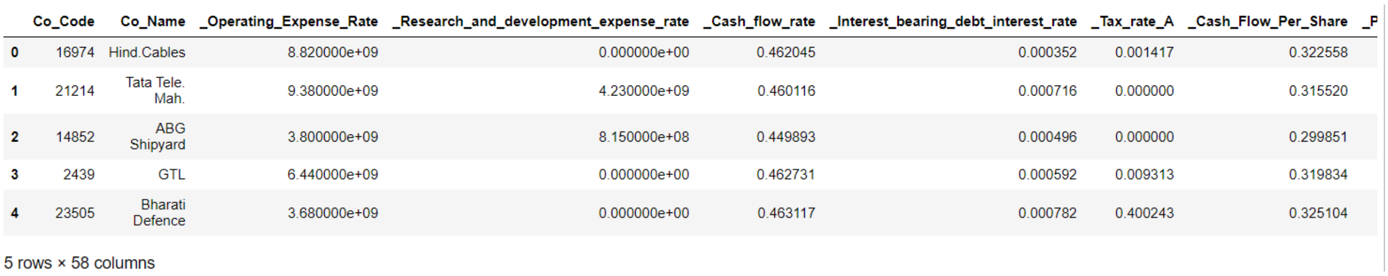
2. Credit Risk Evaluation: Evaluate credit risk exposure by analyzing liquidity ratios, debt-to-equity ratios, and other key financial indicators to ascertain the likelihood of default and inform investment decisions.

As a part of the data science team in the organization, you have been provided 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 default on its debt repayments in the next two quarters. The predictive model will help the organization anticipate potential challenges with the financial performance of the companies and enable proactive risk mitigation strategies.

- Sample of the dataset:

## Data Overview

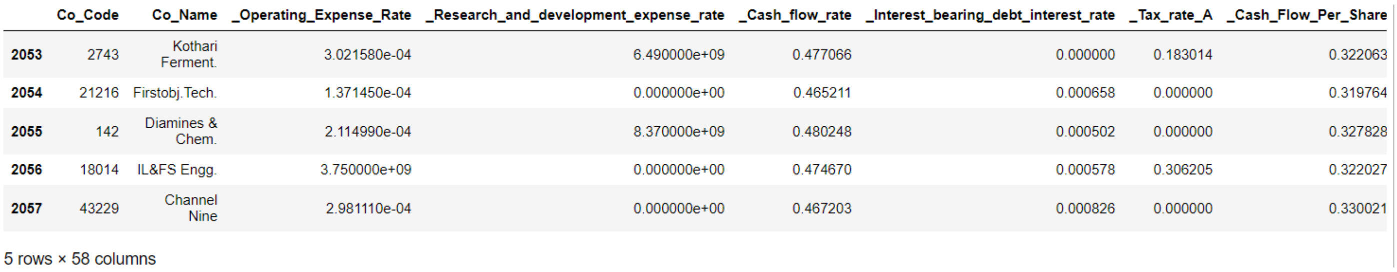
### Head of the dataset



| | Co_Code | Co_Name | _Operating_Expense_Rate | _Research_and_development_expense_rate | _Cash_flow_rate | _Interest_bearing_debt_interest_rate | _Tax_rate_A | _Cash_Flow_Per_Share | _P |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 16974 | Hind.Cables | 8.820000e+09 | 0.000000e+00 | 0.462045 | 0.000352 | 0.001417 | 0.322558 | |
| 1 | 21214 | Tata Tele. Mah. | 9.380000e+09 | 4.230000e+09 | 0.460116 | 0.000716 | 0.000000 | 0.315520 | |
| 2 | 14852 | ABG Shipyard | 3.800000e+09 | 8.150000e+08 | 0.449893 | 0.000496 | 0.000000 | 0.299851 | |
| 3 | 2439 | GTL | 6.440000e+09 | 0.000000e+00 | 0.462731 | 0.000592 | 0.009313 | 0.319834 | |
| 4 | 23505 | Bharati Defence | 3.680000e+09 | 0.000000e+00 | 0.463117 | 0.000782 | 0.400243 | 0.325104 | |

5 rows × 58 columns

### Tail of the dataset

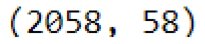


| | Co_Code | Co_Name | _Operating_Expense_Rate | _Research_and_development_expense_rate | _Cash_flow_rate | _Interest_bearing_debt_interest_rate | _Tax_rate_A | _Cash_Flow_Per_Share |
|---|---|---|---|---|---|---|---|---|
| 2053 | 2743 | Kothari Ferment. | 3.021580e-04 | 6.490000e+09 | 0.477066 | 0.000000 | 0.183014 | 0.322063 |
| 2054 | 21216 | Firstobj.Tech. | 1.371450e-04 | 0.000000e+00 | 0.465211 | 0.000658 | 0.000000 | 0.319764 |
| 2055 | 142 | Diamines & Chem. | 2.114990e-04 | 8.370000e+09 | 0.480248 | 0.000502 | 0.000000 | 0.327828 |
| 2056 | 18014 | IL&FS Engg. | 3.750000e+09 | 0.000000e+00 | 0.474670 | 0.000578 | 0.306205 | 0.322027 |
| 2057 | 43229 | Channel Nine | 2.981110e-04 | 0.000000e+00 | 0.467203 | 0.000826 | 0.000000 | 0.330021 |

5 rows × 58 columns

## 2. Check shape, Data types, and Statistical summary

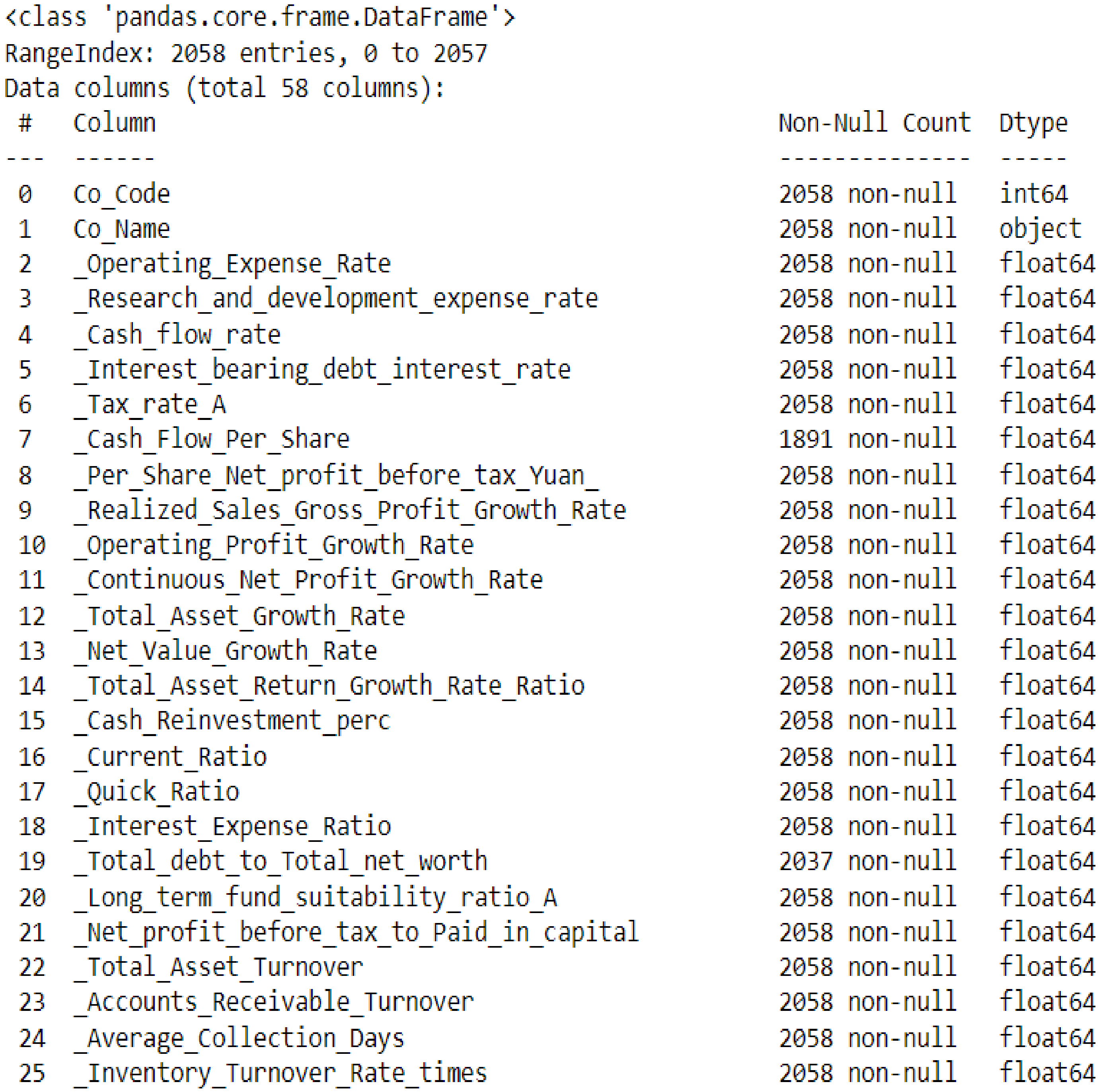
Check shape:



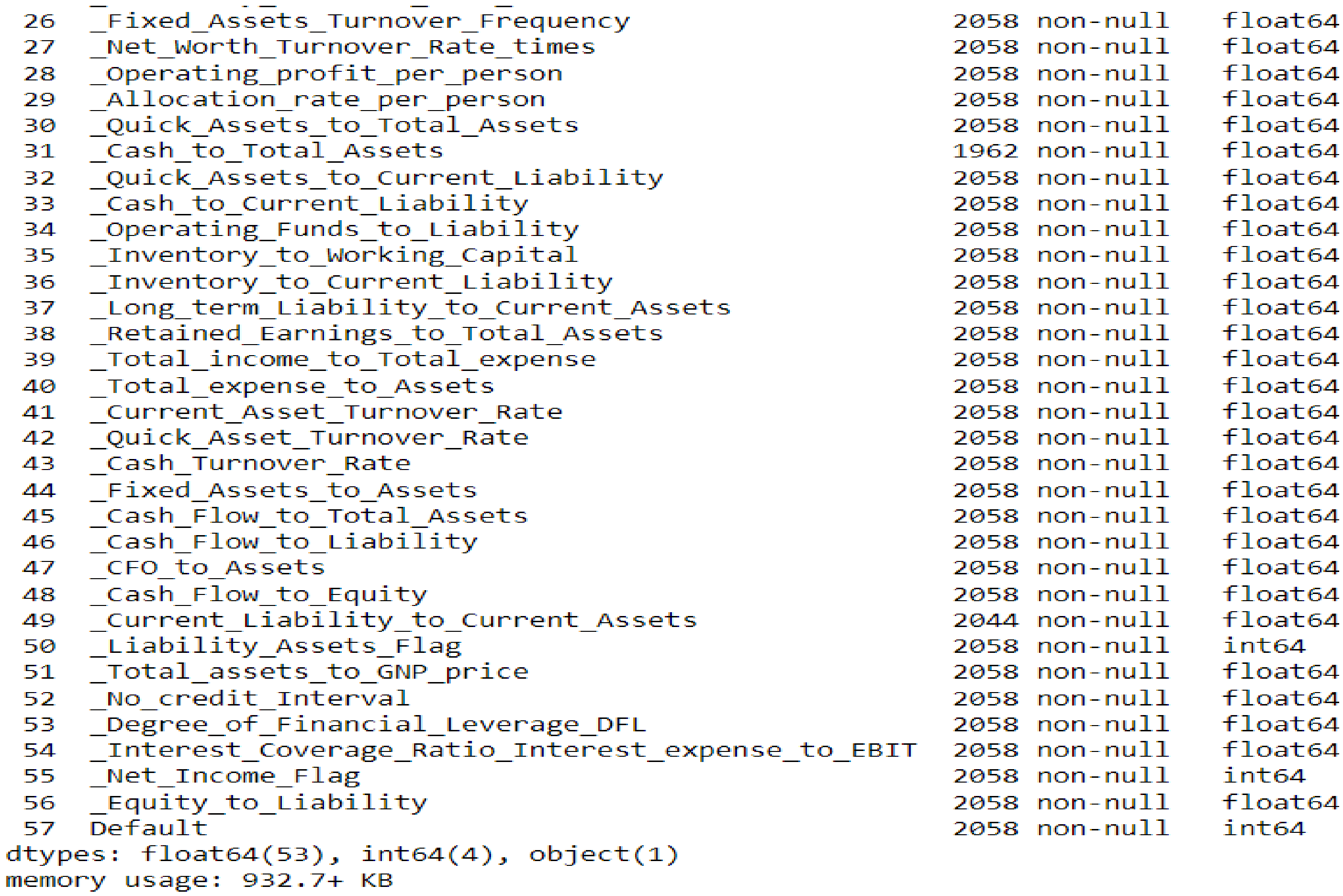
```
(2058, 58)
```

The dataset have 2058 rows and 58 columns.

Data types:



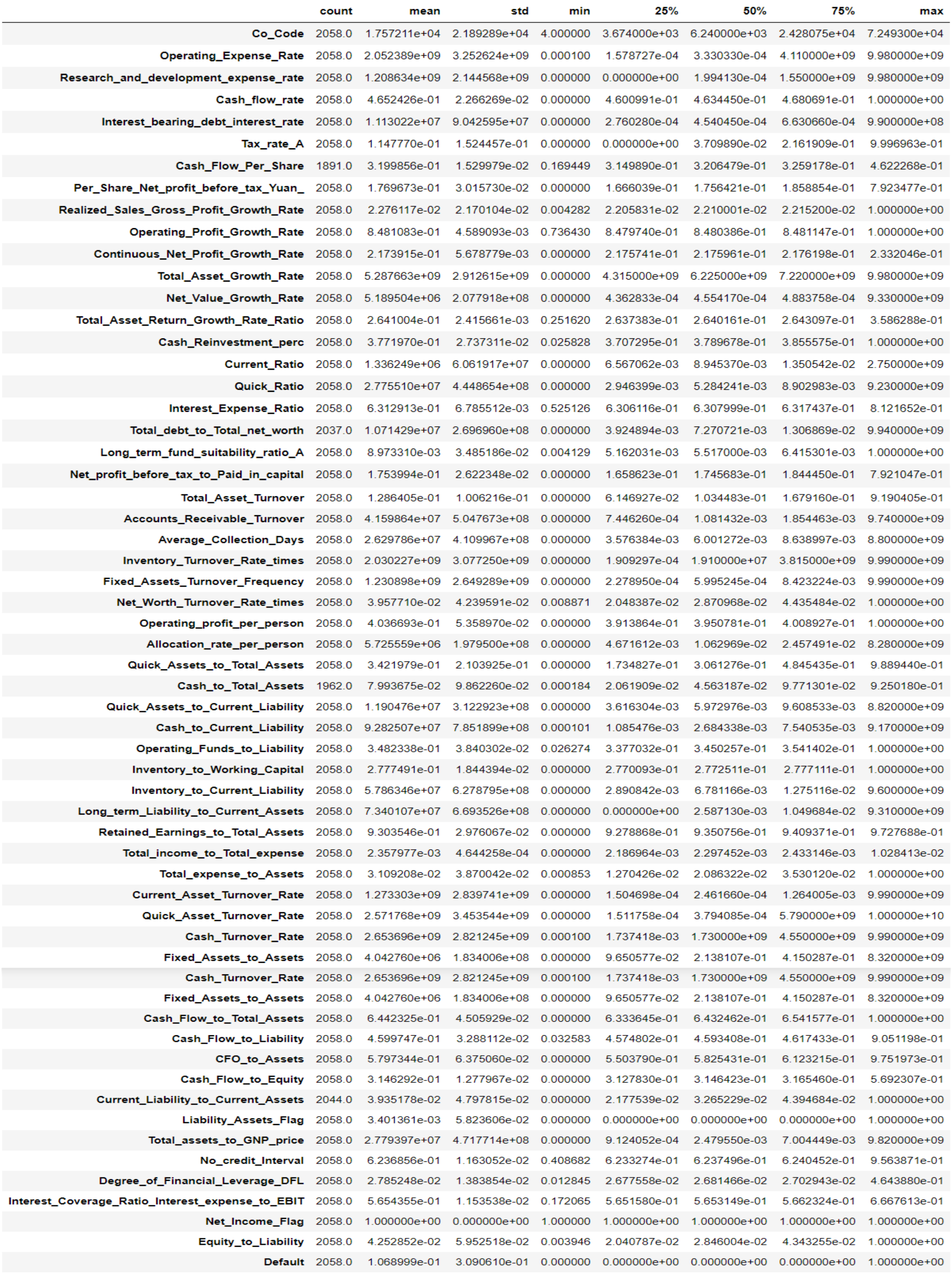
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2058 entries, 0 to 2057
Data columns (total 58 columns):
 #   Column                                   Non-Null Count  Dtype  
---  ------                                   --------------  -----  
 0   Co_Code                                  2058 non-null   int64  
 1   Co_Name                                  2058 non-null   object 
 2   _Operating_Expense_Rate                  2058 non-null   float64
 3   _Research_and_development_expense_rate   2058 non-null   float64
 4   _Cash_flow_rate                          2058 non-null   float64
 5   _Interest_bearing_debt_interest_rate     2058 non-null   float64
 6   _Tax_rate_A                              2058 non-null   float64
 7   _Cash_Flow_Per_Share                     1891 non-null   float64
 8   _Per_Share_Net_profit_before_tax_Yuan_   2058 non-null   float64
 9   _Realized_Sales_Gross_Profit_Growth_Rate 2058 non-null   float64
 10  _Operating_Profit_Growth_Rate            2058 non-null   float64
 11  _Continuous_Net_Profit_Growth_Rate       2058 non-null   float64
 12  _Total_Asset_Growth_Rate                 2058 non-null   float64
 13  _Net_Value_Growth_Rate                   2058 non-null   float64
 14  _Total_Asset_Return_Growth_Rate_Ratio    2058 non-null   float64
 15  _Cash_Reinvestment_perc                  2058 non-null   float64
 16  _Current_Ratio                           2058 non-null   float64
 17  _Quick_Ratio                             2058 non-null   float64
 18  _Interest_Expense_Ratio                  2058 non-null   float64
 19  _Total_debt_to_Total_net_worth           2037 non-null   float64
 20  _Long_term_fund_suitability_ratio_A      2058 non-null   float64
 21  _Net_profit_before_tax_to_Paid_in_capital 2058 non-null   float64
 22  _Total_Asset_Turnover                    2058 non-null   float64
 23  _Accounts_Receivable_Turnover            2058 non-null   float64
 24  _Average_Collection_Days                 2058 non-null   float64
 25  _Inventory_Turnover_Rate_times           2058 non-null   float64
```



```
 26   _Fixed_Assets_Turnover_Frequency                     2058 non-null   float64
 27   _Net_Worth_Turnover_Rate_times                       2058 non-null   float64
 28   _Operating_profit_per_person                         2058 non-null   float64
 29   _Allocation_rate_per_person                          2058 non-null   float64
 30   _Quick_Assets_to_Total_Assets                        2058 non-null   float64
 31   _Cash_to_Total_Assets                                1962 non-null   float64
 32   _Quick_Assets_to_Current_Liability                   2058 non-null   float64
 33   _Cash_to_Current_Liability                           2058 non-null   float64
 34   _Operating_Funds_to_Liability                        2058 non-null   float64
 35   _Inventory_to_Working_Capital                        2058 non-null   float64
 36   _Inventory_to_Current_Liability                      2058 non-null   float64
 37   _Long_term_Liability_to_Current_Assets               2058 non-null   float64
 38   _Retained_Earnings_to_Total_Assets                   2058 non-null   float64
 39   _Total_income_to_Total_expense                       2058 non-null   float64
 40   _Total_expense_to_Assets                             2058 non-null   float64
 41   _Current_Asset_Turnover_Rate                         2058 non-null   float64
 42   _Quick_Asset_Turnover_Rate                           2058 non-null   float64
 43   _Cash_Turnover_Rate                                  2058 non-null   float64
 44   _Fixed_Assets_to_Assets                              2058 non-null   float64
 45   _Cash_Flow_to_Total_Assets                           2058 non-null   float64
 46   _Cash_Flow_to_Liability                              2058 non-null   float64
 47   _CFO_to_Assets                                       2058 non-null   float64
 48   _Cash_Flow_to_Equity                                 2058 non-null   float64
 49   _Current_Liability_to_Current_Assets                 2044 non-null   float64
 50   _Liability_Assets_Flag                               2058 non-null   int64
 51   _Total_assets_to_GNP_price                           2058 non-null   float64
 52   _No_credit_Interval                                  2058 non-null   float64
 53   _Degree_of_Financial_Leverage_DFL                    2058 non-null   float64
 54   _Interest_Coverage_Ratio_Interest_expense_to_EBIT    2058 non-null   float64
 55   _Net_Income_Flag                                     2058 non-null   int64
 56   _Equity_to_Liability                                 2058 non-null   float64
 57   Default                                              2058 non-null   int64
dtypes: float64(53), int64(4), object(1)
memory usage: 932.7+ KB
```

- The data is a pandas DataFrame with 2058 observations and 58 columns.
- The columns represent various financial metrics for each observation, such as operating expense rate, tax rate, cash flow per share, and net profit before tax.
- The data types of the columns are a mix of float64, int64, and object. The object data type typically represents categorical variables or strings.
- There are no missing values in most of the columns, but some columns have missing values, such as _Cash_Flow_Per_Share (1891 non-null), _Total_debt_to_Total_net_worth (2037 non-null), and _Cash_to_Total_Assets (1962 non-null).
- The last column, Default, is an int64 data type, which may represent whether the company has defaulted on its debt or not.

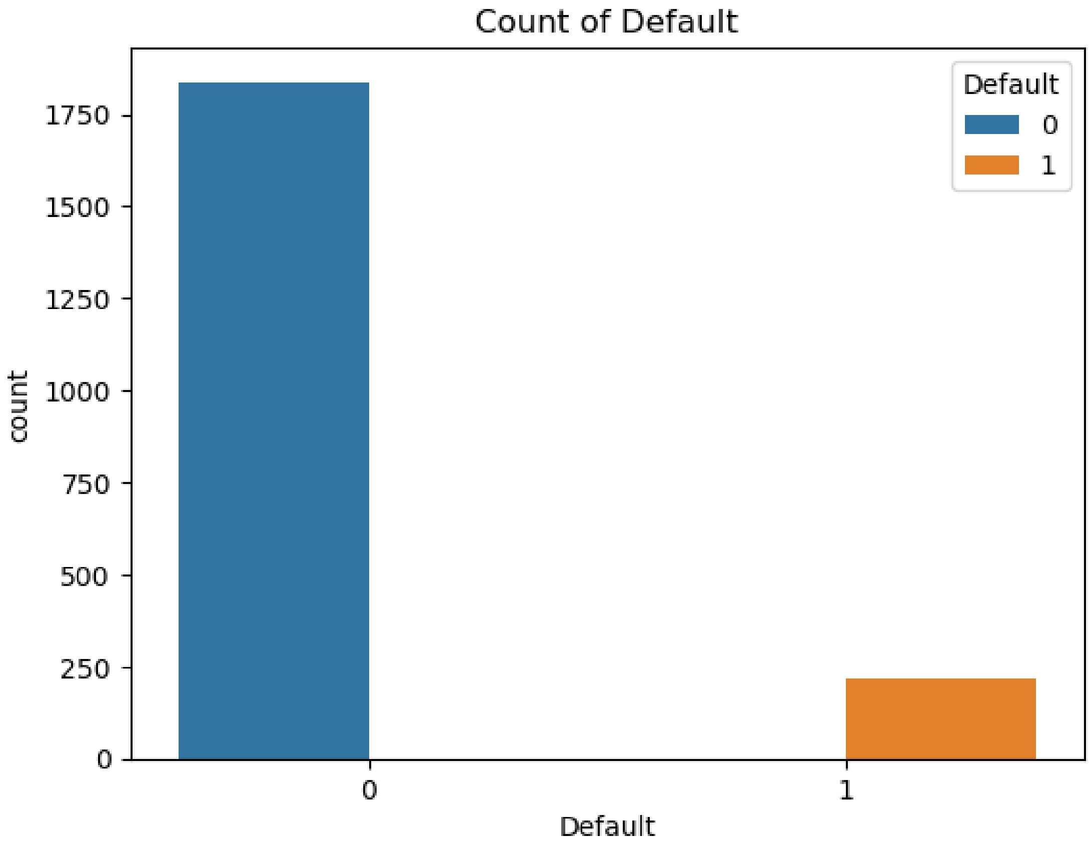
# Statistical summary:



| | count | mean | std | min | 25% | 50% | 75% | max |
|---:|---:|---:|---:|---:|---:|---:|---:|---:|
| **Co_Code** | 2058.0 | 1.757211e+04 | 2.189289e+04 | 4.000000 | 3.674000e+03 | 6.240000e+03 | 2.428075e+04 | 7.249300e+04 |
| **Operating_Expense_Rate** | 2058.0 | 2.052389e+09 | 3.252624e+09 | 0.000100 | 1.578727e-04 | 3.330330e-04 | 4.110000e+09 | 9.980000e+09 |
| **Research_and_development_expense_rate** | 2058.0 | 1.208634e+09 | 2.144568e+09 | 0.000000 | 0.000000e+00 | 1.994130e-04 | 1.550000e+09 | 9.980000e+09 |
| **Cash_flow_rate** | 2058.0 | 4.652426e-01 | 2.266269e-02 | 0.000000 | 4.600991e-01 | 4.634450e-01 | 4.680691e-01 | 1.000000e+00 |
| **Interest_bearing_debt_interest_rate** | 2058.0 | 1.113022e+07 | 9.042595e+07 | 0.000000 | 2.760280e-04 | 4.540450e-04 | 6.630660e-04 | 9.900000e+08 |
| **Tax_rate_A** | 2058.0 | 1.147770e-01 | 1.524457e-01 | 0.000000 | 0.000000e+00 | 3.709890e-02 | 2.161909e-01 | 9.996963e-01 |
| **Cash_Flow_Per_Share** | 1891.0 | 3.199856e-01 | 1.529979e-02 | 0.169449 | 3.149890e-01 | 3.206479e-01 | 3.259178e-01 | 4.622268e-01 |
| **Per_Share_Net_profit_before_tax_Yuan_** | 2058.0 | 1.769673e-01 | 3.015730e-02 | 0.000000 | 1.666039e-01 | 1.756421e-01 | 1.858854e-01 | 7.923477e-01 |
| **Realized_Sales_Gross_Profit_Growth_Rate** | 2058.0 | 2.276117e-02 | 2.170104e-02 | 0.004282 | 2.205831e-02 | 2.210001e-02 | 2.215200e-02 | 1.000000e+00 |
| **Operating_Profit_Growth_Rate** | 2058.0 | 8.481083e-01 | 4.589093e-03 | 0.736430 | 8.479740e-01 | 8.480386e-01 | 8.481147e-01 | 1.000000e+00 |
| **Continuous_Net_Profit_Growth_Rate** | 2058.0 | 2.173915e-01 | 5.678779e-03 | 0.000000 | 2.175741e-01 | 2.175961e-01 | 2.176198e-01 | 2.332046e-01 |
| **Total_Asset_Growth_Rate** | 2058.0 | 5.287663e+09 | 2.912615e+09 | 0.000000 | 4.315000e+09 | 6.225000e+09 | 7.220000e+09 | 9.980000e+09 |
| **Net_Value_Growth_Rate** | 2058.0 | 5.189504e+06 | 2.077918e+08 | 0.000000 | 4.362833e-04 | 4.554170e-04 | 4.883758e-04 | 9.330000e+09 |
| **Total_Asset_Return_Growth_Rate_Ratio** | 2058.0 | 2.641004e-01 | 2.415661e-03 | 0.251620 | 2.637383e-01 | 2.640161e-01 | 2.643097e-01 | 3.586288e-01 |
| **Cash_Reinvestment_perc** | 2058.0 | 3.771970e-01 | 2.737311e-02 | 0.025828 | 3.707295e-01 | 3.789678e-01 | 3.855575e-01 | 1.000000e+00 |
| **Current_Ratio** | 2058.0 | 1.336249e+06 | 6.061917e+07 | 0.000000 | 6.567062e-03 | 8.945370e-03 | 1.350542e-02 | 2.750000e+09 |
| **Quick_Ratio** | 2058.0 | 2.775510e+07 | 4.448654e+08 | 0.000000 | 2.946399e-03 | 5.284241e-03 | 8.902983e-03 | 9.230000e+09 |
| **Interest_Expense_Ratio** | 2058.0 | 6.312913e-01 | 6.785512e-03 | 0.525126 | 6.306116e-01 | 6.307999e-01 | 6.317437e-01 | 8.121652e-01 |
| **Total_debt_to_Total_net_worth** | 2037.0 | 1.071429e+07 | 2.696960e+08 | 0.000000 | 3.924894e-03 | 7.270721e-03 | 1.306869e-02 | 9.940000e+09 |
| **Long_term_fund_suitability_ratio_A** | 2058.0 | 8.973310e-03 | 3.485186e-02 | 0.004129 | 5.162031e-03 | 5.517000e-03 | 6.415301e-03 | 1.000000e+00 |
| **Net_profit_before_tax_to_Paid_in_capital** | 2058.0 | 1.753994e-01 | 2.622348e-02 | 0.000000 | 1.658623e-01 | 1.745683e-01 | 1.844450e-01 | 7.921047e-01 |
| **Total_Asset_Turnover** | 2058.0 | 1.286405e-01 | 1.006216e-01 | 0.000000 | 6.146927e-02 | 1.034483e-01 | 1.679160e-01 | 9.190405e-01 |
| **Accounts_Receivable_Turnover** | 2058.0 | 4.159864e+07 | 5.047673e+08 | 0.000000 | 7.446260e-04 | 1.081432e-03 | 1.854463e-03 | 9.740000e+09 |
| **Average_Collection_Days** | 2058.0 | 2.629786e+07 | 4.109967e+08 | 0.000000 | 3.576384e-03 | 6.001272e-03 | 8.638997e-03 | 8.800000e+09 |
| **Inventory_Turnover_Rate_times** | 2058.0 | 2.030227e+09 | 3.077250e+09 | 0.000000 | 1.909297e-04 | 1.910000e+07 | 3.815000e+09 | 9.990000e+09 |
| **Fixed_Assets_Turnover_Frequency** | 2058.0 | 1.230898e+09 | 2.649289e+09 | 0.000000 | 2.278950e-04 | 5.995245e-04 | 8.423224e-03 | 9.990000e+09 |
| **Net_Worth_Turnover_Rate_times** | 2058.0 | 3.957710e-02 | 4.239591e-02 | 0.008871 | 2.048387e-02 | 2.870968e-02 | 4.435484e-02 | 1.000000e+00 |
| **Operating_profit_per_person** | 2058.0 | 4.036693e-01 | 5.358970e-02 | 0.000000 | 3.913864e-01 | 3.950781e-01 | 4.008927e-01 | 1.000000e+00 |
| **Allocation_rate_per_person** | 2058.0 | 5.725559e+06 | 1.979500e+08 | 0.000000 | 4.671612e-03 | 1.062969e-02 | 2.457491e-02 | 8.280000e+09 |
| **Quick_Assets_to_Total_Assets** | 2058.0 | 3.421979e-01 | 2.103925e-01 | 0.000000 | 1.734827e-01 | 3.061276e-01 | 4.845435e-01 | 9.889440e-01 |
| **Cash_to_Total_Assets** | 1962.0 | 7.993675e-02 | 9.862260e-02 | 0.000184 | 2.061909e-02 | 4.563187e-02 | 9.771301e-02 | 9.250180e-01 |
| **Quick_Assets_to_Current_Liability** | 2058.0 | 1.190476e+07 | 3.122923e+08 | 0.000000 | 3.616304e-03 | 5.972976e-03 | 9.608533e-03 | 8.820000e+09 |
| **Cash_to_Current_Liability** | 2058.0 | 9.282507e+07 | 7.851899e+08 | 0.000101 | 1.085476e-03 | 2.684338e-03 | 7.540535e-03 | 9.170000e+09 |
| **Operating_Funds_to_Liability** | 2058.0 | 3.482338e-01 | 3.840302e-02 | 0.026274 | 3.377032e-01 | 3.450257e-01 | 3.541402e-01 | 1.000000e+00 |
| **Inventory_to_Working_Capital** | 2058.0 | 2.777491e-01 | 1.844394e-02 | 0.000000 | 2.770093e-01 | 2.772511e-01 | 2.777111e-01 | 1.000000e+00 |
| **Inventory_to_Current_Liability** | 2058.0 | 5.786346e+07 | 6.278795e+08 | 0.000000 | 2.890842e-03 | 6.781166e-03 | 1.275116e-02 | 9.600000e+09 |
| **Long_term_Liability_to_Current_Assets** | 2058.0 | 7.340107e+07 | 6.693526e+08 | 0.000000 | 0.000000e+00 | 2.587130e-03 | 1.049684e-02 | 9.310000e+09 |
| **Retained_Earnings_to_Total_Assets** | 2058.0 | 9.303546e-01 | 2.976067e-02 | 0.000000 | 9.278868e-01 | 9.350756e-01 | 9.409371e-01 | 9.727688e-01 |
| **Total_income_to_Total_expense** | 2058.0 | 2.357977e-03 | 4.644258e-04 | 0.000000 | 2.186964e-03 | 2.297452e-03 | 2.433146e-03 | 1.028413e-02 |
| **Total_expense_to_Assets** | 2058.0 | 3.109208e-02 | 3.870042e-02 | 0.000853 | 1.270426e-02 | 2.086322e-02 | 3.530120e-02 | 1.000000e+00 |
| **Current_Asset_Turnover_Rate** | 2058.0 | 1.273303e+09 | 2.839741e+09 | 0.000000 | 1.504698e-04 | 2.461660e-04 | 1.264005e-03 | 9.990000e+09 |
| **Quick_Asset_Turnover_Rate** | 2058.0 | 2.571768e+09 | 3.453544e+09 | 0.000000 | 1.511758e-04 | 3.794085e-04 | 5.790000e+09 | 1.000000e+10 |
| **Cash_Turnover_Rate** | 2058.0 | 2.653696e+09 | 2.821245e+09 | 0.000100 | 1.737418e-03 | 1.730000e+09 | 4.550000e+09 | 9.990000e+09 |
| **Fixed_Assets_to_Assets** | 2058.0 | 4.042760e+06 | 1.834006e+08 | 0.000000 | 9.650577e-02 | 2.138107e-01 | 4.150287e-01 | 8.320000e+09 |
| **Cash_Turnover_Rate** | 2058.0 | 2.653696e+09 | 2.821245e+09 | 0.000100 | 1.737418e-03 | 1.730000e+09 | 4.550000e+09 | 9.990000e+09 |
| **Fixed_Assets_to_Assets** | 2058.0 | 4.042760e+06 | 1.834006e+08 | 0.000000 | 9.650577e-02 | 2.138107e-01 | 4.150287e-01 | 8.320000e+09 |
| **Cash_Flow_to_Total_Assets** | 2058.0 | 6.442325e-01 | 4.505929e-02 | 0.000000 | 6.333645e-01 | 6.432462e-01 | 6.541577e-01 | 1.000000e+00 |
| **Cash_Flow_to_Liability** | 2058.0 | 4.599747e-01 | 3.288112e-02 | 0.032583 | 4.574802e-01 | 4.593408e-01 | 4.617433e-01 | 9.051198e-01 |
| **CFO_to_Assets** | 2058.0 | 5.797344e-01 | 6.375060e-02 | 0.000000 | 5.503790e-01 | 5.825431e-01 | 6.123215e-01 | 9.751973e-01 |
| **Cash_Flow_to_Equity** | 2058.0 | 3.146292e-01 | 1.277967e-02 | 0.000000 | 3.127830e-01 | 3.146423e-01 | 3.165460e-01 | 5.692307e-01 |
| **Current_Liability_to_Current_Assets** | 2044.0 | 3.935178e-02 | 4.797815e-02 | 0.000000 | 2.177539e-02 | 3.265229e-02 | 4.394684e-02 | 1.000000e+00 |
| **Liability_Assets_Flag** | 2058.0 | 3.401361e-03 | 5.823606e-02 | 0.000000 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 1.000000e+00 |
| **Total_assets_to_GNP_price** | 2058.0 | 2.779397e+07 | 4.717714e+08 | 0.000000 | 9.124052e-04 | 2.479550e-03 | 7.004449e-03 | 9.820000e+09 |
| **No_credit_Interval** | 2058.0 | 6.236856e-01 | 1.163052e-02 | 0.408682 | 6.233274e-01 | 6.237496e-01 | 6.240452e-01 | 9.563871e-01 |
| **Degree_of_Financial_Leverage_DFL** | 2058.0 | 2.785248e-02 | 1.383854e-02 | 0.012845 | 2.677558e-02 | 2.681466e-02 | 2.702943e-02 | 4.643880e-01 |
| **Interest_Coverage_Ratio_Interest_expense_to_EBIT** | 2058.0 | 5.654355e-01 | 1.153538e-02 | 0.172065 | 5.651580e-01 | 5.653149e-01 | 5.662324e-01 | 6.667613e-01 |
| **Net_Income_Flag** | 2058.0 | 1.000000e+00 | 0.000000e+00 | 1.000000 | 1.000000e+00 | 1.000000e+00 | 1.000000e+00 | 1.000000e+00 |
| **Equity_to_Liability** | 2058.0 | 4.252852e-02 | 5.952518e-02 | 0.003946 | 2.040787e-02 | 2.846004e-02 | 4.343255e-02 | 1.000000e+00 |
| **Default** | 2058.0 | 1.068999e-01 | 3.090610e-01 | 0.000000 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 1.000000e+00 |

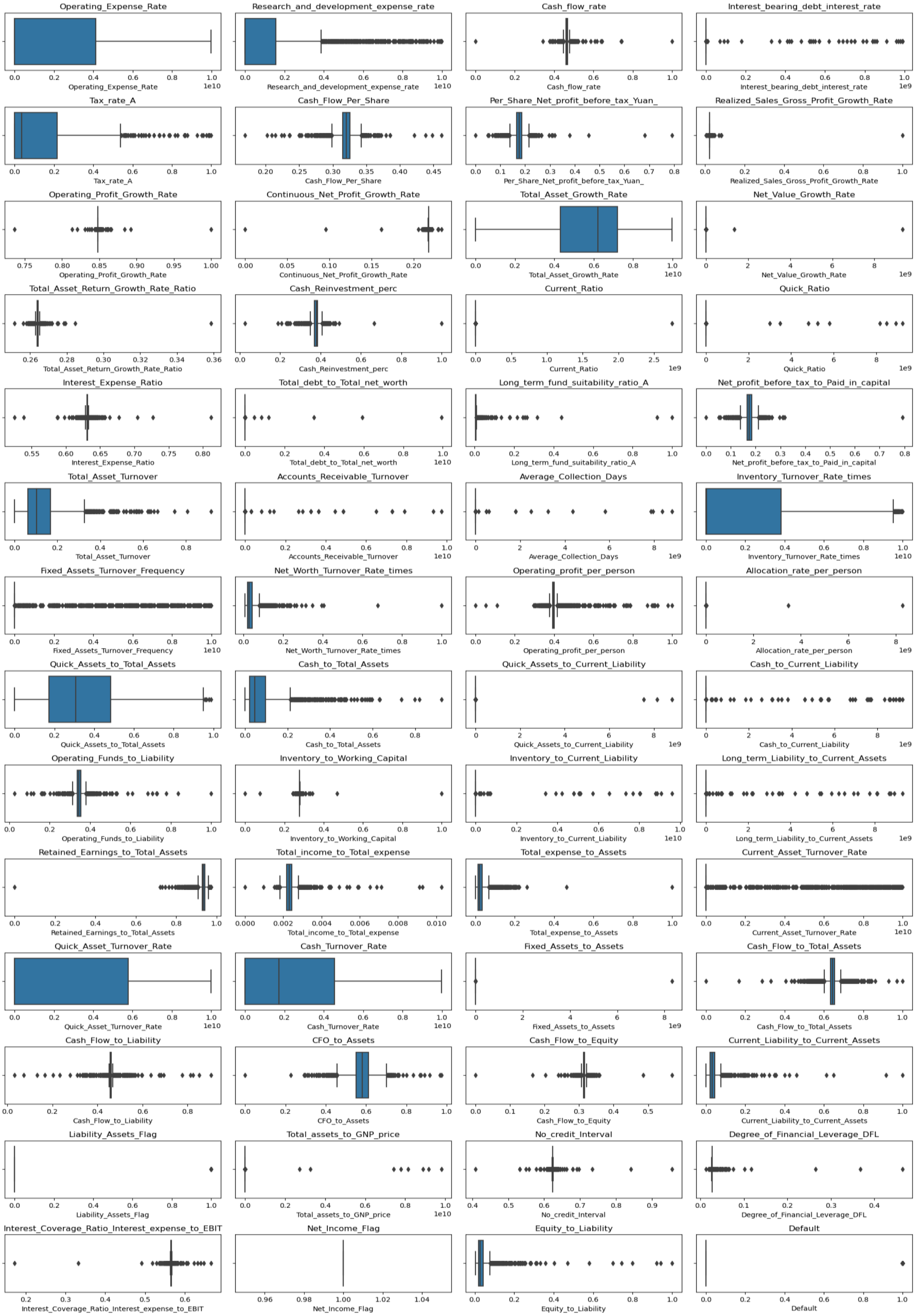
## 3. Univariate analysis

- Count plot of Default



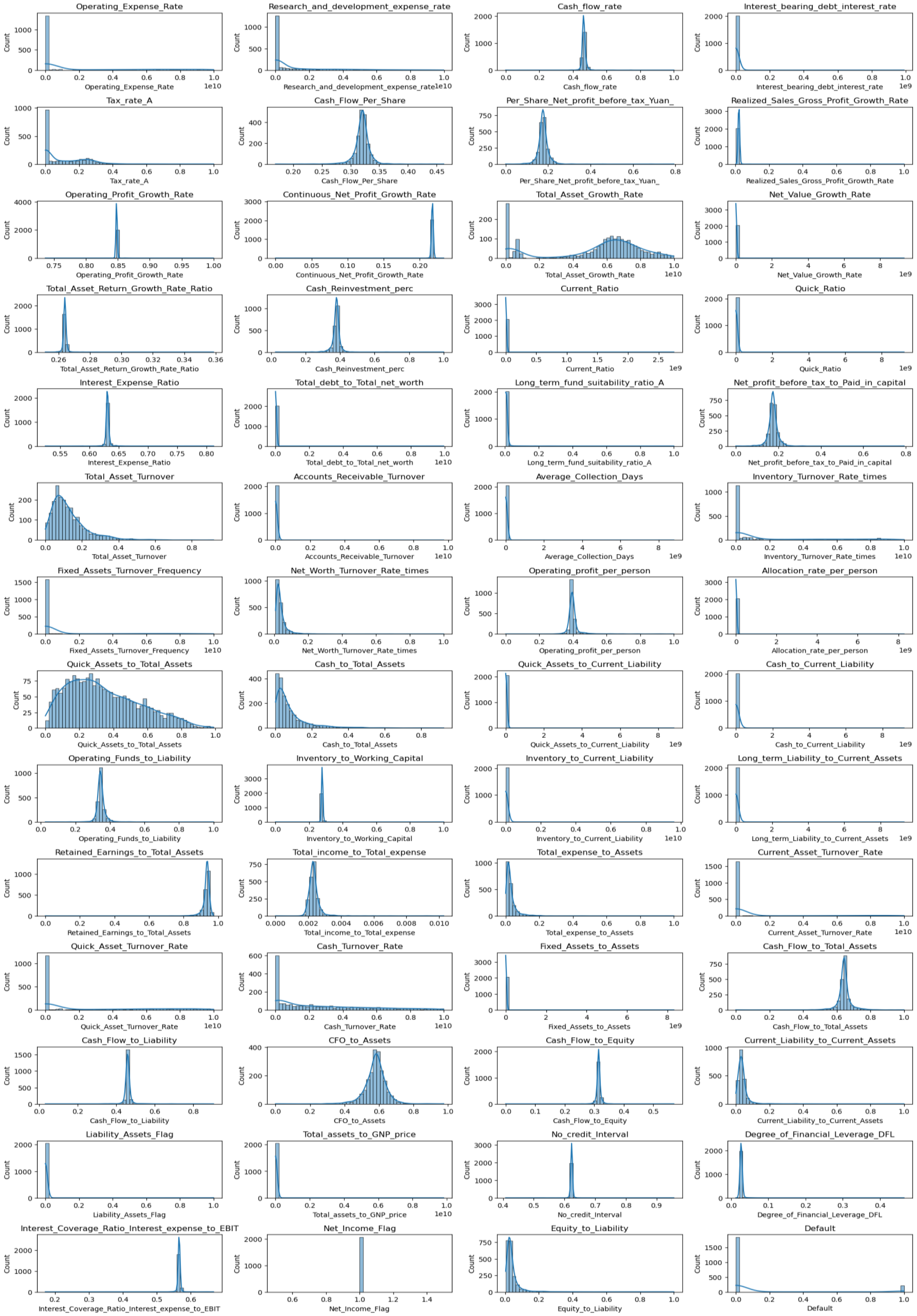
The image shows a bar graph representing the count of default. There are two bars in the graph, representing the count of '0' and '1' for the Default variable. The bar for '0' is much taller than the bar for '1', indicating that there are significantly more instances of non-default (0) compared to default (1) in the dataset.

- Boxplots for all the numerical columns



The boxplots show that the data is highly skewed and has many outliers. Many of the variables have long tails, which suggests that there are some extreme values. This is likely due to the fact that the data is from a variety of sources and may have been collected using different methods. In addition, many of the variables are measured on different scales, which can make it difficult to compare them directly.

- Histplot for all numerical columns in the data

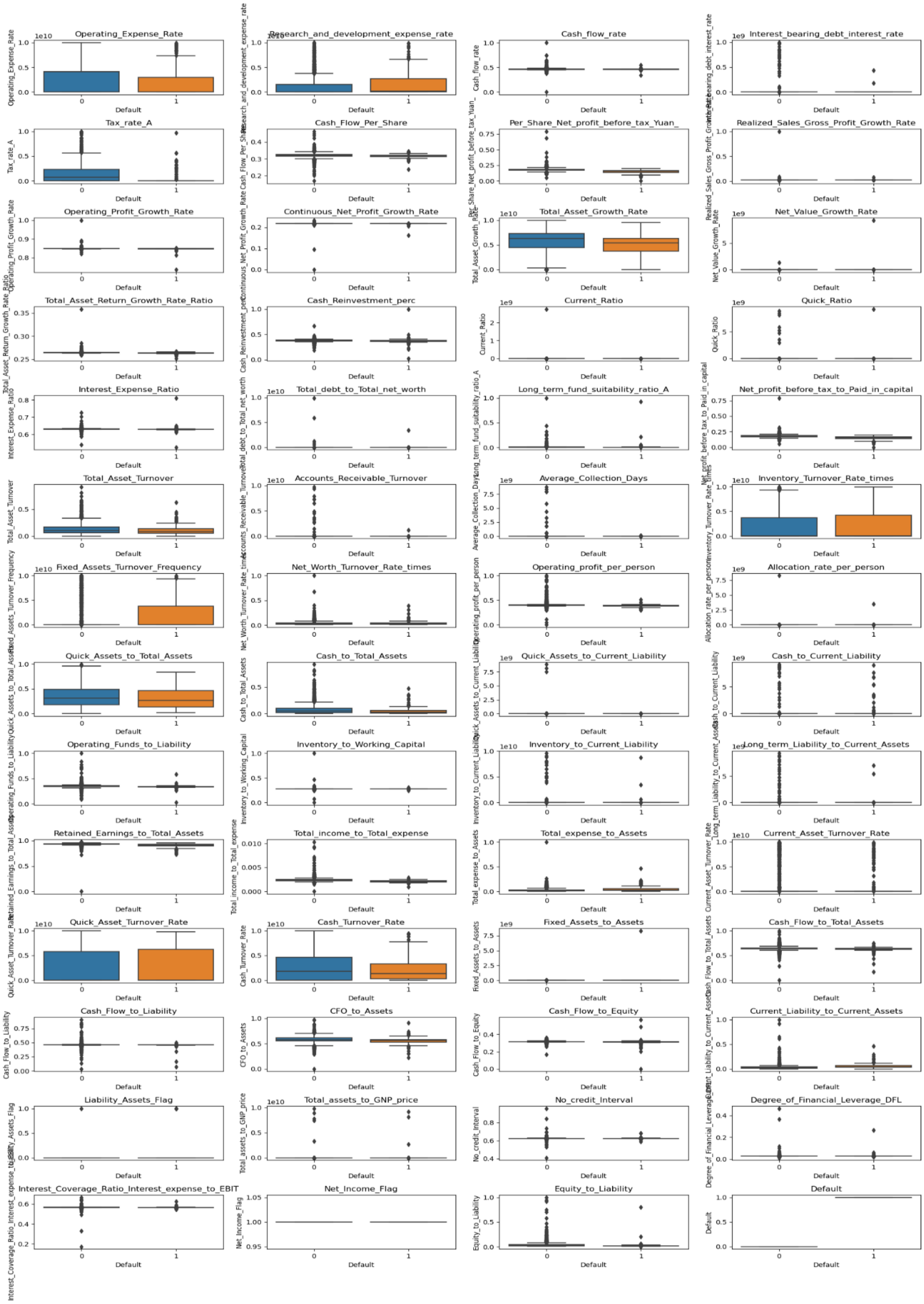


The histograms show that:

- The distribution of several financial ratios for a set of companies. These ratios are commonly used to assess the financial health and performance of a company.
- Operating expense rate is heavily skewed to the left, with a large majority of companies having an operating expense rate below 2 billion.
- Research and development expense rate is also heavily skewed to the left, with most companies having a rate below 0.2 billion.
- Cash flow rate is extremely tightly clustered around a value close to 0.4. This suggests that there is a strong tendency for cash flow rates to be relatively similar across different companies.

## 4. Bivariate Analysis

- Boxplot of all variables with Default column in the data



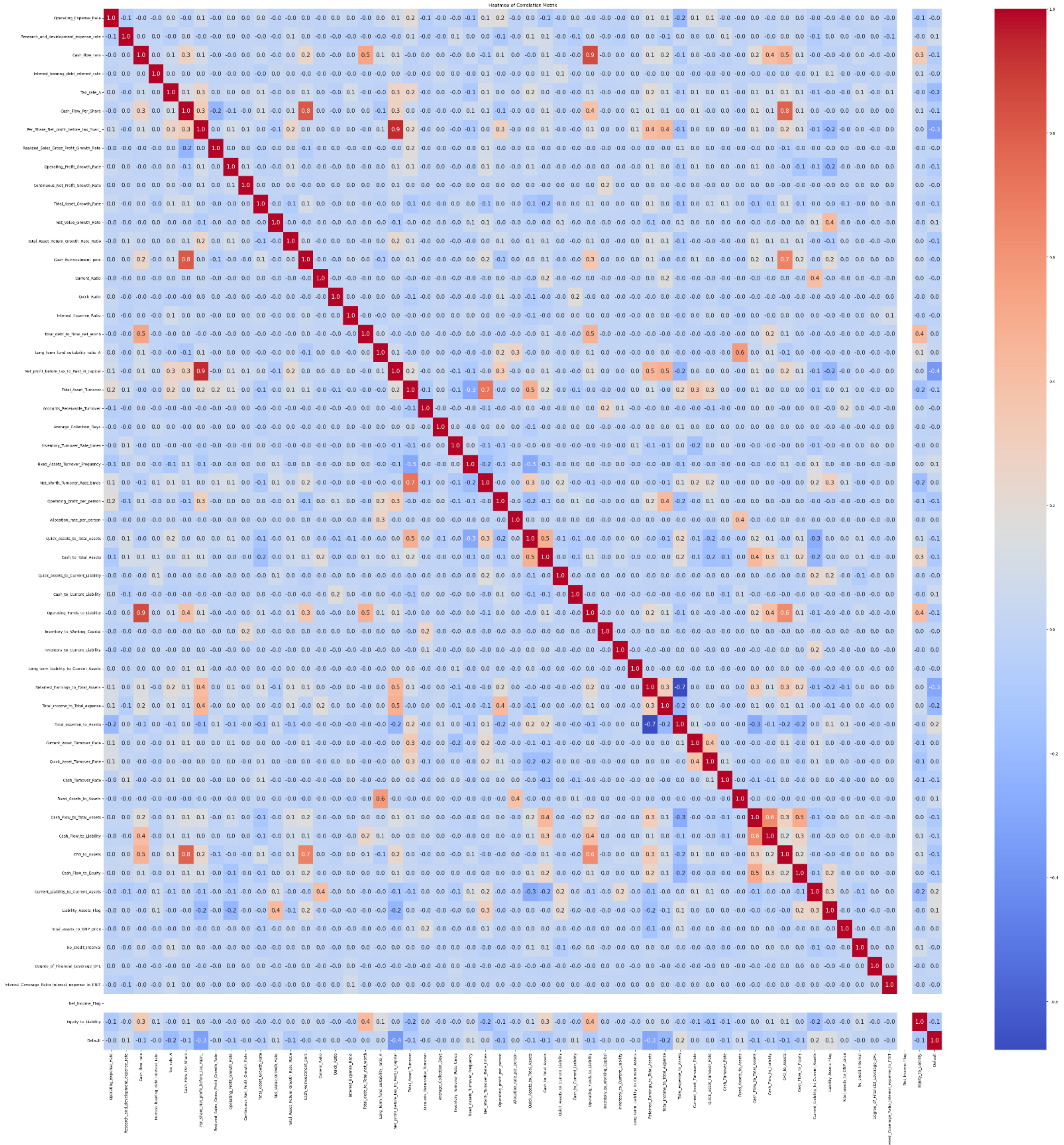
Observations:

Features that appear to be good discriminators:
Operating_Expense_Rate (high values for non-defaults)
Cash_flow_rate (high values for non-defaults)
Net_Value_Growth_Rate (high values for non-defaults)
Cash_to_Total_Assets (high values for non-defaults)
Quick_Assets_to_Total_Assets (high values for non-defaults)
Quick_Asset_Turnover_Rate (high values for non-defaults)
Retained_Earnings_to_Total_Assets (high values for non-defaults)
Current_Asset_Turnover_Rate (high values for non-defaults)

Features that appear to be less discriminators:
Total_Asset_Growth_Rate (similar distributions)
Total_debt_to_Total_net_worth (similar distributions)
Inventory_Turnover_Rate_times (similar distributions)
Total_income_to_Total_expense (similar distributions)
Total_expense_to_Assets (similar distributions)
Cash_Flow_to_Equity (similar distributions)
Net_Income_Flag (similar distributions)

Features with outlier issues:
Operating_Profit_Growth_Rate (lots of outliers)
Continuous_Net_Profit_Growth_Rate (lots of outliers)
Average_Collection_Days (lots of outliers)
Inventory_to_Working_Capital (lots of outliers)

- Heatmap of the correlation matrix



This heatmap shows the correlation between different financial metrics. The brighter the red, the higher the positive correlation, and the brighter the blue, the higher the negative correlation.

Strongest Positive Correlations:

- Cash Flow Per Share and Cash Reinvestment_perc are highly correlated (0.8). This means companies that have a high cash flow per share tend to reinvest a higher portion of their earnings.
- Total Asset Turnover and Net Worth Turnover Rate times are also strongly correlated (0.7). Companies that have a high asset turnover generally have a high net worth turnover, indicating efficient use of assets.
- Total_expense_to_Assets and Current_Asset_Turnover_Rate are positively correlated (0.4). This suggests that companies with higher total expenses relative to their assets tend to have a higher current asset turnover.
- CFO_to_Assets and Cash_Flow_to_Liability are highly positively correlated (0.8). This is expected as CFO (Cash Flow from Operations) typically drives a large portion of cash available to meet liabilities.

Strongest Negative Correlations:

- Total_Asset_Turnover and Degree_of_Financial_Leverage_DFL have a strong negative correlation (-0.3). This means companies with higher asset turnover tend to have lower financial leverage (less debt relative to equity).
- Per_Share_Net_profit_before_tax_Yuan_ and Fixed_Assets_Turnover_Frequency are negatively correlated (-0.3). Companies with high earnings per share may not necessarily have high fixed asset turnover.
- Total_income_to_Total_expense and Total_expense_to_Assets are negatively correlated (-0.7). This suggests that companies with higher total income relative to expenses have a lower proportion of total expenses compared to their assets.
- Cash_Flow_to_Equity and Default are negatively correlated (-0.4).

This suggests that companies with higher cash flow to equity may be less likely to default.

Key Observations:

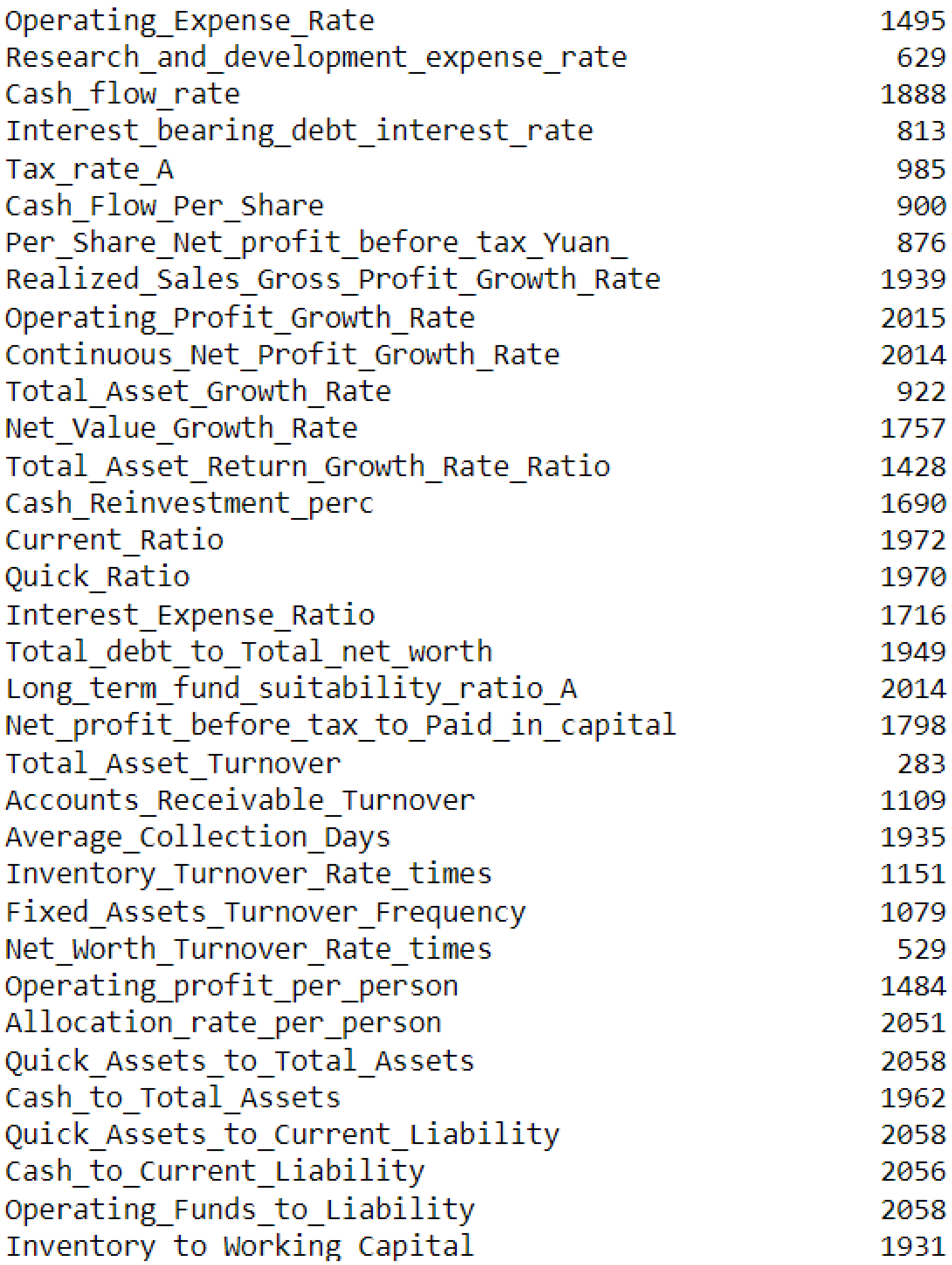
- The relationship between Total_debt_to_Total_net_worth and Interest_Expense_Ratio shows a strong positive correlation (0.5). This highlights the relationship between debt levels and interest expense.
- Operating_Funds_to_Liability has a high positive correlation (0.9) with Cash_flow_rate and Cash_Flow_to_Liability. This implies that companies with strong operating funds to liability ratios often have good cash flow and better management of their liabilities.
- Current_Liability_to_Current_Assets has a strong positive correlation (0.4) with Current_Ratio. This makes sense as higher current liabilities relative to assets often lead to a higher current ratio.

# PART A: Data Pre-processing

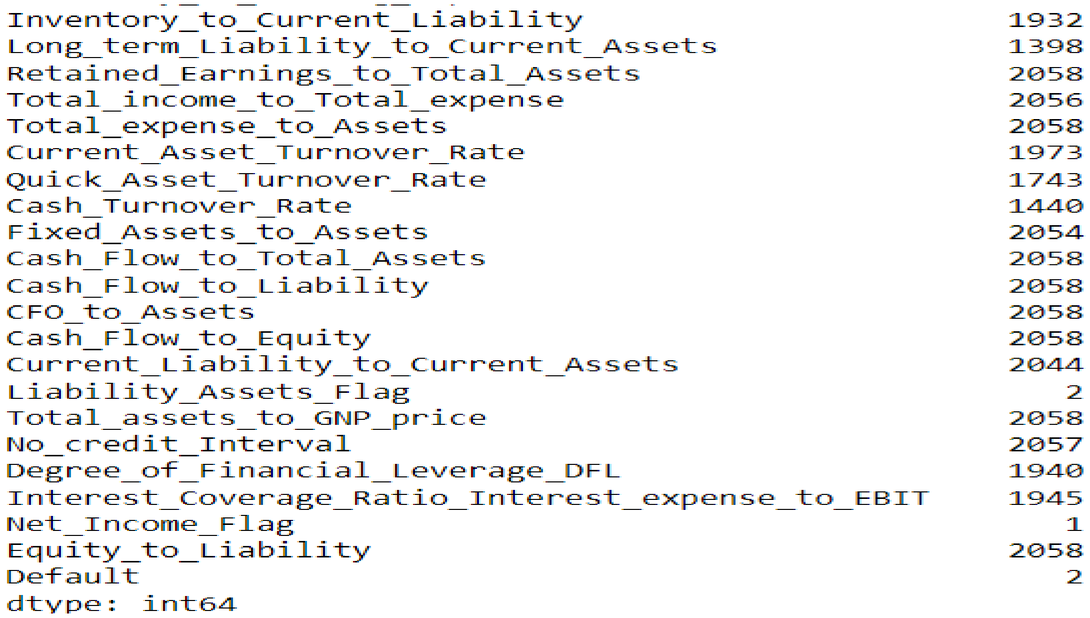
## 5. Prepare the data for modeling:

Dropping columns with few unique values

- Check the unique value

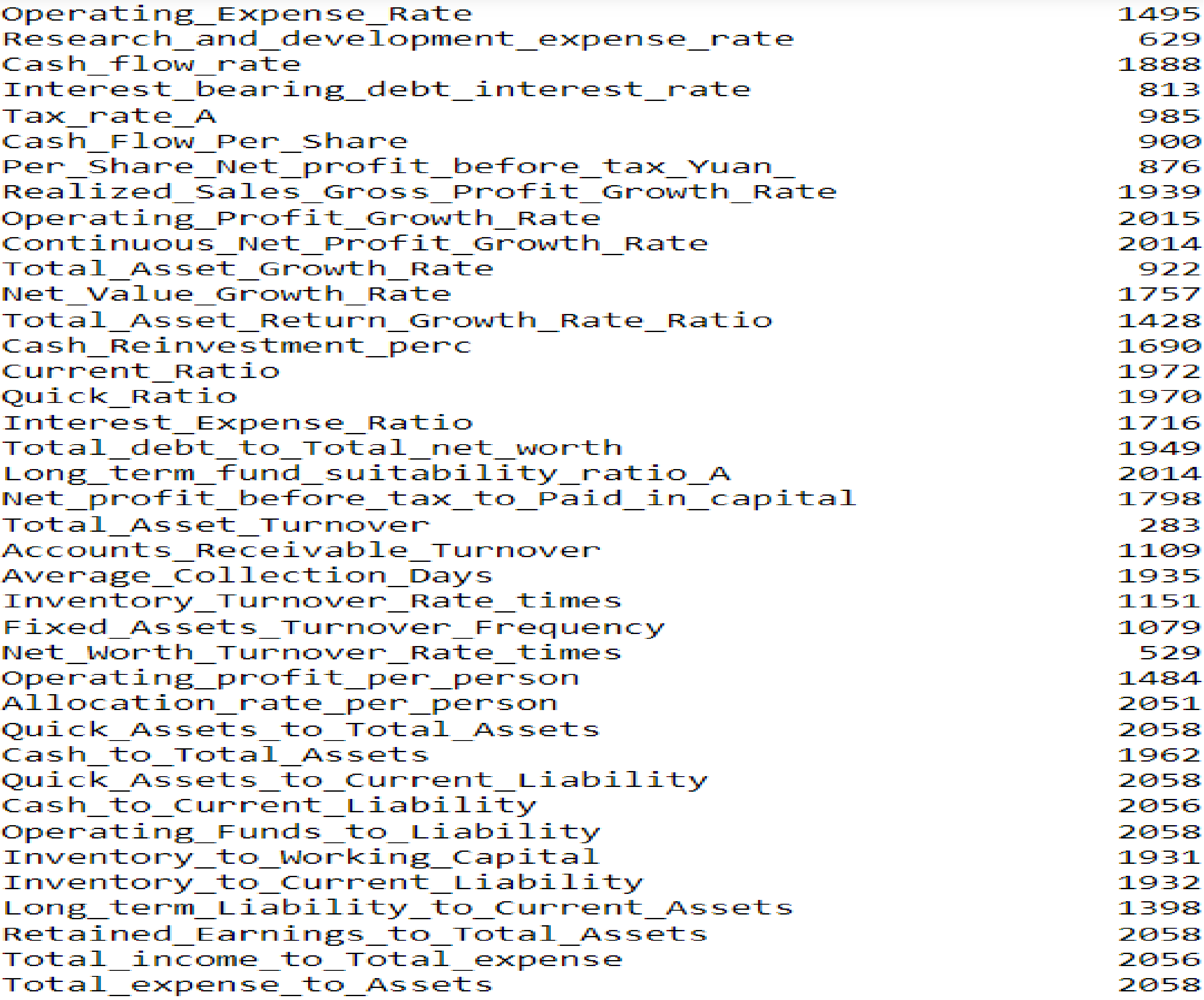


```
Operating_Expense_Rate                              1495
Research_and_development_expense_rate                629
Cash_flow_rate                                      1888
Interest_bearing_debt_interest_rate                  813
Tax_rate_A                                           985
Cash_Flow_Per_Share                                  900
Per_Share_Net_profit_before_tax_Yuan_                876
Realized_Sales_Gross_Profit_Growth_Rate             1939
Operating_Profit_Growth_Rate                        2015
Continuous_Net_Profit_Growth_Rate                   2014
Total_Asset_Growth_Rate                              922
Net_Value_Growth_Rate                               1757
Total_Asset_Return_Growth_Rate_Ratio                1428
Cash_Reinvestment_perc                              1690
Current_Ratio                                       1972
Quick_Ratio                                         1970
Interest_Expense_Ratio                              1716
Total_debt_to_Total_net_worth                       1949
Long_term_fund_suitability_ratio_A                  2014
Net_profit_before_tax_to_Paid_in_capital            1798
Total_Asset_Turnover                                 283
Accounts_Receivable_Turnover                        1109
Average_Collection_Days                             1935
Inventory_Turnover_Rate_times                       1151
Fixed_Assets_Turnover_Frequency                     1079
Net_Worth_Turnover_Rate_times                        529
Operating_profit_per_person                         1484
Allocation_rate_per_person                          2051
Quick_Assets_to_Total_Assets                        2058
Cash_to_Total_Assets                                1962
Quick_Assets_to_Current_Liability                   2058
Cash_to_Current_Liability                           2056
Operating_Funds_to_Liability                        2058
Inventory to Working Capital                        1931
```

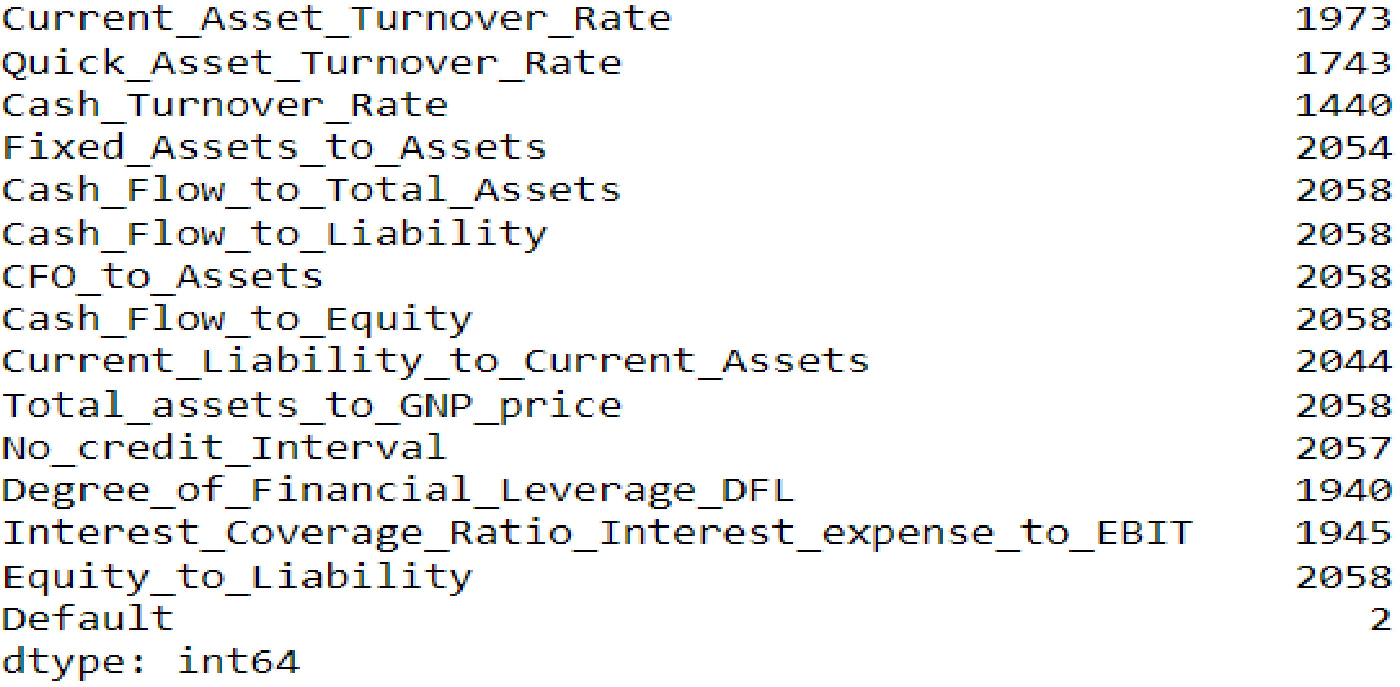


```
Inventory_to_Current_Liability                          1932
Long_term_Liability_to_Current_Assets                   1398
Retained_Earnings_to_Total_Assets                       2058
Total_income_to_Total_expense                           2056
Total_expense_to_Assets                                 2058
Current_Asset_Turnover_Rate                             1973
Quick_Asset_Turnover_Rate                               1743
Cash_Turnover_Rate                                      1440
Fixed_Assets_to_Assets                                  2054
Cash_Flow_to_Total_Assets                               2058
Cash_Flow_to_Liability                                  2058
CFO_to_Assets                                           2058
Cash_Flow_to_Equity                                     2058
Current_Liability_to_Current_Assets                     2044
Liability_Assets_Flag                                      2
Total_assets_to_GNP_price                               2058
No_credit_Interval                                      2057
Degree_of_Financial_Leverage_DFL                        1940
Interest_Coverage_Ratio_Interest_expense_to_EBIT        1945
Net_Income_Flag                                            1
Equity_to_Liability                                     2058
Default                                                    2
dtype: int64
```

We can drop the columns Net_Income_Flag and Liability_Assets_Flag as they have very few unique values.

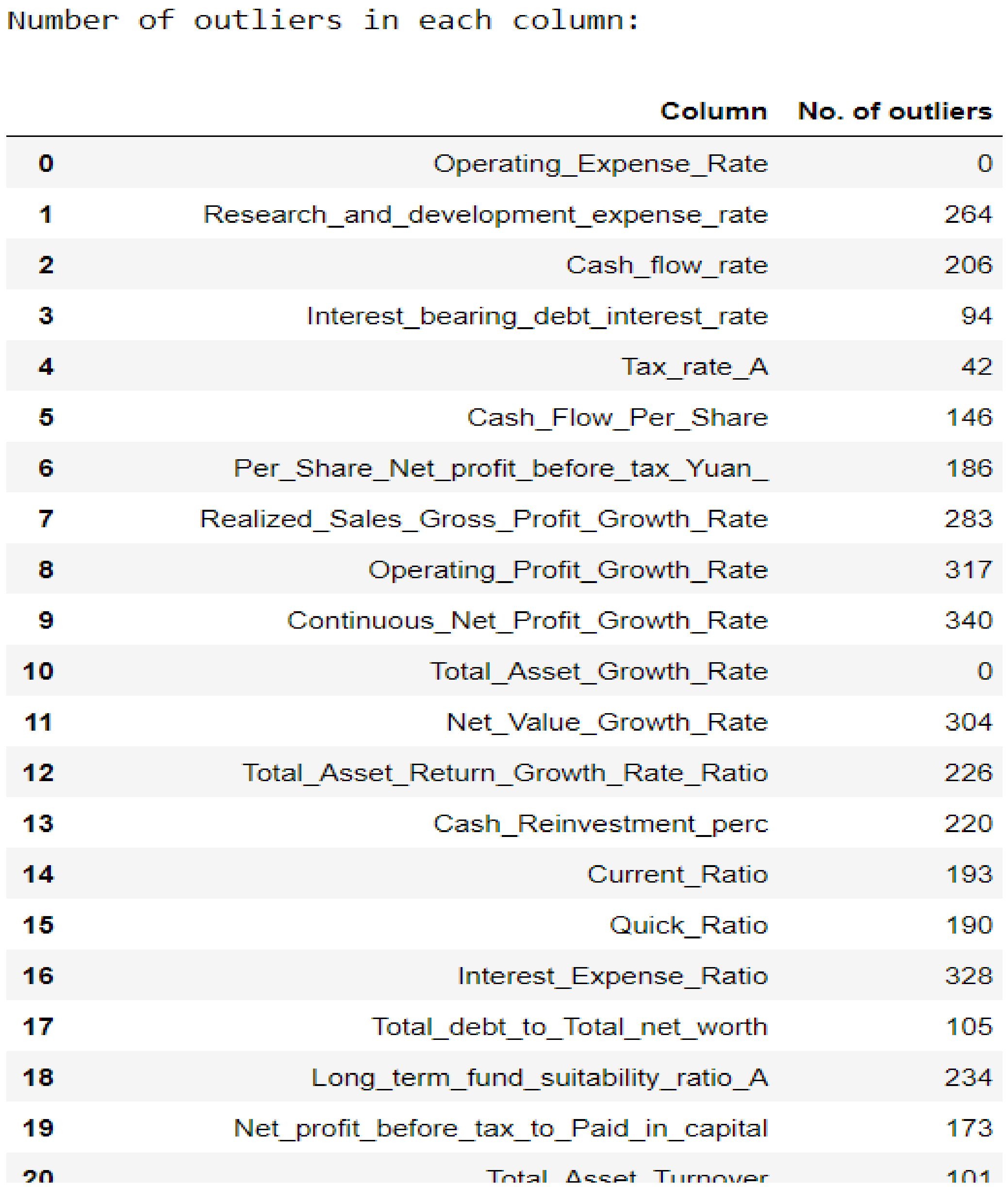


```
Operating_Expense_Rate                                      1495
Research_and_development_expense_rate                        629
Cash_flow_rate                                              1888
Interest_bearing_debt_interest_rate                          813
Tax_rate_A                                                   985
Cash_Flow_Per_Share                                          900
Per_Share_Net_profit_before_tax_Yuan_                        876
Realized_Sales_Gross_Profit_Growth_Rate                     1939
Operating_Profit_Growth_Rate                                2015
Continuous_Net_Profit_Growth_Rate                           2014
Total_Asset_Growth_Rate                                      922
Net_Value_Growth_Rate                                       1757
Total_Asset_Return_Growth_Rate_Ratio                        1428
Cash_Reinvestment_perc                                      1690
Current_Ratio                                               1972
Quick_Ratio                                                 1970
Interest_Expense_Ratio                                      1716
Total_debt_to_Total_net_worth                               1949
Long_term_fund_suitability_ratio_A                          2014
Net_profit_before_tax_to_Paid_in_capital                    1798
Total_Asset_Turnover                                         283
Accounts_Receivable_Turnover                                1109
Average_Collection_Days                                     1935
Inventory_Turnover_Rate_times                               1151
Fixed_Assets_Turnover_Frequency                             1079
Net_Worth_Turnover_Rate_times                                529
Operating_profit_per_person                                 1484
Allocation_rate_per_person                                  2051
Quick_Assets_to_Total_Assets                                2058
Cash_to_Total_Assets                                        1962
Quick_Assets_to_Current_Liability                           2058
Cash_to_Current_Liability                                   2056
Operating_Funds_to_Liability                                2058
Inventory_to_Working_Capital                                1931
Inventory_to_Current_Liability                              1932
Long_term_Liability_to_Current_Assets                       1398
Retained_Earnings_to_Total_Assets                           2058
Total_income_to_Total_expense                               2056
Total_expense_to_Assets                                     2058
```



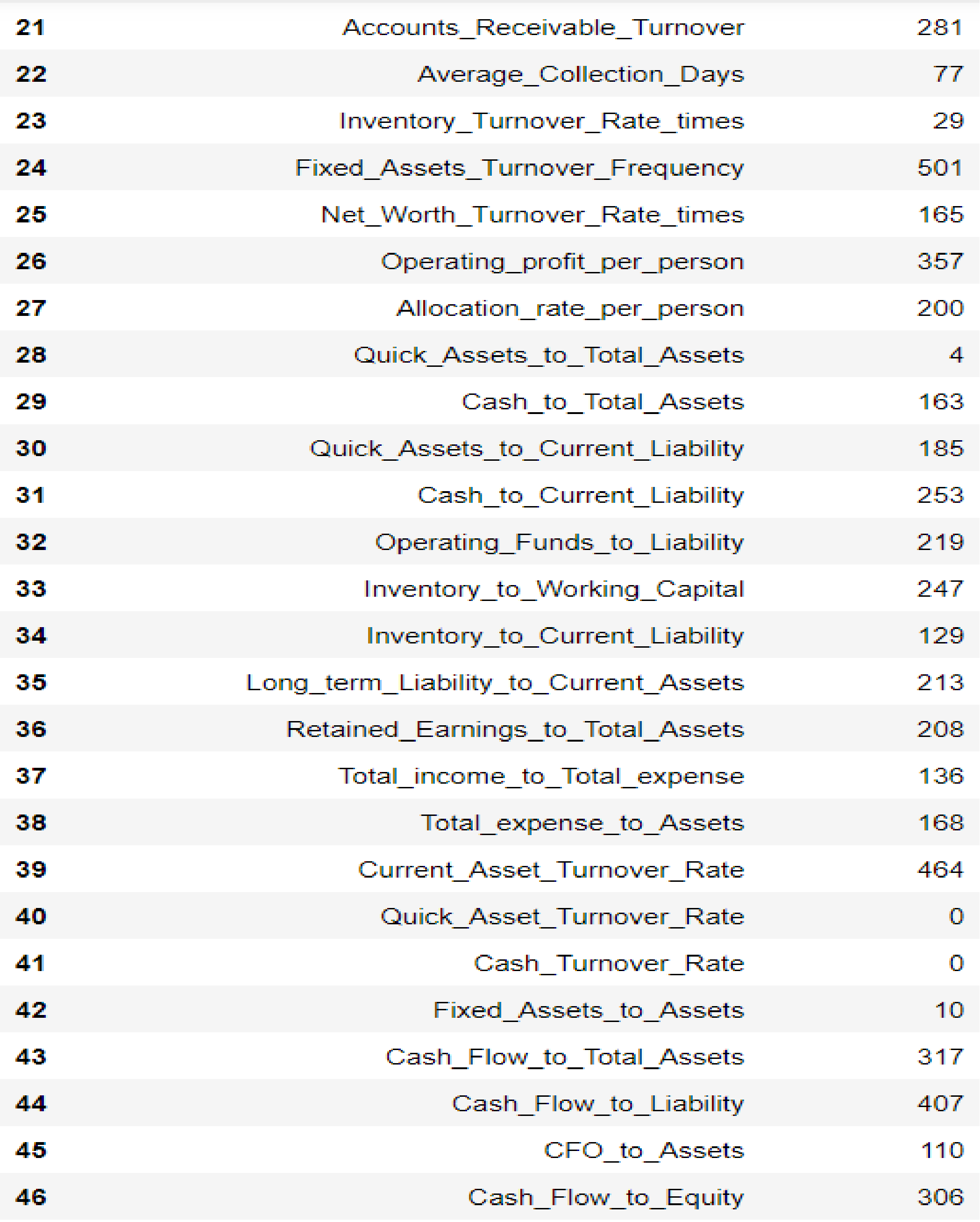
```
Current_Asset_Turnover_Rate                         1973
Quick_Asset_Turnover_Rate                           1743
Cash_Turnover_Rate                                  1440
Fixed_Assets_to_Assets                              2054
Cash_Flow_to_Total_Assets                           2058
Cash_Flow_to_Liability                              2058
CFO_to_Assets                                       2058
Cash_Flow_to_Equity                                 2058
Current_Liability_to_Current_Assets                 2044
Total_assets_to_GNP_price                           2058
No_credit_Interval                                  2057
Degree_of_Financial_Leverage_DFL                    1940
Interest_Coverage_Ratio_Interest_expense_to_EBIT    1945
Equity_to_Liability                                 2058
Default                                                2
dtype: int64
```

# 6. Outlier Detection (treat, if needed)

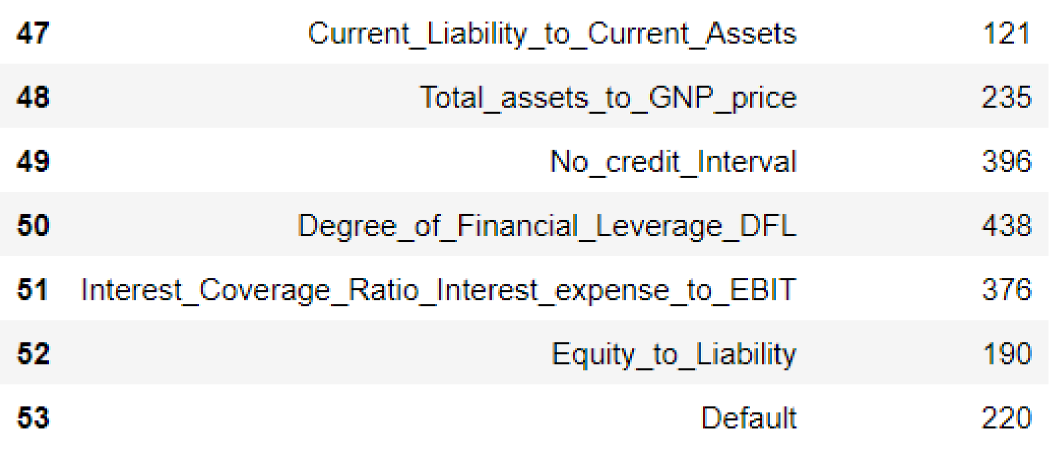


```
Number of outliers in each column:
```

| | Column | No. of outliers |
|---|---|---|
| **0** | Operating_Expense_Rate | 0 |
| **1** | Research_and_development_expense_rate | 264 |
| **2** | Cash_flow_rate | 206 |
| **3** | Interest_bearing_debt_interest_rate | 94 |
| **4** | Tax_rate_A | 42 |
| **5** | Cash_Flow_Per_Share | 146 |
| **6** | Per_Share_Net_profit_before_tax_Yuan_ | 186 |
| **7** | Realized_Sales_Gross_Profit_Growth_Rate | 283 |
| **8** | Operating_Profit_Growth_Rate | 317 |
| **9** | Continuous_Net_Profit_Growth_Rate | 340 |
| **10** | Total_Asset_Growth_Rate | 0 |
| **11** | Net_Value_Growth_Rate | 304 |
| **12** | Total_Asset_Return_Growth_Rate_Ratio | 226 |
| **13** | Cash_Reinvestment_perc | 220 |
| **14** | Current_Ratio | 193 |
| **15** | Quick_Ratio | 190 |
| **16** | Interest_Expense_Ratio | 328 |
| **17** | Total_debt_to_Total_net_worth | 105 |
| **18** | Long_term_fund_suitability_ratio_A | 234 |
| **19** | Net_profit_before_tax_to_Paid_in_capital | 173 |
| **20** | Total_Asset_Turnover | 101 |



| | | |
|---|---|---|
| 21 | Accounts_Receivable_Turnover | 281 |
| 22 | Average_Collection_Days | 77 |
| 23 | Inventory_Turnover_Rate_times | 29 |
| 24 | Fixed_Assets_Turnover_Frequency | 501 |
| 25 | Net_Worth_Turnover_Rate_times | 165 |
| 26 | Operating_profit_per_person | 357 |
| 27 | Allocation_rate_per_person | 200 |
| 28 | Quick_Assets_to_Total_Assets | 4 |
| 29 | Cash_to_Total_Assets | 163 |
| 30 | Quick_Assets_to_Current_Liability | 185 |
| 31 | Cash_to_Current_Liability | 253 |
| 32 | Operating_Funds_to_Liability | 219 |
| 33 | Inventory_to_Working_Capital | 247 |
| 34 | Inventory_to_Current_Liability | 129 |
| 35 | Long_term_Liability_to_Current_Assets | 213 |
| 36 | Retained_Earnings_to_Total_Assets | 208 |
| 37 | Total_income_to_Total_expense | 136 |
| 38 | Total_expense_to_Assets | 168 |
| 39 | Current_Asset_Turnover_Rate | 464 |
| 40 | Quick_Asset_Turnover_Rate | 0 |
| 41 | Cash_Turnover_Rate | 0 |
| 42 | Fixed_Assets_to_Assets | 10 |
| 43 | Cash_Flow_to_Total_Assets | 317 |
| 44 | Cash_Flow_to_Liability | 407 |
| 45 | CFO_to_Assets | 110 |
| 46 | Cash_Flow_to_Equity | 306 |



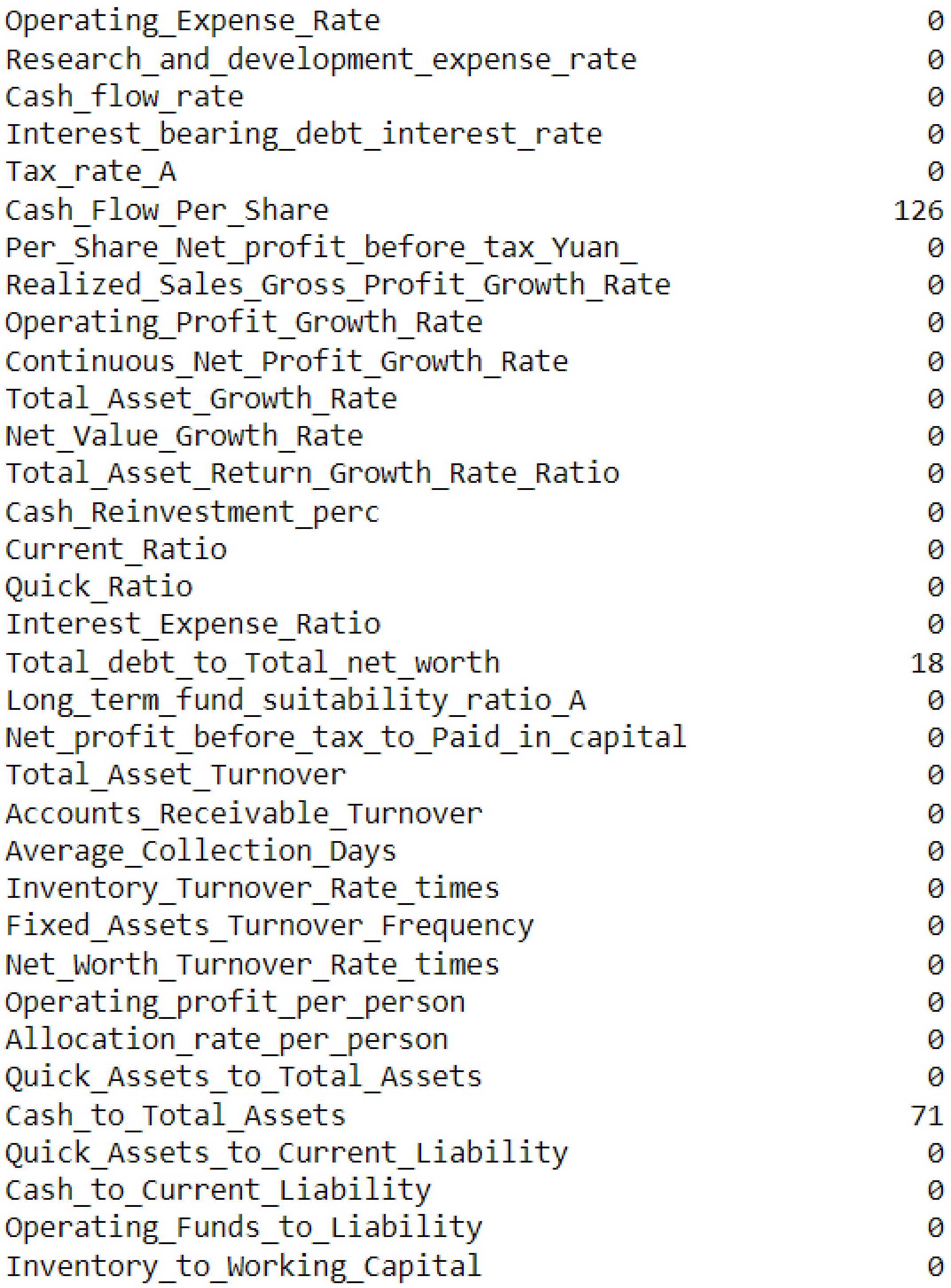
| | | |
|---|---|---|
| 47 | Current_Liability_to_Current_Assets | 121 |
| 48 | Total_assets_to_GNP_price | 235 |
| 49 | No_credit_Interval | 396 |
| 50 | Degree_of_Financial_Leverage_DFL | 438 |
| 51 | Interest_Coverage_Ratio_Interest_expense_to_EBIT | 376 |
| 52 | Equity_to_Liability | 190 |
| 53 | Default | 220 |

## 7. Data split

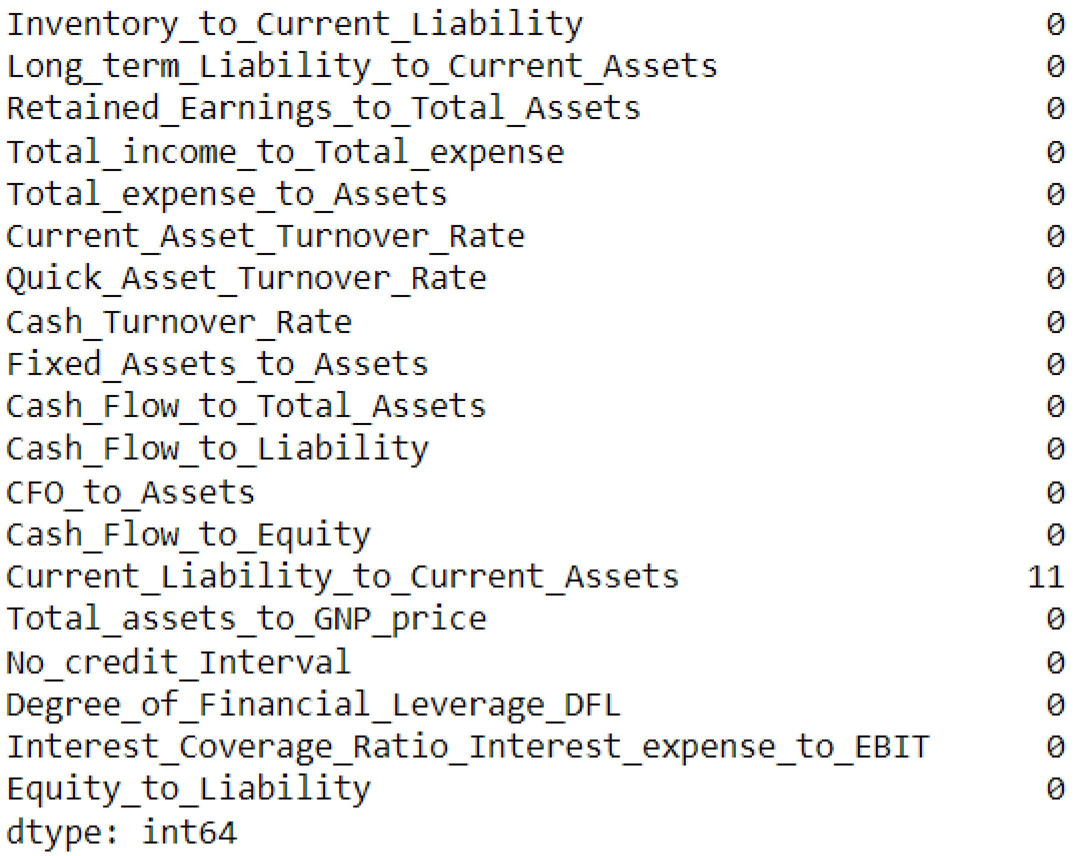
Seperating target variable from the rest of the data. Then, split the data into train and test in the ratio 75:25..

Missing Values Detection and Treatment

Check missing values of Train Dataset

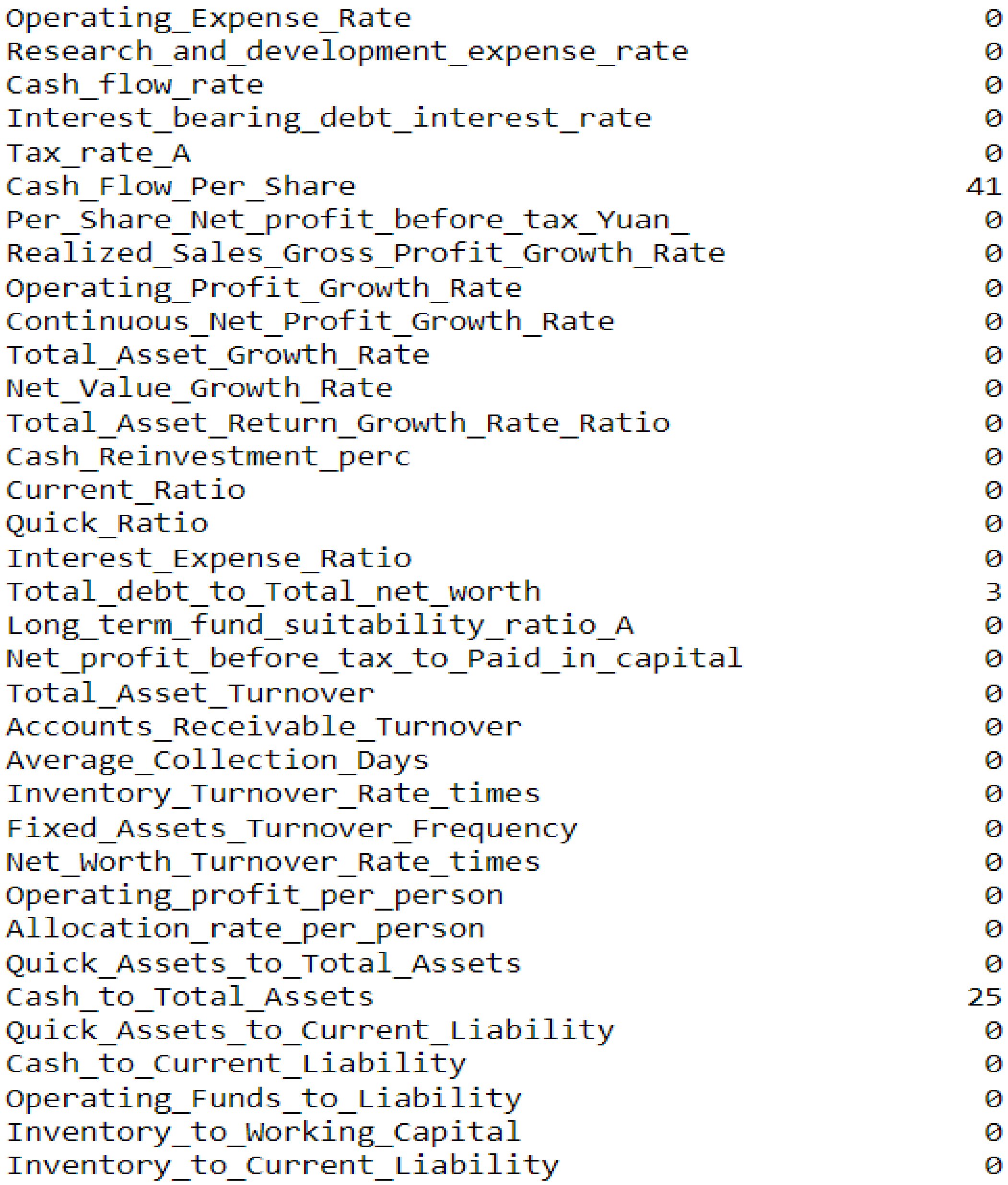


```
Operating_Expense_Rate                                0
Research_and_development_expense_rate                 0
Cash_flow_rate                                        0
Interest_bearing_debt_interest_rate                   0
Tax_rate_A                                            0
Cash_Flow_Per_Share                                 126
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                        18
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                                 71
Quick_Assets_to_Current_Liability                     0
Cash_to_Current_Liability                             0
Operating_Funds_to_Liability                          0
Inventory_to_Working_Capital                          0
```

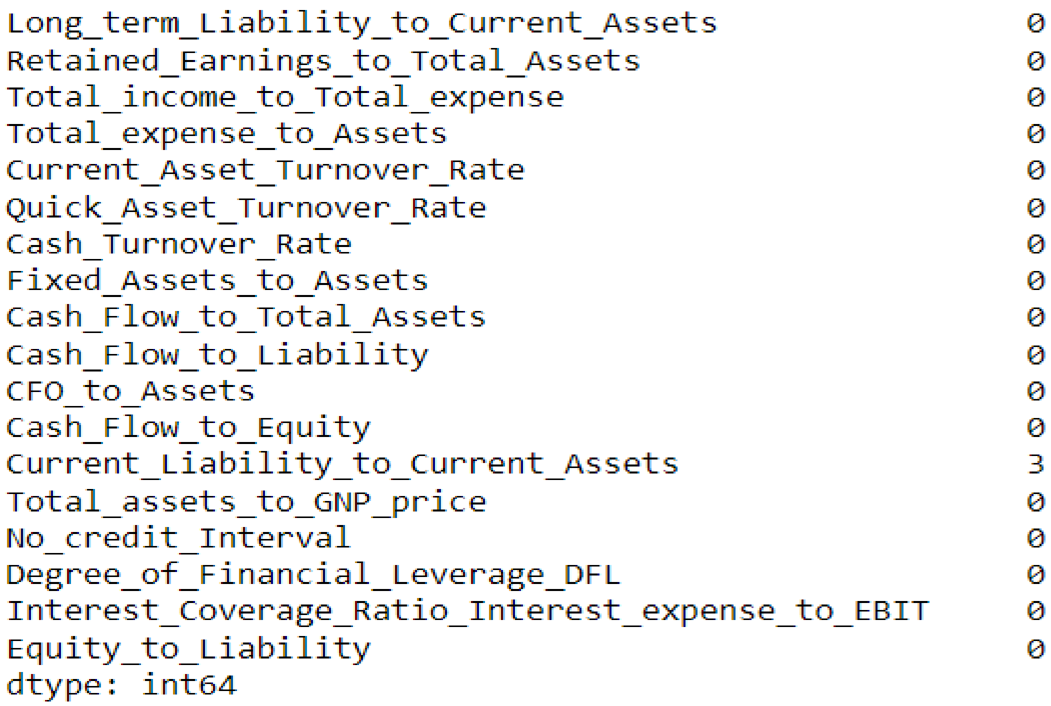


```
Inventory_to_Current_Liability                          0
Long_term_Liability_to_Current_Assets                   0
Retained_Earnings_to_Total_Assets                       0
Total_income_to_Total_expense                           0
Total_expense_to_Assets                                 0
Current_Asset_Turnover_Rate                             0
Quick_Asset_Turnover_Rate                               0
Cash_Turnover_Rate                                      0
Fixed_Assets_to_Assets                                  0
Cash_Flow_to_Total_Assets                               0
Cash_Flow_to_Liability                                  0
CFO_to_Assets                                           0
Cash_Flow_to_Equity                                     0
Current_Liability_to_Current_Assets                    11
Total_assets_to_GNP_price                               0
No_credit_Interval                                      0
Degree_of_Financial_Leverage_DFL                        0
Interest_Coverage_Ratio_Interest_expense_to_EBIT        0
Equity_to_Liability                                     0
dtype: int64
```

# Check missing values of Test Dataset

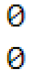


```
Operating_Expense_Rate                                 0
Research_and_development_expense_rate                  0
Cash_flow_rate                                         0
Interest_bearing_debt_interest_rate                    0
Tax_rate_A                                             0
Cash_Flow_Per_Share                                   41
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                          3
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                                  25
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
Total_income_to_Total_expense                         0
Total_expense_to_Assets                               0
Current_Asset_Turnover_Rate                           0
Quick_Asset_Turnover_Rate                             0
Cash_Turnover_Rate                                    0
Fixed_Assets_to_Assets                                0
Cash_Flow_to_Total_Assets                             0
Cash_Flow_to_Liability                                0
CFO_to_Assets                                         0
Cash_Flow_to_Equity                                   0
Current_Liability_to_Current_Assets                   3
Total_assets_to_GNP_price                             0
No_credit_Interval                                    0
Degree_of_Financial_Leverage_DFL                      0
Interest_Coverage_Ratio_Interest_expense_to_EBIT      0
Equity_to_Liability                                   0
dtype: int64
```

Replace the missing values in the data using KNN Imputer. Then, check the missing value of Train and test data.

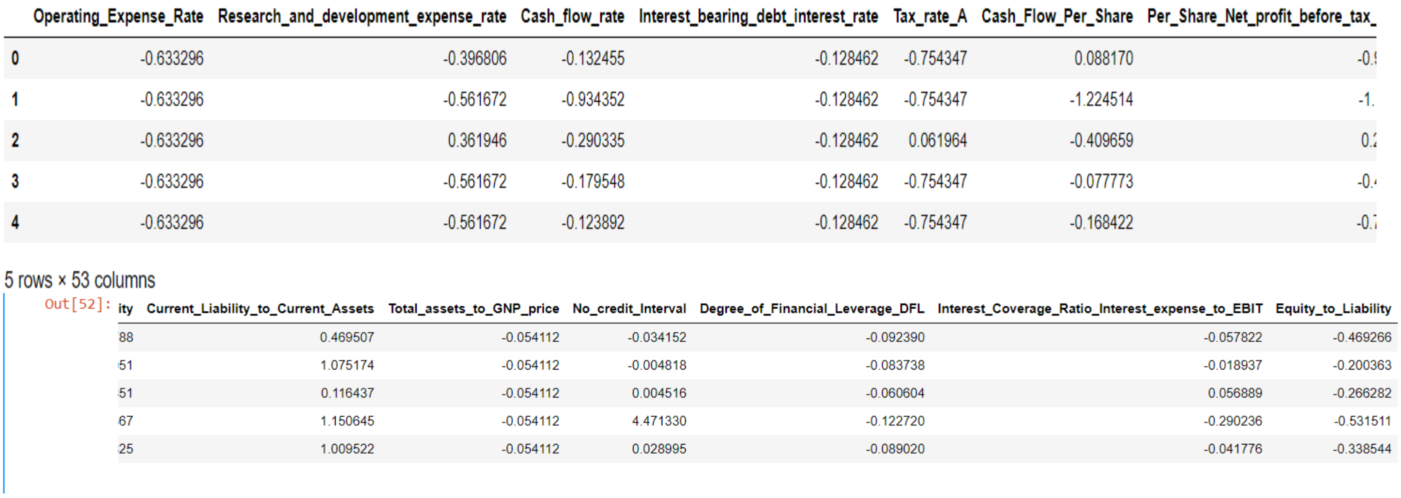


```
0
0
```

# 8. Scale the data

Scaling of features is done to bring all the features to the same scale.

## Scale Train data



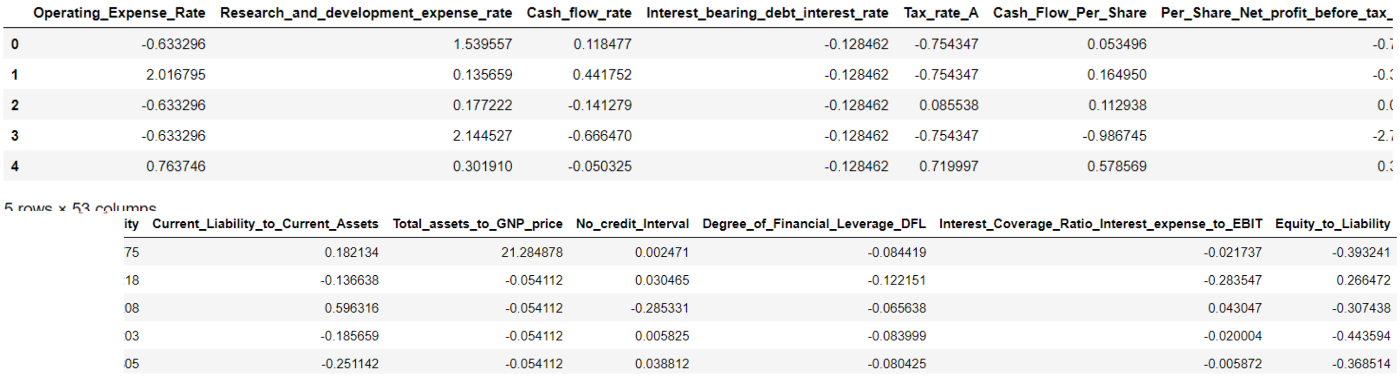
| | Operating_Expense_Rate | Research_and_development_expense_rate | Cash_flow_rate | Interest_bearing_debt_interest_rate | Tax_rate_A | Cash_Flow_Per_Share | Per_Share_Net_profit_before_tax_ |
|---|---|---|---|---|---|---|---|
| 0 | -0.633296 | -0.396806 | -0.132455 | -0.128462 | -0.754347 | 0.088170 | -0.9 |
| 1 | -0.633296 | -0.561672 | -0.934352 | -0.128462 | -0.754347 | -1.224514 | -1. |
| 2 | -0.633296 | 0.361946 | -0.290335 | -0.128462 | 0.061964 | -0.409659 | 0.2 |
| 3 | -0.633296 | -0.561672 | -0.179548 | -0.128462 | -0.754347 | -0.077773 | -0.4 |
| 4 | -0.633296 | -0.561672 | -0.123892 | -0.128462 | -0.754347 | -0.168422 | -0.7 |

5 rows × 53 columns

Out[52]:

| ity | Current_Liability_to_Current_Assets | Total_assets_to_GNP_price | No_credit_Interval | Degree_of_Financial_Leverage_DFL | Interest_Coverage_Ratio_Interest_expense_to_EBIT | Equity_to_Liability |
|---|---|---|---|---|---|---|
| 88 | 0.469507 | -0.054112 | -0.034152 | -0.092390 | -0.057822 | -0.469266 |
| 51 | 1.075174 | -0.054112 | -0.004818 | -0.083738 | -0.018937 | -0.200363 |
| 51 | 0.116437 | -0.054112 | 0.004516 | -0.060604 | 0.056889 | -0.266282 |
| 67 | 1.150645 | -0.054112 | 4.471330 | -0.122720 | -0.290236 | -0.531511 |
| 25 | 1.009522 | -0.054112 | 0.028995 | -0.089020 | -0.041776 | -0.338544 |

## Scale Test data



| | Operating_Expense_Rate | Research_and_development_expense_rate | Cash_flow_rate | Interest_bearing_debt_interest_rate | Tax_rate_A | Cash_Flow_Per_Share | Per_Share_Net_profit_before_tax_ |
|---|---|---|---|---|---|---|---|
| 0 | -0.633296 | 1.539557 | 0.118477 | -0.128462 | -0.754347 | 0.053496 | -0.7 |
| 1 | 2.016795 | 0.135659 | 0.441752 | -0.128462 | -0.754347 | 0.164950 | -0.3 |
| 2 | -0.633296 | 0.177222 | -0.141279 | -0.128462 | 0.085538 | 0.112938 | 0.0 |
| 3 | -0.633296 | 2.144527 | -0.666470 | -0.128462 | -0.754347 | -0.986745 | -2.7 |
| 4 | 0.763746 | 0.301910 | -0.050325 | -0.128462 | 0.719997 | 0.578569 | 0.3 |

5 rows × 53 columns

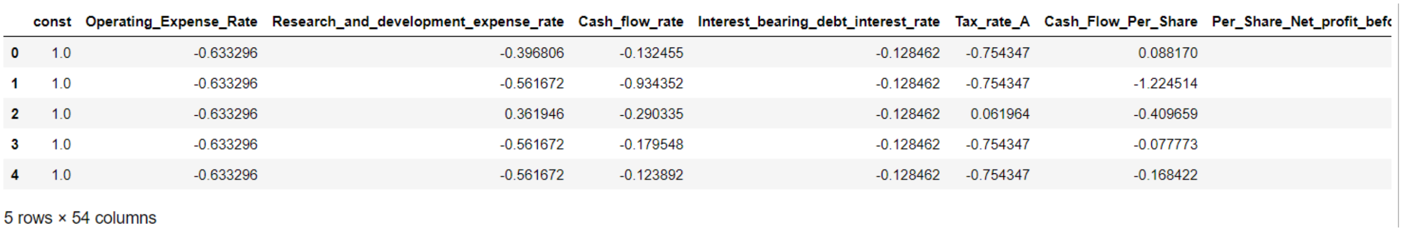
| ity | Current_Liability_to_Current_Assets | Total_assets_to_GNP_price | No_credit_Interval | Degree_of_Financial_Leverage_DFL | Interest_Coverage_Ratio_Interest_expense_to_EBIT | Equity_to_Liability |
|---|---|---|---|---|---|---|
| 75 | 0.182134 | 21.284878 | 0.002471 | -0.084419 | -0.021737 | -0.393241 |
| 18 | -0.136638 | -0.054112 | 0.030465 | -0.122151 | -0.283547 | 0.266472 |
| 08 | 0.596316 | -0.054112 | -0.285331 | -0.065638 | 0.043047 | -0.307438 |
| 03 | -0.185659 | -0.054112 | 0.005825 | -0.083999 | -0.020004 | -0.443594 |
| 05 | -0.251142 | -0.054112 | 0.038812 | -0.080425 | -0.005872 | -0.368514 |

## 9. Model Building - Metrics of Choice (Justify the evaluation metrics)

Defining a function to compute different metrics to check performance of a classification model built using sklearn.

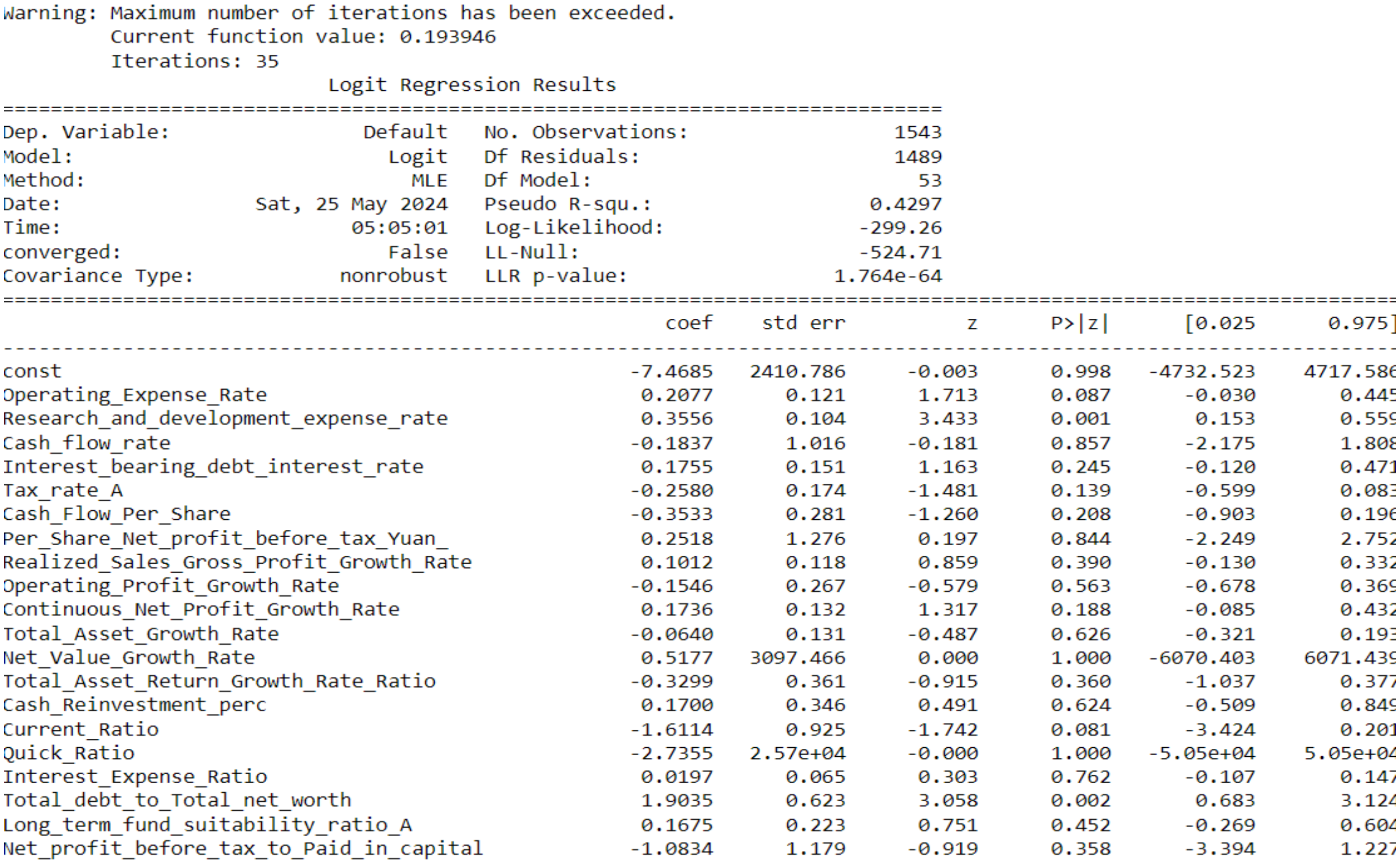
## 10. Model Building (Logistic Regression, Random Forest)

Adding constant to data for Logistic Regression

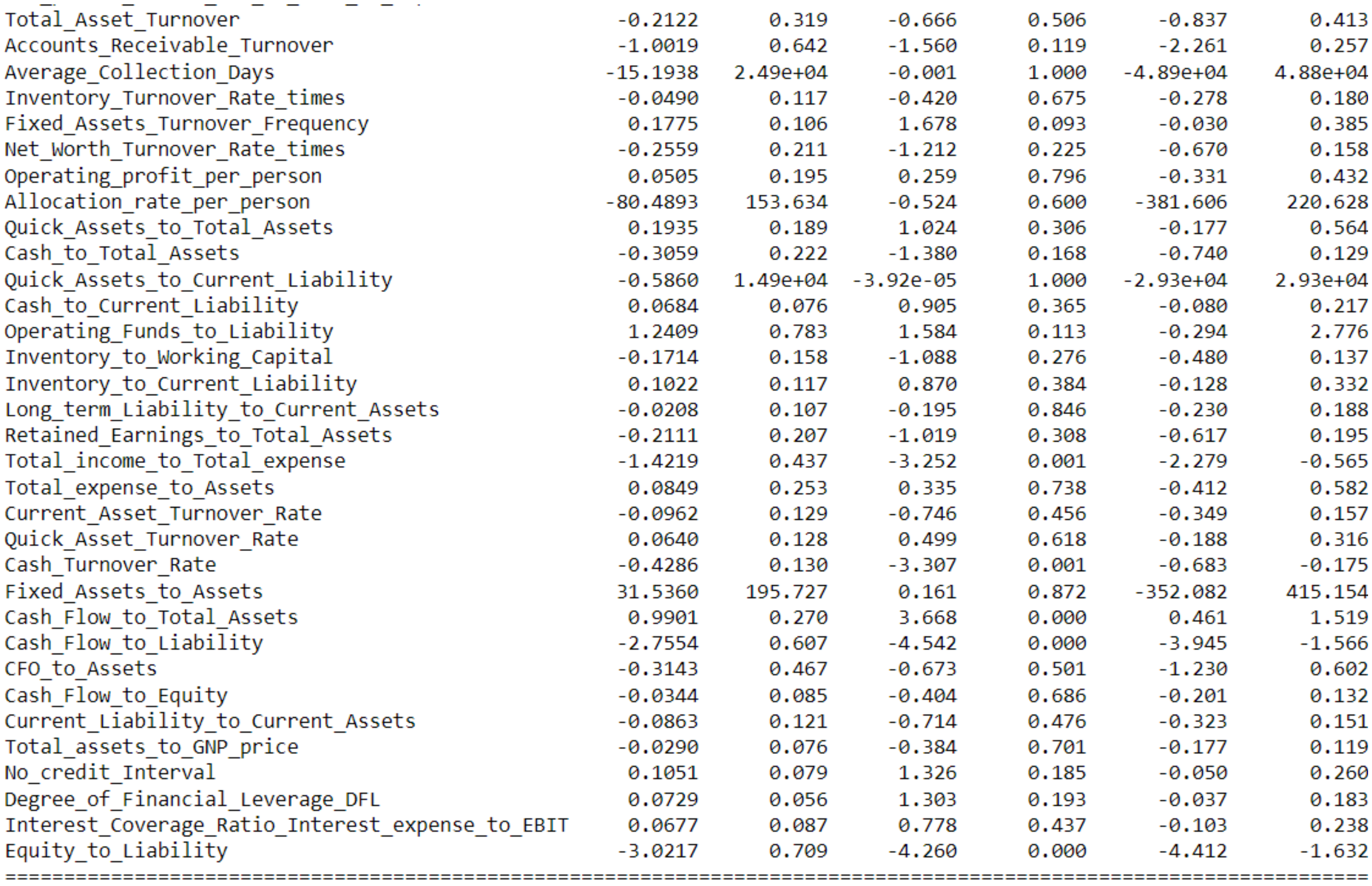


| | const | Operating_Expense_Rate | Research_and_development_expense_rate | Cash_flow_rate | Interest_bearing_debt_interest_rate | Tax_rate_A | Cash_Flow_Per_Share | Per_Share_Net_profit_befo |
|---|---|---|---|---|---|---|---|---|
| 0 | 1.0 | -0.633296 | -0.396806 | -0.132455 | -0.128462 | -0.754347 | 0.088170 | |
| 1 | 1.0 | -0.633296 | -0.561672 | -0.934352 | -0.128462 | -0.754347 | -1.224514 | |
| 2 | 1.0 | -0.633296 | 0.361946 | -0.290335 | -0.128462 | 0.061964 | -0.409659 | |
| 3 | 1.0 | -0.633296 | -0.561672 | -0.179548 | -0.128462 | -0.754347 | -0.077773 | |
| 4 | 1.0 | -0.633296 | -0.561672 | -0.123892 | -0.128462 | -0.754347 | -0.168422 | |

5 rows × 54 columns

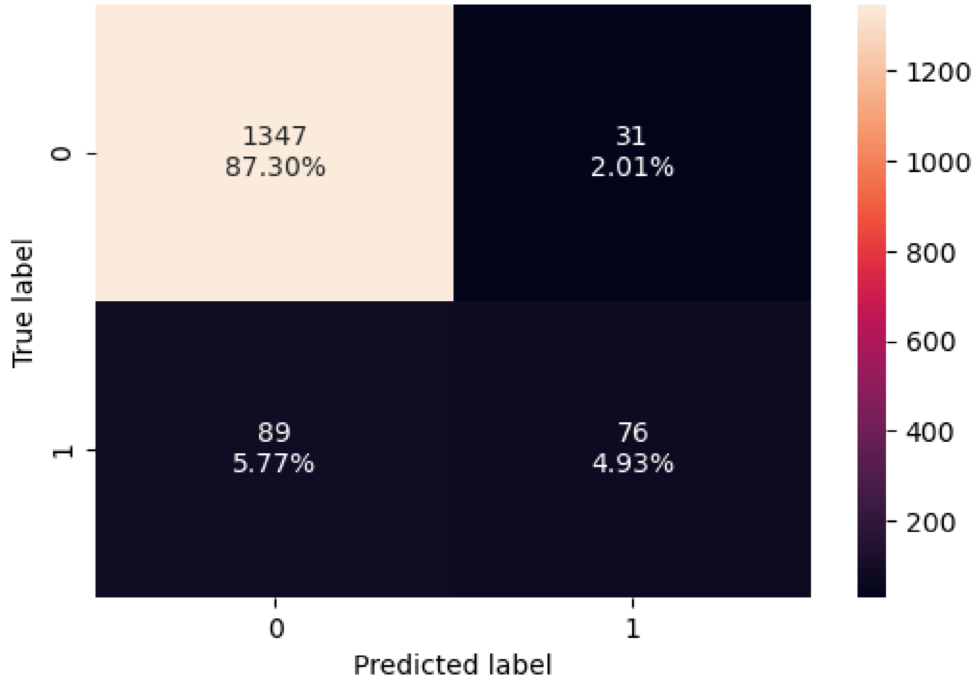


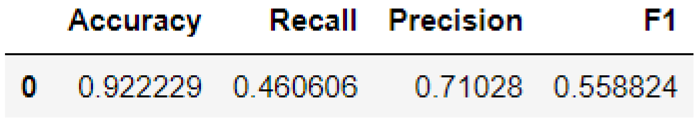
```
Warning: Maximum number of iterations has been exceeded.
         Current function value: 0.193946
         Iterations: 35
                           Logit Regression Results
==============================================================================
Dep. Variable:                Default   No. Observations:                 1543
Model:                          Logit   Df Residuals:                     1489
Method:                           MLE   Df Model:                           53
Date:                Sat, 25 May 2024   Pseudo R-squ.:                  0.4297
Time:                        05:05:01   Log-Likelihood:                -299.26
converged:                      False   LL-Null:                       -524.71
Covariance Type:            nonrobust   LLR p-value:                 1.764e-64
====================================================================================================================
                                                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------------------------------------------
const                                           -7.4685   2410.786     -0.003      0.998   -4732.523    4717.586
Operating_Expense_Rate                           0.2077      0.121      1.713      0.087      -0.030       0.445
Research_and_development_expense_rate            0.3556      0.104      3.433      0.001       0.153       0.559
Cash_flow_rate                                  -0.1837      1.016     -0.181      0.857      -2.175       1.808
Interest_bearing_debt_interest_rate              0.1755      0.151      1.163      0.245      -0.120       0.471
Tax_rate_A                                      -0.2580      0.174     -1.481      0.139      -0.599       0.083
Cash_Flow_Per_Share                             -0.3533      0.281     -1.260      0.208      -0.903       0.196
Per_Share_Net_profit_before_tax_Yuan_            0.2518      1.276      0.197      0.844      -2.249       2.752
Realized_Sales_Gross_Profit_Growth_Rate          0.1012      0.118      0.859      0.390      -0.130       0.332
Operating_Profit_Growth_Rate                    -0.1546      0.267     -0.579      0.563      -0.678       0.369
Continuous_Net_Profit_Growth_Rate                0.1736      0.132      1.317      0.188      -0.085       0.432
Total_Asset_Growth_Rate                         -0.0640      0.131     -0.487      0.626      -0.321       0.193
Net_Value_Growth_Rate                            0.5177   3097.466      0.000      1.000   -6070.403    6071.439
Total_Asset_Return_Growth_Rate_Ratio            -0.3299      0.361     -0.915      0.360      -1.037       0.377
Cash_Reinvestment_perc                           0.1700      0.346      0.491      0.624      -0.509       0.849
Current_Ratio                                   -1.6114      0.925     -1.742      0.081      -3.424       0.201
Quick_Ratio                                     -2.7355   2.57e+04     -0.000      1.000   -5.05e+04    5.05e+04
Interest_Expense_Ratio                           0.0197      0.065      0.303      0.762      -0.107       0.147
Total_debt_to_Total_net_worth                    1.9035      0.623      3.058      0.002       0.683       3.124
Long_term_fund_suitability_ratio_A               0.1675      0.223      0.751      0.452      -0.269       0.604
Net_profit_before_tax_to_Paid_in_capital        -1.0834      1.179     -0.919      0.358      -3.394       1.227
```



```
Total_Asset_Turnover                              -0.2122      0.319     -0.666      0.506      -0.837       0.413
Accounts_Receivable_Turnover                      -1.0019      0.642     -1.560      0.119      -2.261       0.257
Average_Collection_Days                          -15.1938   2.49e+04     -0.001      1.000   -4.89e+04    4.88e+04
Inventory_Turnover_Rate_times                     -0.0490      0.117     -0.420      0.675      -0.278       0.180
Fixed_Assets_Turnover_Frequency                    0.1775      0.106      1.678      0.093      -0.030       0.385
Net_Worth_Turnover_Rate_times                     -0.2559      0.211     -1.212      0.225      -0.670       0.158
Operating_profit_per_person                        0.0505      0.195      0.259      0.796      -0.331       0.432
Allocation_rate_per_person                       -80.4893    153.634     -0.524      0.600    -381.606     220.628
Quick_Assets_to_Total_Assets                       0.1935      0.189      1.024      0.306      -0.177       0.564
Cash_to_Total_Assets                              -0.3059      0.222     -1.380      0.168      -0.740       0.129
Quick_Assets_to_Current_Liability                 -0.5860   1.49e+04  -3.92e-05      1.000   -2.93e+04    2.93e+04
Cash_to_Current_Liability                          0.0684      0.076      0.905      0.365      -0.080       0.217
Operating_Funds_to_Liability                       1.2409      0.783      1.584      0.113      -0.294       2.776
Inventory_to_Working_Capital                      -0.1714      0.158     -1.088      0.276      -0.480       0.137
Inventory_to_Current_Liability                     0.1022      0.117      0.870      0.384      -0.128       0.332
Long_term_Liability_to_Current_Assets             -0.0208      0.107     -0.195      0.846      -0.230       0.188
Retained_Earnings_to_Total_Assets                 -0.2111      0.207     -1.019      0.308      -0.617       0.195
Total_income_to_Total_expense                     -1.4219      0.437     -3.252      0.001      -2.279      -0.565
Total_expense_to_Assets                            0.0849      0.253      0.335      0.738      -0.412       0.582
Current_Asset_Turnover_Rate                       -0.0962      0.129     -0.746      0.456      -0.349       0.157
Quick_Asset_Turnover_Rate                          0.0640      0.128      0.499      0.618      -0.188       0.316
Cash_Turnover_Rate                                -0.4286      0.130     -3.307      0.001      -0.683      -0.175
Fixed_Assets_to_Assets                            31.5360    195.727      0.161      0.872    -352.082     415.154
Cash_Flow_to_Total_Assets                          0.9901      0.270      3.668      0.000       0.461       1.519
Cash_Flow_to_Liability                            -2.7554      0.607     -4.542      0.000      -3.945      -1.566
CFO_to_Assets                                     -0.3143      0.467     -0.673      0.501      -1.230       0.602
Cash_Flow_to_Equity                               -0.0344      0.085     -0.404      0.686      -0.201       0.132
Current_Liability_to_Current_Assets               -0.0863      0.121     -0.714      0.476      -0.323       0.151
Total_assets_to_GNP_price                         -0.0290      0.076     -0.384      0.701      -0.177       0.119
No_credit_Interval                                 0.1051      0.079      1.326      0.185      -0.050       0.260
Degree_of_Financial_Leverage_DFL                   0.0729      0.056      1.303      0.193      -0.037       0.183
Interest_Coverage_Ratio_Interest_expense_to_EBIT   0.0677      0.087      0.778      0.437      -0.103       0.238
Equity_to_Liability                               -3.0217      0.709     -4.260      0.000      -4.412      -1.632
===================================================================================================================
```

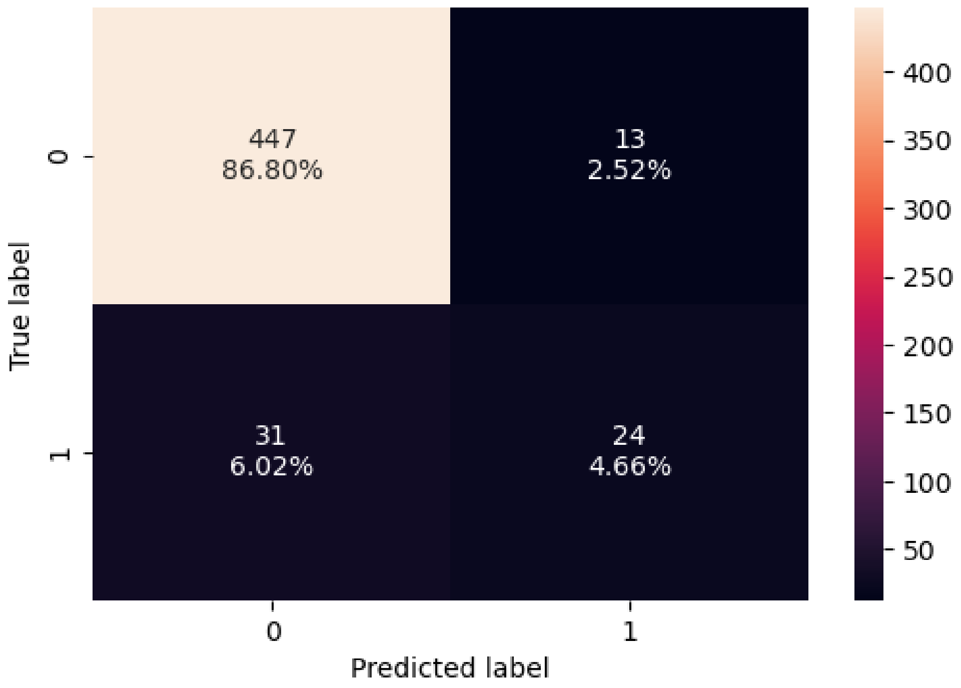
- Logistic Regression Model - Training Performance

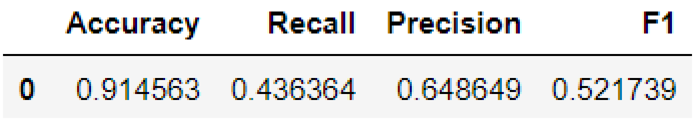




|   | Accuracy | Recall | Precision | F1 |
|---|---|---|---|---|
| **0** | 0.922229 | 0.460606 | 0.71028 | 0.558824 |

- Logistic Regression Model - Test Performance



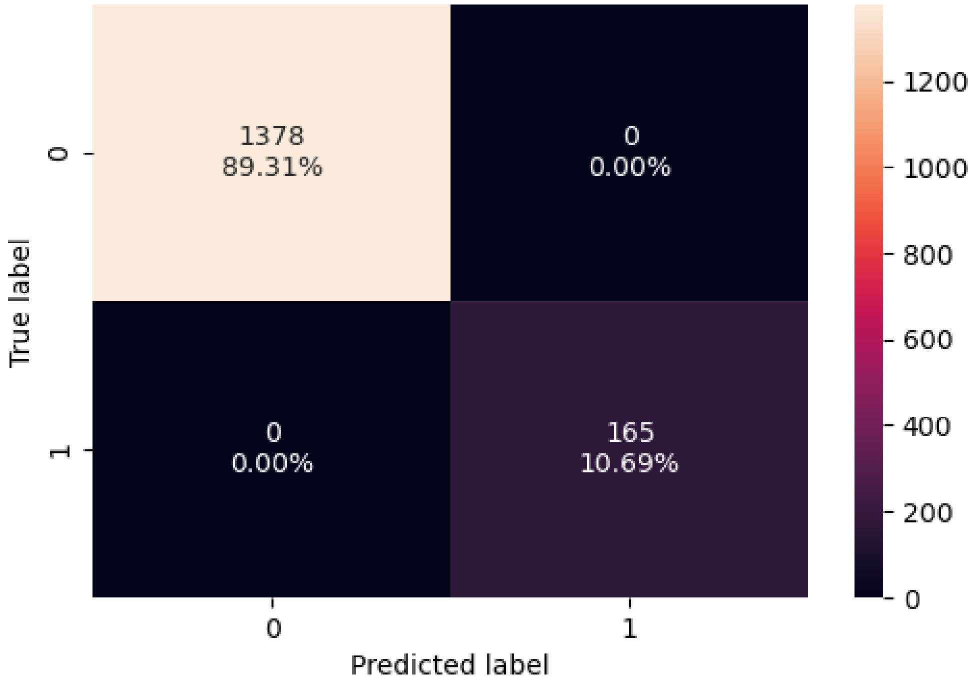


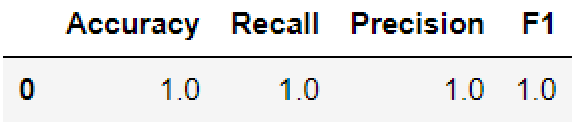
| | Accuracy | Recall | Precision | F1 |
|---|---|---|---|---|
| **0** | 0.914563 | 0.436364 | 0.648649 | 0.521739 |

- Random Forest

Define random forest with random state = 42

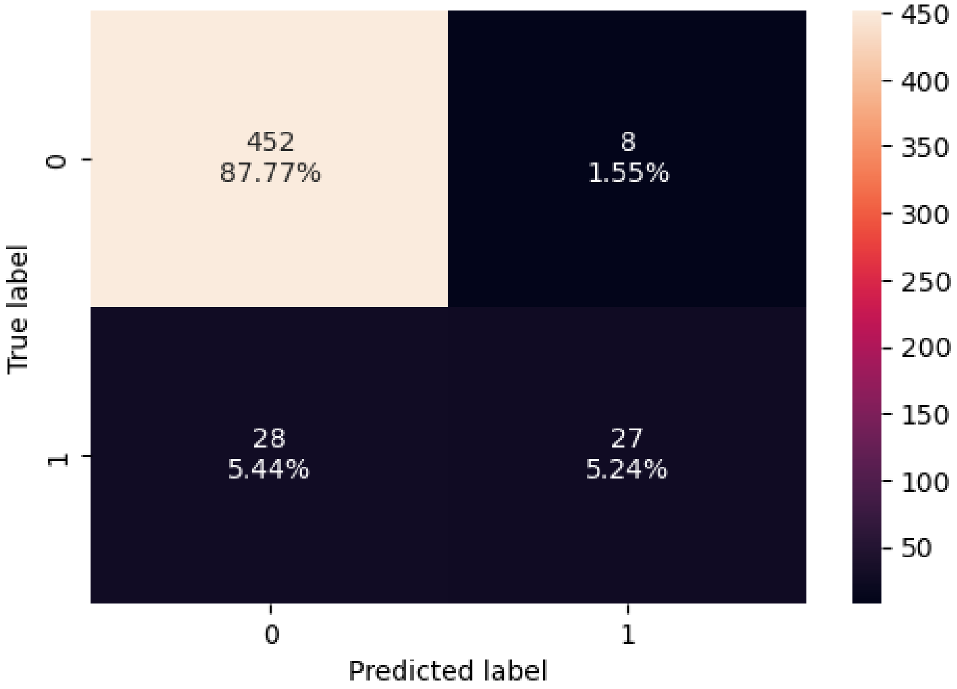
- Random Forest Model - Test Performance

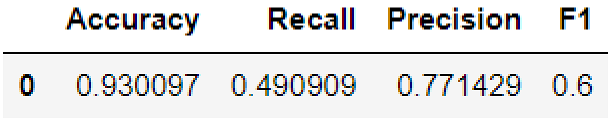




| | Accuracy | Recall | Precision | F1 |
|---|---|---|---|---|
| **0** | 1.0 | 1.0 | 1.0 | 1.0 |

- Random Forest Model - Test Performance



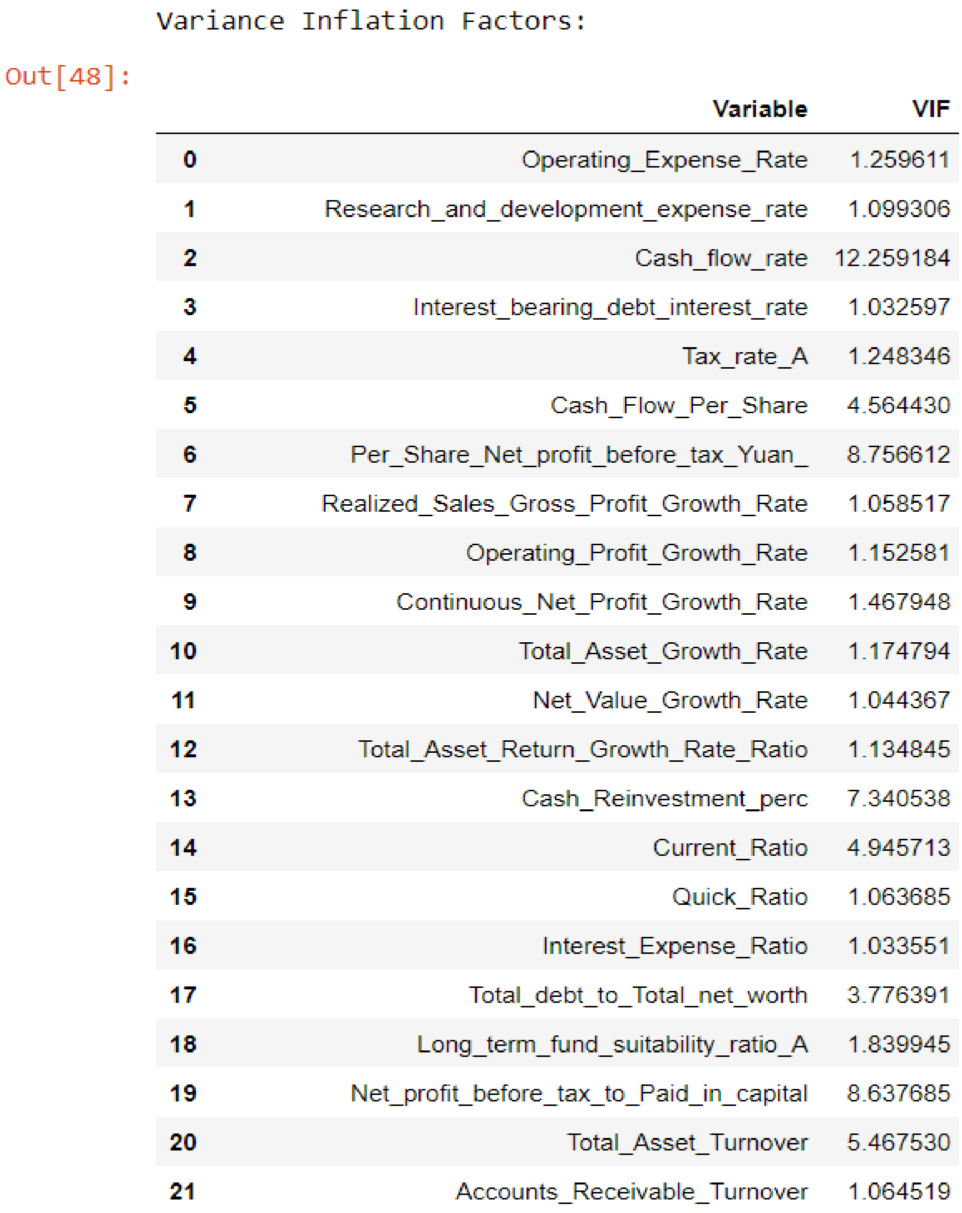


|   | Accuracy | Recall | Precision | F1 |
|---|---|---|---|---|
| **0** | 0.930097 | 0.490909 | 0.771429 | 0.6 |

# PART A: Model Performance Improvement

## 11. Dealing with multicollinearity using VIF

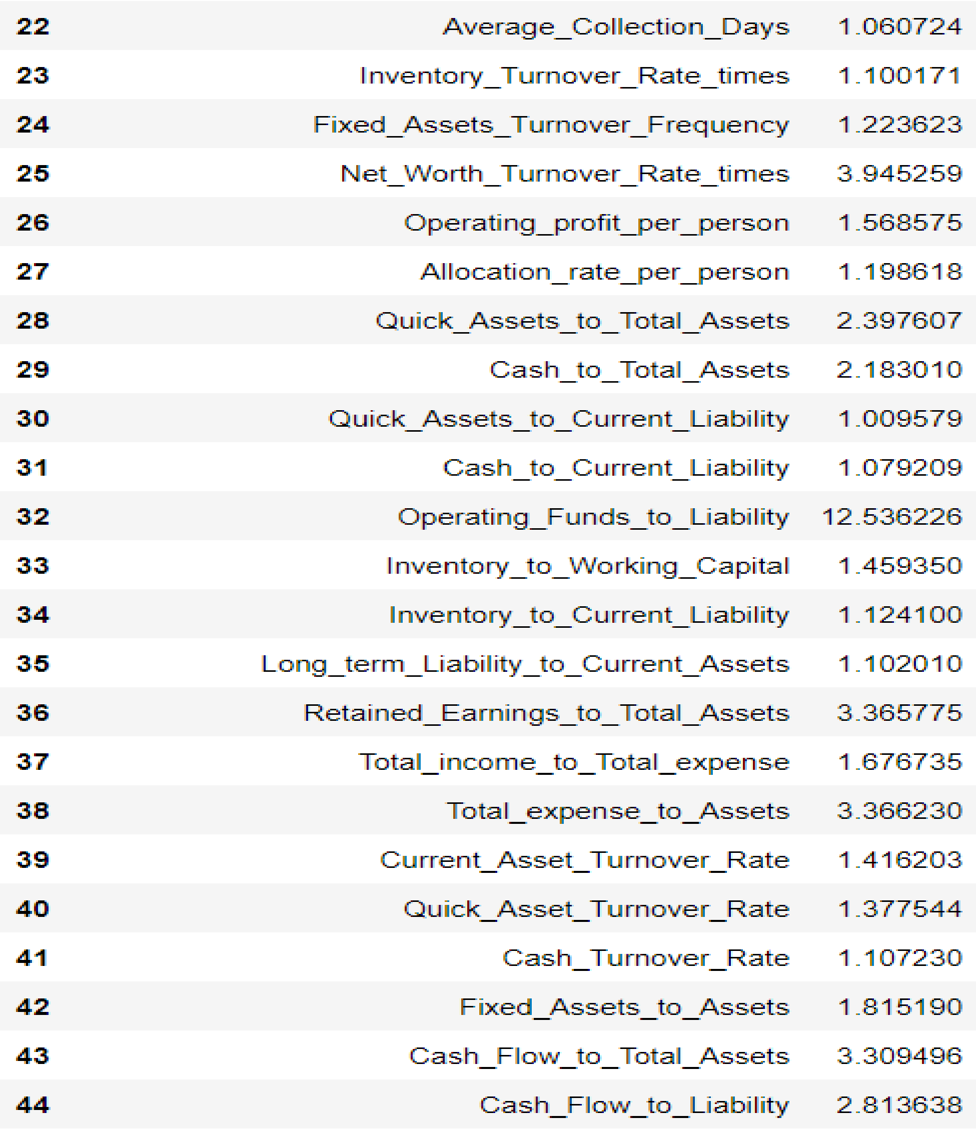
### Model Performance Improvement - Logistic Regression



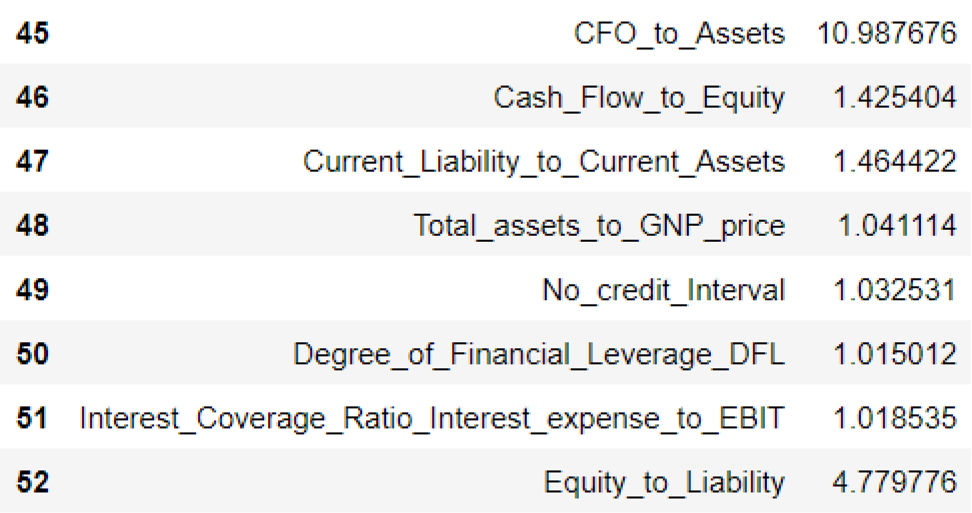
```
Variance Inflation Factors:
```

Out[48]:

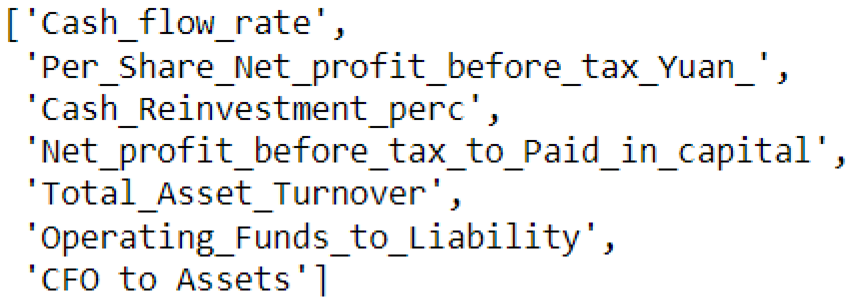
|  | Variable | VIF |
|---|---|---|
| **0** | Operating_Expense_Rate | 1.259611 |
| **1** | Research_and_development_expense_rate | 1.099306 |
| **2** | Cash_flow_rate | 12.259184 |
| **3** | Interest_bearing_debt_interest_rate | 1.032597 |
| **4** | Tax_rate_A | 1.248346 |
| **5** | Cash_Flow_Per_Share | 4.564430 |
| **6** | Per_Share_Net_profit_before_tax_Yuan_ | 8.756612 |
| **7** | Realized_Sales_Gross_Profit_Growth_Rate | 1.058517 |
| **8** | Operating_Profit_Growth_Rate | 1.152581 |
| **9** | Continuous_Net_Profit_Growth_Rate | 1.467948 |
| **10** | Total_Asset_Growth_Rate | 1.174794 |
| **11** | Net_Value_Growth_Rate | 1.044367 |
| **12** | Total_Asset_Return_Growth_Rate_Ratio | 1.134845 |
| **13** | Cash_Reinvestment_perc | 7.340538 |
| **14** | Current_Ratio | 4.945713 |
| **15** | Quick_Ratio | 1.063685 |
| **16** | Interest_Expense_Ratio | 1.033551 |
| **17** | Total_debt_to_Total_net_worth | 3.776391 |
| **18** | Long_term_fund_suitability_ratio_A | 1.839945 |
| **19** | Net_profit_before_tax_to_Paid_in_capital | 8.637685 |
| **20** | Total_Asset_Turnover | 5.467530 |
| **21** | Accounts_Receivable_Turnover | 1.064519 |



| | | |
|---|---|---|
| 22 | Average_Collection_Days | 1.060724 |
| 23 | Inventory_Turnover_Rate_times | 1.100171 |
| 24 | Fixed_Assets_Turnover_Frequency | 1.223623 |
| 25 | Net_Worth_Turnover_Rate_times | 3.945259 |
| 26 | Operating_profit_per_person | 1.568575 |
| 27 | Allocation_rate_per_person | 1.198618 |
| 28 | Quick_Assets_to_Total_Assets | 2.397607 |
| 29 | Cash_to_Total_Assets | 2.183010 |
| 30 | Quick_Assets_to_Current_Liability | 1.009579 |
| 31 | Cash_to_Current_Liability | 1.079209 |
| 32 | Operating_Funds_to_Liability | 12.536226 |
| 33 | Inventory_to_Working_Capital | 1.459350 |
| 34 | Inventory_to_Current_Liability | 1.124100 |
| 35 | Long_term_Liability_to_Current_Assets | 1.102010 |
| 36 | Retained_Earnings_to_Total_Assets | 3.365775 |
| 37 | Total_income_to_Total_expense | 1.676735 |
| 38 | Total_expense_to_Assets | 3.366230 |
| 39 | Current_Asset_Turnover_Rate | 1.416203 |
| 40 | Quick_Asset_Turnover_Rate | 1.377544 |
| 41 | Cash_Turnover_Rate | 1.107230 |
| 42 | Fixed_Assets_to_Assets | 1.815190 |
| 43 | Cash_Flow_to_Total_Assets | 3.309496 |
| 44 | Cash_Flow_to_Liability | 2.813638 |

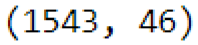


| | | |
|---|---|---|
| 45 | CFO_to_Assets | 10.987676 |
| 46 | Cash_Flow_to_Equity | 1.425404 |
| 47 | Current_Liability_to_Current_Assets | 1.464422 |
| 48 | Total_assets_to_GNP_price | 1.041114 |
| 49 | No_credit_Interval | 1.032531 |
| 50 | Degree_of_Financial_Leverage_DFL | 1.015012 |
| 51 | Interest_Coverage_Ratio_Interest_expense_to_EBIT | 1.018535 |
| 52 | Equity_to_Liability | 4.779776 |

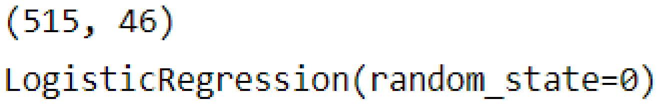


```
['Cash_flow_rate',
 'Per_Share_Net_profit_before_tax_Yuan_',
 'Cash_Reinvestment_perc',
 'Net_profit_before_tax_to_Paid_in_capital',
 'Total_Asset_Turnover',
 'Operating_Funds_to_Liability',
 'CFO to Assets']
```

## Dropping columns with VIF > 5

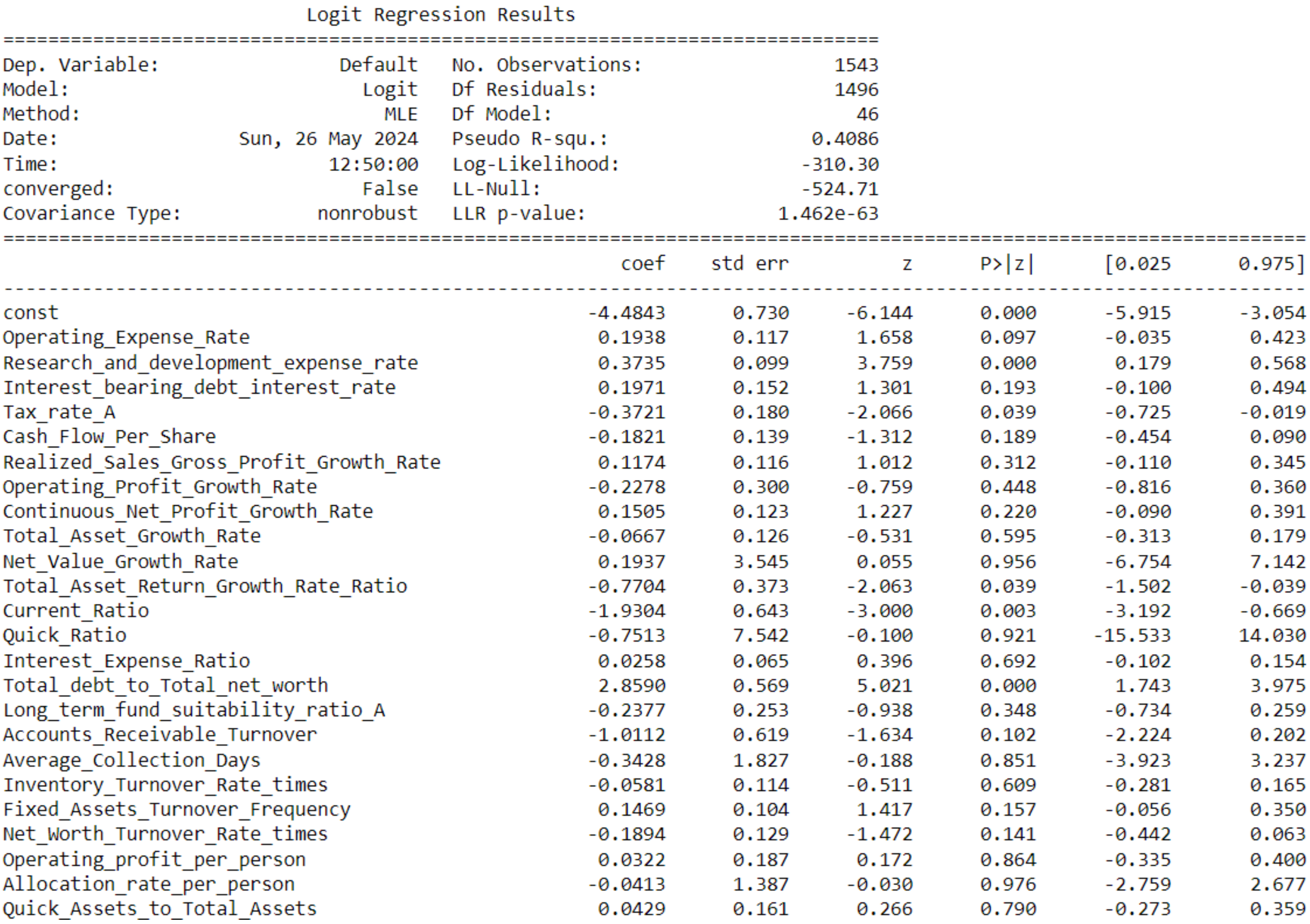


```
(1543, 46)
```

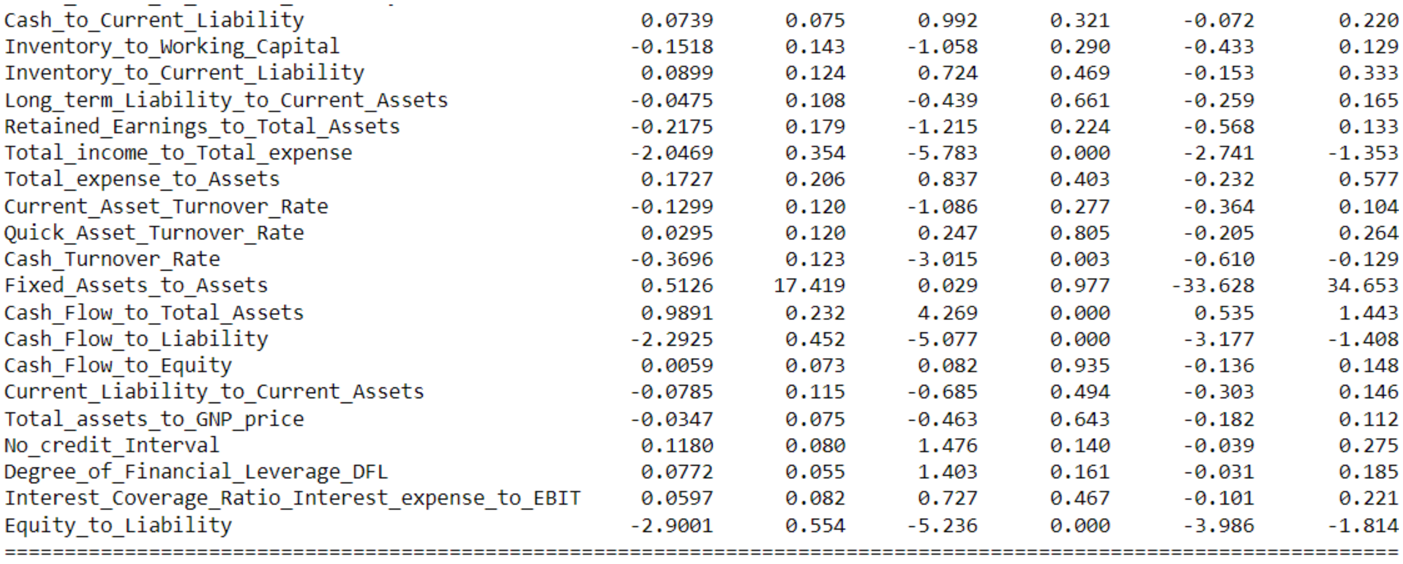


```
(515, 46)
LogisticRegression(random_state=0)
```

## Logistic Regression Model on new train data with intercept



```
                           Logit Regression Results
==============================================================================
Dep. Variable:                Default   No. Observations:                 1543
Model:                          Logit   Df Residuals:                     1496
Method:                           MLE   Df Model:                           46
Date:                Sun, 26 May 2024   Pseudo R-squ.:                  0.4086
Time:                        12:50:00   Log-Likelihood:                -310.30
converged:                      False   LL-Null:                       -524.71
Covariance Type:            nonrobust   LLR p-value:                 1.462e-63
===================================================================================================================
                                                 coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------------------------------------
const                                         -4.4843      0.730     -6.144      0.000      -5.915      -3.054
Operating_Expense_Rate                         0.1938      0.117      1.658      0.097      -0.035       0.423
Research_and_development_expense_rate          0.3735      0.099      3.759      0.000       0.179       0.568
Interest_bearing_debt_interest_rate            0.1971      0.152      1.301      0.193      -0.100       0.494
Tax_rate_A                                    -0.3721      0.180     -2.066      0.039      -0.725      -0.019
Cash_Flow_Per_Share                           -0.1821      0.139     -1.312      0.189      -0.454       0.090
Realized_Sales_Gross_Profit_Growth_Rate        0.1174      0.116      1.012      0.312      -0.110       0.345
Operating_Profit_Growth_Rate                  -0.2278      0.300     -0.759      0.448      -0.816       0.360
Continuous_Net_Profit_Growth_Rate              0.1505      0.123      1.227      0.220      -0.090       0.391
Total_Asset_Growth_Rate                       -0.0667      0.126     -0.531      0.595      -0.313       0.179
Net_Value_Growth_Rate                          0.1937      3.545      0.055      0.956      -6.754       7.142
Total_Asset_Return_Growth_Rate_Ratio          -0.7704      0.373     -2.063      0.039      -1.502      -0.039
Current_Ratio                                 -1.9304      0.643     -3.000      0.003      -3.192      -0.669
Quick_Ratio                                   -0.7513      7.542     -0.100      0.921     -15.533      14.030
Interest_Expense_Ratio                         0.0258      0.065      0.396      0.692      -0.102       0.154
Total_debt_to_Total_net_worth                  2.8590      0.569      5.021      0.000       1.743       3.975
Long_term_fund_suitability_ratio_A            -0.2377      0.253     -0.938      0.348      -0.734       0.259
Accounts_Receivable_Turnover                  -1.0112      0.619     -1.634      0.102      -2.224       0.202
Average_Collection_Days                       -0.3428      1.827     -0.188      0.851      -3.923       3.237
Inventory_Turnover_Rate_times                 -0.0581      0.114     -0.511      0.609      -0.281       0.165
Fixed_Assets_Turnover_Frequency                0.1469      0.104      1.417      0.157      -0.056       0.350
Net_Worth_Turnover_Rate_times                 -0.1894      0.129     -1.472      0.141      -0.442       0.063
Operating_profit_per_person                    0.0322      0.187      0.172      0.864      -0.335       0.400
Allocation_rate_per_person                    -0.0413      1.387     -0.030      0.976      -2.759       2.677
Quick_Assets_to_Total_Assets                   0.0429      0.161      0.266      0.790      -0.273       0.359
```

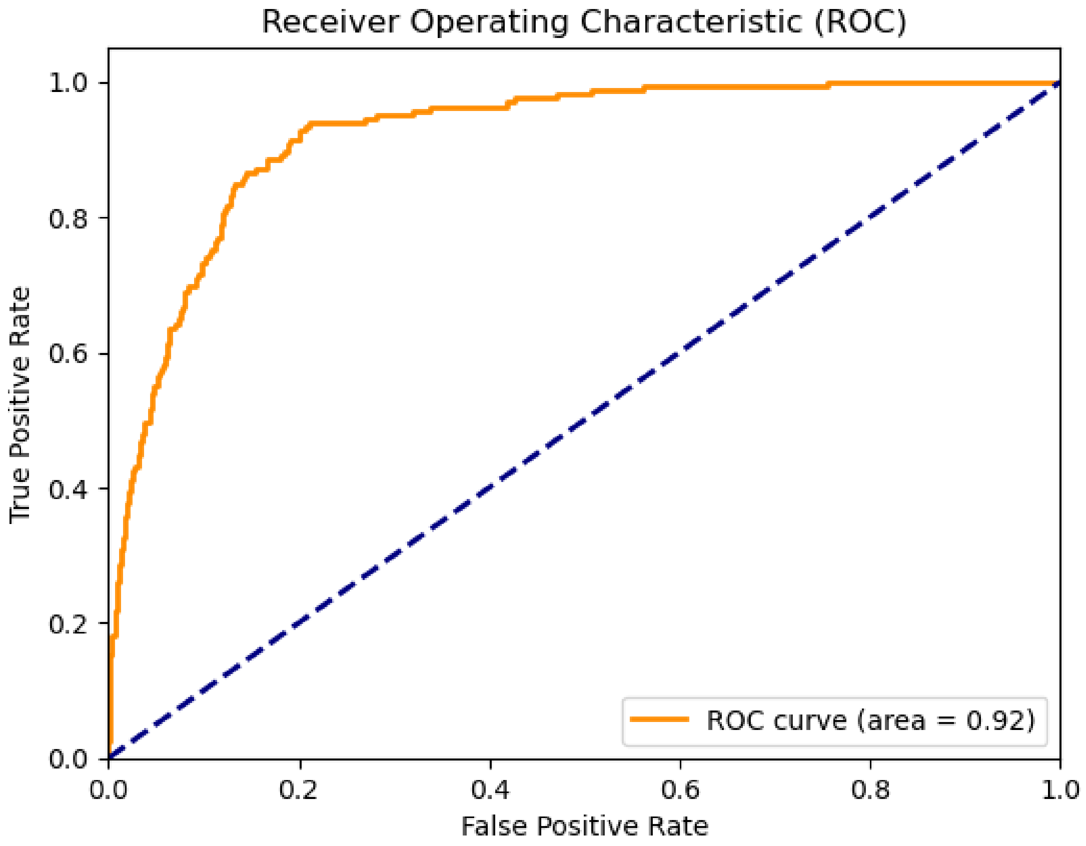


```
Cash_to_Current_Liability                        0.0739      0.075      0.992      0.321      -0.072       0.220
Inventory_to_Working_Capital                    -0.1518      0.143     -1.058      0.290      -0.433       0.129
Inventory_to_Current_Liability                   0.0899      0.124      0.724      0.469      -0.153       0.333
Long_term_Liability_to_Current_Assets           -0.0475      0.108     -0.439      0.661      -0.259       0.165
Retained_Earnings_to_Total_Assets               -0.2175      0.179     -1.215      0.224      -0.568       0.133
Total_income_to_Total_expense                   -2.0469      0.354     -5.783      0.000      -2.741      -1.353
Total_expense_to_Assets                          0.1727      0.206      0.837      0.403      -0.232       0.577
Current_Asset_Turnover_Rate                     -0.1299      0.120     -1.086      0.277      -0.364       0.104
Quick_Asset_Turnover_Rate                        0.0295      0.120      0.247      0.805      -0.205       0.264
Cash_Turnover_Rate                              -0.3696      0.123     -3.015      0.003      -0.610      -0.129
Fixed_Assets_to_Assets                           0.5126     17.419      0.029      0.977     -33.628      34.653
Cash_Flow_to_Total_Assets                        0.9891      0.232      4.269      0.000       0.535       1.443
Cash_Flow_to_Liability                          -2.2925      0.452     -5.077      0.000      -3.177      -1.408
Cash_Flow_to_Equity                              0.0059      0.073      0.082      0.935      -0.136       0.148
Current_Liability_to_Current_Assets             -0.0785      0.115     -0.685      0.494      -0.303       0.146
Total_assets_to_GNP_price                       -0.0347      0.075     -0.463      0.643      -0.182       0.112
No_credit_Interval                               0.1180      0.080      1.476      0.140      -0.039       0.275
Degree_of_Financial_Leverage_DFL                 0.0772      0.055      1.403      0.161      -0.031       0.185
Interest_Coverage_Ratio_Interest_expense_to_EBIT 0.0597      0.082      0.727      0.467      -0.101       0.221
Equity_to_Liability                             -2.9001      0.554     -5.236      0.000      -3.986      -1.814
==============================================================================================================
```

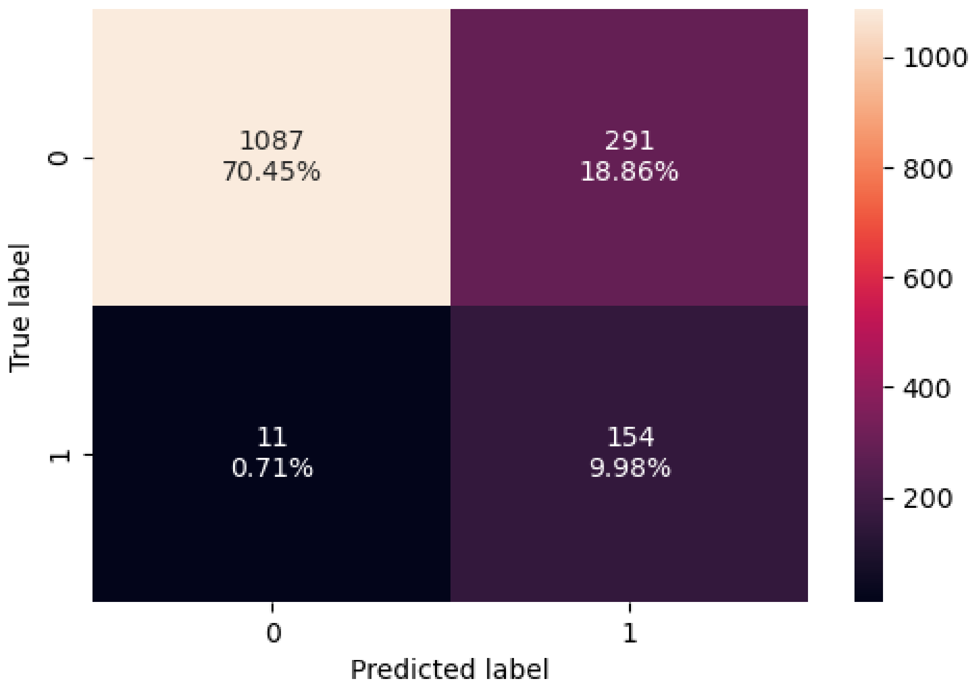
## 12. Identify optimal threshold for Logistic Regression using ROC curve

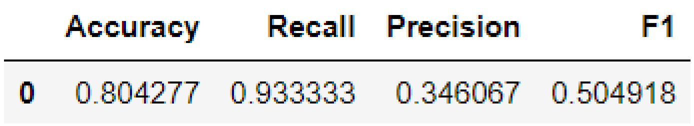
### Finding Optimal Threshold value

```
0.084
```



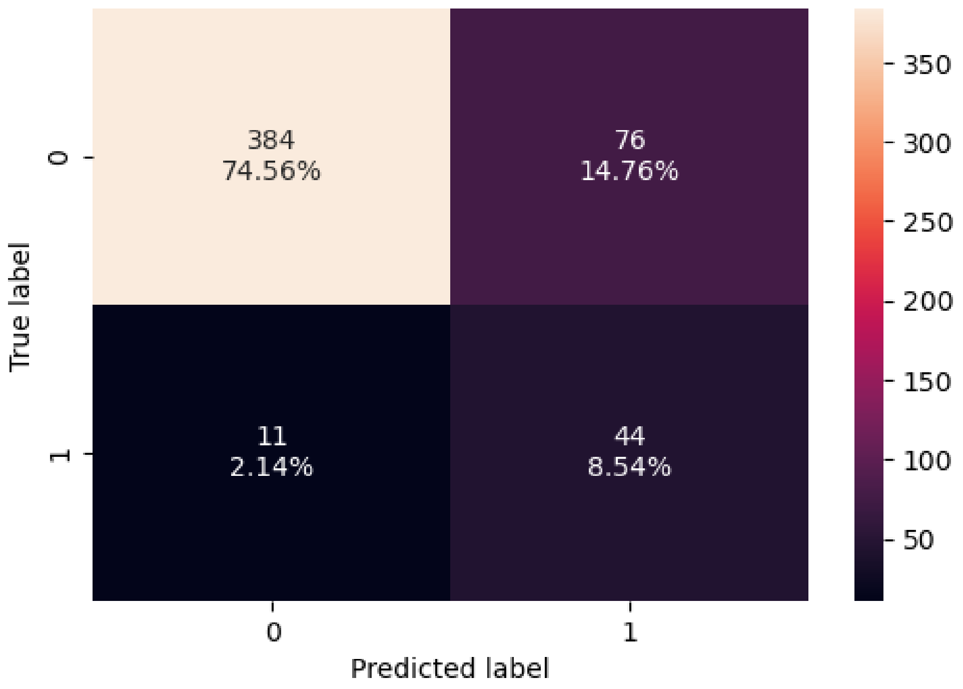
# Logistic Regression Performance - Training Set¶

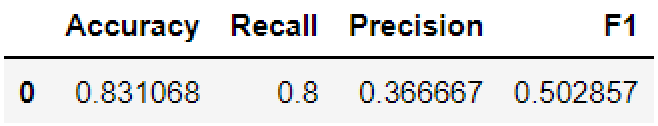




| | Accuracy | Recall | Precision | F1 |
|---|---|---|---|---|
| **0** | 0.804277 | 0.933333 | 0.346067 | 0.504918 |

# Logistic Regression Performance - Test Set





| | Accuracy | Recall | Precision | F1 |
|---|---|---|---|---|
| **0** | 0.831068 | 0.8 | 0.366667 | 0.502857 |

# 13. Hyperparameter Tuning for Random Forest

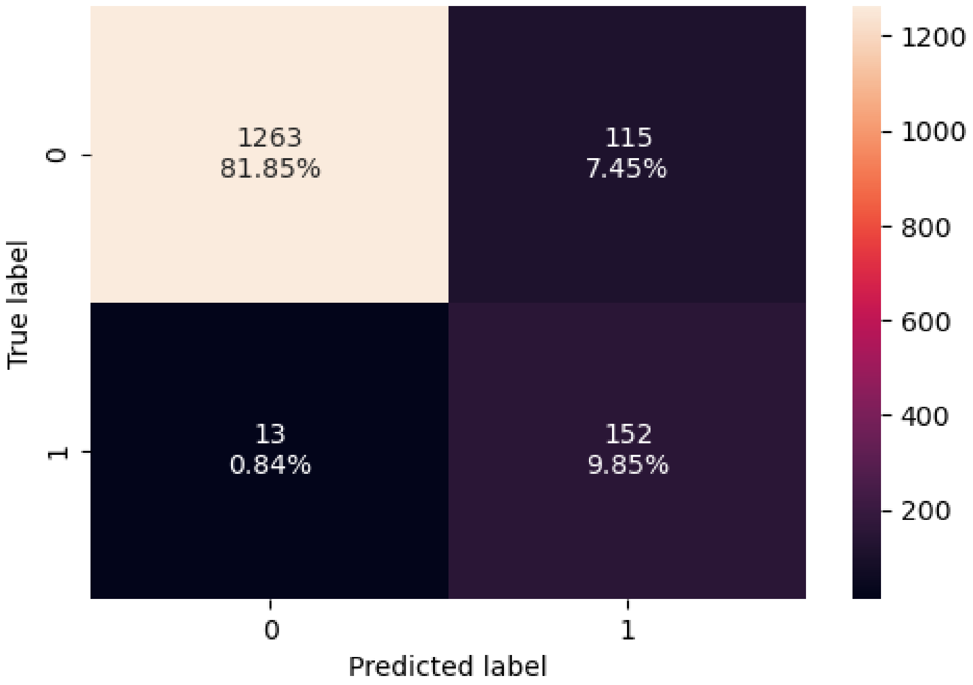
## Model Performance Improvement - Random Forest

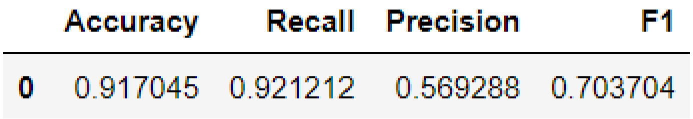
```
Best parameters: {'max_depth': 5, 'min_samples_leaf': 7, 'min_samples_split': 2, 'n_estimators': 200}
```

## Access the best estimator directly if needed

```
Parameters used in the Random Forest Classifier:
bootstrap: True
ccp_alpha: 0.0
class_weight: balanced
criterion: gini
max_depth: 5
max_features: auto
max_leaf_nodes: None
max_samples: None
min_impurity_decrease: 0.0
min_samples_leaf: 7
min_samples_split: 2
min_weight_fraction_leaf: 0.0
n_estimators: 200
n_jobs: None
oob_score: False
random_state: 42
verbose: 0
warm_start: False
```

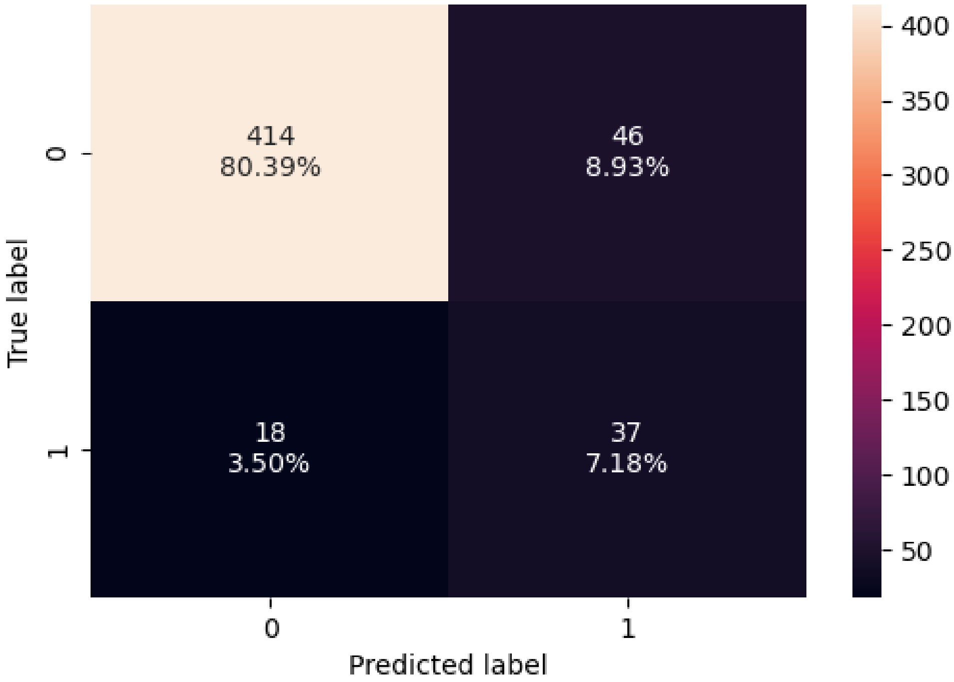
## Random Forest Performance - Training Set

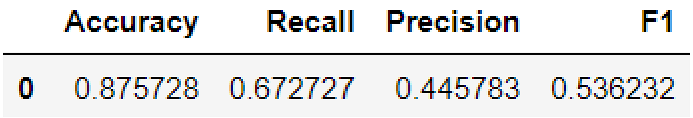




| | Accuracy | Recall | Precision | F1 |
|---|---|---|---|---|
| **0** | 0.917045 | 0.921212 | 0.569288 | 0.703704 |

## Random Forest Performance - Test Set

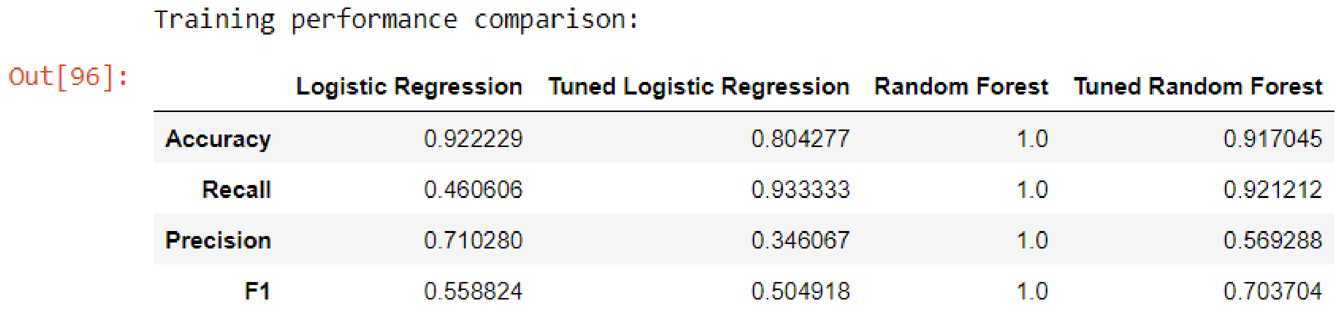




| | Accuracy | Recall | Precision | F1 |
|---|---|---|---|---|
| **0** | 0.875728 | 0.672727 | 0.445783 | 0.536232 |

# PART A: Model Performance Comparison and Final Model Selection

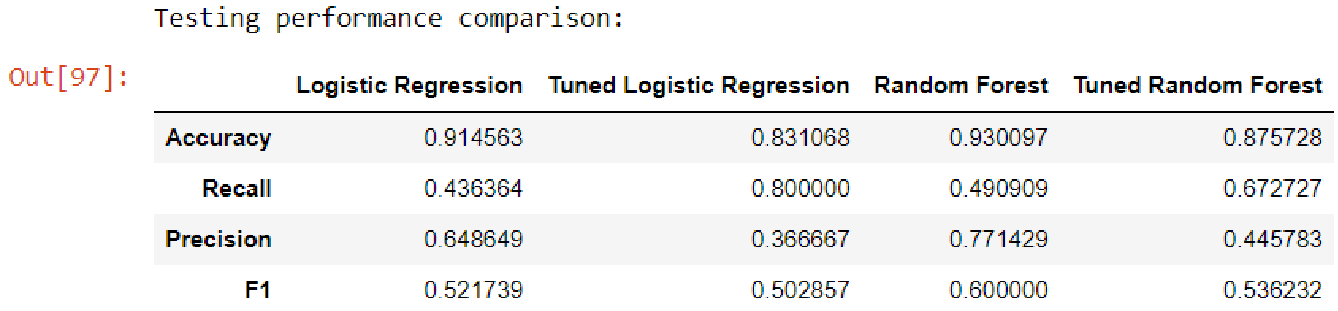
## 14. Compare all the models built



Training performance comparison:

Out[96]:

| | Logistic Regression | Tuned Logistic Regression | Random Forest | Tuned Random Forest |
|---|---|---|---|---|
| Accuracy | 0.922229 | 0.804277 | 1.0 | 0.917045 |
| Recall | 0.460606 | 0.933333 | 1.0 | 0.921212 |
| Precision | 0.710280 | 0.346067 | 1.0 | 0.569288 |
| F1 | 0.558824 | 0.504918 | 1.0 | 0.703704 |



Testing performance comparison:

Out[97]:

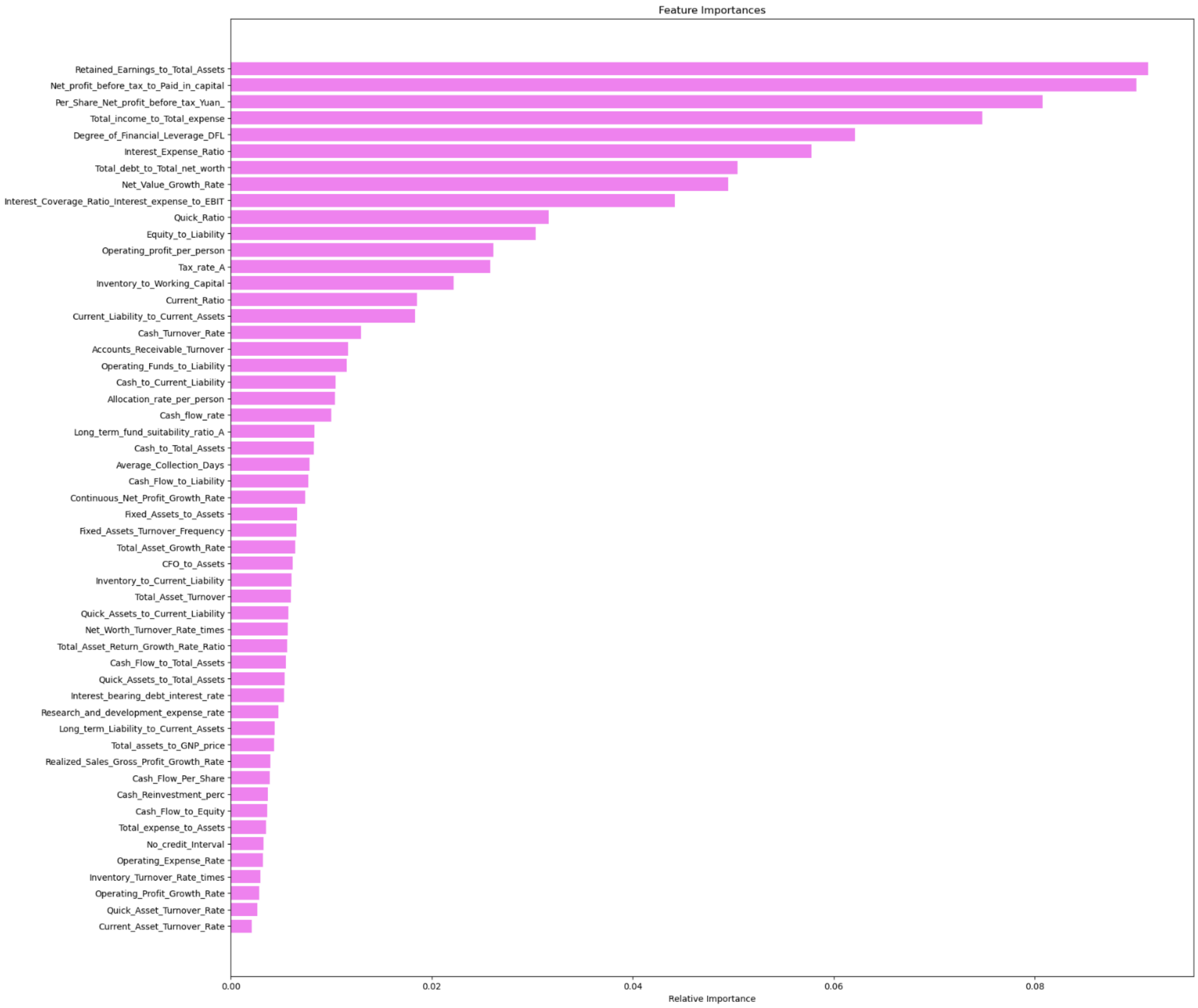
| | Logistic Regression | Tuned Logistic Regression | Random Forest | Tuned Random Forest |
|---|---|---|---|---|
| Accuracy | 0.914563 | 0.831068 | 0.930097 | 0.875728 |
| Recall | 0.436364 | 0.800000 | 0.490909 | 0.672727 |
| Precision | 0.648649 | 0.366667 | 0.771429 | 0.445783 |
| F1 | 0.521739 | 0.502857 | 0.600000 | 0.536232 |

## 15. Select the final model with the proper justification

The Tuned Random Forest model achieves the highest accuracy (0.8757), recall (0.6727), precision (0.4458), and F1 score (0.5362) on the test dataset. These results suggest that the Tuned Random Forest model is a reliable and accurate predictor of the target variable.

Therefore, it is recommended to use the Tuned Random Forest model for financial analysis purposes. For future improvement, consider tuning the model's hyperparameters further, incorporating additional relevant features into the model, or exploring ensemble methods to enhance the model's predictive power.

## 16. Check the most important features in the final model and draw inferences



The most important features in the final model are:
Retained_Earnings_to_Total_Assets
Net_profit_before_tax_to_Paid_in_capital
Per_Share_Net_profit_before_tax_Yuan_
Total_income_to_Total_expense
Degree_of_Financial_Leverage_DFL
Interest_Expense_Ratio
Total_debt_to_Total_net_worth
Net_Value_Growth_Rate
Interest_Coverage_Ratio_Interest_expense_to_EBIT

Inferences:
The model is likely to be good at predicting the financial health of businesses.
The model is likely to be less effective at predicting other aspects of business performance, such as customer satisfaction or employee morale.

**PART A: Actionable Insights & Recommendations**

17. Actionable Insights

- The chart shows the relative importance of different financial metrics for predicting a company's financial health. The metrics with the highest relative importance are:

- Retained Earnings to Total Assets: This metric indicates how much of a company's earnings are being reinvested back into the business. A high value suggests a company is investing in its growth, which is a positive sign.
- Net Profit Before Tax to Paid-in Capital: This metric measures a company's profitability relative to the amount of capital invested by shareholders. A high value indicates that the company is efficiently using its capital to generate profits.
- Per Share Net Profit Before Tax (Yuan): This metric measures the profitability of the company on a per-share basis. A high value suggests that the company is generating significant profits for its shareholders.
- These metrics are important because they provide insights into a company's profitability, growth potential, and financial stability.

Recommendations:

- Focus on improving the metrics with the highest relative importance. This means making sure that the company is reinvesting its earnings wisely, using its capital efficiently, and generating strong profits for its shareholders.

- Use the chart to identify areas where the company can improve. For example, if the company has a low cash flow to total assets ratio, it might need to focus on improving its cash management practices.
- Monitor the performance of the key metrics over time. This will help you track the company's progress and identify any potential problems early on.
- By focusing on the most important financial metrics, companies can improve their chances of achieving financial success.

## PART B: Define the problem and perform Exploratory Data Analysis

### 18. Problem definition

Executive Summary:
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

Introduction:
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.

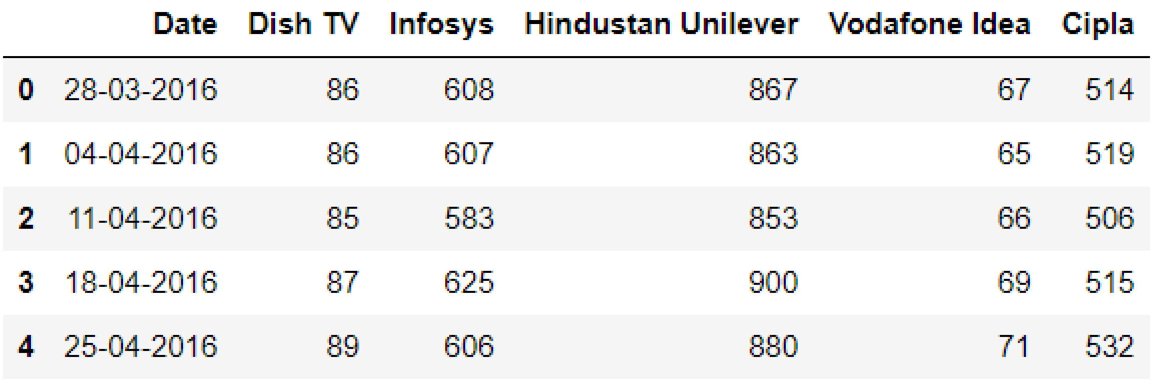
Through this analysis, investors can aim to achieve the following objectives:

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

Data Description:

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.

Sample of the dataset

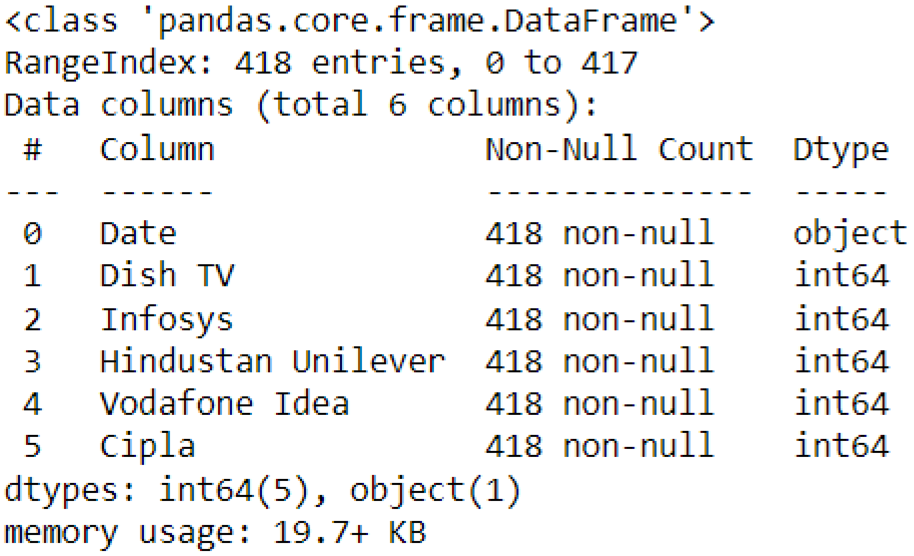


| | Date | Dish TV | Infosys | Hindustan Unilever | Vodafone Idea | Cipla |
|---|---|---|---|---|---|---|
| **0** | 28-03-2016 | 86 | 608 | 867 | 67 | 514 |
| **1** | 04-04-2016 | 86 | 607 | 863 | 65 | 519 |
| **2** | 11-04-2016 | 85 | 583 | 853 | 66 | 506 |
| **3** | 18-04-2016 | 87 | 625 | 900 | 69 | 515 |
| **4** | 25-04-2016 | 89 | 606 | 880 | 71 | 532 |

19. <u>Check shape, Data types, and statistical summary</u>

```
(418, 6)
```

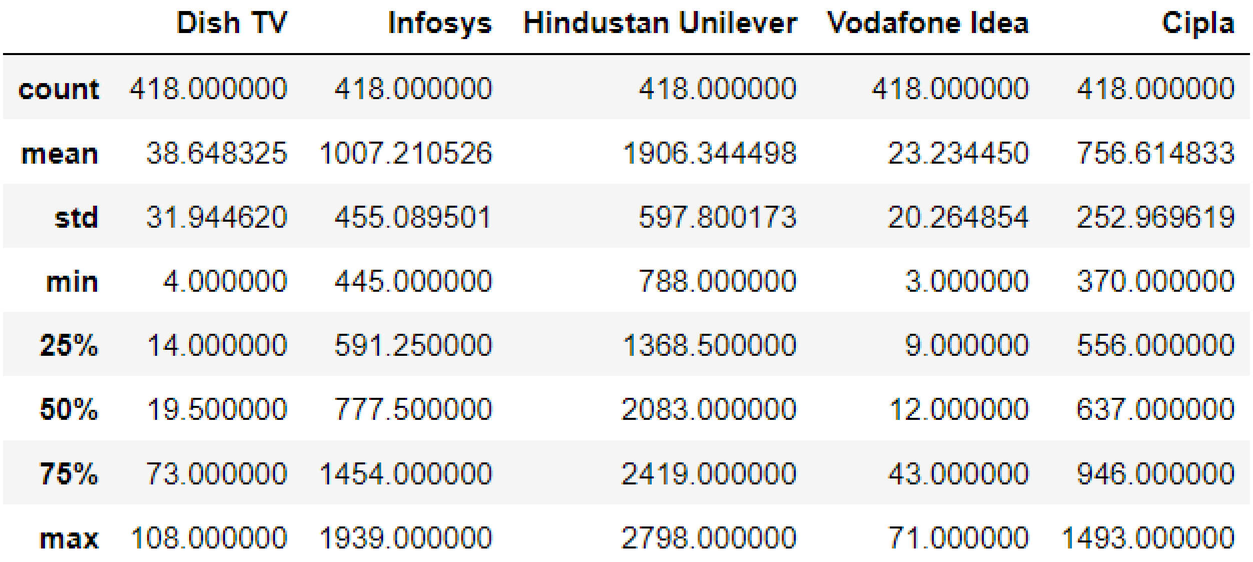
dataset has 418 rows and 6 columns.



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 6 columns):
 #   Column              Non-Null Count  Dtype 
---  ------              --------------  ----- 
 0   Date                418 non-null    object
 1   Dish TV             418 non-null    int64 
 2   Infosys             418 non-null    int64 
 3   Hindustan Unilever  418 non-null    int64 
 4   Vodafone Idea       418 non-null    int64 
 5   Cipla               418 non-null    int64 
dtypes: int64(5), object(1)
memory usage: 19.7+ KB
```

- The dataset has 418 observations (rows) and 6 variables (columns).
- The columns are:

- Date: a string column representing dates
- Dish TV, Infosys, Hindustan Unilever, Vodafone Idea, and Cipla: integer columns representing values for each company



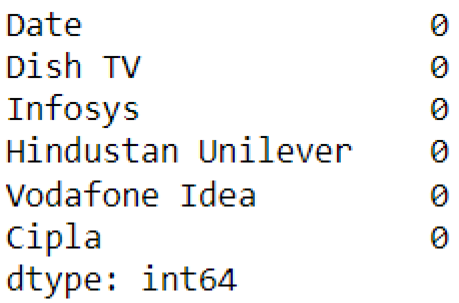
| | Dish TV | Infosys | Hindustan Unilever | Vodafone Idea | Cipla |
|---|---|---|---|---|---|
| **count** | 418.000000 | 418.000000 | 418.000000 | 418.000000 | 418.000000 |
| **mean** | 38.648325 | 1007.210526 | 1906.344498 | 23.234450 | 756.614833 |
| **std** | 31.944620 | 455.089501 | 597.800173 | 20.264854 | 252.969619 |
| **min** | 4.000000 | 445.000000 | 788.000000 | 3.000000 | 370.000000 |
| **25%** | 14.000000 | 591.250000 | 1368.500000 | 9.000000 | 556.000000 |
| **50%** | 19.500000 | 777.500000 | 2083.000000 | 12.000000 | 637.000000 |
| **75%** | 73.000000 | 1454.000000 | 2419.000000 | 43.000000 | 946.000000 |
| **max** | 108.000000 | 1939.000000 | 2798.000000 | 71.000000 | 1493.000000 |

- The dataset has 418 observations and 6 variables.
- The variables are:
- Date: a non-null object (string) column representing dates.
- Dish TV, Infosys, Hindustan Unilever, Vodafone Idea, and Cipla: non-null integer columns representing values for each company.

Here are some key statistics for each variable:

- Dish TV: count=418, mean=38.65, std=31.94, min=4, max=108.
- Infosys: count=418, mean=1007.21, std=455.09, min=445, max=1939.
- Hindustan Unilever: count=418, mean=1906.34, std=597.80, min=788, max=2798.
- Vodafone Idea: count=418, mean=23.23, std=20.26, min=3, max=71.
- Cipla: count=418, mean=756.61, std=252.97, min=370, max=1493.

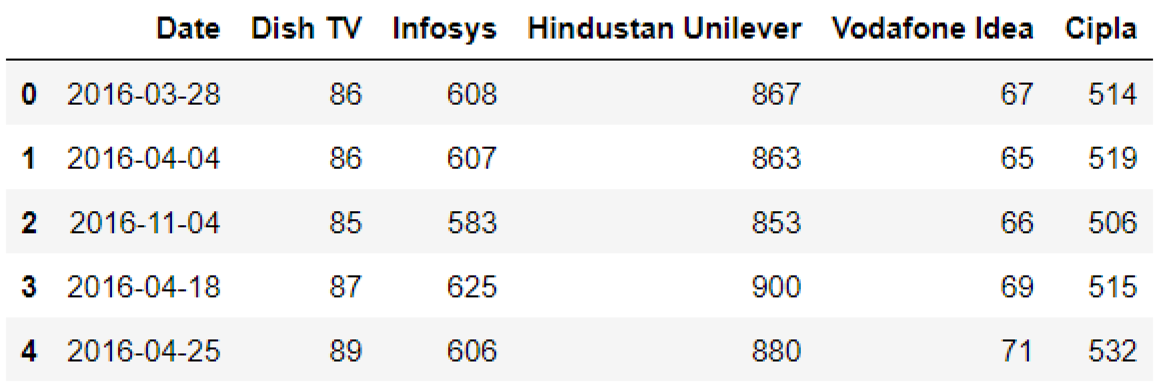
number of null or NaN values in each column



```
Date                 0
Dish TV              0
Infosys              0
Hindustan Unilever   0
Vodafone Idea        0
Cipla                0
dtype: int64
```

- There is no null value.

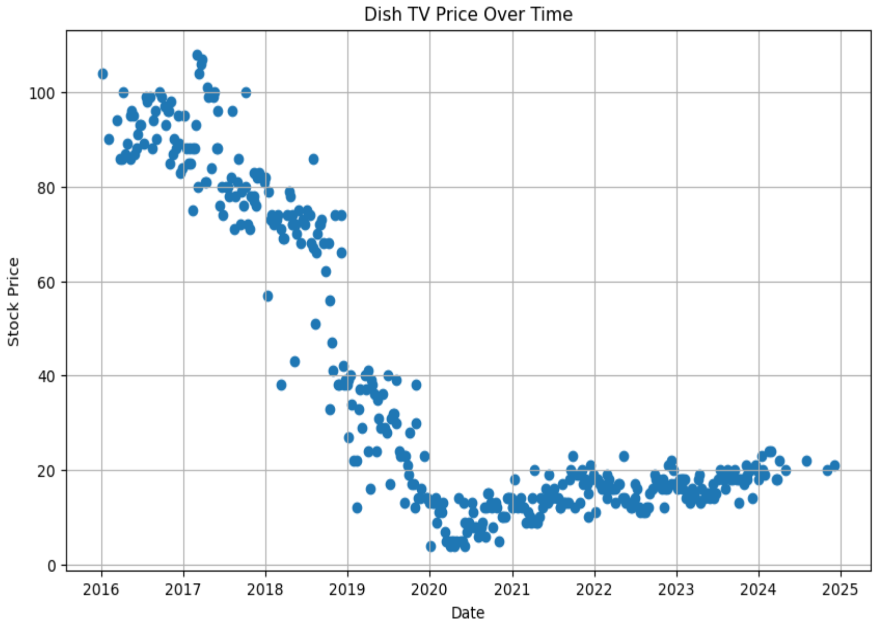
Convert Date column from object to datetime

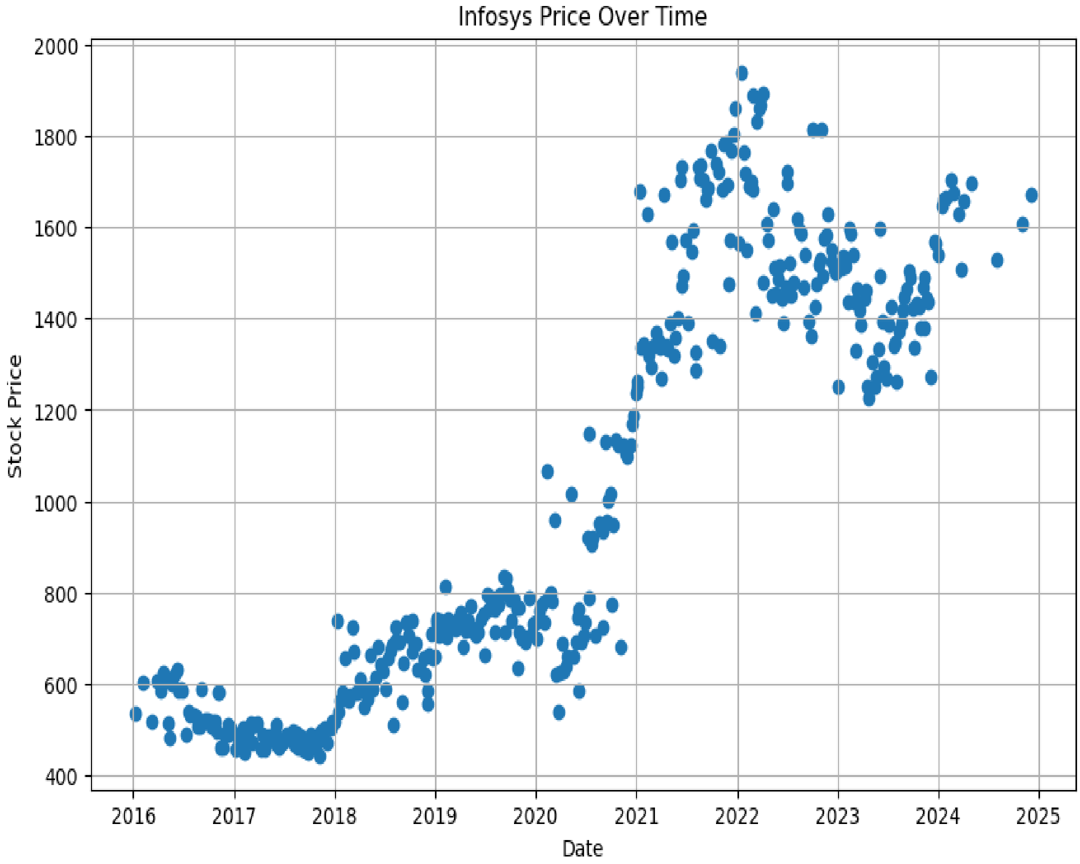


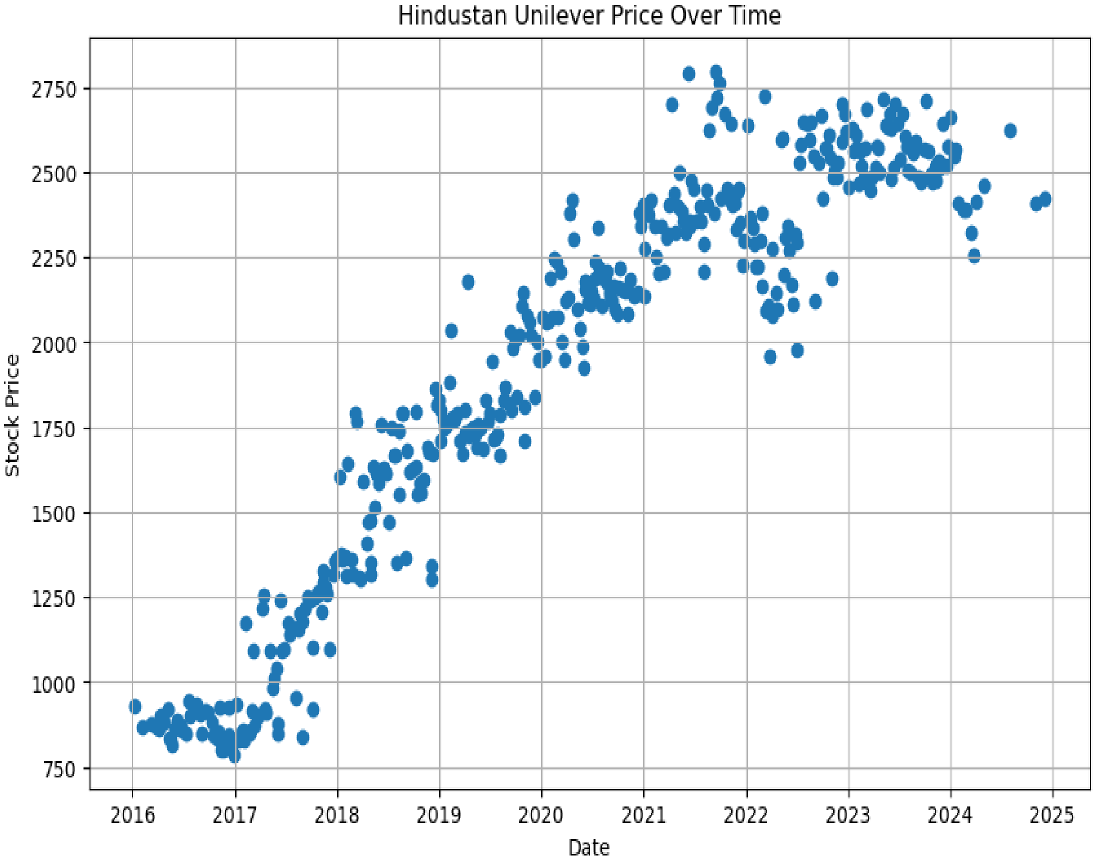
| | Date | Dish TV | Infosys | Hindustan Unilever | Vodafone Idea | Cipla |
|---|---|---|---|---|---|---|
| **0** | 2016-03-28 | 86 | 608 | 867 | 67 | 514 |
| **1** | 2016-04-04 | 86 | 607 | 863 | 65 | 519 |
| **2** | 2016-11-04 | 85 | 583 | 853 | 66 | 506 |
| **3** | 2016-04-18 | 87 | 625 | 900 | 69 | 515 |
| **4** | 2016-04-25 | 89 | 606 | 880 | 71 | 532 |

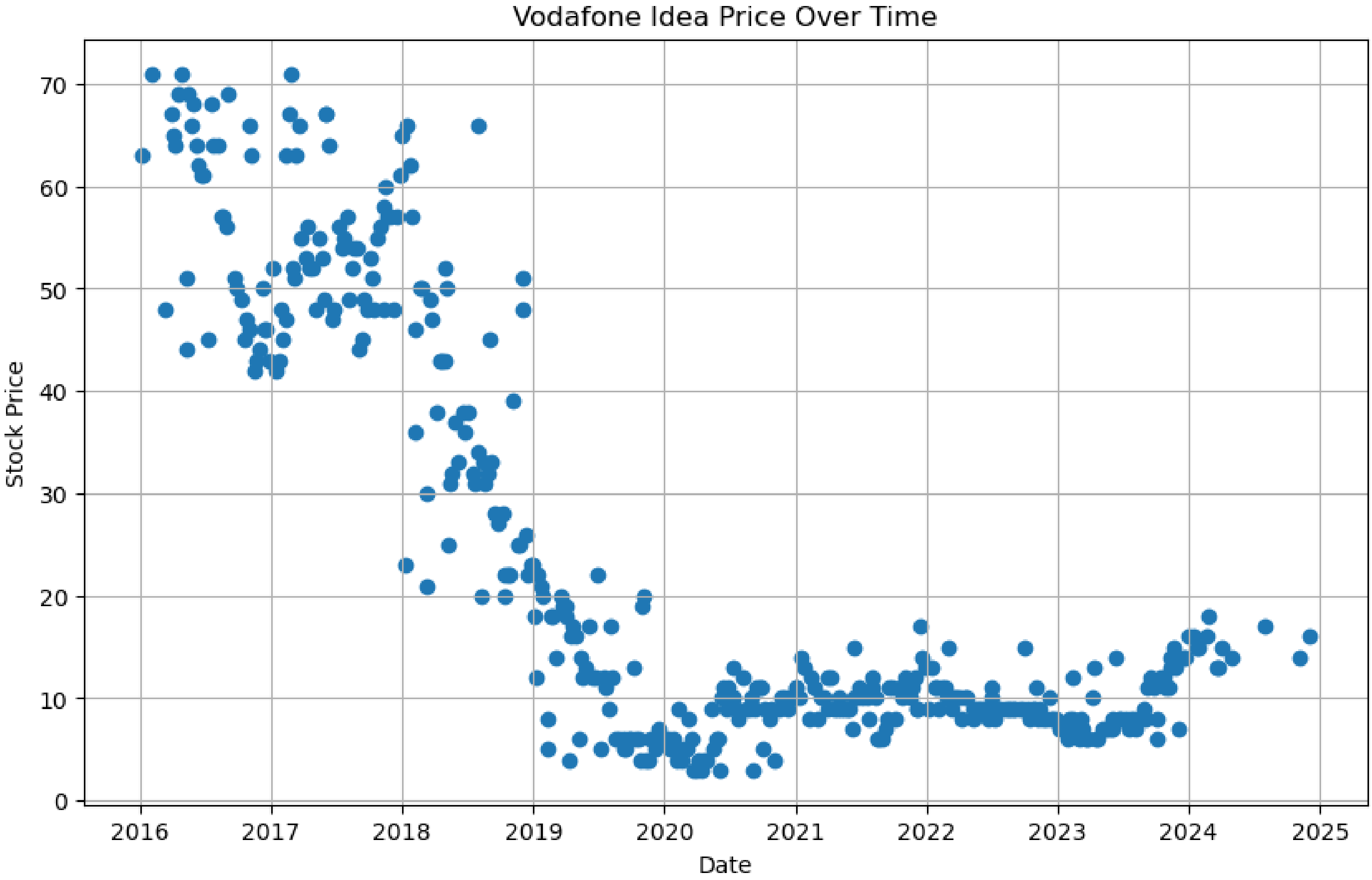
# PART B: Stock Price Graph Analysis

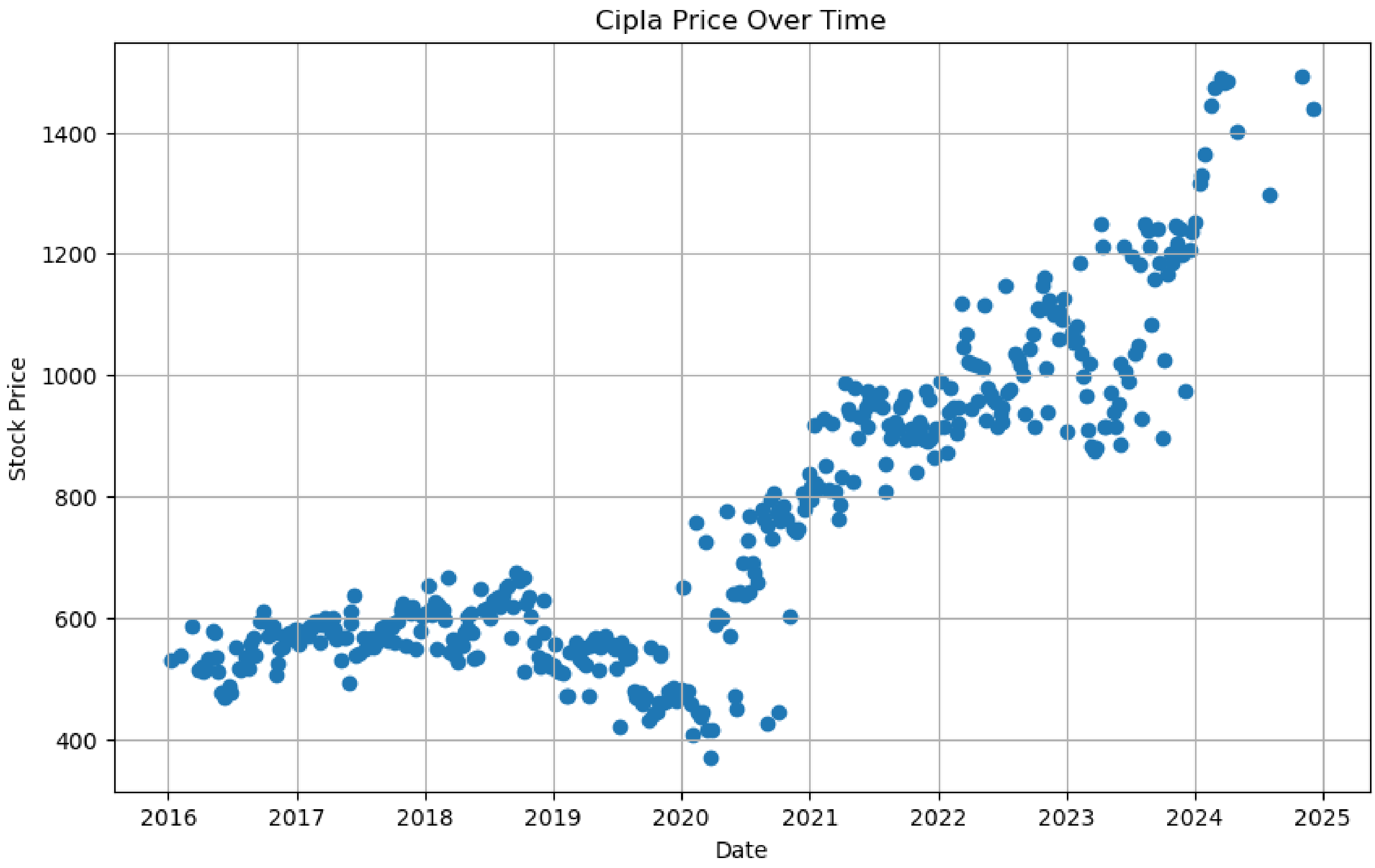
20. Draw Stock Price Graph (Stock Price vs Time) for the given stocks






Hindustan Unilever Price Over Time
Stock Price
2750
2500
2250
2000
1750
1500
1250
1000
750
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
Date


Vodafone Idea Price Over Time
Stock Price
70
60
50
40
30
20
10
0
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
Date



Observations

- The scatter plot shows the stock price of Dish TV over time. The data points show the stock price going down over time, particularly around the year 2019 and onward. The graph shows that the stock price has generally been on a downward trend, with some fluctuations.
- Scatter plot of stock price over time for Infosys. The x-axis represents the date, and the y-axis represents the stock price. The data spans from 2016 to 2025. Overall, there is an upward trend in the stock price. The plot shows some fluctuations and volatility in the stock price over time. The data suggests that Infosys stock has been steadily growing over the past few years.
- Scatter plot of the stock price of Hindustan Unilever over time. The x-axis represents the date, and the y-axis represents the stock price. The plot shows that the stock price has been generally

increasing over time, with some fluctuations.

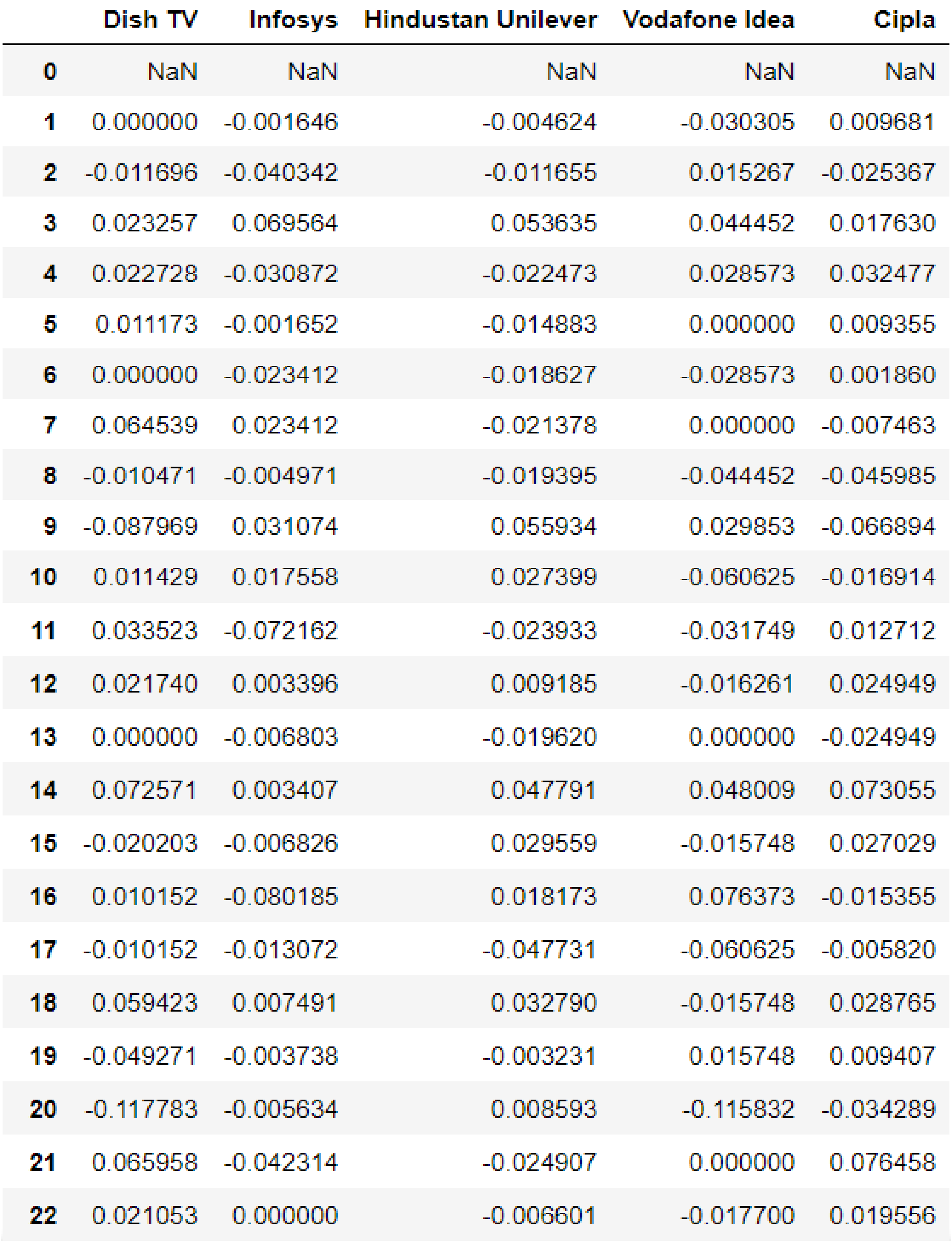
- Scatter plot of the Vodafone Idea stock price over time. The x-axis represents the date, and the y-axis represents the stock price. The plot shows that the stock price has been declining over time. There is a clear drop in price in 2019, and the price has been fluctuating around a lower level since then.
- Scatter plot of Cipla's stock price over time, from 2016 to 2025. The stock price generally increased over time, with some fluctuations. The plot can be used to analyze the stock price trends and identify any potential patterns.

# PART B: Stock Returns Calculation and Analysis

## 21. Calculate Returns for all stocks

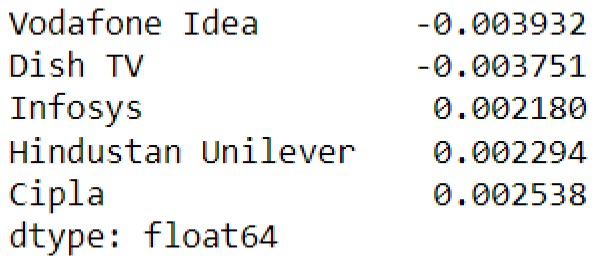
Returns and Volatility Analysis

- Return Calculation



| | Dish TV | Infosys | Hindustan Unilever | Vodafone Idea | Cipla |
|---|---|---|---|---|---|
| **0** | NaN | NaN | NaN | NaN | NaN |
| **1** | 0.000000 | -0.001646 | -0.004624 | -0.030305 | 0.009681 |
| **2** | -0.011696 | -0.040342 | -0.011655 | 0.015267 | -0.025367 |
| **3** | 0.023257 | 0.069564 | 0.053635 | 0.044452 | 0.017630 |
| **4** | 0.022728 | -0.030872 | -0.022473 | 0.028573 | 0.032477 |
| **5** | 0.011173 | -0.001652 | -0.014883 | 0.000000 | 0.009355 |
| **6** | 0.000000 | -0.023412 | -0.018627 | -0.028573 | 0.001860 |
| **7** | 0.064539 | 0.023412 | -0.021378 | 0.000000 | -0.007463 |
| **8** | -0.010471 | -0.004971 | -0.019395 | -0.044452 | -0.045985 |
| **9** | -0.087969 | 0.031074 | 0.055934 | 0.029853 | -0.066894 |
| **10** | 0.011429 | 0.017558 | 0.027399 | -0.060625 | -0.016914 |
| **11** | 0.033523 | -0.072162 | -0.023933 | -0.031749 | 0.012712 |
| **12** | 0.021740 | 0.003396 | 0.009185 | -0.016261 | 0.024949 |
| **13** | 0.000000 | -0.006803 | -0.019620 | 0.000000 | -0.024949 |
| **14** | 0.072571 | 0.003407 | 0.047791 | 0.048009 | 0.073055 |
| **15** | -0.020203 | -0.006826 | 0.029559 | -0.015748 | 0.027029 |
| **16** | 0.010152 | -0.080185 | 0.018173 | 0.076373 | -0.015355 |
| **17** | -0.010152 | -0.013072 | -0.047731 | -0.060625 | -0.005820 |
| **18** | 0.059423 | 0.007491 | 0.032790 | -0.015748 | 0.028765 |
| **19** | -0.049271 | -0.003738 | -0.003231 | 0.015748 | 0.009407 |
| **20** | -0.117783 | -0.005634 | 0.008593 | -0.115832 | -0.034289 |
| **21** | 0.065958 | -0.042314 | -0.024907 | 0.000000 | 0.076458 |
| **22** | 0.021053 | 0.000000 | -0.006601 | -0.017700 | 0.019556 |

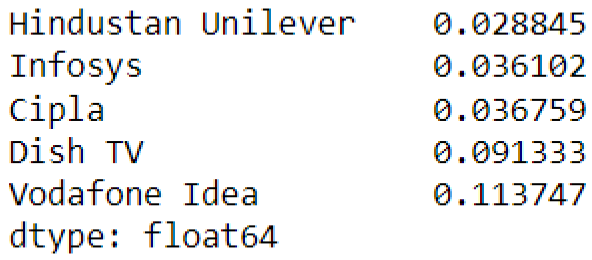
- Average Returns



```
Vodafone Idea        -0.003932
Dish TV              -0.003751
Infosys               0.002180
Hindustan Unilever    0.002294
Cipla                 0.002538
dtype: float64
```

  - Vodafone Idea has the lowest average return at -0.003932, indicating underperformance.
  - Dish TV has the second-lowest average return at -0.003751, also indicating underperformance.
  - Infosys has the third-highest average return at 0.002180, indicating outperformance.
  - Hindustan Unilever has the second-highest average return at 0.002294, also indicating outperformance.
  - Cipla has the highest average return at 0.002538, indicating the highest level of outperformance among the five stocks.
  - The average returns for all five stocks are relatively small, indicating stable performance over the given period.

- Volatility

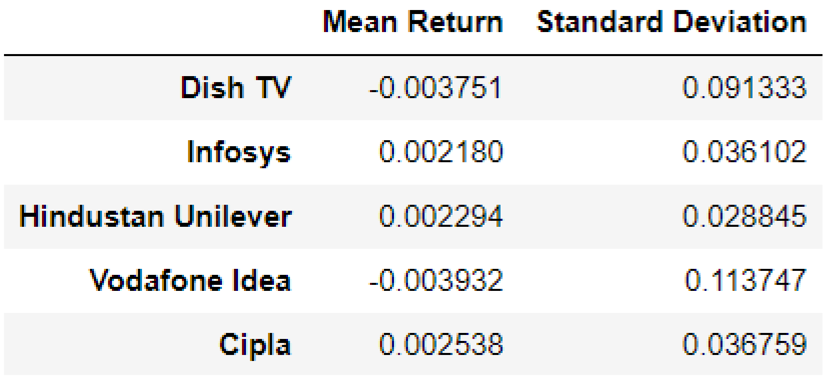


```
Hindustan Unilever    0.028845
Infosys               0.036102
Cipla                 0.036759
Dish TV               0.091333
Vodafone Idea         0.113747
dtype: float64
```

- Vodafone Idea is that it has the highest volatility at 0.113747, indicating a higher risk compared to the other four stocks.
- Dish TV is that it has the highest volatility at 0.091333, indicating a higher risk compared to the other four stocks.
- Infosys is that it has the second-lowest volatility at 0.036102, indicating a lower risk compared to the other four stocks.
- Hindustan Unilever is that it has the lowest volatility at 0.028845, indicating a lower risk compared to the other four stocks.
- Cipla is that it has the third-lowest volatility at 0.036759, indicating a lower risk compared to the other four stocks.

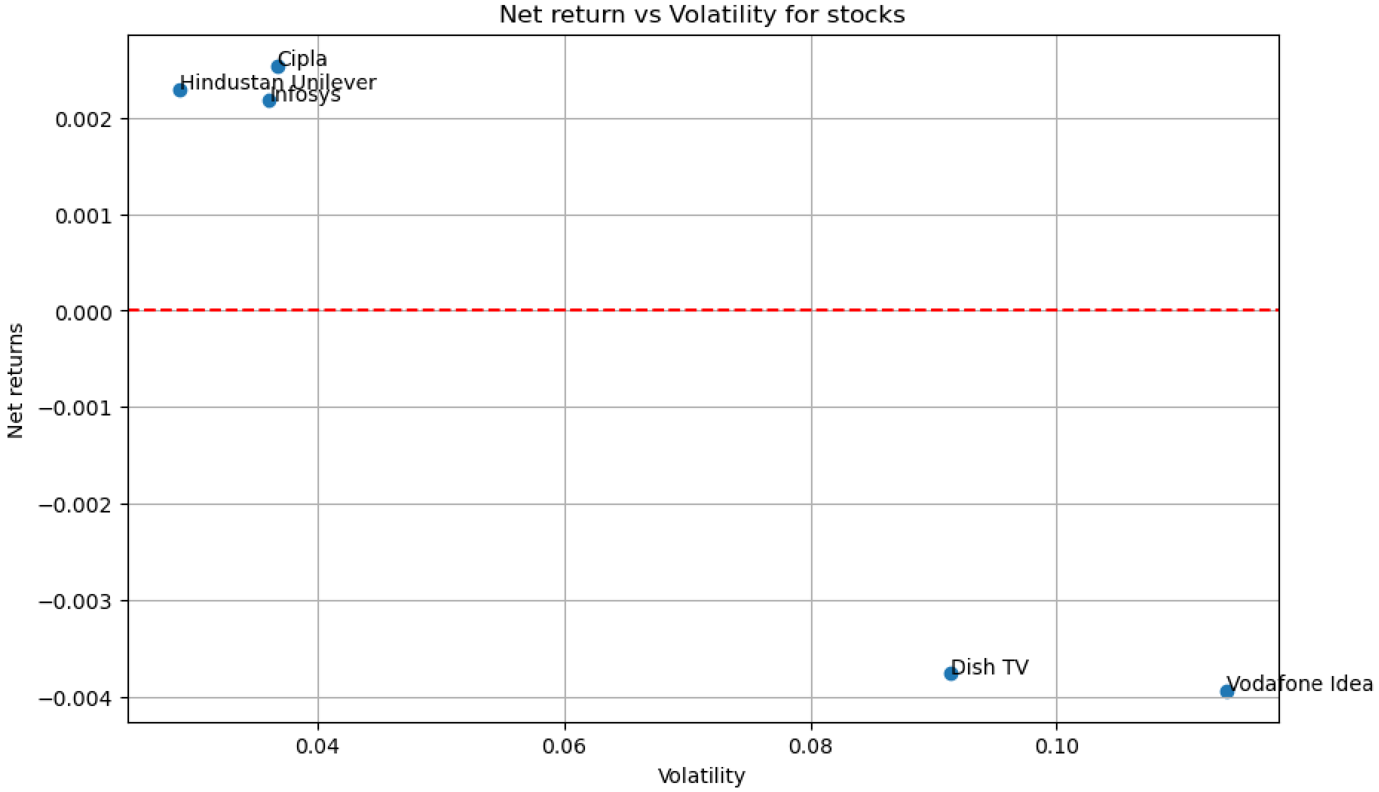
## 22. <u>Calculate the Mean and Standard Deviation for the returns of all stocks</u>

Visualizing Returns and Volatility



| | Mean Return | Standard Deviation |
|---|---|---|
| **Dish TV** | -0.003751 | 0.091333 |
| **Infosys** | 0.002180 | 0.036102 |
| **Hindustan Unilever** | 0.002294 | 0.028845 |
| **Vodafone Idea** | -0.003932 | 0.113747 |
| **Cipla** | 0.002538 | 0.036759 |

## 23. Draw a plot of Mean vs Standard Deviation for all stock returns



Inferences:

- Risk and return are positively correlated: Stocks with higher volatility (risk) tend to have lower returns.
- Investors are risk-averse: Investors are willing to accept lower returns for lower risk.
- There is a trade-off between risk and return: Investors must choose a portfolio that balances their desired level of risk with their desired level of return.
- This graph effectively conveys these inferences. The scatter plot shows that stocks with higher volatility (risk) tend to have lower returns. This confirms the principle of risk and return being positively correlated.

## PART B: Actionable Insights & Recommendations

### 24. Actionable insights and recommendations

- Investors seeking higher returns should consider stocks with lower volatility (risk), as they tend to have higher returns.
- Investors who are risk-averse should consider stocks with lower volatility (risk), even if the returns are lower.
- When building a portfolio, it's important to balance the desired level of risk with the desired level of return.
- Regularly reviewing the risk and return of a portfolio can help ensure it remains aligned with the investor's goals and risk tolerance.
- Diversification can help reduce the overall risk of a portfolio, as the risk of one investment may be offset by the performance of another.