Finance and Risk Analytics Project :

-Prapthi Pandian

Table of Contents

1. Part A-

1.1 Problem Statement-

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

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

1.2 Summary –

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	_Per_Share_Net_profit_before_tax_Yuan_	_Realized_Sale
0	16974	Hind.Cables	8820000000.00	0.00	0.46	0.00	0.00	0.32	
1	21214	Tata Tele. Mah.	9380000000.00	4230000000.00	0.46	0.00	0.00	0.32	
2	14852	ABG Shipyard	3800000000.00	815000000.00	0.45	0.00	0.00	0.30	
3	2439	GTL	6440000000.00	0.00	0.46	0.00	0.01	0.32	
4	23505	Bharati Defence	3680000000.00	0.00	0.46	0.00	0.40	0.33	

5 rows × 58 columns

Shape-

```
The number of rows (observations) is 2058
The number of columns (variables) is 58
```

Summary -

```
<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 dataset contains 2058 rows (observations) and 58 columns (variables).
- The majority of columns are of float type (53 columns), followed by int64 (4 columns) and 1 column of object type.

- Some columns have missing values (NaN):
- The column ‘Default’ is the target variable and contains binary values indicating default or non-default.
- Columns ‘_Liability_Assets_Flag’ and ‘_Net_Income_Flag’ are binary flags containing values 0 or 1.

We have dropped columns- 'Co_Code','Co_Name' for our analysis.

Descriptive statistics -

	_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
count	2058.00	2058.00	2058.00	2058.00	2058.00	1891.00	
mean	2052388835.76	1208634256.56	0.47	11130223.52	0.11	0.32	
std	3252623690.29	2144568158.08	0.02	90425949.04	0.15	0.02	
min	0.00	0.00	0.00	0.00	0.00	0.17	
25%	0.00	0.00	0.46	0.00	0.00	0.31	
50%	0.00	0.00	0.46	0.00	0.04	0.32	
75%	4110000000.00	1550000000.00	0.47	0.00	0.22	0.33	
max	9980000000.00	9980000000.00	1.00	990000000.00	1.00	0.46	

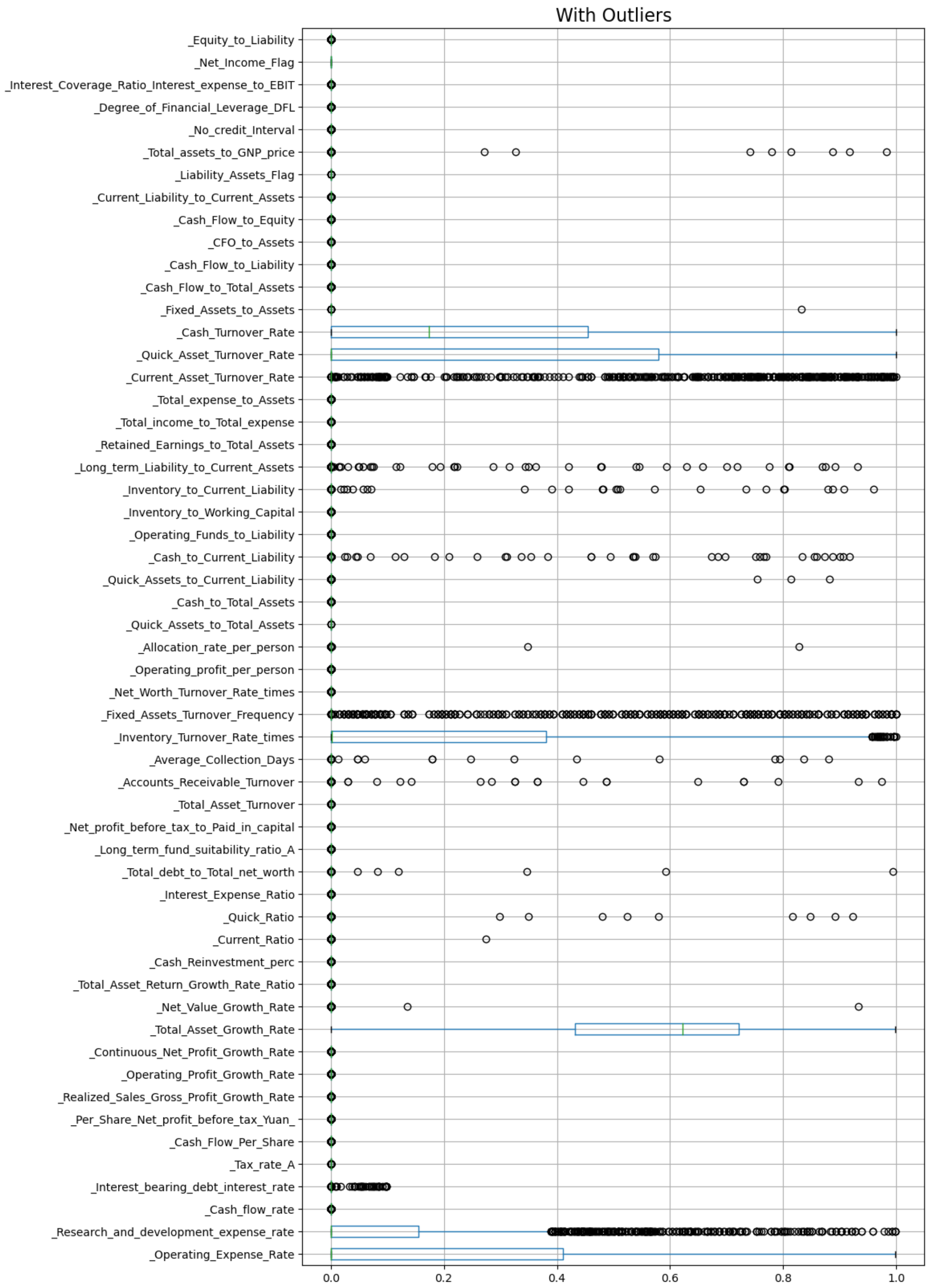
No. of defaulters-

```
Default
0    1838
1     220
Name: count, dtype: int64

Default
0   0.89
1   0.11
Name: proportion, dtype: float64
```

- We can observe that 11% of the company is defaulting.

1.3 Outliers-


With Outliers
_Equity_to_Liability
_Net_Income_Flag
_Interest_Coverage_Ratio_Interest_expense_to_EBIT
_Degree_of_Financial_Leverage_DFL
_No_credit_Interval
_Total_assets_to_GNP_price
_Liability_Assets_Flag
_Current_Liability_to_Current_Assets
_Cash_Flow_to_Equity
_CFO_to_Assets
_Cash_Flow_to_Liability
_Cash_Flow_to_Total_Assets
_Fixed_Assets_to_Assets
_Cash_Turnover_Rate
_Quick_Asset_Turnover_Rate
_Current_Asset_Turnover_Rate
_Total_expense_to_Assets
_Total_income_to_Total_expense
_Retained_Earnings_to_Total_Assets
_Long_term_Liability_to_Current_Assets
_Inventory_to_Current_Liability
_Inventory_to_Working_Capital
_Operating_Funds_to_Liability
_Cash_to_Current_Liability
_Quick_Assets_to_Current_Liability
_Cash_to_Total_Assets
_Quick_Assets_to_Total_Assets
_Allocation_rate_per_person
_Operating_profit_per_person
_Net_Worth_Turnover_Rate_times
_Fixed_Assets_Turnover_Frequency
_Inventory_Turnover_Rate_times
_Average_Collection_Days
_Accounts_Receivable_Turnover
_Total_Asset_Turnover
_Net_profit_before_tax_to_Paid_in_capital
_Long_term_fund_suitability_ratio_A
_Total_debt_to_Total_net_worth
_Interest_Expense_Ratio
_Quick_Ratio
_Current_Ratio
_Cash_Reinvestment_perc
_Total_Asset_Return_Growth_Rate_Ratio
_Net_Value_Growth_Rate
_Total_Asset_Growth_Rate
_Continuous_Net_Profit_Growth_Rate
_Operating_Profit_Growth_Rate
_Realized_Sales_Gross_Profit_Growth_Rate
_Per_Share_Net_profit_before_tax_Yuan_
_Cash_Flow_Per_Share
_Tax_rate_A
_Interest_bearing_debt_interest_rate
_Cash_flow_rate
_Research_and_development_expense_rate
_Operating_Expense_Rate
0.0
0.2
0.4
0.6
0.8
1.0
1e10

Post outlier treatment –

Outliers are identified and removed based on the IQR method. They are replaced by the lower range and upper range range.

1.4 **Missing values-**

```
_Operating_Expense_Rate                                0
_Research_and_development_expense_rate                 0
_Cash_flow_rate                                        0
_Interest_bearing_debt_interest_rate                   0
_Tax_rate_A                                            0
_Cash_Flow_Per_Share                                 167
_Per_Share_Net_profit_before_tax_Yuan_                 0
_Realized_Sales_Gross_Profit_Growth_Rate               0
_Operating_Profit_Growth_Rate                          0
_Continuous_Net_Profit_Growth_Rate                     0
_Total_Asset_Growth_Rate                               0
_Net_Value_Growth_Rate                                 0
_Total_Asset_Return_Growth_Rate_Ratio                  0
_Cash_Reinvestment_perc                                0
_Current_Ratio                                         0
_Quick_Ratio                                           0
_Interest_Expense_Ratio                                0
_Total_debt_to_Total_net_worth                        21
_Long_term_fund_suitability_ratio_A                    0
_Net_profit_before_tax_to_Paid_in_capital              0
_Total_Asset_Turnover                                  0
_Accounts_Receivable_Turnover                          0
_Average_Collection_Days                               0
_Inventory_Turnover_Rate_times                         0
_Fixed_Assets_Turnover_Frequency                       0
_Net_Worth_Turnover_Rate_times                         0
_Operating_profit_per_person                           0
_Allocation_rate_per_person                            0
_Quick_Assets_to_Total_Assets                          0
_Cash_to_Total_Assets                                 96
_Quick_Assets_to_Current_Liability                     0
_Cash_to_Current_Liability                             0
_Operating_Funds_to_Liability                          0
_Inventory_to_Working_Capital                          0
_Inventory_to_Current_Liability                        0
_Long_term_Liability_to_Current_Assets                 0
_Retained_Earnings_to_Total_Assets                     0
_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                  14
_Liability_Assets_Flag                                 0
_Total_assets_to_GNP_price                             0
_No_credit_Interval                                    0
_Degree_of_Financial_Leverage_DFL                      0
_Interest_Coverage_Ratio_Interest_expense_to_EBIT      0
_Net_Income_Flag                                       0
_Equity_to_Liability                                   0
Default                                                0
dtype: int64
```

- Approximately 0.25% of the data is missing.

Treating null values-

- StandardScaler is used to scale the features.
- KNNImputer is used to impute missing values in both the training and testing sets separately.
- The imputation is performed using k-nearest neighbors algorithm with n_neighbors=5.

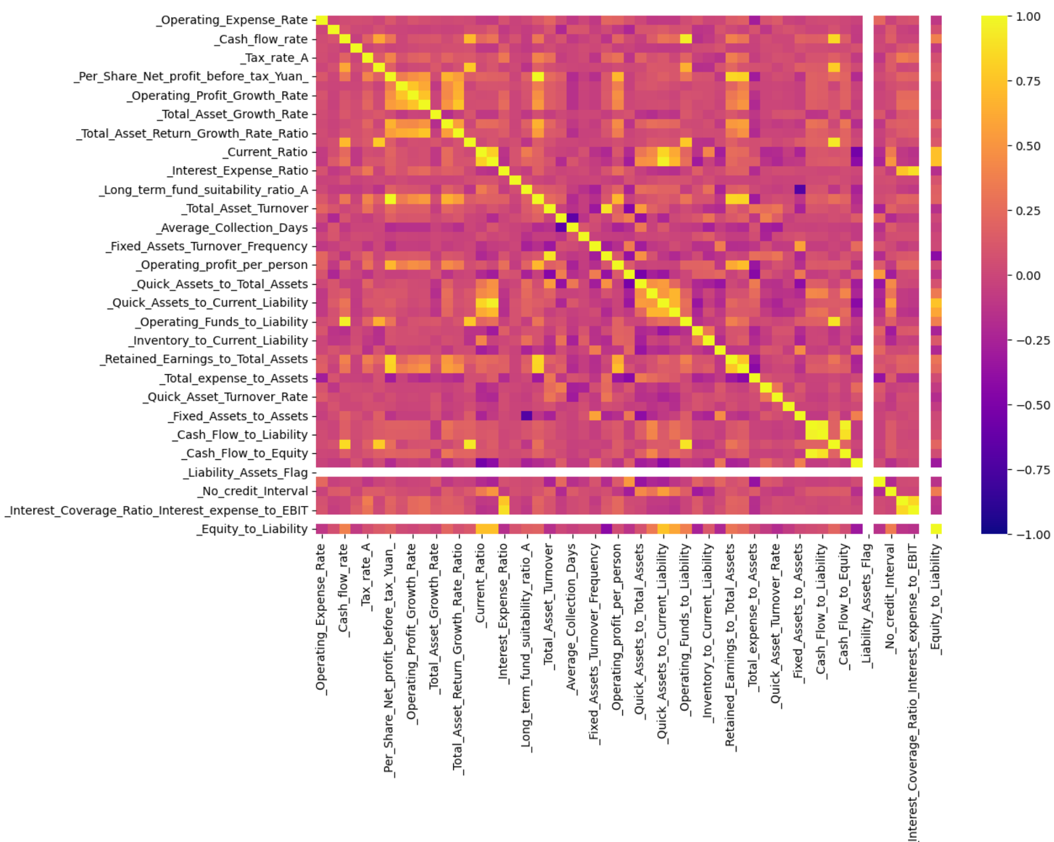
Printing null values post imputation-

```
No. of missing values in imputed train set: 0
No. of missing values in imputed test set: 0
```

1.5 Train-Test Split

Post scaling the features, the data is split into training and testing sets with random state= 42, such that the training set contains 67% of the data and the test set contains the remaining 33%.

Heatmap of the correlation matrix of the features in the training dataset after imputation-



We can observe certain variables being highly correlated with the other.

To avoid multicollinearity, we are calculating the VIF factor.

VIF - explains how good independent variable can be defined as a linear combination of other independent variables.

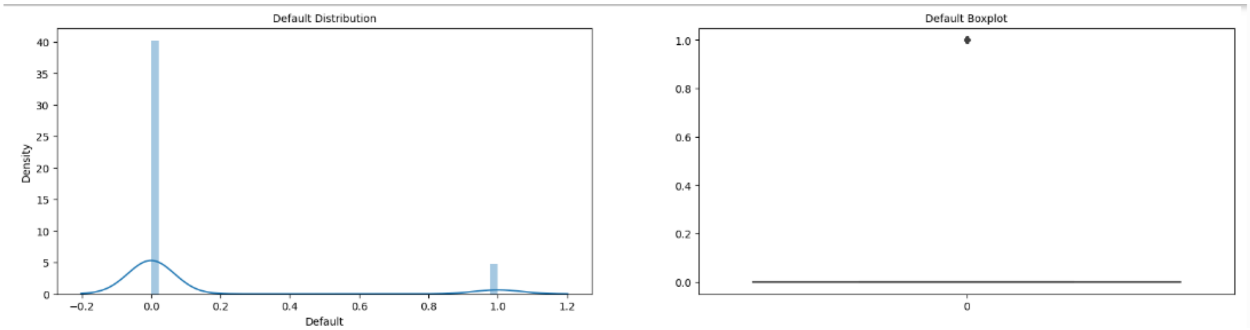
If VIF > 5 for a variable, we can eliminate it to avoid redundancy.

```
Removing '_Per_Share_Net_profit_before_tax_Yuan_' with highest VIF value of 105.10119148199128
Removing '_Cash_Flow_to_Total_Assets' with highest VIF value of 56.22953401156523
Removing '_Quick_Assets_to_Current_Liability' with highest VIF value of 32.31624463076571
Removing '_CFO_to_Assets' with highest VIF value of 25.699898439953564
Removing '_Operating_Funds_to_Liability' with highest VIF value of 18.714442671288015
Removing '_Total_Asset_Turnover' with highest VIF value of 11.11368526613855
Removing '_Current_Ratio' with highest VIF value of 9.411603968266595
Removing '_Net_profit_before_tax_to_Paid_in_capital' with highest VIF value of 7.821382662983197
Removing '_Interest_Coverage_Ratio_Interest_expense_to_EBIT' with highest VIF value of 6.922579024462722
Removing '_Cash_Flow_to_Equity' with highest VIF value of 5.428998718270422
Removing '_Quick_Assets_to_Total_Assets' with highest VIF value of 5.197089894746536

Final VIF Results:
                                     Feature  VIF
35                     _Fixed_Assets_to_Assets 4.38
30              _Total_income_to_Total_expense 4.12
13                                _Quick_Ratio 4.06
43                      _Equity_to_Liability 4.00
7               _Operating_Profit_Growth_Rate 3.75
12                     _Cash_Reinvestment_perc 3.72
8          _Continuous_Net_Profit_Growth_Rate 3.50
29         _Retained_Earnings_to_Total_Assets 3.48
2                              _Cash_flow_rate 3.32
25                 _Cash_to_Current_Liability 3.29
11      _Total_Asset_Return_Growth_Rate_Ratio 3.10
6    _Realized_Sales_Gross_Profit_Growth_Rate 2.92
22                _Operating_profit_per_person 2.90
16         _Long_term_fund_suitability_ratio_A 2.83
23                 _Allocation_rate_per_person 2.78
21              _Net_Worth_Turnover_Rate_times 2.76
5                         _Cash_Flow_Per_Share 2.75
24                      _Cash_to_Total_Assets 2.55
10                      _Net_Value_Growth_Rate 2.54
14                      _Interest_Expense_Ratio 2.51
17             _Accounts_Receivable_Turnover 2.50
41           _Degree_of_Financial_Leverage_DFL 2.49
18                    _Average_Collection_Days 2.25
31                    _Total_expense_to_Assets 2.13
20           _Fixed_Assets_Turnover_Frequency 1.92
39                   _Total_assets_to_GNP_price 1.77
27            _Inventory_to_Current_Liability 1.70
28     _Long_term_Liability_to_Current_Assets 1.66
37       _Current_Liability_to_Current_Assets 1.65
40                         _No_credit_Interval 1.60
32                _Current_Asset_Turnover_Rate 1.56
4                                  _Tax_rate_A 1.47
26              _Inventory_to_Working_Capital 1.47
33                  _Quick_Asset_Turnover_Rate 1.40
36                     _Cash_Flow_to_Liability 1.36
0                       _Operating_Expense_Rate 1.31
19             _Inventory_Turnover_Rate_times 1.23
1       _Research_and_development_expense_rate 1.19
9                      _Total_Asset_Growth_Rate 1.16
3         _Interest_bearing_debt_interest_rate 1.10
34                          _Cash_Turnover_Rate 1.10
15              _Total_debt_to_Total_net_worth 1.06
38                      _Liability_Assets_Flag  NaN
42                            _Net_Income_Flag  NaN
```

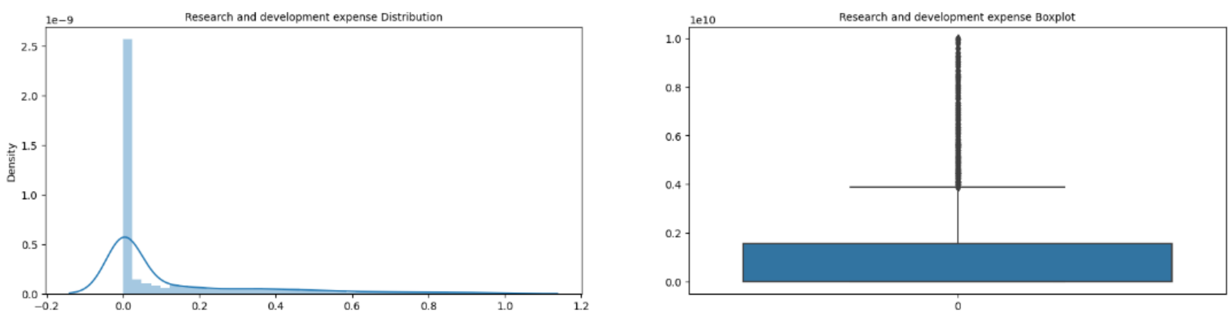
Also, we are dropping variables with vif value as NaN since they do not add any additional information to the model and might be redundant information.

1.6 Univariate Analysis–



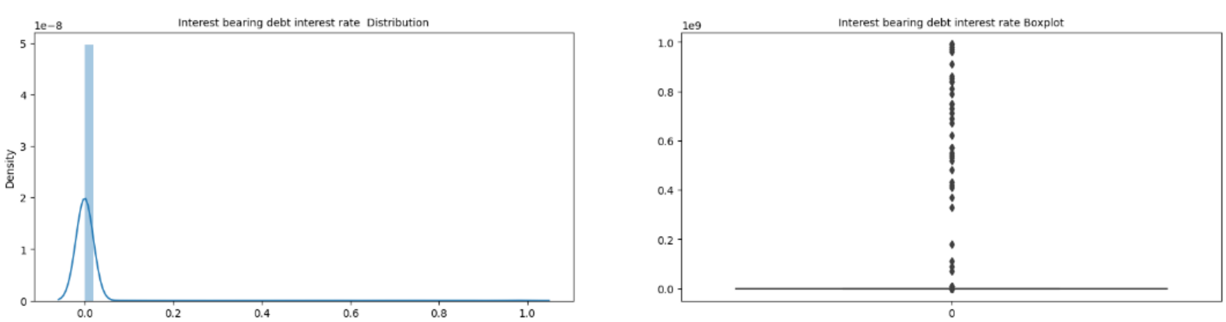
Skewness = 2.55

Distribution of the "Default" variable is positively skewed. This indicates that there are more instances of non-default compared to default.

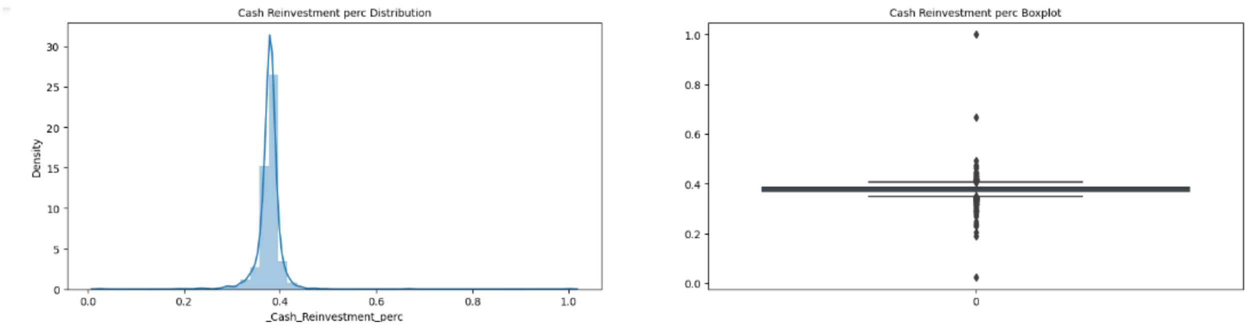


Skewness = 1.99

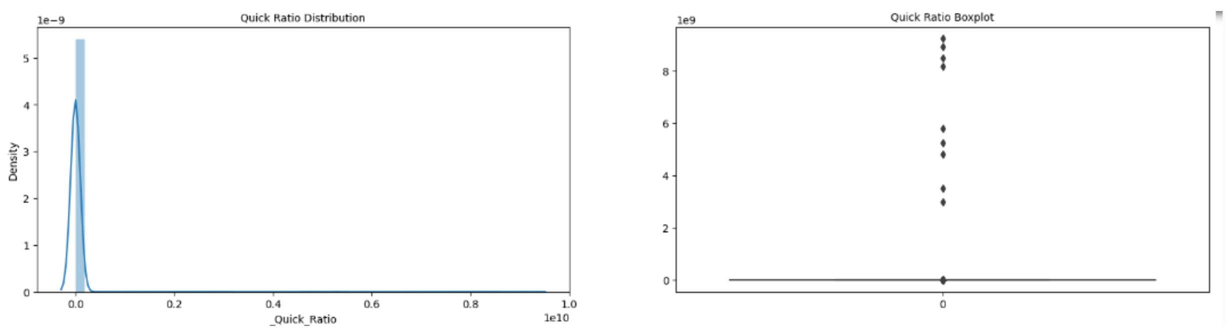
Has varying distribution with major under 0.0-0.18 value.



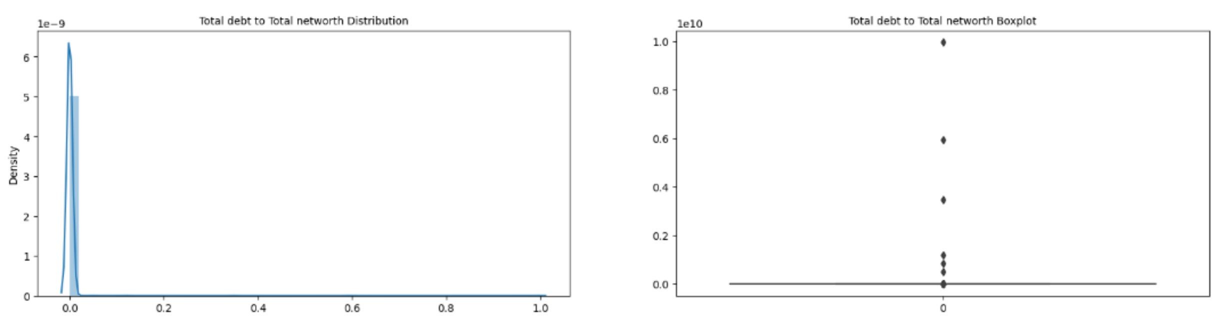
The skewness value of 8.67 indicates significant positive skewness. There are few instances of very high interest-bearing debt interest rates compared to the majority of lower rates.



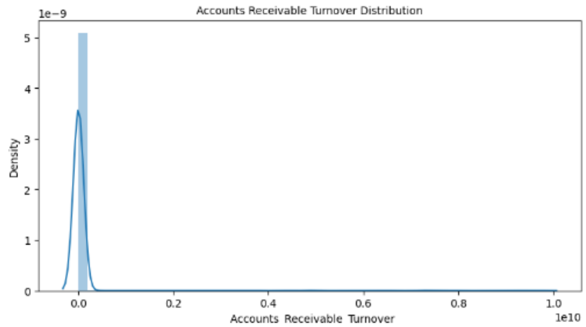
Skewness = 4.42, suggesting that majority of annual cash flow that the company invests back into the business as a new investment falls under 0.3-0.4.

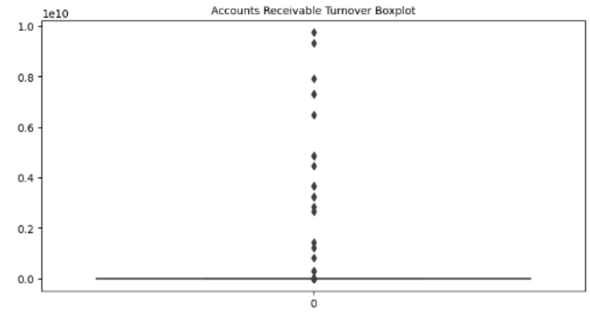


A skewness value of 17.33, indicates extremely high positive skewness with major quick ratios under 0.05.

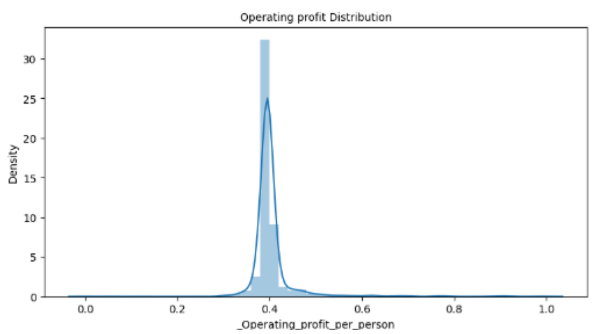


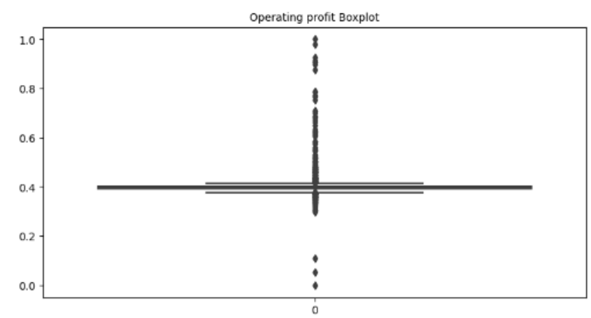
This variable exhibits extremely high positive skewness of 30.83. It indicates that there are very few instances of high debt-to-net worth ratios compared to the majority of lower ratios.



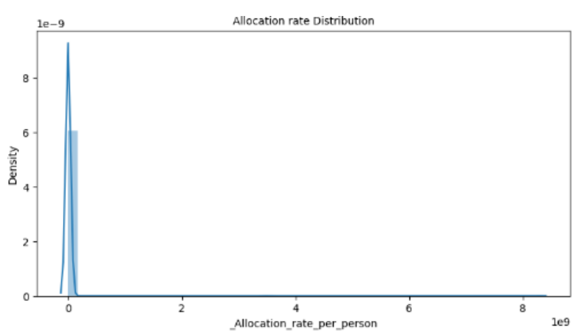


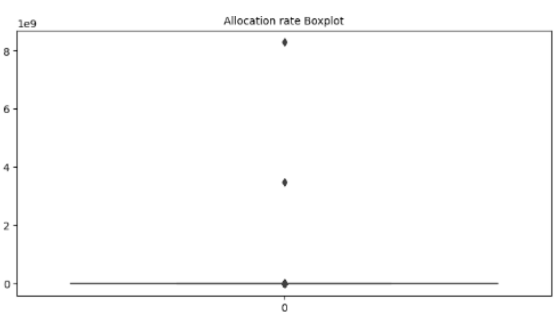
Positive skewness of 14.19. This suggests relatively higher instances of low accounts receivable turnover compared to lower turnover rates.



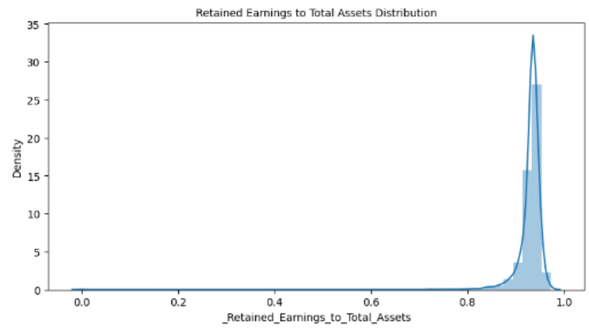


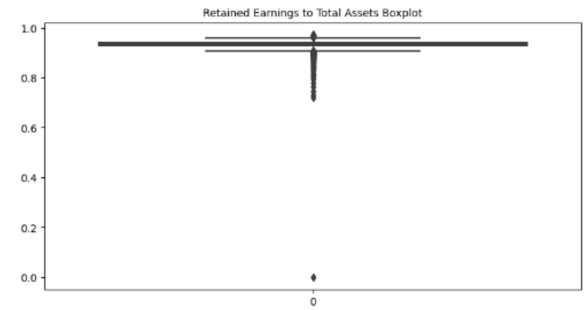
Skewness = 5.34. It indicates that for the Operating Income/ per employee, very few fall below 0.3 and a major of them fall under 0.3-0.5.



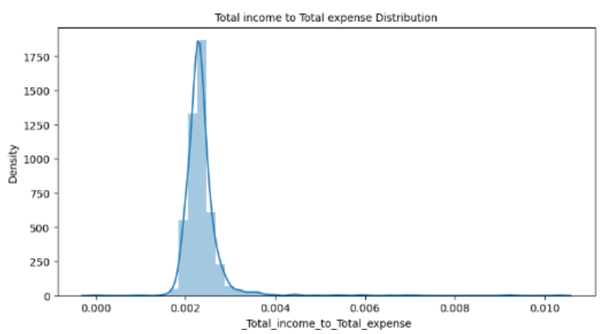


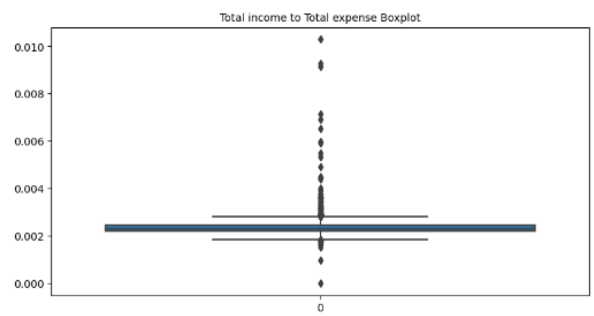
Extremely high positive skewness of 38.17. It indicates that there may be very few instances of high allocation rates per person compared to the majority of lower rates.



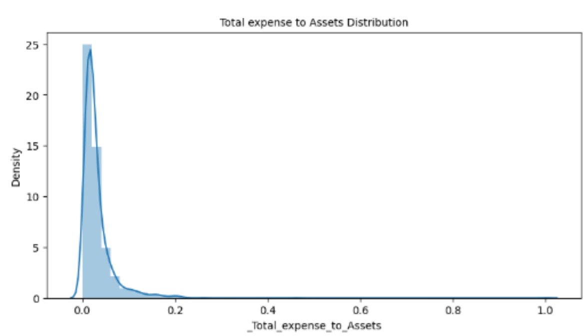


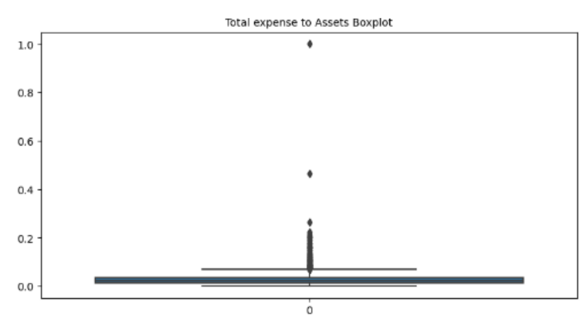
Negative skewness of -16.14. It indicates that majority of high retained earnings to total assets ratios fall under 0.8-1.0



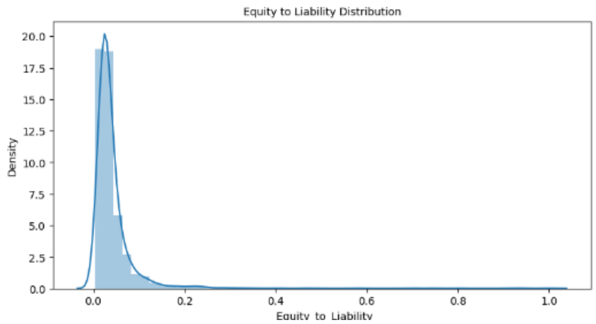


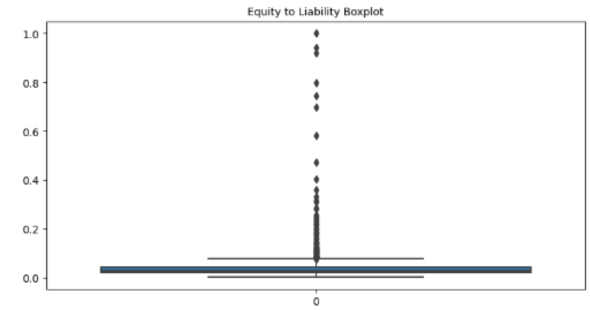
Positive skewness of 8.02 suggests that majority of total income to total expense ratios fall under 0.2-0.3%





Positive skewness of 9.75 indicates that majority of observations have lower total expense to assets ratios.

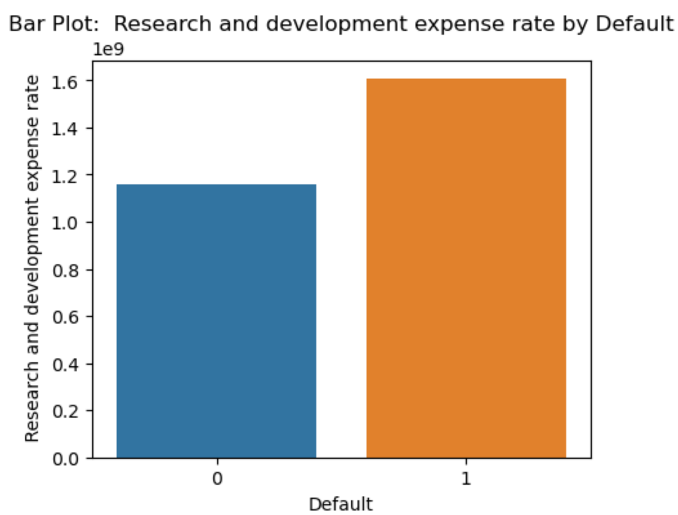


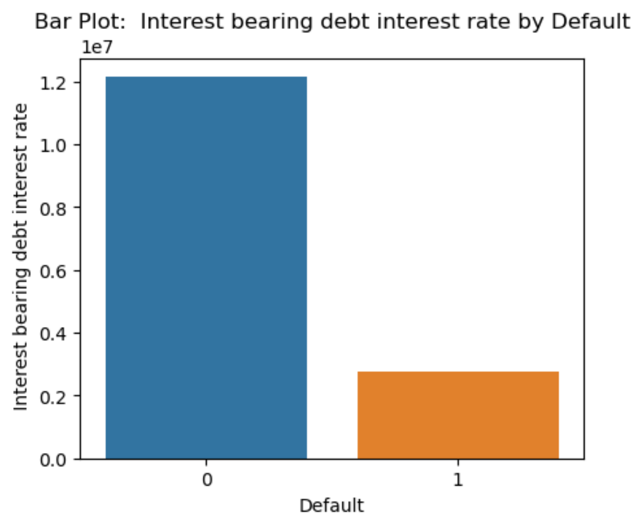


Positive skewness of 9.14. A peak in the range between 0.0 and 0.2, indicates that a significant proportion of the data points have equity to liability ratios falling within this range.

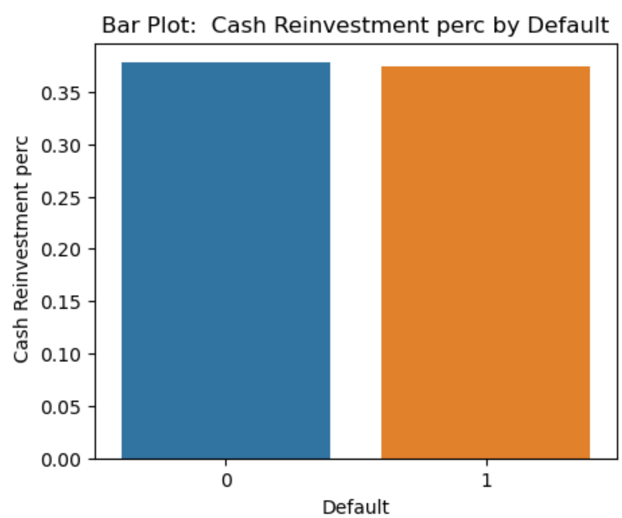
There are a notable number of data points that deviate significantly from this trend. These outliers represent observations with much higher equity to liability ratios compared to the majority of the dataset.

Bivariate Analysis -

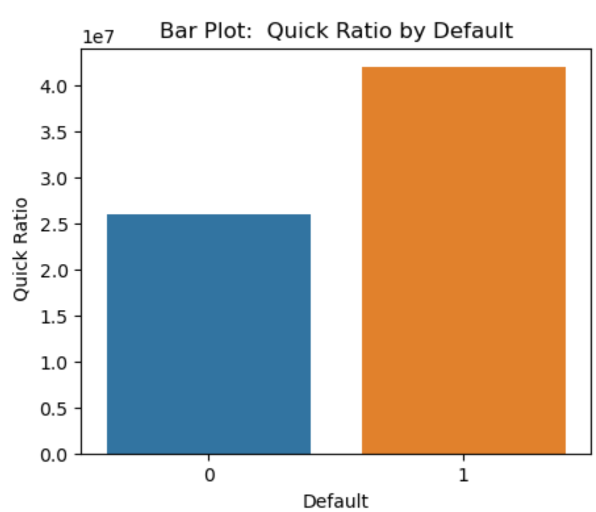




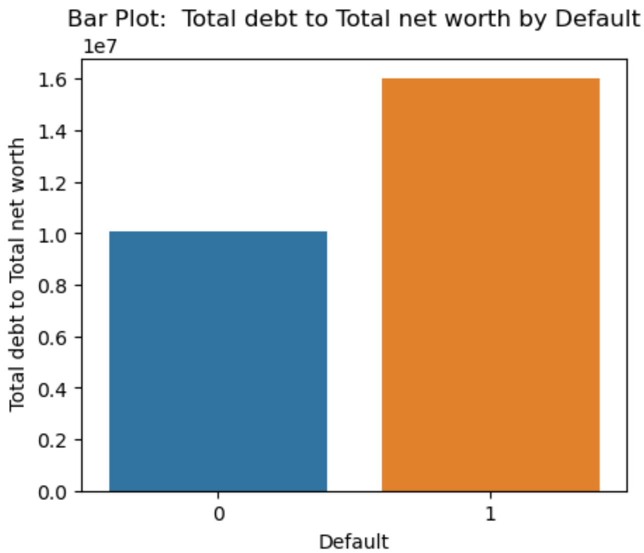
Bar Plot: Interest bearing debt interest rate by Default
1e7
Interest bearing debt interest rate
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
1
Default

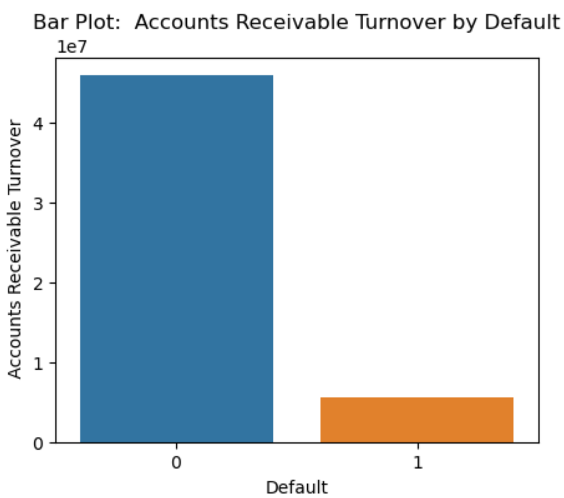


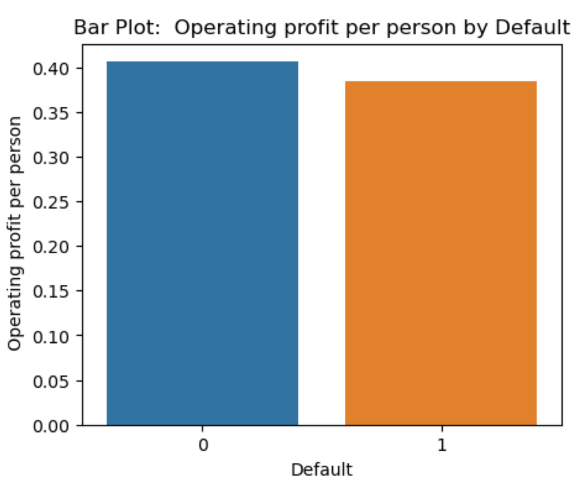
Bar Plot: Cash Reinvestment perc by Default
Cash Reinvestment perc
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
1
Default

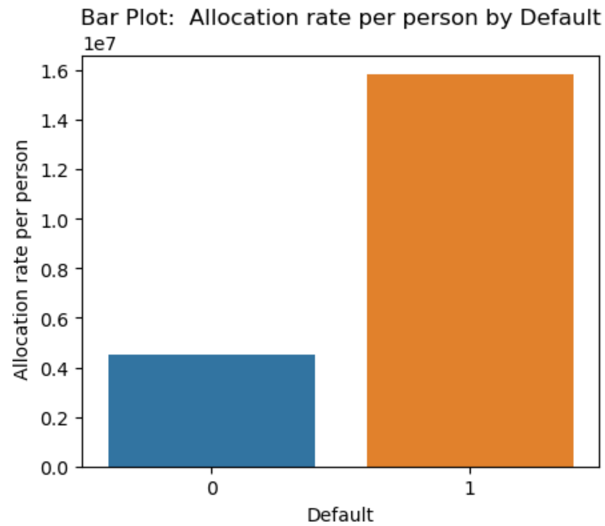


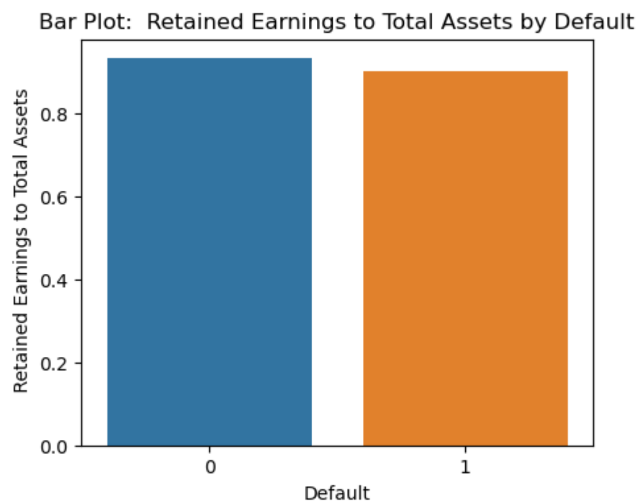
1e7
Bar Plot: Quick Ratio by Default
Quick Ratio
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0
1
Default

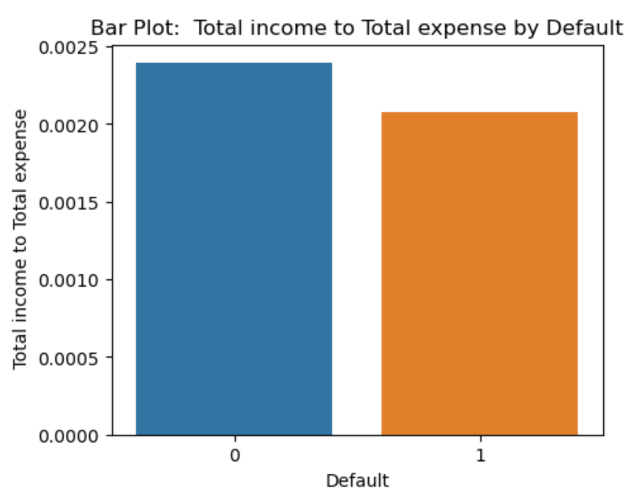

Bar Plot: Total debt to Total net worth by Default
1e7
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Total debt to Total net worth
0
1
Default

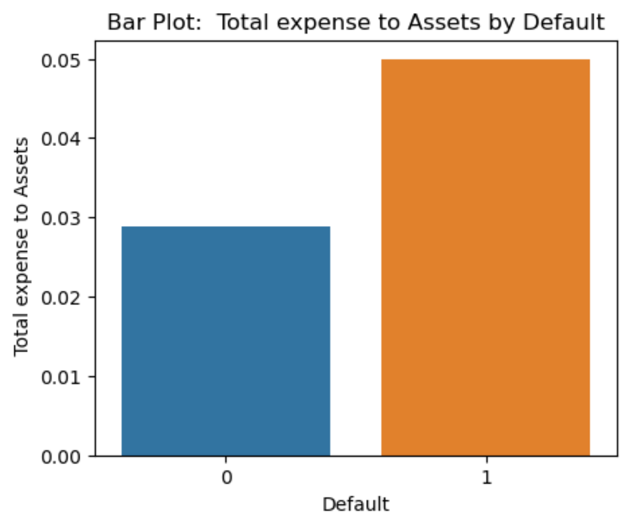

Bar Plot: Accounts Receivable Turnover by Default
1e7
4
3
2
1
0
Accounts Receivable Turnover
0
1
Default

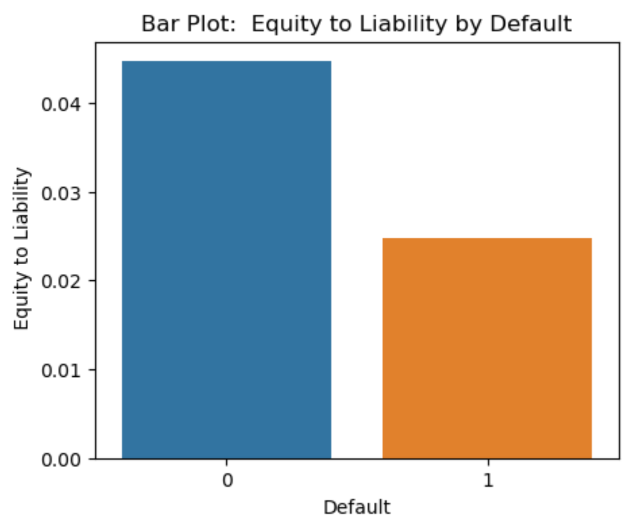

Bar Plot: Operating profit per person by Default
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Operating profit per person
0
1
Default


Bar Plot: Allocation rate per person by Default
1e7
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Allocation rate per person
0
1
Default


Bar Plot: Retained Earnings to Total Assets by Default
0.8
0.6
0.4
0.2
0.0
Retained Earnings to Total Assets
0
1
Default


Bar Plot: Total income to Total expense by Default
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
Total income to Total expense
0
1
Default





- More instances of Default are observed when the research and development expense rate is higher
- Instances of Default are more prevalent when the interest-bearing debt interest rate is lower
- The distribution of cash reinvestment percentages between default and non-default companies is somewhat equal, though slightly higher instances of default are observed for lower reinvestment percentages
- Default companies tend to have higher quick ratios
- Default companies have higher total debt to total net worth ratios compared to non-default companies
- Default companies tend to have lower accounts receivable turnover compared to non-default companies
- Both Default and non-default companies exhibit high operating profit per person, with slightly higher values observed for non-bankrupt companies.

- Default companies have higher allocation rates per person compared to non-default companies
- Both default and non-default companies have high retained earnings to total assets ratios, with slightly higher values observed for non-default companies.
- The ratio of total income to total expense is higher for non-default companies compared to default companies
- The ratio of total expense to assets is higher for default companies compared to non-default companies
- The ratio of equity to liability is higher for non-default companies compared to default companies

In summary, higher research and development expenses, lower interest-bearing debt interest rates, and higher cash reinvestment percentages are associated with lower likelihoods of defaulting. Conversely, higher total debt to total net worth ratios and total expense to asset ratios are associated with higher likelihoods of defaulting.

1.7 Logistic Regression Model-

We have built logistic regression model using statsmodels library and defining a function which describes Default using all independent variables.

We have fit the model to the training data and here's the summary information of model 1-

```
Optimization terminated successfully.
         Current function value: 0.185745
         Iterations 9
```

Logit Regression Results

Dep. Variable:	Default	**No. Observations:**	1378
Model:	Logit	**Df Residuals:**	1335
Method:	MLE	**Df Model:**	42
Date:	Fri, 05 Apr 2024	**Pseudo R-squ.:**	0.4673
Time:	23:00:57	**Log-Likelihood:**	-255.96
converged:	True	**LL-Null:**	-480.46
Covariance Type:	nonrobust	**LLR p-value:**	1.508e-69

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.3328	0.305	-14.209	0.000	-4.930	-3.735
_Operating_Expense_Rate	0.0822	0.141	0.582	0.561	-0.195	0.359
_Research_and_development_expense_rate	0.4434	0.126	3.510	0.000	0.196	0.691
_Cash_flow_rate	-0.0536	0.309	-0.174	0.862	-0.658	0.551
_Interest_bearing_debt_interest_rate	0.4558	0.153	2.987	0.003	0.157	0.755
_Tax_rate_A	-0.1703	0.168	-1.011	0.312	-0.500	0.160
_Cash_Flow_Per_Share	0.2365	0.250	0.945	0.344	-0.254	0.727
_Realized_Sales_Gross_Profit_Growth_Rate	-0.0587	0.157	-0.373	0.709	-0.367	0.250
_Operating_Profit_Growth_Rate	0.1266	0.188	0.672	0.502	-0.243	0.496
_Continuous_Net_Profit_Growth_Rate	-0.2766	0.204	-1.356	0.175	-0.676	0.123
_Total_Asset_Growth_Rate	-0.2425	0.140	-1.731	0.083	-0.517	0.032
_Net_Value_Growth_Rate	-0.4228	0.198	-2.135	0.033	-0.811	-0.035
_Total_Asset_Return_Growth_Rate_Ratio	0.3275	0.199	1.647	0.100	-0.062	0.717
_Cash_Reinvestment_perc	-0.4782	0.225	-2.127	0.033	-0.919	-0.038
_Quick_Ratio	-1.1682	0.315	-3.706	0.000	-1.786	-0.550
_Interest_Expense_Ratio	0.0087	0.145	0.060	0.952	-0.276	0.293

_Long_term_fund_suitability_ratio_A	0.2180	0.192	1.135	0.257	-0.159	0.595
_Accounts_Receivable_Turnover	-0.5044	0.194	-2.604	0.009	-0.884	-0.125
_Average_Collection_Days	0.2055	0.170	1.210	0.226	-0.127	0.538
_Inventory_Turnover_Rate_times	0.0455	0.129	0.351	0.725	-0.208	0.299
_Fixed_Assets_Turnover_Frequency	0.2165	0.153	1.416	0.157	-0.083	0.516
_Net_Worth_Turnover_Rate_times	0.2245	0.194	1.157	0.247	-0.156	0.605
_Operating_profit_per_person	0.5488	0.206	2.660	0.008	0.144	0.953
_Allocation_rate_per_person	0.6737	0.198	3.399	0.001	0.285	1.062
_Cash_to_Total_Assets	-0.0470	0.259	-0.182	0.856	-0.554	0.460
_Cash_to_Current_Liability	0.2819	0.195	1.447	0.148	-0.100	0.664
_Inventory_to_Working_Capital	-0.0935	0.112	-0.837	0.403	-0.313	0.126
_Inventory_to_Current_Liability	-0.0970	0.203	-0.478	0.632	-0.494	0.300
_Long_term_Liability_to_Current_Assets	-0.2265	0.157	-1.444	0.149	-0.534	0.081
_Retained_Earnings_to_Total_Assets	-0.7812	0.250	-3.127	0.002	-1.271	-0.292
_Total_income_to_Total_expense	-0.8809	0.337	-2.615	0.009	-1.541	-0.221
_Total_expense_to_Assets	0.4072	0.187	2.177	0.029	0.041	0.774
_Current_Asset_Turnover_Rate	-0.0554	0.139	-0.400	0.689	-0.327	0.216
_Quick_Asset_Turnover_Rate	-0.0337	0.135	-0.250	0.803	-0.298	0.231
_Cash_Turnover_Rate	-0.1929	0.136	-1.419	0.156	-0.459	0.074
_Fixed_Assets_to_Assets	0.1284	0.229	0.562	0.574	-0.320	0.577
_Cash_Flow_to_Liability	-0.1789	0.185	-0.964	0.335	-0.542	0.185
_Current_Liability_to_Current_Assets	-0.1738	0.160	-1.085	0.278	-0.488	0.140
_Total_assets_to_GNP_price	0.1046	0.148	0.709	0.479	-0.185	0.394
_No_credit_Interval	-0.0116	0.128	-0.091	0.928	-0.262	0.238
_Degree_of_Financial_Leverage_DFL	0.0521	0.155	0.336	0.737	-0.252	0.356
_Equity_to_Liability	-0.9825	0.370	-2.657	0.008	-1.707	-0.258

As high p-value indicates that the independent variable may not be statistically significant in predicting the dependent variable. We are building models by removing independent variables for which the associated p-value is greater than 0.05.

This process helps in reducing overfitting and improving the interpretability of the model by focusing on the most relevant predictors.

By continuing the process, here's our final model, and we have cut down to the most important features for our prediction.

```
Optimization terminated successfully.
         Current function value: 0.196506
         Iterations 9
```

Logit Regression Results

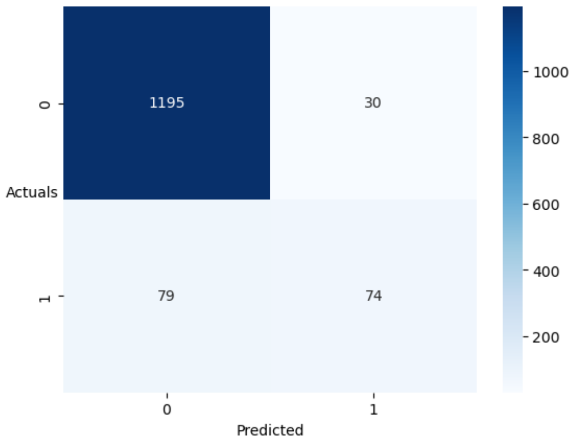
Dep. Variable:	Default	**No. Observations:**	1378
Model:	Logit	**Df Residuals:**	1365
Method:	MLE	**Df Model:**	12
Date:	Fri, 05 Apr 2024	**Pseudo R-squ.:**	0.4364
Time:	22:05:41	**Log-Likelihood:**	-270.79
converged:	True	**LL-Null:**	-480.46
Covariance Type:	nonrobust	**LLR p-value:**	3.010e-82

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.2623	0.273	-15.597	0.000	-4.798	-3.727
_Research_and_development_expense_rate	0.3966	0.112	3.556	0.000	0.178	0.615
_Interest_bearing_debt_interest_rate	0.4014	0.143	2.808	0.005	0.121	0.682
_Cash_Reinvestment_perc	-0.3675	0.110	-3.350	0.001	-0.582	-0.153
_Quick_Ratio	-0.7906	0.245	-3.228	0.001	-1.271	-0.311
_Total_debt_to_Total_net_worth	0.2572	0.065	3.980	0.000	0.131	0.384
_Accounts_Receivable_Turnover	-0.6406	0.140	-4.570	0.000	-0.915	-0.366
_Operating_profit_per_person	0.4699	0.190	2.474	0.013	0.098	0.842
_Allocation_rate_per_person	0.7036	0.139	5.070	0.000	0.432	0.976
_Retained_Earnings_to_Total_Assets	-0.8771	0.206	-4.258	0.000	-1.281	-0.473
_Total_income_to_Total_expense	-1.0932	0.274	-3.995	0.000	-1.630	-0.557
_Total_expense_to_Assets	0.4129	0.150	2.755	0.006	0.119	0.707
_Equity_to_Liability	-1.1364	0.275	-4.139	0.000	-1.674	-0.598

We are predicting on train set and converting predicted probabilities to class labels based on a threshold of 0.5 such that-

- If the predicted probability is greater than 0.5, classify it as 1.
- If the predicted probability is less than or equal to 0.5, classify it as 0.

Confusion Matrix on Train set -



Classification report on Train set-

	precision	recall	f1-score	support
0.0	0.938	0.976	0.956	1225
1.0	0.712	0.484	0.576	153
accuracy			0.921	1378
macro avg	0.825	0.730	0.766	1378
weighted avg	0.913	0.921	0.914	1378

Choosing optimal Threshold-

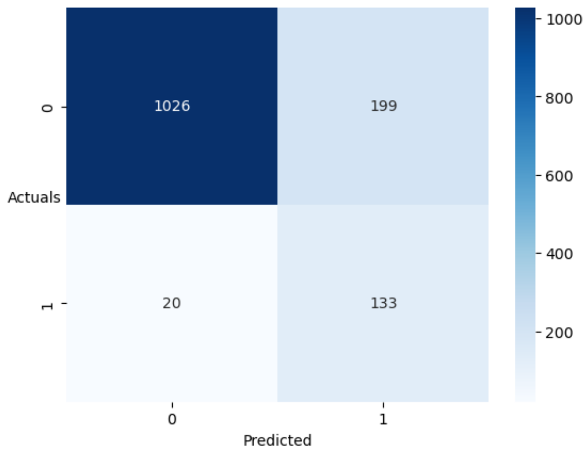
We worked with default threshold of 0.5.

Now, using ROC curve, we are building threshold such that it ensures there is maximum difference between TPR and FPR i.e. it maximizes True Positive and minimizes False Positive.

The threshold we've derived is **0.11**

Working on model with revised threshold-

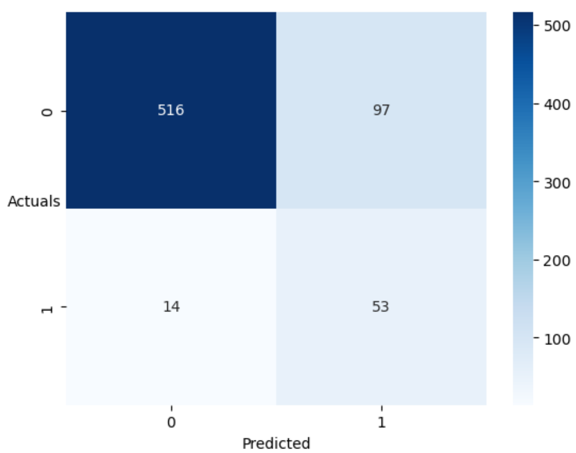
<table>
<tr><td></td><td>Confusion Matrix</td><td>Classification report</td></tr>
<tr><td>Train</td><td>



</td><td>

<table>
<tr><td></td><td>precision</td><td>recall</td><td>f1-score</td><td>support</td></tr>
<tr><td>0.0</td><td>0.981</td><td>0.838</td><td>0.904</td><td>1225</td></tr>
<tr><td>1.0</td><td>0.401</td><td>0.869</td><td>0.548</td><td>153</td></tr>
<tr><td>accuracy</td><td></td><td></td><td>0.841</td><td>1378</td></tr>
<tr><td>macro avg</td><td>0.691</td><td>0.853</td><td>0.726</td><td>1378</td></tr>
<tr><td>weighted avg</td><td>0.916</td><td>0.841</td><td>0.864</td><td>1378</td></tr>
</table>

</td></tr>
<tr><td>Test</td><td>



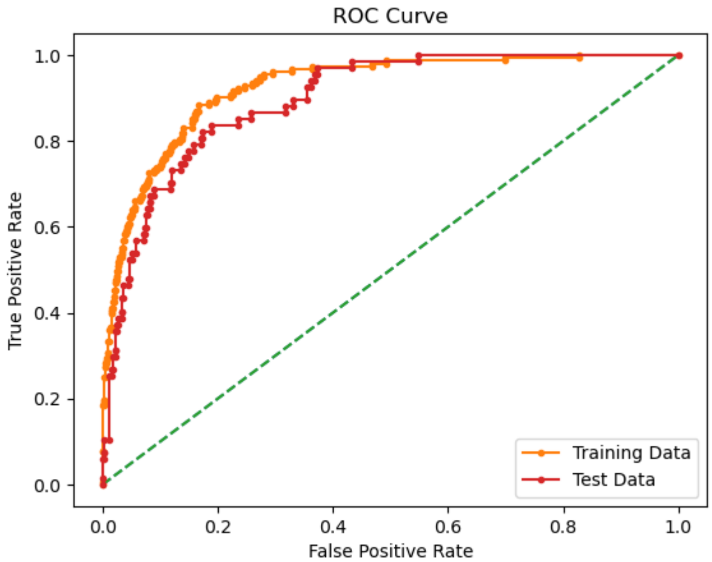
</td><td>

<table>
<tr><td></td><td>precision</td><td>recall</td><td>f1-score</td><td>support</td></tr>
<tr><td>0.0</td><td>0.974</td><td>0.842</td><td>0.903</td><td>613</td></tr>
<tr><td>1.0</td><td>0.353</td><td>0.791</td><td>0.488</td><td>67</td></tr>
<tr><td>accuracy</td><td></td><td></td><td>0.837</td><td>680</td></tr>
<tr><td>macro avg</td><td>0.663</td><td>0.816</td><td>0.696</td><td>680</td></tr>
<tr><td>weighted avg</td><td>0.912</td><td>0.837</td><td>0.862</td><td>680</td></tr>
</table>

</td></tr>
</table>

AUC, ROC –

AUC for the Training Data: 0.925
AUC for the Test Data: 0.897



1.8 Random Forest Model-

Performing Grid Search and tuning few hyper-parameters for the Random Forest classifier

```
GridSearchCV(estimator=RandomForestClassifier(),
             param_grid={'max_depth': [3, 5, 7],
                         'min_samples_leaf': [5, 10, 15],
                         'min_samples_split': [15, 30, 45],
                         'n_estimators': [25, 50]})
```

Choosing best params and predicting using best estimators-

```
{'max_depth': 5,
 'min_samples_leaf': 5,
 'min_samples_split': 30,
 'n_estimators': 25}
```

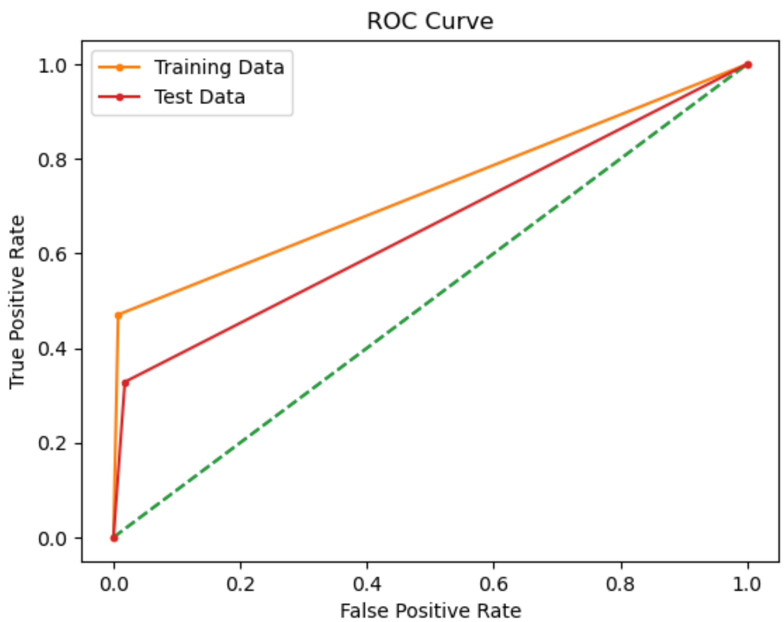
Classification report for Train-

	precision	recall	f1-score	support
0.0	0.94	0.99	0.96	1225
1.0	0.89	0.47	0.62	153
accuracy			0.93	1378
macro avg	0.91	0.73	0.79	1378
weighted avg	0.93	0.93	0.93	1378

Classification report for Test-

	precision	recall	f1-score	support
0.0	0.93	0.98	0.96	613
1.0	0.67	0.33	0.44	67
accuracy			0.92	680
macro avg	0.80	0.66	0.70	680
weighted avg	0.90	0.92	0.90	680

AUC for the Training Data: 0.732
AUC for the Test Data: 0.655



1.9 Linear Discriminant Analysis -

```
▾ LinearDiscriminantAnalysis
LinearDiscriminantAnalysis()
```

Classification Report for Train-

	precision	recall	f1-score	support
0.0	0.95	0.96	0.95	1225
1.0	0.64	0.58	0.61	153
accuracy			0.92	1378
macro avg	0.79	0.77	0.78	1378
weighted avg	0.91	0.92	0.92	1378

Classification Report for Test-

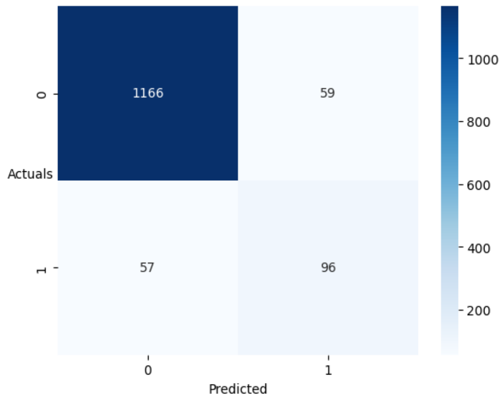
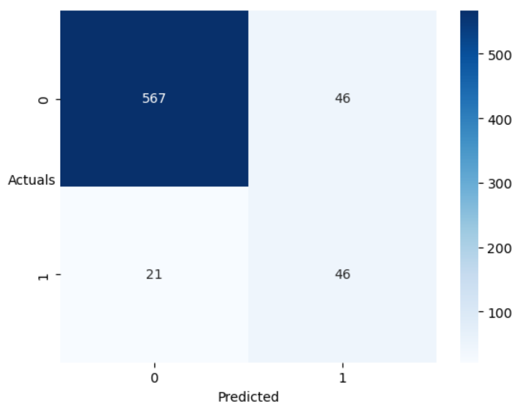
	precision	recall	f1-score	support
0.0	0.96	0.94	0.95	613
1.0	0.55	0.63	0.58	67
accuracy			0.91	680
macro avg	0.75	0.78	0.77	680
weighted avg	0.92	0.91	0.91	680

Adjusting threshold-

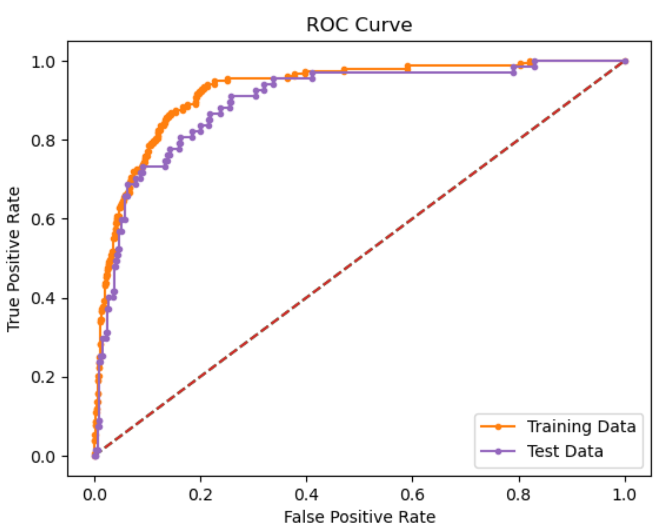
We are separately predicting probabilities and taking only probability of 1

Threshold derived- **0.378**

Modifying classification on the basis of revised threshold.

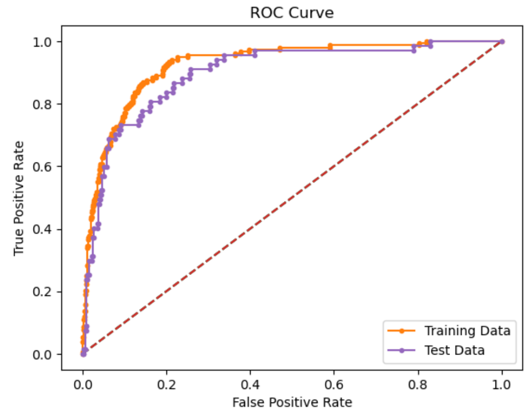
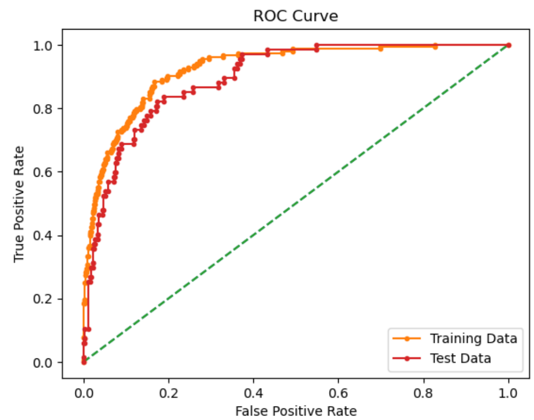
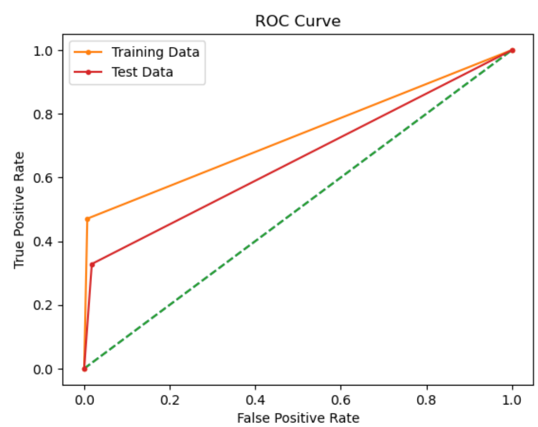
<table>
<tr><th></th><th>Confusion Matrix</th><th>Classification Report</th></tr>
<tr><td>Train</td><td>
</td><td>precision recall f1-score support
0.0 0.953 0.952 0.953 1225
1.0 0.619 0.627 0.623 153
accuracy 0.916 1378
macro avg 0.786 0.790 0.788 1378
weighted avg 0.916 0.916 0.916 1378</td></tr>
<tr><td>Test</td><td>
</td><td>precision recall f1-score support
0.0 0.964 0.925 0.944 613
1.0 0.500 0.687 0.579 67
accuracy 0.901 680
macro avg 0.732 0.806 0.761 680
weighted avg 0.919 0.901 0.908 680</td></tr>
</table>

AUC for the Training Data: 0.925
AUC for the Test Data: 0.898



1.10 Comparison of the models-

Model	Set		precision	recall	f1-score	support
LDA	Train set	0.0	0.953	0.952	0.953	1225
		1.0	0.619	0.627	0.623	153
		accuracy			0.916	1378
		macro avg	0.786	0.790	0.788	1378
		weighted avg	0.916	0.916	0.916	1378
LDA	Test set	0.0	0.964	0.925	0.944	613
		1.0	0.500	0.687	0.579	67
		accuracy			0.901	680
		macro avg	0.732	0.806	0.761	680
		weighted avg	0.919	0.901	0.908	680
Logistic Regression	Train set	0.0	0.981	0.838	0.904	1225
		1.0	0.401	0.869	0.548	153
		accuracy			0.841	1378
		macro avg	0.691	0.853	0.726	1378
		weighted avg	0.916	0.841	0.864	1378
Logistic Regression	Test set	0.0	0.974	0.842	0.903	613
		1.0	0.353	0.791	0.488	67
		accuracy			0.837	680
		macro avg	0.663	0.816	0.696	680
		weighted avg	0.912	0.837	0.862	680
Random Forest	Train set	0.0	0.94	0.99	0.96	1225
		1.0	0.89	0.47	0.62	153
		accuracy			0.93	1378
		macro avg	0.91	0.73	0.79	1378
		weighted avg	0.93	0.93	0.93	1378
Random Forest	Test set	0.0	0.93	0.98	0.96	613
		1.0	0.67	0.33	0.44	67
		accuracy			0.92	680
		macro avg	0.80	0.66	0.70	680
		weighted avg	0.90	0.92	0.90	680

	AUC	**ROC**
LDA	AUC for the Training Data: 0.925 AUC for the Test Data: 0.898	
Logistic Regression	AUC for the Training Data: 0.925 AUC for the Test Data: 0.897	
Random Forest	AUC for the Training Data: 0.732 AUC for the Test Data: 0.655	

Let's analyze the three models by considering performance metrices :

1. Accuracy: This indicates the overall correctness of the model predictions.

2. Precision: Precision is the ratio of correctly predicted positive observations to the total predicted positives. It measures the correctness of positive predictions.

3. Recall: Recall is the ratio of correctly predicted positive observations to all observations in actual class. It measures the ability of the model to find all the relevant cases within a dataset.

4. F1-score: F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall.

5. AUC (Area Under the ROC Curve): AUC measures the ability of the model to distinguish between positive and negative classes. Higher AUC values indicate better performance.

- Based on the comparison, the **Linear Discriminant Analysis (LDA) model appears to be the optimum choice for this problem.**
- LDA shows a good balance between precision, recall, and F1-score on both the train and test sets, indicating better generalization.
- LDA achieves the highest AUC on the test set among the three models suggesting better discrimination power.
- Random Forest performs well on the train set but shows a decrease in performance on the test set, indicating potential overfitting.
- Logistic Regression also shows a decrease in performance on the test set compared to LDA.

1.11 Conclusions and Recommendations-

Model Performance:

- Linear Discriminant Analysis (LDA) demonstrates the best overall performance among the three models, with consistently high precision, recall, and F1-score on both the train and test sets. LDA also achieves the highest Area Under the Curve (AUC) on the test set, indicating superior discrimination power.
- Logistic Regression shows decent performance but slightly lower than LDA, especially in terms of recall and F1-score for the minority class (Default).
- Random Forest performs well on the train set but exhibits a decrease in performance on the test set, suggesting potential overfitting.

Important Features:

- Certain features have significant impact on the likelihood of default. For instance, higher research and development expense rates, lower interest-bearing debt interest rates, and higher cash reinvestment percentages are associated with lower likelihoods of defaulting.

- While, higher total debt to total net worth ratios and total expense to asset ratios are associated with higher likelihoods of defaulting.

Business Recommendation-

- **Risk Assessment:** By understanding the companies that might struggle financially, we can make smart decisions about giving loans or investing money.
- **Invest in Research and Development:** Companies should prioritize spending on research and development to improve their products or services. This investment not only drives innovation but also lowers the risk of default.
- **Manage Debt**: Be cautious with borrowing and ensure that interest-bearing debt remains at manageable levels. High debt can strain finances and increase the chances of defaulting. Strive to keep the ratio of total debt to total net worth within reasonable limits. Excessive debt relative to net worth increases financial risk and the likelihood of default.
- **Reinvestment:** Use available cash to reinvest in the business wisely. This could involve upgrading equipment, expanding operations, or investing in new opportunities. Strategic reinvestment can enhance growth and financial stability.
- **Liquidity Management:** Focus on maintaining adequate liquidity levels, as indicated by the quick ratio to ensure the ability to meet short-term obligations. Maintaining sufficient liquidity safeguards against default during unforeseen circumstances.
- **Efficiency Improvement:** Improve efficiency in collecting accounts receivable. As a high turnover ratio indicates prompt collection of payments, this helps maintain steady cash flow and reduces the risk of default.
- **Operational Efficiency:** Companies that use their resources well and make good profits per employee are stronger. We can look for ways to improve how efficiently the company works.

By implementing these recommendations, businesses can enhance their ability to identify and mitigate default risks, thereby safeguarding financial stability and position themselves for sustainable growth and success.

2 Part B-

2.1 Problem Statement-

The dataset contains 6 years of information(weekly stock information) on the stock prices of 10 different Indian Stocks. Calculate the mean and standard deviation on the stock returns and share insights. You are expected to do the Market Risk Analysis using Python.

2.2 Summary-

Head-

	Date	Infosys	Indian Hotel	Mahindra & Mahindra	Axis Bank	SAIL	Shree Cement	Sun Pharma	Jindal Steel	Idea Vodafone	Jet Airways
0	31-03-2014	264	69	455	263	68	5543	555	298	83	278
1	07-04-2014	257	68	458	276	70	5728	610	279	84	303
2	14-04-2014	254	68	454	270	68	5649	607	279	83	280
3	21-04-2014	253	68	488	283	68	5692	604	274	83	282
4	28-04-2014	256	65	482	282	63	5582	611	238	79	243

Shape-

The number of rows (observations) is 314
The number of columns (variables) is 11

Summary-

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 314 entries, 0 to 313
Data columns (total 11 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Date                 314 non-null    object
 1   Infosys              314 non-null    int64
 2   Indian Hotel         314 non-null    int64
 3   Mahindra & Mahindra  314 non-null    int64
 4   Axis Bank            314 non-null    int64
 5   SAIL                 314 non-null    int64
 6   Shree Cement         314 non-null    int64
 7   Sun Pharma           314 non-null    int64
 8   Jindal Steel         314 non-null    int64
 9   Idea Vodafone        314 non-null    int64
 10  Jet Airways          314 non-null    int64
dtypes: int64(10), object(1)
memory usage: 27.1+ KB
```

Descriptive statistics-

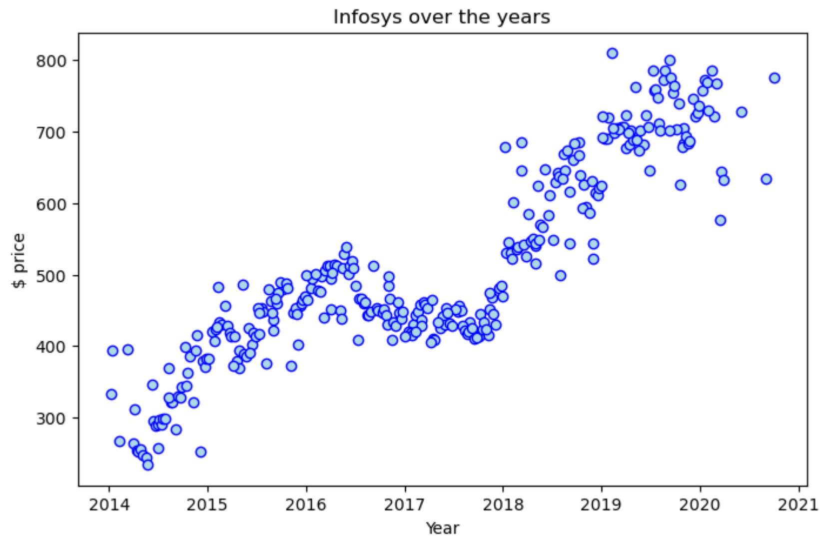
	Infosys	Indian Hotel	Mahindra & Mahindra	Axis Bank	SAIL	Shree Cement	Sun Pharma	Jindal Steel	Idea Vodafone	Jet Airways
count	314.00	314.00	314.00	314.00	314.00	314.00	314.00	314.00	314.00	314.00
mean	511.34	114.56	636.68	540.74	59.10	14806.41	633.47	147.63	53.71	372.66
std	135.95	22.51	102.88	115.84	15.81	4288.28	171.86	65.88	31.25	202.26
min	234.00	64.00	284.00	263.00	21.00	5543.00	338.00	53.00	3.00	14.00
25%	424.00	96.00	572.00	470.50	47.00	10952.25	478.50	88.25	25.25	243.25
50%	466.50	115.00	625.00	528.00	57.00	16018.50	614.00	142.50	53.00	376.00
75%	630.75	134.00	678.00	605.25	71.75	17773.25	785.00	182.75	82.00	534.00
max	810.00	157.00	956.00	808.00	104.00	24806.00	1089.00	338.00	117.00	871.00

- The dataset contains information on the weekly stock prices of 10 different Indian stocks over a period of 6 years.
- There are a total of 314 observations (rows) and 11 variables (columns) in the dataset.
- The 'Date' column contains date values indicating the week for which the stock prices are recorded.
- The other 10 columns represent the stock prices for the respective companies: Infosys, Indian Hotel, Mahindra & Mahindra, Axis Bank, SAIL, Shree Cement, Sun Pharma, Jindal Steel, Idea Vodafone, and Jet Airways.
- All stock price columns are of integer type.
- The mean stock prices vary across different companies, ranging from 53.71 for Idea Vodafone to 14806.41 for Shree Cement.
- Companies like Infosys, Mahindra & Mahindra, and Axis Bank have relatively higher mean stock prices compared to others.
- The stock prices for all companies exhibit a wide range of values, as indicated by the difference between the minimum and maximum values.

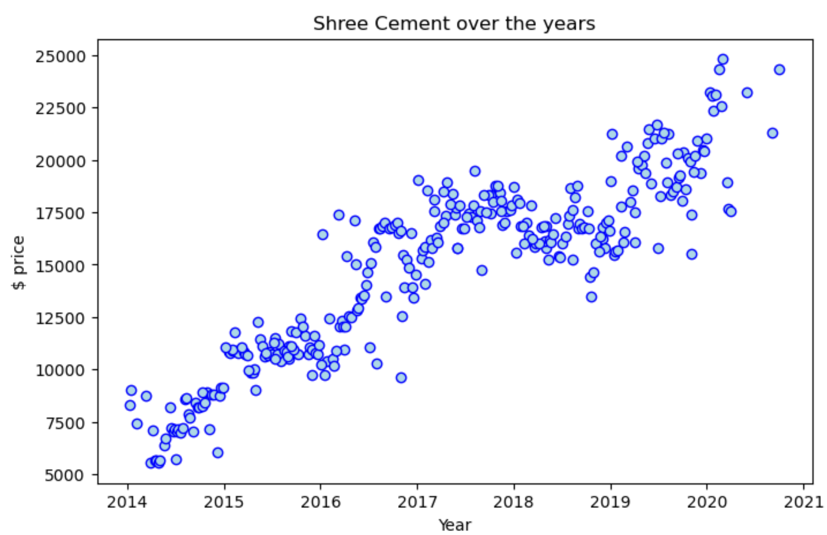
‘Date’ is as object data type so we create new field - ‘dates’ and converted it to Datetime.

2.3 Stock Price Graph

We are considering 2 stocks – Infosys and Shree Cement



- Infosys exhibits an upward trend.
- Although a slight decrease in stock price was observed in 2016-2018, the price has increased over the years.
- From 2014, when stock prices ranged between $250 and $400, there has been substantial growth with prices expanding significantly to reach $700-$800 by the year 2020.



- Highest mean stock price has been observed for Shree Cement.
- The stock price of Shree Cement has shown significant growth over the observed period.
- In 2014, the stock price ranged between 5500 and 9200, and by 2020, it had surged to a range of 17500 to 20200. This indicates a substantial increase in the value of Shree Cement stocks over the years.

2.4 Returns-

Calculating Logarithmic return from prices. It is the difference between 2 consecutive day prices.

Since the data is collected on a weekly basis, it is the difference between prices of 2 consecutive weeks.

Shape of stock returns dataset- `(314, 10)`

Head-

	Infosys	Indian Hotel	Mahindra & Mahindra	Axis Bank	SAIL	Shree Cement	Sun Pharma	Jindal Steel	Idea Vodafone	Jet Airways
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	-0.03	-0.01	0.01	0.05	0.03	0.03	0.09	-0.07	0.01	0.09
2	-0.01	0.00	-0.01	-0.02	-0.03	-0.01	-0.00	0.00	-0.01	-0.08
3	-0.00	0.00	0.07	0.05	0.00	0.01	-0.00	-0.02	0.00	0.01
4	0.01	-0.05	-0.01	-0.00	-0.08	-0.02	0.01	-0.14	-0.05	-0.15

1st row has value of Nan as this observation do not have previous values to be converted to return.

2.5 Stock Means and Standard Deviation -

We now look at Means & Standard Deviations of these returns.

Stock Means: Average returns that the stock is making on a week to week basis

```
Infosys               0.00
Indian Hotel          0.00
Mahindra & Mahindra  -0.00
Axis Bank             0.00
SAIL                 -0.00
Shree Cement          0.00
Sun Pharma           -0.00
Jindal Steel         -0.00
Idea Vodafone        -0.01
Jet Airways          -0.01
dtype: float64
```

Stock Standard Deviation : It is a measure of volatility, meaning, the more a stock's returns vary from the stock's average return, the more volatile the stock.

```
Infosys              0.04
Indian Hotel         0.05
Mahindra & Mahindra  0.04
Axis Bank            0.05
SAIL                 0.06
Shree Cement         0.04
Sun Pharma           0.05
Jindal Steel         0.08
Idea Vodafone        0.10
Jet Airways          0.10
dtype: float64
```

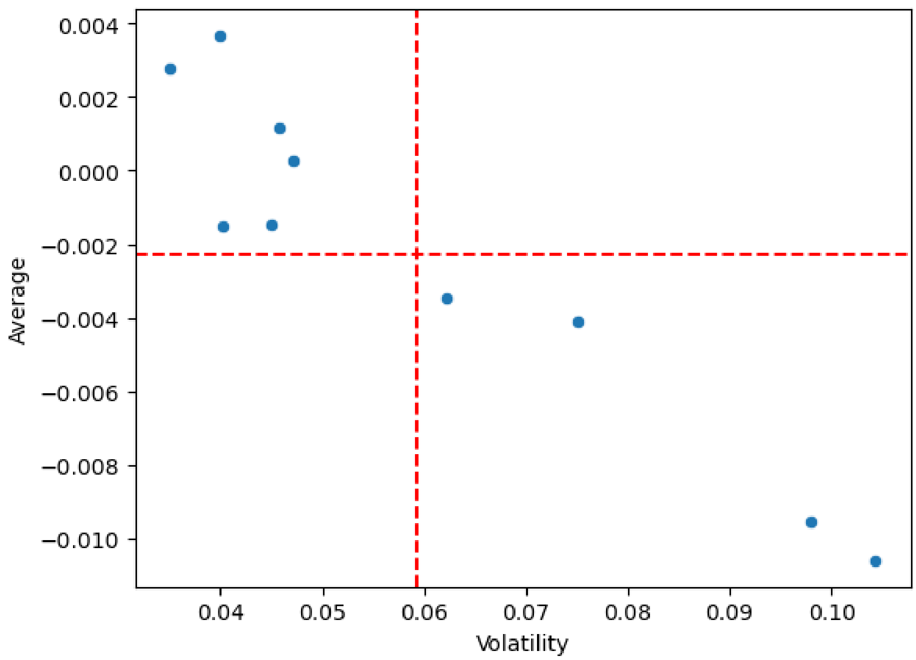
2.6 Plot of Stock Means vs Standard Deviation

We are combining these values into a dataframe-

	Average	Volatility
Infosys	0.00	0.04
Indian Hotel	0.00	0.05
Mahindra & Mahindra	-0.00	0.04
Axis Bank	0.00	0.05
SAIL	-0.00	0.06
Shree Cement	0.00	0.04
Sun Pharma	-0.00	0.05
Jindal Steel	-0.00	0.08
Idea Vodafone	-0.01	0.10
Jet Airways	-0.01	0.10

Now we will observe how each of these stocks perform compared to a reference point (calculated by taking mean of Average returns and Volatility)

Scatterplot-



- Stocks with lower volatility are considered less risky, while those with higher volatility are considered more risky.
- In our case, stocks with lower volatility are giving higher returns while those with higher volatility offer lower returns.
- And our aim would be to have as low risk as possible and get high return as possible.
- Thus, the ones with higher return for a comparative or lower risk are considered better.

2.7 Conclusions and Recommendations

Stocks with average returns greater than the mean average returns-

	Average	Volatility
Infosys	0.00	0.04
Shree Cement	0.00	0.04
Mahindra & Mahindra	-0.00	0.04
Sun Pharma	-0.00	0.05
Axis Bank	0.00	0.05
Indian Hotel	0.00	0.05

- These stocks represent relatively stable investment options with consistent average returns and manageable volatility compared to the overall market.
- The volatility values for the selected stocks range from 0.04 to 0.05. This suggests that these stocks exhibit relatively low to moderate levels of price fluctuations over time.
- Stocks like Infosys, Shree Cement, Mahindra & Mahindra exhibit relatively stable average returns (around 0) with moderate volatility (0.04).
- Sun Pharma, Axis Bank, and Indian Hotel also have stable average returns around 0 but slightly higher volatility (around 0.05).
- In general, all the selected stocks have an 'Average' return value of around 0.00. This indicates that, on average, these stocks have not shown significant positive or negative returns during the analyzed period.
- Investors seeking stable investments with lower risk may find these stocks attractive as they offer the potential for modest returns while minimizing exposure to significant price fluctuations.

Recommendations-

- **Focus on Low Volatility Stocks**: Given the preference for lower risk, investors should prioritize stocks with lower volatility. These stocks are expected to provide more stable returns over time and are suitable for risk-averse investors.
- **Portfolio Optimization:** Construct portfolios that balance risk and return by combining stocks with different risk profiles. This helps optimize returns while minimizing overall portfolio volatility.
- **Risk Management:** Monitor and manage risk exposure by regularly assessing the volatility and performance of portfolio holdings and implement risk management strategies to mitigate potential losses.

- Investors seeking long-term growth may prefer stocks with stable average returns like Infosys and Shree Cement. Their consistent performance over time can contribute to wealth accumulation.
- **Long-term approach:** Adopt a long-term investment approach when investing in stable, low-risk stocks. Focus on the fundamentals of the companies and their growth prospects rather than short-term market fluctuations.
- **Market monitoring:** Continuously monitor market conditions and stock performance to identify opportunities and threats. Regularly review portfolio holdings and adjust strategies based on changing market dynamics.

By following these recommendations, investors can construct portfolios that prioritize stability and minimize risk while aiming to achieve satisfactory returns over the long term.

3. Dataset:

3.1 Part A-

Dataset: Credit Risk Dataset

Data Dictionary: Data Dictionary

3.2 Part B-

Dataset: Market Risk Dataset

THE END.