

FRA

Part A

Vignesh Subramaniam

Table of Contents

<u>1. 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 HIGER 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.</u>	<u>4</u>
<u>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.</u>	<u>4</u>
<u>DATA THAT IS AVAILABLE INCLUDES INFORMATION FROM THE FINANCIAL STATEMENT OF THE COMPANIES FOR THE PREVIOUS YEAR.</u>	<u>4</u>
<u>2. PART A: OUTLIER TREATMENT</u>	<u>10</u>
<u>3. PART A: MISSING VALUE TREATMENT</u>	<u>16</u>
<u>4. PART A: UNIVARIATE (4 MARKS) & BIVARIATE (6 MARKS) ANALYSIS WITH PROPER INTERPRETATION. (YOU MAY CHOOSE TO INCLUDE ONLY THOSE VARIABLES WHICH WERE SIGNIFICANT IN THE MODEL BUILDING)</u>	<u>16</u>
<u>5. PART A: TRAIN TEST SPLIT</u>	<u>25</u>
<u>6. PART A: BUILD LOGISTIC REGRESSION MODEL (USING STATSMODELS LIBRARY) ON MOST IMPORTANT VARIABLES ON TRAIN DATASET AND CHOOSE THE OPTIMUM CUT-OFF. ALSO SHOWCASE YOUR MODEL BUILDING APPROACH</u>	<u>25</u>
<u>7. PART A: VALIDATE THE MODEL ON TEST DATASET AND STATE THE PERFORMANCE METRICS. ALSO STATE INTERPRETATION FROM THE MODEL</u>	<u>25</u>
<u>8. PART A: BUILD A RANDOM FOREST MODEL ON TRAIN DATASET. ALSO SHOWCASE YOUR MODEL BUILDING APPROACH</u>	<u>29</u>
<u>9. PART A: VALIDATE THE RANDOM FOREST MODEL ON TEST DATASET AND STATE THE PERFORMANCE METRICS. ALSO STATE INTERPRETATION FROM THE MODEL</u>	<u>30</u>
<u>10. PART A: BUILD A LDA MODEL ON TRAIN DATASET. ALSO SHOWCASE YOUR MODEL BUILDING APPROACH</u>	<u>30</u>
<u>11. PART A: VALIDATE THE LDA MODEL ON TEST DATASET AND STATE THE PERFORMANCE METRICS. ALSO STATE INTERPRETATION FROM THE MODEL</u>	<u>30</u>

12. PART A: CONCLUSIONS AND RECOMMENDATIONS 31

Table of Figures

FIGURE 1.1	4
FIGURE 1.2	4
FIGURE 1.3	4
FIGURE 1.4	5
FIGURE 1.5	6
FIGURE 1.6	6
FIGURE 1.7	7
FIGURE 1.8	8
FIGURE 1.9	8
FIGURE 1.10	9
FIGURE 1.11	10
FIGURE 2.1	11
FIGURE 2.2	12
FIGURE 2.3	12
FIGURE 2.4	13
FIGURE 2.5	13
FIGURE 2.6	13
FIGURE 2.7	13
FIGURE 2.8	13
FIGURE 2.9	14
FIGURE 2.10	14
FIGURE 2.11	15
FIGURE 2.12	15
FIGURE 2.13	16
FIGURE 3.1	16
FIGURE 4.1	17
FIGURE 4.2	19
FIGURE 4.3	19
FIGURE 4.4	21
FIGURE 4.5	21
FIGURE 4.6	23
FIGURE 4.7	23
FIGURE 4.8	25
FIGURE 4.9	25
FIGURE 6.1	25
FIGURE 7.1	26
FIGURE 7.2	26
FIGURE 7.3	26
FIGURE 7.4	26
FIGURE 7.5	27
FIGURE 7.6	28
FIGURE 7.7	29
FIGURE 8.1	29
FIGURE 8.2	29
FIGURE 8.3	30
FIGURE 9.1	30
FIGURE 10.1	30
FIGURE 11.1	31

FRA Part A

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. Data that is available includes information from the financial statement of the companies for the previous year.

Head

IP_price	No_credit_Interval	Degree_of_Financial_Leverage_DFL	Interest_Coverage_Ratio	Interest_expense_to_EBIT	Net_Income_Flag	Equity_to_Liability	Default
0.03	0.62	0.03		0.57	1	0.02	0
0.01	0.62	0.03		0.56	1	0.03	1
0.00	0.62	0.03		0.57	1	0.04	0
0.00	0.62	0.03		0.57	1	0.03	0
0.01	0.62	0.03		0.57	1	0.02	0

Figure 1.1

Tail

Co_Code	Co_Name	Operating_Expense_Rate	Research_and_development_expense_rate	Cash_flow_rate	Interest_bearing_debt_interest_rate	Tax_rate
2053	2743 Kothari Ferment.	0.00	6490000000.00	0.48	0.00	0
2054	21216 Firstobj.Tech.	0.00	0.00	0.47	0.00	0
2055	142 Diamines & Chem.	0.00	8370000000.00	0.48	0.00	0
2056	18014 IL&FS Engg.	3750000000.00	0.00	0.47	0.00	0
2057	43229 Channel Nine	0.00	0.00	0.47	0.00	0

5 rows × 58 columns

Figure 1.2

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

Figure 1.3

```

<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  _CF0_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

```

Figure 1.4

	<code>Co_Code</code>	<code>_Operating_Expense_Rate</code>	<code>_Research_and_development_expense_rate</code>	<code>_Cash_flow_rate</code>	<code>_Interest_bearing_debt_interest_rate</code>	<code>_Tax_rate_A</code>	<code>_Cas</code>
<code>count</code>	2058.00	2058.00		2058.00	2058.00	2058.00	2058.00
<code>mean</code>	17572.11	2052388835.76		1208634256.56	0.47	11130223.52	0.11
<code>std</code>	21892.89	3252623690.29		2144568158.08	0.02	90425949.04	0.15
<code>min</code>	4.00	0.00		0.00	0.00	0.00	0.00
<code>25%</code>	3674.00	0.00		0.00	0.46	0.00	0.00
<code>50%</code>	6240.00	0.00		0.00	0.46	0.00	0.04
<code>75%</code>	24280.75	4110000000.00		1550000000.00	0.47	0.00	0.22
<code>max</code>	72493.00	9980000000.00		9980000000.00	1.00	990000000.00	1.00

8 rows × 57 columns

Figure 1.5

From Describe command it is observed that there are number of columns which has constant values.

	<code>Co_Code</code>	<code>Operating_Expense_Rate</code>	<code>_Research_and_development_expense_rate</code>	<code>_Cash_flow_rate</code>	<code>_Interest_bearing_debt_interest_rate</code>	<code>_Tax_rate_A</code>	<code>_Cas</code>
<code>count</code>	2058.00	2058.00		2058.00	2058.00	2058.00	2058.00
<code>mean</code>	17572.11	2052388835.76		1208634256.56	0.47	11130223.52	0.11
<code>std</code>	21892.89	3252623690.29		2144568158.08	0.02	90425949.04	0.15
<code>min</code>	4.00	0.00		0.00	0.00	0.00	0.00
<code>25%</code>	3674.00	0.00		0.00	0.46	0.00	0.00
<code>50%</code>	6240.00	0.00		0.00	0.46	0.00	0.04
<code>75%</code>	24280.75	4110000000.00		1550000000.00	0.47	0.00	0.22
<code>max</code>	72493.00	9980000000.00		9980000000.00	1.00	990000000.00	1.00

8 rows × 57 columns

Figure 1.6

From this function it looks like many of the columns are having constant values in 25%,50%,75%

Replace the or clean up the column names which has special characters or space etc..

After Fixing messy columns

```

<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

```

Figure 1.7

Checking for the variable default

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

Figure 1.8

Checking for the proportion of default

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

Figure 1.9

Checking for the Missing Values

Co_Code	0
Co_Name	0
_Operating_Expense_Rate	0
_Research_and_development_expense_rate	0
_Cash_flow_rate	0
_Interest_bearing_debt_interest_rate	0
_Tax_rate_A	0
_Cash_Flow_Per_Share	167
_Per_Share_Net_profit_before_tax_Yuan_	0
_Realized_Sales_Gross_Profit_Growth_Rate	0
_Operating_Profit_Growth_Rate	0
_Continuous_Net_Profit_Growth_Rate	0
_Total_Asset_Growth_Rate	0
_Net_Value_Growth_Rate	0
_Total_Asset_Return_Growth_Rate_Ratio	0
_Cash_Reinvestment_perc	0
_Current_Ratio	0
_Quick_Ratio	0
_Interest_Expense_Ratio	0
_Total_debt_to_Total_net_worth	21
_Long_term_fund_suitability_ratio_A	0
_Net_profit_before_tax_to_Paid_in_capital	0
_Total_Asset_Turnover	0
_Accounts_Reivable_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

Figure 1.10

There are 2% of missing values present in the dataset

Split the data set to identify the outliers present in the independent records

	Co_Code	Co_Name	Operating_Expense_Rate	_Research_and_development_expense_rate	_Cash_flow_rate	_Interest_bearing_debt_interest_rate	_Tax_
0	16974	Hind.Cables	8820000000.00		0.00	0.46	0.00
1	21214	Tata Tele. Mah.	9380000000.00		4230000000.00	0.46	0.00
2	14852	ABG Shipyard	3800000000.00		815000000.00	0.45	0.00
3	2439	GTL	6440000000.00		0.00	0.46	0.00
4	23505	Bharati Defence	3680000000.00		0.00	0.46	0.00
...
2053	2743	Kothari Ferment.	0.00		6490000000.00	0.48	0.00
2054	21216	Firstobj.Tech.	0.00		0.00	0.47	0.00
2055	142	Diamines & Chem.	0.00		8370000000.00	0.48	0.00
2056	18014	IL&FS Engg.	3750000000.00		0.00	0.47	0.00
2057	43229	Channel Nine	0.00		0.00	0.47	0.00

2058 rows × 57 columns

Figure 1.11

2. PART A: Outlier Treatment

Checking for the number of outliers

Adding the Upper and Lower outliers present in each columns

Co_Code	241
Co_Name	0
_Accounts_Receivable_Turnover	281
_Allocation_rate_per_person	200
_Average_Collection_Days	77
_CFO_to_Assets	110
_Cash_Flow_Per_Share	146
_Cash_Flow_to_Equity	306
_Cash_Flow_to_Liability	407
_Cash_Flow_to_Total_Assets	317
_Cash_Reinvestment_perc	220
_Cash_Turnover_Rate	0
_Cash_flow_rate	206
_Cash_to_Current_Liability	253
_Cash_to_Total_Assets	163
_Continuous_Net_Profit_Growth_Rate	340
_Current_Asset_Turnover_Rate	464
_Current_Liability_to_Current_Assets	121
_Current_Ratio	193
_Degree_of_Financial_Leverage_DFL	438
_Equity_to_Liability	190
_Fixed_Assets_Turnover_Frequency	501
_Fixed_Assets_to_Assets	10
_Interest_Coverage_Ratio_Interest_expense_to_EBIT	376
_Interest_Expense_Ratio	328
_Interest_bearing_debt_interest_rate	94
_Inventory_Turnover_Rate_times	29
_Inventory_to_Current_Liability	129
_Inventory_to_Working_Capital	247
_Liability_Assets_Flag	7
_Long_term_Liability_to_Current_Assets	213
_Long_term_fund_suitability_ratio_A	234
_Net_Income_Flag	0
_Net_Value_Growth_Rate	304
_Net_Worth_Turnover_Rate_times	165
_Net_profit_before_tax_to_Paid_in_capital	173
_No_credit_Interval	396
_Operating_Expense_Rate	0
_Operating_Funds_to_Liability	219
_Operating_Profit_Growth_Rate	317
_Operating_profit_per_person	357
_Per_Share_Net_profit_before_tax_Yuan_	186
_Quick_Asset_Turnover_Rate	0
_Quick_Assets_to_Current_Liability	185
_Quick_Assets_to_Total_Assets	4
_Quick_Ratio	190
_Realized_Sales_Gross_Profit_Growth_Rate	283
_Research_and_development_expense_rate	264
_Retained_Earnings_to_Total_Assets	208
_Tax_rate_A	42
_Total_Asset_Growth_Rate	0
_Total_Asset_Return_Growth_Rate_Ratio	226
_Total_Asset_Turnover	101
_Total_assets_to_GNP_price	235
_Total_debt_to_Total_net_worth	105
_Total_expense_to_Assets	168
_Total_income_to_Total_expense	136
dtype:	int64

Figure 2.1

Outliers Present in the data set

11105

Figure 2.2

Converting the outliers to Null values

Co_Code	241
Co_Name	0
_Operating_Expense_Rate	0
_Research_and_development_expense_rate	264
_Cash_flow_rate	206
_Interest_bearing_debt_interest_rate	94
_Tax_rate_A	42
_Cash_Flow_Per_Share	313
_Per_Share_Net_profit_before_tax_Yuan	186
_Realized_Sales_Gross_Profit_Growth_Rate	283
_Operating_Profit_Growth_Rate	317
_Continuous_Net_Profit_Growth_Rate	340
_Total_Asset_Growth_Rate	0
_Net_Value_Growth_Rate	304
_Total_Asset_Return_Growth_Rate_Ratio	226
_Cash_Reinvestment_perc	220
_Current_Ratio	193
_Quick_Ratio	190
_Interest_Expense_Ratio	328
_Total_debt_to_Total_net_worth	126
_Long_term_fund_suitability_ratio_A	234
_Net_profit_before_tax_to_Paid_in_capital	173
_Total_Asset_Turnover	101
_Accounts_Receivable_Turnover	281
_Average_Collection_Days	77
_Inventory_Turnover_Rate_times	29
_Fixed_Assets_Turnover_Frequency	501
_Net_Worth_Turnover_Rate_times	165
_Operating_profit_per_person	357
_Allocation_rate_per_person	200
_Quick_Assets_to_Total_Assets	4
_Cash_to_Total_Assets	259
_Quick_Assets_to_Current_Liability	185
_Cash_to_Current_Liability	253
_Operating_Funds_to_Liability	219
_Inventory_to_Working_Capital	247
_Inventory_to_Current_Liability	129
_Long_term_Liability_to_Current_Assets	213
_Retained_Earnings_to_Total_Assets	208
_Total_income_to_Total_expense	136
_Total_expense_to_Assets	168
_Current_Asset_Turnover_Rate	464
_Quick_Asset_Turnover_Rate	0
_Cash_Turnover_Rate	0
_Fixed_Assets_to_Assets	10
_Cash_Flow_to_Total_Assets	317
_Cash_Flow_to_Liability	407
_CFO_to_Assets	110
_Cash_Flow_to_Equity	306
_Current_Liability_to_Current_Assets	135
_Liability_Assets_Flag	7
_Total_assets_to_GNP_price	235
_No_credit_Interval	396
_Degree_of_Financial_Leverage_DFL	438
_Interest_Coverage_Ratio_Interest_expense_to_EBIT	376
_Net_Income_Flag	0
_Equity_to_Liability	190
dtype: int64	

Figure 2.3

Total added values of Null and outliers

```
11403
```

Figure 2.4

Removing the Columns which has constant values in the dataset

After removing the columns which has constant values in the dataset

```
(2058, 50)
```

Figure 2.5

Checking for total missing value for each row

```
0      4  
1      8  
2      1  
3      5  
4      1  
..  
2053   10  
2054   4  
2055   6  
2056   0  
2057   4  
Length: 2058, dtype: int64
```

Figure 2.6

Total missing values in terms of rows

```
10382
```

Figure 2.7

Filtering the data which is 90% or more records at the row level

```
(1262, 51)
```

Figure 2.8

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1262 entries, 0 to 2057
Data columns (total 51 columns):
 #   Column           Non-Null Count Dtype  
--- 
 0   _Operating_Expense_Rate    1262 non-null   float64 
 1   _Research_and_development_expense_rate 1115 non-null   float64 
 2   _Cash_flow_rate            1232 non-null   float64 
 3   _Tax_rate_A               1237 non-null   float64 
 4   _Cash_Flow_Per_Share     1154 non-null   float64 
 5   _Per_Share_Net_profit_before_tax_Yuan_ 1233 non-null   float64 
 6   _Operating_Profit_Growth_Rate        1132 non-null   float64 
 7   _Continuous_Net_Profit_Growth_Rate   1117 non-null   float64 
 8   _Total_Asset_Growth_Rate          1262 non-null   float64 
 9   _Net_Value_Growth_Rate          1182 non-null   float64 
 10  _Total_Asset_Return_Growth_Rate_Ratio 1215 non-null   float64 
 11  _Cash_Reinvestment_perc        1211 non-null   float64 
 12  _Current_Ratio              1244 non-null   float64 
 13  _Quick_Ratio                1246 non-null   float64 
 14  _Interest_Expense_Ratio      1101 non-null   float64 
 15  _Total_debt_to_Total_net_worth 1238 non-null   float64 
 16  _Long_term_fund_suitability_ratio_A 1184 non-null   float64 
 17  _Net_profit_before_tax_to_Paid_in_capital 1232 non-null   float64 
 18  _Total_Asset_Turnover        1237 non-null   float64 
 19  _Accounts_Receivable_Turnover 1178 non-null   float64 
 20  _Average_Collection_Days    1235 non-null   float64 
 21  _Inventory_Turnover_Rate_times 1241 non-null   float64 
 22  _Fixed_Assets_Turnover_Frequency 1030 non-null   float64 
 23  _Net_Worth_Turnover_Rate_times 1196 non-null   float64 
 24  _Operating_profit_per_person 1152 non-null   float64 
 25  _Allocation_rate_per_person 1200 non-null   float64 
 26  _Quick_Assets_to_Total_Assets 1262 non-null   float64 
 27  _Cash_to_Total_Assets       1188 non-null   float64 
 28  _Quick_Assets_to_Current_Liability 1245 non-null   float64 
 29  _Cash_to_Current_Liability 1221 non-null   float64 
 30  _Operating_Funds_to_Liability 1231 non-null   float64 
 31  _Inventory_to_Working_Capital 1135 non-null   float64 
 32  _Inventory_to_Current_Liability 1227 non-null   float64 
 33  _Long_term_Liability_to_Current_Assets 1183 non-null   float64 
 34  _Retained_Earnings_to_Total_Assets 1211 non-null   float64 
 35  _Total_income_to_Total_expense 1219 non-null   float64 
 36  _Total_expense_to_Assets     1218 non-null   float64 
 37  _Current_Asset_Turnover_Rate 1036 non-null   float64 
 38  _Quick_Asset_Turnover_Rate   1262 non-null   float64 
 39  _Cash_Turnover_Rate         1262 non-null   float64 
 40  _Fixed_Assets_to_Assets    1261 non-null   float64 
 41  _Cash_Flow_to_Total_Assets 1190 non-null   float64 
 42  _Cash_Flow_to_Liability     1149 non-null   float64 
 43  _CFO_to_Assets             1252 non-null   float64 
 44  _Cash_Flow_to_Equity       1196 non-null   float64 
 45  _Current_Liability_to_Current_Assets 1221 non-null   float64 
 46  _Total_assets_to_GNP_price 1165 non-null   float64 
 47  _Degree_of_Financial_Leverage_DFL 992 non-null   float64 
 48  _Interest_Coverage_Ratio_Interest_expense_to_EBIT 1058 non-null   float64 
 49  _Equity_to_Liability       1229 non-null   float64 
 50  Default                   1262 non-null   int64 

dtypes: float64(50), int64(1)
memory usage: 512.7 KB

```

Figure 2.9

```

0    1162
1    100
Name: Default, dtype: int64

```

Figure 2.10

```

0    0.92
1    0.08
Name: Default, dtype: float64

```

Figure 2.11

- Initially we had more than 10% of defaulters now it has been reduced to just 7% where we are missing out on some of the dataset
- From the below table it is clear that null values are below 30% for all the columns now.
- So there is no need to remove or delete any of the columns

_Fixed_Assets_Turnover_Frequency	0.24
_Current_Asset_Turnover_Rate	0.23
_Degree_of_Financial_Leverage_DFL	0.21
_Cash_Flow_to_Liability	0.20
_Interest_Coverage_Ratio_Interest_expense_to_EBIT	0.18
_Operating_profit_per_person	0.17
_Continuous_Net_Profit_Growth_Rate	0.17
_Interest_Expense_Ratio	0.16
_Operating_Profit_Growth_Rate	0.15
_Cash_Flow_to_Total_Assets	0.15
_Cash_Flow_Per_Share	0.15
_Cash_Flow_to_Equity	0.15
_Net_Value_Growth_Rate	0.15
_Accounts_Reivable_Turnover	0.14
_Research_and_development_expense_rate	0.13
_Cash_to_Total_Assets	0.13
_Cash_to_Current_Liability	0.12
_Inventory_to_Working_Capital	0.12
_Total_assets_to_GNP_price	0.11
_Long_term_fund_suitability_ratio_A	0.11
_Total_Asset_Return_Growth_Rate_Ratio	0.11
_Cash_Reinvestment_perc	0.11
_Operating_Funds_to_Liability	0.11
_Long_term_Liability_to_Current_Assets	0.10
_Retained_Earnings_to_Total_Assets	0.10
_Cash_flow_rate	0.10
_Allocation_rate_per_person	0.10
_Current_Ratio	0.09
_Quick_Ratio	0.09
_Equity_to_Liability	0.09
_Per_Share_Net_profit_before_tax_Yuan_	0.09
_Quick_Assets_to_Current_Liability	0.09
_Net_profit_before_tax_to_Paid_in_capital	0.08
_Total_expense_to_Assets	0.08
_Net_Worth_Turnover_Rate_times	0.08
_Total_income_to_Total_expense	0.07
_Current_Liability_to_Current_Assets	0.07
_Inventory_to_Current_Liability	0.06
_Total_debt_to_Total_net_worth	0.06
_CFO_to_Assets	0.05
_Total_Asset_Turnover	0.05
_Average_Collection_Days	0.04
_Tax_rate_A	0.02
_Inventory_Turnover_Rate_times	0.01
_Fixed_Assets_to_Assets	0.00
_Quick_Assets_to_Total_Assets	0.00
_Operating_Expense_Rate	0.00
_Cash_Turnover_Rate	0.00
_Quick_Asset_Turnover_Rate	0.00
_Total_Asset_Growth_Rate	0.00
Default	0.00
dtype:	float64

Figure 2.12

Segerigate the Predictor and response

Scale the predictors

```
Index(['Operating_Expense_Rate', 'Research_and_development_expense_rate',
       'Cash_flow_rate', 'Tax_rate_A', 'Cash_Flow_Per_Share',
       'Per_Share_Net_profit_before_tax_Yuan_',
       'Operating_Profit_Growth_Rate', 'Continuous_Net_Profit_Growth_Rate',
       'Total_Asset_Growth_Rate', 'Net_Value_Growth_Rate',
       'Total_Asset_Return_Growth_Rate_Ratio', 'Cash_Reinvestment_perc',
       'Current_Ratio', 'Quick_Ratio', 'Interest_Expense_Ratio',
       'Total_debt_to_Total_net_worth', 'Long_term_fund_suitability_ratio_A',
       'Net_profit_before_tax_to_Paid_in_capital', 'Total_Asset_Turnover',
       'Accounts_Reivable_Turnover', 'Average_Collection_Days',
       'Inventory_Turnover_Rate_times', 'Fixed_Assets_Turnover_Frequency',
       'Net_Worth_Turnover_Rate_times', 'Operating_profit_per_person',
       'Allocation_rate_per_person', 'Quick_Assets_to_Total_Assets',
       'Cash_to_Total_Assets', 'Quick_Assets_to_Current_Liability',
       'Cash_to_Current_Liability', 'Operating_Funds_to_Liability',
       'Inventory_to_Working_Capital', 'Inventory_to_Current_Liability',
       'Long_term_Liability_to_Current_Assets',
       'Retained_Earnings_to_Total_Assets', 'Total_income_to_Total_expense',
       'Total_expense_to_Assets', 'Current_Asset_Turnover_Rate',
       'Quick_Asset_Turnover_Rate', 'Cash_Turnover_Rate',
       'Fixed_Assets_to_Assets', 'Cash_Flow_to_Total_Assets',
       'Cash_Flow_to_Liability', 'CFO_to_Assets', 'Cash_Flow_to_Equity',
       'Current_Liability_to_Current_Assets', 'Total_assets_to_GNP_price',
       'Degree_of_Financial_Leverage_DFL',
       'Interest_Coverage_Ratio_Interest_expense_to_EBIT',
       'Equity_to_Liability', 'Default'],
      dtype='object')
```

Figure 2.13

3. PART A: Missing Value Treatment

Performing the Train and Test Split

Treating the missing values

```
0  
0
```

Figure 3.1

4. PART A: Univariate (4 marks) & Bivariate (6 marks) analysis with proper interpretation. (You may choose to include only those variables which were significant in the model building)

- Checking for correlations between variables
- Here the correlations are being checked only for the Independent variables
- Variable which are significant to the models

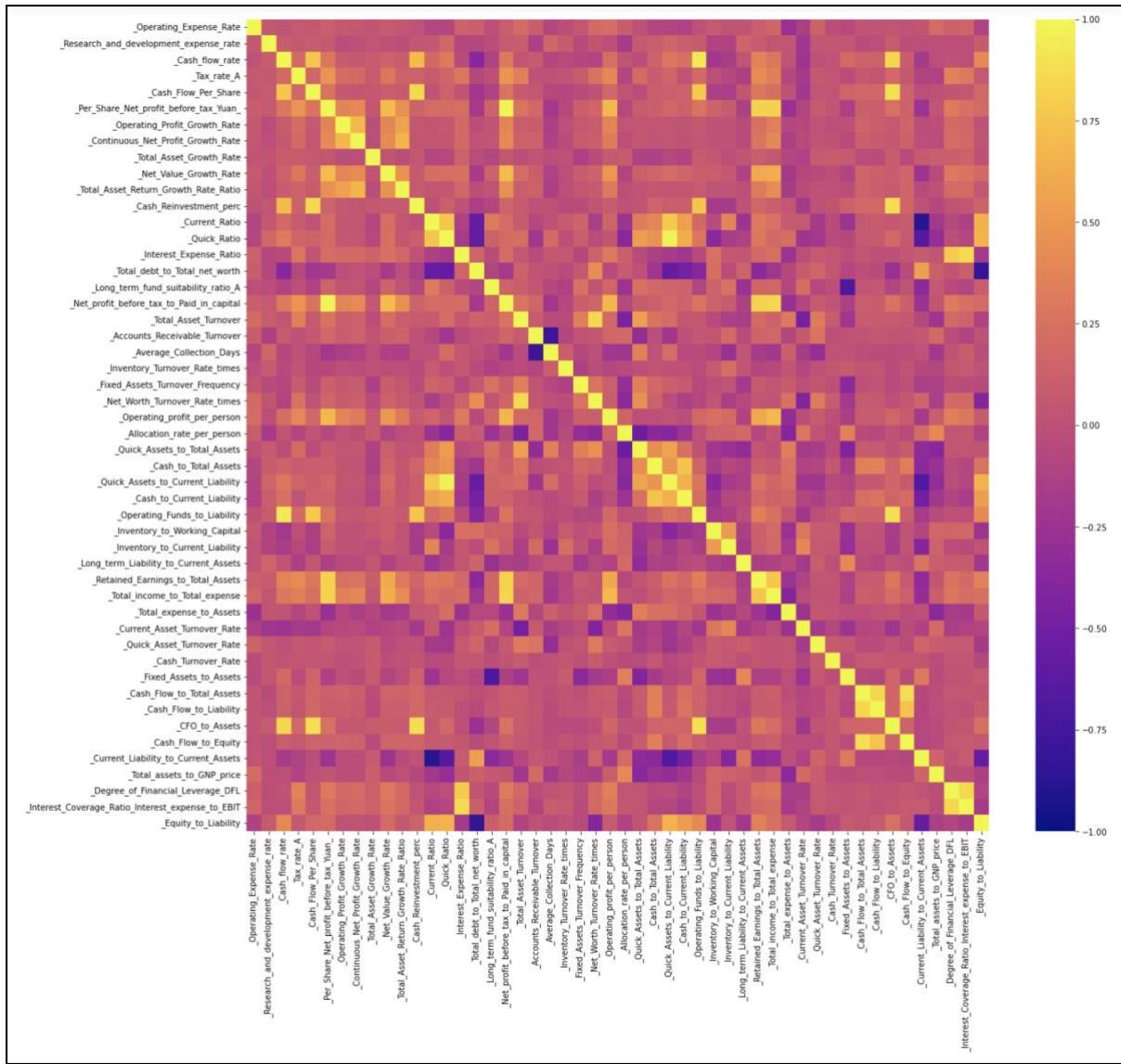


Figure 4.1

- Presence of correlation is clearly visible from the above plot
- Since there is a presence of correlation, Applying the VIF and sorting the columns which has VIF score of less than or equal to 5

	variables	VIF
17	_Net_profit_before_tax_to_Paid_in_capital	45.65
5	_Per_Share_Net_profit_before_tax_Yuan_	44.58
28	_Quick_Assets_to_Current_Liability	34.25
13	_Quick_Ratio	23.34
30	_Operating_Funds_to_Liability	11.89
43	_CFO_to_Assets	11.28
12	_Current_Ratio	10.53
2	_Cash_flow_rate	9.75
41	_Cash_Flow_to_Total_Assets	7.58
18	_Total_Asset_Turnover	7.50
23	_Net_Worth_Turnover_Rate_times	6.86
48	_Interest_Coverage_Ratio_Interest_expense_to_EBIT	6.38
26	_Quick_Assets_to_Total_Assets	5.98
4	_Cash_Flow_Per_Share	5.81
11	_Cash_Reinvestment_perc	5.78
45	_Current_Liability_to_Current_Assets	5.67
49	_Equity_to_Liability	5.38
35	_Total_income_to_Total_expense	4.74
14	_Interest_Expense_Ratio	4.73
44	_Cash_Flow_to_Equity	4.64
47	_Degree_of_Financial_Leverage_DFL	4.42
15	_Total_debt_to_Total_net_worth	4.29
34	_Retained_Earnings_to_Total_Assets	4.20
32	_Inventory_to_Current_Liability	4.13
29	_Cash_to_Current_Liability	4.08
42	_Cash_Flow_to_Liability	3.95
27	_Cash_to_Total_Assets	3.40
40	_Fixed_Assets_to_Assets	3.26
20	_Average_Collection_Days	3.15
24	_Operating_profit_per_person	2.79
19	_Accounts_Receivable_Turnover	2.71
7	_Continuous_Net_Profit_Growth_Rate	2.55
9	_Net_Value_Growth_Rate	2.50
37	_Current_Asset_Turnover_Rate	2.37
16	_Long_term_fund_suitability_ratio_A	2.27
25	_Allocation_rate_per_person	2.21

Figure 4.2

10	_Total_Asset_Return_Growth_Rate_Ratio	2.20
6	_Operating_Profit_Growth_Rate	2.17
31	_Inventory_to_Working_Capital	2.02
36	_Total_expense_to_Assets	1.99
33	_Long_term_Liability_to_Current_Assets	1.67
3	_Tax_rate_A	1.66
46	_Total_assets_to_GNP_price	1.55
38	_Quick_Asset_Turnover_Rate	1.48
22	_Fixed_Assets_Turnover_Frequency	1.41
0	_Operating_Expense_Rate	1.34
21	_Inventory_Turnover_Rate_times	1.28
1	_Research_and_development_expense_rate	1.20
8	_Total_Asset_Growth_Rate	1.12
39	_Cash_Turnover_Rate	1.09

Figure 4.3

	variables	VIF
27	_Quick_Assets_to_Current_Liability	34.24
13	_Quick_Ratio	23.34
29	_Operating_Funds_to_Liability	11.88
42	_CFO_to_Assets	11.28
12	_Current_Ratio	10.52
2	_Cash_flow_rate	9.74
40	_Cash_Flow_to_Total_Assets	7.58
17	_Total_Asset_Turnover	7.50
22	_Net_Worth_Turnover_Rate_times	6.86
47	_Interest_Coverage_Ratio_Interest_expense_to_EBIT	6.38
5	_Per_Share_Net_profit_before_tax_Yuan_	6.33
25	_Quick_Assets_to_Total_Assets	5.96
4	_Cash_Flow_Per_Share	5.81
11	_Cash_Reinvestment_perc	5.78
44	_Current_Liability_to_Current_Assets	5.67
48	_Equity_to_Liability	5.37
14	_Interest_Expense_Ratio	4.73
43	_Cash_Flow_to_Equity	4.64
34	_Total_income_to_Total_expense	4.63
46	_Degree_of_Financial_Leverage_DFL	4.42
15	_Total_debt_to_Total_net_worth	4.29
33	_Retained_Earnings_to_Total_Assets	4.15
31	_Inventory_to_Current_Liability	4.13
28	_Cash_to_Current_Liability	4.08
41	_Cash_Flow_to_Liability	3.93
26	_Cash_to_Total_Assets	3.39
39	_Fixed_Assets_to_Assets	3.26
19	_Average_Collection_Days	3.15
23	_Operating_profit_per_person	2.79
18	_Accounts_Receivable_Turnover	2.70
7	_Continuous_Net_Profit_Growth_Rate	2.55
9	_Net_Value_Growth_Rate	2.50
36	_Current_Asset_Turnover_Rate	2.37
16	_Long_term_fund_suitability_ratio_A	2.27
24	_Allocation_rate_per_person	2.21
10	_Total_Asset_Return_Growth_Rate_Ratio	2.20

Figure 4.4

6	_Operating_Profit_Growth_Rate	2.17
30	_Inventory_to_Working_Capital	2.02
35	_Total_expense_to_Assets	1.99
3	_Tax_rate_A	1.66
32	_Long_term_Liability_to_Current_Assets	1.66
45	_Total_assets_to_GNP_price	1.54
37	_Quick_Asset_Turnover_Rate	1.48
21	_Fixed_Assets_Turnover_Frequency	1.41
0	_Operating_Expense_Rate	1.34
20	_Inventory_Turnover_Rate_times	1.28
1	_Research_and_development_expense_rate	1.20
8	_Total_Asset_Growth_Rate	1.12
38	_Cash_Turnover_Rate	1.09

Figure 4.5

	variables	VIF
28	_Operating_Funds_to_Liability	11.87
41	_CFO_to_Assets	11.27
2	_Cash_flow_rate	9.74
12	_Current_Ratio	8.94
13	_Quick_Ratio	7.80
39	_Cash_Flow_to_Total_Assets	7.58
17	_Total_Asset_Turnover	7.36
22	_Net_Worth_Turnover_Rate_times	6.84
46	_Interest_Coverage_Ratio_Interest_expense_to_EBIT	6.35
5	_Per_Share_Net_profit_before_tax_Yuan_	6.33
4	_Cash_Flow_Per_Share	5.80
11	_Cash_Reinvestment_perc	5.78
43	_Current_Liability_to_Current_Assets	5.66
25	_Quick_Assets_to_Total_Assets	5.28
47	_Equity_to_Liability	5.19
14	_Interest_Expense_Ratio	4.72
42	_Cash_Flow_to_Equity	4.64
33	_Total_income_to_Total_expense	4.62
45	_Degree_of_Financial_Leverage_DFL	4.42
15	_Total_debt_to_Total_net_worth	4.26
32	_Retained_Earnings_to_Total_Assets	4.15
27	_Cash_to_Current_Liability	4.05
40	_Cash_Flow_to_Liability	3.92
30	_Inventory_to_Current_Liability	3.83
26	_Cash_to_Total_Assets	3.39
38	_Fixed_Assets_to_Assets	3.26
19	_Average_Collection_Days	3.15
23	_Operating_profit_per_person	2.77
18	_Accounts_Receivable_Turnover	2.70
7	_Continuous_Net_Profit_Growth_Rate	2.55
9	_Net_Value_Growth_Rate	2.50
35	_Current_Asset_Turnover_Rate	2.35
16	_Long_term_fund_suitability_ratio_A	2.26
24	_Allocation_rate_per_person	2.20
10	_Total_Asset_Return_Growth_Rate_Ratio	2.20
6	_Operating_Profit_Growth_Rate	2.17

Figure 4.6

29	_Inventory_to_Working_Capital	2.02
34	_Total_expense_to_Assets	1.98
3	_Tax_rate_A	1.66
31	_Long_term_Liability_to_Current_Assets	1.64
44	_Total_assets_to_GNP_price	1.54
36	_Quick_Asset_Turnover_Rate	1.48
21	_Fixed_Assets_Turnover_Frequency	1.40
0	_Operating_Expense_Rate	1.34
20	_Inventory_Turnover_Rate_times	1.28
1	_Research_and_development_expense_rate	1.20
8	_Total_Asset_Growth_Rate	1.12
37	_Cash_Turnover_Rate	1.09

Figure 4.7

	variables	VIF
27	_Operating_Funds_to_Liability	11.83
40	_CFO_to_Assets	11.26
2	_Cash_flow_rate	9.71
12	_Current_Ratio	8.93
13	_Quick_Ratio	7.72
38	_Cash_Flow_to_Total_Assets	7.54
45	_Interest_Coverage_Ratio_Interest_expense_to_EBIT	6.35
5	_Per_Share_Net_profit_before_tax_Yuan_	6.33
4	_Cash_Flow_Per_Share	5.79
11	_Cash_Reinvestment_perc	5.77
42	_Current_Liability_to_Current_Assets	5.65
46	_Equity_to_Liability	5.19
24	_Quick_Assets_to_Total_Assets	5.01
14	_Interest_Expense_Ratio	4.71
32	_Total_income_to_Total_expense	4.60
41	_Cash_Flow_to_Equity	4.60
44	_Degree_of_Financial_Leverage_DFL	4.42
31	_Retained_Earnings_to_Total_Assets	4.11
15	_Total_debt_to_Total_net_worth	4.10
26	_Cash_to_Current_Liability	4.03
39	_Cash_Flow_to_Liability	3.92
29	_Inventory_to_Current_Liability	3.79
21	_Net_Worth_Turnover_Rate_times	3.43
25	_Cash_to_Total_Assets	3.38
37	_Fixed_Assets_to_Assets	3.26
18	_Average_Collection_Days	3.12
22	_Operating_profit_per_person	2.76
17	_Accounts_Receivable_Turnover	2.68
7	_Continuous_Net_Profit_Growth_Rate	2.55
9	_Net_Value_Growth_Rate	2.50
16	_Long_term_fund_suitability_ratio_A	2.26
34	_Current_Asset_Turnover_Rate	2.25
23	_Allocation_rate_per_person	2.20
10	_Total_Asset_Return_Growth_Rate_Ratio	2.18
6	_Operating_Profit_Growth_Rate	2.17
28	_Inventory_to_Working_Capital	2.00

Figure 4.8

33	_Total_expense_to_Assets	1.97
3	_Tax_rate_A	1.64
30	_Long_term_Liability_to_Current_Assets	1.64
43	_Total_assets_to_GNP_price	1.54
35	_Quick_Asset_Turnover_Rate	1.41
20	_Fixed_Assets_Turnover_Frequency	1.39
0	_Operating_Expense_Rate	1.32
19	_Inventory_Turnover_Rate_times	1.27
1	_Research_and_development_expense_rate	1.20
8	_Total_Asset_Growth_Rate	1.12
36	_Cash_Turnover_Rate	1.08

Figure 4.9

5. PART A: Train Test Split

6. PART A: Build Logistic Regression Model (using statsmodels library) on most important variables on train dataset and choose the optimum cut-off. Also showcase your model building approach

Have performed the Logistic regression using scikit as well as stats model

Logistic Regression

	precision	recall	f1-score	support
0.0	0.95	0.98	0.96	827
1.0	0.73	0.55	0.63	96
accuracy			0.93	923
macro avg	0.84	0.76	0.79	923
weighted avg	0.93	0.93	0.93	923

Figure 6.1

7. PART A: Validate the Model on Test Dataset and state the performance metrics. Also state interpretation from the model

	precision	recall	f1-score	support
0.0	0.92	0.97	0.94	398
1.0	0.66	0.40	0.50	57
accuracy			0.90	455
macro avg	0.79	0.69	0.72	455
weighted avg	0.89	0.90	0.89	455

Figure 7.1

- We need to consider only the recall value here for both train and test split.
- It is very much clear that recall value for both train and test split is very low.
- This could also be because of the distribution of data between train and test.
- Hence we need to try to implement the SMOTE technique.

	precision	recall	f1-score	support
0.0	0.91	0.88	0.89	827
1.0	0.84	0.88	0.86	620
accuracy			0.88	1447
macro avg	0.87	0.88	0.88	1447
weighted avg	0.88	0.88	0.88	1447

Figure 7.2

	precision	recall	f1-score	support
0.0	0.96	0.90	0.93	398
1.0	0.51	0.72	0.60	57
accuracy			0.88	455
macro avg	0.73	0.81	0.76	455
weighted avg	0.90	0.88	0.89	455

Figure 7.3

- Hence after applying the smote technique in the Logistic regression model we could clearly see that the recall value for train and test split has improved drastically.
- This model can be termed as a better model considering the fact that the given data set has lots of missing values, outliers and correlation problems.

Logistic Regression using stats model have used stats model only to obtain the Optimization

```
Optimization terminated successfully.
    Current function value: 0.171847
    Iterations 10
```

Figure 7.4

Logit Regression Results			
Dep. Variable:	Default	No. Observations:	923
Model:	Logit	Df Residuals:	875
Method:	MLE	Df Model:	47
Date:	Sun, 26 Nov 2023	Pseudo R-squ.:	0.4852
Time:	20:13:58	Log-Likelihood:	-158.61
converged:	True	LL-Null:	-308.10
Covariance Type:	nonrobust	LLR p-value:	2.233e-38

Figure 7.5

		coef	std err	z	P> z	[0.025	0.975]
	Intercept	-4.5454	0.448	-10.154	0.000	-5.423	-3.668
	_Operating_Funds_to_Liability	-0.1188	0.686	-0.173	0.862	-1.464	1.226
	_CFO_to_Assets	-0.9538	0.562	-1.698	0.089	-2.055	0.147
	_Cash_flow_rate	0.2850	0.646	0.441	0.659	-0.981	1.551
	_Current_Ratio	0.2200	0.712	0.309	0.757	-1.175	1.615
	_Quick_Ratio	-2.3447	0.652	-3.599	0.000	-3.622	-1.068
	_Cash_Flow_to_Total_Assets	0.5330	0.524	1.016	0.309	-0.495	1.561
	_Interest_Coverage_Ratio_Interest_expense_to_EBIT	-0.0260	0.306	-0.085	0.932	-0.626	0.574
	_Per_Share_Net_profit_before_tax_Yuan_	-0.0491	0.402	-0.122	0.903	-0.837	0.739
	_Cash_Flow_Per_Share	0.1753	0.363	0.483	0.629	-0.536	0.887
	_Cash_Reinvestment_perc	0.0841	0.332	0.254	0.800	-0.566	0.735
	_Current_Liability_to_Current_Assets	0.0257	0.335	0.076	0.939	-0.632	0.683
	_Equity_to_Liability	-1.0002	0.582	-1.719	0.086	-2.141	0.140
	_Quick_Assets_to_Total_Assets	0.5929	0.432	1.372	0.170	-0.254	1.440
	_Interest_Expense_Ratio	-0.0724	0.289	-0.251	0.802	-0.639	0.494
	_Total_income_to_Total_expense	-0.7115	0.401	-1.775	0.076	-1.497	0.074
	_Cash_Flow_to_Equity	0.2270	0.331	0.686	0.493	-0.422	0.876
	_Degree_of_Financial_Leverage_DFL	0.0014	0.263	0.005	0.996	-0.514	0.517
	_Retained_Earnings_to_Total_Assets	0.0638	0.312	0.204	0.838	-0.548	0.676
	_Total_debt_to_Total_net_worth	0.2929	0.289	1.014	0.310	-0.273	0.859
	_Cash_to_Current_Liability	0.9325	0.433	2.156	0.031	0.085	1.780
	_Cash_Flow_to_Liability	-1.0191	0.445	-2.290	0.022	-1.891	-0.147
	_Inventory_to_Current_Liability	0.1622	0.461	0.352	0.725	-0.741	1.066
	_Net_Worth_Turnover_Rate_times	-0.2810	0.263	-1.068	0.286	-0.797	0.235
	_Cash_to_Total_Assets	-0.0432	0.323	-0.134	0.894	-0.676	0.590
	_Fixed_Assets_to_Assets	0.0442	0.280	0.158	0.874	-0.504	0.592
	_Average_Collection_Days	0.3731	0.277	1.349	0.177	-0.169	0.915
	_Operating_profit_per_person	0.3291	0.253	1.298	0.194	-0.168	0.826
	_Accounts_Receivable_Turnover	-0.5358	0.306	-1.748	0.080	-1.136	0.065
	_Continuous_Net_Profit_Growth_Rate	0.0547	0.228	0.240	0.810	-0.392	0.501
	_Net_Value_Growth_Rate	-0.5821	0.268	-2.171	0.030	-1.108	-0.057
	_Long_term_fund_suitability_ratio_A	-0.2203	0.249	-0.885	0.376	-0.708	0.267
	_Current_Asset_Turnover_Rate	-0.4055	0.275	-1.475	0.140	-0.944	0.133
	_Allocation_rate_per_person	0.4357	0.227	1.921	0.055	-0.009	0.880
	_Total_Asset_Return_Growth_Rate_Ratio	-0.1589	0.239	-0.666	0.506	-0.627	0.309

Figure 7.6

_Operating_Profit_Growth_Rate	0.1912	0.197	0.968	0.333	-0.196	0.578
_Inventory_to_Working_Capital	-0.2304	0.188	-1.225	0.221	-0.599	0.138
_Total_expense_to_Assets	0.2987	0.212	1.407	0.160	-0.117	0.715
_Tax_rate_A	-0.5353	0.263	-2.036	0.042	-1.051	-0.020
_Long_term_Liability_to_Current_Assets	-0.1673	0.199	-0.839	0.402	-0.558	0.224
_Total_assets_to_GNP_price	0.0525	0.181	0.289	0.772	-0.303	0.408
_Quick_Asset_Turnover_Rate	0.1662	0.184	0.905	0.365	-0.194	0.526
_Fixed_Assets_Turnover_Frequency	0.1682	0.170	0.988	0.323	-0.166	0.502
_Operating_Expense_Rate	0.1209	0.175	0.692	0.489	-0.222	0.463
_Inventory_Turnover_Rate_times	0.0354	0.176	0.202	0.840	-0.309	0.380
_Research_and_development_expense_rate	0.2477	0.168	1.478	0.139	-0.081	0.576
_Total_Asset_Growth_Rate	-0.2137	0.182	-1.175	0.240	-0.570	0.143
_Cash_Turnover_Rate	-0.1926	0.174	-1.105	0.269	-0.534	0.149

Figure 7.7

Possibly complete quasi-separation:

- A fraction 0.11 of observations can be perfectly predicted.
- This might indicate that there is complete quasi-separation.
- In this case some parameters will not be identified.

8. PART A: Build a Random Forest Model on Train Dataset. Also showcase your model building approach

```
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]})
```

Figure 8.1

```
{'max_depth': 7,
 'min_samples_leaf': 10,
 'min_samples_split': 30,
 'n_estimators': 50}
```

Figure 8.2

	precision	recall	f1-score	support
0.0	0.94	0.99	0.97	827
1.0	0.88	0.46	0.60	96
accuracy			0.94	923
macro avg	0.91	0.73	0.78	923
weighted avg	0.93	0.94	0.93	923

Figure 8.3

9. PART A: Validate the Random Forest Model on test Dataset and state the performance metrics. Also state interpretation from the model

	precision	recall	f1-score	support
0.0	0.90	0.98	0.94	398
1.0	0.67	0.28	0.40	57
accuracy			0.89	455
macro avg	0.79	0.63	0.67	455
weighted avg	0.88	0.89	0.87	455

Figure 9.1

- The recall value for test data set is very low compared to that of train data set, The RF model looks underfit.
- As been tried for Logistic regression model SMOTE technique can also be implemented The data still consists of high Pvalues which also could be the cause of poor prediction by the model.

10. PART A: Build a LDA Model on Train Dataset. Also showcase your model building approach

	precision	recall	f1-score	support
0.0	0.95	0.96	0.95	827
1.0	0.61	0.53	0.57	96
accuracy			0.92	923
macro avg	0.78	0.75	0.76	923
weighted avg	0.91	0.92	0.91	923

Figure 10.1

11. PART A: Validate the LDA Model on test Dataset and state the performance metrics. Also state interpretation from the model

	precision	recall	f1-score	support
0.0	0.93	0.95	0.94	398
1.0	0.57	0.47	0.52	57
accuracy			0.89	455
macro avg	0.75	0.71	0.73	455
weighted avg	0.88	0.89	0.89	455

Figure 11.1

Compared to Random Forest model LinearDiscriminant model has performed well but still the Recall is not the best.

12. PART A: Conclusions and Recommendations

- The data set is about predicting the customers who will be defaulter or not Hence this can be considered as the classification model since being an classification approach we need to check for data split as well if it unbalanced the model wont perform as per the expectation.
- Hence have performed the SMOTE technique in one of the models.
- The targeted column here is the Default and if we check for the proportion of default is 10% which is high in banking perspective.
- From our model perspective Linear Regression model has performed well after implementing the SMOTE technique compared to all other models.
- Missing values Outlier and correlations were observed in the dataset which were treated but still the pvalues for some of the columns were observed to be high.
- LinearDiscrim model has performed better when compared to RandomForest.