Business Report
On
Indian credit risk

Name: Anthony Prasath

Course: PGP-DSBA Online – FRA Project

Date: 07/Aug/2022

Table of Contents

1. Introduction of the Business Problem

1.1 OBJECTIVE

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 interests 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 PROBLEM STATEMENT

Need to create an Indian credit risk(default) model, using the data provided in the spreadsheet.

.

1.3 SCOPE

Data that is available includes information from the financial statement of the companies for the previous year. Also, information about the Net worth of the company in the following year is provided which can be used to drive the labeled field.

Explanation of data fields available in Data Dictionary, 'Company (FRA).csv'.

Understanding how data was collected in terms of time, frequency and methodology.

Model has been built and tested using the Juypter Notebook by cleansing the data, fixing the missing values using Median method, scaling the variables to understand the business and problem statement.

Data Spitted into 70:30 range for test and training.

2. EDA and Data Cleaning and Pre-processing

2.1 DATA DICTIONARY:

We have got the data from the Companies having columns of 51 and 4256 rows. This data contains with financial information and other parameters.

Obtained this data to understand the business and do analytical study on that.

Variable Name	Description
Networth Next Year	Net worth of the customer in next year
Total assets	Total assets of customer
Net worth	Net worth of the customer of present year
Total income	Total income of the customer
Change in stock	difference between value of current stock and the value of stock in last trading day
Total expenses	Total expense done by customer
Profit after tax	Profit after tax deduction
PBDITA	Profit before depreciation, income tax and amortization
PBT	Profit before tax deduction
Cash profit	Total Cash profit
PBDITA as % of total income	PBDITA / Total income
PBT as % of total income	PBT / Total income
PAT as % of total income	PAT / Total income
Cash profit as % of total income	Cash Profit / Total income
PAT as % of net worth	PAT / Net worth
Sales	Sales done by customer
Income from financial services	Income from financial services
Other income	Income from other sources
Total capital	Total capital of the customer
Reserves and funds	Total reserves and funds of the customer
Deposits (accepted by commercial	All blank values
Borrowings	Total amount borrowed by customer
Current liabilities & provisions	current liabilities of the customer
Deferred tax liability	Future income tax customer will pay because of the current transaction
Shareholders funds	Amount of equity in a company, which is belong to shareholder
Cumulative retained profits	Total cumulative profit retained by customer
Capital employed	Current asset minus current liabilities
TOL/TNW	Total liabilities of the customer divided by Total net worth
Total term liabilities / tangible net worth	Short + long term liabilities divided by tangible net worth
Contingent liabilities / Net worth (%)	Contingent liabilities / Net worth
Contingent liabilities	Liabilities because of uncertain events
Net fixed assets	purchase price of all fixed assets
Investments	Total invested amount
Current assets	Assets that are expected to be converted to cash within a year
Net working capital	Difference of current liabilities and current assets
Quick ratio (times)	Total cash divided by current liabilities
Current ratio (times)	Current assets divided by current liabilities
Debt to equity ratio (times)	Total liabilities divided by its shareholder equity
Cash to current liabilities (times)	Total liquid cash divided by current liabilities
Cash to average cost of sales per day	Total cash divided by average cost of the sales
Creditors turnover	Net credit purchase divided to average trade creditors
Debtors turnover	Net credit sales divided by average accounts receivable
Finished goods turnover	Annual sales divided by average inventory
WIP turnover	The cost of goods sold for a period divided by the average inventory for that period
Raw material turnover	Cost of goods sold is divided by the average inventory for the same period
Shares outstanding	Number of issued shares minus the number of share held in the company
Equity face value	cost of the equity at the time of issuing
EPS	Net income divided by total number of outstanding share
Adjusted EPS	Adjusted net earning divided by the weighted average number of common share outstanding on a diluted basis during the plan year
Total liabilities	Sum of all type of liabilities
PE on BSE	Company current stock price divided by its earning per share

2.2 READING AND UNDERSTANDING THE DATA

Below is the sample data set which is provided for an analysis.

	Num	Networth Next Year	Total assets	Net worth	Total income	Change in stock	Total expenses	Profit after tax	PBDITA	PBT	Cash profit	PBDITA as % of total income	PBT as % of total income	PAT as % of total income	Cash profit as % of total income	PAT as % of net worth	Sales	Income from fincial services
0	1	395.30	827.60	336.50	534.10	13.50	508.70	38.90	124.40	64.60	95.20	23.29	12.10	7.28	17.82	12.27	533.50	0.60
1	2	36.20	67.70	24.30	137.90	-3.70	131.00	3.20	5.50	1.00	3.80	3.99	0.73	2.32	2.76	0.00	135.50	nan
2	3	84.00	238.40	78.90	331.20	-18.10	309.20	3.90	25.80	10.50	9.40	7.79	3.17	1.18	2.84	5.07	330.60	0.60
3	4	2041.40	6883.50	1443.30	8448.50	212.20	8482.40	178.30	418.40	185.10	178.00	4.95	2.19	2.11	2.11	13.17	8444.20	2.00
4	5	41.80	90.90	47.00	388.60	3.40	392.70	-0.70	7.20	-0.60	3.90	1.85	-0.15	-0.18	1.00	-1.48	387.60	0.20

	Num	Networth Next Year	Total assets	Net worth	Total income	Change in stock	Total expenses	Profit after tax	PBDITA	PBT	Cash profit	PBDITA as % of total income	PBT as % of total income	PAT as % of total income	Cash profit as % of total income	PAT as % of net worth	Sales	Income from fincial services
4251	4252	0.20	0.40	0.20	nan	nan	nan	nan	nan	nan	nan	0.00	0.00	0.00	0.00	0.00	nan	nan
4252	4253	93.30	159.60	86.70	172.90	0.10	169.70	3.30	18.40	3.70	12.60	10.64	2.14	1.91	7.29	3.88	172.10	0.40
4253	4254	932.20	833.80	664.60	2314.70	32.10	2151.60	195.20	348.40	303.00	219.50	15.05	13.09	8.43	9.48	33.55	2309.40	3.00
4254	4255	64.60	95.00	48.50	110.50	4.60	113.50	1.60	9.70	2.60	6.70	8.78	2.35	1.45	6.06	4.08	110.00	0.10
4255	4256	0.00	384.60	111.30	345.80	11.30	341.70	15.40	57.60	20.70	34.80	16.66	5.99	4.45	10.06	16.04	338.30	1.10

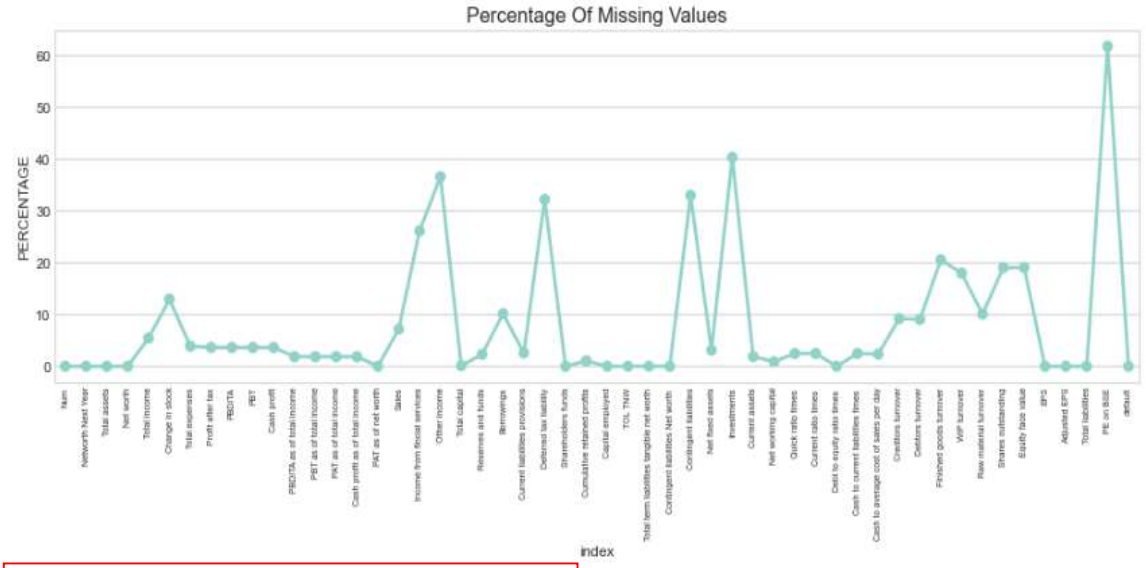
Observation: *Data looks good based on initial records seen in top 5 and bottom 5 and There are 51 variables and 4256 records.*

2.3 DATA INFO AND TYPES / MISSING VALUES:

Data Type **Missing Values**

```
<class 'pandas.core.frame.DataFrame'>
RangeInt64Index: 4256 entries, 0 to 4255
Data columns (total 51 columns):
 #   Column                                     Non-Null Count  Dtype
---  ------                                     --------------  -----
 0   Num                                        4256 non-null   int64
 1   Networth Next Year                         4256 non-null   float64
 2   Total assets                               4256 non-null   float64
 3   Net worth                                  4256 non-null   float64
 4   Total income                               4025 non-null   float64
 5   Change in stock                            3706 non-null   float64
 6   Total expenses                             4091 non-null   float64
 7   Profit after tax                           4102 non-null   float64
 8   PBDITA                                     4102 non-null   float64
 9   PBT                                        4102 non-null   float64
 10  Cash profit                                4102 non-null   float64
 11  PBDITA as of total income                  4177 non-null   float64
 12  PBT as of total income                     4177 non-null   float64
 13  PAT as of total income                     4177 non-null   float64
 14  Cash profit as of total income             4177 non-null   float64
 15  PAT as of net worth                        4256 non-null   float64
 16  Sales                                      3951 non-null   float64
 17  Income from fincial services               3145 non-null   float64
 18  Other income                               2700 non-null   float64
 19  Total capital                              4251 non-null   float64
 20  Reserves and funds                         4158 non-null   float64
 21  Borrowings                                 3825 non-null   float64
 22  Current liabilities provisions             4146 non-null   float64
 23  Deferred tax liability                     2887 non-null   float64
 24  Shareholders funds                         4256 non-null   float64
 25  Cumulative retained profits                4211 non-null   float64
 26  Capital employed                           4256 non-null   float64
 27  TOL TNW                                    4256 non-null   float64
 28  Total term liabilities tangible net worth  4256 non-null   float64
 29  Contingent liabilities Net worth           4256 non-null   float64
 30  Contingent liabilities                     2854 non-null   float64
 31  Net fixed assets                           4124 non-null   float64
 32  Investments                                2541 non-null   float64
 33  Current assets                             4176 non-null   float64
 34  Net working capital                        4219 non-null   float64
 35  Quick ratio times                          4151 non-null   float64
 36  Current ratio times                        4151 non-null   float64
 37  Debt to equity ratio times                 4256 non-null   float64
 38  Cash to current liabilities times          4151 non-null   float64
 39  Cash to average cost of sales per day      4156 non-null   float64
 40  Creditors turnover                         3865 non-null   float64
 41  Debtors turnover                           3871 non-null   float64
 42  Finished goods turnover                    3382 non-null   float64
 43  WIP turnover                               3492 non-null   float64
 44  Raw material turnover                      3828 non-null   float64
 45  Shares outstanding                         3446 non-null   float64
 46  Equity face value                          3446 non-null   float64
 47  EPS                                        4256 non-null   float64
 48  Adjusted EPS                               4256 non-null   float64
 49  Total liabilities                          4256 non-null   float64
 50  PE on BSE                                  1629 non-null   float64
dtypes: float64(50), int64(1)
memory usage: 1.7 MB
```

```
Num                                          0
Networth Next Year                         624
Total assets                               585
Net worth                                  595
Total income                               739
Change in stock                           1300
Total expenses                             683
Profit after tax                           866
PBDITA                                     738
PBT                                        858
Cash profit                                781
PBDITA as of total income                  425
PBT as of total income                     625
PAT as of total income                     689
Cash profit as of total income             505
PAT as of net worth                        427
Sales                                      805
Income from fincial services              1628
Other income                              1945
Total capital                              556
Reserves and funds                         741
Borrowings                                 963
Current liabilities provisions             691
Deferred tax liability                    1775
Shareholders funds                         588
Cumulative retained profits                744
Capital employed                           572
TOL TNW                                    414
Total term liabilities tangible net worth  486
Contingent liabilities Net worth           478
Contingent liabilities                    1795
Net fixed assets                           701
Investments                               2106
Current assets                             612
Net working capital                        843
Quick ratio times                          476
Current ratio times                        502
Debt to equity ratio times                 381
Cash to current liabilities times          644
Cash to average cost of sales per day      683
Creditors turnover                         833
Debtors turnover                           793
Finished goods turnover                   1273
WIP turnover                              1142
Raw material turnover                      724
Shares outstanding                        1286
Equity face value                         1343
EPS                                        638
Adjusted EPS                               694
Total liabilities                          585
PE on BSE                                 2864
dtype: int64
```




```
The total number of populaiton is (including blanks): 221312
The total number of missing values is: 17778
```

Observation:

- *Total of 51 columns with 4256 rows are present.*
- *Data looks good based on initial records seen in top 5 and bottom 5.*
- *The present variables are only numerics types in nature - i.e. float, int.*
- *There are missing values in most of the variables which PE on BSE has the highest.*
- *The total number of population are: 221312*
- *Total number of missing values are: 17778 and PE on BSE column has high missing values.*
- *Total 8% of missing values are noted on the total population.*
- *"Total income","Change in stock","Total expenses","Profit after tax","PBDITA","PBT","Cash profit","PBDITA as % of total income","PBT as % of total income","PAT as % of total income","Cash profit as % of total income","Sales","Income from fincial services","Other income","Total capital","Reserves and funds","Borrowings","Current liabilities & provisions","Deferred tax liability","Cumulative retained profits","Contingent liabilities","Net fixed assets","Investments","Current assets","Net working capital","Quick ratio (times)","Current ratio (times)","Cash to current liabilities (times)","Cash to average cost of sales per day","Creditors turnover","Debtors turnover","Finished goods turnover","WIP turnover","Raw material turnover","Shares outstanding","Equity face value","PE on BSE" columns are less line counts.*

2.4 GETTING THE SUMMARY STATISTICS

	count	mean	std	min	25%	50%	75%	max
Num	4256.00	2128.50	1228.75	1.00	1064.75	2128.50	3192.25	4256.00
Networth Next Year	4256.00	1544.74	15036.74	-74265.80	3.98	72.10	330.83	805773.40
Total assets	4256.00	3573.82	30074.44	0.10	91.30	315.50	1120.80	1178509.20
Net worth	4256.00	1351.95	12981.31	0.00	31.48	104.80	389.85	613151.60
Total income	4025.00	4888.19	53916.95	0.00	107.10	455.10	1485.00	2442628.20
Change in stock	3706.00	43.70	406.92	-3029.40	-1.80	1.80	15.40	14185.50
Total expenses	4091.00	4356.30	51398.09	-0.10	96.80	426.80	1395.70	2388035.30
Profit after tax	4102.00	295.05	3079.90	-3908.30	0.50	9.00	53.50	119439.10
PBDITA	4102.00	805.94	9846.23	-440.70	8.93	36.50	156.70	2089678.50
PBIT	4102.00	410.28	4217.42	-3894.80	0.80	12.60	74.18	145282.60
Cash profit	4102.00	408.27	4143.93	-2245.70	2.90	19.40	96.25	178911.80
PBDITA as of total income	4177.00	3.18	172.28	-6400.00	4.97	9.68	16.47	100.00
PBT as of total income	4177.00	-78.20	419.91	-21340.00	0.56	3.34	6.94	100.00
PAT as of total income	4177.00	-20.03	423.58	-21340.00	0.35	2.37	6.42	150.00
Cash profit as of total income	4177.00	-9.02	209.98	-15020.00	2.00	5.88	10.73	100.00
PAT as of net worth	4256.00	10.17	81.53	-748.72	0.00	5.04	20.20	2466.67
Sales	3951.00	4845.88	53880.90	0.10	113.35	468.60	1481.20	2384984.40
Income from fincial services	3145.00	81.38	1042.78	0.00	0.50	1.90	9.80	51938.20
Other income	2700.00	66.95	1178.42	0.00	0.40	1.50	6.20	42898.70
Total capital	4251.00	224.56	1884.56	0.10	13.20	42.60	103.15	75273.20
Reserves and funds	4158.00	1210.56	12818.23	-6525.90	5.30	55.15	282.52	625137.80
Borrowings	3525.00	1178.25	9981.25	0.10	24.40	89.80	356.30	278257.30
Current liabilities provisions	4148.00	980.83	9140.54	0.10	17.50	70.30	285.92	352240.30
Deferred tax liability	2887.00	234.50	2106.25	0.10	3.20	13.50	51.30	72798.60
Shareholders funds	4256.00	1306.49	13010.69	0.00	32.30	107.60	408.90	613151.60
Cumulative retained profits	4211.00	937.18	9853.10	-6534.30	1.10	37.40	206.20	390133.60
Capital employed	4256.00	2433.82	20496.40	0.00	61.30	221.20	780.30	891408.90
TOL TNW	4256.00	4.03	20.88	-350.48	0.80	1.42	2.83	473.00
Total term liabilities tangible net worth	4256.00	1.85	15.88	-325.60	0.05	0.34	1.00	458.00
Contingent liabilities Net worth	4256.00	55.71	369.17	0.00	0.00	5.58	31.01	14704.27
Contingent liabilities	2954.00	948.55	12096.74	0.10	8.00	37.85	156.33	559508.80
Net fixed assets	4124.00	1209.49	12502.40	0.00	26.20	93.55	352.83	636804.60
Investments	2541.00	721.87	8793.88	0.00	1.00	8.20	83.80	199978.80
Current assets	4178.00	1350.38	10155.57	0.10	38.60	148.35	515.00	354815.20
Net working capital	4219.00	182.87	3182.03	-83839.00	-1.10	16.70	86.50	85782.80
Quick ratio times	4151.00	1.50	9.33	0.00	0.41	0.87	1.03	341.00
Current ratio times	4151.00	2.28	12.48	0.00	0.93	1.23	1.72	505.00
Debt to equity ratio times	4256.00	2.87	15.60	0.00	0.22	0.79	1.75	466.00
Cash to current liabilities times	4151.00	0.53	4.80	0.00	0.02	0.07	0.19	185.00
Cash to average cost of sales per day	4156.00	145.16	2521.59	0.00	2.88	6.04	21.97	128040.76
Creditors turnover	3885.00	16.81	75.67	0.00	3.72	6.17	11.69	2401.00
Debtors turnover	3871.00	17.93	90.16	0.00	3.81	6.47	11.85	3135.20
Finished goods turnover	3382.00	84.37	982.64	-0.09	8.19	17.32	40.01	17947.80
WIP turnover	3492.00	28.68	169.65	-0.16	5.10	9.86	20.24	5851.40
Raw material turnover	3528.00	17.73	343.13	-2.00	3.02	6.41	11.82	21092.00
Shares outstanding	3446.00	23764809.56	170975941.33	-214746364.7.00	1308382.50	4750000.00	10998020.00	4130400646.00
Equity face value	3446.00	-1094.83	34101.38	-999999.90	10.00	10.00	10.00	10000.00
EPS	4256.00	-198.22	13081.95	-843181.82	0.00	1.49	10.00	34922.53
Adjusted EPS	4256.00	-197.53	13081.93	-843181.82	0.00	1.24	7.82	34922.53
Total liabilities	4256.00	3573.82	30074.44	0.10	91.30	315.50	1120.80	1178509.20
PE on BSE	1629.00	66.48	1304.46	-1116.64	2.97	8.69	17.00	51002.74

Observation: -.

- Based on summary descriptive, the data looks good.
  -. In given data not a single column is normally distributed, and data is skewed.
  -. The range of minimum value range and maximum value range is close at some columns and at higher level at some places
  -. On first look we can say that count of all columns is not same. Some of the missing values are noted.
  -. We see for most of the variable, mean/medium is nearly equal.
  -. Std Deviation is high for Shares outstanding variable and also all the parameters.

.

2.5 DUPLICATE CHECKS

```
Number of duplicate rows = 0
```

Num	Networth Next Year	Total assets	Net worth	Total income	Change in stock	Total expenses	Profit after tax	PBDITA	PBT	Cash profit	PBDITA as of total income	PBT as of total income	PAT as of total income	Cash profit as of total income	PAT as of net worth	Sales	Income from fincial services	Other income

Observation: - *There are no duplicate rows are present in the data.*

2.6 FIXING MESSY COLUMN NAMES (CONTAINING %, &, (,), / SPECIAL CHARACTERS) FOR EASE OF USE

ıcome from fincial rvices	Other income	Total capital	Reserves and funds	Borrowings	Current liabilities provisions	Deferred tax liability	Shareholders funds	Cumulative retained profits	Capital employed	TOL TNW	Total term liabilities tangible net worth	Contingent liabilities Net worth	Contingent liabilities	Net fixed assets	Inves
0.60	nan	87.60	249.00	390.70	43.90	56.40	336.50	248.90	727.20	1.28	0.99	186.21	626.60	461.10	
nan	0.20	11.90	4.30	16.60	23.70	3.10	24.30	-8.20	40.90	1.53	0.21	47.74	11.60	18.50	
0.60	nan	25.00	56.70	44.70	102.20	9.80	78.90	53.10	123.60	1.70	0.33	30.42	24.00	56.80	
2.00	nan	100.00	1343.30	2789.30	2650.80	0.10	1443.30	593.30	4232.60	3.69	0.22	10.79	155.70	8.60	
0.20	0.80	10.70	35.80	25.50	14.10	4.30	47.00	35.80	72.50	0.81	0.44	0.00	nan	36.30	

Observation:

- Replaced the special characters from the column name to " " (blank space).

2.7 CREATING A BINARY TARGET VARIABLE USING 'NETWORTH NEXT YEAR).

	default	Networth Next Year
0	0	395.30
1	0	36.20
2	0	84.00
3	0	2041.40
4	0	41.80
5	0	291.50
6	0	93.30
7	0	985.10
8	0	188.60
9	0	229.60
10	0	292.80
11	0	12300.00
12	0	3232.30
13	1	-5.30
14	0	26.70
15	0	560.30
16	0	283.20
17	0	28.20
18	0	1.60
19	0	3.90

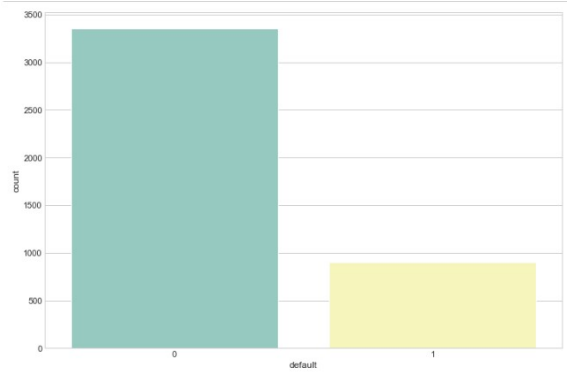
Observation:

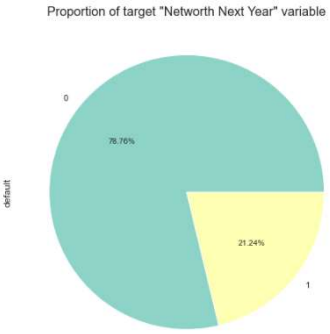
-. We need to create a default variable that take the value of 1 when net worth next year is negative & 0 when net worth next year is positive.

- Top 20 records are validated in the above tables.

2.8 'DEFAULT' VARIABLE LOOKS LIKE:

```
0    3352
1     904
Name: default, dtype: int64
```





Observation:

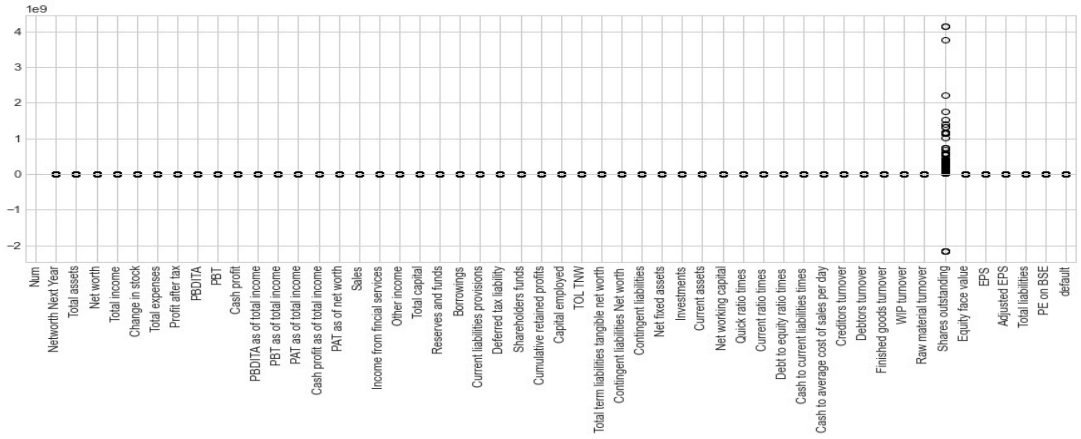
- *Creating a binary Target Variable Using "Network_Next_Year" as per the business case.*

- *Checked the proportion of default for "0" and "1" has 78.8% and 21.2% respectively.*

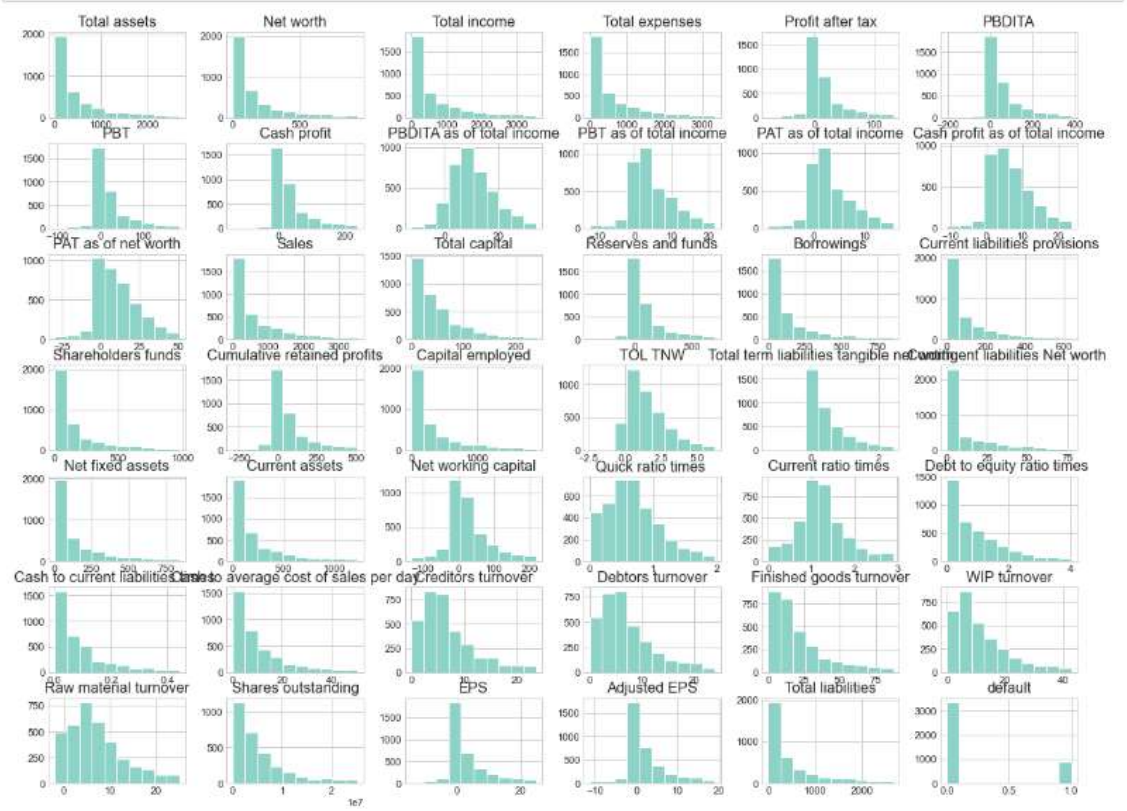
- *Checked the proportion of default for "0" has 3352 and "1" has 904 counts from the total population.*

2.9 UNIVARIATE ANALYSIS.

- ***Outlier Check:*** As we noticed that there are outlier for the variables Shares Outstanding and most them are not much outliers are found.



We did the Univariate analysis of all columns as per there numeric be seen as below

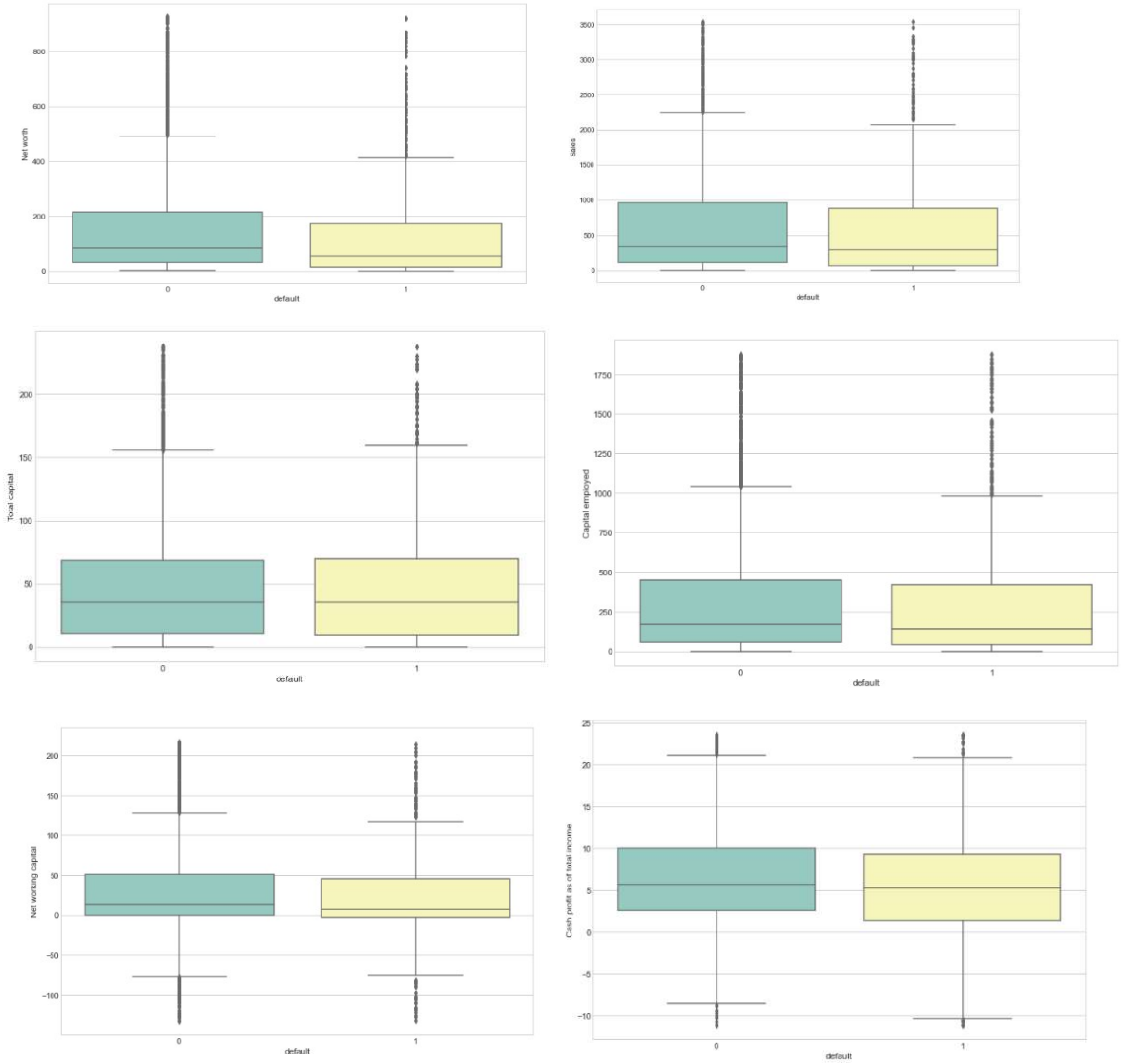



```
Shareholders funds                        1.82
Net worth                                 1.81
Borrowings                                1.80
Capital employed                          1.77
Net fixed assets                          1.76
Current liabilities provisions            1.75
Total liabilities                         1.74
Total assets                              1.74
PBDITA                                    1.72
Current assets                            1.71
Total income                              1.64
Total expenses                            1.63
Contingent liabilities Net worth          1.62
Cash to average cost of sales per day     1.62
Sales                                     1.61
PBT                                       1.61
Cash to current liabilities times         1.61
Cash profit                               1.60
Reserves and funds                        1.58
Shares outstanding                        1.56
Profit after tax                          1.51
Total capital                             1.49
Finished goods turnover                   1.44
Cumulative retained profits               1.41
default                                   1.41
EPS                                       1.34
Total term liabilities tangible net worth 1.30
WIP turnover                              1.27
Debt to equity ratio times                1.19
Adjusted EPS                              1.15
Creditors turnover                        1.11
TOL TNW                                   1.07
Debtors turnover                          1.01
Raw material turnover                     0.96
Net working capital                       0.83
Quick ratio times                         0.69
PAT as of total income                    0.61
PBT as of total income                    0.54
PBDITA as of total income                 0.52
PAT as of net worth                       0.50
Current ratio times                       0.47
Cash profit as of total income            0.46
dtype: float64
```

Insight:

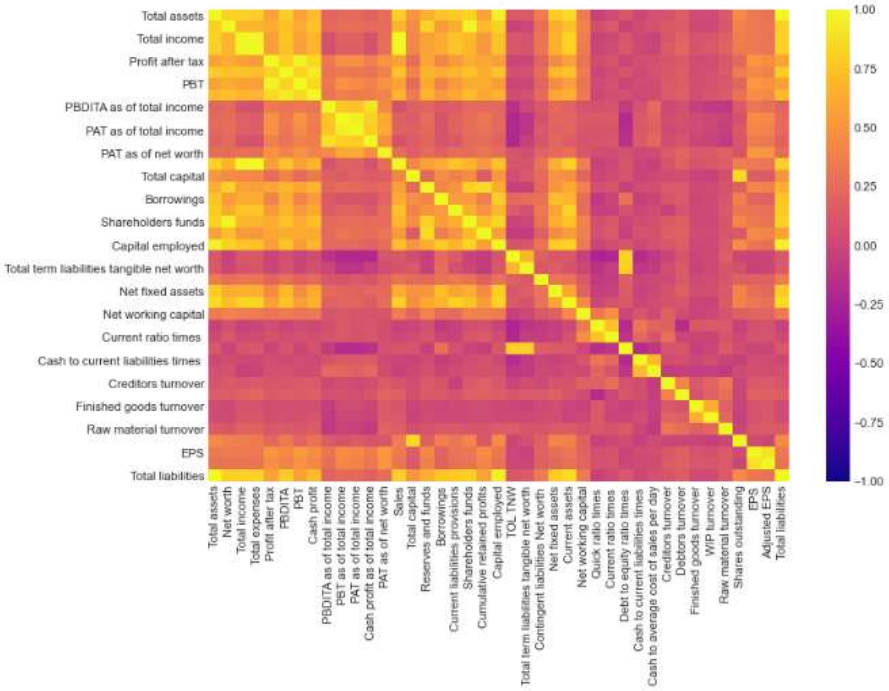
In given data most of them are skewed to right side.

2.10 UNDERSTANDING ON THE TARGET VARIABLE I.E. DEFALUT WITH RELATED DEPENDENT VARIABLES:



- *Bivariate analysis variables "Net worth", "Sales", "Total capital", "Capital employed", "Net working capital","Cash profit as of total income", "Profit after tax" with proper interpretation.*
- *This will be the analysis done between any two variables which looks quite significant or just to check what is the correlation between them.*
- *The above variables having out outliers and most of them 3Q's are not having much differences.*

2.11 MULTIVARIATE ANALYSIS BEFROE FIXING THE OUTLIER AND MISSING VALUES:

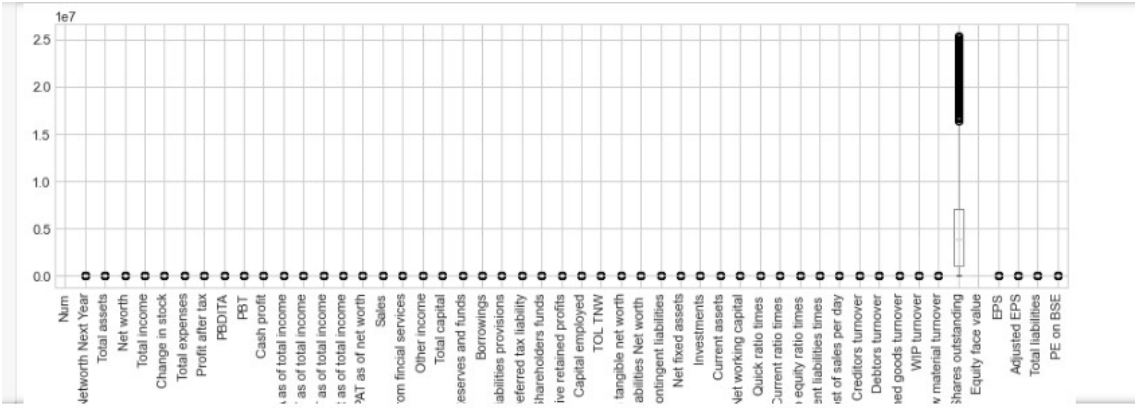


		correlation
Sales	Total income	0.99
	Total expenses	0.99
Total expenses	Total income	0.99
Shareholders funds	Net worth	0.99
PBT	Profit after tax	0.96
Total liabilities	Capital employed	0.96
PBT as of total income	PAT as of total income	0.95
PBDITA	Cash profit	0.93
PBDITA as of total income	Cash profit as of total income	0.90
EPS	Adjusted EPS	0.90
Reserves and funds	Cumulative retained profits	0.87

Insight: *Top five variables are highly correlated which has more than 95%*

- *"Sales" with "Total Income", "Total expenses"*
- *"Total expenses" highly correlated with "Total Income"*
- *"Shareholders funds" highly correlated with "Net Worth"*
- *"PBT" highly correlated with "Profit after tax"*
- *"Total liabilities" highly correlated with "Captial employed".*

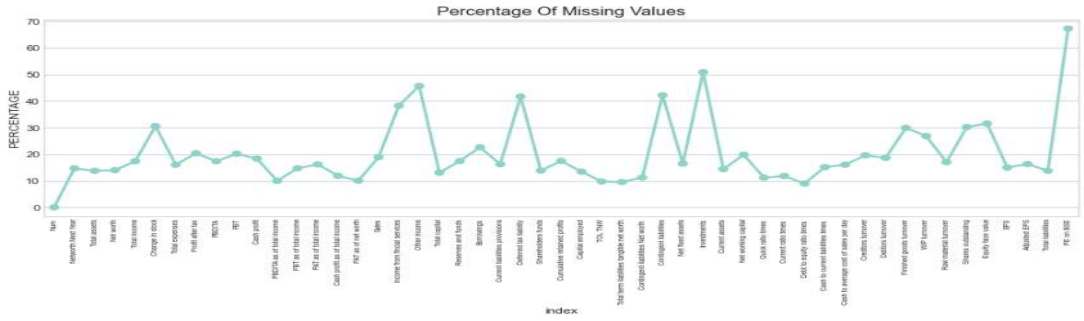
2.12 OUTLIERS CHECK TREATMENT:



Insight:

Usually, the outlier is treated with the help of the IQR, the IQR will help us find the difference between the highest range and the lowest range near to mean of the attribute. If the outliers are higher and do not fall in the IQR, they will not be treated and will only treat the values which fall in this range. Still we can see outliers for "Shares Outstanding" variable.

2.13 MISSING VALUES CHECK AFTER TREATING OUTLIERS:




```
The total number of populaiton is (including blanks): 217056
The total number of missing values is: 43724
```

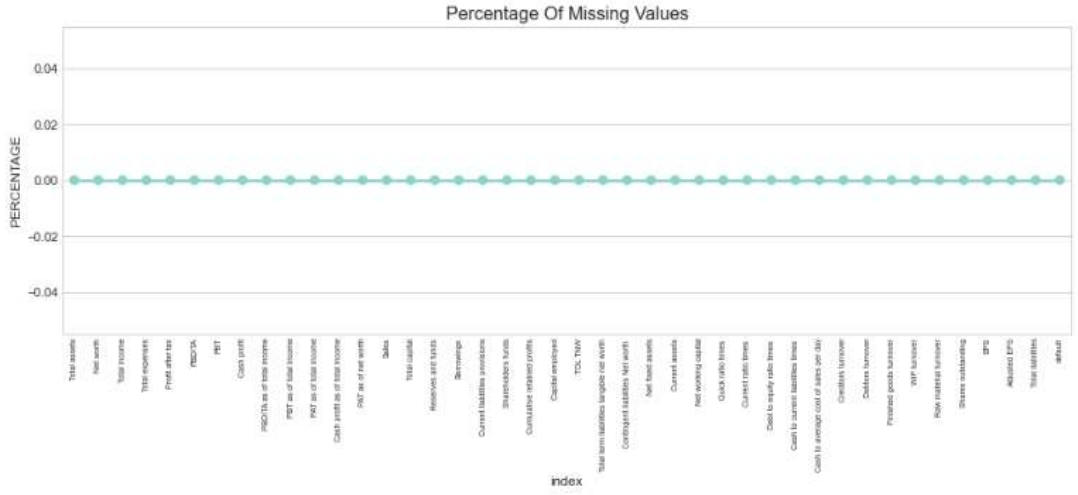
PE on BSE	0.67
Investments	0.51
Other income	0.46
Contingent liabilities	0.42
Deferred tax liability	0.42
Income from fincial services	0.38
Equity face value	0.32
Change in stock	0.31
Shares outstanding	0.30
Finished goods turnover	0.30
WIP turnover	0.27
Borrowings	0.23

Insight:

- *After fixing the Outliers, still the missing values are increased some of the columns has more than 30%.*

We are planned to drop these columns 'PE on BSE', 'Investments', 'Other income', 'Contingent liabilities', 'Deferred tax liability', 'Income from finical services', 'Equity face value','Change in stock' as these are more than 30% missing values. After that it has been noticed with 4256 rows and 42 columns in the data.

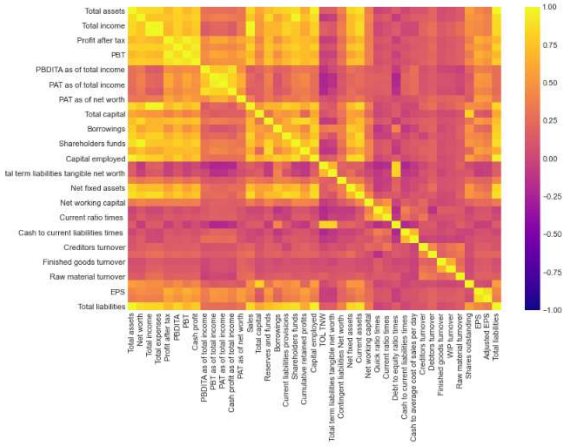
2.14 IMPUTING THE REMAINING MISSING VALUES



Insight:

Using "KNNImputer" (mean value) filled the missing values in the data. Now we can see there is no missing values in the data.

2.15 MULTIVARIATE ANALYSIS ON TARGET VS CATEGORICAL VARIABLES:



Insight:

After fixitng the missing values, still we can see Top five variables are highly correlated which has more than 95%."Sales" with "Total Income", "Total expenses"; "Total expenses" highly correlated with "Total Income"; "Shareholders funds" highly correlated with "Net Worth"; "PBT" highly correlated with "Profit after tax"; "Total liabilities" highly correlated with "Captial employed"

3. Model building & Model validation

3.1 LOGISTIC REGRESSION:

```
array([ 9,  1, 19,  1,  2,  1,  3,  1,  8, 24,  7,  1, 27,  1, 17,  1, 22,
       26,  1,  5,  1, 14, 21,  1, 15,  1, 25, 20,  1, 23, 10, 11, 12, 13,
        1,  6,  4, 16, 18,  1,  1])
```

	Feature	Rank
1	Net worth	1
3	Total expenses	1
5	PBDITA	1
7	Cash profit	1
11	Cash profit as of total income	1
13	Sales	1
15	Reserves and funds	1
18	Shareholders funds	1
20	Capital employed	1
23	Contingent liabilities Net worth	1
28	Current ratio times	1
34	Finished goods turnover	1
36	Raw material turnover	1
39	Adjusted EPS	1
40	Total liabilities	1

Insights:
Above variables are top 15 which are ranked as "1".

3.2 VALIDATING THE MODEL ON TRAIN AND TEST SET

Tain Data Set:

Acutal	***2349***	***3***
	620	***7***
	Predict	

	precision	recall	f1-score	support
0.0	0.79	1.00	0.88	2352
1.0	0.70	0.01	0.02	627
accuracy			0.79	2979
macro avg	0.75	0.50	0.45	2979
weighted avg	0.77	0.79	0.70	2979

Test Data Set:

Acutal	***995***	***5***
	275	***2***
	Predict	

	precision	recall	f1-score	support
0.0	0.78	0.99	0.88	1000
1.0	0.29	0.01	0.01	277
accuracy			0.78	1277
macro avg	0.53	0.50	0.45	1277
weighted avg	0.68	0.78	0.69	1277

Insights:

- *Here, we can see that right predictions are around 2349 (there are true positives), true negatives are only 7 observations in the test data. Were in train data 995 has true positives and only 2 are true negatives.*
- *False negative is also higher number 620 in test and 275 in training.*
- *Therefore, we can still have a 78% of accuracy, this might be missing values.*

.

3.3 MODELS WILL BE USING THE BACKWARD ELIMINATION METHOD:

Details of 0 and 1:

0	***2245***
1	***606***

0.21255699754472115

Insights:
When we are tyring out backward elimination method, we haven noticed that "0" has 2245 and "1" has 606 with 21% of defaults.

Model 1:

Logit Regression Results

Dep. Variable:	default	No. Observations:	2851
Model:	Logit	Df Residuals:	2841
Method:	MLE	Df Model:	9
Date:	Sun, 07 Aug 2022	Pseudo R-squ.:	0.01948
Time:	17:29:31	Log-Likelihood:	-1446.2
converged:	True	LL-Null:	-1474.9
Covariance Type:	nonrobust	LLR p-value:	4.121e-09

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-1.3437	0.047	-28.492	0.000	-1.436	-1.251
Q("Net worth")	-0.2298	0.272	-0.844	0.399	-0.763	0.304
Q("Total expenses")	0.7090	0.344	2.059	0.040	0.034	1.384
Q("PBDITA")	-0.0024	0.147	-0.016	0.987	-0.290	0.286
Q("Cash profit")	0.0785	0.140	0.561	0.575	-0.196	0.353
Q("Cash profit as of total income")	-0.2849	0.062	-4.570	0.000	-0.407	-0.163
Q("Sales")	-0.7356	0.362	-2.033	0.042	-1.445	-0.027
Q("Reserves and funds")	-0.2772	0.090	-3.070	0.002	-0.454	-0.100
Q("Shareholders funds")	0.3433	0.263	1.306	0.192	-0.172	0.858
Q("Capital employed")	0.0874	0.110	0.797	0.425	-0.128	0.302

```
Optimization terminated successfully.
         Current function value: 0.507247
         Iterations 6
```

Insight: *Top 9 Feature variable and ranked 1, current function value has 51% with Iterations of 6 and we have seen the highest P(Z) value is there for PBDITA. Hence, we have removed this value in the next model.*

Model 2:

Logit Regression Results

Dep. Variable:	default	No. Observations:	2851
Model:	Logit	Df Residuals:	2842
Method:	MLE	Df Model:	8
Date:	Sun, 07 Aug 2022	Pseudo R-squ.:	0.01948
Time:	17:39:09	Log-Likelihood:	-1446.2
converged:	True	LL-Null:	-1474.9
Covariance Type:	nonrobust	LLR p-value:	1.463e-09

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-1.3437	0.047	-28.496	0.000	-1.436	-1.251
Q("Net worth")	-0.2294	0.271	-0.846	0.398	-0.761	0.302
Q("Total expenses")	0.7097	0.342	2.073	0.038	0.039	1.381
Q("Cash profit")	0.0770	0.105	0.736	0.462	-0.128	0.282
Q("Cash profit as of total income")	-0.2850	0.062	-4.596	0.000	-0.407	-0.163
Q("Sales")	-0.7367	0.355	-2.076	0.038	-1.432	-0.041
Q("Reserves and funds")	-0.2773	0.090	-3.090	0.002	-0.453	-0.101
Q("Shareholders funds")	0.3431	0.263	1.307	0.191	-0.171	0.858
Q("Capital employed")	0.0871	0.108	0.805	0.421	-0.125	0.299

Insight:

Here, we have seen the highest P(Z) value is there for "Cash profit". Hence, we have removed this value in the next model.

Model 3:

Logit Regression Results

Dep. Variable:	default	No. Observations:	2851
Model:	Logit	Df Residuals:	2843
Method:	MLE	Df Model:	7
Date:	Sun, 07 Aug 2022	Pseudo R-squ.:	0.01930
Time:	17:49:33	Log-Likelihood:	-1446.4
converged:	True	LL-Null:	-1474.9
Covariance Type:	nonrobust	LLR p-value:	6.205e-10

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-1.3431	0.047	-28.502	0.000	-1.435	-1.251
Q("Net worth")	-0.2241	0.271	-0.826	0.409	-0.756	0.308
Q("Total expenses")	0.6765	0.336	2.014	0.044	0.018	1.335
Q("Cash profit as of total income")	-0.2623	0.054	-4.896	0.000	-0.367	-0.157
Q("Sales")	-0.6727	0.340	-1.979	0.048	-1.339	-0.006
Q("Reserves and funds")	-0.2613	0.087	-3.002	0.003	-0.432	-0.091
Q("Shareholders funds")	0.3475	0.263	1.324	0.186	-0.167	0.862
Q("Capital employed")	0.0942	0.108	0.875	0.382	-0.117	0.305

Insight:

Here, we have seen the highest P(Z) value is there for "Net worth" & "Capital employed". Hence, we have removed this value in the next model.

Model 4:

Logit Regression Results

Dep. Variable:	default	No. Observations:	2851
Model:	Logit	Df Residuals:	2845
Method:	MLE	Df Model:	5
Date:	Sun, 07 Aug 2022	Pseudo R-squ.:	0.01885
Time:	17:51:44	Log-Likelihood:	-1447.1
converged:	True	LL-Null:	-1474.9
Covariance Type:	nonrobust	LLR p-value:	9.829e-11

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-1.3428	0.047	-28.508	0.000	-1.435	-1.251
Q("Total expenses")	0.7024	0.335	2.098	0.036	0.046	1.359
Q("Cash profit as of total income")	-0.2620	0.053	-4.901	0.000	-0.367	-0.157
Q("Sales")	-0.6688	0.342	-1.955	0.051	-1.339	0.002
Q("Reserves and funds")	-0.2941	0.081	-3.627	0.000	-0.453	-0.135
Q("Shareholders funds")	0.2139	0.093	2.296	0.022	0.031	0.396

Insight:

After the removal of the attributes with higher values, we have reached to these values where the values which are more significant and give better results.

3.4 CHECKING FOR THE PROBABILITY PREDICTIONS FOR TRAIN AND TEST MODEL

Train Data.

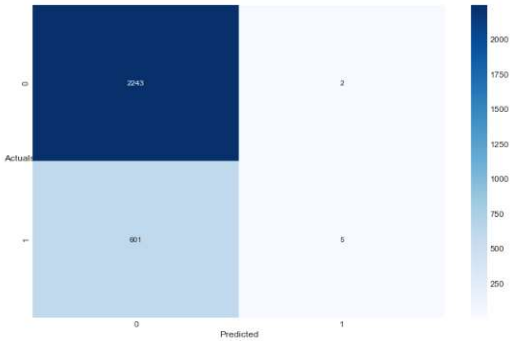
```
2893   0.17
4243   0.26
3345   0.25
332    0.23
3265   0.23
       ...
1923   0.22
124    0.22
3740   0.21
3985   0.10
3159   0.20
Length: 2851, dtype: float64
```

Test Data.

```
2302   0.21
2296   0.28
4075   0.16
3416   0.21
3609   0.23
       ...
1128   0.18
3516   0.18
3002   0.19
1760   0.21
4052   0.25
Length: 1405, dtype: float64
```

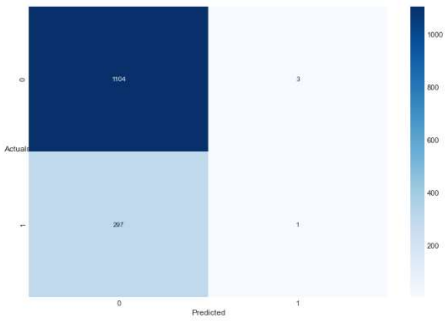
3.5 CONFUSION_MATRIX TRAIN AND TEST DATA

Train Data



	precision	recall	f1-score	support
0.0	0.789	0.999	0.882	2245
1.0	0.714	0.008	0.016	606
accuracy			0.788	2851
macro avg	0.751	0.504	0.449	2851
weighted avg	0.773	0.788	0.698	2851

Test Data



	precision	recall	f1-score	support
0.0	0.788	0.997	0.880	1107
1.0	0.250	0.003	0.007	298
accuracy			0.786	1405
macro avg	0.519	0.500	0.444	1405
weighted avg	0.674	0.786	0.695	1405

Insights:

- *Here, we can see that right predictions are around 2243 (there are true positives), true negatives are only 2 observations in the test data. Were in train data 1104 has true positives and only 2 are true negatives.*

- *False negative is also higher number 601 in test and 207 in training.*

- *Therefore, we can still have a approximately 78% of accuracy, this might be missing values.*

3.6 RANDOM FOREST REGRESSOR

```
MSE train data: 0.036, MSE test data: 0.240
R2 train data: 0.000, R2 test data: 0.000
MAE train data: 0.141, MAE test data: 0.383
```

Insights:
MAE, MSE and RMSE are not better than LOGISTIC REGRESSION.

3.7 PREDICTING THE FEATURE_IMPORTANCES TRAINING DATA:

	Imp
Finished goods turnover	0.04
Shares outstanding	0.04
WIP turnover	0.03
PBT	0.03
Creditors turnover	0.03
Debtors turnover	0.03
Raw material turnover	0.03
Cumulative retained profits	0.03
EPS	0.03
Net worth	0.03
Reserves and funds	0.03
Cash profit	0.03
Cash to average cost of sales per day	0.03
Net working capital	0.03
Current ratio times	0.03
TOL TNW	0.03
Shareholders funds	0.03
Quick ratio times	0.03
Total capital	0.02
Current liabilities provisions	0.02
Profit after tax	0.02
Adjusted EPS	0.02
Contingent liabilities Net worth	0.02
Cash to current liabilities times	0.02
PBT as of total income	0.02
Cash profit as of total income	0.02
Borrowings	0.02
PAT as of net worth	0.02
PBDITA as of total income	0.02
Net fixed assets	0.02
Total term liabilities tangible net worth	0.02
PAT as of total income	0.02
Debt to equity ratio times	0.02
Total expenses	0.02
PBDITA	0.02
Current assets	0.02
Sales	0.01
Total income	0.01
Capital employed	0.01
Total assets	0.01
Total liabilities	0.01

Insights:
The feature importance are suggested "Finished Goods Turnover" and followed by others

4. Final Interpretation

- *Initially the Data was looking good with record counts of 51 variables and 4256 records.*
- *There were lots of missing values in the data.*
- Shares outstanding variable having high in Std. Deviation and other parameters i.e. 3Q's and mean etc.
- In given data not a single column is normally distributed, and data is skewed.
- *There are proportion of default for "0" has 3352 and "1" has 904 counts from the total population. After defending variable that take the value of 1 when net worth next year is negative & 0 when net worth next year is positive.*
- *21% of the population has showing as defaults.*
- *There were outliers in the column "*Shares outstanding".
- *Top five variables are highly correlated which has more than 95%.*

5. Recommendations

- ***As we noticed the data given to us some of the details are not completely populated manily "PE on BSE" which has high missing values. And also there are many more missing values. Hence, if the data provided with missing vlaues it will help us to predict with more accurtly.***
- ***Logistic Regression given 78% of accuracy which is less than industrial standards which is will be 95% will be the best.***
- ***Random Forest is not giving good results with provided data sets.***
- ***Finally, we are able to achieve a descent recall value without overfitting. Considering the opportunities such as outliers, missing values and correlated features this is a fairly good model.***
- ***It can be improved if we get better quality data where the features explaining the default are not missing to this extent.***
- ***Of course we can try other techniques which are not sensitive towards missing values and outliers.***

Thank you