Finance & Risk Analytics

PROJECT REPORT

By

Raghvendra Singh

Email: raghavsingh0027@gmail.com

Phone: +91-8130670022

Index

Contents

List of Figures

List of Tables

Data dictionary (Part A)

COLUMN NAME	DICTIONARY
Networth Next Year	Net worth of the customer in the next year
Total assets	Total assets of customer
Net worth	Net worth of the customer of the present year
Total income	Total income of the customer
Change in stock	Difference between the current value of the stock and the value of stock in the last trading day
Total expenses	Total expenses done by the 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 the 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
Borrowings	Total amount borrowed by the 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 belongs to shareholders
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	The 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 between the 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 the average cost of the sales
Creditors turnover	Net credit purchase divided by 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 shares held in the company
Equity face value	cost of the equity at the time of issuing
EPS	Net income divided by the total number of outstanding share
Adjusted EPS	Adjusted net earnings divided by the weighted average number of common shares outstanding on a diluted basis during the plan year
Total liabilities	Sum of all types of liabilities
PE on BSE	Company's current stock price divided by its earnings per share

Data Dictionary (Part B)

Date	Date of stock price movement
ITC Limited	Stock price of ITC Limited
Bharti Airtel	Stock price of Bharti Airtel
Tata Motors	Stock price of Tata Motors
DLF Limited	Stock price of DLF Limited
Yes Bank	Stock price of Yes Bank

PROBLEM STATEMENT 1

Part A

Context

In the realm of modern finance, businesses encounter the perpetual challenge of managing debt obligations effectively to maintain a favorable credit standing and foster sustainable growth. Investors keenly scrutinize companies capable of navigating financial complexities while ensuring stability and profitability. A pivotal instrument in this evaluation process is the balance sheet, which provides a comprehensive overview of a company's assets, liabilities, and shareholder equity, offering insights into its financial health and operational efficiency. In this context, leveraging available financial data, particularly from preceding fiscal periods, becomes imperative for informed decision-making and strategic planning.

Objective

A group of venture capitalists want to develop a Financial Health Assessment Tool. With the help of the tool, it endeavors to empower businesses and investors with a robust mechanism for evaluating the financial well-being and creditworthiness of companies. By harnessing machine learning techniques, they aim to analyze historical financial statements and extract pertinent insights to facilitate informed decision-making via the tool. Specifically, they foresee facilitating the following with the help of the tool:

- Debt Management Analysis: Identify patterns and trends in debt management practices to assess the ability of businesses to fulfill financial obligations promptly and efficiently, and identify potential cases of default.
- Credit Risk Evaluation: Evaluate credit risk exposure by analyzing liquidity ratios, debt-to-equity ratios, and other key financial indicators to ascertain the likelihood of default and inform investment decisions.

They have hired you as a data scientist and provided you with the financial metrics of different companies. The task is to analyze the data provided and develop a predictive model leveraging machine learning techniques to identify whether a given company will be tagged as a defaulter in terms of net worth next year. The predictive model will help the organization anticipate potential challenges with the financial performance of the companies and enable proactive risk mitigation strategies.

PART A: Define the problem and perform Exploratory Data Analysis

- Problem definition - Check shape, Data types, and statistical summary - Univariate analysis - Multivariate analysis - Use appropriate visualizations to identify the patterns and insights - Key meaningful observations on individual variables and the relationship between variables

	Num	Networth Next Year	Total assets	Net worth	Total income	Change in stock	Total expenses	Profit after tax	PBDITA	PBT	...	Debtors turnover	Finished goods turnover	WIP turnover	Raw material turnover	Shares outstanding	Equity face value	EPS
4251	4252	0.2	0.4	0.2	NaN	NaN	NaN	NaN	NaN	NaN	...	0.00	NaN	NaN	0.00	NaN	NaN	0.00
4252	4253	93.3	159.6	86.7	172.9	0.1	169.7	3.3	18.4	3.7	...	1.80	11.00	8.28	9.88	8162700.0	10.0	0.42
4253	4254	932.2	833.8	664.6	2314.7	32.1	2151.6	195.2	348.4	303.0	...	6.08	59.28	31.14	9.87	7479762.0	10.0	26.58
4254	4255	64.6	95.0	48.5	110.5	4.6	113.5	1.6	9.7	2.6	...	3.71	78.99	11.51	14.95	NaN	NaN	0.00
4255	4256	0.0	384.6	111.3	345.8	11.3	341.7	15.4	57.6	20.7	...	4.71	53.37	8.33	3.74	960000.0	10.0	15.63

5 rows × 51 columns

	PAT as % of net worth	Sales	Income from fincial services	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 (%)
0	12.27	533.5	0.6	NaN	87.6	249.0	390.7	43.9	56.4	336.5	248.9	727.2	1.28	0.99	186.21
1	0.00	135.5	NaN	0.2	11.9	4.3	16.6	23.7	3.1	24.3	-8.2	40.9	1.53	0.21	47.74
2	5.07	330.6	0.6	NaN	25.0	56.7	44.7	102.2	9.8	78.9	53.1	123.6	1.70	0.33	30.42
3	13.17	8444.2	2.0	NaN	100.0	1343.3	2789.3	2650.8	0.1	1443.3	593.3	4232.6	3.69	0.22	10.79
4	-1.48	387.6	0.2	0.8	10.7	35.8	25.5	14.1	4.3	47.0	35.8	72.5	0.81	0.44	0.00

	Contingent liabilities	Net fixed assets	Investments	Current assets	Net working capital	Quick ratio (times)	Current ratio (times)	Debt to equity 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
0	626.6	461.1	18.1	257.6	163.1	0.99	2.52	1.16	0.06	5.41	11.60	5.65	3.99	3.37	14.87
1	11.6	18.5	0.2	39.0	3.9	0.67	1.11	0.68	0.02	1.62	NaN	NaN	NaN	NaN	NaN
2	24.0	56.8	0.2	158.3	38.3	1.11	1.31	0.57	0.19	26.42	2.24	2.51	17.67	8.76	8.35
3	155.7	8.6	NaN	6576.4	1455.1	0.99	1.28	1.93	0.07	15.93	3.48	1.91	18.14	18.62	11.11
4	NaN	36.3	NaN	39.8	20.8	0.35	2.09	0.54	0.05	0.85	21.67	68.00	45.87	28.67	19.93

	Shares outstanding	Equity face value	EPS	Adjusted EPS	Total liabilities
0	8760056.0	10.0	4.44	4.44	827.6
1	NaN	NaN	0.00	0.00	67.7
2	NaN	NaN	0.00	0.00	238.4
3	10000000.0	10.0	17.60	17.60	6883.5
4	107315.0	100.0	-6.52	-6.52	90.9

Figure 1 Head of data features

Shape: (4256, 51)

- The dataset has 51 columns and 4256 rows. The fist column Num is redundant and not important for analysis and model building.

INFO:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 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

 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
```

Figure 2 Info

- We have 51 numerical values and absence of any categorical values.
- Change in stock, Sales, Other income, Reserve and funds, Borrowings, Deferred tax liability etc have lot of missing values which needs to be treated.

DESCRIBE:

	count	mean	std	min	25%	50%	75%	max
Num	4256.0	2.128500e+03	1.228746e+03	1.000000e+00	1064.750	2128.500	3.192250e+03	4.256000e+03
Networth Next Year	4256.0	1.344741e+03	1.593674e+04	-7.426560e+04	3.975	72.100	3.308250e+02	8.057734e+05
Total assets	4256.0	3.573617e+03	3.007444e+04	1.000000e-01	91.300	315.500	1.120800e+03	1.176509e+06
Net worth	4256.0	1.351950e+03	1.296131e+04	0.000000e+00	31.475	104.800	3.898500e+02	6.131516e+05
Total income	4025.0	4.688190e+03	5.391895e+04	0.000000e+00	107.100	455.100	1.485000e+03	2.442828e+06
Change in stock	3706.0	4.370248e+01	4.369150e+02	-3.029400e+03	-1.800	1.600	1.840000e+01	1.418550e+04
Total expenses	4091.0	4.356301e+03	5.139809e+04	-1.000000e-01	96.800	426.800	1.395700e+03	2.366035e+06
Profit after tax	4102.0	2.950506e+02	3.079902e+03	-3.908300e+03	0.500	9.000	5.330000e+01	1.194391e+05
PBDITA	4102.0	6.059406e+02	5.646231e+03	-4.407000e+02	6.925	36.900	1.587000e+02	2.085765e+05
PBT	4102.0	4.102590e+02	4.217415e+03	-3.894800e+03	0.800	12.600	7.417500e+01	1.452926e+05
Cash profit	4102.0	4.082675e+02	4.143926e+03	-2.245700e+03	2.900	19.400	9.625000e+01	1.769118e+05
PBDITA as % of total income	4177.0	3.179892e+00	1.722566e+02	-6.400000e+03	4.970	9.680	1.647000e+01	1.000000e+02
PBT as % of total income	4177.0	-1.819683e+01	4.199111e+02	-2.134000e+04	0.560	3.340	8.940000e+00	1.000000e+02
PAT as % of total income	4177.0	-2.003367e+01	4.235762e+02	-2.134000e+04	0.350	2.370	6.420000e+00	1.500000e+02
Cash profit as % of total income	4177.0	-9.021278e+00	2.999574e+02	-1.502000e+04	2.000	5.660	1.073000e+01	1.000000e+02
PAT as % of net worth	4256.0	1.016786e+01	6.153240e+01	-7.487200e+02	0.000	8.040	2.020250e+01	2.466670e+03
Sales	3951.0	4.645685e+03	5.308090e+04	1.000000e-01	113.350	468.600	1.481200e+03	2.384984e+06
Income from fincial services	3145.0	8.136006e+01	1.042759e+03	0.000000e+00	0.500	1.900	9.800000e+00	5.193820e+04
Other income	2700.0	5.595289e+01	1.178415e+03	0.000000e+00	0.400	1.500	6.200000e+00	4.285670e+04
Total capital	4251.0	2.245577e+02	1.684951e+03	1.000000e-01	13.200	42.600	1.031500e+02	7.827320e+04
Reserves and funds	4158.0	1.210562e+03	1.281623e+04	-6.525900e+03	5.300	55.150	2.825250e+02	6.251378e+05
Borrowings	3825.0	1.176248e+03	8.581249e+03	1.000000e-01	24.400	99.800	3.583000e+02	2.782573e+05
Current liabilities & provisions	4146.0	9.606314e+02	9.140536e+03	1.000000e-01	17.500	70.300	2.659250e+02	3.522403e+05
Deferred tax liability	2887.0	2.344951e+02	2.106253e+03	1.000000e-01	3.200	13.500	5.130000e+01	7.279660e+04
Shareholders funds	4256.0	1.376487e+03	1.301069e+04	0.000000e+00	32.300	107.600	4.089000e+02	6.131516e+05
Cumulative retained profits	4211.0	9.371820e+02	9.853096e+03	-6.534300e+03	1.100	37.400	2.062000e+02	3.901338e+05
Capital employed	4256.0	2.433618e+03	2.049640e+04	0.000000e+00	61.300	221.200	7.903000e+02	8.914089e+05
TOL/TNW	4256.0	4.025343e+00	2.087909e+01	-3.504800e+02	0.600	1.420	2.830000e+00	4.730000e+02
Total term liabilities / tangible net worth	4256.0	1.854288e+00	1.587506e+01	-3.256000e+02	0.050	0.345	1.000000e+00	4.560000e+02
Contingent liabilities / Net worth (%)	4256.0	5.570750e+01	3.691657e+02	0.000000e+00	0.000	5.360	3.101250e+01	1.470427e+04

Figure 3 Describe

Contingent liabilities	2854.0	9.485522e+02	1.205674e+04	1.000000e-01	6.000	37.850	1.953250e+02	5.595068e+05
Net fixed assets	4124.0	1.209487e+03	1.250240e+04	0.000000e+00	26.200	93.850	3.528250e+02	6.366046e+05
Investments	2541.0	7.218659e+02	6.793860e+03	0.000000e+00	1.000	8.200	6.380000e+01	1.999786e+05
Current assets	4176.0	1.350360e+03	1.015557e+04	1.000000e-01	36.600	148.350	5.150000e+02	3.548152e+05
Net working capital	4219.0	1.628742e+02	3.182030e+03	-6.383900e+04	-1.100	16.700	8.650000e+01	8.578280e+04
Quick ratio (times)	4151.0	1.497355e+00	9.327519e+00	0.000000e+00	0.410	0.670	1.030000e+00	3.410000e+02
Current ratio (times)	4151.0	2.257398e+00	1.247829e+01	0.000000e+00	0.930	1.230	1.720000e+00	5.050000e+02
Debt to equity ratio (times)	4256.0	2.871563e+00	1.559997e+01	0.000000e+00	0.220	0.790	1.750000e+00	4.560000e+02
Cash to current liabilities (times)	4151.0	5.284197e-01	4.796342e+00	0.000000e+00	0.020	0.070	1.900000e-01	1.650000e+02
Cash to average cost of sales per day	4156.0	1.451579e+02	2.521992e+03	0.000000e+00	2.880	8.040	2.197000e+01	1.280408e+05
Creditors turnover	3865.0	1.681226e+01	7.567492e+01	0.000000e+00	3.720	6.170	1.169000e+01	2.401000e+03
Debtors turnover	3871.0	1.792903e+01	9.016443e+01	0.000000e+00	3.810	6.470	1.185000e+01	3.135200e+03
Finished goods turnover	3382.0	8.436999e+01	5.626374e+02	-9.000000e-02	8.190	17.320	4.001250e+01	1.794760e+04

WIP turnover	3492.0	2.868451e+01	1.696509e+02	-1.800000e-01	5.100	9.860	2.024000e+01	5.651400e+03
Raw material turnover	3828.0	1.773393e+01	3.431259e+02	-2.000000e+00	3.020	6.410	1.182250e+01	2.109200e+04
Shares outstanding	3446.0	2.376491e+07	1.709790e+08	-2.147484e+09	1308382.500	4750000.000	1.090602e+07	4.130401e+09
Equity face value	3446.0	-1.094829e+03	3.410136e+04	-9.999989e+05	10.000	10.000	1.000000e+01	1.000000e+05
EPS	4256.0	-1.962175e+02	1.306195e+04	-8.431818e+05	0.000	1.490	1.000000e+01	3.452253e+04
Adjusted EPS	4256.0	-1.975276e+02	1.306193e+04	-8.431818e+05	0.000	1.240	7.615000e+00	3.452253e+04
Total liabilities	4256.0	3.573617e+03	3.007444e+04	1.000000e-01	91.300	315.500	1.120800e+03	1.176509e+06
PE on BSE	1629.0	5.546229e+01	1.304445e+03	-1.116640e+03	2.970	8.690	1.700000e+01	5.100274e+04

Figure 4 Describe 2

- The description shows count of all features which is not same due to missing values.

```
Num                                      0
Networth Next Year                       0
Total assets                             0
Net worth                                0
Total income                           231
Change in stock                        550
Total expenses                         165
Profit after tax                       154
PBDITA                                 154
PBT                                    154
Cash profit                            154
PBDITA as % of total income             79
PBT as % of total income                79
PAT as % of total income                79
Cash profit as % of total income        79
PAT as % of net worth                    0
Sales                                  305
Income from fincial services          1111
Other income                          1556
Total capital                            5
Reserves and funds                      98
Borrowings                             431
Current liabilities & provisions       110
Deferred tax liability                1369
Shareholders funds                       0
Cumulative retained profits             45
Capital employed                         0
```

```
TOL/TNW                                              0
Total term liabilities / tangible net worth          0
Contingent liabilities / Net worth (%)               0
Contingent liabilities                            1402
Net fixed assets                                   132
Investments                                       1715
Current assets                                      80
Net working capital                                 37
Quick ratio (times)                                105
Current ratio (times)                              105
Debt to equity ratio (times)                         0
Cash to current liabilities (times)                105
Cash to average cost of sales per day              100
Creditors turnover                                 391
Debtors turnover                                   385
Finished goods turnover                            874
WIP turnover                                       764
Raw material turnover                              428
Shares outstanding                                 810
Equity face value                                  810
EPS                                                  0
Adjusted EPS                                         0
Total liabilities                                    0
PE on BSE                                         2627
dtype: int64
```

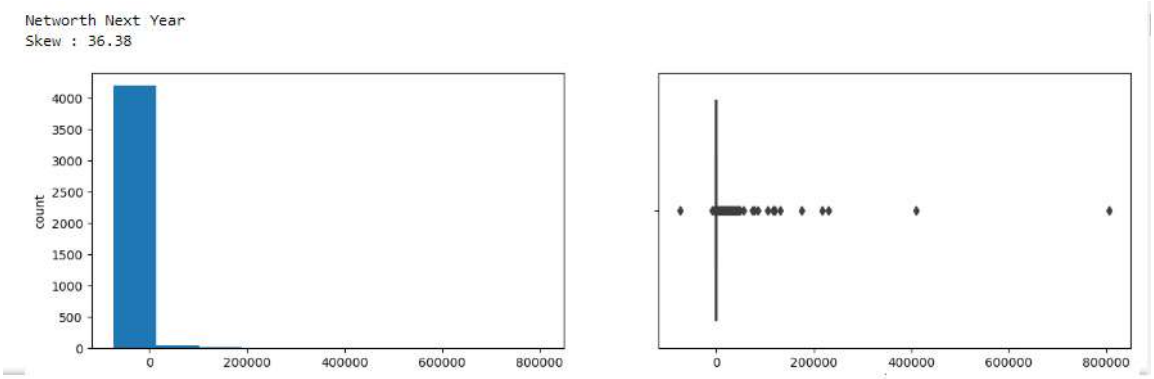
Figure 5 Null Values

- Deferred tax liability, Net fixed assets, Contingent liabilities and Other income have very high missing information.

Missing Values: 17778

- 8.19% is the total missing data which is significant enough to be treated. We will treat the missing values later after Univariate analysys.

UNIVARIATE ANALYSIS:



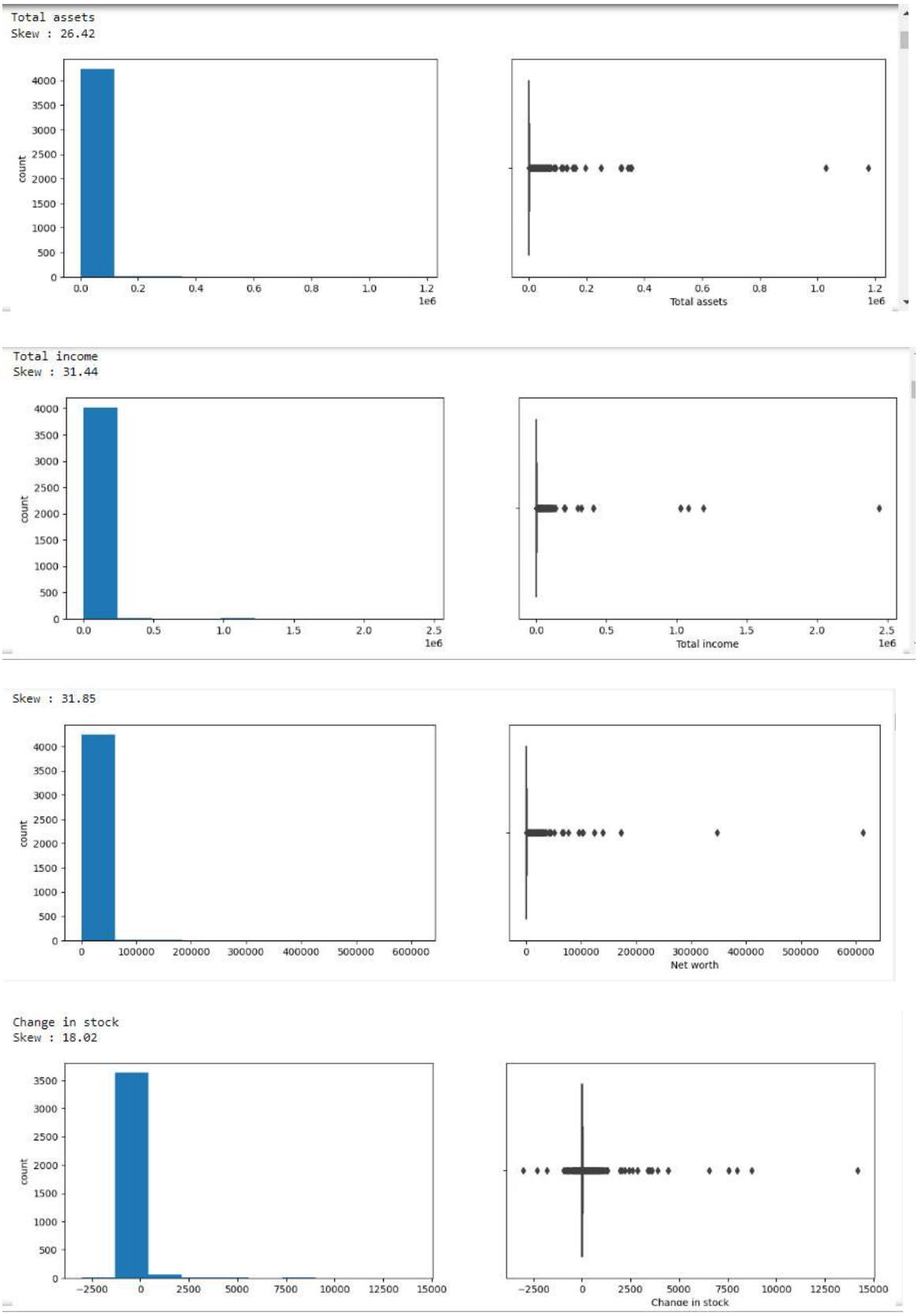


Figure 6 Univariate Analysis: Box & Histplot

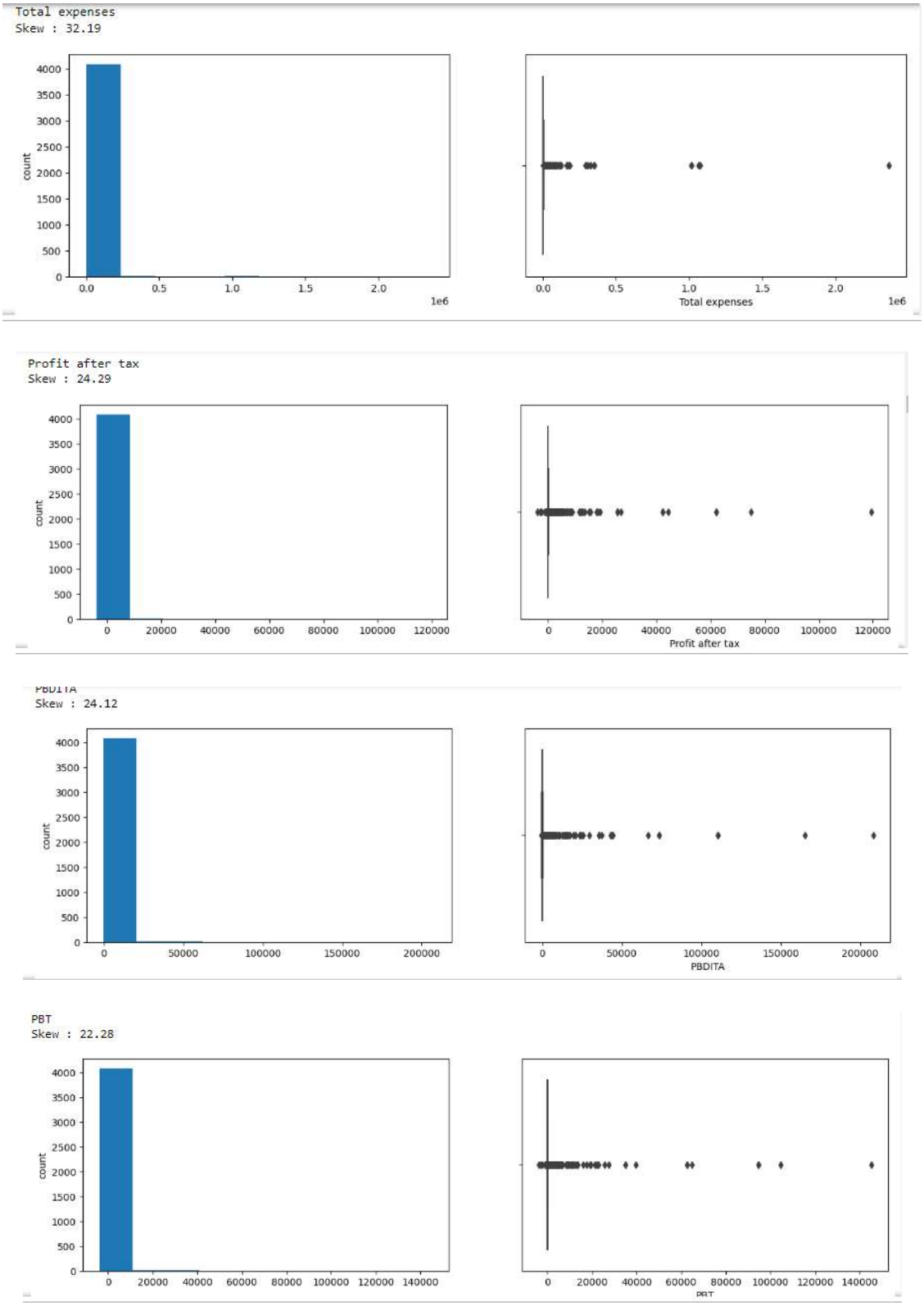


Figure 7 Univariate Analysis: Box & Histplot 2

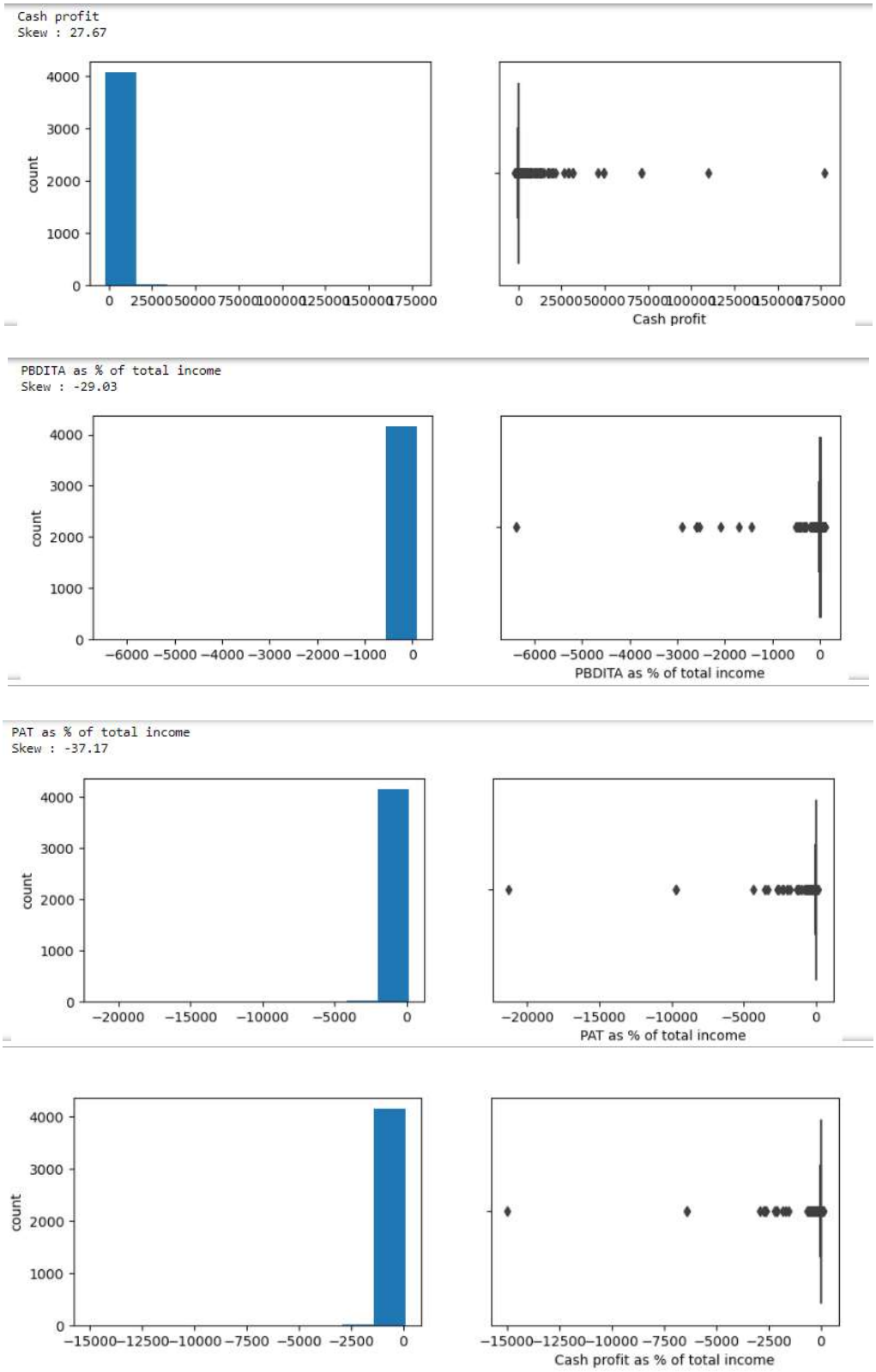


Figure 8 Univariate Analysis: Box & Histplot 3

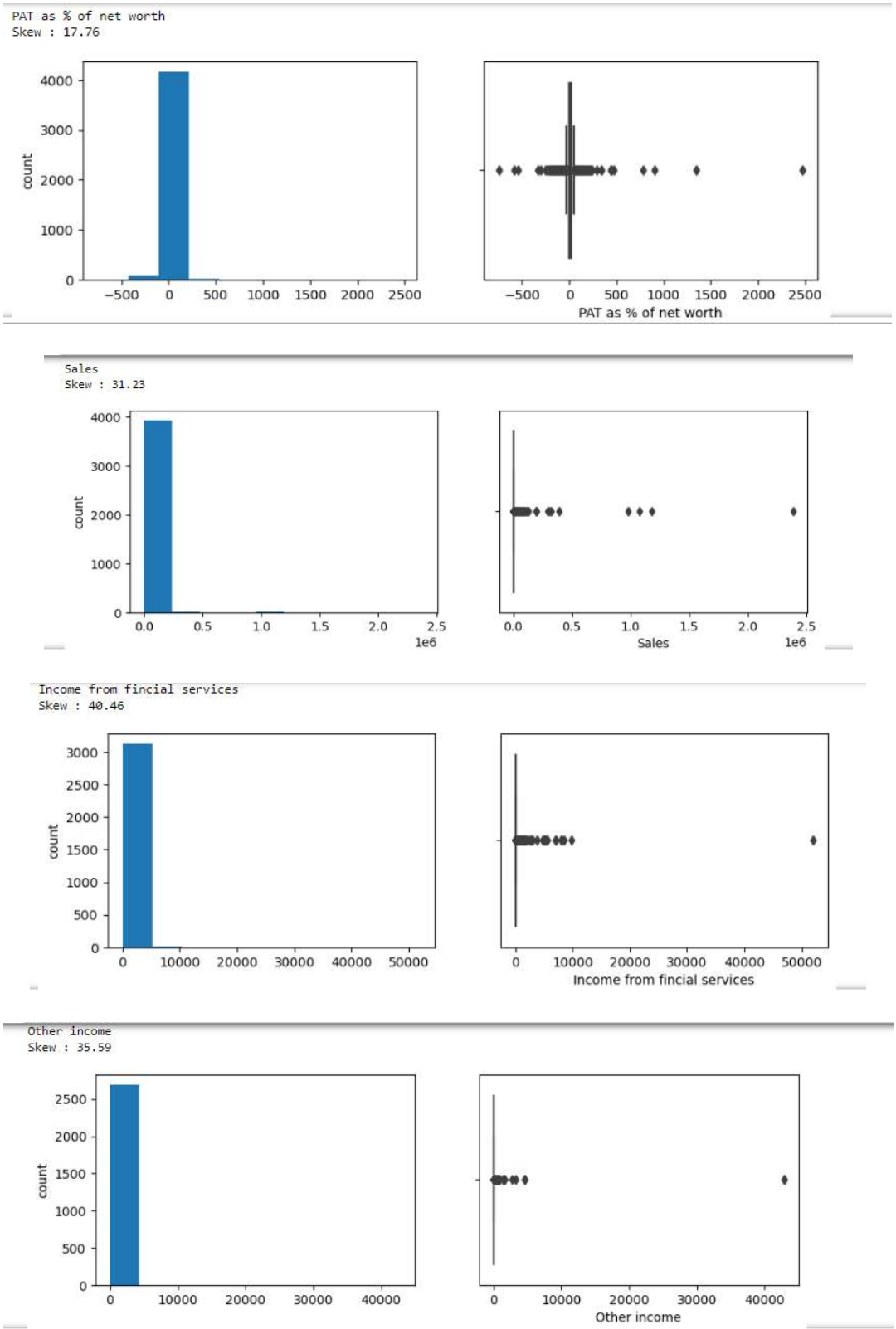
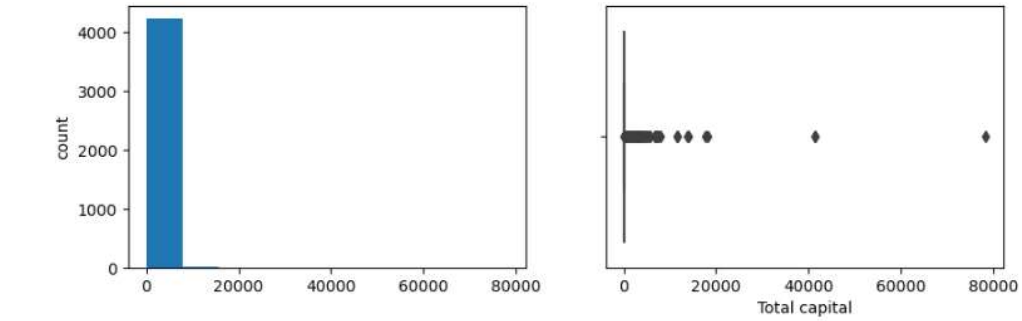
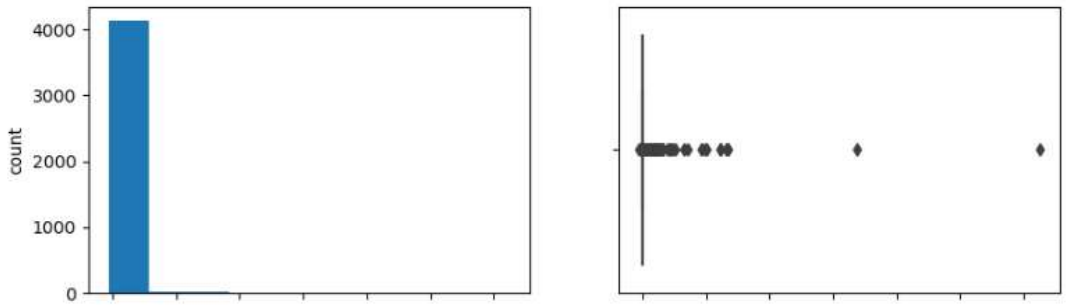


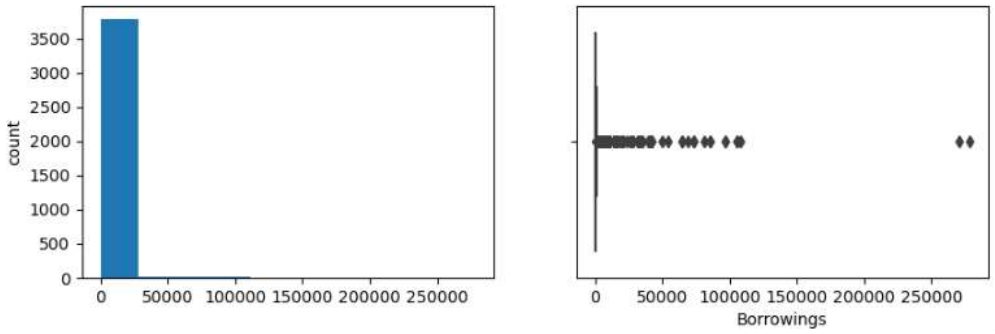
Figure 9 Univariate Analysis: Box & Histplot 4



Reserves and funds
Skew : 34.11



Borrowings
Skew : 20.89



Current liabilities & provisions
Skew : 26.51

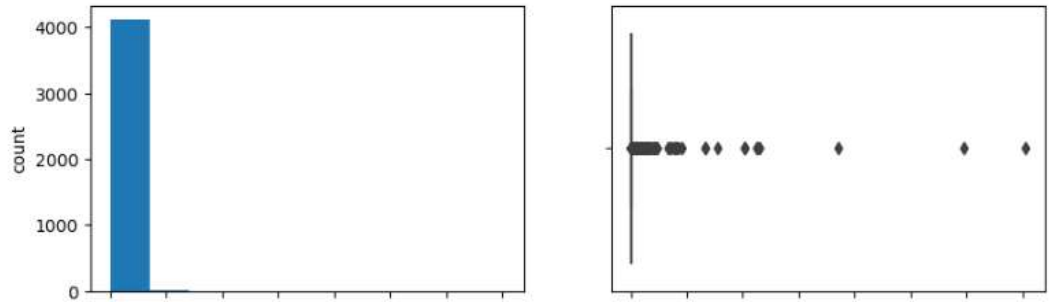


Figure 10 Univariate Analysis: Box & Histplot 5

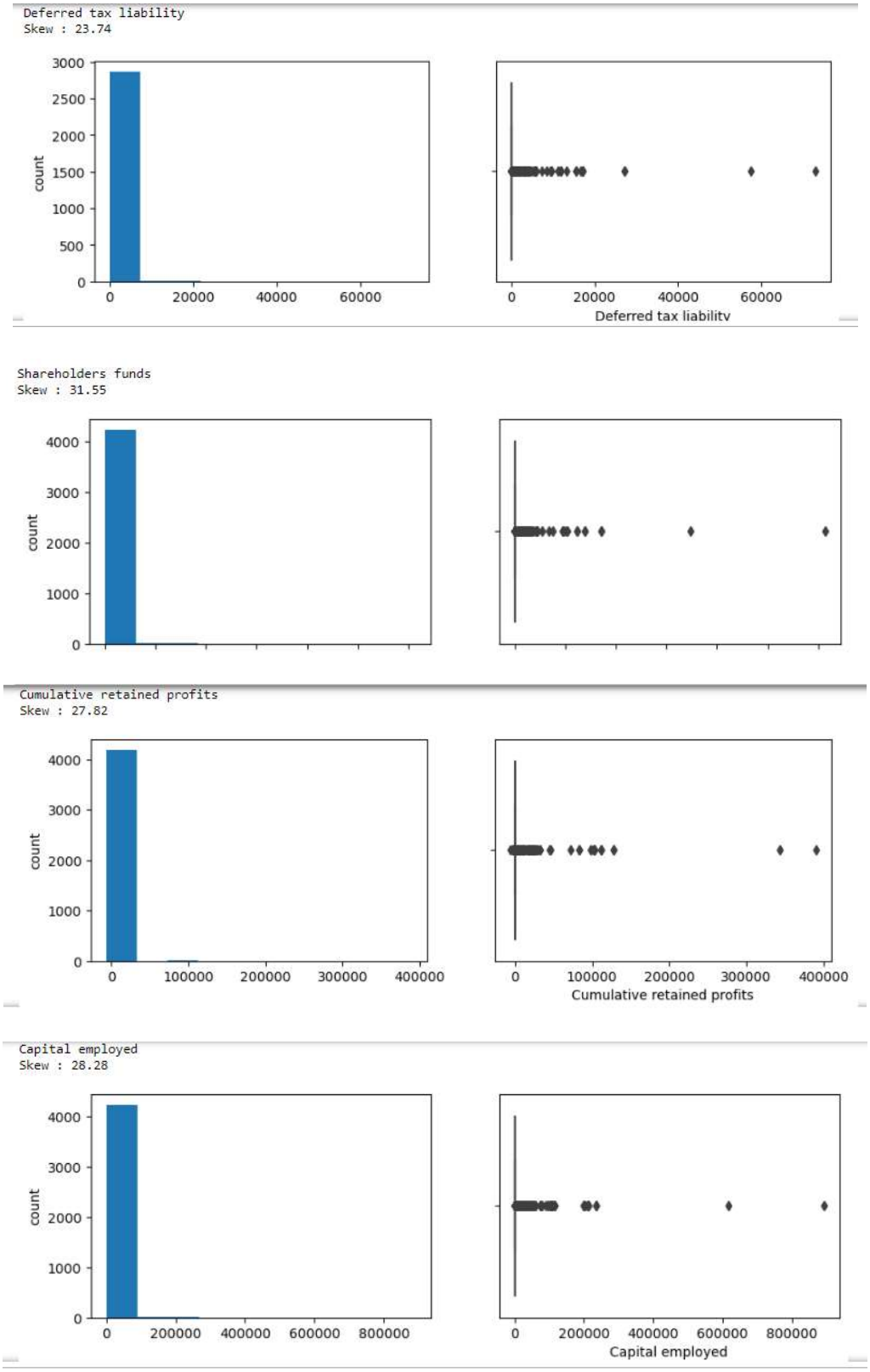


Figure 11 Univariate Analysis: Box & Histplot 6

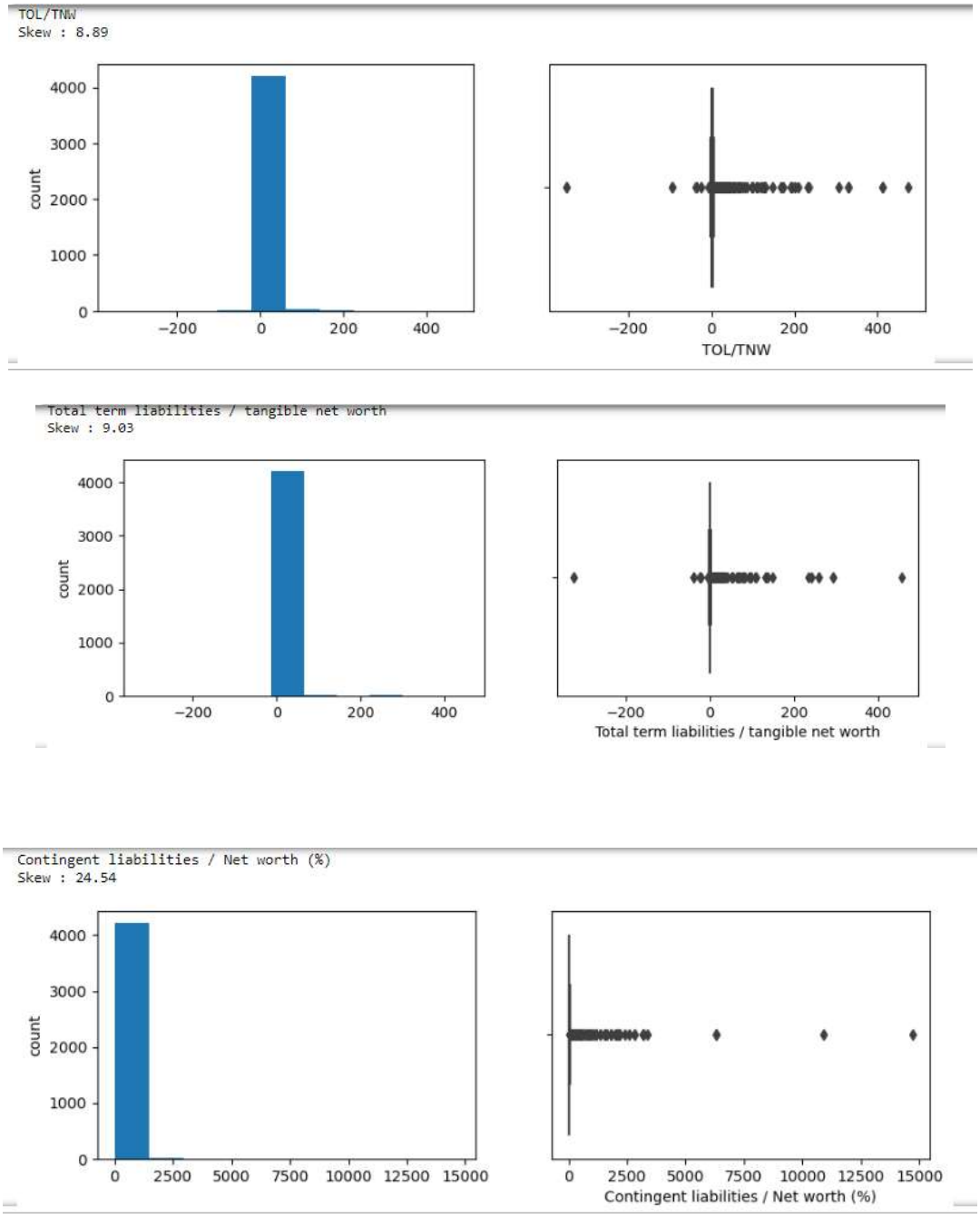
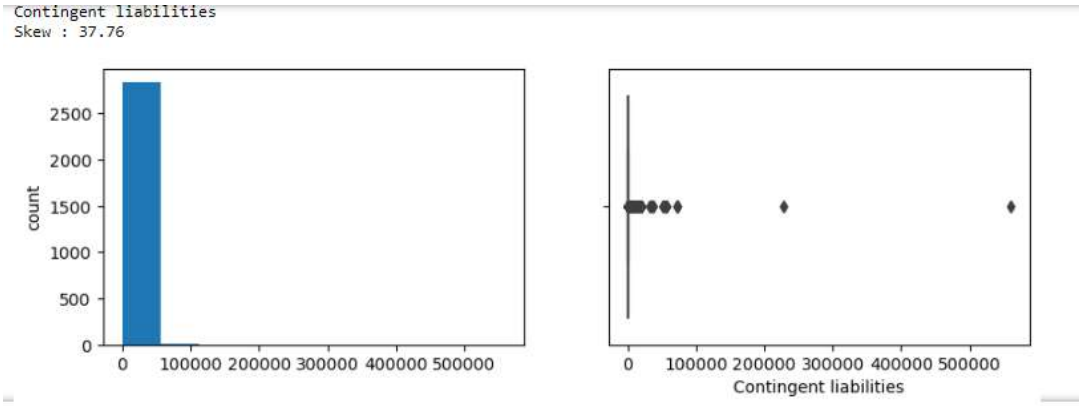
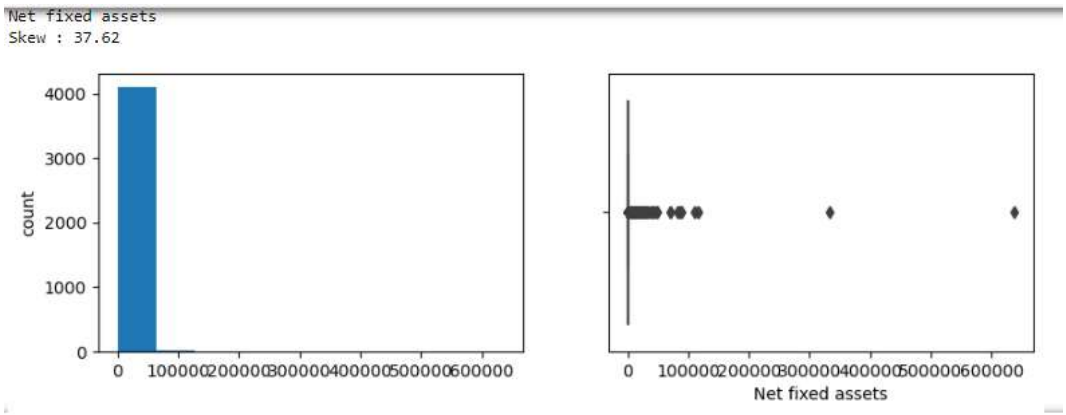
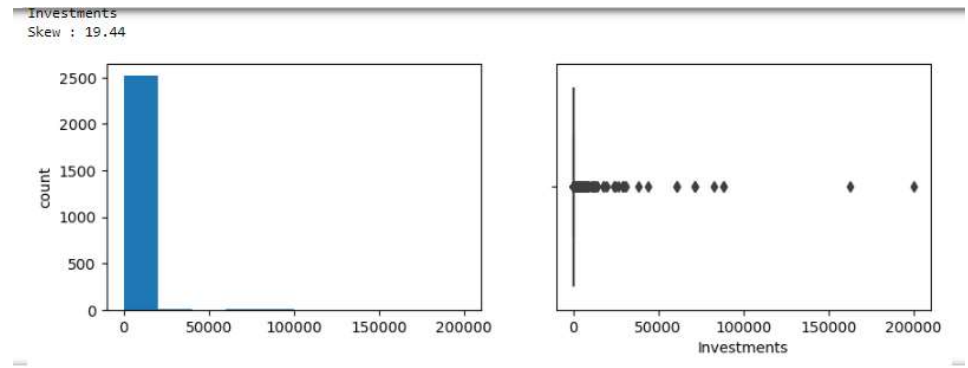


Figure 12 Univariate Analysis: Box & Histplot 7







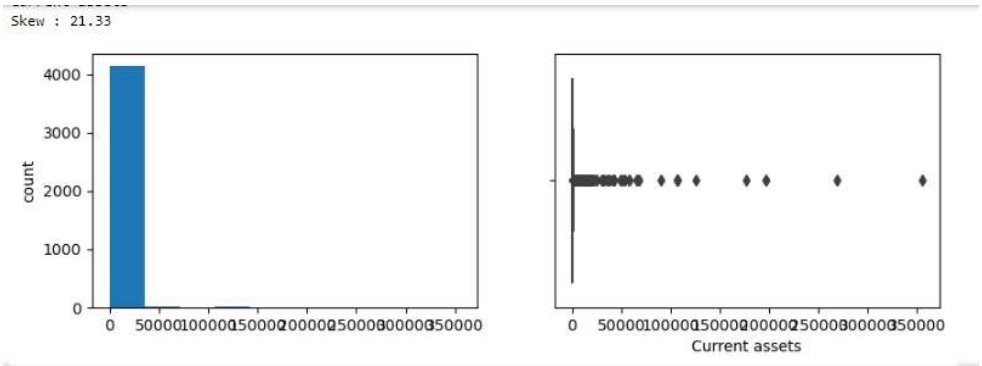


Figure 13 Univariate Analysis: Box & Histplot 8

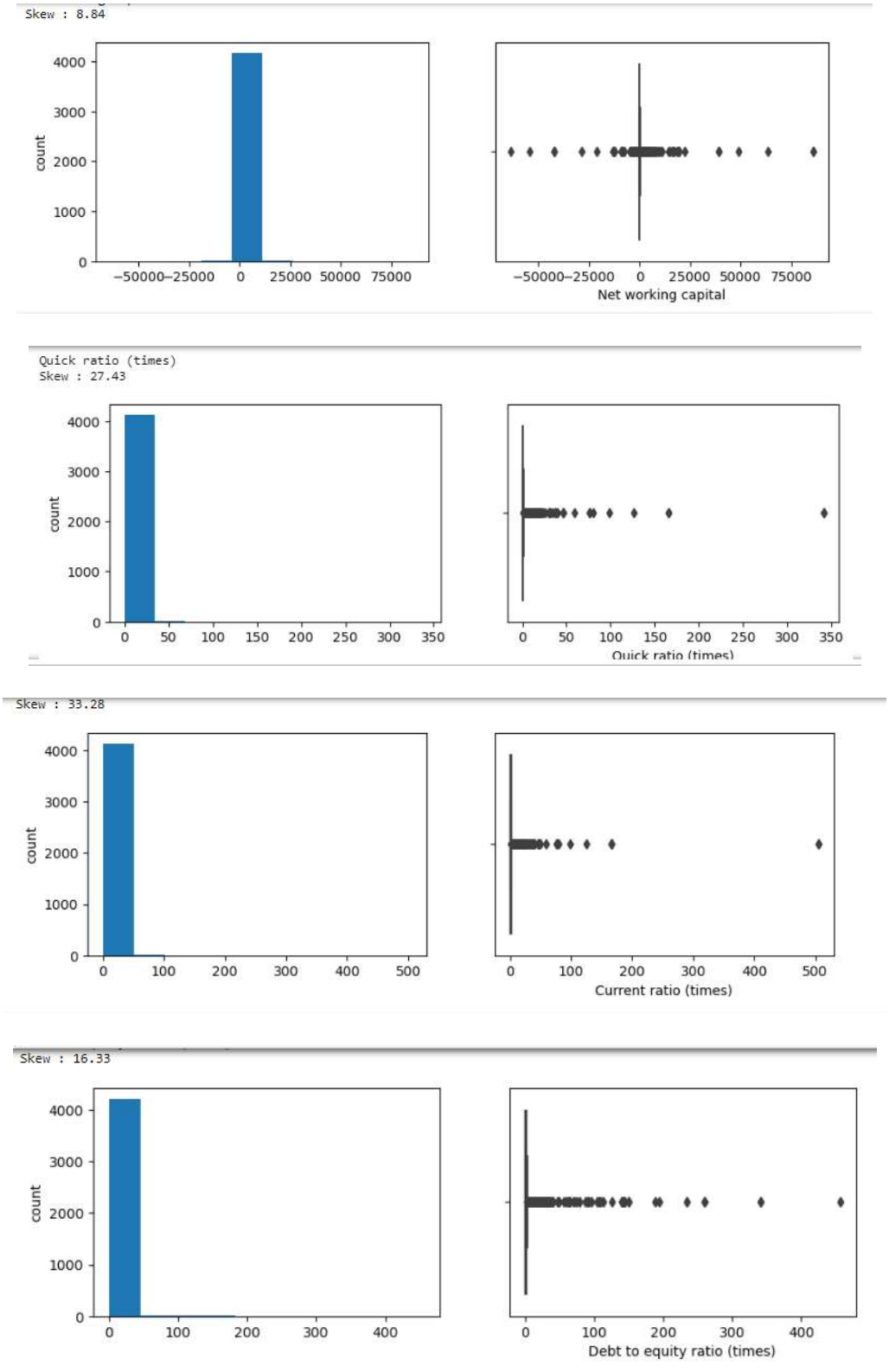


Figure 14 Univariate Analysis: Box & Histplot 9

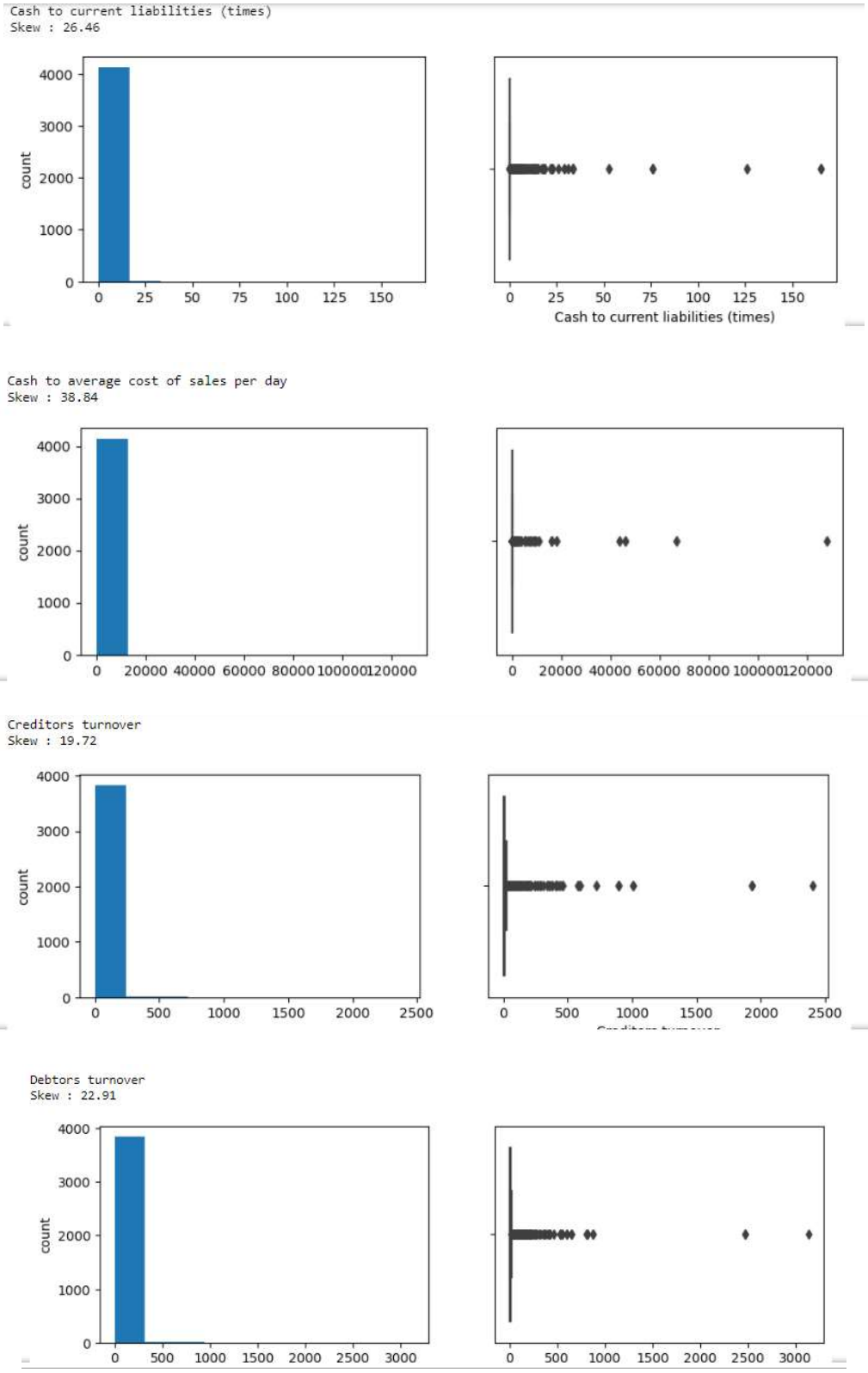


Figure 15 Univariate Analysis: Box & Histplot 10

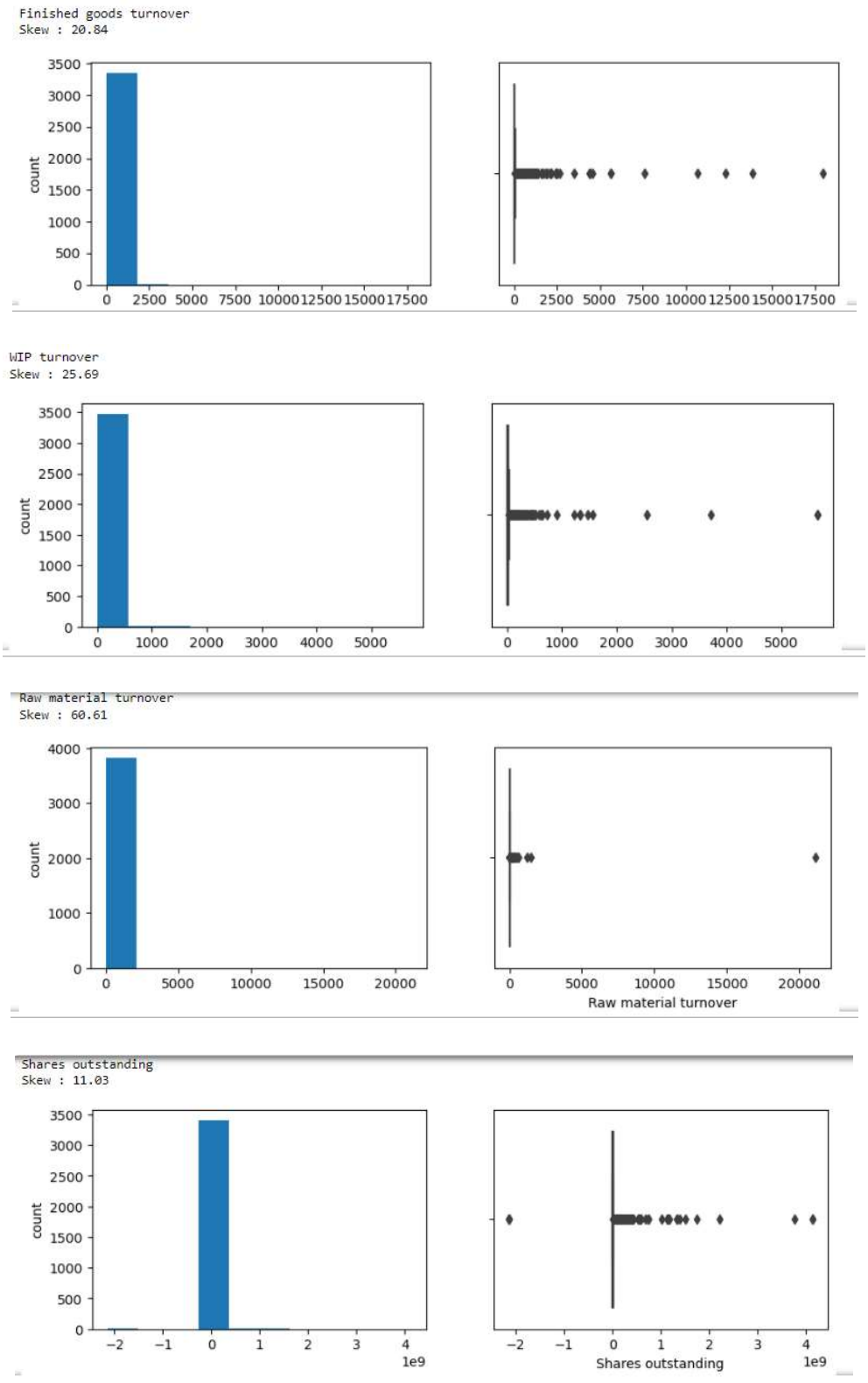


Figure 16 Univariate Analysis: Box & Histplot 11

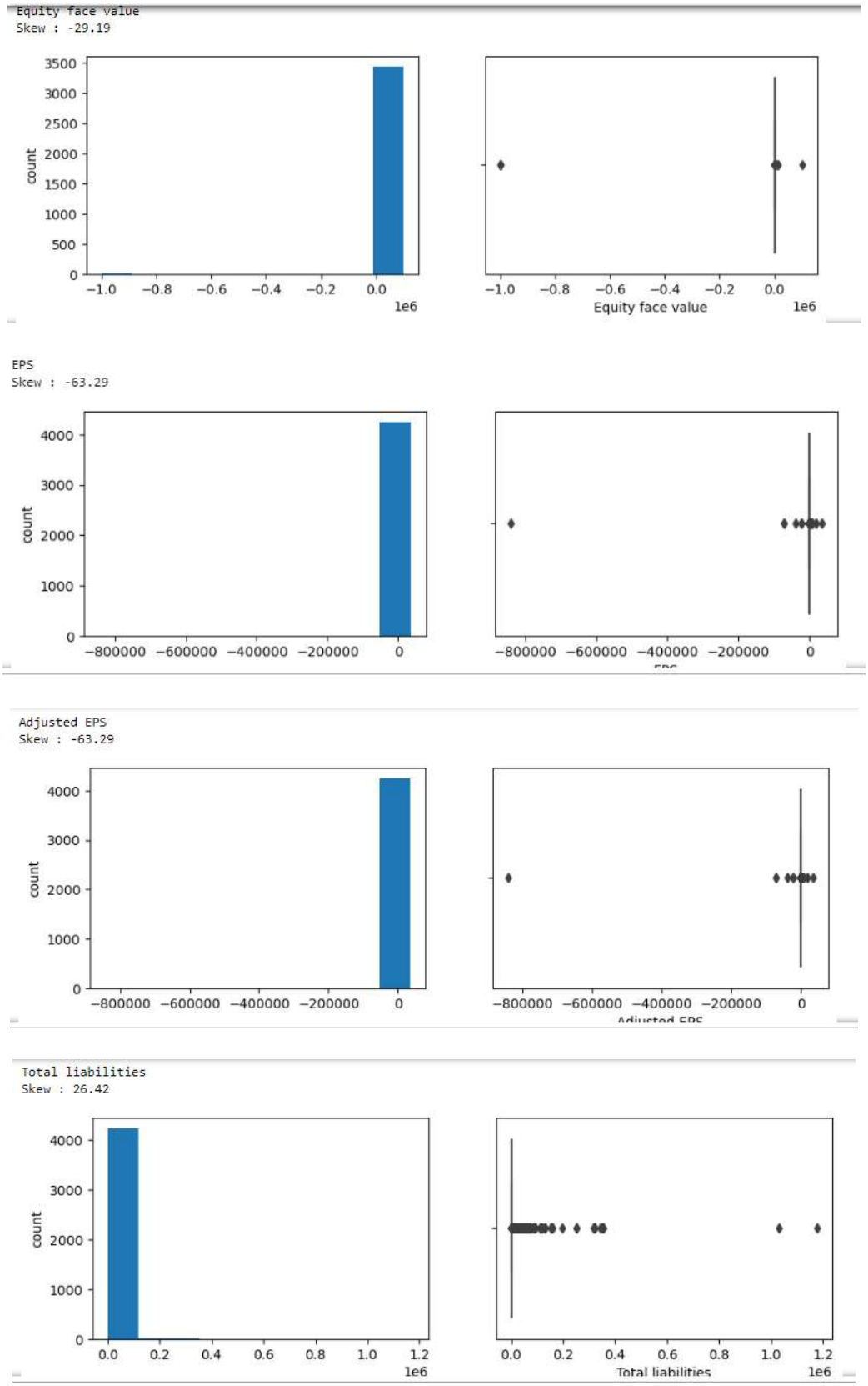


Figure 17 Univariate Analysis: Box & Histplot 12

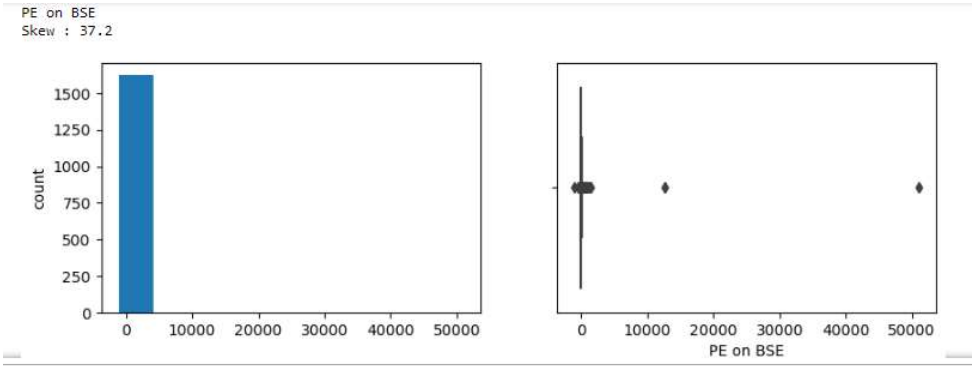


Figure 18 Univariate Analysis: Box & Histplot 12

Inferences from Univariate Analysis:

- Univariate analysis shows asymmetric data for most of the features.
- Univariate analysis shows that most of the features have outliers which needs to be treated.
- Treating outliers can exclude important data points for model building.
- Univariate analysis gives a strong reason to not treat the outliers which will exclude lot of data points for model building.

Bivariate Analysis:

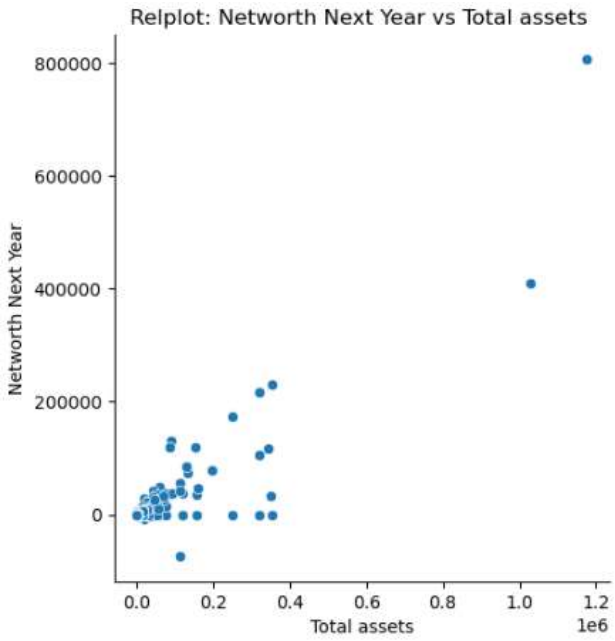


Figure 19 Bivariate Analysis: Networth Next year vs Total assets

- Networth Next Year vs Total assets shows a positive relation. Most data points are in range of 0 to .4 of total assests and 0 to 200000.

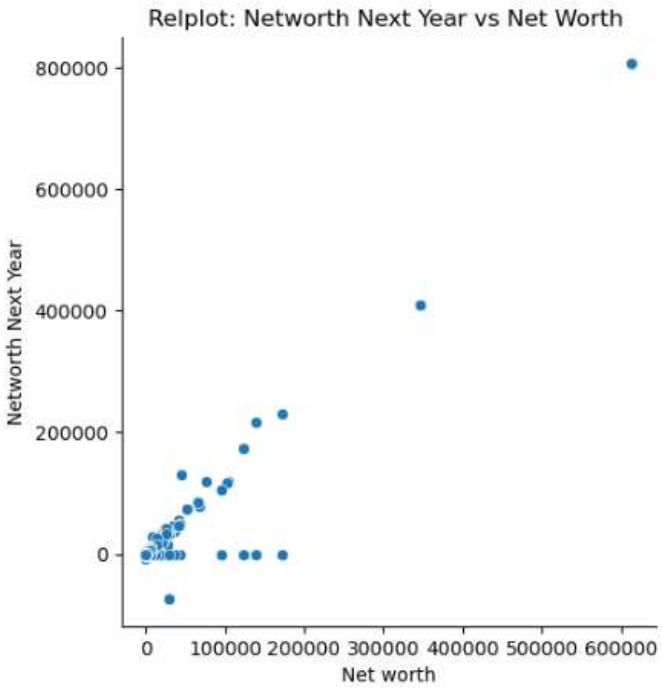


Figure 20 Bivariate Analysis: Networth Next year vs Net Worth

- Networth Next Year vs Net worth shows a positive relation. Most data points are in range of 0 to 100000 of total assests and 0 to 200000.

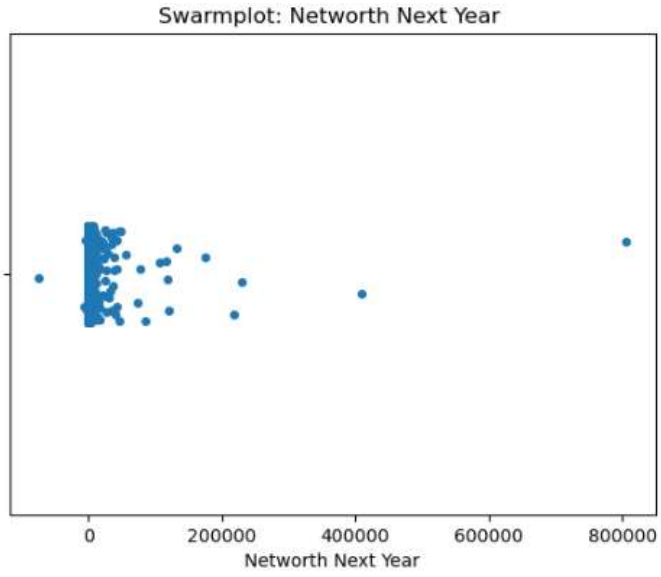


Figure 21 Swarmplot Networth Next Year

- Networth Next Year swarmplot shows high concentration between 0 to 50000 of networth.

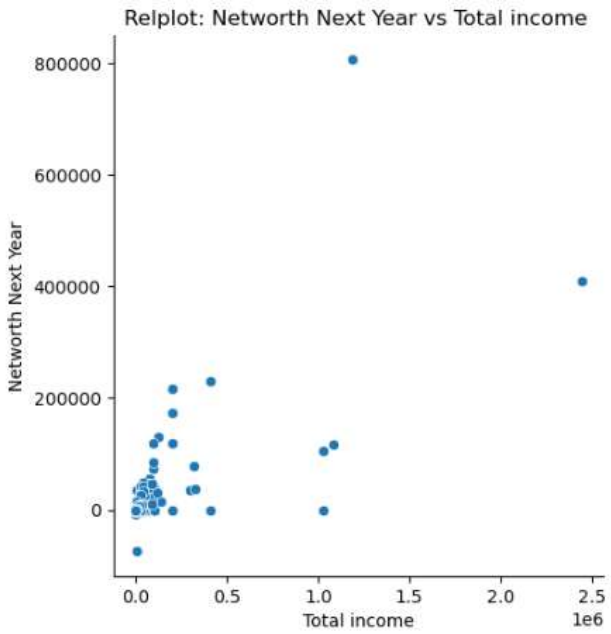


Figure 22 Bivariate Analysis: Networth Next year vs Total income

- Networth Next Year vs Total Income shows a unclear pattern. Most data points are in range of 0 to 0.25 of total income and 0 to 100000. Low Networth do not indicate that Income is less.

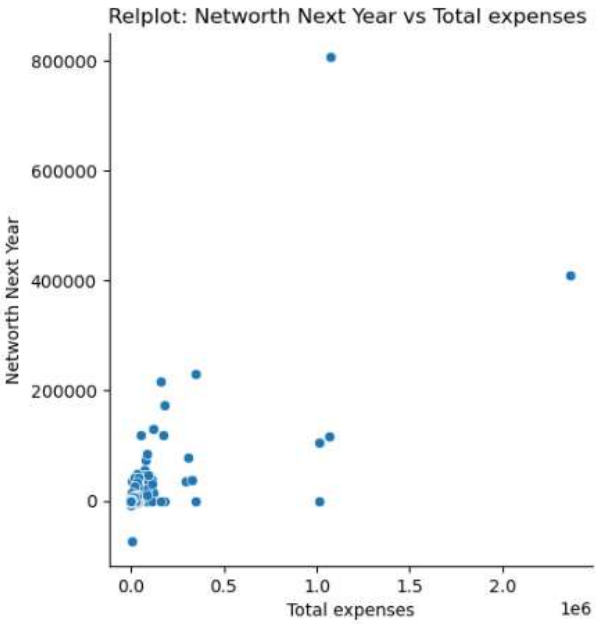


Figure 23 Bivariate Analysis: Networth Next year vs Total expenses

- Networth Next Year vs Total expenses shows a positive relation. Most data points are in range of 0 to 100000 of total expenses and 0 to 200000.

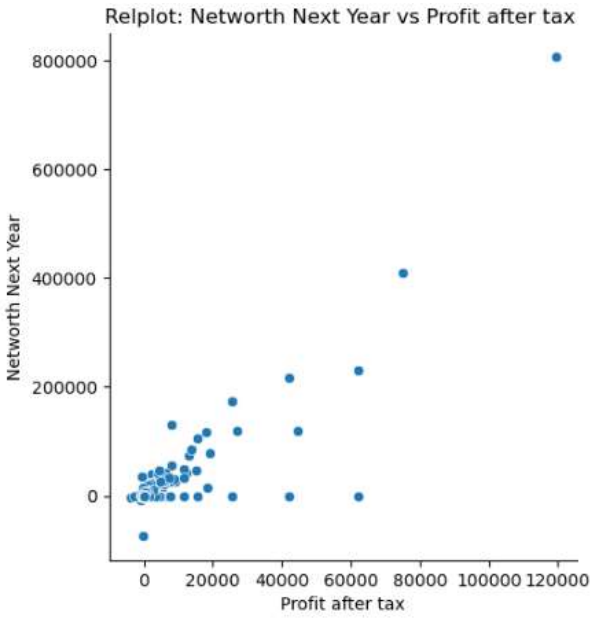


Figure 24 Bivariate Analysis: Networth Next year vs Profit AfterTax

- Networth Next Year vs Net worth shows a positive relation. Most data points are in range of 0 to 100000 of total assets and 0 to 200000.

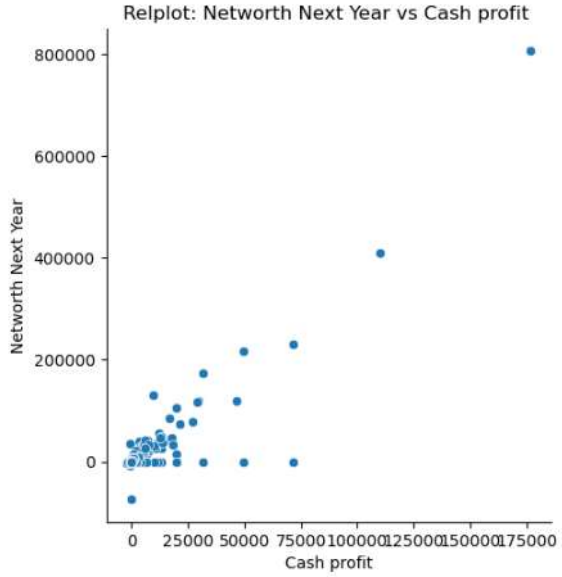


Figure 25 Bivariate Analysis: Networth Next year vs Cash Profit

- Networth Next Year vs Cash profit shows a positive relation. Most data points are in range of 0 to 100000 of Cash Profit and 0 to 100000 of Networth Next Year.

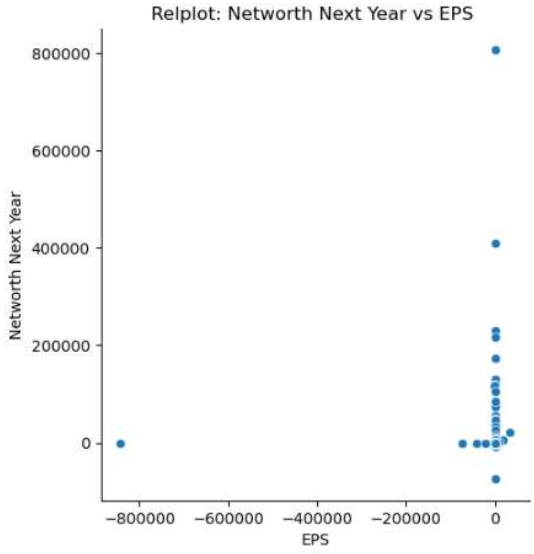


Figure 26 Bivariate Analysis: Networth Next year vs EPS

- Networth Next Year vs EPS shows a unclear relation. Most data points have zero EPS for value in range 0 to 200000 of Networth Next Year.

Multivariate Analysis:

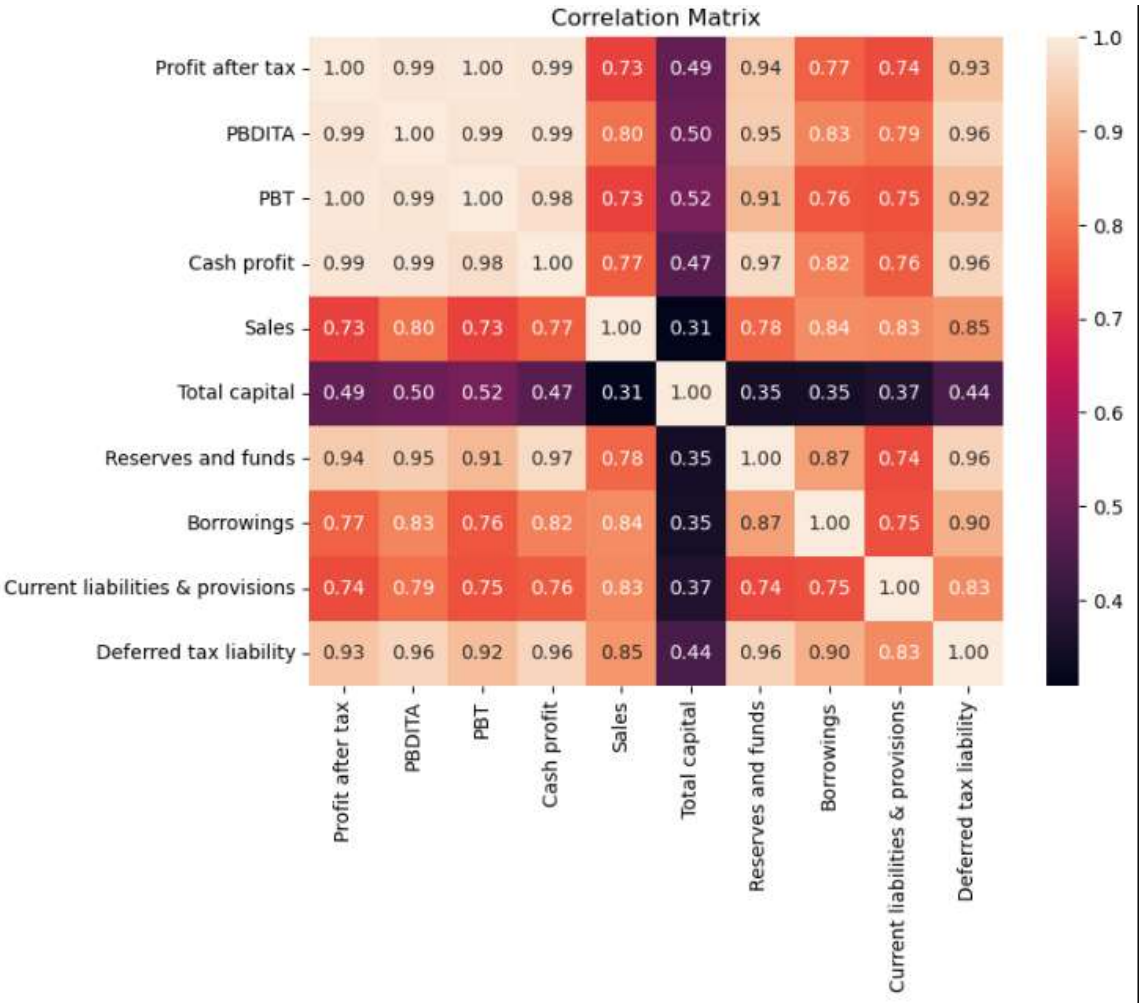


Figure 27 Multivariate Analysis: Heat map

- We can see high multicollinierity between features.
- Sales is having strong correlation with every variable except Total Capital
- Total Capital is having low correlation with other features making it independent.
- Profit after Tax and PBDITA have very high correlation.
- We need to focus on only independent features to build the model.

PART A: Data Pre-processing

*Prepare the data for 32odelling: - Outlier Detection (treat, if needed) – Encode the data – Data split – Scale the data – Target variable creation * The target variable is default and should take the value 1 when net worth next year is negative & 0 when net worth next year is positive*

Cheking outliers:

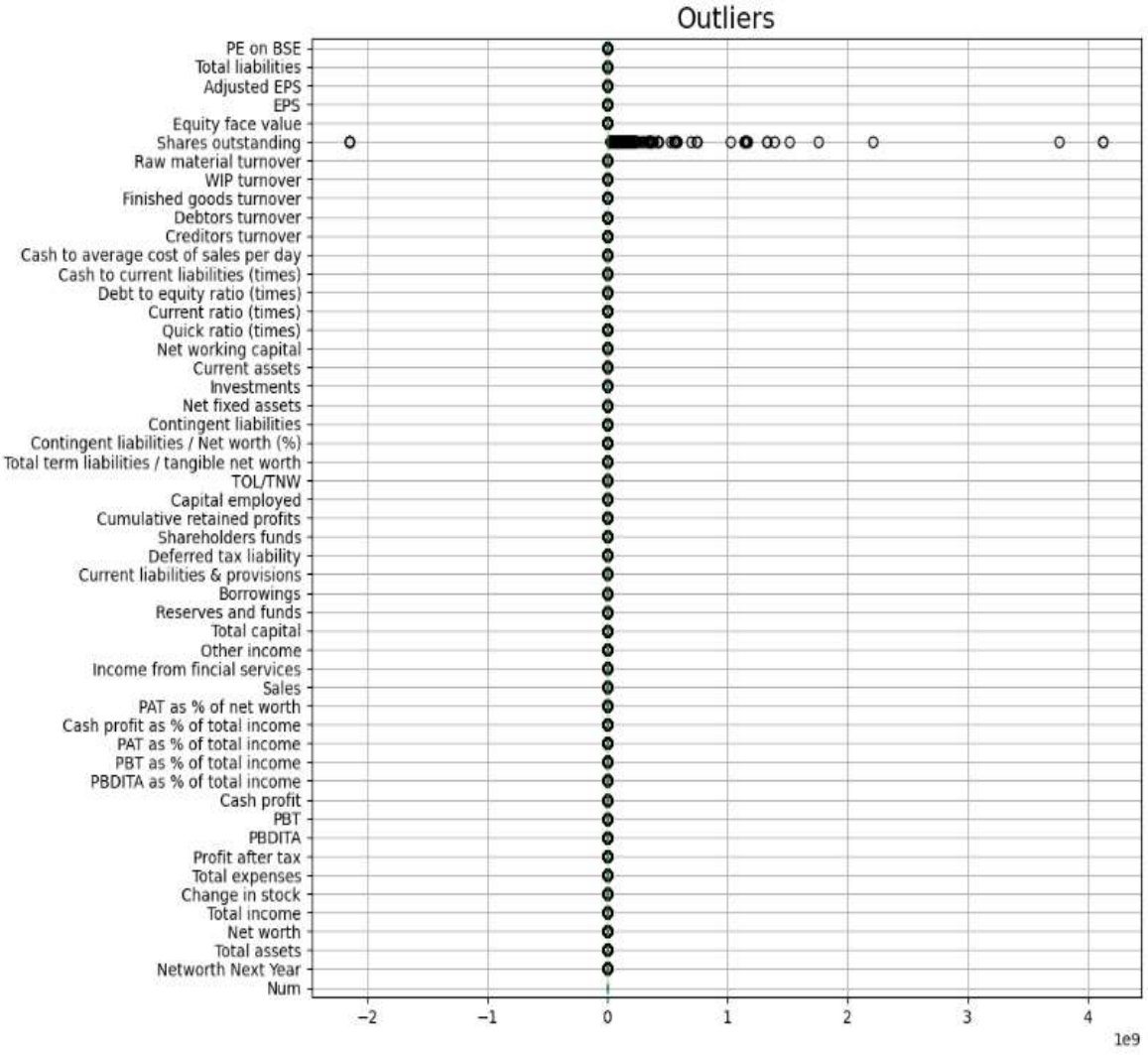


Figure 28 Outliers check

- We can see high number of outliers in shares outstanding. Based on the univariate analysis it does not seem suitable to treat the outliers.

After Treating the outliers the performance of the models were highly unsatisfactory. Therefore, we will not do any outlier treatment to ensure that model performance is maintained. WHEN WE DID OUTLIER TREATMENT WE FOUND THAT THE CONFUSION MATRIX WAS UNABLE TO PREDICT FN & TN VALUES APPROPRIATELY.

We will impute missing values with mean.

Num	0
Networth Next Year	0
Total assets	0
Net worth	0
Total income	0
Change in stock	0
Total expenses	0
Profit after tax	0
PBDITA	0
PBT	0
Cash profit	0
PBDITA as % of total income	0
PBT as % of total income	0
PAT as % of total income	0
Cash profit as % of total income	0
PAT as % of net worth	0
Sales	0
Income from fincial services	0
Other income	0
Total capital	0
Reserves and funds	0
Borrowings	0
Current liabilities & provisions	0
Deferred tax liability	0
Shareholders funds	0
Cumulative retained profits	0
Capital employed	0
TOL/TNW	0
Total term liabilities / tangible net worth	0
Contingent liabilities / Net worth (%)	0
Contingent liabilities	0
Net fixed assets	0
Investments	0
Current assets	0
Net working capital	0
Quick ratio (times)	0
Current ratio (times)	0
Debt to equity ratio (times)	0
Cash to current liabilities (times)	0
Cash to average cost of sales per day	0
Creditors turnover	0
Debtors turnover	0
Finished goods turnover	0
WIP turnover	0
Raw material turnover	0
Shares outstanding	0
Equity face value	0
EPS	0

```
Adjusted EPS                                  0
Total liabilities                             0
PE on BSE                                     0
dtype: int64
```

Figure 29 Missing Value Impute

There are no null/ missing values after imputation.

Scaling data:

	Total assets	Net worth	Total income	Change in stock	Total expenses	Profit after tax	PBDITA	PBT	Cash profit	PBDITA as % of total income	...	Debtors turnover	Finished goods turnover	WIP turnover
0	-0.091318	-0.078354	-7.923304e-02	-7.408876e-02	-0.076363	-8.472554e-02	-8.688209e-02	-0.083494	-0.076963	0.117858	...	-1.428151e-01	-1.602867e-01	-1.647552e-01
1	-0.116588	-0.102444	-8.678996e-02	-1.162815e-01	-0.083859	-9.653384e-02	-1.083347e-01	-0.098857	-0.099432	0.004748	...	4.132095e-17	2.833804e-17	-2.312224e-17
2	-0.110912	-0.098231	-8.310305e-02	-1.516057e-01	-0.080322	-9.630231e-02	-1.046720e-01	-0.096562	-0.098056	0.027018	...	-1.793358e-01	-1.330073e-01	-1.296753e-01
3	0.110069	0.007049	7.172228e-02	4.133360e-01	0.081890	-3.861696e-02	-3.383707e-02	-0.054387	-0.056608	0.010374	...	-1.863143e-01	-1.320700e-01	-6.550319e-02
4	-0.115817	-0.100692	-8.200823e-02	-9.886476e-02	-0.078665	-9.782382e-02	-1.080279e-01	-0.099244	-0.099408	-0.007794	...	5.823661e-01	-7.677331e-02	-9.445654e-05
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4251	-0.118826	-0.104303	-1.734725e-17	1.743010e-17	0.000000	1.880179e-17	-2.051198e-17	0.000000	0.000000	-0.018636	...	-2.085292e-01	2.833804e-17	-2.312224e-17
4252	-0.113532	-0.097629	-8.612239e-02	-1.069599e-01	-0.083091	-9.650076e-02	-1.060072e-01	-0.098205	-0.097269	0.043721	...	-1.875937e-01	-1.463080e-01	-1.327993e-01
4253	-0.091112	-0.053037	-4.527076e-02	-2.846169e-02	-0.043756	-3.302704e-02	-4.646683e-02	-0.025909	-0.046406	0.069566	...	-1.378138e-01	-5.003226e-02	1.598112e-02
4254	-0.115680	-0.100576	-8.731257e-02	-9.592107e-02	-0.084206	-9.706306e-02	-1.075769e-01	-0.098471	-0.098719	0.032820	...	-1.653789e-01	-1.072830e-02	-1.117774e-01
4255	-0.106050	-0.095731	-8.282458e-02	-7.948551e-02	-0.079677	-9.249851e-02	-9.893450e-02	-0.094099	-0.091811	0.079002	...	-1.537480e-01	-6.181746e-02	-1.324739e-01

4256 rows × 49 columns

Figure 30 Data Scaling

Target Variable:

The task is to analyze the data provided and develop a predictive model leveraging machine learning techniques to identify whether a given company will be tagged as a defaulter in terms of net worth next year.

A company will not be tagged as a defaulter if its net worth next year is positive, or else, it'll be tagged as a defaulter.

We will create **'Defaulter' as Target variable** where Default is '0' if net worth of company next year is positive else if it is negative or zero its value is '1' which means Defaulter . Lets see the value count of True values:

```
Default
0    3352
1     904
Name: count, dtype: int64
```

- We see that 3352 companies seem to be positive and 904 companies have negative networth and have been categorized as defaulter.

Data Split:

We split the test train data in 70:30 ratio for model building and validation.

X_train

	const	Total assets	Change in stock	PBDITA as % of total income	Cash profit as % of total income	PAT as % of net worth	Income from fincial services	Other income	Total capital	Contingent liabilities / Net worth (%)	...	Cash to current liabilities (times)	Cash to average cost of sales per day	Creditors turnover	Debtors turnover	Fir ! tur
2045	1.0	-0.039348	0.449151	0.037274	0.045305	-0.000940	-0.066232	-0.052378	-0.098029	0.008270	...	-0.098901	-0.051723	0.099961	-0.152352	-0.1
523	1.0	-0.116622	-0.106715	0.044835	0.061796	-0.055552	-0.090668	-0.059518	-0.103434	-0.144173	...	-0.048228	-0.050603	-0.099746	-0.162704	-0.0
853	1.0	-0.085339	0.260264	0.135088	0.096159	0.600114	-0.090445	0.000000	-0.041786	-0.003569	...	-0.098901	-0.048978	-0.200293	0.396040	-0.1
2738	1.0	0.087855	0.224940	0.117448	0.095419	0.319254	-0.056190	-0.024885	0.158123	-0.124749	...	-0.022892	-0.040009	-0.181016	-0.135255	-0.0
209	1.0	-0.052769	-0.255371	0.039794	0.048166	0.122586	-0.089329	-0.058559	-0.100702	-0.112774	...	-0.107347	-0.056274	-0.197658	-0.189222	-0.0

5 rows × 21 columns

Figure 31 Split Data: X_train

y_train

```
2045    1
523     1
853     1
2738    0
209     1
       ..
450     1
443     1
2187    0
1616    1
277     1
Name: Default, Length: 2979, dtype: int32
```

Figure 32 Split Data: y_train

X_test

	Total assets	Net worth	Total income	Change in stock	Total expenses	Profit after tax	PBDITA	PBT	Cash profit	PBDITA as % of total income	...	Debtors turnover	Finished goods turnover	WIP turnover
221	-0.083959	-0.098640	-6.059441e-02	-1.511151e-01	-0.057167	-9.180391e-02	-8.435614e-02	-0.094630	-0.085764	0.035047	...	-1.095510e-01	-1.179118e-01	-1.145109e-01
3427	-0.118048	-0.103030	-8.888232e-02	1.743010e-17	-0.085909	-9.742691e-02	-1.090924e-01	-0.098978	-0.100047	0.008381	...	-1.706127e-01	2.833804e-17	-2.312224e-17
1493	-0.111650	-0.096394	-8.510195e-02	-6.697486e-02	-0.081650	-9.742691e-02	-1.061695e-01	-0.098881	-0.098203	0.026667	...	-1.717758e-01	-1.204244e-01	-1.524545e-01
2839	-0.071897	-0.065229	-1.734725e-17	1.743010e-17	0.000000	1.880179e-17	-2.051198e-17	0.000000	0.000000	-0.018636	...	4.132095e-17	2.833804e-17	-2.312224e-17
1531	0.069159	0.168024	5.622888e-02	-1.374327e-02	0.020282	6.618763e-01	3.821875e-01	0.535046	0.431252	0.190413	...	-1.529339e-01	-7.539737e-02	-3.940478e-02

5 rows × 49 columns

Figure 33 Split Data: X_test

y_test

```
221     1
3427    1
1493    1
2839    1
1531    0
       ..
3750    1
1551    1
1381    1
1867    1
50      1
Name: Default, Length: 1277, dtype: int32
```

Figure 34 Split Data: y_test

PART A: Model Building

- Metrics of Choice (Justify the evaluation metrics) - Model Building (Logistic Regression, Random Forest) - Model performance check across different metrics

Which metric is most important?

- We need to make sure that False-Positives (FP) are less for model
- Large FP values is critical as it reflects those defaulter companies which have been predicted as non-defaulter by the model.
- We will use **precision** to evaluate the performance which is a measure of FP
- Precision = TP/(TP+FP)
- Higher the precision better is the model and lesser is the FP values.
- We will discuss the metrics later as well after building the model.

Logistics Regression Model:

We will build a model using Logit() and will fit the train data into model. We found that the model had high multicolliniearity and without dealing with such dependent variables model was not working. Multicollinearity has been dealt extensively before building the model which is explained later in the multicollinearity section. Lets see the first Logistic Regression model performance.

Model:	Logit	Method:	MLE
Dependent Variable:	Default	Pseudo R-squared:	0.037
Date:	2024-09-27 18:14	AIC:	2997.2599
No. Observations:	2979	BIC:	3123.2461
Df Model:	20	Log-Likelihood:	-1477.6
Df Residuals:	2958	LL-Null:	-1534.3
Converged:	1.0000	LLR p-value:	4.9608e-15
No. Iterations:	9.0000	Scale:	1.0000

Figure 35 Logistic Model 1 Performance

- **Logistic Model 1:**
  - The R-squared value is 0.037 and AIC is 2997.
  - We will build more models after removing those features which are insignificant by calculating their p-values.

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
const	-1.3976	0.0525	-26.6061	0.0000	-1.5005	-1.2946
Total assets	-0.0608	0.1461	-0.4159	0.6775	-0.3471	0.2256
Change in stock	-0.0173	0.0880	-0.1969	0.8439	-0.1897	0.1551
PBDITA as % of total income	0.0063	0.0790	0.0793	0.9368	-0.1486	0.1612
Cash profit as % of total income	-0.1052	0.1202	-0.8753	0.3814	-0.3409	0.1304
PAT as % of net worth	-0.3418	0.0784	-4.3606	0.0000	-0.4955	-0.1882
Income from fincial services	0.0155	0.1058	0.1466	0.8834	-0.1919	0.2229
Other income	-0.0186	0.1005	-0.1851	0.8532	-0.2157	0.1785
Total capital	-0.0192	0.0689	-0.2791	0.7801	-0.1544	0.1159
Contingent liabilities / Net worth (%)	0.1648	0.0755	2.1829	0.0290	0.0168	0.3128
Debt to equity ratio (times)	0.2857	0.0717	3.9859	0.0001	0.1452	0.4262
Cash to current liabilities (times)	0.0286	0.0364	0.7868	0.4314	-0.0427	0.1000
Cash to average cost of sales per day	0.1682	0.1202	1.3991	0.1618	-0.0674	0.4039
Creditors turnover	0.0779	0.0430	1.8108	0.0702	-0.0064	0.1622
Debtors turnover	0.0391	0.0336	1.1634	0.2447	-0.0267	0.1049
Finished goods turnover	-0.0251	0.0640	-0.3927	0.6946	-0.1505	0.1002
WIP turnover	0.0569	0.0473	1.2026	0.2291	-0.0358	0.1495
Raw material turnover	-1.0188	0.8380	-1.2156	0.2241	-2.6613	0.6238
Shares outstanding	0.0095	0.0785	0.1205	0.9041	-0.1443	0.1632
Equity face value	-0.0247	0.0504	-0.4896	0.6244	-0.1234	0.0741
PE on BSE	-1.5286	0.8523	-1.7935	0.0729	-3.1990	0.1418

Figure 36 P-values of Logistic Model 1

We will check the significant variables where p-value is greater than 0.05. We will eliminate other variables and build a model using only significant variables. The insignificant features which will be removes are:

```
['const',
 'PAT as % of net worth',
 'Contingent liabilities / Net worth (%)',
 'Debt to equity ratio (times)']
```

Logistic Regression Model 2:

Model:	Logit	Method:	MLE
Dependent Variable:	Default	Pseudo R-squared:	0.029
Date:	2024-09-27 18:15	AIC:	2987.8040
No. Observations:	2979	BIC:	3011.8014
Df Model:	3	Log-Likelihood:	-1489.9
Df Residuals:	2975	LL-Null:	-1534.3
Converged:	1.0000	LLR p-value:	4.1168e-19
No. Iterations:	6.0000	Scale:	1.0000

Figure 37 Logistic Model 2 performnace

- The R-squared value is 0.029 and AIC is 2987. This model is better than Logistic Model 1 as the RMSE and AIC values have decreased.

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
const	-1.3471	0.0462	-29.1482	0.0000	-1.4377	-1.2565
PAT as % of net worth	-0.3722	0.0779	-4.7775	0.0000	-0.5249	-0.2195
Contingent liabilities / Net worth (%)	0.1638	0.0760	2.1536	0.0313	0.0147	0.3128
Debt to equity ratio (times)	0.2925	0.0708	4.1311	0.0000	0.1537	0.4313

Figure 38 P Values Logistic Model 2 performnace

We will keep the remaining features and use Logistic Model 2 for checking predictive power using confusion matrix.

Model Evaluation on Train Data using significant variables:

Getting the predited values for trained data and building a classification matrix will help us evaluate the model performance using precision metric.

	actual	predicted_prob	predicted
1798	1	0.172291	0
3759	0	0.204053	0
2326	0	0.241843	0
3234	0	0.347511	0
3487	1	0.204659	0
1361	0	0.198862	0
1686	1	0.999245	1
1136	1	0.195407	0
3270	0	0.202762	0
3284	0	0.199943	0

Figure 39 Predicted values Logistic Model 2

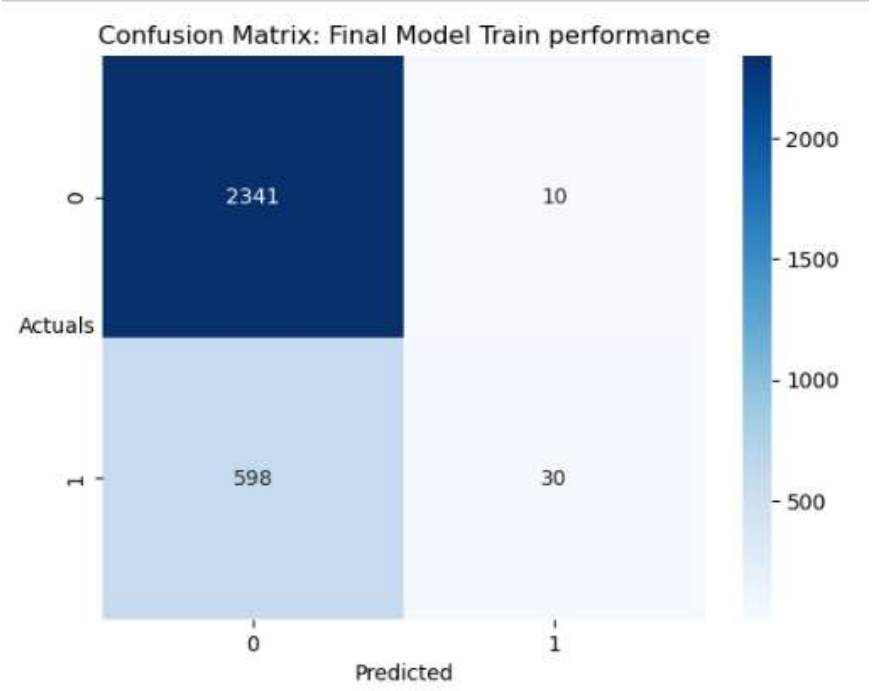


Figure 40 Confusion Matrix (Train) Logistic Model 2

- Here, 1 is Defaulter means Positive & 0 is non- defaulter means Negative
- TN: 2341, FP: 10, FN:598, TP:30.

- False Negatives are a matter of concern as they the the ones who are defaulters but have been predicted as Non-defaulter by the model.
- Precision is the metric to evaluate TN. We see that the precision is decent for predicting defaulters at 75%. Accuracy score is 79.6% which is also decent.

	precision	recall	f1-score	support
0	0.797	0.996	0.885	2351
1	0.750	0.048	0.090	628
accuracy			0.796	2979
macro avg	0.773	0.522	0.487	2979
weighted avg	0.787	0.796	0.717	2979

Figure 41 Classification Scores Logistic Model 2

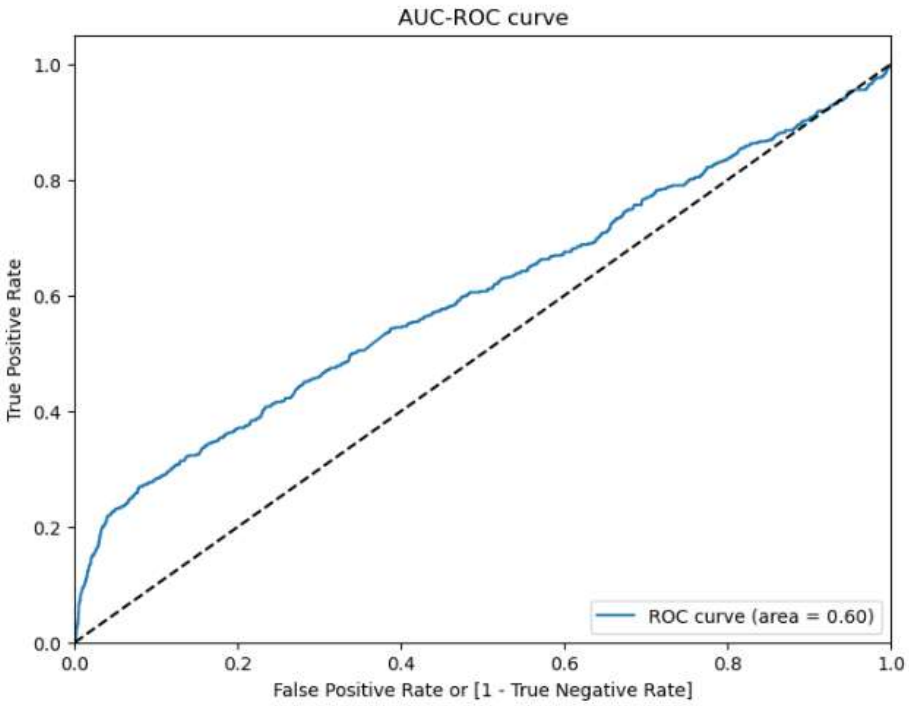


Figure 42 AUC-ROC curve(Train) Logistic Model 2

- The area of AUC-ROC curve is .60 which is not very good.

Model Evaluation on Train Data:

	actual	predicted_prob	predicted
963	0	0.245335	0
1920	1	0.178589	0
1743	0	0.172307	0
2209	0	0.203562	0
2509	0	0.182954	0
1363	1	0.207085	0
3473	0	0.193790	0
1801	1	0.206119	0
1615	1	0.187953	0
2088	0	0.193766	0

Figure 43 Predicted values Test set

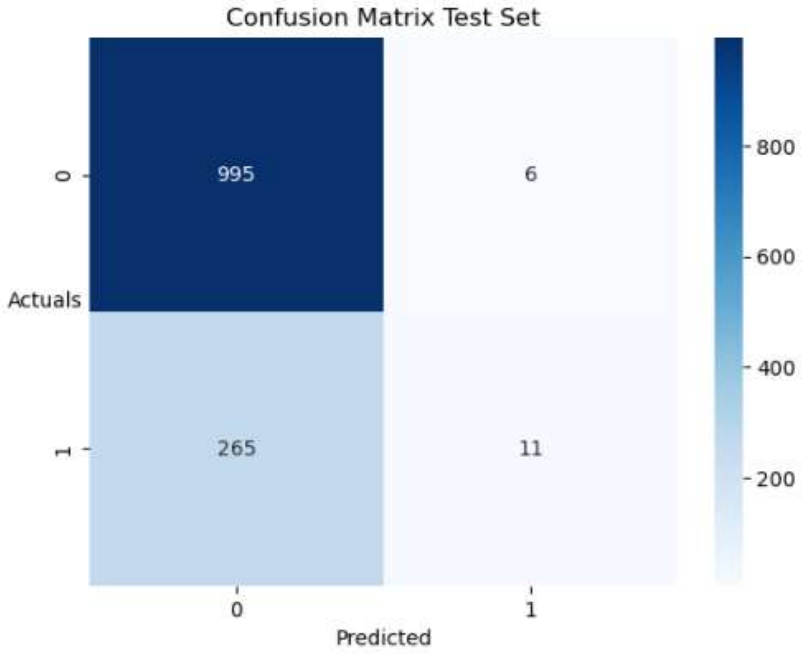


Figure 44 Confusion Matric Test Set

- TN: 995, FP: 6, FN:265, TP: 11.
- False Negatives are a matter of concern as they the companies who are defaulters but have been predicted as Non-defaulter by the model.

- Precision is the metric to evaluate TN. We see that the precision is decent for predicting defaulters at 64.7%. The model can predict non-defaulters with precision of 79%.
- The accuracy of test model is 78.8% which is very close to accuracy score of train set. Thus the model is free from overfitting/underfitting.

	precision	recall	f1-score	support
0	0.790	0.994	0.880	1001
1	0.647	0.040	0.075	276
accuracy			0.788	1277
macro avg	0.718	0.517	0.478	1277
weighted avg	0.759	0.788	0.706	1277

Figure 45 Classification score Test set

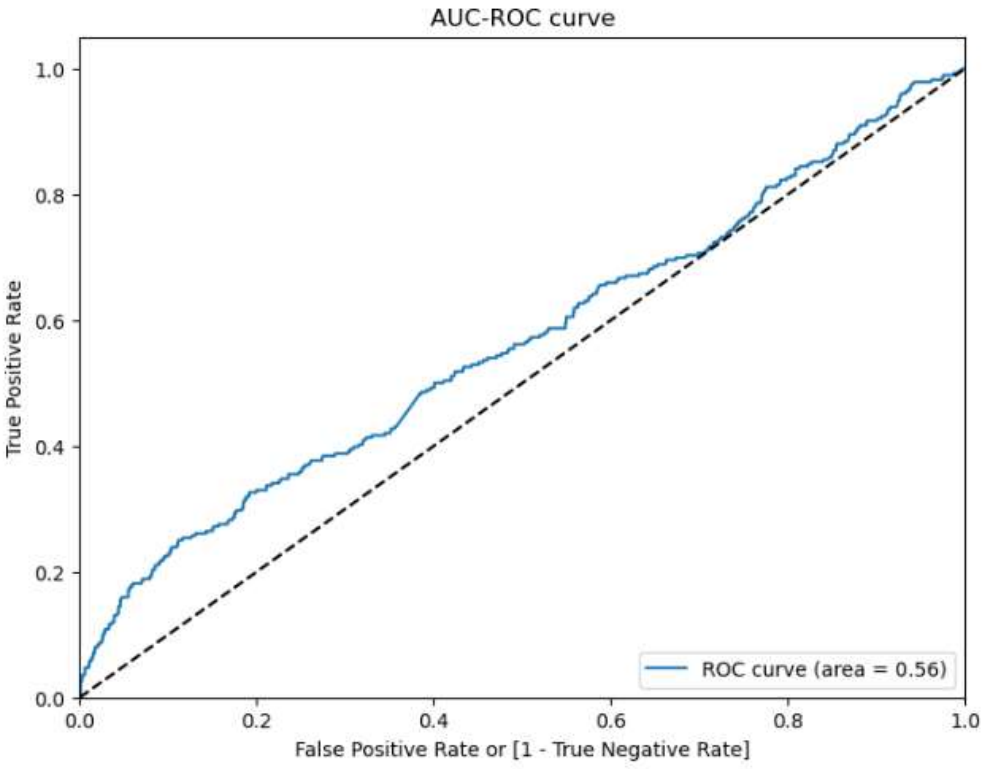


Figure 46 AUC ROC curve Test Set

- AUC-ROC curve area is 0.56% which is not good for the test model.

This model has low predictive power mainly due to lack of missing values. We will compare it with Random Forrest model and check the metrics.

Random Forest Model using only 10 classifiers:

RandomForestClassifier

```
RandomForestClassifier(random_state=0)
```

- We have built a Random Forest model using only 10 decision trees and can see that the accuracy scores are better than Logistic Regression model.

Train Accuracy score: 0.873

Test Accuracy score: 0.705

Building Random Forest Classifier Model with 100 trees:

RandomForestClassifier

```
RandomForestClassifier(random_state=0)
```

We have built a Random Forest model with 100 decision trees. There is no significant change in accuracy score with 10 and 100 classifiers and gives almost similar accuracy scores.

Model accuracy score for Train Set with 100 decision-trees : 0.8731

Model accuracy score for Test set with 100 decision-trees : 0.7048

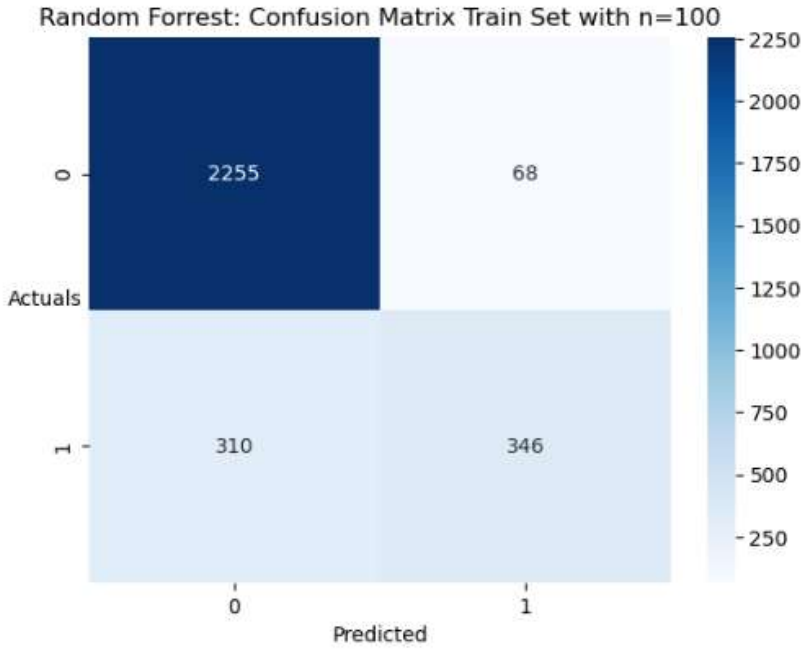


Figure 47 Confusion Matrix Random Forest (Train)

- TN: 2255, FP: 68, FN:310, TP: 346. The False positive values are a concern as in actual it is Defaulter but predicted as Non-Defaulter.

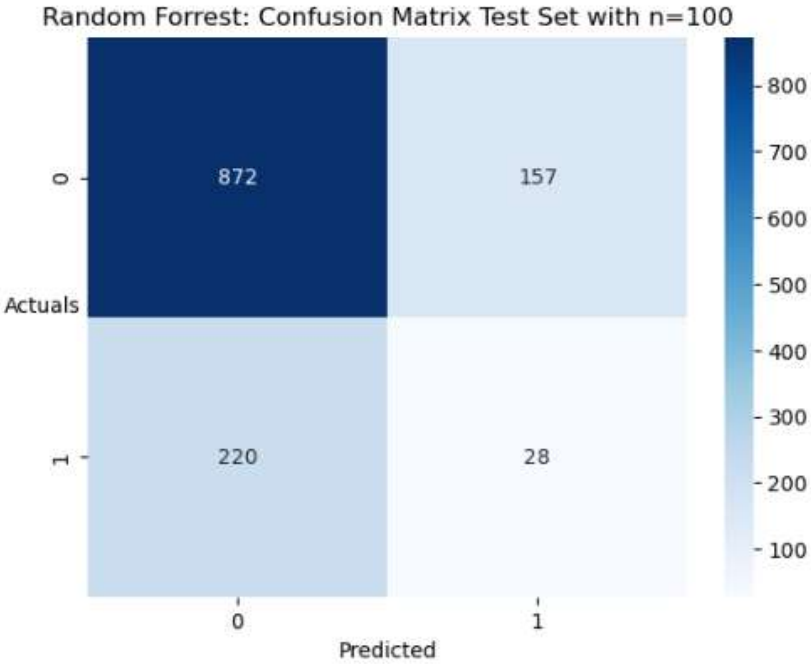


Figure 48 Confusion Matrix Random Forest (Test)

- TN: 872, FP: 157, FN:220, TP: 28. The False positive values are a concern as in actual it is Defaulter but predicted as Non-Defaulter.

Classification Metrics for Train Set:

	precision	recall	f1-score	support
0	0.88	0.97	0.92	2323
1	0.84	0.53	0.65	656
accuracy			0.87	2979
macro avg	0.86	0.75	0.78	2979
weighted avg	0.87	0.87	0.86	2979

Figure 49 Classification score Random Forest (Train)

Classification Metrics for Test Set:

	precision	recall	f1-score	support
0	0.80	0.85	0.82	1029
1	0.15	0.11	0.13	248
accuracy			0.70	1277
macro avg	0.47	0.48	0.48	1277
weighted avg	0.67	0.70	0.69	1277

Figure 50 Classification score Random Forest (Test)

- The precision value for predicting Defaulters for Test set is 84%. However, we have very poor predictability of only 15% in case of test set. We will perform Hyperparameter tuning to improvise and further compare the model.

PART A: Model Performance Improvement

- Dealing with multicollinearity using VIF - Identify optimal threshold for Logistic Regression using ROC curve - Hyperparameter Tuning for Random Forest - Model performance check across different metrics

Dealing with Multicolliniarity:

Calculate VIF Values:

	Feature	VIF
0	Total assets	inf
1	Net worth	4837.6
2	Total income	14030.5
3	Change in stock	4.0
4	Total expenses	6994.6
5	Profit after tax	1410.7
6	PBDITA	1174.9
7	PBT	1659.0
8	Cash profit	1123.9
9	PBDITA as % of total income	2.2
10	PBT as % of total income	163.0
11	PAT as % of total income	130.6
12	Cash profit as % of total income	34.1
13	PAT as % of net worth	1.1
14	Sales	7844.5
15	Income from fincial services	18.5
16	Other income	7.7
17	Total capital	42.8
18	Reserves and funds	1469.1
19	Borrowings	597.9
20	Current liabilities & provisions	1231.7
21	Deferred tax liability	83.3
22	Shareholders funds	9169.8
23	Cumulative retained profits	226.0
24	Capital employed	11255.6
25	TOL/TNW	14.1
26	Total term liabilities / tangible net worth	11.9
27	Contingent liabilities / Net worth (%)	1.2
28	Contingent liabilities	47.7
29	Net fixed assets	225.2
30	Investments	25.2
31	Current assets	152.2
32	Net working capital	14.3
33	Quick ratio (times)	54.8
34	Current ratio (times)	46.9
35	Debt to equity ratio (times)	4.7
36	Cash to current liabilities (times)	2.7
37	Cash to average cost of sales per day	2.0
38	Creditors turnover	1.0
39	Debtors turnover	1.0
40	Finished goods turnover	1.1
41	WIP turnover	1.1

42	Raw material turnover	1.0
43	Shares outstanding	5.2
44	Equity face value	2.0
45	EPS	1691936.6
46	Adjusted EPS	1691928.4
47	Total liabilities	inf
48	PE on BSE	1.0

Table 1 VIF Values

Based on the VIF Values we will drop highly dependent features having VIF>2 as follows:

['Net worth', 'Total income', 'Total expenses', 'Profit after tax', 'PBDITA', 'PBT', 'Cash profit', 'PBT as % of total income', 'PAT as % of total income', 'Sales', '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 fixed assets', 'Investments', 'Current assets', 'Net working capital', 'Quick ratio (times)', 'Current ratio (times)', 'EPS', 'Adjusted EPS', 'Total liabilities']

After removing the multicollinearity only we were able to build the Logistic Regression model which we will evaluate based on precision values. We have built Logistic Regression Model 1 and Logistic Regression model 2 after dealing with multicolliniearity.

Identifying optimal Threshold for Logistic Regression Model with ROC curve:

- Sensitivity is also known as True Positive Rate (TPR) and specificity is known as True Negative Rate (TNR). That is, select the cut-off probability for which (TPR + TNR – 1) is maximum.
- tpr and fpr values, which we have stored in variables tpr, fpr, respectively.
- The variable thresholds captures the corresponding cut-off probabilities. We can take difference of tpr and fpr and then sort the values in descending

order. The thresholds value, for which is maximum, should be the optimal cut-off.

	tpr	fpr	thresholds	diff
175	0.250000	0.111888	0.213858	0.138112
180	0.253623	0.115884	0.213063	0.137739
176	0.250000	0.112887	0.213716	0.137113
181	0.253623	0.116883	0.213057	0.136740
177	0.250000	0.113886	0.213422	0.136114

Figure 51 Youden's Index Thresholds

The cutoff probability is 0.2138. Therefore, we will create classification report and check the metrics again.

:

	actual	predicted_prob	predicted	predicted_new
221	0	0.203033	0	0
3427	0	0.200847	0	0
1493	0	0.205590	0	0
2839	0	0.222638	0	1
1531	0	0.132231	0	0
...	...	...	...	...
3750	0	0.198470	0	0
1551	0	0.196077	0	0
1381	0	0.180552	0	0
1867	1	0.196996	0	0
50	0	0.191487	0	0

1277 rows × 4 columns

Figure 52 Youden's Cutoff Predicted values

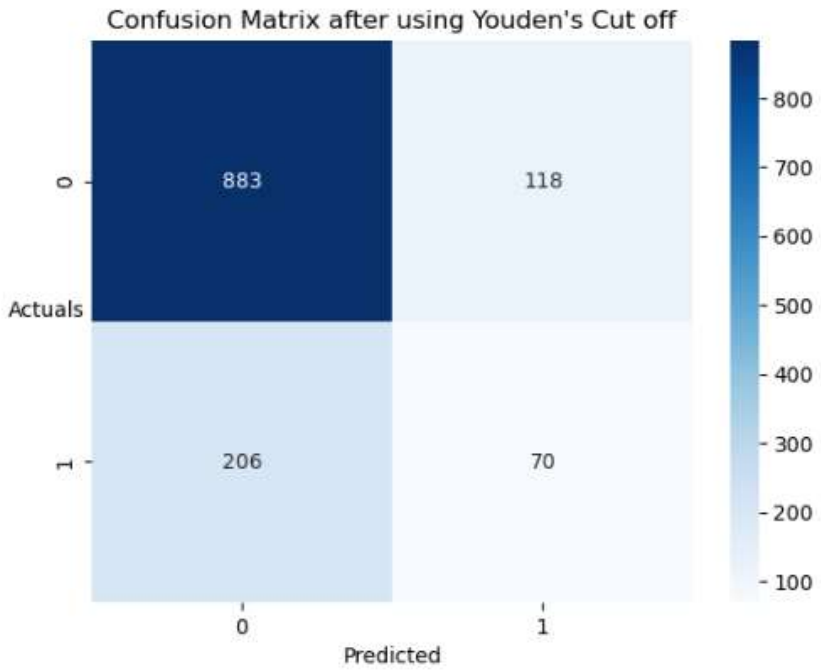


Figure 53 Youden's cutoff Confusion Matrix

- TN: 883, FP: 118, FN:206, TP: 70. The False positive values are a concern as in actual it is Defaulter but predicted as Non-Defaulter.

	precision	recall	f1-score	support
0	0.811	0.882	0.845	1001
1	0.372	0.254	0.302	276
accuracy			0.746	1277
macro avg	0.592	0.568	0.573	1277
weighted avg	0.716	0.746	0.728	1277

Figure 54 Classification Table Logistic Regression Youden's Cutoff Index

- The precision value after selecting optimal threshold limit from Youden's Cut Off Index has dropped to only 37.1% from 64.7% from Test performance of Logistic Regression Model 2.

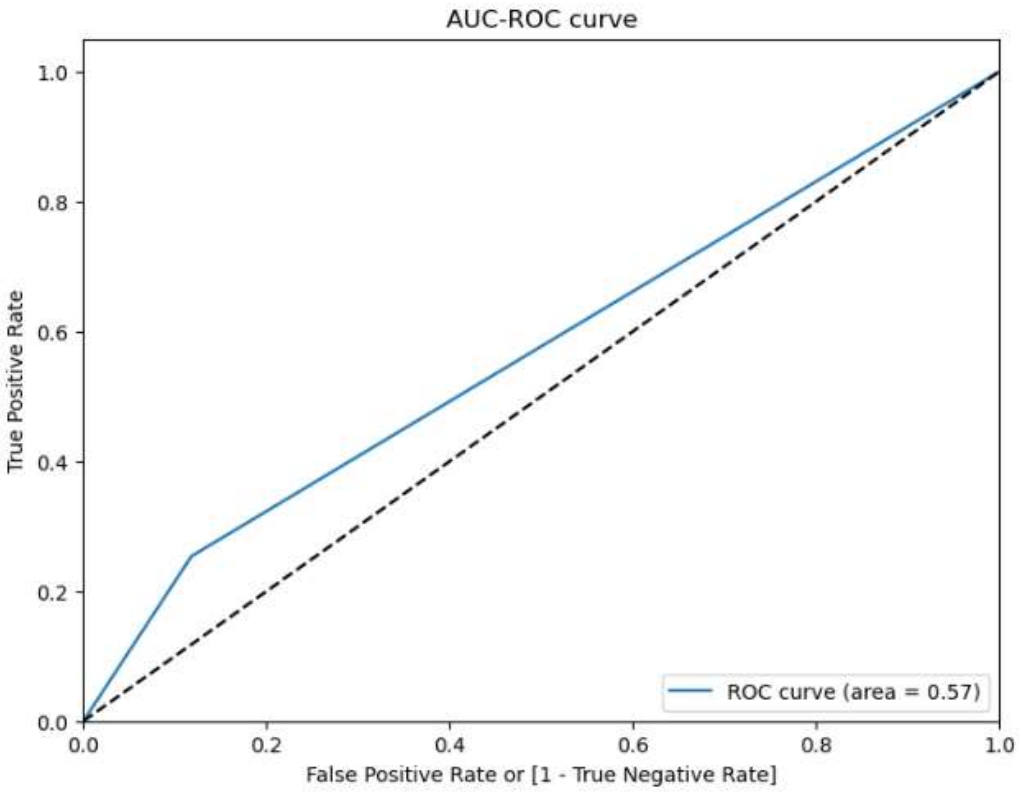


Figure 55 AUC-ROC Youden's Cutoff index Logistic Regression

- The AUC-ROC curve area is 0.57 indicating low predictive power of the model.

Hyperparameter Tuning for Random Forrest:

We will do feature selection to improve predictive power of Random Forrest. Lets calculate feature scores for each feature:

```
PAT as % of net worth                              0.033756
TOL/TNW                                            0.030531
Debt to equity ratio (times)                       0.029176
Net worth                                          0.026758
PBT as % of total income                           0.025824
Creditors turnover                                 0.025583
PAT as % of total income                           0.025372
Total term liabilities / tangible net worth        0.025048
Shareholders funds                                 0.024627
Quick ratio (times)                                0.023618
Reserves and funds                                 0.023598
Cumulative retained profits                        0.023003
Profit after tax                                   0.022934
PBDITA as % of total income                        0.022164
Cash profit as % of total income                   0.022068
Net working capital                                0.022012
Cash to average cost of sales per day              0.021955
```

Current ratio (times)	0.021920
Finished goods turnover	0.021819
Debtors turnover	0.021138
WIP turnover	0.021089
Cash profit	0.020970
EPS	0.020255
Current assets	0.020203
Total capital	0.020032
Contingent liabilities / Net worth (%)	0.019940
Current liabilities & provisions	0.019485
Net fixed assets	0.019422
Raw material turnover	0.019394
PBT	0.019153
Shares outstanding	0.019058
Cash to current liabilities (times)	0.018690
Capital employed	0.018612
Change in stock	0.018331
Adjusted EPS	0.018233
Borrowings	0.017858
Contingent liabilities	0.017782
PBDITA	0.017684
Total liabilities	0.017312
Income from fincial services	0.016575
Total assets	0.016573
Deferred tax liability	0.016472
Total income	0.016004
Other income	0.015315
Investments	0.015308
Sales	0.015239
Total expenses	0.015186
PE on BSE	0.012447
Equity face value	0.004479

dtype: float64

Table 2 Feature scores

After seeing the feature scores of all variables we will eliminate ones with lower scores and will build a model with better predictive power.

We will sort the values based on their scores.

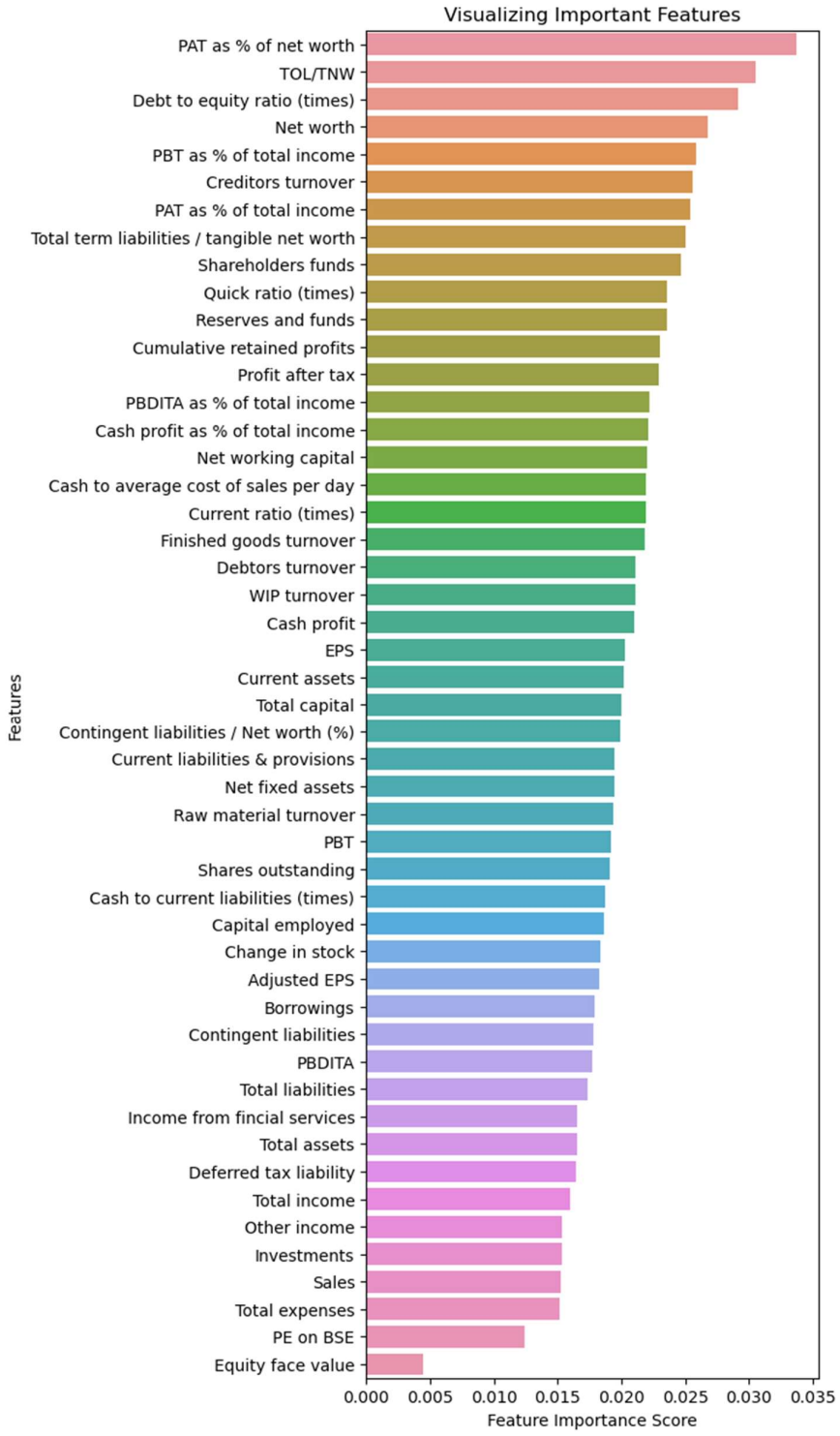


Figure 56 Feature scoes

PAT as %age of net worth has highest feature score. We will drop the last columns and again build the model. Dropping features with least scores----'Default', 'Networth Next Year','Num', 'Equity face value'

Model accuracy score for train data with variable removed : 0.8731

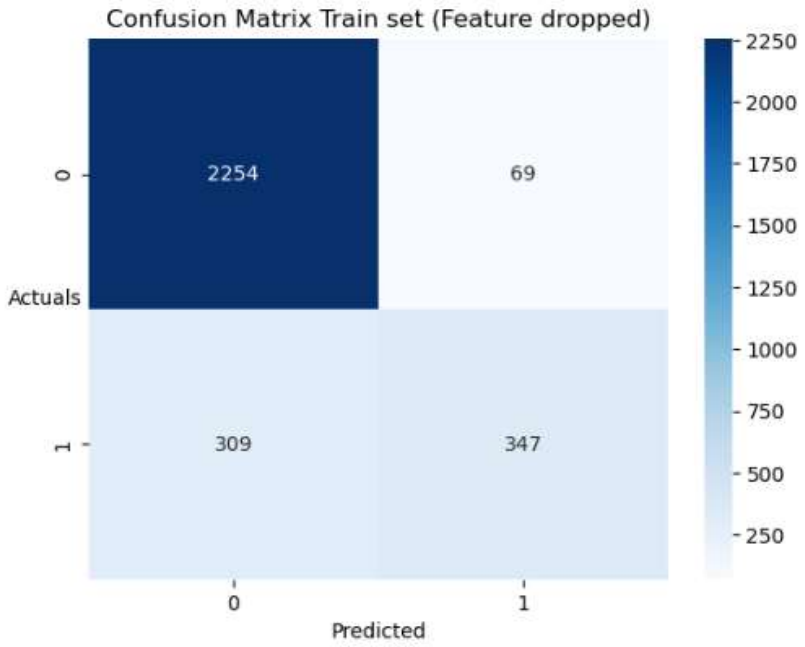


Figure 57 Confusion Matrix Random forest Tain with feature drop

- TN: 2254, FP: 69, FN:309, TP: 347. The False positive values are a concern as in actual it is Defaulter but predicted as Non-Defaulter.

	precision	recall	f1-score	support
0	0.88	0.97	0.92	2323
1	0.83	0.53	0.65	656
accuracy			0.87	2979
macro avg	0.86	0.75	0.79	2979
weighted avg	0.87	0.87	0.86	2979

Figure 58 Classification Table Random forest Tain with feature drop

- The precision value for predicting Defaulter companies is 83% with model accuracy of 87%. This train model has good performance.

Model accuracy score for test set with low score variable removed : 0.7048

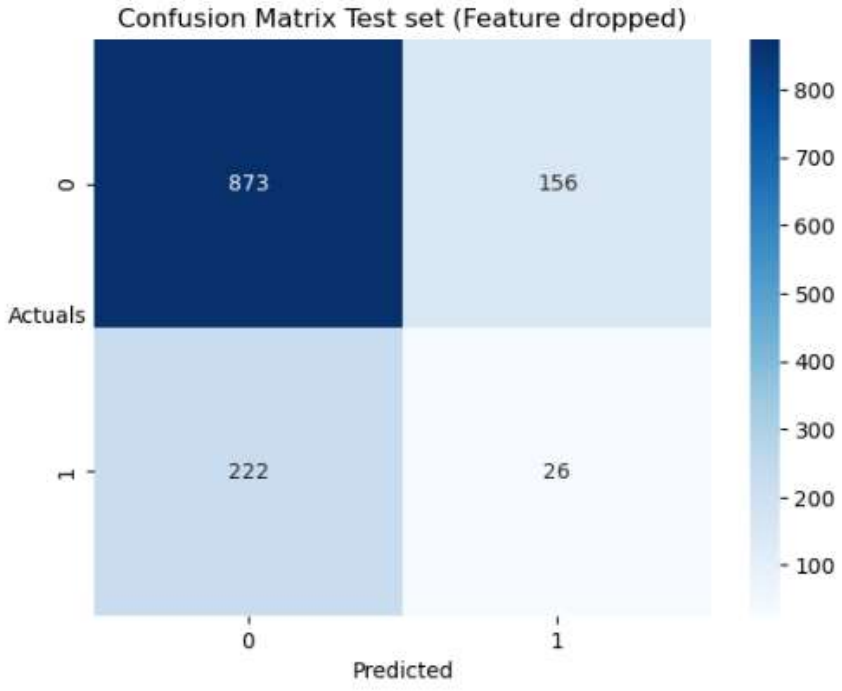


Figure 59 Confusion Matrix Random forest Test with feature drop

- TN: 873, FP: 156, FN:222, TP: 26. The False positive values are a concern as in actual it is Defaulter but predicted as Non-Defaulter.

	precision	recall	f1-score	support
0	0.80	0.85	0.82	1029
1	0.14	0.10	0.12	248
accuracy			0.70	1277
macro avg	0.47	0.48	0.47	1277
weighted avg	0.67	0.70	0.69	1277

Figure 60 Classification Table Random forest Test with feature drop

The precision value on Test data set for predicting Defaulter companies is very dismal at 14%. This predictive power is not sufficient for model building even after removing the low score features.

Hyper parameter tuning using Grid search CV:

Using Gridsearch CV for hyperparameter tuning to find the best parameters for Random Forest model building.

Best Parameters:

```
{'bootstrap': True,
 'max_depth': 2,
 'max_features': 0.6,
 'max_samples': 0.75,
 'n_estimators': 100}
```

Grid Best Score: 78.88% is the best score for the best parameters identified.

Model performance evaluation using best parameters:

Train Performance:

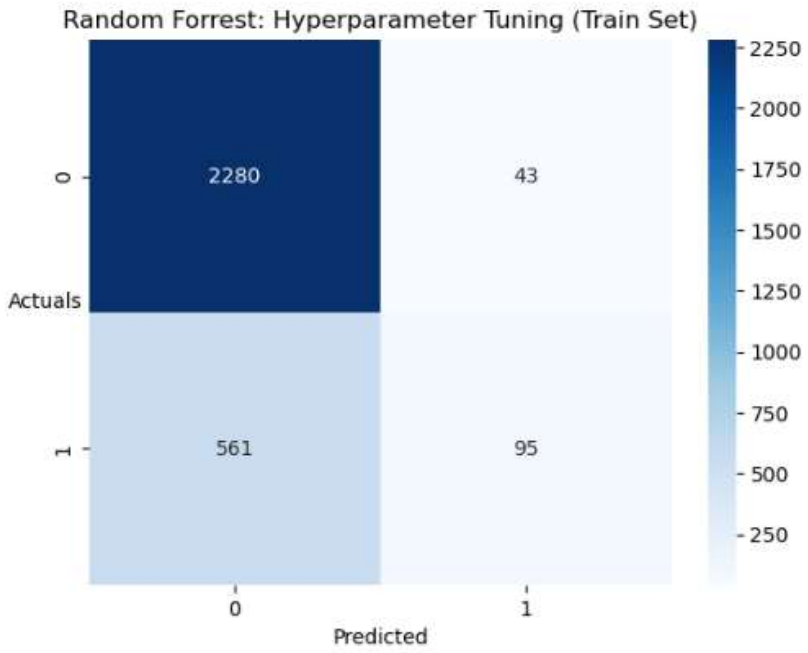


Figure 61 Confusion matrix: Random Forest Hyper tuning Train

- TN: 2280, FP: 43, FN:561, TP: 95. The False positive values are a concern as in actual it is Defaulter but predicted as Non-Defaulter.

	precision	recall	f1-score	support
0	0.80	0.98	0.88	2323
1	0.69	0.14	0.24	656
accuracy			0.80	2979
macro avg	0.75	0.56	0.56	2979
weighted avg	0.78	0.80	0.74	2979

Figure 62 Classification Table: Random Forest Hyper tuning Train

- The precision score for predicting defaulter is 69% using the best parameters for Random Forrest model. The accuracy is 80% for the model which is also good for the model train performance. Let see the test performance after hyper parameter tuning.

Test Performance:

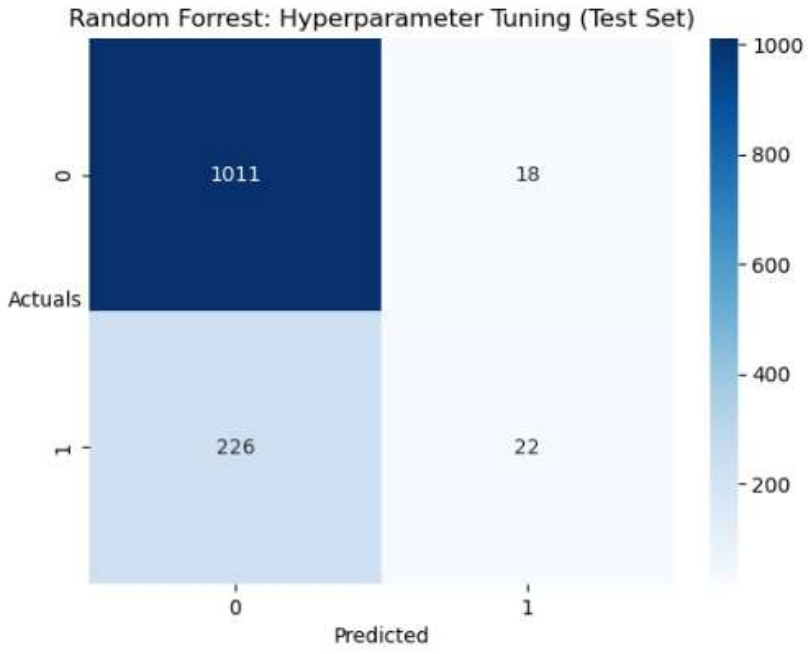


Figure 63 Confusion matrix: Random Forest Hyper tuning Test

- TN: 1011, FP: 18, FN:226, TP: 22. The False positive values are a concern as in actual it is Defaulter but predicted as Non-Defaulter.

	precision	recall	f1-score	support
0	0.82	0.98	0.89	1029
1	0.55	0.09	0.15	248
accuracy			0.81	1277
macro avg	0.68	0.54	0.52	1277
weighted avg	0.77	0.81	0.75	1277

Figure 64 Classification Table: Random Forest Hyper tuning Test

- The model test performance for train set after hyperparameter tuning gives 55% precision in classifying the defaulters.

PART A: Model Performance Comparison and Final Model Selection

- Compare all the models built - Select the final model with the proper justification - Check the most important features in the final model and draw inferences

Lets see the precision scores for all models built.

S. I.	Model Name	Data Set	Precision (0)	Precision (1)	Accuracy Score
1	Logistic Regression Model Final	Train	79.7%	75.0%	79.6%
1	Logistic Regression Model Final	Test	79.0%	64.7%	78.8%
2	Logistic Regression Model (Youden's Index)	Test	81.1%	37.2%	74.6%
3	Random Forrest (n=100)	Train	88.0%	84.0%	87.0%
3	Random Forrest (n=100)	Test	80.0%	15.0%	70.0%
4	Random Forrest (n=100) with Feature drop	Train	88.0%	83.0%	87.0%
4	Random Forrest (n=100) with Feature drop	Test	80.0%	14.0%	70.0%
5	Random Forrest (n=100) with Best Params	Train	80.0%	69.0%	80.0%
5	Random Forrest (n=100) with Best Params	Test	82.0%	55.0%	81.0%

Table 3 Model Perfomance Comparision

- The Train performance is best for the Random Forest model with highest precision for predicting Defaulters at 84%. However, the predictive ability has sharply declined for test set to 15% which is very poor performance for

predicting the Defaulters. This model cannot be used as final model among all models built and validated.

- Logistic Regression Model which we built after dropping the features with high dependency and VIF values seems to have better performance with no overfitting. The model Train precision score is 75% for predicting Defaulter companies. The model test precision score is 64.7% which is decent enough for predicting the defaulters better than other models.
- Logistic Regression Final Model is the best model among all validated models for prediting the Deaulter company with a precision of 64.7% and accuracy of 78.8%.

Most Important features in Final Model:

Logitic Regression Final Model with best features are:

1. PAT as % of net worth
2. Contingent liabilities / Net worth (%)
3. Debt to equity ratio (times)

Inferences:

1. PAT as % of net worth describes ratio of PAT to Total Income is often used to assess a company's efficiency in converting its income into profit after accounting for taxes. It provides insight into the company's profitability and overall operational performance. A higher ratio indicates better profitability, while a lower ratio may suggest higher costs or lower efficiency in generating profit from income. The mean is -20.03.
2. Contingent liabilities / Net worth (%) helps assess the risk profile of a company by comparing its potential liabilities that may arise in the future (contingent liabilities) to its shareholders' equity (net worth). The mean value is 55.71.
3. Debt to equity ratio (times) indicates the relative proportion of a company's debt to its shareholders' equity. It is used to assess a company’s financial leverage and risk. The mean is 2.87.

PART A: Actionable Insights & Recommendations

- Actionable insights and recommendations

1. The venture Capital Firms should focus on 3 key variables for predicting the performance of organizations which is PAT as % of net worth, Contingent liabilities / Net worth (%) and Debt to equity ratio (times).
2. For predicting the Defaulting companies the organization needs to make sure that the data collection is done properly. When data collection is not done properly the model might not be having high predictive power.
3. Need to focus on parameters which are more of independent in nature as well like Manpower costs, total manpower and Cash burn rate to get a better picture of spending.
4. We see that the mean of Debt to Equity Ratio is 2.87. This infers that most of the companies have High debt compared to Equity. A healthy debt to equity ratio suggests that company is not raising debt funds for operations but for creating assets which will ensure that company is not a defaulter.

Recommendation:

- Predicting the defaulter is a crucial step to ensure that the VC firm is not risk averse by providing handholding to companies who are likely to default. However, the information for predicting needs to be more accurate and reliable. Build better data collection pipelines to ensure better prediction ability and avoid financial risks by reducing the credit to organizations who have low debt to Equity ratio and have high volume of assets.

PART-B: PROBLEM STATEMENT

Context

Investors face market risk, arising from asset price fluctuations due to economic events, geopolitical developments, and investor sentiment changes. Understanding and analyzing this risk is crucial for informed decision-making and optimizing investment strategies.

Objective

The objective of this analysis is to conduct Market Risk Analysis on a portfolio of Indian stocks using Python. It uses historical stock price data to understand market volatility and riskiness. Using statistical measures like mean and standard deviation, investors gain a deeper understanding of individual stocks' performance and portfolio variability.

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

Risk Assessment: Analyze the historical volatility of individual stocks and the overall portfolio.

Portfolio Optimization: Use Market Risk Analysis insights to enhance risk-adjusted returns.

Performance Evaluation: Assess portfolio management strategies' effectiveness in mitigating market risk.

Portfolio Performance Monitoring: Monitor portfolio performance over time and adjust as market conditions and risk preferences change.

Data Description:

	Date	ITC Limited	Bharti Airtel	Tata Motors	DLF Limited	Yes Bank
0	28-03-2016	217	316	386	114	173
1	04-04-2016	218	302	386	121	171
2	11-04-2016	215	308	374	120	171
3	18-04-2016	223	320	408	122	172
4	25-04-2016	214	319	418	122	175

Figure 65 Part B Head

Shape: (418, 6)

- The number of rows (observations) is 418
- The number of columns (variables) is 6

PART B: Stock Price Graph Analysis

- Draw a Stock Price Graph (Stock Price vs Time) for the given stocks - Write observations

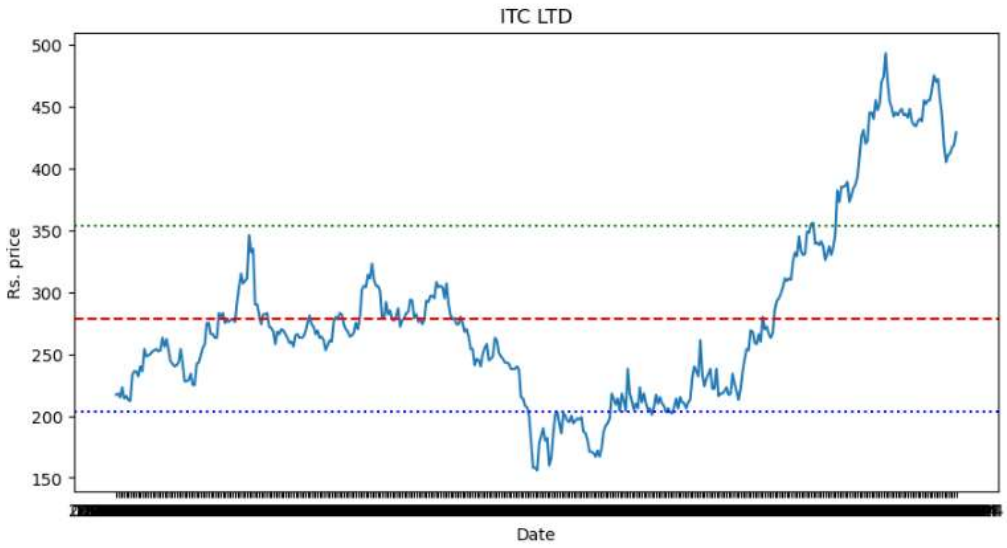


Figure 66 ITC Ltd Stock Price

- The ITC LTD stock has given a decent growth during the period of 28-03-2016 to 25-03-2024 with a return of Rs. 212 per share.
- Last trading price has been between Rs.400-450. The mean price has been between Rs. 250-300.

- The upper standard deviation is near Rs. 350 and Lower standard deviation is close to Rs. 200.
- The stock is a good performer in recent years.

Figure 67 Bharti Airtel Stock Price

- The Bharti Airtel stock has given a decent growth during the period of 28-03-2016 to 25-03-2024 with a return of Rs. 920 per share.
- Last trading price was near Rs. 1200. The mean price has been between Rs. 500 to 600.
- The upper standard deviation is near Rs. 800 and Lower standard deviation is close to Rs. 300.
- The stock is a good and consistent performer over the period.

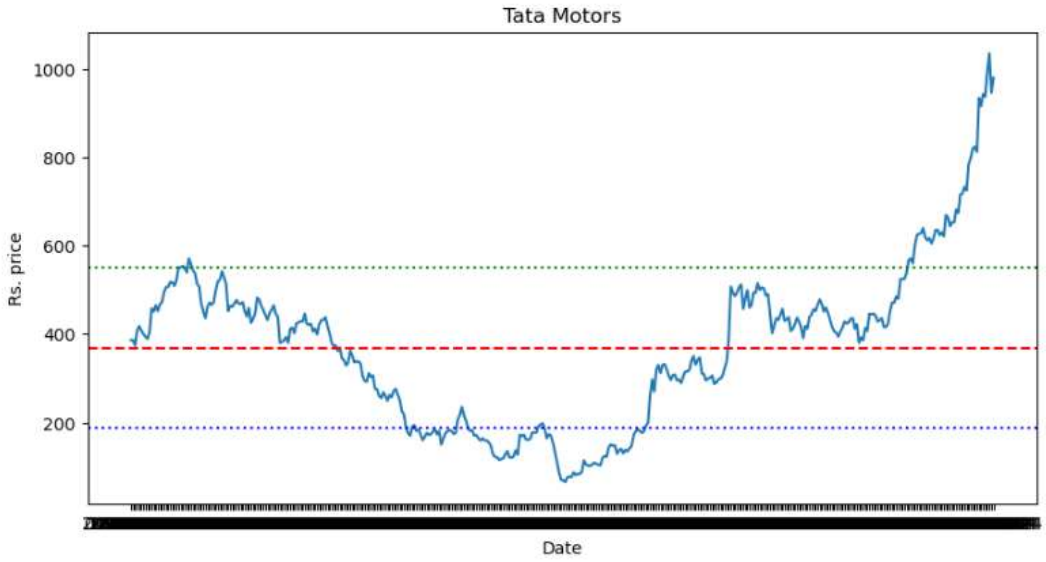


Figure 68 Tata Motors Stock Price

- The Tata Motors stock has given a decent growth during the period of 28-03-2016 to 25-03-2024 with a return of Rs. 594 per share.
- Last trading price was near Rs. 1000. The mean price has been between Rs. 300 to 400.
- The upper standard deviation is near Rs. 600 and Lower standard deviation is close to Rs. 200.
- The stock is a good performer in recent years.

Figure 69 DLF Ltd Stock Price

- The DLF Ltd stock has given a decent growth during the period of 28-03-2016 to 25-03-2024 with a return of Rs. 752 per share.
- Last trading price was near Rs. 800. The mean price has been between Rs. 200 to 300.
- The upper standard deviation is near Rs. 450 and Lower standard deviation is close to Rs. 125.
- The stock is a high performer in recent years.

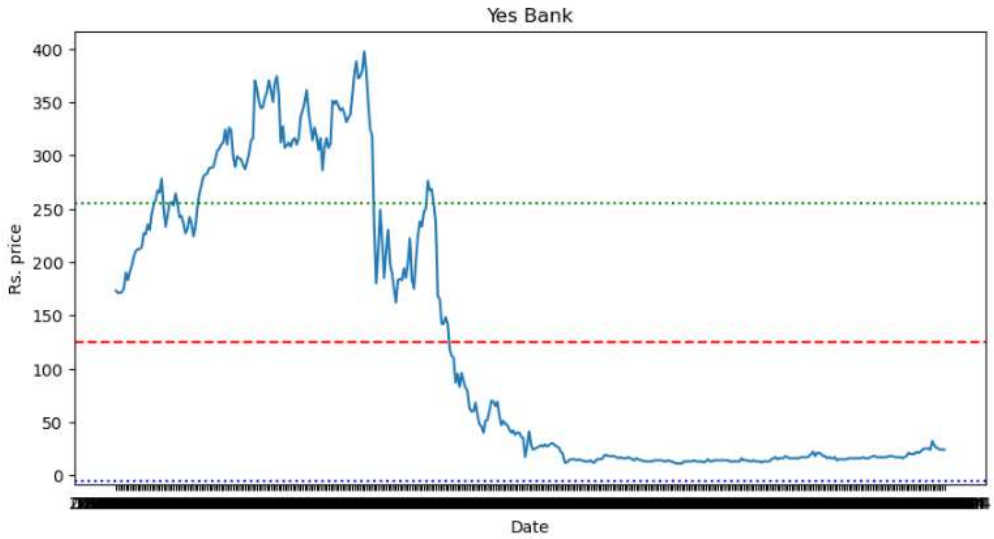


Figure 70 Yes Bank Stock Price

- The Yes Bank stock has given a decent growth during the period of 28-03-2016 to 25-03-2024 with a return of Rs. -149 per share.
- Last trading price was near Rs. 25. The mean price has been between Rs. 100 to 150.
- The upper standard deviation is near Rs. 250 and Lower standard deviation is close to Rs.0.
- The stock is a poor performer in recent years.

PART B: Stock Returns Calculation and Analysis

- Calculate Returns for all stocks - Calculate the Mean and Standard Deviation for the returns of all stocks - Draw a plot of Mean vs Standard Deviation for all stock returns - Write observations and inferences

The Table below shows the daily returns value for each stock which is the difference of price on each subsequent day.

	ITC Limited	Bharti Airtel	Tata Motors	DLF Limited	Yes Bank
0	NaN	NaN	NaN	NaN	NaN
1	1.0	-14.0	0.0	7.0	-2.0
2	-3.0	6.0	-12.0	-1.0	0.0
3	8.0	12.0	34.0	2.0	1.0
4	-9.0	-1.0	10.0	0.0	3.0
...	...	...	...	...	...
413	6.0	-3.0	-6.0	34.0	-2.0
414	1.0	14.0	56.0	27.0	-1.0
415	5.0	54.0	42.0	3.0	-1.0
416	2.0	39.0	-89.0	-102.0	0.0
417	10.0	11.0	34.0	40.0	0.0

418 rows × 5 columns

Figure 71 Returns stock

The data available is for the period 28-03-2016 to 25-03-2024. The Net returns over the period is:

```
{'ITC Limited': 212,
'Bharti Airtel': 920,
 'Tata Motors': 594,
 'DLF Limited': 752,
  'Yes Bank': -149}
```

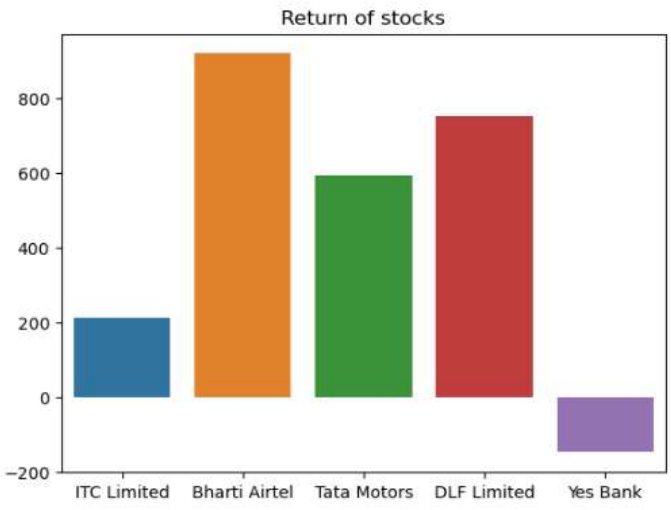


Figure 72 Net Returns over period / share

- The highest returns is given by Bharti Airtel and Lowest is by Yes Bank.

Lets calculate the avg. change in stock price for each stock.

Mean:

ITC Limited	0.508393
Bharti Airtel	2.206235
Tata Motors	1.424460
DLF Limited	1.803357
Yes Bank	-0.357314

Table 4 Mean stock prices

- The mean price change for ITC is Rs. 0.5 per day
- The mean price change for Airtel is Rs. 2.2 per day
- The mean price change for Tata Motors is Rs. 1.42 per day
- The mean price change for DLF Ltd is Rs. 1.80 per day
- The mean price change for Yes Bank is Rs. -0.35 per day

Standard Deviation:

ITC Limited	9.309866
Bharti Airtel	19.587959
Tata Motors	19.428609
DLF Limited	16.087044
Yes Bank	11.361389
dtype:	float64

Table 5 Standard Deviation Stock Prices

- The Standard Deviation for ITC is Rs. 9.3 per day
- The Standard Deviation for Airtel is Rs. 19.58 per day

- The Standard Deviation for Tata Motors is Rs. 19.42 per day
- The Standard Deviation for DLF Ltd is Rs. 16.04 per day
- The Standard Deviation for Yes Bank is Rs. 11.36 per day

The Average Price change and Valatality is shown below for each stock:

	Average	Volatility
ITC Limited	0.508393	9.309866
Bharti Airtel	2.206235	19.587959
Tata Motors	1.424460	19.428609
DLF Limited	1.803357	16.087044
Yes Bank	-0.357314	11.361389

Figure 73 Avg and Volatility

We will make a graph for mean of all stocks:

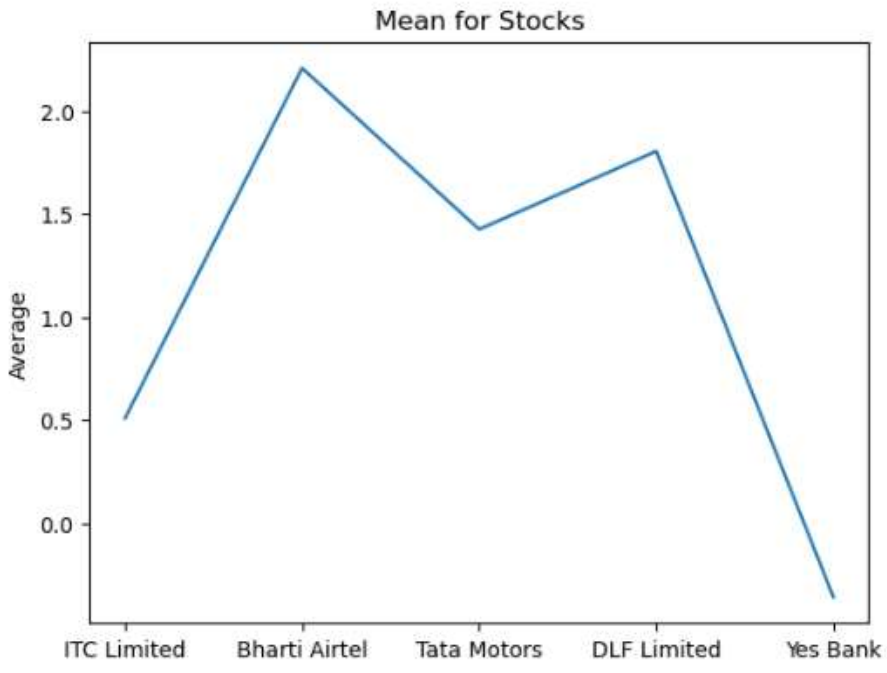


Figure 74 Mean price change

1. The highest mean value is for Bharti Airtel and the lowest is for Yes Bank.

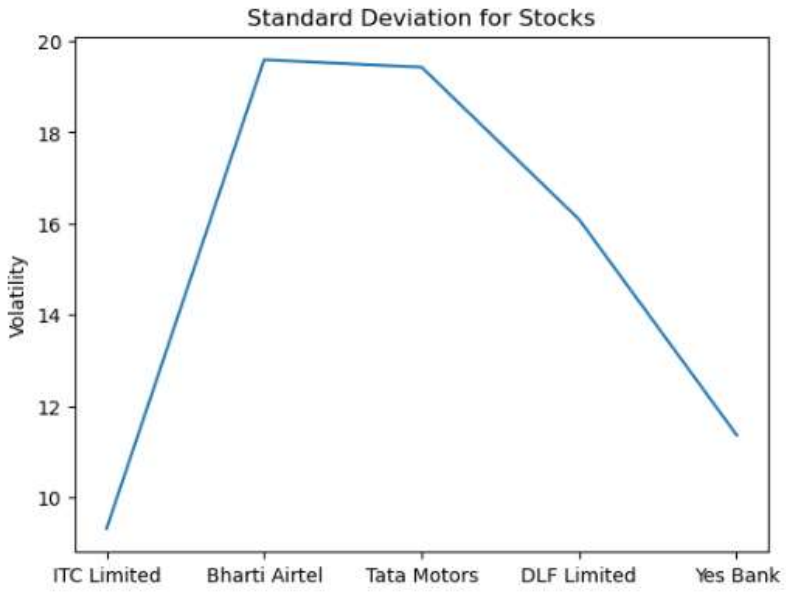


Figure 75 Volatility price change

2. The standard deviation plot for all stocks shows that the highest standard deviation on daily price change is for Bharti Airtel and the lowest volatility is for ITC Ltd.

PART B: Actionable Insights & Recommendations

- Actionable insights and recommendations

1. The best stock to invest is Bharti Airtel as it has given highest returns over a period of time. Bharti Airtel stock has also high volatility of 19.58 which also makes it high risk stock. Therefore, investments should be done carefully in such stock.
2. DLF Ltd has also shown high returns of 752 over a period. However, the volatility of 16.08 is not very high. This is a moderate risk stock and with good returns potential.
3. ITC Limited is a safe stock with Average returns of Rs. 212 over a period. Due to its low volatility it makes is low risk and low return stock.
4. Yes Bank is a highly risky stock with poor returns showing degrowth trends. This stock has high potential to grow based on market demands. However this stock must be avoided due to its low return and moderate volatility.

Recommendation:

- Depending on the high risk potential of an investor may invest for best returns in Bharti Airtel and DLF Ltd.
- Inventing in Yes Bank must be avoided as the stock has not performed well.
- ITC Limited stock is a good long term investment suggestion due to low volatility and moderate returns to balance the overall portfolio.

THE END