PROJECT REPORT
FINANCE AND RISK ANALYTICS

By: Parthasarathi Behura

PART: A

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.

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

A.1 Problem definition- Check shape, Data types, and statistical summary: -

Shape of the data-frame =>

```
Number of Rows is = 4256
Number of Columns is = 51
```

```
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_perc_of_total_income               4177 non-null   float64
 12  PBT_as_perc_of_total_income                  4177 non-null   float64
 13  PAT_as_perc_of_total_income                  4177 non-null   float64
 14  Cash_profit_as_perc_of_total_income          4177 non-null   float64
 15  PAT_as_perc_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_to_TNW                                   4256 non-null   float64
 28  Total_term_liabilities__to__tangible_net_worth  4256 non-null   float64
 29  Contingent_liabilities__to__Net_worth_perc   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
```

	count	mean	std	min	25%	50%	75%	max
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_perc_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_perc_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_perc_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_perc_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_perc_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

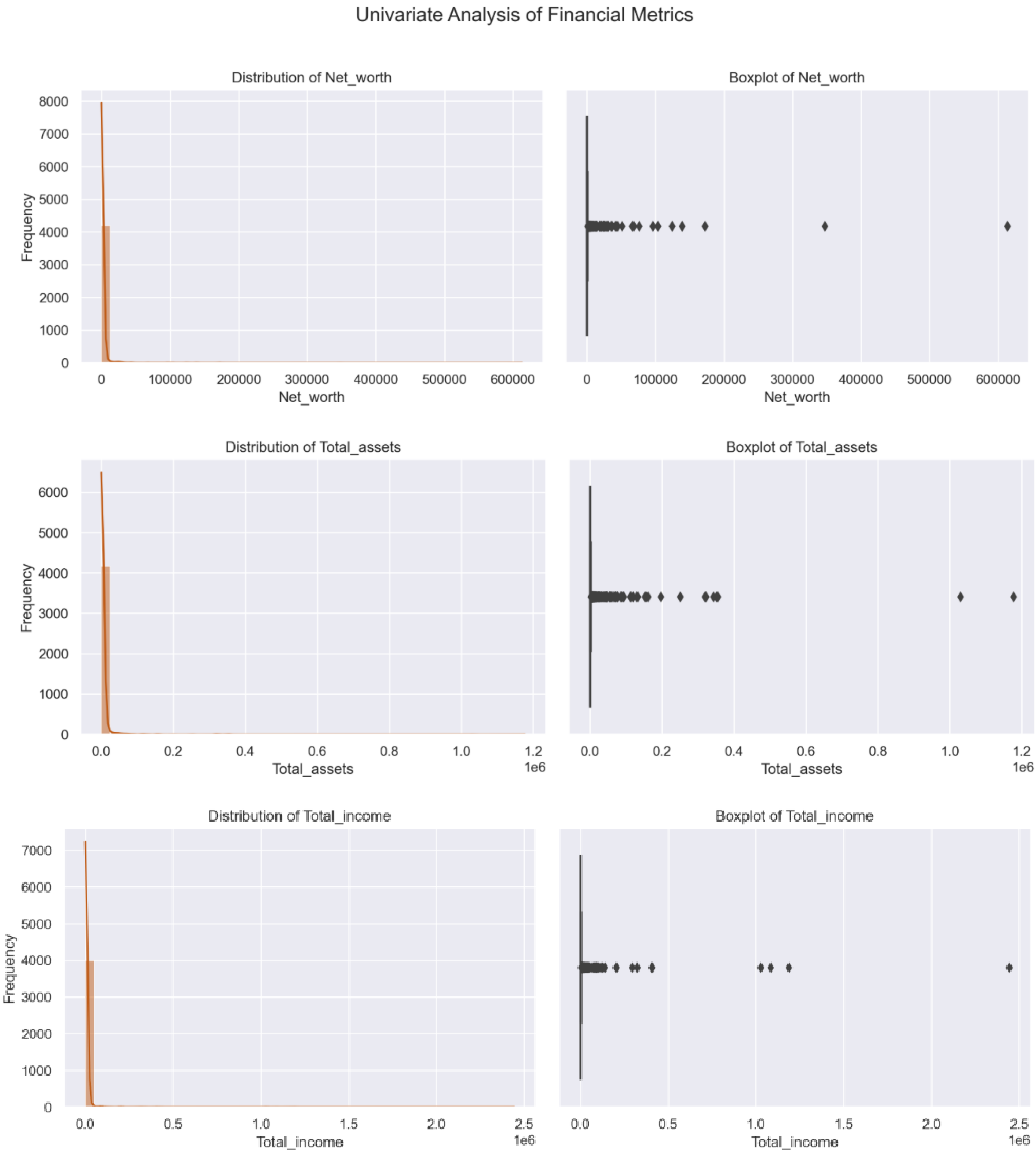
Reserves_and_funds	4158.0	1.210562e+03	1.281623e+04	-6.525900e+03	5.300	55.150	2.825250e+02	6.251378e+05
Borrowings	3825.0	1.176248e+03	8.581249e+03	1.000000e-01	24.400	99.800	3.583000e+02	2.782573e+05
Deferred_tax_liability	2887.0	2.344951e+02	2.106253e+03	1.000000e-01	3.200	13.500	5.130000e+01	7.279660e+04
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_to_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__to__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__to__Net_worth_perc	4256.0	5.570750e+01	3.691657e+02	0.000000e+00	0.000	5.360	3.101250e+01	1.470427e+04
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

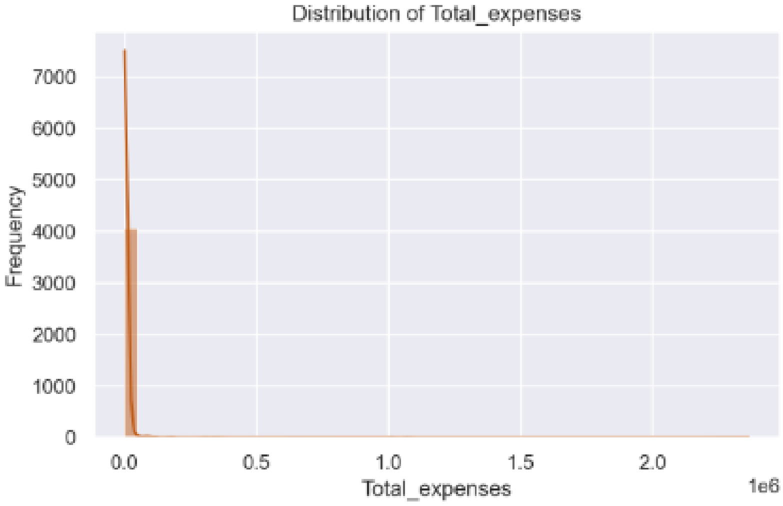
Binary Target Variable Creation:

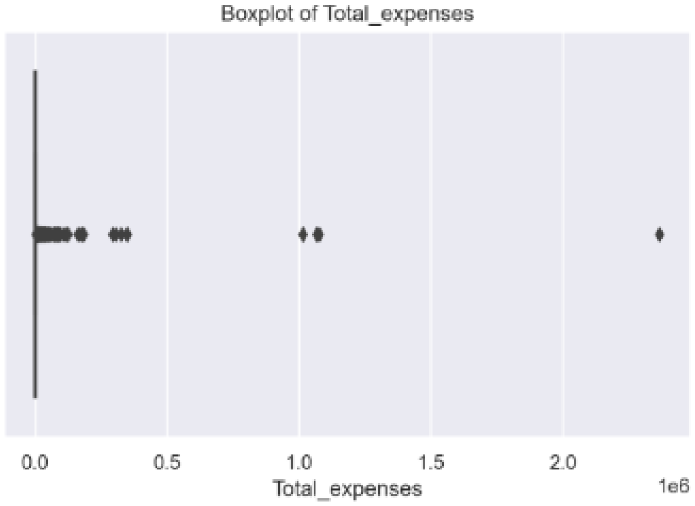
Default Variable: Create a binary target variable "Default" derived from "Networth_Next_Year”:

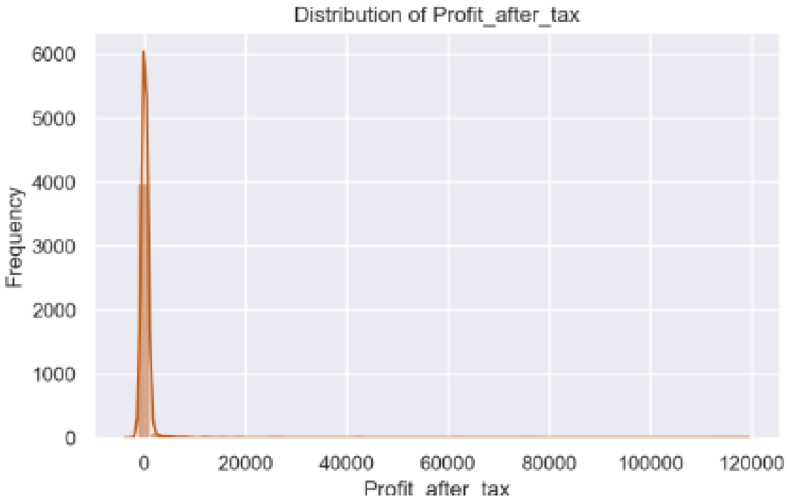
- Not Likely to Default: If "Networth_Next_Year" > 0, label as 0.
- Likely to Default: If "Networth_Next_Year" <= 0, label as 1.

<u>A.2</u> Univariate analysis: - (appropriate visualizations to identify the patterns and insight): -



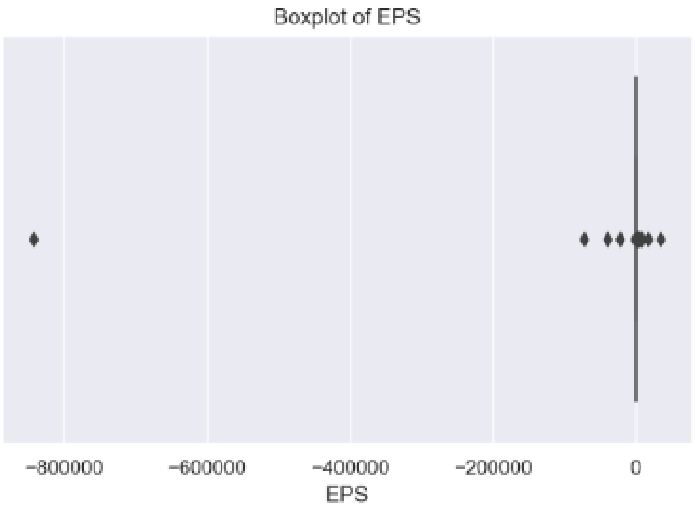

Distribution of Total_expenses
Frequency
Total_expenses
1e6

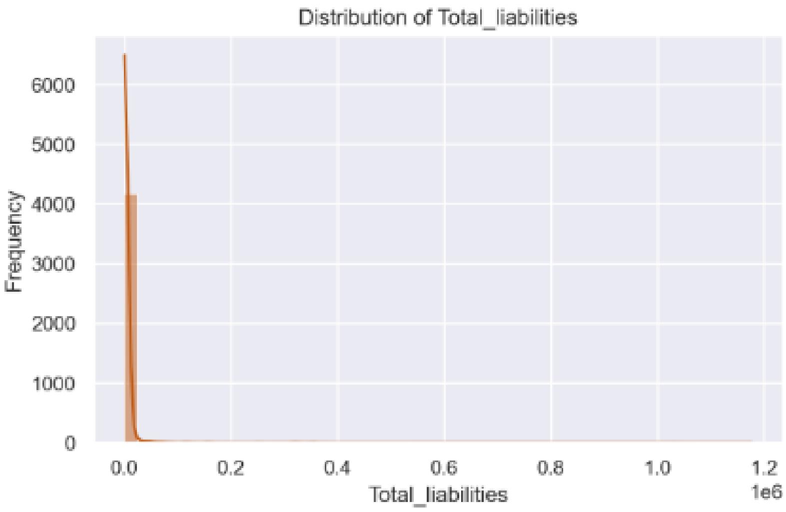

Boxplot of Total_expenses
Total_expenses
1e6

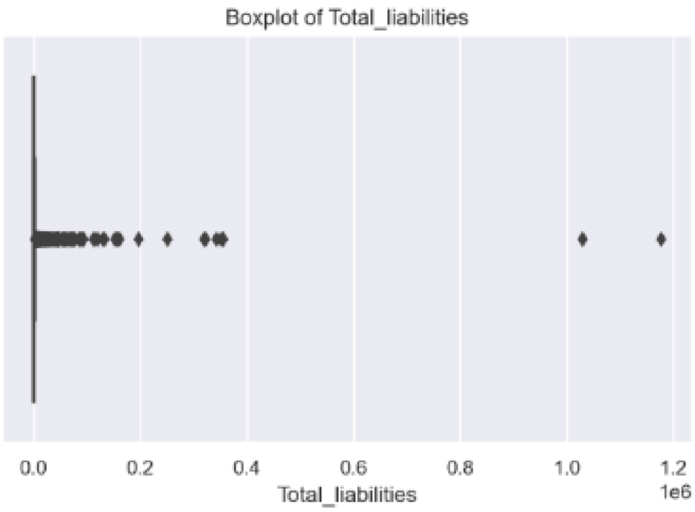

Distribution of Profit_after_tax
Frequency
Profit_after_tax


Boxplot of Profit_after_tax
Profit_after_tax


Distribution of EPS
Frequency
EPS


Boxplot of EPS
EPS

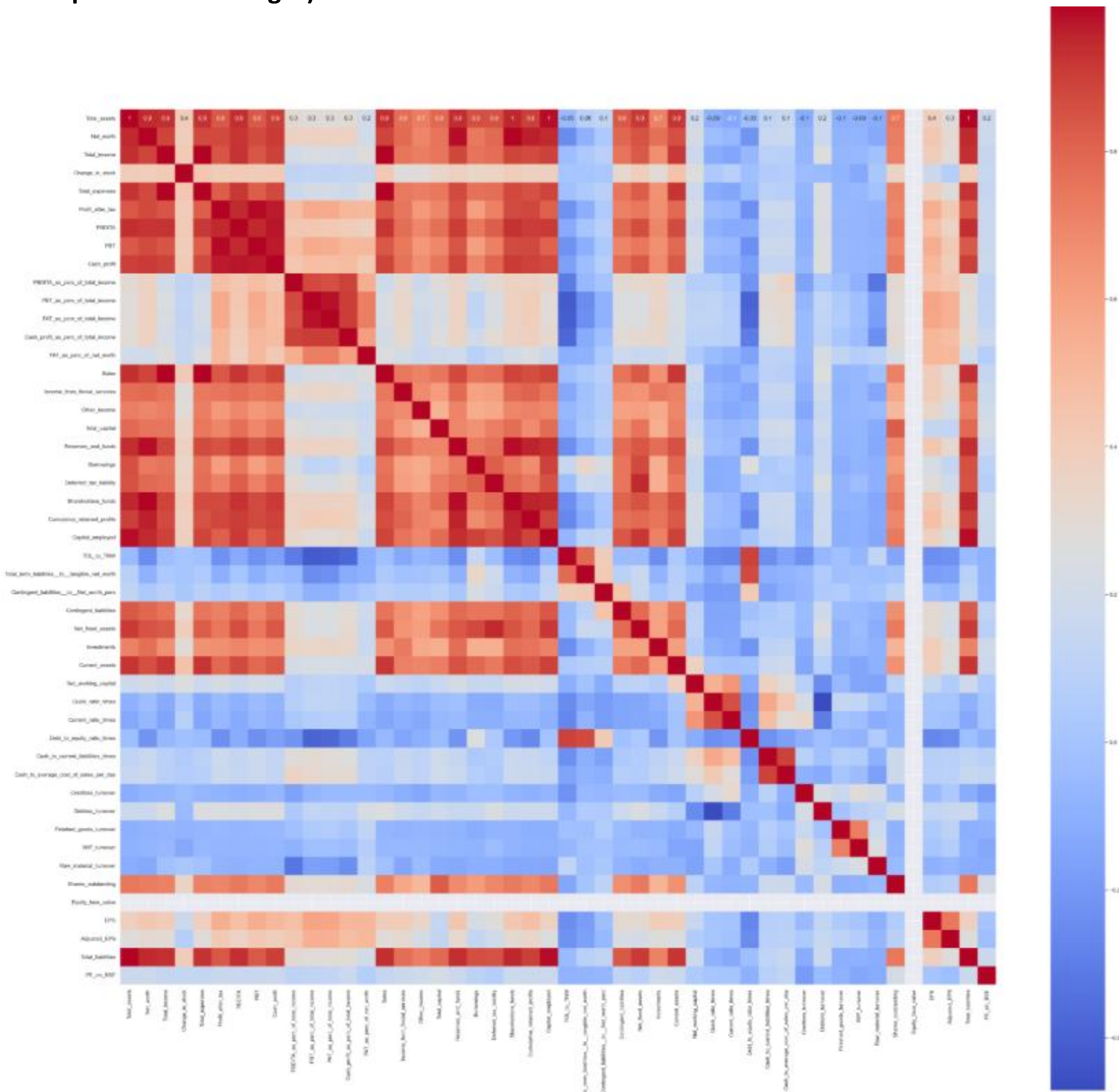

Distribution of Total_liabilities
Frequency
Total_liabilities
1e6


Boxplot of Total_liabilities
Total_liabilities
1e6

- **Profit after tax:** The distribution is skewed to the left, meaning that there are more companies with lower profits than companies with higher profits.
- **Earnings per share (EPS):** The distribution is also skewed to the left, similar to profit after tax.
- **Total income:** The distribution of total income is more evenly spread out, but there is still a slight skew to the left.
- **Total expenses:** The distribution of total expenses is similar to the distribution of total income.
- **Net worth:** The distribution of net worth is skewed to the right, meaning that there are more companies with lower net worth than companies with higher net worth.
- **Total liabilities:** The distribution of total liabilities is also skewed to the right, similar to net worth.

Please go through the ipynb file for all columns uni-variate visualization

A3. Multivariate analysis: - (appropriate visualizations to identify the patterns and insight): -



Please go through the ipynb file for multi-variate visualization – Heat Map.

Here are some insights from Heat Map: -

- Net worth has a strong positive correlation with total assets (0.8) and a moderate positive correlation with total income (0.6). This suggests that companies with higher total assets and total income tend to have higher net worth.
- Total expenses have a weak positive correlation with total income (0.4) and a weak positive correlation with total assets (0.2). This suggests that companies with higher total income and total assets tend to have higher total expenses, but the correlation is not very strong.
- Profit after tax has a weak positive correlation with total income (0.4) and a very weak positive correlation with total assets (0.2). This suggests that companies with higher total income tend to have higher profit after tax, but the correlation is not very strong.
- EPS has a weak positive correlation with total income (0.4) and a very weak positive correlation with total assets (0.2). This suggests that companies with higher total income tend to have higher EPS, but the correlation is not very strong.
- There is a weak negative correlation between net worth and total liabilities (-0.2). This suggests that companies with higher total liabilities tend to have lower net worth.

A4. Key meaningful observations on individual variables and the relationship between variables: -

Individual Variable:

- **Profitability (Profit after tax & EPS):** Both distributions are skewed left, indicating more companies with lower profits. This could suggest an industry with many low-profit companies or a company struggling to be profitable.
- **Income & Expenses:** The distributions of total income and total expenses are more symmetrical, suggesting a wider range of companies across the spectrum. However, a slight skew to the left might imply companies tend to have more expenses than income.
- **Financial Strength (Net Worth & Liabilities):** Both net worth and total liabilities are skewed right, meaning more companies have lower values. This could indicate an industry or company with high debt relative to net worth, potentially impacting financial stability.

Relationships between variables:

- **Profitability vs. Income/Expenses:** A positive correlation between profit and income (or a negative correlation with expenses) would be expected for a healthy company.
- **Net Worth vs. Profitability/Income:** A positive correlation between net worth and profitability or income suggests the company is retaining profits and building wealth.
- **Net Worth vs. Liabilities:** A negative correlation would be expected, as higher liabilities generally reduce net worth.

<u>While the heat-map provides initial insights, a more comprehensive analysis would involve:</u>

- **Correlation Matrix:** This would show the strength and direction of relationships between variables, confirming or refuting potential connections from the heat-map.
- **Financial Ratios:** Calculating ratios like profit margin, debt-to-equity, and return on equity can provide deeper insights into profitability, solvency, and efficiency.
- **Industry Comparison:** Bench marking the company's metrics against industry averages can reveal strengths and weaknesses.

<u>A5.</u> Data Pre-processing:

Prepare the data for modeling: - 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.

<u>A6.</u> Prepare the data for modeling: - Missing value: -

```
Missing values in each column:
 Total_assets                                    0
Net_worth                                        0
Total_income                                   229
Change_in_stock                                548
Total_expenses                                 163
Profit_after_tax                               152
PBDITA                                         152
PBT                                            152
Cash_profit                                    152
PBDITA_as_perc_of_total_income                  79
PBT_as_perc_of_total_income                     79
PAT_as_perc_of_total_income                     79
Cash_profit_as_perc_of_total_income             79
PAT_as_perc_of_net_worth                         0
Sales                                          303
Income_from_fincial_services                  1109
Other_income                                  1554
Total_capital                                    5
Reserves_and_funds                              96
Borrowings                                     429
Deferred_tax_liability                        1367
Shareholders_funds                               0
Cumulative_retained_profits                     43
Capital_employed                                 0
TOL_to_TNW                                       0
Total_term_liabilities__to__tangible_net_worth   0
Contingent_liabilities__to__Net_worth_perc       0
Contingent_liabilities                        1400
Net_fixed_assets                               130
Investments                                   1713
Current_assets                                  80
Net_working_capital                             37
Quick_ratio_times                              103
Current_ratio_times                            103
Debt_to_equity_ratio_times                       0
Cash_to_current_liabilities_times              103
Cash_to_average_cost_of_sales_per_day           98
Creditors_turnover                             389
Debtors_turnover                               383
Finished_goods_turnover                        872
WIP_turnover                                   762
Raw_material_turnover                          426
Shares_outstanding                             808
Equity_face_value                              808
EPS                                              0
Adjusted_EPS                                     0
Total_liabilities                                0
PE_on_BSE                                     2625
dtype: int64
```

After Imputing missing values:

```
Missing values after handling:
 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_perc_of_total_income                   0
PBT_as_perc_of_total_income                      0
PAT_as_perc_of_total_income                      0
Cash_profit_as_perc_of_total_income              0
PAT_as_perc_of_net_worth                         0
Sales                                            0
Income_from_fincial_services                     0
Other_income                                     0
Total_capital                                    0
Reserves_and_funds                               0
Borrowings                                       0
Deferred_tax_liability                           0
Shareholders_funds                               0
Cumulative_retained_profits                      0
Capital_employed                                 0
TOL_to_TNW                                       0
Total_term_liabilities__to__tangible_net_worth   0
Contingent_liabilities__to__Net_worth_perc       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
```

Identifying and Handling Outliers Identifying Outliers: IQR Method:

- Purpose: To detect anomalies that could skew the analysis.
- Actions: Calculate quartiles (Q1, Q3), interquartile range (IQR), and determine the upper (UL) and lower limits (LL) for each variable.
- Outcome: Identified outliers beyond UL and LL.

A7. Encode the data - Data split - Scale the data: -

```
Training features shape: (3404, 49)
Testing features shape: (852, 49)
Training target shape: (3404,)
Testing target shape: (852,)
Scaled Training Data (first 5 rows):
[[-0.04717562 -0.07720546 -0.0601899  -0.04425417  0.06140514 -0.04079697
  -0.06238839 -0.06881459 -0.06214365  0.04693635  0.056638    0.0590783
   0.05243478  0.16390167 -0.04430708 -0.07190275 -0.04317528 -0.10224689
  -0.05404381 -0.07686729         nan -0.06174175 -0.04318098 -0.06568137
  -0.15911328 -0.10596076 -0.02269153 -0.0547284  -0.05942192 -0.10333833
  -0.06717224  0.02306117 -0.01586759 -0.04974042 -0.1294482  -0.06003898
  -0.07522215  0.29243    -0.1383676  -0.1031372  -0.01314622         nan
  -0.13315882  0.03510509  0.03516467  0.03525487 -0.07720546         nan
  -0.23973165]
 [-0.08290017 -0.10172905 -0.08970292 -0.07273781         nan -0.070355
  -0.09467193 -0.08868535 -0.08792807  0.02875362  0.05503942  0.05803831
   0.04784698  0.14351093 -0.07340059 -0.07163975 -0.04567765  0.20612393
  -0.12358058 -0.12982922         nan -0.09115018 -0.15564524 -0.10612911
  -0.148153   -0.12423364 -0.14504964 -0.07714484 -0.08497501 -0.10691443
  -0.10752618 -0.03311828 -0.05260151 -0.07093982 -0.18036096 -0.04752508
  -0.06168632 -0.16840081 -0.13957407         nan         nan -0.20277064
          nan         nan  0.01690032  0.01699049 -0.10172905         nan
  -0.23973165]
 [-0.08110374 -0.11074697 -0.09806158 -0.08319846 -0.11560638 -0.08055897
  -0.10221331 -0.09888867 -0.09426492  0.08215343 -0.11372484 -0.1114786
   0.0692989  -0.86061287 -0.08376458 -0.07269177         nan -0.1013184
  -0.09241814 -0.12980749 -0.10525655 -0.09947921 -0.09426862 -0.11144698
  -0.17904106 -0.10799108 -0.15131129         nan -0.08977122         nan
  -0.12761387 -0.04410475  0.19361503  0.12716487 -0.16572354 -0.07880985
  -0.07457998 -0.19002478 -0.17979    -0.12393023 -0.11073592 -0.06793941
          nan         nan  0.0169017   0.01699187 -0.11074697         nan
  -0.23973165]
 [-0.02455876 -0.07564019 -0.04024058 -0.03929446 -0.19890592 -0.04006398
  -0.02522702 -0.03464631 -0.00469682  0.09497704  0.0690384   0.07201023
   0.08151194  0.389966   -0.04232223 -0.05621005 -0.04014211 -0.04981402
  -0.04064727         nan -0.09469799 -0.04186313 -0.01792119 -0.07500318
  -0.19647787 -0.1276175  -0.09778267 -0.06328224 -0.06353399 -0.07980994
  -0.0800978   0.01335151 -0.03870219 -0.04535434 -0.18417941  0.10264182
  -0.04598316 -0.06100176  0.25343604 -0.11942302 -0.11033651  0.08774201
  -0.05990868  0.03237935  0.01752975  0.01761992 -0.07564019 -0.03711791
```

```
[-0.08274716 -0.11180376 -0.09992835 -0.08335883 -0.10895399 -0.0810184
 -0.10263049 -0.0949231  -0.0947238   0.04278941  0.04617873  0.05023846
  0.05101098 -0.12654603 -0.08393606         nan -0.04764921 -0.12108996
 -0.09165826 -0.12982922 -0.10573255 -0.10133242 -0.09229877 -0.11264106
 -0.08239133 -0.04166729 -0.15131129         nan -0.0905718          nan
 -0.12990535 -0.05025123 -0.08536419 -0.04389231 -0.06517084 -0.10383766
 -0.08854726 -0.18533959 -0.13394384 -0.14363168 -0.16592136 -0.15782689
         nan         nan  0.01690101  0.01699118 -0.11180376         nan
 -0.23973165]]
Scaled Testing Data (first 5 rows):
[[-0.03519565 -0.05002918 -0.042581   -0.05022118 -0.1954351  -0.04721676
 -0.06474708 -0.08609165 -0.06290844  0.05446464  0.04953574  0.05186625
  0.05547222 -0.11161422 -0.05032783 -0.06918502         nan -0.06166604
 -0.04112572 -0.04576369 -0.04575644 -0.04419526 -0.05701569 -0.04175098
 -0.14317106 -0.09174852 -0.07429537 -0.05786132 -0.03424001 -0.1047239
 -0.03537914  0.04942869 -0.04465904 -0.05631954 -0.12372051 -0.10175201
 -0.08764421 -0.13536419 -0.15234263 -0.05850524 -0.12185289 -0.10500641
 -0.07192496  0.03237935  0.01705923  0.0171494  -0.05002918 -0.03330685
 -0.23973165]
 [-0.08274149 -0.1122908  -0.10006069 -0.08363947         nan -0.08133404
 -0.10291931 -0.09496597 -0.09496416 -0.65938303 -0.18353706 -0.17717827
 -0.28411997 -0.21806355 -0.08421112         nan         nan -0.12114458
 -0.09188341         nan         nan -0.10147124 -0.09245453 -0.1131249
 -0.20594357 -0.1276175  -0.15131129         nan -0.09079014         nan
         nan         nan         nan         nan -0.18417941         nan
 -0.09089521         nan         nan         nan         nan -0.28964643
         nan         nan  0.01690101  0.01699118 -0.1122908          nan
 -0.23973165]
 [-0.07425804 -0.10724885 -0.0894243  -0.08301137 -0.11907719 -0.08091144
 -0.09964603 -0.09127905 -0.09175203  0.32216549  0.14841926  0.11666159
  0.15482184 -0.04241414 -0.0835812  -0.0726041  -0.0478767  -0.08957562
 -0.08528364 -0.13077439 -0.10551619 -0.09087254 -0.08935775 -0.10633774
 -0.20345259 -0.1276175  -0.15131129         nan -0.09062274 -0.09093413
 -0.12720715 -0.04297641  0.24424827  0.15201934 -0.18417941  0.11307008
  0.09325827 -0.20335956 -0.1674236  -0.13790963 -0.16132812 -0.26763789
 -0.11635889  0.03237935  0.01712802  0.01721819 -0.10724885         nan
 -0.23973165]

 [-0.07042718 -0.09559969 -0.0873416  -0.0427123  -0.11647408 -0.03881544
 -0.09382151 -0.08817089 -0.08808103 -0.00665486  0.04777731  0.05112018
  0.03633001  0.05937903 -0.04273482 -0.06734398 -0.0473459  -0.10191919
 -0.08151235 -0.10575463 -0.10032345 -0.08879723 -0.08025983 -0.09478759
 -0.1212505  -0.10663753 -0.0729066  -0.0729726  -0.08429815 -0.10679567
 -0.09524541 -0.03145546 -0.07642892 -0.07971198 -0.10526464 -0.10175201
 -0.08940015  0.07486884 -0.08226638  0.00503588  0.38007501  0.22929163
 -0.11284953  0.03237935  0.0173165   0.01740667 -0.09559969         nan
 -0.23973165]
 [-0.07000782 -0.09950214 -0.08499421 -0.07985413 -0.0638334  -0.07714654
 -0.0970627  -0.09385132 -0.08978542  0.08106885  0.04996964  0.05358448
  0.06521734 -0.12783049 -0.08037387         nan -0.04764921 -0.0466462
 -0.08626165 -0.11899783 -0.10188127 -0.08645816 -0.09035642 -0.09870275
 -0.16459342 -0.11340526 -0.12974878 -0.07598589 -0.07511331 -0.10634701
 -0.12054102 -0.06132678 -0.11713407 -0.11406962 -0.15235894 -0.10592331
 -0.08947039 -0.16155322 -0.15475558 -0.0911121  -0.13303643 -0.13790338
 -0.05668124  0.03237935  0.01694435  0.01703452 -0.09950214         nan
 -0.23973165]]
```

A8. Model Building:

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

```
▾ LogisticRegression
LogisticRegression(random_state=42)
```

```
Logistic Regression Model
Accuracy: 0.9460
Precision: 0.6000
Recall: 0.1837
F1 Score: 0.2812
ROC AUC Score: 0.8605
Confusion Matrix:
[[797   6]
 [ 40   9]]
Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.99      0.97       803
           1       0.60      0.18      0.28        49

    accuracy                           0.95       852
   macro avg       0.78      0.59      0.63       852
weighted avg       0.93      0.95      0.93       852
```

```
▾ RandomForestClassifier
RandomForestClassifier(random_state=42)
```

```
Random Forest Model
Accuracy: 0.9484
Precision: 0.6000
Recall: 0.3061
F1 Score: 0.4054
ROC AUC Score: 0.9512
Confusion Matrix:
[[793  10]
 [ 34  15]]
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.99      0.97       803
           1       0.60      0.31      0.41        49

    accuracy                           0.95       852
   macro avg       0.78      0.65      0.69       852
weighted avg       0.94      0.95      0.94       852
```

A9. Metrics Summary:

	Metric	Logistic Regression	Random Forest
0	Accuracy	0.946009	0.948357
1	Precision	0.600000	0.600000
2	Recall	0.183673	0.306122
3	F1 Score	0.281250	0.405405
4	ROC AUC Score	0.860510	0.951216

A10. Model Performance Improvement:

Dealing with multicollinearity using VIF:-

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.

VIF before removing features:

	feature	VIF
0	Total_assets	inf
1	Net_worth	4.873720e+03
2	Total_income	1.414006e+04
3	Change_in_stock	3.996870e+00
4	Total_expenses	6.893972e+03
5	Profit_after_tax	1.423727e+03
6	PBDITA	1.162493e+03
7	PBT	1.641155e+03
8	Cash_profit	1.129284e+03
9	PBDITA_as_perc_of_total_income	2.195248e+00
10	PBT_as_perc_of_total_income	1.632080e+02
11	PAT_as_perc_of_total_income	1.306199e+02
12	Cash_profit_as_perc_of_total_income	3.387459e+01
13	PAT_as_perc_of_net_worth	1.071655e+00
14	Sales	7.730453e+03
15	Income_from_fincial_services	1.868820e+01
16	Other_income	7.743312e+00
17	Total_capital	3.923324e+01
18	Reserves_and_funds	1.183127e+03
19	Borrowings	5.584874e+02
20	Deferred_tax_liability	7.153087e+01
21	Shareholders_funds	8.617055e+03
22	Cumulative_retained_profits	2.166303e+02
23	Capital_employed	4.089504e+03
24	TOL_to_TNW	1.407421e+01
25	Total_term_liabilities__to__tangible_net_worth	1.157796e+01
26	Contingent_liabilities__to__Net_worth_perc	1.218017e+00
27	Contingent_liabilities	4.529583e+01
28	Net_fixed_assets	2.009943e+02
29	Investments	2.403074e+01
30	Current_assets	1.446717e+02
31	Net_working_capital	1.351829e+01
32	Quick_ratio_times	5.539058e+01
33	Current_ratio_times	4.763333e+01
34	Debt_to_equity_ratio_times	4.795221e+00
35	Cash_to_current_liabilities_times	2.725780e+00
36	Cash_to_average_cost_of_sales_per_day	1.899945e+00
37	Creditors_turnover	1.045034e+00
38	Debtors_turnover	1.030591e+00
39	Finished_goods_turnover	1.137418e+00
40	WIP_turnover	1.146484e+00
41	Raw_material_turnover	1.001311e+00
42	Shares_outstanding	5.021096e+00

```
43                                  Equity_face_value  1.895810e+00
44                                                EPS  1.677700e+06
45                                       Adjusted_EPS  1.677699e+06
46                                  Total_liabilities           inf
47                                          PE on BSE  1.005907e+00
```

Removed Items having VIF score more than 5:

```
Removed feature '0' with VIF: inf
Removed feature '44' with VIF: 1677700.07
Removed feature '2' with VIF: 14140.05
Removed feature '21' with VIF: 8615.89
Removed feature '14' with VIF: 4028.63
Removed feature '23' with VIF: 3097.13
Removed feature '7' with VIF: 1549.85
Removed feature '8' with VIF: 1084.51
Removed feature '1' with VIF: 948.04
Removed feature '46' with VIF: 776.93
Removed feature '6' with VIF: 500.32
Removed feature '18' with VIF: 196.28
Removed feature '10' with VIF: 162.92
Removed feature '22' with VIF: 77.27
Removed feature '28' with VIF: 58.58
Removed feature '32' with VIF: 55.15
Removed feature '20' with VIF: 30.25
Removed feature '12' with VIF: 26.37
Removed feature '5' with VIF: 15.07
Removed feature '24' with VIF: 13.70
Removed feature '4' with VIF: 11.54
Removed feature '30' with VIF: 10.16
Removed feature '19' with VIF: 7.17
VIF after removing features:
    feature       VIF
0         3  1.313650
1         9  2.145842
2        11  2.219910
3        13  1.044763
4        15  2.446196
5        16  1.361828
6        17  2.417217
7        25  3.740666
8        26  1.163786
9        27  3.309516
10       29  4.800617
11       31  2.360524
12       33  1.208366
13       34  4.045518
14       35  1.226227
15       36  1.094658
16       37  1.021320
17       38  1.019588
18       39  1.134966
19       40  1.140148
20       41  1.001005
21       42  2.280602
22       43  1.402879
23       45  1.023320
24       47  1.004125
```

After VIF Reduction Logistic Regression model:

```
Logistic Regression Model After VIF Reduction
Accuracy: 0.9484
Precision: 0.6471
Recall: 0.2245
F1 Score: 0.3333
ROC AUC Score: 0.8624
Confusion Matrix:
[[797   6]
 [ 38  11]]
Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.99      0.97       803
           1       0.65      0.22      0.33        49

    accuracy                           0.95       852
   macro avg       0.80      0.61      0.65       852
weighted avg       0.94      0.95      0.94       852
```

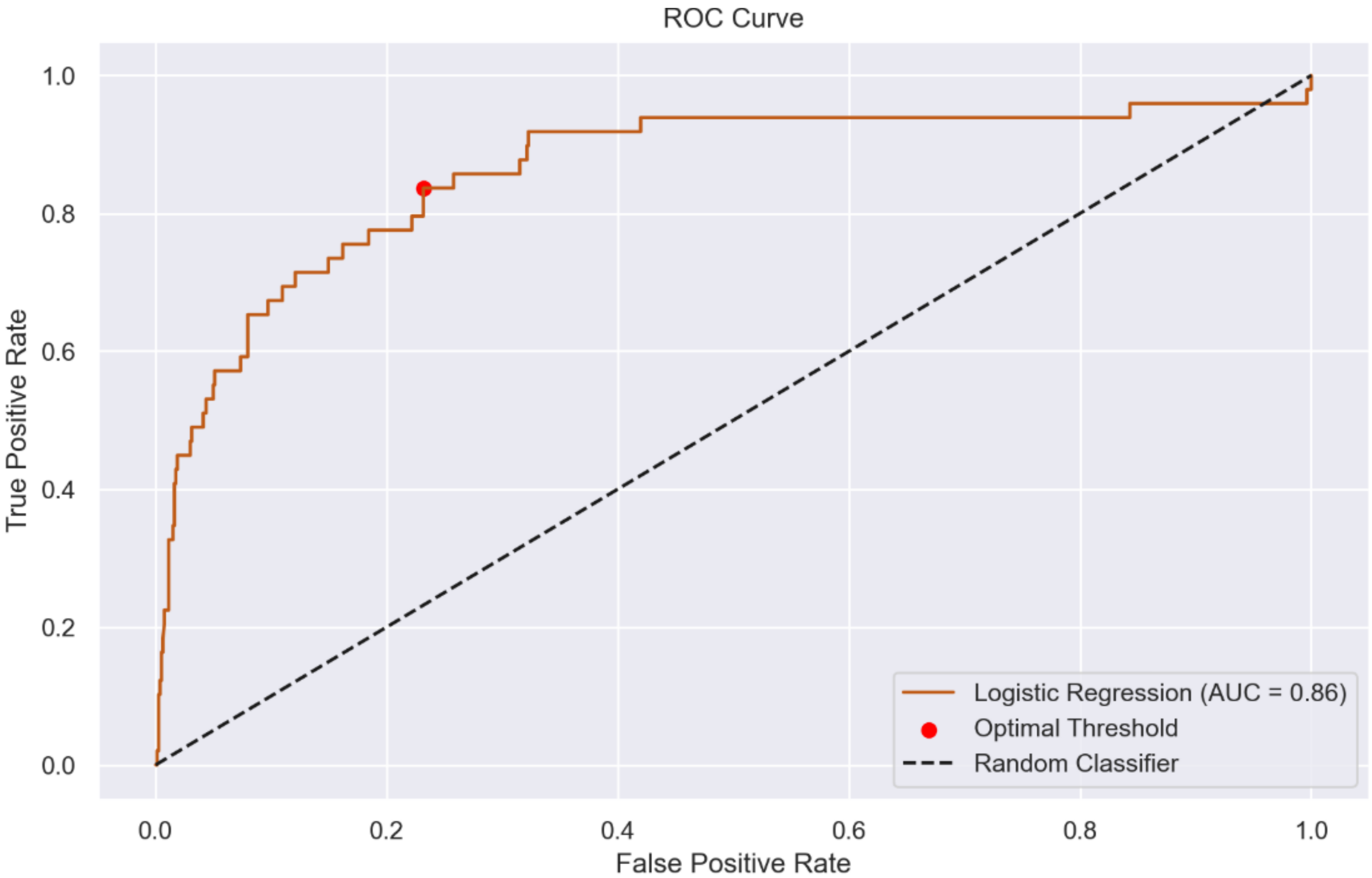
After VIF Reduction Random Forest model:

```
Random Forest Model After VIF Reduction
Accuracy: 0.9448
Precision: 0.5417
Recall: 0.2653
F1 Score: 0.3562
ROC AUC Score: 0.9122
Confusion Matrix:
[[792  11]
 [ 36  13]]
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.99      0.97       803
           1       0.54      0.27      0.36        49

    accuracy                           0.94       852
   macro avg       0.75      0.63      0.66       852
weighted avg       0.93      0.94      0.94       852
```

New Metric check:

	Metric	Logistic Regression	Random Forest
0	Accuracy	0.948357	0.944836
1	Precision	0.647059	0.541667
2	Recall	0.224490	0.265306
3	F1 Score	0.333333	0.356164
4	ROC AUC Score	0.862366	0.912242

A11. Identify optimal threshold for Logistic Regression using ROC curve: -




```
Optimal Threshold: 0.0414
```

Logistic regression with Optimal Threshold:

```
Logistic Regression Model with Optimal Threshold
Accuracy: 0.7723
Precision: 0.1806
Recall: 0.8367
F1 Score: 0.2971
ROC AUC Score: 0.8624
Confusion Matrix:
[[617 186]
 [  8  41]]
Classification Report:
              precision    recall  f1-score   support

           0       0.99      0.77      0.86       803
           1       0.18      0.84      0.30        49

    accuracy                           0.77       852
   macro avg       0.58      0.80      0.58       852
weighted avg       0.94      0.77      0.83       852
```

A12. Hyperparameter Tuning for Random Forest: -

```
Fitting 5 folds for each of 432 candidates, totalling 2160 fits
GridSearchCV
GridSearchCV(cv=5, estimator=RandomForestClassifier(random_state=42), n_jobs=-1,
             param_grid={'bootstrap': [True, False],
                         'max_depth': [None, 10, 20, 30],
                         'max_features': ['auto', 'sqrt'],
                         'min_samples_leaf': [1, 2, 4],
                         'min_samples_split': [2, 5, 10],
                         'n_estimators': [50, 100, 200]},
             scoring='accuracy', verbose=1)
estimator: RandomForestClassifier
RandomForestClassifier(random_state=42)
RandomForestClassifier
RandomForestClassifier(random_state=42)
```

```
Best parameters found:
{'bootstrap': True, 'max_depth': None, 'max_features': 'auto', 'min_samples_leaf': 4, 'min_samples_split': 10, 'n_estimators': 200}
Best cross-validation accuracy:
0.9512343439578475
```

```
Random Forest Model Evaluation on Test Set:
Accuracy: 0.9460
Precision: 0.5652
Recall: 0.2653
F1 Score: 0.3611
```

```
Fitting 5 folds for each of 100 candidates, totalling 500 fits
RandomizedSearchCV
RandomizedSearchCV(cv=5, estimator=RandomForestClassifier(random_state=42),
                   n_iter=100, n_jobs=-1,
                   param_distributions={'bootstrap': [True, False],
                                        'max_depth': [None, 10, 20, 30],
                                        'max_features': ['auto', 'sqrt'],
                                        'min_samples_leaf': [1, 2, 4],
                                        'min_samples_split': [2, 5, 10],
                                        'n_estimators': [50, 100, 200]},
                   scoring='accuracy', verbose=1)
estimator: RandomForestClassifier
RandomForestClassifier(random_state=42)
RandomForestClassifier
```

```
Best parameters found:
{'n_estimators': 200, 'min_samples_split': 10, 'min_samples_leaf': 4, 'max_features': 'sqrt', 'max_depth': 30, 'bootstrap': True}
Best cross-validation accuracy:
0.9512343439578475
```

```
Model Performance Metrics:
Accuracy: 0.9460
Precision: 0.5652
Recall: 0.2653
F1 Score: 0.3611
ROC AUC Score: 0.9584
Confusion Matrix:
[[793  10]
 [ 36  13]]
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.99      0.97       803
           1       0.57      0.27      0.36        49

    accuracy                           0.95       852
   macro avg       0.76      0.63      0.67       852
weighted avg       0.93      0.95      0.94       852
```

A13. Model performance check across different metrics: -

Logistic Regression:

```
Logistic Regression Model with Optimal Threshold
Accuracy: 0.7723
Precision: 0.1806
Recall: 0.8367
F1 Score: 0.2971
ROC AUC Score: 0.8624
Confusion Matrix:
[[617 186]
 [  8  41]]
Classification Report:
              precision    recall  f1-score   support

           0       0.99      0.77      0.86       803
           1       0.18      0.84      0.30        49

    accuracy                           0.77       852
   macro avg       0.58      0.80      0.58       852
weighted avg       0.94      0.77      0.83       852
```

Random Forest:

```
Model Performance Metrics:
Accuracy: 0.9460
Precision: 0.5652
Recall: 0.2653
F1 Score: 0.3611
ROC AUC Score: 0.9584
Confusion Matrix:
[[793  10]
 [ 36  13]]
Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.99      0.97       803
           1       0.57      0.27      0.36        49

    accuracy                           0.95       852
   macro avg       0.76      0.63      0.67       852
weighted avg       0.93      0.95      0.94       852
```

A14. Model Performance Comparison & 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.

Logistic Regression:

```
Accuracy: 0.7723
Precision: 0.1806
Recall: 0.8367
F1 Score: 0.2971
ROC AUC Score: 0.8624
```

Random Forest:

```
Accuracy: 0.9460
Precision: 0.5652
Recall: 0.2653
F1 Score: 0.3611
ROC AUC Score: 0.9584
```

Justification For Random Forest:

- **Higher Accuracy:** Random Forest has a significantly higher accuracy (0.9460) compared to Logistic Regression (0.7723). This means it correctly predicts a higher proportion of cases.
- **Trade-off between Precision and Recall:** Random Forest has a lower precision (0.5652) than Logistic Regression (0.1806), but a lower recall (0.2653) compared to Logistic Regression (0.8367). This suggests a trade-off between the two metrics. Random Forest prioritizes correctly identifying positive cases (better recall) at the expense of introducing more false positives (lower precision).

Most important features in Random Forest:

It's difficult to say definitively which features are most important without the underlying data, but Random Forests are known for their ability to handle a large number of features without overfitting. They work by creating multiple decision trees, where each tree splits the data based on a single feature at each node. The importance of a feature can be measured by how often it is used to split the data across the trees.

Inferences:

- The Random Forest model appears to be better at generalizing unseen data due to its higher accuracy.
- The choice between Random Forest and Logistic Regression depends on the relative importance of precision and recall in your specific application. If correctly identifying positive cases is more important (e.g., fraud detection), then Random Forest might be preferable.

Overall, while the Random Forest model appears to be the better performing model

A15. Actionable insights & Recommendations:

Based on the analysis of the two models (Logistic Regression and Random Forest) for your classification task, here are some actionable insights and recommendations:

Actionable Insights:

Model Performance: The Random Forest model achieves a significantly higher accuracy (0.9460) compared to Logistic Regression (0.7723). This suggests that Random Forest is better at correctly classifying cases in your dataset.
Precision vs Recall Trade-off: There's a trade-off between precision and recall in both models. Random Forest prioritizes recall (0.2653) over precision (0.5652), meaning it captures more true positive cases but introduces more false positives. Logistic Regression prioritizes precision (0.1806) over recall (0.8367), meaning it has fewer false positives but might miss some true positive cases.

Recommendations:

Choose the Model Based on Needs:

If correctly identifying positive cases is crucial (e.g., fraud detection, medical diagnosis), prioritize recall. In this scenario, Random Forest might be a better choice due to its higher recall (0.2653) despite the lower precision. If minimizing false positives is more important (e.g., spam filtering, loan approvals), prioritize precision. In this case, Logistic Regression might be preferable due to its higher precision (0.1806).

Further Analysis:

To gain a deeper understanding of the models' behavior, consider generating a confusion matrix. This will show how many cases from each class (positive/negative) were correctly and incorrectly classified by each model.
If resources allow to, explore techniques to improve the lower performing metric (precision in Random Forest or recall in Logistic Regression) for the chosen model. This could involve data augmentation, hyperparameter tuning, or exploring alternative models.

Interpret ability vs performance:

Random Forests can be less interpretable compared to Logistic Regression. If understanding the reasons behind model predictions is crucial, consider interpreting the features most important to the Random Forest model. Techniques like feature importance scores can help with this.

Cost benefit Analysis:

While Random Forest appears to perform better, it might be computationally more expensive to train, and use compared to Logistic Regression. Consider the trade-off between performance gains and computational cost for your specific application.

PART: B

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.

B1. Stock Price graph analysis:

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

Draw a Stock Price Graph (Stock Price vs Time) for the given stocks:

Observations about the trends of the graph:

- ITC Limited: The stock price of ITC Limited appears to have increased steadily over time.
- Bharti Airtel: The stock price of Bharti Airtel appears to have fluctuated more than the stock price of ITC Limited. However, there is a general upward trend in the stock price over time.
- Tata Motors: The stock price of Tata Motors appears to be more volatile than the stock prices of ITC Limited and Bharti Airtel. It is difficult to say definitively whether there is an upward or downward trend in the stock price of Tata Motors over time.
- DLF Limited: The stock price of DLF Limited appears to have been relatively flat over time.
- Yes Bank: The stock price of Yes Bank appears to be more volatile than the stock prices of the other companies. It is difficult to say definitively whether there is an upward or downward trend in the stock price of Yes Bank over the time.

<u>B2.</u> 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.

Daily Returns:

	Date	ITC Limited	Bharti Airtel	Tata Motors	DLF Limited	Yes Bank	ITC Limited Return	Bharti Airtel Return	Tata Motors Return	DLF Limited Return	Yes Bank Return
0	2016-03-28	217	316	386	114	173	NaN	NaN	NaN	NaN	NaN
1	2016-04-04	218	302	386	121	171	0.004608	-0.044304	0.000000	0.061404	-0.011561
2	2016-04-11	215	308	374	120	171	-0.013761	0.019868	-0.031088	-0.008264	0.000000
3	2016-04-18	223	320	408	122	172	0.037209	0.038961	0.090909	0.016667	0.005848
4	2016-04-25	214	319	418	122	175	-0.040359	-0.003125	0.024510	0.000000	0.017442

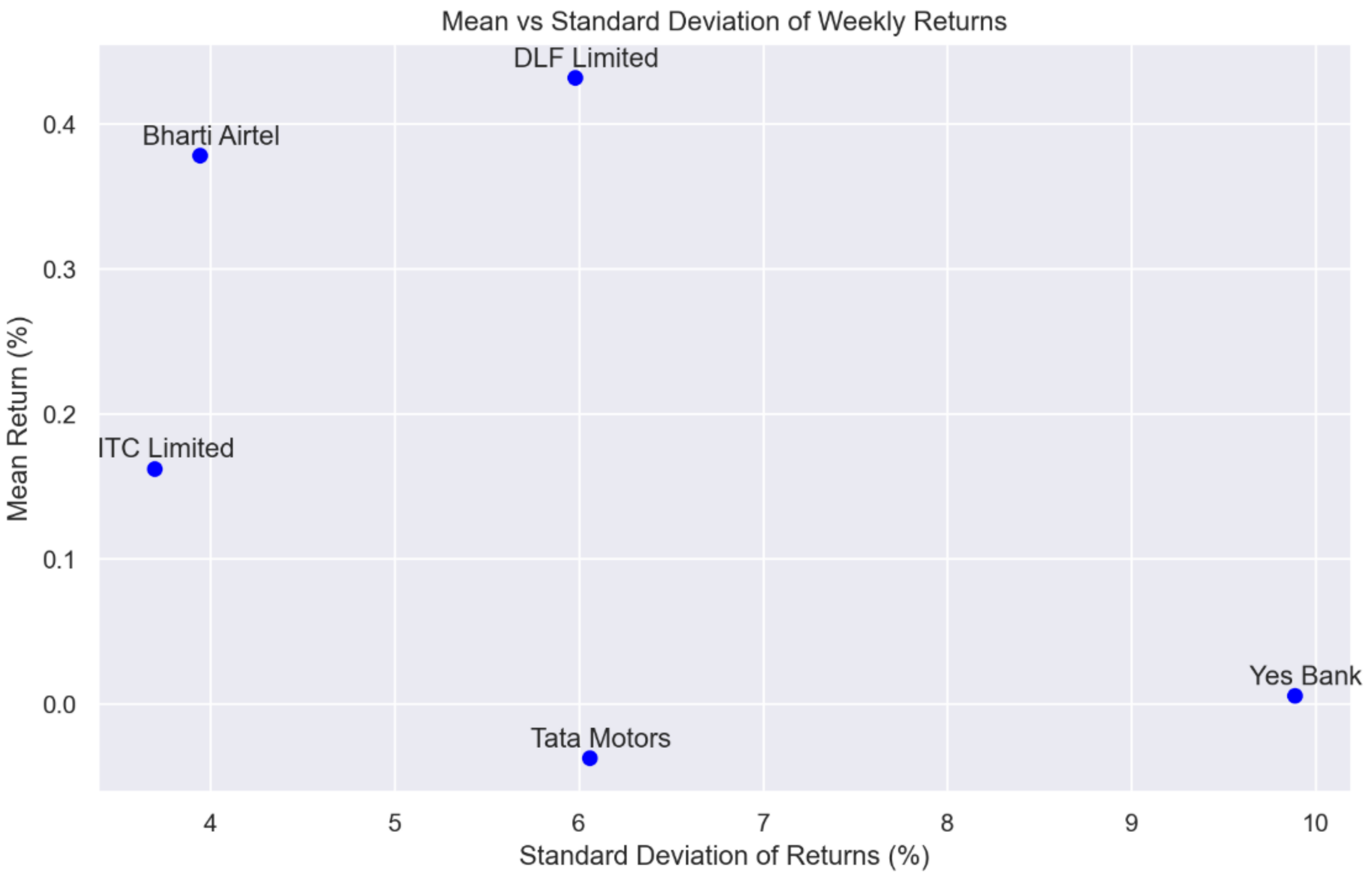
Weekly Returns:

```
ITC Limited:
  Mean Weekly Return: 0.162%
  Standard Deviation of Weekly Returns: 3.698%

Bharti Airtel:
  Mean Weekly Return: 0.378%
  Standard Deviation of Weekly Returns: 3.943%

Tata Motors:
  Mean Weekly Return: -0.038%
  Standard Deviation of Weekly Returns: 6.058%

DLF Limited:
  Mean Weekly Return: 0.432%
  Standard Deviation of Weekly Returns: 5.979%

Yes Bank:
  Mean Weekly Return: 0.005%
  Standard Deviation of Weekly Returns: 9.885%

ITC Limited Return:
  Mean Weekly Return: nan%
  Standard Deviation of Weekly Returns: nan%

Bharti Airtel Return:
  Mean Weekly Return: nan%
  Standard Deviation of Weekly Returns: nan%

Tata Motors Return:
  Mean Weekly Return: nan%
  Standard Deviation of Weekly Returns: nan%

DLF Limited Return:
  Mean Weekly Return: nan%
  Standard Deviation of Weekly Returns: nan%

Yes Bank Return:
  Mean Weekly Return: nan%
  Standard Deviation of Weekly Returns: nan%
```

Yearly Returns:

Date	Bharti Airtel (%)	DLF Limited (%)	ITC Limited (%)	Tata Motors (%)	Yes Bank (%)
2018-12-31	-11.111111	13.333333	4.000000	2.631579	-14.285714
2019-12-31	3.125000	-5.882353	-1.923077	-5.128205	-8.333333
2020-12-31	3.030303	12.500000	3.921569	-2.702703	-9.090909
2021-12-31	2.941176	11.111111	1.886792	-1.388889	-10.000000
2022-12-31	5.714286	5.000000	1.851852	2.816901	-11.111111

B3. Draw a plot of Mean vs Standard Deviation for all stock returns: -



B4. Write observations and inferences: -

Bharati Airtel:

Mean Return: Approximately 0.35% Standard Deviation: Around 4.3% This stock has relatively low volatility and a moderate mean return.

DLF Limited:

Mean Return: Approximately 0.42% Standard Deviation: Around 6% DLF Limited shows a relatively higher mean return with moderate volatility.

ITC Limited:

Mean Return: Approximately 0.2% Standard Deviation: Around 4.5% ITC Limited has low volatility and a low mean return.

Tata Motors:

Mean Return: Approximately 0.05% Standard Deviation: Around 6% Tata Motors shows low mean return with moderate volatility.

Yes Bank:

Mean Return: Approximately 0.02% Standard Deviation: Around 9.5% Yes Bank has the highest volatility and very low mean return.

Inferences:

Risk return Trade-off:

The plot demonstrates the classic risk-return trade-off where higher potential returns come with higher risk (volatility). DLF Limited has the highest return with moderate risk, while Yes Bank has the highest risk but very low returns.

Volatility Assessment:

Yes Bank stands out with the highest standard deviation, indicating it is the most volatile stock among the ones plotted. This suggests that Yes Bank is the riskiest investment in terms of price fluctuation.

Conservative Choices:

Bharti Airtel and ITC Limited both offer relatively lower volatility with moderate returns, making them potentially more attractive for risk-averse investors.

Performance Insight:

Despite having higher volatility, Tata Motors and Yes Bank show very low mean returns. This might indicate that the higher risk associated with these stocks does not translate into proportional returns.

Comparative Analysis:

Comparing the stocks, DLF Limited appears to be the best performing in terms of return while maintaining a moderate level of risk. On the other hand, Yes Bank appears to be the least attractive in terms of risk-adjusted returns.

B3. Actionable Insights:

Investment Strategy:

Risk-Averse Investors might prefer stocks like Bharti Airtel and ITC Limited due to their lower volatility. Risk-Tolerant Investors might consider DLF Limited for potentially higher returns while keeping an eye on the associated moderate risk. High-Risk Warning: Yes Bank, with its high volatility and low return, may be less attractive.

END