Finance and Risk Analytics Project

Name: Aishwariya Hariharan
PGP-DSBA Online September’ 23
Date: 04 Aug 2024

Contents

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:

1. 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.

2. 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.

Data Description

Sample of the Dataset:

	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	Adjusted EPS	Total liabilities	PE on BSE
0	1	395.3	827.6	336.5	534.1	13.5	508.7	38.9	124.4	64.6	...	5.65	3.99	3.37	14.87	8760056.0	10.0	4.44	4.44	827.6	NaN
1	2	36.2	67.7	24.3	137.9	-3.7	131.0	3.2	5.5	1.0	...	NaN	NaN	NaN	NaN	NaN	NaN	0.00	0.00	67.7	NaN
2	3	84.0	238.4	78.9	331.2	-18.1	309.2	3.9	25.8	10.5	...	2.51	17.67	8.76	8.35	NaN	NaN	0.00	0.00	238.4	NaN
3	4	2041.4	6883.5	1443.3	8448.5	212.2	8482.4	178.3	418.4	185.1	...	1.91	18.14	18.62	11.11	10000000.0	10.0	17.60	17.60	6883.5	NaN
4	5	41.8	90.9	47.0	388.6	3.4	392.7	-0.7	7.2	-0.6	...	68.00	45.87	28.67	19.93	107315.0	100.0	-6.52	-6.52	90.9	NaN

- 50 columns of the float type.
- One column of the integer type.
- The number of rows (observations) is 4256
- The number of columns (variables) is 51
- There are no duplicate rows
- The data has a lot of missing values (NaN)

Data Dictionary:

The data consists of financial metrics from the balance sheets of different companies. The detailed data dictionary is given below.

- 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

Note: **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.**

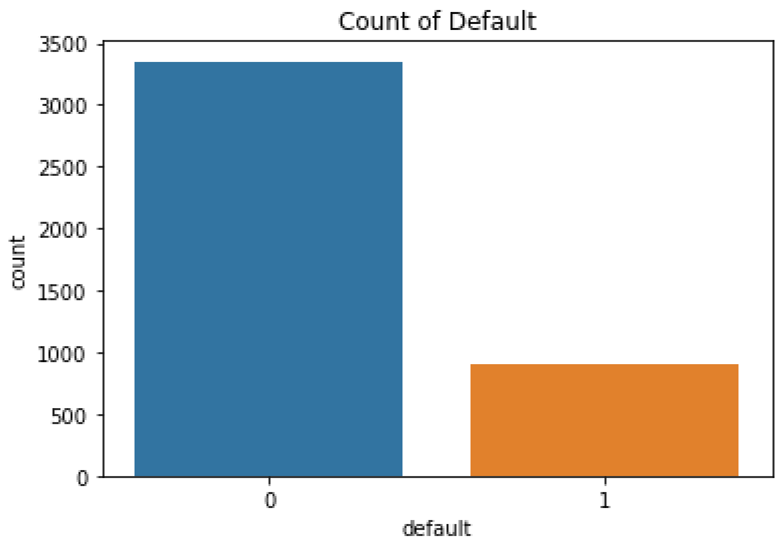
Statistical Summary of the Dataset

	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

Exploratory Data Analysis

Univariate Analysis

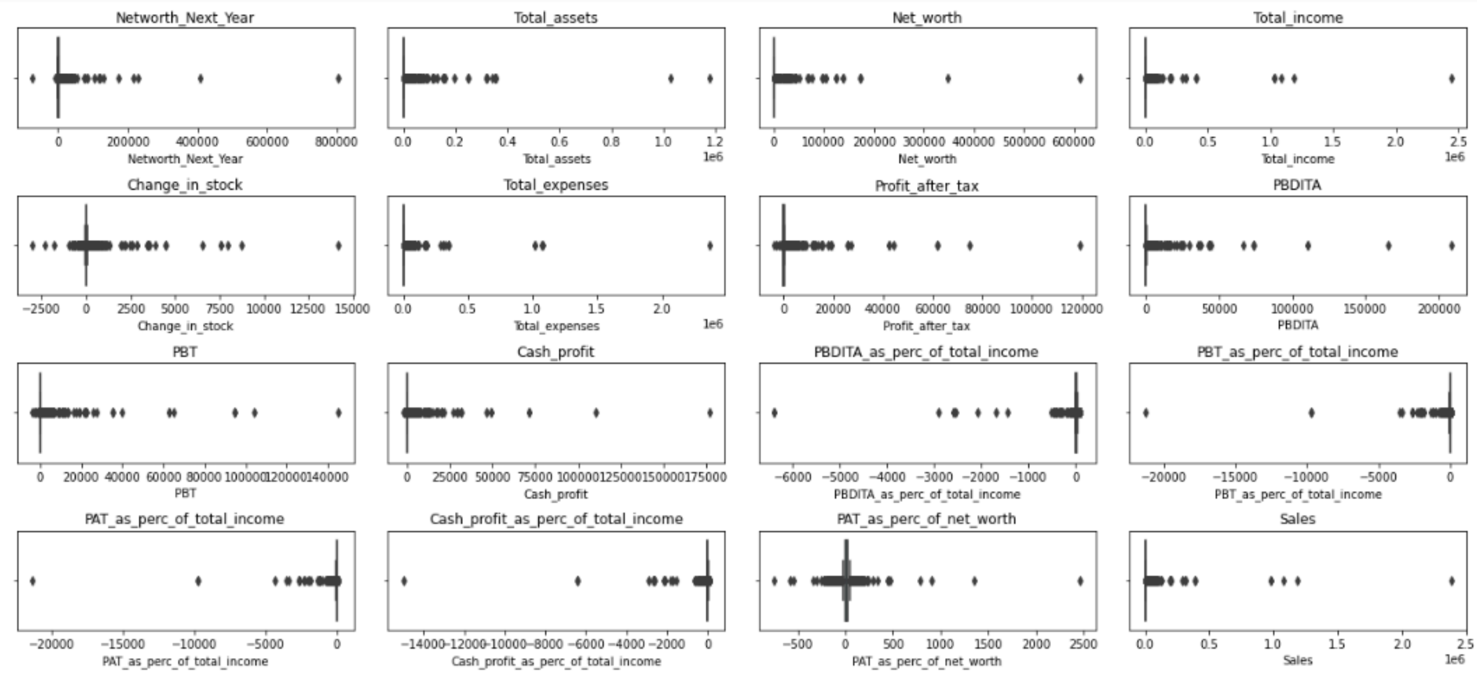
Number of defaulters

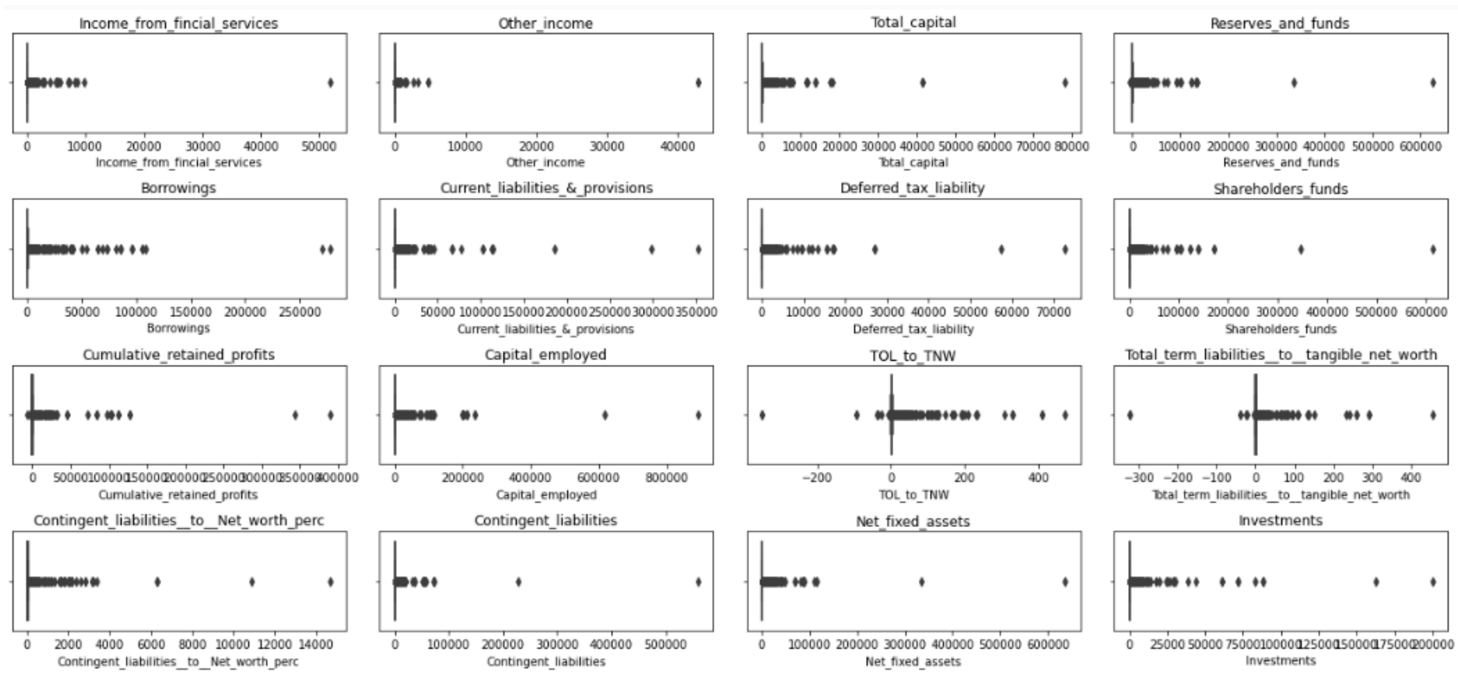


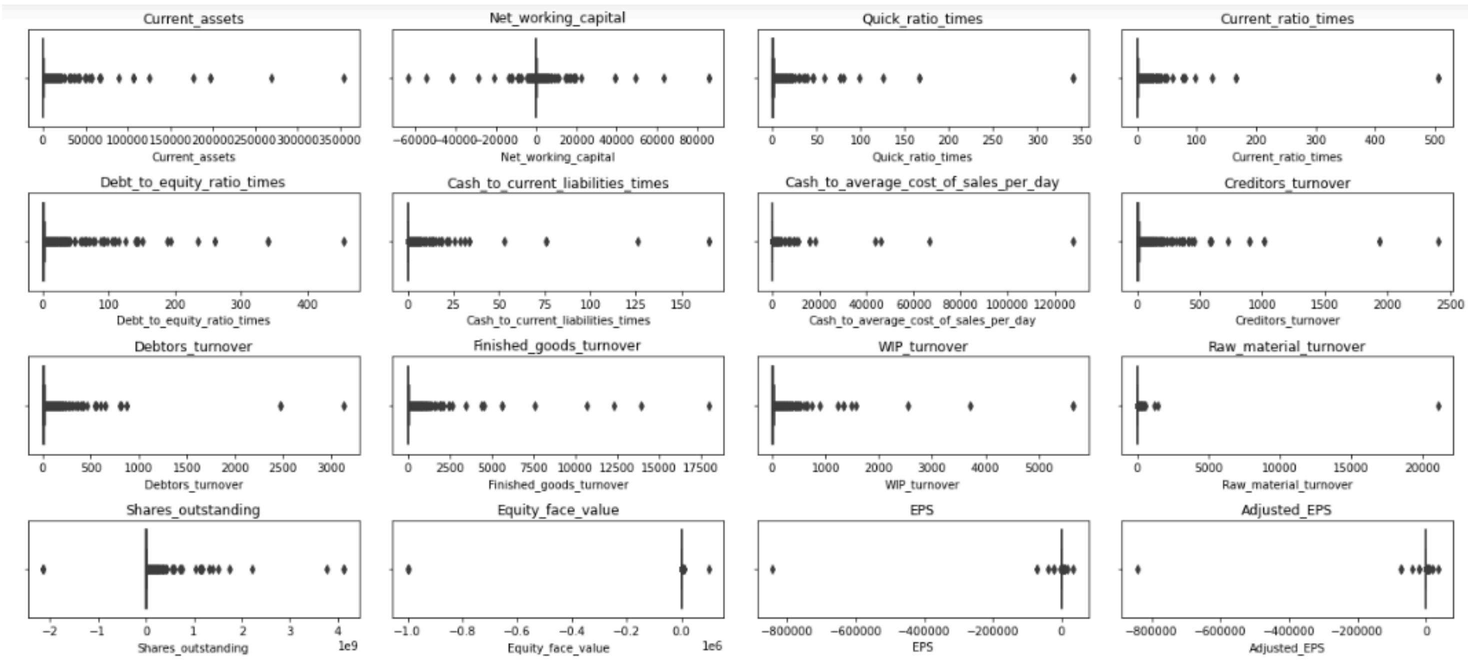

```
0    3352
1     904
```

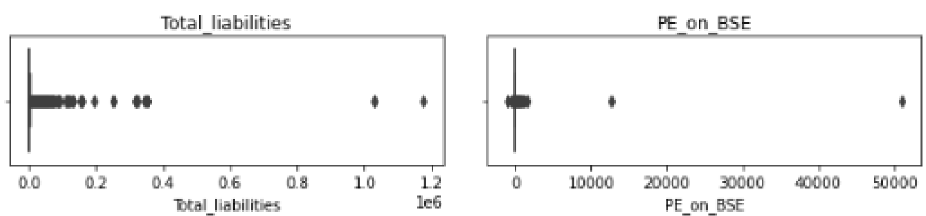
- There are 904 defaulters (21.2%) and 3352 (78.7%) non-defaulters.

Data distribution





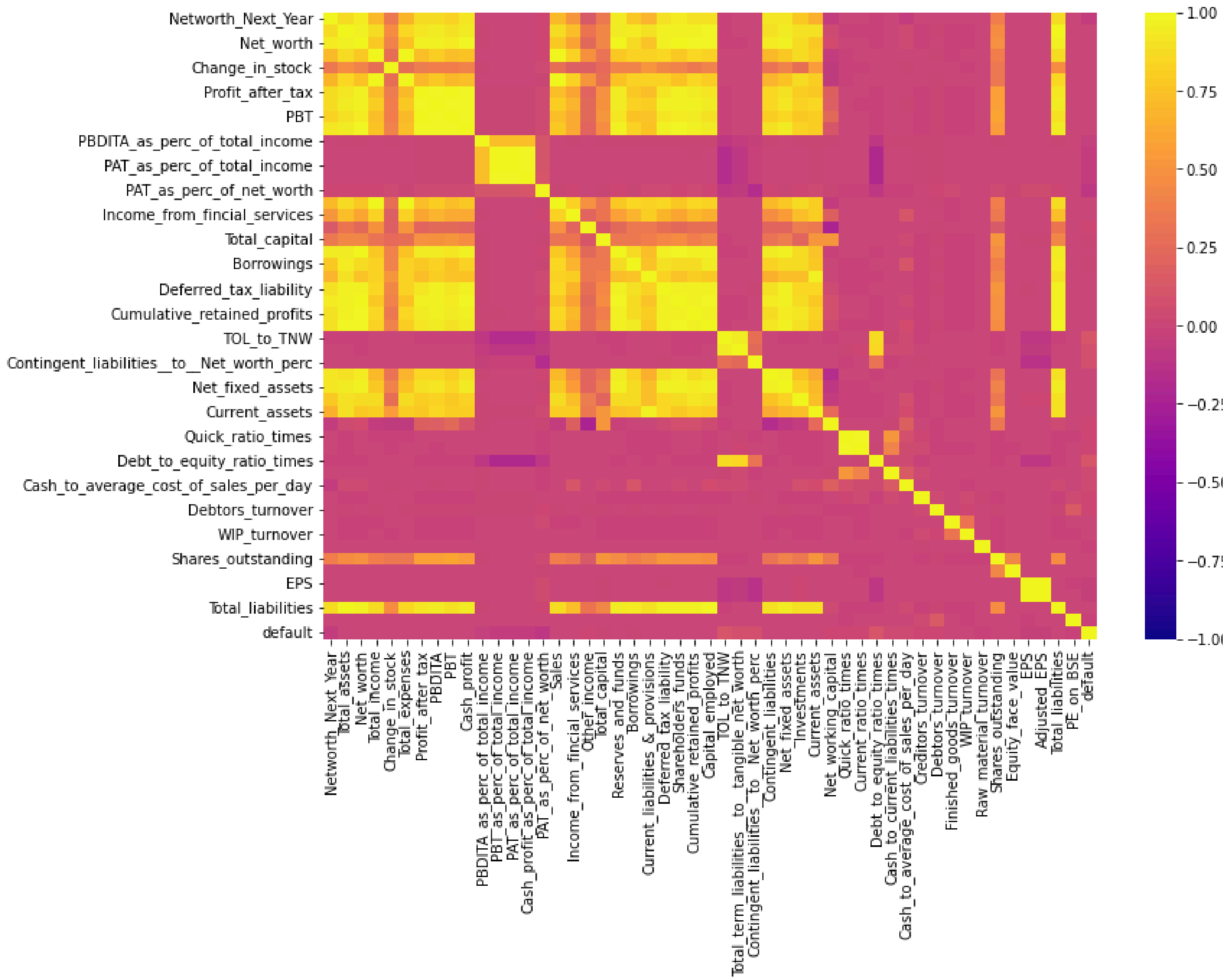




- We can observe that data in each column is skewed and has outliers.

Bivariate and Multivariate Analysis

<u>**Correlation between independent variables**</u>



Data preprocessing

- We drop the 'Num' column since it is a categorical variable.
- We drop the column 'Equity_face_value' since it has very few unique values, thus not impacting much to companies defaulting or not defaulting.
- The target variable is default and takes the value 1 when net worth next year is negative & 0 when net worth next year is positive.
- The proportion of default:
  - 0: 78.7%
  - 1: 21.2%
- We can see that the data is imbalanced.
- Missing values:
  - Number of missing values: 17778
  - Proportion of missing values: 8.19%
- The data has a lot of outliers
  - Number of outliers: 17477
  - Outliers can have a disproportionate effect on logistic regression models, leading to inaccurate predictions, biased coefficients, and poor model performance. Therefore, we replace the outliers with missing values and impute the new missing values using KNN imputer.
- We filter the data which is 90% or more complete at the row level.
- We drop columns with more than 30% missing values.
- We scale the data using the StandardScaler.

Model building

- We split the data for training and testing in the 75-25 ratio respectively.
- We compute different metrics– such as Accuracy, Recall, Precision, and F1 – to check the performance of a classification model built using sklearn

Justifying the evaluation metrics

In credit risk analysis, the most commonly used metric for evaluating models like logistic regression and random forest is the **Area Under the Receiver Operating Characteristic Curve (AUC-ROC)**.

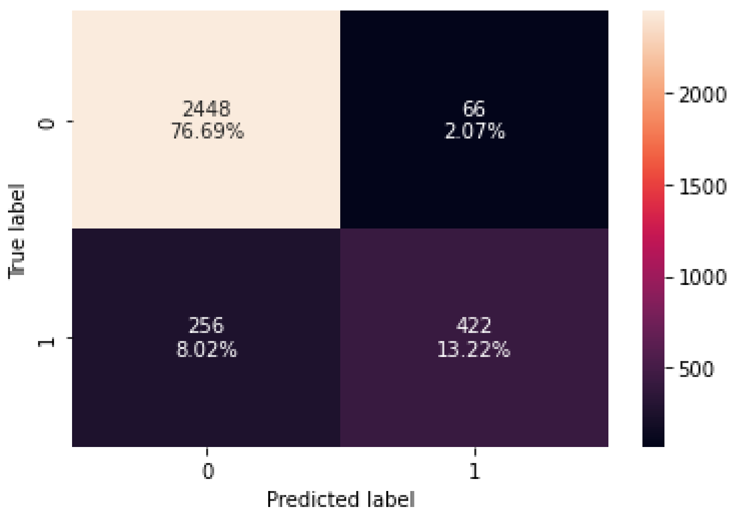
- AUC-ROC measures the model's ability to discriminate between the positive class (e.g., default) and the negative class (e.g., non-default) across all possible classification thresholds.
- It's a robust metric because it considers both true positive and false positive rates, providing a single number that summarizes the model's performance.
- It also handles class imbalance well.

Other metrics:

- **Accuracy**
  - It measures the proportion of correct predictions out of all predictions made but can be misleading in cases of class imbalances since in credit risk analysis, defaults might be rare, accuracy can be high even if the model fails to identify most defaults, making it less useful.
  - Hence we shall not use this metric.
- **Precision**:
  - This metric indicates how many of the predicted positives (defaults) are actually positive. It's important when the cost of false positives is high.
- **Recall (Sensitivity)**:
  - This measures how many of the actual positives (defaults) are correctly identified by the model. It's crucial in situations where identifying all defaults is a priority.
- **F1 score:**
  - The F1-score is the harmonic mean of precision and recall. It is useful when you need a balance between precision and recall.
  - The F1-score is helpful when dealing with imbalanced datasets, providing a single metric that accounts for both false positives and false negatives.

Logistic Regression

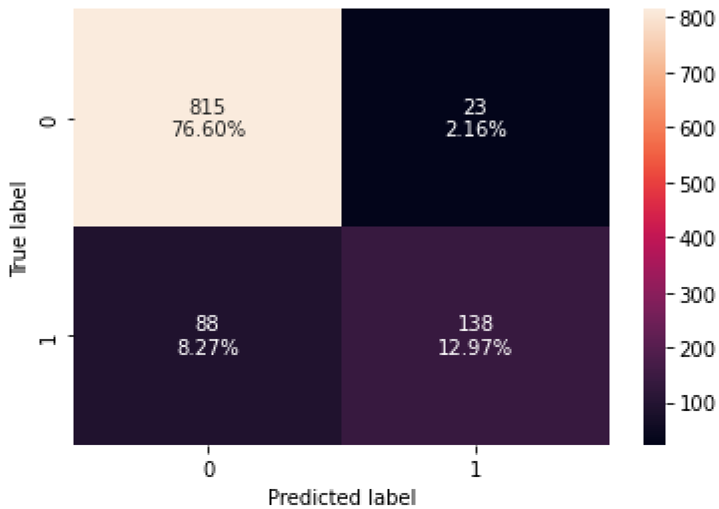
Training data



	Accuracy	Recall	Precision	F1
1	0.899123	0.622419	0.864754	0.723842

- A recall of 62.2% means that the model successfully identified 62.2% of the actual positive cases, i.e. the defaults.
- A precision of 86.4% indicates that 86.4% of the cases the model predicted as defaults were indeed defaults. This suggests that there are relatively few false positives (non-defaults that the model incorrectly identified as defaults), which is important for minimizing unnecessary interventions or actions on borrowers who are not actually at risk of default.
- An F1-score of 72.3% reflects a moderate balance between precision and recall, i.e. balancing between minimizing false positives and maximizing true positives.

Testing data

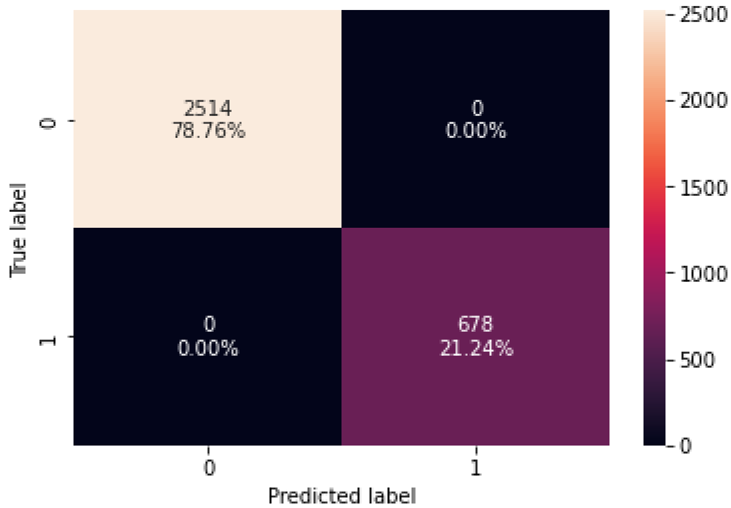


	Accuracy	Recall	Precision	F1
1	0.895677	0.610619	0.857143	0.713178

- A recall of 61% means that the model is correctly identifying 61% of all actual default cases. This value may be considered inadequate.The remaining 39% of actual positive cases were not identified by the model. These are false negatives, indicating that the model missed about one-third of the positive cases.
    - We need to improve or tune the model.
- A precision of 85.7% indicates that 85.7% of the cases the model predicted as defaults were indeed defaults. This high precision suggests that there are relatively few false positives (non-defaults that the model incorrectly identified as defaults), which is important for minimizing unnecessary interventions or actions on borrowers who are not actually at risk of default.
- An F1-score of 71.3% reflects a moderate balance between precision and recall. Neither metric is overly dominating, and the model is performing reasonably well in both aspects. This F1 score indicates that the model is somewhat reliable but not perfect.

Random forest

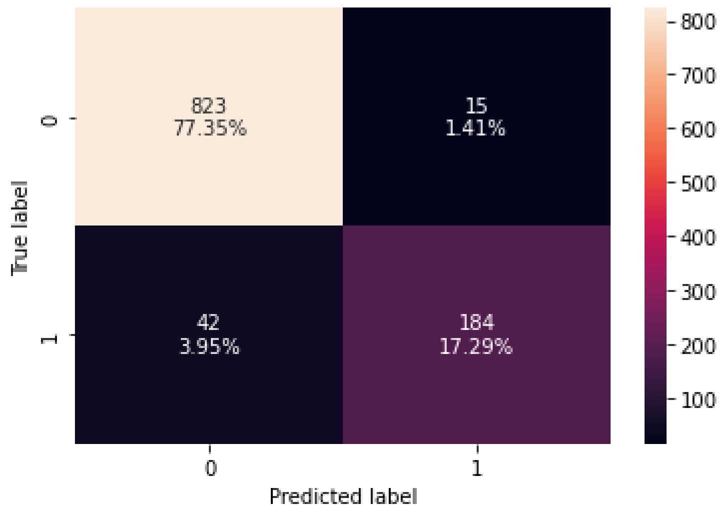
Training data



	Accuracy	Recall	Precision	F1
0	1.0	1.0	1.0	1.0

- A value of 1 for all metrics on the training data means that the model is making zero mistakes on the training set. It has perfectly learned the patterns (and possibly the noise) in the training data.
- This could also be due to overfitting of data when the model is too complex (e.g., too many trees, deep trees, or too many features) and ends up memorizing the training data instead of learning general patterns.

Testing data



	Accuracy	Recall	Precision	F1
1	0.946429	0.814159	0.924623	0.865882

- A recall of 81.4% indicates that the model is successfully identifying 81% of the actual positive cases. It's missing only 19% of the actual positive cases.
- A 92% precision means that when your model predicts a positive case, it is correct 92% of the time. In other words, there are relatively few false positives.
- The combination of 81% recall and 92% precision indicates that your model is performing well, both in identifying actual positives and in making correct positive predictions.
- 86.5% F1 score seems to be a relatively high score reflecting a good balance between precision and recall.

Model performance improvement

- We deal with multicollinearity using the variation inflation factor (VIF).
- We drop column with VIF greater than 5.

Identify optimal threshold for Logistic Regression using ROC curve

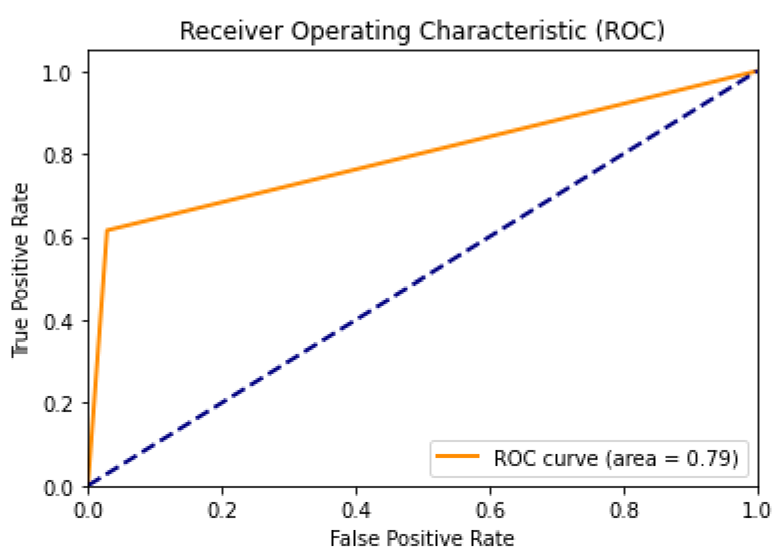
The Receiver Operating Characteristic (ROC) curve plots the true positive rate (recall) against the false positive rate at various threshold settings. The goal is to find a threshold that balances these rates.

Typically, the optimal threshold is chosen to maximize a metric such as the Youden's J statistic (sensitivity + specificity - 1), or it might be selected based on where the ROC curve is closest to the top-left corner of the plot.

The optimal threshold value from our model = 1

A threshold of 1 means that the model only predicts the positive class if the predicted probability is exactly 1. Since this is highly unlikely for most cases, it would result in the model predicting the negative class for almost all instances.

This could be due to the imbalance in the data and model overfitting.

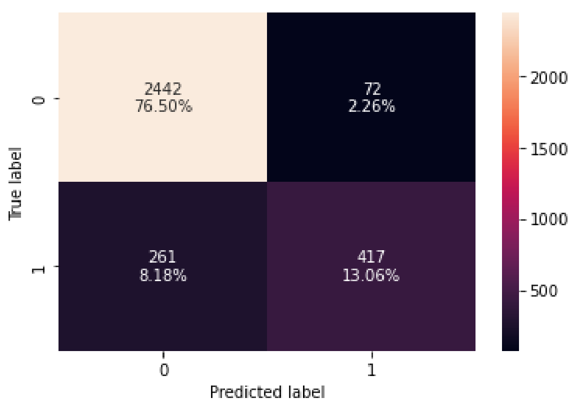


- A curve that is closer to the top-left corner (i.e., higher TPR and lower FPR) indicates a better performing model. The ideal point is (0, 1), representing 100% true positive rate and 0% false positive rate.

- For our model, the point is about (0.01, 0.6). This means that the model correctly identifies 60% of the actual positive cases and incorrectly identifies 1% of the actual negative cases as positive.
- This is a reasonable model.

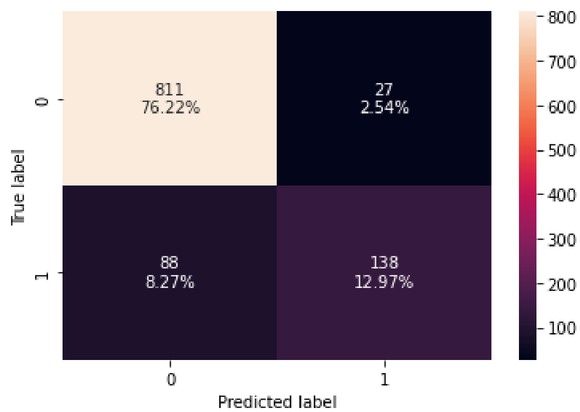
Tuned logistic regression model

Training data



	Accuracy	Recall	Precision	F1
1	0.895677	0.615044	0.852761	0.714653

Testing data



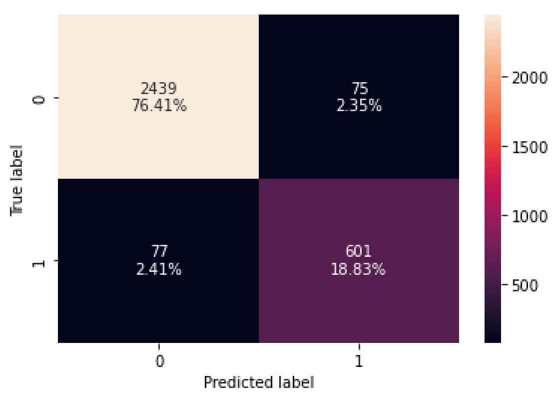
	Accuracy	Recall	Precision	F1
1	0.891917	0.610619	0.836364	0.705882

Model Performance Improvement - Random Forest

We perform hyperparameter tuning for the Random Forest Model using the GridSearchCV function.

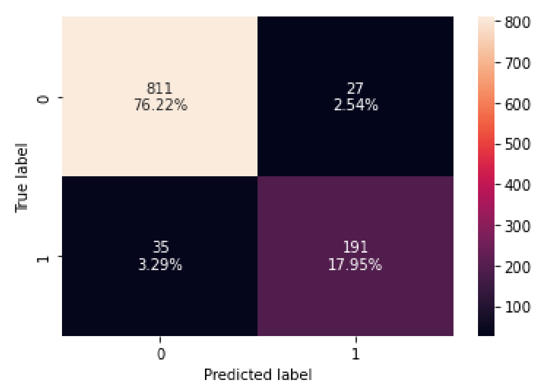
Tuned Random Forest Model

Training data



	Accuracy	Recall	Precision	F1
1	0.952381	0.886431	0.889053	0.88774

Testing data



	Accuracy	Recall	Precision	F1
1	0.941729	0.845133	0.876147	0.86036

Model Comparison and Final Model Selection

Training data

	Logistic Regression	Tuned Logistic Regression	Random Forest	Tuned Random Forest
Accuracy	0.899123	0.895677	1.0	0.952381
Recall	0.622419	0.615044	1.0	0.886431
Precision	0.864754	0.852761	1.0	0.889053
F1	0.723842	0.714653	1.0	0.887740

The untuned random forest model has a perfect score of 1 for all metrics, but this could be due to overfitting or data imbalance.

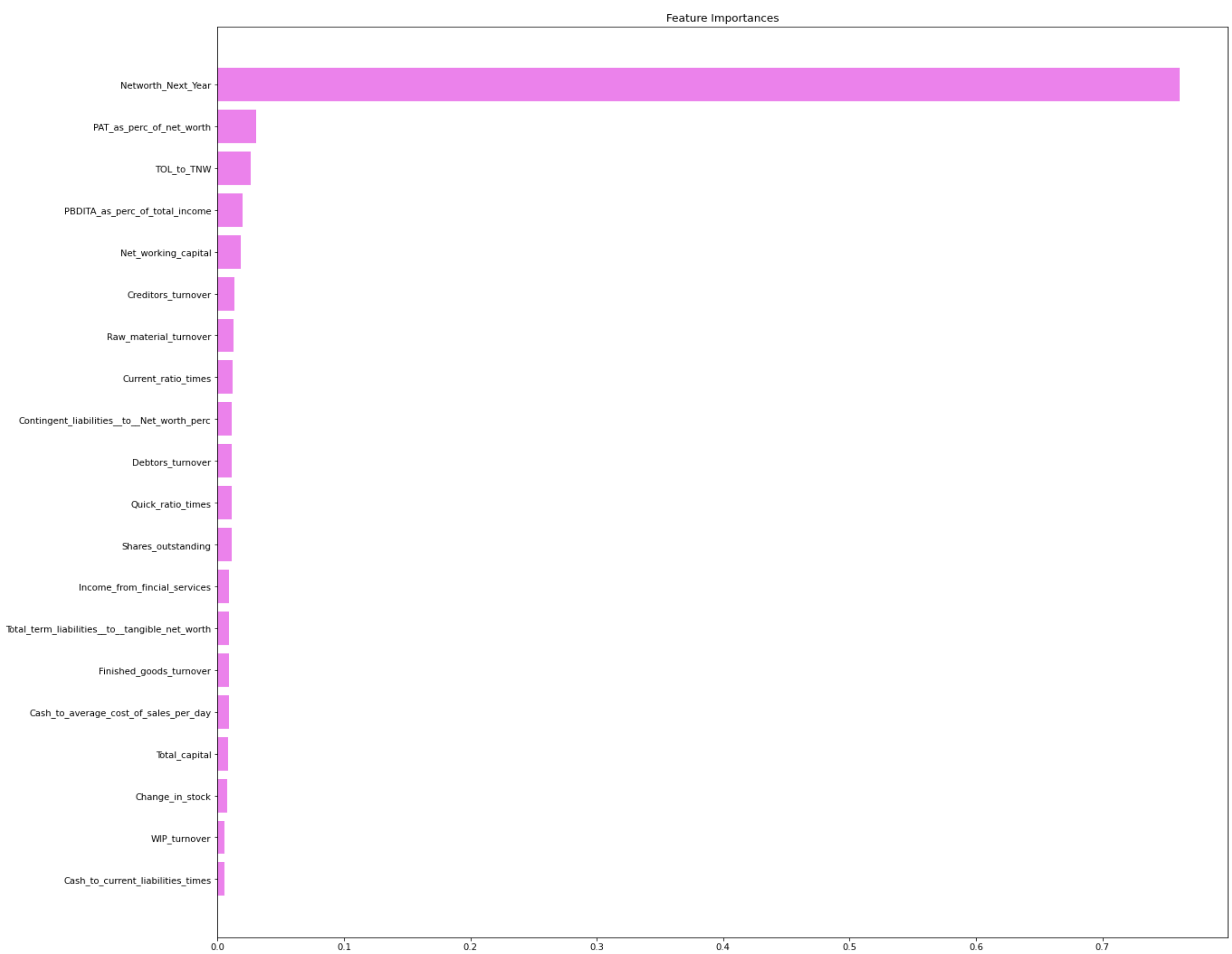
We can see that the tuned random forest model has a better recall, precision, and F1 score.

Testing data

	Logistic Regression	Tuned Logistic Regression	Random Forest	Tuned Random Forest
Accuracy	0.895677	0.891917	0.946429	0.941729
Recall	0.610619	0.610619	0.814159	0.845133
Precision	0.857143	0.836364	0.924623	0.876147
F1	0.713178	0.705882	0.865882	0.860360

For the testing data as well, the tuned random forest model has a better recall, precision, and F1 score.

Important features:



- Net worth next year, Profit after tax as percentage of net worth, Ratio of Total liabilities of the customer by Total net worth are the top three important features impacting the model performance.

Insights and recommendations

1. Model Performance and Reliability

- **Model Effectiveness**: The model demonstrates strong performance with high recall, precision, and F1 scores. This suggests that it effectively identifies both defaulters and non-defaulters with balanced trade-offs between sensitivity and specificity.
- **Recommendation**: The model can be reliably used in decision-making for credit risk assessment. However, continuous monitoring and recalibration are recommended to ensure sustained performance over time, especially as the data and market conditions evolve.

2. Top Contributing Features

- **Net Worth Next Year**: This feature's importance suggests that the projected financial stability of a customer is a key determinant of their creditworthiness.
- **Profit After Tax as a Percentage of Net Worth**: This indicates that profitability relative to the customer's net worth plays a critical role in assessing their ability to meet financial obligations.
- **Ratio of Total Liabilities to Total Net Worth**: A higher ratio signifies higher leverage, which increases the risk of default. This metric's importance underlines the need for careful monitoring of customer debt levels relative to their equity.
- **Recommendation**: Focus on these key financial metrics during the credit assessment process. Implement stronger monitoring and stress-testing for customers with concerning levels in these areas. Additionally, consider developing targeted financial products or interventions to support customers with high leverage or declining profitability.

3. Data Quality Management

- **Outliers and Missing Values**: The presence of outliers and missing data can distort model predictions and reduce reliability.
- **Recommendation**: Implement robust data cleaning processes, including the handling of missing values and outliers. Use imputation techniques for missing data, and consider transforming or capping outliers to mitigate their impact. Regularly update and validate the dataset to maintain its integrity.

4. Addressing Data Imbalance

- **Class Imbalance**: The majority of the dataset comprises non-defaulters, which can bias the model toward predicting non-defaults.
- **Recommendation**: Although your model is performing well, consider further techniques to address class imbalance, such as using resampling methods (e.g., SMOTE, under-sampling), or adjusting decision thresholds based on business objectives. Regularly evaluate the model's performance on minority classes to ensure it continues to accurately predict defaulters.

5. Operational Considerations

- **Integration into Business Processes**: Ensure that the model's predictions are effectively integrated into the credit decision-making workflow. Develop clear guidelines for interpreting model outputs and thresholds for action.
- **Recommendation**: Train relevant teams on how to use the model's insights and establish protocols for cases flagged as high-risk. Consider setting up automated alerts for customers who exhibit worsening trends in the top contributing features.

Part B

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 Dictionary

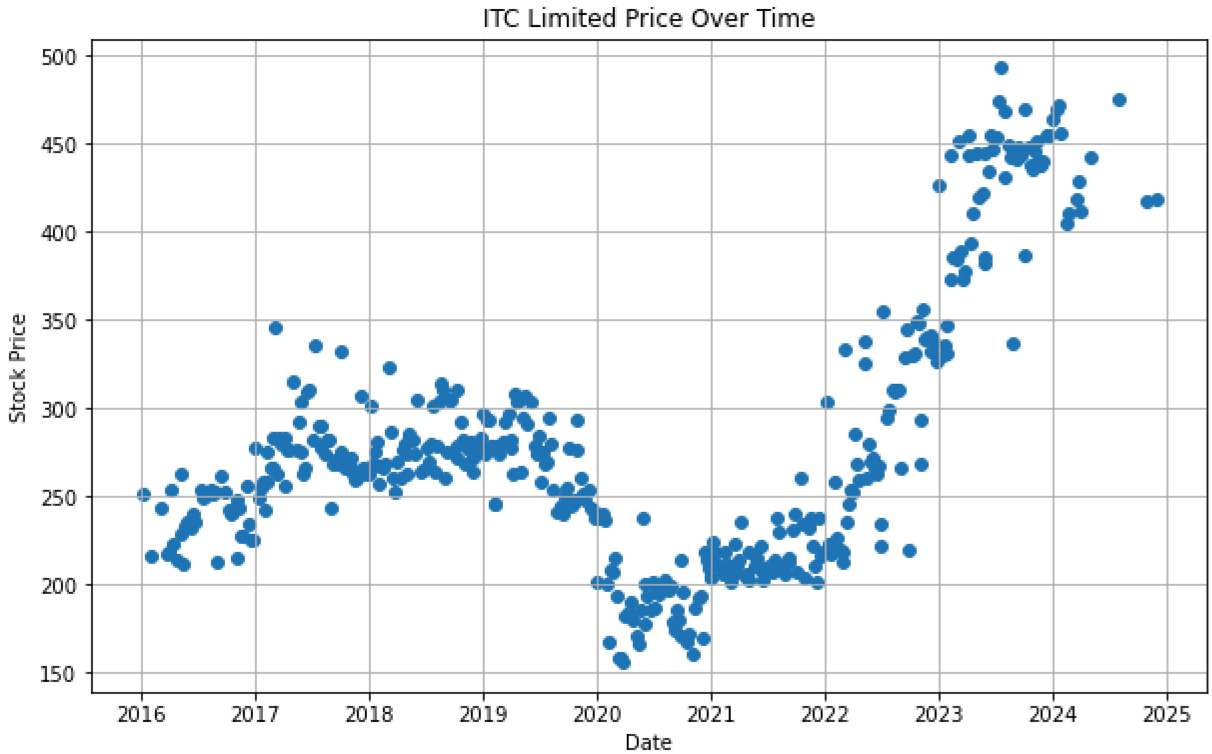
The dataset contains weekly stock price data for 5 Indian stocks over an 8-year period. The dataset enables us to analyze the historical performance of individual stocks and the overall market dynamics.

- There are 6 columns with 418 rows
- The date column is converted from object to date format and the other columns are of the integer type.
- There are no null values in the data.

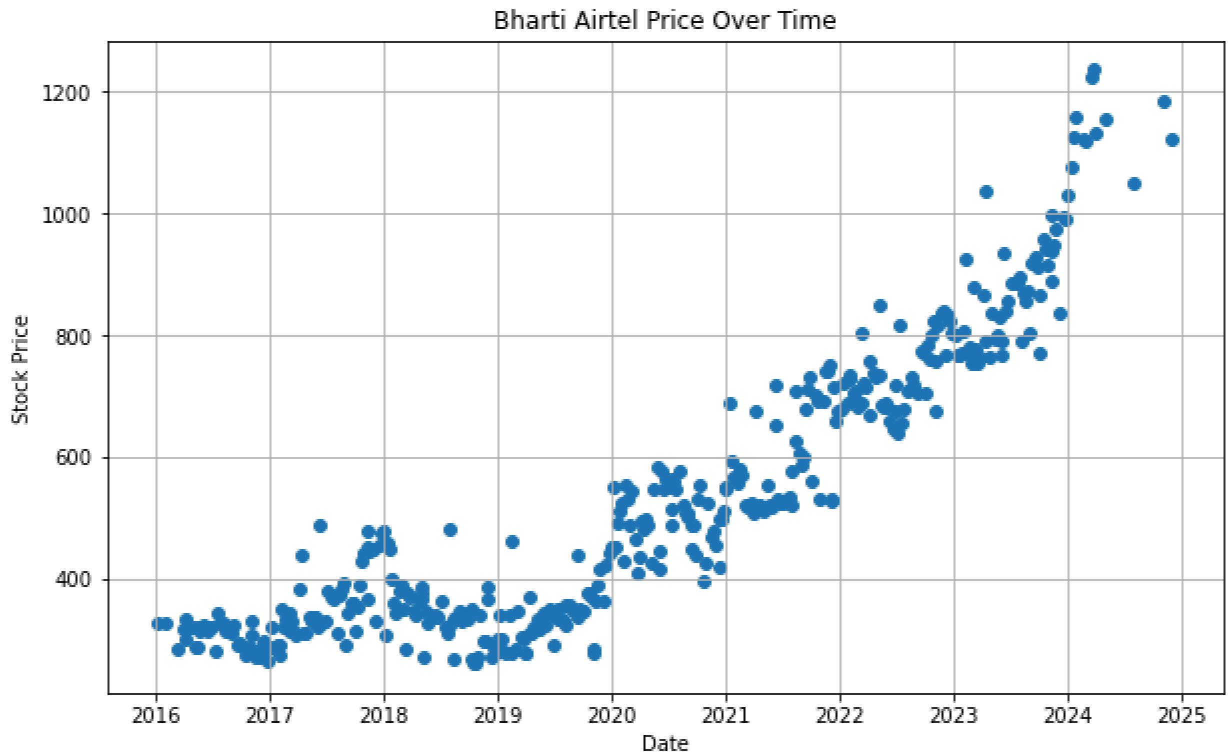
Statistical summary

	ITC Limited	Bharti Airtel	Tata Motors	DLF Limited	Yes Bank
count	418.000000	418.000000	418.000000	418.000000	418.000000
mean	278.964115	528.260766	368.617225	276.827751	124.442584
std	75.114405	226.507879	182.024419	156.280781	130.090884
min	156.000000	261.000000	65.000000	110.000000	11.000000
25%	224.250000	334.000000	186.000000	166.250000	16.000000
50%	265.500000	478.000000	399.500000	213.000000	30.000000
75%	304.000000	706.750000	466.000000	360.500000	249.750000
max	493.000000	1236.000000	1035.000000	928.000000	397.000000

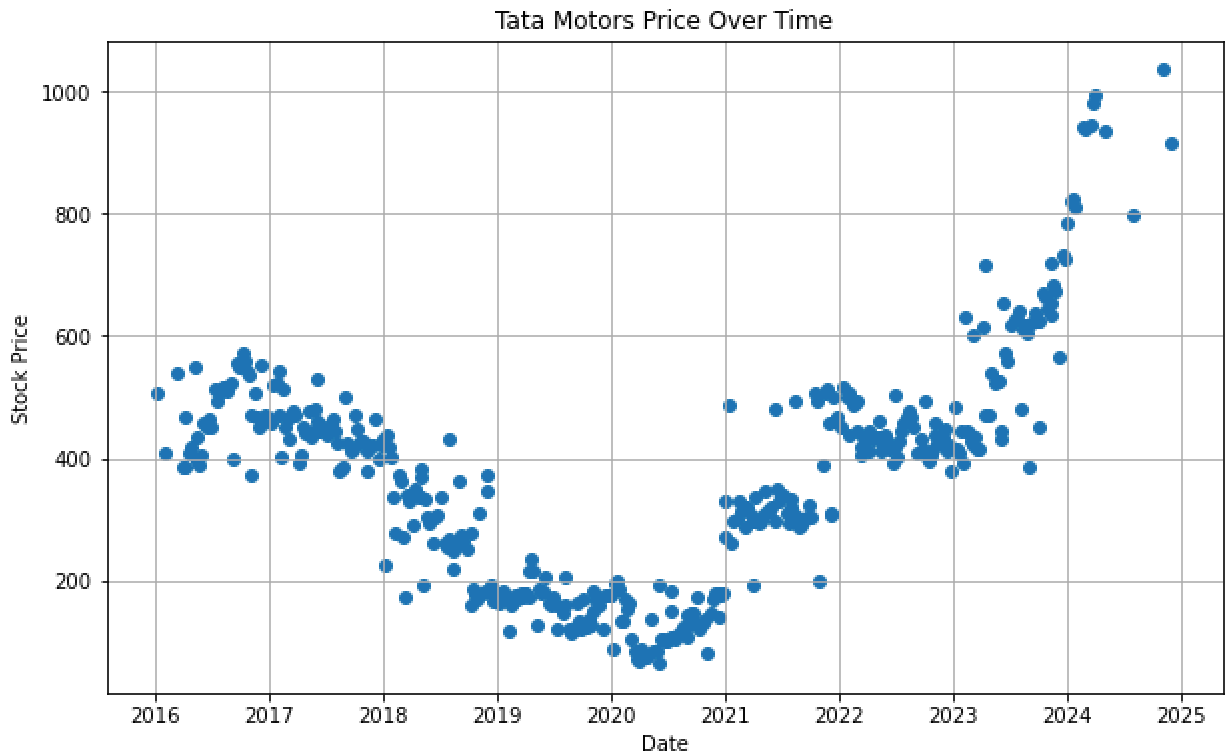
Stock Price Graph Analysis



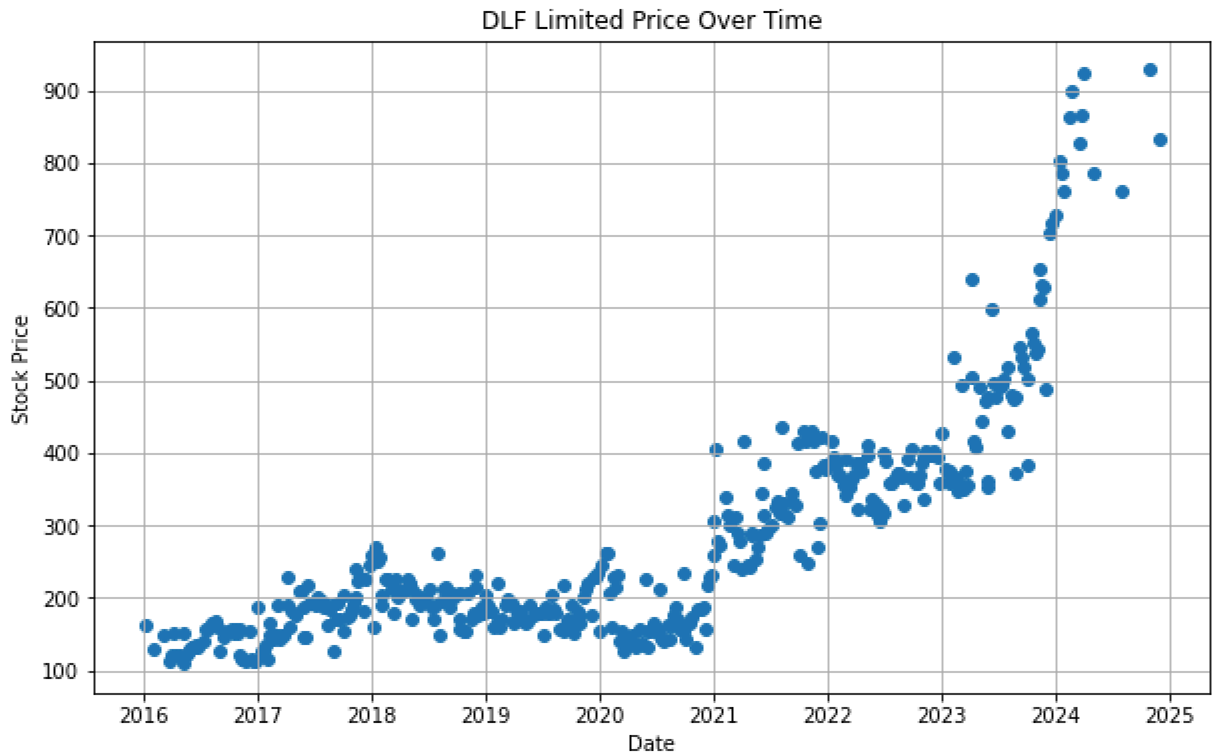
- Between 2016 to 2025, the ITC stock price hovered between Rs. 200 and Rs. 500.
- After the 2020 COVID market crash, the ITC stock price has steadily increased. Almost linearly.



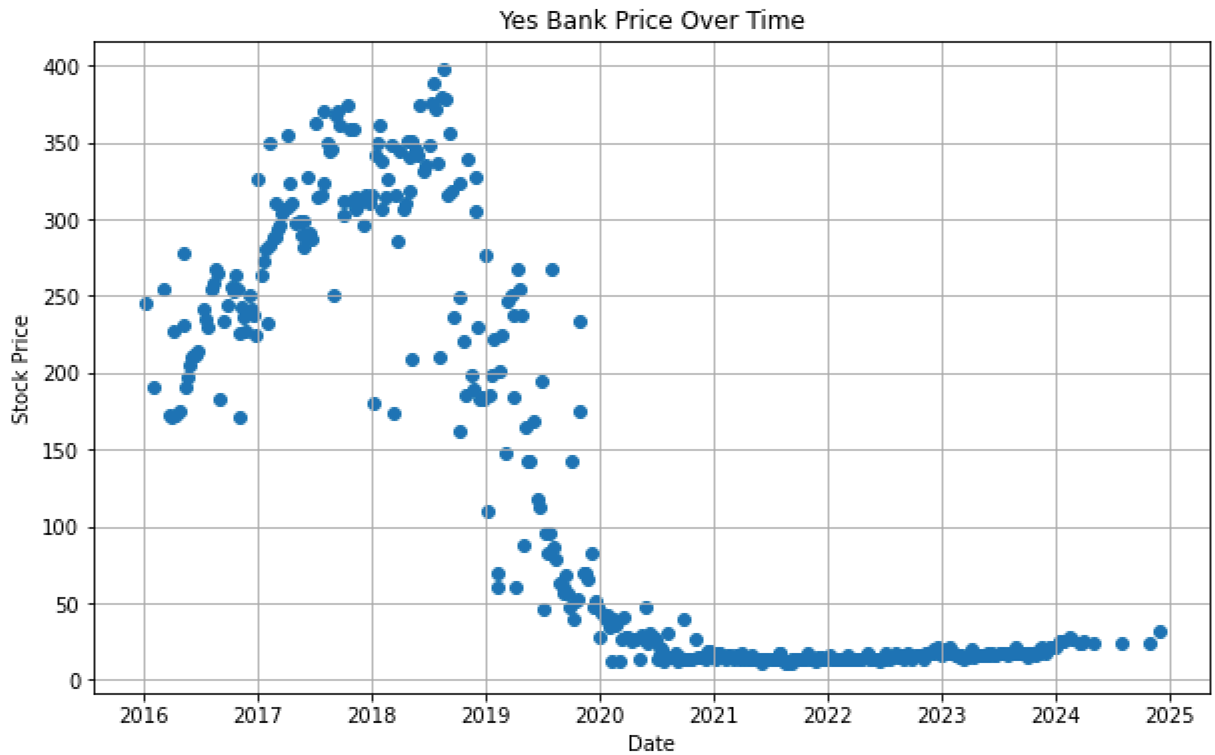
- From 2016 to 2020 the stock price has moved side ways between the range Rs. 300 to Rs. 500.
- Post 2020, there seems to be a linear increase in the stock price, hitting an all time high of about Rs. 1200.



- From 2016 to 2020, the Tata Motors stock price fell from 600 to 100.
- Post COIVD, the price has by and large moved upwards, i.e. from 2021 to 2025.
- However, the price movement each year has been sideway.



- The DLF stock price has remained stagnant from 2016 to 2021.
- Post 2021, there seems to be an exponential increase in the price of the stock, touching a price limit above Rs. 900



- The Yes Bank stock price has been volatile between 2106 to 2019.
- After the 2020 crash and the bank crisis due to bad loans, the stock price crashed below Rs. 50 and stagnated ever since.

Stock Returns Calculation and Analysis

Return of stocks

	ITC Limited	Bharti Airtel	Tata Motors	DLF Limited	Yes Bank
0	NaN	NaN	NaN	NaN	NaN
1	0.004598	-0.045315	0.000000	0.059592	-0.011628
2	-0.013857	0.019673	-0.031582	-0.008299	0.000000
3	0.036534	0.038221	0.087011	0.016529	0.005831
4	-0.041196	-0.003130	0.024214	0.000000	0.017291
5	0.009302	0.024769	-0.024214	0.055791	0.082238
6	-0.013986	-0.006135	-0.019803	-0.015625	-0.037538
7	-0.004706	-0.015504	-0.015114	-0.040166	0.042787
8	0.094452	-0.025318	-0.012772	0.016261	0.030930
9	0.012793	0.019048	0.040308	0.039531	0.039806
10	0.000000	0.012500	0.122982	0.030537	0.024098

Average returns

```
Yes Bank        -0.004737
ITC Limited      0.001634
Tata Motors      0.002234
Bharti Airtel    0.003271
DLF Limited      0.004863
```

- DLF Limited has the highest return and YES Bank the lowest return, in fact makes a loss.

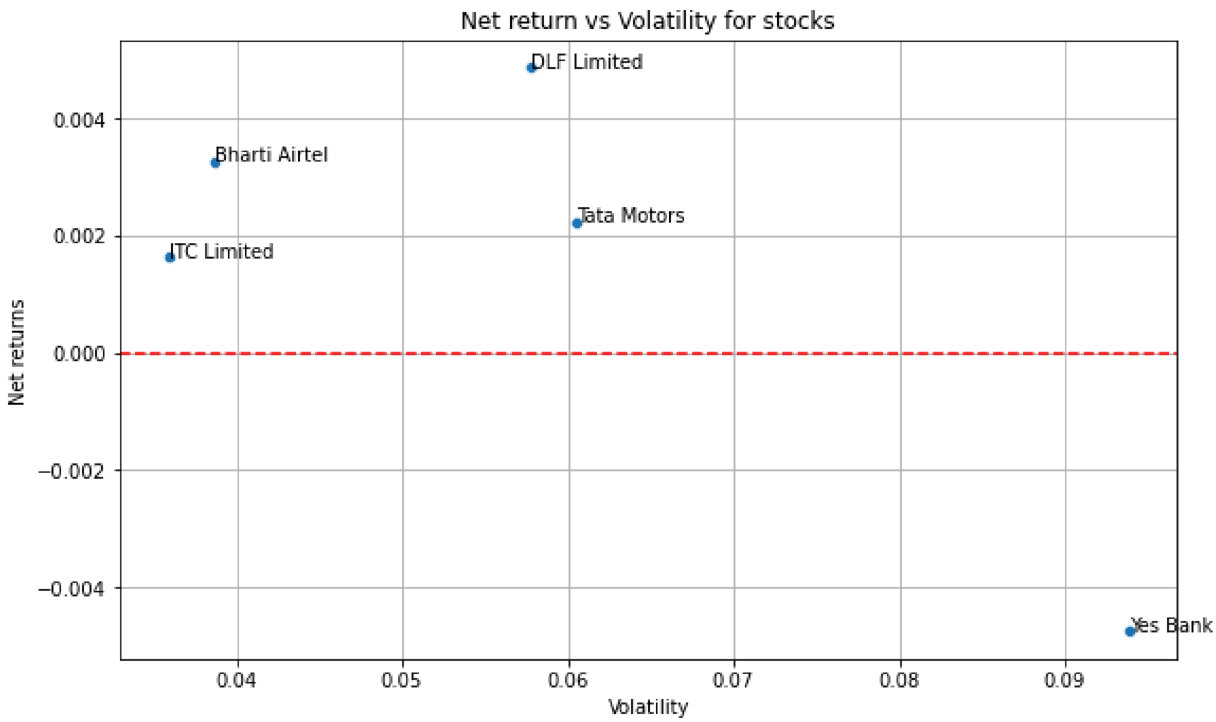
Volatility in stock price (standard deviation)

```
ITC Limited      0.035904
Bharti Airtel    0.038728
DLF Limited      0.057785
Tata Motors      0.060484
Yes Bank         0.093879
```

- The standard deviation of 0.035904 indicates that the returns of ITC stock are relatively less variable or dispersed. This means that ITC's returns tend to fluctuate within a narrower range compared to more volatile stocks
- The standard deviation of 0.093875 indicates that the returns of YES Bank stock are more variable or dispersed. This means YES Bank’s returns experience greater fluctuations and are more unpredictable..

Returns and Volatility

	Mean	StdDev
ITC Limited	0.001634	0.035904
Bharti Airtel	0.003271	0.038728
Tata Motors	0.002234	0.060484
DLF Limited	0.004863	0.057785
Yes Bank	-0.004737	0.093879



- The YES bank stock is highly volatile and also has a negative return.
- ITC is least volatile with a modest return of about 0.0015.
- Compared to ITC, Bharti Airtel has higher returns and comparable volatility.
- DLF has the highest net returns with a medium volatility.

Actionable insights and recommendations

Risk Assessment: YES Bank has higher volatility than ITC. This means that YES Bank's stock returns are subject to greater swings, indicating higher risk. Investors might experience larger gains or losses with YES Bank compared to ITC.

Investment Stability: ITC's lower volatility suggests a more stable investment with less risk of significant deviations in returns. This might be preferable for risk-averse investors seeking steadier performance.

Market Behavior: The higher volatility of YES Bank may reflect broader issues or market conditions affecting the bank, such as the crisis it faced. In contrast, the lower volatility of ITC suggests more stable market conditions for that stock.