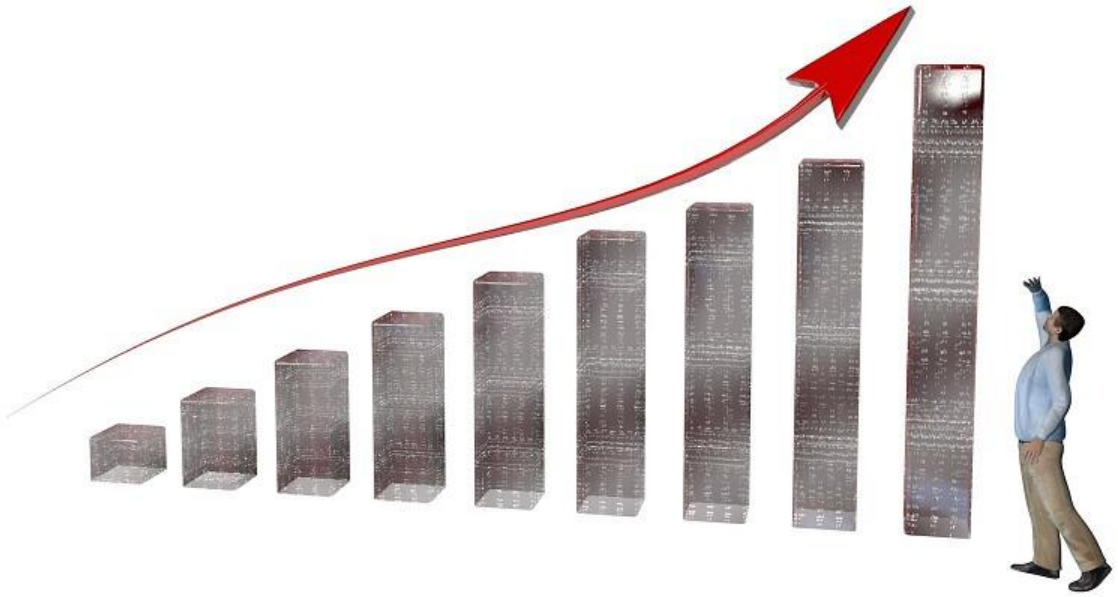
Financial Risk Analysis



Rasika Pande | 08/06/2023

Table of Contents

List of Figures

List of Tables

Problem Statement

Businesses or companies can fall prey to default if they are not able to keep up their debt obligations. Defaults will lead to a lower credit rating for the company which in turn reduces its chances of getting credit in the future and may have to pay higher interest on existing debts as well as any new obligations. From an investor's point of view, he would want to invest in a company if it is capable of handling its financial obligations, can grow quickly, and is able to manage the growth scale. A balance sheet is a financial statement of a company that provides a snapshot of what a company owns, owes, and the amount invested by the shareholders. Thus, it is an important tool that helps evaluate the performance of a business.

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

Test Train Split - Split the data into train and test datasets in the ratio of 6733 and use a random state of 42 (random_state=42). Model building is to be done on the train dataset and model validation is to be done on the test dataset.

Part A

Data

The data provided by companies contains 58 columns (features / variables) and 2058 observations of data from the previous financial year. There are 53 columns with the float data type, 4 integer data type columns and 1 object data-type column. Using the info function (table 1 below), we can also see that columns have missing values such as the 'Cash_Flow_Per_Share' column. Before proceeding with any outlier treatment or exploratory data analysis we have fixed the column names using the replace functions and dropped the Co_Code and Co_Name variables as they will not be useful to our analysis.

#	Column	Non-Null Count	Dtype
0	Co_Code	2058 non-null	int64
1	Co_Name	2058 non-null	object
2	Operating_Expense_Rate	2058 non-null	float64

3	Research_and_development_expense_rate	2058	non-null	float64
4	Cash_flow_rate	2058	non-null	float64
5	Interest_bearing_debt_interest_rate	2058	non-null	float64
6	Tax_rate_A	2058	non-null	float64
7	Cash_Flow_Per_Share	1891	non-null	float64
8	Per_Share_Net_profit_before_tax_Yuan	2058	non-null	float64
9	Realized_Sales_Gross_Profit_Growth_Rate	2058	non-null	float64
10	Operating_Profit_Growth_Rate	2058	non-null	float64
11	Continuous_Net_Profit_Growth_Rate	2058	non-null	float64
12	Total_Asset_Growth_Rate	2058	non-null	float64
13	Net_Value_Growth_Rate	2058	non-null	float64
14	Total_Asset_Return_Growth_Rate_Ratio	2058	non-null	float64
15	Cash_Reinvestment_perc	2058	non-null	float64
16	Current_Ratio	2058	non-null	float64
17	Quick_Ratio	2058	non-null	float64
18	Interest_Expense_Ratio	2058	non-null	float64
19	Total_debt_to_Total_net_worth	2037	non-null	float64
20	Long_term_fund_suitability_ratio_A	2058	non-null	float64
21	Net_profit_before_tax_to_Paid_in_capital	2058	non-null	float64
22	Total_Asset_Turnover	2058	non-null	float64
23	Accounts_Receivable_Turnover	2058	non-null	float64
24	Average_Collection_Days	2058	non-null	float64
25	Inventory_Turnover_Rate_times	2058	non-null	float64
26	Fixed_Assets_Turnover_Frequency	2058	non-null	float64
27	Net_Worth_Turnover_Rate_times	2058	non-null	float64
28	Operating_profit_per_person	2058	non-null	float64
29	Allocation_rate_per_person	2058	non-null	float64
30	Quick_Assets_to_Total_Assets	2058	non-null	float64
31	Cash_to_Total_Assets	1962	non-null	float64
32	Quick_Assets_to_Current_Liability	2058	non-null	float64
33	Cash_to_Current_Liability	2058	non-null	float64
34	Operating_Funds_to_Liability	2058	non-null	float64
35	Inventory_to_Working_Capital	2058	non-null	float64
36	Inventory_to_Current_Liability	2058	non-null	float64
37	Long_term_Liability_to_Current_Assets	2058	non-null	float64
38	Retained_Earnings_to_Total_Assets	2058	non-null	float64
39	Total_income_to_Total_expense	2058	non-null	float64
40	Total_expense_to_Assets	2058	non-null	float64
41	Current_Asset_Turnover_Rate	2058	non-null	float64
42	Quick_Asset_Turnover_Rate	2058	non-null	float64
43	Cash_Turnover_Rate	2058	non-null	float64
44	Fixed_Assets_to_Assets	2058	non-null	float64
45	Cash_Flow_to_Total_Assets	2058	non-null	float64
46	Cash_Flow_to_Liability	2058	non-null	float64
47	CFO_to_Assets	2058	non-null	float64

```
 48  Cash_Flow_to_Equity                                    2058 non-null   float64
 49  Current_Liability_to_Current_Assets                    2044 non-null   float64
 50  Liability_Assets_Flag                                  2058 non-null   int64
 51  Total_assets_to_GNP_price                              2058 non-null   float64
 52  No_credit_Interval                                     2058 non-null   float64
 53  Degree_of_Financial_Leverage_DFL                       2058 non-null   float64
 54  Interest_Coverage_Ratio_Interest_expense_to_EBIT       2058 non-null   float64
 55  Net_Income_Flag                                        2058 non-null   int64
 56  Equity_to_Liability                                    2058 non-null   float64
 57  Default                                                2058 non-null   int64
dtypes: float64(53), int64(4), object(1)
```

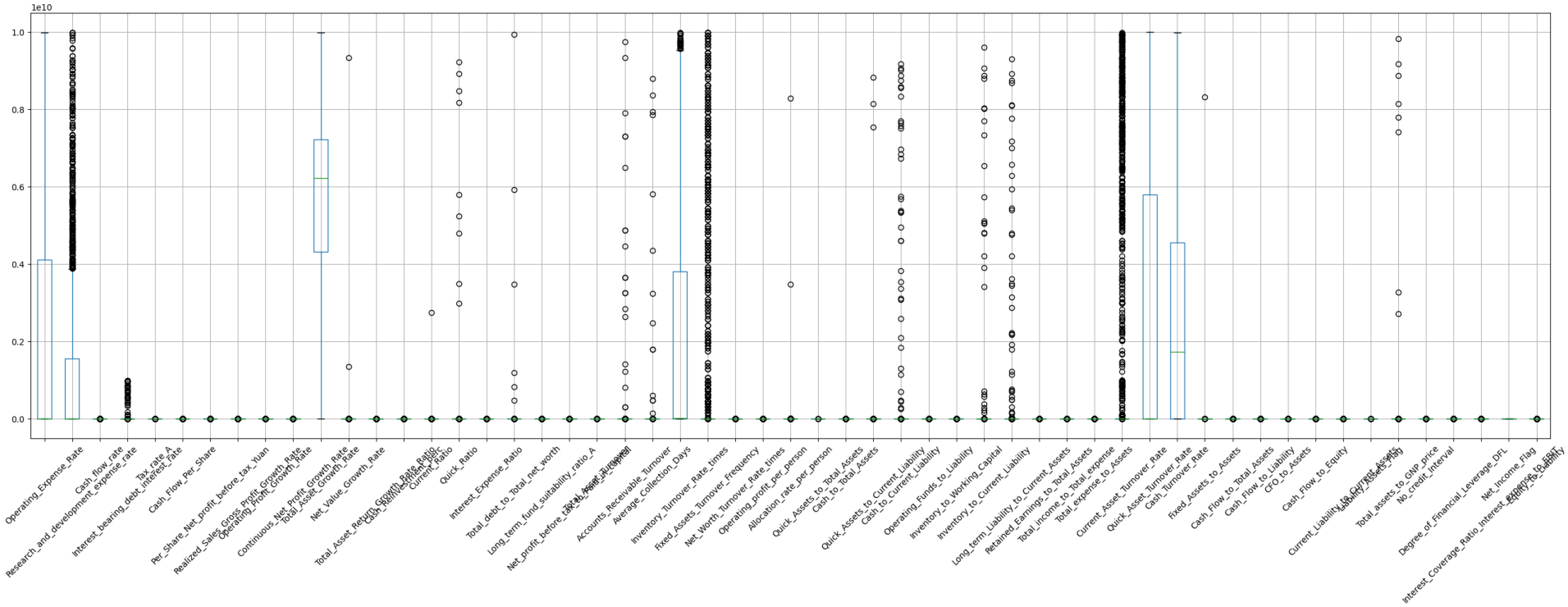
Table 1 Table of Data Info

	count	mean	std	min	25%	50%	75%	max
Operating_Expense_Rate	2058.00	2052388835.76	3252623690.29	0.00	0.00	0.00	4110000000.00	9980000000.00
Research_and_development_expense_rate	2058.00	1208634256.56	2144568158.08	0.00	0.00	0.00	1550000000.00	9980000000.00
Cash_flow_rate	2058.00	0.47	0.02	0.00	0.46	0.46	0.47	1.00
Interest_bearing_debt_interest_rate	2058.00	11130223.52	90425949.04	0.00	0.00	0.00	0.00	990000000.00
Tax_rate_A	2058.00	0.11	0.15	0.00	0.00	0.04	0.22	1.00
Cash_Flow_Per_Share	1891.00	0.32	0.02	0.17	0.31	0.32	0.33	0.46
Per_Share_Net_profit_before_tax_Yuan	2058.00	0.18	0.03	0.00	0.17	0.18	0.19	0.79
Realized_Sales_Gross_Profit_Growth_Rate	2058.00	0.02	0.02	0.00	0.02	0.02	0.02	1.00
Operating_Profit_Growth_Rate	2058.00	0.85	0.00	0.74	0.85	0.85	0.85	1.00
Continuous_Net_Profit_Growth_Rate	2058.00	0.22	0.01	0.00	0.22	0.22	0.22	0.23
Total_Asset_Growth_Rate	2058.00	5287663257.05	2912614769.58	0.00	4315000000.00	6225000000.00	7220000000.00	9980000000.00
Net_Value_Growth_Rate	2058.00	5189504.37	207791797.86	0.00	0.00	0.00	0.00	9330000000.00
Total_Asset_Return_Growth_Rate_Ratio	2058.00	0.26	0.00	0.25	0.26	0.26	0.26	0.36
Cash_Reinvestment_perc	2058.00	0.38	0.03	0.03	0.37	0.38	0.39	1.00
Current_Ratio	2058.00	1336248.80	60619173.20	0.00	0.01	0.01	0.01	2750000000.00
Quick_Ratio	2058.00	27755102.05	444865390.47	0.00	0.00	0.01	0.01	9230000000.00
Interest_Expense_Ratio	2058.00	0.63	0.01	0.53	0.63	0.63	0.63	0.81
Total_debt_to_Total_net_worth	2037.00	10714285.73	269696017.59	0.00	0.00	0.01	0.01	9940000000.00
Long_term_fund_suitability_ratio_A	2058.00	0.01	0.03	0.00	0.01	0.01	0.01	1.00
Net_profit_before_tax_to_Paid_in_capital	2058.00	0.18	0.03	0.00	0.17	0.17	0.18	0.79
Total_Asset_Turnover	2058.00	0.13	0.10	0.00	0.06	0.10	0.17	0.92
Accounts_Receivable_Turnover	2058.00	41598639.46	504767266.59	0.00	0.00	0.00	0.00	9740000000.00
Average_Collection_Days	2058.00	26297862.01	410996733.83	0.00	0.00	0.01	0.01	8800000000.00
Inventory_Turnover_Rate_times	2058.00	2030227259.48	3077250265.27	0.00	0.00	19100000.00	3815000000.00	9990000000.00

Table 2 Data Summary of Features

Outlier Treatment

There are multiple outliers present in the dataset.



In the provided dataset, we will proceed with transforming the outliers to missing values and then using the KNN imputer. Outliers are extreme data points that can distort analysis and model performance and using the KNN imputer is a good way to bring outliers/missing data to values closer to their dataset. Other common outlier handling techniques, such as removal, winsorization, or transformations, can be considered based on data characteristics and analysis goals.

We will find any outliers using the IQR method and then replace values above the IQR range with null values.

Missing Value Treatment

Before treating the outliers, there were only 3 columns with missing values however after outlier treatment there are 11162 missing values in the entire dataset. We will visualize the missing values in each row/column using a heatmap before proceeding with missing value treatment. Since there are too many missing values we first will check if we can only proceed with a dataset based on the availability of data.

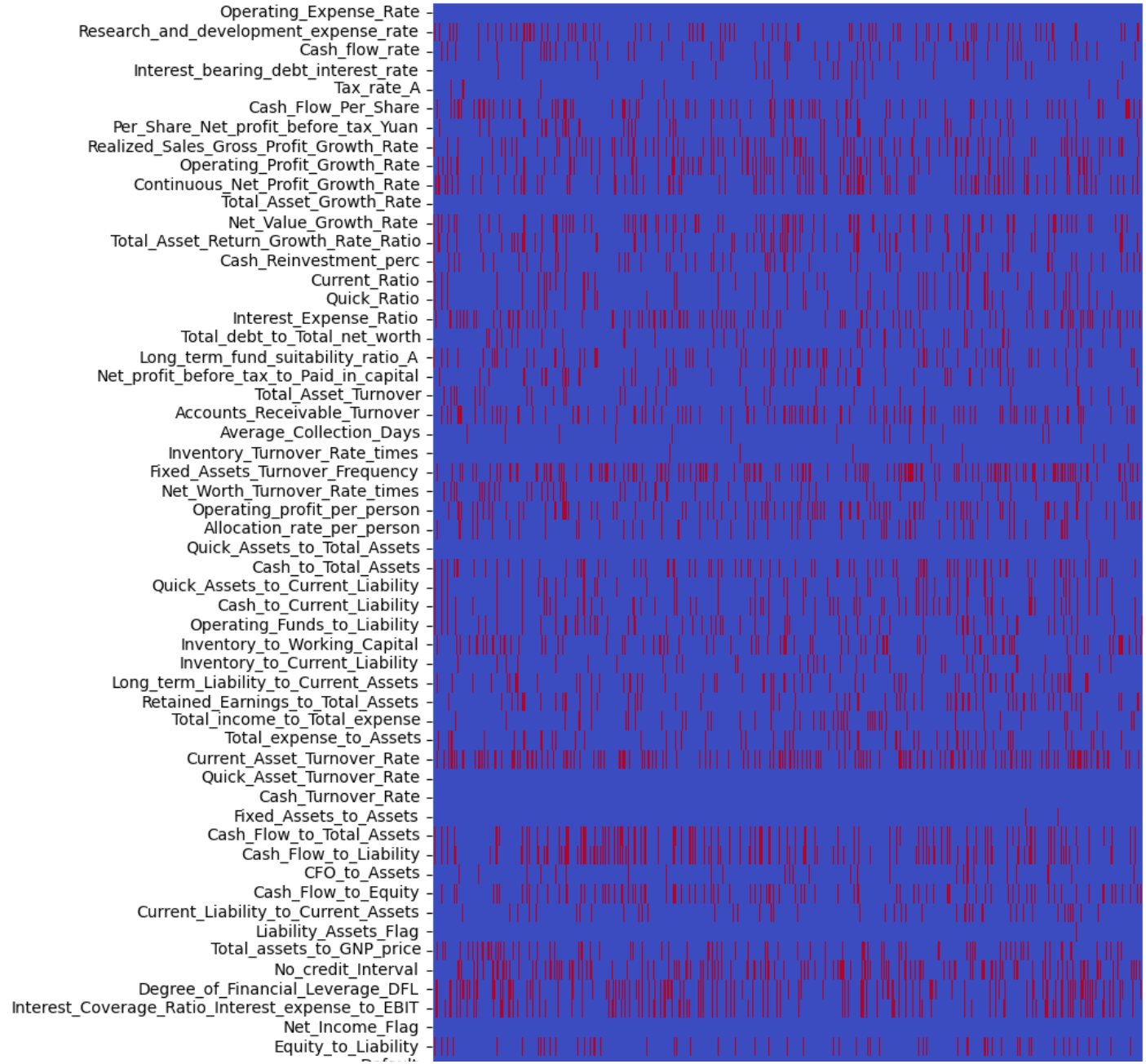


Figure 1 Heatmap of Missing Values

If we only took the rows where there were less than 5 null values, we would be almost halving the dataset and it would not be a good sample of data as there are only 0.07% of defaulters based on data availability. So we will proceed with using a KNN imputer.

Missing Values were imputed using a KNN imputer.

Train Test Split

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

Validate the Model on Test Dataset and state the performance metrics. Also state interpretation from the model

Build a Random Forest Model on Train Dataset. Also showcase your model building approach

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

Build a LDA Model on Train Dataset. Also showcase your model building approach

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

Compare the performances of Logistic Regression, Random Forest, and LDA models (include ROC curve)

Conclusions and Recommendations

Part B

Draw Stock Price Graph (Stock Price vs Time) for any 2 given stocks with inference

The 2 stocks plotted in the graph below are for Idea-Vodafone and Jet Airways. The two stocks are plotted from 2014-2021 and while Idea-Vodafone shows a slow but steady decline across the years with some peaks in each month, Jet Airways has a much more volatile decline. The stock of Jet Airways peaked at the end of 2017/ beginning of 2021 at over 800 but then decreased drastically with peaks and dips. For an investor who is looking for a stable stock, they should not invest in either of these companies.

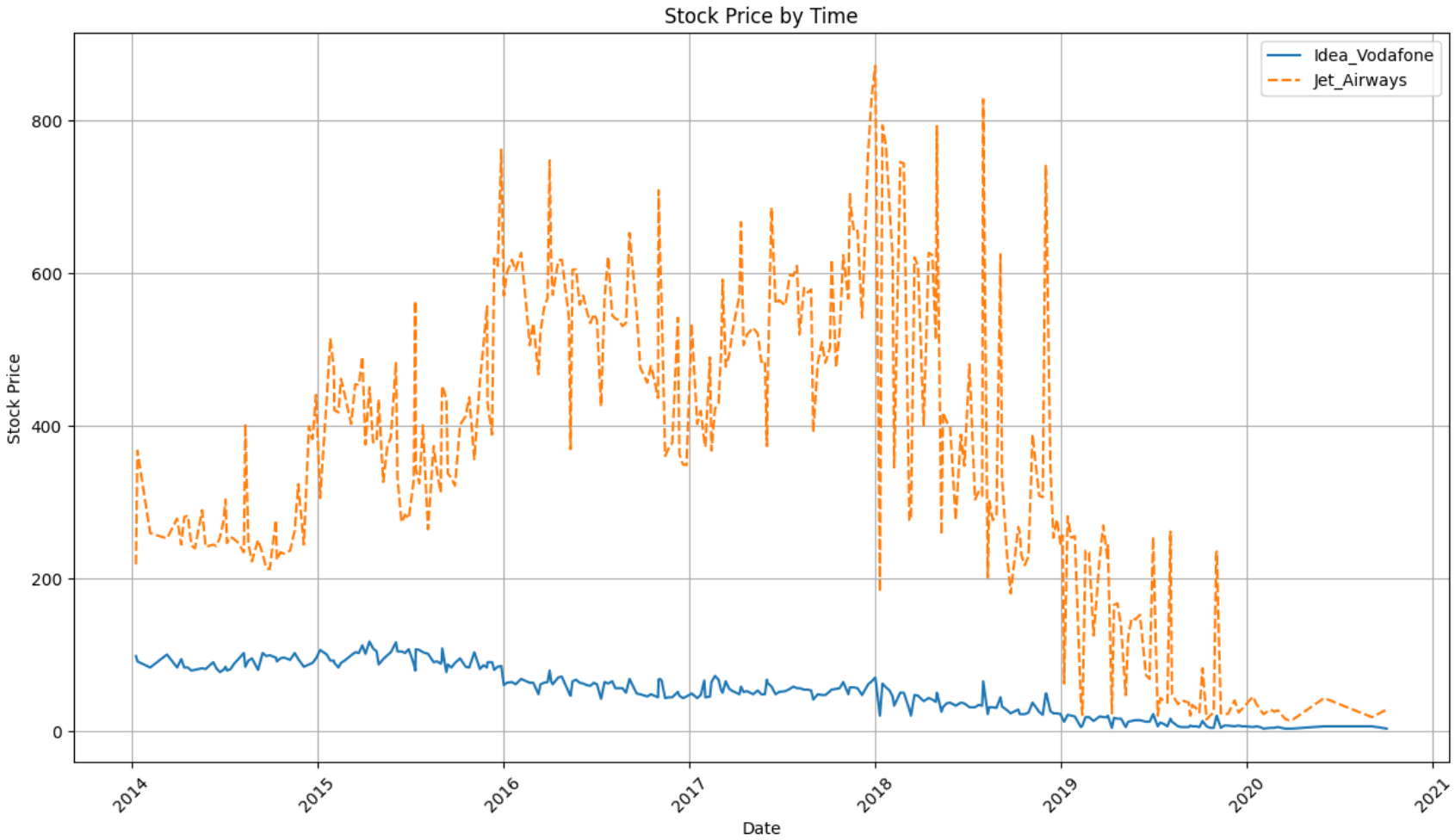


Figure 2 Stock Price by Time (2 Stocks)

Calculate Returns for all stocks with inference

When calculating returns for all the stocks, the calculation featured a daily returns, monthly returns and yearly returns to better understand how the stock performed between 2014 – 2020. Since most investors do not sell a stock in a day or two after purchase it is important to understand how the stock performs monthly or yearly.

Date	Infosys	Indian_Hotel	Mahindra_and_Mahindra	Axis_Bank	SAIL	Shree_Cement	Sun_Pharma	Jindal_Steel	Idea_Vodafone	Jet_Airways
2014-03-31	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2014-07-04	-0.03	-0.01	0.01	0.05	0.03	0.03	0.10	-0.06	0.01	0.09
2014-04-14	-0.01	0.00	-0.01	-0.02	-0.03	-0.01	-0.00	0.00	-0.01	-0.08
2014-04-21	-0.00	0.00	0.07	0.05	0.00	0.01	-0.00	-0.02	0.00	0.01
2014-04-28	0.01	-0.04	-0.01	-0.00	-0.07	-0.02	0.01	-0.13	-0.05	-0.14
...	...	...	...	...	...	...	...	...	...	...
2020-02-03	0.01	-0.10	0.03	-0.06	-0.08	0.02	0.08	-0.05	-0.25	-0.12
2020-09-03	-0.13	-0.05	-0.09	-0.14	-0.09	-0.08	-0.04	-0.17	1.00	-0.18
2020-03-16	-0.09	-0.21	-0.25	-0.25	-0.10	-0.11	-0.05	-0.13	-0.50	-0.11
2020-03-23	0.12	-0.17	-0.09	-0.16	-0.22	-0.07	-0.07	-0.15	0.00	-0.12
2020-03-30	-0.02	0.00	-0.03	0.05	0.10	-0.01	0.04	-0.08	0.00	0.00

Table 3 Daily Returns

Based on the return numbers we can observe that the change is often minimal to insignificant on a daily basis and often greater on the yearly if the stock is volatile. For example, Jindal Steel had negative returns on their stock during years 2014 and 2015 but saw a significant increase in price in 2017 similar to Jet Airways and SAIL. Shree Cement on the other hand based on their yearly returns year on year seems to be a stable stock to invest in with little fluctuations in the first 5 years.

Table 4 Monthly Returns

Date	Infosys	Indian_Hotel	Mahindra_and_Mahindra	Axis_Bank	SAIL	Shree_Cement	Sun_Pharma	Jindal_Steel	Idea_Vodafone	Jet_Airways
2014-01-31	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2014-02-28	-0.32	-0.23	-0.07	-0.25	0.32	-0.18	-0.27	1.21	-0.09	-0.29
2014-03-31	-0.01	-0.22	-0.19	-0.28	-0.35	-0.26	-0.05	-0.12	0.00	0.07
2014-04-30	-0.03	-0.06	0.06	0.07	-0.07	0.01	0.10	-0.20	-0.05	-0.13
2014-05-31	-0.09	0.26	0.17	0.20	0.33	0.20	-0.03	0.24	0.03	-0.01
...	...	...	...	...	...	...	...	...	...	...
2020-06-30	0.15	0.83	0.92	0.95	1.17	0.32	0.25	1.16	1.00	2.07
2020-07-31	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2020-08-31	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2020-09-30	-0.13	-0.17	-0.22	-0.23	-0.40	-0.08	-0.13	-0.32	0.00	-0.58
2020-10-31	0.22	0.17	0.22	0.30	0.47	0.14	0.08	0.58	-0.50	0.56

Table 5 Yearly Returns

Date	Infosys	Indian_Hotel	Mahindra_and_Mahindra	Axis_Bank	SAIL	Shree_Cement	Sun_Pharma	Jindal_Steel	Idea_Vodafone	Jet_Airways
2014-12-31	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2015-12-31	0.22	0.00	0.05	-0.12	-0.41	0.22	-0.01	-0.42	-0.11	0.73
2016-12-31	-0.04	-0.18	-0.05	0.01	0.04	0.30	-0.23	-0.27	-0.47	-0.54
2017-12-31	0.08	0.27	0.28	0.27	0.86	0.23	-0.09	1.97	0.44	1.39
2018-12-31	0.29	0.22	-0.03	0.10	-0.43	-0.07	-0.24	-0.26	-0.66	-0.71
2019-12-31	0.18	0.02	-0.26	0.20	-0.10	0.27	0.03	0.13	-0.73	-0.86
2020-12-31	0.05	-0.08	-0.02	-0.01	-0.06	0.16	-0.06	0.11	-0.50	-0.18

Calculate Stock Means and Standard Deviation for all stocks with inference

Stock Name	Mean	Standard Deviation
Infosys	511.340764	135.952051
Indian_Hotel	114.560510	22.509732
Mahindra_and_Mahindra	636.678344	102.879975
Axis_Bank	540.742038	115.835569
SAIL	59.095541	15.810493
Shree_Cement	14806.410828	4288.275085
Sun_Pharma	633.468153	171.855893
Jindal_Steel	147.627389	65.879195
Idea_Vodafone	53.713376	31.248985
Jet_Airways	372.659236	202.262668

Table 6 Table of Stock Means & Standard Deviation

The mean stock of the companies indicates how their stock fared over 7 years and the standard deviation indicates how far from the mean the stock moved overall. We would ideally want to see a smaller standard deviation for a larger mean as that should indicate stability. Indian Hotel/ Shree Cement are good examples of this as they have a much smaller standard deviation compared to their mean. By comparison, stocks like Jet Airways or Idea Vodafone that have higher standard deviations might not be considered good stocks to invest in.

Draw a plot of Stock Means vs Standard Deviation and state your inference

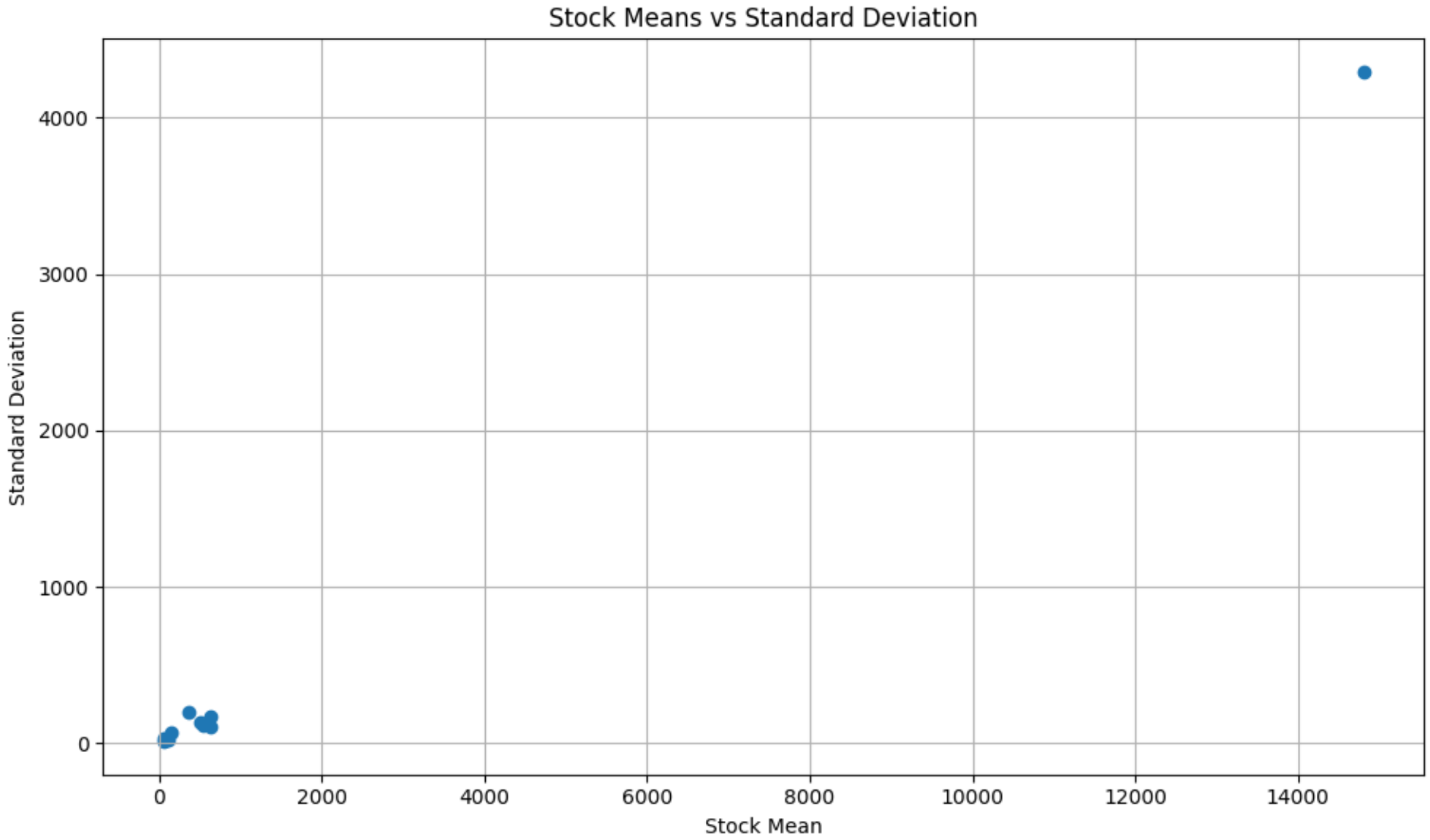


Figure 3 Graph of Stock Means vs Standard Deviation

The stock mean vs standard deviation chart was made to be a scatter plot to identify large distances between companies however we were able to see that most companies have a stock mean / standard deviation close to 0. The outlier in this case is Shree Cement with a large stock mean and standard deviation. The chart can also represent that companies who have a stock mean less than 200, will also have a smaller standard deviation making them less risky stocks to invest in because the fluctuations in return may be lower as compared to stocks like Shree Cement who have a large stock mean and a large standard deviation.

Conclusions and Recommendations

1. It is better to research stocks in a monthly or yearly period as you will be able to see the trend of the stock over a period of time since daily returns are fluctuated too much.
2. A higher standard deviation over a period of time may indicate a volatile stock and may be more of a risk to invest in if the stock mean is low.
3. A high standard deviation and high stock mean can be a high risk or reward.