

**Interim Project report on**  
**Factor Investing in Indian Equities**  
**Submitted By**  
**Group No. 8 [Batch: Oct'2019, Location: Bangalore]**

**Group Members**

- 1. Shivkumar M**
- 2. Jithesh Janardhanan**
- 3. Vinod Kumar D**
- 4. Tarique Zaman**

**Research Supervisor: Pranov Mishra**

**Great Lakes Institute of Management**



## Contents

1.	Introduction .....	3
2.	Problem Statement, Scope and Objective.....	3
3.	Data Source and Description .....	4
4.	Approach to Problem Statement .....	7
5.	Consumer Macro Economic Sector.....	8
6.	Commodities Macro Economic Sector .....	16
7.	Financial Services Macro Economic Sector .....	26
8.	Manufacturing Macro Economic Sector.....	44
10	Actionable insights and recommendations to the stakeholder .....	56
8.	References and Bibliography .....	56
9.	Appendix.....	58

## 1. Introduction

The project is an attempt to enhance the Factor-Based Investment Model. Factor investing is a strategy that chooses securities on attributes that are associated with higher returns.

The concept of using historical data and identifying factors that influence stock price or returns has been in the forefront of research to predict stock prices and maximize returns. Traditionally either fundamental or technical analysis were used to pick stocks but there was a need for a scalable and integrated approach to pick a set of stocks for maximum returns. It started with the Capital Pricing Asset Model (CAPM) and later the Arbitrage pricing theory which basically established the linear relationship to different indices and factors. The Fama-French 3 factor model expanded on this to initially prove that a combination of small cap, a high book to market ratio and risk-free returns over time, yields better returns. Two other factors (Momentum and Quality) were later added to explain the variation better. A lot of research and academic work has already been performed to predict stock prices / returns, however, the financial products in the market today limited to sector or style based.

The industry standard factors are Value, Size, Momentum, Volatility and Quality. However, Factor-Based Investment strategy does not take a holistic view, i.e. all the factors and the underlying variables and its impact on the Earnings per Share of a stock. It is this gap that the project is looking to address.

## 2. Problem Statement, Scope and Objective

The main objective is to build a model which combines various factors and its underlying variables to predict the Adjusted Earnings Per Share on an individual stock. Initially the study is limited to nifty stocks - National Stock Exchange (NSE) (top 500 stocks).

Following are the 5 Steps to achieve the defined objectives: -

### Step 1 – Collection of data: Collecting historical data on Nifty stocks:

Data points related to the Company Financial Results (historical financial results) collected are: -

Free Cash flow	Price to Cash Flow Ratio	EBITM (%)	Payable days
Adjusted PE (x)	Free Cash Flow per Share	Pre Tax Margin(%)	Net Sales Growth(%)
PCE(x)	Price to Free Cash Flow	PATM (%)	Core EBITDA Growth(%)
Price / Book Value(x)	Free Cash Flow Yield	CPM(%)	EBIT Growth(%)
Dividend Yield(%)	Sales to cash flow ratios	ROA (%)	PAT Growth(%)
EV/Net Sales(x)	Earnings Per Share (Rs)	ROE (%)	Adj. EPS Growth(%)
EV/EBITDA(x)	Adjusted EPS (Rs.)	ROCE (%)	Total Debt/Equity(x)
EV/EBIT(x)	CEPS(Rs)	Asset Turnover(x)	Current Ratio(x)
EV/CE(x)	DPS(Rs)	Inventory Turnover(x)	Quick Ratio(x)
M Cap / Sales	Adj DPS(Rs)	Debtors Turnover(x)	Interest Cover(x)
High PE	Book Value (Rs)	Fixed Asset Turnover (x)	Total Debt/Mcap(x)
Low PE	Adjusted Book Value (Rs)	Sales(x)/Working Capital	
Net Sales	Tax Rate(%)	Fixed Capital/Sales(x)	

Profit After Tax	Dividend Pay Out Ratio(%)	Receivable days	
Cash Flow Per share	PBIDTM (%)	Inventory Days	

[Refer to Data Dictionary in Appendix](#)

Data structure details:

- Number of records – 4234 rows, 60 columns
- Number of Companies – 402
- Number of Industry – 37
- Number of Macroeconomic – 9
- Years – 10 Years Data (2009 – 2019)

### **Step 2 – Data Pre-processing step:**

- Perform a preliminary EDA
- Check for Outliers
- Missing Data Check
- Check for Correlation between predictor variables & between target variables and predictor variables.

### **Step 3 – Create a model (linear / non-linear)**

Create Linear Regression, Ridge Regression, Lasso Regression, Random Forest, Support Vector Machine (SVM) based models to understand variable of importance, select the best model based on Model Evaluation

### **Step 4 –Building a portfolio of stocks**

Build a portfolio of stocks based on the maximum adjusted EPS values thereby maximizing returns.

## **3. Data Source and Description**

The data used for this study is of top 500 companies in National Stock Exchange of India (NSE). Nifty Index acts as one of the barometers for Indian Stock market health. The data used will be financial results of Nifty500 Companies.

The data used for this analysis is primarily sourced from [AceEquity](#) with help from one of the executives of company who is supporting us for this academic project. AceEquity is one of the top vendors in India for sourcing data intended for purposes of Wealth management, Portfolio Management, Investment Banking, Institutional Stock broking, Management consulting and Business Advisory.

As stated earlier that the data sourced comprises of financial results of Nifty500 companies.

Let us delve a little deeper into the data:

### **Nifty500 Company Financial Results data:**

#### **Key Variable definitions:**

1. Free Cash Flow (FCF): this is cash that a company generates after accounting for outflows to support its operations. FCF is generally a measure of profitability and often a supplemental tool for analysis

Free Cash Flow per Share (FCF\_Per\_Share) is the FCF divided by the total number of shares

outstanding

2. Cash Flow Per Share(CF\_Per\_share): is the after tax earnings plus depreciation on a per share basis which serves as a measure of firm's financial strength
3. Book Value: generally, refers to a company's total assets minus its outstanding liabilities. It represents the total amount of equity worth to its shareholders after liquidating all its tangible assets and paying off all its liabilities. Its useful in determining the value of a company and is often expressed as book value per share
4. Adjusted Book Value – is a measure of the company's valuation after liabilities including off-balance sheet liabilities and assets adjusted to reflect its fair market value.  
Generally, both Book Value and Adjusted book value are very highly correlated.
5. Enterprise Value (EV): is a measure of a company's total value it's often used as an alternative to market capitalization. EV includes market capitalization of a company but also short term and long-term debt as well as any cash on balance sheet. It's generally used for valuing a company for potential takeover.
6. EV/CE(%) (VR\_CE): is a financial ratio between Enterprise Values and Capital employed expressed as a percentage.
7. Netsales: is the Gross Sales minus returns, allowances & discounts. Changes in Netsales do affect the gross profit of a company
8. Profit After Taxes (PAT): is the earning of a company after deducting all its taxes. PAT is watched closely by investors to see profit generating ability of the company
9. Return on Capital Employed (ROCE): this is a financial ratio that measures a company's profitability & efficiency with which the capital is used. A higher ROCE means the company has efficiently utilized its capital. Investors generally prefer companies with stable and gradually rising ROCE

$$\text{ROCE} = \frac{\text{EBIT}}{\text{Capital Employed}}$$

Where: EBIT is Earnings before Interest & Taxes

Capital Employed = Total Assets – Current Liabilities

10. Return on Equity (ROE): is a measure of financial performance & expressed as a percentage. It can be calculated by dividing net income by Shareholders equity. It is an important measure to see how effectively the company management is using company's assets to create profits.

$$\text{ROE} = \frac{\text{Net Income}}{\text{Average Shareholders' Equity}}$$

11. Return on Assets (ROA): is an indicator of how profitable a company is relative to its total assets. This gives an idea to investors how efficient is a company in using its assets to generate earnings.

It's a useful ratio to compare similar companies or comparing a company to its earlier performance. Higher ROA is an indication of assets efficiency.

$$\text{ROE} = \frac{\text{Net Income}}{\text{Total Assets}}$$

12. Earnings Per Share (EPS): is an important measure which reflect how much money does a company make for each share of its stock. Its calculated by company's net profit divided by outstanding shares of its common stockholder.

$$\text{EPS} = \frac{\text{Net Profit}}{\text{End of Period Common Shares Outstanding}}$$

13. Debt Equity: is a ratio of company total liabilities divided by its shareholder equity. A higher debt to Equity ratio for a company generally indicates a risk to shareholders. However, comparing Debt Equity ratios across industries is not ideal since they may vary by industry.

$$\text{Debt Equity} = \frac{\text{Total Liabilities}}{\text{Total Shareholders' Equity}}$$

14. Dividend Yield: is always expressed as a percentage. It's a financial ratio – Dividend by Price, essentially how much a company pays out in dividends each year relative to it's stock price. Mature companies are most likely to pay dividends.

$$\text{Dividend Yield} = \frac{\text{Annual Dividends per Share}}{\text{Price per Share}}$$

15. Price to Earnings Ratio (PE): as the name suggests it is a ratio company share price to its earning per share. A **high PE** ratio could mean overvaluation or high growth expected by investors. A company with no PE ratio would mean that they have no earnings.

$$\text{PE Ratio} = \frac{\text{Market Value Per Share}}{\text{Earnings per Share}}$$

16. Adjusted PE: also know as CAPE ratio is a valuation measure that uses real earning per share (EPS) over a 10-year period to smooth out the fluctuations in corporate profits that occur over different periods of business cycle. This ratio is generally to broad equity indices to whether market is undervalued or overvalued.

$$\text{Adjusted PE} = \frac{\text{Price}}{\text{Average earnings for 10 years}} \\ \text{adjusted for inflation}$$

17. Price to Book Ratio (PB): it is a ratio company's market capitalization to its book value. Calculated by dividing company stocks value per share by it book value per share. Typically, the market value of an equity is higher than the book value. PB ratios under 1 are considered as solid investments.

$$\text{PB Ratio} = \frac{\text{Market Value Per Share}}{\text{Book Value Per Share}}$$

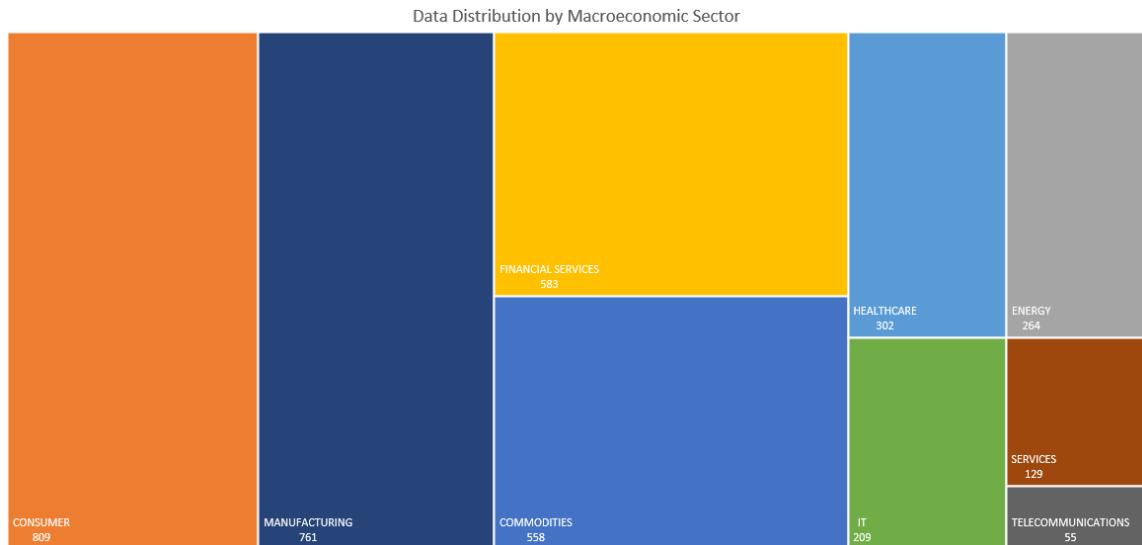
18. Profit before Interest Depreciation and Tax Margin (PBIDTM(%)) : is a financial ratio which is calculated as Adjusted Gross profit + Interest by sales ratio, and then expressed as %

$$\text{PBIDTM}(%) = \frac{\text{Adjusted Gross Profit} + \text{Interest}}{\text{Sales}} * 100$$

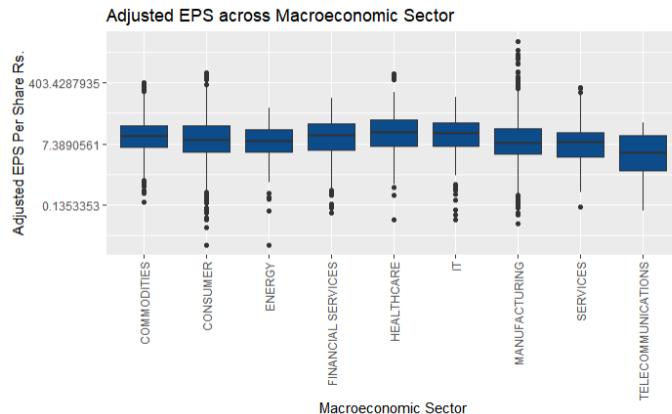
19. Earning Before Interests Taxes and Management (EBITM (%)) : is financial ratio which is calculated by revenue before interest and taxes and management expenses
20. Profit After Taxes Margin (PATM(%)): this is a profit margin ratio which shows the percentage of net sales that remains after deducting cost of goods sold and all other expenses including taxes.
21. Dividend Per Share (DPS): is sum of declared dividends issued to every share outstanding. This is an important metric to investors as it translates to direct income for shareholder.

#### 4. Approach to Problem Statement

As the problem statement is to find the factors that influence Earning per Share (EPS) for different Macro Economic sector. The decision on Macro Economic sector is because the number of observations is more compared to by industry study. Data is distributed more around the following 4 sectors – Consumer, Commodities, Financial Services & Manufacturing.



Hence, the exploratory data analysis and modelling techniques are being focused on the 4 Macro Economic sectors. Since the target variable of this analysis is Adjusted Earnings Per Share below is a snapshot by Macro Economic Sector



Our approach is following below steps for each of the 4 Macro Economic Sector and then compare the results of each sector:

- 1- Data Preprocessing
- 2- Exploratory Data Analysis
- 3- Modelling Approach

## 5. Consumer Macro Economic Sector

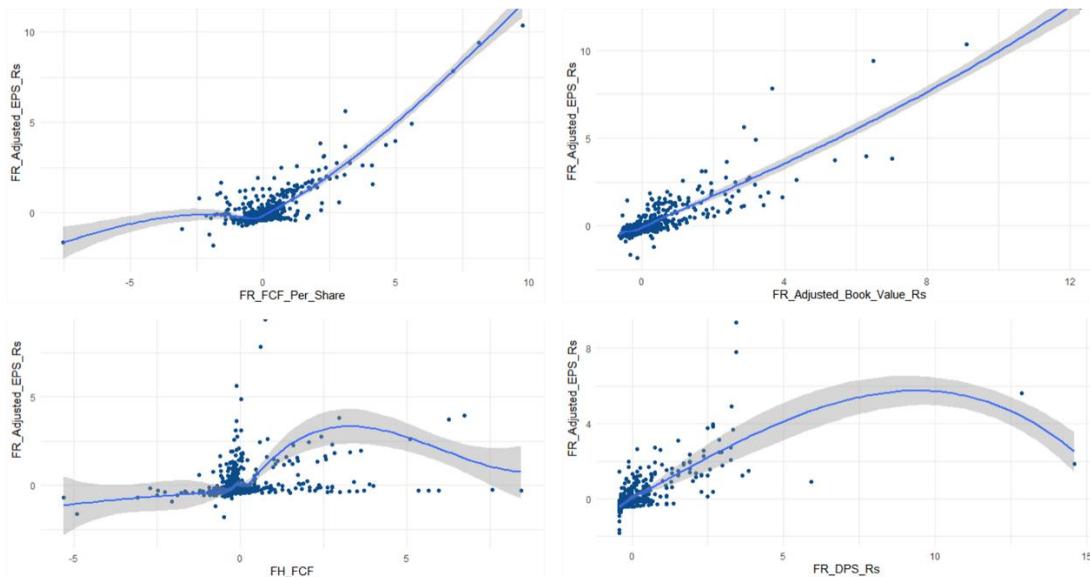
### Data Preprocessing:

- There are 809 observations with 59 variables (including Target variable)
- There are 75 companies in the data set spread across 8 industry for Consumer Macro Economic Sector.
- Data has financial statement information for past 10 Years Data (2009 – 2019)
- Missing Values – No missing values are present
- There is a presence of outliers on the **higher side**, however a decision not to treat the outliers as they could be related to the target variable (as it is a continuous variable)

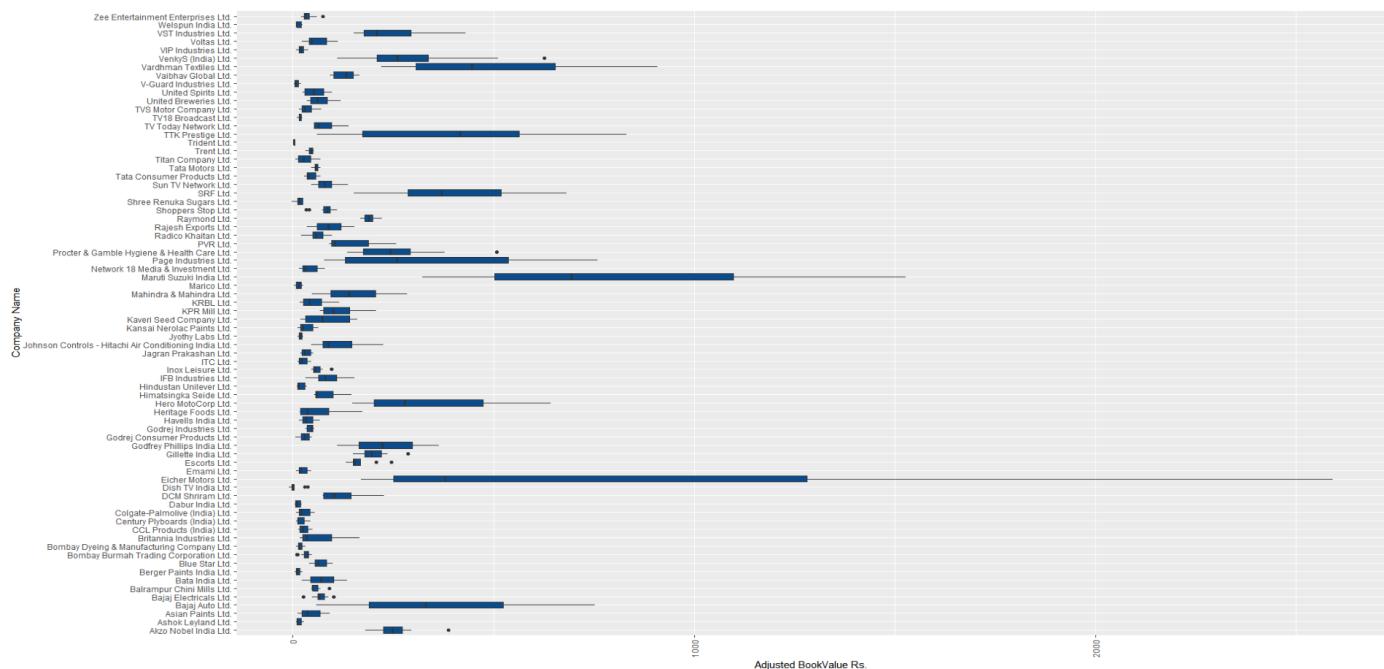
### Exploratory Data Analysis:

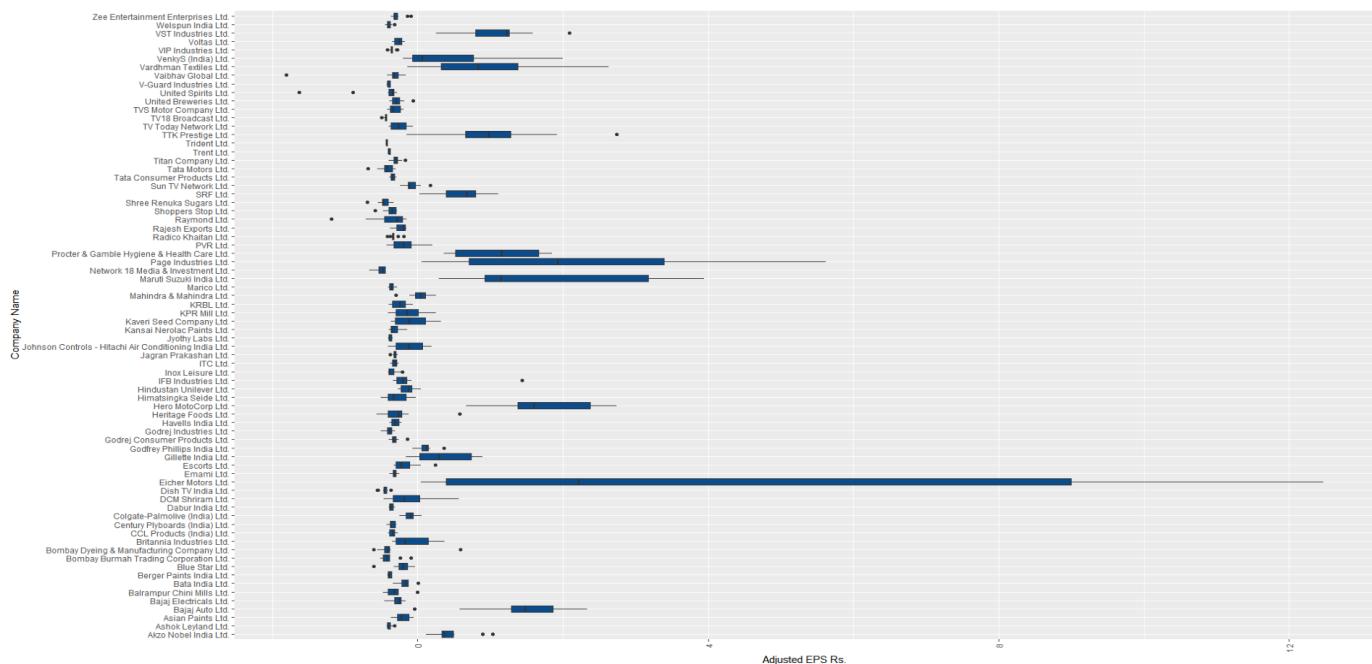
**Univariate Analysis:** on the scaled data would show that majority variables are not normally distributed which is due two reasons 1. Data has outliers on the higher side 2. Number of observations (note – the graph for the univariate analysis is included in Appendix)

### Bivariate Analysis:



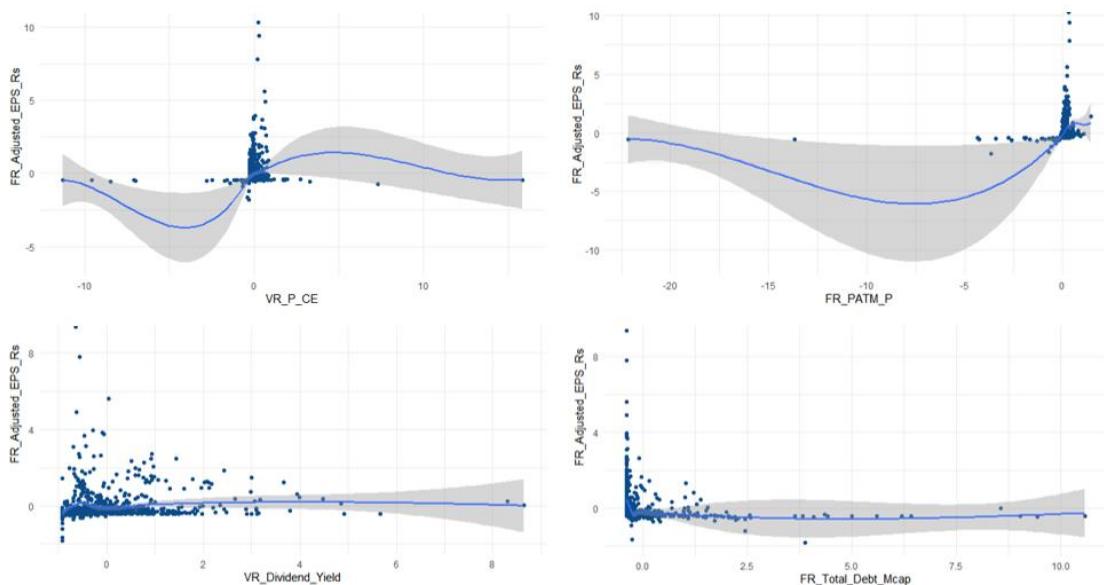
- Companies having a higher book value seems to have a higher Adjusted Earnings per share





- Following are select companies with higher EPS & high book value are

Eicher Motors Ltd.  
Maruti Suzuki Ltd.  
Page Industries Ltd.  
Vardhman Textiles Ltd.



- The above variables Dividend Yield, Profit After Tax Margin, Total Debt Market Capital have a weak correlation with Earnings Per Share

Correlation :

- Running a quick correlation reveals some variables are correlated with the target variable which is observed in the above plots as well
- Below tables reveals some variables which have high correlation with target variables
- Correlation also reveals some variables which are highly correlated amongst themselves. We will however not remove them now but remove them based on the VIF check on the models which should show the multicollinearity

Correlation with Target Variable	
	FR_Adjusted_EPS_Rs
FR_FCF_Per_Share	0.81
FR_Adj_DPS_Rs	0.67
FR_Book_Value_Rs	0.81
FR_Adjusted_Book_Value_Rs	0.89
FH_FCF	0.26
VR_Adjusted_PE	-0.02
VR_P_CE	0.04
VR_Dividend_Yield	0.06
FR_Total_Debt_Mcap	-0.13
PL_NetSales	0.21
PL_PAT	0.26
FR_ROA_P	0.37

Variables with High Correlation	
	FR_EBITM_P
FR_Pre_Tax_Margin_P	0.95
FR_PATM_P	0.94
FR_CPM_P	0.94
FR_Adj_EPS_Growth_P	0.9972

### Modelling Approach:

The approach will be to split data into train & test (70:30 Split), create multiple models using Linear regression, Ridge, Lasso, Random Forest & SVM. Model Evaluation based on RMSE and compare the top 10 variables of importance

Control Parameters: 10-fold Cross Validation with 3 repetitions

### Linear Regression (LR):

LR Model 1: This is a full model.

- The full provides the below list of variables as significant
- Adjusted R<sup>2</sup>: 0.9881

- The full model has a very high adjusted R<sup>2</sup> which is due to presence of Multicollinearity

Estimate	Value
FR_DPS_Rs	0.0252
FR_Asset_Turnover	0.021233
FH_FCF	0.006175
PL_PAT	0.0002718
VR_EBITDA	0.001657
FR_CEPS_Rs	< 2e-16
FR_Interest_Cover	0.000393
FR_CPM_P	0.0000209
FR_Fixed_Asset_Turnover	2.89E-05
FR_CF_Per_share	8.41E-07
PL_NetSales	8.13E-08
FR_FCF_Per_Share	5.87E-10
FR_Adjusted_Book_Value_Rs	2.27E-14
FR_Net_Sales_Growth_P	0.097341
FR_Core_EBITDA_Growth_P	0.091532
FR_PATM_P	0.069057
FR_Payable_days	0.061019
VR_Dividend_Yield	0.053314
FR_Quick_Ratio	0.966864
FR_FCF_Yield	0.900853
FR_ROE_P	0.889813
FR_Current_Ratio	0.881154

Correlated Ratio Pairs	
FR1	FR2
Cash Earnings Per share (FR_CEPS_RS)	Earnings Per Share (FR_EPS_Rs)
	Adjusted EPS Growth (FR_Adj_EPS_Growth_P)
	Earning Before Interest Tax & Management (FR_EBITM_P)
	Earnings Before Interest Tax Depreciation & Amortization (VR_EBITDA)
Asset Turnover	Fixed Asset Turnover
Profit After Tax Margin (FR_PATM_P)	Profit After Tax (PL_PAT)
	FR_CPM_P
Free Cash Flow Per Share (FR_FCF_Per_Share)	Cash Flow Per Share (FR_CF_Per_share)
	Free Cash Flow (FH_FCF)
Market Capitalization by Sales (VR_Mcap_Sales)	VR_NetSales
	PL_NetSales
Dividend Yield	Adjusted Dividend Per Share (FR_Adj_DPS_Rs)
FR_Quick_Ratio	FR_Current_Ratio
FR_Adjusted_Book_Value_Rs	FR_Book_Value_Rs

LR Model 2: This model is created with variables dropped from the above table (FR2)

- Running VIF on the Model 2 would show quite few variables (highlighted in table) with VIF of > 15
- Next step is to run another model after eliminating the variables VIF >15

	VIF		VIF
FR_Pre_Tax_Margin_P	648.1672	FR_Quick_Ratio	1.723073
FR_PATM_P	628.3273	FR_PAT_Growth_P	1.650386
FR_Adj_DPS_Rs	63.07593	FR_Inventory_Turnover	1.626055
FR_DPS_Rs	61.38151	VR_Dividend_Yield	1.522461
FR_CEPS_Rs	25.66528	VR_P_CE	1.52186
FR_Adjusted_Book_Value_Rs	15.52035	FR_Price_CF_Ratio	1.514389
FR_ROA_P	7.104024	VR_High_PE	1.473297
FR_ROCE_P	6.85098	FR_Debtors_Turnover	1.393377
VR_CE	5.372378	FR_Inventory_Days	1.343456
VR_Mcap_Sales	4.423708	FR_Payable_days	1.29124
VR_P_BV	4.149725	FR_Dividend_Pay_Out_Ratio_P	1.280193
FR_FCF_Per_Share	3.598053	FR_Sales_to_CF_ratios	1.22339
FR_PBIDTM_P	3.175334	VR_Adjusted_PE_	1.2137
FR_Asset_Turnover	2.78073	FR_Core_EBITDA_Growth_P	1.20295
FR_Total_Debt_Equity	2.602097	FR_Net_Sales_Growth_P	1.18176
FR_ROE_P	2.360867	VR_EBIT	1.168008
FR_Fixed_Capital_Sales	2.113681	FR_FCF_Yield	1.161535
FR_Receivable_days	2.065681	FR_Interest_Cover	1.129363
FR_EBIT_Growth_P	1.896178	FR_Price_to_FCF	1.042413
FR_Total_Debt_Mcap	1.748306	FR_Tax_Rate_P	1.036926
VR_Low_PE	1.739468	FR_Sales_Working_Capital	1.010421

LR Model 3 – Creating new LR model after dropping the above variables with VIF >15

Model Output summary:

- VIF check shows all variables have VIF of <10
- In this model there are 12 variables that are significant
- **Adjusted R<sup>2</sup>: 0.9237, RMSE: 0.3538176**
- Next step is to run model on only significant variables

	Pr(> t )		Pr(> t )
VR_CE	1.53E-09	FR_Fixed_Capital_Sales	0.472473
FR_Total_Debt_Equity	1.26E-06	FR_ROA_P	0.483734
FR_Debtors_Turnover	0.000216	FR_PAT_Growth_P	0.494647
FR_FCF_Per_Share	< 2e-16	FR_Inventory_Turnover	0.528767
FR_Adjusted_Book_Value_Rs	< 2e-16	FR_Sales_Working_Capital	0.589768
FR_ROCE_P	0.00182	VR_P_CE	0.621182
VR_P_BV	0.003707	VR_Adjusted_PE_	0.623961

FR_Net_Sales_Growth_P	0.013148	FR_Interest_Cover	0.664008
FR_Receivable_days	0.022712	FR_Price_to_FCF	0.72956
FR_FCF_Yield	0.025396	(Intercept)	0.735692
VR_Dividend_Yield	0.039395	FR_Price_CF_Ratio	0.770352
FR_ROE_P	0.048254	FR_Inventory_Days	0.797265
FR_Quick_Ratio	0.076842	VR_High_PE	0.820683
FR_Sales_to_CF_ratios	0.082343	VR_EBIT	0.870621
VR_Low_PE	0.109485	FR_Dividend_Pay_Out_Ratio_P	0.950711
FR_EBIT_Growth_P	0.135421	VR_Mcap_Sales	0.996232
FR_Core_EBITDA_Growth_P	0.228623		
FR_PBIDTM_P	0.286499		
FR_Asset_Turnover	0.301851		
FR_Tax_Rate_P	0.34465		
FR_Payable_days	0.365835		
FR_Total_Debt_Mcap	0.373798		

#### LR Model 4 – New LR model with significant variables

- Model 4 has the best model where all variables are significant and VIF is <10 for all variables with no multicollinearity
- **Adjusted R<sup>2</sup>: 0.9244, RMSE: 0.3533869**
- The list of variables along based in order of their significance are as follows:  
 Adjusted Book value  
 Free Cash Flow Per share  
 Net Sales Growth %  
 Return on Capital Employed  
 Price by Book Value  
 Dividend Yield  
 Return on Equity  
 Total Debt Equity  
 Valuation Ratio of Enterprise Value by Capital Employed expressed as a percentage(VR\_CE)

Based on the significant variables from Final Linear Regression Model (LR model 4), we run the Ridge, Lasso, Random Forest & Support Vector Machine model.

#### Comparison of Various Model on RMSE for Consumer MacroEconomic Sector:

	Linear Regression	Ridge	Lasso	Random Forest	SVM
RMSE on Test Data	0.3533869	0.3475785	0.3528797	0.2166468	0.3681683

- Based on above table Random Forest gives the best model in terms of RMSE followed by Ridge
- A quick look at variables of importance for these all model; VarImp function from Caret Package which is a generic function to find the variables of importance based on absolute value of T statistic

Linear Regression Model	Variable of Importance		
	Ridge Model	Lasso Model	Random Forest
Adjusted Book Value	Adjusted Book Value	Adjusted Book Value	Adjusted Book Value
Free Cash Flow Per_Share			
Enterprise Value by Capital Employed (%)	Enterprise Value by Capital Employed (%)	Enterprise Value by Capital Employed (%)	Return on Capital Employed (ROCE)
Return on Capital Employed (ROCE)	Return on Capital Employed (ROCE)	Return on Capital Employed (ROCE)	Return on Equity (ROE)
Total Debt Equity	Total Debt Equity	Total Debt Equity	Enterprise Value by Capital Employed (%)
Price by Book Value	Debtors Turn Over	Price by Book Value	Debtors Turn Over
Return on Equity (ROE)	Price by Book Value	Netsales Growth (%)	Free Cash Flow Yield
Debtors Turn Over	Netsales Growth (%)	Debtors Turn Over	Price by Book Value
Dividend Yield	Dividend Yield	Return on Equity (ROE)	Netsales Growth (%)
Netsales Growth (%)	Free Cash Flow Yield	Dividend Yield	Total Debt Equity
Free Cash Flow Yield	Return on Equity (ROE)	Free Cash Flow Yield	Dividend Yield

- Financial Ratio's like Adjusted Book Value, Return on Equity, Return on Capital Employed, Free Cash Flow Per share, Total Debt Equity & Valuation Ratio like Enterprise Value by Capital Employed (%) are key variables for Consumer Macro Economic Sector

## 6. Commodities Macro Economic Sector

### Data Preprocessing:

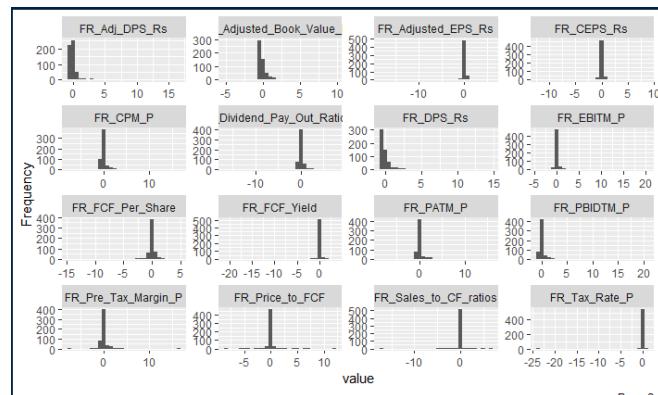
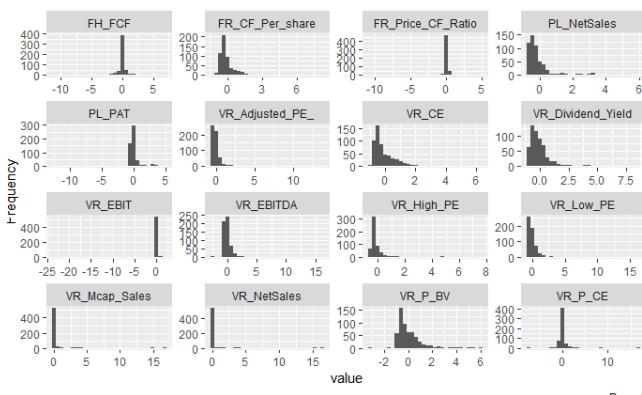
- Below is a snapshot of the data. We have 558 observations for the Commodities macro sector and 68 gross variables
- There are 8 different industries and 51 companies' data considered for the commodities sector modelling for last 10 years (2009 to 2019)
- There are no missing data
- All the variables are numerical except for momentum which is a derived variable and will not be considered in the analysis
- There are extreme values in the data for majority of the variables, but since they are not outliers, they are to be considered in the model building process, bring treated as **Extreme Values**

### Exploratory Data Analysis:

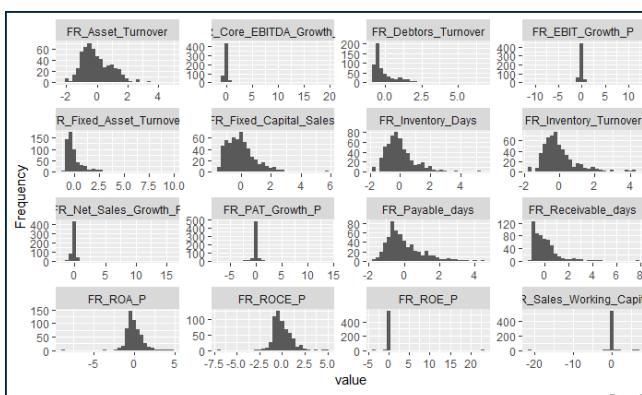
#### Univariate Analysis

- This is done to understand the distribution of the variables when considered independently. The variables are expected to be normally distributed to have an accurate modelling in the later stages
- Histogram plots are drawn for univariate analysis for every variable in the model as shown below

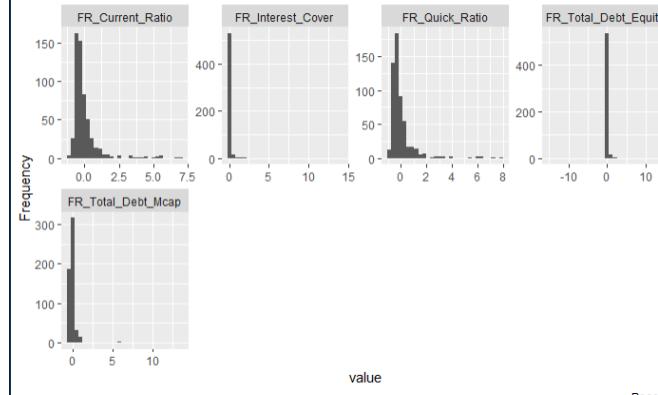
Snapshots:



Page 2



Page 3



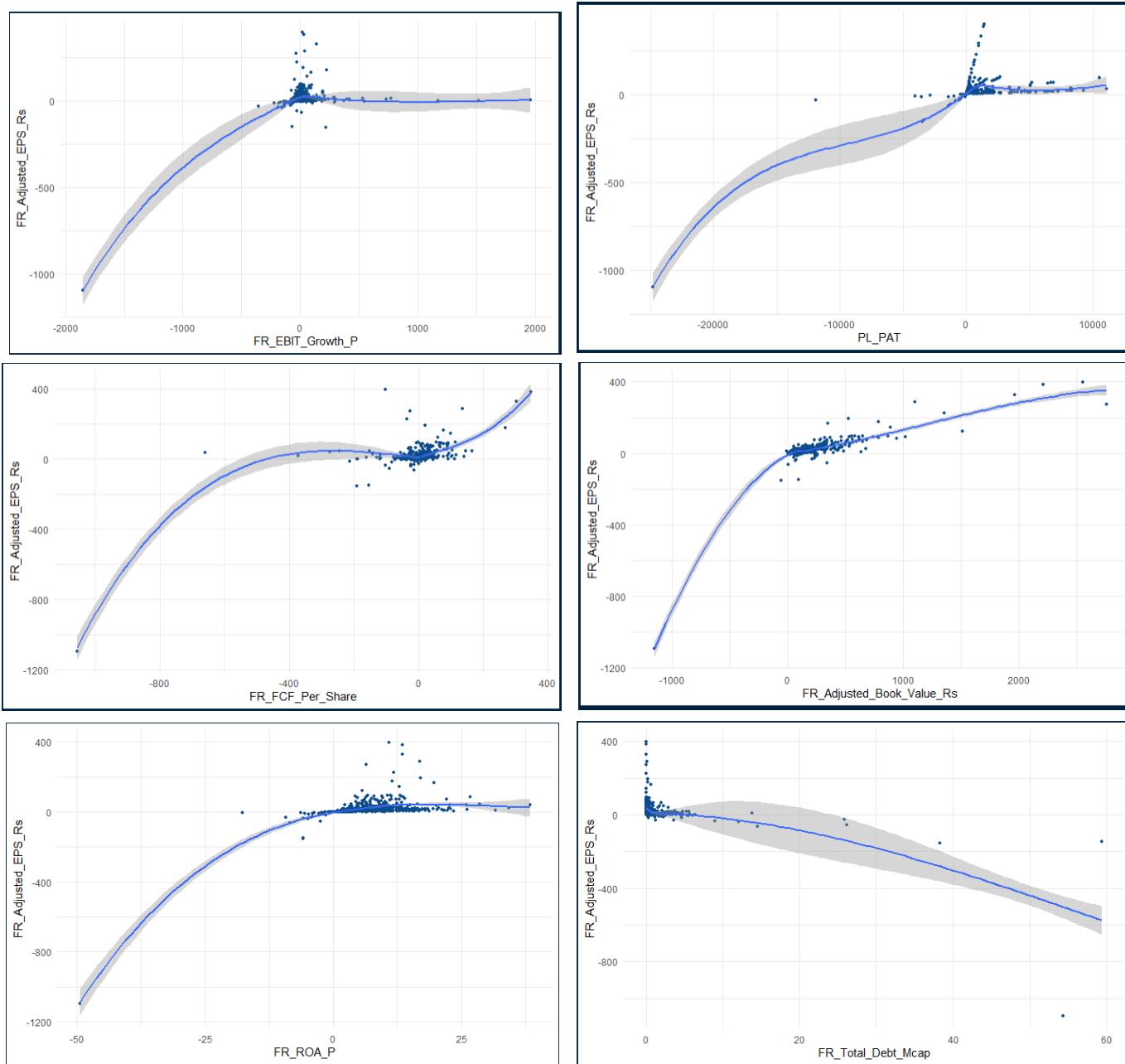
Page 4

There are violations of normality from the histogram plots above for a lot of variables under consideration, but we are going to include these in the model building without any treatment as they are considered to be extreme values which is affecting the normality of the distribution of variables

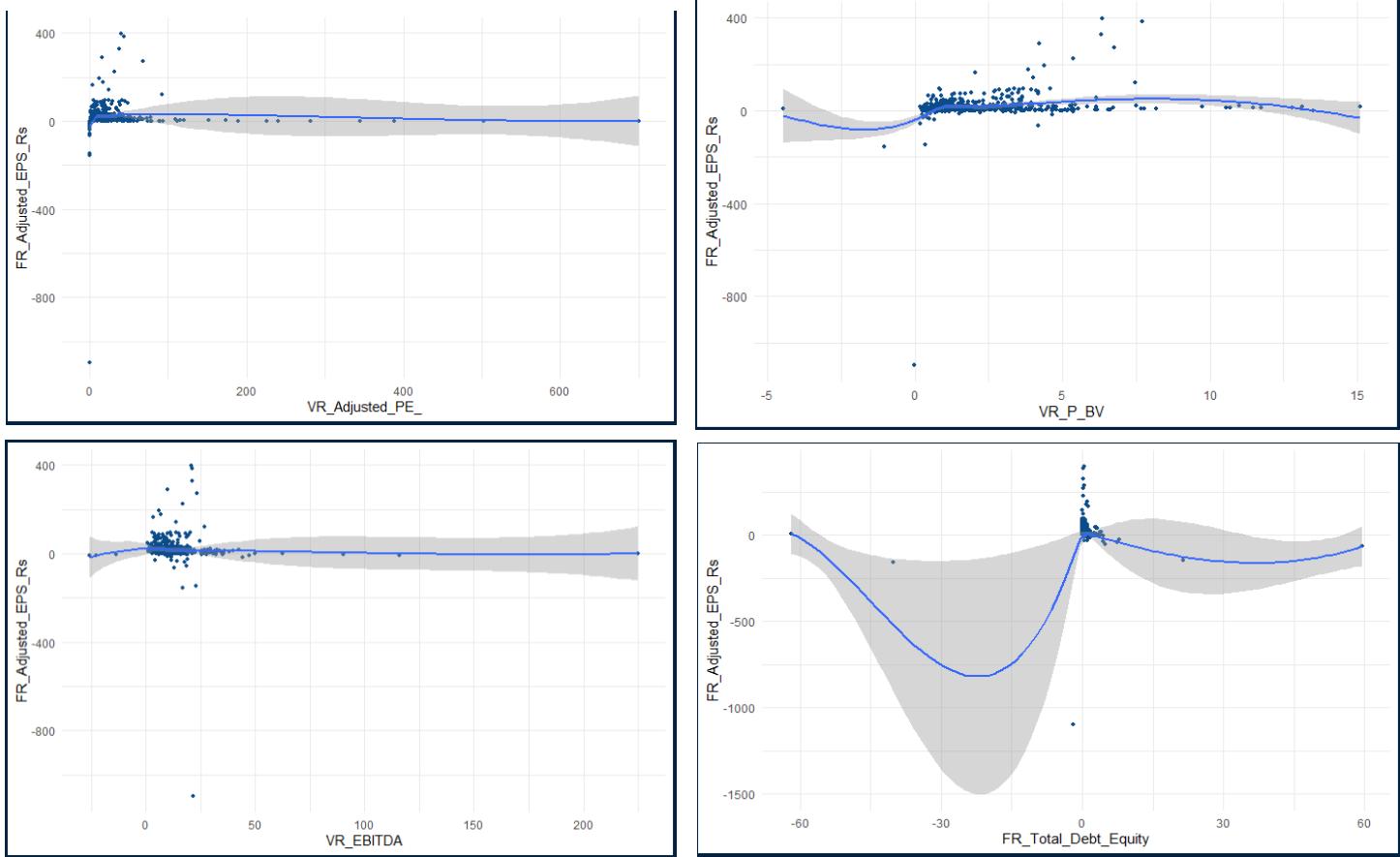
### Bivariate Analysis

- The dependent variable here is “FR\_Adjusted\_EPS\_Rs”, which is adjusted earnings per share of the company for the year.
- Very few variables are showing a correlated linear relationship with the dependent variable as we can see in few of the scatter plots displayed below
- The slope of the line shows the strength of the relation with the dependent variable

Strong relationships:



### Weak or no Relationships



### Correlations

- There are highly correlated variables in the data, which will be omitted based on the VIF values of the model later
- Below are the variables that have more than 0.25 correlation with the dependent variable (highlighted variables have very high correlation (above .6) with the dependent variable **“FR\_Adjusted\_EPS\_Rs”**)
- The correlation output from the below table is somewhat a reflection of relations we got from the bivariate EDA above
- Also we remove the variables “FR\_Book Value (Rs)” as we have one high correlated adjusted book value variable and “PL\_Earnings Per Share”, “PL\_Adjusted EPS”, “FR\_Earnings Per Share (Rs)”, on the basis of dependence with the dependent variable

Predictor Variables	FR_Adjusted_EPS_Rs	Predictor Variables	FR_Adjusted_EPS_Rs
FH_FCF	0.44	FR_Pre_Tax_Margin_P	0.29
PL_PAT	0.51	FR_PATM_P	0.32
FR_CF_Per_share	0.49	FR_CPM_P	0.33
FR_FCF_Per_Share	0.64	FR_ROA_P	0.48
FR_FCF_Yield	0.71	FR_ROCE_P	0.38
FR_EPS_Rs	0.94	FR_Fixed_Capital_Sales	-0.25

FR_CEPS_Rs	0.93	FR_EBIT_Growth_P	0.38
FR_DPS_Rs	0.44	FR_Total_Debt_Mcap	-0.58
FR_Adj_DPS_Rs	0.50		
FR_Book_Value_Rs	0.62		
FR_Adjusted_Book_Value_Rs	0.71		

### Modelling Approach

- In this part of the project, we are trying to find the top factors that affects the earnings of the investors. This will be the exploratory part of the project and based on the exploratory outputs, the stakeholders of the company operating in a particular macro-economic sector would be able to identify the top factors to be focused to derive improved earnings for the investors
- We run and compare different model outputs and performance parameters to arrive at the best method

### Linear Regression

The initial LR model is run with all the variables except the dropped variables as mentioned in the data preparation stage and below.

- There are some correlated variables in the data as shown below in the table

Correlated variables	
FR1	FR2
FR_CEPS_RS	FR_EPS_Rs
	FR_Adj_EPS_Growth_P
	FR_EBITM_P
	VR_EBITDA
FR_Asset_Turnover	FR_Fixed_Asset_Turnover
FR_PATM_P	PL_PAT
	FR_CPM_P
	FR_CF_Per_share
FR_FCF_Per_Share	FH_FCF
	VR_NetSales
VR_Mcap_Sales	PL_NetSales
FR_Quick_Ratio	FR_Current_Ratio
FR_Adjusted_Book_Value_Rs	FR_Book_Value_Rs
FR_Adjusted_EPS_Rs	FR_CEPS_Rs

- There is high degree of multicollinearity between the variables (Column FR1 & FR2)
- So, we consider only the variables in Column FR1 for model building

### Model#1

Coefficients:	Estimate	Std.Error	t value	Pr(> t )	Significance
FR_Adjusted_Book_Value_Rs	0.6	0.0	17.3	0.0	***
FR_FCF_Yield	0.2	0.0	6.5	0.0	***
FR_EBIT_Growth_P	0.2	0.0	6.5	0.0	***
FR_Quick_Ratio	-0.1	0.0	-3.4	0.0	***
FR_PBIDTM_P	-0.3	0.1	-3.2	0.0	**
FR_ROA_P	0.3	0.1	3.2	0.0	**
FR_Total_Debt_Mcap	-0.1	0.0	-3.0	0.0	**
FR_Core_EBITDA_Growth_P	-0.1	0.0	-3.0	0.0	**
FR_PATM_P	0.5	0.2	2.3	0.0	*
FR_Total_Debt_Equity	0.1	0.0	2.2	0.0	*
VR_P_BV	0.2	0.1	2.1	0.0	*
VR_Low_PE	0.0	0.0	-2.0	0.0	*
VR_CE	-0.2	0.1	-2.0	0.1	.
VR_Mcap_Sales	0.1	0.0	1.9	0.1	.
FR_FCF_Per_Share	0.1	0.0	1.6	0.1	
FR_Dividend_Pay_Out_Ratio_P	0.0	0.0	-1.4	0.2	
FR_PAT_Growth_P	0.0	0.0	-1.4	0.2	
FR_ROCE_P	-0.1	0.1	-1.2	0.2	
FR_Fixed_Capital_Sales	0.0	0.0	-1.2	0.2	
VR_P_CE	-0.1	0.0	-1.2	0.2	
FR_Pre_Tax_Margin_P	-0.2	0.2	-1.1	0.3	
VR_EBIT	-0.2	0.2	-1.1	0.3	
FR_ROE_P	0.0	0.0	1.0	0.3	
FR_Tax_Rate_P	0.2	0.2	1.0	0.3	
FR_Asset_Turnover	0.0	0.0	-1.0	0.3	
FR_Price_to_FCF	0.0	0.0	-0.9	0.4	
FR_Receivable_days	0.0	0.0	0.8	0.4	
FR_Sales_Working_Capital	0.0	0.0	0.7	0.5	
FR_Debtors_Turnover	0.0	0.0	0.7	0.5	
FR_Net_Sales_Growth_P	0.0	0.0	-0.6	0.5	
FR_Sales_to_CF_ratios	0.0	0.0	0.6	0.5	
FR_DPS_Rs	0.0	0.0	-0.6	0.6	
FR_Price_CF_Ratio	0.0	0.0	-0.6	0.6	
FR_Payable_days	0.0	0.0	0.5	0.6	
FR_Interest_Cover	0.0	0.0	-0.5	0.7	
FR_Inventory_Days	0.0	0.0	-0.4	0.7	
VR_High_PE	0.0	0.0	0.4	0.7	
FR_Adj_DPS_Rs	0.0	0.1	-0.3	0.8	
VR_Adjusted_PE	0.0	0.0	-0.2	0.8	
VR_Dividend_Yield	0.0	0.0	-0.2	0.9	
FR_Inventory_Turnover	0.0	0.0	-0.1	0.9	

- The model is statistically significant at 95% confidence level, with an **Adjusted R squared value of 0.90**, which might be due to some high correlated variables as we can see from the VIF outputs below
- The table below shows the output of full model excluding the variables dropped above

The VIF values of the full model are as listed below

Variables	VIF Value	Variables	VIF Value
FR_Pre_Tax_Margin_P	90.77	FR_Inventory_Turnover	2.96
FR_PATM_P	83.34	FR_Adjusted_Book_Value_Rs	2.69
FR_ROA_P	26.82	FR_Total_Debt_Equity	2.44
VR_P_BV	22.70	FR_ROE_P	2.37
VR_CE	21.19	VR_Dividend_Yield	2.12
FR_ROCE_P	20.65	FR_Debtors_Turnover	2.01
FR_PBIDTM_P	8.74	FR_Price_CF_Ratio	1.99
VR_P_CE	7.86	VR_Low_PE	1.91
FR_FCF_Yield	6.88	FR_Receivable_days	1.89
VR_Mcap_Sales	5.90	FR_Core_EBITDA_Growth_P	1.85
FR_FCF_Per_Share	4.81	FR_Quick_Ratio	1.69
FR_Adj_DPS_Rs	4.06	FR_Sales_to_CF_ratios	1.54
FR_Asset_Turnover	3.85	FR_Tax_Rate_P	1.54
VR_High_PE	3.58	FR_PAT_Growth_P	1.49
FR_Total_Debt_Mcap	3.56	FR_Net_Sales_Growth_P	1.47
VR_Adjusted_PE_	3.45	FR_Dividend_Pay_Out_Ratio_P	1.44
VR_EBIT	3.39	FR_Interest_Cover	1.31
FR_EBIT_Growth_P	3.24	FR_Payable_days	1.20
FR_Fixed_Capital_Sales	3.15	FR_Sales_Working_Capital	1.06
FR_DPS_Rs	3.08	FR_Price_to_FCF	1.05
FR_Inventory_Days	3.00		

- Below is the list of variables which are coming out to significant in the model at 95% confidence interval, for the full model
- The **Adjusted R squared value** of the model is **.90** and **AIC** value of **305**

Coefficients:	Estimate	Std.Error	t value	Pr(> t )	Significance
FR_Adjusted_Book_Value_Rs	0.6	0.0	17.3	0.0	***
FR_FCF_Yield	0.2	0.0	6.5	0.0	***
FR_EBIT_Growth_P	0.2	0.0	6.5	0.0	***
FR_Quick_Ratio	-0.1	0.0	-3.4	0.0	***
FR_PBIDTM_P	-0.3	0.1	-3.2	0.0	**
FR_ROA_P	0.3	0.1	3.2	0.0	**
FR_Total_Debt_Mcap	-0.1	0.0	-3.0	0.0	**
FR_Core_EBITDA_Growth_P	-0.1	0.0	-3.0	0.0	**
FR_PATM_P	0.5	0.2	2.3	0.0	*

FR_Total_Debt_Equity	0.1	0.0	2.2	0.0	*
VR_P_BV	0.2	0.1	2.1	0.0	*
VR_Low_PE	0.0	0.0	-2.0	0.0	*

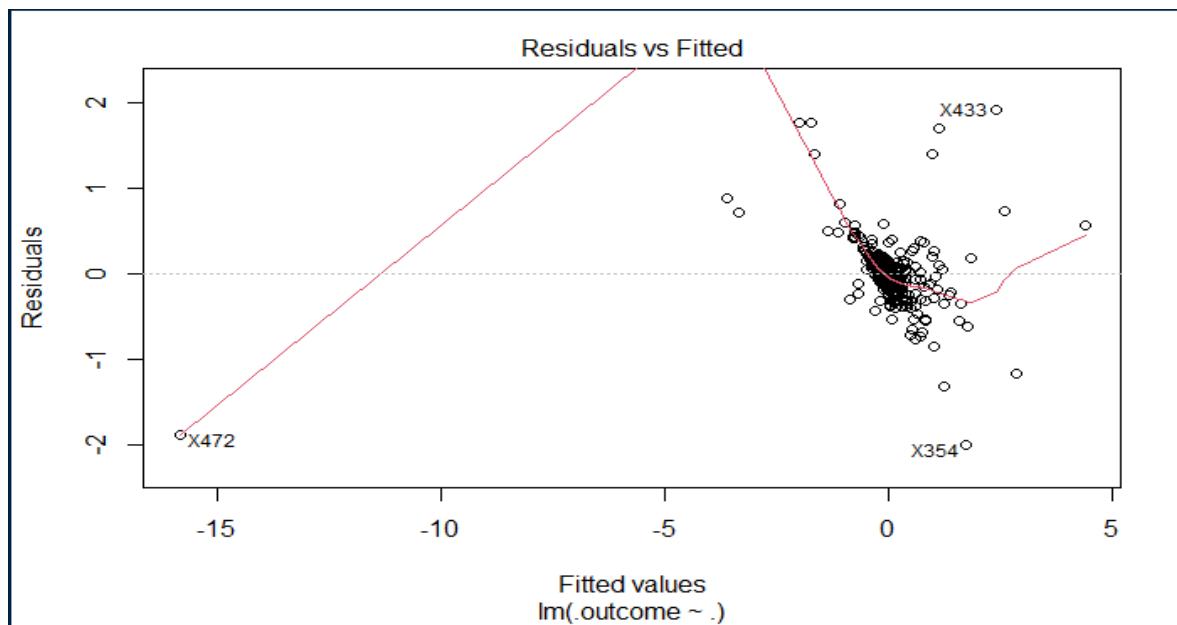
- The variables with above 15 VIF values are removed and a second model is built using the remaining variables to obtain an output as shown below

### Model2 output

- Below are the significant variables at 95% confidence interval.

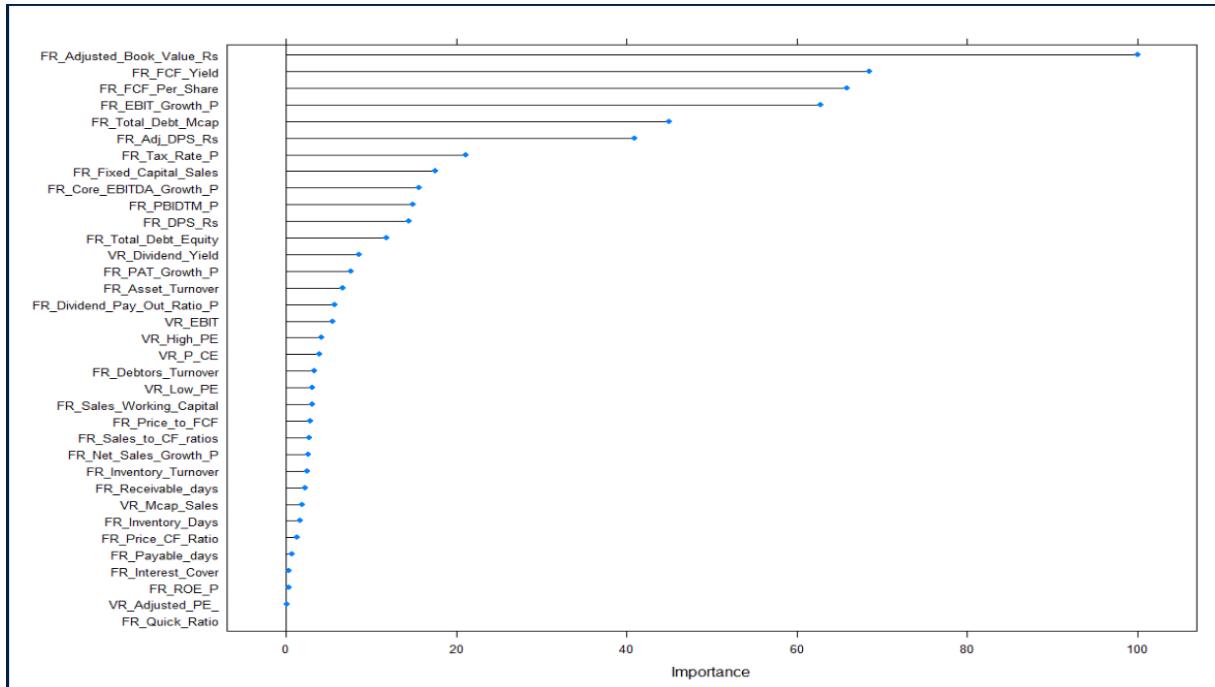
Coefficients:	Estimate	Std.Error	t value	Pr(> t )	Significance
FR_Adjusted_Book_Value_Rs	0.56	0.03	16.03	0.00	***
FR_FCF_Yield	0.33	0.04	8.34	0.00	***
FR_EBIT_Growth_P	0.28	0.03	8.20	0.00	***
FR_Total_Debt_Mcap	-0.14	0.03	-4.95	0.00	***
FR_Core_EBITDA_Growth_P	-0.17	0.04	-4.11	0.00	***
FR_Total_Debt_Equity	0.08	0.03	2.87	0.00	**
FR_PBIDTM_P	0.15	0.06	2.44	0.02	*
FR_Asset_Turnover	0.07	0.03	2.29	0.02	*
FR_Fixed_Capital_Sales	-0.07	0.03	-2.23	0.03	*

- The model is statistically significant with an adjusted R squared value of **0.88**, AIC of **360** and RMSE of **0.59**
- A residual plot of the model after removing the high collinearity variables suggests that the residuals does not show any pattern and are randomly dispersed around the parallel axis to horizontal and the linear regression is appropriate



- Further to this, we run a Ridge and Lasso regression to find the variable importance

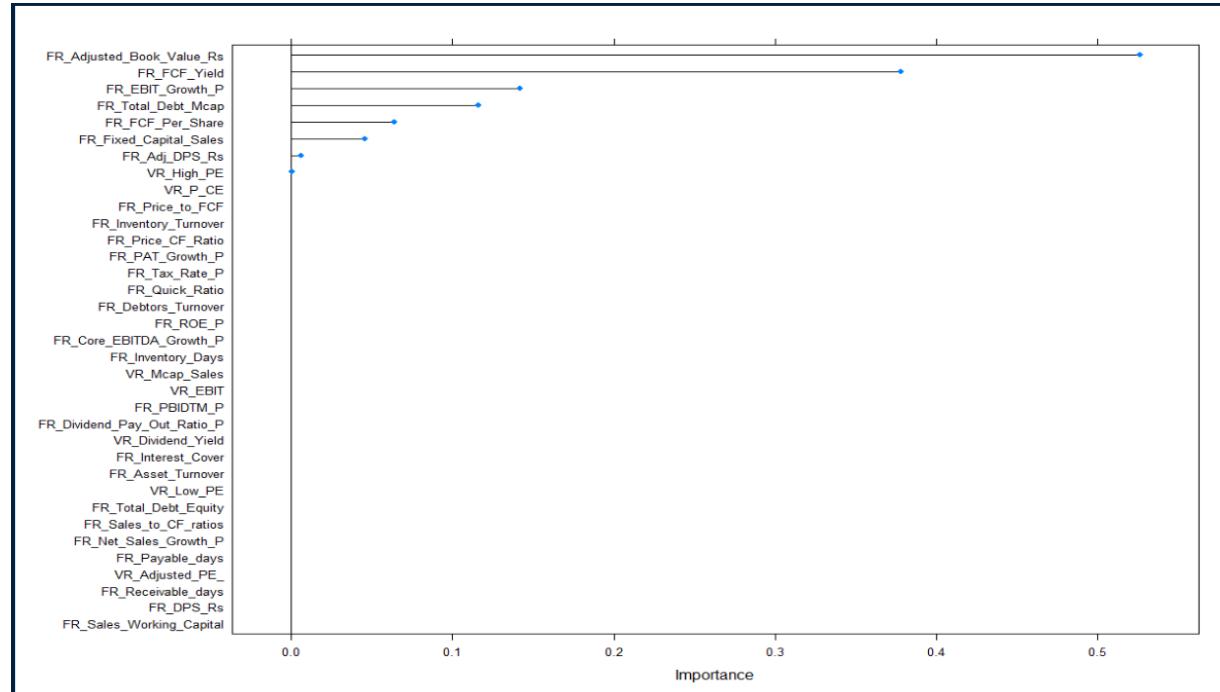
#### Variable importance from Ridge regression



Variables	Variable Importance Value
FR_Adjusted_Book_Value_Rs	100.00
FR_FCF_Yield	68.50
FR_FCF_Per_Share	65.86
FR_EBIT_Growth_P	62.76
FR_Total_Debt_Mcap	44.97
FR_Adj_DPS_Rs	40.92
FR_Tax_Rate_P	21.09
FR_Fixed_Capital_Sales	17.57
FR_Core_EBITDA_Growth_P	15.65
FR_PBIDTM_P	14.95
FR_DPS_Rs	14.41
FR_Total_Debt_Equity	11.80
VR_Dividend_Yield	8.55
FR_PAT_Growth_P	7.62
FR_Asset_Turnover	6.69
FR_Dividend_Pay_Out_Ratio_P	5.69
VR_EBIT	5.48
VR_High_PE	4.12
VR_P_CE	3.96
FR_Debtors_Turnover	3.33

**RMSE on ridge regression = 0.44**

### Lasso Regression outputs



We get a more condensed list of final variables from the variable importance list from Lasso regression as given below which is again comparable to model outputs of Ridge regression

Variables	Variable Importance Value
FR_Adjusted_Book_Value_Rs	100.00
FR_FCF_Yield	71.85
FR_EBIT_Growth_P	26.99
FR_Total_Debt_Mcap	22.12
FR_FCF_Per_Share	12.20
FR_Fixed_Capital_Sales	8.68
FR_Adj_DPS_Rs	1.21
VR_High_PE	0.04

**RMSE on Lasso Regression = 0.35**

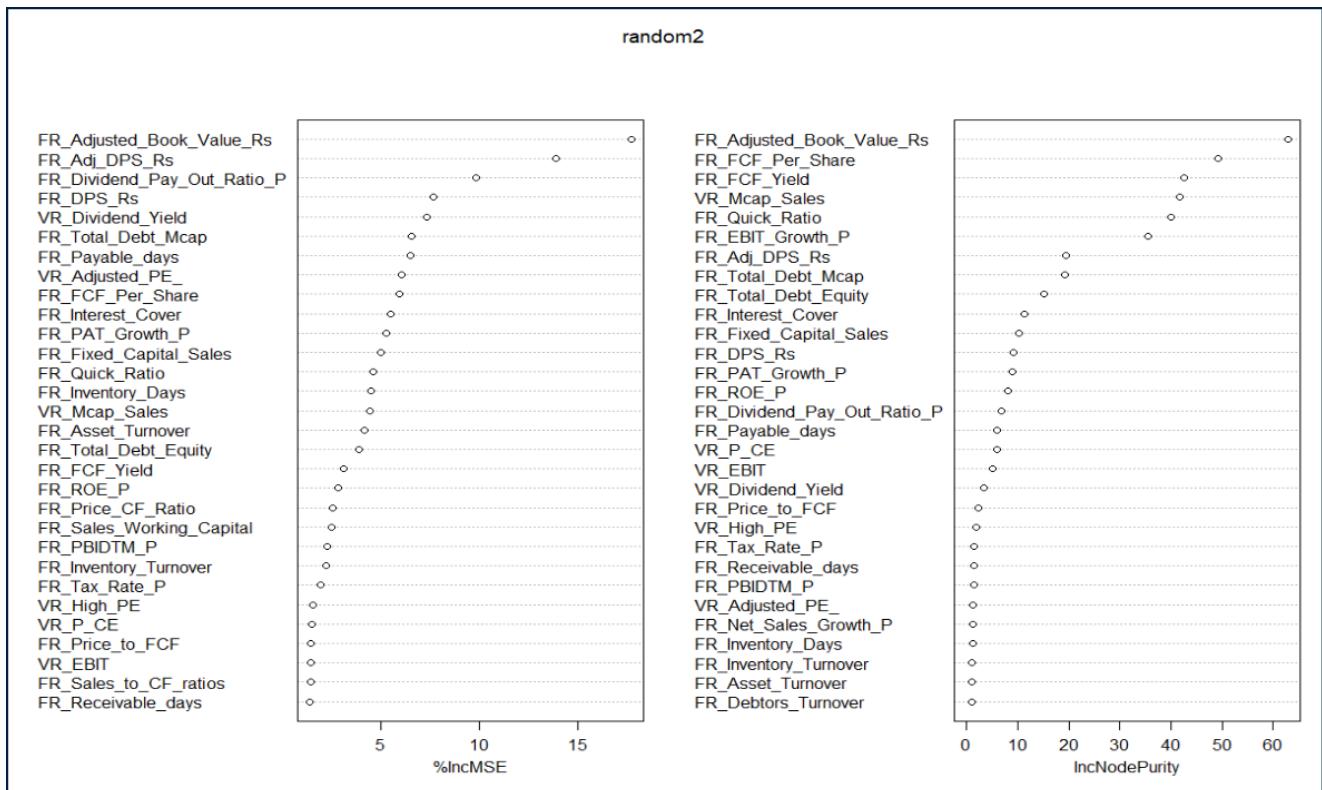
### **Random Forest Algorithm**

In addition to the linear regression, we ran a random forest to see which one gives the better results in terms of the performance matrices

- The Random Forest algorithm produces the below variable importance table

Variable Importance		Variable Importance	
Variable	Value	Variable	Value
FR_Adjusted_Book_Value_Rs	17.74	FR_ROE_P	2.83
FR_Adj_DPS_Rs	13.91	FR_Price_CF_Ratio	2.51
FR_Dividend_Pay_Out_Ratio_P	9.85	FR_Sales_Working_Capital	2.49
FR_DPS_Rs	7.66	FR_PBIDTM_P	2.23
VR_Dividend_Yield	7.31	FR_Inventory_Turnover	2.17
FR_Total_Debt_Mcap	6.52	FR_Tax_Rate_P	1.92
FR_Payable_days	6.51	VR_High_PE	1.51
VR_Adjusted_PE_	6.07	VR_P_CE	1.48
FR_FCF_Per_Share	5.93	FR_Price_to_FCF	1.41
FR_Interest_Cover	5.48	VR_EBIT	1.41
FR_PAT_Growth_P	5.29	FR_Sales_to_CF_ratios	1.39
FR_Fixed_Capital_Sales	4.96	FR_Receivable_days	1.38
FR_Quick_Ratio	4.58	FR_Core_EBITDA_Growth_P	1.29
FR_Inventory_Days	4.47	FR_Debtors_Turnover	1.12
VR_Mcap_Sales	4.42	FR_EBIT_Growth_P	0.47
FR_Asset_Turnover	4.13	VR_Low_PE	0.10
FR_Total_Debt_Equity	3.89	FR_Net_Sales_Growth_P	-0.64
FR_FCF_Yield	3.08		

Variable importance chart from RF method



The random forest method variable importance list is comparable to the list we got from the regression models

**RMSE on Random Forest = 0.47**

### **Support Vector Machine Algorithm**

Support Vector machine algorithm is also used as one of the algorithms to see the model performance  
The RMSE value of the SVM output is **0.46**

### **Comparison of Various Model on RMSE for Consumer Macroeconomic Sector:**

- The below table compared the RMSE outputs of different models run so far
- The Lasso regression comes out to be the best model based on RMSE value

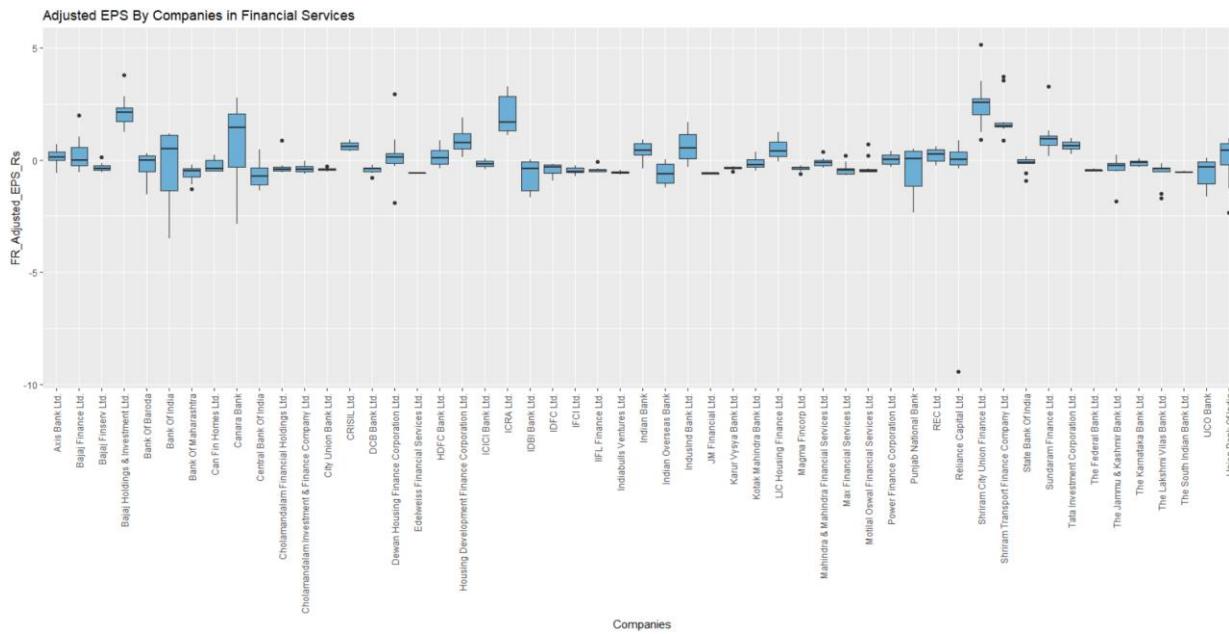
	LR	Ridge	Lasso	RF	SVM
RMSE	0.59	0.44	0.35	0.47	0.46

## **7. Financial Services Macro Economic Sector**

### **Data Pre-Processing**

- There are 583 observations with 67 variables (including Target variable)
- There are 53 companies in the data set spread across 2 industries – Banks and Finance
- Data has financial statement information for past 10 Years Data (2009 – 2019)
- Missing Values – No missing values are present
- There is a presence of outliers on the **higher side**, however a decision not to treat the outliers as they could be related to the target variable (as it is a continuous variable)

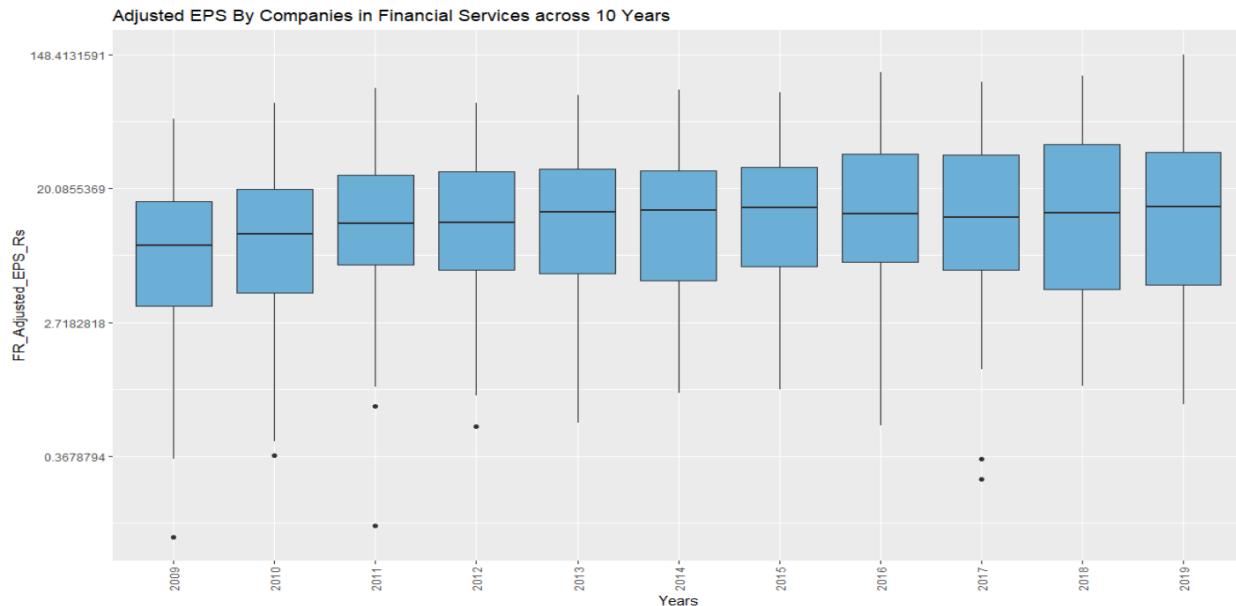
### **Exploratory Data Analysis**



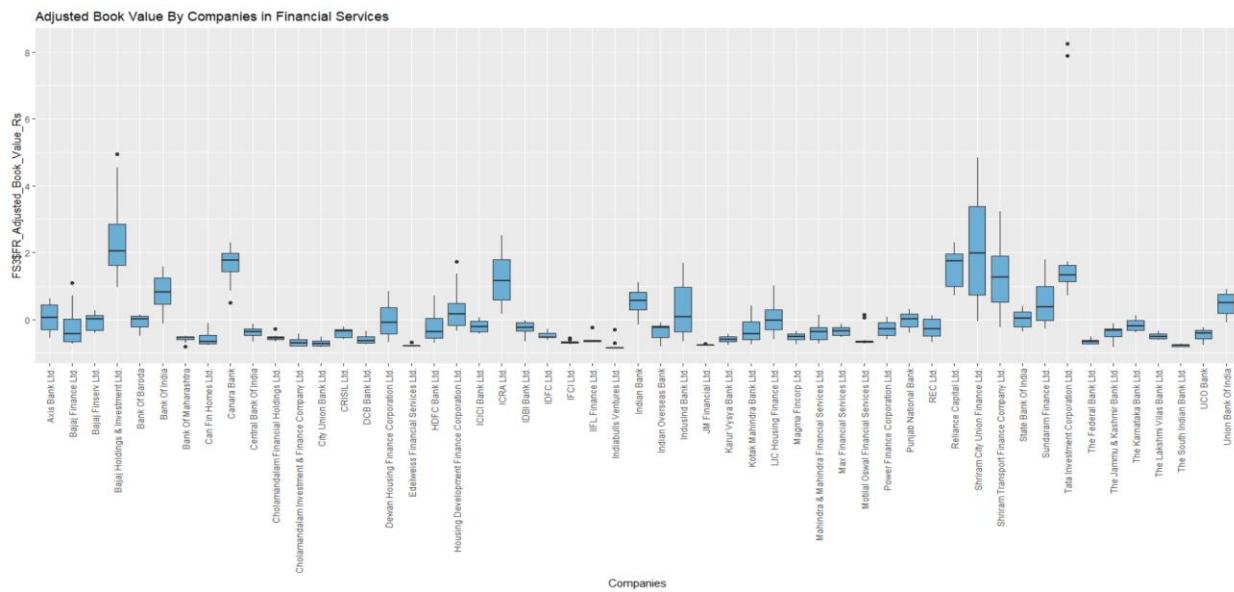
The following Companies belonging to Financial Services Macro Sector have a higher Adjusted EPS compared to the rest: -

- Shriram City Union Finance LTD.
- ICRA
- Canara Bank
- Bajaj Holding and Investment LTD.

Rest of the companies have a median Adjusted EPS close to the average.



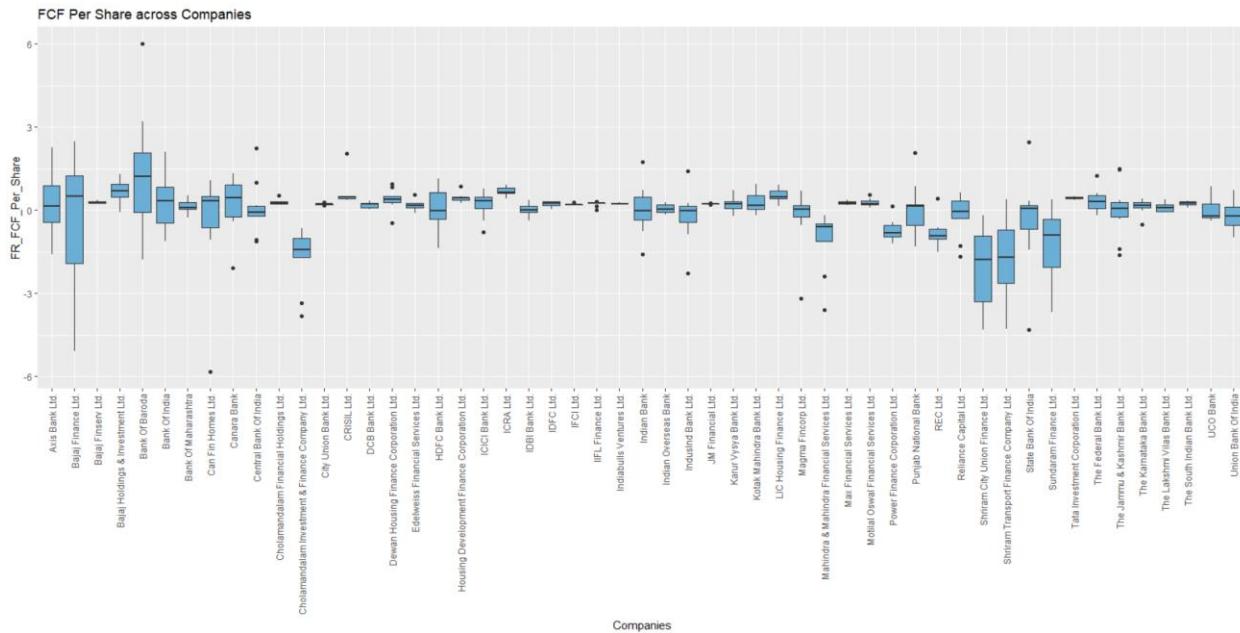
The Adjusted EPS across Companies belonging to the Financial Services Industry increased in the first two years and has stabilized over the last 8 years.



The Adjusted Book Value is higher for the following companies belonging to the Financial Services Macro Sector as compared to the rest: -

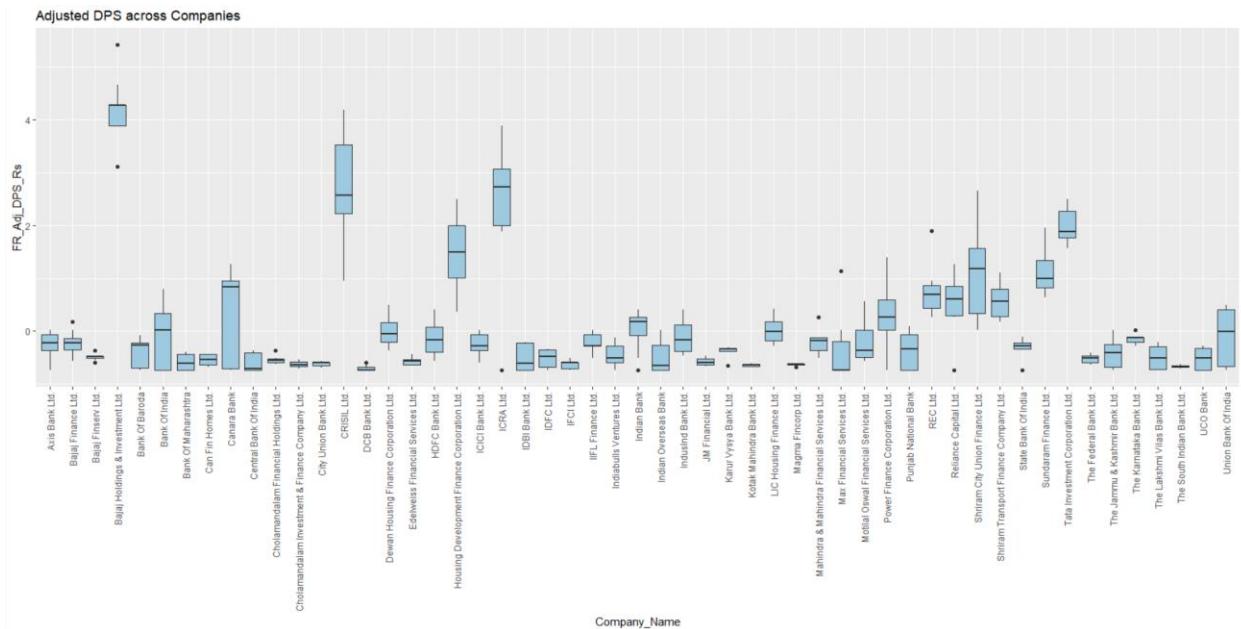
- Shriram City Union Finance LTD.
  - ICRA
  - Canara Bank
  - Bajaj Holding and Investment LTD.
  - Reliance Capital

Adjusted book value looks at the value of a company in terms of the current market values of its assets and liabilities. Specifically, book value concerns the total value of company assets minus the total value of company liabilities.



Free cash flow per share (FCF) is a measure of a company's financial flexibility that is determined by dividing free cash flow by the total number of shares outstanding. Free Cash Flow per share is low for the following: -

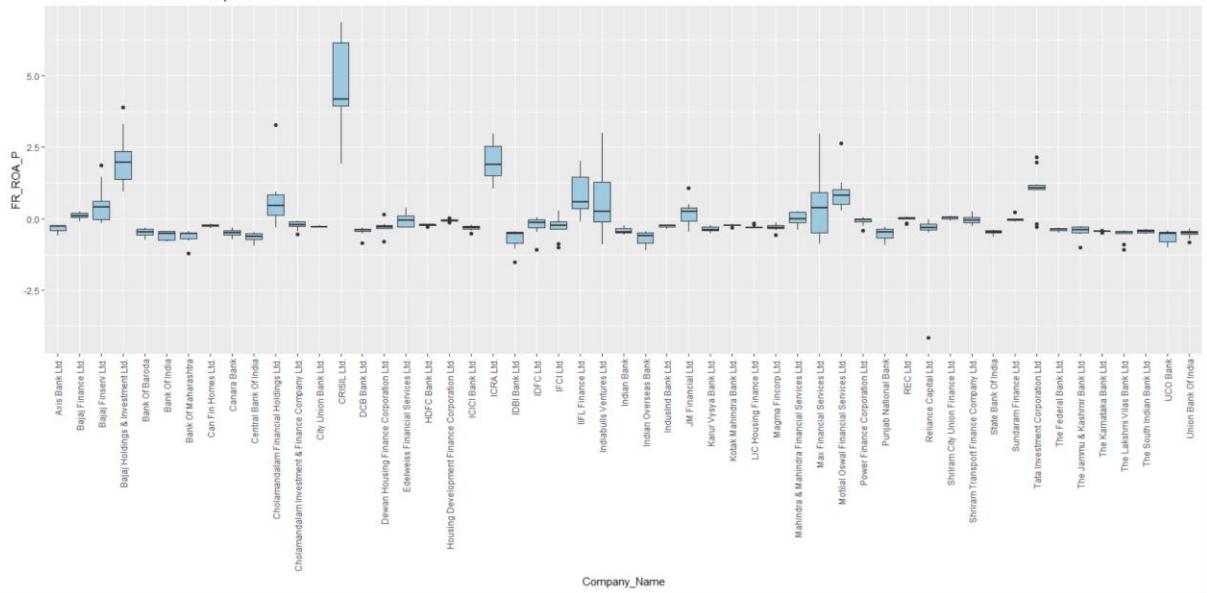
- Shriram City Union Finance LTD
- Shriram Transport Finance LTD.
- Sundaram Finance LTD.



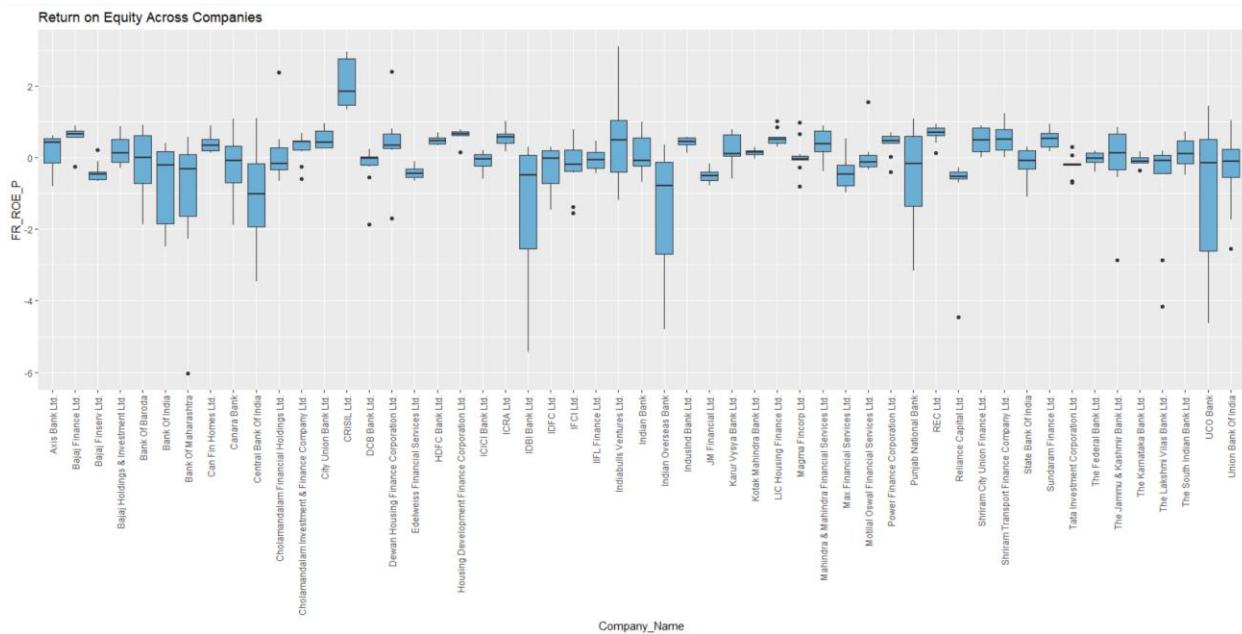
Adjusted DPS is high for the following: -

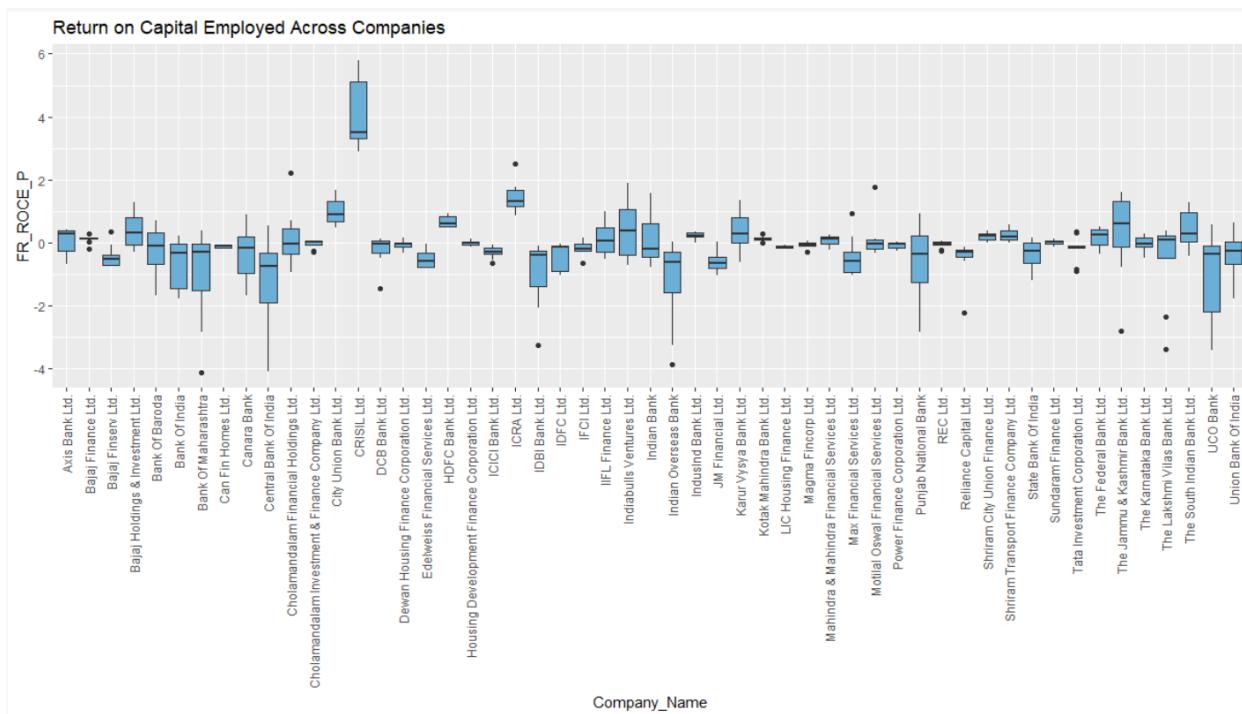
- Bajaj Holdings and Investments
- CRISIL
- ICRA
- HDFC
- Tata Investment Corporation
- Shriram City Union

Return on Assets Across Companies



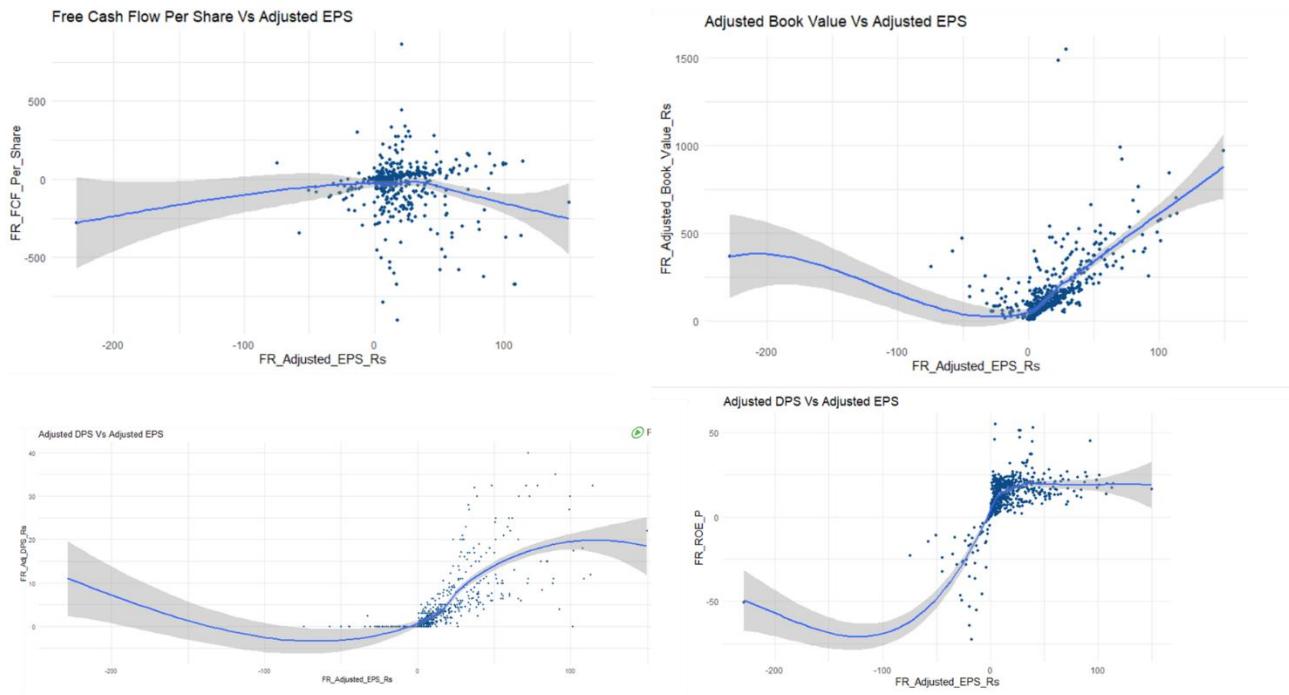
Return on Equity Across Companies



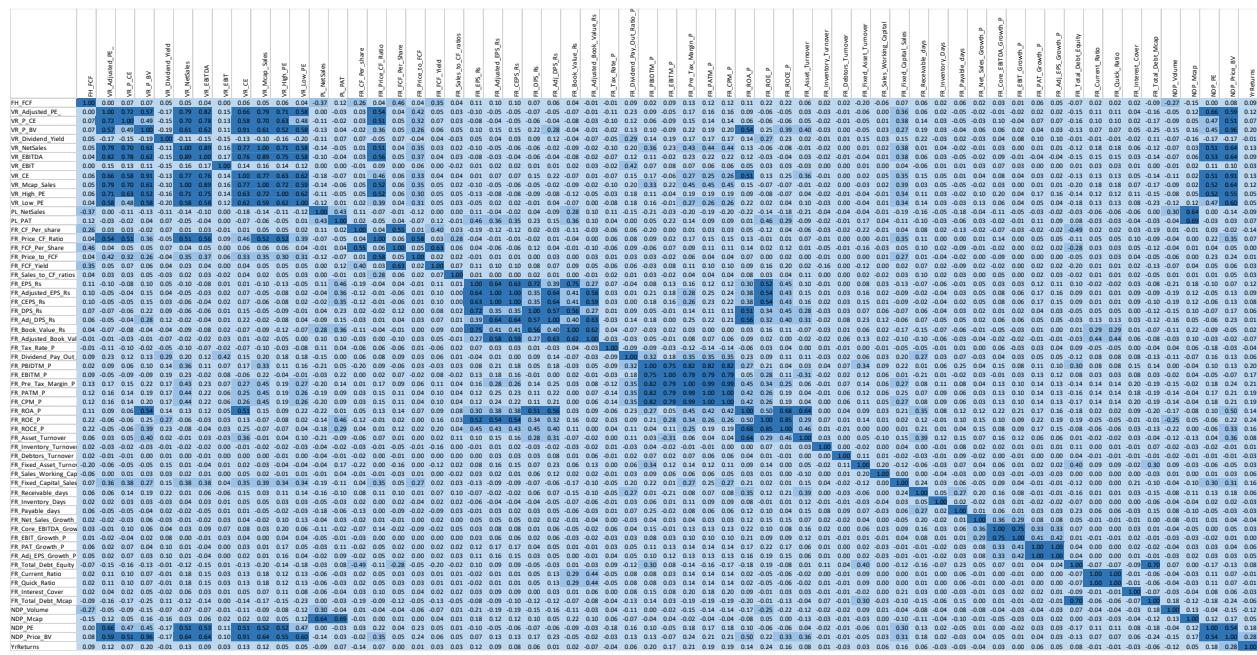


## Bivariate Analysis

Following are the plots of Adjusted Earnings per Share with select variables:-



## Correlation Plot



Those correlation values between ratios / variables greater than 0.5 or less than -0.5 are high (positive or negative correlation).

	FR_Adjusted_EPS_Rs		FR_Adjusted_EPS_Rs		FR_Adjusted_EPS_Rs
FH_FCF	0.10	FH_FCF	<b>0.011767</b>	FR_PBIDTM_P	0.21
VR_Adjusted_PE_	-0.05	VR_Adjusted_PE_	<b>0.246031</b>	FR_EBITM_P	<b>5.95E-07</b>
VR_P_CE	-0.04	VR_P_CE	<b>0.323138</b>	FR_Pre_Tax_Margin_P	<b>1.61E-05</b>
VR_P_BV	0.15	VR_P_BV	<b>0.000334</b>	FR_PATM_P	<b>4.35E-12</b>
VR_Dividend_Yield	0.04	VR_Dividend_Yield	<b>0.384046</b>	FR_CPM_P	<b>1.77E-09</b>
VR_NetSales	-0.05	VR_NetSales	<b>0.261766</b>	FR_ROA_P	<b>4.90E-09</b>
VR_EBITDA	-0.03	VR_EBITDA	<b>0.411306</b>	FR_ROE_P	<b>0</b>
VR_EBIT	0.02	VR_EBIT	<b>0.6553</b>	FR_ROCE_P	<b>0</b>
VR_CE	0.07	VR_CE	<b>0.070445</b>	FR_Asset_Turnover	<b>0.00022</b>
VR_Mcap_Sales	-0.05	VR_Mcap_Sales	<b>0.218164</b>	FR_Inventory_Turnov	<b>0.760315</b>
VR_High_Pe	-0.08	VR_High_Pe	<b>0.067538</b>	FR_Debtors_Turnover	<b>0.412165</b>
VR_Low_Pe	0.02	VR_Low_Pe	<b>0.570753</b>	FR_Fixed_Asset_Turn	<b>6.73E-05</b>
PL_NetSales	-0.04	PL_NetSales	<b>0.381383</b>	FR_Sales_Working_Cap	<b>0.598432</b>
PL_PAT	0.36	PL_PAT	<b>0</b>	FR_Fixed_Capital_Sale	<b>0.037669</b>
FR_CF_Per_share	-0.12	FR_CF_Per_share	<b>0.002844</b>	FR_Receivable_days	<b>-0.02</b>
FR_Price_CF_Ratio	-0.01	FR_Price_CF_Ratio	<b>0.79322</b>	FR_Inventory_Days	<b>-0.04</b>
FR_FCF_Per_Share	-0.06	FR_FCF_Per_Share	<b>0.168199</b>	FR_Payable_days	<b>0.03</b>
FR_Price_to_FCF	0.01	FR_Price_to_FCF	<b>0.836243</b>	FR_Net_Sales_Growth	<b>0.05</b>
FR_FCF_Yield	0.10	FR_FCF_Yield	<b>0.015611</b>	FR_Core_EBITDA_Grov	<b>0.08</b>
FR_Sales_to_CF_ratios	0.00	FR_Sales_to_CF_ratio	<b>0.994245</b>	FR_EBIT_Growth_P	<b>0.06</b>
FR_EPS_Rs	0.64	FR_EPS_Rs	<b>0</b>	FR_PAT_Growth_P	<b>0.17</b>
FR_Adjusted_EPS_Rs	1.00	FR_Adjusted_EPS_Rs	<b>NA</b>	FR_Adj_EPS_Growth_P	<b>0.16</b>
FR_CEPS_Rs	1.00	FR_CEPS_Rs	<b>0</b>	FR_Total_Debt_Equity	<b>0.09</b>
FR_DPS_Rs	0.35	FR_DPS_Rs	<b>0</b>	FR_Current_Ratio	<b>0.01</b>
FR_Adj_DPS_Rs	0.64	FR_Adj_DPS_Rs	<b>0</b>	FR_Quick_Ratio	<b>0.01</b>
FR_Book_Value_Rs	0.41	FR_Book_Value_Rs	<b>0</b>	FR_Interest_Cover	<b>0.09</b>
FR_Adjusted_Book_Val	0.58	FR_Adjusted_Book_Val	<b>0</b>	FR_Total_Debt_Mcap	<b>-0.09</b>
FR_Tax_Rate_P	0.03	FR_Tax_Rate_P	<b>0.470075</b>	YrReturns	<b>0.09</b>
FR_Dividend_Pay_Out	0.01	FR_Dividend_Pay_Out	<b>0.892431</b>		

Target Variable Vs Predictor Variables – Correlation Values and P-Values

Ratios such as Earnings Per Share, Cash Earnings Per Share have a high degree of correlation with Adjusted Earnings per share. Adjusted Earnings per share is a derived variable from EPS and CEPS. The strength of the correlation is significant between Adjusted EPS and Dividend Per Share, Book Value and Free Cash Flow.

### **Linear Regression Model**

#### Full Model

The data related to Financial Services has been split to form Train Data (70%) and Test Data (30%). Linear Regression Model has been run on the Scaled Train Data and the following table highlights in blue the significant variables ( $p\text{-value} < 0.05$ ).

Variables	Coefficients	P-Value
FR_CEPS_Rs	0.99215	4.24E-264
FR_PATM_P	0.55180	5.56E-10
(Intercept)	0.39998	0.8376
FR_Quick_Ratio	0.20234	0.47176
VR_NetSales	0.07441	0.079193
FR_EPS_Rs	0.02349	0.085234
FR_PAT_Growth_P	0.02151	0.445605
FR_ROCE_P	0.02134	0.001258
VR_EBITDA	0.01974	0.082454
PL_PAT	0.01301	0.000497
FR_Fixed_Capital_Sales	0.01190	0.003104
FR_Adj_DPS_Rs	0.01081	0.036765
FR_Pre_Tax_Margin_P	0.00827	0.757619
FR_PBIDTM_P	0.00748	0.3833
FR_FCF_Per_Share	0.00711	0.025649
FR_Total_Debt_Equity	0.00702	0.192281
FR_Fixed_Asset_Turnover	0.00701	0.007072
MomentumStrong Positive	0.00647	0.452416
FR_ROA_P	0.00598	0.342513
VR_P_CE	0.00370	0.29769
FR_EBIT_Growth_P	0.00192	0.864298
FR_Payable_days	0.00189	0.525734
FR_FCF_Yield	0.00168	0.540731
VR_High_PE	0.00135	0.777079
VR_Dividend_Yield	0.00118	0.668699
FR_Receivable_days	0.00115	0.661363
FR_Dividend_Pay_Out_Ratio_P	0.00083	0.756113
FR_Debtors_Turnover	0.00080	0.621345
FR_Net_Sales_Growth_P	0.00075	0.804781
FR_Sales_to_CF_ratios	0.00045	0.811078
FR_Price_to_FCF	0.00032	0.905229
FR_Sales_Working_Capital	-0.00013	0.946295
FR_Year_End	-0.00020	0.838116
FR_Inventory_Days	-0.00081	0.765543
FR_Core_EBITDA_Growth_P	-0.00150	0.694716
FR_Price_CF_Ratio	-0.00199	0.629711
FR_Interest_Cover	-0.00211	0.587842
FR_Inventory_Turnover	-0.00224	0.533775
VR_EBIT	-0.00225	0.303715
FH_FCF	-0.00283	0.284373
MomentumNegative	-0.00292	0.751359
VR_CE	-0.00312	0.729617
VR_Low_PE	-0.00357	0.310584
FR_Tax_Rate_P	-0.00418	0.295084
MomentumPositive	-0.00424	0.636419
VR_Adjusted_PE_	-0.00489	0.244433
FR_Total_Debt_Mcap	-0.00601	0.076649
FR_CF_Per_share	-0.00628	0.047788
MomentumStrong Negative	-0.00837	0.382848
VR_P_BV	-0.00873	0.191697
FR_DPS_Rs	-0.00889	0.107139
PL_NetSales	-0.00935	0.003132
FR_Book_Value_Rs	-0.01555	0.143849
FR_Adjusted_Book_Value_Rs	-0.01575	0.079006
FR_Adj_EPS_Growth_P	-0.01911	0.494984
FR_Asset_Turnover	-0.02339	4.68E-11
FR_EBITM_P	-0.02445	0.001723
FR_ROE_P	-0.02978	6.70E-05
VR_Mcap_Sales	-0.09154	0.028832
FR_Current_Ratio	-0.19190	0.494933
FR_CPM_P	-0.53083	7.06E-11

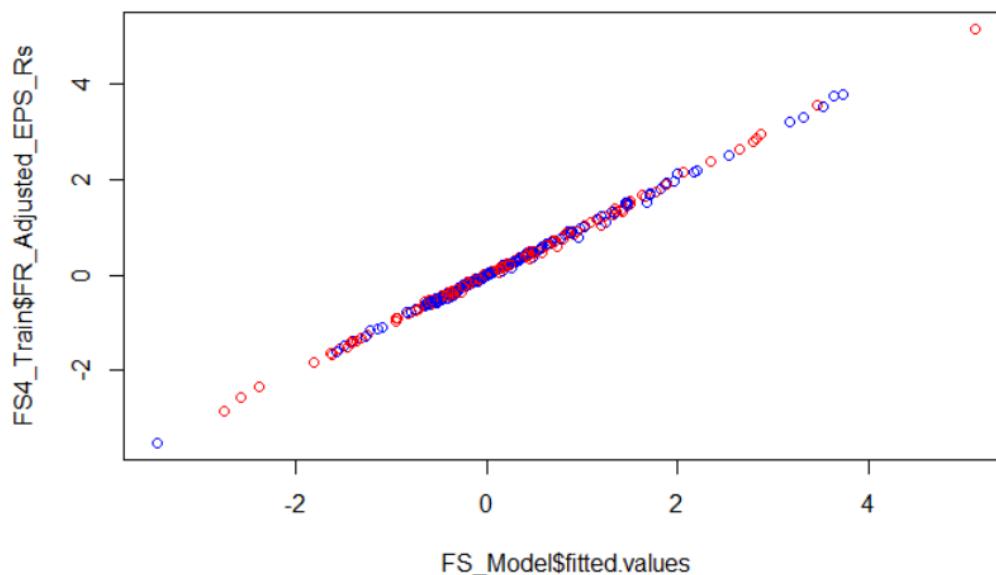
The following ratios are *significant and have the highest coefficients*: -

- Cash Earnings per share (FR\_CEPS\_Rs)
- After Tax Profit Margin (FR\_PATM\_P)

Other ratios which are significant are: -

- Asset Turnover
- CPM
- Return on Equity
- Profit After Tax
- Return on Capital Employed
- Earnings before Income, Tax and Management
- Net Sales (PL)
- Fixed Asset Turnover
- Free Cash Flow per Share
- Mcap Sales
- Adjusted Dividend Per Share
- Cash Flow Per Share

### Fitted Vs Actual Values - Full Model



The Adjusted R<sup>2</sup> of the Full Model is **99.85%** and is highly significant. However, the model performance measures are mid-leading on account of the presence of variables / ratios which are multi-collinear.

The following table represents the multi-collinearity between ratios (predictor variables): -

Variables	Coefficients	P-Value	VIF					
FR_Current_Ratio	-0.19190	0.494933	33039.03	FR_Total_Debt_Equity	0.00702	0.192281	8.938985	
FR_Quick_Ratio	0.20234	0.47176	33033.09	FR_Adj_DPS_Rs	0.01081	0.036765	8.748811	
FR_PATM_P	0.55180	5.56E-10	1712.774	FR_EBIT_Growth_P	0.00192	0.864298	4.620283	
FR_CPM_P	-0.53083	7.06E-11	1416.327	VR_High_PE	0.00135	0.777079	4.595357	
VR_NetSales	0.07441	0.079193	330.1723	FR_Fixed_Capital_Sales	0.01190	0.003104	3.874946	
VR_Mcap_Sales	-0.09154	0.028832	324.1974	VR_Adjusted_PE	-0.00489	0.244433	3.863055	
FR_PAT_Growth_P	0.02151	0.445605	215.8806	FR_Asset_Turnover	-0.02339	4.68E-11	3.828152	
FR_Adj_EPS_Growth_P	-0.01911	0.494984	213.6515	PL_PAT	0.01301	0.000497	3.816409	
FR_Pre_Tax_Margin_P	0.00827	0.757619	169.3205	FR_Net_Sales_Growth_P	0.00075	0.804781	3.66078	
FR_EPS_Rs	0.02349	0.085234	53.24696	FR_FCF_Per_Share	0.00711	0.025649	3.356799	
FR_Book_Value_Rs	-0.01555	0.143849	32.57124	PL_NetSales	-0.00935	0.003132	3.353006	
FR_Adjusted_Book_Value_Rs	-0.01575	0.079006	27.66235	FR_Total_Debt_Mcap	-0.00601	0.076649	3.231218	
FR_CEPS_Rs	0.99215	4.24E-264	25.29082	FR_Core_EBITDA_Growth_P	-0.00150	0.694716	3.188563	
VR_EBITDA	0.01974	0.082454	21.75641	FR_Year_End	-0.00020	0.838116	2.892327	
VR_CE	-0.00312	0.729617	18.01057	FR_CF_Per_share	-0.00628	0.047788	2.841625	
FR_PBIDTM_P	0.00748	0.3833	18.00054	VR_Dividend_Yield	0.00118	0.668699	2.777401	
FR_ROE_P	-0.02978	6.70E-05	14.438	FR_Dividend_Pay_Out_Ratio_P	0.00083	0.756113	2.732928	
FR_EBITM_P	-0.02445	0.001723	13.2461	VR_P_CE	0.00370	0.29769	2.730835	
FR_ROCE_P	0.02134	0.001258	12.37839	FR_FCF_Ratio	-0.00199	0.629711	2.710177	
FR_ROA_P	0.00598	0.342513	12.00929	VR_Low_PE	-0.00357	0.310584	2.593157	
FR_DPS_Rs	-0.00889	0.107139	11.34254	FR_Inventory_Turnover	-0.00224	0.533775	2.563002	
VR_P_BV	-0.00873	0.191697	10.86094	FR_FCF	-0.00283	0.284373	2.456112	
				FR_FCF_Yield	0.00168	0.540731	2.420346	
				FR_Tax_Rate_P	-0.00418	0.295084	2.359385	
				FR_Receivable_days	0.00115	0.661363	2.328814	
				FR_Fixed_Asset_Turnover	0.00701	0.007072	2.324954	
				VR_EBIT	-0.00225	0.303715	1.980424	
				FR_Payable_days	0.00019	0.525734	1.912381	
				FR_Price_to_FCF	0.00032	0.905229	1.748185	
				FR_Inventory_Days	-0.00081	0.765543	1.30246	
				FR_Sales_to_CF_ratios	0.00045	0.811078	1.282581	
				FR_Interest_Cover	-0.00021	0.587842	1.26897	
				FR_Sales_Working_Capital	-0.00013	0.946295	1.178585	
				FR_Debtors_Turnover	0.00080	0.621345	1.091029	

Following tables represents the variables that are high

Correlated Ratio Pairs	
FR 1	FR 2
Cash Earnings Per Share ( <i>FR_CEPS_Rs</i> )	Earnings per Share ( <i>FR_EPS_Rs</i> )
	Adjusted EPS Growth ( <i>FR_Adj_EPS_Growth_P</i> )
	Earnings before Income, Tax and Management ( <i>FR_EBITM_P</i> )
	Earnings before interest, taxes, depreciation, and amortization ( <i>VR_EBITDA</i> )
Asset Turnover ( <i>FR_Asset_Turnover</i> )	Fixed Asset Turnover ( <i>FR_Fixed_Asset_Turnover</i> )
After Tax Profit Margin ( <i>FR_PATM_P</i> )	Profit After Tax ( <i>PL_PAT</i> )
	<i>FR_CPM_P</i> ((Net Profit + Depreciation)/Sales)
Free Cash Flow per Share ( <i>FR_FCF_Per_Share</i> )	Cash Flow Per Share ( <i>FR_CF_Per_share</i> )
VR_Mcap_Sales	VR_NetSales
Quick Ratio ( <i>FR_Quick_Ratio</i> )	Current Ratio ( <i>FR_Current_Ratio</i> )
FR_Adjusted_Book_Value_Rs	FR_Book_Value_Rs
Adjusted Dividend Per Share ( <i>FR_Adj_DPS_Rs</i> )	Dividend Per Share ( <i>FR_DPS_Rs</i> )

Following ratios (variables) are required to be removed based on correlation and multi-collinearity: -

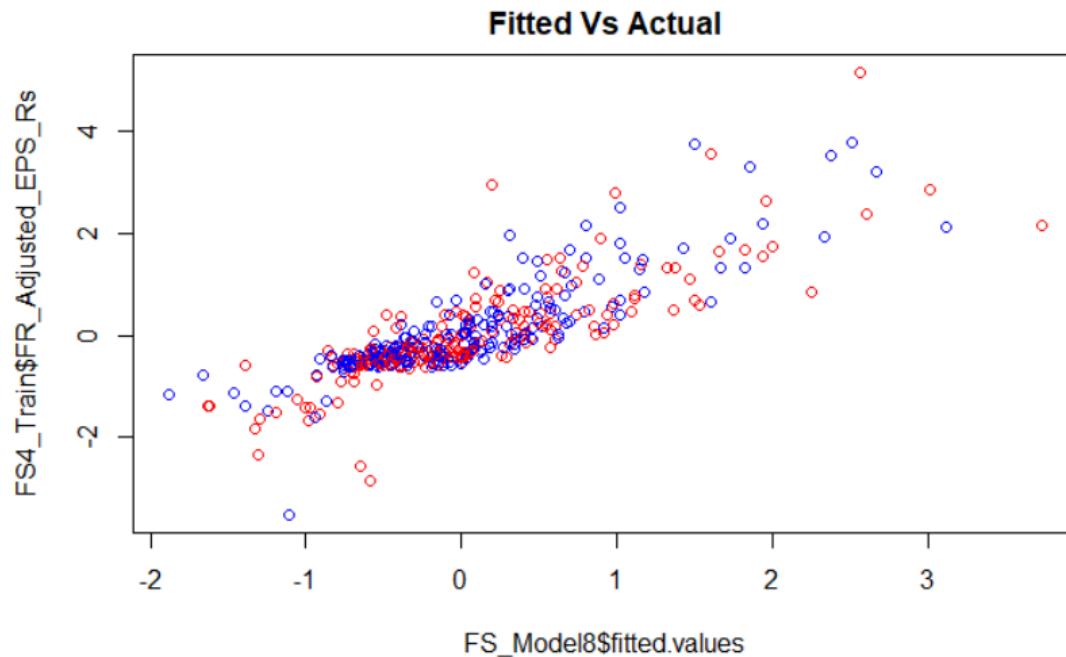
- Cash Earnings Per Share and Earnings Per Share has a high degree of correlation with Adjusted Earnings Per Share
- Fixed Asset Turnover has a high degree of multi-collinearity with Asset Turnover
- Cash Flow Per Share has a high degree of multi-collinearity with Free Cash Flow Per Share
- Book Value has a high degree of multi-collinearity with Adjusted Book Value
- Adjusted Dividend Per Share has a high degree of multi-collinearity with Dividend Per Share and other Dividend related ratios
- Current Ratio has a high degree of multi-collinearity with Quick Ratio
- Market Capitalization to Sales Ratio has a high degree of multi-collinearity with Earnings before interest, taxes, depreciation, and amortization
- After Tax Profit Margin has a high multi-collinearity with CPM Ratio (Adjusted Net Profit + Depreciation / Sales) and

## Model#2

Based on the above ratios / variables that are deleted the and on applying 8 iterations, the following ratios are shortlisted. These variables have a low degree of multi-collinearity.

Variables	Coefficients	Pr(> t )	VIF
FR_Adj_DPS_Rs	0.6852	0.0000	2.502936
FR_ROCE_P	0.3548	0.0000	2.431356
FR_Total_Debt_Equity	0.2053	0.0000	2.659024
VR_Mcap_Sales	0.1853	0.0365	6.711074
FR_Book_Value_Rs	0.1657	0.0000	1.671297
VR_Low_PE	0.0650	0.1247	1.732233
FR_Interest_Cover	0.0415	0.4394	1.115995
VR_High_PE	0.0376	0.5161	3.12451
VR_P_CE	0.0280	0.5390	2.075809
FR_Inventory_Days	0.0252	0.4893	1.074133
FR_Inventory_Turnover	0.0241	0.4684	1.010453
(Intercept)	0.0147	0.5914	
FR_Asset_Turnover	-0.0026	0.9452	2.159181
FR_Receivable_days	-0.0040	0.8984	1.495642
VR_Adjusted_PE_	-0.0099	0.8462	2.61566
FR_Debtors_Turnover	-0.0131	0.5714	1.039617
FR_Payable_days	-0.0164	0.6544	1.338538
FR_PATM_P	-0.0225	0.6934	3.428682
FR_Price_to_FCF	-0.0258	0.4639	1.377546
FR_Price_CF_Ratio	-0.0341	0.4884	1.779205
FR_Tax_Rate_P	-0.0432	0.3534	1.484805
FR_FCF_Per_Share	-0.1040	0.0002	1.174635
FR_Quick_Ratio	-0.1143	0.0000	1.306285
FR_Total_Debt_Mcap	-0.1351	0.0023	2.511909
VR_P_BV	-0.1923	0.0020	4.318624
FR_ROA_P	-0.2457	0.0001	5.533808

The Adjusted R<sup>2</sup> based on the above model is **67.3%**.

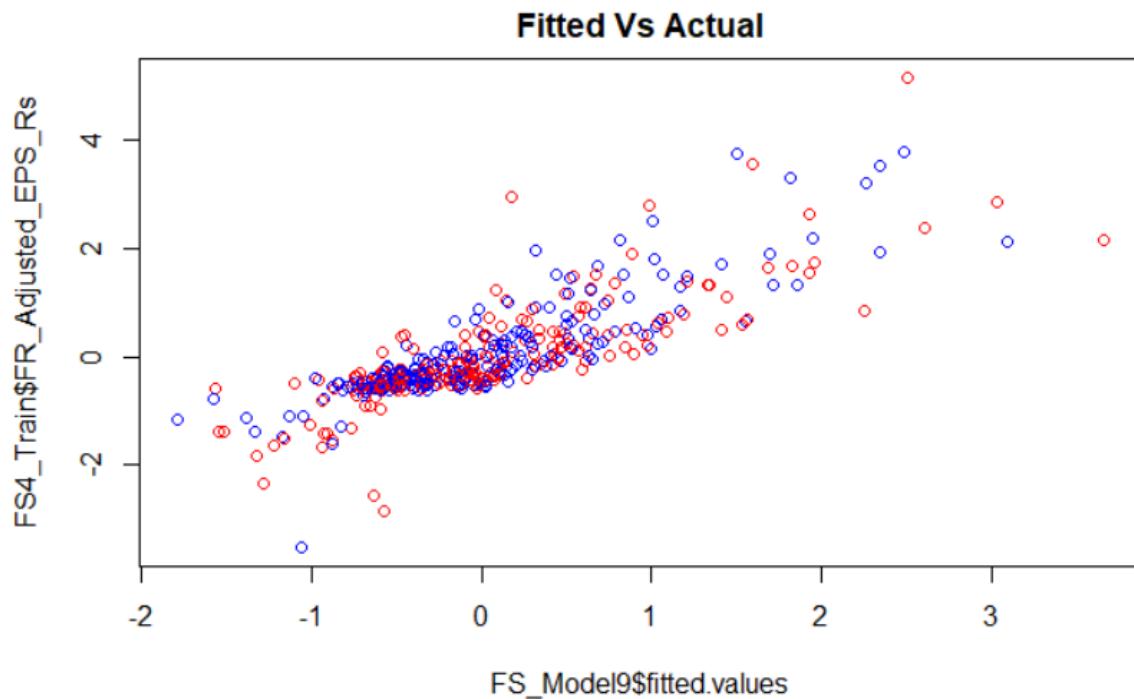


#### Model#3 – Significant Variables

Following are the model outputs based on the significant variables (p-value <0.05) filtered out from the previous model. The table also displays the multi-collinearity values (VIF)

Variables	Estimate	Pr(> t )	VIF
(Intercept)	0.013655	0.6105224	
FR_Adj_DPS_Rs	0.679075	0.0000000	2.296166
FR_ROCE_P	0.342295	0.0000000	2.288791
FR_Total_Debt_Equity	0.184762	0.0000028	2.195462
VR_Mcap_Sales	0.203128	0.0000202	1.938147
FR_Book_Value_Rs	0.16001	0.0000049	1.61652
FR_FCF_Per_Share	-0.10467	0.0001348	1.152115
FR_Quick_Ratio	-0.11056	0.0000160	1.25842
FR_Total_Debt_Mcap	-0.12578	0.0014402	2.031351
VR_P_BV	-0.15544	0.0011795	2.58767
FR_ROA_P	-0.25243	0.0000005	3.480794

The Adjusted R<sup>2</sup> of the model is 67.78%.



#### Model#4

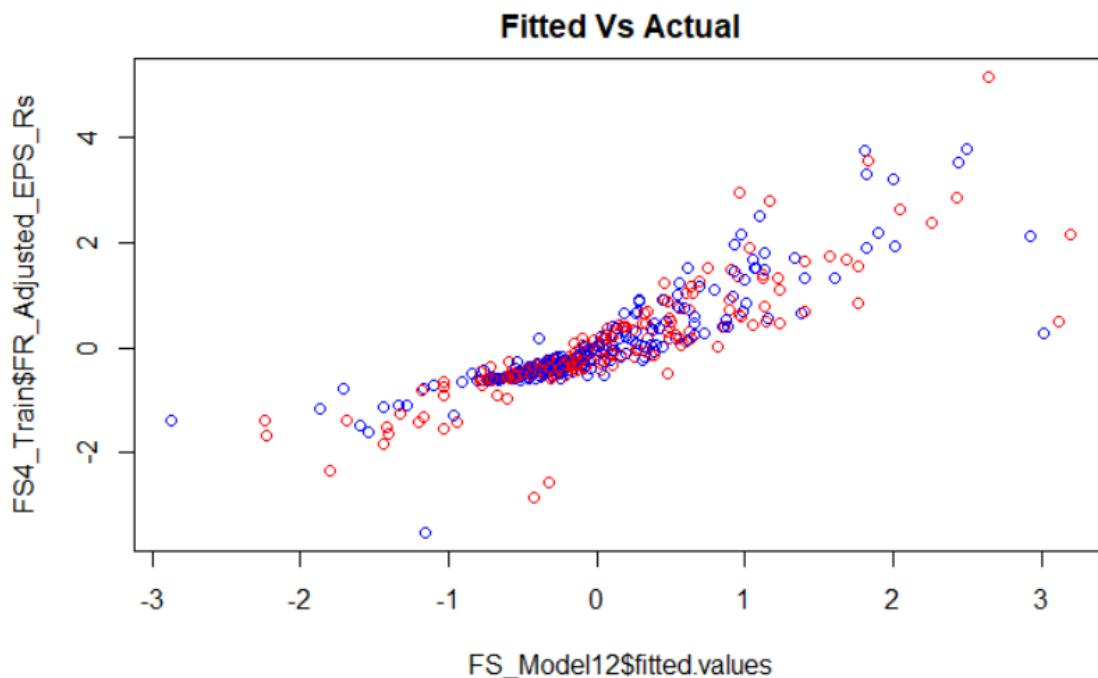
Following table shows the important variables and highlights those variables which will be considered in this model along with ensuring significance of variables (p-value) and minimizing multi-collinearity

Variables	Coefficients	P-Value	VIF	Variable Imp
FR_CEPS_Rs	0.99215	4.24E-264	25.29082	102.807221
FR_Asset_Turnover	-0.02339	4.68E-11	3.828152	6.79410115
FR_CPM_P	-0.53083	7.06E-11	1416.327	6.72673104
FR_PATM_P	0.55180	5.56E-10	1712.774	6.38100597
FR_ROE_P	-0.02978	6.70E-05	14.438	4.0351882
PL_PAT	0.01301	0.000497	3.816409	3.51546443
FR_ROCE_P	0.02134	0.001258	12.37839	3.25189996
FR_EBITM_P	-0.02445	0.001723	13.2461	3.1585938
FR_Fixed_Capital_Sales	0.01190	0.003104	3.874946	2.97788106
PL_NetSales	-0.00935	0.003132	3.353006	2.97504582
FR_Fixed_Asset_Turnover	0.00701	0.007072	2.324954	2.70934258
FR_FCF_Per_Share	0.00711	0.025649	3.356799	2.24102591
VR_Mcap_Sales	-0.09154	0.028832	324.1974	2.19481027
FR_Adj_DPS_Rs	0.01081	0.036765	8.748811	2.09636742
FR_CF_Per_share	-0.00628	0.047788	2.841625	1.98620299
FR_Total_Debt_Mcap	-0.00601	0.076649	3.231218	1.77569741
FR_Adjusted_Book_Value_Rs	-0.01575	0.079006	27.66235	1.76160766
VR_NetSales	0.07441	0.079193	330.1723	1.76050696
VR_EBITDA	0.01974	0.082454	21.75641	1.74160346
FR_EPS_Rs	0.02349	0.085234	53.24696	1.72596573
FR_DPS_Rs	-0.00889	0.107139	11.34254	1.61532448
FR_Book_Value_Rs	-0.01555	0.143849	32.57124	1.46488082
VR_P_BV	-0.00873	0.191697	10.86094	1.30809544

Following table displays the model outputs: -

Variables	Estimate	Pr(> t )	VIF
(Intercept)	0.006203	0.798404	
FR_Asset_Turnover	-0.04805	0.072774	1.311313
FR_ROE_P	0.348469	4.69E-22	1.7459
FR_FCF_Per_Share	-0.09783	4.08E-05	1.05421
FR_Adj_DPS_Rs	0.366387	5.85E-22	2.408047
FR_Adjusted_Book_Value_Rs	0.308943	1.61E-19	2.075334
PL_PAT	0.103475	0.000745	1.469697

The Adjusted R<sup>2</sup> of this model is **73.45%**



#### Linear Regression Model Evaluation Parameters

Model Types	RMSE
Full Model	0.044747
Model#2	1.003964
Model#3	1.002625
Model#4	0.804857
Lasso	1.000953
Ridge	0.9666798
Elasticnet	1.000573

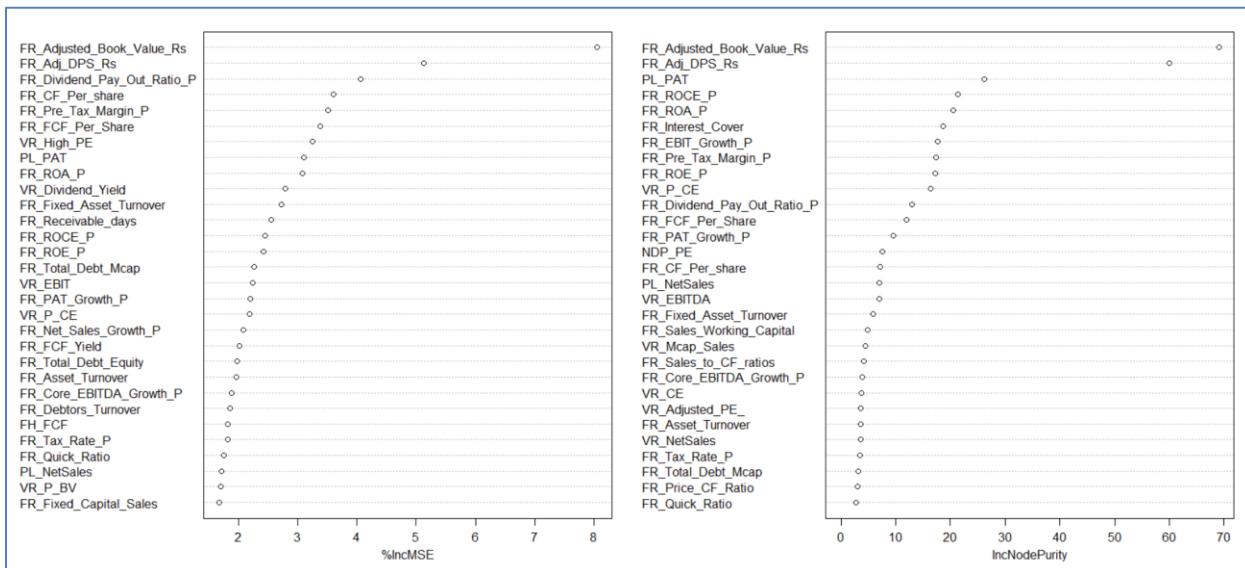
Model#4 has the best Model Evaluation Parameters and makes up the following ratios / variables: -

- **Asset Turnover Ratio** - measures the value of a company's sales or revenues relative to the value of its assets
- **Return on Equity** - measure of the profitability of a business in relation to the equity
- **Free Cash Flow per Share** - measure of a company's financial flexibility that is determined by dividing free cash flow by the total number of shares outstanding
- **Adjusted Dividend Per Share** - dividing the total dividends paid out by a business, including interim dividends, over a period of time by the number of outstanding ordinary shares issued
- **Adjusted Book Value** - measure of a company's valuation after liabilities—including off-balance sheet liabilities—and assets adjusted to reflect true fair market value
- **Profit After Tax**

## Random Forest

Random Forest Model has been run on the variables and following is the output based on 50 Trees and mtry value of 8: -

Variable Importance Plot



The variable importance plot gives the following variables as important

- Adjusted Book Value
- Adjusted Dividend Per Share
- Cash Flow per Share
- Pre-Tax Profit Margin
- Profit after Tax
- Return on Asset
- Return on Equity
- Cash Flow per share

The importance of variables is provided by the below table: -

Variables	Overall	
FR_Adjusted_Book_Value_Rs	8.053965	
FR_Adj_DPS_Rs	5.129441	
FR_Dividend_Pay_Out_Ratio_P	4.057803	
FR_CF_Per_share	3.607815	
FR_Pre_Tax_Margin_P	3.504581	
FR_FCF_Per_Share	3.376364	
VR_High_PE	3.246146	
PL_PAT	3.105273	
FR_ROA_P	3.071522	
VR_Dividend_Yield	2.782327	
FR_Fixed_Asset_Turnover	2.725296	
FR_Receivable_days	2.544626	
FR_ROCE_P	2.444434	
FR_ROE_P	2.425468	
FR_Total_Debt_Mcap	2.267134	
VR_EBIT	2.234029	
FR_PAT_Growth_P	2.200612	
VR_P_CE	2.182742	
FR_Net_Sales_Growth_P	2.071508	
FR_FCF_Yield	2.013511	
FR_Total_Debt_Equity	1.967417	
FR_Asset_Turnover	1.961157	
		FR_Core_EBITDA_Growth_P 1.873986
		FR_Debtors_Turnover 1.850782
		FH_FCF 1.820349
		FR_Tax_Rate_P 1.808866
		FR_Quick_Ratio 1.752527
		PL_NetSales 1.708861
		VR_P_BV 1.700129
		FR_Fixed_Capital_Sales 1.666609
		VR_Adjusted_PE_ 1.642795
		VR_EBITDA 1.543441
		FR_Inventory_Turnover 1.459817
		VR_NetSales 1.375948
		VR_CE 1.342345
		NDP_PE 1.320991
		VR_Low_PE 1.30578
		VR_Mcap_Sales 1.257958
		FR_Sales_to_CF_ratios 1.191066
		FR_Sales_Working_Capital 1.146736
		FR_Payable_days 0.908039
		FR_EBIT_Growth_P 0.239023
		FR_Price_CF_Ratio -0.15265
		FR_Current_Ratio -0.29607
		FR_Price_to_FCF -0.55378
		FR_Interest_Cover -1.00187
		FR_Inventory_Days -1.01015

## Overall Model Evaluation Parameters

Model Types	RMSE
Full Model	0.044747
Model#2	1.003964
Model#3	1.002625
Model#4	0.804857
Lasso	1.000953
Ridge	0.9666798
Elasticnet	1.000573
Random Forest	0.7265402
SVM	0.7803777

The RMSE score based on Random Forest model is 0.7265402. The important variables got out of the Random Forest Model matches with the significant variables got out of the Linear Regression Model

## **8. Manufacturing Macro Economic Sector**

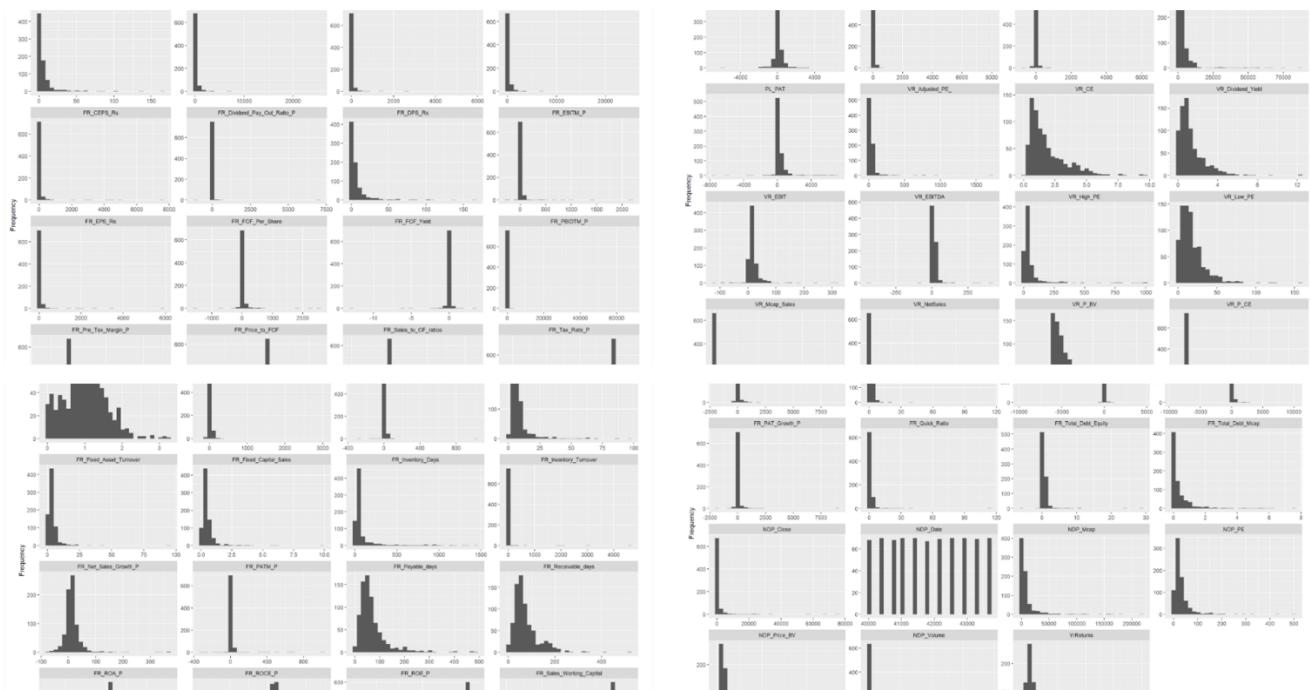
## Data Preprocessing:

- Below is a snapshot of the data. We have 761 observations for the manufacturing macro sector and 68 gross variables
  - There are 5 different industries and 70 companies' data considered for the manufacturing sector modelling for last 10 years (2009 to 2019)
  - There are no missing data
  - All the variables are numerical except for momentum which is a derived variable and will not be considered in the analysis
  - There are extreme values in the data for majority of the variables, but since they are not outliers, they are to be considered in the model building process, bring treated as **Extreme Value**

## Exploratory Data Analysis:

## Univariate Analysis

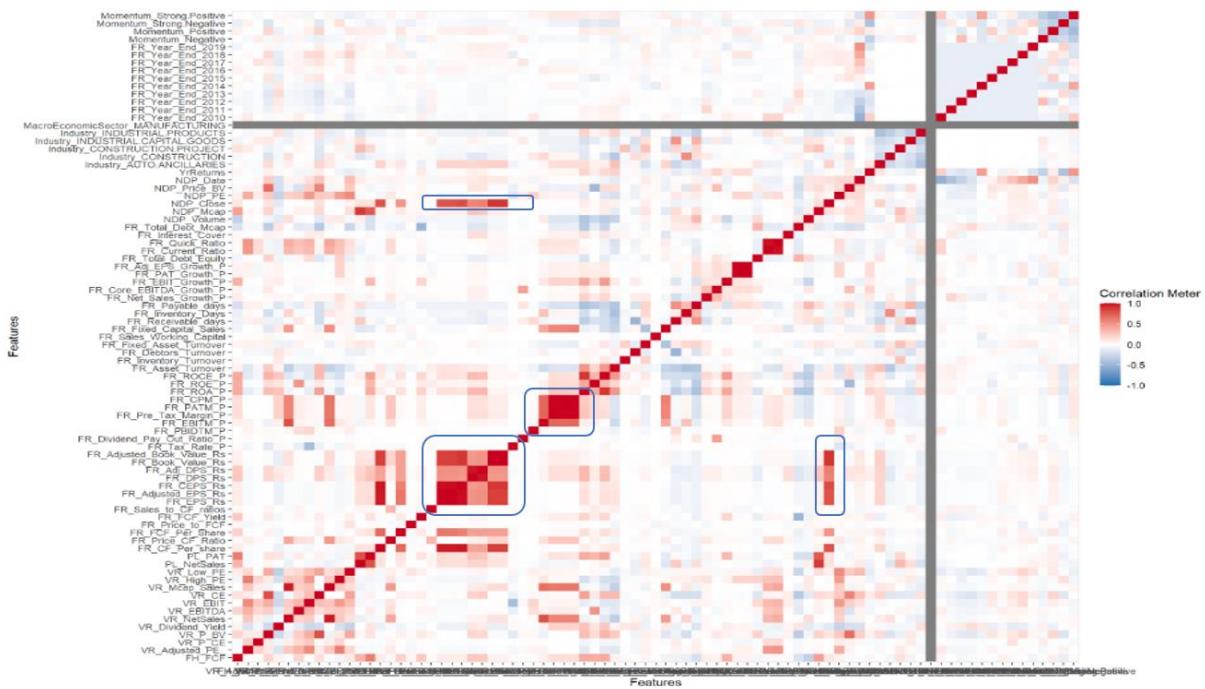
Below are the snapshots of univariate analysis:



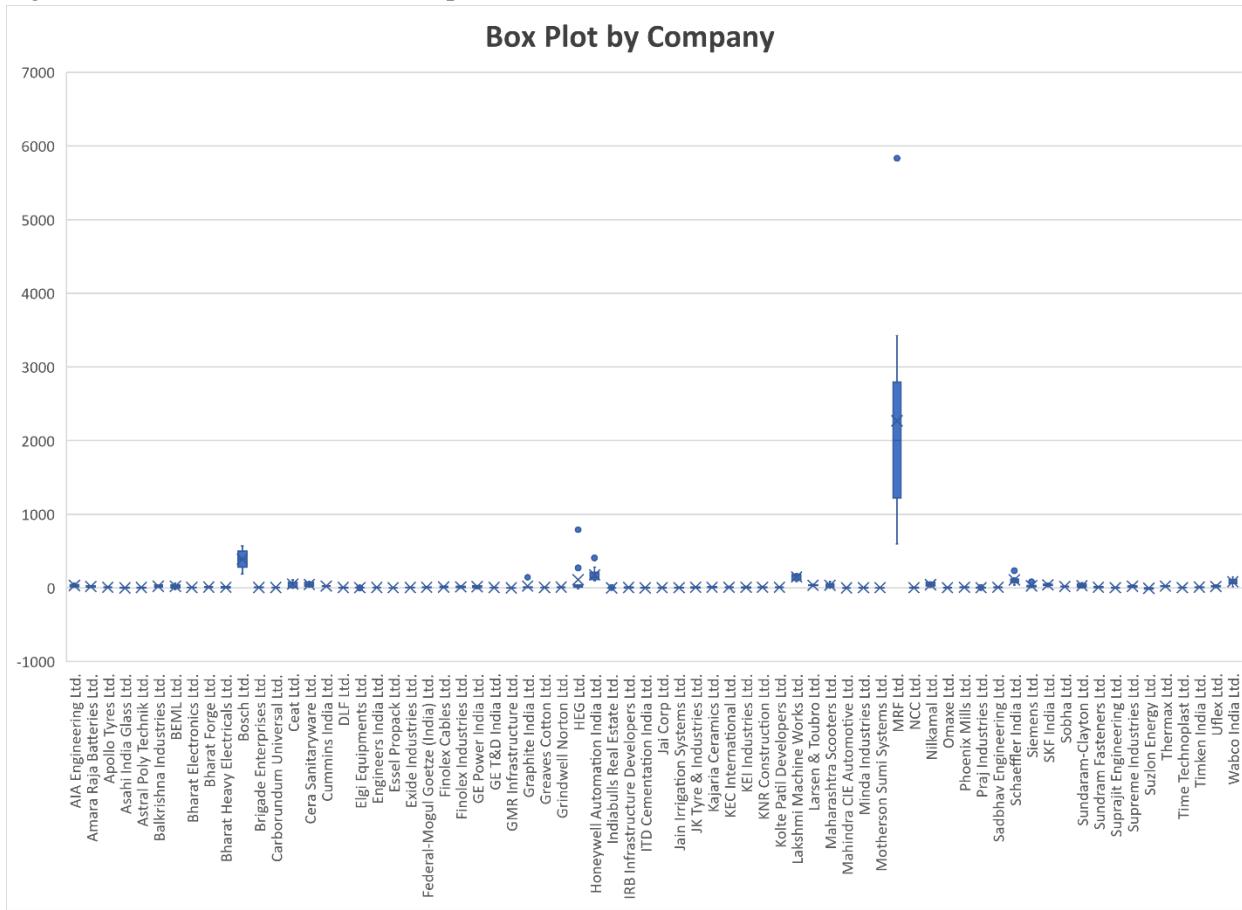
There are violations of normality from the histogram plots above for a lot of variables under consideration, but we are going to include these in the model building without any treatment.

### Initial Correlation Plot

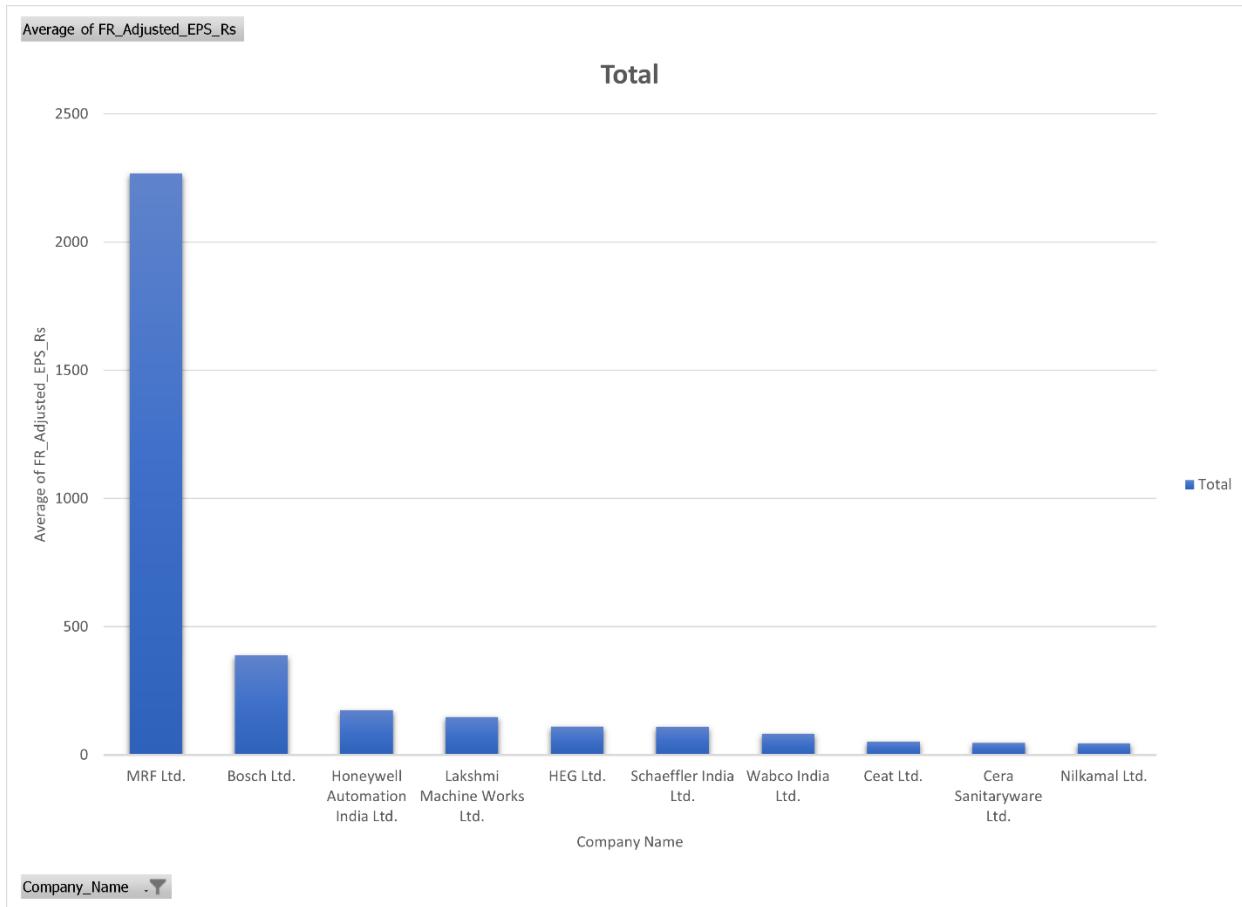
A correlation plot is generated as part of the report extract using the dataexplorer package. As depicted (Blue boxes) below, there may be a few cases of multicollinearity that needs to be analyzed further using the VIF analysis.



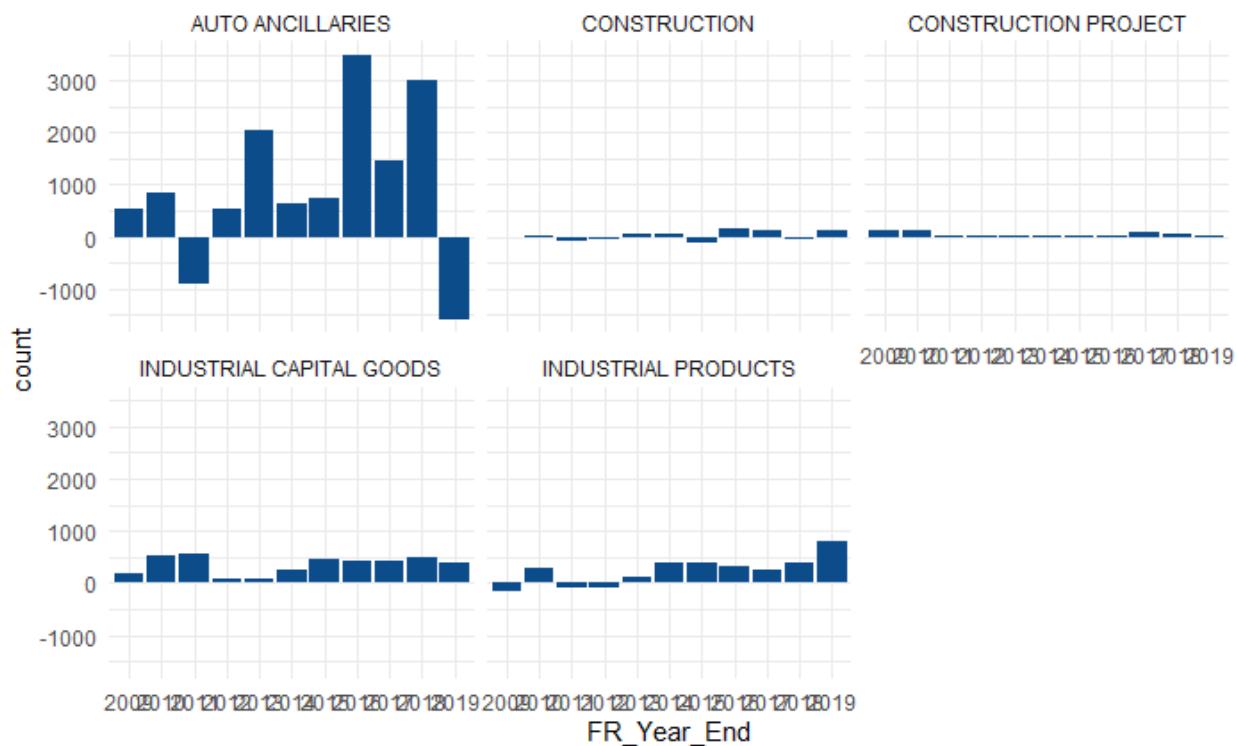
Below is a plot by companies and the FR\_Adjusted\_EPS\_Rs values. MRF seems to have significantly higher values than the rest of the companies.



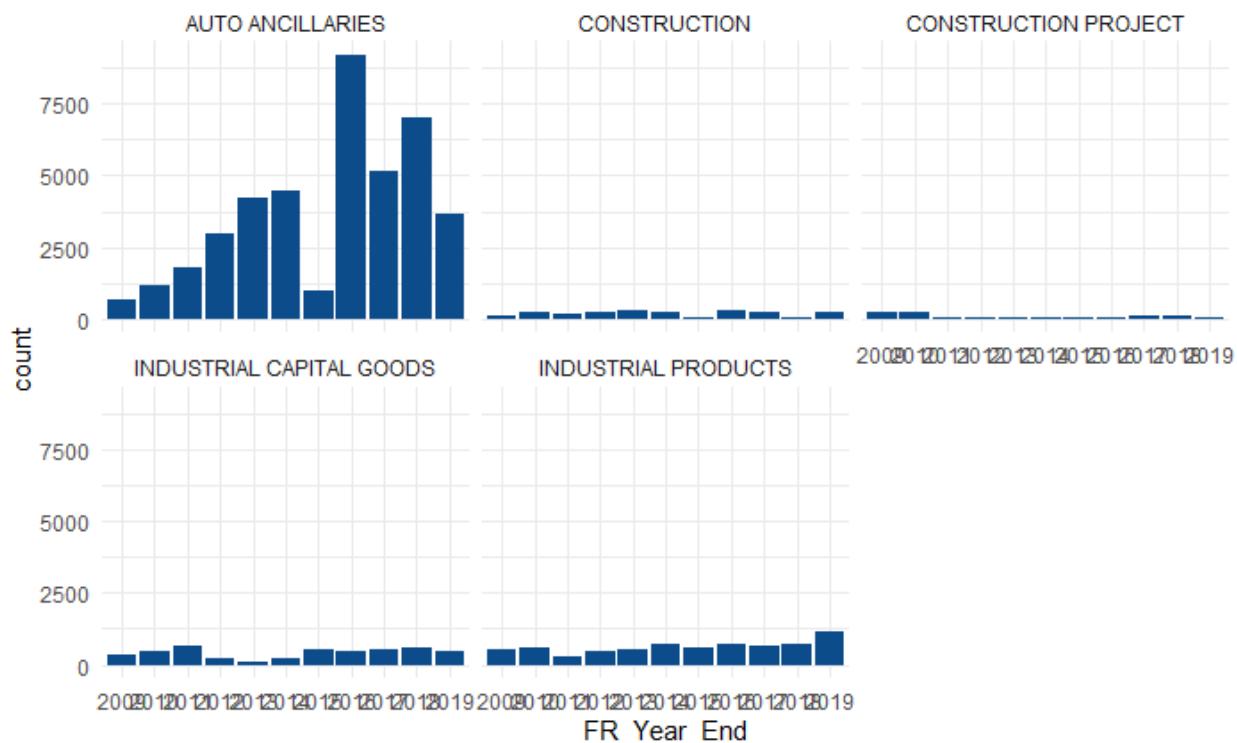
- Below are the top 10 companies by average FR\_Adjusted\_EPS\_Rs in the manufacturing sector. As we can see, MRF has significant higher averages and median values, followed by Bosch and Honeywell.



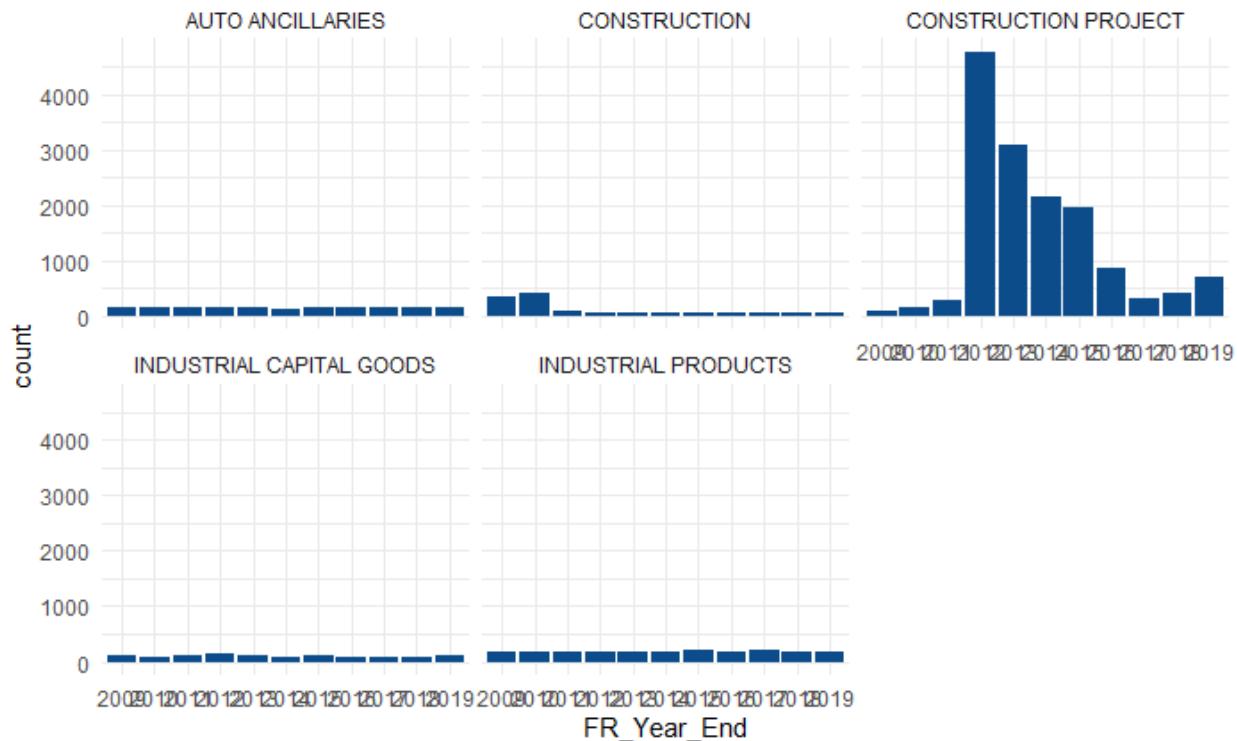
- Company Name ▾
- A deeper dive into the industry type reveals that FR\_CF\_Per\_share and average FR\_FCF\_Per\_Share for Auto Ancillaries throughout the year is significantly greater than the rest of the industry.



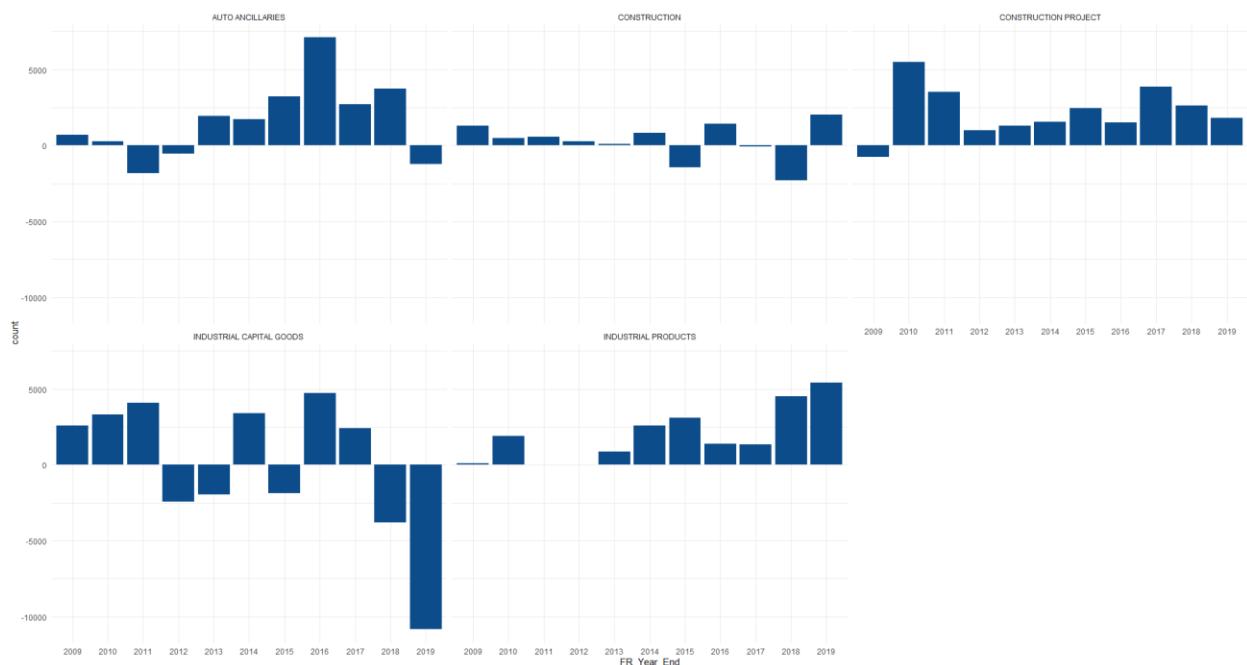
- Furthermore, the FR\_Adjusted\_Book\_Value\_Rs for Auto Ancillaries Grew significantly greater over the years than any other industry type.



- While FR\_ROE\_P Declined briskly through the years for Construction Industry, whereas others were fairly stable.



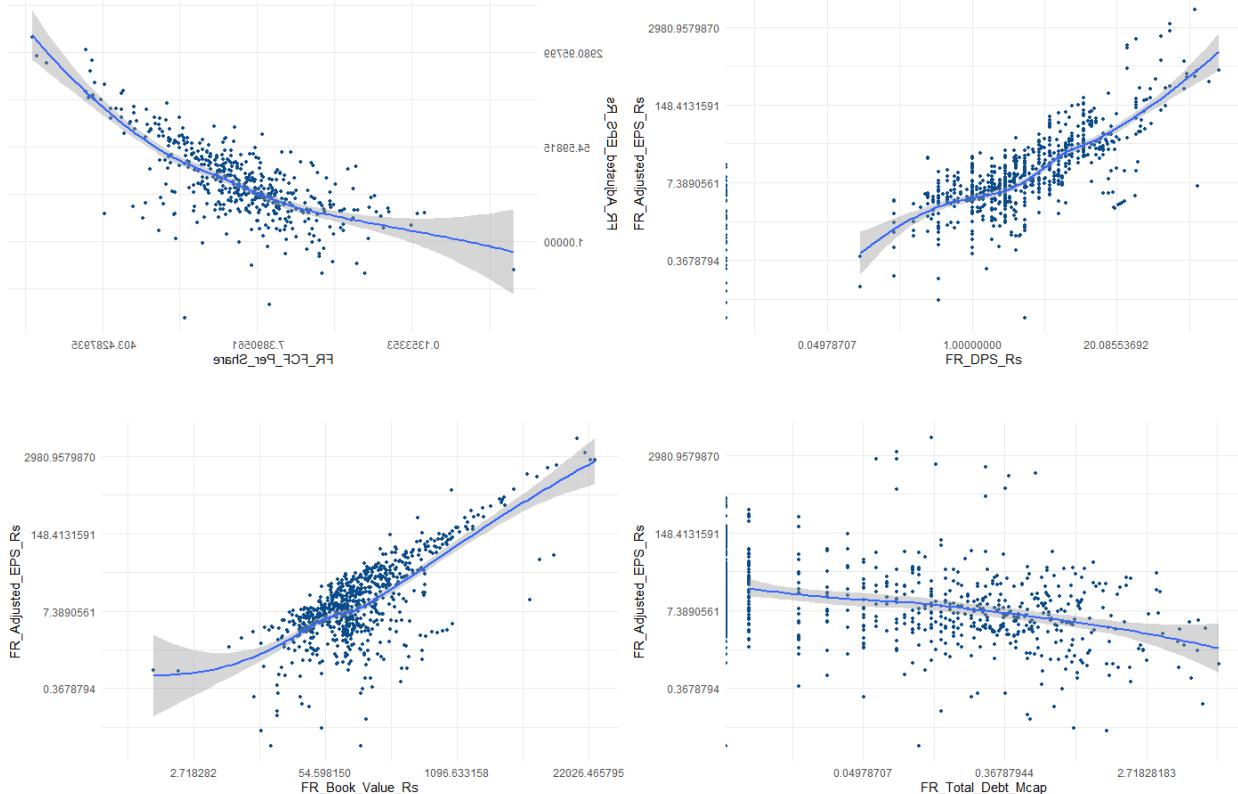
- One unique observation was that FH\_FCF had one big fall in 2019 mainly for Industrial Capital goods, larger than any other categories when the average values were fairly consistent.



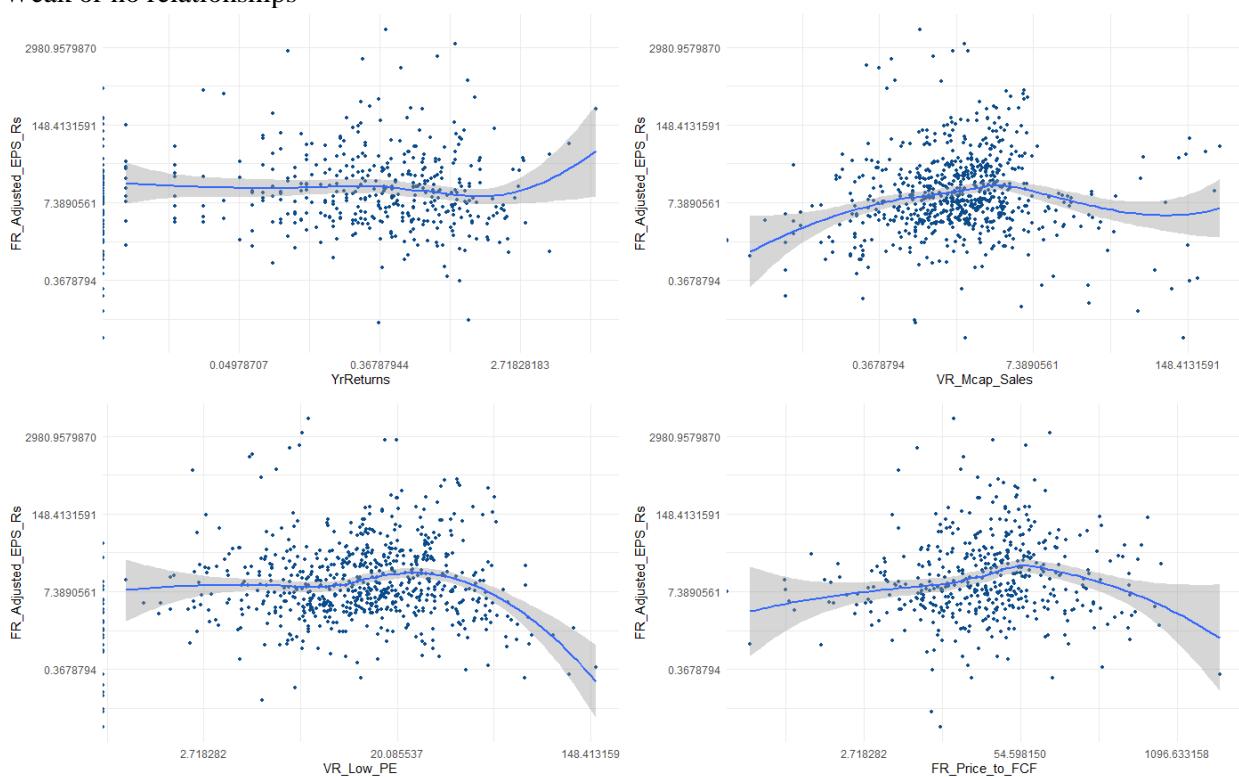
## Bivariate Analysis

- The dependent variable here is “FR\_Adjusted\_EPS\_Rs”, which is adjusted earning per share of the company for the year.
- Very few variables are actually showing a correlated linear relationship with the dependent variable as we can see in few of the scatter plots displayed below
- As seen from the graphs below, there may be a non linear relationships with the variables.
- The last graph shows that there may be some inverse relationship with FR\_Total\_Debt\_Mcap

## Strong Relationships



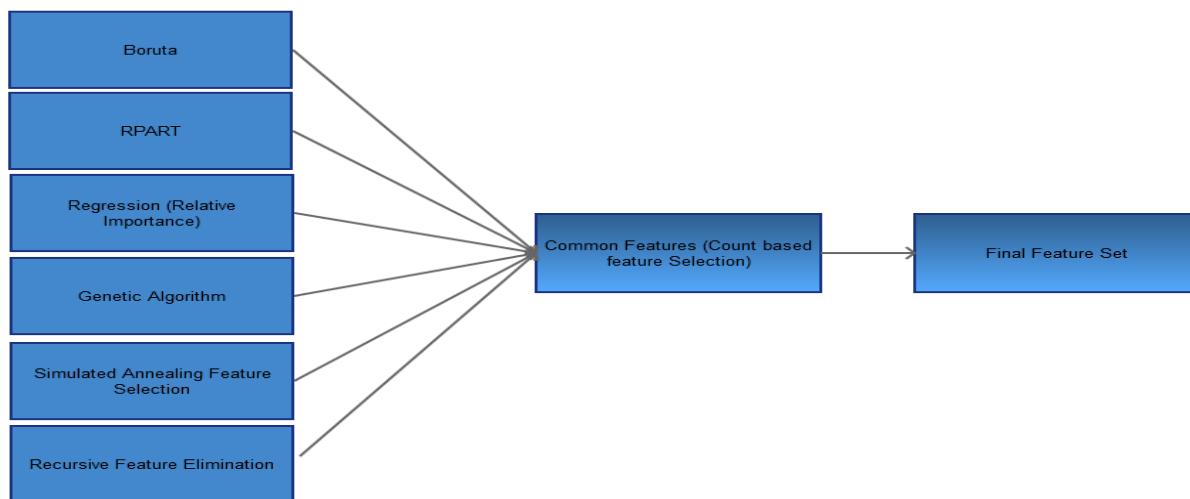
### Weak or no relationships



### Modelling Approach

- There were broadly two approaches used for the modeling exercise.
  - o Blackbox Feature Selection algorithms. In this approach, we are using algorithms explicitly with the goal of feature selection, using a combination of variables that are significant (per the models) and using them to compute model metrics
  - o Traditional model building and evaluation. Build full models and use traditional variable selection techniques like AIC, VIF and p values to arrive at the final model

### Blackbox Feature Selection algorithms



Below are the steps for Variable Selection Technique

- The models that were used for **variables selection** were
  - o Boruta
  - o Rpart
  - o Relative Importance from regression models
  - o Genetic Algorithm
  - o Simulated Annealing Feature Selection
  - o Recursive feature elimination
- Each of these algorithms mentioned above we select variables using a linear / non-linear relationship.
- The intent here is to select the variables that were significant across each of these models (Count based)
- Another approach could have been selecting the top 3 or 5 from each model.
- The output (significant variables) are then used as final variables in a linear regression model.
- The variables in this list is selected with a score >50% which signifies they were selected by at least 4 out of 7 models.
- The FR\_EPS\_Rs & FR\_CEPS\_Rs has a high multicollinearity, Similarly other variables need to be analysed from a multicollinearity perspective.

Variable Name	#Models	Score(%)	Variable Name	#Models	Score(%)
FR_Book_Value_Rs	7	100%	VR_CE	4	57%
PL_PAT	6	86%	VR_High_PE	4	57%
FR_EPS_Rs	6	86%	FR_CF_Per_share	4	57%
FR_CEPS_Rs	6	86%	FR_FCF_Per_Share	4	57%
FR_Adj_DPS_Rs	6	86%	FR_FCF_Yield	4	57%
FR_Adjusted_Book_Value_Rs	6	86%	FR_Dividend_Pay_Out_Ratio_P	4	57%
VR_Low_PE	5	71%			
FR_DPS_Rs	5	71%			
FR_Total_Debt_Mcap	5	71%			

- The following model performance measures were arrived at by running a linear regression model with the above variables, but excluding the FR\_EPS\_Rs and FR\_CEPS\_Rs

Metric	Value
RMSE	.266
R <sup>2</sup>	.967

### Linear Regression Model:

The approach is to start with a full model, analyze the significant variables, their p values, AIC and VIF scores to then select on the final model.

- Partition the dataset into test and train based on a 70/30 ratio

### Full Model Performance Metrics

Metric	Value
RMSE	1.99
R <sup>2</sup>	.997

- As seen from the metrics the model's explanatory power is only because of inclusion of all variables, yet the RMSE is slightly on the higher side for a relatively high R<sup>2</sup>
- We then move towards selecting the variables based on three parameters
  - o Significant values with p value <.05
  - o VIF <10
  - o Lowest AIC
- First a full model was run in step mode to get a sense of the significant variables that result in the lowest AIC.
- As we can see from the below output, there are a number of variables that may have multicollinearity but the % of significant variables in context has gone up.

#### Coefficients:

```

Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.292e-17 8.769e-04 0.000 1.000000 .
FH_FCF -2.114e-03 1.207e-03 -1.752 0.080174 .
VR_Dividend_Yield 3.606e-03 1.082e-03 3.333 0.000903 ***
VR_Netsales 2.063e-03 1.290e-03 1.599 0.110166
VR_EBITDA 2.758e-03 1.001e-03 2.756 0.006000 **
VR_CE 2.022e-03 1.314e-03 1.539 0.124227
VR_Low_PE 2.469e-03 1.093e-03 2.258 0.024240 *
PL_PAT -7.135e-03 1.197e-03 -5.962 3.86e-09 ***
FR_CF_Per_share -1.948e-02 7.288e-03 -2.672 0.007698 **
FR_FCF_Per_Share 5.453e-03 2.211e-03 2.466 0.013899 *
FR_EPS_Rs 9.633e-01 6.445e-03 149.470 < 2e-16 ***
FR_CEPS_Rs 5.567e-02 1.002e-02 5.554 3.90e-08 ***
FR_DPS_Rs -1.531e-01 4.383e-03 -34.936 < 2e-16 ***
FR_Adj_DPS_Rs 1.524e-01 4.596e-03 33.162 < 2e-16 ***
FR_Adjusted_Book_value_Rs -4.223e-03 2.414e-03 -1.750 0.080600 .
FR_PATM_P 1.152e-01 4.314e-02 2.671 0.007734 **
FR_CPM_P -1.163e-01 4.392e-02 -2.647 0.008287 **
FR_ROA_P -3.571e-03 1.545e-03 -2.312 0.021065 *
FR_Asset_Turnover 3.274e-03 1.269e-03 2.580 0.010072 *
FR_Fixed_Capital_sales 2.992e-03 1.827e-03 1.638 0.101862
FR_Receivable_days -2.319e-03 1.100e-03 -2.108 0.035327 *
FR_Payable_days 3.574e-03 1.145e-03 3.122 0.001865 **
FR_EBIT_Growth_P 2.661e-03 1.066e-03 2.497 0.012747 *
FR_Current_Ratio 3.801e-02 1.041e-02 3.652 0.000279 ***
FR_Quick_Ratio -3.879e-02 1.048e-02 -3.700 0.000232 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02419 on 736 degrees of freedom
Multiple R-squared: 0.9994, Adjusted R-squared: 0.9994
F-statistic: 5.409e+04 on 24 and 736 DF, p-value: < 2.2e-16

```

- The above variables are selected based on the lowest AIC values of -5640

- The next step is to recursively remove all variables with VIF>10

The following variables were selected as part of this logic that had a VIF < 10

FH_FCF	VR_CE	FR_EPS_Rs	FR_ROA_P	FR_Payable_days
VR_Dividend_Yield	VR_Low_PE	FR_DPS_Rs	FR_Asset_Turnover	FR_EBIT_Growth_P
VR_NetSales	PL_PAT	FR_Adjusted_Book_Value_Rs	FR_Fixed_Capital_Sales	FR_Current_Ratio
VR_EBITDA	FR_FCF_Per_Share	FR_PATM_P	FR_Receivable_days	

- After the process to remove variables with VIF > 10, the statistically insignificant variables (with p value >=.05) were removed
- The resulting variables that were selected post the process of selecting statistically significant variables

VR_EBITDA	FR_DPS_Rs
VR_Low_PE	FR_Adjusted_Book_Value_Rs
PL_PAT	FR_Receivable_days
FR_EPS_Rs	FR_Payable_days

- The above model still had an R2 of .998 as the FR\_EPS\_Rs was also selected. This variable needs to be removed from the final model as the Y variable was derived from it.
- Final model was run with the following variables
  - o Earnings before interest, tax, depreciation, amortization
  - o Dividend per share
  - o Adjusted Book Value
  - o Profit After Tax
  - o Price to Earnings Ratio
  - o Receivable Days
  - o Payable days

Model performance measures of the final model

Metric	Value
RMSE	.026
R <sup>2</sup>	0.96

### Random Forest:

A random forest model was also run to validate if non linear relationships exist that may be able to explain the Y variable better.

- In looking at the variable importance plot the random forest model is suggesting the following variables.
- The FR\_DPS\_Rs, FR\_payable\_days are picked up as common variables in comparison to the linear regression model

FR_Adj_DPS_Rs	VR_Dividend_Yield
FR_Book_Value_Rs	FR_Inventory Turnover
FR_DPS_Rs	FR_FCF_Per_Share
FR_Payable_Days	

Model performance measures of the Random Forest Model

Metric	Value
RMSE	0.87
R <sup>2</sup>	0.86

Summary of all models for the manufacturing data:

Model Types	RMSE	R <sup>2</sup>
Full Model	1.99	.99
Linear Regression Model (Variables selected with Blackbox approach)	.26	.96
Model post selecting significant variables	.26	.96
Random Forest	.87	.86

### Observations:

- The linear regression model explains the variation in the target variable better, therefore it's the model of choice here.
- The model with variables selected using the black box approach and via the traditional methods perform on par.

Final Conclusion: Owing to simplicity in process, compute and explainability of the model, the linear regression model with significant variables selected using statistical significance is the final choice.

## 10 Actionable insights and recommendations to the stakeholder

Based on the models created for different Macro Economic Sectors the following Ratios are key to determine Adjusted Earnings Per Share:

- Adjusted Book value
- Free Cash Flow Per Share
- Return on Equity
- Dividend yield

The random forest is the best model for all macro-economic sectors except for manufacturing and Commodities for which, the linear regression model was selected.

Macro-economic sector	Unique Ratios
<b>Consumer</b>	-Return on capital Employed
<b>Financial Services</b>	-Asset Turn over
<b>Commodities</b>	-Total Debt Market Capital
<b>Manufacturing</b>	-Earnings before interest, tax, depreciation, amortization (EBIDTA)

### Next Steps:

- Build a portfolio of stocks based on the maximum adjusted EPS by running clustering techniques
- Build predictive models by macro-economic sectors
- Use significant variables to link industry standard factor definitions viz Quality, Momentum, Value etc

## 11 References and Bibliography

- Fischer Black, Michael C. Jensen, Myron Scholes, The Capital Asset Pricing Model: Some Empirical Tests (1972)
- Stephen A. Ross, The arbitrage theory of capital asset pricing (1976)
- E. F. Fama, K. R. French, The Cross-Section of Expected Stock Returns (1992)
- Wen Long, Zhichen Lu , Lingxiao Cui, Deep learning-based feature engineering for stock price movement prediction (2018)
- Nikola Milosevic, Equity forecast: Predicting long term stock price movement using machine learning (2016)
- Yuxuan Huang, Machine Learning for Stock Prediction Based on Fundamental Analysis (2019)
- Fama French Three Factor Model  
<https://www.investopedia.com/terms/f/famaandfrenchthreefactormodel.asp>

- Size, Value and Momentum in Indian equities  
<https://journals.sagepub.com/doi/pdf/10.1177/0256090917733848>
- Corporate Finance & Accounting – terms  
<https://www.investopedia.com/terms/r/roce.asp>
- An Overview of Factor Investing - Fidelity Investments  
[https://www.fidelity.com/bin-public/060/www\\_fidelity\\_com/documents/brokerage/overview-factor-investing.pdf](https://www.fidelity.com/bin-public/060/www_fidelity_com/documents/brokerage/overview-factor-investing.pdf)
- Foundations of factor investing – MSCI  
<https://www.msci.com/www/research-paper/foundations-of-factor-investing/016381488>
- Correlation and Regression with R - Contributing Authors: Ching-Ti Liu, PhD, Associate Professor, Biostatistics ; Jacqueline Milton, PhD, Clinical Assistant Professor, Biostatistics ; Avery McIntosh, doctoral candidate
- Deep learning-based feature engineering for stock price movement prediction - Wen Long, Zhichen Lu \*; Lingxiao Cui
-

## 12 Appendix

### 1. Data Dictionary –

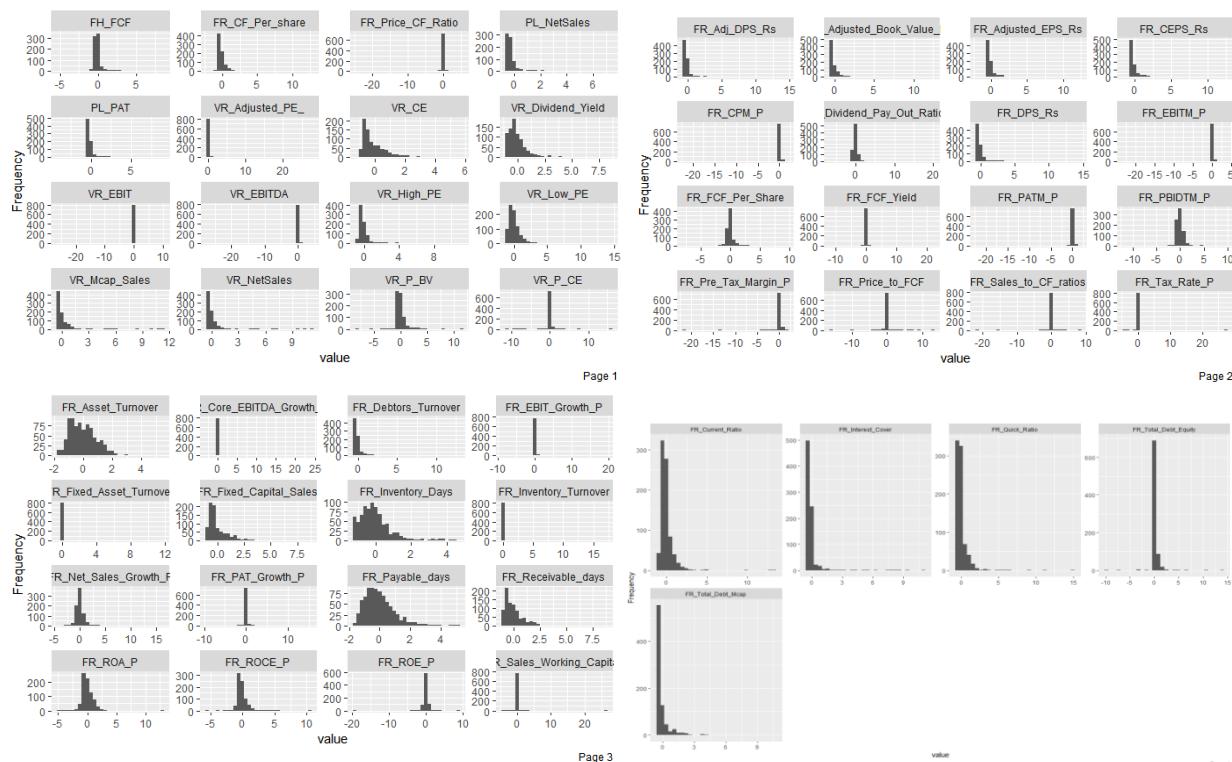
Prefix Legend
FR - Financial Ratio
VR - Valuation Ratio
PL - Profit & Loss detail

Field Name used in Code	Original Field name
Company_Name	Company_Name
Industry	Industry
MacroEconomicSector	MacroEconomicSector
FR_Year_End	FR_Year_End
FH_FCF	FH_Free Cash flow
VR_Adjusted_PE	FR_Adjusted PE (x)
VR_P_CE	FR_PCE(x)
VR_P_BV	FR_Price / Book Value(x)
VR_Dividend_Yield	FR_Dividend Yield(%)
VR_NetSales	FR_EV/Net Sales(x)
VR_EBITDA	FR_EV/EBITDA(x)
VR_EBIT	FR_EV/EBIT(x)
VR_CE	FR_EV/CE(x)
VR_Mcap_Sales	FR_M Cap / Sales
VR_High_PE	FR_High PE
VR_Low_PE	FR_Low PE
PL_NetSales	PL_Net Sales
PL_PAT	PL_Profit After Tax
FR_CF_Per_share	FR_Cash Flow Per share
FR_Price_CF_Ratio	FR_Price to Cash Flow Ratio
FR_FCF_Per_Share	FR_Free Cash Flow per Share
FR_Price_to_FCF	FR_Price to Free Cash Flow
FR_FCF_Yield	FR_Free Cash Flow Yield
FR_Sales_to_CF_ratios	FR_Sales to cash flow ratios
FR_EPS_Rs	FR_Earnings Per Share (Rs)
FR_Adjusted_EPS_Rs	FR_Adjusted EPS (Rs.)
FR_CEPS_Rs	FR_CEPS(Rs)

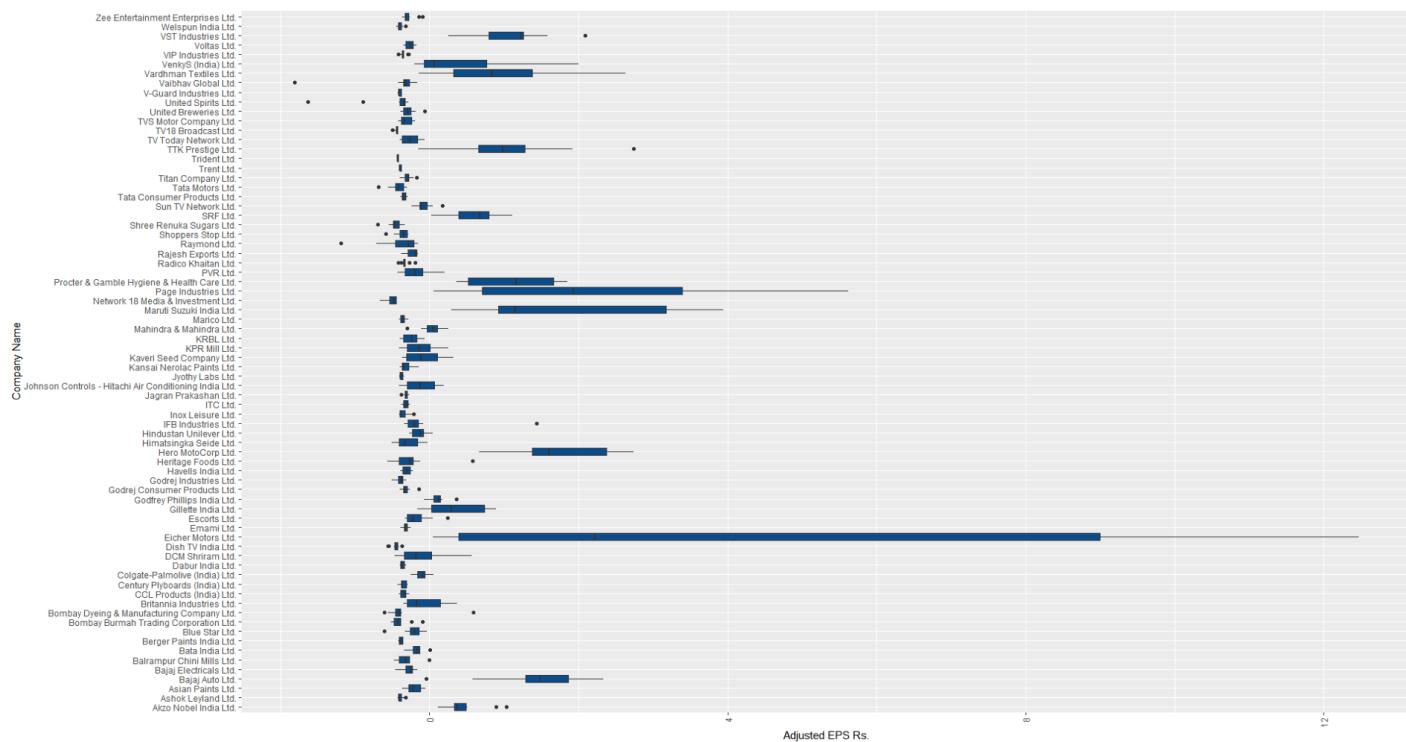
FR_DPS_Rs	FR_DPS(Rs)
FR_Adj_DPS_Rs	FR_Adj DPS(Rs)
FR_Book_Value_Rs	FR_Book Value (Rs)
FR_Adjusted_Book_Value_Rs	FR_Adjusted Book Value (Rs)
FR_Tax_Rate_P	FR_Tax Rate(%)
FR_Dividend_Pay_Out_Ratio_P	FR_Dividend Pay Out Ratio(%)
FR_PBIDTM_P	FR_PBIDTM (%)
FR_EBITM_P	FR_EBITM (%)
FR_Pre_Tax_Margin_P	FR_Pre Tax Margin(%)
FR_PATM_P	FR_PATM (%)
FR_CPM_P	FR_CPM(%)
FR_ROA_P	FR_ROA (%)
FR_ROE_P	FR_ROE (%)
FR_ROCE_P	FR_ROCE (%)
FR_Asset_Turnover	FR_Asset Turnover(x)
FR_Inventory_Turnover	FR_Inventory Turnover(x)
FR_Debtors_Turnover	FR_Debtors Turnover(x)
FR_Fixed_Asset_Turnover	FR_Fixed Asset Turnover (x)
FR_Sales_Working_Capital	FR_Sales(x)/Working Capital
FR_Fixed_Capital_Sales	FR_Fixed Capital/Sales(x)
FR_Receivable_days	FR_Receivable days
FR_Inventory_Days	FR_Inventory Days
FR_Payable_days	FR_Payable days
FR_Net_Sales_Growth_P	FR_Net Sales Growth(%)
FR_Core_EBITDA_Growth_P	FR_Core EBITDA Growth(%)
FR_EBIT_Growth_P	FR_EBIT Growth(%)
FR_PAT_Growth_P	FR_PAT Growth(%)
FR_Adj_EPS_Growth_P	FR_Adj. EPS Growth(%)
FR_Total_Debt_Equity	FR_Total Debt/Equity(x)
FR_Current_Ratio	FR_Current Ratio(x)
FR_Quick_Ratio	FR_Quick Ratio(x)
FR_Interest_Cover	FR_Interest Cover(x)
FR_Total_Debt_Mcap	FR_Total Debt/Mcap(x)

[back to source](#)

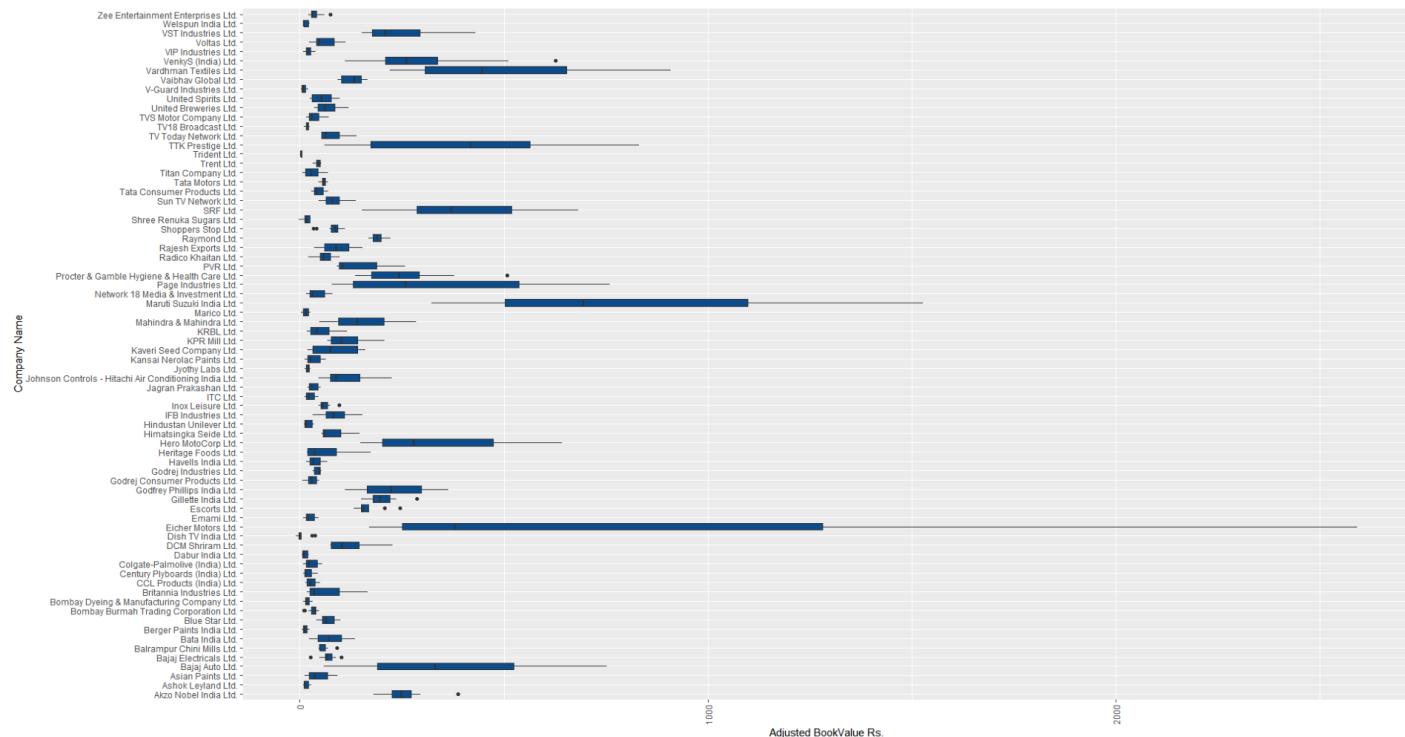
## 2. Consumer Univariate Analysis Graph output



## 3. Adjusted EPS vs Company Name



#### 4. Adjusted Book value Vs Comapny



## 5. Linear regression output (full model)

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.0103278	0.0113136	-0.913	0.361727
FH_FCF	-0.0589703	0.0153690	-3.837	0.000140 ***
VR_Adjusted_PE_	0.0206665	0.0103365	1.999	0.046078 *
VR_P_CE	-0.0166420	0.0154362	-1.078	0.281474
VR_Dividend_Yield	0.0319037	0.0139768	2.283	0.022848 *
VR_EBITDA	-0.0397583	0.0117880	-3.373	0.000799 ***
VR_EBIT	-0.0341955	0.0751944	-0.455	0.649467
VR_CE	0.0806757	0.0167821	4.807	2.00e-06 ***
VR_High_PE	0.0119273	0.0137948	0.865	0.387637
VR_Low_PE	-0.0076593	0.0125961	-0.608	0.543401
PL_NetSales	-0.0411859	0.0160869	-2.560	0.010737 *
FR_CF_Per_share	0.0593293	0.0276137	2.149	0.032123 *
FR_FCF_Per_Share	0.2475814	0.0242266	10.219	< 2e-16 ***
FR_Price_to_FCF	-0.0017199	0.0097101	-0.177	0.859475
FR_FCF_Yield	-0.0152263	0.0105060	-1.449	0.147843
FR_Adjusted_Book_Value_Rs	0.6460033	0.0191431	33.746	< 2e-16 ***
FR_Tax_Rate_P	-0.0029644	0.0094306	-0.314	0.753391
FR_Dividend_Pay_Out_Ratio_P	0.0071710	0.0204525	0.351	0.726013
FR_PBIDTM_P	-0.0035123	0.0176604	-0.199	0.842436
FR_ROA_P	0.1378728	0.0297197	4.639	4.41e-06 ***
FR_ROE_P	0.0768439	0.0371453	2.069	0.039055 *
FR_ROCE_P	-0.0230955	0.0315073	-0.733	0.463869
FR_Asset_Turnover	-0.0080645	0.0179554	-0.449	0.653516
FR_Inventory_Turnover	-0.0187467	0.0152926	-1.226	0.220794
FR_Debtors_Turnover	0.0150931	0.0122052	1.237	0.216776
FR_Fixed_Asset_Turnover	0.0033725	0.0126327	0.267	0.789597
FR_Sales_Working_Capital	-0.0090642	0.0093802	-0.966	0.334330
FR_Fixed_Capital_Sales	0.0103274	0.0149625	0.690	0.490360
FR_Receivable_days	0.0003716	0.0148084	0.025	0.979987
FR_Inventory_Days	-0.0118199	0.0128980	-0.916	0.359866
FR_Payable_days	0.0285506	0.0129966	2.197	0.028469 *
FR_Net_Sales_Growth_P	0.0205822	0.0113305	1.817	0.069854 .
FR_Core_EBITDA_Growth_P	0.0223163	0.0109877	2.031	0.042750 *
FR_EBIT_Growth_P	0.0092277	0.0184529	0.500	0.617234
FR_PAT_Growth_P	-0.0094260	0.0168730	-0.559	0.576641
FR_Total_Debt_Equity	0.0958468	0.0256710	3.734	0.000209 ***
FR_Interest_Cover	0.0184146	0.0114022	1.615	0.106904
FR_Total_Debt_Mcap	-0.0150949	0.0153652	-0.982	0.326349

---  
Signif. codes: 0 \*\*\* 0.001 \*\* 0.01 \* 0.05 . 0.1 ' ' 1

Residual standard error: 0.2617 on 530 degrees of freedom  
Multiple R-squared: 0.9174, Adjusted R-squared: 0.9117  
F-statistic: 159.2 on 37 and 530 DF, p-value: < 2.2e-16

## 6. Consumer - Final Linear Model Summary & variable importance

```

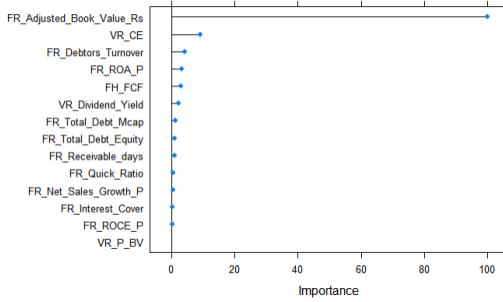
Call:
lm(formula = .outcome ~ ., data = dat)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.63443 -0.12157  0.03341  0.14434  2.90739 

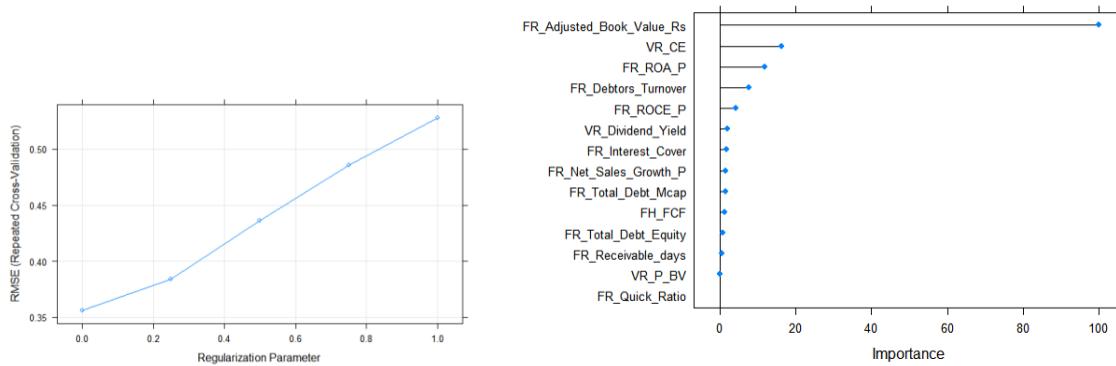
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.0006162  0.0146069 -0.042 0.966367    
FR_Adjusted_Book_Value_Rs 0.8715094  0.0144223  60.428 < 2e-16 ***
FH_FCF      -0.0541085  0.0159802 -3.386 0.000760 ***  
VR_CE       0.1774897  0.0250903  7.074 4.57e-12 ***  
FR_Debtors_Turnover 0.0815427  0.0198402  4.110 4.56e-05 ***  
FR_ROA_P     0.1156565  0.0320176  3.612 0.000331 ***  
FR_Total_Debt_Equity 0.0427257  0.0180480  2.367 0.018260 *  
VR_Dividend_Yield 0.0451005  0.0152401  2.959 0.003215 **  
FR_Net_Sales_Growth_P 0.0415701  0.0202934  1.950 0.051694 .  
FR_Total_Debt_Mcap 0.0417104  0.0184444  2.459 0.013864 *  
FR_Receivable_days 0.0442723  0.0194889  2.273 0.023090 *  
VR_P_BV      -0.0466605  0.0270744 -1.723 0.085372 .  
FR_ROCE_P     0.0606578  0.0333310  1.820 0.069321 .  
FR_Quick_Ratio -0.0312867  0.0155907 -2.007 0.045261 *  
FR_Interest_Cover 0.0299237  0.0162047  1.847 0.065339 .  
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3468 on 553 degrees of freedom
Multiple R-squared:  0.8954, Adjusted R-squared:  0.8928 
F-statistic: 338.3 on 14 and 553 DF, p-value: < 2.2e-16

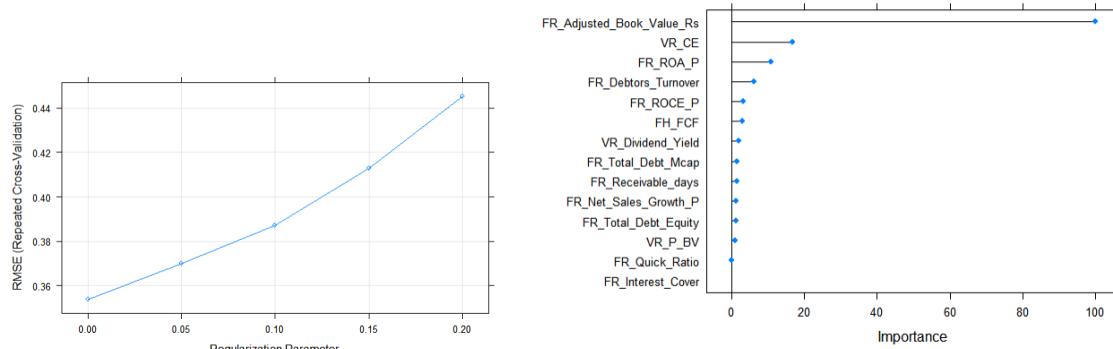
```



## 7. Consumer – Ridge Model & variable importance



## 8. Consumer – Lasso Model & variable importance



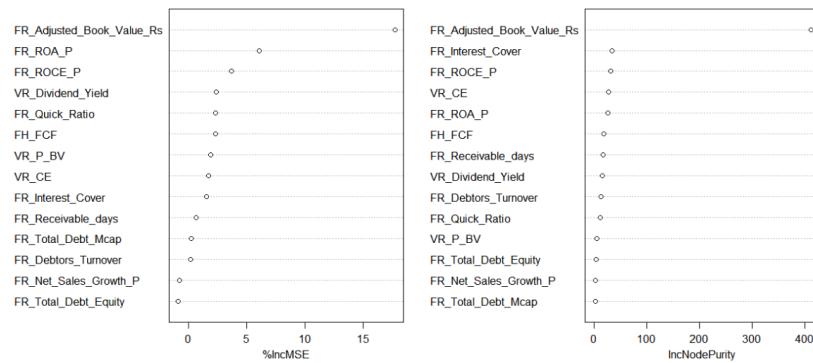
## 9. Consumer – Random Forest Model & variable importance

```

Call:
randomForest(formula = FR_Adjusted_EPS_Rs ~ FR_Adjusted_Book_Value_Rs + FH_FCF + VR_CE +
FR_Debtors_Turnover + FR_ROA_P + FR_Total_Debt_Equity + VR_Dividend_Yield + FR_Net_Sales_Growth_P +
FR_Total_Debt_Mcap + FR_Receivable_days + VR_P_BV + FR_ROCE_P + FR_Quick_Ratio + FR_Interest_Cover,
data = trainData, ntree = 90, mtry = 7, importance = TRUE)
Type of random forest: regression
Number of trees: 90
No. of variables tried at each split: 7

Mean of squared residuals: 0.2076134
% var explained: 81.46

```



## Manufacturing Data

```
> print(summary(lmfullModel1))
```

```
call:
lm(formula = FR_Adjusted_EPS_Rs ~ ., data = lm_train)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.160369	-0.002310	0.000841	0.003081	0.087166

coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1.086e-04	6.261e-04	-0.174	0.862315
FH_FCF	5.490e-04	1.069e-03	0.514	0.607700
VR_Adjusted_PE_	2.860e-04	1.131e-03	0.253	0.800388
VR_P_CE	6.277e-05	5.810e-04	0.108	0.914019
VR_P_BV	2.777e-04	2.077e-03	0.134	0.893722
VR_Dividend_Yield	-4.550e-04	8.419e-04	-0.540	0.589189
VR_Netsales	-3.580e-02	2.256e-02	-1.587	0.113241
VR_EBITDA	2.811e-04	9.002e-04	0.312	0.754964
VR_EBIT	2.864e-05	9.914e-04	0.029	0.976964
VR_CE	-1.401e-04	2.018e-03	-0.069	0.944675
VR_Mcap_Sales	2.953e-02	1.874e-02	1.576	0.115739
VR_High_PE	-5.335e-04	9.552e-04	-0.559	0.576760
VR_Low_PE	6.799e-04	8.269e-04	0.822	0.411373
PL_Netsales	5.235e-05	1.966e-03	0.027	0.978772
PL_PAT	-4.337e-03	2.093e-03	-2.072	0.038760 *
FR_CF_Per_share	-3.432e-03	4.686e-03	-0.732	0.464364
FR_Price_CF_Ratio	2.266e-04	1.160e-03	0.195	0.845123
FR_FCF_Per_Share	-4.619e-05	1.643e-03	-0.028	0.977583
FR_Price_to_FCF	-9.354e-05	6.054e-04	-0.155	0.877268
FR_FCF_Yield	-1.200e-04	6.526e-04	-0.184	0.854246
FR_Sales_to_CF_ratios	1.454e-04	8.735e-04	0.166	0.867852
FR_EPS_Rs	9.866e-01	5.309e-03	185.821	< 2e-16 ***
FR_CEPS_Rs	1.433e-02	6.913e-03	2.072	0.038774 *
FR_DPS_Rs	-5.922e-02	3.795e-03	-15.605	< 2e-16 ***
FR_Adj_DPS_Rs	5.947e-02	3.913e-03	15.199	< 2e-16 ***
FR_Book_value_Rs	-6.702e-01	2.060e-02	-32.540	< 2e-16 ***
FR_Adjusted_Book_Value_Rs	6.722e-01	2.068e-02	32.501	< 2e-16 ***
FR_Tax_Rate_P	3.711e-05	6.595e-04	0.056	0.955148
FR_Dividend_Pay_Out_Ratio_P	2.396e-04	6.300e-04	0.380	0.703841
FR_PBIDTM_P	1.339e-04	6.034e-04	0.222	0.824525
FR_EBITM_P	1.596e-02	1.037e-02	1.538	0.124710
FR_Pre_Tax_Margin_P	-1.339e-02	1.379e-02	-0.971	0.332120
FR_PATM_P	2.749e-02	3.342e-02	0.823	0.411155
FR_CPM_P	-2.372e-02	3.040e-02	-0.780	0.435615
FR_ROA_P	2.163e-03	2.345e-03	0.922	0.356772
FR_ROE_P	9.775e-05	9.514e-04	0.103	0.918213
FR_ROCE_P	-2.691e-03	2.223e-03	-1.211	0.226683
FR_Asset_Turnover	-8.028e-04	1.083e-03	-0.741	0.458960
FR_Inventory_Turnover	-5.482e-04	5.877e-04	-0.933	0.351456
FR_Debtors_Turnover	6.158e-04	7.488e-04	0.822	0.411289

```

FR_Debtors_Turnover      6.158e-04  7.488e-04  0.822  0.411289
FR_Fixed_Asset_Turnover 9.199e-04  9.688e-04  0.950  0.342847
FR_Sales_working_Capital -1.299e-03 5.277e-04 -2.462  0.014162 *
FR_Fixed_Capital_sales   2.041e-04  1.291e-03  0.158  0.874465
FR_Receivable_days       6.917e-04  1.013e-03  0.683  0.495094
FR_Inventory_Days        6.046e-06  9.992e-04  0.006  0.995175
FR_Payable_days          -4.002e-06 1.063e-03 -0.004  0.996997
FR_Net_Sales_Growth_P    4.035e-05  7.083e-04  0.057  0.954592
FR_Core_EBITDA_Growth_P  -3.759e-04 8.220e-04 -0.457  0.647682
FR_EBIT_Growth_P         1.879e-03  9.972e-04  1.885  0.060066 .
FR_PAT_Growth_P          1.006e-02  1.175e-02  0.856  0.392432
FR_Adj_EPS_Growth_P     -1.019e-02 1.172e-02 -0.869  0.385255
FR_Total_Debt_Equity    -5.307e-04 9.875e-04 -0.537  0.591228
FR_Current_Ratio         1.179e-02  9.325e-03  1.265  0.206663
FR_Quick_Ratio           -1.238e-02 9.562e-03 -1.295  0.195913
FR_Interest_Cover        -3.341e-04 8.285e-04 -0.403  0.686991
FR_Total_Debt_Mcap       3.972e-03  1.045e-03  3.800  0.000163 ***
```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0142 on 477 degrees of freedom
Multiple R-squared: 0.9997, Adjusted R-squared: 0.9997
F-statistic: 2.872e+04 on 55 and 477 DF, p-value: < 2.2e-16

```

### Logic to recursively remove VIF>10

```

Recursively remove variables with VIF >10
```
signif_all <- names(all_vifs)
# Remove vars with VIF>10 and re-build model until none of VIFs don't exceed 10.
while(any(all_vifs > 10)){
  var_with_max_vif <- names(which(all_vifs == max(all_vifs))) # get the var with max vif
  signif_all <- signif_all[!(signif_all) %in% var_with_max_vif] # remove
  myForm <- as.formula(paste("FR_Adjusted_EPS_RS ~ ", paste(signif_all, collapse=" + ")), sep=""))
  selectedMod <- lm(myForm, data=manufacturing3) # re-build model with new formula
  all_vifs <- car::vif(selectedMod)
}
print(summary(selectedMod))
print(car::vif(selectedMod))
signif_all
as.data.frame(signif_all)
```

```

### Logic to remove non statistically significant variables

```

Remove non significant variables
```
all_vars <- names(selectedMod[[1]])[-1] # names of all x variables
# Get the non-significant vars
summ <- summary(selectedMod) # model summary
pvals <- summ[[4]][, 4] # get all p values
not_significant <- character() # init variables that aren't statsitically significant
not_significant <- names(which(pvals > 0.05))
not_significant <- not_significant[!not_significant %in% "(Intercept)"] # remove 'intercept'. optional!

# If there are any non-significant variables,
while(length(not_significant) > 0){
  all_vars <- all_vars[!all_vars %in% not_significant[1]]
  myForm <- as.formula(paste("FR_Adjusted_EPS_RS ~| ", paste(all_vars, collapse=" + ")), sep=""))
  selectedMod <- lm(myForm, data=manufacturing3) # re-build model with new formula

  # Get the non-significant vars.
  summ <- summary(selectedMod)
  pvals <- summ[[4]][, 4]
  not_significant <- character()
  not_significant <- names(which(pvals > 0.1))
  not_significant <- not_significant[!not_significant %in% "(Intercept)"]
}
summary(selectedMod)
```

```

```
```{r}
set.seed(101)
rf_model1 = randomForest(FR_Adjusted_EPS_Rs ~ ., data=lm_train, ntree=101, mtry=5, importance=TRUE)
randomForest::varImpPlot(rf_model1)
randomForest::importance(rf_model1, type=1)
```

```

### Random Forest Variable Importance Plot

