Analysis of 2012’s Financial Data

The task was to identify the factors that were most contributing to predicting **Ev/EBITDA** and **M/Sales** and to also build a regression model that could predict these values with a high accuracy.

# Data Preprocessing:

* The csv file’s columns were converted from str to float format
* The NaN values were replaced with the mean of their respective columns.

# Ev/EBITDA Correlation Matrix:

* As can be seen from the heatmap above, the features which were most influential on the target variable were:
  + 1) FR\_Dividend Pay Out Ratio(%)
  + 2) ROE (%)
  + 3) ROCE (%)
  + 4) Asset Turnover (%)
  + 5) Free Float
* The data was scaled using StandardScaler
* It was then split into a training and test set for the ML model

# Results:

Three models were tried

**a) LassoCV**

**b) XGBoost**

**c) RidgeCV**

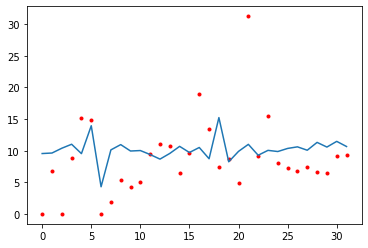
**d) BayesianRidge**

**e) ARDRegression**

**f) ElasticNetCV**

**g) LassoLarsCV**

**h) SVR**

Out for these,the error for BayesianRidge was the least.

# M/Sales Correlation Matrix:

* As can be seen from the heatmap above, the features which were most influential on the target variable were:
  + 1) PBIDTM (%)
  + 2) FR\_Dividend Pay Out Ratio(%)
  + 3) ROCE (%)
  + 4) Cons OCF
  + 5) Cons PAT

# Results:

Three models were tried

**a) LassoCV**

**b) XGBoost**

**c) RidgeCV**

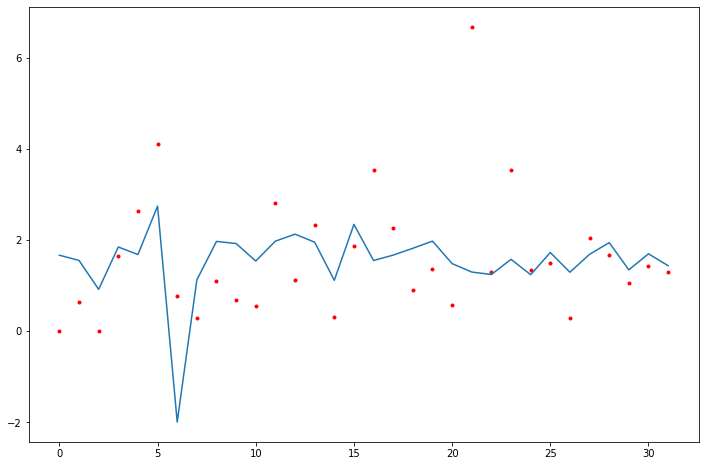
**d) BayesianRidge**

**e) ARDRegression**

**f) ElasticNetCV**

**g) LassoLarsCV**

**h) SVR**

Out for these,the errors for ElasticNetCV and LassoCV was the least.

# Analysis and Observations:

* The data for one year was less, hence the model might overfit the training data
* As the model contained a few outliers, LassoCV performed well as it penalizes outliers for M/Sales
* In general, the models with the given factors worked well for M/Sales while a little inaccurate for Ev/EBITA