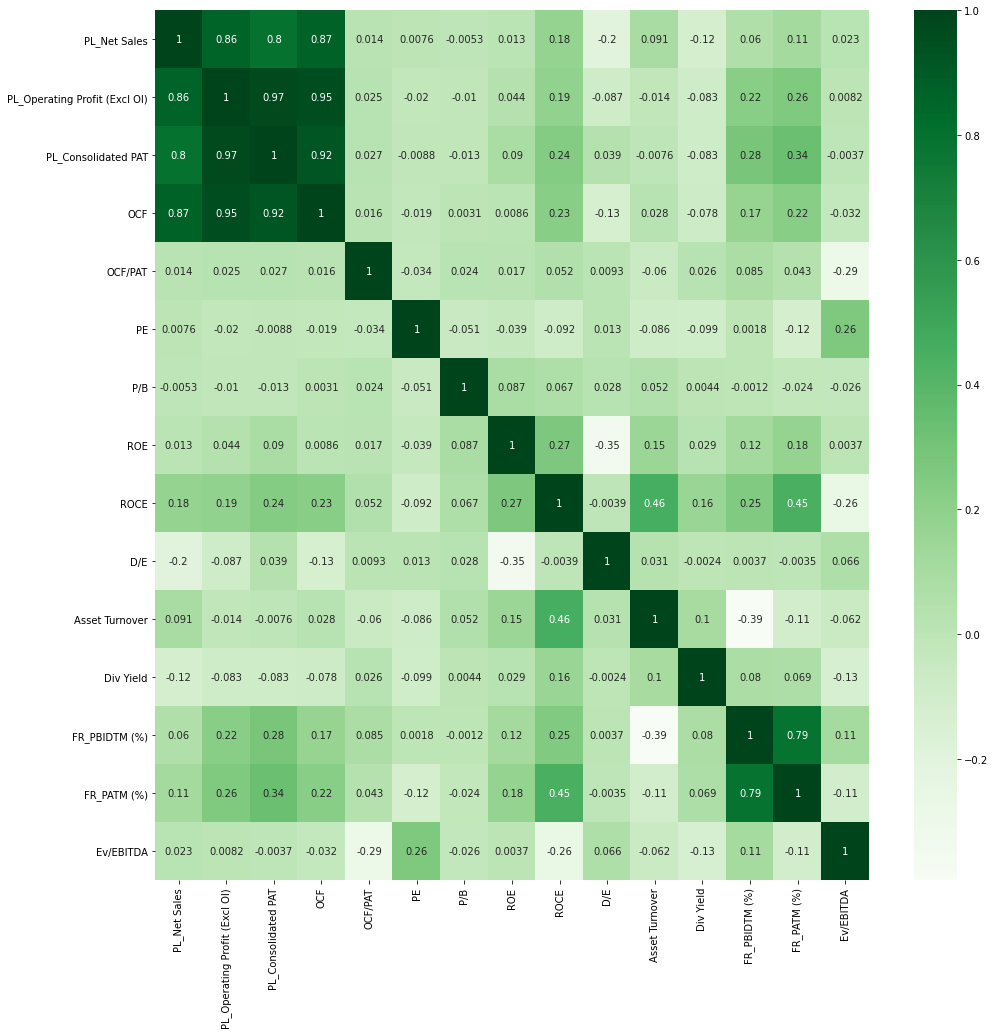
Analysis of 2010’s Financial Data

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) OCF/PAT
  + 2) PE
  + 3) ROCE
  + 4) Div/Yield
  + 5) FR\_PBIDTM (%)
  + 6) FR\_PATM (%)
* The data was scaled using StandardScaler
* It was then split into a training and test set for the ML model

# Results:

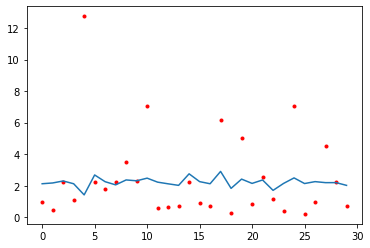
Three models were tried

a) LassoCV

b) XGBoost

c) RidgeCV

Out for these,the error for LassoCV 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) PL\_Consolidated PAT
  + 2) PE
  + 3) Asset Turnover
  + 4) FR\_PBIDTM (%)
  + 5) FR\_PATM (%)

# Results:

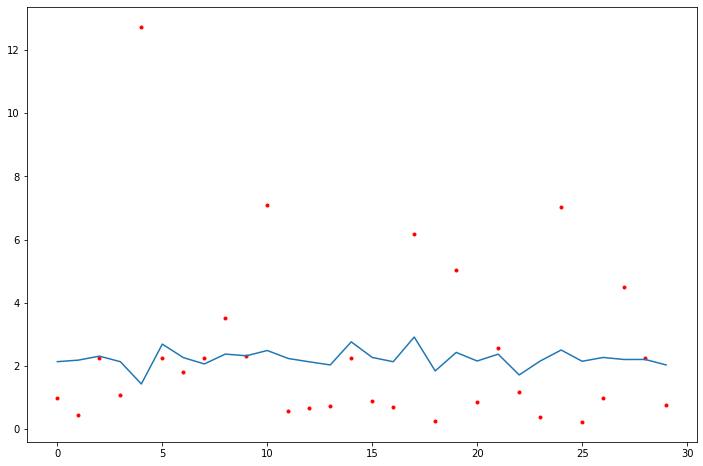
Three models were tried

a) LassoCV

b) XGBoost

c) RidgeCV

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



# Analysis and Observations:

* The data for one year was less, hence the model might overfit the training data
* A lot of values were #N/A and were replaced by the mean. This could cause slight inaccuracy in results.
* As the model contained a few outliers, LassoCV performed well as it penalizes outliers.