**APPROACH**

1. I have targeted the problem as a regression analysis and used Xgboost regressor machine learning algorithm to build my final model.
2. For data pre-processing I checked for missing values, looked for multi collinearity problem, looked for distance magnitude problem that can be caused by columns containing large values and other column containing small values.

Using label Encoder converted Categorical data into numeric form

For feature engineering, Date column looks interesting so I separated month, day, year and weekday from it

X\_train=train1[['Store\_id','Store\_Type','Location\_Type','Region\_Code','Holiday','Discount','weekday']]

y\_train = train['Sales']

1. Before reaching my final model, I have used Random forest and Gradient boosting algorithms, however Xgboost gave a bit better results as compared.

Fine tuning and using extracted weekday column played a crucial role in getting a descent model.

xgb\_grid = {'n\_estimators':[131], 'subsample':[0.73],'max\_depth':[12],'learning\_rate':[.12],"colsample\_bytree":[0.72]}

Local accuracy (using neg\_mean\_squared\_log\_error and cross validation): mean: -0.09945, std: 0.05365