## Objective

Our Project goal was build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads.

## Procedure Followed

After loading dataset we have performed various Exploratory Data Analysis both at univariate and bivariate levels to come up with insights. Some are listed below:

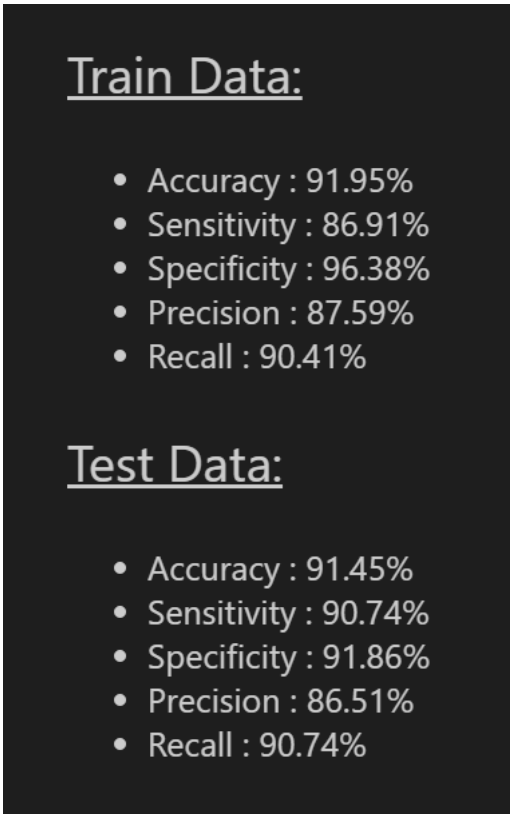
- Among Lead Origin Lead Ad Form show high Conversion Rates
- Google and Reference are high yielding (more conversion rates) Lead Sources
- Conversion Rates High Among people who want Emails
- The current conversion rates are at 38.5 percent
- SMS Sent, Email Opened these Last Activities are high yielding
- The leads are predominantly from India and from in and around Mumbai(this may also be due to data gathering from in and around Mumbai)
- Many of the MBA Specializations have high Conversion Rates
- Working Professions are high yielding
- Most people choosing Course for better career Prospects
- Tags like will revert after reading email, Closed By Horizon are high yielding

Also we followed the appropriate processes of

- Missing value imputation and dropping.
- Outlier Analysis was done and handled appropriately.
- Categorical variables were converted to dummies and named appropriately.
- Scaling was done to ensure all variables are at same numerical scale.

For Modelling we used logistic regression and tried out multiple model. Used

- Recursive Feature elimination to shortlist 20 most important variables
- Then iteratively dropped more variables by seeing the p-values and also the variance inflation factor to finally shortlist 16 variables.
- We prepared 5 logistic regression models
- We sensed importance of the variables from the magnitude of their coefficients
- The top 5 important variables were Tags_Will revert after reading the email,Tags_Closed by Horizzon,Tags_Ringing,Total Time Spent on Website and Tags_Lost to EINS
- Various threshold Metric were tried out to ensure selection of optimal model threshold.
- We got around 97% ROC- Score




- We can see performances are good and are similar for train and test sets and thus no major overfitting.
- Model can thus be used for production.

## Outcomes and Insights

Model was performing quite well and the output list of people ranked by their lead scores can be used to prioritize the people to contact. Also for the features found important we can devise strategies so that conversion becomes even higher.

Like for example:

- More engaging websites to increase time spend on websites and closely monitoring website section where there is more bounce rates and improve them.
- More advertisements in google since high yielding and also referral bonuses to past converted leads so that they recommend more people.
- Working Class people, MBA Professionals or students can be focussed on for more conversion. Even specialized courses can be designed for MBA professionals.
- SMS and Emails can be improved since they are very important factors in ensuring conversion. So proper framing and design can be focussed on.