

SUMMARY OF LEAD SCORE

This analysis is done for X-Education and find way to get more industry lead to join their courses. The basic data provided gave us lot of information about how the potential customers visit the site the time they spend there, how they reached the site and the conversion rate .

Following steps we did in the process :

- **Data cleaning** : in data cleaning we removed major null values and replace minor null values with mode , also did necessary steps to recheck data and prepare it.
- **EDA**: did all necessary steps of EDA which is very important.
- **Dummy Variable** : Dummy Variables are created and later dummies with 'abc' element removed. For numerical values we use minmax scaler.
- **Train-Test Split** : the split was done at 70% and 30% for train and test data.
- **Model building** : firstly RFE was done to attain the top 15 relevant variables. Later the rest of the variables were removed manually depending on the VIF and P value (the variable with VIF <5 and p value <0.05 were kept.)
- **Model Evaluation** : A confusion matrix was made. Later on the optimum cut off value (using ROC curve) was used to find the accuracy , sensitivity, and specify which came to around 81%.
- **Prediction** : Prediction was done on the test data frame and with the optimum cut off as 0.35.
- **Precision Recall** : This method is used to recheck and cut off of 0.41.

From all of above we can say that X-Education has very high chances to get almost all the potential buyers to buy their courses .