

## Summary

This analysis is done for X Education to help find **identify the most potential leads, also known as 'Hot Leads'**. The provided data have lots of features that helps to solve the problem

The following are the steps used:

1. Cleaning data: First we did data cleaning, dropped the unnecessary columns, and filled null values with appropriate values
2. EDA: To check the overall condition of the data we did EDA. And it was found that some columns are irrelevant, so we have dropped it. The numeric variables didn't have many outliers
3. Dummy Variables: Dummy variables were created and used Minmax scaler
4. Train-Test split: The split was done at 70% and 30% for train and test data set
5. Model Building and Model Evaluation: Build a logistic regression model, a confusion and found accuracy, sensitivity and specificity
6. Prediction: Prediction was done on the test data frame with accuracy, sensitivity and specificity of more than 90%

From the analysis, the following are the most important variables:

1. The total time spend on the Website.
2. Total number of visits.
3. When the lead source was: a. Google b. Direct traffic c. Organic search d. reference