# Summary Report for Lead Score Case Study By Rudresh, Ruchi and Sri Nath

As a data scientist, our project for X-Education aimed to optimize the lead conversion process and increase course sales. We developed a logistic regression model with rigorous analysis and consideration of key variables, resulting in a robust solution for identifying potential customers with a higher likelihood of conversion. We followed the following steps to develop the logistic regression model which fits the best to achieve the goal for our problem statement:

## Step 1: Reading and understanding the dataset (Leads.csv):

## Step 2: Data cleaning:

## Step 3: Visualization of the Data:

In this step, we used various graphical representations like percent plot, boxplots and others for the visualization of the data and drew conclusions from these graphs.

- We used boxplots to detect outliers in the dataframe and treated the accordingly.
- We plotted numerical variables and categorical variables against target variables simultaneously to analyze relation between these variables.
- After observing the visualized data, we combined some of the columns into one single category i.e., "Others" and also dropped some columns from the dataframe which are not significant for analysis and will not add any significant information to the model.

## Step 4: Data Preparation:

- We created dummy variables for the categorical variables.
- We removed all the repeated and redundant variables.
- Then, we divided the dataframe into test and train sections with a proportion of 70- 30% values.
- For Feature Scaling, we used the Standard Scaling to scale the all variables and then, we plot the heatmap to check the correlations among the variables.

## Step 5: Model Building:

- We used the Recursive Feature Elimination and went ahead and selected the 15 top important features.
- We used the statistics generated, we recursively tried looking at the P-values in order to select the most significant values that should be present and dropped the insignificant values.
- Finally, we arrived at the 12 most significant variables. The VIF's for these variables were also found to be good.
- For our final model we checked the optimal probability cut off by finding points and checking the accuracy, sensitivity and specificity.

## Step 6: Model Evaluation:

- We then plotted the ROC curve for the features and the curve came out be pretty decent with an area coverage of 96% which further solidified the accuracy of the model.
- Then, we found the optimal cut off point to get balanced sensitivity and specificity. Finally, prepared classification report based on the Accuracy, Confusion Metrics, Sensitivity, Specificity, False Positive Rate, Positive Predictive Value, Negative Predictive Value on final prediction on train set.

## Step 7: Precision and Recall:

- We checked the precision and recall with accuracy, sensitivity and specificity for our final model on train set.
- Next, Based on the Precision and Recall trade-off, we got a cut off value of approximately 0.25.
- Then we implemented the learning to the test model and calculated the conversion probability based on the Sensitivity and Specificity metrics and found out the accuracy value to be 89.68%; Sensitivity= 85.74%; Specificity= 91.92%.

## Step 8: Determining Feature Importance:

- We assigned lead score in this step and used the following formula:
  Lead Score = 100 * Conversion Probability which needed to be calculated for all the leads from the original dataset (train + test).
- Then, for the final step we drew conclusions from the overall analysis and visualizations.

Through the implementation of our model and recommendations, X-Education can expect improved conversion rates and enhanced sales performance. Our data-driven approach and robust model provide a reliable foundation for making informed decisions and optimizing the lead conversion process.

Based on the technical data, strong model performance, and the influence of key variables, we can confidently justify the effectiveness and value of our model in helping X-Education achieve their sales targets.