**Summary**

1. **EDA:**

* We dropped the columns with more than 45% missing values, we could have dropped the columns with more than 30% missing values but we might have lost lot more data on that so we replaced Nan Values with the more repetitive values. .
* We did analysis on Numerical variables, managed outliers and also with Dummy variables.

1. **Train-Test split & Scaling :**

* Out train and test data were 70% and 30% respectively.
* We did min max scaling in the following variables ['Page Views Per Visit', 'TotalVisits', 'Total Time Spent on Website']

1. **Model Building**

* We used REF for feature selection
* REF was then performed to get the top 15 variables
* Then we manually removed the variables depending upon their REF Value and P Value.
* We created a confusion matrix and checked overall accuracy which is 80.91%

1. **Model Evaluation**

* **Sensitivity – Specificity**
* On **Training Data**
* The optimum cut off value was found with the help of ROC curve. The area under ROC curve was 0.88.
* After Plotting the cutoff was **0.35** which gave us the following

Accuracy to be 80.91%

Sensitivity to be 79.94%

Specificity to be 81.50%.

* Prediction on **Test Data**
* We got

Accuracy to be 80.02%

Sensitivity to be 79.23%

Specificity to be 80.50%

* **Precision – Recall:**

When we do precision -Recall On **Training Data**

* With the cutoff of 0.35 we get the Precision & Recall of 79.29% & 70.22% respectively.
* So to increase the above percentage we need to change the cut off value. After plotting we found the optimum cut off value of **0.44** which gave

Accuracy was 81.80%

Precision was 75.71%

Recall was 76.32%

When we do precision -Recall On

Accuracy was 80.57%

Precision was 74.87%

Recall was 73.26%

1. So if we go with Sensitivity-Specificity Evaluation the optimal cut off value would be **0.35**

&

If we go with Precision – Recall Evaluation the optimal cut off value would be **0.44**

**CONCLUSION**

TOP VARIABLE CONTRIBUTING TO CONVERSION:

* LEAD SOURCE:
  + - Total Time Spent on Website
    - Total Visits
  + Lead Origin:
    - Lead Add Form
  + Lead source:
    - Direct traffic
    - Google
    - Welingak website
    - Organic search
    - Referral Sites

Last Activity:

* + Do Not Email\_Yes
  + Last Activity\_Email Bounced
  + Olark chat conversation

The model was good in terms of prediction and we can definitely give a green light in using to improve the business.