Summary

The model building and prediction is being done for company X Education and to find ways to convert potential users. We will further understand and validate the data to reach a conclusion to target the correct group and increase conversion rate. Let us discuss steps followed:

1. **EDA:**
   - Quick check was done on null value and we dropped columns with more than 3000 missing values.
   - Then we saw the Number of Values for India were quite high (nearly 90% of the Data), so this column was dropped.
   - We also worked on numerical variable, outliers and dummy variables.
2. **Visualizations:**
   - Histograms, boxplots, and heat maps were used to visualize data distributions, outliers, and correlations.
   - Feature importance charts were used to illustrate the relative influence of each feature on conversion.

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

   - The split was done at 70% and 30% for train and test data respectively.
   - We will do min-max scaling on the variables ['TotalVisits', 'Page Views Per Visit', 'Total Time Spent on Website']
4. **Model Building**

   - RFE was used for feature selection.
   - Then RFE was done to attain the top 15 relevant variables.
   - Later the rest of the variables were removed manually depending on the VIF values and p-value.
   - A confusion matrix was created, and overall accuracy was checked which came out to be 78.86%.
5. **Model Evaluation**

   - **Sensitivity – Specificity**
     If we go with Sensitivity- Specificity Evaluation. We will get:
     - On **Training Data**

- The optimum cut off value was found using ROC curve. The area under ROC curve was 0.86. After Plotting we found that optimum cutoff was **0.42** which gave
  Accuracy 79.08%
  Sensitivity 79.34%
  Specificity 78.84%.

- Prediction on **Test Data**
  - We get
    Accuracy 78.45%
    Sensitivity 77.95%
    Specificity 78.92%

- **Precision – Recall:**

  If we go with Precision – Recall Evaluation

  - On **Training Data**
    - With the cutoff of 0.42 we get the Precision & Recall of 80.57% & 73.94% 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 78.95%
      Precision 78.40%
      Recall 77.71%

  - Prediction on **Test Data**
    - We get
      Accuracy 78.66%
      Precision 78.28%
      Recall 76.75%

6. So, if we go with Sensitivity-Specificity Evaluation the optimal cut off value would be **0.42**
   &
   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 Visits
  - Total Time Spent on Website
- Lead Origin:
  - Lead Add Form
- Lead source:
  - Direct traffic
  - Google
  - Welingak website
  - Organic search
  - Referral Sites

This lead scoring case study has shown how important it is to score our leads in order to increase conversions. By understanding the needs and wants of our target customer, we can better qualify them as a potential buyer. This will save you time and money in the long run. The Model seems to predict the Conversion Rate very well and we should be able to give the Company confidence in making good calls based on this model.

1. The leads are joined course for Better Career Prospects, most of having Specialization from Finance Management. Leads from HR, Finance & marketing management specializations are high probability to convert.
2. Most of leads current occupation is Unemployed, which means gave more focus on unemployed leads.
3. The high number of total visits & Total time spent on platform may increasing the chances of lead to be converted.
4. Talking to last notable Activity, making improvement in customer engagement through email & calls will help to convert leads. As the leads which are opening email have high probability to convert, Same as Sending SMS will also benefit.