Summary

In order to convert potential consumers, business X Education is constructing models and making predictions. In order to target the right audience and boost conversion rates, we will further analyze and confirm the data. Let's go over the subsequent steps:

1. **EDA:**

- We performed a quick check on the percentage of null values and removed columns that had more than 35% missing values.
- Additionally, we observed that the rows with null values would cost us a significant amount of data and were crucial columns. Therefore, we changed the NaN values to read "not provided."
- We substituted all missing values with India because that was the non-missing data' most frequent occurrence.
- After discovering that India accounted for a substantial number of values (almost 97% of the data), this column was removed.
- Additionally, we dealt with dummy variables, outliers, and numerical variables.

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

- For train and test data, the split was done at 70% and 30%, respectively.
- On the variables ("TotalVisits," "Page Views Per Visit," and "Total Time Spent on Website"),we will do min-max scaling.

3. **Model Building**

- RFE was used for feature selection.
- The top 15 pertinent variables were then obtained by RFE.
- Later, based on the VIF values and p-value, the remaining variables were manually deleted.
- The correctness of the confusion matrix, which was constructed, was tested, and it was found to be 80.91%.

4. **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. Thearea under ROC curve was 0.89.
- After Plotting we found that optimum cutoff was **0.38** which gave

  Accuracy 80.88%
  Sensitivity 78.74%
  Specificity 82.23%.

- Prediction on Test Data

- We get

  Accuracy 81.63%
  Sensitivity 79.36%
  Specificity 82.91%

- **Precision — Recall:**

  If we go with Precision — Recall Evaluation

- On **Training Data**

- With the cutoff of 0.38 we get the Precision & Recall of79.77% & 69.86% respectively.
- So to increase the above percentage we need to change the cut off value. After plotting we found the optimum cut offvalue of **0.41** which gave

Accuracy 81.13%
Precision 74.92%
Recall 76.99%

- Prediction on Test DataoWe get

Accuracy 81.74%
Precision 73.44%
Recall 77.11%

5. So if we go with Sensitivity-Specificity Evaluation the optimal cut offvalue would be **0.38**
&
If we go with Precision — Recall Evaluation the optimal cut off valuewould be **0.41**

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

Last Activity:

- Do Not Email_Yes
- Last Activity_EmaiI Bounced
- Olark chat conversation

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.