

## Summary

This analysis focuses on X Education's efforts to attract more industry professionals to enroll in their courses. The provided data offered insights into potential customers' website visits, the duration of their stay, their entry sources, and the conversion rates. The following steps were undertaken in the analysis:

1. **Data Cleaning:**

The dataset required some cleaning, addressing a few null values. Options with minimal information were replaced with null values to preserve data integrity. Some null entries were marked as 'not provided' to avoid data loss. Given the diverse geographic distribution of the leads, entries were categorized as 'India', 'Outside India', and 'not provided'.

2. **Exploratory Data Analysis (EDA):**

A preliminary EDA was conducted to assess the dataset's condition. It revealed that many categorical variable elements were irrelevant, while numeric values appeared sound, with no outliers detected.

3. **Dummy Variables Creation:**

Dummy variables were generated, and those associated with 'not provided' were subsequently removed. Numeric values were scaled using the MinMaxScaler.

4. **Train-Test Split:**

The dataset was divided into training (70%) and testing (30%) subsets.

5. **Model Building:**

Recursive Feature Elimination (RFE) was performed to identify the top 15 relevant variables. Additional variables were excluded based on Variance Inflation Factor (VIF) and p-value criteria, retaining those with  $VIF < 5$  and  $p\text{-value} < 0.05$ .

6. **Model Evaluation:**

A confusion matrix was created, and the optimal cutoff value was determined using the ROC curve, achieving accuracy, sensitivity, and specificity around 80%.

7. **Prediction:**

Predictions were made on the test dataset using an optimum cutoff value of 0.35, maintaining the accuracy, sensitivity, and specificity at 80%.

8. **Precision-Recall Analysis:**

This method was employed for further validation, revealing a cutoff of 0.41, with precision at approximately 73% and recall at 75%.

The analysis identified the most significant factors influencing potential buyers, ranked in descending order:

1. Total time spent on the website.
2. Total number of visits.
3. Lead Source:
  - Google
  - Direct traffic
  - Organic search
  - Welingak website
4. Last activity:
  - SMS
  - Olark chat conversation
5. Lead origin in 'Lead add' format.
6. Current occupation as a working professional.

With these insights, X Education is well-positioned to effectively convert potential buyers into course enrollments.