

# SUMMARY

This analysis was conducted for X Education to explore strategies for increasing enrollment from industry professionals in their courses. The initial dataset provided insights into how prospective customers interacted with the website, including their browsing patterns, the amount of time spent on the site, their methods of accessing the site, and the overall conversion rate. The following steps were undertaken in the analysis:

1. **Data Cleaning:** The dataset was mostly clean, with only a few missing values. The "Select" option was replaced with a null value as it lacked informative value. Some null values were replaced with 'not provided' to retain as much data as possible, although these were later excluded when creating dummy variables. Given the predominance of users from India, the location data was categorized into 'India,' 'Outside India,' and 'Not Provided.'
2. **Exploratory Data Analysis (EDA):** A brief EDA was performed to assess the state of the data. It revealed that many categorical variable elements were irrelevant, while the numeric data appeared well-distributed with no significant outliers.
3. **Dummy Variables:** Dummy variables were generated, and those associated with 'Not Provided' were removed. For numerical variables, the MinMaxScaler was applied.
4. **Train-Test Split:** The data was divided into 70% for training and 30% for testing.
5. **Model Building:** Recursive Feature Elimination (RFE) was used to identify the top 15 relevant variables. Variables were manually removed based on Variance Inflation Factor (VIF) values and p-values, retaining those with  $VIF < 5$  and  $p\text{-value} < 0.05$ .
6. **Model Evaluation:** A confusion matrix was constructed. The optimal cutoff value was determined using the ROC curve, resulting in accuracy, sensitivity, and specificity scores of around 80%.
7. **Prediction:** Predictions were made on the test dataset with an optimal cutoff of 0.35, achieving approximately 80% accuracy, sensitivity, and specificity.

8. **Precision-Recall Analysis:** This method was used for validation, revealing an optimal cutoff of 0.41, with precision around 73% and recall around 75% on the test dataset.

Key factors influencing potential buyers, in order of importance, were identified as:

1. Total time spent on the website
2. Total number of visits
3. Lead source, particularly:
  - Google
  - Direct traffic
  - Organic search
  - Welingak website
4. Last activity, particularly:
  - SMS
  - Olark chat conversation
5. Lead origin, particularly when it is a Lead Ad Form
6. Current occupation, specifically as a working professional

With these insights, X Education can significantly increase its chances of converting potential buyers into course enrollees.