

# Summary

**Objective:** The analysis aims to help X Education attract more industry professionals to their courses by leveraging insights from the provided data. The data includes information on how potential customers interact with the site, their duration of stay, access routes, and conversion rates.

## **Analysis Steps:**

### **1. Data Cleaning:**

- The dataset was mostly clean, with minor null values. The option "select" was replaced with null values due to its lack of information.
- Some null values were changed to "not provided" to retain data, although these were later removed when creating dummy variables.
- Location data was categorized into "India", "Outside India", and "not provided".

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

- A preliminary EDA revealed some irrelevant elements in categorical variables, while numeric values were clean with no outliers.

### **3. Creating Dummy Variables:**

- Dummy variables were generated, and those with "not provided" elements were removed. Numeric values were normalized using the MinMaxScaler.

### **4. Train-Test Split:**

- The data was split into 70% training and 30% testing sets.

### **5. Model Building:**

- Recursive Feature Elimination (RFE) was used to select the top 15 relevant variables.
- Variables with Variance Inflation Factor (VIF) less than 5 and p-value less than 0.05 were retained.

### **6. Model Evaluation:**

- A confusion matrix was created, and the optimal cut-off value was determined using the ROC curve. The model achieved around 80% accuracy, sensitivity, and specificity.

### **7. Prediction:**

- Predictions on the test set used an optimal cut-off of 0.35, maintaining 80% accuracy, sensitivity, and specificity.

### **8. Precision-Recall Analysis:**

- A cut-off of 0.41 was identified, yielding approximately 73% precision and 75% recall on the test set.

### **Key Influencing Factors for Potential Buyers:**

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

By focusing on these factors, X Education can significantly increase the likelihood of converting potential buyers into course enrollees, enhancing their chances of success.