

# Summary

The objective of this analysis is to support X Education in boosting enrolment among industry professionals for their courses. The initial dataset provides valuable insights into visitor behaviour, including website interactions, session durations, referral sources, and conversion rates.

**The approach involved the following stages:**

## **1. Data Cleaning:**

- The dataset was relatively clean, with only a few missing values.
- The "select" option was replaced with null values due to its limited information value.
- Some missing records were removed.
- The "Country" column was dropped to address data imbalance.
- The "City" column was also dropped as it contained mostly "select" options, which did not provide meaningful information.

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

- A brief EDA was conducted to gain insights into the data.
- It was observed that many elements in categorical variables were irrelevant.
- Numeric values appeared to be suitable, with no outliers identified.

## **3. Dummy Variables:**

- Dummy variables were created for categorical variables.
- MinMaxScaler was applied to numeric values.

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

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

## **5. Model Building:**

- Feature selection was performed using Recursive Feature Elimination (RFE), resulting in the top 20 relevant variables.
- The remaining variables were manually removed based on VIF values and p-values.
- Variables with  $VIF < 5$  and  $p\text{-value} < 0.05$  were retained.

## **6. Model Assessment:**

- A confusion matrix was generated to evaluate the model's performance.
- The ROC curve was used to determine the optimal threshold value, resulting in approximately 80% accuracy, sensitivity, and specificity.

## **7. Making Predictions:**

- Predictions were made on the test dataset using an optimal threshold of 0.43, yielding an accuracy, sensitivity, and specificity of nearly 80%.

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

- Precision-recall analysis was employed to further verify the model's performance.
- Using a threshold of 0.43, the model achieved a precision of approximately 78.8% and a recall of around 77.6% on the test dataset.

### **To enhance conversion rates, the following strategies are recommended:**

- Focus efforts on leads from "Welingak Website" and "Reference" sources, known for their higher conversion potential.
- Prioritize working professional leads, as they demonstrate a stronger inclination to convert.
- Engage leads who spend more time on the website, as their active involvement suggests a higher likelihood of conversion.
- Pay special attention to "Olark Chat" leads, as they have a higher chance of conversion.
- Utilize "SMS Sent" activity by reaching out to leads with this status for better conversion prospects.
- Consider avoiding leads with "Olark Chat Conversation" activity, which may have lower conversion rates.
- Exercise caution with "Landing Page Submission" leads due to their historically lower conversion rates.
- Selectively approach leads with "Others" specialization, considering their lower conversion probability.
- Minimize contact with "Do Not Email" leads, as they have a lower likelihood of conversion.