Summary

This analysis was conducted for X Education to identify strategies to attract more industry professionals to take admissions in their courses. The initial data provided offered insights into potential customers' website behaviour, including visit frequency, time spent on the site, how they arrived at the site, and their conversion rates.

Steps in the Analysis

Data Cleaning

The data was mostly clean, with few null values. The placeholder option "select" was present in many columns. Dropped columns which have higher count of "select" values because they were not helpful during analysis.

Exploratory Data Analysis (EDA)

A quick EDA was conducted to assess the data quality. Many categorical variable elements were found to be irrelevant. The numeric data was in good condition, with no significant outliers detected.

Dummy Variables

Dummy variables were created for categorical data. For numeric data, the MinMaxScaler was applied to normalize the values.

Train-Test Split

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

Model Building

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

Model Evaluation

A confusion matrix was generated, and the optimum cut-off value was determined using the ROC curve. The model achieved approximately 78% accuracy, sensitivity, and specificity.

Prediction

Predictions were made on the test dataset using an optimum cut-off of 0.35, maintaining around 80% accuracy, sensitivity, and specificity.

Precision-Recall Analysis

A precision-recall analysis was conducted, leading to a cut-off of 0.41, with precision at 73% and recall at 75% on the test dataset.

Key Findings

The most influential factors for potential buyers, in descending order, were:

1. Total time spent on the website.

2. Total number of visits.

3. Lead sources such as:

- Google
- Direct traffic
- Organic search
- Welingak website

4. Last activities including:

- 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 has a strong opportunity to convert a high percentage of potential buyers into enrolled students, enhancing their overall success.