Summary

During the analysis, Landing page submissions show the highest conversion rates among lead origins, while Google is the most effective lead source. Notably, some conversions occurred even among those who opted for "Do Not Email," suggesting areas for further investigation regarding communication preferences.

SMS communications are identified as highly effective, leading to the most conversions. This suggests SMS might be a critical touchpoint in the conversion process. Additionally, a significant proportion of converted leads are unemployed, indicating potential targeting strategies that resonate with this demographic.

Engagement metrics such as email and SMS opens strongly correlate with conversion likelihood, emphasizing the importance of crafting compelling messages that encourage engagement.

Numeric variables reveal that page views per visit show an equal distribution between conversions and non-conversions, suggesting this metric alone may not be a strong predictor. However, higher conversion rates are observed among visitors with more total visits and time spent on the website, indicating that more engaged visitors are more likely to convert.

To refine the model, categorical variables were converted into dummy variables, and 20 features were selected using Recursive Feature Elimination (RFE). Further refinement was done using Variance Inflation Factor (VIF) and P-values to eliminate non-contributing features.

The model's performance was evaluated using the ROC Curve and Precision-Recall Tradeoff to determine the optimal cut-off value, balancing accuracy, sensitivity, specificity, precision, and recall. The final model achieves a conversion rate of 81%, surpassing the target set by X Education's CEO. This indicates the model is well-calibrated and effective in identifying and converting leads, aligning with the company's goals, and demonstrating potential for continued optimization.

The training set accuracy of the model stands at 81%, with a sensitivity of 81.08% and a specificity of 80.91%. Precision is recorded at 80% and recall at 70%. These metrics were evaluated using both Sensitivity-Specificity and Precision-Recall metrics, with the optimal cutoff selected based on Sensitivity and Specificity for final predictions. The test set shows accuracy, sensitivity, and specificity values of approximately 81.5%, 80.5%, and 82%, respectively. Precision is recorded at 74%, and recall at 80.54%, closely matching the training set. The lead scoring on the training data indicates a conversion rate of around 80.57% for the final model.

In summary, the analysis highlights key factors influencing conversion rates, including lead origin, lead source, communication preferences, and engagement metrics. The refined model, built using a combination of feature selection and performance evaluation techniques, achieves a high conversion rate, meeting the company's objectives. This comprehensive approach to understanding and optimizing conversion rates provides valuable insights and actionable strategies for enhancing lead conversion efforts.