Summary

The task was to increase X Education's lead conversion rate from 30% to 80% by building a predictive model that identifies "hot leads" with higher conversion potential, allowing the sales team to focus on them and improve efficiency.

Approach:

Problem Understanding: X Education generates leads from various sources, such as website visits, form submissions, and referrals. Despite a large number of leads, only 30% convert into paying customers. The goal was to build a model that predicts which leads are most likely to convert, optimizing sales efforts.

Data Preparation:

- **Data Sources**: The data included demographic information, browsing behaviour, referral sources, and sales interactions.
- **Preprocessing**: Missing values and outliers were handled, and categorical variables (like lead source) were encoded into dummy variables for analysis.

Exploratory Data Analysis (EDA): EDA revealed important patterns, showing that features such as "Total Time Spent on Website," "Lead Source," and "Page Views per Visit" were significant contributors to lead conversion. These insights guided the feature selection for the model.

Model Building:

- **Logistic Regression**: This model was chosen to predict the probability of lead conversion, assigning scores to leads based on their likelihood to convert.
- **Evaluation**: The model's performance was assessed using accuracy, precision, recall, and AUC scores, confirming its ability to distinguish high-potential leads.

Lead Scoring: The model assigned scores to leads based on conversion probability. Leads with higher scores were prioritized for follow-ups, while lower-scoring leads were nurtured with automated emails and marketing content, ensuring efficient use of sales resources.

Implementation: Leads with higher scores were followed up directly by the sales team to boost conversion rates, while lower-scoring leads received automated content. This structured approach allowed the team to focus on promising leads, improving conversion efficiency.

Learnings:

1. **Feature Importance**: Variables such as "Total Time Spent on Website," "Lead Source," and "Last Activity" were found to be the most influential predictors of conversion, and focusing on these enhanced the model's accuracy.
2. **Handling Data Imbalance**: Since the data was imbalanced (few leads converted), oversampling techniques were applied to ensure the model didn't favour non-converting leads.
3. **Business Impact**: The model streamlined the sales process by focusing on high-potential leads, enabling the sales team to approach the 80% conversion target. This resulted in better resource allocation and potential revenue growth for X Education.

In conclusion, the predictive model provided a data-driven method to optimize X Education's lead conversion process, boosting efficiency and guiding the company towards improved conversion rates and revenue growth.