Summary of Lead Scoring Model for X Education

The goal of this project was to develop a logistic regression model that assigns a lead score between 0 and 100, enabling X Education to prioritize leads based on their likelihood to convert. A higher score means a higher chance of conversion, while a lower score indicates a reduced likelihood. The company targeted a conversion rate of around 80%, with key features contributing to conversion probability, such as **Lead Origin (Lead Add Form)**, **Occupation (Working Professional)**, and **Total Time Spent on Website**.

Process Followed:

Data Acquisition & Preparation:
The first step involved sourcing and cleaning the data for consistency. This included handling missing values, filtering outliers, and creating dummy variables for categorical fields. A test-train split was performed to evaluate the model reliably.

Feature Selection & Model Building:
Recursive Feature Elimination (RFE) was used to eliminate variables with high p-values and to reduce multicollinearity by evaluating the variance inflation factors (VIFs) of feature variables. The logistic regression model was trained on the cleaned dataset, and performance was assessed using metrics like accuracy, sensitivity, specificity, and the ROC curve. The optimal cutoff was determined based on the sensitivity-specificity trade-off.

Model Evaluation:
The model's performance yielded an accuracy of 77%, with sensitivity at 83% and specificity at 74%. These results indicate the model is effective at identifying high-conversion leads, achieving a conversion rate near 80%. Sensitivity ensures potential leads are not missed, while specificity minimizes false positives.

Key Insights & Inferences

1. **Lead Origin & Occupation:** Leads from the "Lead Add Form" and working professionals had a higher likelihood of conversion and should be prioritized by the sales team.
2. **Website Engagement:** Leads spending more time on the website were more likely to convert. Strategies like retargeting and real-time engagement tools could further improve conversion rates.
3. **Marketing Source Efficiency:** Organic referrals and past customer recommendations generated higher-converting leads than paid ads, suggesting a need to reallocate the marketing budget.

Recommendations

1. **Prioritize High-Scoring Leads:** Focus on leads from the Lead Add Form and working professionals. Personalized communication and prompt follow-ups should be prioritized.
2. **Refine Lead Scoring & Cutoff Selection:** Fine-tune the cutoff threshold and adjust the precision-recall trade-off to better meet the target conversion rate of 80%.
3. **Improve Engagement Strategies:** High-engagement leads should be prioritized for immediate follow-up. Personalized communication and chatbots can guide high-intent visitors toward conversion.
4. **Optimize Marketing Spend:** Reallocate the marketing budget to high-converting channels and reduce spending on low-quality leads. A/B testing landing pages and CTAs can further improve engagement.
5. **Data & Operational Improvements:** Retraining the model with updated data and addressing data quality issues, such as "Select" values in categorical fields, is essential for accurate predictions.

Learning Outcomes

This project emphasized the importance of data quality and feature selection in creating an effective lead-scoring model. Key learnings include focusing on high-conversion features, utilizing real-time engagement tools, and continuously refining the model for optimal performance. Following a structured methodology, the team successfully built a model that exceeds the target conversion rate and provides actionable insights for X Education’s sales and marketing teams.