Lead Score Case Study Summary

Objective:
The case study focuses on improving X Education's lead conversion rate from 30% to 80% by developing a model to assign lead scores between 0 and 100. These scores help prioritize high-potential leads, enabling the sales team to optimize efforts and resources.

Solution Approach:

1. **Data Preparation:**
    - Removed irrelevant columns like Lead Number and Prospect ID.
    - Addressed missing values and eliminated columns with over 35% missing data.
    - Handled outliers in features such as Total Visits and Page Views Per Visit.
2. **Exploratory Data Analysis (EDA):**
    - Identified an imbalance in the dataset with a higher percentage of non-converted leads.
    - Key insights showed that website interaction and lead origin significantly influenced conversions.
3. **Feature Engineering:**
    - Dummy variables were created for categorical features.
    - Scaled numerical variables using MinMaxScaler to ensure uniformity.
4. **Model Building and Evaluation:**
    - Logistic regression was selected as the predictive technique.
    - Recursive Feature Elimination (RFE) identified the top 15 features, refined to 13 key predictors through statistical significance and multicollinearity checks.
    - Evaluation metrics, including accuracy (81%), sensitivity (76%), and specificity (84%), demonstrated strong model performance.
5. **ROC Curve Analysis:**
    - The Area Under the Curve (AUC) was 0.89, confirming the model's discriminative ability**.**

Key Predictors of Conversion:

1. **Positive Influencers:**

- Time spent on the website: Strongest predictor, with higher engagement indicating higher likelihood of conversion.
- Lead origin from Lead Add Form and occupation of working professionals showed high conversion probabilities**.**
- Phone conversations were a significant factor for conversions.

2. **Negative Influencers:**
    - "Do Not Email" preference and engagement via Olark Chat reduced the likelihood of conversion.
3. **Secondary Predictors:**
    - Moderate impact was noted for total website visits and lead sources such as Olark Chat.

Final Results:

- The model successfully classified leads using a probability threshold of 0.40, achieving an accuracy of 81%.
- It provides actionable insights for targeted strategies, aligning with the business goal of increasing the conversion rate.

Recommendations:

1. **Sales Team Focus:**
    - Prioritize leads with high website interaction and from specific origins like Lead Add Form.
    - Tailor strategies for working professionals and follow up persistently with unreachable leads.
2. **Marketing Optimization:**
    - Enhance the effectiveness of lead sources such as Olark Chat and Welingak Website.
    - Address negative interactions with alternate communication channels like phone or SMS.
3. **Improved Engagement:**
    - Reduce reliance on less effective channels like Olark Chat for critical leads.
    - Enrich website content to encourage prolonged user interaction.
4. **Data-Driven Lead Prioritization:**

- Leverage the model-generated lead scores to focus efforts on "Hot Leads" and allocate resources effectively.

By employing these strategies, X Education is poised to achieve its ambitious target conversion rate of 80%, ensuring sustained business growth and optimized resource utilization.