Lead Score Case Study Summary

Objective:
The case study aims to enhance X Education's lead conversion rate from 30% to 80% by developing a model that assigns lead scores between 0 and 100. These scores allow the sales team to focus their efforts on high-potential leads, optimizing resource allocation and improving efficiency.

Solution Approach:

1. **Data Preparation:**
    - Addressed missing values and eliminated features with over 35% missing data.
    - Removed irrelevant columns such as Lead Number and Prospect ID.
    - Removed columns having only one unique value or having very skewed data
    - Treated outliers in features like Total Visits and Page Views Per Visit.

2. **Exploratory Data Analysis (EDA):**
    - Observed an imbalance in the dataset, with a majority of leads not converting.
    - Identified significant relationships between website engagement, lead origin, and conversion rates.

3. **Data Pre-processing:**
    - Created dummy variables for categorical features.
    - Split dataset into train(70%) and test(30%) datasets
    - Scaled numerical variables using MinMaxScaler to ensure consistent scaling across features.

4. **Model Building and Evaluation:**
    - Logistic regression was chosen for its interpretability and effectiveness in binary classification.
    - Recursive Feature Elimination (RFE) was used to identify the top 15 key predictors.

- After that by considering statistical significance and avoiding multicollinearity we left with 13 features.
- The model achieved strong performance metrics at cut-off probability of 0.5 , including:
    - Accuracy: 81%
    - Sensitivity: 69%
    - Specificity: 88%

5. **ROC Curve Analysis:**
    - The Area Under the Curve (AUC) was 0.88, confirming the model's strong ability to distinguish between converted and non-converted leads.

6. **Sensitivity-Specificity Curve:**
    - By analysing Sensitivity-Specificity curve we get optimal cut-off probability as **0.332** , by using this cut-off we get
        - Accuracy: 80.27%
        - Sensitivity: 80.33%
        - Specificity: 80.23%

7. **Precision-Recall Trade off:**
    - By analysing Precision-Recall curve we get optimal cut-off probability as **0.413** , by using this cut-off we get
        - Accuracy: 81.39%
        - Sensitivity: 75.59%
        - Specificity: 84.96%

    - Using a precision-recall cut-off of 0.413, the True Positive Rate (Sensitivity/Recall) has dropped to about 75%, while the business objective requires this metric to be closer to 80%.

    - We have successfully attained the desired metrics of 80% for both Sensitivity and Specificity by implementing a cut-off threshold of 0.332. Consequently, we shall continue with the sensitivity-specificity analysis to determine the optimal cut-off for making final predictions.

8. **Making Predictions on Test dataset:**
    - Predicted values on test dataset using cut-off probability of 0.332 we get

- Accuracy: 81.24%
- Sensitivity: 80.64%
- Specificity: 81.63%

Key Predictors of Conversion:

1. **Positive Influencers:**
    - Time Spent on Website: The strongest predictor, with higher engagement significantly increasing conversion likelihood.
    - Lead Origin - Lead Add Form and Occupation - Working Professionals: Strongly associated with conversions.
    - Last Notable Activity - Phone Conversations: Demonstrated a significant positive impact on lead conversion.

2. **Negative Influencers:**
    - Do Not Email: Preference reduced the likelihood of conversion.
    - Engagement via Olark Chat: Negatively correlated with conversions.
3. **Secondary Predictors:**
    - Moderate impact from Total Visits and Lead Sources like Olark Chat and Welingak Website.

Recommendations:

1. **Sales Team Focus:**
    - Prioritize leads with high engagement metrics and specific origins, such as Lead Add Form.
    - Tailor outreach strategies for working professionals and persistently follow up with unreachable leads.
2. **Marketing Optimization:**
    - Enhance the impact of lead sources like Olark Chat and Welingak Website through targeted campaigns.
    - Use alternate communication channels, such as phone calls or SMS, for leads preferring not to receive emails.

3. **Improved Engagement:**
    - Focus on enriching website content to increase engagement time.
    - Reduce reliance on less effective channels like Olark Chat for high-priority leads.