Summary Report: Lead Scoring Model for X Education

Objective:

X Education aimed to improve its lead-to-sale conversion rate by identifying 'hot leads'—users most likely to convert into paying customers. The goal was to improve conversion efficiency from 30% toward 80% by enabling the sales team to focus on high-potential leads.

Approach:

A dataset of ~9,240 leads was provided, containing information such as time spent on the website, lead source, and last activity. The objective was to build a logistic regression model to score leads based on their likelihood to convert (lead score from 0 to 100).

Data Cleaning:

Columns with placeholder values like 'Select' were treated as missing and handled accordingly. Redundant or low-variance columns such as 'Magazine' and 'Newspaper Article' were dropped. Relevant categorical columns were retained or imputed where possible.

Exploratory Data Analysis (EDA):

Univariate, bivariate, and multivariate analyses were conducted. For example, higher conversion rates were observed among users spending more than 500 seconds on the website. Heatmaps showed low multicollinearity, supporting clean modeling.

Feature Engineering:

Categorical variables were encoded using dummy variables. Recursive Feature Elimination (RFE) identified key features, including 'Total Time Spent on Website' and 'TotalVisits', which had a strong correlation with conversion.

Modeling:

A logistic regression model was trained and tested. The cutoff probability of 0.44 was selected based on a precision-recall tradeoff. The model assigns a probability score to each lead, aiding decision-making.

Performance Metrics:

- Accuracy: 65.59%
- Precision: 65.96%
- Recall: 59.05%
- ROC-AUC: 86%

Insights & Recommendations:

• Focus on leads spending more time on the website.
• Leads with too many visits but low engagement might reflect indecision.
• Apply dynamic cutoff: use 0.3–0.4 for aggressive campaigns, and 0.6–0.7 for conservative filtering.

Conclusion:

The lead scoring model enables X Education to effectively prioritize high-value leads, reducing effort and increasing conversions. The model is scalable and can adapt to changing sales strategies with threshold tuning.