Lead Scoring Model Presentation

Analysis of Lead Conversion
Prediction using Logistic Regression

Problem Statement

- X Education receives a lot of leads daily, but their conversion rate is only 30%.
- The company aims to increase the conversion rate to 80% by focusing on hot leads.
- The goal is to build a model that identifies which leads are most likely to convert.

Data Preprocessing

- 1. Handled missing values by imputing or dropping columns.
- 2. Dropped irrelevant features with high missing values (> 40%).
- 3. Encoded categorical variables using one-hot encoding.
- 4. Scaled the data using StandardScaler for logistic regression.

Model Building

- 1. Logistic regression was chosen for its interpretability and lead scoring capability.
- 2. The dataset was split into training (80%) and testing (20%) sets.
- 3. A logistic regression model was built with max_iter=2000 and solver='lbfgs'.
- 4. The model was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC.

Model Evaluation

- Model Performance Metrics:
- - Accuracy: 93.77%
- - Precision: 93.35%
- - Recall: 90.96%
- - F1-Score: 92.14%
- - ROC-AUC Score: 97.76%
- The model performs well in distinguishing between hot and cold leads.

Key Insights and Recommendations

- 1. 'Total Time Spent on Website' is a critical indicator of lead conversion.
- 2. Leads sourced from Google and Direct Traffic should be prioritized.
- 3. Focus on engaging leads with high email activity (e.g., 'Email Opened').
- 4. Use the lead scores from the model to prioritize follow-ups and reduce wasted effort.

Conclusion

- The logistic regression model effectively identifies high-potential leads, enabling X Education to focus its efforts on the most likely conversions. The model improves efficiency, leading to a higher conversion rate and better resource allocation.