Lead Scoring Case Study Summary

1. Problem Statement X Education, an online course provider, faces a low lead conversion rate of around 30%. The company seeks to improve efficiency by identifying high-potential leads (Hot Leads) to focus sales efforts and achieve a target conversion rate of 80%.

2. Data Understanding & Cleaning

- The dataset contains 9,000+ leads with features like Lead Source, Total Time Spent on Website, Last Activity, etc.
- Missing values in categorical features were imputed with mode, and numerical features with median.
- Categorical variables were encoded using dummy variables.
- Features with excessive missing values (>50%) were removed.

3. Model Development

- Logistic Regression was selected for classification.
- The dataset was split into training (80%) and testing (20%).
- Features were standardized using StandardScaler.

4. Model Performance

- Accuracy: ~80%
- ROC AUC Score: ~0.85
- Confusion Matrix: Shows a good balance between true positives and true negatives.
- ROC Curve: Indicates strong model performance in distinguishing lead conversion likelihood.

5. Business Insights & Recommendations

- Leads with high engagement (e.g., higher time spent on the website) show a higher likelihood of conversion.
- The most influential features include Lead Source, Last Activity, and Total Time Spent on Website.
- The model assigns a Lead Score between 0-100, helping the sales team prioritize high-scoring leads.
- By focusing on leads with scores above a threshold (e.g., 70), X Education can enhance conversion rates while optimizing resources.

6. Next Steps

- Further refine the model with feature engineering.

- Deploy the model and integrate it with the sales pipeline.
- Monitor model performance and retrain periodically for better results.

This approach ensures a data-driven lead prioritization strategy, improving efficiency and business outcomes.