Lead Scoring Analysis - Summary Report

1. Problem Statement

X Education faces challenges with a **low lead conversion rate** (~30%). The objective is to build a data-driven model to **assign lead scores**, identifying potential leads ("hot leads") to improve the conversion rate closer to 80%.

---

2. Data Understanding and Preparation

- **Dataset**: 9000+ leads with features like:
    - Lead Source
    - Total Time Spent on Website
    - Last Activity
    - Specialization
- **Target Variable**: Converted (1 = Converted, 0 = Not Converted)

Data Preparation Steps:

1. Dropped columns with more than 50% missing data.
2. Imputed missing values:
    - **Numerical**: Median
    - **Categorical**: Mode
3. Standardized numerical features (e.g., TotalVisits).
4. Created dummy variables for categorical features.
5. Removed duplicates.

---

3. Exploratory Data Analysis (EDA)

- **Key Insights**:
    - **Total Time Spent on Website** positively correlates with conversions.
    - Significant categorical contributors include **Lead Source** and **Last Activity**.
    - Imbalance in Converted (only ~30% leads converted).

---

4. Model Building and Evaluation

- **Model**: Logistic Regression

- **Train-Test Split**: 70% training, 30% testing

- **Performance Metrics**:

  - **Accuracy**: 90.98%

  - **Precision, Recall, and F1-Score**: High performance for both converted and non-converted leads.

  - **ROC-AUC Score**: 91%

Evaluation Tools:

- Confusion Matrix

- ROC Curve

---

5. Results and Recommendations

- **Lead Scoring**: Lead scores (0–100) were generated based on the model's probabilities.

- **Results**:

  - Leads with **scores > 80** are "hot leads" with high conversion potential.

  - Output saved in Lead_Scores.csv.

Recommendations:

1. Focus on leads with **Lead Scores > 80**.

2. Develop tailored strategies for **warm leads** (scores between 60-80).

3. Automate lead prioritization using model output.

4. Continuously retrain the model with new data for improved performance.

---

6. Conclusion

The lead scoring model enables X Education to:

- Optimize resource allocation.

- Prioritize leads most likely to convert.

- Increase overall conversion rates by focusing on "hot leads."