Case Study – Lead Scoring Summary

Analysis Approach :

Data Import and Inspection: The dataset was imported, and an initial inspection was performed to identify missing values and inconsistencies in both categorical and numerical variables.

1. **Exploratory Data Analysis (EDA) and Data Preparation**:
    - The data was cleaned by handling missing values and creating dummy variables for categorical features.
    - Non-informative columns like "Lead Profile" and "How did you hear about X Education" were dropped.
    - Visualizations revealed that leads with more landing page submissions and those sourced through Google or direct traffic had a higher chance of converting.
2. **Model Building**:
    - Recursive Feature Elimination (RFE), Rsquared, VIF, and p-values were used to select important features such as **Total Visits**, **Total Time Spent on Website**, and **Lead Source**.
    - Categorical variables like **Lead Origin** and **Last Activity** were also critical for predicting lead conversions.
3. **Model Evaluation**:
    - The model's ROC curve had an AUC of **0.86**, indicating good performance. Precision and recall were both **77%**, and overall accuracy was **78%**, reflecting a well-balanced model that minimized false positives.
4. **Prediction on Test Data**: The model was validated using a test dataset to ensure its ability to generalize well to new, unseen data.

Key Learnings

- Data Cleaning was essential for improving model accuracy by removing irrelevant columns.
- Feature Selection through RFE helped focus on impactful variables, making the model more efficient.
- Model Optimization by adjusting the cut-off threshold between 0.42 and 0.44 improved performance metrics.
- Business Insights indicated that leads from Google, direct traffic, and those with more landing page submissions were more likely to convert, offering actionable strategies for the company.

This structured analysis and model-building process helped create a reliable lead-scoring model that can enhance the company’s marketing efficiency and improve lead conversion.