Summary

The lead scoring case study uses **Logistic Regression** to enhance lead conversion rates by assigning each lead a score from 0 to 100. Here's a summary of the key steps and insights:

Data Understanding and Cleaning:

- The data initially contained 9240 rows and 37 columns.
- Missing data was handled by converting "Select" values to `NaN`, removing columns with more than 35% missing data, and dropping unnecessary columns.
- After cleaning, the dataset was reduced to 9074 rows and 20 columns.

Exploratory Data Analysis:

- **Univariate Analysis**: Visualized categorical variables using count plots and numerical variables with histograms.
- **Bivariate Analysis**: Assessed the relationship between categorical variables and the target variable (converted) using count plots.

Data Preparation for Modeling:

- Created dummy variables for categorical features and dropped redundant ones.
- The data was split into a 70-30% training and test set.
- Applied MinMax scaling to numerical variables and checked correlations using a heatmap.

Building the Logistic Regression Model:

- Feature selection was done using **Recursive Feature Elimination** (RFE).
- **Variance Inflation Factor (VIF)** was used to detect multicollinearity, and variables with high VIF (>5) or high p-values (>0.05) were eliminated iteratively.

Model Evaluation:

- Predictions were made on the training set, and a confusion matrix was generated to calculate accuracy, sensitivity, and specificity.
- The **optimal cut-off** was determined using both the **ROC curve** and **Precision-Recall tradeoff**. Since the focus was on maximizing true positives, the optimal cut-off from the Precision-Recall tradeoff (0.44) was selected.

Key Insights:

1. **Lead Origin**: Most leads are generated from Landing Page Submissions, while the **Lead Add Form** has the highest conversion rate.
2. **Lead Source**: Google and Direct Traffic generate the most leads.
3. **Demographics**: Most leads are from India, and many are looking to improve career prospects.

4. **Occupation**: The majority of leads are unemployed, but **Working Professionals** have the highest conversion rate.
5. **Activity**: Leads with "SMS sent" or "Email opened" as their last activity show high conversion rates.
6. **Lead Source**: Sources like "Reference" and "Welingak Website" lead to higher conversion rates.
7. **Contact Preferences**: Leads opting to receive calls or emails convert at a higher rate, though few leads are generated through recommendations.

These insights provide actionable directions for focusing marketing and sales efforts to boost lead conversion.