SUMMARY

Lead scoring using logistic regression is a methodology used to predict the probability of a lead converting into a customer. By leveraging historical data and various lead features, a logistic regression model can classify leads as either converted or not converted. Here is a summary of the lead scoring logistic regression approach:

1. Problem: The goal is to build a logistic regression model that accurately predicts the probability of a lead converting into a customer.

2. Dataset: The dataset consists of lead data, including demographic information, website interactions, marketing campaign data, and more.

3. Methodology:

- Exploratory Data Analysis (EDA): Analyze the dataset to understand variable distributions, identify missing values or outliers, and gain insights into feature-target relationships.

- Data Preprocessing: Handle missing values, encode categorical variables, scale numerical features, and split the dataset into training and testing sets.

- Feature Engineering: Create new features if needed, such as interaction terms or dummy variables for categorical features.

- Model Development: Train a logistic regression model using the training dataset and optimize hyperparameters if necessary.

- Model Evaluation: Evaluate the trained model using the testing dataset. Calculate metrics like accuracy, precision, recall, and F1-score. Visualize the results using a confusion matrix and ROC curve.

4. Dependencies: The project requires Python 3.x and libraries like Pandas, NumPy, Scikit-learn, Matplotlib, and Seaborn.