Lead Scoring Case Study Summary

Here is a summary of the steps that should be taken to develop a lead scoring model in the correct order:

Step 1: Reading and understanding data

Review the provided historic data and understand the variables and their significance

Step 2: Data cleaning and preparation

Clean the data by handling missing values, duplicates, and outliers

Transform the data as needed, such as scaling or encoding categorical variables

Step 3: Exploratory data analysis (EDA)

Explore the data to identify trends, patterns, and outliers

Step 4: Data analysis

Use statistical methods to analyze the relationships between variables and the conversion rate

Step 5: Creating dummy variables

Convert categorical variables into binary variables to include them in the model

Step 6: Test-train split

Split the data into training and testing sets to train and evaluate the model

Step 7: Feature rescaling

Scale the features to ensure that they have the same magnitude and distribution

Step 8: Feature selection using Recursive Feature Elimination (RFE)

Select the most relevant features that impact the conversion rate

Step 9: Plotting the Receiver Operating Characteristic (ROC) curve

Evaluate the model's performance by plotting the ROC curve and calculating the area under the curve (AUC)

Step 10: Finding the optimal cutoff point

Determine the cutoff point that maximizes the model's accuracy and precision

Step 11: Computing precision and recall matrics

Calculate the precision and recall metrics to evaluate the model's performance

Step 12: Making predictions on the test set

Use the trained model to predict the conversion probability for the test set and assign lead scores based on the probability.

By following these steps, a lead scoring model can be developed and used to identify fruitful customers with higher probabilities of conversion.