Summary

X Education Lead Scoring and Conversion Rate Improvement

Objective

To assign a score to each lead, improving the lead conversion rate from the current 30% to the CEO's target of 80%.

Data Cleaning

1. Column Management:

- ❖ Dropped columns with more than 40% missing values.
- ❖ Assessed categorical columns for imputation, creating new categories when necessary, or dropping if imputation would skew data.
- ❖ Imputed numerical categorical data with mode; dropped columns with a single unique value.
- ❖ Addressed outliers, invalid data, and grouped low-frequency values.
- ❖ Mapped binary categorical values appropriately.

Exploratory Data Analysis (EDA)

1. Data Imbalance:

- ❖ Only 38.5% of leads were converted.

2. Univariate and Bivariate Analysis:

- ❖ Analyzed impact of variables like 'Lead Origin,' 'Current Occupation,' and 'Lead Source' on lead conversion.
- ❖ Noted positive impact of duration of visit on the website on lead conversion.

Data Preparation

1. Feature Engineering:

❖ One-hot encoded categorical variables.

2. Data Splitting:

❖ Split data into training and test sets in a 70:30 ratio.

3. Feature Scaling:

❖ Min Max Scalling.

4. Feature Selection:

❖ Used Recursive Feature Elimination (RFE) to reduce variables from 48 to 15.
❖ Manually reduced features, retaining those with p-values < 0.05 and ensuring no multicollinearity (VIF < 5).

Model Building

1. Initial Models:

❖ Built three models before stabilizing on the final model.

2. Final Model (logm4):

❖ Selected 12 variables with stable performance and p-values < 0.05.
❖ Evaluated using confusion matrix and chose a cutoff point of 0.42 based on accuracy, sensitivity, and specificity plots.

Model Evaluation

1. Performance Metrics:

❖ Achieved approximately 80% accuracy, specificity, and precision.
❖ Precision-recall view showed lower performance (~75%).

2. Lead Score Assignment:

❖ Assigned lead scores to training data using a 0.42 cut-off.

Predictions on Test Data

1. Scaling and Predictions:

❖ Applied final model to test data.

2. Evaluation Metrics:

❖ Achieved consistent performance metrics (~80%) for both train and test sets.

3. Lead Score Assignment:

❖ Assigned lead scores based on model predictions.

Top Features

1. Lead Source:

❖ Welingak Website
❖ Reference

2. Current Occupation:

❖ Working Professional

Recommendations

1. Marketing and Budget Allocation:

❖ Increase budget for advertising on the Welingak Website.

2. Incentives for Referrals:

❖ Offer discounts or incentives for references that convert to leads to encourage more referrals.

3. Targeting Working Professionals:

❖ Focus marketing efforts on working professionals who have higher conversion rates and potentially better financial capacity to afford higher fees.