Summary: Optimizing Lead Conversion for X Education

Introduction:

X Education, while generating numerous leads, faces a challenge with a low lead conversion rate of approximately 30%. The company aims to improve this rate to around 80% and has enlisted our help in building a model to assign lead scores. These scores will prioritize leads with higher conversion potential.

Exploratory Data Analysis (EDA):

- Addressed data imbalance, with only 38.5% of leads converting.
- Conducted univariate and bivariate analysis for categorical and numerical variables.

Data Cleaning:

- Dropped features with over 40% null values.
- Transformed the 'lead quality' variable with 51% nulls into a new category.
- Eliminated columns with 100% nulls and irrelevant flag columns.
- Dropped 'City' and 'Country' due to high nulls and skewed data.
- Removed the 'Tags' column due to sentence data and a significant number of nulls.
- Handled numerical outliers using Interquartile Range (IQR).

Data Preparation:

- Created dummy features for categorical variables.
- Split the dataset into training and test sets (70:30 ratio).
- Applied MinMax scaling for feature standardization.

Model Building:

- Initially built a model with all features, identified high collinearity and p-values.
- Used Recursive Feature Elimination (RFE) to reduce variables from 75 to 15, mitigating collinearity.
- Employed a manual feature reduction process through iterative model building, resulting in a final model with 13 features.
- Determined the optimal cutoff of 0.39 based on the tradeoff between sensitivity and specificity.

Model Evaluation:

- Achieved an overall accuracy of 85% for both training and test sets.

- Balanced sensitivity and specificity, with values exceeding 80% for optimal model performance.

- Selected the cutoff based on the sensitivity-specificity view to align with the company's priorities.

Feature Importance:

- Identified key features influencing predictions, including 'LeadQuality_Worst,' 'Total Time Spent on Website,' and 'LeadSource_Welingak Website.'

- Recognized the significance of working professionals in the conversion process.

Recommendations:

- Allocate additional budget to Welingak Website advertising for increased visibility.

- Emphasize Olark Chat for enhanced engagement.

- Encourage employee input for lead quality assessment.

- Target working professionals due to their higher conversion rates and financial capability.

In summary, the developed logistic regression model provides actionable insights and recommendations for X Education to strategically enhance lead conversion rates. The recommendations span advertising strategies, employee involvement, and targeted audience focus, aiming to align with the company's overarching goal of optimizing lead conversion.