<u>Summary</u>

<u>Problem Statement</u>

X Education faces a significant challenge with its low lead conversion rate, hovering around 30%. The company's CEO has set an ambitious target of achieving an 80% conversion rate. To address this issue, our task was to develop a predictive model that assigns lead scores to prioritize leads with higher conversion potential.

<u>Data Cleaning</u>:

The initial phase of our project involved thorough data cleaning to ensure the quality of our dataset. Columns with over 40% missing values were dropped, and categorical columns underwent careful scrutiny.

Imputation strategies were employed, including dropping columns, creating new categories, or imputing the most frequent values. Additionally, numerical categorical data were imputed with the mode, and columns with only one unique response were removed.

Outliers were treated, invalid data fixed, and low-frequency values were grouped or mapped to enhance data quality.

<u>EDA:</u>

Exploratory data analysis (EDA) revealed a considerable class imbalance, with only 38.5% of leads converting. Univariate and Bi-variate analyses were conducted on both categorical and numerical variables, providing valuable insights into factors influencing lead conversion.

<u>Data Preparation:</u>

Data preparation involved creating dummy features for categorical variables, splitting the dataset into training and testing sets, and standardizing features. Highly correlated columns were dropped to prevent multicollinearity issues.

<u>Model Building:</u>

Model building commenced with Recursive Feature Elimination (RFE) to reduce the number of variables from 50 to 13 for improved manageability.

A manual feature reduction process was then employed, dropping variables with p-values greater than 0.05. Three models were iteratively built and evaluated, culminating in the selection of the third model (logm3) as the final model. This model exhibited stability with p-values below 0.05 and no signs of multicollinearity (VIF < 5).

<u>Model Evaluation and Prediction:</u>

Model evaluation was conducted using a confusion matrix, with a cutoff point of 0.358 selected based on accuracy, sensitivity, and specificity considerations. This cutoff point yielded balanced performance metrics, with accuracy, specificity, and precision all hovering around 80%. Given the business objective of boosting conversion rates to 80%, the model seemed a good fit.

Predictions were made on the test dataset using the final model, with evaluation metrics closely mirroring those of the training set.

Conclusion:

In conclusion, actionable recommendations were proposed based on the model insights. These included allocating more budget to advertising on the Welingak Website, targeting working professionals more aggressively due to their higher conversion rates and potentially better financial capacity and especially those Who spent good amount of time on the Website.