**SUMMARY**

Initially, the primary challenge lay in enhancing the lead conversion rate of X Education, which stood at approximately 30%. The company aimed to develop a model for assigning lead scores, ensuring that leads with higher scores had a greater likelihood of conversion, aligning with the CEO's goal of achieving an 80% conversion rate.

The data cleansing phase entailed removing columns with over 40% null values, employing suitable methods for imputing categorical data, addressing outliers and invalid entries, consolidating low-frequency values, and encoding binary categorical variables. Exploratory data analysis (EDA) involved assessing data imbalance, conducting univariate and bivariate analyses for categorical and numerical variables, and pinpointing variables significantly influencing the target variable.

Data preparation encompassed generating dummy features for categorical variables, partitioning the dataset into training and testing sets, standardizing feature scales, and eliminating highly correlated columns.

Model construction involved feature reduction using recursive feature elimination (RFE) and manual methods. Three models were developed before arriving at the stable final model, characterized by p-values < 0.05 and VIF < 5, containing 12 variables denoted by logm4. This final model was employed to predict outcomes on both the training and testing datasets.

Evaluation of the model involved constructing a confusion matrix and determining a cut-off point of 0.345, derived from accuracy, sensitivity, and specificity considerations. Lead scores were assigned to the training data based on this cut-off, with the top three contributing features identified as Lead Source Welingak Website, Lead Source\_Reference, and Current\_occupation\_Working Professional.

Recommendations arising from the analysis included increasing budget allocation for Welingak Website advertising, offering incentives or discounts for successful references leading to conversions, and targeting working professionals aggressively due to their high conversion rates and favorable financial circumstances.

In summary, the project provided hands-on exposure to data cleansing, exploratory data analysis, data preparation, model construction, and evaluation. It underscored the importance of selecting suitable evaluation metrics aligned with the business problem and understanding the trade-offs between different metrics. Moreover, the analysis shed light on factors influencing lead conversion rates and proposed strategies for enhancing conversions.