Summary:

In this project, we partnered with X Education, a company facing challenges in achieving a satisfactory lead conversion rate. With a current conversion rate of around 30%, our goal was to develop a predictive model that assigns lead scores to optimize the conversion process. The company's CEO set an ambitious target of achieving an 80% lead conversion rate.

Our project journey can be summarized into key phases:

1. **Data Cleaning:**
   We began by addressing missing values, dropping columns with excessive nulls, and imputing others using appropriate strategies. Categorical columns were treated based on their value distributions, and numerical categorical data were imputed with mode. Outliers, invalid data, and low-frequency values were handled to ensure the quality of the dataset.

2. **Exploratory Data Analysis (EDA):**
   An in-depth analysis of the dataset was conducted to gain insights into the relationship between various attributes and the target variable. We identified patterns, correlations, and potential indicators of lead conversion. EDA revealed significant variables such as 'Lead Origin', 'Current Occupation', and 'Lead Source' that impact the likelihood of lead conversion.

3. **Data Preparation:**
   To prepare the data for modeling, we created dummy variables for categorical features using one-hot encoding. The dataset was split into training and testing sets in a 70:30 ratio. Feature scaling was applied using standardization to ensure that all variables were on the same scale. Correlation analysis led to the removal of some highly correlated columns, simplifying the dataset.

4. **Model Building and Selection:**
   We employed Recursive Feature Elimination (RFE) and manual feature reduction to build several models. The aim was to achieve a parsimonious yet effective model. After rigorous testing and variable elimination, Model 7 emerged as the most stable and promising. Its 17 variables demonstrated statistical significance with p-values below 0.05 and showed no signs of multicollinearity with Variance Inflation Factor (VIF) values under 5.

5. **Model Evaluation and Selection of Optimal Cut-off:**
   The model's performance was evaluated using confusion matrices and various cut-off points. The optimal cut-off of 0.354 was chosen based on a trade-off between accuracy, sensitivity, and specificity. This cut-off aligned with the business objective of achieving an 80% conversion rate. By focusing on sensitivity-specificity metrics, we ensured that our predictions aligned with the CEO's goal.

6. **Making Predictions and Lead Scoring:**
   Our final model, log_model7, was used to predict lead conversion on both the training and test datasets. The model demonstrated consistent performance, with accuracy

metrics close to 80% on both sets. Lead scores were assigned based on the optimal cut-off to identify high-potential leads for aggressive follow-up.

7. **Recommendations:**
   Based on the insights gained from our model, we provide strategic recommendations for improving lead conversion rates:
   - Allocate additional budget to promote the Welingak Website, leveraging its effectiveness in attracting quality leads.
   - Introduce incentives or discounts for leads who provide references that convert, encouraging lead referrals.
   - Implement targeted marketing campaigns for working professionals, capitalizing on their high conversion potential due to their financial capacity and interest.

In conclusion, our project successfully addressed X Education's challenge of optimizing lead conversion rates. By leveraging data-driven insights, strategic modeling, and careful evaluation, we have provided actionable recommendations to assist the company in achieving its ambitious lead conversion target.