LEAD SCORING CASE STUDY SUMMARY

As part of the Lead Scoring case study, we have been provided with information on how X Education pursues customer leads from various sources and attempts to convert them into potential customers. Currently, the conversion rate is relatively low at 30%. Our task is to analyze the data and develop a model that can predict lead conversions with an accuracy of up to 80%.

The whole solution is divided into the following sections:

1. **Data Understanding and Exploration**
2. **Data Visualization**
3. **Data Preparation**
4. **Model Building and Evaluation**

Data Understanding and Exploration
To achieve our goal, we began with a comprehensive analysis of the provided dataset. Initially, we identified the relevant columns based on the Data Dictionary and eliminated any invalid or redundant columns. We then removed records with more than 30% missing data and imputed missing values in certain columns to ensure data completeness.

Data Visualization
Next, we identified potential columns that could contribute to accurate predictions and analyzed the relationships and distributions of the column data using various graphs. Outliers in numerical variables were removed to enhance data quality.

Data Preparation
Following this, we proceeded with encoding the categorical data into dummy variables, making them compatible for model input. We also dropped columns representing 'Others' and 'Unknown' values from the dummy variables to maintain data integrity. Since the classes were imbalanced, we decided to use the SMOTE-NC (Synthetic Minority Over-sampling Technique for Nominal and Continuous) analysis technique for balancing. The dataset was then split into training and test sets in a 70:30 ratio, and the training data was scaled to avoid any disparities in data magnitude that could impact model predictions.

Model Building and Evaluation
For model development, we trained a Generalized Linear Model (GLM) using the prepared training data. The steps performed for model development included:

- Model building based on scaled features and final features selected by Recursive Feature Elimination (RFE).
- Model evaluation.
- Checking the model's ROC (Receiver Operating Characteristic) Curve.
- Checking the Variance Inflation Factor (VIF) of the variables.
- Ineffective variables were eliminated through RFE and VIF analysis, leaving us with a model comprising 14 variables and 1 constant. This model achieved upto 80% accuracy and precision on the training dataset. We then applied the same model to the scaled test dataset and observed upto 80% accuracy and precision there as well, validating the model's effectiveness. Lead score was assigned.

The Top features are:

- Total Time Spent on Website

- Lead Source _Reference
- Last Notable Activity_SMS Sent
- Lead Source_Welingak Website

Recommendations

- **Focus on High-Engagement Visitors:** Pay close attention to the total time spent on the website. When a visitor invests significant time exploring the site, it indicates a strong interest and intent to pursue further. Prioritize these high-engagement visitors, as they are more likely to convert into leads. Tailor your follow-up strategies to engage these individuals effectively, ensuring personalized and timely communication to maximize conversion potential.
- **Prioritize Leads with Recent SMS Activity:** Pay special attention to leads whose last notable activity was receiving an SMS. When a lead engages with an SMS, it indicates a level of interest and responsiveness. Focus on these leads by following up promptly with personalized communication.
- **Offer Incentives for Referrals:** Provide discounts or other incentives for referrals that convert into leads, encouraging more customers to provide references.
- **Increase Advertising Budget for Welingak Website:** Allocate more funds towards advertising on the Welingak Website to attract more leads.
- **Target Working Professionals:** Focus marketing efforts on working professionals, as they have a higher conversion rate and are likely to have the financial capacity to pay higher fees.