Summary

Lead Scoring Analysis and Model Building

Objective: The primary goal of the lead scoring analysis was to develop a predictive model that could accurately determine the probability of a lead converting into a customer. This model would help X Education prioritize their efforts and resources towards leads with the highest likelihood of conversion, thereby improving overall efficiency and effectiveness.

Data Overview: The dataset provided for analysis included leads data from the past three years, comprising 9,240 records and 37 features. These features encompassed various aspects such as lead origin, lead source, last activity, and demographic information.

Data Cleaning and Preparation: The initial step involved cleaning and preparing the data. This included removing irrelevant columns, handling missing values, and creating dummy variables for categorical columns. This process ensured that the data was in a suitable format for analysis and model building.

Exploratory Data Analysis (EDA): EDA was conducted to uncover key insights and patterns within the data. Some significant findings included:

- Leads spending more time on the website had higher conversion rates.

- The Welingak Website and Olark Chat were identified as significant lead sources.

- Activities such as SMS Sent and Email Opened were crucial in influencing lead conversion.

Feature Selection: To build an effective predictive model, it was essential to identify the most influential features. Recursive Feature Elimination (RFE) was used to select the top 20 features. Among these, the most impactful were:

- Total Time Spent on Website

- Lead Source_Welingak Website

- Last Activity_SMS Sent

- Current Occupation_Working Professional

Model Building: A logistic regression model was chosen for its simplicity and interpretability. The model was trained on the selected features, and its performance was evaluated using metrics such as accuracy, precision, recall, and F1 score. The model demonstrated high accuracy and precision in predicting lead conversions.

Model Evaluation: The evaluation metrics indicated that the model was effective in distinguishing between leads that were likely to convert and those that were not. This validation provided confidence in the model's ability to guide decision-making processes.

Recommendations: Based on the insights derived from the analysis, several recommendations were made to enhance lead conversion rates:

- **Enhance Website Engagement**: Increase the time leads spend on the website through engaging content and user-friendly design.

- **Target High-Converting Sources**: Focus marketing efforts on lead sources with higher conversion rates, such as the Welingak Website.

- **Utilize Effective Communication Channels**: Leverage SMS and email campaigns to engage leads, as these activities were found to be influential in conversions.

- **Tailor Strategies for Working Professionals**: Develop targeted strategies for working professionals, who were identified as having a higher likelihood of conversion.

Conclusion: The lead scoring model provides valuable insights that can significantly improve lead conversion rates for X Education. By implementing the recommended strategies and continuously monitoring performance, the company can optimize its sales efforts and achieve better outcomes.