Summary Report of Lead Scoring Case Study

X Education, an online education company, seeks to improve its lead conversion rate from 30% to 80%. The goal is to build a model that assigns a lead score, enabling the sales team to focus on high-potential leads.

The dataset underwent extensive pre-processing, including handling missing values, dropping irrelevant columns, and treating outliers. Recursive Feature Elimination (RFE) reduced the dataset to the 15 most relevant features, and logistic regression was used to build and refine the model. The final model, Logistic Regression Model 4 (logm4), was chosen for its significant p-values and lack of multicollinearity.

Key factors influencing lead conversion included **'Lead Source', 'Current_Occupation_Working Professional', and 'Total Time Spent on Website'.** Recommendations based on these insights are:

Several strategic recommendations were derived from the analysis:

- Emphasise features like 'Total Time Spent on Website' and 'Current_Occupation_Working Professional' in lead generation efforts due to their high conversion rates.
- Target working professionals aggressively, as they are more likely to convert and have better financial stability.
- Incentivise referral leads, which have a significantly higher conversion rate, through discounts or rewards.
- Increase the frequency of digital marketing efforts, such as Google ads and email campaigns, to enhance conversion rates.
- Prioritise leads whose last activity was 'SMS Sent' or 'Email Opened', as these actions correlate with higher conversion probabilities.
- Enhance the user experience by analysing the behaviour of customers who spend more time on the website and creating engaging content and user-friendly navigation.
- Tailor course offerings and marketing campaigns to the most popular specialisations, such as Marketing Management and HR Management, to attract and retain customers in these fields.

The data cleaning process involved addressing columns with high null values, imputing categorical data appropriately, and treating outliers. Exploratory Data Analysis (EDA) identified significant variables affecting the target variable, and the data preparation phase included creating dummy features, feature scaling, and dropping highly correlated columns.

The final model, logm2, consisting of 15 variables, was used to predict lead scores with a cut-off point of 0.41. This model highlighted the importance of sources like the Welingak Website, references, and the occupation of the lead. Recommendations based on these findings included increasing the advertising budget for the Welingak Website, incentivizing referrals, and targeting working professionals.

This case study provided valuable hands-on experience in data cleaning, EDA, data preparation, model building, and evaluation. It underscored the significance of selecting appropriate evaluation metrics tailored to business needs and highlighted factors influencing lead conversion rates. The resulting insights and strategies are aimed at substantially improving X Education's conversion rates