SUMMARY

X Education is a provider of online courses catering to industry professionals. To enhance efficiency, X Education seeks to identify high-potential leads, known as "Hot Leads," aiming to improve its lead conversion rate by concentrating efforts on the most promising prospects.

To address this, a lead scoring model is required, assigning scores indicating a lead's conversion potential. This model's deployment for future use is also a key objective. In the analysis, an outlier was spotted in the "TotalVisits" column.

In the process of scrutinizing object type variables, features lacking significant variance have been identified and subsequently dropped. Notable among these are "Do Not Call," "What matters most to you in choosing a course," "Search," "Newspaper Article," "X Education Forums," "Newspaper," and "Digital Advertisement."

Columns containing more than 40% missing values, such as 'How did you hear about X Education' and 'Lead Profile,' have been discarded as well. This systematic approach of feature selection and data refinement ensures a more focused and relevant dataset for subsequent analysis.

For analysis, the data underwent a 70:30 train-test split, normalization of numerical variables, and generation of dummy variables for categorical ones. The dataset comprises 9157 rows and 78 columns.

Logistic regression was applied, yielding promising accuracy (92%), sensitivity (80%), and specificity (93%) on the training data. The model's precision on the training data was 89%. On test data, the model demonstrated accuracy (93%), sensitivity (91%), and specificity (94%).

Random forest feature selection identified key predictors for lead conversion, including "Tags_Will revert after reading the email," "Total Time Spent on Website," and "Last Notable Activity_SMS Sent." These insights form a solid foundation for X Education's lead conversion optimization efforts, offering data-driven strategies to enhance customer targeting and lead conversion success.