Summary Report

X Education, an online course provider, attracts professionals to their website for course exploration. Lead generation occurs through form submissions and referrals. Despite numerous leads, the current conversion rate of 30% is unsatisfactory. To enhance efficiency, a logistic model is needed to assign lead scores (0 to 100) to every lead. Higher scores indicate 'Hot Leads' with better conversion potential. The objective is to achieve an 80% conversion rate by focusing on leads with higher scores. This model aims to optimize lead conversion and boost overall effectiveness.

We have followed the below steps to get achieve the objective:

1. Reading and Inspecting Data
   Read, load, understand and analyse the data.

2. Data Cleaning (Dropping Insignificant Columns)
   We found missing values and imputed them with appropriate values. For some redundant variables we have chosen to eliminate from the final data set.

3. Data Visualization and Exploratory Data Analysis
   We have used the given data and visualised it compared to our target variable. We have chosen appropriate univariate / bivariate analyses to explore the data.

4. Data Pre-processing for Model Building
   Here we scaled the numerical variables using MinMax Scaler.
   For Model Building, RFE was iteratively used thrice to filter out the most insignificant variables. RFE was used to find out the 20 most significant variables, and subsequent feature elimination was done manually. We check VIF and P values to eliminated any unnecessary variables, and arrived at a final predictor variable count of 13.

5. Making Predictions and Model Evaluation – Conversion probability was calculated for both the training and test data.

6. We assigned lead scores by multiplying the conversion probability by 100 and rounding to the nearest whole number. The final model was assessed using confusion matrix, and after plotting accuracy, sensitivity, and specificity, we determined that 0.35 serves as the optimal cut-off probability.

7. Recommendations
   It is recommended to give more priority to the leads coming via below variables as with 0.35 optimum cut-off is giving 80% of conversion of the lead.

As per the final model, below are the top 5 variables that will help the company convert the leads.

I. Total Time Spent on Website
II. Lead Origin_Lead Add Form
III. Current Occupation_Working Professional
IV. Lead Source_Welingak Website
V. Last Notable Activity_SMS Sent

The top 5 variables that have a negative effect on the lead conversion are as below:

VI. Last Notable Activity_Olark Chat Conversation
VII. Do Not Email
VIII. Specialization_Unknown
IX. Lead Origin_Landing Page Submission
X. Specialization_Hospitality Management