Summary Report

X Education is a provider of professional online courses that receives a large number of leads each month from its marketing channels. However, only about one out of three of these leads eventually converts into a paying customer. This low conversion ratio means that the sales team invests significant time and energy in following up with individuals who are unlikely to purchase. To overcome this challenge, the company needed a data-driven method to identify and prioritize the leads that are most likely to enroll.

To address this, a **lead scoring system** was developed using logistic regression.

➢ The model calculates the probability that a given lead will convert and then converts that probability into a score between 0 and 100.
➢ A higher score corresponds to a greater likelihood of conversion.
➢ By using this scoring approach, the sales staff can focus their attention on high-scoring leads, which improves efficiency and increases the chances of achieving sales targets.

The dataset used in this analysis contained close to 9,000 past lead records. These included demographic details, online engagement measures such as time spent on the website and number of visits, as well as records of sales interactions. The target field was Converted, coded as 1 for successful conversion and 0 otherwise. Before modeling, several data preparation steps were performed. These included dealing with missing information, treating "Select" responses in categorical variables as null values, creating dummy variables for categorical fields, and standardizing numerical variables to ensure consistency.

The logistic regression model was built and refined through stepwise elimination, removing predictors that were either statistically insignificant or highly correlated with others.

The final model was assessed using multiple measures beyond accuracy—namely precision, recall, specificity, and the ROC–AUC score. Since a fixed cutoff of 0.5 is not always optimal, different thresholds were tested. A cutoff range of about 0.35–0.40 was found to provide the best trade-off between sensitivity and specificity.

The model revealed several important predictors of lead conversion.

➢ The top three were: **Tags: Closed by Horizon**, **Tags: Lost to EINS**, and **Last Notable Activity: Had a Phone Conversation**.
➢ The first two were strong negative predictors, indicating leads that were unlikely to convert.
➢ In contrast, the presence of a phone conversation was a strong positive indicator, reinforcing the value of personal engagement in the sales process.

From a business standpoint, adjusting the cutoff point provides flexibility. When additional interns are available to assist with outreach, the cutoff can be lowered to around 0.3.

➢ This increases recall and ensures that nearly all potential conversions are contacted, even at the cost of calling more non-converting leads.
➢ On the other hand, when the company has already met its sales objectives for the quarter, the cutoff can be raised to around 0.7.
➢ This maximizes precision, meaning the team will only contact highly promising prospects, minimizing wasted effort.

In summary, the lead scoring model equips X Education with a practical decision-support framework. It allows the sales team to target hot leads more effectively, improves the conversion rate, and ensures resources are deployed efficiently. The approach is also transparent and interpretable, making it easy for business stakeholders to understand and trust. Looking ahead, the model can be updated with fresh data to maintain relevance, and more advanced algorithms can be explored to further enhance predictive power.