SUMMARY:

In this assignment, the primary objective was to build a logistic regression model to assign lead scores, aiding X Education in identifying potential leads more effectively. The dataset, containing around 9000 data points, was first preprocessed. Conversion values were encoded to 0 and 1, 'Select' options were replaced with NaN values, columns with high missing values were dropped, and some missing values were imputed. Rows with missing values were also dropped to enhance data quality.

The data was split into training and test sets using a 70:30 ratio. Recursive Feature Elimination (RFE) was applied to choose the top 15 variables, and a logistic regression model was built by removing variables with p-values > 0.05 and VIF > 5. The model achieved an impressive overall accuracy of 92.0%.

Further, I calculated accuracy, sensitivity, and specificity for various probability cutoffs, ranging from 0.1 to 0.9. Based on the evaluation metrics and graphical representation, the optimal cutoff was determined to be 0.19. This cutoff struck a balance between sensitivity and specificity, providing a reliable threshold for classifying leads.

Interpreting the coefficients of the logistic regression model, the top three influential features were identified: 'Tags Lost to EINS,' 'Tags Closed by Horizzon,' and 'Lead Quality Worst.' These features significantly impacted the probability of conversion, providing actionable insights for the sales team.

The logistic regression model's success lies in its ability to predict potential leads accurately. The chosen features, especially those related to tags and lead quality, played a crucial role in determining a lead's conversion probability.

For X Education's specific scenarios, I proposed strategies based on the company's objectives. During periods of aggressive lead conversion, increasing the outreach to almost all potential leads with a higher cutoff (e.g., 0.7) can be effective. Conversely, during quarters where targets are achieved early, lowering the cutoff (e.g., 0.1) can help minimize unnecessary calls while still identifying valuable prospects.

The journey of this assignment enhanced my understanding of logistic regression modeling and its application in lead scoring. Exploring strategies for different business scenarios highlighted the flexibility and adaptability required in implementing machine learning models in real-world scenarios. The iterative process of model evaluation and refinement underscored the importance of aligning data-driven insights with business goals.

In conclusion, this assignment provided hands-on experience in building and fine-tuning logistic regression models, emphasizing their significance in optimizing lead conversion processes. The learnings gained contribute to a broader understanding of how machine learning can empower businesses in making informed decisions and achieving their objectives.