SUMMARY

The Lead Scoring Model was developed to help X Education identify and prioritize potential "Hot Leads" to improve their conversion rate. The dataset used in this model contained various attributes such as lead origin, source, specialization, and occupation, among others.

The main objective of the Lead Scoring Model was to identify potential "Hot Leads" to increase conversion rates by focusing on effective lead nurturing. The model helped X Education prioritize their leads and improve their sales communication, leading to increased conversions.

The approach that was adopted to conduct the case study is as follows :

- Reading and Understanding the data : This step involved importing the data from the dataset, reading and understanding the various nuances of the dataset and later cleaning the data for further analysis. The null values were removed, outliers were checked and handled and many imbalance data were balanced.
- Visualizing Data : Exploratory data analysis for figuring out most helpful attributes was conducted including univariate, bivariate and multivariate analysis.
- Data Preparation : Creation of dummy variables and scaling of features was done in this step.
- Splitting the Data into Train and Test Set : For further analysis, the data was split into train and test set.
- Model Building : Preparation of Logistic Regression Model was done by various iterations. First, a holistic model was built, then the second model was built using Recursive Feature Elimination (RFE) technique, then fine tuning of the model was conducted by eliminating variables with high p-values and high VIFs (Variance Inflation Factor).
- Model Evaluation : Confusion Matrix, Accuracy, Sensitivity, Specificity, Precision and Recall scores were evaluated on the train set.
- Plotting the ROC Curve : The ROC (Receiver Operating Characteristic) Curve was plotted to understand the fit of the model.
- Finding the Optimal Cut off : Optimal Cut off was obtained which was used to obtain the probability scores.
- Making Predictions on Test Set : Predictions on the test set were made and metrics were evaluated.
- Precision and Recall Analysis : This analysis was conducted to obtain the better optimal cut off.
- Making Predictions on Test Set Using the New Cut off : Further, predictions were made with the new cut off and lead scoring was done.

The model was trained on the dataset, and the ROC curve was plotted, which showed a good model with a value of 0.97. The accuracy of the model was found to be 92.79%, with sensitivity and specificity values of 88.57% and 95.37%, respectively.

Furthermore, the model was tested on the dataset, and it showed an accuracy of 92.18%, with sensitivity and specificity values of 89.96% and 93.53%, respectively.

Overall, the Lead Scoring Model proved to be an effective solution for X Education to identify potential leads and improve their conversion rates, ultimately leading to increased revenue.