LEAD SCORING CASE STUDY SUMMARY :-

This analysis is done for X Education and to find ways to get more industry professionals to join their courses. The basic data provided gave us a lot of information about how the potential customers visit the site, the time they spend there, how they reached the site and the conversion rate.

The following are the steps used:

1**. Reading and Understanding the data**: - Load, read and analyze the data and its shape.

2. **Cleaning data**: - The data needs to be cleaned for null values and the value select. The select value had to be replaced with 'nan' as it indicated that the user did not give any information for that feature. Some columns with high percentage of NULL values were dropped while some missing values were imputed with median/mode as needed for numerical variables and creation of new classification variable (eg. 'not provided') for categorical variables.

3. **EDA**: - Then we did exploratory data analysis. It was found that a lot of elements in the categorical variables were irrelevant. The numeric values seem good with no outliers.

4. **Dummy Variables**: - Then we created the dummy variables for categorical values.

5. **Train-Test split**: - In the next step the data was split into train and test data set with a proportion of 70-30% respectively.

6. **Feature Scaling**: - Then we used MinMaxScaler to scale the original numerical values.

7. **Model Building**: - We use the Recursive Feature Elimination to get the top 15 most relevant variables. Now depending on the VIF values and p-value the rest of the variables are removed.

8. **Model Evaluation**: - Taking an initial assumption that a probability of more than 0.5 means 1 else 0, we created the data frame having the converted probability values. Based on the initial assumption, we derived the Confusion Metrics. We then calculated the overall 'Accuracy', the 'Sensitivity' and the 'Specificity' of the model matrices to evaluate how reliable the model is.