Summary

This Analysis has been done for an education company named X Education that sells online courses to industry professionals. The goal was to help them select the most promising leads i.e., the leads that are most likely to convert into paying customers.

Steps followed for the modelling:

Step 1: Clean Data

- ⇨ Categorical values as “Select” was equivalent to null, had converted to null.
- ⇨ Removed columns with null % of 50 and above.
- ⇨ Imputed with appropriate values for other variables.
- ⇨ Removed variables that were insignificant for the case study.

Step 2: EDA

- ⇨ Performed Univariate Analysis for each variable against target variable “Converted”.
- ⇨ Removed variables that were not contributing towards classification of the lead conversion.
- ⇨ Encoded the binary categorical variables to 1/0 for Yes/No values.

Step 3: Prepare Data

- ⇨ Created dummy variables using one-hot encoding technique.
- ⇨ Removed the original variables after creation of dummy variables

Step 4: Model Creation

- ⇨ Created x, y variables
- ⇨ Performed a Train-Test Split
- ⇨ Performed Feature Scaling
    - o Standardize the variables with high Variance
- ⇨ Train the model
- ⇨ Feature Selection using techniques:
    - o RFE
    - o VIF
    - o P Values
- ⇨ Asses the models using metrics and ROC.
- ⇨ Calculated the optimum cutoff probability
- ⇨ Test the Model
- ⇨ Generate the Lead scores between 0-100

Final Model Metrics

Train Set:

Overall Accuracy	:	79.24%
Sensitivity	:	77.51%
Specificity	:	80.33%
Precision	:	71.17%
Recall	:	77.51%

Test Set:

Overall Accuracy	:	79.47%
Sensitivity	:	76.84%
Specificity	:	80.96%
Precision	:	69.72%
Recall	:	76.84%