LEAD SCORING CASESTUDY SUMMARY

Problem Statement:

An education company named X Education sells online courses to industry professionals. Company wishes to identify the most potential leads, also known as **'Hot Leads'.**

For this company wants to create a model in which we need **to** assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Business Objective:

Build a logistic regression model to assign a lead score to find "Hot Leads" so that the lead conversion rate at X education is around 30%.

Methodology Used:

*Reading and understanding the data

*Data Cleaning and Manipulation

. Check and handle the duplicate data
. Verify and handle the SELECT values, NA values and missing values
. Drop the columns which has large number of missing values
. Imputation of values, if necessary
. Handle the outliers

*Exploratory Data Analysis

. Univariate Analysis & Bivariate Analysis
. Outliers are identified and removed

*Data Transformation
. Changed binary variables

*Reading and understanding the data

*Dummy Variable Creation

. Created dummy variables for Categorical values
. Removed all redundant variables

*Classification Technique

. Logical Regression used for making models and predictions

*Test-Train Split

*Validation of models -Standard scaling used

*Model Building

. Selected 15 top most features

. Created 4 models and looked for p-values in order to select most significant values.

. Concluded at 12 significant variables for which VIF was good.

. Checked for Accuracy, Specificity and Sensitivity.

. ROC Curve had a coverage of 88%.

. Checked Precision, Recall Trade-off.

. Implemented learning to test model as well and calculated conversion probability.

Conclusion:

- Projected Values from Train Data:

Accuracy : 80.3%
Sensitivity :80.78%
Specificity : 80.15%

- Projected Values from Test Data:

Accuracy : 79.2%
Sensitivity :82%
Specificity : 77.5%

- The lead score calculated in the trained set of data shows the Conversion rate on the final predicted model is around 80% .
- Top 3 important features responsible for good conversion rate or the ones' which contributes more towards the probability of a lead getting converted in decreasing order are :

  #1) Lead Source_Welingak Website

  #2) Lead Source_Reference

  #3) What is your current occupation_Working Professional