SUMMARY

This analysis is done for X Education to find ways to get more industrials professionals to join their course. The data provided gives us a lot of information about how the potential customers visit the site , how they reached site and the conversion rate .

Steps used for this case study :-

- ## Cleaning data = Removed columns with high missing values .Imputed missing values for columns with fewer missing values . Removed Outliers .

- ## EDA = Started with the Exploratory Data Analysis of the data set to get a feel of how the data is oriented in relation to the target variable – 'Converted' .Identified variables that were having only single values and dropped them .

- ## Data preparation = Converted Binary Variables – Yes / No to 1 / 0 .Create Dummy Variables – for Categorical variables.

- ## Dummy variables = The dummy variables were created for categorical features , then the original features dropped after concatenating the newly created dummy variables .

- ## Scaling = Used Standard scalar to scale the data for Continuous variables.

- ## Train-Test splitting = The split was done at 70% and 30% for train and test respectively.

- **Model Building:** By using RFE. It gives top relevant variables. Later the irrelevant features were removed manually depending on the VIF values and p-value (The variables with VIF < 3 and p-value 0.05 were kept).

- **Model evaluation** = An ROC curve demonstrates several things:

It shows the tradeoff between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity).The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test. The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.

Quick Stats:-	
Accuracy: 84%	Specificity: 85%
Sensitivity: 83	Roc: 91%

- **Recommendation=** The company should make calls to the leads coming from the lead sources
  - ➢ "Welingak Websites" and "Reference".
  - ➢ "working professionals".
  - ➢ "more time on the websites".
  - ➢ "whose last activity was SMS Sent".