**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. Cleaning data: The data was partially clean except for a few null values and the option select had to be replaced with a null value since it did not give us much information. Few of the null values were changed to 'not provided' so as to not lose much data. Although they were later removed while making dummies. Since there were many from India and few from outside, the elements were changed to 'India', 'Outside India' and 'not provided'.

2. EDA: A quick EDA was done to check the condition of our data. It was found that a lot of elements in the categorical variables were irrelevant. The numeric values seems good and no outliers were found.

Data Transformation:

- Changed the multicategory labels into binary variables in the form of '0' and '1'.
- Created dummy variables for some variables.
- Checked the outliers and removed some of the numbers using 0.99- 0.1% analysis. Data Preparation:
- Splitting the dataset into train and test dataset.
- Scaled the dataset using the StandardScaler().
- Plotted heatmap for finding the correlations and dropping them.

**Model Building:**

- We build our model with the help of RFE with 19 variables.
- Checked the VIF Score for each variables, as all of the variables are having VIF Score < 5.0, we proceed to our next step.
- We then removed the insignificant variables using the P-Value Score.
- For our final model we checked the optimal probability cutoff by finding points and checking the accuracy, sensitivity and specificity.
- We found one convergent points and we chose that point for cutoff and predicted our final outcomes.
- We checked the precision and recall with accuracy, sensitivity and specificity for our final model and the tradeoffs.
- Prediction made now in test set and predicted value was recoded.

- We did model evaluation on the test set like checking the accuracy, recall/sensitivity to find how the model is.
- We found the score of accuracy and sensitivity from our final test model is in acceptable range.
- We have given lead score to the test dataset for indication that high lead score are hot leads and low lead score are not hot leads.

**Conclusion:**

Learning gathered below: ϖ Test set is having accuracy, recall/sensitivity in an acceptable range. ϖ In business terms, our model is having stability and accuracy with adaptive environment skills. Means it will adjust with the company’s requirement changes made in coming future. ϖ Top features for good conversion rate:

- 1. Last Notable Activity_Had a Phone Conversation
- 2. Lead Origin_Lead Add Form
- 3. What is your current occupation_Working Professional