

## Summary

X education wants to have a more precise way to figure out which potential lead are going to convert into actual students for their online course. The data provided is very extensive and is inclusive of lead generation source to the amount of activity the lead performs on the website. The data is a little messy and needs some clean up as well as some preparation before we can jump onto analysing it.

Following are the steps that were performed in order to reach conclusive observations and a model to predict the conversion score.

### 1. Missing Data treatment and EDA:

We worked on the missing data treatment and basic EDA together as that gives us a better picture as to if we can drop a column or not. Also it tell as the influence of a particular feature on the target variable.

Some of the features had huge chunks of data missing and had no real effect on the target variable so they were directly dropped.

Some features like TotalVisits had very few rows missing which suggested the lead never visited the website, so they were treated accordingly.

### 2. Outlier Treatment:

Columns `TotalVisits`, `Total Time Spent on Website` & `Page Views Per Visit` have outliers in the upper quantiles and have been treated accordingly

### 3. Dummy Features:

Majority of the features are in categorical values and hence were encoded

### 4. Scaling Features:

Since the page visits and time spent on website are have some really huge values we have used normalizations to have all the features in same range

### 5. Model Building:

We used RFE on the initial model to choose top 20 parameters and then further worked through the list to have VIF < 5 and p-value close to 0

Following are the model features:

- a. Accuracy of the model is 88%
- b. Sensitivity for the model is 85%
- c. Specificity for the model is 88%
- d. Recall of the model is 85%
- e. Precision of the model is 81%

**Key Insights:**

- We can invest more on the google traffic building activities as not only they generate the highest number of lead they also have great conversion rate
- Employed leads are most likely to convert
- We can invest more on building on our website even more informative and engaging as people spending more time online are converting better.
- People who revert back on once the email is read tend to have high success rate for converting, which resides with the ideology that we usually interact more when we are actually interested