**Summary**

**Problem Statement**

* X Education, an online education provider for professionals, attracts daily website visitors interested in its courses. These leads come from various marketing channels, including websites and search engines like Google. After engaging with the site, they may explore courses or submit contact information, becoming leads. The sales team then contacts them, aiming for a 30% conversion rate.
* Despite many leads, X Education struggles to convert them efficiently. They want to identify 'Hot Leads' for better conversion, focusing on the most promising prospects. Nurturing leads through education and personalized communication is crucial.

**Objectives**

* Build a model that gives leads a score of between 0 and 100.
* Use Logistic Regression to help increase the conversion rate to 80%

**Steps**

1. Read data
2. Clean and Visualise Data
3. Build Model and Evaluate model

The dataset given was useful in providing actionable insights. First when I deep dive into the data, I noticed a lot of columns had plenty of missing values. This means that I now have to decide between dropping them or removing the rows. I decided that for columns with more than 1/3 missing values, I will remove these columns. For the rest, I shall keep.

I also noticed that the word “Select” appeared a lot of times perhaps as an error in the dataset. However, this time, I decided to remove the rows instead of the columns as this will allow me a bigger dataset to run my model and also by not removing the column, I will keep the rest of the data which may prove useful.

After running my model I evaluated the result. I tried looking at the VIF to determine which values to drop further due to high coef. Which increased the accuracy and performance of my model. Once that is done I proceeded to plot the ROC curve and use that to decide on the cut off. My result was 0.2 for the cut off. Finally I ran the model again to get a result of 60% for sensitivity and 78% for Specificity.