# Lead Score - Case study Summary

**Problem Statement:**

An education company named X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google. When people fill up a form providing their email address or phone number, they are classified to be a lead. The typical lead conversion rate at X education is around 30%. Although X Education gets a lot of leads, its lead conversion rate is very poor.

As a Data scientist we need to identify the 'Hot leads' so that lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

**Solution Approach:**

Starting-off with data inspection we have identified unnecessary columns and dropped them. Checked for missing values and handled by dropping the columns. We treated the columns which had 'select' value separately. We have done outlier analysis and used power Transformation to remove the skewness from the numerical variables.

As part of Data pre-processing we did train test split and scaling for the train set. Used RFE to find top 20 features for the model building.

On an iterative approach we have looked at P-values and VIF values to arrive at the final model.

We did Predictions on the Test set using an arbitrary cut-off as 0.5. Evaluated the model using ROC curve and plotted accuracy, specificity and sensitivity to find the optimal cut-off which is 0.42.

We did Predictions on test set from both Specificity-Sensitivity view and Precision- Recall view. We have obtained a good accuracy in both the views.