Summary:

This Analysis is done for X education and to find ways to get more industry professionals to join their course. 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 conversion rate.

The following are the steps used:

1.Cleaning data:

We cleaned up the data by dropping the columns where there is no diverse data , which have null values . We identified the columns where the data points have 'Select' , as this is as good as null value , necessary imputation has been done .

2. EDA:

We did pair plot of the numerical variable through which we understood how they are variating against each other. We used the heat map to find out the correlation between features .

3. Dummy Variables:
For all the categorical variables we extracted the dummy variables

Eg : Lead Origin_Lead Add Form , Lead Source_Welingak Website ,Last Activity_Email Bounced , Last Activity_Had a Phone Conversation etc.

4. Train-Test split:

The split was done at 70% and 30% for train and test data respectively.

5.Model Building:

We used RFE to attain the top 15 features , then we did the summary and verified VIF against the features . We dropped the features with high p values and high VIF (which shows they are highly irrelevant and correlated with other variables). Later we took a cut off 0.5 and build a logistic model using the features we are

left with , post which we evaluated the model by building confusion matrix and extracting specificity and sensitivity of the model . Later we evaluated the model with different cut off values and identified the best cut off by using the cut off plot , and we extracted a cut off of 0.4 .

7. Precession Recall view:
build the training model using the precision-recall view and calculate precision , we still arrived at 0.4 as cut off

Recall.

6. Model Evaluation:
This cut off gave us better results over our default cut off 0.5. We also evaluated the model using this cut off and built a model on the test data to see the performance of the model . This cut off turned off to be good for us

8.Conclusion:

In conclusion we can see that we can make use of features like ,'Lead Origin_Lead Add Form' ,'Last Activity_Had a Phone Conversation ','Specialization_Banking, Investment And Insurance,Marketing Management ,Rural and Agribusiness'

can be used as hot leads , which can be quickly converted .

These leads need to be curated well to achieve better conversion , on top we have some features which would reduce our conversion , we need to pay closer attention to these leads . We can add new features to the Specialisations which can boost our lead conversion