An Educated Approach to help imporve X Education to increase their lead convesion rate

X Education wants to increase their target lead conversion rate to be around 80%. For this they have chosen to rely on past data to derive trends and thus determine the most promising leads.

Outcome of this analysis is to help select the most promising leads.

The first step inthis direction was to prepare the data, the following were carried out on the dataset

- Few columns had ‘Select’ value : 'Specialization', 'How did you hear about X Education', 'Lead Profile', 'City' : these were converted to ‘Missing’ and merged with other null values in the column
- These fields have outliers: TotalVisits, Total Time Spent on Website, Page Views Per Visit
- Columns having more than 45% null values were dropped, except for Lead Quality
- Columns having unique values were dropped, some of which were Magazine, Receive More Updates About Our Courses etc
- Other missing value columns were imputed with the Median or Mode or a new category was created.
- Irrelevant Columns were dropped
- Binary variables were converted to 0’s and 1’s
- Other categorical variables were hot encoded.

Next in the model buling process: the following steps were taken

- RFE and Manual feature selection technique was employed to arrive at the final logistic regression model.
- Data was split into train and test sets & model was built on train set
- Data was scaled
- The following features were removed because of high P - value : What is your current occupation_Housewife, Newspaper, Lead Source_NC_EDM.
- 0.33 is the optimum point as a cutoff probability
- The Model was tested on test set
- Results discussed on next slide
- The original data set was assigned a lead score for each lea, a value between 0 -100, 100 being the hottest lead.

Results

1. Three most important features for lead conversion are:

    -Lead Source_Welingak Website

-Lead Origin_Lead Add Form

-What is your current occupation_Working Professional

2. Comparison of metrics for Test and Train data

Train Data Set metrics
Sensitivity: 0.84
Specificity: 0.81
Precision: 0.75
Recall: 0.83
Accuracy: 0.82

Test Data Set metrics:
Sensitivity: 0.77
Specificity: 0.85
Precision: 0.74
Recall: 0.77
Accuracy: 0.82

3. For an aggressive approach to increase the conversion rate : Choose a lower threshold value for Conversion Probability

4. For a frugal approach to increase the conversion rate : Choose a higher threshold value for Conversion Probability
Thankyou.