

Group Participants		
Snehal Kedar	Siva Sankare	Smith Desai

## Lead Scoring Case Study

This is a brief summary report explaining process followed and the learnings that are gathered.

### ❖ Problem Understanding

First step to solve a problem is to understanding the business objective defined by X Education Company correctly.

### ❖ Data Analysis & Preparation

Doing initial data analysis is also highly important as having correct data in place is another step towards achieving goal of this activity. Hence a detailed look on the data set was taken to understand kind of information present in it.

Post that, we tried to make sure data isn't erroneous to the possible extent, by doing following cleaning activities:

- handling missing values
- checking outliers
- dropping or imputing columns
- creating of dummy variables for further processing.

Next, Exploratory data analysis is done by plotting different graphs to understand trend in data across different features.

### ❖ Train & Test Data Set:

Data split was done at 70% & 30% into train and test data set respectively using split function.

Feature Scaling was done for numerical columns to fit data in the range of 0 to 1.

### ❖ Model Building:

1. RFE feature selection method used to select top 15 features for which model was created using Logistics regression.
2. Conditions of P-value  $\leq 0.05$  and VIF  $\leq 5$  used to select/remove Features manually.
3. Model accuracy was evaluated using different parameters like confusion matrix, sensitivity, specificity, True & False Positive and Negative rate, ROC and accuracy - sensitivity – specificity plot to decide the probability cut-off, precision and recall to confirm accuracy.
4. Prediction done for testing data once expected accuracy on training is achieved. Accuracy checks on testing data is done. Lead score is calculated for test data.

### ❖ **Conclusion/ Recommendations:**

Top 3 Features that emerged from the analysis which can be used to convert leads into paid customers are:

1. TotalVisits
2. Total Time Spent on Website
3. Lead Origin to be more specific - Lead Origin\_Lead Add Form

X Education company should make calls to the leads following in below category as chances of they are getting converted to paid customer is more.

- Leads with more total visits.
- Leads who are spending more time on website.
- Leads whose origin is Lead add form.
- Leads whose current occupation is Working Professional.
- Leads whose source is Welingak Website.
- Leads whose last activity is 'Had a Phone Conversation'.
- Leads whose source is Olark Chat.

### ❖ **Learnings from the case study**

1. This case study helped us to get a glimpse of what real life problems could look like and how Systematic approaches like 4 - Step approach can help us in real life to tackle business issues and consideration we should keep in mind.
2. This also helped us learning step by step issue analysis, data preparation and building model with feature selections, choosing best fit model based on accuracy parameters and derive recommendation to the business from the outcomes.

This was a really good learning experience.