

## **SUMMARY FOR THE CASE STUDY**

### **SOLUTION:**

The case study is on the X Education data set, we need to provide the Lead score of the students who can be their prime candidates such that they will be converted as valuable clients. From the given info, we can derive the insights like, how the leads can be converted.

**Roadmap followed, to achieve the solution with the below mentioned:**

#### Data preparation:

Preparing the data involves checking the duplicates and null values and there is one more category which is “select” left by the leads and were good as null values. So those “select” options from all the variables and replace it as null values. Later removed columns which are having more missing and null values.

#### EDA:

EDA is done to know how leads are acting basing on different aspects and to know what were the aspects that leads are more likely to be converted and which are not. Checked for the outliers and treated accordingly.

#### Dummy variables:

Dummy variables were created using the categorical variables and the variables which were felt not needed were dropped. Later scaling is done to bring all the variables to the same scale.

### Test-Train-split:

Splitting is done on the data to train the model which is 70% and then test set is the remaining 30% to check how the model is working.

### Building the model:

Firstly, built the model and then used RFE to get the top 15 relevant and further removed 2 more columns which are having high p value.

### Evaluation of model:

Using the confusion matrix, we got the specificity of 83.9 % sensitivity and accuracy are 94.3 % and 87.7% respectively.

We can depict the features provide more valuable are:

1. Total Number of visits
2. Lead Source's like: a) Google b) Direct Traffic c) Organic Search d) Welingak website
3. Total time spent on website.
4. Lead Origin: Lead add format
5. SMS Olark chat conversation
6. Working professional

Using these features X Education can make Leads converted.