Summary:

The logistic regression binary analysis is done for X Education company. The main aim is to find ways to convert the potential leads into hot leads.

The following steps were performed:

1. Reading Data

While reading the data, there are missing values, and some data having select option which needs to be converted into null. Once data was cleaned, need to observe if any outliers in data. For this, EDA on our data is required.

2. Exploratory Data Analysis

Based on EDA, i.e., Heat map/ Box plot/ count plot which gives a fine understanding of variables with target variable.

3. Dummy Variables

The dummy variables were created as it becomes more focused on variable w.r.t target variable.

4. Split into Train – Test

The Final data after cleaning, scaling, dummy variable creation, the whole data is divided into train and test data in the ratio of 70:30 respectively.

5. Scaling of Data

Standard Scaling is used on numeric variables in data.

6. Model Building

As Heatmap of whole data is very huge and will not get any observation. Hence recursive feature elimination (RFE) is done to reduce the data variables. It is reduced from 58 to 20. Based on p-value and VIF, some variable were removed.

7. Train Model Evaluation

First confusion matrix is made and optimum value is found by plotting ROC curve.
Based on Specificity and Sensitivity train model is evaluated.

8. Test Model Evaluation

Based on Specificity and Sensitivity test model is evaluated.

Below are the variables which contributed the most are (in decreasing order)
   i.   Tags_Closed by Horizzon
   ii.   Tags_Lost to EINS
   iii. Lead Source_Welingak Website
   iv.   Tags_Will revert after reading the email
   v.   Tags_Already a student

Below are the variables focused on more:
Specialization – Management
Specialization – Business
Working Professional
Lead through
        Google
        Chat
        SMS Chat