Summary

Analysis is being done for X Education Company to increase the number of people join their courses. At present, sales teams conversion rate is around 39 %, CEO wants to make it atleast 80%.

Using the data provided analysis was done using the steps mentioned below:

1. Reading of data
    a. Data was provided in form of csv which was read to find out basic details like number of rows and columns
2. Cleaning of data
    a. The data set contained missing values
    b. Columns with more than 75% missing values were dropped
    c. Some columns were replaced with mean, some with mode and some with na
3. Data Analysis
    a. EDA helped in getting imbalnce data
    b. Relationship with conversion
4. Data Preparation for Modelling
    a. Binary results were converted to 1/0
    b. Dummy variables were introduced
    c. Outliers were capped
    d. Correlated features were removed
5. Data Modelling
    a. Data was split into training and test data set in ratio 70:30
    b. Feature selection was done using RFE
    c. Model was improved by using statsmodel
    d. Model was build
    e. Test data were executed on the model
6. Prediction was done for the data and lead scores were calculated
7. Accuracy, Sensitivity, Specificity was calculated
    a. Training data:
        i. accuracy : 90%
        ii. sensitivity: 89%
        iii. specificity: 91%
    b. Test data:
        i. accuracy : 90%
        ii. sensitivity: 88%
        iii. specificity:91%

Based on the model it was found that:

- Tags to focus on:
    - switched off
    - Busy
    - Lost to EINS
- Identify people with null values for tags and Leads profile
- Focus on Lead Sources from Welingak Website

- Focus on people whose last activity was email opened or sms sent.

With accuracy of model being around 90%, focusing on above can help X Education company to get more leads and convert them too.