Summary

1. The lead score case study objective is to target potential leads, a model built through logistic regression
2. Started with checking the data frame and dropped irrelevant columns due to brevity reason and dropped columns which have more than 3000 missing values
3. The columns which have less than 3000 missing values, removed the rows to refine the data frame
4. Did univariate and multivariate analysis to understand the relationship of features
5. Dropped some categorical columns where one level has a high percentage of data inclined to it such as Newspaper column No and Yes have 9239 and 1 values respectively
6. For the remaining categorical values applied dummy variable technique to change the data to 1's and 0's
7. Assigned the converted variable to y and remaining to X and Splitted the data to test and train parts
8. Built logistic regression model by eliminating variables till p-value and VIF value of all variables in the model decreased below 0.05 and 5 respectively
9. For model evaluation, created a data frame with the actual converted column and conversion predicted and mapped the later to predicted column with 0.5 cutoff as 0 for less than or equal to 0.5 and 1 to the other part
10. Later 0.42 optimal cutoff was found by the intersection of accuracy, sensitivity and specificity lines on probability cutoffs x-axis
11. Made predictions on the test data and finally recall achieved approximately 80% which is the target lead conversion rate ballpark set by the X education CEO.