1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?
    a. Tags_Closed by Horizzon
    b. Tags_Lost to EINS
    c. Tags_Will revert after reading the email

2. What are the top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion?
    a. Tags_Will revert after reading the email
    b. Lead Source_Welingak
    c. Total Time Spent on Website

3. X Education has a period of 2 months every year during which they hire some interns. The sales team, in particular, has around 10 interns allotted to them. So during this phase, they wish to make the lead conversion more aggressive. So they want almost all of the potential leads (i.e. the customers who have been predicted as 1 by the model) to be converted and hence, want to make phone calls to as much of such people as possible. Suggest a good strategy they should employ at this stage.
    a. We know that a high sensitivity value indicates a low rate of missed positive instances (false negatives), suggesting that the test is effective at capturing true positive cases. To achieve this we will lower the probability cut off for conversion to a lower value like 0.2 instead of the actual cut off of 0.36. The result we get are as follows
        i. Training set
            1. Sensitivity / Recall : 0.9386753883892068
            2. Specificity : 0.8865556978233035
            3. False Positive rate : 0.11344430217669654
            4. Positive Predictive Value / Precision : 0.8382621394669587
            5. Negative Predictive Value : 0.9584717607973422
        ii. Test set
            1. Sensitivity / Recall : 0.9241658240647118
            2. Specificity : 0.8765859284890427
            3. False Positive rate : 0.12341407151095732
            4. Positive Predictive Value / Precision : 0.8102836879432624
            5. Negative Predictive Value : 0.9529780564263323
    b. As we can see the for the training set the Sensitivity has increased to 93 percent, which means we will be able to capture maximum conversion

4. Similarly, at times, the company reaches its target for a quarter before the deadline. During this time, the company wants the sales team to focus on some new work as well. So during this time, the company's aim is to not make phone calls unless it's extremely necessary, i.e. they want to minimize the rate of useless phone calls. Suggest a strategy they should employ at this stage.

a. We know that a high specificity value indicates a low rate of false positives, suggesting that the test is effective at avoiding misclassifying negative cases as positive. To achieve this we will raise the probability cut off for conversion to a higher value like 0.4 instead of the actual cut off of 0.36. The result we get are as follows

b.
    i. Training set
        1. Sensitivity / Recall : 0.9043336058871627
        2. Specificity : 0.9480153649167734
        3. False Positive rate : 0.051984635083226635
        4. Positive Predictive Value / Precision : 0.9159420289855073
        5. Negative Predictive Value : 0.9405487804878049
    ii. Test set
        1. Sensitivity / Recall : 0.8827098078867543
        2. Specificity : 0.9382929642445214
        3. False Positive rate : 0.06170703575547866
        4. Positive Predictive Value / Precision : 0.8908163265306123
        5. Negative Predictive Value : 0.9334480780263913

c. As we can see, the training set the Specificity has increased to 93 percent, which means we will be able to identify false positives more accurately.