LEAD SCORE ASSIGNMENT – SUBJECTIVE QUESTIONS

Question 1: Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

Answer: Based on correlation matrix after running model, top three variables contributing the most towards the probability of a lead getting converted are:

1. Tags_Will revert after reading the email: 0.65
2. Total time spent on website: 0.35
3. Last Activity_SMS Sent: 0.34

These 3 factors impact the most positively to drive for the successful conversion which are from , "Total Time Spent On Website", "Tags" and "Last Activity" columns. Therefore, when collecting the data, we need to pay more attention to improve the quality of inputs for these 3 columns by:

1. Tags: There are 35% of tags data is unknown, meaning missing values, we need to minimize the missing values of this column by set a compulsory rule for this filter.
2. Last Activity_SMS Sent: similarly to the case of 'Tags_Close by Horizzon', we need to explore further to trigger customers to SMS Sent Step.
3. Total Time Spent On Website: is moderately important, therefore, we need to improve our information quality on our website to trigger customers browsing information more often, the more time they spent on website, the higher chance they will be converted. Based on exploratory data from this column, the group of customers that spend more than 3 hours on the website have significantly higher chance to be converted.

Question 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?

Answer: Similarly based on correlation matrix, top 3 categorical/ dummy variables we should focus on in order to increase the probability of lead conversions are:

1. Tags_Will revert after reading the email: 0.65
2. Last Activity_SMS Sent: 0.34
3. Tags_Closed by Horizzon: 0.23

'Close by Horizzon' though being moderately important, the percentage of this group is very small (~5%) with nearly 100% is converted, therefore, we need to explore further into this group to expand it further.

Question 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.

Answer: After running logistic regression model, we have found a good model with high accuracy rate (92%), sensitivity (93%) and specificity (91%) to classify whether a lead is a potential one or not. Based on this model, we can generate a good list of potential leads for sales team to maximize the conversion rate with the cut off at 0.3.

Question 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.

Answer: To minimize the rate of useless phone calls, sales team also should not waste time to those customers who have status marked as below variables because these variables impact negatively to the conversion which includes:

1. Tags_Ringging: -0.28
2. Last Activity_Modified: -0.26