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

   Ans:
   As per the residual summary shared for the final model (4th model), the model coefficients were:

```
 2 res.params

const                                                      -10.153529
Lead Number                                                  0.000009
Total Time Spent on Website                                  1.091771
Lead Origin_Lead Add Form                                    3.998760
Lead Source_Olark Chat                                       1.155039
Lead Source_Reference                                       -2.332098
Last Activity_Olark Chat Conversation                       -0.749769
Last Activity_SMS Sent                                       1.705052
Specialization_Not Specified                                 0.448399
What is your current occupation_Working Professional         0.956823
Tags_Closed by Horizzon                                      9.534920
Tags_Not Specified                                           2.875372
Tags_Other                                                   2.759852
Tags_Will revert after reading the email                     6.813244
Lead Profile_Potential Lead                                  0.970168
Last Notable Activity_Modified                              -1.146408
dtype: float64
```

   Also, the heatmap showing the correlation between target variable and numerical variable is shown below:
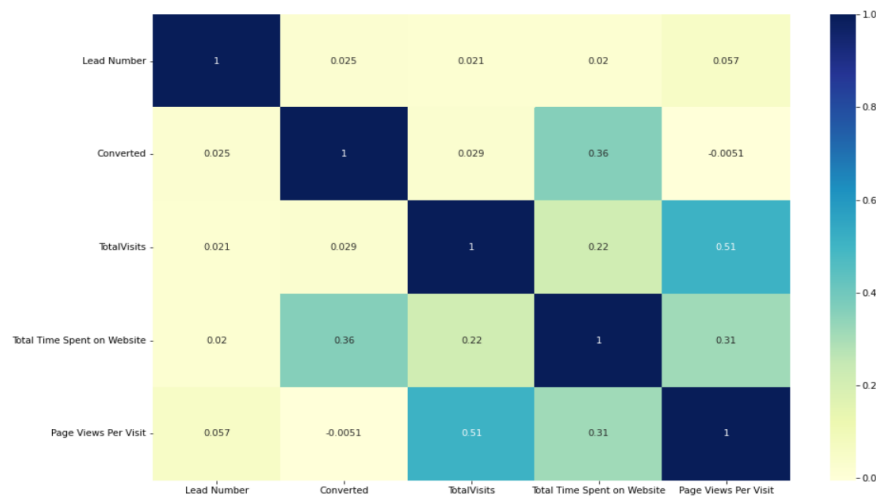
   As per the heatmap and model coefficients, the top 3 variables positively influencing the lead conversion rate would be:

- Lead Origin_Lead Add Form
- Total Time spent on the website

- Last Activity_SMS sent

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?

   Ans:
   As per the residual coefficients obtained in the model summary, the top 3 categorical/dummy variables would be:

Variable	Coefficient
Tags_Closed by Horizzon	9.53
Tags_Will revert after reading the email	6.81
Lead Origin_Lead Add Form	3.99

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.

   Ans:
   A generic sales engagement strategy, which is highlighted by the top variables positively influencing conversion rate through our highly accurate model would be:
   1. The calls can be made to the customers who spent lot of time on the website. Based on their browsing history, these customers can be targeted to the specific programs offered by the X Education.
   2. The calls can be made to the customers who are identified through the 'Lead add form' as it has higher conversion rate than rest of the identifiers.
   3. The calls can be made to the customers whose last activity is sending SMS.
   4. The call can be made to the customers whose Leads received through **references** and **Wellingak Website.**

   However, a more specific answer for this case in terms of specificity and sensitivity metrics, which are perfect to add to the goodness of a highly accurate model is:

   Sensitivity can be defined as the total number of conversions which were correctly predicted as conversions divided by the total number of actual conversions

Specificity, on the other hand, is the total number of non-conversions correctly predicted as non-conversions divided by the total number of actual non-conversions.

A good model desires both high sensitivity (to only reach out to potential 'HOT' leads) and high specificity (to avoid wasting time and resources on trash leads).
However, for any model, as the specificity increases, the sensitivity decreases and vice versa.

As per the graph shown below:
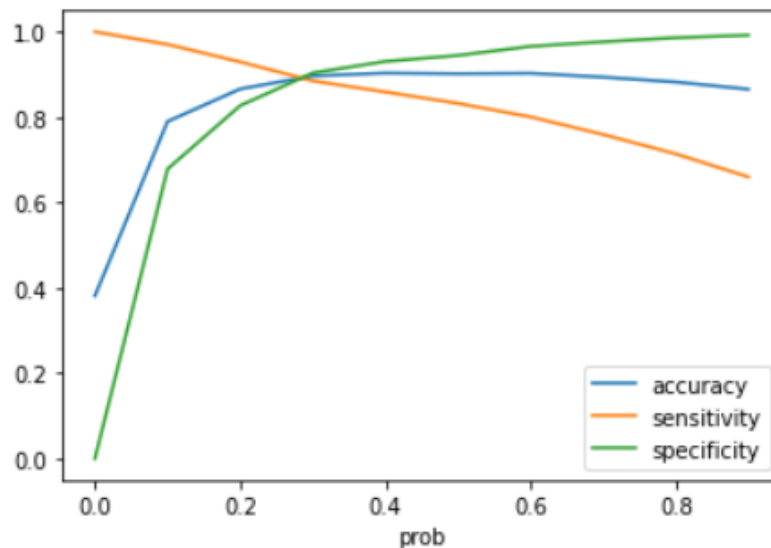
The cutoff taken for the model was 0.3. If we take a higher cut-off, the specificity would increase and sensitivity would decrease and for a lower cut-off, the sensitivity would increase and specificity would decrease.

The task for X education is to be more aggressive in lead conversion. The team wants to convert all the potential leads. In such case, a model with high **Sensitivity** is required. Hence, we will choose a lower cut-off than the current 0.3 in this case to increase the sensitivity.

This model will correctly predict all the hot leads that can be converted. It may also lead to some trash leads to be predicted as hot leads but that will be small proportion.

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.

Ans:

In this case, X education has already reached its target for lead conversion in the quarter well before the deadline. So, it need not be overly aggressive in making phone calls to every possible lead. This will only lead to waste in time, effort and resources and is not recommended. X education would want to be prudent in such a case.

For the above case, a model with high **Specificity** is desired. We can choose a higher cutoff than the current 0.3 and the specificity would increase for this model.
Such a model will correctly predict the most likely non-conversions before making any calls. This will ensure that leads with a low lead score that are least likely to get converted are not selected by the model as HOT leads. The company's salesperson need not make any unnecessary calls since the target is well achieved and can focus on the new work, as desired.