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

➔ Following are the different variables and its coefficients in the final model:

Variable_Name	Coefficient
Const	-3.1351
Do Not Email	-1.4010
TotalVisits	1.7680
Total Time Spent on Website	4.6546
Last Activity_Had a Phone Conversation	2.1145
Last Activity_SMS Sent	1.5764
Lead Origin_Lead Add Form	3.8776
Lead Source_Olark Chat	1.2890
Lead Source_Welingak Website	2.4821
Occupation_Working Professional	2.7065
Lead Profile_Potential Lead	1.8078
Leaf Profile_Student of SomeSchool	-2.1685

➔ From the co-efficient of all the variable, highest absolute coefficients are for the features Total Time Spent on Website, Lead Origin_Lead Add Form and occupation_Working Professional.

➔ Higher absolute co-efficient means that particular variable is more important for the given model. Top 3 variables in the model are:

1) Total Time Spent on Website

2) Lead Add Form (From Lead Origin)

3) Working Professional (From What is your current occupation)

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?**

➔ Following are the different variables and its coefficients in the final model:

Variable_Name	Coefficient
Const	-3.1351
Do Not Email	-1.4010
TotalVisits	1.7680
Total Time Spent on Website	4.6546
Last Activity_Had a Phone Conversation	2.1145
Last Activity_SMS Sent	1.5764
Lead Origin_Lead Add Form	3.8776
Lead Source_Olark Chat	1.2890
Lead Source_Welingak Website	2.4821
Occupation_Working Professional	2.7065
Lead Profile_Potential Lead	1.8078
Leaf Profile_Student of SomeSchool	-2.1685

➔ From the co-efficient of all the variable, highest absolute coefficients are for the features Lead Origin_Lead Add Form, occupation_Working Professional and Lead Source_Welingak Website which are of the categorical / dummy type.

➔ Higher absolute co-efficient means that particular variable is more important for the given model. Top 3 categorical / dummy variables in the model are:

1) Lead Add Form (From Lead Origin)

2) Working Professional (From What is your current occupation)

3) Welingak Website (From Lead Source)

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.**

➔ For making the sales more aggressive, so that almost all the potential leads are contacted via phone calls, we can use low threshold value.

➔ On making the threshold value low, there is very less chance of missing any potential leads.

➔ Below is the table for different cut-off values for the test set and its related different evaluation measures:

Cut-off	Accuracy	Sensitivity	Specificity	Precision	Recall
0.1	0.605915	0.967423	0.389480	0.486836	0.967423
0.2	0.757116	0.889437	0.677896	0.623098	0.889437
0.3	0.790388	0.836130	0.763002	0.678686	0.836130
0.4	0.809242	0.769990	0.832742	0.733772	0.769990
0.5	0.803697	0.663376	0.887707	0.779582	0.663376
0.6	0.800000	0.615005	0.910757	0.804910	0.615005
0.7	0.786322	0.536032	0.936170	0.834101	0.536032
0.8	0.771904	0.449161	0.965130	0.885214	0.449161
0.9	0.734935	0.314906	0.986407	0.932749	0.314906

➔ For lower cut-off value, sensitivity is very high which is given by the formula: **Sensitivity = True Positive / (True Positive + False Negative)**

➔ For lower cut-off value, False Negative records will be very less as most of the records will be categorized as potential leads.

➔ So, in case company wants to be more aggressive and had a more phone conversation, they should opt for the lower cut-off value like 0.2. For the test set, with cut-off 0.2, sensitivity value is close to 89% which means there is only around 11% chance of missing successfully converted lead.

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.**

➔ When company reaches its target and wants to only focus on minimizing the rate of useless phone calls, company could increase the cut-off value.

➔ On making the threshold value high, there is very less chance of selecting lead which is not converted.

➔ Below is the table for different cut-off values for the test set and its related different evaluation measures:

Cut-off	Accuracy	Sensitivity	Specificity	Precision	Recall
0.1	0.605915	0.967423	0.389480	0.486836	0.967423
0.2	0.757116	0.889437	0.677896	0.623098	0.889437
0.3	0.790388	0.836130	0.763002	0.678686	0.836130
0.4	0.809242	0.769990	0.832742	0.733772	0.769990
0.5	0.803697	0.663376	0.887707	0.779582	0.663376
0.6	0.800000	0.615005	0.910757	0.804910	0.615005
0.7	0.786322	0.536032	0.936170	0.834101	0.536032
0.8	0.771904	0.449161	0.965130	0.885214	0.449161
0.9	0.734935	0.314906	0.986407	0.932749	0.314906

➔ For higher cut-off value, specificity is very high which is given by the formula: **Specificity = True Negative / (True Negative + False Positive)**

➔ For higher cut-off value, False Positive records will be very less as only most promising lead for conversion is chosen.

➔ So, in case company want to have minimal useless phone calls, they should opt for the higher cut-off value like 0.8. For the test set, specificity value is close to 96.5% which means there are only 3.5% chances of selected lead is not converted and company makes useless phone calls.