SUBJECTIVE QUESTION AND ANSWERS

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

Ans:

The following are the top three variables in our model that have the greatest impact on the likelihood of a lead being converted:

Predictor Variable	Coefficient
Lead_Source_Welingak website	5.8439
Lead_Source_Welingak website	3.8622
Current Occupation_Working Professional	3.7716

We have the following two predictor variables in the model from non-categorical variables:

Do Not Email: -1.4877 (negatively related)

Total Time Spent on Website: 1.1434

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:

The top three categorical/dummy variables in the model on which the greatest emphasis should be placed in order to increase the likelihood of lead conversion are as follows:

Predictor Variable	Coefficient
Lead_Source_Welingak website	5.8439
Lead_Source_Welingak website	3.8622
Current Occupation_Working Professional	3.7716

As a result, leads with the origin "Welingak website" or "Reference" and who are "Working Professional" should be given greater attention in order to maximise the likelihood of lead conversion.

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:

Confusion matrix

	Predicted Negative	**Predicted Positive**
Actual Negative	TN (True Negative)	FP (False Positive)
Actual Positive	FN (False Negative)	TP (True Positive)

According to the aforementioned confusion matrix, if all Actual Positives are likewise forecasted as positives, i.e. a high True Positive (TP) rate, all potential consumers are correctly identified.

Sensitivity, often known as Recall, is a metric that shows true positive rate.

Sensitivity = (TP/(TP+FN))

As a result, to identify all potential leads, Sensitivity (Recall) should be high.

If we use a lower cutoff, all probability above it will be forecasted as 1 (i.e. TP will increase), and sensitivity will increase as well.

As a result, use a lower cutoff value to increase sensitivity. This will ensure that practically all leads who are likely to convert are appropriately identified, allowing agents to call as many of them as possible.

We found the following specificity values for different probability cutoffs in the Lead scoring case study:

Cutoff	Accuracy	Sensitivity	Specificity
0.0	0.385136	1.000000	0.000000
0.1	0.532042	0.985282	0.248143
0.2	0.695166	0.902698	0.565173
0.3	0.769643	0.829109	0.732394
0.4	0.801134	0.702371	0.862996
0.5	0.797355	0.643908	0.893470
0.6	0.780822	0.567866	0.914213
0.7	0.762400	0.477514	0.940845
0.8	0.735002	0.363451	0.967734
0.9	0.689340	0.203189	0.993854

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:
The requirement here is to accurately forecast potential leads, with very few incorrect predictions, i.e., extremely few false positives.

Confusion matrix

	Predicted Negative	Predicted Positive
Actual Negative	TN (True Negative)	FP (False Positive)
Actual Positive	FN (False Negative)	TP (True Positive)

False positive rate (FPR) is a metric that measures this:
FPR = (FP/(TN+FP))

FPR = 1-Specificity,
Whereas Specificity = (TN/(TN+FP))

As a result, we require a low FPR or a high specificity.

If we use a higher cutoff, only the probabilities above the cutoff will be forecasted as 1 (total predicted positives will be very low, i.e. FP will decrease and TN will increase), increasing specificity.

As a result, choose a higher cutoff value to increase the Specificity. This ensures that practically all leads who are on the verge of being converted are not chosen. As a result, agents will no longer be required to make useless phone calls.

We found the following specificity values for different probability cutoffs in the Lead scoring case study:

Cutoff	Accuracy	Sensitivity	Specificity
0.0	0.385136	1.000000	0.000000
0.1	0.532042	0.985282	0.248143
0.2	0.695166	0.902698	0.565173
0.3	0.769643	0.829109	0.732394
0.4	0.801134	0.702371	0.862996
0.5	0.797355	0.643908	0.893470
0.6	0.780822	0.567866	0.914213
0.7	0.762400	0.477514	0.940845
0.8	0.735002	0.363451	0.967734
0.9	0.689340	0.203189	0.993854