Subjective Questions and Answers

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

Approach 01	Approach 02
tags	ttime_on_site
last_activity	lead_origin_Lead Add Form
ttime_on_site	curr_occupation_Working Professional

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?

Approach 01	Approach 02
tags_Will revert after reading the email	lead_origin_Lead Add Form
tags_Closed by Horizzon	curr_occupation_working Professional
last_activity_SMS Sent	lead_profile_Potential Lead

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.

```
2  cutoff_df = pd.DataFrame( columns = ['prob','accuracy','recall', 'preci'])
3
4  num = [0.0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9]
5  for i in num:
6      cm1 = confusion_matrix(y_train_pred_final.Converted, y_train_pred_final[i])
7      total1=sum(sum(cm1))
8      accuracy = (cm1[0,0]+cm1[1,1])/total1
9
10     recall = cm1[0,0]/(cm1[0,0]+cm1[0,1])
11     preci = cm1[1,1]/(cm1[0,1]+cm1[1,1])
12     cutoff_df.loc[i] =[ i ,accuracy, recall, preci]
13 print(cutoff_df)
```

coring-case-study.ipynb (output) ⊗

```
           prob  accuracy  recall  preci
0.000 0.000       0.385   0.000   0.385
0.100 0.100       0.846   0.765   0.722
0.200 0.200       0.900   0.880   0.830
0.300 0.300       0.916   0.914   0.870
0.400 0.400       0.923   0.945   0.910
0.500 0.500       0.927   0.963   0.936
0.600 0.600       0.924   0.969   0.945
0.700 0.700       0.910   0.977   0.956
0.800 0.800       0.904   0.982   0.964
0.900 0.900       0.889   0.991   0.980
```

- Since they wish to make the lead conversion more aggressive
- Set the threshold such that output results in **Higher Recall score**,
- So the model preditcs more potential leads, with less accuracy.
- Recall is also referred as Sensitivity or True Positive Rate.
- Recall = TP/TP+FN
- **Recall:** Total number of correctly classified Positive points upon total number of positive points in the dataset

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.

```python
cutoff_df = pd.DataFrame( columns = ['prob','accuracy','recall', 'preci'])

num = [0.0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9]
for i in num:
    cm1 = confusion_matrix(y_train_pred_final.Converted, y_train_pred_final[i])
    total1=sum(sum(cm1))
    accuracy = (cm1[0,0]+cm1[1,1])/total1

    recall = cm1[0,0]/(cm1[0,0]+cm1[0,1])
    preci = cm1[1,1]/(cm1[0,1]+cm1[1,1])
    cutoff_df.loc[i] =[ i ,accuracy, recall, preci]
print(cutoff_df)
```

coring-case-study.ipynb (output) ⊗

	prob	accuracy	recall	preci
0.000	0.000	0.385	0.000	0.385
0.100	0.100	0.846	0.765	0.722
0.200	0.200	0.900	0.880	0.830
0.300	0.300	0.916	0.914	0.870
0.400	0.400	0.923	0.945	0.910
0.500	0.500	0.927	0.963	0.936
0.600	0.600	0.924	0.969	0.945
0.700	0.700	0.910	0.977	0.956
0.800	0.800	0.904	0.982	0.964
0.900	0.900	0.889	0.991	0.980

- As they want to minimize the rate of useless phone calls
- Set the cutoff value, such that the conversions have a **Higher precision** score,
- Precision = TP/TP+FP
- **Precision:** Out of total number of points predicted as positive what % of points is actually positive