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

   The features used to build the model have been represented below based on their importance in lead conversion as per their coefficient values.

   ```
   new_params.sort_values(ascending=False)

   Tags_Closed by Horizzon                              7.58
   Tags_Lost to EINS                                    6.65
   Lead Source_Welingak Website                         5.87
   Tags_Will revert after reading the email             4.99
   Last Activity_SMS Sent                               2.24
   Tags_Busy                                            0.59
   What matters most to you in choosing a course_Other  -0.73
   Last Notable Activity_Olark Chat Conversation        -1.65
   Last Notable Activity_Modified                       -2.01
   Tags_Ringing                                         -3.35
   Tags_switched off                                    -3.87
   dtype: float64
   ```

   As per the above diagram, the top 3 dummy variables that contribute most towards the probability of a lead getting converted are:

   1. **Tags_Closed by Horizzon**
   2. **Tags_Lost to EINS**
   3. **Lead Source_Welingak website**

   If we consider columns with out any dummy variables then the top 3 variables that contribute towards the probability of lead getting converted are:
   1. Tags
   2. Lead Source
   3. Lead Activity

2. What are the top 3 categorical/dummy variables in the model which get maximum focus in order to increase the probability of lead conversion?

   Answer:

   As per the above diagram, the top 3 categorical/dummy variables that contribute the most towards the probability of a lead getting converted are also :

   1. **Tags_Lost to EINS**
   2. **Tags_Closed by Horizzon**
   3. **Tags_Will revert after reading the email**

3. X Education has a period of 2 months every year during which they hire few 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:

   Here our requirement is for almost all potential leads we want make a phone call and try as much as possible.
   So here we want to increase prediction for a lead to be converted even if the lead in actual case is not converted.

   Sensitivity with respect to our model can be defined as the ratio of total number of actual Conversions correctly predicted to the total no of actual Conversions.

   Similarly, Specificity can be defined as the ratio of total no of actual non-Conversions correctly predicted to the total number of actual non-Conversions.
   Here in our case.

   When the probability thresholds are very low, the sensitivity is very high and specificity is very low. Similarly, for larger probability thresholds, the sensitivity values are very low but the specificity values are very high.

   High sensitivity implies that our model will correctly identify almost all leads who are likely to Convert. It will do that by over-estimating the Conversion likelihood, i.e. it will misclassify some non-Conversion cases as Conversions.

   Now, since X Education has more man-power for these 2 months and they wish to make the lead conversion more aggressive by wanting almost all of the potential leads, we can choose a lower threshold value for Conversion Probability.

   This will ensure the Sensitivity rating is very high which in turn will make sure almost all leads who are likely to Convert are identified correctly and the agents can make phone calls to as much of such people as possible.

   For our model, the below graph shows how the Sensitivity and Specificity rating changes with change in the threshold value:

```
# Let's plot accuracy sensitivity and specificity for various probabiliti
cutoff_df.plot.line(x='probability', y=['accuracy','sensitivity','specifi
plt.show()
```
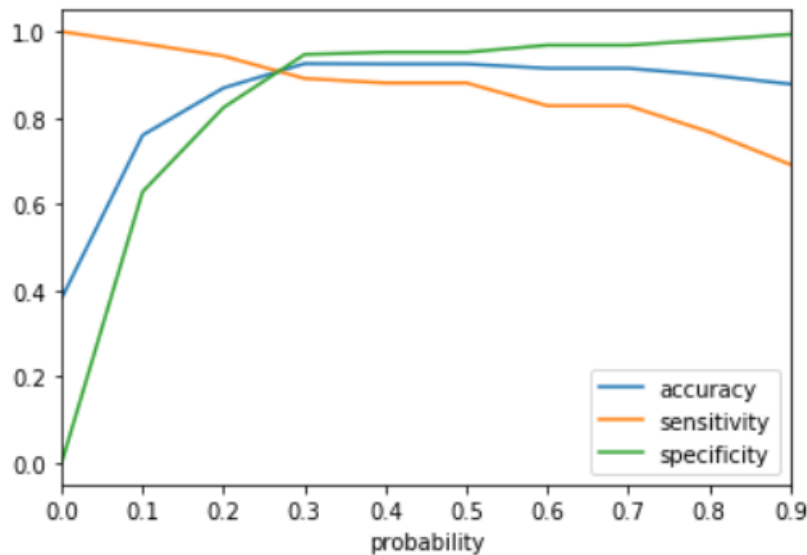


From the curve above, 0.28 is the optimum point to take it as a cutoff probability

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:
Here our requirement is just the opposite of the previous requirement.
So here we want to increase correct Positive prediction for a lead to be converted.

Following the similar logic and context from the previous question, High Specificity implies that our model will correctly identify almost all leads who are not likely to Convert. It will do that at the cost of losing out some low Conversion rate risky leads to the competition, i.e. it will misclassify some Conversion cases as non-Conversions.

Therefore, since X Education has already reached its target for a quarter and doesn't want to make phone calls unless it's extremely necessary, i.e. they want to minimize the rate of useless phone calls, we can choose a higher threshold value for Conversion Probability.

This will ensure the Specificity rating is very high, which in turn will make sure almost all leads who are on the brink of the probability of getting Converted or not are not selected. As a result the agents won't have to make unnecessary phone calls and can focus on some new work.