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

   Solution 1: Here we could be considering top 3 variables in 2 ways

   Top Three Variables from **Original List - *not considering the dummy variables***

   1. **Lead Source:** We see that lead sources - **Welingak Website** and **Reference** have high positive coefficients. This signifies that they contribute most towards getting a high probability to get lead converted

   2. **What is your current occupation:** Again we see that prospects with current occupation as **Working Professionals** are have high coefficient thus contributing towards higher probability and getting leads converted

   3. **Last Notable Activity:** We observe that Last Notable Activities – **SMS Sent, Email Opened** and Other Activities having high positive coefficient and contributing towards higher probability to get leads converted

   The Three variable in **final model – *including the dummy variables***

   1. **Lead Source_Welingak Website**: Coefficient = 6.48
      This is a variable created while creating dummy variables for categorical data. This refers to those leads which are sourced from Welingak Website.

      In Exploratory data analysis it was found that the leads where Lead Source is Welingak Website, we have a very high (more than 90%) Conversion rate.

   2. **Lead Source_Reference**: Coefficient 4.08
      Again this is a variable created while creating dummy variables for categorical data. This refers to those leads which are sourced from Reference.

      In Exploratory data analysis it was found that the leads where Lead Source is Reference, we have a very high Conversion rate.

   3. **What is your current occupation_Working Professional:** Coefficient 2.22
      This is a variable created while creating dummy variables for categorical data. This refers to those leads where the current occupation of prospect is Working Professional

      In EDA if was observed that leads with current occupation as Working Professional have a very high conversion Rate.

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?

Solution 2: The top three categorical / dummy variables are variable in final model which contribute most towards a higher probability of leads getting converted are

1. **Lead Source_Welingak Website**: Coefficient = 6.48
   This is a variable created while creating dummy variables for categorical data. This refers to those leads which are sourced from Welingak Website.

   In Exploratory data analysis it was found that the leads where Lead Source is Welingak Website, we have a very high (more than 90%) Conversion rate. Please refer to snapshot of the EDA done on **Lead Source**
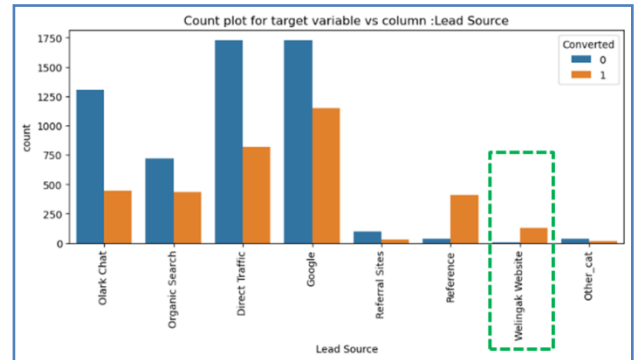
2. **Lead Source_Reference**: Coefficient 4.08
   Again this is a variable created while creating dummy variables for categorical data. This refers to those leads which are sourced from Reference.

   In Exploratory data analysis it was found that the leads where Lead Source is Reference, we have a very high Conversion rate. Please refer to snapshot from EDA done on **Lead Source**
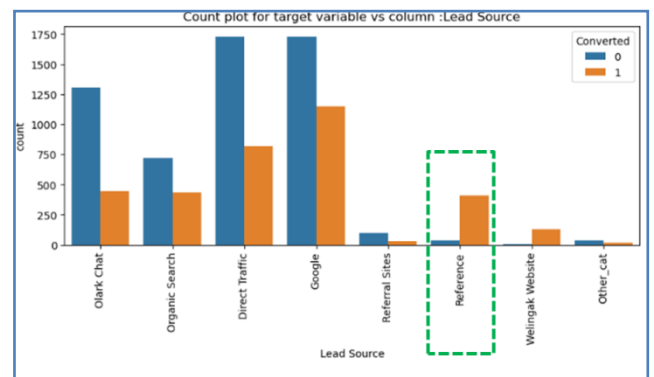
3. **What is your current occupation_Working Professional:** Coefficient 2.22
   This is a variable created while creating dummy variables for categorical data. This refers to those leads where the current occupation of prospect is Working Professional

   In EDA if was observed that leads with current occupation as Working Professional have a very high conversion Rate. Please refe to the attached snapshot from EDA on **What is your current occupation**
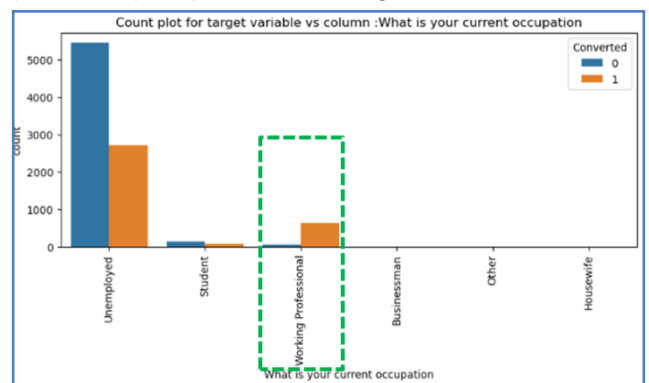
3. X Education has a period of 2 months every year during which they hire some interns. The sales team, in particular, has around 10 interns allotted to them. So during this phase, they wish to make the lead conversion more aggressive. So they want almost all of the potential leads (i.e. the customers 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.

Solution 3:
In order to ensure that no potential lead is missed out we could **lower the Probability cutoff to an optimal level which gives us best ROI**. This way we may be calling some prospects which will not get converted but we will ensure that we do not miss out on prospects which have a high chance of converting
Additionally, with 10 interns joining, thus we have more manpower to make calls and this would be a good strategy

By lowering the probability cutoff for this 2 month period we will be proactively reaching out to a wider set of leads and increase the chance of converting them.

However, we should keep in mind that we cannot drastically lower the probability as it will offer a very high number of potential leads and the return on investment could be diminished.

We should lower the probability keeping in mind what effort we can put, This can be combined with a modeling exercise on accuracy, Sensitivity and Specificity values to make a more calculated decision vs. a random decision.

This can be understood with the below ROC and Accuracy, Sensitivity and Specificity graphs to help us make an informed decision
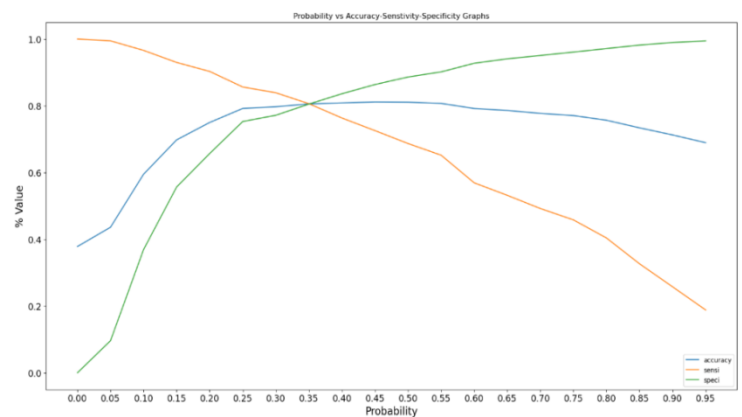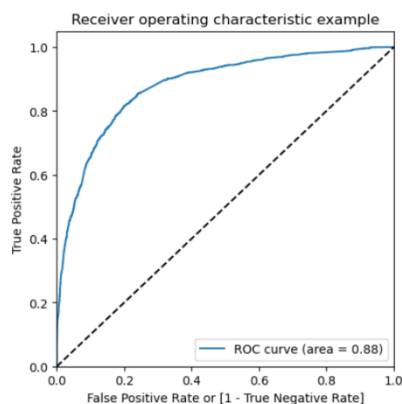
Considering **Sensitivity = True positive / (True Positive + False Negatives)**
It refers to our accuracy of correctly predicted conversions out of actual conversions. We see that sensitivity decreases after a certain cutoff. Thus **we need to lower the cutoff to the point where we are able to see the maximum sensitivity**.

Additionally, This can be combined with intelligent and proactive outreach keeping the important variables in mind. Example
- Reaching out to leads who have **higher probability scores first**
- Reaching out to **Working Professionals** first (very high conversion %)
- Focusing on leads from **Welingak Website and Reference**

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.

Solution 4:

As the Company wants sales team to focus on new work and want to make a phone call to extremely necessarily leads. Thus, they want to work on fewer pool of leads which have very high probability of conversion.

Here we can **increase the Probability cutoff to an optimal level which gives us best ROI.**

- Identify a cutoff probability which is higher and offers the amount of leads which can be effectively worked in the limited time

This can be effectively done by focusing on specificity using the below graph
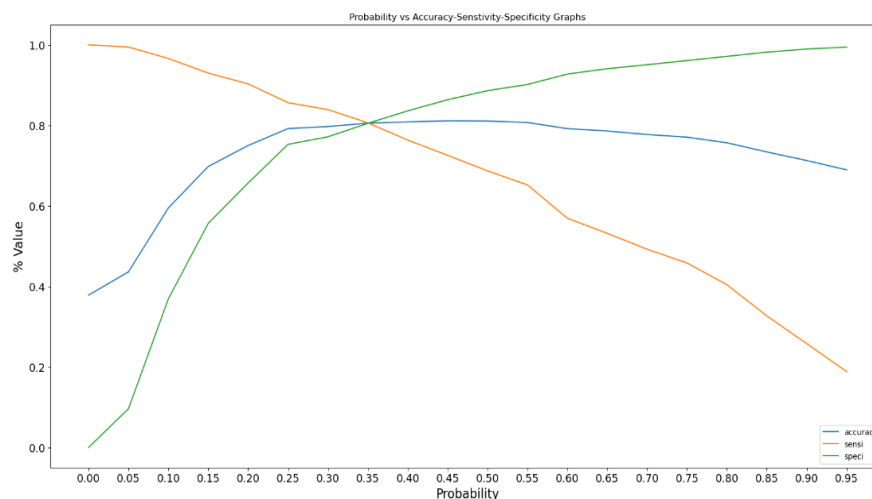
Specificity = True Negative / (True Negative + False Positive)

It measure that how accurate the model is in identifying negative instances (here non Conversion). If we increase the probability cutoff , we will more accuracy on identifying non Conversion cases and we will not call them.

Thus, we will have only more relevant leads which are predicted as conversions and we can be more efficient in reaching out to them in limited time.

Additionally, This can be combined with intelligent and proactive outreach keeping the important variables in mind. Example

- Reaching out to leads who have **higher probability scores first**
- Reaching out to **Working Professionals** first (very high conversion %)
- Focusing on leads from **Welingak Website and Reference**