Assignment Questions

Q1: Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

Top 3 variables that **positively** contribute towards the probability of a lead getting converted:

- Total Time Spent on Website
  - The most important predictor for the model
  - More time spent on website translates to higher chance of conversion
- Lead Origin
  - The encoded variable "Lead Origin_Lead Add Form" is the impactful variable
  - Leads originating from "Lead Add Form" are very likely to be converted
- Last Notable Activity
  - The encoded variable "Last Notable Activity_SMS Sent" is the impactful variable
  - Leads whose last notable activity was "SMS Sent" are very likely to be converted

Top 3 variables that **negatively** contribute towards the probability of a lead getting converted:

- Lead Quality
  - The encoded variable "Lead Quality_Worst" is the impactful variable
  - Leads which have been labelled as "Worst" are less likely to be converted
- What is your current occupation
  - The encoded variable "What is your current occupation_No Information" is the impactful variable
  - Leads that have not filled this field are less likely to convert
- Last Activity
  - The encoded variable "Last Activity_Olark Chat Conversation" is the impactful variable
  - Leads whose last activity was "Olark Chat Conversation" are unlikely to be converted

Q2: 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?

The top 3 categorical/dummy variables that should be focused on to **increase the probability of lead conversion**:

- Lead Origin_Lead Add Form
  - Leads originating from "Lead Add Form" *are very likely* to be converted
- Last Notable Activity_SMS Sent
  - If the lead's last notable activity was "SMS Sent", they *are likely* to be converted
- Lead Source_Olark Chat
  - Leads sourced from "Olark Chat" *are likely* to be converted

The top 3 categorical/dummy variables that should be focused on to **identify leads that should not be prioritized:**

- Lead Quality_Worst
    - Leads which have been labelled as "Worst" *are less likely* to be converted
- What is your current occupation_No Information
    - If the lead has not filled this field, they *are less likely* to convert
- Last Activity_Olark Chat Conversation
    - If the lead's last activity was "Olark Chat Conversation" *are unlikely* to be converted

Q3: 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.

To maximize lead conversion even if losing time by trying to convert a few non-convertible leads is allowed, we can:

- Lower the decision threshold of the model, currently the ideal threshold where Accuracy, Sensitivity and Specificity are equal is 0.34, we can lower this threshold to make sure that more leads can be identified as "Convertible", since we have additional manpower to handle these leads.
- Different leads will have different model confidence, a lead can have 0.9 and 0.6 as the model probability, we can further prioritize the leads have a higher model confidence score to spend more resources on them to make sure they convert. This can be done by binning the model confidence scores or by starting with the leads with the highest model confidence.

Q4: 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.

To only focus on leads that are almost a guaranteed conversion, we can:

- Raise the decision threshold of the model, currently the ideal threshold where Accuracy, Sensitivity and Specificity are equal is 0.34, we can raise this threshold to make sure that only the leads which are the most likely to convert are even being considered by the sales team.
- We can implement additional filters by using an additional model on top of the base model to filter out all but only the most valuable leads for conversion.
- Similar to the previous scenario, we can also start by contacting leads for which the model is more confident.