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

   To identify the top three variables contributing most towards the probability of a lead getting converted, we look at the variables with the highest absolute coefficients in the logistic regression model.

```
# Extract the coefficients and p-values from the final model
coefficients = lreg_model_7.params
p_values = lreg_model_7.pvalues

# Create a DataFrame for better visualization
summary_df = pd.DataFrame({
    'Variable': coefficients.index,
    'Coefficient': coefficients.values,
    'P-Value': p_values.values
})

# Order the variables by the absolute value of their coefficients
summary_df['Abs_Coefficient'] = summary_df['Coefficient'].abs()
summary_df = summary_df.sort_values(by='Abs_Coefficient', ascending=False)

# Select the top 3 variables
top_3_variables = summary_df.head(3).drop(columns='Abs_Coefficient')

print(top_3_variables)
```

```
                          Variable  Coefficient        P-Value
3      Total Time Spent on Website     3.979086  1.903845e-182
10  Occupation_Working Professional     3.814033   3.159241e-88
4                Lead Origin_Other     3.782189   5.201395e-93
```

   Top 3 Variables are:
   - Total Time Spent on Website
   - Occupation_Working Professional
   - Lead Origin_Other

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?

```
# Filter categorical variables (assuming they are encoded as dummy variables)
categorical_vars = summary_df[summary_df['Variable'].str.contains('_')]

# Order by absolute value of coefficients
categorical_vars = categorical_vars.sort_values(by='Abs_Coefficient', ascending=False)

# Select the top 3 categorical variables
top_3_categorical_vars = categorical_vars.head(3).drop(columns='Abs_Coefficient')

print(top_3_categorical_vars)
```

```
                          Variable  Coefficient       P-Value
10  Occupation_Working Professional     3.814033  3.159241e-88
4                Lead Origin_Other     3.782189  5.201395e-93
7                 Occupation_Other     1.586532  1.889389e-03
```

   Top 3 categorical /dummy variables are,
   - Occupation_Working Professional
   - Lead Origin_Other
   - Occupation_Other

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.

To make lead conversion more aggressive during the intern hiring period, the sales team should focus on leads that have a high probability of conversion. Here's a suggested strategy.

- **Lower the Probability Threshold**: Adjust the probability threshold for classifying leads as 'converted' to a lower value (e.g., from 0.5 to 0.3). This will increase the number of leads classified as potential conversions.
- **Prioritize High-Probability Leads**: Prioritize calling leads with the highest predicted probabilities first.
- **Use Top Predictors**: Focus on leads with high values in the top predictors (e.g., high Total Time Spent on Website, Lead Source_Google, etc.).

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.

   When the company wants to minimize the rate of useless phone calls, the sales team should focus on leads with the highest certainty of conversion. Here's a suggested strategy:

- **Raise the Probability Threshold**: Adjust the probability threshold for classifying leads as 'converted' to a higher value (e.g., from 0.5 to 0.7). This will reduce the number of leads classified as potential conversions, focusing only on the most likely ones.
- **Focus on High-Probability Leads**: Only call leads with the highest predicted probabilities.
- **Review Lead Scores**: Regularly review and update the lead scoring model to ensure it accurately reflects the most current data and trends.