Assignment Subjective Questions

Question 1:

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

Answer:

Top 3 Important Features:

Total Time Spent on Website: 0.287145

Last Notable Activity_SMS Sent: 0.111887

Lead Origin_Lead Add Form: 0.102962

Code Snippet for the obtained Answer:

```
# Retrieve the trained Random Forest model

rf_model = models["Random Forest"]["model"]

rf_model.set_params(**results["Random Forest"]["Best Params"])  # Set the best parameters

rf_model.fit(X_train, y_train)  # Refit the model on the training data


# Get feature importances

importance = rf_model.feature_importances_

important_features = pd.Series(importance, index=X.columns).sort_values(ascending=False)


# Display the top 3 important features

print("\nTop 3 Important Features:\n", important_features.head(3))
```

Question 2:

What are the top 3 categorical/dummy variables in the model which should be focused the most on to increase the probability of lead conversion?

Answer:

Top 3 Categorical/Dummy Variables:

Lead Source_Direct Traffic: 0.008027

Lead Source_Facebook: 0.000464

Lead Source_Google: 0.004314

Code Snippet for the obtained Answer:

```
# Retrieve the trained Random Forest model
rf_model = models["Random Forest"]["model"]
rf_model.set_params(**results["Random Forest"]["Best Params"])  # Set the best parameters
rf_model.fit(X_train, y_train)  # Refit the model on the training data

# Get feature importances
importance = rf_model.feature_importances_
important_features = pd.Series(importance, index=X.columns).sort_values(ascending=False)

# Filter for categorical/dummy variables
categorical_features = [col for col in X.columns if col.startswith('Lead Source_') or col.startswith('Last Activity_') or col.startswith('Other_Categorical_Columns')]  # Adjust based on your dataset
categorical_importance = important_features[categorical_features]
```

```
# Display the top 3 categorical/dummy variables

print("\nTop 3 Categorical/Dummy Variables:\n", categorical_importance.head(3))
```

Question 3:

X Education has a period of 2 months every year during which they hire some interns. The sales team 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 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:

Strategy for Aggressive Lead Conversion:

1. Focus on the top leads with the highest conversion probabilities.

2. Assign high-priority leads to experienced sales representatives.

3. Use interns to follow up on medium-priority leads.

Top 10 Leads for Follow-Up:

Lead ID	Score
6243	0.999989
4123	0.995666
6712	0.995327
7053	0.992116
8086	0.989924
3428	0.988605
6362	0.988450

```
1803        0.988338

267         0.988096

7327        0.986650
```

Code Snippet for the obtained Answer:

```
# Assuming `best_model` is the trained model and `X_test` is the test dataset
# Step 1: Predict probabilities for the test set
lead_scores = best_model.predict_proba(X_test)[:, 1]  # Probabilities for class 1 (converted)

# Step 2: Create a DataFrame with Lead IDs and their scores
# Use the index of the original DataFrame for Lead IDs
leads_with_scores = pd.DataFrame({'Lead ID': y_test.index, 'Score': lead_scores})

# Step 3: Sort leads by their scores in descending order
prioritized_leads = leads_with_scores.sort_values(by='Score', ascending=False)

# Step 4: Select the top leads for aggressive follow-up
top_leads = prioritized_leads.head(50)  # Adjust the number based on available resources

# Step 5: Print the strategy and top leads
print("Strategy for Aggressive Lead Conversion:")
print("1. Focus on the top leads with the highest conversion probabilities.")
print("2. Assign high-priority leads to experienced sales representatives.")
print("3. Use interns to follow up on medium-priority leads.")
print("\nTop Leads for Follow-Up:")
print(top_leads)
```

Question 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:

Strategy to Minimize Useless Phone Calls:

1. Focus only on leads with a conversion probability above 90%.

2. Use alternative communication methods (e.g., email, SMS) for lower-probability leads.

3. Prioritize high-value leads based on additional data.

Code Snippet for the obtained Answer:

```
# Filter leads with a high probability of conversion
high_threshold = 0.9
high_probability_leads = leads_with_scores[leads_with_scores['Score'] > high_threshold]

# Print the strategy and high-probability leads
print("Strategy to Minimize Useless Phone Calls:")
print("1. Focus only on leads with a conversion probability above 90%.")
print("2. Use alternative communication methods (e.g., email, SMS) for lower-probability leads.")
print("3. Prioritize high-value leads based on additional data.")
print("\nHigh-Probability Leads for Follow-Up:")
print(high_probability_leads)
```