Results

1. Logistic Regression (Top Features Tuned)

- **Accuracy:** 87.08%
- **Confusion Matrix:** True Positives = 493, True Negatives = 2728, False Positives = 349, False Negatives = 129.
- **Sensitivity (True Positive Rate):** 79.26%
- **Specificity (True Negative Rate):** 88.66%
- **Cross-Validation Mean Accuracy:** 87.60%
- **Cross-Validation Std Deviation:** 0.52%

2. Random Forest (Top Features Tuned)

- **Accuracy:** 89.19%
- **Confusion Matrix:** True Positives = 342, True Negatives = 2957, False Positives = 120, False Negatives = 280.
- **Sensitivity (True Positive Rate):** 55.00%
- **Specificity (True Negative Rate):** 96.10%
- **Cross-Validation Mean Accuracy:** 90.56%
- **Cross-Validation Std Deviation:** 0.34%

3. XGBoost (Top Features Tuned)

- **Accuracy:** 87.00%
- **Confusion Matrix:** True Positives = 499, True Negatives = 2719, False Positives = 358, False Negatives = 123.
- **Sensitivity (True Positive Rate):** 80.23%
- **Specificity (True Negative Rate):** 88.37%
- **Cross-Validation Mean Accuracy:** 87.70%
- **Cross-Validation Std Deviation**: 0.58%

Accuracy:
The Balanced Random Forest model has the highest accuracy (89.19%), followed closely by Logistic Regression (87.08%) and XGBoost (87.00%). This suggests that Random Forest is slightly better at correctly classifying both positive and negative instances.

Sensitivity:
XGBoost leads in sensitivity, correctly identifying 80.23% of actual positive cases, followed by Logistic Regression (79.26%) and Random Forest (55.00%). Higher sensitivity in XGBoost and Logistic Regression indicates better performance in identifying positive instances (purchases).

Specificity:
Random Forest excels in specificity with 96.10%, indicating its strength in correctly identifying negative instances (non-purchases). Logistic Regression and XGBoost have lower specificity scores (88.66% and 88.37%, respectively).

Cross-Validation Results:
The Random Forest model shows the highest mean cross-validation accuracy (90.56%) with the lowest standard deviation (0.34%), suggesting it is the most stable and generalizable model among the three. Logistic Regression and XGBoost show similar mean accuracies with slightly higher variability.

Conclusion:
The Balanced Random Forest model outperforms the others in overall accuracy and specificity. XGBoost and Logistic Regression is superior in sensitivity, making it a good choice if the priority is to capture as many potential purchasers as possible.