Summary Report

Objective

X Education aims to improve its lead conversion rate, which currently stands at 30%. By building a logistic regression model, the company seeks to assign lead scores to predict the likelihood of conversion. The primary goal is to identify high-potential leads ("Hot Leads") so that the sales team can focus on converting them, ultimately improving efficiency and achieving a target conversion rate of 80%.

Approach

1. **Data Preprocessing**:
    - Cleaned the dataset by handling missing values and removing irrelevant levels such as "Select" in categorical variables.
    - Standardized continuous variables and created dummy variables for categorical features.
2. **Model Development**:
    - A logistic regression model was built and optimized using feature selection techniques such as Recursive Feature Elimination (RFE) and statistical significance testing (p-values).
    - Performance metrics like accuracy, precision, recall, and the area under the ROC curve (AUC-ROC) were used to evaluate the model.
3. **Insights from the Model**:
    - **Top Predictive Variables**:
        - **Lead Origin_Lead Add Form**: Capturing lead-specific statuses, this variable showed thestrongest contribution to conversion prediction.
        - **Last Notable Activity_Unreachable**: Reflecting the latest interaction with the lead, itwas critical in assessing engagement levels.
        - **Lead Source**: Identified high-performing channels (e.g., referrals, Google ads) contributing to conversions.
    - **Key Dummy Variables**:
        - **Lead Origin_Lead Add Form**: Strongly associated with successfulconversions.
        - **Last Notable Activity_Unreachable**: Important to follow up with unreachable leads via different mode of communication.

- **Lead Source_Welingak Website**: Highlighted leads from this website requiring proactive follow-ups.

Recommendations

1. **Aggressive Conversion Phase**:
    - During periods with additional sales resources, prioritize leads predicted to convert with a probability near 1 (high-confidence predictions).
    - Adjust the probability threshold downward to ensure most potential leads are contacted, even at the cost of some False Positives.
2. **Minimizing Useless Phone Calls**:
    - In times of reduced focus on lead conversion, use the **Precision-Recall curve** to identify an optimal threshold to minimize False Positives.
    - Focus only on leads with extremely high conversion probabilities, reducing unnecessary outreach.
3. **Focus Areas**:
    - Increase engagement for leads tagged as "Will Revert After Reading the Email" with targeted, time-sensitive communication.
    - Analyze and address gaps identified for leads which are not reachable
    - Replicate strategies from successful cases tagged as "Closed by Horizzon" across similar lead segments.

Conclusion

The logistic regression model offers actionable insights into lead prioritization and conversion. By leveraging high-impact features like **Lead origin**, **Lead Source**, and **Last NotableActivity**, X Education can significantly improve its efficiency and conversion rate. Additionally, customized strategies for aggressive conversion and call minimization phases ensure the model adapts well to changing business needs.