**Lead Conversion Analysis - Summary Report**

**Introduction**

X Education, an online course provider, aimed to enhance its **lead conversion rate** by identifying high-potential leads, also known as **‘Hot Leads.’** The company's current conversion rate is **30%**, and the goal was to build a logistic regression model to **assign lead scores** (0-100), allowing the sales team to prioritize outreach efforts efficiently. This report summarizes the methodology, key findings, and business insights derived from the analysis.

**Methodology**

**1. Data Understanding & Preprocessing**

* The dataset contained **9000 records** with features like Lead Source, Total Time Spent on Website, Last Activity, etc.
* The target variable was **‘Converted’** (1 = Lead converted, 0 = Not converted).
* Handled missing values and removed irrelevant ‘Select’ responses.
* Converted categorical variables into **dummy variables** for model compatibility.
* Normalized numerical features to ensure uniform scaling.

**2. Model Selection & Training**

* Split the data into **70% training** and **30% testing sets**.
* Used **Recursive Feature Elimination (RFE)** and **p-values** to select the most impactful features.
* Trained a **logistic regression model**, which is well-suited for binary classification problems like lead conversion.
* Evaluated the model using **accuracy, precision, recall, F1-score, and AUC-ROC Curve**.

**Key Findings**

**1. Model Performance**

* **Accuracy: 94.01%** (indicating a high predictive capability).
* **Precision & Recall:** Balanced results ensuring minimal false positives and false negatives.
* **AUC-ROC Curve:** Demonstrated strong separability between converted and non-converted leads.

**2. Top Features Impacting Lead Conversion**

The three most significant predictors of lead conversion were:

1. **Tags\_Lost to EINS** (Coefficient: 6.51)
2. **Tags\_Closed by Horizzon** (Coefficient: 5.65)
3. **Tags\_Will revert after reading the email** (Coefficient: 4.98) These indicate that past lead interactions and status significantly impact conversion probability.

**3. Key Categorical Features to Focus On**

1. **Lead Source\_Welingak Website** (High conversion probability)
2. **Tags\_Unknown** (Needs further analysis)
3. **Tags\_switched off** (Negative impact, meaning such leads should be deprioritized)

**Business Insights & Strategies**

**1. Intern Hiring Phase (Aggressive Lead Conversion Strategy)**

* Assign interns to focus on **high-scoring leads**.
* Implement an **email-first approach** before making calls.
* Create urgency with **exclusive discounts** and limited-time offers.

**2. Minimizing Calls When Quarterly Target is Met**

* Filter out **low probability leads** and avoid unnecessary calls.
* Prioritize **recently engaged leads** through emails instead.
* Focus the sales team on **other business development tasks**.

**Conclusion & Recommendations**

✅ Prioritize **high-impact lead sources** (e.g., Welingak Website). ✅ Optimize **sales team outreach** by focusing on strong predictors. ✅ Implement **automated lead nurturing strategies**. ✅ Adapt strategies dynamically **based on business cycle needs**.

This study successfully built a robust logistic regression model to predict **lead conversion probabilities**, equipping X Education with data-driven strategies to enhance conversion efficiency. The findings and strategies outlined in this report should guide the company toward achieving its **80% conversion target** effectively.

**End of Report**