

# Lead Scoring Analysis Report

## 1. Objective

In today's competitive business landscape, lead conversion is essential for driving revenue. This analysis aims to create a predictive model that estimates the probability of a lead becoming a customer. By doing so, the sales team can strategically allocate resources, prioritize high-potential leads, and enhance overall conversion rates.

To achieve this, the study employs logistic regression to develop a lead scoring model that predicts conversion likelihood. The insights gained from this model will support data-driven decision-making regarding customer engagement strategies.

## 2. Data Preparation & Feature Selection

To build a reliable model, the dataset underwent thorough preprocessing before training the logistic regression model.

### 2.1 Data Splitting

- The dataset was split into training (70%) and test (30%) sets to assess the model's performance on new data.

### 2.2 Feature Selection

- Recursive Feature Elimination (RFE) was used to identify the 17 most relevant features.
- Variables with a p-value greater than 0.05 (statistically insignificant) and a Variance Inflation Factor (VIF) above 5 (indicating multicollinearity) were removed to improve model interpretability and prevent overfitting.

## 3. Model Performance & Evaluation Metrics

The trained logistic regression model was tested on the unseen dataset, and its performance was evaluated based on key metrics.

PREDICTED ACTUAL	NOT CONVERTED	CONVERTED
NOT CONVERTED	1582	105
CONVERTED	105	931

---

## 4. Key Findings

### 4.1 Model Performance & Evaluation Metrics

The trained logistic regression model was tested on unseen data, achieving the following results:

- **Accuracy:** 92.29% (Measures overall correctness of predictions)
- **Precision:** 89.86% (Proportion of predicted conversions that were actually converted)
- **Recall (Sensitivity):** 89.86% (Model's ability to identify actual conversions)
- **Specificity:** 93.78% (Model's ability to distinguish non-converting leads)
- **False Positive Rate (FPR):** 6.22% (Proportion of non-converting leads incorrectly predicted as conversions)

These results indicate that the model effectively differentiates between converted and non-converted leads.

### 4.2 Lead Score & Conversion Probability

- The lead scoring model achieves a **92% conversion rate on test data** compared to **95% on training data**, confirming that it generalizes well without overfitting.
- An **optimal probability cutoff of 0.4** was chosen to balance sensitivity and specificity.

### 4.3 Key Factors Influencing Lead Conversion

The top three factors affecting conversion probability are:

1. **Tags\_Lost to EINS** – Leads labeled as "Lost to EINS" are crucial for targeted follow-ups.
2. **Tags\_Closed by Horizon** – Leads marked as "Closed by Horizon" exhibit a strong intent to purchase.
3. **Lead Quality\_Worst** – This factor significantly reduces conversion likelihood, emphasizing the importance of a better lead qualification process.

These insights help refine the lead prioritization strategy for improved efficiency.

---

## 5. Business Recommendations

### 5.1 High Sales Periods (Intern Hiring Phase)

During the two-month intern hiring phase, additional interns support the sales team in driving lead conversion. The following strategy should be implemented:

- Focus on **high-probability leads** (Convert\_Probability > 0.4).
- Prioritize **phone calls** to leads with a **high Lead Score (> 70)**.
- Automate **email and SMS follow-ups** for medium-probability leads (Lead Score: 40–70).
- **Track and analyze** lead response rates to optimize call timing.

### 5.2 Low Sales Periods (After Achieving Targets Early)

Once quarterly sales targets are met, the sales team's focus should shift toward efficiency rather than aggressive conversions. Recommended approach:

- **Limit phone calls** to only high-scoring leads (**Lead Score > 80**).
- Shift focus to **customer retention, upselling, and market expansion**.
- Conduct **internal sales training and process optimizations**.

These strategies ensure optimal resource utilization across different business cycles.

---

## 6. Conclusion

The logistic regression model effectively predicts lead conversion with **high accuracy (92.29%) and precision (89.86%)**. By leveraging data-driven insights, the organization can:

- **Prioritize high-potential leads** for better sales outcomes.
- **Optimize resource allocation** during peak and off-peak periods.
- **Enhance lead nurturing strategies** to boost conversion rates.

This analysis provides a solid foundation for **strategic sales decisions**, ensuring a balance between efficiency and effectiveness in the lead conversion process.