# Lead Scoring Case Study – Summary Report

## 1. Introduction

X Education is an online course provider that generates leads through multiple channels such as website visits, advertisements, emails, and search engines. However, despite generating a large number of leads, only a small percentage convert into paying customers. The company aims to improve its lead conversion rate by identifying high-potential leads and prioritizing sales efforts efficiently.

The objective of this project is to build a lead scoring model that assigns a probability score to each lead, enabling the sales team to focus on leads that are most likely to convert.

## 2. Business Objective

The primary goals of this project are:

- To build a predictive model that estimates the probability of lead conversion
- To identify key factors influencing lead conversion
- To support business decision-making through data-driven lead prioritization
- To optimize sales calling strategies based on different business scenarios

## 3. Data Understanding and Preprocessing

The dataset contains demographic information, lead source details, and user activity-related features. The target variable indicates whether a lead was converted or not.

Key preprocessing steps included:

- Handling missing values appropriately
- Dropping irrelevant and high-missing-value columns
- Converting categorical variables into dummy variables
- Removing redundant dummy variables to avoid multicollinearity
- Ensuring overall data consistency

# 4. Exploratory Data Analysis (EDA)

Exploratory Data Analysis was performed to understand patterns and relationships within the data. Key insights include:

- Conversion rates vary significantly across different lead sources
- Google and Direct traffic contribute the highest number of leads
- Housewives and working professionals show higher conversion rates
- Leads with higher engagement, such as more time spent on the website, are more likely to convert
- The dataset shows class imbalance between converted and non-converted leads

EDA insights guided feature selection and model development.

# 5. Feature Selection and Multicollinearity

To ensure model stability and interpretability:

- Variance Inflation Factor (VIF) was used to detect multicollinearity
- Features with high VIF values were removed iteratively
- Statistical significance was checked using p-values
- Only relevant and interpretable predictors were retained

# 6. Model Building

A logistic regression model was chosen due to:

- High interpretability
- Suitability for probability-based prediction
- Alignment with business decision-making

The model was refined through multiple iterations by removing insignificant and highly correlated variables. A single, well-optimized model was used instead of multiple algorithms to maintain transparency and business relevance.

# 7. Model Evaluation

The model was evaluated using the following metrics:

- Accuracy
- Precision
- Recall
- ROC-AUC

The ROC curve stayed well above the diagonal, indicating strong discriminative power. Precision and recall were analyzed carefully to align model performance with business requirements.

# 8. Business Scenarios and Cutoff Strategy

## Scenario 1: Aggressive Calling Phase

- Objective: Maximize conversions
- Strategy: Use a lower probability cutoff
- Outcome: Higher recall, ensuring most potential leads are contacted

## Scenario 2: Conservative Phase

- Objective: Minimize unnecessary calls
- Strategy: Use a higher probability cutoff
- Outcome: Higher precision, focusing only on high-quality leads

This flexible cutoff strategy enables the business to adapt the model based on operational needs.

# 9. Conclusion

The project successfully developed an interpretable lead scoring model that helps X Education prioritize high-potential leads. By aligning model decisions with business objectives, the solution improves sales efficiency, reduces wasted effort, and supports data-driven decision-making. The model provides flexibility to adjust strategies based on changing business conditions, making it a practical and effective solution.