

# Summary Report: Boosting Lead Conversion with Logistic Regression

## Introduction:

We were tasked with enhancing lead conversion rates for X Education Company, aiming to increase the current 30% rate to an ambitious 80%. Our goal was twofold: to identify high-potential leads and refine the lead conversion process. This report outlines our approach, methodologies, and key findings from this project.

## Approach:

### a. Thorough Data Cleaning:

- It has begun with a comprehensive data cleaning process, addressing missing values, outliers, and irrelevant columns. We carefully transformed categorical variables like "Select" into null values, maintaining data integrity at an impressive 98%.

### b. Exploratory Data Analysis (EDA):

- EDA revealed important insights, including a data imbalance with a 39% conversion rate. We identified Landing Page Submissions, Google, and Direct Traffic as major sources of conversions. Key drivers were SMS Sent an Email Opened, along with influential factors such as Occupation and the time spent on the website.

### c. Data Preparation and Transformation:

- For data preparation, we converted binary categorical variables into numerical values (0 and 1) and created dummy variables for multi-category columns. We used a 70:30 train-test split and applied MinMax scaling to ensure compatibility with our model.

### d. Optimized Model Building:

- We selected features using Recursive Feature Elimination (RFE) and manually reduced variables with insignificant p-values, resulting in 11 key variables. These variables showed significant p-values and acceptable levels of multicollinearity, ensuring the stability of our model.

### e. Strategic Model Evaluation:

- We set a prediction threshold of 0.39 based on thorough evaluation metrics to align with the CEO's 80% target. The model demonstrated promising results with accuracy, sensitivity, and specificity values of 81%, 77%, and 83%, respectively.

## Key Recommendations and Insights:

- **Target High-Score Leads:** Focus on leads with higher scores to optimize resource allocation.
- **Personalized Communication:** Develop tailored communication scripts to increase engagement.
- **Enhance Customer Interaction:** Improve the customer portal experience to boost conversions.
- **Expand Strategies:** Utilize lead feedback and target working professionals for broader reach.

## Notable Learnings:

- **Collaborative Effort:** Teamwork enriched our analysis through diverse perspectives.
- **Data Integrity:** Rigorous data cleaning and preparation were crucial for accurate modelling.
- **Valuable EDA Insights:** Exploratory Data Analysis provided essential insights into conversion patterns.
- **Feature Importance:** Selecting key features enhanced model interpretability.
- **Effective Evaluation:** Understanding metrics enabled precise model assessment.
- **Actionable Strategies:** Recommendations were based on predictions, offering practical strategies for improvement.

## Conclusion:

Our collaborative effort effectively addressed X Education's lead conversion challenges. By deploying a logistic regression model, we successfully assigned lead scores to optimize resource allocation and improve conversion rates. Our approach highlighted the importance of data-driven decision-making, from cleaning and analysis to strategic modelling. With these insights, X Education is well-positioned to aim for a 78% lead conversion rate, aligning with the CEO's vision.