Summary Report: Enhancing Lead Conversion with Logistic Regression

Introduction:

Our team—Antara, Ashirwad, Anand, and Shubhendu—undertook the challenge of elevating X Education Company's lead conversion rates. We aimed to boost the current 30% rate to an ambitious 80%. Our mission was twofold: identify promising leads and refine the lead conversion process. This report outlines our systematic approach, methodologies, and pivotal takeaways from this assignment.

Approach:

1. Thorough Data Cleaning: We initiated with meticulous data cleaning. We tackled missing values, outliers, and irrelevant columns. Categorical variables like "Select" were tactically transformed to null values. This rigorous process retained data integrity at 98%.

2. Exploratory Data Analysis (EDA): EDA unveiled crucial insights. The data imbalance, with a 39% conversion rate, caught our attention. Landing Page Submissions, Google, and Direct Traffic emerged as dominant conversion sources. Key drivers included SMS Sent and Email Opened, alongside influential factors like Occupation and website engagement duration.

3. Data Preparation and Transformation: Our focus on data preparation led to binary categorical variables being mapped to numerical values (0 and 1). Dummy variables were created for multi-category columns. A 70:30 train-test split ratio was applied, and MinMax scaling was adopted for model compatibility.

4. Optimized Model Building: Model building commenced with feature selection through Recursive Feature Elimination (RFE). Manual feature reduction eliminated variables with insignificant p-values, resulting in 11 variables post RFE. These variables demonstrated significant p-values and acceptable multicollinearity, ensuring model stability.

5. Strategic Model Evaluation: A strategic model evaluation involved setting a threshold at 0.39 based on rigorous metrics. This aligned predictions with CEO's 80% target. Accuracy, sensitivity, and specificity metrics for the test dataset showed promising values: 81%, 77%, and 83%, respectively.

Key Recommendations and Insights:

- Our strategic recommendations emerged from model insights. We emphasized resource allocation by targeting high-score leads. Personalized communication scripts were suggested to enhance engagement, while customer portal interaction was highlighted to boost conversions. Exploring expansion strategies included leveraging lead feedback and targeting working professionals.

Notable Learnings:

- Collaborative Synergy: Team collaboration enriched our analysis with diverse perspectives.

- Data Integrity: Data cleaning and preparation laid a robust foundation for accurate modeling.

- Illuminating EDA: Exploratory Data Analysis provided critical insights into conversion patterns.

- Feature Significance: Selecting influential features contributed to model interpretability.

- Strategic Evaluation: Understanding metrics facilitated accurate model assessment.

- Actionable Insights: Recommendations were grounded in predictions, offering actionable strategies.

Conclusion:

Our collaborative effort addressed X Education's lead conversion challenges comprehensively. Deploying a logistic regression model allowed effective lead score assignment, optimizing resource allocation and conversion rates. Our learnings emphasized data-driven decision-making, from data cleaning and analysis to strategic modeling. Ultimately, our insights empower X Education to pursue an 78% lead conversion rate, aligning seamlessly with the CEO's vision.