Summary Report: Gaining higher lead conversion with Logistic Regression

Introduction:

Our objective was to enhance lead conversion rates for X Education Company, from the current low 30% rate to a high 80%. We were to identify high-potential leads and refine the lead conversion process. In this report, we show our approach, methodologies, and key findings from this project.

Approach:

1. Data Cleaning:

We did a detailed data cleaning process, addressing missing values, outliers, and irrelevant columns. We also transformed categorical variables like "Select" into null values, maintaining data integrity as a result we could retain 98% of data.

2. Exploratory Data Analysis (EDA):

By doing the EDA, we saw a data imbalance with a 39% conversion rate. We identified Landing Page Submissions, Google, and Direct Traffic as major sources of conversions. Important factors were SMSSent and Email Opened, along with important factors such as Occupation and the time spent on the website.

3. Data Preparation & Transformation:

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

4. Model Building:

Recursive Feature Elimination (RFE) was used to select features 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.

5. Model Evaluation:

A prediction threshold of 0.39 was set based on evaluation metrics so that we can go closer to target 80%. The model demonstrated promising results with accuracy, sensitivity, and specificity values of 81%, 77%, and 83%, respectively.

Recommendations and Findings:

- The company should focus on higher scores leads to optimize resource allocation.
- The concerned team should develop tailored communication dialogues to increase engagement.
- It should improve the customer portal experience so that conversions are high.
- It should target working professionals for better reach.

Key Takeaways:

- Our analysis were better due to diversified team effort.
- A thorough data cleaning was important so that the best model came out.
- Exploratory Data Analysis provided important insights into conversion patterns.
- Selecting key features improved model interpretability.
- Understanding metrics enabled precise model assessment.
- Recommendations were based on predictions, which can be considered as strategies for improvement.

Conclusion:

Our joint efforts successfully tackled X Education's lead conversion challenges. By implementing a logistic regression model, we were able to assign lead scores that optimized resource allocation and boosted conversion rates. This approach emphasized the value of data-driven decision-making, from data cleaning and analysis to strategic modeling. With these insights in hand, X Education is poised to achieve a 78% lead conversion rate, which is the target to achieve.