Summary Report of Group Assignment on Logistic Regression on Lead Scoring

In the realm of modern business strategies, lead scoring has emerged as a crucial technique to discern the potential of leads for conversion. This involves assigning lead scores, ranging from 0 to 100, which enable companies to distinguish between promising and less promising leads. As a tool for precise targeting, a higher lead score indicates a hotter lead with a higher likelihood of conversion, while a lower score suggests a colder lead is less likely to convert. We embarked on an insightful journey through various stages, culminating in the development of an effective logistic regression model for lead scoring. Here is a summary of our key learnings:

1. Business Problem Understanding:

Before delving into data analysis, a profound comprehension of the business problem was vital. We understood that the objective was to craft a logistic regression model capable of assigning meaningful lead scores. This would empower the company to prioritize resources, focusing on leads with higher scores to maximize conversion potential.

2. Exploratory Data Analysis (EDA):

EDA was an essential phase to unravel insights from our dataset. We delved into lead attributes, behaviors, and interactions. By visualizing data distributions, correlations, and identifying trends, we gained a comprehensive understanding of our data's nuances.

3. Data Preparation for Logistic Regression and Modelling:

The effectiveness of a model hinges on data quality. We undertook meticulous data preparation, creating the dummy variables, handling missing values, encoding categorical variables, and scaling features where necessary. This ensured that our input was pristine and suitable for the logistic regression algorithm.

4. Logistic Regression on Statemodel and Logistic Regression on SkLearn:

We employed two approaches to logistic regression: one using statistical packages and the other through the popular Sklearn library. Both methods enabled us to model the relationship between predictor variables and lead conversion probability, giving us diverse perspectives.

5. Logistic Regression Model Building and Evaluation:

Constructing the logistic regression model was a pivotal step. Through an iterative process, we fine-tuned hyper-parameters with p-values and VIF score and selected features with RFE, and achieved convergence. Evaluating the model was equally critical, involving metrics such as accuracy, precision, recall, and F1-score. We assessed model performance on both training and test data to ensure its generalizability.

6. Concluding with a Brief Summary for the Customer:

In the concluding phase, we summarized our findings for the customer. We highlighted the essence of the model's performance, showcasing accuracy and its ability to distinguish between lead categories effectively. We offered insights into precision, recall, and F1-score, demonstrating our model's ability to classify leads accurately and minimize false positives/negatives. The brief summary showcased the practical utility of our logistic regression model in lead scoring and its potential application in lead targeting efforts.

In summary, our journey encompassed critical stages of business problem understanding, data analysis, data preparation, modelling, and evaluation. Through logistic regression, we forged a robust model for lead scoring, enabling the company to identify the hottest leads and allocate resources judiciously. Our comprehensive approach ensured the model's accuracy, reliability, and applicability. As the business landscape continues to evolve, our proficiency in logistic regression equips us to tackle challenges and harness opportunities through strategic lead scoring.