SUMMARY REPORT

In this assignment, I undertook the task of predicting the probability of lead conversion, following a systematic approach involving data preprocessing, model building, and analysis of feature importance. Here's a summary of how I proceeded and the key learnings gathered along the way.

Data Import and Inspection: I started by importing the lead data and performing an initial inspection. This step allowed me to understand the dataset's structure, the types of variables present, and the general characteristics of the data.

Data Preparation: During the data preparation phase, I focused on converting categorical variables into numerical format using dummy variables. This transformation enabled the model to work effectively with categorical features. Additionally, I conducted outlier analysis to identify extreme values that might affect the model's performance. Handling missing values through imputation was also an integral part of this phase.

Feature Selection and Scaling: To improve model efficiency and interpretability, I applied Recursive Feature Elimination (RFE) to select the most influential features. This step helped streamline the model by retaining the most relevant predictors. Feature scaling was also carried out to ensure that all features were on a similar scale, preventing any dominance by variables with larger magnitudes.

Correlation Analysis and Model Building: Exploring correlations between features helped uncover potential relationships that could impact lead conversion. With a solid understanding of the data, I proceeded to build a predictive model. I considered various algorithms such as logistic regression, decision trees, random forests, and gradient boosting, selecting the most appropriate one based on its performance.

Performance Evaluation and Optimization: After training the model on the training data, I evaluated its performance using an ROC curve to visualize the trade-off between sensitivity and specificity. The optimal cutoff point was determined to achieve a balance between true positives and false positives. This step allowed for more informed decision-making in lead prioritization.

Strategies for Conversion Phases: One of the most insightful aspects of this assignment was formulating strategies for two different lead conversion phases. During the aggressive conversion phase, I recommended focusing on potential leads predicted by the model, prioritizing direct calls to those not open to email communication. High

engagement indicators like "Total Time Spent on Website" and "Page Views Per Visit" were identified as crucial factors for targeted outreach.

Conversely, during the phase of achieving targets ahead of schedule, a conservative approach was advised. The emphasis shifted to selective phone calls, focusing on leads demonstrating strong conversion potential through recent and interactive activities like "Last Notable Activity," "Last Activity," and "Asymmetrique Activity Index."

Key Learnings: Through this assignment, I gained a deeper understanding of the end-to-end process of predictive modeling for lead conversion. The importance of data preprocessing, including variable conversion and handling outliers, was reinforced. The significance of feature selection and scaling in building efficient models became apparent, enhancing the model's performance and interpretability.

Furthermore, I realized the strategic value of tailoring outreach based on different business contexts. Crafting specific approaches for aggressive conversion and target-fulfilled phases highlighted the importance of resource optimization and effective engagement with potential leads.

In conclusion, this assignment provided hands-on experience in dealing with real-world data for predictive modeling, emphasizing the iterative nature of data analysis and the critical role of informed decision-making based on model insights. The learnings garnered will undoubtedly prove valuable in approaching similar business challenges, where data-driven strategies can drive lead conversion and sales success.