This report outlines the development and implementation of a Logistic Regression model designed to assign a lead score between 0 and 100. The model aims to classify leads based on their likelihood of conversion, with higher scores indicating a greater probability of conversion. This scoring system will assist the company in prioritizing and targeting potential leads more effectively.

Approach-

Problem Statement- The goal was to build a Logistic Regression model that assigns a numerical score (0 to 100) to each lead, reflecting the probability of conversion. A higher score denotes a "hot" lead with a high likelihood of converting, while a lower score represents a "cold" lead with a lower probability of conversion.

Data Collection: Data was sourced from historical lead records.

Data Preparation:

- **Data Cleaning:** Addressed missing values, removed duplicates, and ensured consistency. **Normalization:** Scaled numerical features to ensure that all features contributed equally to the model.

Model Development:

- **Data Splitting:** Divided the dataset into training (70%) and testing (30%) sets to build and validate the model
- **Model Training:** Applied Logistic Regression to the training set. This technique was selected for its effectiveness in binary classification and ease of interpretation.
- **Probability Calibration:** The output of the Logistic Regression model is a probability score, which was scaled to a 0-100 range for practical use in lead scoring.

Model Evaluation:

- **Performance Metrics:** Evaluated the model using metrics such as accuracy, precision, recall, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) to ensure robust performance.
- **Scoring Validation:** Tested the model on the test set to verify its ability to correctly score leads and predict conversion likelihood.

**Implementation:**

- **Scoring Leads:** Used the model to generate a lead score for each new lead, converting the predicted probabilities into a 0-100 score.

- **Prioritization:** Implemented a system where leads are prioritized based on their scores, allowing sales and marketing teams to focus on leads with the highest potential for conversion.

The Logistic Regression model successfully provided a lead scoring system that ranges from 0 to 100, reflecting the likelihood of lead conversion. This model aids in prioritizing leads, with higher scores indicating hotter leads. The project highlighted the importance of effective data preparation, model calibration, and continuous validation to ensure the model's practical utility and accuracy. Overall, this scoring system represents a valuable tool for optimizing lead management and boosting sales efficiency.