Summary Report

1. Problem Statement and Objective:

 - Improve lead conversion rate from 30% to the CEO's target of 80%.

 - Build a model to assign lead scores, prioritizing leads with higher conversion chances.

2. Data Cleaning:

 - Drop columns with more than 40% null values.

 - Impute categorical data using appropriate methods.

 - Treat outliers, fix invalid data, group low-frequency values, and map binary categorical values.

3. Exploratory Data Analysis (EDA):

 - Check data imbalance.

 - Perform univariate and bivariate analysis for categorical and numerical variables.

 - Identify variables with a significant effect on the target variable.

4. Data Preparation:

 - Create dummy features for categorical variables.

 - Split data into train and test sets.

 - Standardize features and drop highly correlated columns.

5. Model Building:

 - Reduce the number of variables using recursive feature elimination (RFE) and manual feature reduction.

 - Build three models before arriving at the final model (logm4).

 - Ensure stability with p-values < 0.05 and no multicollinearity (VIF < 5).

6. Model Evaluation:

 - Create a confusion matrix.

 - Select a cut-off point of 0.345 based on accuracy, sensitivity, and specificity plot.

 - Assign lead scores to train data using the chosen cut-off.

7. Top Features and Recommendations:

   - Top three features in the final model: Lead Source_Welingak Website, Lead Source_Reference, and Current_occupation_Working Professional.

   - Recommendations include allocating more budget to Welingak Website advertising, providing incentives for reference leads, and targeting working professionals aggressively.

8. Conclusion:

   - Hands-on experience gained in various aspects of the data science process.

   - Importance of selecting appropriate evaluation metrics based on the business problem.

   - Understanding trade-offs between different metrics.

   - Insights into factors affecting lead conversion rates and recommended strategies for improvement.

This comprehensive summary provides a clear overview of the project's objectives, methodologies, findings, and recommendations.