**Summary of Credit EDA Case Study**

This analysis, conducted for a consumer finance company, focuses on using Exploratory Data Analysis (EDA) to mitigate the risk of loan defaults and improve lending decisions. The primary objective is to understand the factors influencing loan repayment difficulties and to ensure that capable applicants are not unjustly rejected.

1. **DATA CLEANING:**
   * **Missing Values:** Addressed missing values in multiple columns.
   * **Columns Dropped:** Columns with more than 25% null values were removed.
   * **Outlier Treatment:** Managed outliers and fixed invalid data.
   * **Data Mapping:** Binary categorical values were mapped, and low-frequency values were grouped.
2. **EDA:**
   * **Univariate Analysis:** Conducted for categorical and numerical variables. Variables like 'income', 'loan amount', and 'age' provided significant insights into loan default tendencies.
   * **Bivariate Analysis:** Explored relationships between pairs of variables to understand their combined impact on loan default.
   * **Multivariate Analysis:** Analyzed multiple variables together to identify complex interactions and patterns.
3. **DATA PREPARATION:**
   * **Feature Engineering:** Created dummy features (one-hot encoding) for categorical variables.
   * **Train-Test Split:** Data was split into training and testing sets in a 70:30 ratio.
   * **Feature Scaling:** Applied standardization to numerical features.
   * **Correlation Management:** Highly correlated features were dropped to prevent multicollinearity.
4. **MODEL BUILDING:**
   * **Feature Selection:** Reduced the number of features using Recursive Feature Elimination (RFE).
   * **Manual Feature Reduction:** Built models by iteratively removing variables with high p-values.
   * **Model Stability:** Ensured stability in the final model with no multicollinearity (VIF < 5).
5. **MODEL EVALUATION:**
   * **Confusion Matrix:** Used to evaluate model performance. Selected a cut-off point of 0.37 based on accuracy, sensitivity, and specificity.
   * **Performance Metrics:** Achieved balanced metrics around 80% for accuracy, specificity, and precision.
   * **Lead Scoring:** Assigned scores to training data using the optimal cut-off point.

**RECOMMENDATIONS:** To reduce loan default risks, the company should focus on the following key factors:

1. **Income Level:** Pay special attention to lower-income applicants who have a higher default risk.
2. **Loan Amount and Annuity:** Carefully evaluate the relationship between requested loan amounts and annuity payments.
3. **Applicant Age:** Younger applicants under 30 and older applicants above 60 should be scrutinized due to varying credit utilization behaviors.
4. **Demographic Attributes:** Gender, occupation, and family status significantly influence repayment capabilities and should be considered in the risk assessment.