CREDIT EDA ASSIGNMENT

BY: SURYA VARDHAN REDDY

Problem Statement

The purpose is to evaluate loan application data to discover indicators that increase the chance of loan default. This research attempts to assist a consumer financing organization in making

educated judgments when accepting or rejecting loan applications, hence reducing financial risk.

Step 1: Data understanding

Objective : Use the data dictionary (columns_description.csv) to become familiar with the dataset and grasp the significance of each variable.
Action : Load and examine the data, identifying relevant factors and understanding their importance in the context of loan defaults.

Step 2 Data Cleaning

Objective : Address any data quality concerns, including missing values and outliers.
Action: Determine a suitable handling approach for missing data (for example, imputation or column removal). In addition, identify outliers and describe their possible influence on the analysis.

Step 3: Exploratory Data Analysis (EDA).

Objective : Use univariate and bivariate analysis to identify patterns and correlations between variables.
Action: Univariate Analysis: Examine individual variables' distributions.
Bivariate Analysis: Investigate the links between variables, particularly how they relate to the target variable (loan default).
Derived measures: Develop new measures that better reflect the applicant's risk profile.

Step 4: Handling Data Imbalance.

Objective: Correct any imbalances in the dataset, particularly those involving the target variable.

Action: Use approaches such as resampling or weighted analysis to guarantee that the model is not influenced by an imbalance.

Step 5: Correlation Analysis.

Objective: Determine the strongest connections across variables to identify significant causes of loan default.
Action: section data based on the target variable and identify the strongest relationships for each section.

Missing Data Handling

- Rundown of Missing Information Taking care of

- 1. Find out what's missing: Utilized isnull() to examine the dataset for missing values and determine the percentage of missing values in each column.
- 2. Choose an Approach: High Level of Missing Qualities: Sections with a high level of missing qualities (e.g., >50%) were considered for expulsion. Low Level of Missing Qualities: For sections with a lower level of missing information, attribution procedures were utilized: Mathematical Sections: Ascribed missing qualities with the middle or mean, contingent upon the conveyance. Straight out Sections: Credited missing qualities with the mode or another class demonstrating missing information.
- 3. The Missing Data Handling Summary: Sections Eliminated: Segments with unreasonable missing information were taken out to try not to present predisposition. Imputation: Appropriate imputation techniques were used to address data quality issues and retain as much information as possible.

Conclusion

1. The research focuses on identifying the factors that drive loan defaults, which is critical for a consumer financing firm looking to avoid financial risks.
2. The primary purpose was to discover characteristics that signal a higher risk of default, allowing for more accurate loan approval choices.
3. It was noted that applicants with lower salaries and bigger loan amounts were more likely to default, emphasizing the significance of aligning loan amounts to income.
4. It Showed that younger applicants with bigger loans were more likely to default, emphasizing the necessity for stronger lending standards for younger applicants with high loan needs.

Income level and loan amount are critical factors in determining default risk. This suggests that the company should consider adjusting loan amounts based on income to mitigate risk.

Age and Loan Terms: Younger applicants, especially those with larger loans, are more likely to default.

Offering shorter loan terms or smaller loan amounts to younger borrowers could reduce the likelihood of default.

Credit History and Risk: A thorough review of applicants' credit history and financial behaviors is essential to accurately assess their default risk.

Ensuring that only applicants with stable financial backgrounds receive loans could minimize defaults.

I. The study successfully used Exploratory Data Analysis (EDA) to get valuable insights into the factors that influence loan defaults.
II. The complete strategy, which included univariate, segmented univariate, and bivariate analysis, enabled a thorough knowledge of the links between consumer variables, loan attributes, and default risk.
III. These insights will help the organization improve its loan approval process and ensure that creditworthy candidates are accepted while high-risk applicants are properly managed.

IV. By applying the suggestions from this research, the firm may improve its lending strategy, lower default rates, and maintain a healthy loan portfolio.

Thank you