Presentation title

Akshar Patel

Business Understanding:-

When a company receives a loan application, it must assess the applicant's profile to decide whether to approve the loan. This decision involves considering two types of risks:

1. If the applicant is deemed capable of repaying the loan, rejecting the application means losing potential business for the company.
2. Conversely, if the applicant is unlikely to repay the loan (i.e., they may default), approving the loan could result in financial loss for the company.

There are four possible scenarios that occur when a loan is applied for:

- Approval of the loan
- Cancellation of the loan
- Refusal of the loan Offer
- remains unused

Problem Statement

Consumer finance companies face challenges in accurately assessing the risk of loan defaults, which can lead to significant financial losses. In order to make informed lending decisions and minimize risks, it is crucial to identify patterns in loan application data that can serve as indicators of potential defaults.

Objective:

This presentation aims to analyse a dataset of loan applications to identify key patterns and trends that correlate with loan defaults. By understanding these patterns, we can uncover insights that help in better risk assessment and decision-making processes for loan approvals.

We used 4 prior techniques/steps for solving this problem

01

Data cleaning

02

Handling outliers

03

binning

04

Data analysis including univariate, bivariate & multivariate.

Step-wise Process:-

1. Library Import: We began by importing the necessary libraries.
2. Data Cleaning: We initiated the data cleaning process by examining information and identifying null values in the dataframe using the ‘isnull()’ function.
3. Merge Data: After cleaning both datasets, we merged them to identify outliers. Outliers were detected in the dataset.
4. Outlier Removal: We employed three primary techniques—mean, median, and mode—to eliminate outliers. Additionally, we utilized binning processes for certain columns to categorize data into bins and stored them in new columns.
5. Analysis and Conclusion: We proceeded to analyse the data and draw conclusions, utilizing three main steps:

- Univariate Analysis: Examining individual variables to understand their distributions and characteristics.
- Bivariate Analysis: Analysing the relationships between pairs of variables to uncover potential correlations or patterns.
- Multivariate Analysis: Exploring interactions among multiple variables to gain deeper insights into complex relationships.

This structured approach ensured a thorough examination of the data and facilitated the derivation of meaningful conclusions.

Data cleaning

- We commenced the data cleaning process by examining the dataset's information and deriving statistical summaries using the describe function.
- We identified and managed missing values by dropping columns containing more than 45% missing data.
- Additionally, columns that were considered unnecessary or had minimal impact on the target variables were eliminated.
- After addressing missing values, we proceeded to impute the remaining missing values using appropriate methods.
- We ensured that the data types of columns were correct and aligned with their respective values to maintain data integrity.
- Any negative values present in the dataset were rectified as necessary to ensure consistency and accuracy.
- The same data cleaning process was replicated for the previous dataset to maintain consistency in our analyses.

Merge data

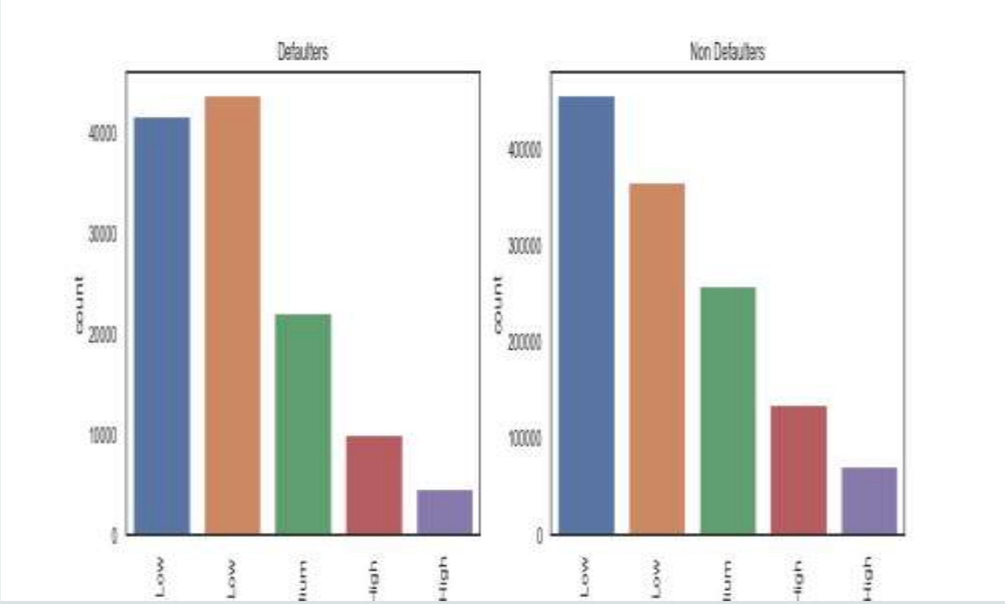
- We merged the clean datasets using an inner join to combine both datasets into one, ensuring that both sets of fields were included in a single column.
- Following the merge, we utilized the drop function to remove columns predominantly populated by 0s and 1s, with the exception of the target column as, it is containing essential data on defaulters and non-defaulters.
- After dropping columns with predominant 0s and 1s, we discovered 37 columns with 0 values that had minimal impact on the dataset.
- Among the remaining 64 columns, we conducted outlier detection to assess and manage any outliers present in the data..

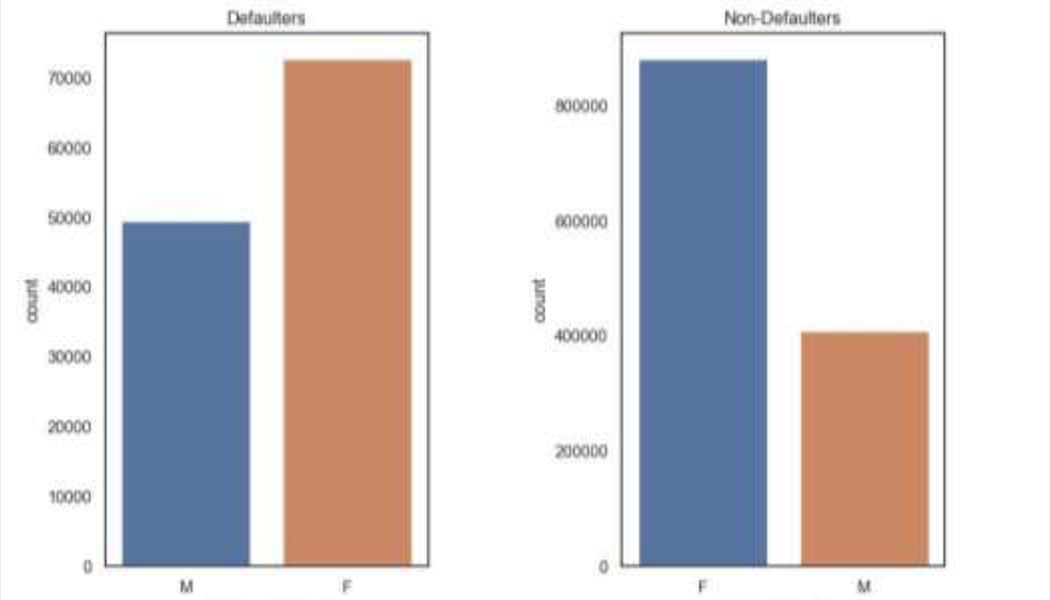
Handling Outliers

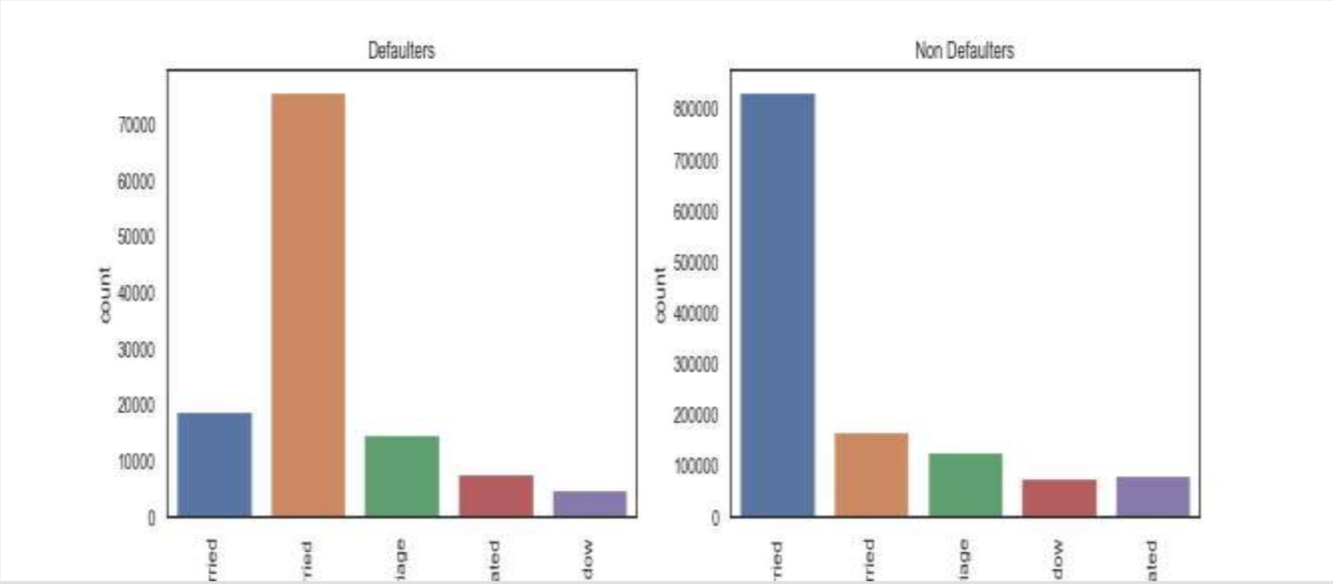
- We identified integer columns with significant outlier presence and addressed them by applying appropriate outlier treatment methods such as mean, median, and mode.
- Following outlier handling, we implemented the binning method for selected columns to categorize data into distinct bins. For instance, we utilized the birthday column to calculate age-based bins and categorized income and credit into "high," "low," and "very high" income categories.
- We examined the TARGET column to assess the distribution of defaulters and non-defaulters within the dataset. This column served as a crucial component for subsequent analysis.
- To enhance visualization and analysis, we partitioned the dataset into two subsets based on the TARGET column, distinguishing between defaulters and non-defaulters. This segmentation facilitated a clearer understanding of the data and its implications.

ANALYSIS

Single-variate analysis:-







Summary:-

- Comparison of Credit Category and Gender with TARGET Column:

The image displays credit category information on the left and gender information on the right, both compared with the TARGET column to identify defaulters and non-defaulters.

- Analysis of Credit Category:

Individuals with very low and low credit amounts represent a significant portion of defaulters. Similarly, non-defaulters also show a notable presence among those with very low credit amounts.

- Analysis of Gender:

Among defaulters, approximately 72,836 are female and 49,524 are male. Non-defaulters consist of approximately 882,358 females and 408,983 males. This suggests that males have a higher tendency for loan default compared to females.

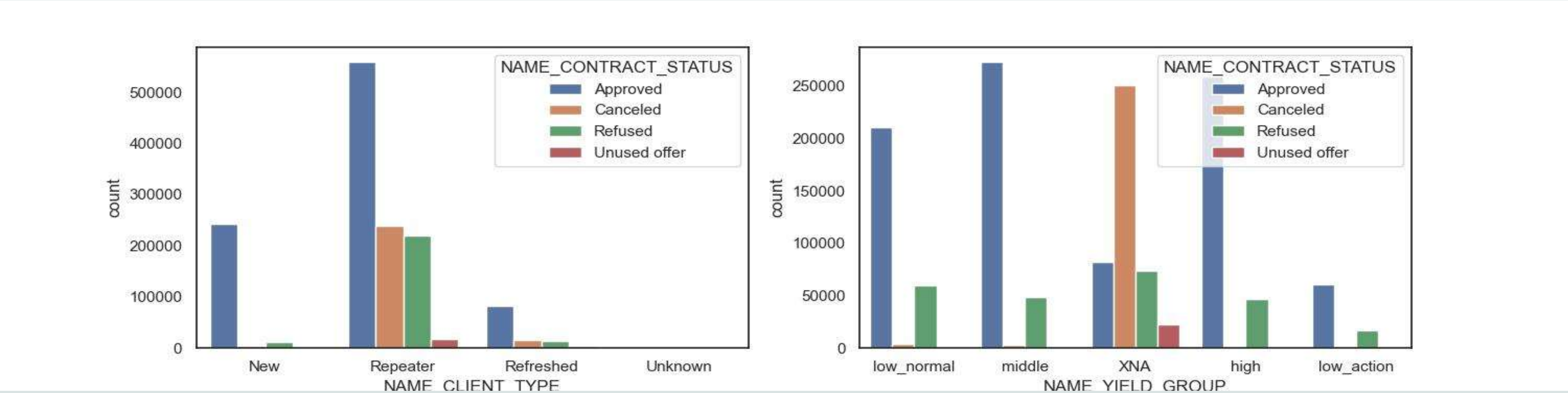
- Family Status Analysis:

The central graph illustrates family status information. Defaulters and non-defaulters are predominantly married. This trend is consistent across both categories, suggesting that family status plays a significant role, likely influenced by total income.

- Iteration with TARGET Column:

All columns are iterated with the TARGET column to accurately identify defaulters and non- defaulters, providing valuable insights into the dataset.

ANALYSIS
Bivariate analysis:-



•The plots reveal that individuals who have made multiple loan applications tend to have their loans approved, with the highest count of approved loans observed among this group.

•Among the various yield groups, those categorized as "middle" show the highest frequency of approved loans, closely followed by individuals in the "high" yield group.

conclusion:-

•Both datasets reveal that cash loans are more prevalent in current applications, whereas consumer loans were predominant in previous applications.

• In the non-defaulters there are businessmen and students which are not there in defaulters, so this indicates that banks trust businessmen and students in giving loans.

• A notable trend indicates that men have a higher likelihood of defaulting on loans compared to females, based on the analysis of both datasets.

• Also, the people who have more income get more credit and people with more income are less in defaulters.

• Pensioners emerge as a reliable option for banks when considering loan approvals, indicating a level of trust associated with this demographic group.

• Women on maternity leave are identified as the topmost defaulters, signalling a need for caution by banks when lending to individuals in this category.

•A strong positive correlation is observed between goods price and credit, suggesting a close relationship between these two variables.

• it's noticeable that males aged 50-60 and 40-50 typically opt for higher credit amounts and are also more likely to default. Likewise, among females, those in the age brackets of 50-60 and 40-50 exhibit higher credit utilization, followed by older individuals (>60) and young adults (30-40). The demographic of young adult males and females seems to pose lower default risks, making them potentially safer lending prospects for banks. Nonetheless, comprehensive checks and background assessments are still necessary in this regard.

Thank you

Akshar Patel

patelakshar2025@gmail.com