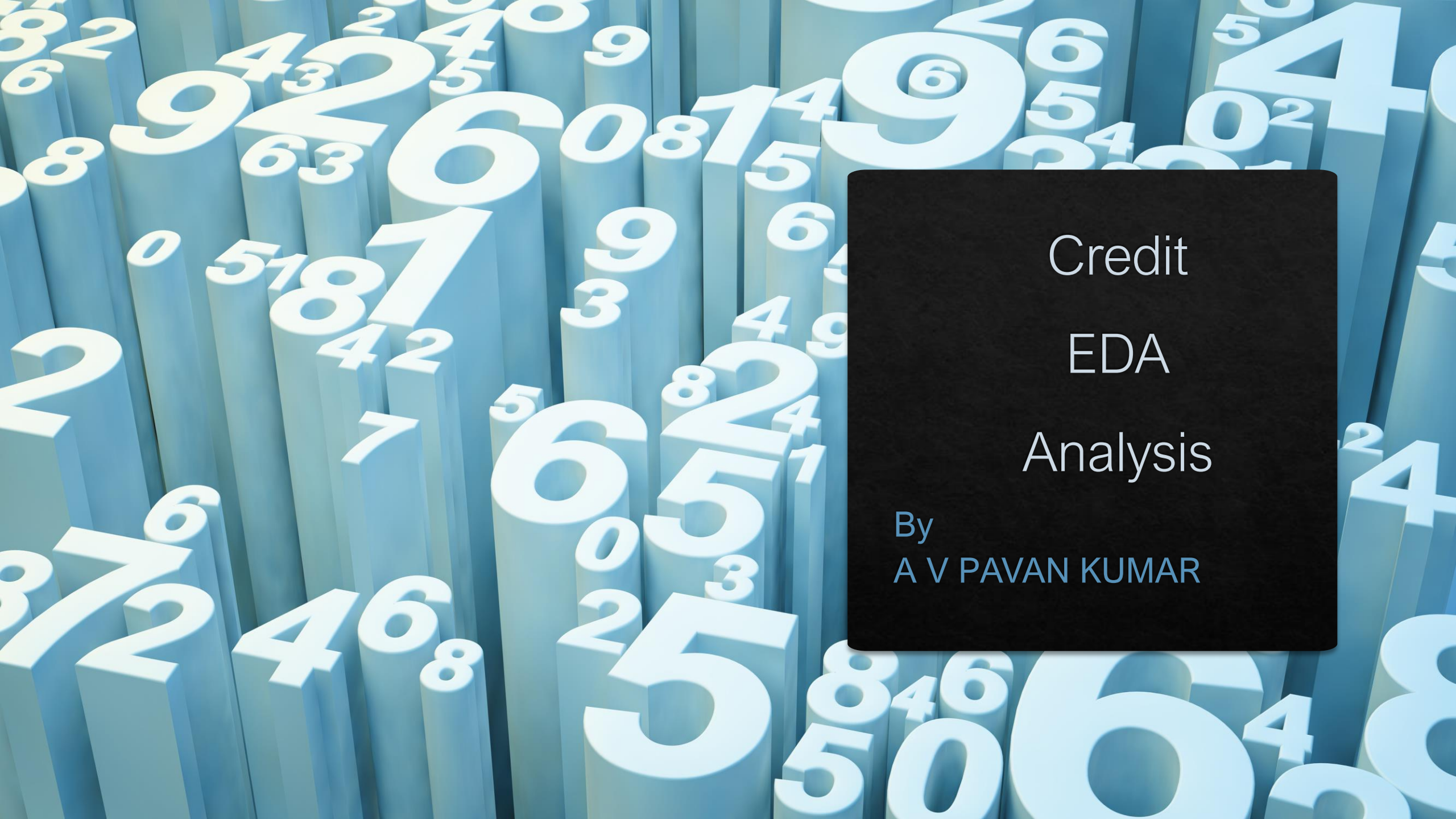


Credit EDA Analysis

By
A V PAVAN KUMAR

CREDIT EDA ANALYSIS

AIM

To do

Proper Exploratory

Data analysis on BANK data

INSIGHTS

Providing all the possible and required insights from the analysis done.

TASKS

Identifying the costumers with payment difficulties.

Content

Challenge Statement

Target 0 aims to identify clients who are expected to have a high likelihood of loan repayment, but declining their loan application may lead to a missed business opportunity for the company.

Target 1, on the other hand, aims to identify clients who have missed at least one installment payment by more than X days. This helps in detecting clients who may be at a higher risk of defaulting on their loan in the future, allowing the company to take necessary measures to reduce their credit risk.

Objective

In this case study, exploratory data analysis (EDA) will be employed to examine how consumer and loan attributes impact the probability of loan default. The loan application process includes four potential outcomes: approval, cancellation, refusal, and unused offer. Approval refers to the company's acceptance of the loan application, while cancellation occurs when the client decides not to proceed with the loan or receives unfavorable pricing due to increased risk. Refusal refers to the rejection of the loan application by the company, usually due to the client's failure to meet their requirements. Finally, an unused offer refers to a loan that was cancelled by the client at various stages of the process.

Lending protocol

Credit risk analysis involves assessing the probability that a borrower may default on their loan or other financial obligations. This entails examining the borrower's credit history, financial condition, and other relevant factors that could influence their capacity to repay the debt. The analysis is commonly utilized by lenders to decide whether or not to extend credit to a borrower and the terms and conditions to be offered.

Action plan

1. Getting Information and Understanding.
2. Issues with binning and data quality.
3. Data imbalance, correlation, and univariate, segmented univariate, and bivariate analysis.
4. Applying previous data to application data.
5. Analysis employing segmented univariate analysis, bivariate analysis, and correlation.
6. Risks and Recommendations

Data Analysis:- Data comprehension

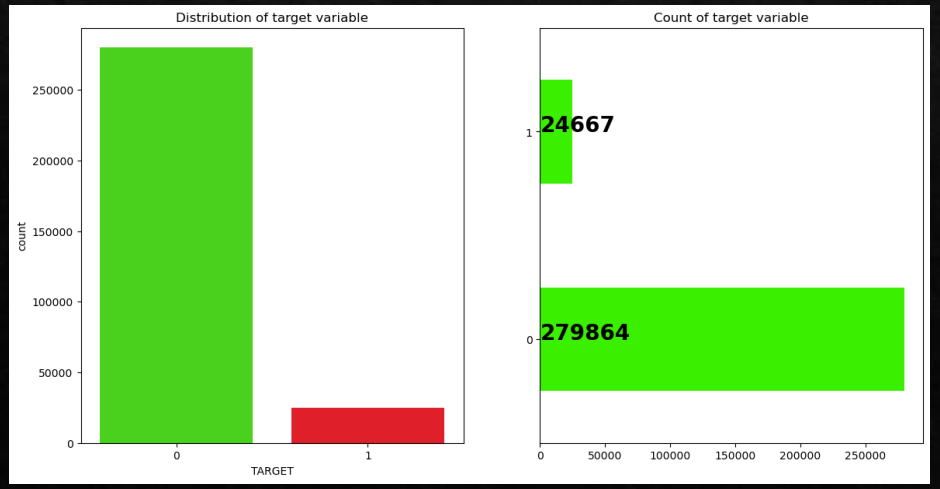
The available data for this study consists of three CSV files. The first file, 'application_data.csv', includes client information at the time of their loan application, specifically whether they had payment difficulties or not. The second file, 'previous_application.csv', provides data on the client's previous loan applications, indicating whether they were approved, cancelled, refused, or had an unused offer. Finally, the 'columns_description.csv' file serves as a data dictionary, explaining the meaning of the variables used in the other two files.

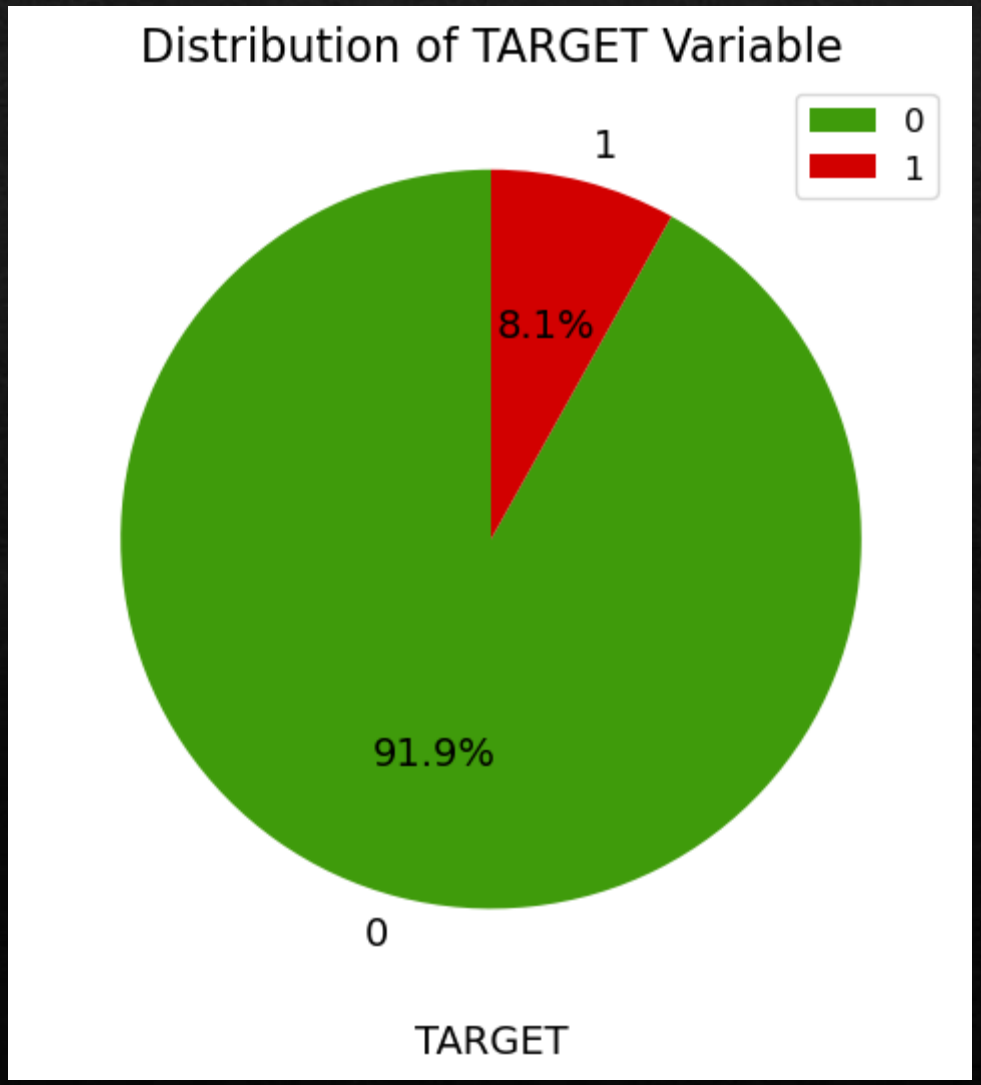
Data cleaning and Formatting

1. Analyze the data to find outliers, abnormalities, and missing values.
2. Deal with missing data by either eliminating them from the dataset or imputing them.
3. Identify outliers and abnormalities, then get rid of or alter them as needed.
4. Identify duplicate records in the dataset and remove them.
5. If required, standardize or normalize the data to make it simpler to compare and analyze.
6. Check for inconsistencies and errors in the data by validating against external sources or using domain knowledge.
7. If required, convert data types to ensure that the data is in the proper format for analysis.
8. Create new variables from the existing data, if necessary, to gain more insights or improve predictive accuracy.

Data skewness

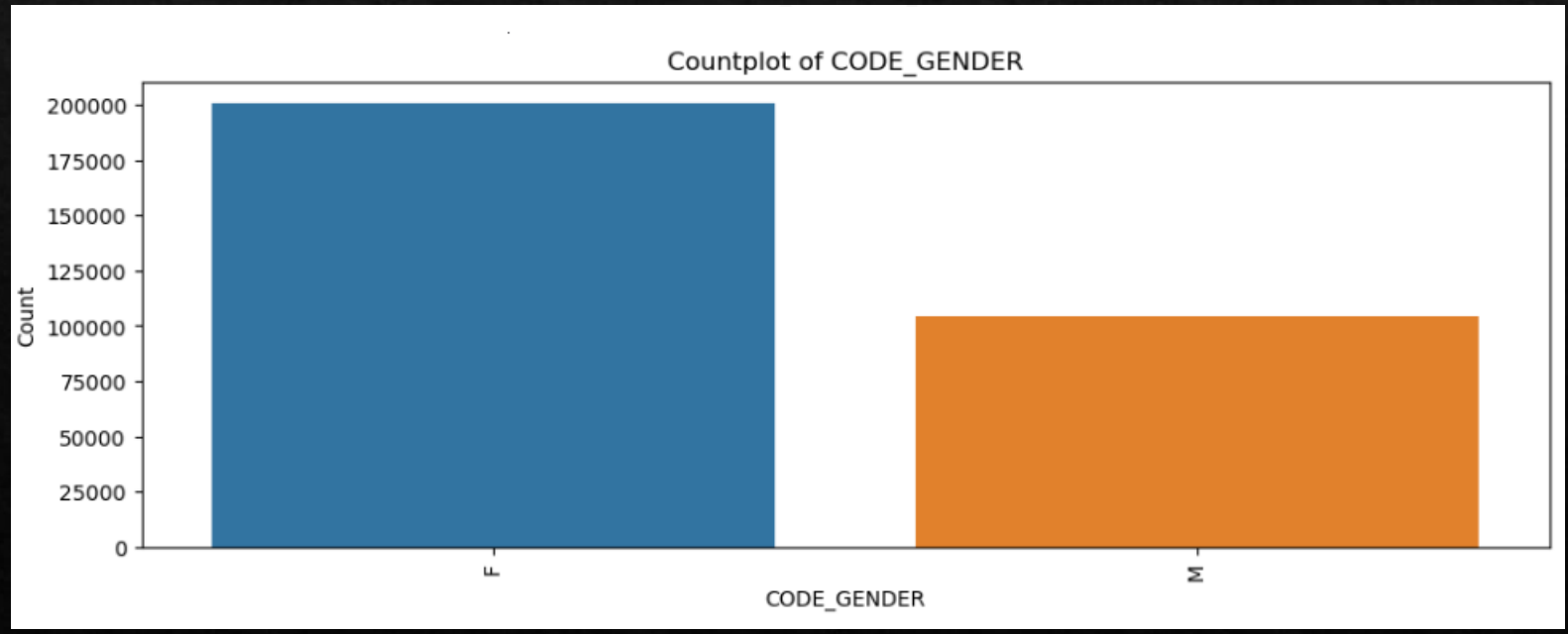
- 91.9% Applicants are Non-defaulters.
- 8.1% Applicants have issues with loan repayment.





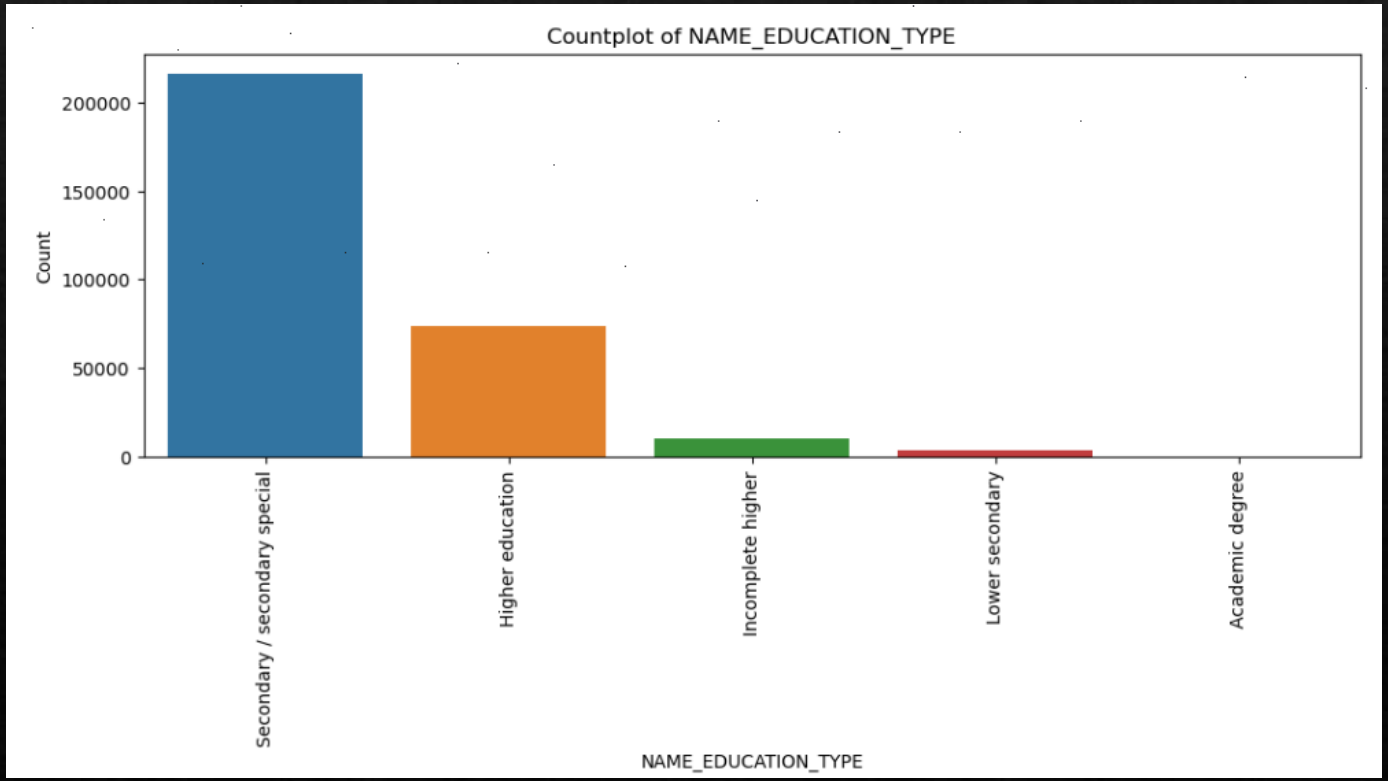
Univariate Analysis: Gender

There is a higher frequency of loan applications from females compared to males, possibly due to a lower interest rate charged by banks for female applicants.



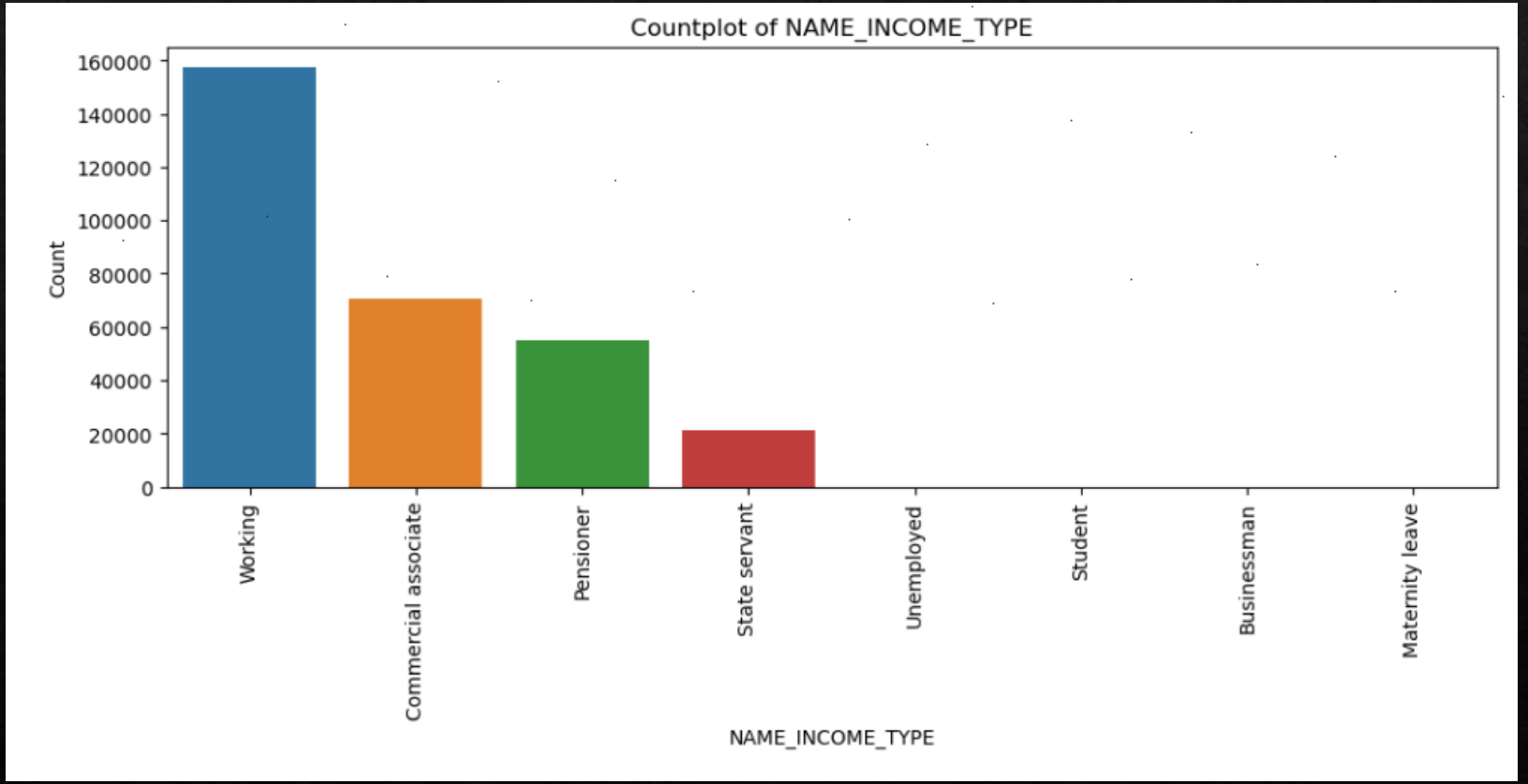
Univariate: Education type

The majority of loan applicants, approximately 71%, have a Secondary/Secondary Special education type.



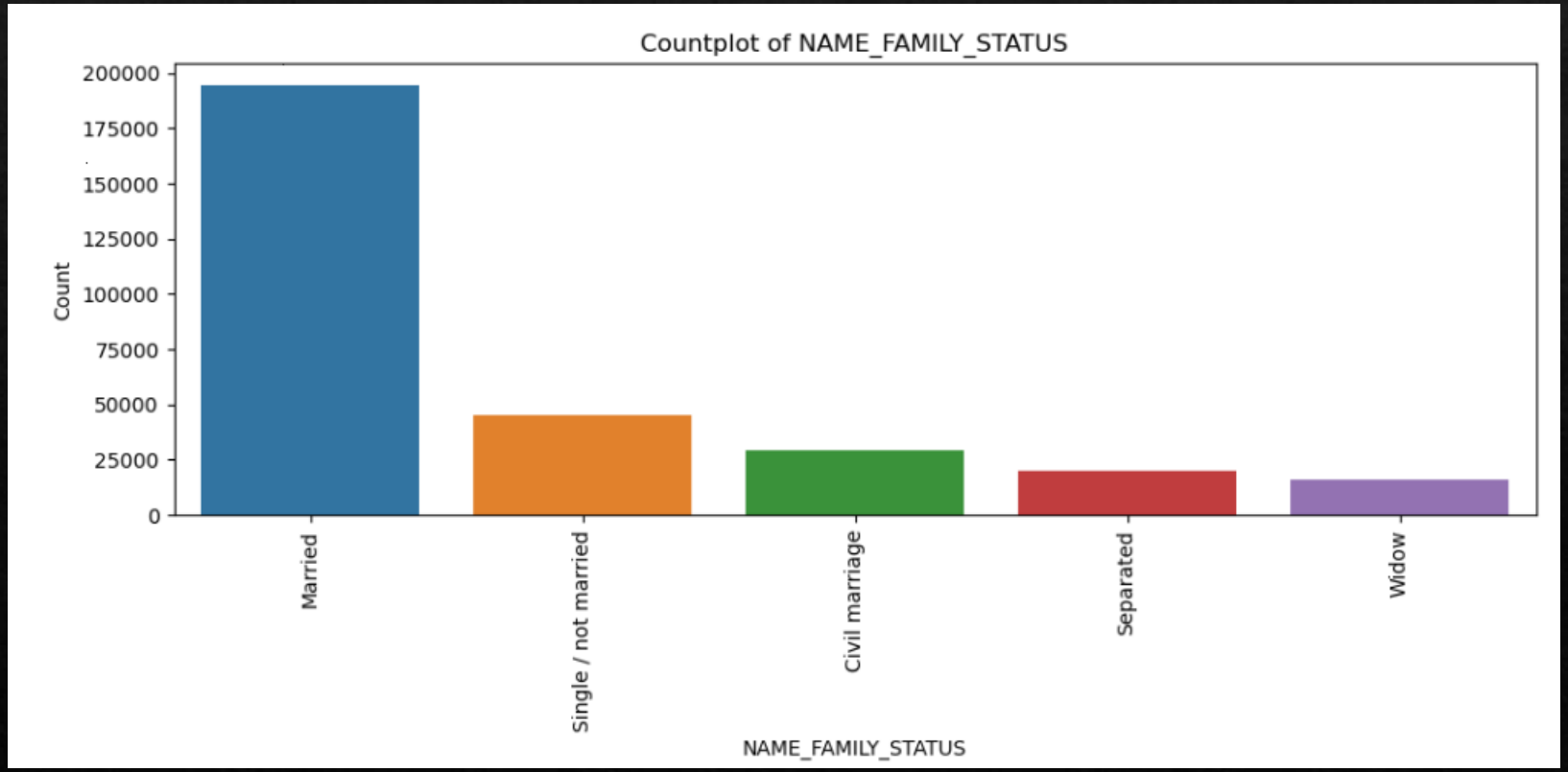
Univariate: Income Type

Over 50% of loan applicants belong to the Working Income Type category.



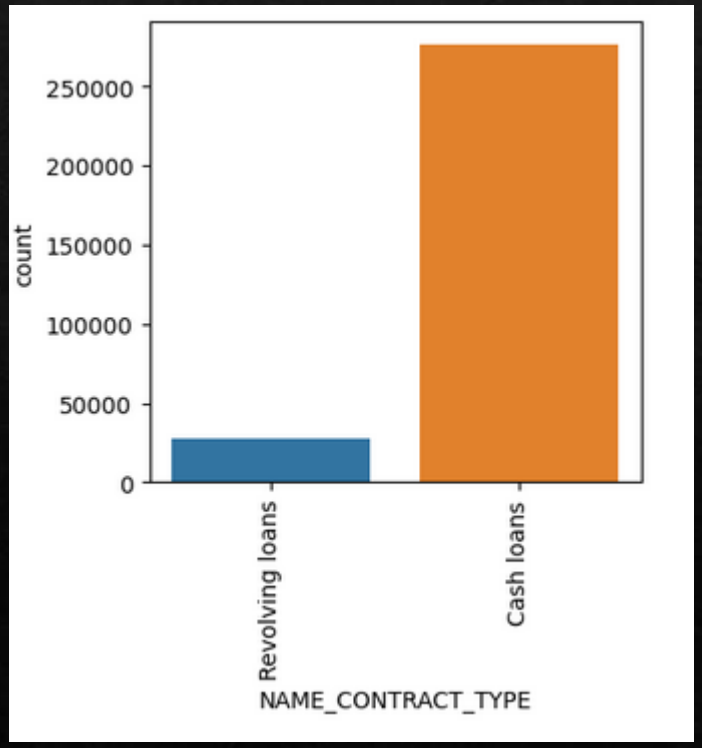
Univariate: Family status

There is a higher frequency of loan applications from married individuals, as evidenced by 63.9% of loan applicants being married.



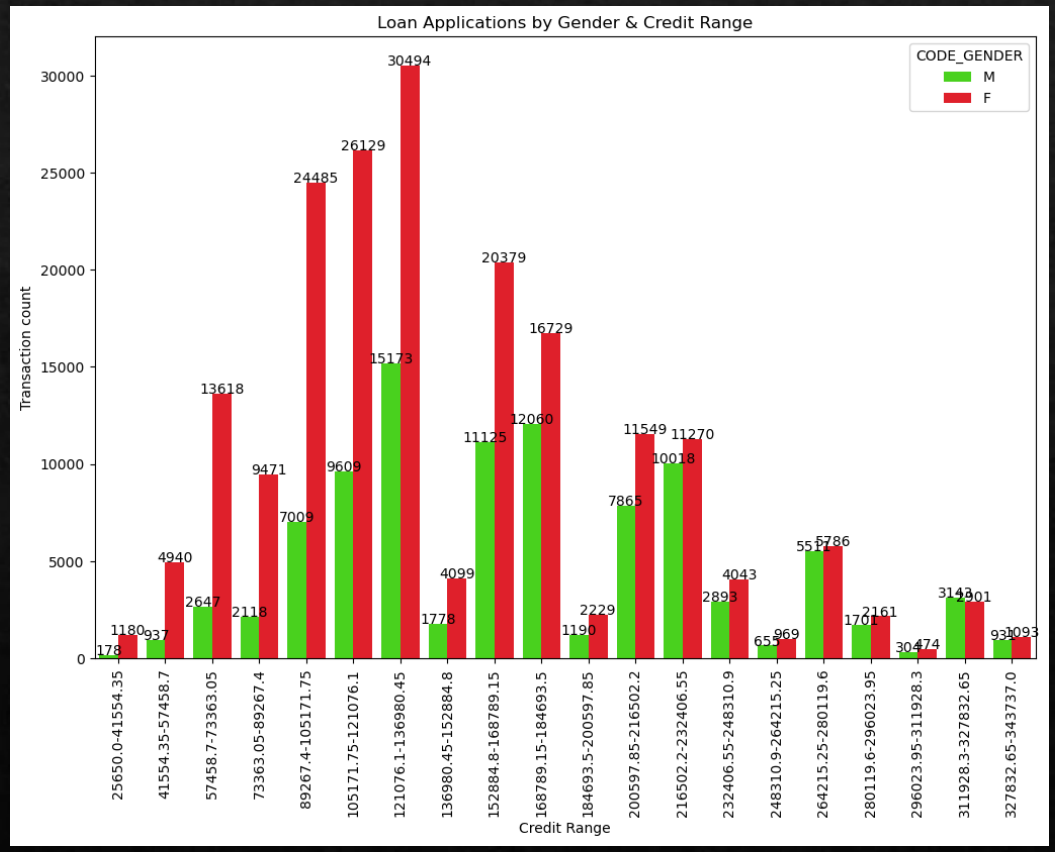
Univariate: Contract type

90.5% Applicants have requested for Cash loans.



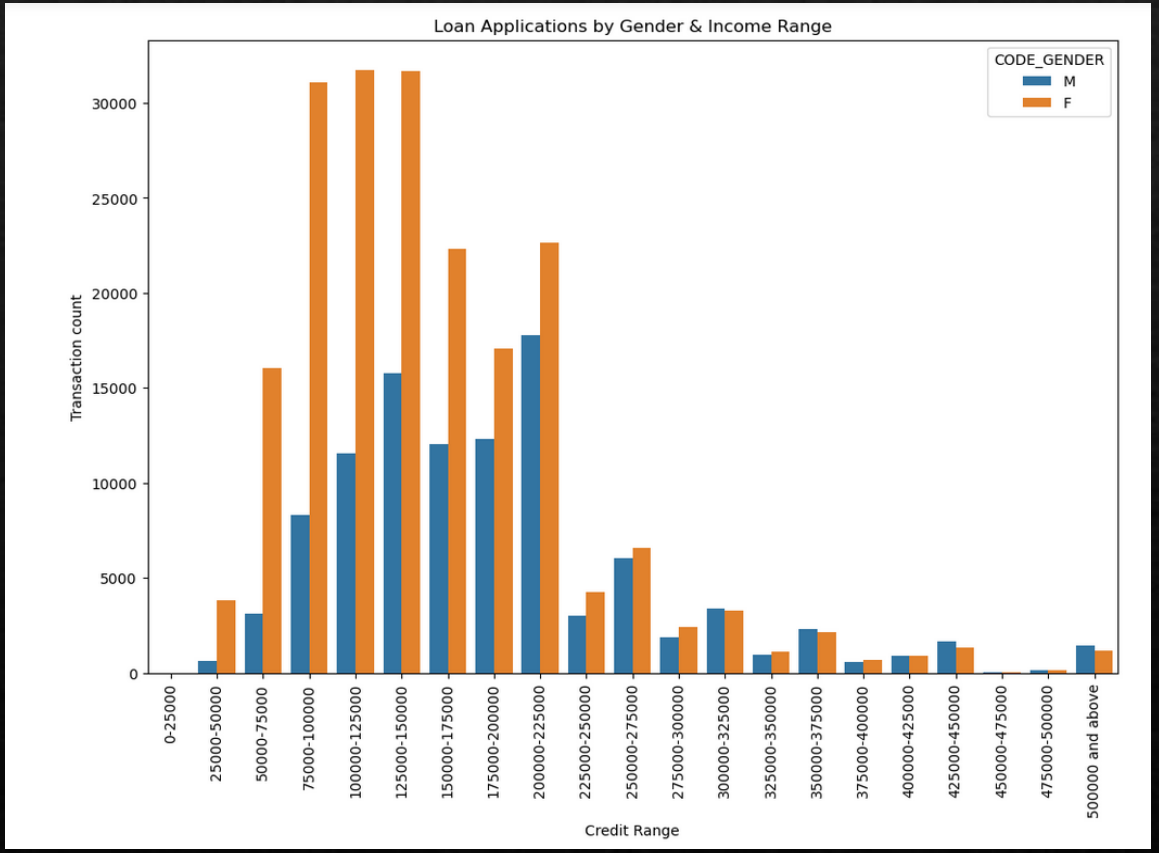
Bivariate: Gender & Credit range

1. The majority of Very Low credit loans are being applied for by females.
2. Males tend to apply for Medium and High credit loans.



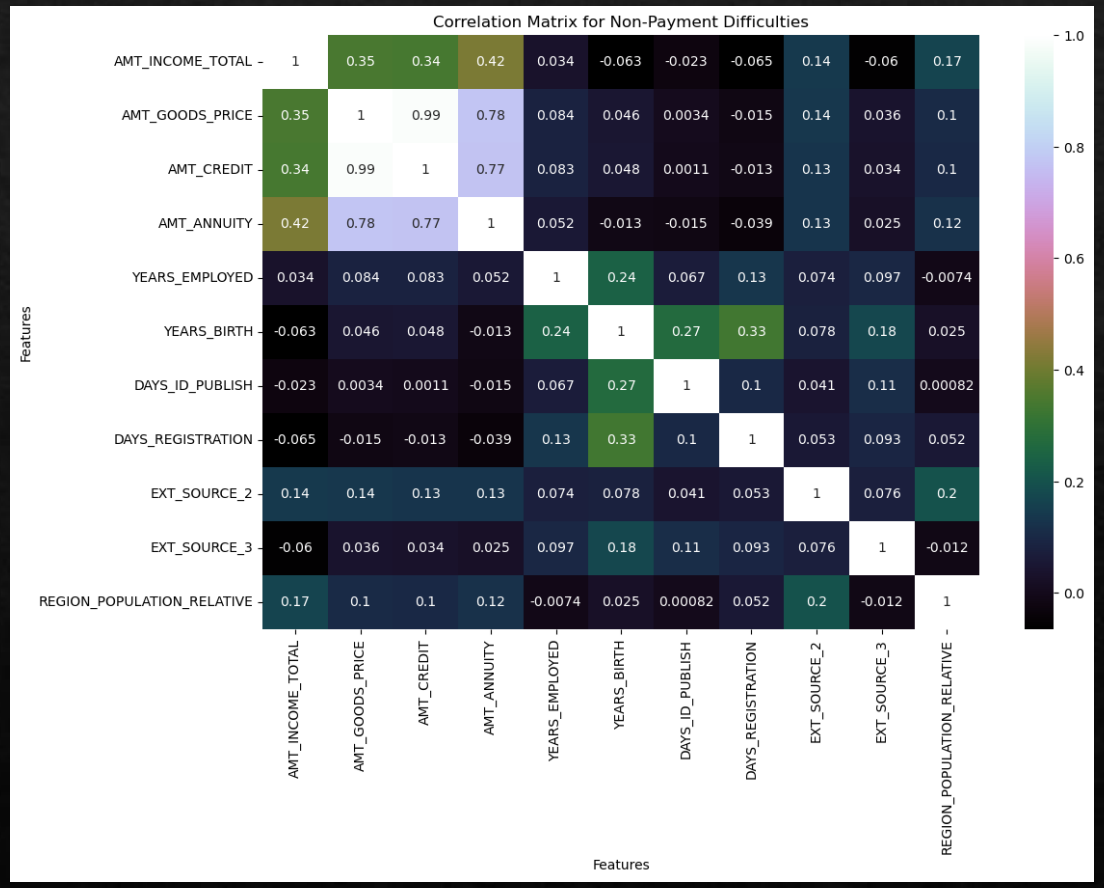
Bivariate: Gender & Income range

Females in average having more income than men but they applied for loan with very less incomes also.



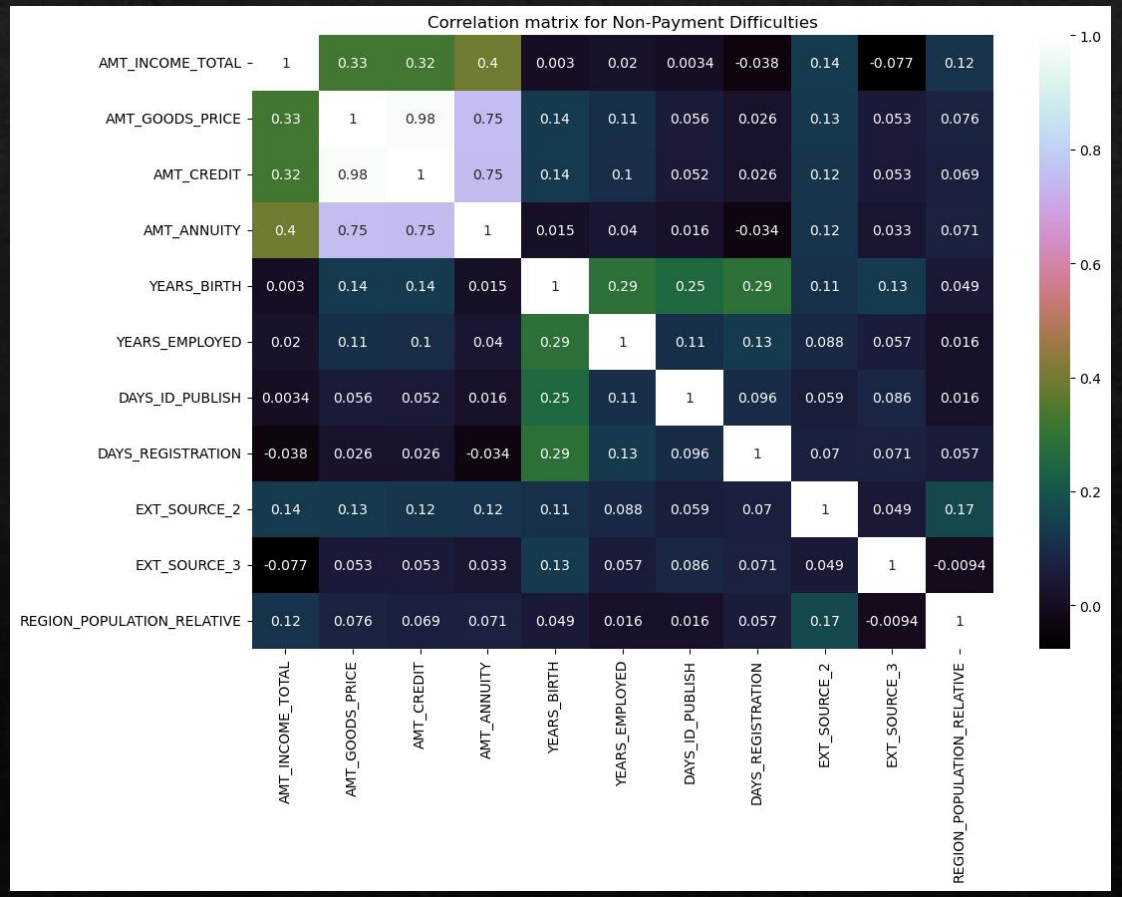
Multivariate: **Correlation in target0**

- AMT_GOODS_PRICE & AMT_CREDIT has highest correlation.
- YEARS_BIRTH, YEARS_EMPLOYED & AMT_INCOME_TOTAL has Negative correlation.



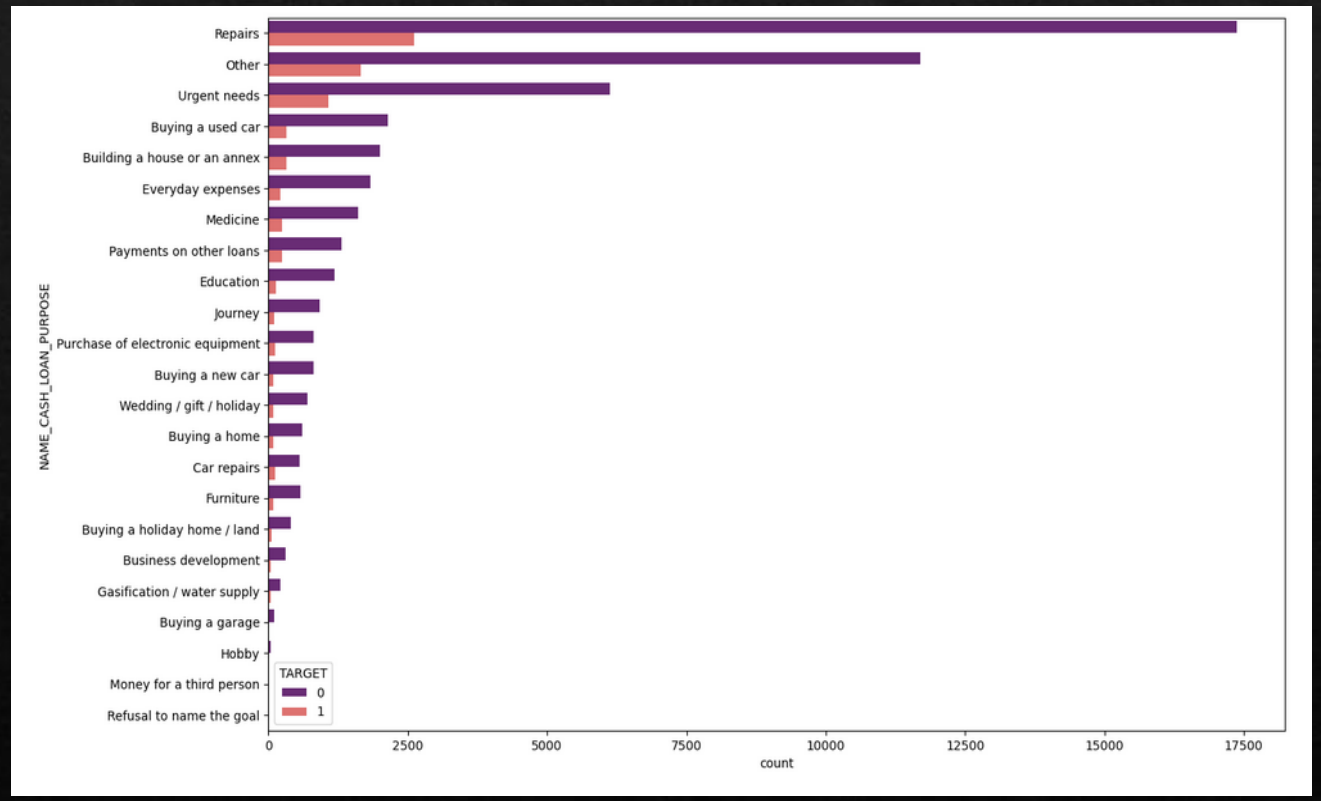
Multivariate: **Correlation in target1**

Females in average having more income than men but they applied for loan with very less incomes also.



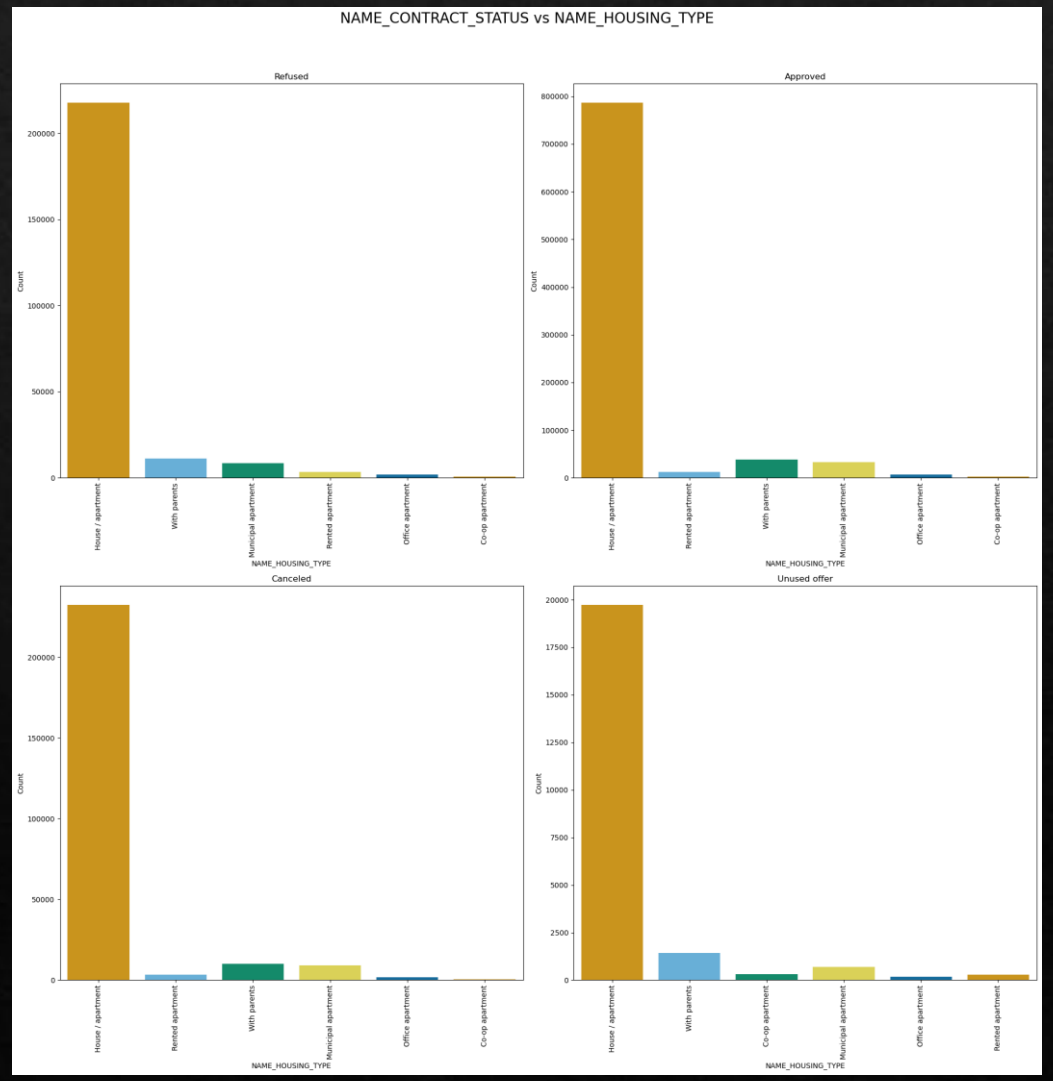
Univariate analysis on Merged data

1. Loan applicants who cite "Repairs" as the purpose of their loan are more likely to experience payment difficulties.
2. Loan payment performance is notably better for loan purposes such as "Buying a garage," "Business development," "Buying land," "Buying a new car," and "Education," suggesting that these loan purposes may be a strategic focus area for minimizing payment difficulties.



Bivariate analysis on Merged data

Here, For 'office apartment' is having high credit of target 0 and 'co-op apartment' is having high credit of target 1. By this, we can conclude that bank should avoid giving loans to co-op apartment housing type, as they are having difficulties in payment and can focus mostly on housing type with parents or House\apartment or municipal apartment for successful payments. rephrase it please



Findings

Certainly! Based on the analysis, here are the main insights:

•Banks mostly approve Consumer Loans.
•Most of the Refused and Cancelled loans are cash loans.
•Most of the approved, refused, and canceled loans belong to old clients, with only 27.4% of loans provided to new customers.
•The percentage of loans approved for females is higher than the percentage refused.
•Most of the approved loans belong to applicants with Secondary/Secondary Special education type.
•The percentage of loans approved for married applicants is higher than for other contract status categories (refused, canceled, etc.).
•Most of the loans that were previously approved belong to the POS name portfolio.
•Credit and cash offices channel type has the highest number of refused and canceled loans.
•Most of the approved loans have a medium grouped interest rate.
•Most of the approved loans belong to Very Low and High Credit range.
•Most of the loans are approved for applicants with a low income range.
•Across all contract statuses (Approved, Refused, Canceled, Unused Offer), people with the Working income type are leading.

It is important to note that these insights are based on the data available and may not necessarily apply in all situations. It is also recommended to conduct further analysis and modeling to gain deeper insights and inform decision-making.