Credit EDA Analysis

BY GARVIT SUNEJA

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.

Challenge Statement

Target 0 endeavors to identify clients who demonstrate a substantial likelihood of fulfilling their loan obligations with utmost diligence. Nonetheless, declining their loan application may inadvertently entail a missed prospect for business expansion and growth, thereby warranting prudent evaluation.

Conversely, Target 1 aspires to discern clients who have exhibited a propensity for delayed installment payments, surpassing a specified threshold of X days. By scrutinizing and identifying such individuals, this approach facilitates the recognition of borrowers who may harbor an elevated risk of loan default in the future. Consequently, the organization can proactively implement requisite measures to mitigate potential credit risks and fortify their risk management framework.

Objective

This case study employs exploratory data analysis (EDA) to investigate the relationship between consumer and loan attributes and their impact on the probability of loan default. The loan application process encompasses four potential outcomes: approval, cancellation, refusal, and unused offer. Approval denotes the acceptance of the loan application by the company, while cancellation occurs when the client decides not to proceed with the loan or faces unfavorable pricing due to increased risk. Refusal indicates the rejection of the loan application by the company, typically due to the client's failure to meet the necessary requirements. Lastly, an unused offer refers to a loan that was cancelled by the client at various stages of the process.

Lending protocol

Credit risk analysis encompasses the evaluation of the likelihood of a borrower defaulting on their loan or other financial obligations. It involves a thorough assessment of the borrower's credit history, financial situation, and other pertinent factors that can impact their ability to repay the debt. Lenders commonly utilize this analysis to make informed decisions regarding extending credit to borrowers, including determining the appropriate terms and conditions for the credit arrangement.

Action plan

1) Information Acquisition and Comprehension
2) Binning and Data Quality Challenges
3) Data Imbalance, Correlation, and Analysis Techniques
4) Univariate, Segmented Univariate, and Bivariate Analysis
5) Application of Previous Data to Current Data
6) Risks and Recommendations

Data Comprehension

The dataset used in this study comprises three CSV files. The first file, named 'application_data.csv', contains client information recorded during their loan application process. This information includes details on whether the clients experienced any payment difficulties.

The second file, labeled 'previous_application.csv', presents data related to the client's previous loan applications. It provides insights into the outcomes of those applications, indicating whether they were approved, cancelled, refused, or resulted in an unused offer.

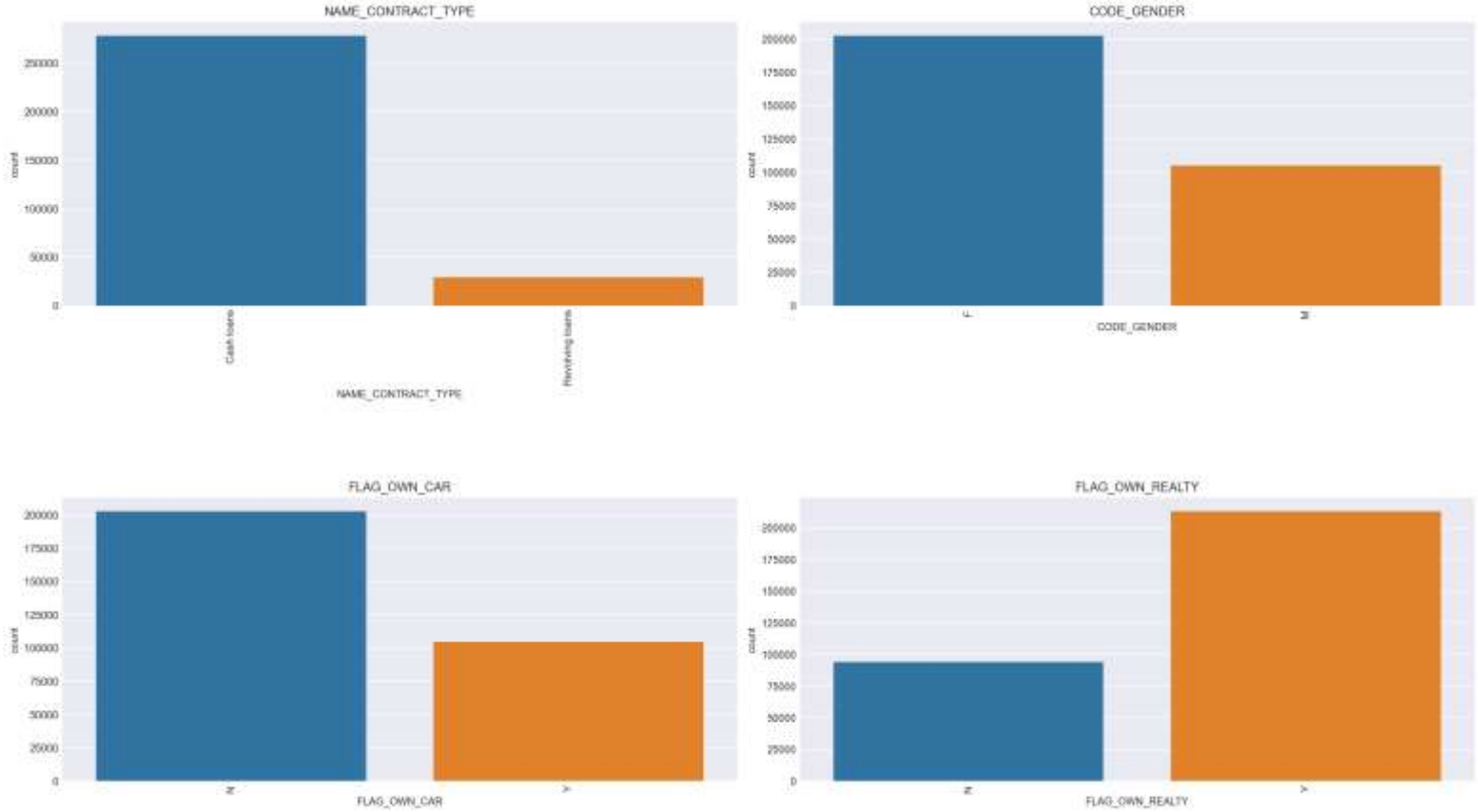
Lastly, the file 'columns_description.csv' functions as a data dictionary, offering an explanation of the variables used in the other two files. It provides clarity and understanding regarding the meaning and context of the variables.

These datasets collectively form the foundation for conducting a comprehensive analysis of the loan application data and its associated attributes.

Refining Data Quality

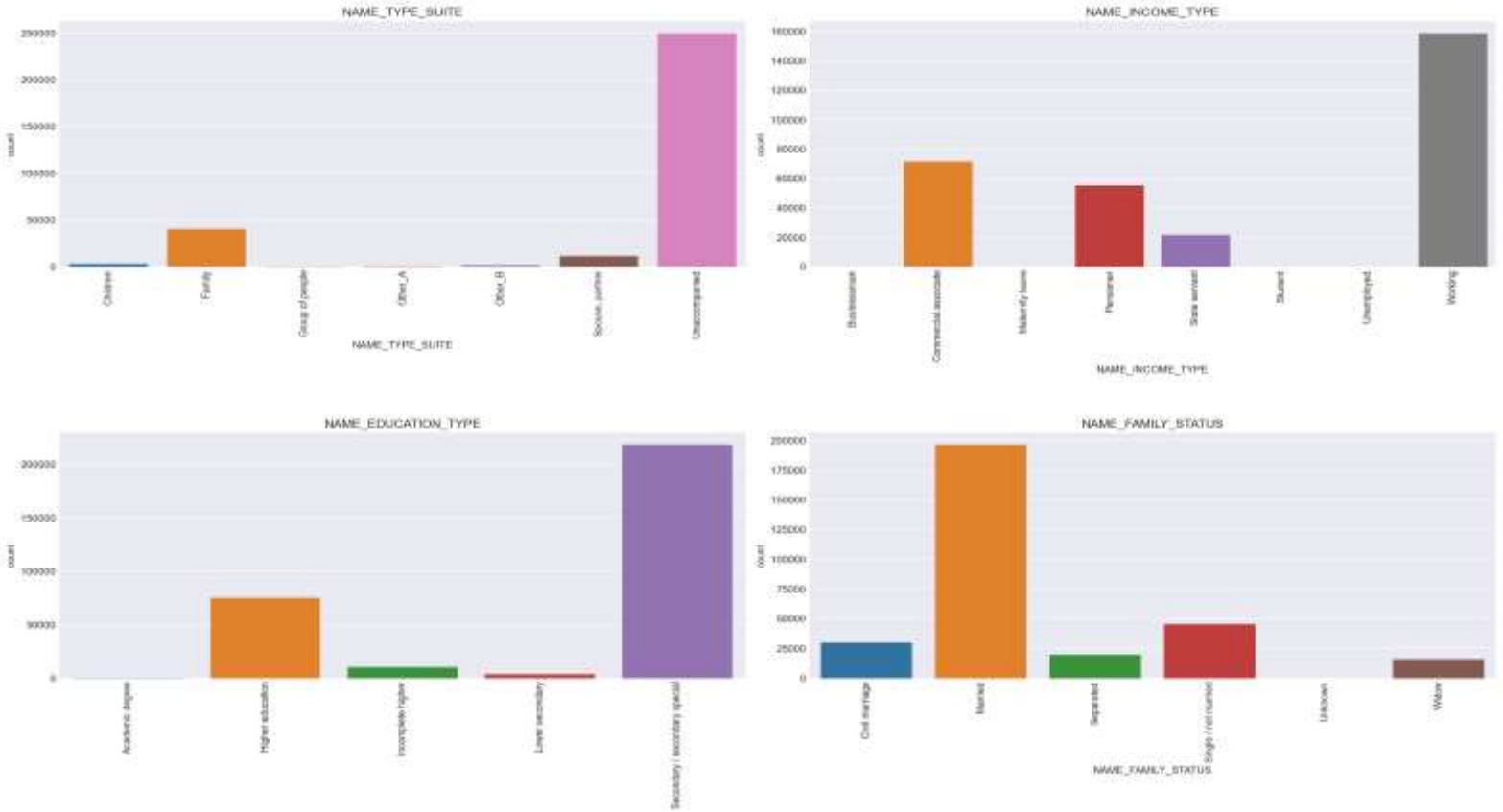
1. Identify outliers, abnormalities, and missing values.
2. Handle missing data through elimination or imputation.
3. Identify and address outliers and abnormalities.
4. Remove duplicate records.
5. Standardize or normalize the data for better comparison and analysis.
6. Validate data consistency and accuracy against external sources or domain knowledge.
7. Convert data types, if needed, for proper format alignment.
8. Perform feature engineering to create new variables for enhanced insights or predictive accuracy.

Univariate Analysis - 1



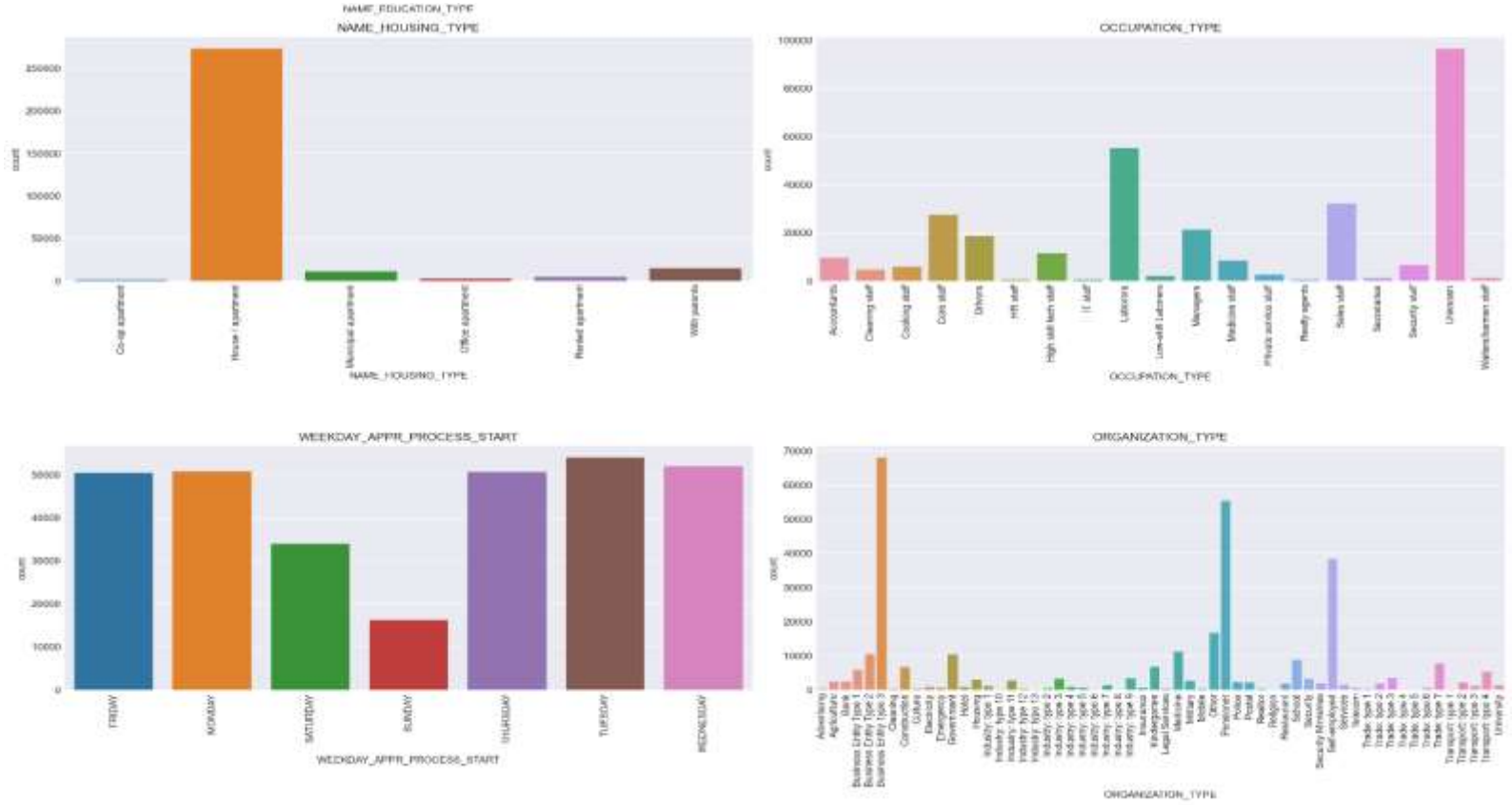
- Around 90% of applicants have taken cash loans, while the remaining have chosen resolving loans.
- There is a higher frequency of loan applications from females compared to males,
- About 65% of applicants do not own a car.
- Around 69% of applicants own their living quarters.

Univariate Analysis - 2



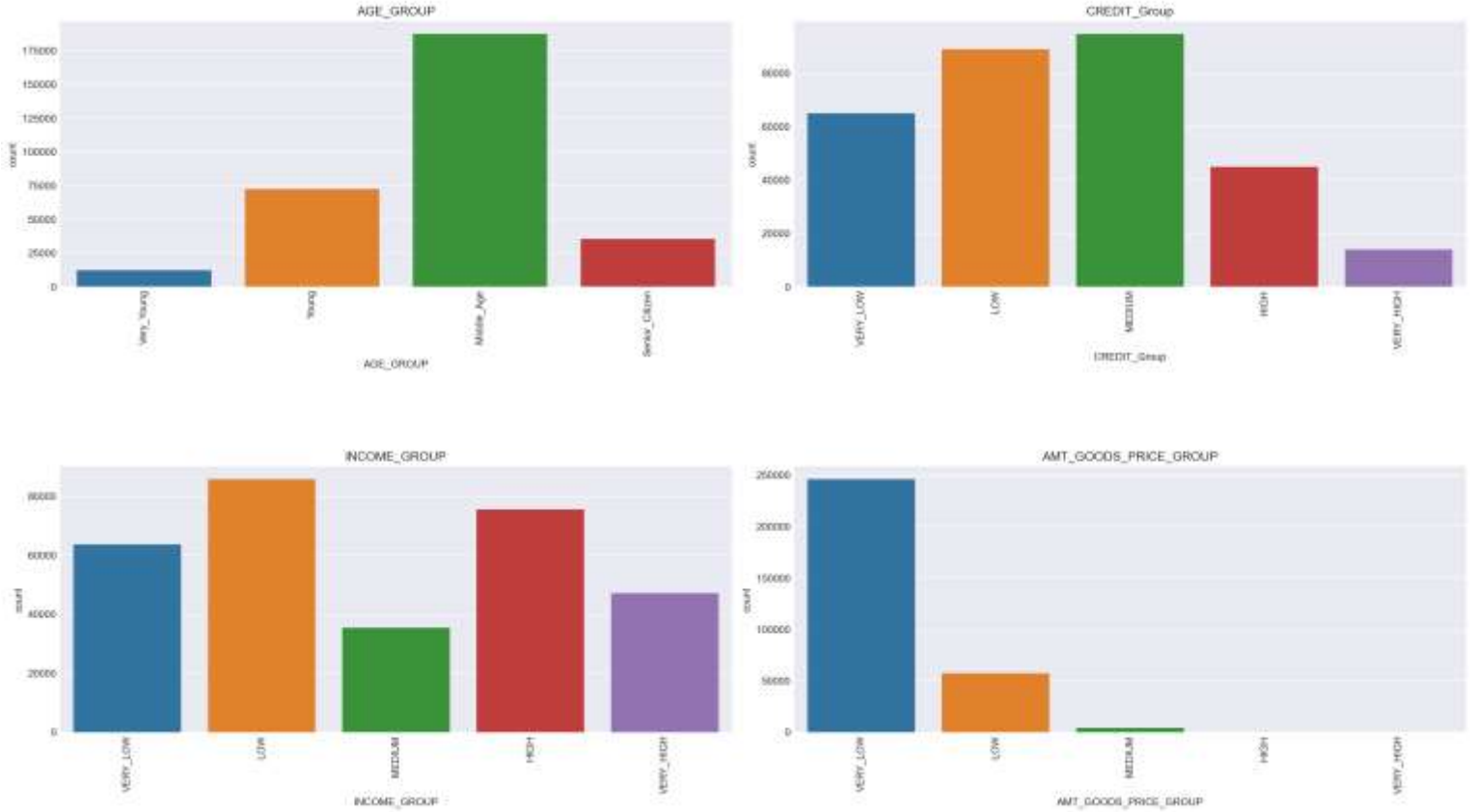
- The majority (about 81%) of applicants came unaccompanied for the loan application.
- The largest proportion of applicants belong to the working class, followed by commercial associates and pensioners.
- Around 71% of applicants have completed secondary education.
- The most common family status among applicants is married, accounting for 63%.

Univariate Analysis - 3



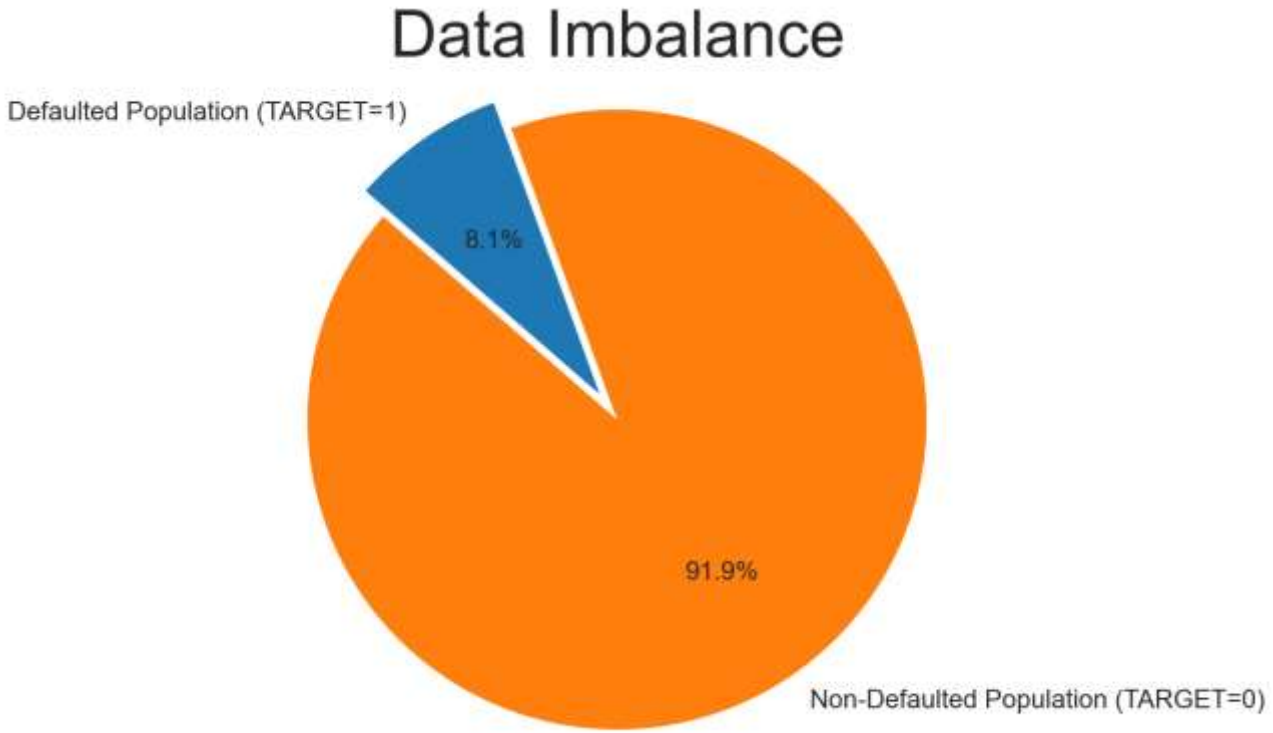
- A majority of applicants have taken loans for a house or apartment.
- Approximately 31% of applicants have not specified their occupation type, followed by laborers.

Univariate Analysis - 4



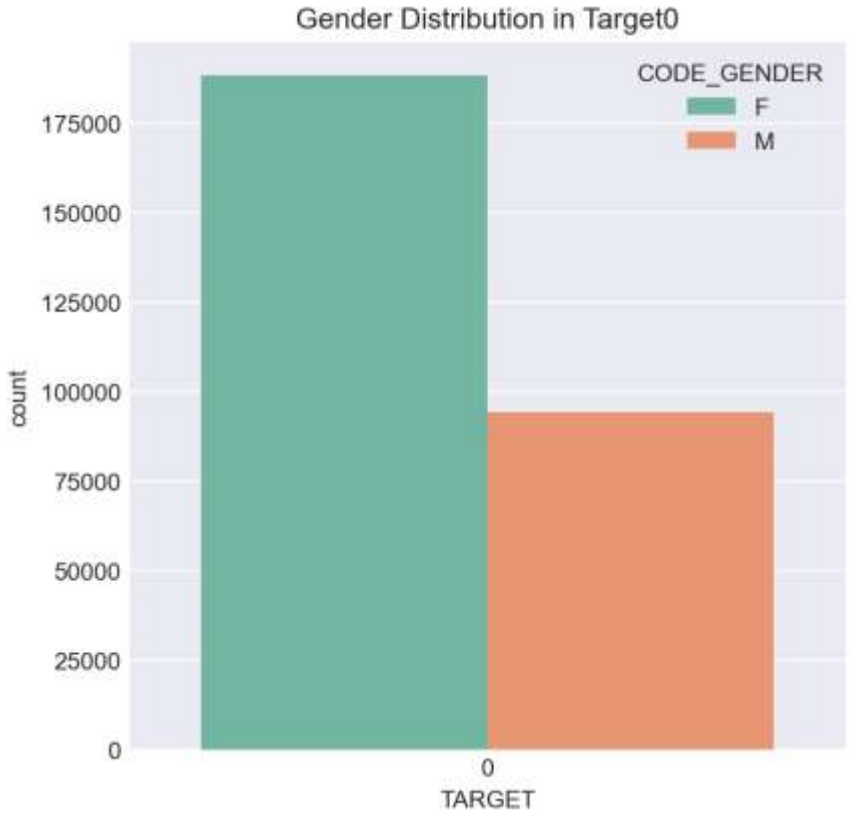
- The majority (60%) of applicants fall into the middle-aged category.
- Approximately 59% of applicants opted for medium to low loan amounts.
- Among the applicants, 28% belong to the low-income group, followed by those in the high-income category.
- It is observed that a significant number of applicants are purchasing goods with very low prices.

Data skewness



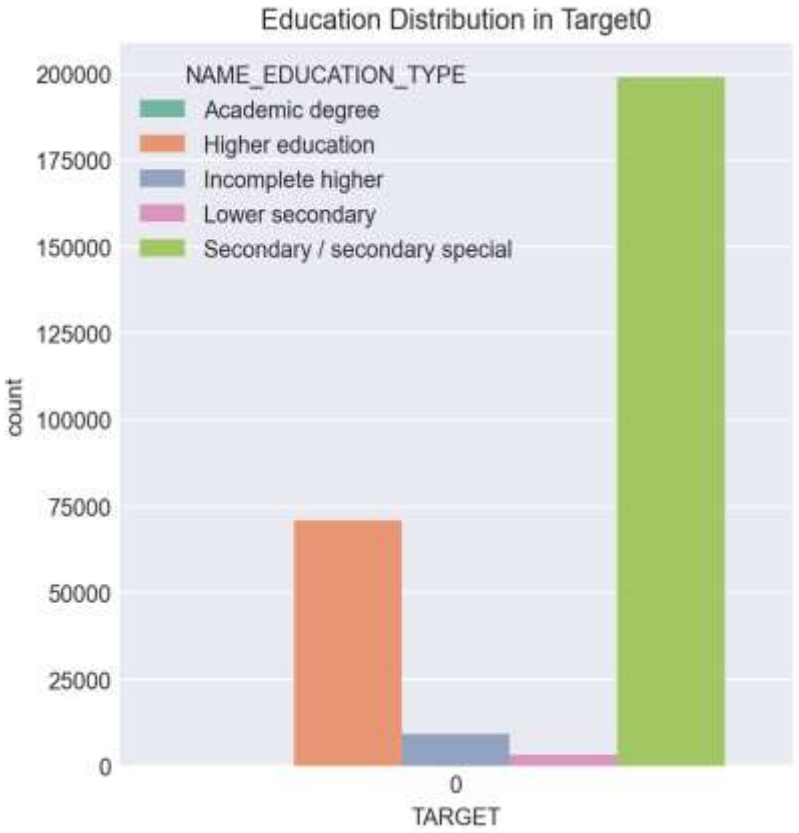
Based on the analysis, it is found that 91.9% of the applicants are classified as non-defaulters, while 8.1% of the applicants have encountered issues with loan repayment.

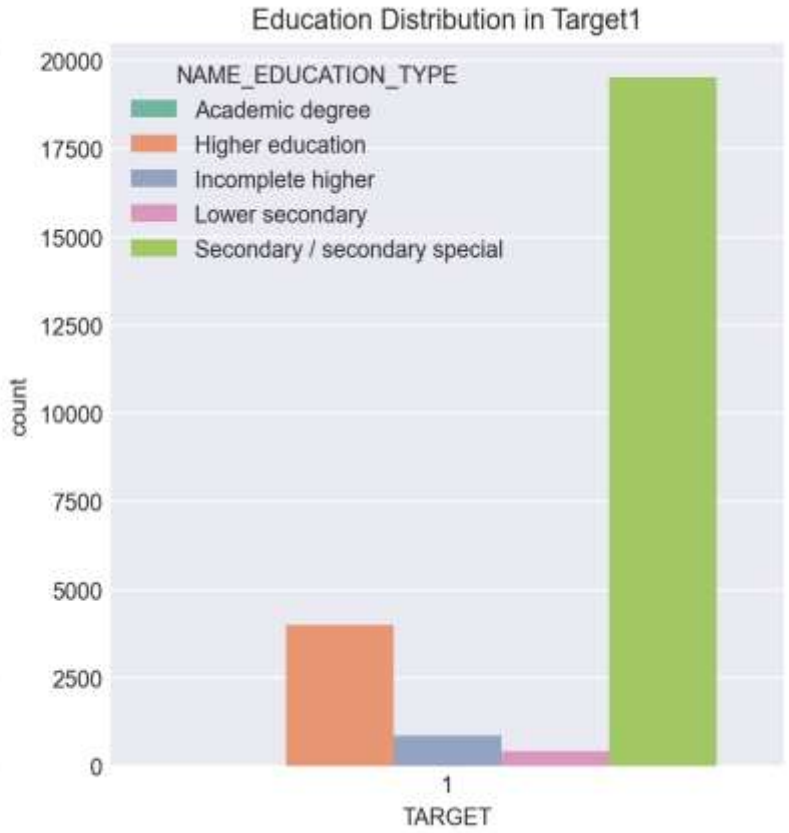
Bivariate Analysis - Gender



- It appears that female clients have a higher loan application rate compared to male clients.
- Among the female clients, 66.6% are classified as non-defaulters, while 33.4% are classified as defaulters.
- For male clients, 57% are classified as defaulters, while 42% are classified as non-defaulters.

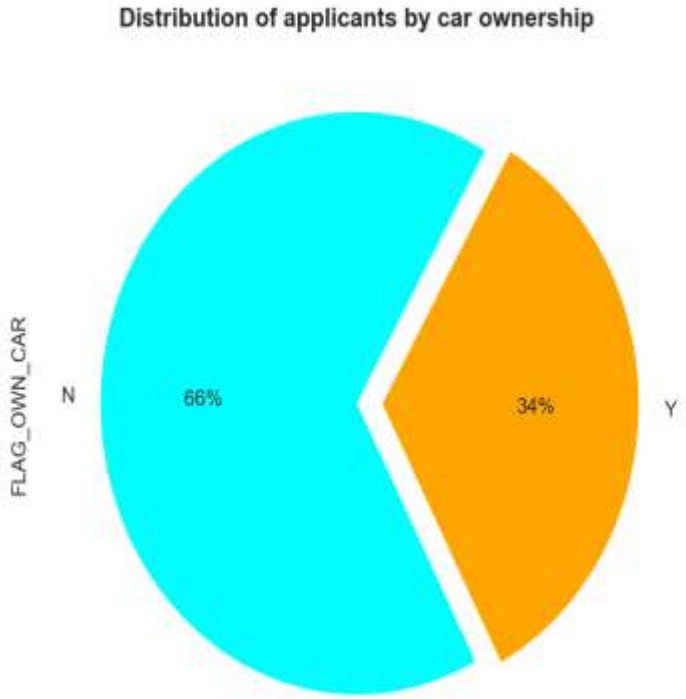
Bivariate Analysis – Education

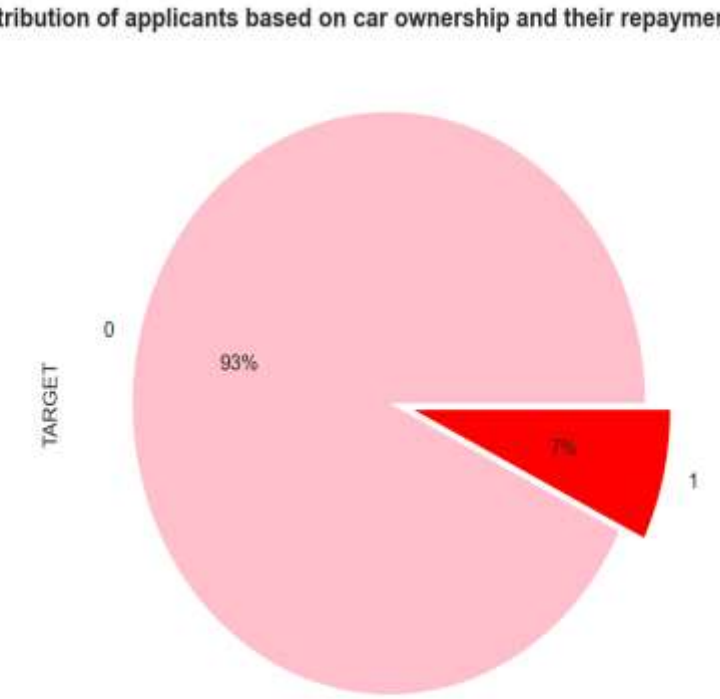




- applicants with education levels of Secondary or Secondary Special are more likely to apply for a loan.
- applicants with education levels of Secondary or Secondary Special have a higher risk of loan default. On the other hand, clients with other education types have a minimal risk of default.

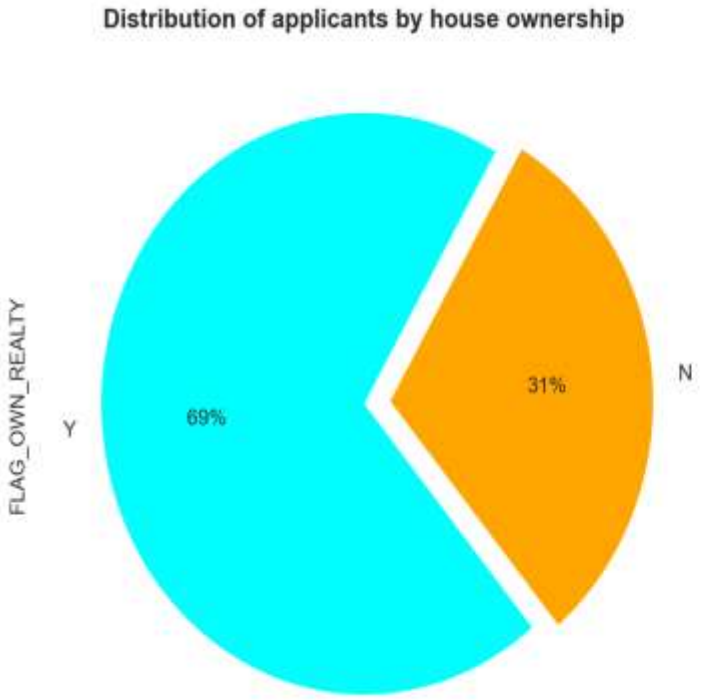
Bivariate Analysis – Own a Car

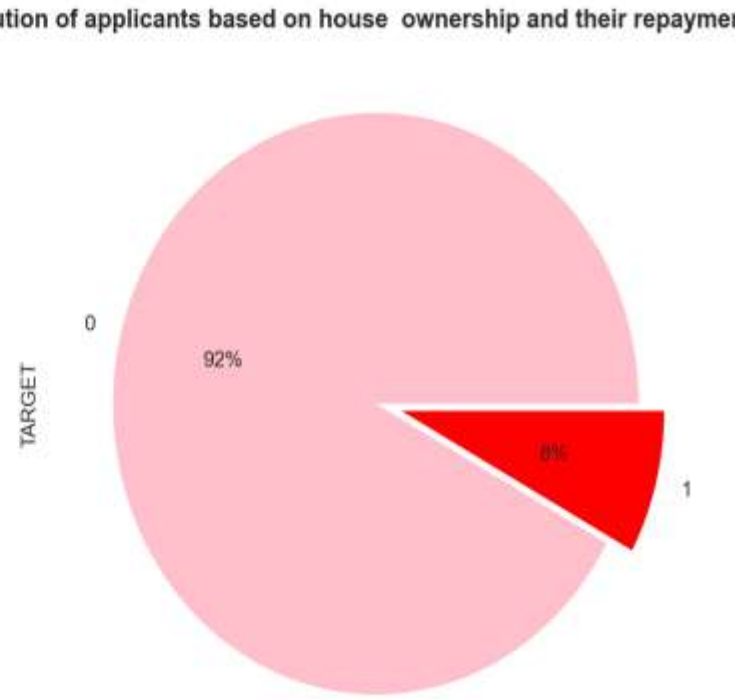




- The majority of clients (62%) do not own a car.
- Among clients who own a car, only a small percentage (8%) experience difficulty in making payments.

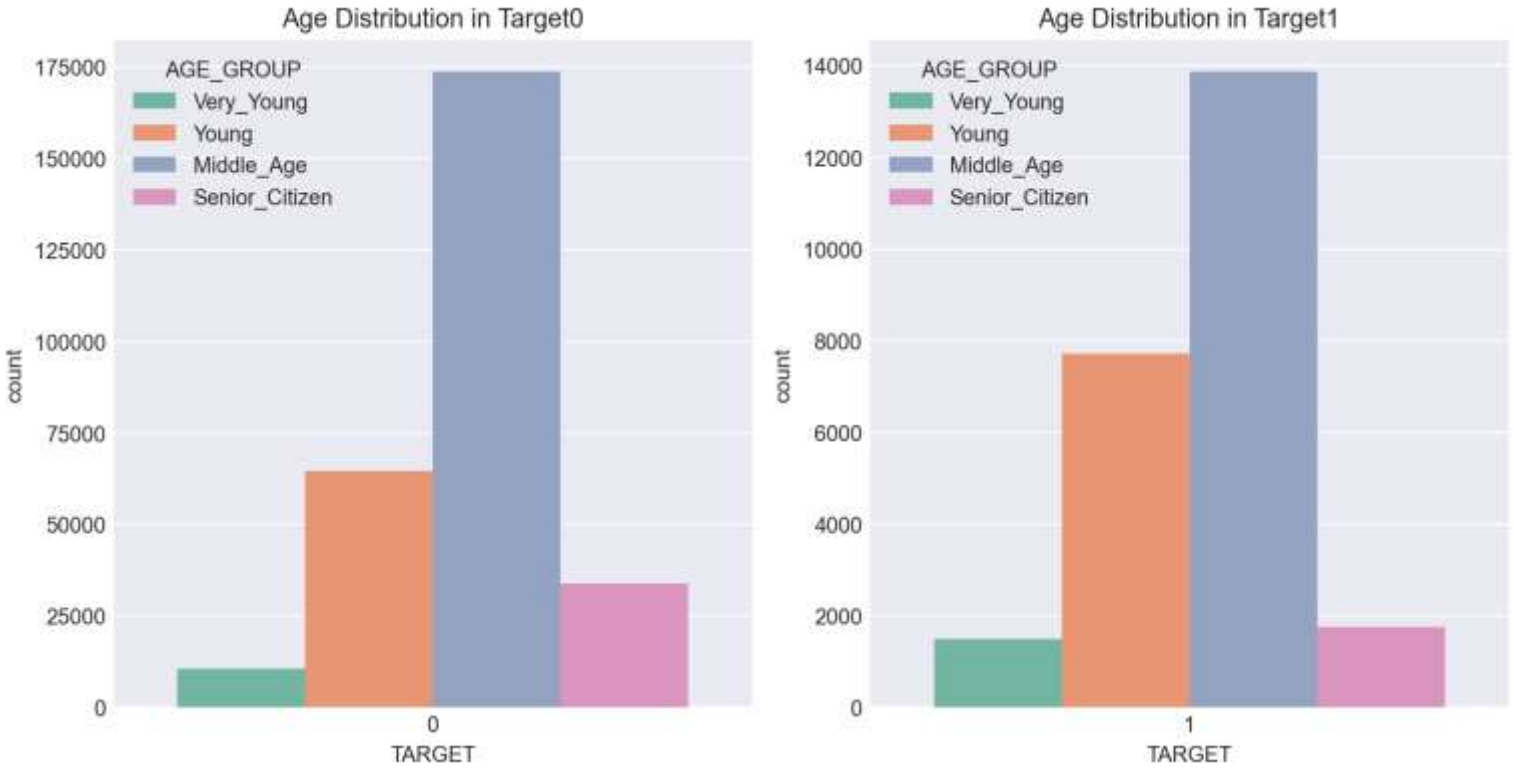
Bivariate Analysis – Own a House





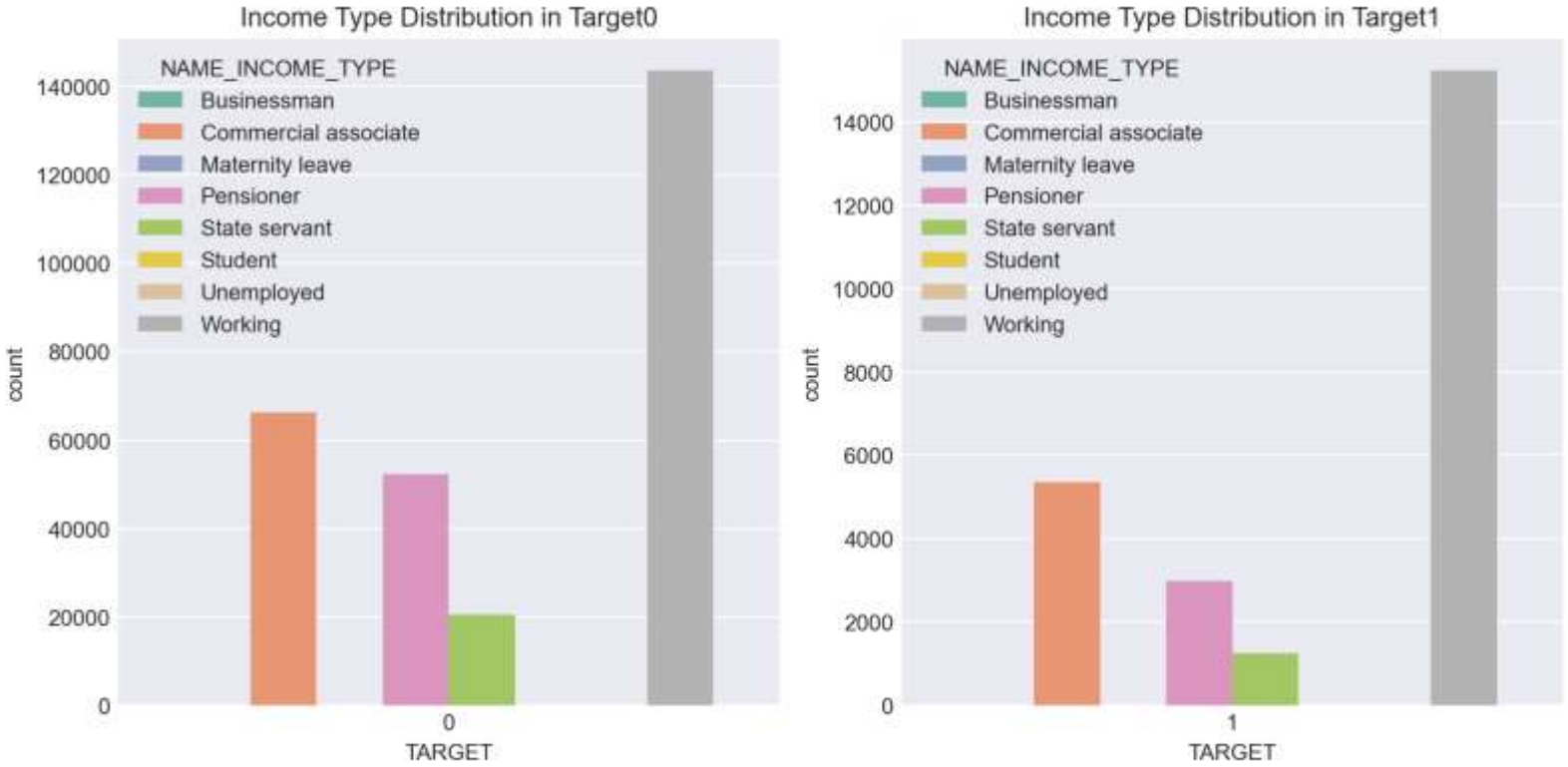
- The pie chart shows that 69% of clients own a house or a flat.
- Among clients who own a house or a flat, 8% of them experience difficulty in making payments.

Bivariate Analysis - Age



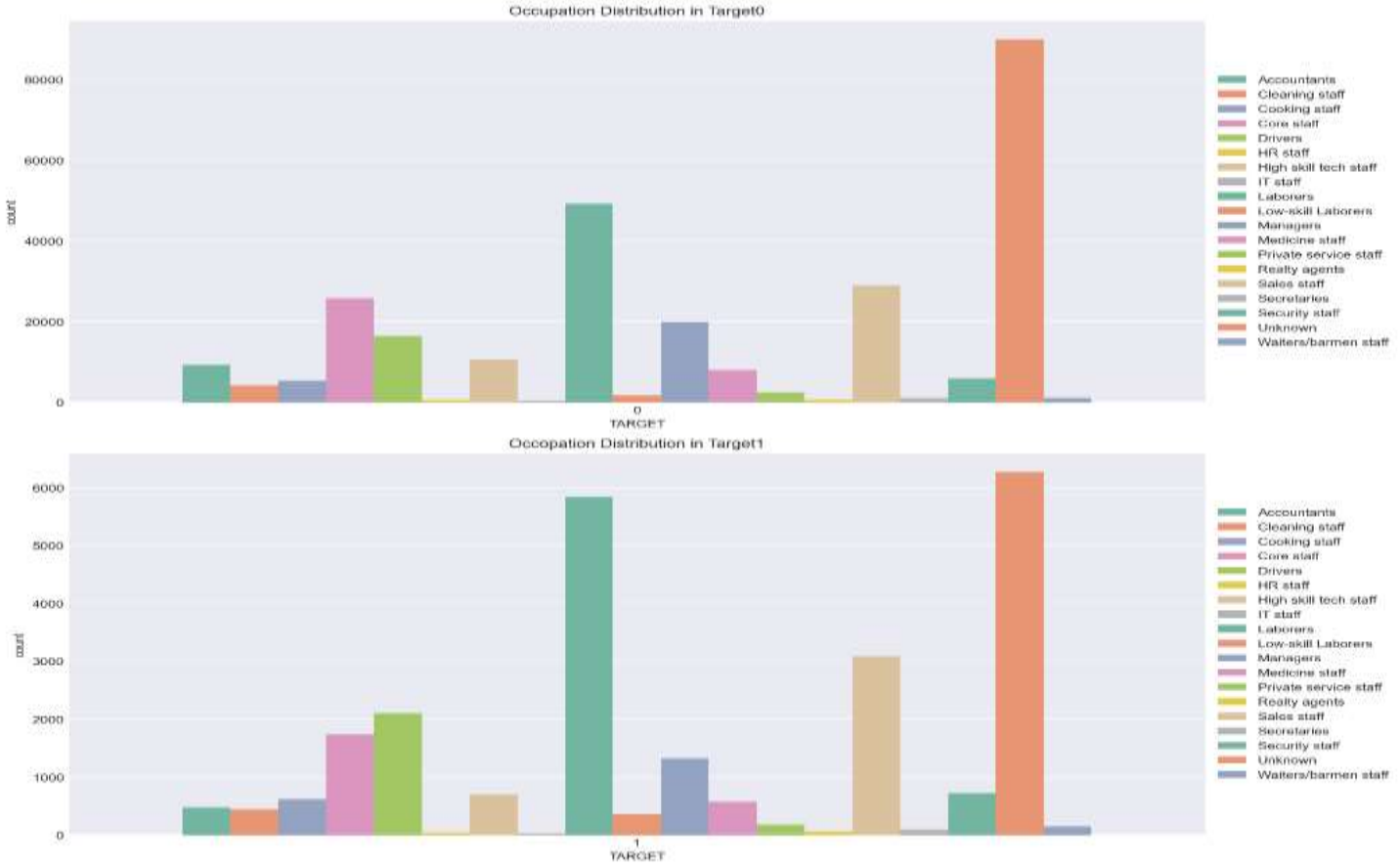
- The middle age group (35-60 years) has the highest number of loan applications compared to other age groups, both for defaulters and non-defaulters.Interestingly, the middle age group also faces the most payment difficulties.
- On the other hand, senior citizens (60-100 years) and very young individuals (19-25 years) experience fewer payment difficulties compared to other age groups.

Bivariate Analysis - Income



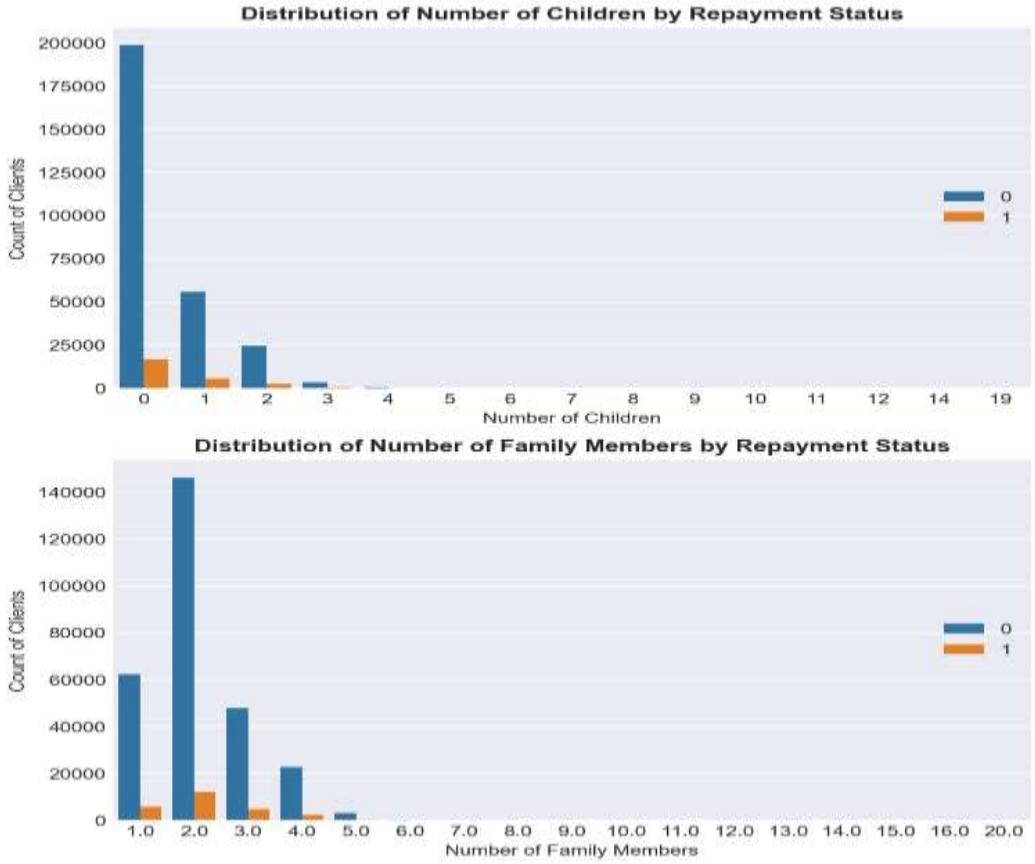
- Applicants who are businessmen, students, or unemployed are unlikely to apply for a loan.
- The working category exhibits a higher risk of loan default.
- State servants have a minimal risk of loan default.

Bivariate Analysis – Occupation



- Labours are clearly having some difficulty while making payment
- Around 96000 people haven't filled their occupation and a lot of them have pension as income type

Bivariate Analysis – Family Members



plot 1

The majority of clients, regardless of their repayment status, have no children.

Clients with more than two children do not experience significant difficulties in making payments.

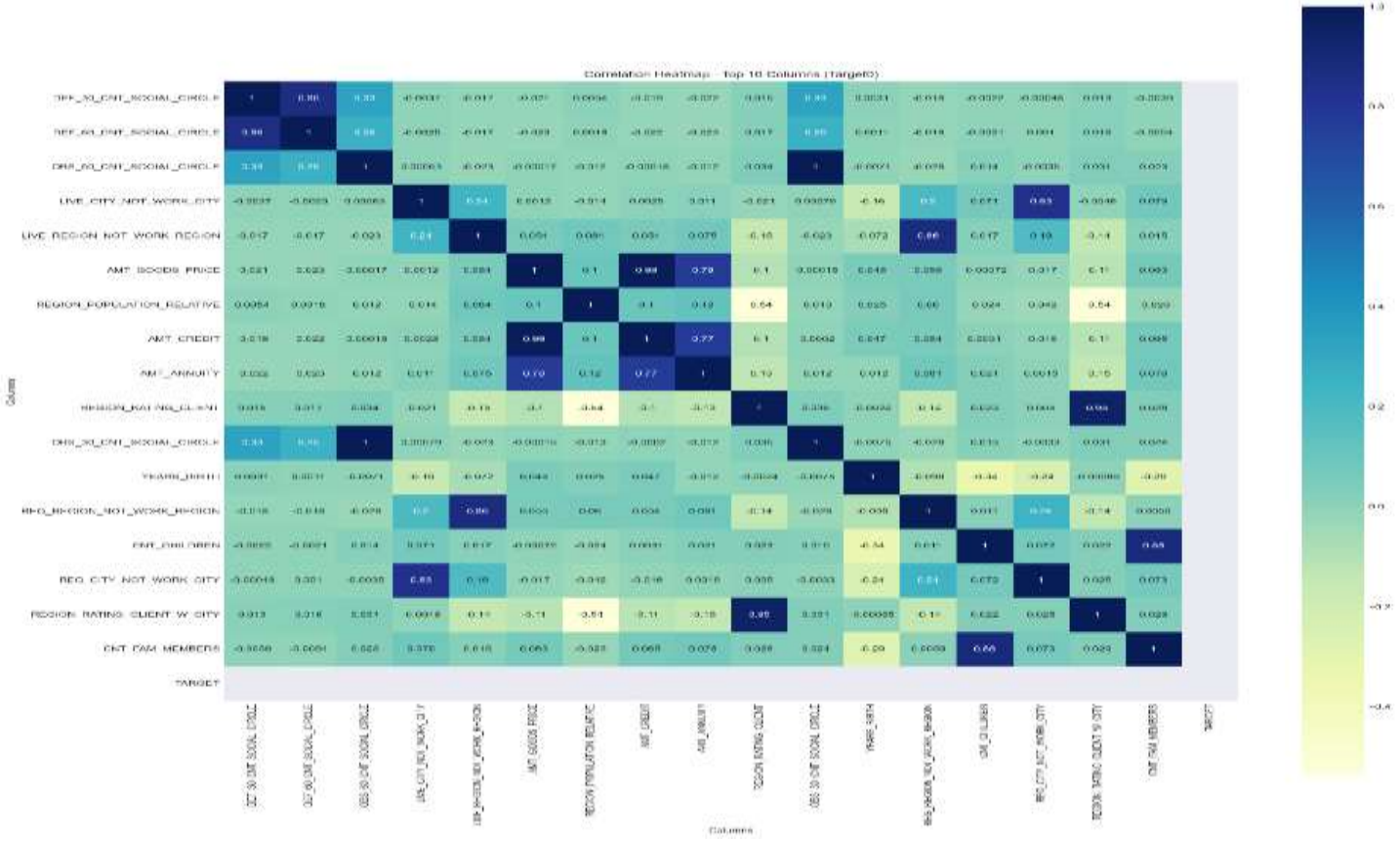
Clients with no children form the majority of those facing challenges in payment.

plot 2

Clients living with two family members are the most prevalent group, irrespective of their repayment status.

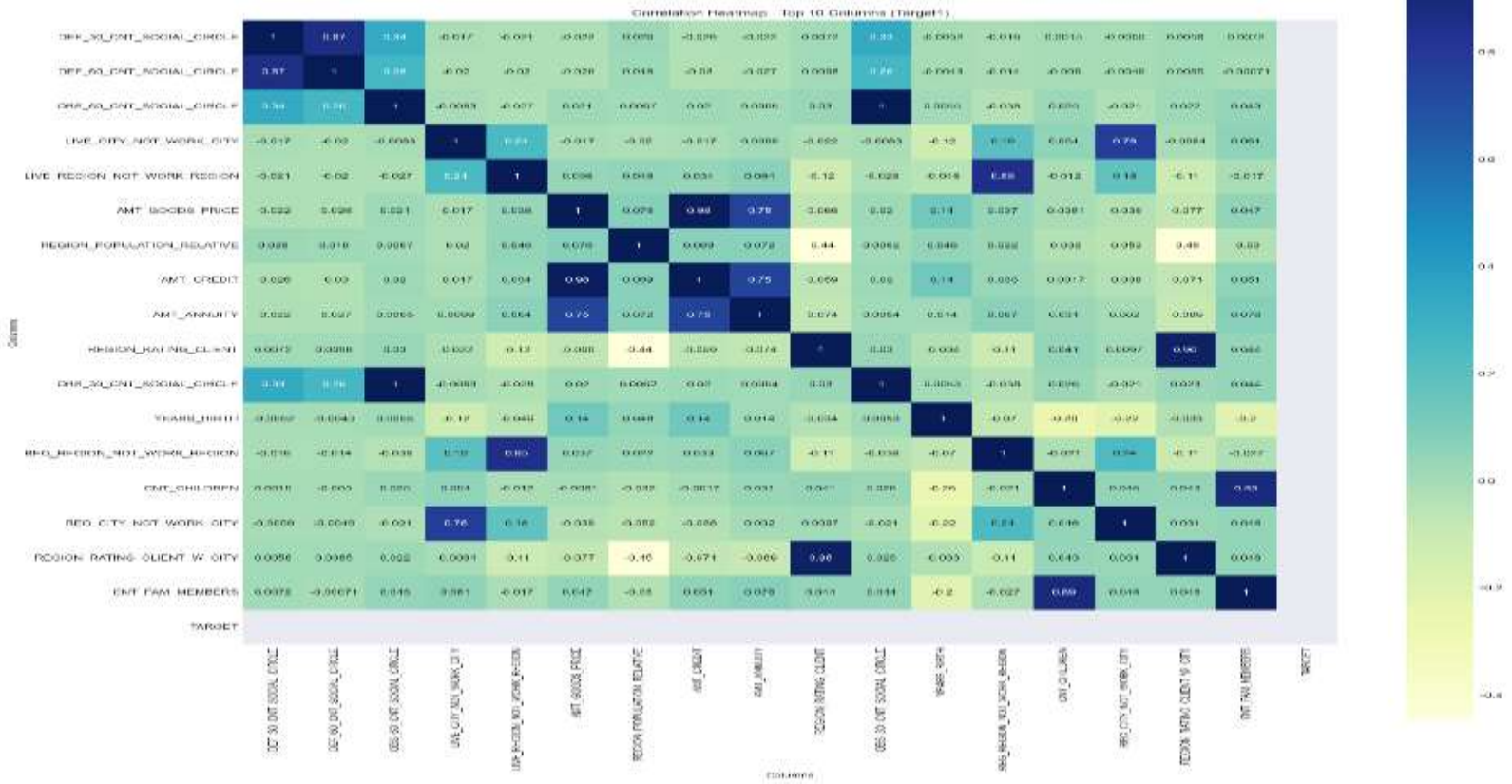
Additionally, as mentioned earlier, the majority of clients experiencing payment difficulties have two family members.

Multivariate: Correlation in target0



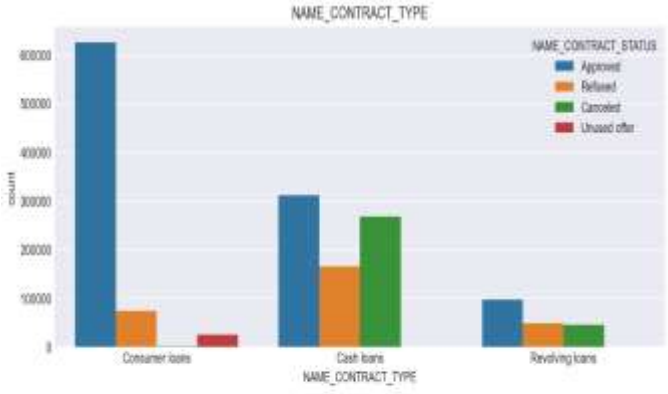
- There is a strong positive correlation between the amount credited and the price of goods.
- CNT_CHILDREN and CNT_FAM_MEMBERS exhibit a strong positive correlation, indicating that clients with children are highly likely to have more family members.
- There is a positive correlation between the annuity amount and both the price of goods and the loan amount.
- If the contact address of a client does not match their work address, it is highly likely that their permanent address also does not match the work address.
- Rating is negatively related to region population

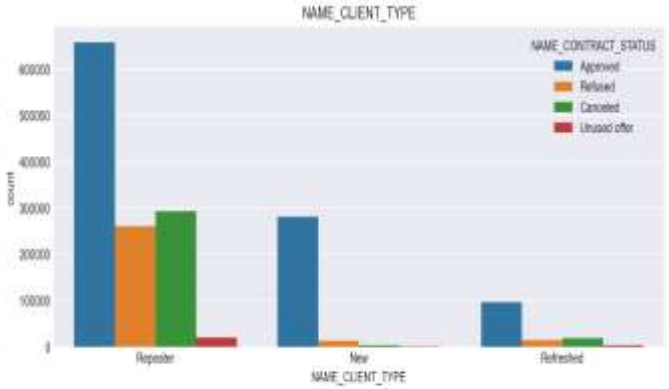
Multivariate: Correlation in target1

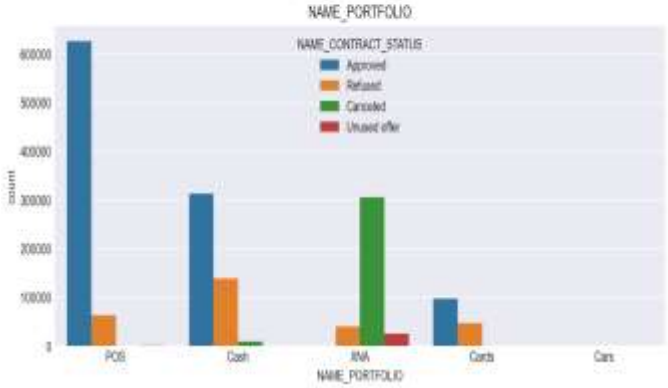


- We can see similar sort of Correlation in Target 1 also

Previous Application - Bivariate



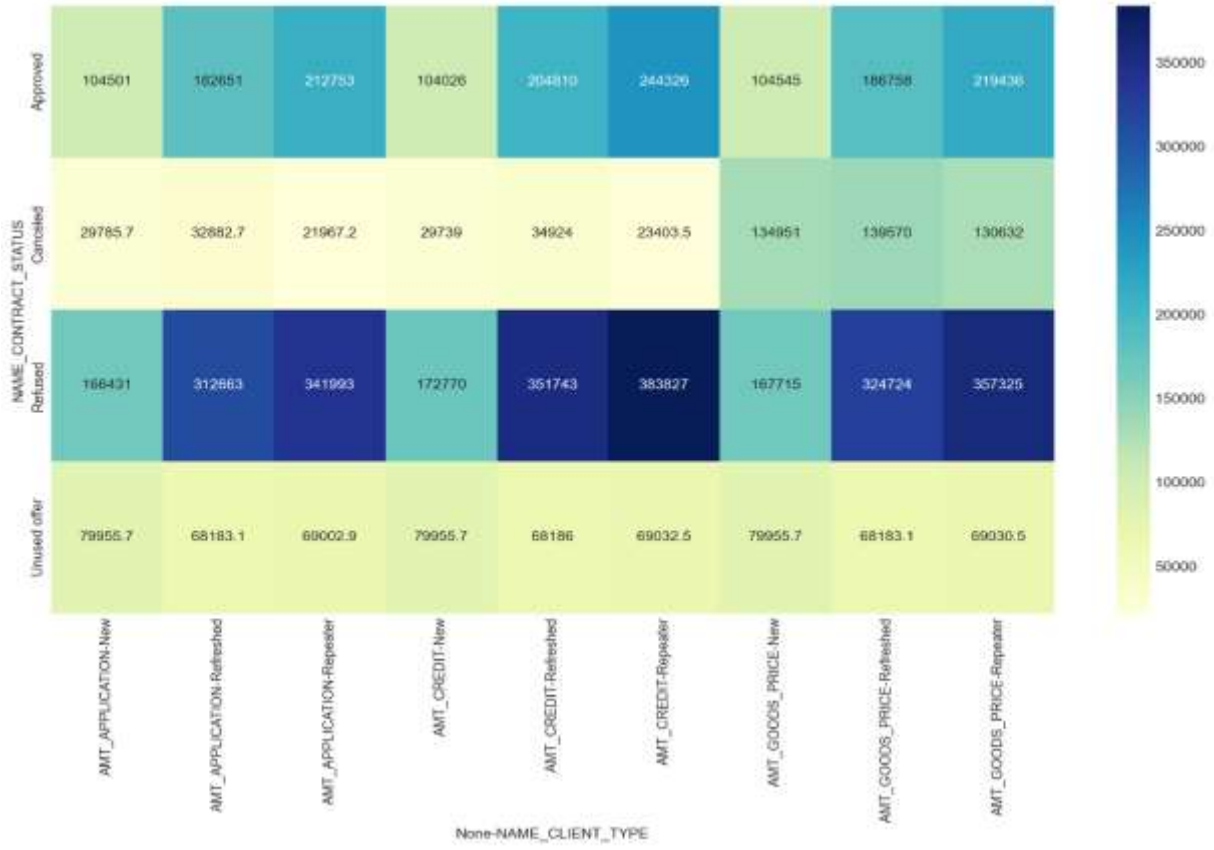




- The bank has a greater number of repeat customers in all categories, including approved, refused, unused, and cancelled loans.
- Transactions related to point-of-sale (POS) also appear to be consumer loans, and similar to the previous point, there is a higher number of cash loans being refused compared to POS loans.

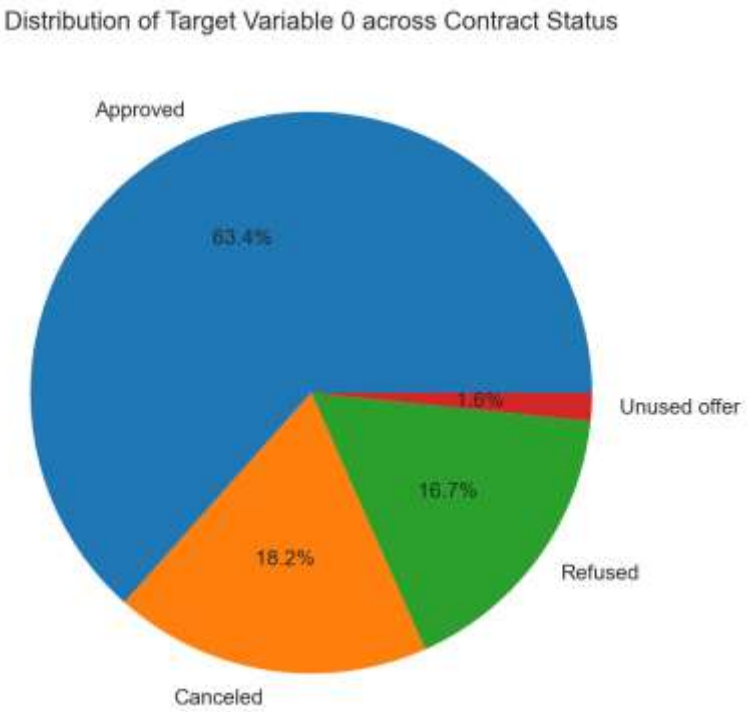
- Among the approved loan applications, the consumer loan category has the highest number of applicants.
- It appears that there are no cancelled loans in the cash loan category compared to the consumer loan category.
- The number of cash loans that have been refused is higher than the number of consumer loans refused.

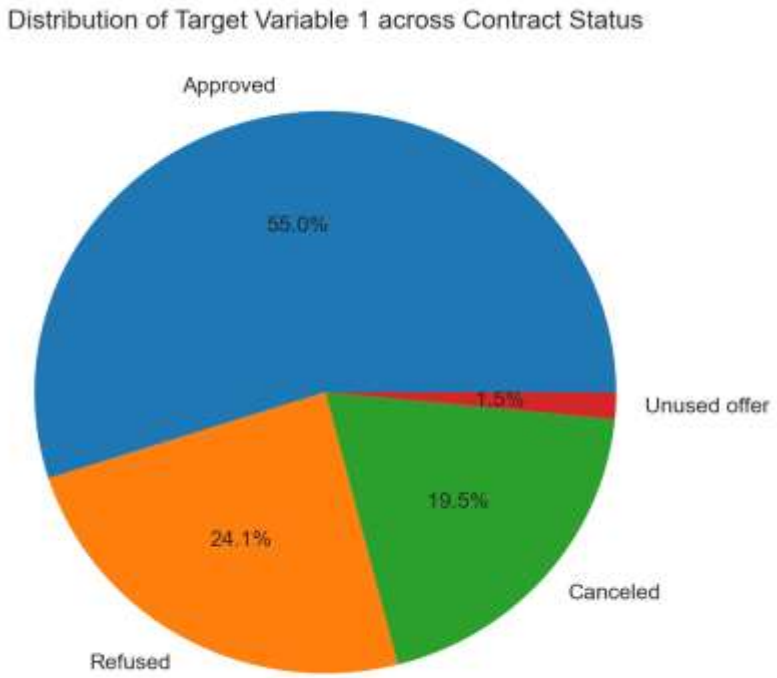
Previous Application - Multivariate



- The amount of cancelled applications is notably high. It is possible that the bank refuses these applications due to concerns about the consumer's high debt-to-liability ratio resulting from the large application amount. This could indicate a higher credit default risk associated with such applications.
- Repeat applicants tend to have higher application amounts compared to new customers. This observation suggests that the bank may have more favorable policies or interest rates for repeat applicants, incentivizing them to apply for larger loan amounts.
- The data shows that all cases that were cancelled or refused have higher values of goods compared to other categories.

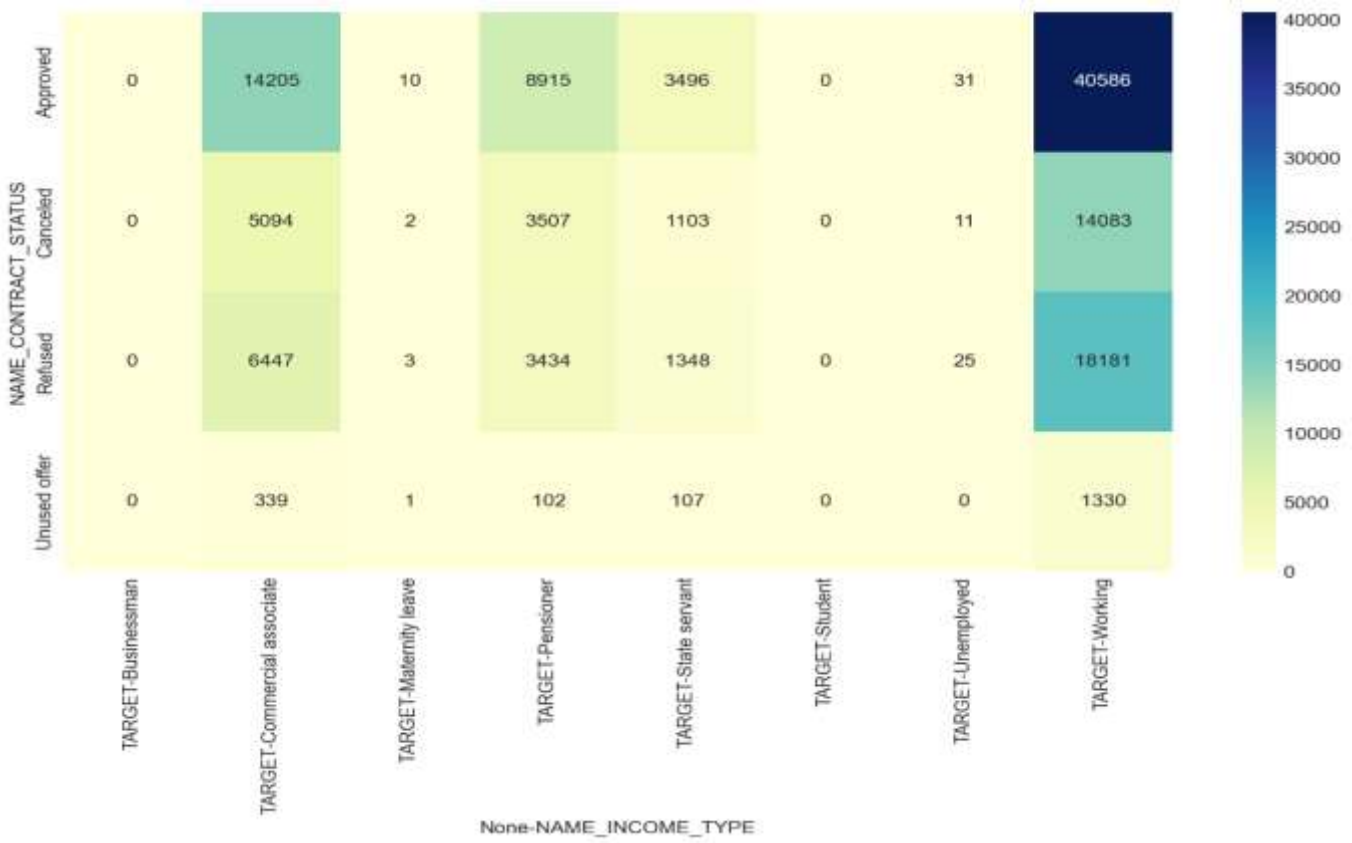
Merged Data





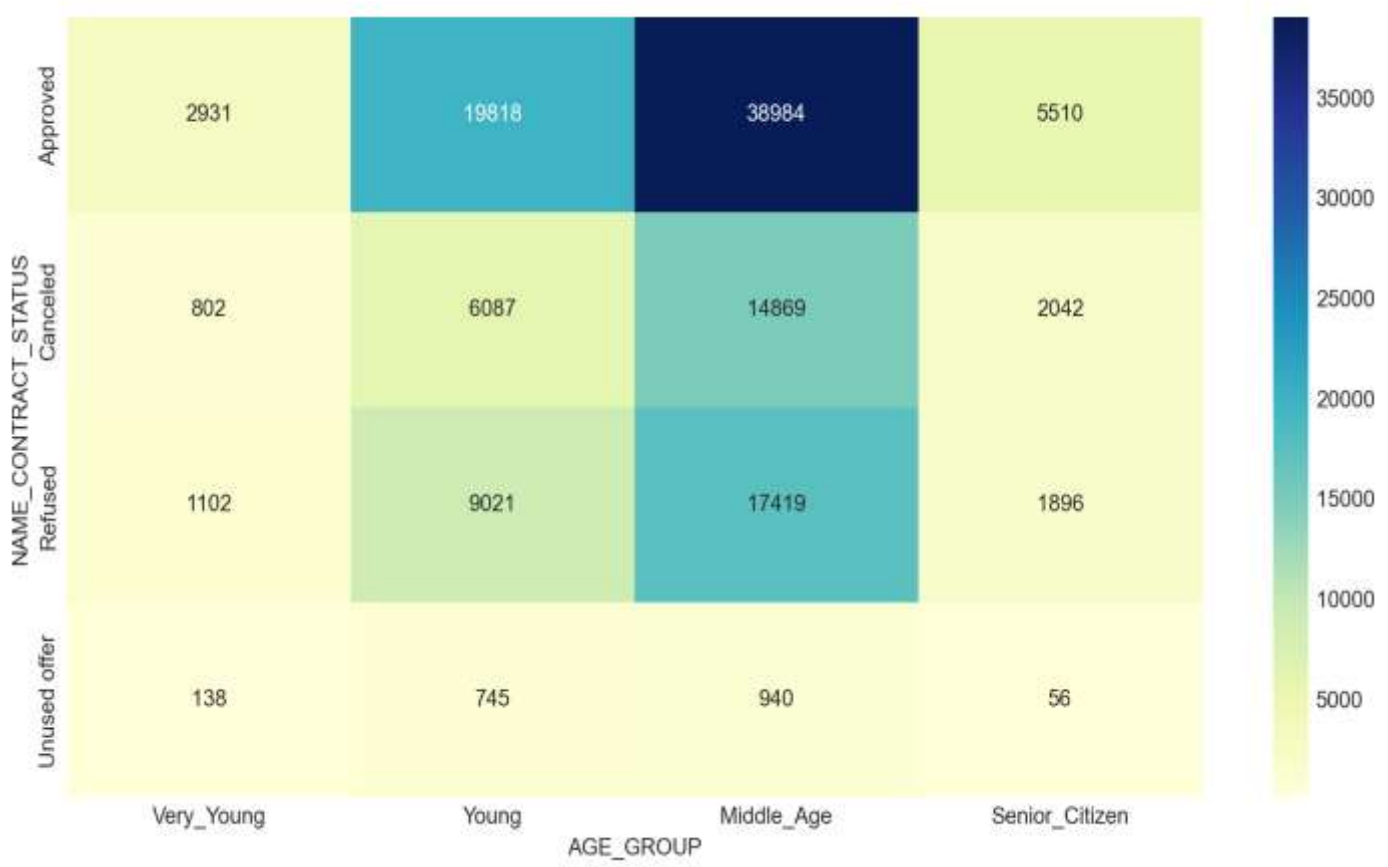
- 45% of population who were previously refused, cancelled or have unused offer have now defaulted their loan

Merged Data - Multivariate



- Applicants with an approved status who are currently employed have a higher number of defaults. Default cases are observed in previous applications that were refused, cancelled, or unused, raising concerns about the financial company's approval decisions. Among the working-class applicants, 18,181 individuals who were previously refused have now defaulted on their loans.

Merged Data - Multivariate



- Applicants in the age groups of 25-35 and 35-60 with approved loans exhibit higher default rates. Default cases are observed in current applications where previous applications were refused or cancelled, suggesting a potential link between previous loan outcomes and current defaults.

Conclusion

When comparing the percentages of individuals facing payment difficulties and those without payment difficulties, we notice a decline in the proportion of pensioners and state servants experiencing payment difficulties, while there is a rise in the percentage of working individuals facing payment difficulties.

The bank should prioritize targeting working applicants to a lesser extent, as they have the highest number of defaulters. While it is important to consider the higher number of defaulters among working applicants, it does not necessarily mean that they should be automatically refused. Instead, it is crucial to conduct a thorough assessment of various parameters and factors to make informed decisions regarding loan approvals.

Conclusion

When comparing the percentages of individuals facing loan payment difficulties and those without payment difficulties, we notice a rise in the proportion of individuals with secondary/secondary special educational qualifications among the group experiencing loan payment difficulties. Conversely, there is a decline in the percentage of individuals who have completed higher education among those facing loan payment difficulties when compared to both the group with loan payment difficulties and the group without payment difficulties.

It can be beneficial for banks to prioritize targeting clients who own a car or a house.

Out of the population who were previously refused, cancelled, or had an unused loan offer, 45% of them have now defaulted on their loan.