# CREDIT EDA CASE STUDY ASSIGNMENT

By: Aditi Marwaha

# INTRODUCTION

This case study is designed to provide you with a practical application of Exploratory Data Analysis (EDA) within a real-world business context.

Throughout this study, you will not only apply the EDA techniques you've learned but also gain a fundamental understanding of risk analytics in the banking and financial services sector. Additionally, you'll learn how data analysis is leveraged to mitigate the risk of financial losses when extending loans to customers.

# BACKGROUND AND PROBLEM STATEMENT

## Background:

We've received a dataset related to the banking or non-banking financial institution (NBFC) sector. This dataset primarily includes:

1.The loans extended by the institution.

2.A critical variable labeled as “TARGET," which categorizes clients into either defaulters or regular, good clients based on their payment behavior.

3.Several other variables collected during the loan approval process.

## Problem Statement:

The dataset encompasses details of loan applications submitted at the time of requesting a loan. It has two distinct scenarios:

1.Clients facing payment difficulties: These clients have experienced late payments exceeding X days on at least one of the initial instalments. (They are denoted as “TARGET = 1.")

2.All other instances: This category includes cases where payments are made punctually. (They are represented as “TARGET = 0.")

# OBJECTIVE

This case study is intended to uncover patterns that can help determine whether a client faces difficulties in repaying their instalments. These insights can inform various actions, such as loan denial, reducing the loan amount, or offering loans to high-risk applicants with higher interest rates, among others. The goal is to ensure that those capable of repaying the loan are not unfairly rejected. The primary objective of this case study is to use Exploratory Data Analysis (EDA) to identify these applicants effectively.

In simpler terms, the study aims to pinpoint the key factors, often referred to as driver variables, that strongly indicate loan default. This knowledge can then be applied to assess and manage the portfolio and risk effectively.

# ANALYSIS FLOW

Given the problem statement, try to identify the most reliable predictors of defaulter behavior within the provided dataset. The approach involves the following key steps:

1.Data Loading and Cleansing: Employ Pandas libraries to load the data and carry out initial data cleansing.

2.Exploring Distributions: The focus will be on comprehending the various data distributions.

3.Data Visualization: Use various visualization libraries to create graphs that reveal patterns and anomalies within the dataset.

4.Identifying Variables with High Null Values: Identify variables with over 50% missing data since such variables tend to have reduced predictive power concerning the “TARGET" variable. Consequently, these variables will be excluded from the subsequent analysis.

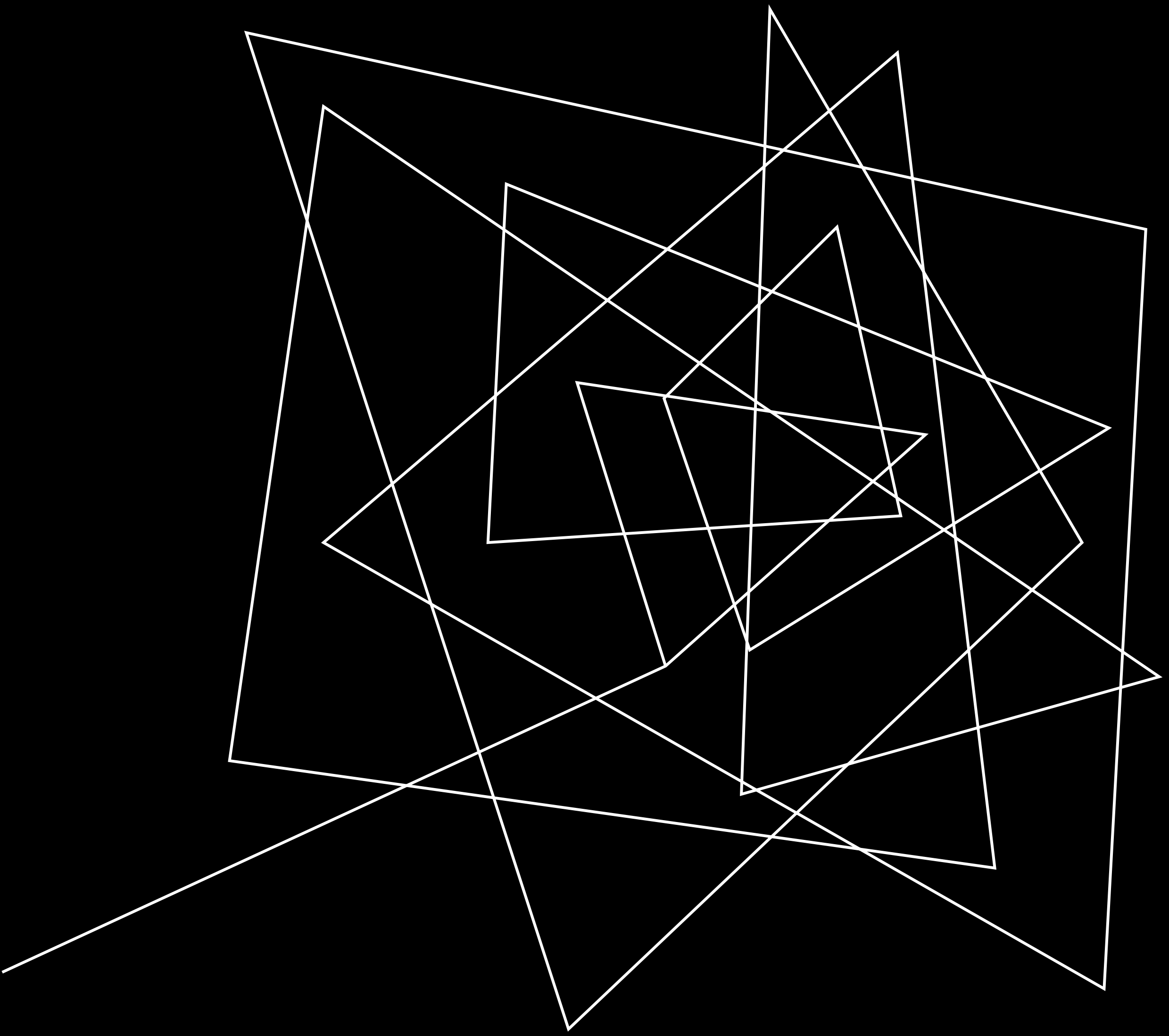
5.Variable Analysis: Among the remaining variables, conduct a thorough analysis, considering them either as groups of related variables or as individual variables to assess their impact on the “TARGET" variable.

6.Data Cleaning: The data for the retained variables will be cleaned, and univariate and bivariate analysis will commence.

7.Data Integration: Merge the cleansed application dataset with the previous application dataset.

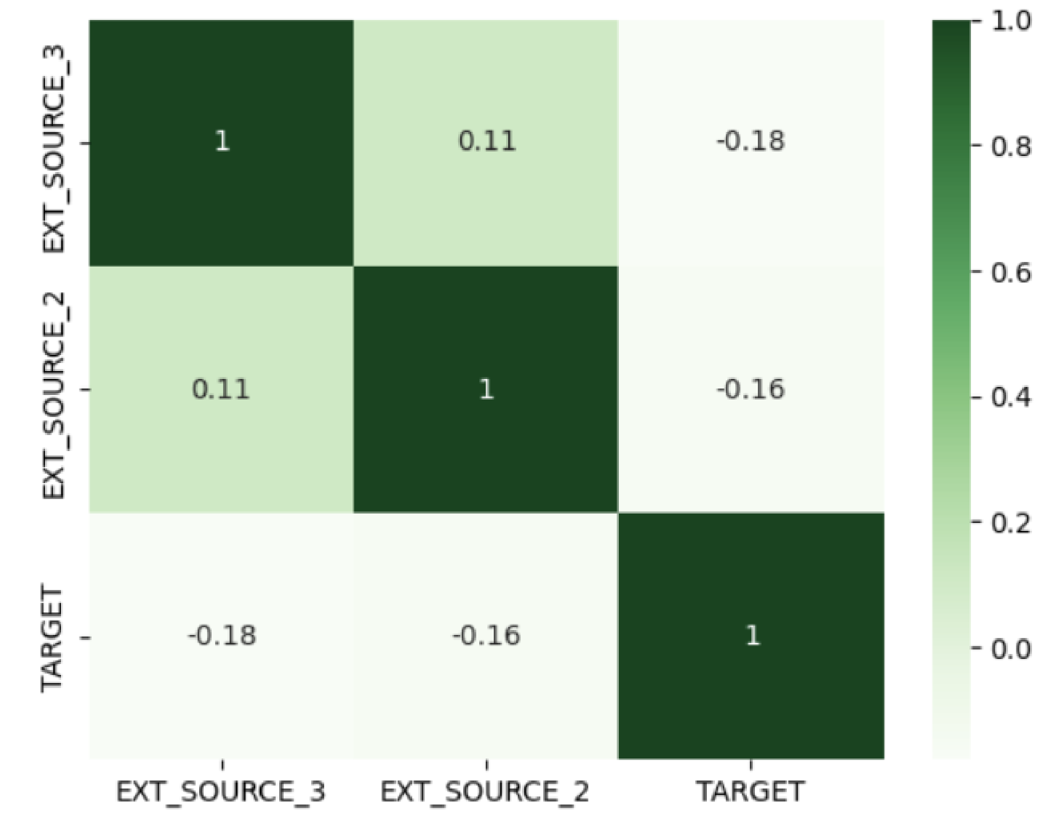
8.Correlation Analysis: Identifying the pertinent variables for correlation analysis will be the next step, leading us to determine the top-performing variables.

Through this comprehensive process, the aim is to pinpoint the most significant variables for predicting defaulter behavior.



# APPLICATION DATASET CLEANING

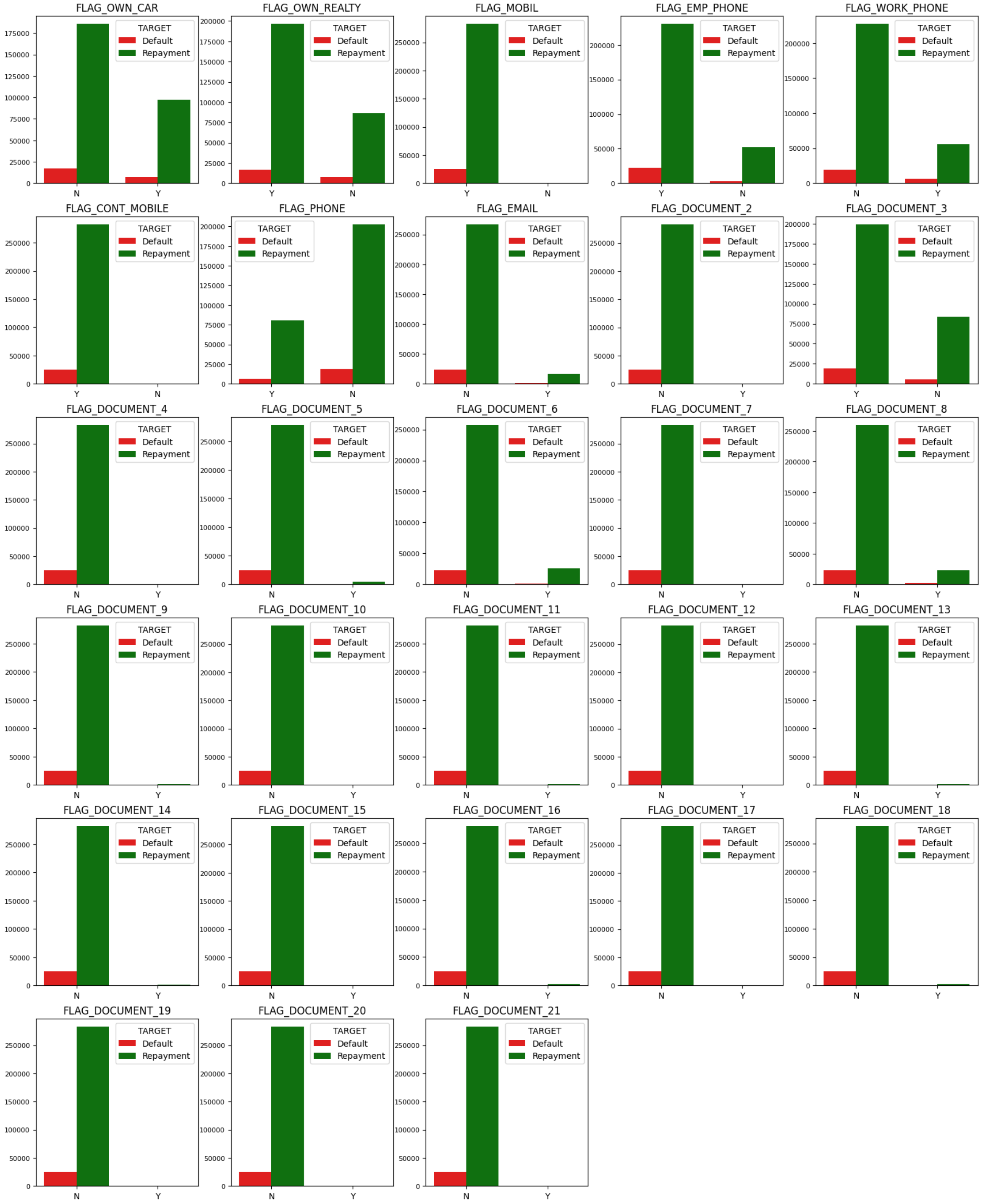
Establishing a correlation between "EXT_SOURCE_3", "EXT_SOURCE_2" and "TARGET"



'EXT_SOURCE_2', 'EXT_SOURCE_3', and 'TARGET' have no linear correlation and won't cause any errors. Hence, 'EXT_SOURCE_2' and 'EXT_SOURCE_3’ are dropped from dataset.

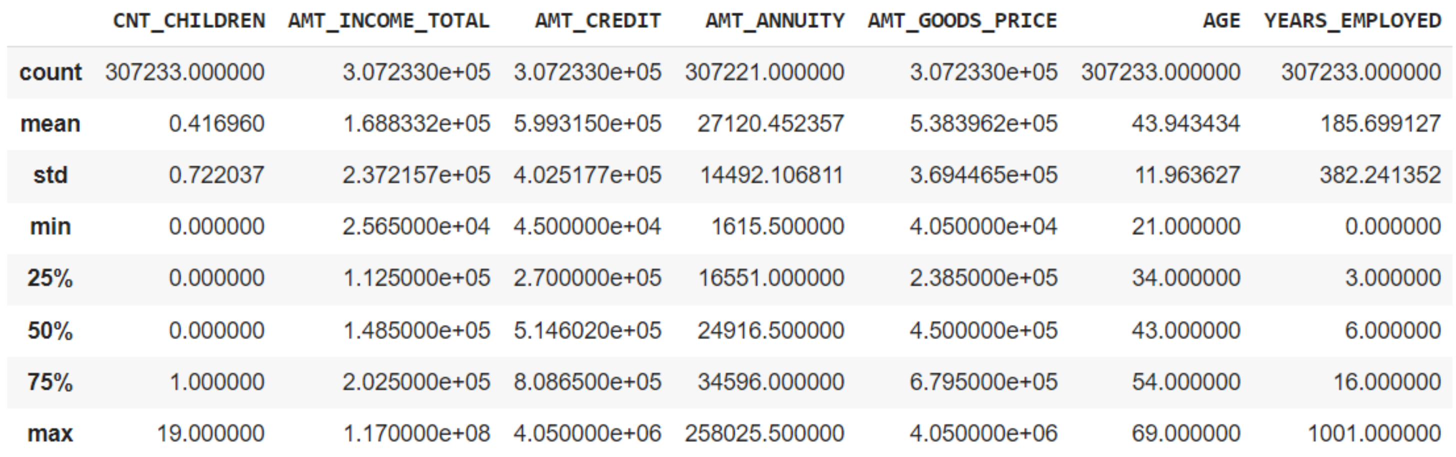
# Column analysis having FLAG.

Upon closer examination of these columns, the majority of columns’ potential contribution to the analysis remained uncertain. Consequently, a detailed analysis was conducted and was concluded that most of the columns can be safely eliminated. The columns that were not dropped are: "FLAG_OWN_REALTY", "FLAG_MOBIL", "FLAG_EMP_PHONE", "FLAG_CONT_MOBILE", "FLAG_DOCUMENT_3“.

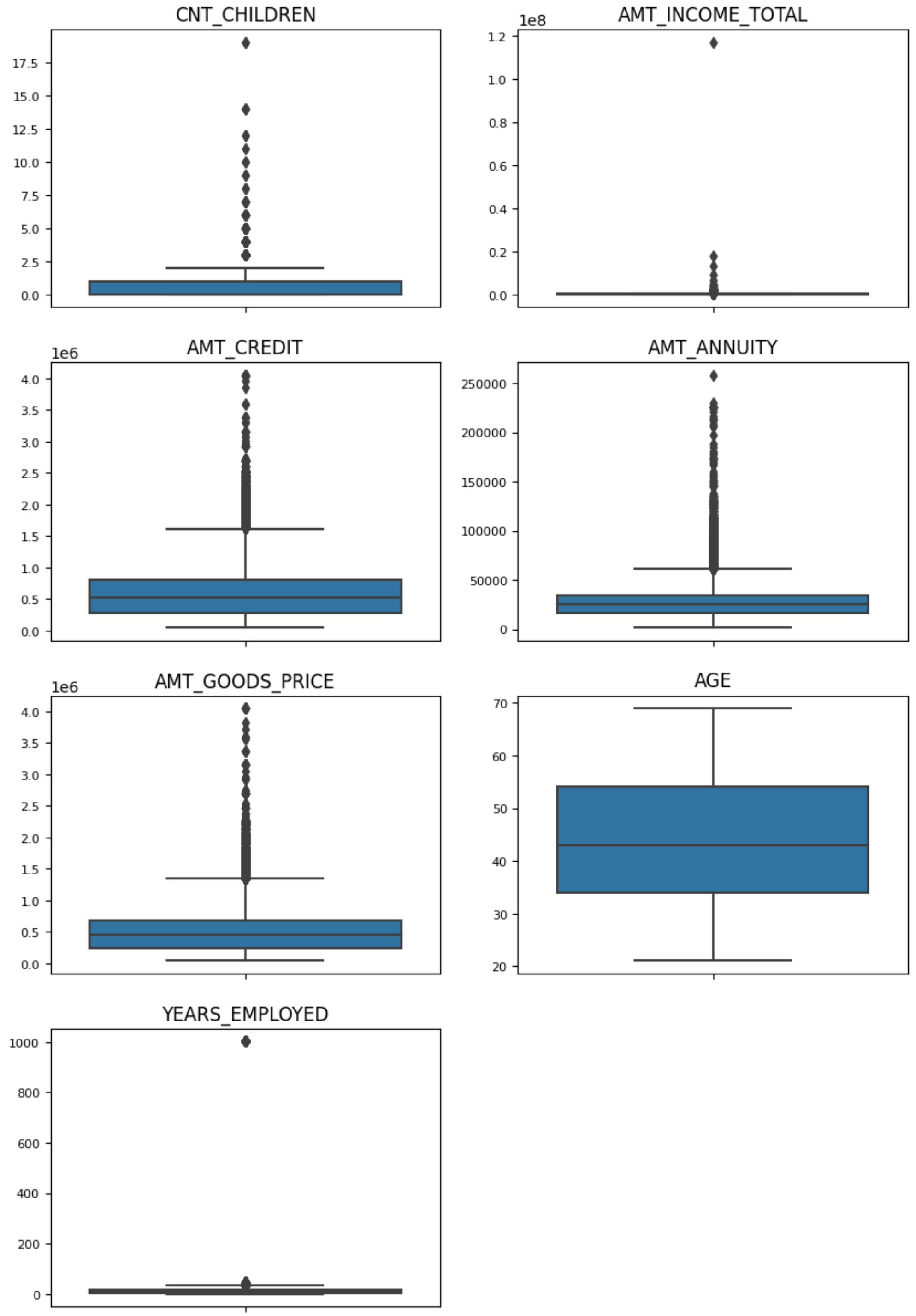


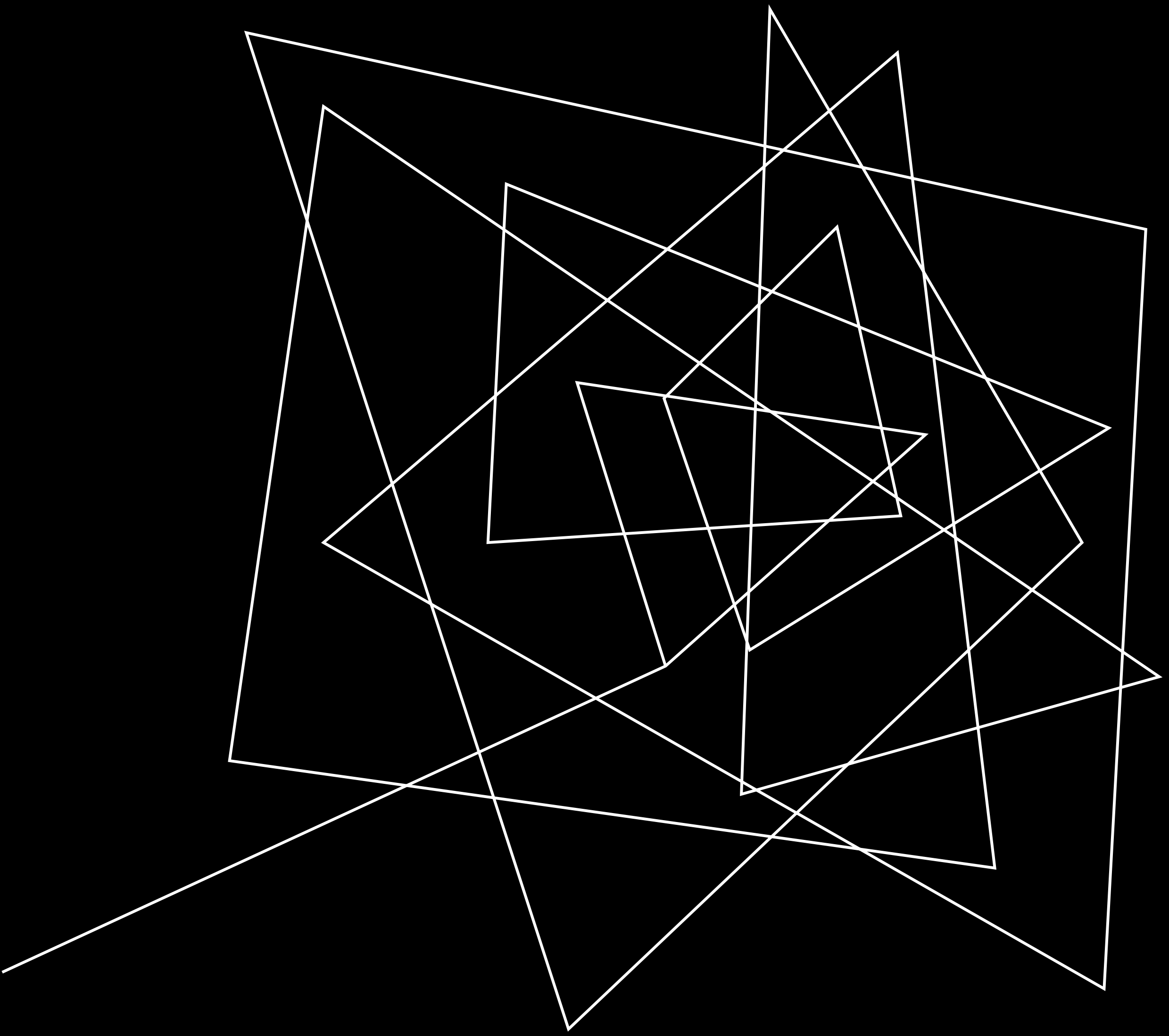
# Checking Outliers:

To address outliers in the numerical column, an examination using boxplots was conducted. In cases where the disparity between the maximum value and the 75th percentile was substantial, corrective action was taken. These outliers were either replaced with the mean values or removed from the dataset.



| | CNT_CHILDREN | AMT_INCOME_TOTAL | AMT_CREDIT | AMT_ANNUITY | AMT_GOODS_PRICE | AGE | YEARS_EMPLOYED |
|---|---|---|---|---|---|---|---|
| **count** | 307233.000000 | 3.072330e+05 | 3.072330e+05 | 307221.000000 | 3.072330e+05 | 307233.000000 | 307233.000000 |
| **mean** | 0.416960 | 1.688332e+05 | 5.993150e+05 | 27120.452357 | 5.383962e+05 | 43.943434 | 185.699127 |
| **std** | 0.722037 | 2.372157e+05 | 4.025177e+05 | 14492.106811 | 3.694465e+05 | 11.963627 | 382.241352 |
| **min** | 0.000000 | 2.565000e+04 | 4.500000e+04 | 1615.500000 | 4.050000e+04 | 21.000000 | 0.000000 |
| **25%** | 0.000000 | 1.125000e+05 | 2.700000e+05 | 16551.000000 | 2.385000e+05 | 34.000000 | 3.000000 |
| **50%** | 0.000000 | 1.485000e+05 | 5.146020e+05 | 24916.500000 | 4.500000e+05 | 43.000000 | 6.000000 |
| **75%** | 1.000000 | 2.025000e+05 | 8.086500e+05 | 34596.000000 | 6.795000e+05 | 54.000000 | 16.000000 |
| **max** | 19.000000 | 1.170000e+08 | 4.050000e+06 | 258025.500000 | 4.050000e+06 | 69.000000 | 1001.000000 |

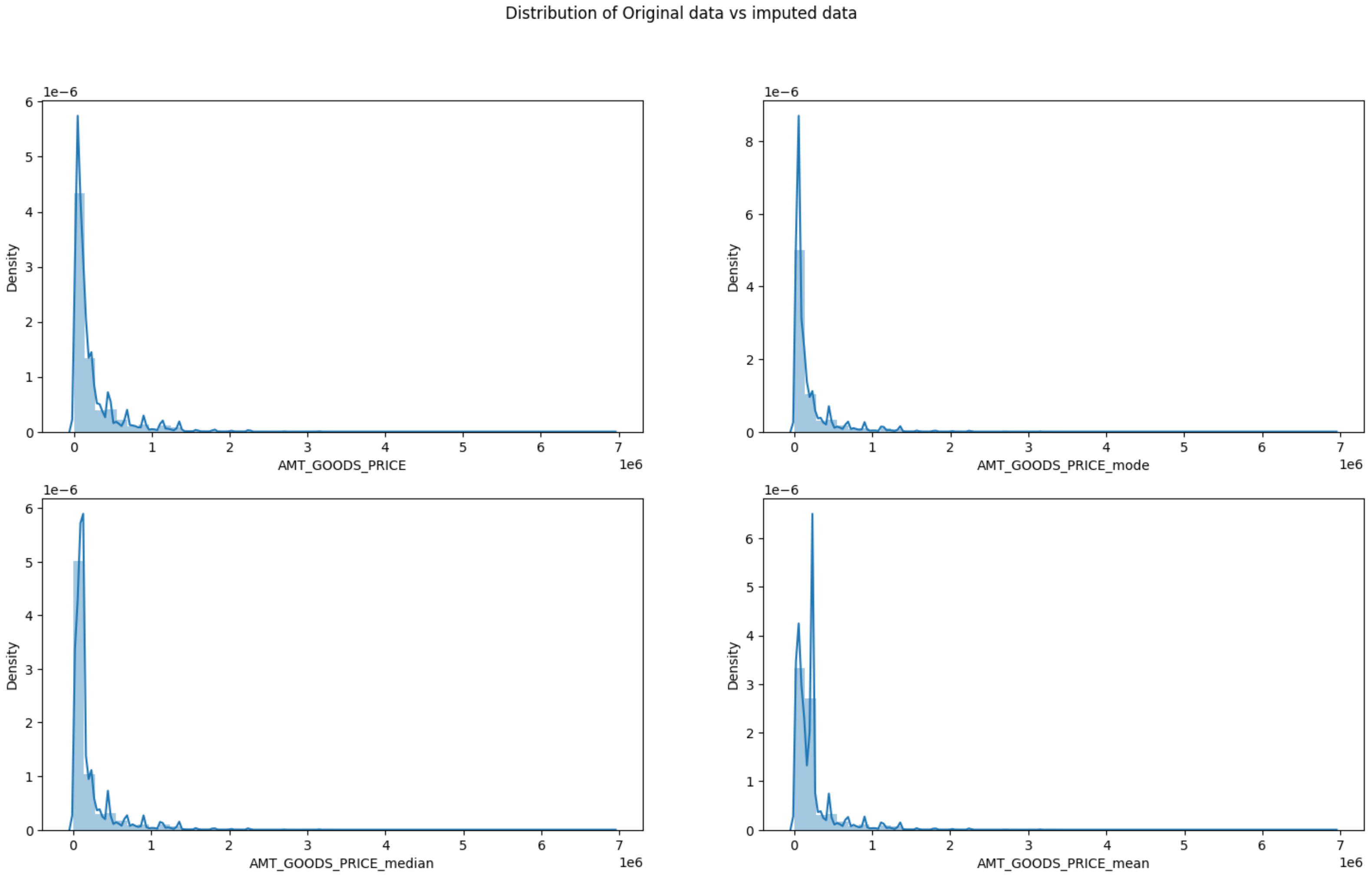




# PREVIOUS APPLICATION DATASET CLEANING

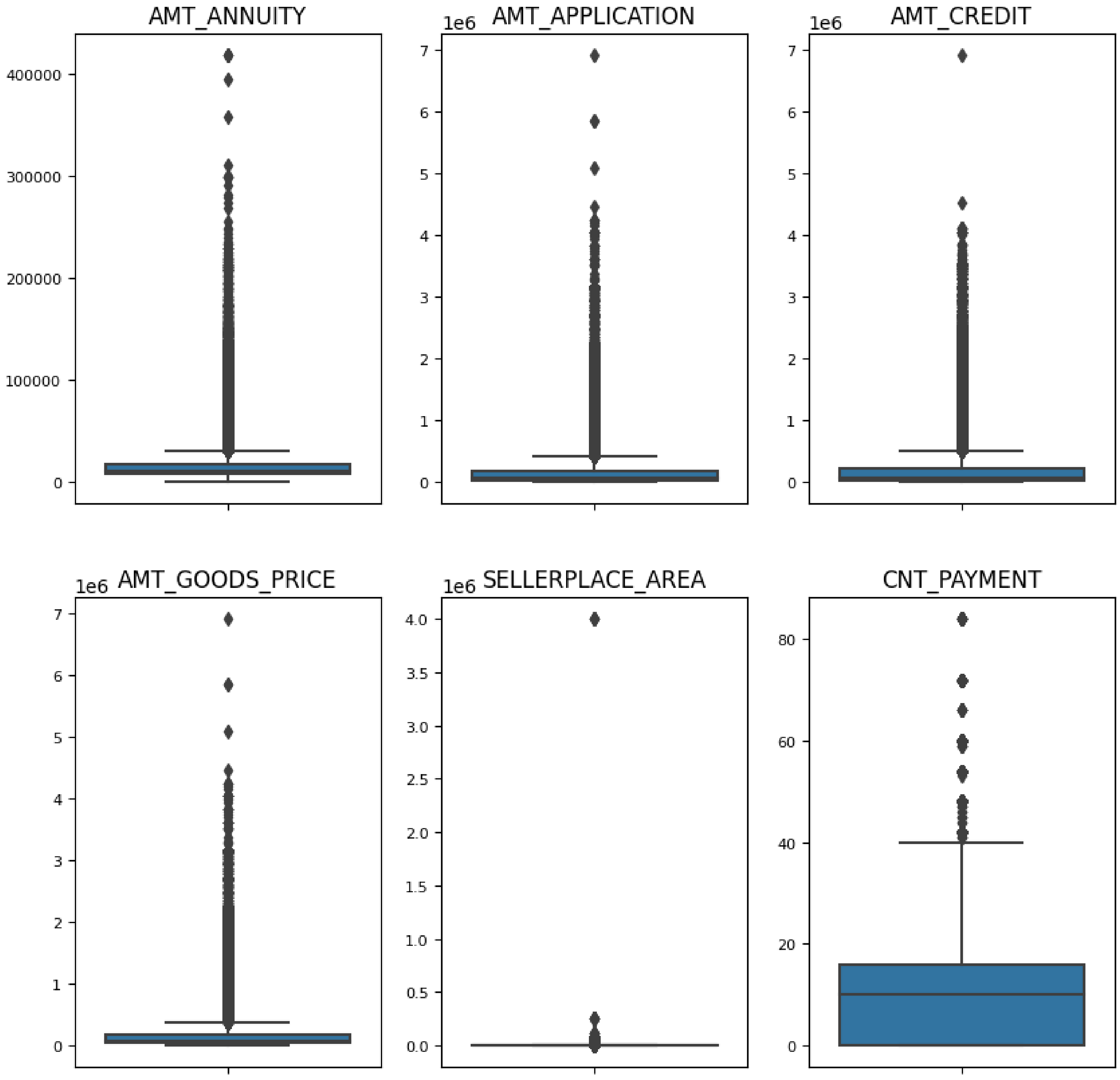
# Imputing Missing values in 'AMT_GOODS_PRICE' column

An examination and analysis of the mean, mode, and median using a kernel density estimation (KDE) plot to address missing values within the column was conducted. Upon assessing the original distribution, it closely resembled a distribution with a prominent mode. Consequently, we opted to impute missing values by utilizing the mode of the dataset.



# Checking Outliers:

AMT_ANNUITY, AMT_APPLICATION, AMT_CREDIT, AMT_GOODS_PRICE, SELLERPLACE_AREA had large number of outlier values, whereas CNT_PAYMENT has few outlier values. Considering that these values were from high-income customers, they were not deleted or changed.



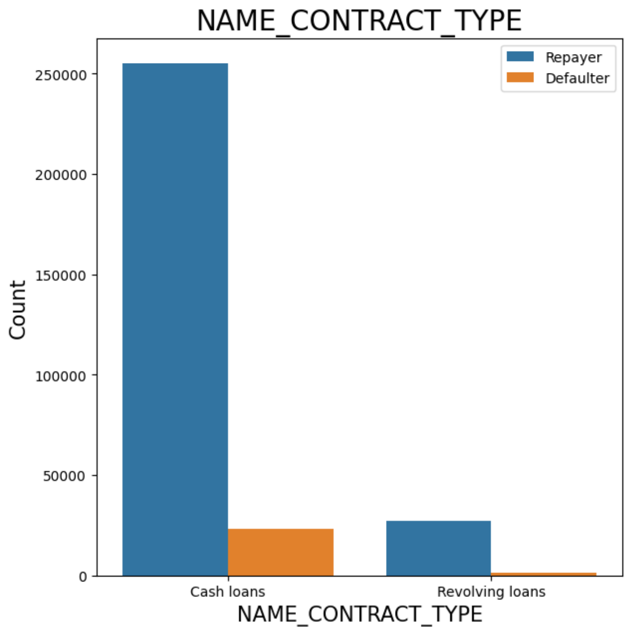
# DATA ANALYSIS

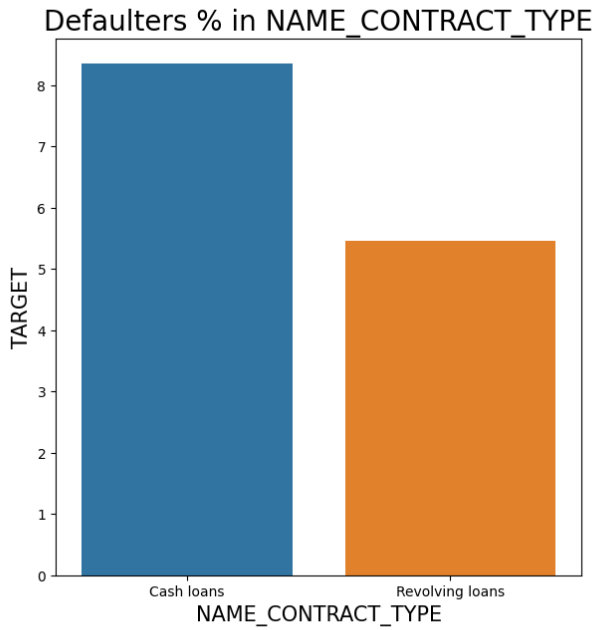
The data analysis followed the following order:

- Data imbalance check
- Categorical segmented Univariate Analysis
- Categorical Bi/Multivariate analysis
- Numeric Data Analysis
- Numerical segmented Univariate Analysis
- Numerical Bi/Multivariate analysis
- Merged Datasets Analysis

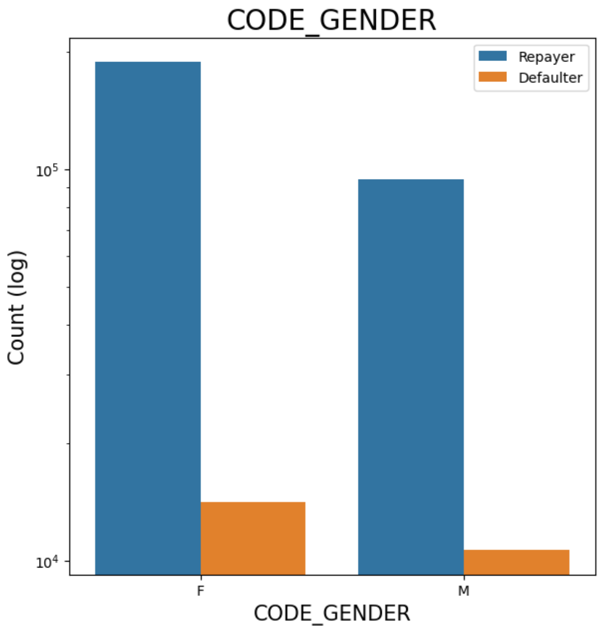
# Categorical Segmented Univariate Analysis

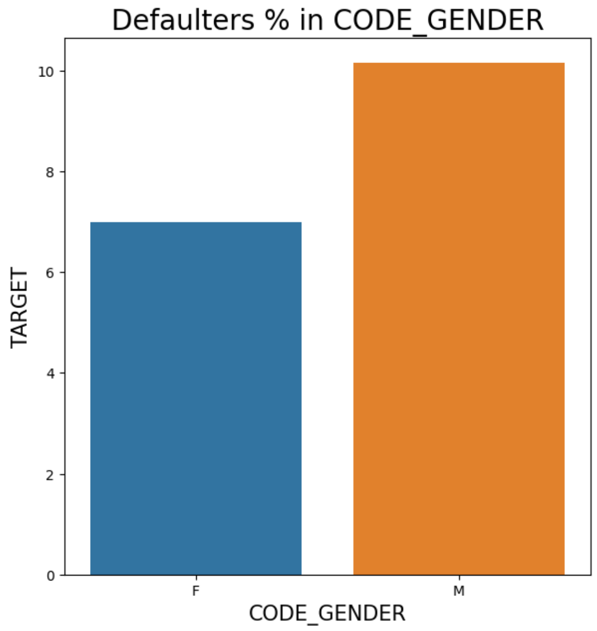
Revolving loans make up just 10% of the total loan portfolio, and among applicants for revolving loans, the default rate stands at 5-6%. Similarly, for cash loans, which constitute the majority, 8-9% of applicants default.



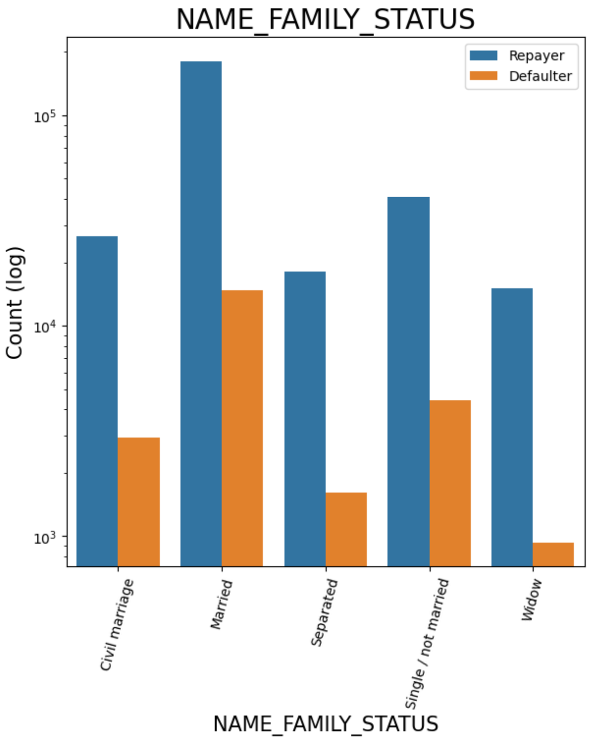


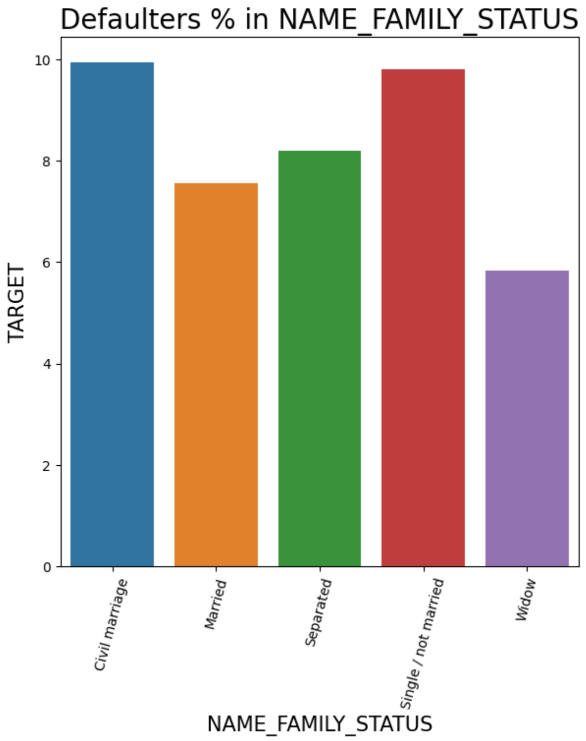
The female customer demographic comprises approximately twice as many individuals as the male customer demographic. In terms of credit default rates, men exhibit a 10% higher likelihood of failing to repay their debts compared to women.



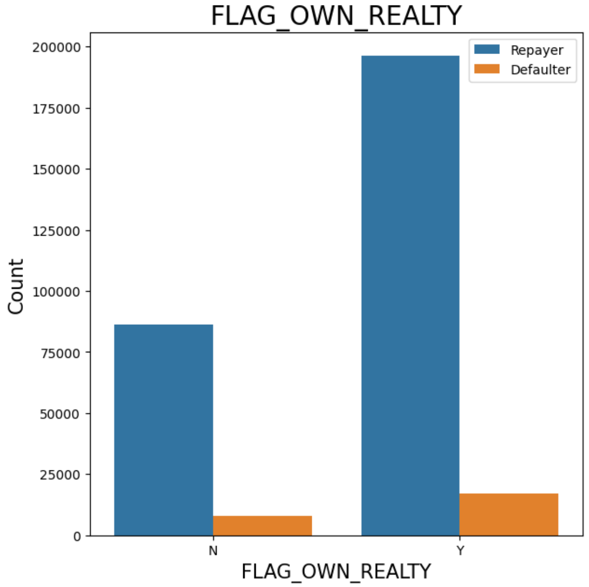


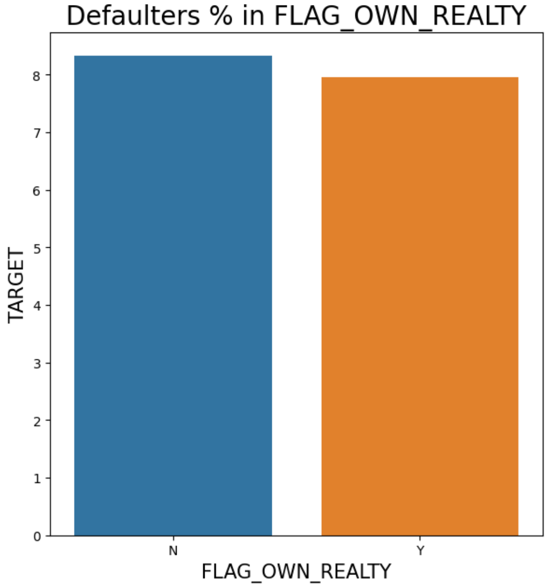
Most borrowers are married, followed by those who are single or not married, and then those in civil marriages. The default rate among borrowers in civil marriages is approximately 10%, while for widows, it stands at around 6%.



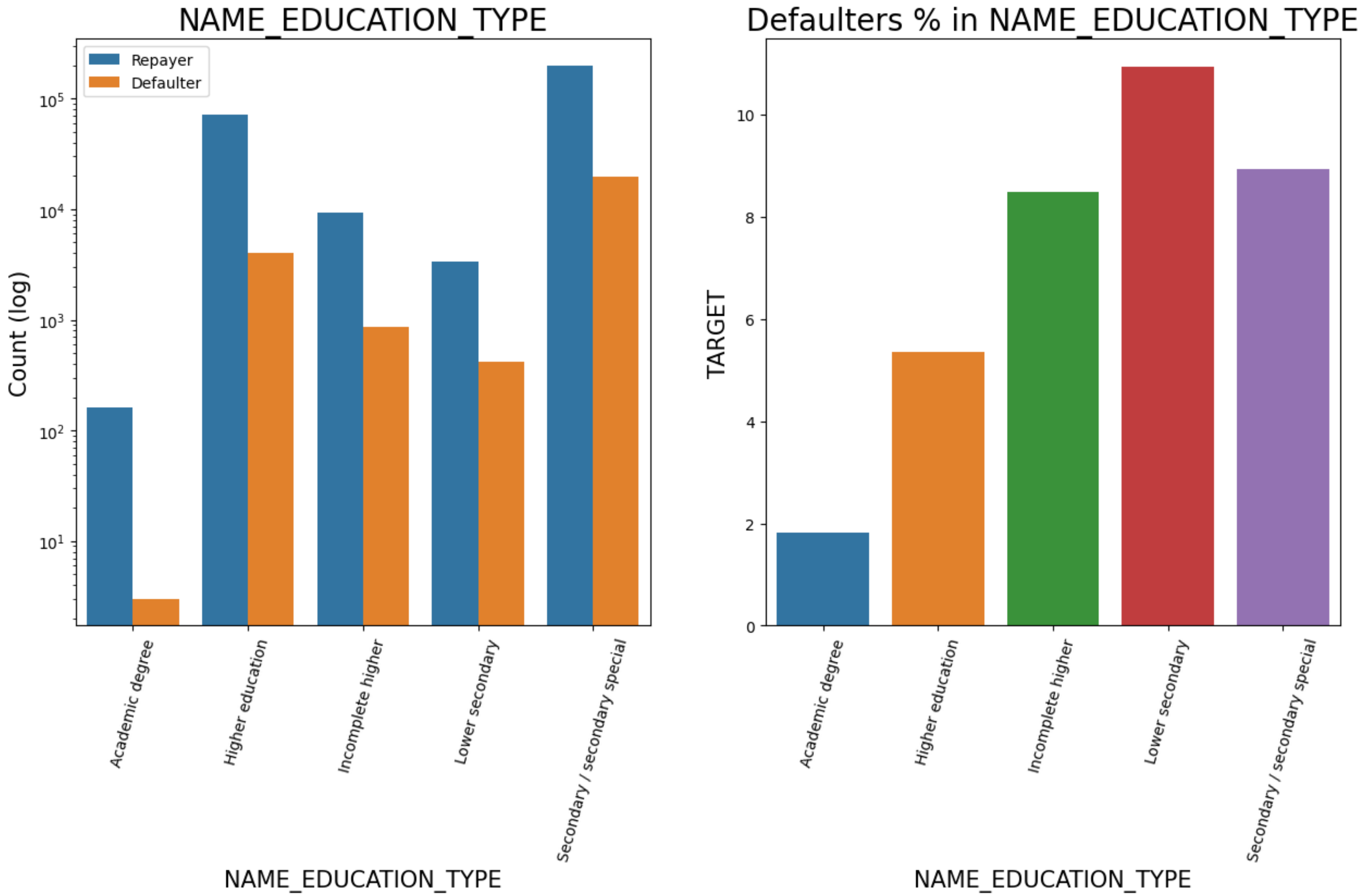


Clients who own real estate outnumber those who do not by more than a two-fold margin. Interestingly, both categories exhibit nearly identical default rates, at around 8%. Consequently, there appears to be no discernible correlation between real estate ownership and the likelihood of defaulting on a debt.

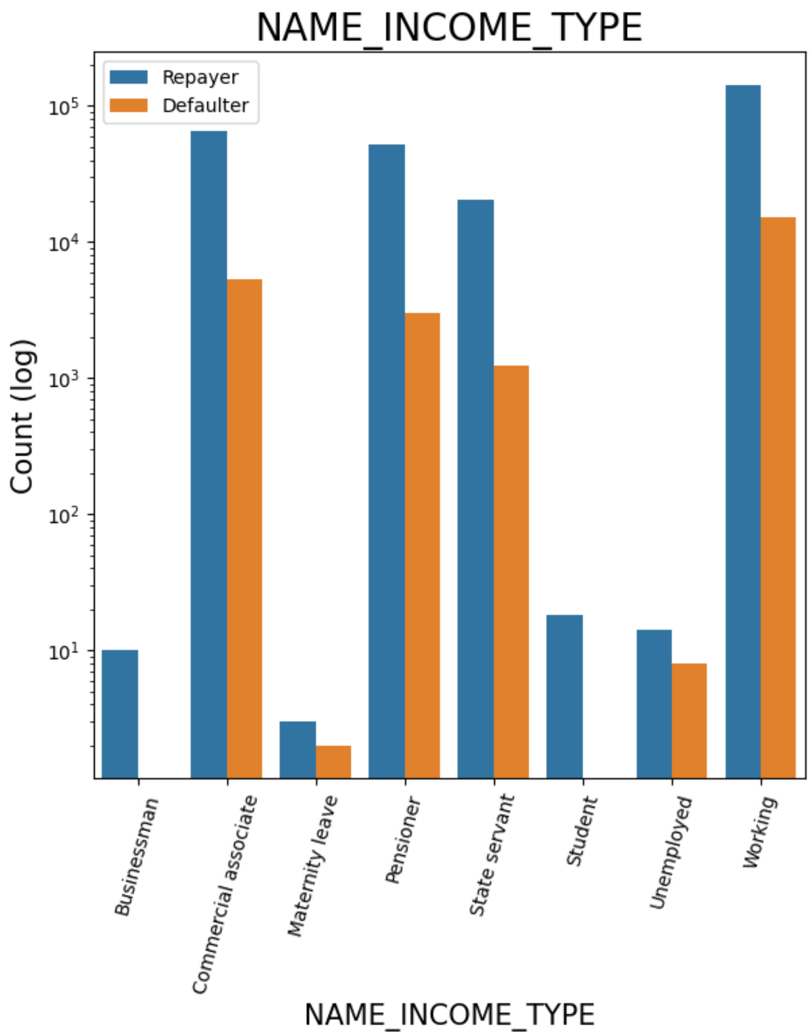


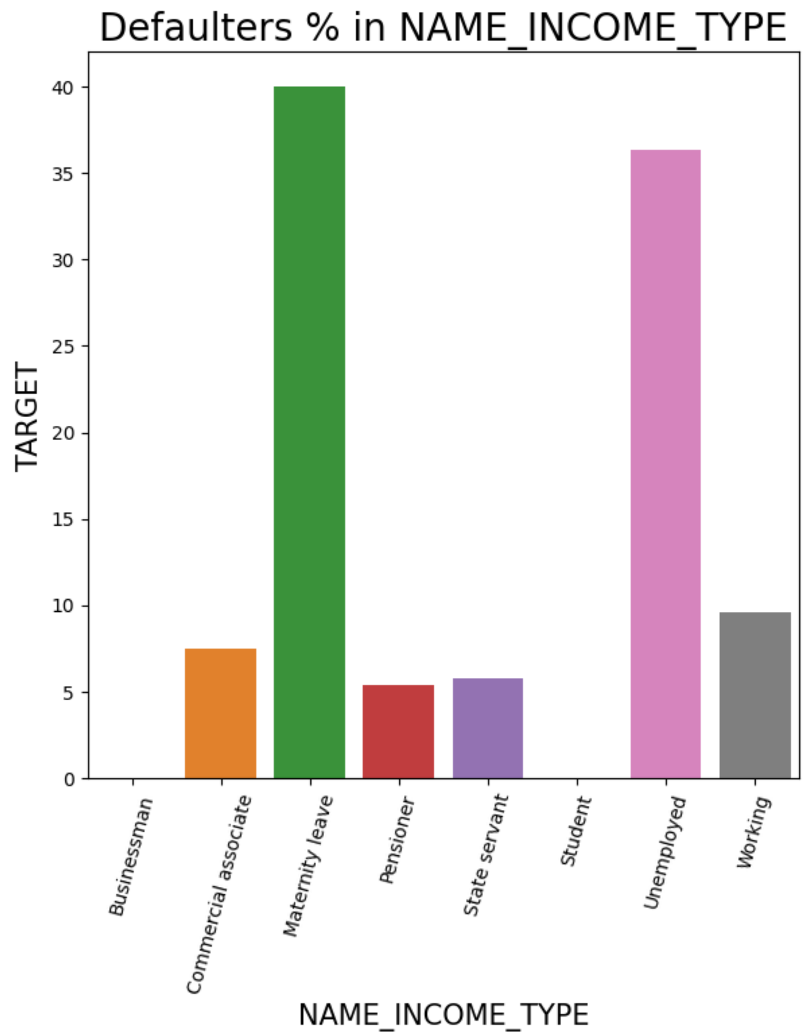


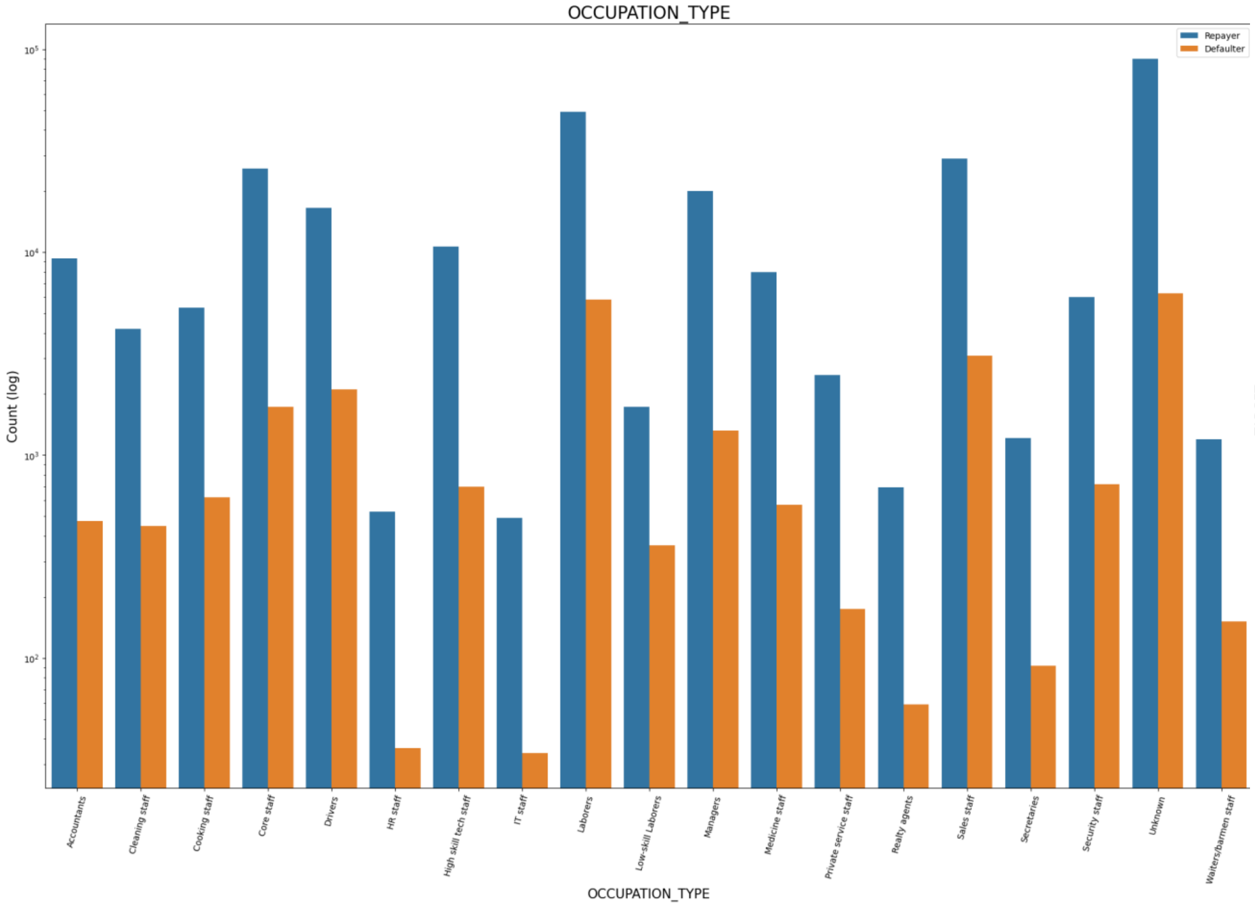
A significant portion of clients have completed either secondary or secondary special education, followed by clients with advanced degrees. Conversely, only a small fraction of clients hold a college degree. Notably, clients with lower secondary education background default at a rate of approximately 11%, while those with academic degrees have the lowest likelihood of defaulting.

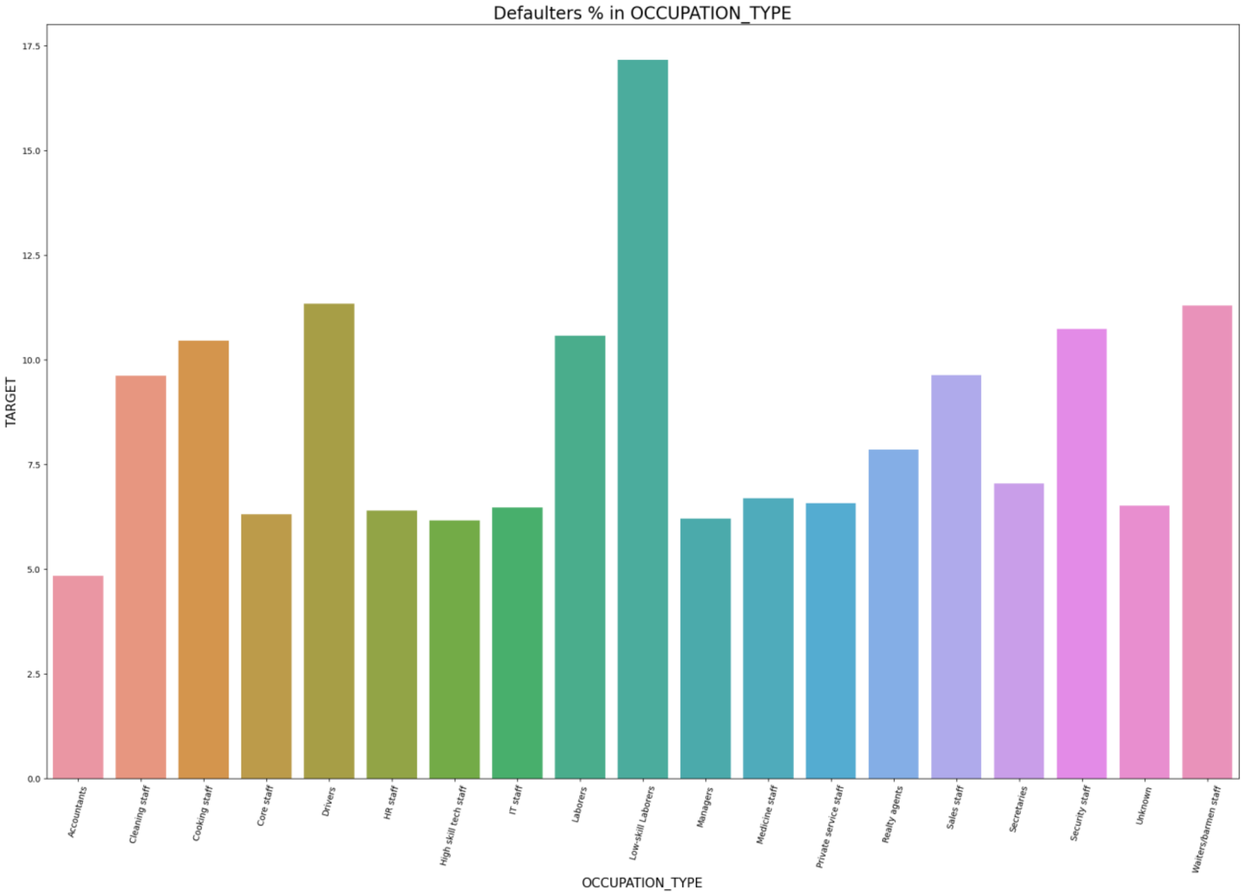


The largest group of loan applicants consists of individuals who are working, followed by those classified as commercial associates, pensioners, and state servants. Notably, the highest default rate, at 40%, is observed among candidates on maternity leave, closely followed by unemployed applicants at 37%. On the other hand, approximately 10% of loans in the remaining categories tend to default on average. It's noteworthy that students and businesspeople have no record of defaults, making them the safest categories for loan offers.



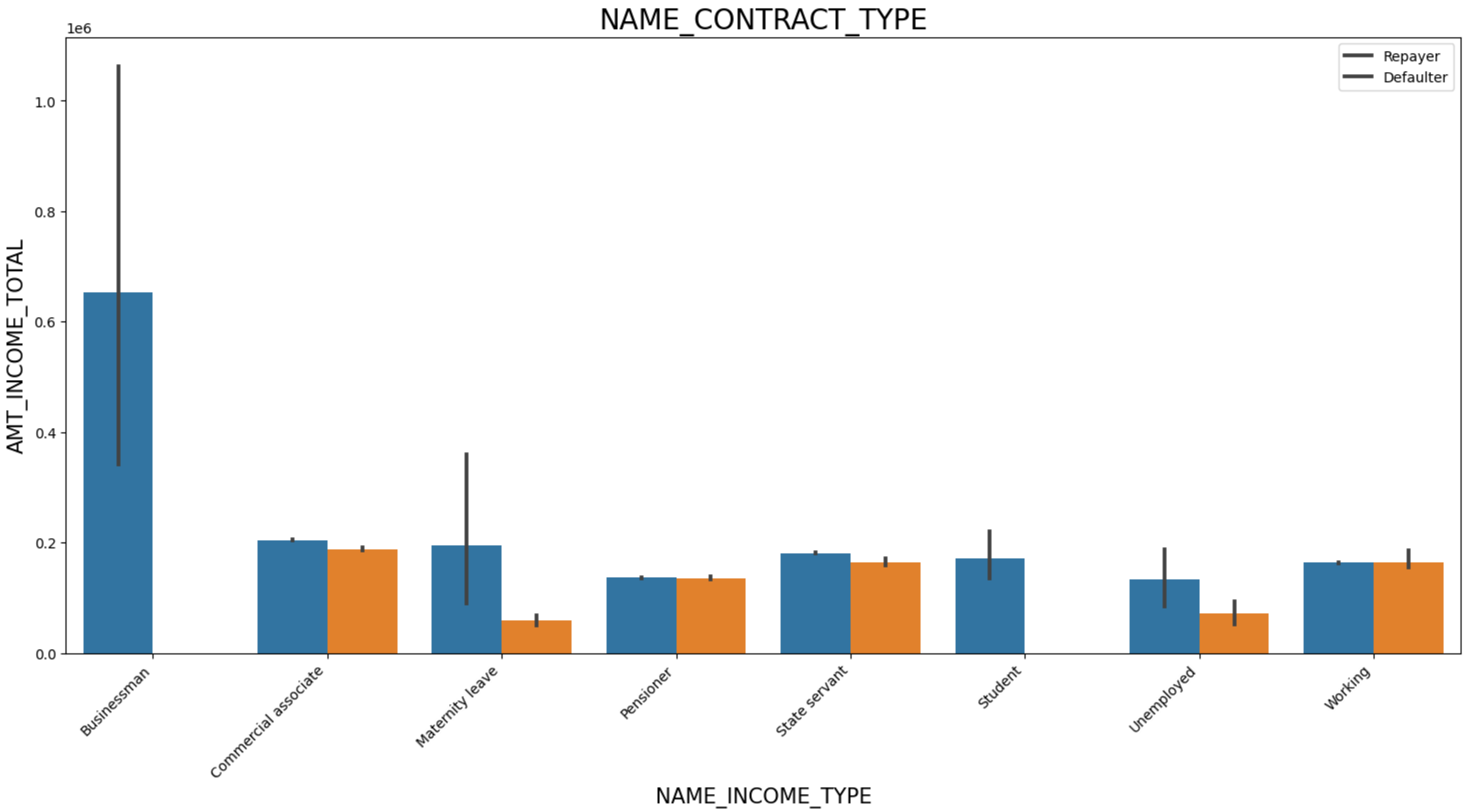






The majority of loans are acquired by laborers, with sales staff being the next most common group of borrowers. Interestingly, among the low-skill laborers, there is a relatively high incidence of fraud, exceeding 17%. Following this group, drivers, waiters/barmen staff, security staff, laborers, and cooking staff also have notable instances of fraud. Conversely, IT staff exhibit a lower likelihood of applying for loans.

# Categorical Bi/Multivariate analysis



We can observe that the income of a businessman is the highest, and the estimated range with a 95% confidence level indicates that a businessman's income may fall within the range of just under 4 lakhs to slightly over 10 lakhs.

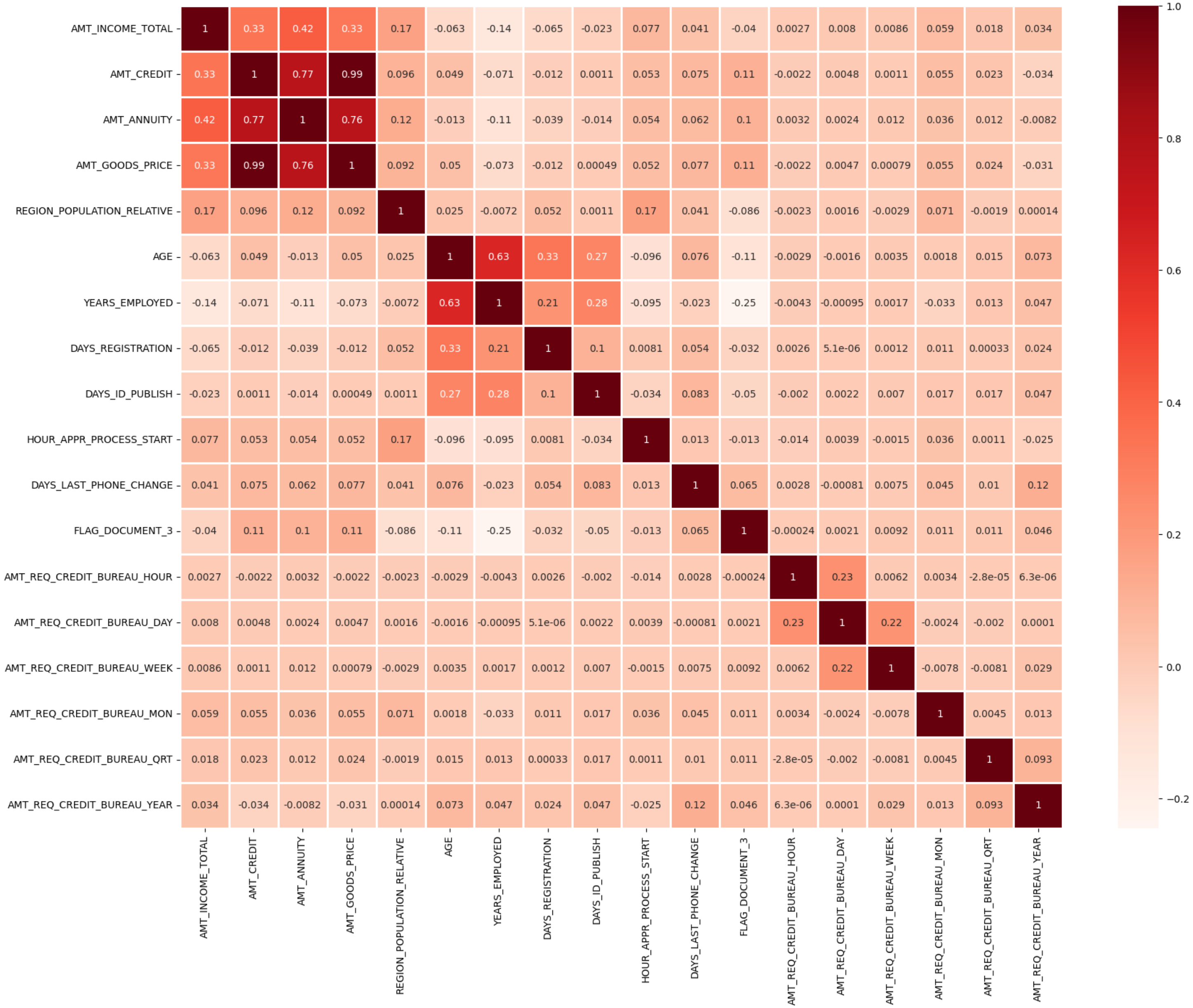
# Numeric Data Analysis

# Repayers (“TARGET” = 0)

**Factors associated with repayers:**

Credit amount is highly correlated with:

- Goods Price Amount
- Annuity Amount
- Total Income Amount



| | AMT_INCOME_TOTAL | AMT_CREDIT | AMT_ANNUITY | AMT_GOODS_PRICE | REGION_POPULATION_RELATIVE | AGE | YEARS_EMPLOYED | DAYS_REGISTRATION | DAYS_ID_PUBLISH | HOUR_APPR_PROCESS_START | DAYS_LAST_PHONE_CHANGE | FLAG_DOCUMENT_3 | AMT_REQ_CREDIT_BUREAU_HOUR | AMT_REQ_CREDIT_BUREAU_DAY | AMT_REQ_CREDIT_BUREAU_WEEK | AMT_REQ_CREDIT_BUREAU_MON | AMT_REQ_CREDIT_BUREAU_QRT | AMT_REQ_CREDIT_BUREAU_YEAR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AMT_INCOME_TOTAL | 1 | 0.33 | 0.42 | 0.33 | 0.17 | -0.063 | -0.14 | -0.065 | -0.023 | 0.077 | 0.041 | -0.04 | 0.0027 | 0.008 | 0.0086 | 0.059 | 0.018 | 0.034 |
| AMT_CREDIT | 0.33 | 1 | 0.77 | 0.99 | 0.096 | 0.049 | -0.071 | -0.012 | 0.0011 | 0.053 | 0.075 | 0.11 | -0.0022 | 0.0048 | 0.0011 | 0.055 | 0.023 | -0.034 |
| AMT_ANNUITY | 0.42 | 0.77 | 1 | 0.76 | 0.12 | -0.013 | -0.11 | -0.039 | -0.014 | 0.054 | 0.062 | 0.1 | 0.0032 | 0.0024 | 0.012 | 0.036 | 0.012 | -0.0082 |
| AMT_GOODS_PRICE | 0.33 | 0.99 | 0.76 | 1 | 0.092 | 0.05 | -0.073 | -0.012 | 0.00049 | 0.052 | 0.077 | 0.11 | -0.0022 | 0.0047 | 0.00079 | 0.055 | 0.024 | -0.031 |
| REGION_POPULATION_RELATIVE | 0.17 | 0.096 | 0.12 | 0.092 | 1 | 0.025 | -0.0072 | 0.052 | 0.0011 | 0.17 | 0.041 | -0.086 | -0.0023 | 0.0016 | -0.0029 | 0.071 | -0.0019 | 0.00014 |
| AGE | -0.063 | 0.049 | -0.013 | 0.05 | 0.025 | 1 | 0.63 | 0.33 | 0.27 | -0.096 | 0.076 | -0.11 | -0.0029 | -0.0016 | 0.0035 | 0.0018 | 0.015 | 0.073 |
| YEARS_EMPLOYED | -0.14 | -0.071 | -0.11 | -0.073 | -0.0072 | 0.63 | 1 | 0.21 | 0.28 | -0.095 | -0.023 | -0.25 | -0.0043 | -0.00095 | 0.0017 | -0.033 | 0.013 | 0.047 |
| DAYS_REGISTRATION | -0.065 | -0.012 | -0.039 | -0.012 | 0.052 | 0.33 | 0.21 | 1 | 0.1 | 0.0081 | 0.054 | -0.032 | 0.0026 | 5.1e-06 | 0.0012 | 0.011 | 0.00033 | 0.024 |
| DAYS_ID_PUBLISH | -0.023 | 0.0011 | -0.014 | 0.00049 | 0.0011 | 0.27 | 0.28 | 0.1 | 1 | -0.034 | 0.083 | -0.05 | -0.002 | 0.0022 | 0.007 | 0.017 | 0.017 | 0.047 |
| HOUR_APPR_PROCESS_START | 0.077 | 0.053 | 0.054 | 0.052 | 0.17 | -0.096 | -0.095 | 0.0081 | -0.034 | 1 | 0.013 | -0.013 | -0.014 | 0.0039 | -0.0015 | 0.036 | 0.0011 | -0.025 |
| DAYS_LAST_PHONE_CHANGE | 0.041 | 0.075 | 0.062 | 0.077 | 0.041 | 0.076 | -0.023 | 0.054 | 0.083 | 0.013 | 1 | 0.065 | 0.0028 | -0.00081 | 0.0075 | 0.045 | 0.01 | 0.12 |
| FLAG_DOCUMENT_3 | -0.04 | 0.11 | 0.1 | 0.11 | -0.086 | -0.11 | -0.25 | -0.032 | -0.05 | -0.013 | 0.065 | 1 | -0.00024 | 0.0021 | 0.0092 | 0.011 | 0.011 | 0.046 |
| AMT_REQ_CREDIT_BUREAU_HOUR | 0.0027 | -0.0022 | 0.0032 | -0.0022 | -0.0023 | -0.0029 | -0.0043 | 0.0026 | -0.002 | -0.014 | 0.0028 | -0.00024 | 1 | 0.23 | 0.0062 | 0.0034 | -2.8e-05 | 6.3e-06 |
| AMT_REQ_CREDIT_BUREAU_DAY | 0.008 | 0.0048 | 0.0024 | 0.0047 | 0.0016 | -0.0016 | -0.00095 | 5.1e-06 | 0.0022 | 0.0039 | -0.00081 | 0.0021 | 0.23 | 1 | 0.22 | -0.0024 | -0.002 | 0.0001 |
| AMT_REQ_CREDIT_BUREAU_WEEK | 0.0086 | 0.0011 | 0.012 | 0.00079 | -0.0029 | 0.0035 | 0.0017 | 0.0012 | 0.007 | -0.0015 | 0.0075 | 0.0092 | 0.0062 | 0.22 | 1 | -0.0078 | -0.0081 | 0.029 |
| AMT_REQ_CREDIT_BUREAU_MON | 0.059 | 0.055 | 0.036 | 0.055 | 0.071 | 0.0018 | -0.033 | 0.011 | 0.017 | 0.036 | 0.045 | 0.011 | 0.0034 | -0.0024 | -0.0078 | 1 | 0.0045 | 0.013 |
| AMT_REQ_CREDIT_BUREAU_QRT | 0.018 | 0.023 | 0.012 | 0.024 | -0.0019 | 0.015 | 0.013 | 0.00033 | 0.017 | 0.0011 | 0.01 | 0.011 | -2.8e-05 | -0.002 | -0.0081 | 0.0045 | 1 | 0.093 |
| AMT_REQ_CREDIT_BUREAU_YEAR | 0.034 | -0.034 | -0.0082 | -0.031 | 0.00014 | 0.073 | 0.047 | 0.024 | 0.047 | -0.025 | 0.12 | 0.046 | 6.3e-06 | 0.0001 | 0.029 | 0.013 | 0.093 | 1 |

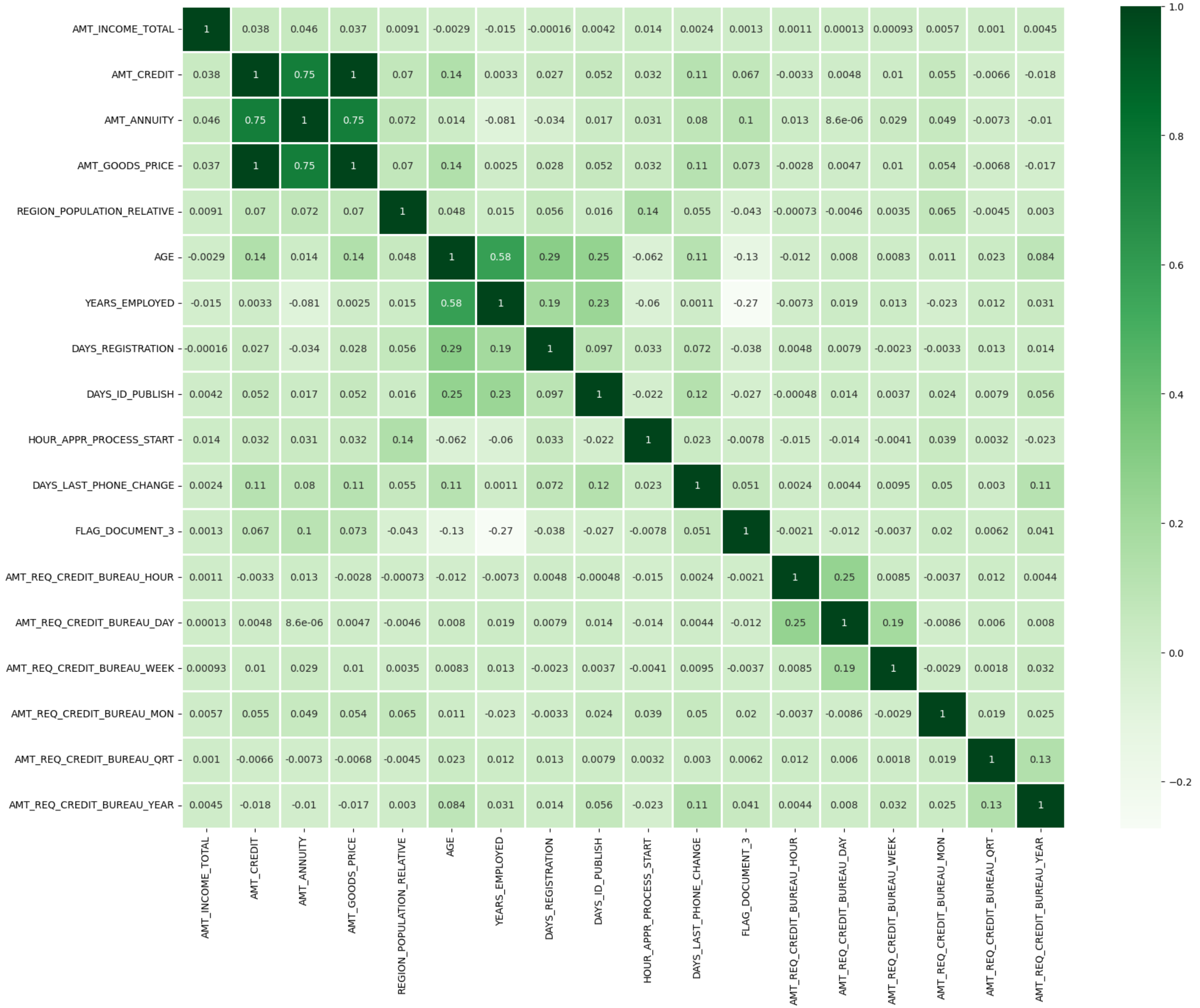
# Defaulters (“TARGET” = 1)

Factors impacting defaulters:

1.There is a strong correlation between the credit amount and the value of the goods purchased.

2.Among defaulters, the correlation between loan annuity and credit amount has slightly decreased to 0.75, compared to 0.77 among repayers.

3.The relationship between the client's total income and the credit amount is significantly weaker among defaulters (0.038) compared to repayers (0.342).



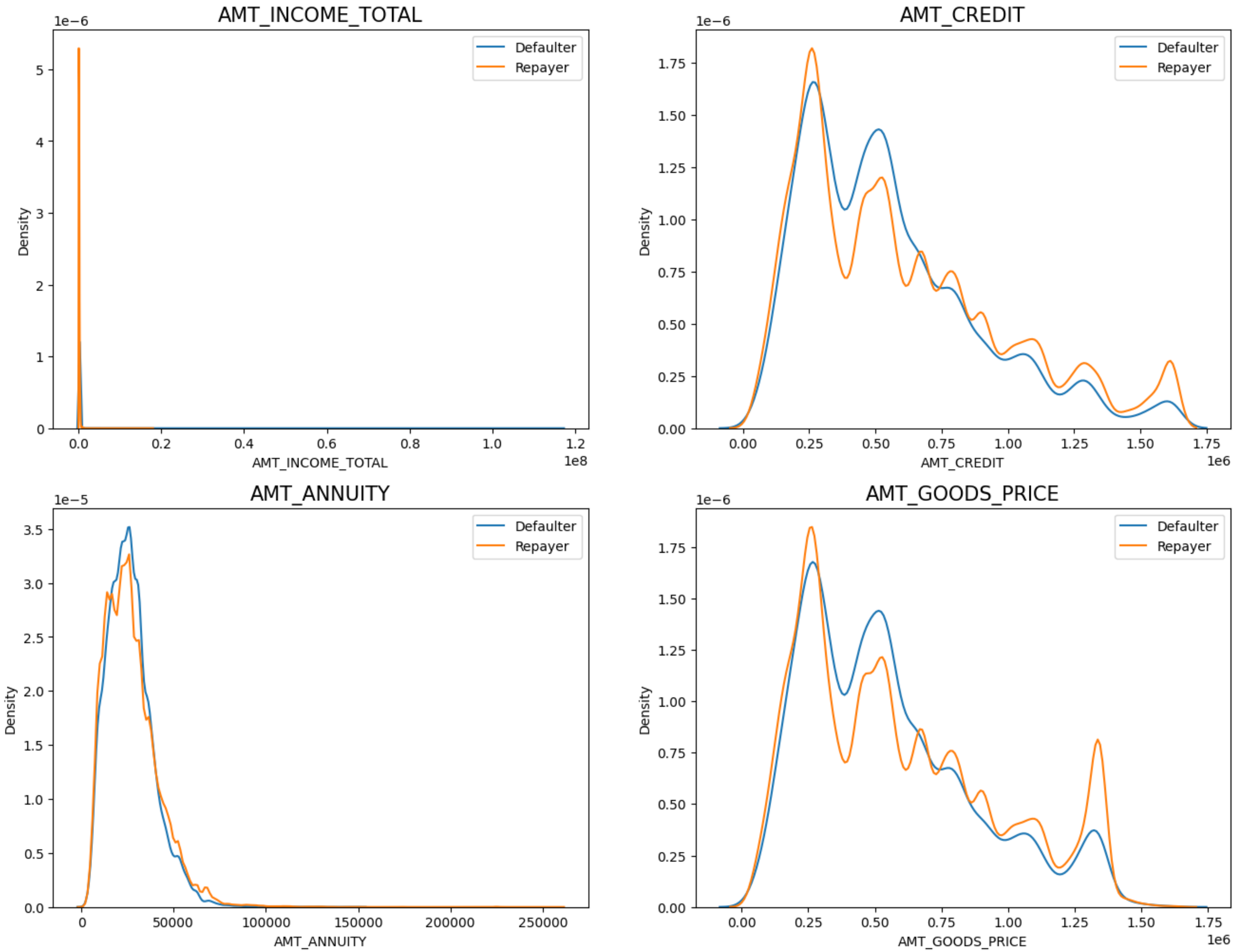
| | AMT_INCOME_TOTAL | AMT_CREDIT | AMT_ANNUITY | AMT_GOODS_PRICE | REGION_POPULATION_RELATIVE | AGE | YEARS_EMPLOYED | DAYS_REGISTRATION | DAYS_ID_PUBLISH | HOUR_APPR_PROCESS_START | DAYS_LAST_PHONE_CHANGE | FLAG_DOCUMENT_3 | AMT_REQ_CREDIT_BUREAU_HOUR | AMT_REQ_CREDIT_BUREAU_DAY | AMT_REQ_CREDIT_BUREAU_WEEK | AMT_REQ_CREDIT_BUREAU_MON | AMT_REQ_CREDIT_BUREAU_QRT | AMT_REQ_CREDIT_BUREAU_YEAR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AMT_INCOME_TOTAL | 1 | 0.038 | 0.046 | 0.037 | 0.0091 | -0.0029 | -0.015 | -0.00016 | 0.0042 | 0.014 | 0.0024 | 0.0013 | 0.0011 | 0.00013 | 0.00093 | 0.0057 | 0.001 | 0.0045 |
| AMT_CREDIT | 0.038 | 1 | 0.75 | 1 | 0.07 | 0.14 | 0.0033 | 0.027 | 0.052 | 0.032 | 0.11 | 0.067 | -0.0033 | 0.0048 | 0.01 | 0.055 | -0.0066 | -0.018 |
| AMT_ANNUITY | 0.046 | 0.75 | 1 | 0.75 | 0.072 | 0.014 | -0.081 | -0.034 | 0.017 | 0.031 | 0.08 | 0.1 | 0.013 | 8.6e-06 | 0.029 | 0.049 | -0.0073 | -0.01 |
| AMT_GOODS_PRICE | 0.037 | 1 | 0.75 | 1 | 0.07 | 0.14 | 0.0025 | 0.028 | 0.052 | 0.032 | 0.11 | 0.073 | -0.0028 | 0.0047 | 0.01 | 0.054 | -0.0068 | -0.017 |
| REGION_POPULATION_RELATIVE | 0.0091 | 0.07 | 0.072 | 0.07 | 1 | 0.048 | 0.015 | 0.056 | 0.016 | 0.14 | 0.055 | -0.043 | -0.00073 | -0.0046 | 0.0035 | 0.065 | -0.0045 | 0.003 |
| AGE | -0.0029 | 0.14 | 0.014 | 0.14 | 0.048 | 1 | 0.58 | 0.29 | 0.25 | -0.062 | 0.11 | -0.13 | -0.012 | 0.008 | 0.0083 | 0.011 | 0.023 | 0.084 |
| YEARS_EMPLOYED | -0.015 | 0.0033 | -0.081 | 0.0025 | 0.015 | 0.58 | 1 | 0.19 | 0.23 | -0.06 | 0.0011 | -0.27 | -0.0073 | 0.019 | 0.013 | -0.023 | 0.012 | 0.031 |
| DAYS_REGISTRATION | -0.00016 | 0.027 | -0.034 | 0.028 | 0.056 | 0.29 | 0.19 | 1 | 0.097 | 0.033 | 0.072 | -0.038 | 0.0048 | 0.0079 | -0.0023 | -0.0033 | 0.013 | 0.014 |
| DAYS_ID_PUBLISH | 0.0042 | 0.052 | 0.017 | 0.052 | 0.016 | 0.25 | 0.23 | 0.097 | 1 | -0.022 | 0.12 | -0.027 | -0.00048 | 0.014 | 0.0037 | 0.024 | 0.0079 | 0.056 |
| HOUR_APPR_PROCESS_START | 0.014 | 0.032 | 0.031 | 0.032 | 0.14 | -0.062 | -0.06 | 0.033 | -0.022 | 1 | 0.023 | -0.0078 | -0.015 | -0.014 | -0.0041 | 0.039 | 0.0032 | -0.023 |
| DAYS_LAST_PHONE_CHANGE | 0.0024 | 0.11 | 0.08 | 0.11 | 0.055 | 0.11 | 0.0011 | 0.072 | 0.12 | 0.023 | 1 | 0.051 | 0.0024 | 0.0044 | 0.0095 | 0.05 | 0.003 | 0.11 |
| FLAG_DOCUMENT_3 | 0.0013 | 0.067 | 0.1 | 0.073 | -0.043 | -0.13 | -0.27 | -0.038 | -0.027 | -0.0078 | 0.051 | 1 | -0.0021 | -0.012 | -0.0037 | 0.02 | 0.0062 | 0.041 |
| AMT_REQ_CREDIT_BUREAU_HOUR | 0.0011 | -0.0033 | 0.013 | -0.0028 | -0.00073 | -0.012 | -0.0073 | 0.0048 | -0.00048 | -0.015 | 0.0024 | -0.0021 | 1 | 0.25 | 0.0085 | -0.0037 | 0.012 | 0.0044 |
| AMT_REQ_CREDIT_BUREAU_DAY | 0.00013 | 0.0048 | 8.6e-06 | 0.0047 | -0.0046 | 0.008 | 0.019 | 0.0079 | 0.014 | -0.014 | 0.0044 | -0.012 | 0.25 | 1 | 0.19 | -0.0086 | 0.006 | 0.008 |
| AMT_REQ_CREDIT_BUREAU_WEEK | 0.00093 | 0.01 | 0.029 | 0.01 | 0.0035 | 0.0083 | 0.013 | -0.0023 | 0.0037 | -0.0041 | 0.0095 | -0.0037 | 0.0085 | 0.19 | 1 | -0.0029 | 0.0018 | 0.032 |
| AMT_REQ_CREDIT_BUREAU_MON | 0.0057 | 0.055 | 0.049 | 0.054 | 0.065 | 0.011 | -0.023 | -0.0033 | 0.024 | 0.039 | 0.05 | 0.02 | -0.0037 | -0.0086 | -0.0029 | 1 | 0.019 | 0.025 |
| AMT_REQ_CREDIT_BUREAU_QRT | 0.001 | -0.0066 | -0.0073 | -0.0068 | -0.0045 | 0.023 | 0.012 | 0.013 | 0.0079 | 0.0032 | 0.003 | 0.0062 | 0.012 | 0.006 | 0.0018 | 0.019 | 1 | 0.13 |
| AMT_REQ_CREDIT_BUREAU_YEAR | 0.0045 | -0.018 | -0.01 | -0.017 | 0.003 | 0.084 | 0.031 | 0.014 | 0.056 | -0.023 | 0.11 | 0.041 | 0.0044 | 0.008 | 0.032 | 0.025 | 0.13 | 1 |

# Numeric Univariate Analysis

The majority of loans are for purchases with prices below 10 lakhs.

A significant number of individuals make annuity payments for credit loans below 50,000.

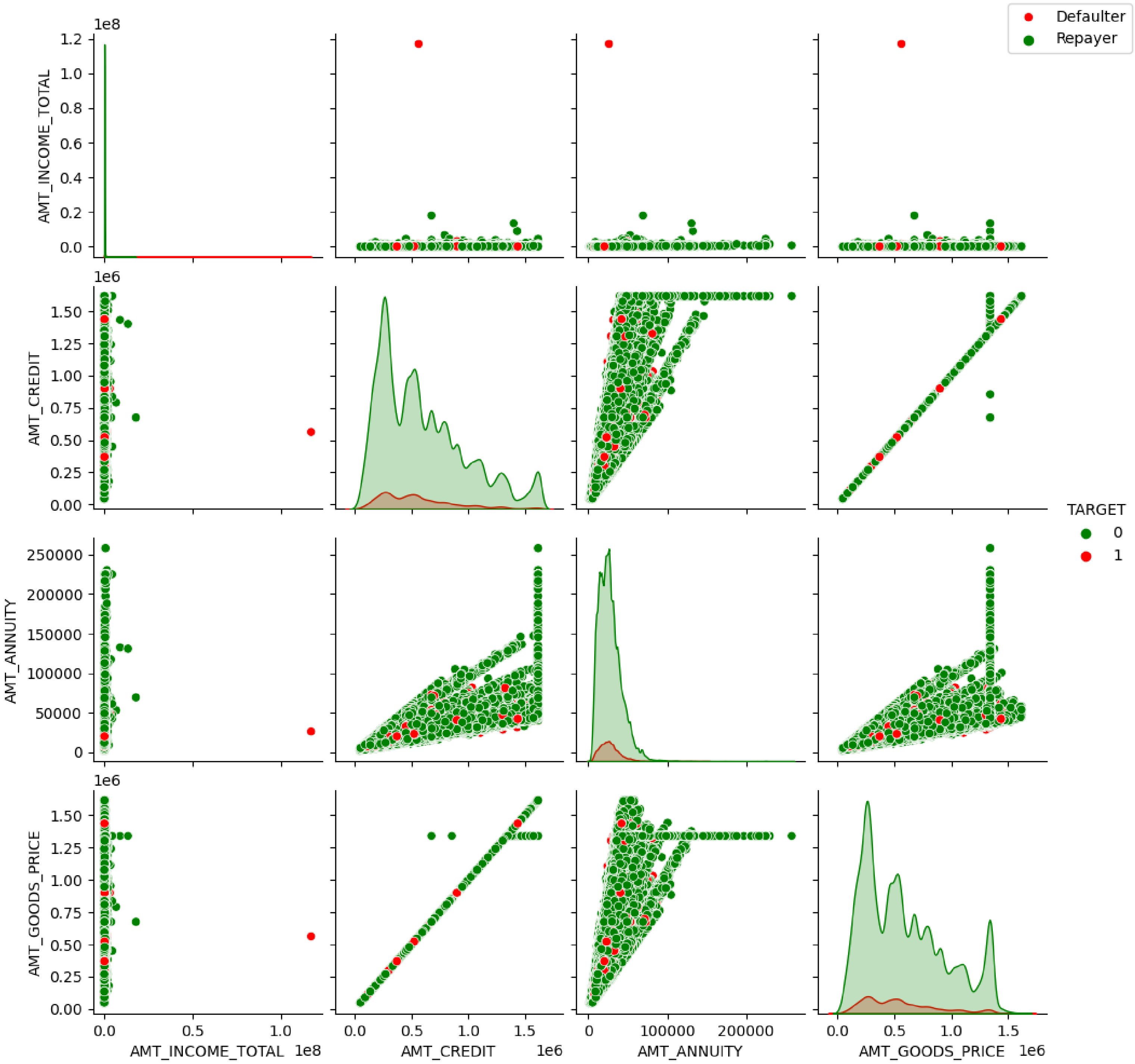
It's important to note that these factors alone cannot be relied upon for decision-making, as the distributions of repayers and defaulters overlap in all the graphs.



# Numeric Bivariate Analysis

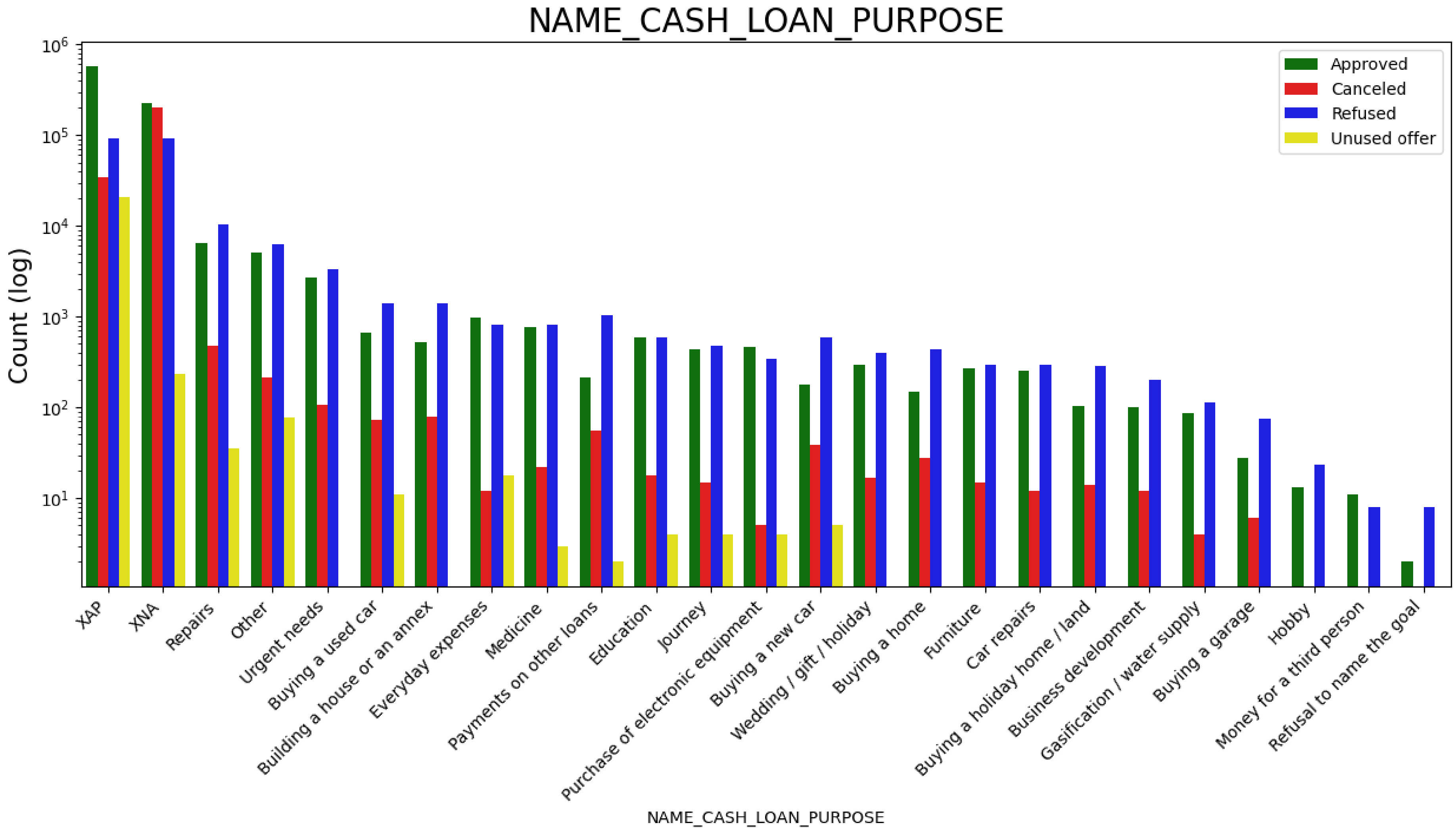
As depicted in the scatterplot, where the data predominantly aligns along a linear pattern, it's evident that there exists a substantial correlation between Loan Amount (AMT_CREDIT) and Goods Price (AMT_GOODS_PRICE).

Additionally, it's noteworthy that for higher AMT_CREDIT values, the number of defaulters is exceptionally low.

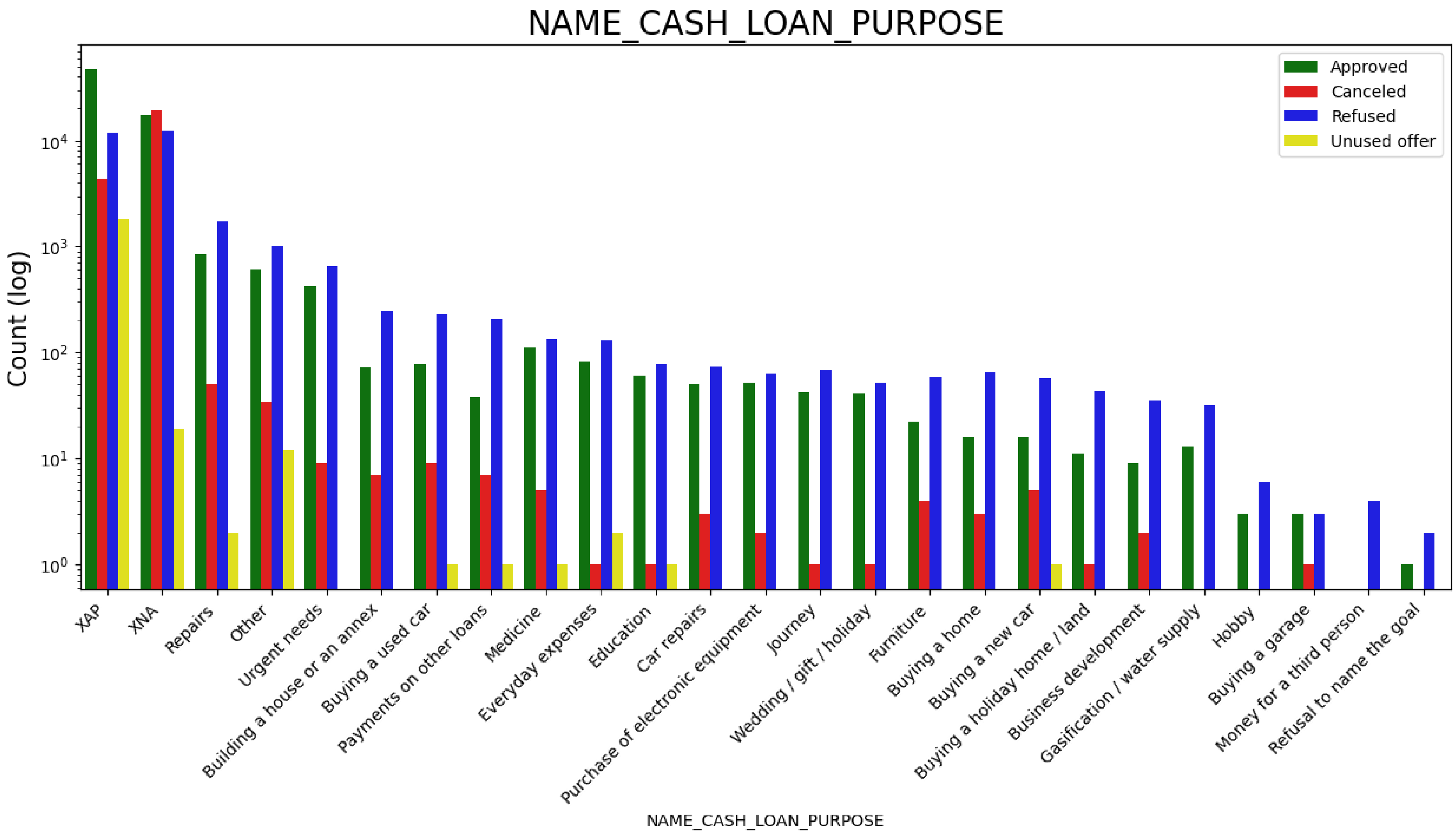


# Merged DataFrame Analysis

# Repayers (“TARGET” = 0)



# Defaulters (“TARGET” = 1)



A significant number of loans have unspecified purposes denoted as "XAP" or "XNA.“

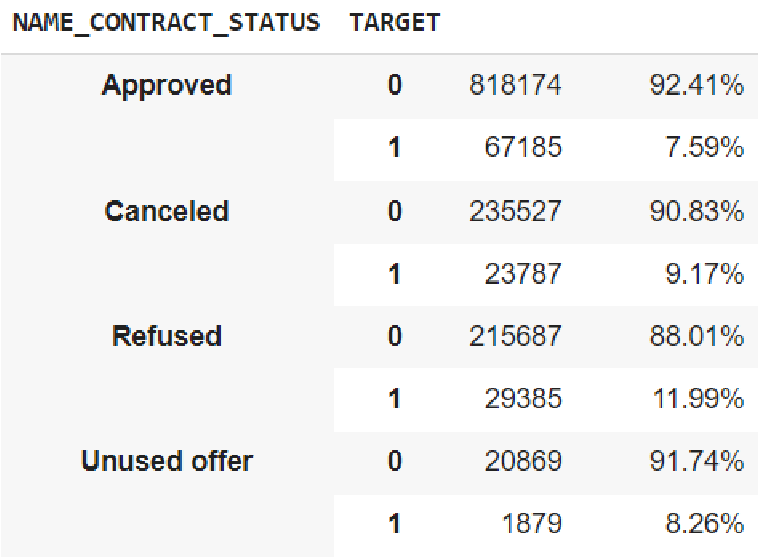
The highest default rate is observed in loans taken for repair purposes.

It's notable that many loan applications, particularly those designated for repair or other unspecified purposes, have faced rejection by both banks and clients. This might imply that the bank perceives repair-related loans as high-risk. Furthermore, these applicants are either denied loans altogether or are offered loans at such high-interest rates that they become unaffordable, ultimately leading to loan rejections.

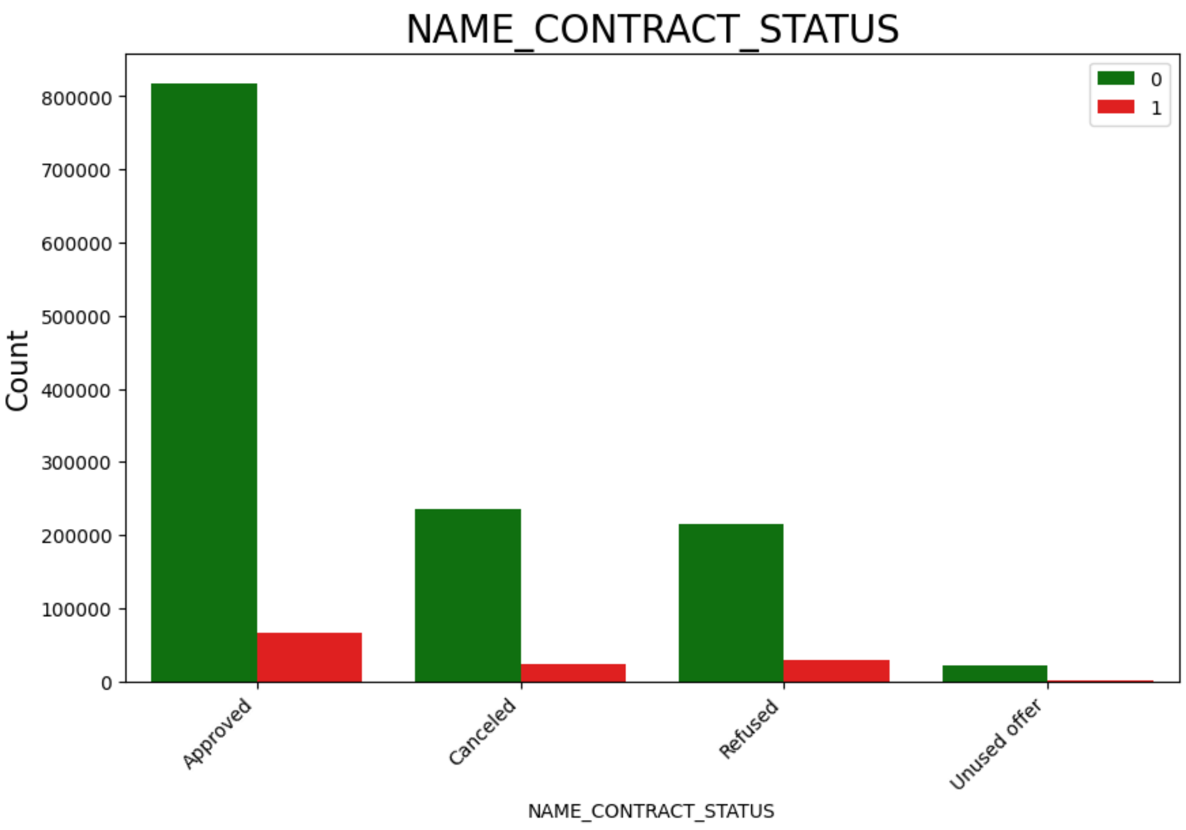
Among clients who were previously rejected, an impressive 90% have successfully repaid their loans. This suggests that adjusting lending rates could potentially enhance business prospects.

In more recent instances, 88% of clients who had previously faced loan denials have exhibited a positive repayment behavior.

It is advisable to take note of the reasons for these rejections for future investigation, as these clients may present valuable opportunities for becoming future customers.



| NAME_CONTRACT_STATUS | TARGET | | |
|---|---|---|---|
| **Approved** | **0** | 818174 | 92.41% |
| | **1** | 67185 | 7.59% |
| **Canceled** | **0** | 235527 | 90.83% |
| | **1** | 23787 | 9.17% |
| **Refused** | **0** | 215687 | 88.01% |
| | **1** | 29385 | 11.99% |
| **Unused offer** | **0** | 20869 | 91.74% |
| | **1** | 1879 | 8.26% |



# CONCLUSION

# SUMMARY

In the category of contract types, it is evident that a higher proportion of female individuals with consumer loans tend to default on their payments.

Married individuals are encountering challenges in meeting their loan obligations in comparison to their single or separated counterparts. Furthermore, a larger number of approved applicants are married. Therefore, it is advisable to prioritize granting loans to single or divorced applicants in the future.

Individuals with a background in secondary or secondary special education are experiencing difficulties in repaying their loans and exhibit a higher likelihood of becoming defaulters.

Both individuals with lower credit amounts and those with significantly higher credit amounts face an elevated risk of defaulting on their loans.

In general, females demonstrate a lower likelihood of experiencing payment difficulties compared to males. Consequently, it is recommended to approve more loans for females, taking into account their higher prevalence among those facing payment challenges.

Applicants who have previously applied for loans (repeated applicants) have a higher probability of both non-defaulting and defaulting when compared to new applicants.

# THANK YOU