



Loan Defaulter EDA

CASE STUDY



Problem Statement

- When the company receives a loan application, the company has to decide for its approval based on the applicant's profile. Two types of risks are associated with bank's decision :
 1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company .
 2. If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.
- This case study aims to identify patterns which indicate if a client has difficulty paying their installments which maybe used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim for this case study.

Approach |

Case study for this huge dataset is to be done by following approach :

Data Cleaning

- Getting an idea of the shape of the Dataset
- Checking for columns with null values more than 45% and dropping those columns
- Checking for columns with null values lower than 15% and imputing them with Mean, Median or Mode accordingly.
- Checking for a few more unwanted columns and dropping them.
- Checking the Datatypes of the columns and making adjustment to it accordingly
- Adjusting the values of few columns to +ve and also naming them accordingly.

Analysis



1.

Checking for outliers with the help of Boxplot .

3.

Dividing the dataset on basis of Target Variable and plotting graphs to get better insights .

5.

Doing Univariate Analysis for numerical and categorical variable in the dataset for both target variables using different plots.

2.

Binning for continuous columns to make them categorical columns which is easier to visualize .

4.

Finding correlations with the help of Heatmap and recording the top 10 correlations for both target variable datasets.

6.

Doing Bivariate Analysis for all Numerical to Numerical, Categorical to Categorical and Numerical to Categorical using different plots.

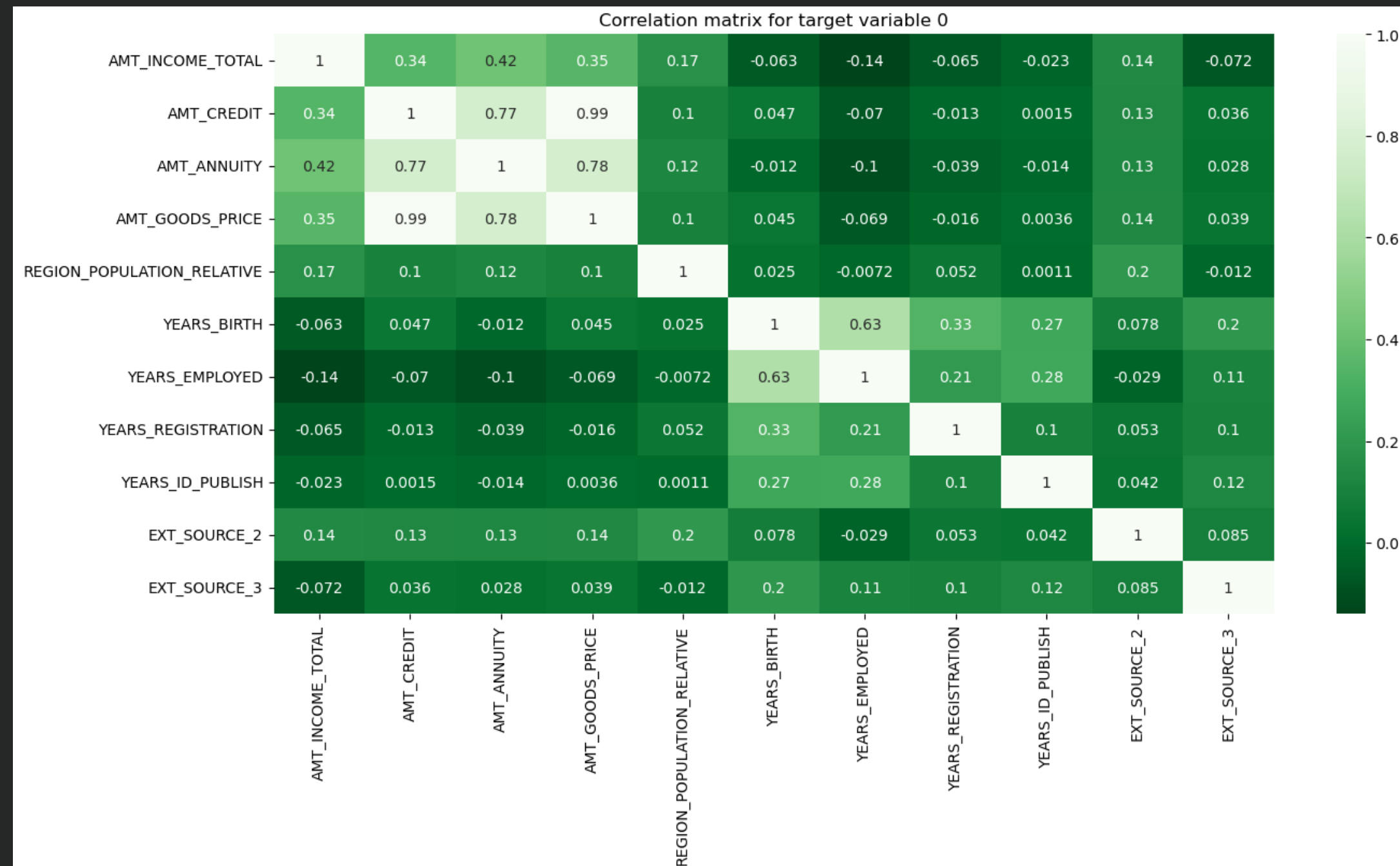


Merging the Datasets

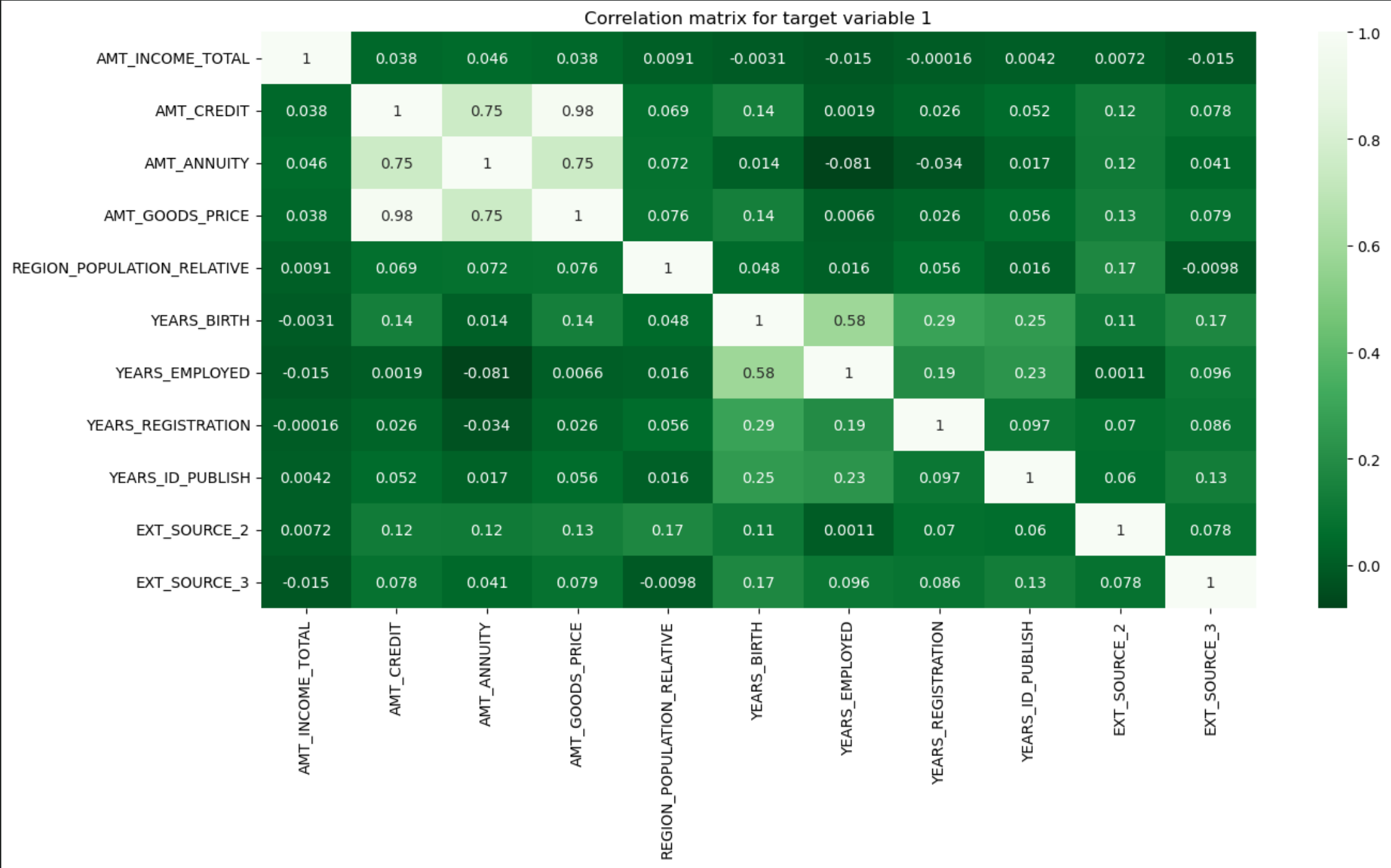
Merged the dataset with the previous application dataset and made some more analysis with the help of subplot to find the different relations between Approved, Refused, Cancelled and Unused contract status. This helped gain further insights on which sectors to target without facing more payment difficulties by the customers.

Visualization

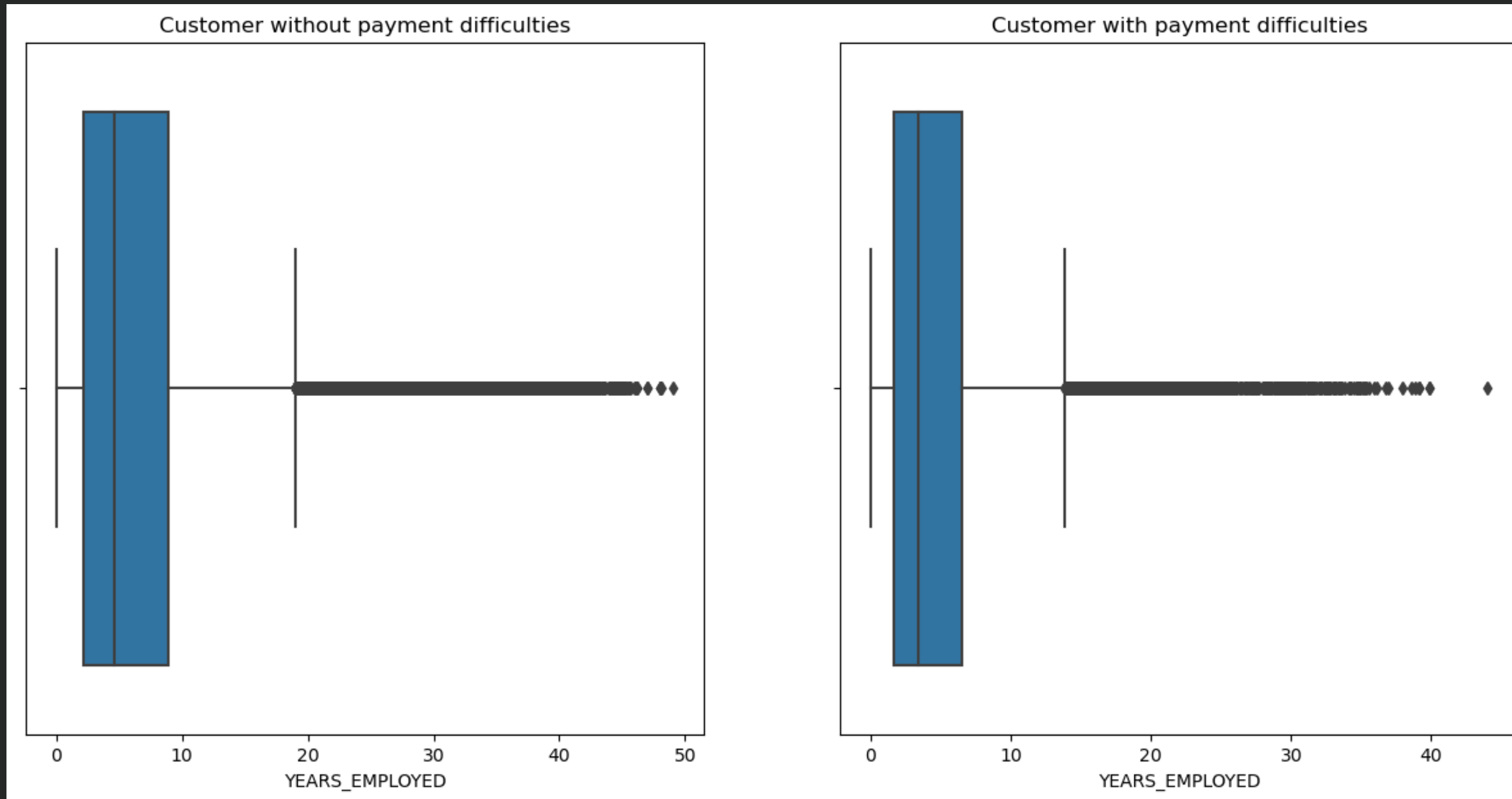
Correlation Matrix for target variable 0 :



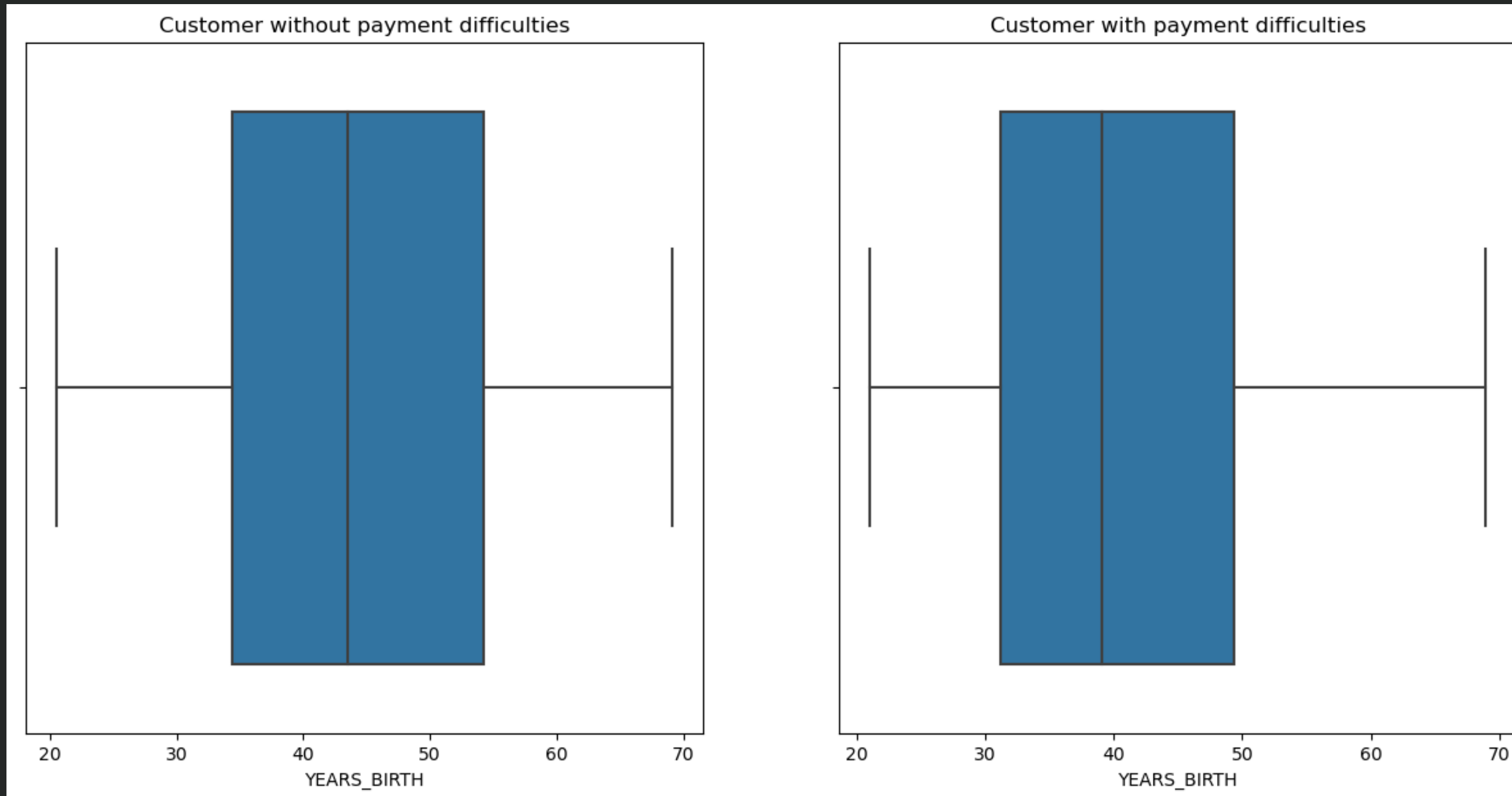
Correlation Matrix for target variable 1:



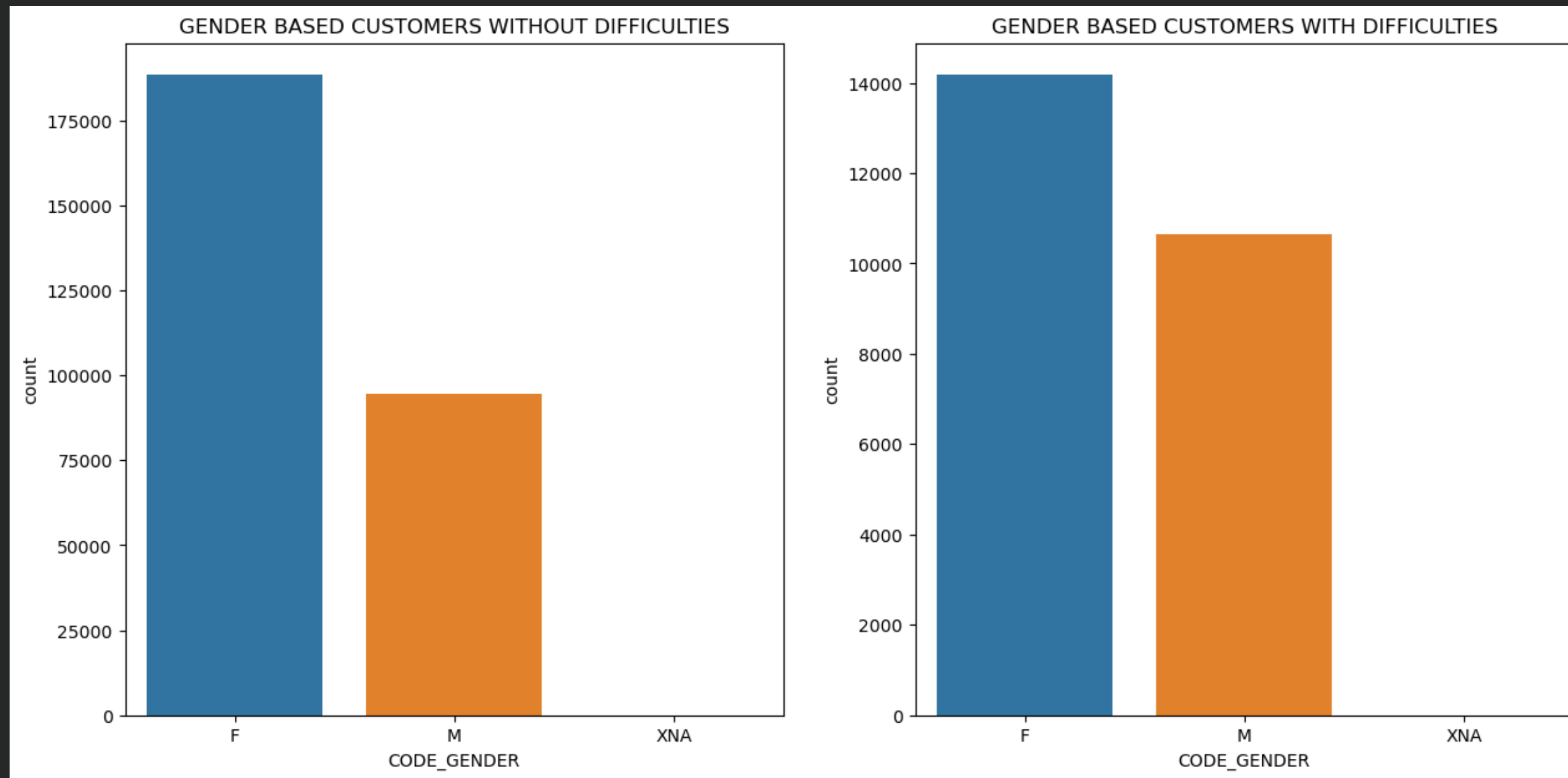
Univariate Analysis on YEARS_EMPLOYED:



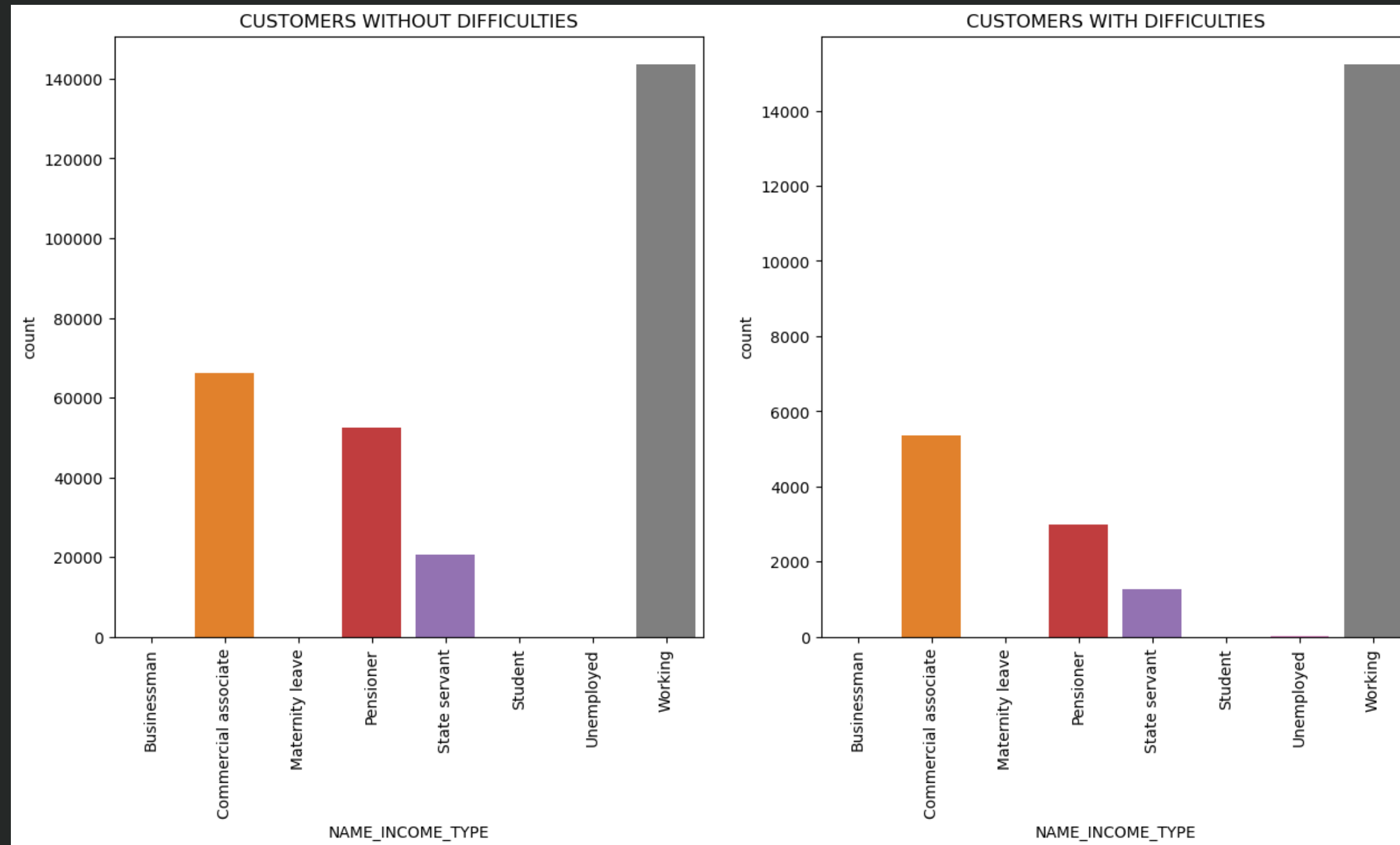
Univariate Analysis on YEARS_BIRTH:



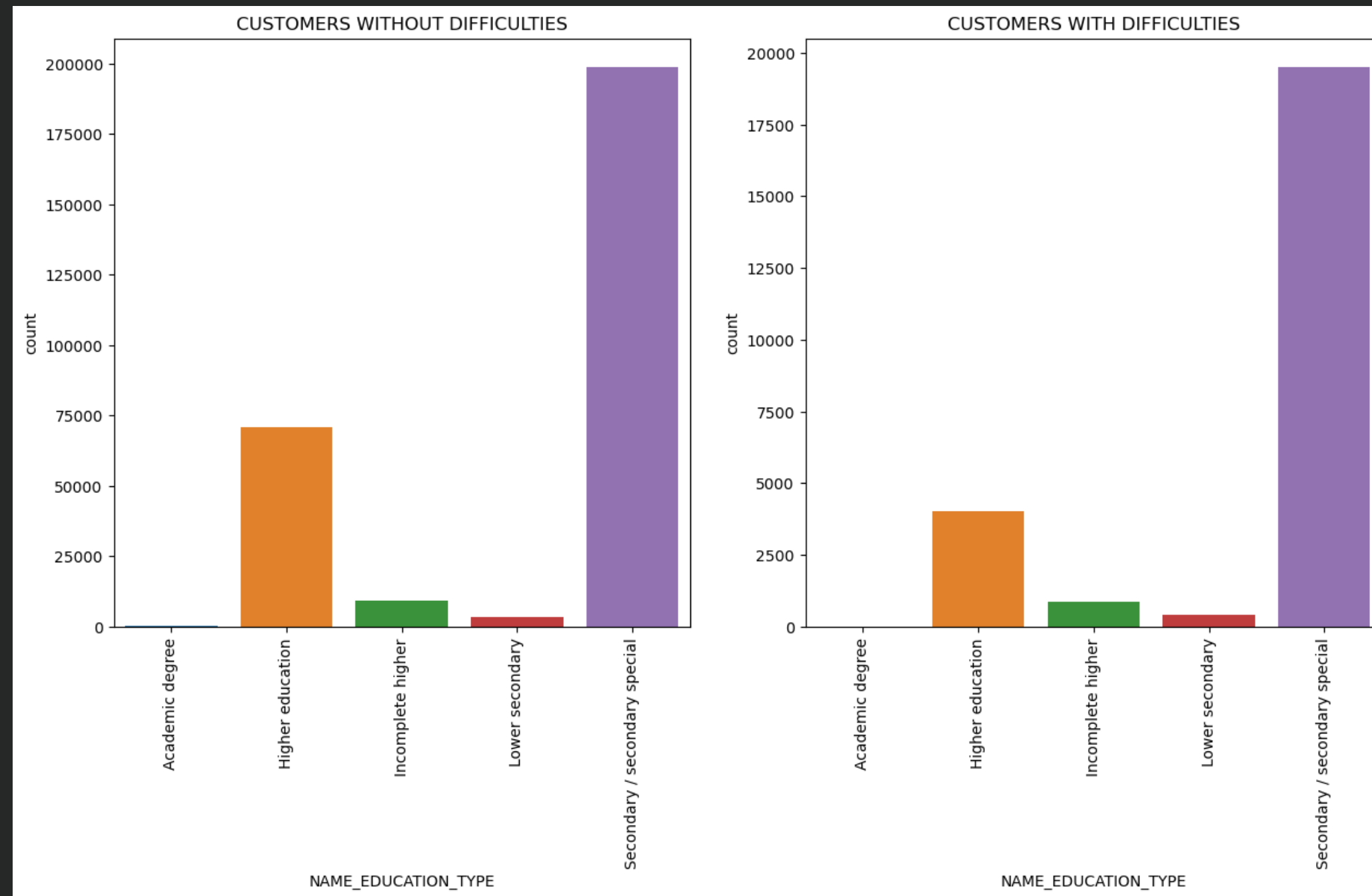
Univariate Analysis on CODE_GENDER:



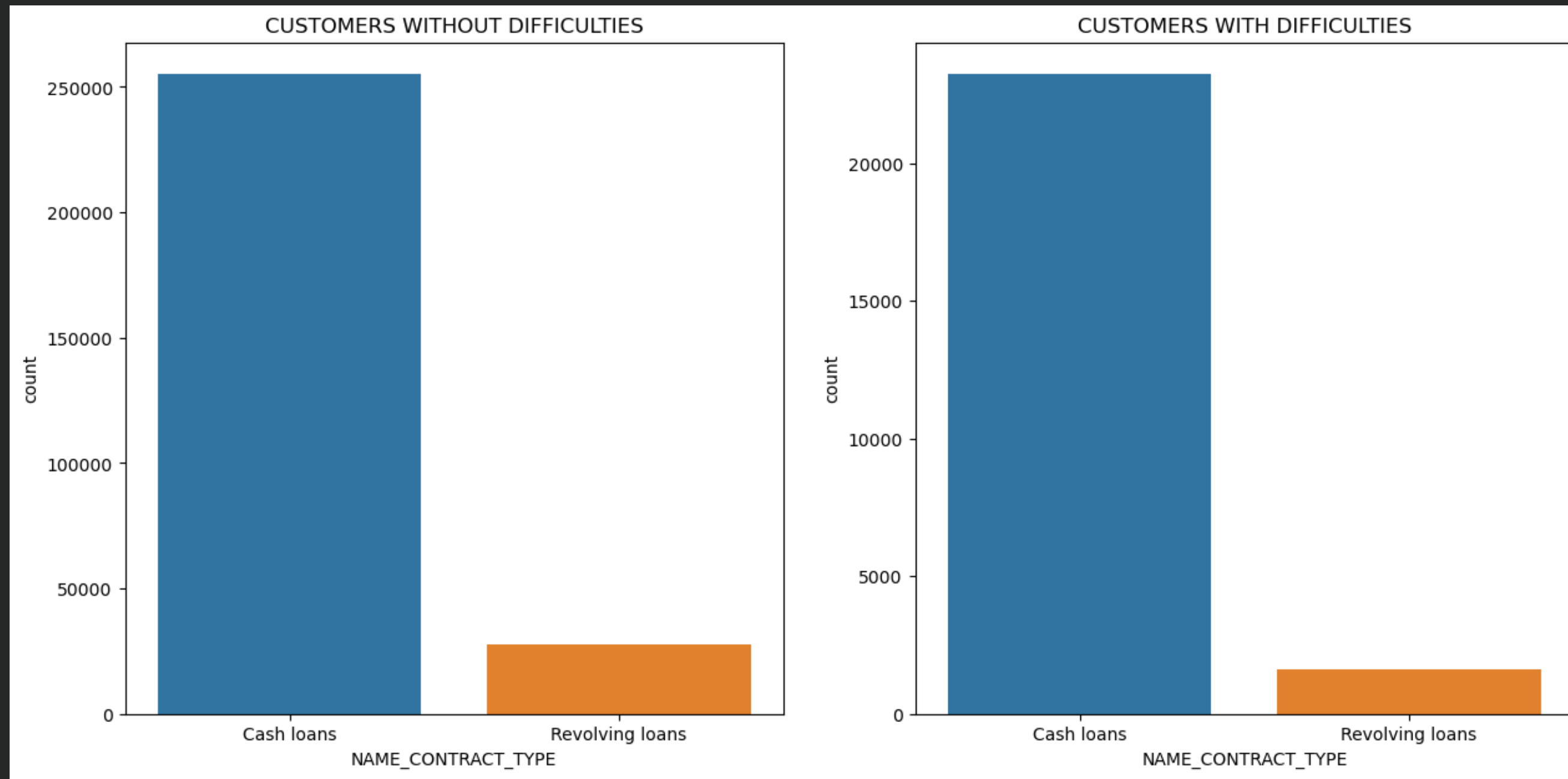
Univariate Analysis on NAME_INCOME_TYPE:



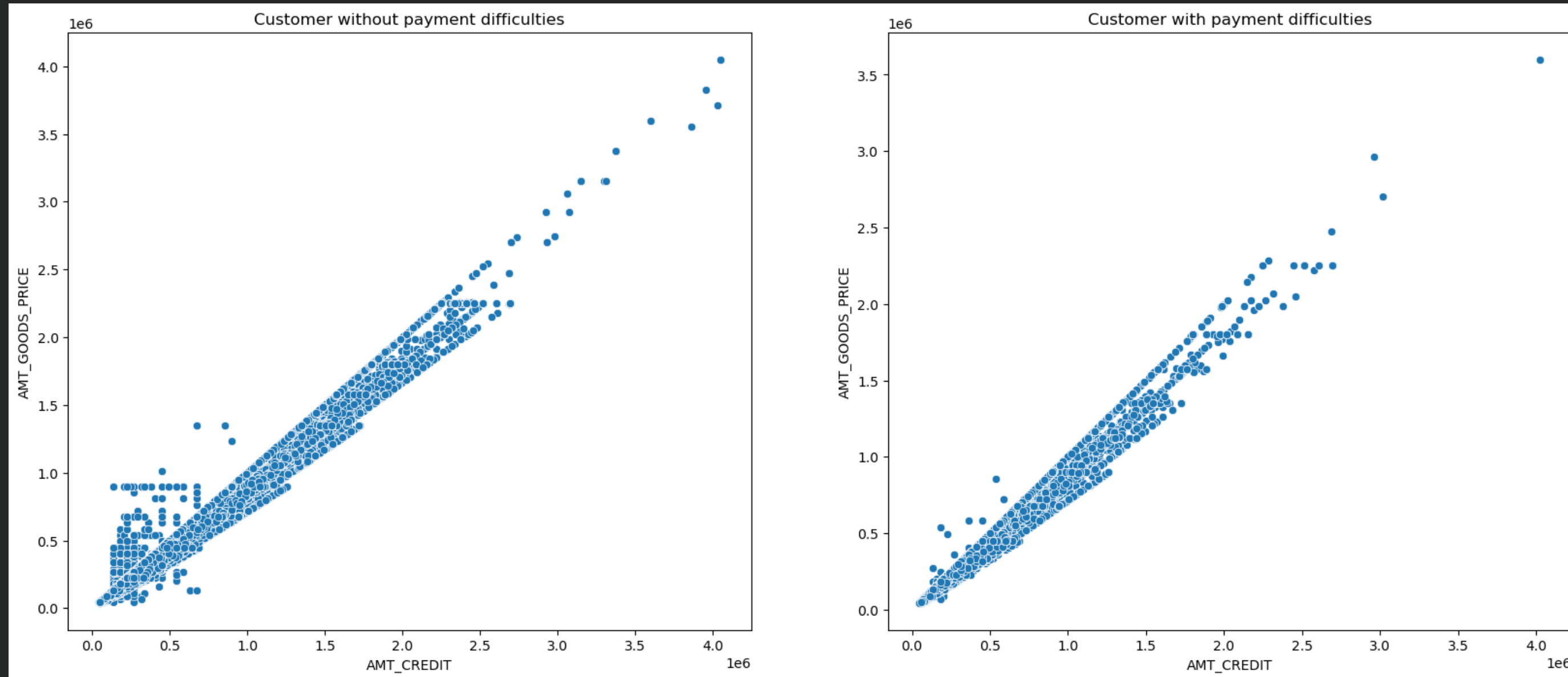
Univariate Analysis on NAME_EDUCATION_TYPE:



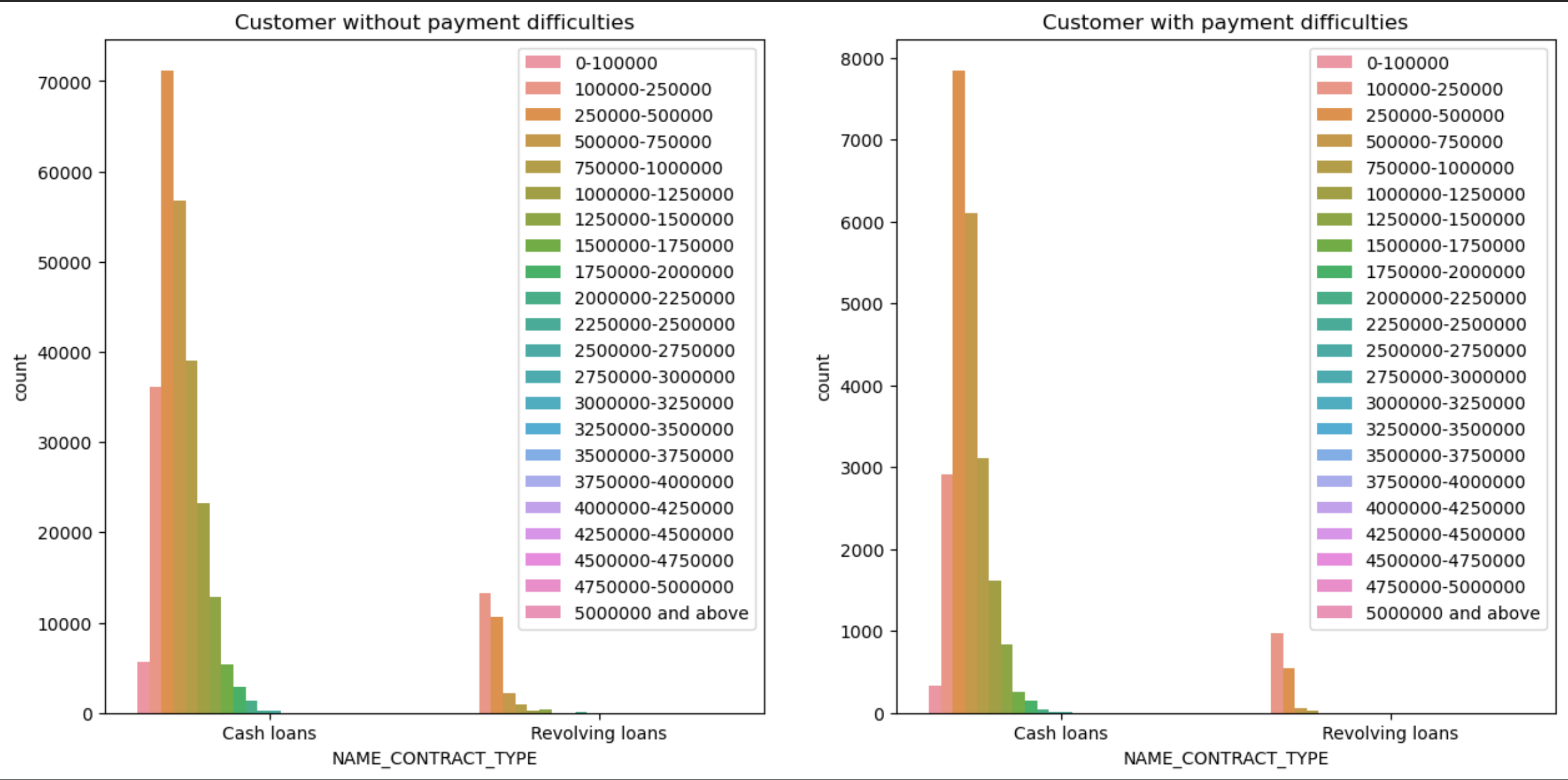
Univariate Analysis on n NAME_CONTRACT_TYPE:



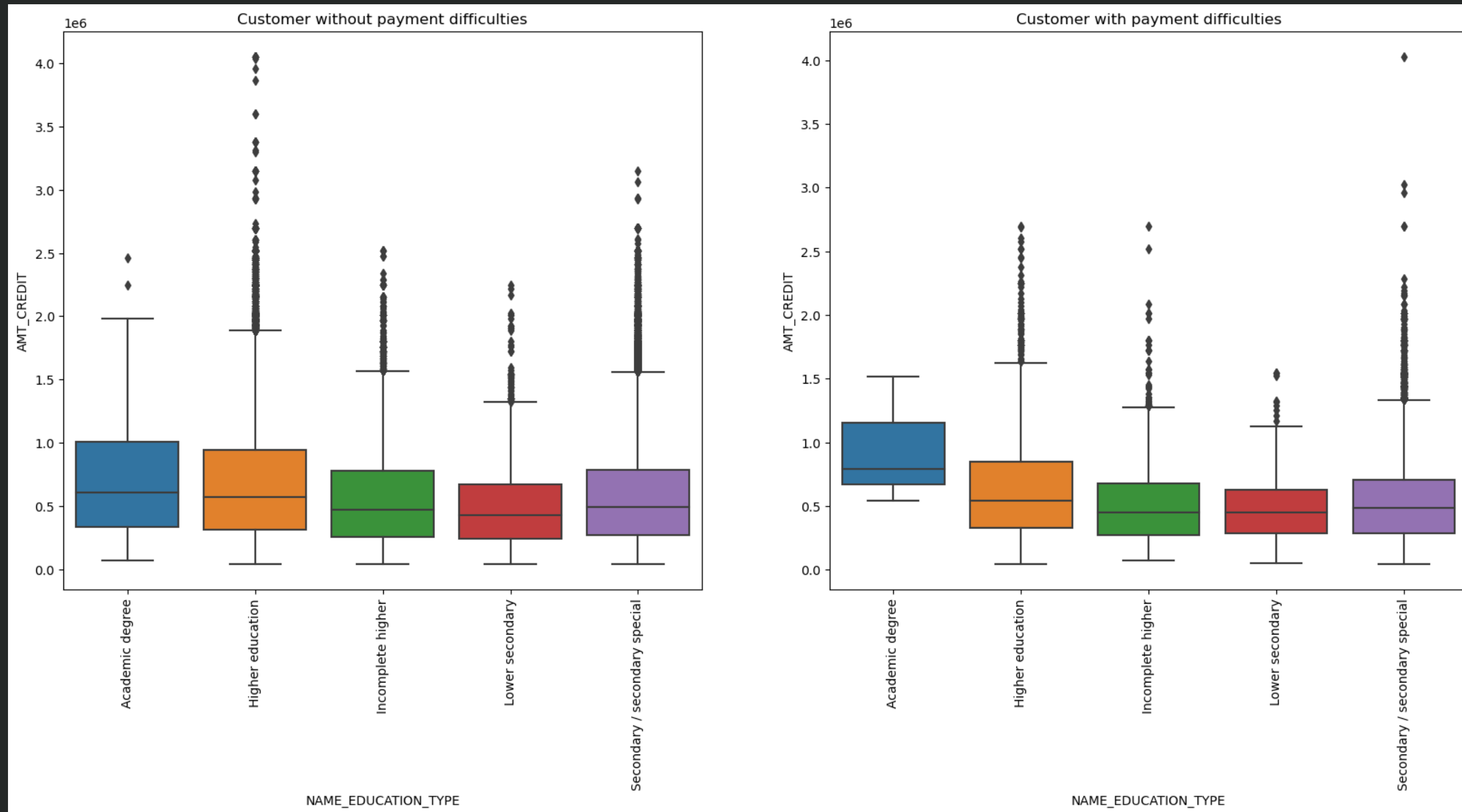
Bivariate Analysis for AMT_CREDIT to AMT_GOODS_PRICE:



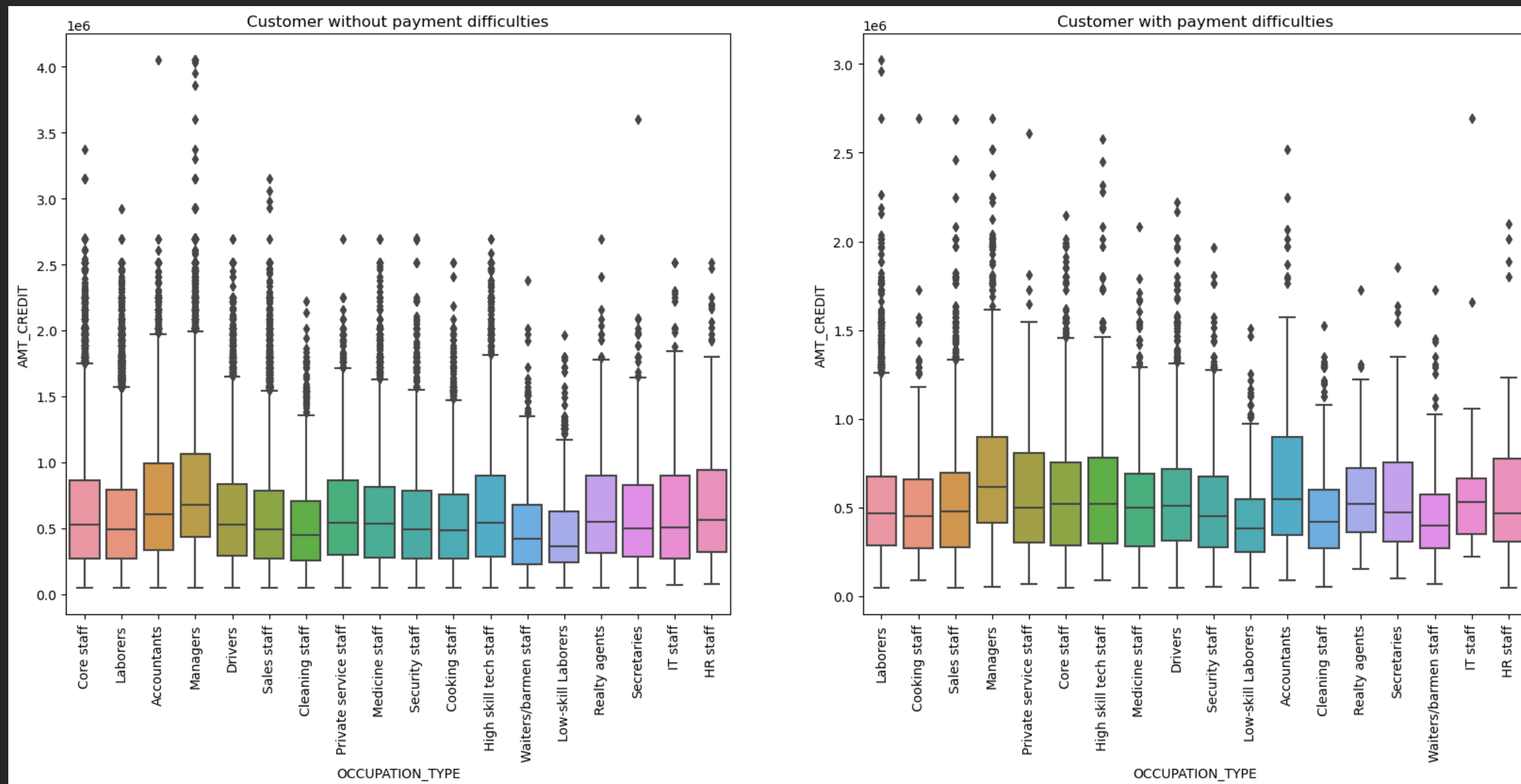
Bivariate analysis for NAME_CONTRACT_TYPE to CREDIT_RANGE:



Bivariate analysis for AMT_CREDIT to NAME_EDUCATION_TYPE:



Bivariate analysis for AMT_CREDIT to OCCUPATION_TYPE:



Observations

- The Working Class customers face more difficulties in payments compared to the other classes.
- Females face more payment difficulties and also apply for more credit and are approved as well.
- Customers with Repairs as reason for Credit tend to face more difficulties with Credit Repayment.
- Married Customers apply for most Credit amongst the other classes and also face difficulties in Payment.

Recommendations

- Banks should focus on Single persons as they have the least payment difficulty and also low refused rate. This may help with successful payments.
- Banks should focus less on 'Married People' as they have the most difficulties with payment and also have the highest approval. They may lead to higher payment difficulties.
- Banks should not focus on Education status 'Academic Degree' as they have the most difficulties with payment. Instead they can focus on 'Lower Secondary' as they have the least difficulties with payment.
- Banks can focus on 'Realty Agents' as they have somewhat lower difficulties with payments than 'Managers' and 'Accountants' but a lower approval rate. So, it might be beneficial to focus on that Occupation type.
- Rather than focussing on 'Working' Income Type, banks should focus on 'Commercial Associate', 'Pensioners' and 'State Servants' as they have the less difficulties with payment compared to 'Working' class.



THANK YOU!