

CREDIT EDA ASSIGNMENT

SUBMITTED BY: ABHISHEK PARAYIL

INTRODUCTION

This case study aims to give you an idea of applying EDA in a real business scenario. In this case study, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers

BUSINESS UNDERSTANDING - 1

- The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specialises in lending various types of loans to urban customers. You have to use EDA to analyse the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.
- When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:
 - If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
 - 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

BUSINESS UNDERSTANDING - 2

The data given below contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:

- The client with payment difficulties: he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample,
- All other cases: All other cases when the payment is paid on time.

When a client applies for a loan, there are four types of decisions that could be taken by the client/company):

1. Approved: The Company has approved loan Application
2. Cancelled: The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.
3. Refused: The company had rejected the loan (because the client does not meet their requirements etc.).
4. Unused offer: Loan has been cancelled by the client but on different stages of the process.

In this case study, we will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

- This case study aims to identify patterns which indicate if a client has difficulty paying their instalments which may be 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 of this case study.
- In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e., the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.
- To develop your understanding of the domain, you are advised to independently research a little about risk analytics - understanding the types of variables and their significance should be enough).

DATA UNDERSTANDING

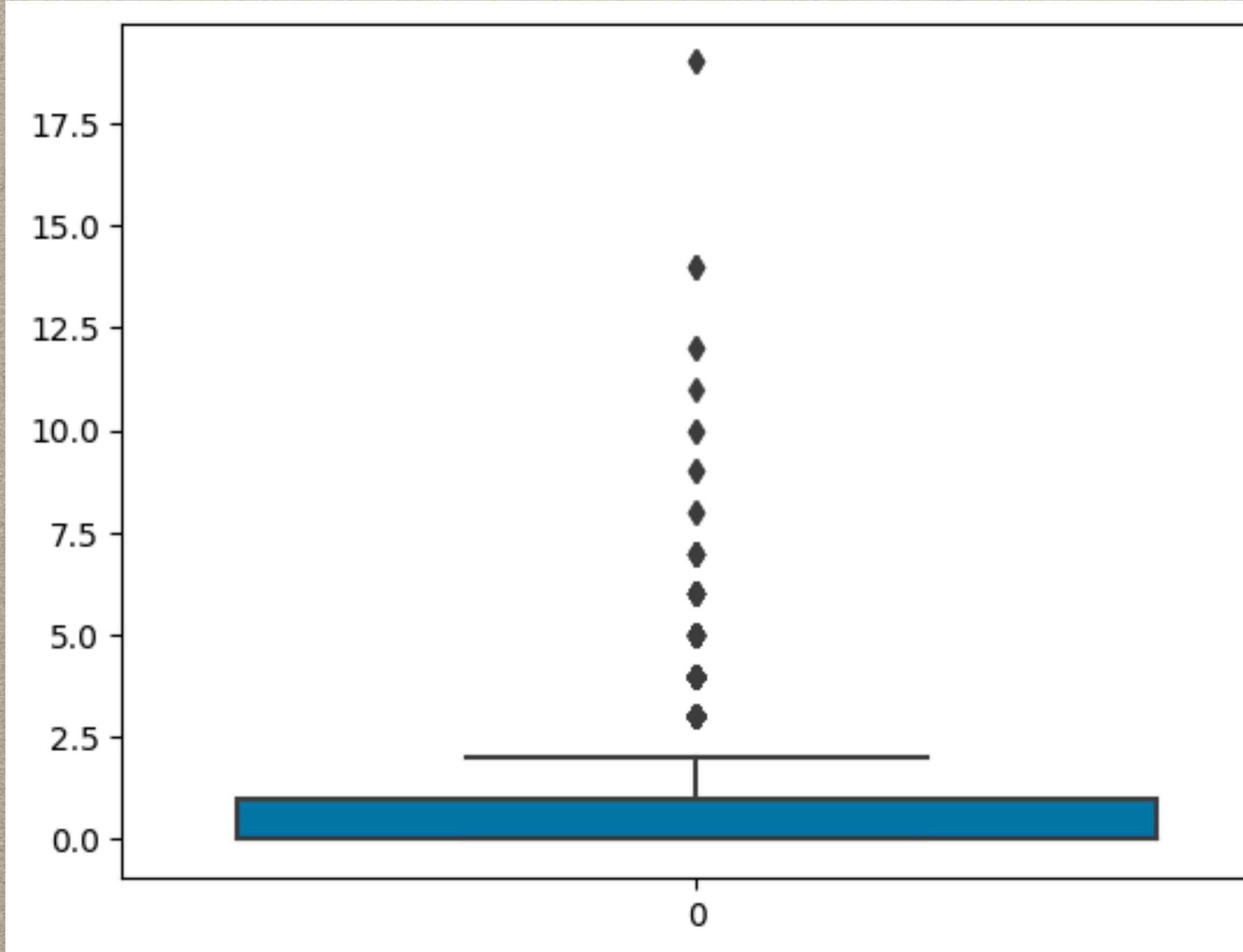
This dataset has 3 files as explained below:

- 1. 'application_data.csv' contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.
- 2. 'previous_application.csv' contains information about the client's previous loan data. It contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.
- 3. 'columns_description.csv' is data dictionary which describes the meaning of the variables.

ANALYSING APPLICATION_DATA .CSV

OUTLIER DETECTION

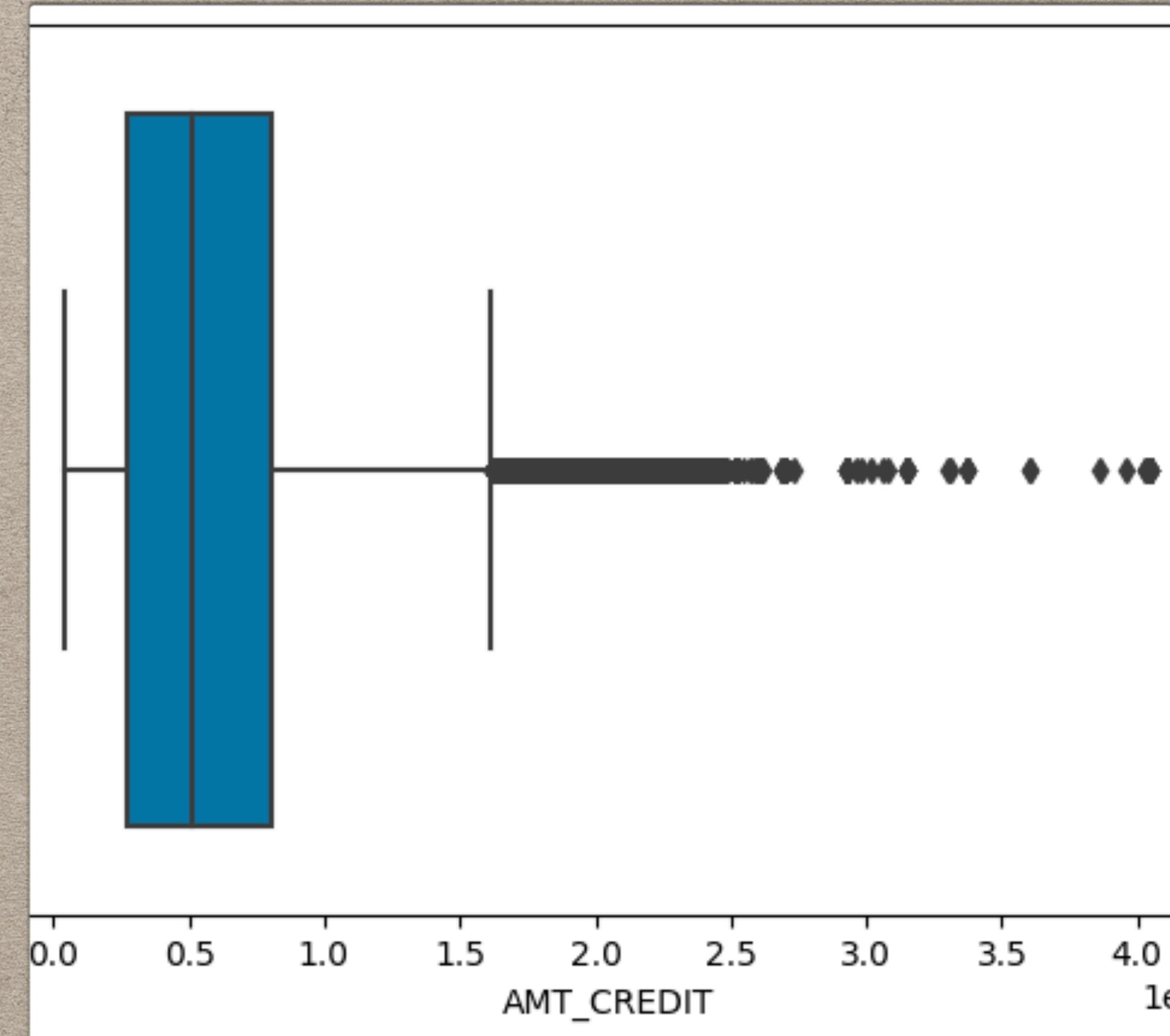
ANALYSING CNT_CHILDREN



- We can find that people with more than 7 children are very minimal and can be considered as an outlier
- Applicants with 3 or more children are outliers

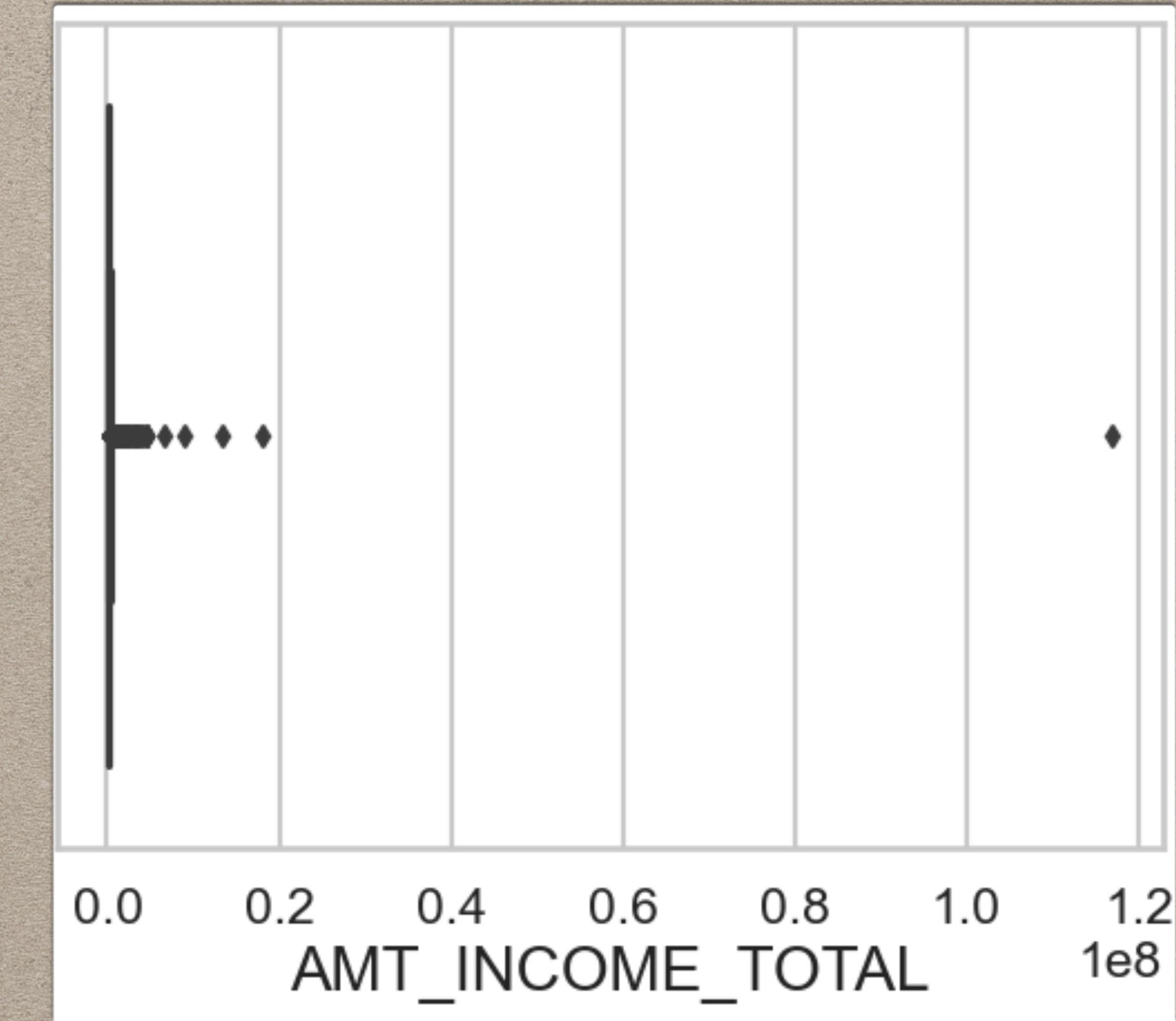
ANALYSING AMT_CREDIT

- As observed in the boxplot we can see that outliers exist after 3500000

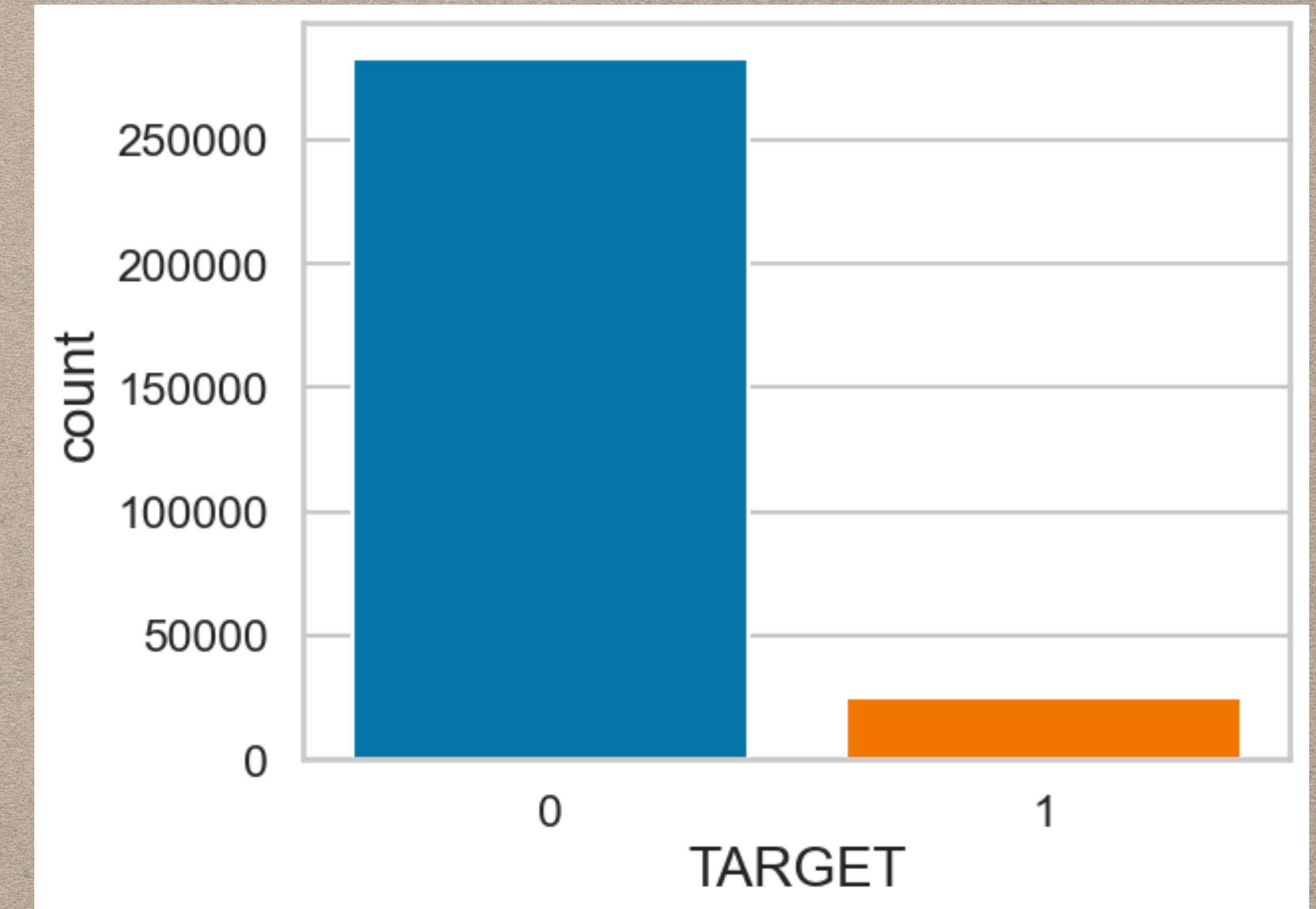
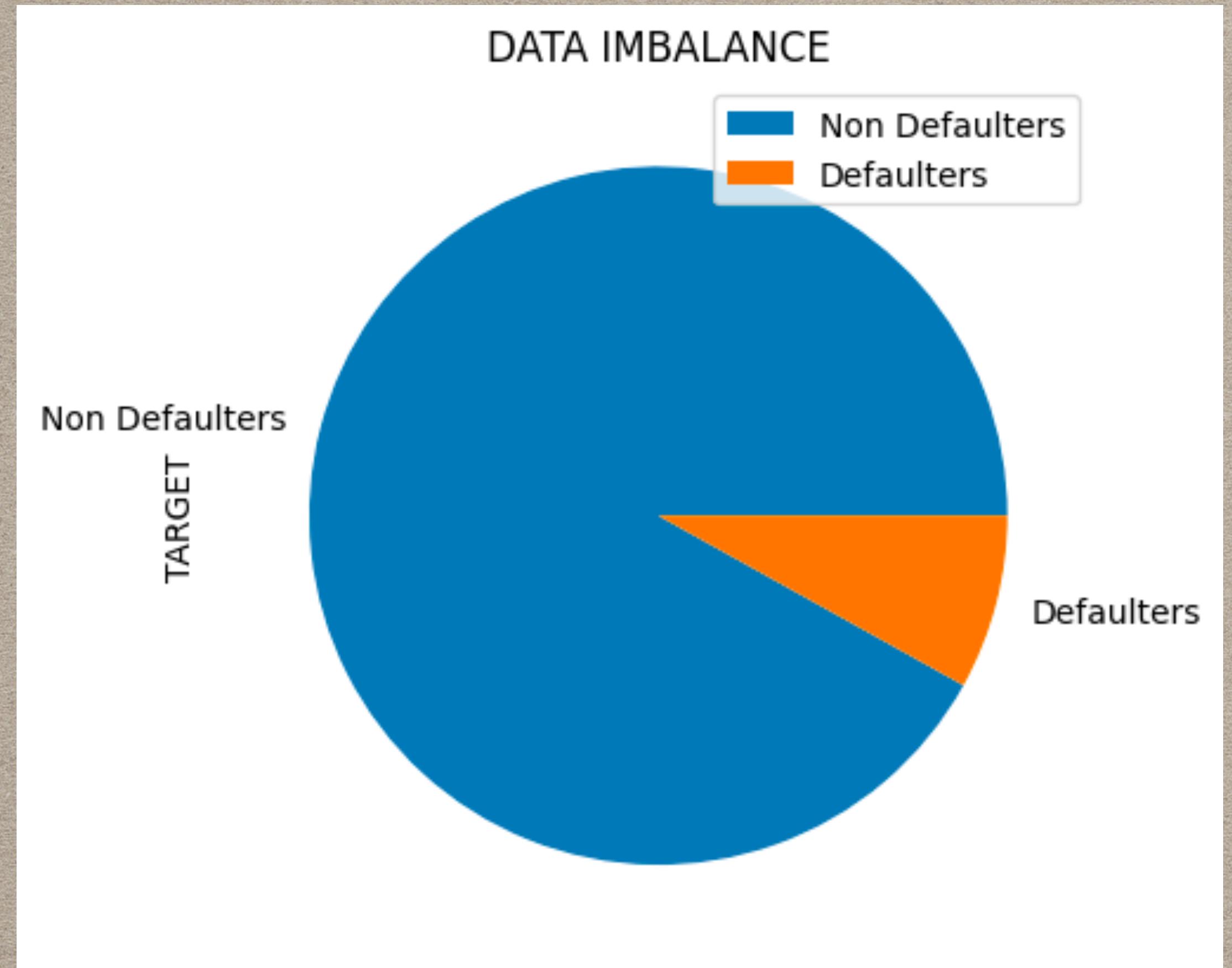


ANALYSING AMT_INCOME_TOTAL

- As seen in this plot we can see that outliers exist after 1000000



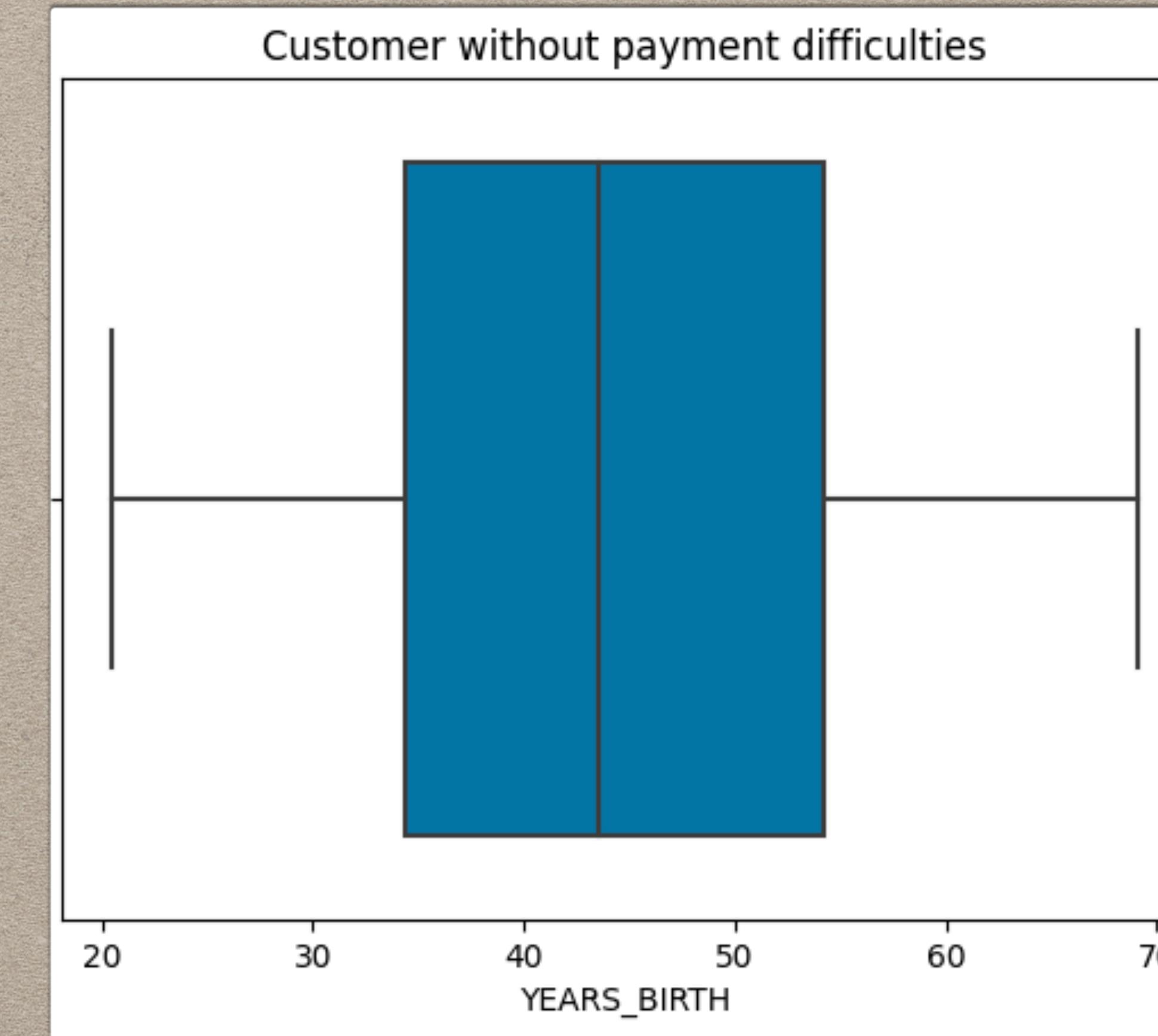
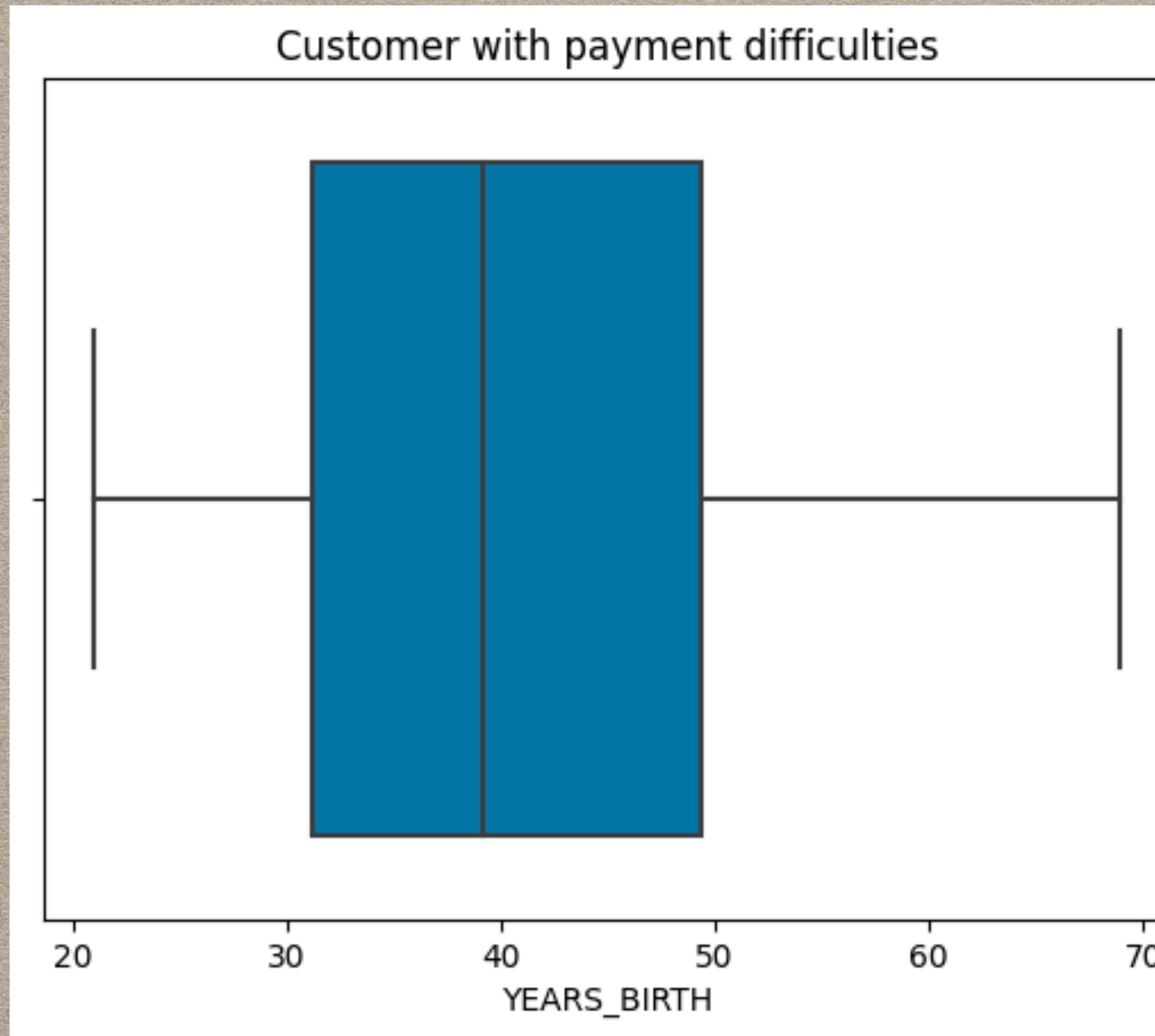
CHECKING FOR IMBALANCE OF DATA



1. There is imbalance of data in Target Variable.
2. Data imbalance ratio is 11.387:1
3. Data imbalance in percentages :
Target value 0 : 91.927
Target value 1 : 8.072 %

UNIVARIATE ANALYSIS

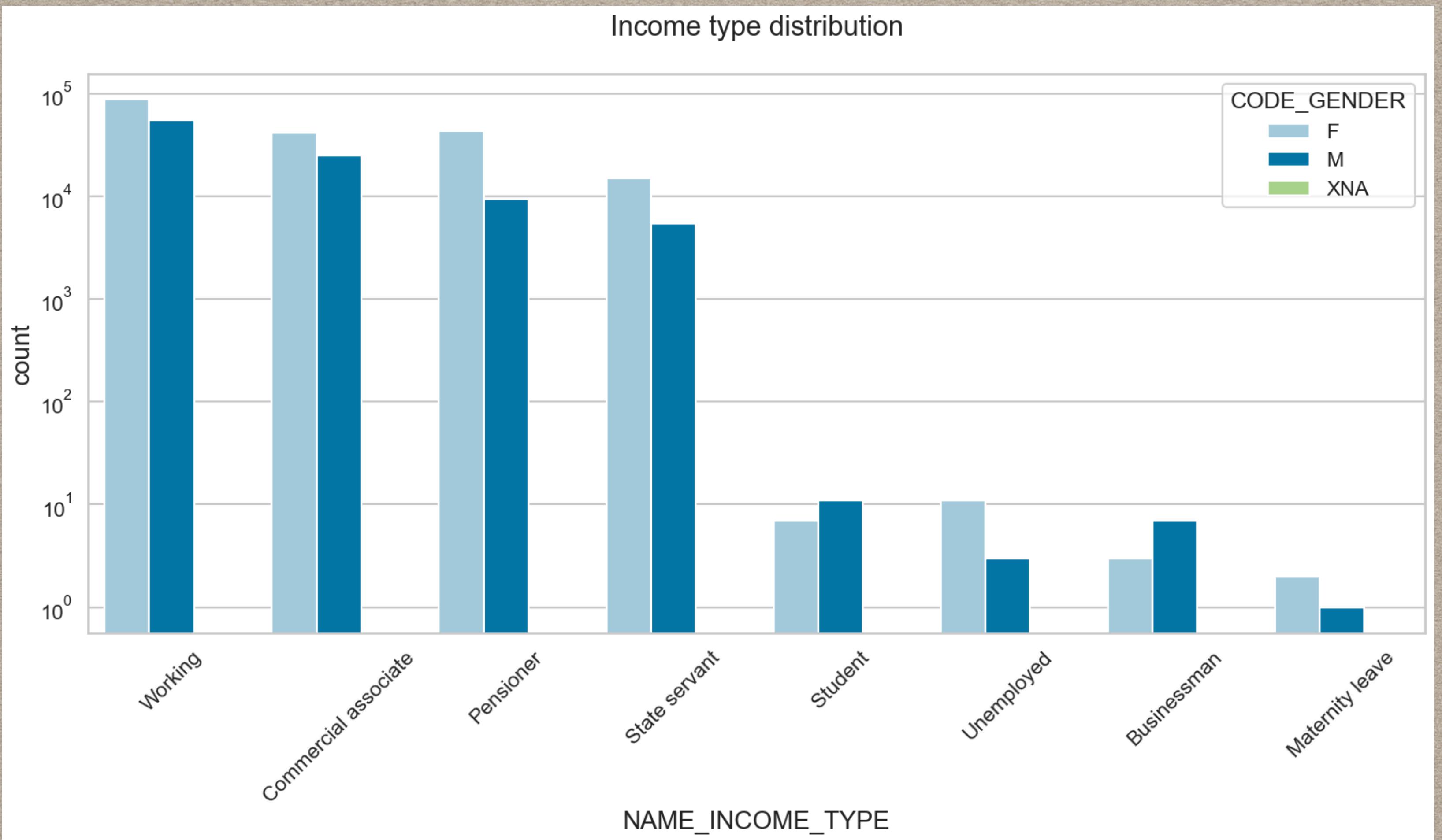
ANALYSIS OF YEARS_BIRTH



From this we can infer that

1. Ages of customer with payment difficulties lies in between 31 and 50
2. Ages of customer without payment difficulties lies in between 34 and 54

Income type distribution

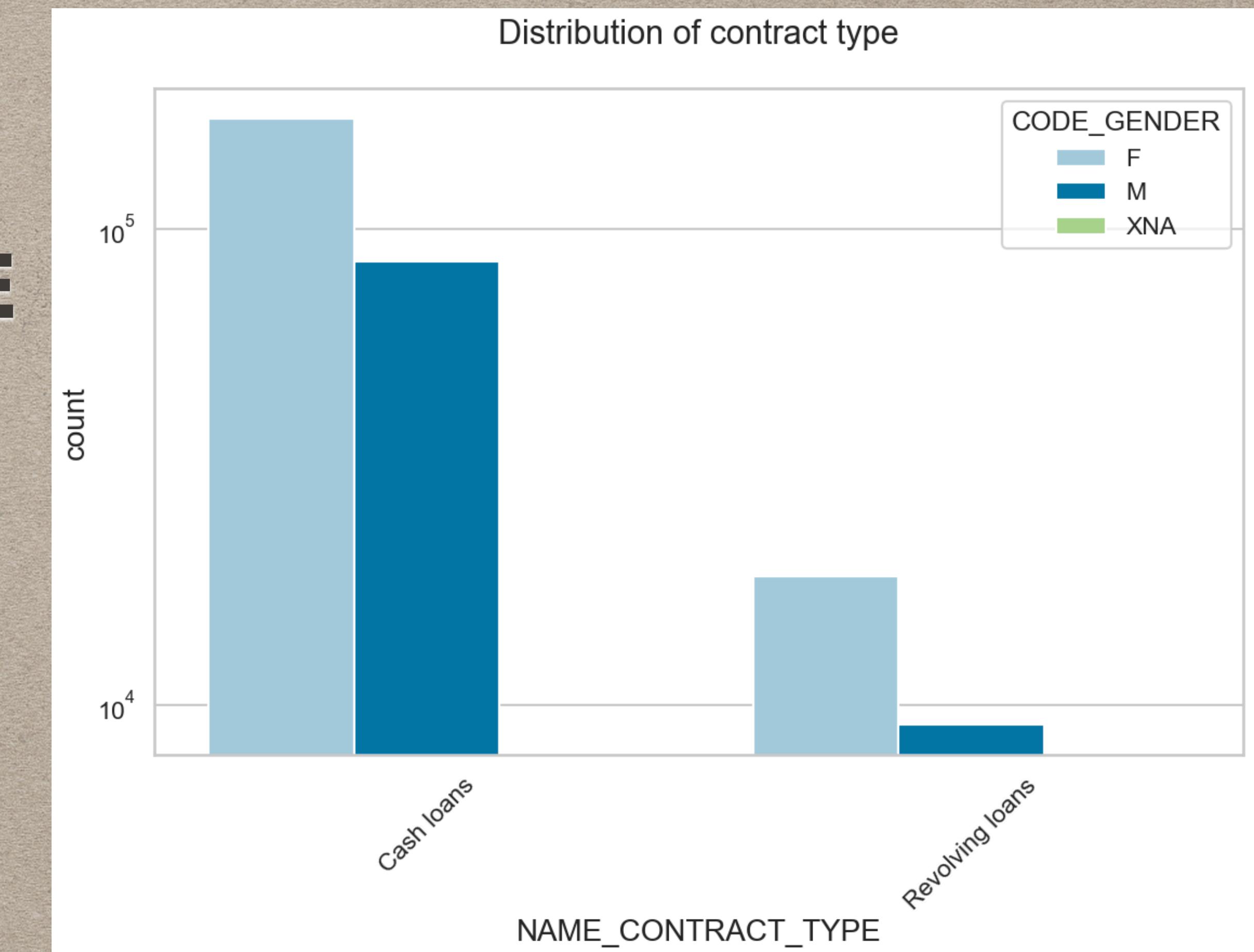


ANALYSING NAME_INCOME_TYPE

1. Student, Unemployed, Businessman has significantly less income than others
2. Working, Commercial associate, Pensioner and State servants have more income
3. Females has more credit than male

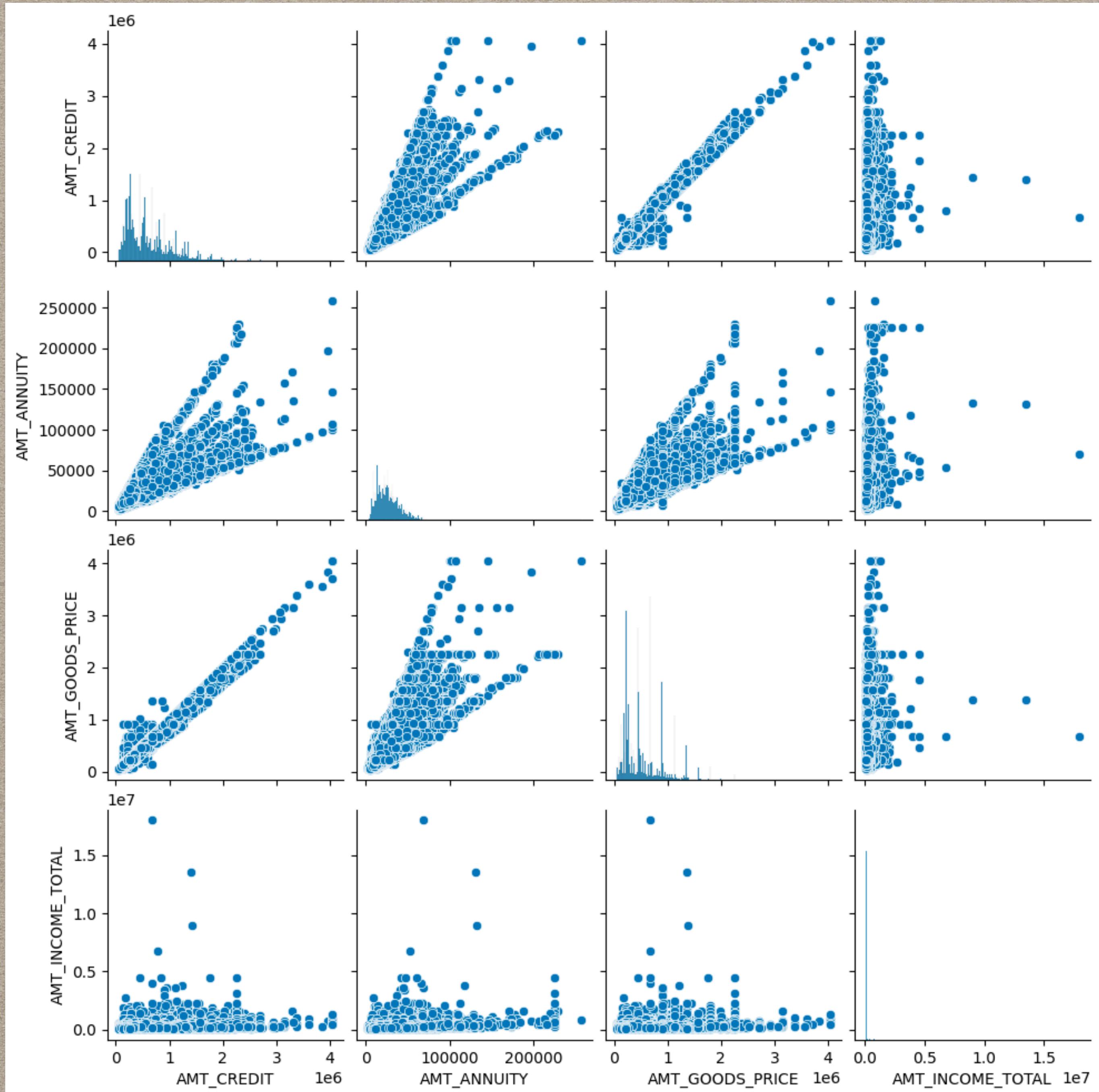
ANALYSING NAME_CONTRACT_TYPE

1. Females has more applications
2. Cash loans have more number than revolving loans

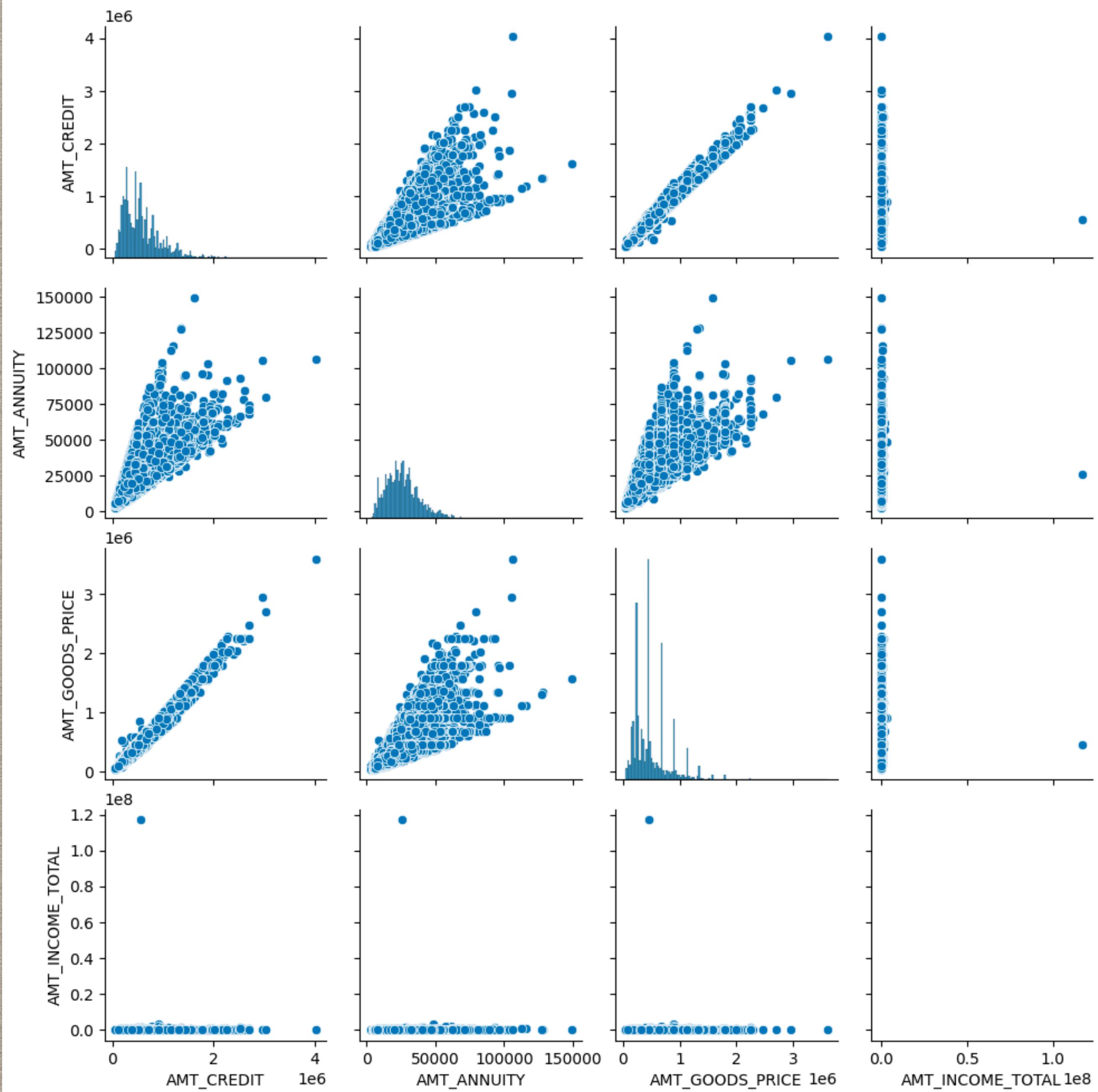


ANALYSING NUMERIC VALUES

**FOR PEOPLE WITHOUT
PAYMENT DIFFICULTY**



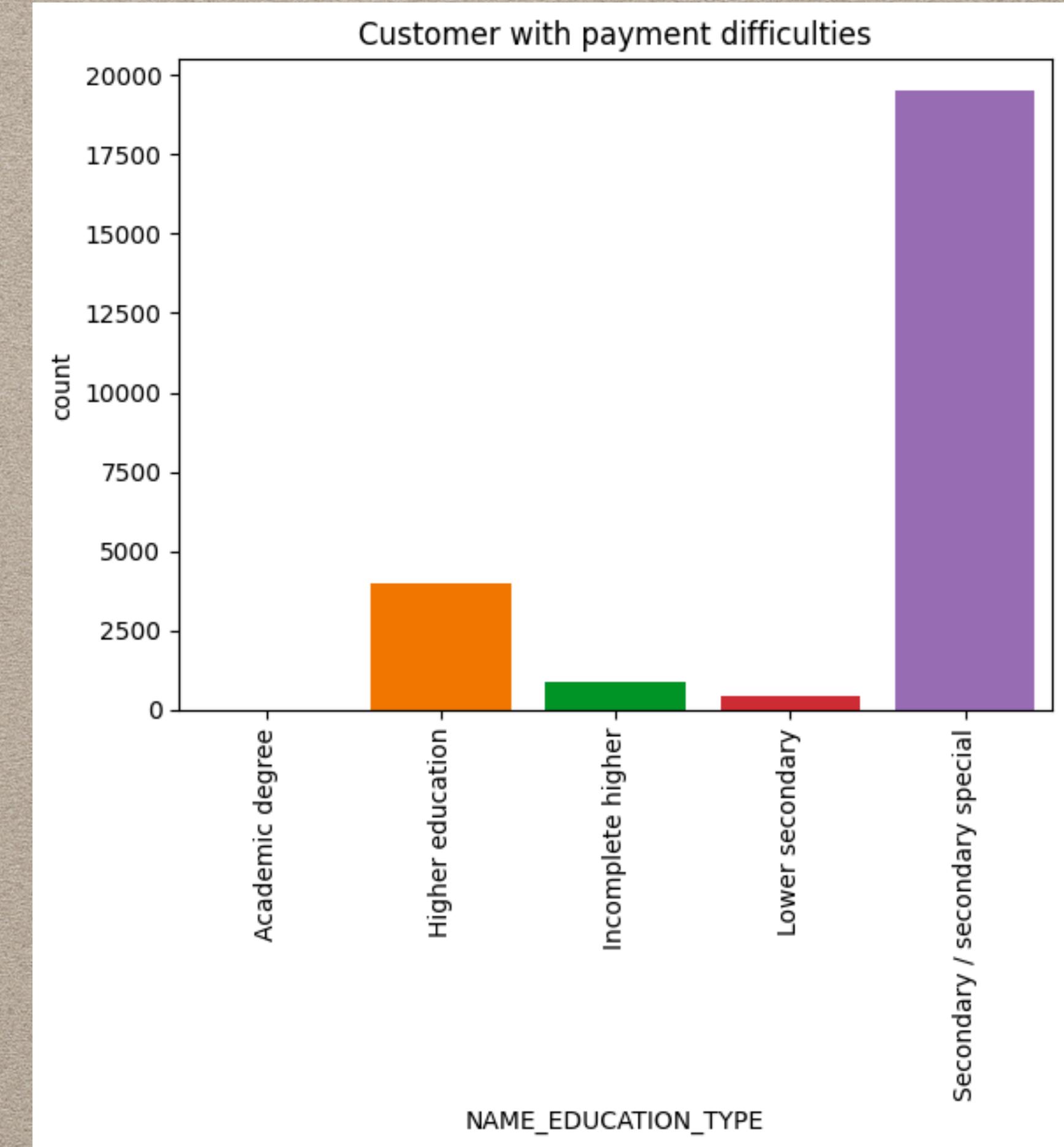
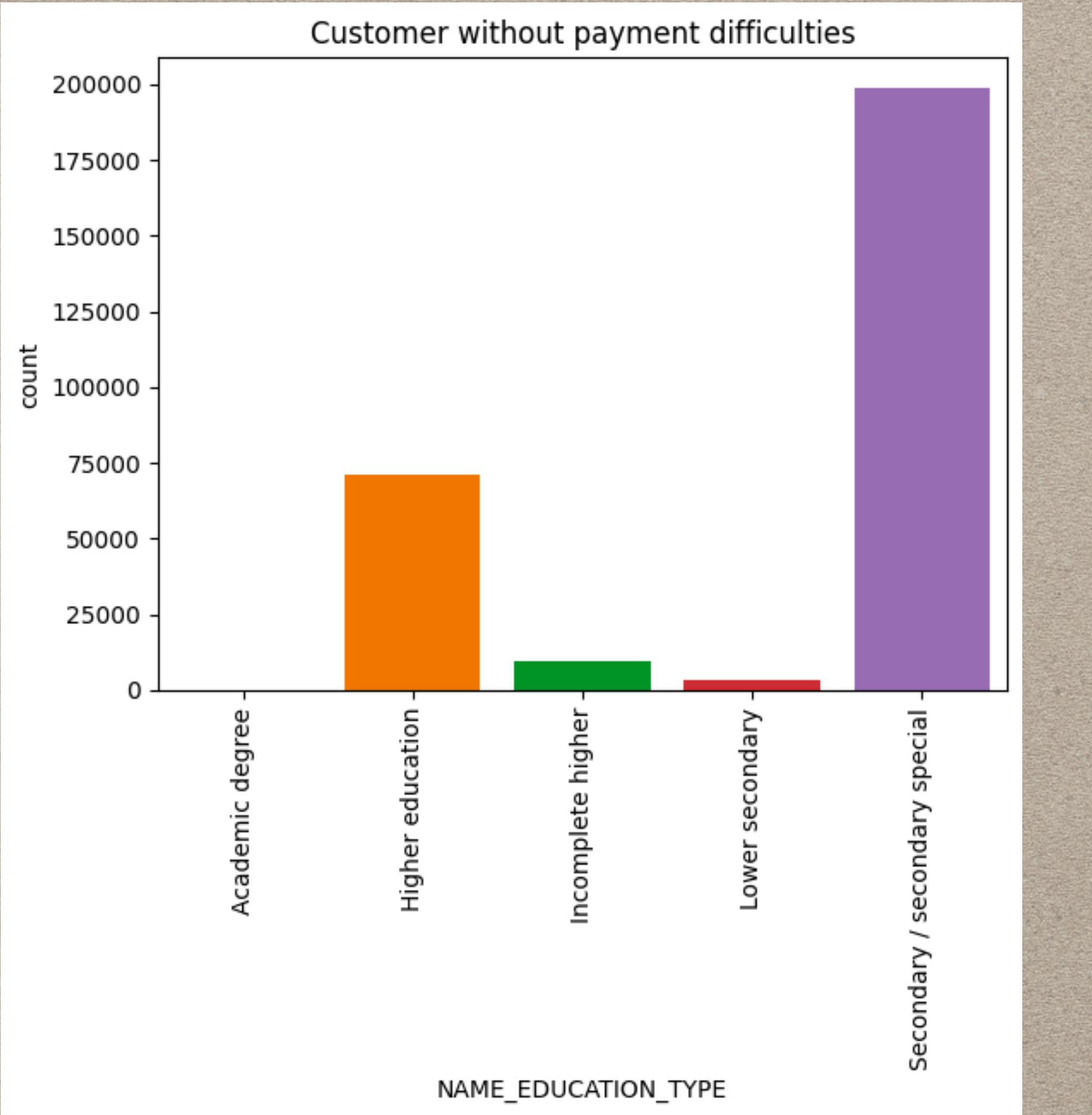
FOR PEOPLE WITH PAYMENT DIFFICULTY



INSIGHTS

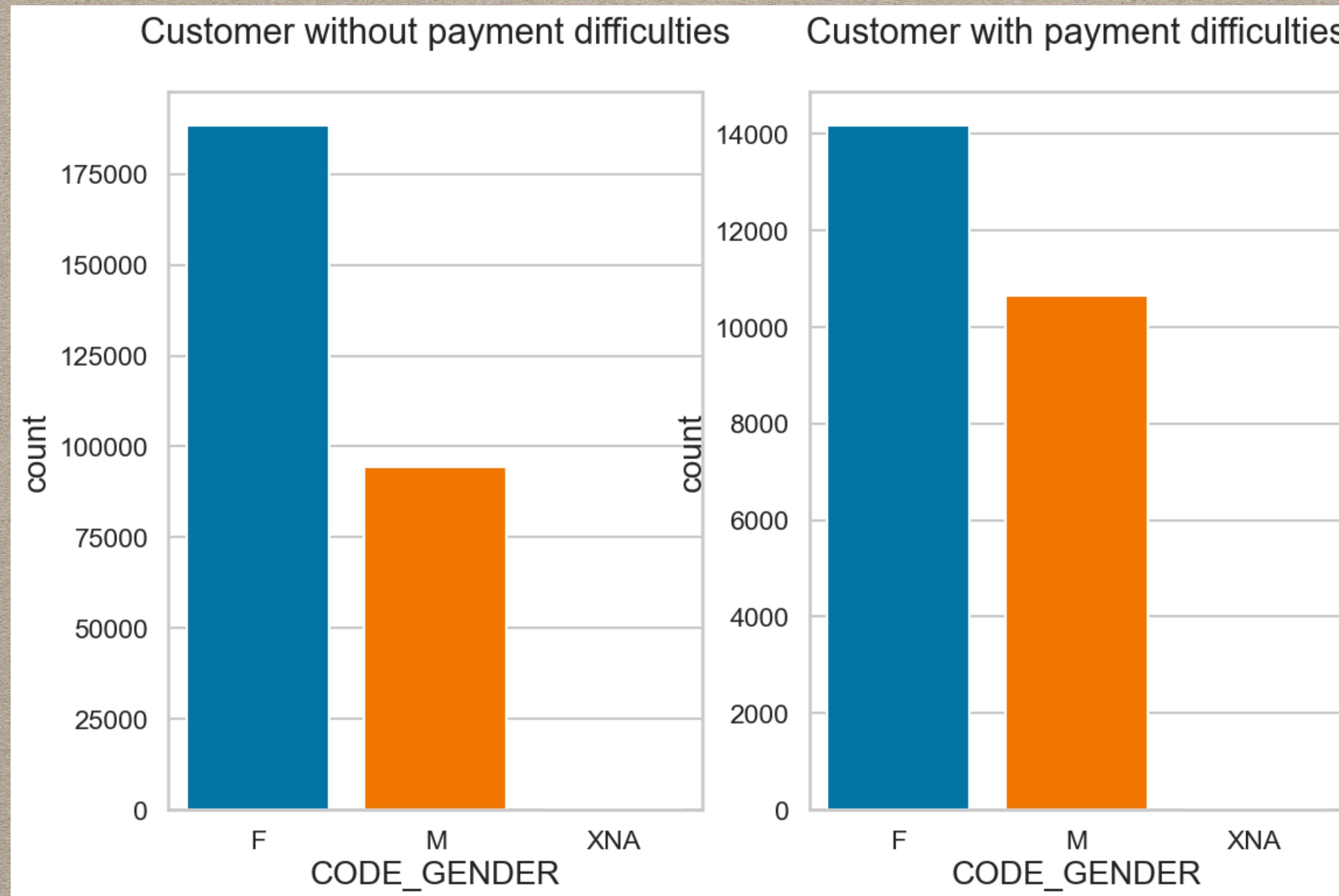
1. *AMT_CREDIT and AMT_ANNUITY share a good correlation, from this we can infer that as the amount of loan goes higher the number of EMI also goes higher*
2. *AMT_CREDIT and AMT_GOODS_PRICE also share a very high correlation between them in both the cases of people having payment difficulties and people not having payment difficulties. This can happen because in most cases of consumer loans the loan amount is provided to buy goods and when the goods amount goes up the loan amount also goes up*

ANALYSING NAME_EDUCATION_TYPE



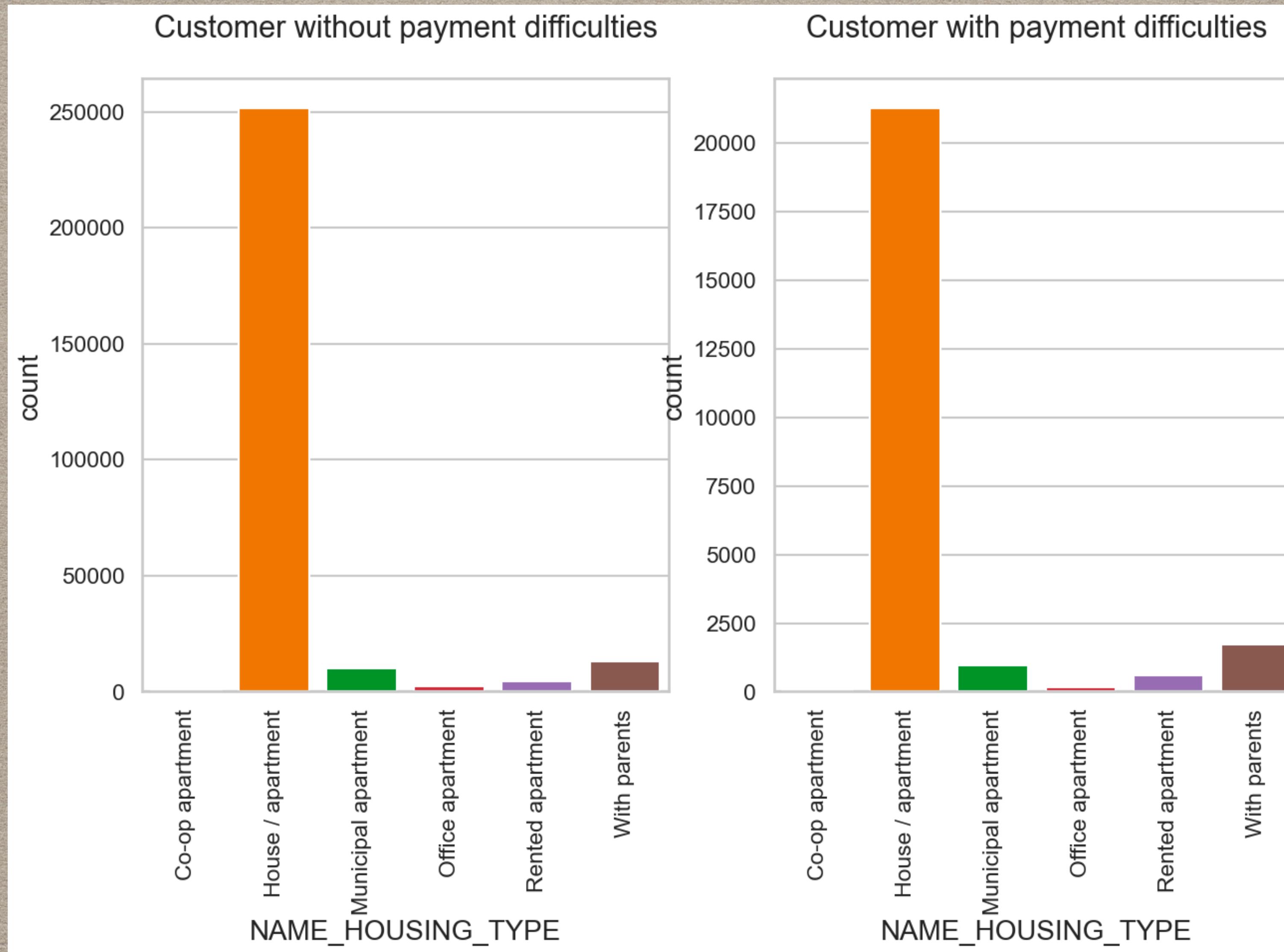
From the above 2 plots we can find that customer with secondary/secondary special education is more in both the cases as more academic people tend to take less amount of loan

ANALYSING CODE_GENDER



We can find that female applicants are better compared to male as there are less defaulters in female than in male applicants

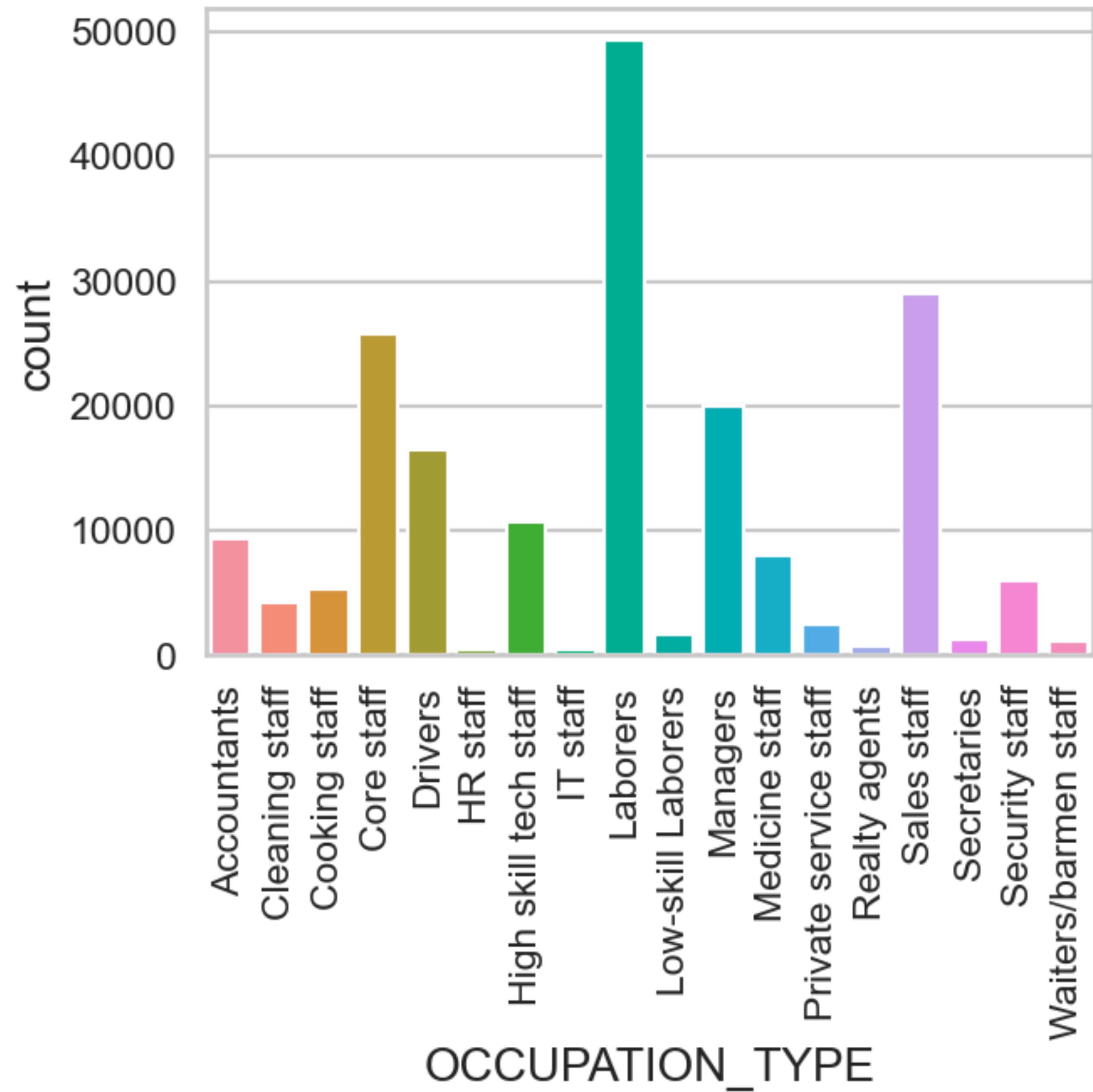
ANALYSING NAME_HOUSING_TYPE



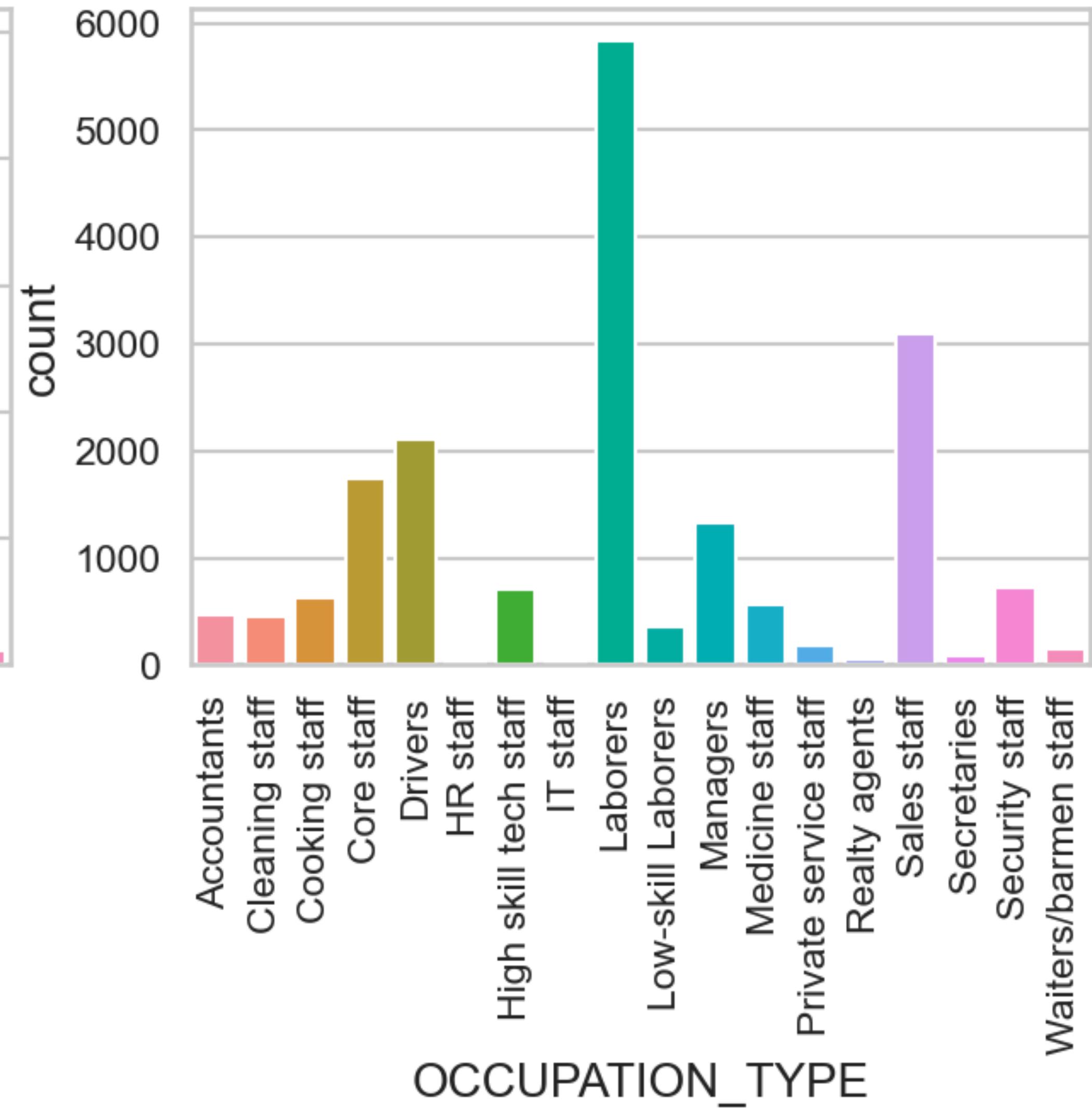
We can infer that people with House / Apartment are more in both the cases.

ANALYSING OCCUPATION_TYPE

Customer without payment difficulties



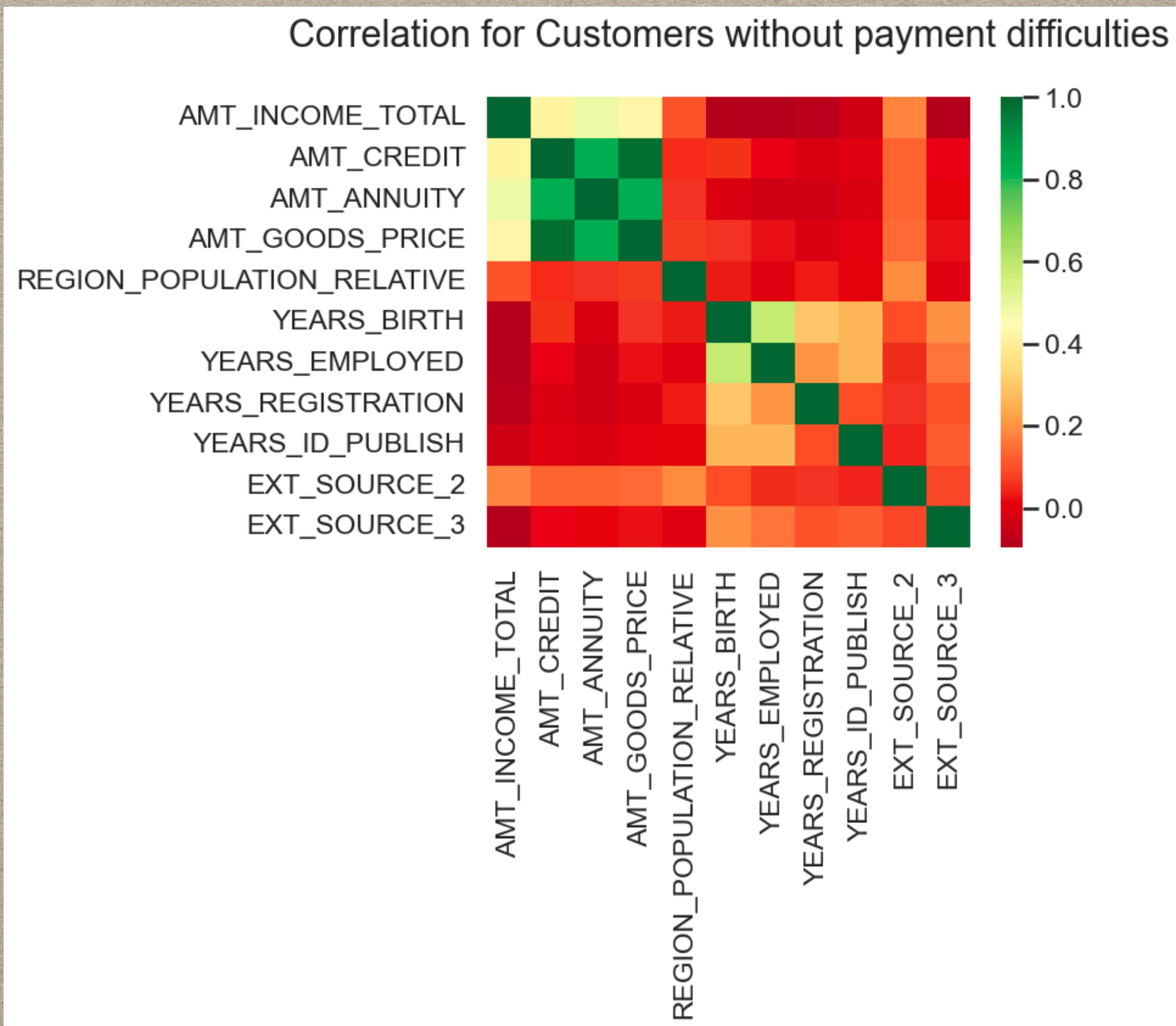
Customer with payment difficulties



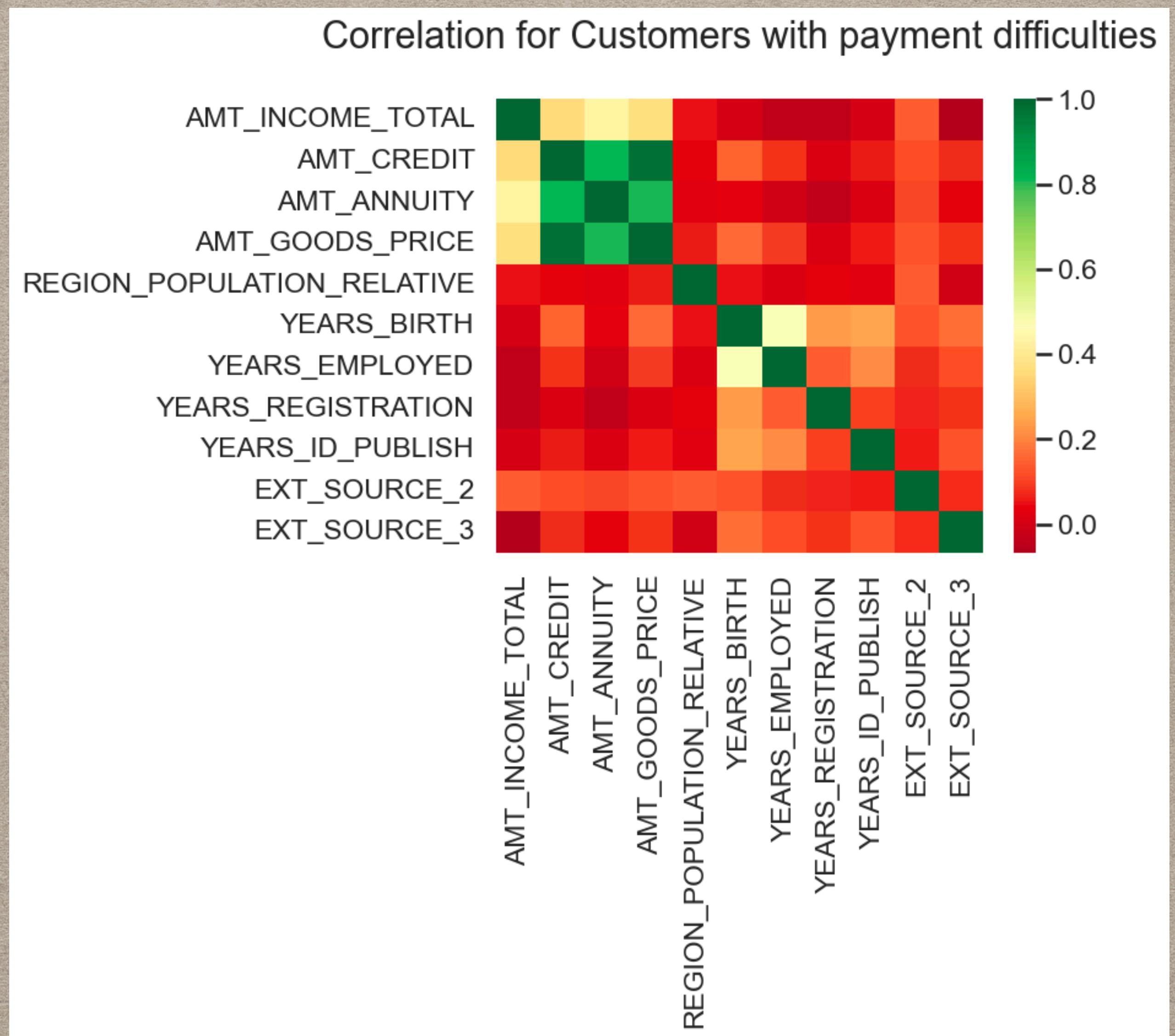
INSIGHTS

- 1. We can find that the labourers are the customers who have difficulty in repaying the loan and then comes sales staff and then core staff**
- 2. But also we can find that labourers are the customers who have higher number in repaying the loan correctly**
-

CORRELATION ANALYSIS FOR TARGET O



CORRELATION ANALYSIS FOR TARGET 1



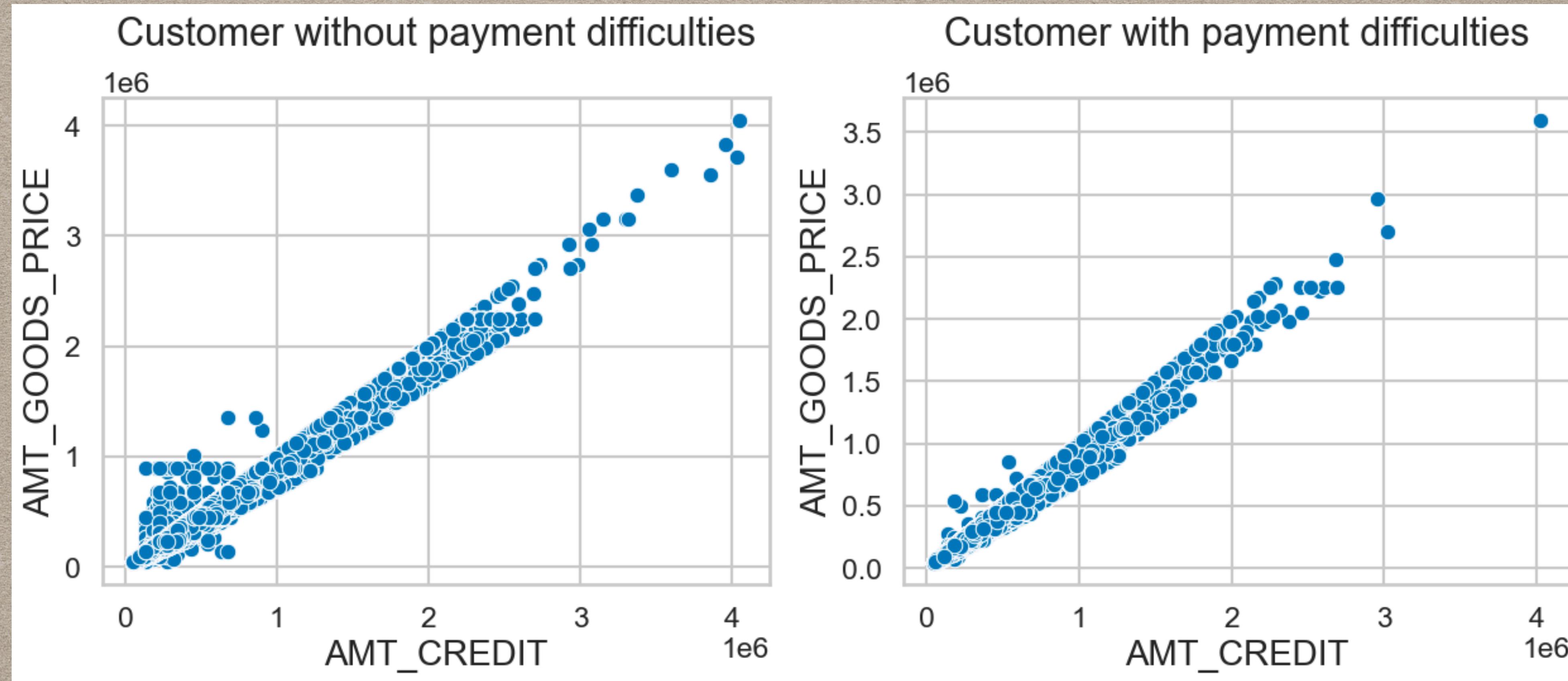
INSIGHTS

- 1. Credit amount is inversely proportional to age**
- 2. Credit amount is higher in densely populated area.**

BI-VARIATE ANALYSIS

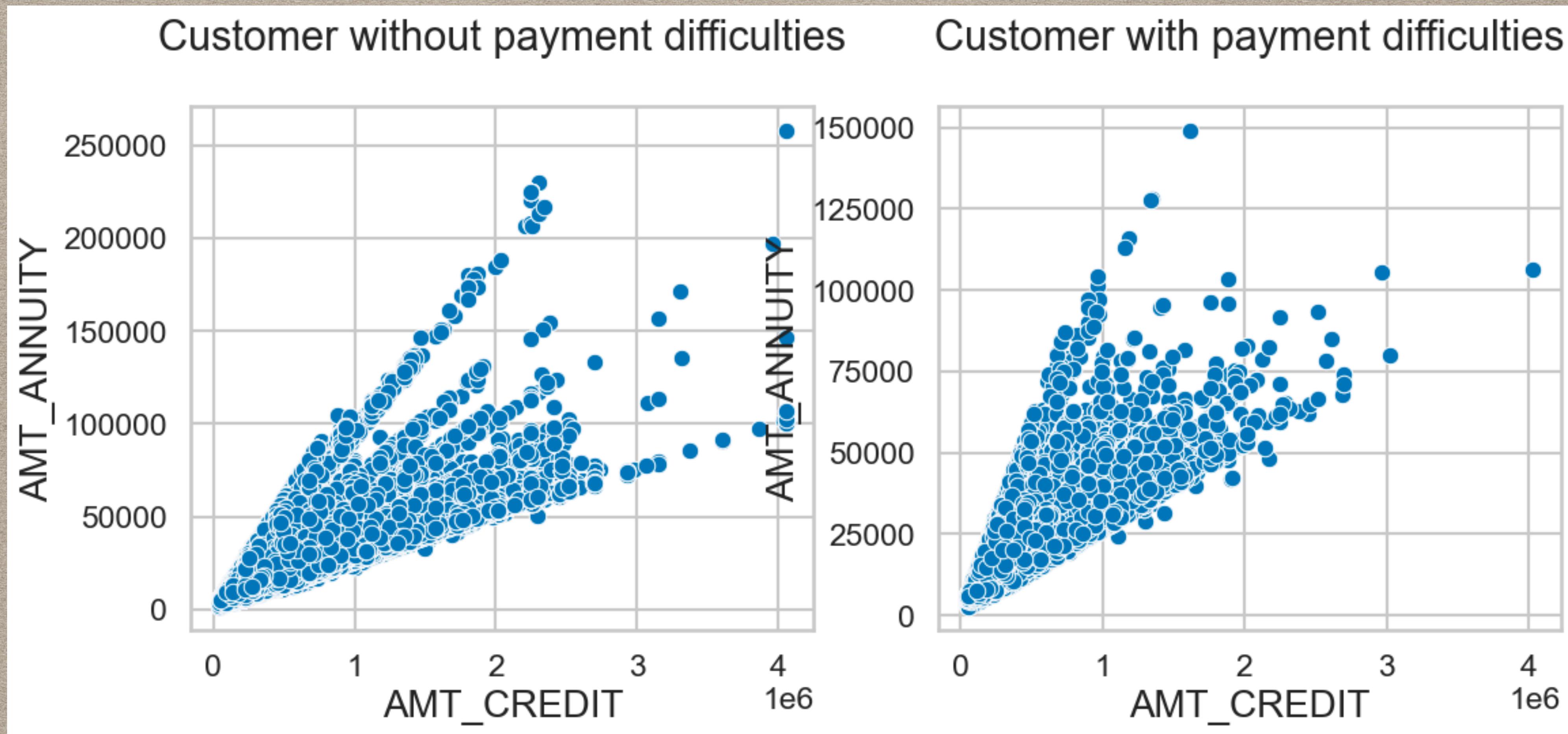
NUMERICAL-NUMERICAL ANALYSIS

ANALYSING AMT_CREDIT AND AMT_GOODS_PRICE



From the above plots we can infer that the goods price is correlated with the credit amount.

ANALYSING AMT_CREDIT AND AMT_ANNUITY

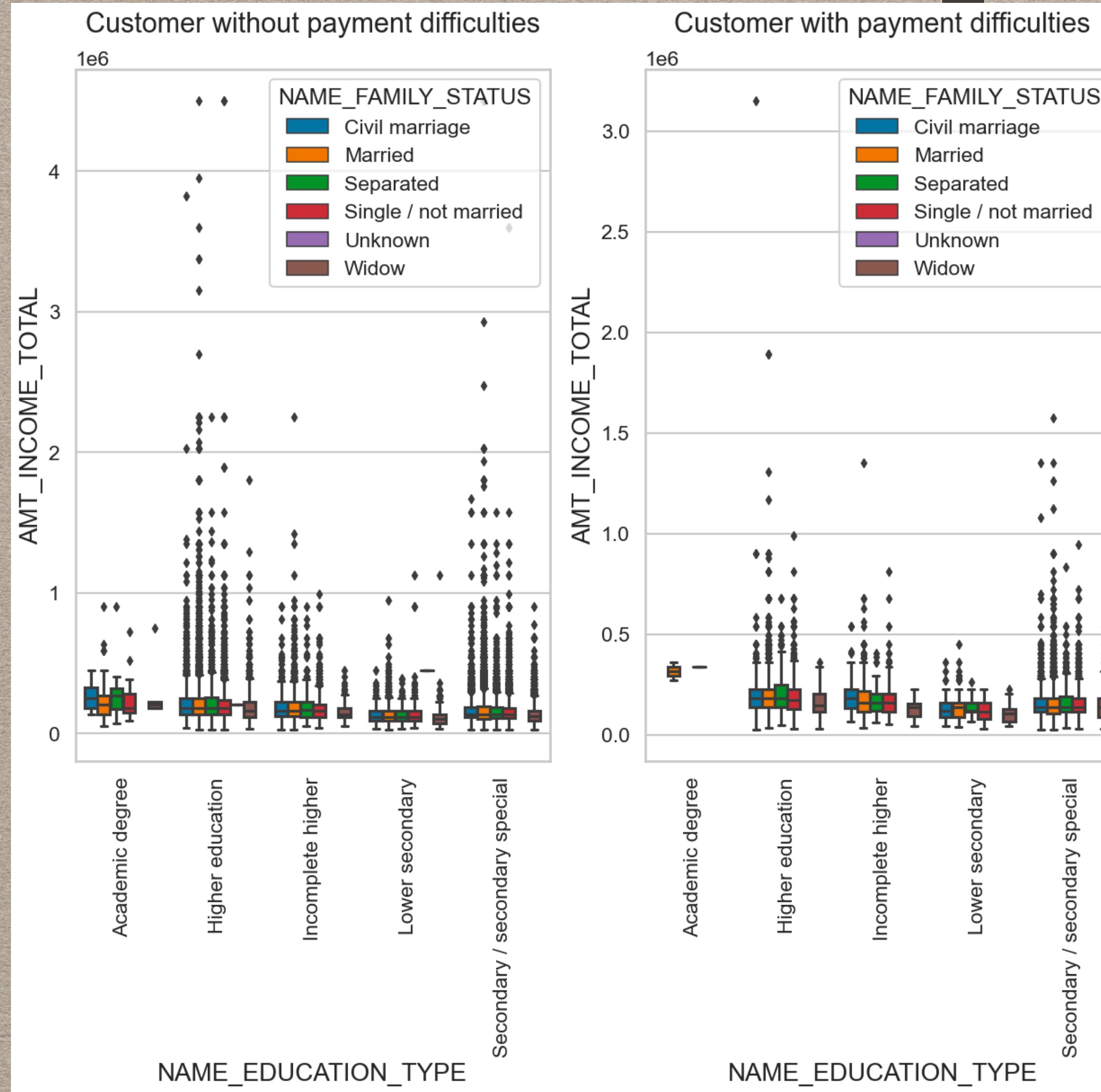


From the above plots we can infer that people without payment difficulties goes for more credit with the annuity that they have

BI-VARIATE ANALYSIS

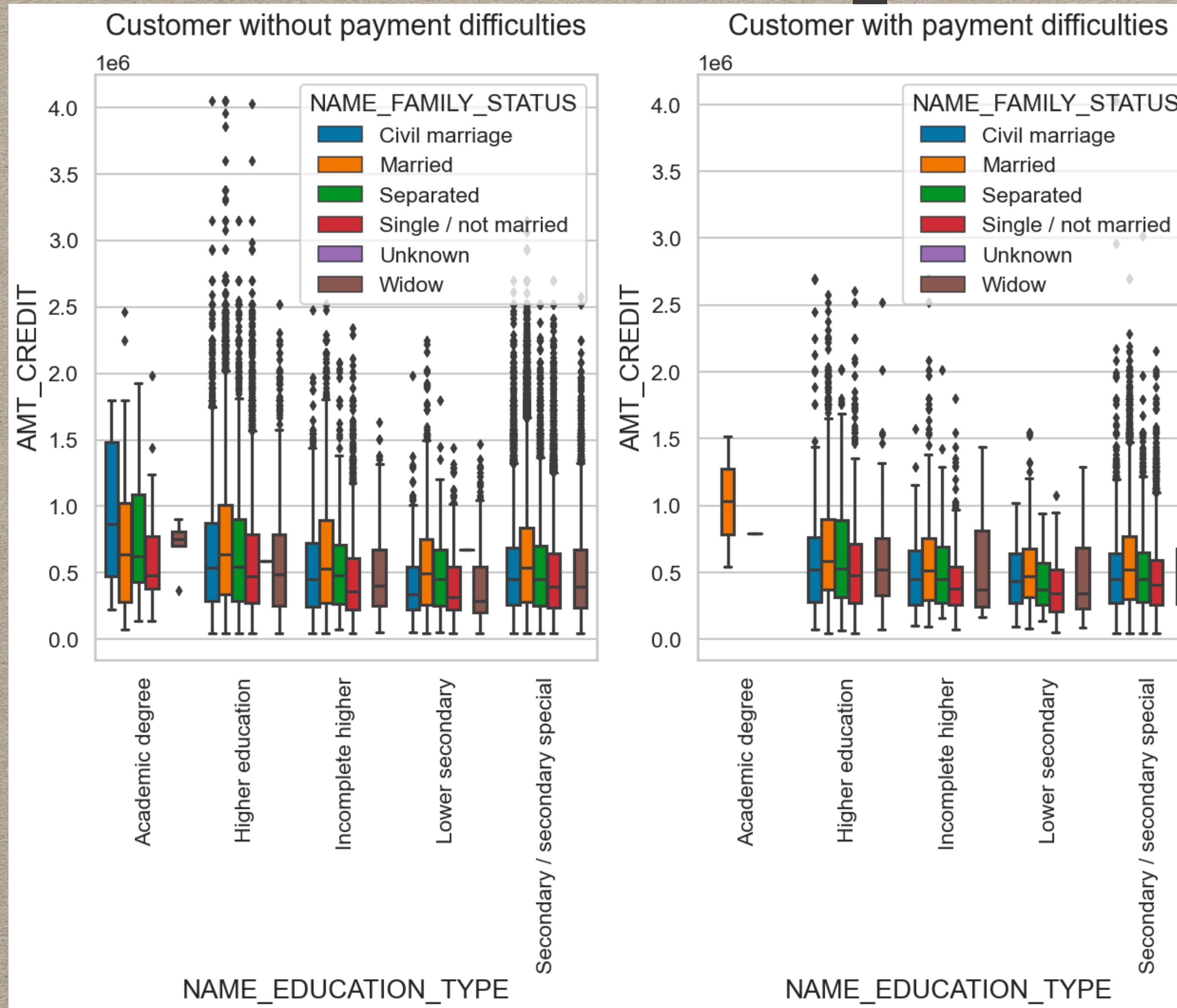
CATEGORICAL-CATEGORICAL ANALYSIS

ANALYSING AMT_INCOME_TOTAL AND NAME_EDUCATION_TYPE



- 1. Lower secondary of civil marriage family status are having less income amount than others.**
- 2. In education type 'Higher education' the income amount is mostly equal with family status. It does contain many outliers.**
- 3. There are less outlier in Academic degree but their income amount is little higher than that of Higher education.**

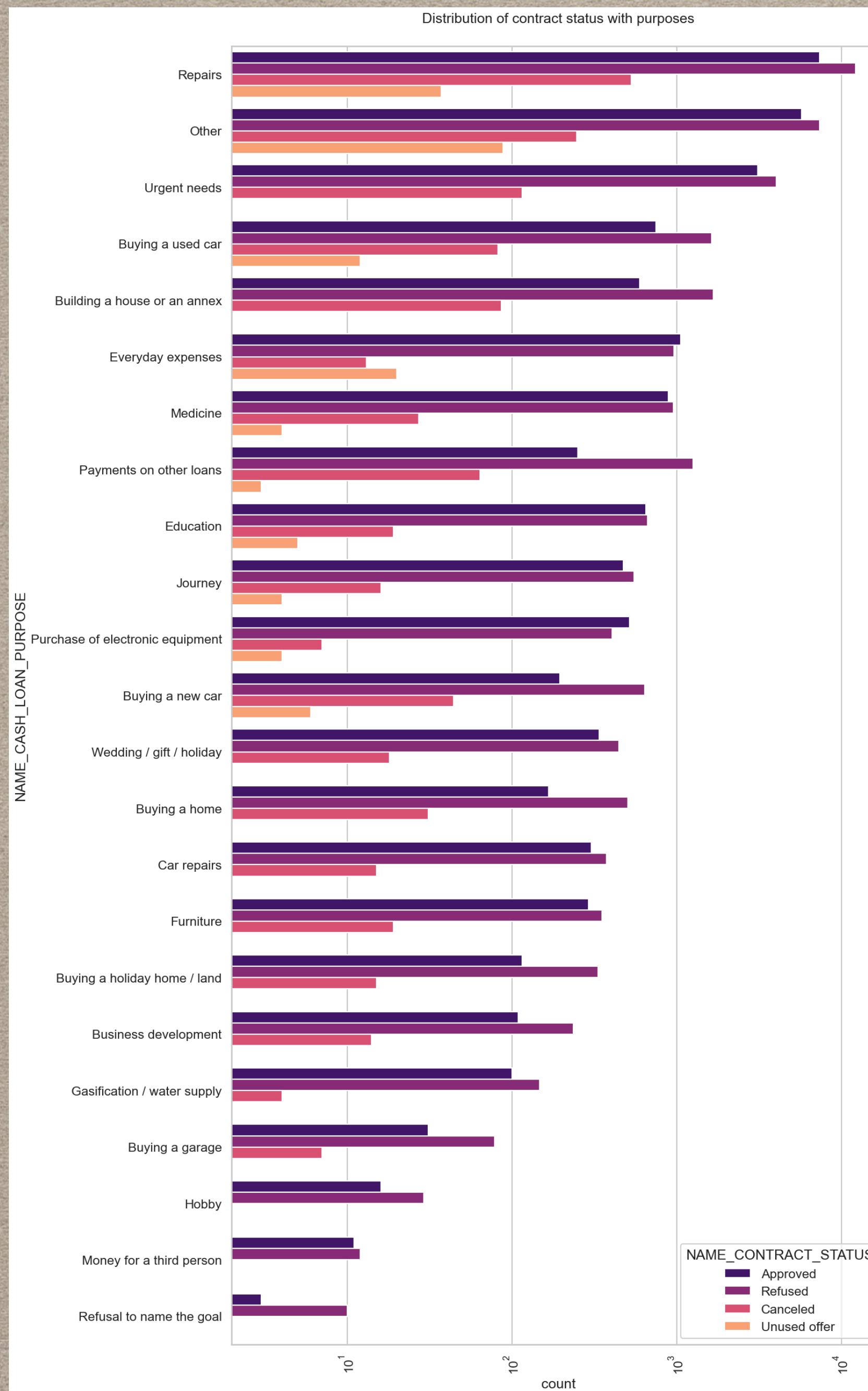
ANALYSING AMT_CREDIT AND NAME_EDUCATION_TYPE



- 1. Higher education of family status of 'marriage', 'single' and 'civil marriage' are having more outliers. Civil marriage for Academic degree is having most of the credits in the third quartile.**
- 2. From the above box plot we can conclude that Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.**

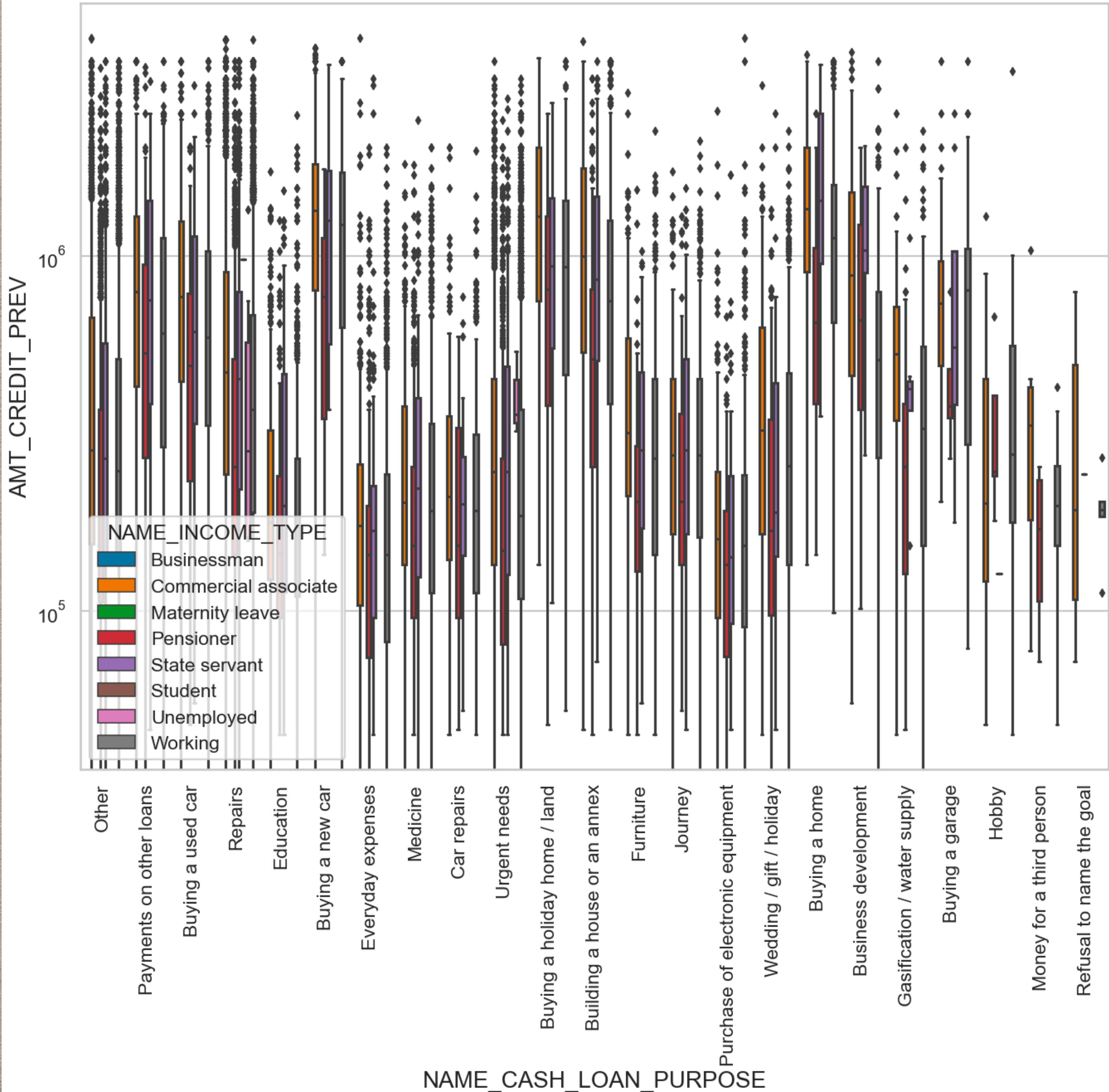
ANALYSIS OF INFORMATION IN PREVIOUS_DATA

UNIVARIATE ANALYSIS : ANALYSING NAME_CASH_LOAN_PURPOSE



- Buying New car and Paying other loans has significantly more rejection than approval*
- Repairs have the most number of rejections*
- Purchase of Electronic equipment and Everyday expenses are the only two Purposes which have more approvals than rejection*
- Education has almost equal amount of Approvals and rejections*

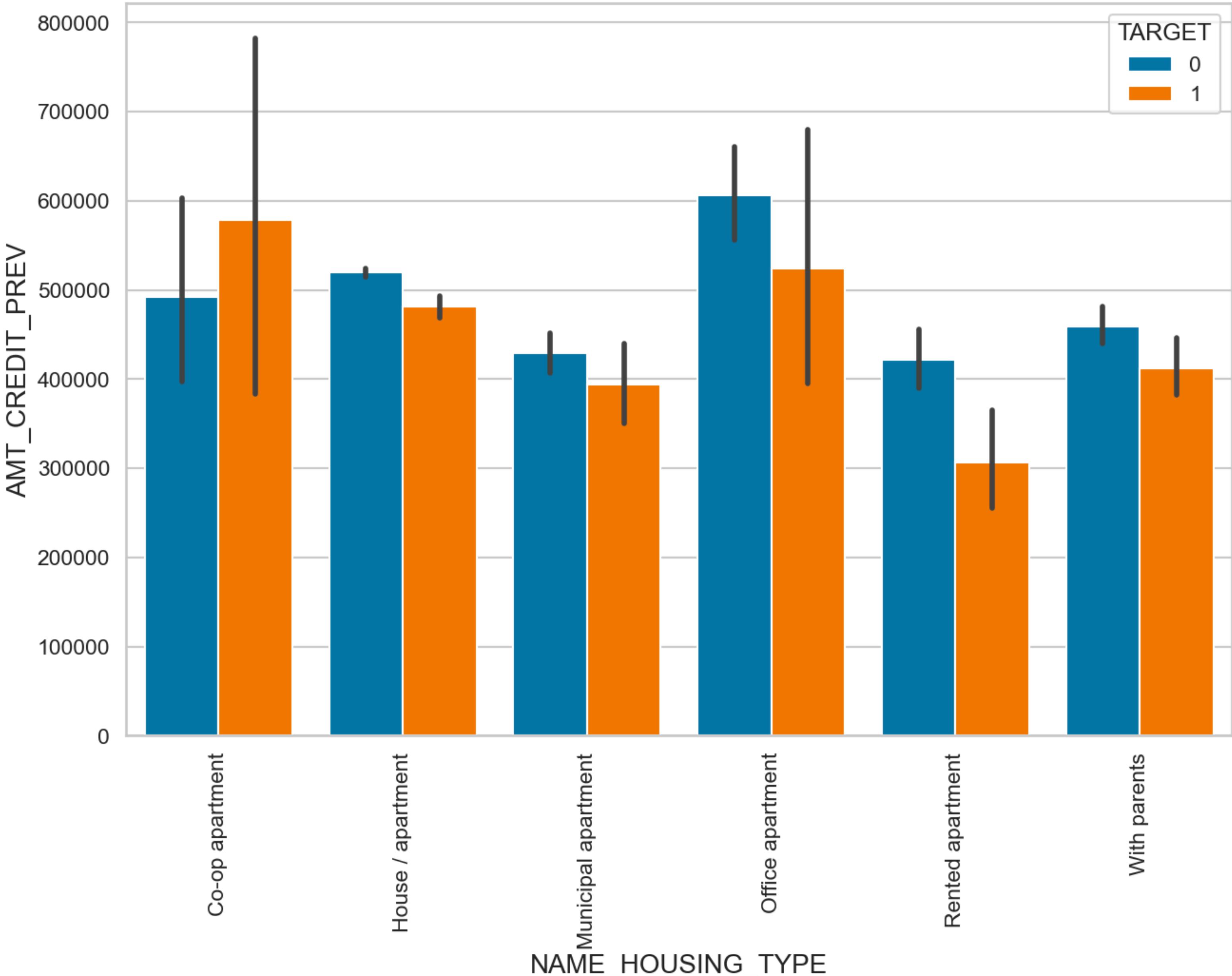
Prev Credit amount vs Loan Purpose



MULTIVARIATE ANALYSIS : ANALYSING NAME_CASH_LOAN_PURPOSE AND AMT_CREDIT_PREV WITH NAME_INCOME_TYPE

- 1. Money for third person or a Hobby is having low number of credits applied for.**
- 2. The credit amount of Loan purposes like 'Buying a home', 'Buying a land', 'Buying a new car' and 'Building a house' is higher.**
- 3. Income type of state servants have a significant amount of credit applied**

Prev Credit amount vs Housing type



MULTIVARIATE ANALYSIS : ANALYSING AMT_CREDIT_PREV AND NAME_HOUSING_TYPE WITH TARGET

Office apartment is having higher credit for customers without payment difficulty and co-op apartment is having higher credit for customers with payment difficulty, So, we can conclude that bank should avoid giving loans to the housing type of co-op apartment as they are having difficulties in payment. Bank can focus mostly on housing type with parents or House\apartment or municipal apartment for successful payments.

CONCLUSION

- 1. Banks must obtain the most clients who fall under the housing category "With Parents," since they have the fewest failed payments.**
- 2. Loan purpose 'Repair' also has higher percentage of unsuccessful payments on time.**
- 3. Banks should avoid giving loan to income type 'Working' as their number of unsuccessful payments are high**
- 4. Banks focus should be more on contract type 'Student' , 'pensioner' and 'Businessman' with housing 'type other than 'Co-op apartment' for successful payments.**