



GOLD ATLANTIS: REPAYING THE CREDIT

OBJECTIVE



The Gold Atlantis is a credit-providing firm, recently they observed a rise in the number of defaulters



The aim is to help the bank identify the customers who have a lower probability of becoming a defaulter.





DATA PREPARATION

&

CLEANING

1000000 rows , 24 columns

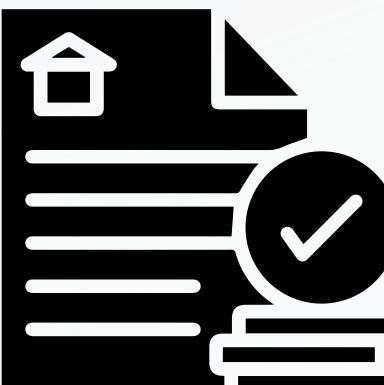
Checked for null values:

Null values are present in

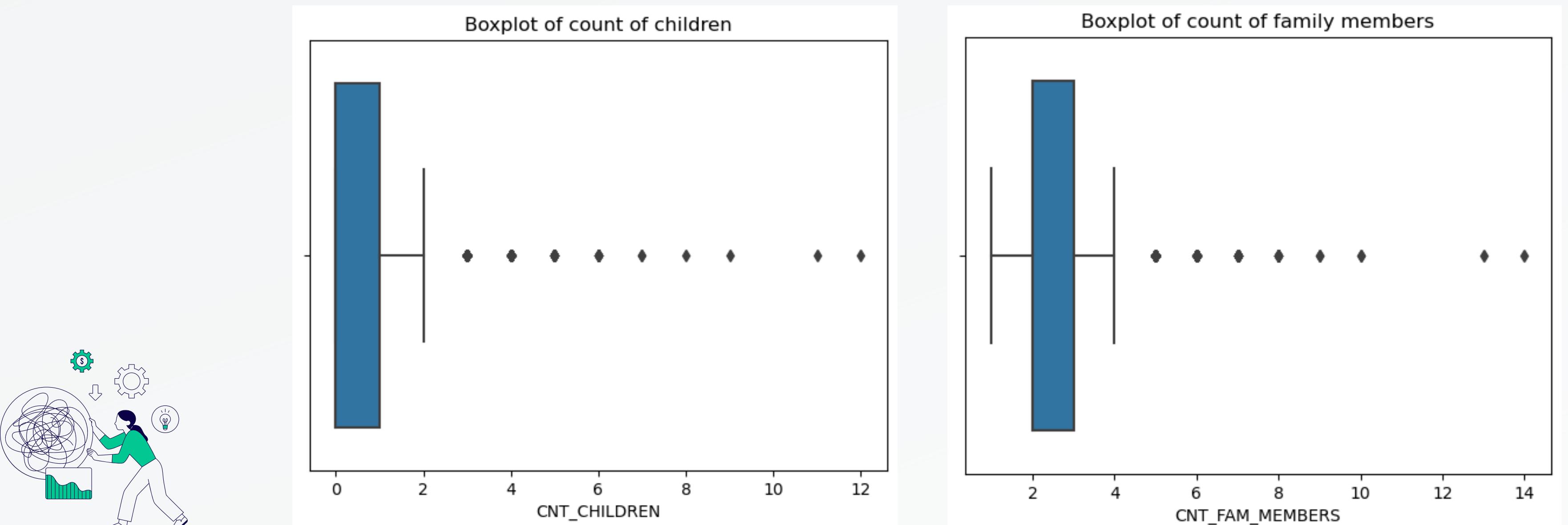
- AMT_GOODS_PRICE
- NAME_TYPE_SUITE
- OCCUPATION_TYPE columns

Treating the null values:

The above three columns are categorical, so replaced the null values with the mode of the data.



- Gender column - ‘xna’ values - Replaced with the mode
- Family status column - Replaced civil marriage values with married
- Occupation type column - too many categories - Added a new column, skill type with categories skilled and unskilled.
- Income type column - Replaced state servant, commercial associate values with working
- Count of children column - Removed outliers
- Count of family members - Removed outliers



After cleaning, the categorical columns are as follows:

Categorical Columns:

- TARGET: 1 - defaulters, 0 - payers
- NAME_CONTRACT_TYPE: Cash loans, Revolving loans
- GENDER: M, F
- Car: N, Y
- House: N, Y
- CNT_CHILDREN: 0 to 3
- NAME_TYPE_SUITE: Unaccompanied, Spouse, etc..
- NAME_INCOME_TYPE: Working, not working

Categorical Columns:

- NAME_EDUCATION_TYPE: Secondary / Higher education
- NAME_FAMILY_STATUS: Single, Married, etc
- MOBILE: 0,1
- WORK_PHONE: 0,1
- HOME_PHONE: 0,1
- MOBILE_REACHABLE: 0,1
- FLAG_EMAIL: 0,1
- SKILL_LEVEL: Skilled/Unskilled
- CNT_FAM_MEMBERS : 0 to 13
- APPLICATION_DAY: S, M, T, W, Th, F, Sa
- TOTAL_DOC_SUBMITTED: 0 to 4

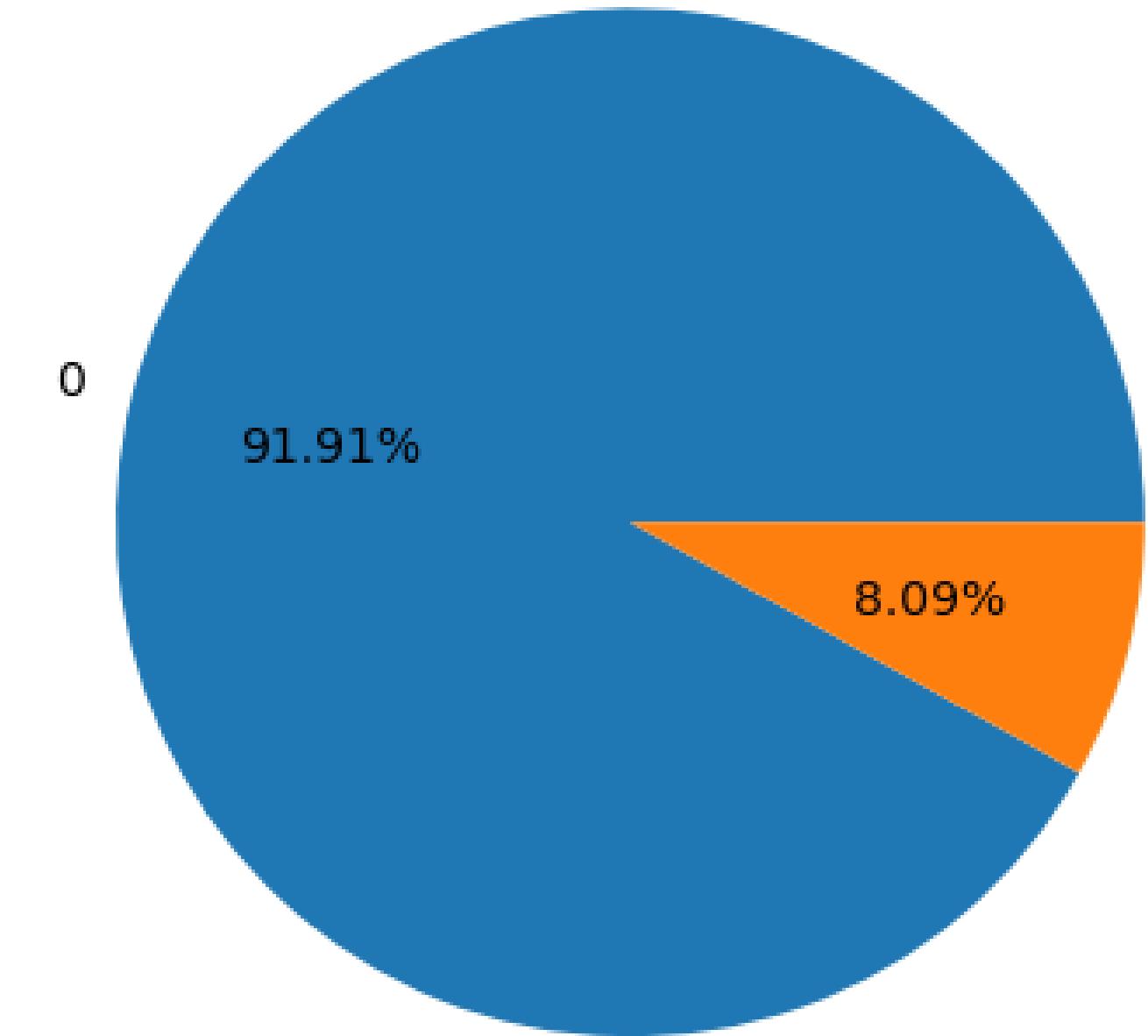
Disclaimer: As the phone number and email ID are the basic information a person will provide, we are not analyzing those columns.

DATA ANALYSIS

UNIVARIATE AND BIVARIATE ANALYSIS

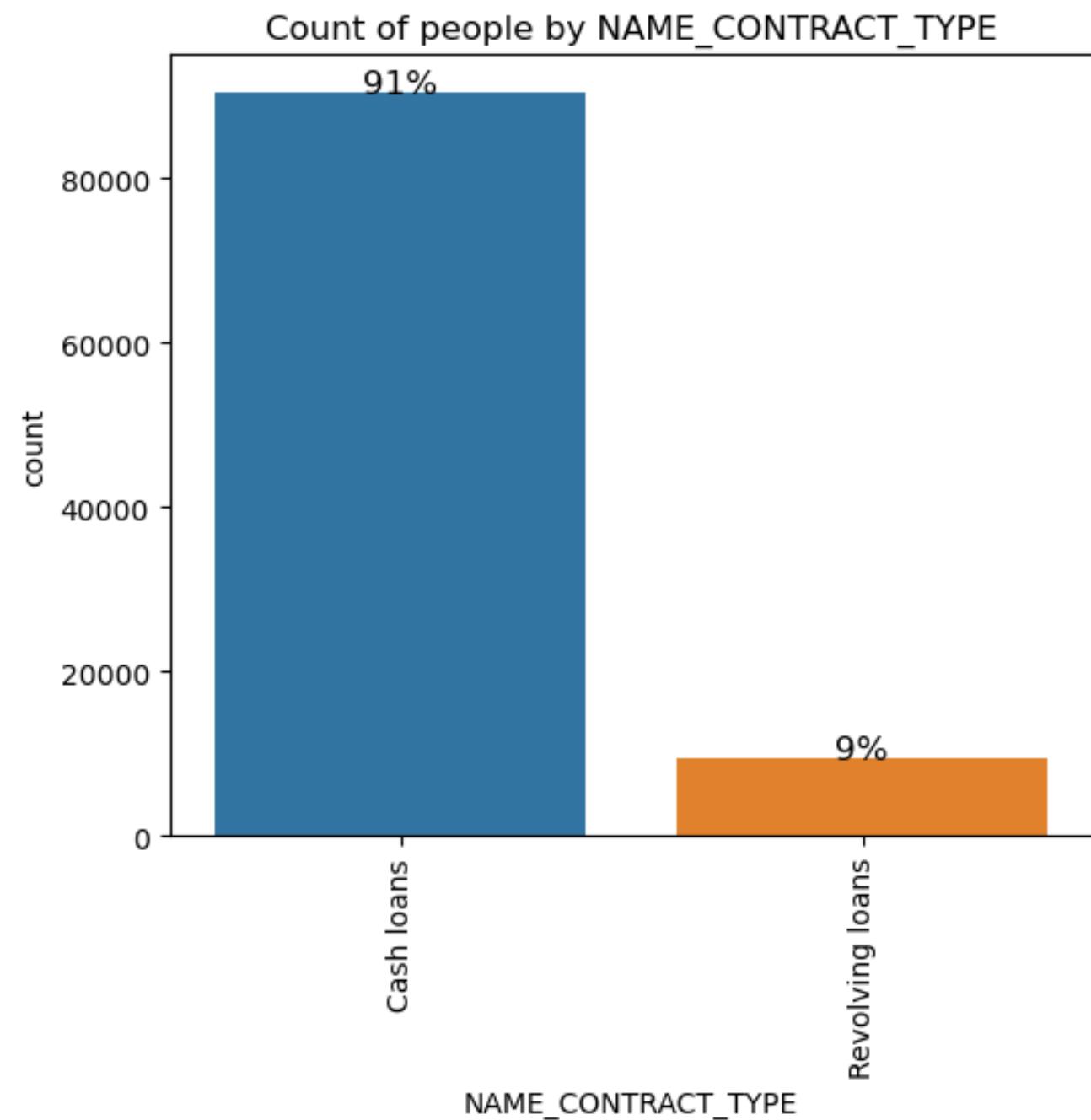
Defaulters : 9%
Payers : 92%

Percentage of customers: Payers or Defaulters

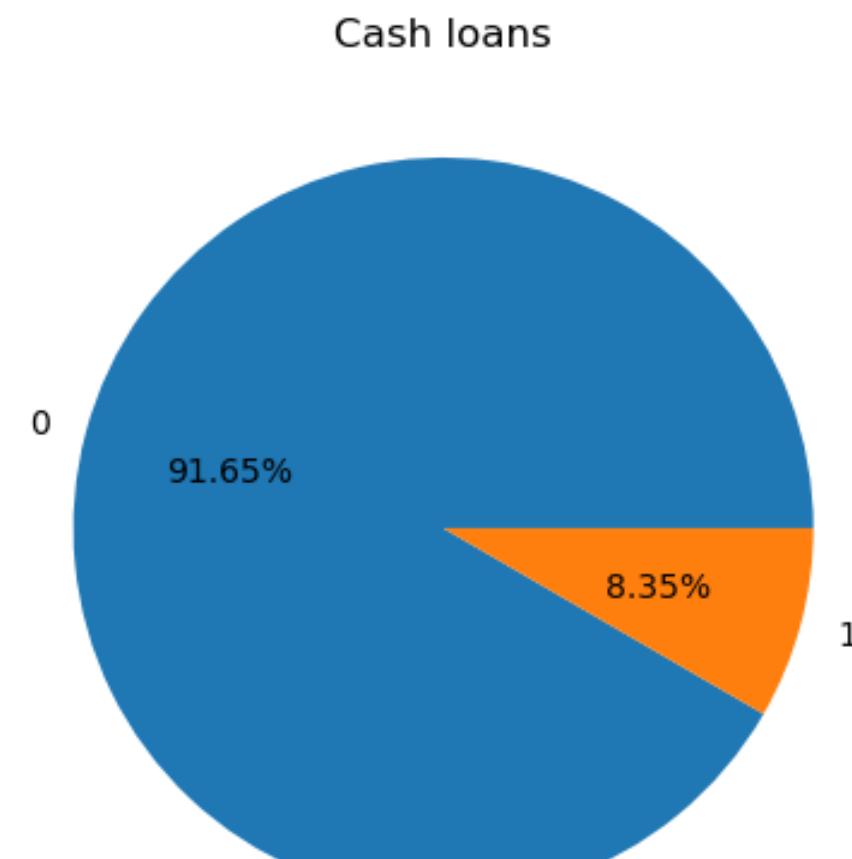


Lets analyse which factors are influencing the defaulters

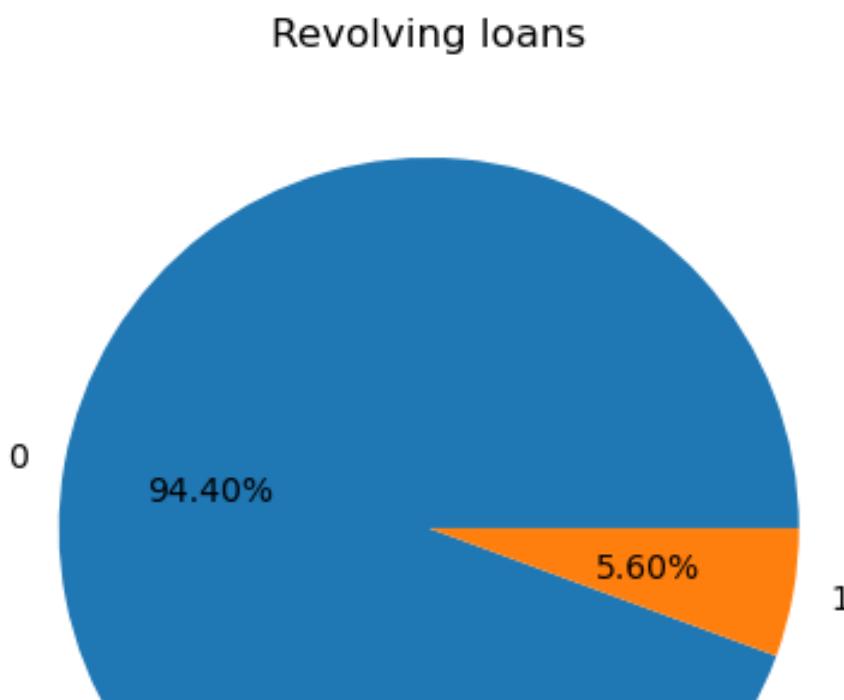
LOAN TYPE (NAME_CONTRACT_TYPE)



More Loans: Cash loans



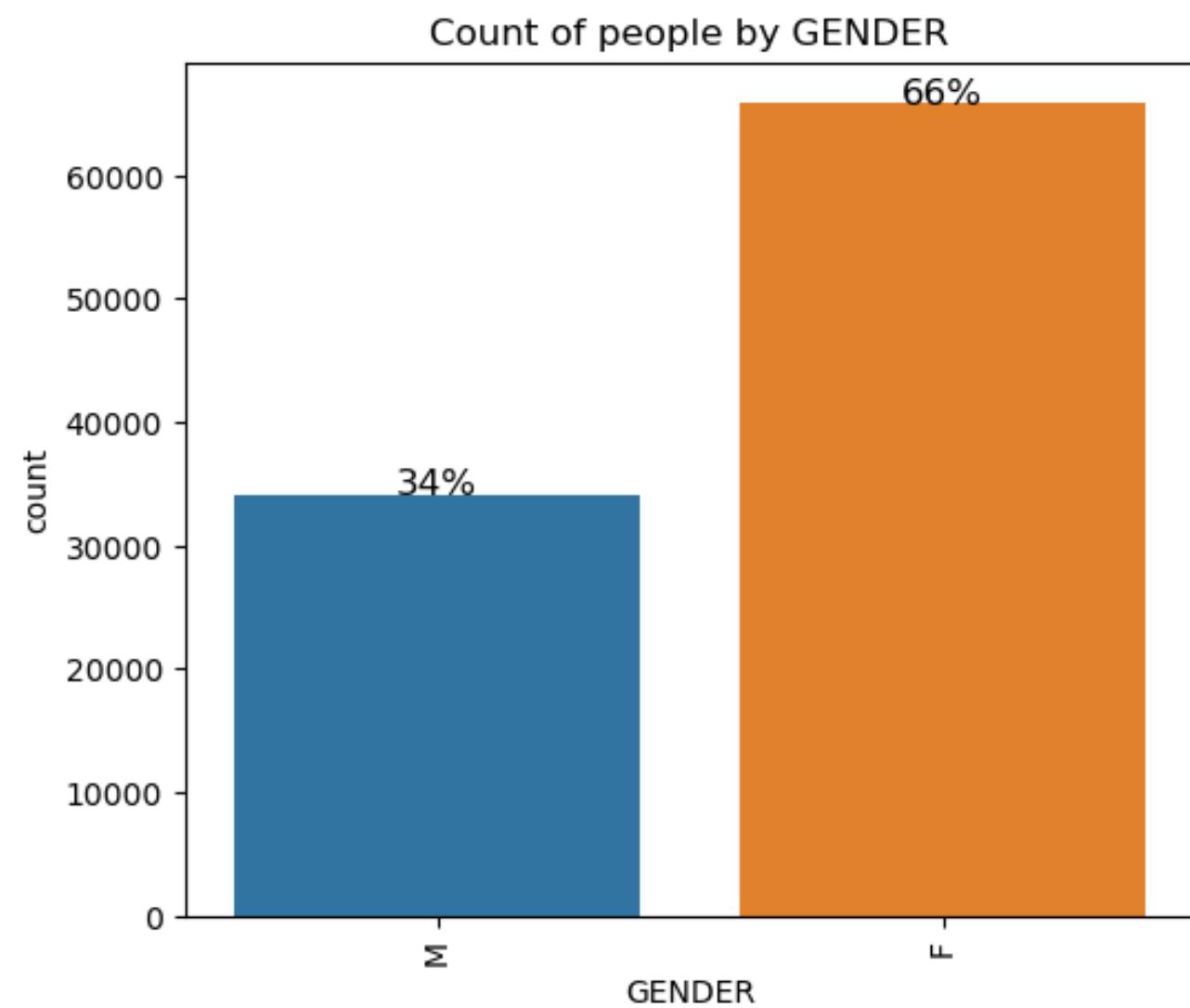
Low defaulters : Revolving loans



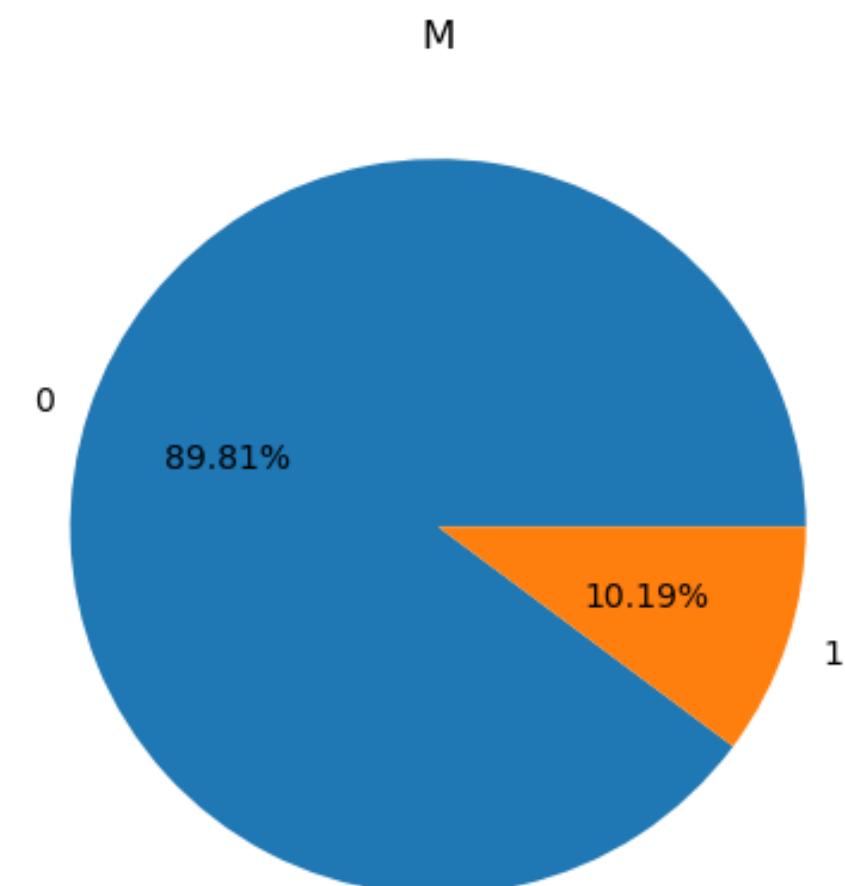
Lets analyse which factors are influencing the defaulters



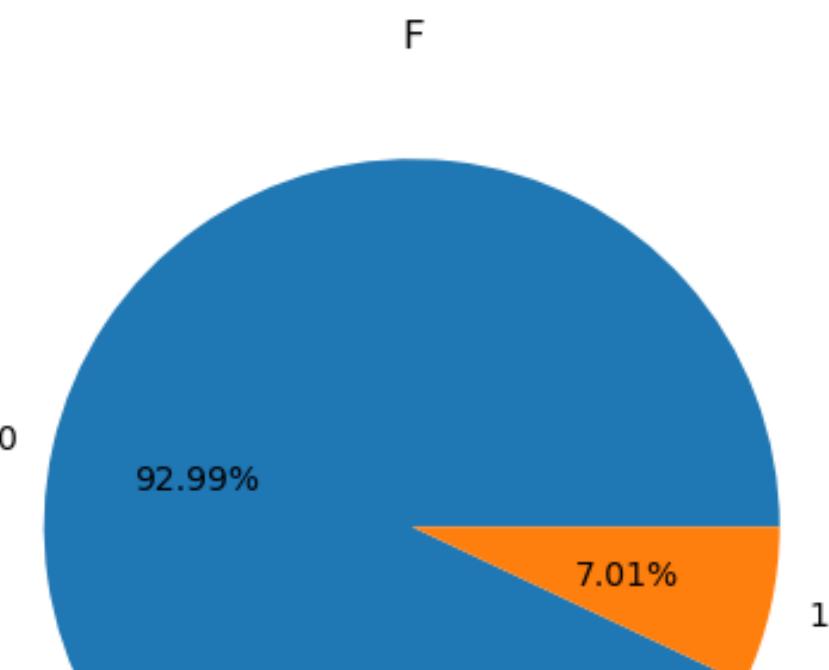
GENDER



More loans: Females

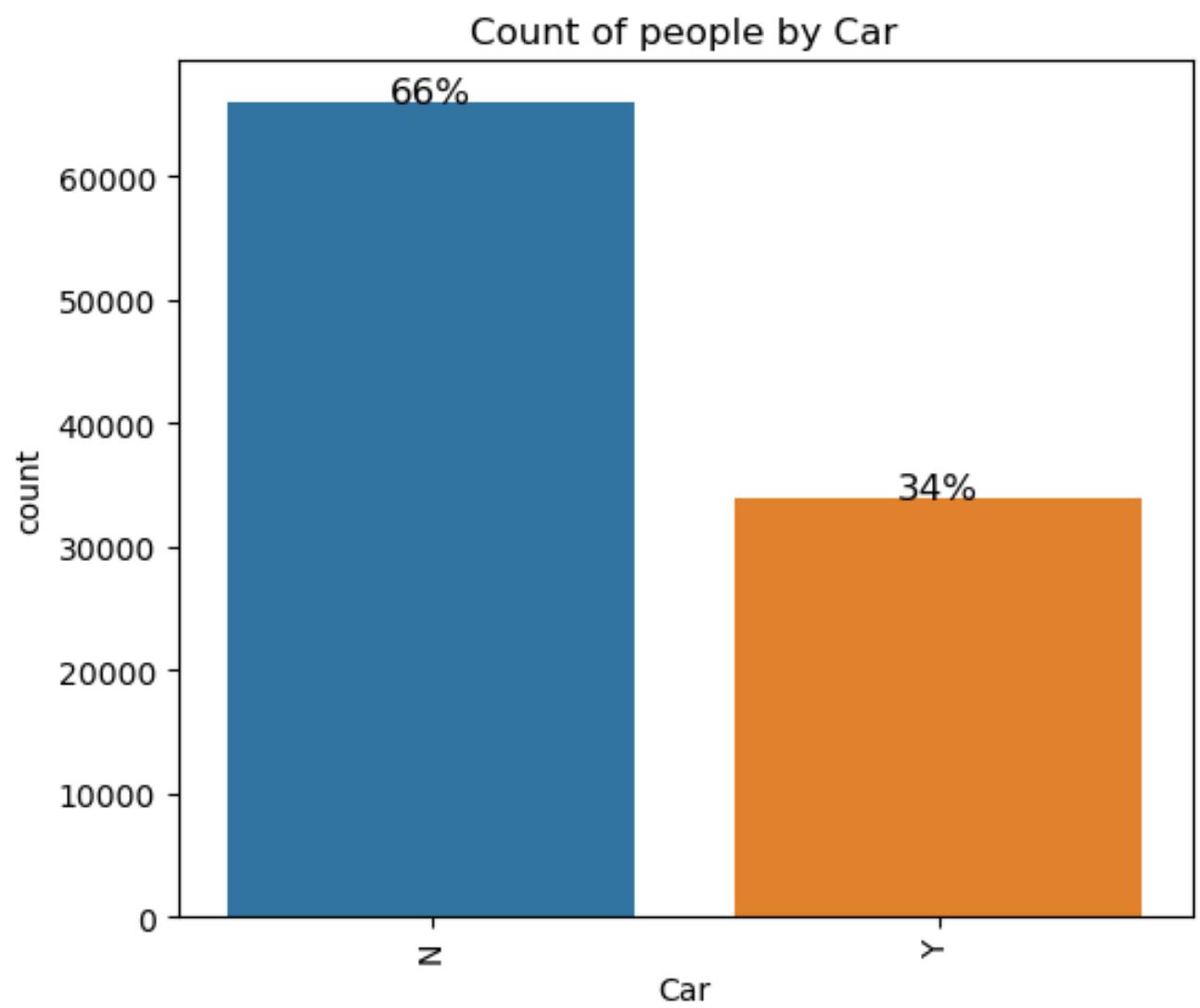


Low defaulters: Females

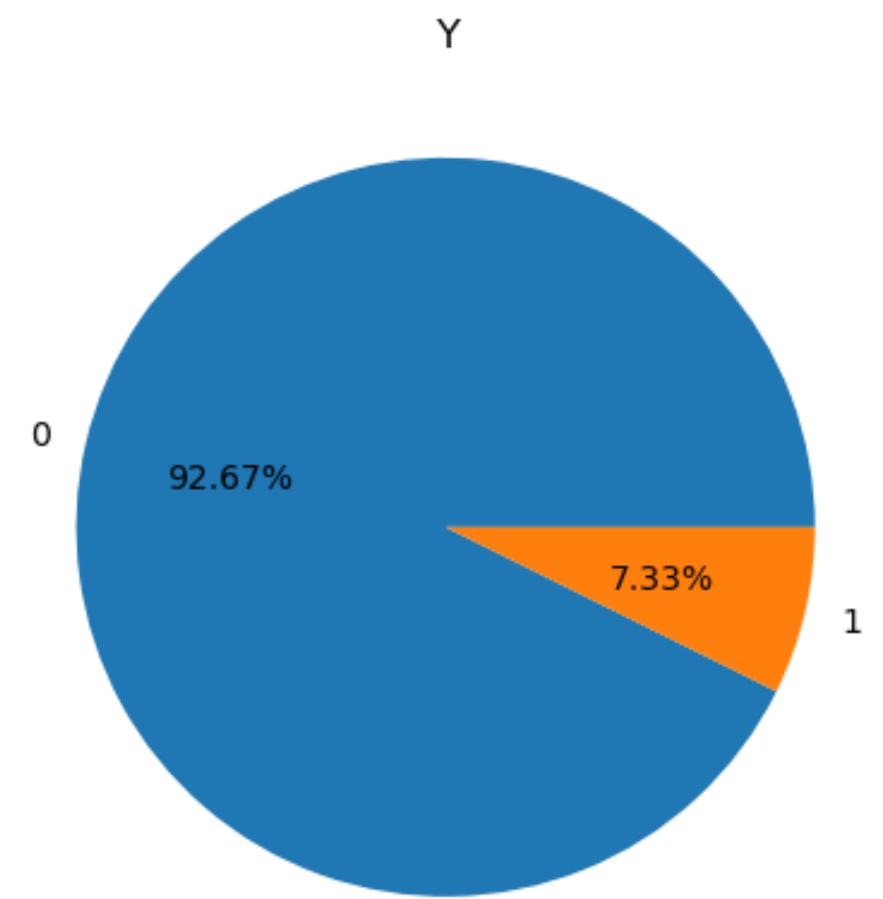


Lets analyse which factors are influencing the defaulters

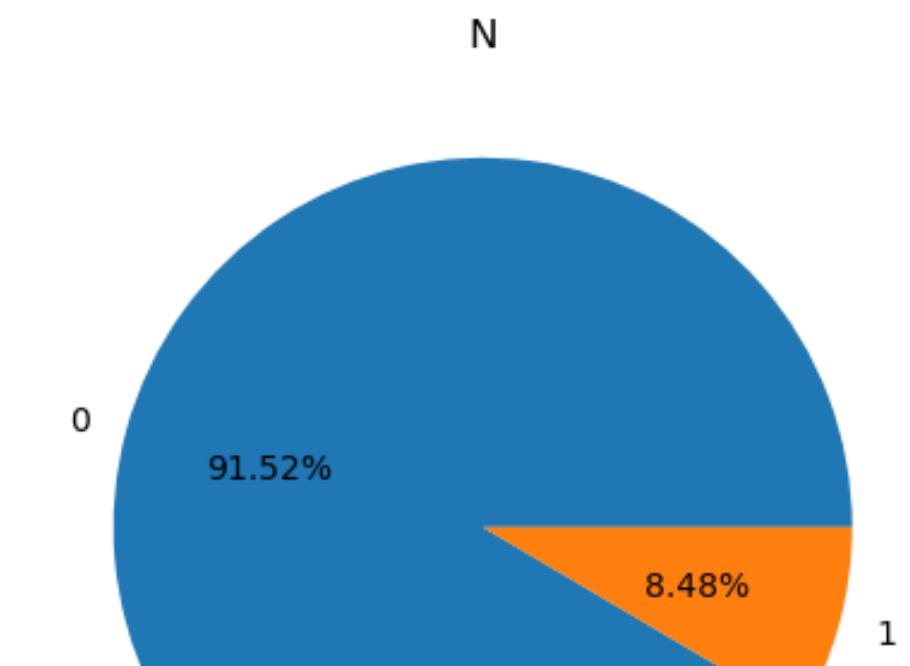
CAR OWNERSHIP



More loans: Do not have car

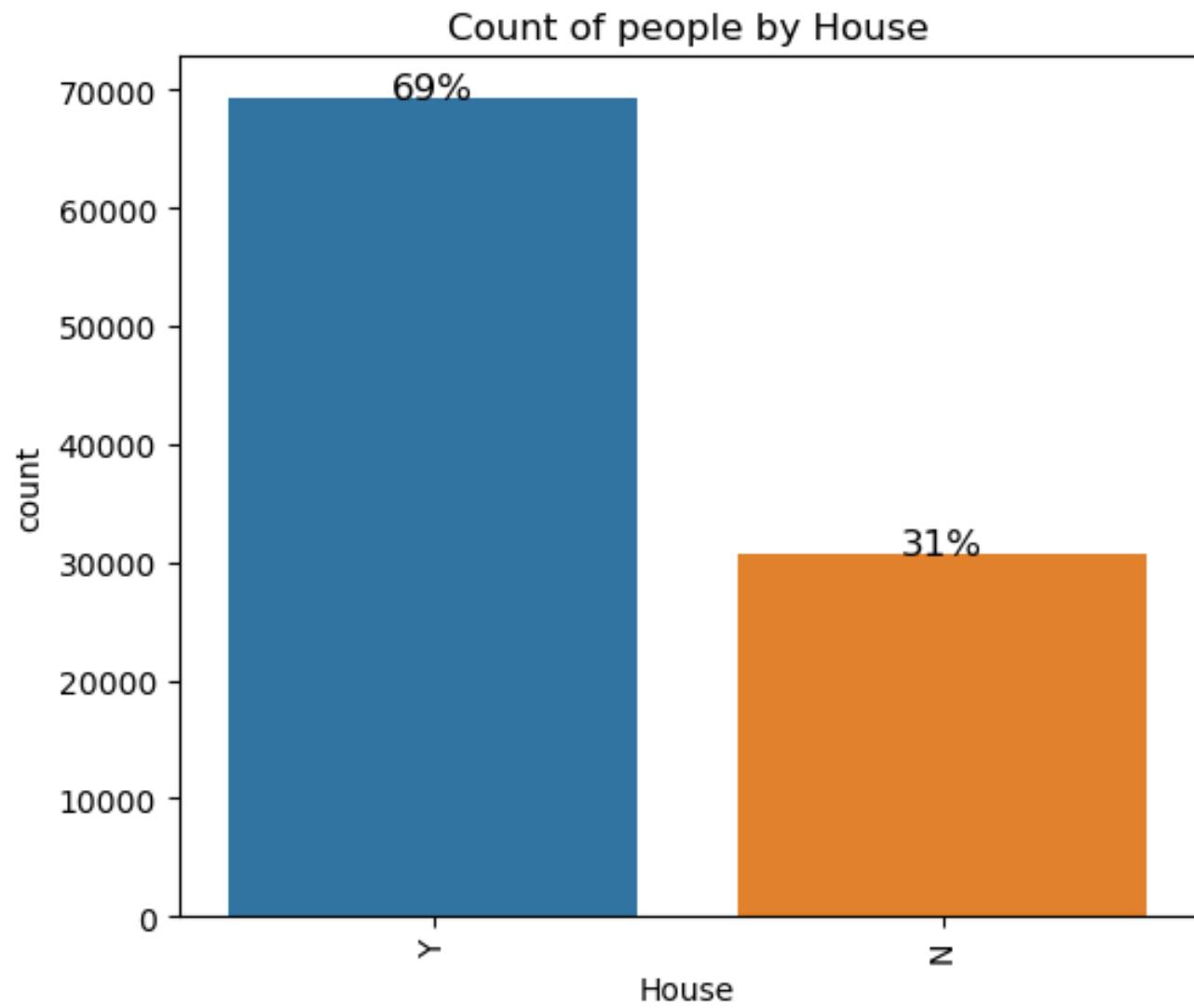


Same number of defaulters

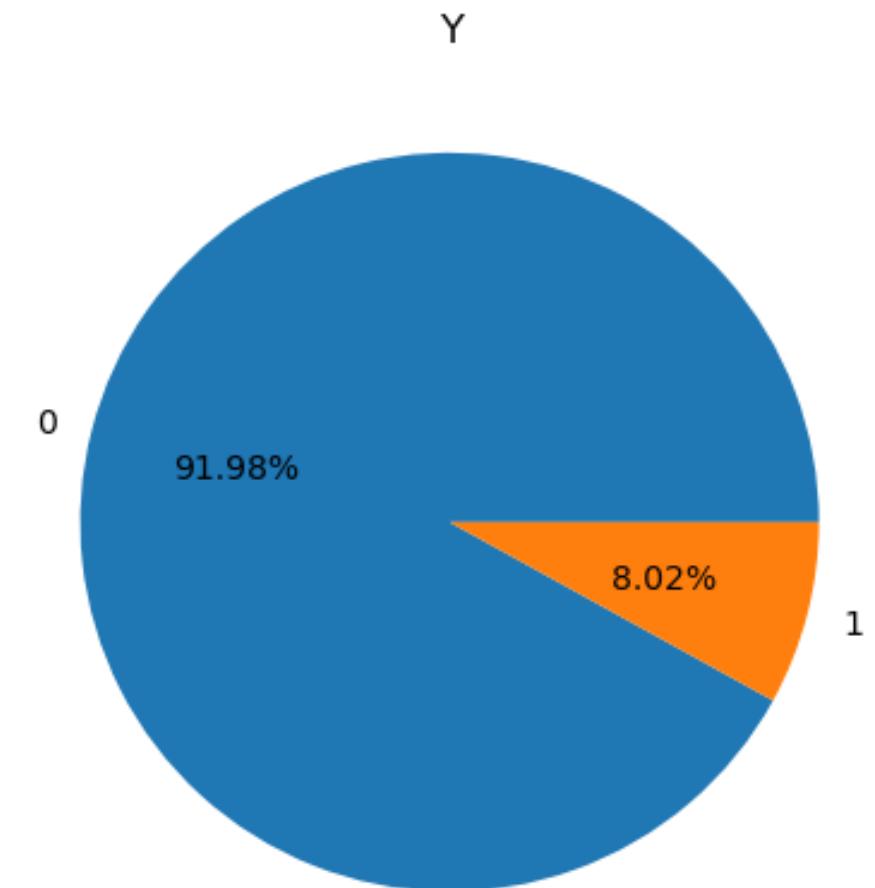


Lets analyse which factors are influencing the defaulters

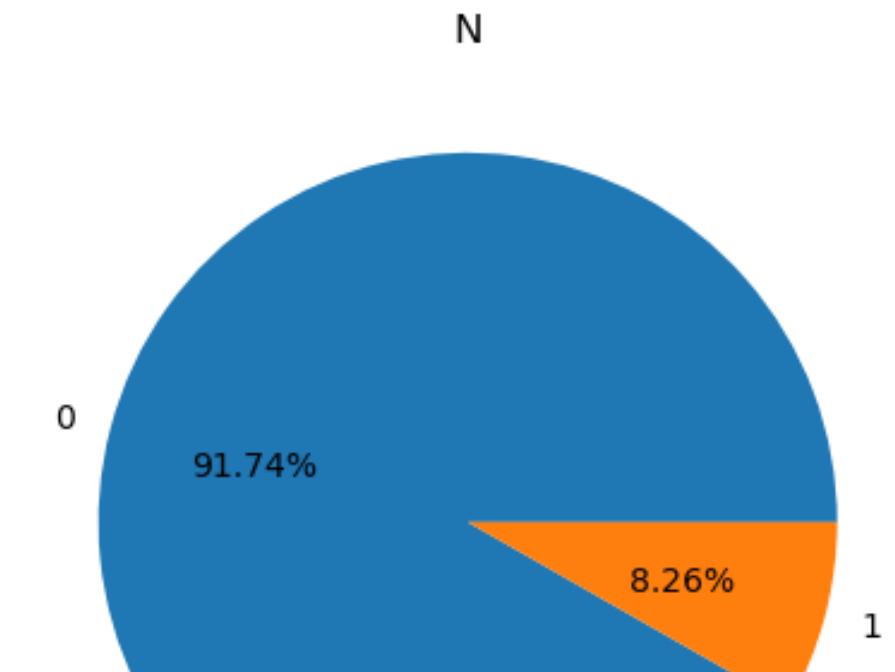
HOUSE OWNERSHIP



More loans: Have house

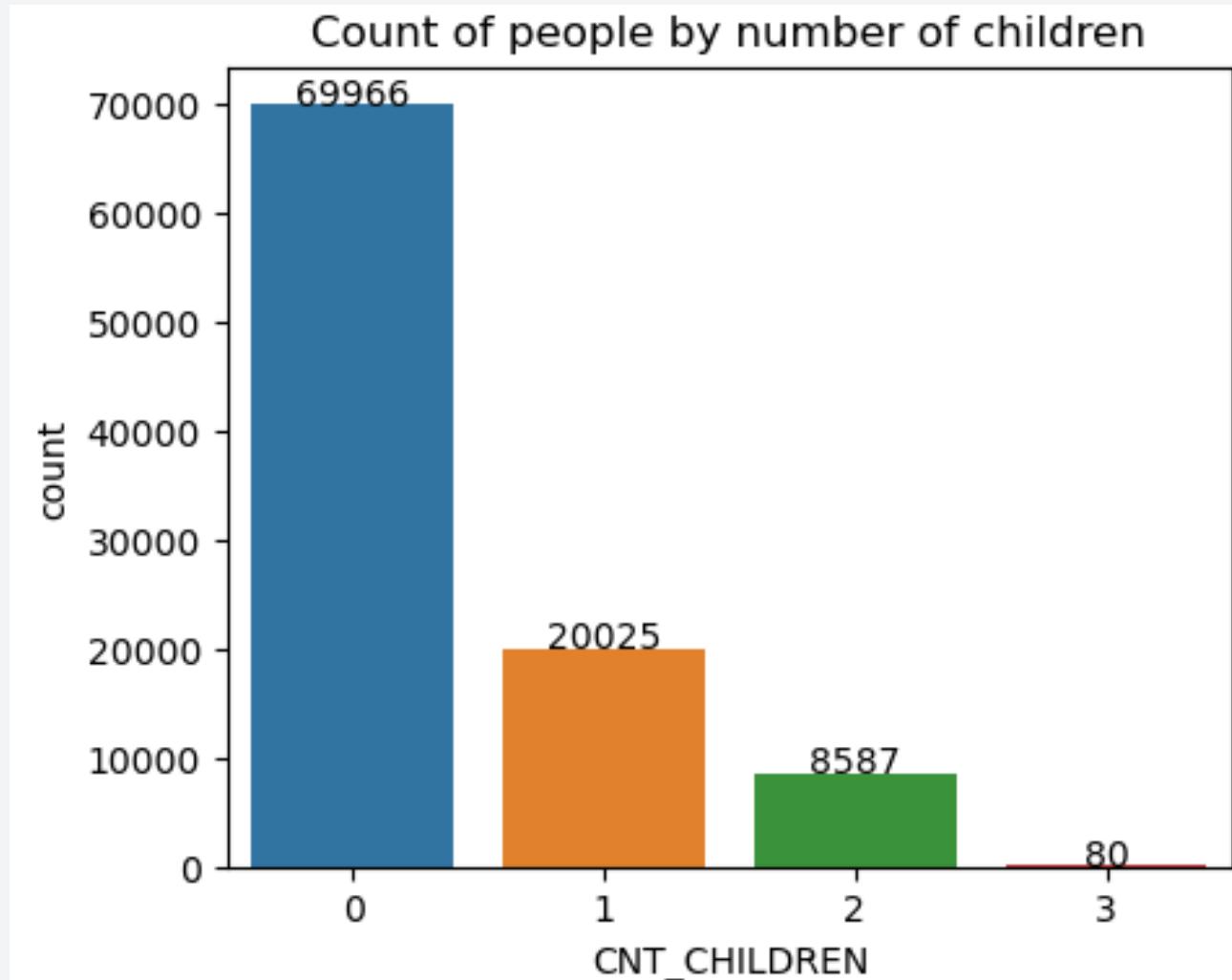


Same number of defaulters



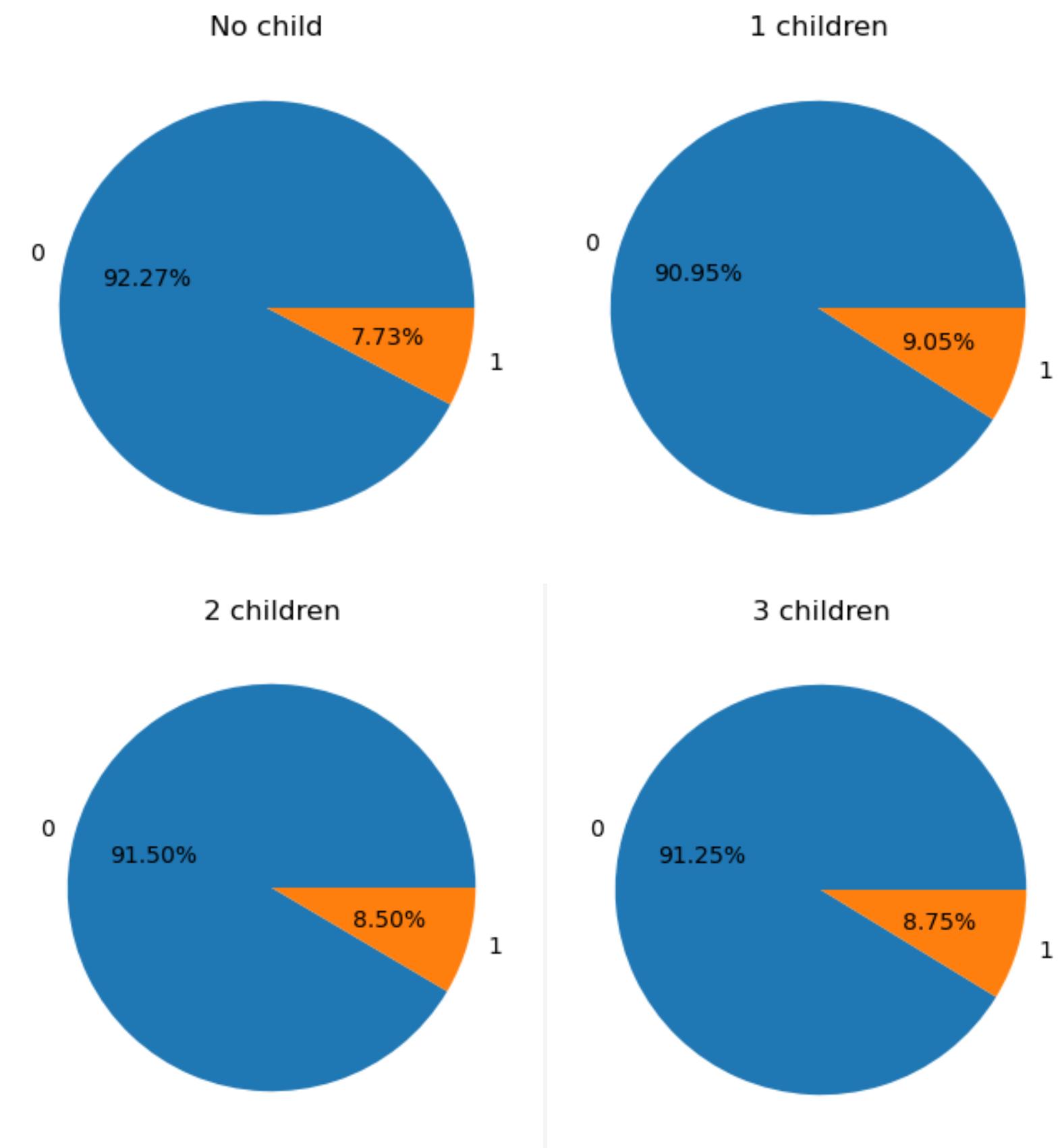
Lets analyse which factors are influencing the defaulters

NUMBER OF CHILDREN



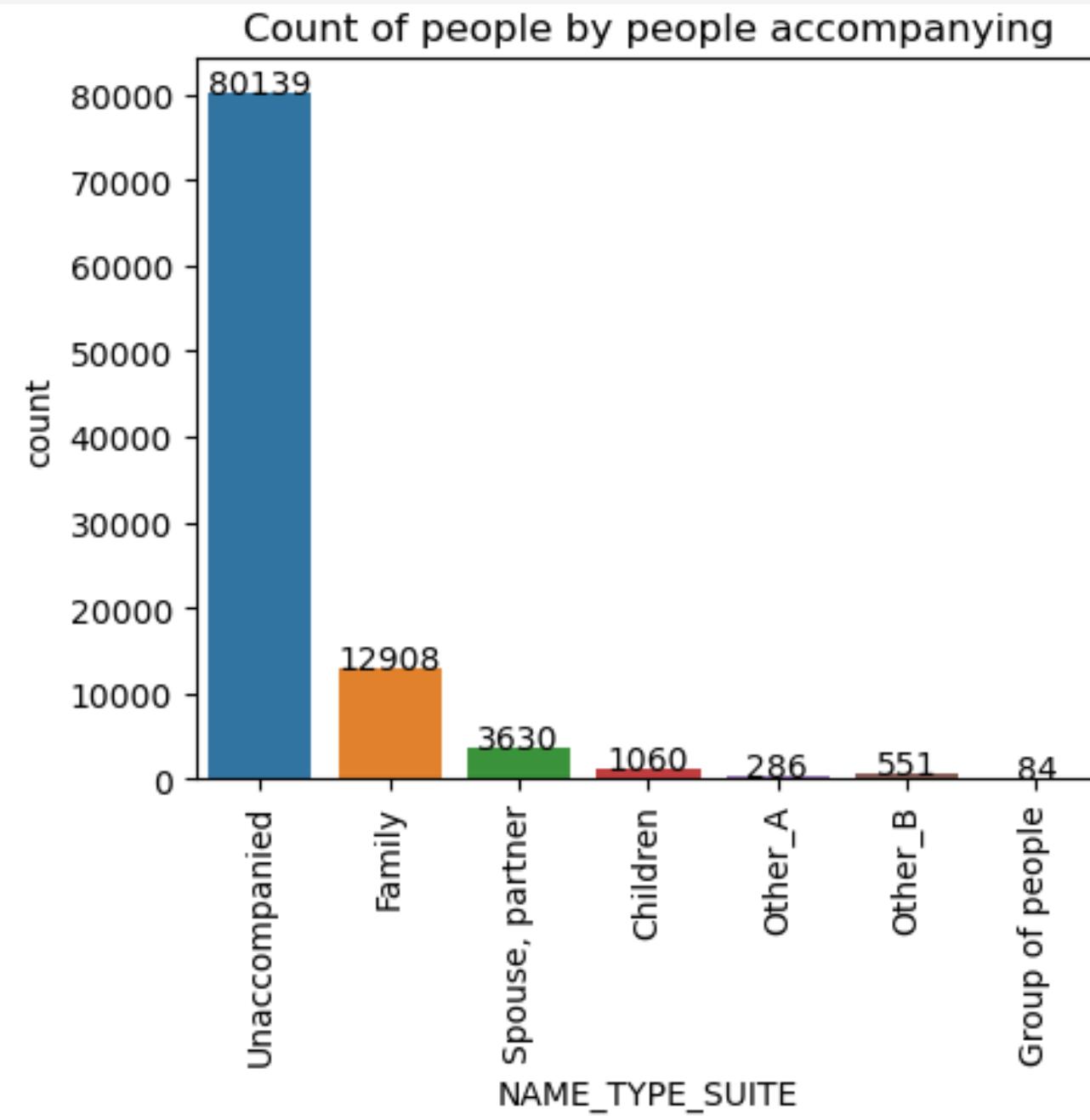
More loans: No child

Similar number of defaulters

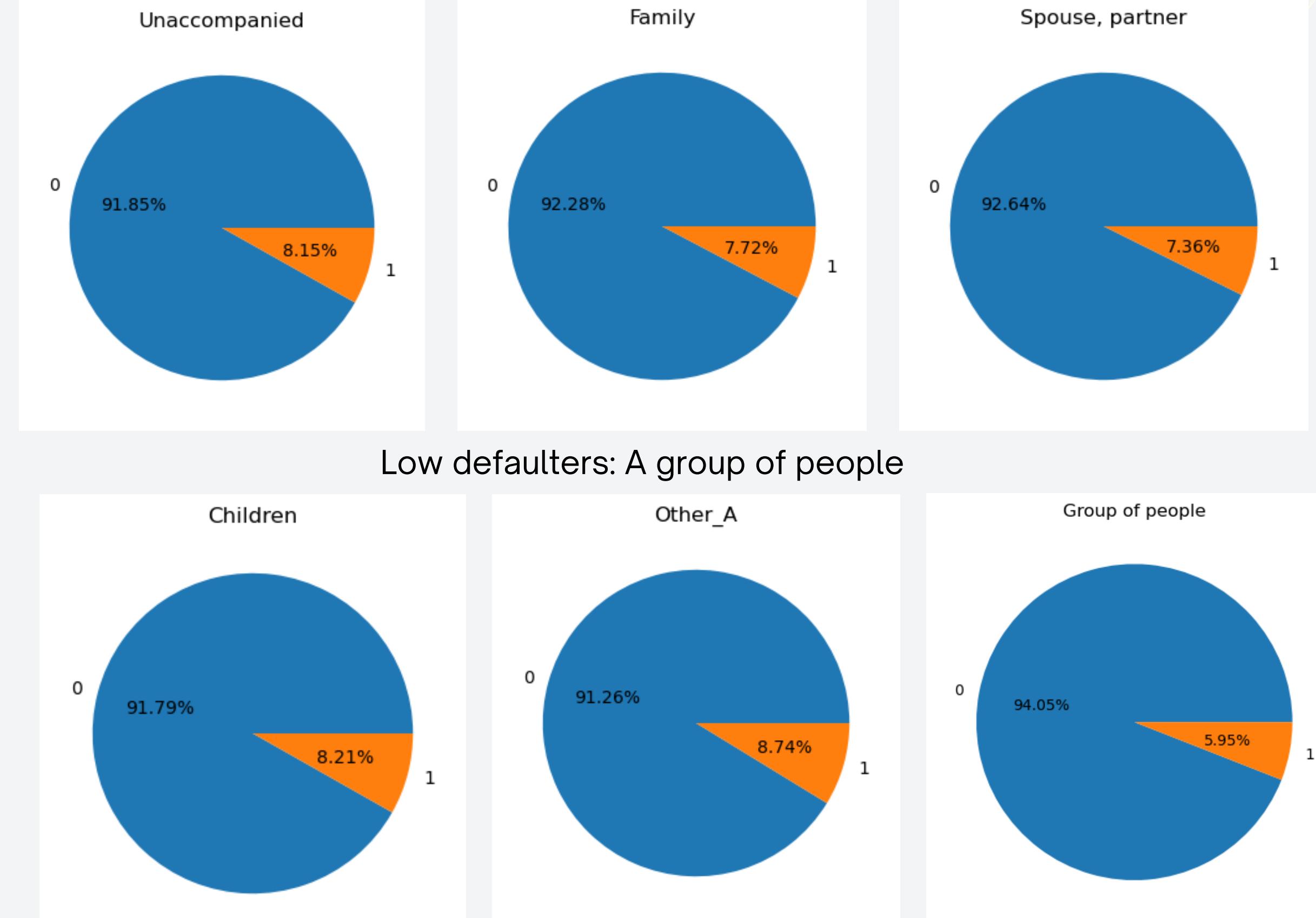


Lets analyse which factors are influencing the defaulters

PEOPLE ACCOMPANYING

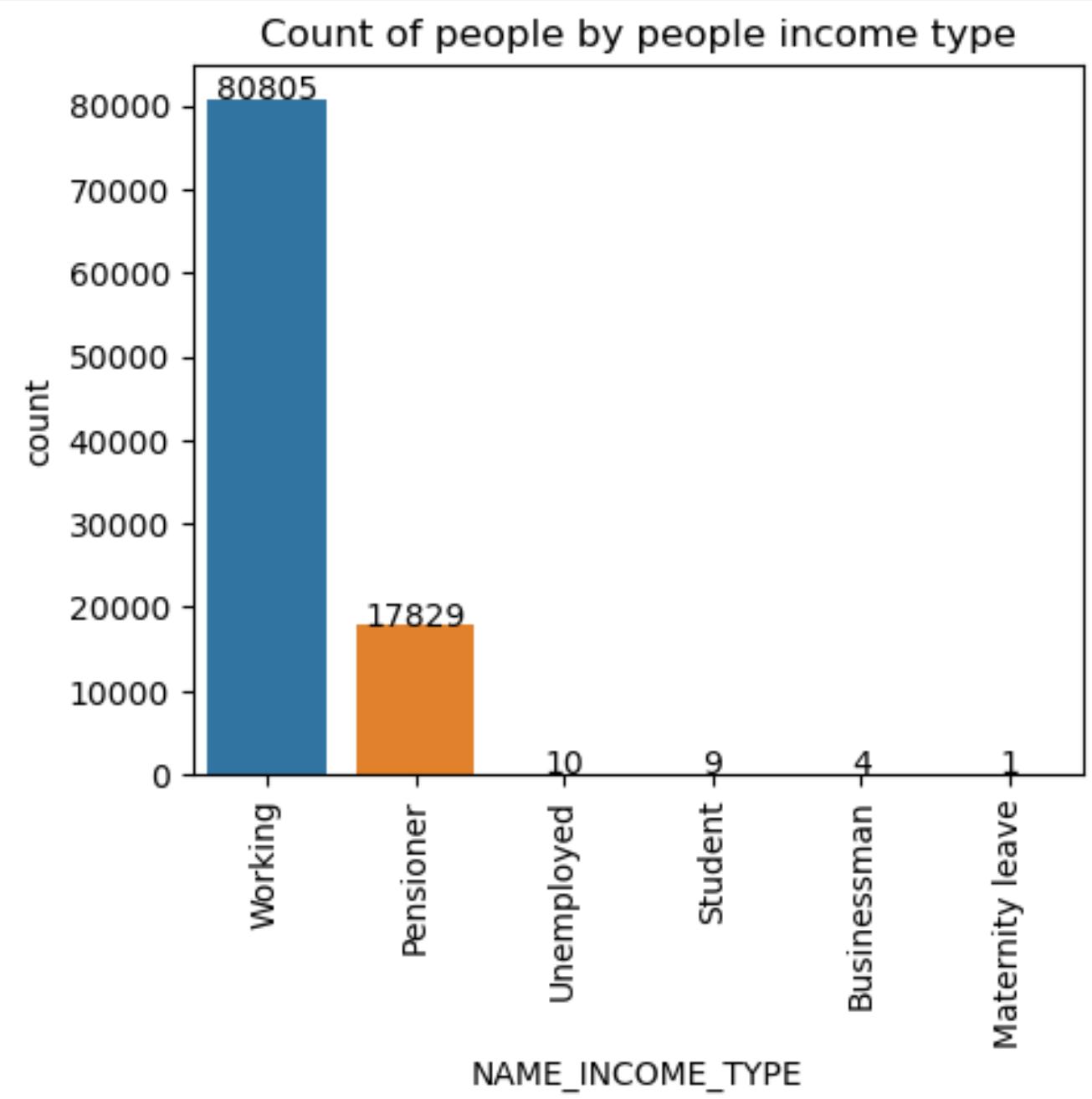


More loans: Unaccompanied

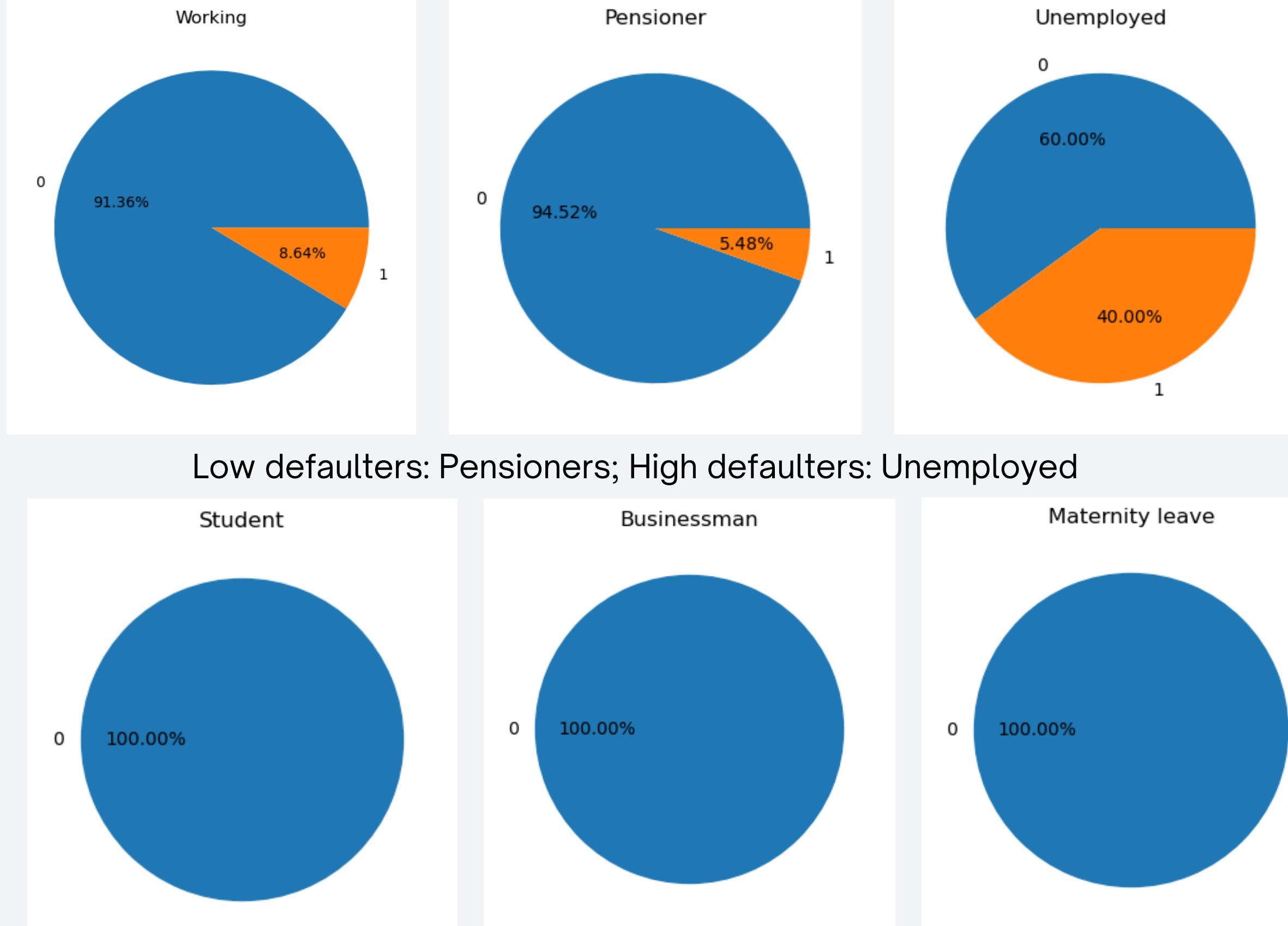


Lets analyse which factors are influencing the defaulters

PEOPLE INCOME TYPE

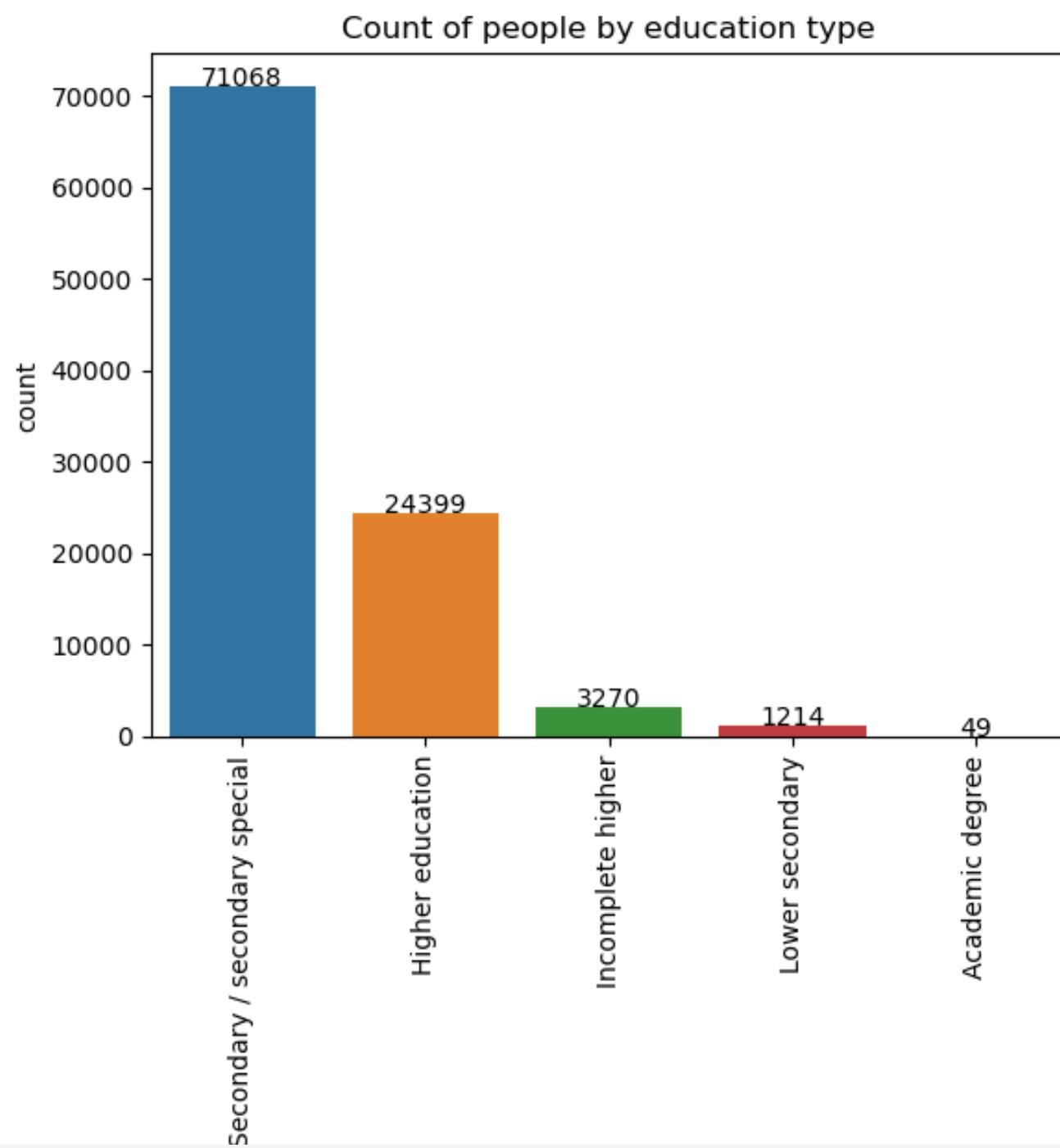


More loans: Working

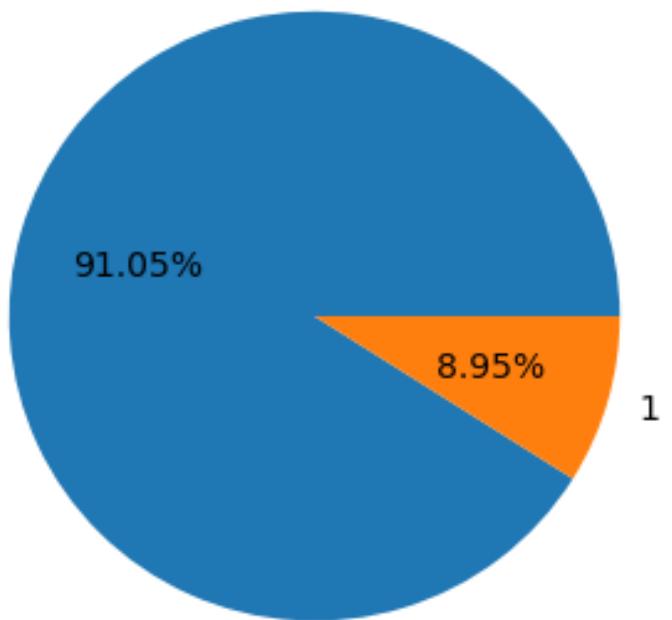


Lets analyse which factors are influencing the defaulters

PEOPLE EDUCATION TYPE

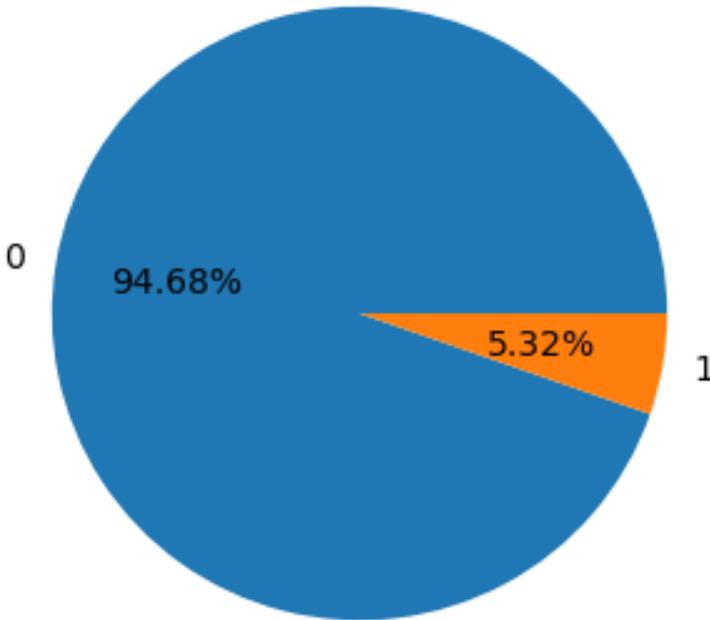


Secondary / secondary special

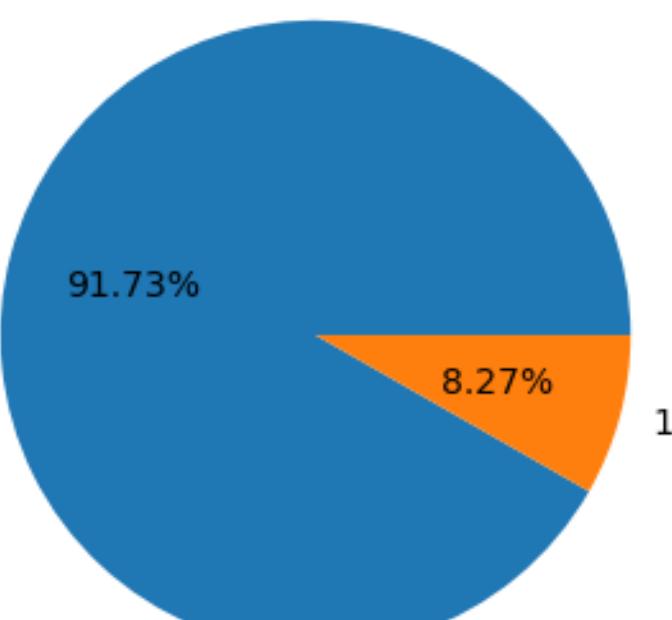


Low defaulters: Academic degree, Higher education

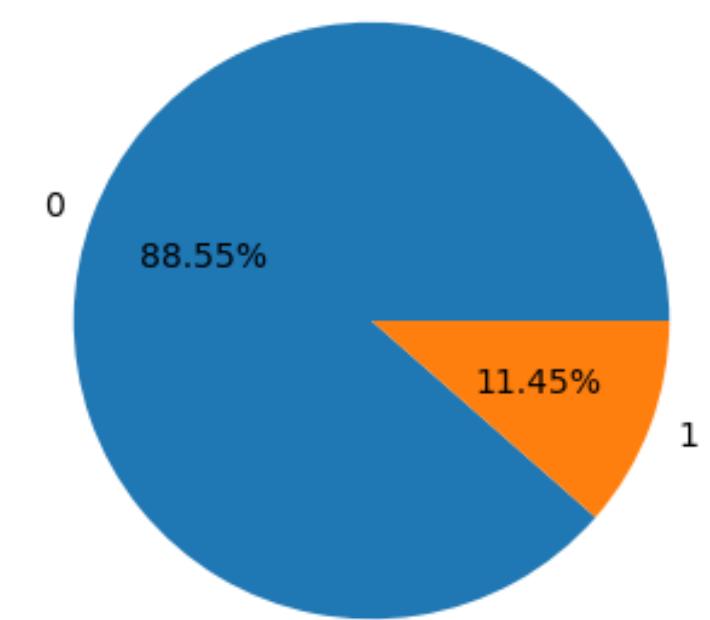
Higher education



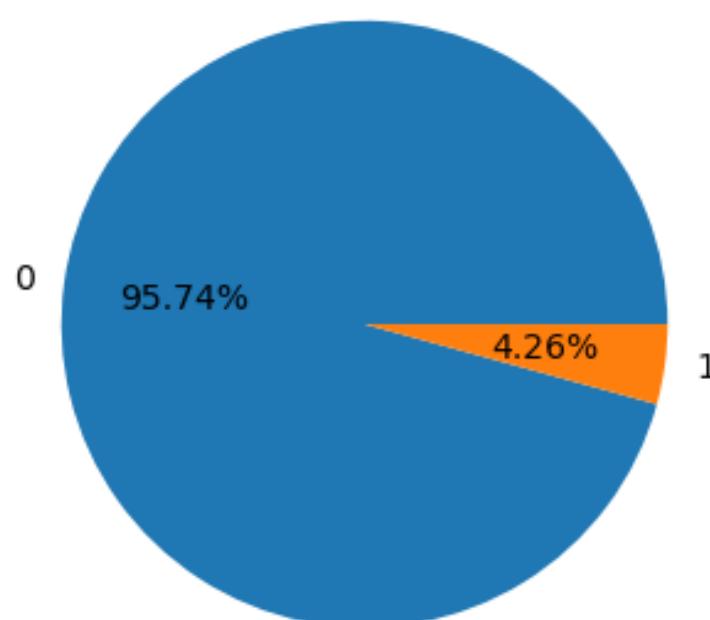
Incomplete higher



Lower secondary

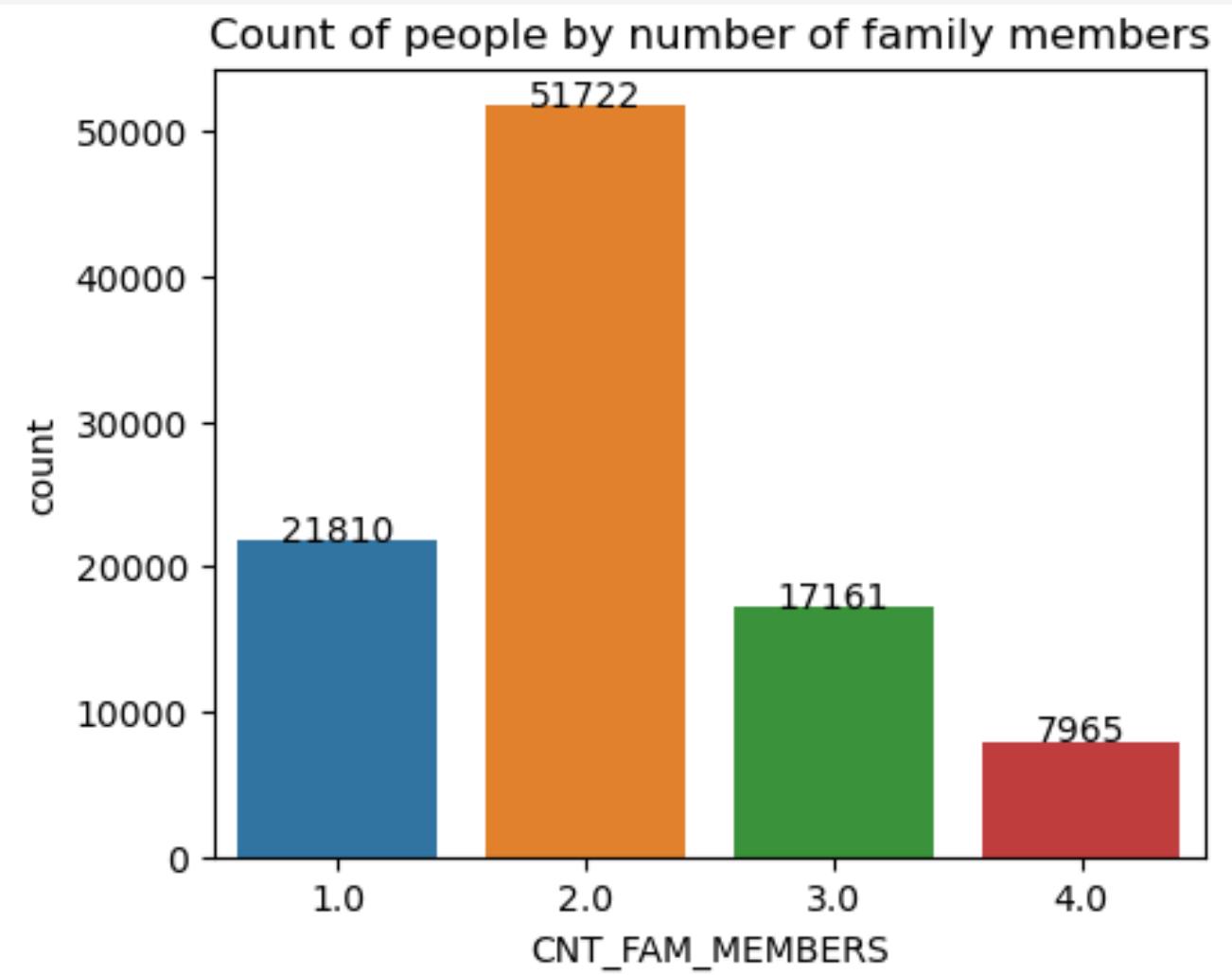


Acedemic degree



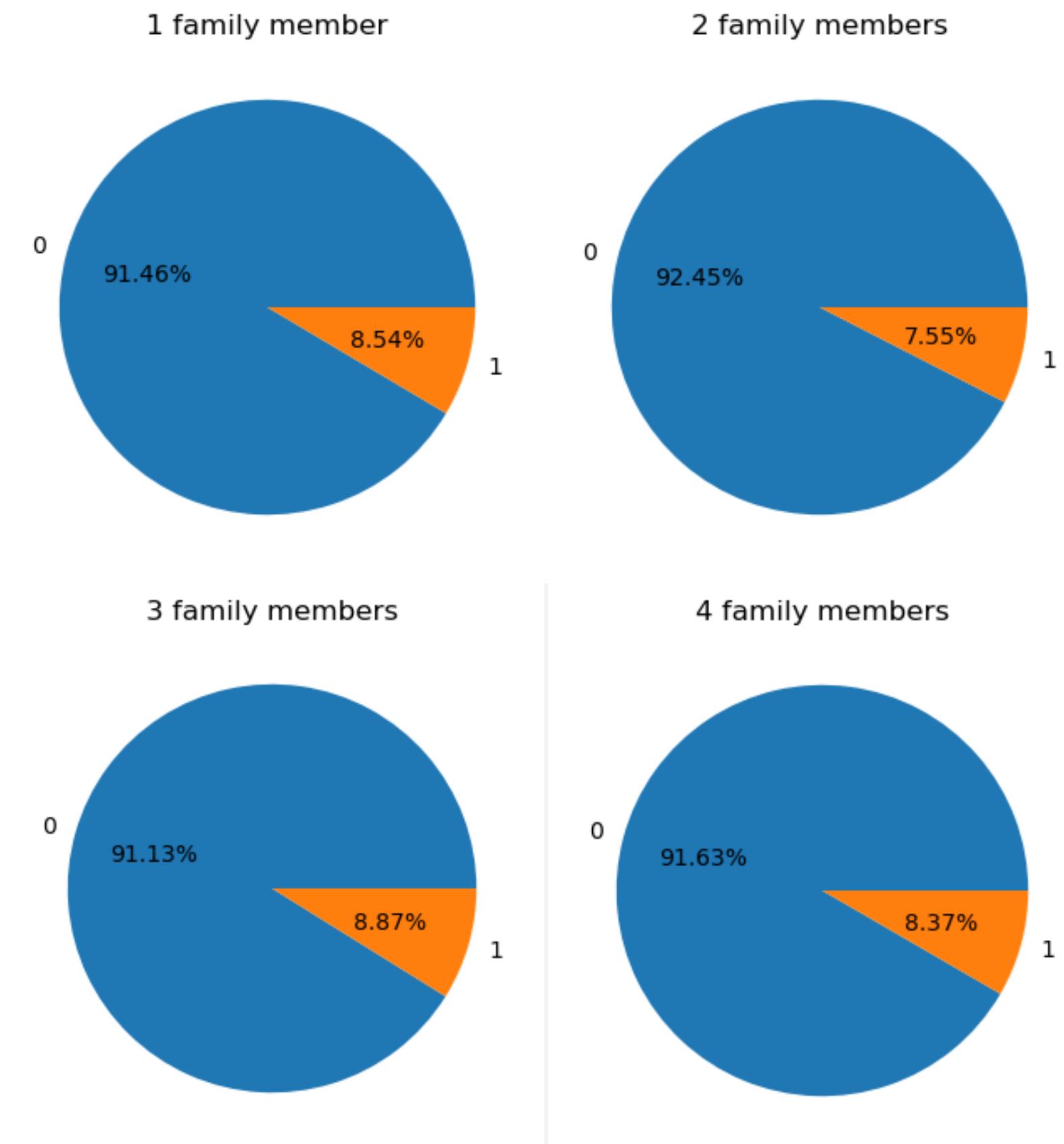
Lets analyse which factors are influencing the defaulters

NUMBER OF FAMILY MEMBERS



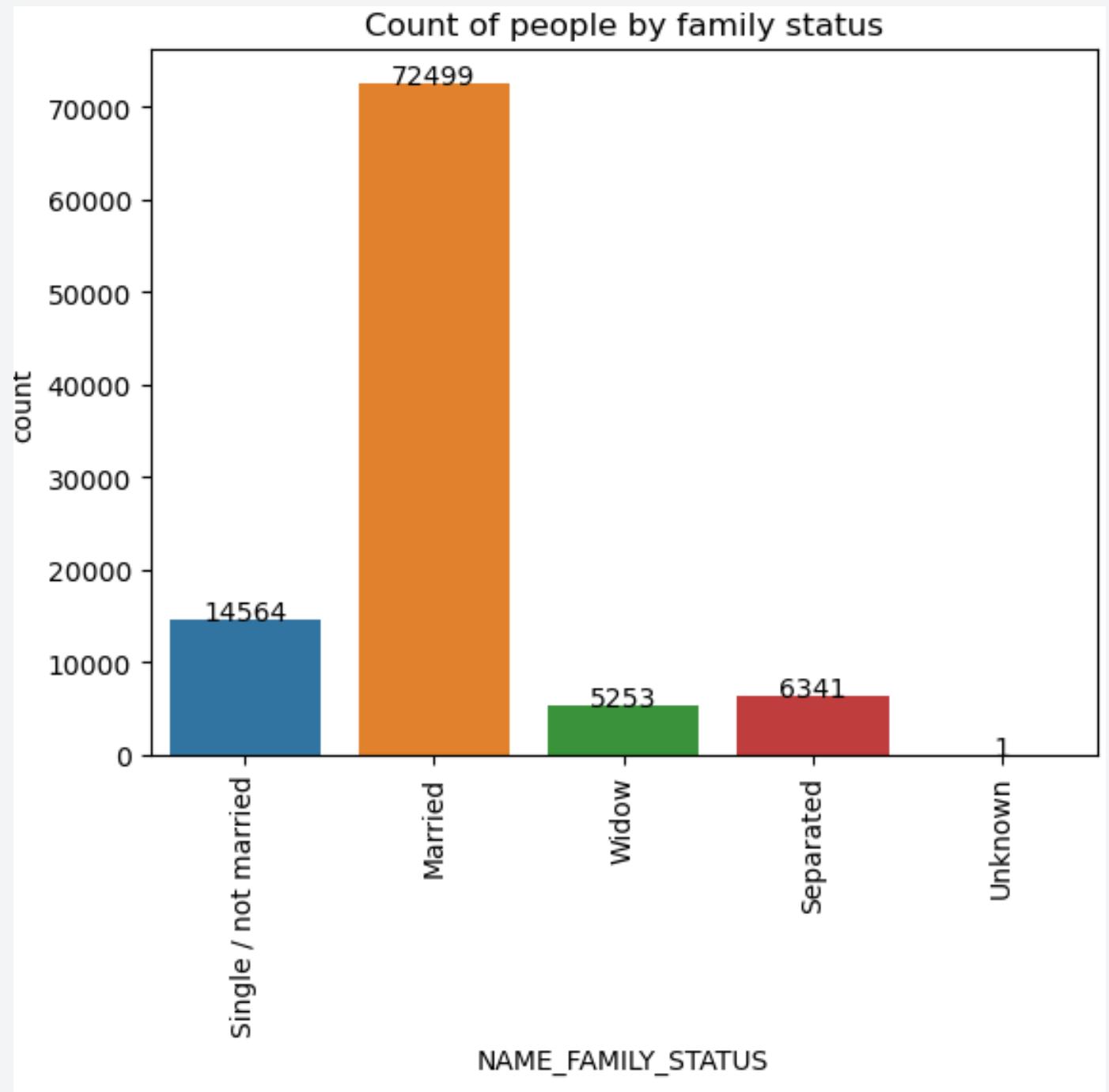
More loans: 2 family members

Similar number of
defaulters



Lets analyse which factors are influencing the defaulters

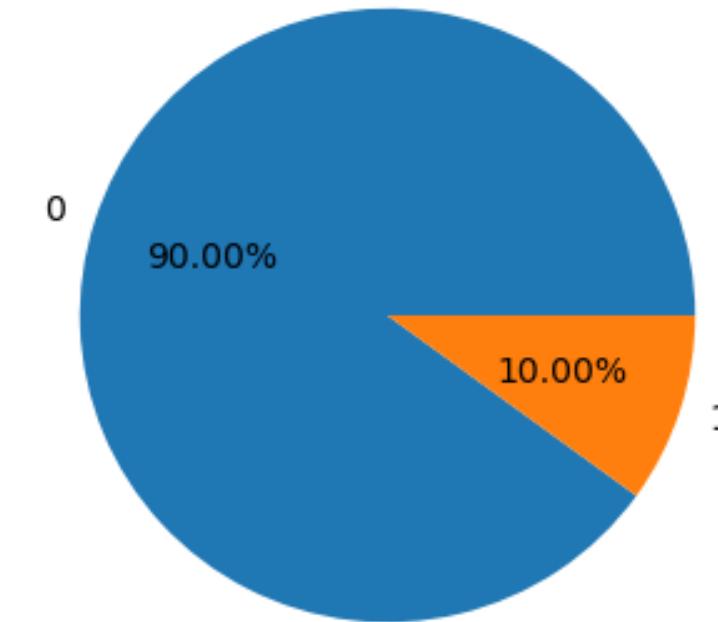
NUMBER OF FAMILY MEMBERS



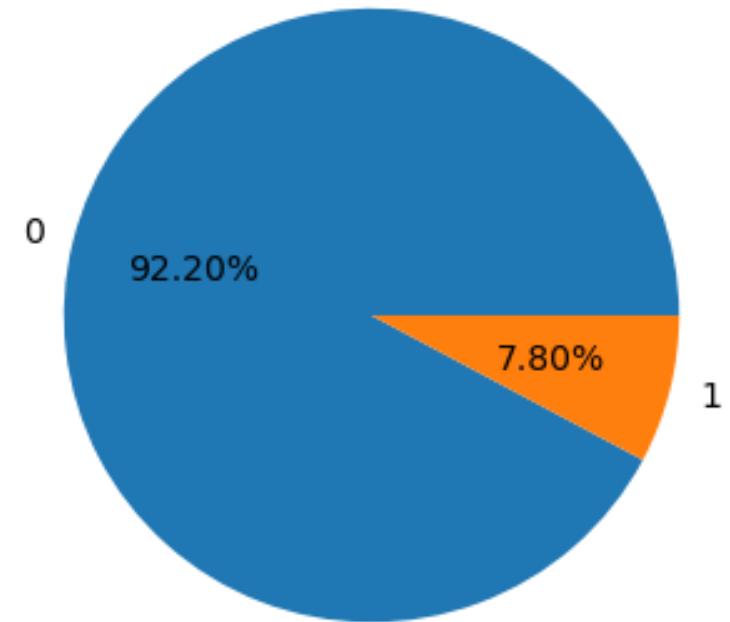
More loans: Married

Low defaulters:
Widows

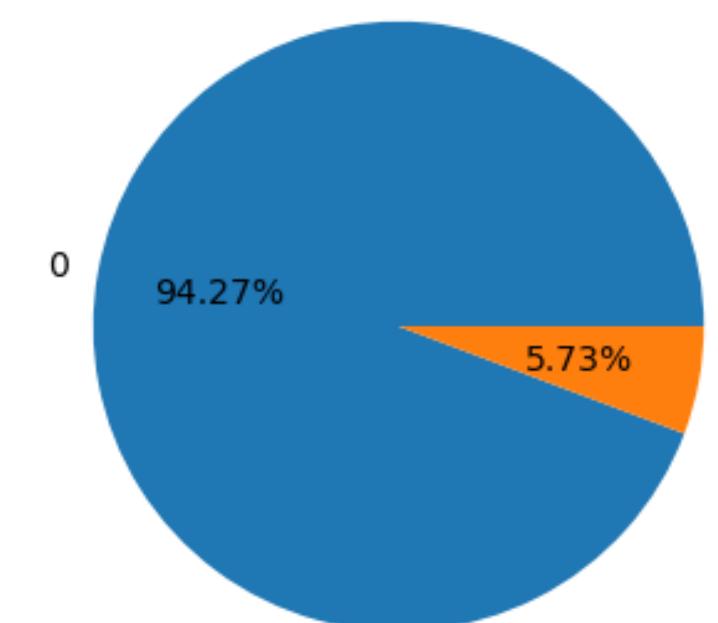
Single / not married



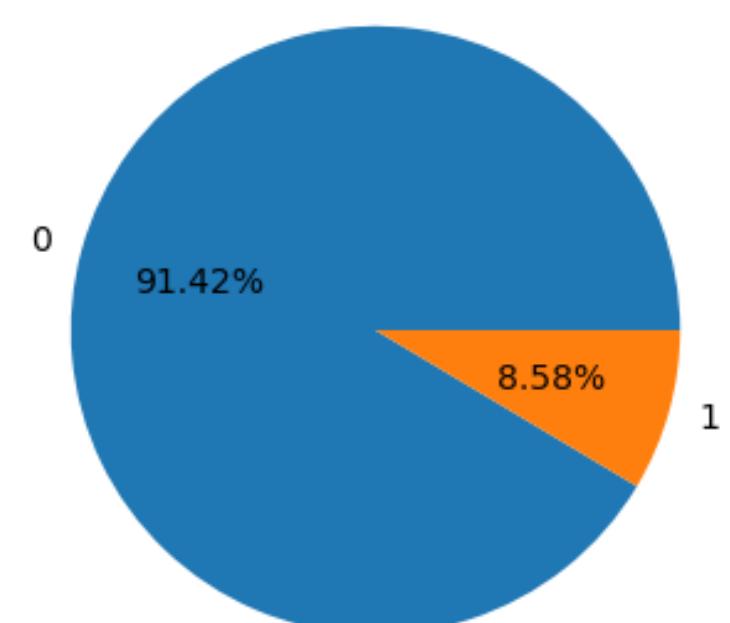
Married



Widow

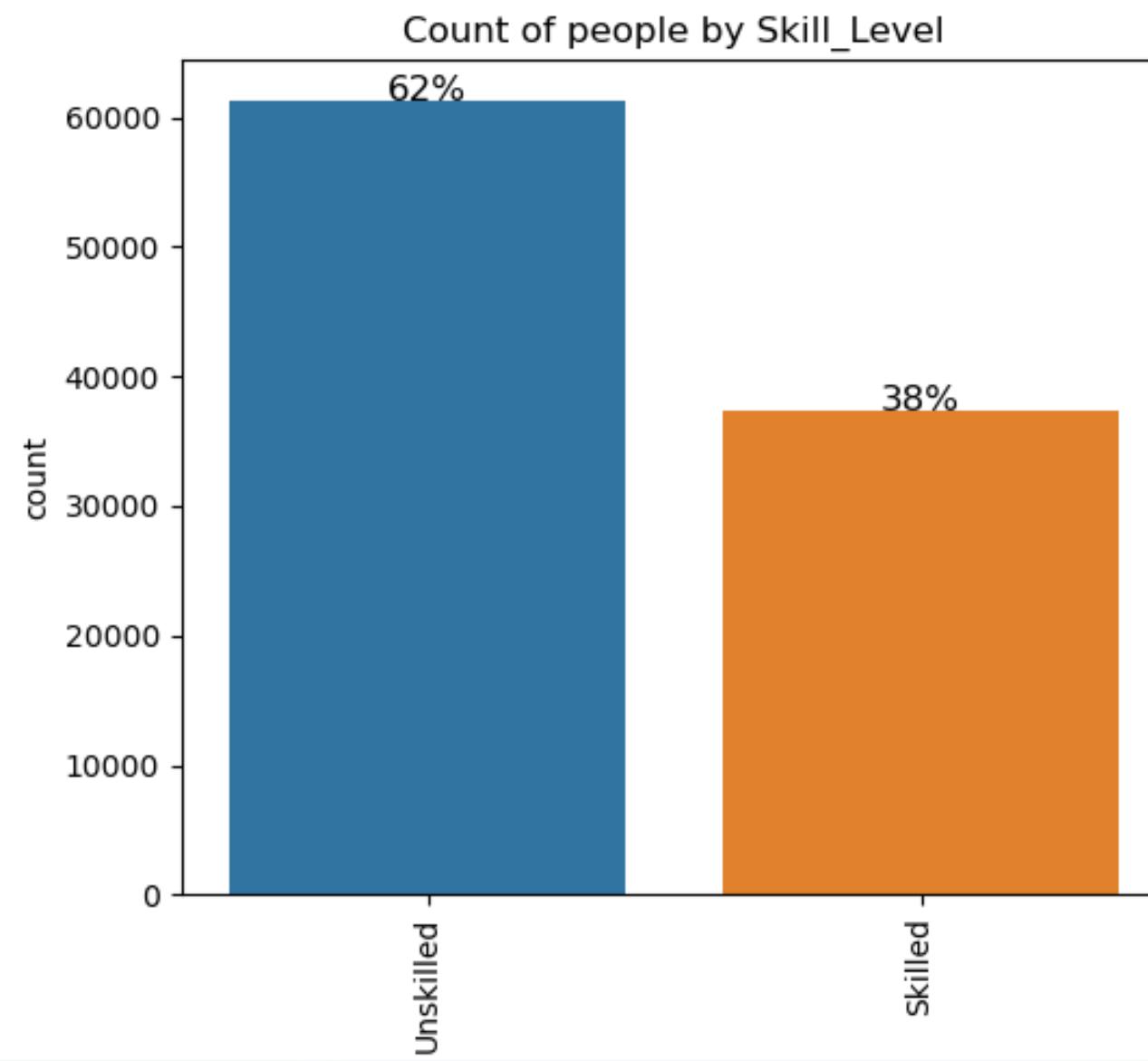


Separated

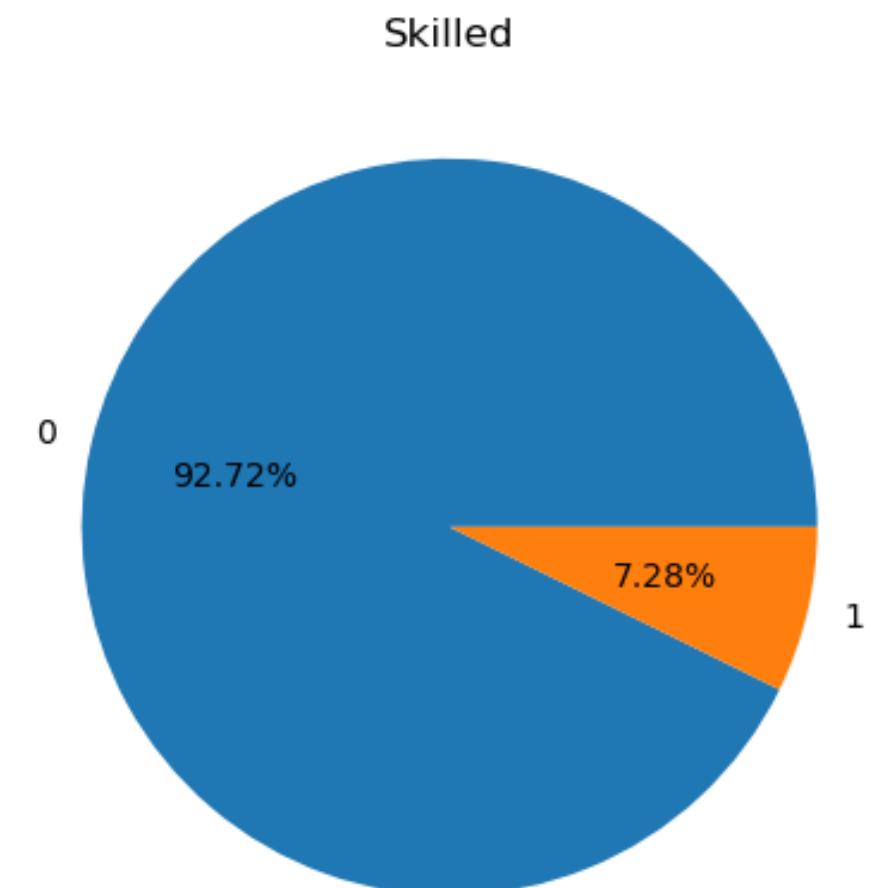


Lets analyse which factors are influencing the defaulters

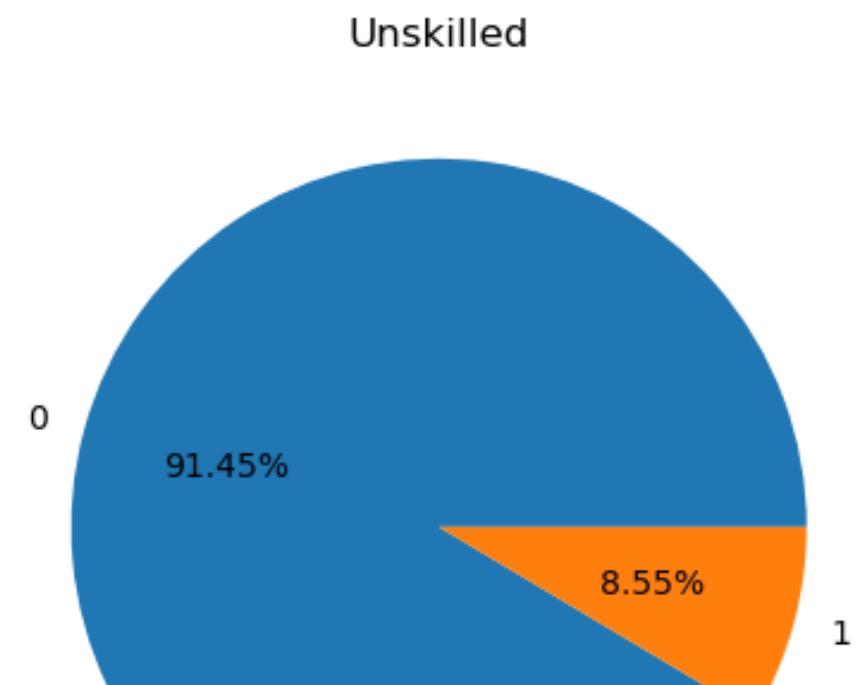
SKILL LEVEL



More loans: Unskilled

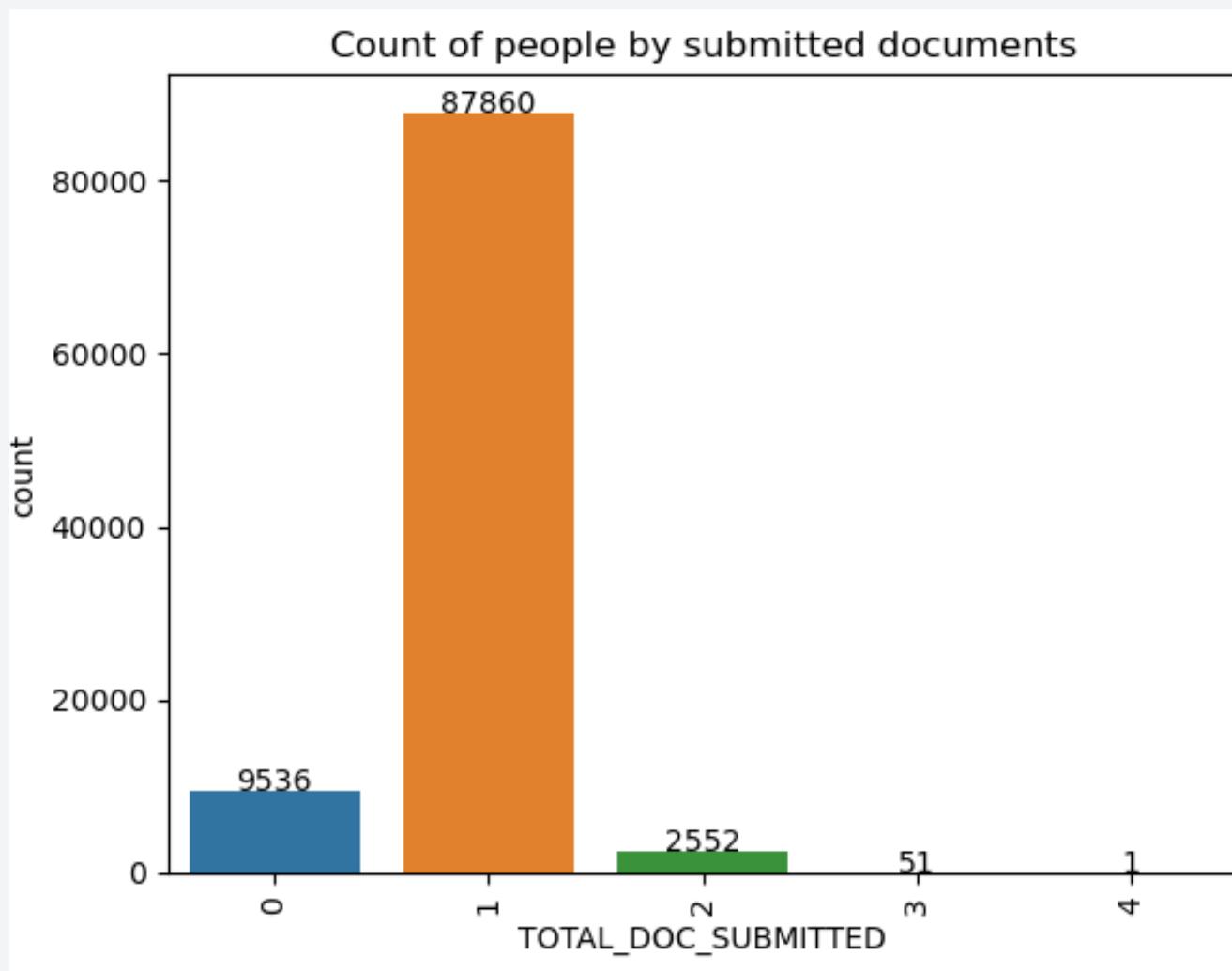


Similar number of defaulters



Lets analyse which factors are influencing the defaulters

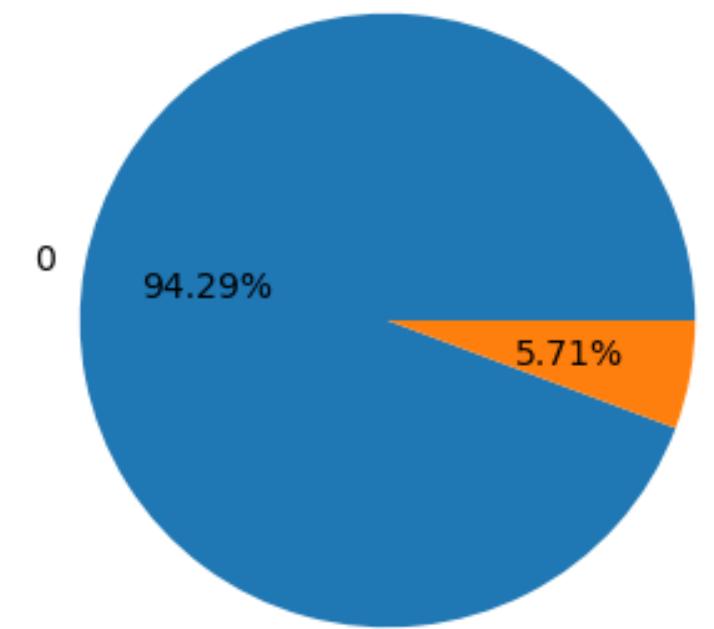
NUMBER OF DOCUMENTS SUBMITTED



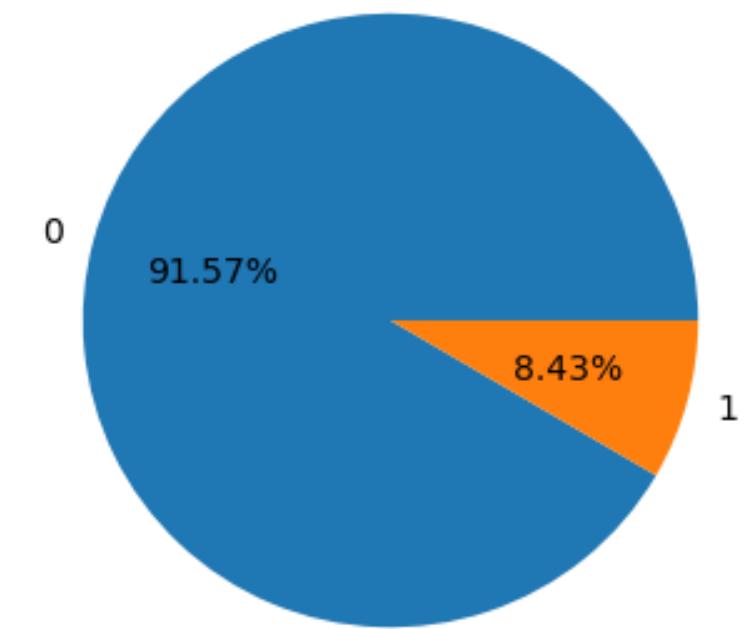
More loans: No child

Low defaulters:
2 documents
submitted

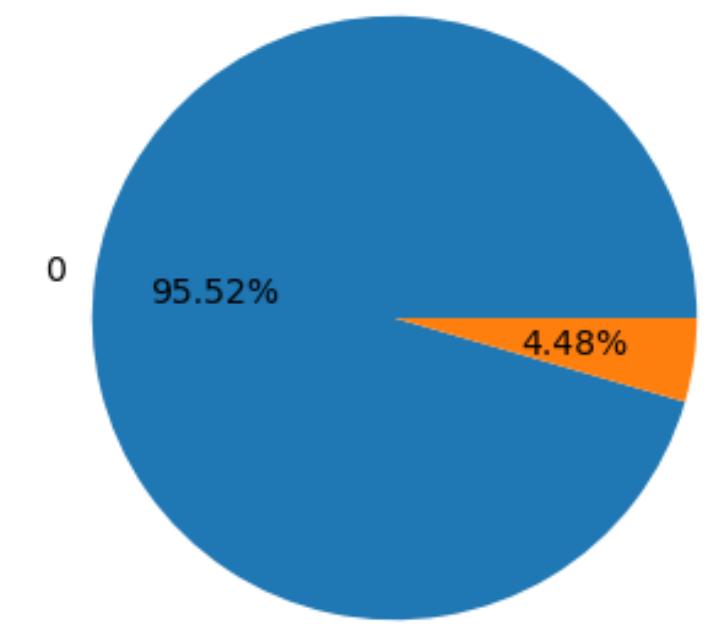
0 documents submitted



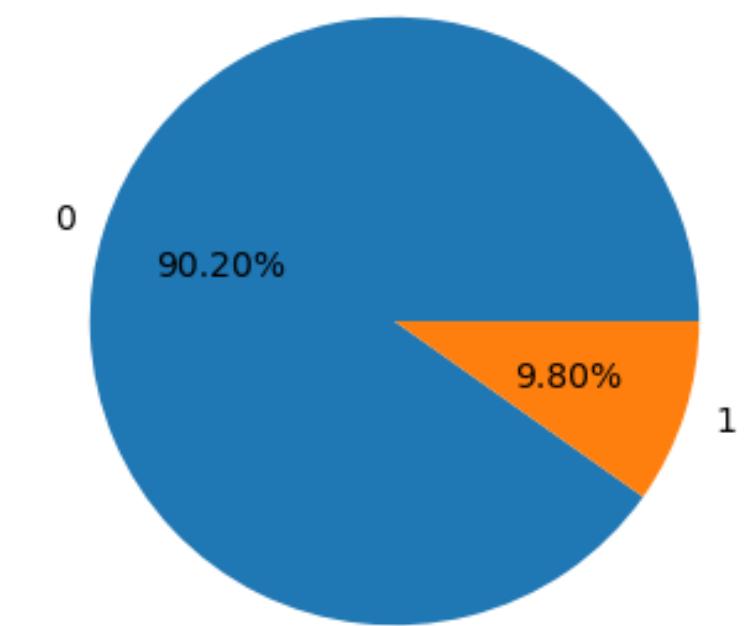
1 documents submitted



2 documents submitted



3 documents submitted

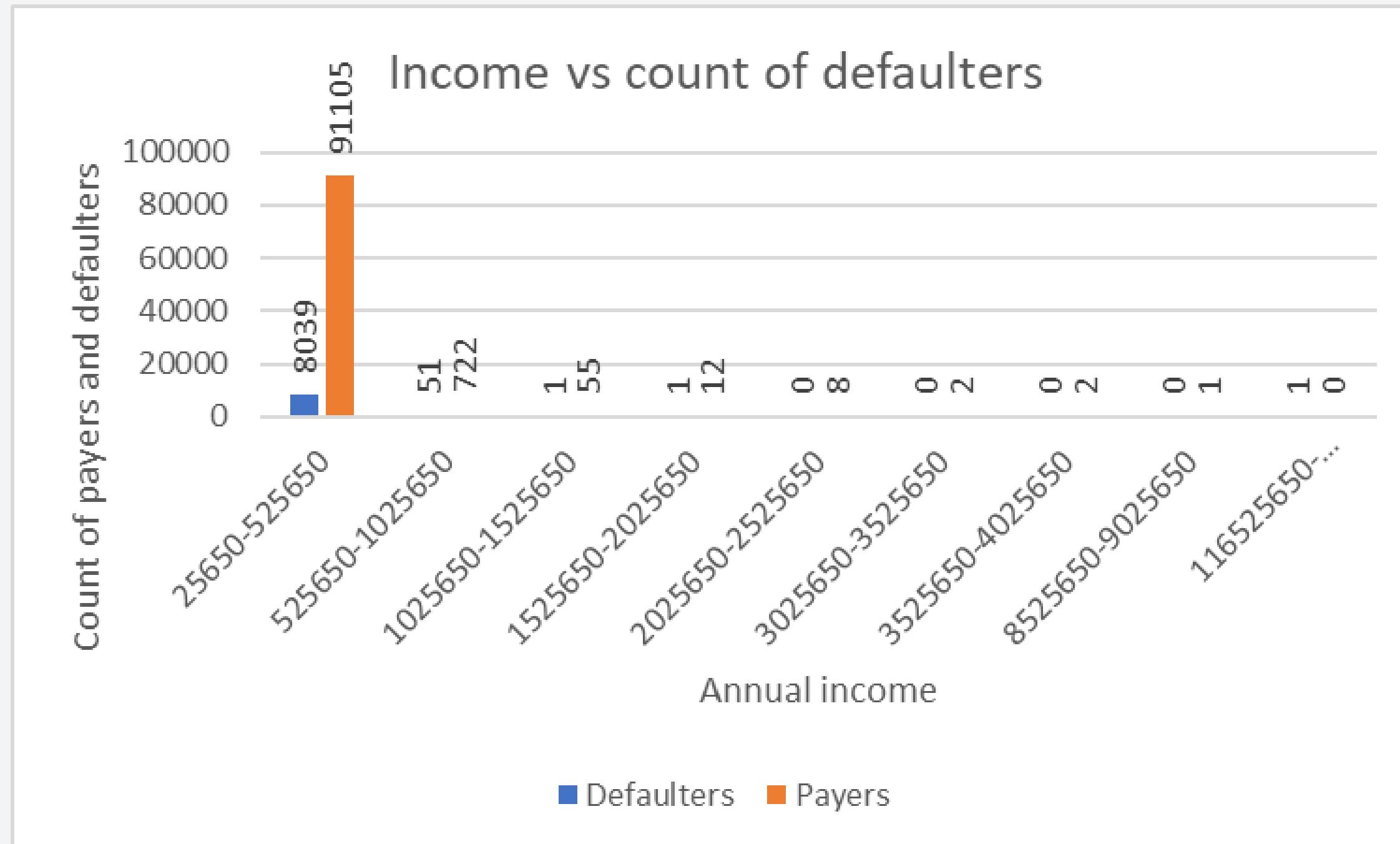


Factors effecting the probability of a customer being a defaulter are:

- Type of loan (Revolving)
- Gender (Females)
- People accompanying (Group of people)
- People's income type (Pensioners)
- People education type (Academic degree/ Higher education)
- People's family status (Widows)
- Number of documents submitted (2)



Let's analyze how numerical columns effect probability of being defaulters:



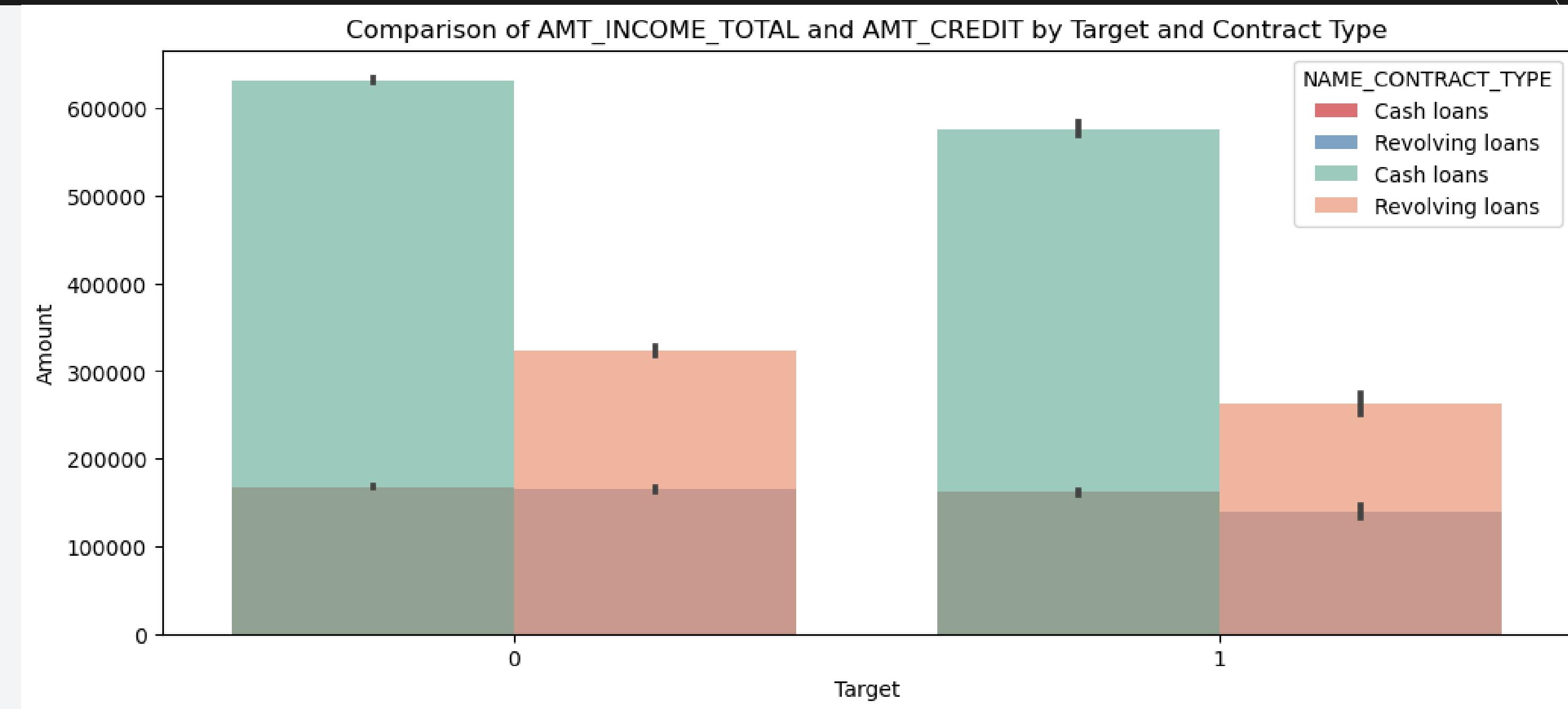
- Most of the people who are taking loans have an income range of 25k to 5L
- Most people who are not repaying are falling in the income range of 25k to 10L
- So, giving loans to people whose income is more than 10 lakhs is advisable.

Now let's analyze how the income is affected to the credit amount

TARGET	Average of AMT_CREDIT
Row Labels	
25650-525650	554726.2396
525650-1025650	852630.1765
1025650-1525650	1006920
1525650-2025650	781920
116525650-117025650	562491
Grand Total	556688.4605

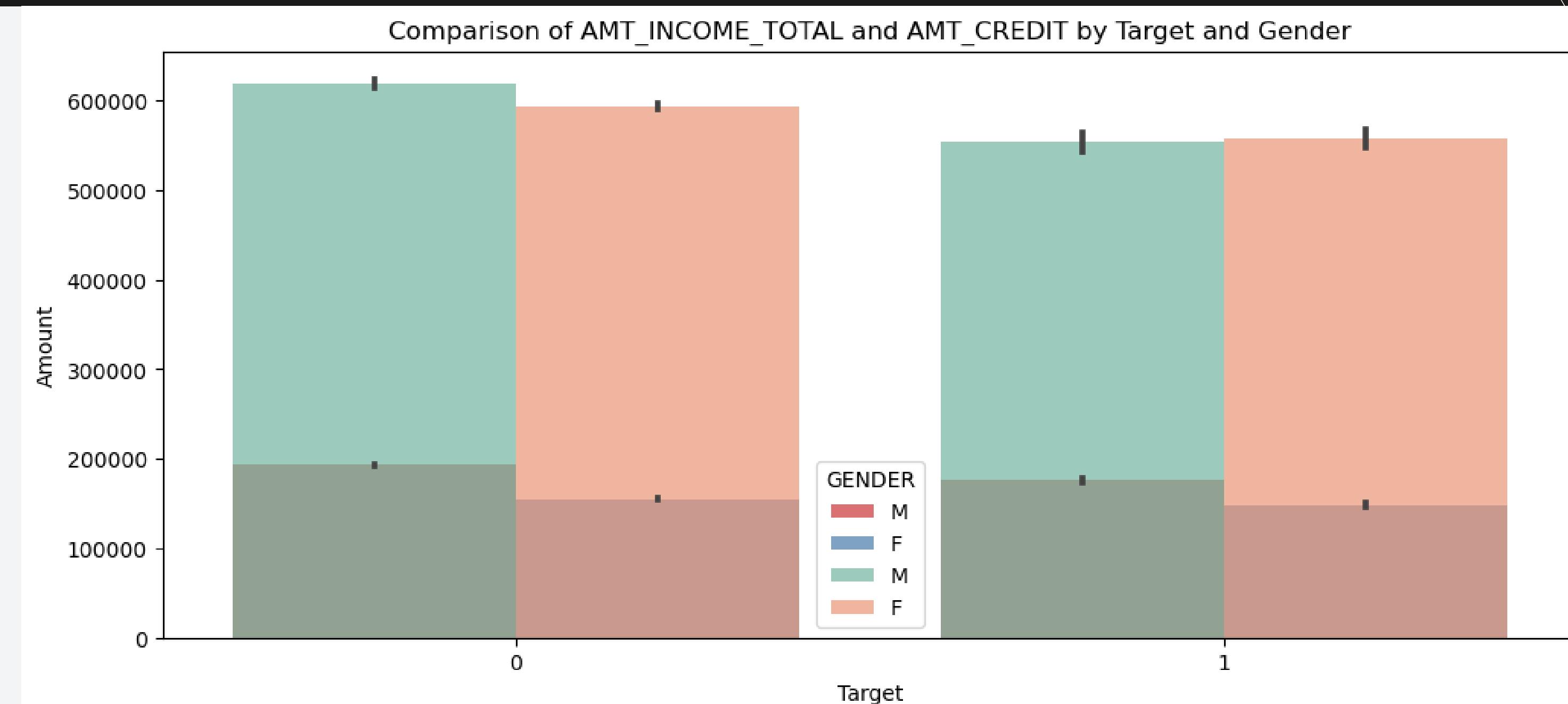
- From this table, we can understand that the bank gives customers more credit than their income.
- But we cannot decide that giving loans to customers more than their income is not advisable as many people cleared the loan though they were given more credit than their income.
- So we also need to collect and analyze the data of the products on which they are investing the money.

Let's analyze how the factors determined are affecting the numerical columns (MULTIVARIATE ANALYSIS)



The people are given way more credit than their income in cash loans.
So, there are more defaulters in people taking cash loans.

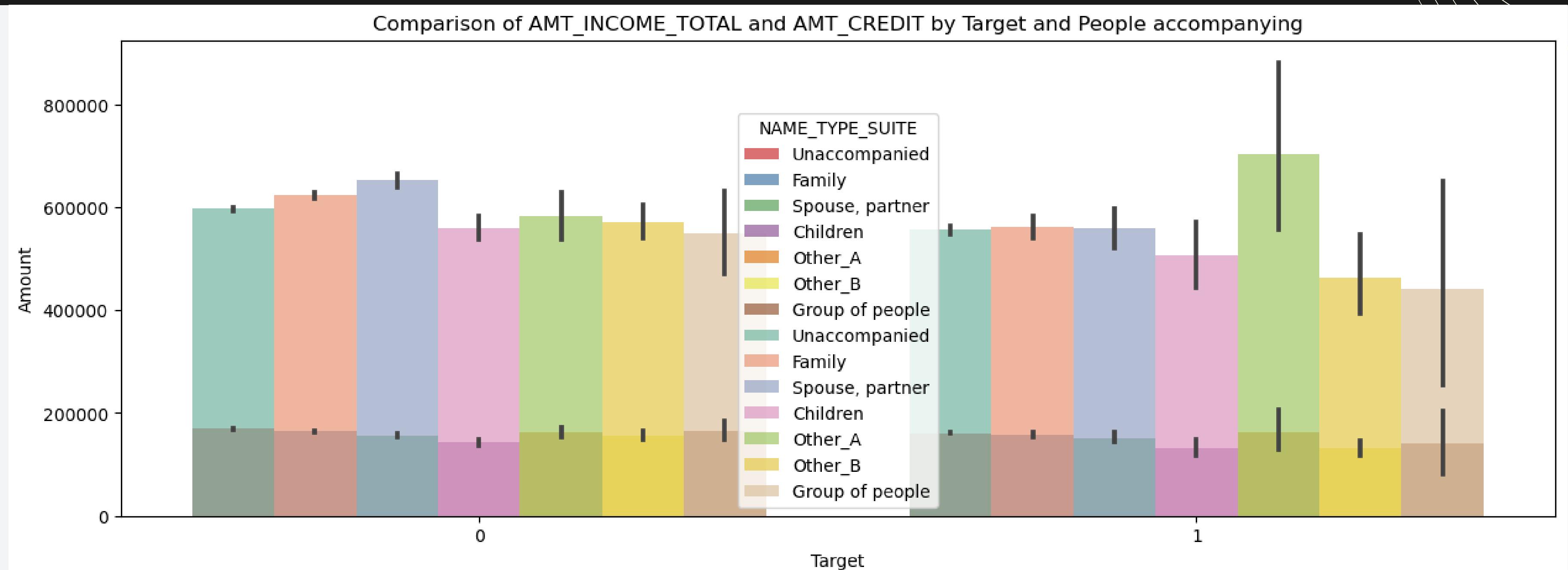
Let's analyze how the factors determined are affecting the numerical columns (MULTIVARIATE ANALYSIS)



Though men and women are given similar amounts of credits, females are able to pay but not males.

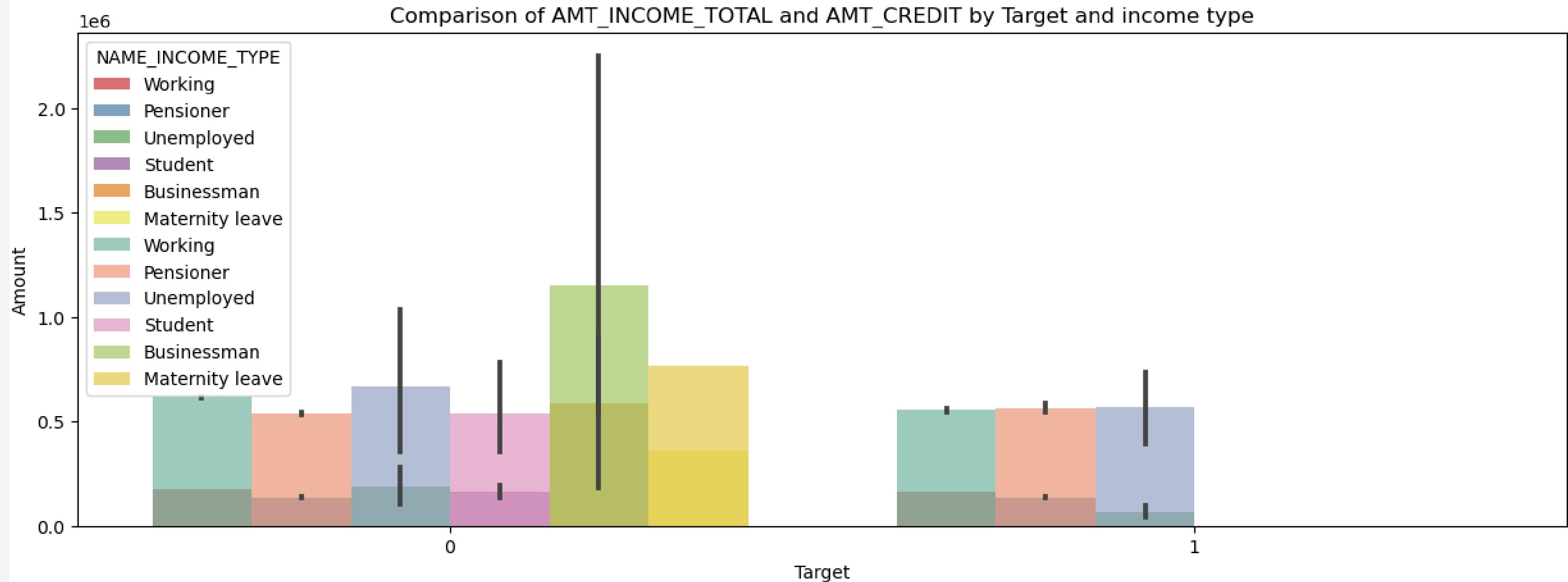
We need to analyze if there are any schemes that are being provided to females that are helping them to repay.

Let's analyze how the factors determined are affecting the numerical columns (MULTIVARIATE ANALYSIS)



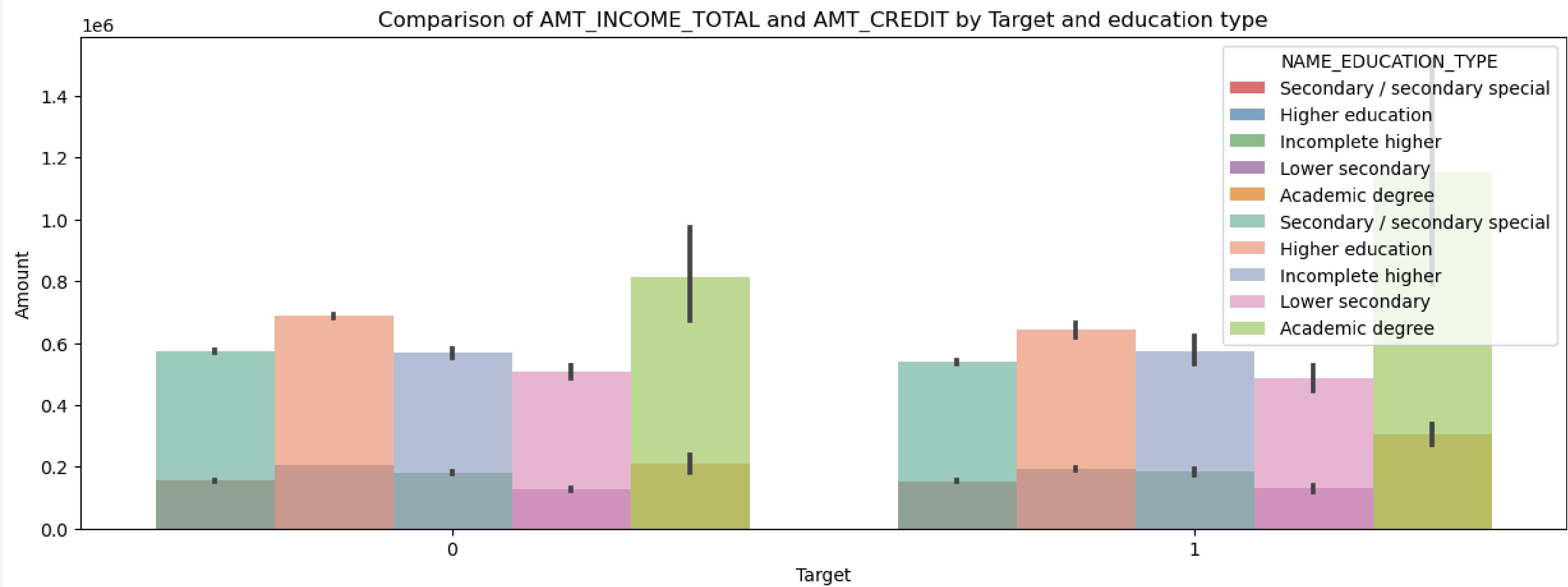
Of all the others, the amount of credit taken by people whom a group of people accompanies is less.
So there are fewer defaulters in them.

Let's analyze how the factors determined are affecting the numerical columns (MULTIVARIATE ANALYSIS)



Pensioners are taking less credit compared to others.
So there are few number of defaulters.

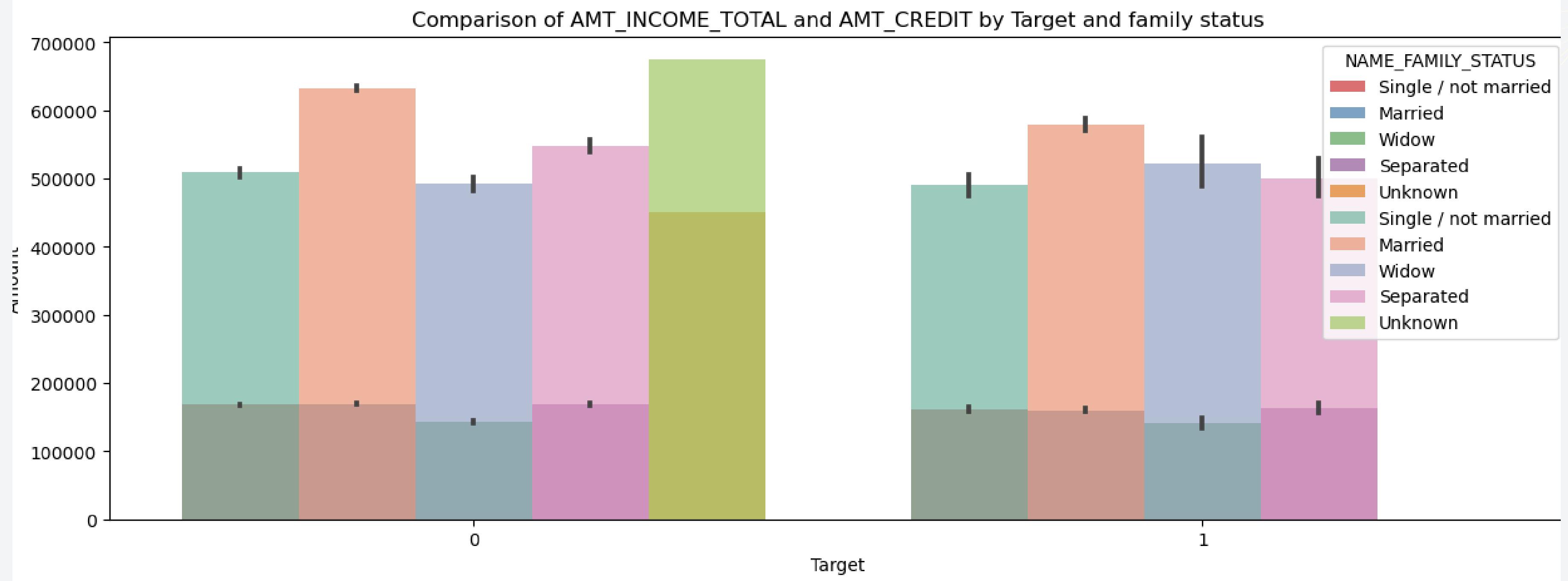
Let's analyze how the factors determined are affecting the numerical columns (MULTIVARIATE ANALYSIS)



Though people with higher education and academic degrees are given more credit than their income, fewer defaulters exist.

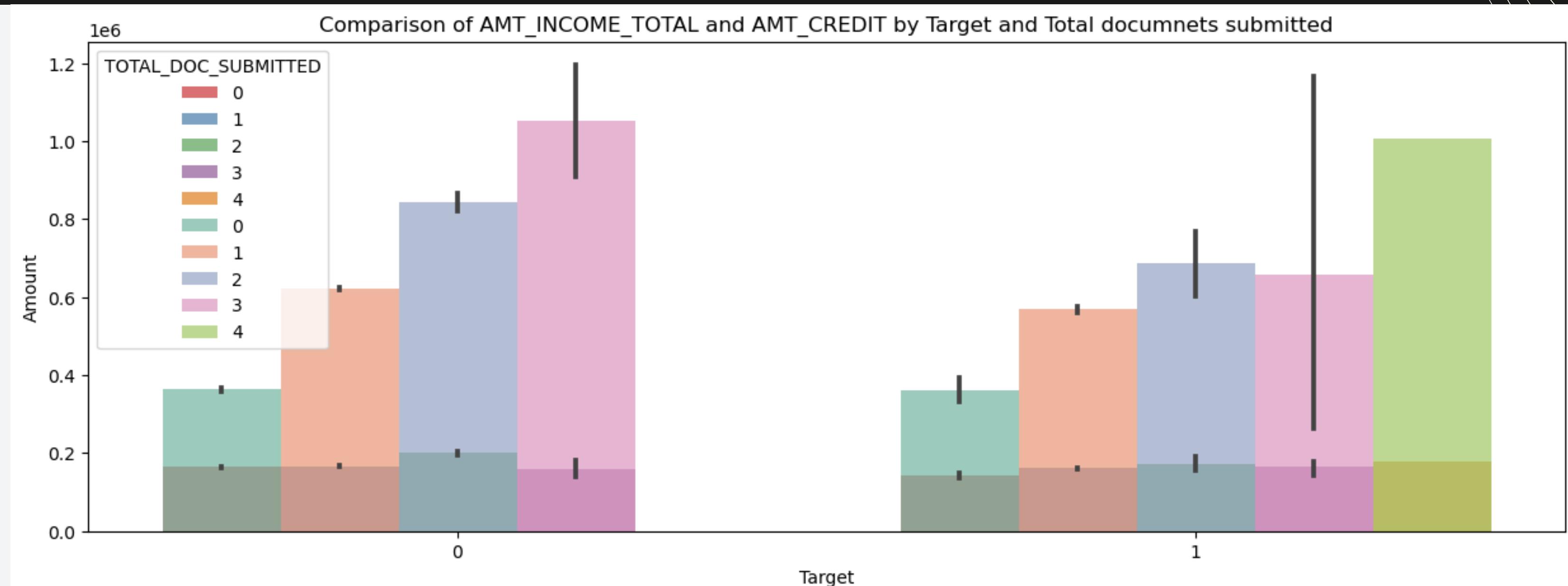
We need to analyze if their companies were giving them any extra benefits.

Let's analyze how the factors determined are affecting the numerical columns (MULTIVARIATE ANALYSIS)



Widows are taking less amount of credit compared to others.
So there are few defaulters.

Let's analyze how the factors determined are affecting the numerical columns (MULTIVARIATE ANALYSIS)



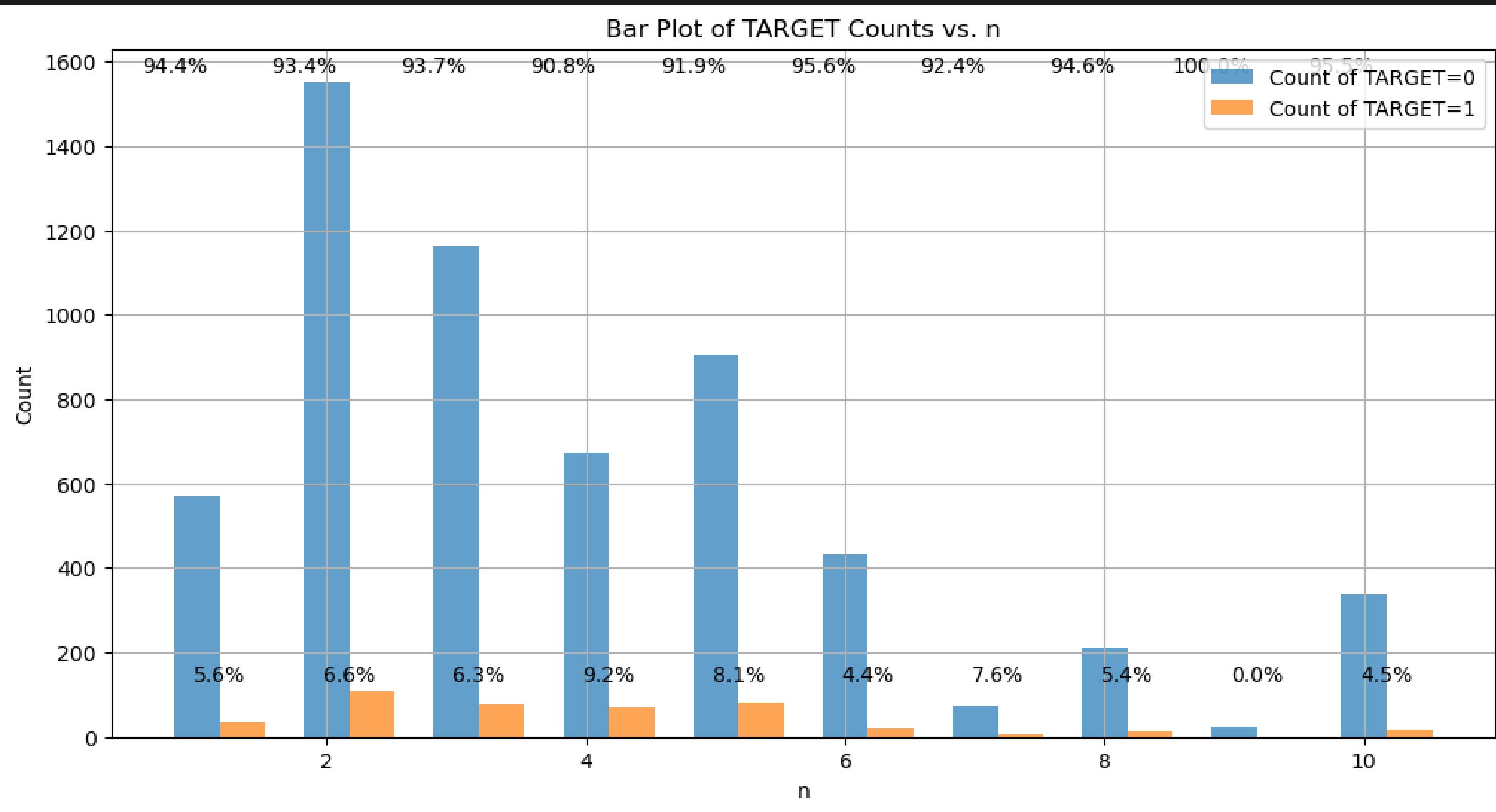
Though the amount of credit is way more than income in people submitting 2 documents, there are few defaulters. So we need to analyse the documents that are being submitted.

As there is huge difference in amount credit and income in people submitting 3 documents. There are more defaulters.

After all this analysis, we can understand that the **income** and the **credit amount** given are very important factors that must be considered.



Let's analyze the relationship between the income and credit amount



$$\text{AMT_CREDIT} = n * \text{AMT_INCOME_TOTAL}$$

CONCLUSION

Low probability of becoming defaulters

- Female customers with a higher education/academic degree and not unemployed could have a low probability of becoming defaulters.
- Giving loans with credit amounts more than the customer's income can lead to defaulters. But as more people take loans, more than income and repay it. We need to analyze the products they invest in after taking the credit.



THANK YOU

