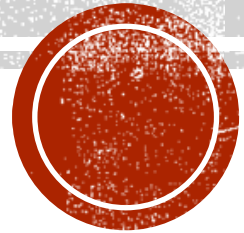
CREDIT EDA CASE STUDY

Credit Risk Analysis



BUSINESS OBJECTIVE :

- 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.

- 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.

PROBLEM STATEMENT :

The data which we have analyzed 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. (in our analysis, it is mentioned as target =1)
- All other cases: All other cases when the payment is paid on time. (in our analysis, it is mentioned as target =0)

ANALYSIS AND STEPS TAKEN :

1. **Data Sourcing (already provided in assignment)**
2. **Data loading and Data Cleaning** :

- *Fixing the rows and columns*
- *Imputing and Removing Missing columns*
- *Handling Outliers*

3. **Univariate Analysis :**

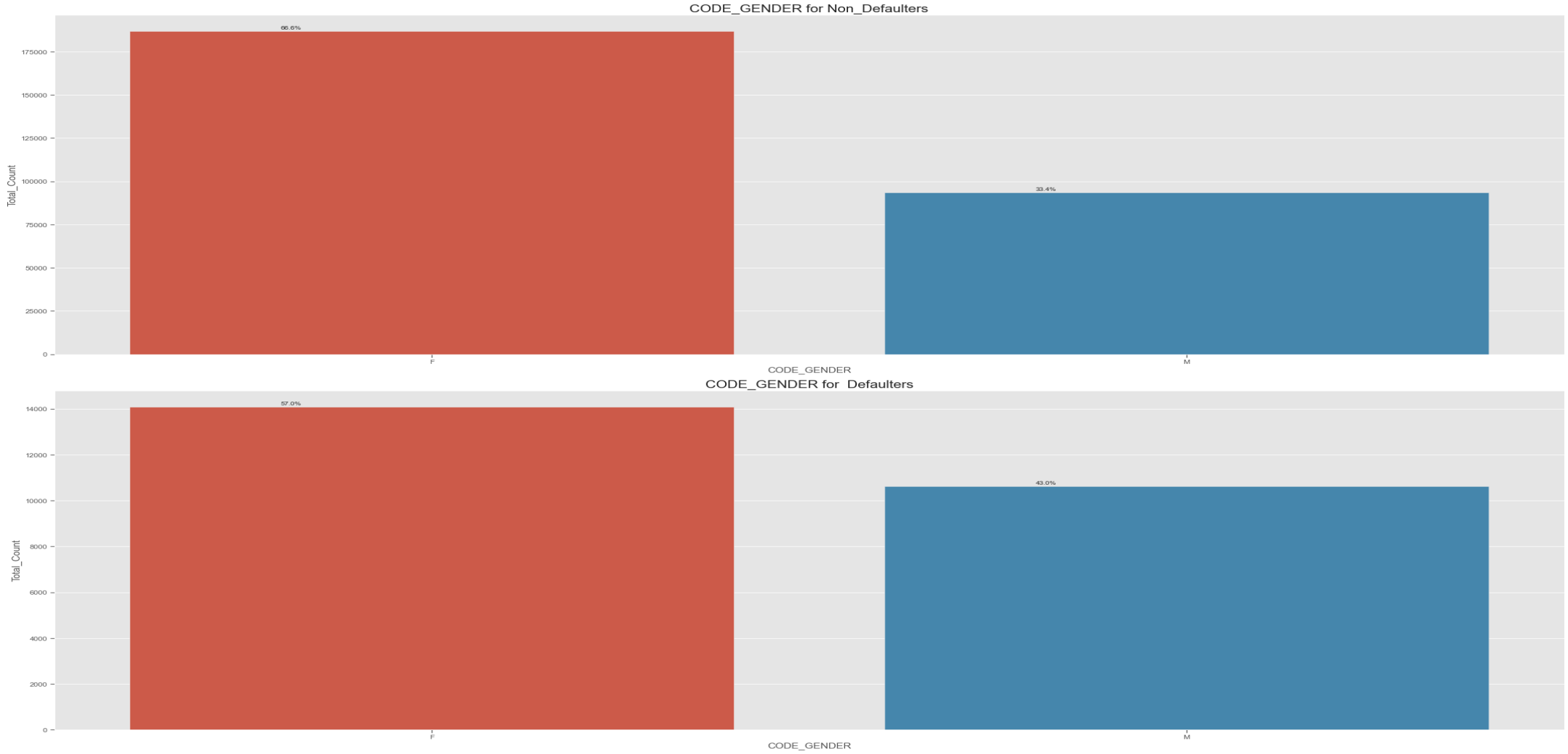
- *Categorical – Numerical Analysis*
- *Numerical- Numerical Analysis*
- *Categorical-Categorical Analysis*

4**. Bivariate and Multivariate Analysis** :

- *Numeric – Numeric Analysis*
- *Correlation*
- *Numerical – Categorical Analysis*
- *Categorical – Categorical Analysis*

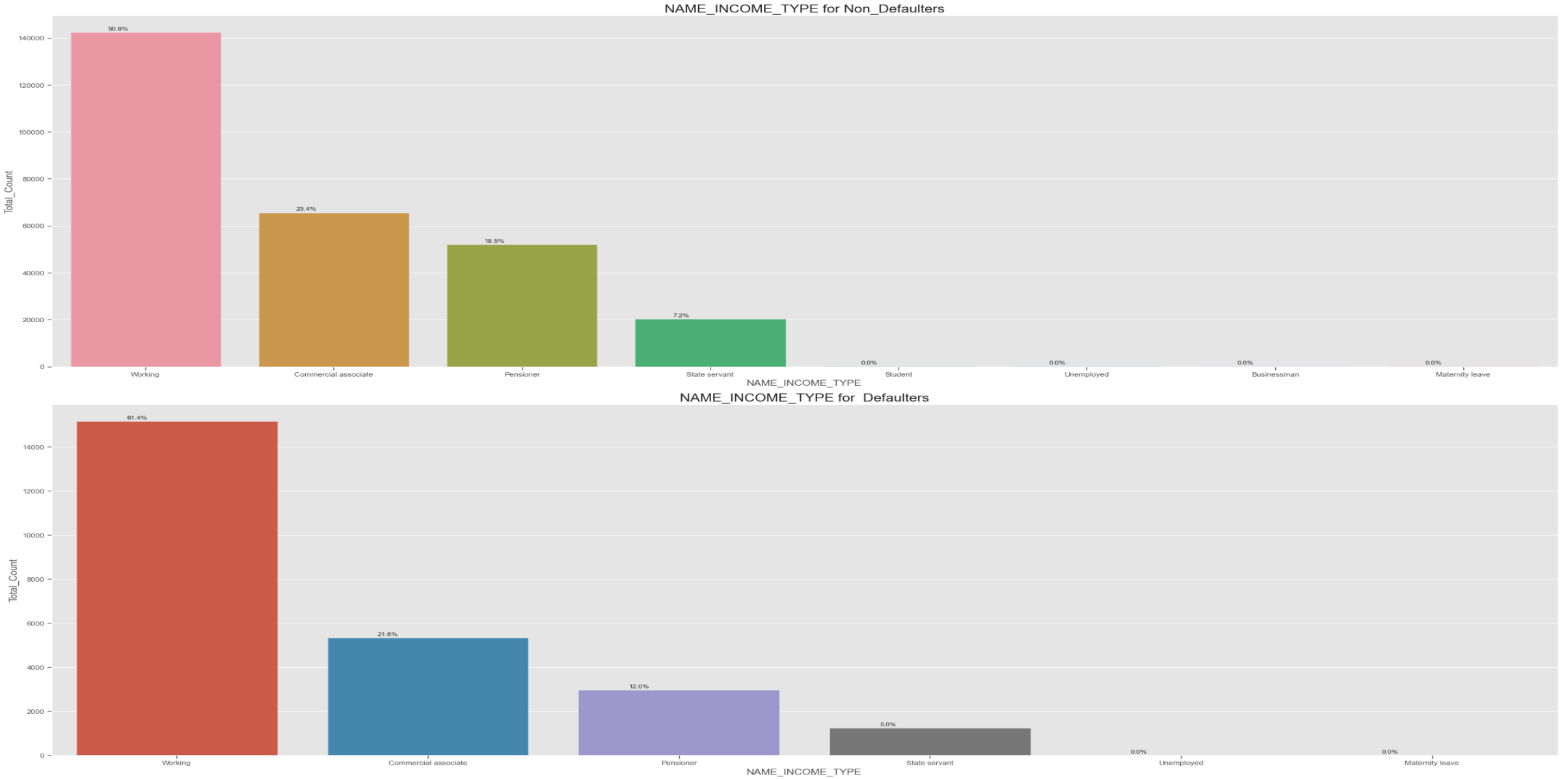
UNIVARIATE ANALYSIS FOR CATEGORICAL VARIABLES :

- GENDER :



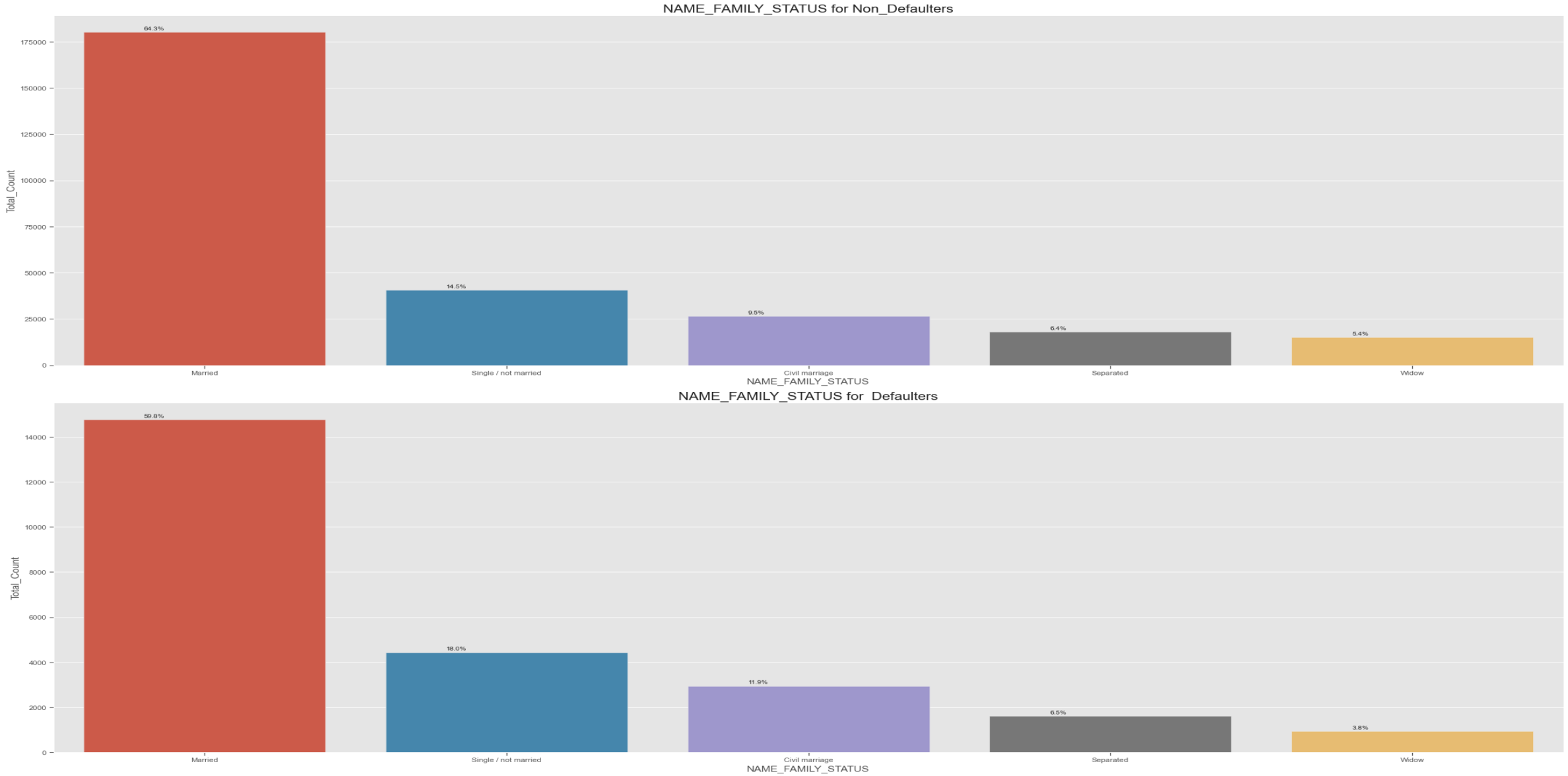
Females are the majority in both the cases, although there is an increase in the percentage in Male Payment Difficulties for target = 1

INCOME SOURCES :



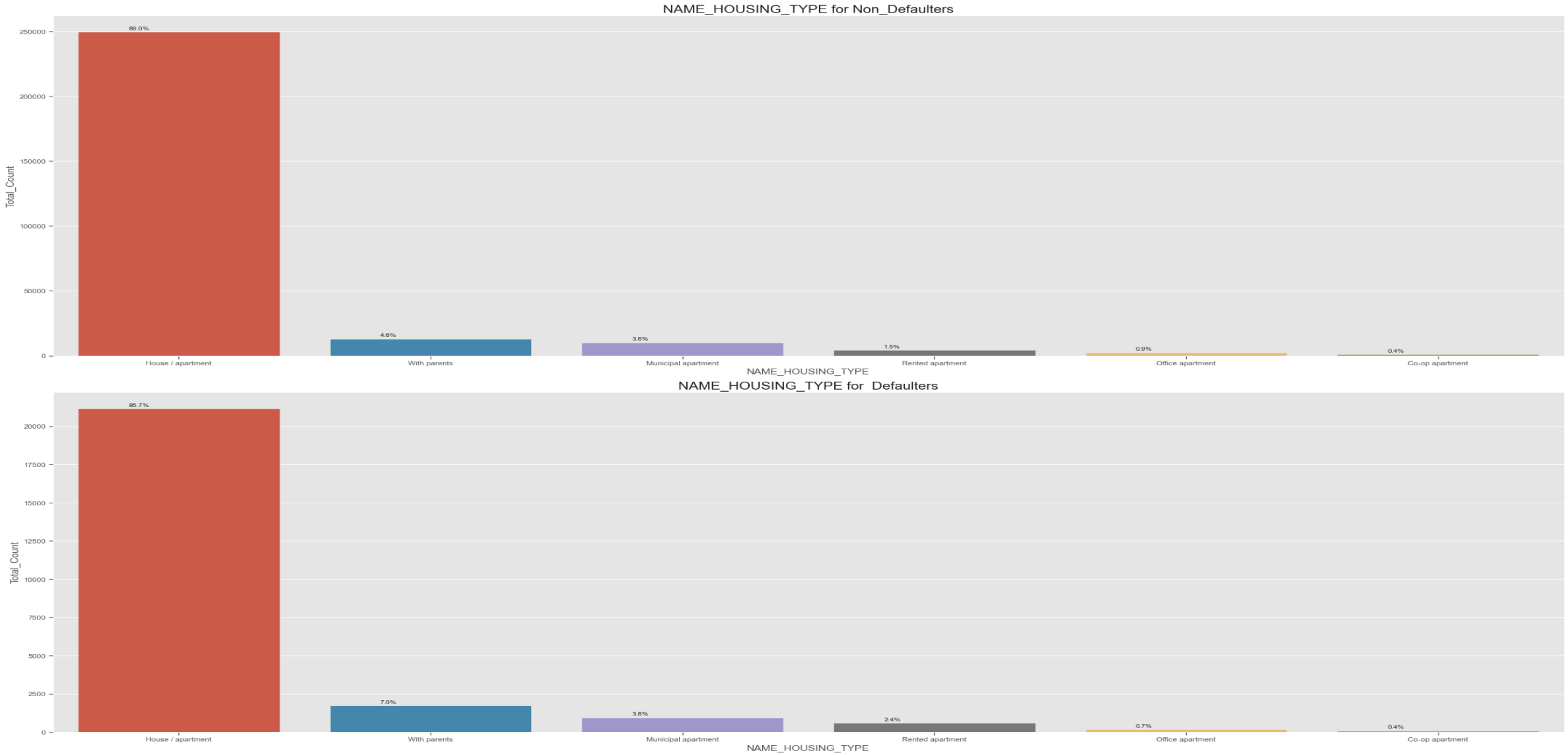
There is a similarity in the percentage of Payment Difficulties who are working when compared to the percentages of both Payment Difficulties and non-Payment Difficulties.

FAMILY STATUS :



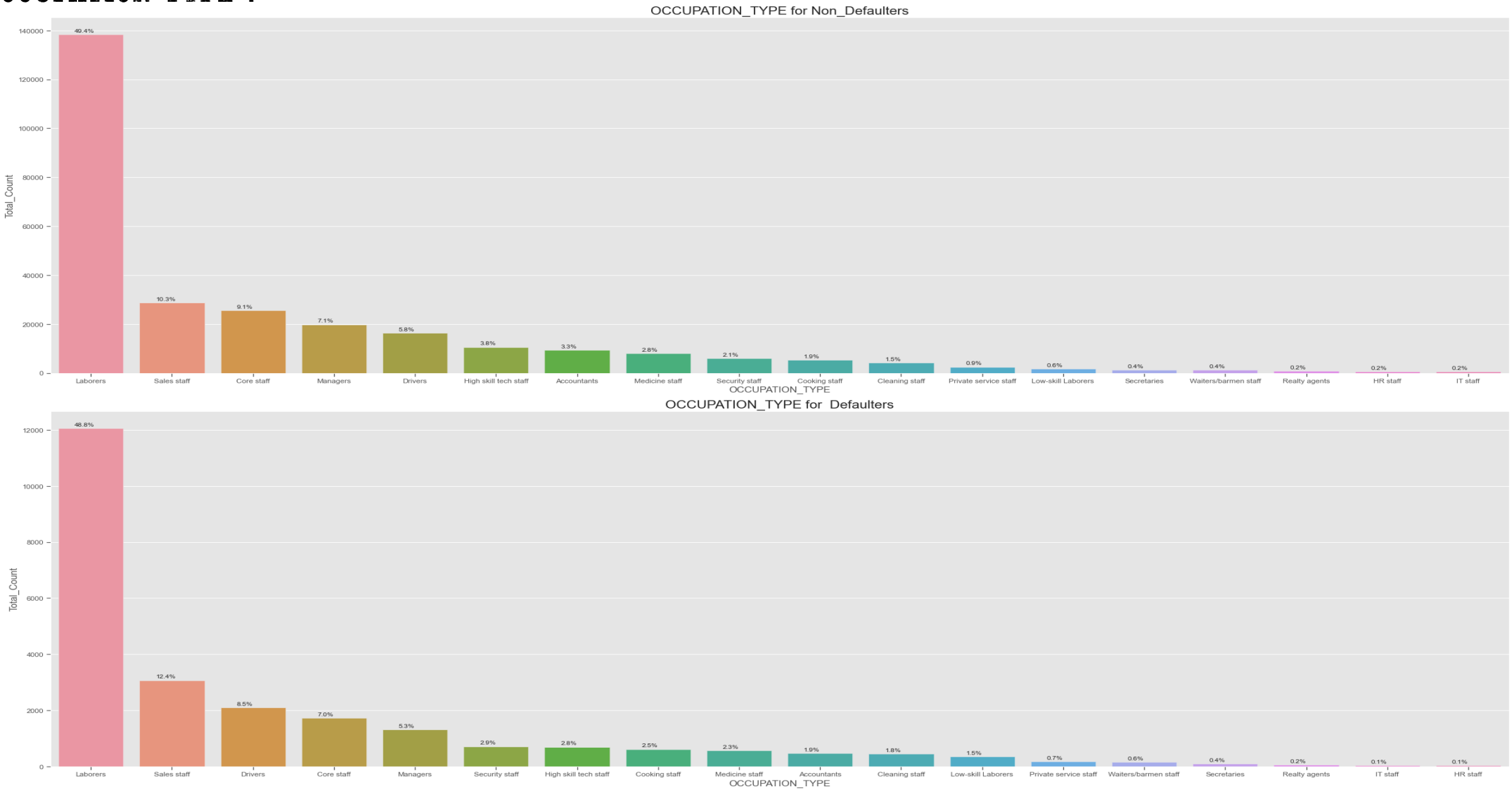
Single and Civil Marriage people percentage increases in defaulters compared to Non defaulters. indicates singles and Civil marriage people have higher chances of defaulting.

HOUSING TYPE :



The order for categories defaulting is similar to categories Not having payment difficulties, But people staying with parents have increased percentage in defaulting.

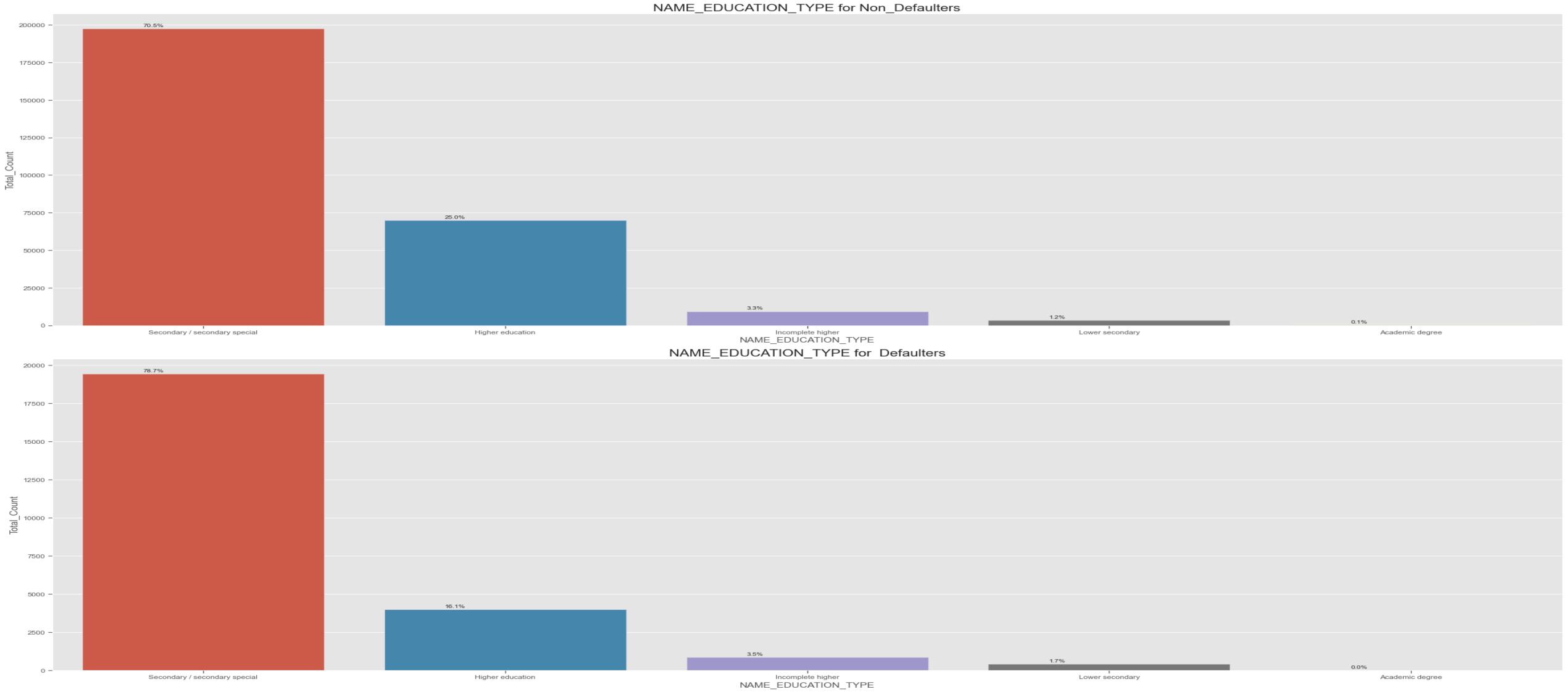
OCCUPATION TYPE :



- Laborers apply for the most number of loans.
- Drivers have increased chance of having payment difficulties.

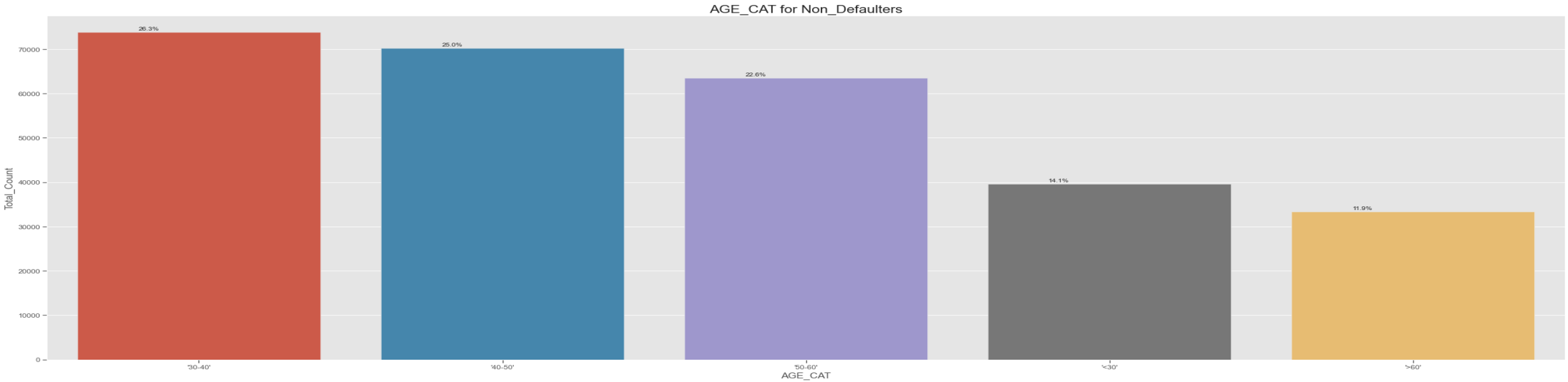
- # UNIVARIATE CATEGORICAL ORDERED ANALYSIS

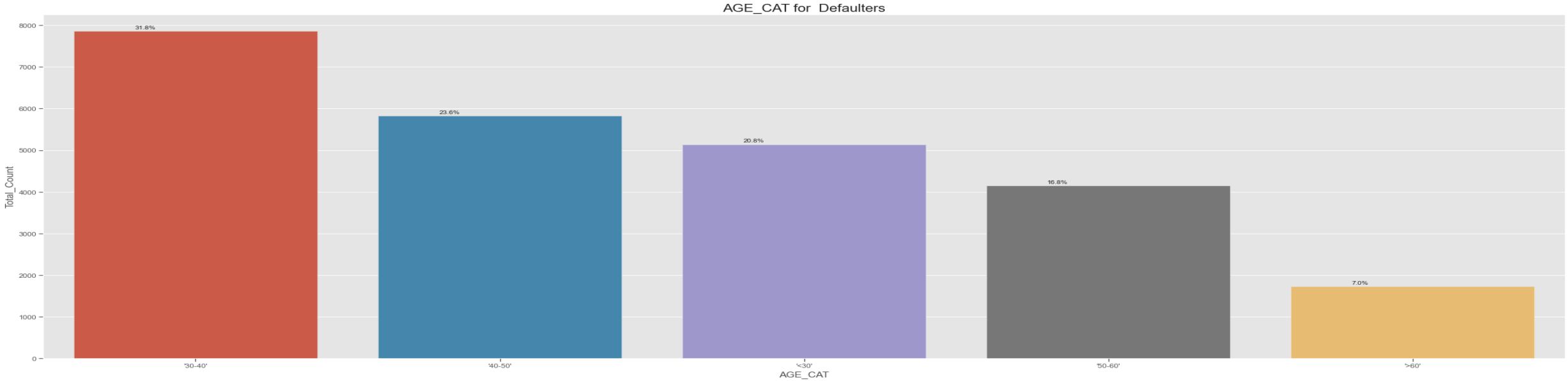
EDUCATION TYPE :



- Secondary Education type apply for more loans compared to all other categories.
- People with higher education have less payment difficulties which may indicate they having better jobs due to better education.

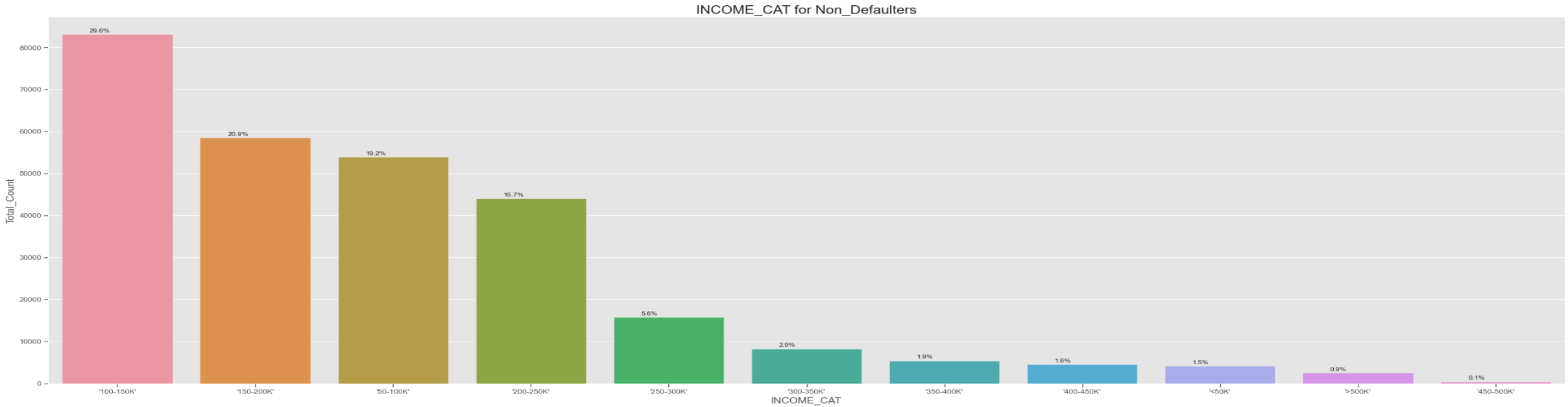
AGE :

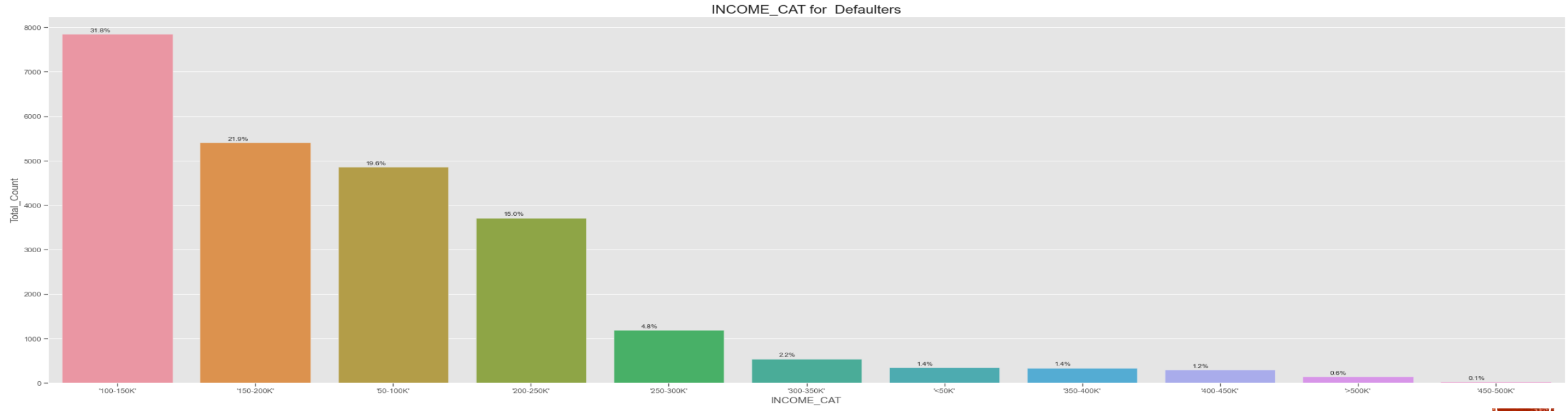




People in age category '30-40' apply for most loans while also have more difficulties paying back loans along with people in category '<30.

- # INCOME RANGE :

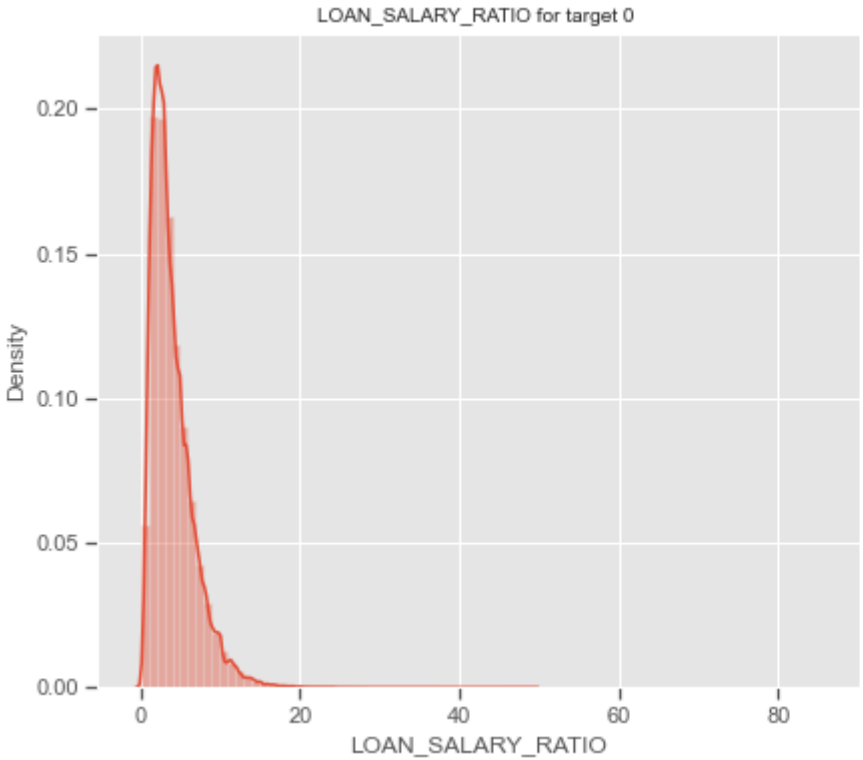


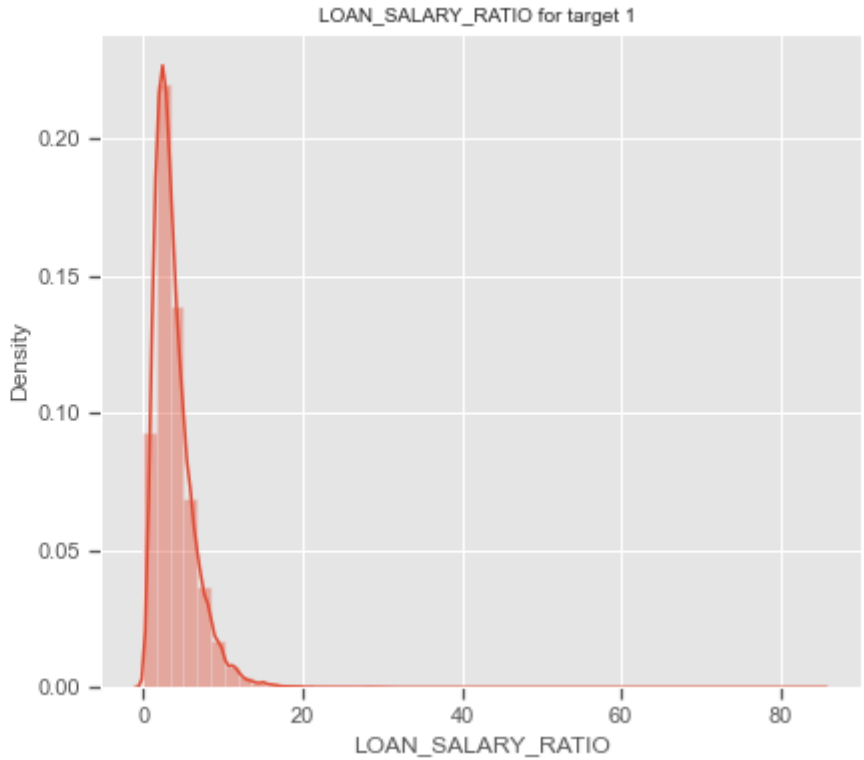


People with in lower range of income('1-1.5L') apply for more loan, but the percentages of categories are similar in both.

FOR CONTINOUS VARIABLES

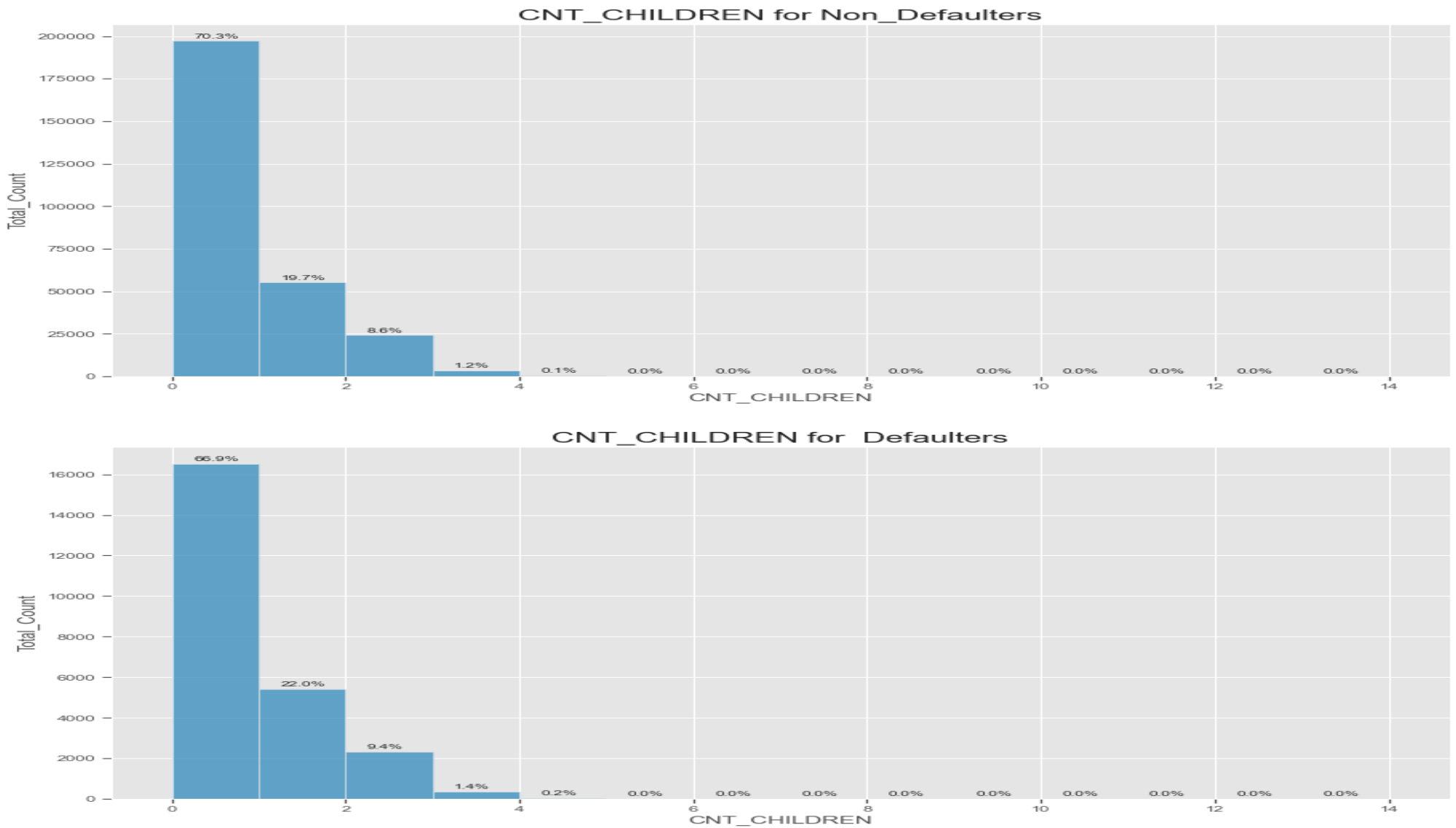
Loan – Salary Ratio :





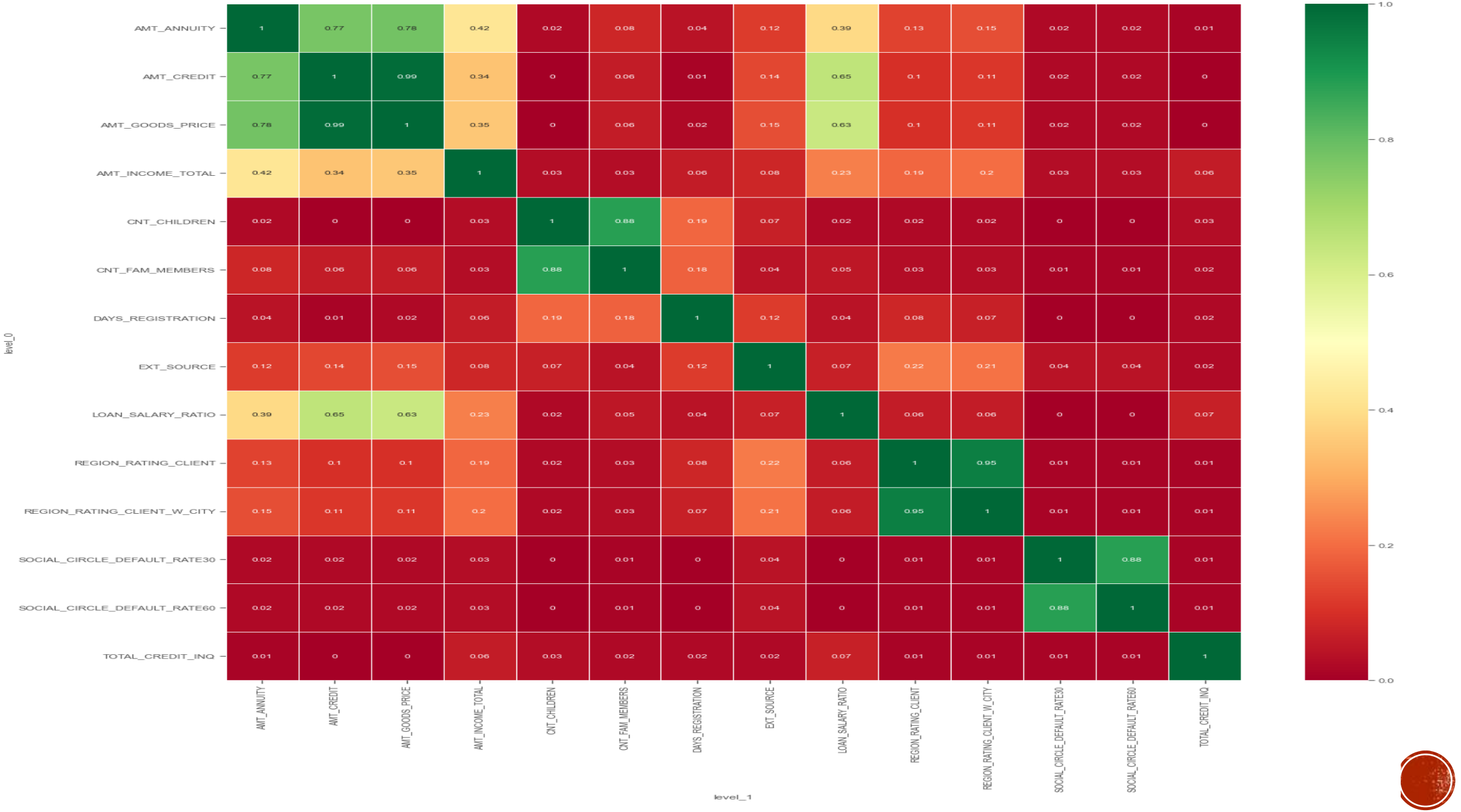
Loan salary ratio is similar for both Targets but having a loan salary ratio of greater than 50 makes it very likely to have payment difficulties.

NUMBER OF CHILDREN :



Clients having zero children have less difficulties in paying loans compared to clients having more than '0' kids.

CORRELATION MATRIX OF TARGET 0 :



THESE ARE THE TOP 10 CORRELATION CATEGORIES FOR TARGET 0. WE CAN SEE THE SAME IN ABOVE CORRELATION MATRIX.

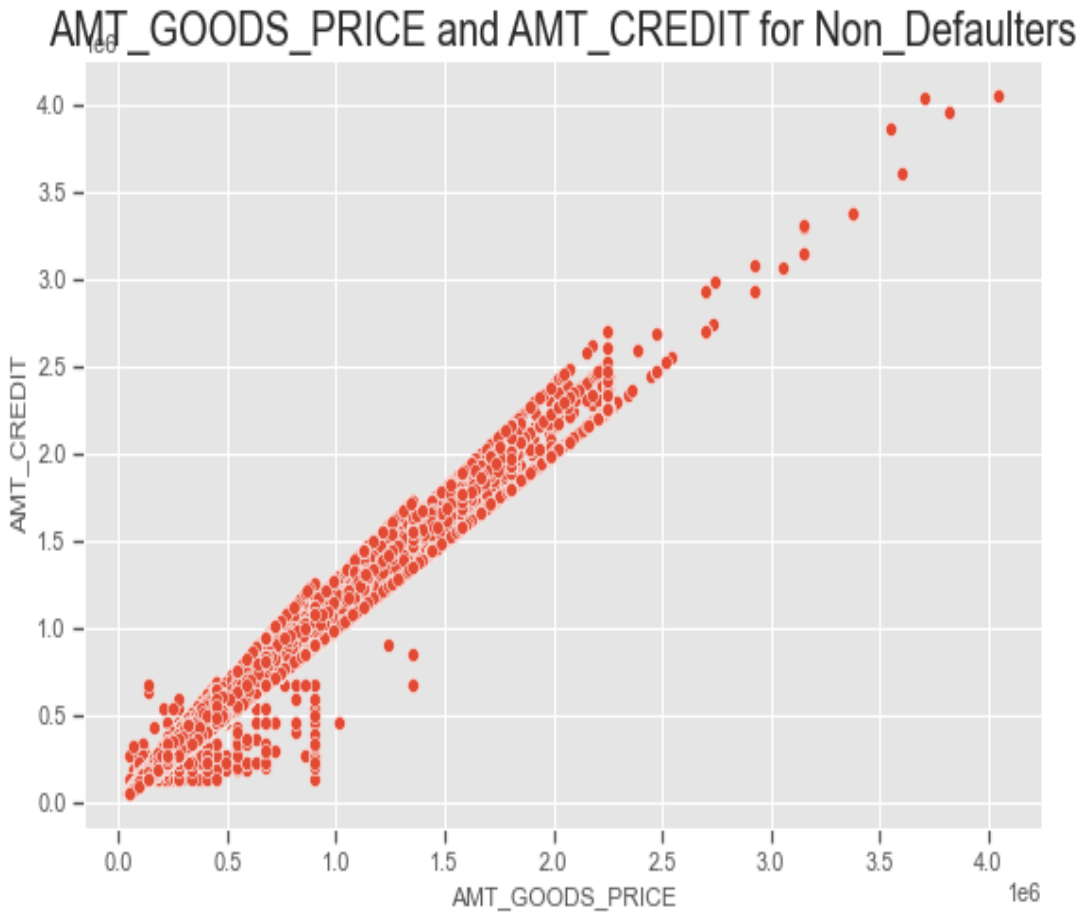
	level_0	level_1	0
210	AMT_GOODS_PRICE	AMT_CREDIT	0.99
124	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.95
152	SOCIAL_CIRCLE_DEFAULT_RATE60	SOCIAL_CIRCLE_DEFAULT_RATE30	0.88
111	CNT_FAM_MEMBERS	CNT_CHILDREN	0.88
219	AMT_GOODS_PRICE	AMT_ANNUITY	0.78
43	AMT_CREDIT	AMT_ANNUITY	0.77
66	LOAN_SALARY_RATIO	AMT_CREDIT	0.65
77	LOAN_SALARY_RATIO	AMT_GOODS_PRICE	0.63
59	AMT_INCOME_TOTAL	AMT_ANNUITY	0.42
75	LOAN_SALARY_RATIO	AMT_ANNUITY	0.39

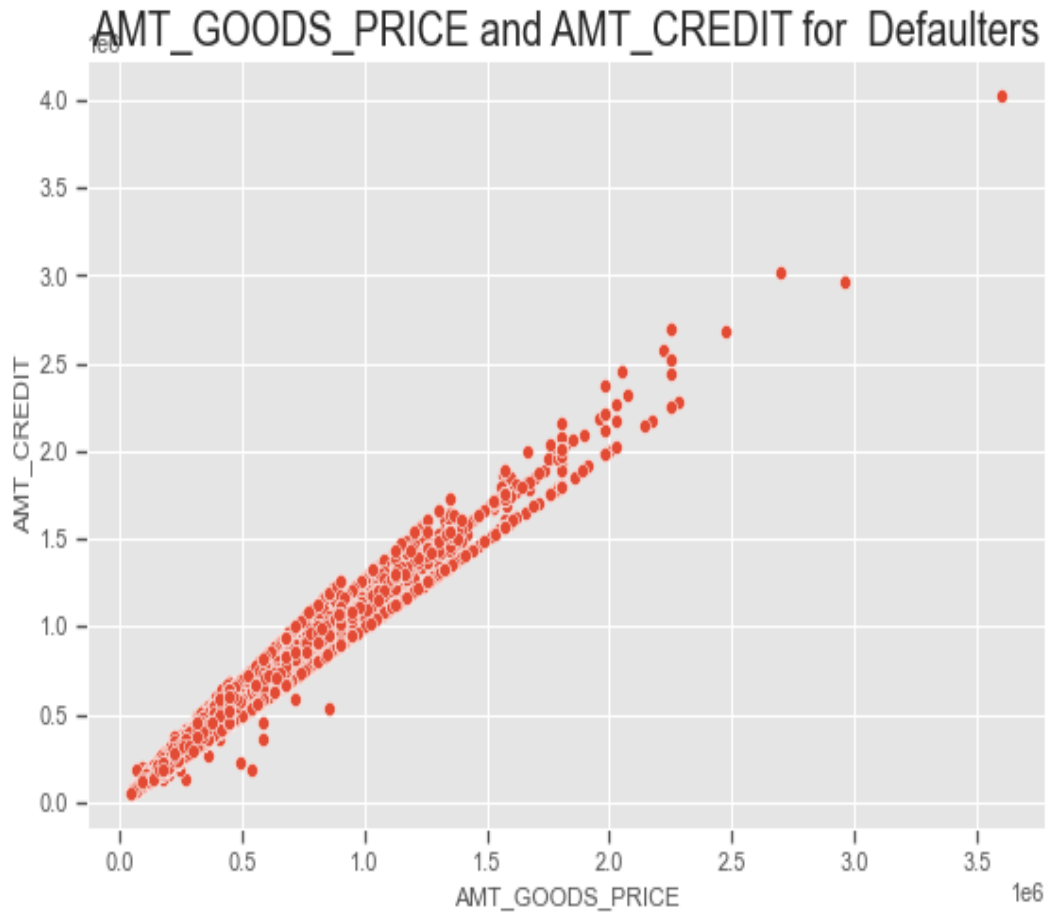
CORRELATION MATRIX OF TARGET 1

THESE ARE THE TOP 10 CORRELATION CATEGORIES FOR TARGET 1. WE CAN SEE THE SAME IN ABOVE CORRELATION MATRIX.

	level_0	level_1	0
210	AMT_GOODS_PRICE	AMT_CREDIT	0.98
124	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.96
152	SOCIAL_CIRCLE_DEFAULT_RATE60	SOCIAL_CIRCLE_DEFAULT_RATE30	0.89
111	CNT_FAM_MEMBERS	CNT_CHILDREN	0.89
219	AMT_GOODS_PRICE	AMT_ANNUITY	0.75
43	AMT_CREDIT	AMT_ANNUITY	0.75
66	LOAN_SALARY_RATIO	AMT_CREDIT	0.64
77	LOAN_SALARY_RATIO	AMT_GOODS_PRICE	0.63
75	LOAN_SALARY_RATIO	AMT_ANNUITY	0.38
206	REGION_RATING_CLIENT	EXT_SOURCE	0.19

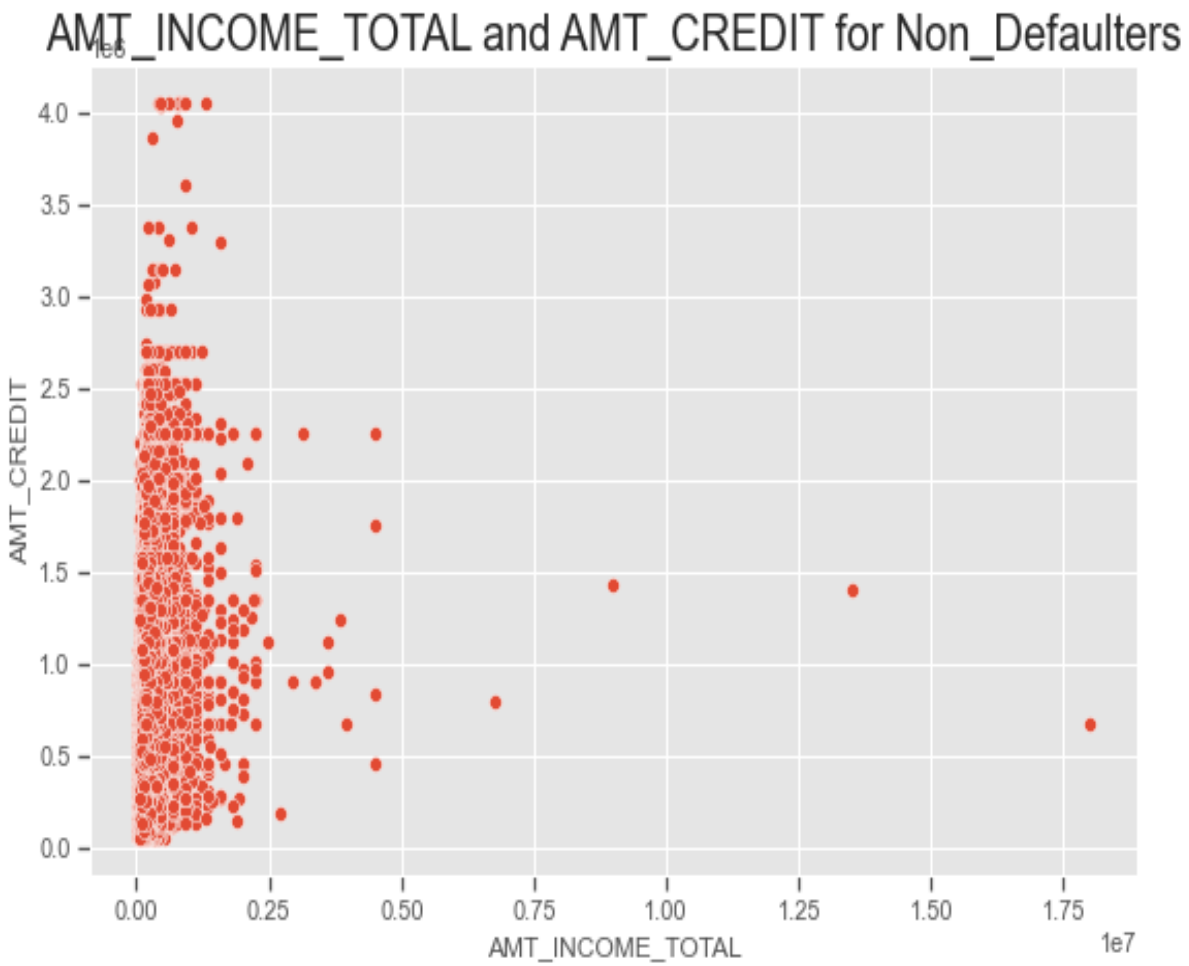
AMT_GOODS_PRICE' and AMT_CREDIT :

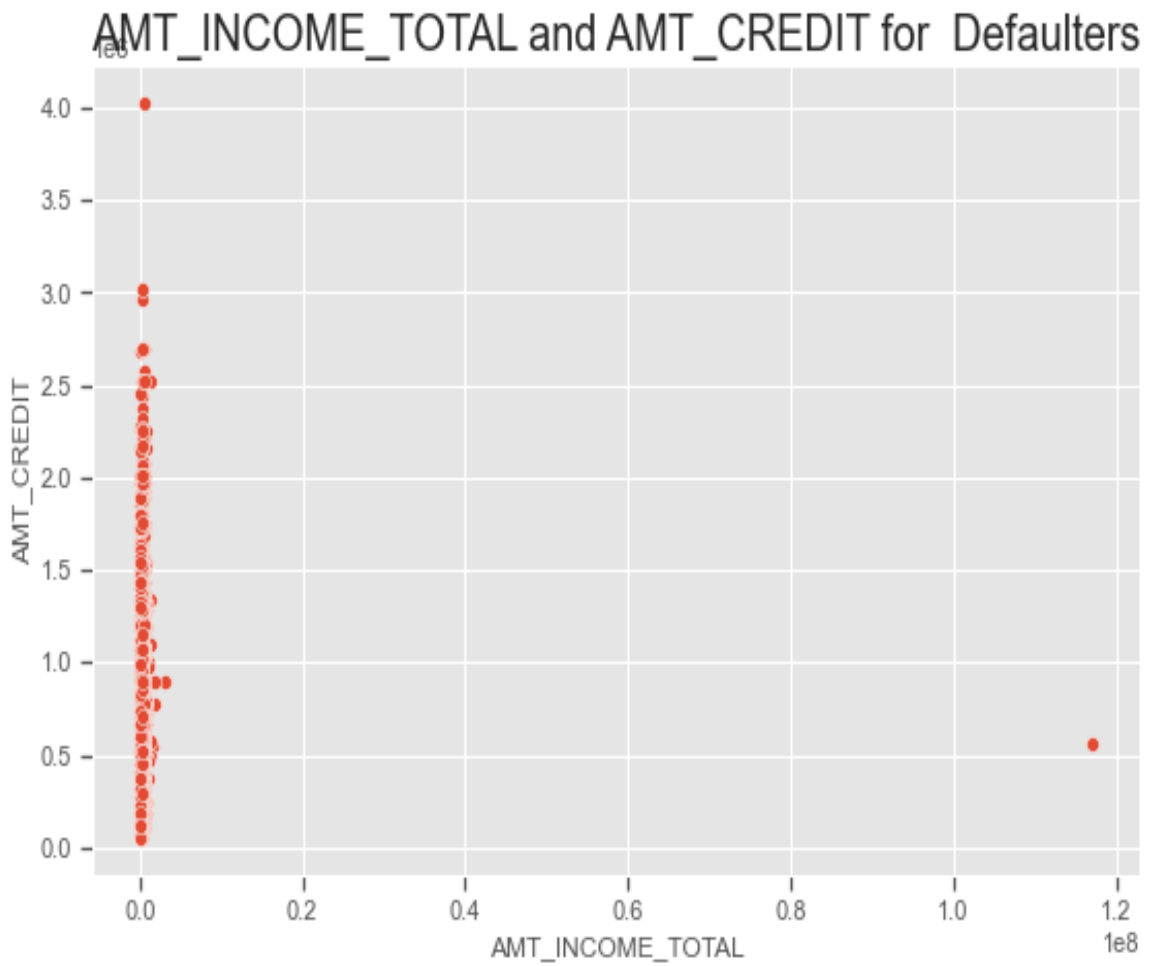




'AMT_GOODS_PRICE','AMT_CREDIT" show high amount of correlation as we can see higher amt_credit is related to high goods_price value and lower amt_credit is related to lower goods price value.
We can observe that large amt_credit and people with larger goods_price default less often.

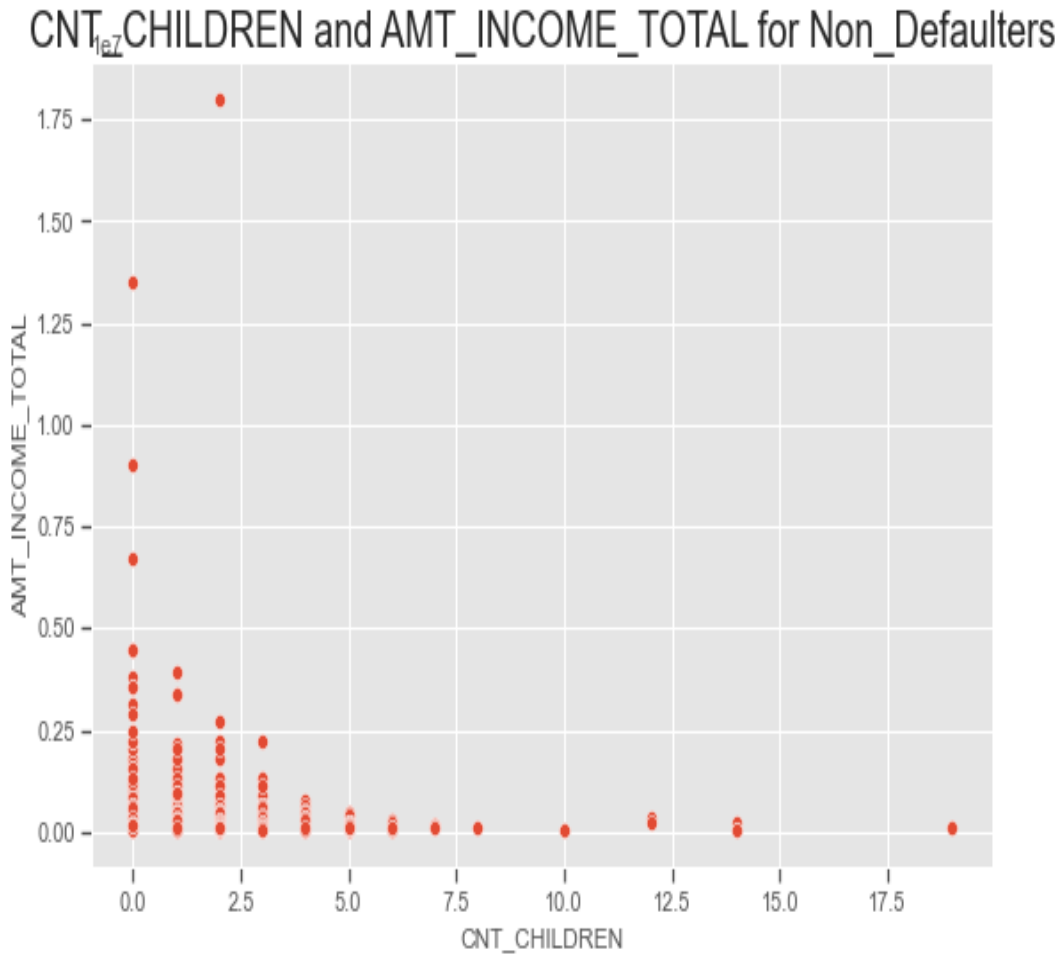
AMT_INCOME_TOTAL AND AMT_CREDIT :

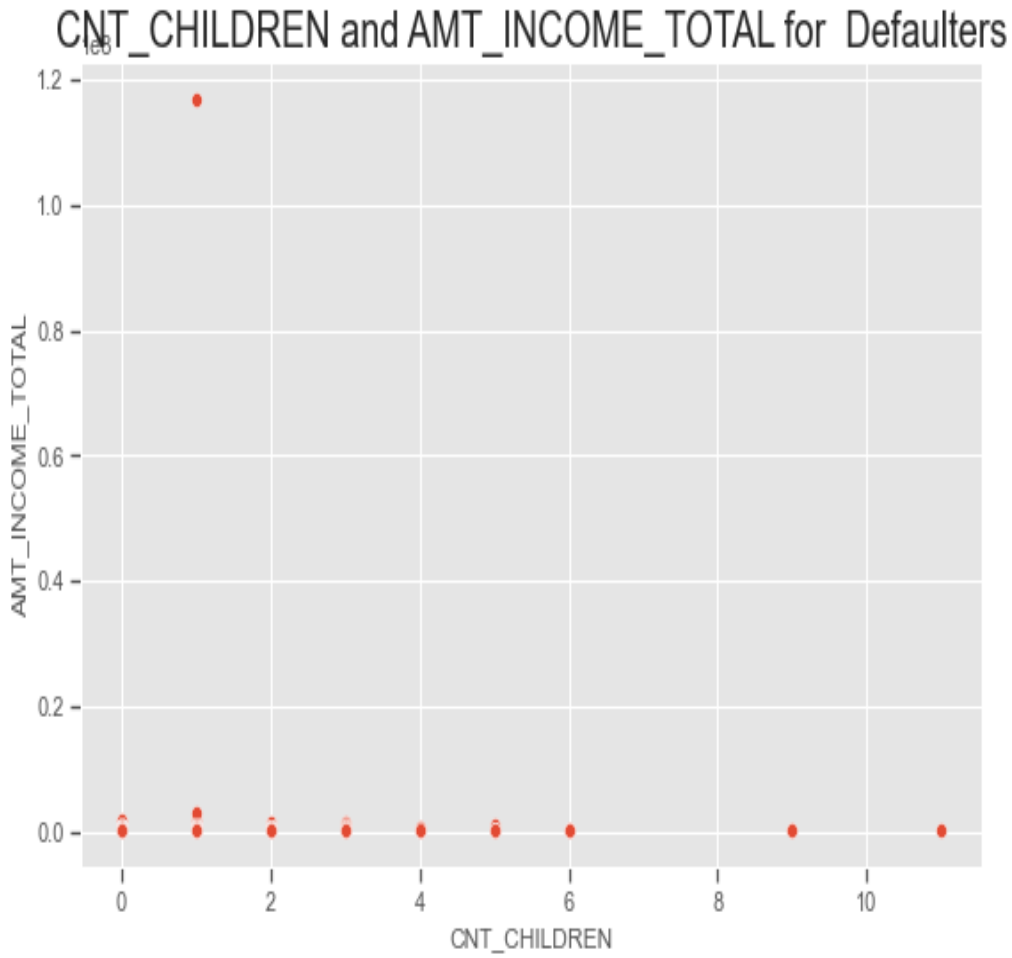




Clients with more income have less payment difficulties and people with lower income take more amount of loans.

CNT_CHILDREN AND AMT_INCOME_TOTAL :

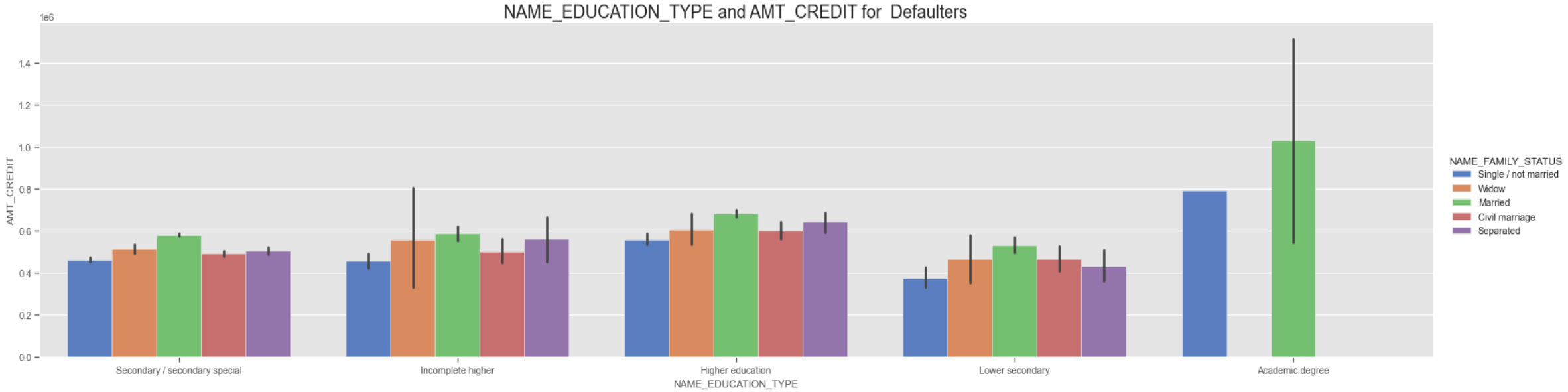


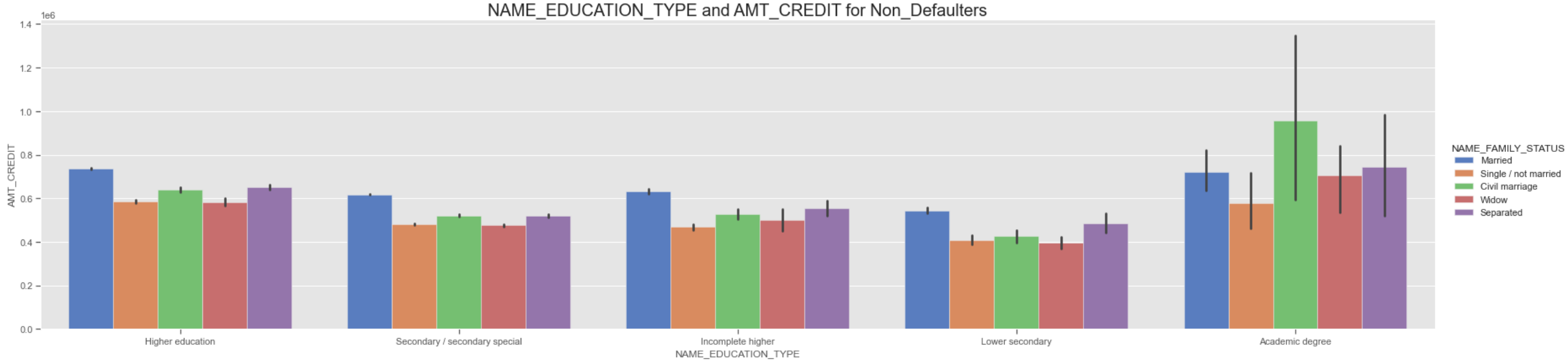


Clients with fewer children have higher total_income and have less payment difficulties compared to clients with more children.

>>>CATEGORICAL TO NUMERICAL VALUES :

NAME_EDUCATION_TYPE VS AMT_CREDIT :





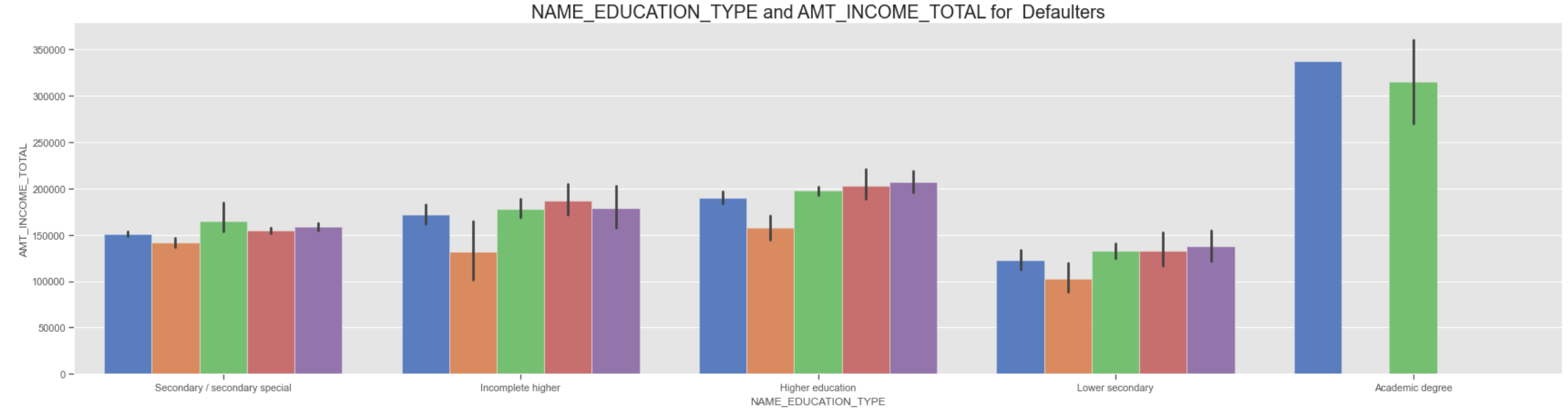
- *Points for target = 0 (Non-Defaulters).*

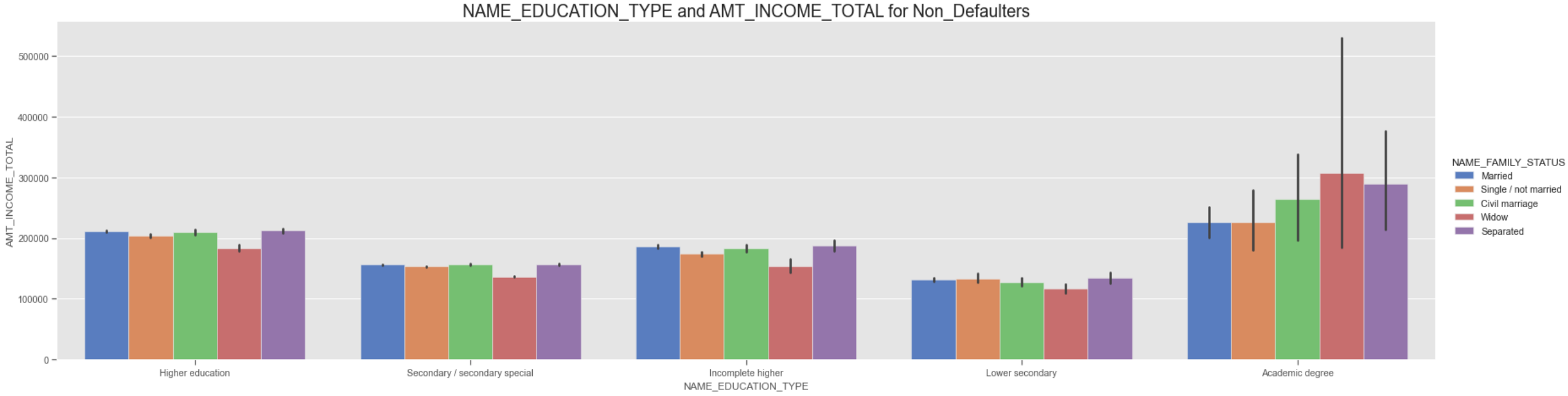
1. Customers holding academic degree have greater credit amount, Civil marriage segment being the highest among them.
2. Lower educated customers tends to have lower credit amount, Widows being the lowest among them
3. Married customers in almost all education segment except lower secondary and academic degrees have a higher credit amount.

- *Points for target = 0 (Non-Defaulters).*

1. Married Academic degree holding customers generally have a higher credit amount and so their defaulting rate is also high
2. Across all education segment married customer tends to have higher credit amount
3. Customers holding lower education tends to have a lower credit amount
4. Single and Married are the only 2 family types present in academic degree .

'NAME_EDUCATION_TYPE','AMT_INCOME_TOTAL' :





- *Points for Target 0*

1. Customer having academic degree have higher income, widows having the highest.
2. Customer having lower level of education earn less.
3. Separated have higher income level in all categories except in academic degree.

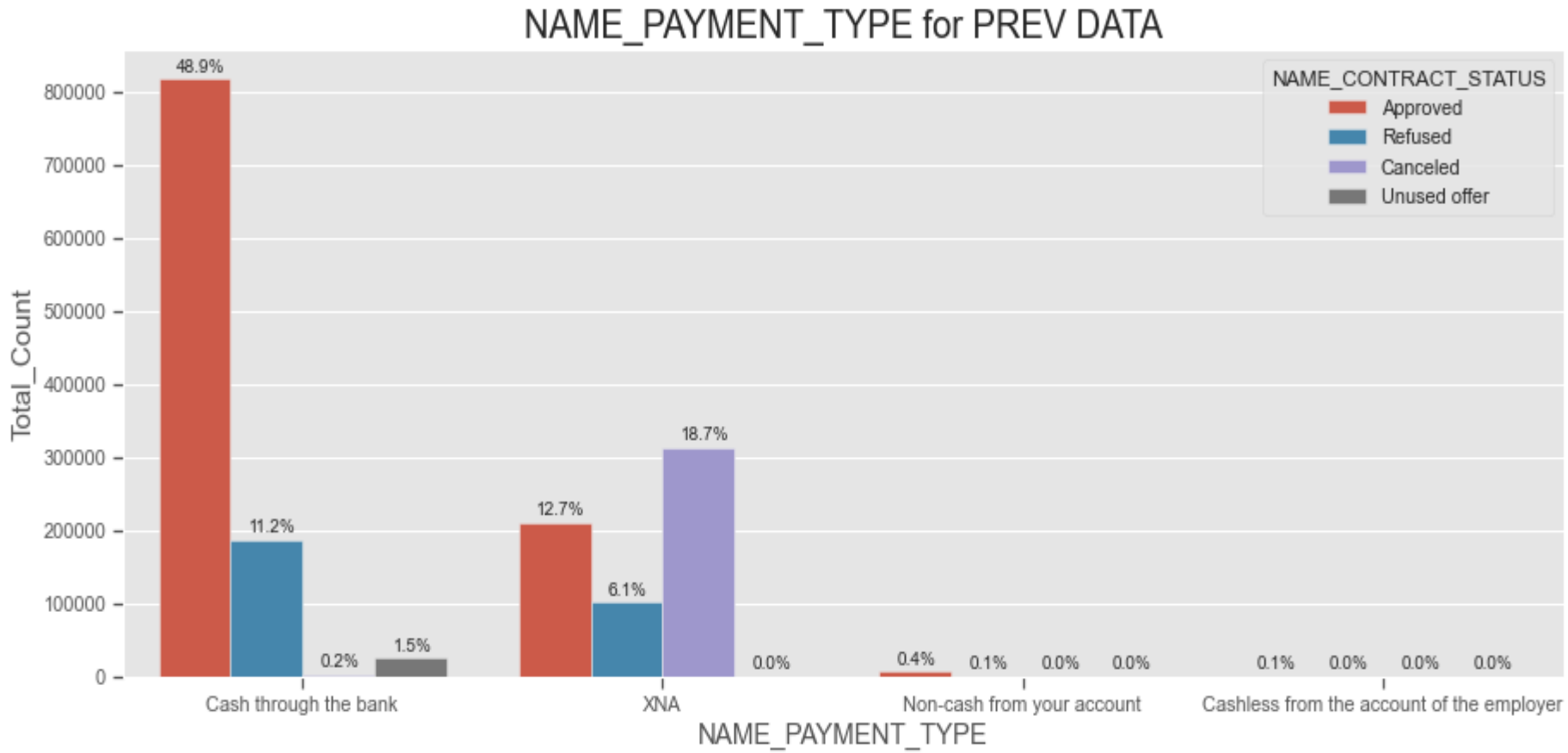
- *Points for Target 1*

1. Single and Married customers with academic degree have higher income and default rate is high.
2. Civil marriage customers tend to have higher total_income.
3. Single and Married are the only 2 family types present in academic degree .
4. Lower education clients have lower income and have more payment difficulties.

DATA ANALYSIS ON PREVIOUS APPLICATION DATA

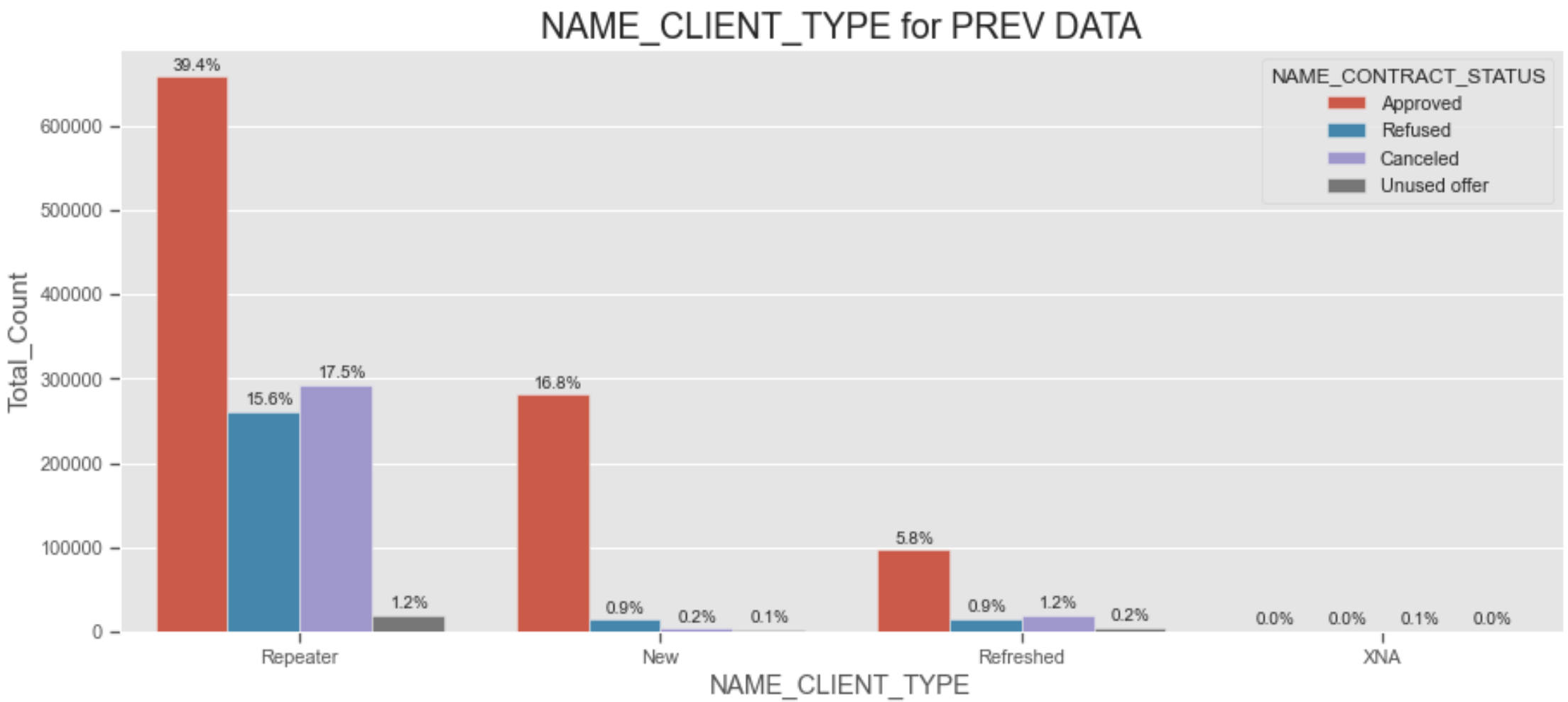
>> *UNIVARIATE ANALYSIS*

A. PAYMENT METHOD



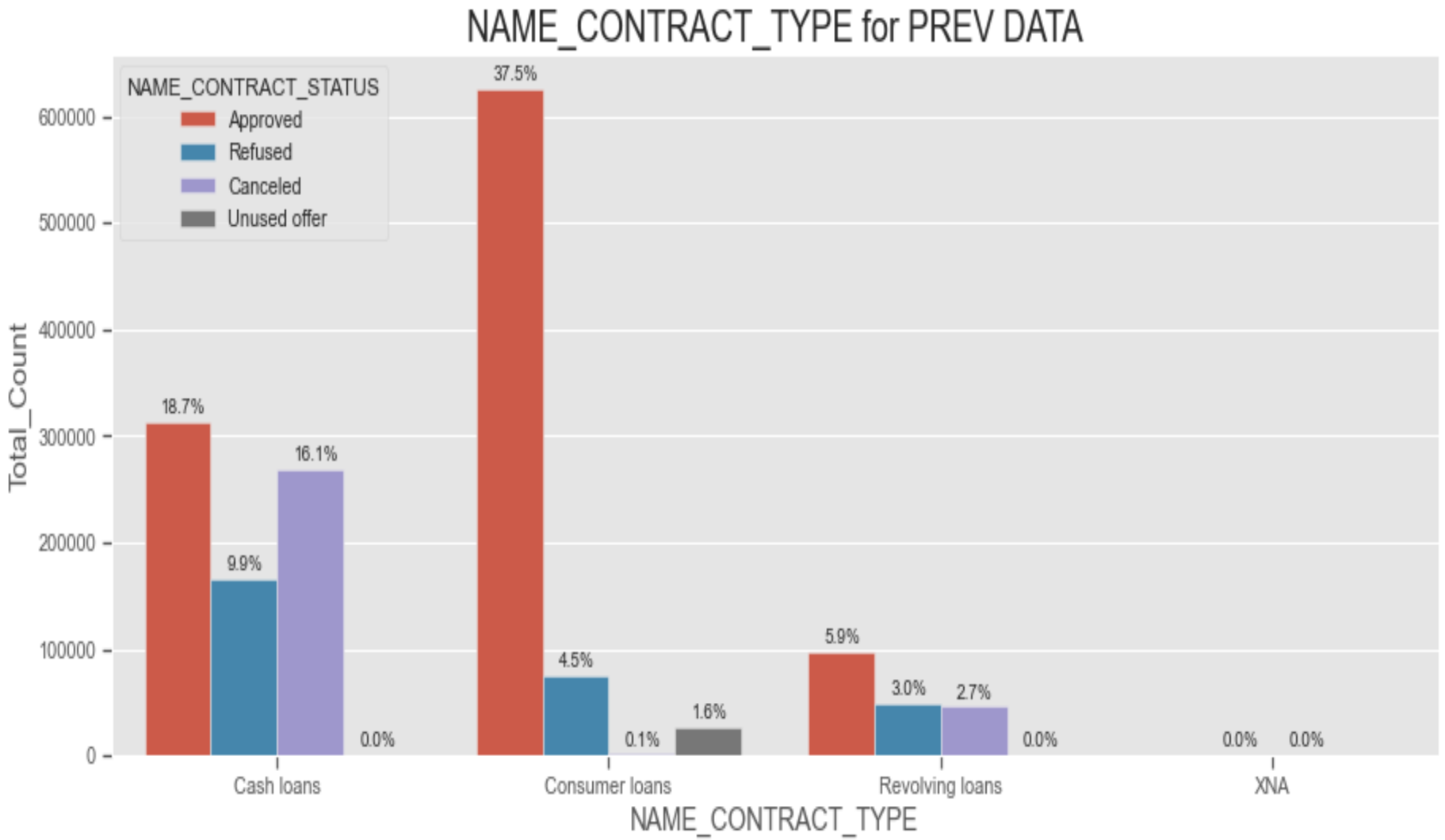
Most of the clients chose to pay cash through bank.

B. CLIENT TYPE :



Majority of the clients are repeaters.

C. CONTRACT TYPE :



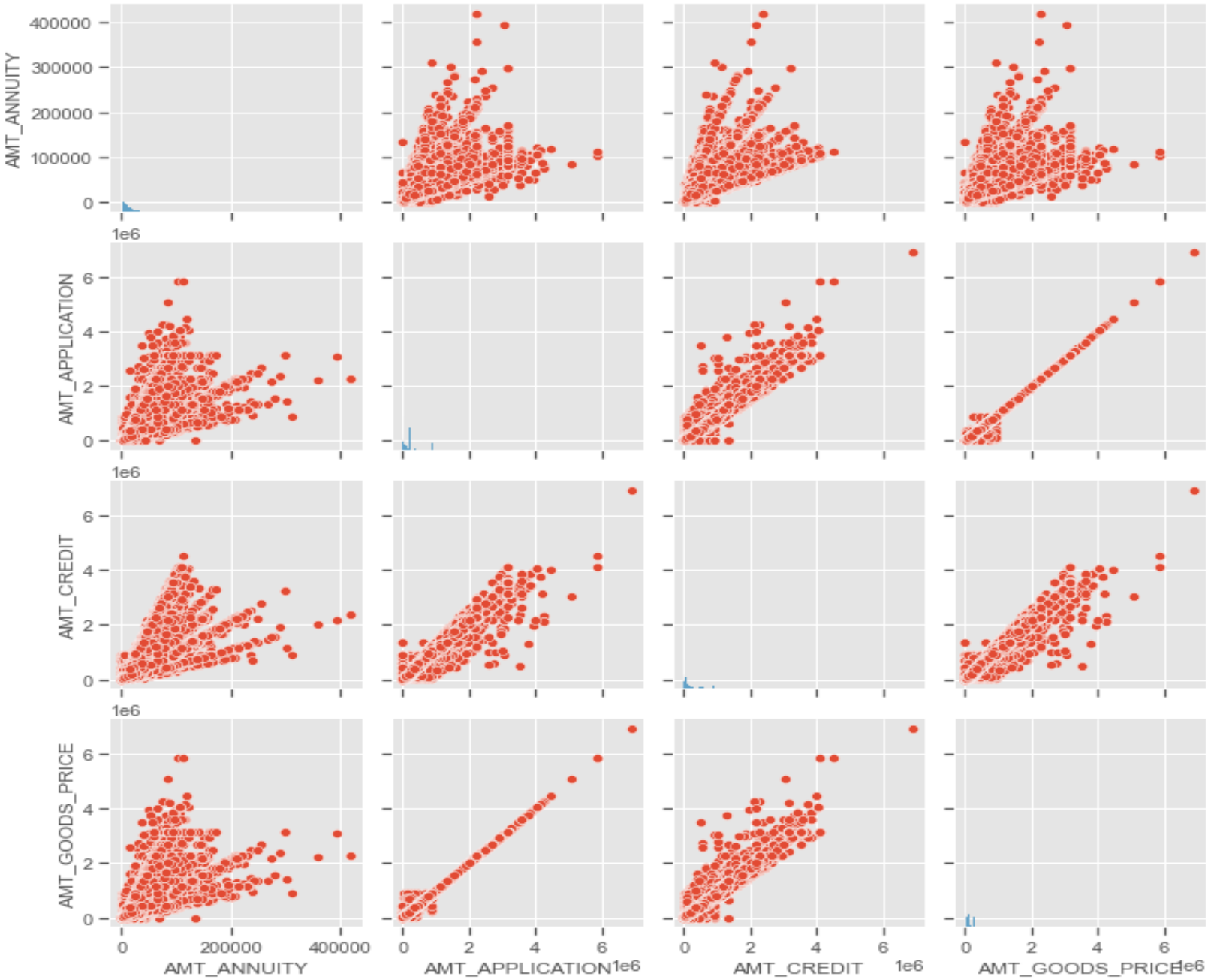
- Number of cash loans and consumer loans were comparable.
- More percentage of consumer loans were accepted.
- Cash loan cancellation was more.

CORRELATION MATRIX FOR PREVIOUS APPLICATION DATA

THESE ARE THE TOP 10 CORRELATION CATEGORIES FOR PREVIOUS DATA.WE CAN SEE THE SAME IN ABOVE MATRIX.

	level_0	level_1	0
58	AMT_GOODS_PRICE	AMT_APPLICATION	1.00
59	AMT_GOODS_PRICE	AMT_CREDIT	0.99
47	AMT_CREDIT	AMT_APPLICATION	0.98
57	AMT_GOODS_PRICE	AMT_ANNUITY	0.82
46	AMT_CREDIT	AMT_ANNUITY	0.82
35	AMT_APPLICATION	AMT_ANNUITY	0.81
113	CNT_PAYMENT	AMT_APPLICATION	0.68
114	CNT_PAYMENT	AMT_CREDIT	0.67
115	CNT_PAYMENT	AMT_GOODS_PRICE	0.67
112	CNT_PAYMENT	AMT_ANNUITY	0.39

HIGHLY CORRELATED COLUMNS TAKEN FOR BIVARIATE ANALYSIS



- ***Annuity of previous application has proportional relationship with these factors.***

1.Amount of credit client asked on the previous application

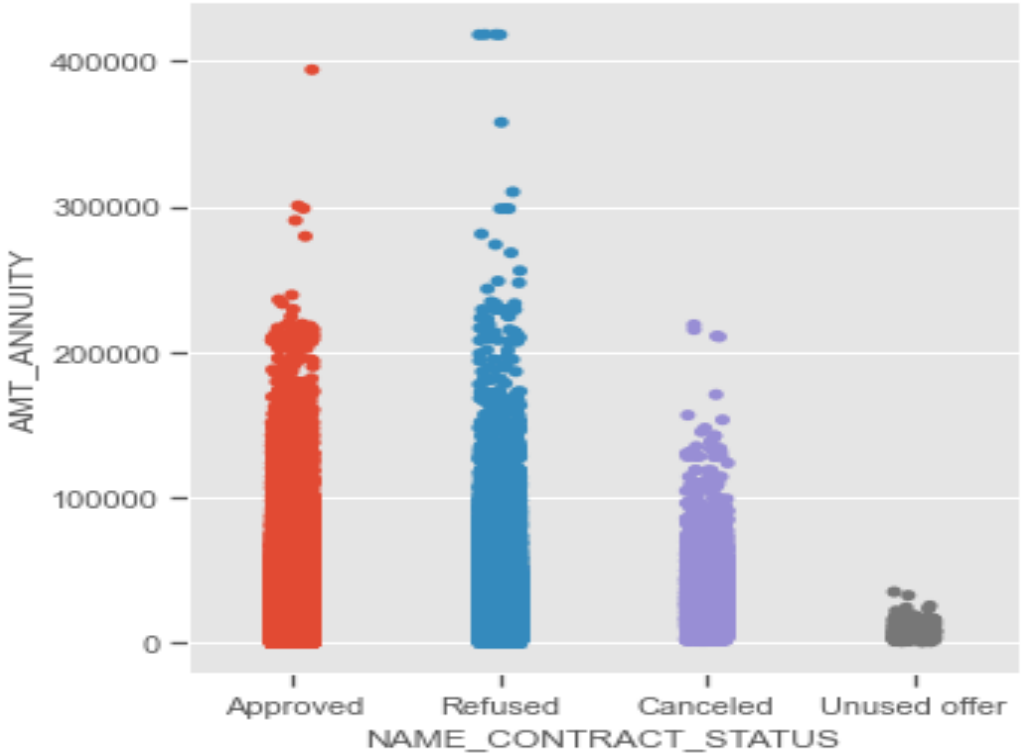
2.Amount that was approved by the bank.

3.Goods Price of the product against which loan was provided

- ***Credit asked by client has positive influence by goods price of the customer.***
- ***Credit given to customer is proportional to amount asked by customer and goods price of the customer.***

BIVARIATE ANALYSIS ON CATEGORICAL VS NUMERIC COLUMNS

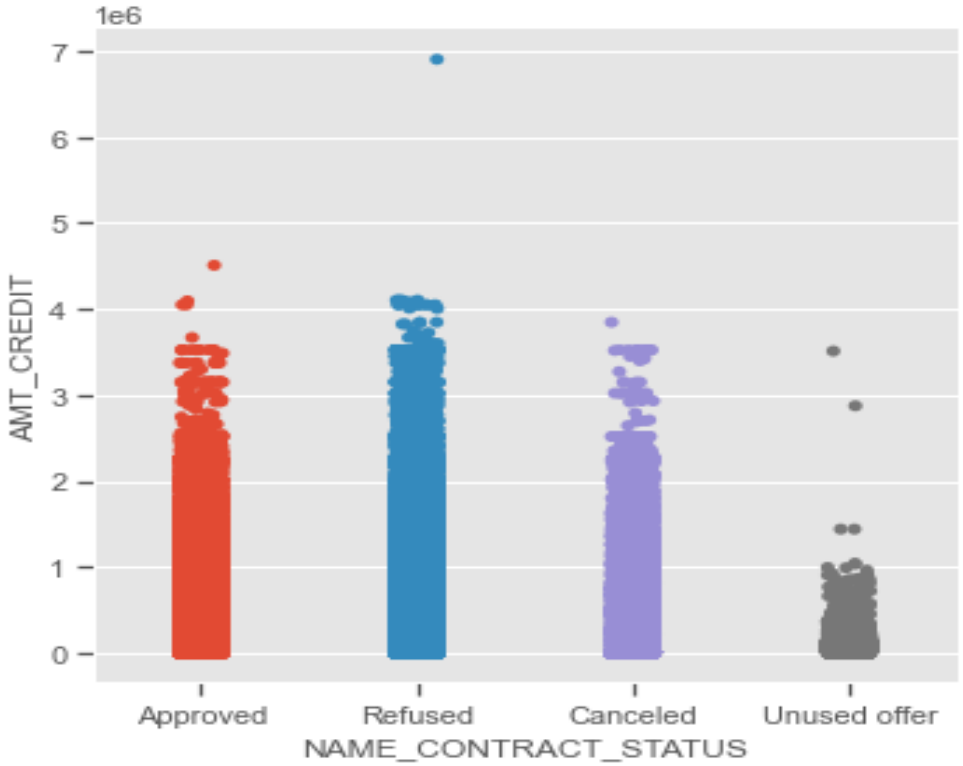
A. NAME_CONTRACT_STATUS' VS 'AMT_ANNUITY



•Client provided with high annuity have high refusal rate.

•Cancelled applicants have lower amount of annuity.

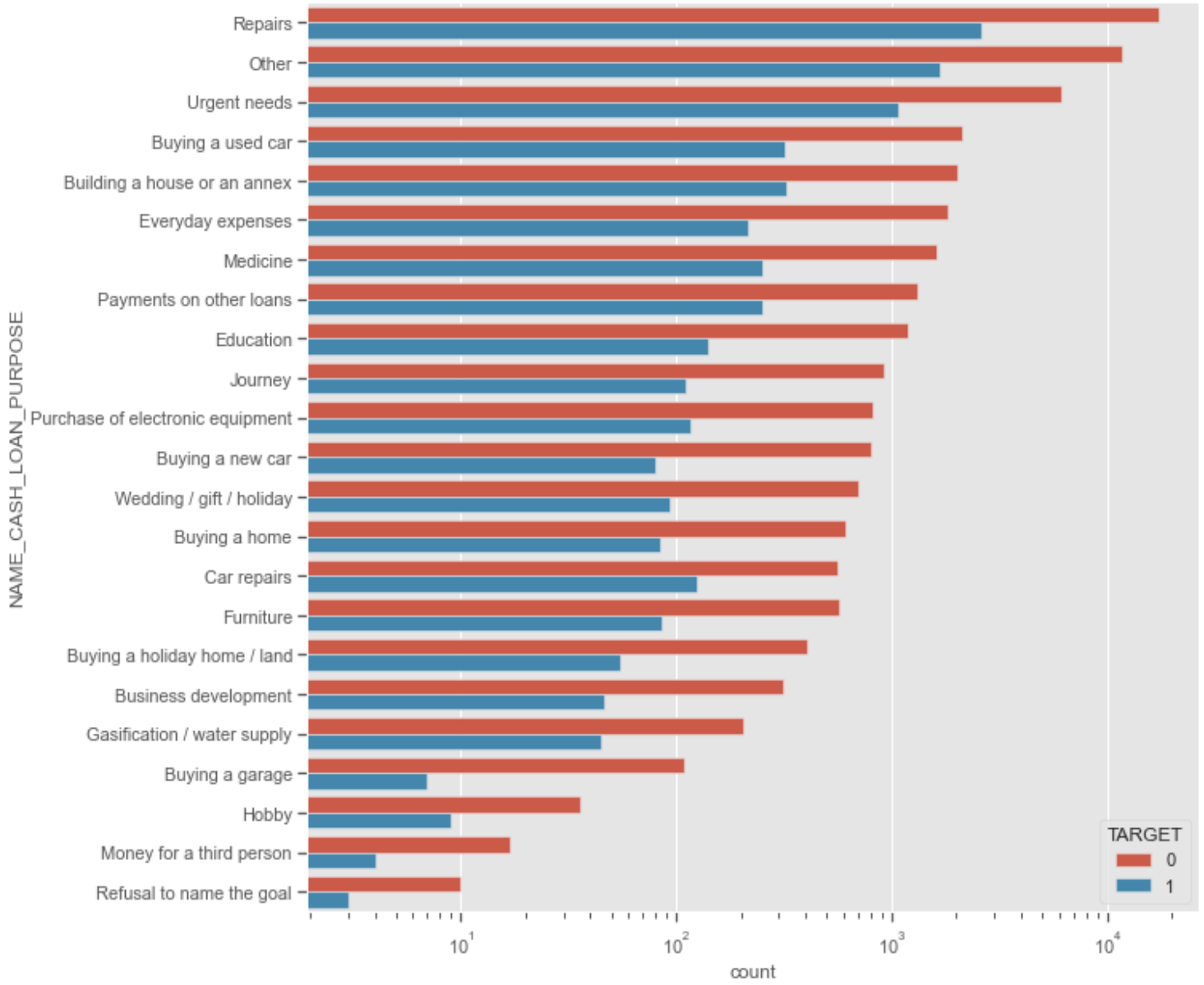
B. NAME_CONTRACT_STATUS' VS 'AMT_CREDIT'



•Lower amount credit has more chance of cancelled or unused status.

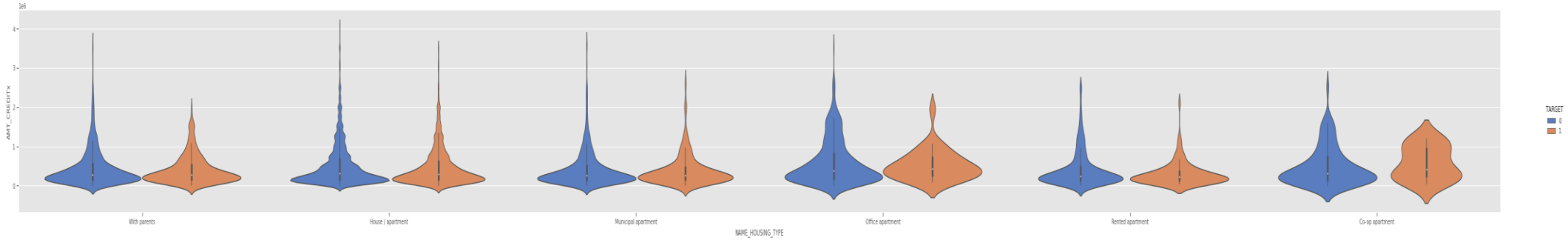
•Amount credit does not have high relationship with approved or refused status.

MERGING THE APPLICATION DATASET WITH PREVIOUS APPLICATION DATASET



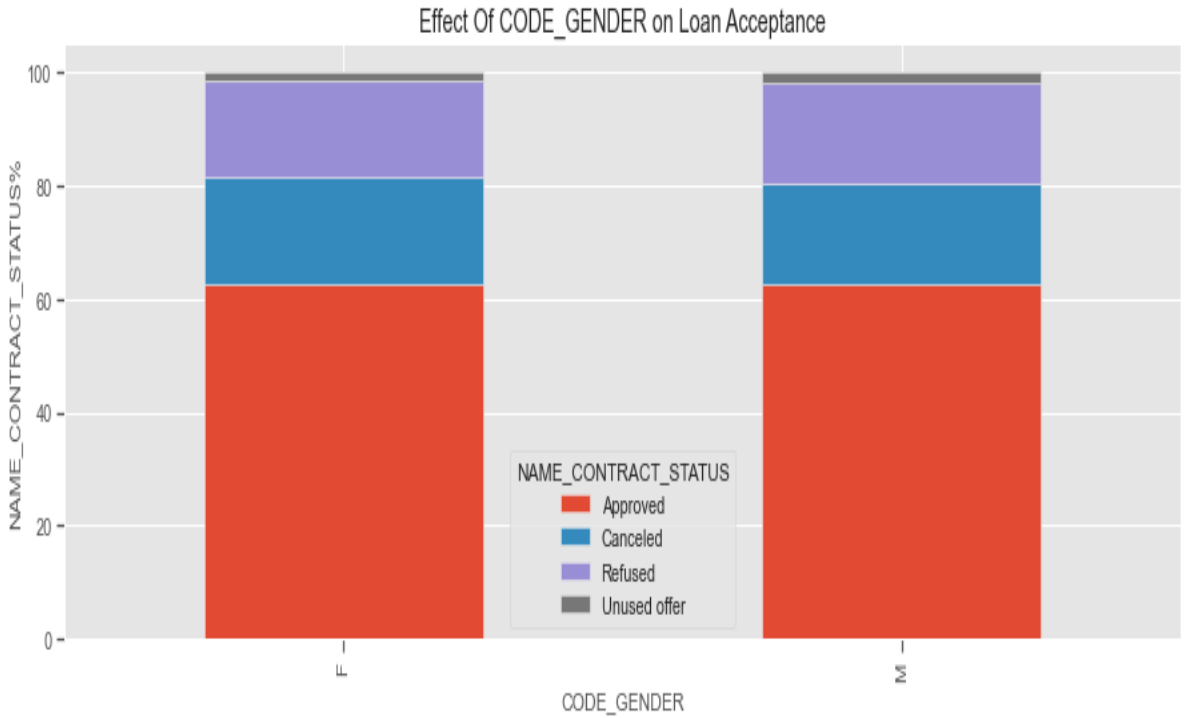
For target = 0 , people use the Loan Purpose for "Repairs", same with target =1, however the percentage is less for defaulters.

>>CREDIT AMOUNT PREV VS HOUSING TYPE



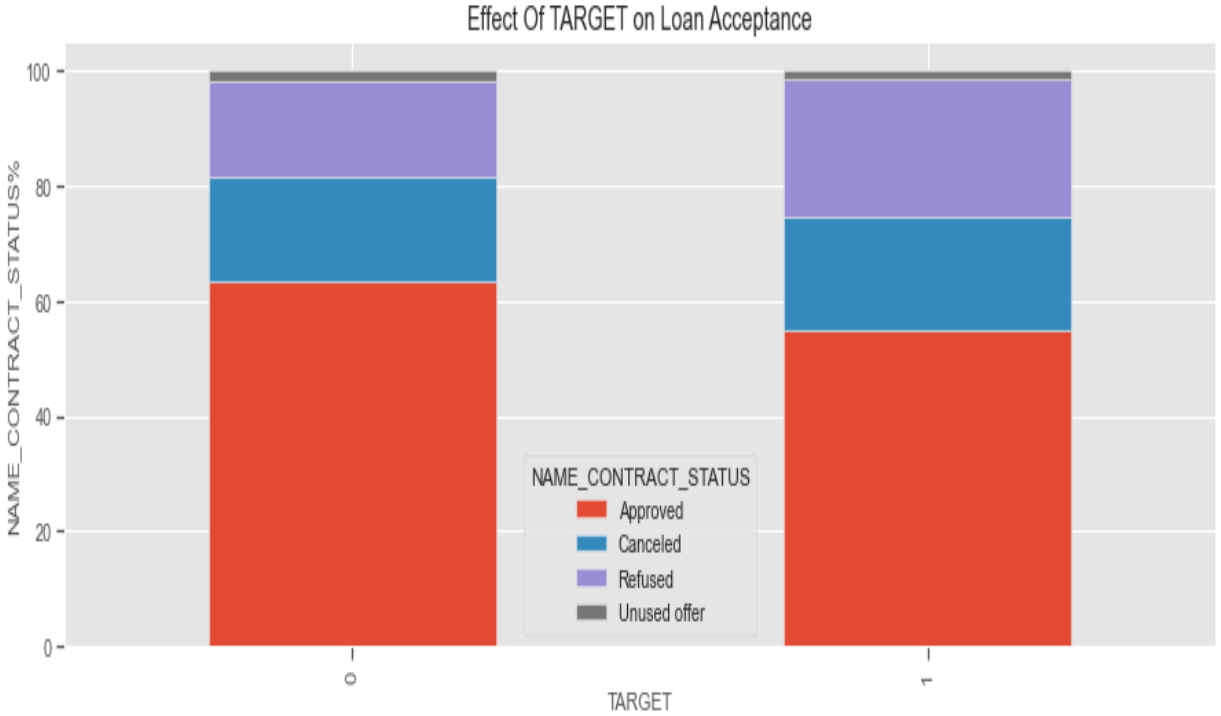
Here for Housing type, office apartment is having higher credit of target 0 and co-op apartment is having higher credit of target 1. So, we can conclude that bank should avoid giving loans to the housing type of co-op apartment as they are having difficulties in payment. Bank can focus mostly on housing type with parents or House/apartment or municipal apartment for successful payments.

EFFECT OF CODE GENDER ON LOAN ACCEPTANCE



- We don't see major difference in distribution between genders but while doing single variate analysis on gender, we saw females have less payment difficulties. So, banks can consider that while giving loans.

EFFECT OF TARGET ON LOAN ACCEPTANCE



- We can see the clients who had gotten previous apporoval have less payment difficulties.
- The clients whose previous loan was refused have more problems in payment of loans.

INSIGHTS

- *Application Data*
- GENDER: *Univariate analysis shows Male Percentage increases when we go from non defaulter to defaulter category.

  * Bank should give females more weightage when considering loans.
- Region Rating: * Univariate analysis shows, region rating of '3' indicates clients percentage of defaulting increases.

  * Bank should be careful providing loans to client with rating '3'.
- Years Worked: * Clients with less than 10 years of experience have more difficulties in loan payments.

  * As years worked increases, clients show better loan paying trend.
- Count of Children: * Univariate analysis shows, as the number of children increase payment difficulties increases.
- Loan/Income: * A value of >50 for loan to income shows default rate will be high.

  * A lower income is also indicative of problem in payment.

- Age: * Clients in <40 age bracket have more problems in repayment of loan.
  As they get older reliability is more.

- Education: * Clients with secondary education and lower secondary have more issues in repayment as it also means a lower total income.

- Housing: * Clients living with parents and in rented houses are more likely to be default.

- *Previous Application Data*

* Most of the customers chose the option "Payment through bank"
* Many repeaters applied for loan.
* New applicants are less but their acceptance rate is better.
* Rate of acceptance of consumer loans are better than cash loans.
* Credit given to customer is proportional to goods price of the customer.
* Repairs are the reason for most amount of loans and they also face more payment difficulties.

- *For the Merged Data:*

- The clients who had gotten previous approval have less payment difficulties so they are better candidates for approval of loan.
- Clients whose application was cancelled are more risk to default.
- Banks should focus more on contract type ‘Student’ ,’Pensioner’ and ‘Businessman’ with housing type other than ‘Co-op apartment’ for successful payments.
- Buying a home and land purposes have a high cancellations.