CREDIT EDA CASE STUDY

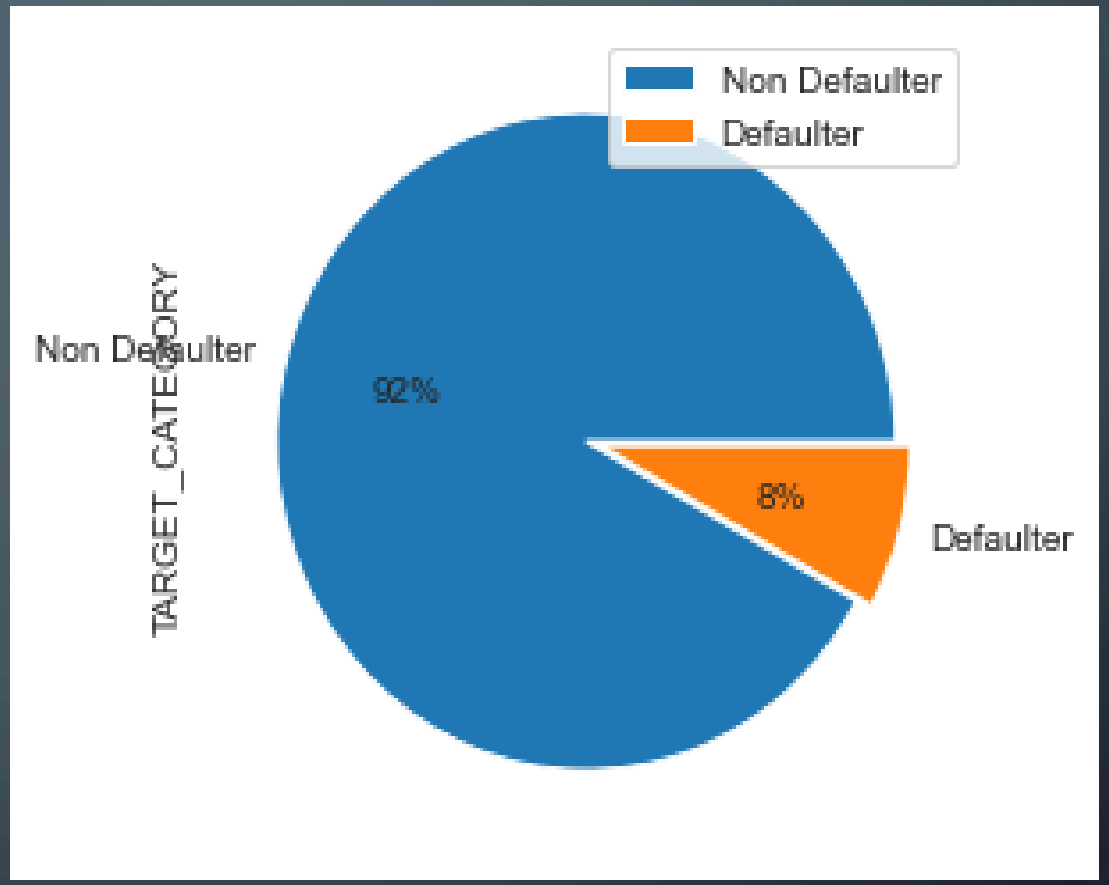
BY PRATIK CHOUHAN

Steps:

- Reading Dataset
- check Missing values which to handle how to handle
- Check outlier ,check data imbalance ratio
- Univariate analysis ,Segmented Univariate analysis and Bivariate analysis ,correlation
- Merging of application data with previous data
- Data analysis by univariate segmented bivariate analysis
- Recommendatio and risks

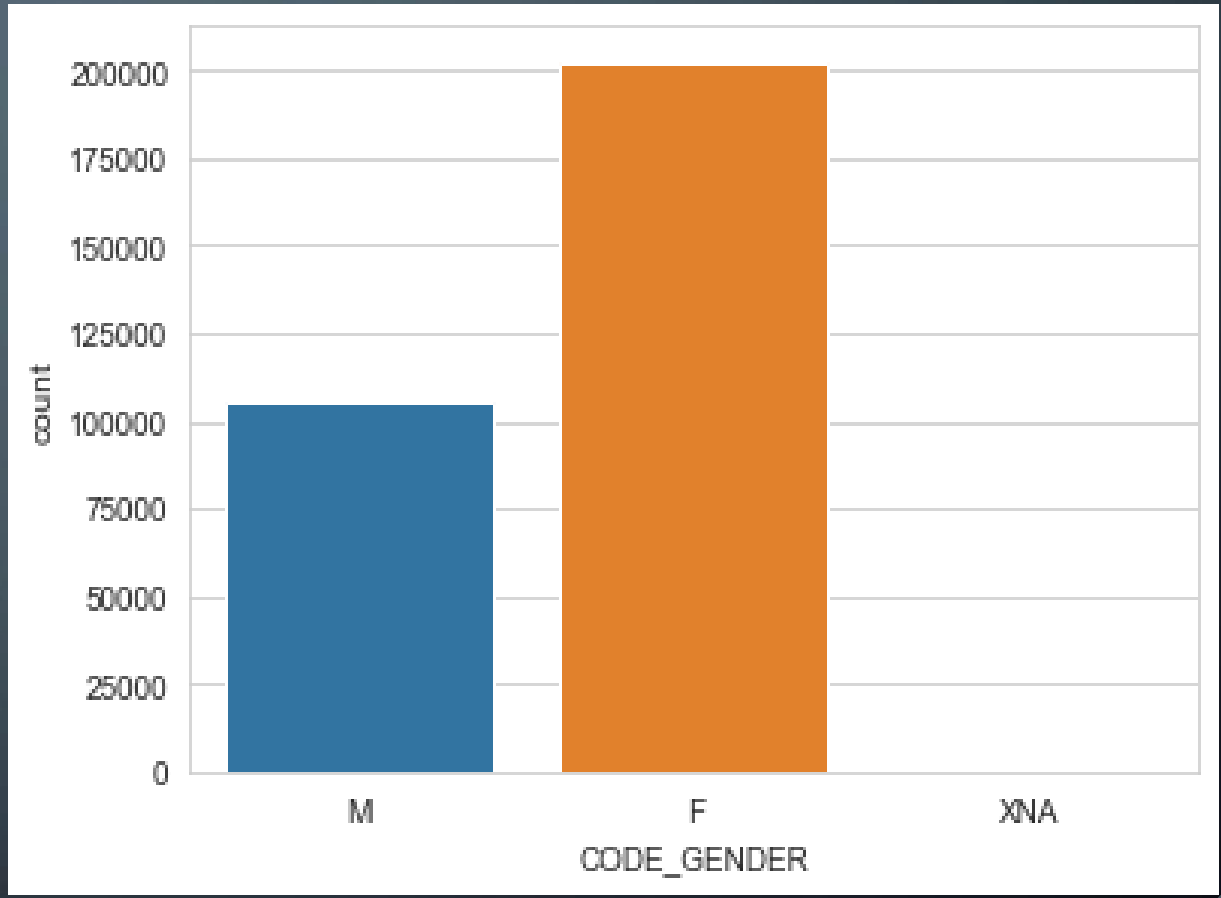
TARGET VARIABLE

- Defaulter are 8% of total
- Nnn Defaulter are 92% in total



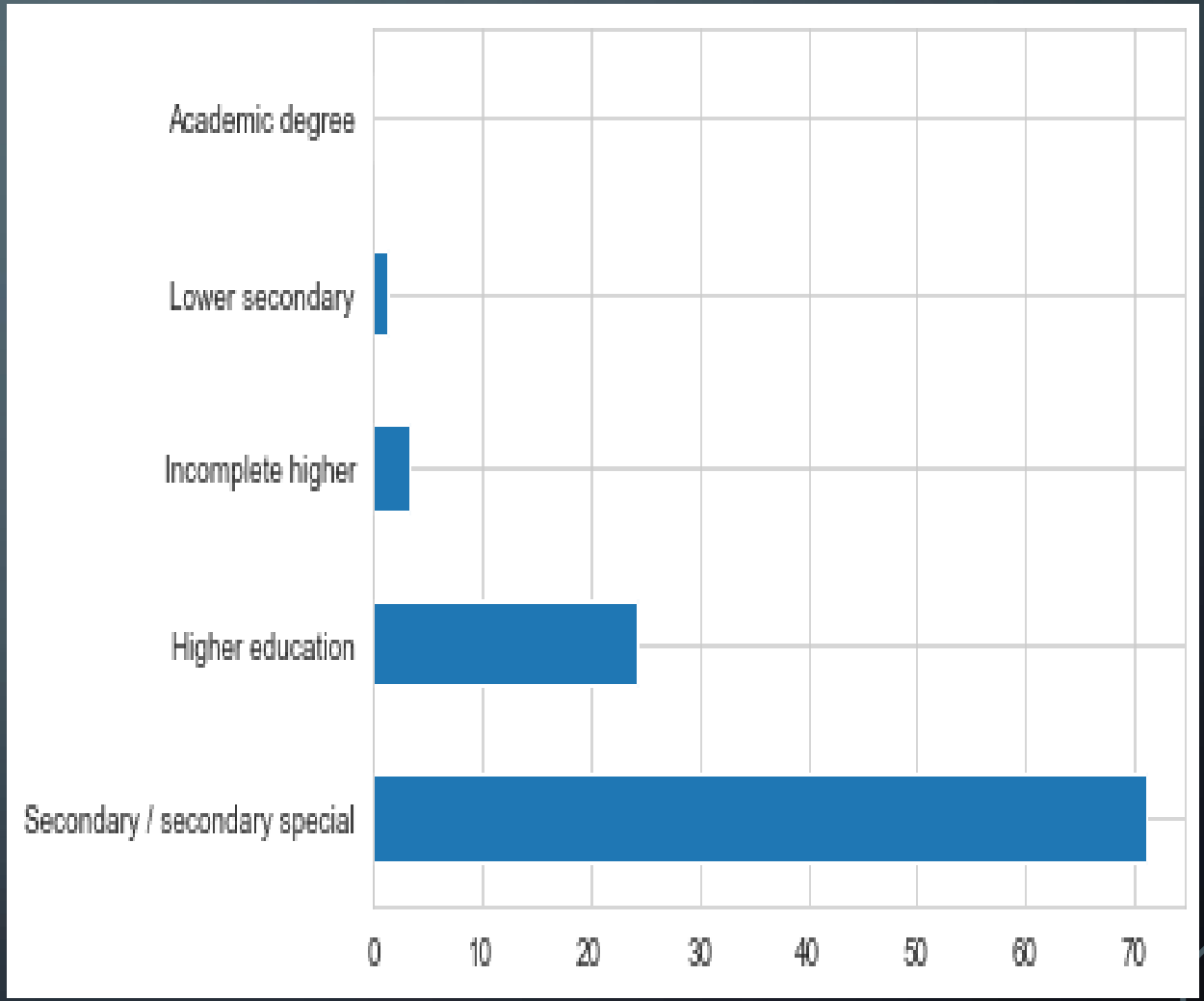
CODE GENDER

- Female is higher than male in application dataset



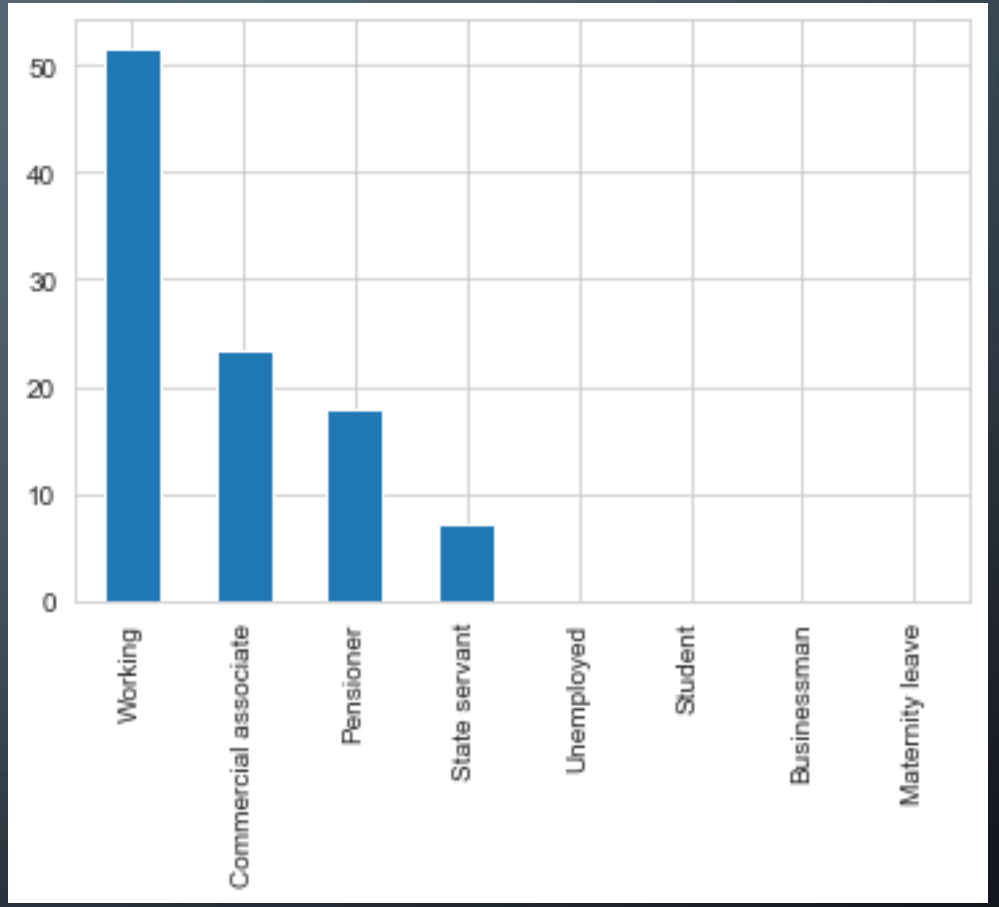
EDUCATION TYPE

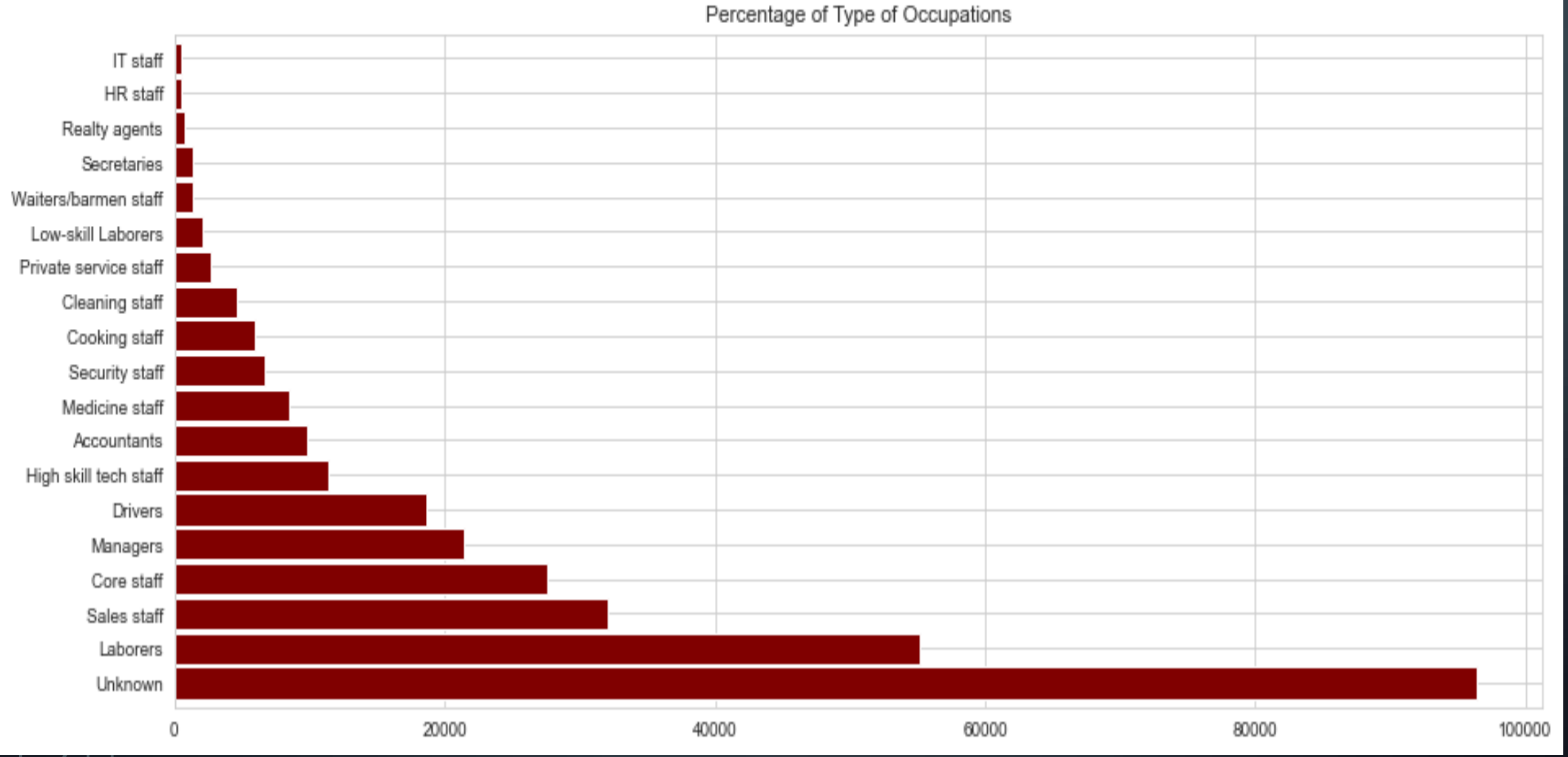
- Number of secondary education is higher than other



NAME INCOME TYPE

- Number of working profession high

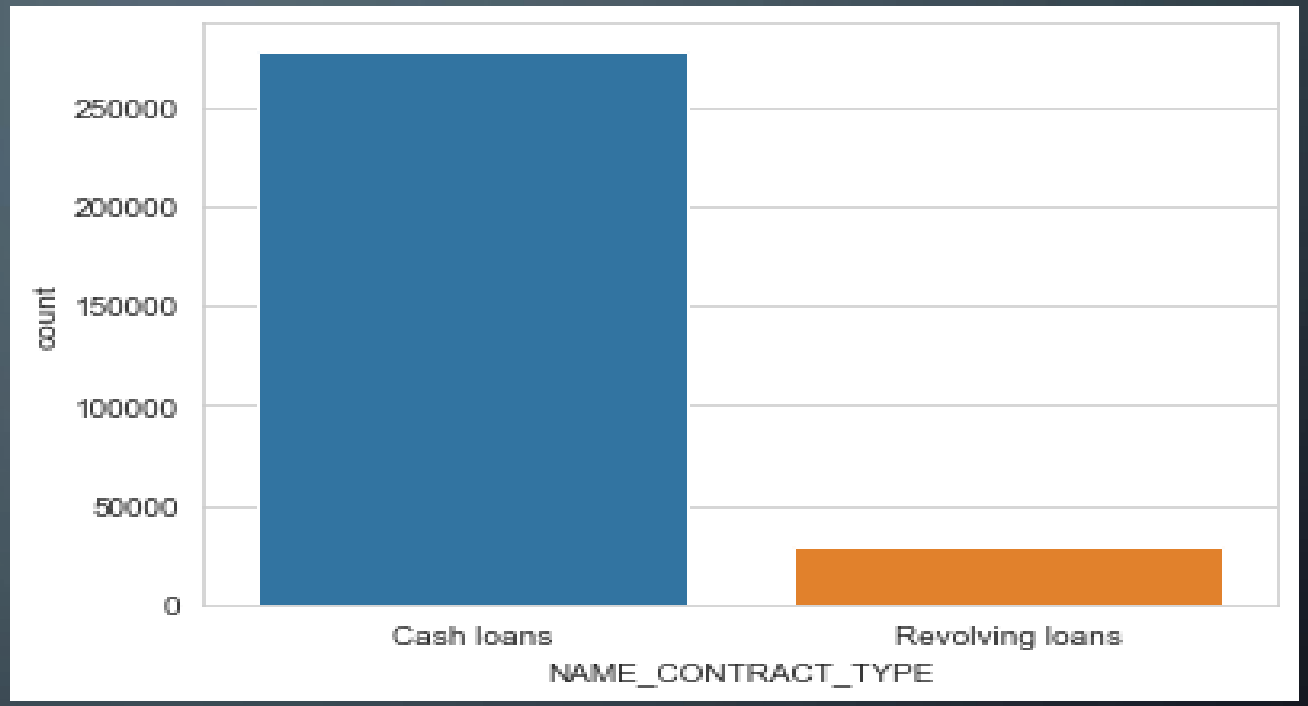




- - OCCUOATION TYPE
- - **Highest percentage of values belongs to Unknown group and Seconds belongs to Laborers**

NAME CONTRACT TYPE

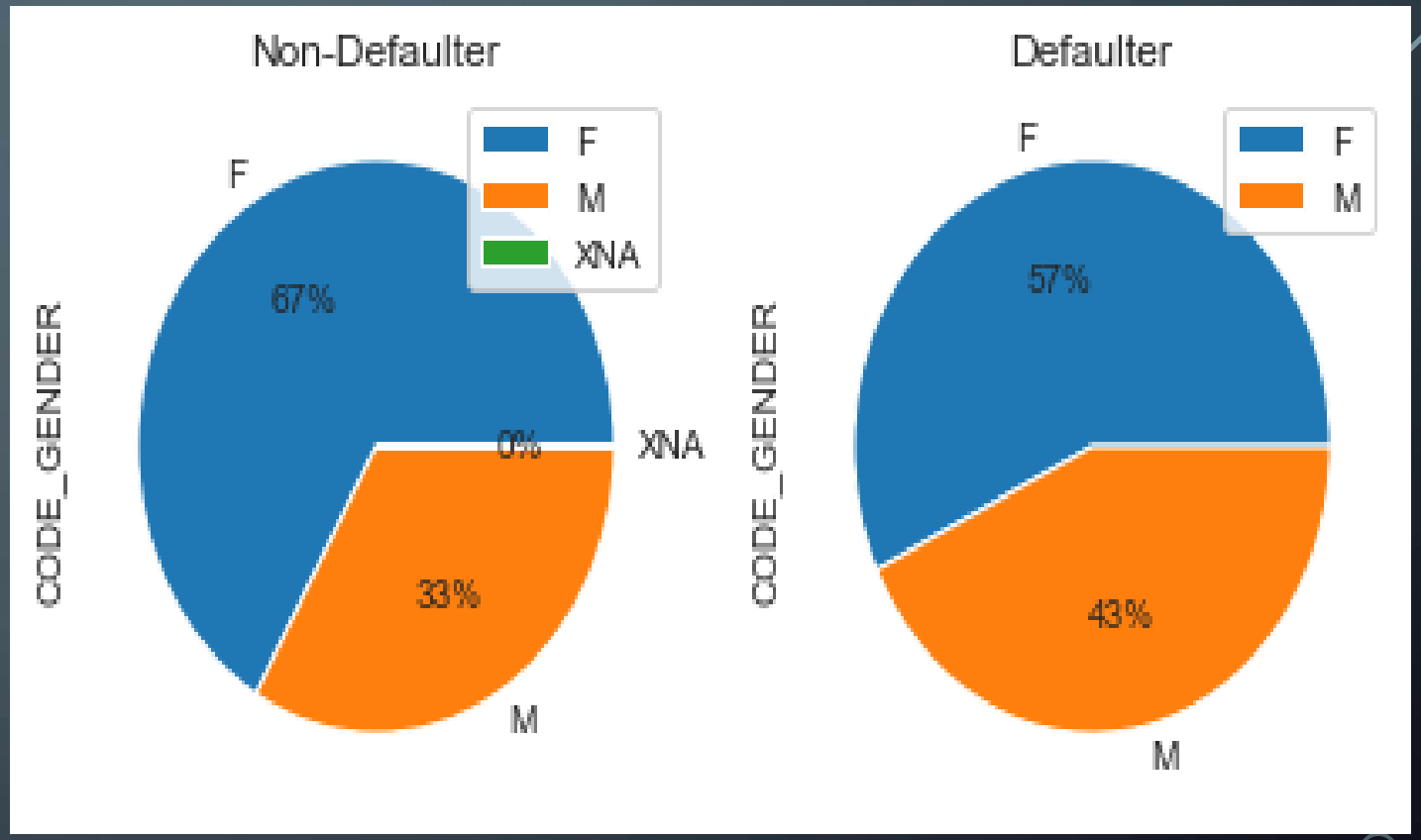
- Cash loan high than revolving loan



- SEGMENTED UNIVARIATE ANALYSIS
- TARGET 1 - DEFAULTER
- TARGET 0 - NON DEFAULTER

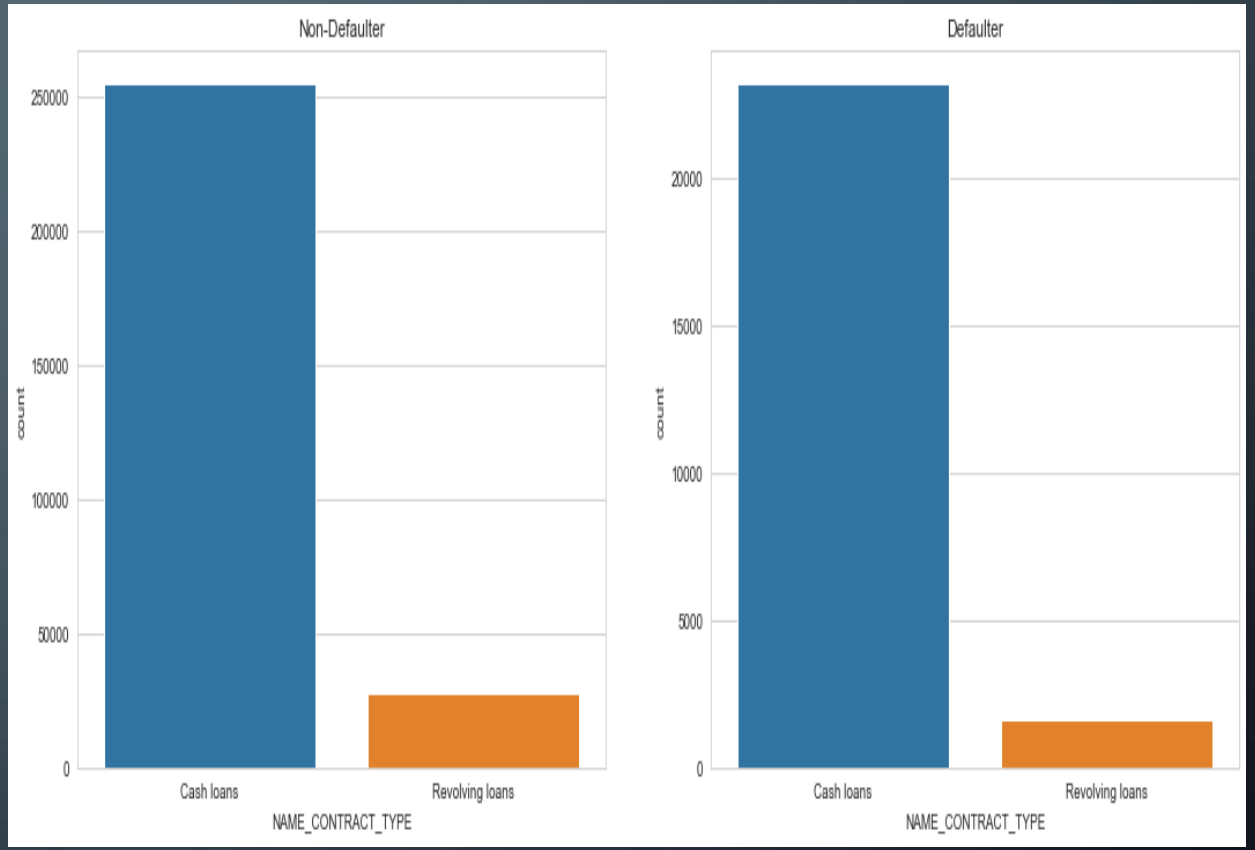
PROPORTION OF DEFAULTER BY GENDER

- Female in defaulter is higher



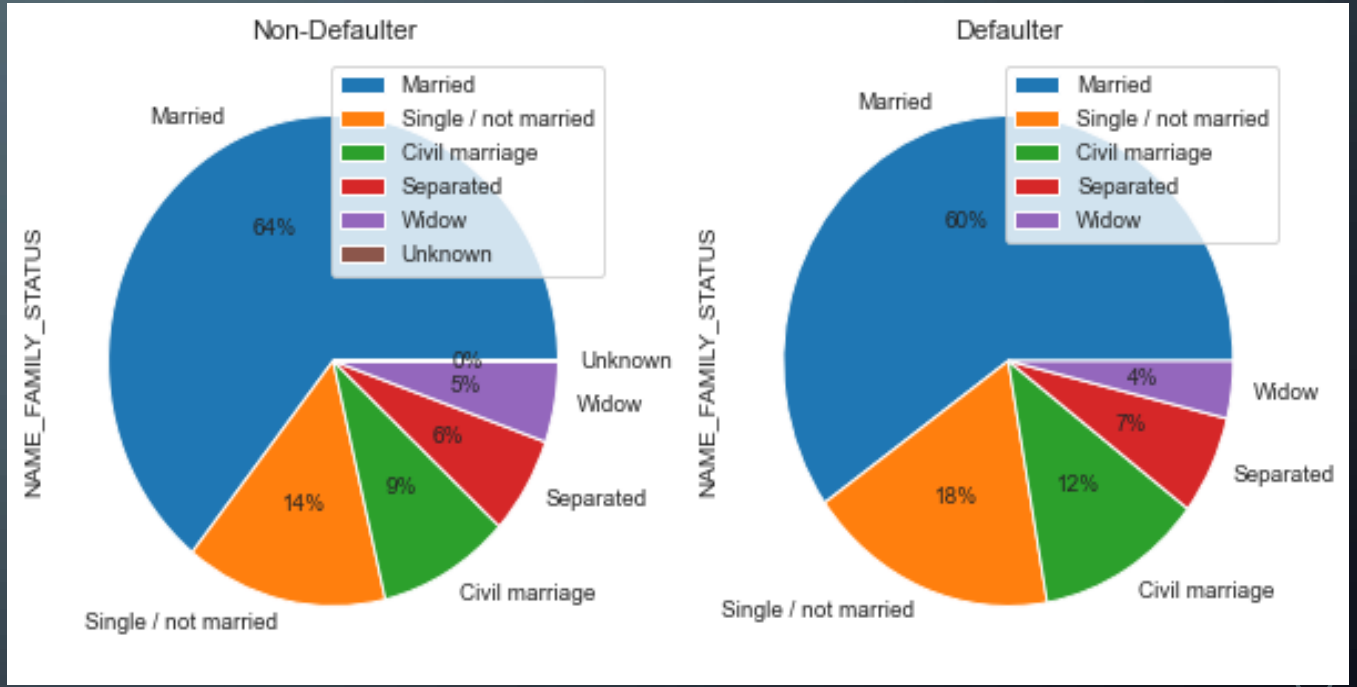
PROPOTION OF DEFAULTER BY PROFESSION

- Cash loan is high both the case Non-defaulter and Defaulter



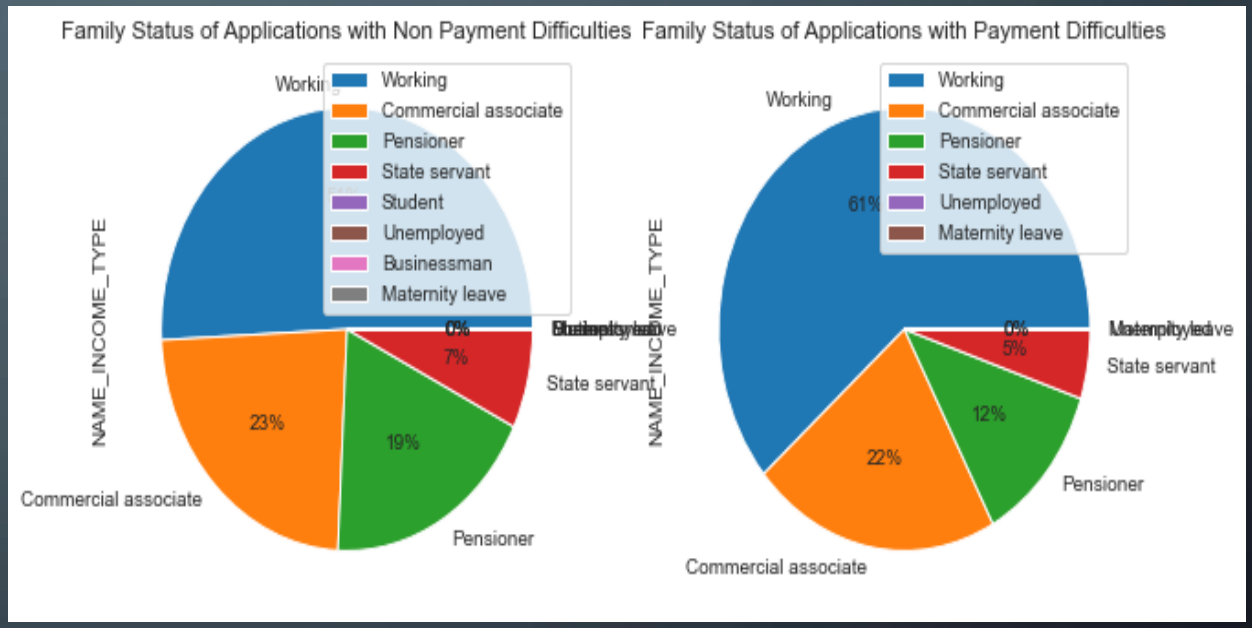
PROPOTION OF DEFAULTER BY PROFESSION FAMILY STATUS

-both the Married is high



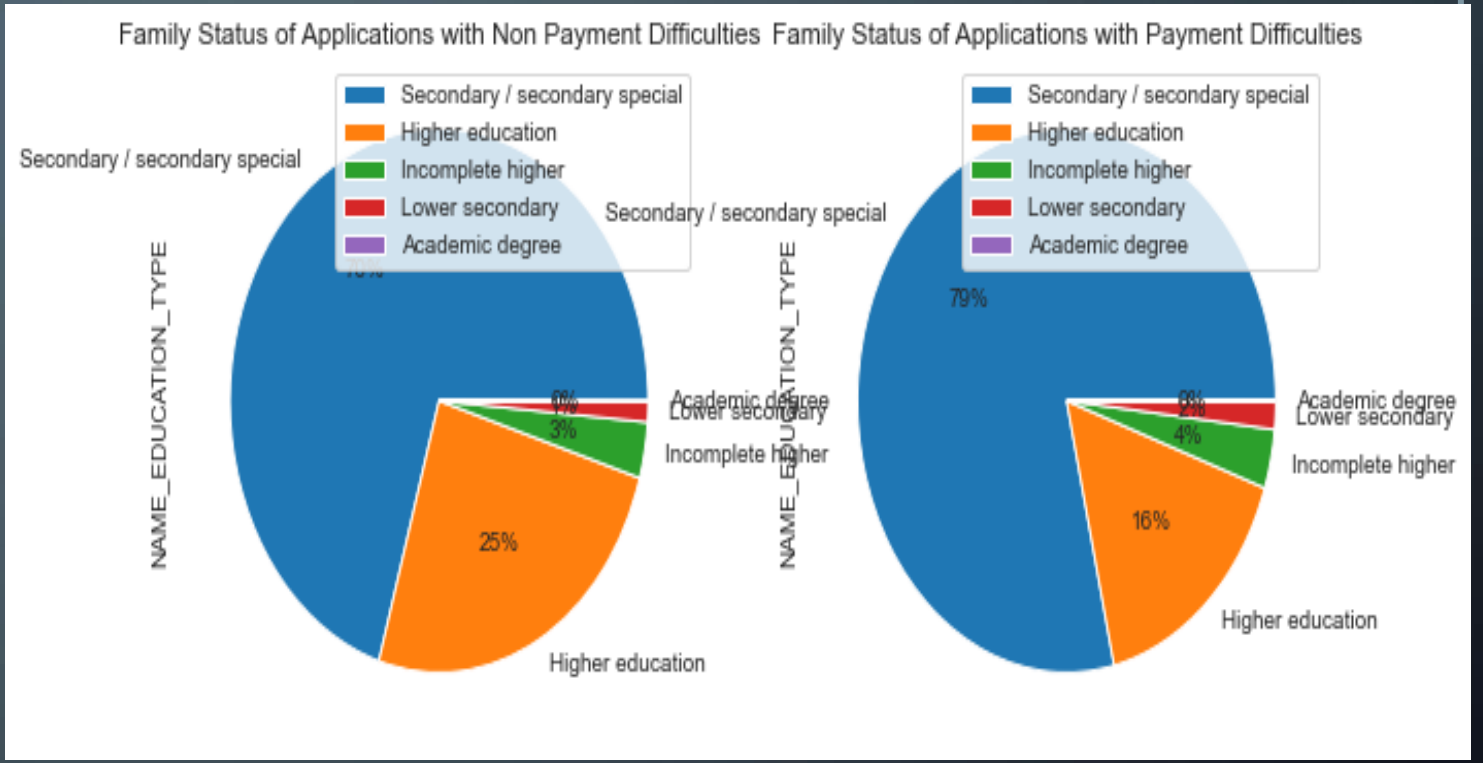
PROPOTION OF DEFAULTER BY PROFESSION INCOME TYPE

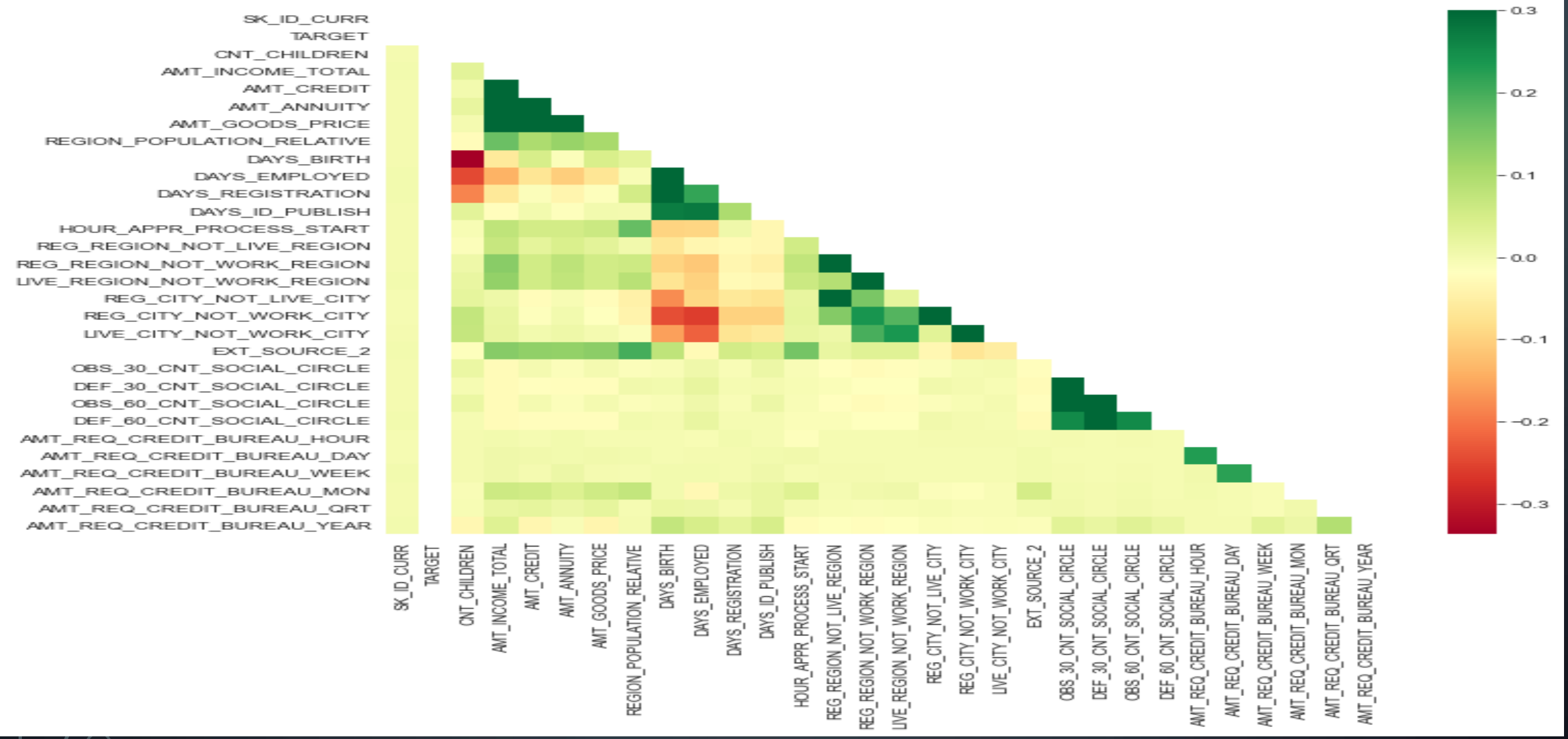
- Propotion working population in defaulter in higher and state servent is lower in defaulter



PROPOTION OF DEFAULTER BY PROFESSION EDUCATION TYPE

- Secondary /secondary special and higher education is default less

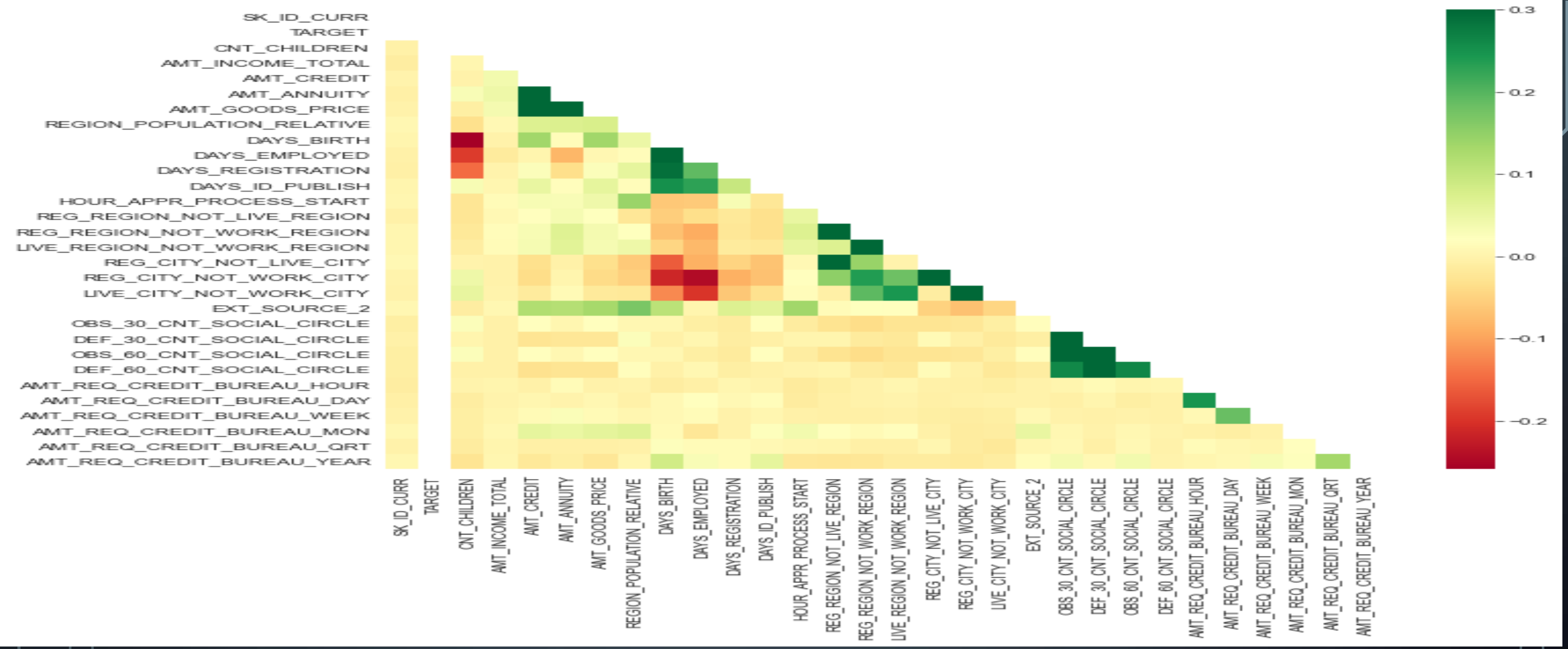




CORRELATION FOR NON DEFAULTER

CORRELATION FOR NON DEFAULTER

- Credit amount is inversely proportional to the date of birth, which means Credit amount is higher for low age and vice-versa.
- Credit amount is inversely proportional to the number of children client have, means Credit amount is higher for less children count client have and vice-versa.
- Income amount is inversely proportional to the number of children client have, means more income for less children client have and vice-versa.
- less children client have in densely populated area.
- Credit amount is higher to densely populated area.
- The income is also higher in densely populated area.



CORRELATION DEFAULTER

CORRELATION DEFAULTER

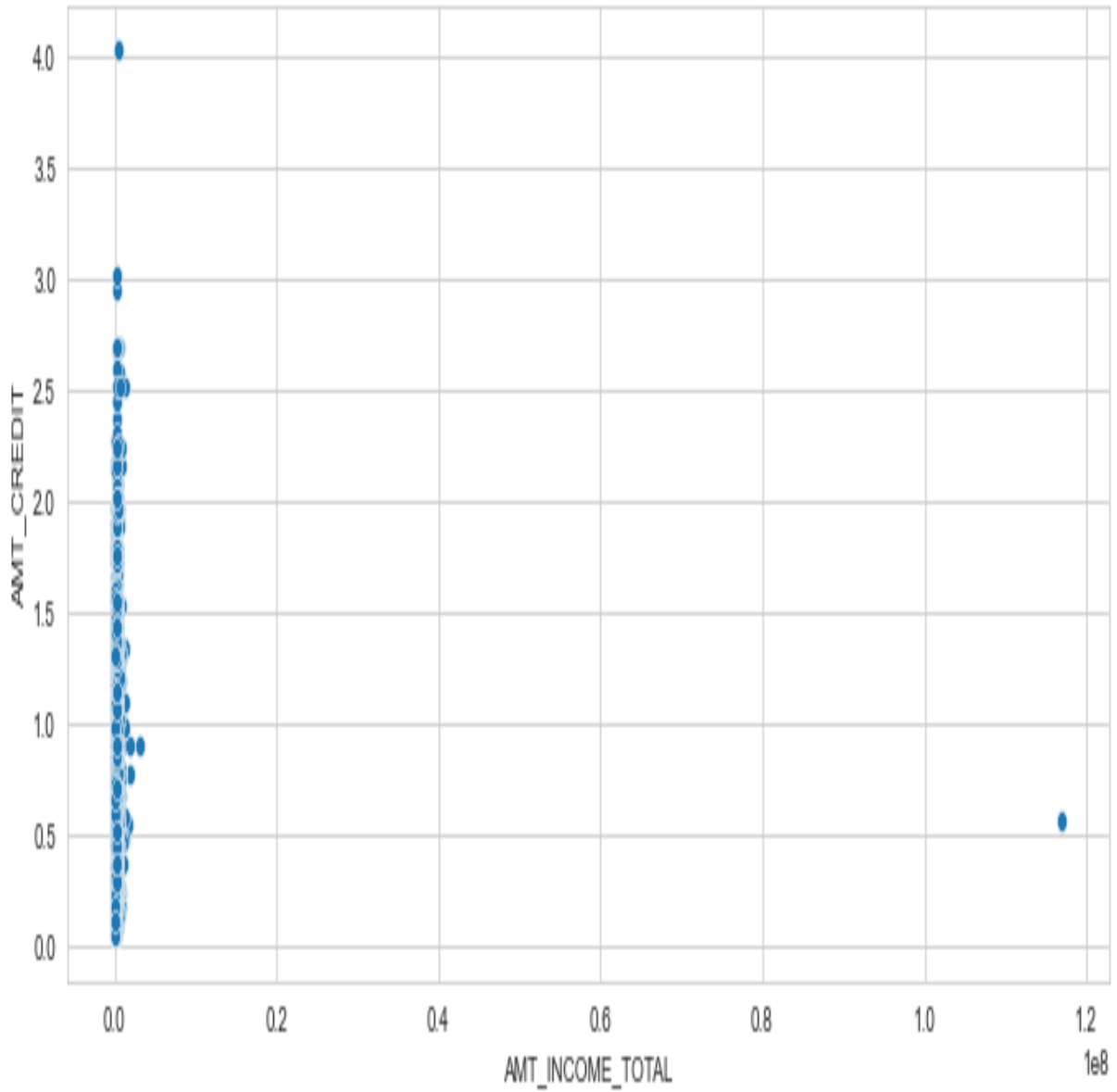
This heat map for Target 1 is also having quite a same observation just like Target 0. But for few points are different. They are listed below

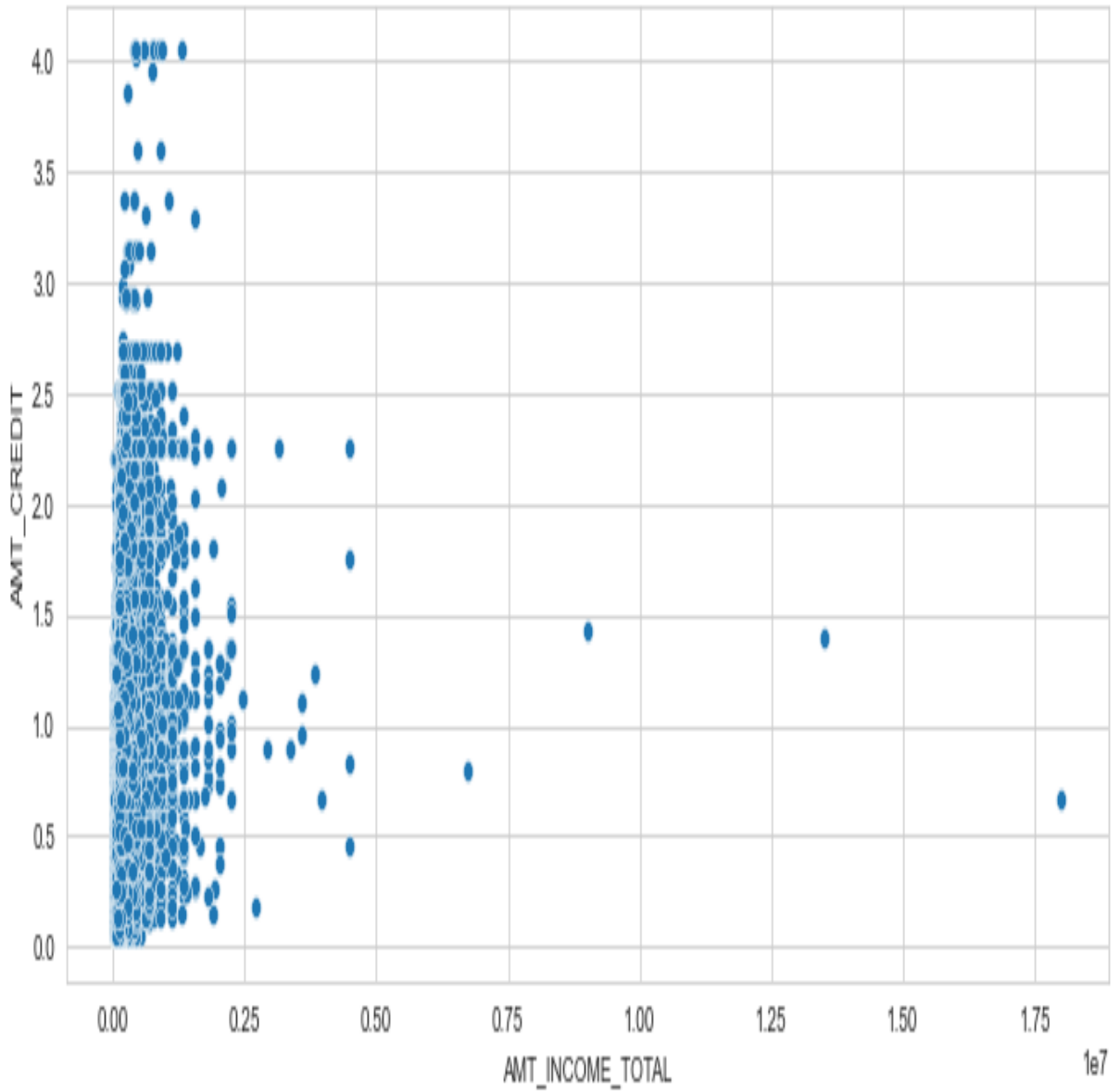
- The client's permanent address does not match contact address are having less children and vice-versa

- The client's permanent address does not match work address are having less children and vice-versa

BIAVARIATE ANALYSIS¶

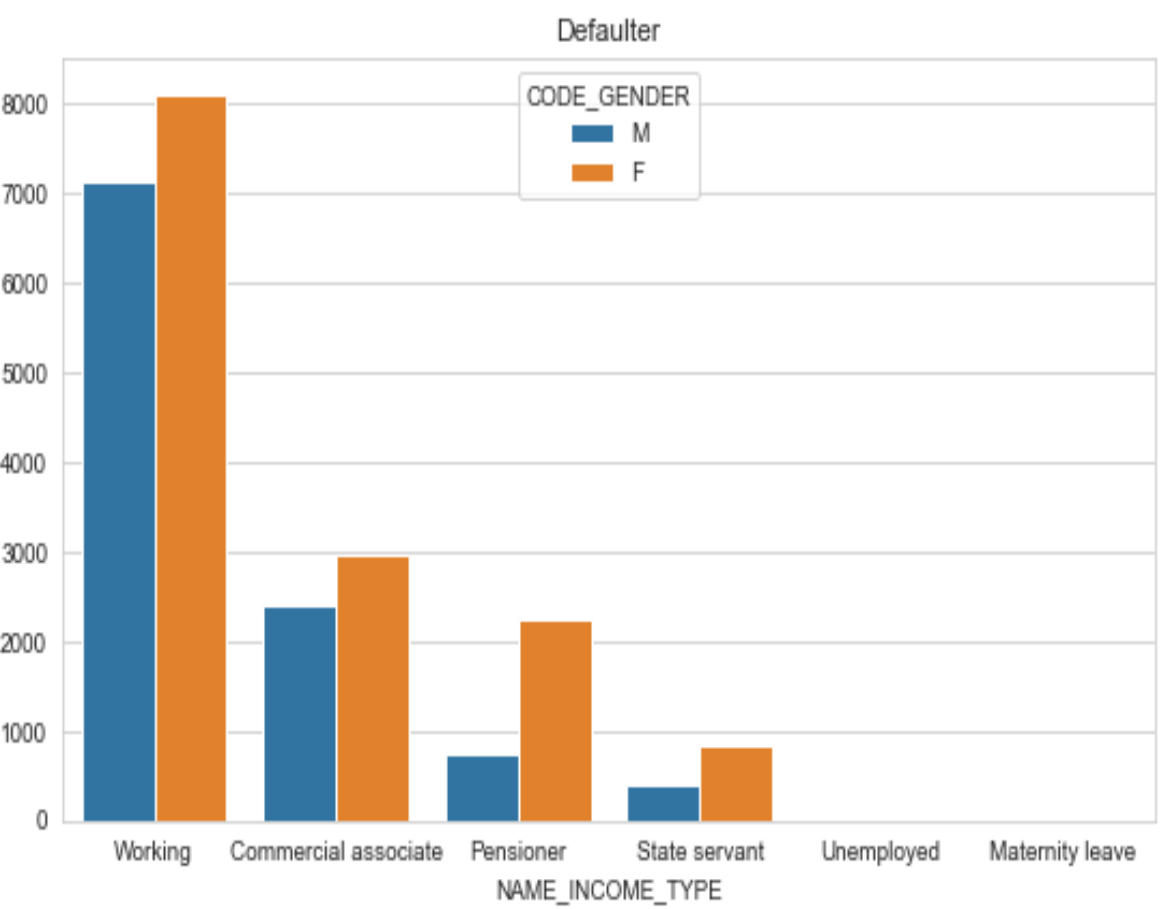
- CONTINUOUS- CONTINUOUS

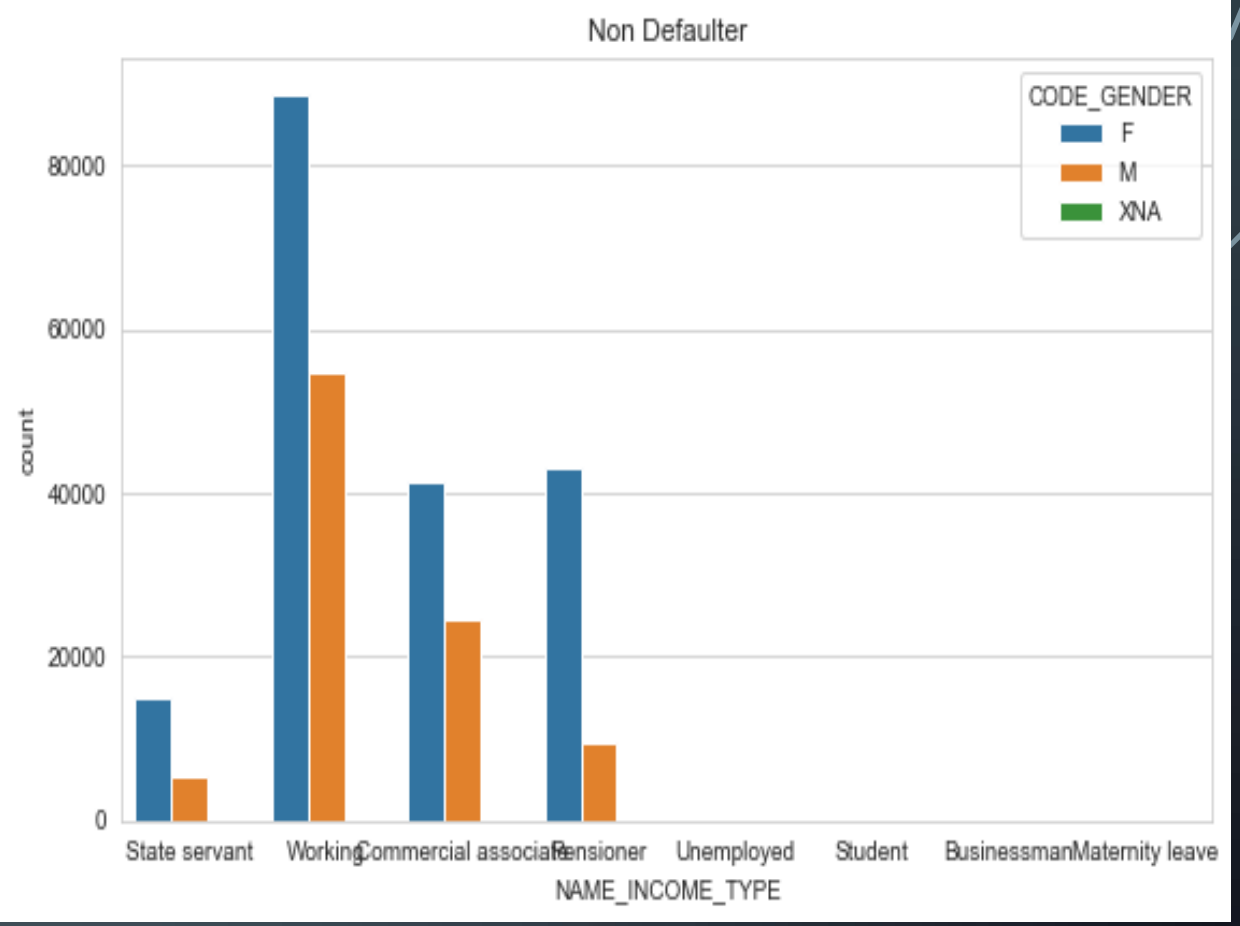




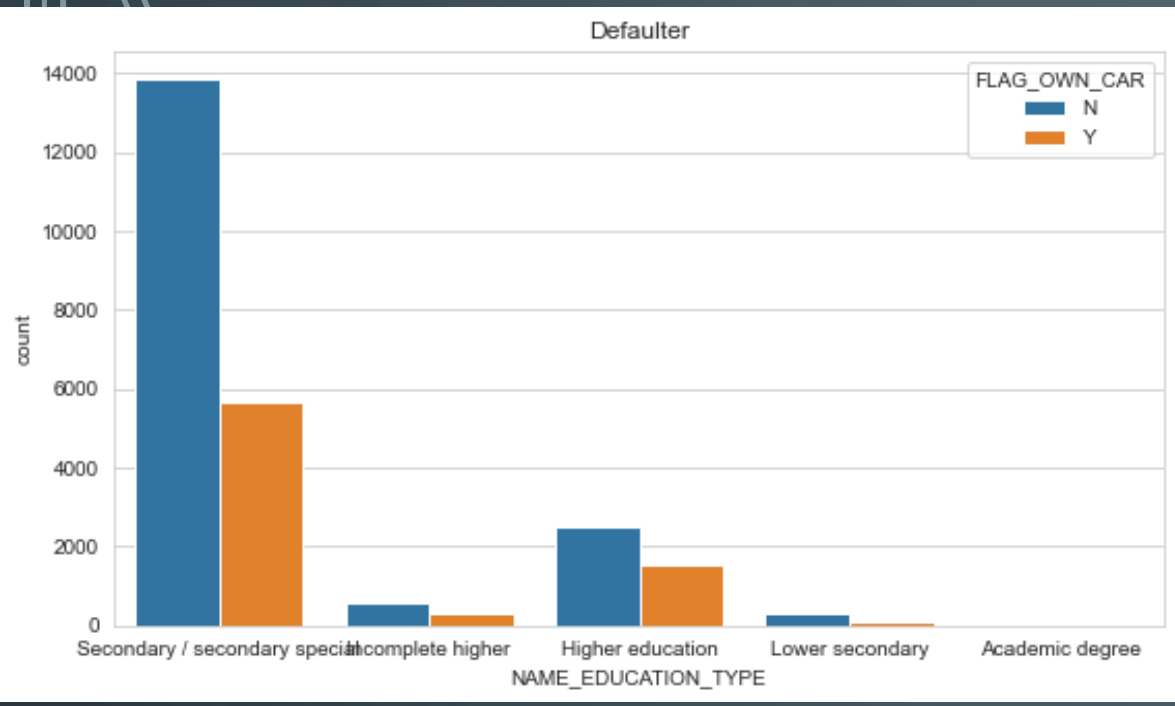
AMT_INCOME_TOTAL VS AMT_CREDIT

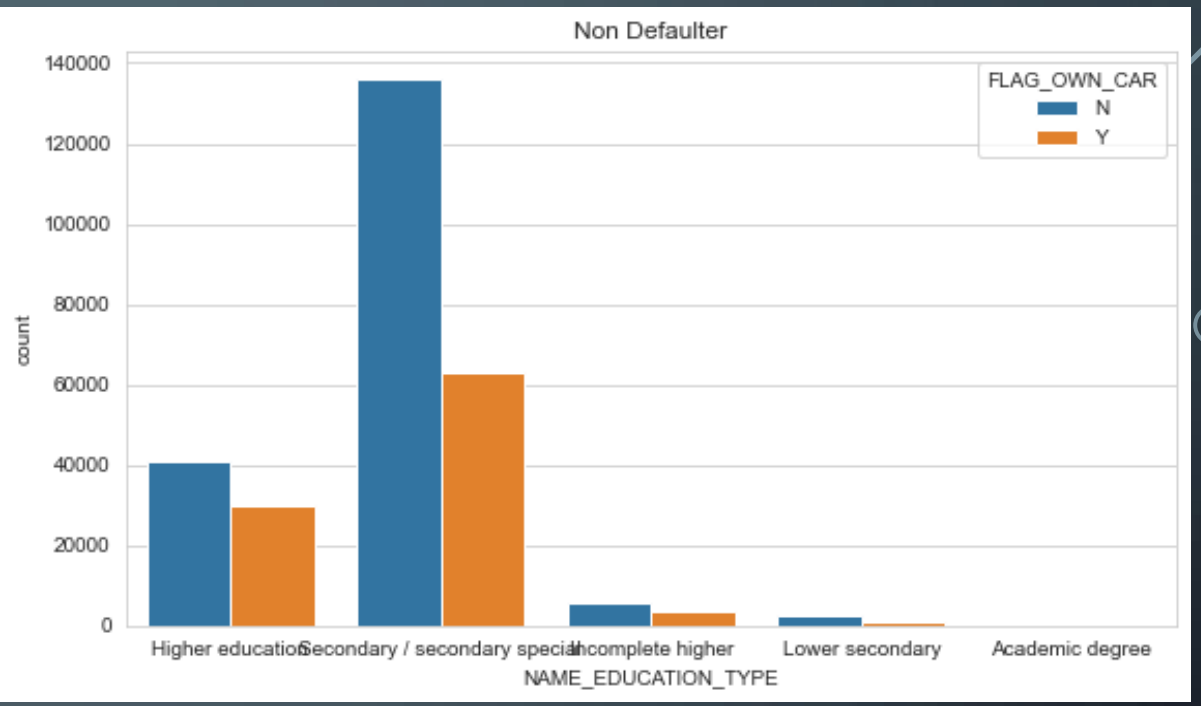
CATEGORICAL-CATEGORICAL COLUMNS





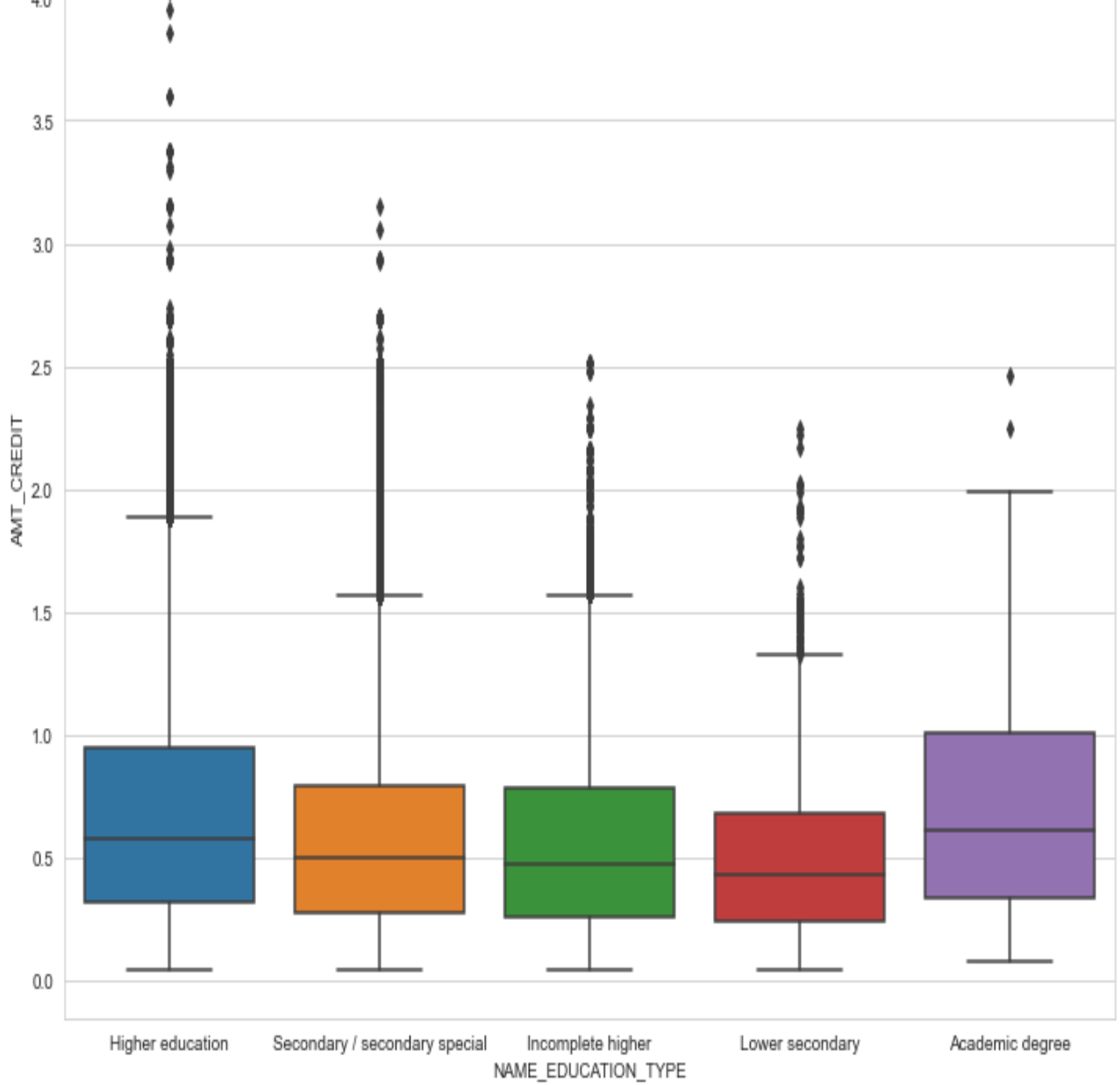
NAME_INCOME_TYPE VS CODE_GENDER

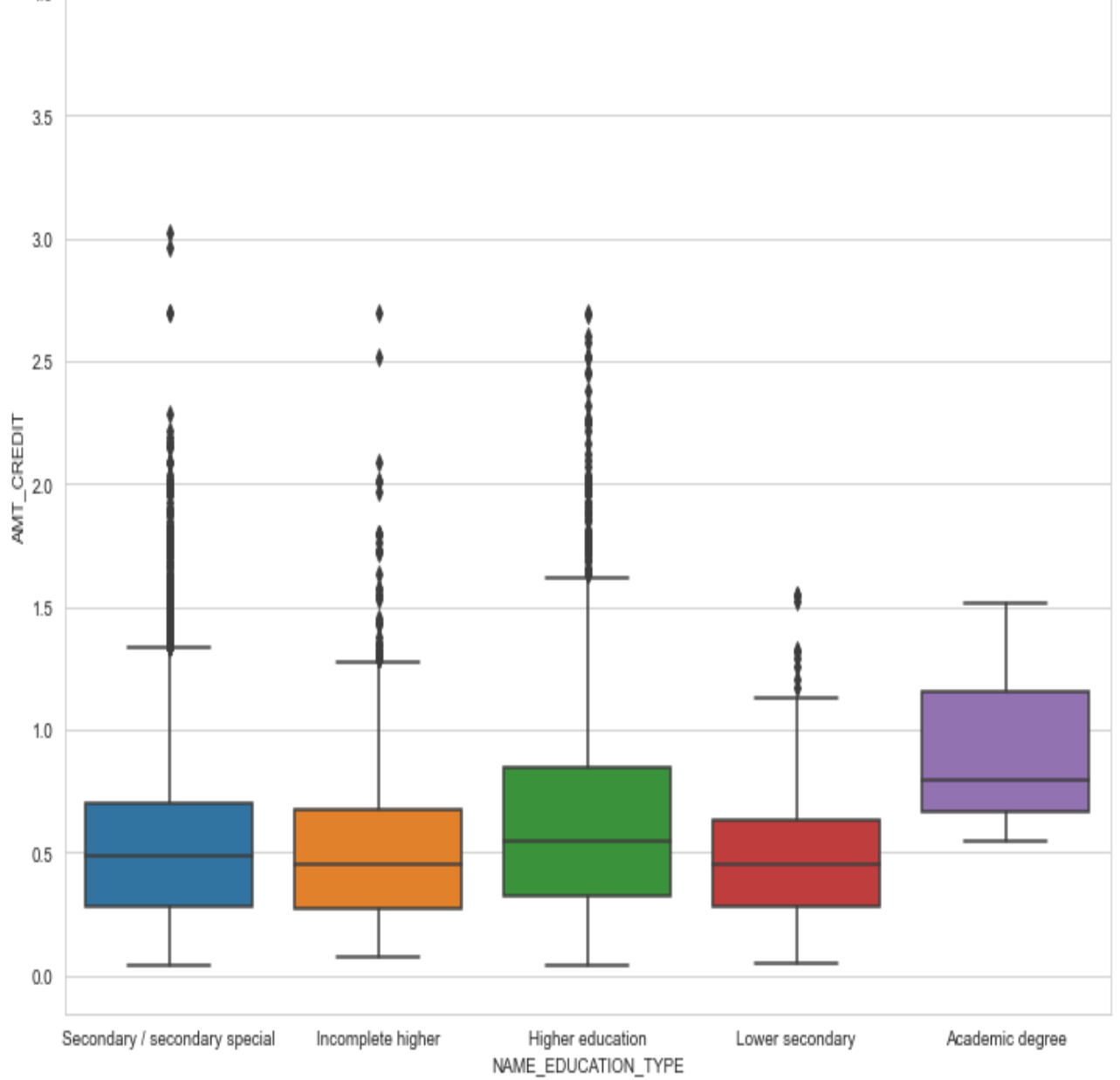




NAME EDUCATION TYPE VS FLAG OWN CAR

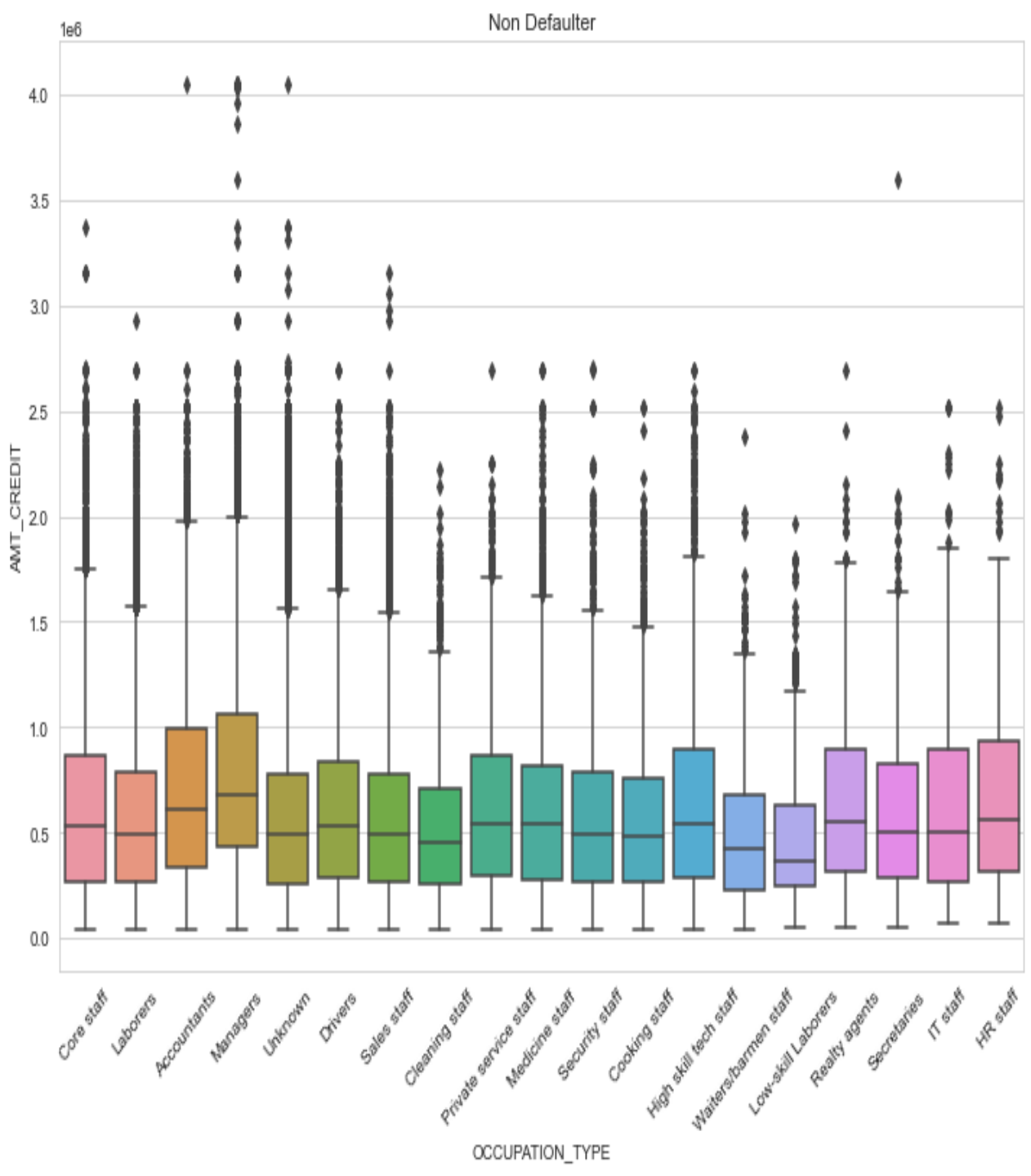
CATEGORICAL VS CONTINOUS VARIABLES

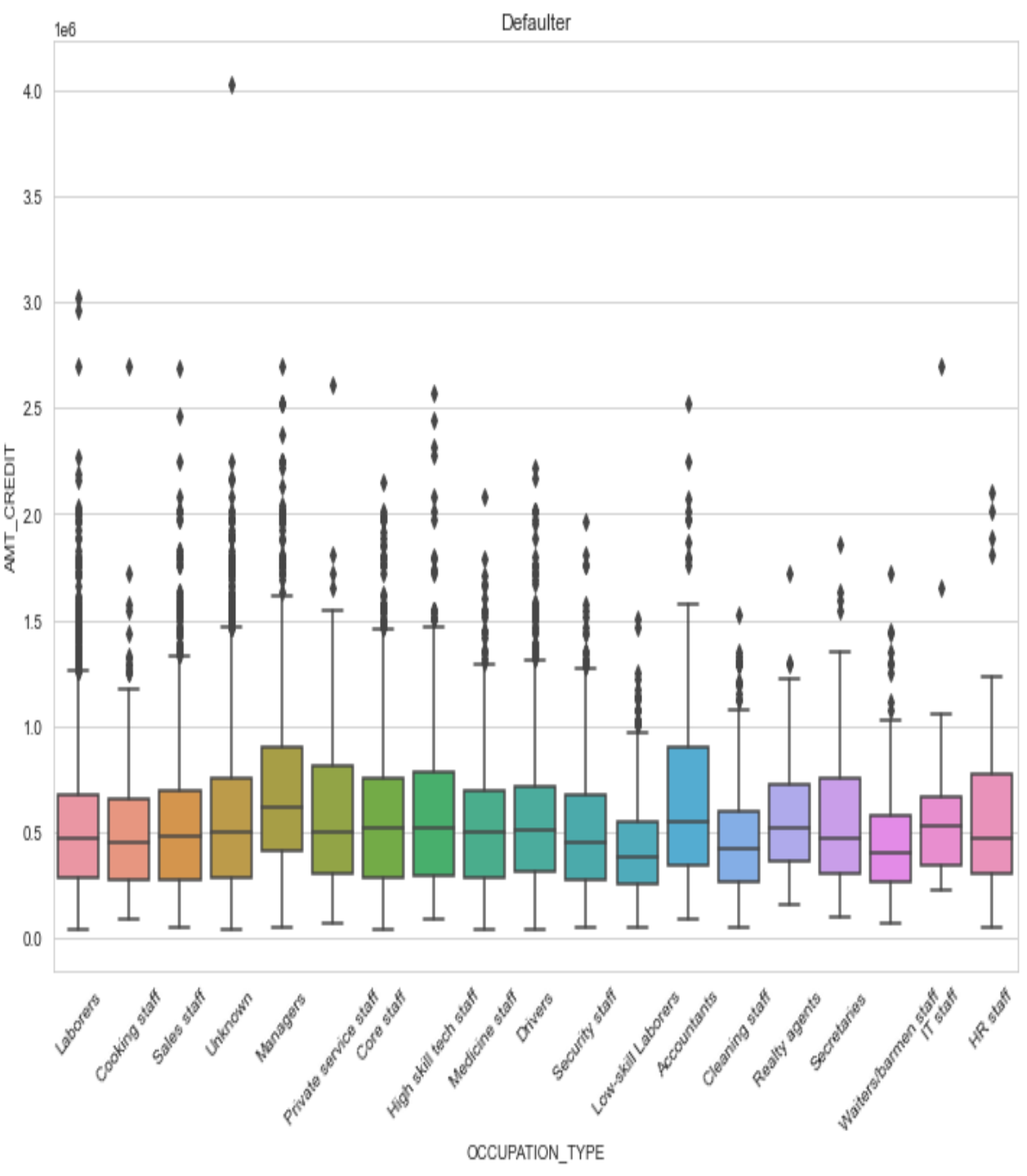




NAME EDUCATION TYPE VS AMT CREDIT

OCCUPATION TYPE VS AMT CREDIT

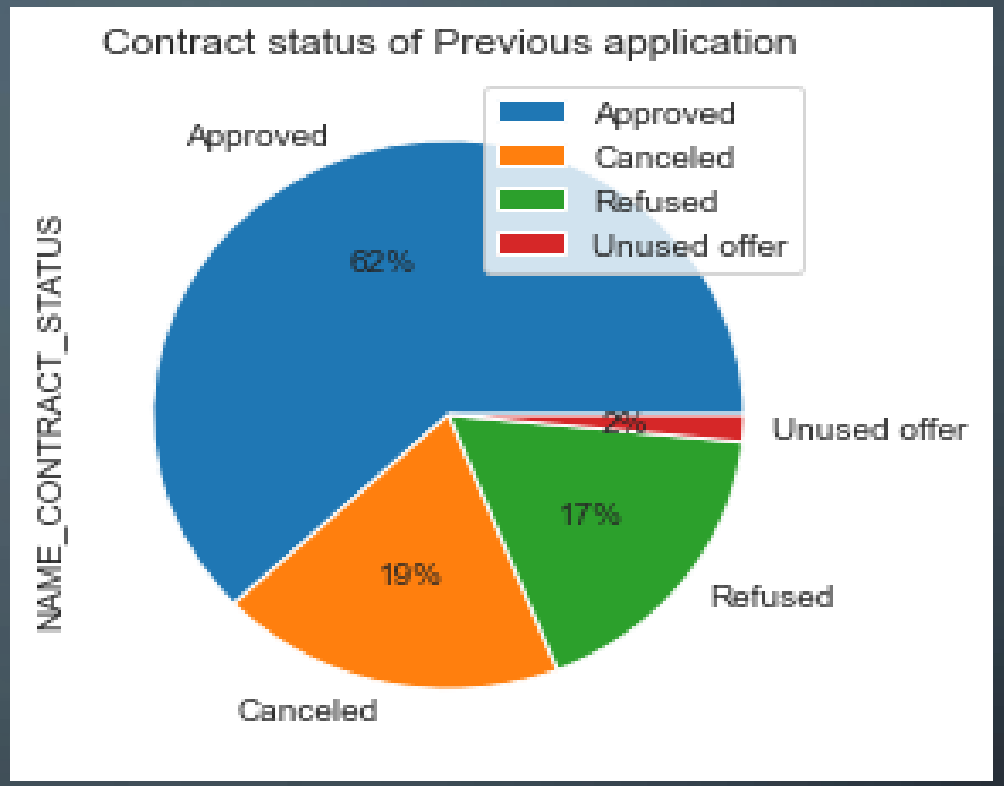

Non Defaulter
1e6
AMT_CREDIT
OCCUPATION_TYPE
Core staff
Laborers
Accountants
Managers
Unknown
Drivers
Sales staff
Cleaning staff
Private service staff
Medicine staff
Security staff
Cooking staff
High skill tech staff
Waiters/barmen staff
Low-skill Laborers
Realty agents
Secretaries
IT staff
HR staff


Defaulter
1e6
AMT_CREDIT
OCCUPATION_TYPE
Laborers
Cooking staff
Sales staff
Unknown
Managers
Private service staff
Core staff
High skill tech staff
Medicine staff
Drivers
Security staff
Low-skill Laborers
Accountants
Cleaning staff
Realty agents
Secretaries
Waiters/barmen staff
IT staff
HR staff

PREVIOUS DATASET - UNIVARIATE ANALYSIS

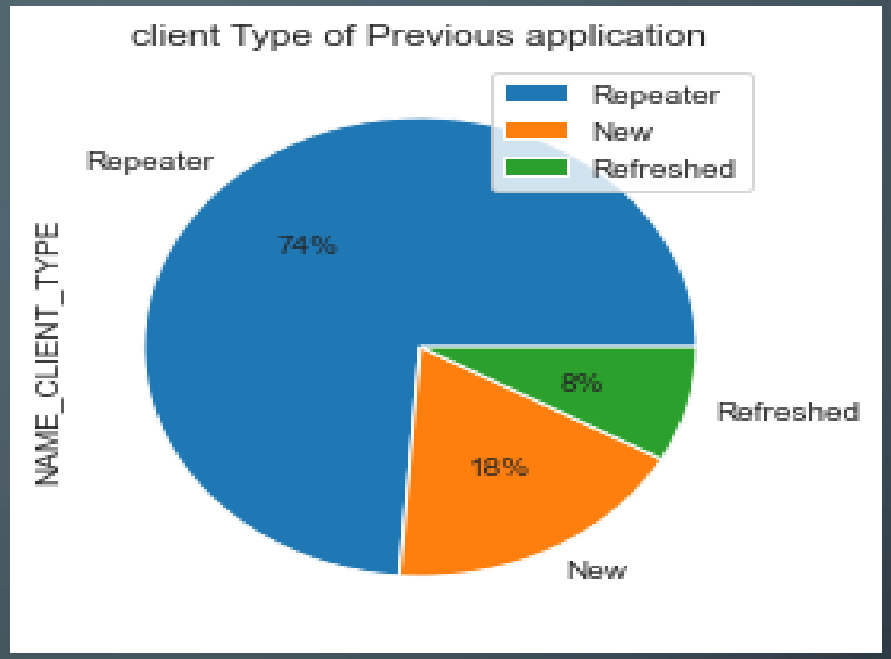
CONTRACT STATUS OF PREVIOUS APPLICATION

- approved loan status is huge than rejected or canceled.



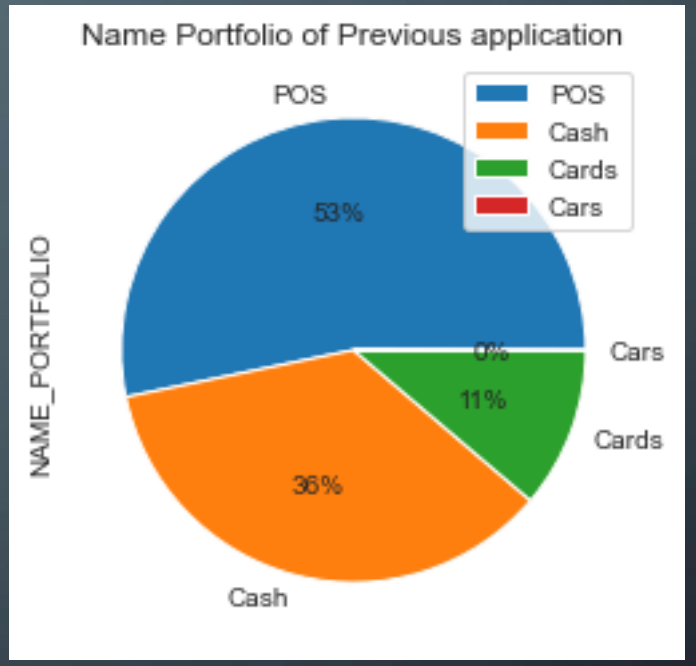
CLIENT TYOE OF PREVIOUS APPLICATION

- Repeater is high than other



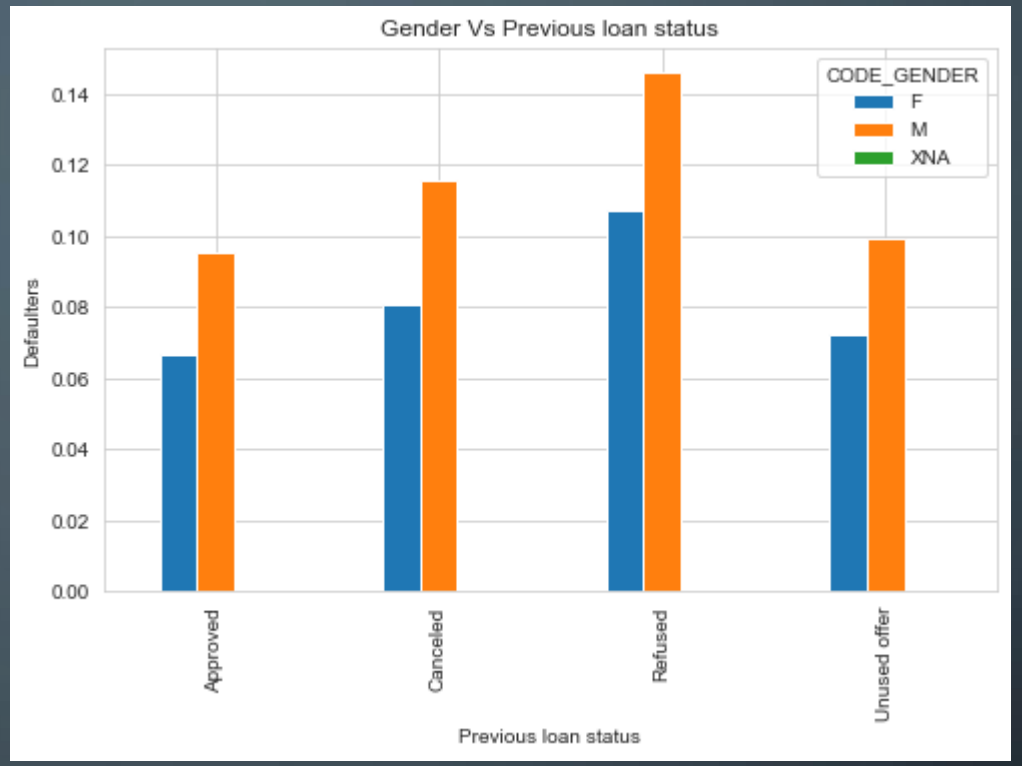
NAME PORTFOLIO VS PREVIOUS APPLICATION

- POS loans are highest rather than cash loans.



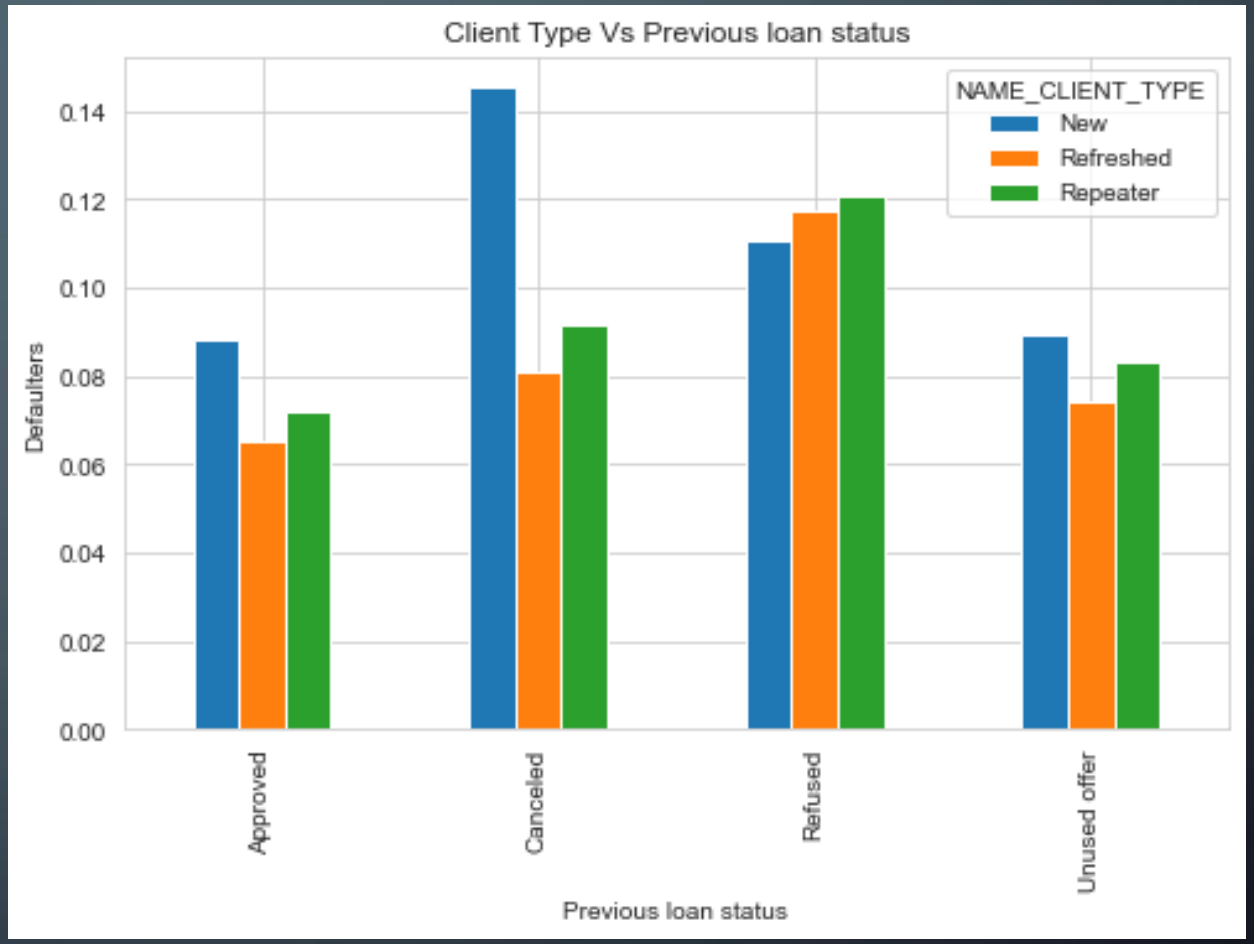
GENDER VS PREVIOUS STATUS

- Male clients are more defaulted than female client. Also, previously refused customer are more defaulted in current application



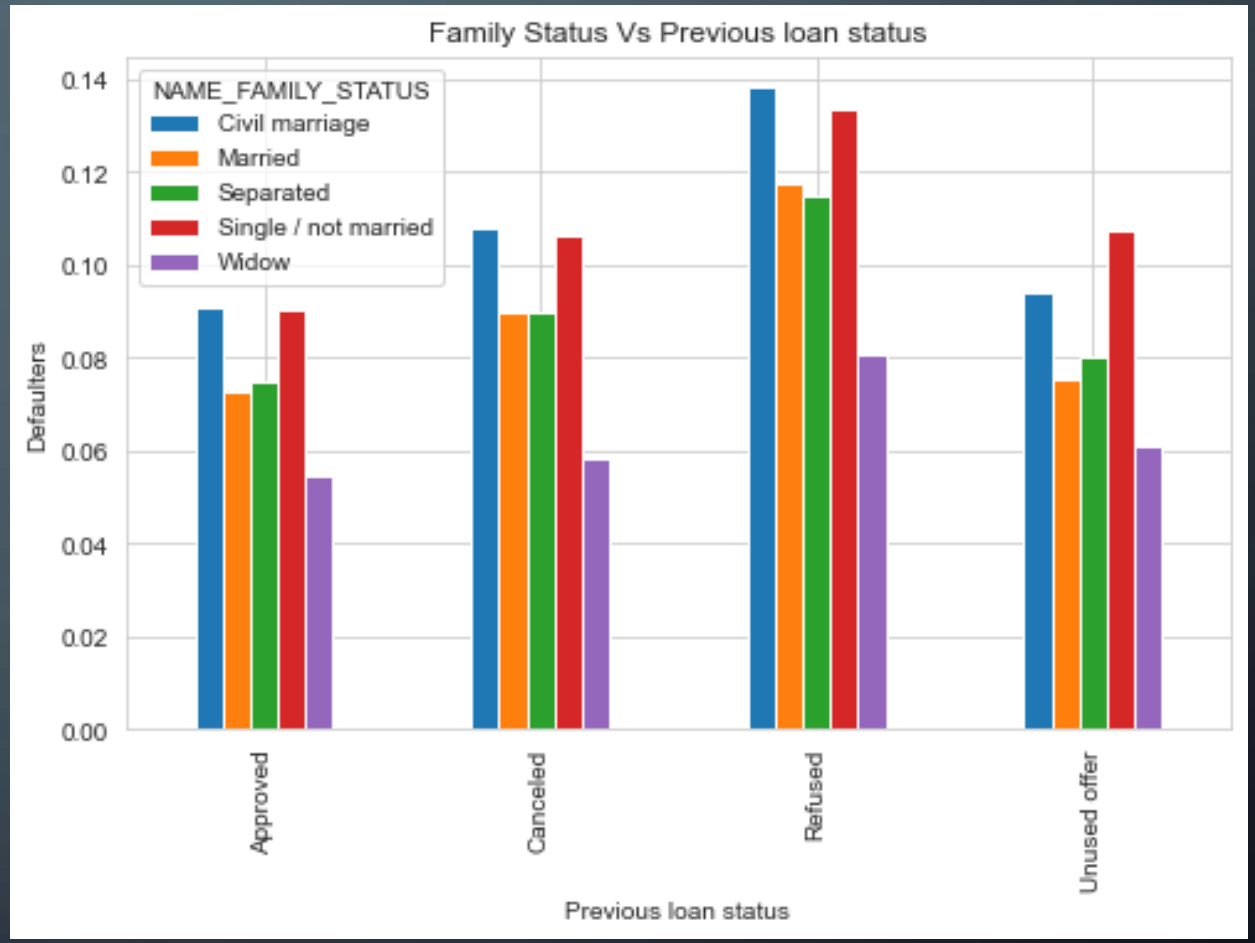
CLIENT TYPE VS LOAN STATUS

- Previously cancelled New and Refreshed clients are more defaulted than repeater clients



FAMILY STATUS VS PREVIOUS LOAN STATUS

- Client who did civil marriage with previously unused loan offers ar more defaulted currently.



CONCLUSION

- Banks should focus less on income type ‘working’ as they are having most number of unsuccessful payments.
- Loan purpose on 'repair' having highest number of unsuccessful repayments.
- Housing type 'with parents' having least number of unsuccesful repayments.
- In genders,'females' are more in number for applying loans.
- Banks should focus on 'students' ,'pensioner' for successful repayments.

THANK YOU