CREDIT CARD EDA CASE STUDY

ANSUMAN PATNAIK

- ansumapatnaik0.ap@gmail.com
- 8763073935

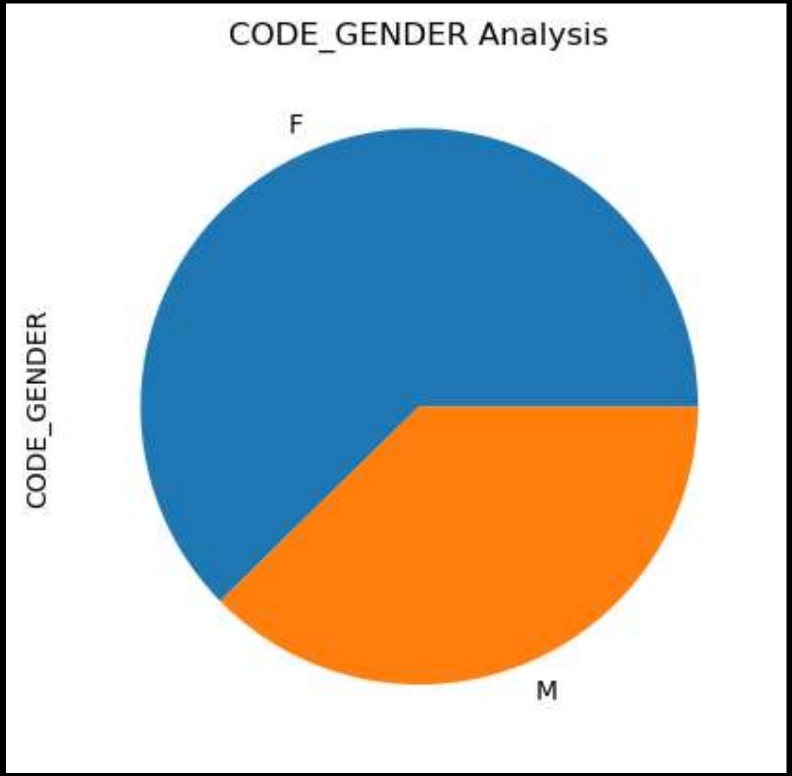
Introduction

Due to their weak or nonexistent credit scores, loan providers find it challenging to grant loans to individuals. Because of this, some customers take advantage of it by defaulting. Imagine you work for a consumer finance business that specializes in providing urban customers with different kinds of loans. To analyze the trends found in the data, you must use EDA. By doing this, it will be ensured that only those applicants who can repay the debt will be accepted.

When a loan application is received, the business must determine whether to approve the loan based on the applicant's profile. The bank's choice is subject to two different kinds of risks:

- If the borrower is likely to repay the loan, refusing to grant it results in the firm losing business.
- If the borrower is not likely to pay back the loan, or is likely to default, then authorizing the loan may result in a loss of revenue for the business.

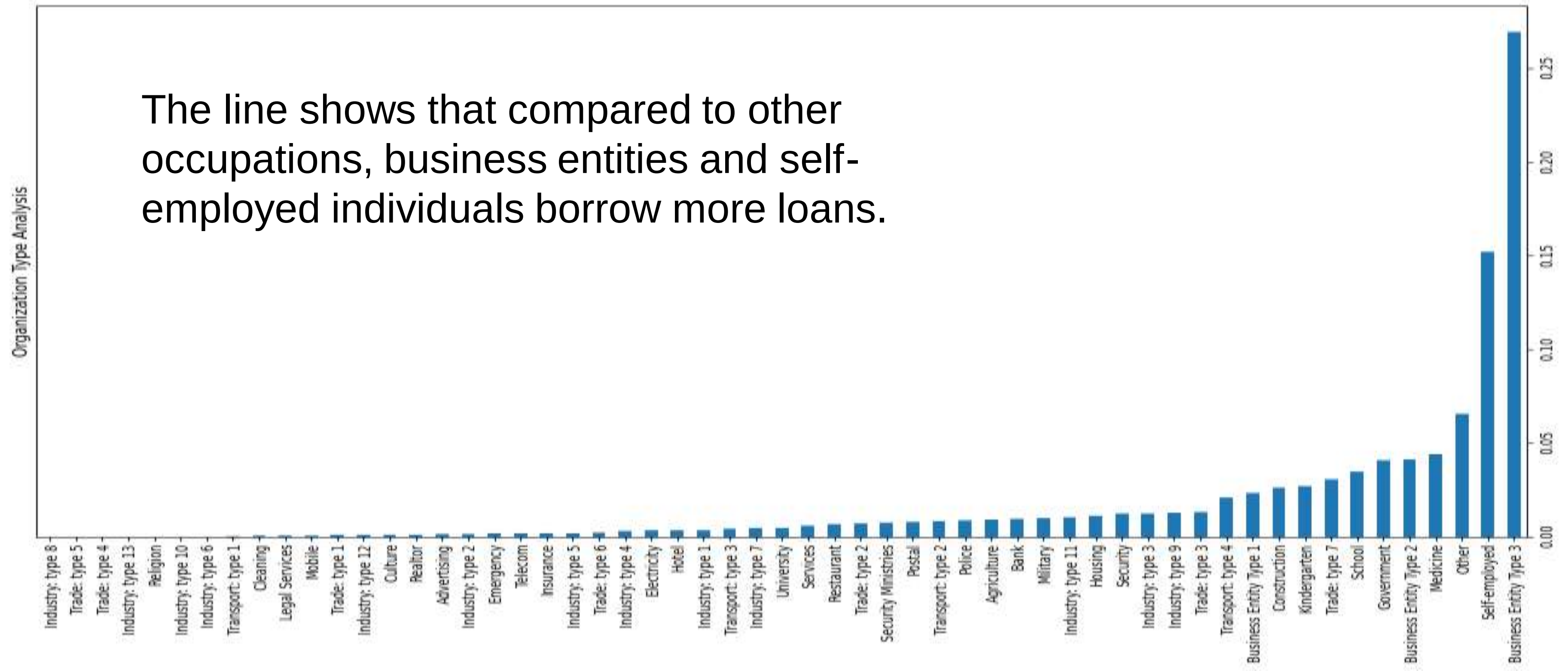
Code Gender Analysis



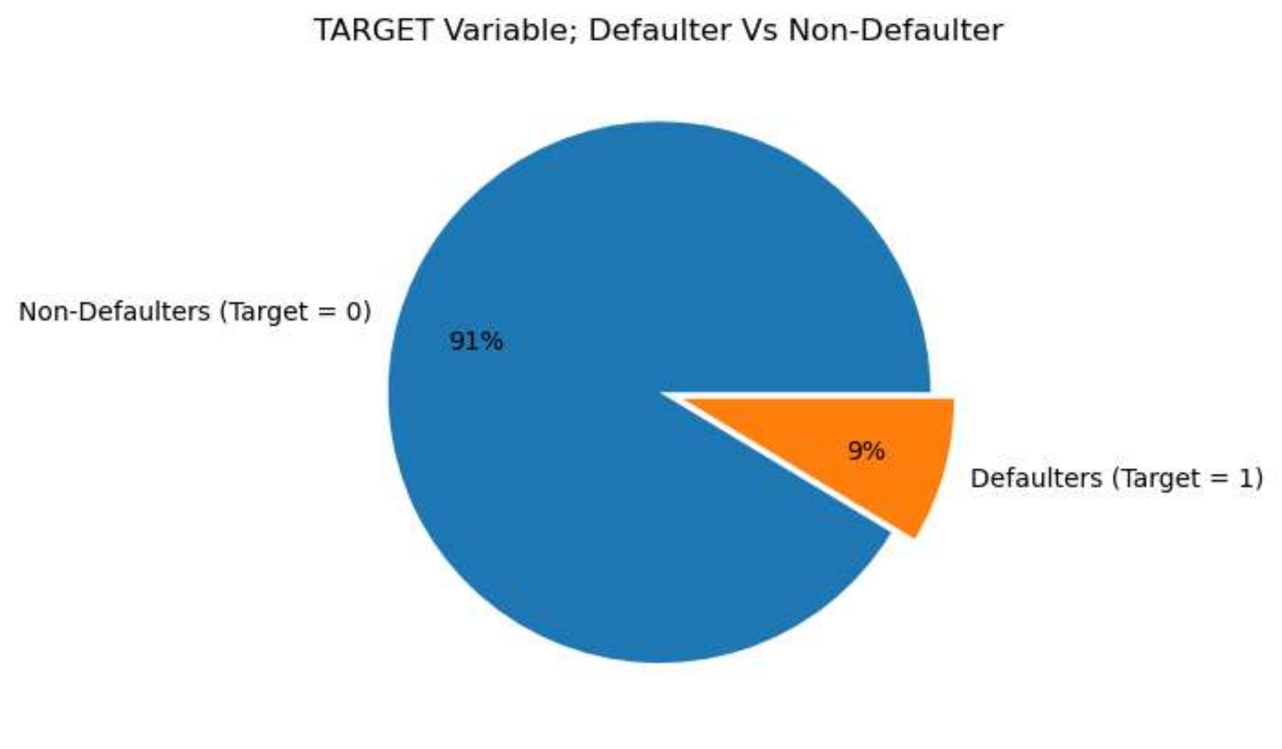
It shows that there are more female loan candidates than male, on average.

Analysis of applicant by Organization type

The line shows that compared to other occupations, business entities and self-employed individuals borrow more loans.

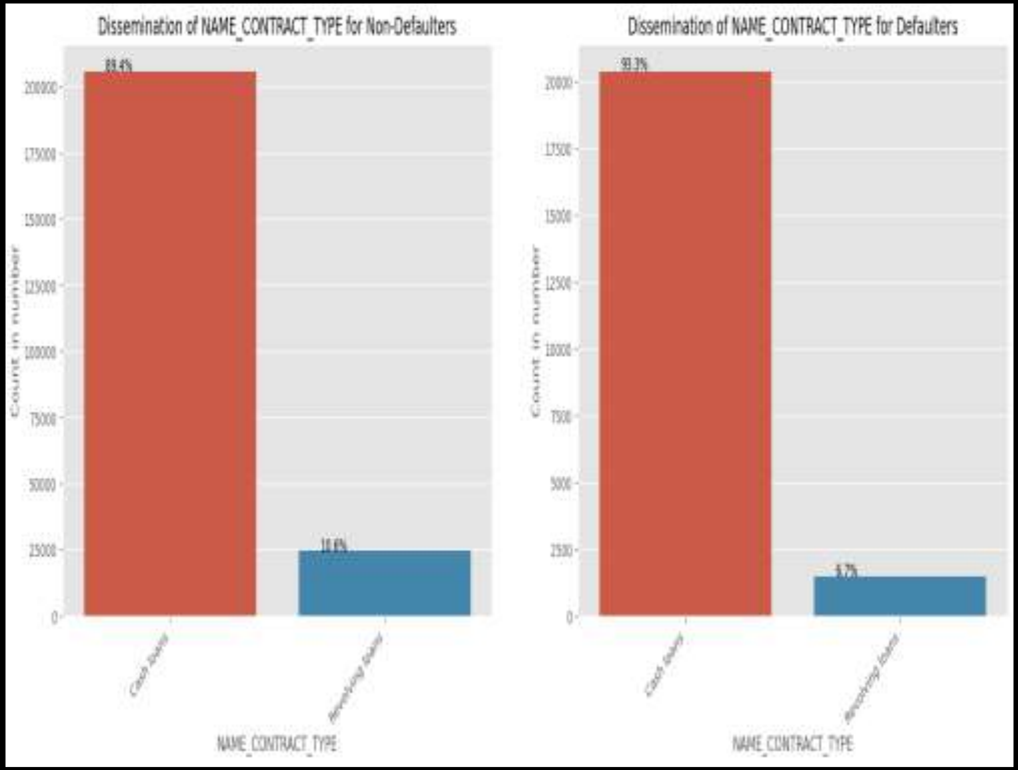


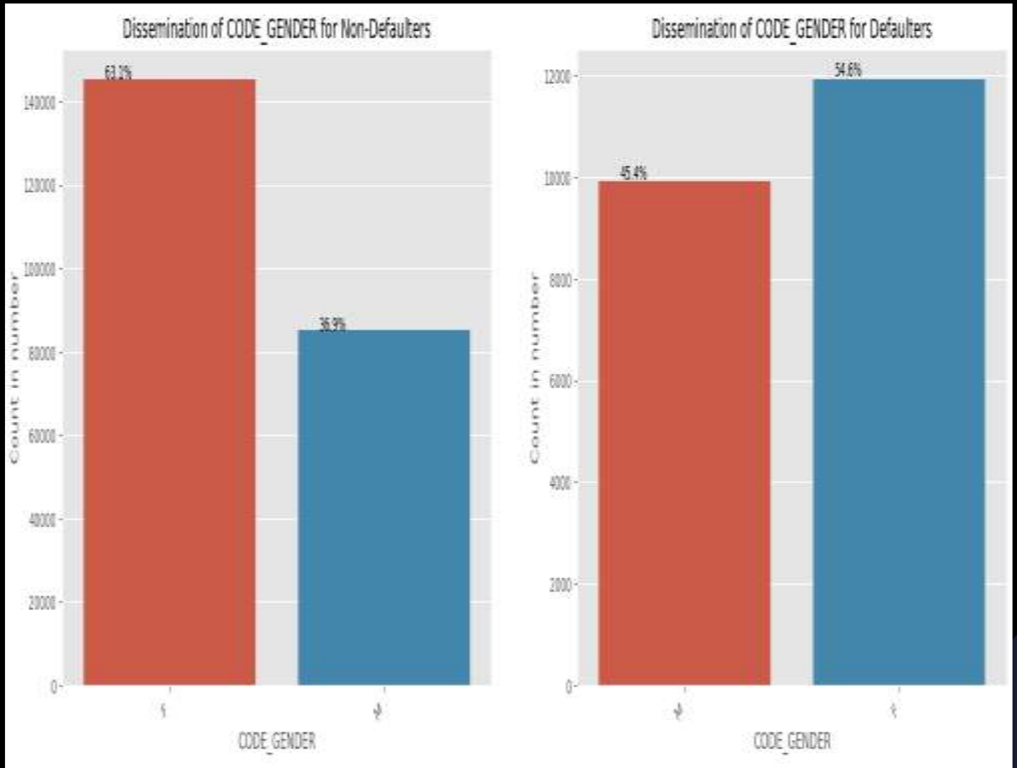
Analysis of applicant by Target Variable (Payment)



Approximately 91% of clients—those on the non-defaulters list—don't have any problems making payments, while the remaining 9% do.

Categorical Analysis based on Gender and Loan Type



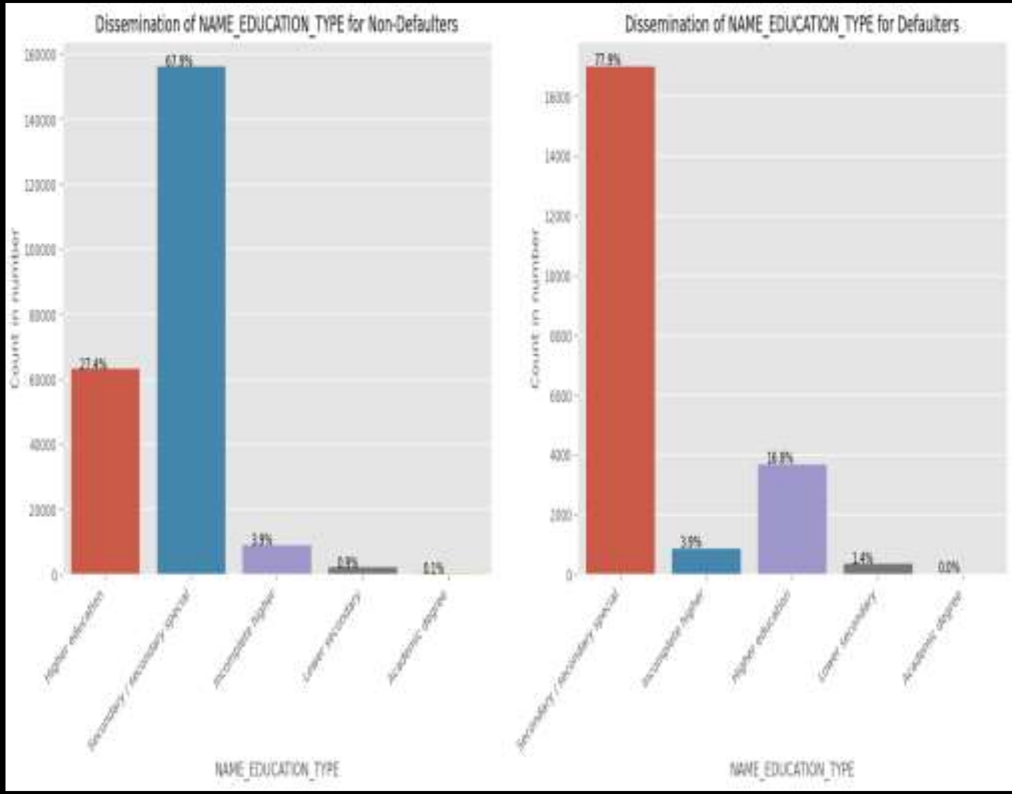


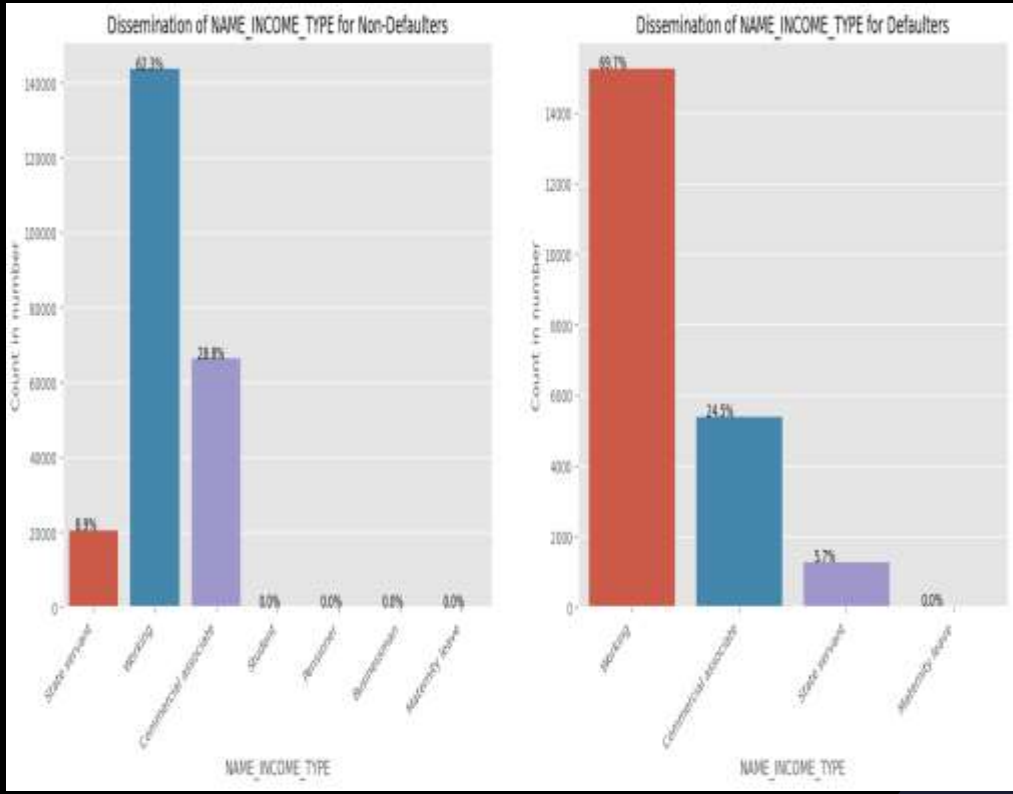
Inference: Categorical analysis based on Gender and Loan Type

- LOAN_TYPE(CONTRACT) DEFAULTER & NON-DEFAULTER ANALYSIS:
  - Most number of defaulters are having cash loans rather than revolving loans.

- CODE_GENDER DEFAULTER & NON-DEFAULTER ANALYSIS:
  - Female contributes to 63% to the non-defaulters whereas Male contributes to 37% of the non-defaulters.
  - To the defaulters list, Female contributes more than Male.
  - Female applying for loans are more than the number of male applying for loan. Hence, we may infer that number of female defaulters are more.

Categorical Analysis based on Income and Education Type

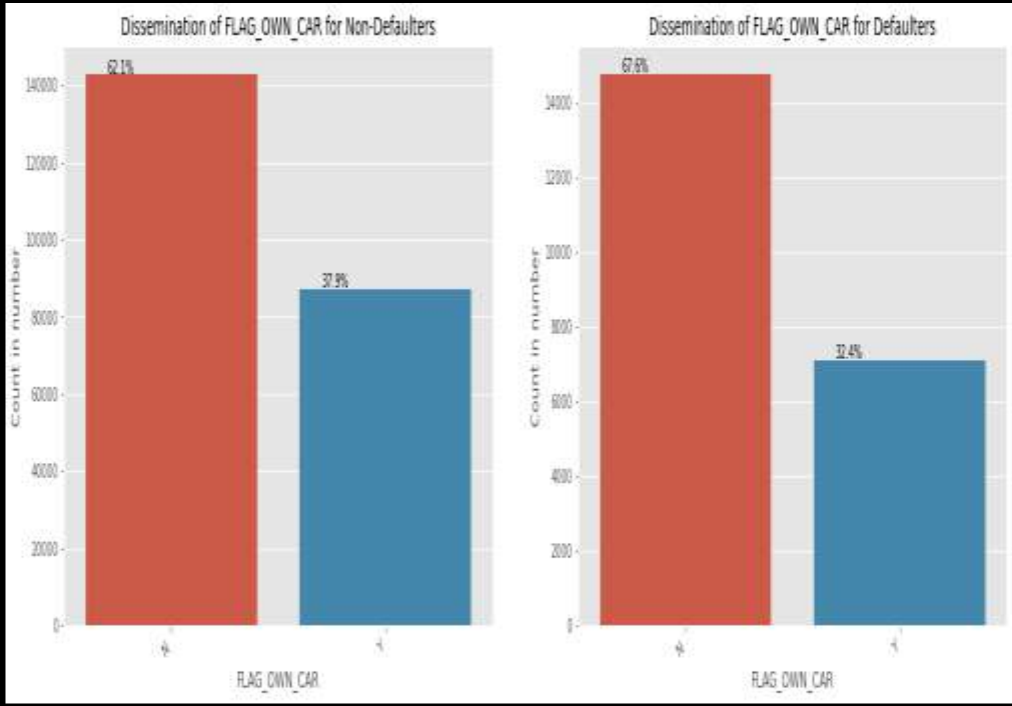


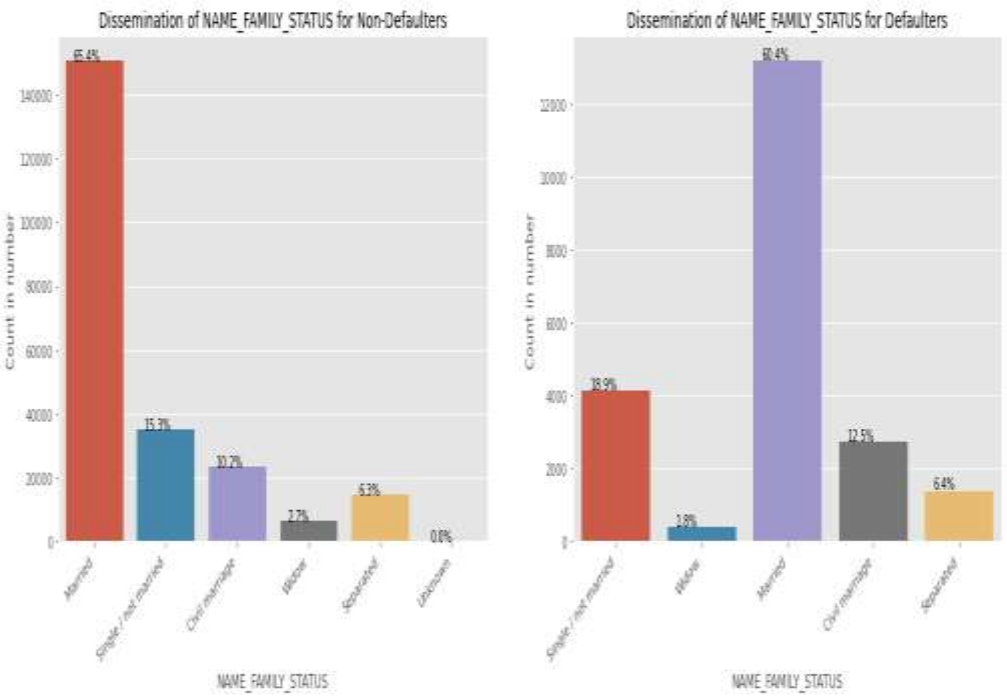


Inference: Categorical analysis based on Income and Education Type

- EDUCATION TYPE DEFAULTERS & NON_DEFAULTERS ANALYSIS:
  - Mostly loans are distributed to "Secondary / Secondary Special Education" group.
  - As the number of loans distributed to Secondary/ Secondary Special Education group is high, secondary/ secondary special group contribution to defaulters list is also high.
- INCOME TYPE DEFAULTERS & NON_DEFAULTERS ANALYSIS:
  - Student, Pensioner, Businessman and women with maternity leaves are not likely to be in the defaulter list.
  - It can be inferred that most of the loans are distributed to working class, commercial associate and state servant.
  - We can also observe that, working class persons contributes 62% to non-defaulters list where as around 70% to the defaulters list. It shows that the working-class clients are more in the defaulters list.

Categorical Analysis based on Own cars and Family status

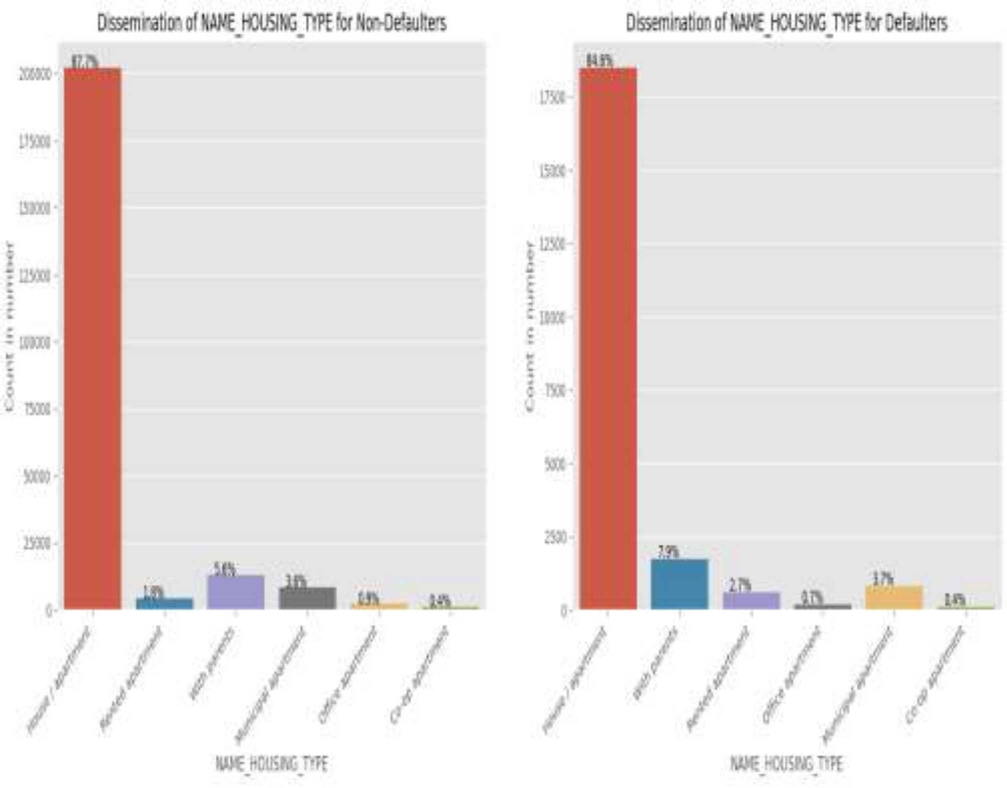


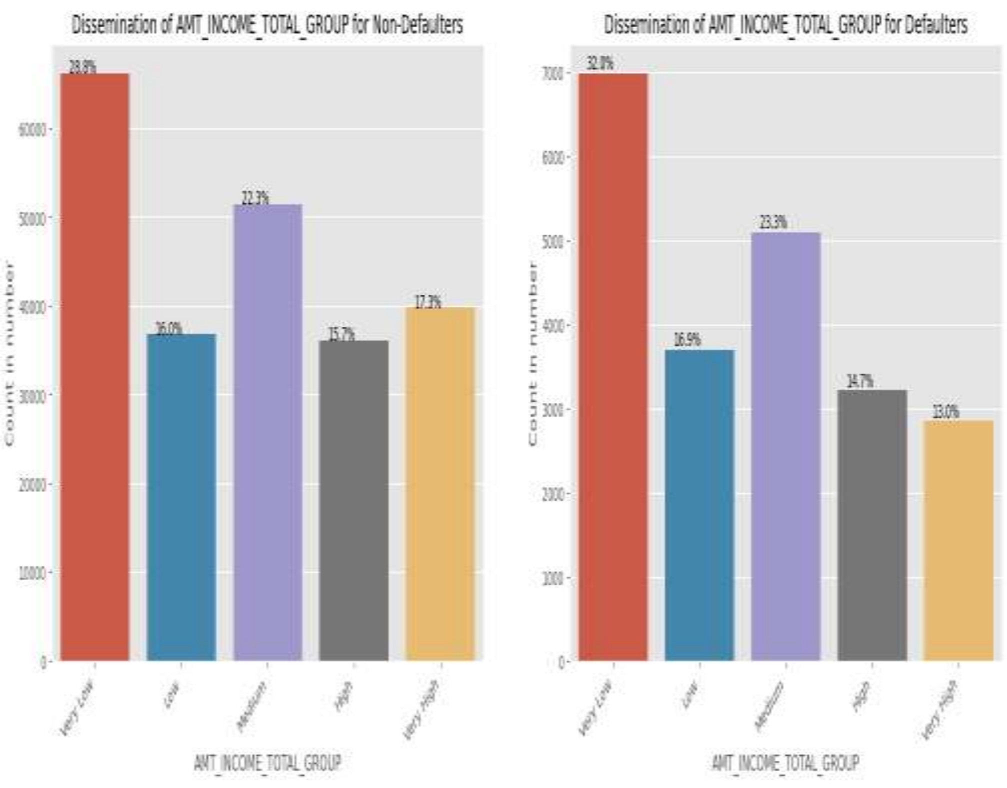


Inference: Categorical analysis based on Own car & Family Status

- FLAG OWN CAR DEFAULTER & NON DEFAULTER ANALYSIS:
  - People who own a vehicle account for 38% of non-defaulters and 32% of defaulters, according to our observations.
  - More individuals who own cars are added to the default list than those who do not.
  - When compared to the percentage of individuals who don't own a car, the default rate for car owners is low.
- FAMILY STATUS DEFAULTER & NON-DEFAULTER ANALYSIS:
  - Married People tend to apply loan more than any other group.
  - We observe that single/ unmarried, civil marriages and separated persons contribute more number to the defaulters list. So, it might be a risk to give loans to these categories.

Categorical Analysis based on Housing type and Total Income Group

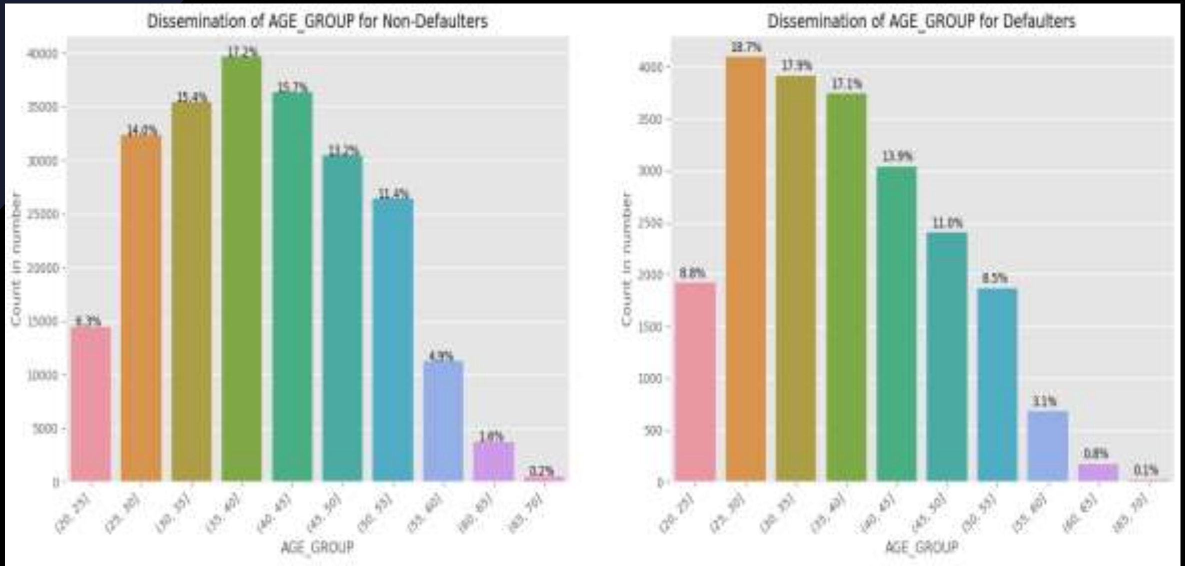




Inference: Categorical analysis based on Housing type and Total Income Group

- HOUSING TYPE DEFAULTERS & NON-DEFAULTERS ANALYSIS:
  - Persons with own house/apartment apply more for loans rather than any other group.
  - Persons living with parents contribute more to defaulters list rather than any other group. It is quite possible that as parents are staying near by, so living expenses are higher and so contributed to defaulters list.
- TOTAL INCOME GROUP DEFAULTERS & NON-DEFAULTERS ANALYSIS:
  - The very high-income group tend to default less than the other class of income group.
  - Very low-income group tend to have more defaulters.

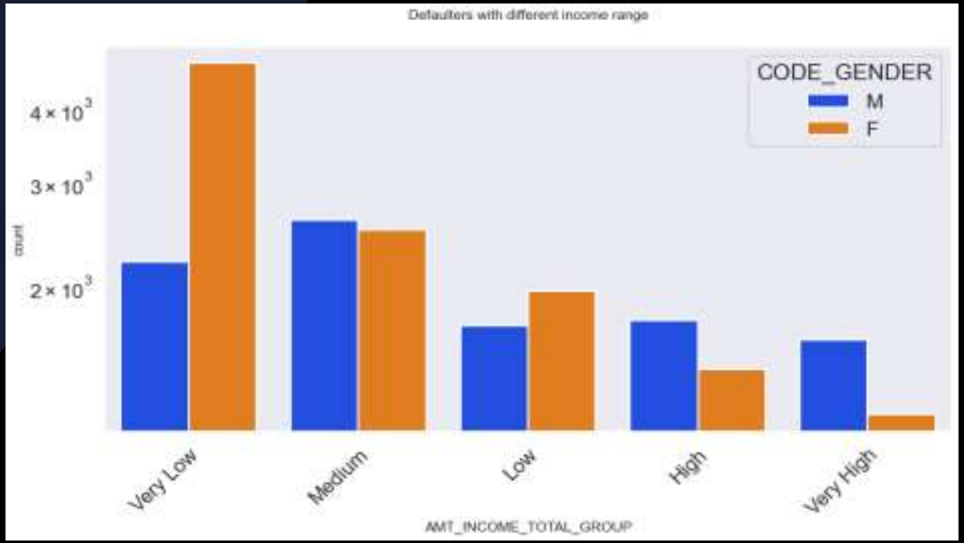
Categorical analysis based on Age Group

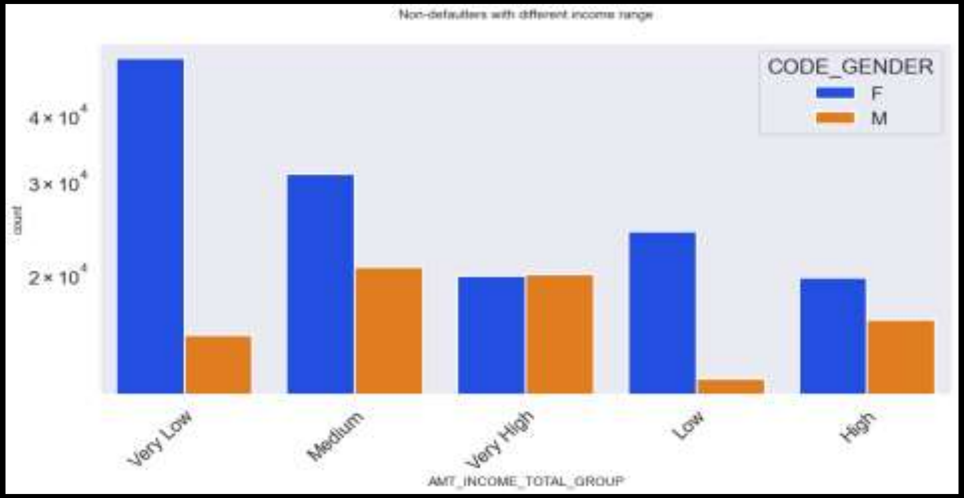


- It is observed that persons between the age of (25-35), tend to default more. So, it is risky to approve loan for this age group.
- With increase in age, persons tend to default less. One of the reason could be they get employed during this age and their salary also increases. So, they repay the loan.

Analysis based on Total Income group and Gender

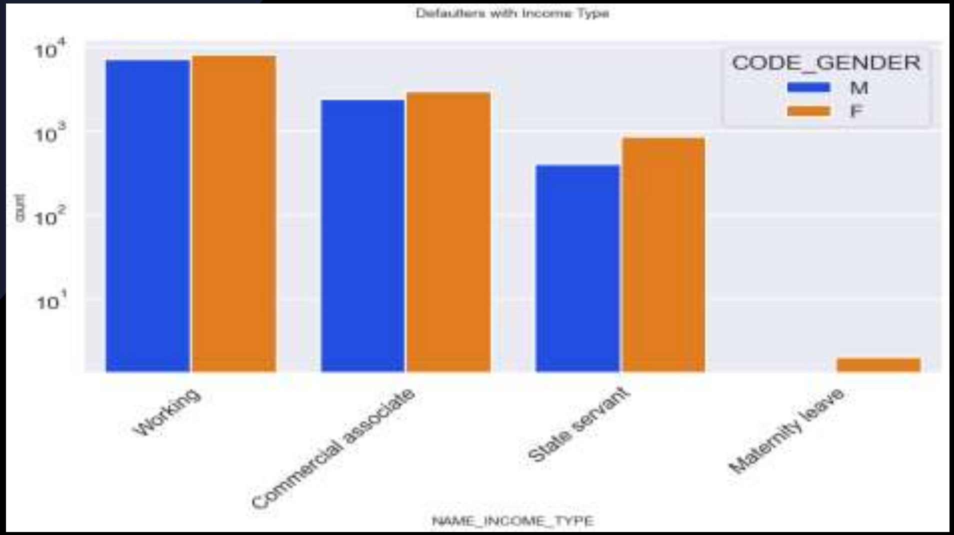
- NON-DEFAULTERS:
  - Count of female is more than that of count of male in the non-defaulters category.
- DEFAULTERS:
  - In the defaulters list, most likely Male in the Medium, High and Very High-income range tend to be defaulters.

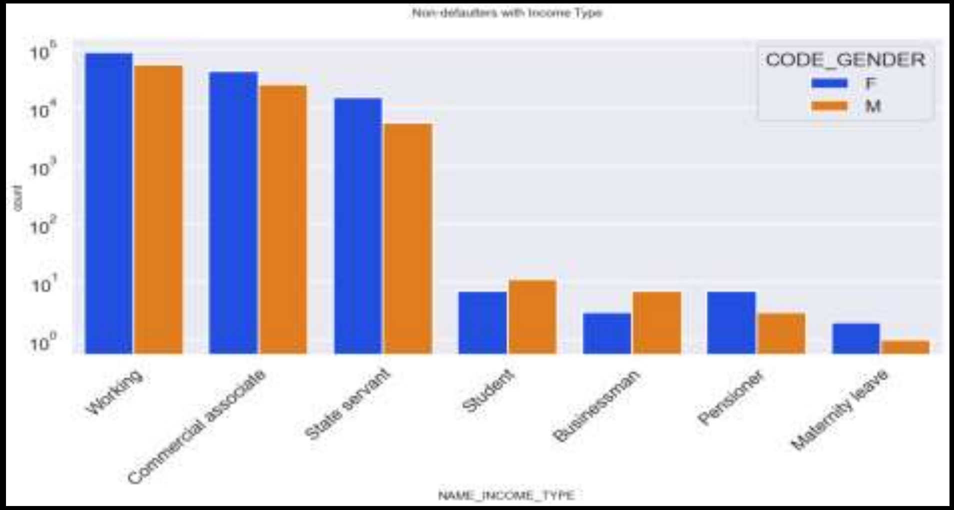




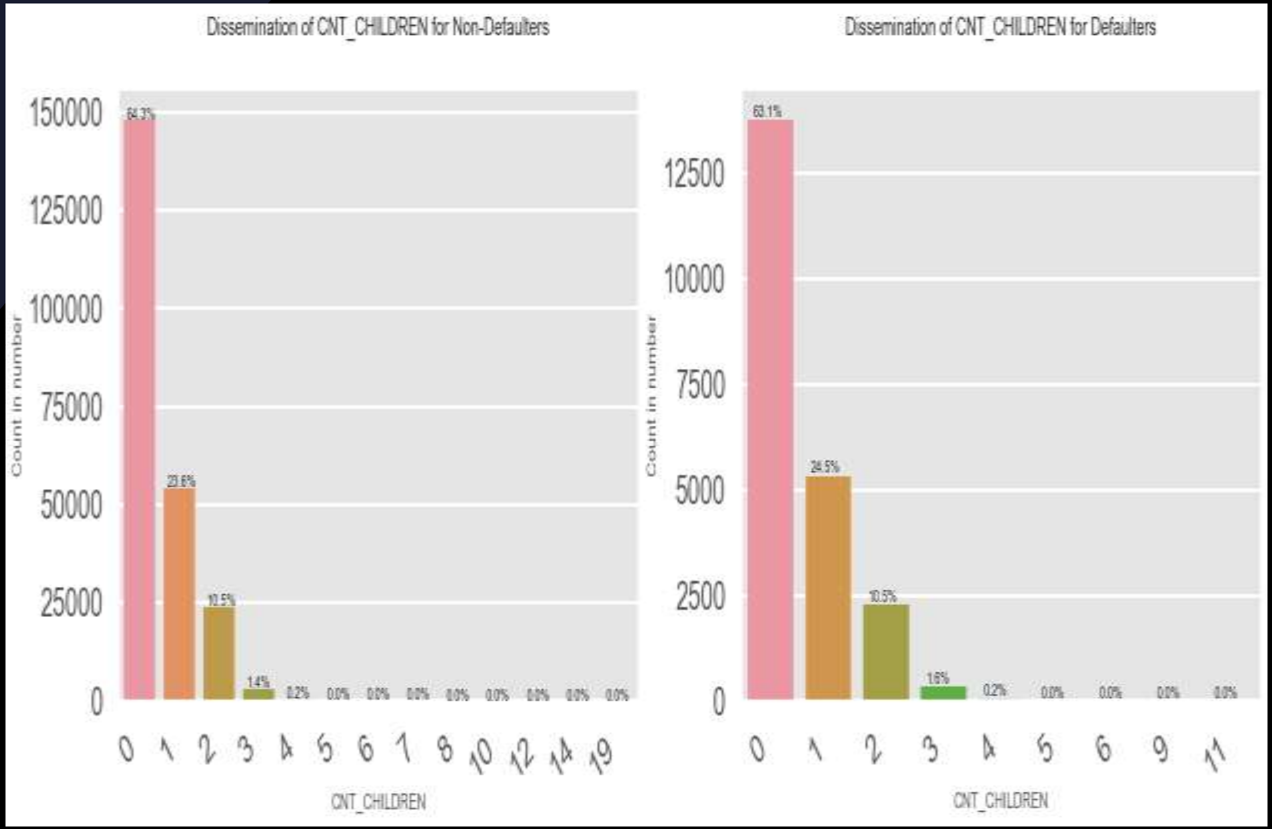
Analysis based on Total Income Type and Gender

- NON-DEFAULTERS:
  - According to working group, Female are having more credit than Male
  - High credit goes for working, commercial associate, state servant
  - Low credit goes for student, pensioner, businessman and maternity leave.
- DEFAULTERS:
  - Mostly Female who are working, commercial associate and state servant tend to be more defaulters than any other categories.



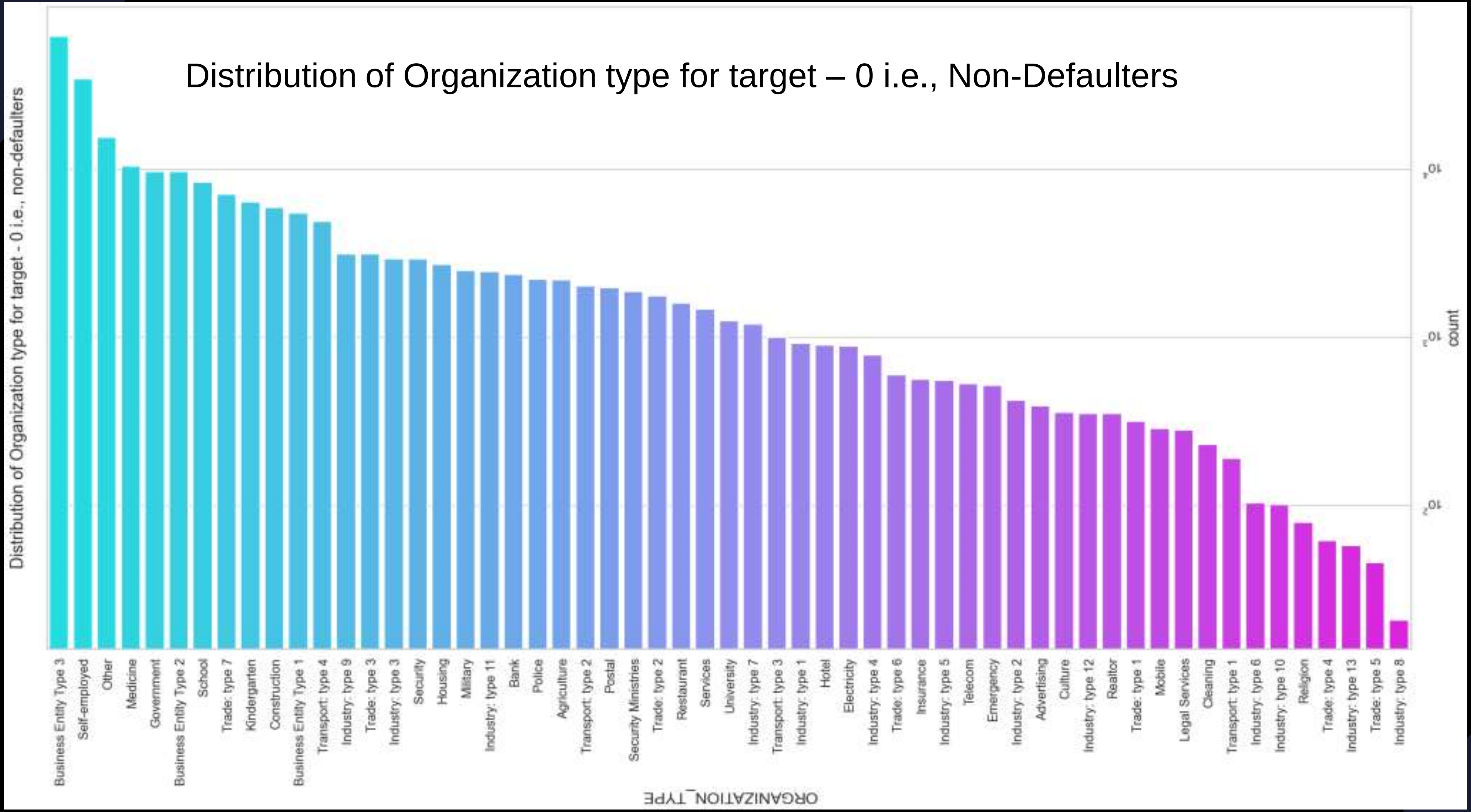


Analysis by count of children for the Applicants



The above chart is unpredictable as low number of children contribute to both defaulters and non-defaulters.

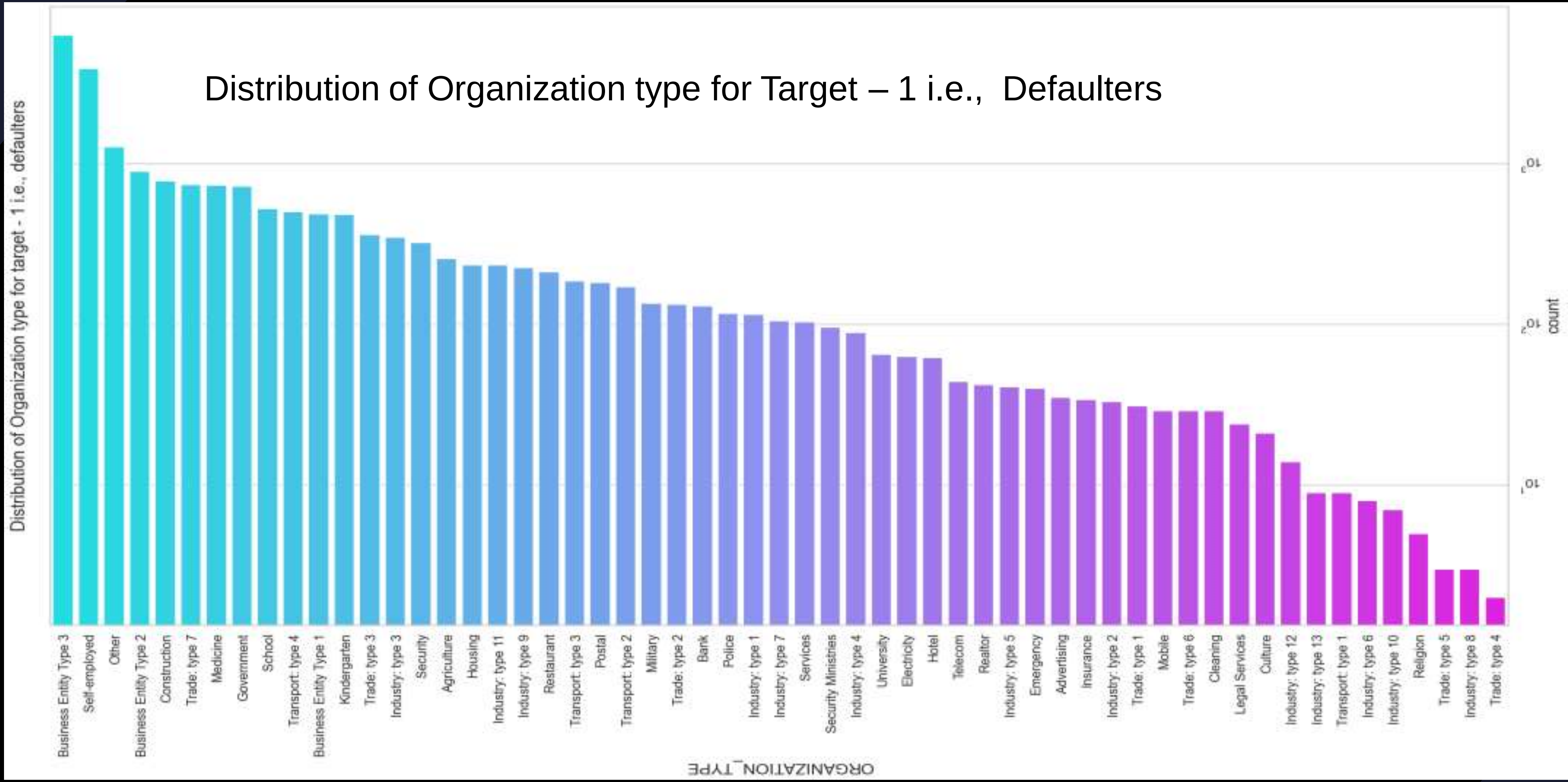
Distribution of Organization type for target – 0 i.e., Non-Defaulters



Inference: Distribution of Organization type for target – 0 i.e., Non-defaulters

- Business Entity, Self-employed, Other, Medicine, Government organization type are the top 5 organizations applying for credit.
- Industry, Trade, Religion organization type are the bottom 5 organizations to apply for credit. We might infer that bank might not provide credit to these type organizations as risk is involved in repayment.

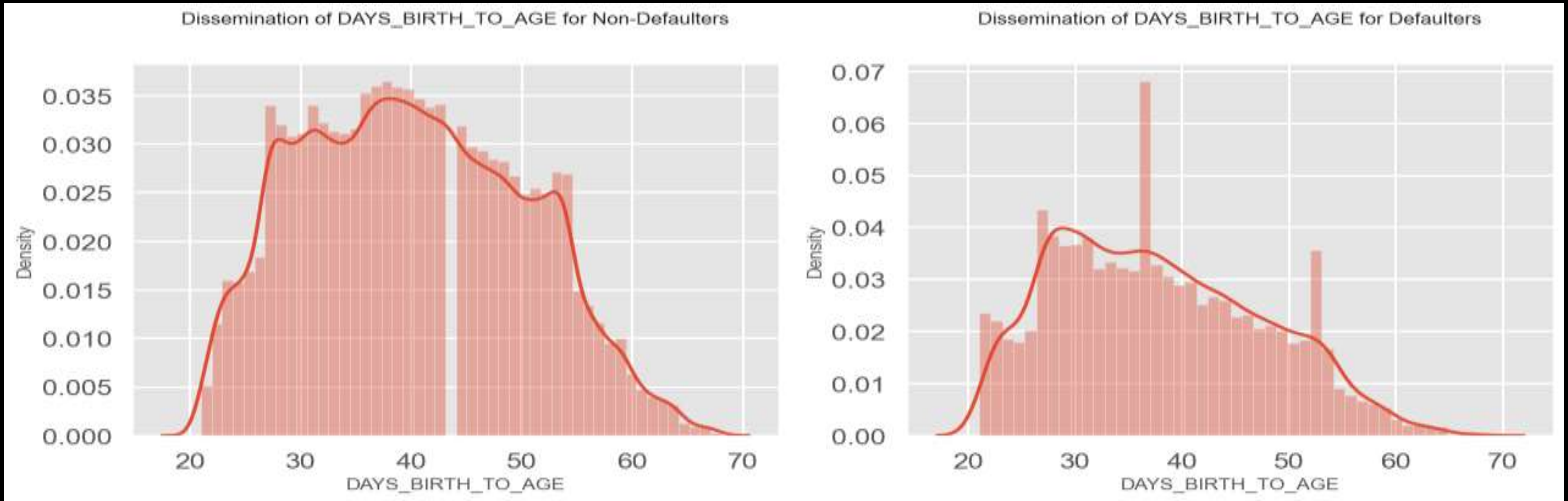
Distribution of Organization type for Target – 1 i.e., Defaulters



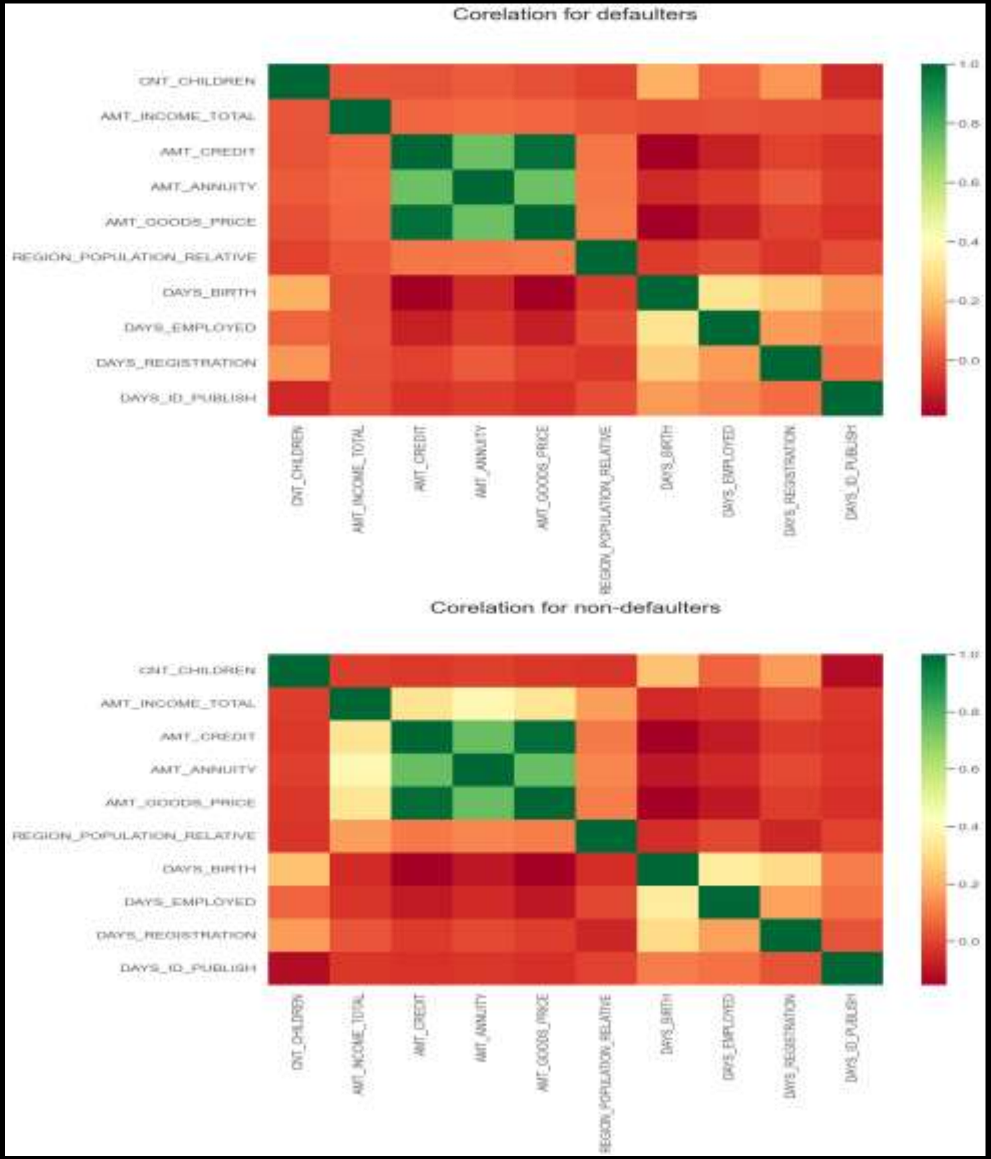
Inference: Distribution of Organization type for target – 1 i.e., Defaulters

- Business Entity, self employed, other, trade and constructions are most likely to get on defaulters list. We may infer that as these type group persons does not get a continuous flow of income as salary, so they are the one's who are likely to get in the defaulters list.
- Trade, Industry, Religion are bottom 5 organizations who are getting into the defaulters list.

Analysis by Days of Birth



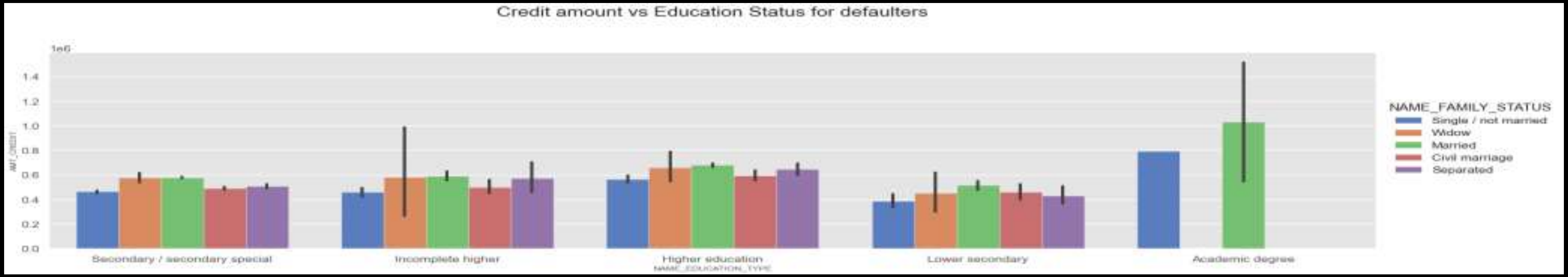
Persons applying for credit are on the age group of (25-55) whereas mostly persons from age group (35-40) and (50-55) are likely to be defaulters.

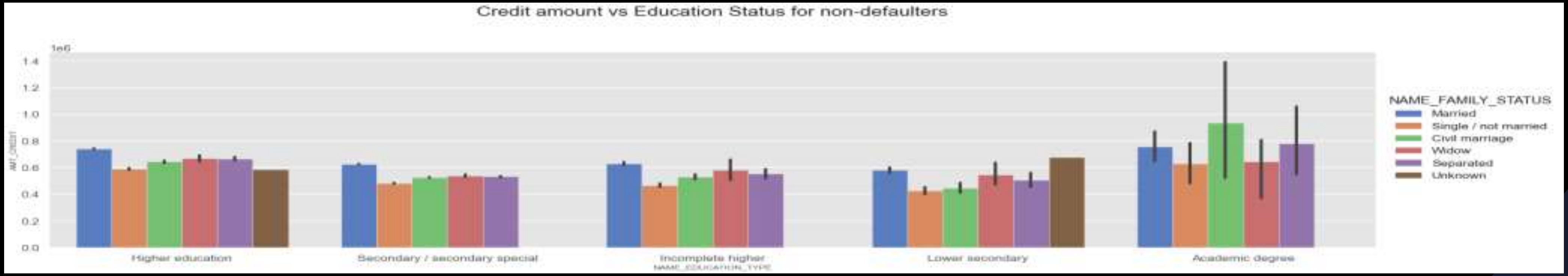


Correlation Analysis

- The following columns are having high correlation:
  - AMT_CREDIT and AMT_GOODS_PRICE
  - AMT_CREDIT and AMT_ANNUITY
  - AMT_ANNUITY and AMT_GOODS_PRICE

Credit amount vs Education status Analysis

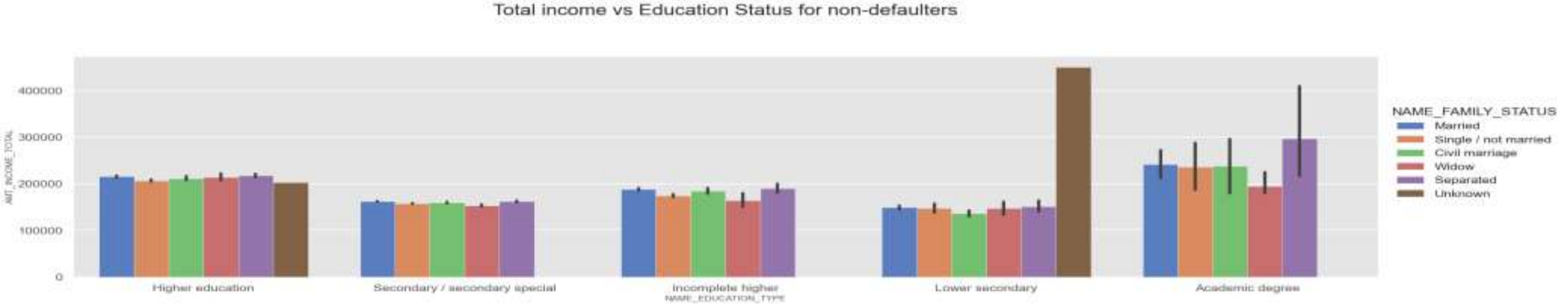


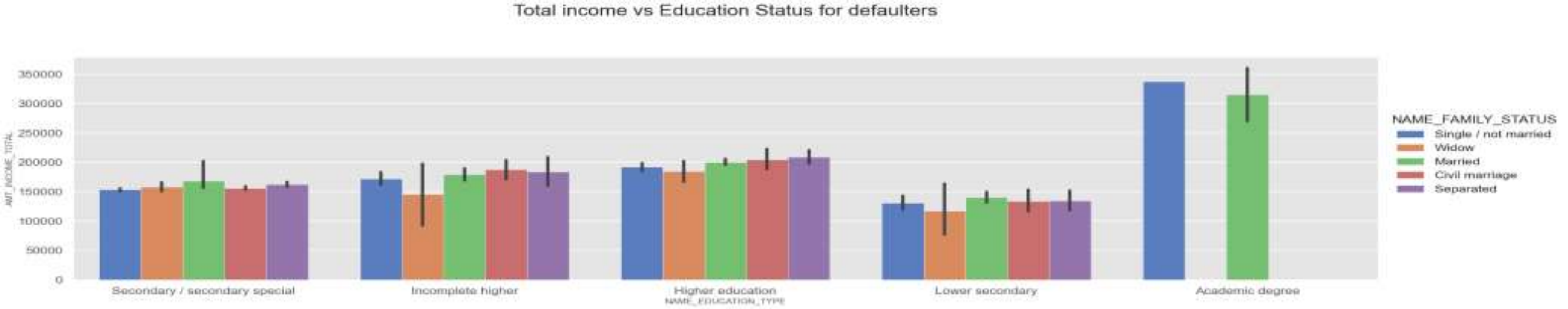


Inference: Credit amount vs Education Status Analysis

- DEFAULTER:
  - Persons falling under married, single and having academic degree are most likely to get on defaulters list.
  - Among all the education categories, married persons are most likely to be on the defaulters list.
  - Persons with lower educational background will tend to have low credit amount.
- NON-DEFAULTER:
  - Among all the categories, Lower secondary single/not married persons are having lower credit amount.
  - Person with Education status as Academic Degree relatively have higher credit amount with civil marriage segment considered the best among all.
  - Married persons are having greater credit amount among all education categories other than academic degree status.

Total Income vs Education status



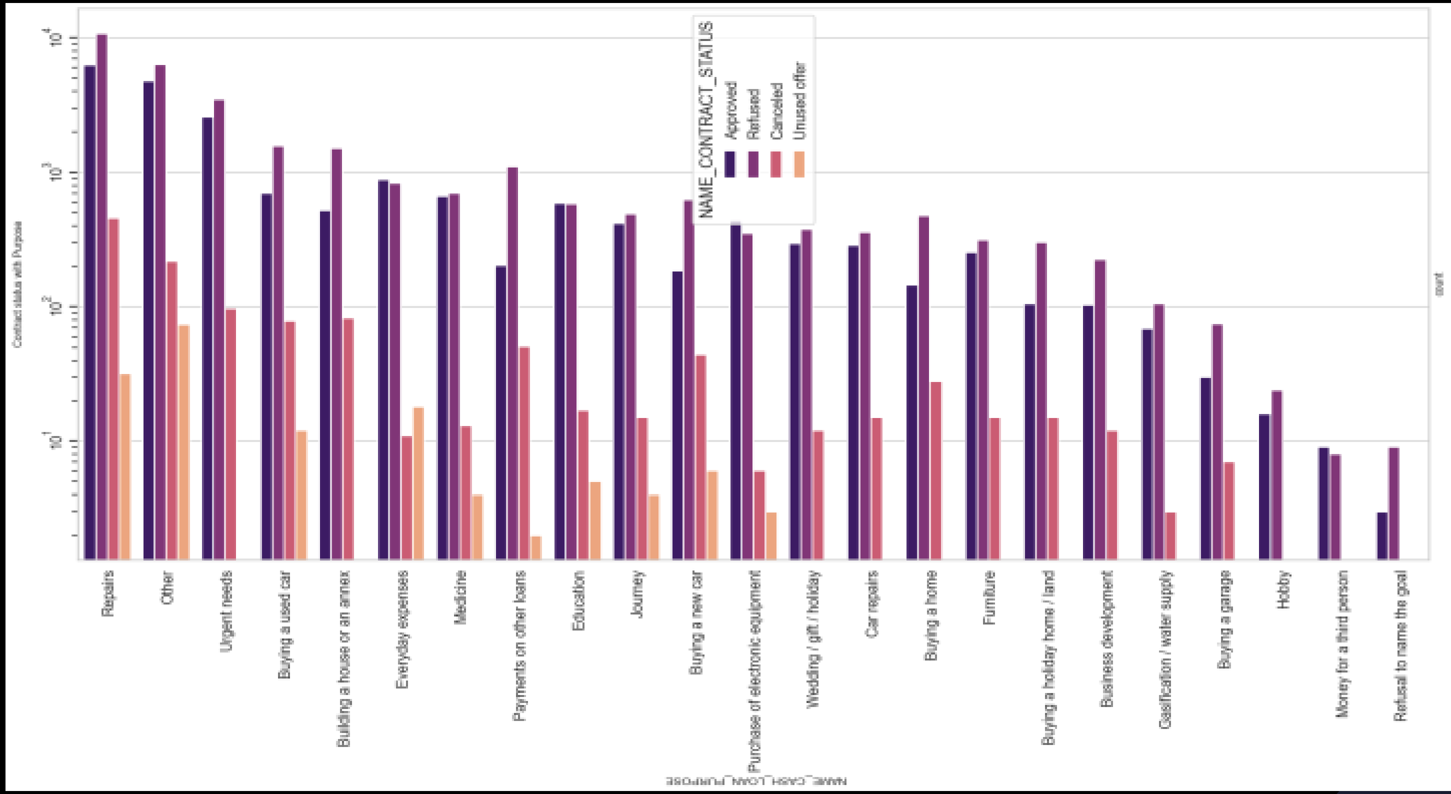


Inference: Total Income vs Education Status

- DEFAULTERS:
  - Under academic degree, we just have two segments i.e., single and married who are likely to be defaulters.
  - Person with lower secondary education are less likely to be defaulters. We may infer that as they are getting less income, so defaulters' segment is also less.
  - Married persons among all education group are more likely to be defaulters but single person from the academic degree is the highest among all.
- NON-DEFAULTERS:
  - We can observe that persons with academic degree are having more income than any other segment but unknown family status with lower education is having the highest income among all.
  - Lower secondary education are comparatively less income than any other sources.
  - Usually separated persons are having more total income than any other status.

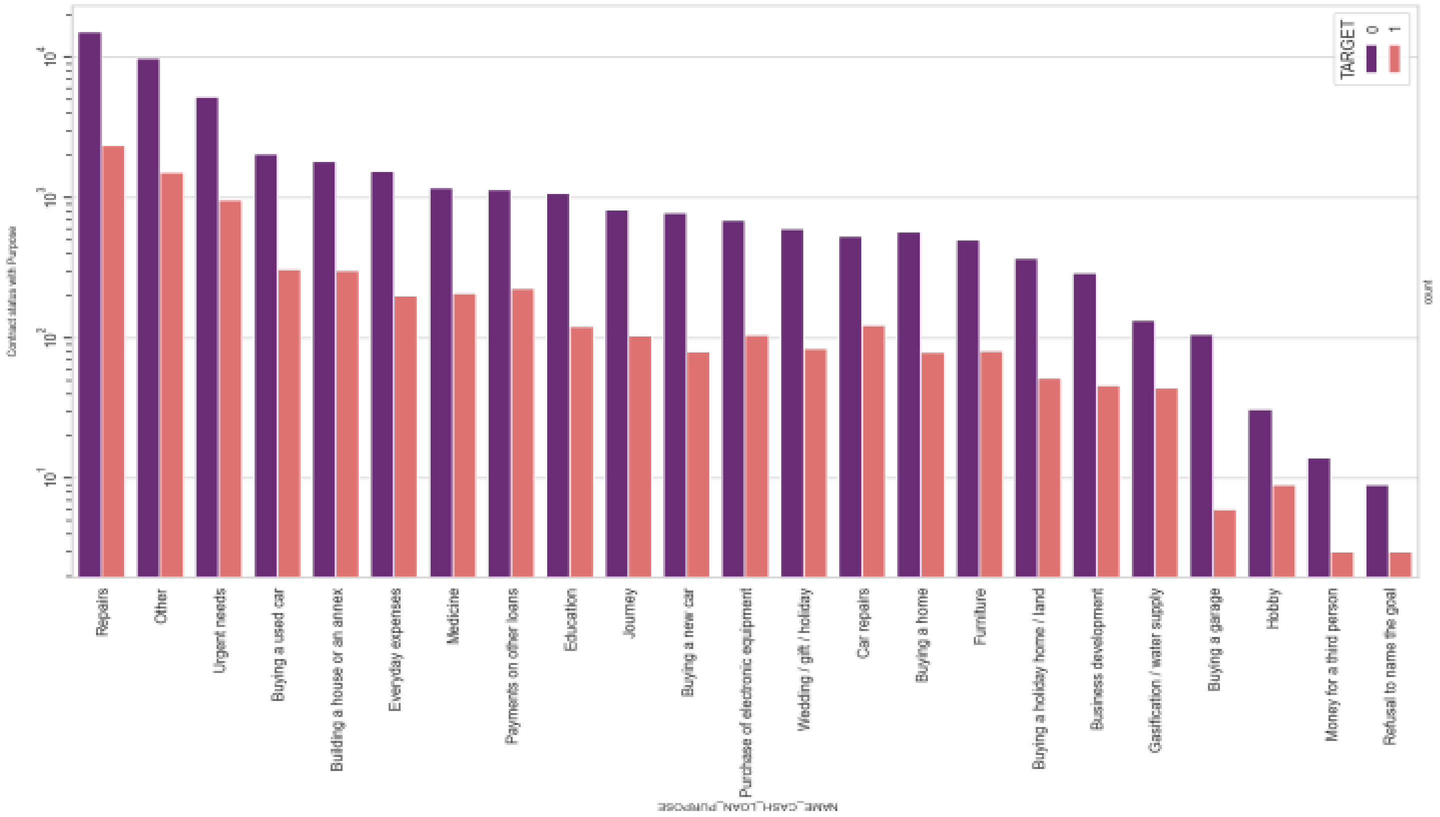
Loan Status vs Loan Purpose

- The greatest number of rejection and approval of loan is for repair purpose.
- Those person who refused to name the goal have least number of approval of loan and a greater number of rejections associated.
- In many cases, rejection is more than approved case except for everyday expense loan, education loan, purchase of electronic equipment loan, and money for third person.



Target vs Loan Purpose

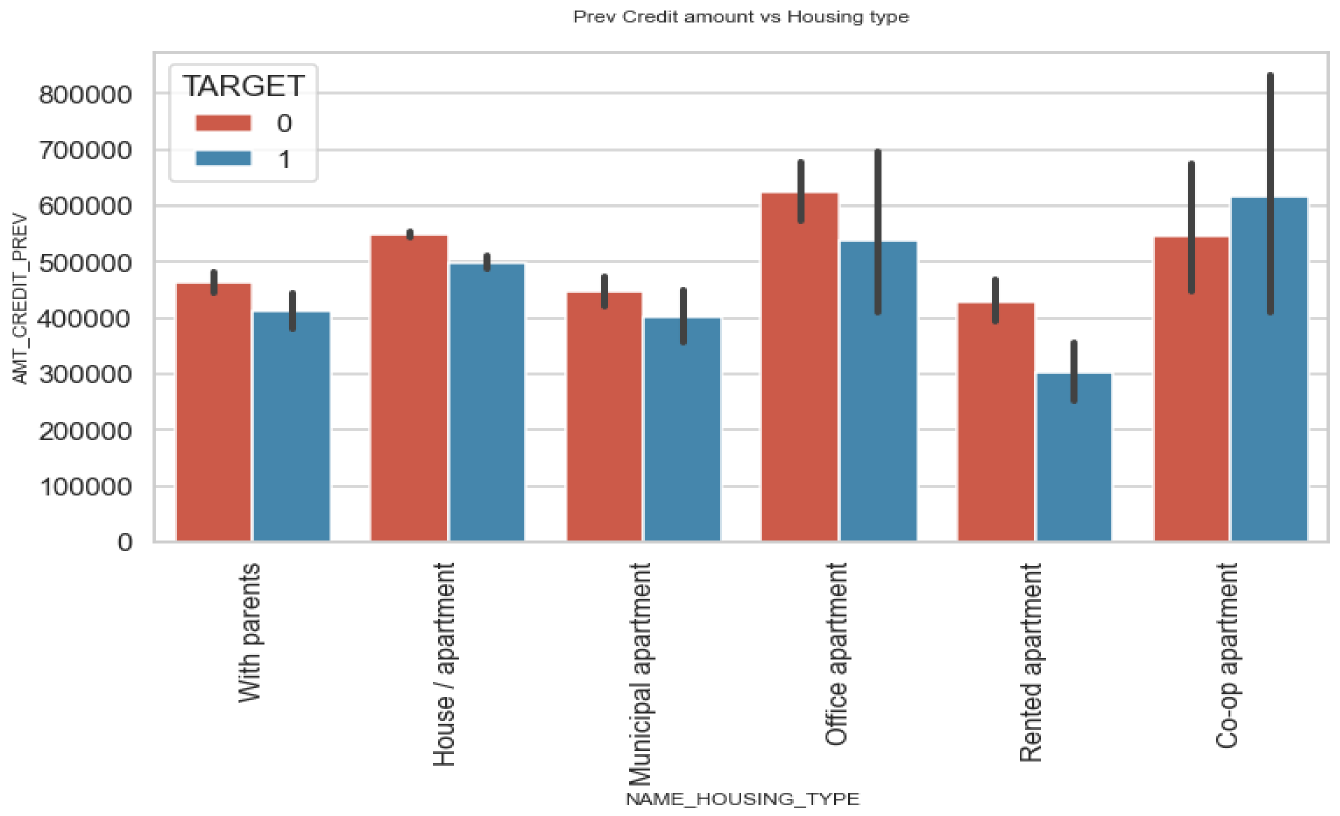
- Usually person taking loan for repair work, other works, urgent needs are most likely to have on defaulter list.
- Persons taking loan for "Refusal to name the goal", "money for third person", "Hobby" are less likely to be on defaulter list. It may be inferred that these persons are getting less approval for loan so less likely to get on defaulter list.



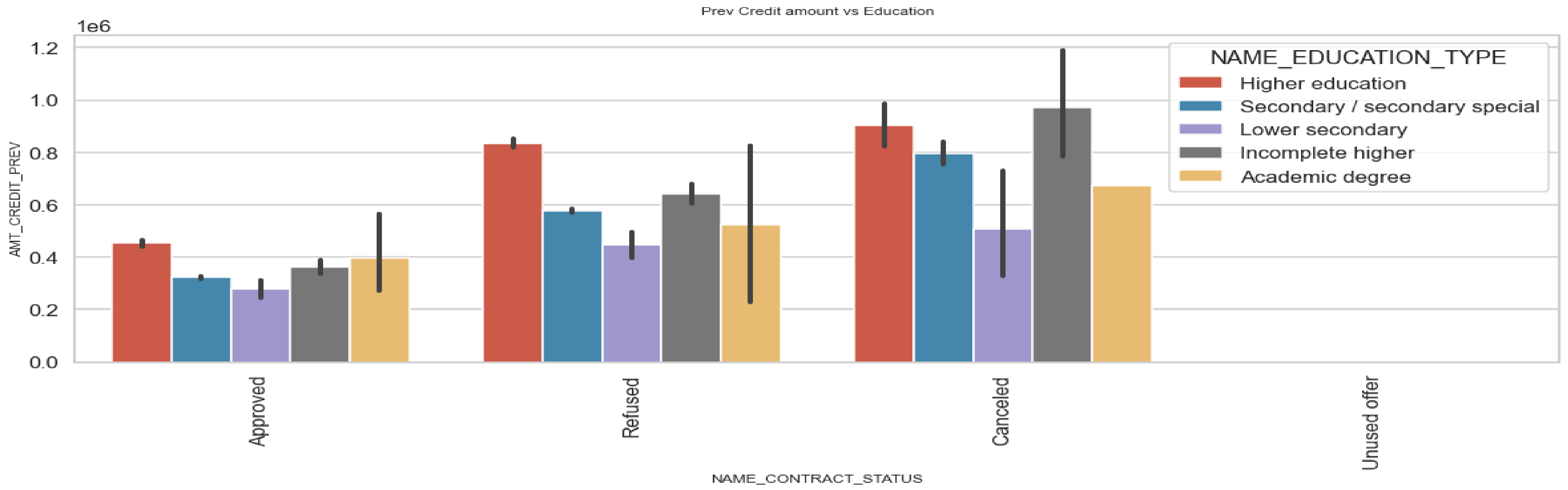
Target vs Previous Credit Amount

- Persons staying in Office apartment are having highest credit. It may be inferred that, persons in office apartment might be getting salary in time are most likely to be non-defaulters than any other group.
- Bank should avoid giving loans to co-op apartment as they are the persons who are most likely to get on to defaulters list.
- Bank can focus on other type of housing like:
    - Staying with parents
    - House/apartment
    - Municipal apartment
    - Rented apartment

in order to get successful payment

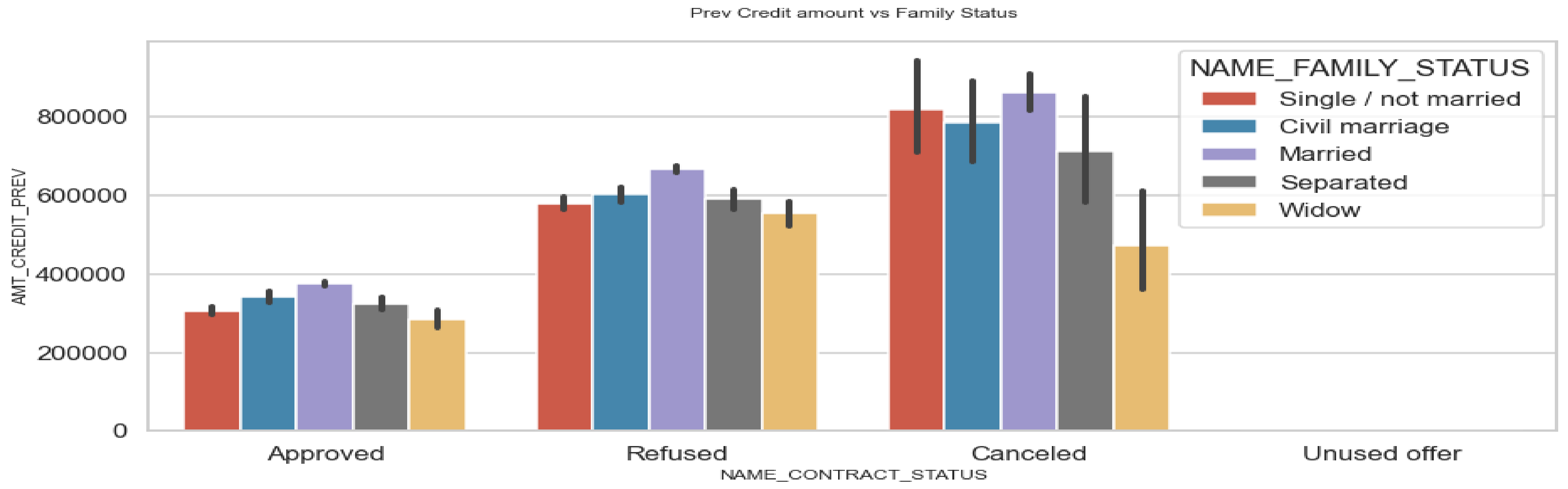


Education vs Previous Credit Amount



- Mostly client with higher educations are applying for loan and it gets approved. It may be inferred that as they are applying for more loan, more number of refusals and cancellations of loan is there.
- Clients with Academic degree are less likely to get refused or cancelled.

Education vs Previous Credit Amount



- Mostly married clients are safe to credit loan.

Conclusion

- If the prior application was approved, clients with better education can obtain loans without risk.
- In order to provide loans, banks should put more of an emphasis on higher education rather than lesser secondary schools, which often have payment issues.
- The bank should concentrate on customers in the 40 to 70 age range as they are more likely to have stable finances and not be on the list of defaulters.
- Bank should focus on the following group to credit loan as they are less likely to be defaulters:
  - Old people of any income group
  - Client whose previous loan was approved
  - Client with high income
  - Staying with parents, house/apartments, municipal apartments
  - Clients who are taking loan for repair work and his/her previous loan was approved
- Bank should avoid the following group to credit loan as they are more likely to be defaulters:
  - Lower secondary educated client
  - Client whose previous loan status was refused
  - Clients staying in Co-op apartment


THANK YOU