CREDIT EDA CASE-STUDY

Riya Mehta

STEPS PERFORMED

1

Data Cleaning of
Dataset 1
(Replacing missing values and
dropping values)

2

Visualization of
Dataset 1
(Observing different
trends)

3

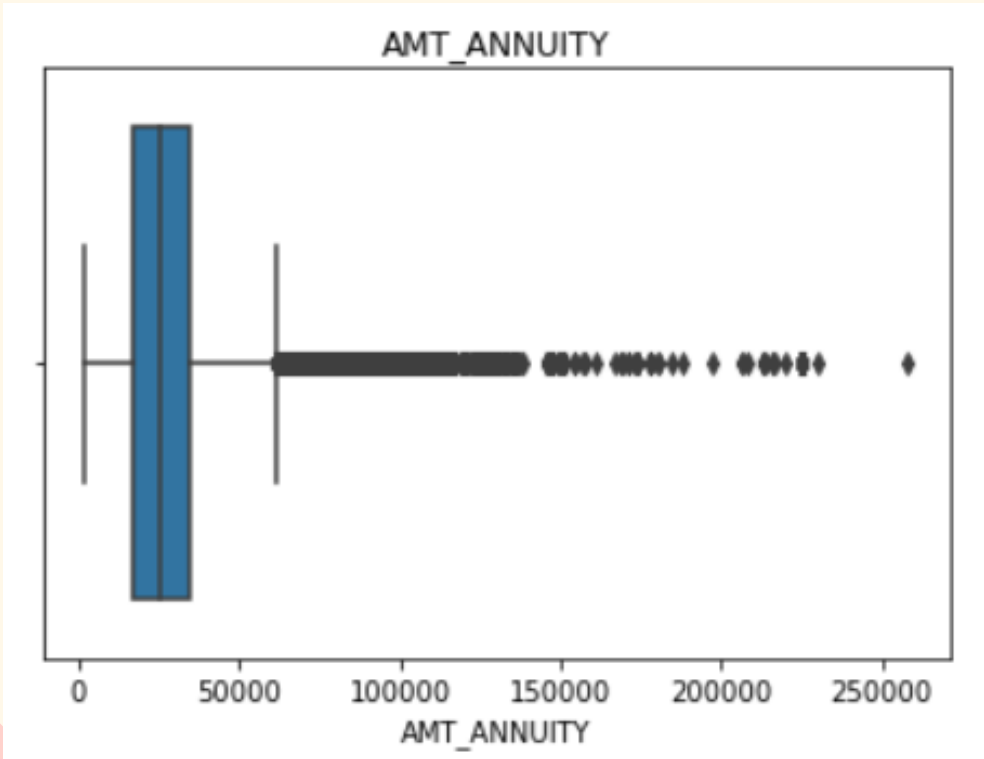Binning, Univariate &
Bivariate Analysis
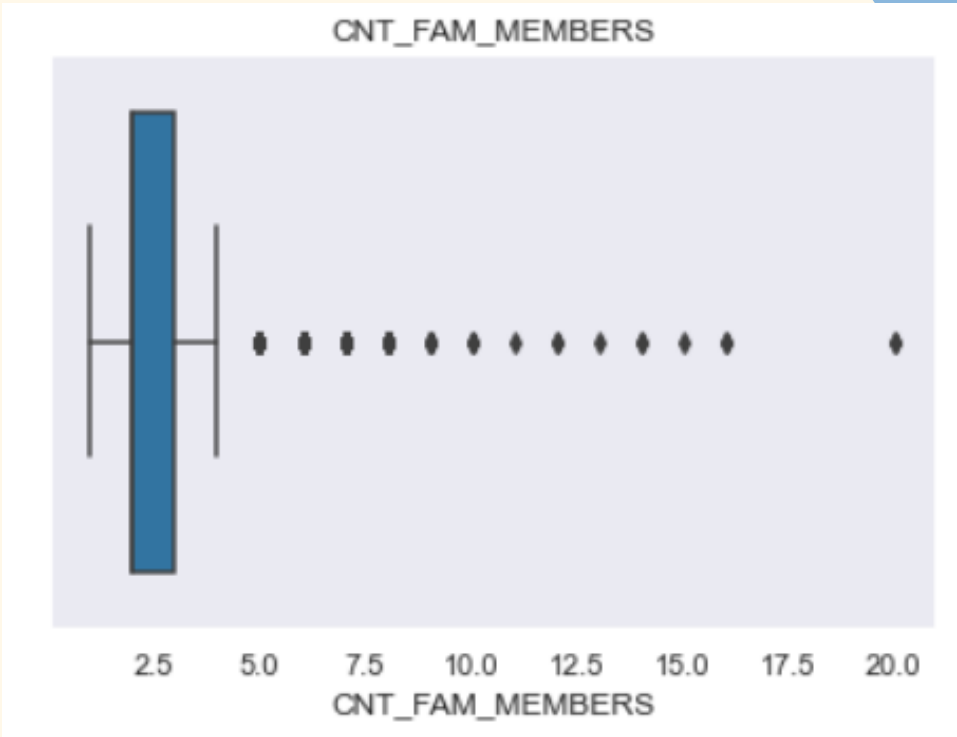(Creating bins and performing
analysis)

4

Data Cleaning of Dataset
2 and Univariate Analysis
(Dropping & replacing
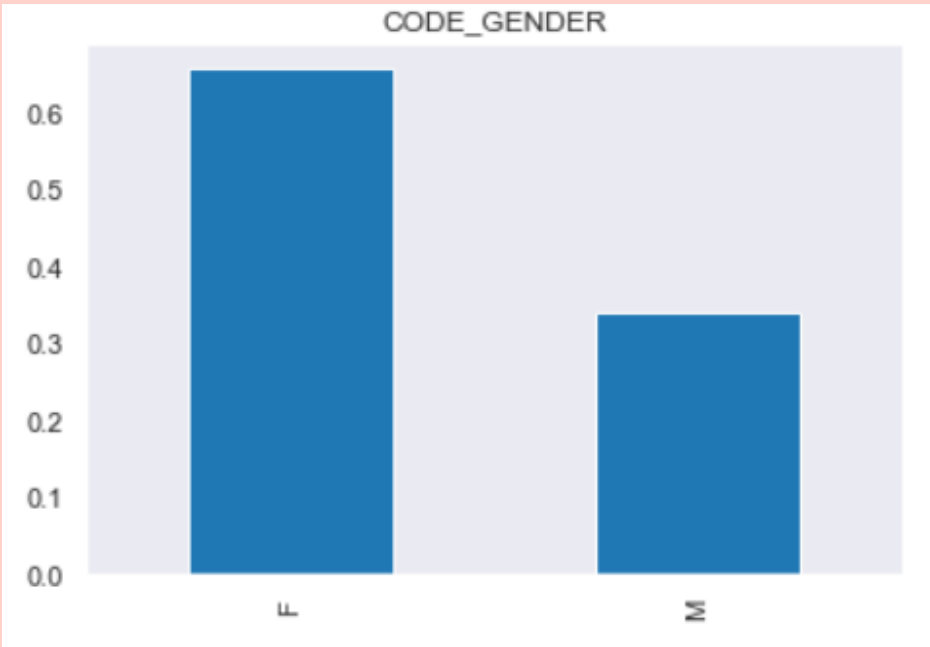missing values and drawing
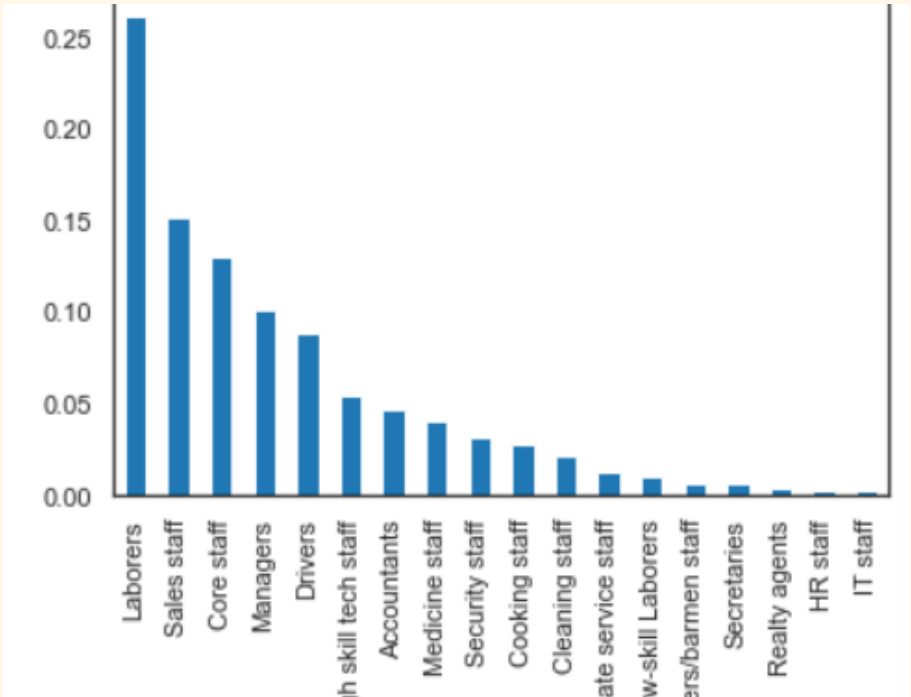conclusion)

Visualisation of Applicaion_data

In AMT_ANNUITY we observe that the
median, 25th and 75th percentile lies between 0-50,000

In CNT_FAM_MEMBERS we can observe that the
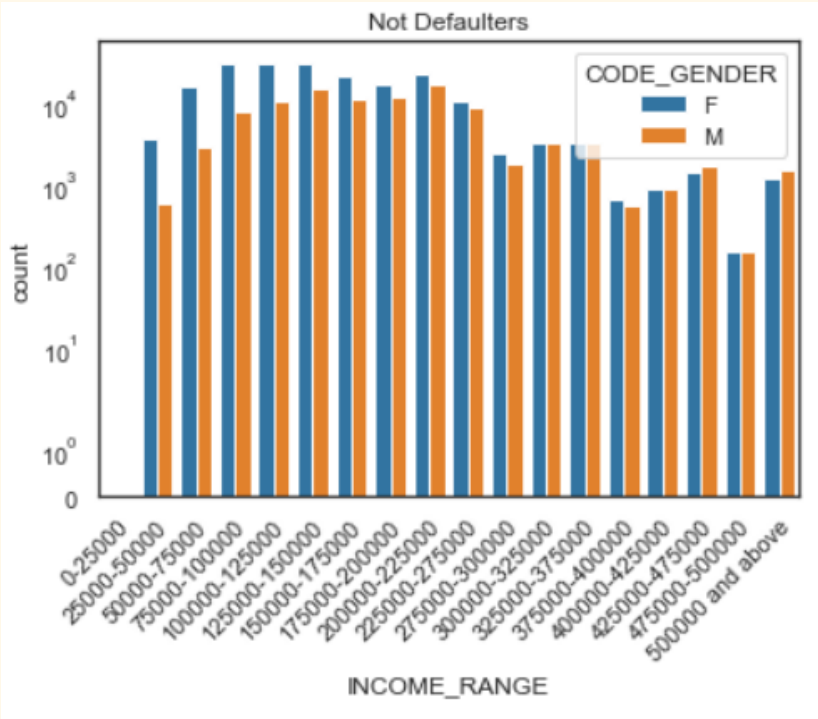majority of family members of the client lie in the
range of 2.5

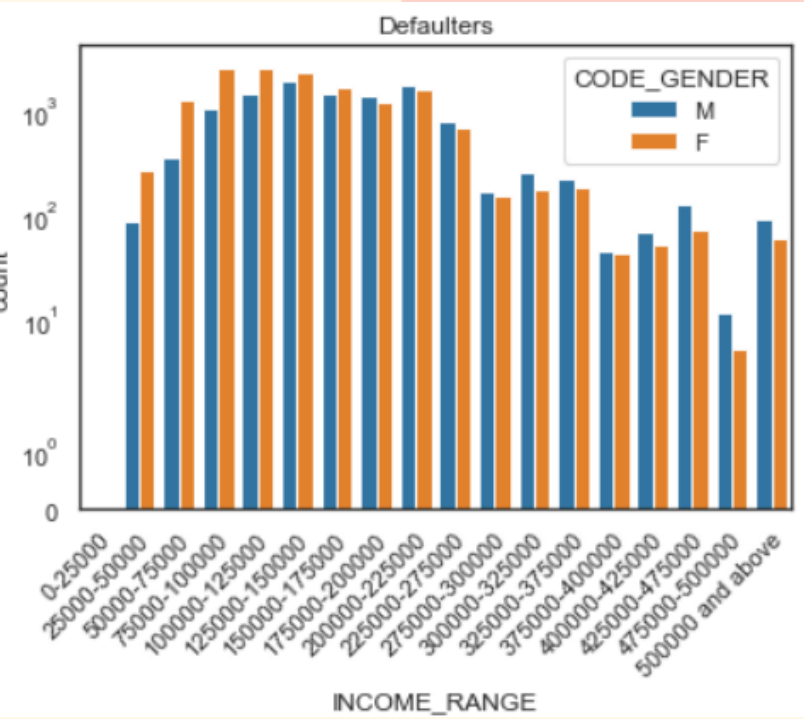In CODE_GENDER we can see that we have majority of Female clients

Here can we can observe that the majority of clients are Laborers and least are IT staff

Univariate Analysis of Application_data

The clients who are not defaulters have their maximum Income ranges between 100000-150000 for Females
& 100000-225000 for Males

The clients who are Defaulters have their maximum Income ranges between 100000-150000 for Females
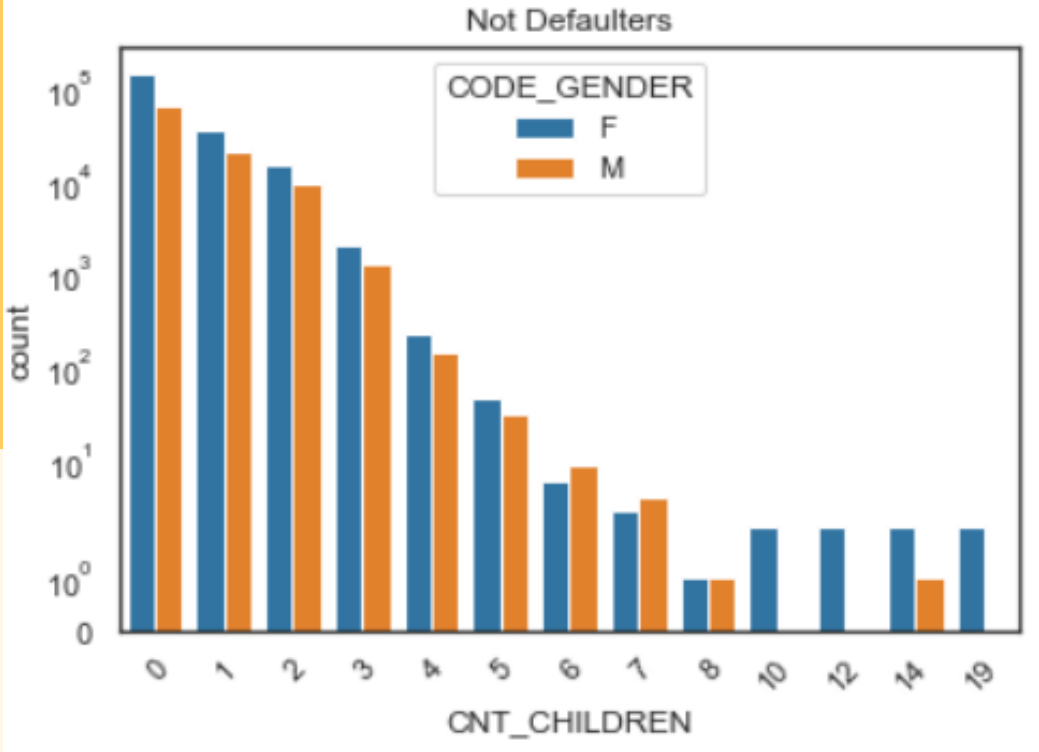& 100000-225000 for Males
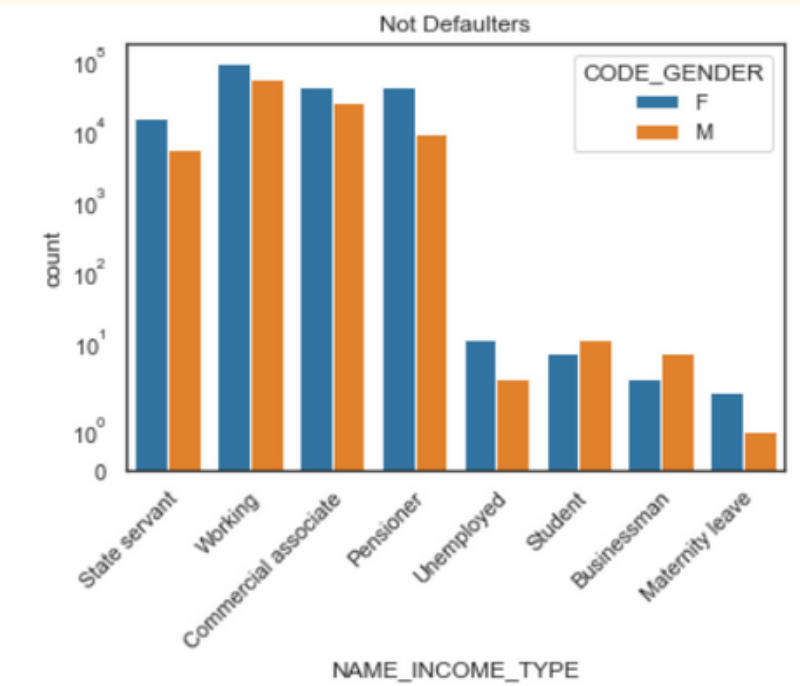
The clients who are Not Defaulters have their maximum income type as
Working, Commercial associate, Pensioner and state servant for Male and Female
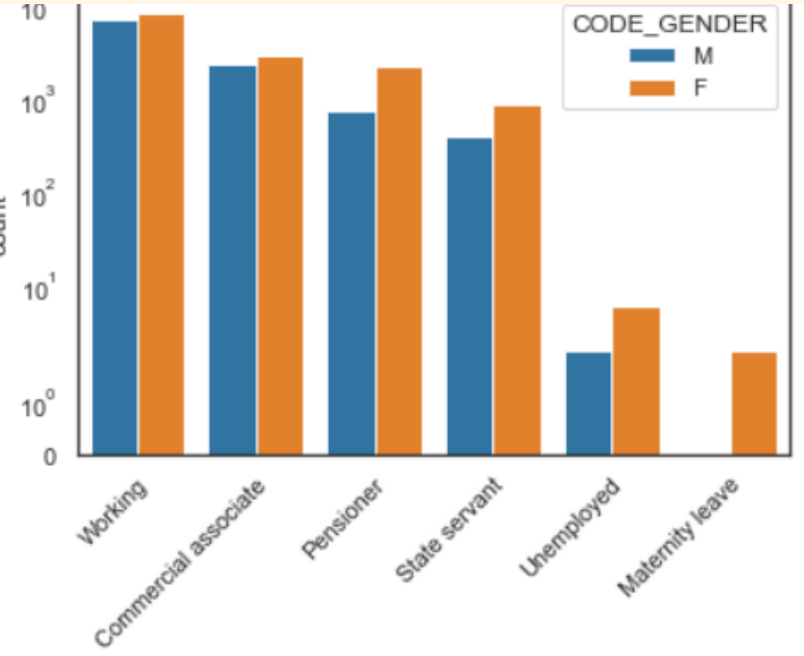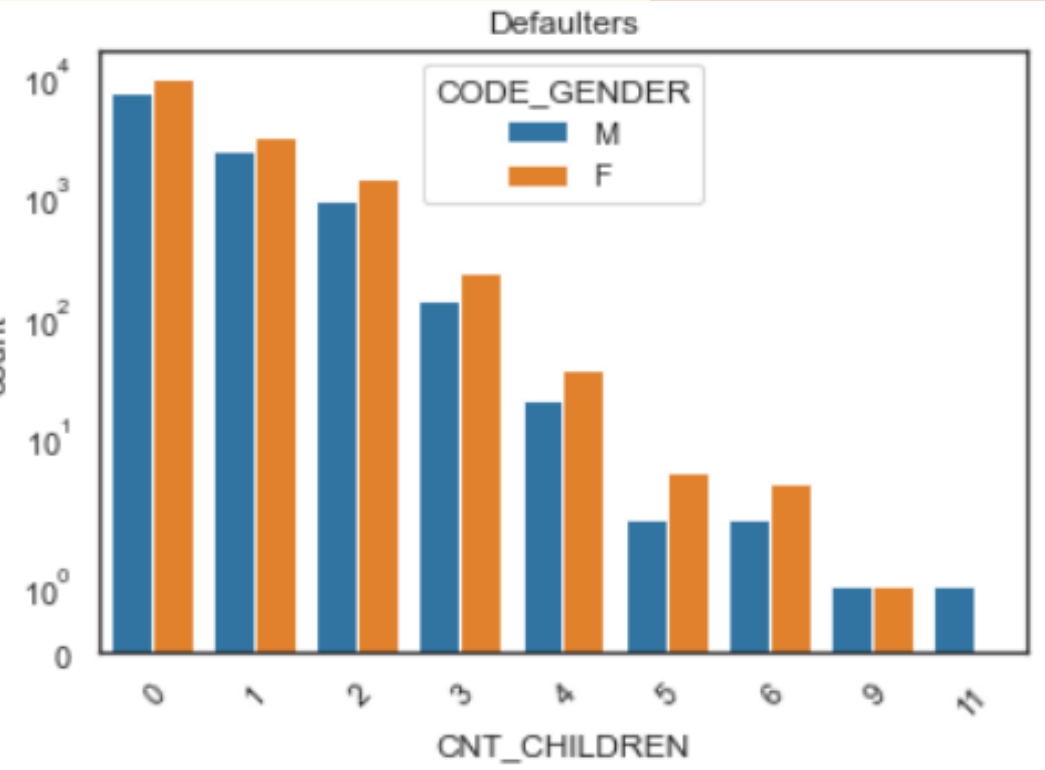

The clients who are Defaulters have their maximum income type as
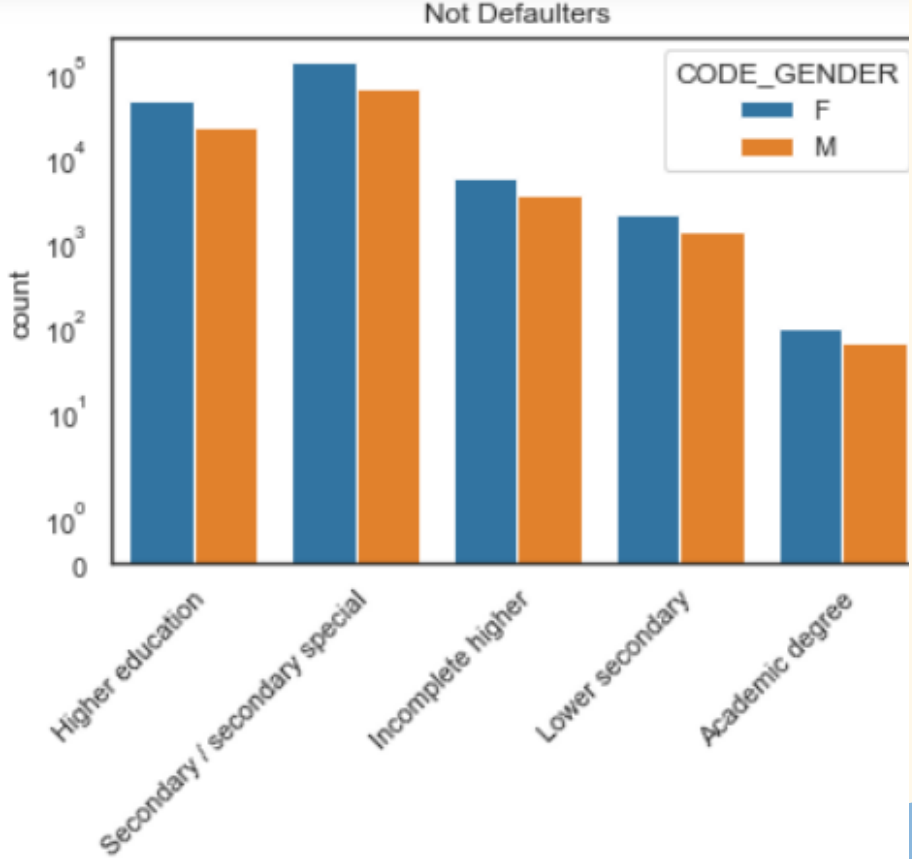Working, Commercial associate and Pensioner for Male and Female
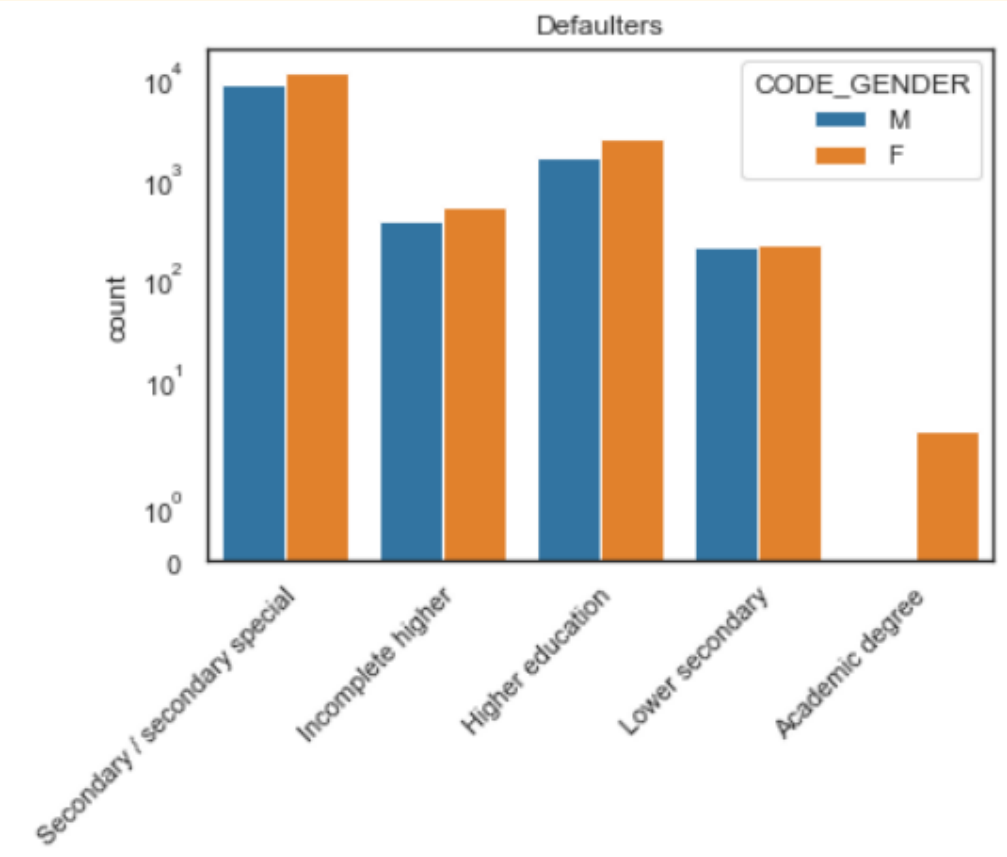
Univariate Analysis

Both Female & Male clients who are Not Defaulters have maximum range of 0 to1 children

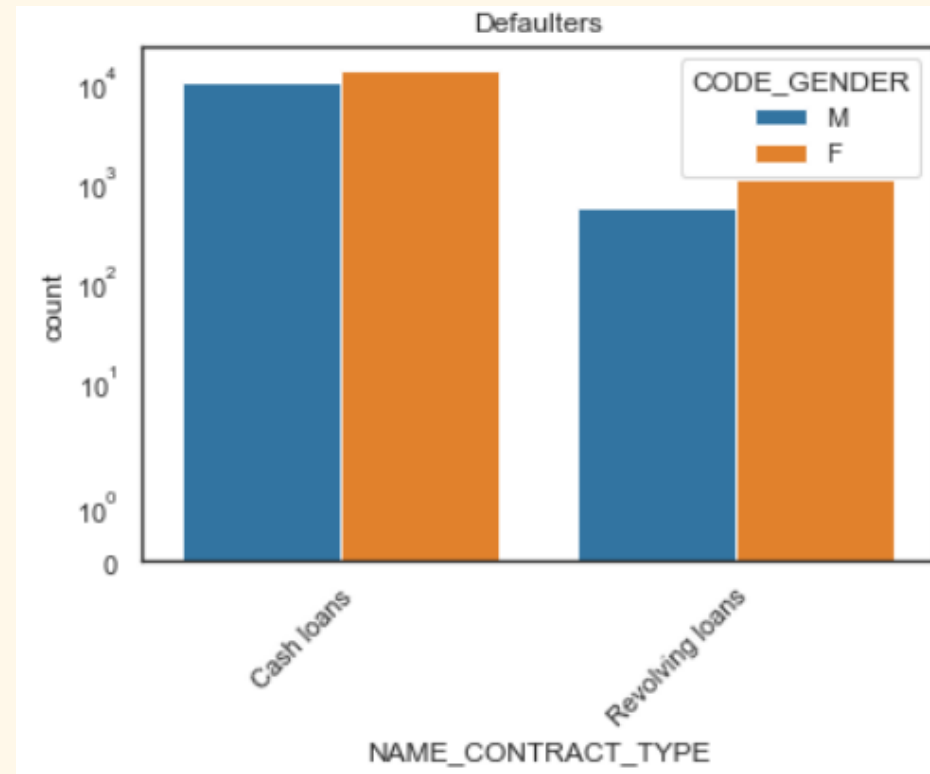Both Female & Male clients who are Defaulters have maximum range of 0 to1 children

For Non Defaulter clients they have a maximum education qualification of Higher and Secondary Education
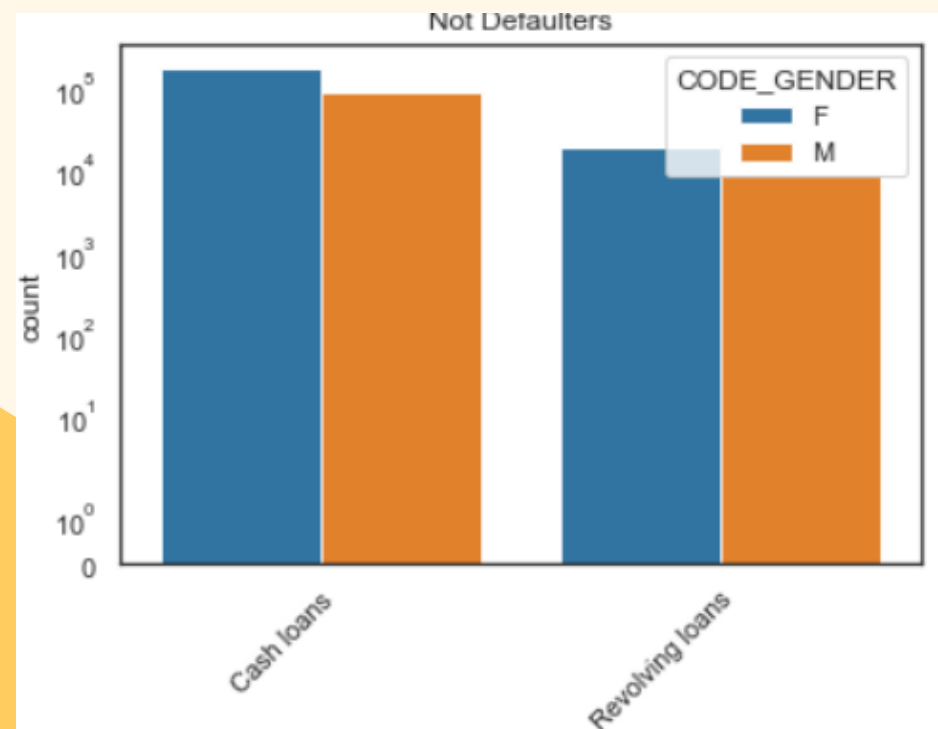

For Defaulter clients they have a maximum education qualification of Secondary Education
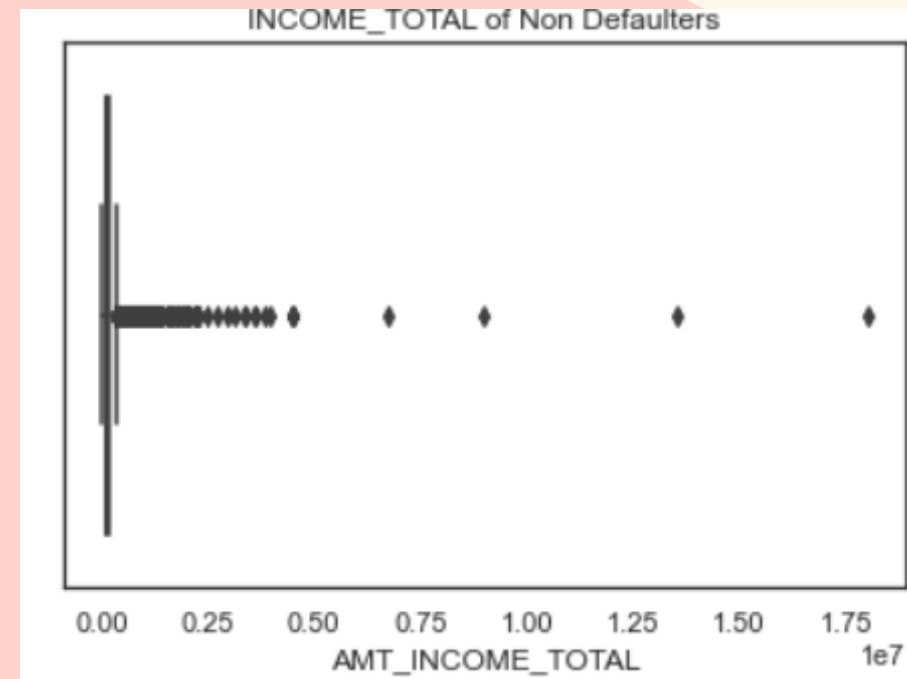
Univariate Analysis

**Maximum clients who are Defaulters have Contract type of Cash loans for both Male & Female clients
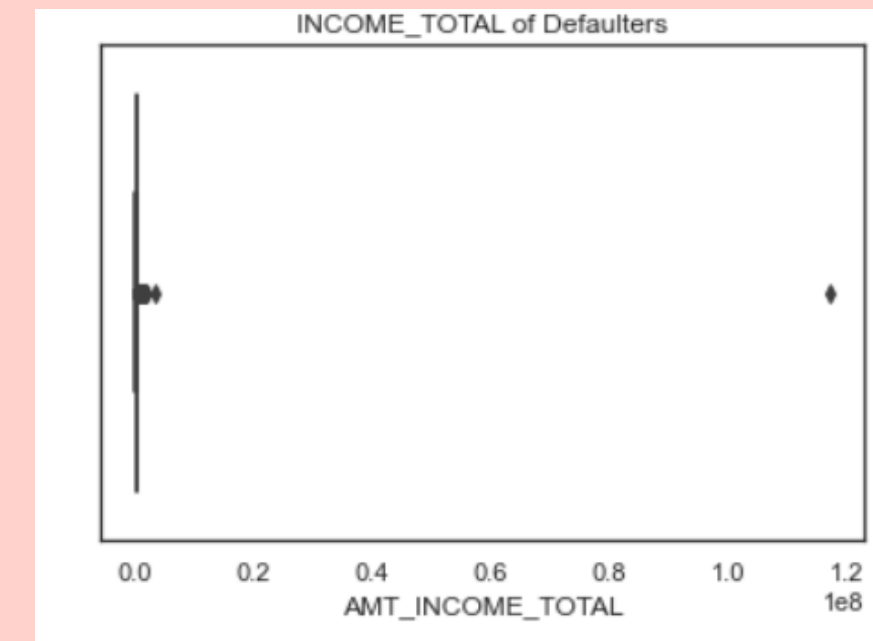In this case Females have higher number than Males**

**Maximum clients who are Non Defaulters have Contract type of Cash loans for both Male & Female clients
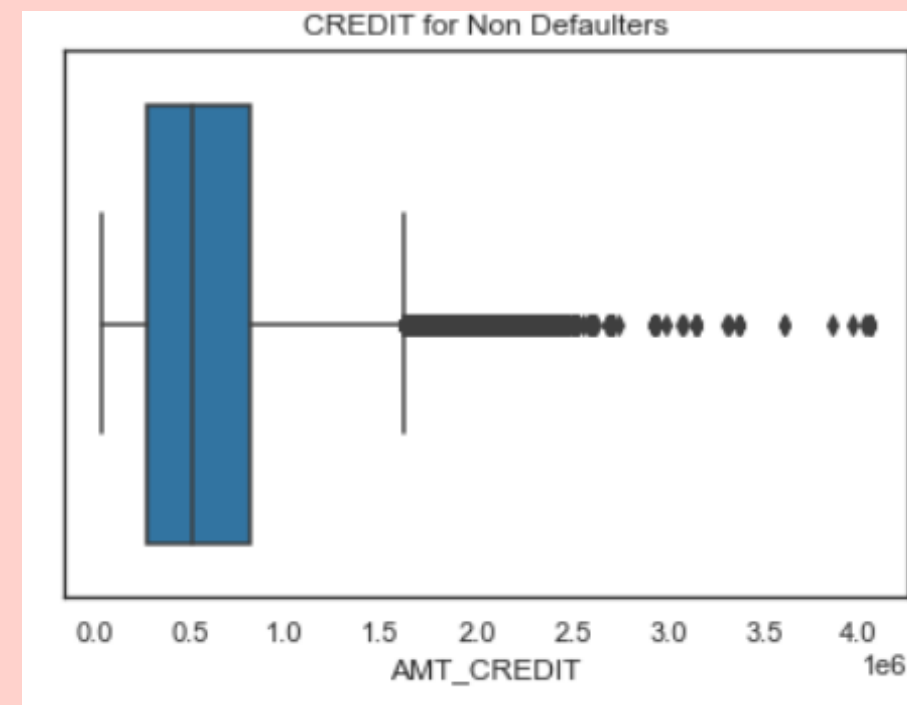In this case Males have higher number than Females**
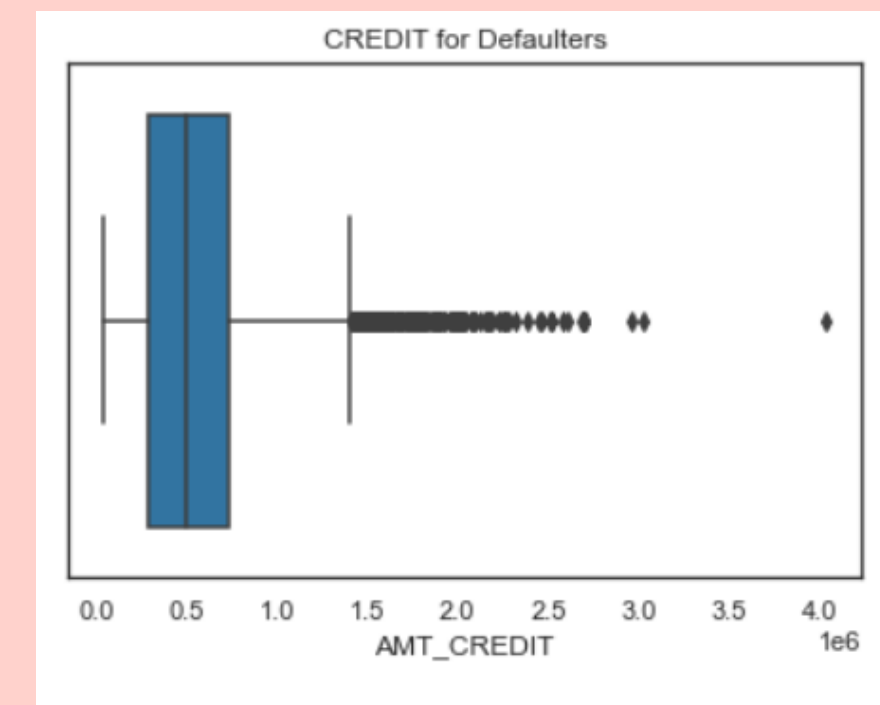
Univariate Analysis using boxplot

The Total income of maximum Non defaulters lies in the First quartile and consists of many Outliers too

The Total income of maximum defaulters lies in the First quartile
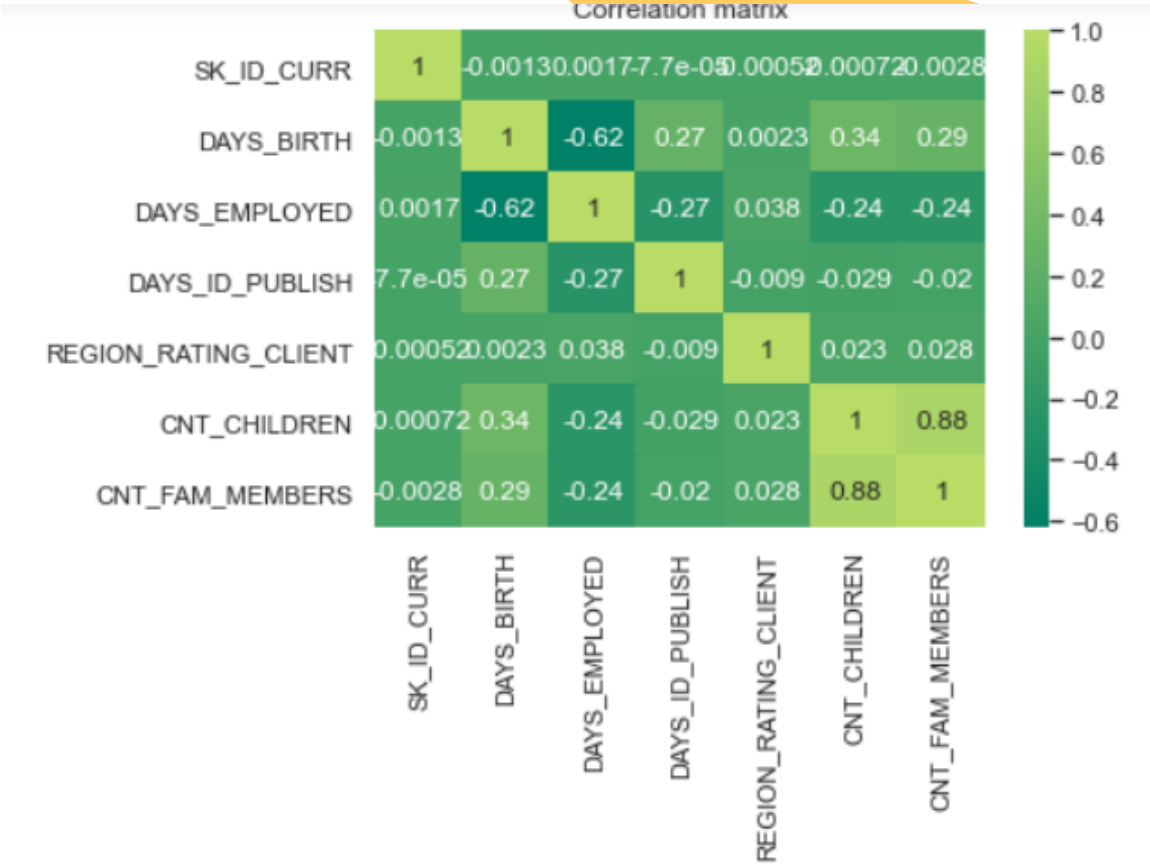
The Credit Amount of Non defaulters median lies on 0.5 and ranges from 0 to1. They lie in the first Quartile

The Credit Amount of Non defaulters median lies on 0.5 and ranges from 0 to1. They lie in the first Quartile

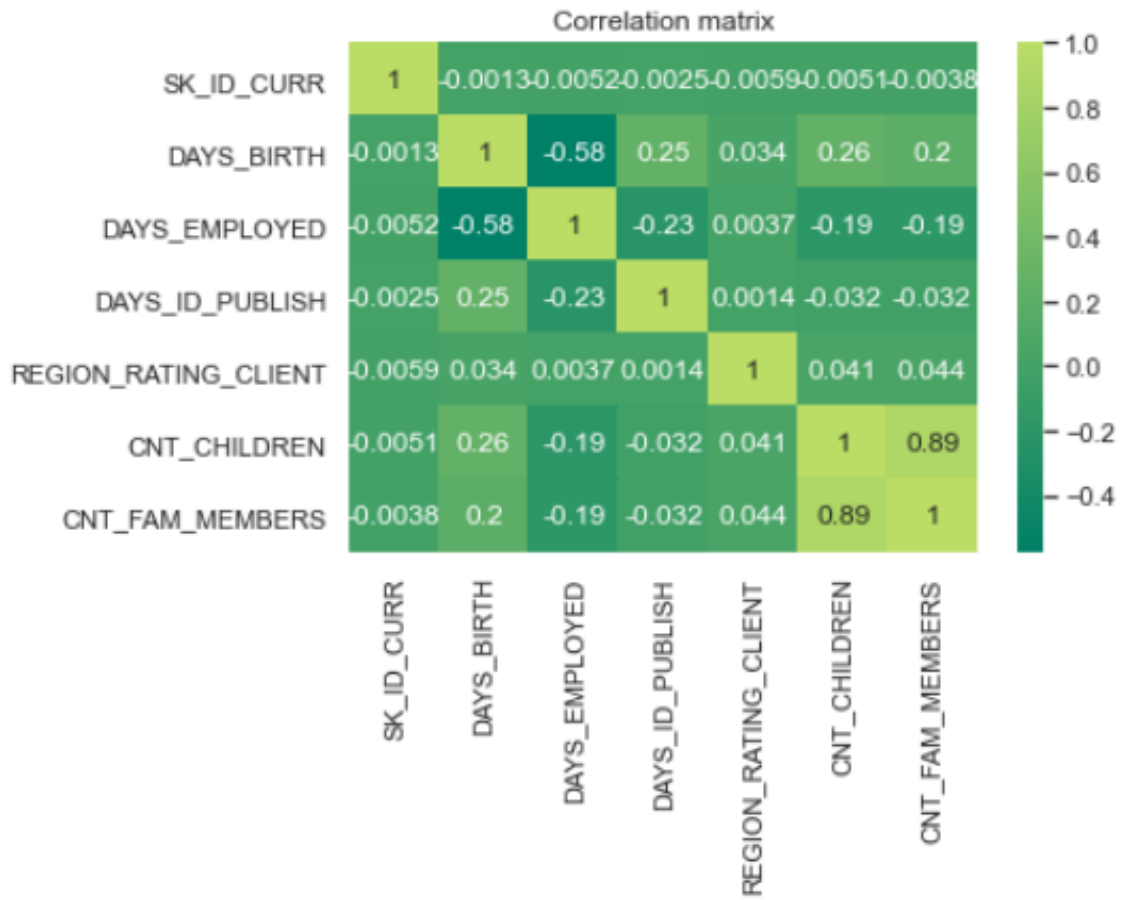# Bivariate Analysis Of Application_data

The High correlation value of Non Defaulters lies in
CNT_CHILDREN and CNT_FAM MEMBERS
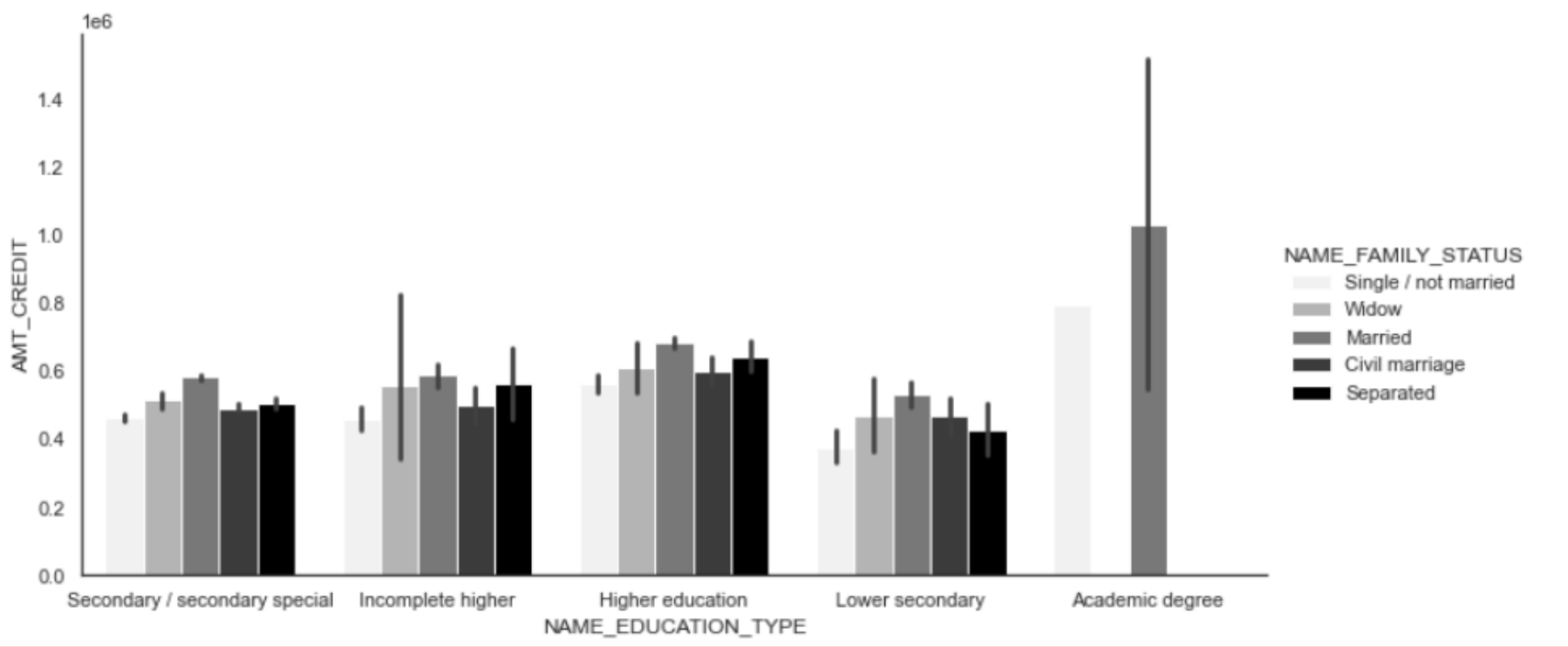CNT_CHILDREN and DAYS_BIRTH

Not Defaulters

The high Correlation value of Defaulters lies in
CNT_CHILDREN and CNT_FAM_MEMBERS

Defaulters

Bivariate Analysis

Defaulters

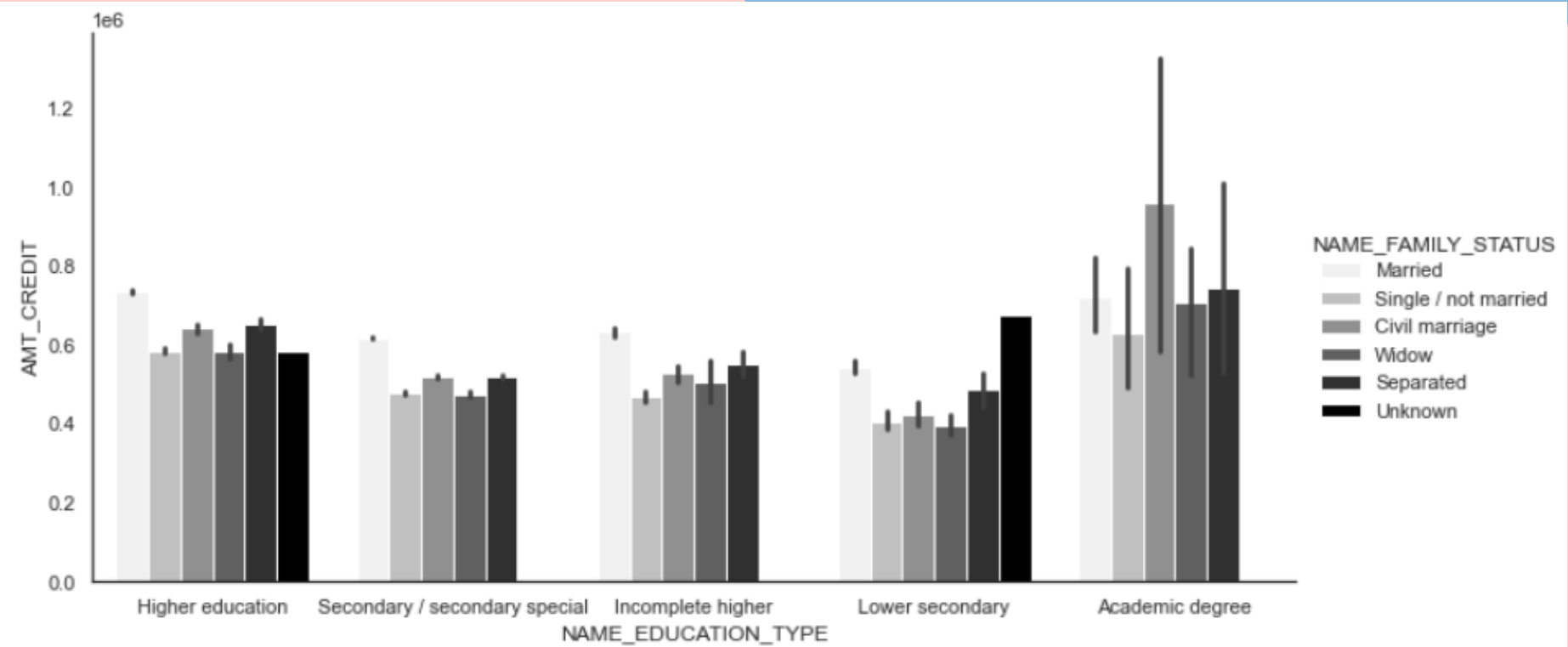**For Defaulters the
Credit amount is highest for those who hold a degree of
Academic degree and it is highest for
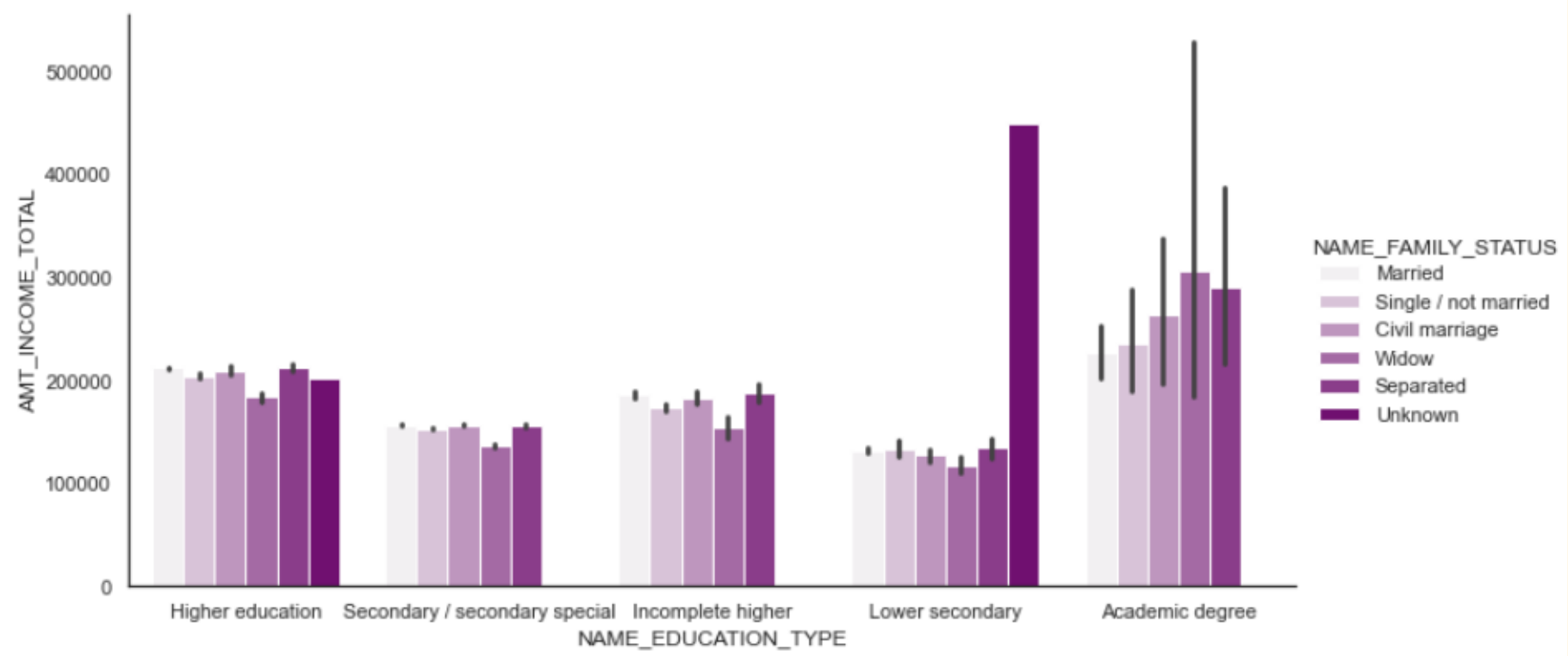Married people and least for Single people**

**For Non Defaulters the
Credit amount is highest for those who hold a degree of
Academic degree and it is highest for
Civil Marriage and least for Single people**

Non Defaulters

Bivariate Analysis


Non Defaulters

For Non defaulters the ones who have
Academic degree have the highest income under which
Widow have the highest income and Married have the least

For Defaulters the ones who have
Academic degree have the highest income under which
Single have the highest income and Married have the least
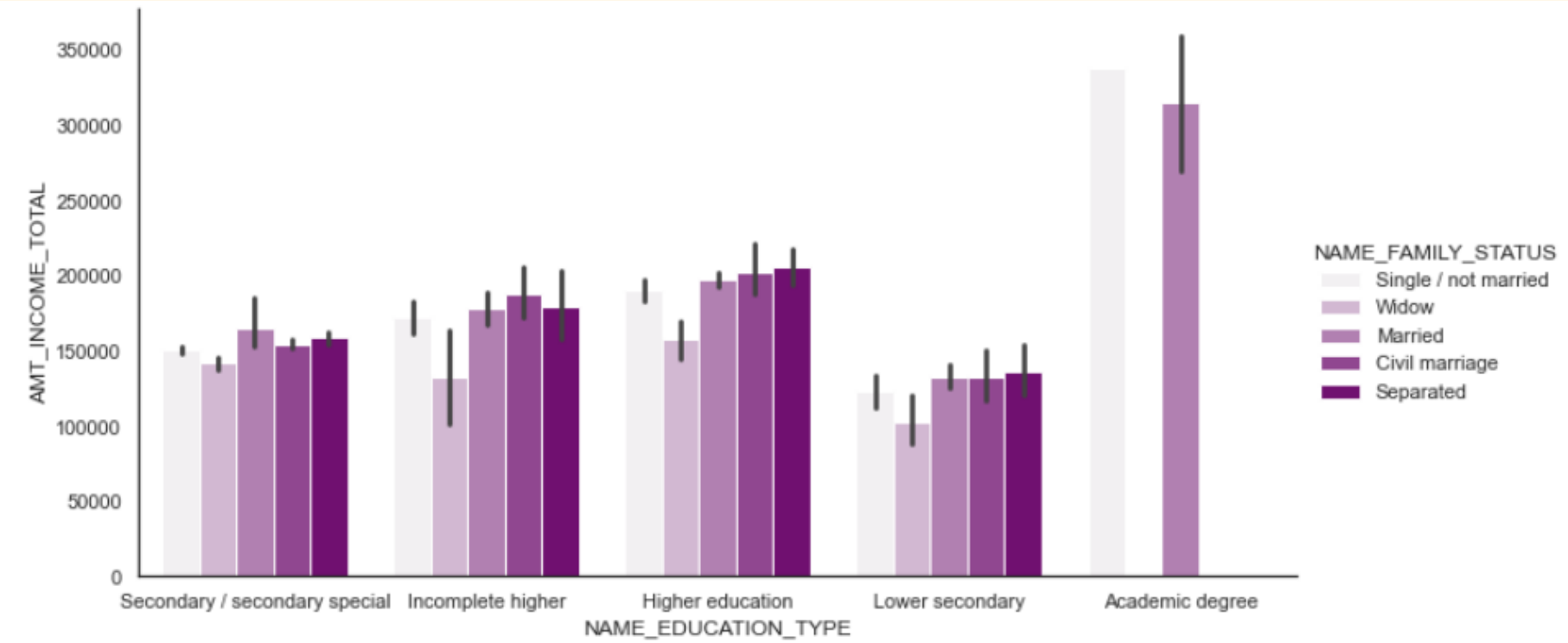
Defaulters

Univariate Analysis of Previous_application

Purpose of Contract Status
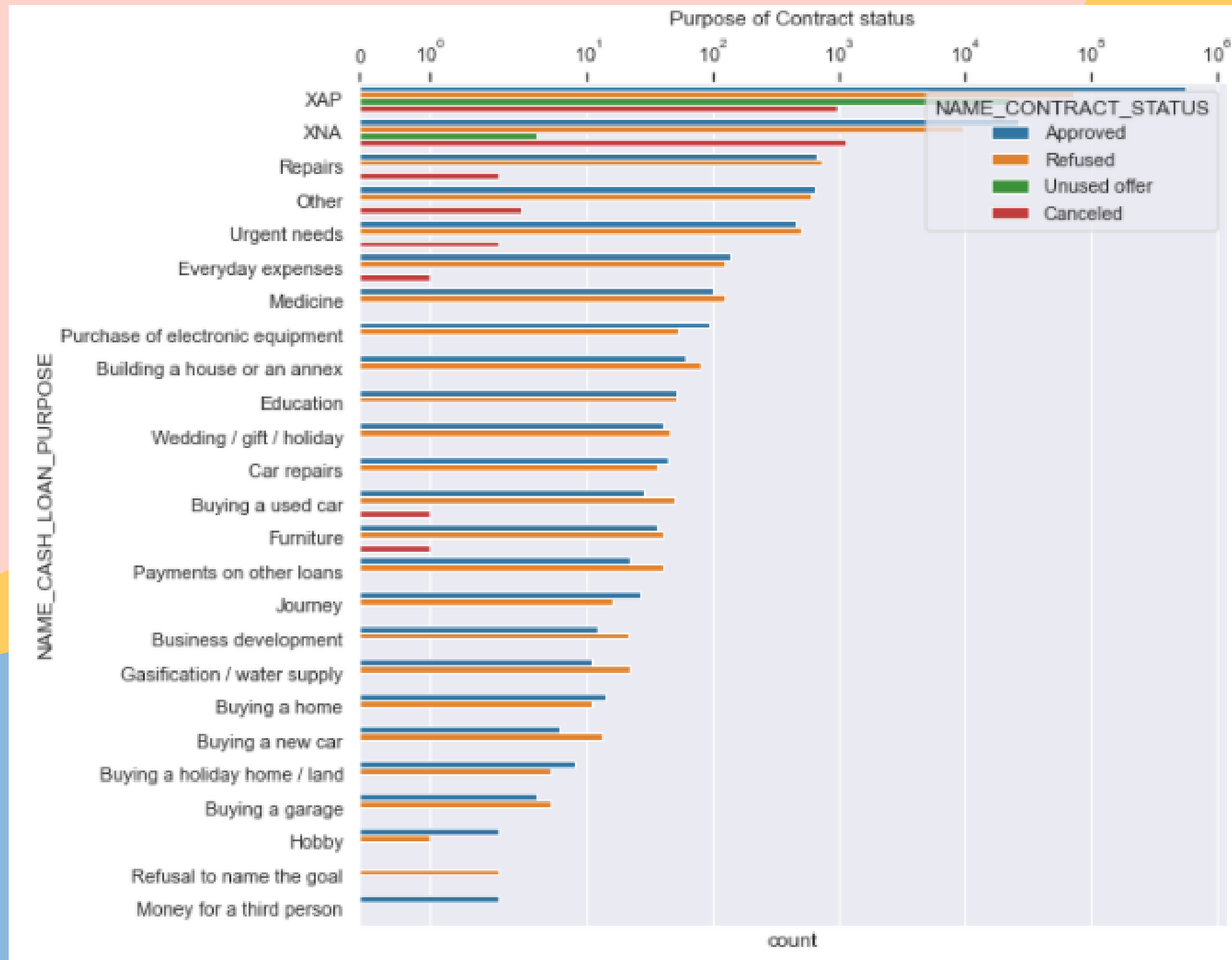
Here we can observe that -

Excluding XNA and XNP
The maximum purpose of the
clients has been Repair
and majority of it has been Refused
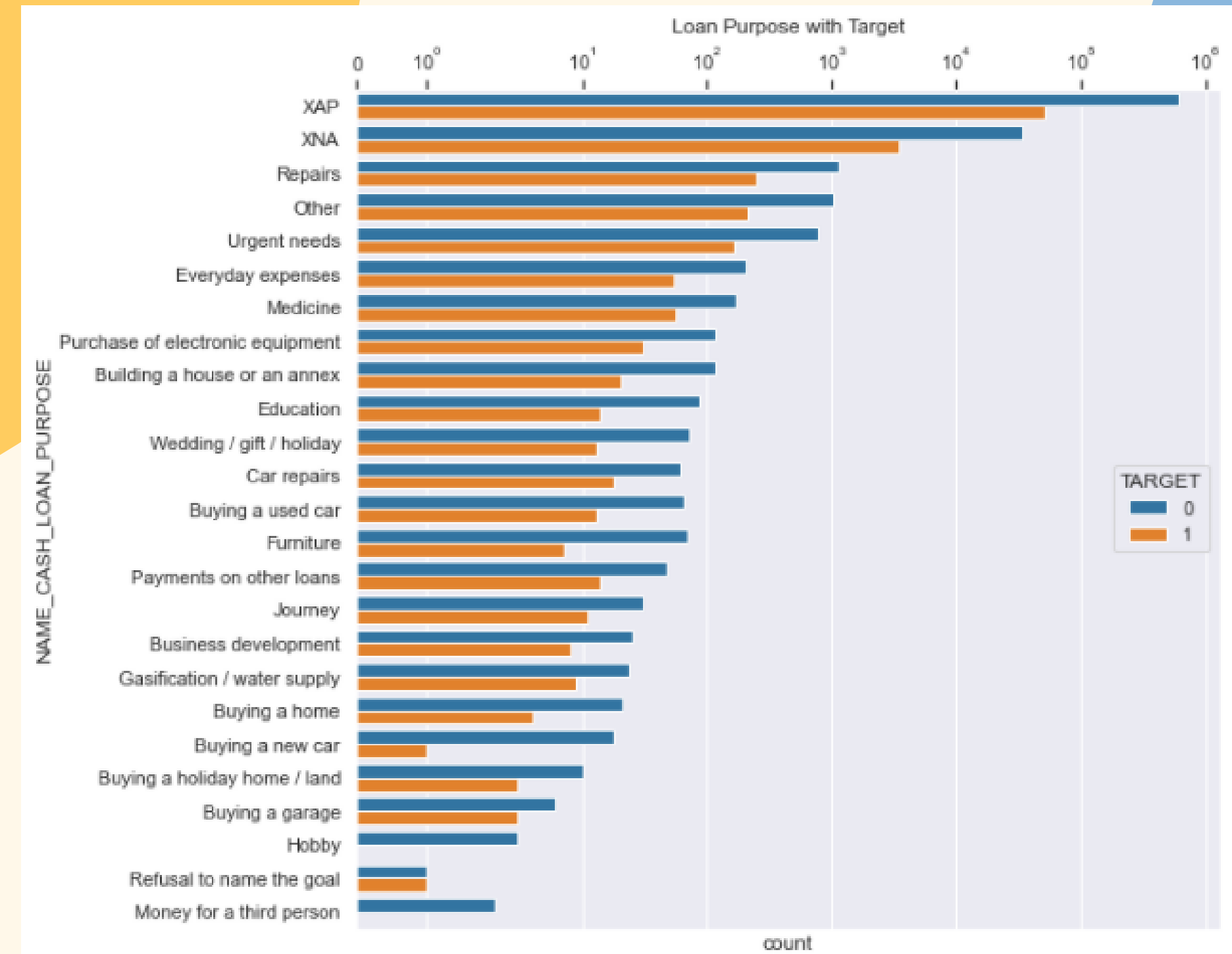Whereas in Education sector the No of Approval
and Rejection is same!

Univariate Analysis of Previous_application

Here we can observe that-

Excluding XNA and XAP
The maximum number of Non Defaulters are
in Repair and their Purpose is high too
whereas in 'Refusal to name goal' the number of
Defaulters and Non Defaulters is same
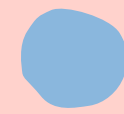

Loan Purpose with Target

CONCLUSION

- From the analysis we got to know that Female clients are more than Males in Non Defaulters
  In case of Defaulter Male clients are more than Female

- The No. of children does not impact the client from being a Defaulter or Non Defaulter, so we cannot conclude anything from this
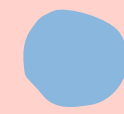
- According to the analysis the clients with Secondary education have defaulted the most.

- For both Defaulters and Non defaulters cash loan contract have higher number of Credits

- Clients having academic degree have high credit amount under which Civil marriage is highest whereas less educated have low credit which is widow for Non defaulters

- Clients having academic degree have high credit amount under which Married are highest defaulters whereas single clients are lowest defaulters
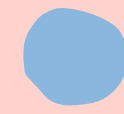
- We can conclude that the purpose 'Repair' has highest no of rejections whereas purpose 'education' has equal no. of approval and Rejection
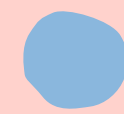
- The purpose 'Repair' have highest no. of Defaulters