



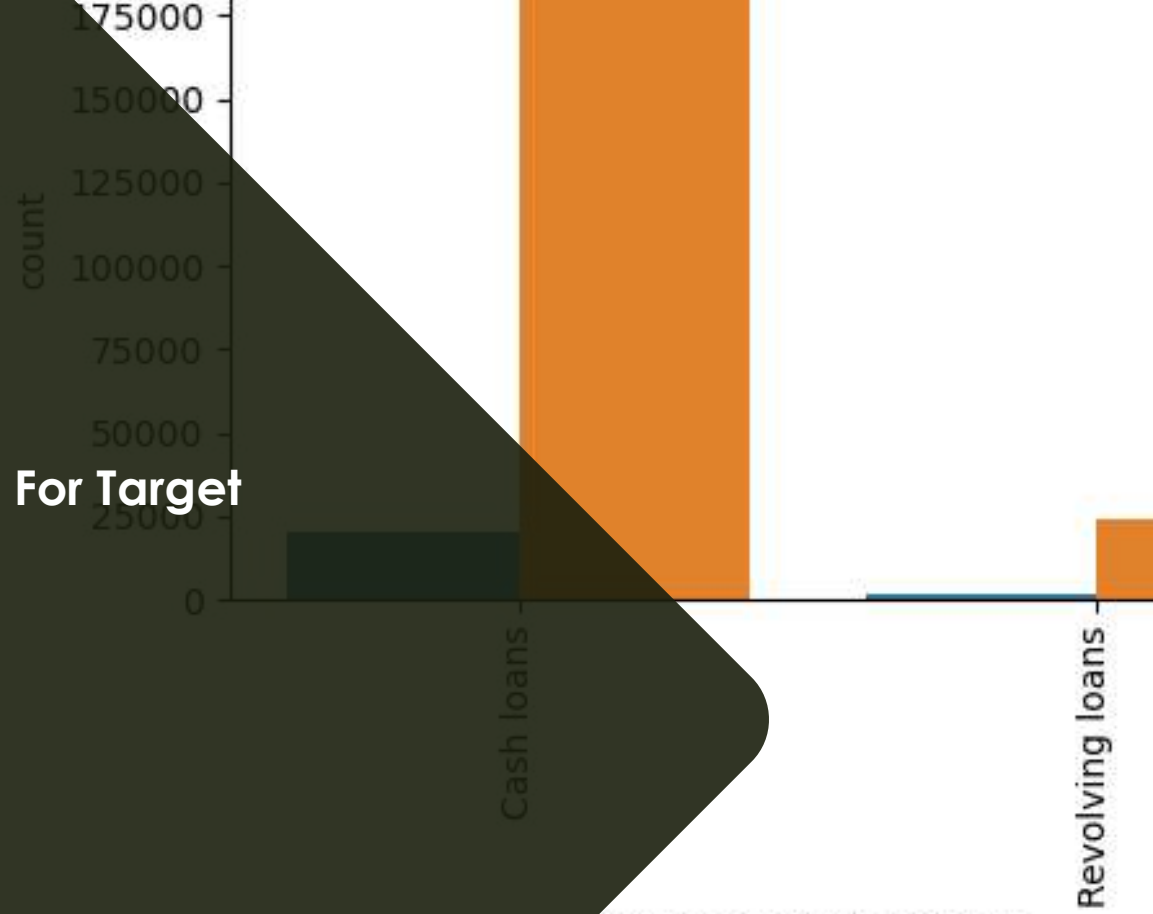
Credit EDA Case Study

By:-

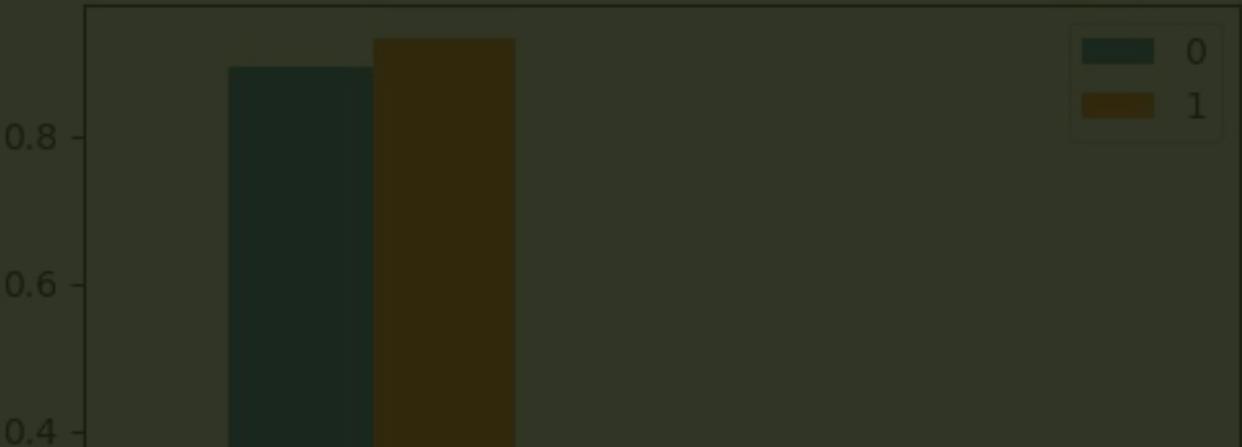
Achleshwar Pratap Singh

NAME_CONTRACT TYPE-

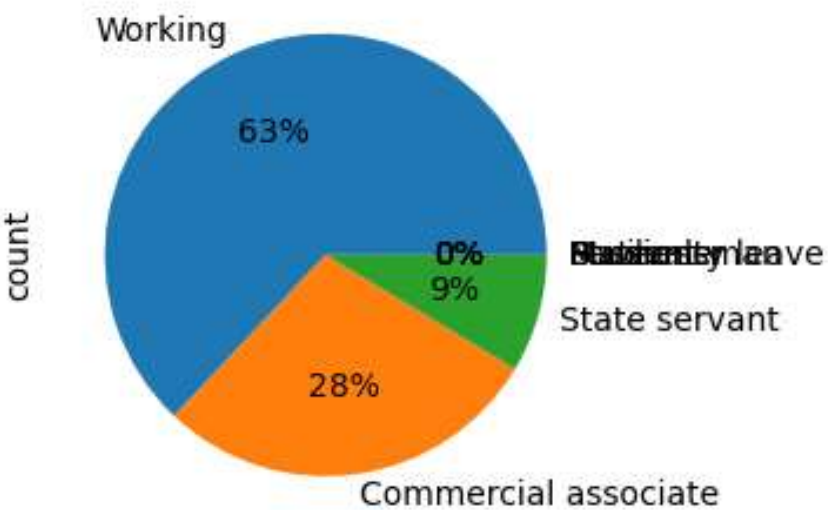
➤ Cash Loans are large part of the company's portfolio. For Target 0 - 85% and almost 95% for Target-1



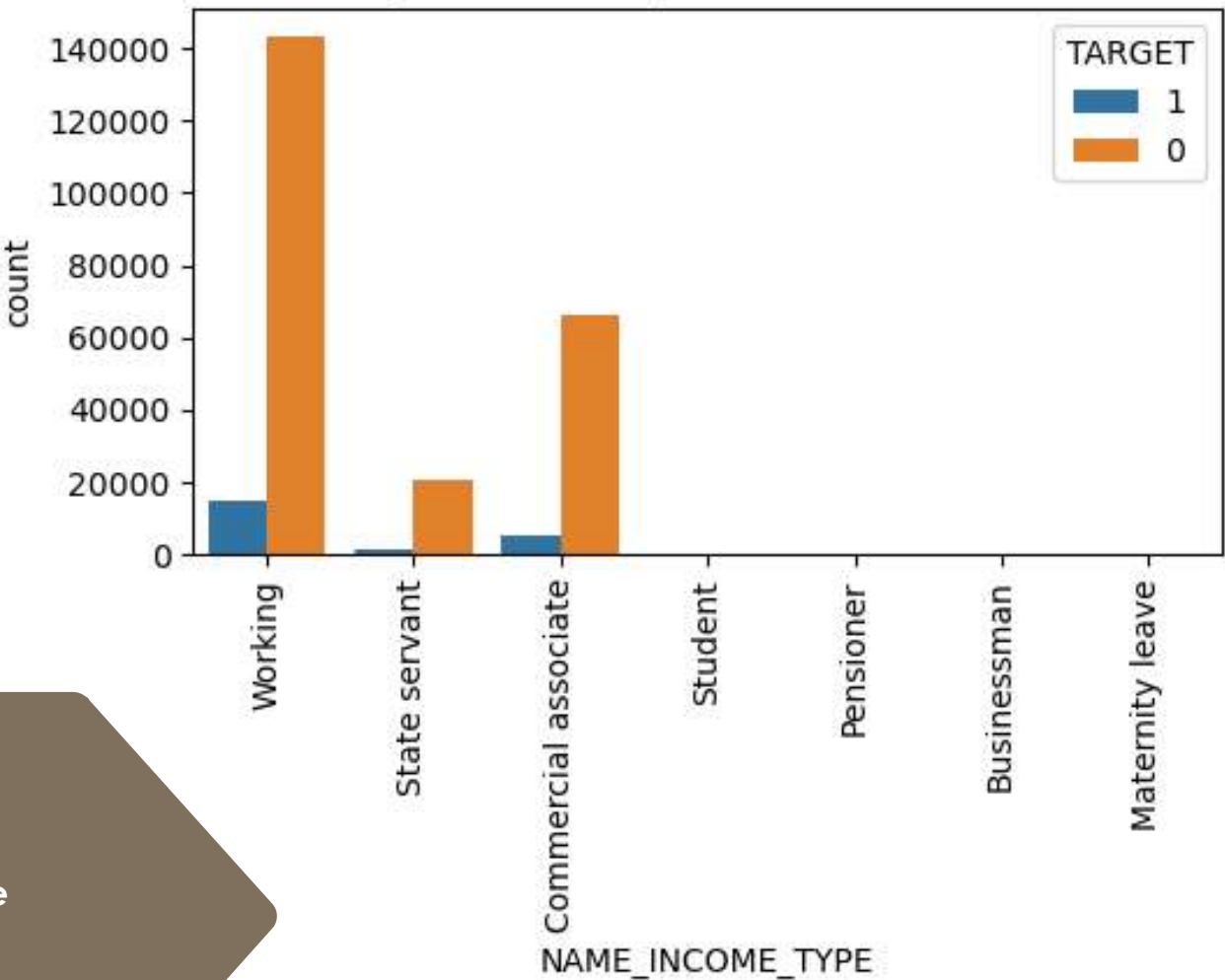
Plotting data for target in terms of percentage



Plotting data for the column: NAME_INCOME_TYPE



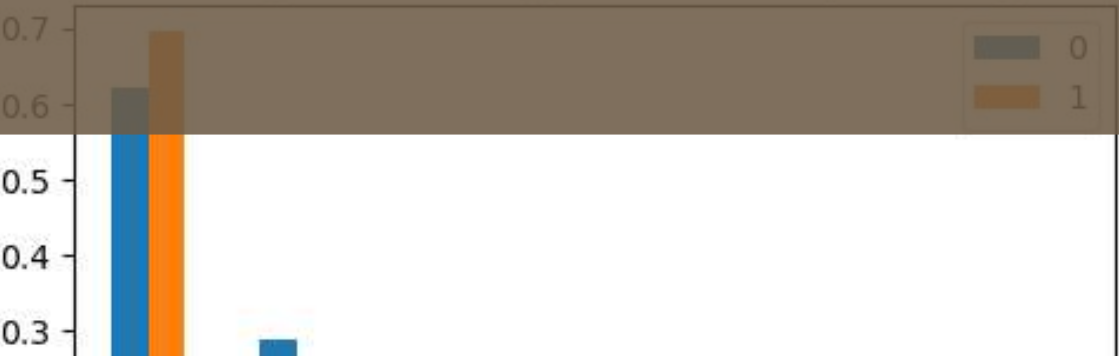
Plotting data for target in terms of total count



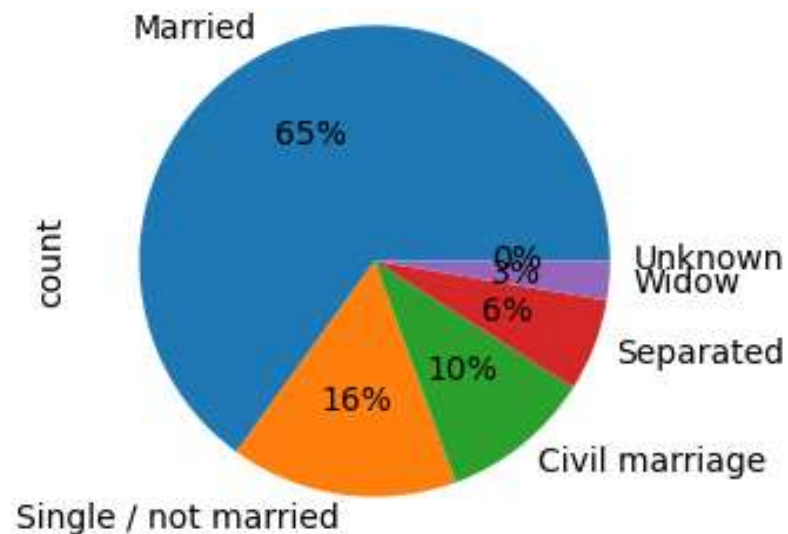
NAME_INCOME TYPE –

50% working in case of Target 0 and 60% in case of Target 1 are working income types.

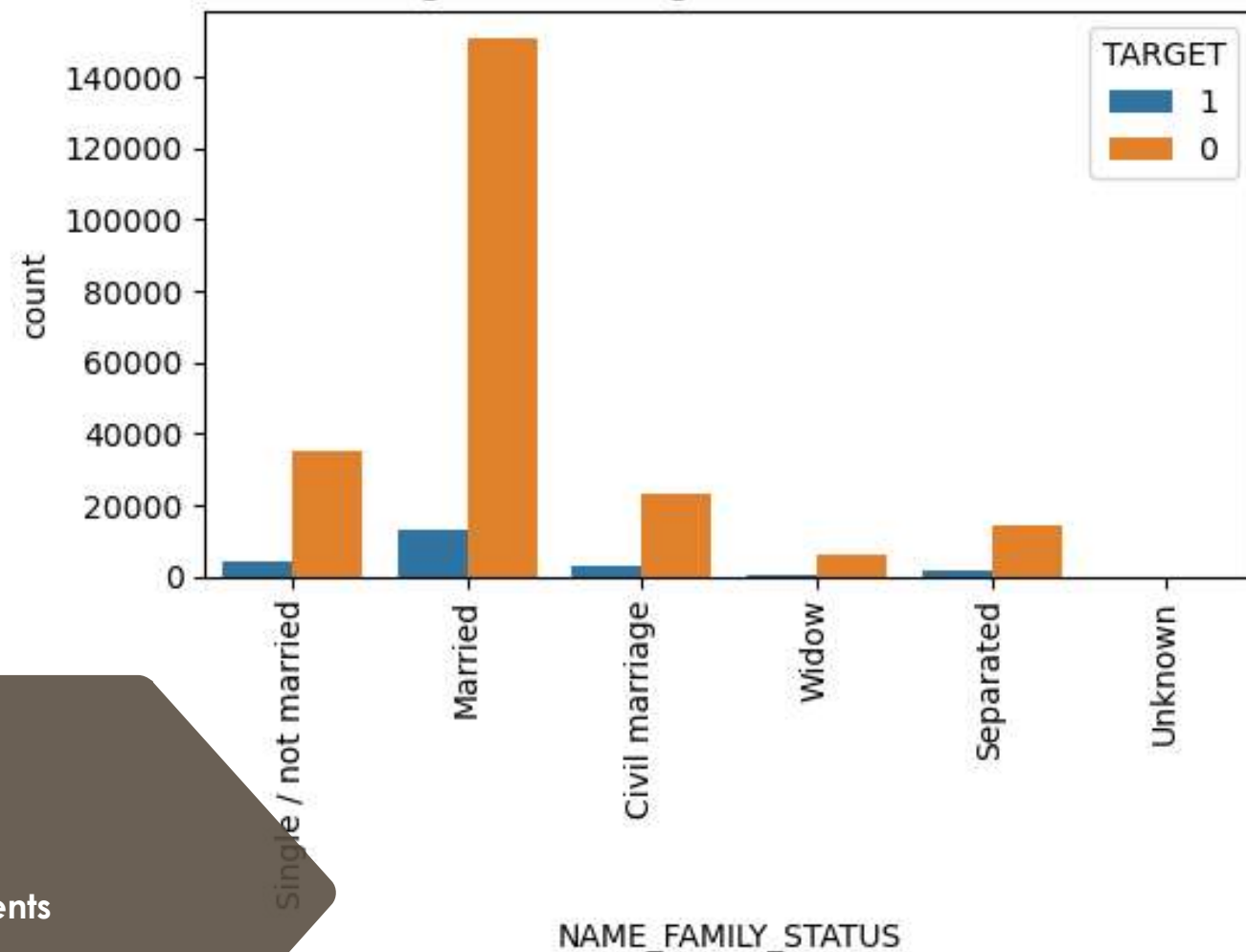
Plotting data for target in terms of percentage



Plotting data for the column: NAME_FAMILY_STATUS



Plotting data for target in terms of total count

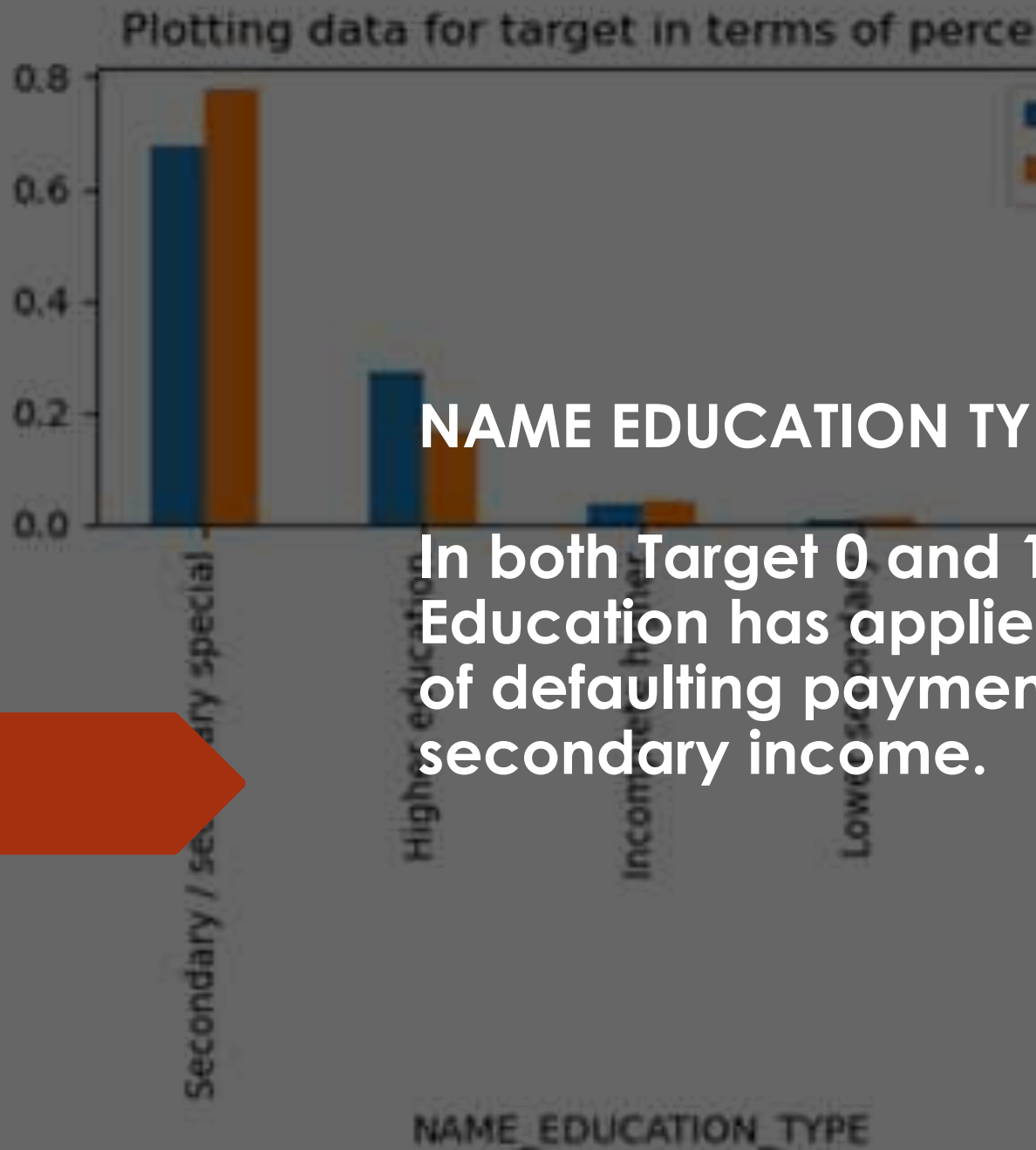


NAME_FAMILY_STATUS -

Married applicants - almost 60% have defaulted on payments

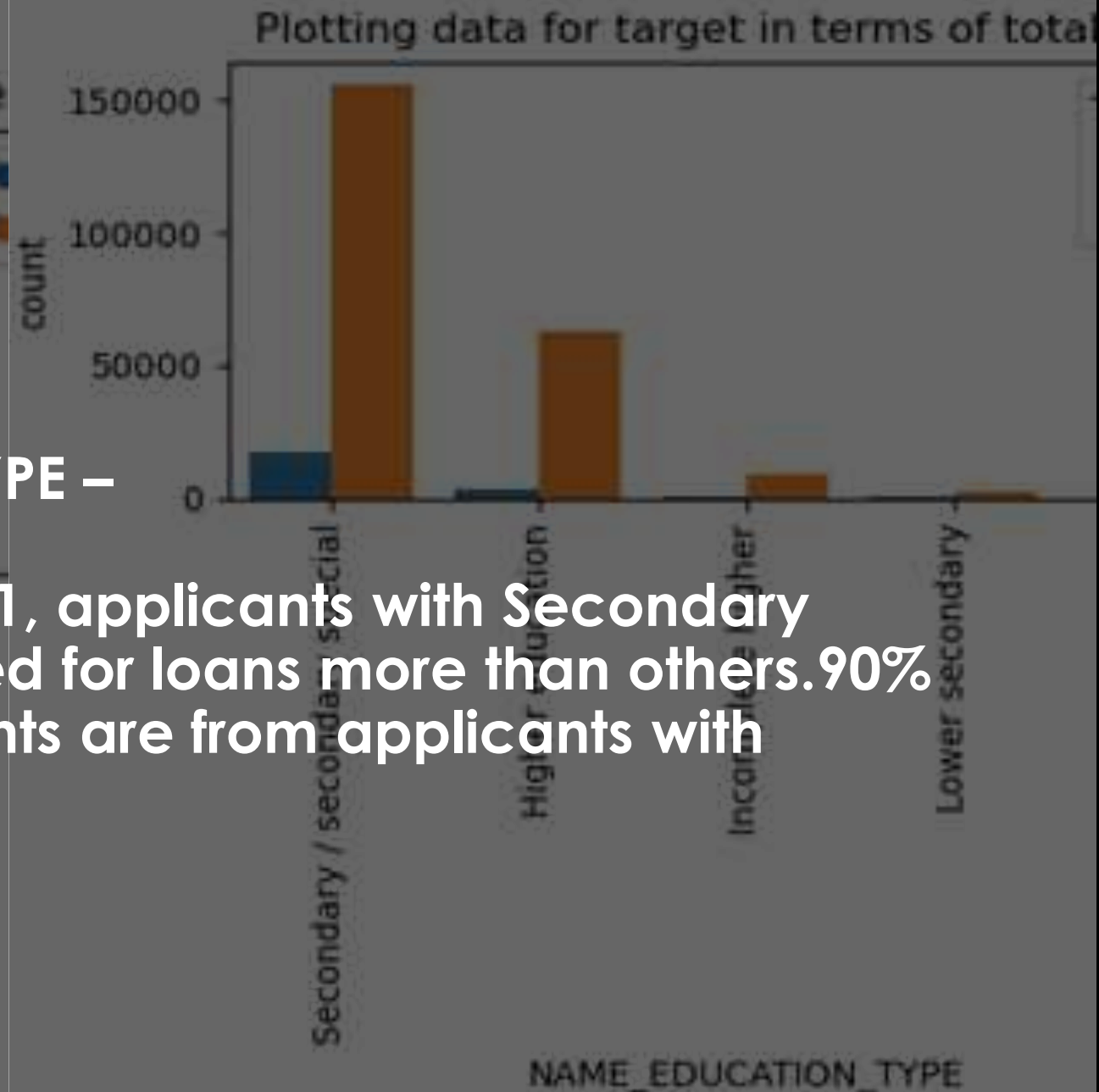
Plotting data for target in terms of percentage

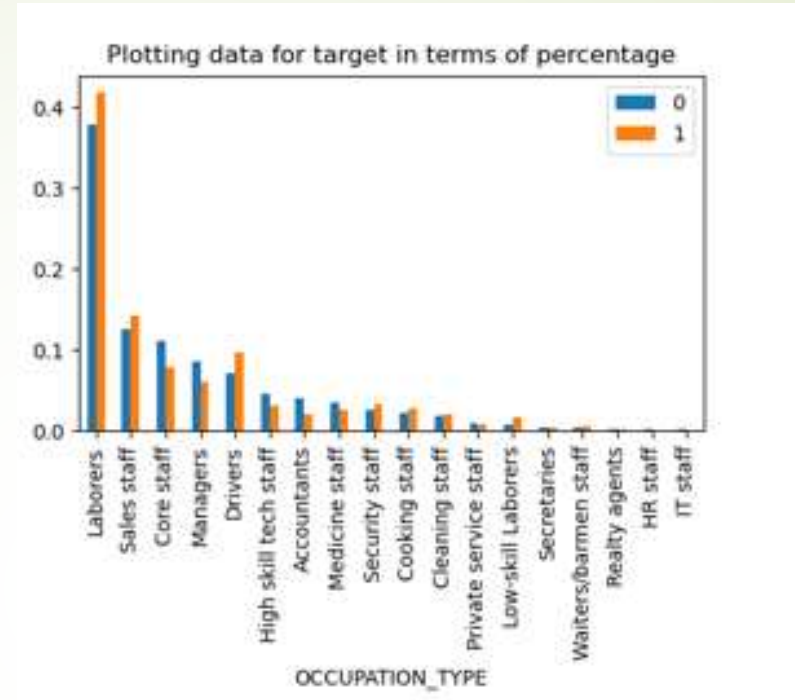
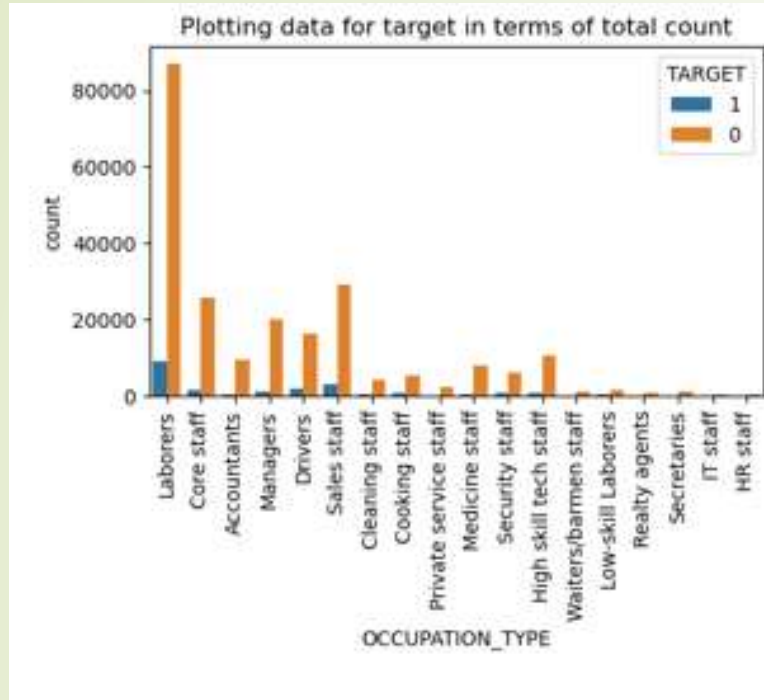




NAME EDUCATION TYPE –

In both Target 0 and 1, applicants with Secondary Education has applied for loans more than others. 90% of defaulting payments are from applicants with secondary income.



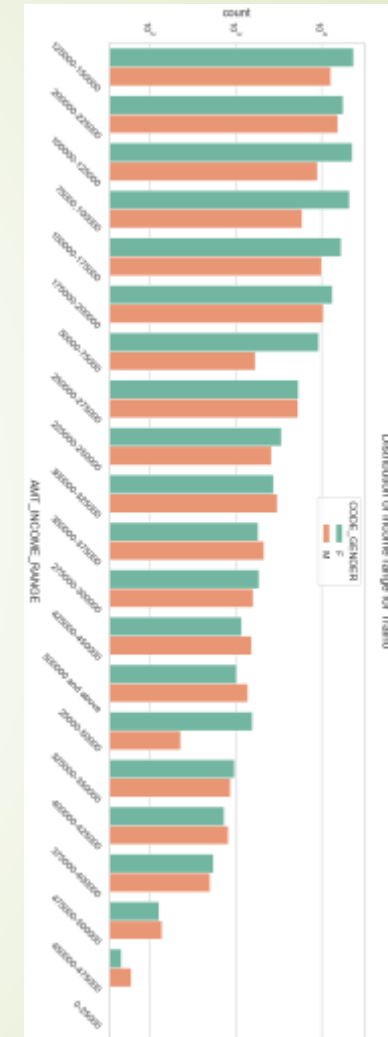


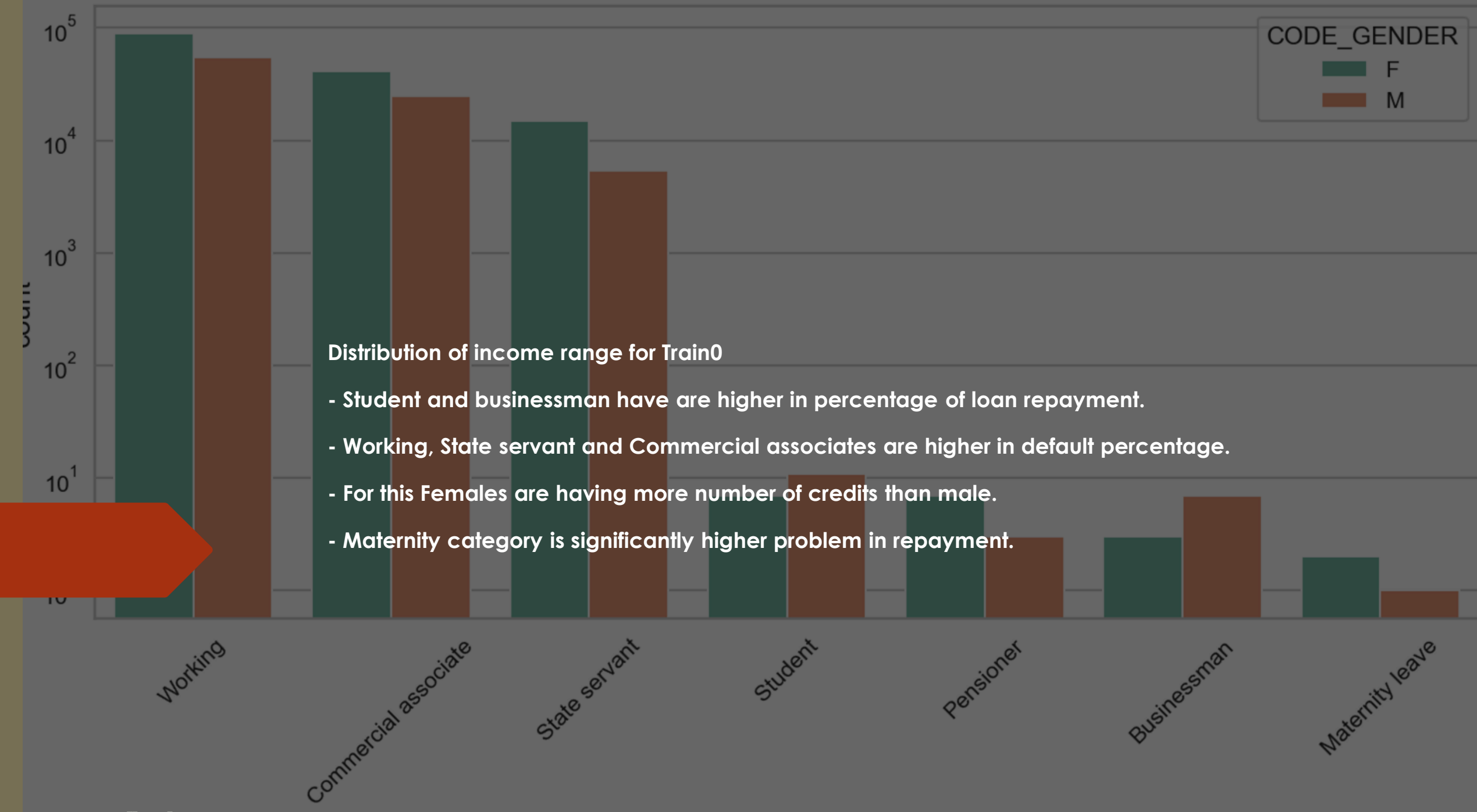
OCCUPATION_TYPE –

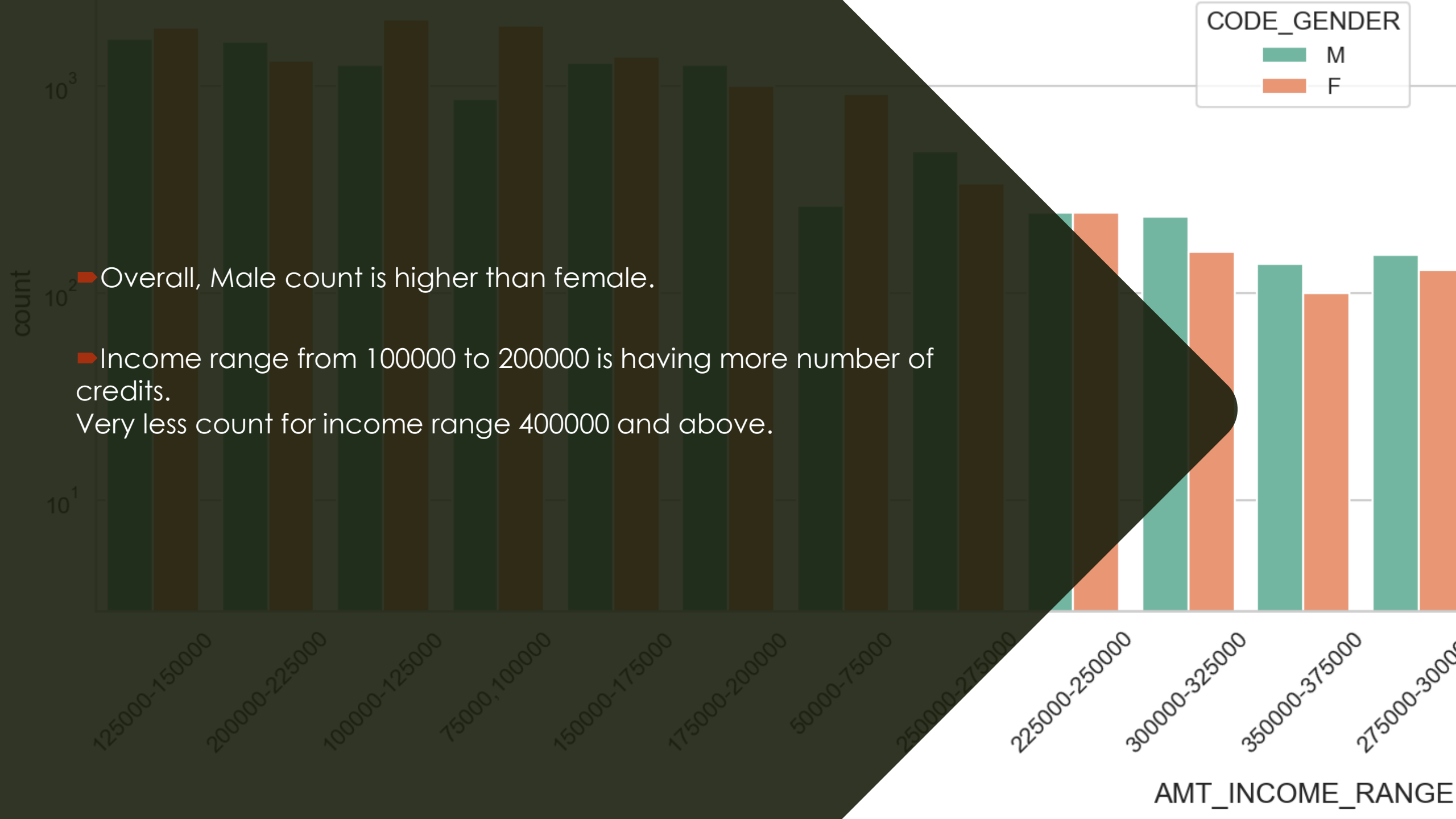
Laborers, sales staff, core staff, drivers constitute of 50% of defaulters. Laborers is the highest percentage of applicants too.

Distribution of income range for Train0

- Female counts are much higher than male.
- This graph show that females are more than male in having credits for that range.
- Very less count for income range 400000 and above.





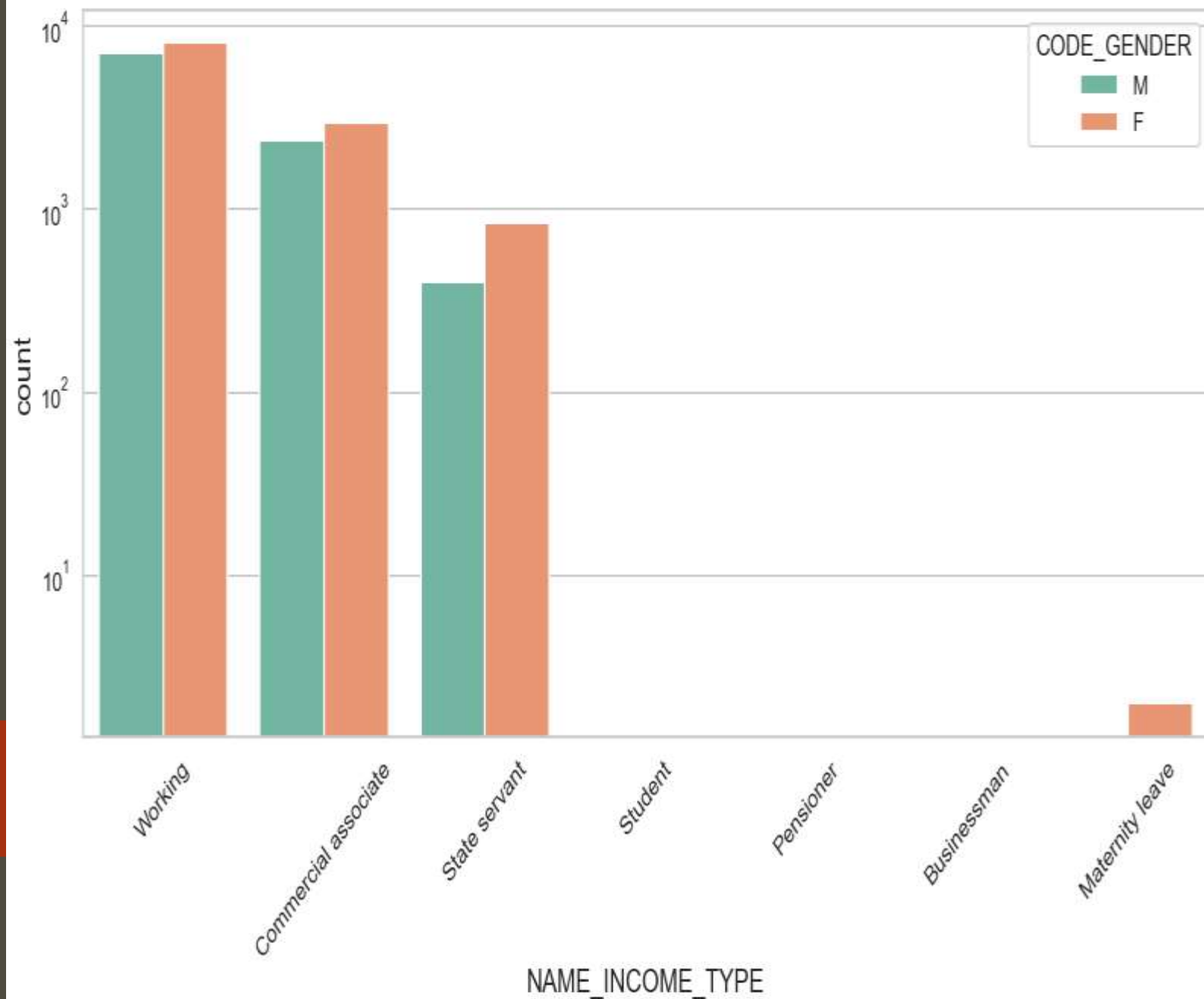


Distribution of income range for Train0

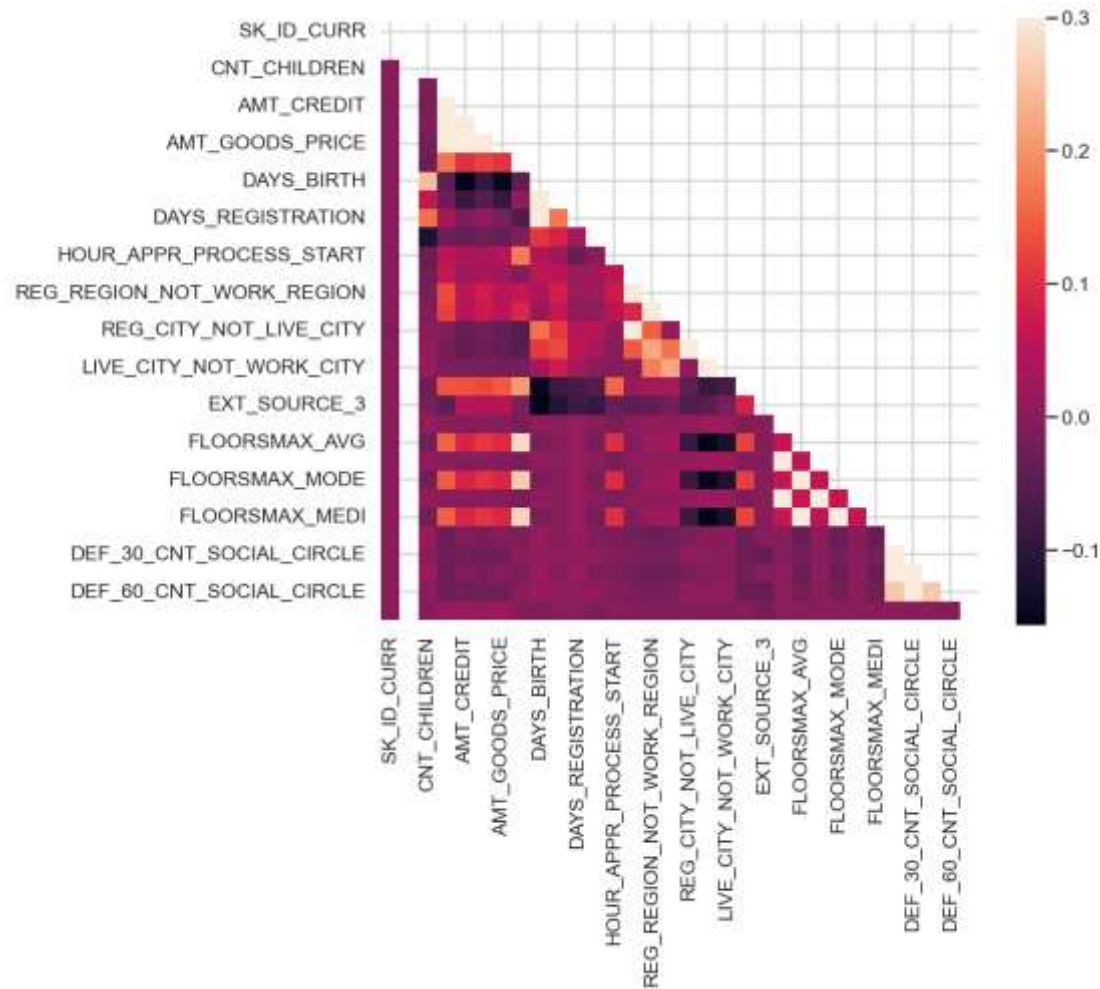
- For income type 'working', 'commercial associate', and 'State Servant' the number of credits are higher than other i.e. 'Maternity leave'.

- For train1: There is no income type for 'student', 'pensioner' and 'Businessman' which means they don't have any late payments.

Distribution of income range for Train1

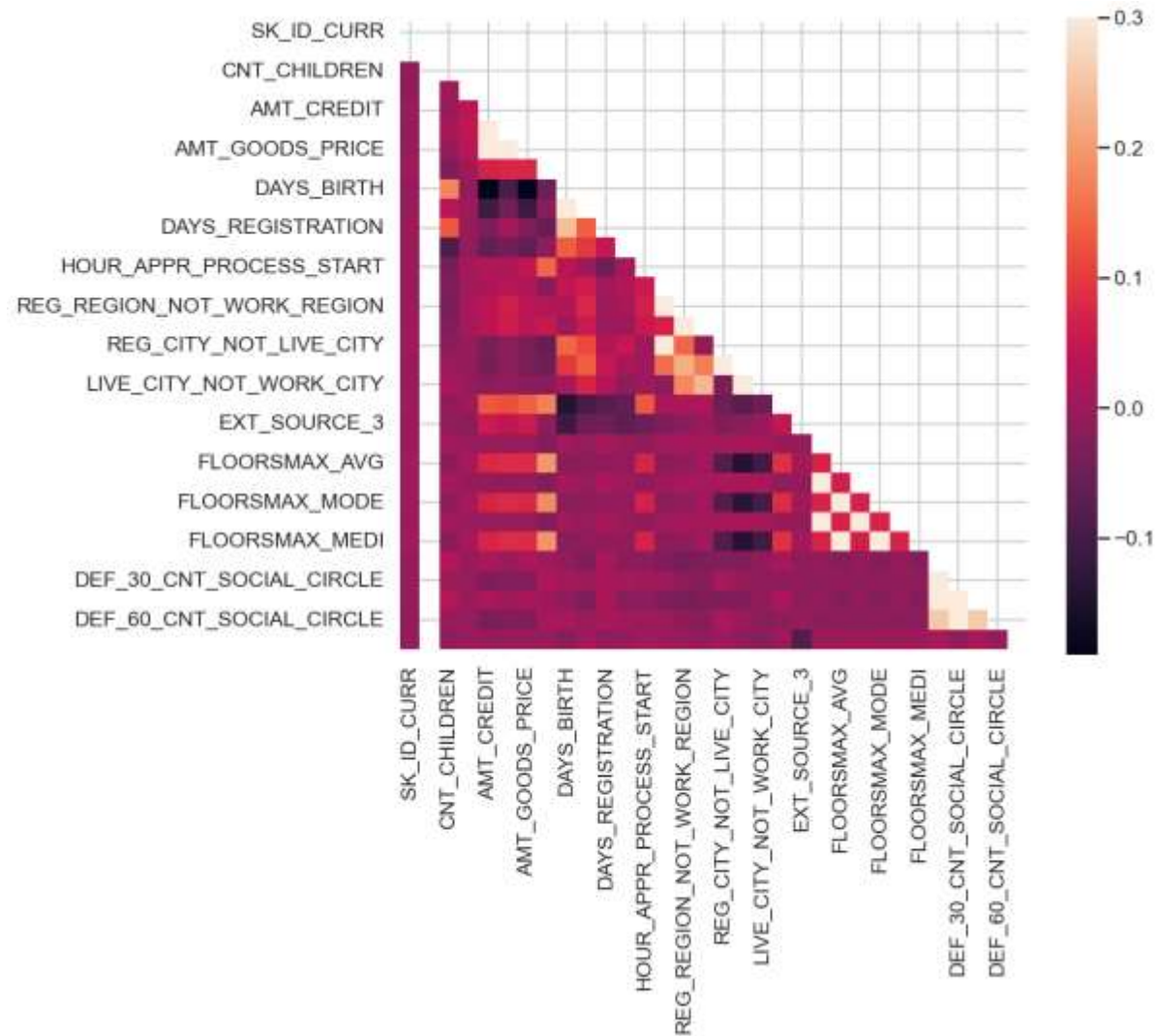


Correlation Train0



- As we can see from above correlation heatmap, There are number of observation we can point out
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.
People have less children in densely populated area.
- Credit and annuity amount is higher in densely populated area.
The income is also higher in densely populated area.

Correlation Train1

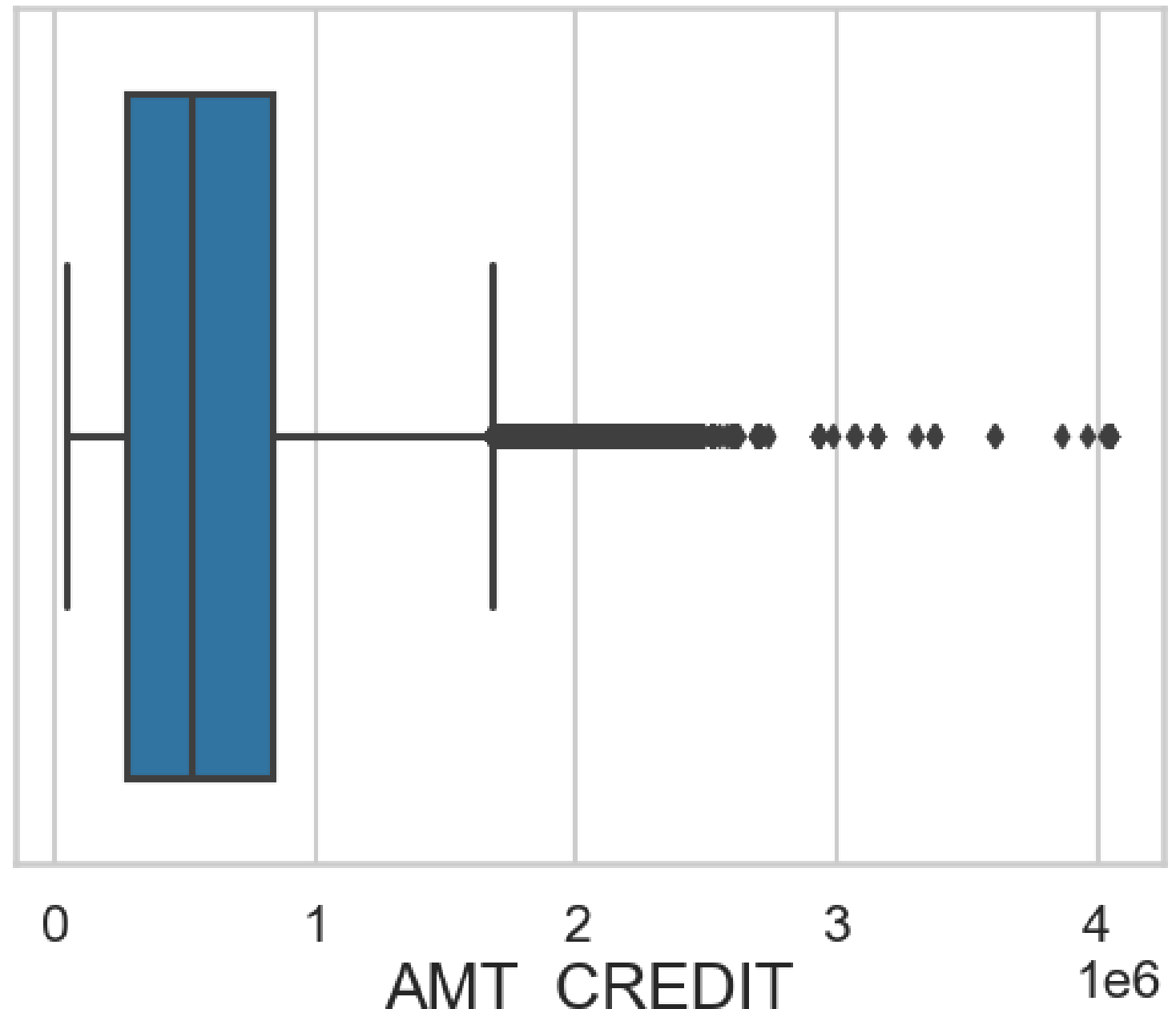


- This heat map for Target 1 is quite similar to Target 0 but with few differences. People with permanent address have a inverse proportion with contact address are having less children.
- People with permanent address have a inverse proportion work address are having less children.

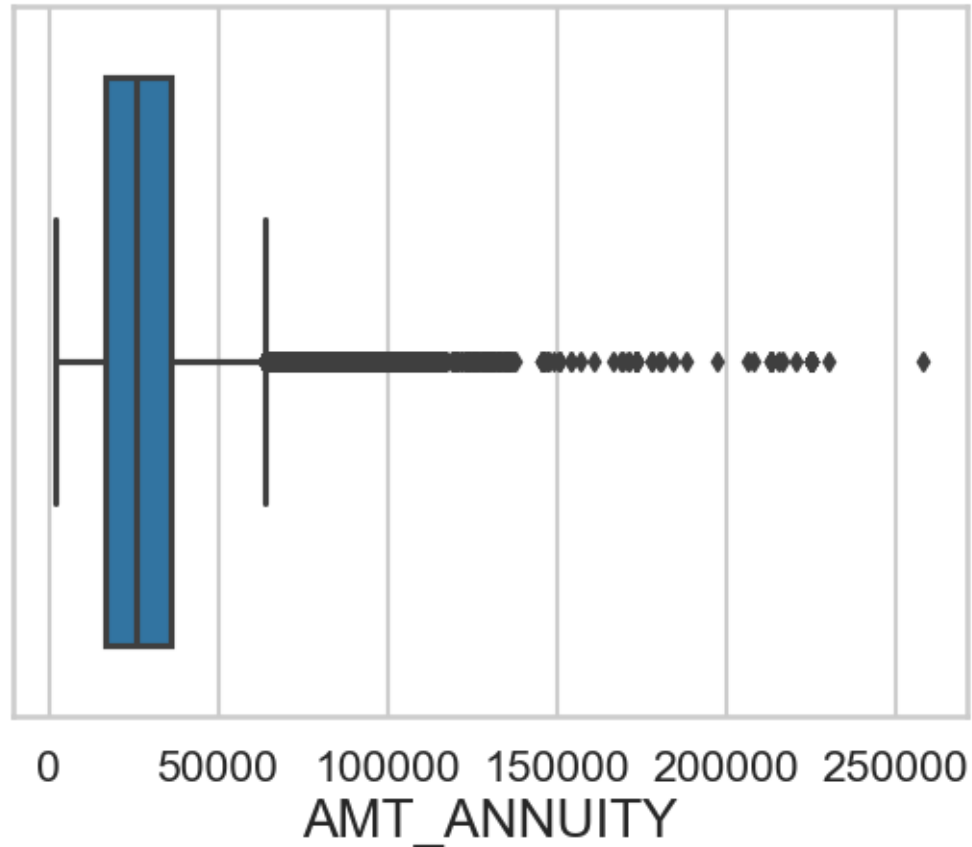
Distribution of credit amount for target0

Credit Amount Box plot analysis-

Some outliers are noticed in credit amount.
The first quartile is bigger than third quartile for credit amount which means most of the credits of clients are present in the first quartile.



Distribution of Annuity amount for target0



Annuity Amount Box plot analysis-

- Some outliers are noticed in annuity amount. The first quartile is bigger than third quartile for annuity amount which means most of the annuity clients are from first quartile.

Some outliers are noticed in income amount.

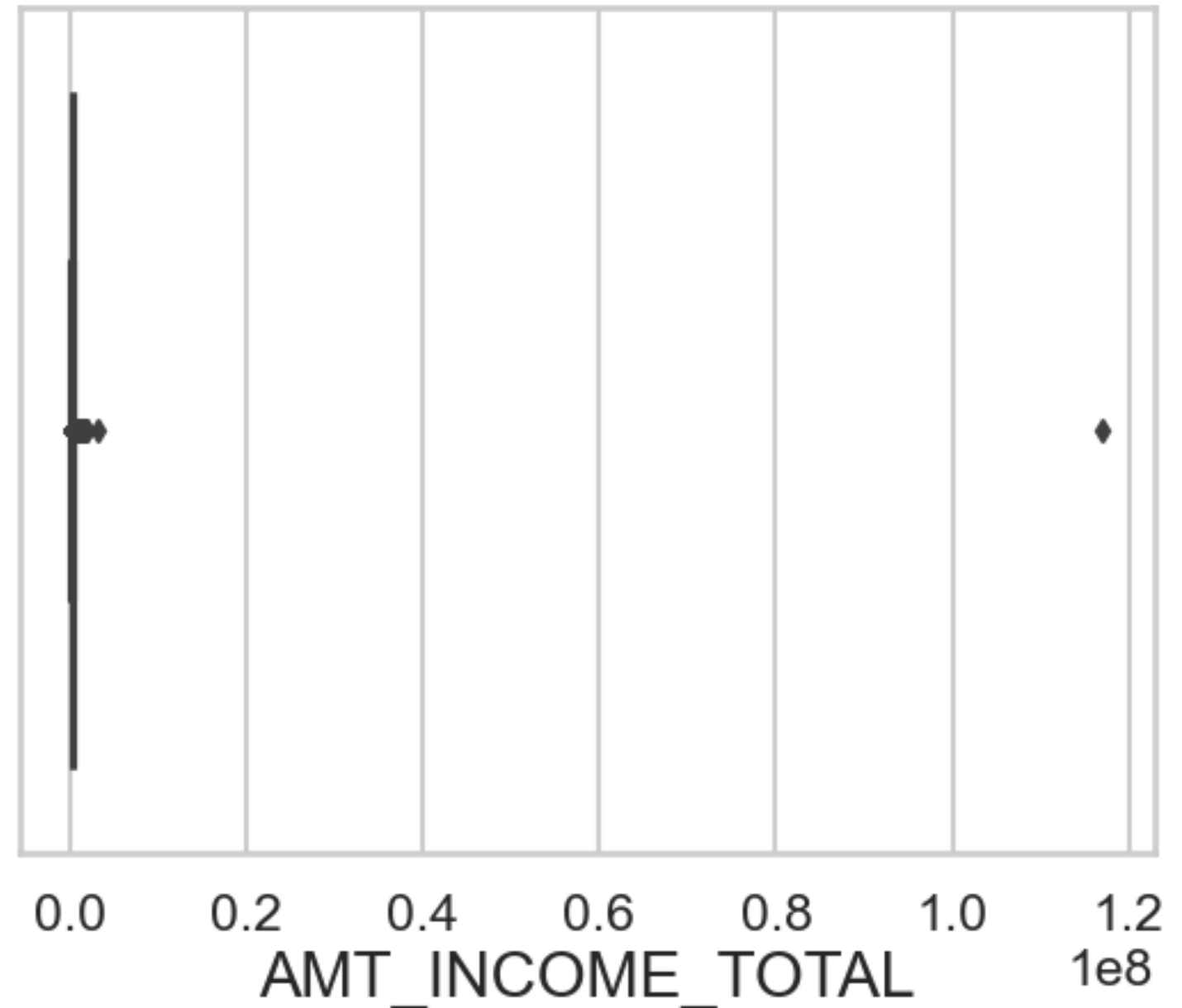
The third quartile is very slim for income amount.

Most of the clients of income are present in first quartile.

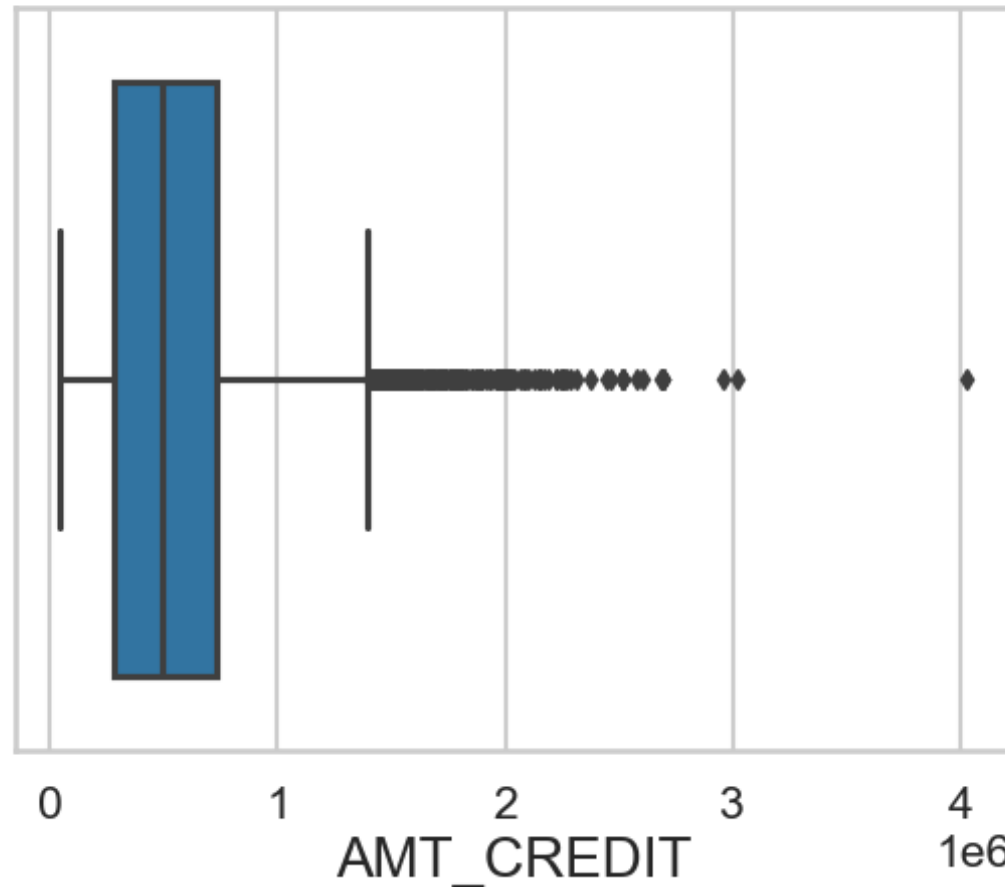
Some outliers are noticed in income amount.

The third quartile is very slim for income amount.

Most of the clients of income are present in first quartile.



Distribution of credit amount for target1



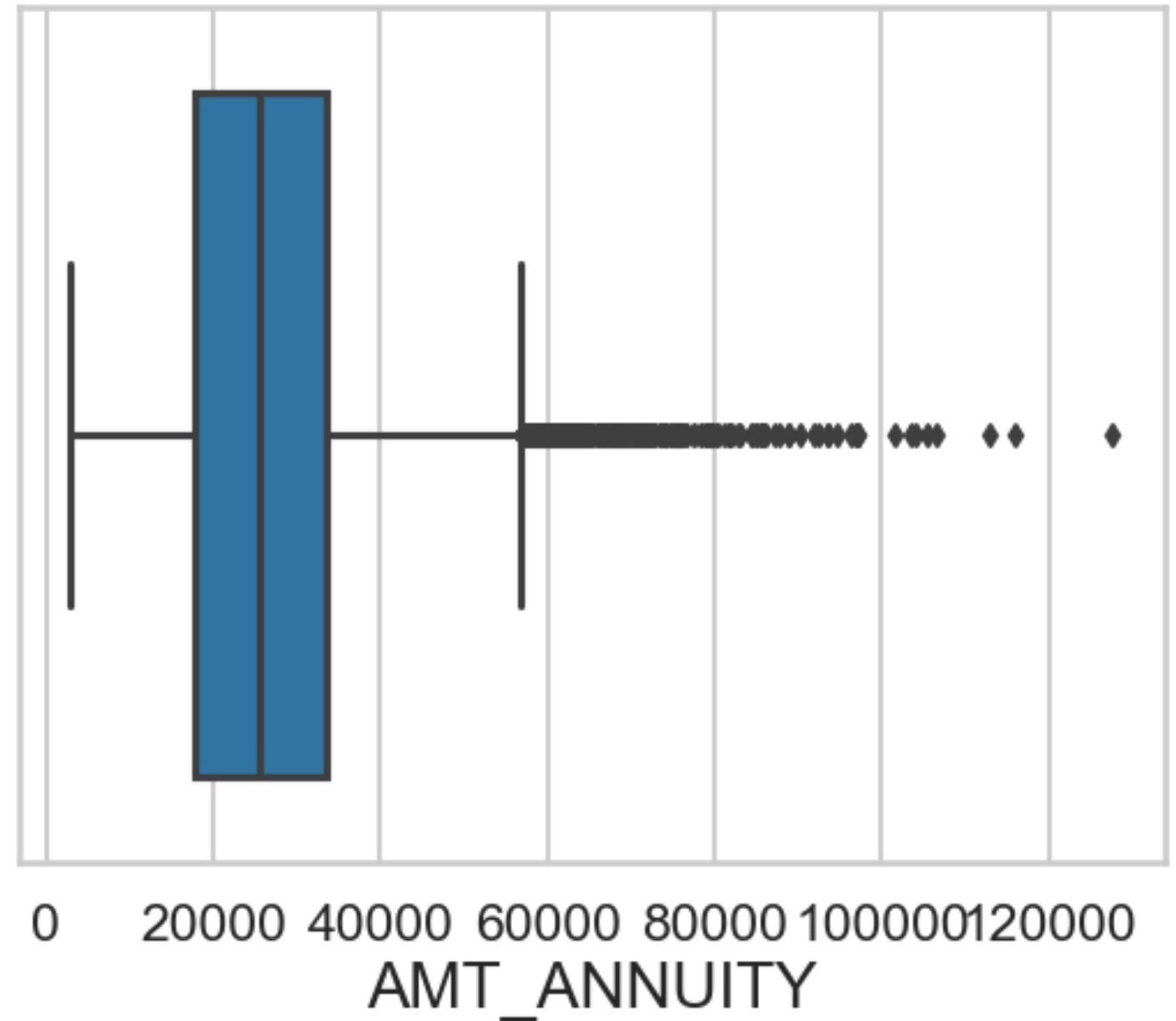
Credit Amount Box plot analysis-

- Some outliers are noticed in credit amount. The first quartile is bigger than third quartile for credit amount which means most of the credits of clients are present in the first quartile.

Distribution of Annuity amount for target1

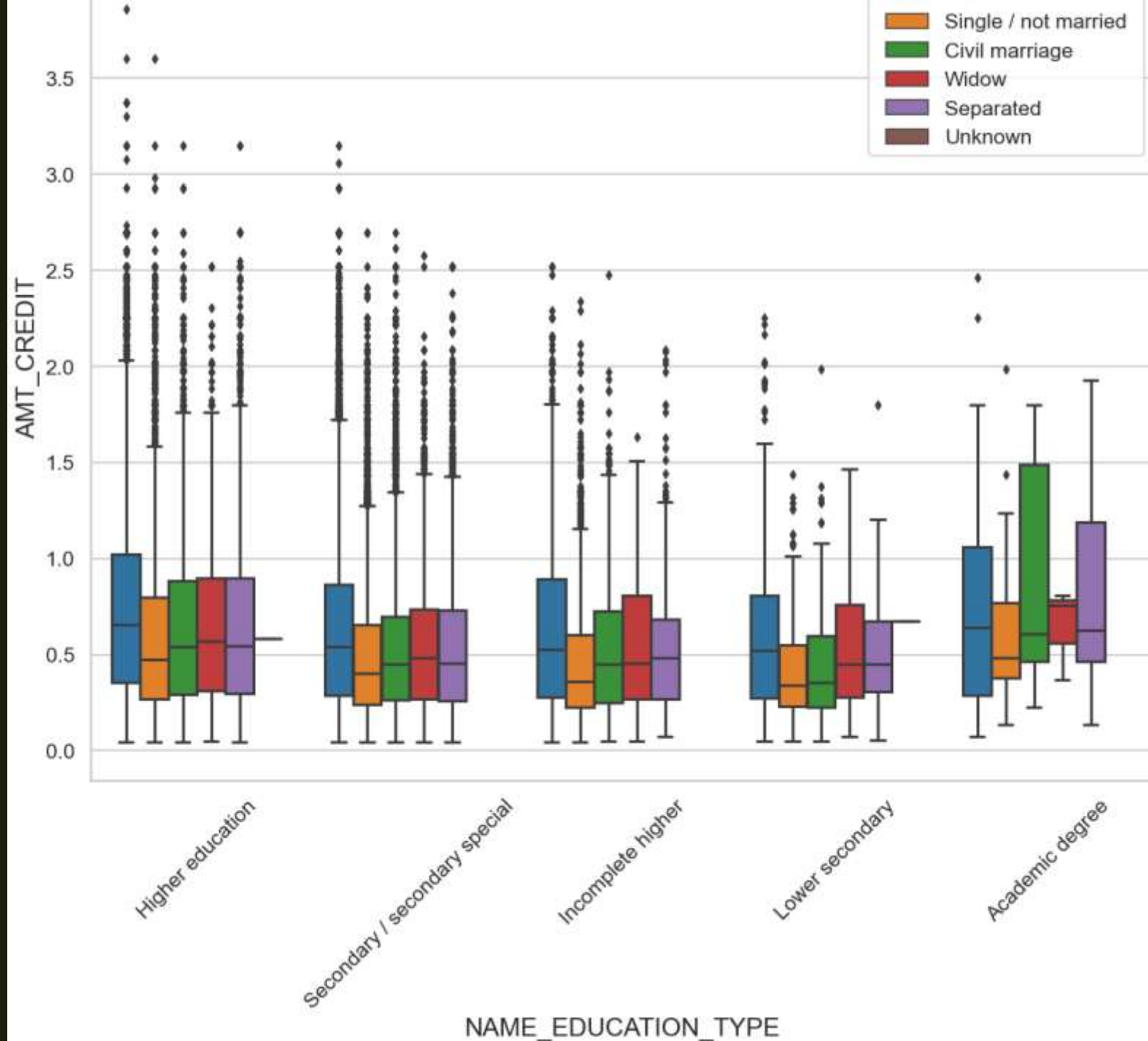
Annuity Amount Box plot analysis-

- Some outliers are noticed in annuity amount. The first quartile is bigger than third quartile for annuity amount which means most of the annuity clients are from first quartile



From the above box plot we can analyze for Train_0 –

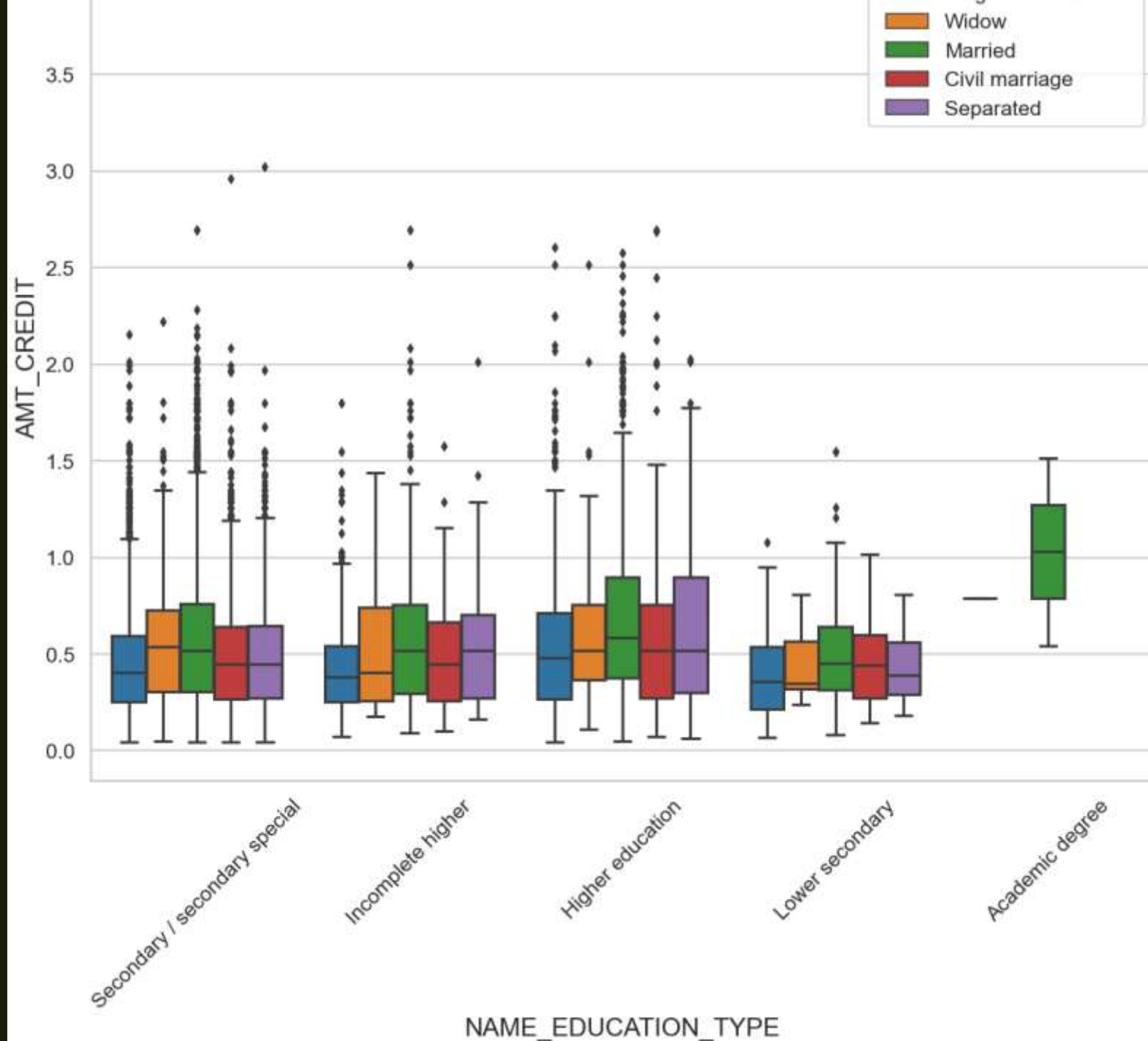
- People with Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.
- Higher education of people with family status of 'marriage', 'single' and 'civil marriage' has more outliers.
- Civil marriage for Academic degree is having most of the credits in the third quartile.



From the above box plot we can analyze for Train_1

- From the above box plot we can say that people with Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.

Most of the outliers are from Education type 'Higher education' and 'Secondary'. Civil marriage for Academic degree is having most of the credits in the third quartile.





Previous Application Data

Few notable points –

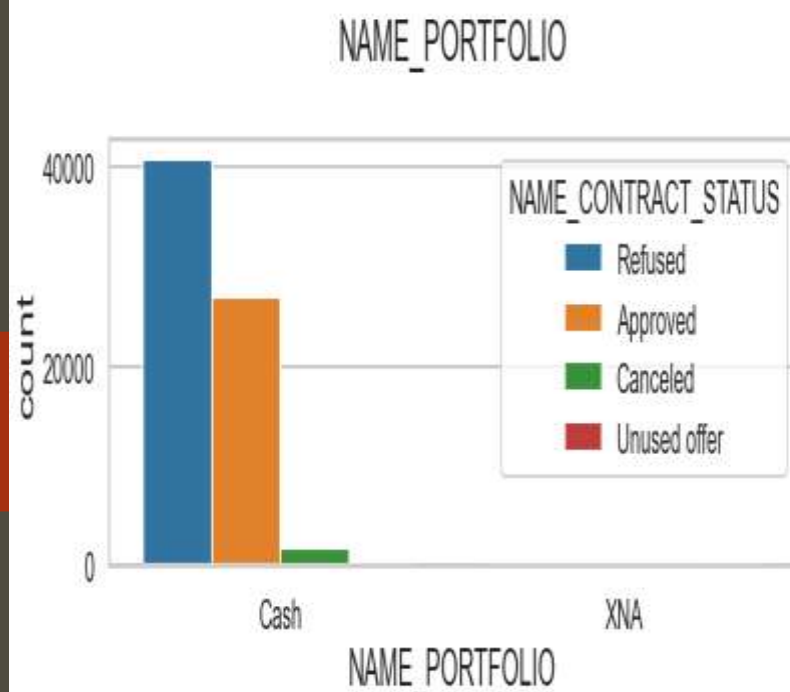
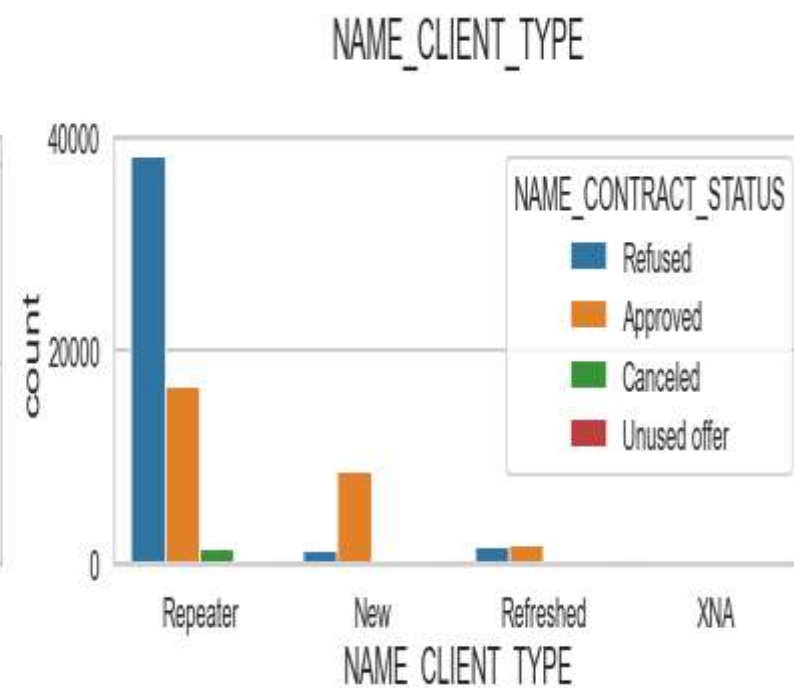
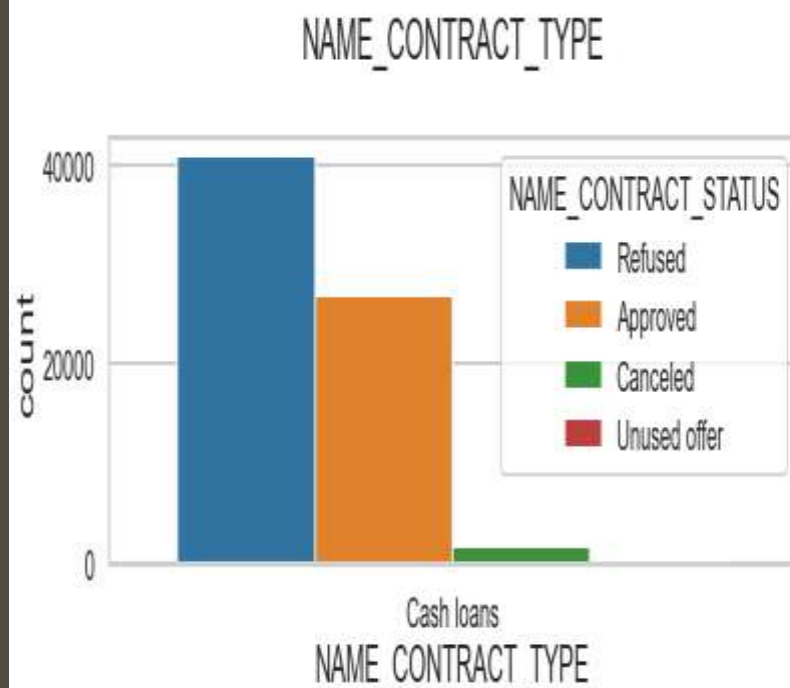
In approved category, consumer loan has largest no of applicants.

There seem to be no cancelled loans in cash loan category than consumer loan.

More cash loans have been refused than consumer loans.

The bank has more repeaters in all approved, refused, unused, cancelled categories

POS transactions seem to be consumer loans and similar to point 2 - more cash loans have been refused than POS.





Heat Map Analysis – Multivariate

- Few notable points –
- Unused offer application amount is low
Cancelled application amount is high.
- The bank may be refusing these possibly as the Debt liability ratio of consumer must be going high due to the high amount and thus credit default risk.
- Repeater's application amount is higher than the New customers. This may indicate that the bank has more conducive policies/rate of interest etc for repeat applicants

Analysis from the below chart above –

- Most rejection of loans came from purpose 'repairs' and then 'other' For education purposes we have equal number of approves and rejection.

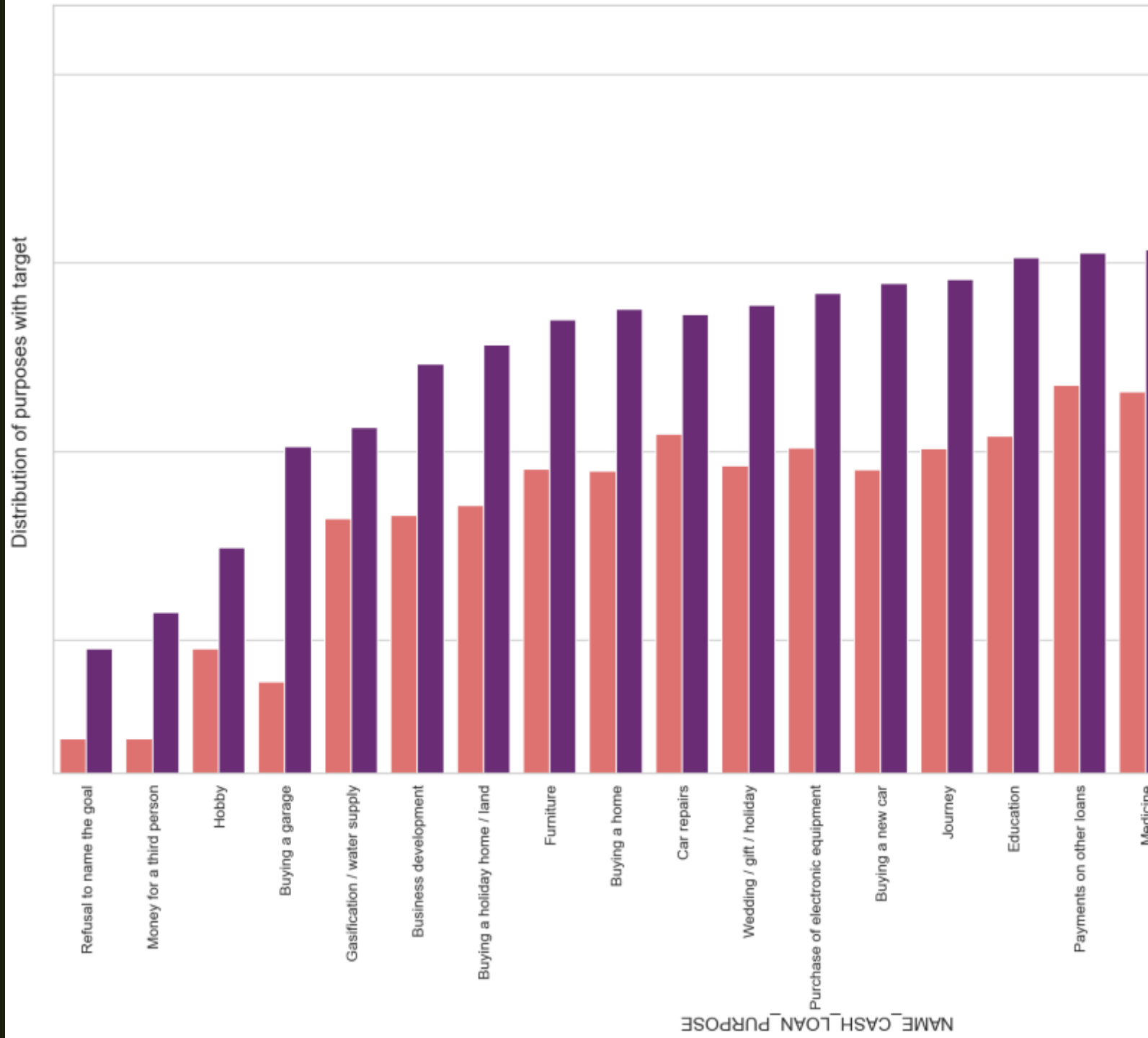
- Paying other loans and buying a new car is having significant higher rejection than approves.

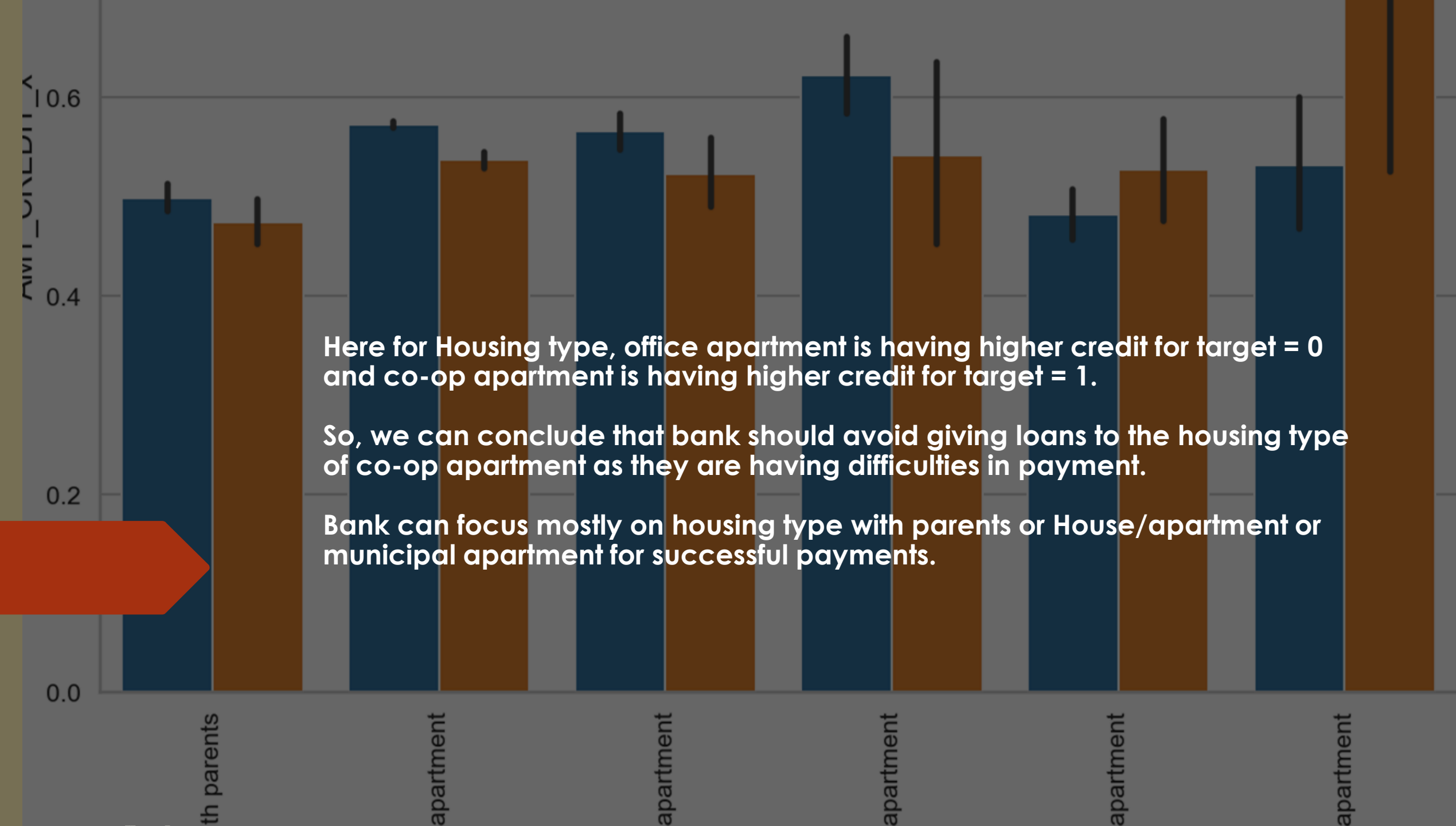


Few points we can conclude from above plot:

- Loan purposes with 'Repairs' are facing most difficulties in payment on time (among all purposes).
- There are few places where loan payment is significantly higher than facing difficulties.
- They are 'Buying a garage', 'Business development', 'Buying land', 'Buying a new car' and 'Education'

Hence, we can focus on these purposes for which the client is having for minimal payment difficulties.





Here for Housing type, office apartment is having higher credit for target = 0 and co-op apartment is having higher credit for 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.

Final Summary for the Case study

Defaulters Characteristics –

- All the below variables were established in analysis of Application dataframe as leading to default:
 - Medium income group can lead for Defaulters.
 - 25-35 years age range can lead for Defaulters and there is a slight chance the 35-45 years age group can also become lead for Defaulters.
 - We have noticed the Male and mostly in the category of Defaulters.- People with housing type 'With parents' can be targeted as they are having least number of unsuccessful payments.
 - Banks should focus less on income type 'Working' as they are having most number of unsuccessful payments.- People like Laborers, Salesman, Drivers also led to become Defaulters.
 - People with Business type and non-Own House can also become Defaulters.
 - Also, with loan purpose 'Repair' is having higher number of unsuccessful payments on time.



Final Summary for the Case study

- **Other IMPORTANT Factors to be considered**
 - Days last phone number changed - Lower figure points at concern.
 - No of Bureau Hits in last week. Month etc – zero hits is good- Amount income not correspondingly equivalent to Good Bought. Income low and good value high is a concern-
 - Previous applications with Refused, Cancelled, Unused loans also have default which is a matter of concern.
 - Focus shall be on people with income type 'Student' , 'pensioner' and 'Businessman' with housing type other than 'Co-op apartment' for successful payments.
 - This indicates that the financial company had Refused/Cancelled previous application but has approved the current and is facing default on those approvals.

Final Summary for the Case study

➤ Credible Applications refused –

- Unused applications have lower loan amount. Is this the reason for no usage
- Female applicants should be given extra weightage as defaults are lesser.
- 60% of defaulters are Working applicants. This does not mean working applicants must be refused. Proper scrutiny of other parameters needed
- Previous applications with Refused, Cancelled, Unused loans also have cases where payments are coming on time in current application. This indicates that possibly wrong decisions were done in those cases.

