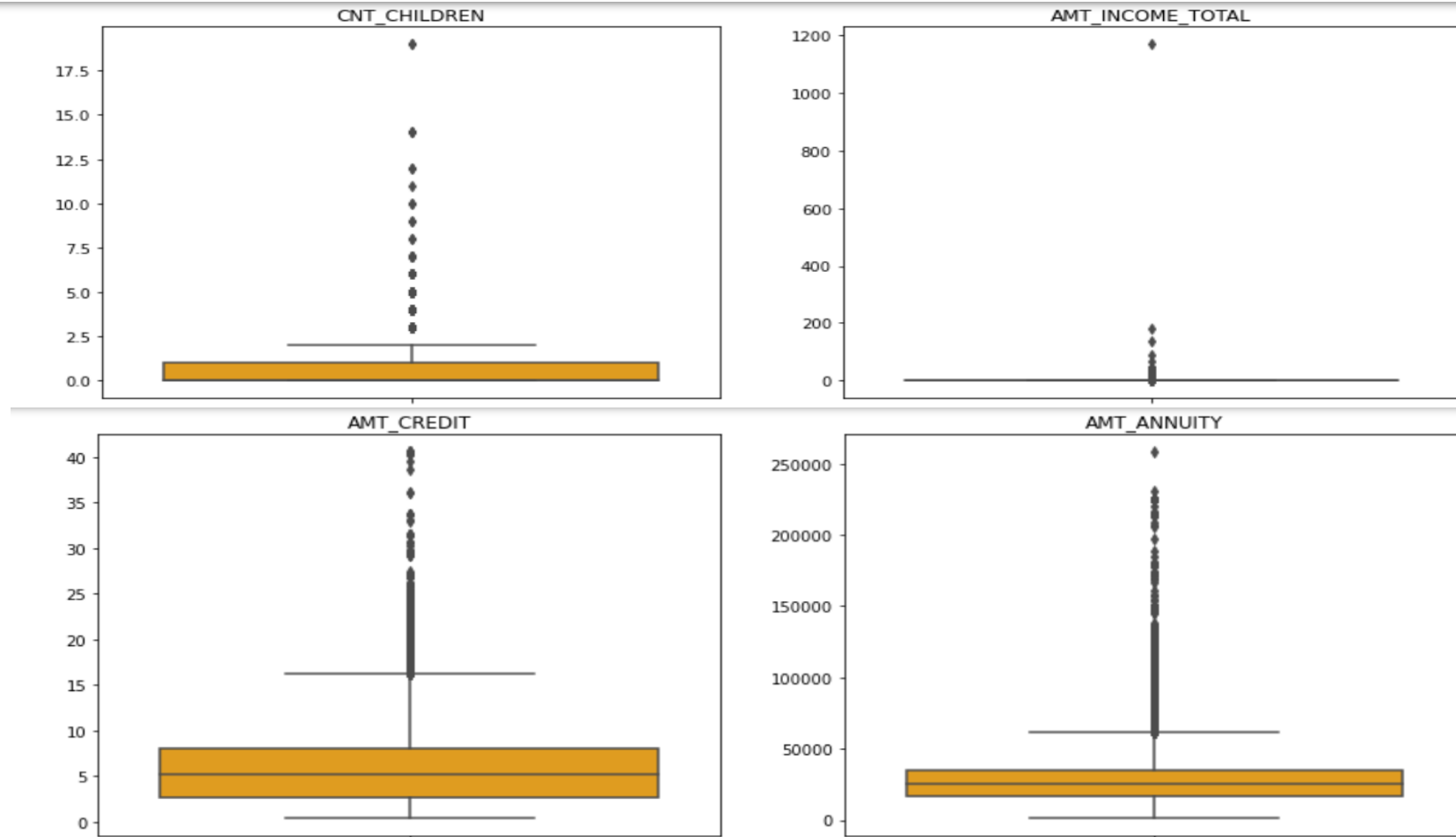


Credit EDA Case Study

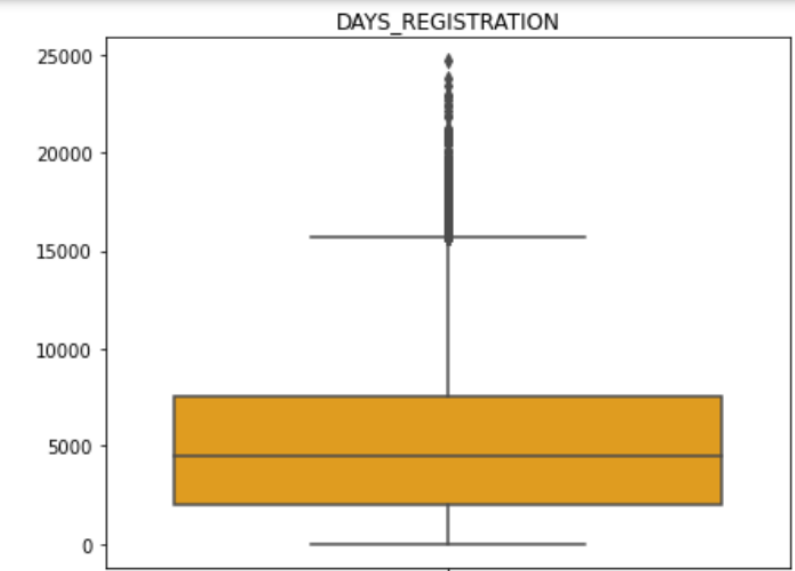
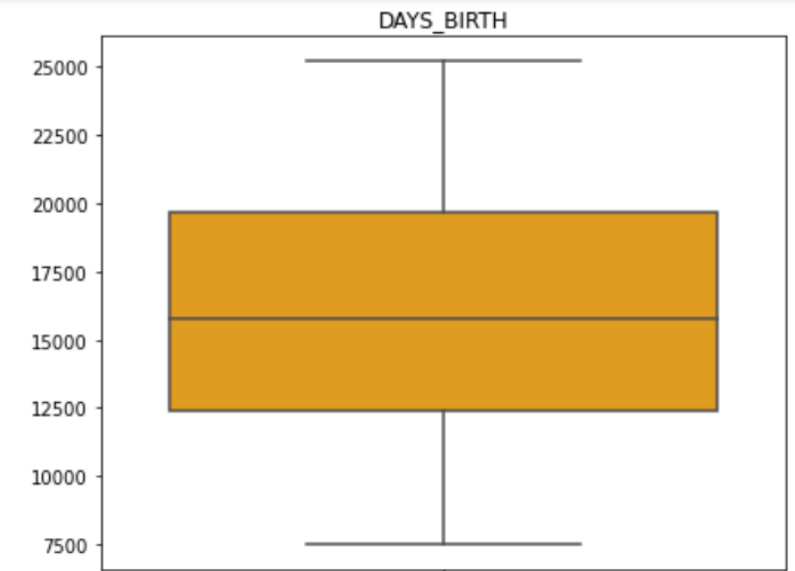
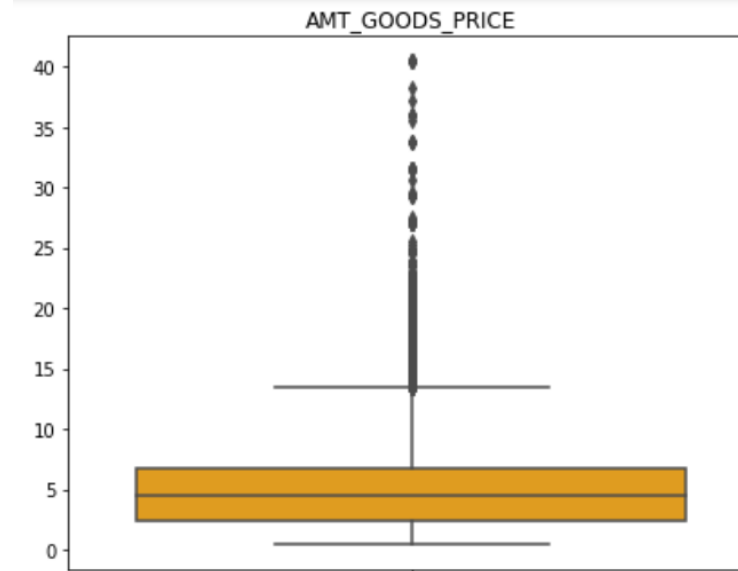
BY – SHREY TOSHNIWAL

Outliers in Application Data



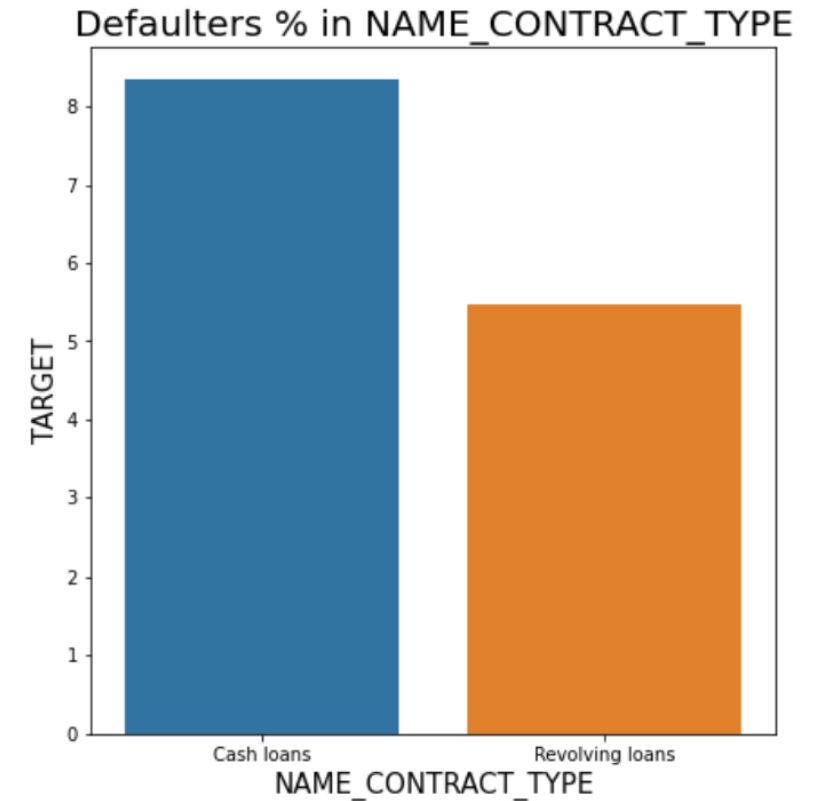
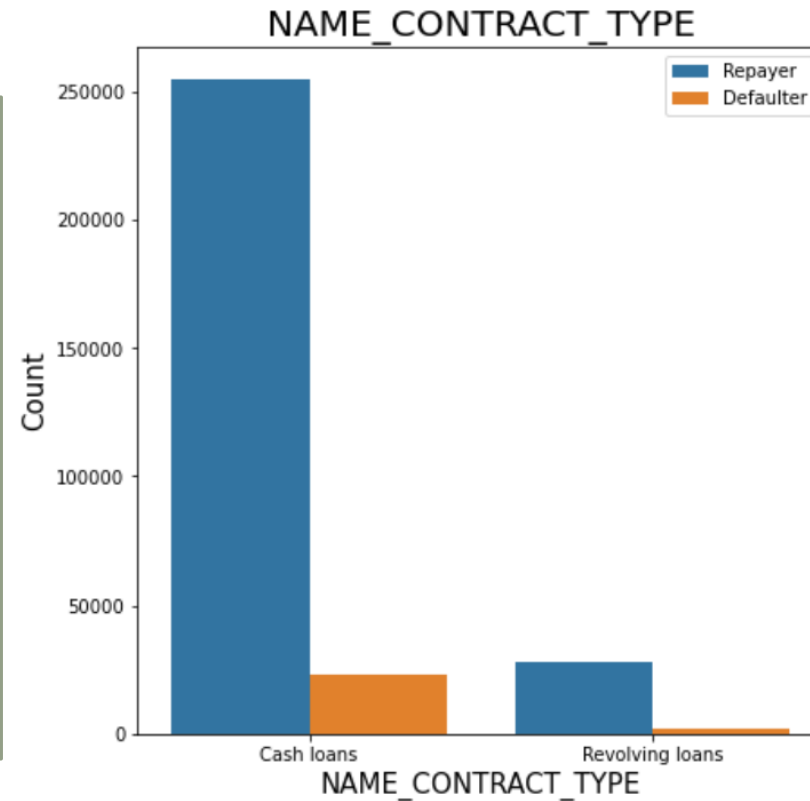
It can be seen that in current application data:

1. AMT_ANNUITY, AMT_CREDIT, AMT_GOODS_PRICE, CNT_CHILDREN have some number of outliers.
2. AMT_INCOME_TOTAL has huge number of outliers which indicate that few of the loan applicants have high income when compared to the others.
3. DAYS_BIRTH has no outliers which means the data available is reliable.
4. DAYS_EMPLOYED has outlier values around 350000(days) which is around 958 years which is impossible and hence this has to be incorrect entry.



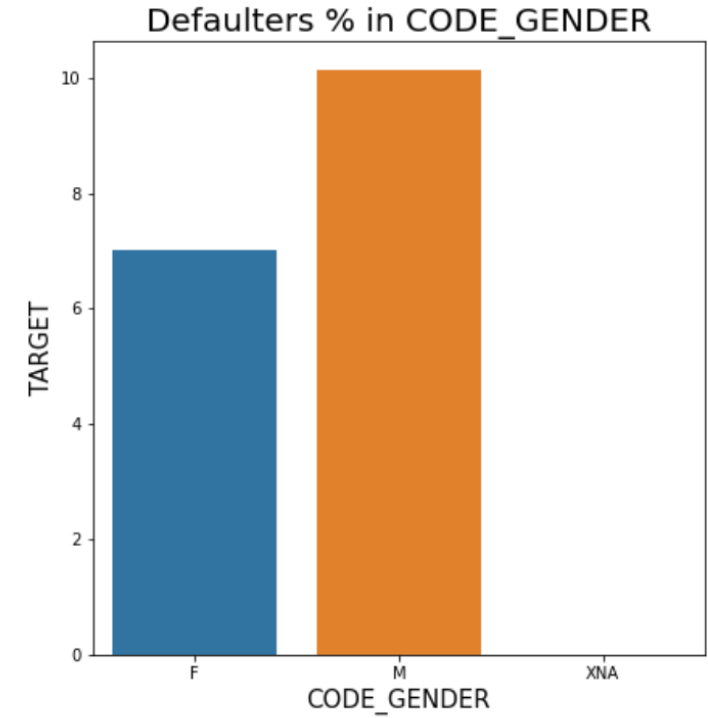
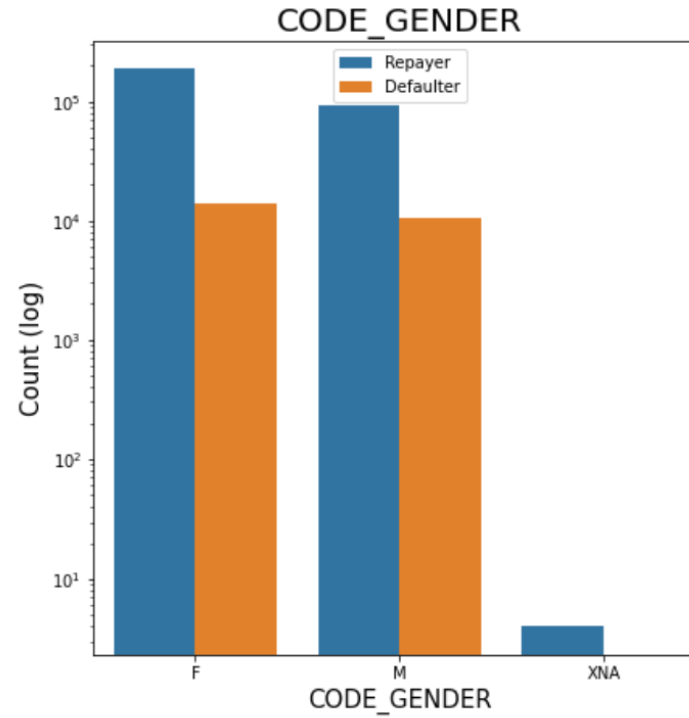
Categorical Variable Analysis

1. Revolving loans are just a small fraction (10%) from the total number of loans
2. Around 8-9% Cash loan applicants and 5-6% Revolving loan applicant are in defaulters

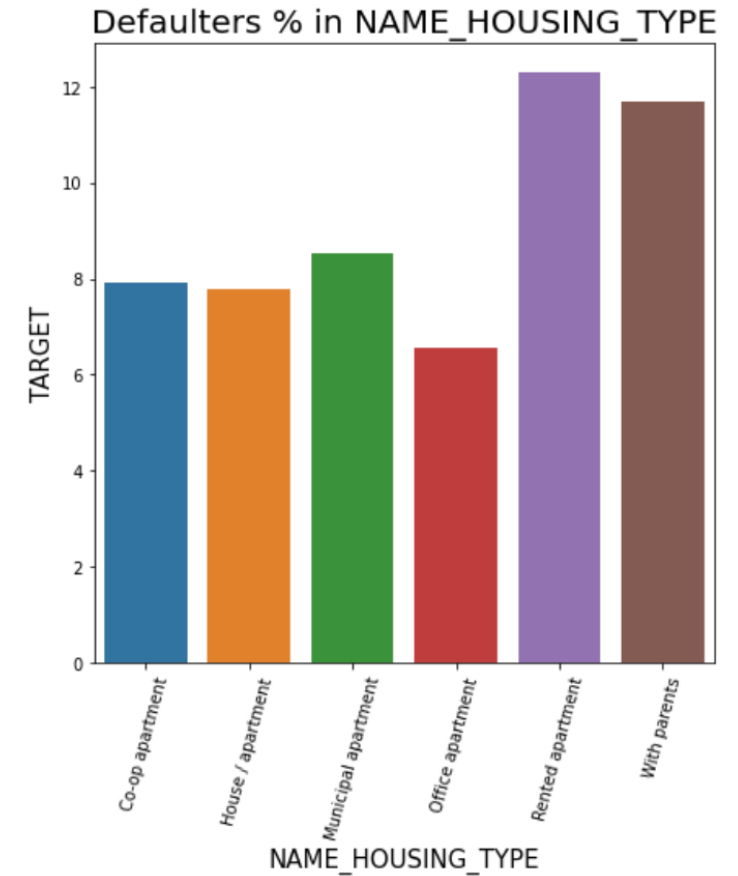
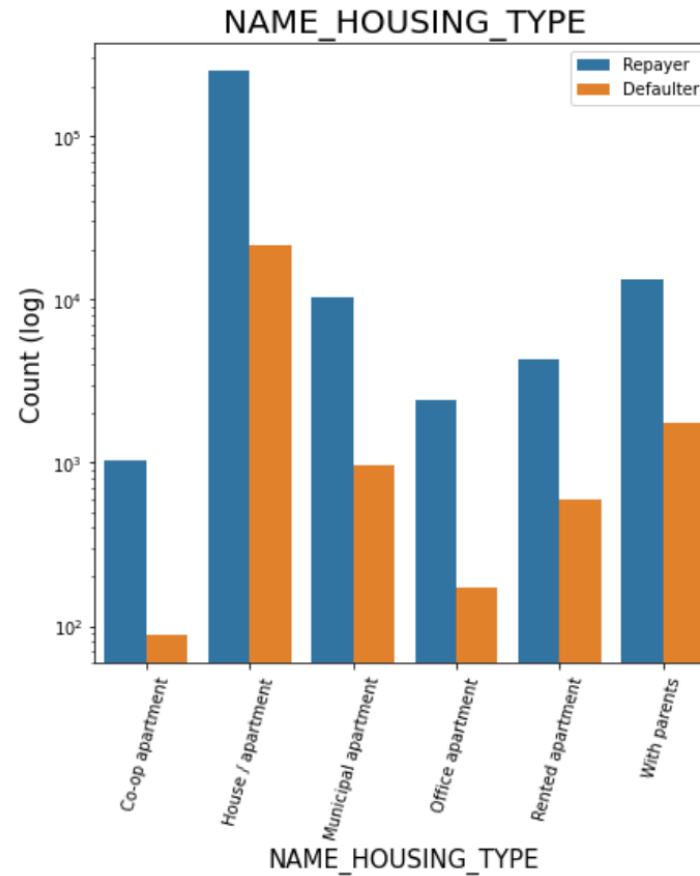


1. The number of female clients is almost double the number of male clients.

2. Based on the percentage of defaulted credits, males have a higher chance of not returning their loans about 10%, comparing with women about 7%

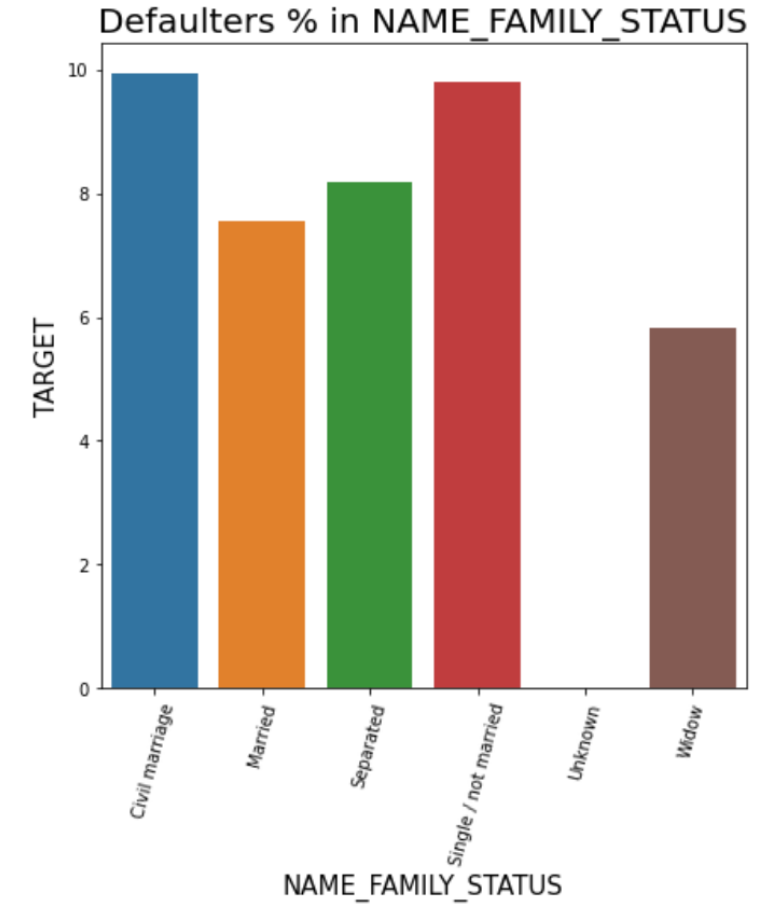
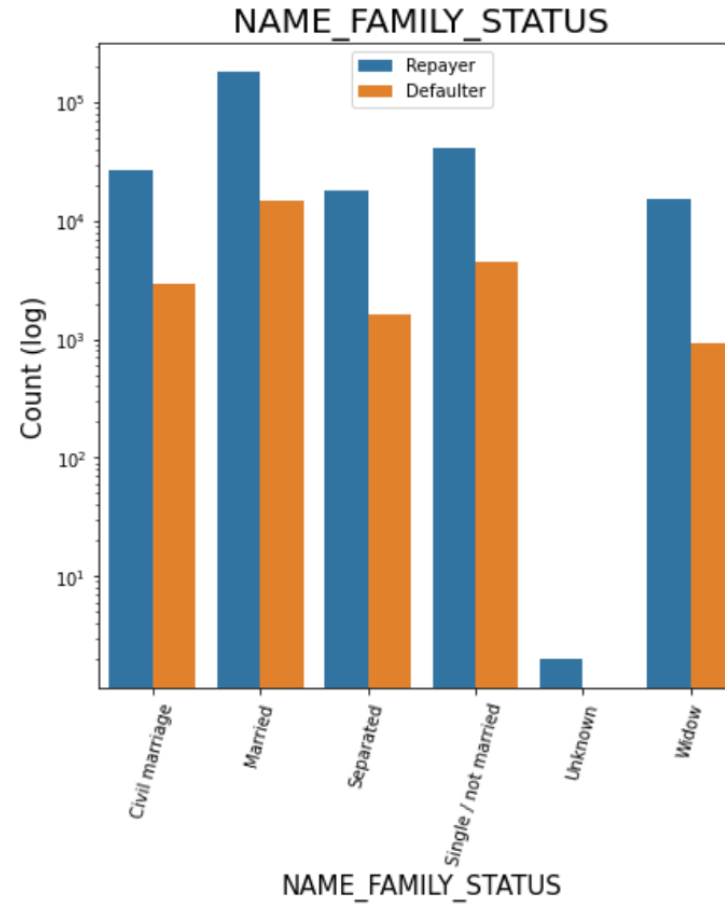


1. Majority of people live in House/apartment
2. People living in office apartments have lowest default rate
3. People living with parents (~11.5%) and living in rented apartments (>12%) have higher probability of defaulting



1. Most of the people who have taken loan are married, followed by Single/not married and civil marriage

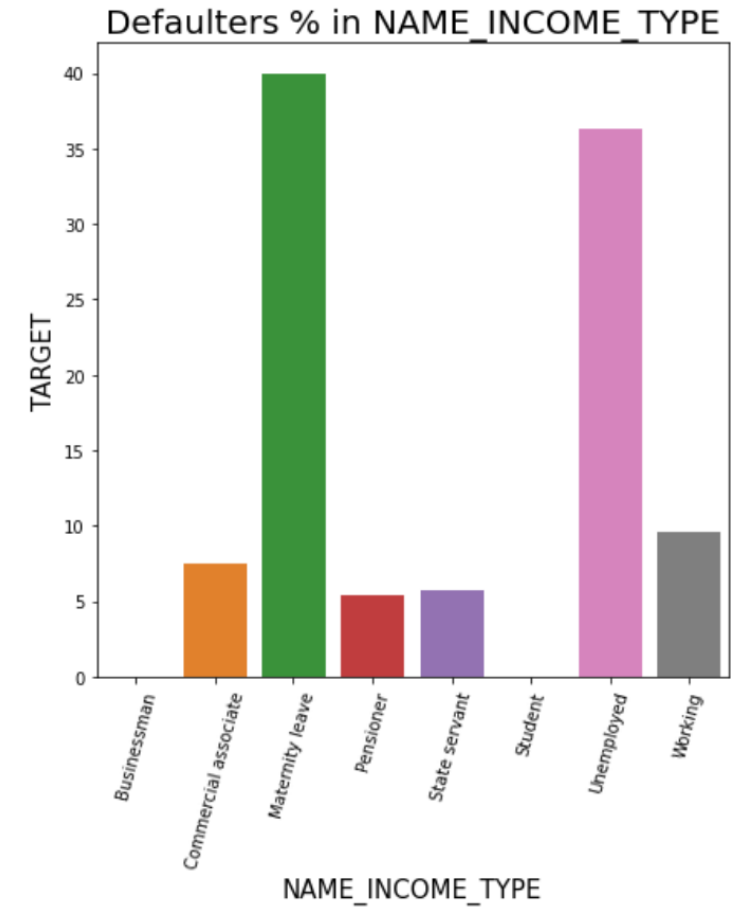
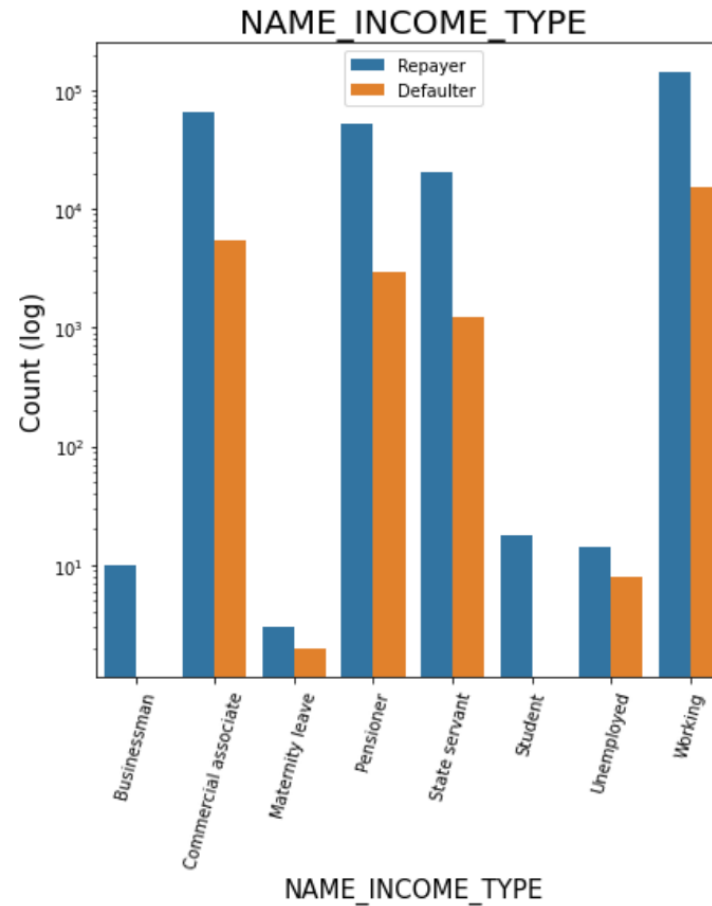
2. In Percentage of defaulters, Civil marriage has the highest percent around (10%) and widow has the lowest around 6% (exception being Unknown).

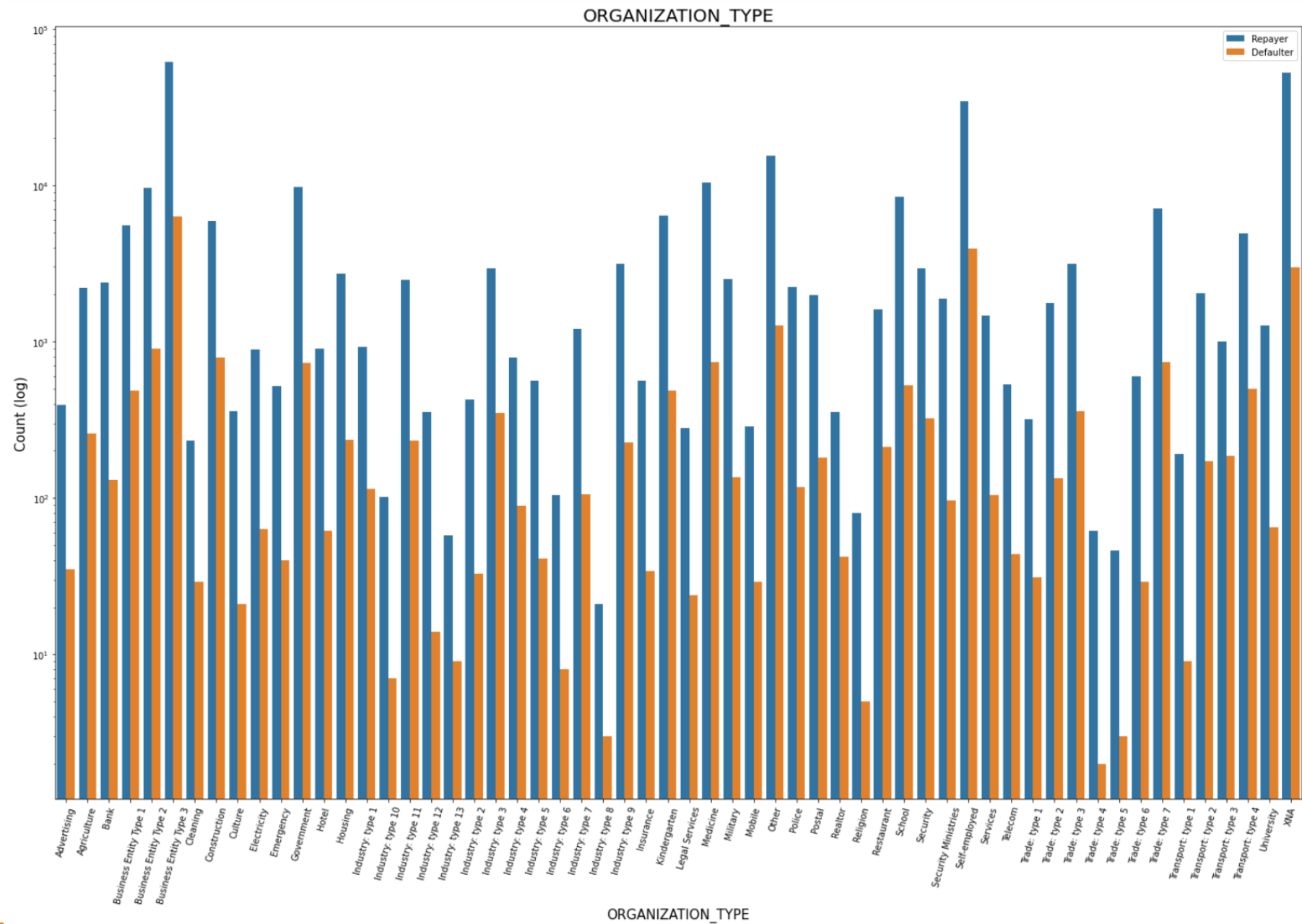


1. Most of applicants for loans income type is Working, followed by Commercial associate, Pensioner and State servant.

2. The applicants who are on Maternity leave have defaulting percentage of 40% which is the highest, followed by Unemployed (37%). The rest under average around 10% defaultees.

3. Student and Businessmen though less in numbers, do not have default record. Safest two categories for providing loan.





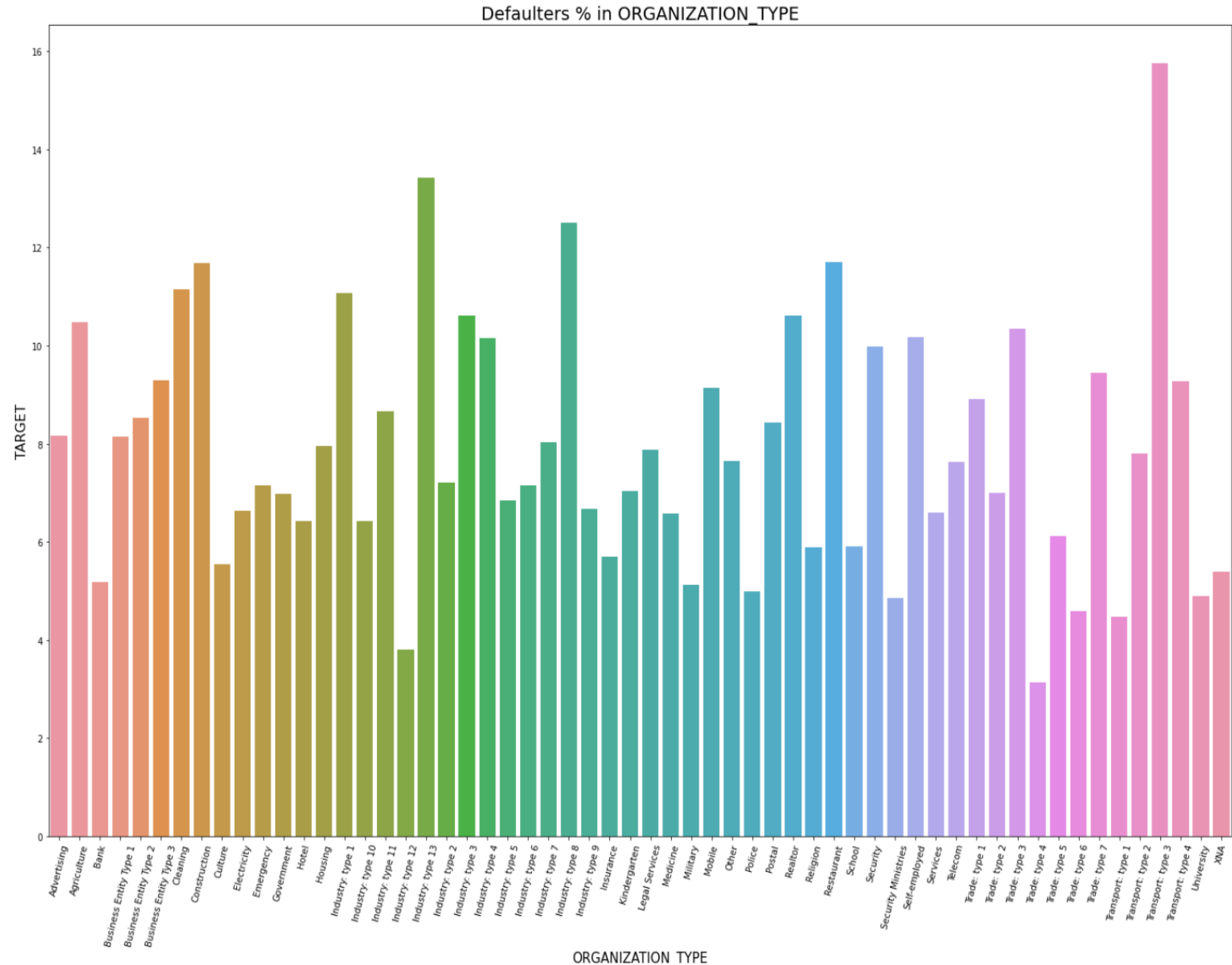
1. Organizations with highest percent of defaulters are Transport: type 3 (16%), Industry: type 13 (13.5%), Industry: type 8 (12.5%) and Restaurant (less than 12%).

2. Self employed people have relative high defaulting rate, to be safer side loan disbursement should be avoided or provide loan with higher interest rate to mitigate the risk of defaulting.

3. Most of the people application for loan are from Business Entity Type 3

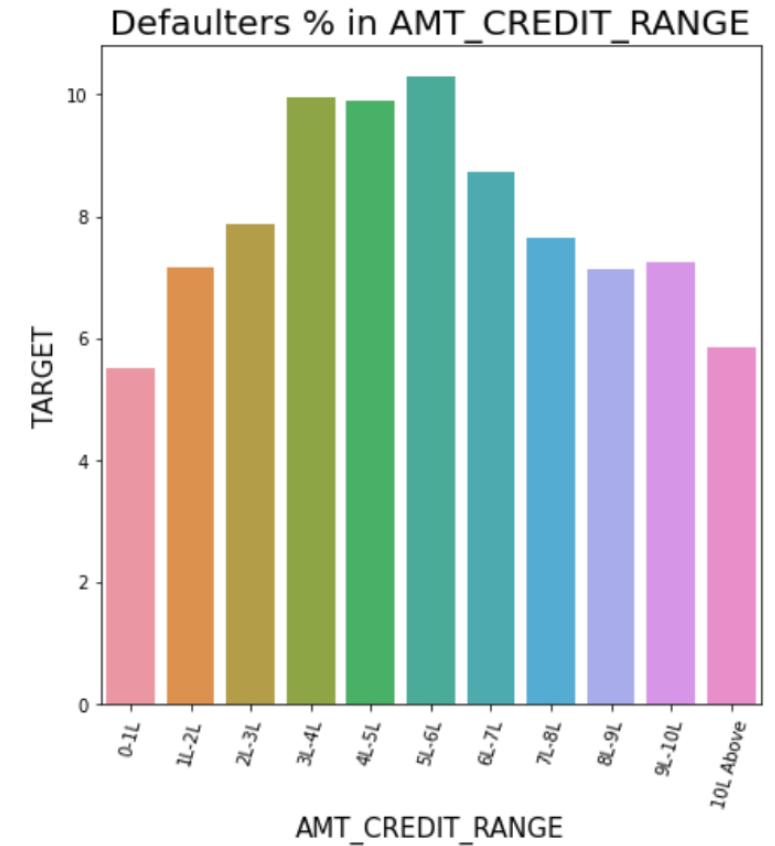
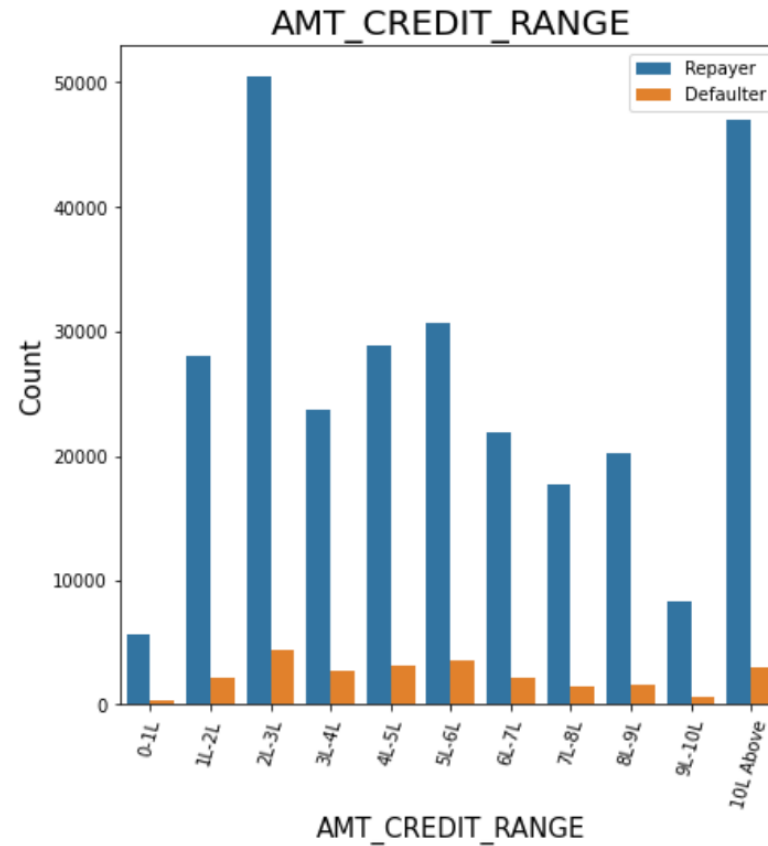
4. For a very high number of applications, Organization type information is unavailable(XNA)

5. It can be seen that following category of organization type has lesser defaulters thus safer for providing loans: Trade Type 4 and 5, Industry type 8



1. There are high number of applicants have loan in range of 2-3 Lakhs followed by 10 Lakh above range

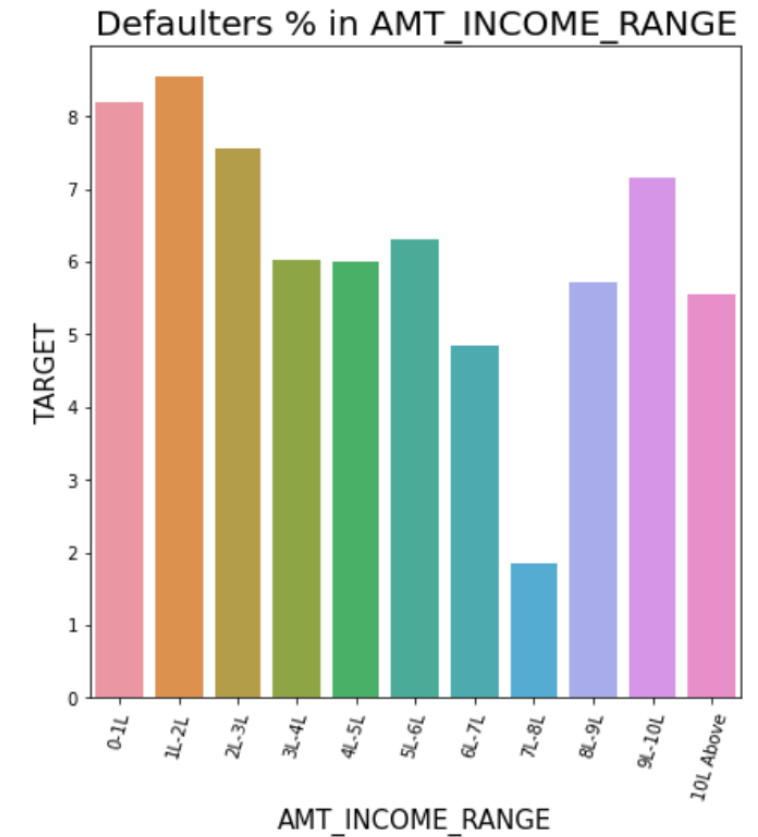
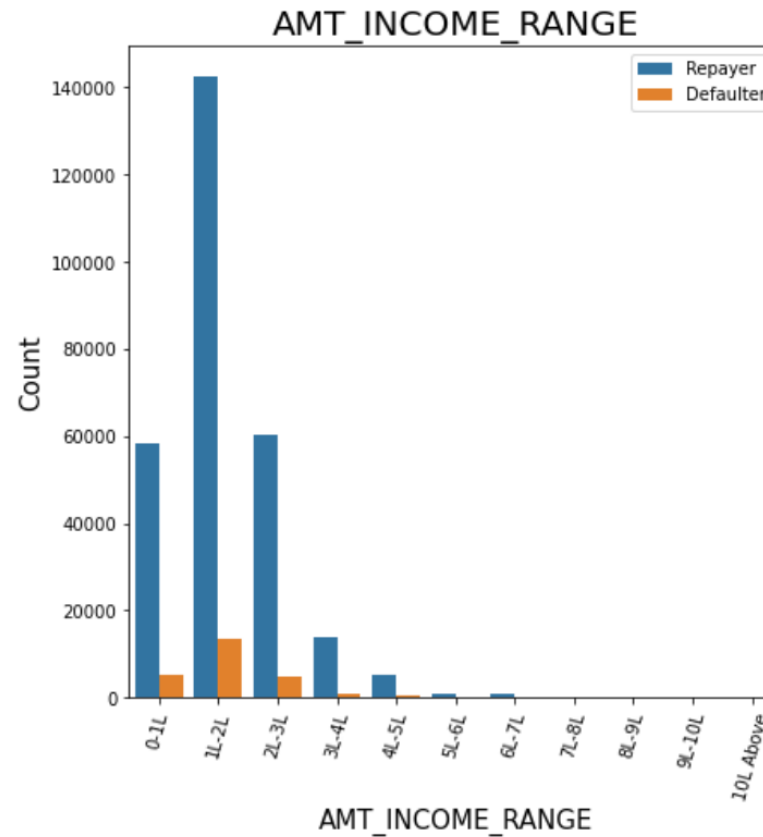
2. People who get loan for 3-6 Lakhs have most number of defaulters than other loan range.



1. Majority of the applications have Income total less than 3 Lakhs.

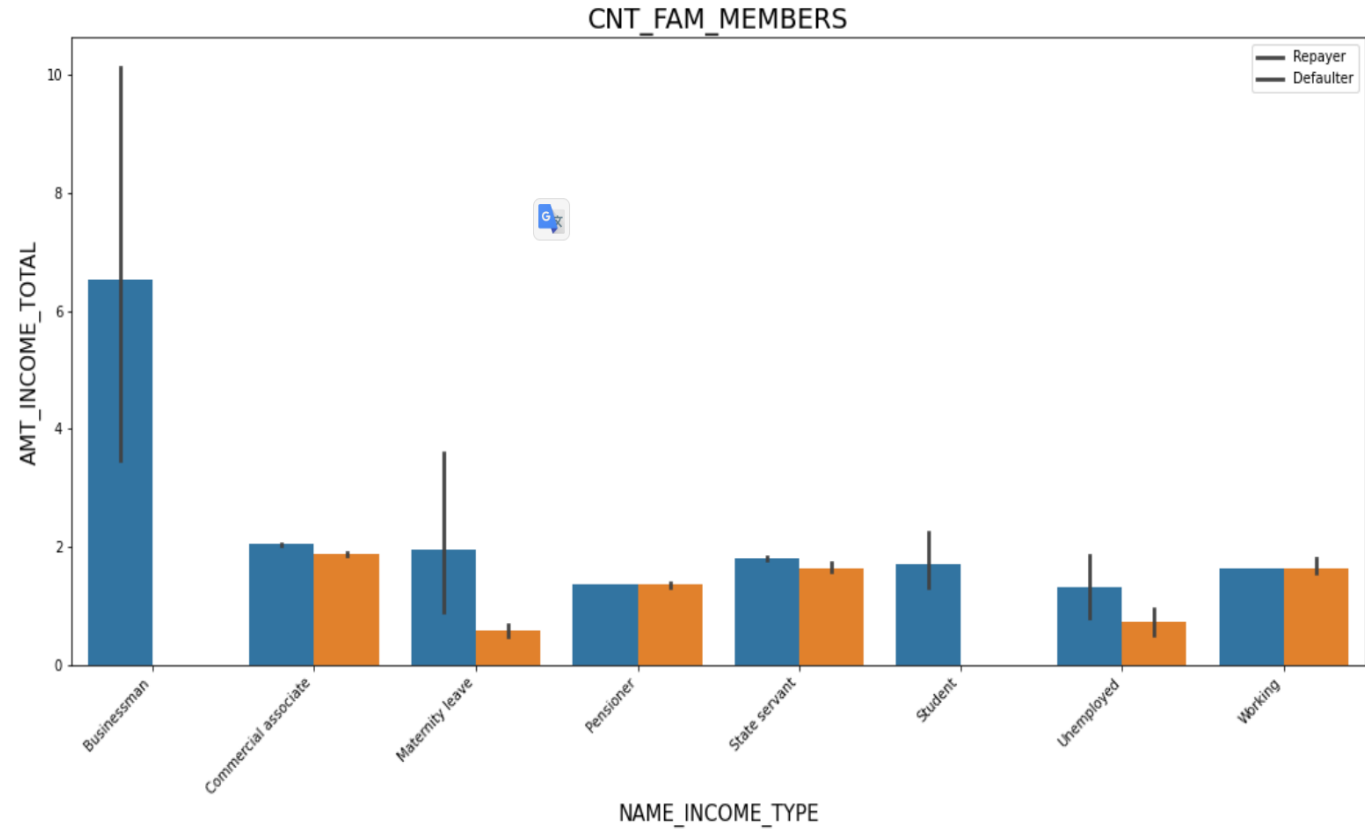
2. Application with Income less than 3 Lakhs has high probability of defaulting

3. Applicant with Income 7-8 Lakhs are less likely to default.



Categorical Bivariate or Multivariate Analysis

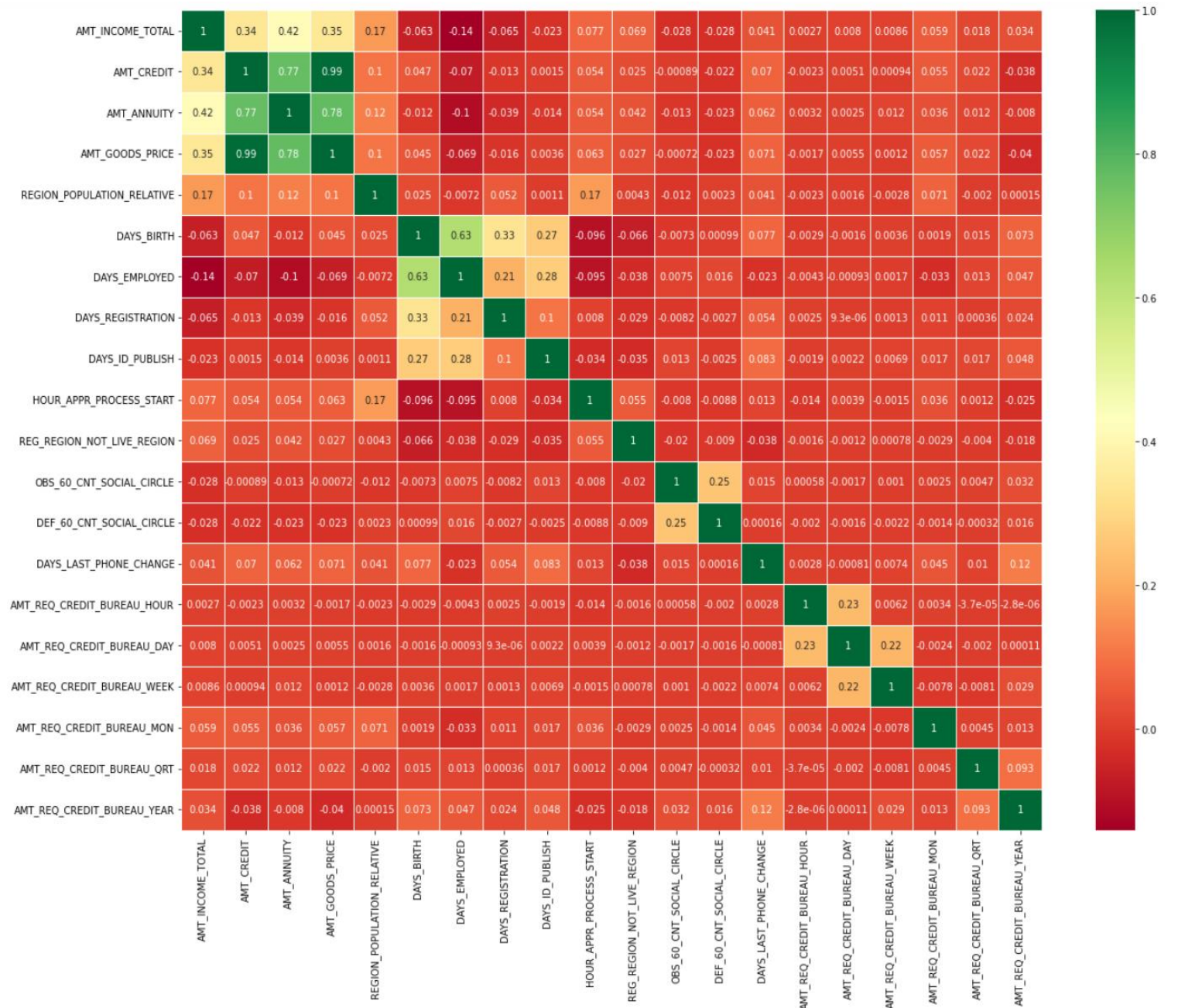
It can be seen that Businessman income is the highest and the estimated range with default 95% confidence level seem to indicate that the income of a Businessman could be in the range of slightly close to 4 lakhs and slightly above 10 lakhs



Correlation of Repayers and Defaulters

1. Credit amount is highly correlated with Goods Price Amount, Loan Annuity, Total Income

2. We can also see that repayers have high correlation in number of days employed.



1. Credit amount is highly correlated with good price amount which is same as repayers.

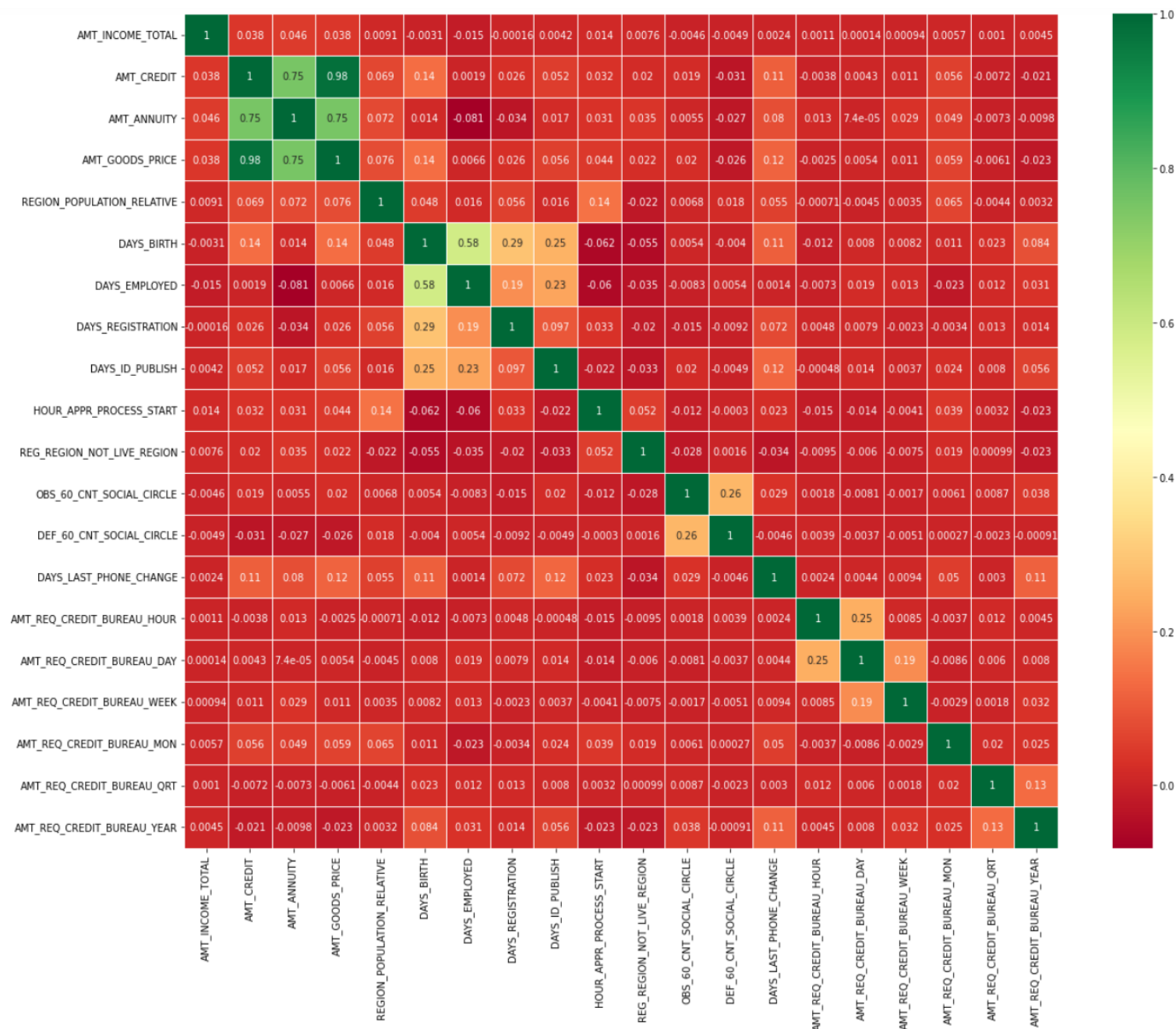
2. Loan annuity correlation with credit amount has slightly reduced in defaulters(0.75) when compared to repayers(0.77)

3. We can also see that repayers have high correlation in number of days employed(0.62) when compared to defaulters(0.58).

4. There is a severe drop in the correlation between total income of the client and the credit amount(0.038) amongst defaulters whereas it is 0.342 among repayers.

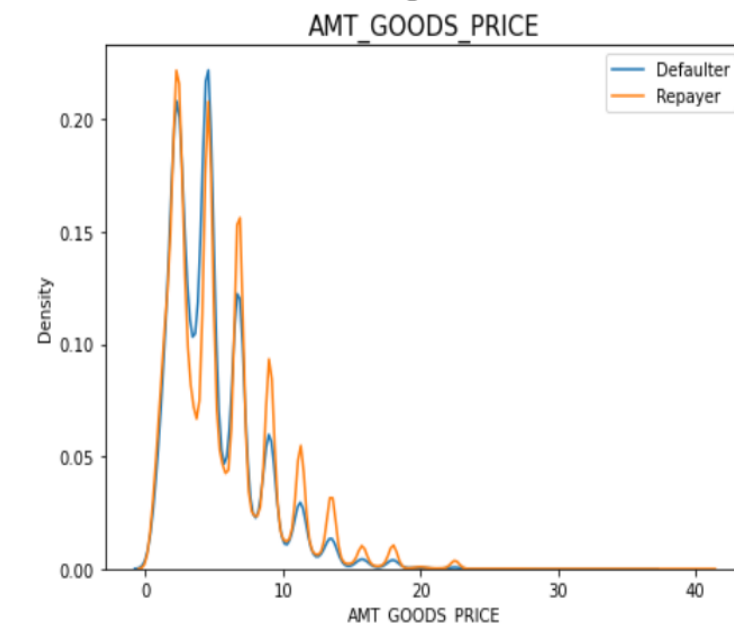
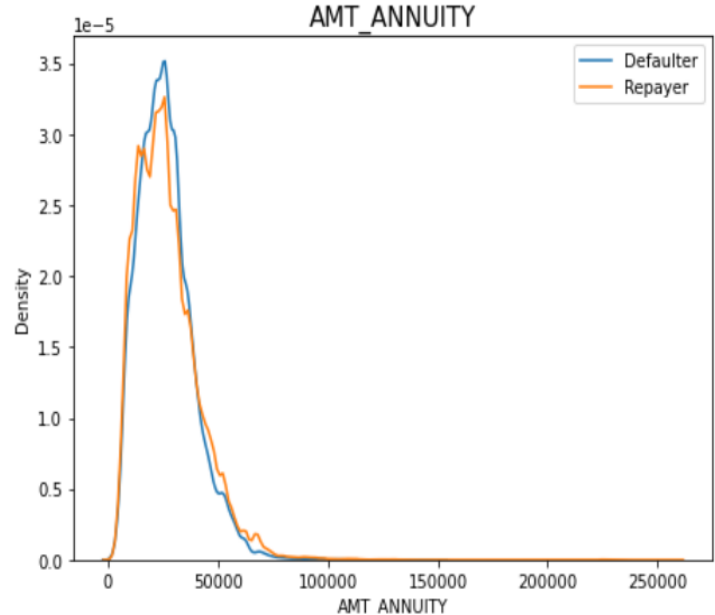
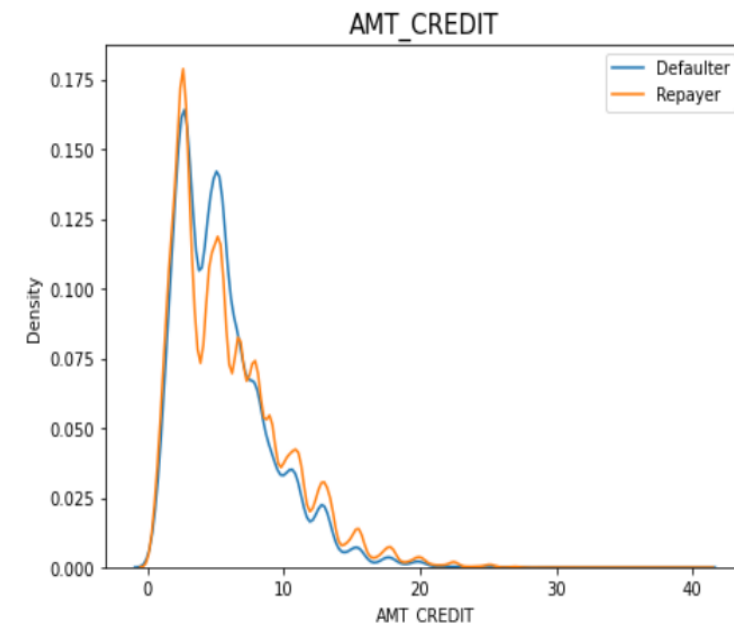
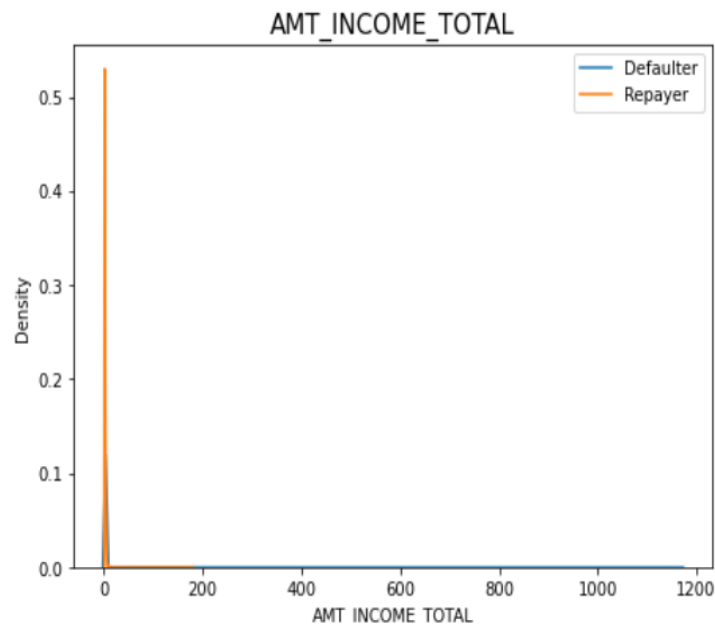
5. Days_birth and number of children correlation has reduced to 0.259 in defaulters when compared to 0.337 in repayers.

6. There is a slight increase in defaulted to observed count in social circle among defaulters(0.264) when compared to repayers(0.254)



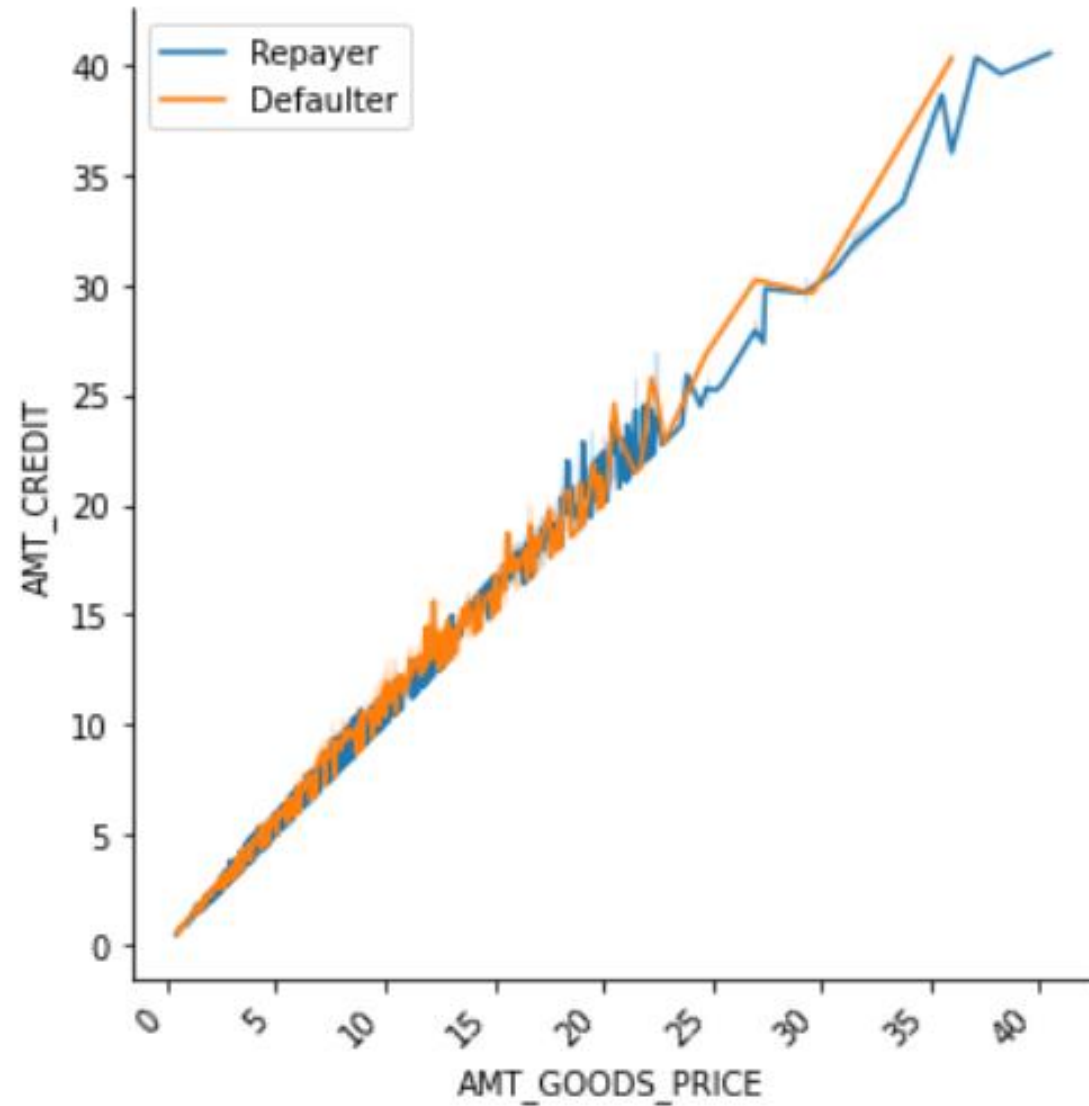
Numerical Univariate Analysis

1. Most no of loans are given for goods price below 10 lakhs
2. Most people pay annuity below 50K for the credit loan
3. Credit amount of the loan is mostly less then 10 lakhs
4. The repayers and defaulters distribution overlap in all the plots and hence we cannot use any of these variables in isolation to make a decision



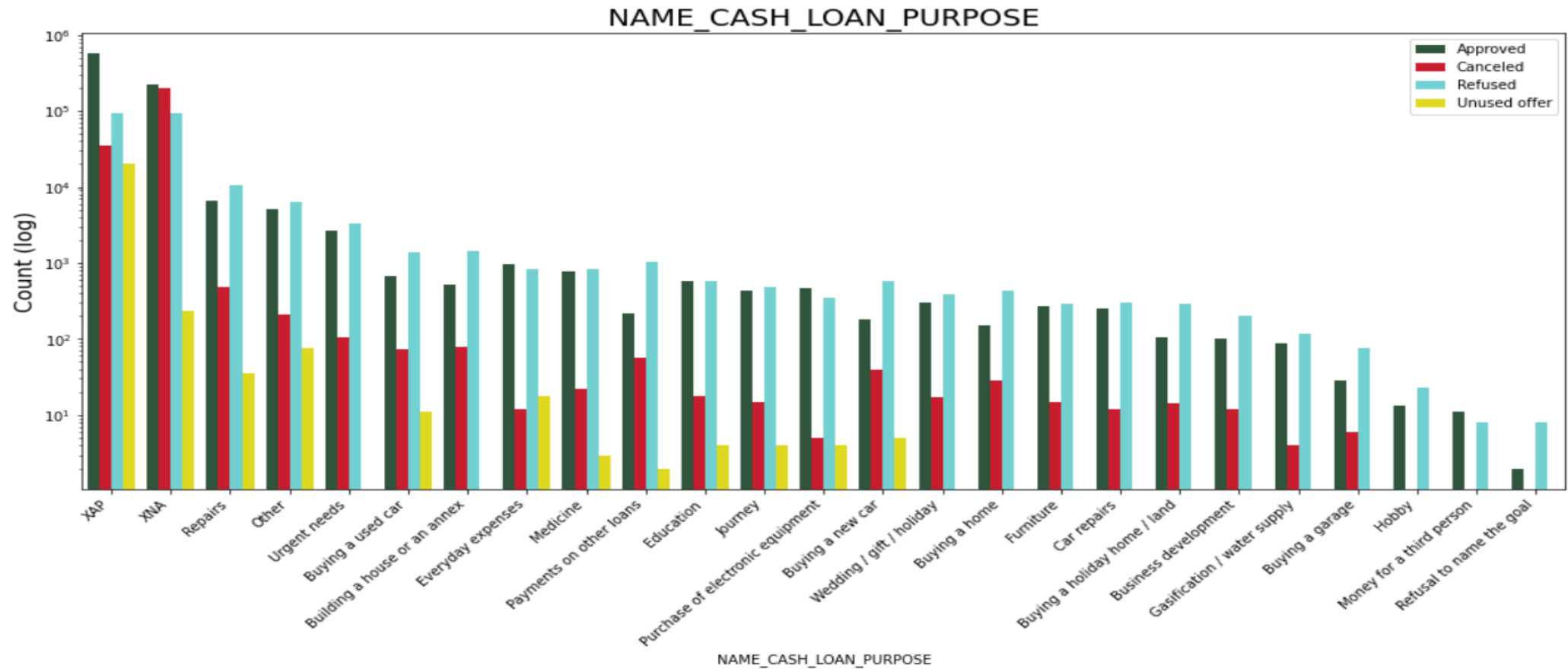
Numerical Bivariate Analysis

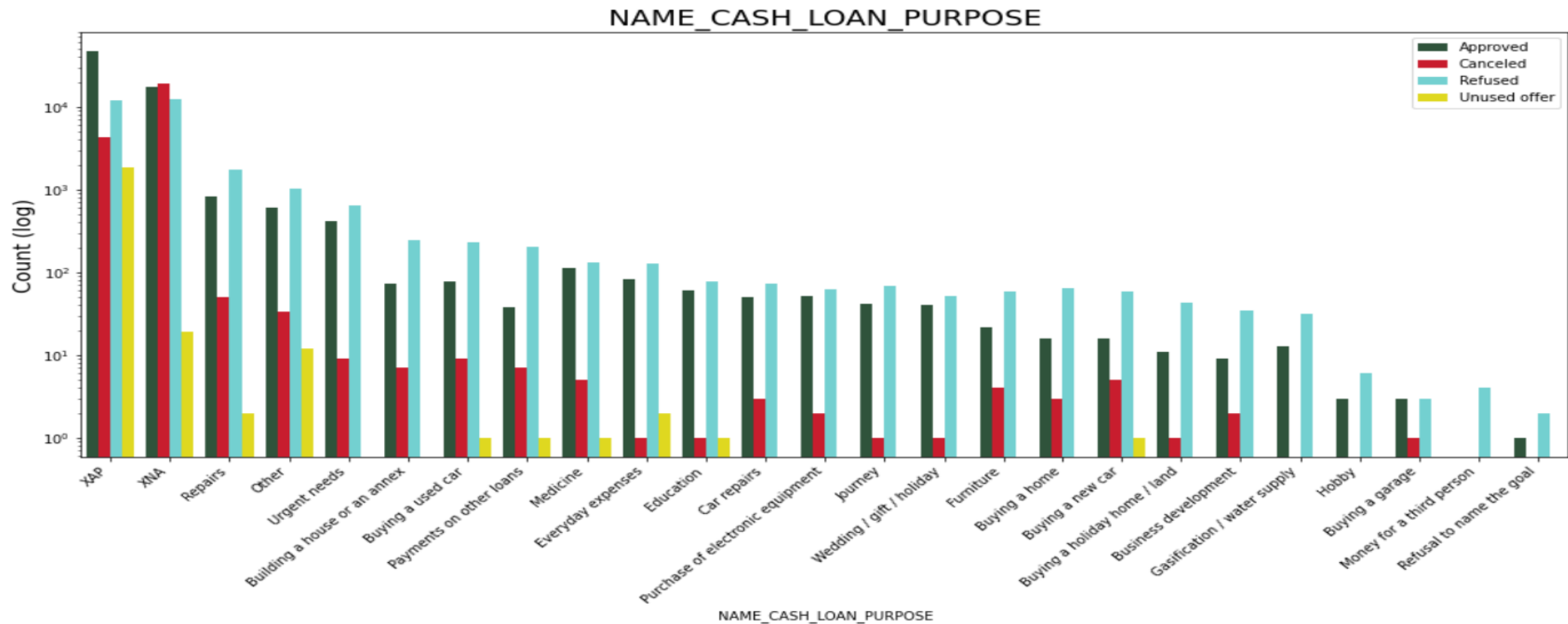
When the credit amount goes beyond 30 Lakhs, there is an increase in defaulters.



Merged Dataframes Analysis

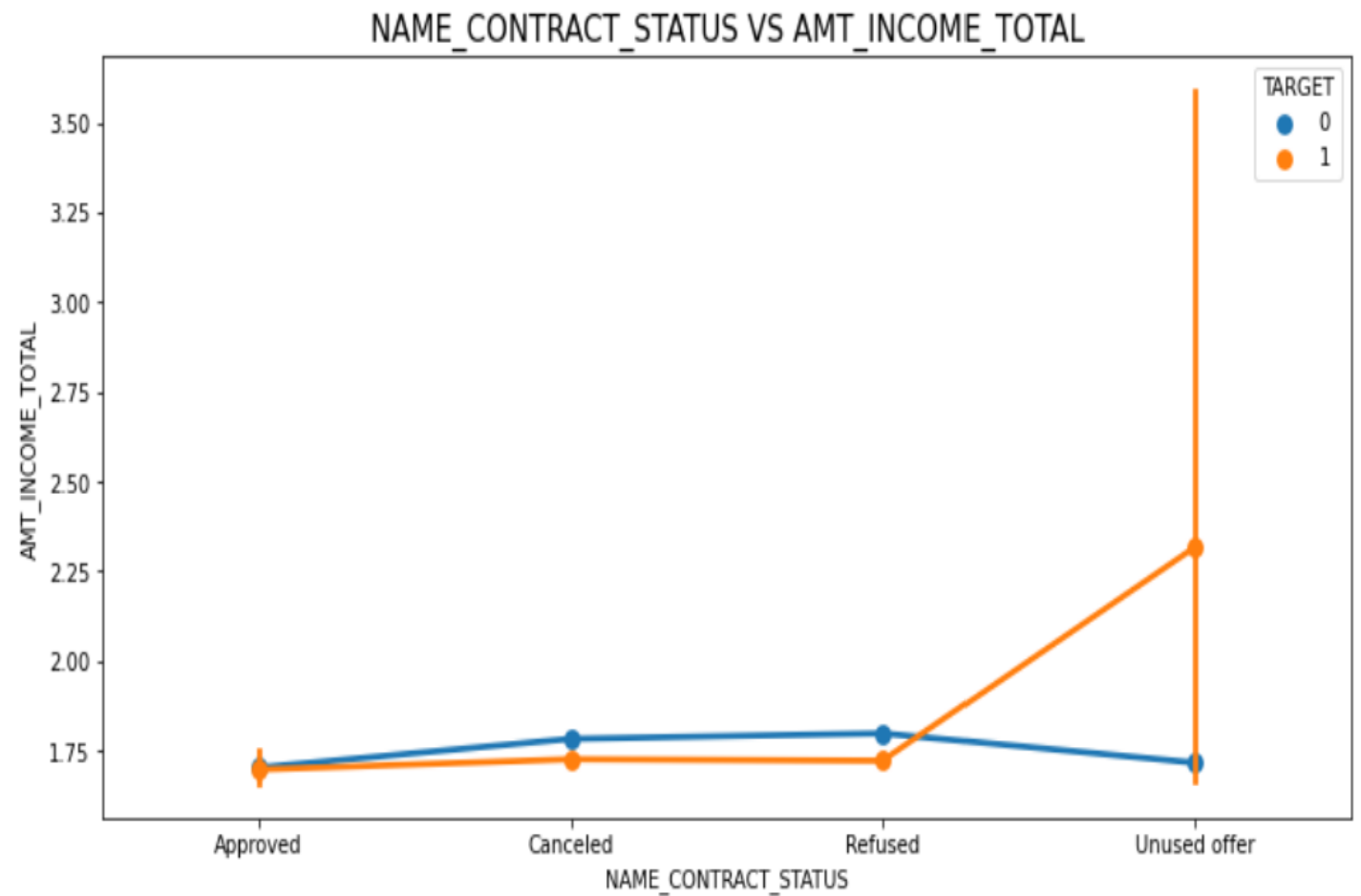
Plotting Contract Status vs purpose of the loan





1. Loan purpose has high number of unknown values (XAP, XNA)
2. Loan taken for the purpose of Repairs looks to have highest default rate
3. Huge number application have been rejected by bank or refused by client which are applied for Repair or Other. from this we can infer that repair is considered high risk by bank. Also, either they are rejected or bank offers loan on high interest rate which is not feasible by the clients and they refuse the loan.

The point plot show that the people who have not used offer earlier have defaulted even when there average income is higher than others



Conclusion

- **88% of the clients who were refused by bank for loan earlier have now turned into a repaying client. Hence documenting the reason for rejection could mitigate the business loss and these clients could be contacted for further loans.**
- **Applicant with Income more than 700,000 are less likely to default**
- **Clients who are either at Maternity leave OR Unemployed default a lot.**
- **People who get loan for 3-6 Lakhs tend to default more than others and hence having higher interest specifically for this credit range would be ideal.**
- **Since 90% of the applications have Income total less than 3Lakhs and they have high probability of defaulting, they could be offered loan with higher interest compared to other income category.**