LOAN CREDIT EDA CASE STUDY

By Pankaj Saha

PROBLEM STATEMENT:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

Note:- Python and Jupyter Notebook were used for this EDA.

Analysis Done:

- Check Missing Values, which to handle and how to handle
- Check Outliers, check Data Imbalance and ratio of Imbalance.
- First Create the correlation of Application and previous application dataset.
- Finally Merging both the correlation table of new application and previous application together.
- Aftermath visualising the merged dataframe to obtain the final result.

RESULT EXPECTED:

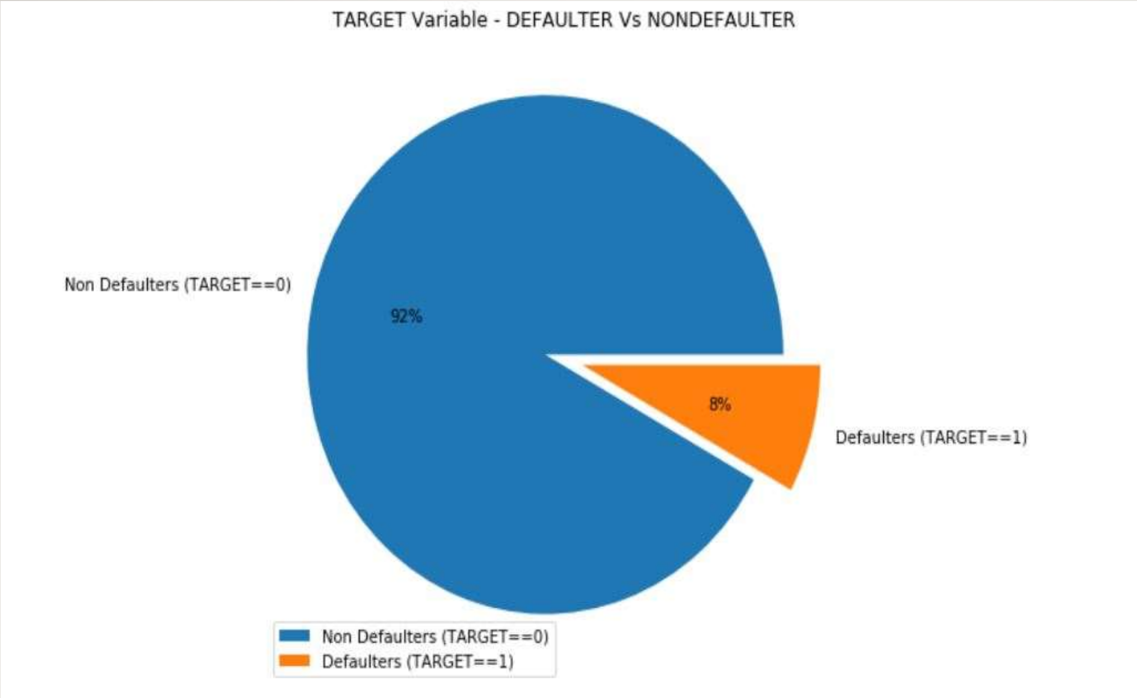
1) Approved: The Company has approved loan Application

2) Cancelled: The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.

3) Refused: The company had rejected the loan (because the client does not meet their requirements etc.

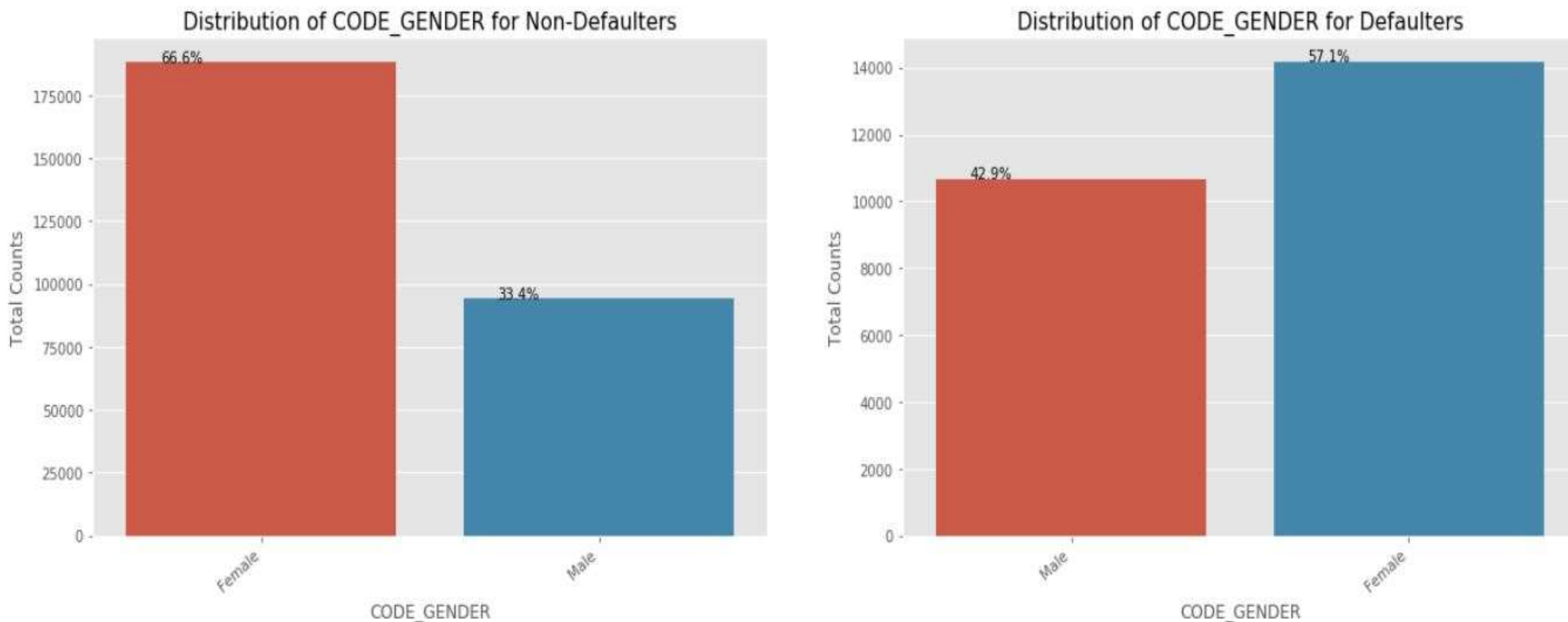
4) Unused offer: Loan has been cancelled by the client but on different stages of the process

PROPORTION OF DEFAULTERS AND NON DEFAULTERS IN THE DATASET:



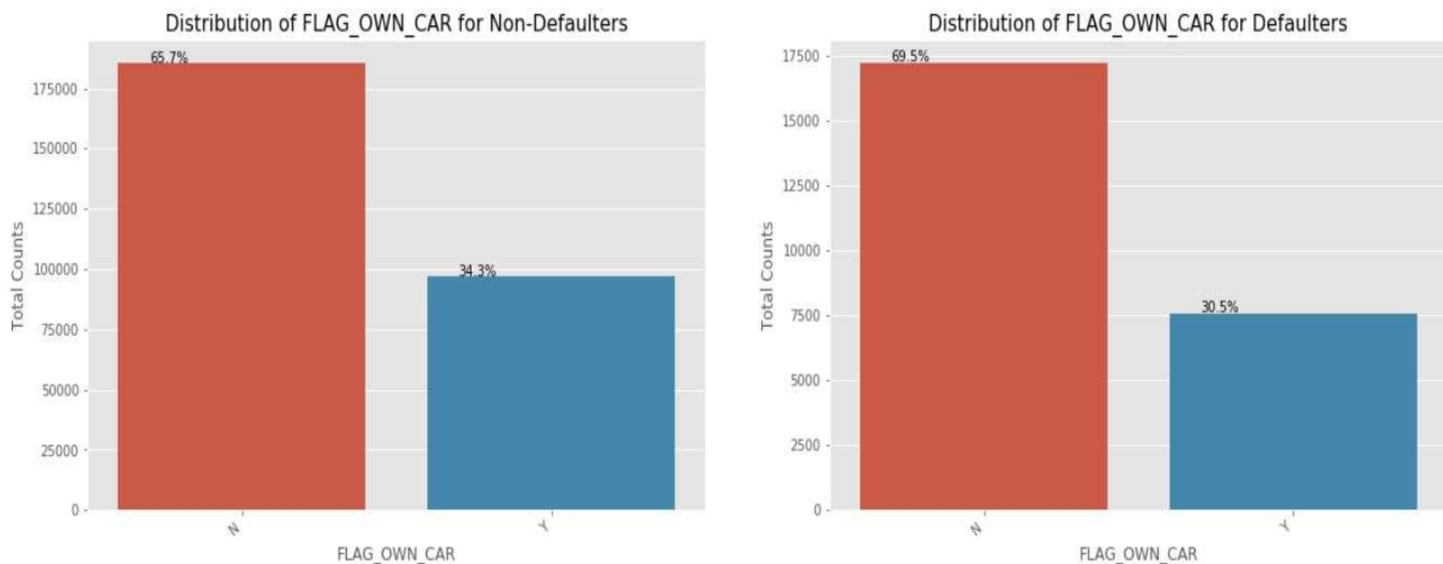
Here we can visualise that approx 8% people defaulted their loan by not paying any installment, where as approx 92% people were geniuengly paying the sum.

PROPORTION OF GENDER BASED ON DEFAULTERS AND NON DEFAULTERS IN THE DATASET:



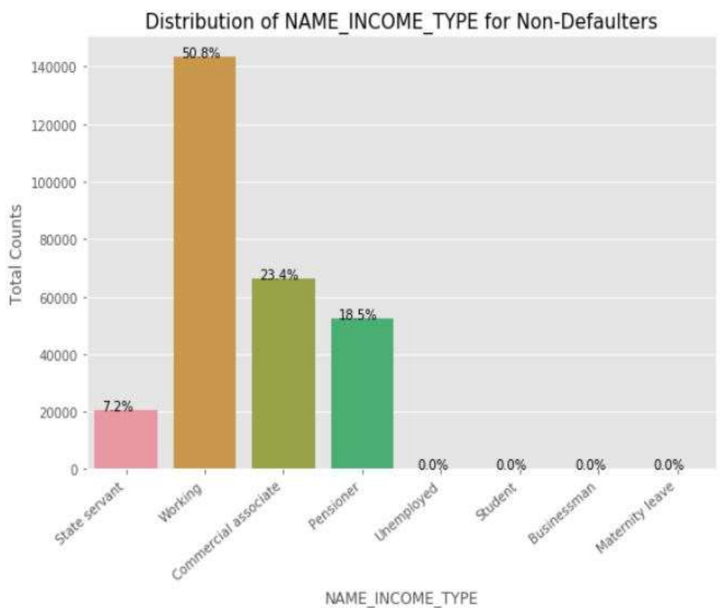
We can see that Female contribute 67% to the non-defaulters while 57% to the defaulters. We see more female applying for loans than males and hence the more number of female defaulters as well. **But the rate of defaulting of FEMALE is much lower compared to their MALE counterparts.**

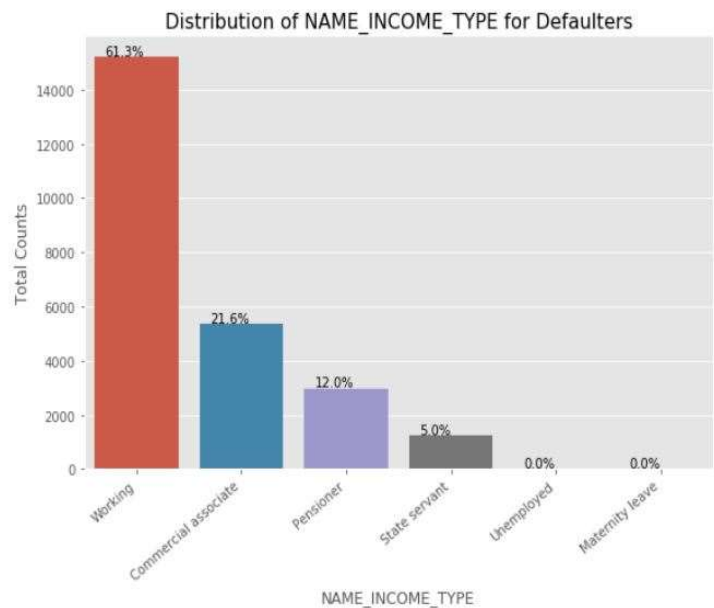
PROPORTION OF CAR OWNED BASED ON DEFAULTERS AND NON DEFAULTERS IN THE DATASET:



We can see that people with cars contribute 65.7% to the non-defaulters while 69.5% to the defaulters. We can conclude that While people who have car default more often, the reason could be there are simply more people without cars Looking at the percentages in both the charts, we can conclude that the rate of default of people having car is low compared to people who don't.

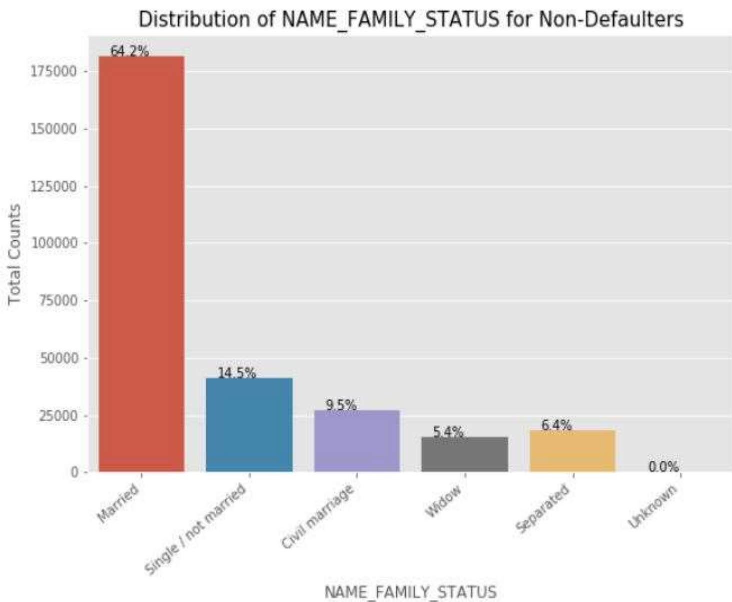
PROPORTION OF INCOME NAME TYPE BASED ON DEFAULTERS AND NON DEFAULTERS IN THE DATASET:

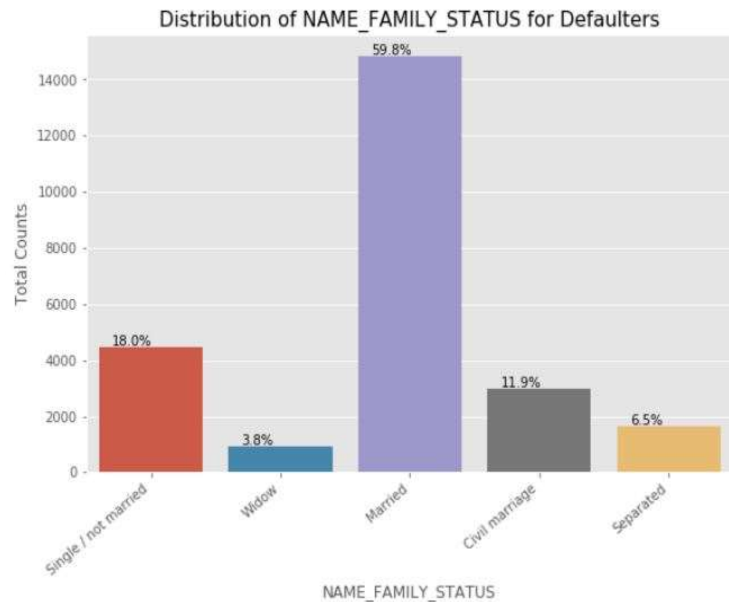




We can notice that the students don't default. The reason could be they are not required to pay during the time they are students. We can also see that the Business Men never default. Most of the loans are distributed to working class people We also see that working class people contribute 51% to non defaulters while they contribute to 61% of the defaulters. Clearly, the chances of defaulting are more in their case.

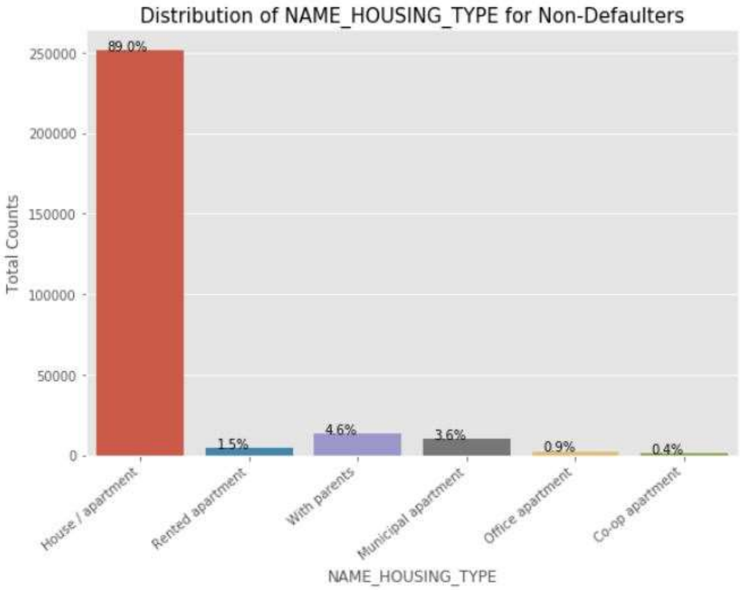
PROPORTION OF NAME-FAMILY-STATUS BASED ON DEFAULTERS AND NON DEFAULTERS IN THE DATASET:

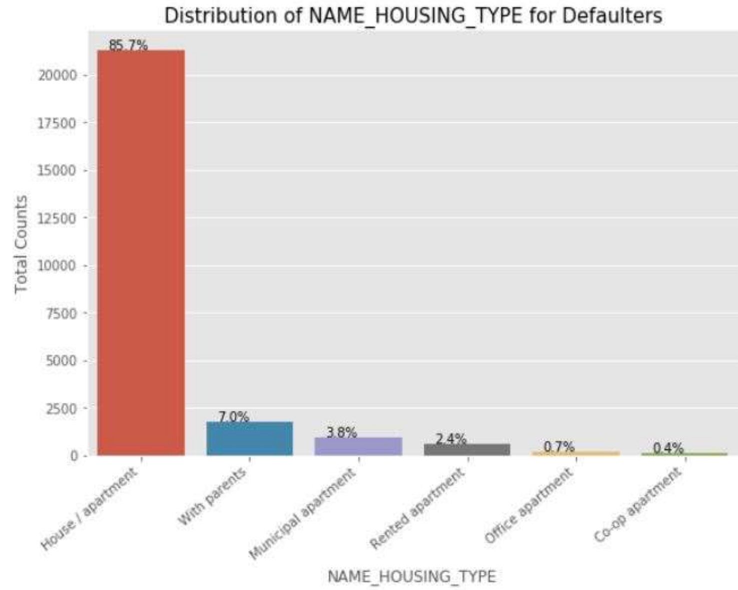




Married people tend to apply for more loans comparatively. But from the graph we see that Single/non Married people contribute 14.5% to Non Defaulters and 18% to the defaulters. So there is more risk associated with them.

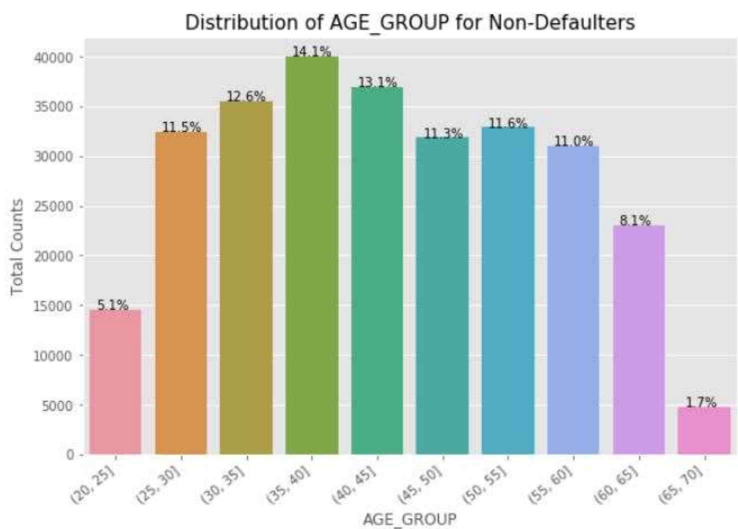
PROPORTION OF NAME-HOUSING_TYPE BASED ON DEFAULTERS AND NON DEFAULTERS IN THE DATASET:

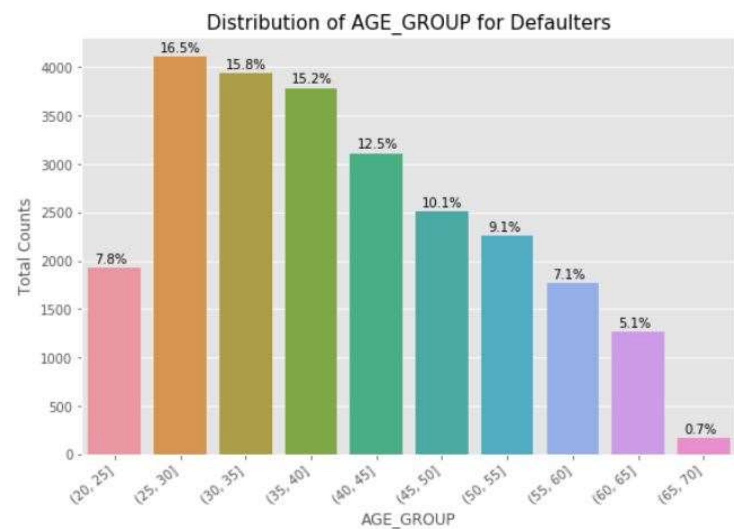




It is clear from the graph that people who have House/Appartment, tend to apply for more loans. People living with parents tend to default more often when compared with others. The reason could be their living expenses are more due to their parents living with them.

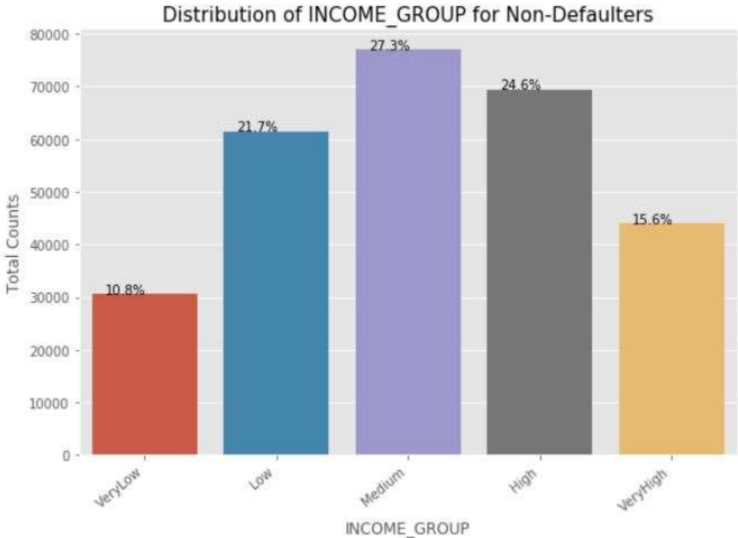
PROPORTION OF AGE-GROUP BASED ON DEFAULTERS AND NON DEFAULTERS IN THE DATASET:

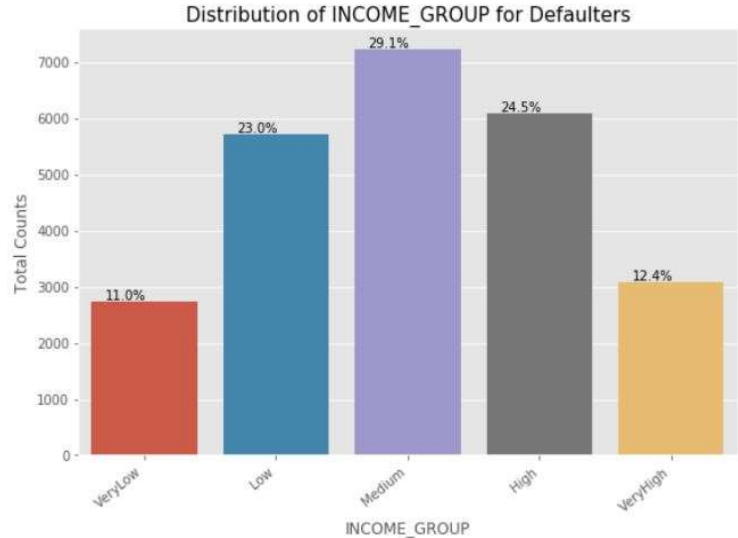




We see that (25,30] age group tend to default more often. So they are the riskiest people to loan to. With increasing age group, people tend to default less starting from the age 25. One of the reasons could be they get employed around that age and with increasing age, their salary also increases.

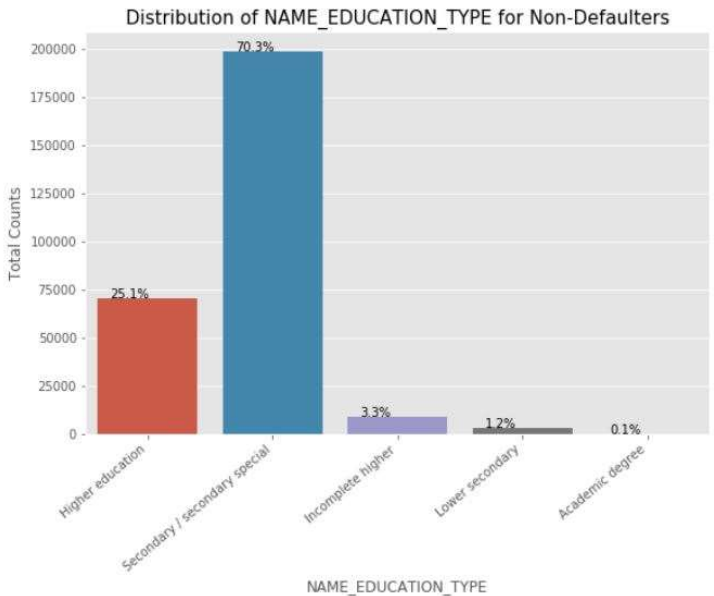
PROPORTION OF INCOME-GROUP BASED ON DEFAULTERS AND NON DEFAULTERS IN THE DATASET:

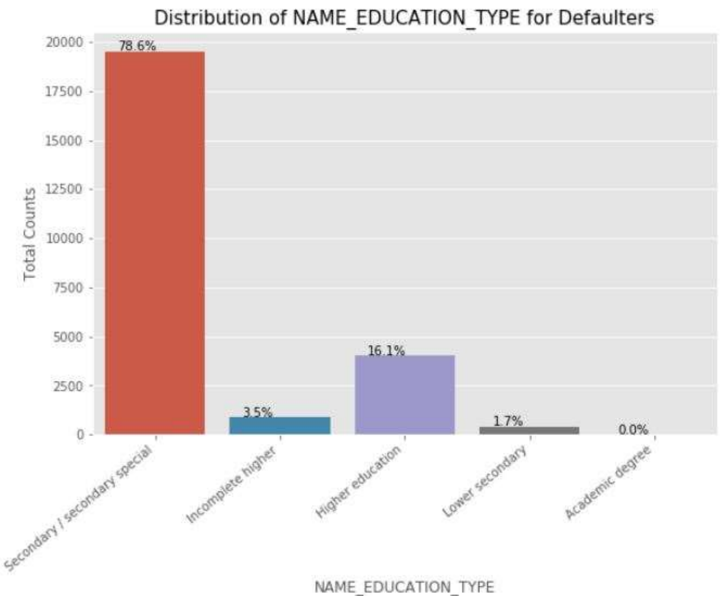




The Very High income group tend to default less often. They contribute 12.4% to the total number of defaulters, while they contribute 15.6% to the Non-Defaulters.

PROPORTION OF NAME-EDUCATION-TYPE BASED ON DEFAULTERS AND NON DEFAULTERS IN THE DATASET:



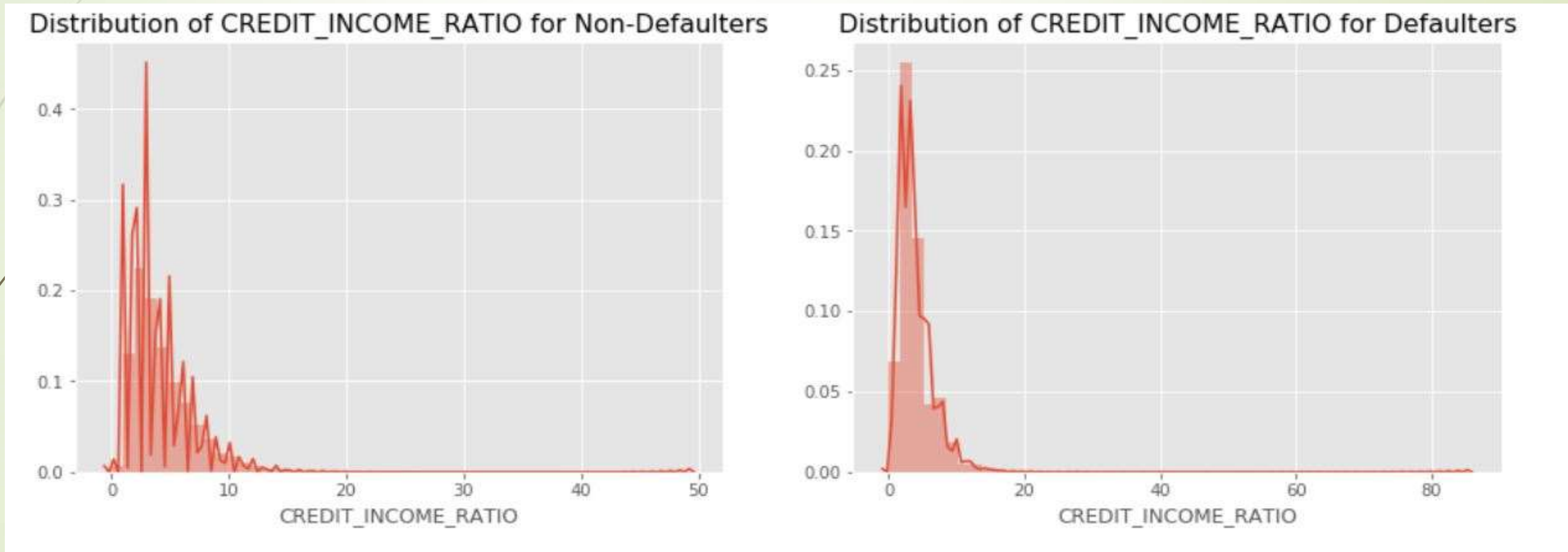


Almost all of the Education categories are equally likely to default except for the higher educated ones who are less likely to default and secondary educated people are more likely to default

PROPORTION OF REGION-RATING-CLIENT BASED ON DEFAULTERS AND NON DEFAULTERS IN THE DATASET:

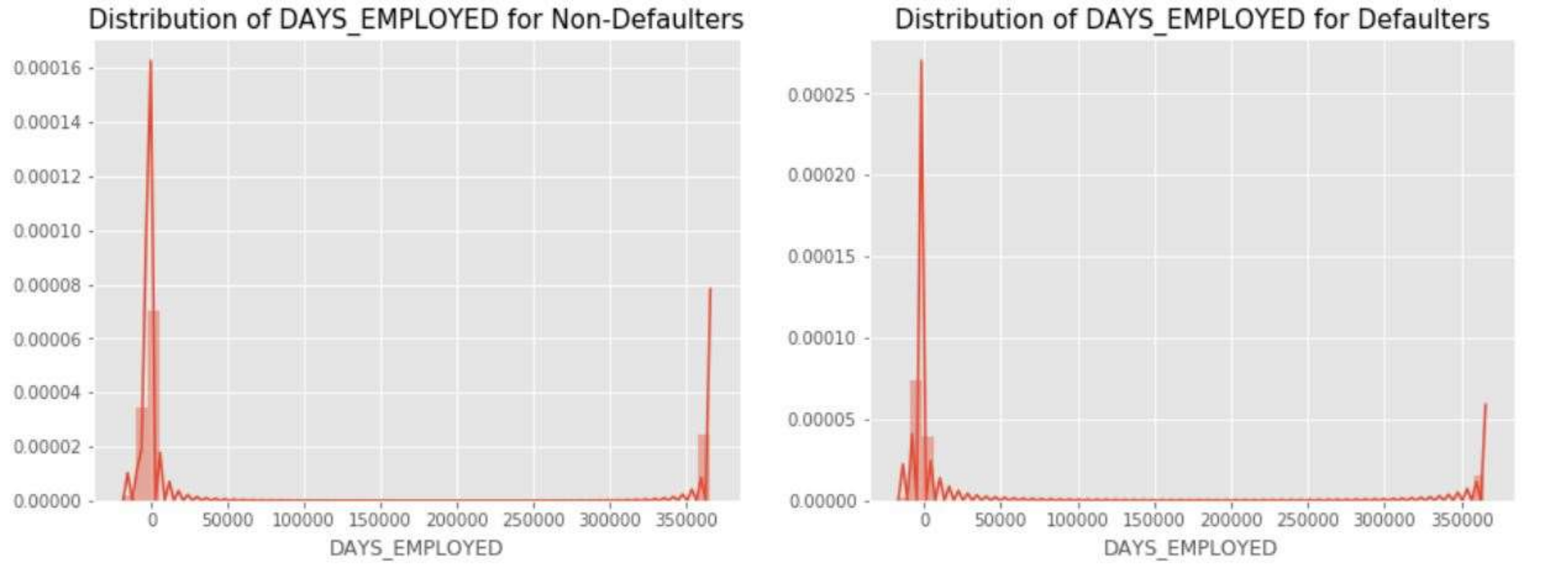
More people from second tier regions tend to apply for loans. We can infer that people living in better areas(Rating 3) tend contribute more to the defaulters by their weightage. People living in 1 rated areas

PROPORTION OF CREDIT-INCOME-RATIO BASED ON DEFAULTERS AND NON DEFAULTERS IN THE DATASET:

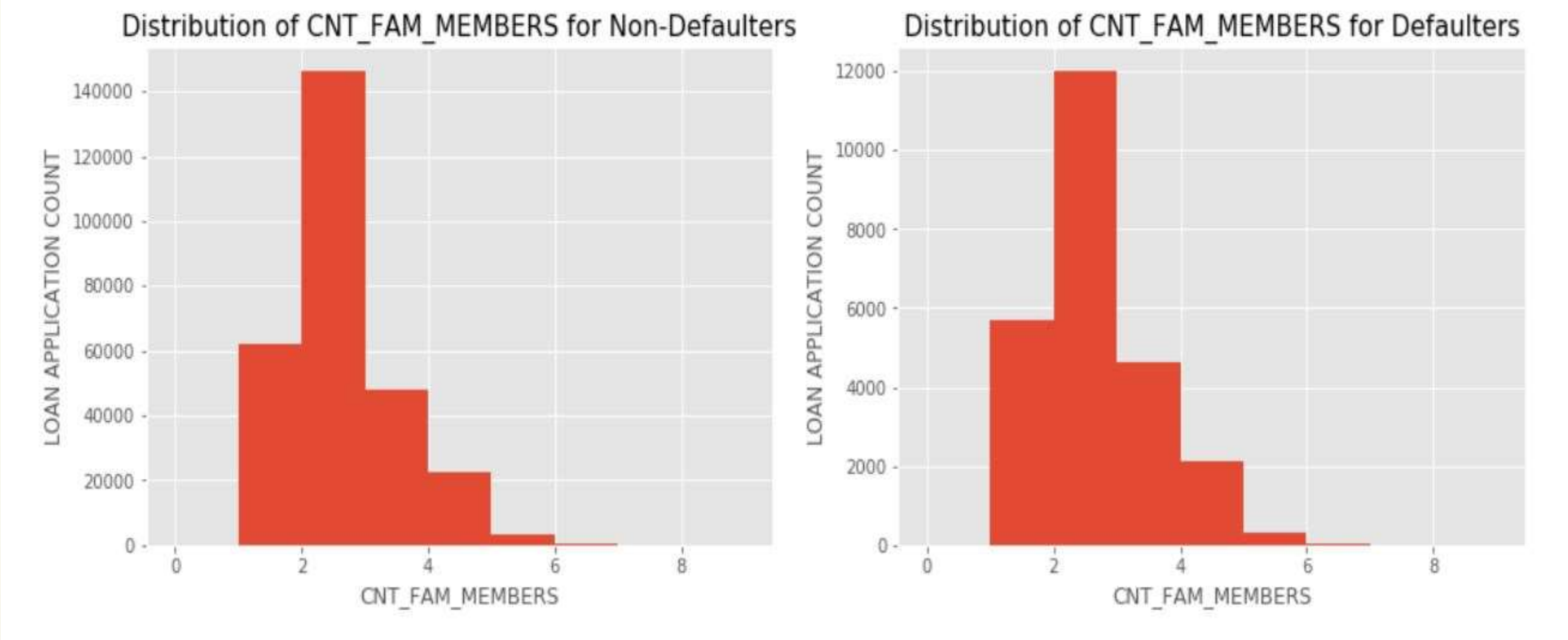


Credit income ratio the ratio of AMT_CREDIT/AMT_INCOME_TOTAL. Although there doesn't seem to be a clear distiguish between the group which defaulted vs the group which didn't when compared using the ratio, we can see that when the CREDIT_INCOME_RATIO is more than 50, people default.

PROPORTION OF DAYS-EMPLOYED BASED ON DEFAULTERS AND NON DEFAULTERS IN THE DATASET:

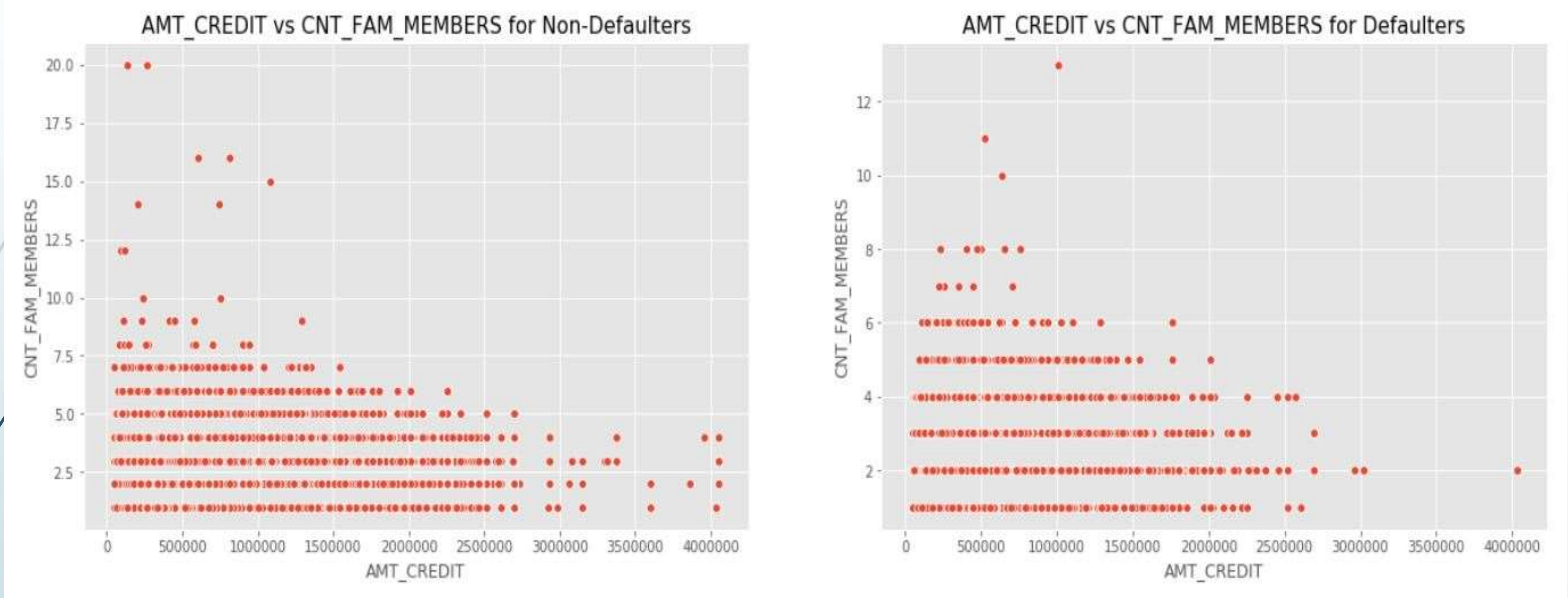


PROPORTION OF CNT-FAM-MEMBERS BASED ON DEFAULTERS AND NON DEFAULTERS IN THE DATASET:



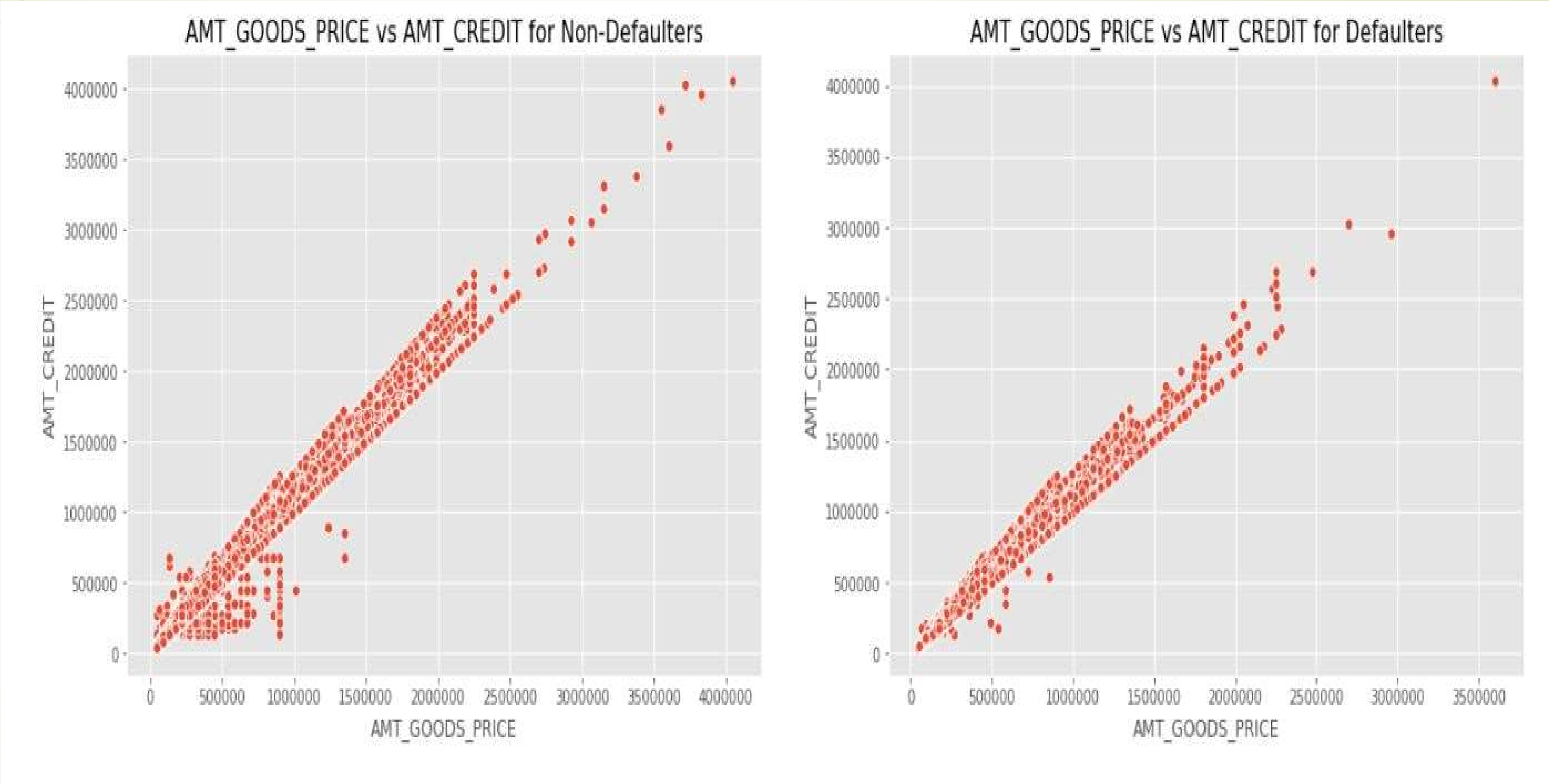
We can see that a family of 3 applies loan more often than the other families

AMT-CREDIT Vs CNT-FAM-MEMBERS

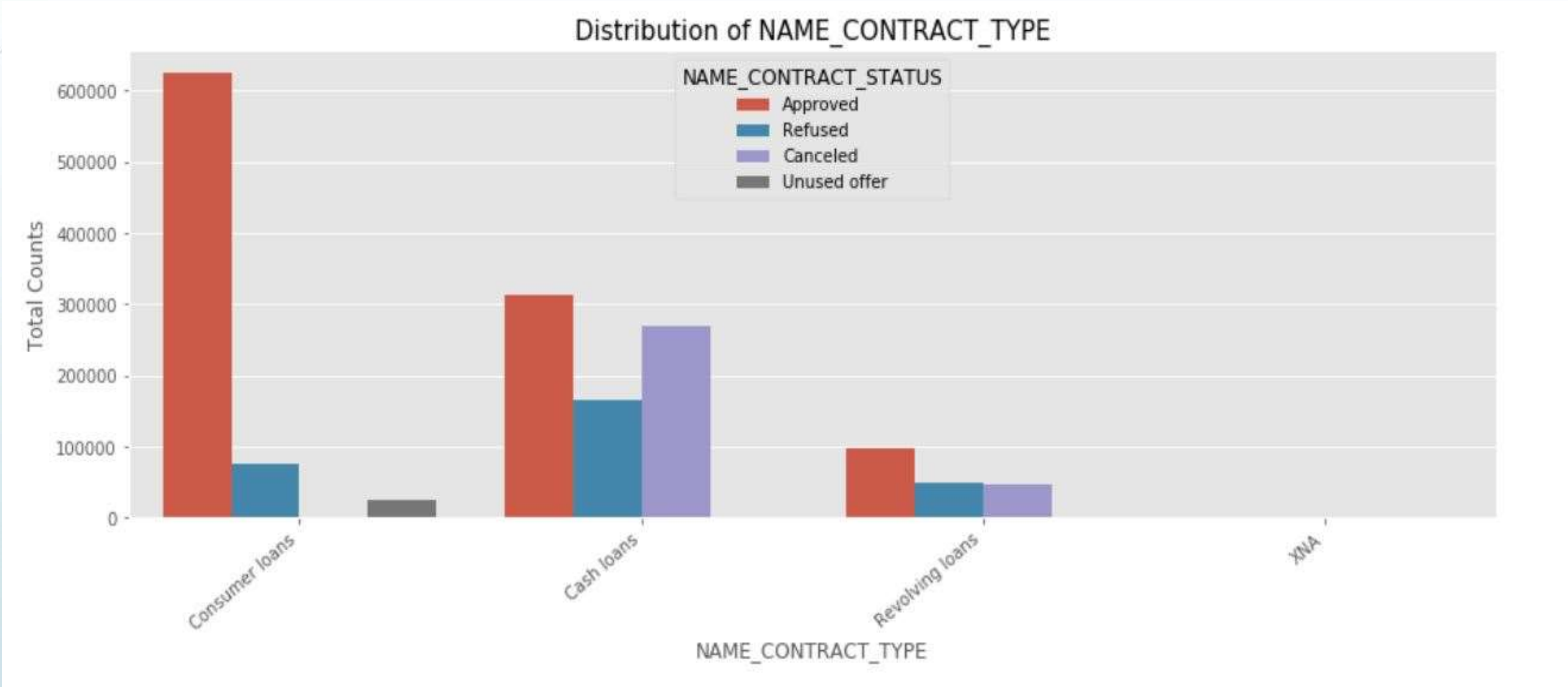


We can see that the density in the lower left corner is similar in both the case, so the people are equally likely to default if the family is small and the AMT_CREDIT is low. We can observe that larger families and people with larger AMT_CREDIT default less often

AMT_GOODS_PRICE Vs AMT_CREDIT

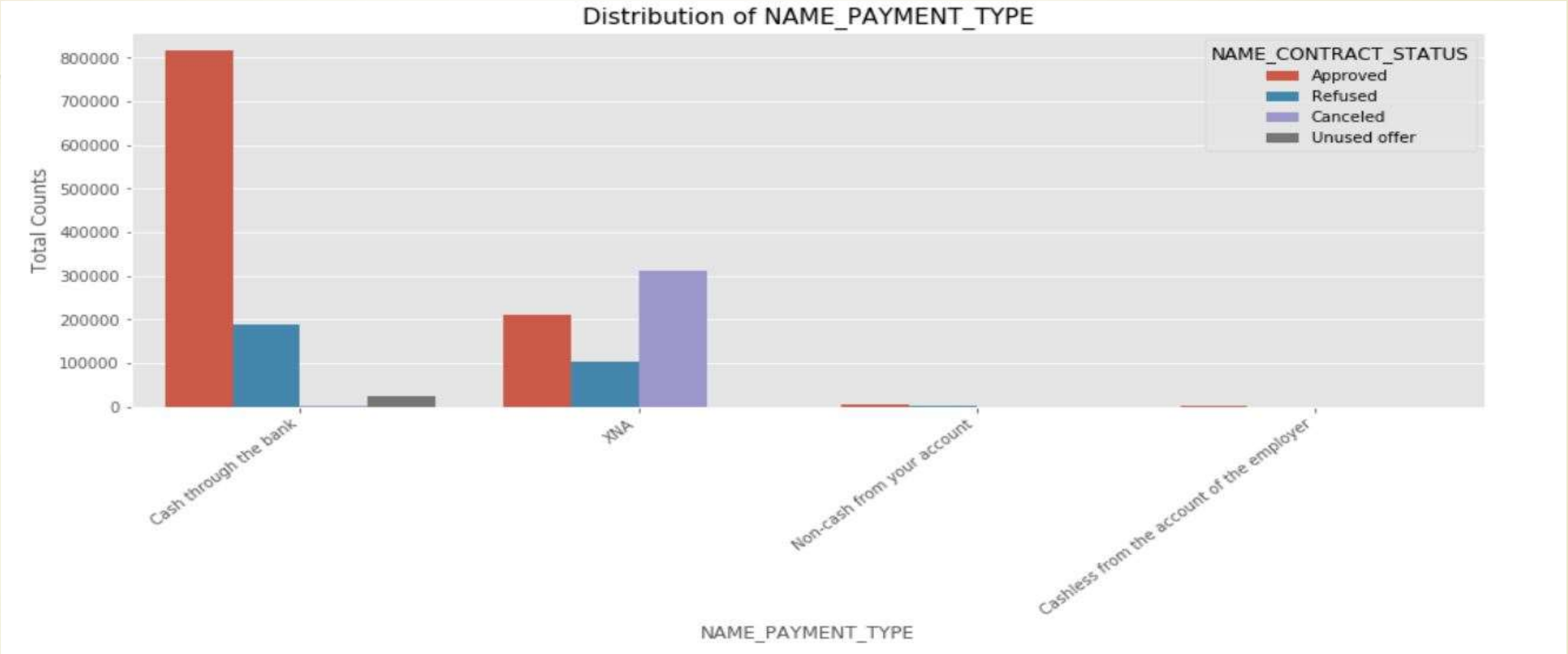


PROPORTION OF NAME_CONTRACT_TYPE



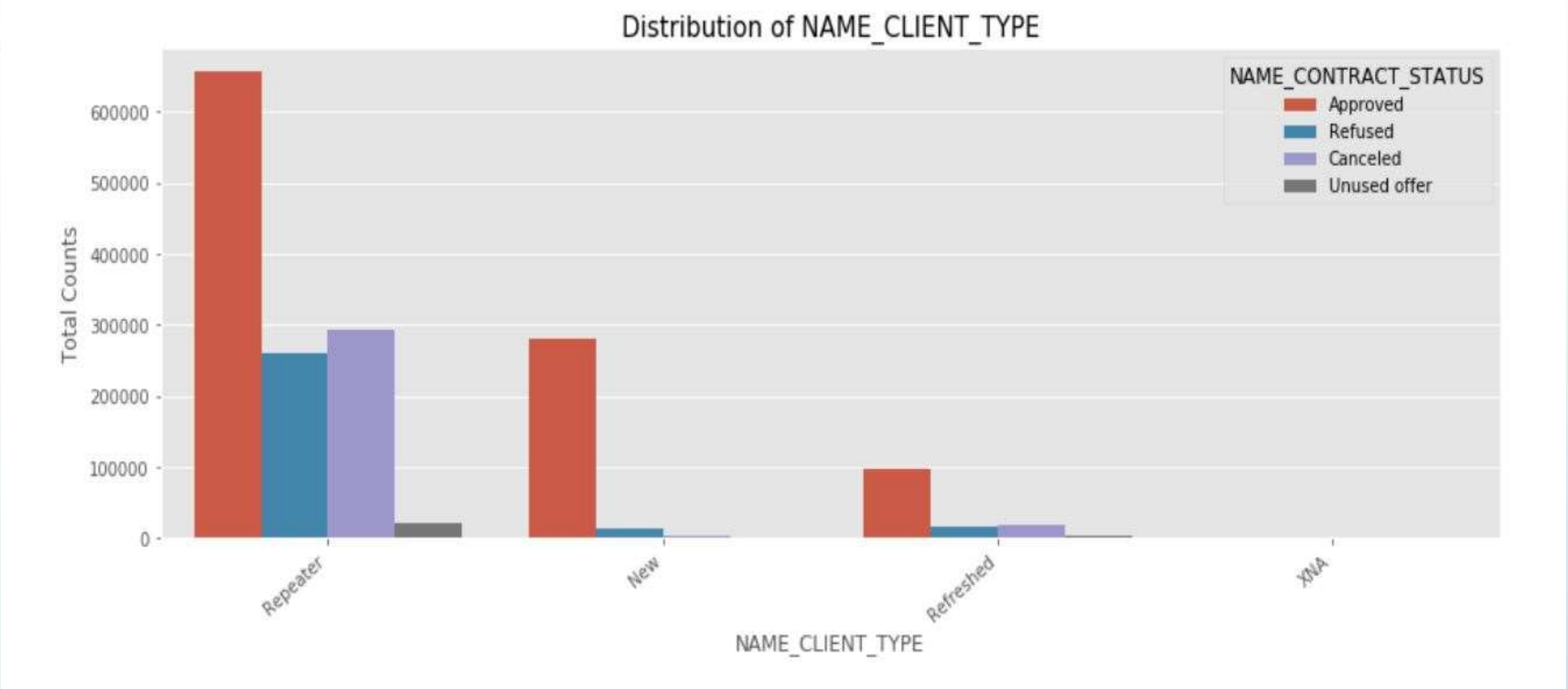
From the above chart, we can infer that, most of the applications are for 'Cash loan' and 'Consumer loan'. Although the cash loans are refused more often than others.

PROPORTION OF NAME_PAYMENT_TYPE



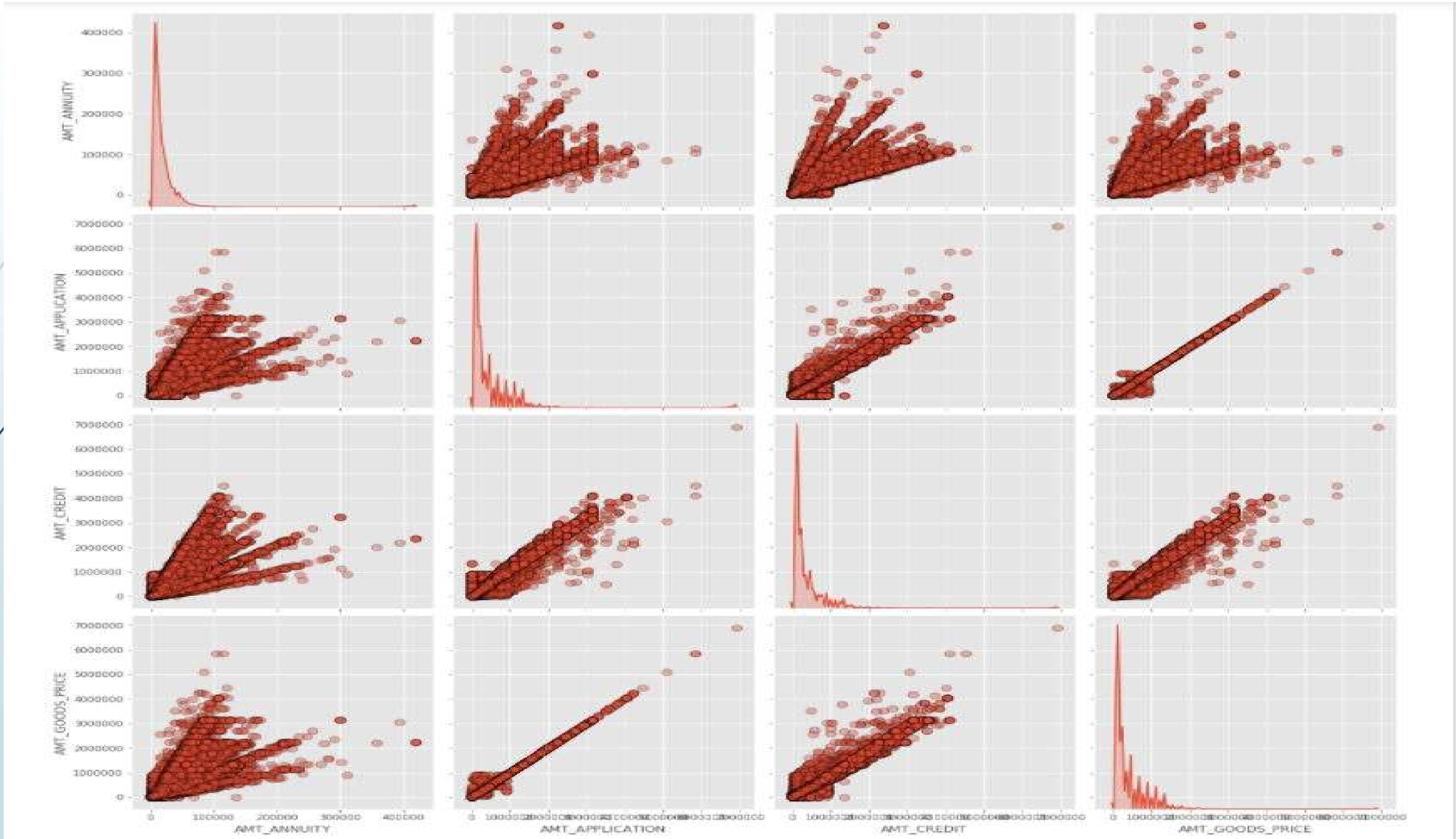
From the above chart, we can infer that most of the clients chose to repay the loan using the 'Cash through the bank' option We can also see that 'Non-Cash from your account' & 'Cashless from the account of the employee' options are not at all popular in terms of loan repayment amongst the customers

PROPORTION OF NAME_CLIENT_TYPE



Most of the loan applications are from repeat customers, out of the total applications 70% of customers are repeaters. They also get refused most often.

Bivariate analysis on numerical columns



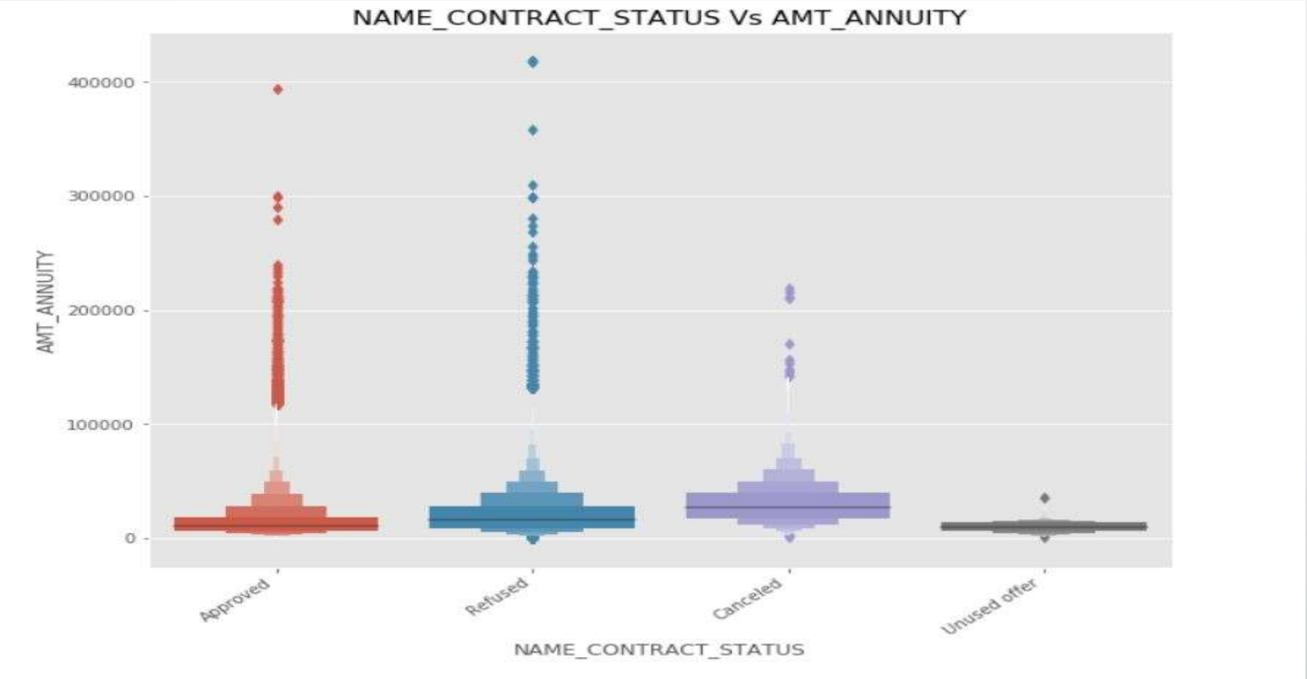
ANNUITY OF PREVIOUS APPLICATION HAS A VERY HIGH AND POSITIVE INFLUENCE OVER:
(INCREASE OF ANNUITY INCREASES BELOW FACTORS)

(1) HOW MUCH CREDIT DID CLIENT ASKED ON THE PREVIOUS APPLICATION
(2)FINAL CREDIT AMOUNT ON THE PREVIOUS APPLICATION THAT WAS APPROVED BY THE BANK
(3) GOODS PRICE OF GOOD THAT CLIENT ASKED FOR ON THE PREVIOUS APPLICATION.

FOR HOW MUCH CREDIT DID CLIENT ASK ON THE PREVIOUS APPLICATION IS HIGHLY INFLUENCED BY THE GOODS PRICE OF GOOD THAT CLIENT HAS ASKED FOR ON THE PREVIOUS APPLICATION

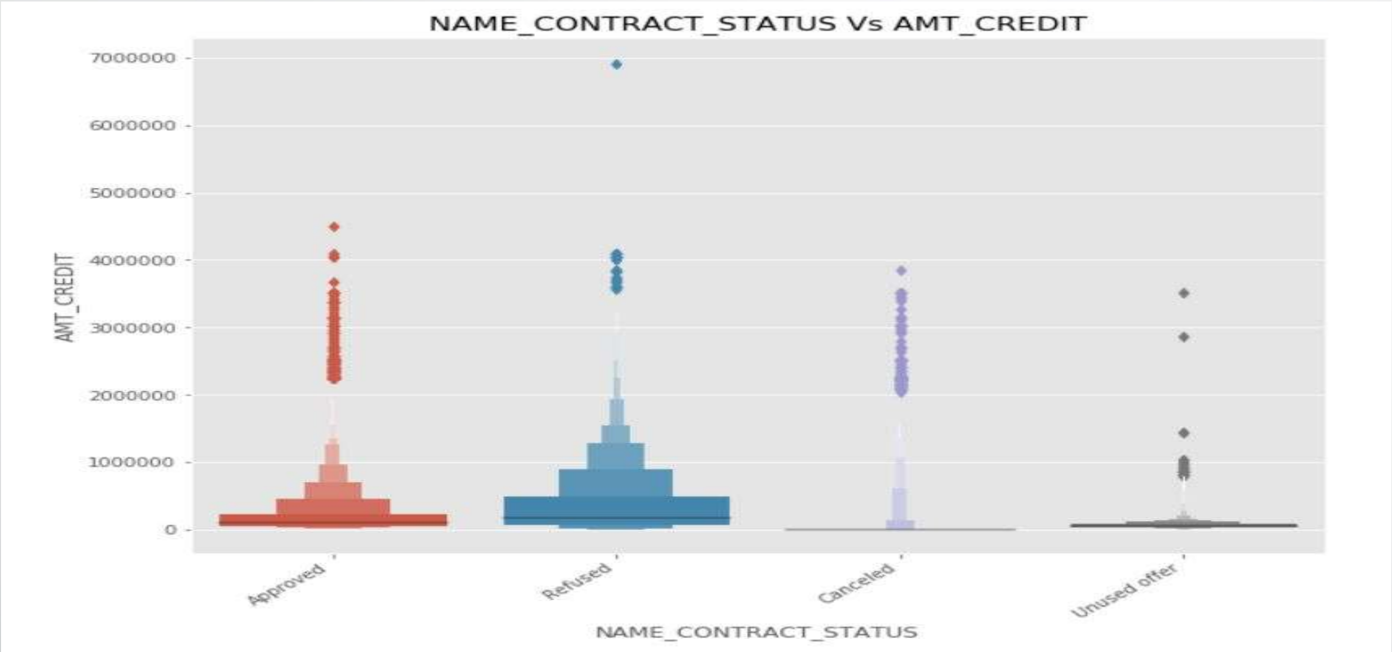
FINAL CREDIT AMOUNT DISBURSED TO THE CUSTOMER PREVIOUSLY, AFTER APPROVAL IS HIGHLY INFLUENCE BY THE APPLICATION AMOUNT AND ALSO THE GOODS PRICE OF GOOD THAT CLIENT ASKED FOR ON THE PREVIOUS APPLICATION.

NAME_CONTRACT_STATUS Vs AMT_ANNUITY



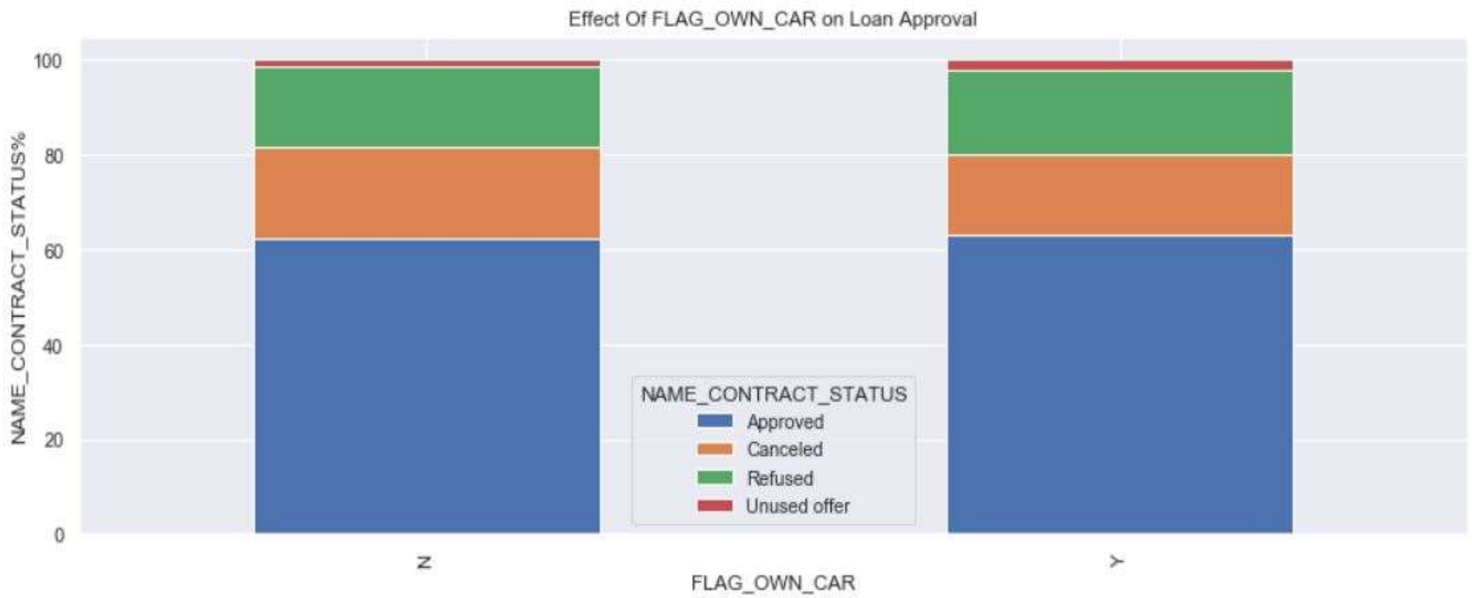
From the above plot we can see that loan application for people with lower AMT_ANNUITY gets cancelled or Unused most of the time. We also see that applications with too high AMT ANNUITY also got refused more often than others.

NAME_CONTRACT_STATUS Vs AMT_CREDIT



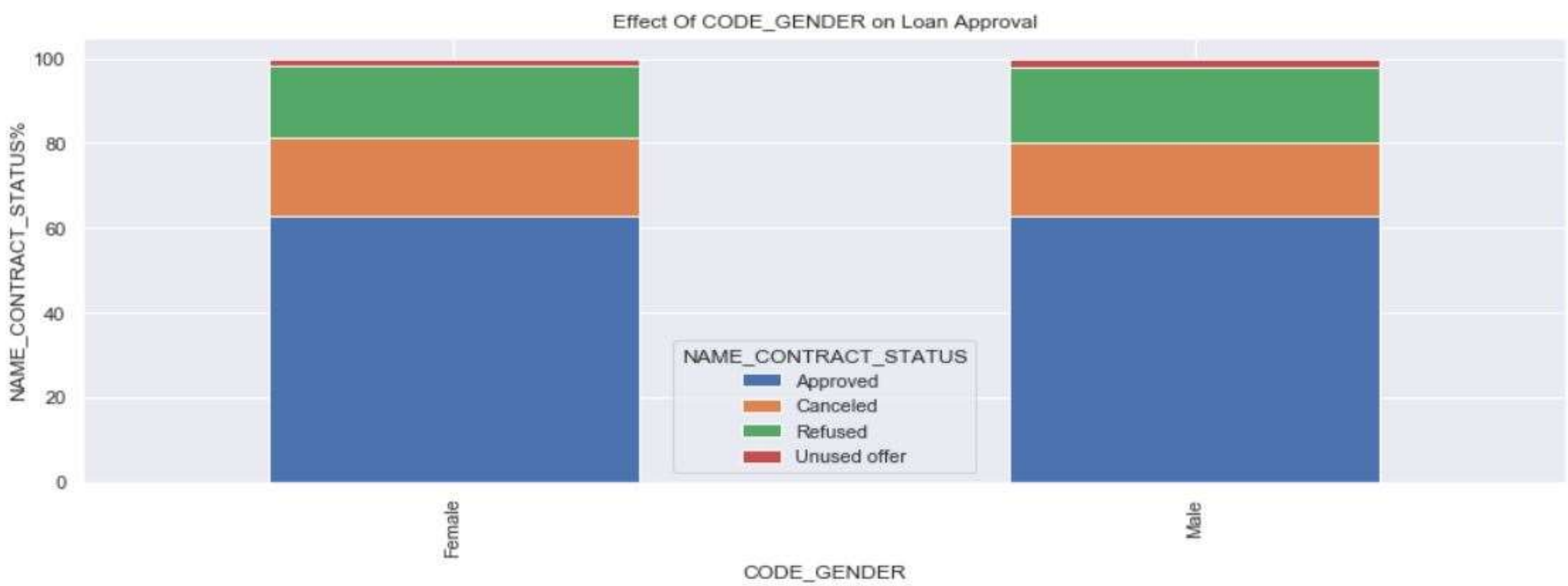
We can infer that when the AMT_CREDIT is too low, it get's cancelled/unused most of the time.

EFFECT OF FLAG_OWN_CAR ON LOAN APPROVAL



We see that car ownership doesn't have any effect on application approval or rejection. But we saw earlier that the people who has a car has lesser chances of default. The bank can add more weightage to car ownership while approving a loan amount

EFFECT OF CODE GENDER ON LOAN APPROVAL



We see that code gender doesn't have any effect on application approval or rejection. But we saw earlier that female have lesser chances of default compared to males. The bank can add more weightage to female while approving a loan amount.

EFFECT OF TARGET ON LOAN APPROVAL

We can see that the people who were approved for a loan earlier, defaulted less often where as people who were refused a loan earlier have higher chances of defaulting.

CONCLUSION

- Banks should focus more on contract type ‘Student’ ,’pensioner’ and ‘Businessman’ with housing ‘type other than ‘Co-op apartment’ for successful payments.
- Banks should focus less on income type ‘Working’ as they are having most number of unsuccessful payments.
- Also with loan purpose ‘Repair’ is having higher number of unsuccessful payments on time.
- Get as much as clients from housing type ‘With parents’ as they are having least number of unsuccessful payments.

THANK YOU