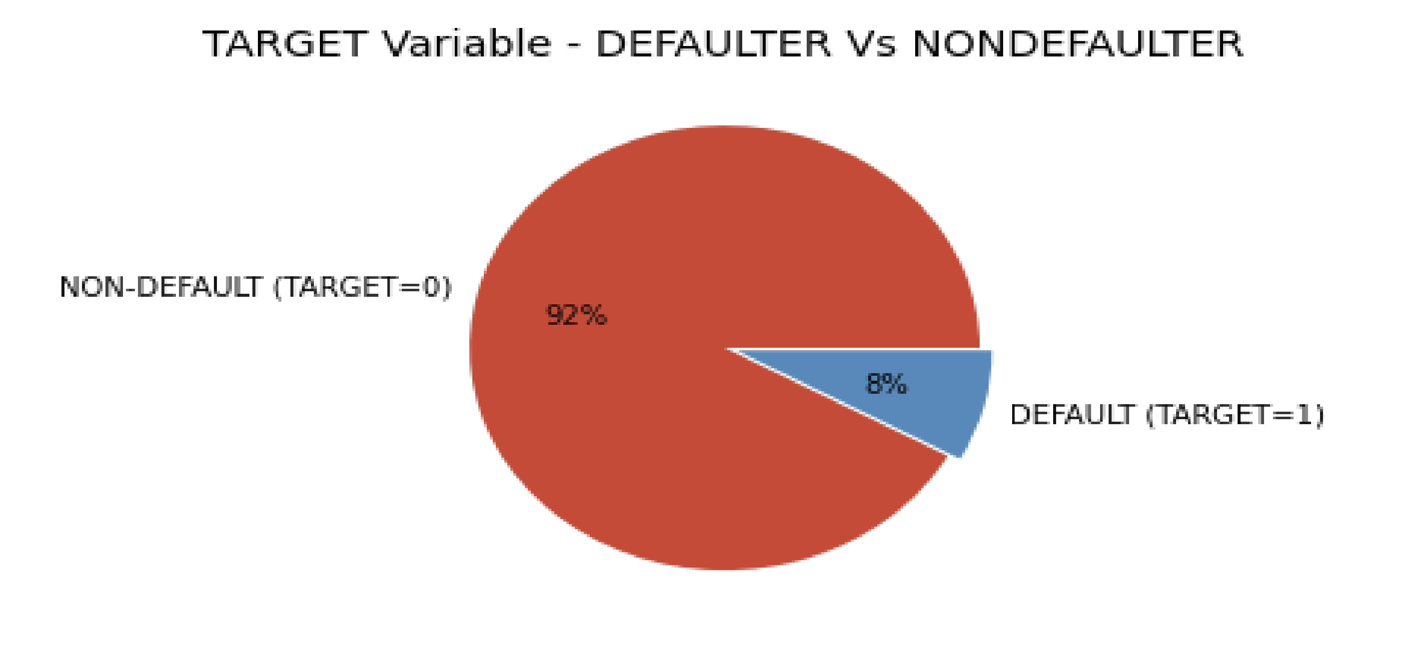
CREDIT EDA CASE STUDY

BY ADITYA KUMAR & TANMAY KURMI

INTRODUCTION

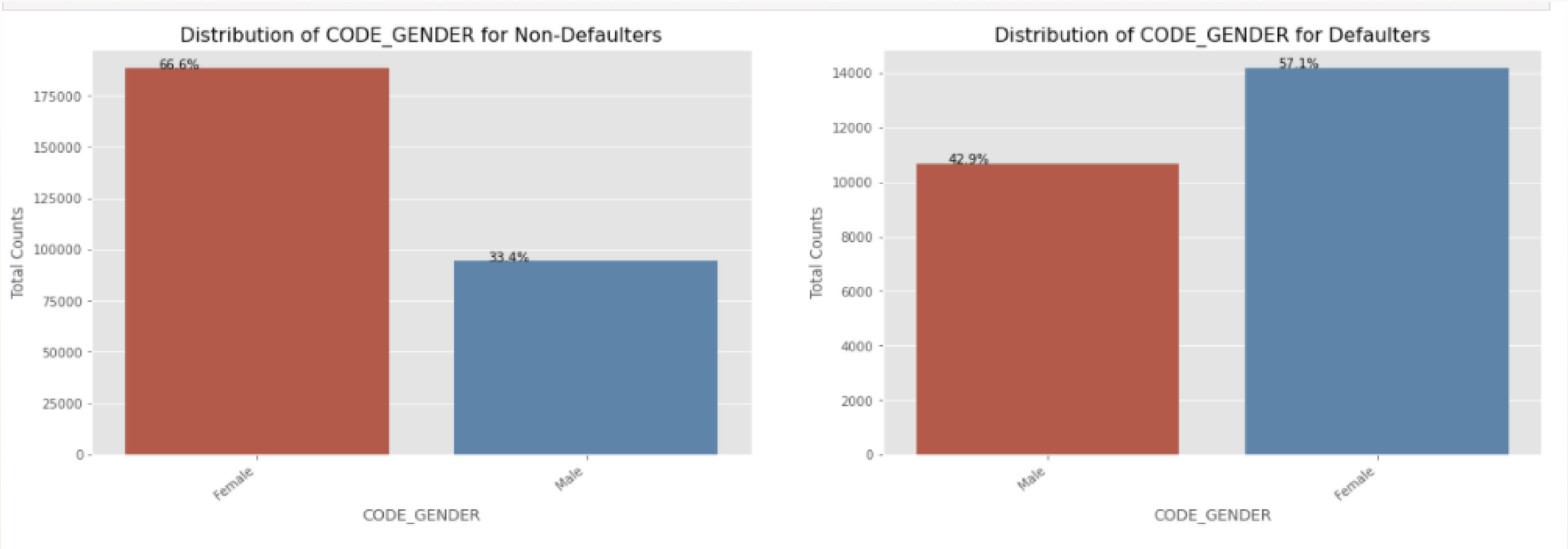
THIS CASE STUDY AIMS TO GIVE YOU AN IDEA OF APPLYING EDA IN A REAL BUSINESS SCENARIO. IN THIS CASE STUDY, APART FROM APPLYING THE TECHNIQUES THAT YOU HAVE LEARNT IN THE EDA MODULE, YOU WILL ALSO DEVELOP A BASIC UNDERSTANDING OF RISK ANALYTICS IN BANKING AND FINANCIAL SERVICES AND UNDERSTAND HOW DATA IS USED TO MINIMIZE THE RISK OF LOSING MONEY WHILE LENDING TO CUSTOMERS.

IMBALANCE IN TARGET

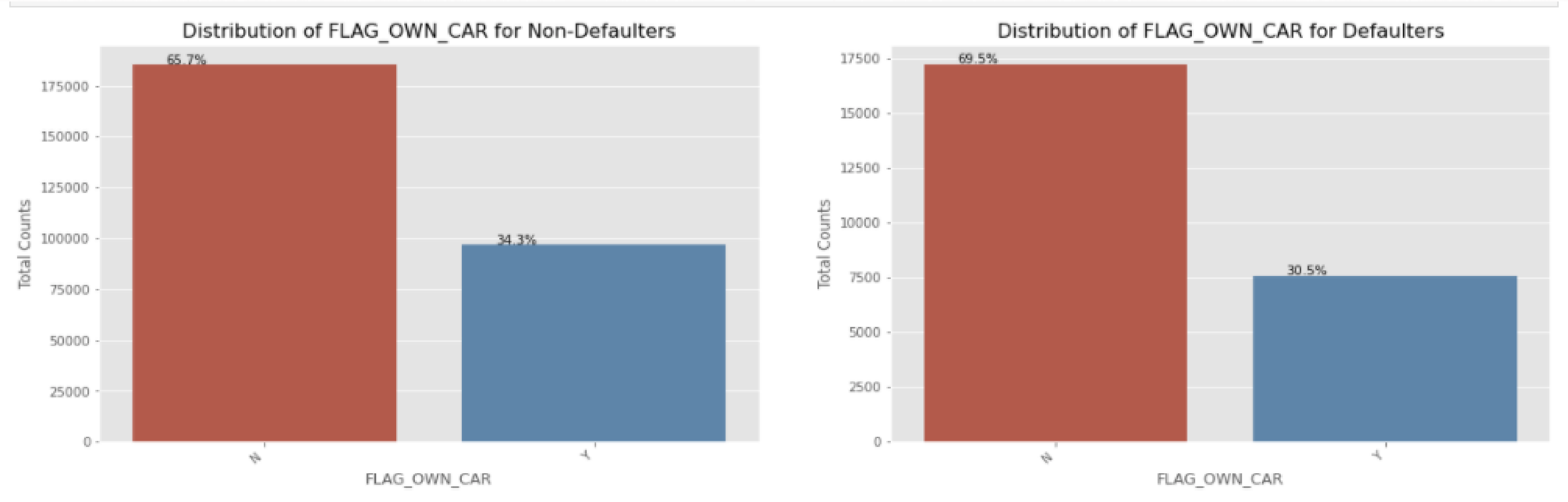


there is an imbalance between people who defaulted and who didn't default. More than 92% of people didn't default as opposed to 8% who defaulted

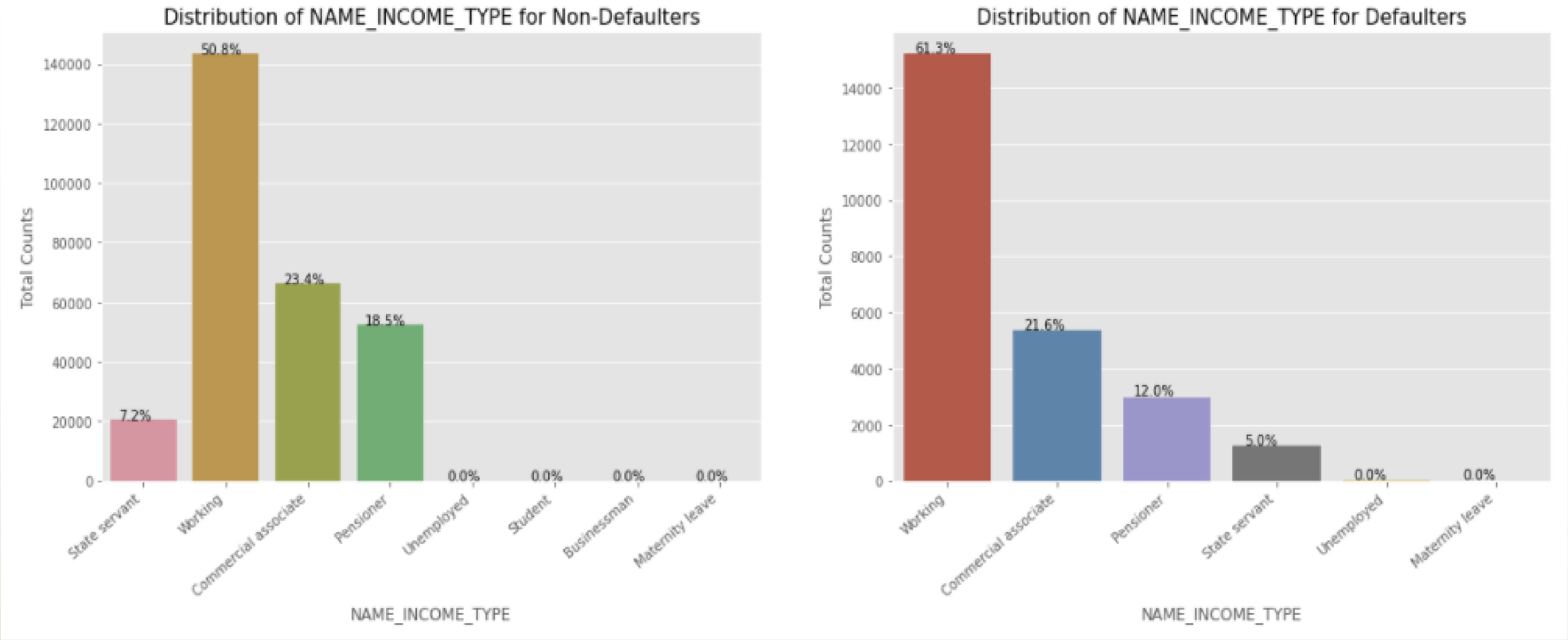
UNIVARIATE CATEGORICAL ORDERED ANALYSIS



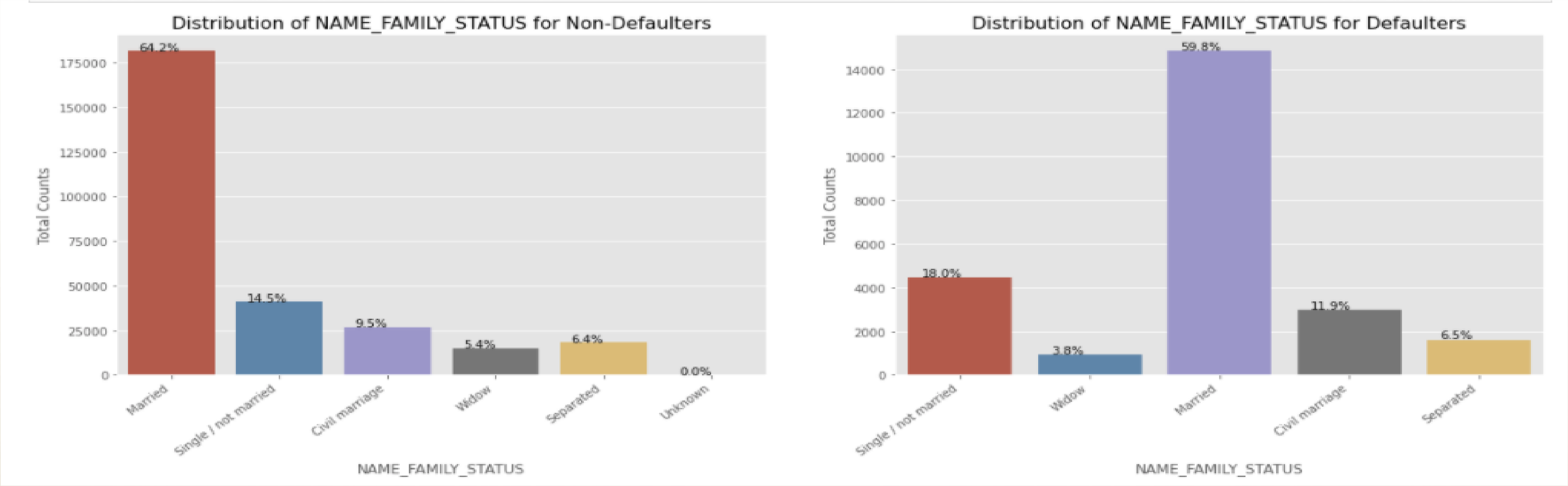
We can see that Female contribute 67% to the non-defaulters while 57% to the defaulters. We can conclude that We see more female applying for loans than males and hence the more number of female defaulters as well



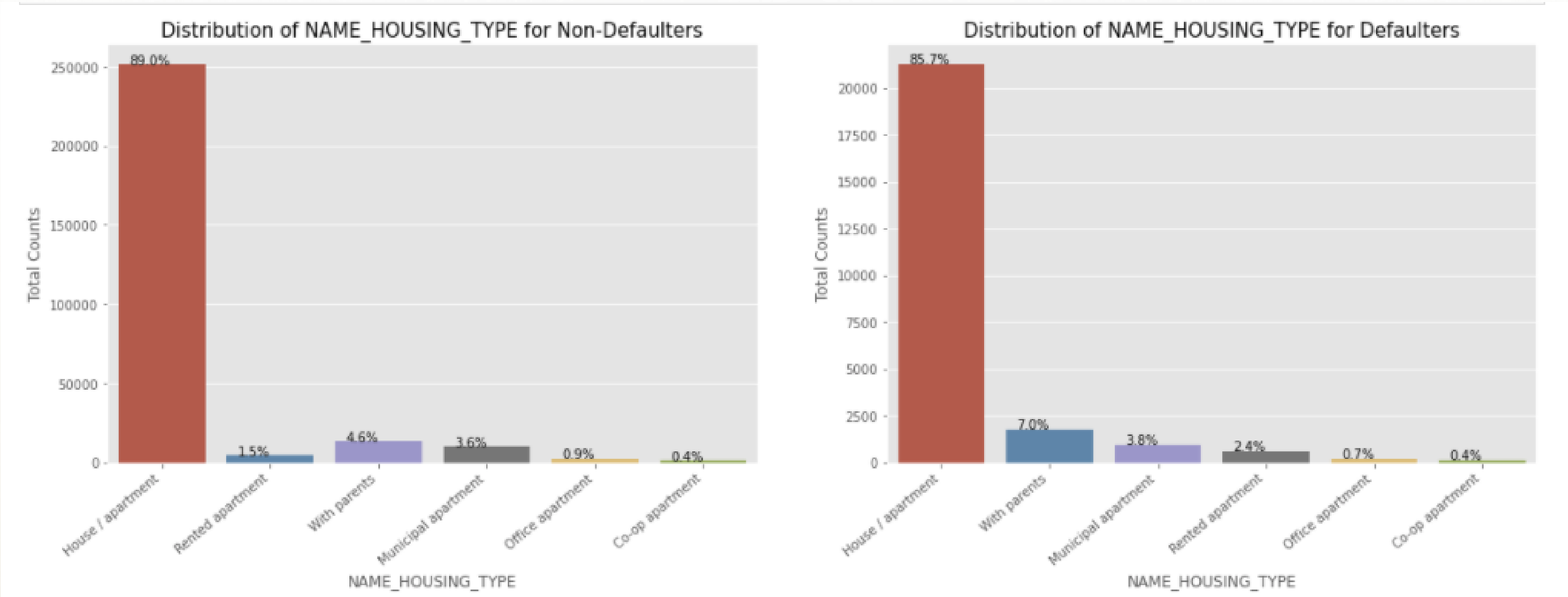
We can see that people with cars contribute 65.7% to the non-defaulters while 69.5% to the defaulters. We can conclude that While people who have car default more often, the reason could be there are simply more people without cars Looking at the percentages in both the charts, we can conclude that the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



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

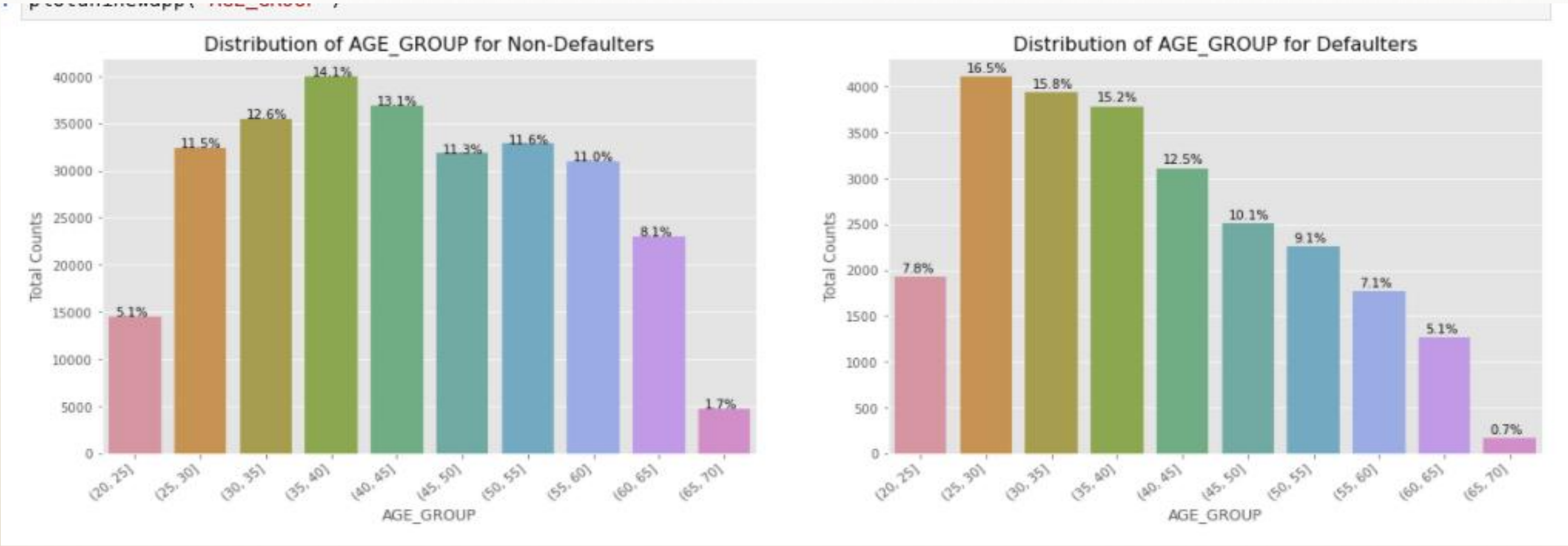


Married people tend to apply for more loans comparatively. But from the graph we see that Single/non Married people contribute 14.5% to Non Defaulters and 18% to the defaulters. So there is more risk associated with them.

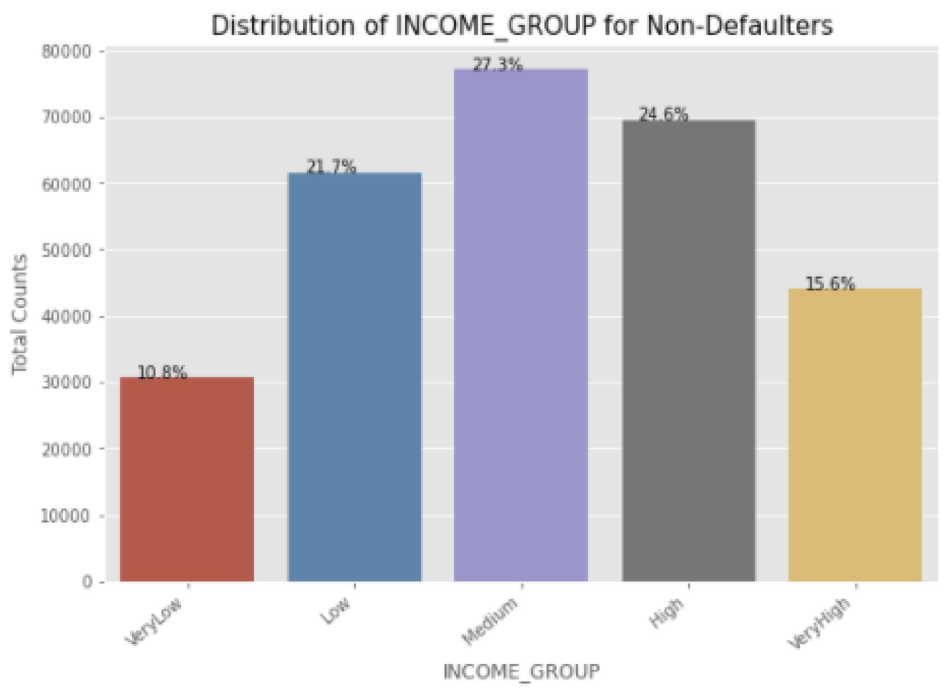


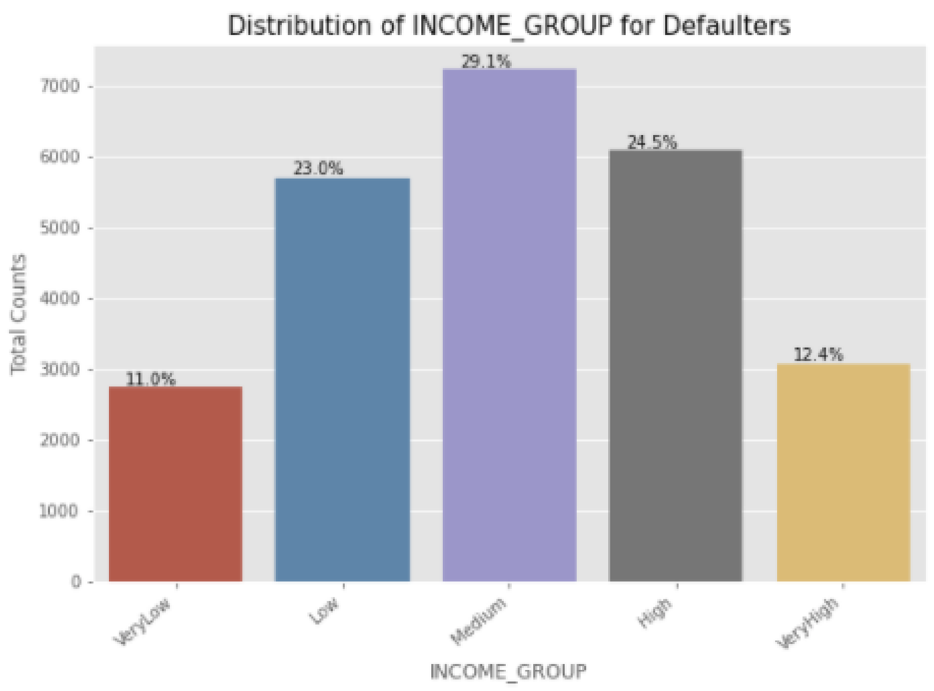
It is clear from the graph that people who have House/Apartments, tend to apply for more loans. People living with parents tend to default more often when compared with others. The reason could be their living expenses are more due to their parents living with them.

UNIVARIATE CATEGORICAL ORDERED ANALYSIS

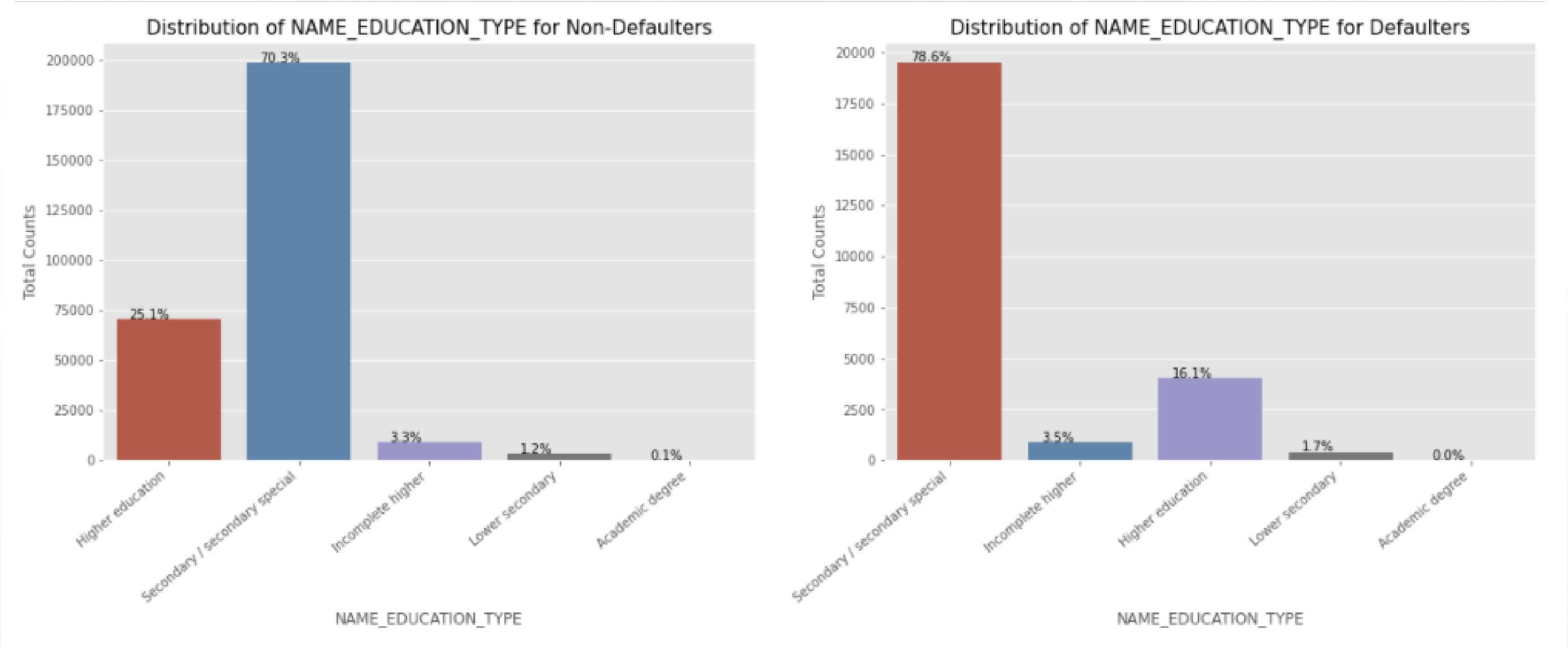


We see that (25,30] age group tend to default more often. So they are the riskiest people to loan to. With increasing age group, people tend to default less starting from the age 25. One of the reasons could be they get employed around that age and with increasing age, their salary also increases.

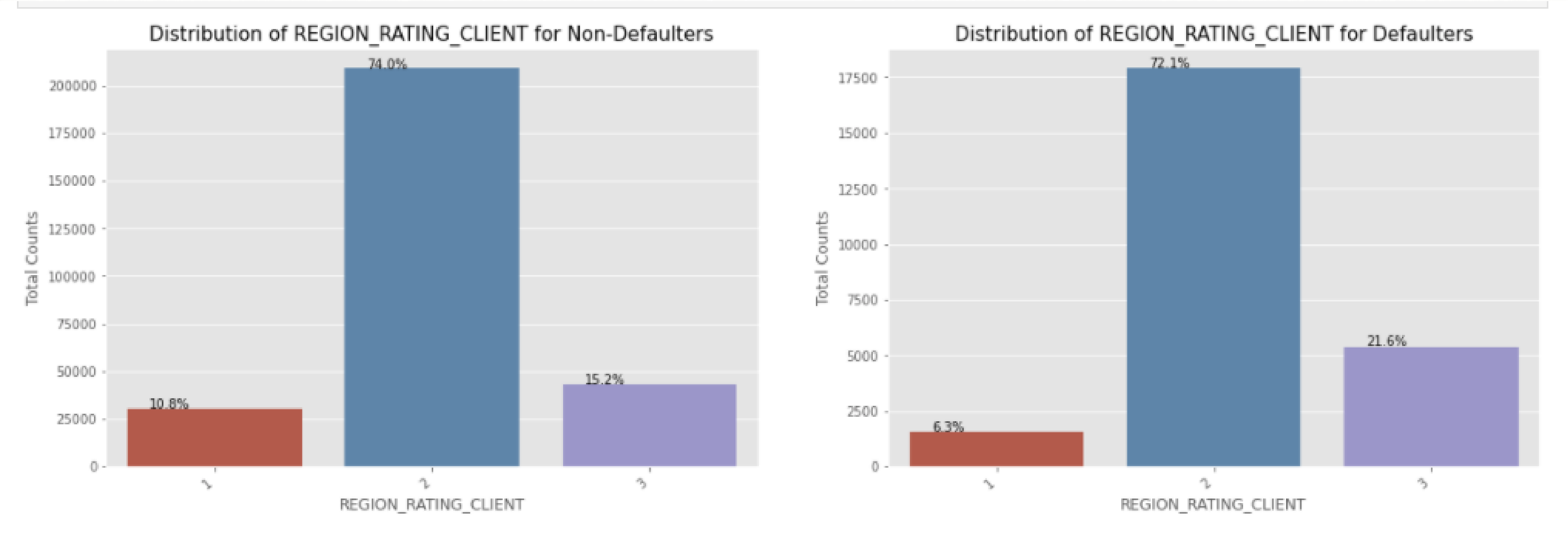




The Very High income group tend to default less often. They contribute 12.4% to the total number of defaulters, while they contribute 15.6% to the Non-Defaulters.

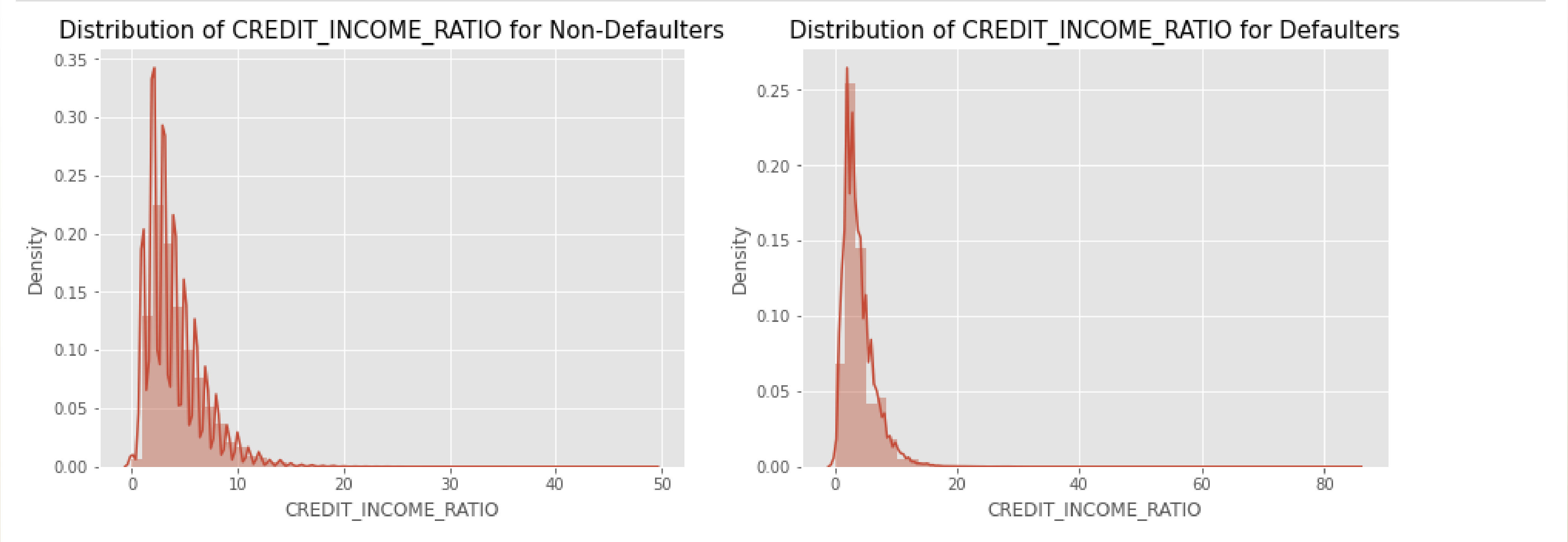


Almost all of the Education categories are equally likely to default except for the higher educated ones who are less likely to default and secondary educated people are more likely to default

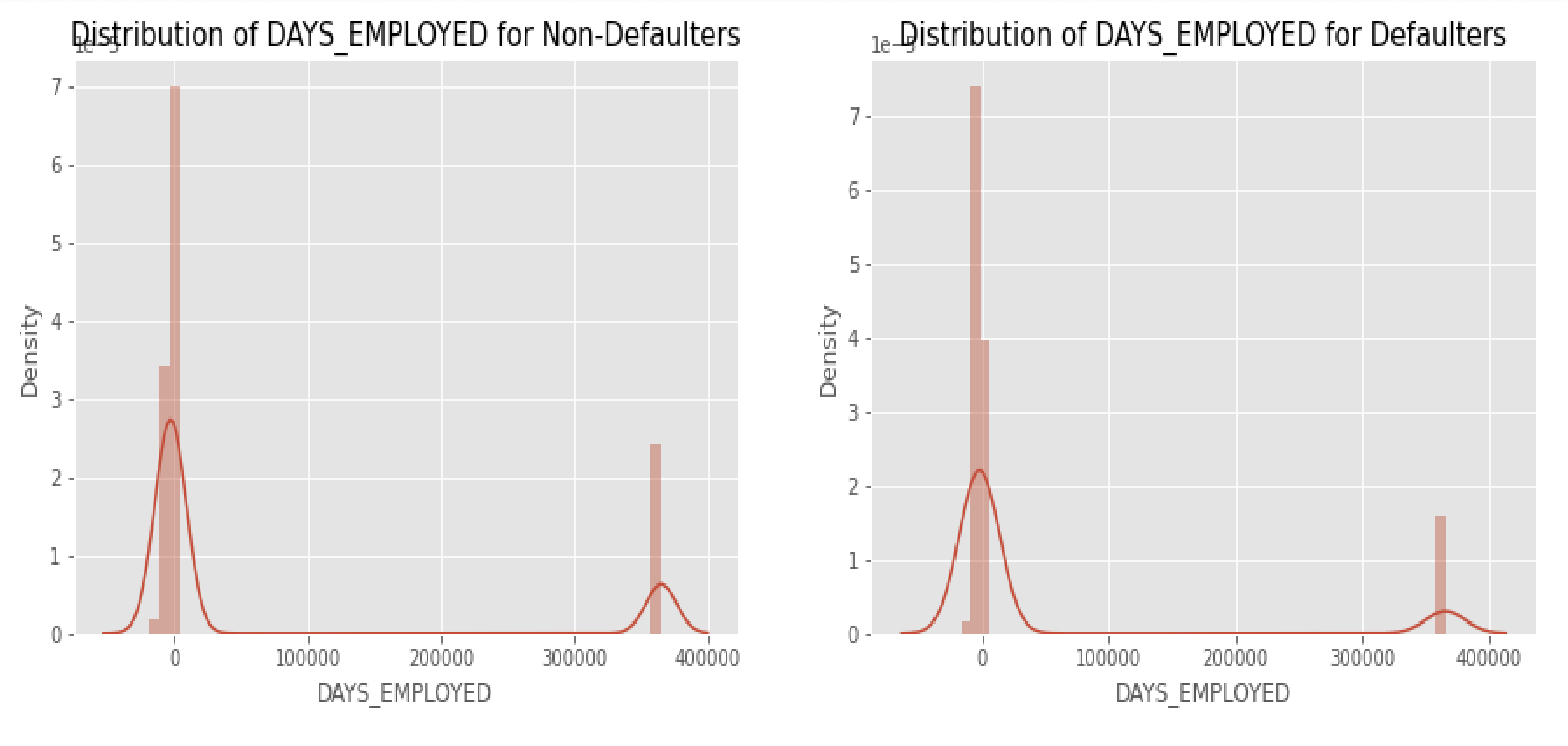


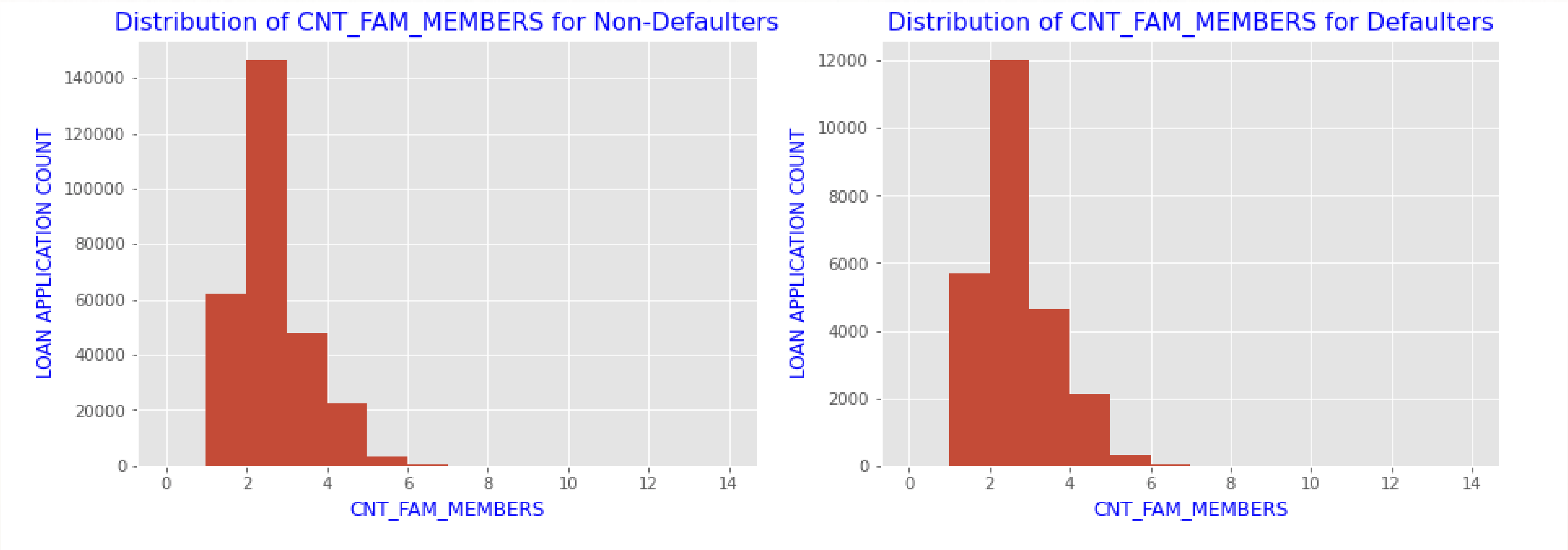
More people from second tier regions tend to apply for loans. We can infer that people living in better areas(Rating 3) tend contribute more to the defaulters by their weightage. People living in 1 rated areas

UNIVARIATE CONTINUOUS VARIABLE ANALYSIS



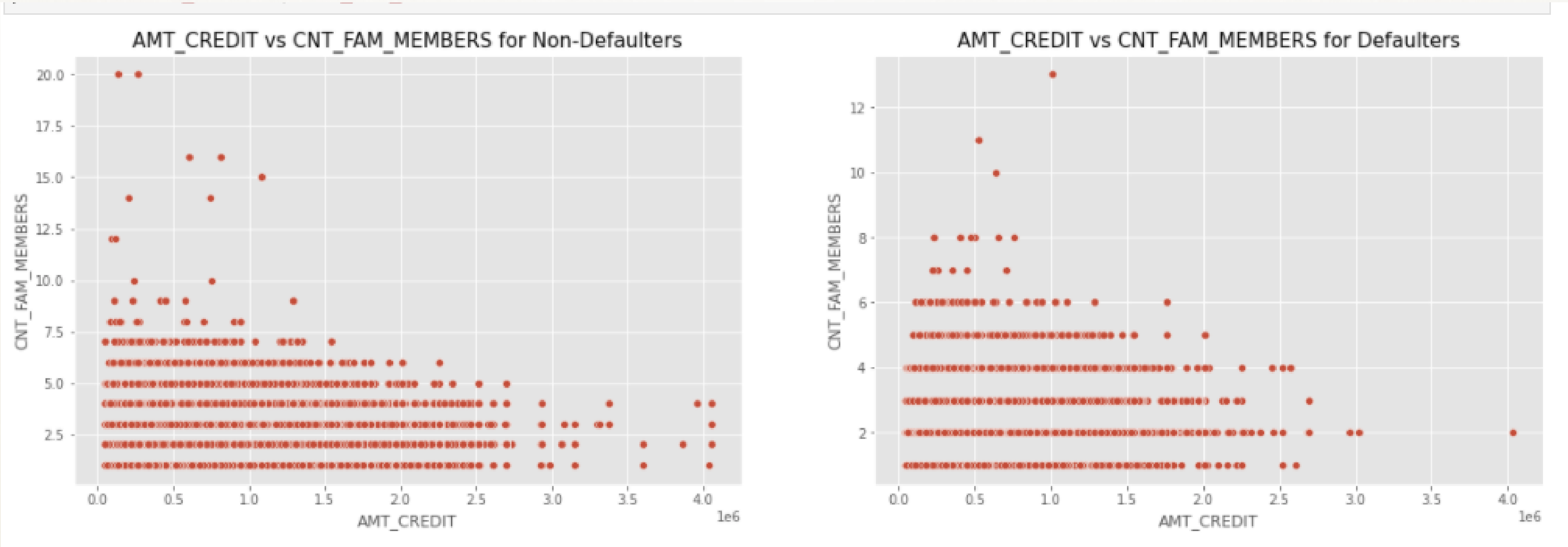
Credit income ratio the ratio of AMT_CREDIT/AMT_INCOME_TOTAL. Although there doesn't seem to be a clear distiguish between the group which defaulted vs the group which didn't when compared using the ratio, we can see that when the CREDIT_INCOME_RATIO is more than 50, people default.


Distribution of DAYS_EMPLOYED for Non-Defaulters
1e−5
Density
DAYS_EMPLOYED
0
100000
200000
300000
400000
Distribution of DAYS_EMPLOYED for Defaulters
1e−5
Density
DAYS_EMPLOYED

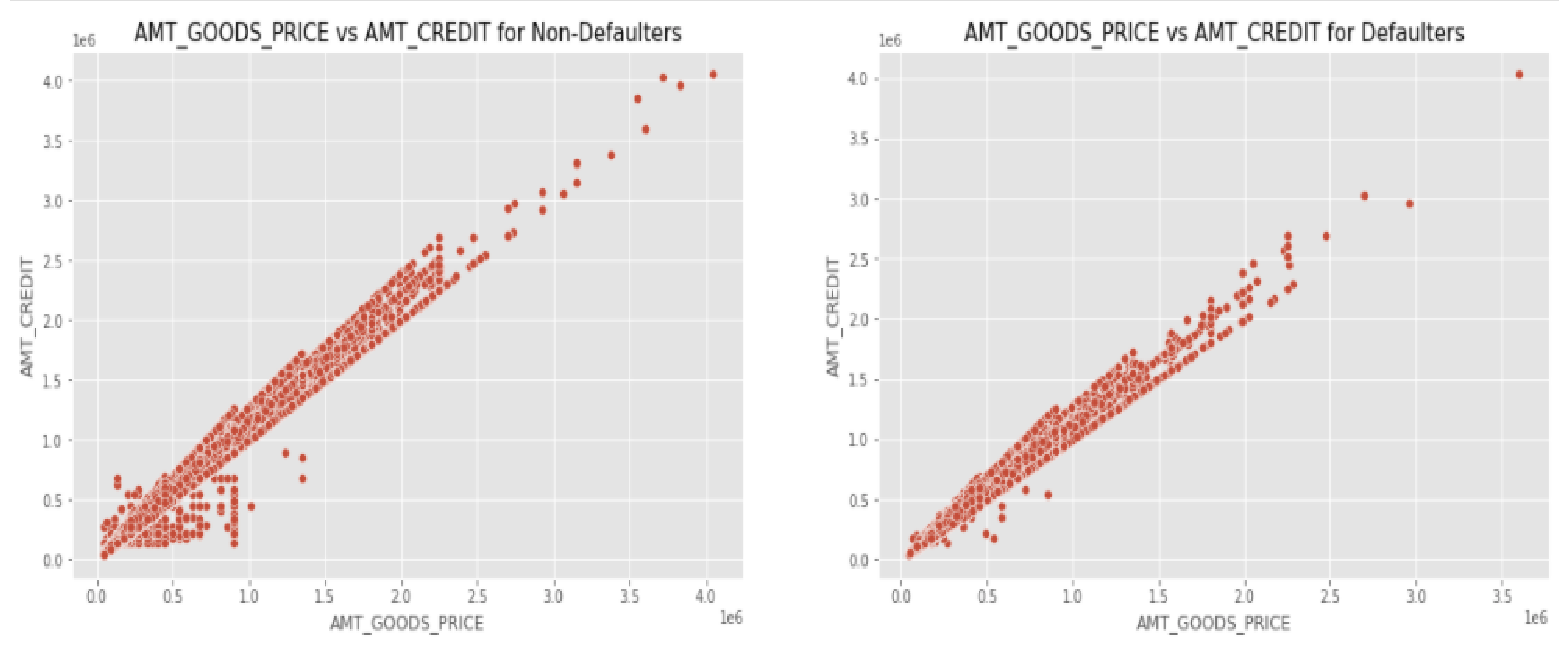


We can see that a family of 3 applies loan more often than the other families

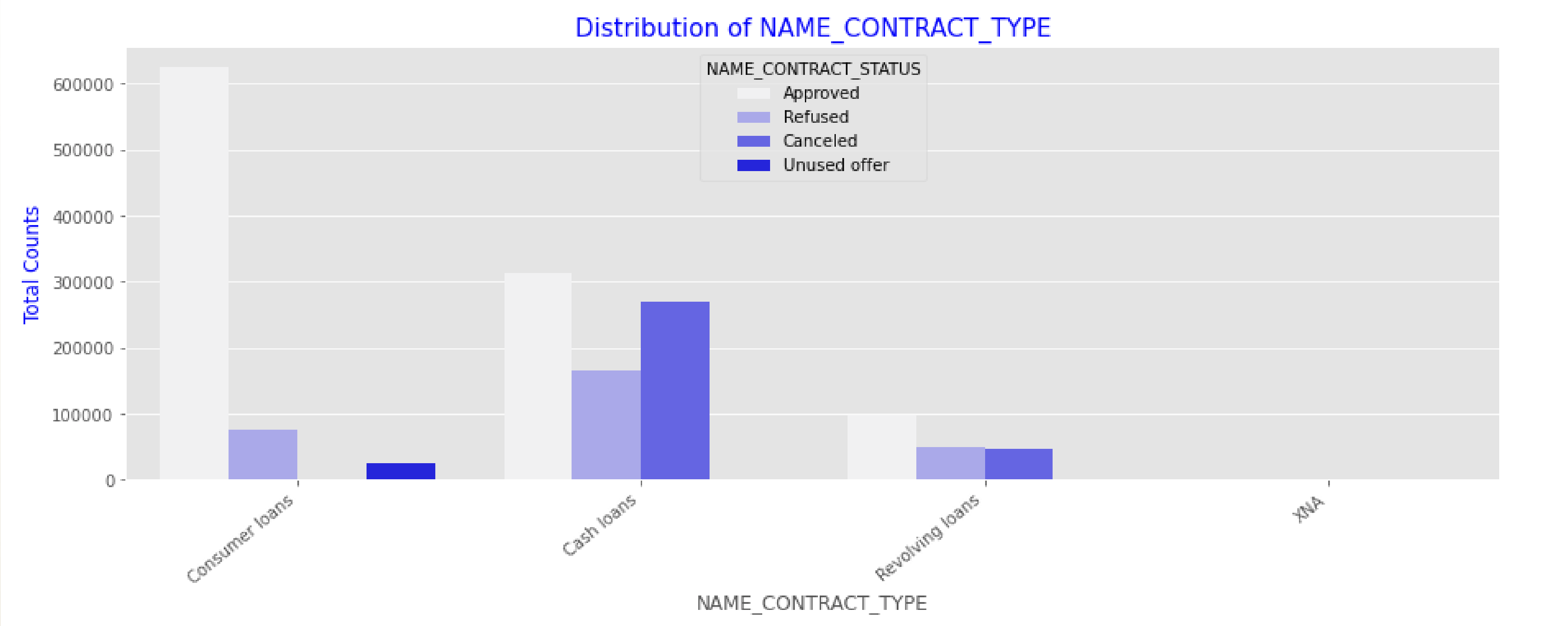
BIVARIATE ANALYSIS OF NUMERICAL VARIABLES



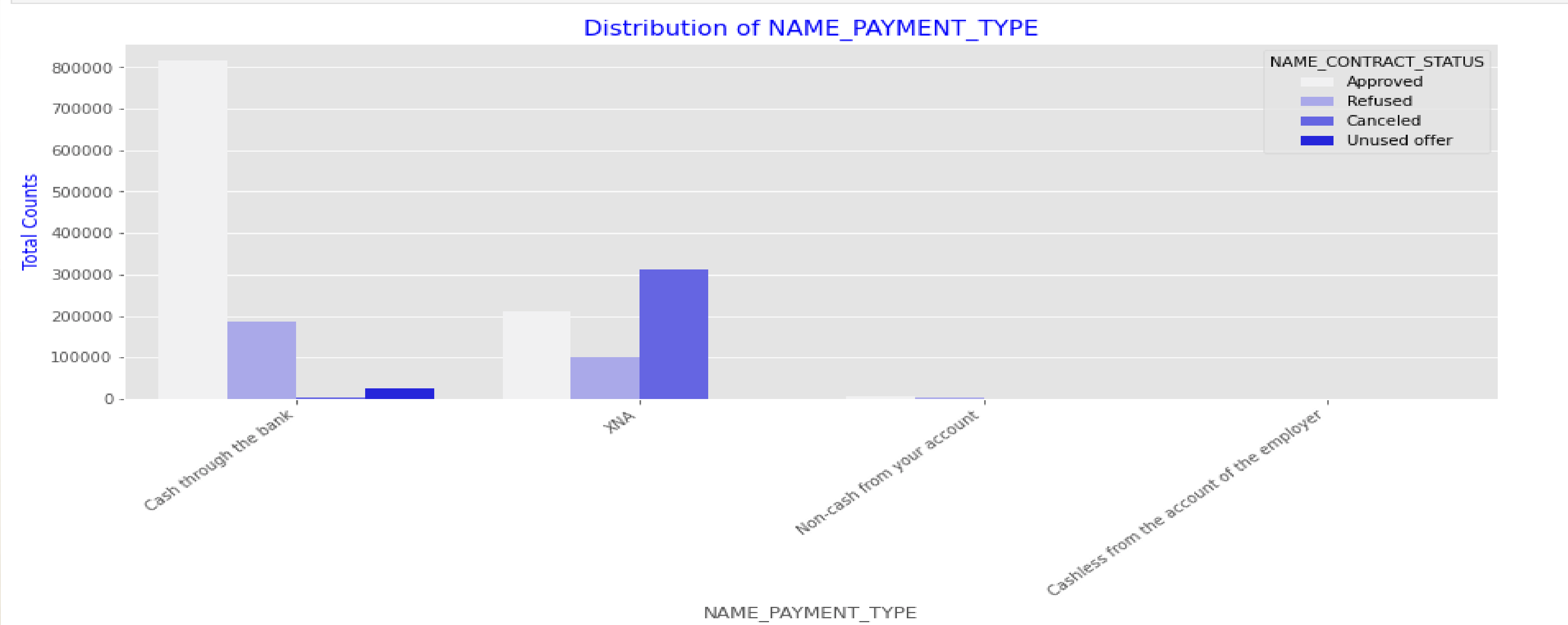
We can see that the density in the lower left corner is similar in both the case, so the people are equally likely to default if the family is small and the AMT_CREDIT is low. We can observe that larger families and people with larger AMT_CREDIT default less often.


AMT_GOODS_PRICE vs AMT_CREDIT for Non-Defaulters
AMT_CREDIT
AMT_GOODS_PRICE
1e6
AMT_GOODS_PRICE vs AMT_CREDIT for Defaulters
AMT_CREDIT
AMT_GOODS_PRICE
1e6

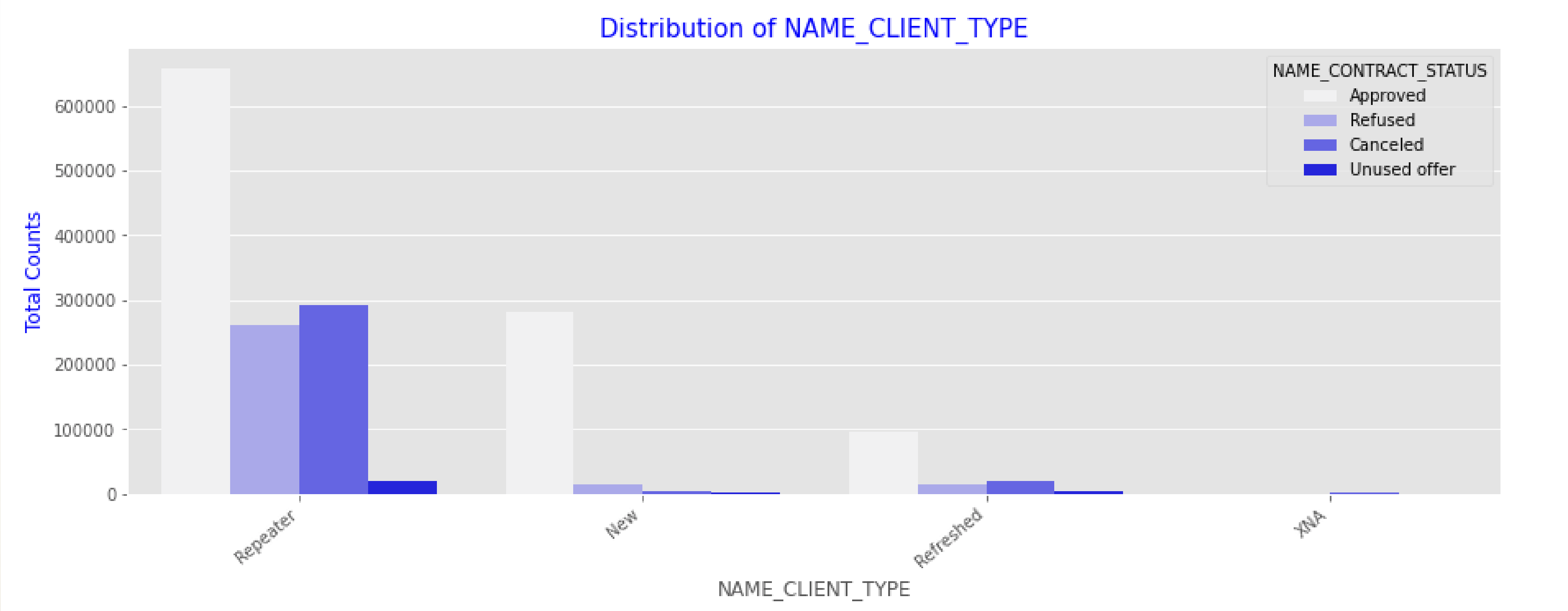
UNIVARIATE ANALYSIS



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.

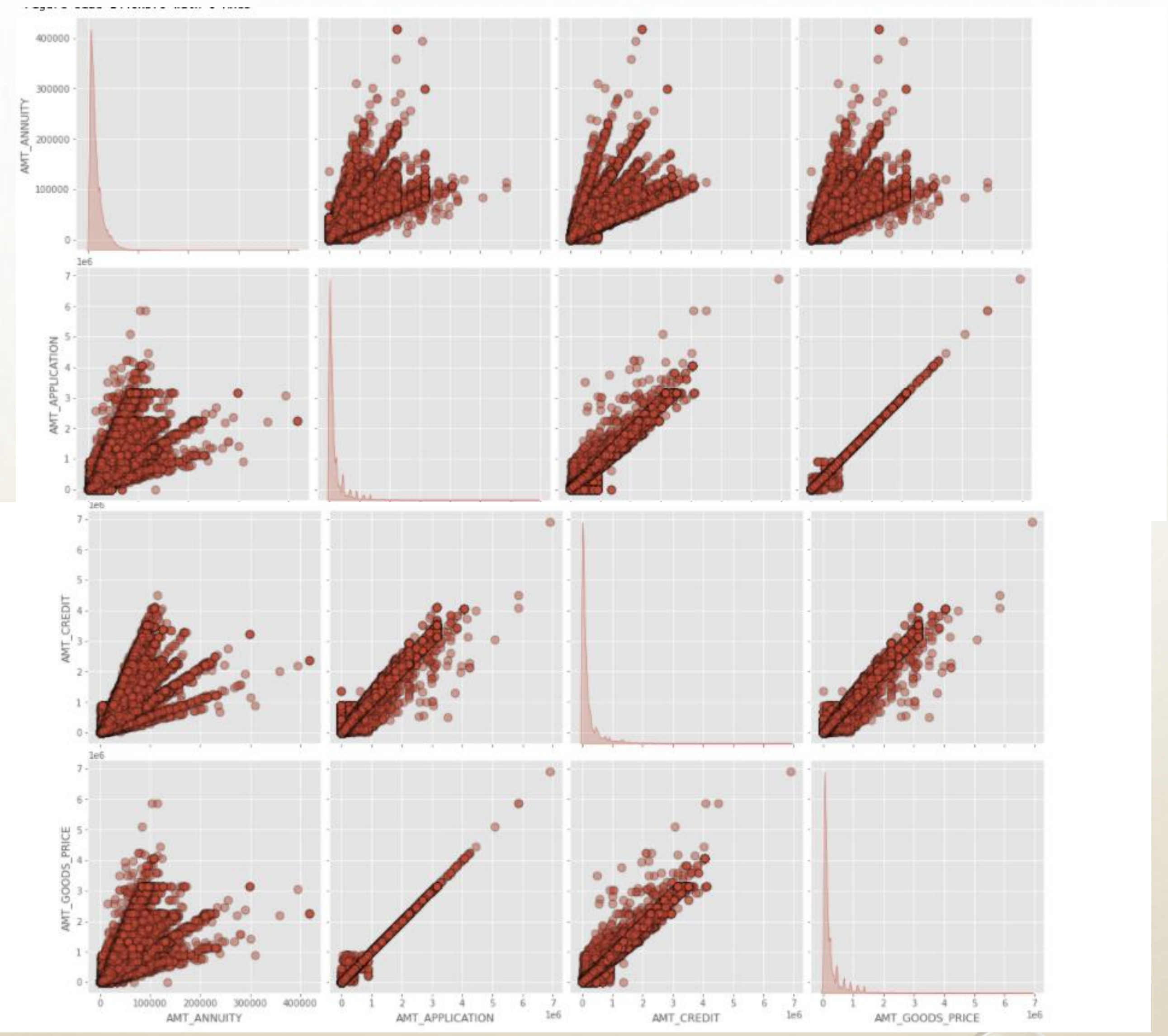


From the above chart, we can infer that most of the clients chose to repay the loan using the 'Cash through the bank' option We can also see that 'Non-Cash from your account' & 'Cashless from the account of the employee' options are not at all popular in terms of loan repayment amongst the customers.

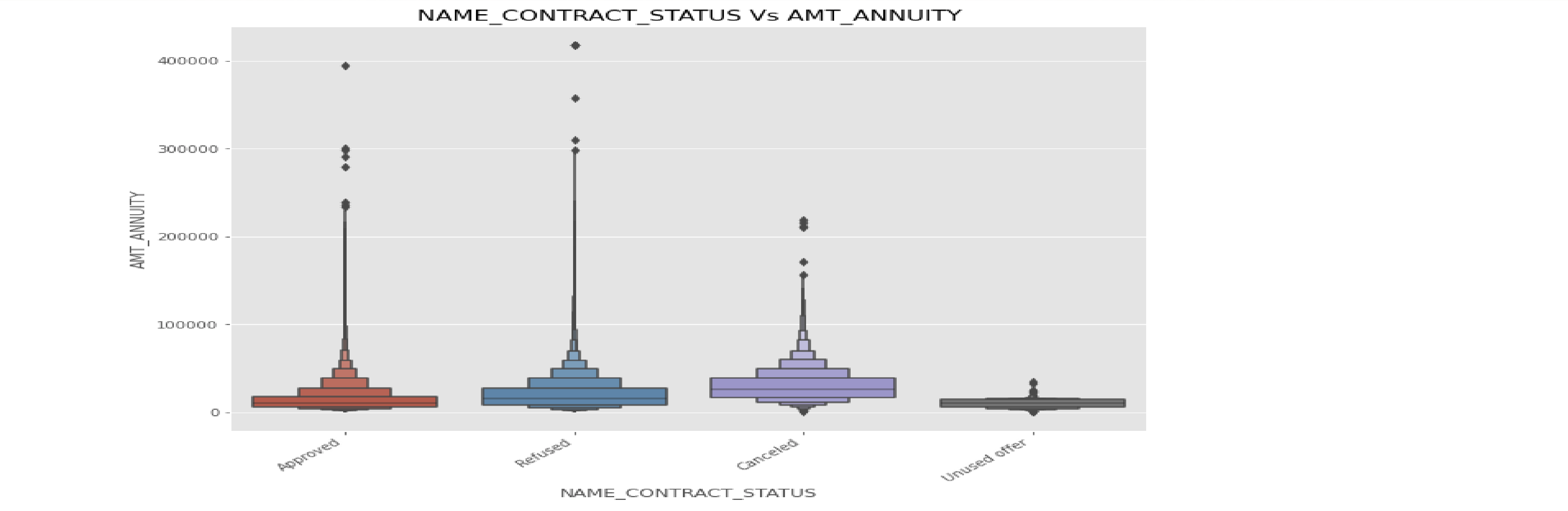


Most of the loan applications are from repeat customers, out of the total applications 70% of customers are repeaters. They also get refused most often.

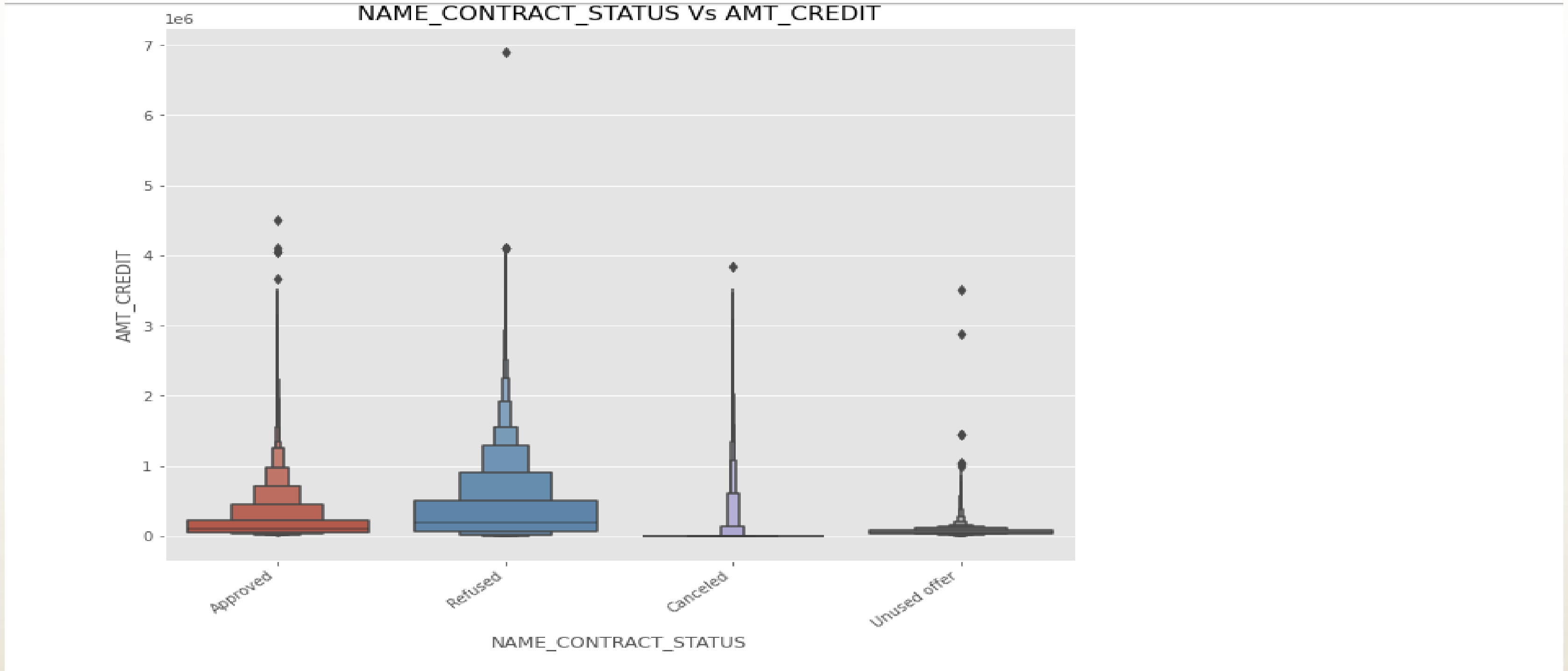
Using Pair plot to Perform Bivariate Analysis on Numerical Columns



USING BOX PLOT TO DO SOME MORE BIVARIATE ANALYSIS ON CATEGORICAL VS NUMERIC COLUMNS

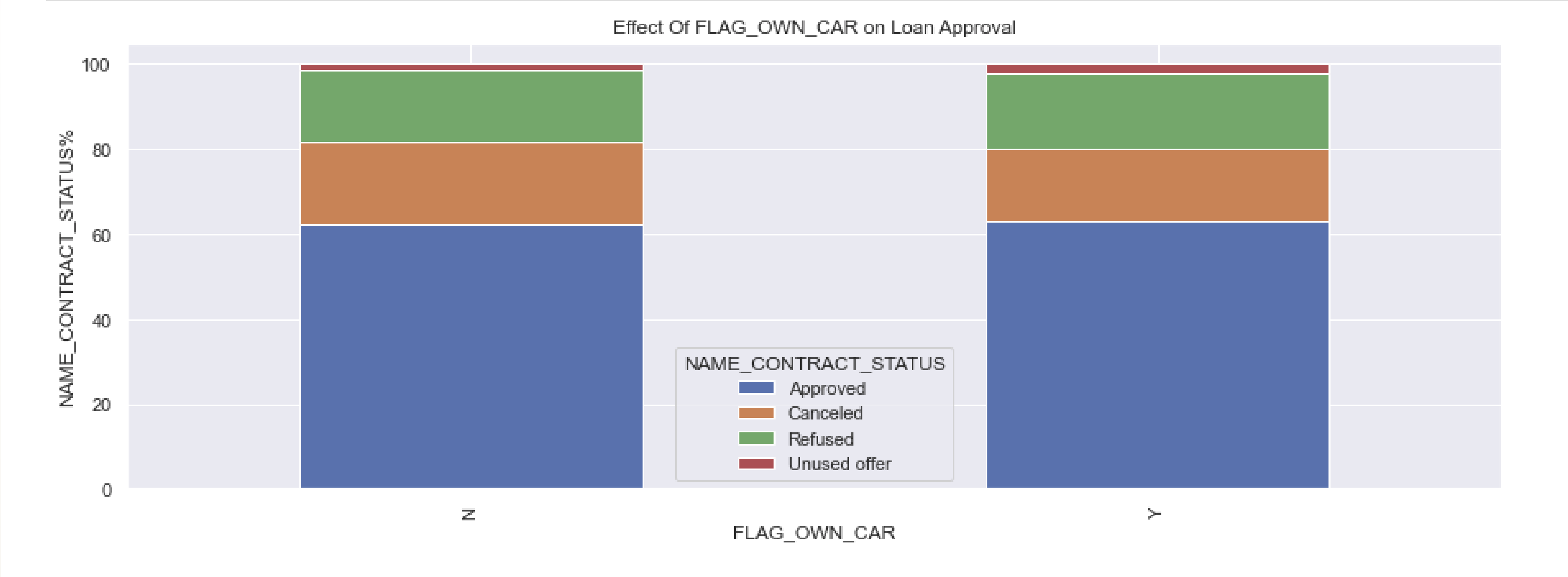


From the above plot we can see that loan application for people with lower AMT_ANNUITY gets cancelled or Unused most of the time. We also see that applications with too high AMT ANNUITY also got refused more often than others.

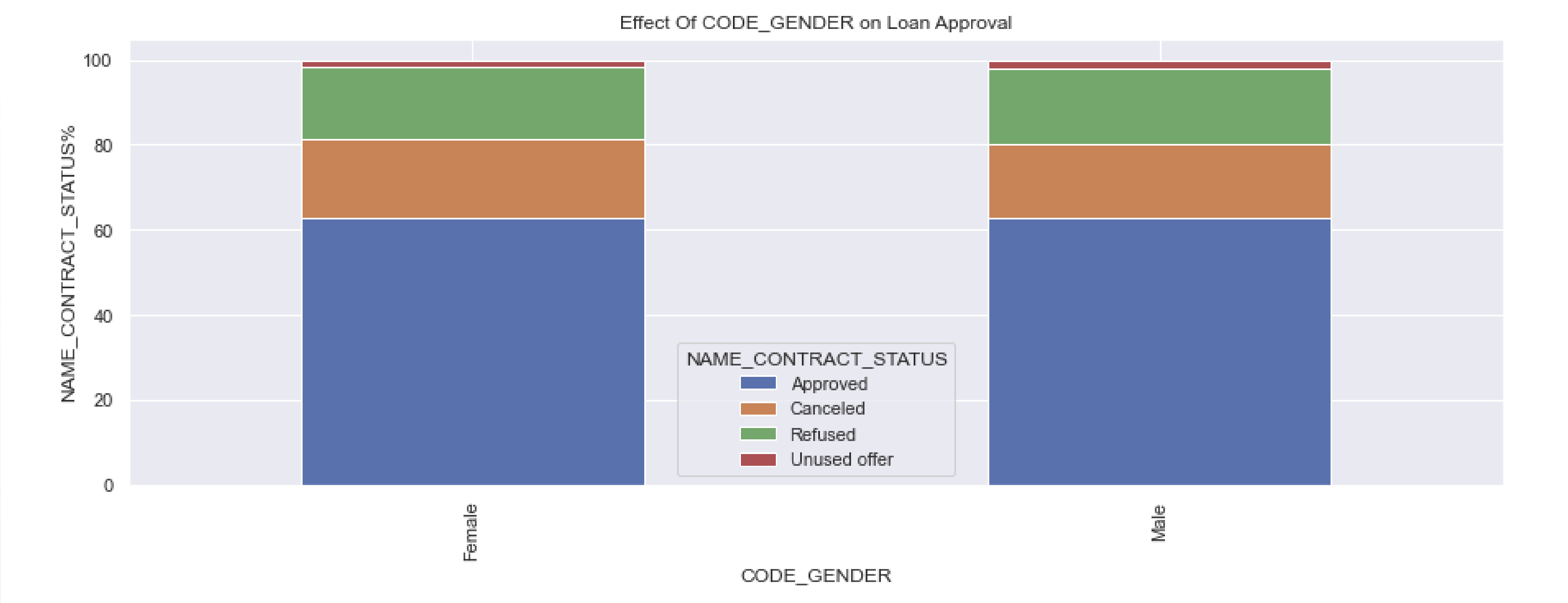


We can infer that when the AMT_CREDIT is too low, it get's cancelled/unused most of the time.

MERGING THE FILES AND ANALYZING THE DATA



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



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.

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.

THANK YOU