Exploratory Data Analysis (Bank Loan Default)

By : Arpit Rakesh Kothiya

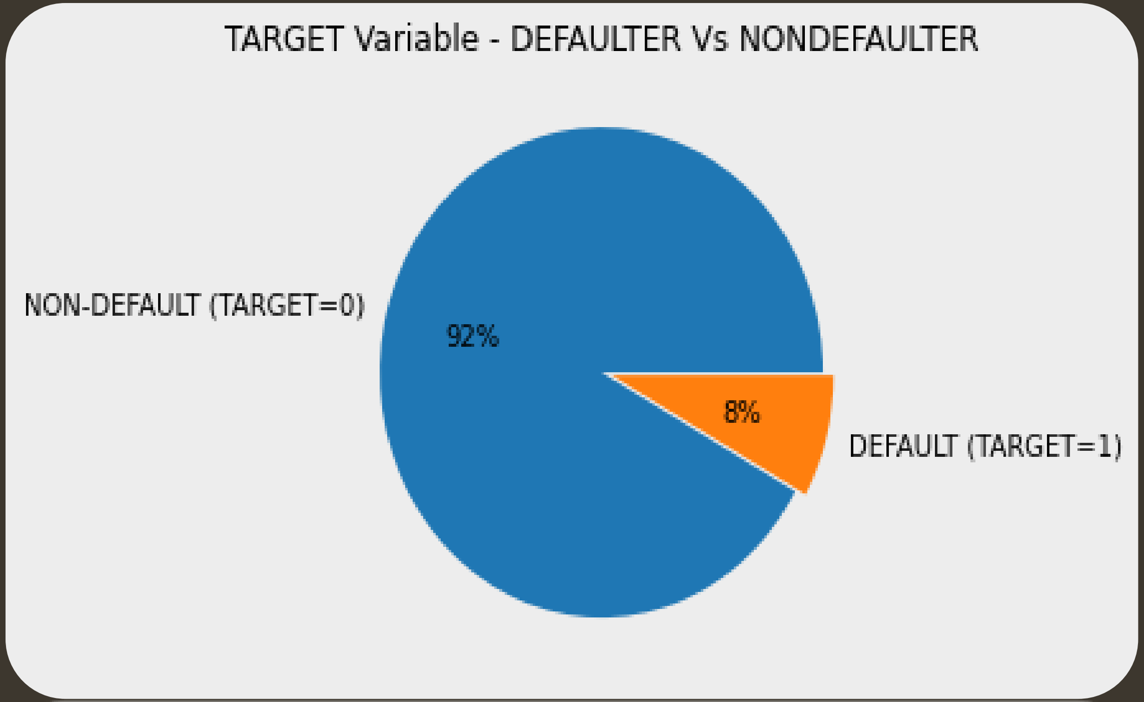
Content

- Percentage of Total Default
- Education Type
- Income Type
- Own Car
- Family Status
- Age Group
- Gender
- Income Group
- Distribution of credit income ratio
- Distribution of Days Employed
- Distribution of Number of Family Members
- Credit Vs Family Member
- Client , Payment , Contract Type
- Bivariate
- Effect of Target on Loan Approval
- Effect of Gender on Loan Approval
- Effect of Car Owner on Loan Approval

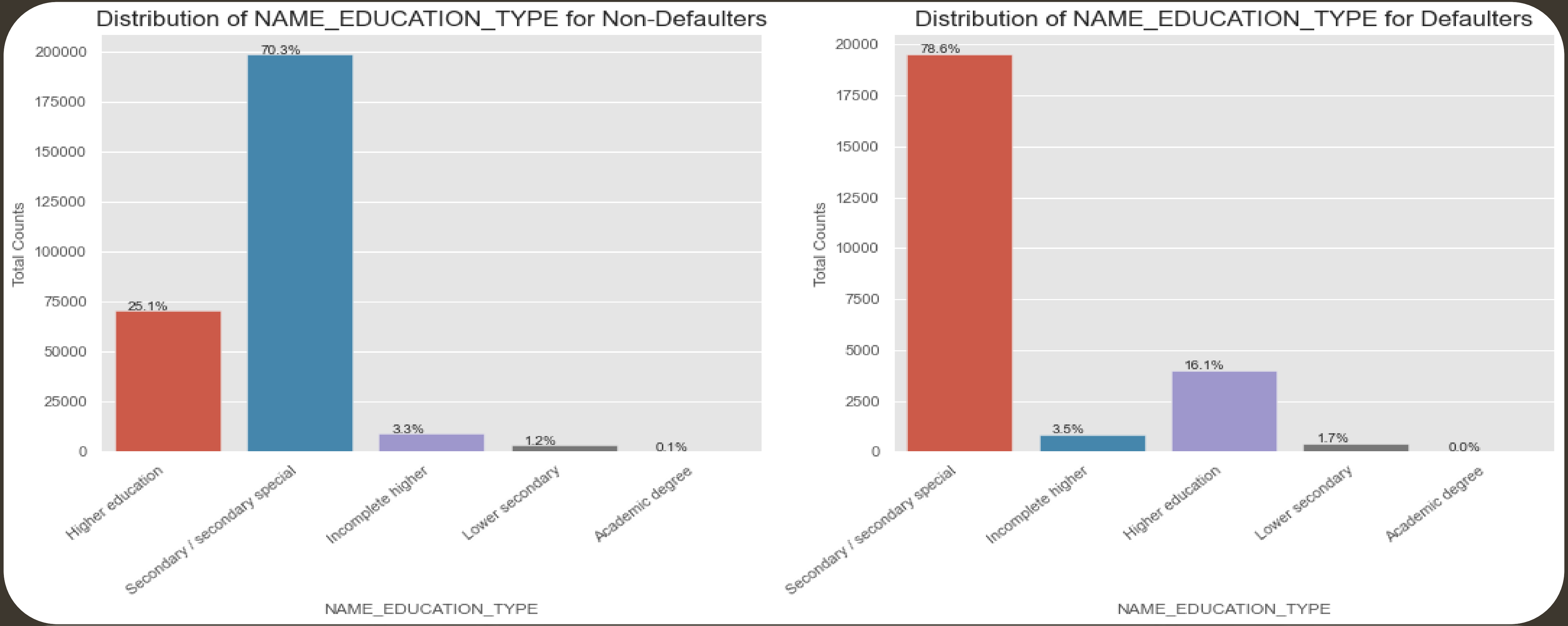
Percentage of Total Default

- Default case only 8%.
- Non- Default case 92%

As it’s clear out of total cases , only 8% are default case , rest 92% are non-default.

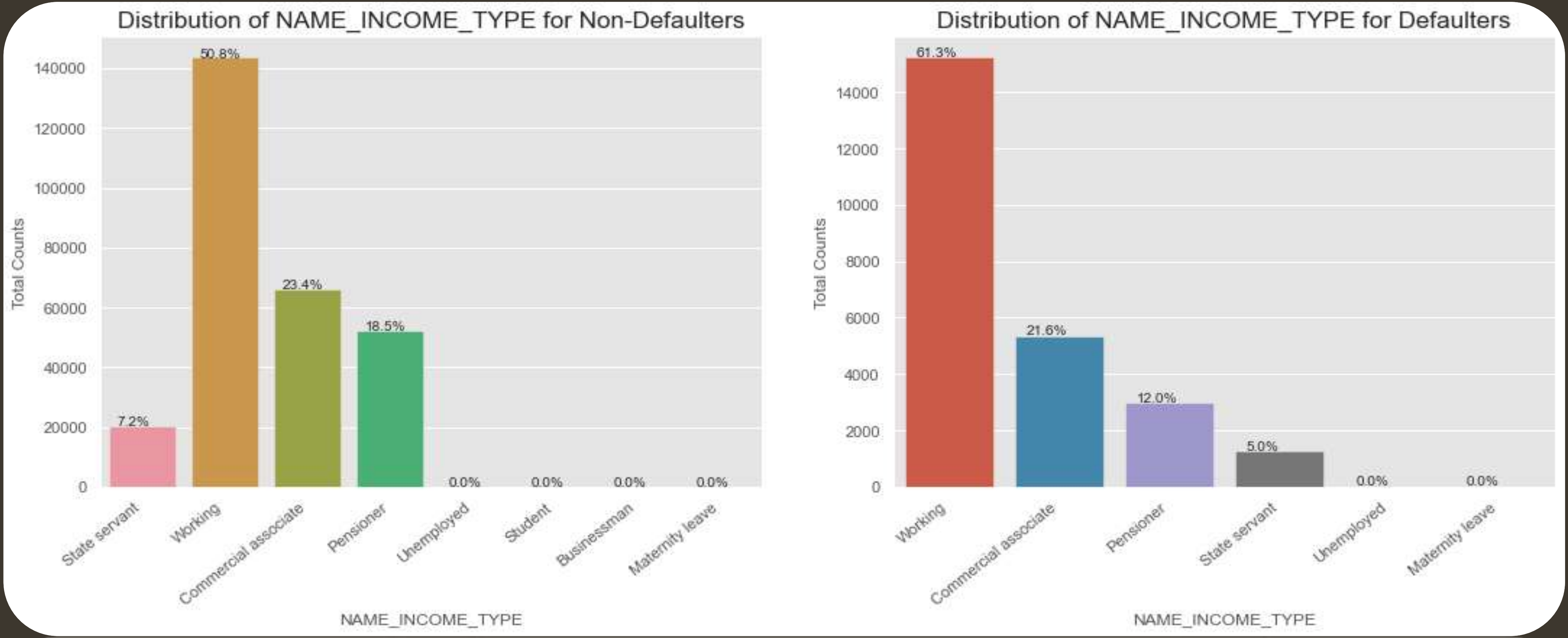


Education Type



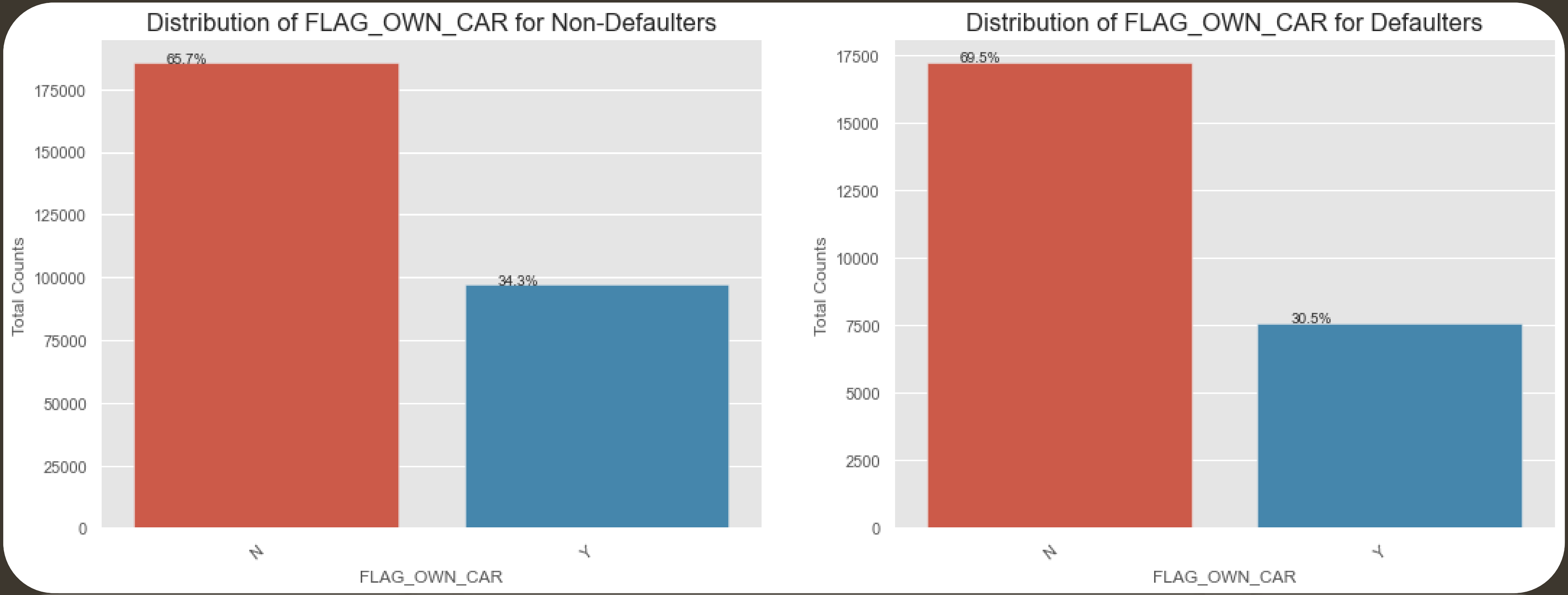
- Almost all of the Education categories are equally likely to default except for the higher educated ones who are less likely to default and secondary educated people are more likely to default
- Secondary educated people also applies the most application for loan.
- Academic degree holder negligibly defaults.
- Secondary / secondary special have 78% default case.
- Higher education have 16% default case

Income Type


Distribution of NAME_INCOME_TYPE for Non-Defaulters
Total Counts
NAME_INCOME_TYPE
State servant
Working
Commercial associate
Pensioner
Unemployed
Student
Businessman
Maternity leave
7.2%
50.8%
23.4%
18.5%
0.0%
0.0%
0.0%
0.0%
0
20000
40000
60000
80000
100000
120000
140000
Distribution of NAME_INCOME_TYPE for Defaulters
Total Counts
NAME_INCOME_TYPE
Working
Commercial associate
Pensioner
State servant
Unemployed
Maternity leave
61.3%
21.6%
12.0%
5.0%
0.0%
0.0%
0
2000
4000
6000
8000
10000
12000
14000

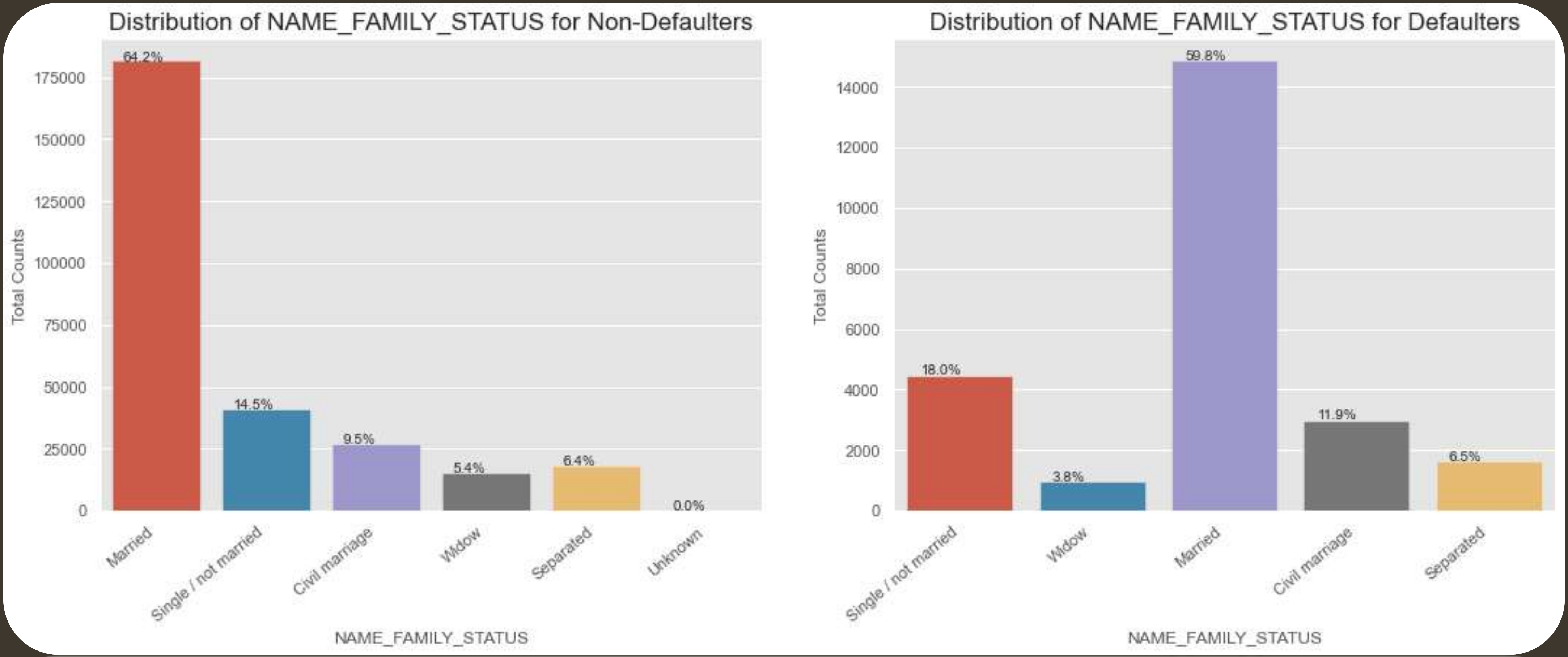
- We can notice that the students don't default. The reason could be they are not required to pay during the time they are students.
- We can also see that the Businessmen never default.
- Most of the loans are distributed to working class people.
- We also see that working class people contribute 51%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

Own Car


Distribution of FLAG_OWN_CAR for Non-Defaulters
Total Counts
175000
150000
125000
100000
75000
50000
25000
0
65.7%
34.3%
N
Y
FLAG_OWN_CAR
Distribution of FLAG_OWN_CAR for Defaulters
Total Counts
17500
15000
12500
10000
7500
5000
2500
0
69.5%
30.5%
N
Y
FLAG_OWN_CAR

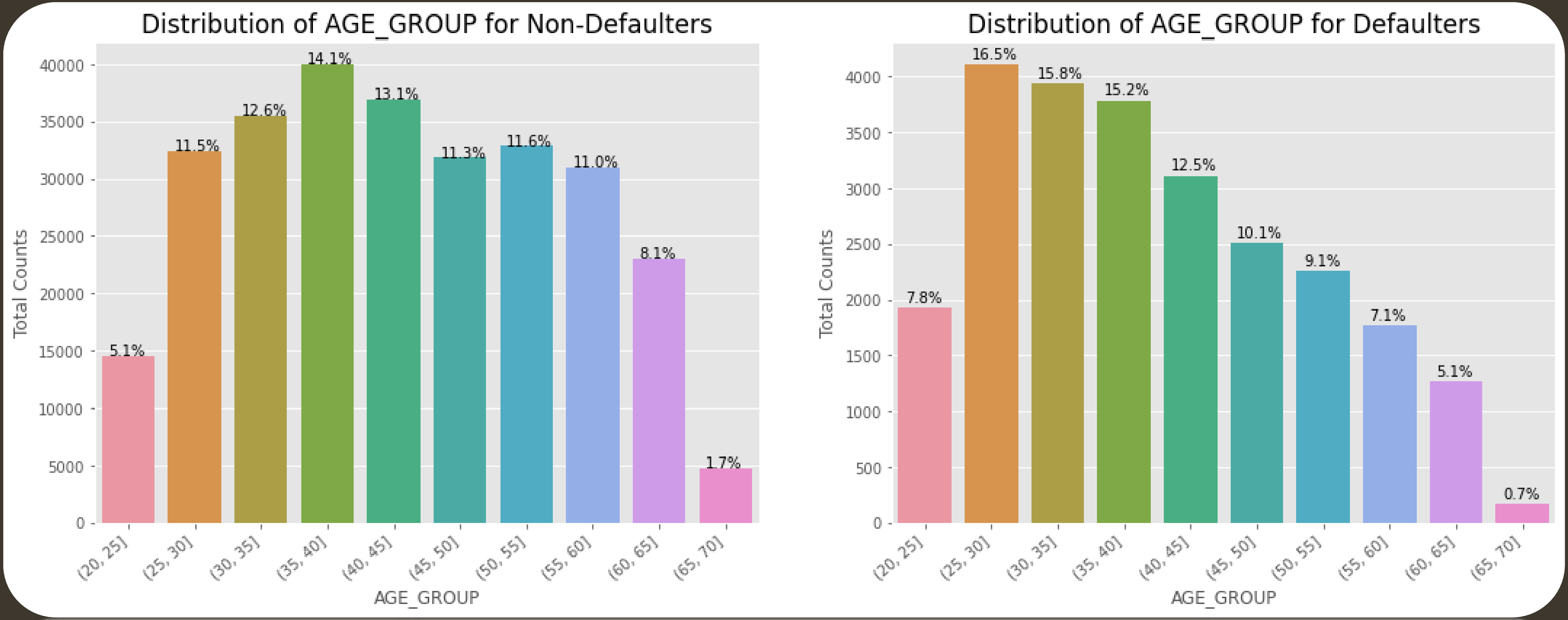
- We can see that people with cars contribute 65.7% to the non-defaulters while 69.5% to the defaulters.
- While people who have car default more often, the reason could be there are simply more people without cars.
- Looking at the percentages in both the charts, we can conclude that the rate of default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

Family Status



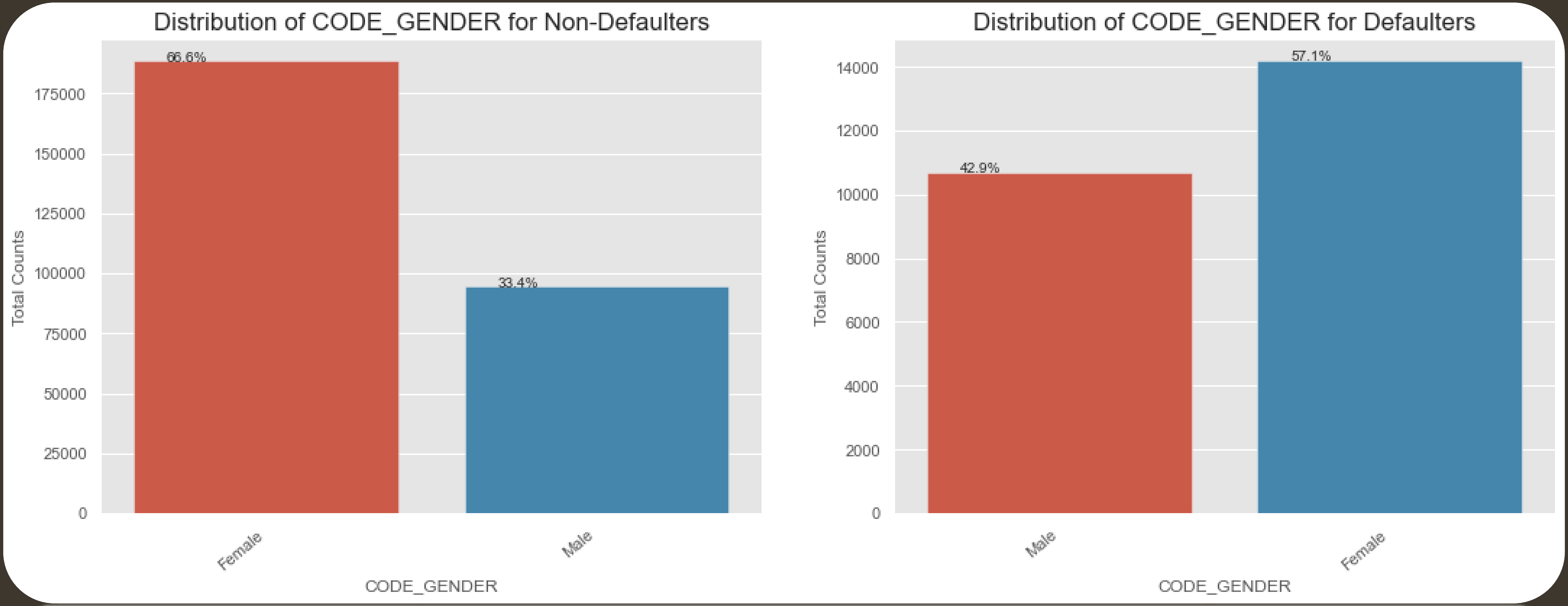
- Married people tend to apply for more loans comparatively.
- But from the graph we see that Single/non Married people contribute 14.5% to Non Defaulters and 18% to the defaulters. So there is more risk associated with them.

Age group


Distribution of AGE_GROUP for Non-Defaulters
Total Counts
40000
35000
30000
25000
20000
15000
10000
5000
0
5.1%
11.5%
12.6%
14.1%
13.1%
11.3%
11.6%
11.0%
8.1%
1.7%
(20, 25]
(25, 30]
(30, 35]
(35, 40]
(40, 45]
(45, 50]
(50, 55]
(55, 60]
(60, 65]
(65, 70]
AGE_GROUP
Distribution of AGE_GROUP for Defaulters
Total Counts
4000
3500
3000
2500
2000
1500
1000
500
0
7.8%
16.5%
15.8%
15.2%
12.5%
10.1%
9.1%
7.1%
5.1%
0.7%
(20, 25]
(25, 30]
(30, 35]
(35, 40]
(40, 45]
(45, 50]
(50, 55]
(55, 60]
(60, 65]
(65, 70]
AGE_GROUP

- 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.

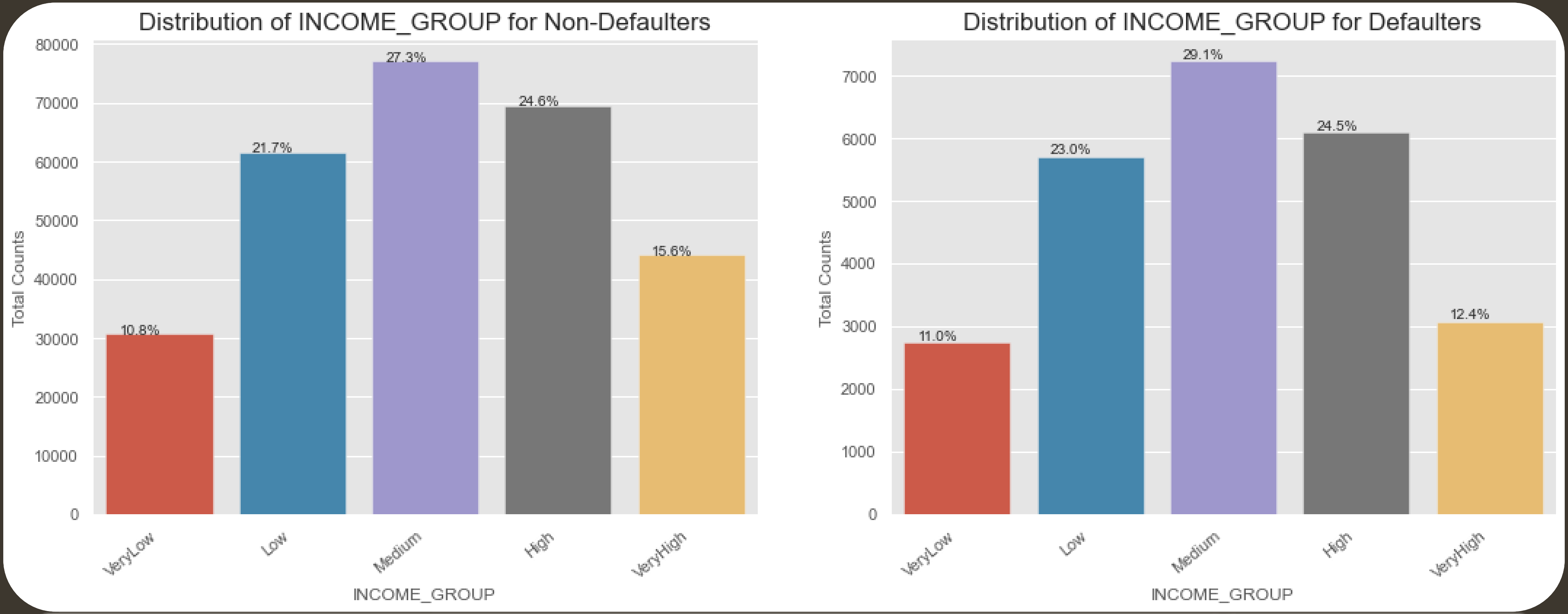
Gender



Distribution of CODE_GENDER for Non-Defaulters
Total Counts
0
25000
50000
75000
100000
125000
150000
175000
66.6%
33.4%
Female
Male
CODE_GENDER
Distribution of CODE_GENDER for Defaulters
Total Counts
0
2000
4000
6000
8000
10000
12000
14000
42.9%
57.1%
Male
Female
CODE_GENDER

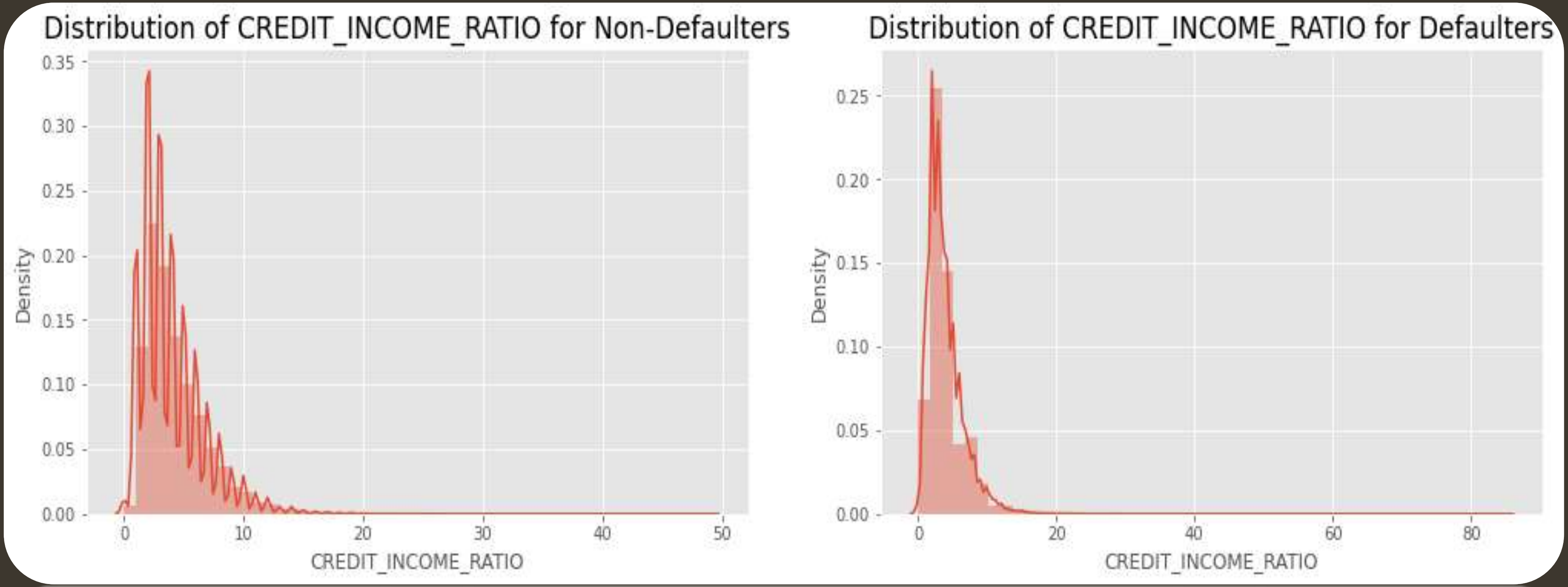
- We can see that Female contribute 67% to the non-defaulters while 57% to the defaulters.
- We see more female applying for loans than males and hence the more number of female defaulters as well.
- But the rate of defaulting of FEMALE is much lower compared to their MALE counterparts.

Income Group



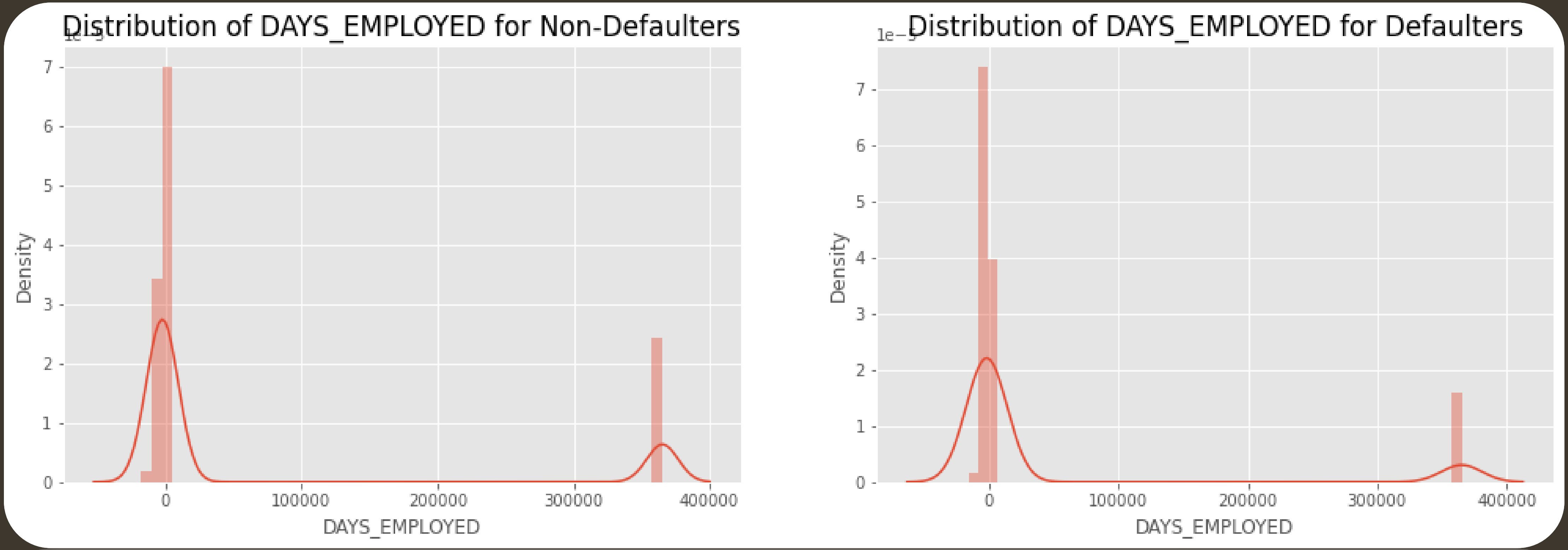
- The Very High income group tend to default less often.
- They contribute 12.4% to the total number of defaulters, while they contribute 15.6% to the Non-Defaulters.

Distribution of Credit Income Ratio

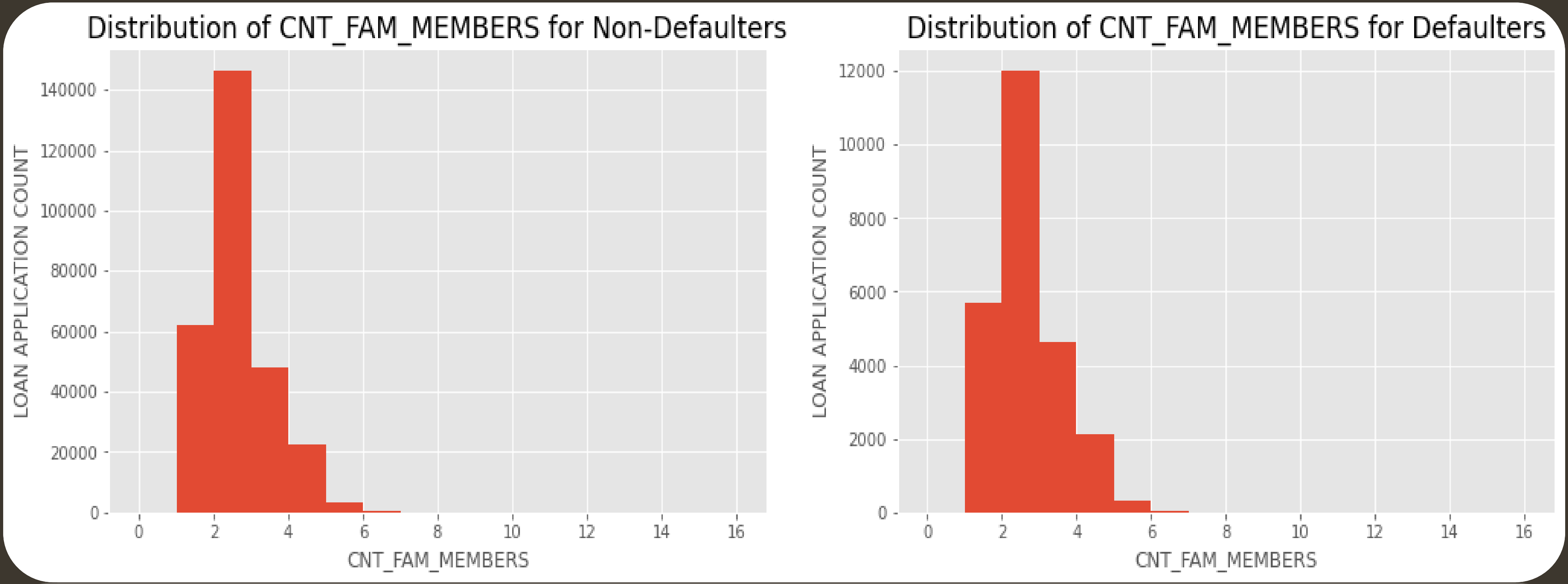


- Credit income ratio the ratio of AMT_CREDIT/AMT_INCOME_TOTAL.
- Although there doesn't seem to be a clear distiguish between the group which defaulted vs the group which didn't when compared using the ratio, we can see that when the CREDIT_INCOME_RATIO is more than 50, people default.

Distribution of Days Employed

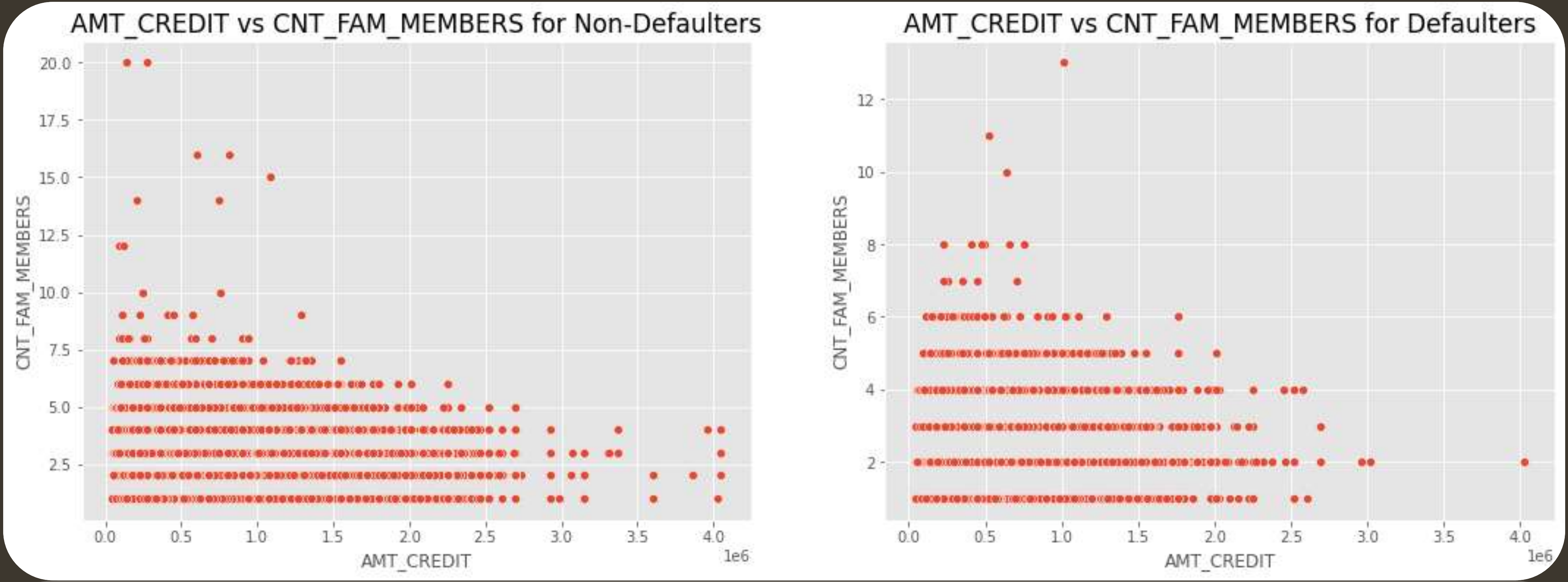


Distribution of Number of Family Members



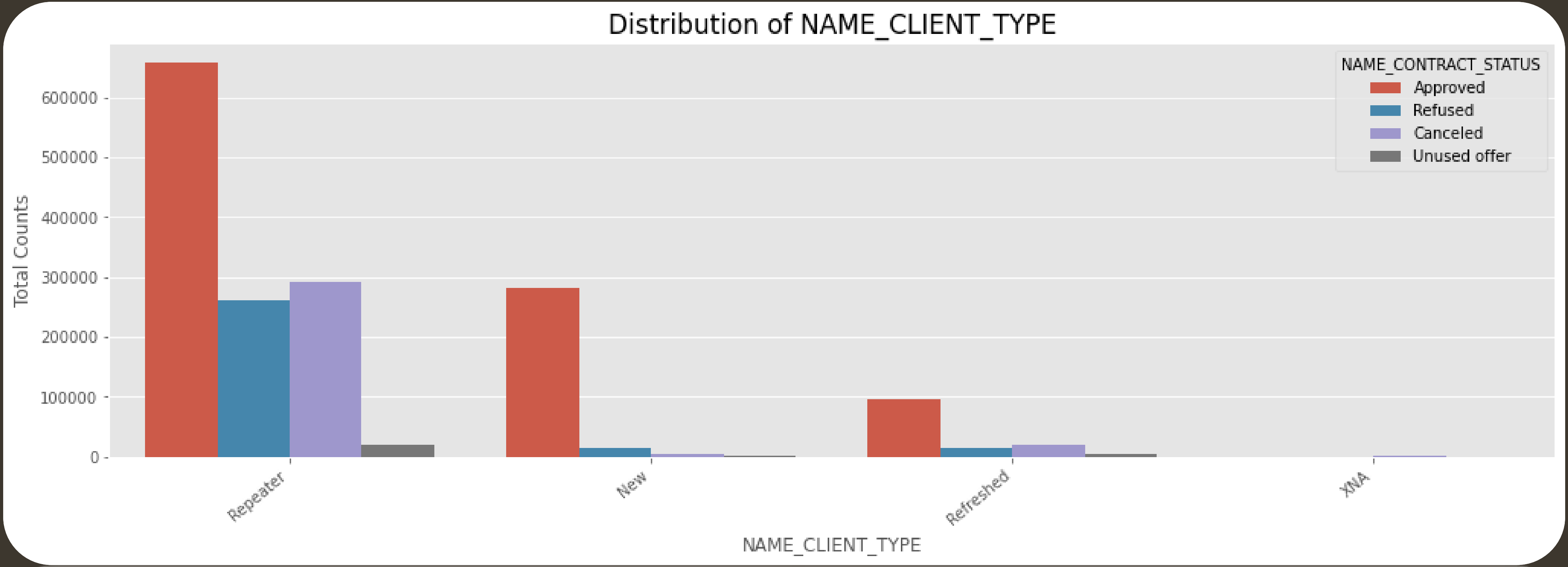
- We can see that a family of 3 applies loan more often than the other families.

Credit Vs Family Member



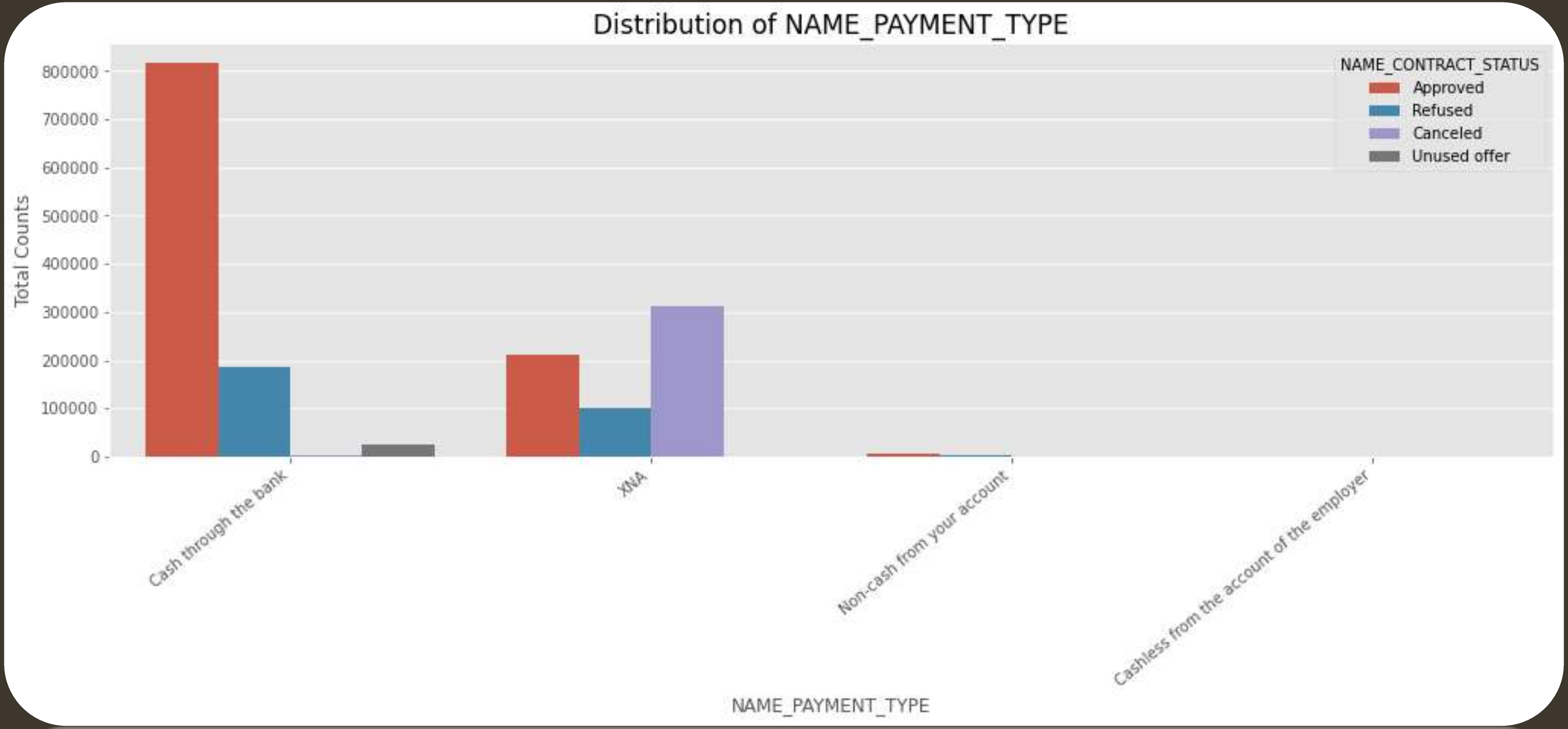
- We can see that the density in the lower left corner is similar in both the case, so the people are equally likely to default if the family is small and the AMT_CREDIT is low.
- We can observe that larger families and people with larger AMT_CREDIT default less often.

Name Client Type



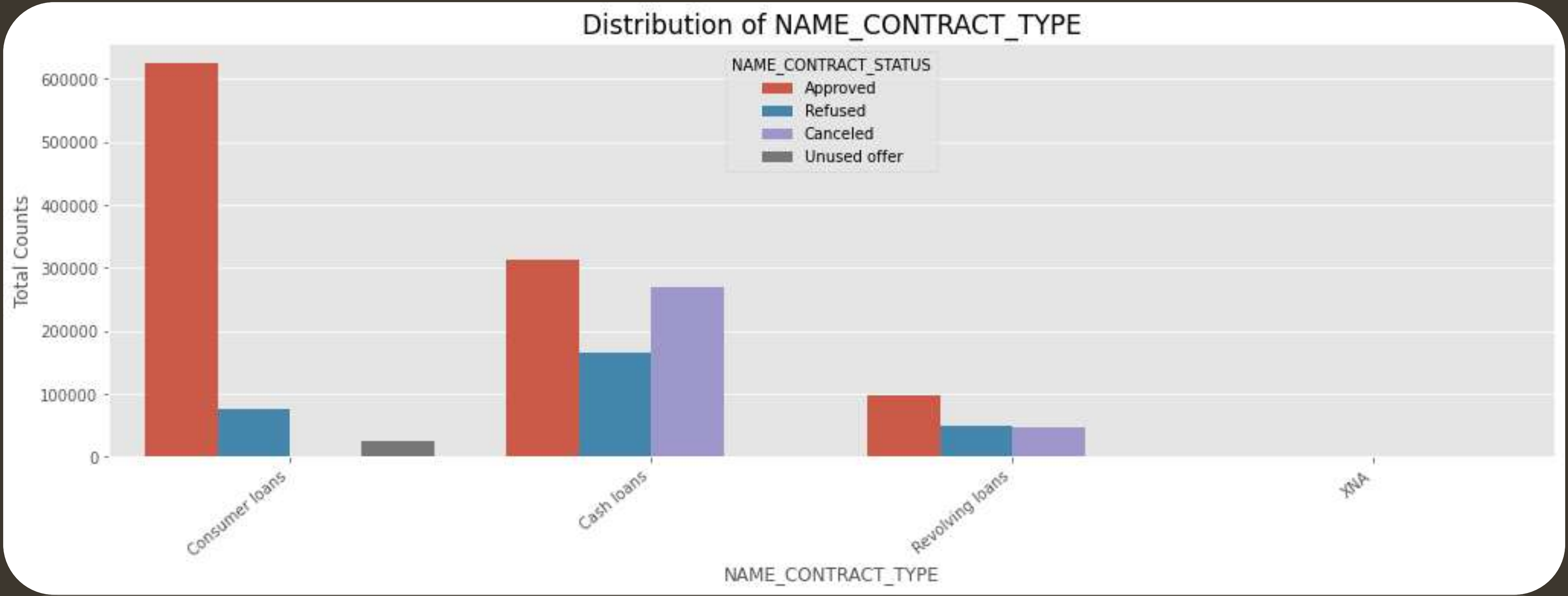
- Most of the loan applications are from repeat customers, out of the total applications 70% of customers are repeaters. They also get refused most often.

Name Payment Type



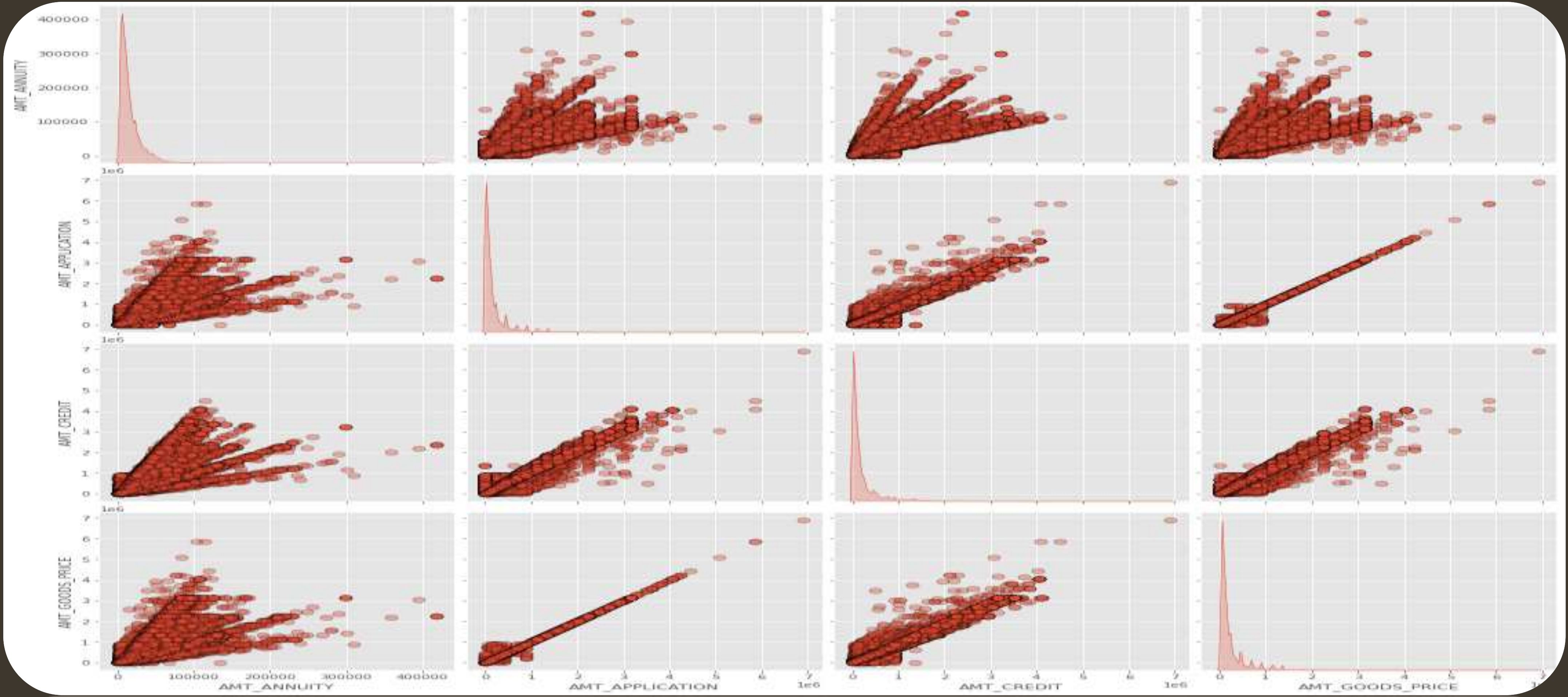
- From the above chart, we can infer that most of the clients chose to repay the loan using the 'Cash through the bank' option.
- We can also see that 'Non-Cash from your account' & 'Cashless from the account of the employee' options are not at all popular in terms of loan repayment amongst the customers.

Contract Type



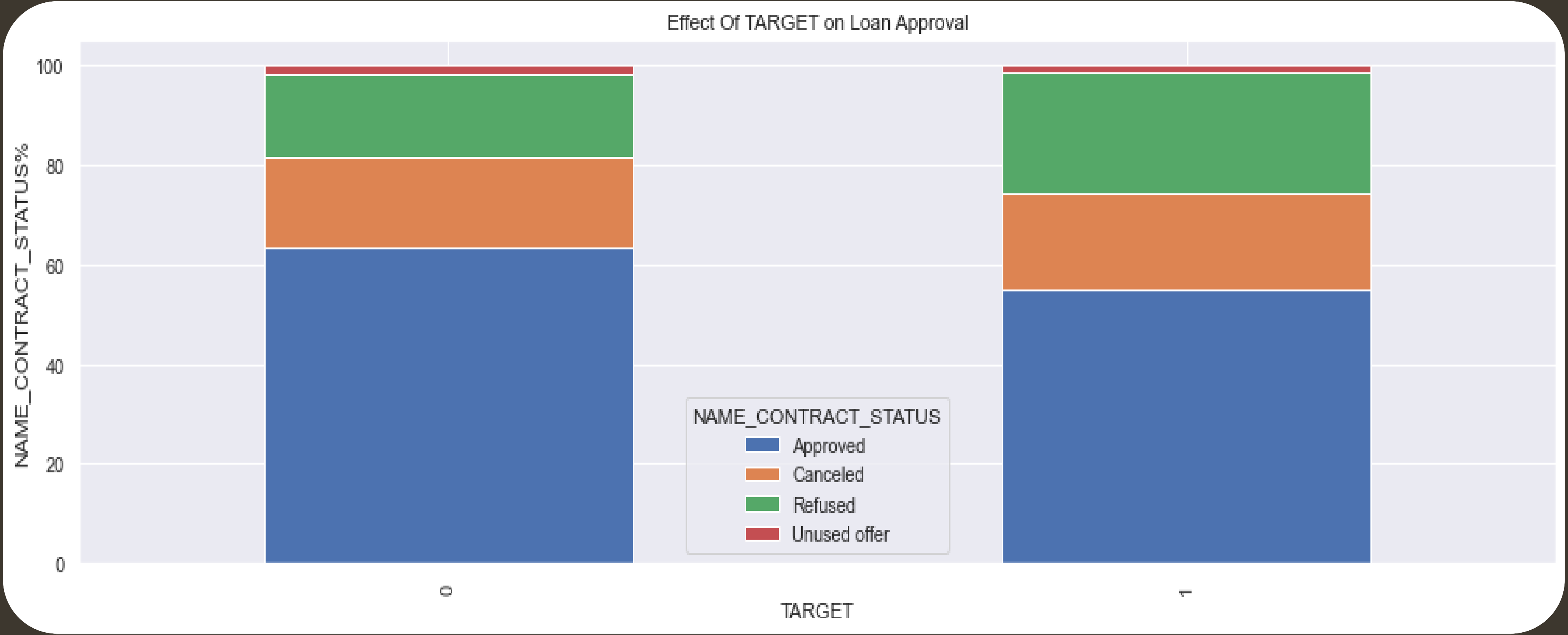
- From the above chart, we can infer that, most of the applications are for 'Cash loan' and 'Consumer loan’.
- Although the cash loans are refused more often than others.

Bivariate Analysis



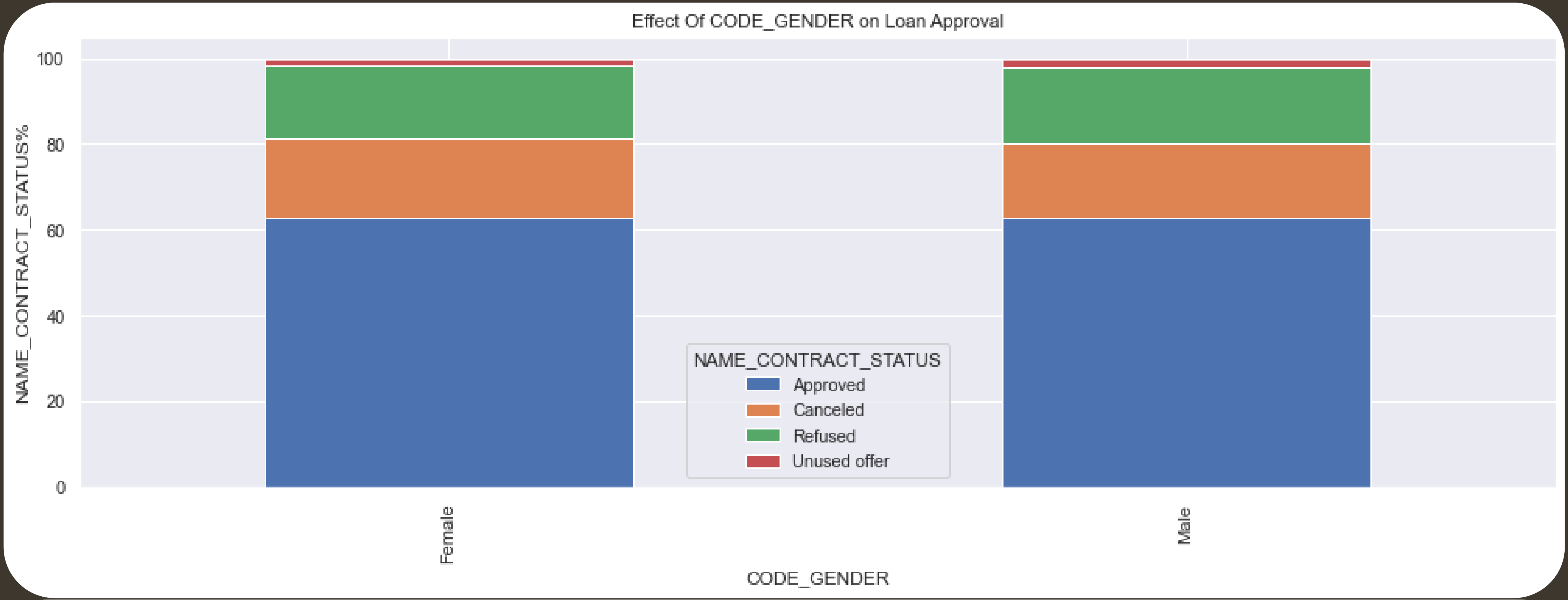
- Annuity of previous application has a very high and positive influence over: (Increase of annuity increases below factors)
  - (1) How much credit did client asked on the previous application
  - (2)Final credit amount on the previous application that was approved by the bank
  - (3) Goods price of 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.

Effect of Target on Loan Approval



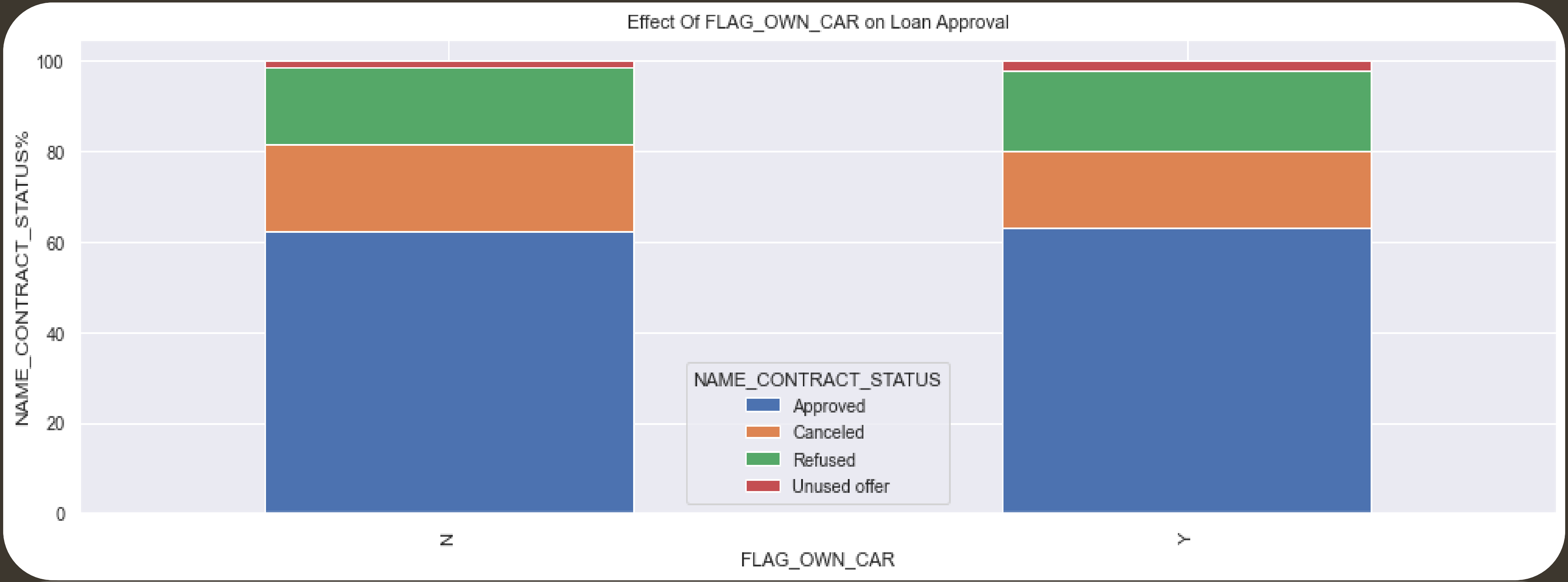
- We can see that the people who were approved for a loan earlier, defaulted less often where as people who were refused a loan earlier have higher chances of defaulting.

Effect of 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 Car Owner 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.

Thank You