CREDIT EDA CASE STUDY

BY PANKAJ MORALWAR

PROBLEM STATEMENT

1. Aim is to identify patterns which indicate if a client had difficulty paying the installments which.

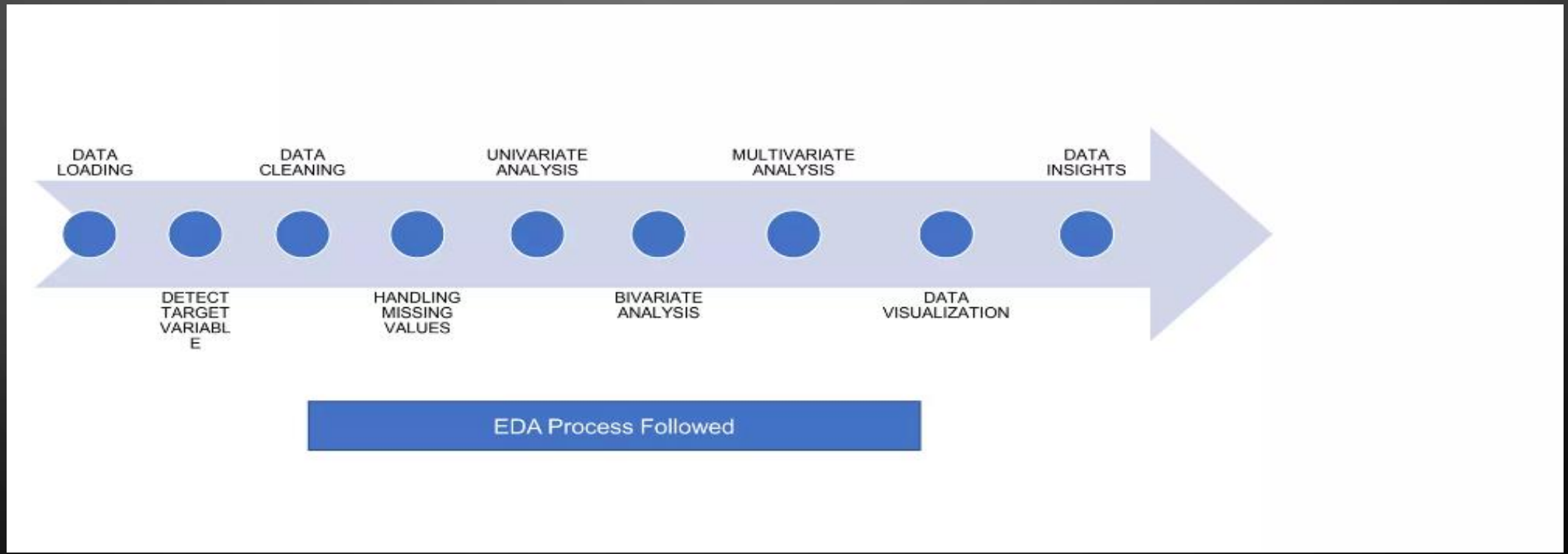
will help the bank in taking following actions:

- Denying the loan
- Reducing the amount of loan
- Lending (to risky applicants) at a higher interest rate, etc.

2. Identifying the co-relation between dependent variables with target variable.

3. To ensure the the consumers capable of repaying the loan are not rejected.

STEPS TO FOLLOW

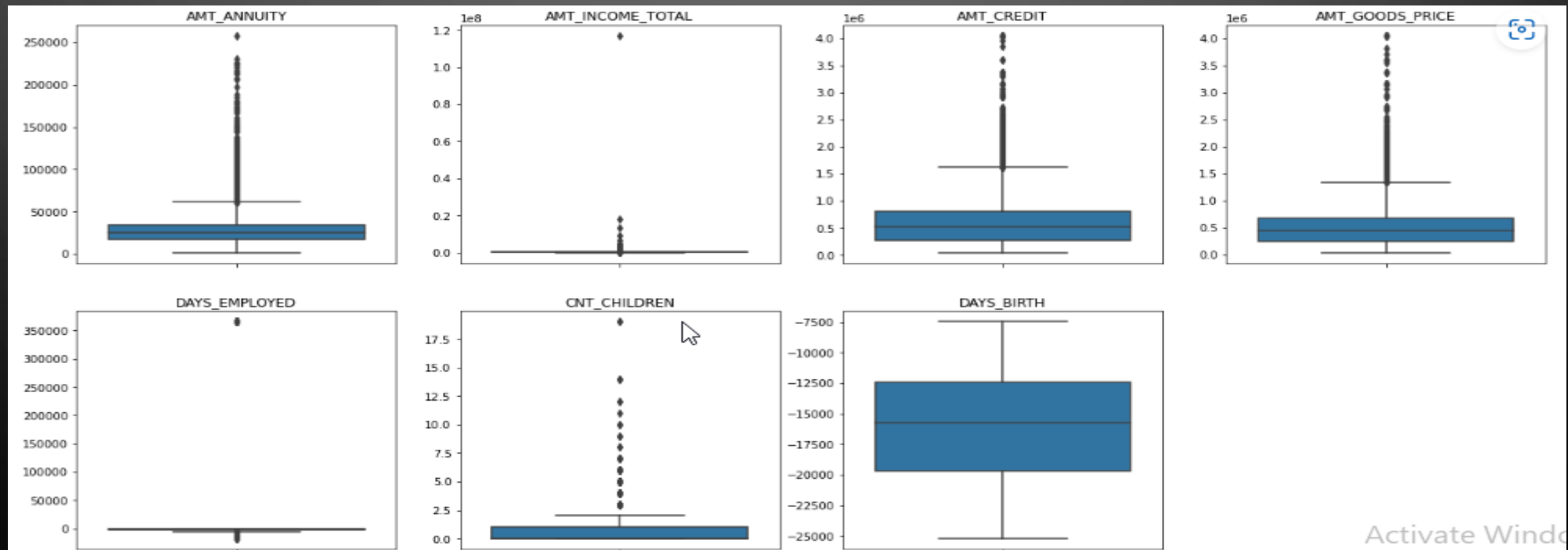


OUTLIER ANALYSIS

OUTLIER ANALYSIS FOR NEW_APPLICATIONS

Points to be concluded from the below graph.

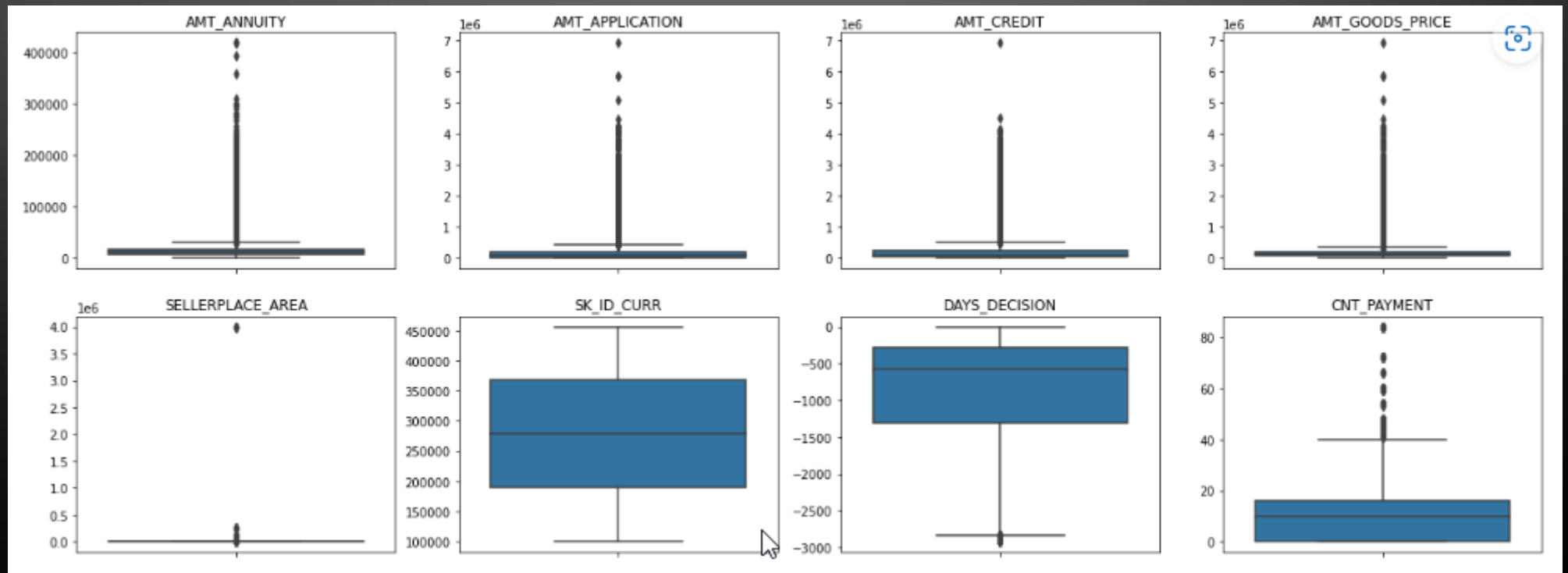
- 1. AMT_ANNUITY, AMT_CREDIT, AMT_GOODS_PRICE,CNT_CHILDREN have some number of outliers.
- 2. AMT_INCOME_TOTAL has huge number of outliers which indicate that few of the loan applicants have high income when compared to the others.
- 3. DAYS_BIRTH has no outliers which means the data available is reliable.
- 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.



OUTLIER ANALYSIS FOR PREVIOUS_APPLICATIONS

Points to be concluded from the below graph.

- 1. AMT_ANNUITY, AMT_APPLICATION, AMT_CREDIT, AMT_GOODS_PRICE, SELLERPLACE_AREA have a greater number of outliers.
- 2. CNT_PAYMENT has few outlier values.
- 3. SK_ID_CURR is an ID column and hence no outliers.
- 4. DAYS_DECISION has little number of outliers indicating that these previous applications decisions were taken long back.

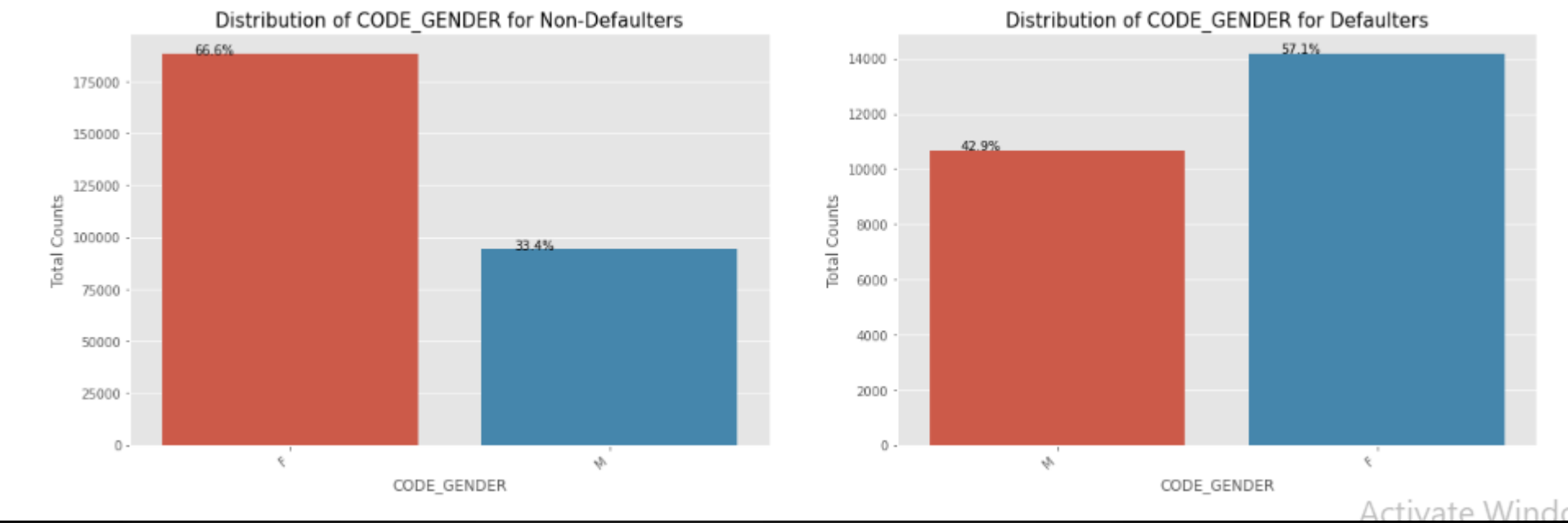


CATEGORICAL UNIVARIATE ANALYSIS FOR NON-DEFAULTER AND DEFAULT

DISTRIBUTION OF THE PEOPLE BASED ON GENDER

Points to be concluded from the below graph.

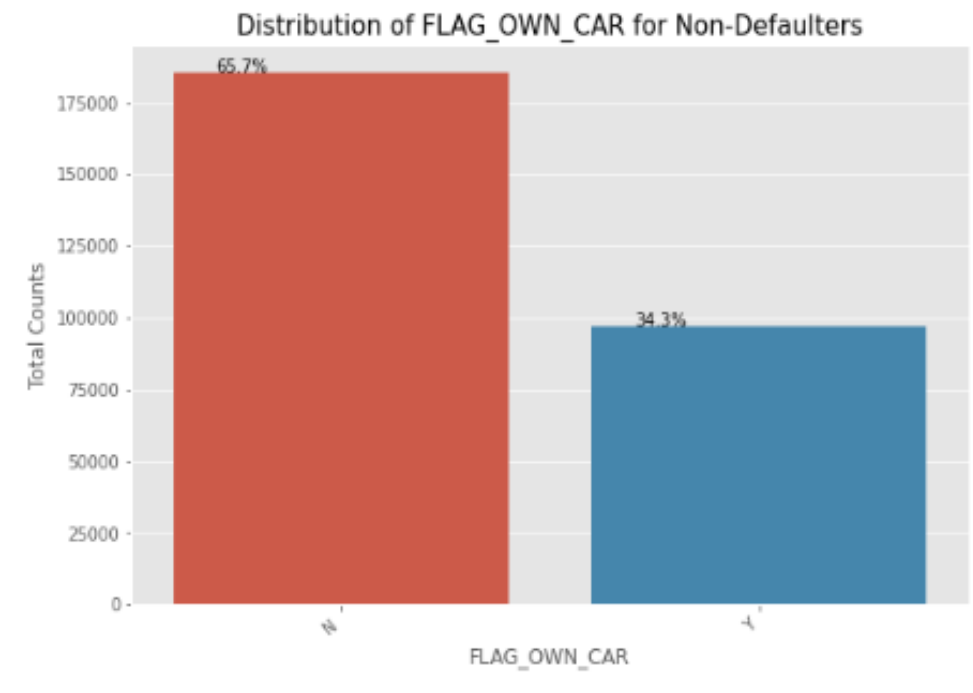
- 1. We can see that Female contribute 67% to the non-defaulters while 57% to the defaulters.
- 2. We see more female applying for loans than males and hence the more number of female defaulters as well.
- 3. **But the rate of defaulting of MALE is much lower compared to their FEMALE counterparts.**

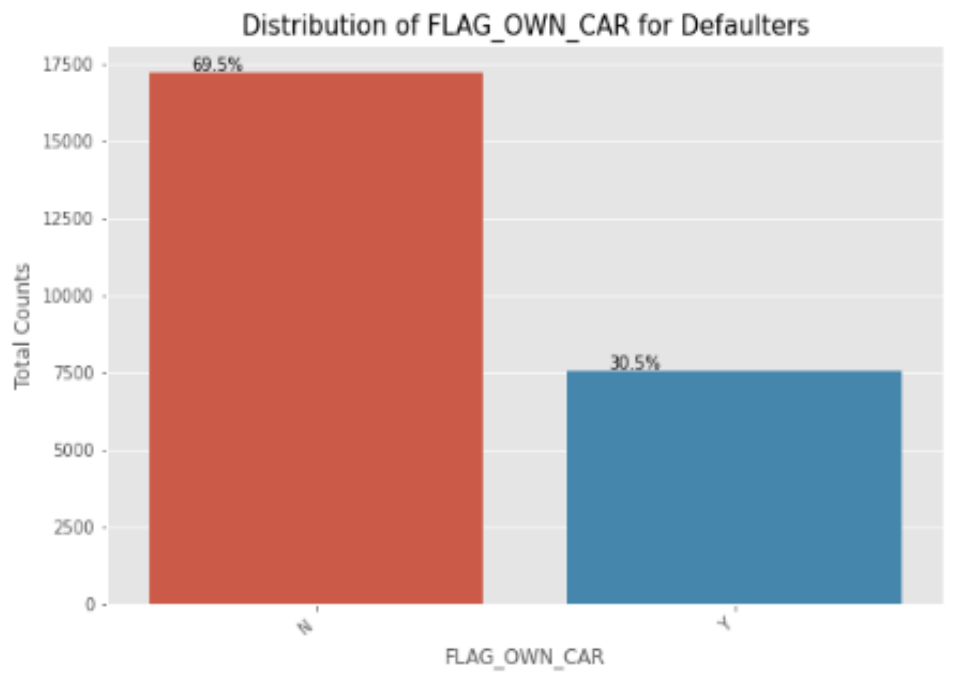


DISTRIBUTION OF THE PEOPLE BASES ON OWN CAR

Points to be concluded from the below graph.

- 1. We can see that people without cars contribute 65.7% to the non-defaulters while 69.5% to the defaulters.
- 2. We can conclude that, The people who don't have car default more often
- 3. **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.**

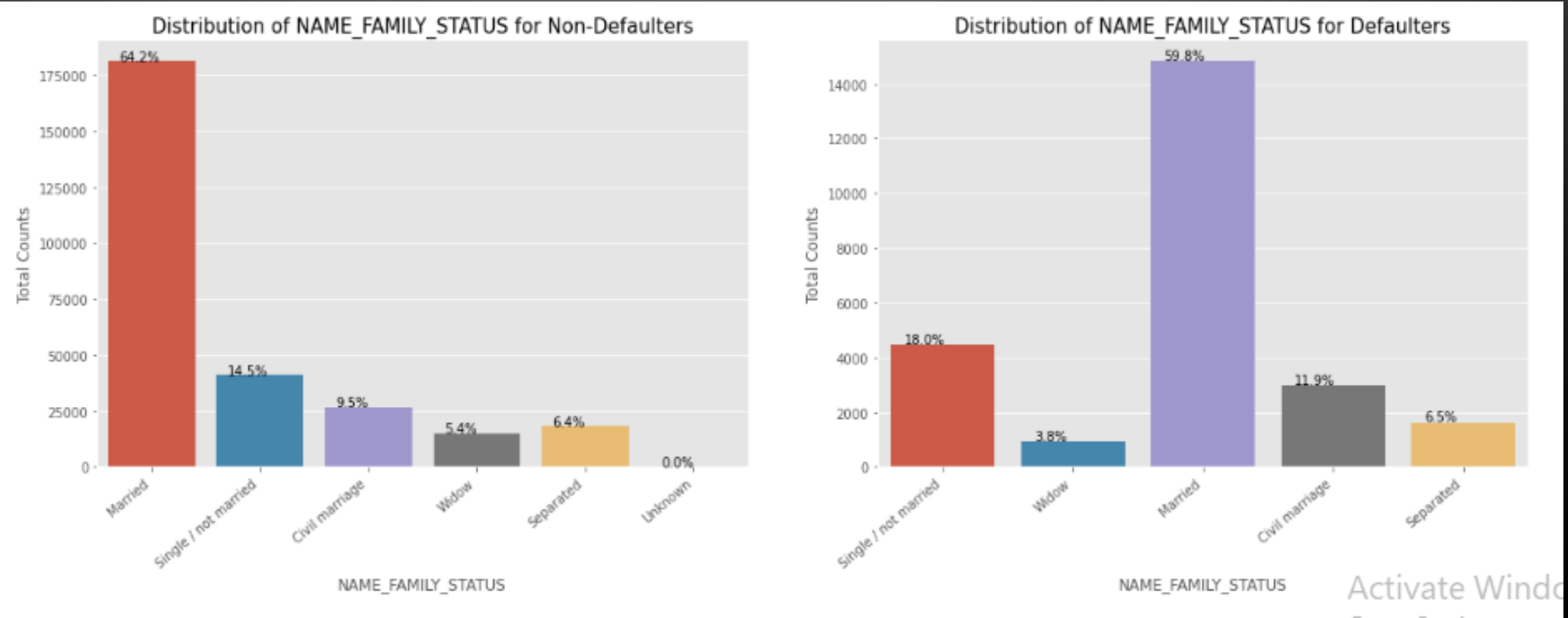




DISTRIBUTION OF THE PEOPLE BASES ON NAME_FAMILY_STATUS

Points to be concluded from the below graph.

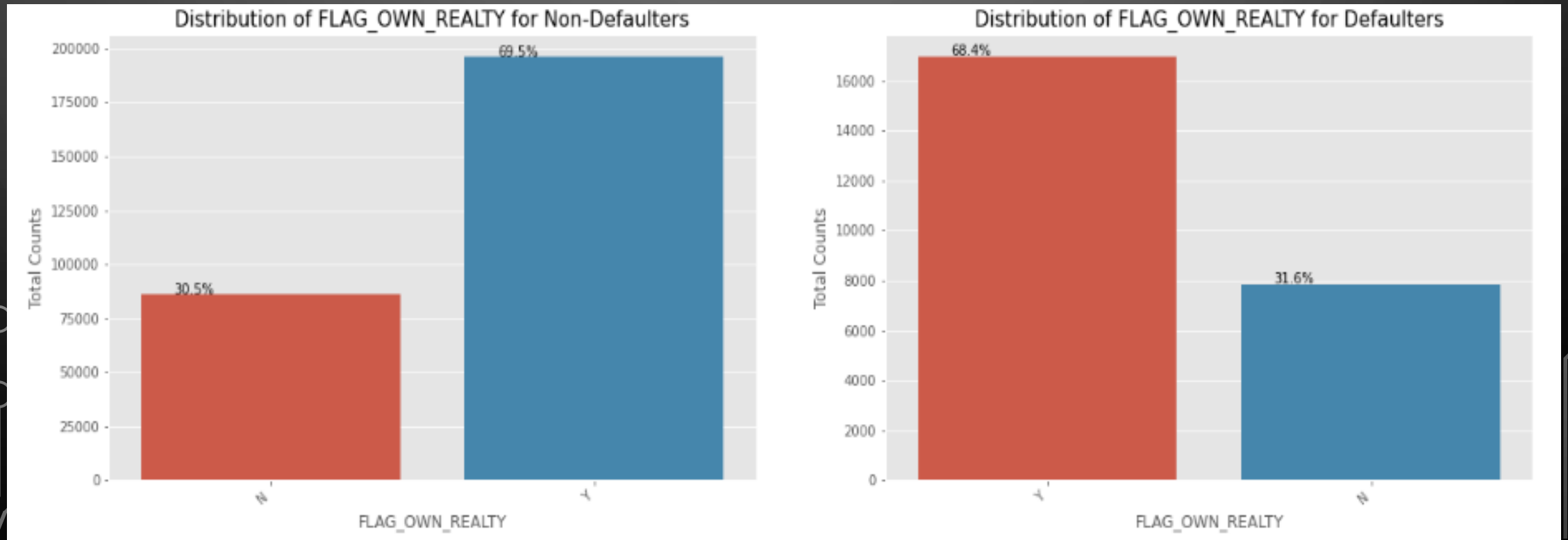
- 1. Married people tend to apply for more loans comparatively.
- 2. But from the graph we see that Single/non Married people contribute 14.5% to Non Defaulters and 18% to the defaulters. So there is more risk associated with them.



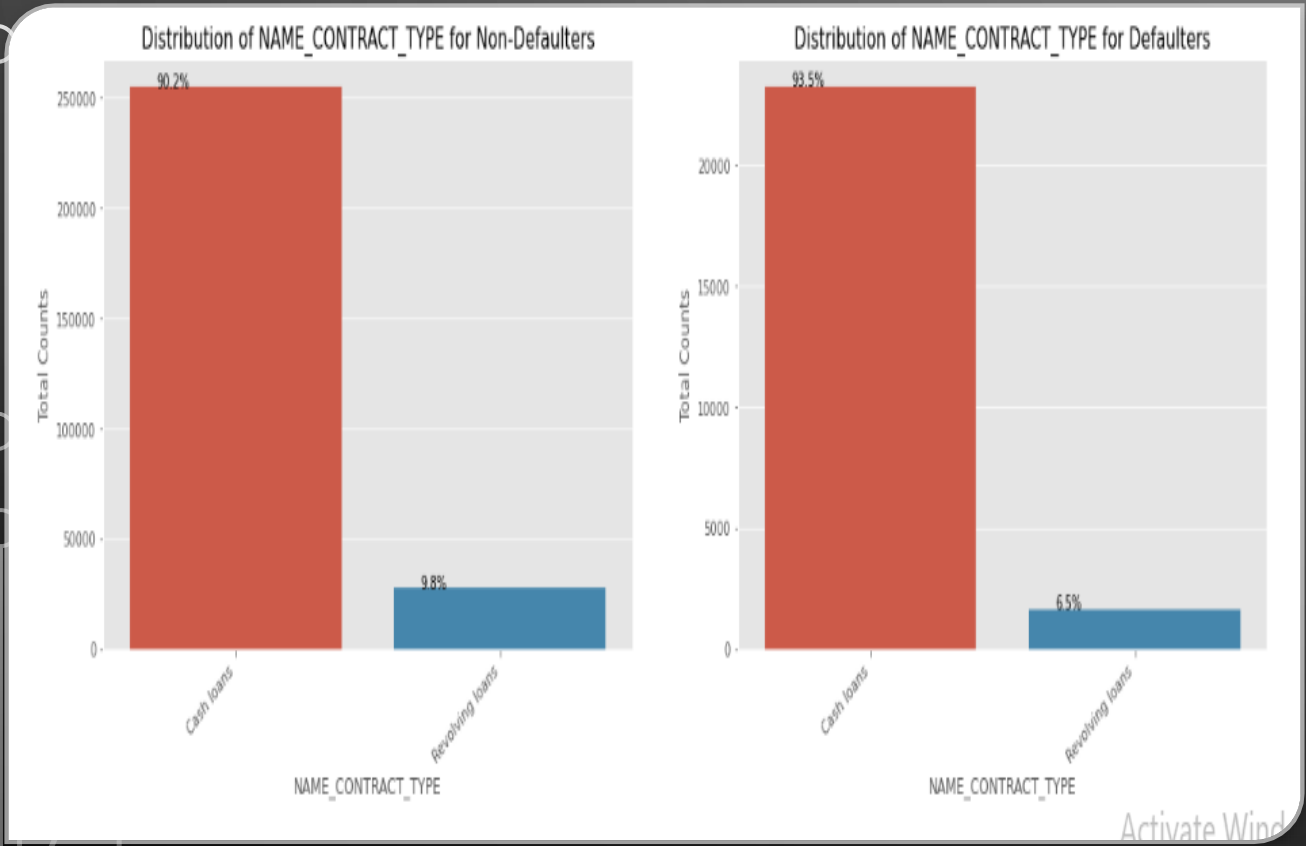
DISTRIBUTION OF PEOPLE REAL-ESTATE VS NOT OWNING REAL-ESTATE

Points to be concluded from the below graph.

- 1. The clients who own real estate are more than double of the ones that don't own.
- 2. But the defaulting rate of both categories are around the same (~8%).
- 3. Thus,there is no correlation between owning a reality and defaulting the loan.



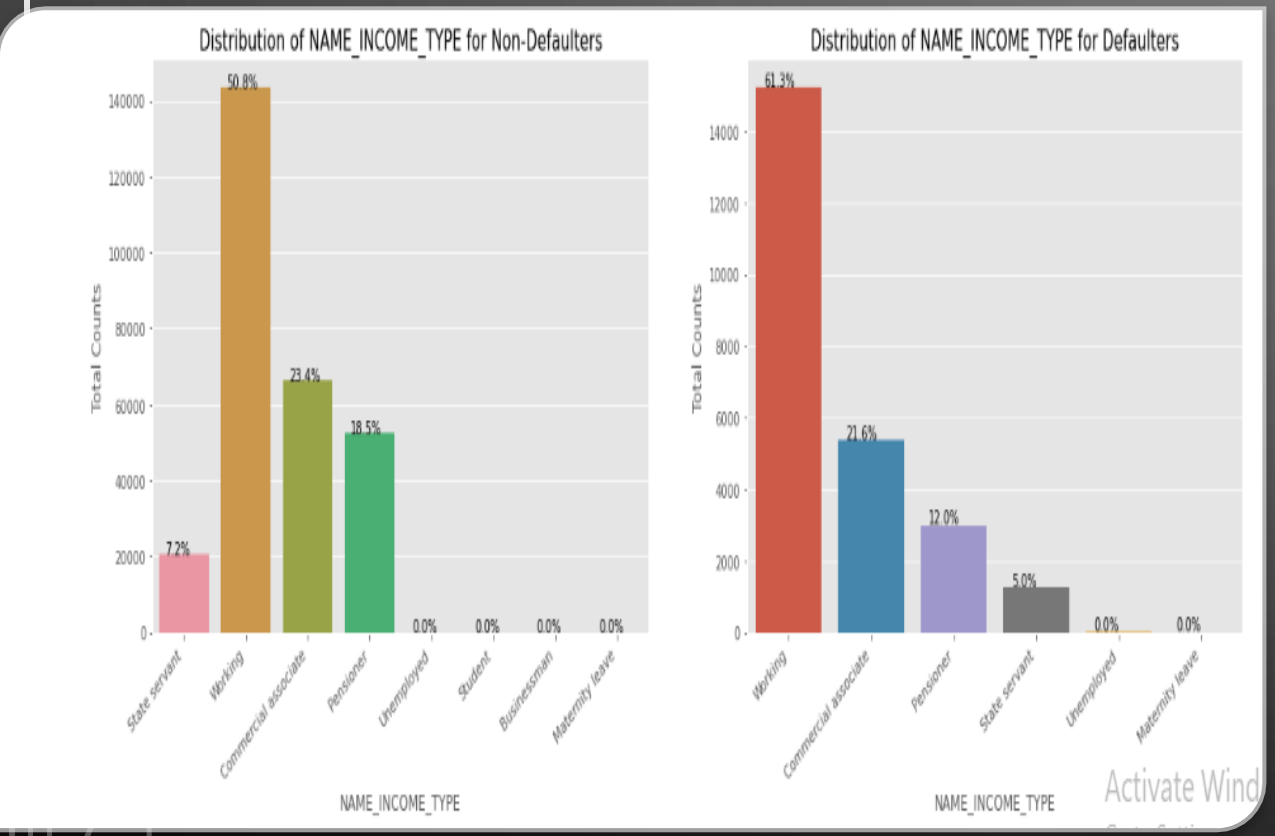
DISTRIBUTION BASED ON CONTRACT TYPE OF LOAN - REVOLVING OR CASH LOANS



Points to be concluded from the below graph.

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

DISTRIBUTION OF INCOME TYPE



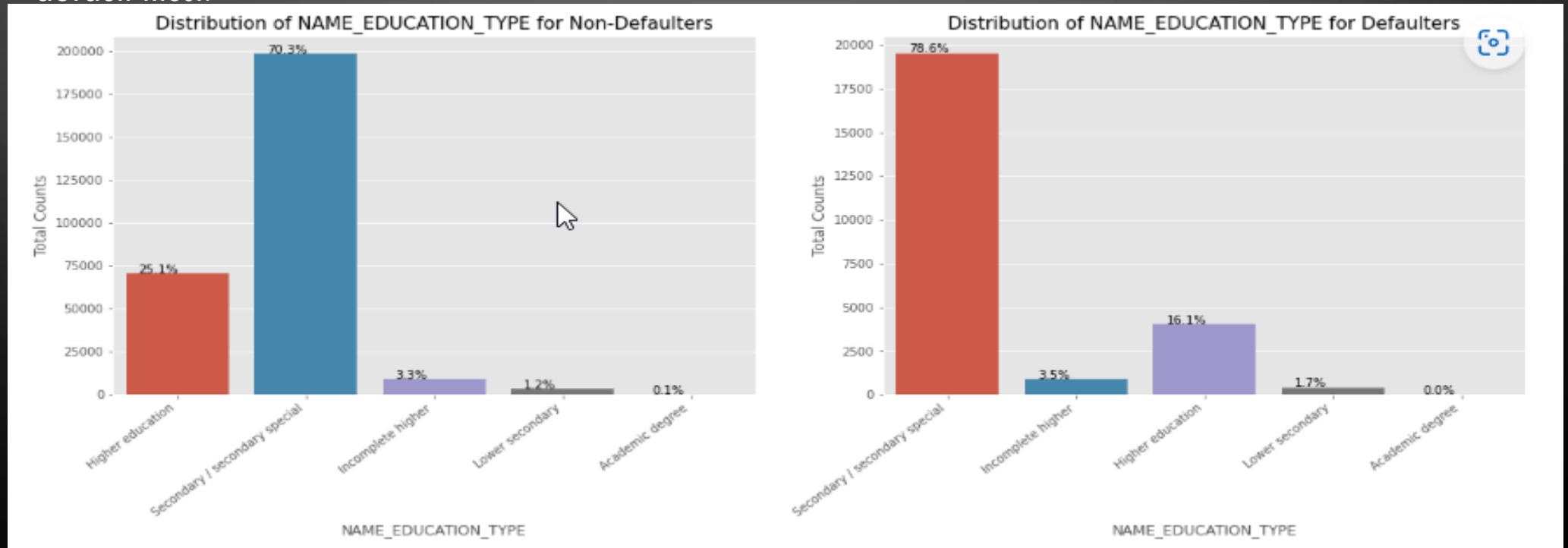
Points to be concluded from the graph on the left.

- 1. 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 because the educational loan will have to pay after education is over.
- 2. We can also see that the Businessman never default.
- 3. Most of the loans are distributed to working class people.
- 4. 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.

DISTRIBUTION OF PEOPLE ON THEIR EDUCATION LEVEL

Points to be concluded from the below graph.

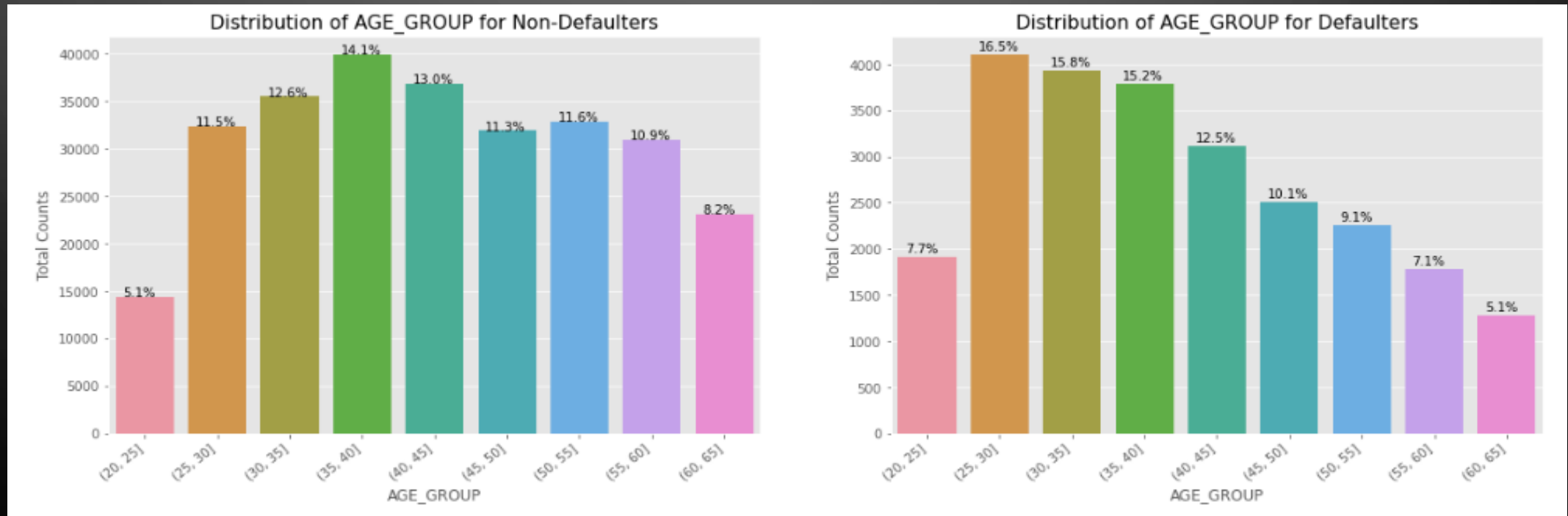
- 1. The clients who have acadamic degree tends to not applying for loan where as lower education clients apply for loan more often.
- 2. The clients who have higher education , 'secondary/secondary special' education are more likely to apply loans.
- 3. Where as the 'secondary/secondary special' & 'Higher education' clients are mostly applying for loans also they tends to default most.



DISTRIBUTION OF PEOPLE ON THEIR AGE_GROUP

Points to be concluded from the below graph.

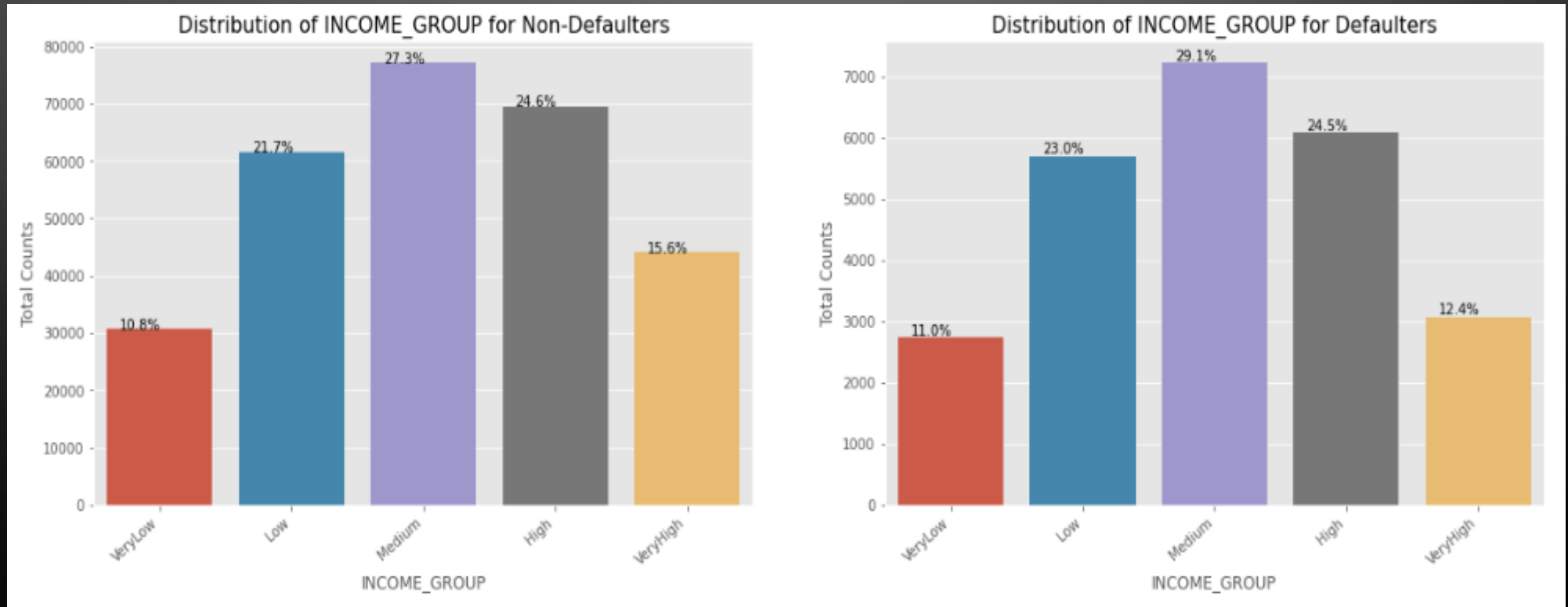
- 1. We see that (25,30] age group tend to default more often. So they are the riskiest people to loan to.
- 2. With increasing age group, people tend to default less starting from the age 25. One of the reasons could be they get employed around that age and with increasing age, their salary also increases.



DISTRIBUTION OF PEOPLE ON THEIR INCOME_GROUP

Points to be concluded from the below graph.

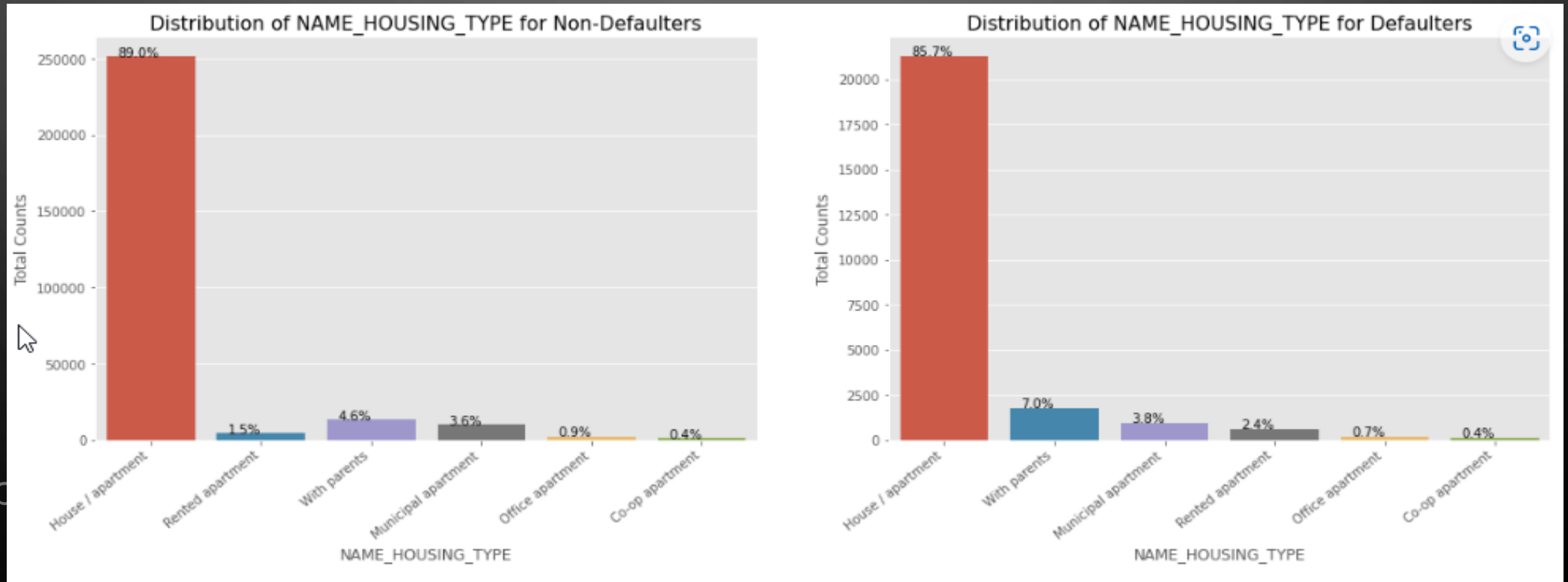
- 1. Very high income group people are less likely to default.
- 2. Medium income group are likely to apply more for the loan as well as they are more likely to Default.



DISTRIBUTION OF PEOPLE ON THEIR HOUSING_TYPE

Points to be concluded from the below graph.

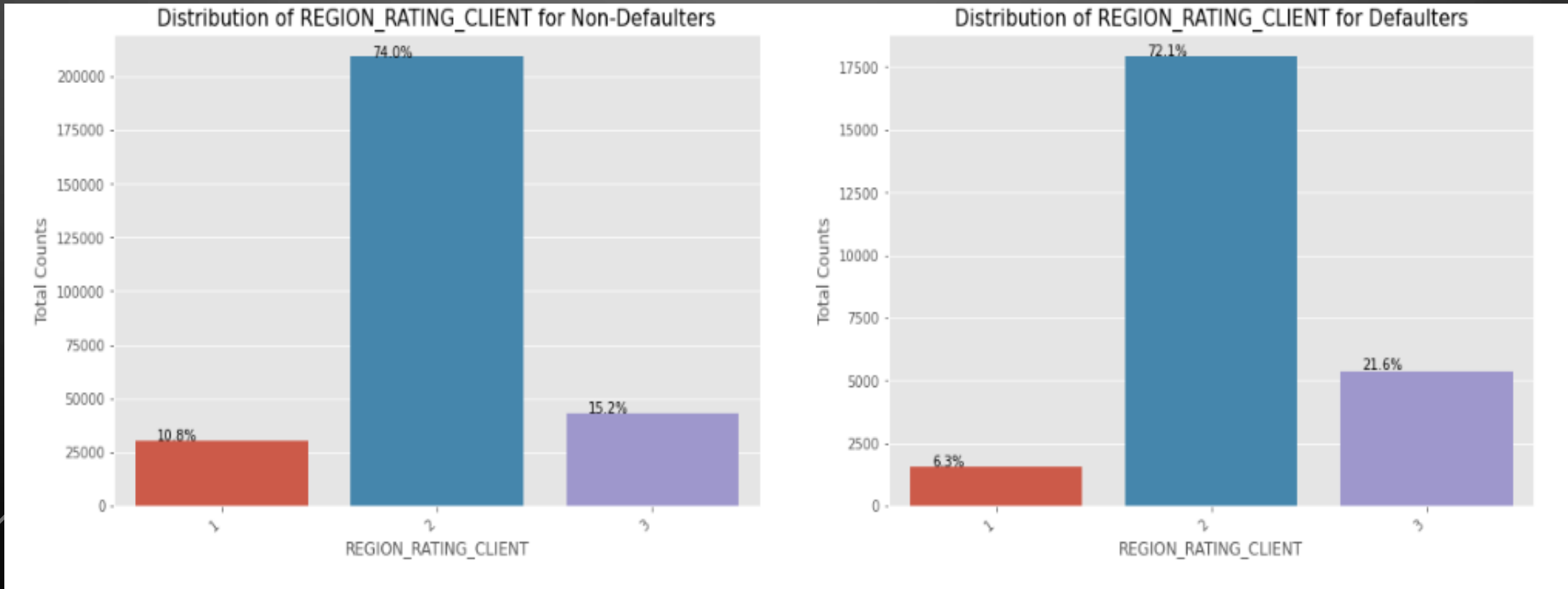
- 1. People with House / Apartment are more likely to apply for loan and they are are also more likely to default.



DISTRIBUTION OF PEOPLE ON THEIR REGION

Points to be concluded from the below graph.

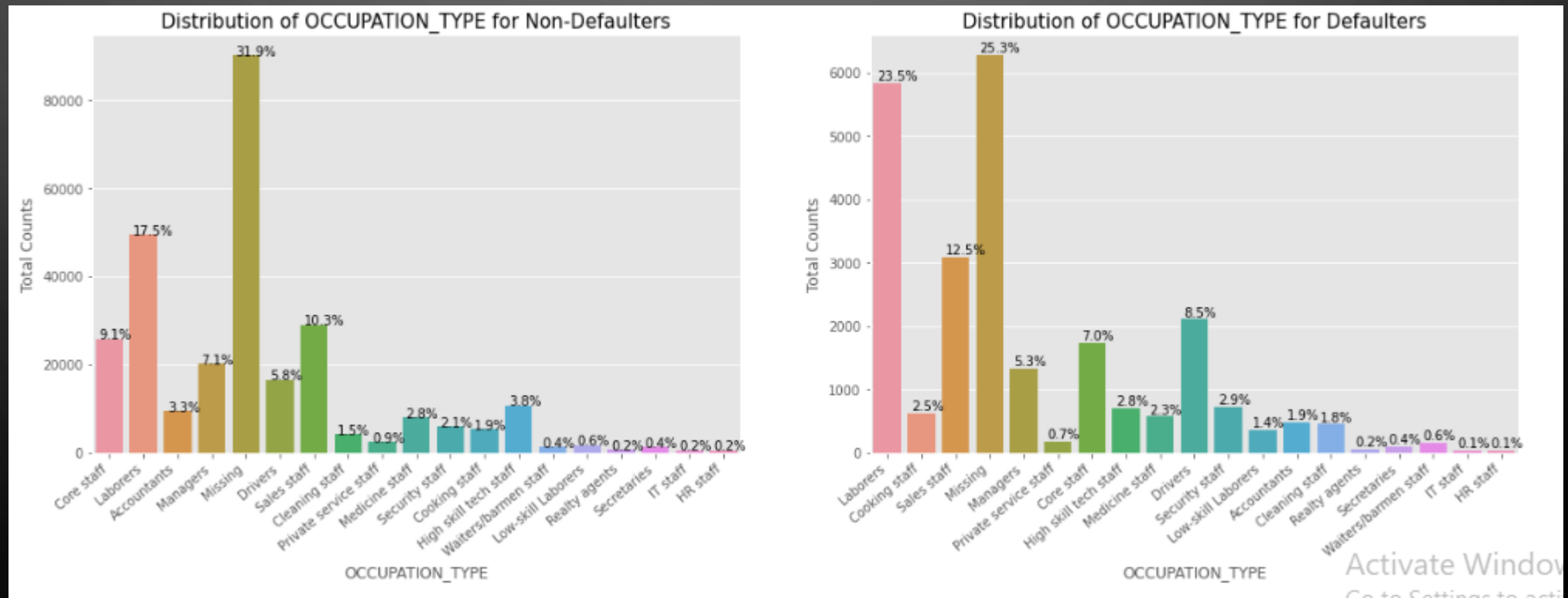
- 1. People who live in region 2 rating has applied mostly for loans.
- 2. Also they are more likely to default most. After them people who live in region 3 are likely to default.



DISTRIBUTION OF PEOPLE ON THEIR OCCUPATION_TYPE

Points to be concluded from the below graph.

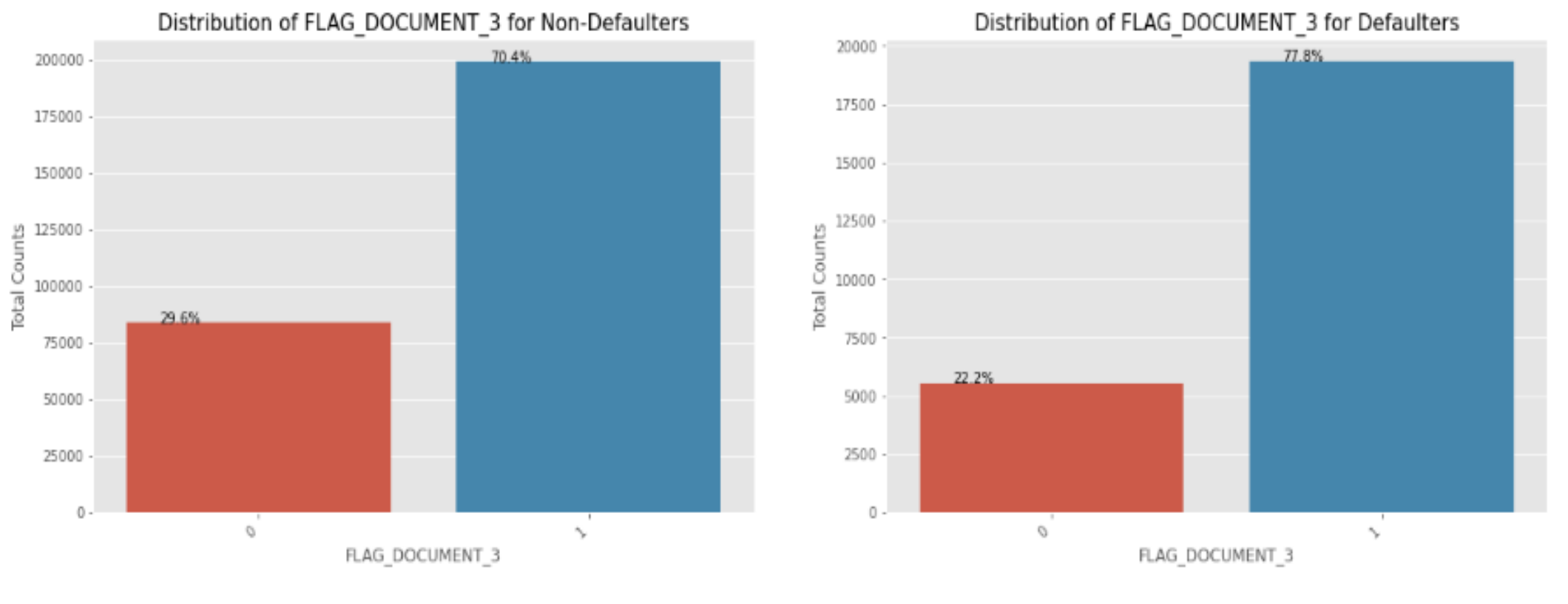
1. People with 'missing' occupation type are applied for more no. of loans followed by 'Labours'
2. Also people with 'missing' & 'Labours' are more often to get default.



DISTRIBUTION OF PEOPLE ON FLAG_DOCUMENT_3

Points to be concluded from the below graph.

1. There is no significant correlation between repayers and defaulters in terms of submitting document 3 as we see even if applicants have submitted the document, they have defaulted with who have not submitted the document.

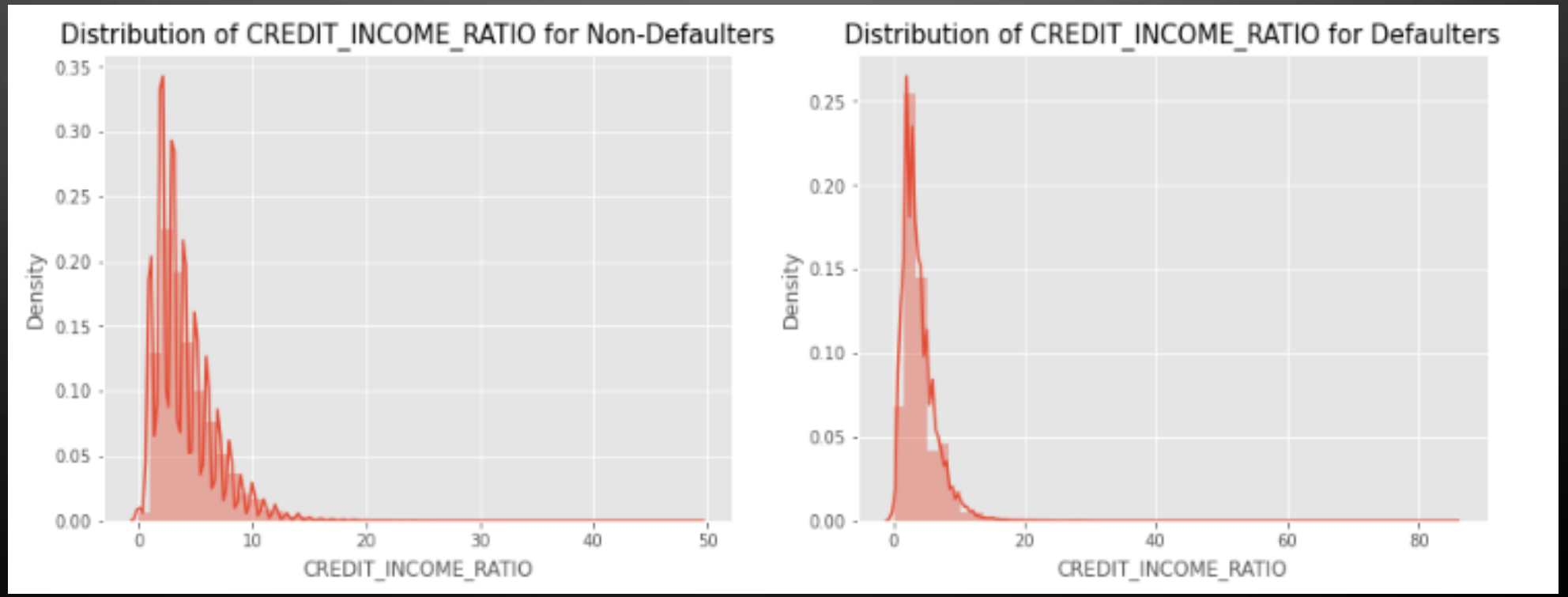


DISTRIBUTION OF PEOPLE ON CREDIT_INCOME_RATIO

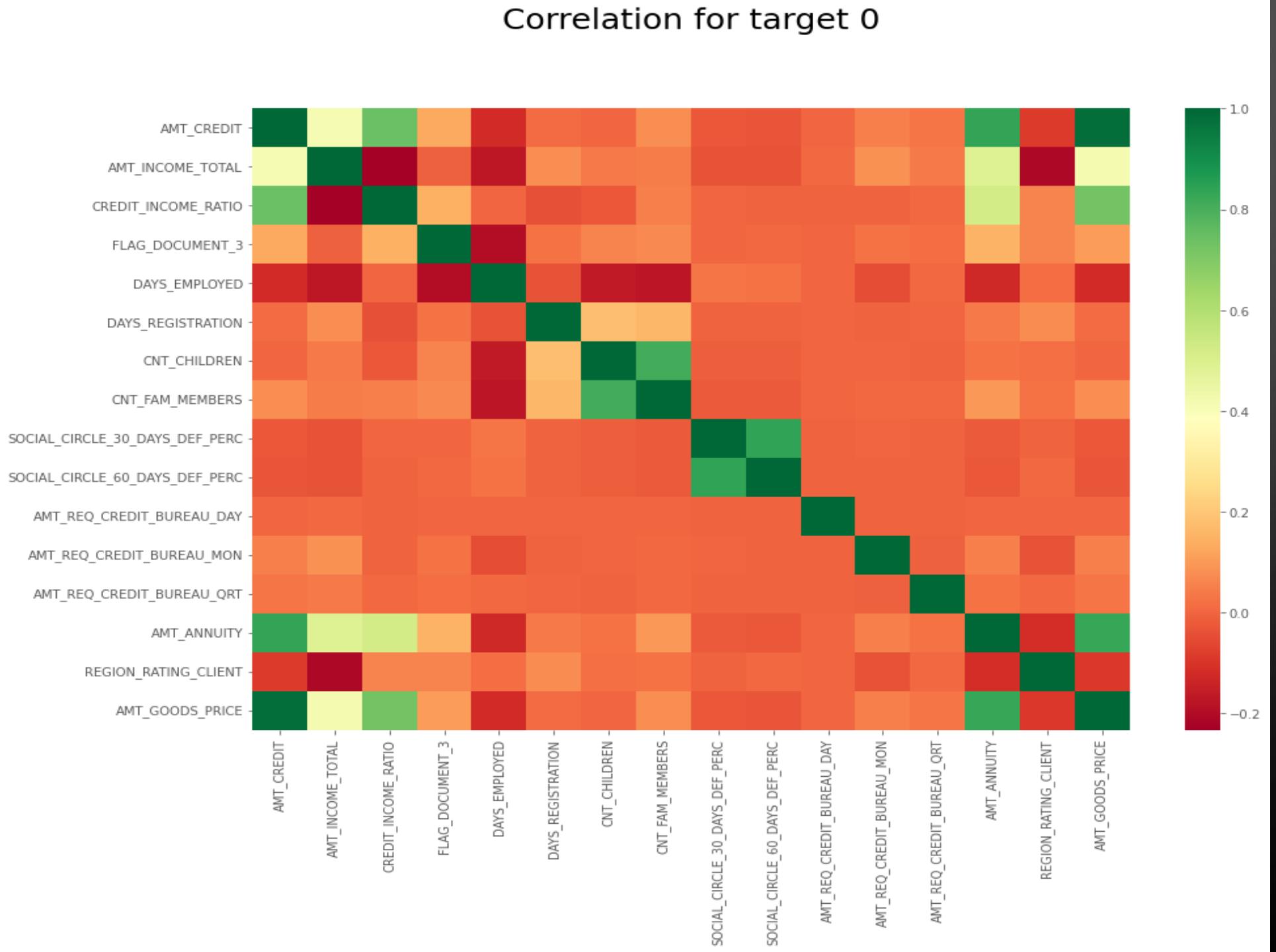
Points to be concluded from the below graph.

1. Credit income ratio the ratio of AMT_CREDIT/AMT_INCOME_TOTAL.

2. 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.



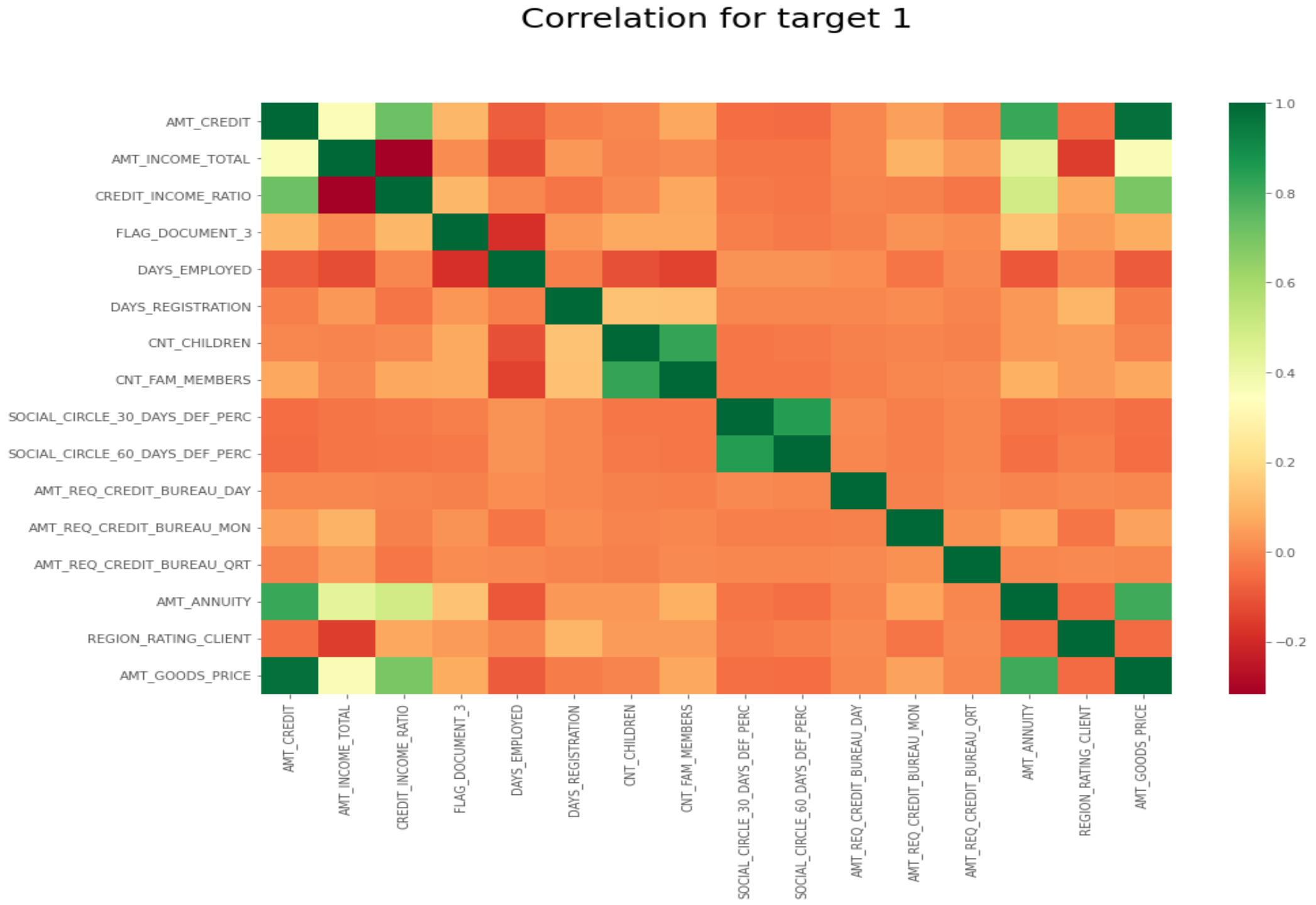
CORRELATION BETWEEN VARIABLES FOR TARGET 0 & TARGET 1


Correlation for target 0
AMT_CREDIT
AMT_INCOME_TOTAL
CREDIT_INCOME_RATIO
FLAG_DOCUMENT_3
DAYS_EMPLOYED
DAYS_REGISTRATION
CNT_CHILDREN
CNT_FAM_MEMBERS
SOCIAL_CIRCLE_30_DAYS_DEF_PERC
SOCIAL_CIRCLE_60_DAYS_DEF_PERC
AMT_REQ_CREDIT_BUREAU_DAY
AMT_REQ_CREDIT_BUREAU_MON
AMT_REQ_CREDIT_BUREAU_QRT
AMT_ANNUITY
REGION_RATING_CLIENT
AMT_GOODS_PRICE
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

CORRELATION FOR TARGET 0

Points to be concluded from the graph presented before.

- Credit amount is inversely proportional to the days employed, which means Credit amount is higher for less days employed 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.
- less children client have in densely populated area.
- Credit amount is higher to densely populated area.
- The income is also higher in densely populated area.


Correlation for target 1
AMT_CREDIT
AMT_INCOME_TOTAL
CREDIT_INCOME_RATIO
FLAG_DOCUMENT_3
DAYS_EMPLOYED
DAYS_REGISTRATION
CNT_CHILDREN
CNT_FAM_MEMBERS
SOCIAL_CIRCLE_30_DAYS_DEF_PERC
SOCIAL_CIRCLE_60_DAYS_DEF_PERC
AMT_REQ_CREDIT_BUREAU_DAY
AMT_REQ_CREDIT_BUREAU_MON
AMT_REQ_CREDIT_BUREAU_QRT
AMT_ANNUITY
REGION_RATING_CLIENT
AMT_GOODS_PRICE
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

CORRELATION FOR TYPE 1

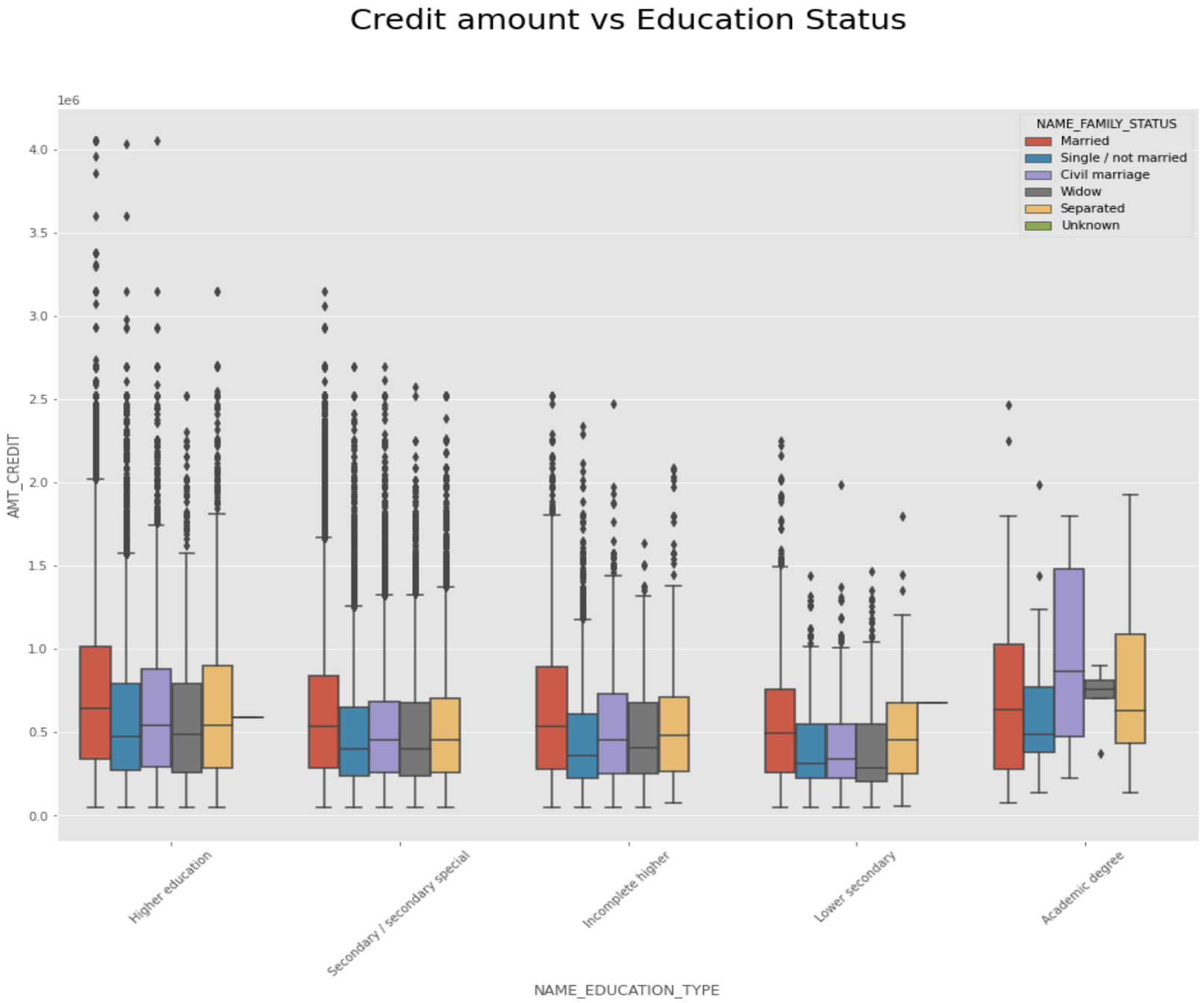
This heat map for Target 1 is also having quite a same observation just like Target 0.

BIVARIATE ANALYSIS

CREDIT AMOUNT VS EDUCATION STATUS

Few points can be concluded from the graph.

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

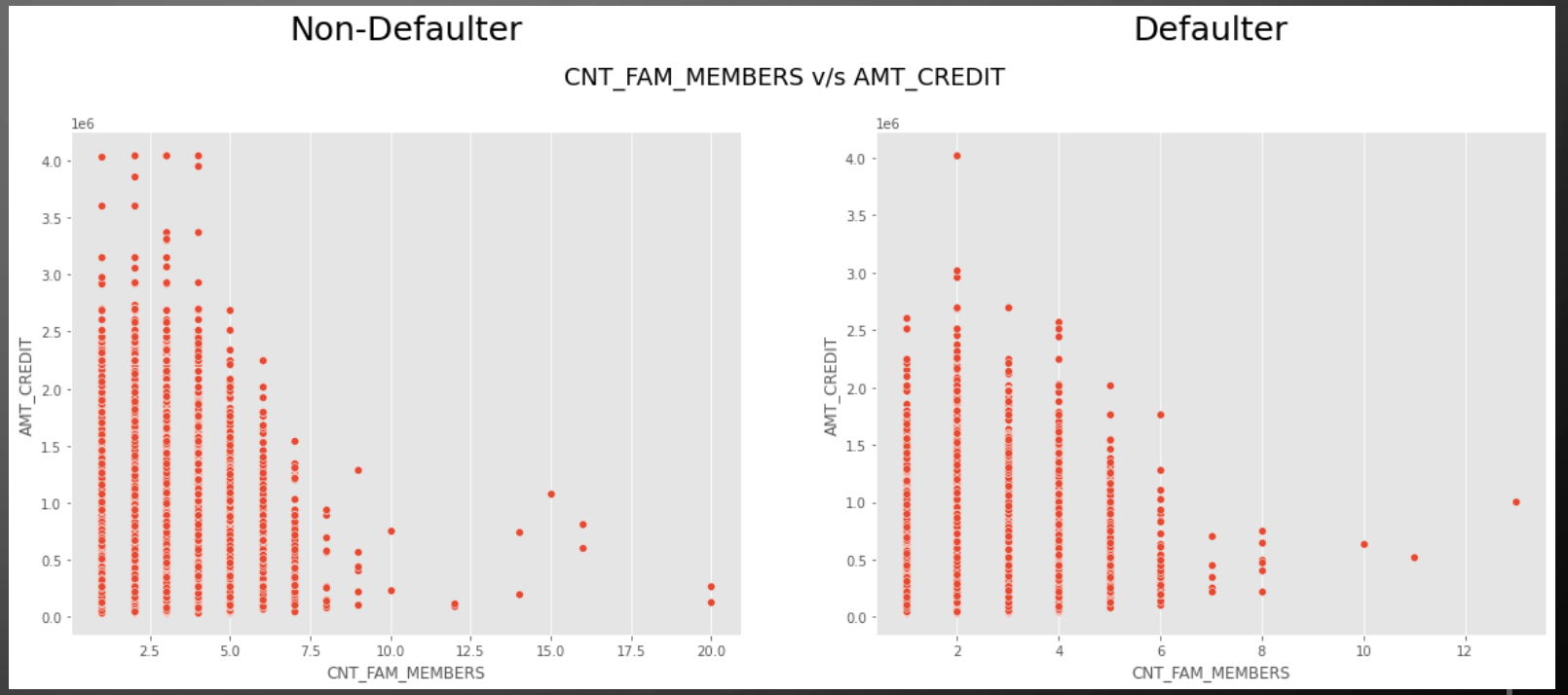


CNT_FAM_MEMBERS VS AMT_CREDIT

Few points can be concluded from the graph.

1. Applicants having larger family size and less Amount credit are likely to default less.

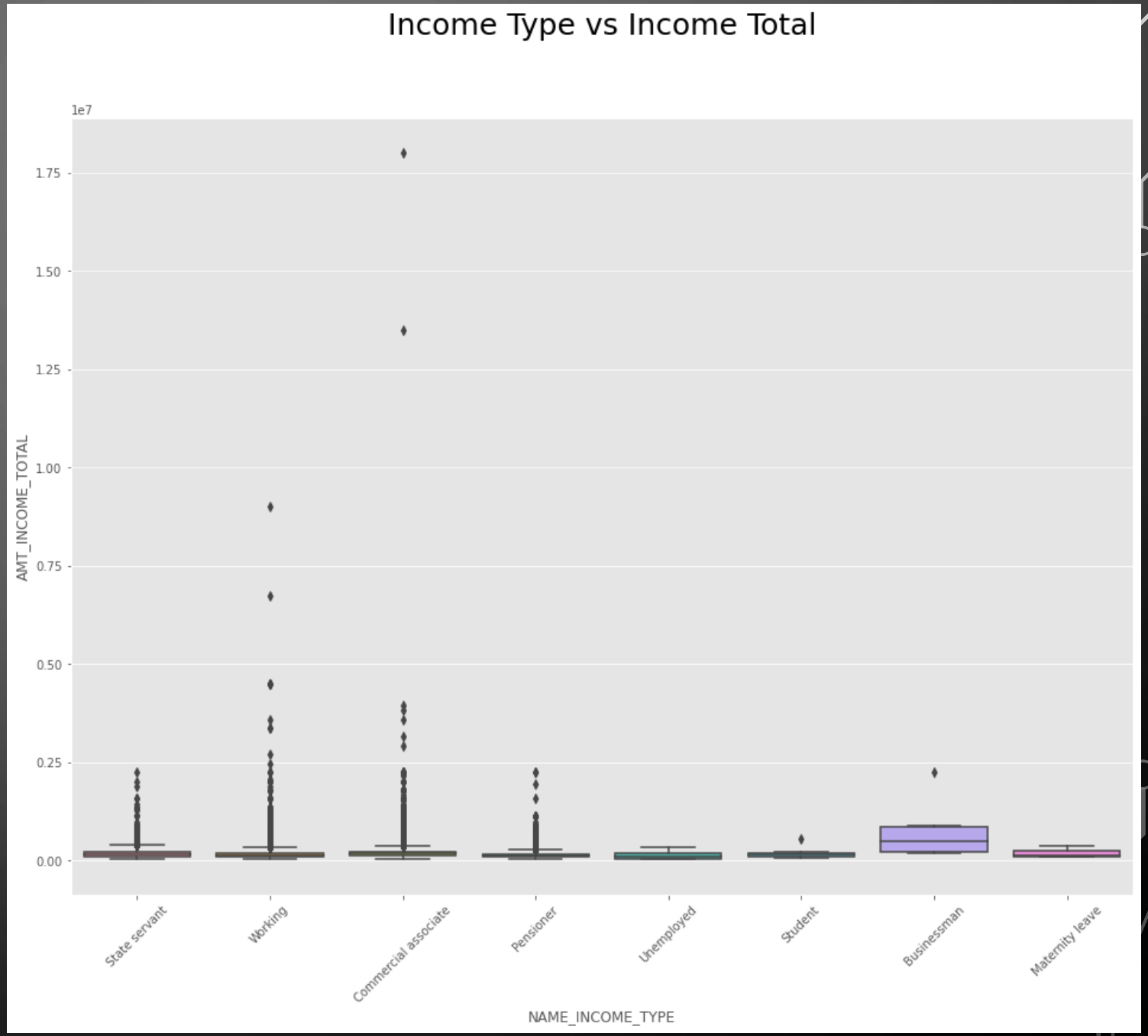
2. Applicants having smaller family size and higher Amount credit are likely to default less.



INCOME TYPE VS INCOME TOTAL

Few points can be concluded from the graph.

- It can be seen that business man's income is the highest.

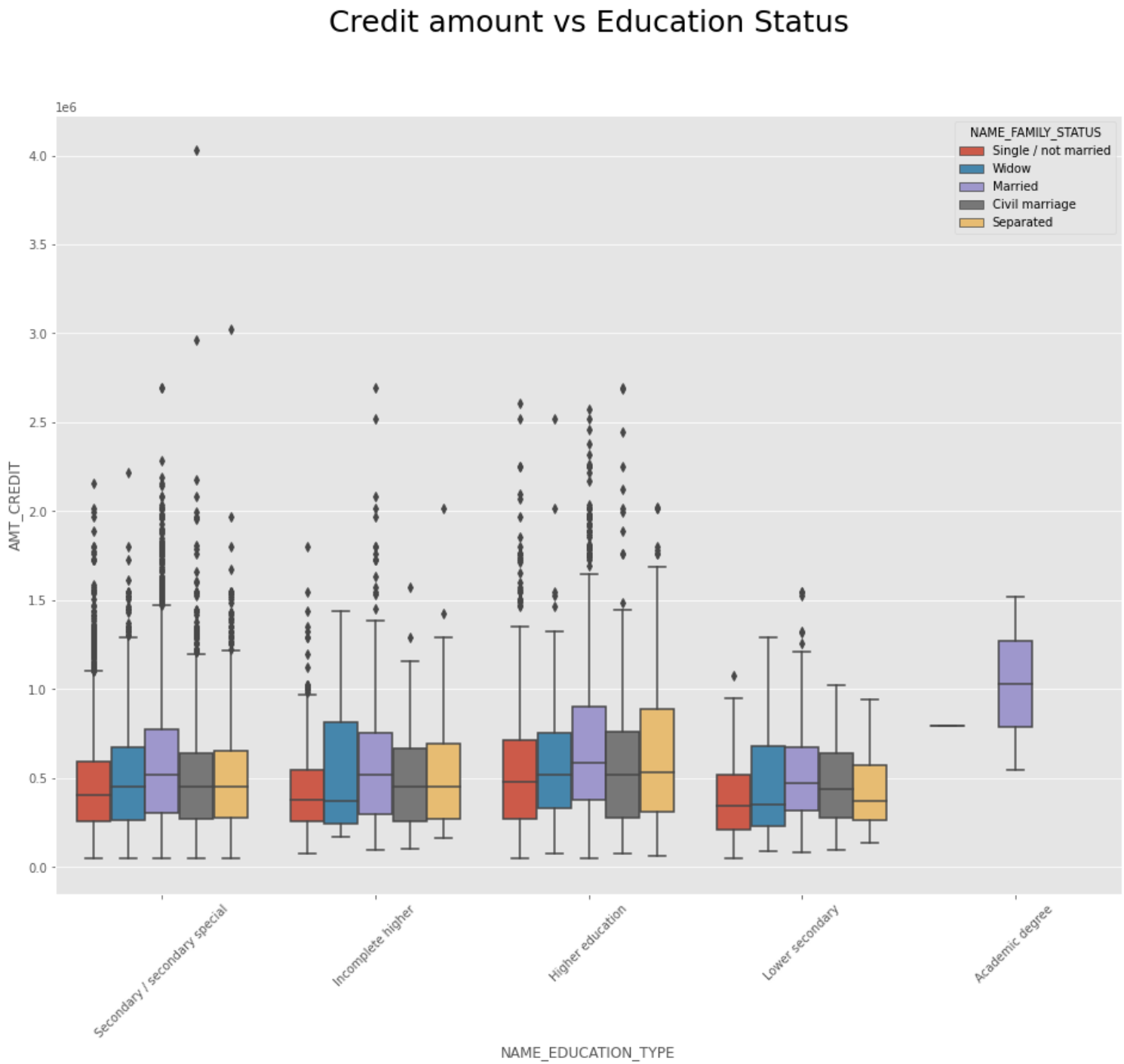


BIVARIATE ANALYSIS FOR TYPE 1

CREDIT AMOUNT VS EDUCATION STATUS

Few points can be concluded from the graph.

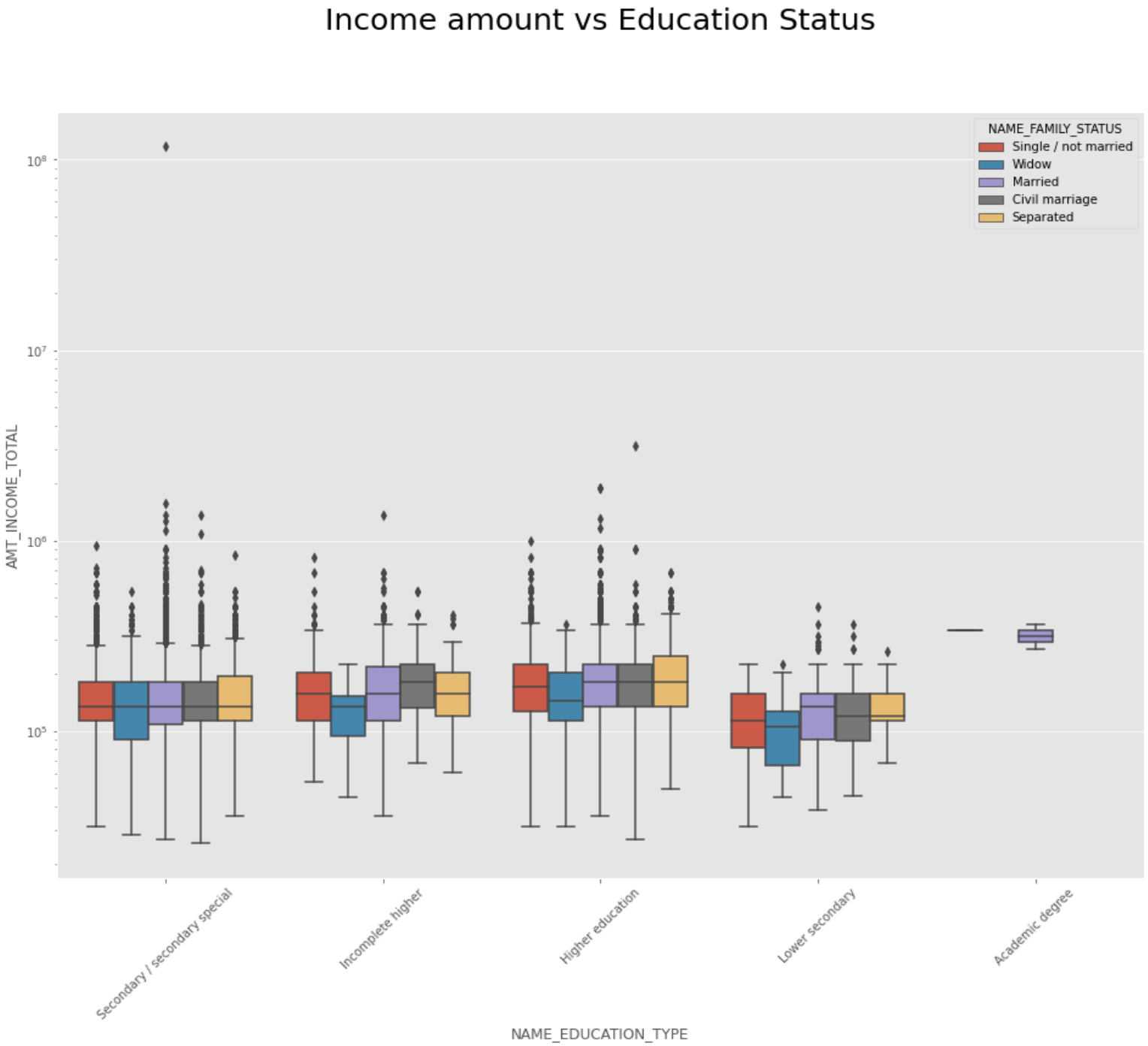
- Quite similar from Target 0, we can say that Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.
- Most of the outliers are from Education type 'Higher education' and 'Secondary'.
- Civil marriage for Academic degree is having most of the credits in the third quartile.



INCOME AMOUNT VS EDUCATION STATUS

Few points can be concluded from the graph.

- Have some similarity with Target0, From above boxplot for Education type 'Higher education' the income amount is mostly equal with family status.
- Less outlier are having for Academic degree but there income amount is little higher that Higher education.
- Lower secondary are have less income amount than others.

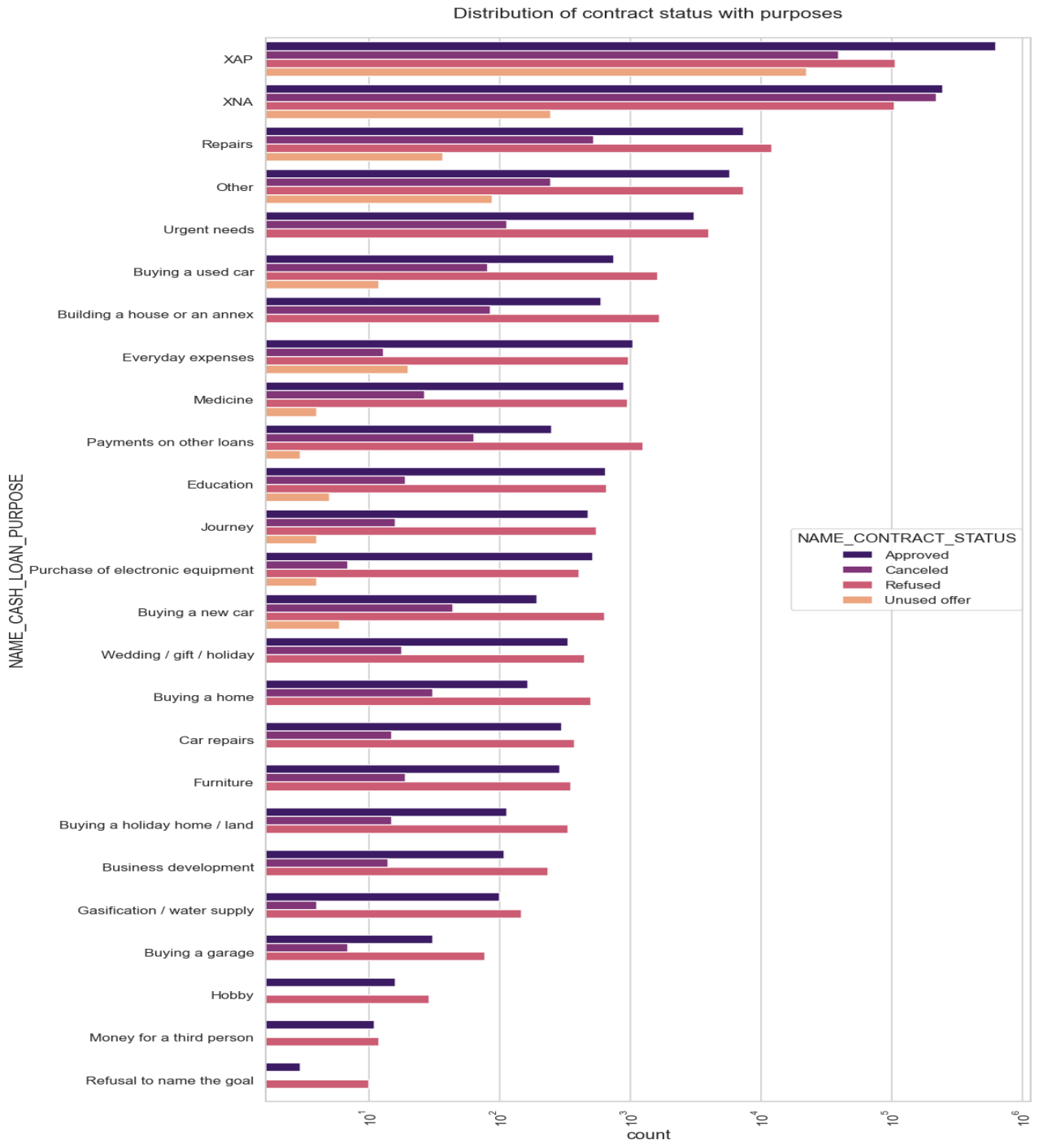


UNIVARIATE ANALYSIS AFTER MERGING PREVIOUS DATA

DISTRIBUTION OF CONTRACT STATUS WITH PURPOSES

Few points can be concluded from the graph.

- Most rejection of loans came from purpose 'repairs'.
- 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.

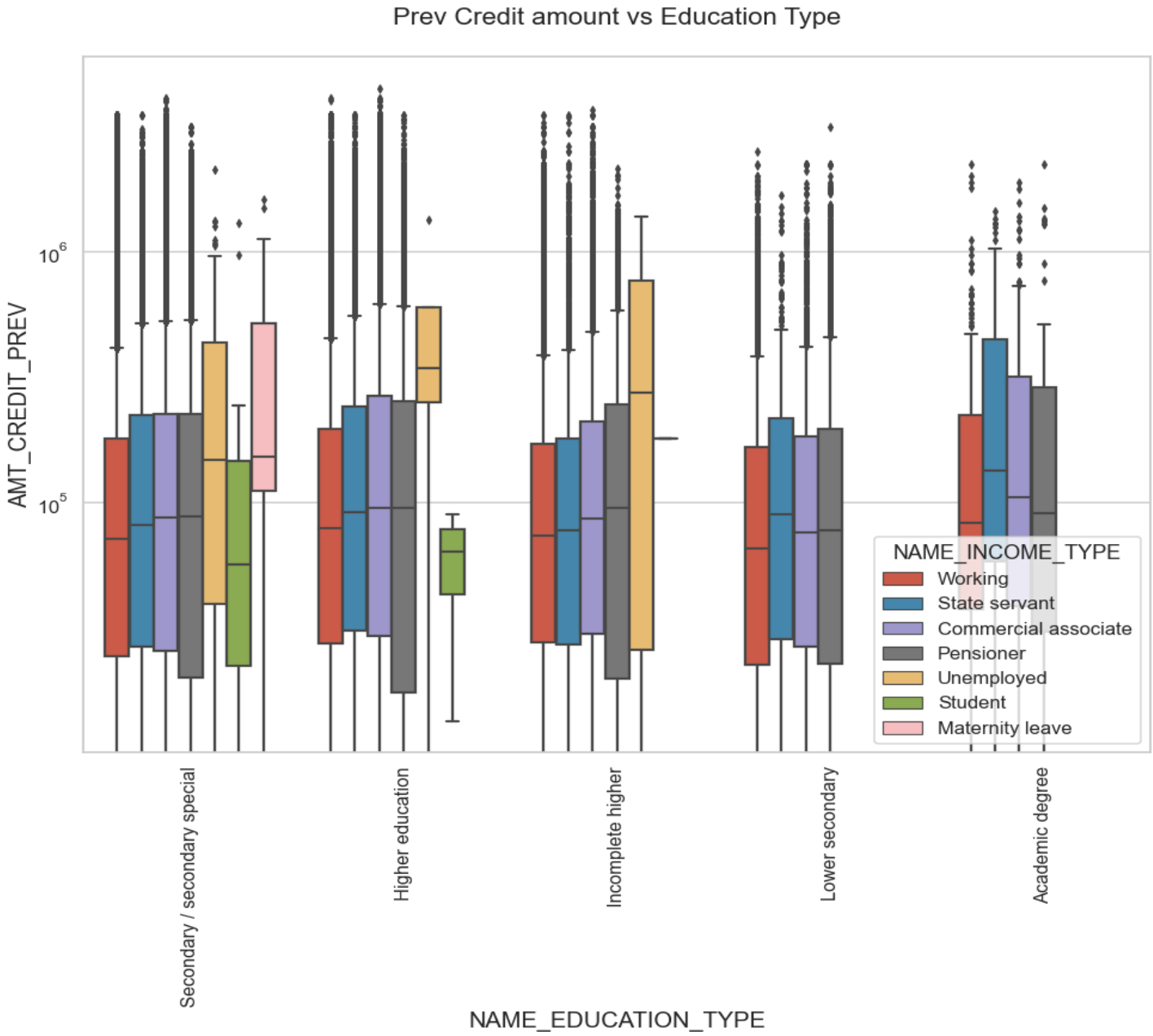


BIVARIATE ANALYSIS AFTER MERGING PREVIOUS DATA

DISTRIBUTION OF PREV_CREDIT ON THEIR EDUCATION TYPE

Few points can be concluded from the graph.

- 1. Here awe can see that people who have incompleted higher education and unemployed have more amount of previous loan credit.
- 2. Also client who is student and completed his Higher education have less amount of previous loan credit.



PERFORMING BIVARIATE ANALYSIS

CONCLUSION FOR APPROVING LOANS

- The rate of default of people having car is low compared to people who don't. So, bank should focus on customers who have cars.
- Student and Businessmen have no defaults.
- Customer who belongs to 'widow' or 'separated' category is less likely to default.
- There is no correlation between owning a reality and defaulting the loan.
- People with academic degree has less defaults.
- Female applicants have relatively lower default rate.

CONCLUSION FOR REJECTING LOANS

- We see that (25,30] age group tend to default more often. So, they are the riskiest people to loan to.
- Applicants who have higher family members (>=11) have higher default rate and their applications can be rejected.
- Applicants in civil marriage or who are single have higher default rate.
- Avoid young applicants who are in age group of 20-40 as they have higher probability of defaulting.
- Applicants who live in areas with Region Rating as 3 has highest defaults.
- Male applicants have relatively higher default rate.

THANK YOU !!