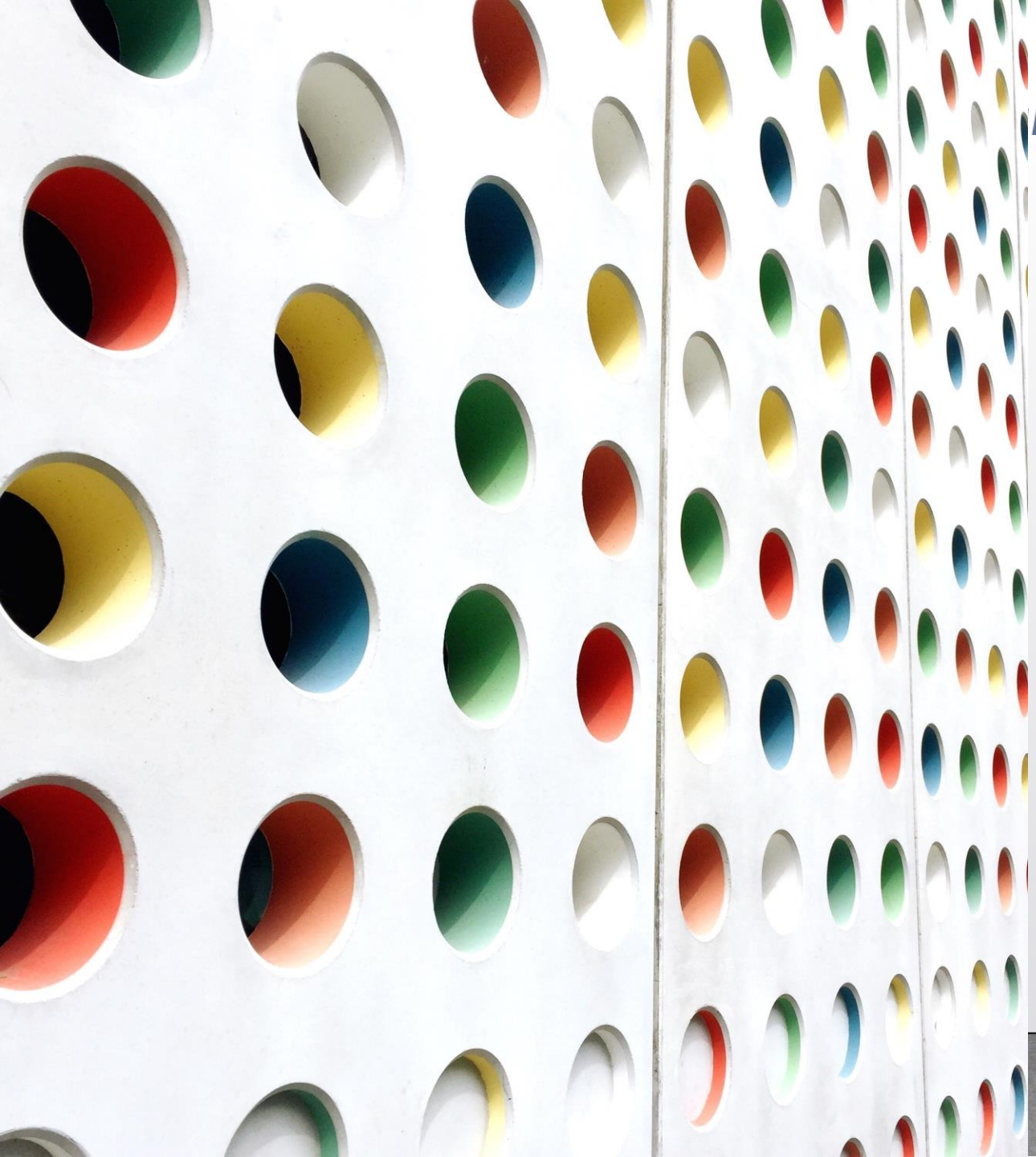
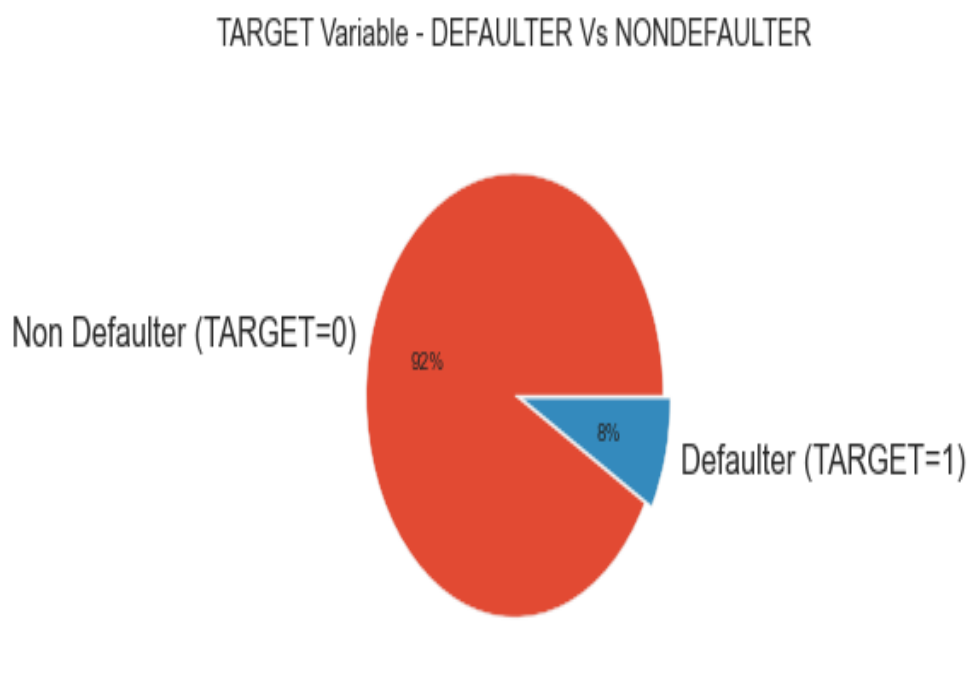


EDA Credit Assignment

Kushagra Saxena: DS-c40

Target Variable

- There is an imbalance between number of defaulter and non defaulters. More than 91% of people didn't default as opposed to 8% who defaulted.



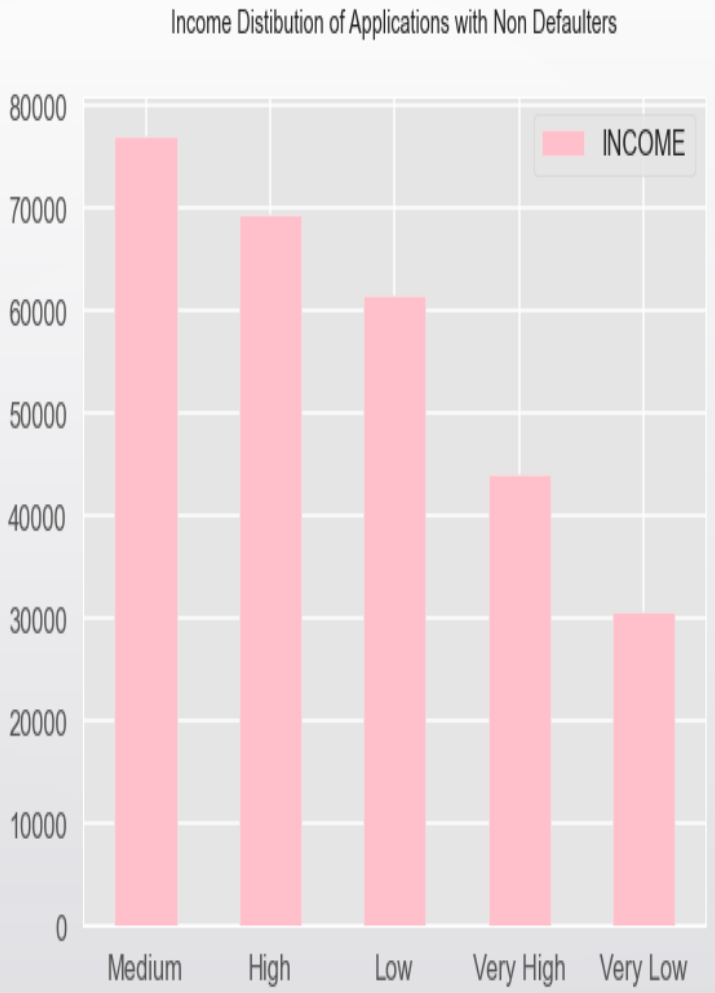
UNIVARIATE ANALYSIS

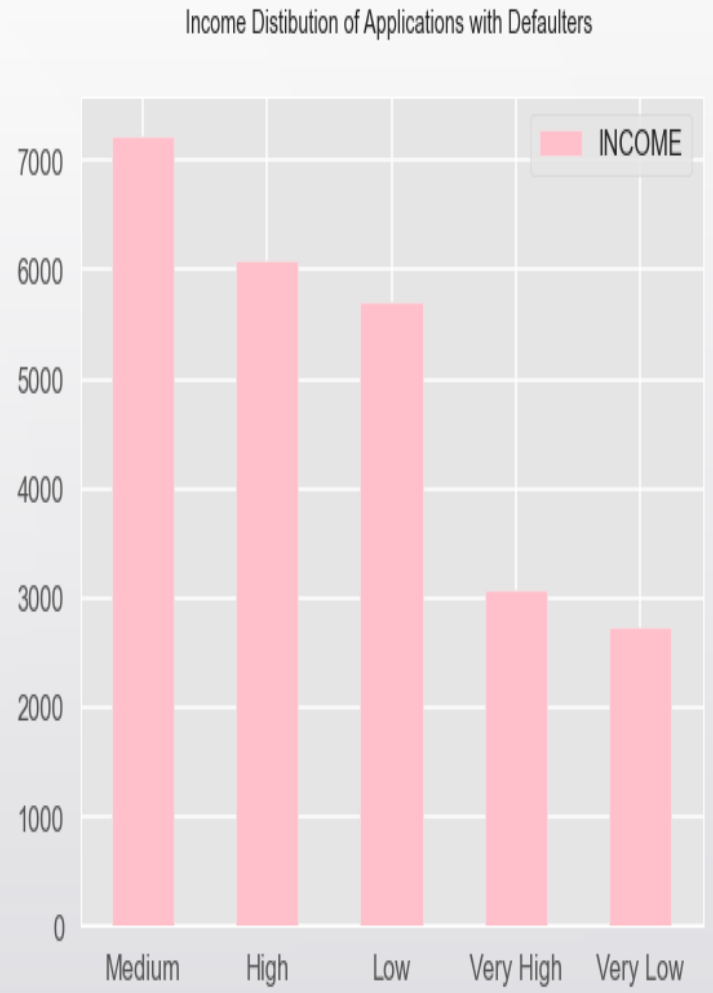
Analyses of INCOME Column

Inference:

1. Medium Salary people in both defaulters and non defaulters are highest.

2. Very low category of income people do not get much loans from the bank.



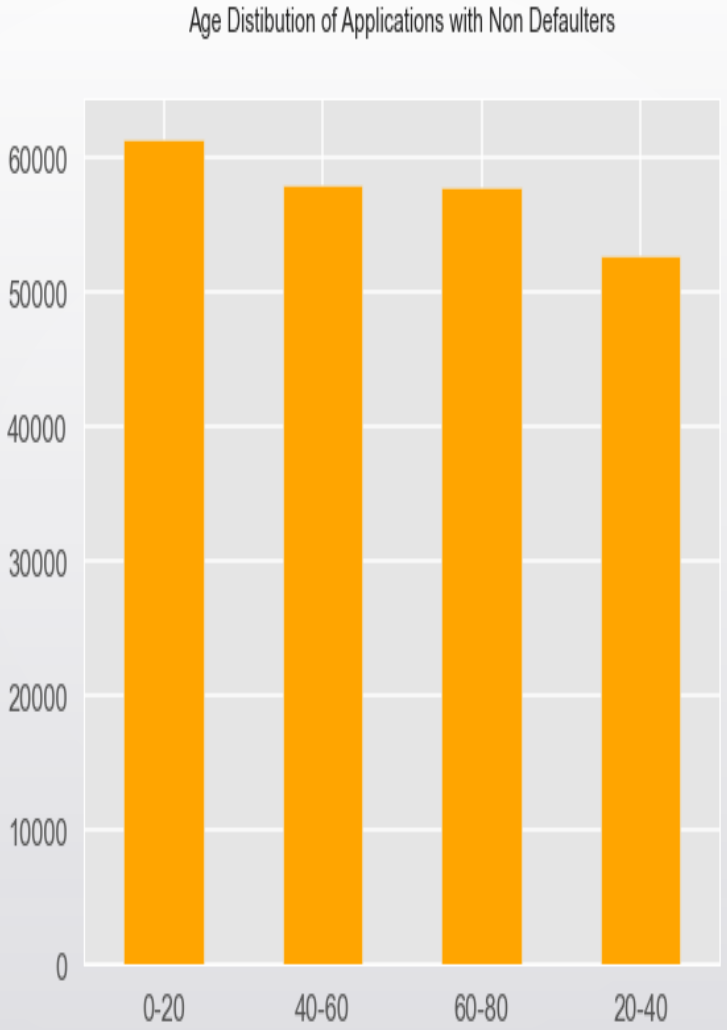


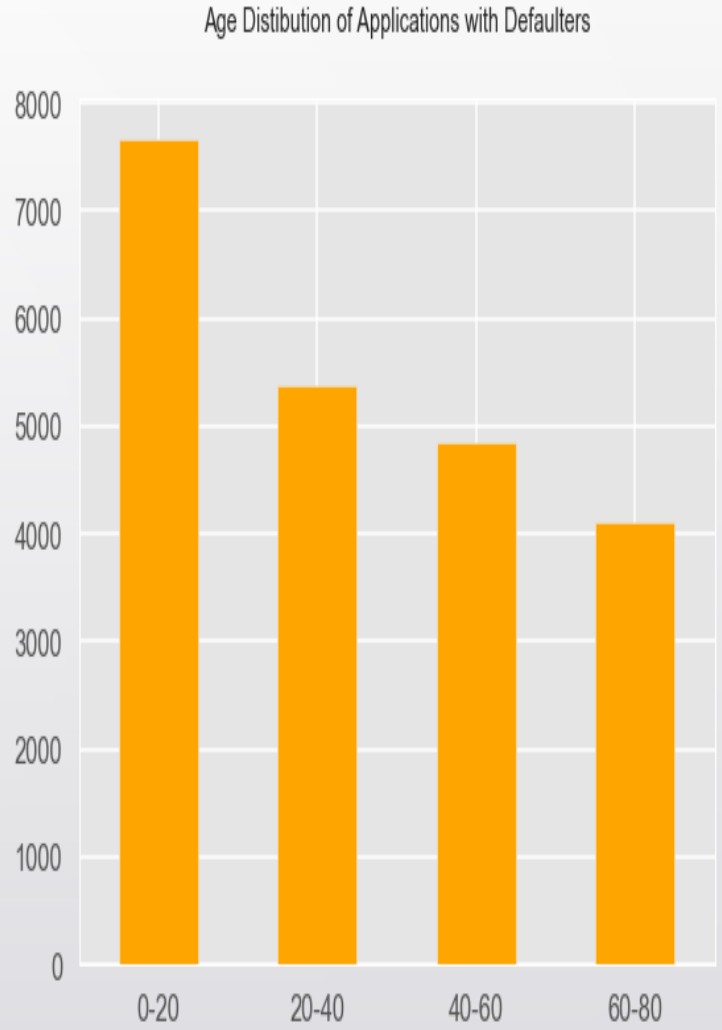
Analyses of INCOME Column

Inference:

1. Below 20 people has the highest count in defaulter and non defaulters. Hence, they are mostly students and have no source of income. They may take loans for studies or something else.

2. Interesting thing is that age group of 20-40 are lowest in Non defaulter category and 60-80 are lowest in defaulter category. hence, 60-80 people can be targeted by the bank for loans.

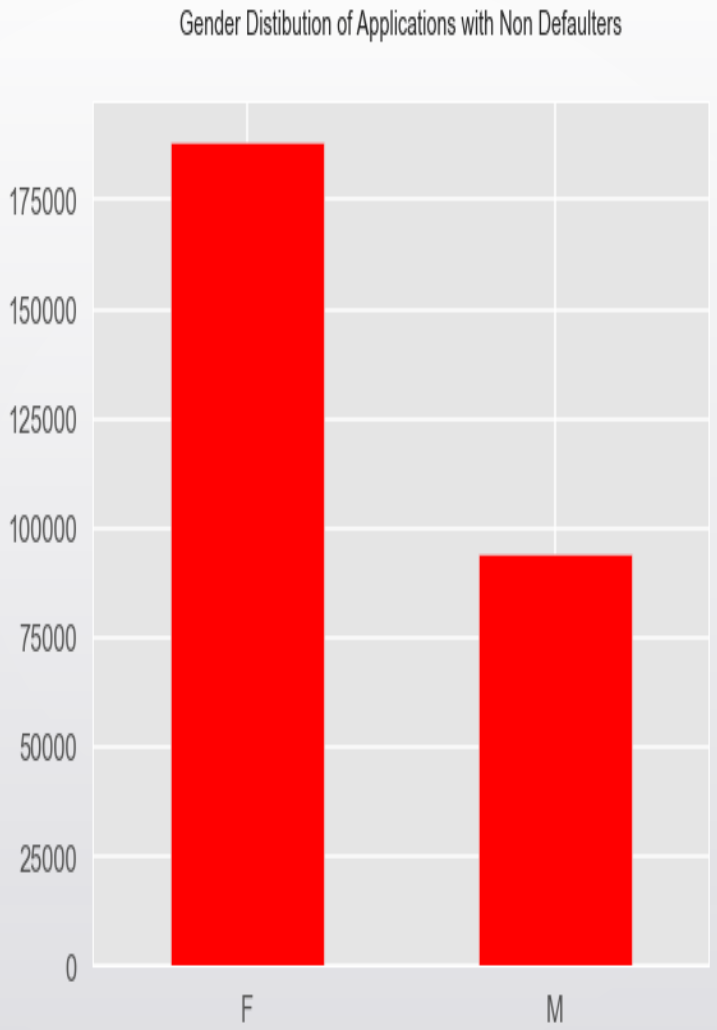


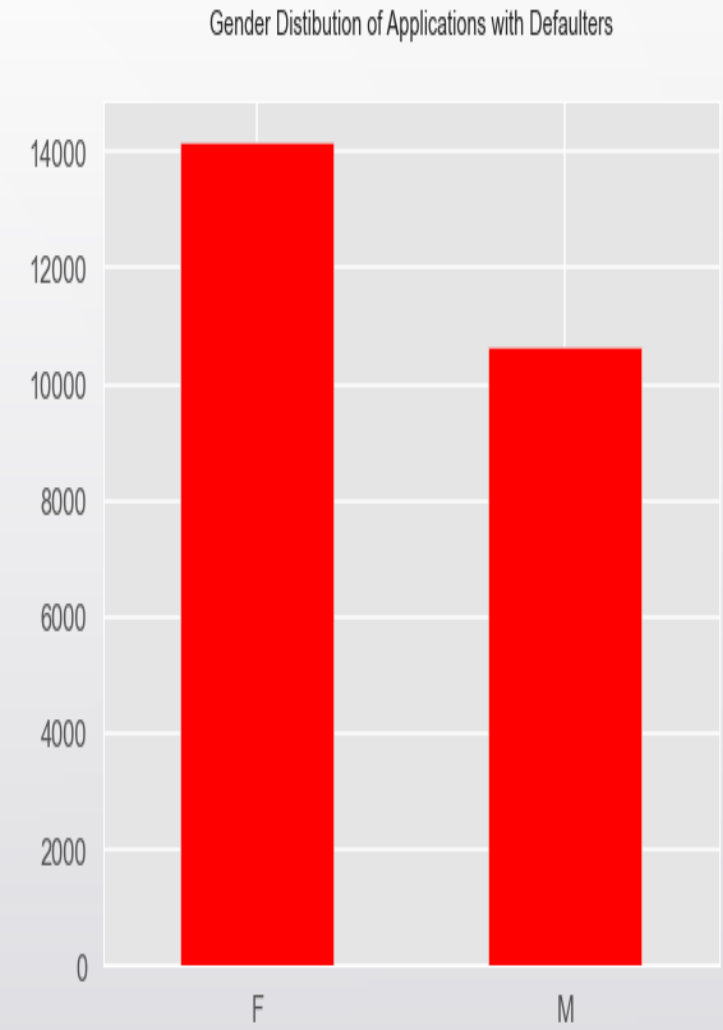


Analyses of GENDER_CODE Column

Inference:

1. Females are more in both the graph. They apply for loans more as compared to men.

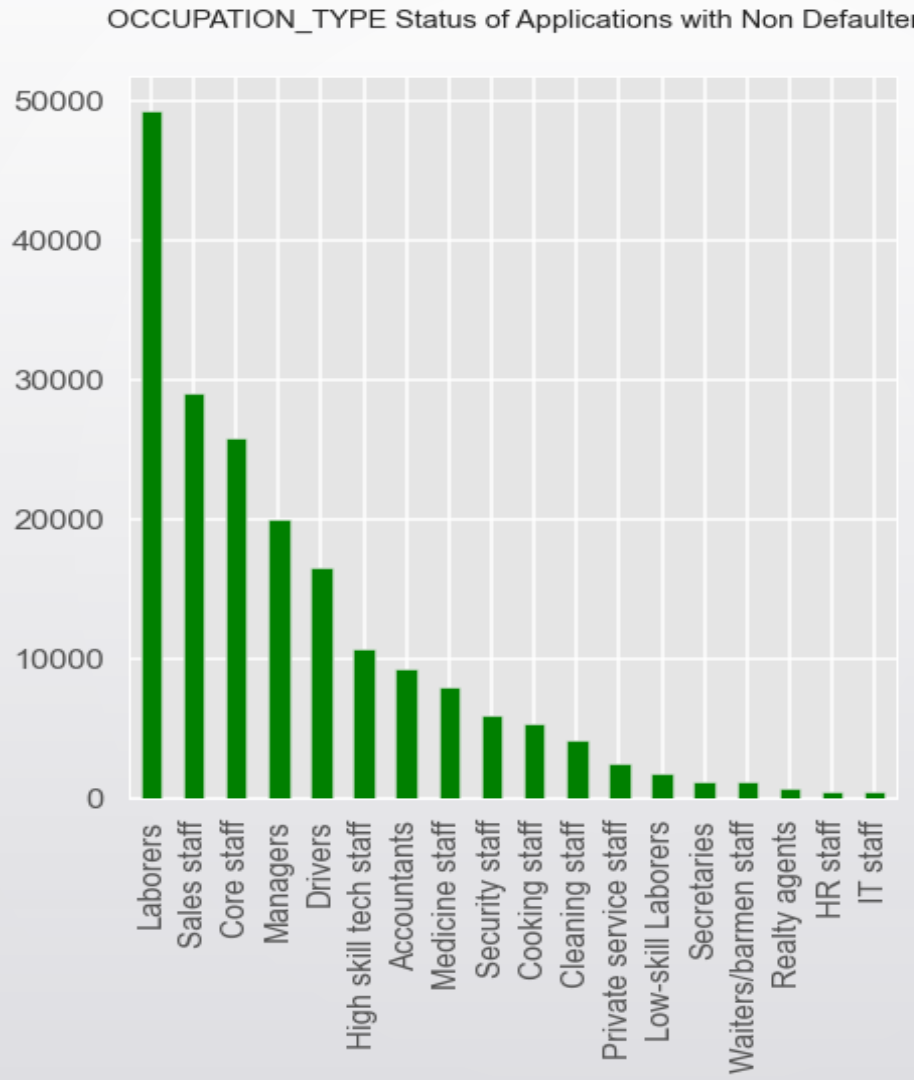


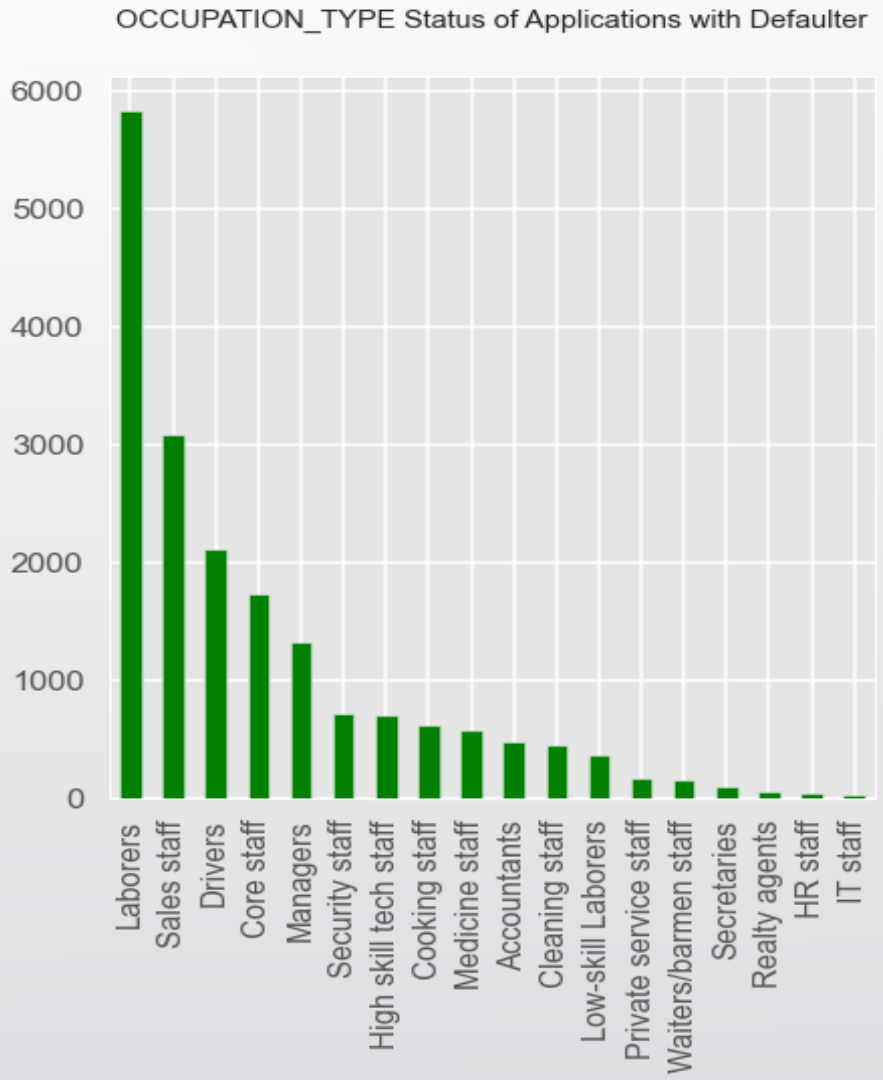


Analyses of OCCUPATION_TYPE Column

Inference:

1. Laborers, Sales Staff, Core Staff, Managers are the highest and contributes to 50 % of the data in Defaulters and Non Defaulters.

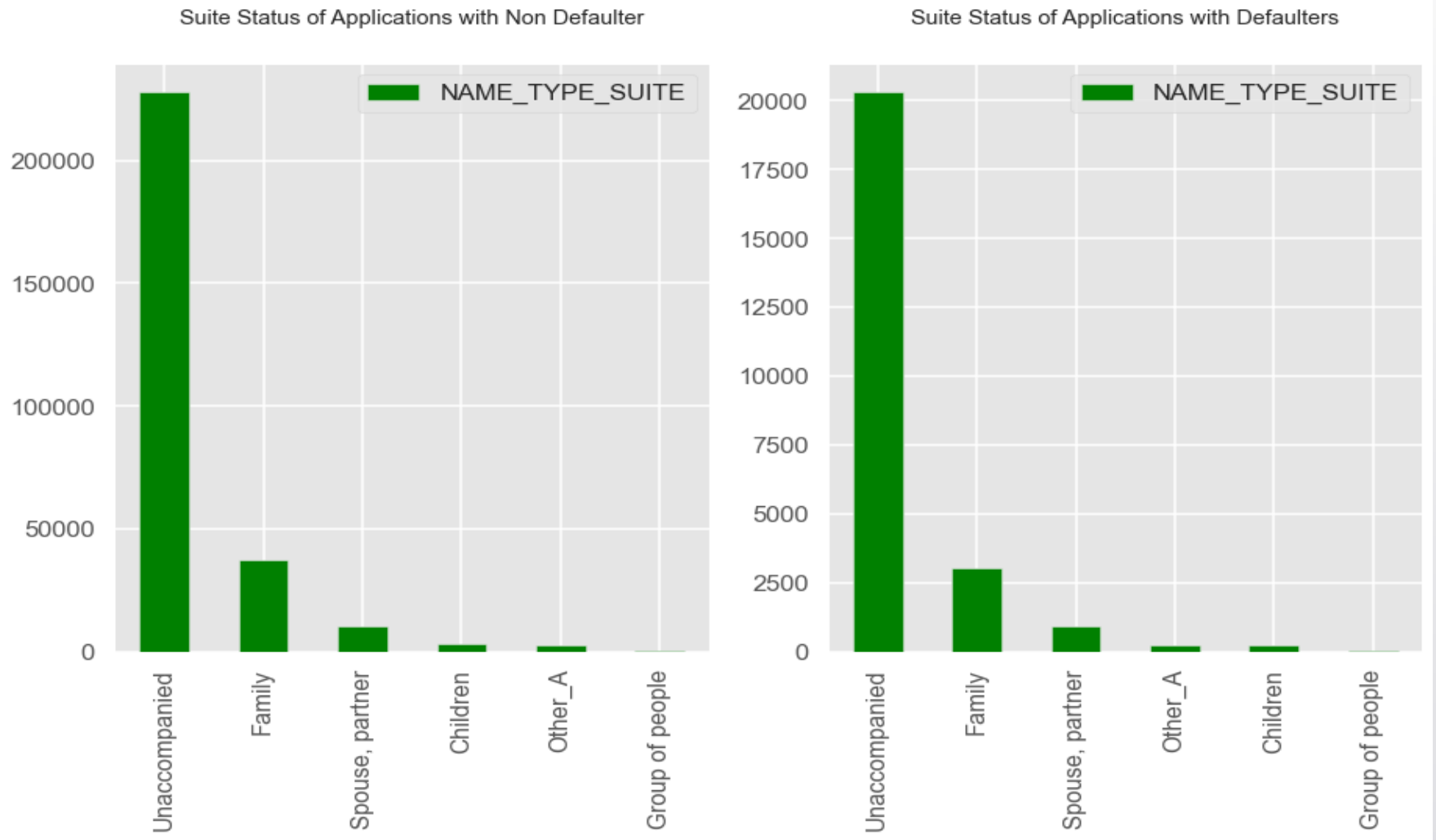




Analyses of NAME_TYPE_SUITE Column

Inference:

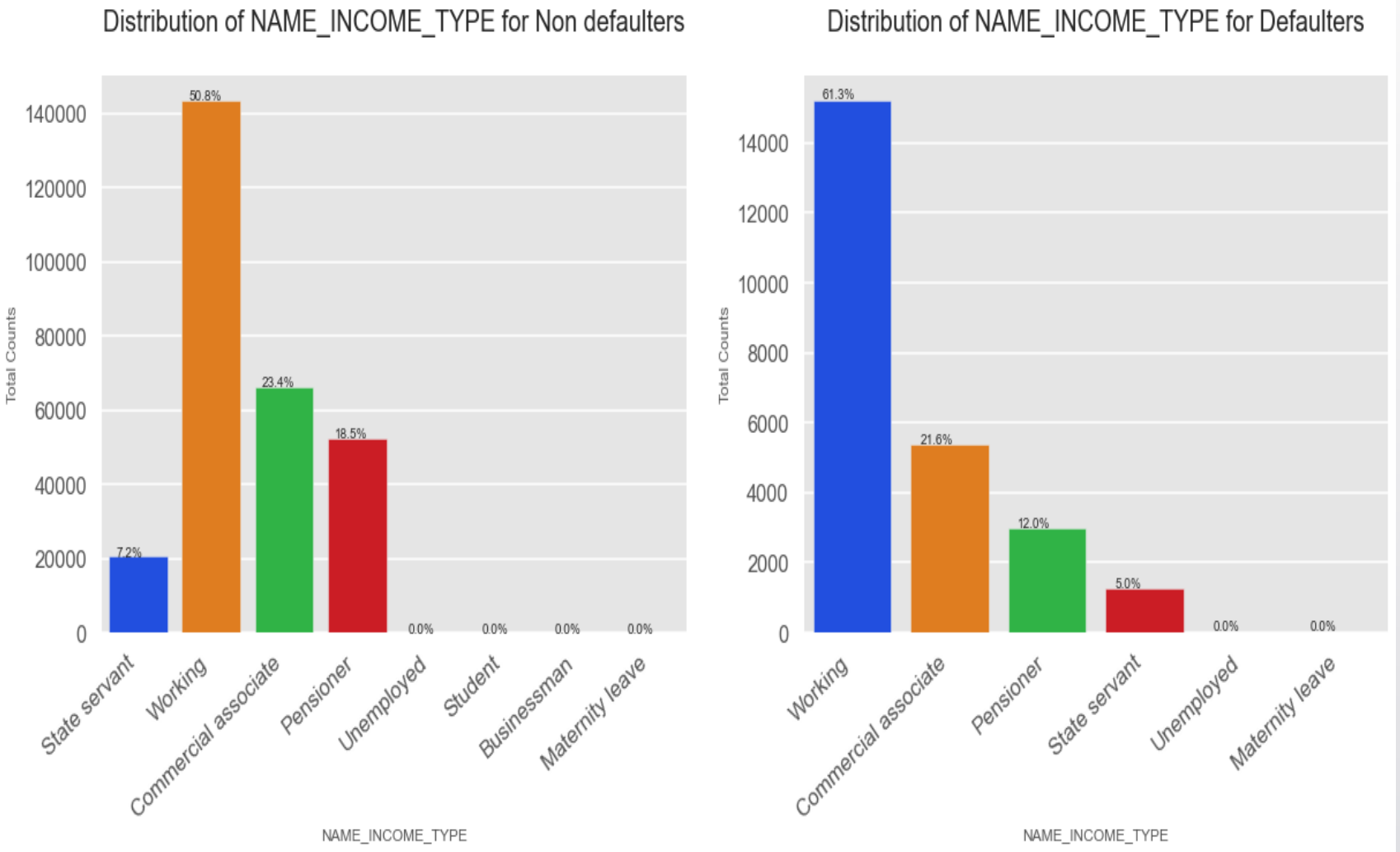
Mostly, there are Loan Unaccompanied in both the data.



Analyses of NAME_FAMILY_STATUS Column

Inference:

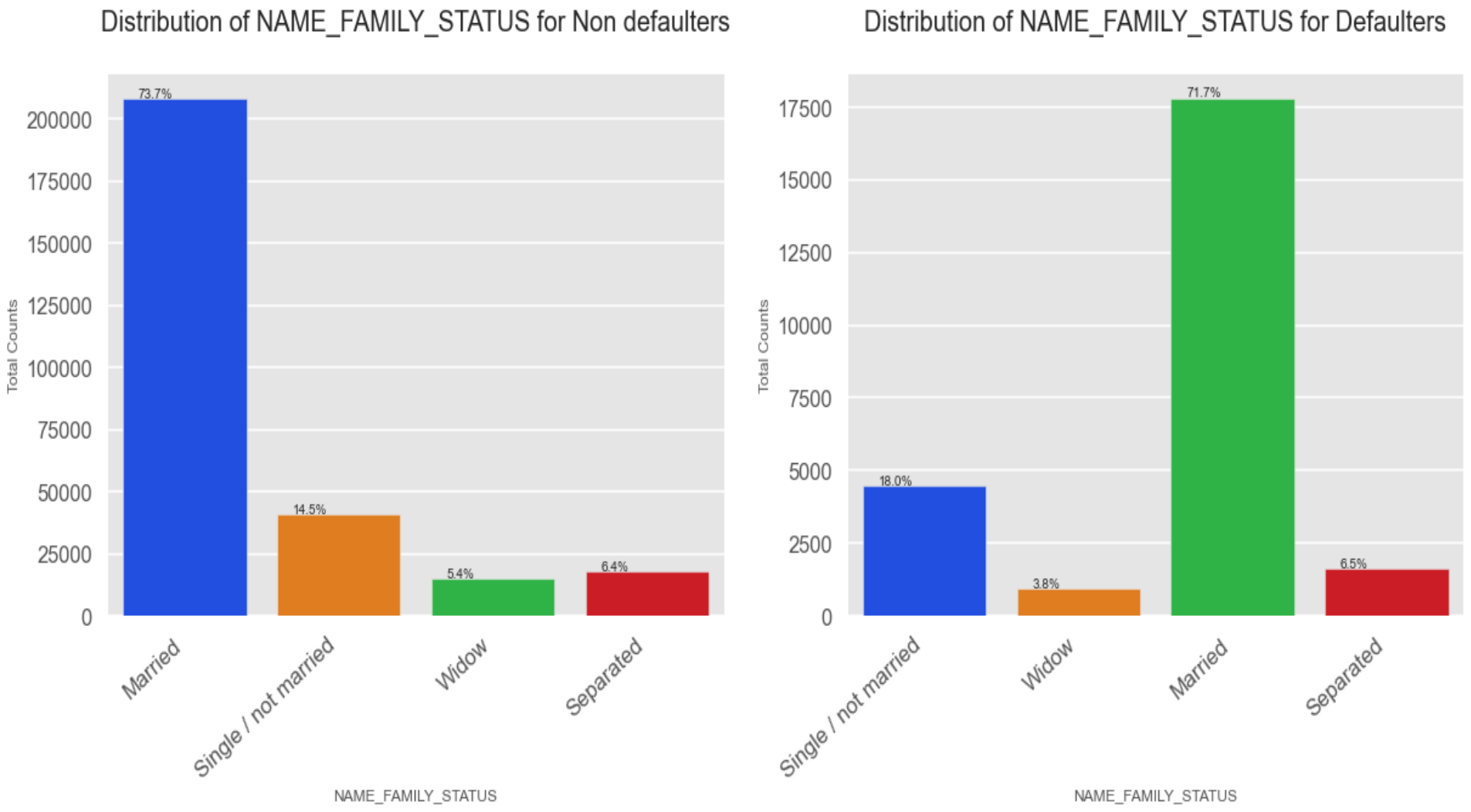
1. Working professionals contribute to 50 % in Non defaulters and 61 % in defaulters.



Analyses of NAME_FAMILY_STATUS Column

Inference:

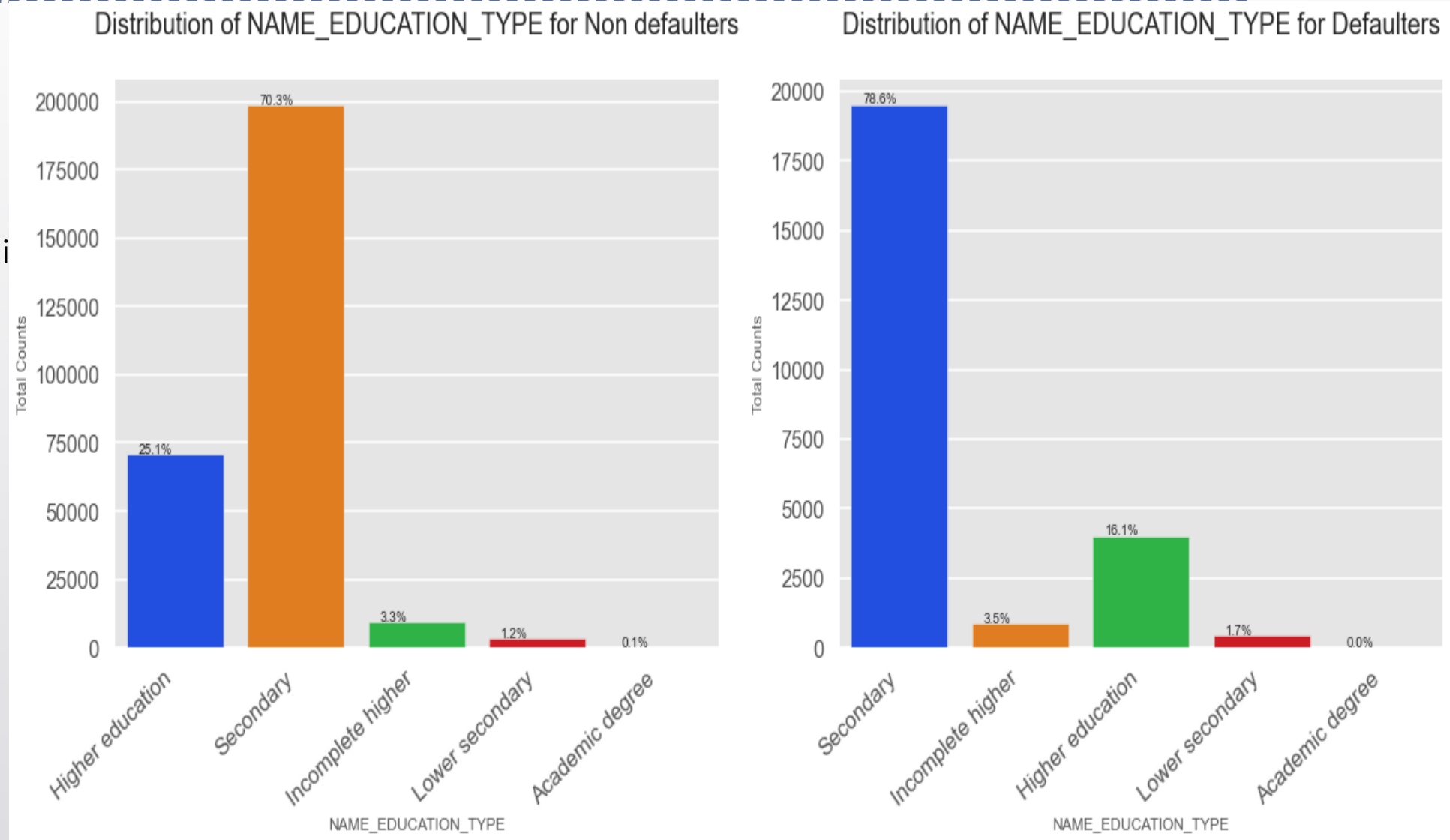
1. Married people, 71 % have Defaulted for payments.



Analyses of NAME_EDUCATION_TYPE Column

Inference:

1. Secondary Education people have applied for loans and further analysis i needed for conclusion

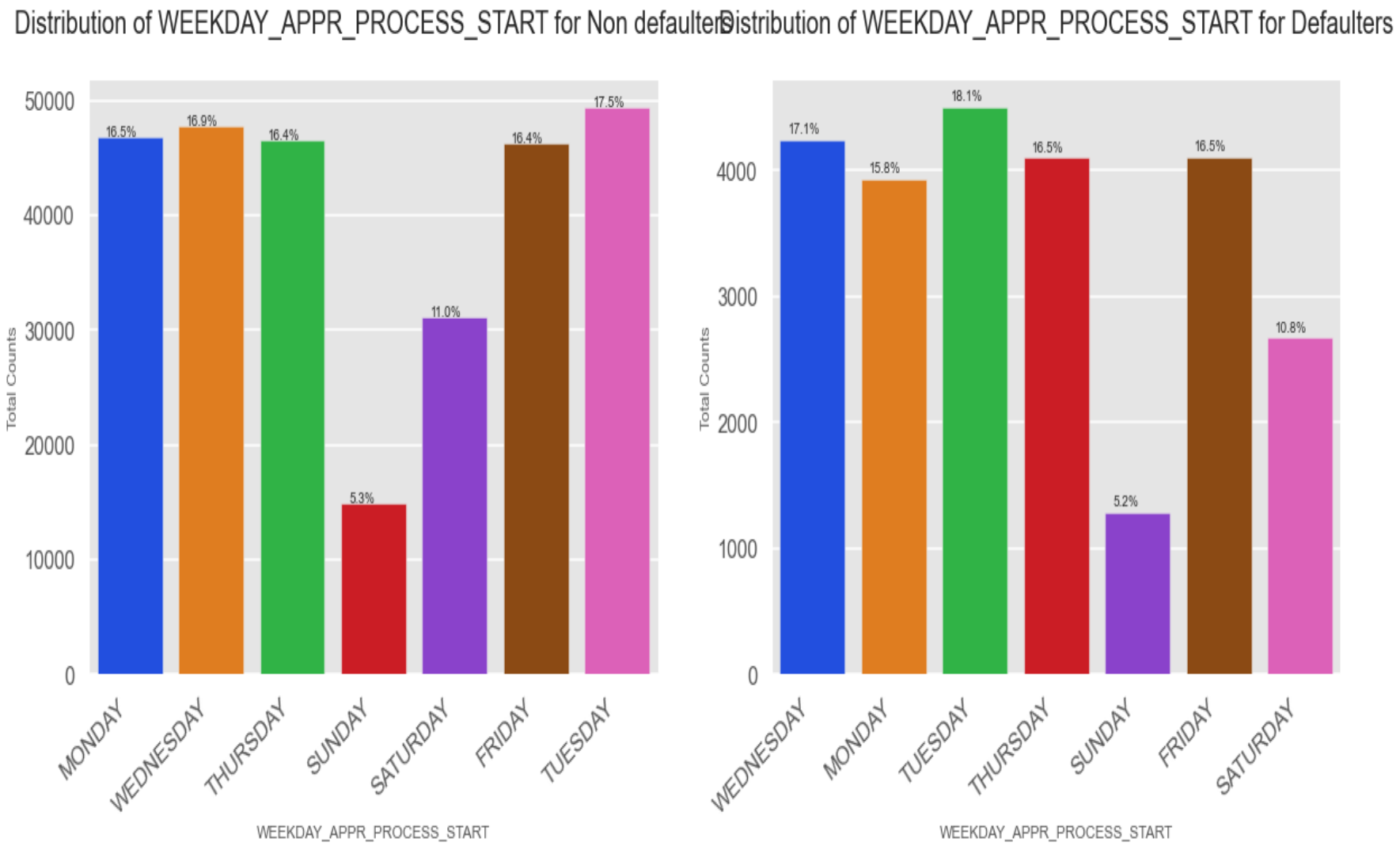


Analyses of WEEKDAY_APPR_PROCESS_START Column

Inference:

1. Tuesday records the highest number of applicants

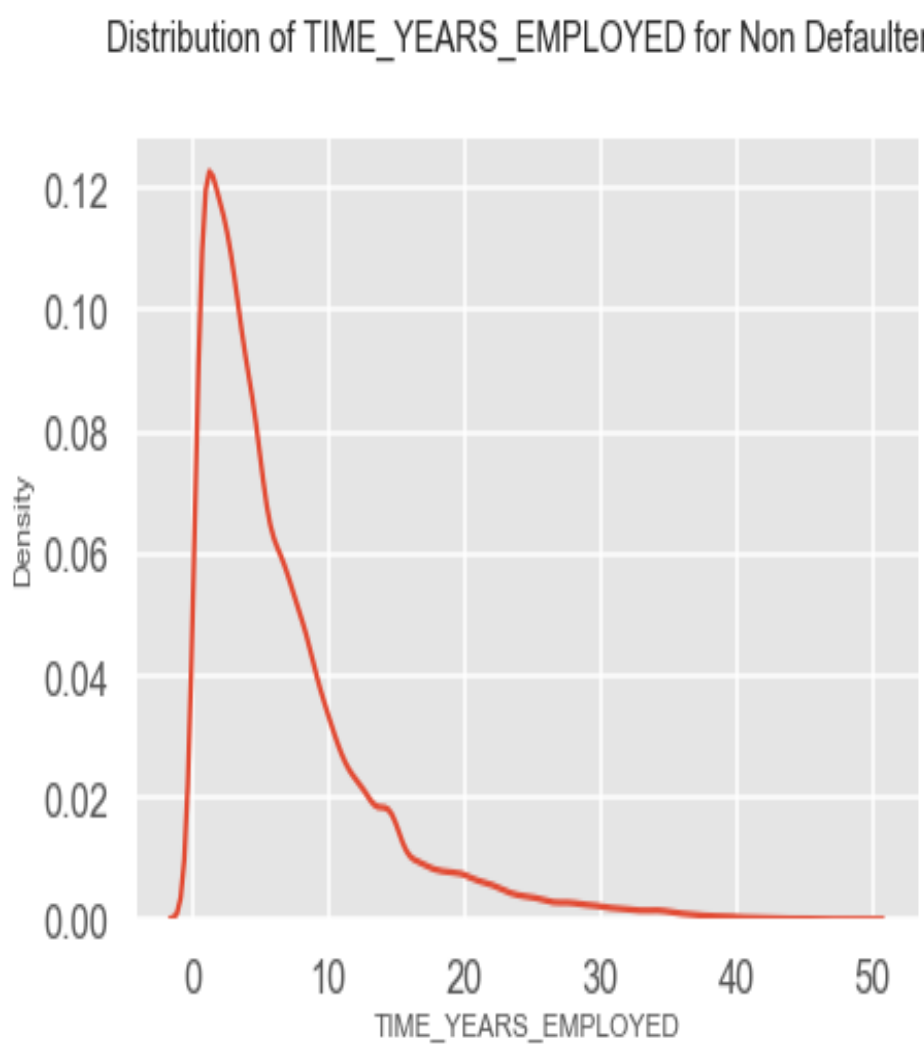
2. Sunday is the lowest

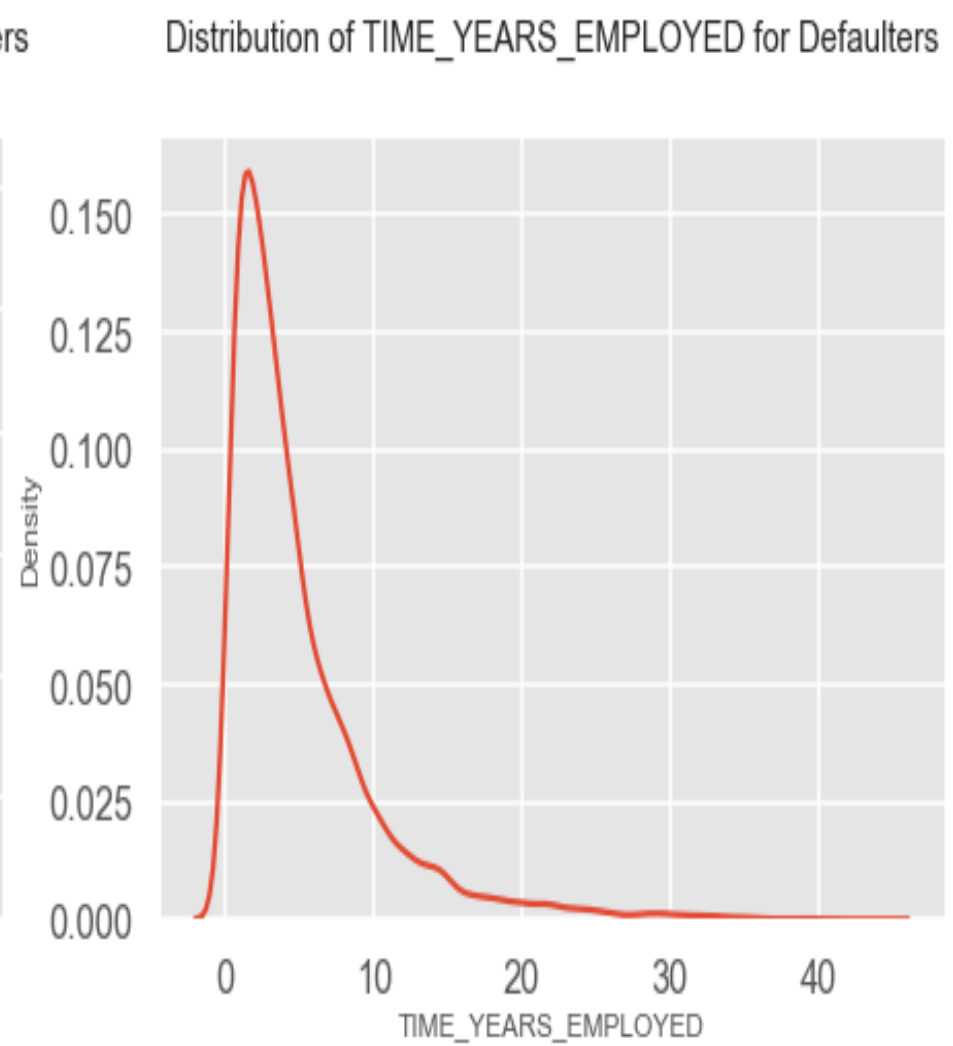


Analyses of TIME_YEARS_EMPLOYED Column

Inference:

People who have experienced bracket of 0-10 years are mostly applying for loans





Analyses of AGE Column

Inference:

1. This is an important finding that there is a sharp decline in Defaulters in age group 40-60.

2. The curve in Non Defaulter seemed to be less sharp.

BIVARIATE ANALYSIS

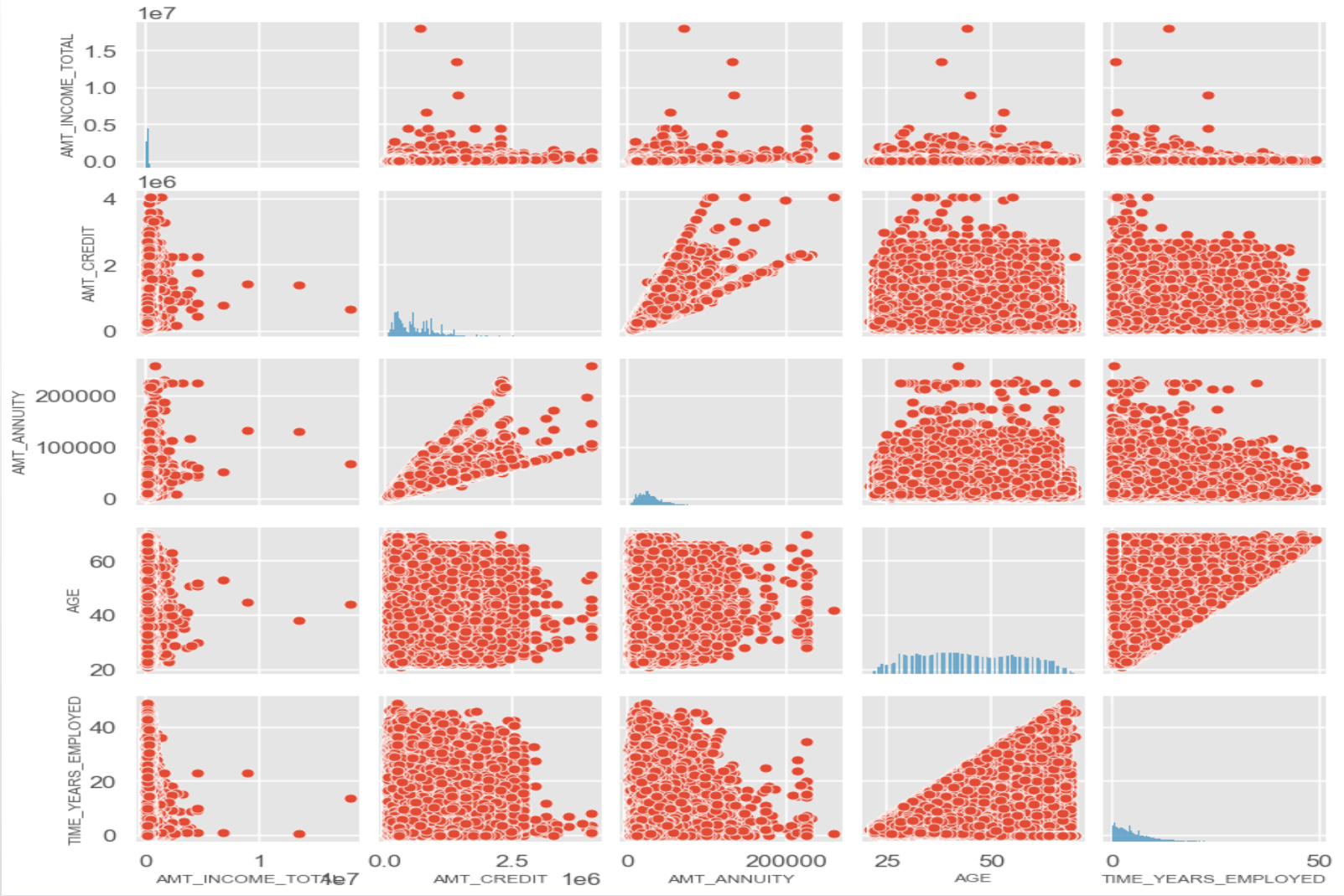
Plotting AMT_INCOME_TOTAL, AMT_CREDIT, AMT_ANNUITY, AGE, TIME_YEARS_EMPLOYED columns for Non Defaulters

Inference:

1. Amount Credit and Amount Annuity shows a strong relation between each other.

2. Amount Annuity decreases with the number of years of experience.

3. Amount Annuity becomes constant after a certain age.



Plotting AMT_INCOME_TOTAL, AMT_CREDIT, AMT_ANNUITY, AGE, TIME_YEARS_EMPLOYED columns for Defaulters

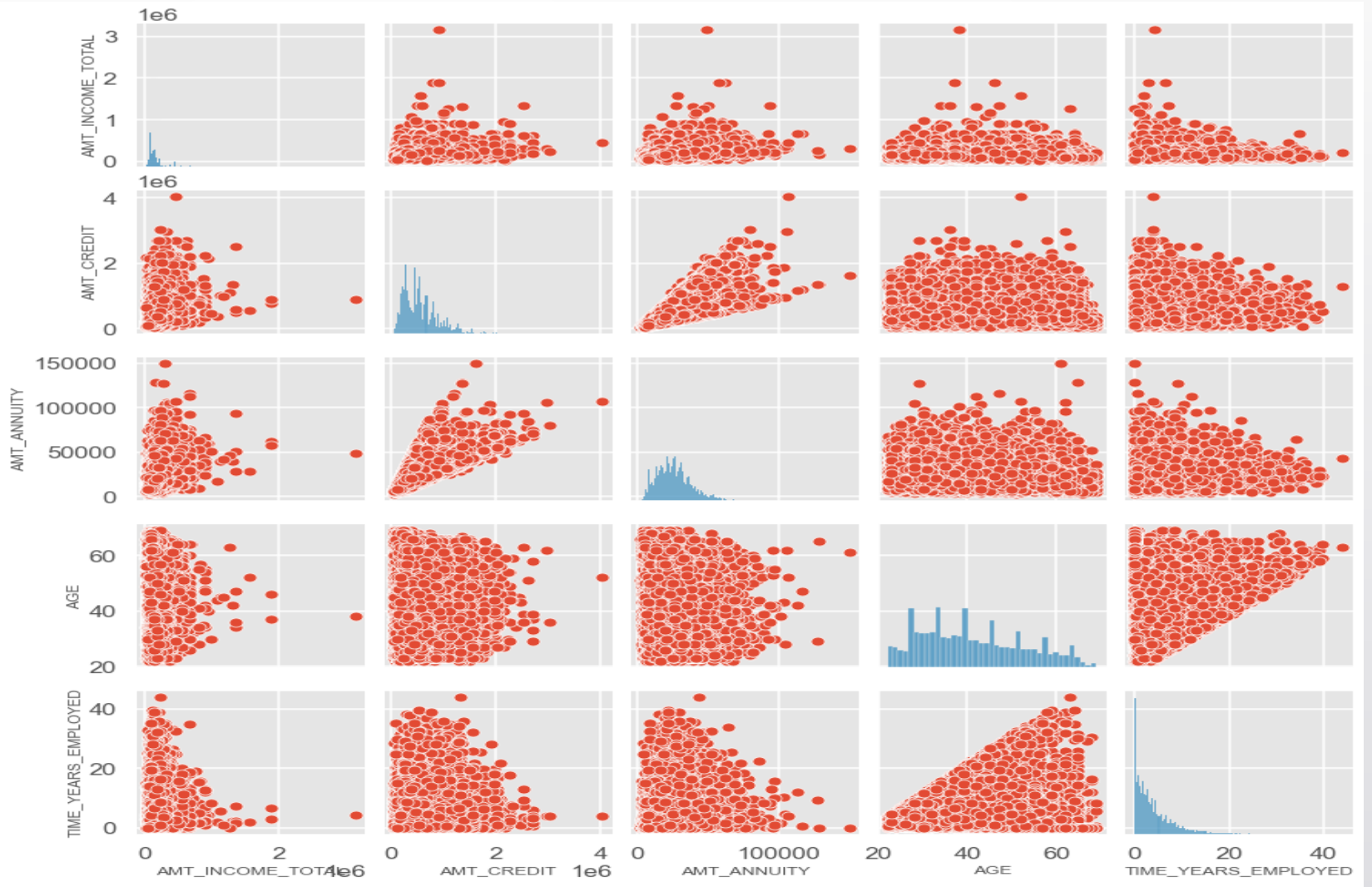
Inference:

Inference:

1. Amount Credit and Amount Annuity shows a strong relation between each other.

2. Amount Annuity decreases with the number of years of experience.

3. Amount Credit is decreased as the age and expereice increases and tend to be higher in the age group of 20-40 years



Plotting NAME_EDUCATION_TYPE vs AMT_CREDIT Column

Inference: Points to be drawn from the graph above for (Non-Defaulters).

1. Customers with an academic degree have a higher credit limit, with c married being the most common.

2. Customers with less education have lower credit limits, with widows having the lowest.

3. Married clients have a larger credit amount in practically all schooling segments except lower secondary and academic degrees.

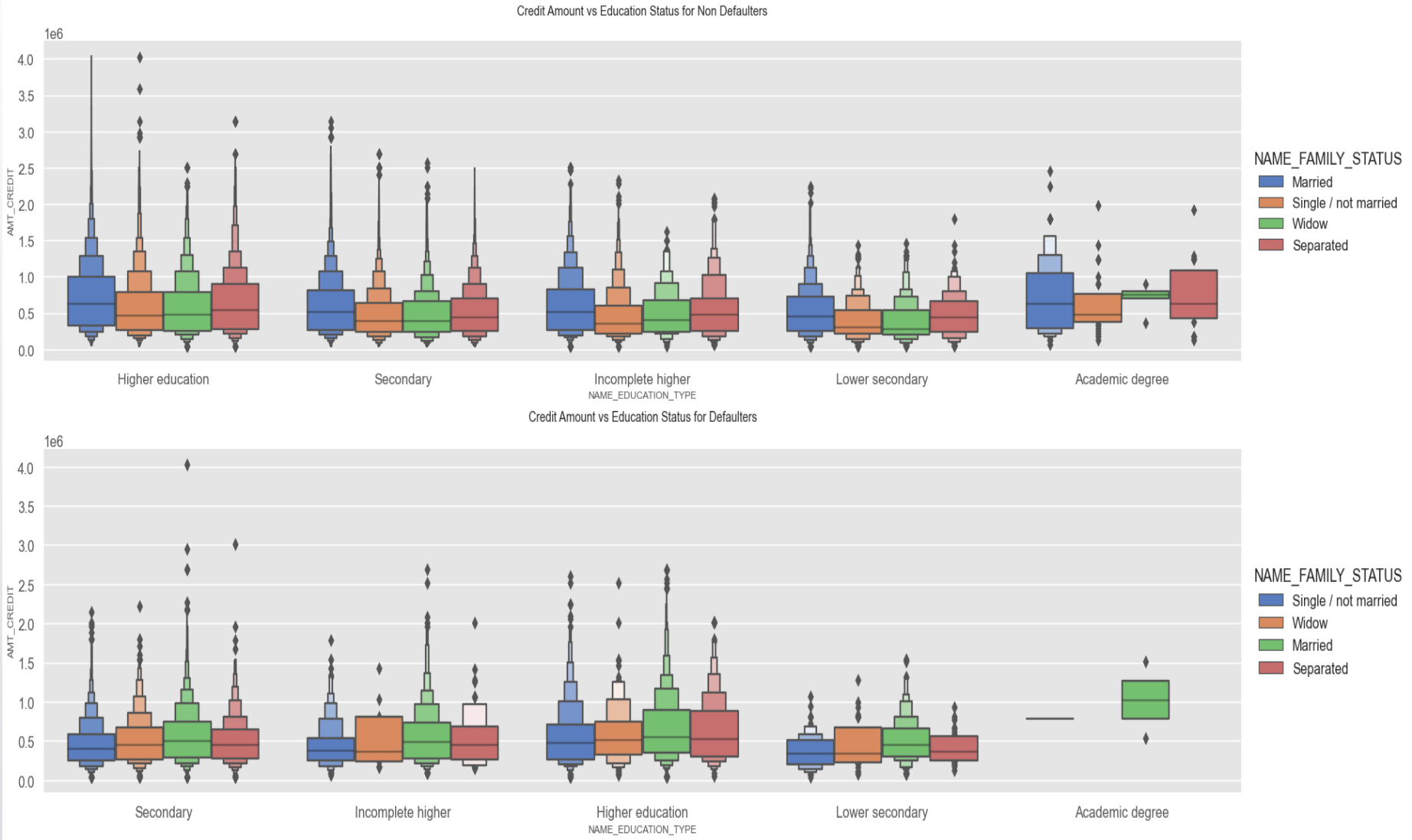
Conclusions to be drawn from the preceding graph for (Defaulters).

1. Customers with a married academic degree have a greater credit limit and, as a result, a higher default rate.

2. Married customers have larger credit amounts across all academic segments.

3. Customers with a lesser eductation have a lower credit limit.

4. The only two family types represented in academic degrees are single and married.



plotting AMT_CREDIT prev vs NAME_HOUSING_TYPE Column

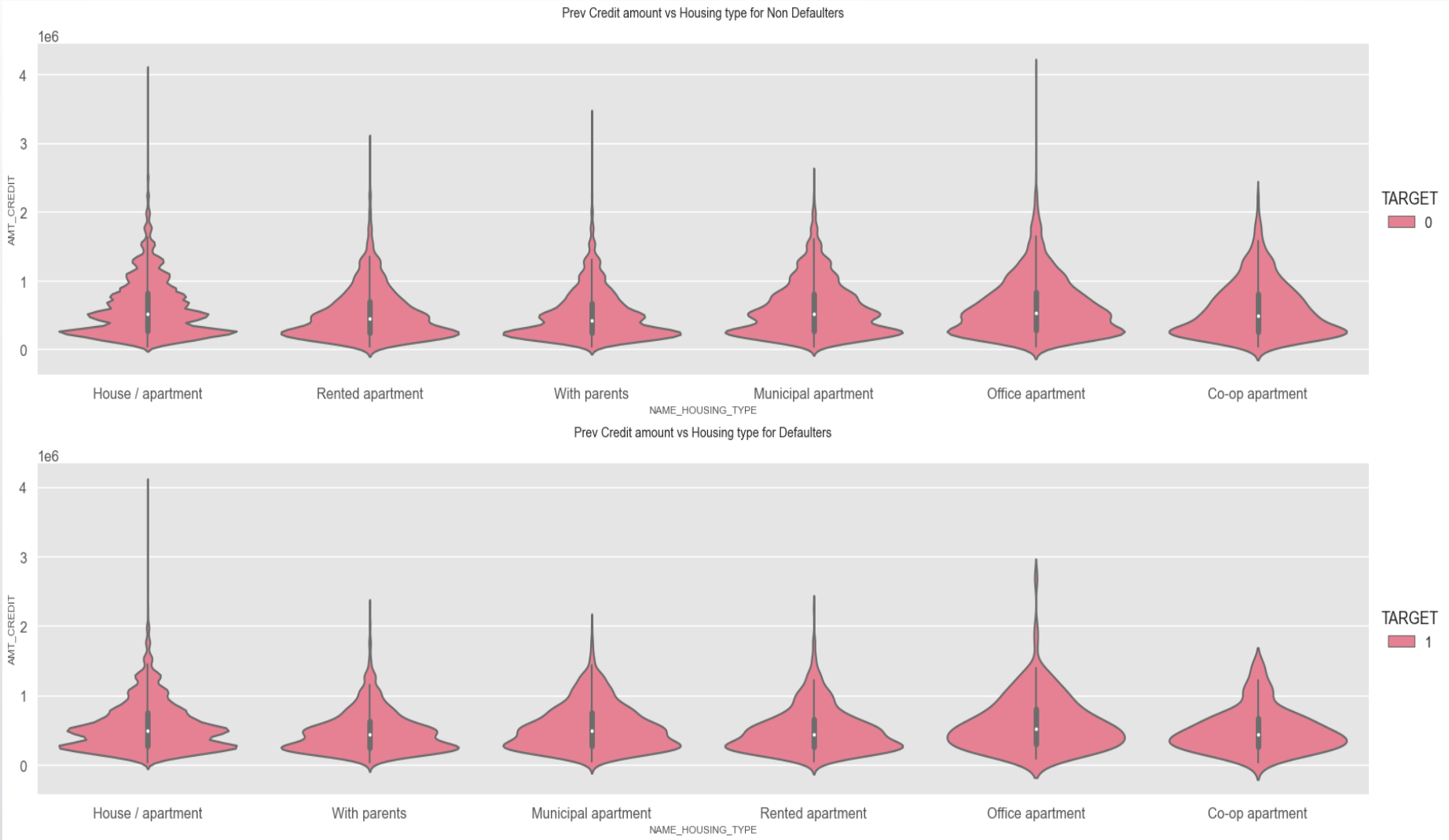
Inference:

Points to be drawn from the graph above for (Non-Defaulters).

1. People who belongs office apartments tend to have higher cresit limit followed by house apartments with Co-op apartment as the lowest

Points to be drawn from the graph above for (Non-Defaulters).

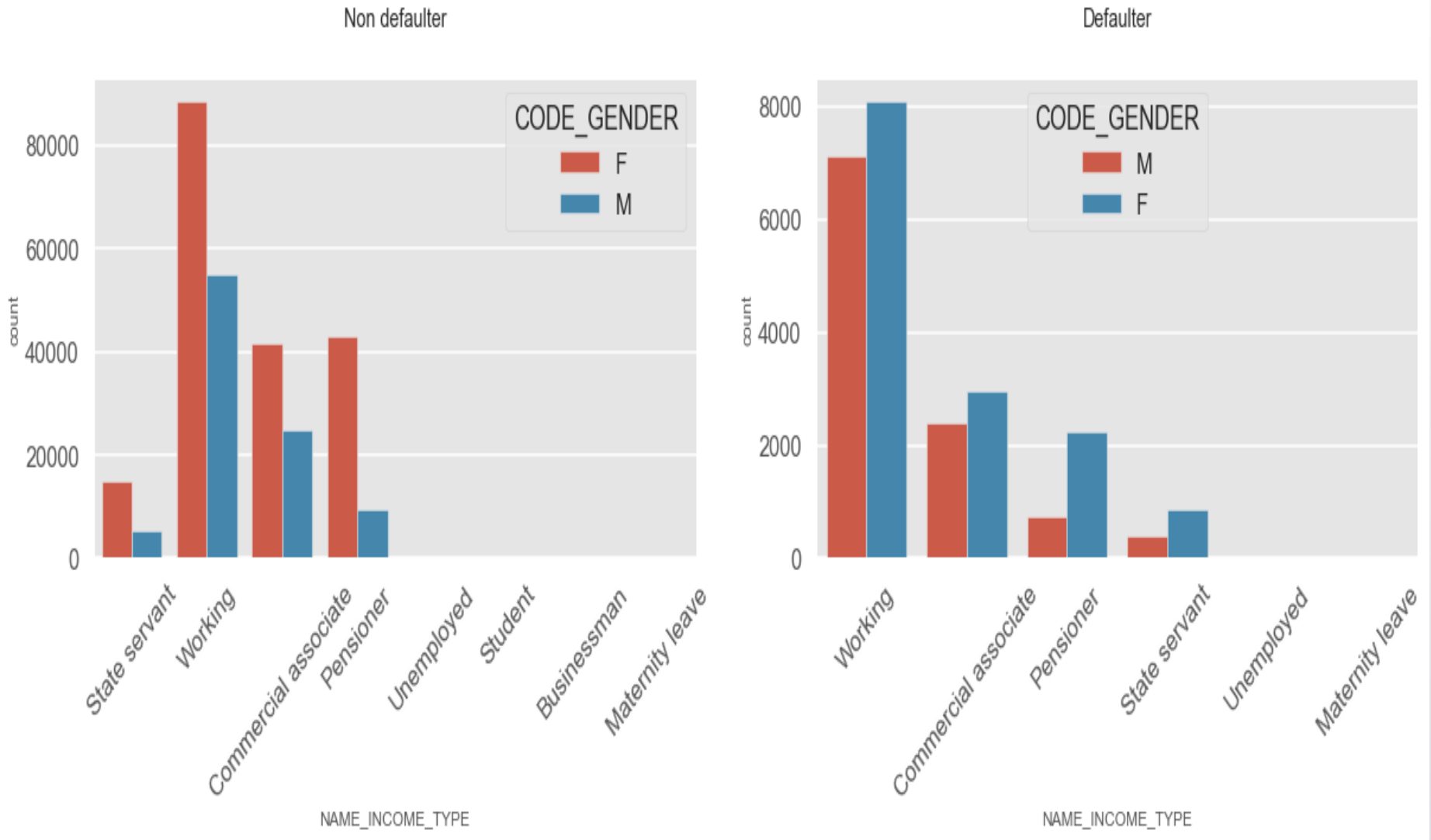
2. People from house apartments most likely to be defaulted.



plotting for NAME_INCOME_TYPE vs CODE_GENDER

Inference:

1. Females in all the categories are the highets.



MULTIVARIATE ANALYSIS

Multivariate Analysis of Continuous Columns

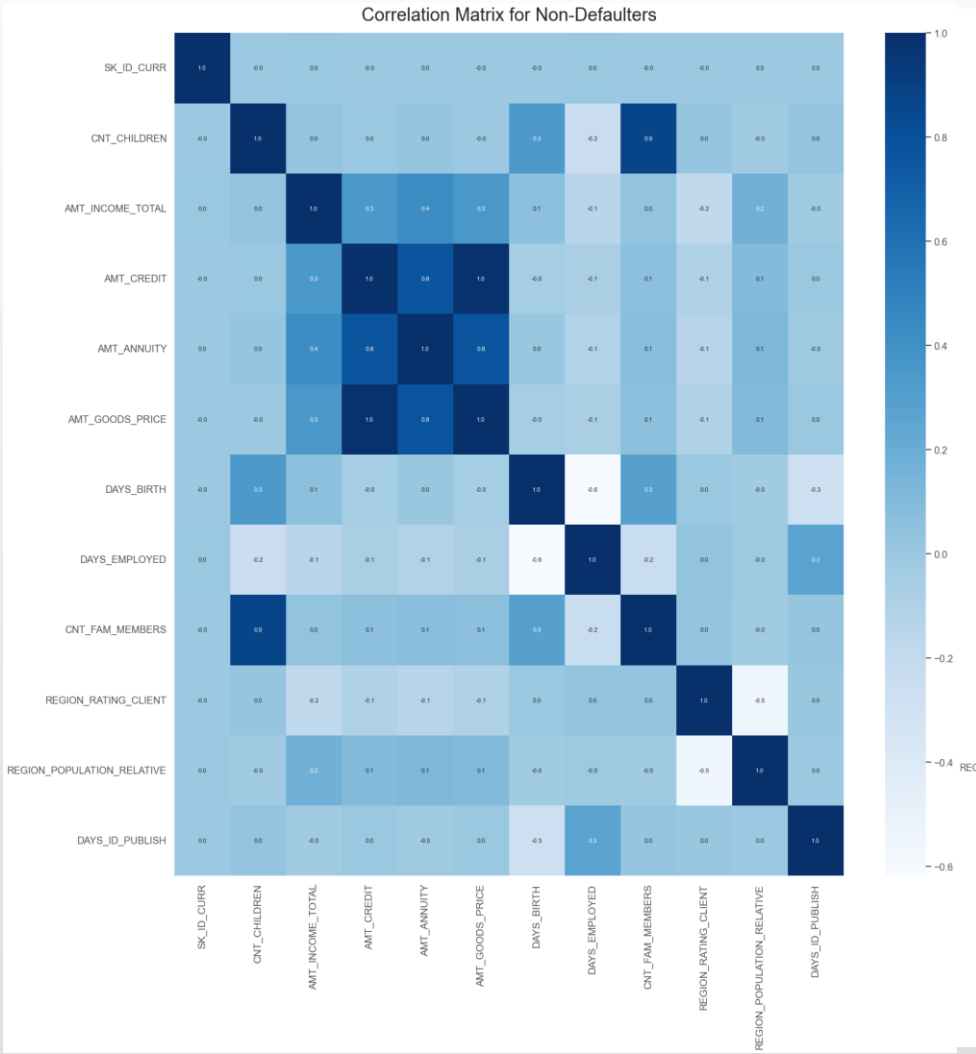
Inference:

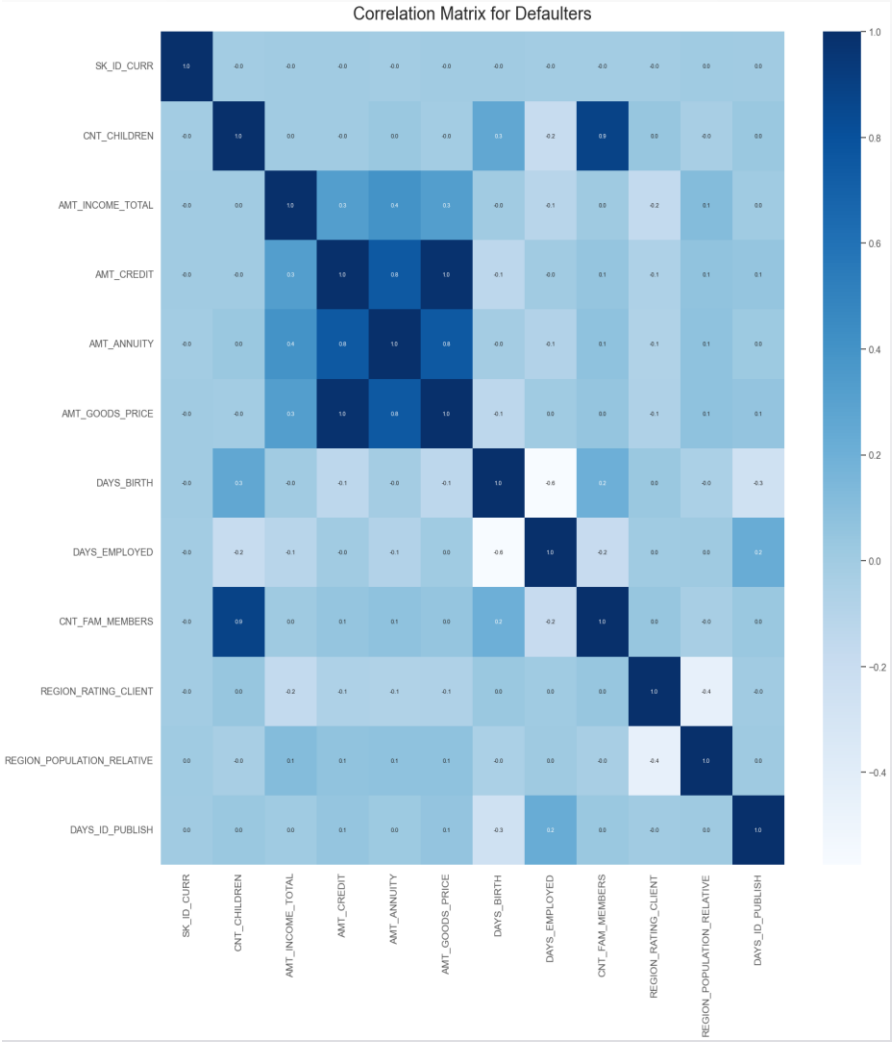
The credit amount is inversely related to the date of birth, with the credit amount increasing as the age decreases.

The credit amount is inversely related to the number of children a client has, with the credit amount increasing as the number of children decreases.

Credit amount is inversely related to days worked, i.e. credit amount is larger for those who have recently begun working.

The amount of income a customer earns is inversely related to the number of children they have, i.e., those with fewer children earn more.





Conclusions:

Conclusions:

- Instead of offering loans to co-op apartment housing, banks should focus on 'House apartment' and 'With parent' housing.

- People with a 'working' income category are more likely to have irregular or failed payments, according to the bank.

- In 'NAME CONTRACT TYPE,' the number of 'Revolving Loans' is quite low, and it also has the highest percentage of payment difficulties—around 10%. As a result, clients whose contract type in the prior application was 'Revolving loans' are the driving forces for Loan Defaulters.