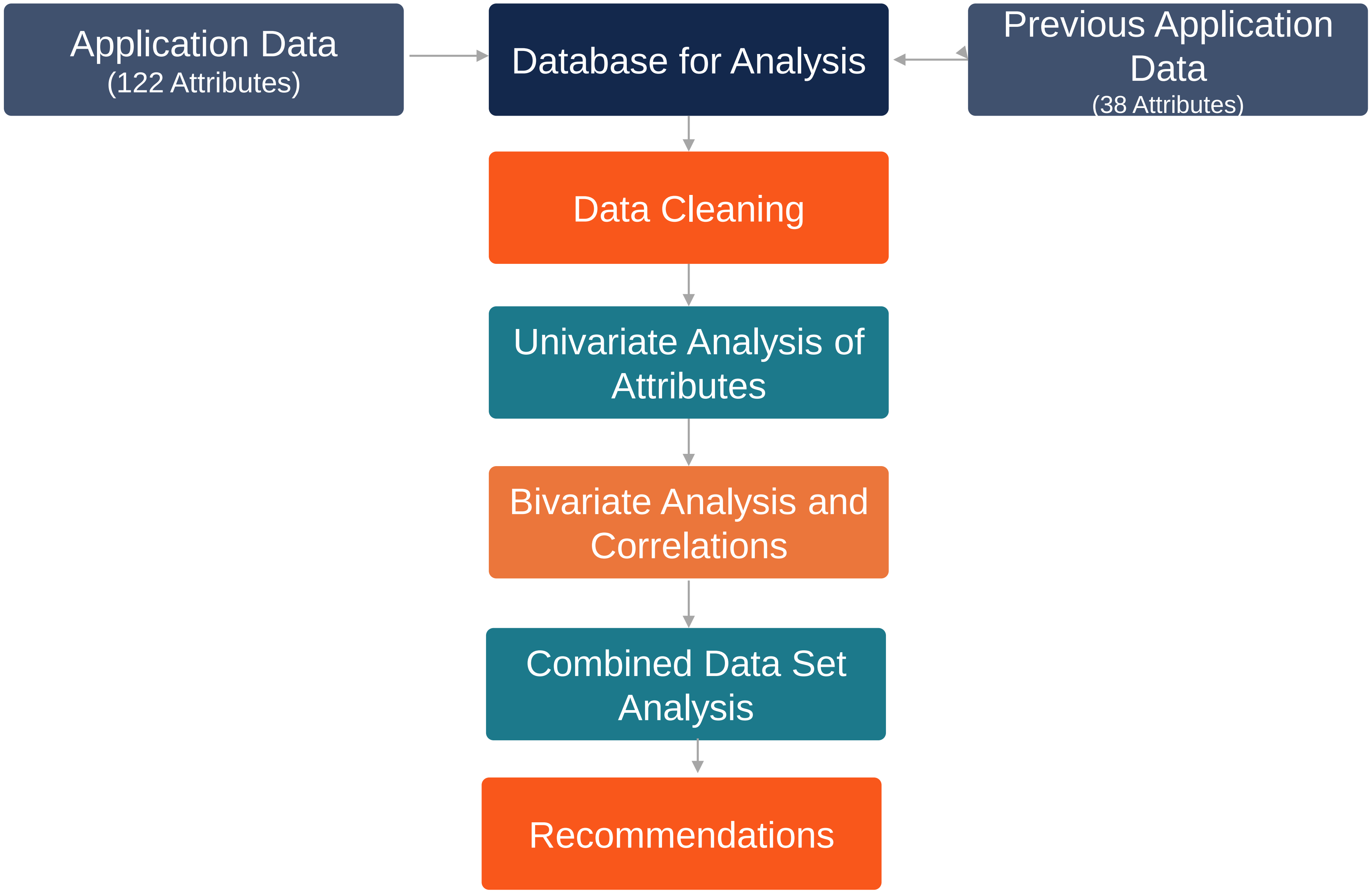
Objective

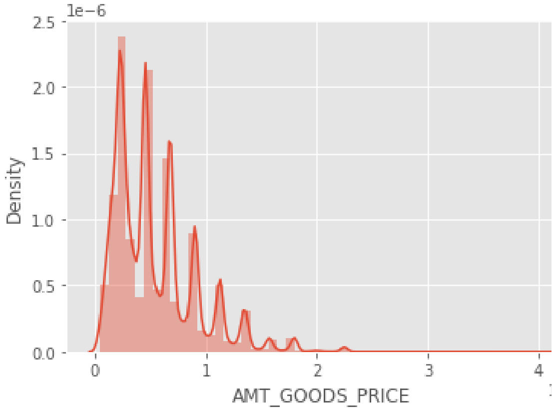
Problem Statement	• Understand how consumer attributes and loan attributes influence the tendency of default.
Business Objectives	• Identify patterns which indicate if a client has difficulty paying their instalments . • Ensure that the consumers capable of repaying the loan are not rejected • understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.
Methodology	• Univariate and bivariate analysis of different attributes etc. • Find top correlations for the defaults with other variables. • Share recommendations for business to make informed decisions on lending.

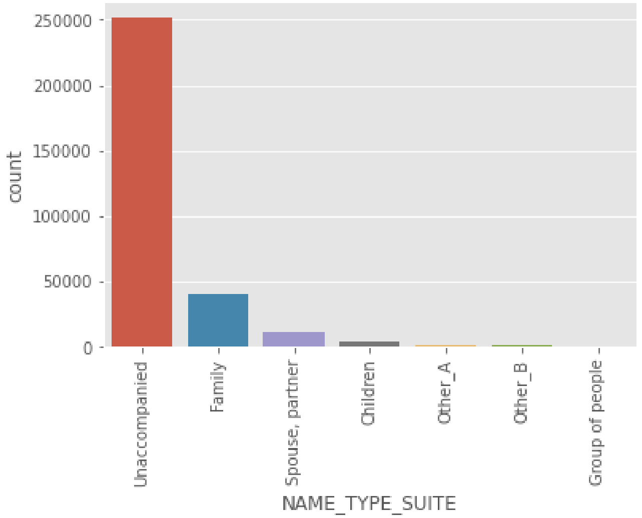
Strategy - Analysis Process

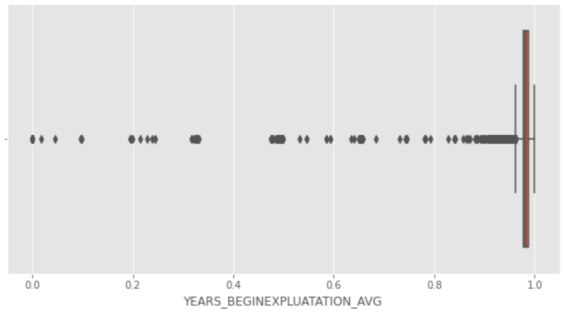


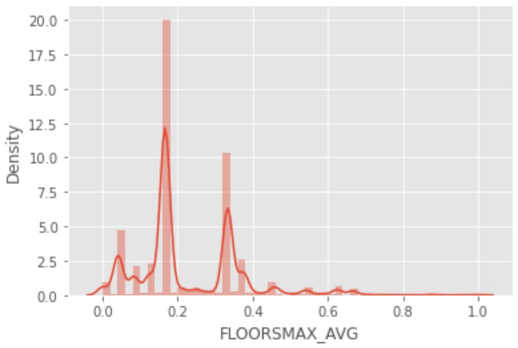
Data Cleaning Assumptions – Modal & Median Values

Mode Based Treatment



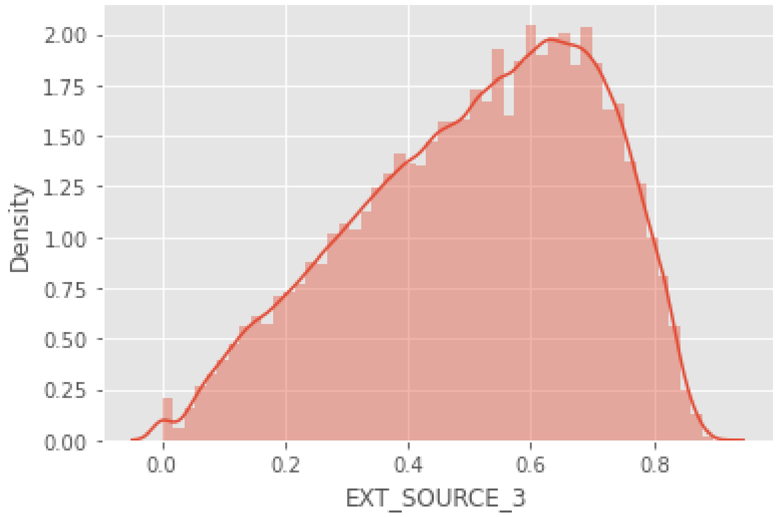


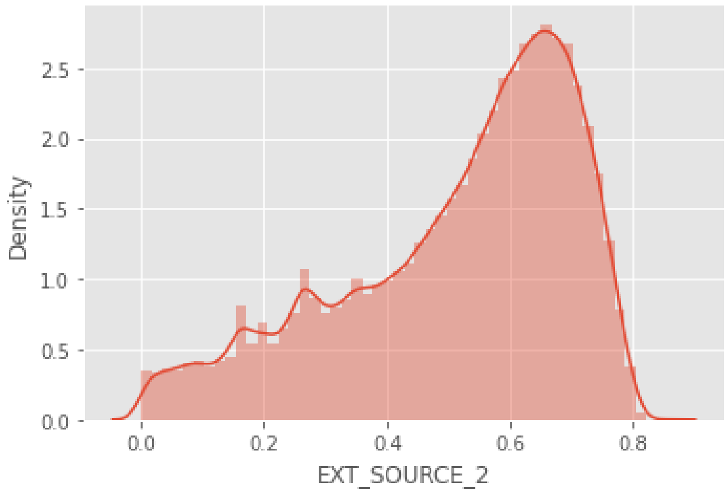




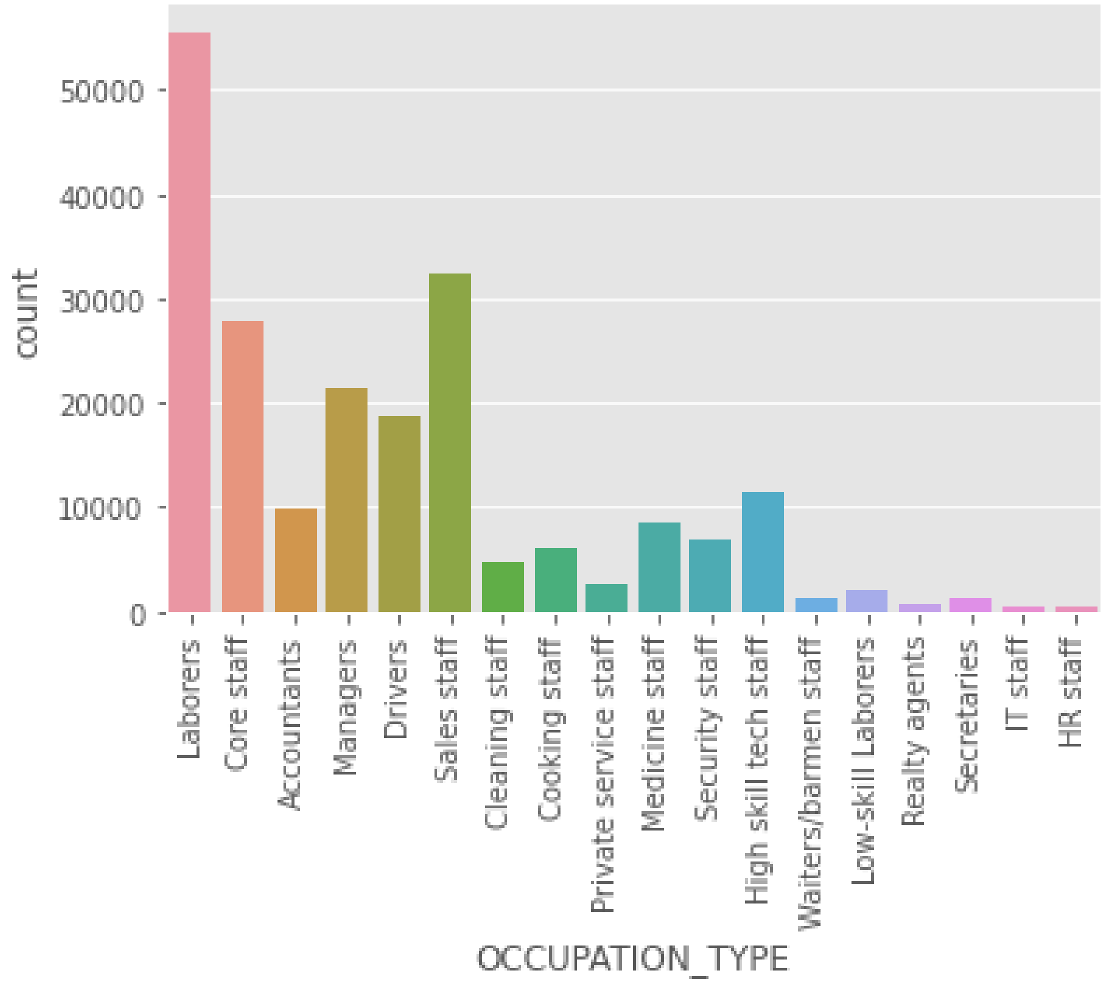
Other Mode Based assumed data points included the attributes -
YEARS_BEGINEXPLUATATION_AVG, FLOORSMAX_MODE, YEARS_BEGINEXPLUATATION_MEDI, TOTALAREA_MODE, OBS_30_CNT_SOCIAL_CIRCLE and AMT_REQ_CREDIT_BUREAU

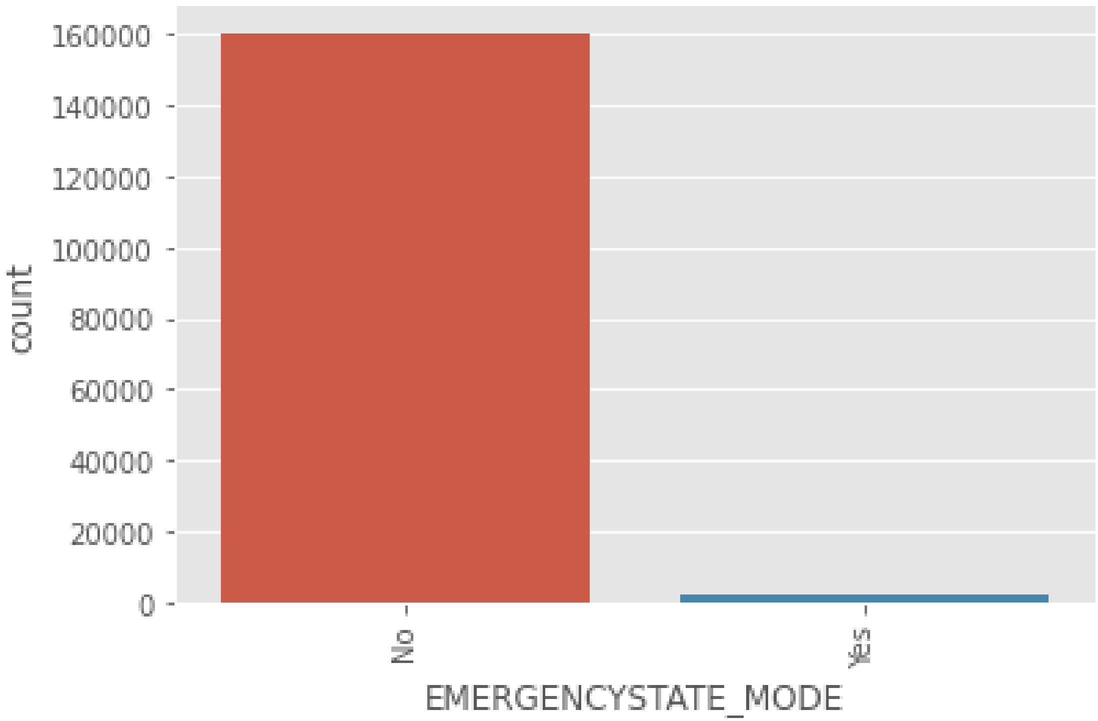
Median Based Treatment





Data Cleaning Assumptions – Unchanged Data





No Assumptions made in the unavailable data for these 2 attributes is the data was found to be sensitive and any assumption would affect the results based on these 2 attributes.

Categorizing & Deriving Numerical Data for Analysis

```
count    309683.000000
mean         43.407662
std          11.945723
min          20.000000
25%          33.000000
50%          43.000000
75%          53.000000
max          69.000000
Name: AGE, dtype: float64
```

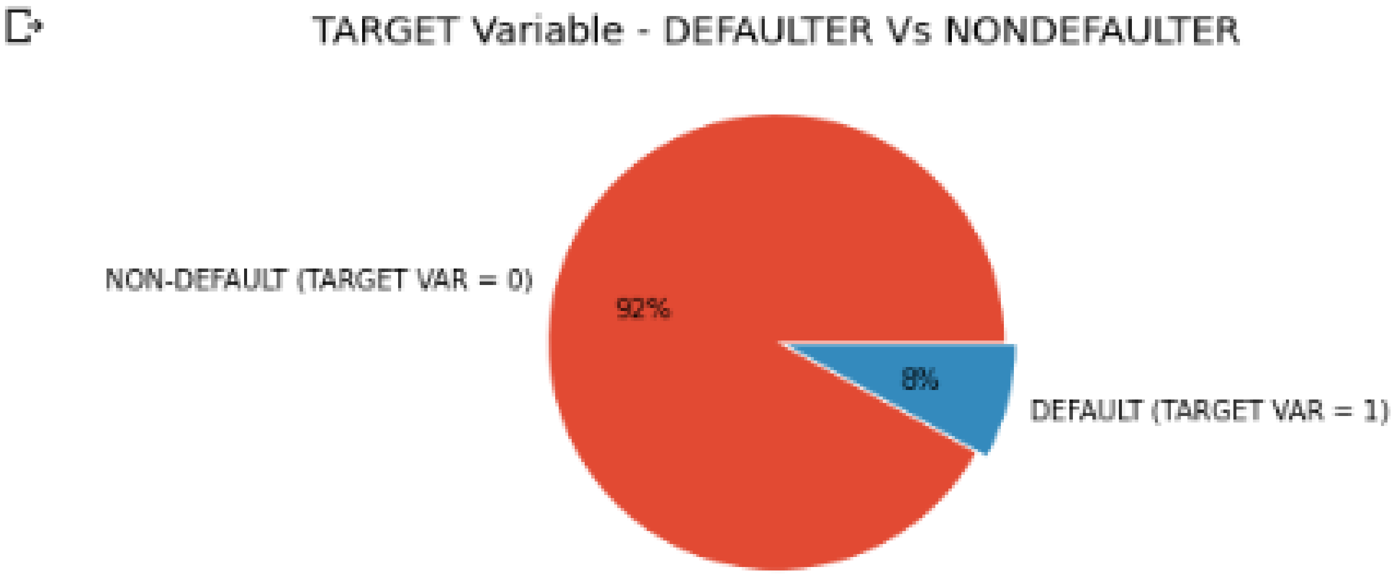
Grouping of Age Groups and
Correcting the negative values

```
appl_db['FOIR']=round((appl_db['AMT_CREDIT']/appl_db['AMT_INCOME_TOTAL']))
appl_db['FOIR'].describe()
```

```
count    309683.000000
mean          3.957373
std           2.705975
min           0.000000
25%           2.000000
50%           3.000000
75%           5.000000
max          85.000000
Name: FOIR, dtype: float64
```

Deriving Fixed Obligations to Income Ratio from the Data for Business Use –
Assessing Extent to be Customer is leveraged/indebted

Target Variable Distribution



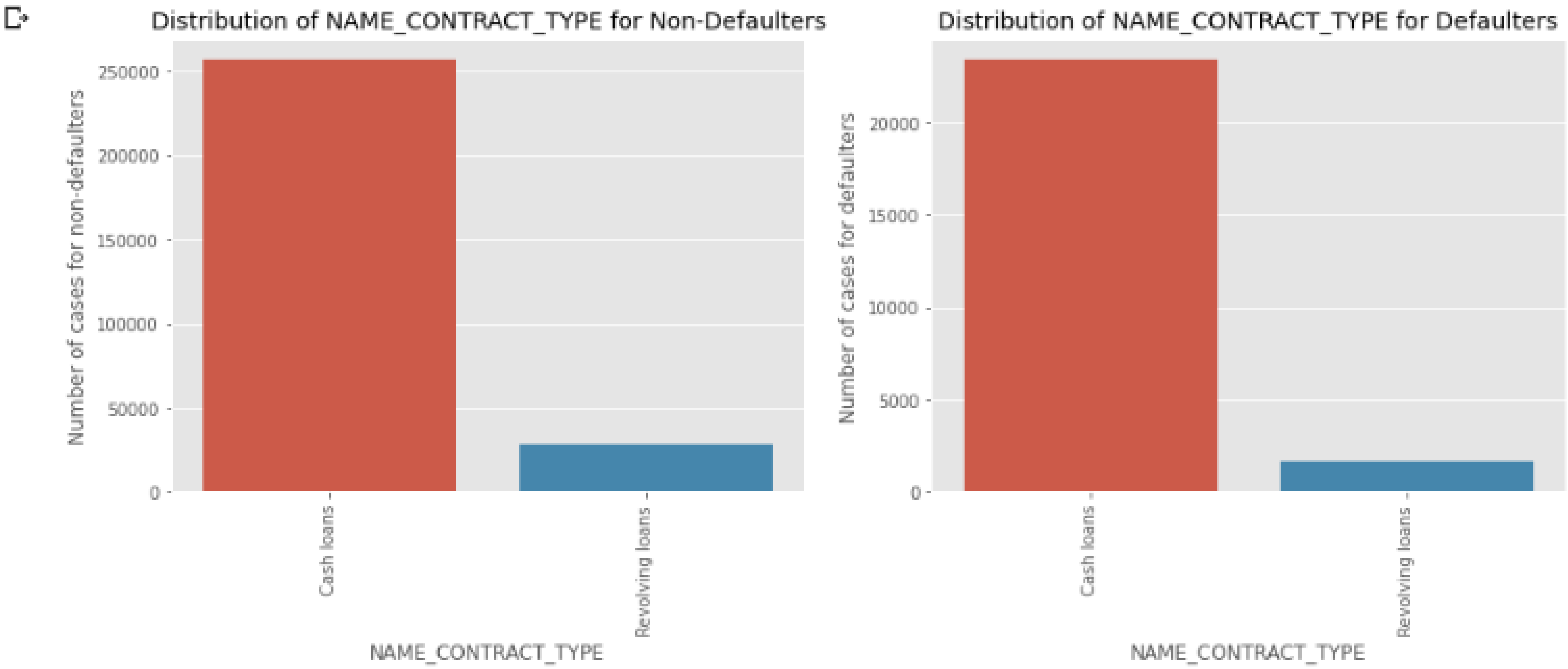
The Imbalance ratio is found to be 11.387

Also, the clear demarkation observed in the above graph of 92% Non defaulters vs. 8% defaulters

Basis the above skewed distribution, databases was split into two – Defaulters and Non Defaulters to understand how various attributes were affecting the Defaulting Tendency of Applicants

Univariate Analysis

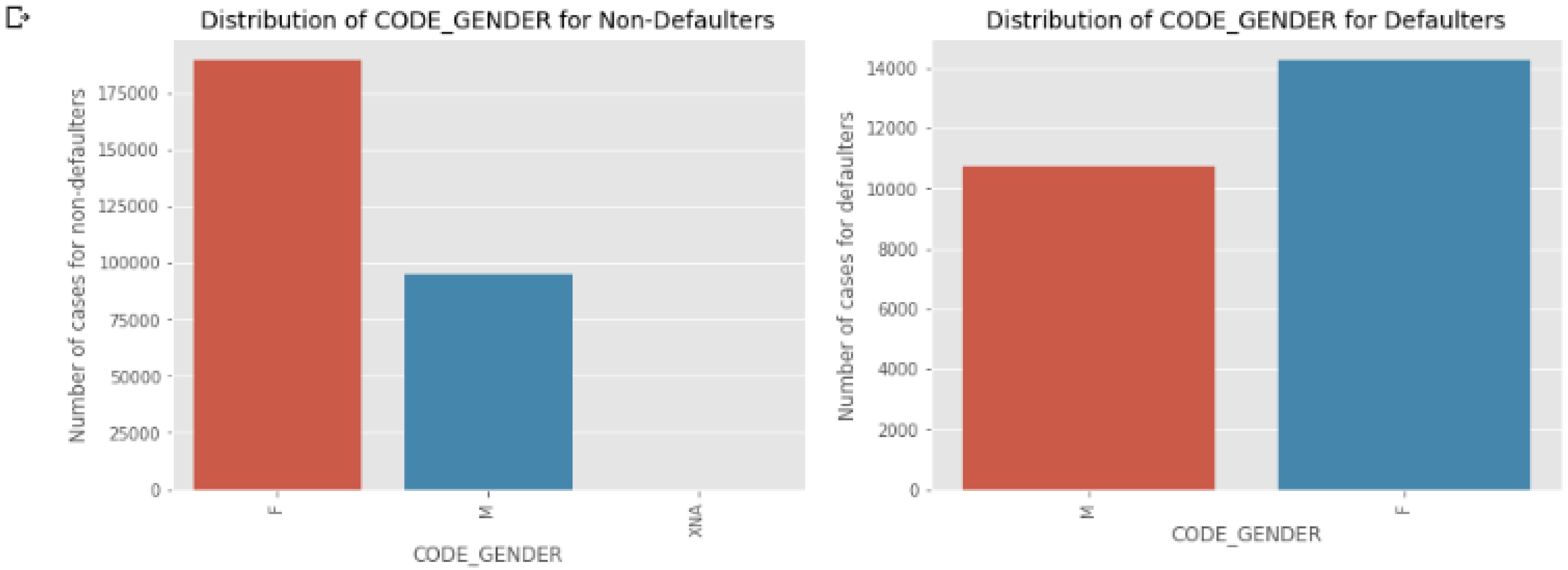
DEFAULTING TENDENCY IN TYPE OF LOANS



The revolving loan comparatively has lower default rate. The risk associated to it would be considered low

Univariate Analysis

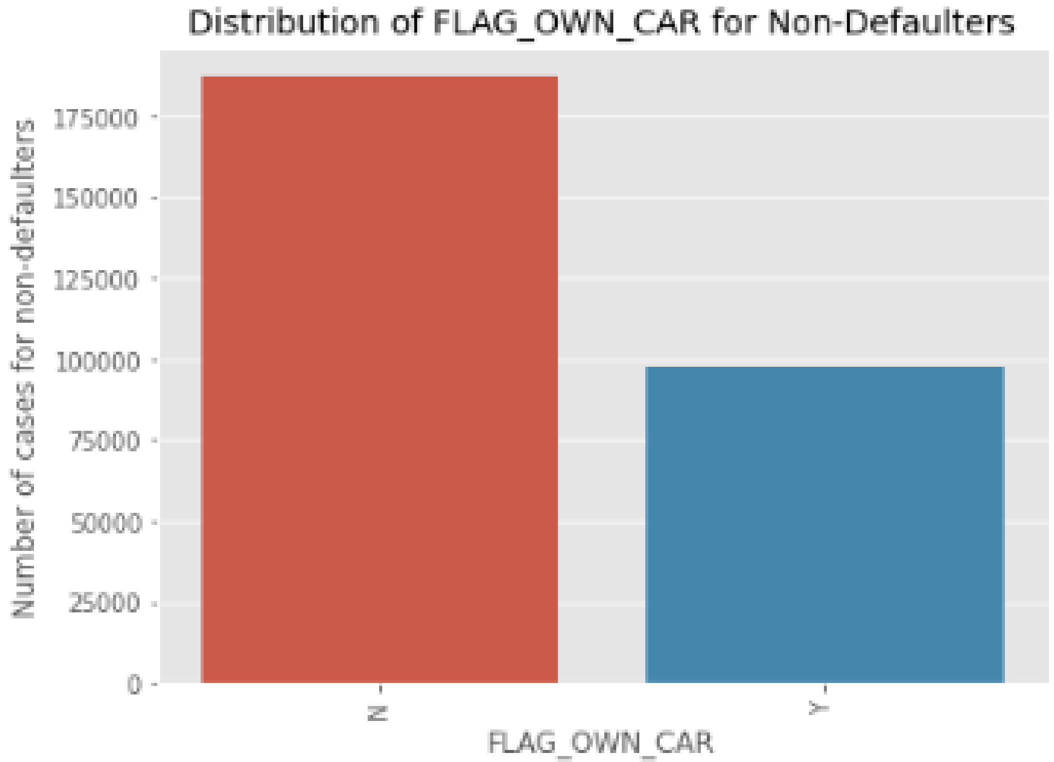
GENDER BASED ANALYSIS

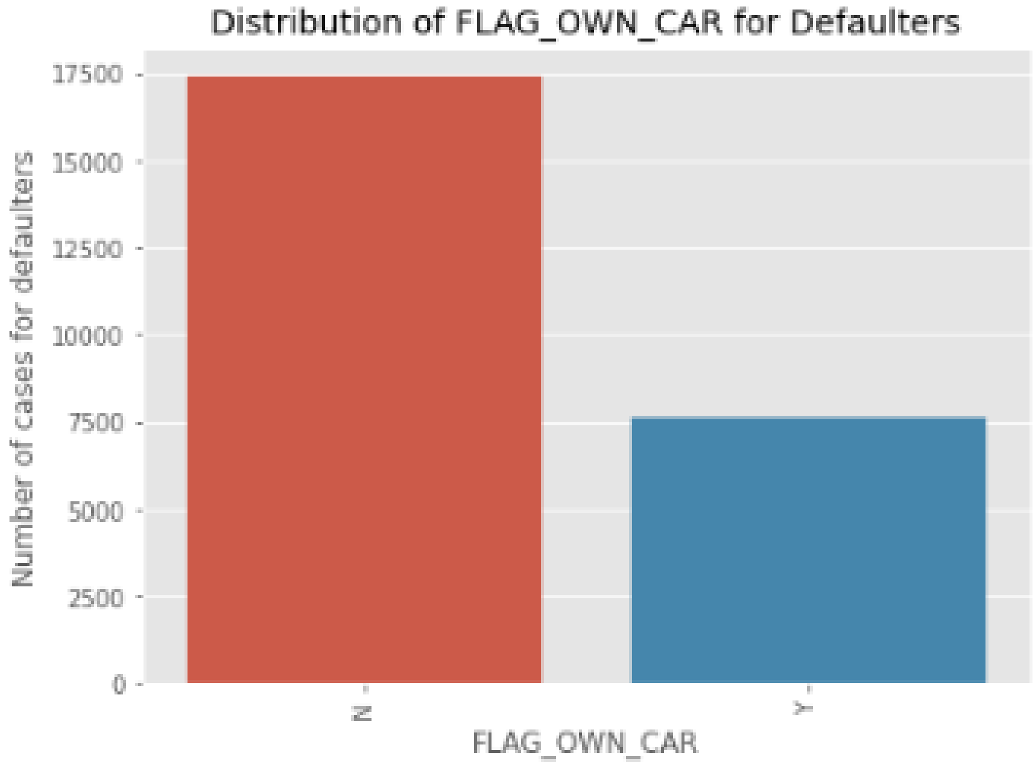


Female non defaulters visibly contribute higher percentage. However not that much difference in the defaulters. So may be there are more numbers female applications hence more number of female defaulters. The percent of defaulting of females is much less when compared to the males

Univariate Analysis

CAR OWNERSHIP



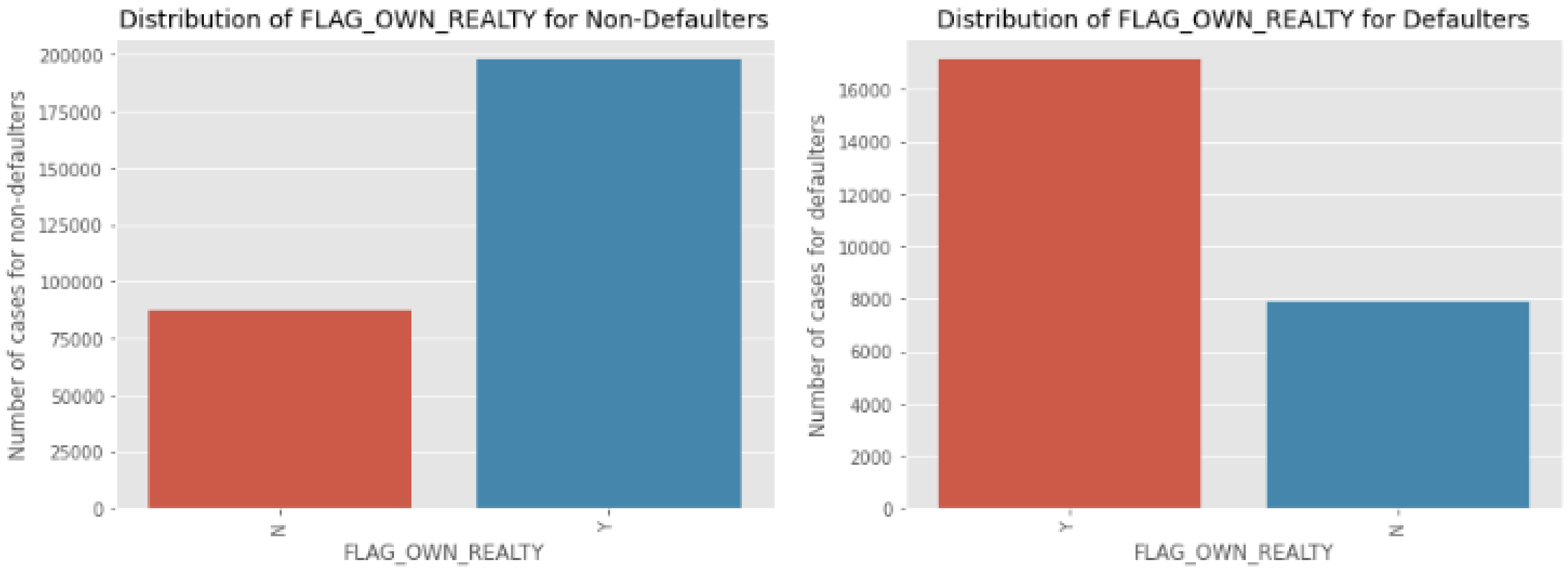


There is visible difference between the defaulters with cars compared to without cars. It simply can mean that the people without cars are more than with car. It seems that people with cars have a defaulter rate of lower compared to those without cars.

Prima Facie it appears car owners have slightly less tendency to default. However, the same would required deeper analysis if there is any correlation between the default rate and car ownership

Univariate Analysis

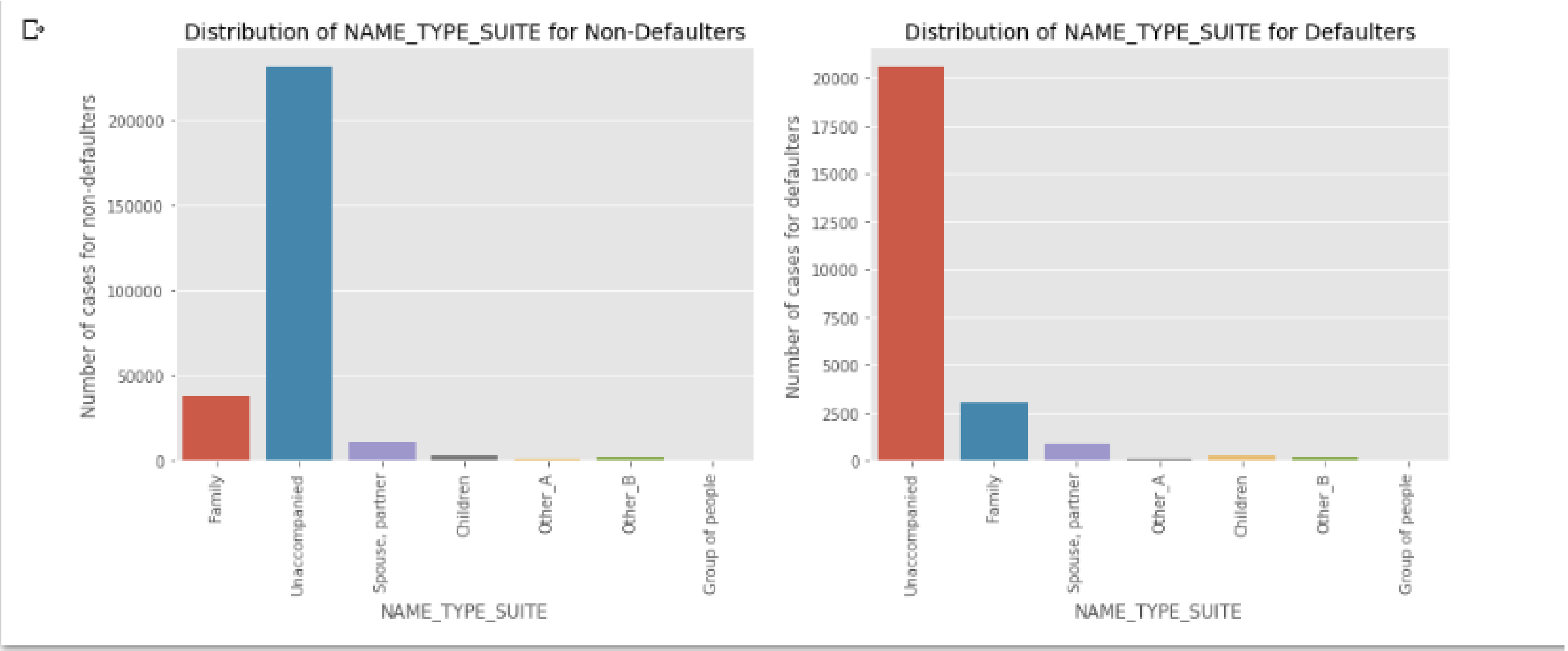
PROPERTY OWNERSHIP



Those without an existing property appear to have higher default rates - This begs deeper analysis if there are factors causing same like better finances management or simply higher income

Univariate Analysis

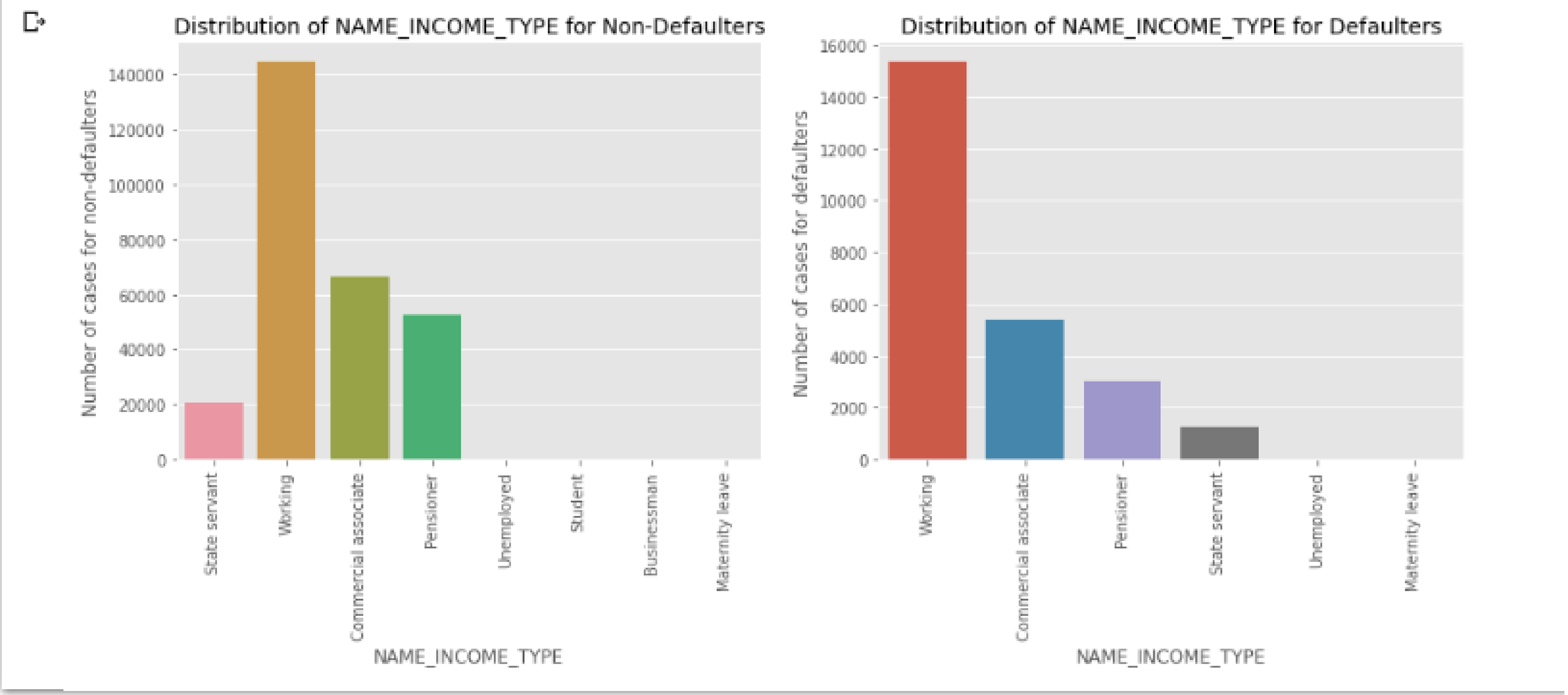
Co Applicant On Loan?



Since the Unaccompanied applicants are more, even the defaulters seem to be more but, there visibly is a dip in defaulter percent of those accompanied by family and spouse. The risk might be considered low in those. Also a possibility of Jointly repaying the loan. This needs to be analysed separately with appropriate dataset

Univariate Analysis

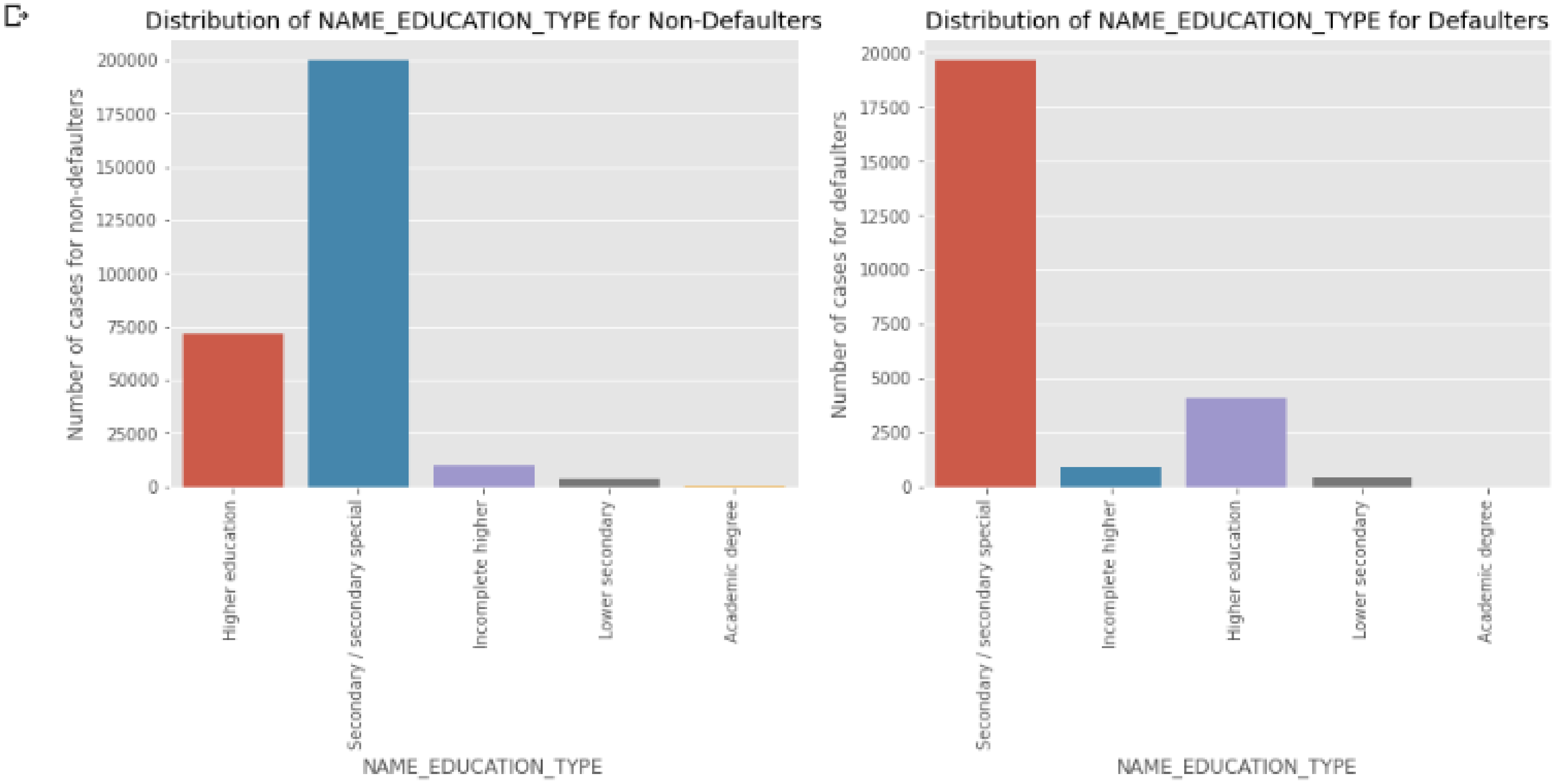
Type of Income



Students don't default. May be because they are not required to pay while their education is on. Even Businessmen never default. Most of the loans are distributed to working class people We also see that working class people contribute lower percent to non defaulters while they contribute to those of the defaulters. The chances of defaulting seems more in their case.

Univariate Analysis

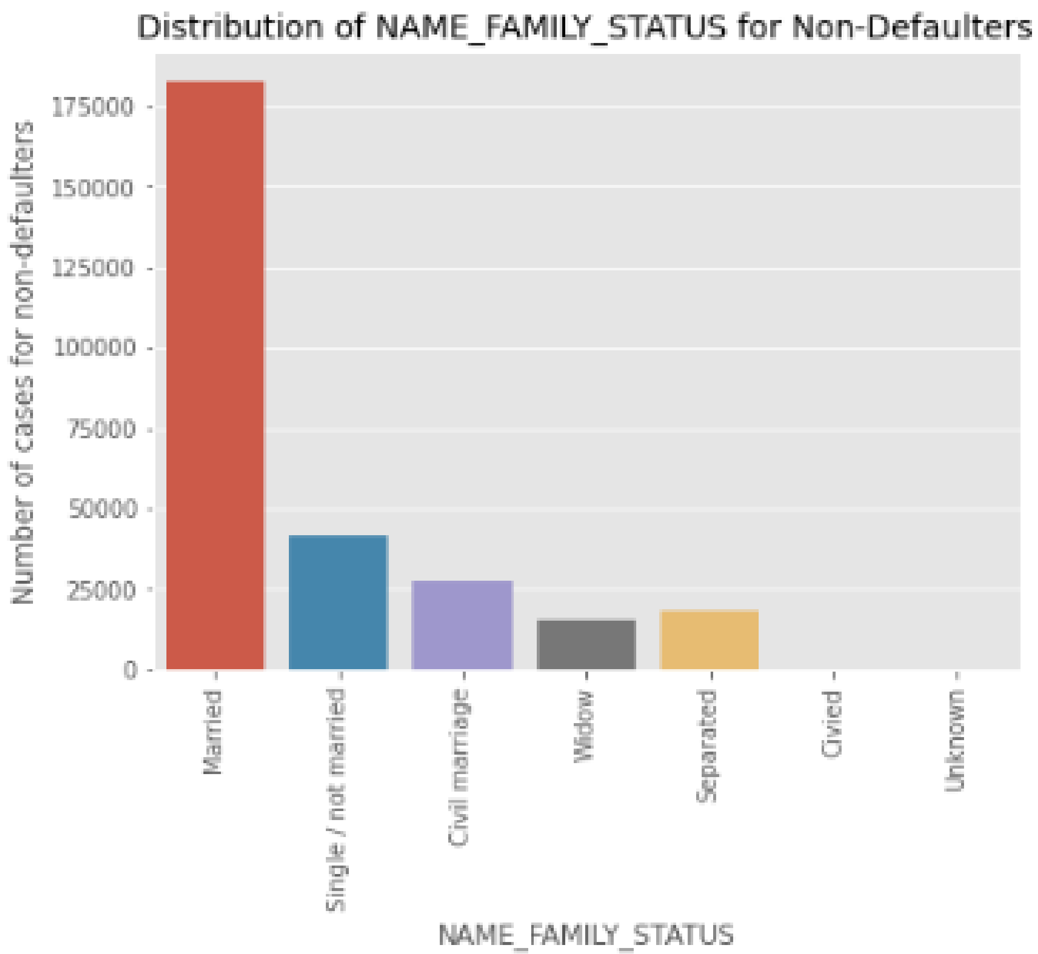
Type of Education

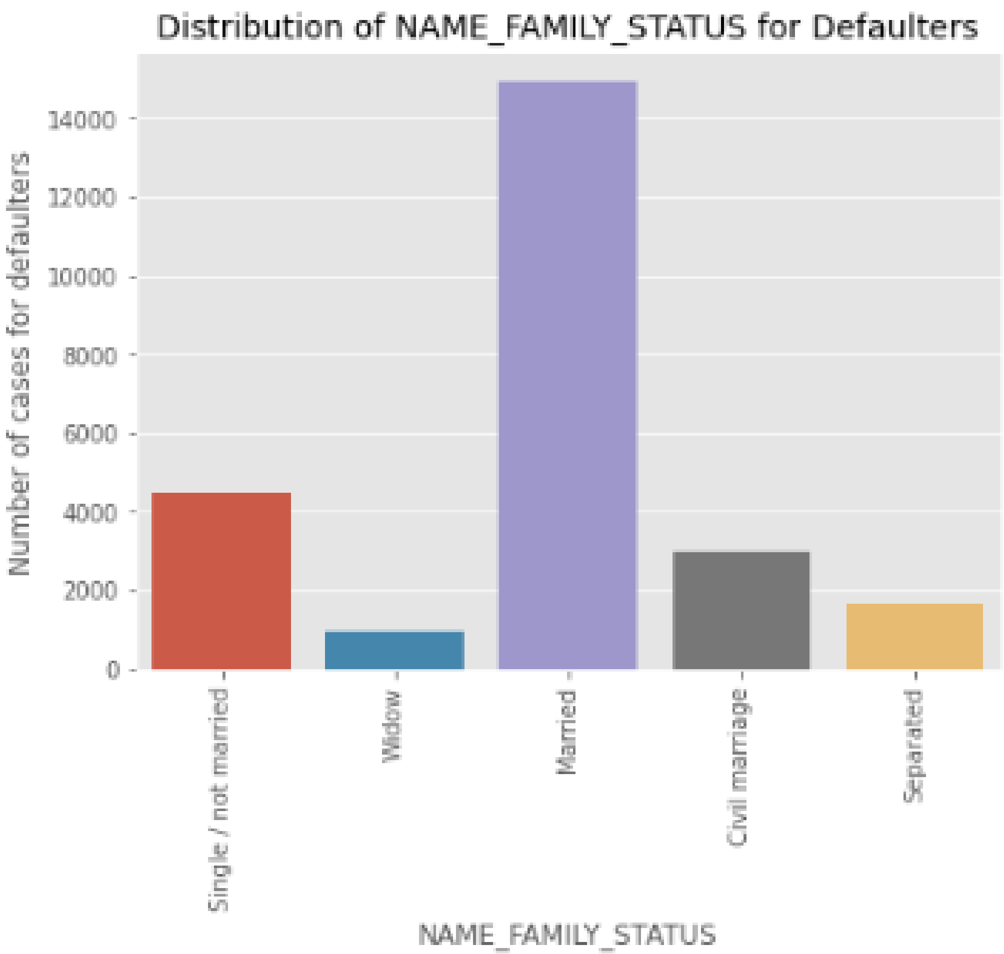


Higher education seem to be less likely to get defaulted with the non defaulter rate lower compared to the defaulter rate

Univariate Analysis

Family Status

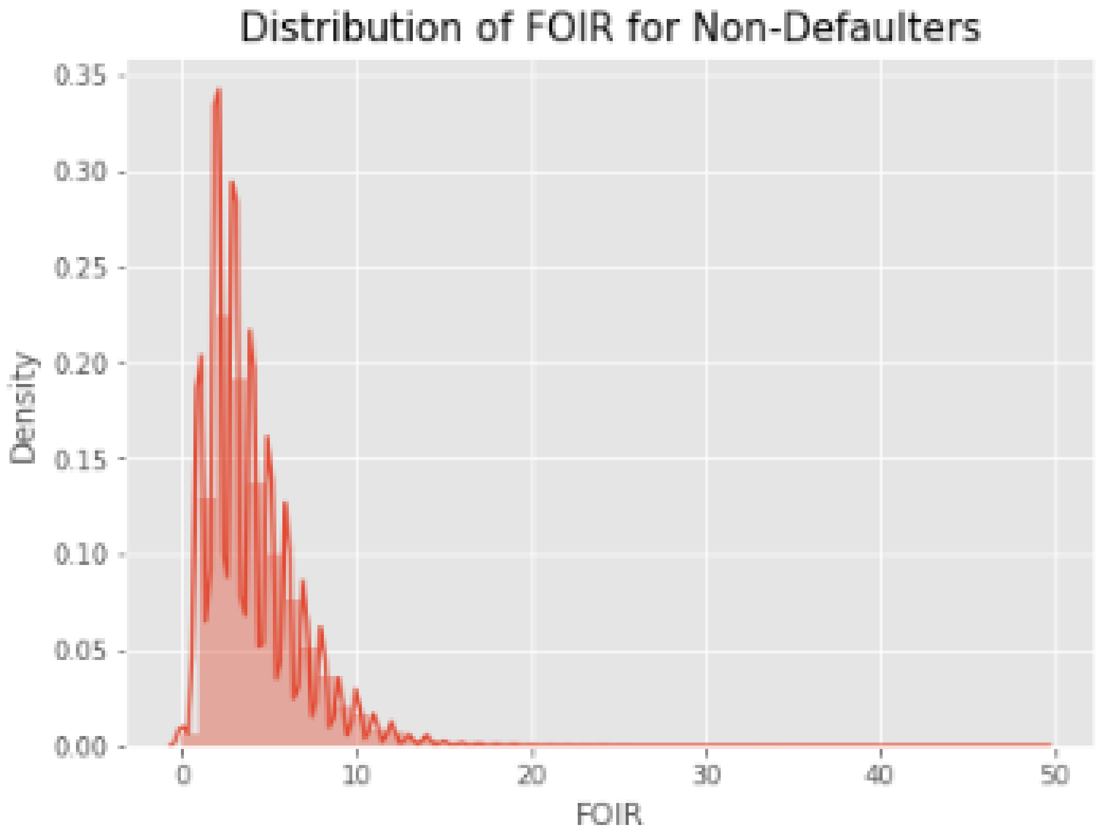


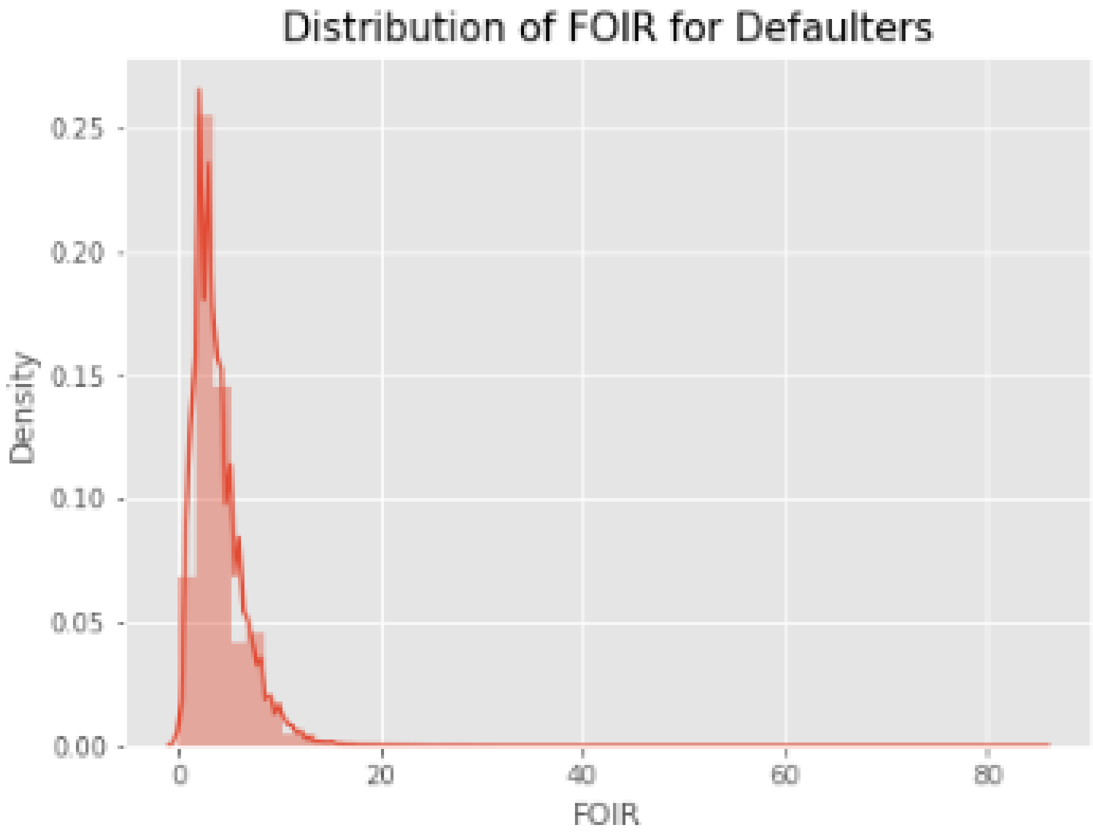


Married people tend to apply for more loans comparatively. But from the graph we see that Single/non Married people contribute higher percent to Non Defaulters than to the defaulters. Hence there can be more risk associated with the Single/non Married group.

Univariate Analysis

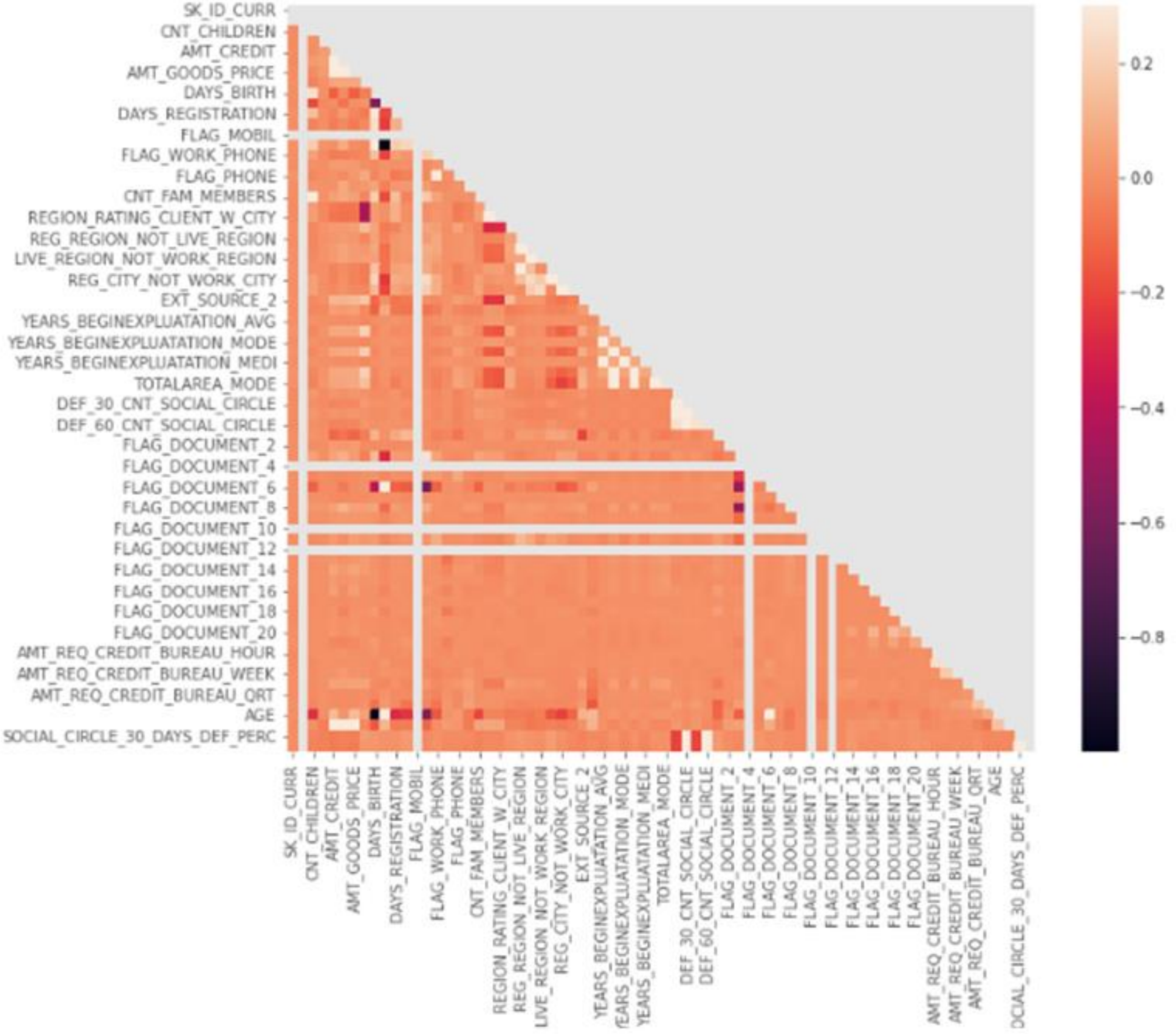
FOIR





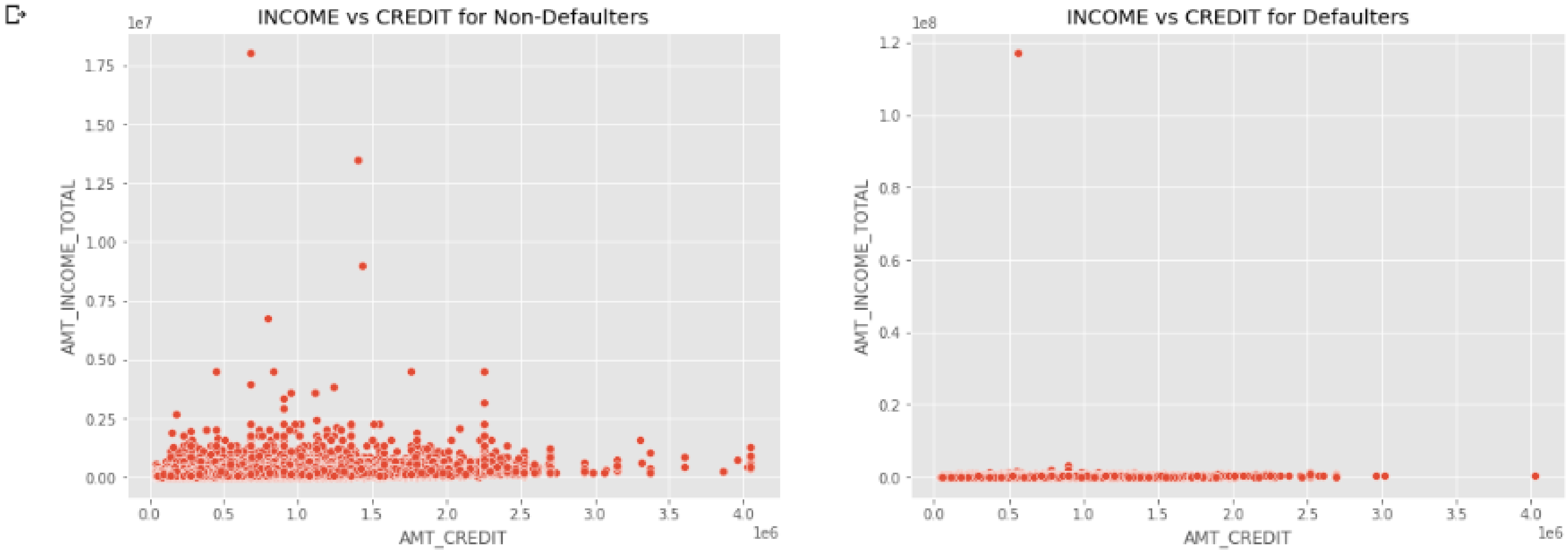
The FOIR ratio for the defaulter is seen close to 10%

Bivariate Analysis



Heat Map for Correlations between Attributes

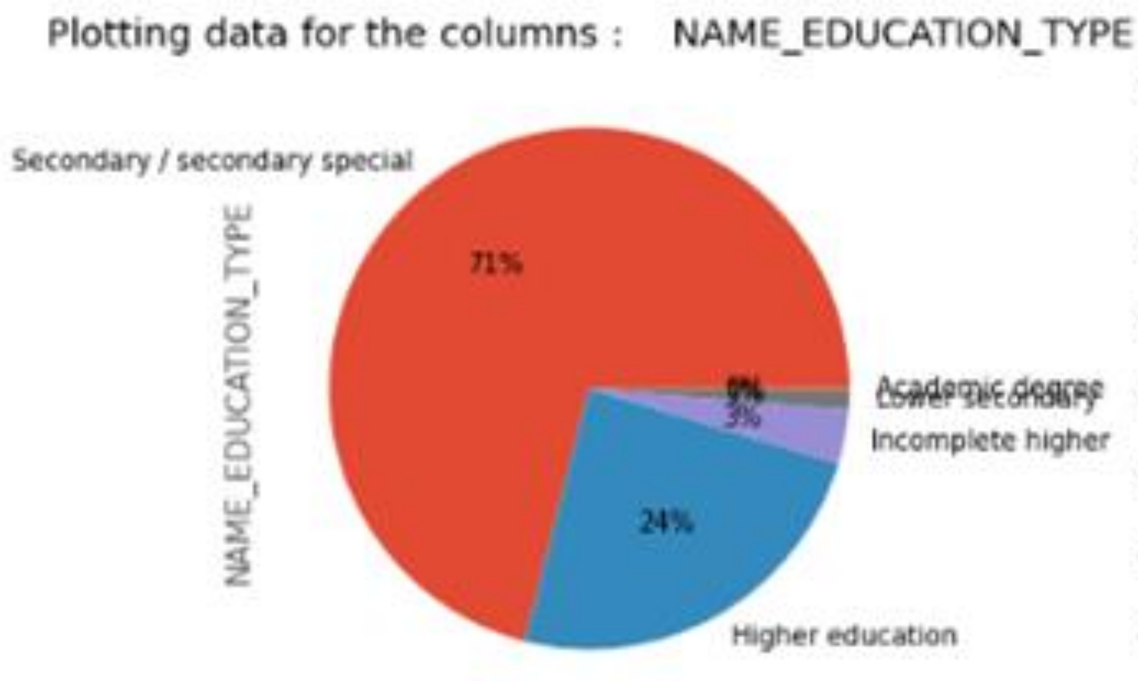
Bivariate Analysis

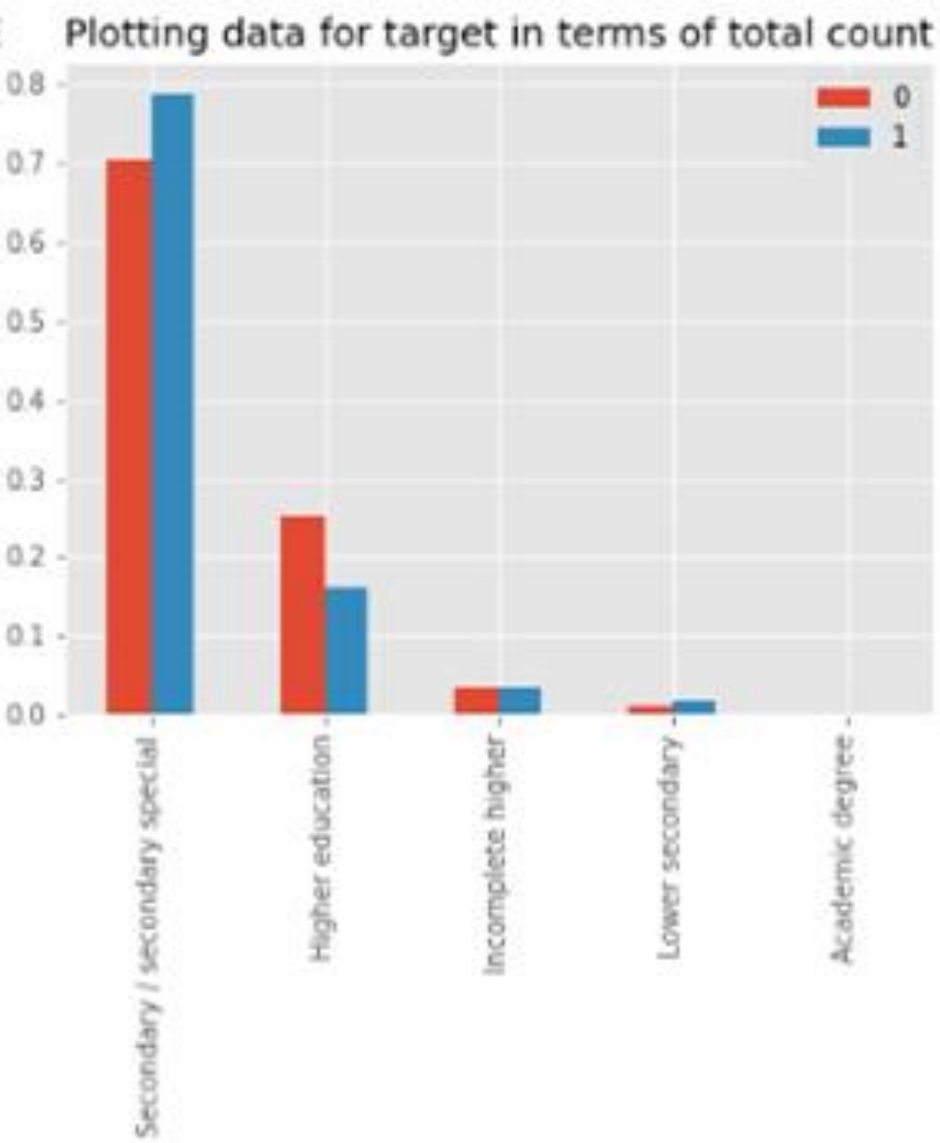


Defaulters are mainly in low income category, with 1 extreme case of defaulter in high income.

Income Based

Combined Data Analysis





Income Based

Recommendations for Business

- Revolving Loans seem to have a lesser tendency of defaults – OD and CC Limits may be explored further as a less risk carrying product , for more market penetration
- Loans in name of female applicants have a lesser default tendency, same may be explored with lucrative offerings in the market.
- Car and Realty ownership showed lesser tendency of defaults.
- The default tendency of applicants who came with someone was less. This was prbably due to co applicants taken on loan. Having co applicants on loan as a mandate would improve serviceability of loan
- Income type – State Servants had the lowest default tendency among defaulters. Business may consider lucrative rates on these loans.
- Higher education also played a role in defaulting tendency. This may be taken as a parameter while assessing a loan application among other critical parameters like FOIR, Income levels.