

# Credit EDA Case Study

Prepared By-  
Amit Kumar Singh  
Suman Kumar Das

# PROBLEM STATEMENT

- To identify the patterns from the given credit data set which indicate whether the client will be able to pay the loan amount or not.

# Missing Values

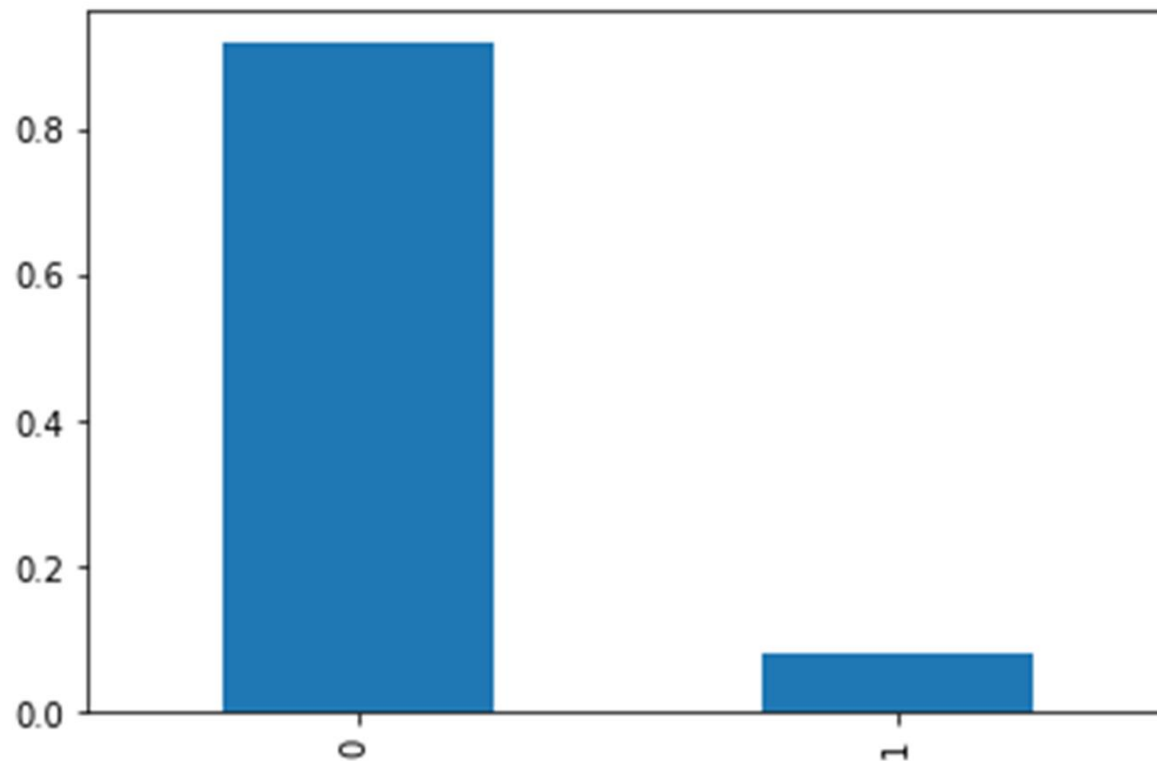
- The data set containing information about application is having total 122 columns.
- 41 columns out of 122 columns were having null value more than 50%. These columns were dropped for further analysis.
- There were some columns having missing values less than around 13%. For these columns missing value, some imputation method can be used.
  - For example: For "AMT\_REQ\_CREDIT\_BUREAU\_HOUR" column mode value can be used for imputation.

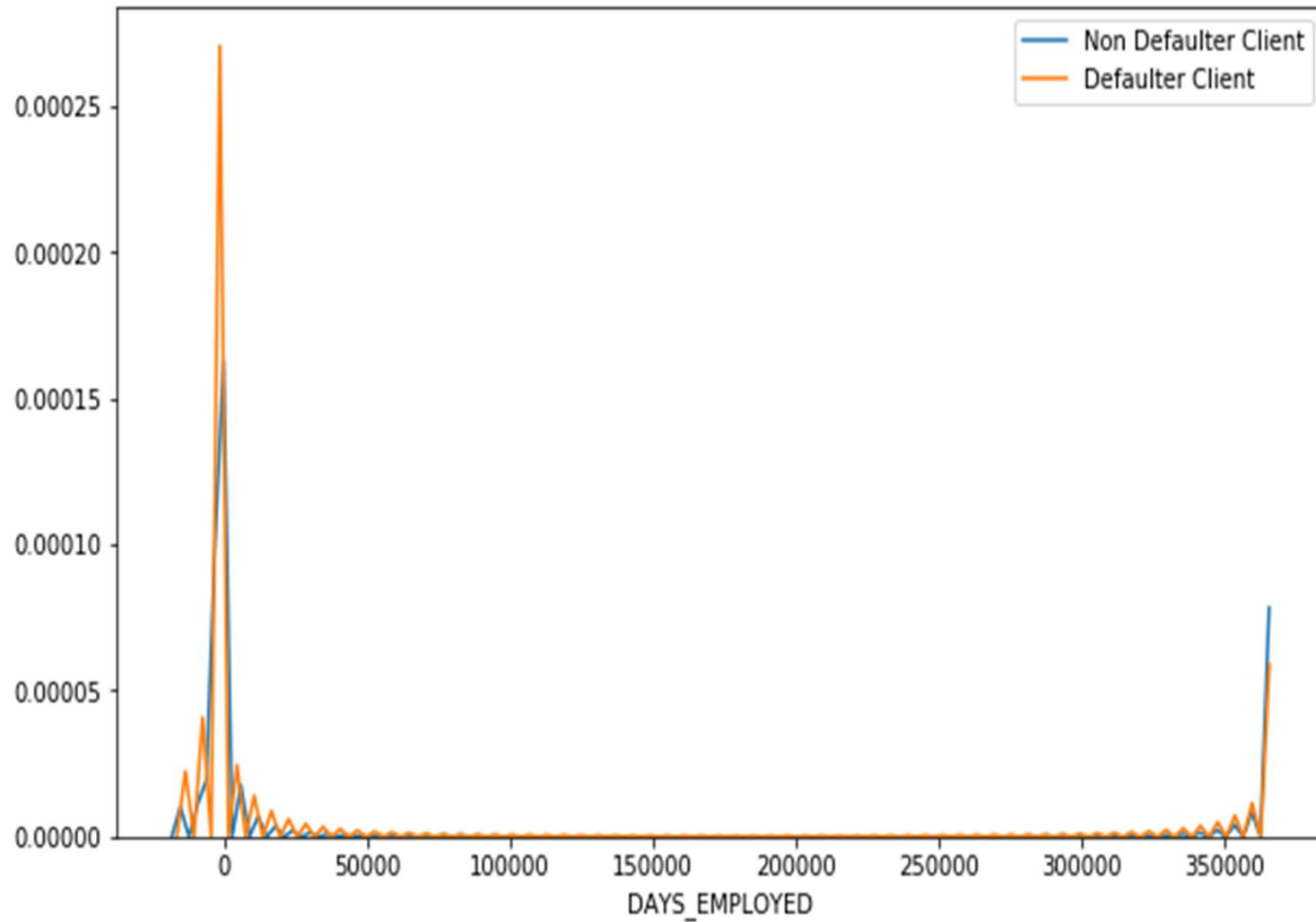
# DATA IMBALANCE

The dataset provided for the analysis was imbalanced as shown in the graph. The ration of imbalance is:

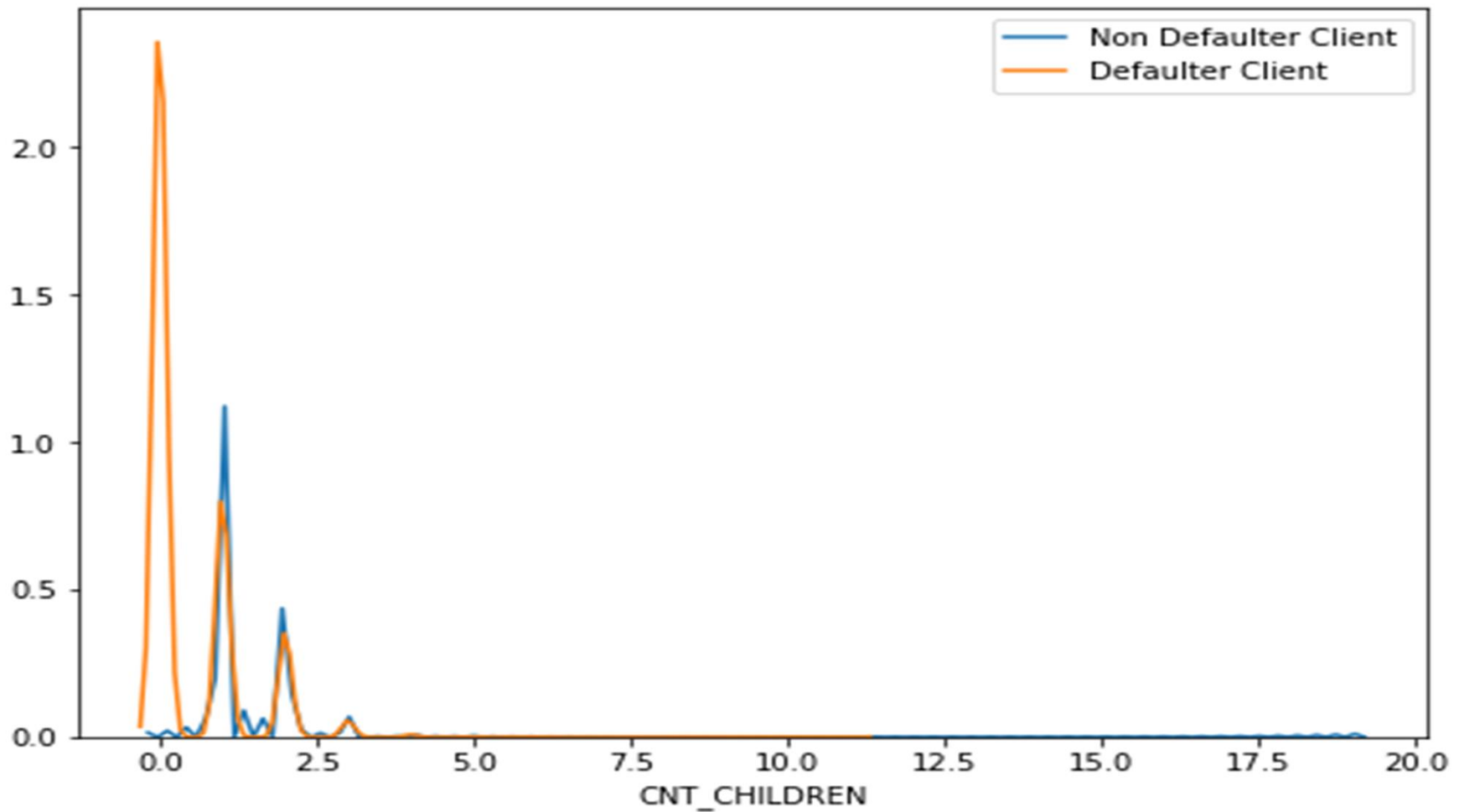
**Target 0- 91.92**

**Target 1- 8.073**

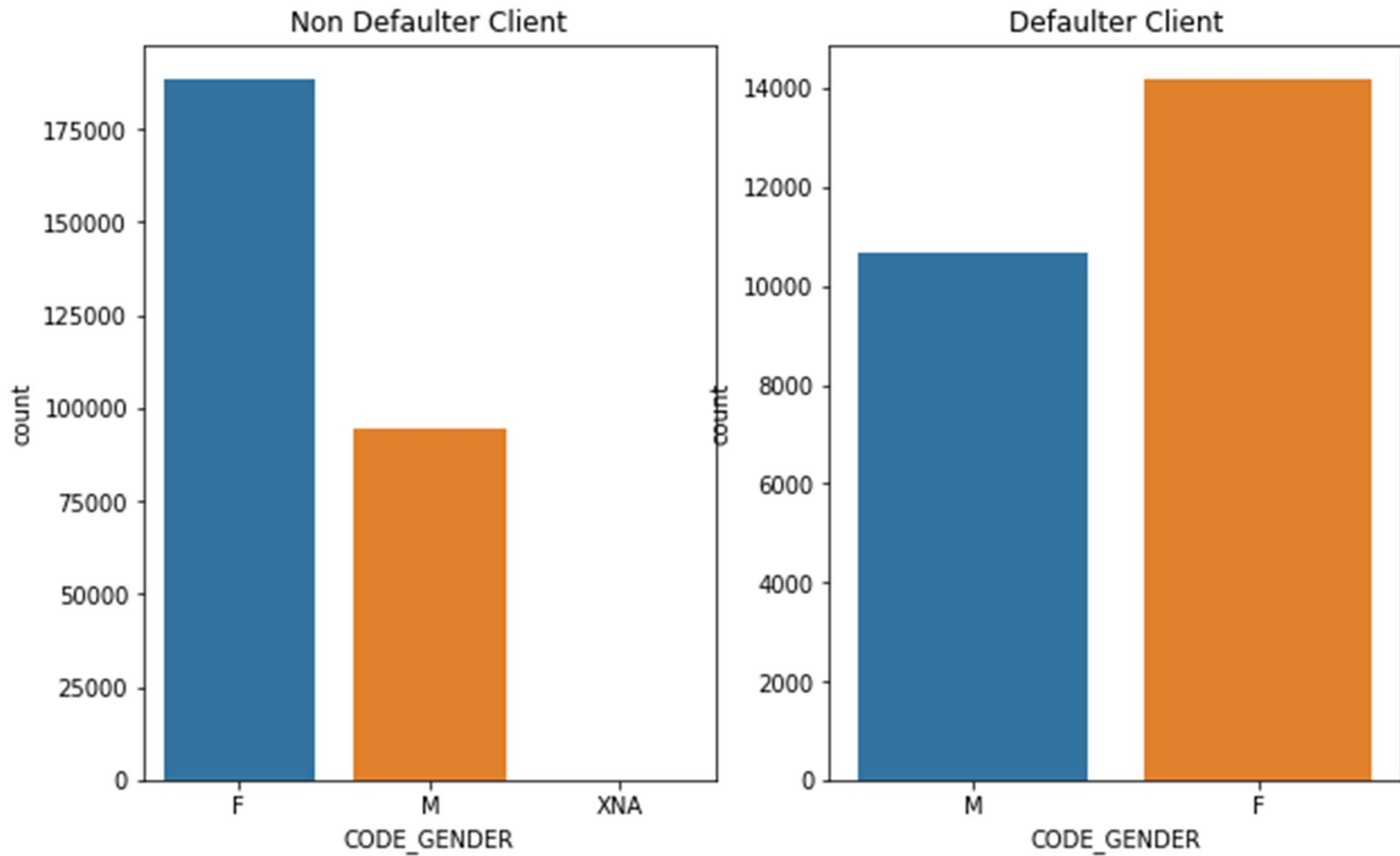




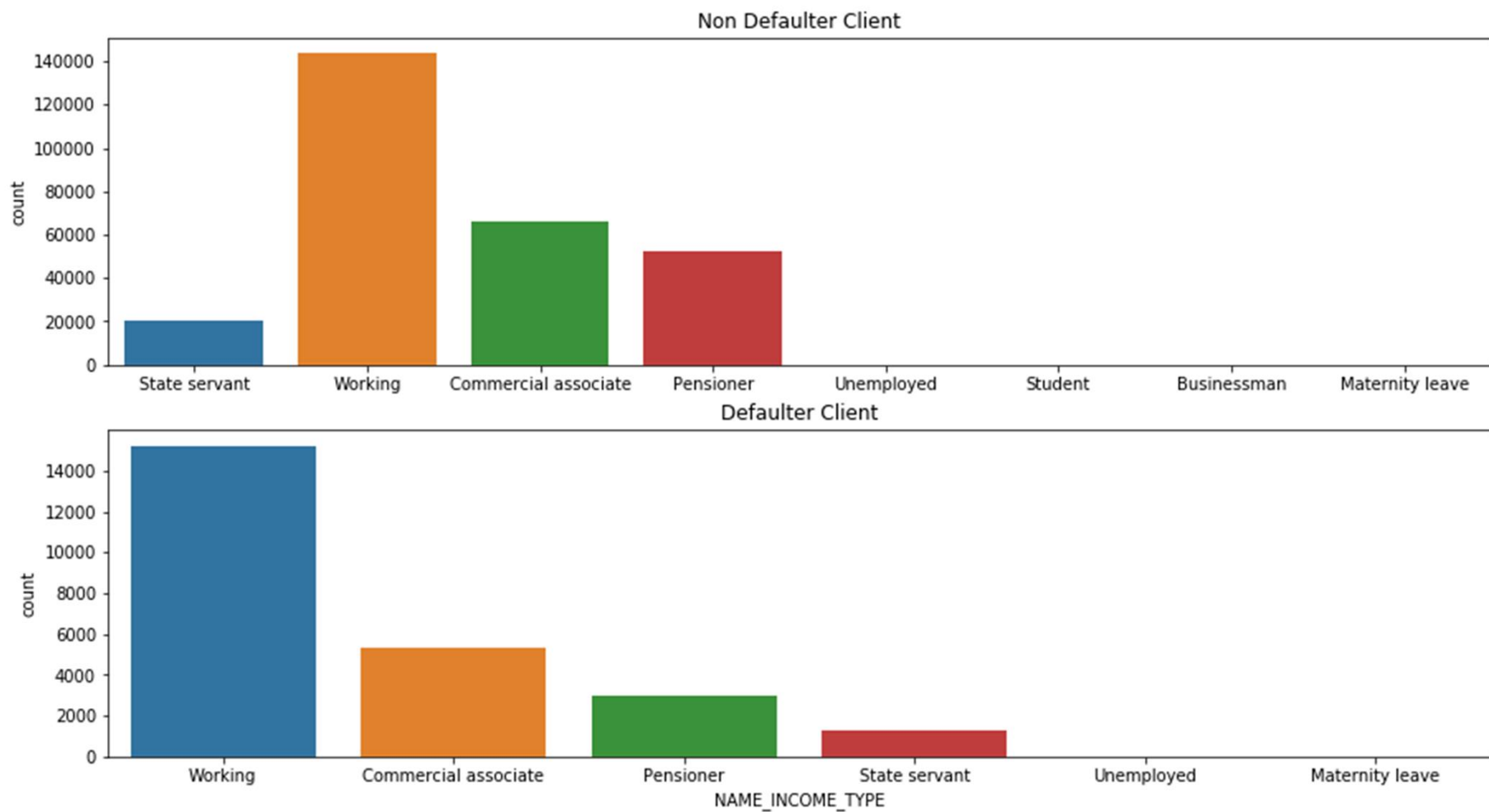
- On analysis of No. of Days Employed column for Target 0 i.e. Non-Defaulter and Target 1 i.e. Defaulter.
- It can be observed from the graph that most of the defaulter are not employed..



It can be observed after analysis of children count for Target 0 and Target 1 (Non-Defaulter and Defaulter) that most of the people who are defaulter are having no children.

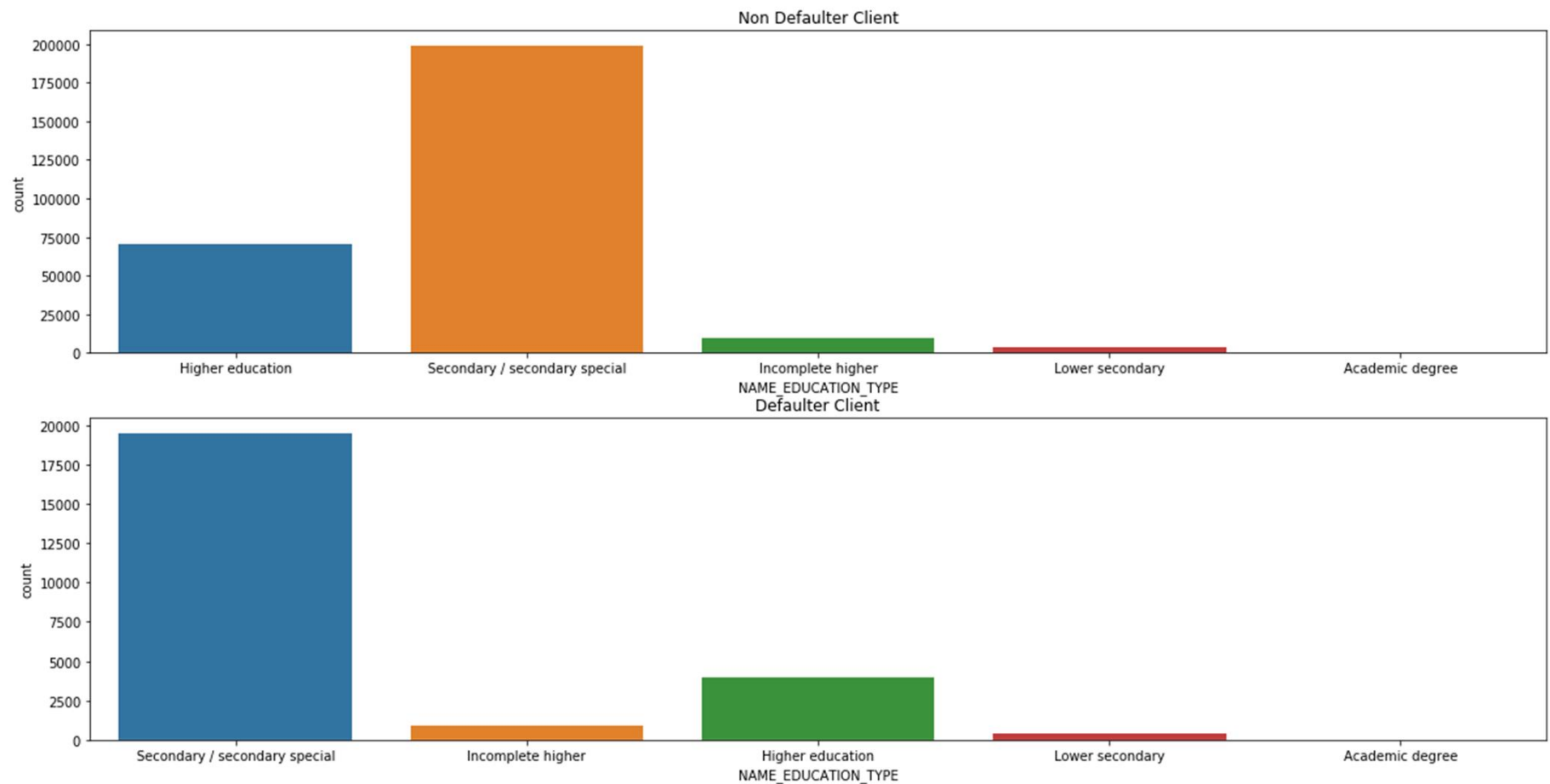


It can be observed that the more no. of Female applicants take loan than the male applicants. But max. percentage defaulters are Male (10.14% Male are defaulter)

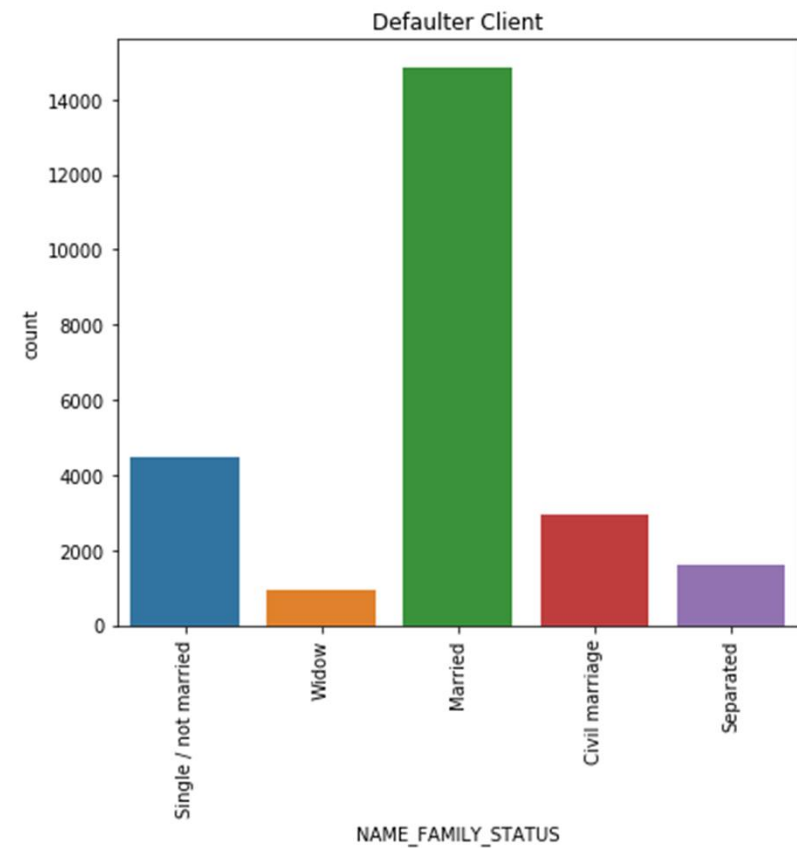
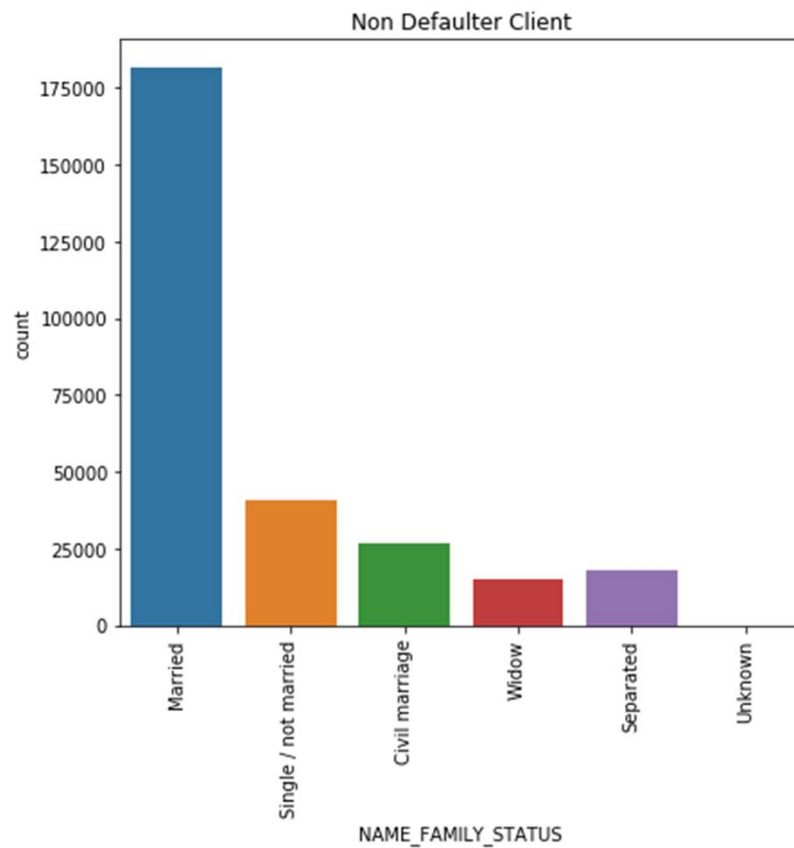


It is observed that very less student and businessman took loan. But it was observed that all the student and businessman who took loan have repaid it. So, it can be inferred that bank need to target Student and Businessman for loan.

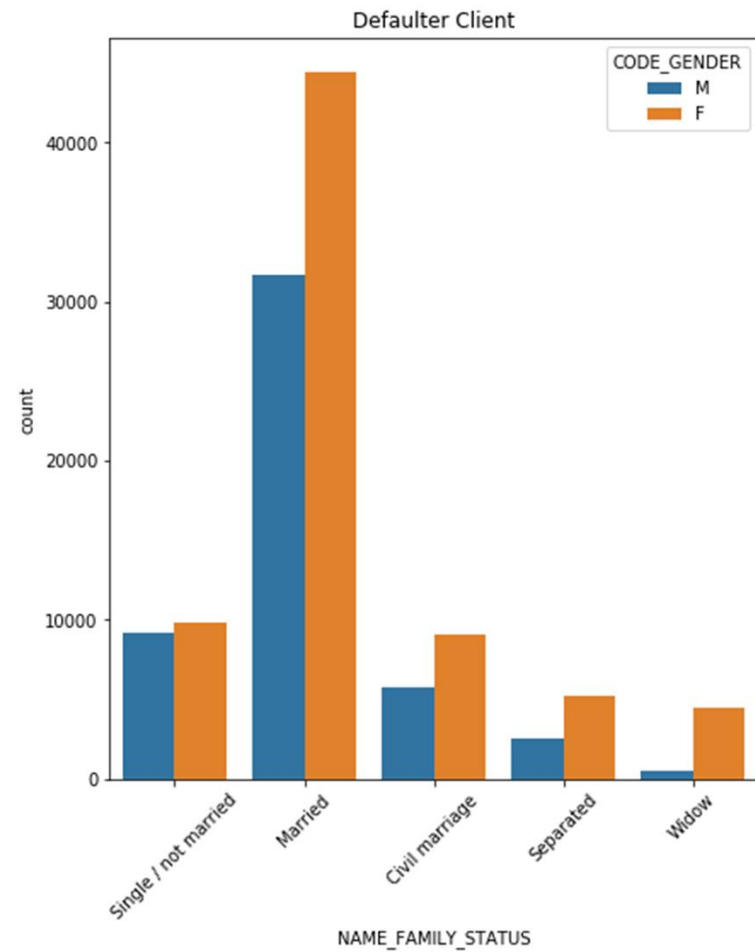
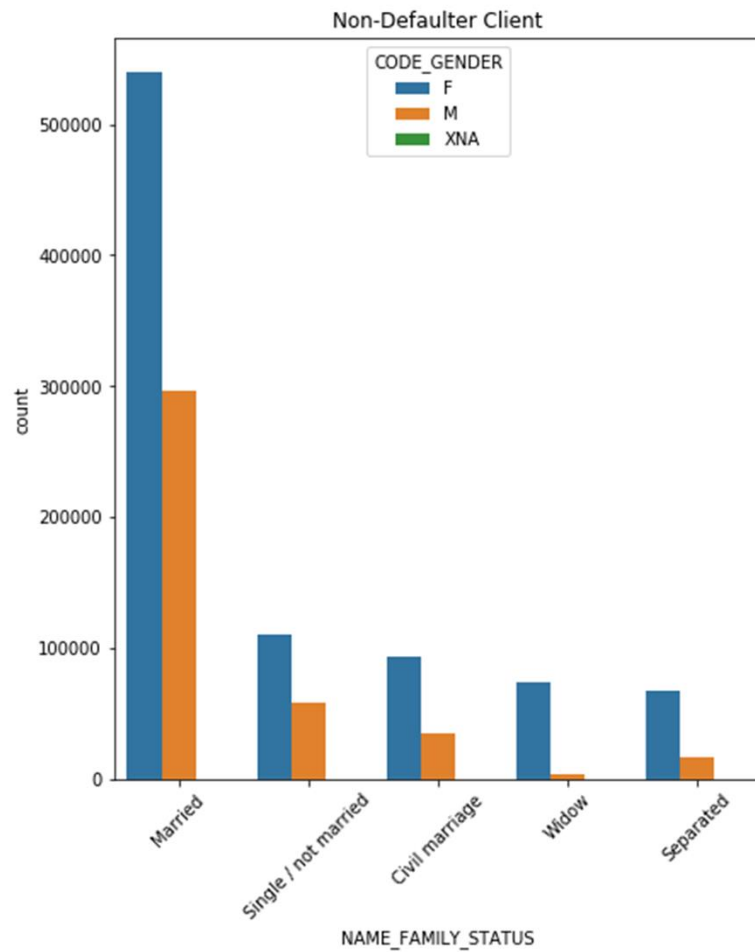




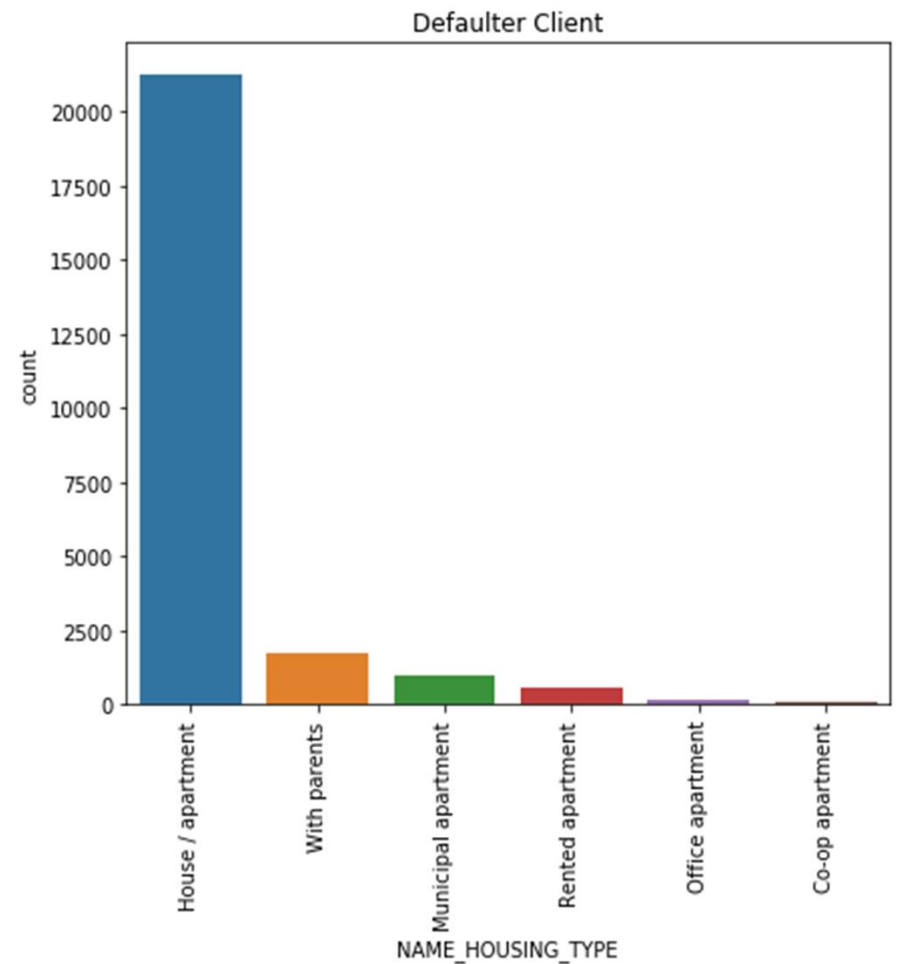
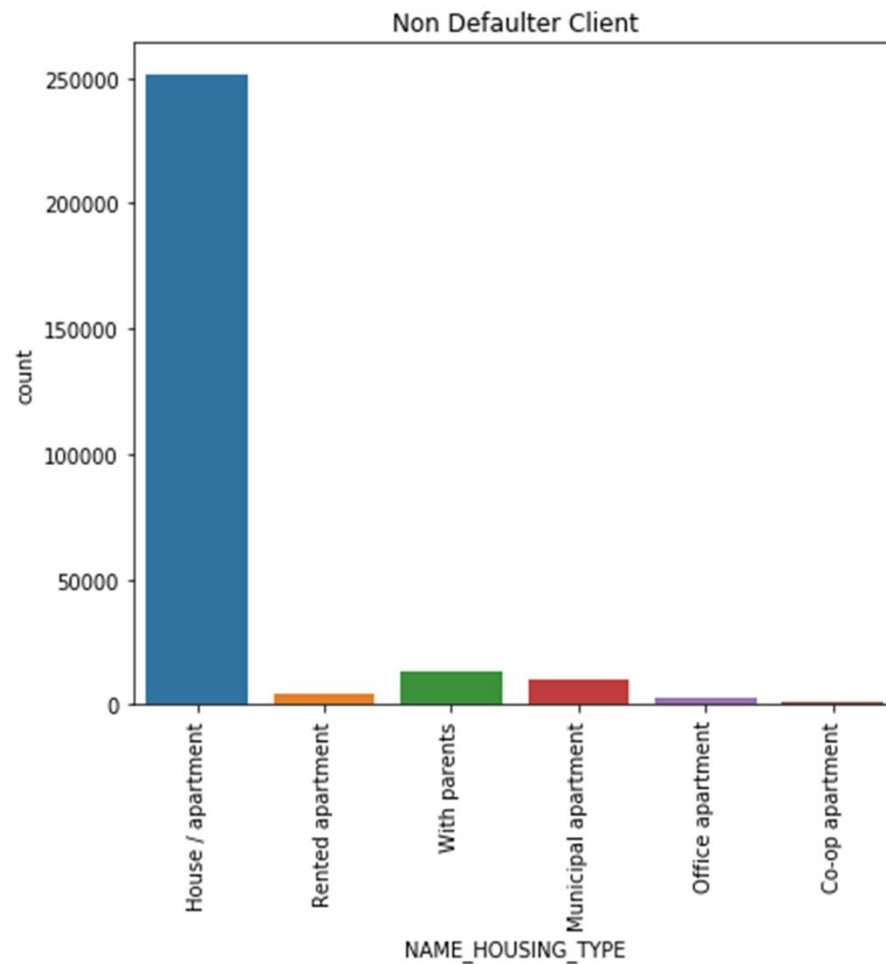
It can be observed that applicants having Secondary/secondary special education tends to take more loan.



It can be observed that married people tend to take more loans than other category.

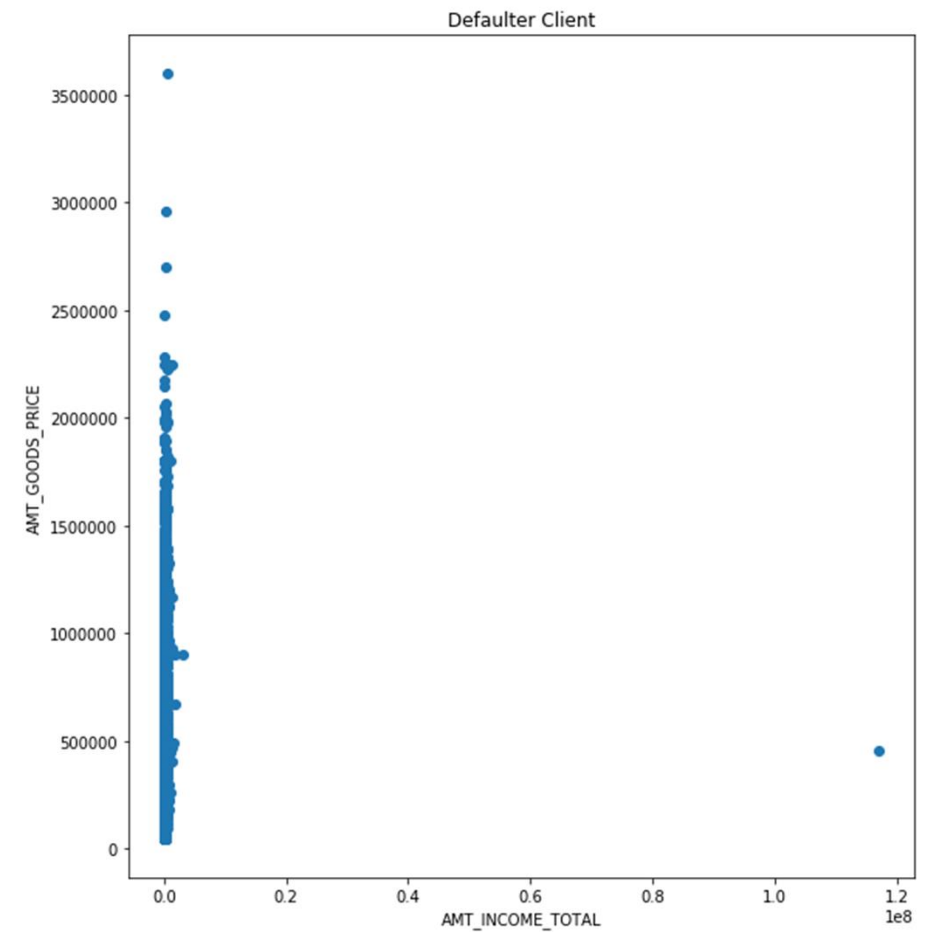
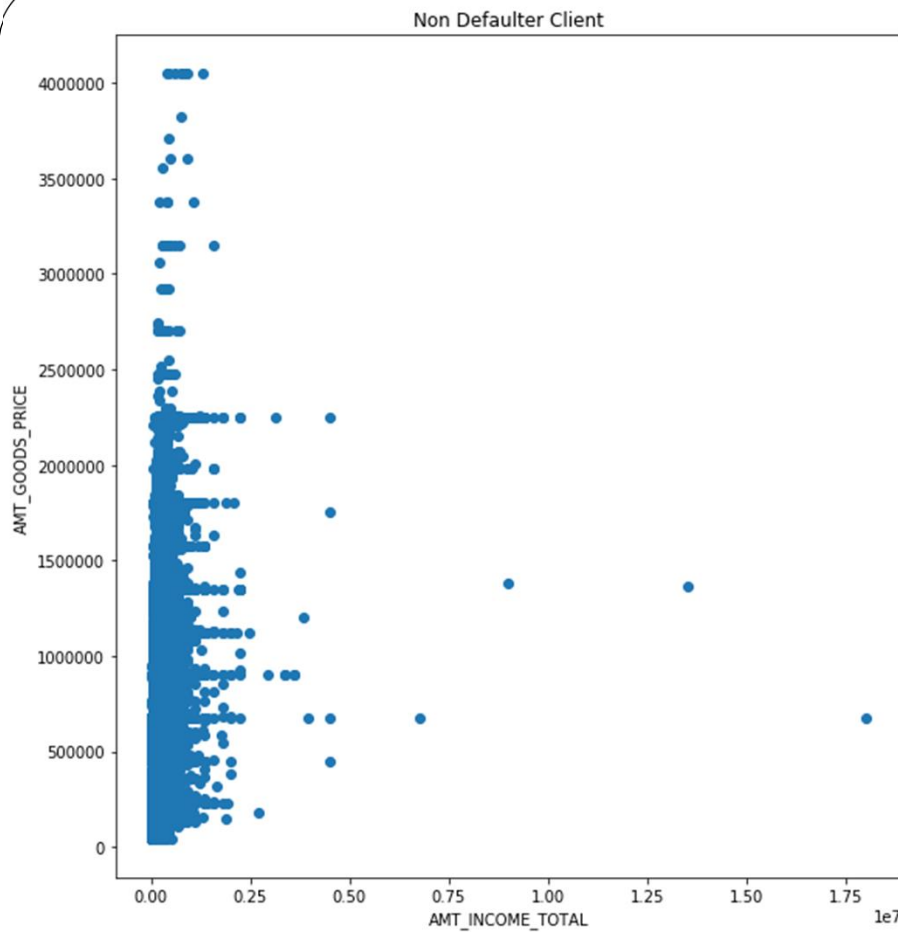


It can be observed that Single/Not Married Male most likely tend to be a defaulter.

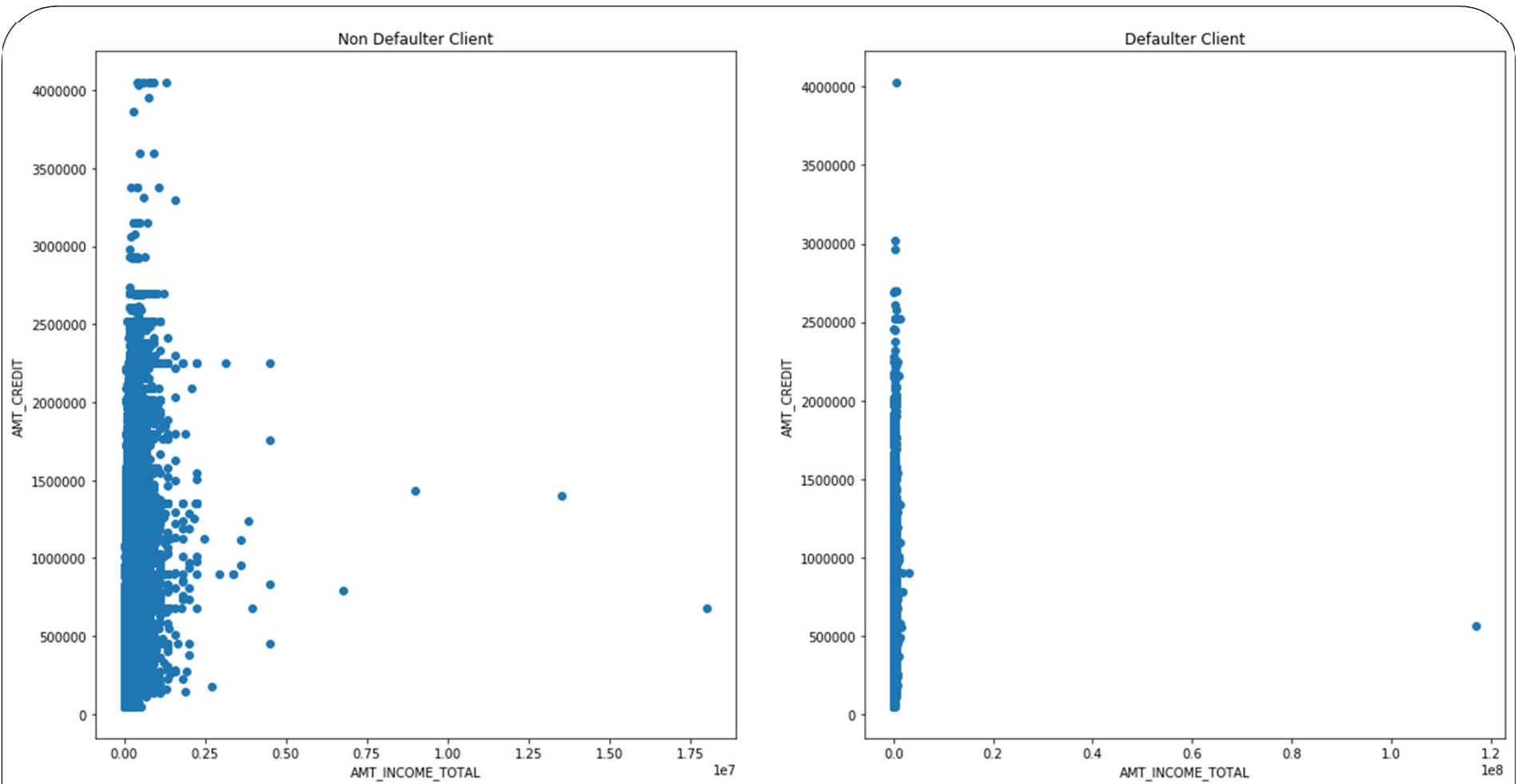


It can be observed that most of the applicant are from House/apartment category who are most likely to take loan.

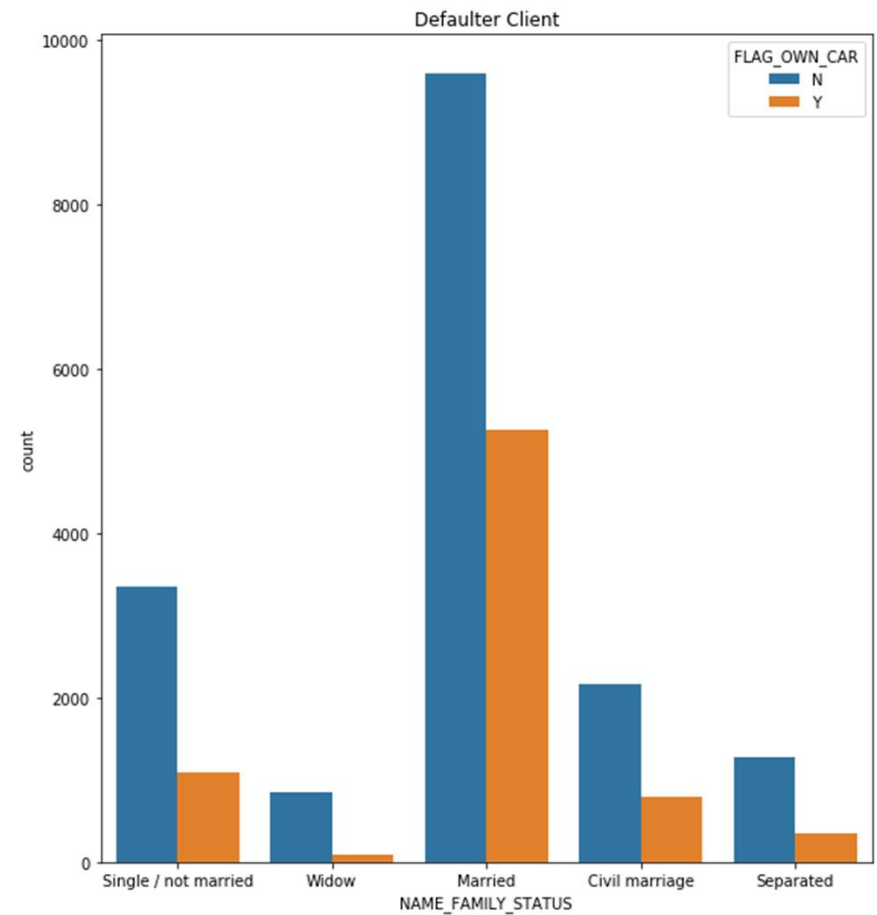
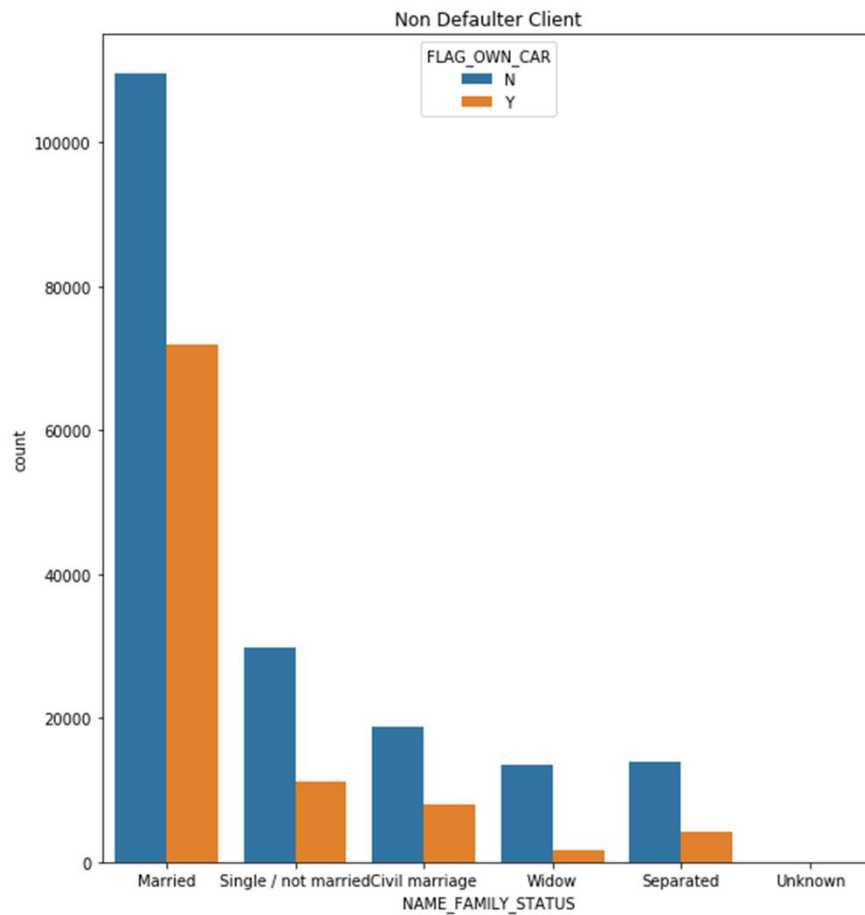
It can also be observed from the graph that more percentage of applicant from rented category who took loan tend to be Defaulter as the bar is high for rented apartment defaulter clients than for Non defaulter client.



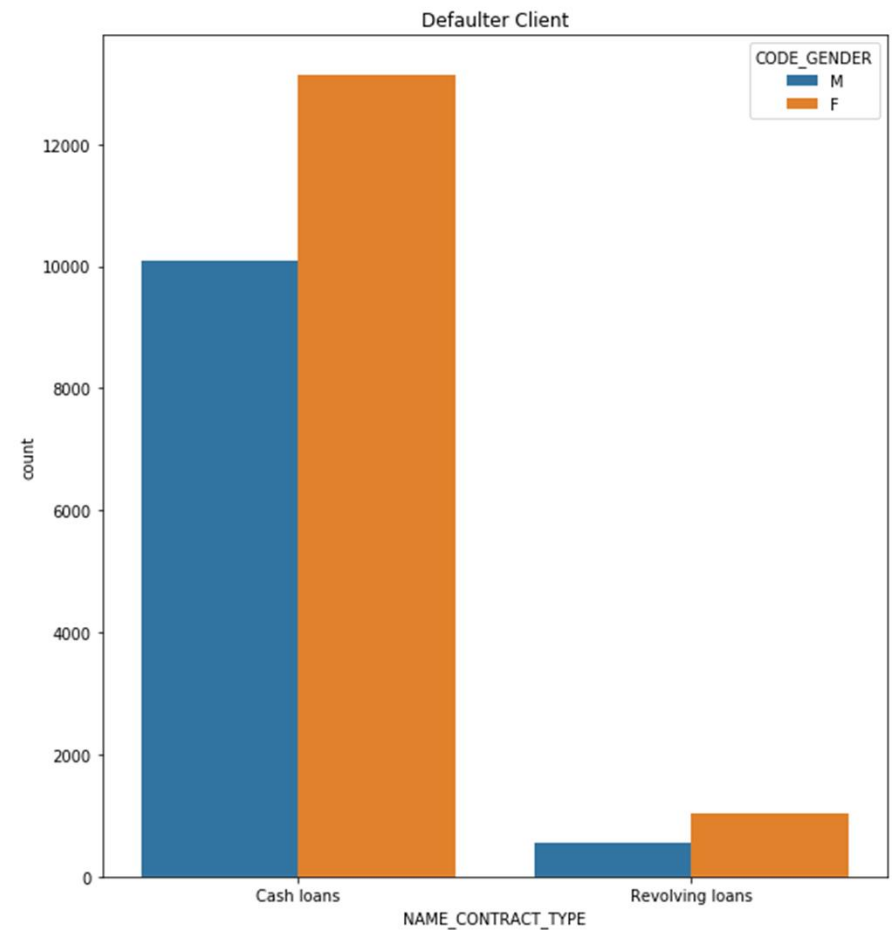
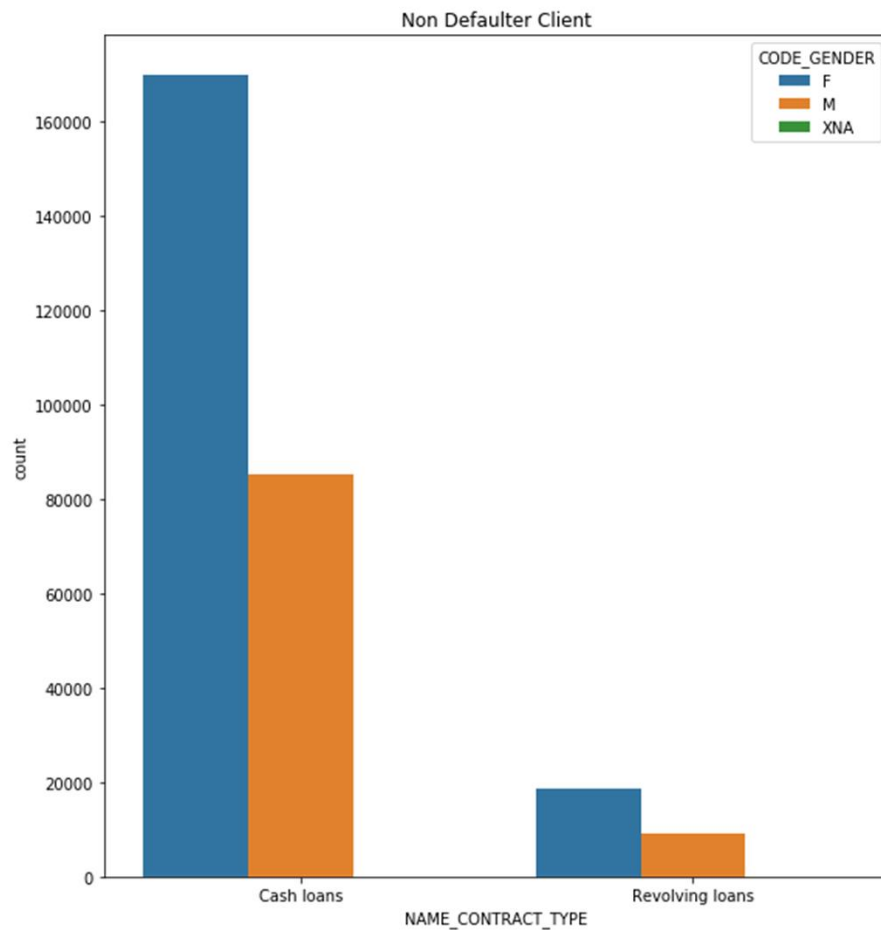
It can be observed that higher income applicant are less likely to take loan for "Goods".  
So, we can infer from the graph that as the income of applicant is less, they are more likely to take loan on Goods



It can be observed that applicant having more income are more likely to pay credit amount.

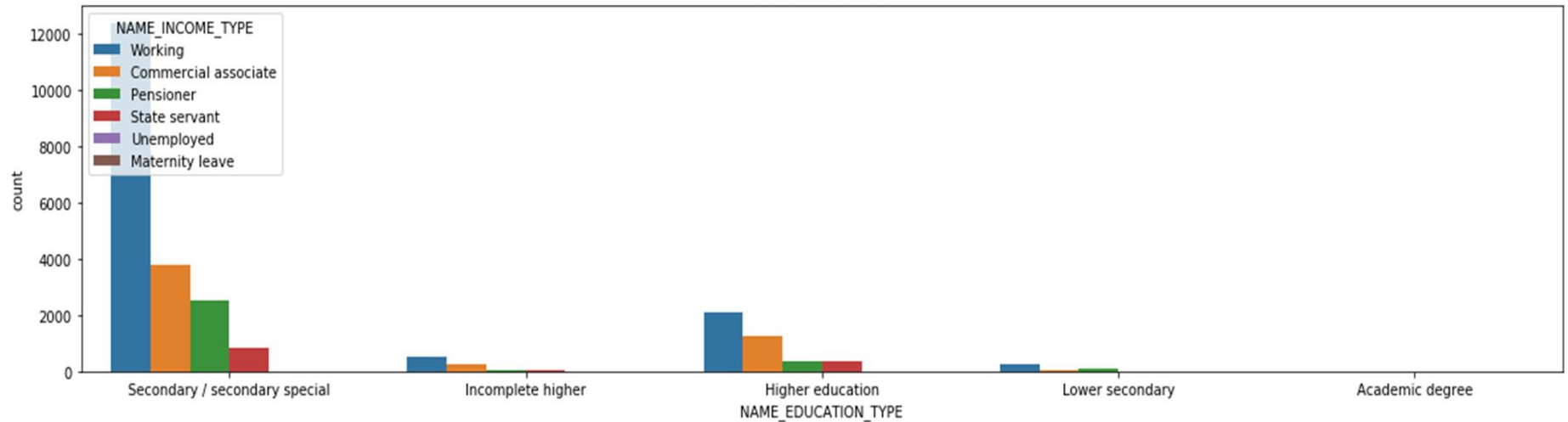
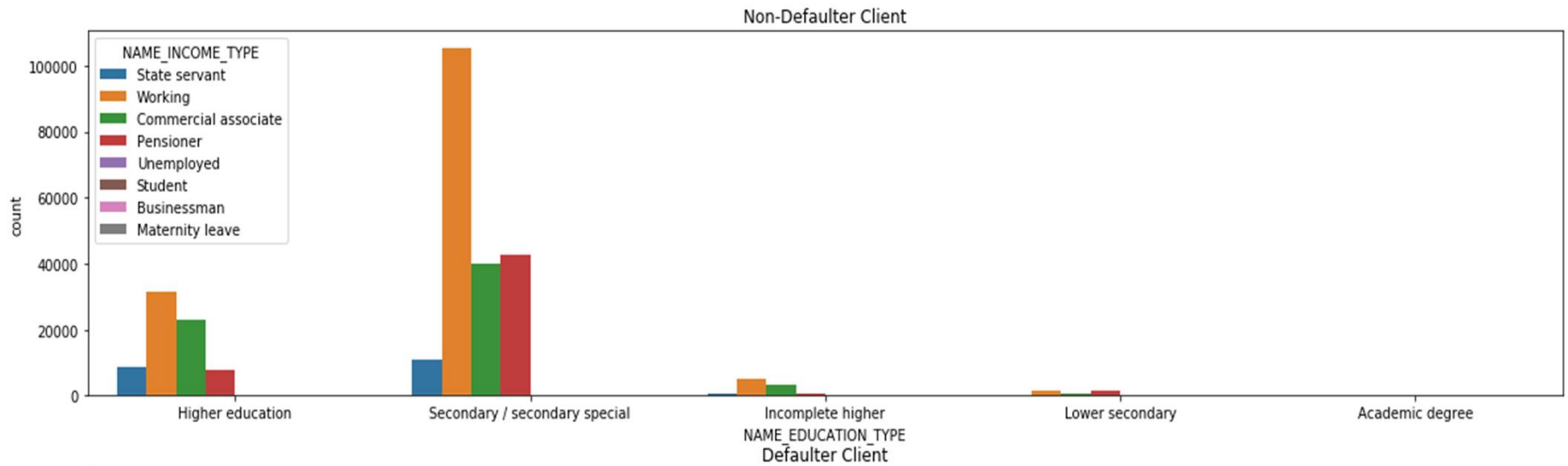


It can be observed that applicant from married category and not having cars are likely to take loans.



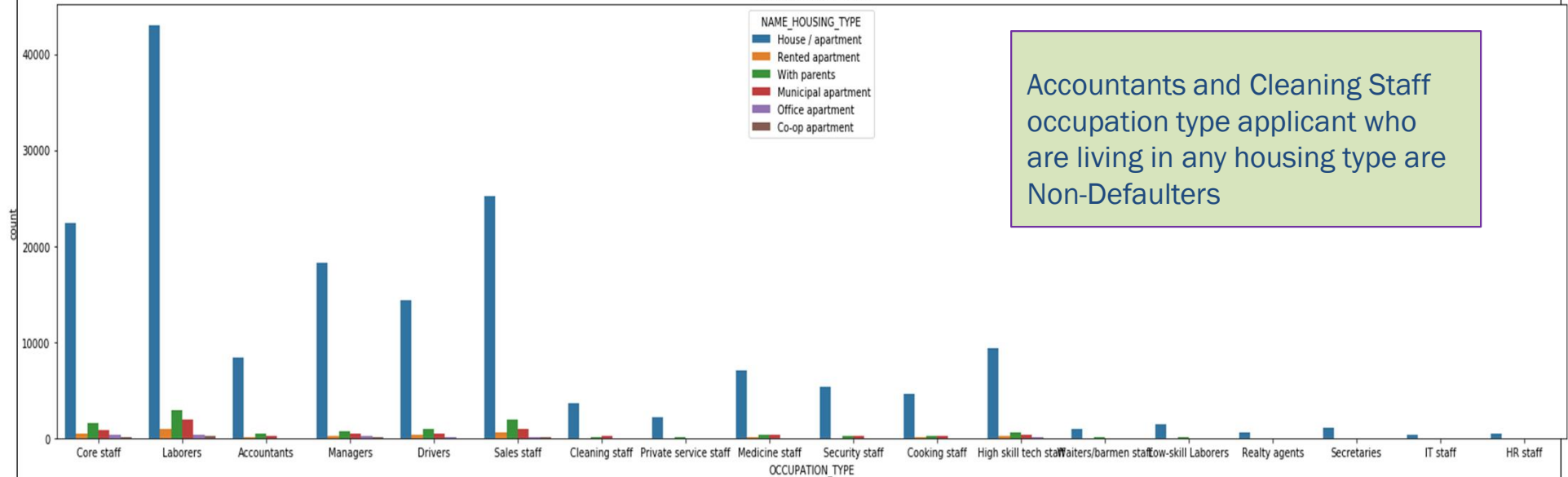
It can be infer from the above graph that Male category applicant taking cash loan are more likely to be defaulters in comparison to Female.





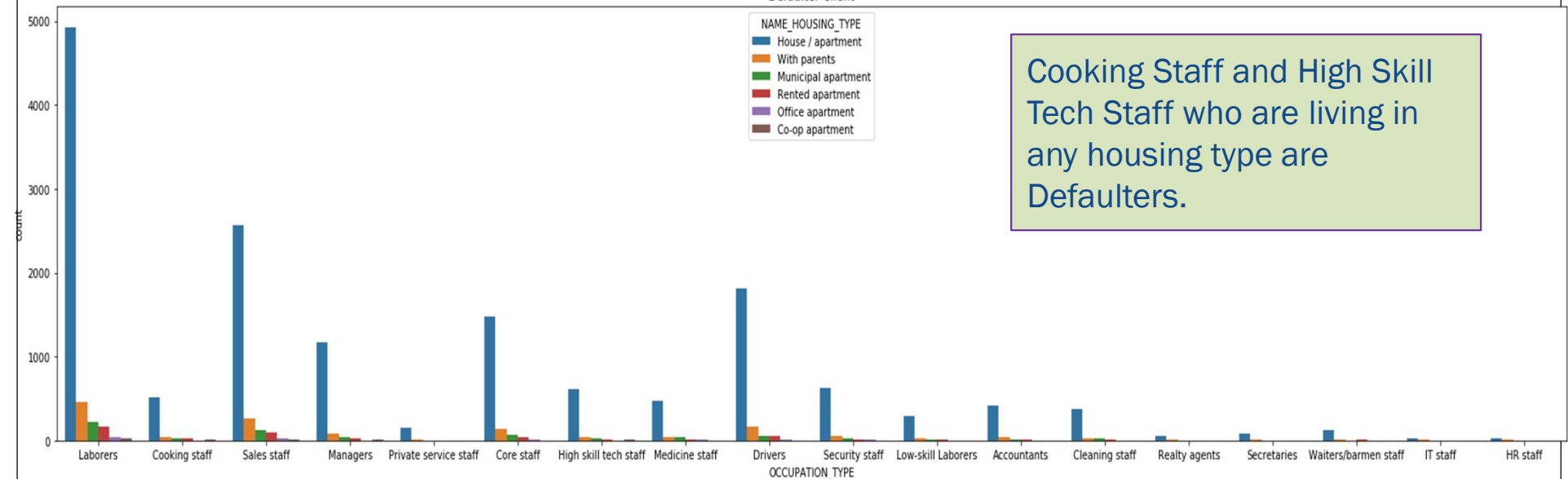
State Servant having Higher Education are less likely to be a defaulter.

Non-Defaulter Client

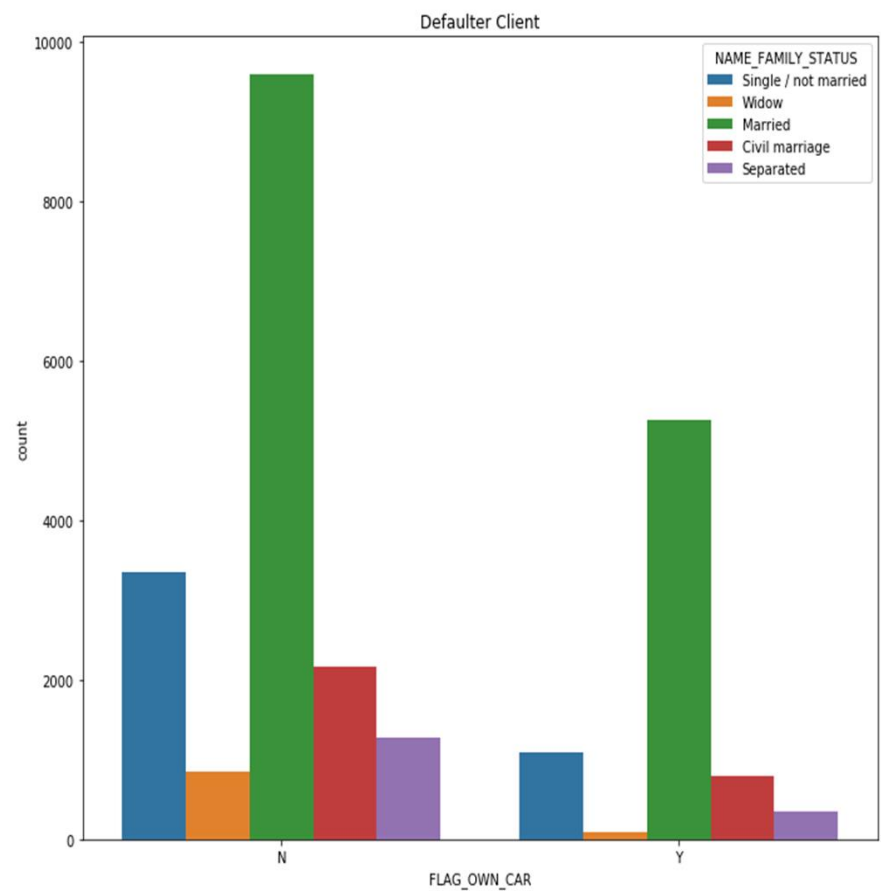
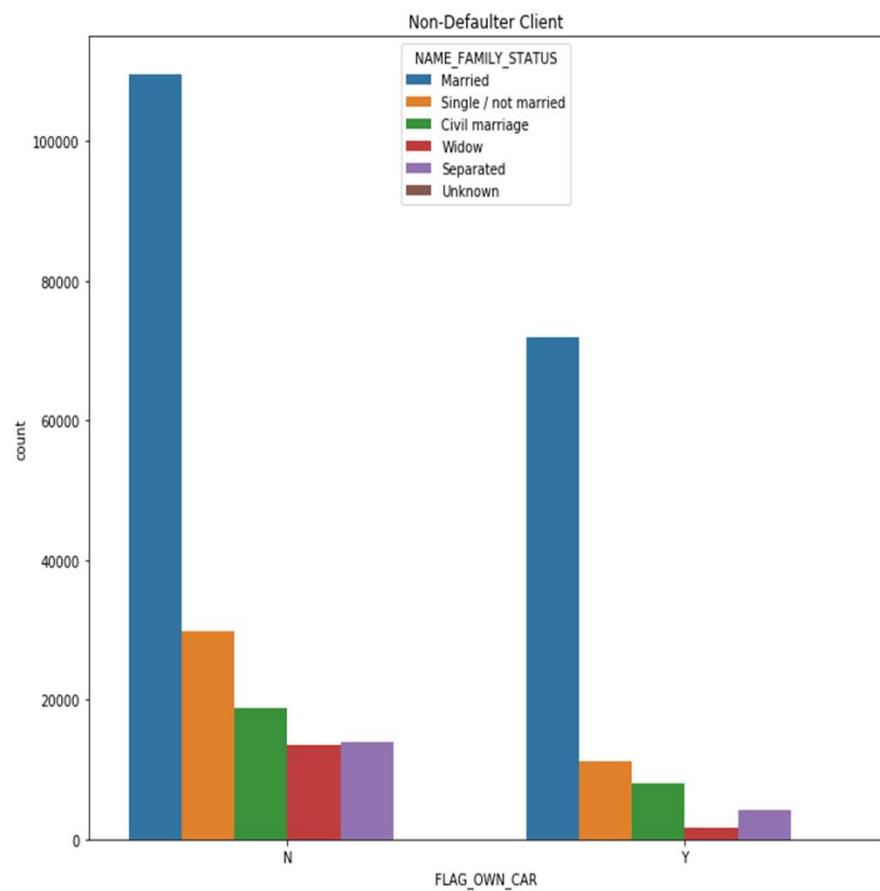


Accountants and Cleaning Staff occupation type applicant who are living in any housing type are Non-Defaulters

Defaulter Client

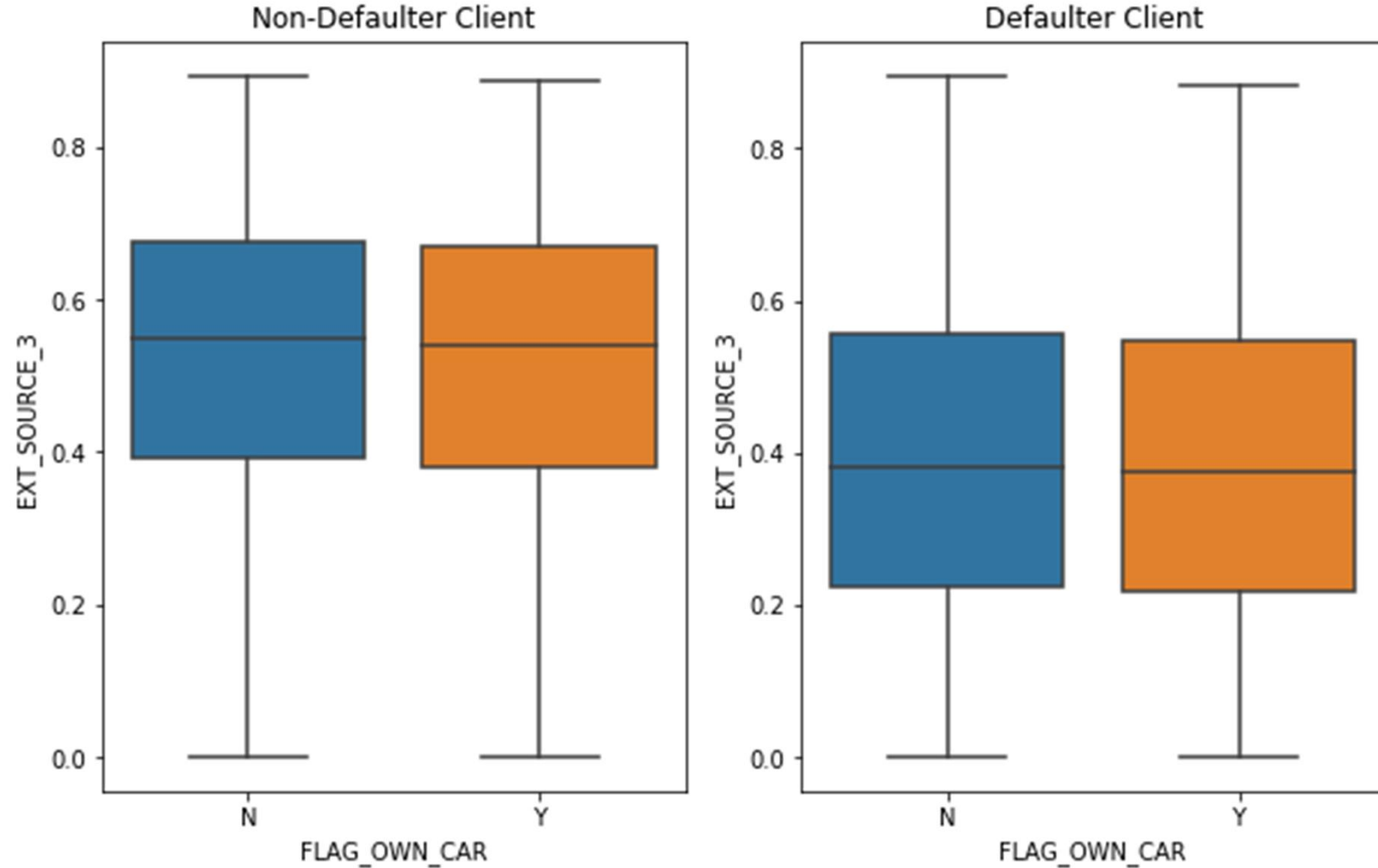


Cooking Staff and High Skill Tech Staff who are living in any housing type are Defaulters.

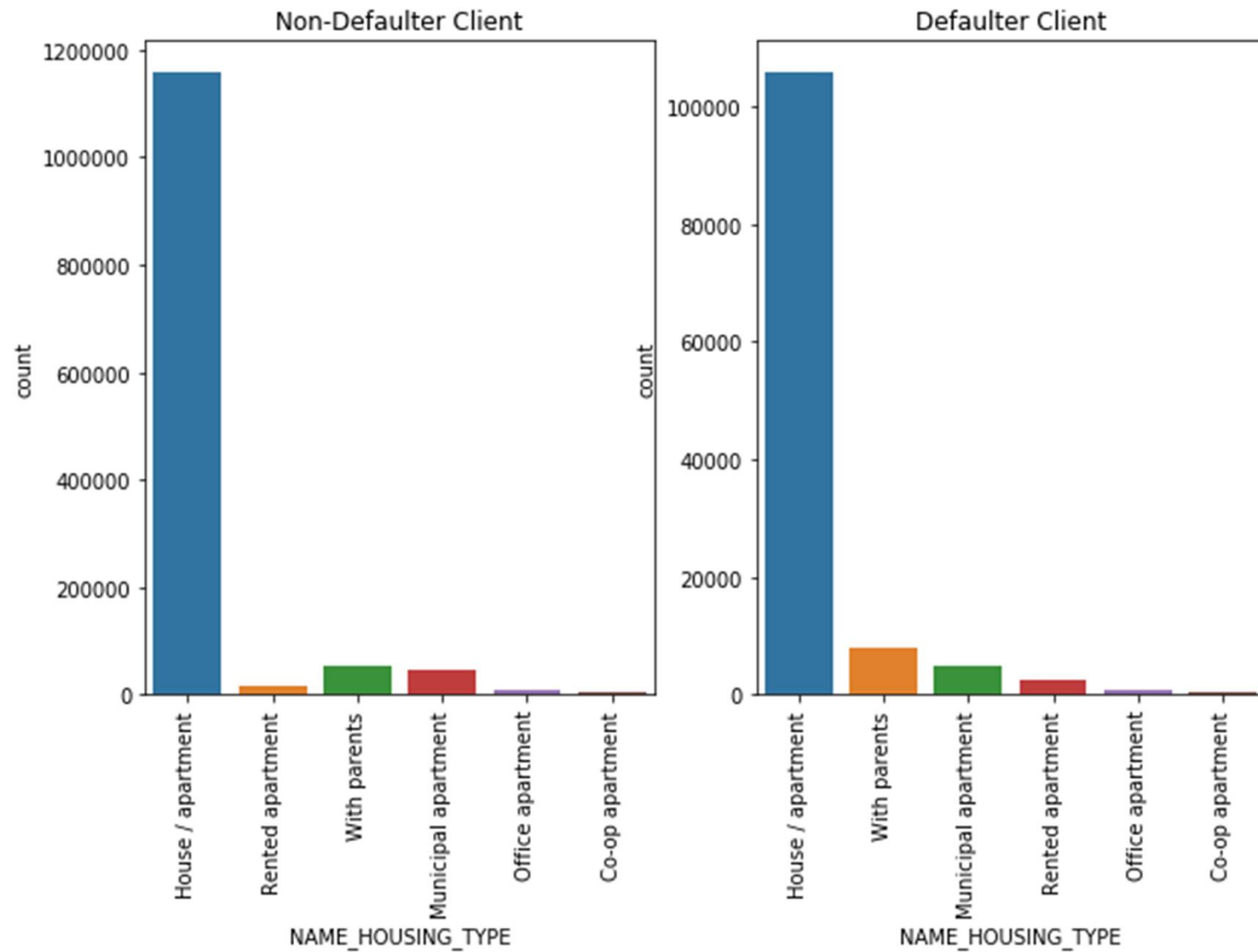


#### Observation:

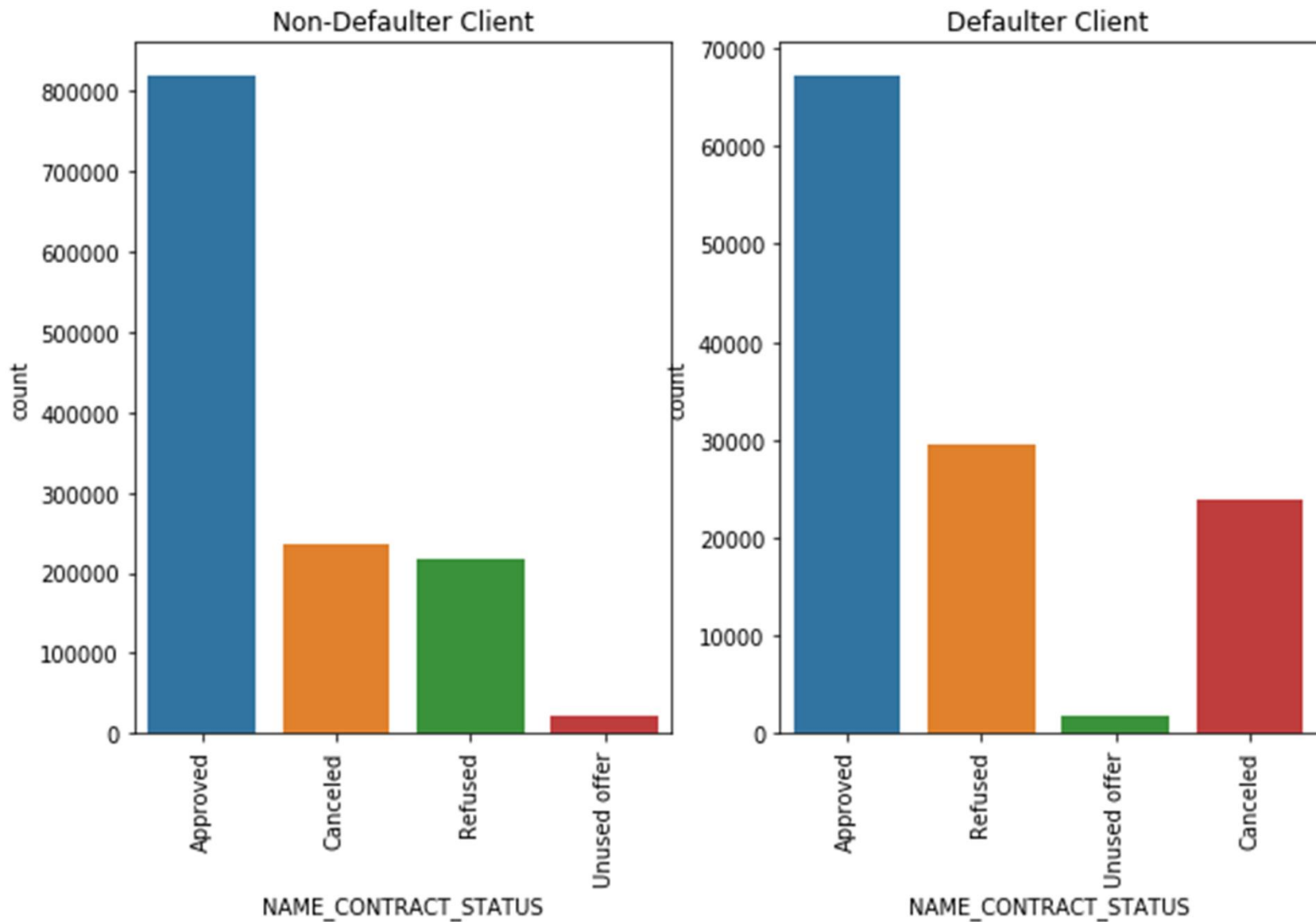
- It is observed that married applicant who don't own a car are most likely to be the defaulter.



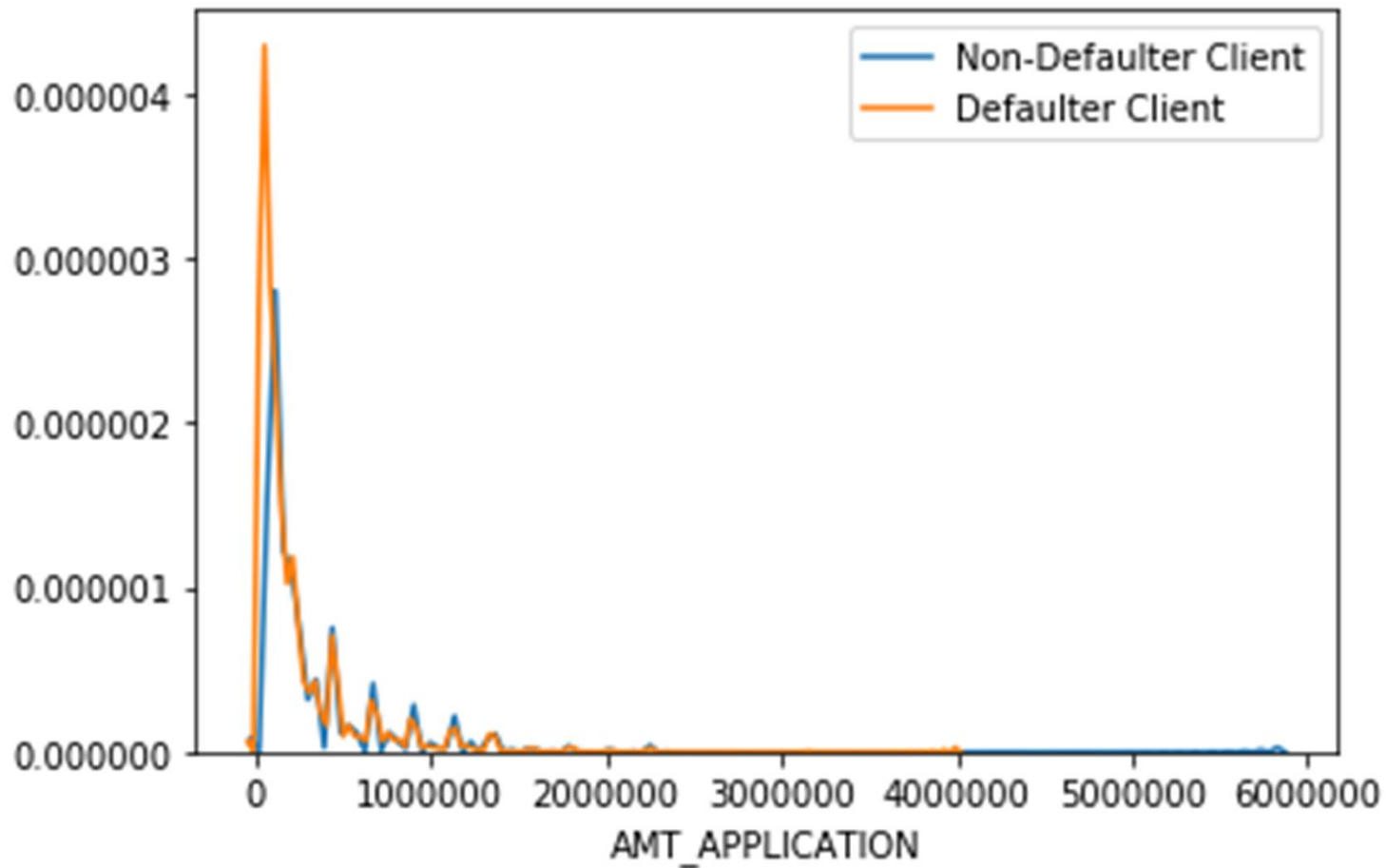
It can be observed that applicant who's EXT\_SOURCE\_3 rating is more than 0.4 are more likely to pay the loan irrespective of the condition that they own a car or not.



It can be observed that the applicant who live in rented apartment are most likely to be defaulter.



The applicant whose previous application status was refused tend to be defaulter.



More number of applicants apply for loan amount less than 1000000.  
The applicant who tend to be defaulter apply for loan of lesser amount.

# CONCLUSION

- Bank should focus on Married applicant, giving preference to Married Female having applicant as they are most likely to pay the loan.
- Bank should focus more on Student and Businessman to take loans, as they are most likely to return the loan.
- Accountants and Cleaning Staff occupation type applicant who are living in any housing type are most likely to be Non-Defaulters. Bank should focus on them.
- Cooking Staff and High Skill Tech Staff who are living in any housing type are most likely to be Defaulters. Bank should avoid giving loan to them.
- Bank should avoid giving loan to applicant living in rented apartment.