



# CREDIT EDA ASSIGNMENT

- ASHISH YADAV  
BATCH NO. - DS C49

# Introduction

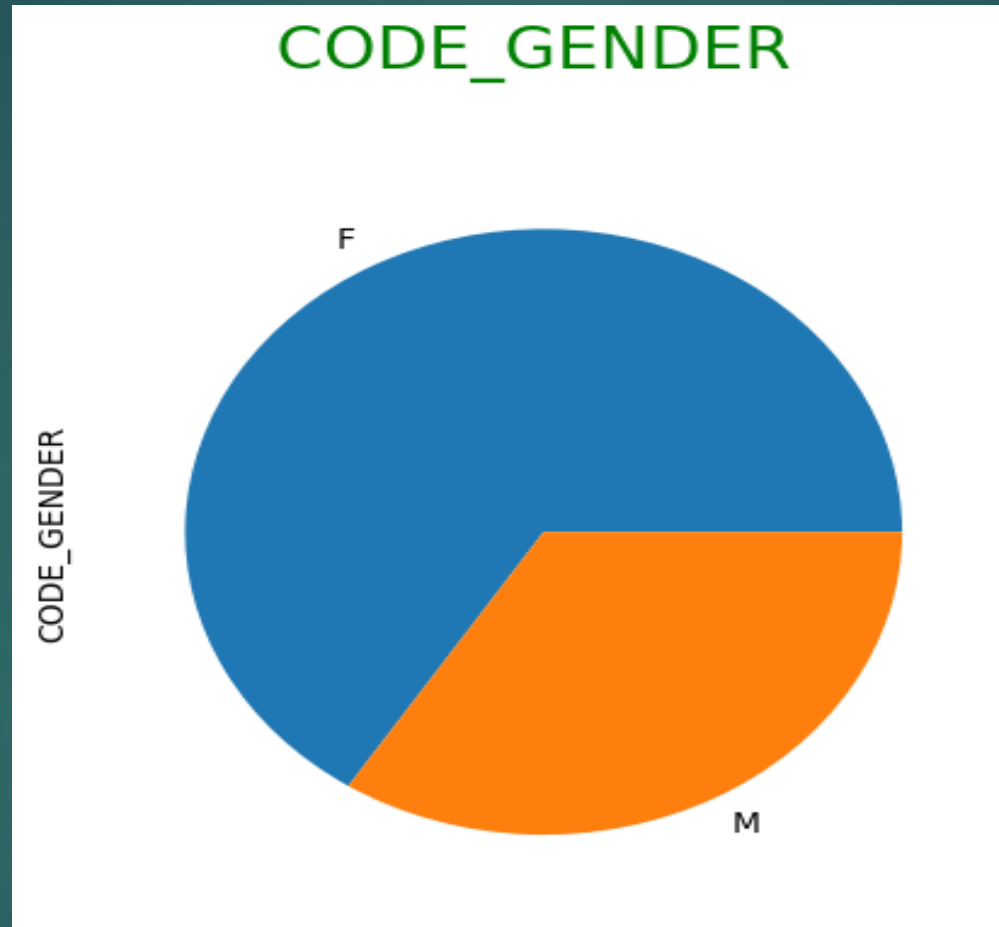
This assignment applies EDA in a real business scenario. In this assignment risk analytics in banking and financial services to understand how data is used to minimize the risk of losing money while lending to customers.

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it to their advantage by becoming a defaulter. Thus using EDA we analyze the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

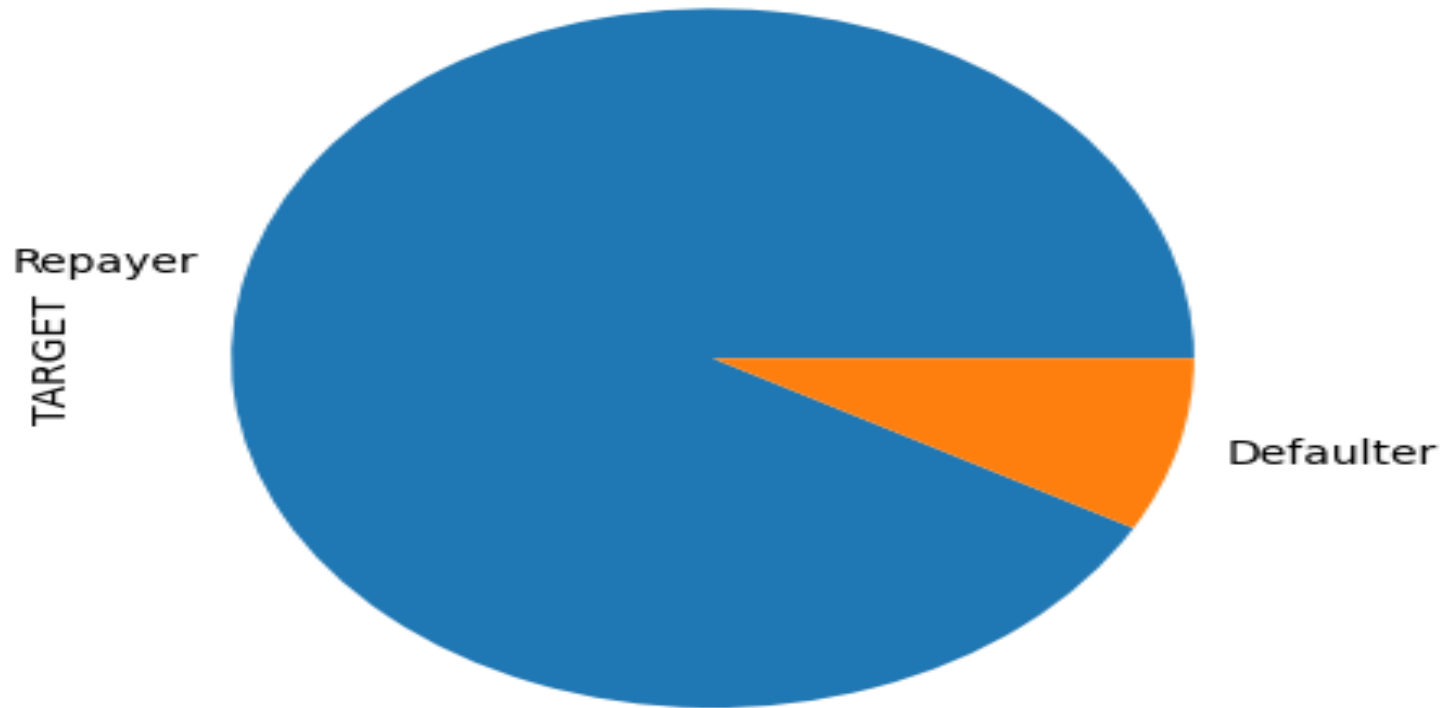
- 1) If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- 2) If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

# Univariate analysis



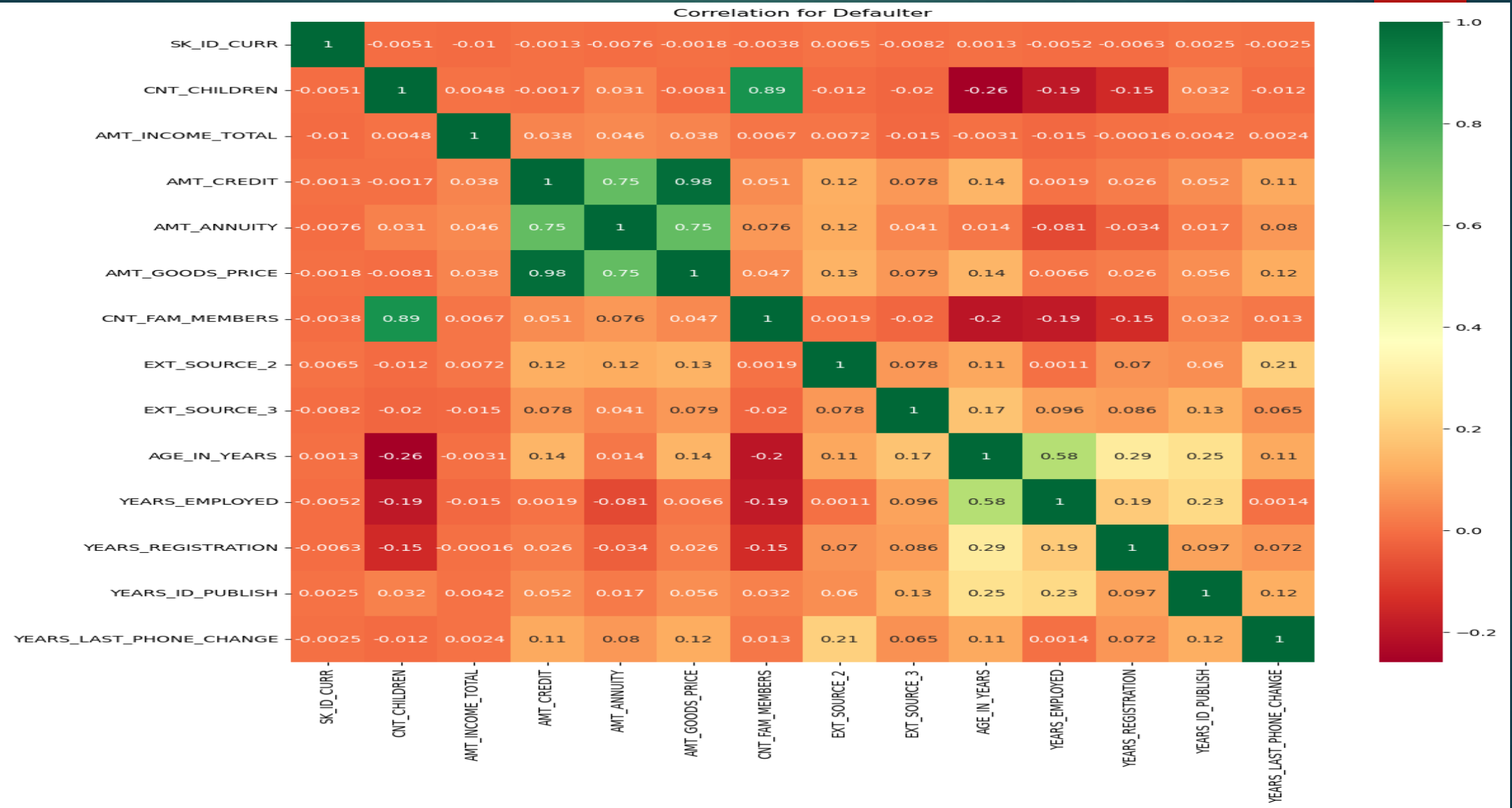
IT GIVES A CLEAR VIEW THAT MORE NO. OF FEMALES ARE APPLYING FOR LOAN AS COMPARED TO MALES.

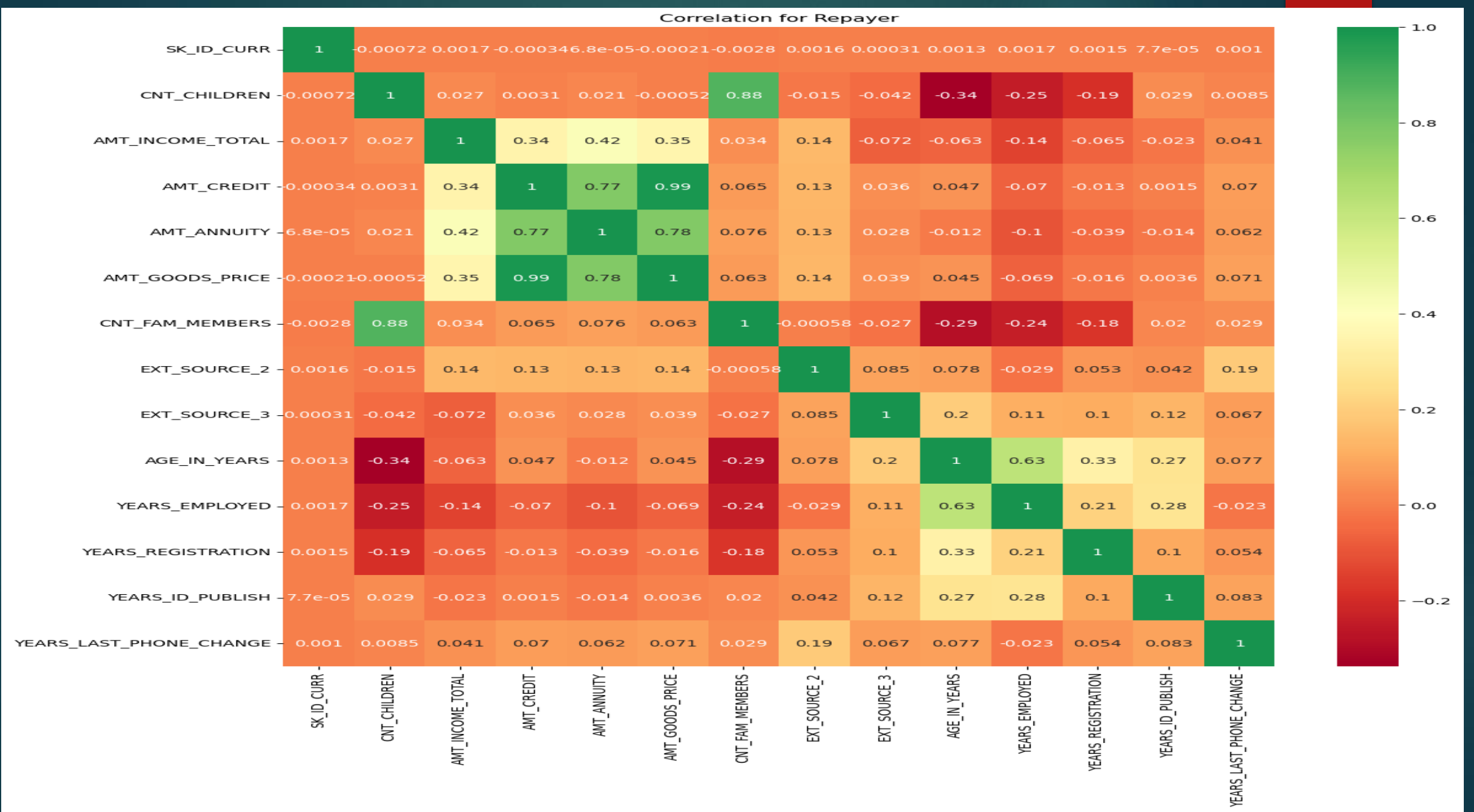
## Repayer vs Defaulter



There is only small group of the person who defaults the loan amount.

# Correlation



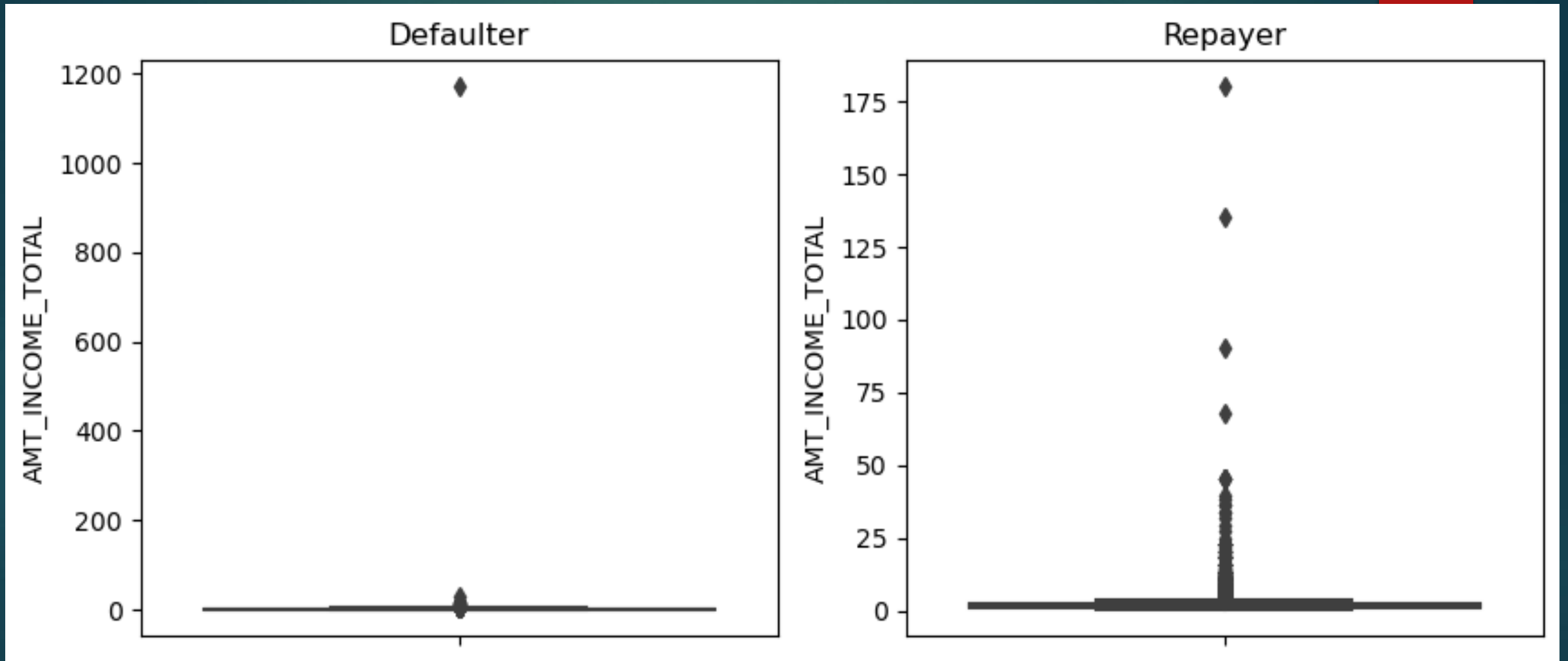


# Insight –

AMT\_INCOME\_TOTAL,  
AMT\_ANNUITY,  
AMT\_GOODS\_PRICE,  
AMT\_CREDIT,  
AGE\_IN\_YEARS,  
YEARS\_EMPLOYED

Above mentioned columns are highly correlated for both the case where client is either defaulter or repayer.

# Outliers



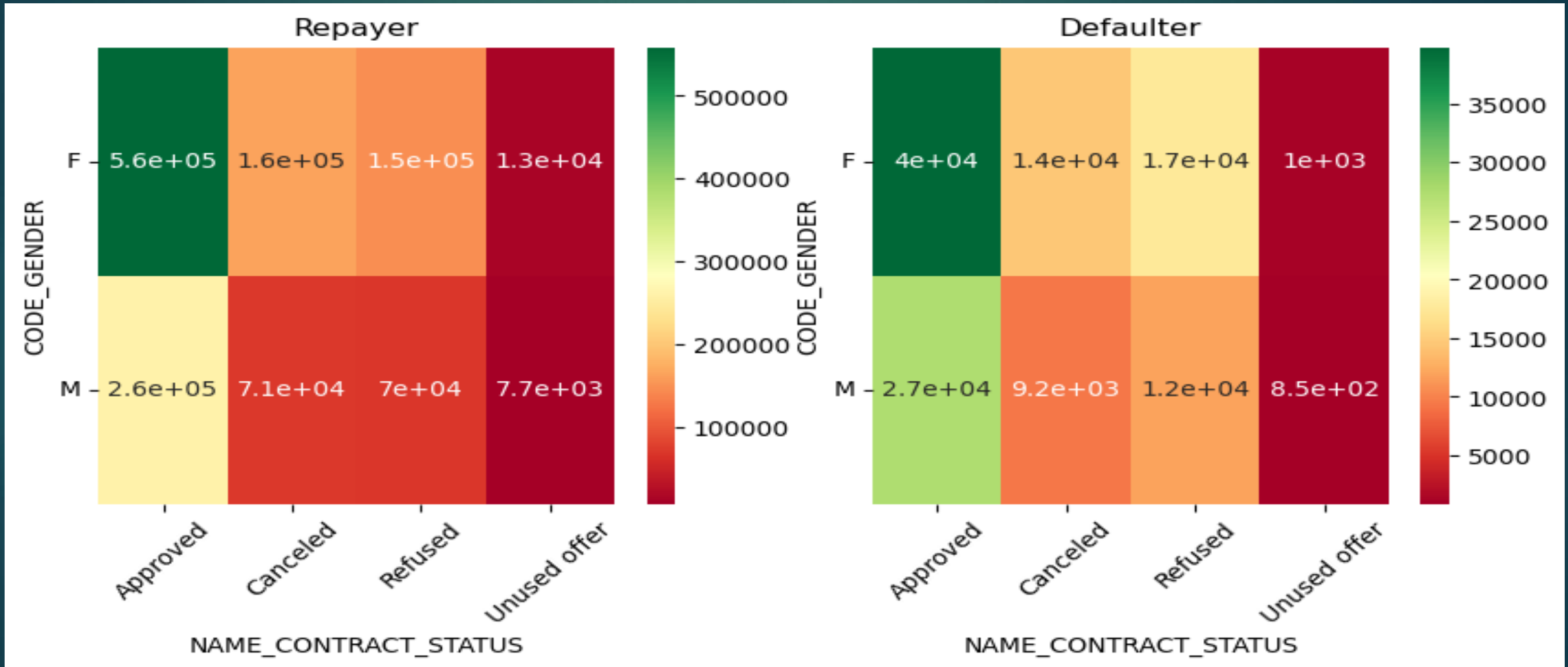
Insight - Both the columns contain outliers but no. of outliers for Repayer category is more as compared to the Defaulter.



# Analysis after merging two dataframes –

- 1) Application\_data and
- 2) Previous\_Application\_Data

# Correlation Between Gender , Status of loan application and status of client -



# Insight -

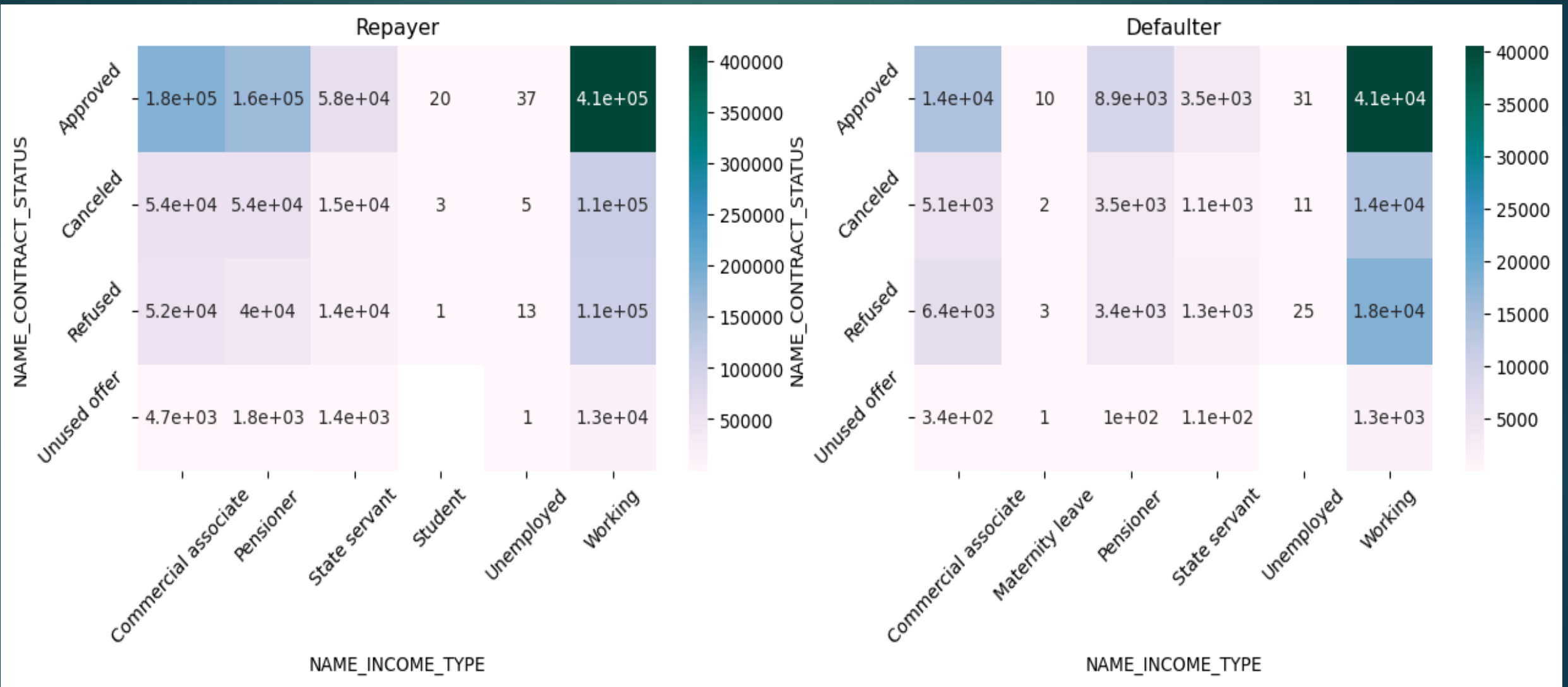
## Repayer -

- ▶ Loan application status of females is approved in large numbers,
- ▶ while Males have high number of unused offer.

## Defaulter -

- ▶ Females have least number of unused offer,
- ▶ while Males have high number of cancelled application status.

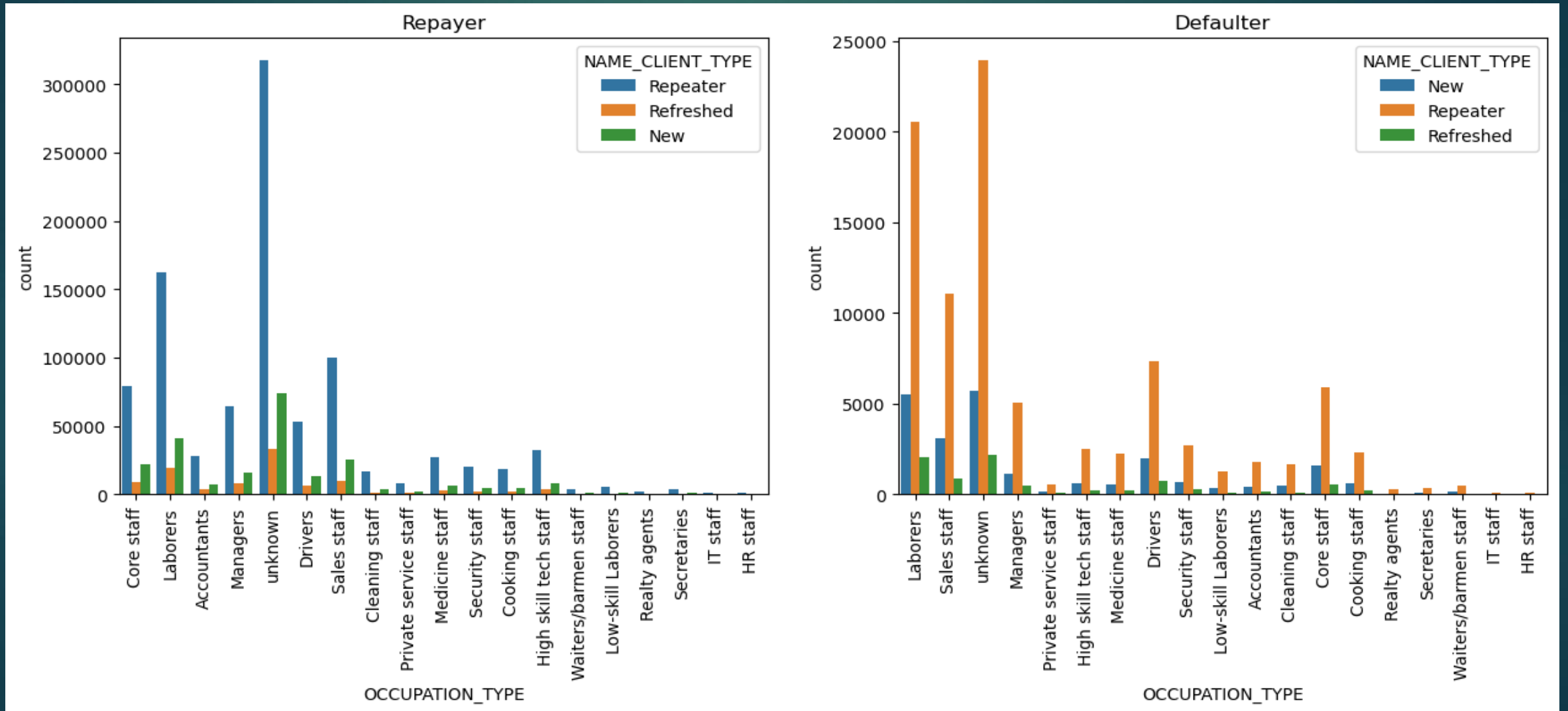
# Correlation between Income type, Status of loan application and Status of client -



## Insight -

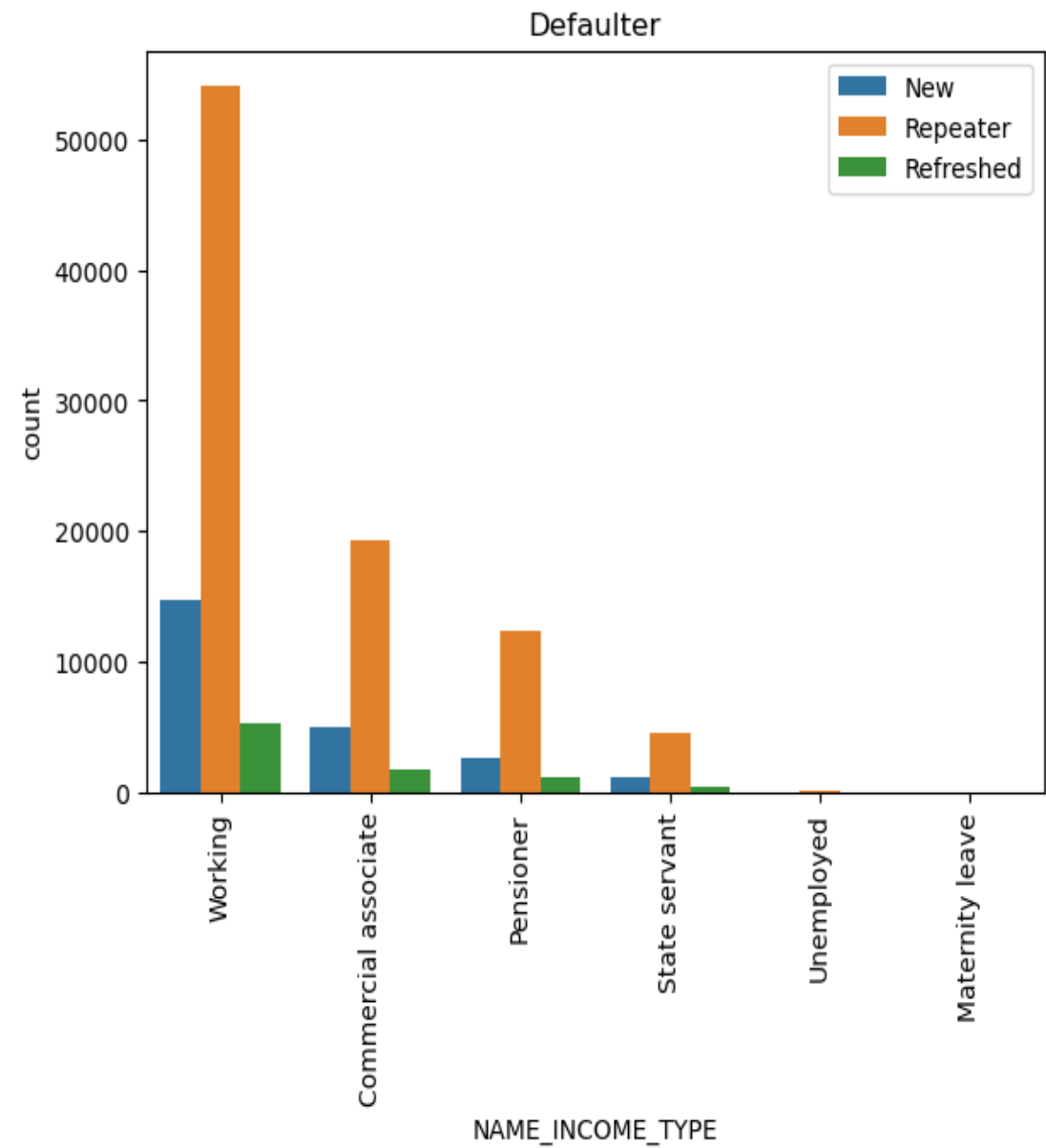
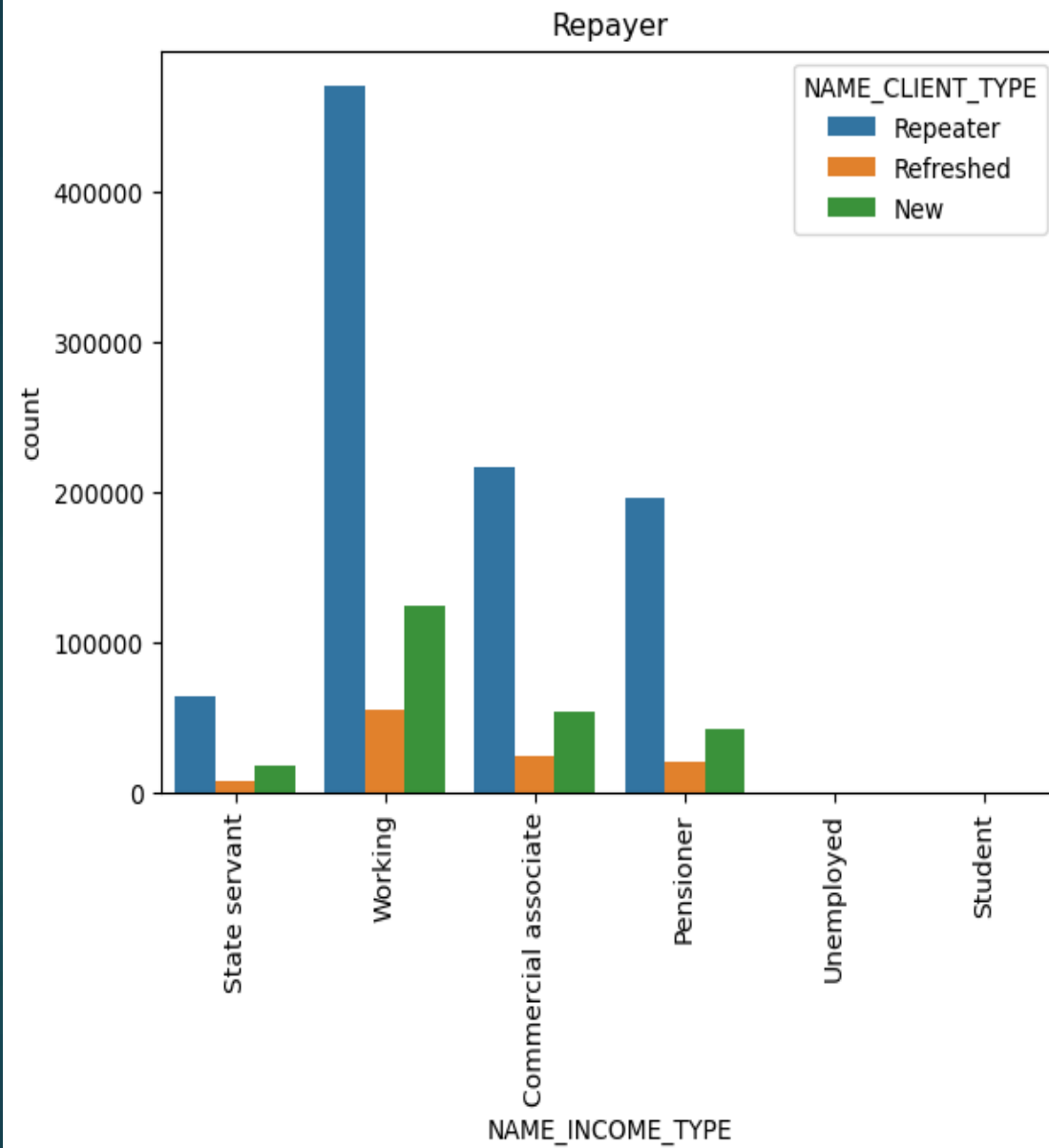
- ▶ Most number of clients are from Working class which are defaulter and Repayer both.
- ▶ There is no correlation between student vs unused offer and unemployed vs unused offer.

# Multivariate analysis -



# Insight –

- ▶ Unknown occupation type clients are in the highest in numbers in both Repayer and Defaulter list.
- ▶ More number of Repeater clients are Defaulter and Repayer as compared with New and Refreshed clients





# Insight -

- ▶ Working and commercial class have the highest number of repeaters as compared to other categories in both Defaulter and repayer list.
- ▶ Student and maternity leave clients are lowest among the category.

# Conclusion –



- ▶ State servants are better at repaying loan and pension clients tend to default more.
- ▶ Most number of clients are from Working class which are defaulter and Repayer both.
- ▶ Females tend to become less defaulter than males.
- ▶ More number of Repeater clients are Defaulter and Repayer as compared with New and Refreshed clients