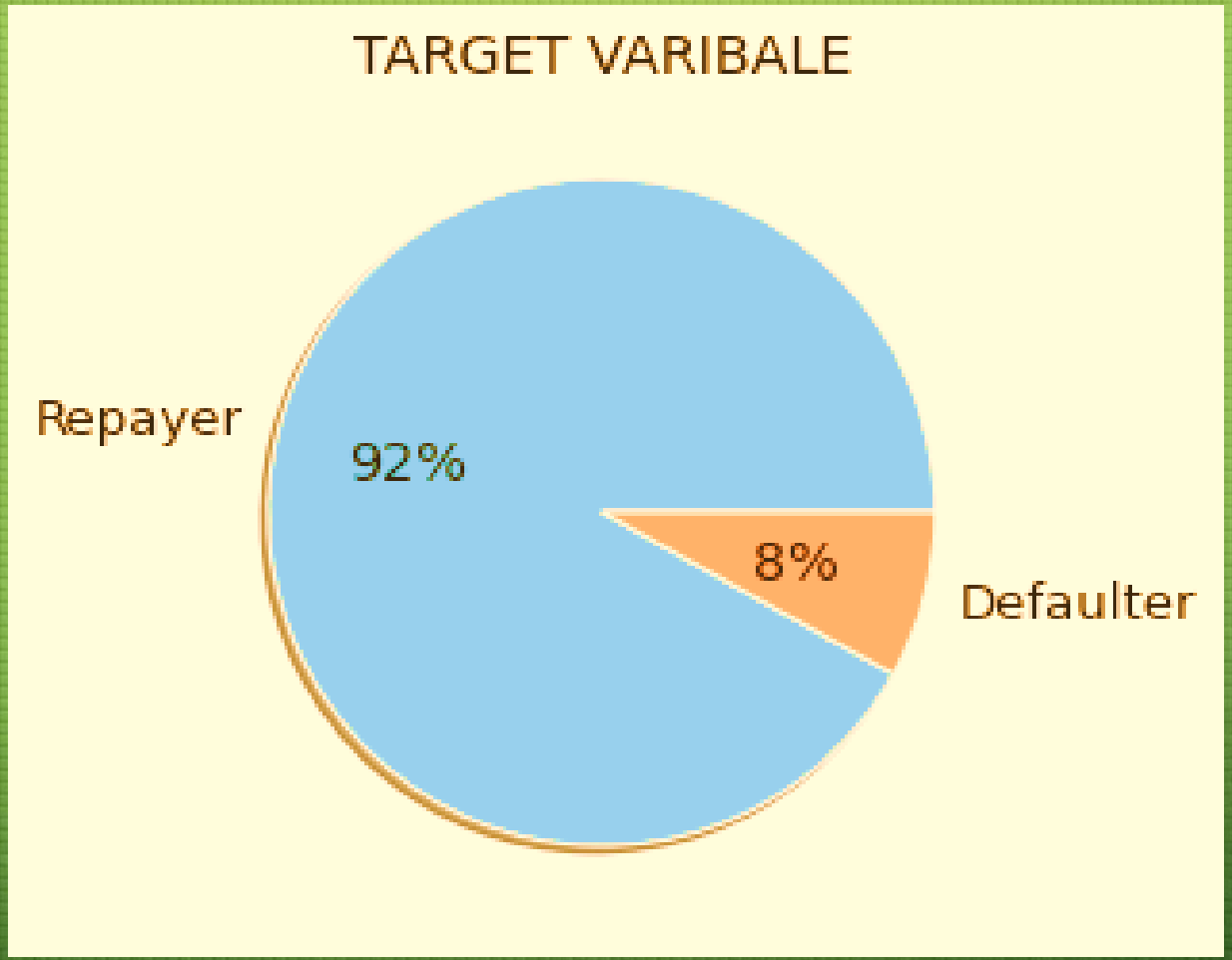
CREDIT EDA CASE STUDY

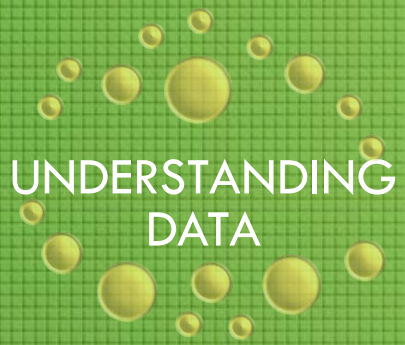
BY MR. NIRANJAN SABLE



PROBLEM STATEMENT -

- To validate the risk associated with a loan applicant –
    1. Disbursing loan to a defaulter.
    2. Rejecting loan to a potential repayer.
- Analysing the various features of data with respect to target variable. The two scenarios –
    1. The Client with Payment difficulties.
    2. All other cases when the payment is made on time.
- With the analyses, the aim is to understand the driving factors behind the loan defaulter

APPROACH FOR ANALYSIS



- Read the data file
- Inspect the data frame

DATA CLEANING

- Deleting and imputing null values
- Identifying and deleting irrelevant columns
- Standardizing
- Data type conversion
- Identifying the outliers

Data Analysis

- Univariate
- Segmented Univariate
- Bi/Multivariate
- Merged dataframe

Conclusion

- Driving factors for defaulters
- Recommendations

DATA ANALYSIS

TOP THREE CORRELATIONS

FOR CLIENTS WITH PAYMENT DIFFICULTIES

	VAR1	VAR2	Correlation
1	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998
2	AMT_GOODS_PRICE	AMT_CREDIT	0.983
3	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.956

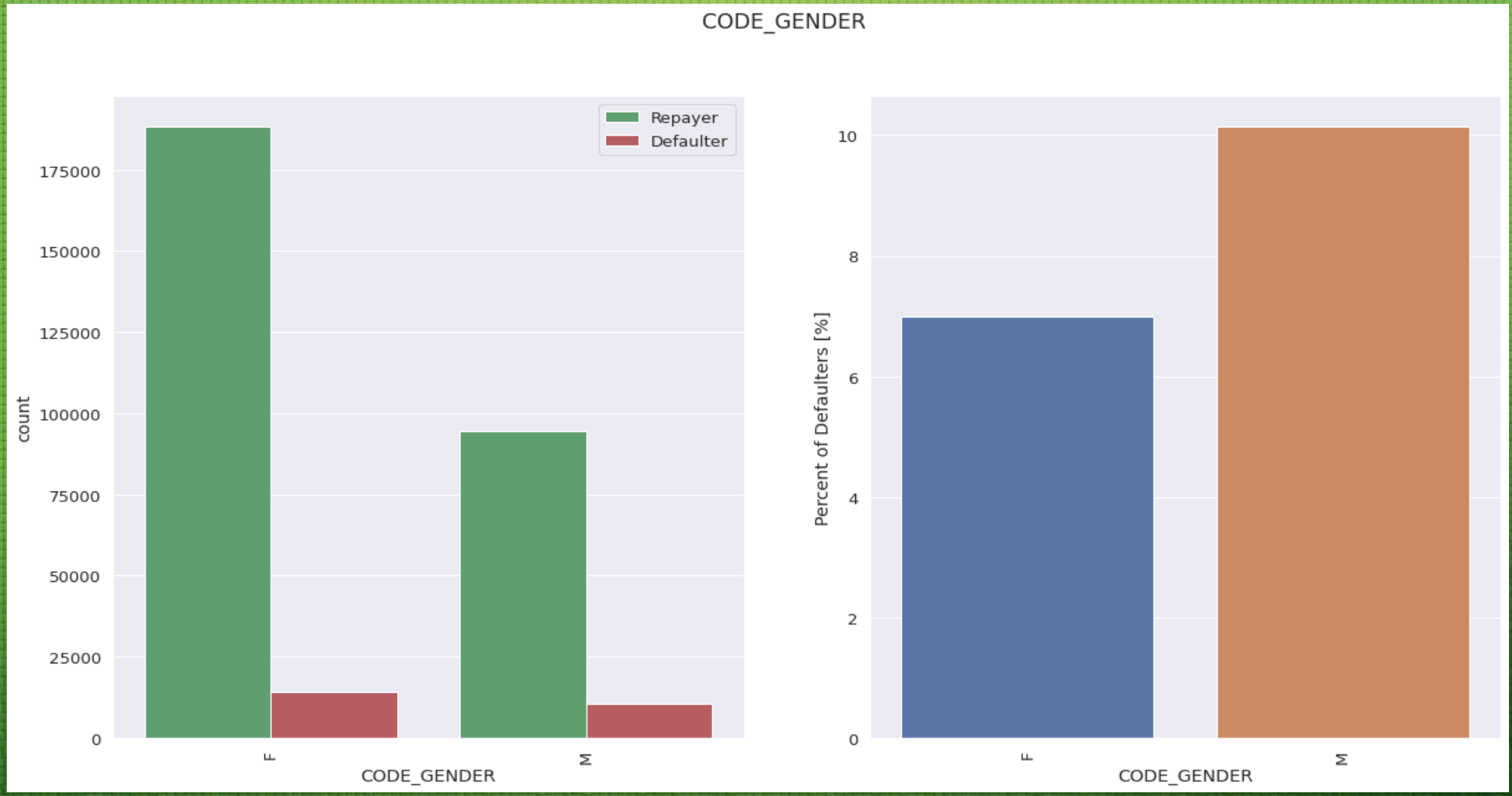
FOR CLIENTS WITH NO PAYMENT DIFFICULTIES

	VAR1	VAR2	Correlation
1	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.990
2	AMT_GOODS_PRICE	AMT_CREDIT	0.982
3	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.956

DRIVING FACTOR BEHIND DEFAULTERS–

1. CODE_GENDER

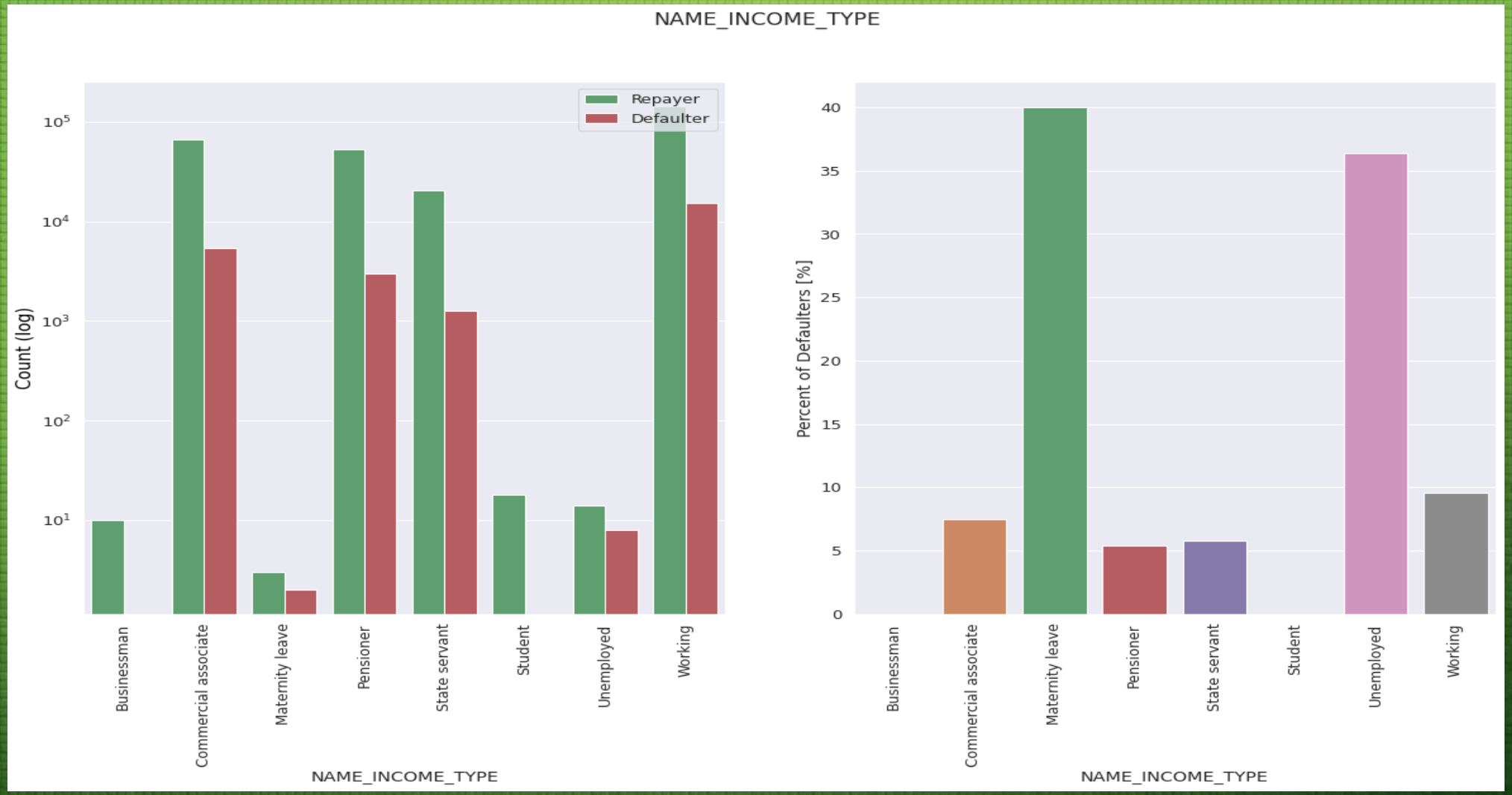
The count of female applicants is more compared to male. Moreover, it is the male applicants who have higher defaulter rate.



DRIVING FACTOR BEHIND DEFAULTERS–

2. INCOME_TYPE

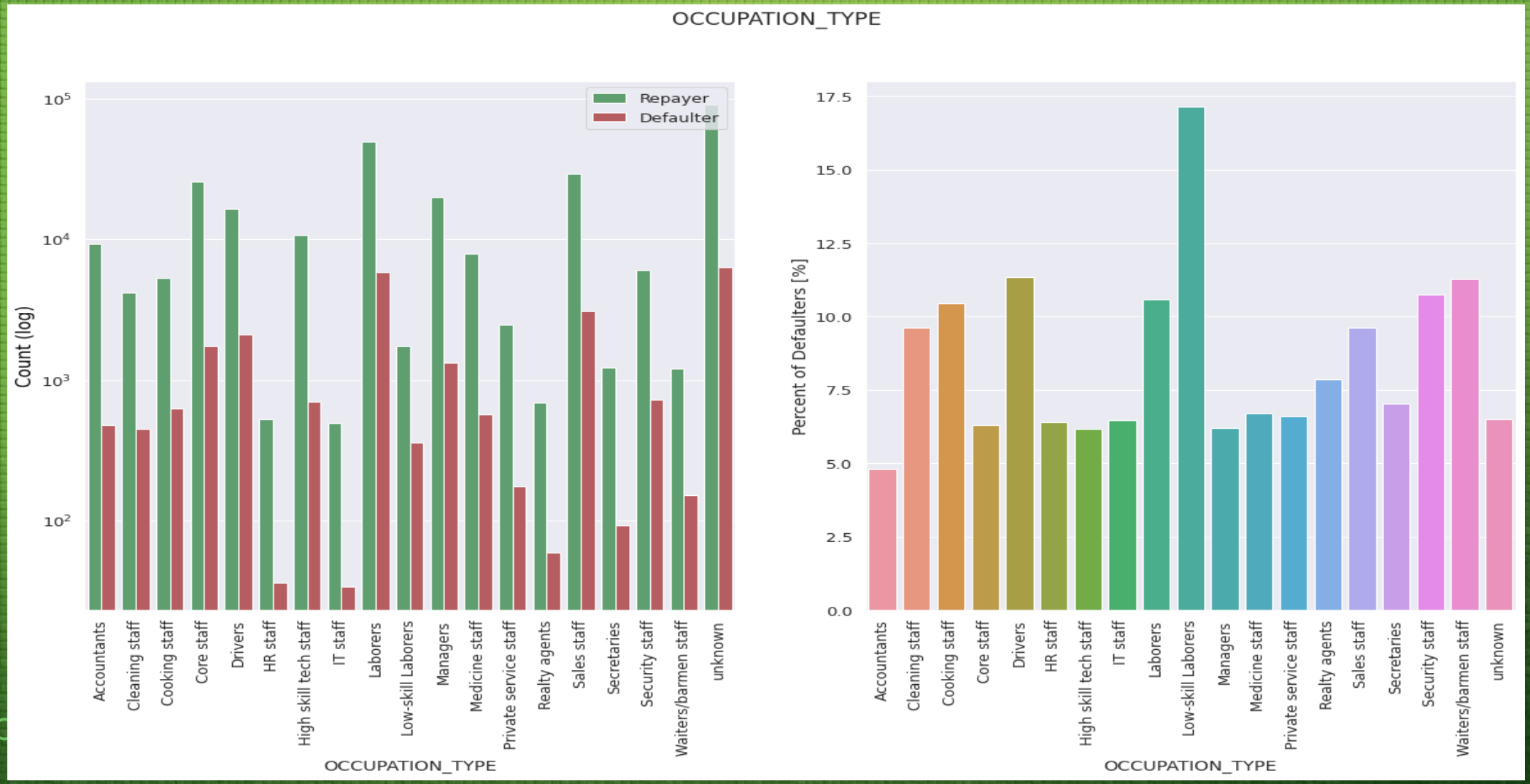
The maternity leave and unemployed income type has higher defaulter rate and the count of the applicants for the income type is also less. Hence we reject the loan applications.



DRIVING FACTOR BEHIND DEFAULTERS–

3. OCCUPATION_TYPE

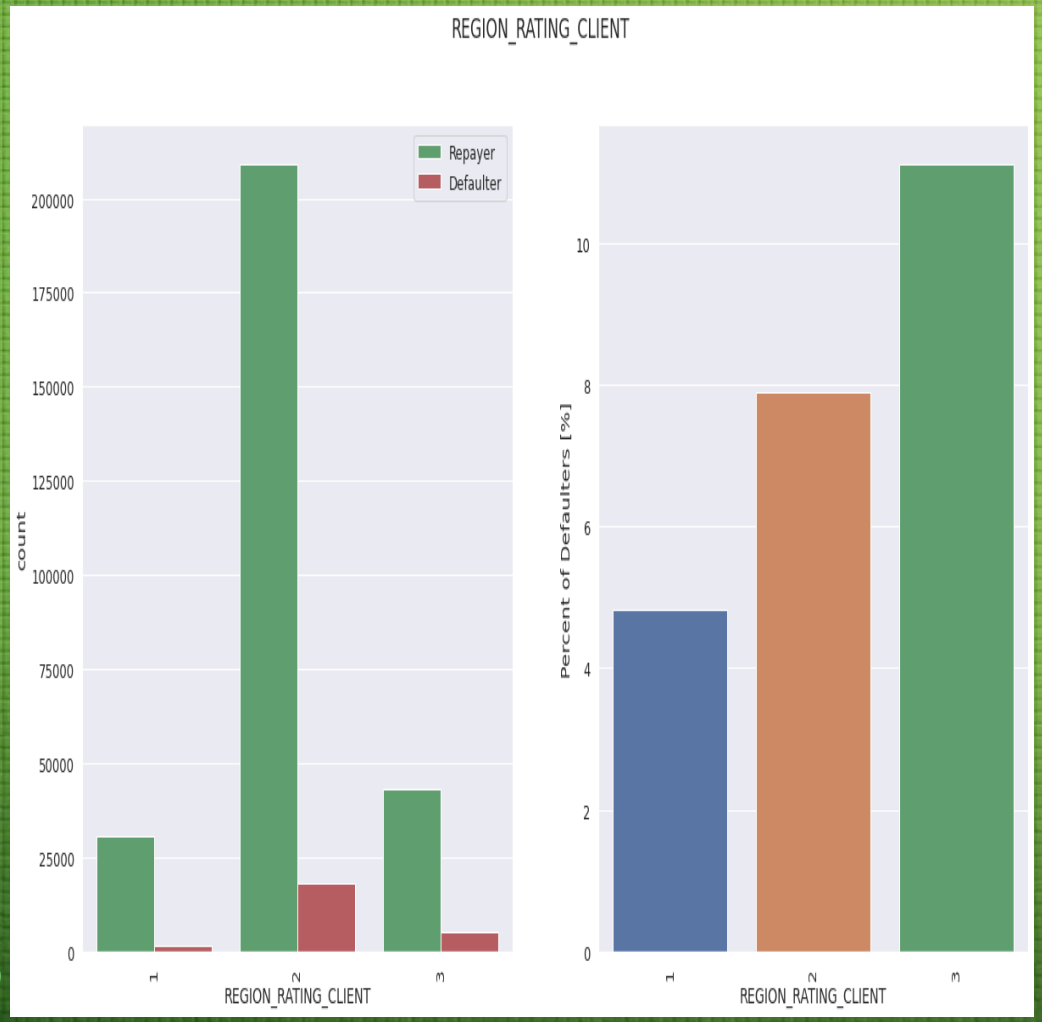
The low skilled labours have the highest default rate. The number of applicant are also significantly high, and therefore we can term it a high risk category.

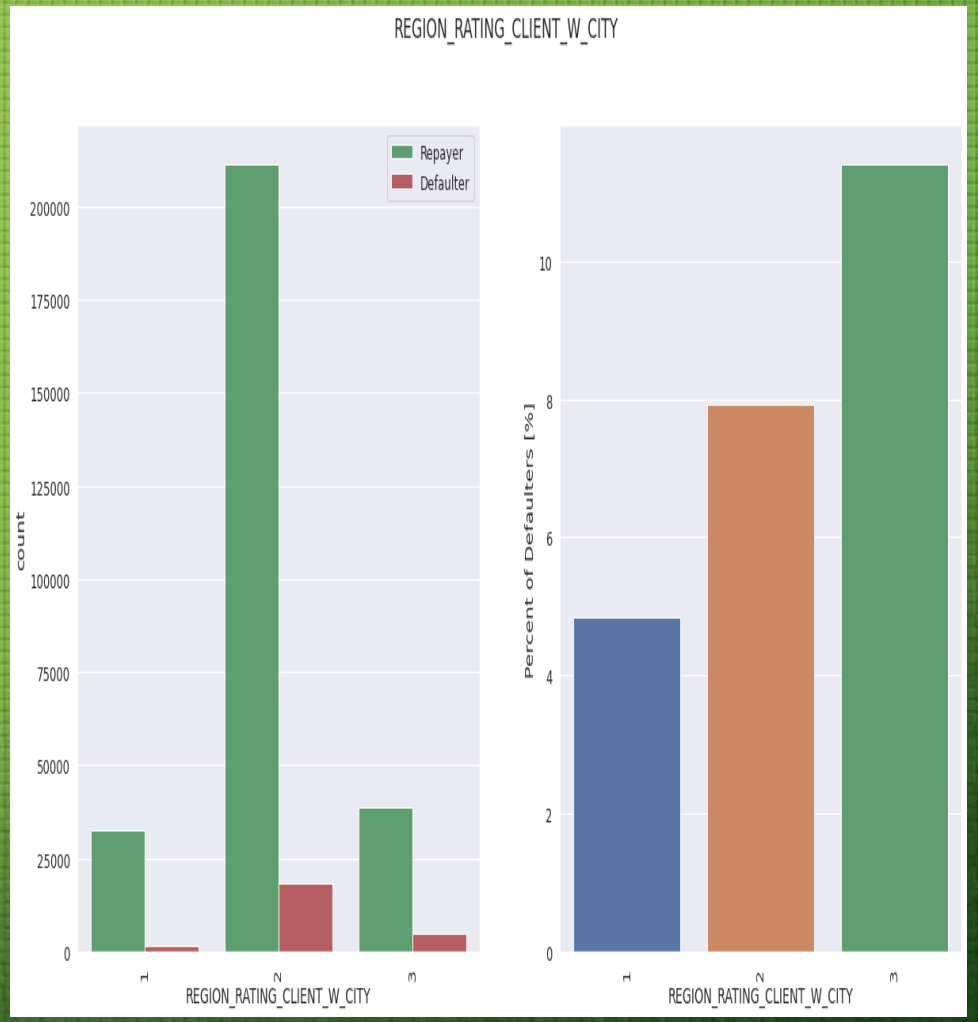


DRIVING FACTOR BEHIND DEFAULTERS–

4/5. REGION_RATING_CLIENT / REGION_RATING_CLIENT_W_CITY

The region with rating of 3,with taking city in account, have comparatively higher defaulter rate. In additions, the number of applications are also less.

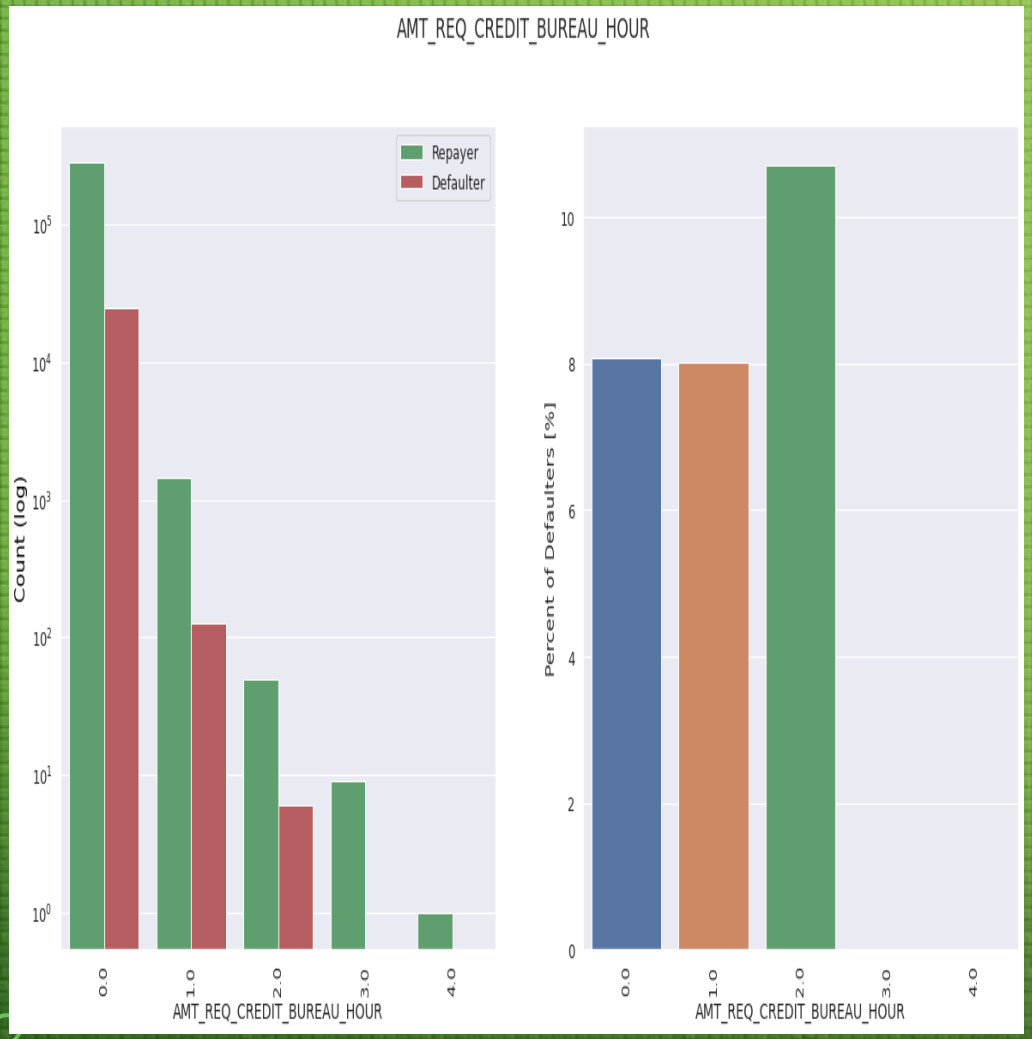


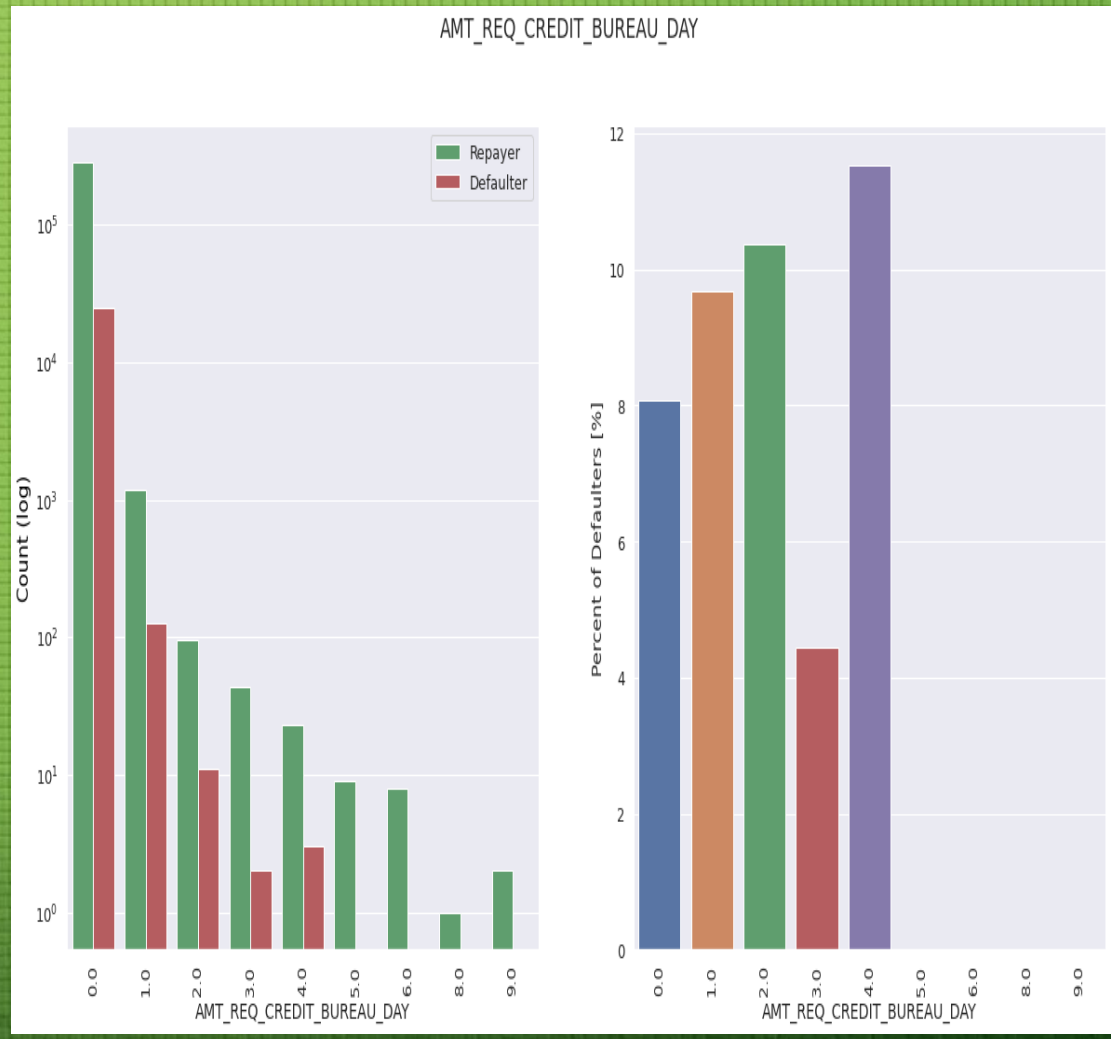


DRIVING FACTOR BEHIND DEFAULTERS–

6/7. AMT_REQ_CREDIT_BUREAU_HOUR / _DAY

For higher number of enquiries before an hour and a day of application, the defaulters are zero.

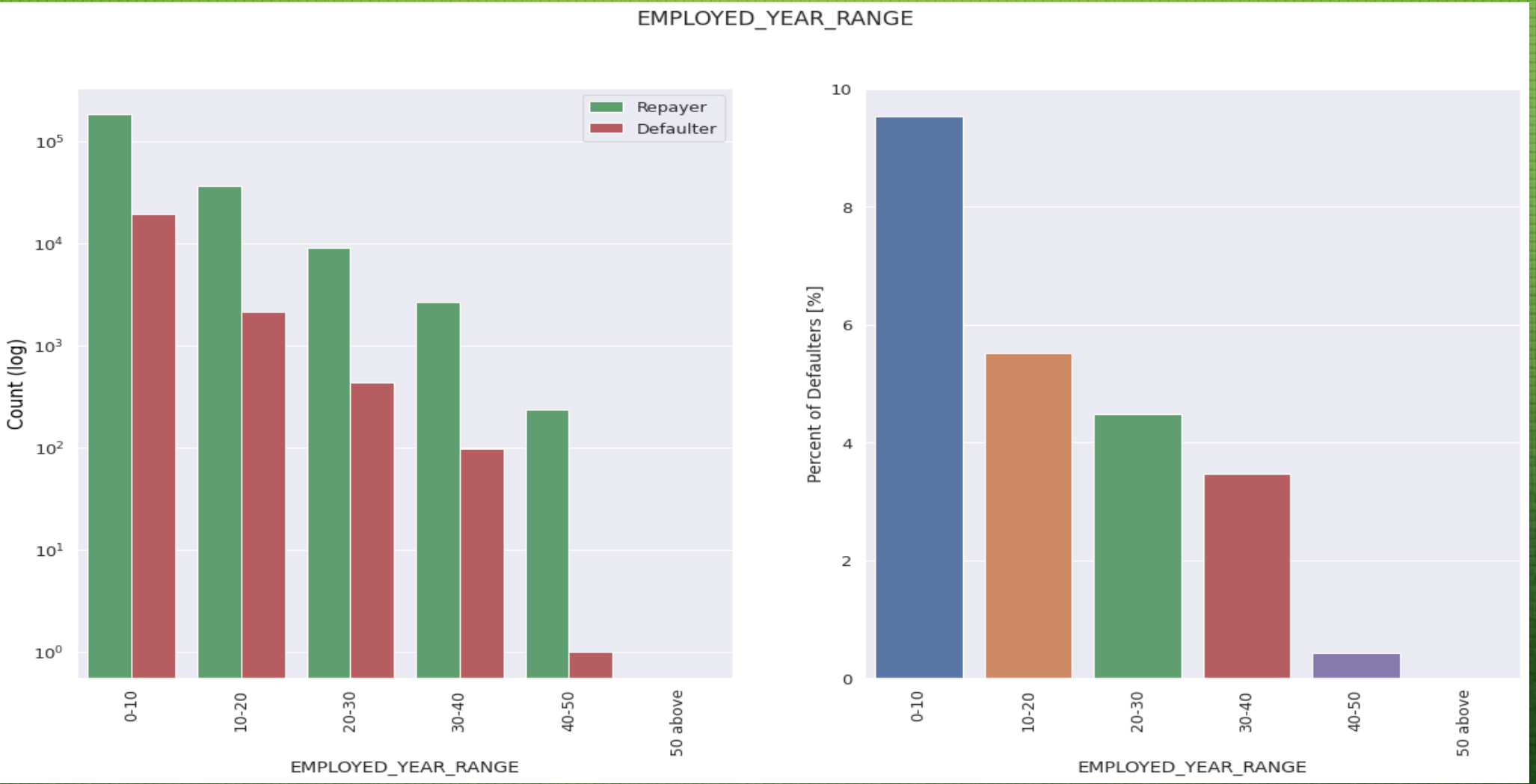




DRIVING FACTOR BEHIND DEFAULTERS–

8. EMPLOYED_YEAR_RANGE

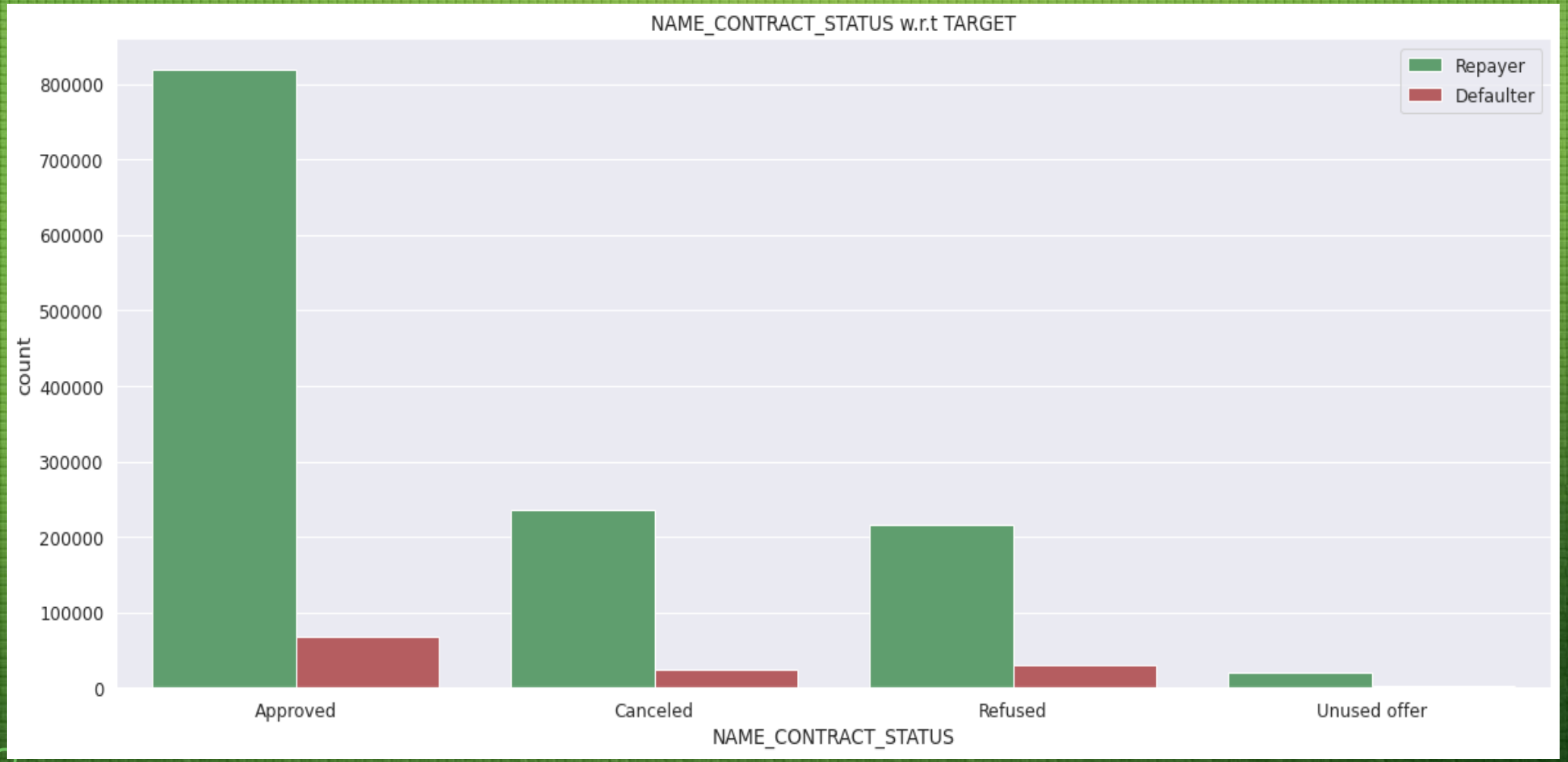
From the bar chart, it is quite evident that as the number of years of experience increases, the rate of defaulter also decreases.



RECOMMENDATIONS –

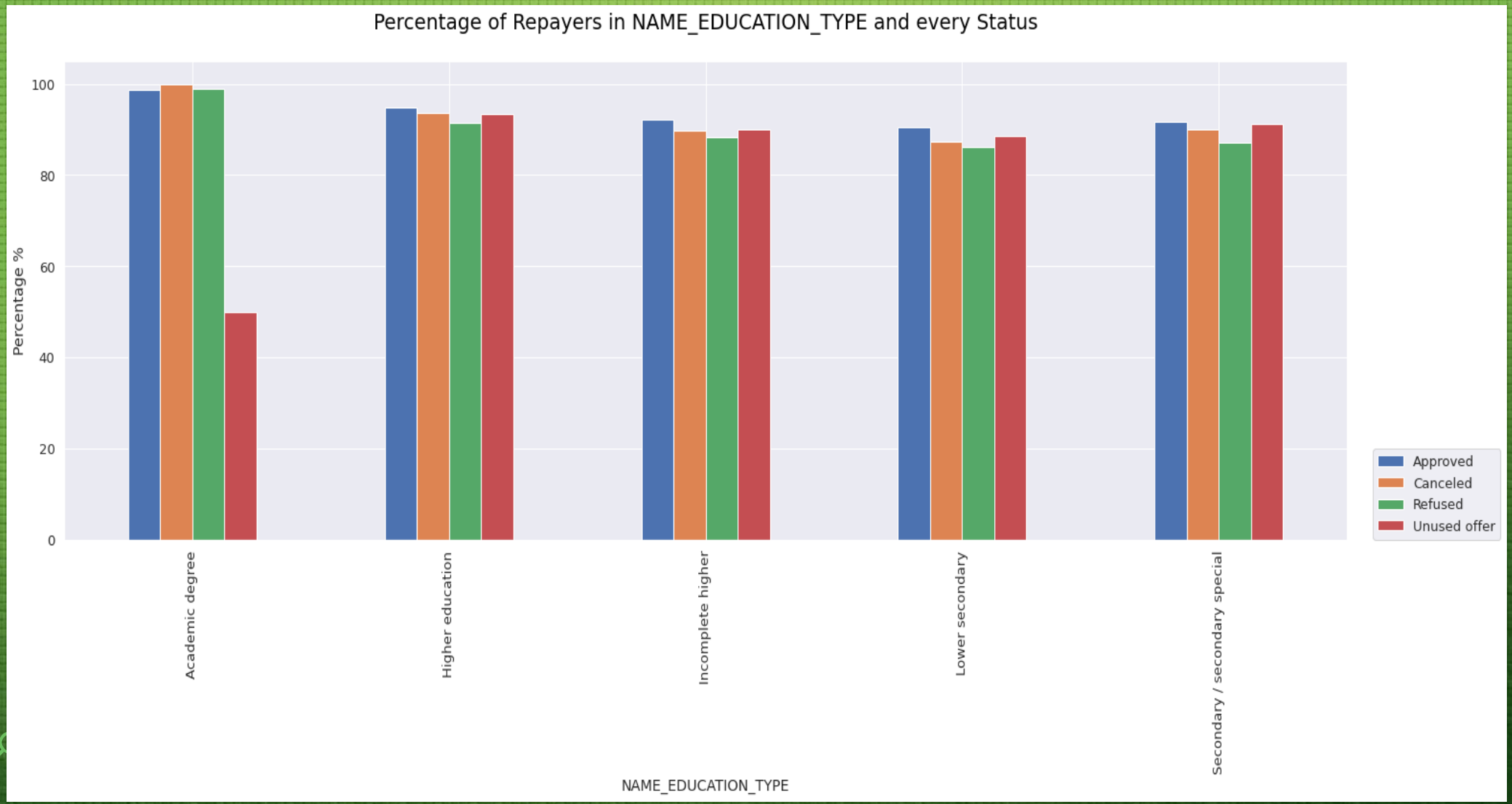
1. NAME_CONTRACT_STATUS with respect to TARGET

There are lot of applicants who are repayers, but still there applications are cancelled. This is a loss in business for the bank. Banks should further analysis the cancelled category.



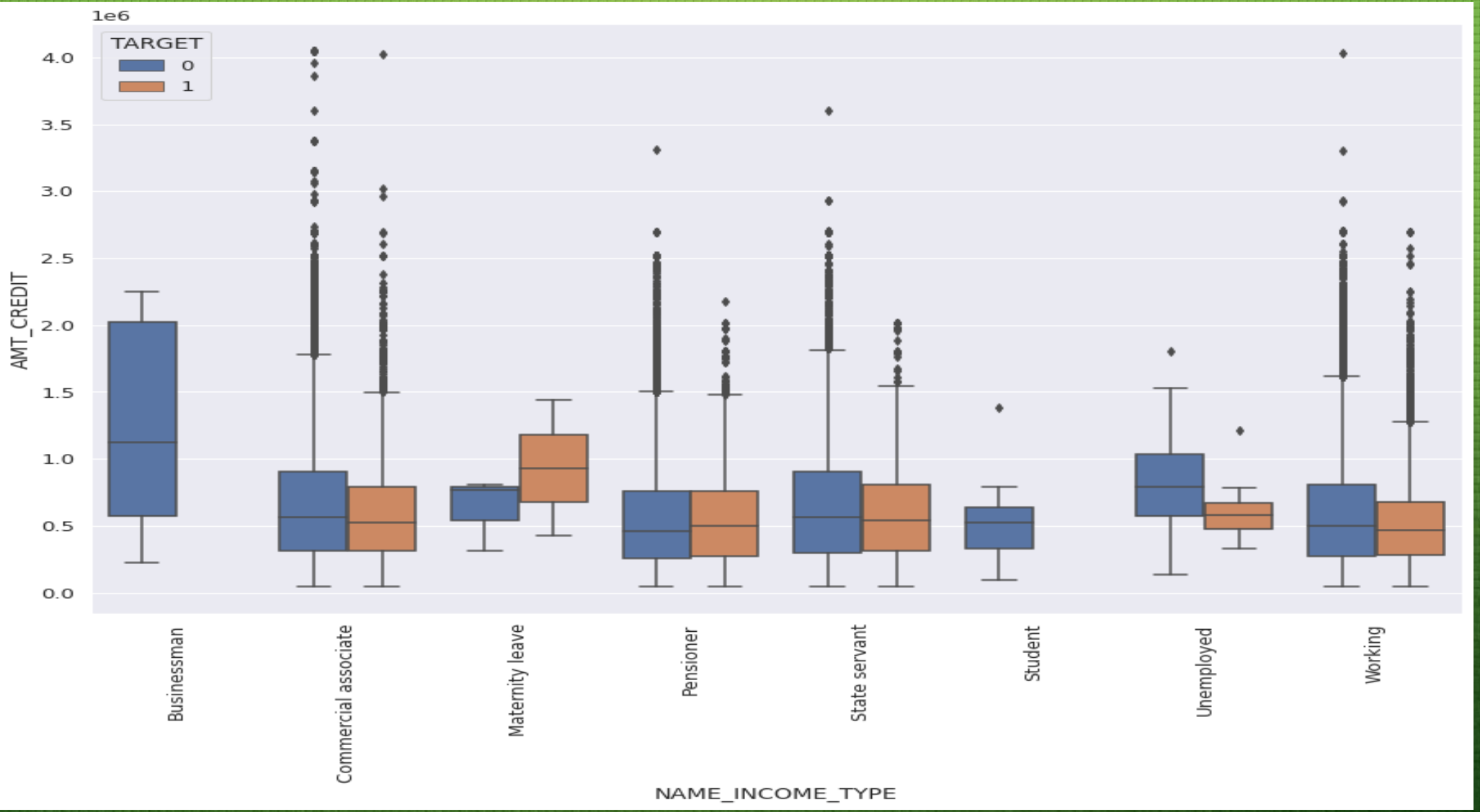
RECOMMENDATIONS –

2. One of the example for point no. 1(recommendation) is - All the applicants whose loan were cancelled and who belonged to Academic degree were repayers. Banks should reconsider this once. In addition, the application belonging to unused and academic degree feature has 98 % repayers. Hence the financial institute can encourage this group to use the loan



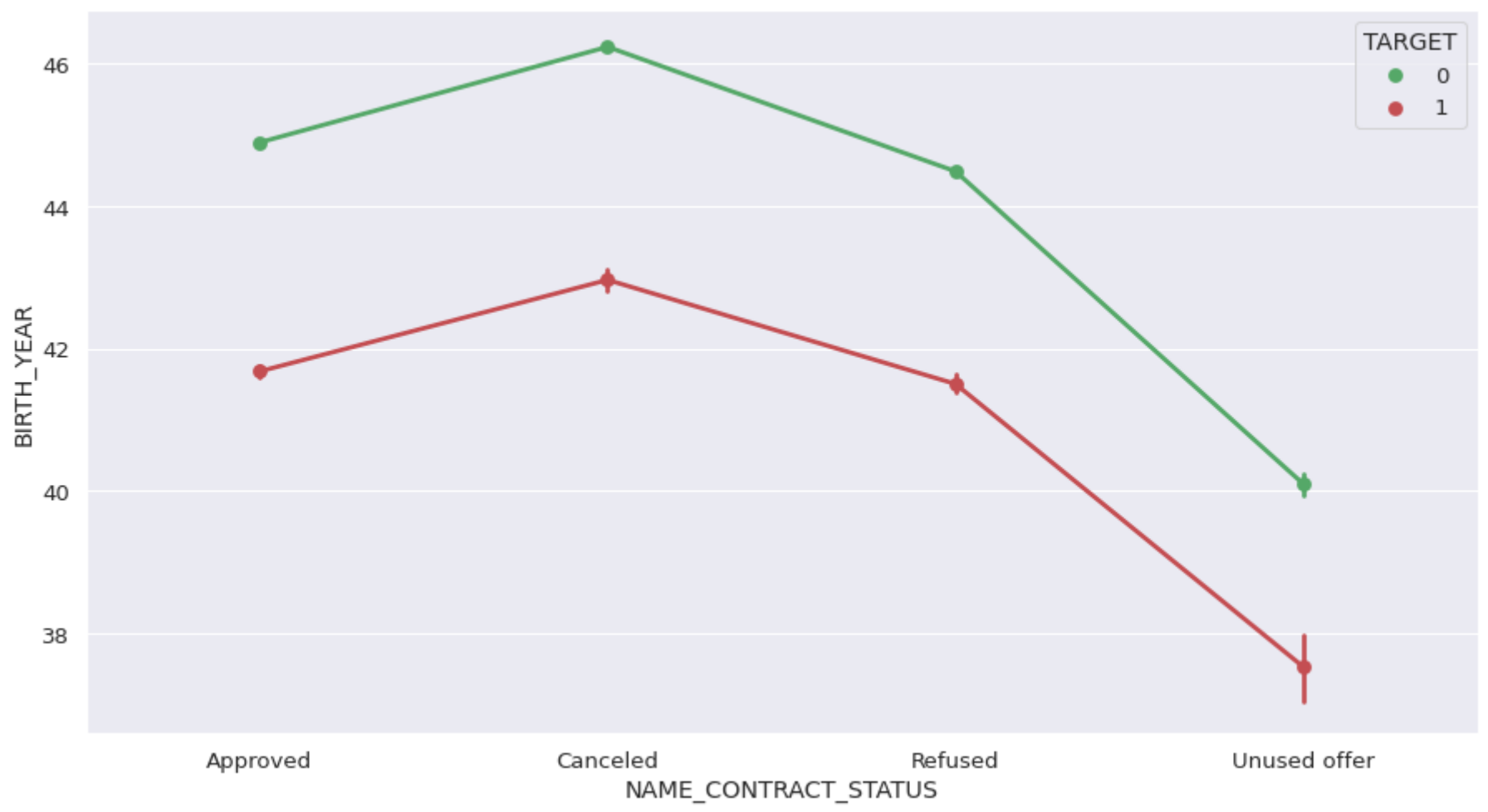
RECOMMENDATIONS –

3. The Business and student income group have no defaulters. Hence the bank should look for similar applicants.



RECOMMENDATIONS –

4. The applicants belonging to lower age group of upto 40 years are defaulting more. But the number of applications for loan are high for this group, so the bank can't ignore it. Rather the bank could give this category a loan with high interest rate, considering the category highly risky.



| | THANK YOU | |