CREDIT EDA CASE STUDY

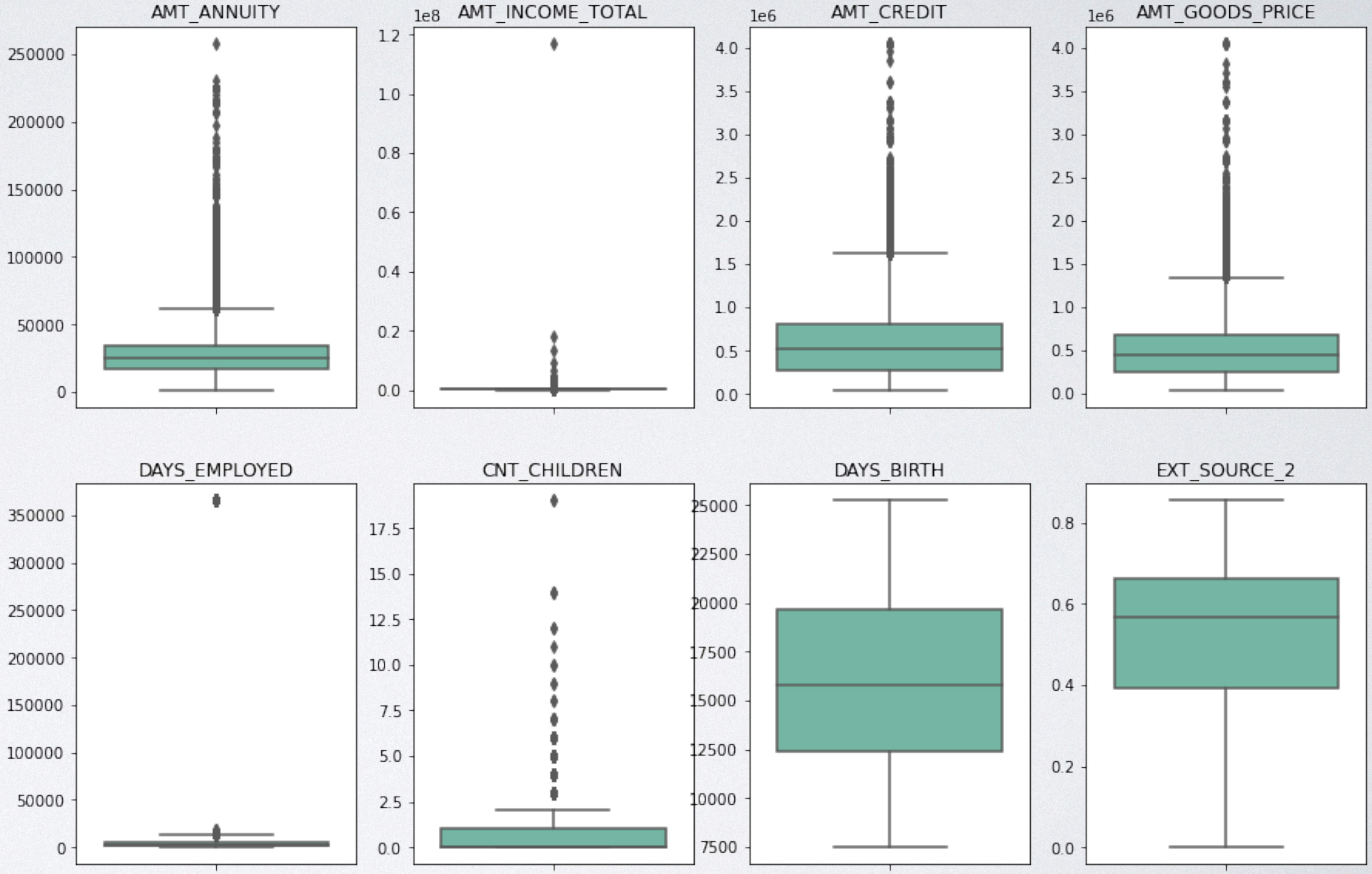
By

- Avanesh Kadu
- Ashish K

30/08/2021

Dealing with outliers

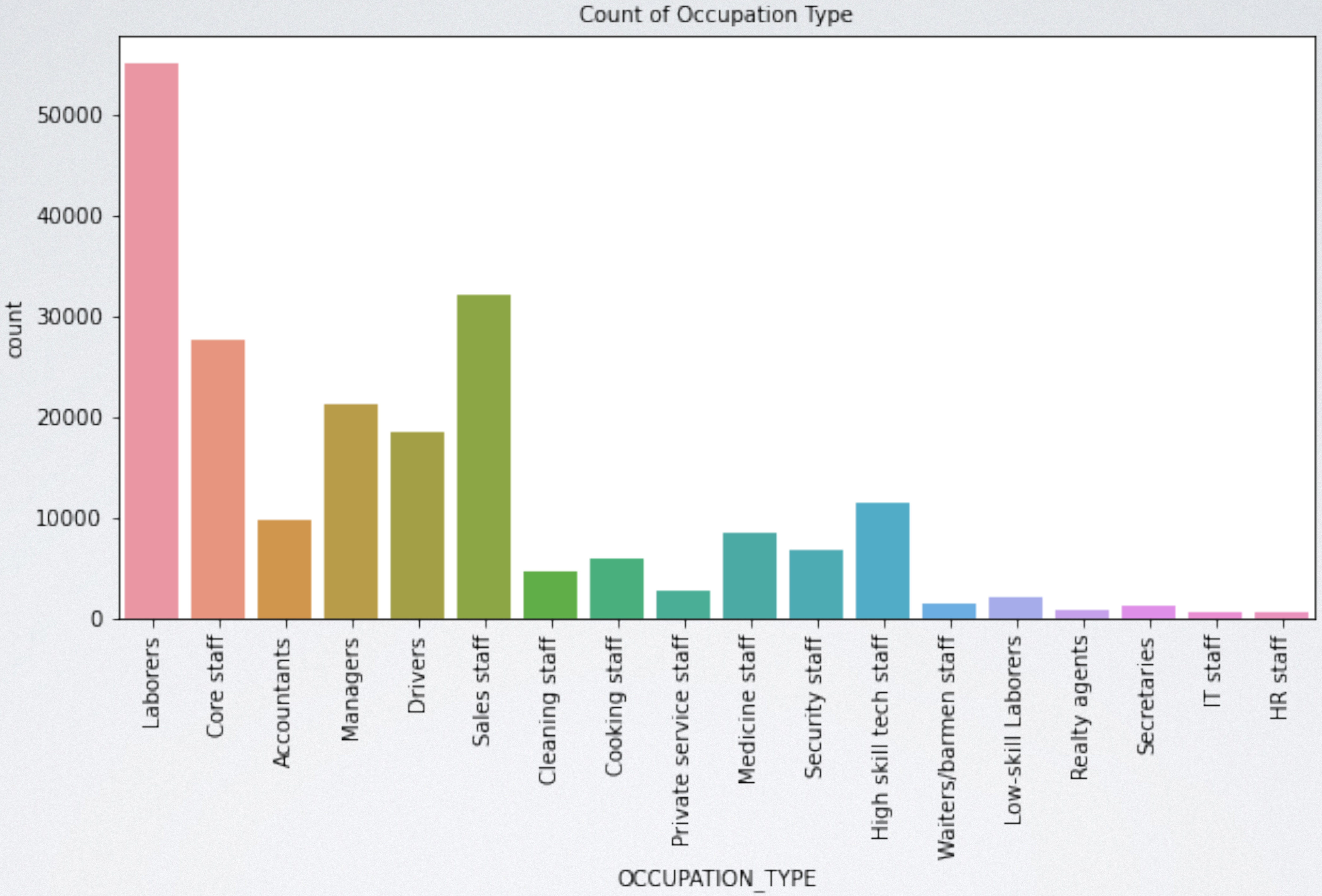
We can observe outliers in AMT_ANUITY , AMT_INCOME , and AMT_CREDIT that means some customers have high income and high credit score comparatively. These columns don't need to be imputed or deleted as these people can be potential customers.



We can also observe that DAYS_EMPLOYED column has heavy outliers and they need to deleted as these numbers are not practical and data will become non reliable .

DAYS_BIRTH do not have any outliers hence data is reliable.

Treating Occupation_Type column



Insights: Majority of people who applied for loan are labourers. Therefore we cannot designate missing value of other occupations like IT staff or HR staff with labourers. Hence we should not treat missing values of this particular column.

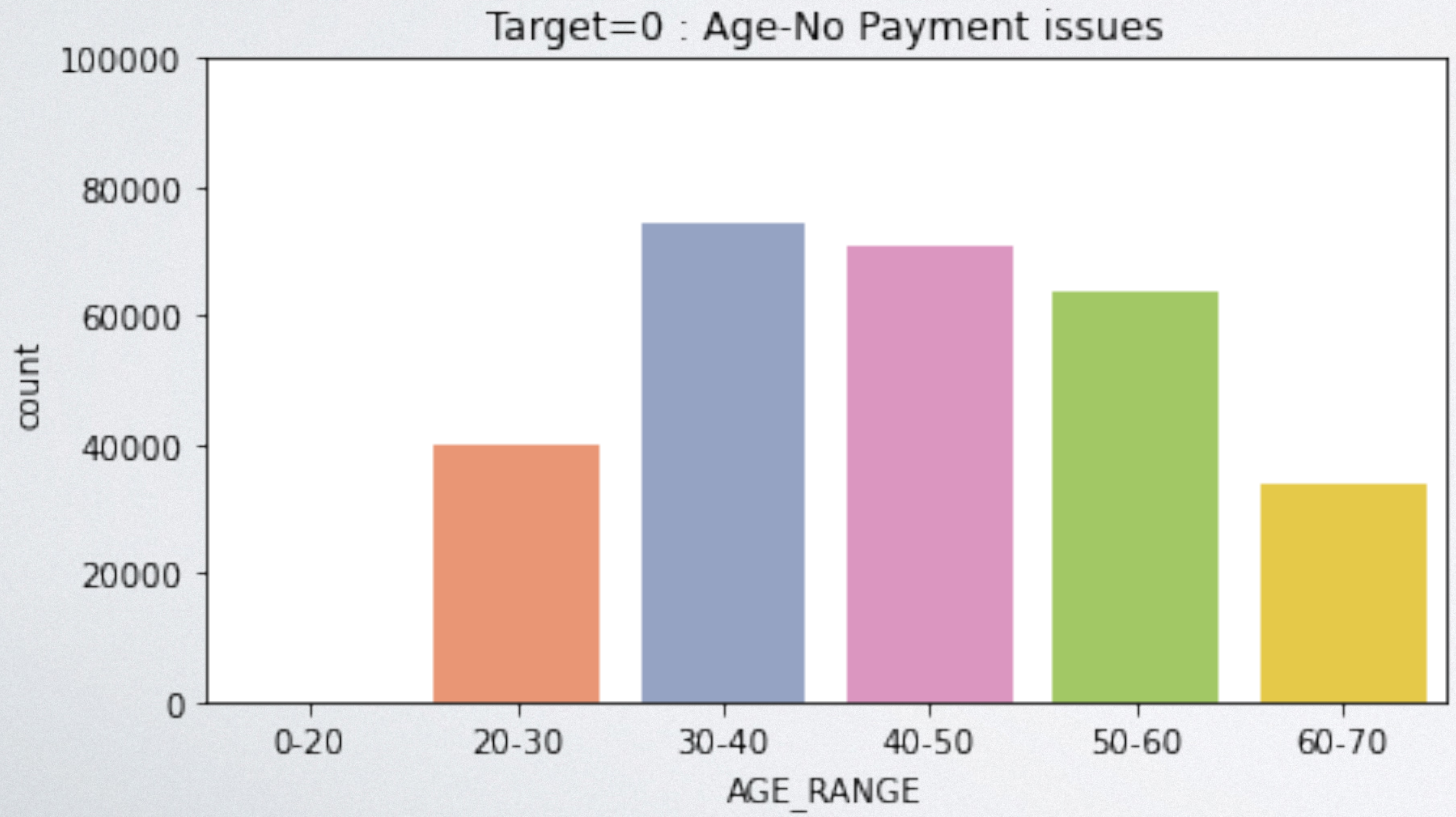
UNIVARIATE ANALYSIS OF TARGET COLUMNS

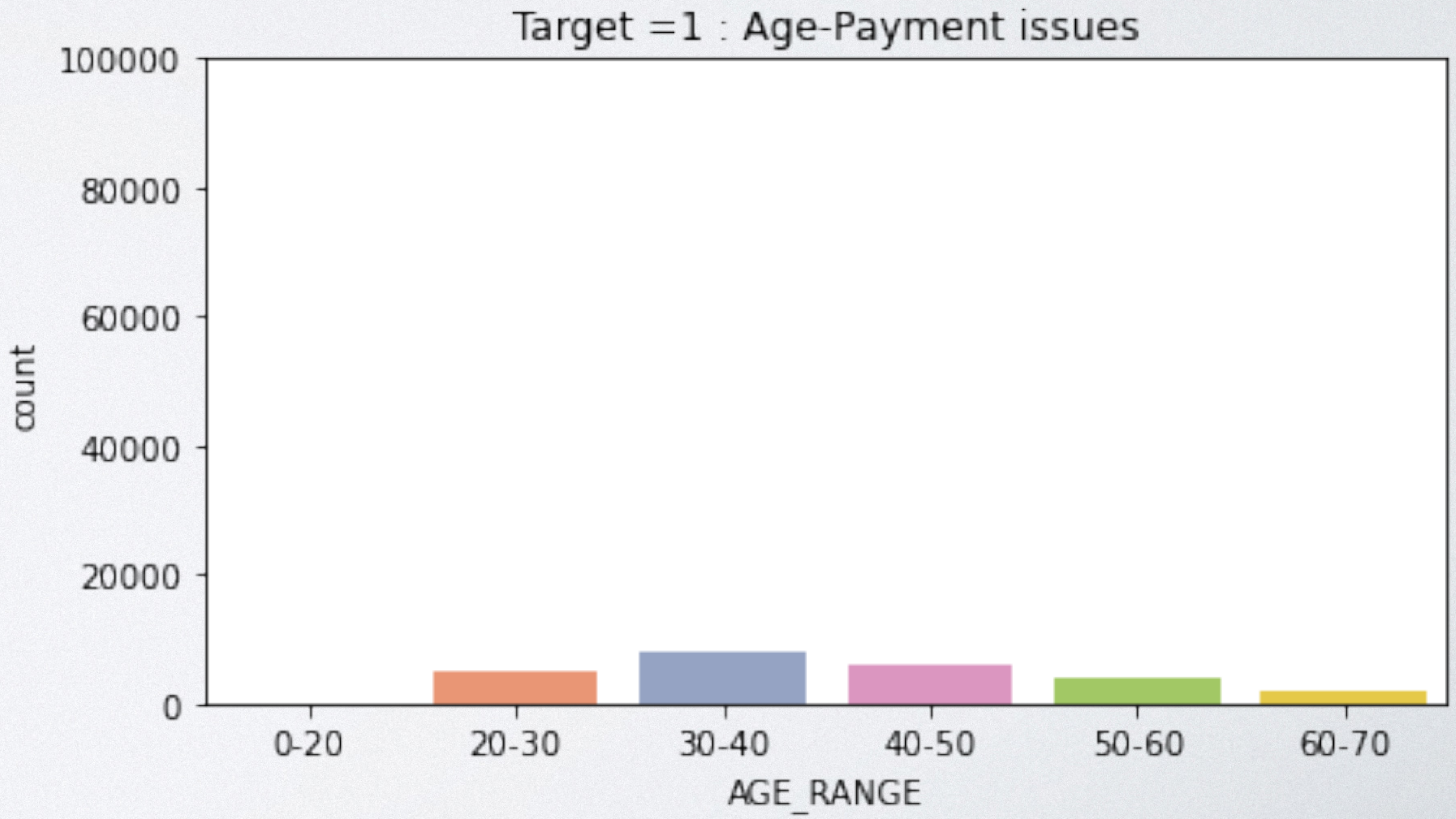
Numeric variable analysis for target_0 & target_1 dataframe

Insights :

1) Millennials of age 0-20 are not and should not be approved for loan as they lack education and employment

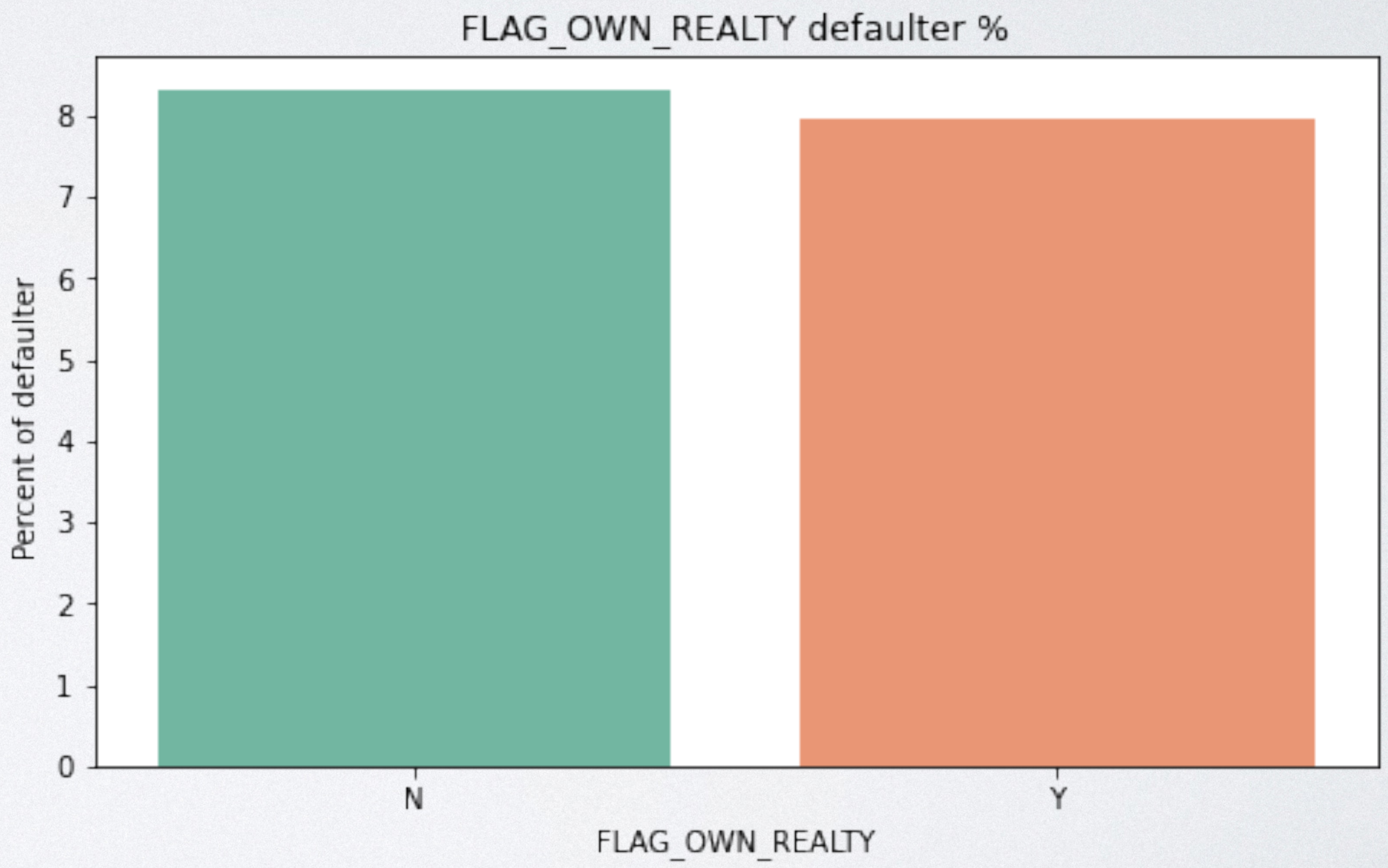
2) Age Group 30-40 should be strongly considered for lending loas as they tend to pay back , followed by 40-50 and 50-60





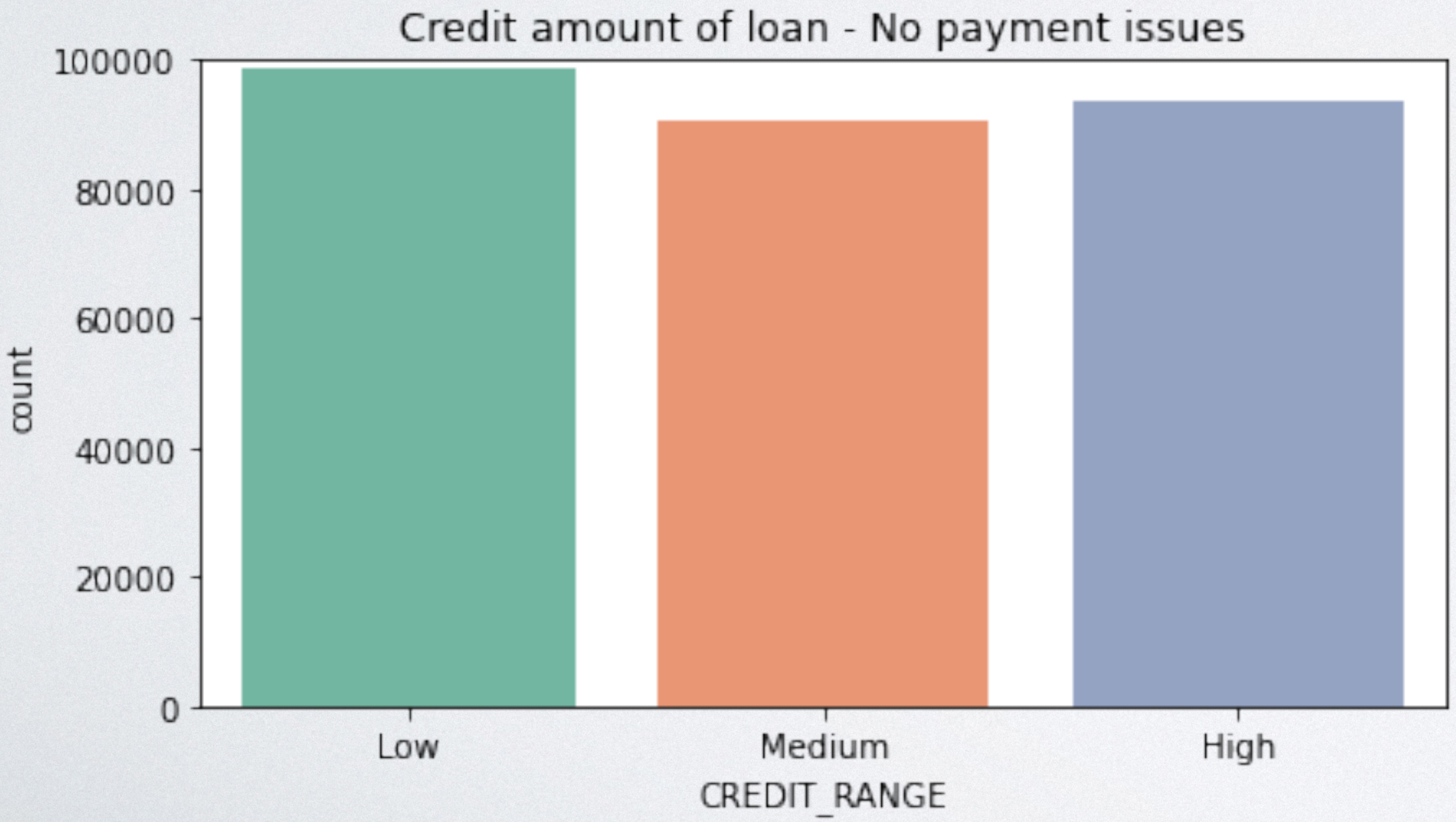
Target vs Flag Own Realty

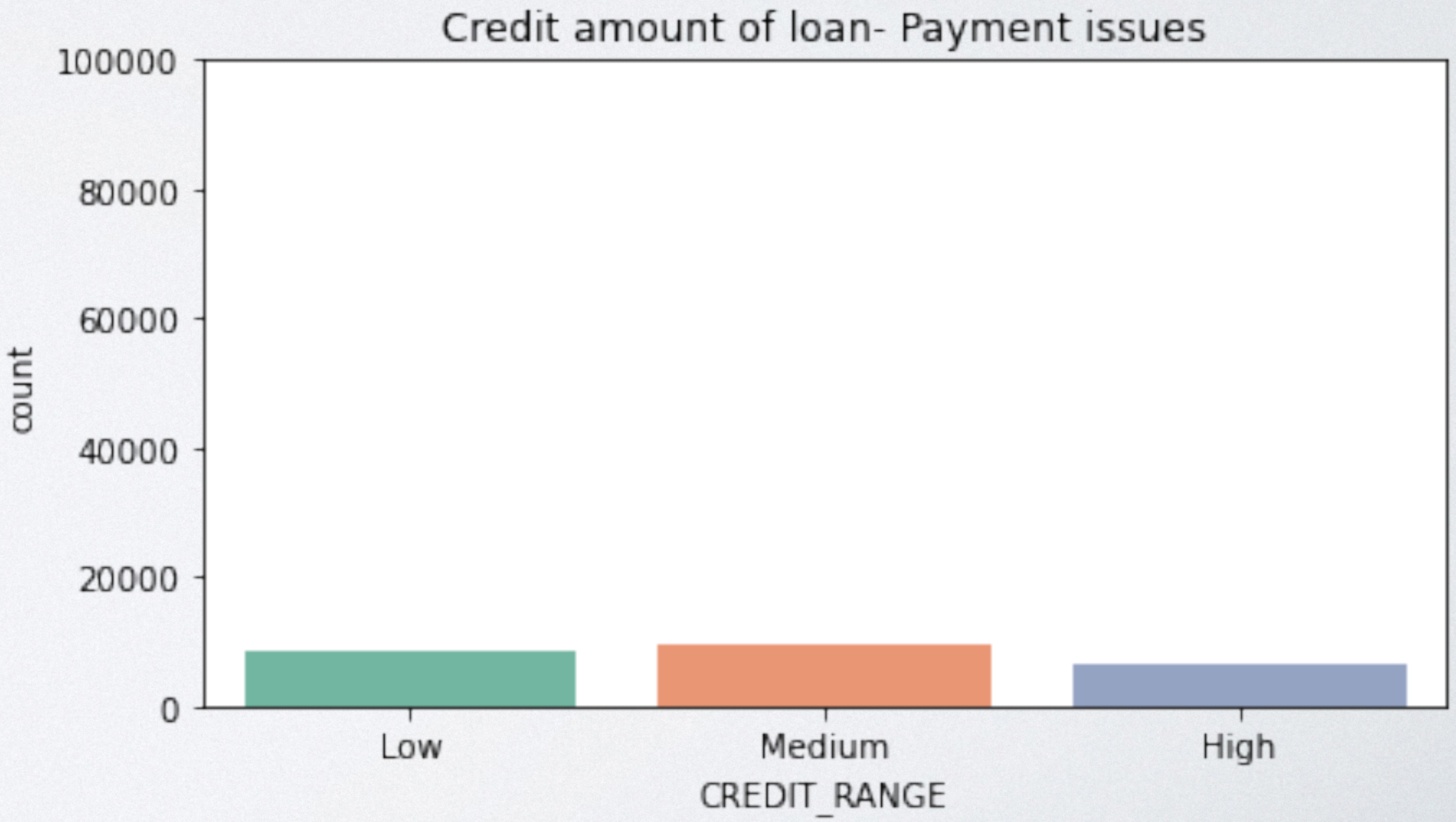
People with own property are almost double then the people who don't own property but the rate of default is very close . Hence Owning own property doesn't guarantee payback , in-fact Owing realty and payback of loan are not related.



Target Vs Credit Range

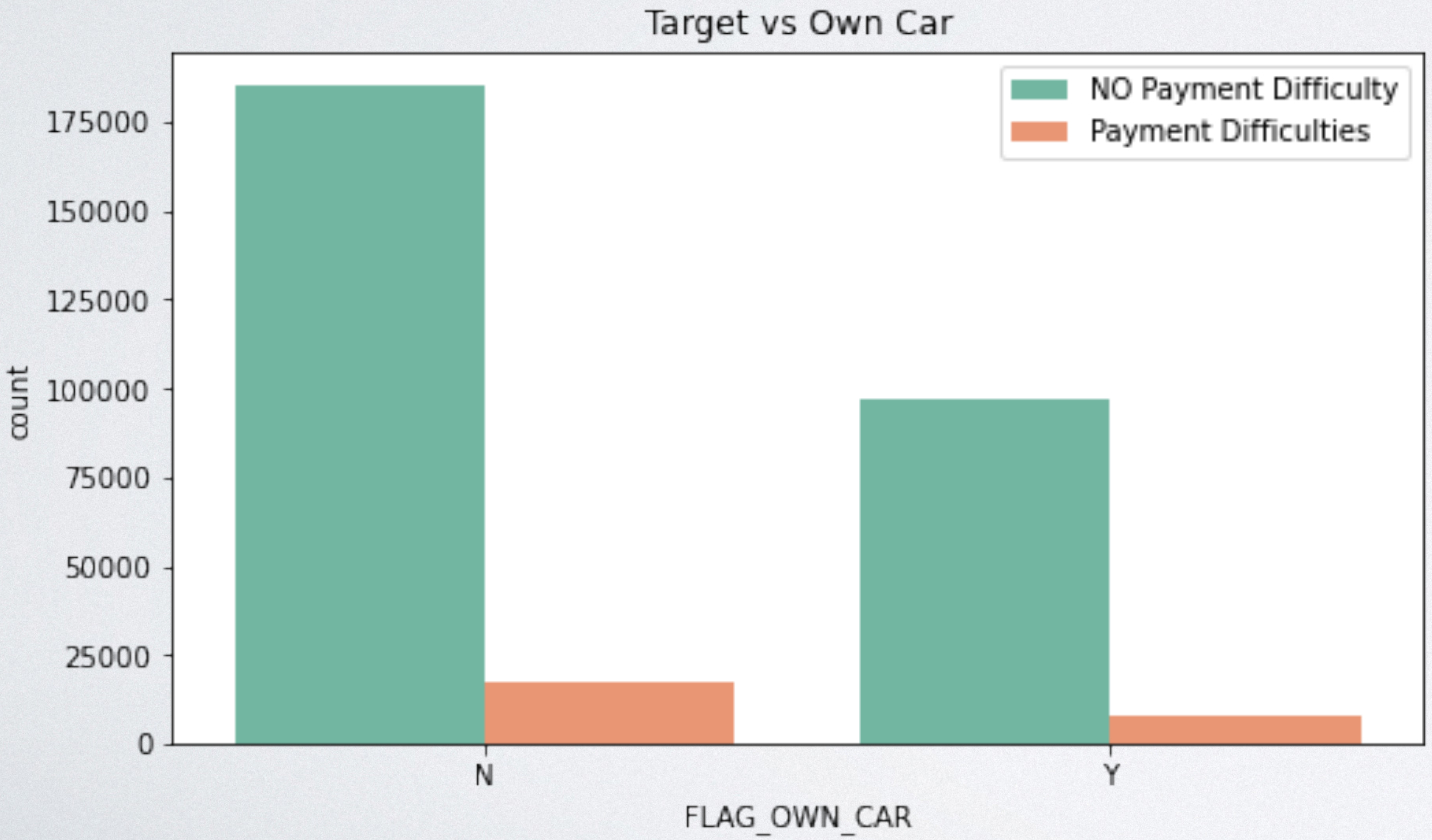
1. People with low Credit range tend to pay back and have less payment issues .
2. People with High and Medium credit range are also upto the mark and can surely be considered for issuing loan.

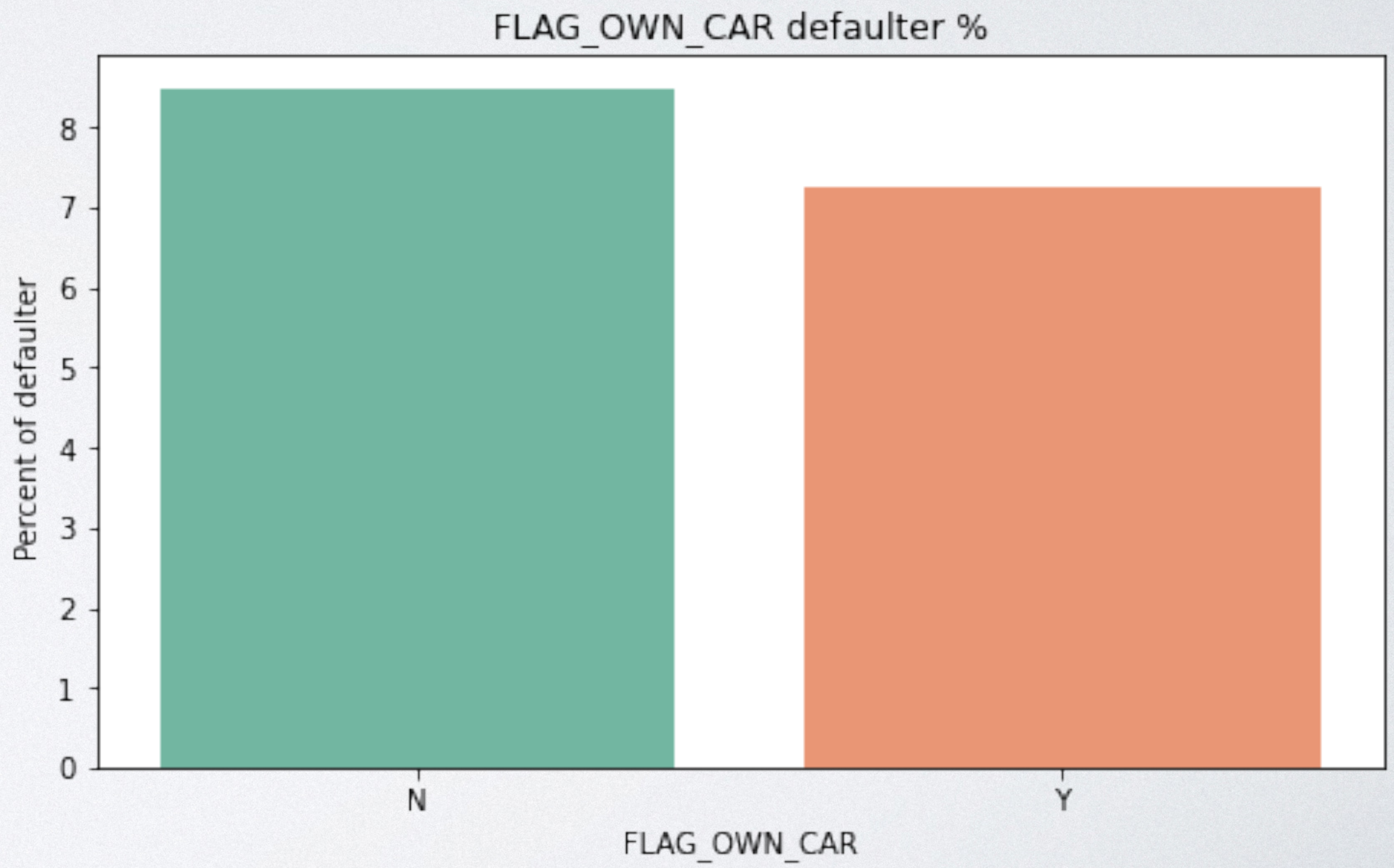




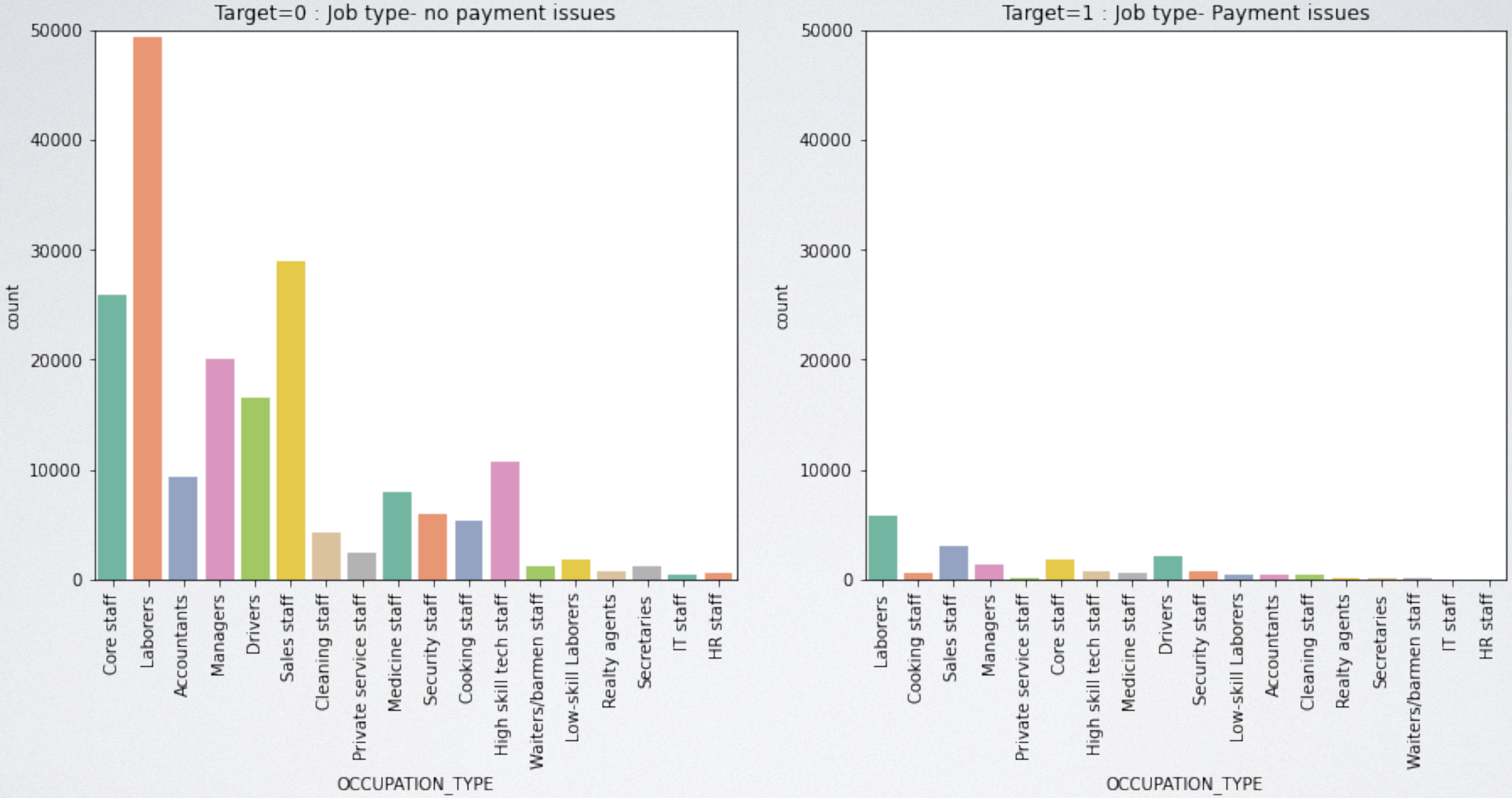
Target vs FLAG_OWN_CAR

The number of people who don't have a car is almost double the number of people who own a car . It can be seen that mostly customers who own a car have more chance to payback the loan , and Customers who don't own car have little high chance of defaulting . This factor should be taken into consideration by banks





Categorical variable analysis - Target vs Occupation Type

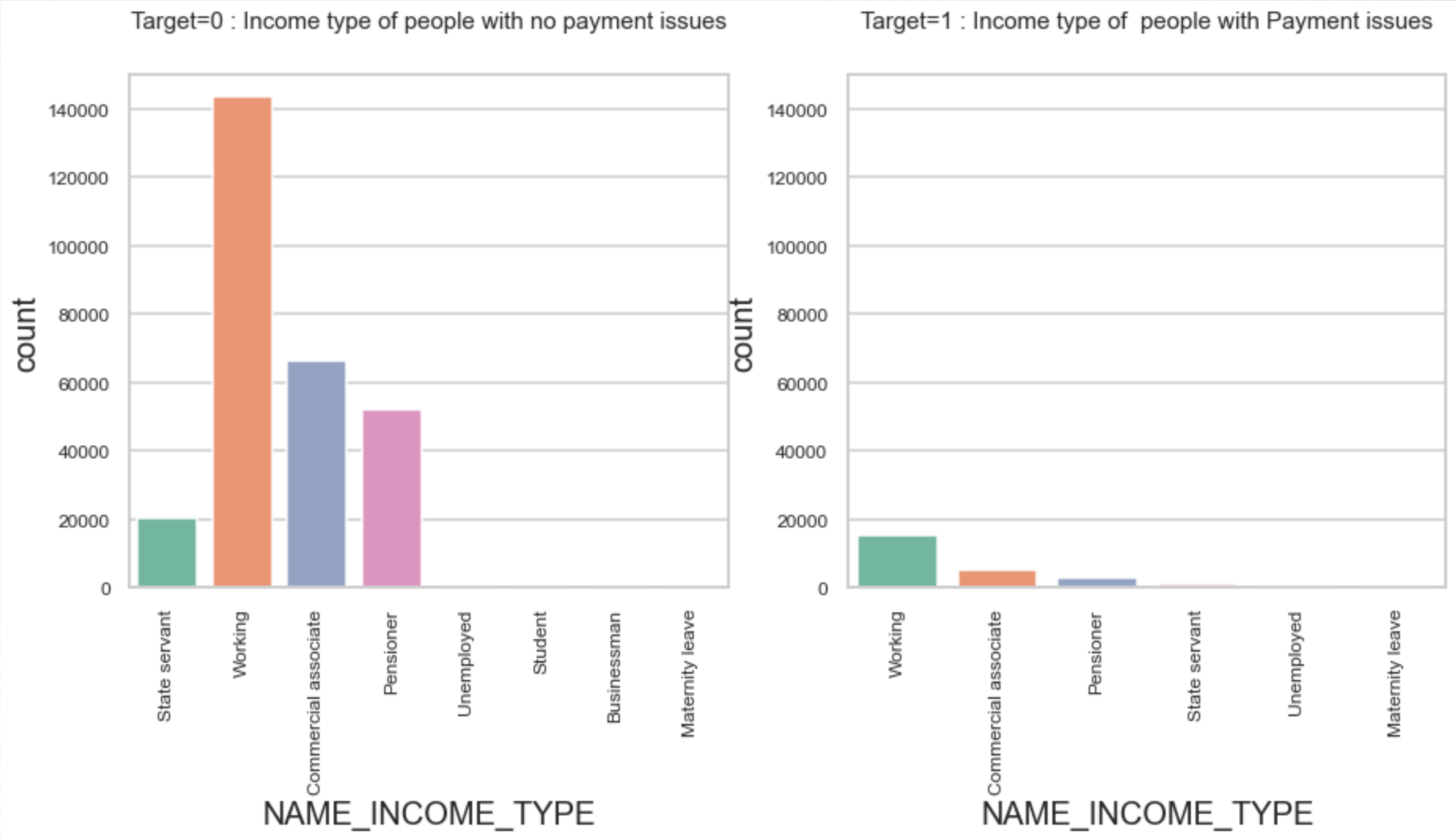


From above analysis we can conclude that Labourers tend to payback the loan on time , while IT Staff mostly likely to don't pay back the loan on time .

Categorical variable analysis - Target vs Income Type

Insights :

Plot shows that working class of people have very high percentage of paying back the loan.

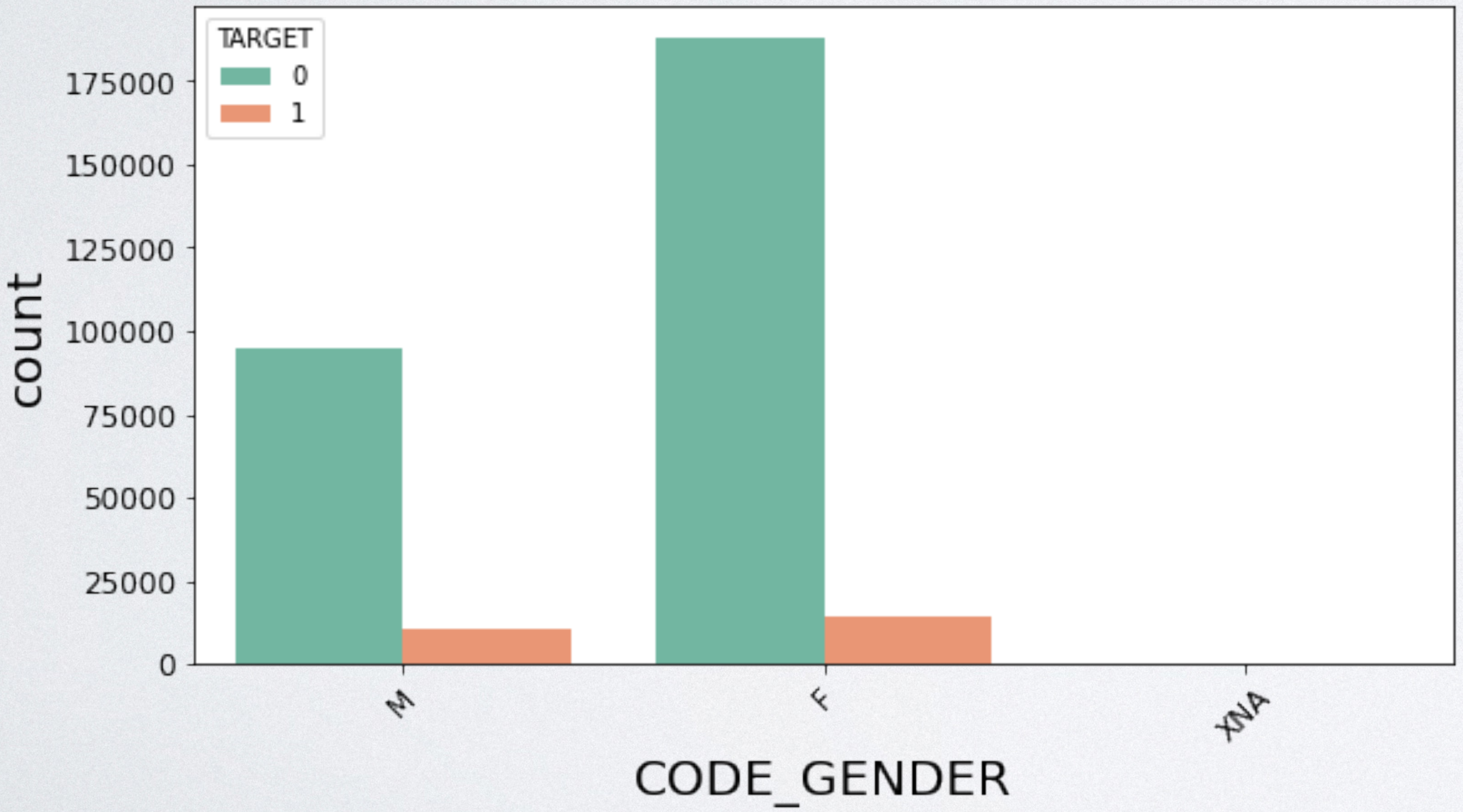


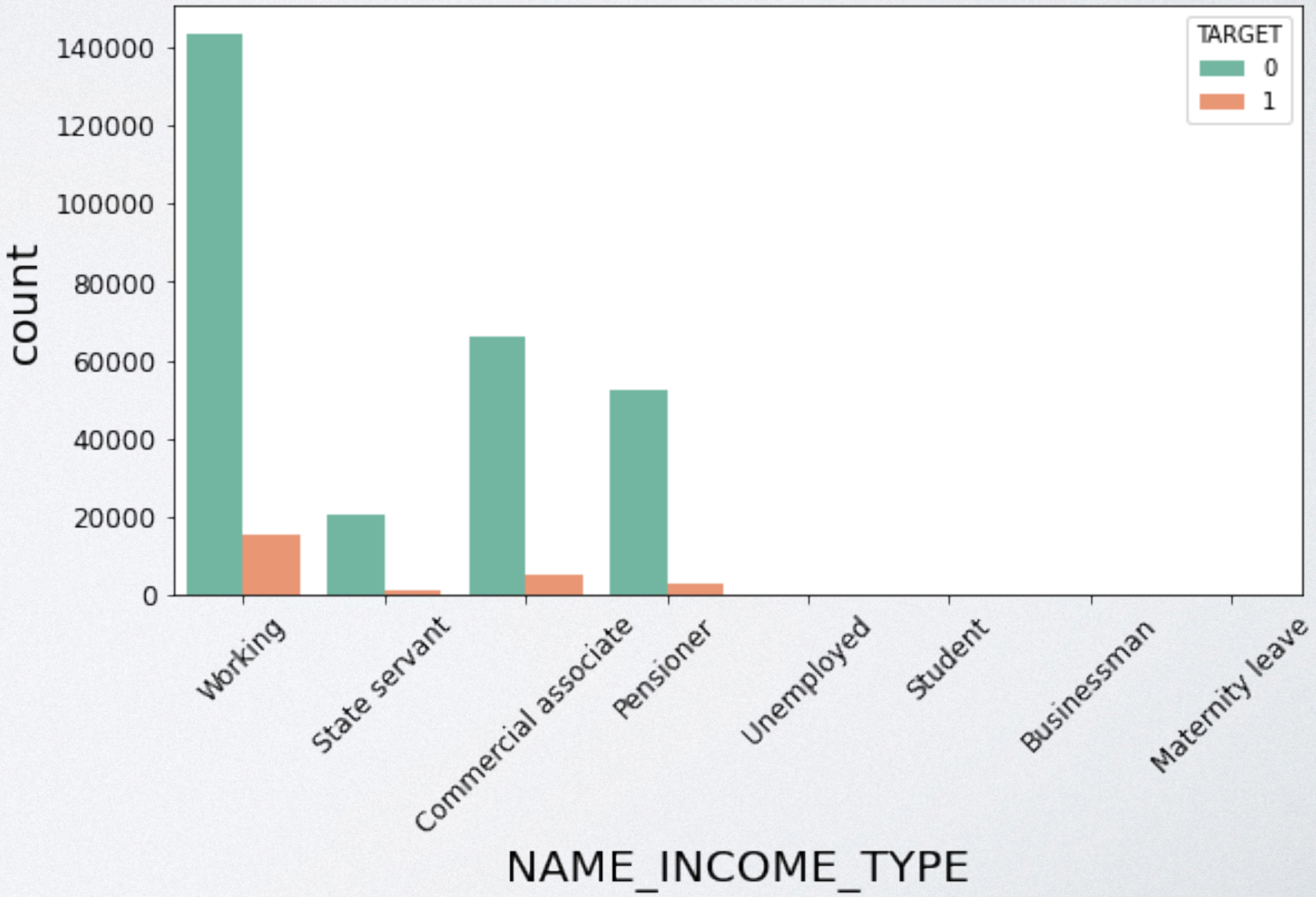
Analysing Categorical variables with respect to Target variable

Insights :

1. Upon observing following plots it is evident that females have high chance of paying back the loan compared to males . Hence gender can be considered while approving or rejecting the loan application.

2. Working people should be given priority.

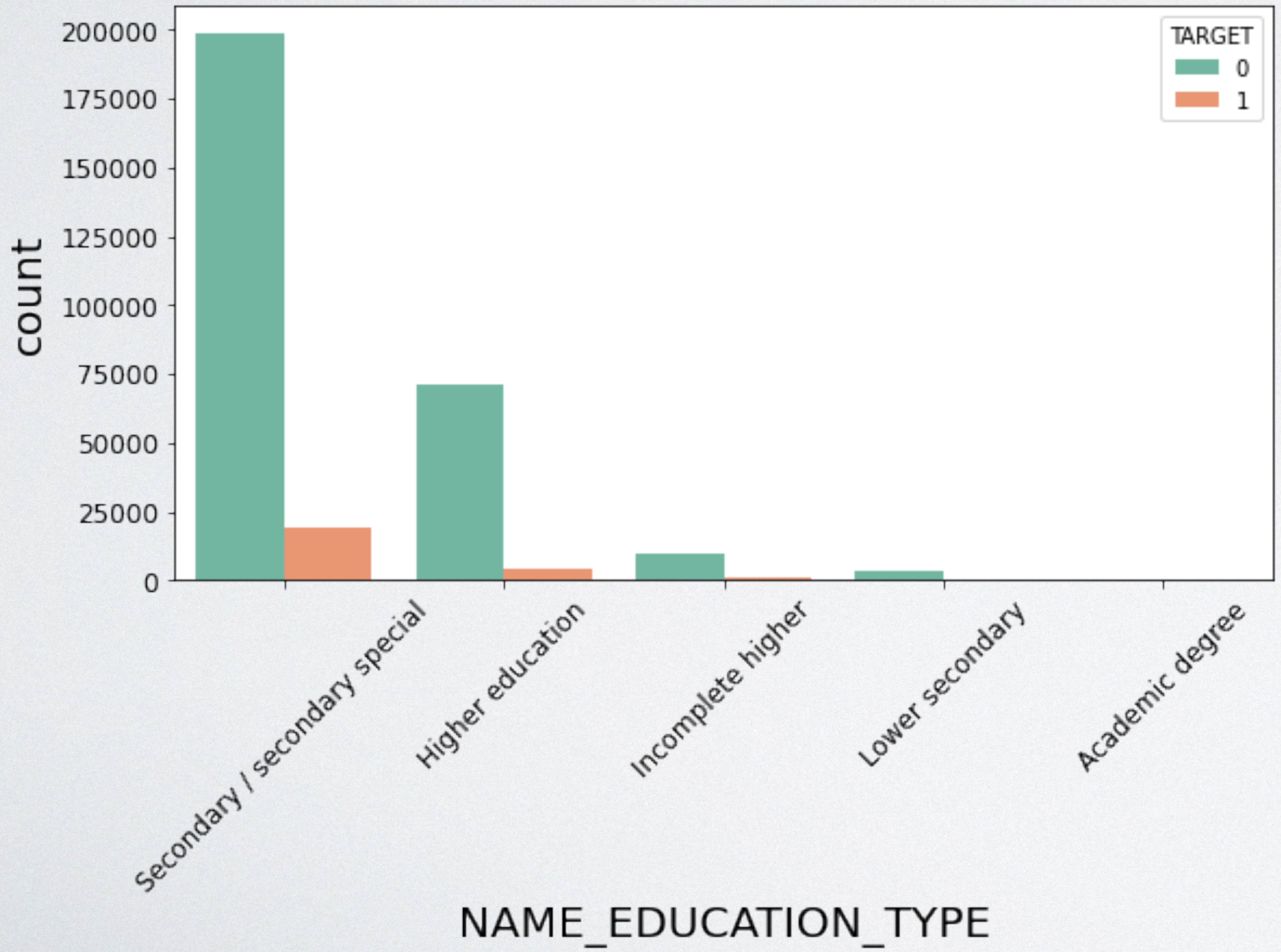


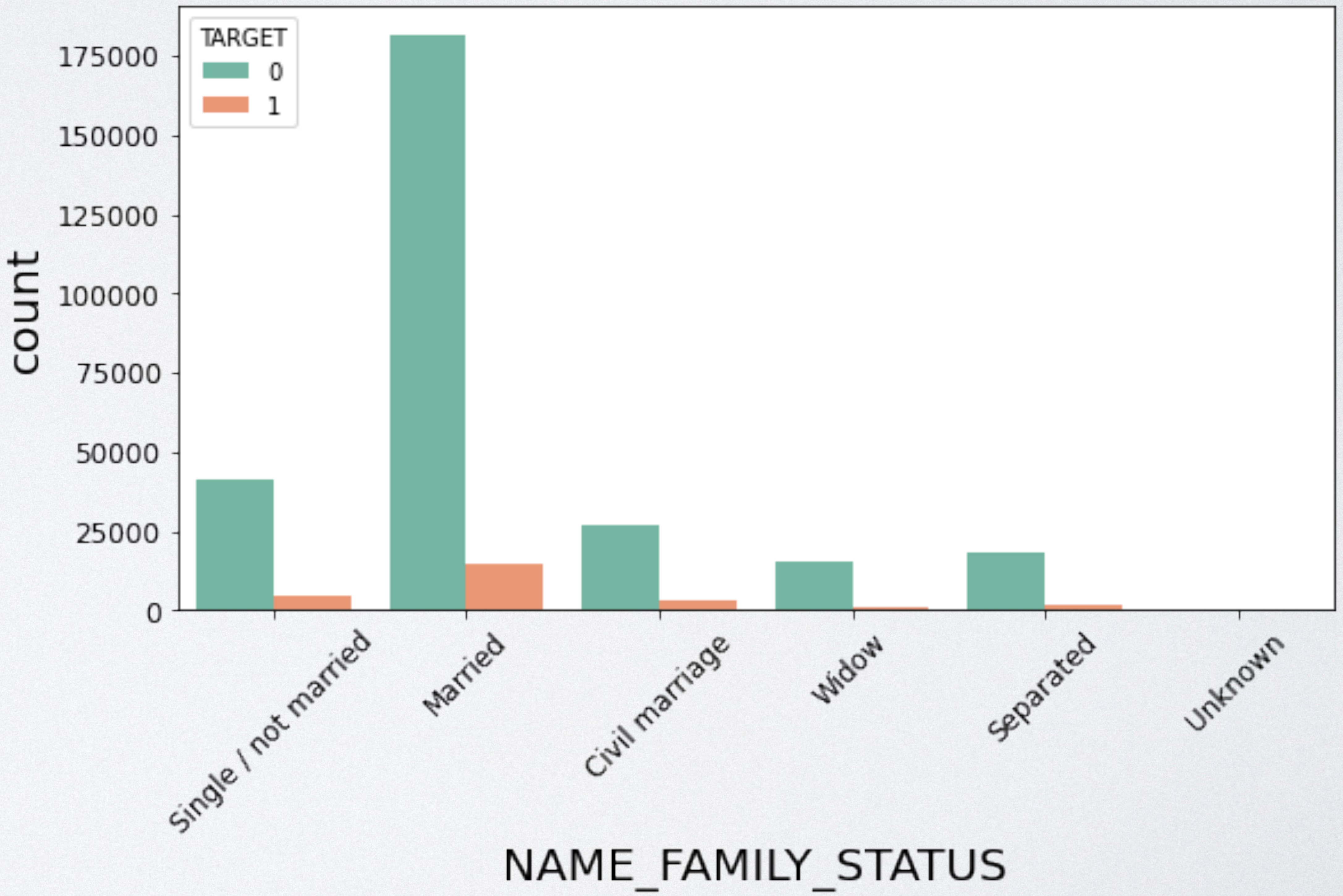


Analysing Categorical variables with respect to Target variable

Insights :

1. People with completed secondary education or Higher Education tend to pay back the loan.
2. Married Couples are the once who apply for the loan the most and they don't tend to have payment difficulties hence they should be considered.

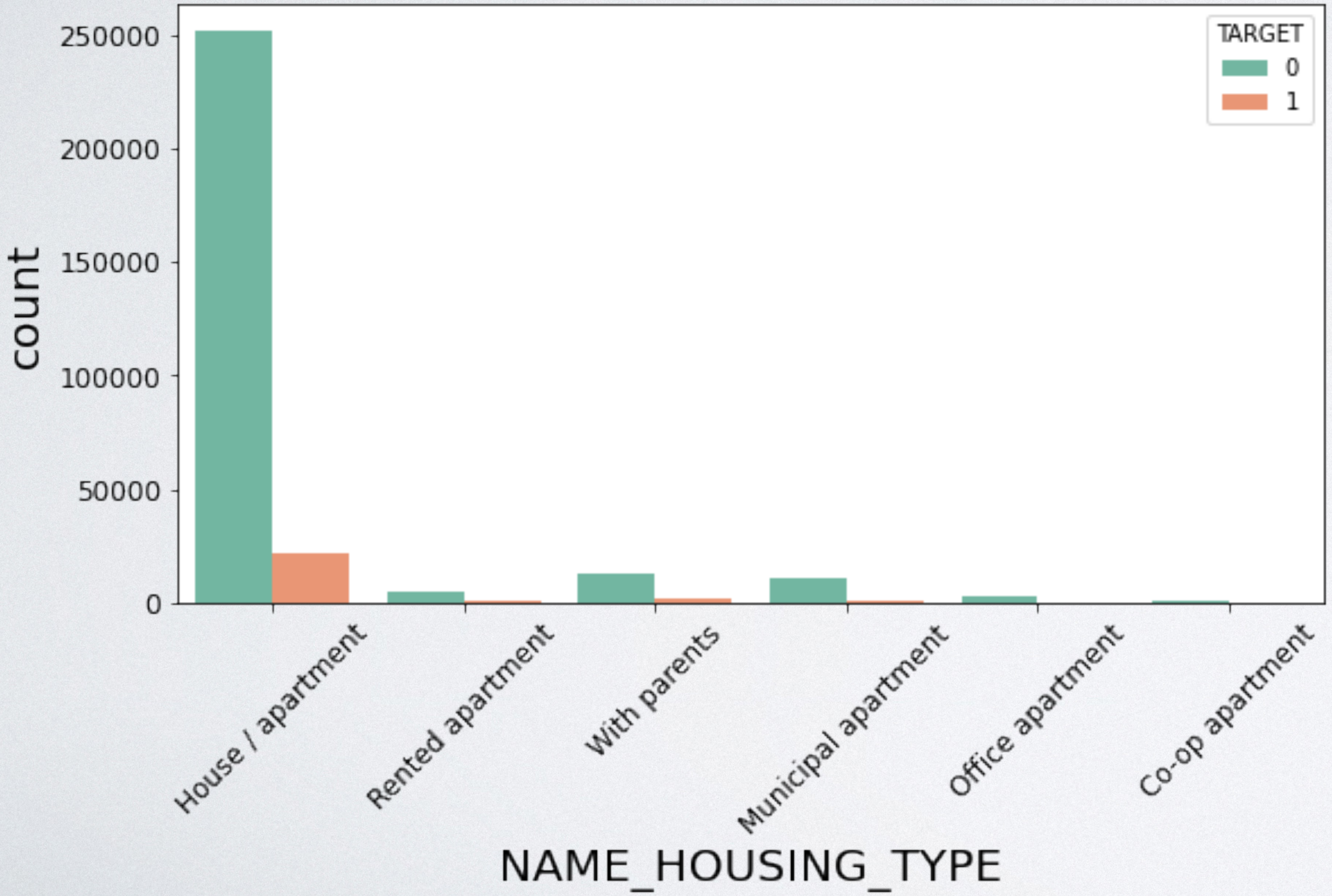


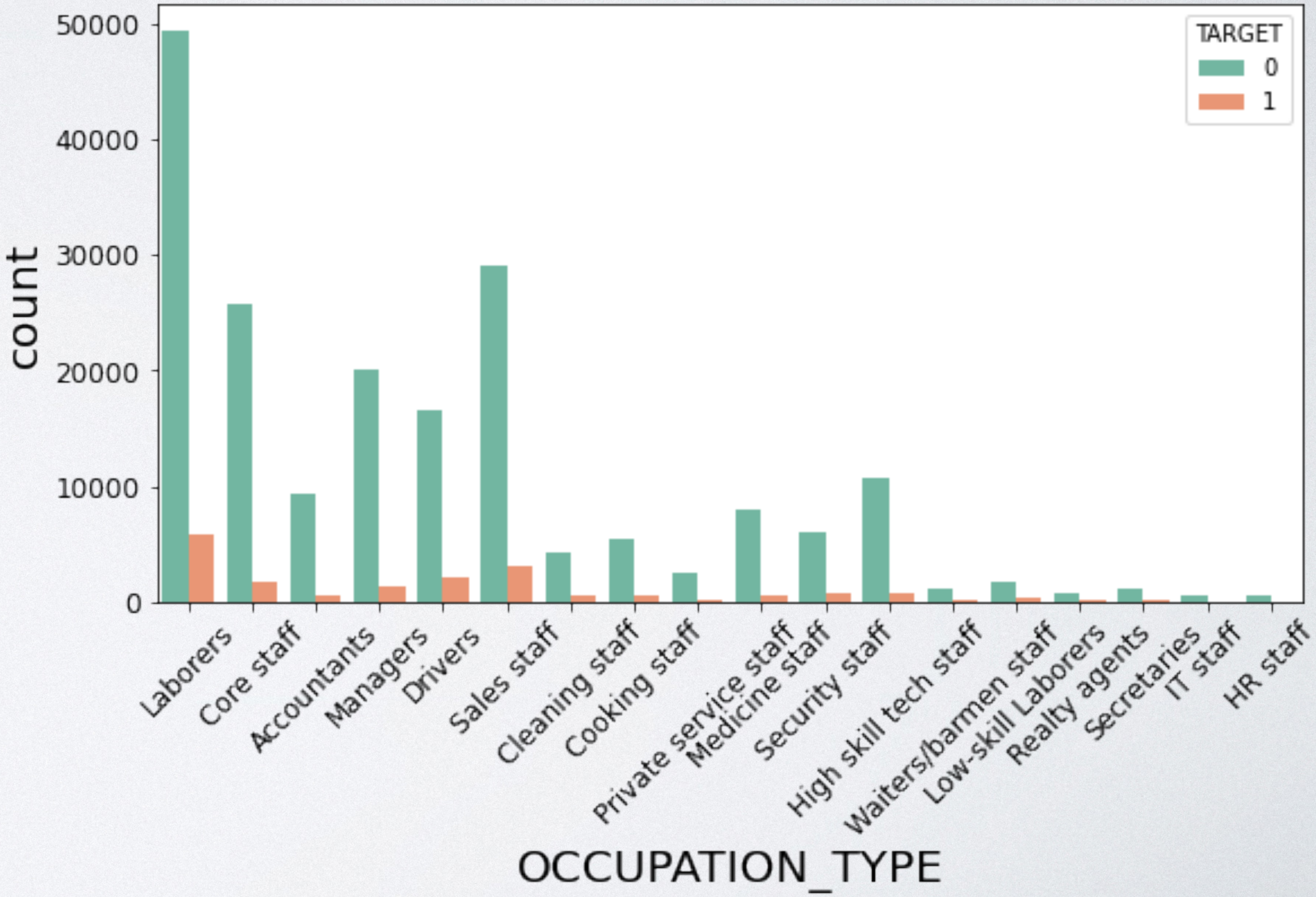


Analysing Categorical variables with respect to Target variable

Insights :

1. People with their own House/Apartment tend to pay back compared to Co-op Apartment.
2. Labourers can surely be granted small loans as they have good payback track (Explained in upcoming analysis).

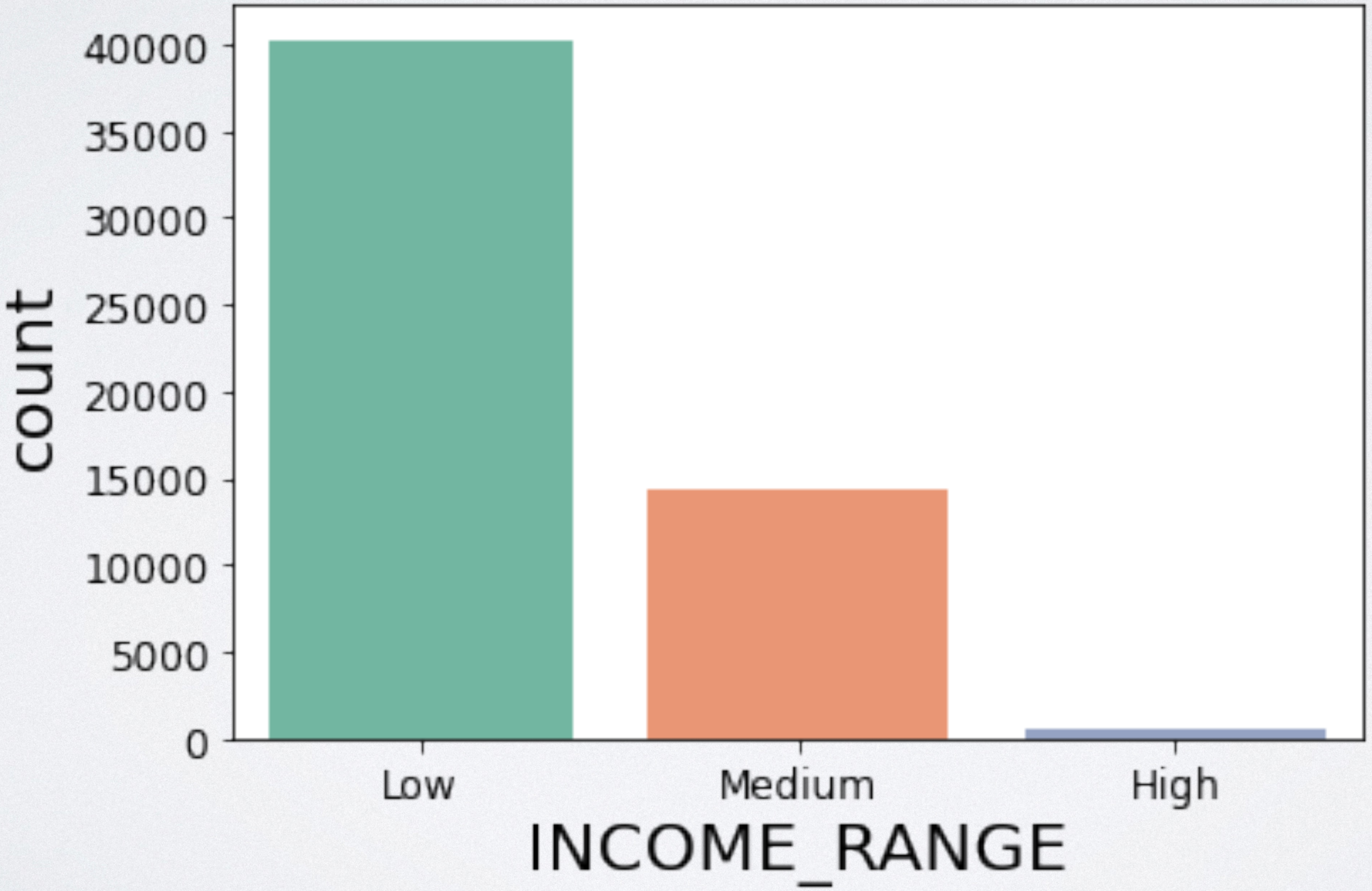




Analysing Categorical variables with respect to Target variable

Insights :

1. As Most of Labourers are at low income range . It is more feasible to give them small Loans, as they have good payback trend they can be considered for small loans.

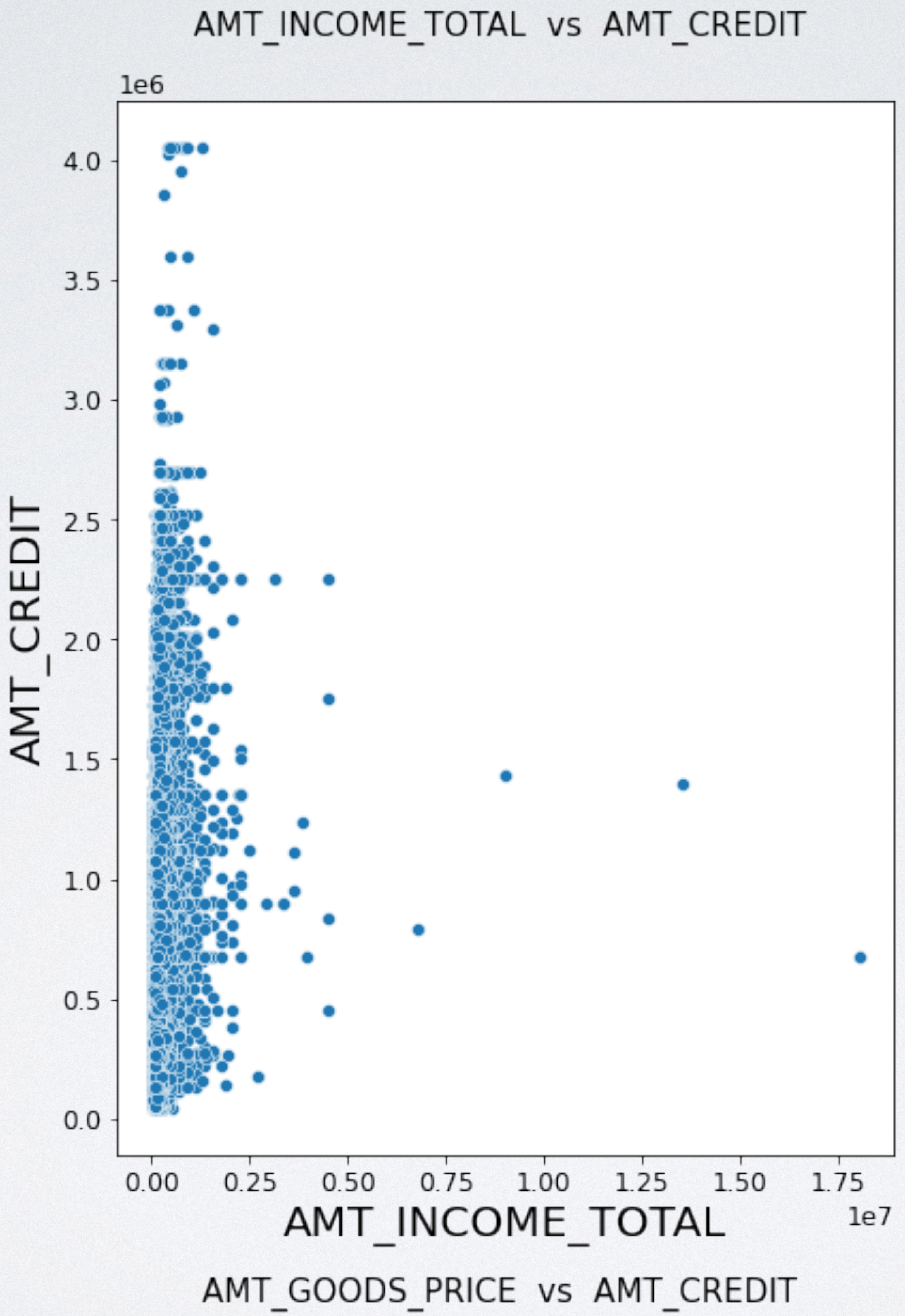


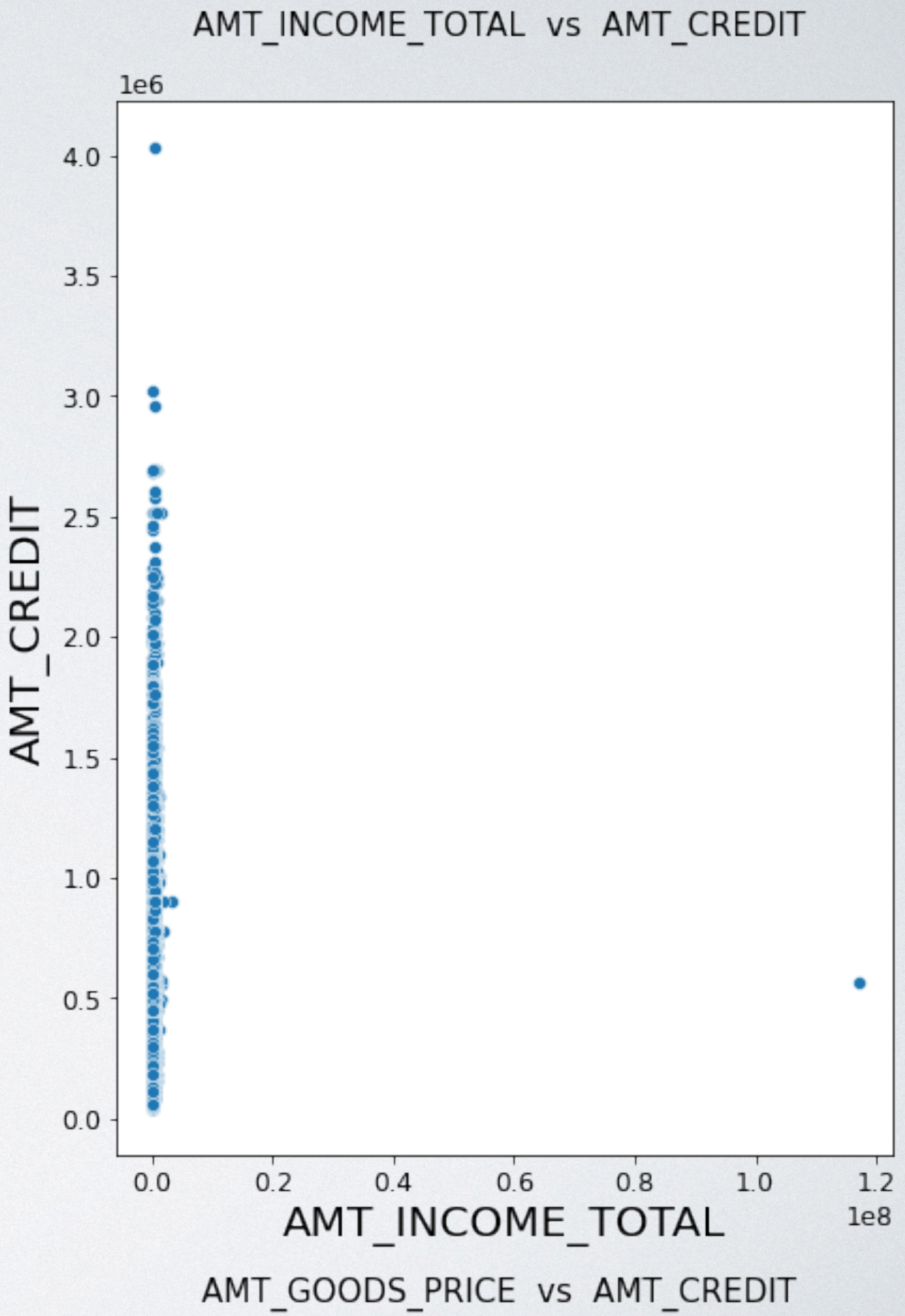
BIVARIATE ANALYSIS FOR TARGET 0 AND TARGET 1

Income vs Credit, Goods price vs Credit

Insights

1) Upon observation it can be seen that customers who have paid back the loan with any income level have good credit score compared to customers who did not pay back the loan .

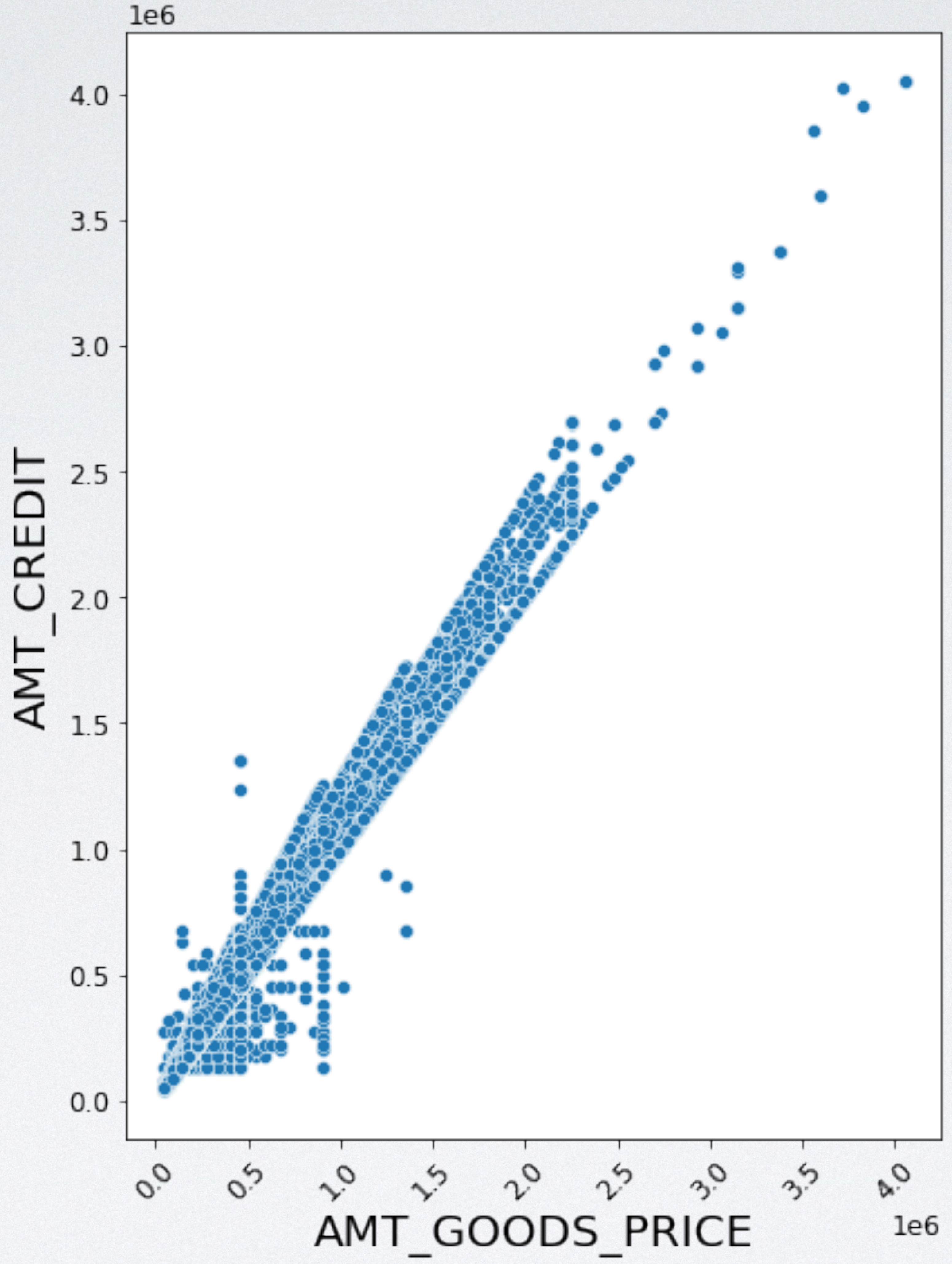


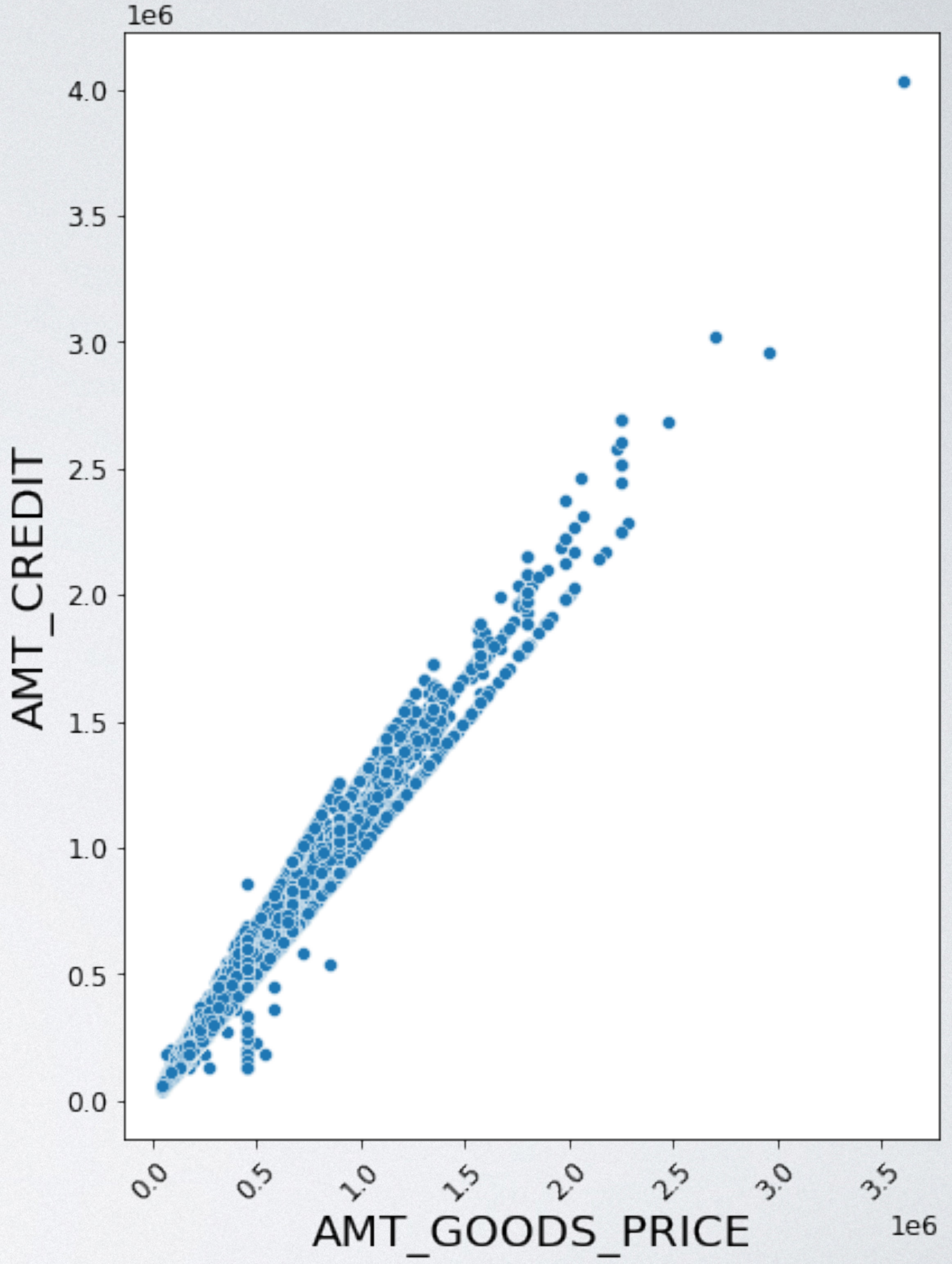


Income vs Credit, Goods price vs Credit

Insights

1) Higher the good price higher the credit amount of loan if and when money is paid back.

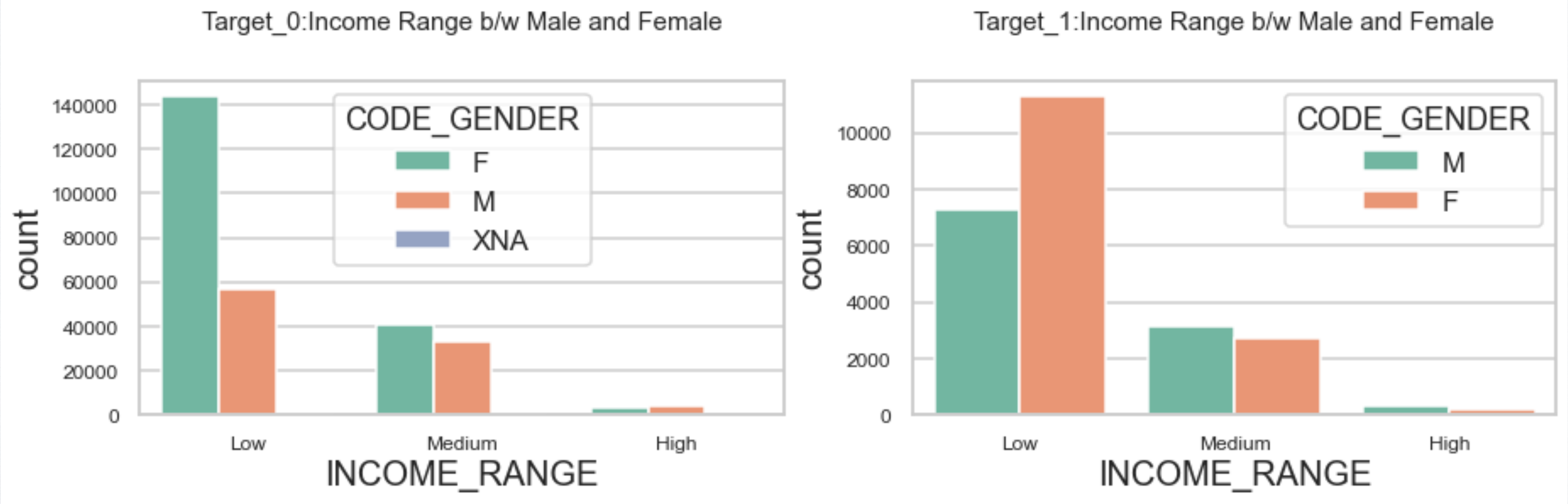




Analysing Categorical variables with respect to Target variable

Insights

1) It is evident that Females Low and Medium income range do not have payment difficulty .



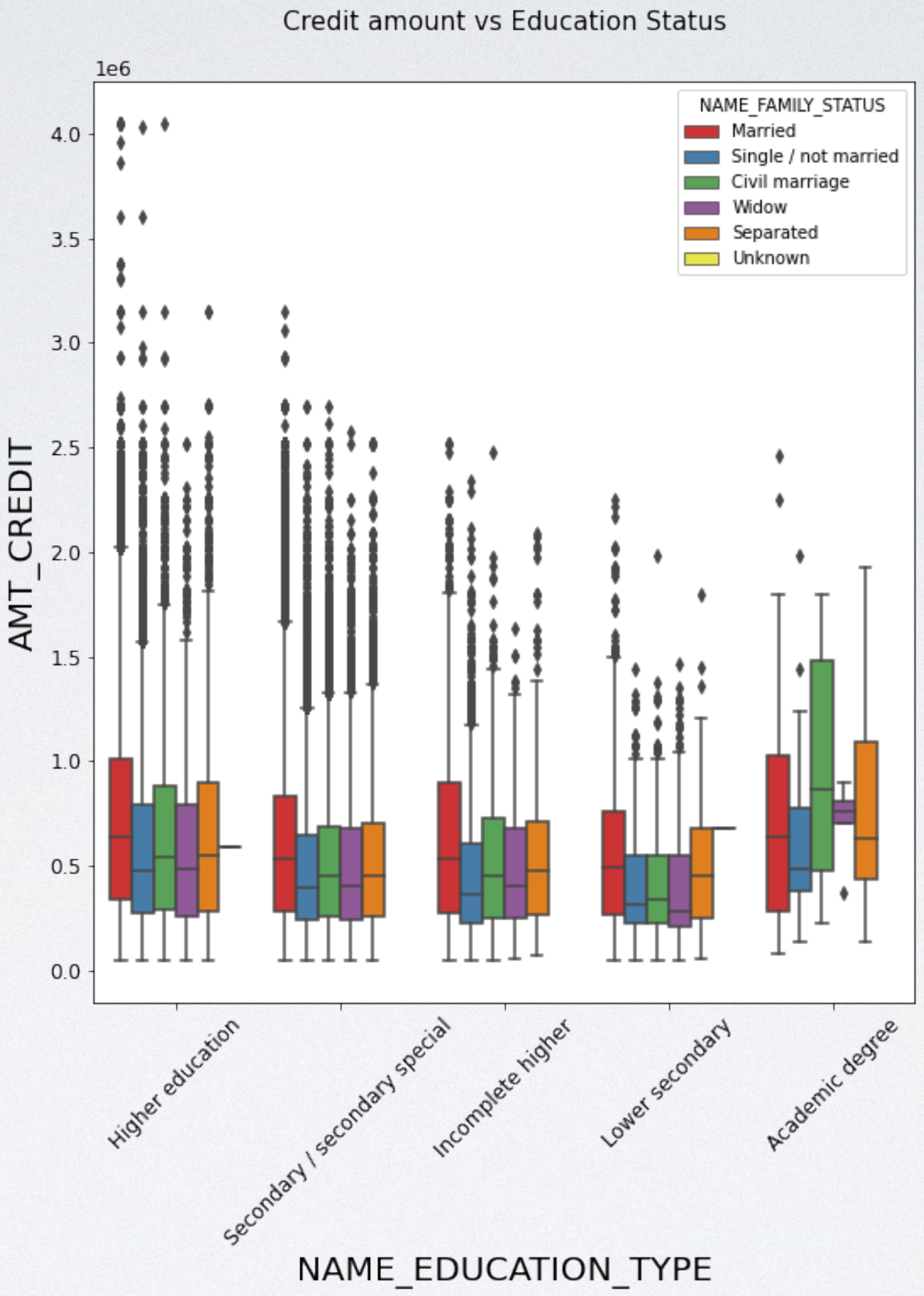
Box plotting for Credit amount

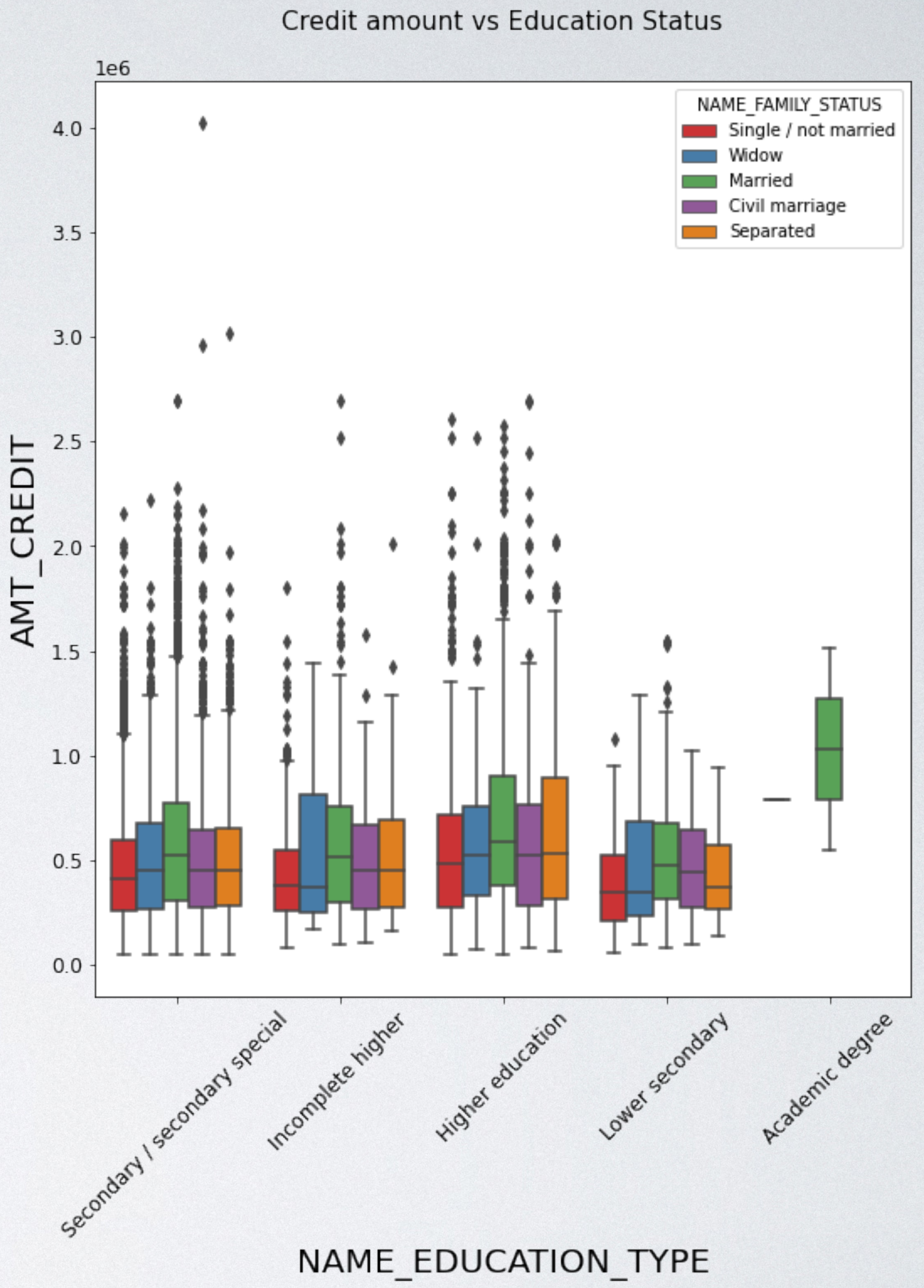
Insights

1) Married Customers with Higher Education have very high credit compared to customers with any other educational backgrounds .

2) Widows with Lower Secondary Education have the lowest credit (2nd Quantile) comparatively .

3) Highly Educated Married Customers have very high chances of paying back the loan without any difficulties .



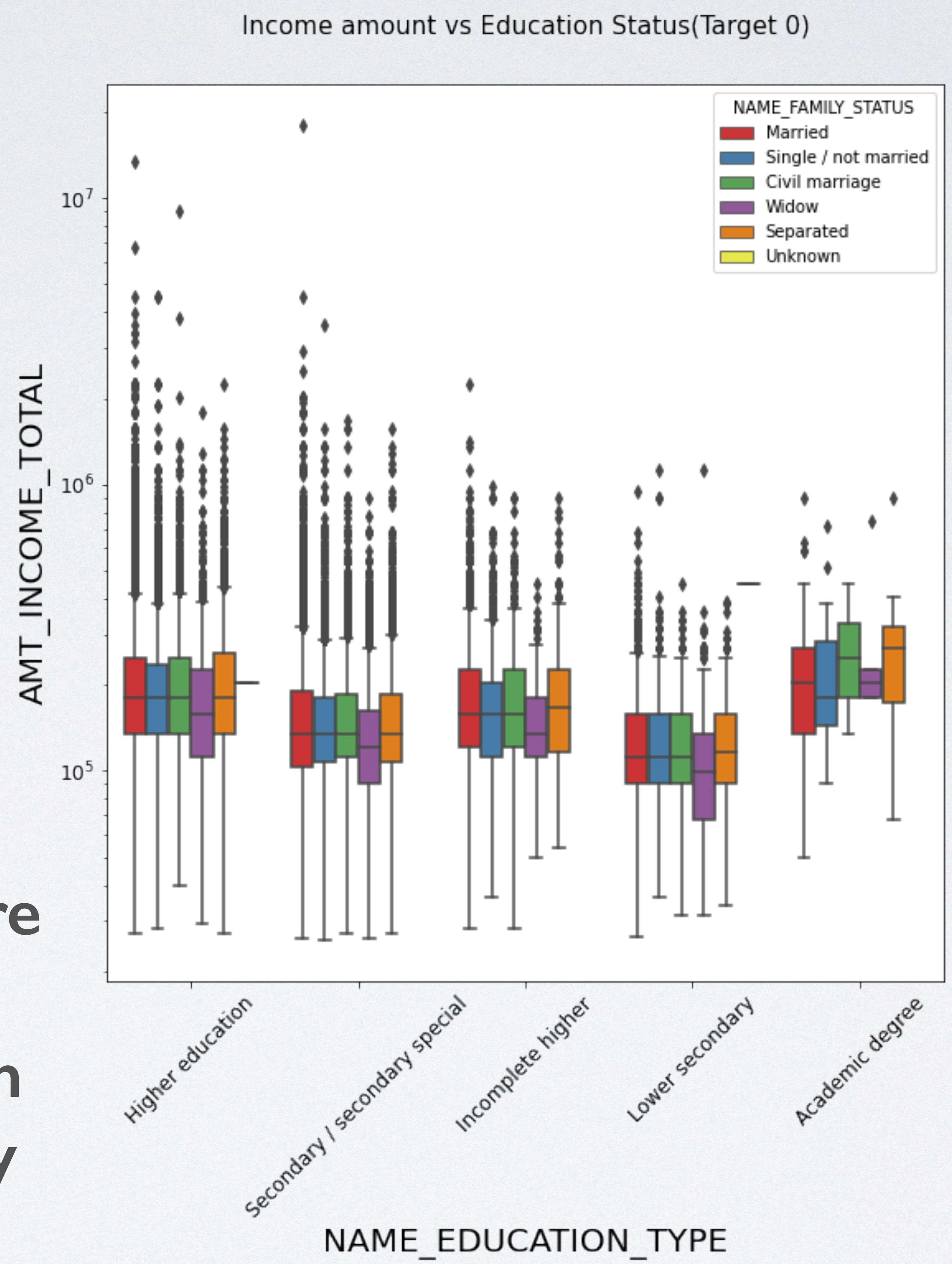


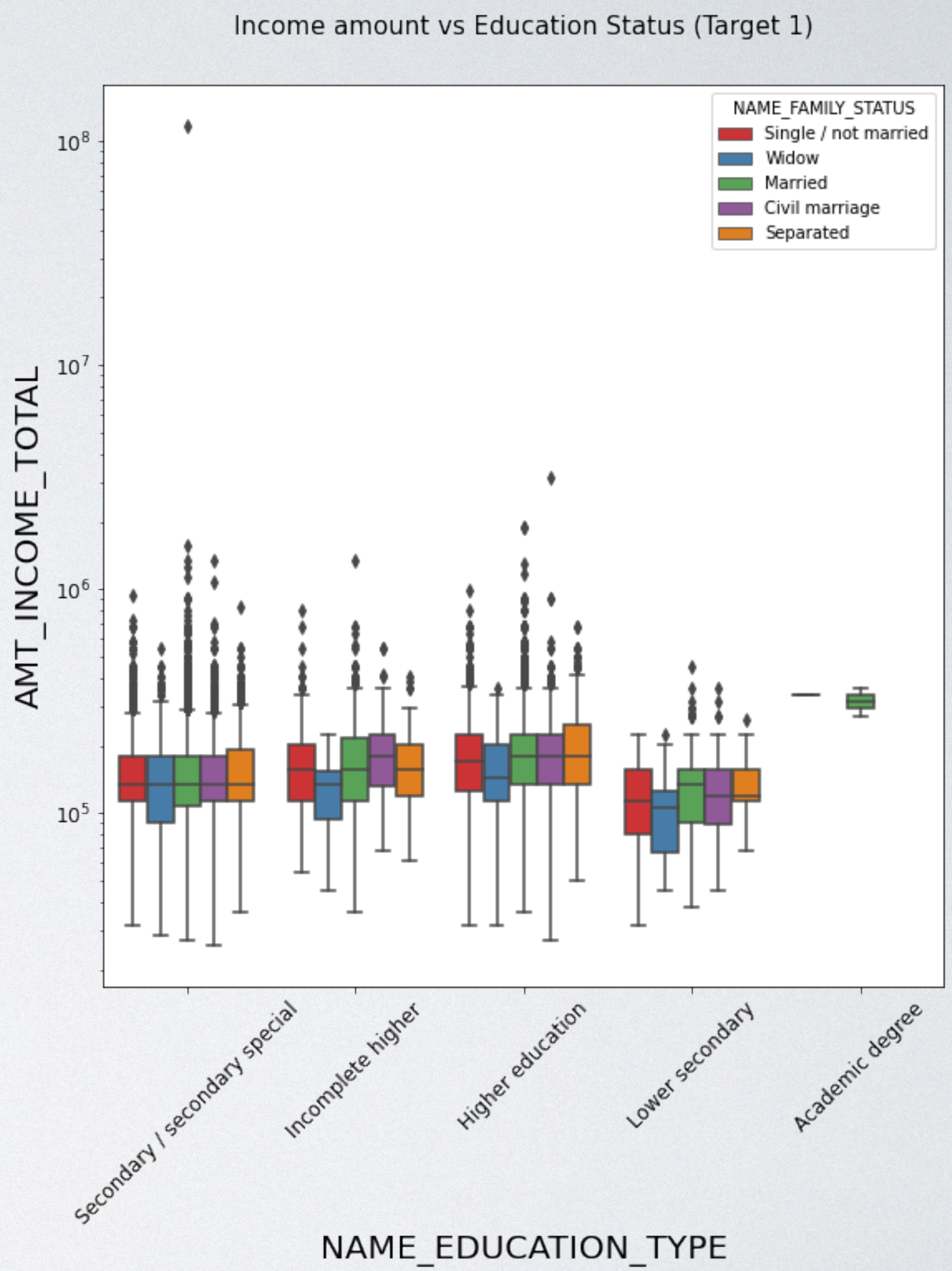
Box plotting for Income amount in logarithmic scale

Insights

1) Again Highly Educated Married Customers have higher income comparatively.
2) Customers with Higher Education but less income have more difficulty paying back the loan.

Conclusion :-
1) Upon Analysing we can conclude that customers with good education specially married customers with high education are most likely to payback the loan.
2) We can also conclude that Widows with lower secondary education are most likely to default.



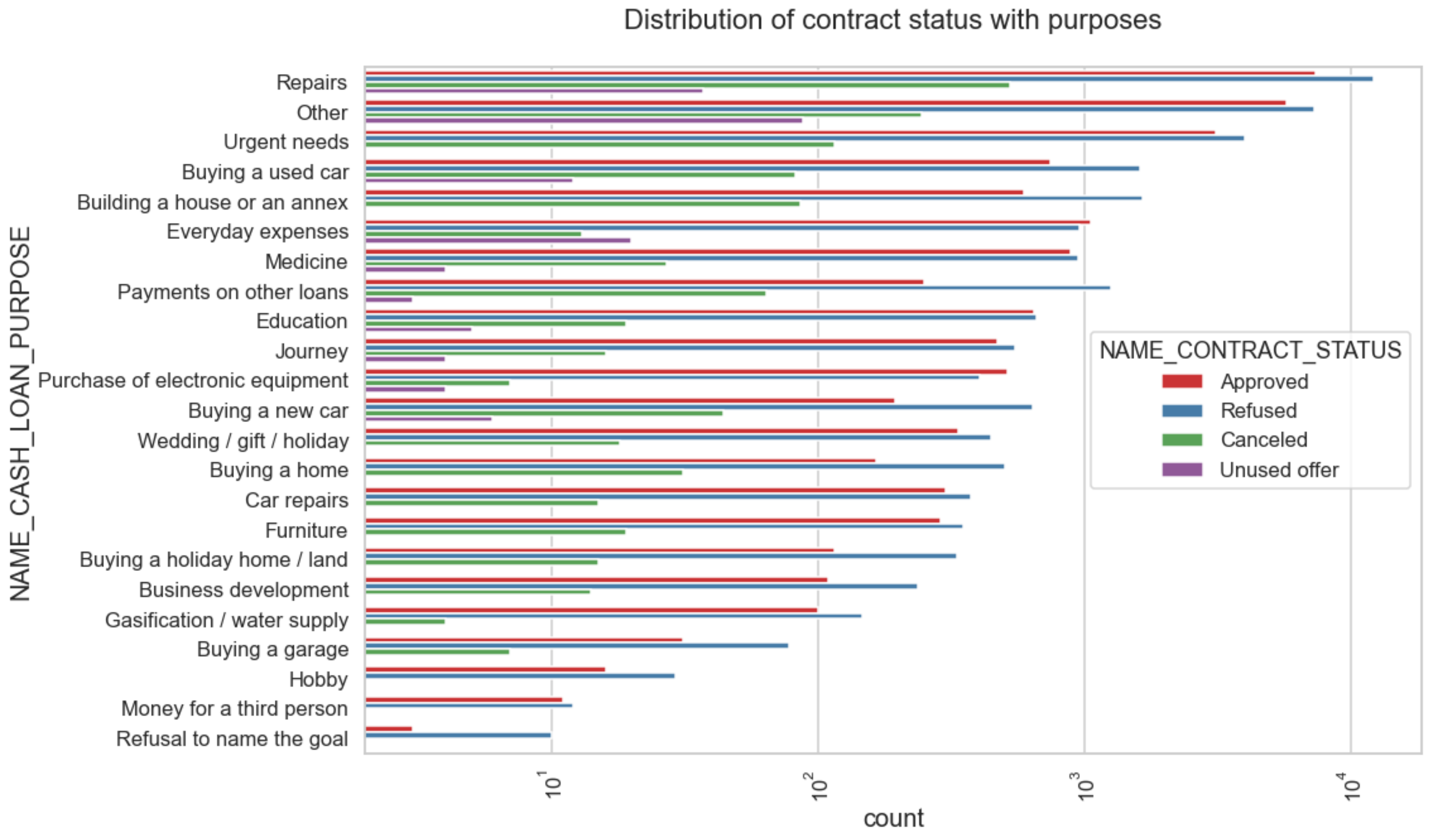


UNIVARIATE ANALYSIS

Distribution of contract status in logarithmic scale

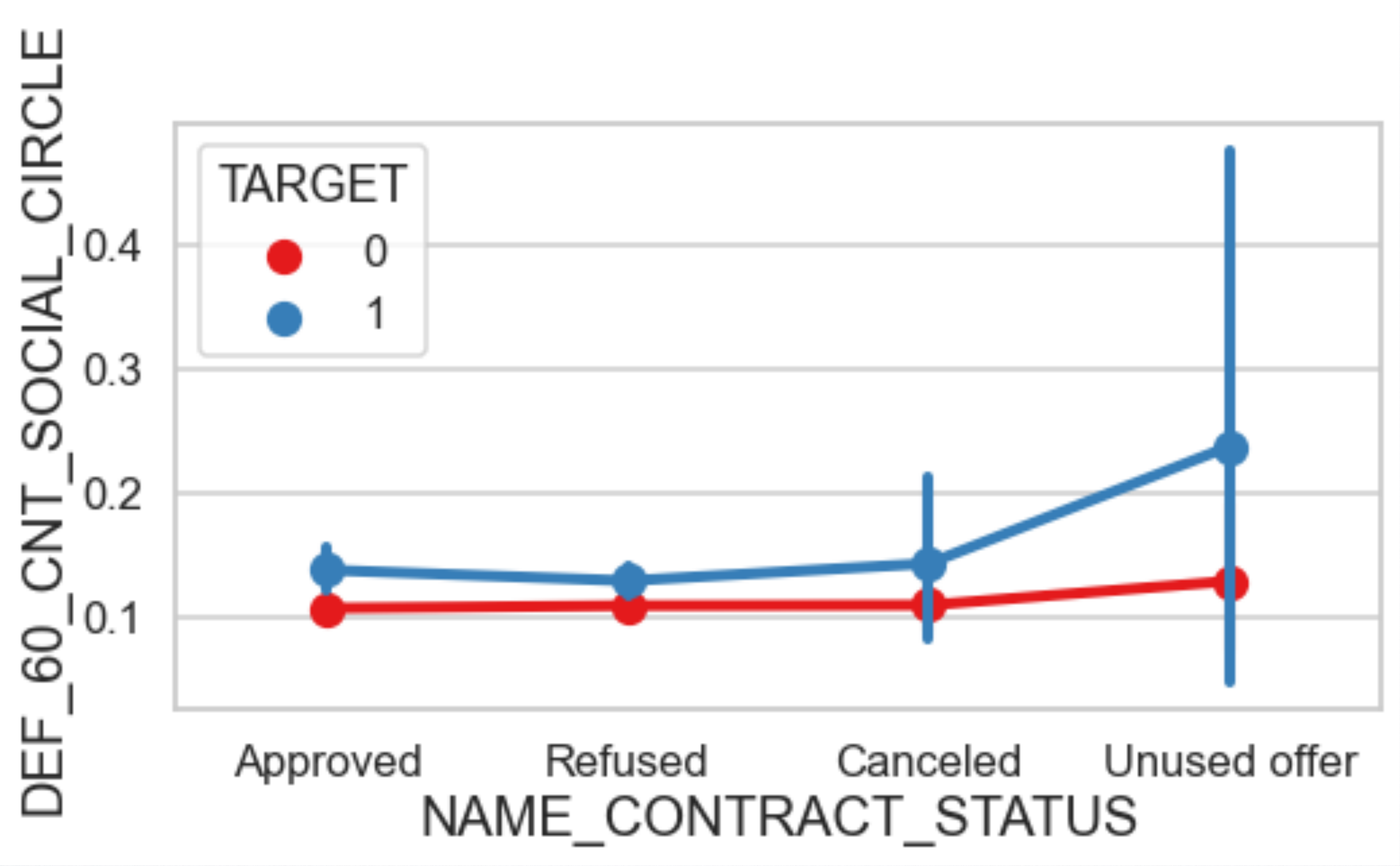
Insights

1. It is evident that the purpose repairs has the maximum loan refusal rate overall
2. Loan required for the purpose pf Medicine , Education , Journey , and Everyday Expenses has almost equal approval and rejection rates .
3. Purposes like Buying car or house or payment of other loan has very high approval rate as compared to their own rejection rate.



Distribution of contract status with Social Circle

Customers who has higher average CNT_SOCIAL_CIRCLE score , precisely more than 0.10 has higher chance to default . Hence Customers social surrounding should be taken into consideration.

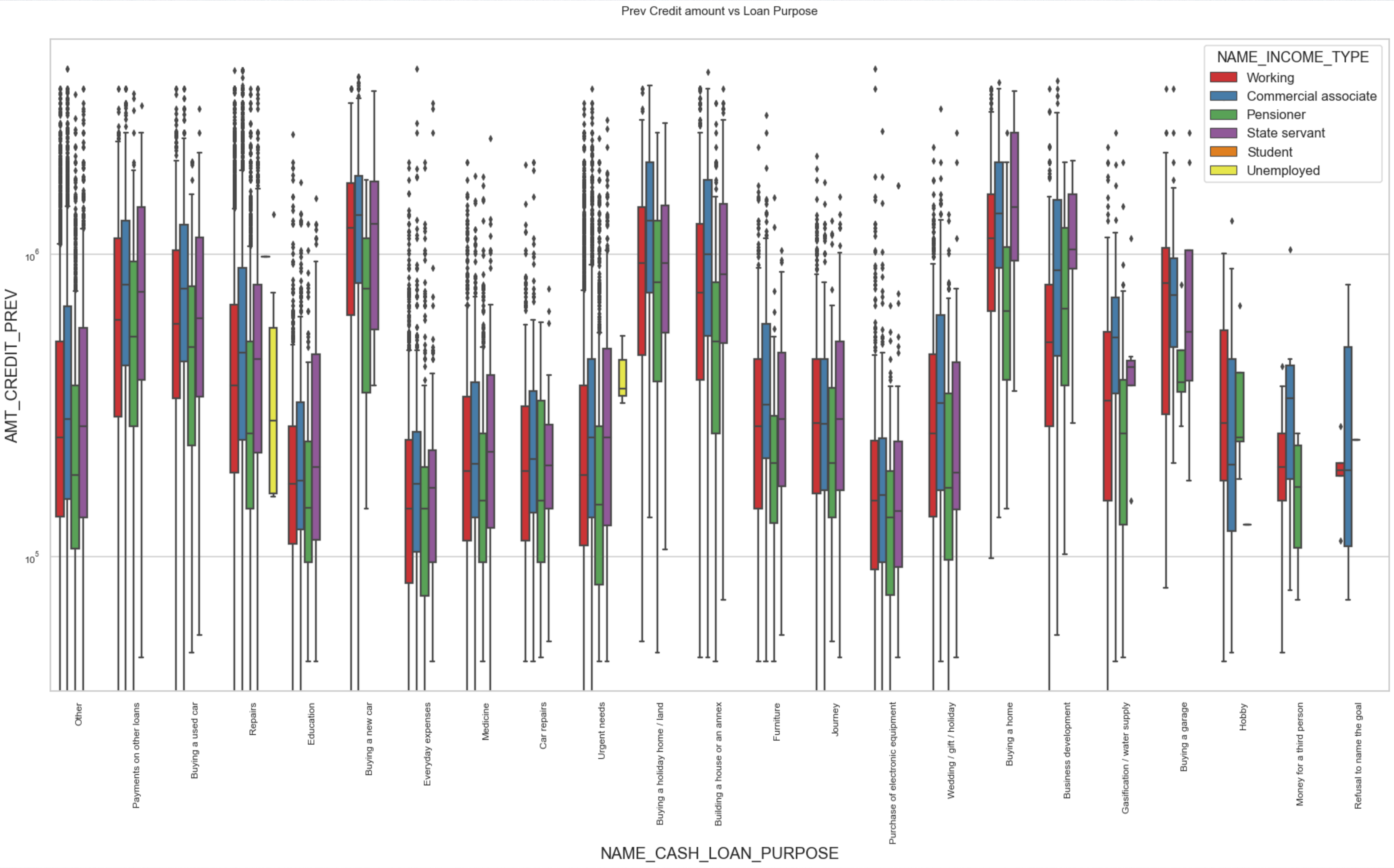


BIVARIATE ANALYSIS

Box plotting for Credit amount in logarithmic scale

Insights

1. The credit amount of Loan purposes like 'Buying a home','Buying a land','Buying a new car' and'Building a house' is higher. Income type of state servants have a significant amount of credit applied Money for third person or a Hobby is having less credits applied for.

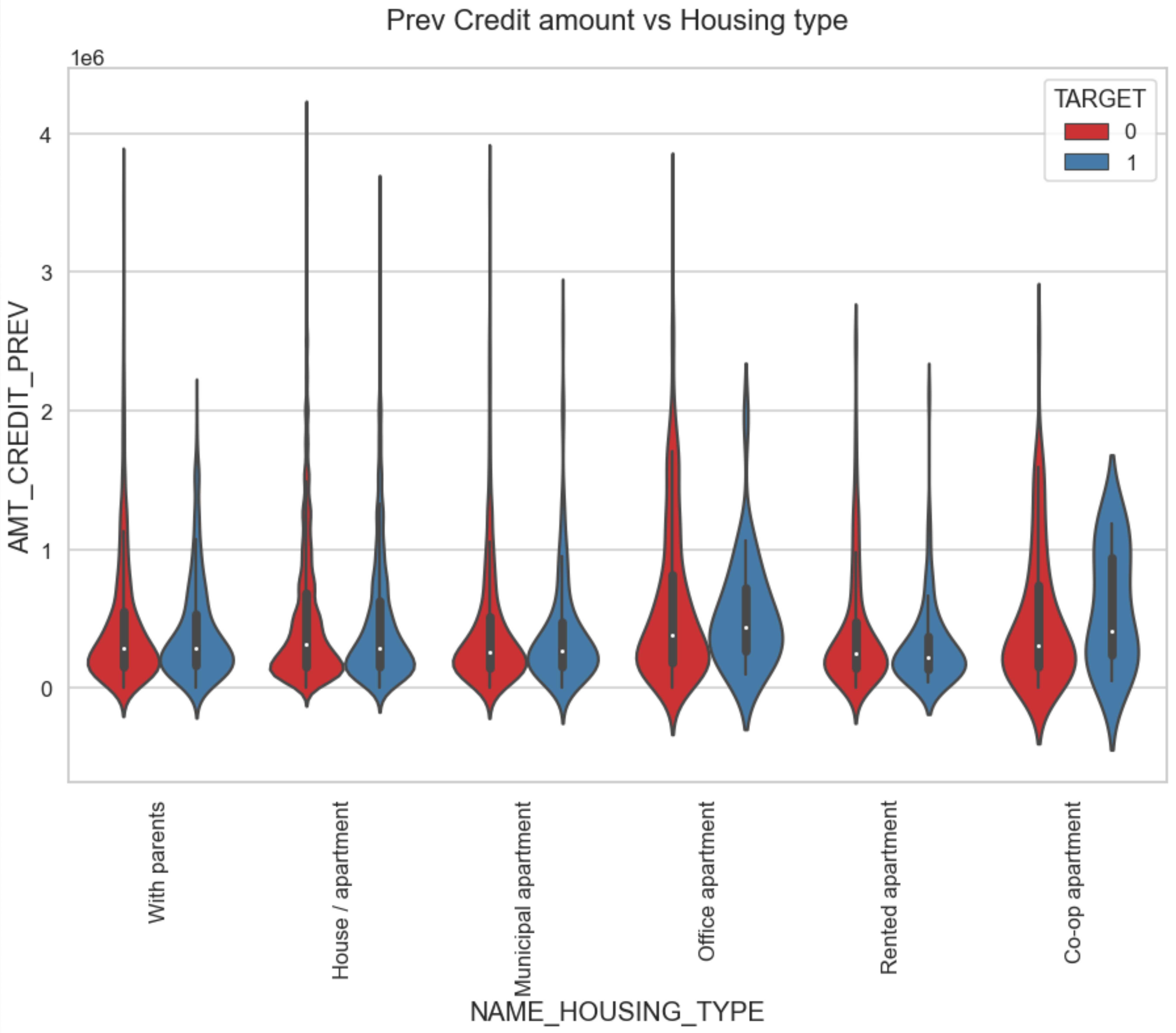


Plotting Credit amount prev vs Housing type in logarithmic scale

We can observe that customers from Housing type office apartment has least payment difficulties, and customers from Co-op apartments has highest rate of payment difficulties.

Customers Living with Parents are perusable as they as they have high rate of paying back and low rate of defaulting.

Conclusion :- Hence customers hailing from Co-op apartments are most likely to be defaulters and bank should avoid accepting their loan applications . Banks should mostly focus on customers hailing from Office Apartments or House / Apartments for assured payback of the Loan.



Insights

1) Customers like less educated widows , workers living in co-op apartments , or males in general have high chances of defaulting . Hence Bank must avoid approving their loan for assured paybacks.

2) Customers matching above mentioned profile with the loan purpose of Repair , Urgent Need , money for third person and Hobby also have very high chance to default . Hence Bank must avoid approving their loan for assured paybacks.

3) Banks should rather focus on married couples with higher education or singles with higher education as they have least payment difficulties.

4) Customers like Businessman , Pensioner and Students with housing type House and Living with Parents should be aggressively pursued as they have least default chances.