

# EDA case study

Presented by -

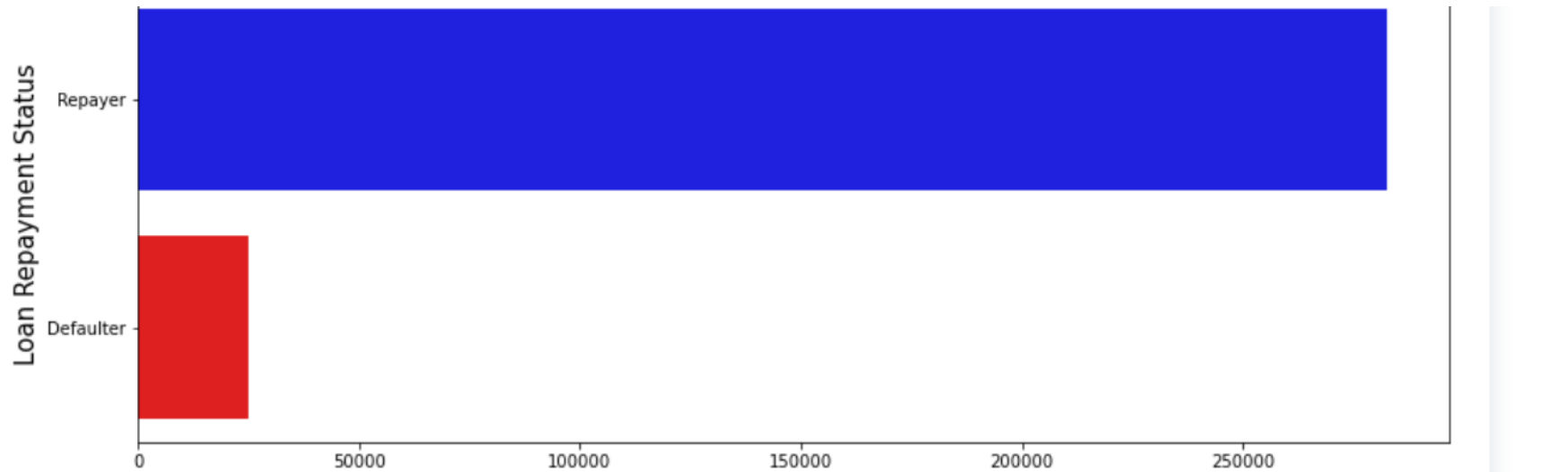
1. Ramya shree T V ( 7892138625 , [ramyashreetv@gmail.com](mailto:ramyashreetv@gmail.com) )
2. Ayushman joshi ( 7827517323 , [erayushman@gmail.com](mailto:erayushman@gmail.com) )

# Exploratory Data Analysis

## Steps

- 1.Data sourcing
- 2.Data understanding
- 3.Checking the missing values and null values
- 4.Imbalance in Data
- 5.Categorical Data Analysis
- 6.Categorical segmented Univariate Analysis
- 7.Categorical Bi/Multivariate analysis
- 8.Numeric Data Analysis
- 9.Bi-furcation of databased based on TARGET
- 10.data Correlation Matrix
- 11.Numerical segmented Univariate Analysis
- 12.Numerical Bi/Multivariate analysis

# Imbalance Plotting (Repayer Vs Defaulter)

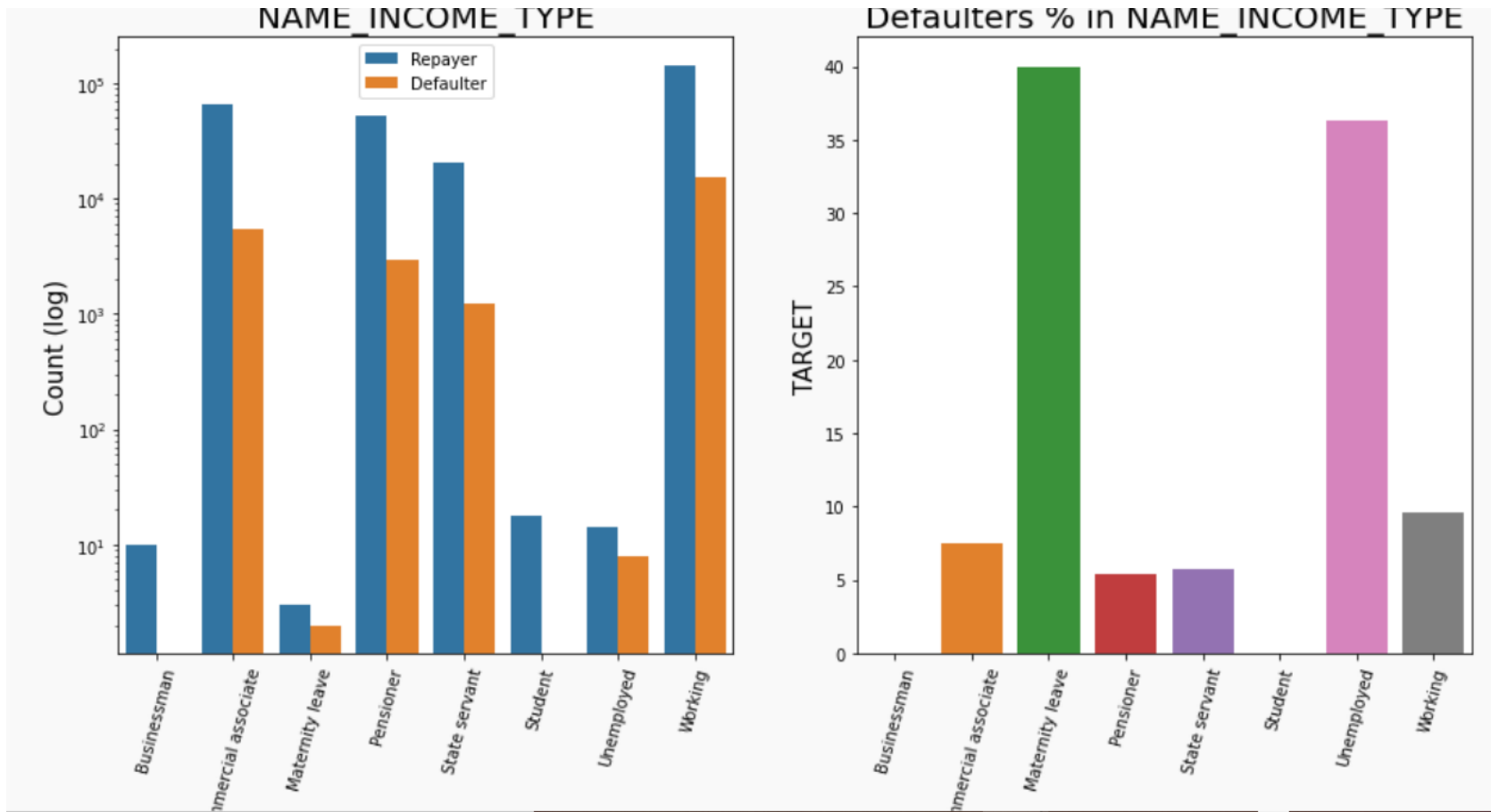


Repayer Percentage is 91.93%.

Defaulter Percentage is 8.07%.

Imbalance Ratio with respect to Repayer and Defaulter is given: 11.39/1 (approx.).

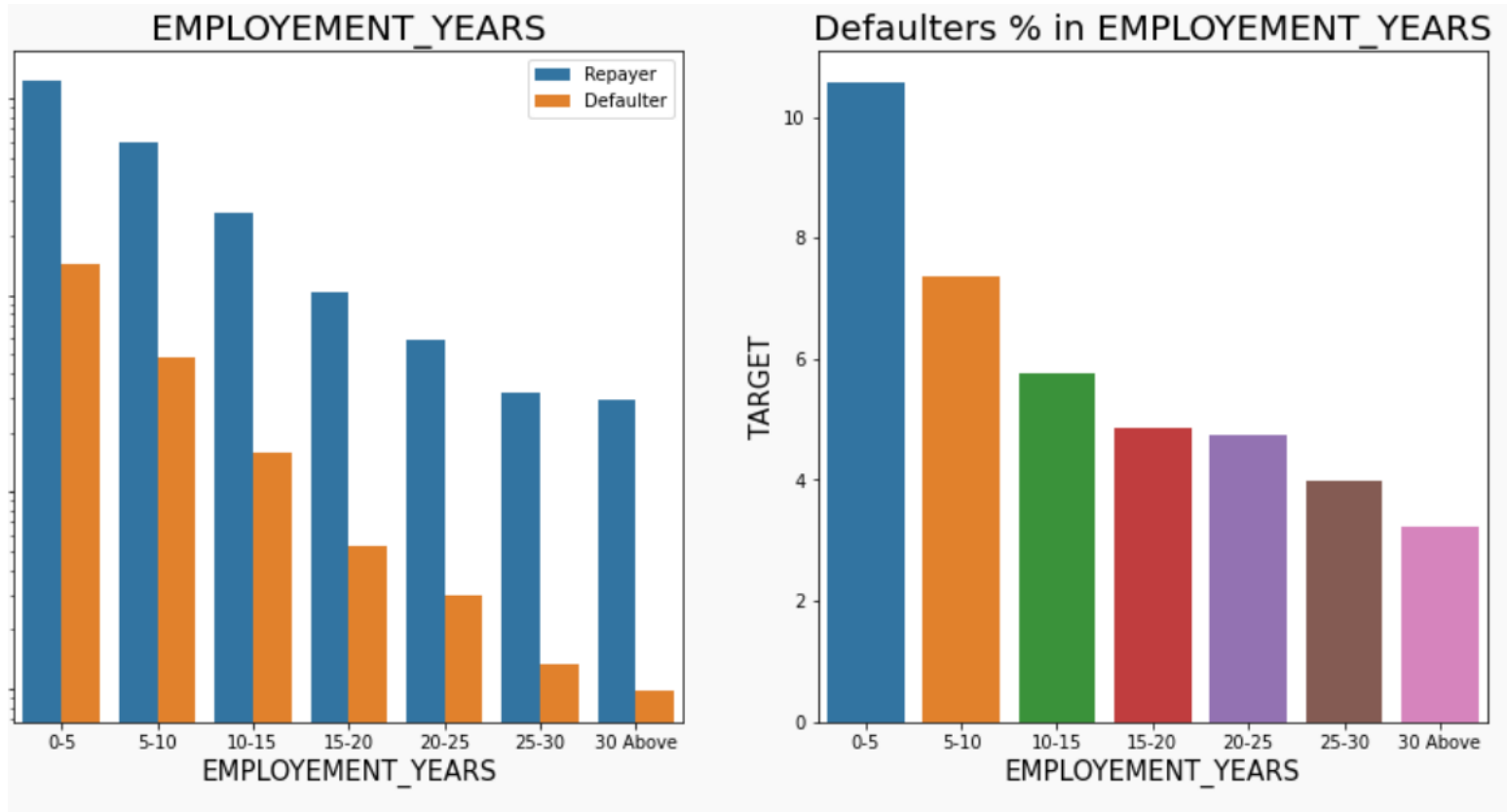
# Analyzing Income Type based on loan repayment status



### **Inferences:**

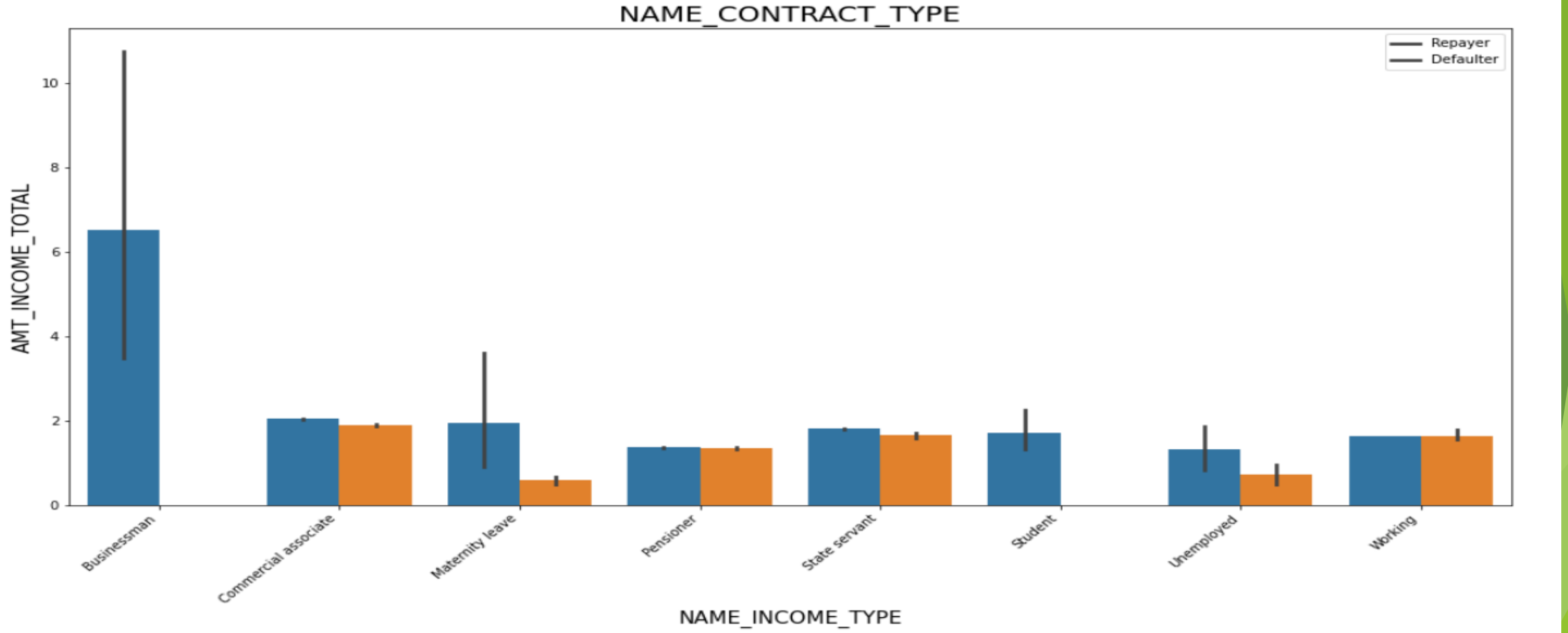
- **Most of applicants for loans income type is Working, followed by Commercial associate, Pensioner and State servant.**
- **The applicants who are on Maternity leave have defaulting percentage of 40% which is the highest, followed by Unemployed (37%). The rest under average around 10% defaulters.**
- **Student and Businessmen though less in numbers, do not have default record. Safest two categories for providing loan.**

# Analyzing Employment\_Year based on loan repayment status



- Majority of the applicants having working experience between 0-5 years are defaulters. The defaulting rating of this group is also the highest which is around 10%.
- With increase of employment year, defaulting rate is gradually decreasing.
- The people who are having 40+ years experience have less than 1% default rate.

# Categorical Bivariate



- It can be seen that Businessman's income is the highest and the estimated range with default 95% confidence level seem to indicate that the income of a Businessman could be in the range of slightly close to 4 lakhs and slightly above 10 lakhs



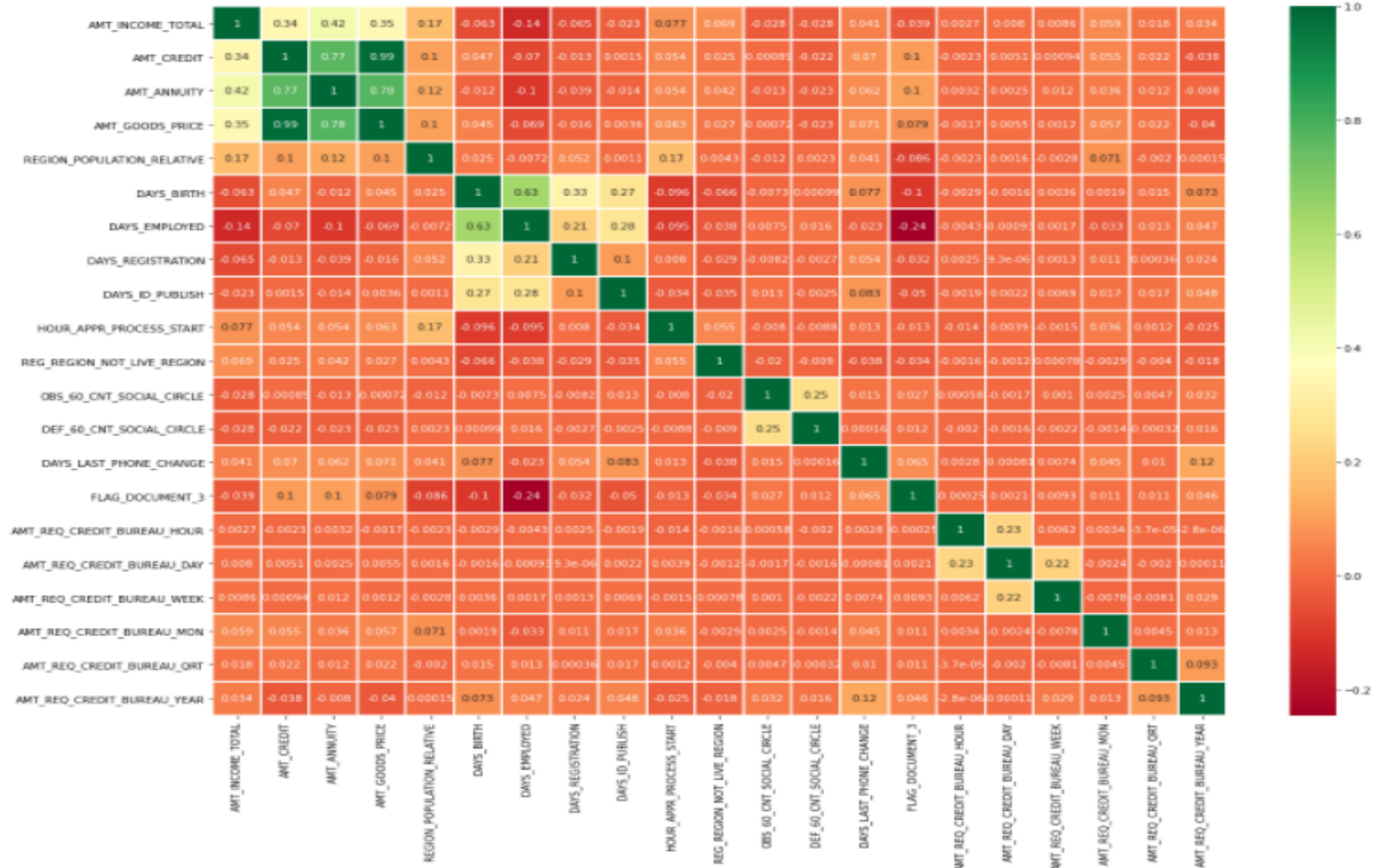
# Correlation between numeric variable

Correlating factors amongst repayers

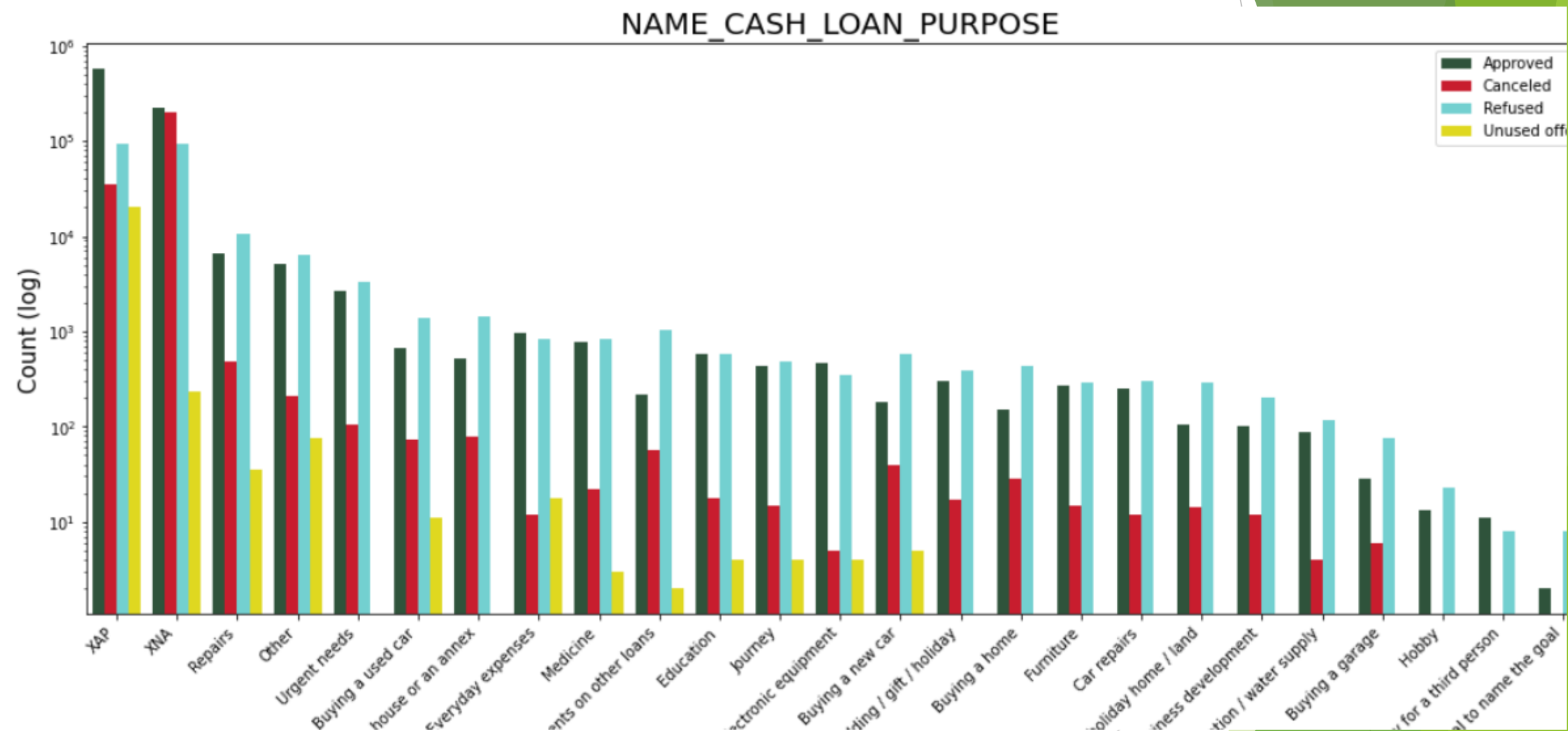
1. Credit amount is highly correlated with:

- Goods Price Amount
- Loan Annuity
- Total Income

2. We can also see that repayers have high correlation in number of days employed.

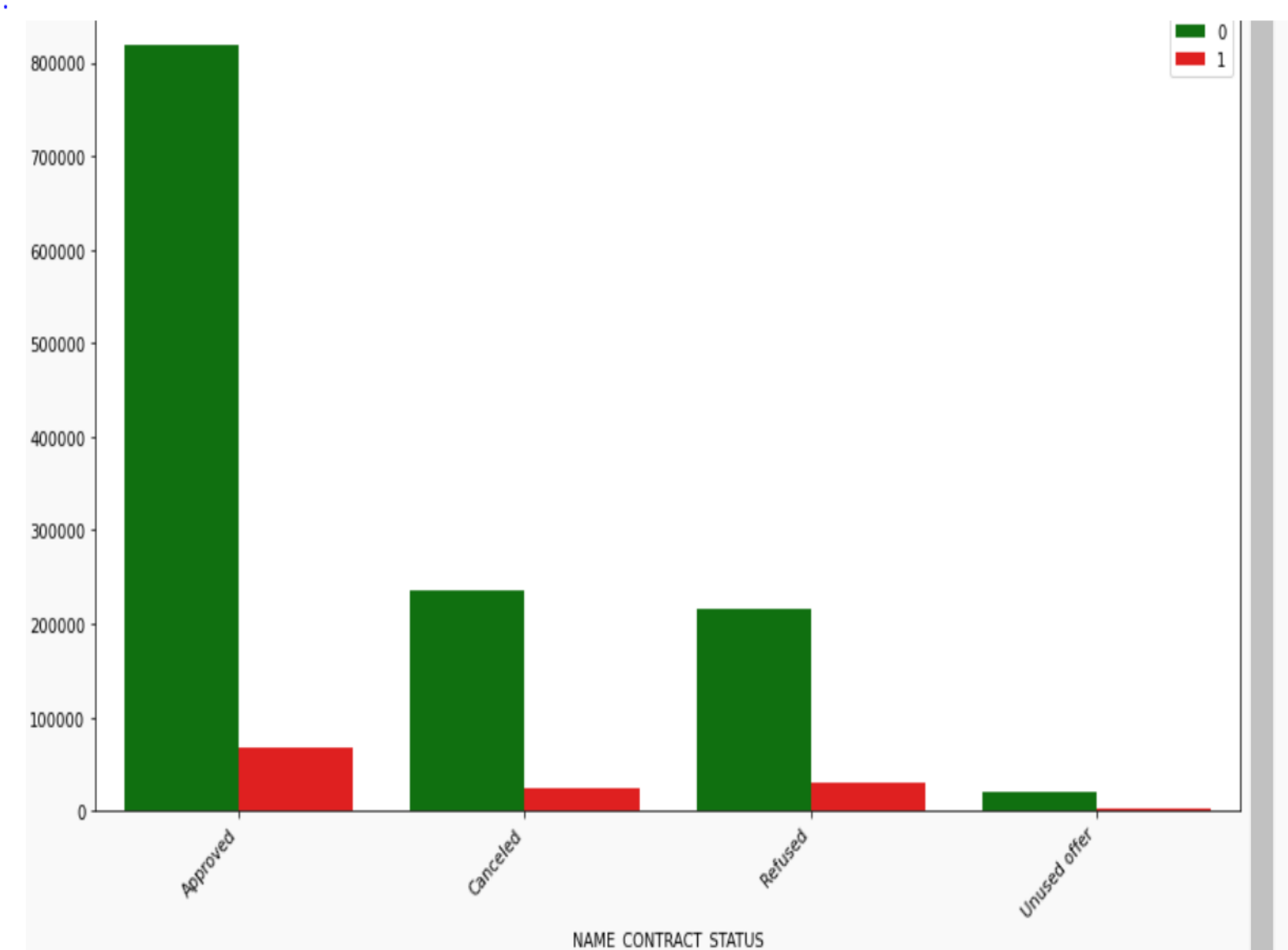


# Plotting Contract Status vs purpose of the loan



- Loan purpose has high number of unknown values (XAP, XNA)
- Loan taken for the purpose of Repairs looks to have highest default rate
- Huge number application have been rejected by bank or refused by client which are applied for Repair or Other. from this we can infer that repair is considered high risk by bank. Also, either they are rejected or bank offers loan on high interest rate which is not feasible by the clients and they refuse the loan.

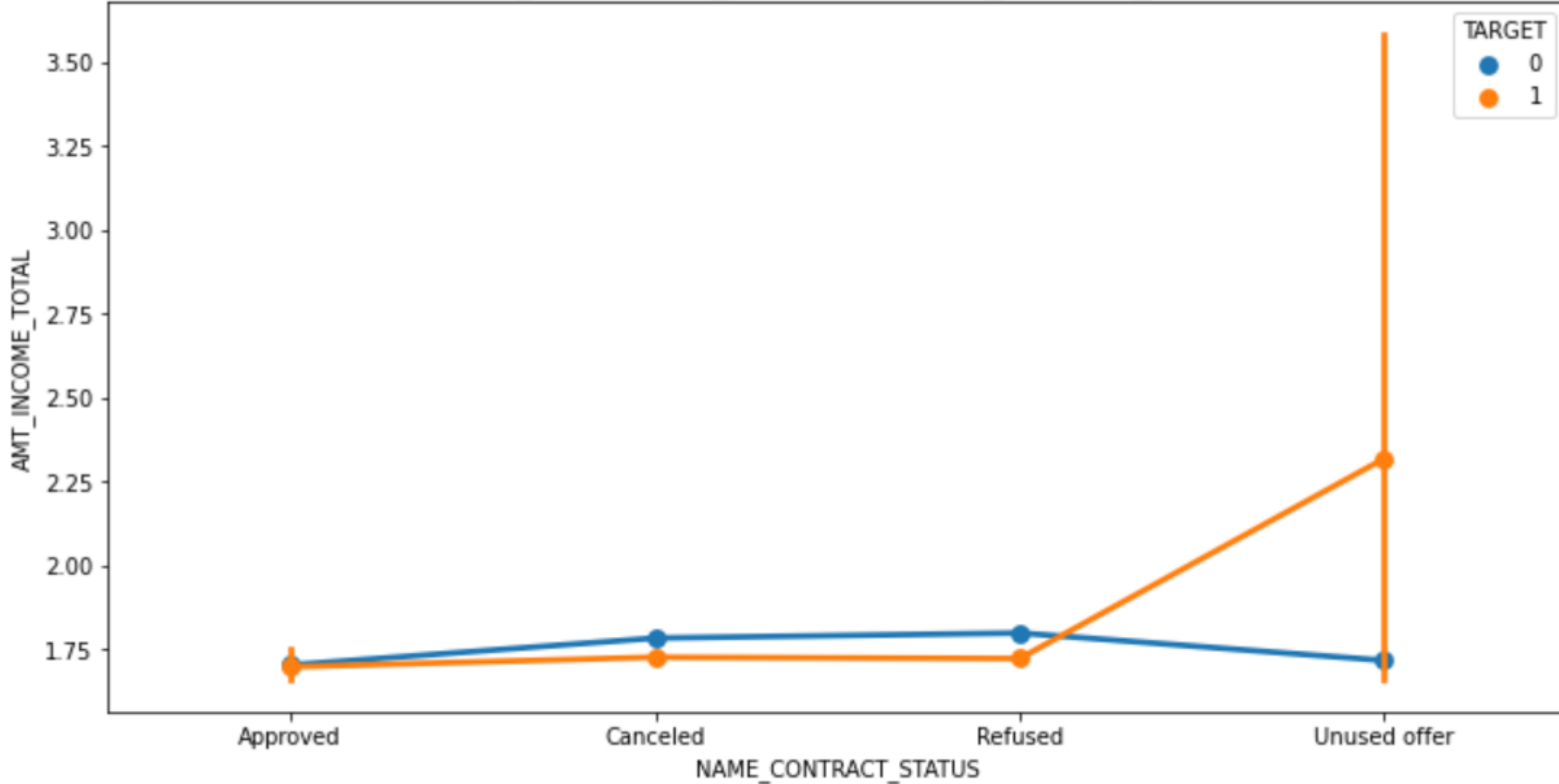
*Contract Status based on loan repayment status whether there is any business loss or financial loss*



- 90% of the previously cancelled client have actually repayed the loan. Revising the interest rates would increase business opportunity for these clients**
- 88% of the clients who have been previously refused a loan has payed back the loan in current case.**
- Refusal reason should be recorded for further analysis as these clients could turn into potential repaying customer.**

## *plotting the relationship between income total and contact status*

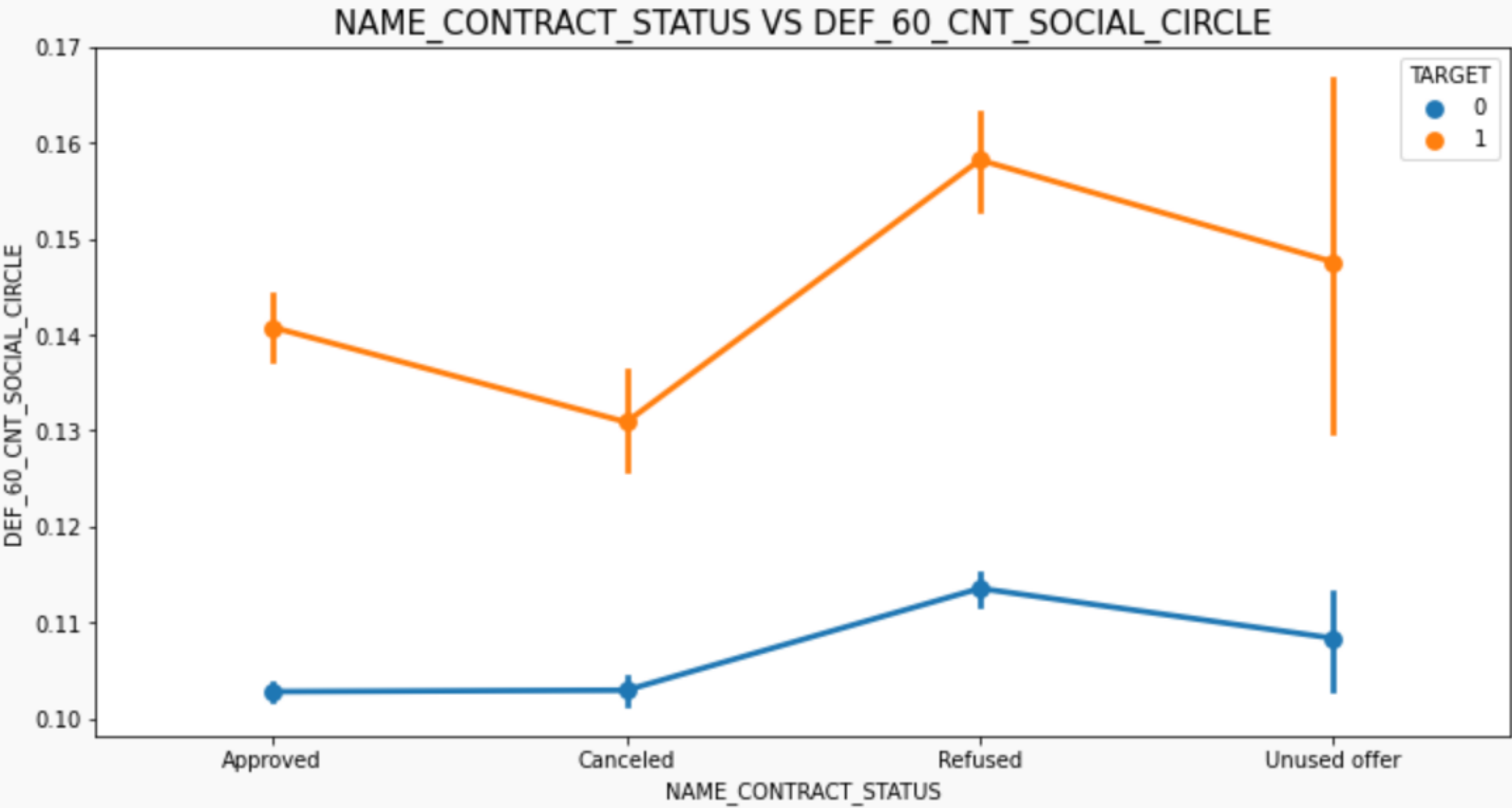
NAME\_CONTRACT\_STATUS VS AMT\_INCOME\_TOTAL



- The point plot show that the people who have not used offer earlier have defaulted even when there average income is higher than others



plotting the relationship between people who defaulted in last 60 days being in client's social circle and contact status



- **Clients who have average of 0.13 or higher their DEF\_60\_CNT\_SOCIAL\_CIRCLE score tend to default more and thus analyzing client's social circle could help in disbursement of the loan.**

# CONCLUSION

- \*Banks should focus more on contract type 'Student' , 'pensioner' and 'Businessman' with housing 'type other than 'Co-op apartment' for successful payments.**
- \*Banks should focus less on income type 'Working' as they are having most number of unsuccessful payments.**
- \*Also with loan purpose 'Repair' is having higher number of unsuccessful payments on time.**
- \*Get as much as clients from housing type 'With parents' as they are having least number of unsuccessful payments.**

End