



CREDIT EDA CASE STUDY

By Anjalika Roychowdhury

Problem Statement

- To understand how the bank approves and refuses loan.
- Find out different patterns and represent the outcomes to help the bank reduce the credit risk and interest risk.

Objective

- Credit risk will help the company to make a decision on loan approval based on the applicant's profile.
- This ultimately controls loss of the company and avoids the financial loss of the company

Steps Followed

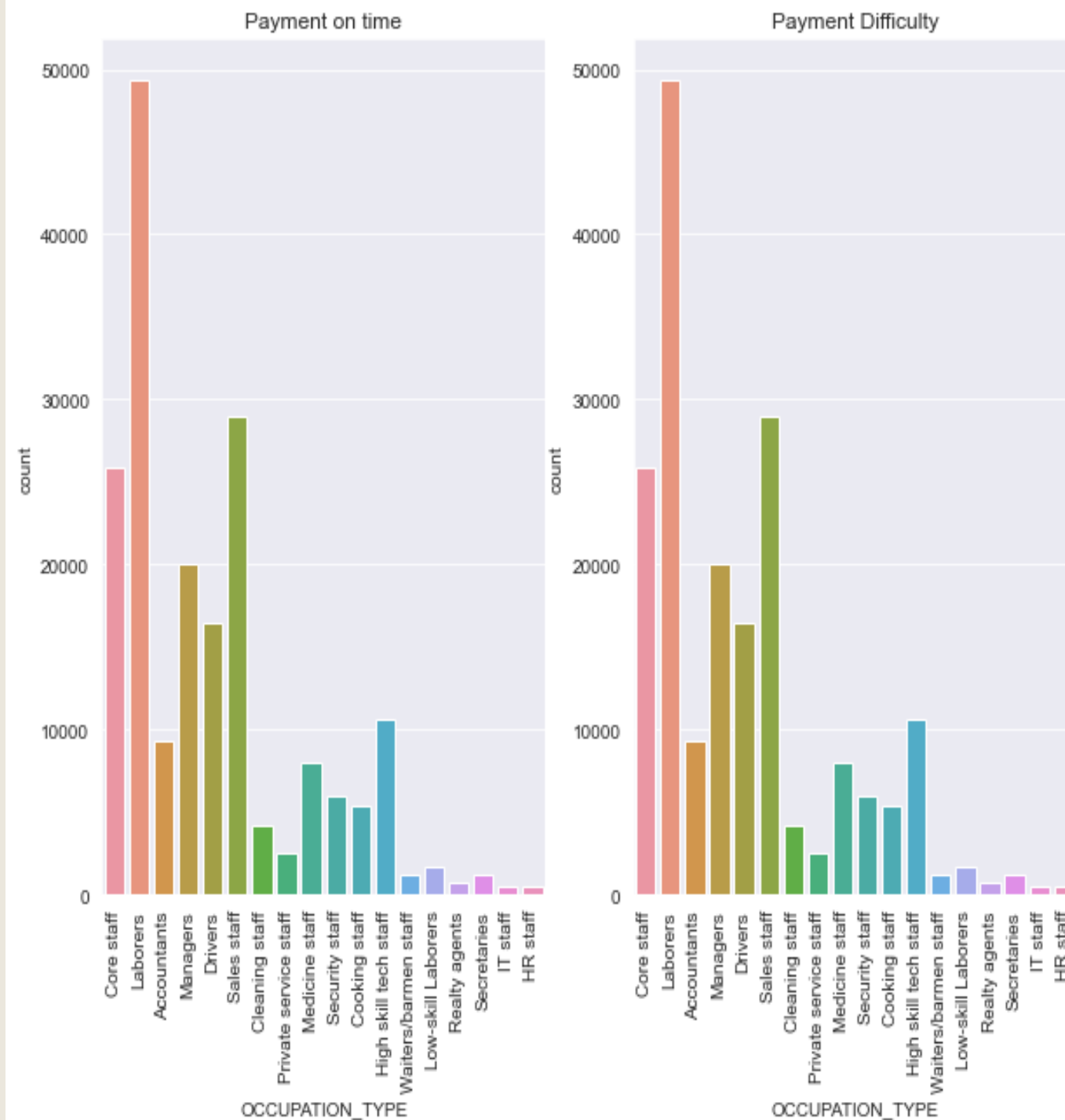
- Data understanding and sourcing.
- Data checking, cleaning and Binning
- Checking the data imbalance.
- Univariate and Bivariate analysis
- Merging the two data sets
- Univariate and Bivariate analysis on merged dataset
- conclusion

Univariate Analysis

Distribution of OCCUPATION_TYPE

Based on the occupation type defaulters and non defaulters has been showed.

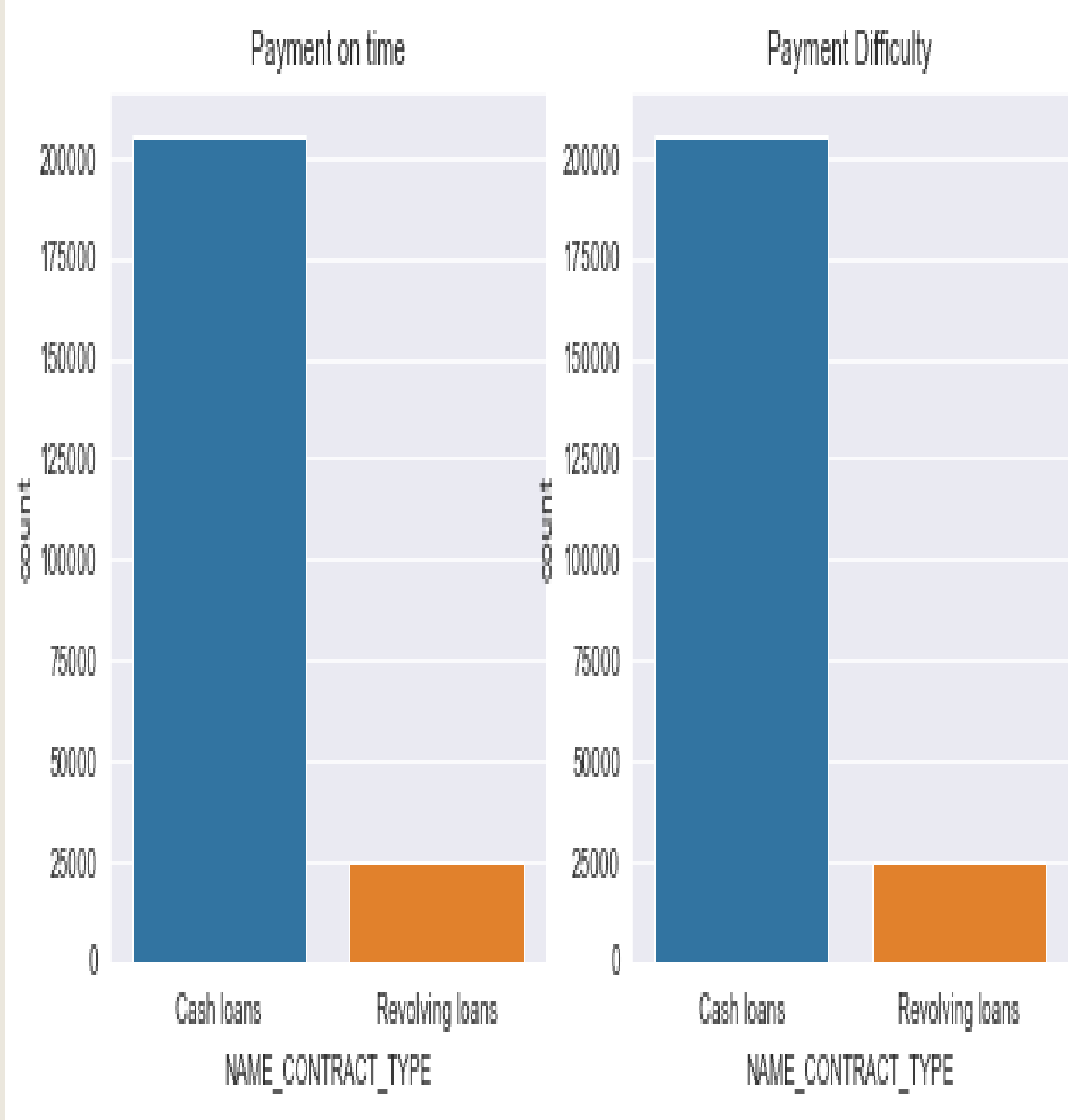
From the distribution it has been observed that laborers are on the higher side on both defaulters and non defaulters



Distribution of Contract Type

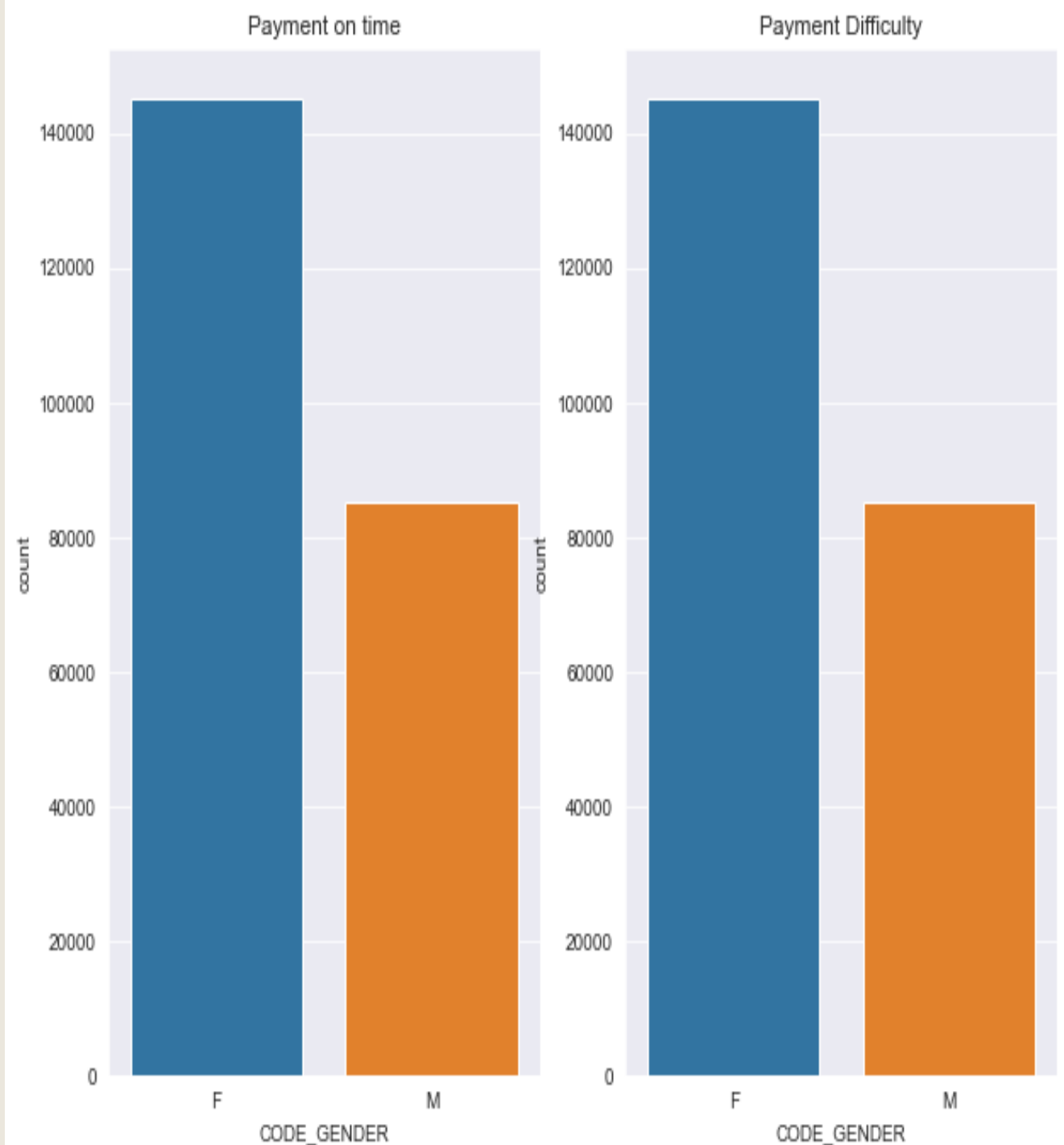
On the basis of contract type it shows that people tend to take cash loan in place of any other loan.

majority of the customer have taken cash loan irrespective of defaulter or non defaulter



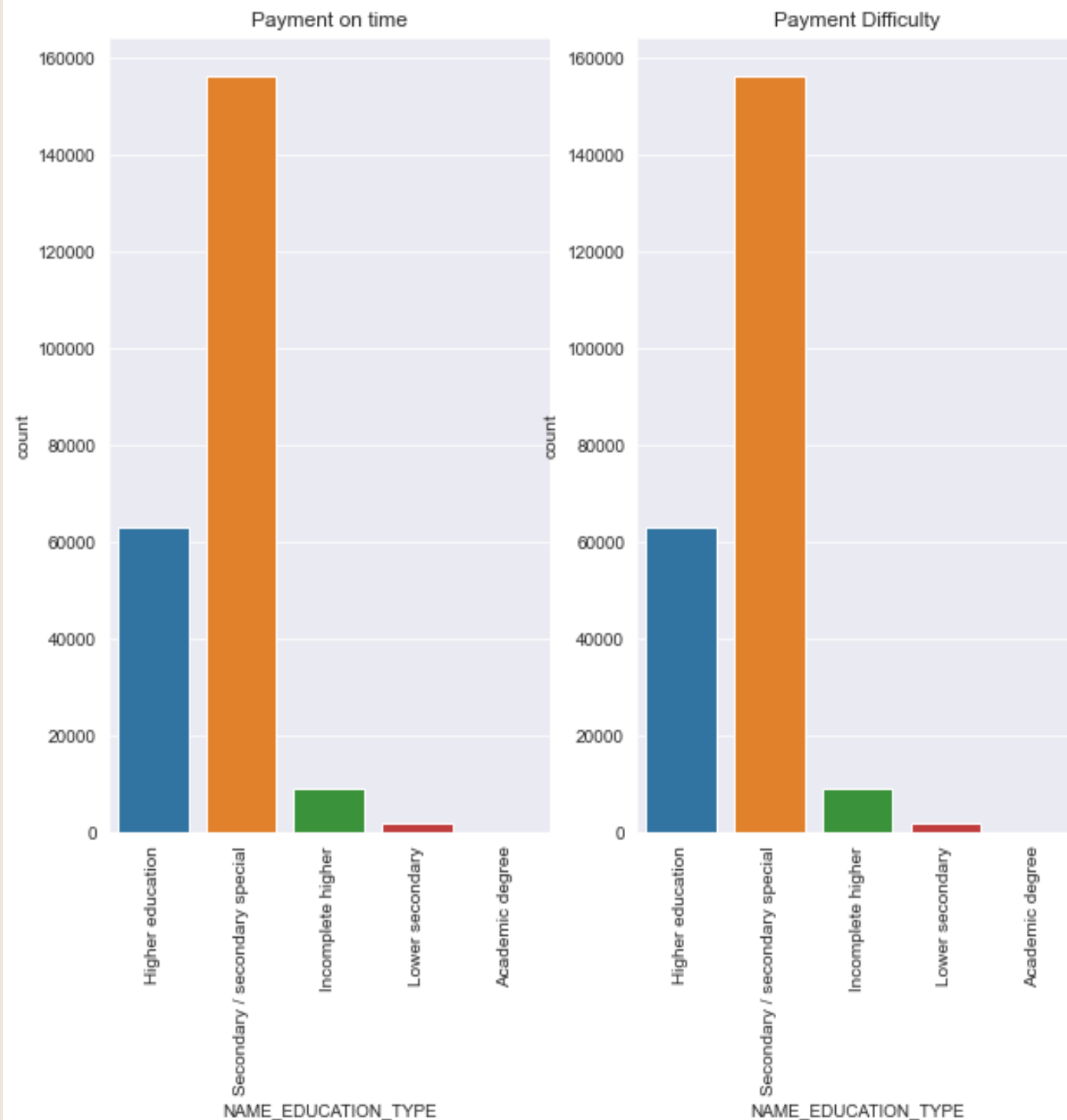
Analysis based on - Gender

Female have high ratio for both defaulter and non defaulter



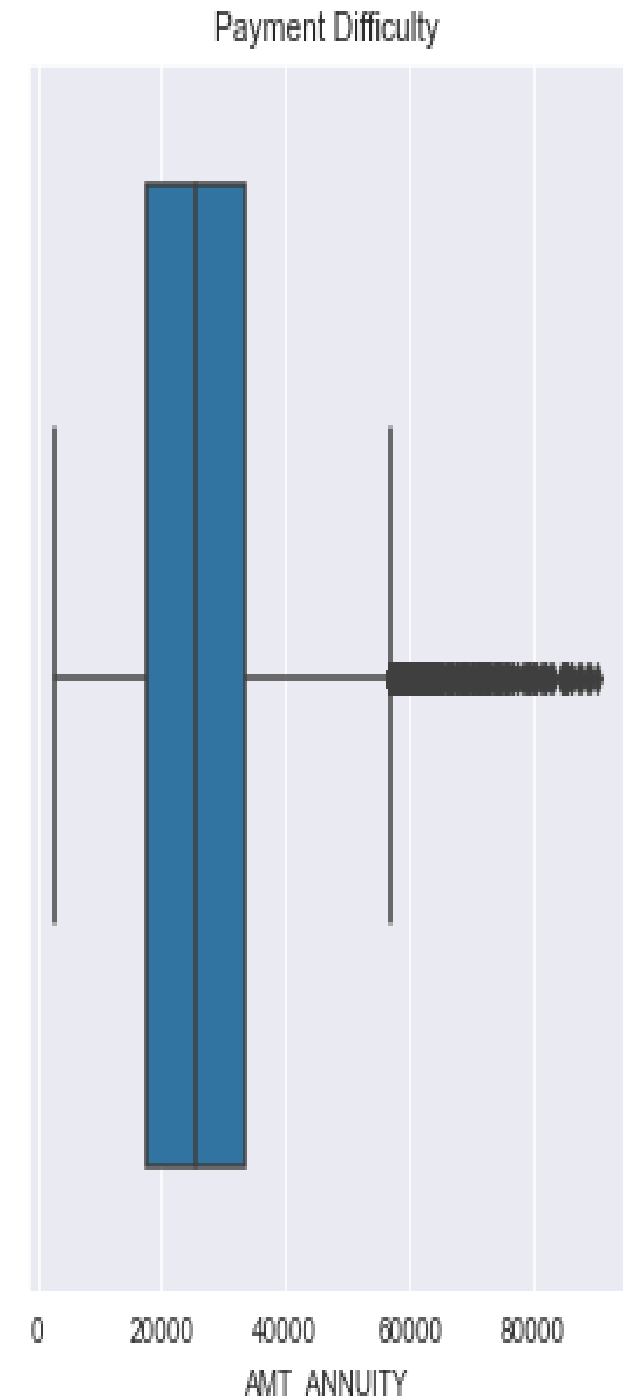
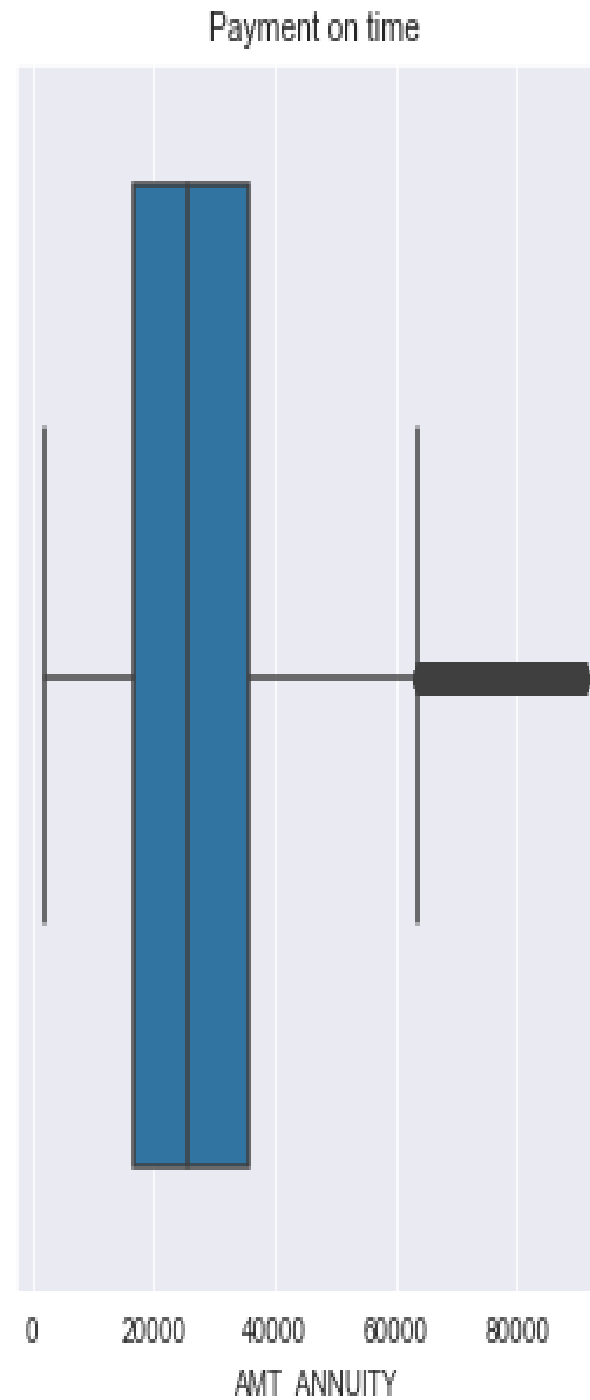
Distribution – Education Types

Highest loan proportion with "Secondary/secondary special" in both the cases.. lower is academic degree



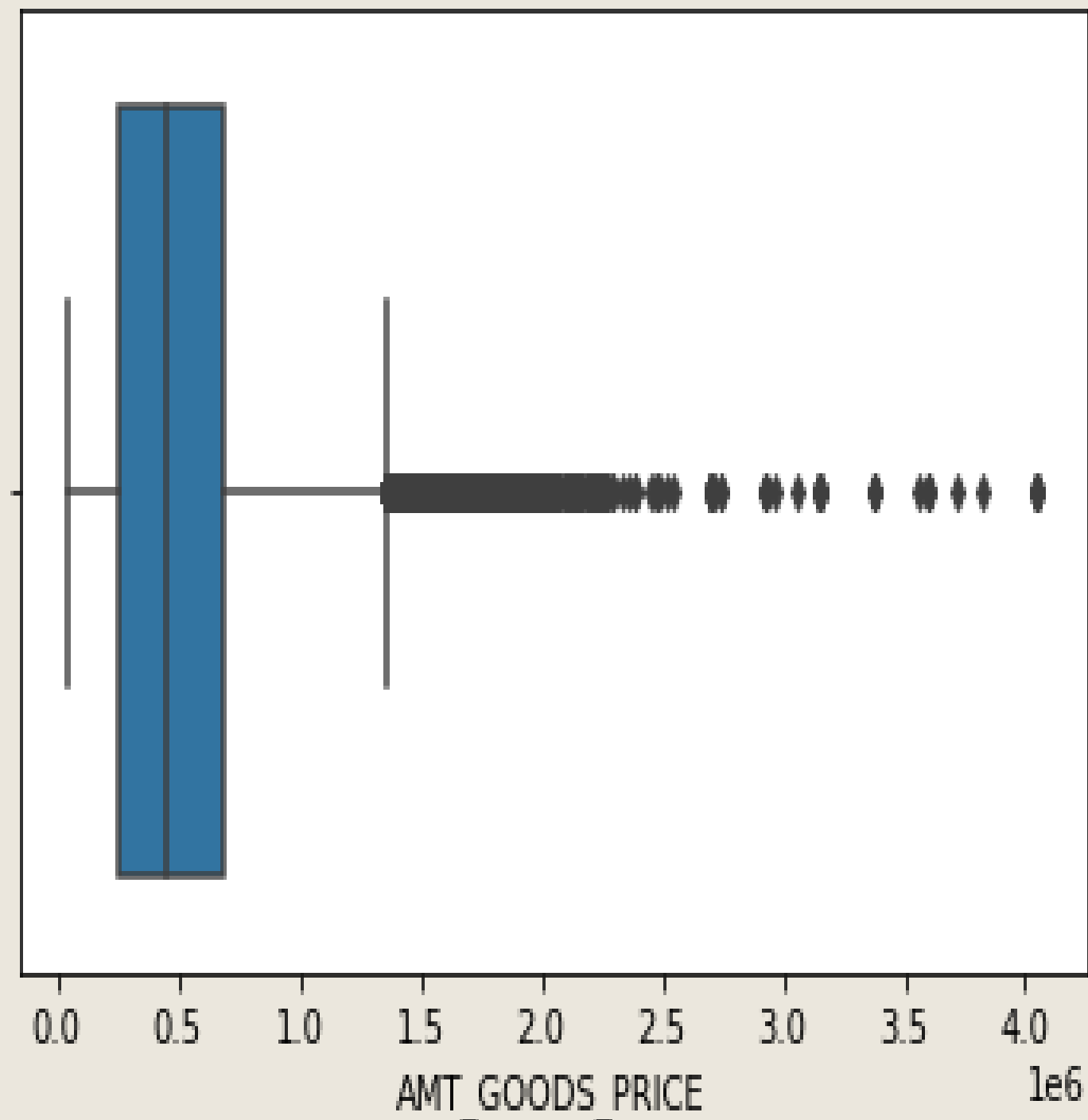
Boxplot for Amt_annuity

This shows that for non defaulters loan of annuity amount lies between 15000-35000 whereas for defaulters lies in the range of 17000-32000



Boxplot for Amt_Goods_Price

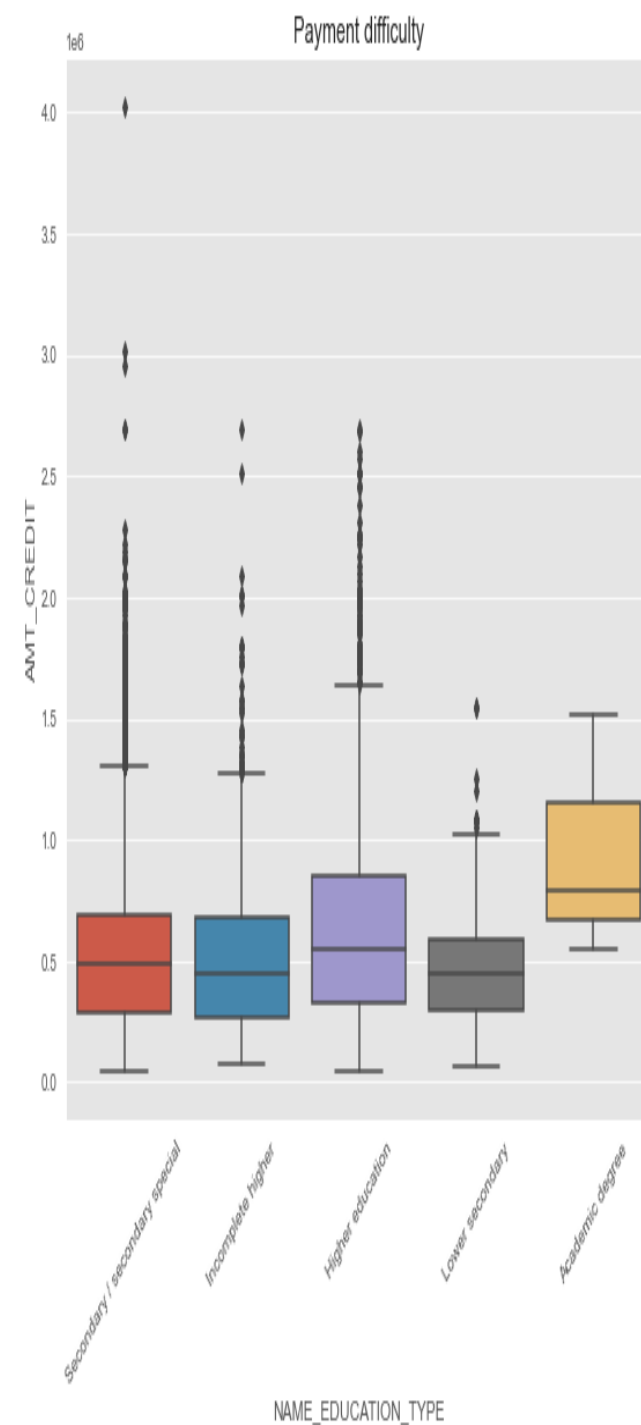
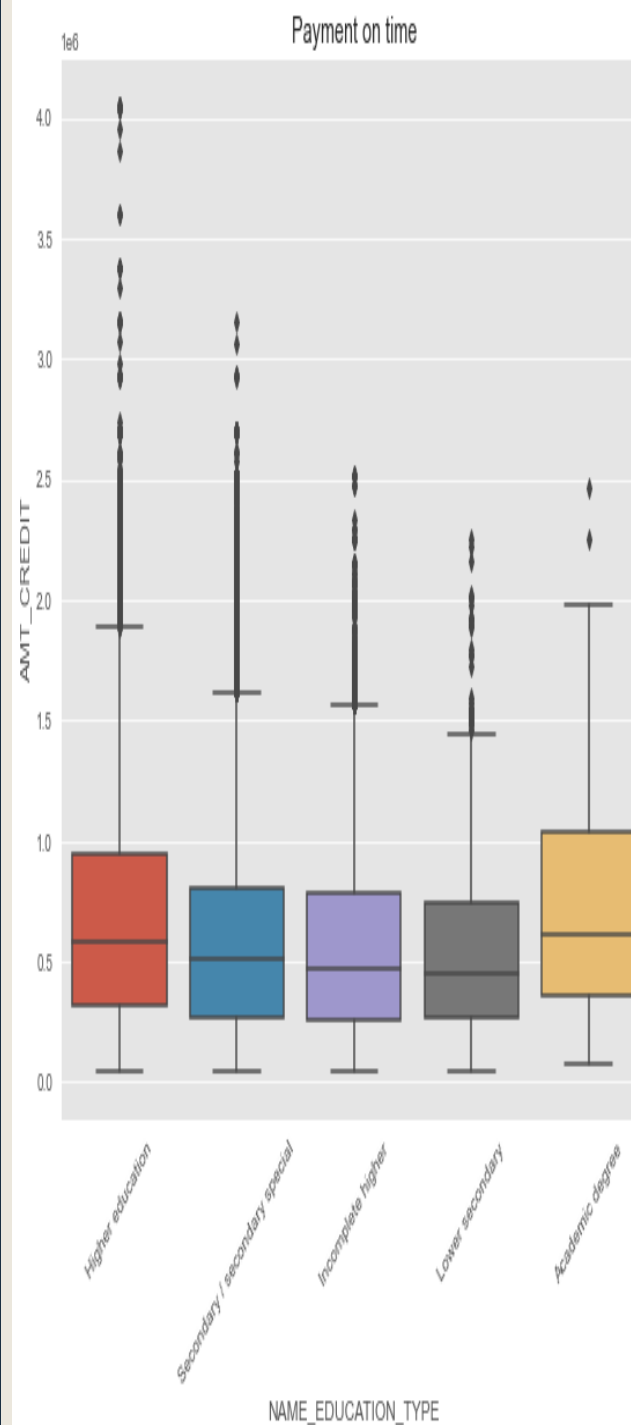
Amt_Goods_Price is a continuous variable and has a outlier . It has been imputed using median of the column



BIVARIATE ANALYSIS

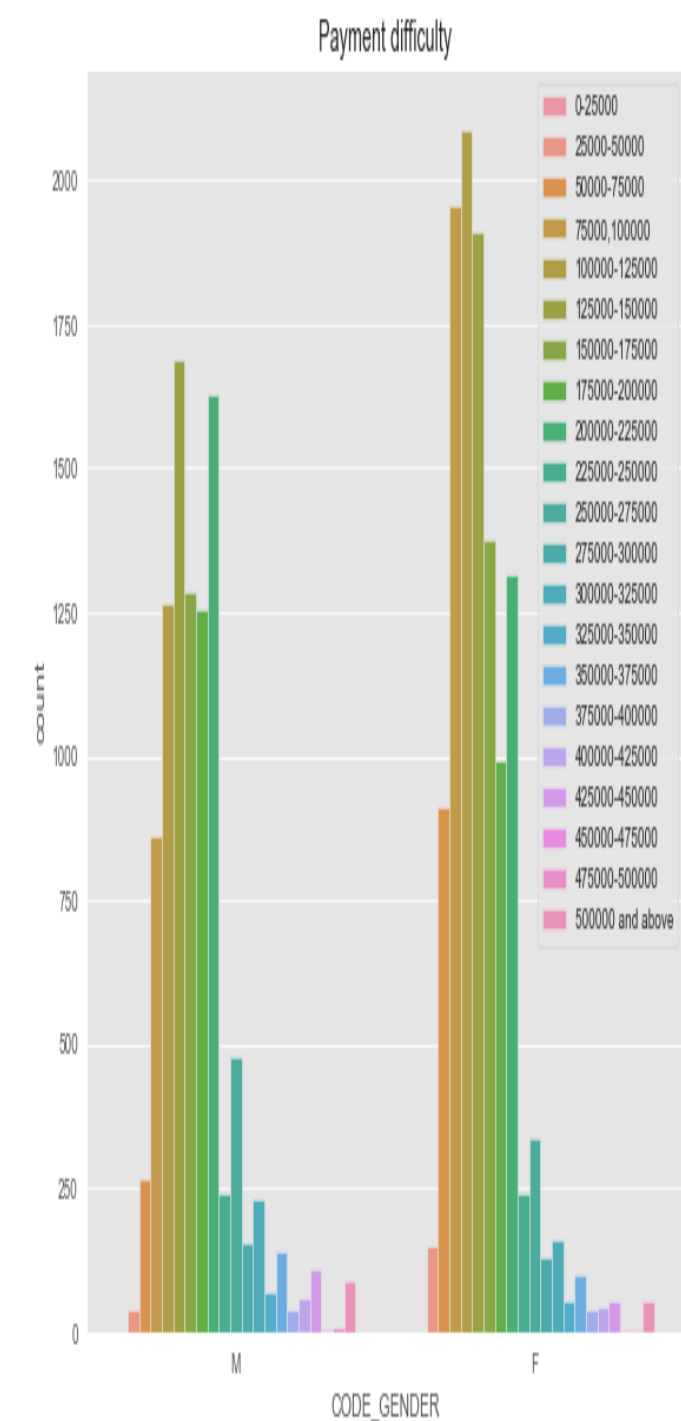
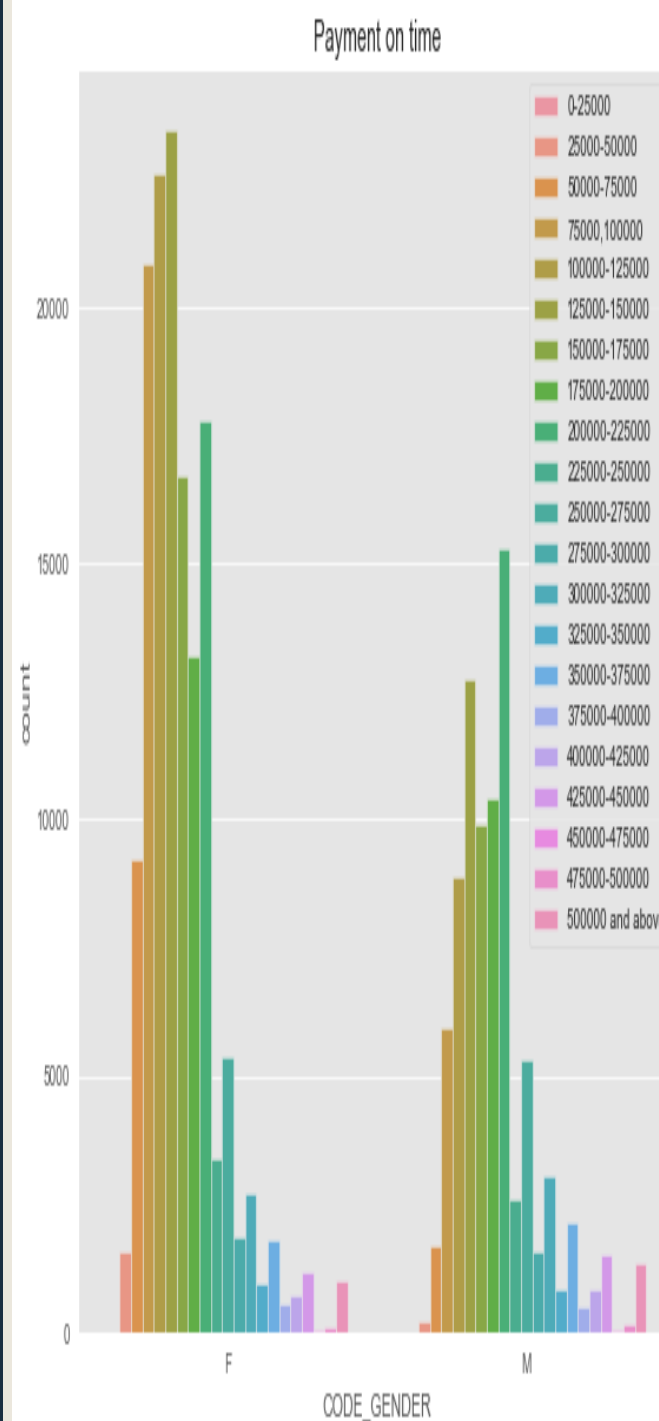
NAME_EDUCATION_TYPE and AMT_CREDIT

On the basis of education and type of loan taken for education..



Analysis on the basis of Gender and Amt_Income

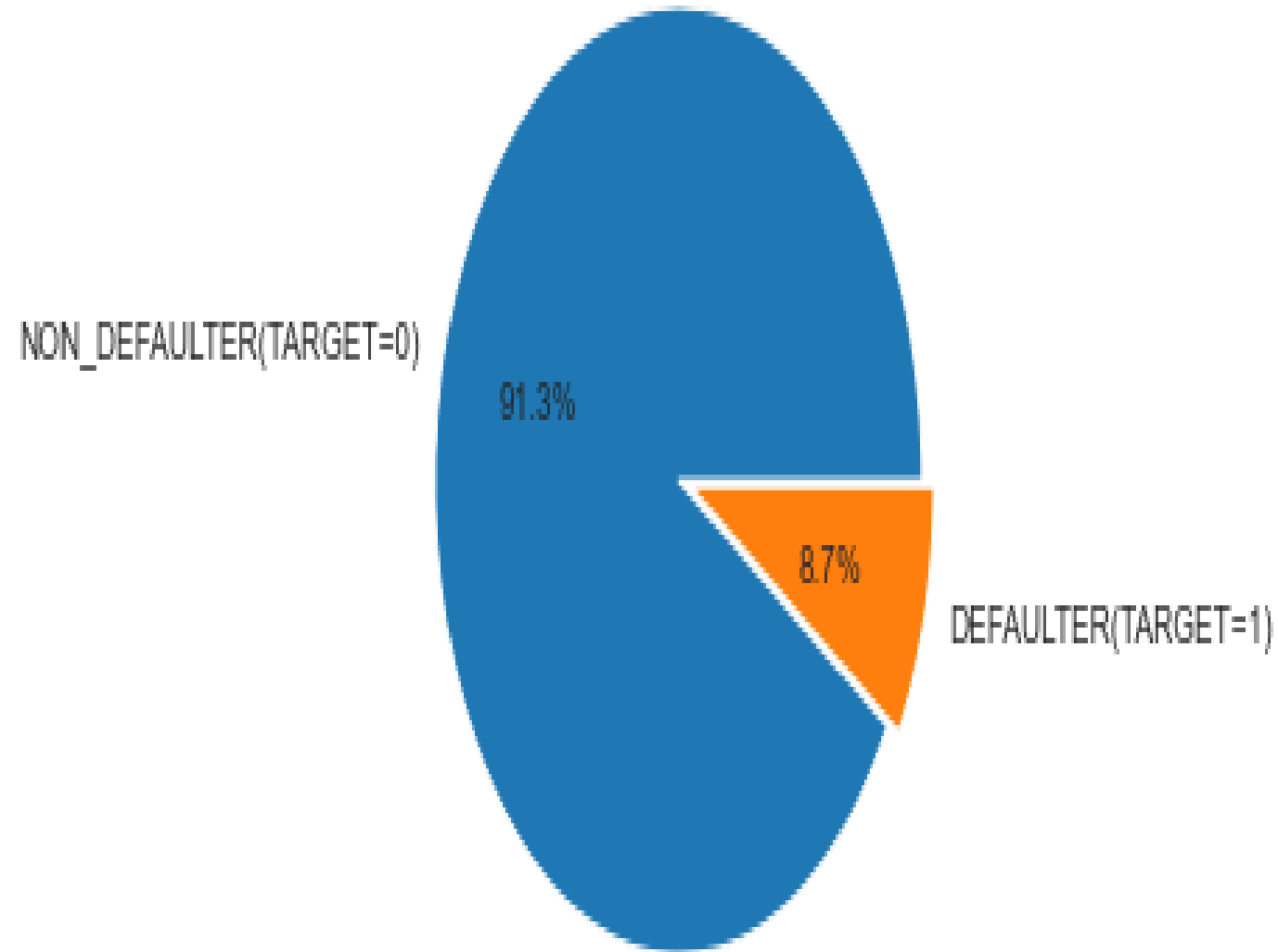
It has been observed that females has high proportion in both cases of defaulter and non defaulter



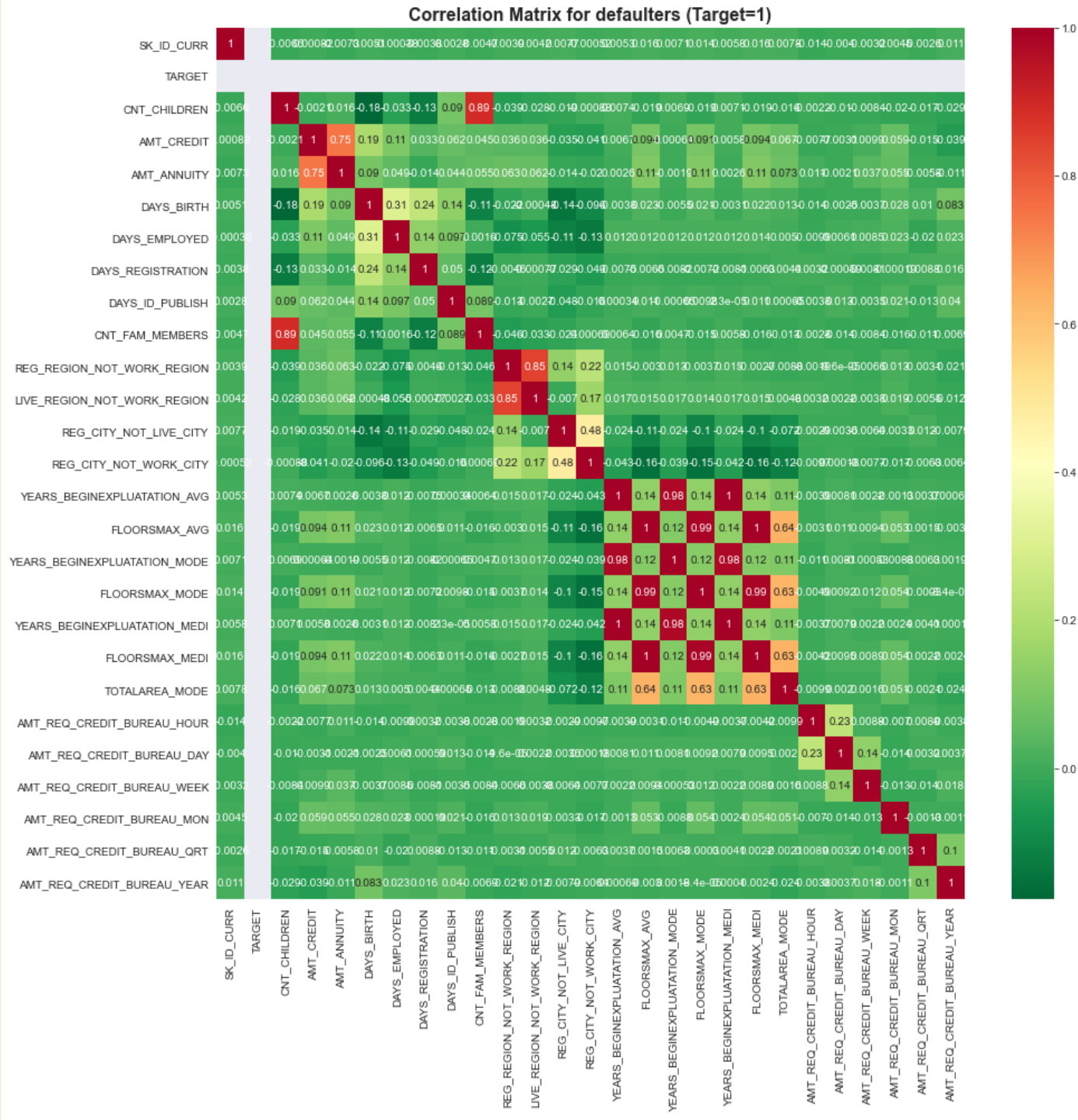
TARGET Varibale- DEFAULTER Vs NONDEFAULTER

Defaulters Vs Non defaulters

As we can observe the number of Non- defaulter clients (Target= 0) is more than defaulter clients (Target=1). it is almost 8.65 % of Non-defaulter clients.



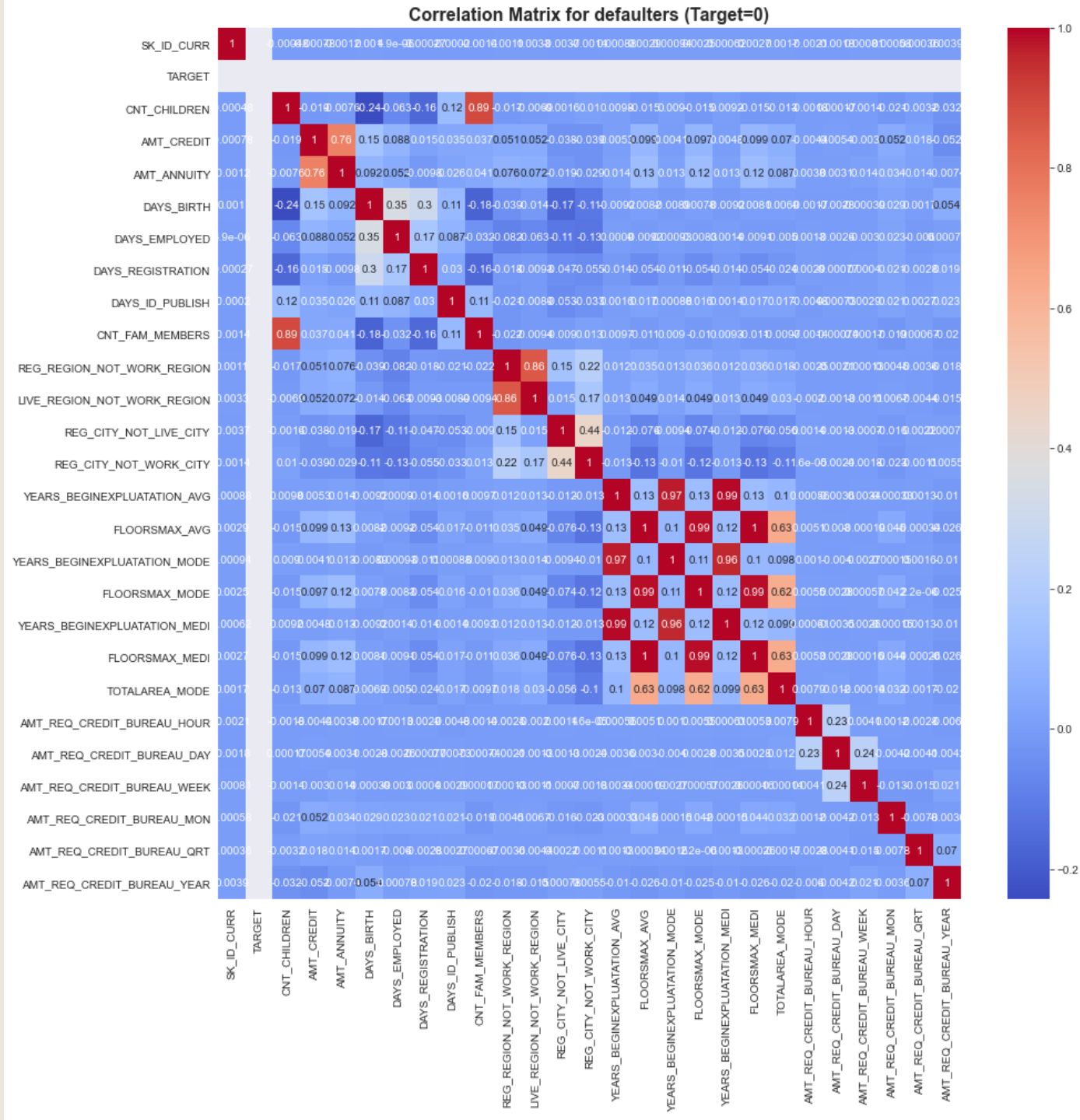
Since target column has null value so it's showing blank rows and columns



Correlation Matrix for defaulters

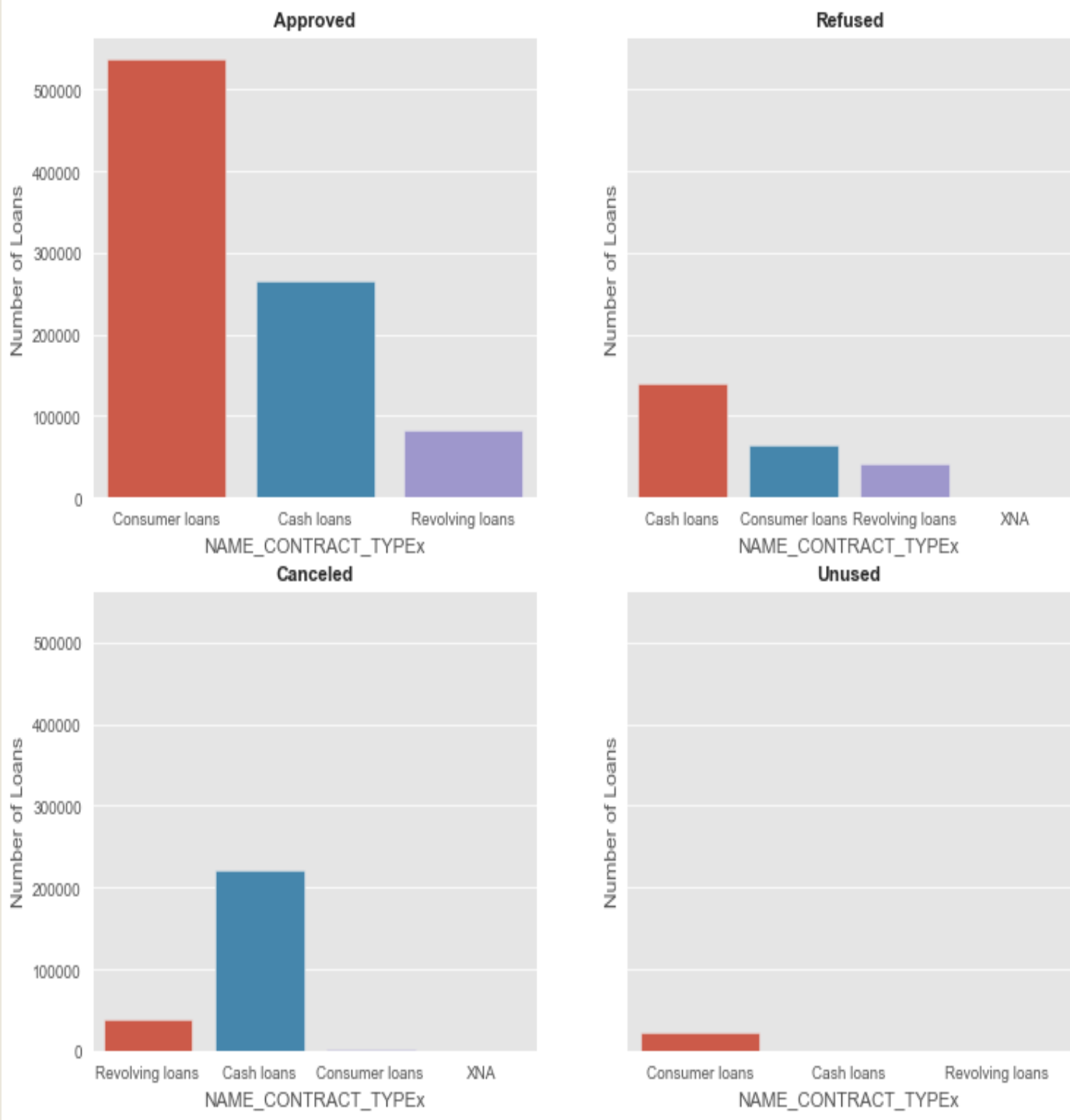
- Since the majority is target 0 than target 1, hence calculating imbalance percentage

- The imbalance ratio is 10.55



Bivariate Analysis on the basis of contract type

Cash loan is mostly taken by people



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.

THANK YOU