

EDA CASE STUDY

SUBMITTED BY:

YOGESH KRISHNA (yogeshkrishna59@gmail.com)

JATIN JAIN (jatinj07@gmail.com)

PROBLEM STATEMENT

- There are two types of risks associated with any loan request :
 - If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
 - 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.
 - In the dataset:
 - TARGET 1 - Client with payment difficulties.
 - TARGET 0 - All other cases.
- Analysis of the data set has been done in Python on a Jupyter Notebook.

ANALYSIS DONE

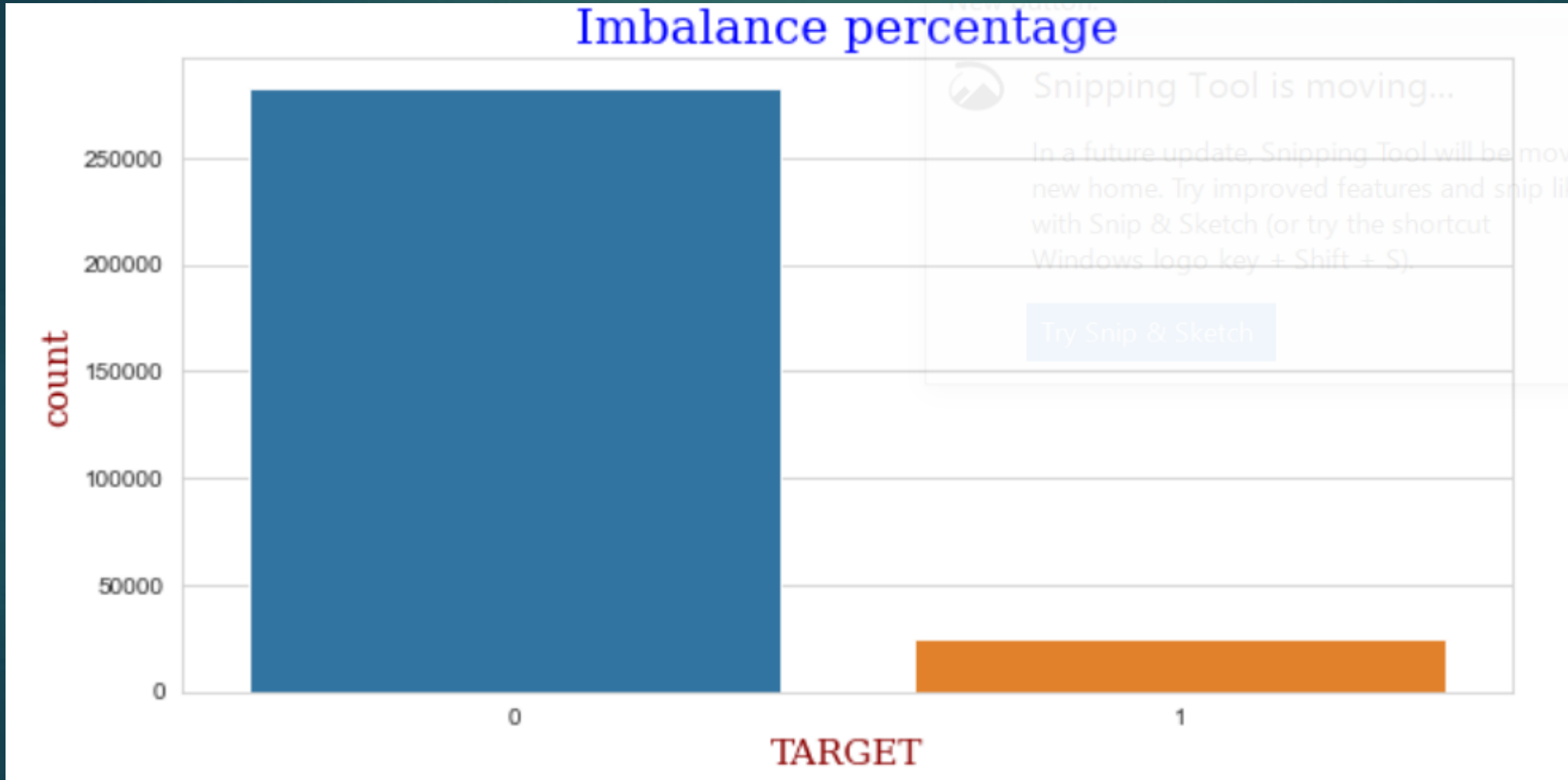
3

- Steps :
 - Check Missing Values
 - which to handle
 - how to handle
 - Check Outliers
 - Check data imbalance; ratio of imbalance
 - Check data types
 - Top correlations for the Client with payment difficulties other variables within Application DF & Previous App DF

APPLICATION_DATA ANALYSIS

Imbalance percentage of TARGET variable

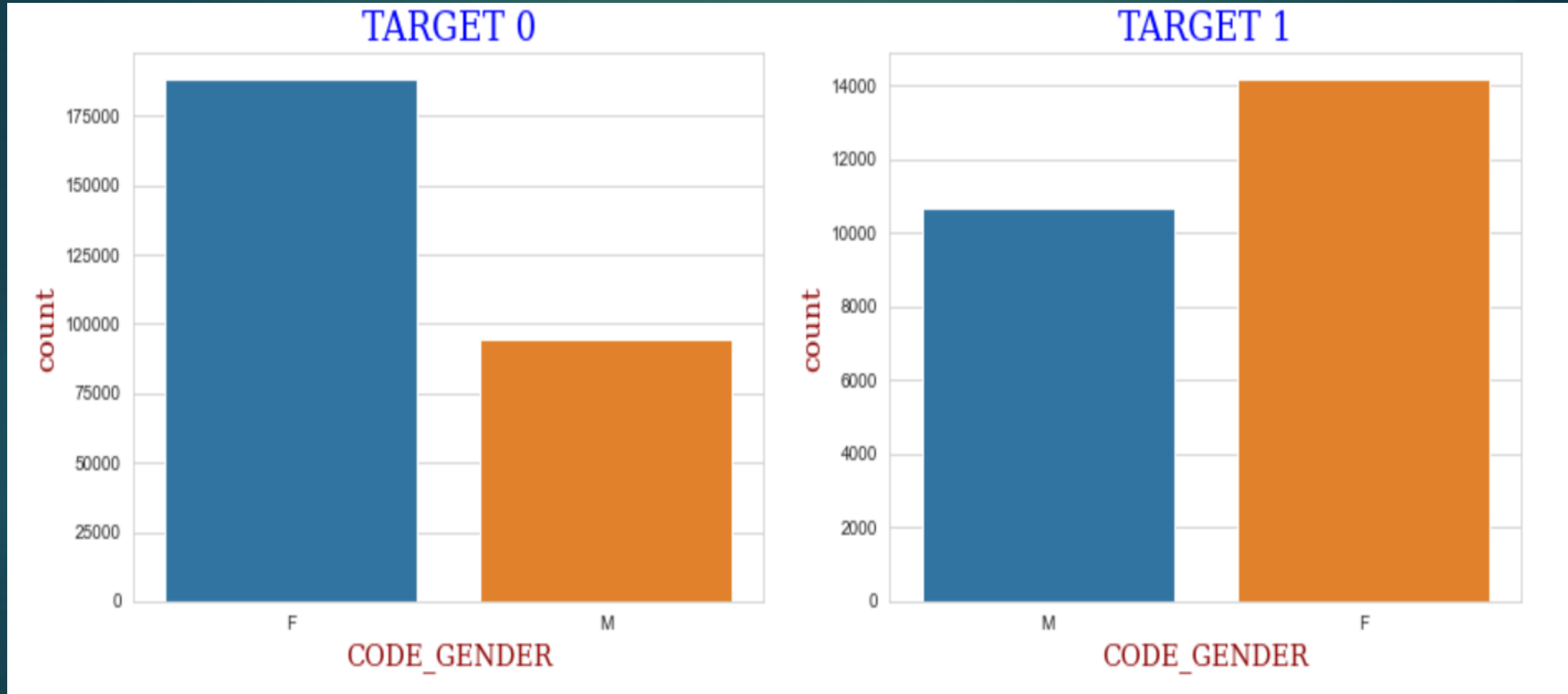
5



TARGET 0 (pay on time) is higher than TARGET 1(Defaulter).

Proportion of Gender in TARGET variable

6

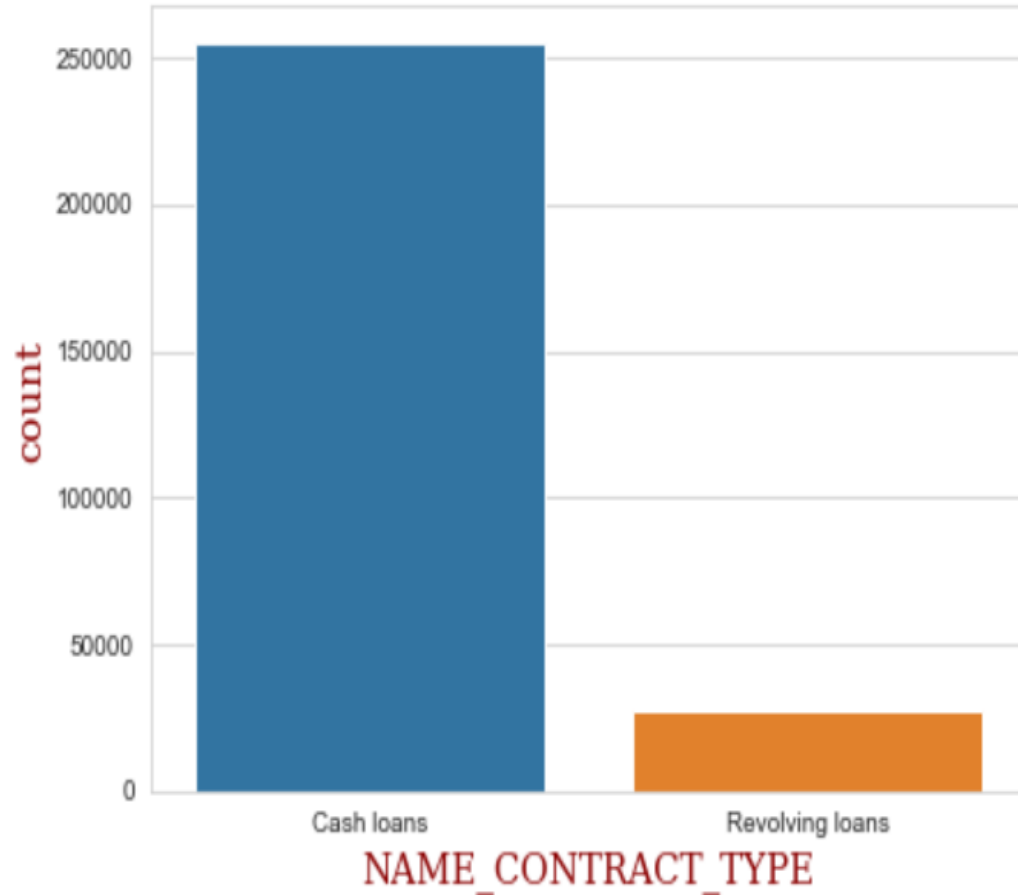


Female defaulters are higher.

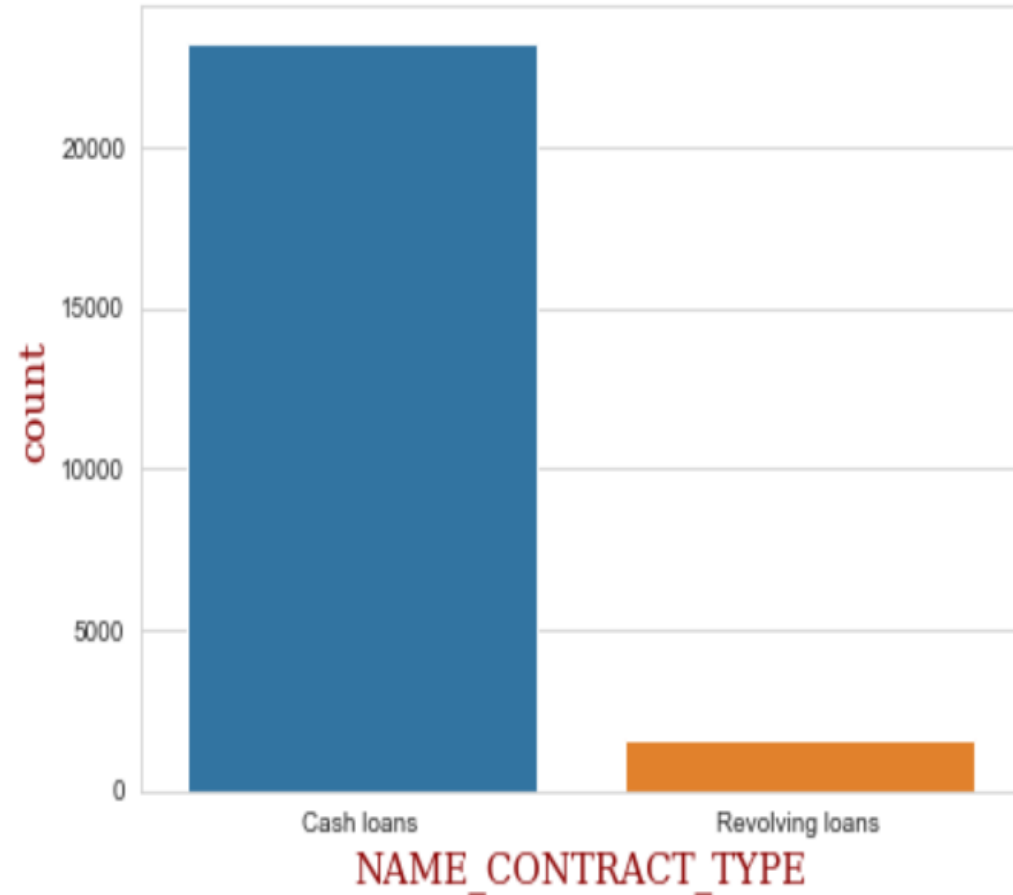
Proportion of Contract type in TARGET variable

7

TARGET 0



TARGET 1

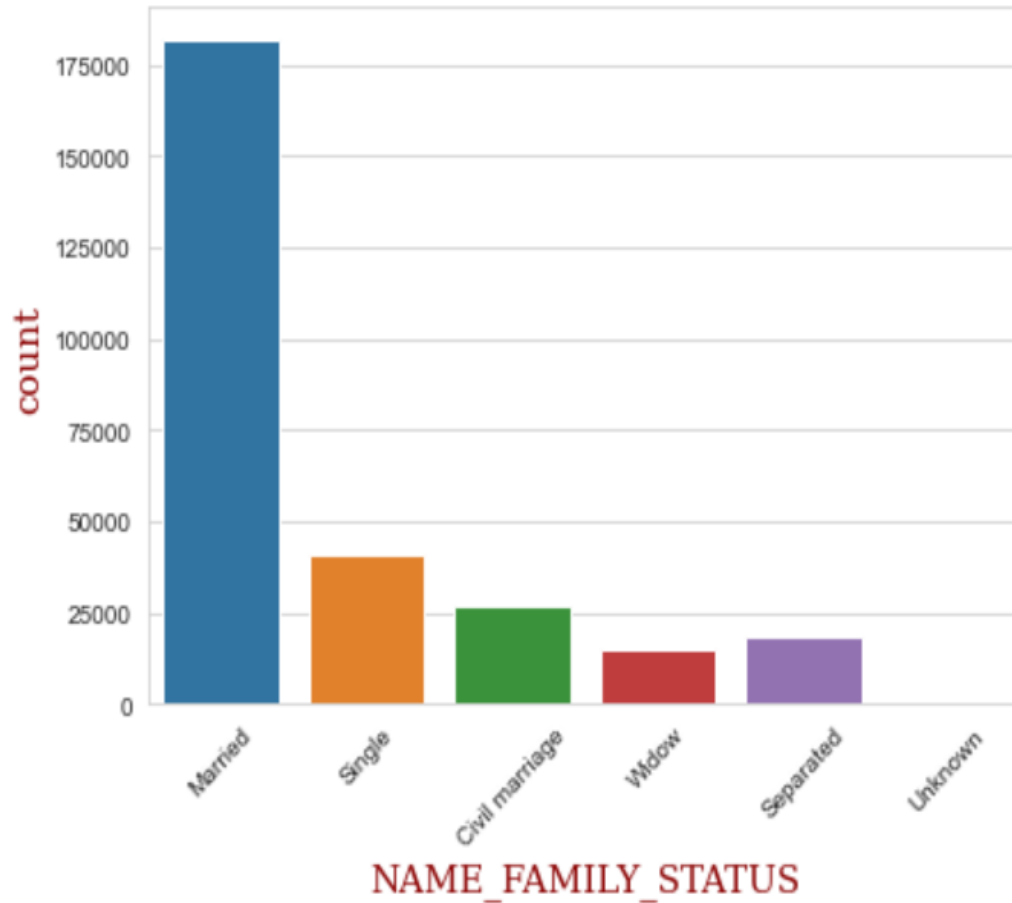


Cash Loans are defaulted more.

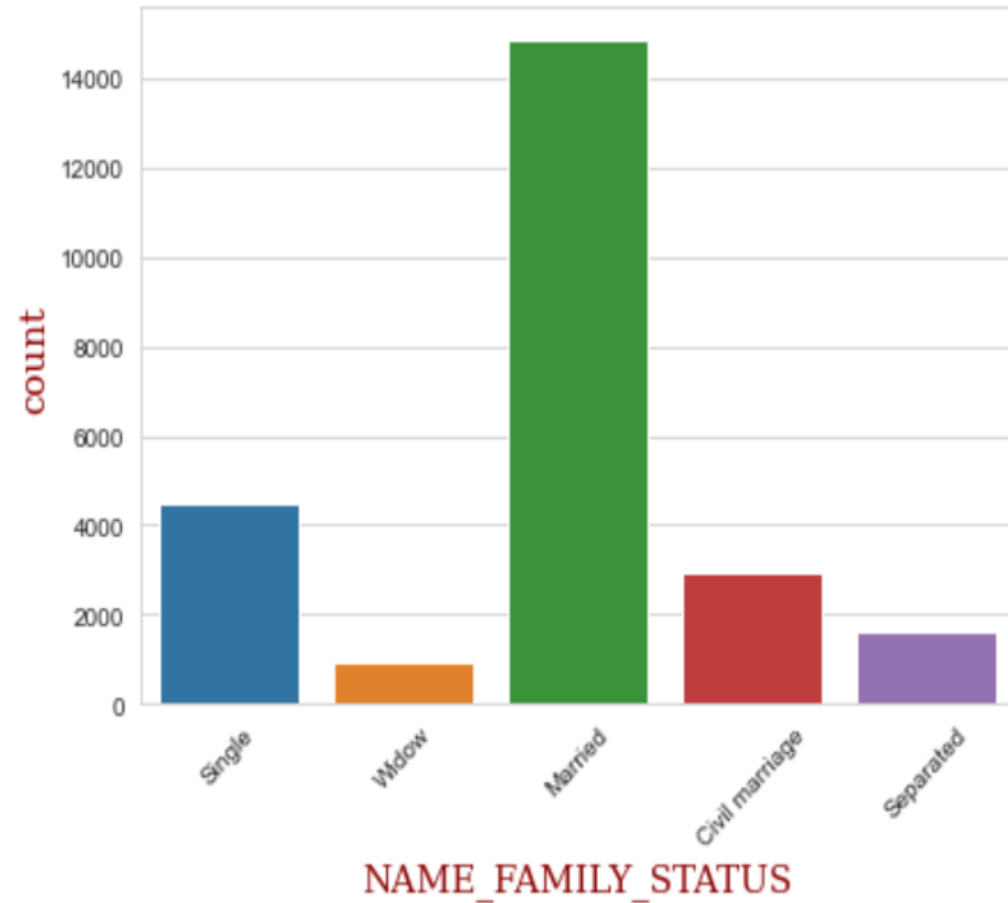
Proportion of family status in TARGET variable

8

TARGET 0



TARGET 1

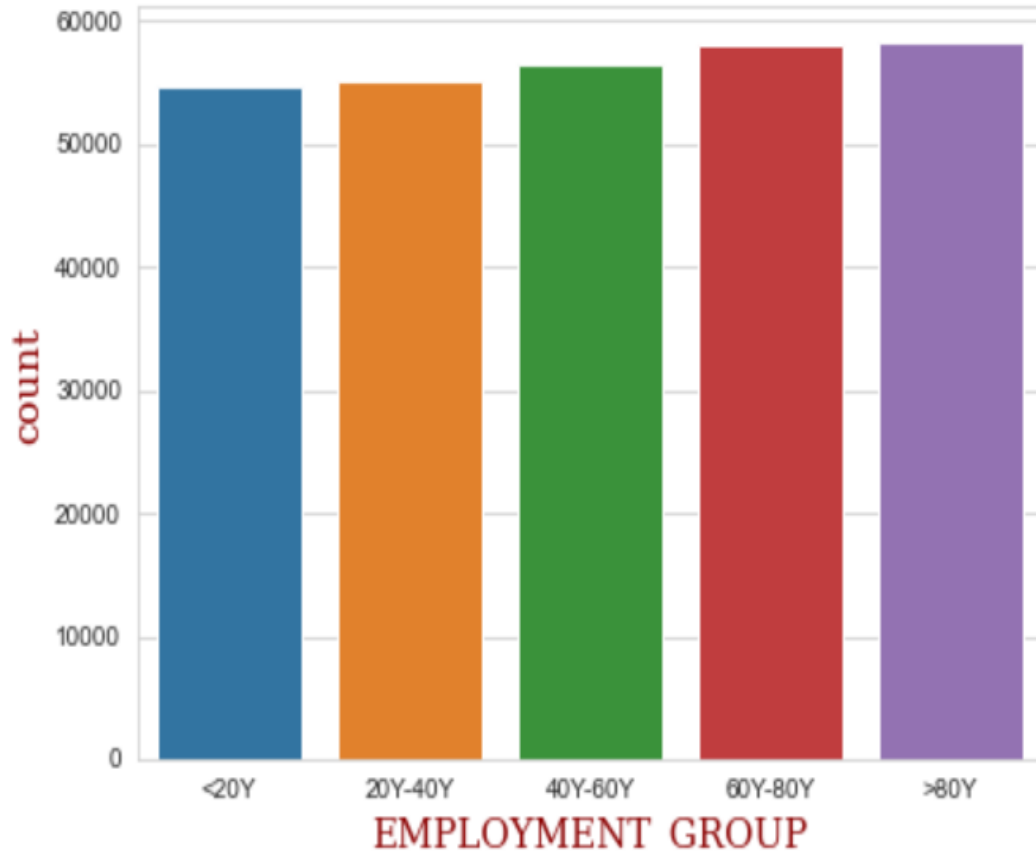


Married applicants are more likely to default.

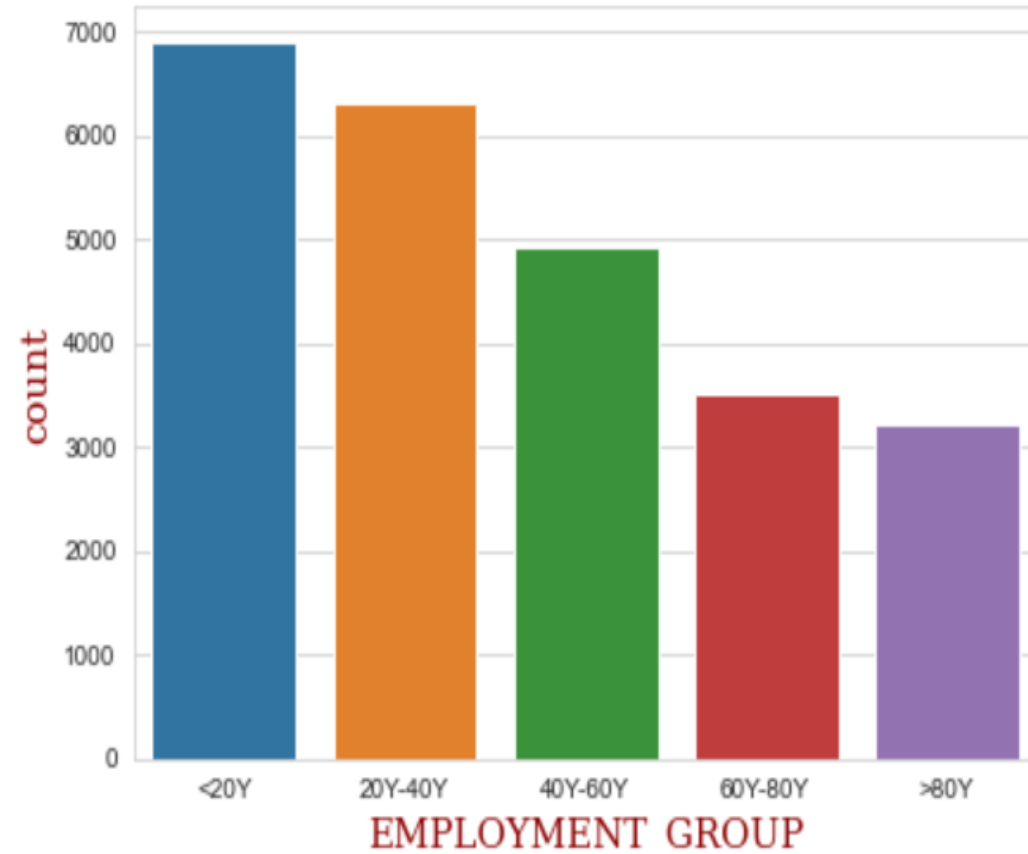
Proportion of employment group in TARGET variable

9

TARGET 0



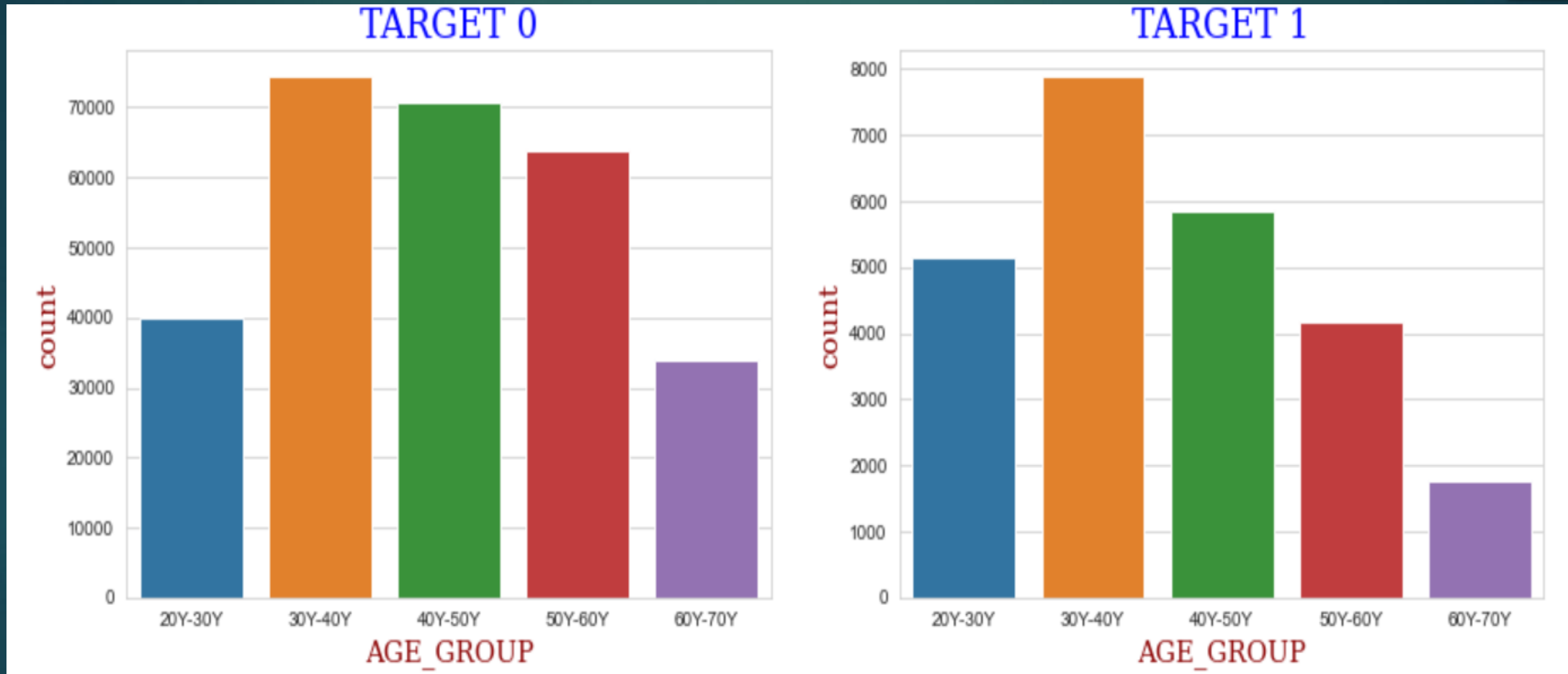
TARGET 1



Employees with less than 20 years of experience are more likely to default.

Proportion of age group in TARGET variable

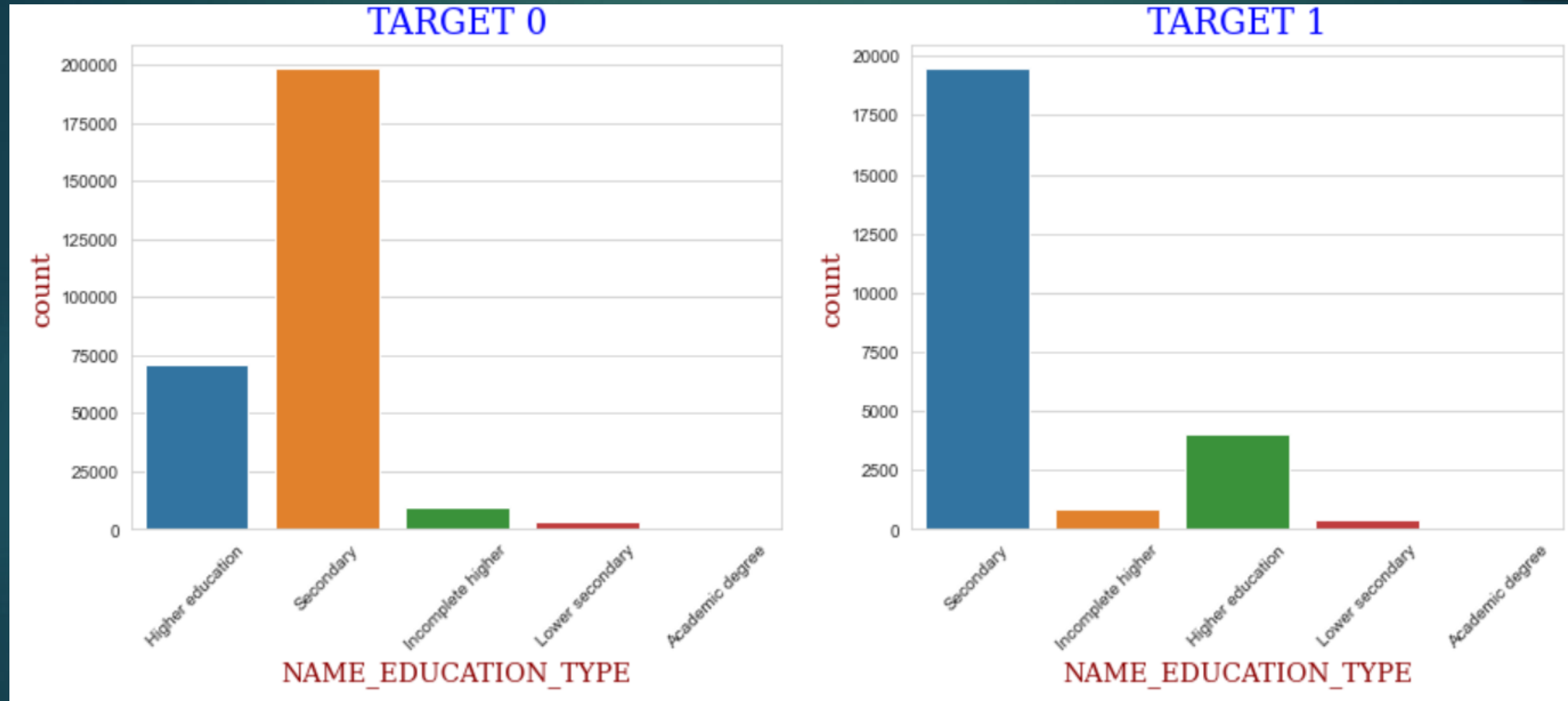
10



Employees in the Age group range of between 30 to 40 are more likely to default.

Proportion of education type in TARGET variable

11

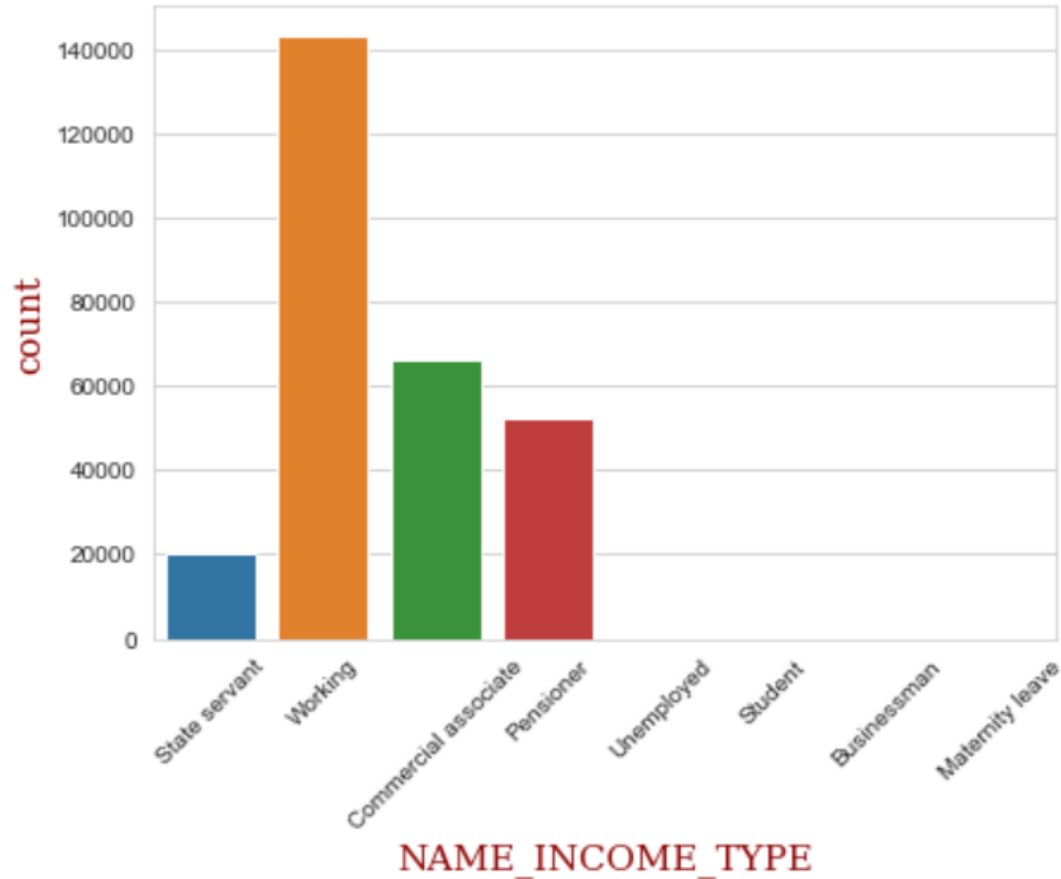


Secondary education type has higher default rate.

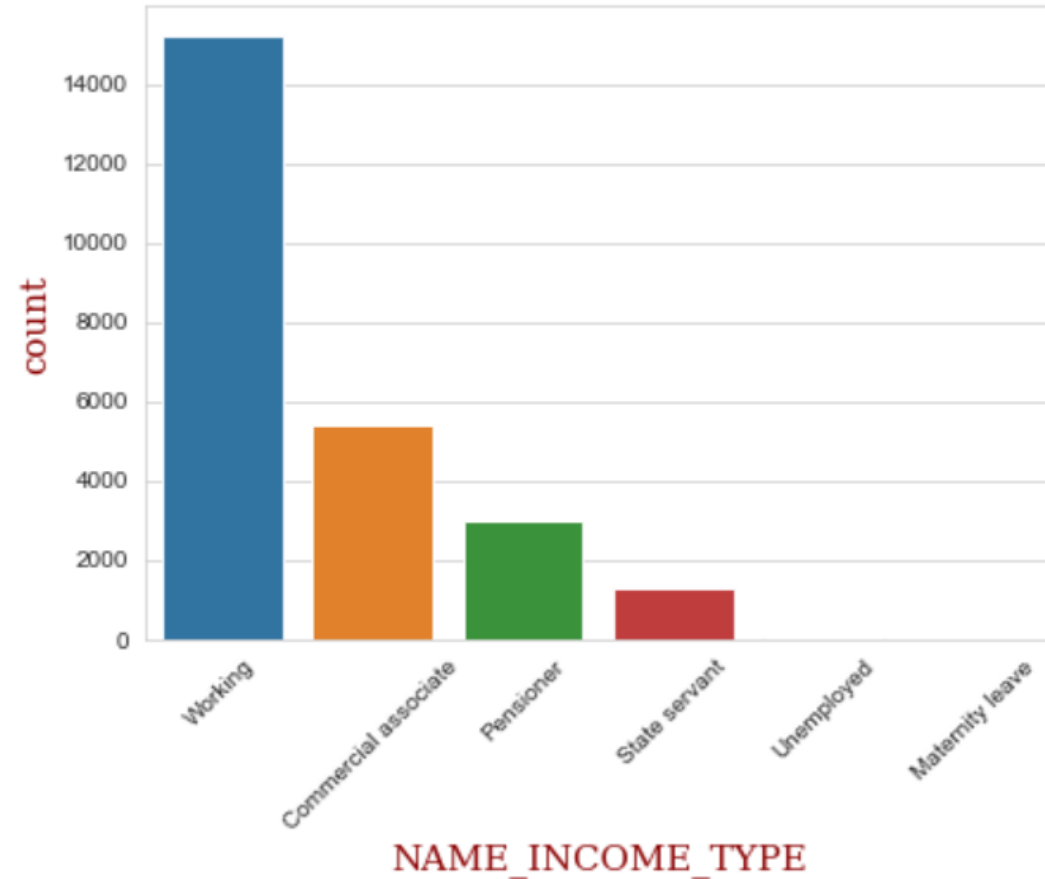
Proportion of income type in TARGET variable

12

TARGET 0



TARGET 1

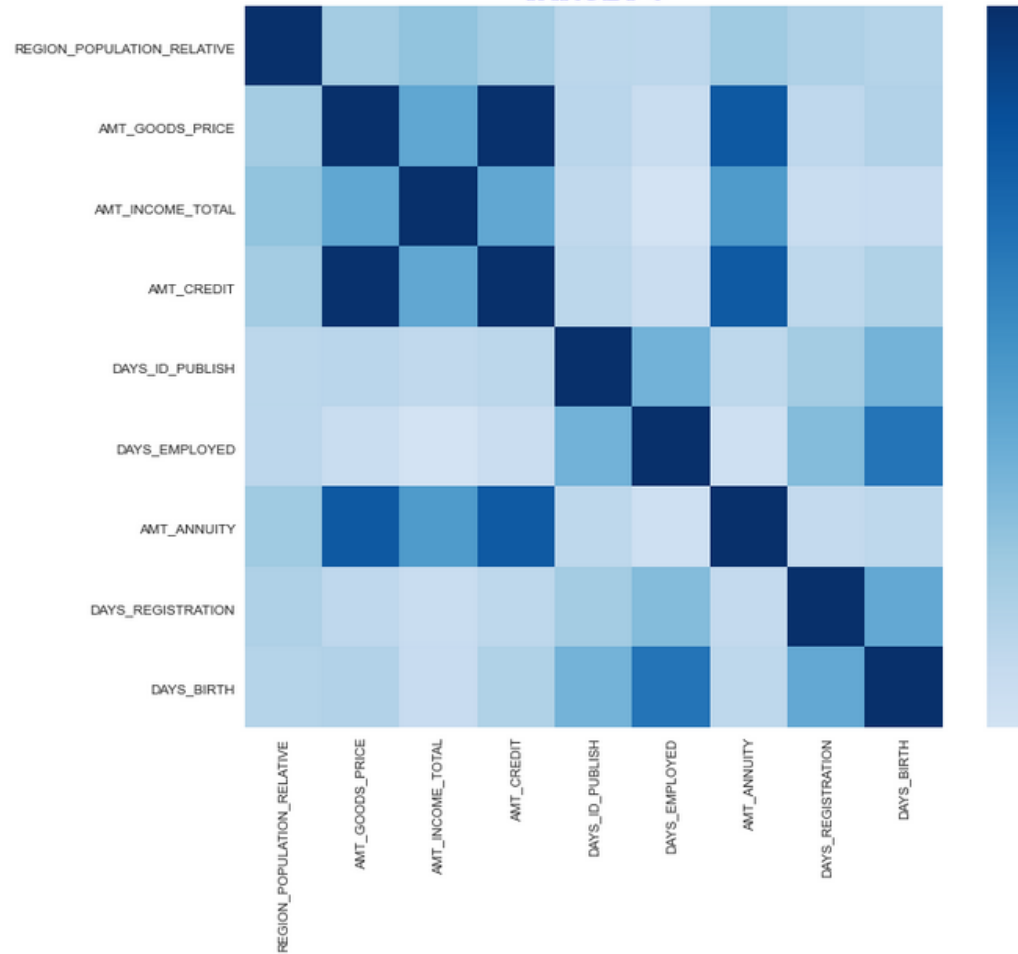


Working applicants are more likely to default.

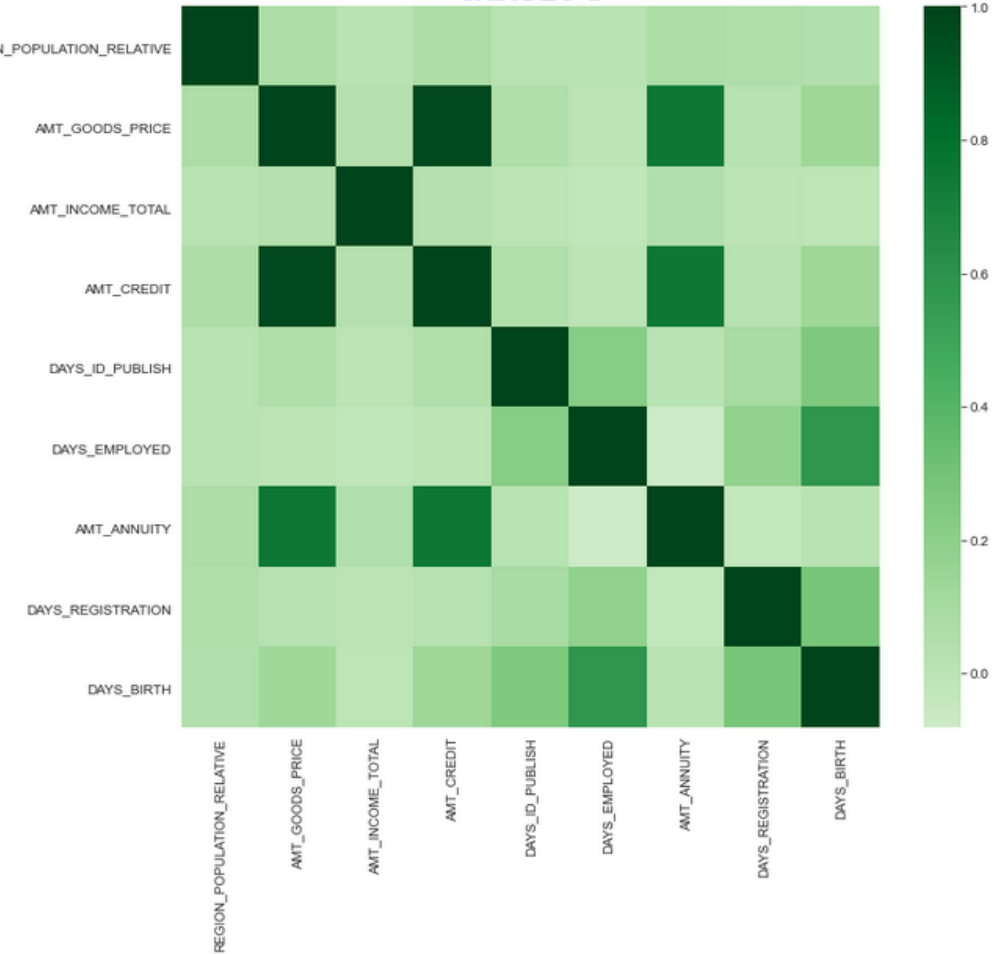
Correlation

13

TARGET 0



TARGET 1

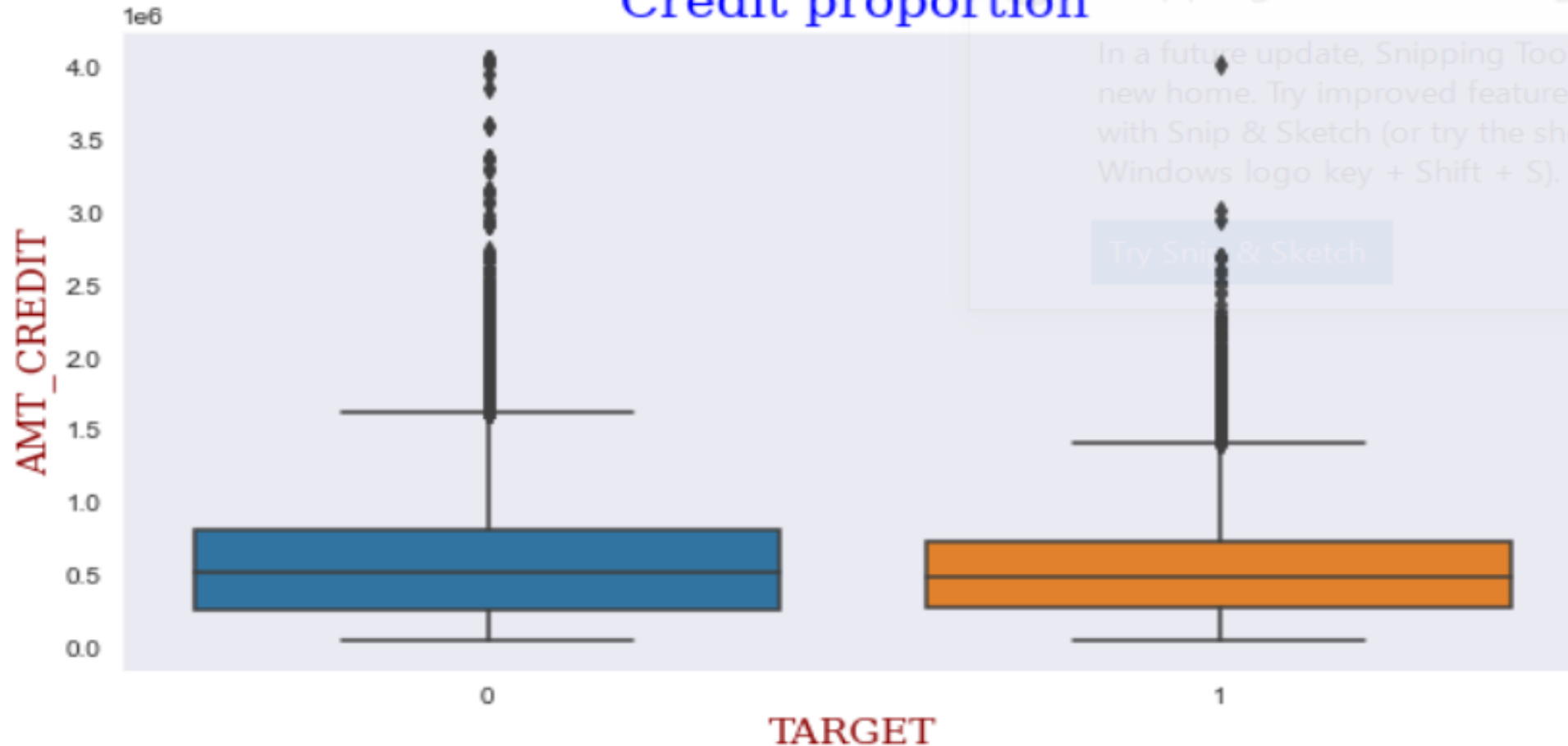


The correlation looks similar for TARGET 0 (who Pays on time) and TARGET 1 (Defaulters).

Distribution of Amount credit of TARGET variable

14

Credit proportion

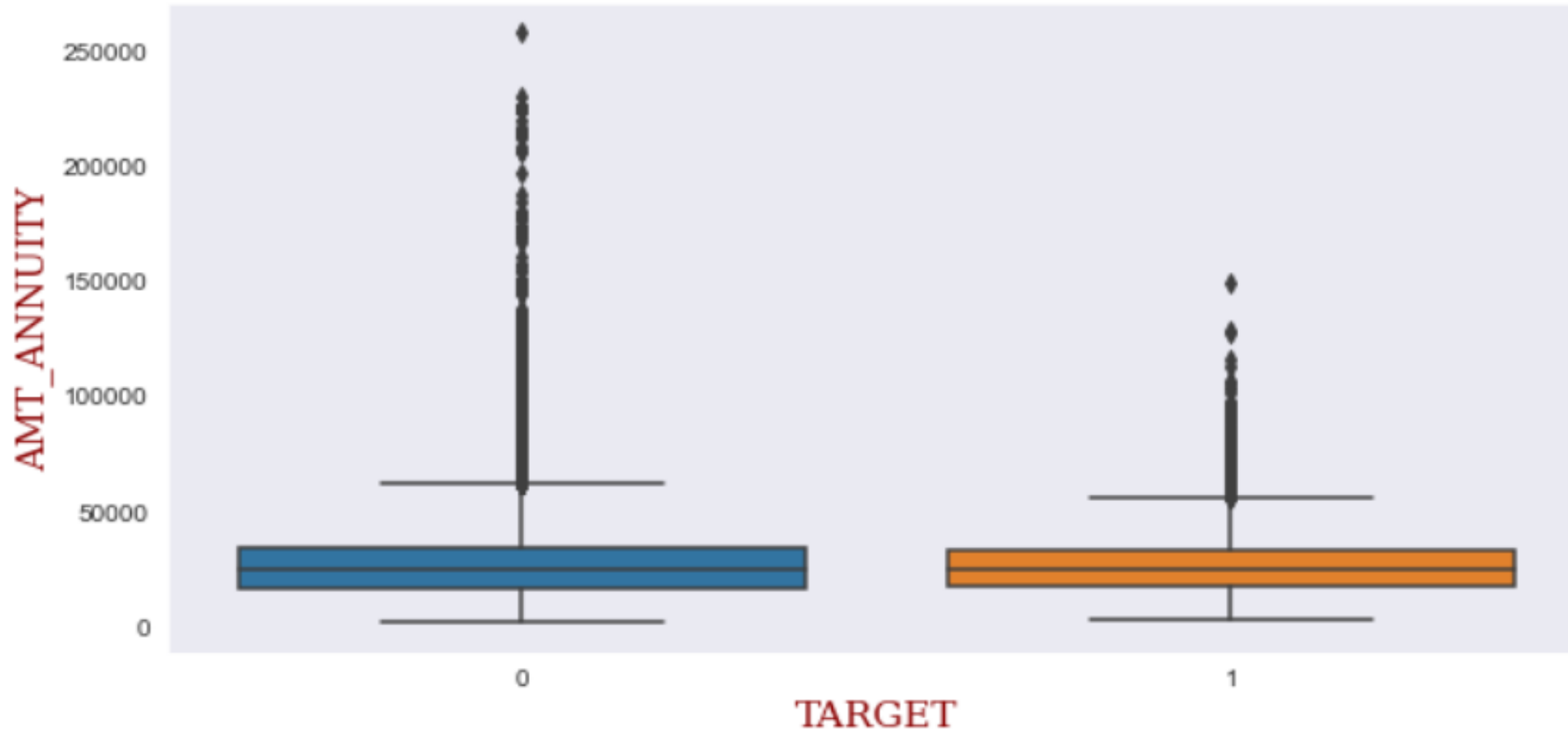


Amount credited has almost same distribution for both TARGET 0 and TARGET 1. There are high values of amount credit for TARGET 0 (pay on time)

Imbalance percentage of TARGET variable

15

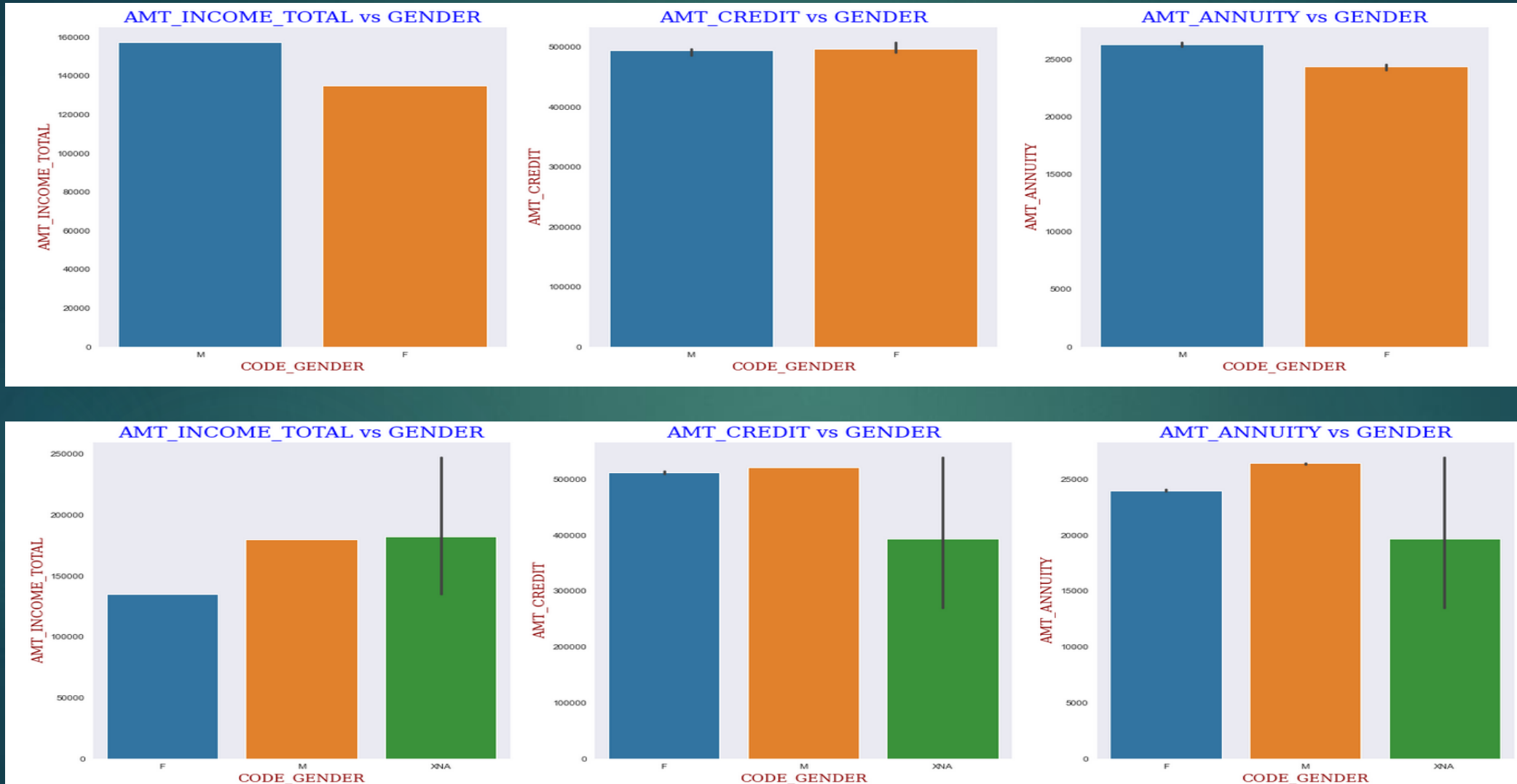
Annuity proportion



Amount annuity has almost same distribution for both TARGET 0 and TARGET 1. There are high values of amount credit for TARGET 0 (pay on time)

Proportion of Gender vs other variables

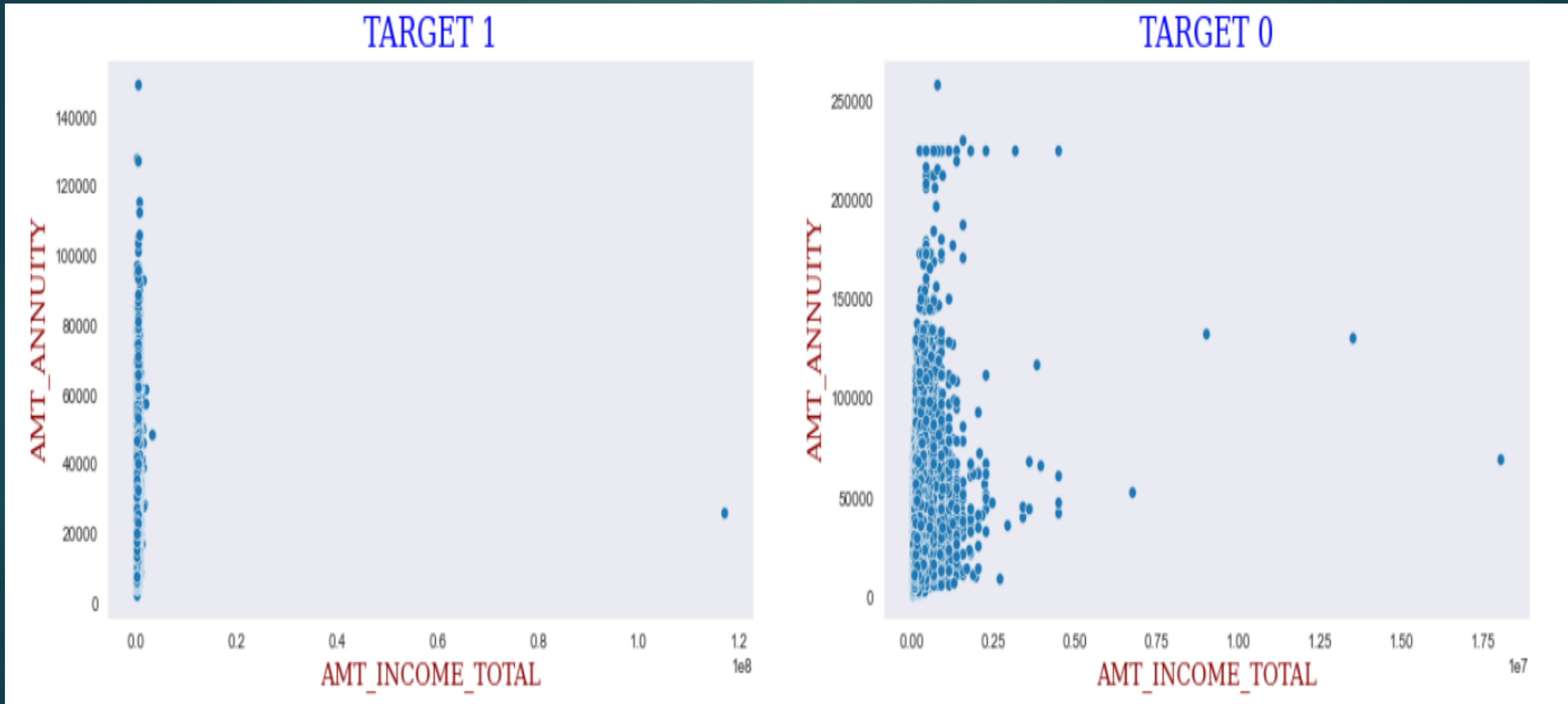
16



Males has higher income total, and so they have high Amount credit and Amount Annuity.

Relationship of Annuity vs Income

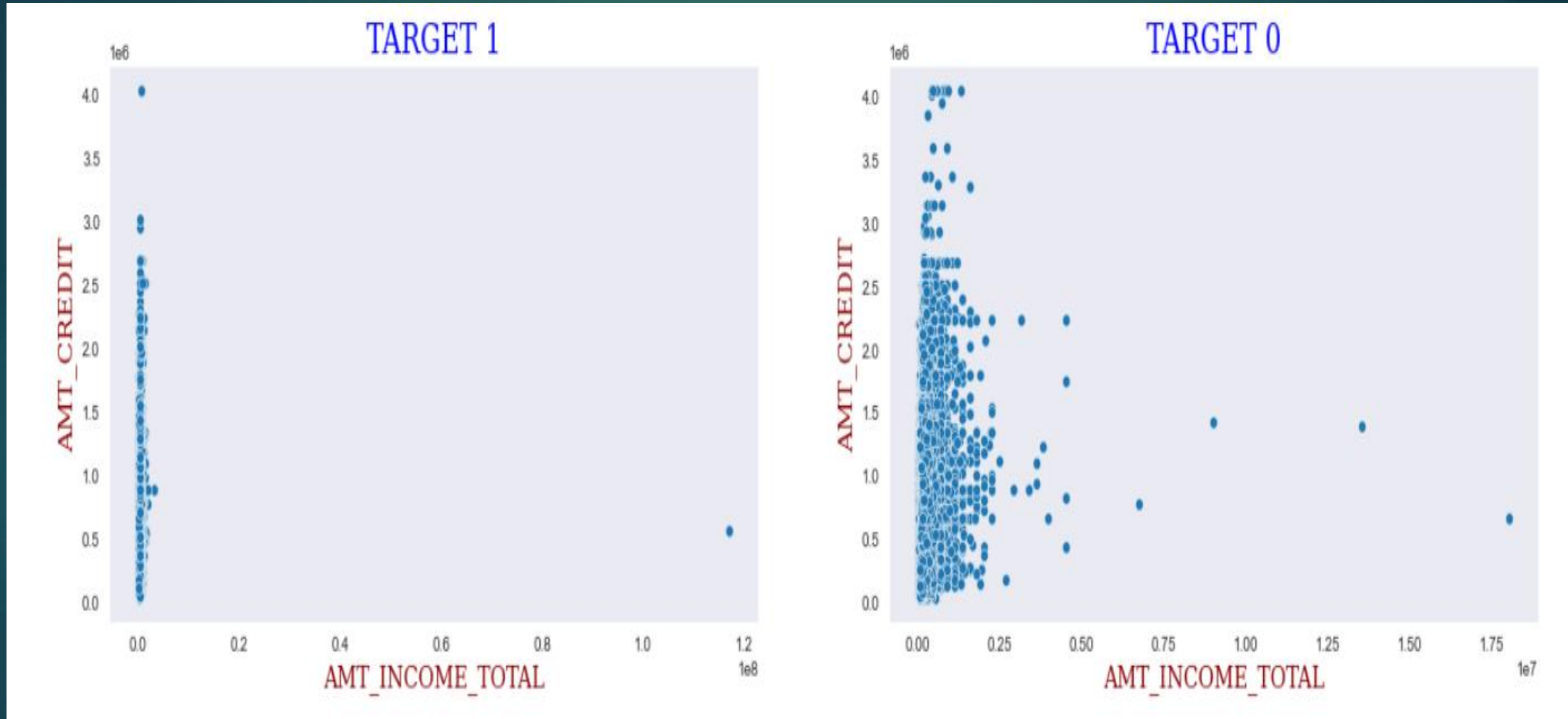
17



Applicants who pay on time have more scattered relationship between Amount Annuity and Income.

Relationship of Credit vs Income total

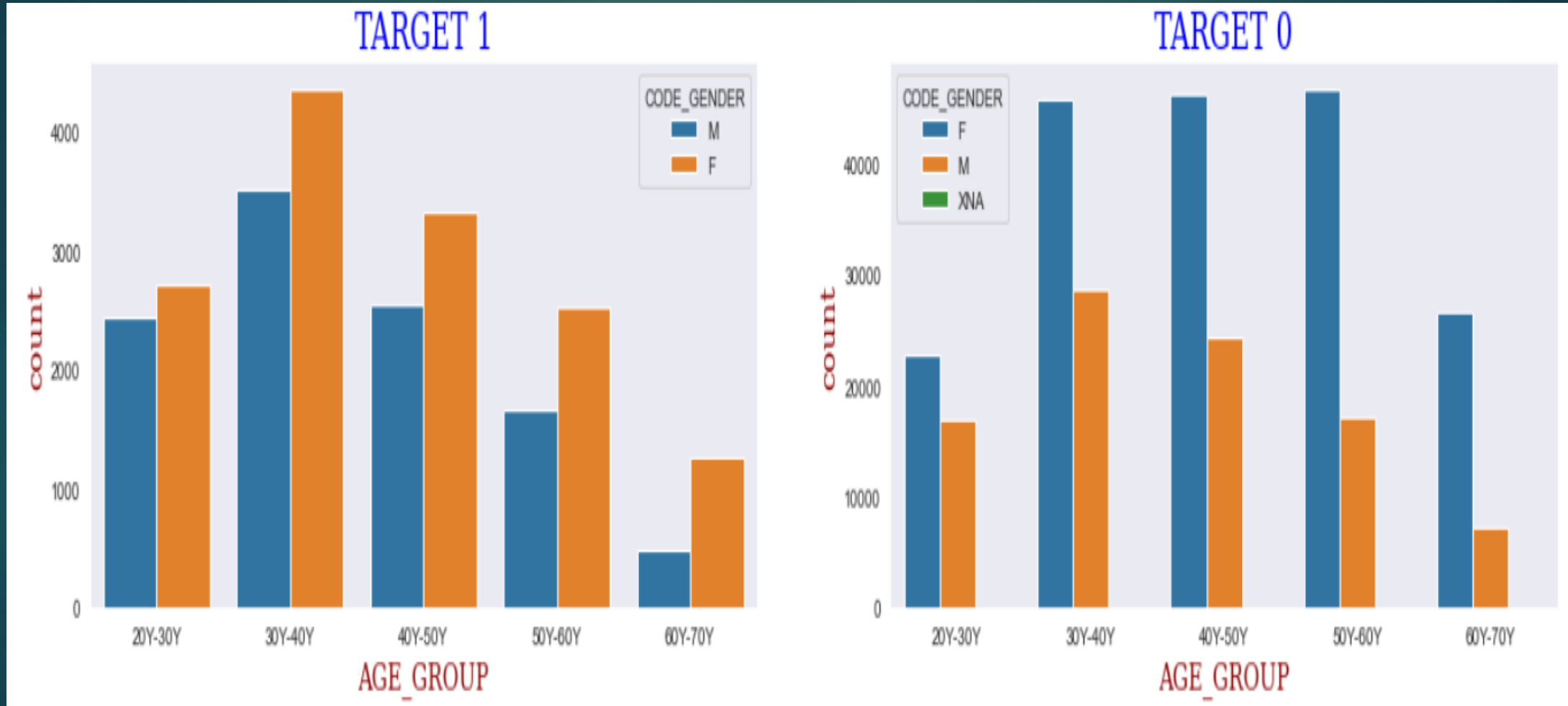
18



There is no linear relationship between Amount Credit and Income.

Proportion of Age group vs Gender

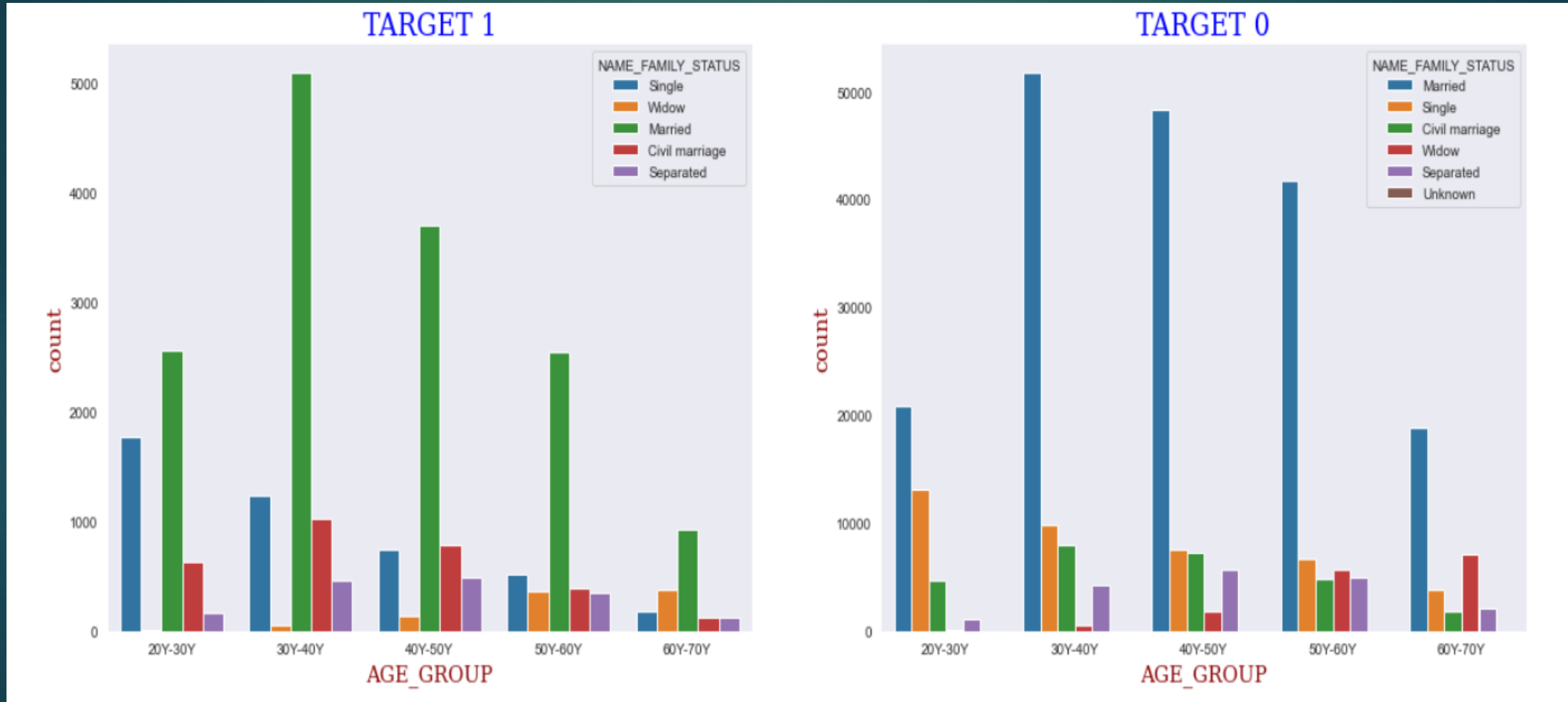
19



Females in the age-group of 30-40 years are more likely to default.

Proportion of Age group vs Family status

20

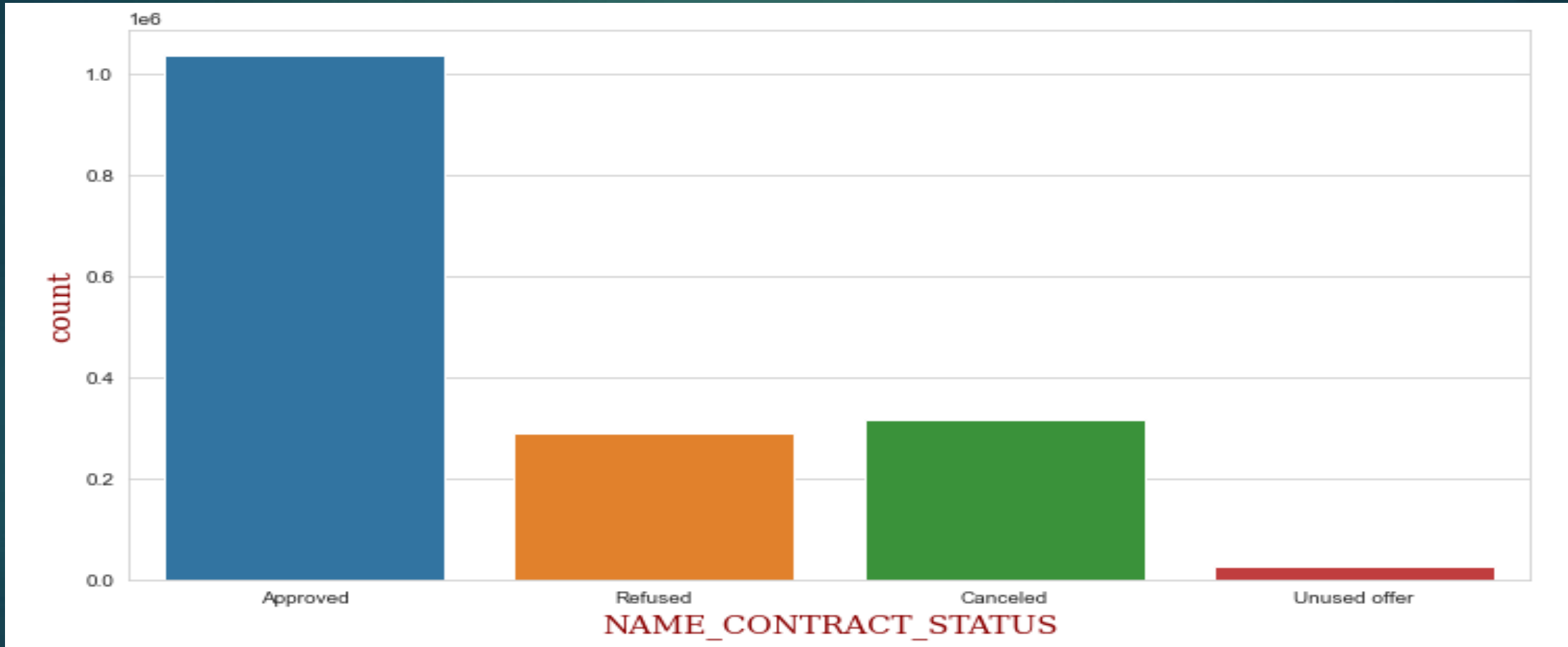


Married applicants in the age group of 30-40 years are more likely to default on their loans.

PREVIOUS APPLICATION ANALYSIS

Proportion of Contract status

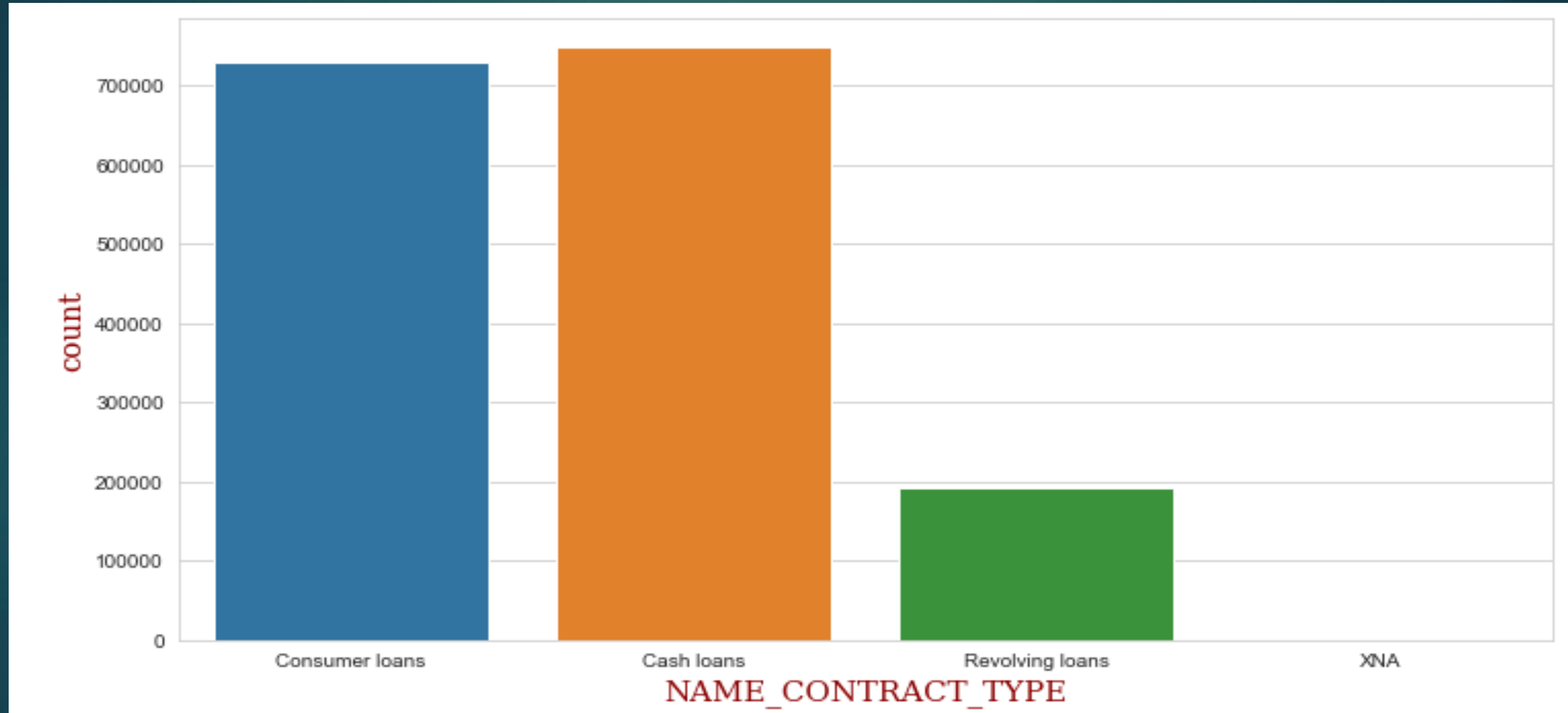
22



Number of **Approved** loans are higher compared to other counterparts. Therefore the percentage of loan which are being approved is high.

Proportion of Contract type

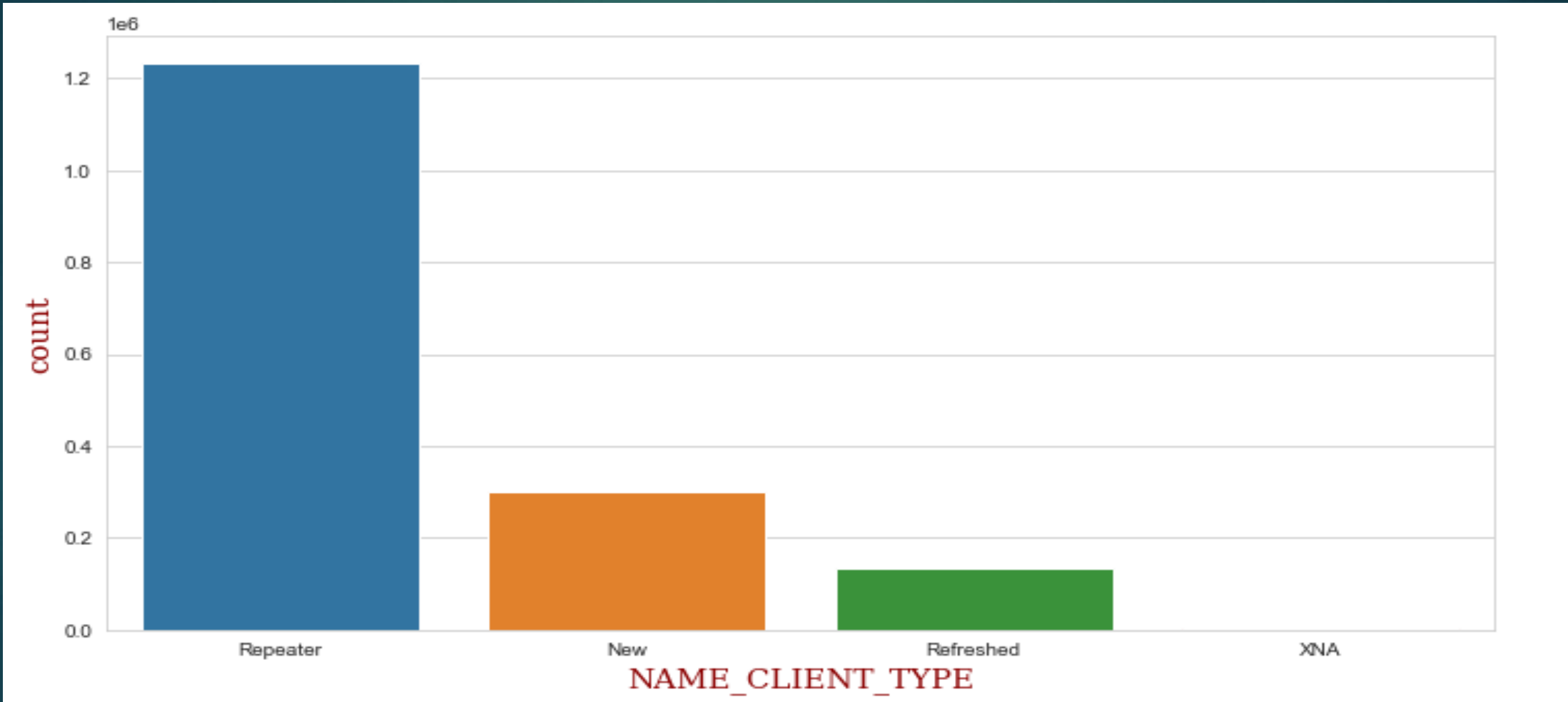
23



Cash Loans are preferred more than other contract types closely followed by consumer loans.

Proportion of Client type

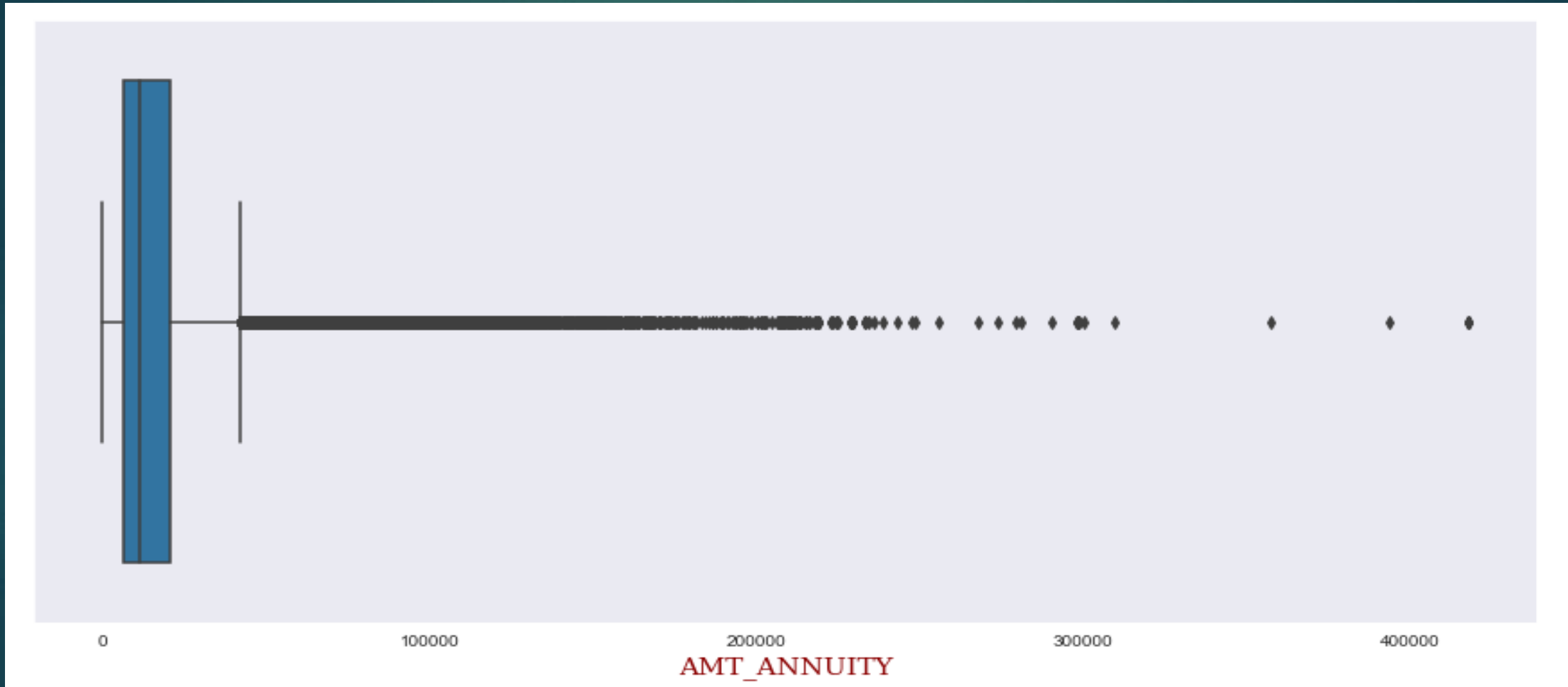
24



Repeater applications are higher in comparison to New applications.

Proportion for Annuity amount

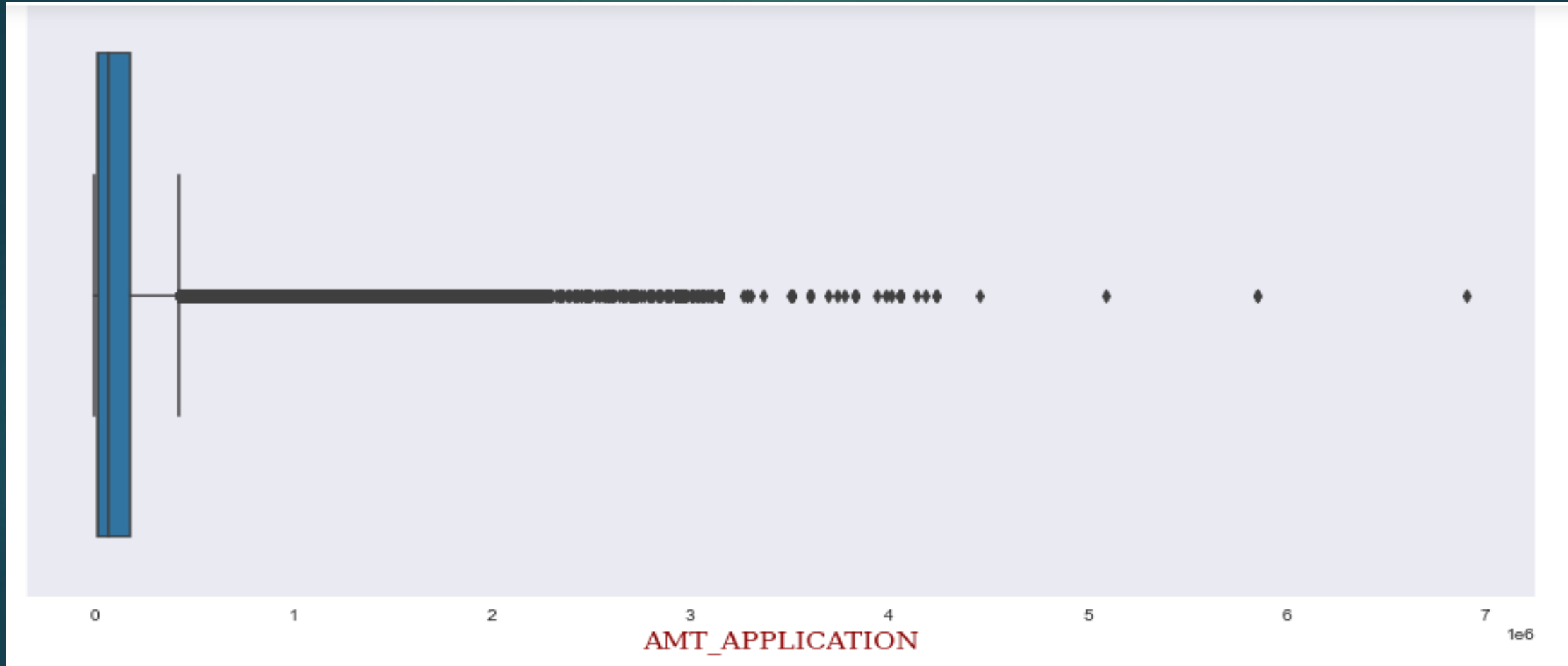
25



The majority of annuity amount is under 50000.

Proportion for Application amount

26



The majority of application amount is under 1000000.

Proportion for Credit amount

27

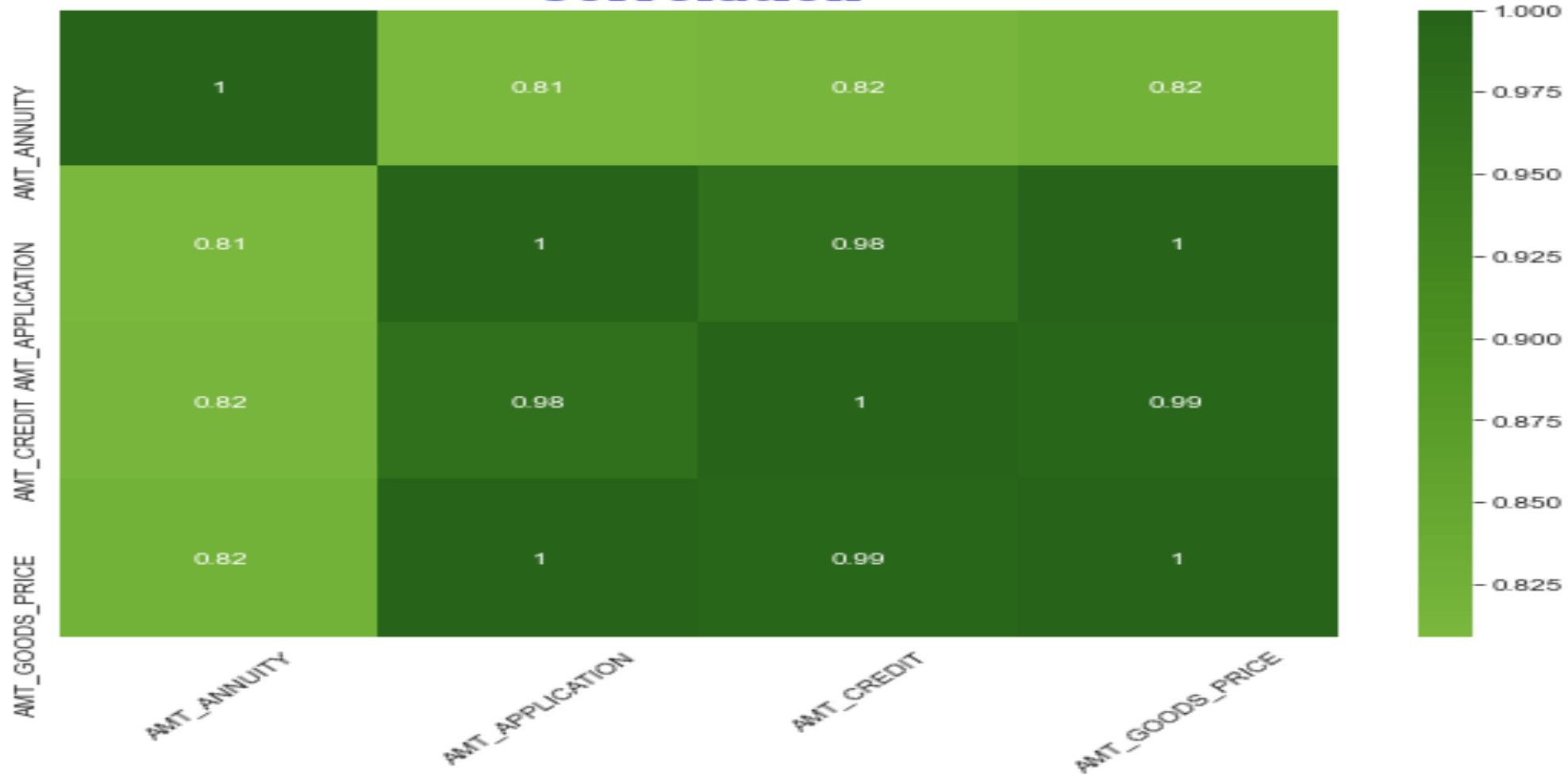


The majority of credit amount is under 1000000 but there are fewer number of outliers in comparison to application amount.

Correlation between continuous variables

28

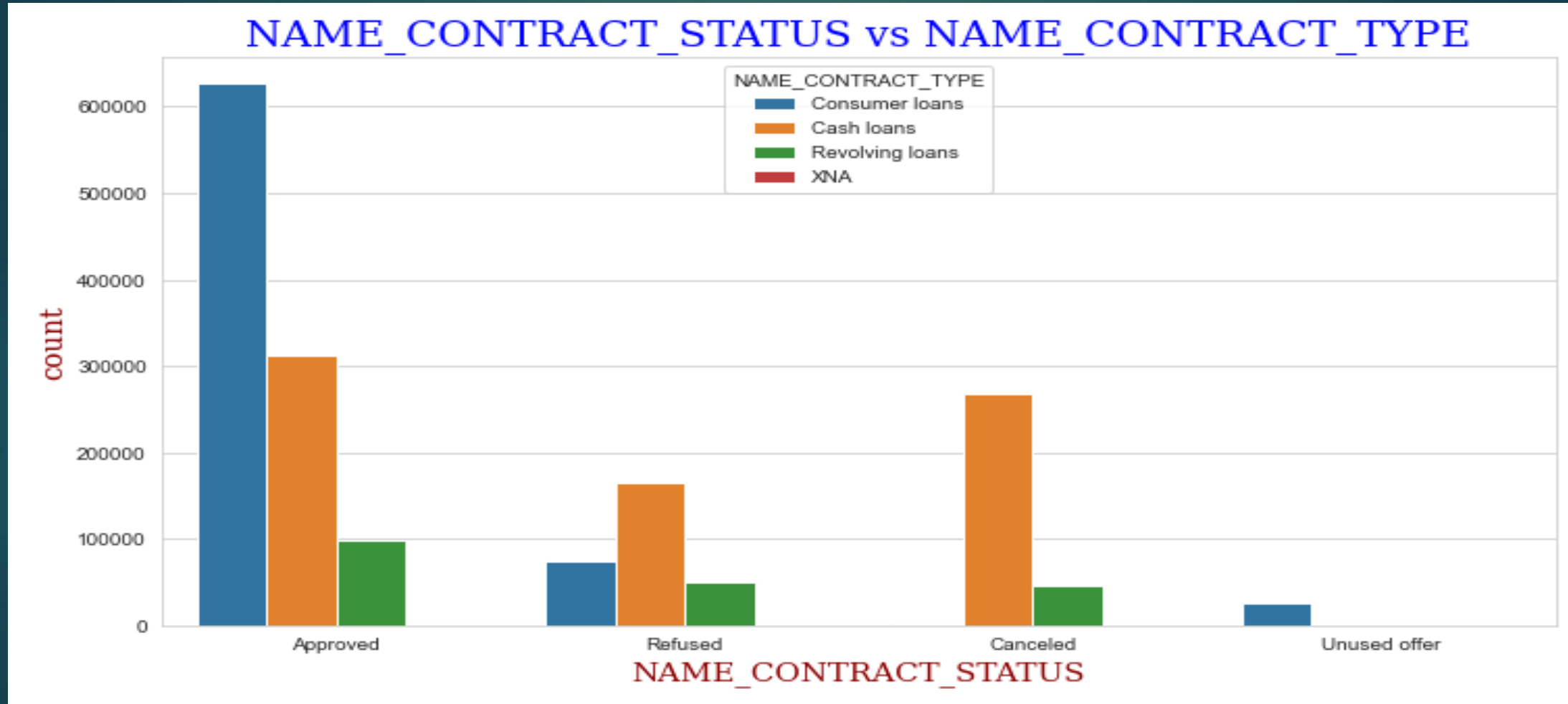
Correlation



Amount credit and amount application has higher co-relation(0.98).

Proportion of Contract status vs Contract type

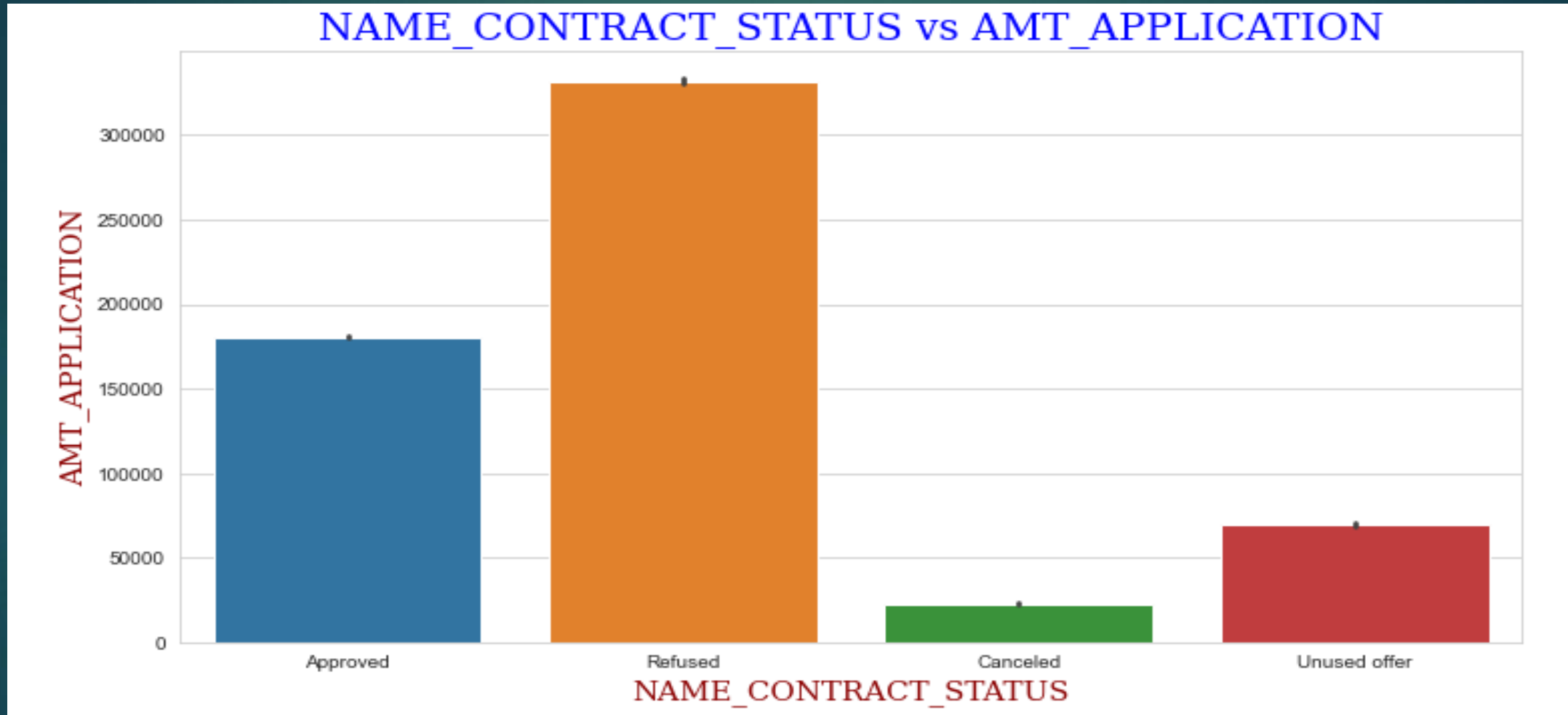
29



Consumer Loans are more in approved category. Cash loans are refused and cancelled more.

Proportion of Contract status vs Application

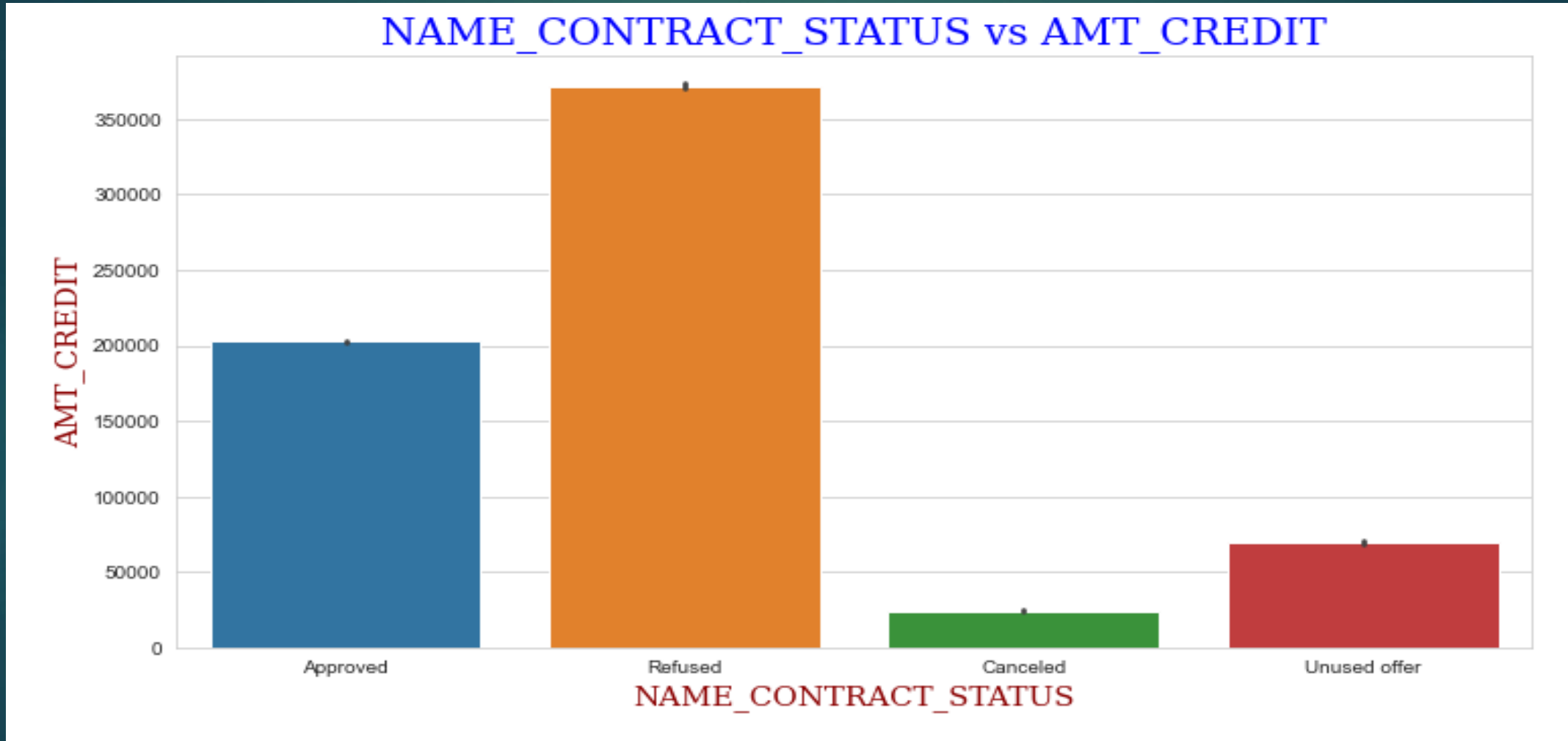
30



High value of amount applied for loans have been refused.

Proportion of Contract status vs Credit

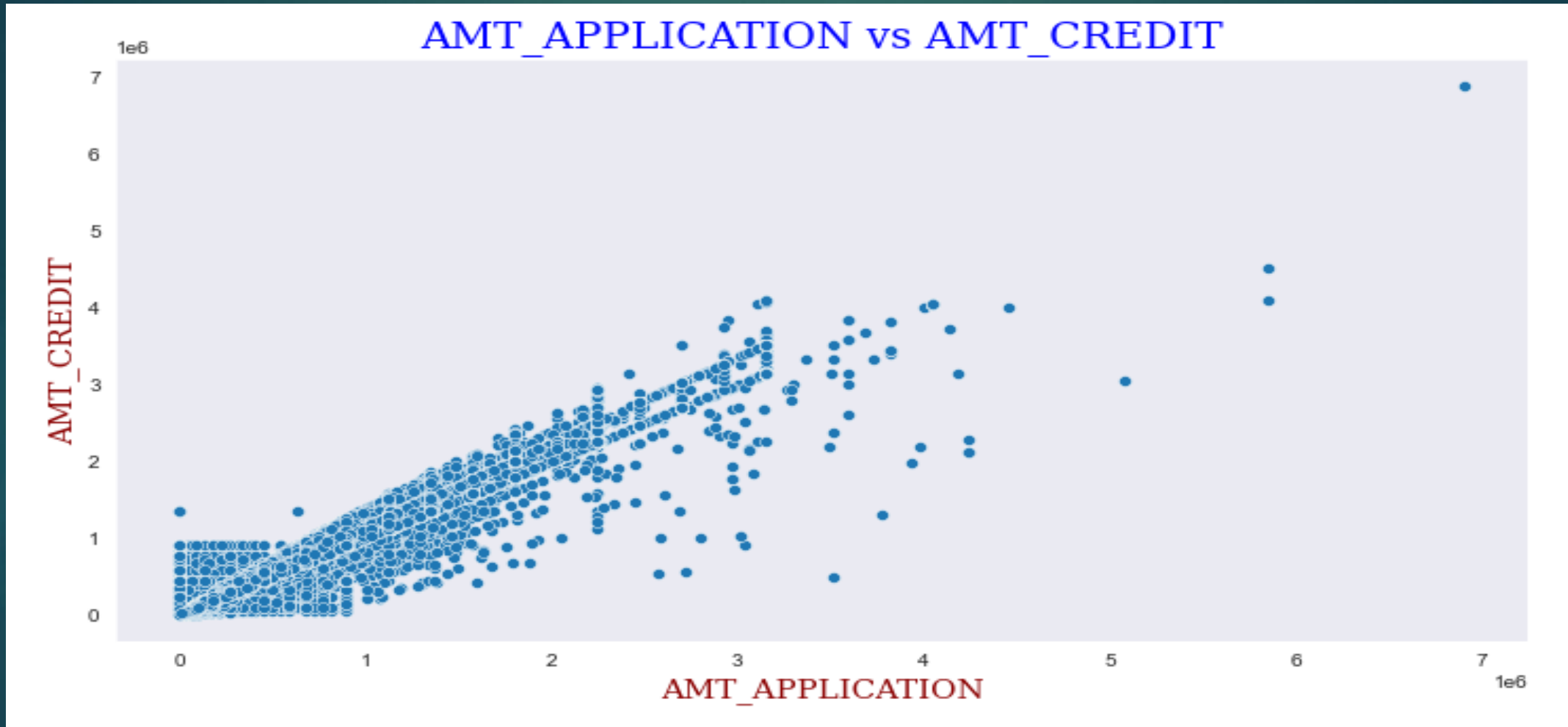
31



High value of amount credit for loans have been refused.

Proportion of Application amount

32



There is strong linear relationship between Amount application and Amount credit.

CONCLUSION

33

- EDA for the banking data set revealed that :
 - TARGET 0 (pay on time) is higher than TARGET 1 (Defaulter).
 - Female defaulters are higher.
 - Cash Loans are defaulted more.
 - Married applicants are more likely to default
 - Employees with less than 20 years of experience are more likely to default
 - Employees in the Age group range of between 30 to 40 are more likely to default
 - Secondary education type has higher default rate.

CONCLUSION continued...

34

- EDA for the banking data set revealed that :
 - Working applicants are more likely to default.
 - Females in the age-group of 30-40 years are more likely to default
 - Married applicants in the age group of 30-40 years are more likely to default on their loans.
 - Approved loans are higher compared to other counterparts
 - Cash Loans are preferred more than other contract types closely followed by consumer loans
 - Repeater applications are higher in comparison to New applications
 - Consumer Loans are more in approved category. Cash loans are refused and cancelled more.
 - There is strong linear relationship between Amount application and Amount credit.

THANK YOU