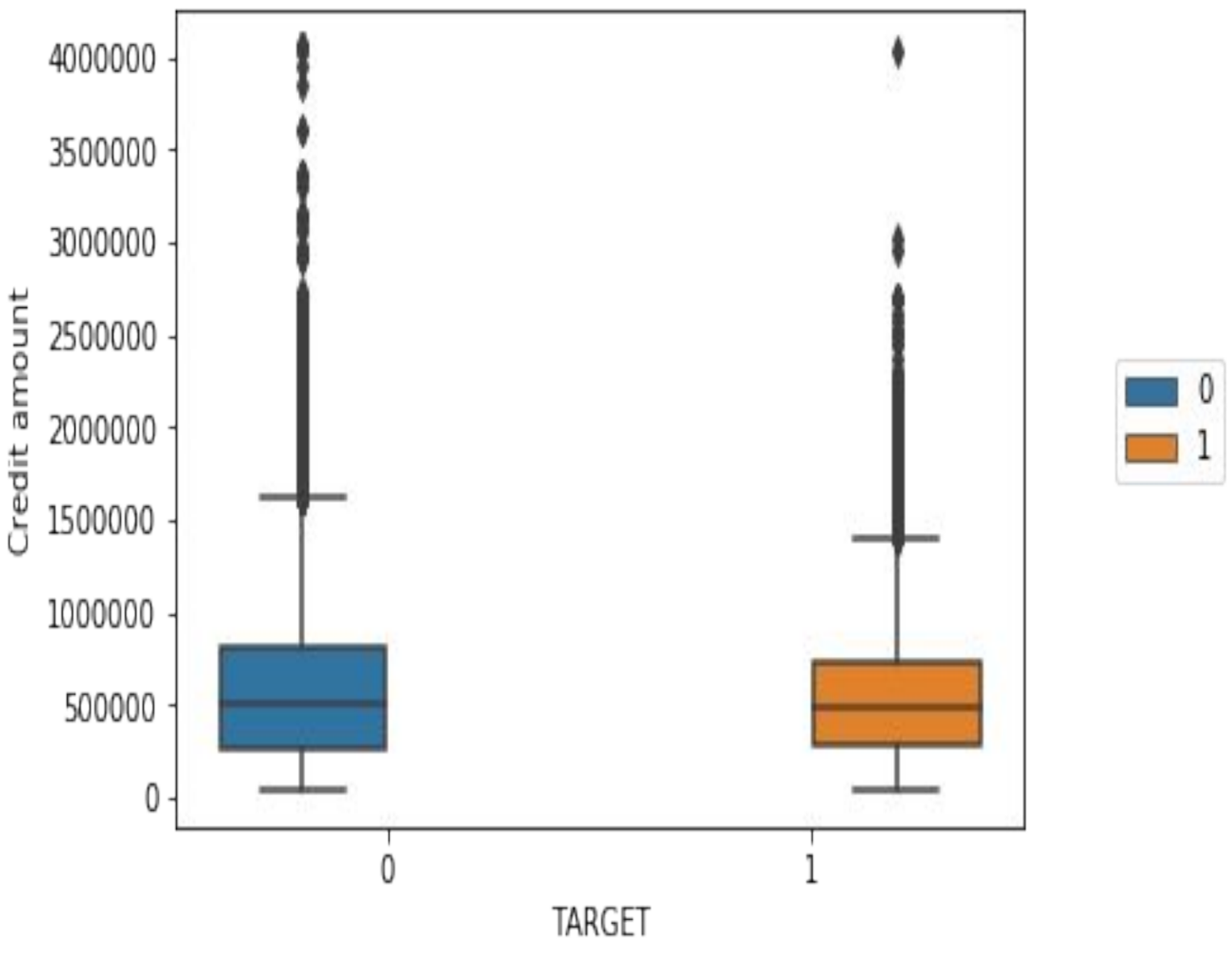
EDA Group case study

Topic: Identify driving factors behind loan default

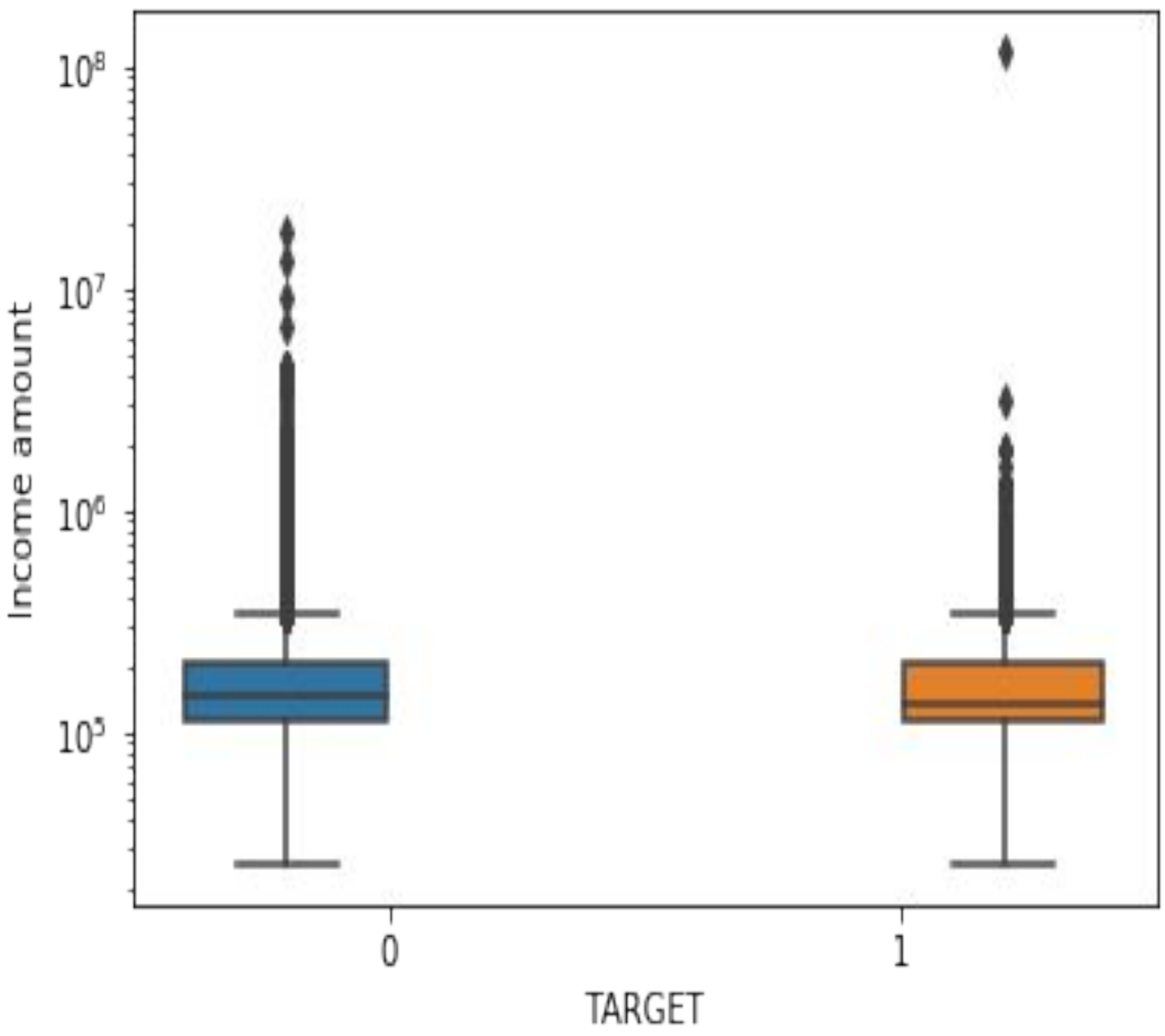
By,
Rucha Shelke
Sneha Velankar

Outliers



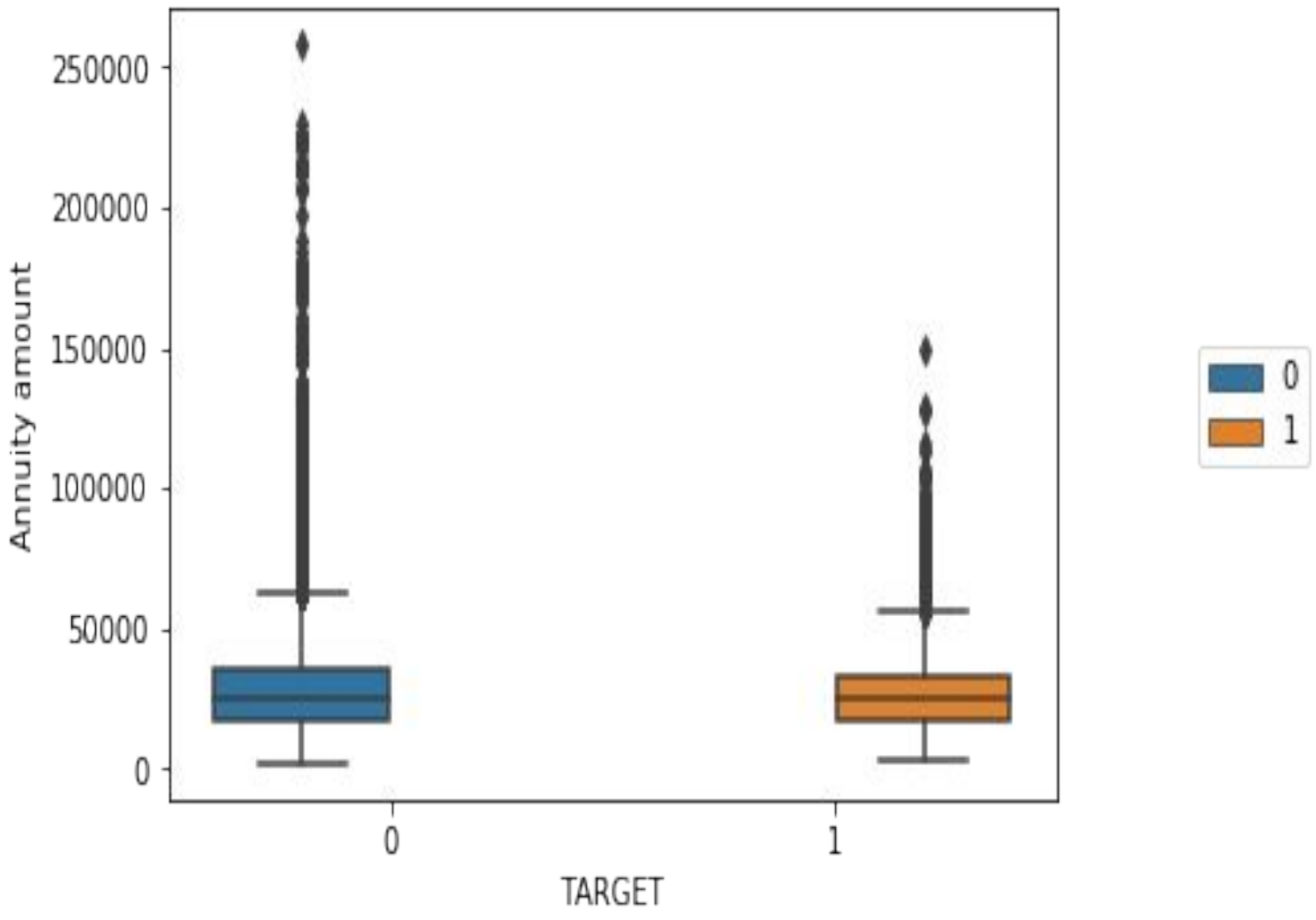
Data has outliers in case of Credit amount for both the targets (Target: 1 difficulty in payment and Target: 0 all other cases)

Outliers



Data has outliers in case of Total income for both the targets (Target: 1 difficulty in payment and Target: 0 all other cases)

Outliers



Data has outliers in case of Annuity amount for both the targets (Target: 1 difficulty in payment and Target: 0 all other cases)

Data Imbalance

1. Target variable:

    Imbalance percentage is 8.073

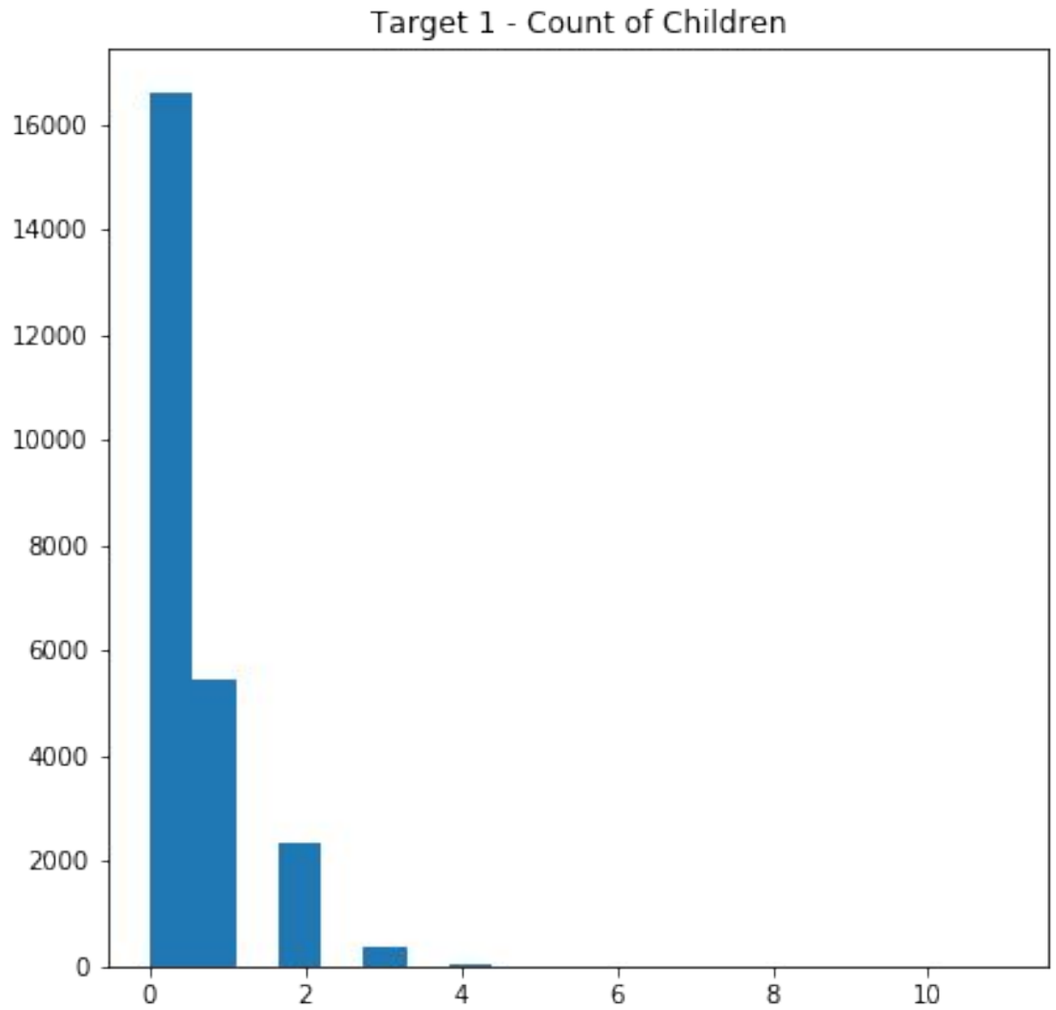
2. Gender variable:

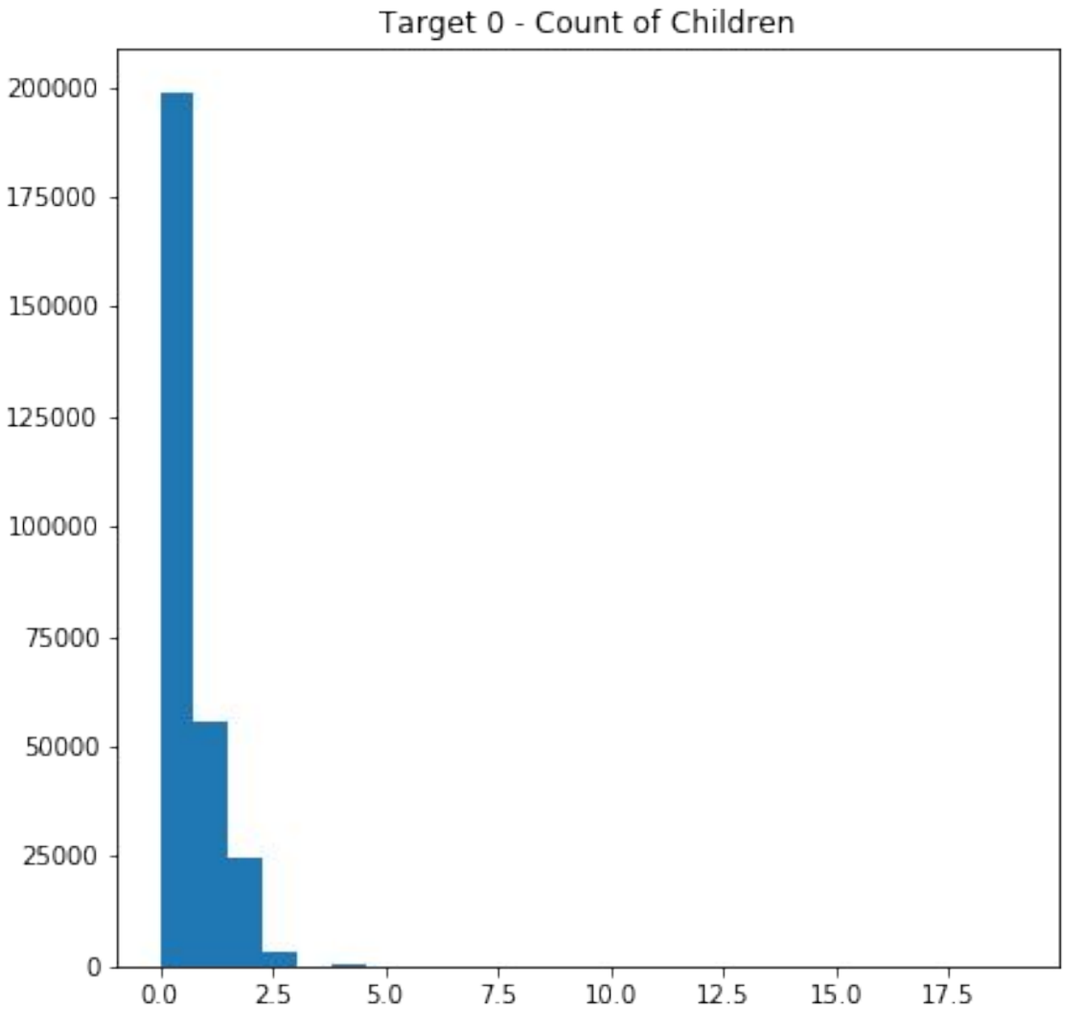
    Gender imbalance Male: 34.16

    Gender imbalance Female: 65.84

Univariate Analysis

Number of Children

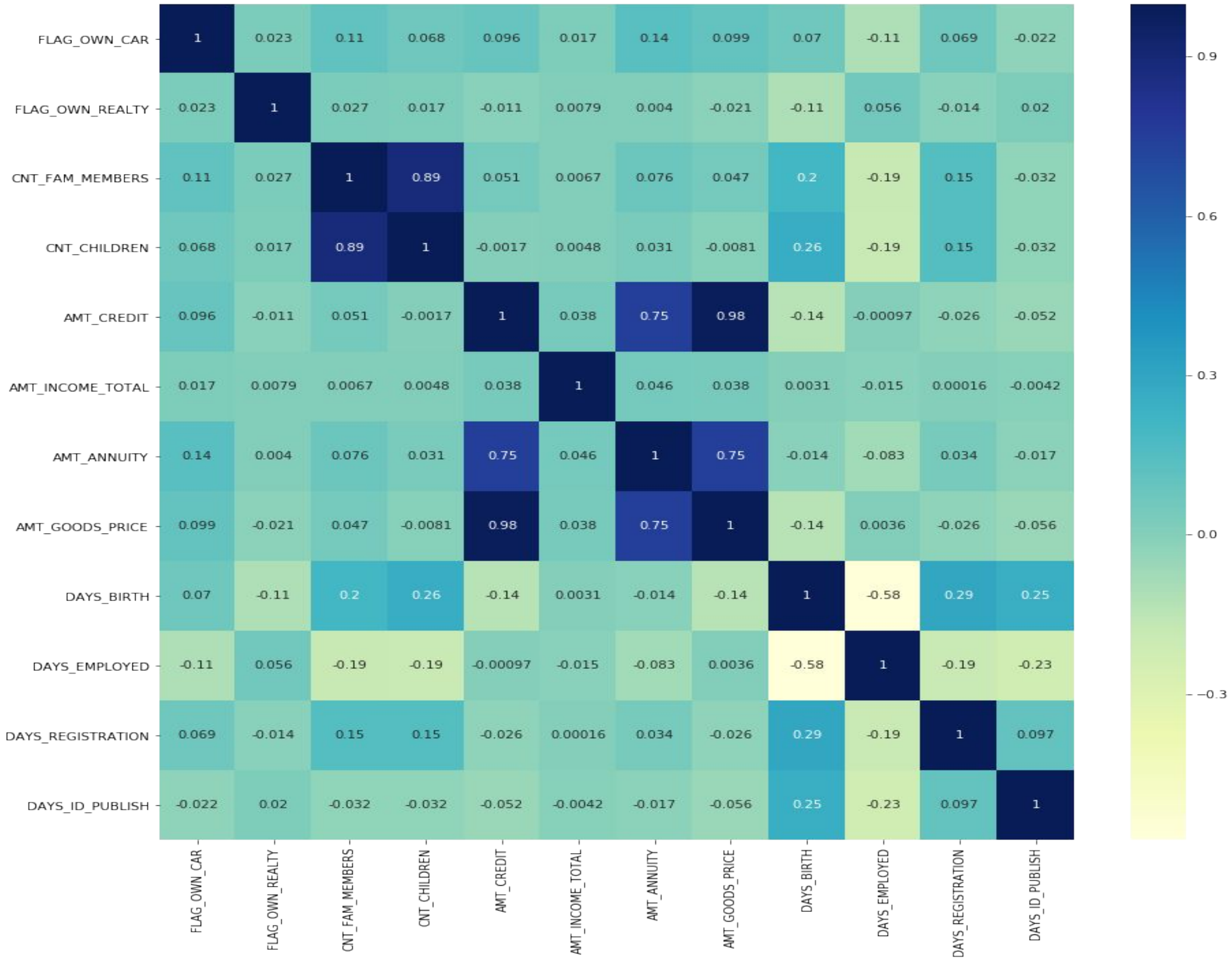




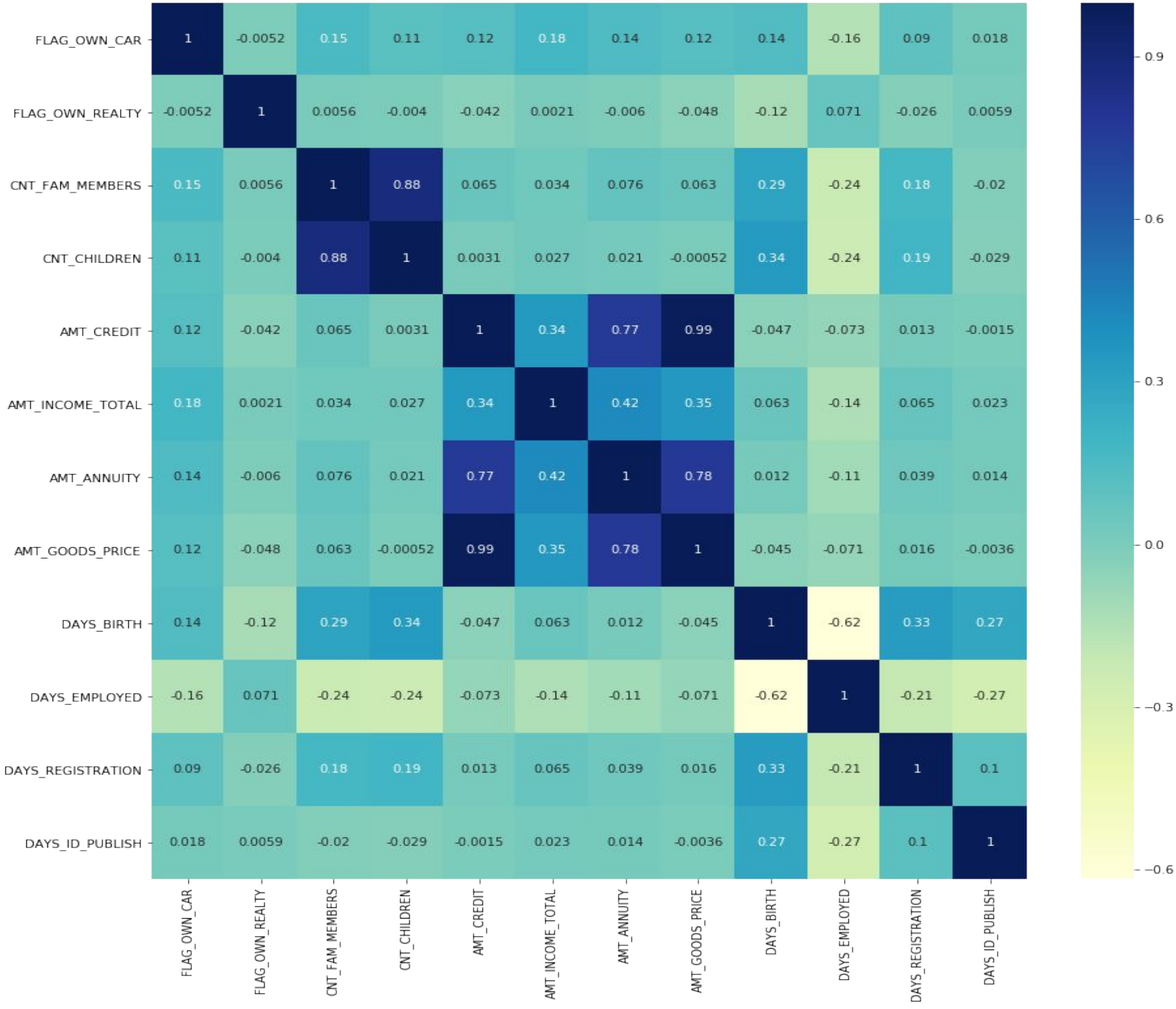
As you can see in previous graph, there is not much difference in case of number of children for clients finding difficulties and clients who can repay the loan.

Correlation Analysis

Correlation analysis on clients finding difficulties in payment



Correlation analysis on clients who are able to repay the loan

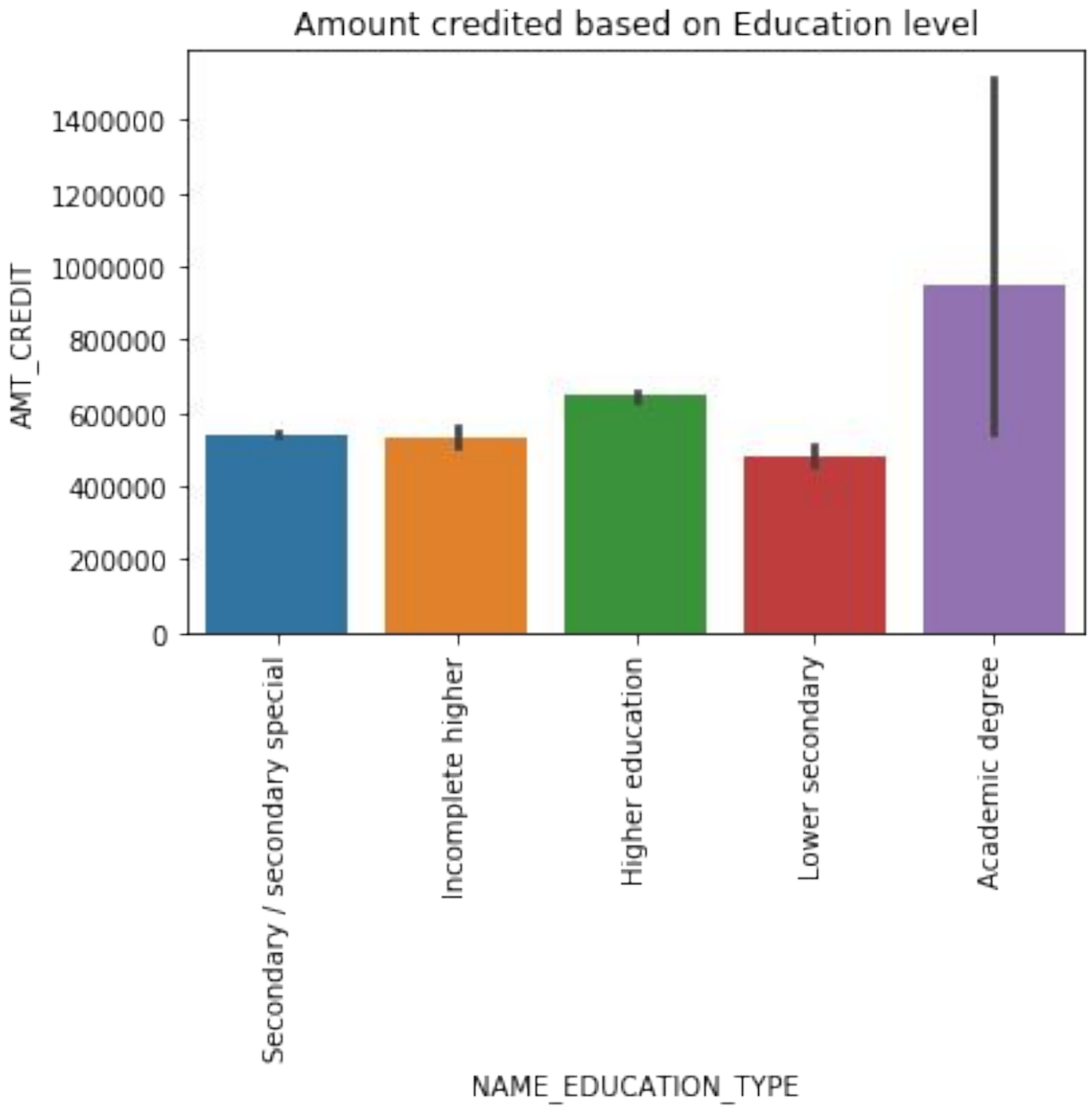


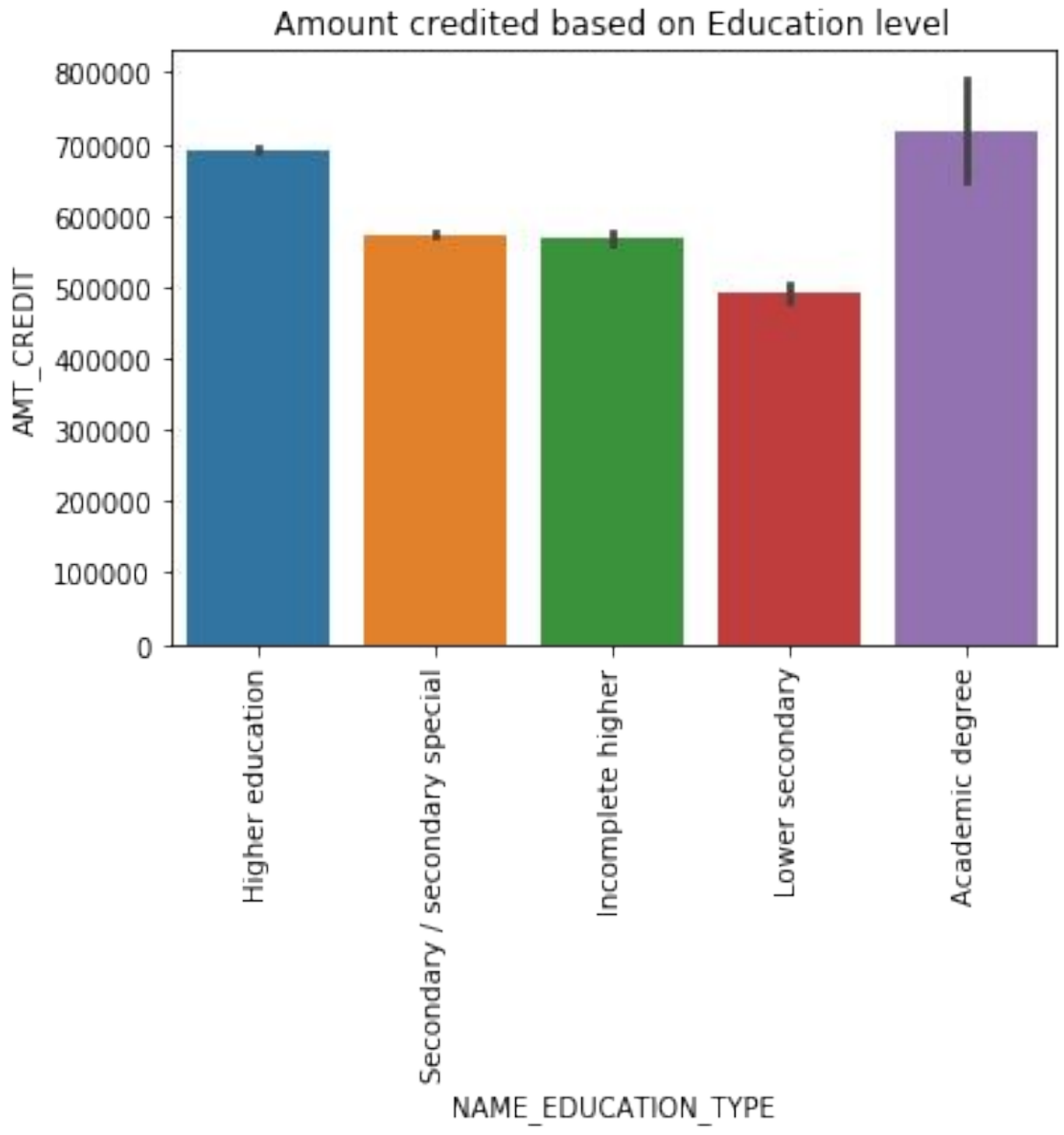
Correlation analysis conclusion

Variables with highest correlation are same in both the cases (Target 1 and 0)

1. Highly positively correlated variables are: Credit amount, Annuity amount, goods price amount
2. Highly negatively correlated variables are: DAYS_BIRTH and experience of employee in current organization

Amount Credited vs Education for Target 1 and 0





Amount Credited vs Income type for Target 1 and 0