EPGP PROGRAMME IN DSc

Credit EDA Assignment

Anki Singha
DSC 56

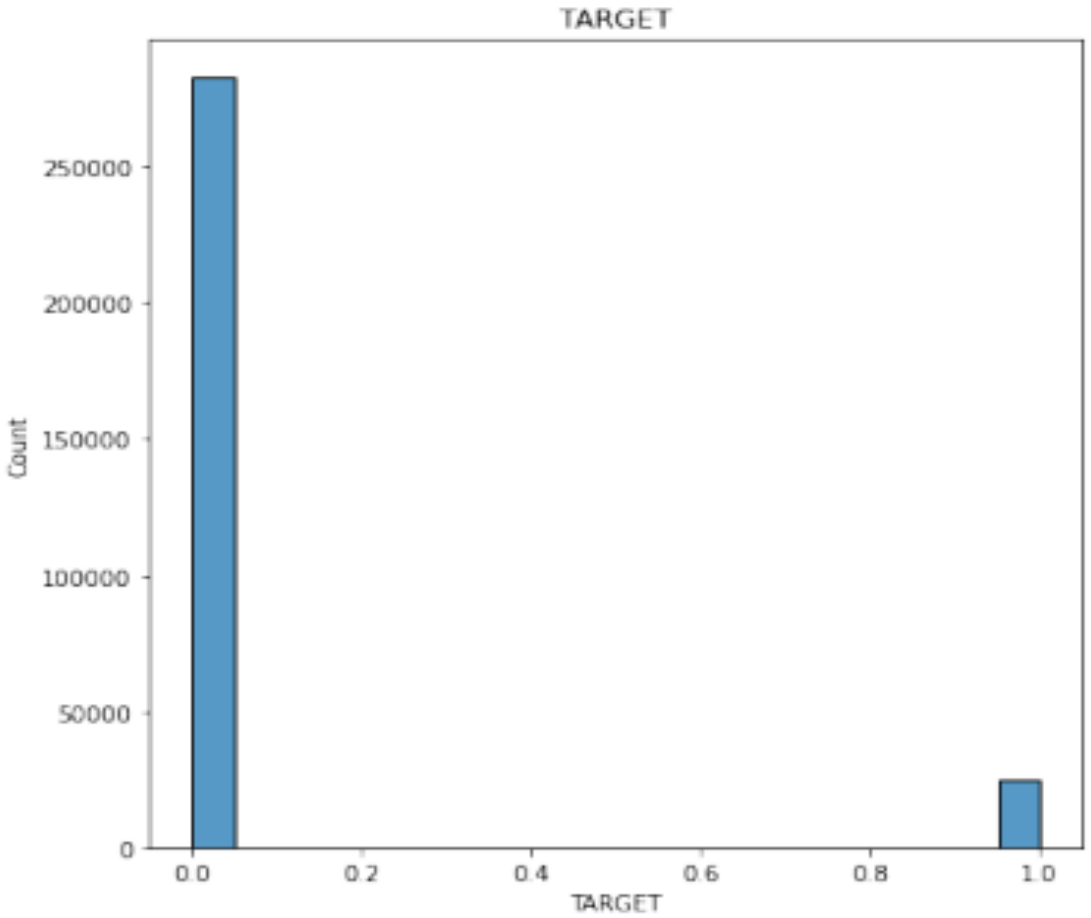
Abstract:

- This assignment aims to give you an idea of applying EDA in a real business scenario. In this assignment, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

- Methodology:
  - Understanding data from data dictionary.
  - Data cleaning :- Dealing with missing values, identifying data types, outliers, data imbalance, sorting and filtering data frames, fixing invalid entries , standardizing column(s).
  - Univariate analysis of segmented data frames using different plots.
  - Bivariate analysis of segmented data frames using different plots.
  - Multivariate analysis of segmented data frames using correlation matrix and heat map.

  “The client with payment difficulties “ has been used as “TARGET” variable denoted by ‘ 1’.

ANALYSIS:-

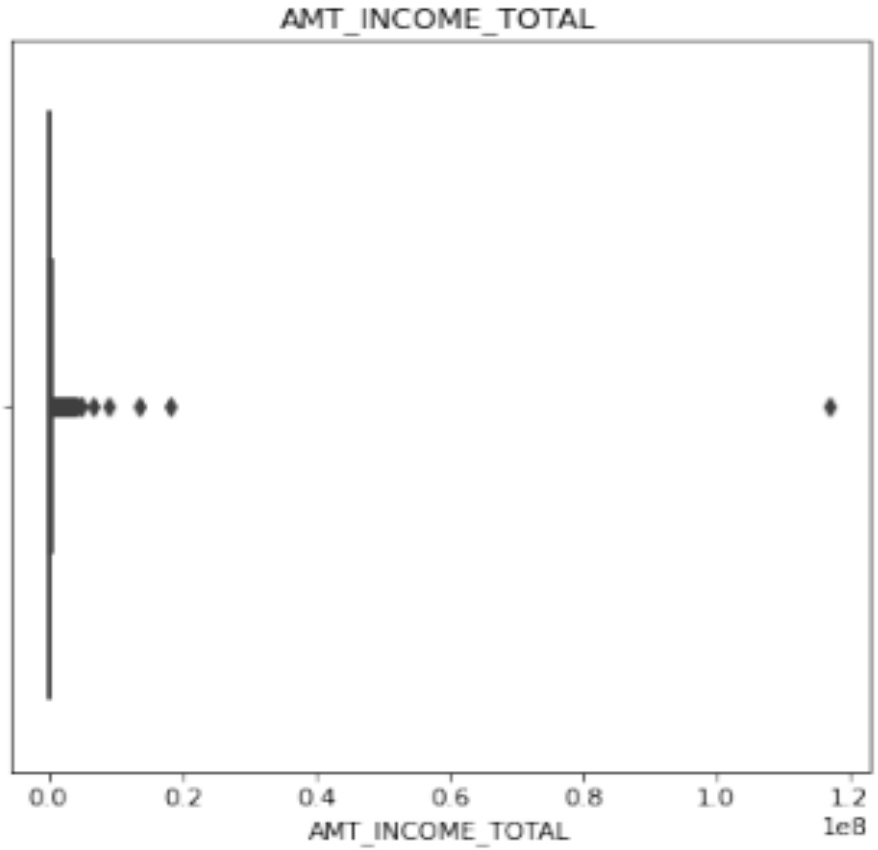


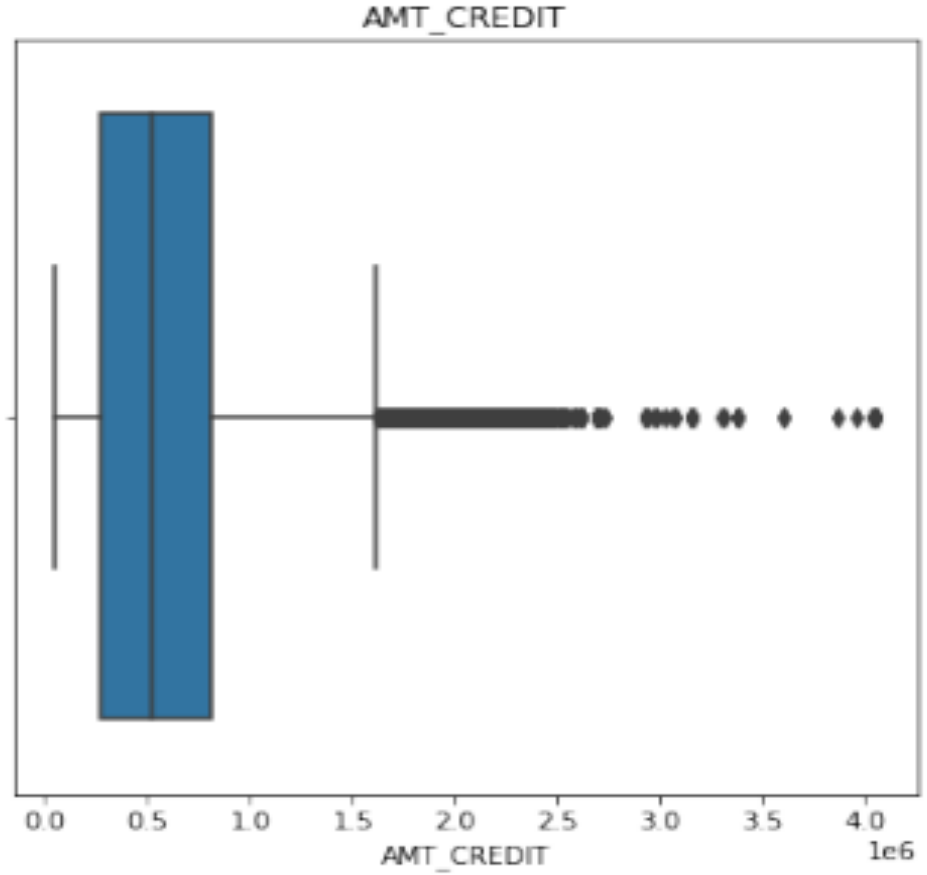
This histogram plot is clearly showing that there is data imbalance .

The ratio of data imbalance is
For data frame 1-
Non defaulter(‘0’) : defaulter(‘1’) = 11.386586102719033 : 1.0

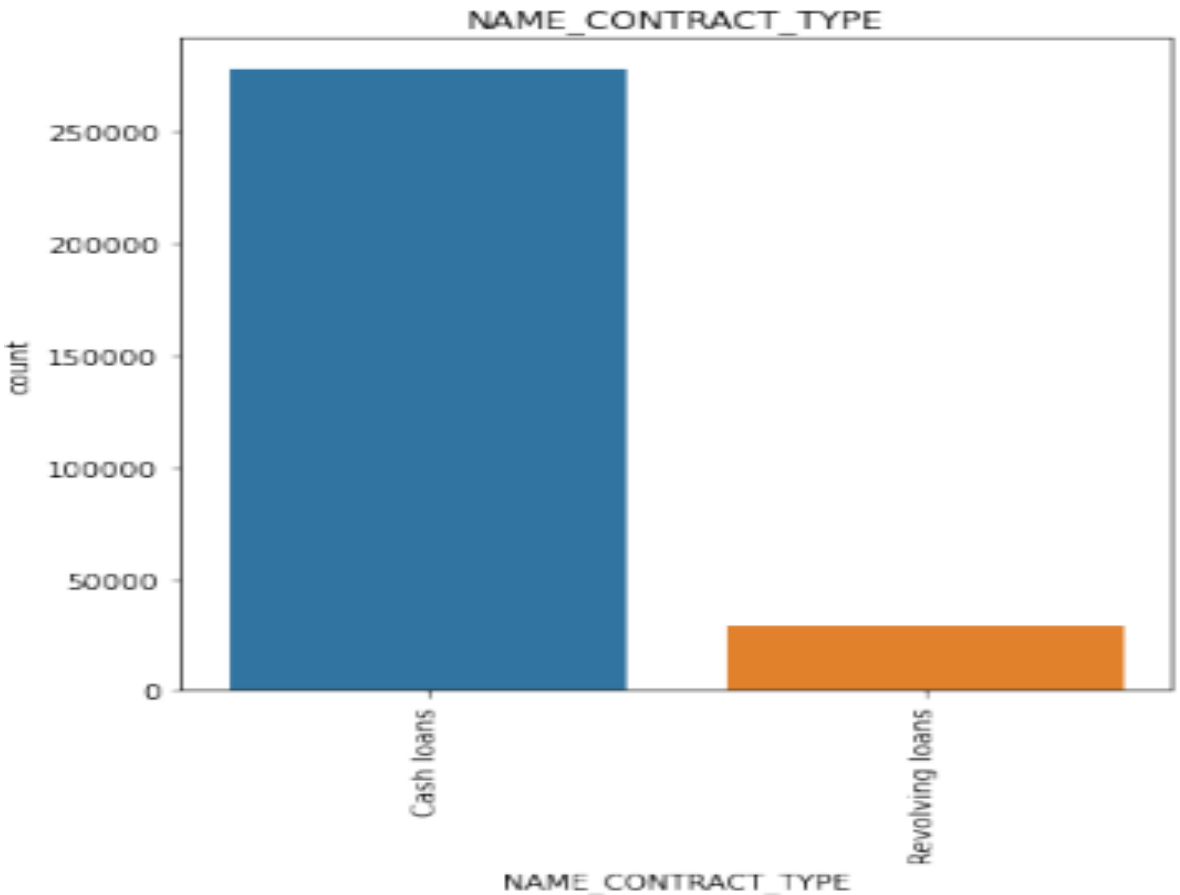
For data frame 2-
Non defaulter(‘0’) : defaulter(‘1’) = 10.7036887452881 : 1.0

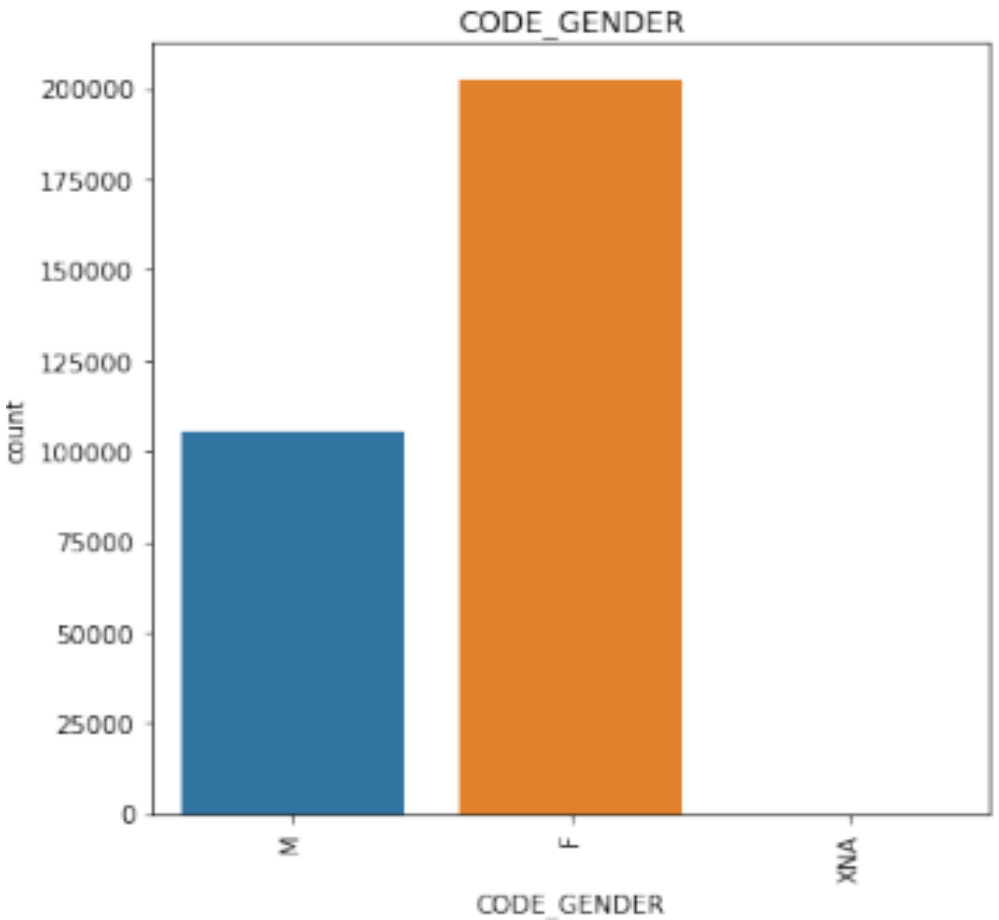
OUTLIERS:-





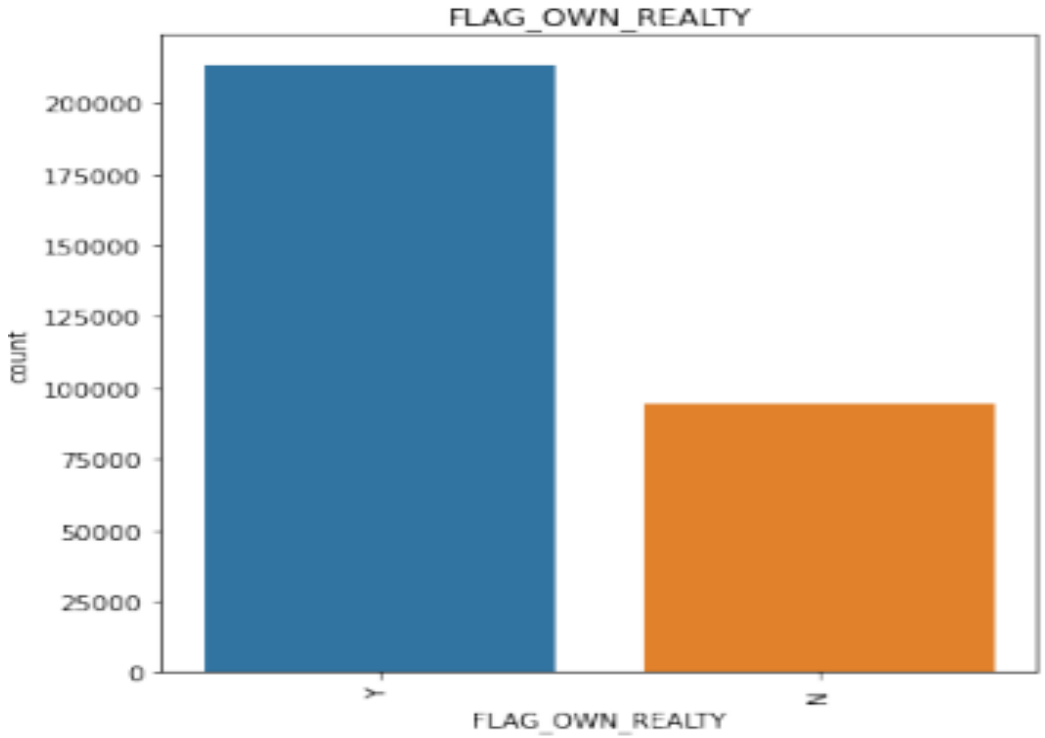
These box plots are showing that there are outliers in both columns. In case of income beyond 0.2*10^8 (Rs) there are outliers including potential outliers.
In case of Credit beyond 2.5*10^8 (Rs) there are outliers including potential outliers.

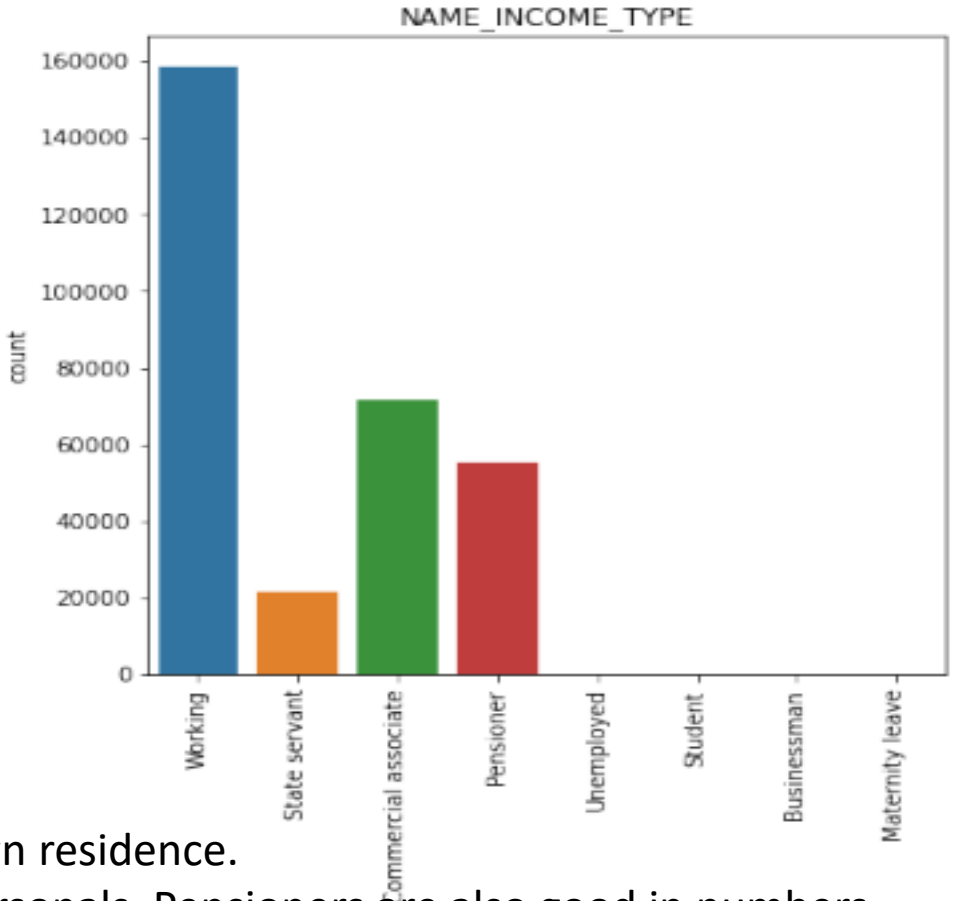




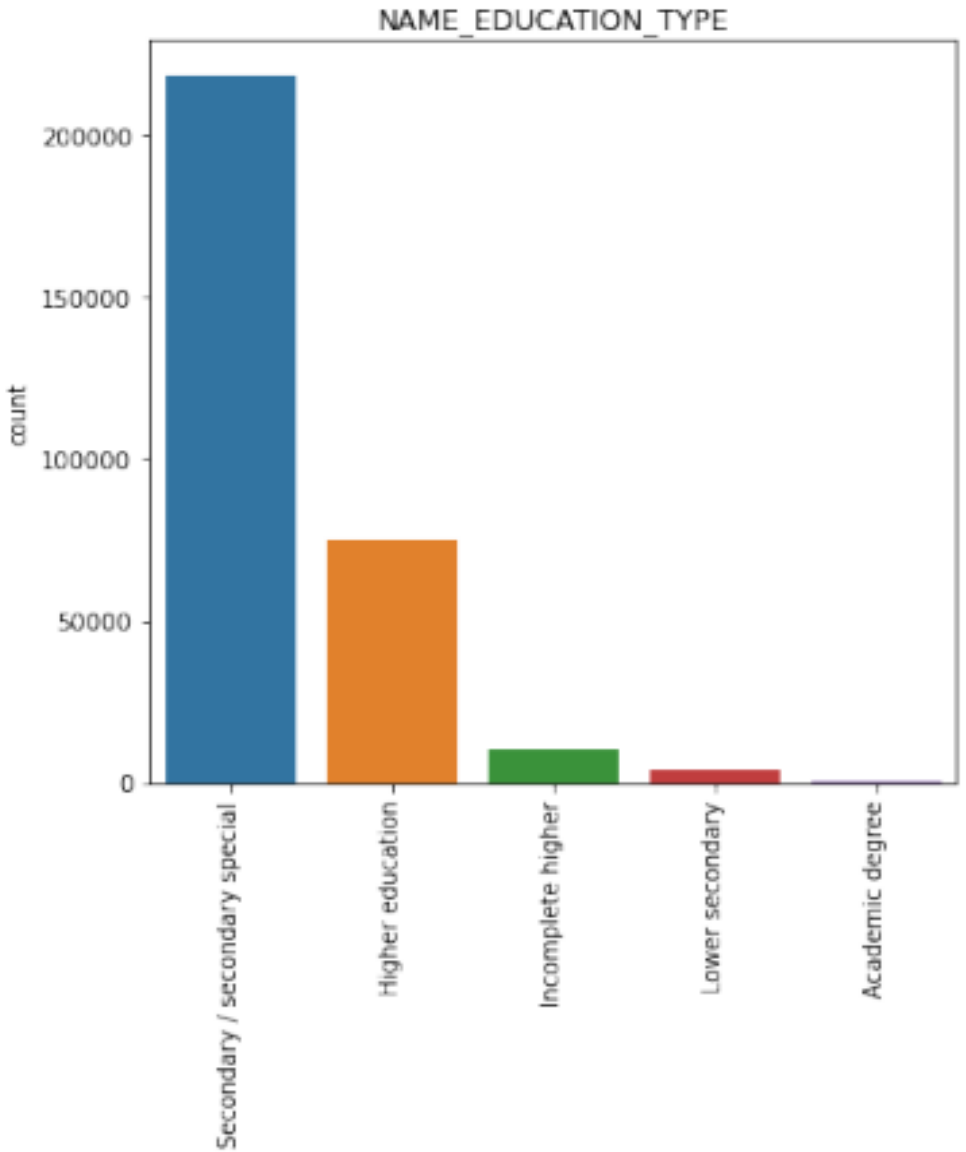
These count plots are showing that

- Cash loan disbursement is higher than revolving loan .
- Number of female borrowers is higher than male borrower.

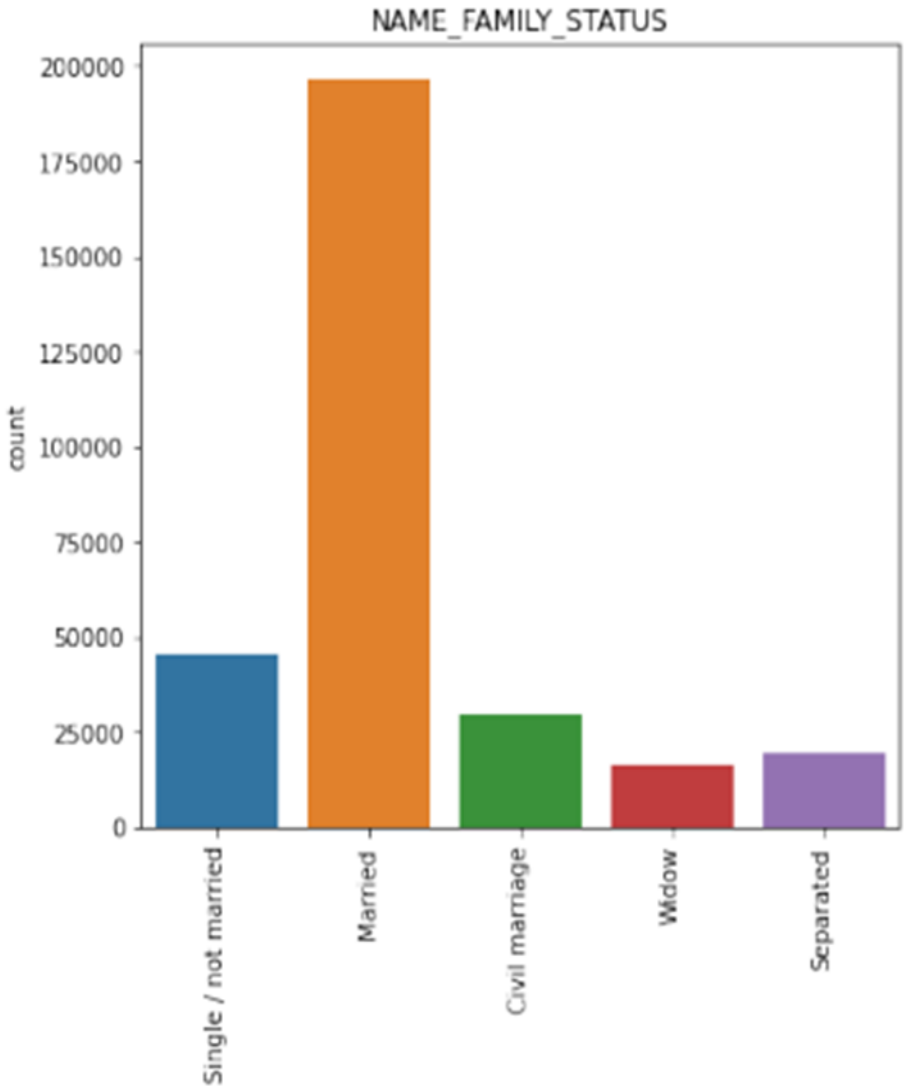




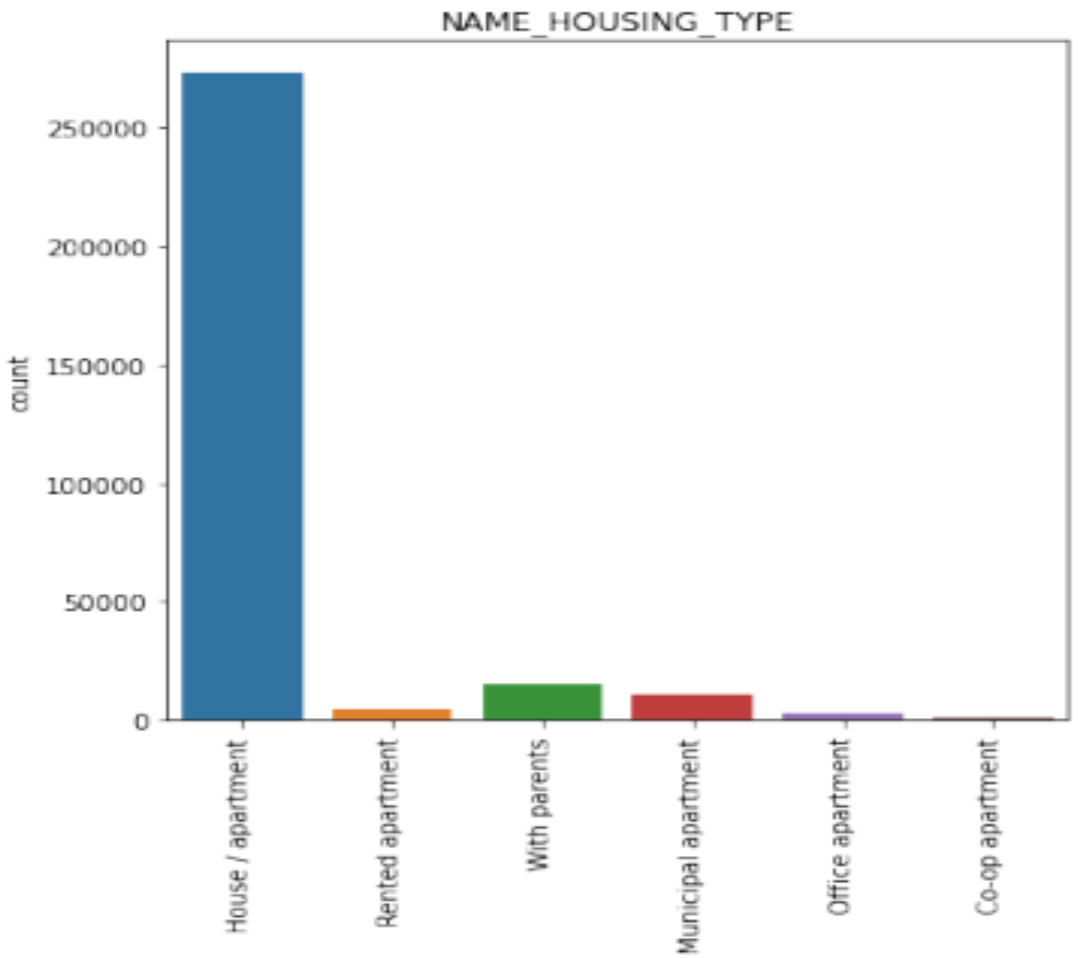
- 1ST graph is showing that the more borrowers are having own residence.
- 2nd graph is showing that most of borrowers are working personals. Pensioners are also good in numbers.

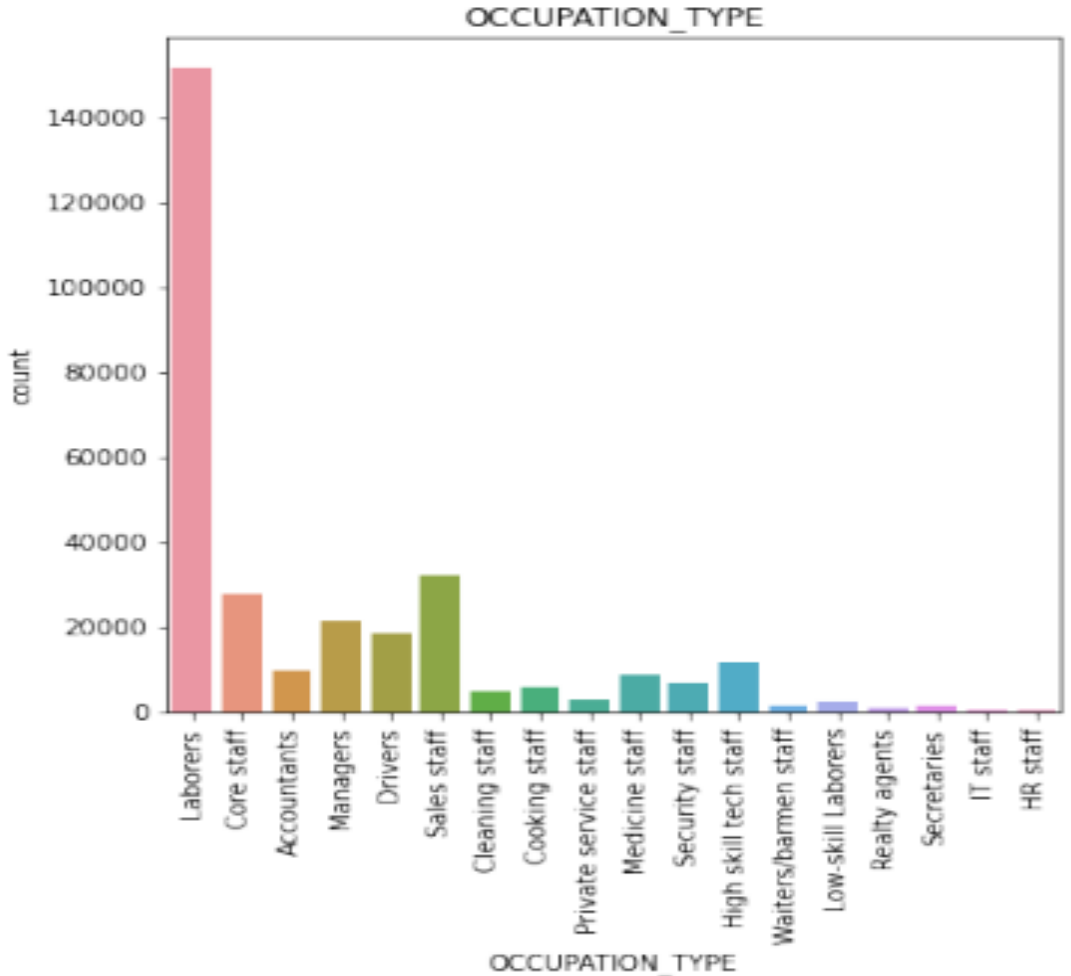


- This count plots implies that most of the borrowers are having education up to secondary level . Highly educated borrowers are almost 1/3 rd of borrowers with secondary education.



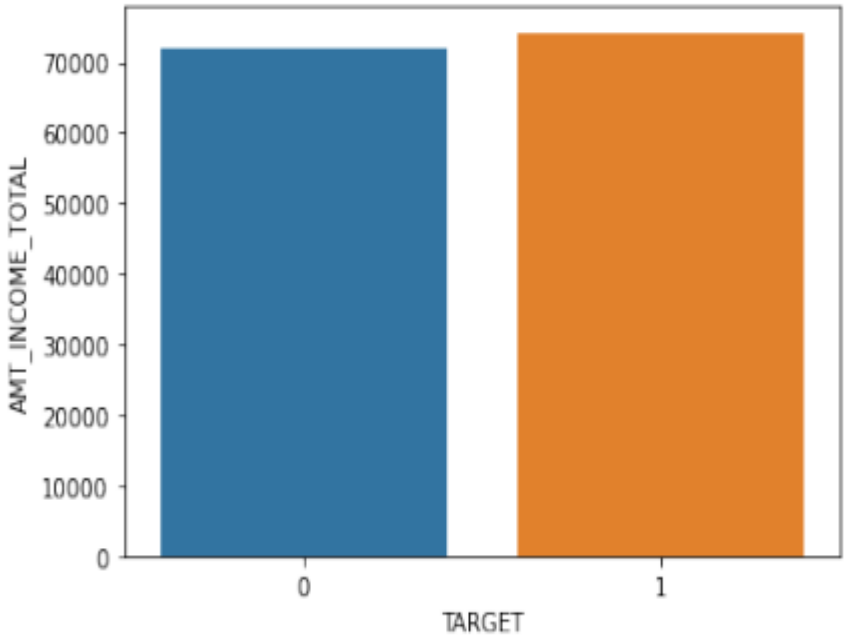
- Married persons are more likely to take loan and widows have less tendency to take loans .

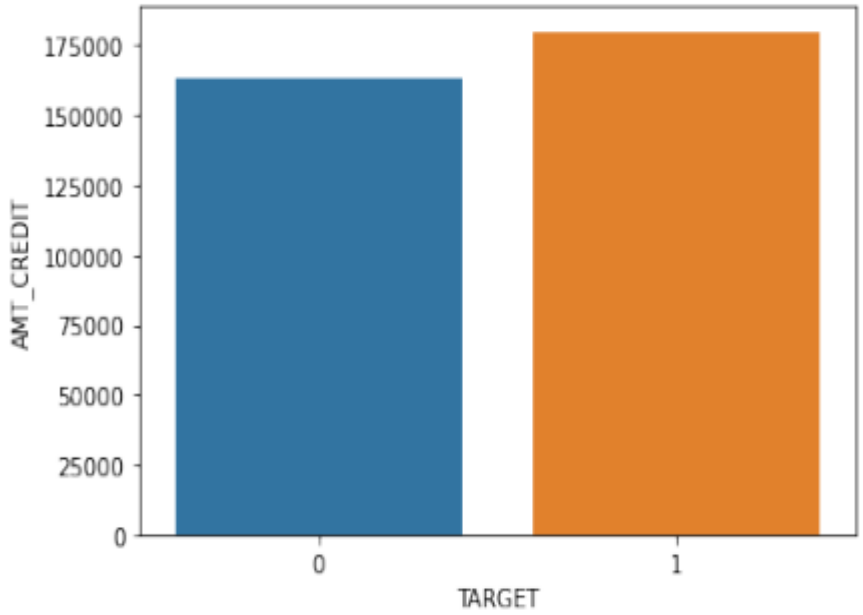




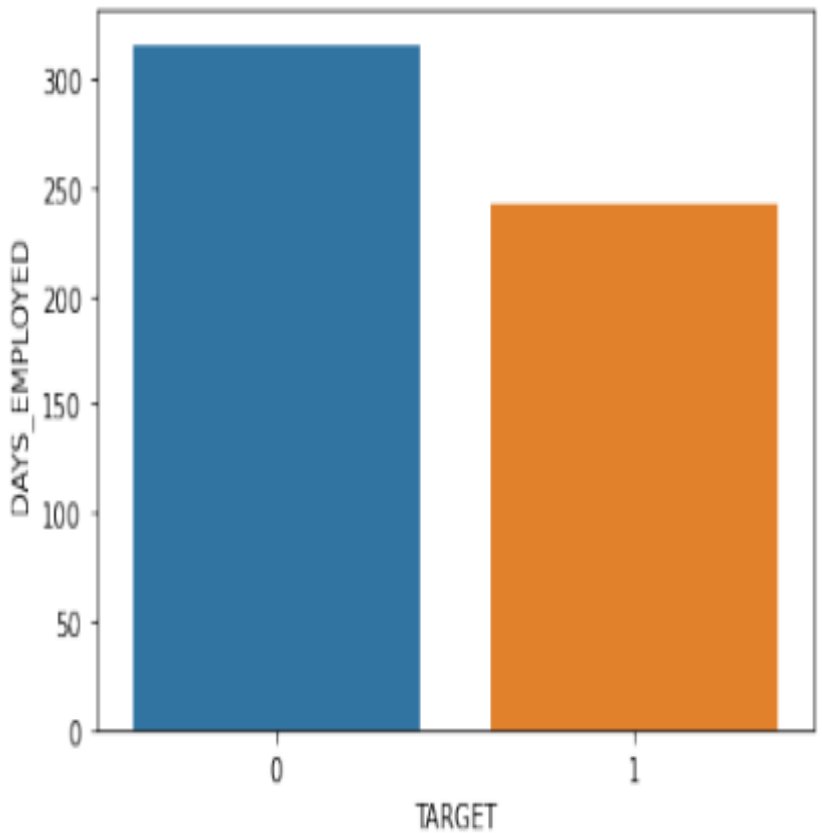
- 1st graph implies that most of the borrowers have own house and number of borrowers with rented one is nominal.
- 2nd graph implies that labour community is highest loan borrower. Managers, accountant, driver, core staff also take loan .

Bivariate Analysis

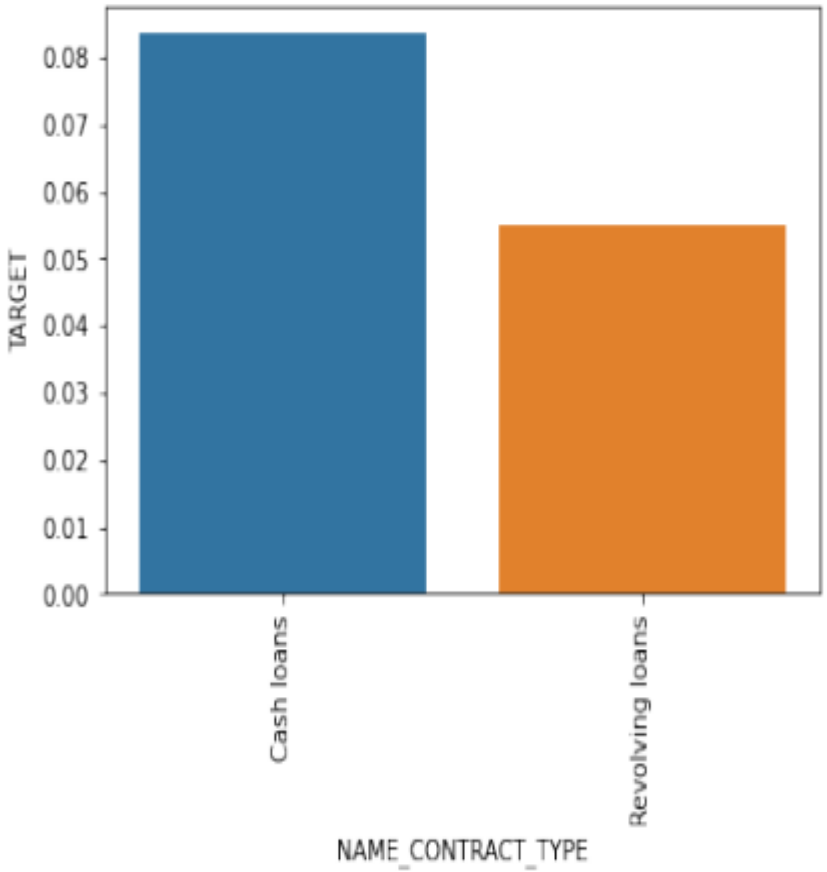


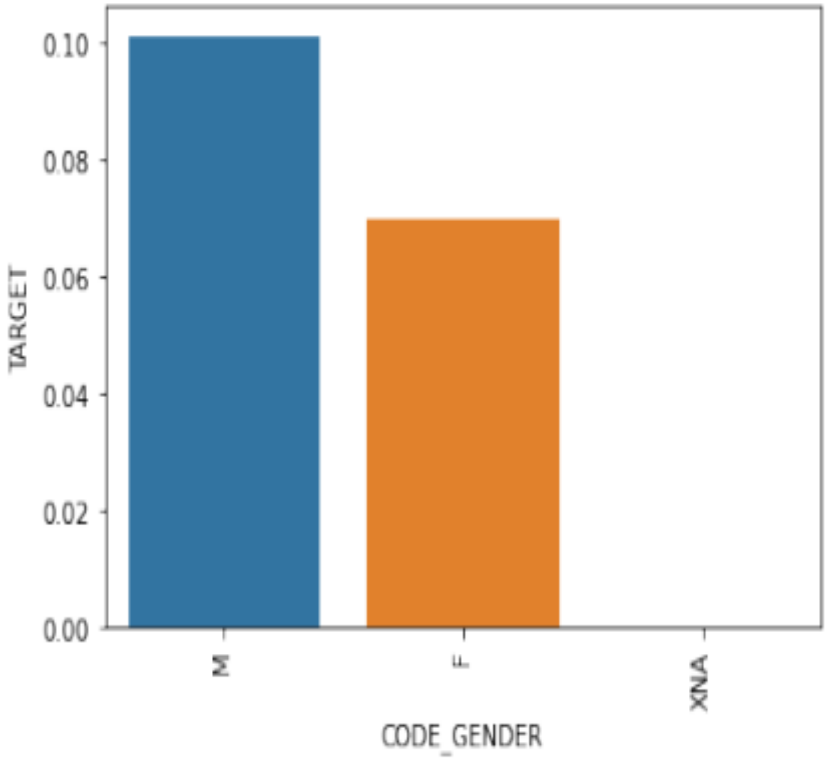


- 1st bar plot implies that 'Defaulter' may be income holder of any amount (high or low) but number of defaulter is higher than non defaulter with rising annual income.
- 2nd plot implies that more credit loan increase the chance of being defaulter.

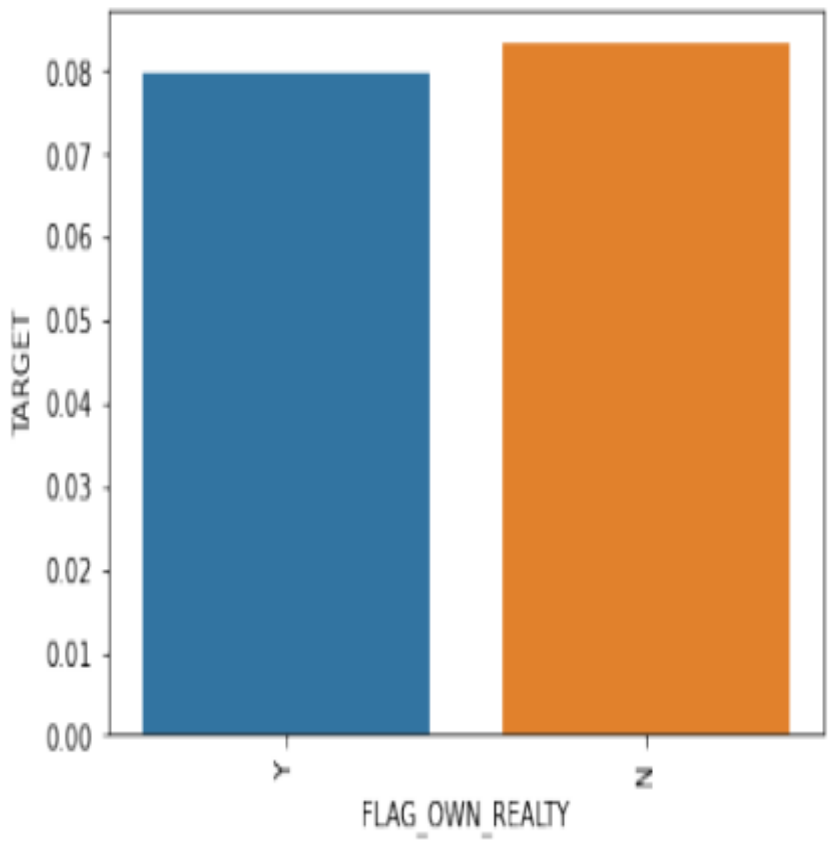


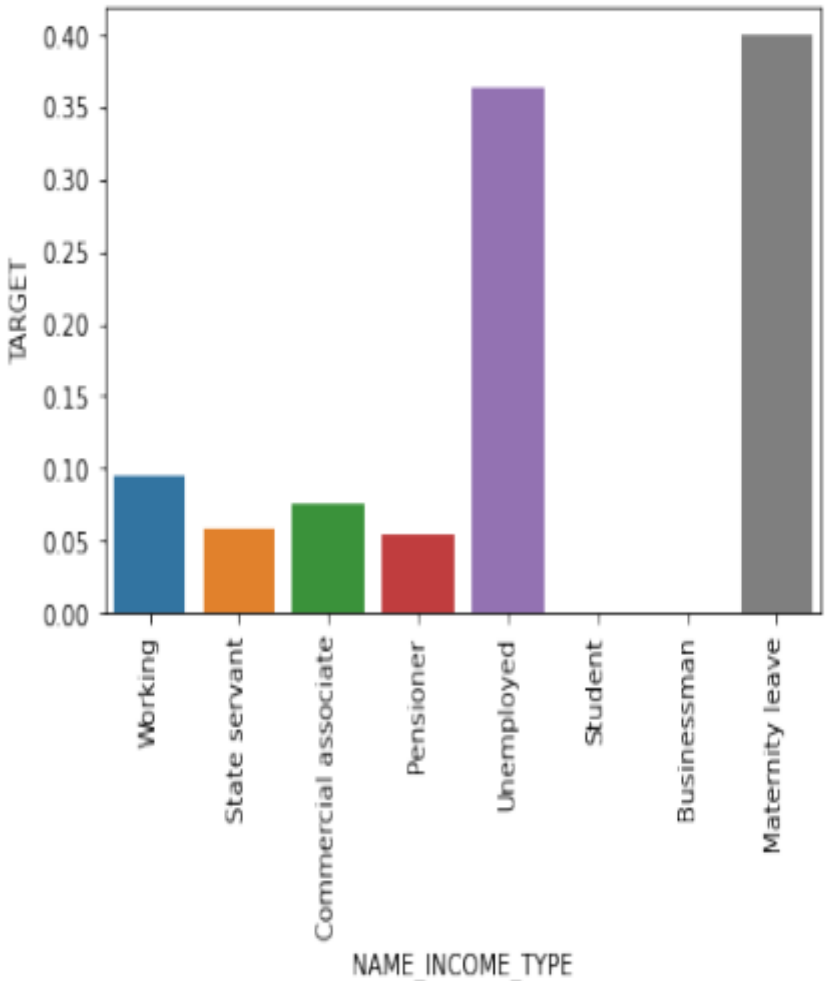
- This graph implies that borrowers who have more job life are the less likely to be defaulter.



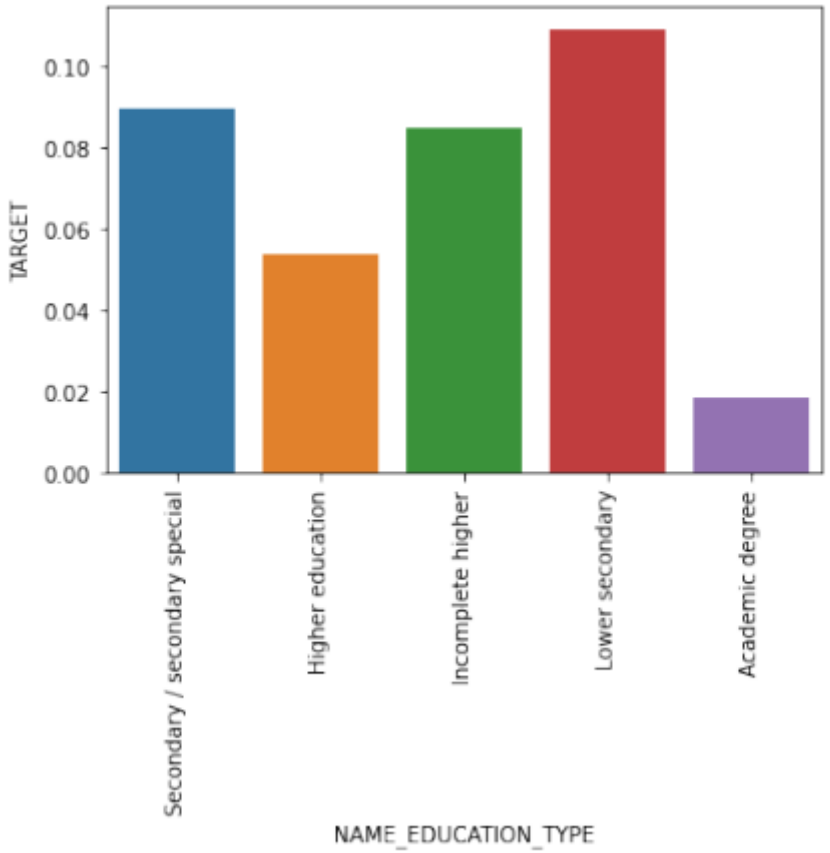


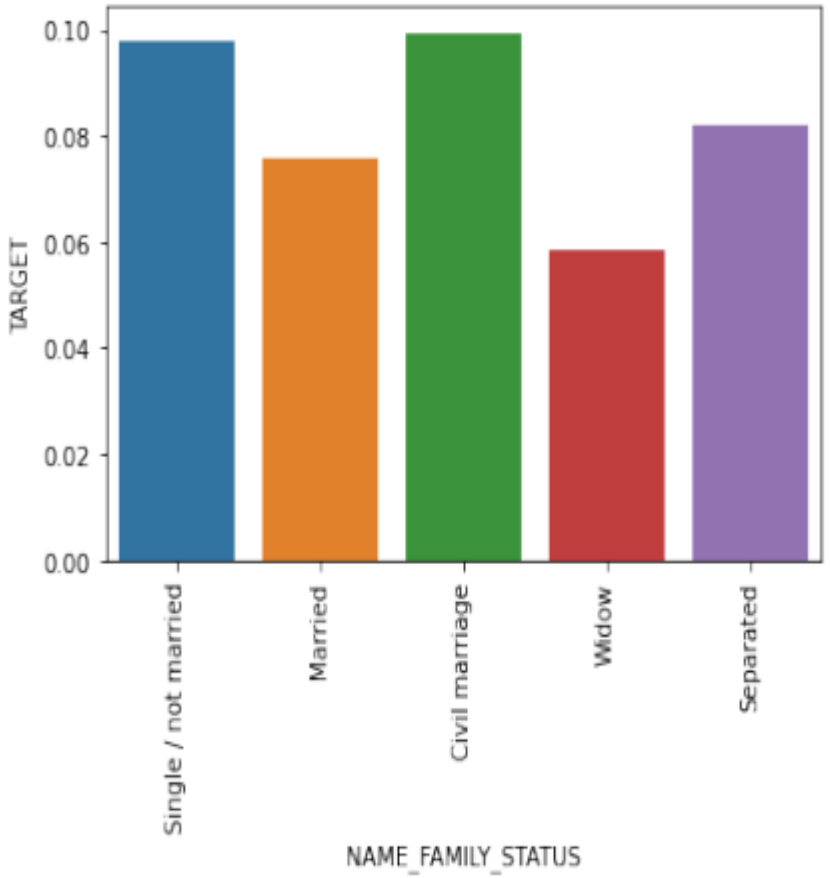
- 1st graph implies that those who take cash loan are more likely to defaulter than the revolving loan borrowers.
- 2nd graph implies that The male clients are more likely to have difficulties to pay off loan than female clients.



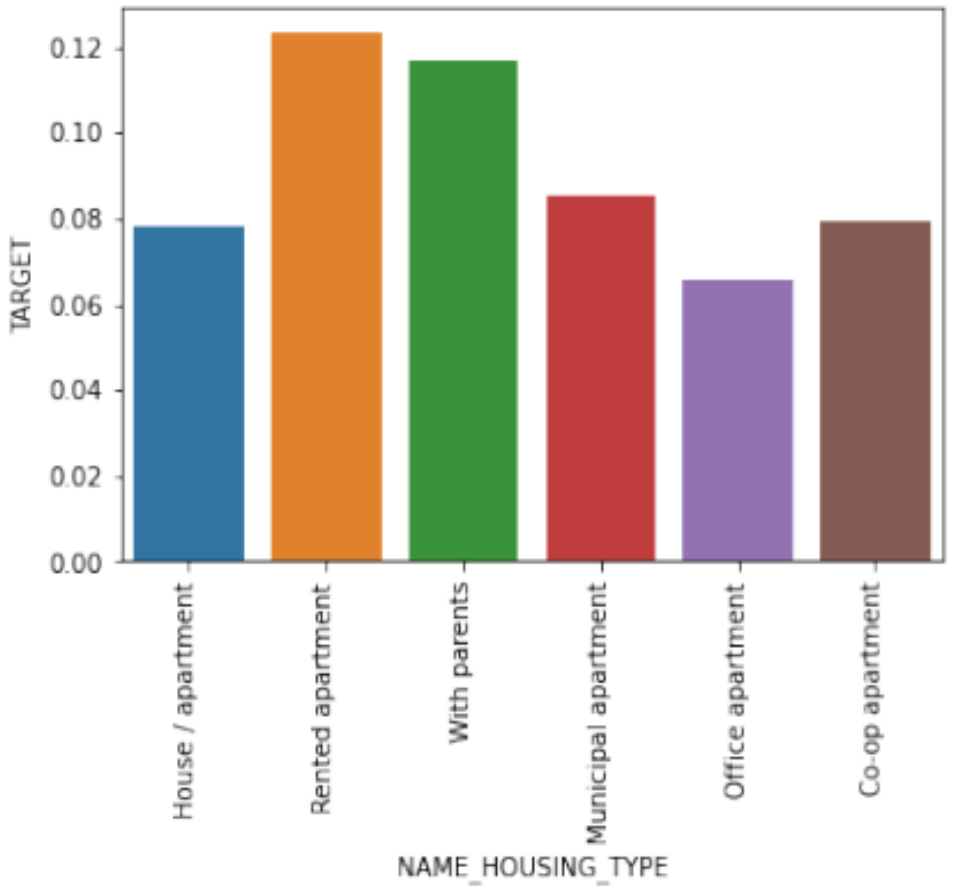


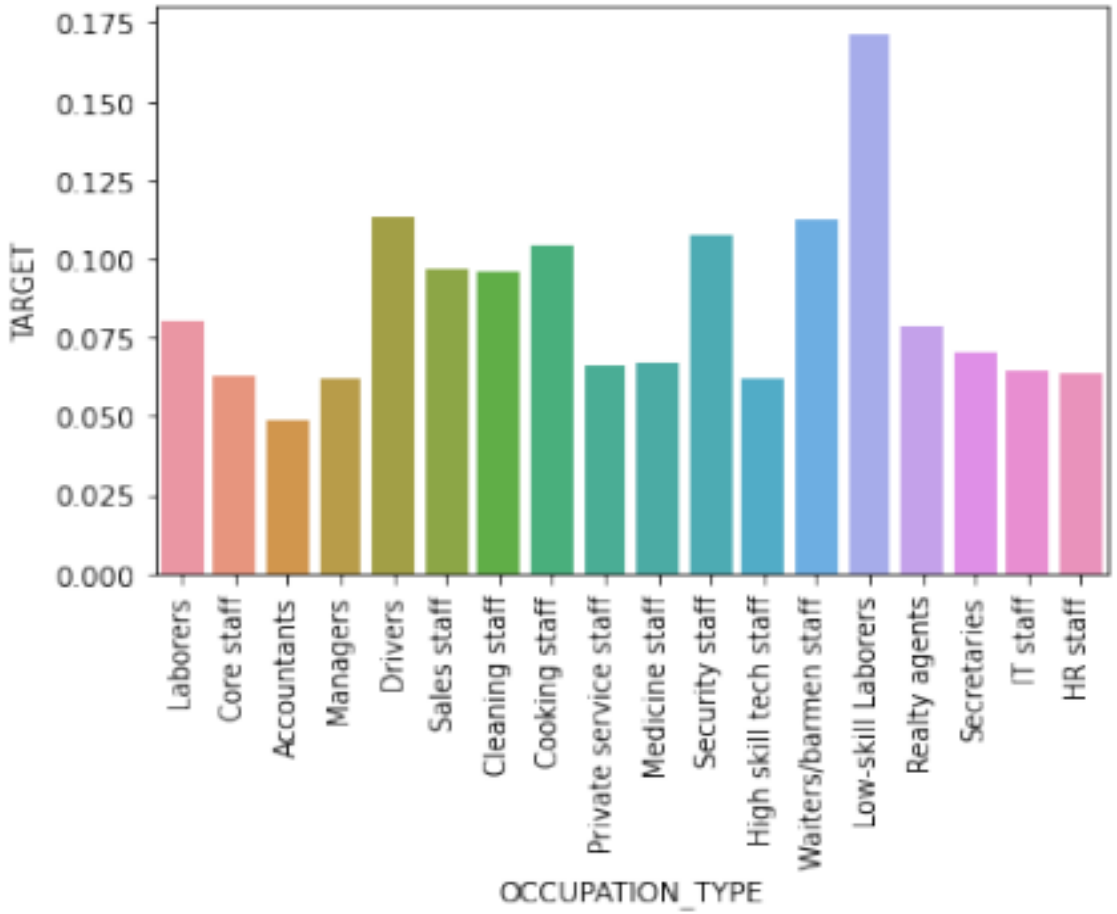
- 1st plot implies that clients not having own residence are more likely to be defaulter than those who have the same but both type of clients may be defaulter.
- 2nd plot shows that working persons have less chance to be defaulter than unemployed persons and female clients in maternity leave .



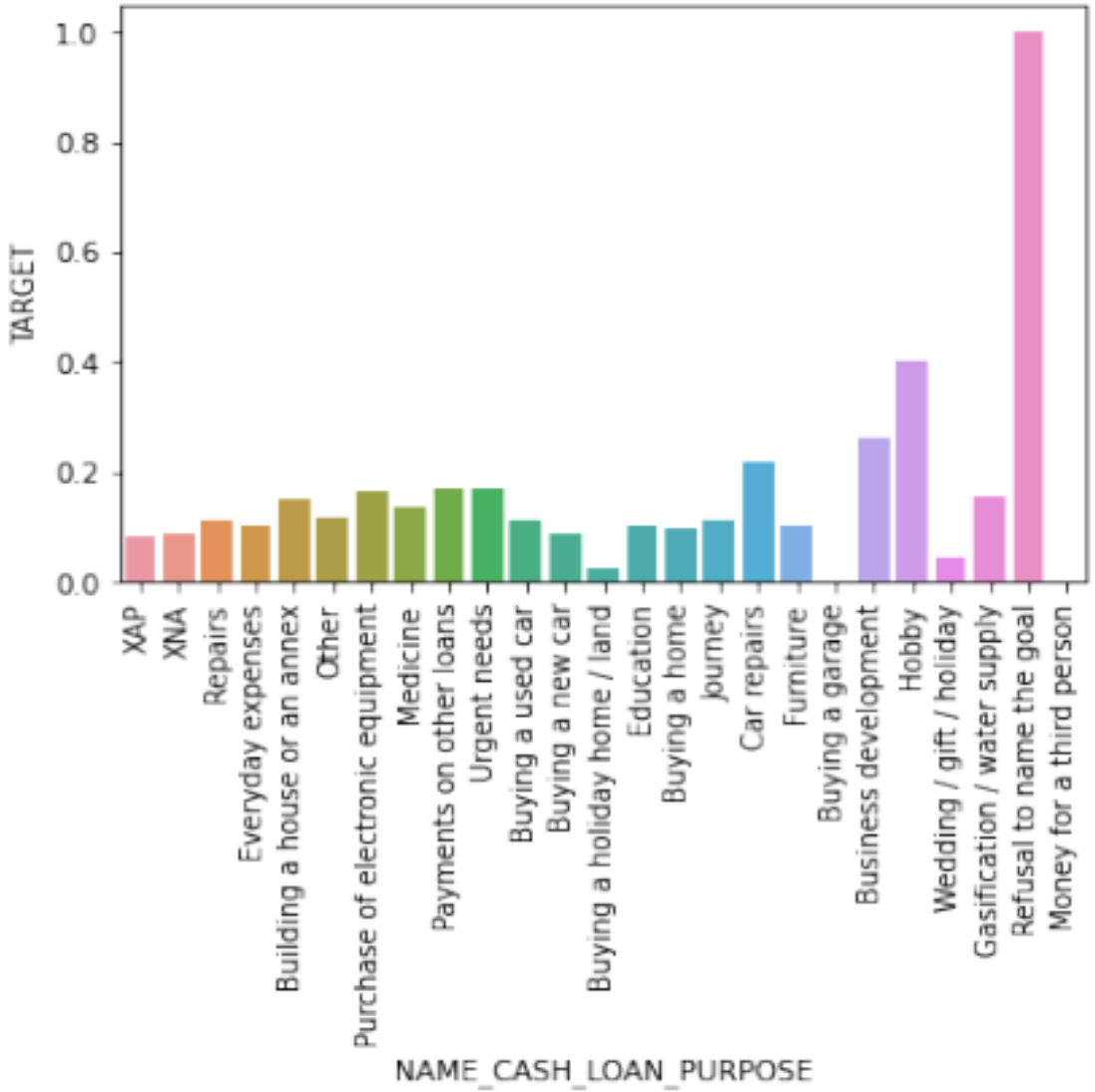


- 1St plot shows that Highly educated people are less likely to be defaulter than people with secondary and people drop up higher education . People with lower secondary education are the most likely to be defaulter.
- 2nd plot shows that single, person having civil marriage and separated people are the most likely to be defaulter than married and widow persons.



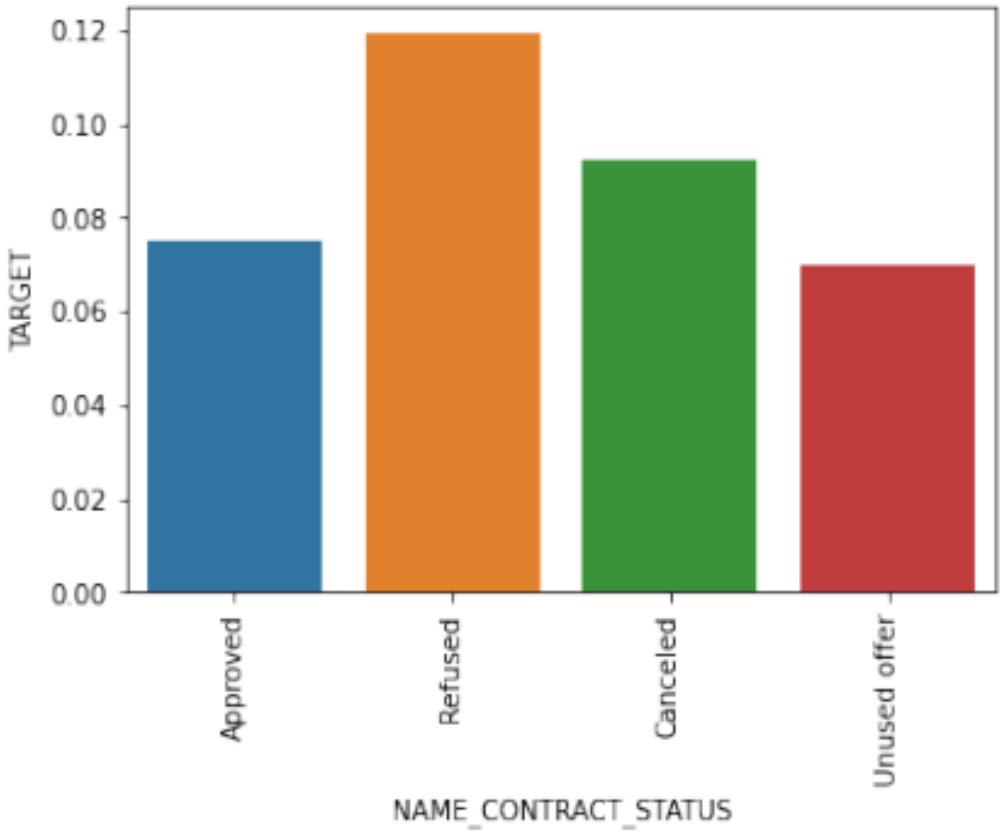


- 1st plot implies that clients who are owner of house and don't have dependent parents with them are less likely to be defaulter.
- 2nd plot shows that there are high chance of loan paying difficulties in case of lower skilled labours and other low paid persons like drivers, security staff, water barmen, cleaning staff, sales staff where managers, accountants, medicine,high skill tech staff have less chance to be defaulter .

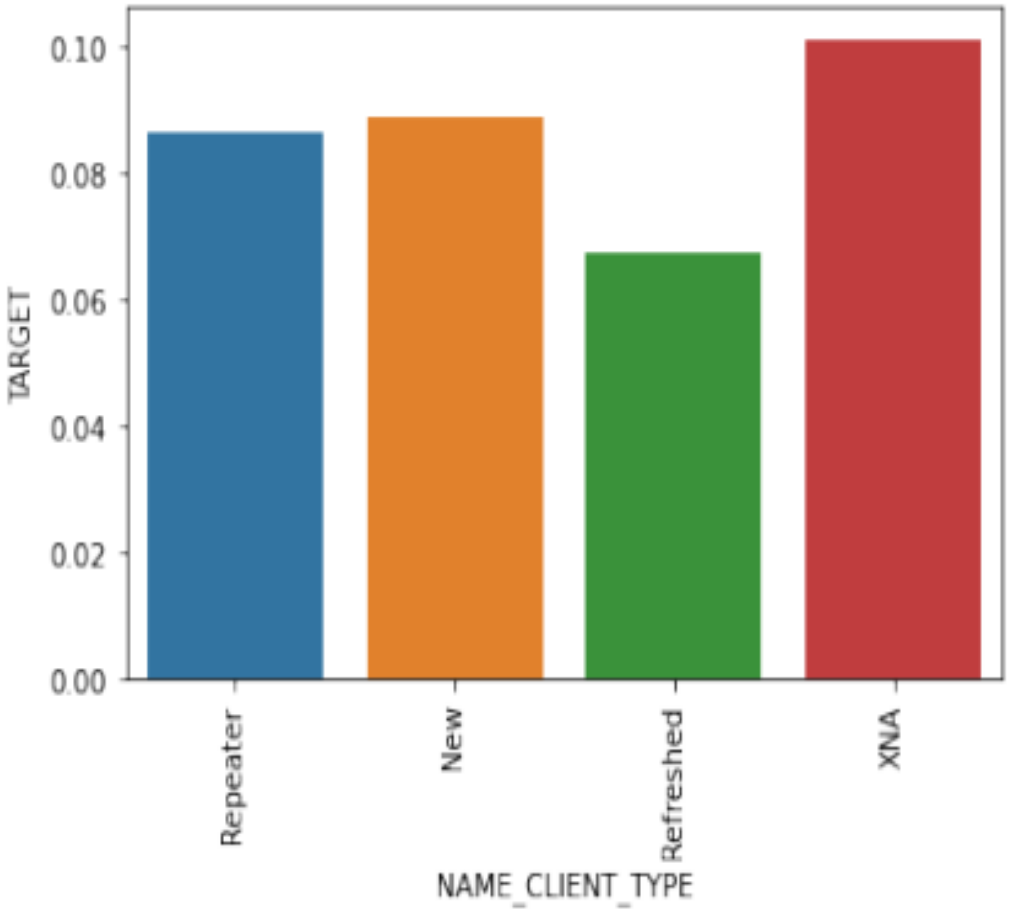


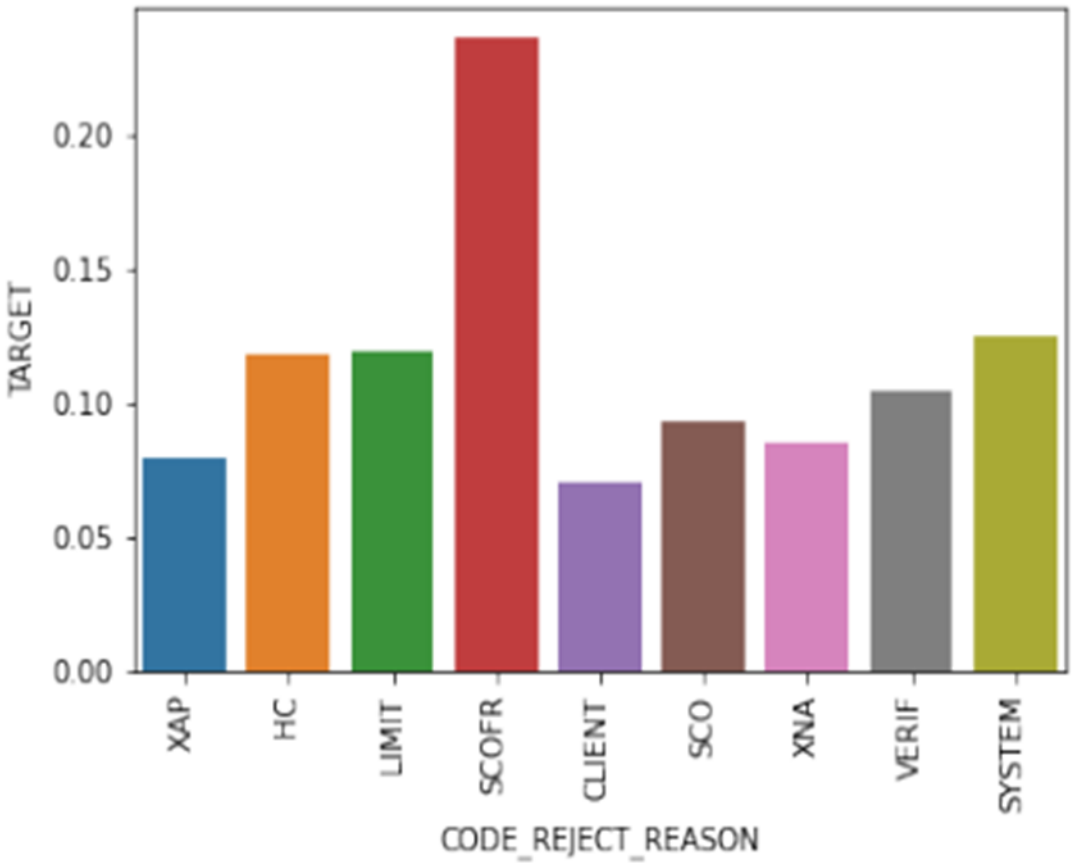
- Borrowers with anonymous reason of taking loan or borrowers don’t want to reveal the purpose are the most likely to be defaulter followed by the borrowers take loan for fulfillment of their hobbies, business development, paying other loans, urgent purpose, buying something etc.

- Those who take loan for marriage purpose are the least likely to be defaulter. Purpose like repairing, medicine, everyday expenses, car purchase etc. are most likely to pay off the loan.



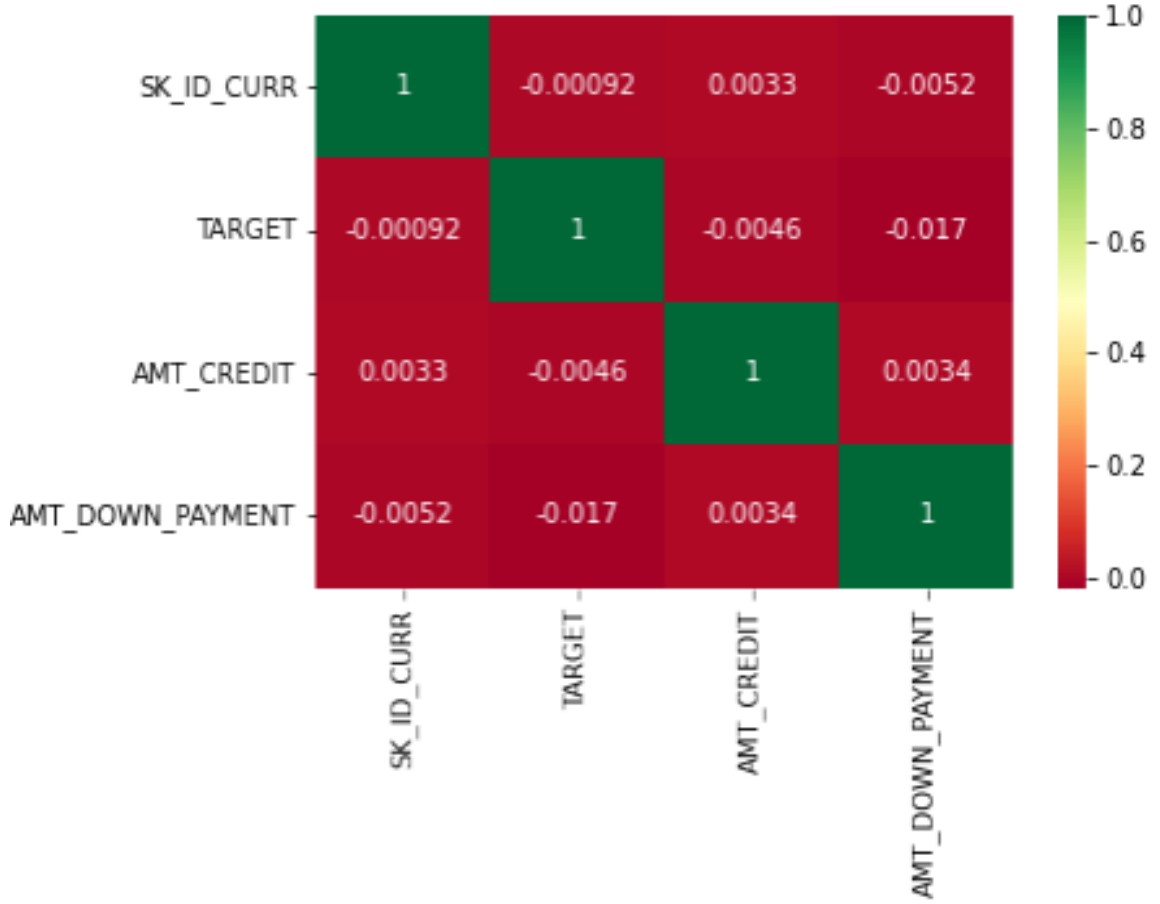
- 1st graph plotted based on the borrower's previous records regarding loan application shows that the borrowers who were previously rejected are the highest defaulter followed by those application were cancelled in previous case. Borrowers those who got approval in previous case are less likely to be defaulter this time also.





- 1st graph implies that there are less chance of non paying loan in case of refreshed borrowers and with code of rejection ‘xna’ followed by new borrowers have more likely to be defaulter. Repeaters are less defaulter comparison to new.
- The borrowers having rejection reason code ‘SCOFR’ in previous case are the highest defaulter again followed by the codes ‘HC’, ‘LIMIT’ etc.

Multivariate Analysis :



- These heat maps are showing that Total income, Amount of down payment, Days_employed are the parameters which have –ve linear correlation with Target variable that means as much as value of these parameter increase the chance of default cases decrease.

Conclusions :

- ✓ Bank should be more judgmental in case of issuing loan to the person with more liabilities and lesser property .
- ✓ Bank should increase Loan approval for marriage repairing, medicine, everyday expenses, car purchase purpose.
- ✓ Bank should avoid previously refused customers.
- ✓ Bank should not approve loan to such person with very low income.
- ✓ Bank can increase loan approval to Female customers , pensioners, widows, state service holder, commercial associates.