Loan Default Risk - EDA Case Study 1

Prepared By,

Harsh Patel

Business Problem

- Consumer finance company wants to identify loan applicants capable of repaying.
- Avoid rejecting responsible applicants and approving applicants likely to default.
- Use EDA to analyze patterns in data to minimize risk.

Information Summary

- Data contains loan application particulars
- Two circumstances:
  - Clients with payment troubles (late X days on Y installments)
  - Clients who remitted on time
- Four application conclusions;
  - Approved
  - Cancelled
  - Refused
 - Unused Proposition

Business Objectives

- Identify patterns indicating if a client will have difficulty paying installments

- Use patterns to take actions like:

  - Denying loan

  - Reducing loan amount

  - Lending at higher interest rate

- Minimize risk of defaults

EDA Approach

- Analyze correlation between attributes.
- Study distribution of key attributes.
- Identify trends across segments.
- Highlight significant variables influencing tendency to default.

Distribution of Annuity Amount

It can be observed from the histogram that annuity amount data is right skewed and maximum value is approximately 175,000.
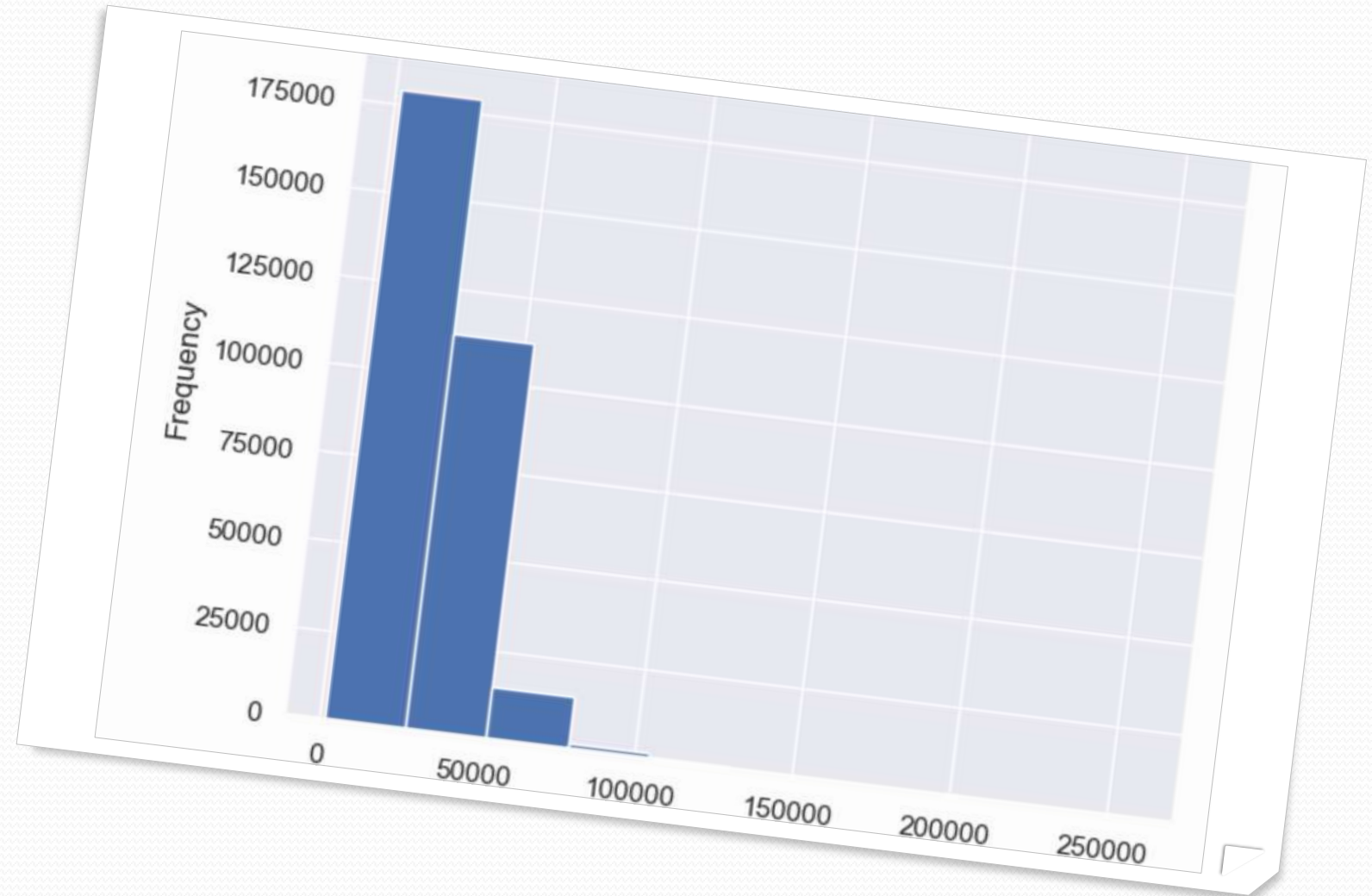
Plotting for distribution analysis

In this count plot, the data is categorized for OCCUPATION_TYPE. It can only be imputed with mode values. Since the column has 31% of missing values, it cannot be replaced with only Laborers. It will be unfair for other customers' occupation type. Thus, I will leave it unchanged.
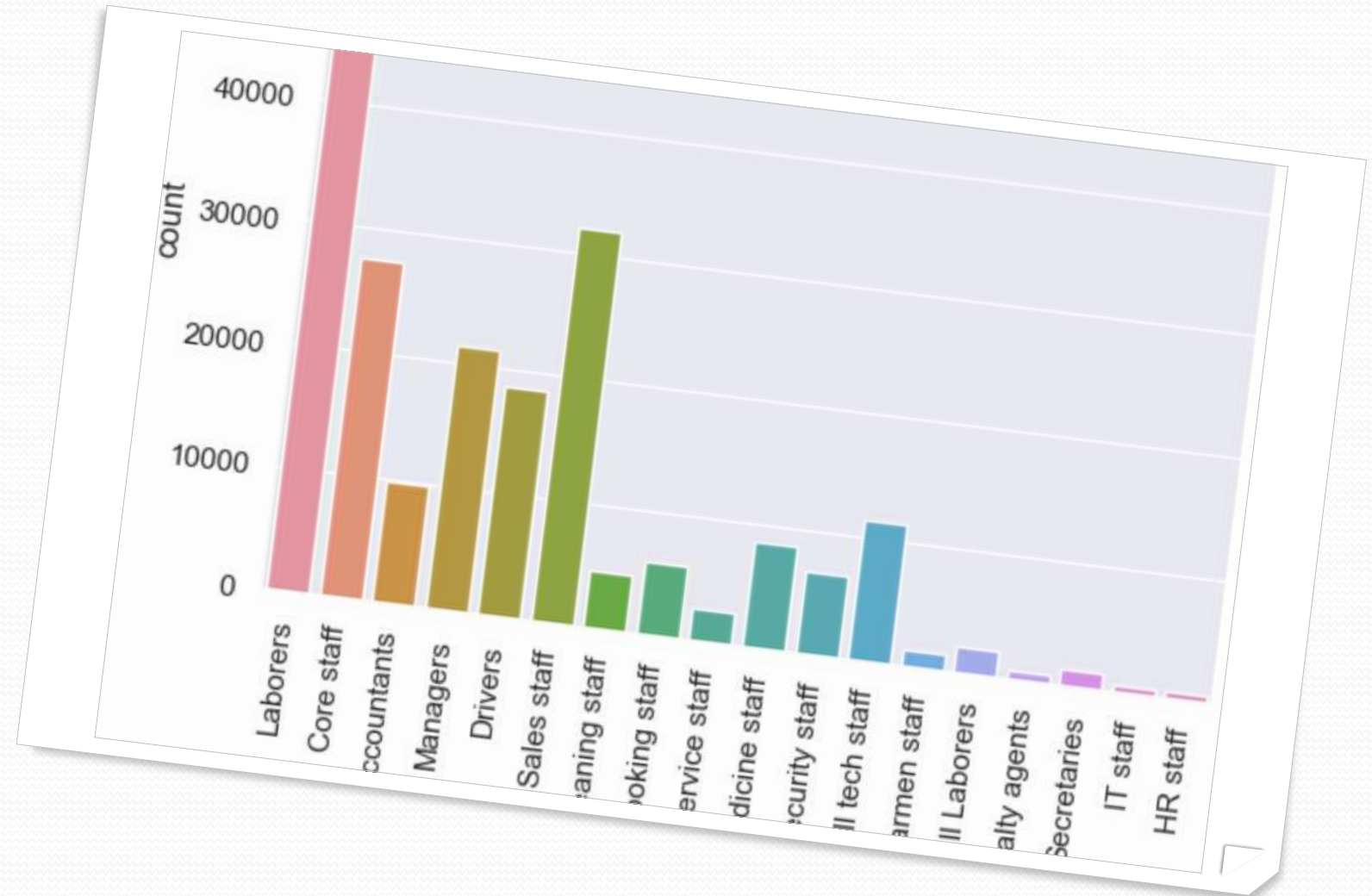
EXT_SOURCE_3 Histogram

The data has mean and median very close to each other. Thus, it can be imputed with anyone of them. The histogram shows, the data has skewed normal distribution.
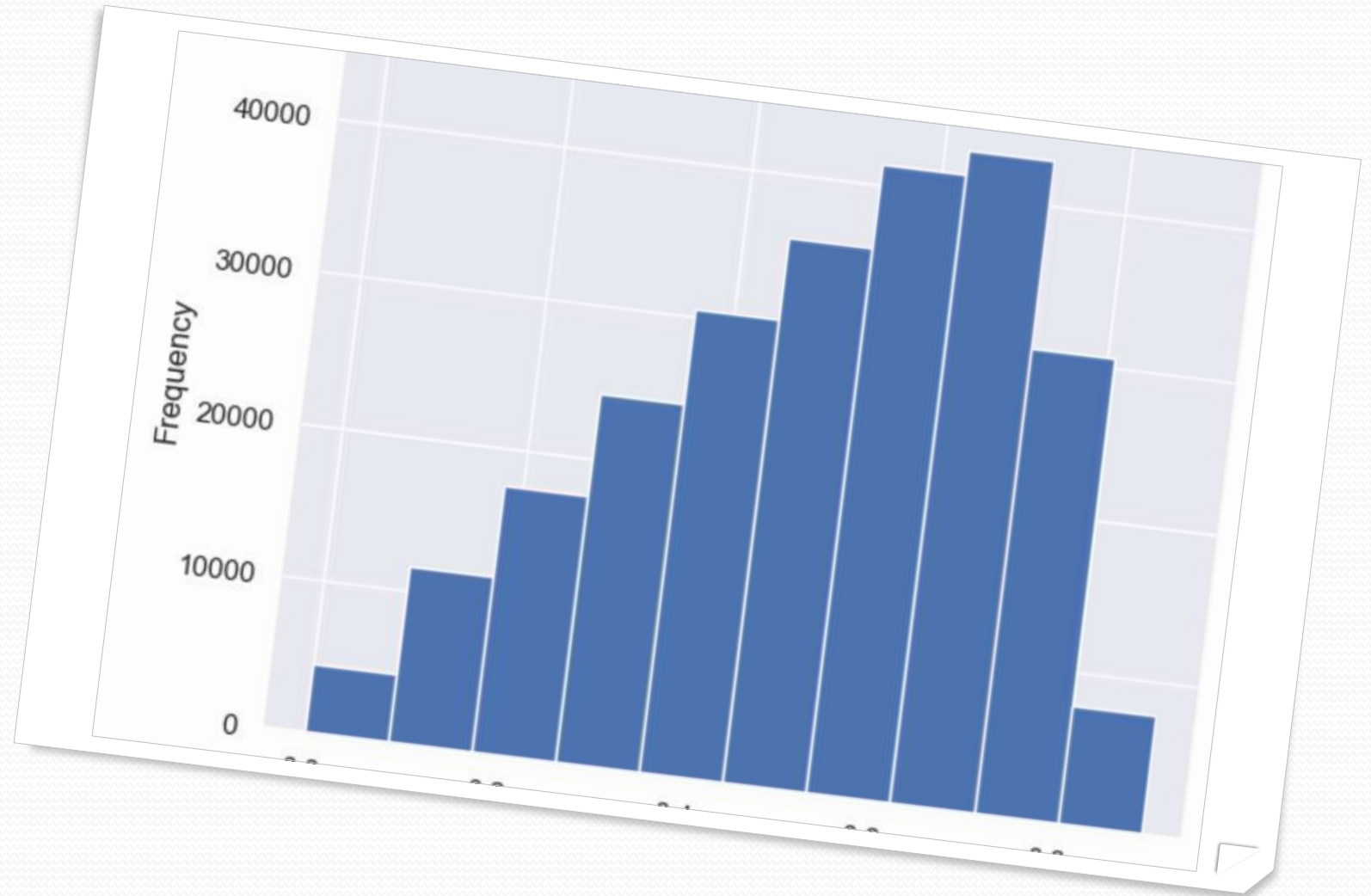
Histogram of AMT_REQ_CREDIT_BUREAU_YEAR

- For this column, the data is distributed between zero(0) to 9.
- The data is distributed between different bins. the maximum distribution is between 0, 1,2, and 3.
- The mean and median is 1.8 and 1 respectively. But it cannot be replaced as the median value is a decimal value.
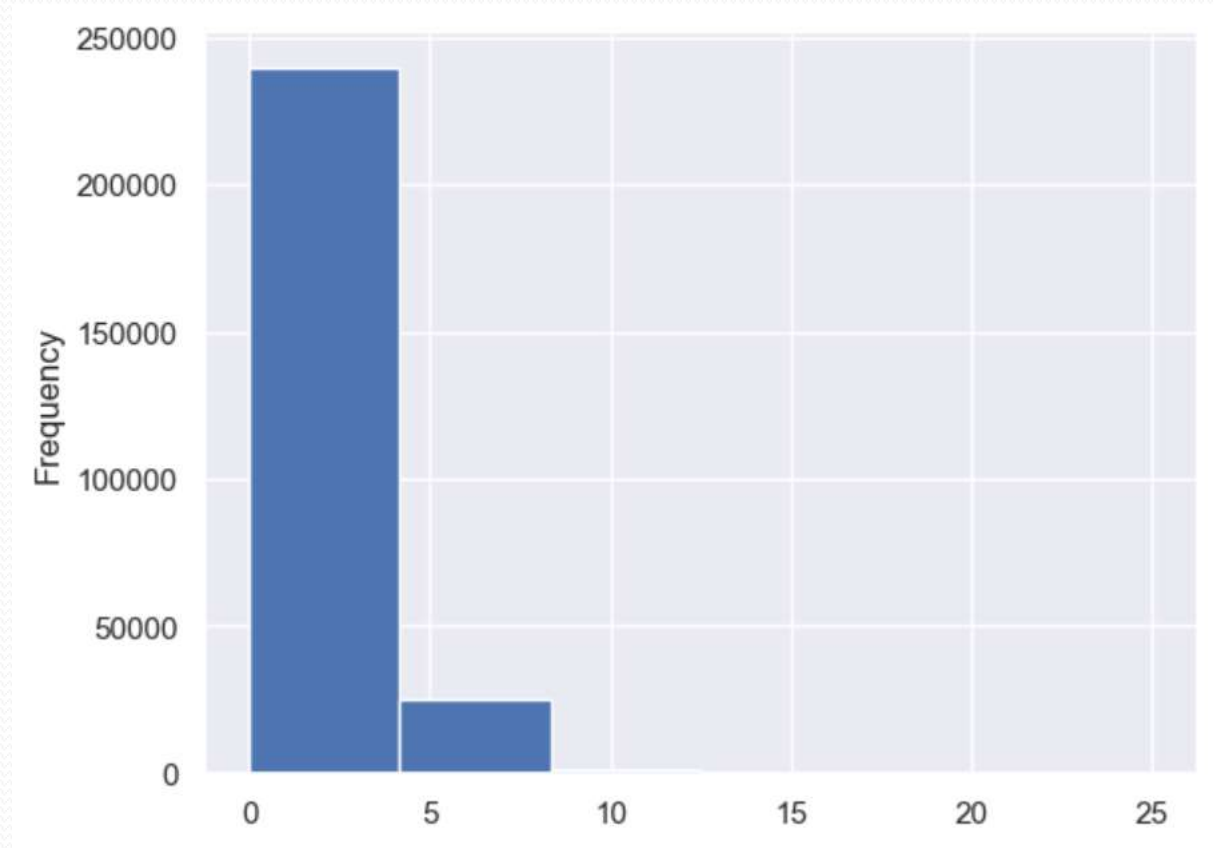
CODE_GENDER Count plot

Here, we can observe that Females are major loan takers in baking sector.

DAYS_BIRTH Histogram

This graph will look a little messy but one can observe that most of the loan takers are between the age of 28-45. Thus, the customers of this age range should be taken into consideration while launching a new product or providing a loan.
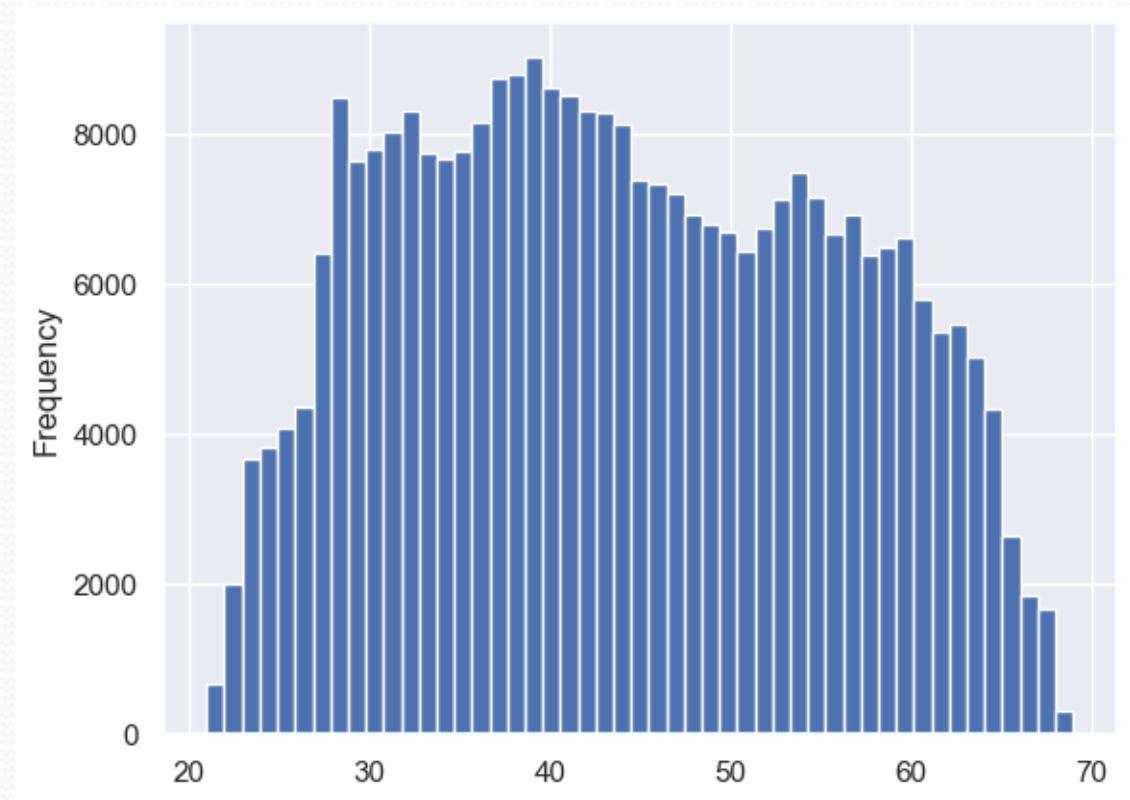
NAME_CONTRACT_TYPE
## Count plot

Here, we can observe that customers are more interested into taking Cash Loans than Revolving Loans.
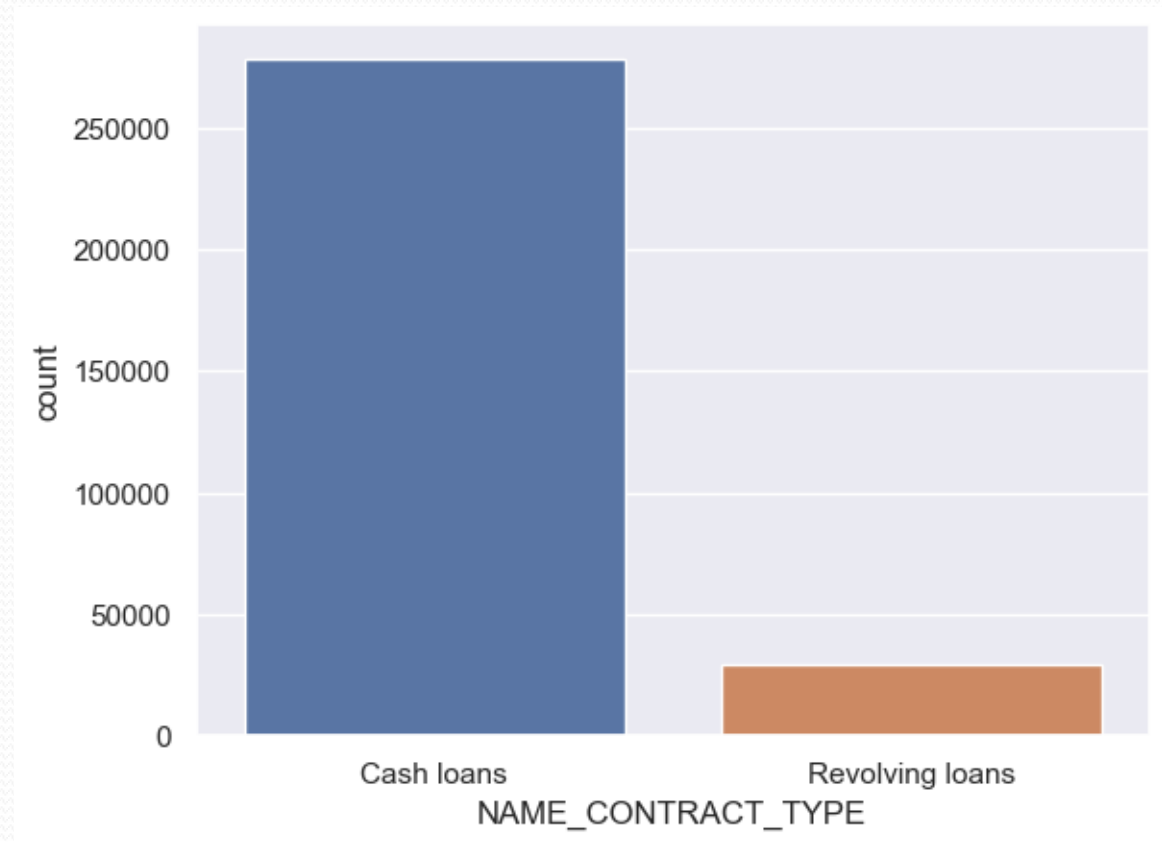
Income Type Distribution

- Working, commercial, pensioners, and state servants have higher income and other. Particularly Female are higher earners.

- From this graph, it can be said that female have more credit than men.

- Other income type people has lower credit. Thus, they are less likely to be future customers.
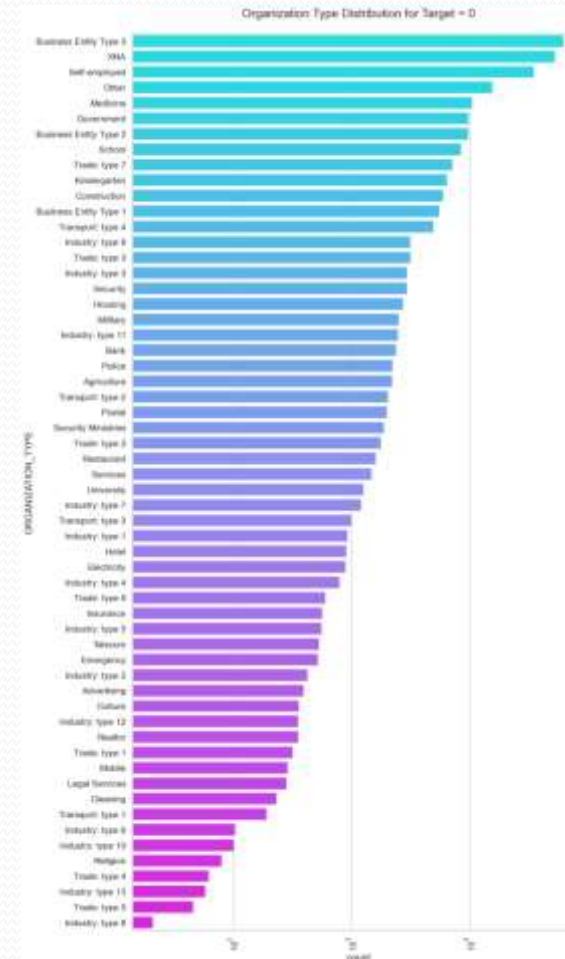

Income Type Distribution
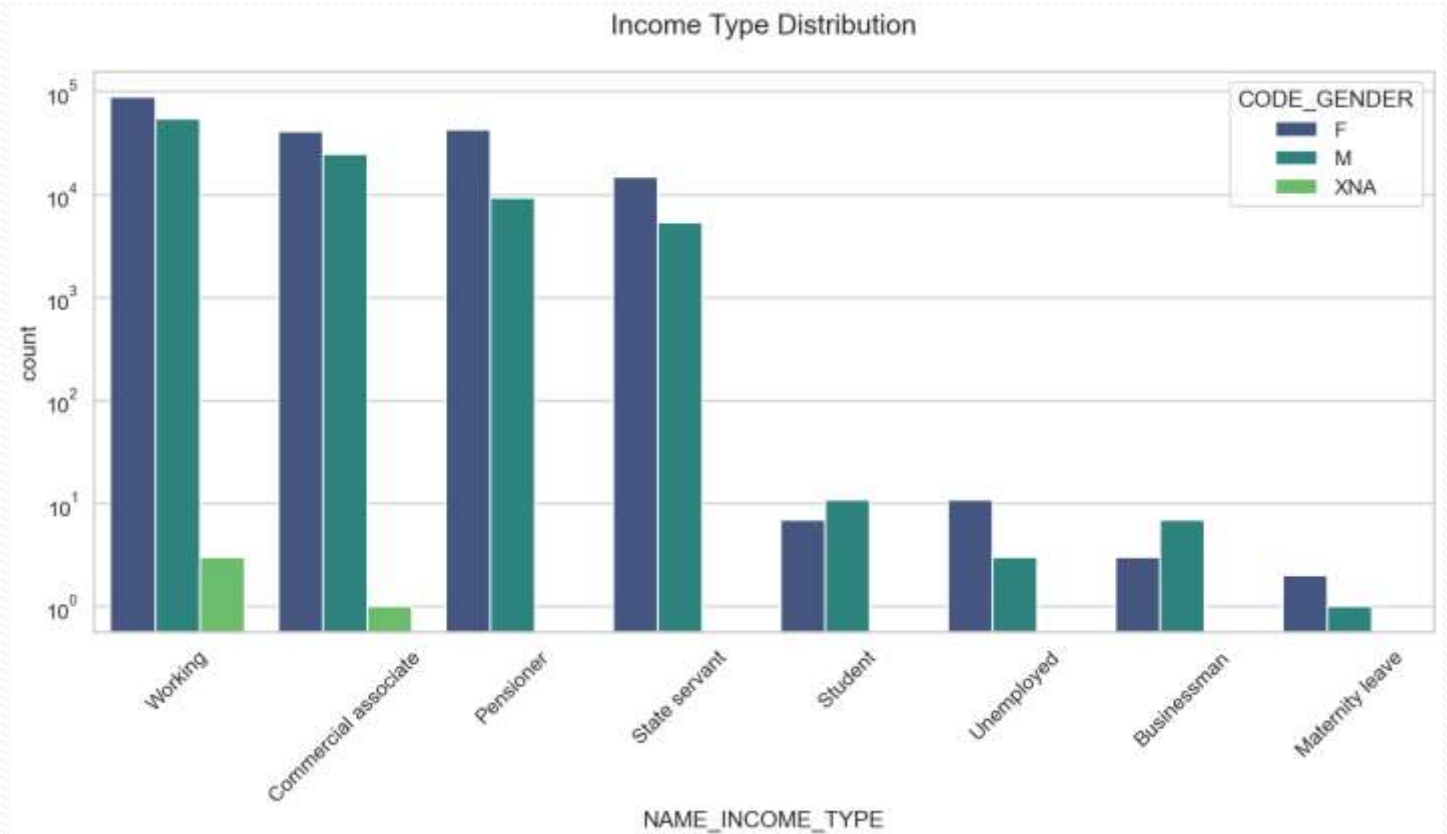
Contract Type Distribution of Non-Defaulters

- Cash loans are taken in higher number than revolving loans.
- Again, female is leading loan applier than male.

Contract Type Distribution
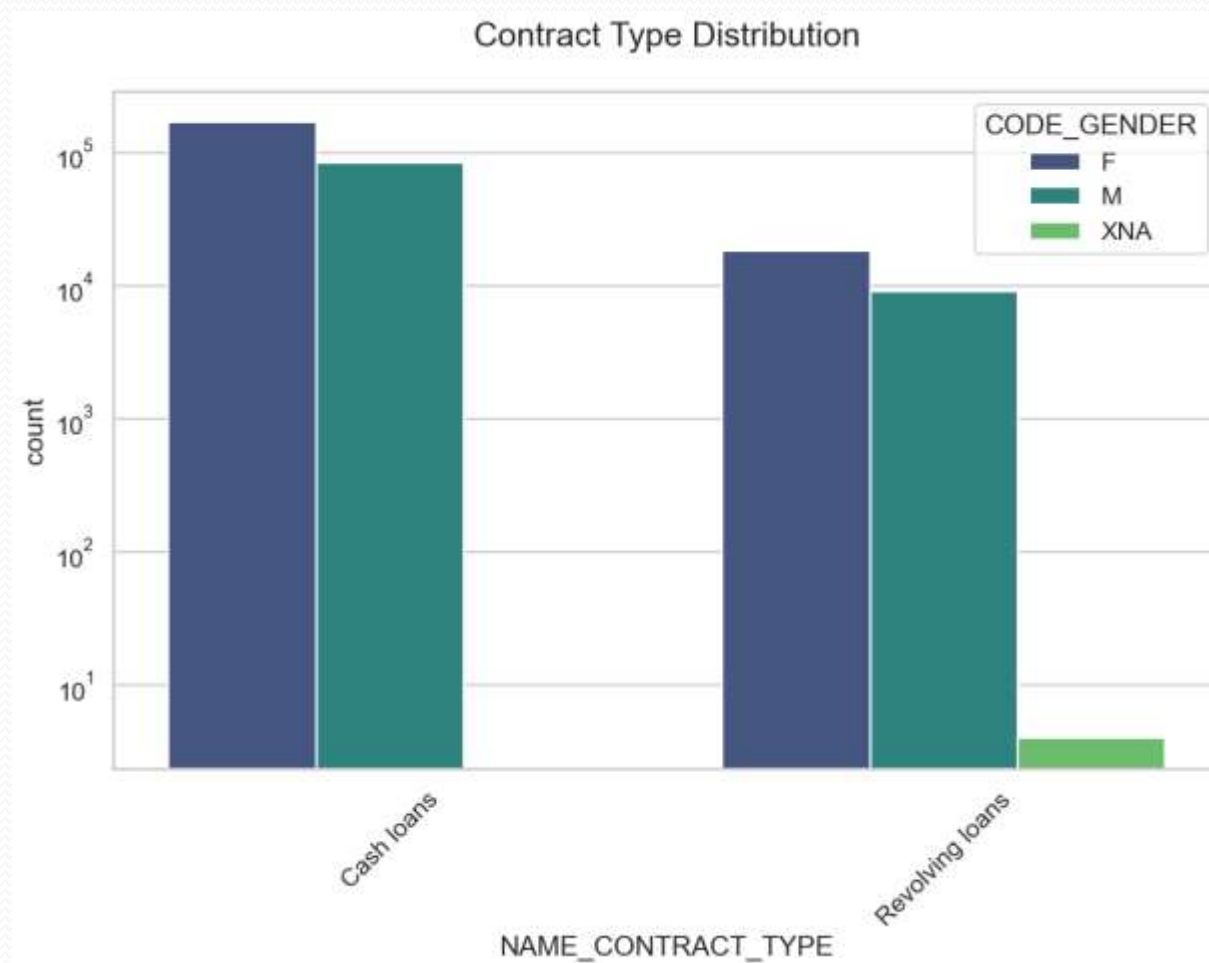
Organization Type Distribution of Non-Defaulters

- Major loan appliers are from Business Entity Type 3, XNA, Self-Employed, Medicine, Government, and Others.

- Least customers are from type 8, type 5, type 13, type 4, religion, type 10, and type 6 job sectors.

Organization Type Distribution for Target = 0
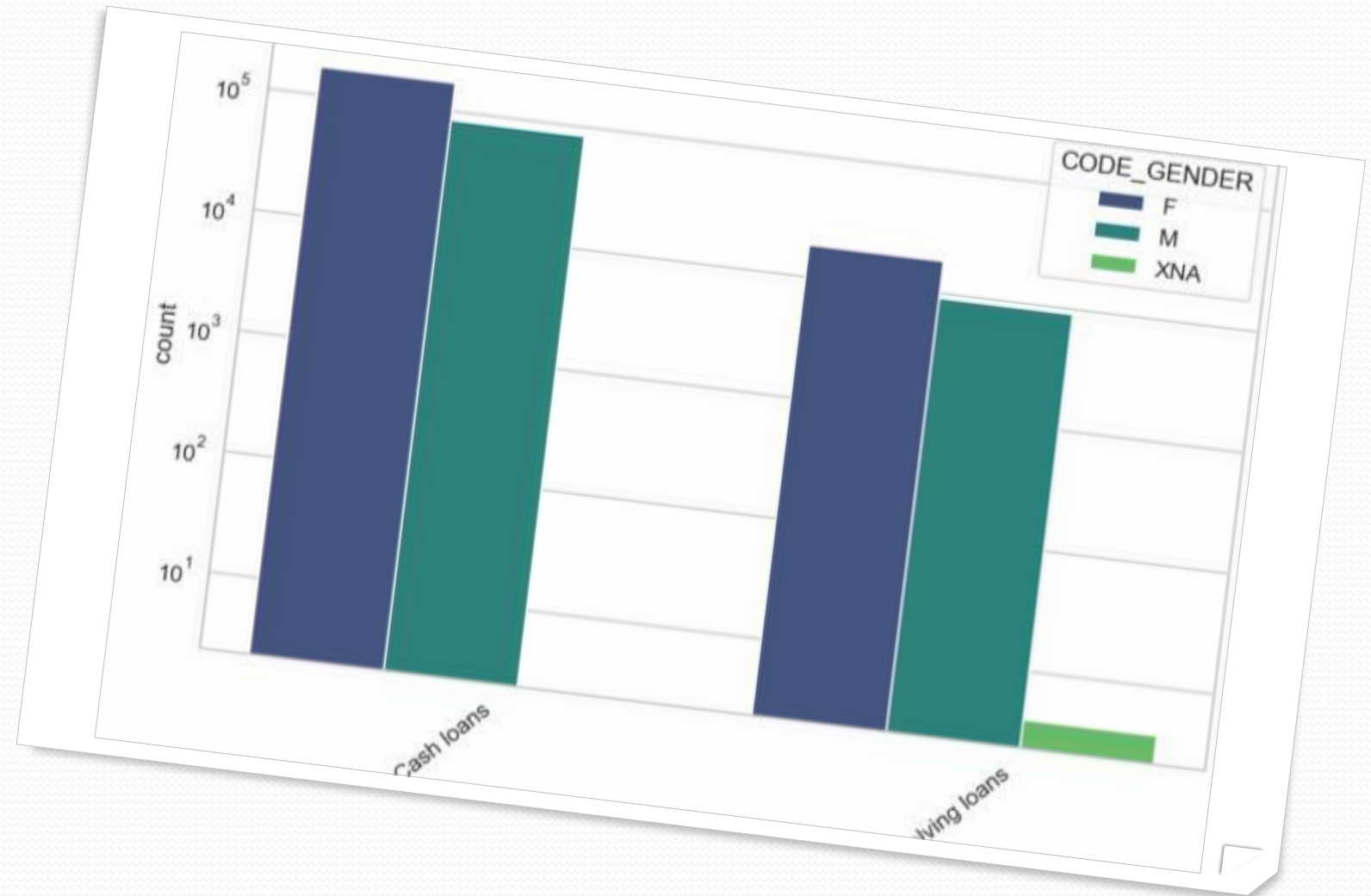
Defaulter Income Type Distribution

Working, Commercial associate, pensioner, and state servant females are more likely to become defaulters.
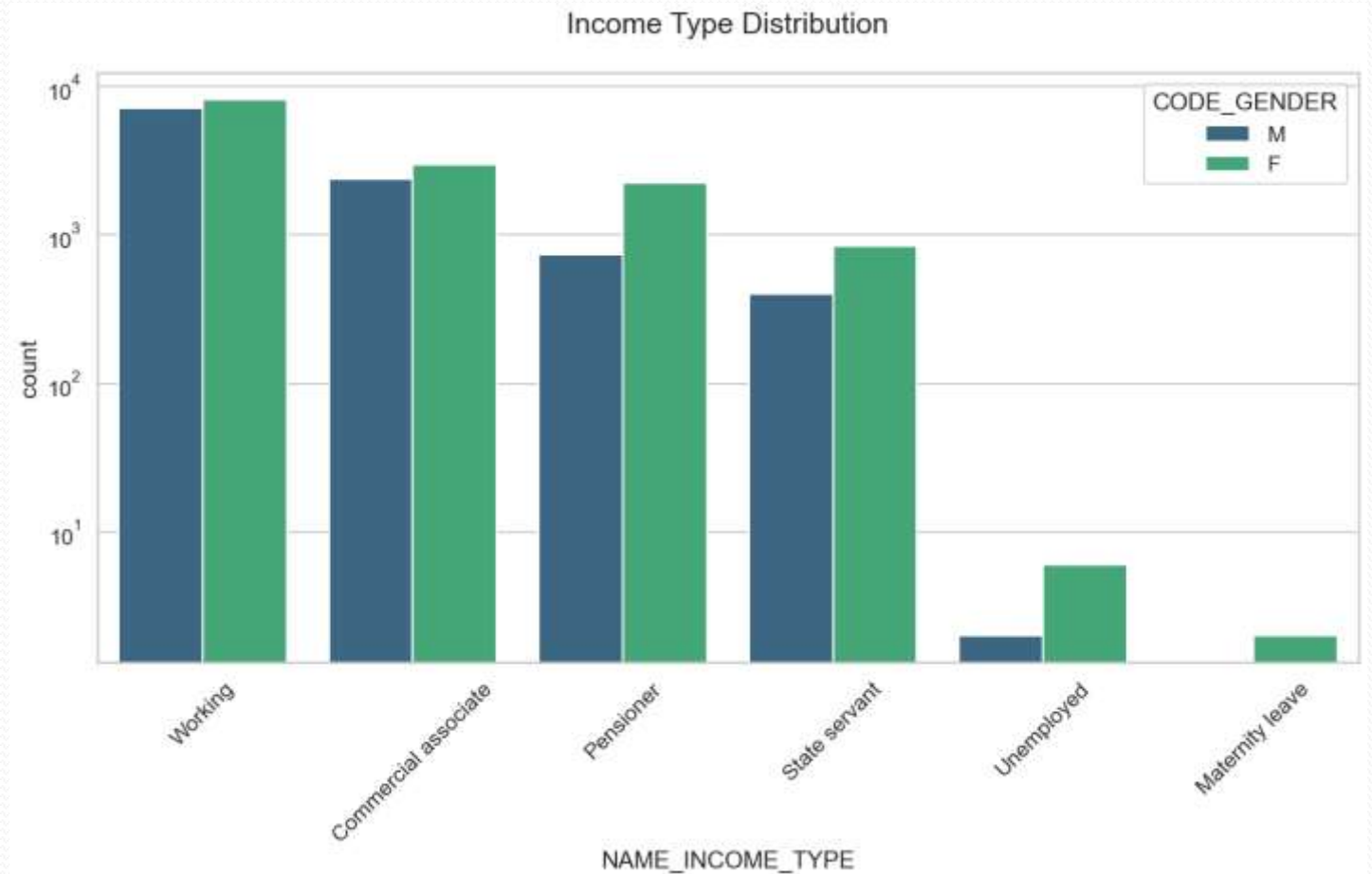
Contract Type Distribution of Defaulters

Here we can see that females are having more difficulty in paying off loan of both types.

Organization Type Distribution of Defaulters

From this graph, it can be observed that Business Entity Type 3, XNA, Self-employed, and other job sectors are more likely to have payment difficulties in the future.

Organization Type Distribution for Target = 1
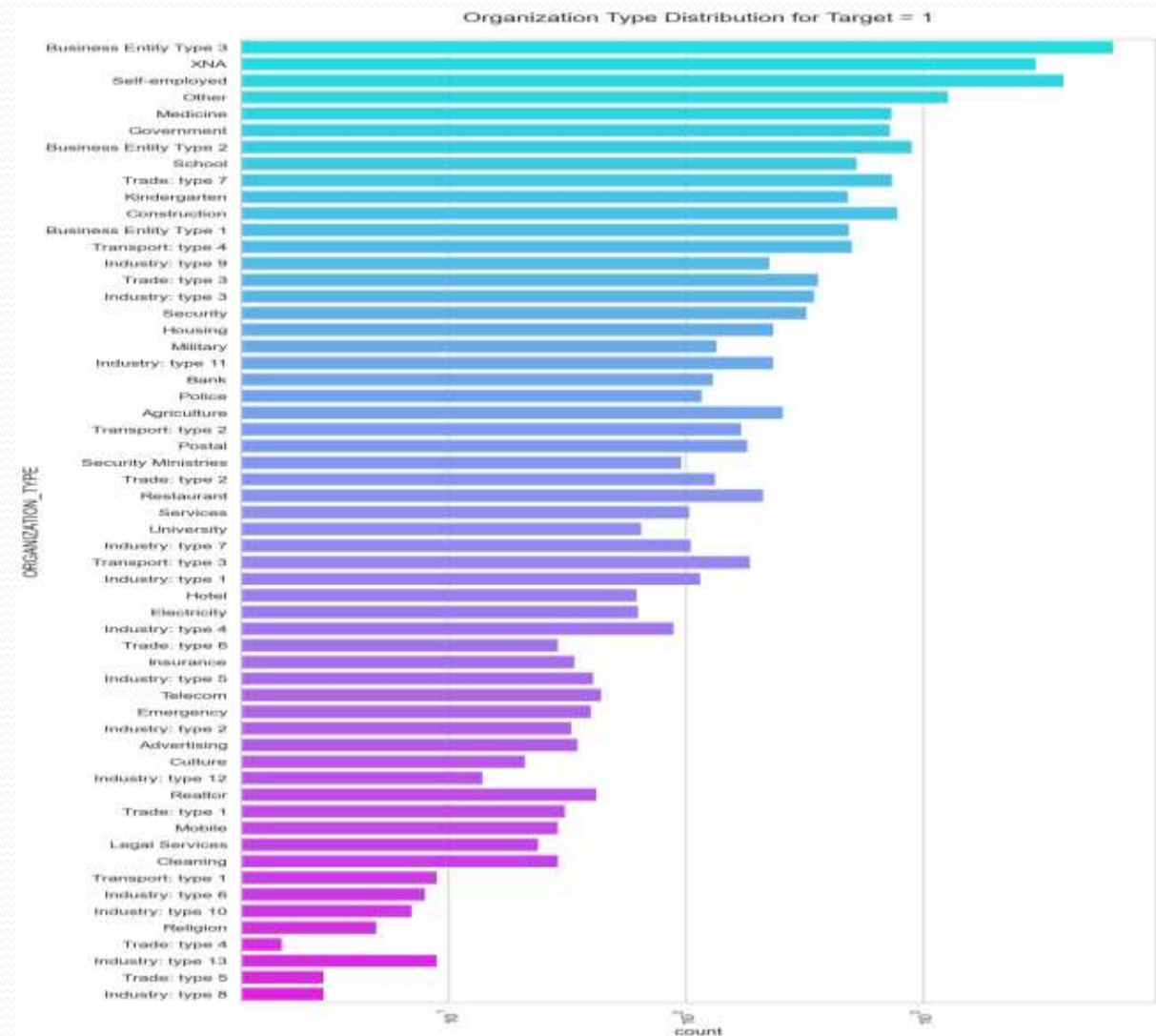
Correlation Analysis of Non-Defaulters

- Age has higher negative correlation with amount credit of loan. That means young people has lower credit than older people.

- Young people are more likely to pay off loans on time.

- People who own an apartment has positive correlation with the among credit of loan.

Correlation Analysis of Target = 0
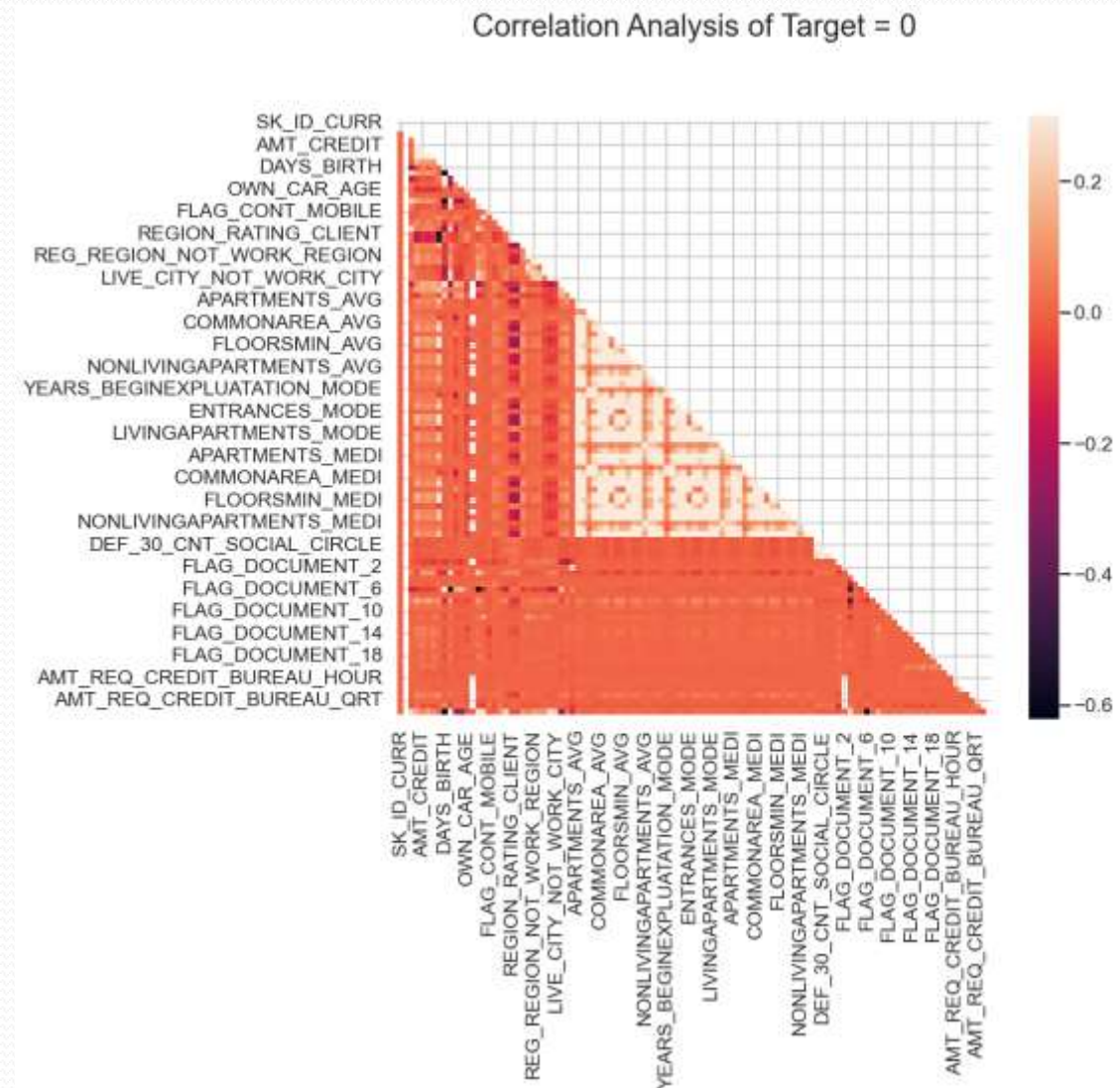
Correlation Analysis of Defaulters

- Older people has positive correlation with loan amount. That means they have higher loans.

- Age and document number has negative correlation. That means younger people tend to give lower documents.

- Amount of credit and car owner has positive correlation. That means if a person owns a car, he should have better credit amount.

Correlation Analysis of Target = 1
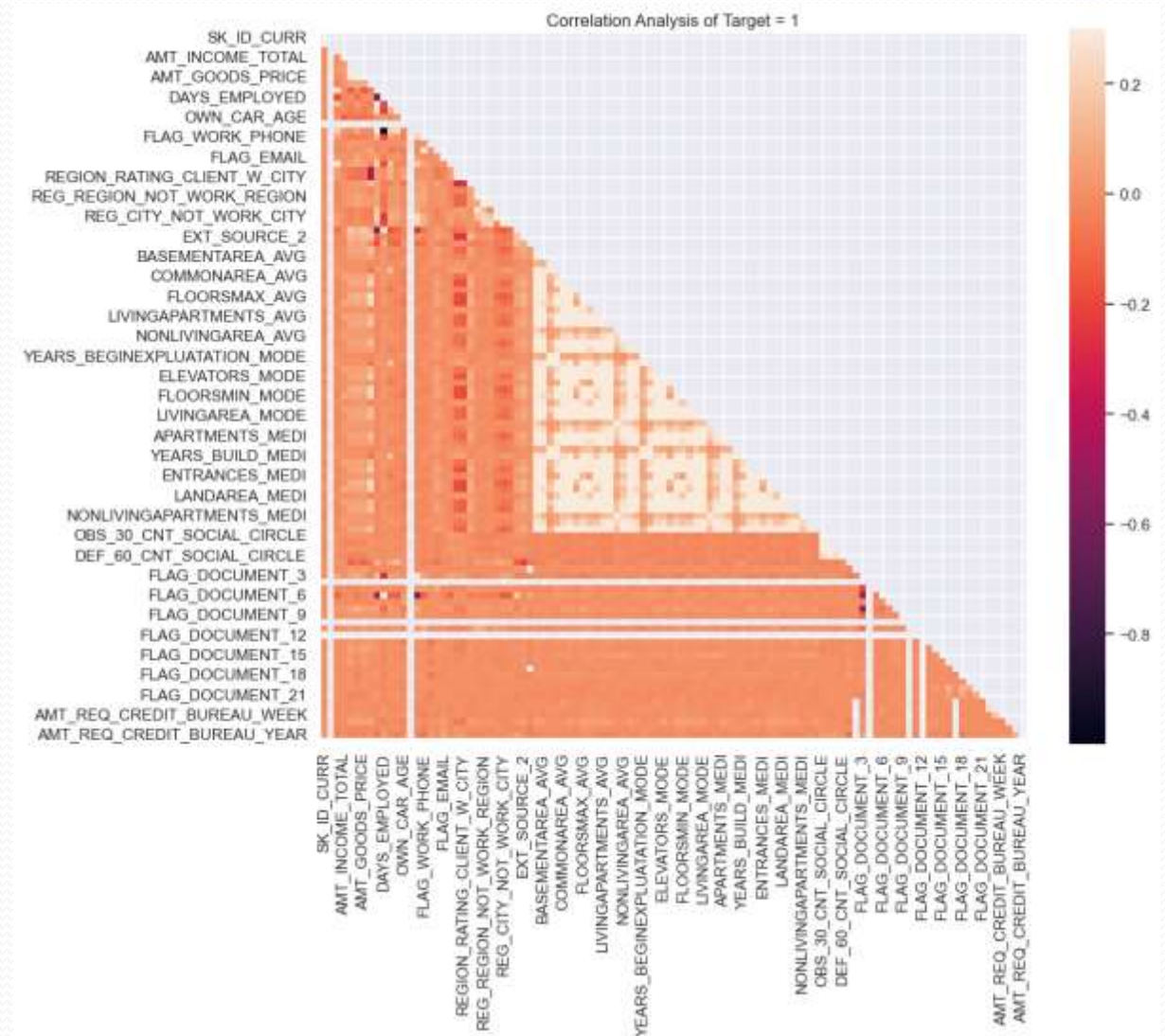
Outlier Analysis For Non-Defaulters

It is noticeable that some outliers are there in the income. Here, the quartiles are very slim, especially the third.
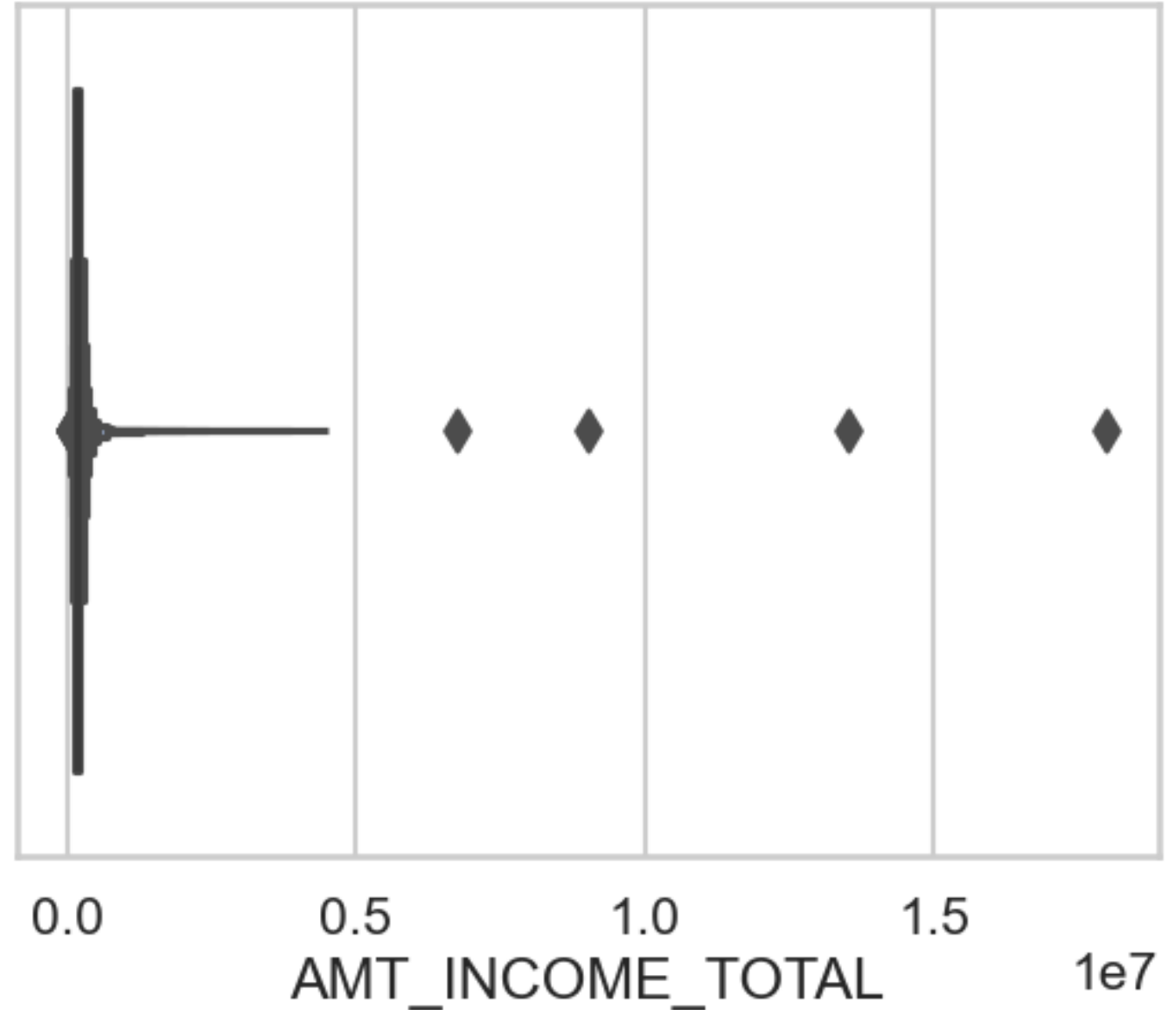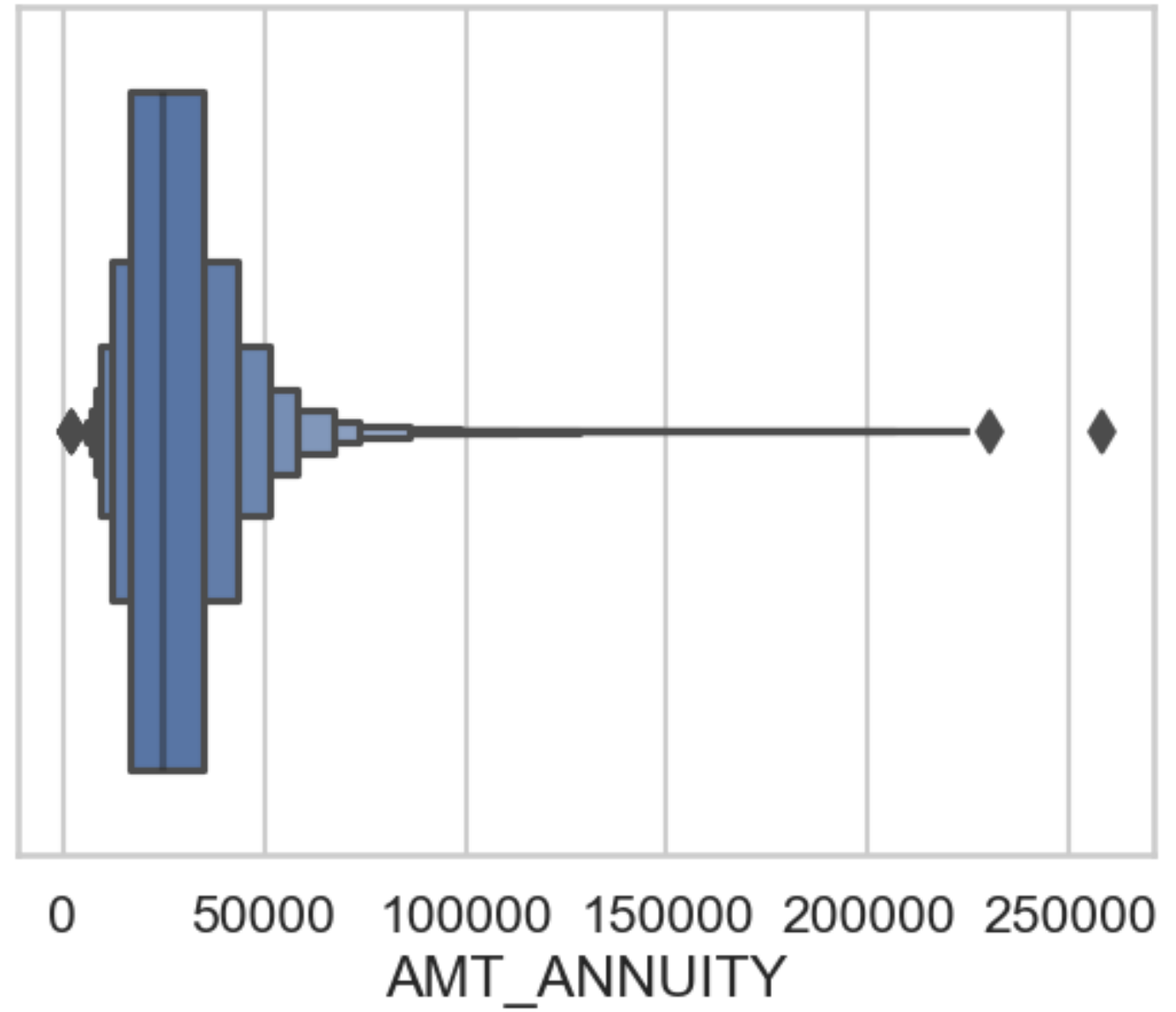
Credit Amount Box Plot For Non-Defaulters

It is noticeable that some outliers are there for credit as well. First quartile is very big, that means most of the credit amounts for customers are in the first quartile.
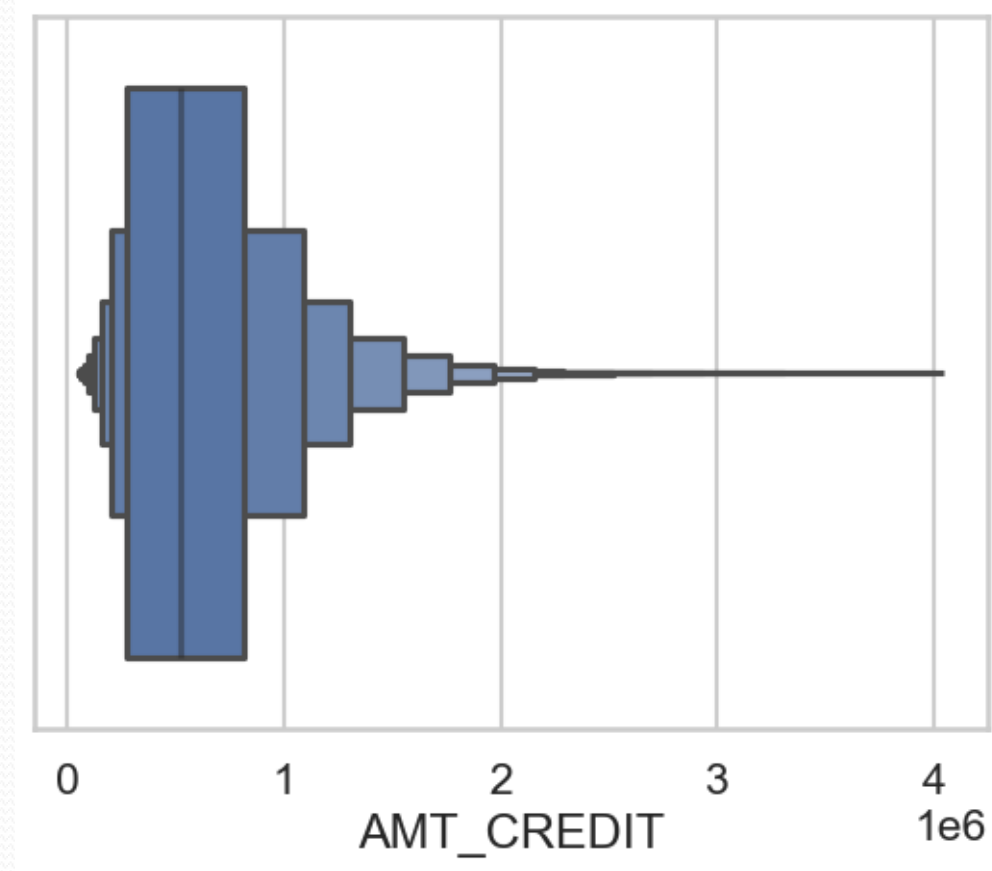
Annuity Amount Box Plot For Non-Defaulters

It is noticeable that there are a few outliers. First quartile is bigger than the third. That means most of the annuity is in the first quartile

## Credit amount vs Education box plotting For Non-Defaulters

It can be observed that in Higher Education, income remains similar for all family status types. Higher education shows more outliers than others. Academic degree shows the least numbers of outliers. Lower Secondary of civil marriage, single, and widow shows lowest income in all.
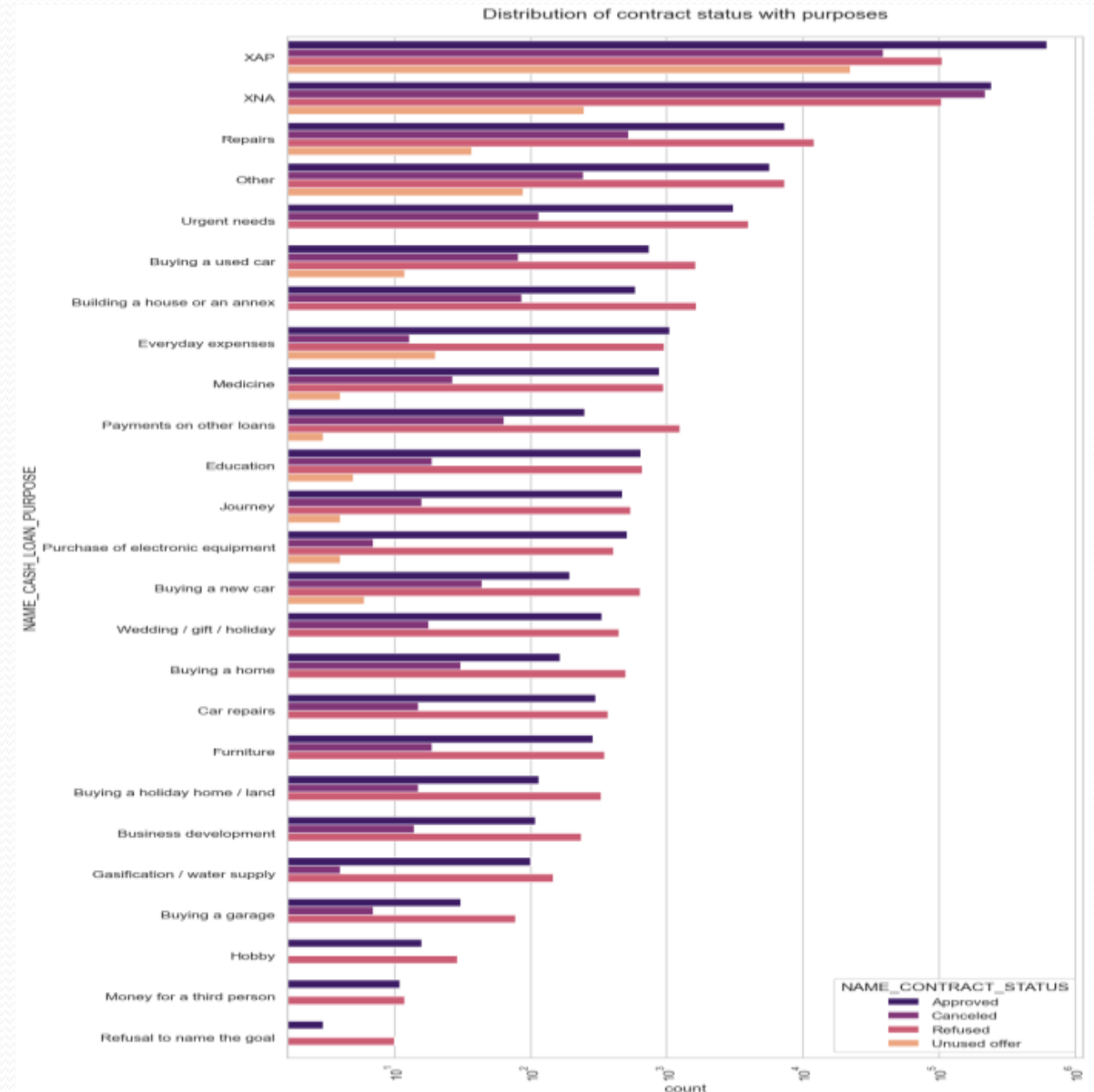

Credit Amount Vs. Education for Target = 0
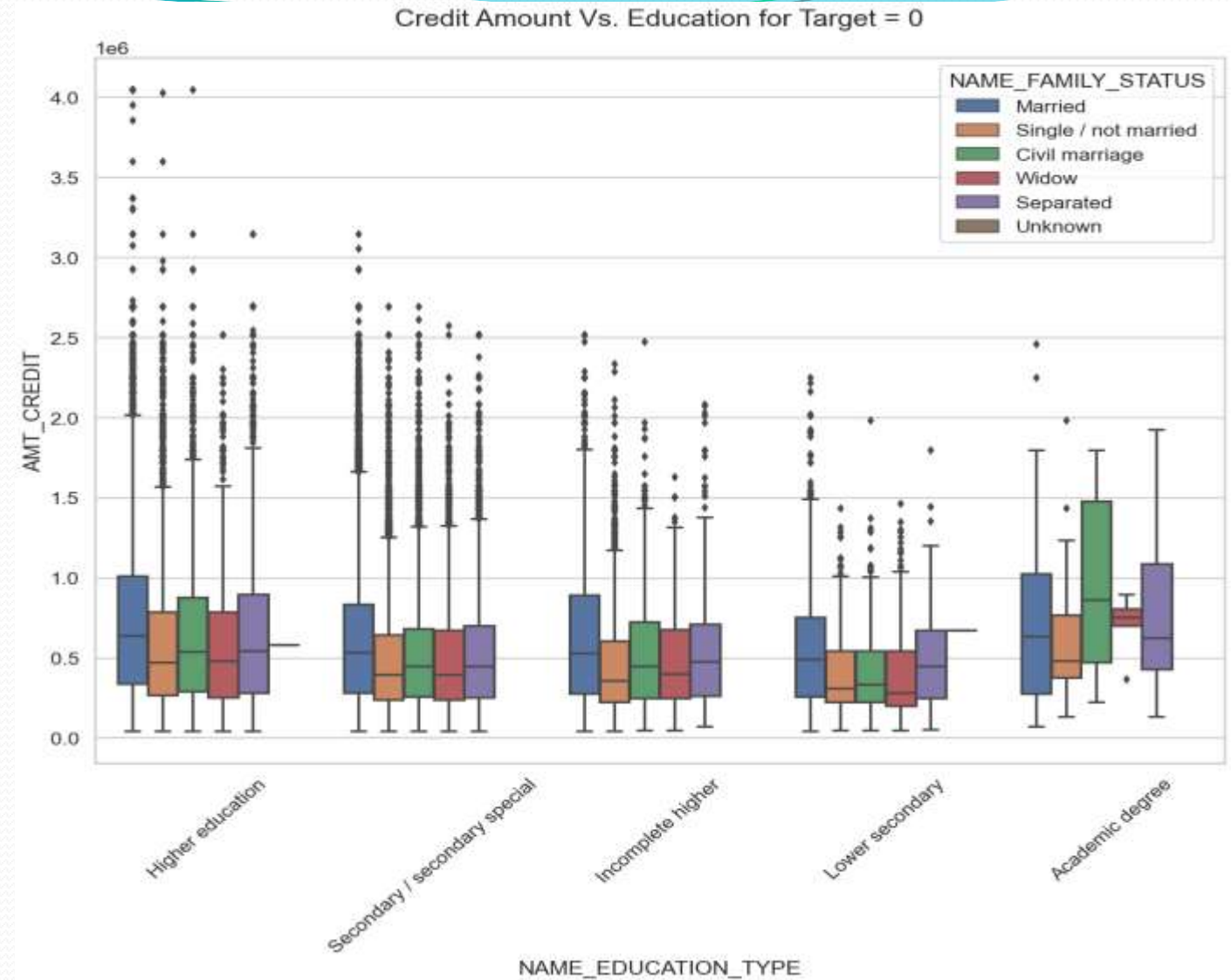
Credit amount vs Education box plotting For Defaulters

Here, married and Secondary special show higher income and higher numbers of outliers. Interestingly, Academic degree show highest income in all with lower numbers of outliers. Incomplete Higher and Lower Secondary of singles show lower income in all.

Credit Amount Vs. Education for Target = 1
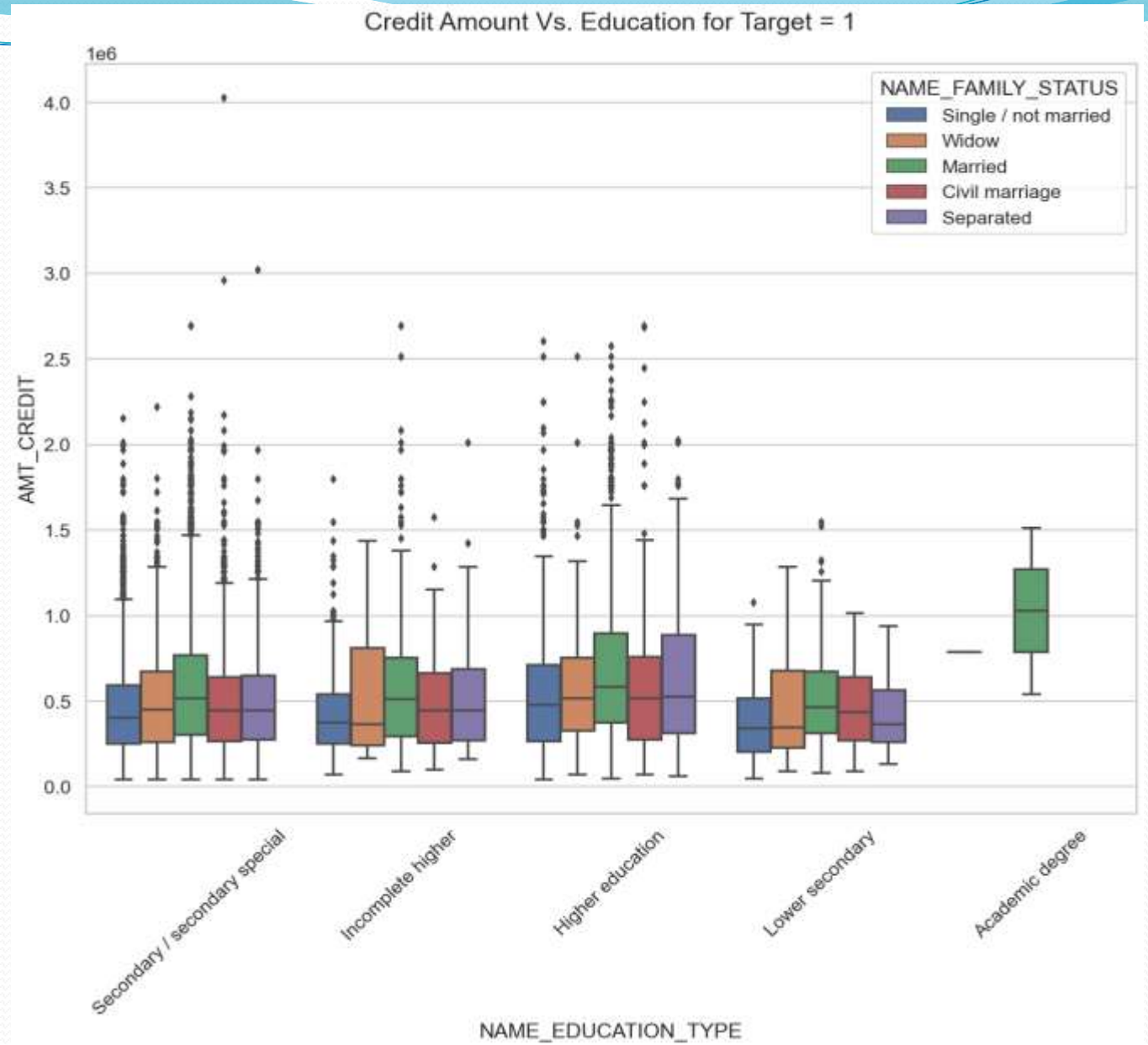
Defaulters Vs. Non-Defaulters Contract Status Distribution Analysis

- Loan purpose of XAP, XNA, and Repairs are having hard time in paying off the loan.

- Buying a garage, Business development, Buying a new car, Education, Buying a holiday home/land are showing higher loan payment than payment difficulties.

Distribution of contract status with purposes

Credit amount vs Housing Type Box Plot Of Both Target Category

- Here, House/apartment shows same credit for target 1 and target 0.

- Co-op apartment shows higher credit for target 0 than target 1. Thus, it can be concluded that bank should not give loan to Co-op apartment as they are having difficulty in paying loans.

- Municipal apartment and with parents are the best options for giving loans and successful payments.

Prev Credit amount vs Housing type
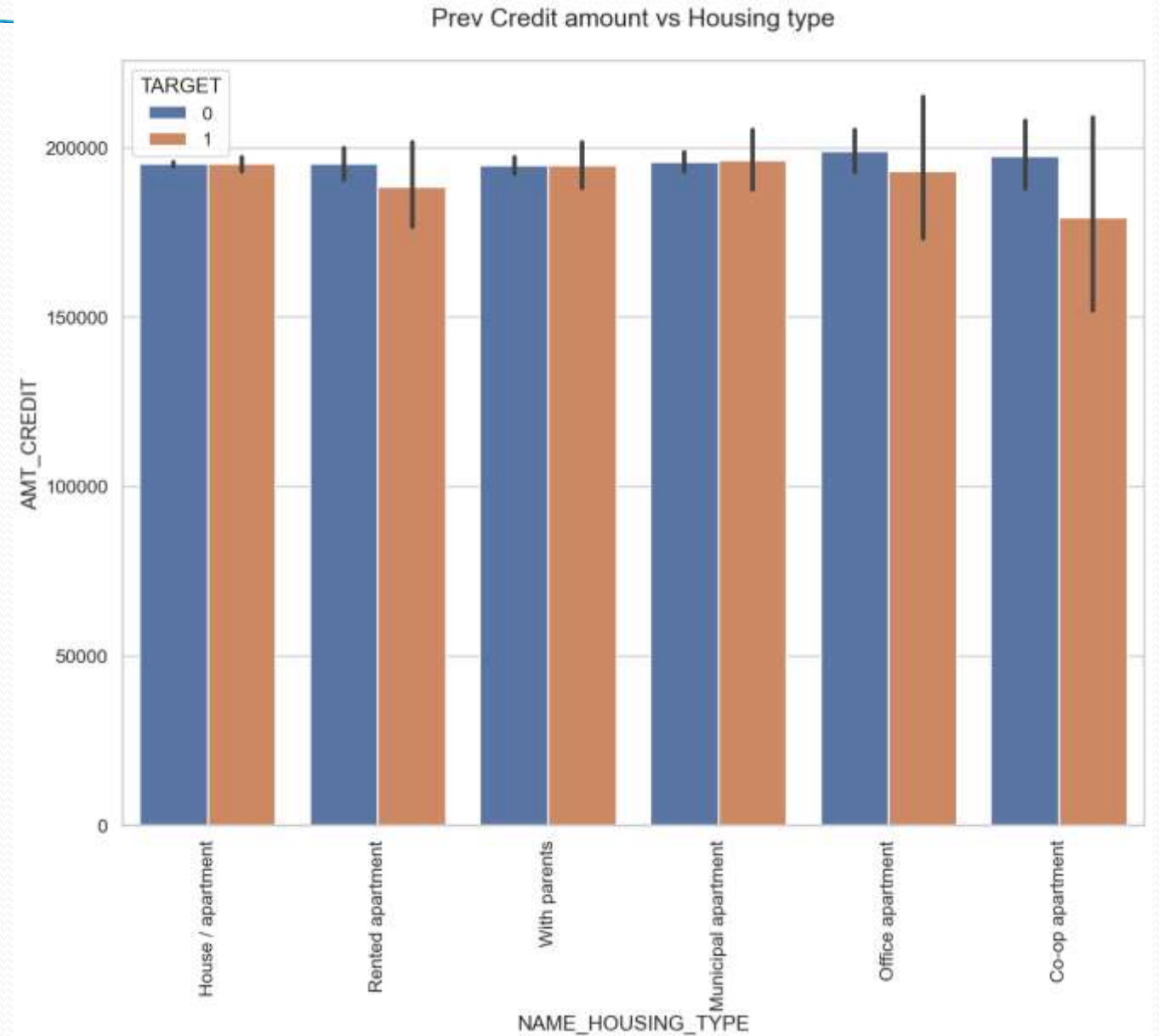
Conclusion

- Bank should find more customers of Student, Pensioners, and Businessman with housing type other than Co-op apartment.
- Bank should focus less on Working income type as they tend to be defaulters.
- Loan purpose of Repair shows higher unsuccessful payments. It should be avoided by the banks.
- Housing type With Parents are the best options as they have higher rate of successful payments.

Thank You