

Credit EDA

Problem Statement

- Due to the insufficient or non-existent credit history of the people, the loan providing companies are finding it hard to give them loans.
- Some consumers use it to their advantage by becoming a defaulter.

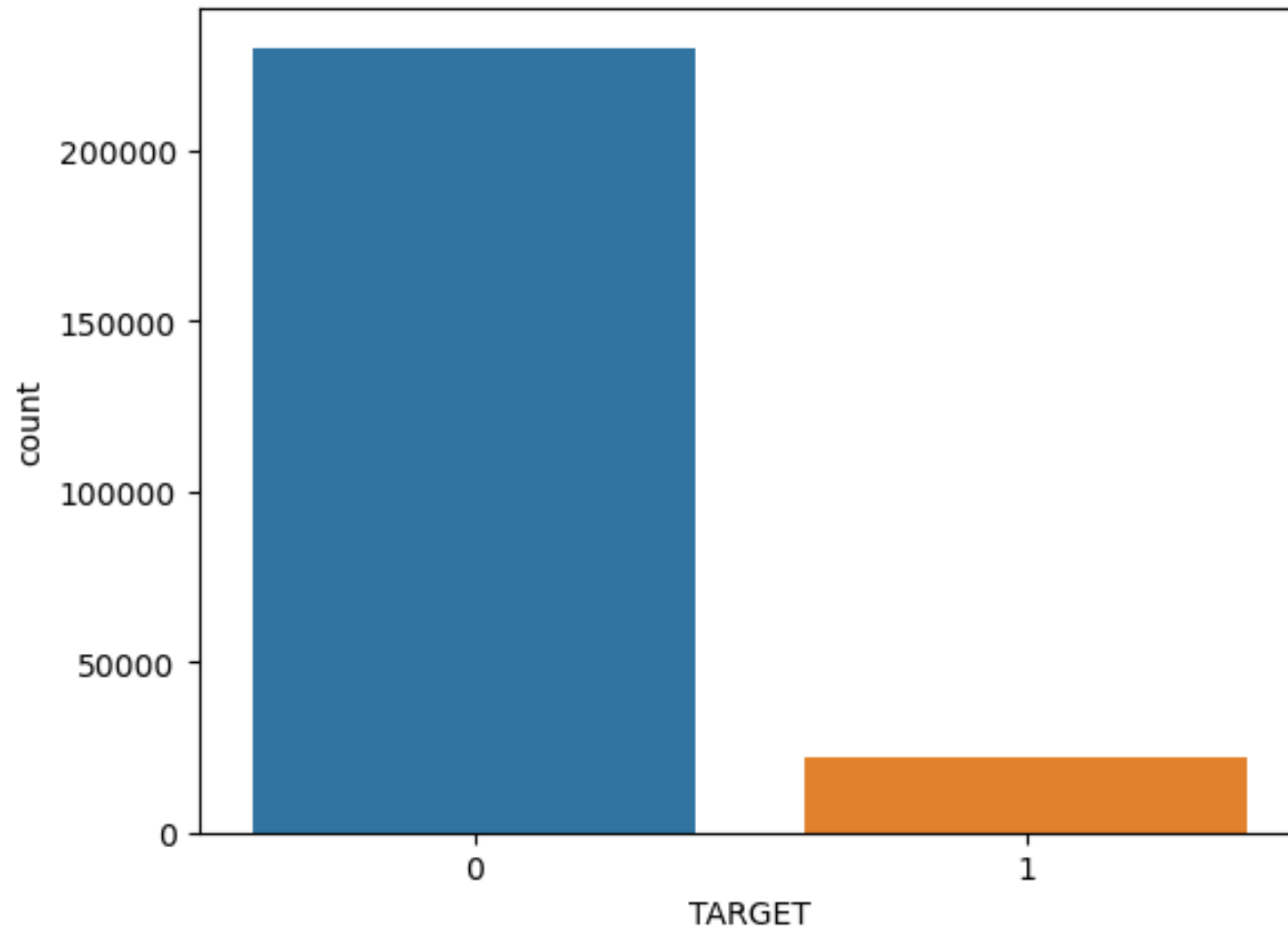
Objective

- Identifying the patterns indicating that client has difficulties paying their installments .
- This Case Study will ensure that the consumers capable of repaying the loan are not rejected

Approach and Methodology

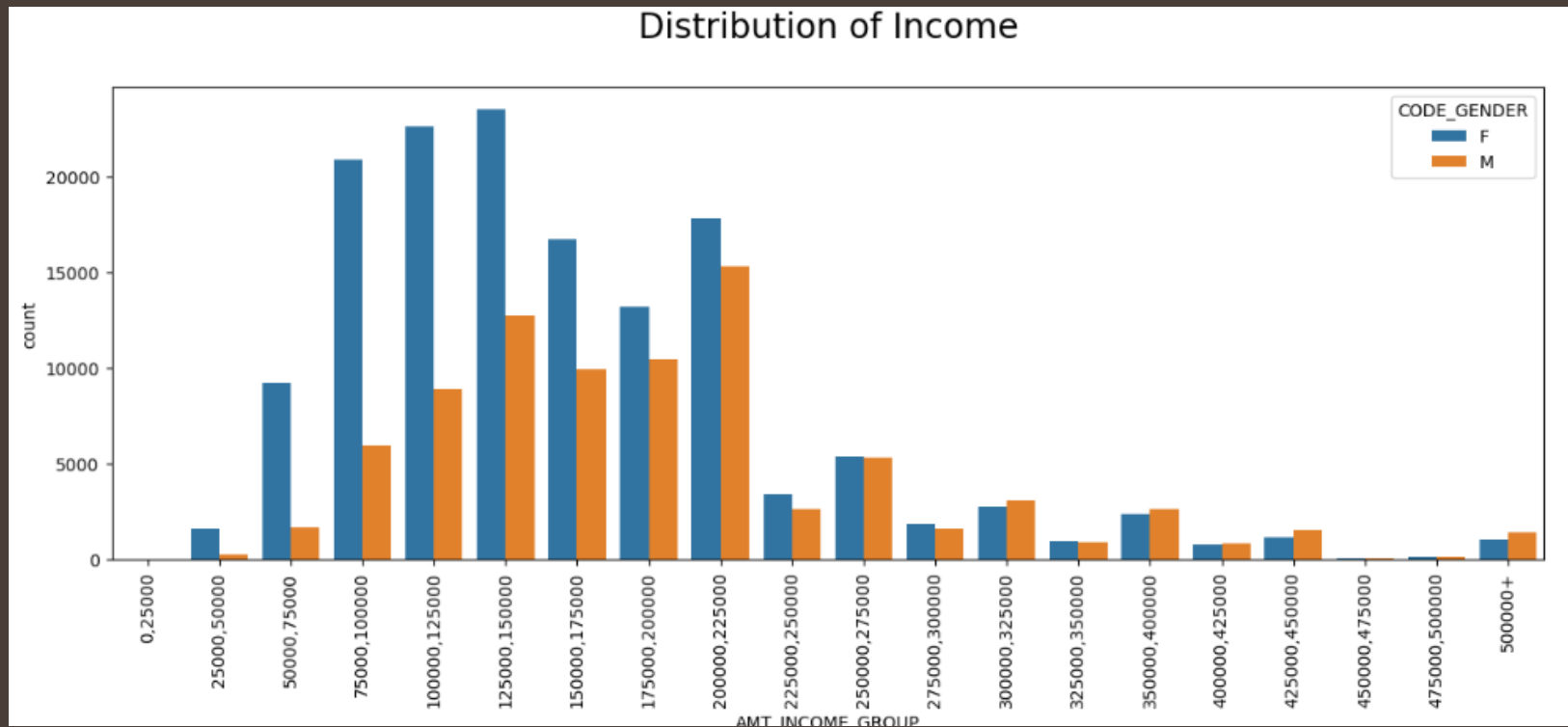
- We will first clean the data by identifying the null or missing value columns.
- We will impute these columns by using appropriate method either by replacing null values or dropping these columns.
- We will check for outliers in the dataset if present.
- Check for Data imbalance and use different plots and graphs to analyze different aspects due to data imbalance.

Imbalance
Ratio of
Target_0 and
Target_1 is
10.55



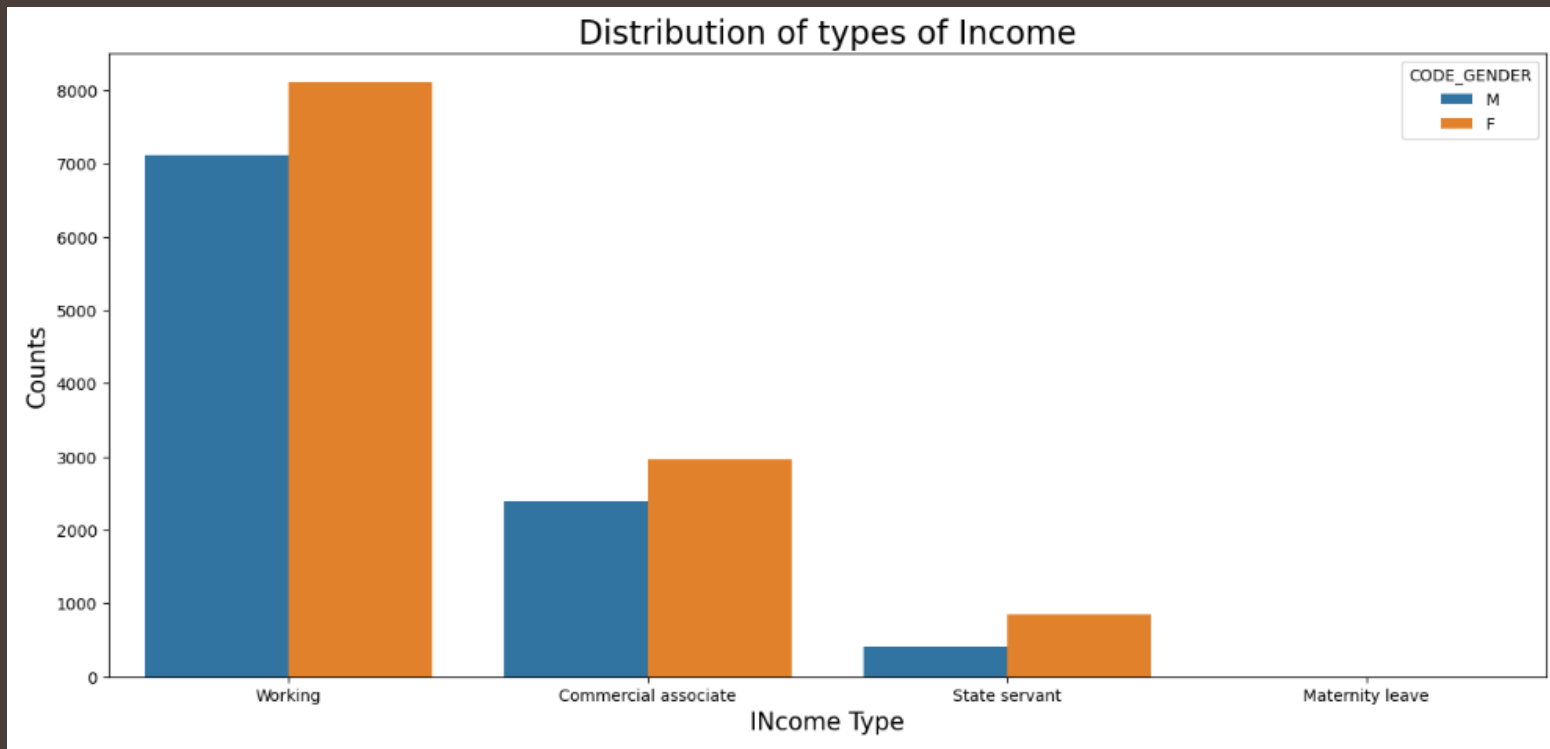
Univariate Analysis Of Target o





Distribution of Income Range

- 1) Income range from 125000-150000 is having more number of credits
- 2) Less count from range 450000-475000
- 3) Females are having more number of credit than Male in range 125000-150000



Distribution of Type of Income

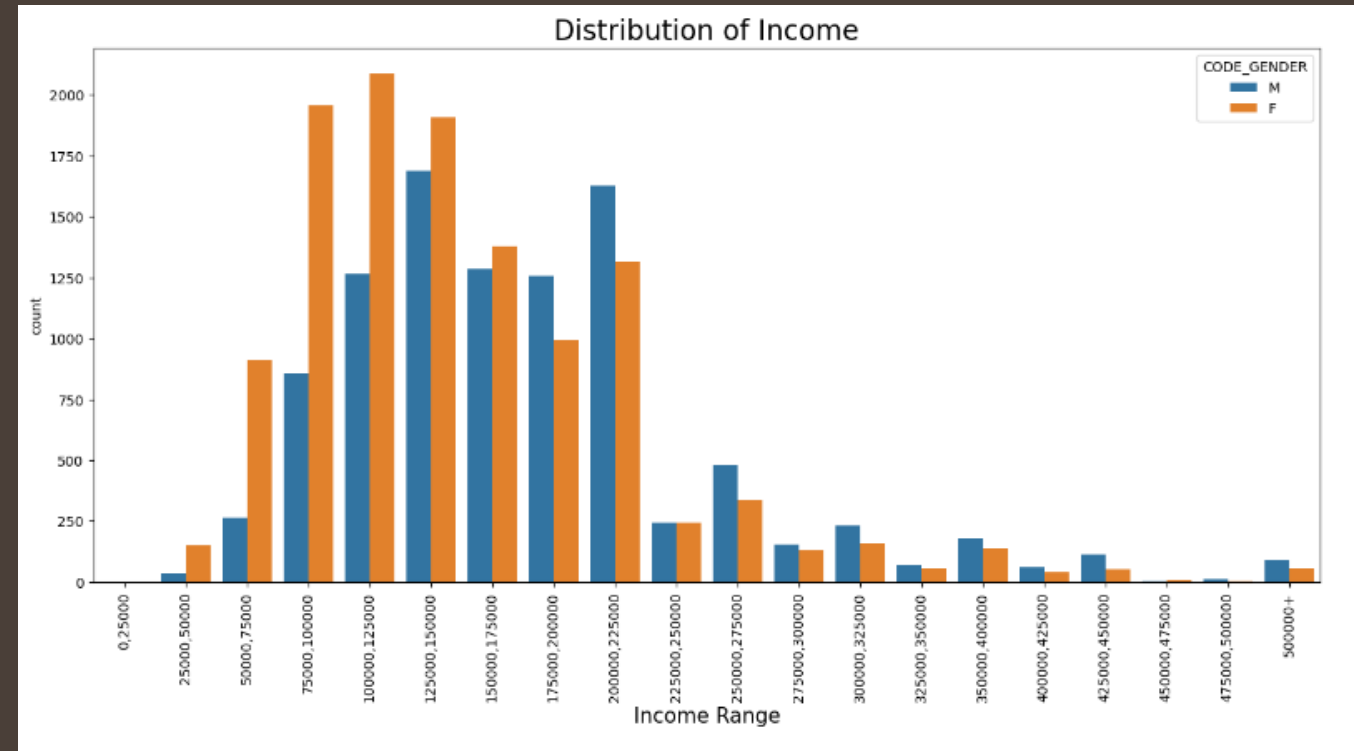
- 1) In working category Females are having more no of credits than male
- 2) Working category, State Servant & Commercial Associate category has more number of credits than any other
- 3) Women on Maternity Leave has less number of credits than others.

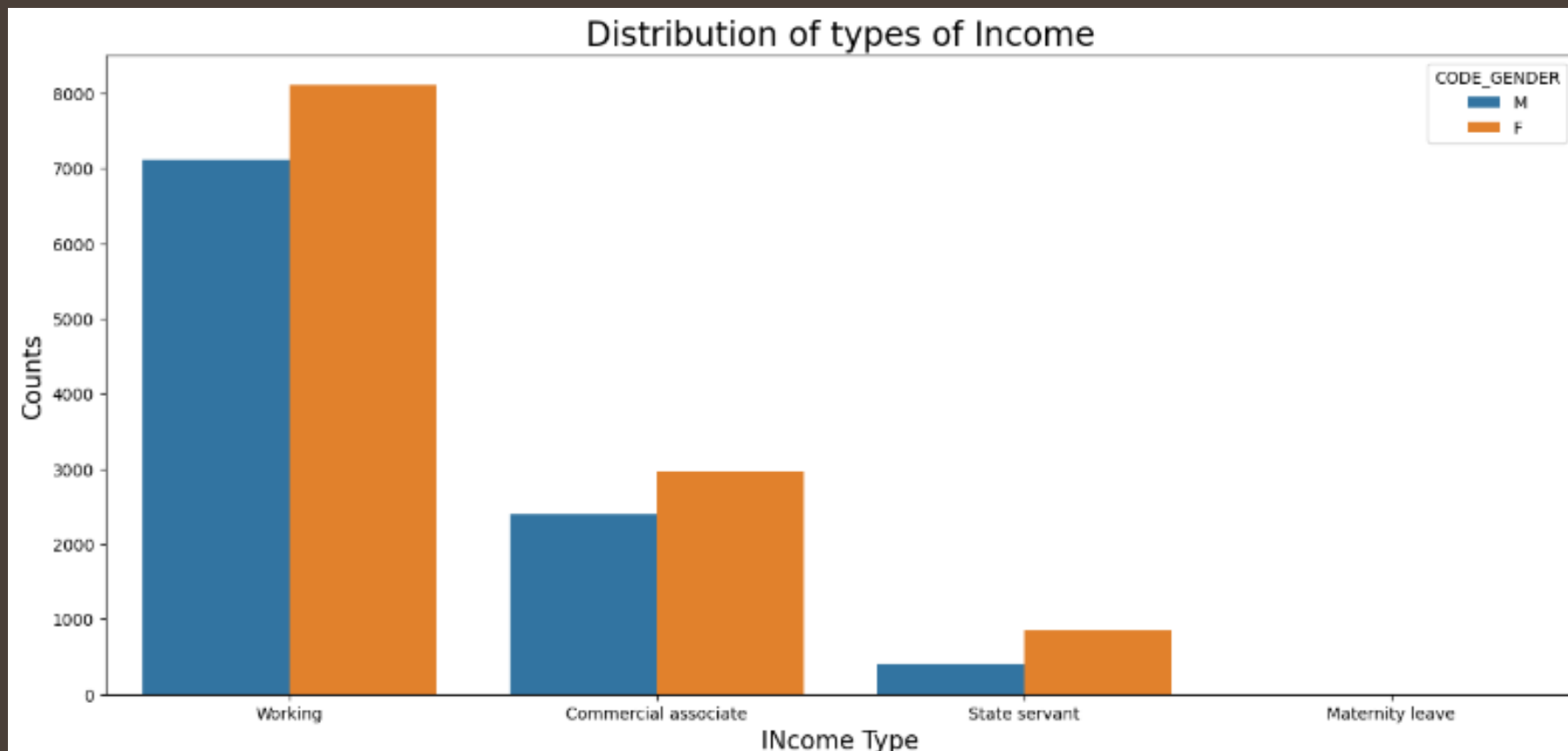
Univariate Analysis of Target_1



Distribution of Income Range.

- 1) Male counts are higher in Target 1
- 2) Income range 450000-475000 has less number of credits
- 3) Income range 100000-125000 has large number of credits



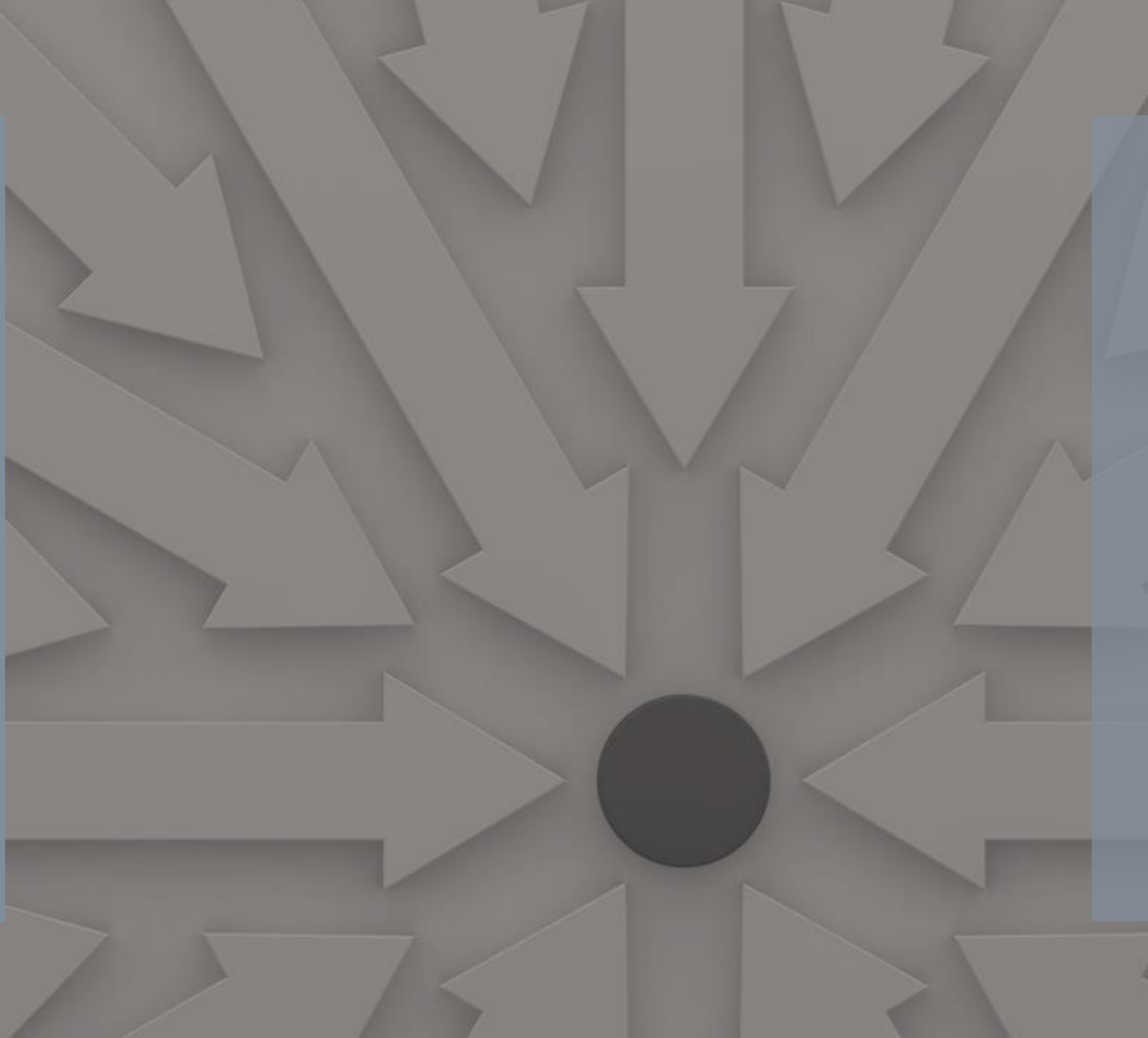


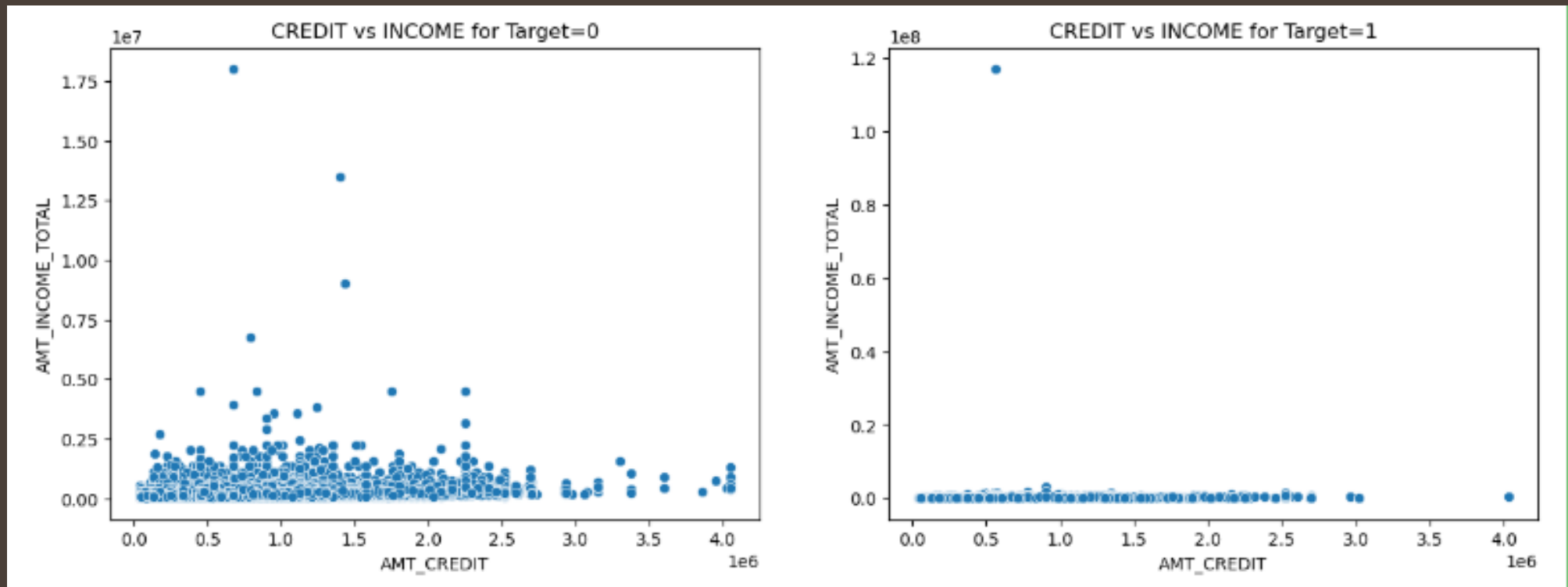
Distribution of Income Type

Conclusions ::

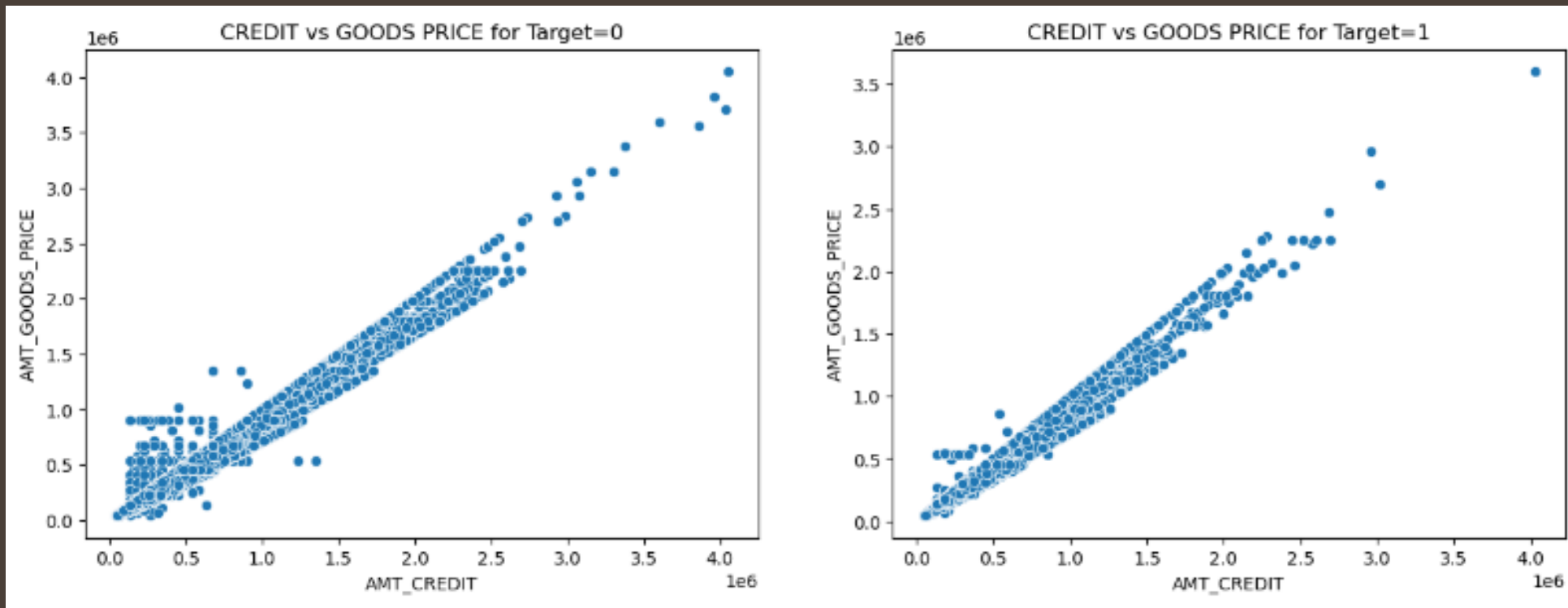
- 1) Except Maternity Leave category all other has almost high credits
- 2) For this Females are having more number of credits

Bivariate Analysis of Target_o and Target_1





Credit Amount Vs Income Amount



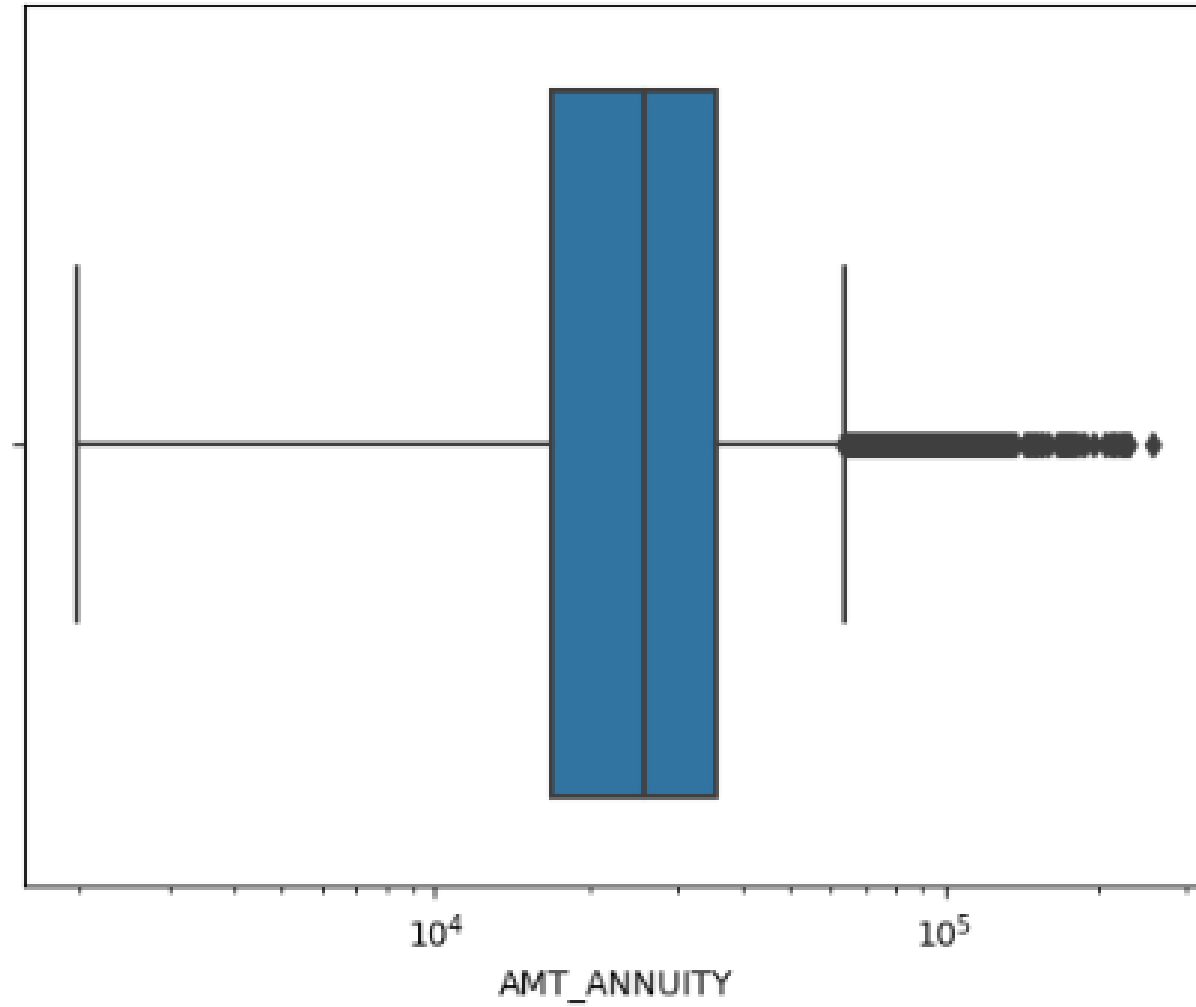
Credit Vs Goods Price.

Conclusion :

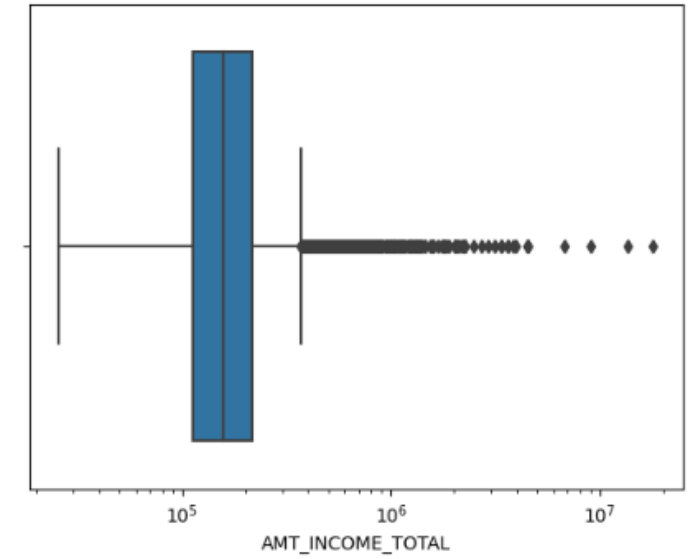
If increase in goods Price credit Increase directly and vice-versa

Outliers for Target_o

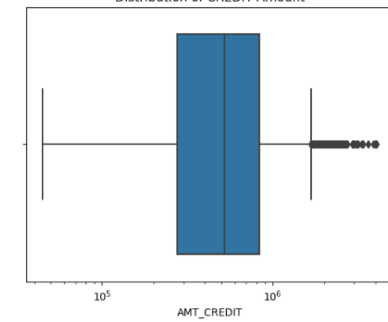
Distribution of Annuity Amount



Distribution of Income Amount

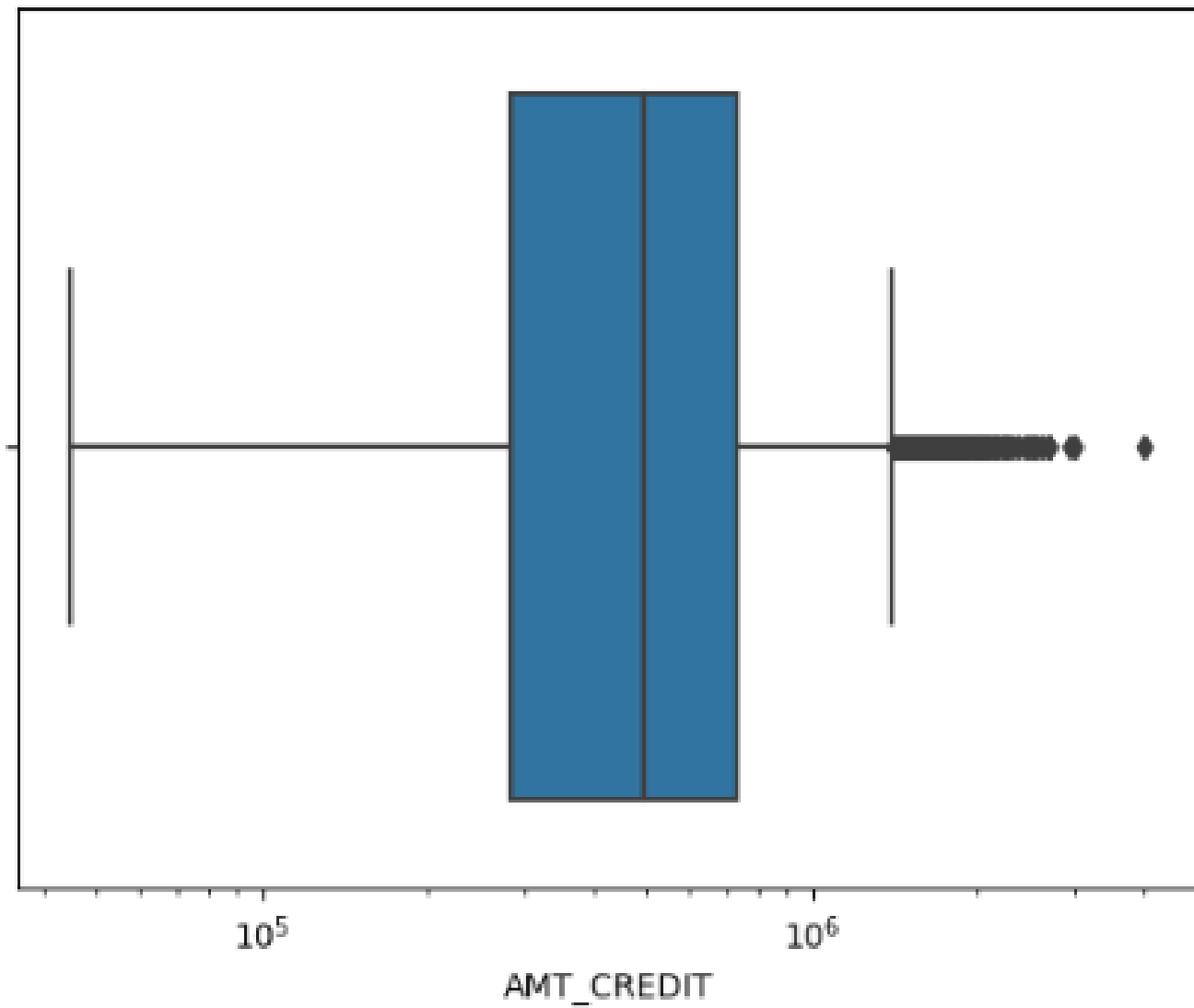


Distribution of CREDIT Amount

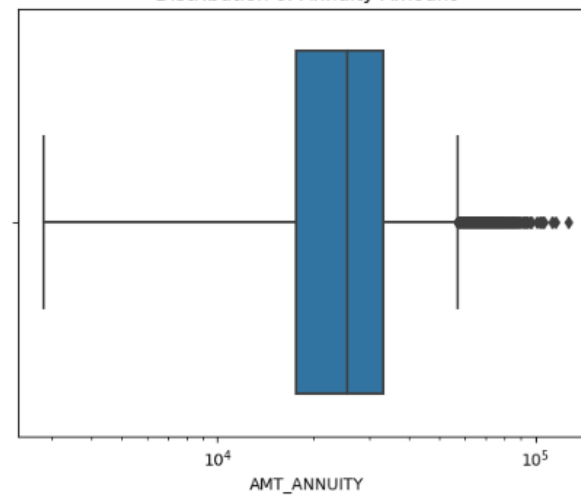


Outliers For Target_1

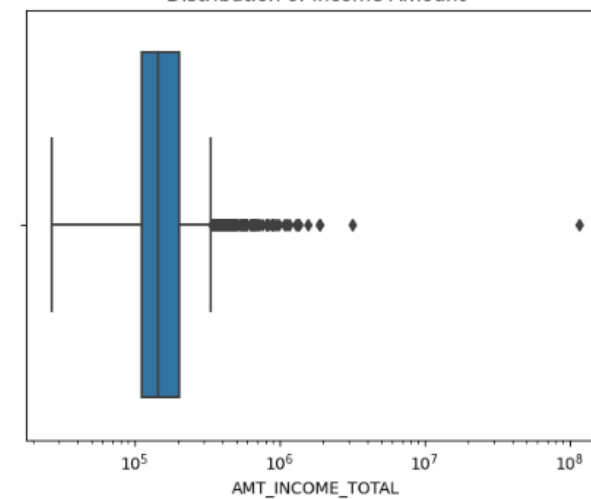
Distribution of CREDIT Amount



Distribution of Annuity Amount



Distribution of Income Amount



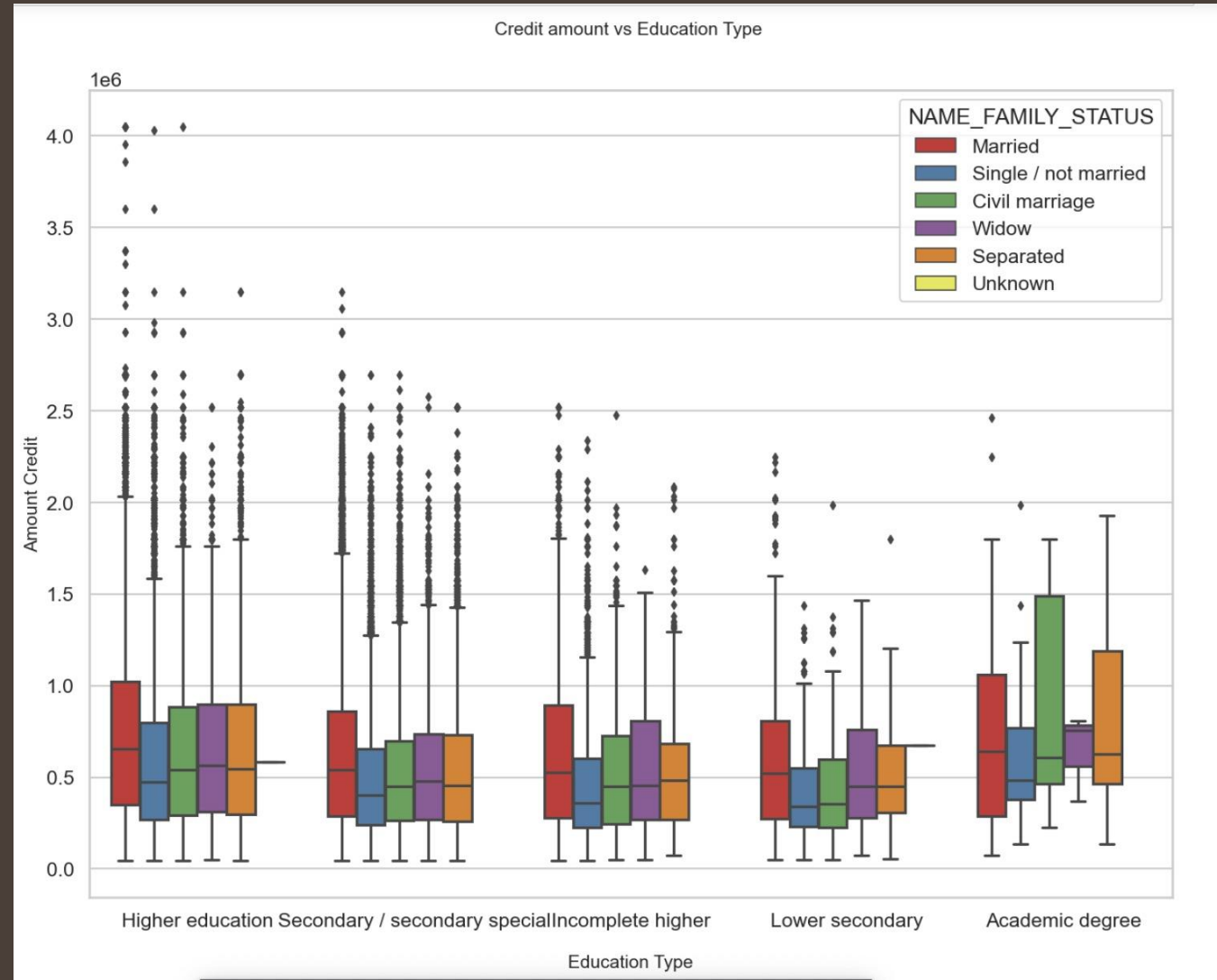


Multivariate Analysis of Target_o

Credit Amount VS Education Type.

Conclusion:

- 1) Family Status of Higher Education is having higher outliers than others.
- 2) Family Status of Academic Degree is having less outliers than others.
- 3) Civil marriage in Academic Degree category has more number of credits.
- 4) Family status of 'Civil Marriage', 'Married', 'Separated' has more credits than others

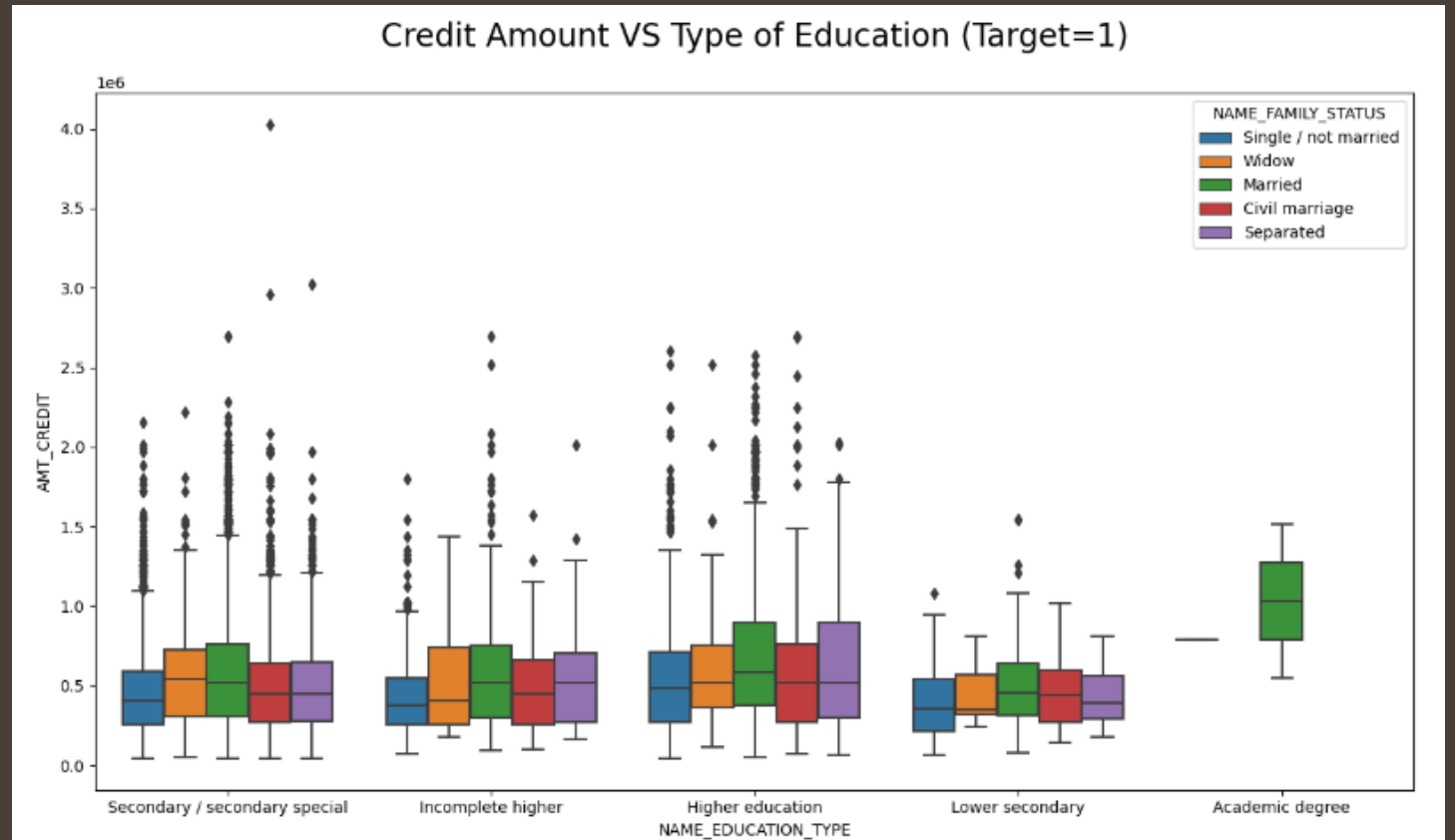


Multivariate Analysis of Target_1



Credit Amount VS Type of Education.

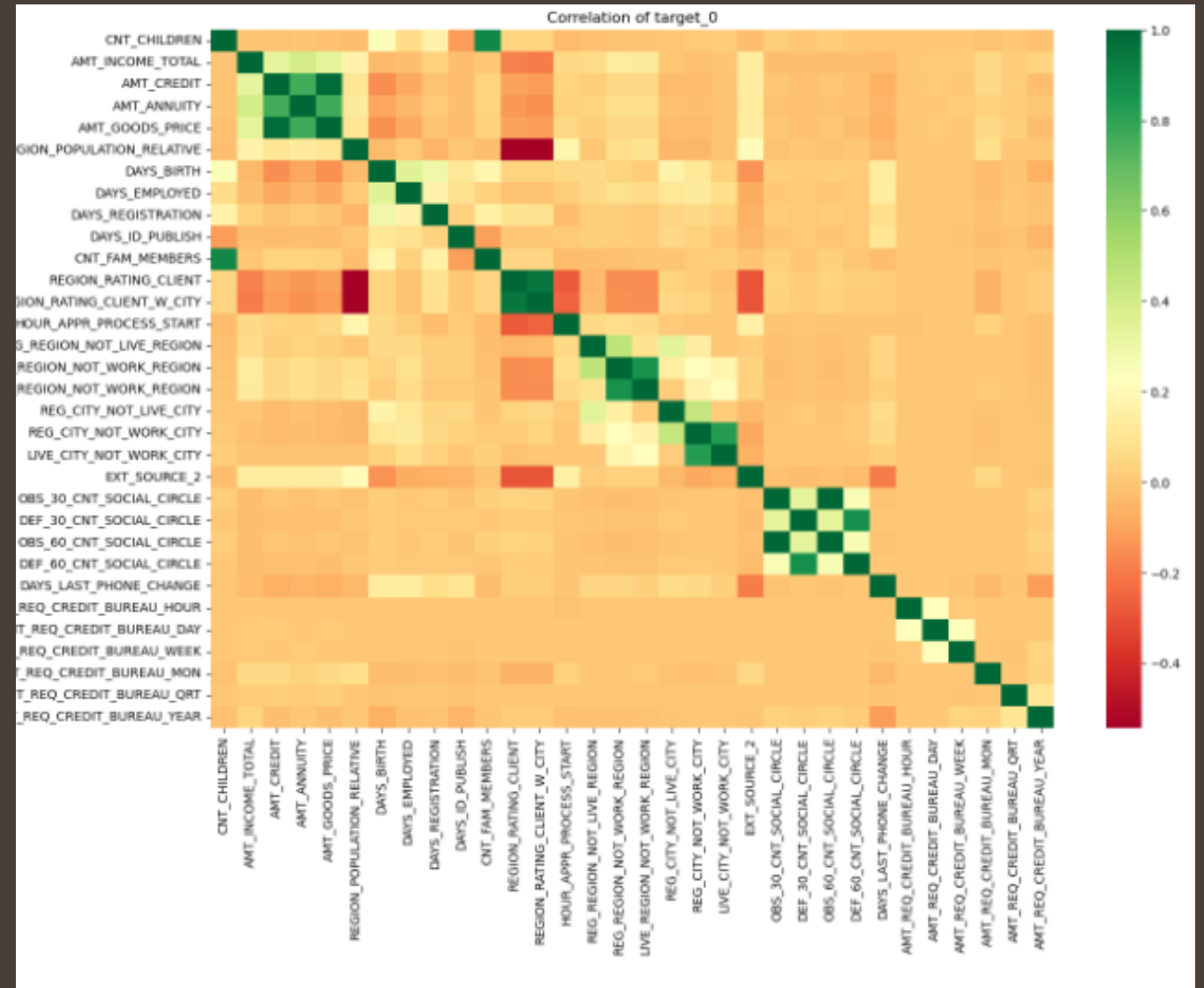
Family Status of Civil Marriage, Married and separated of Academic Degree are having more number of credits



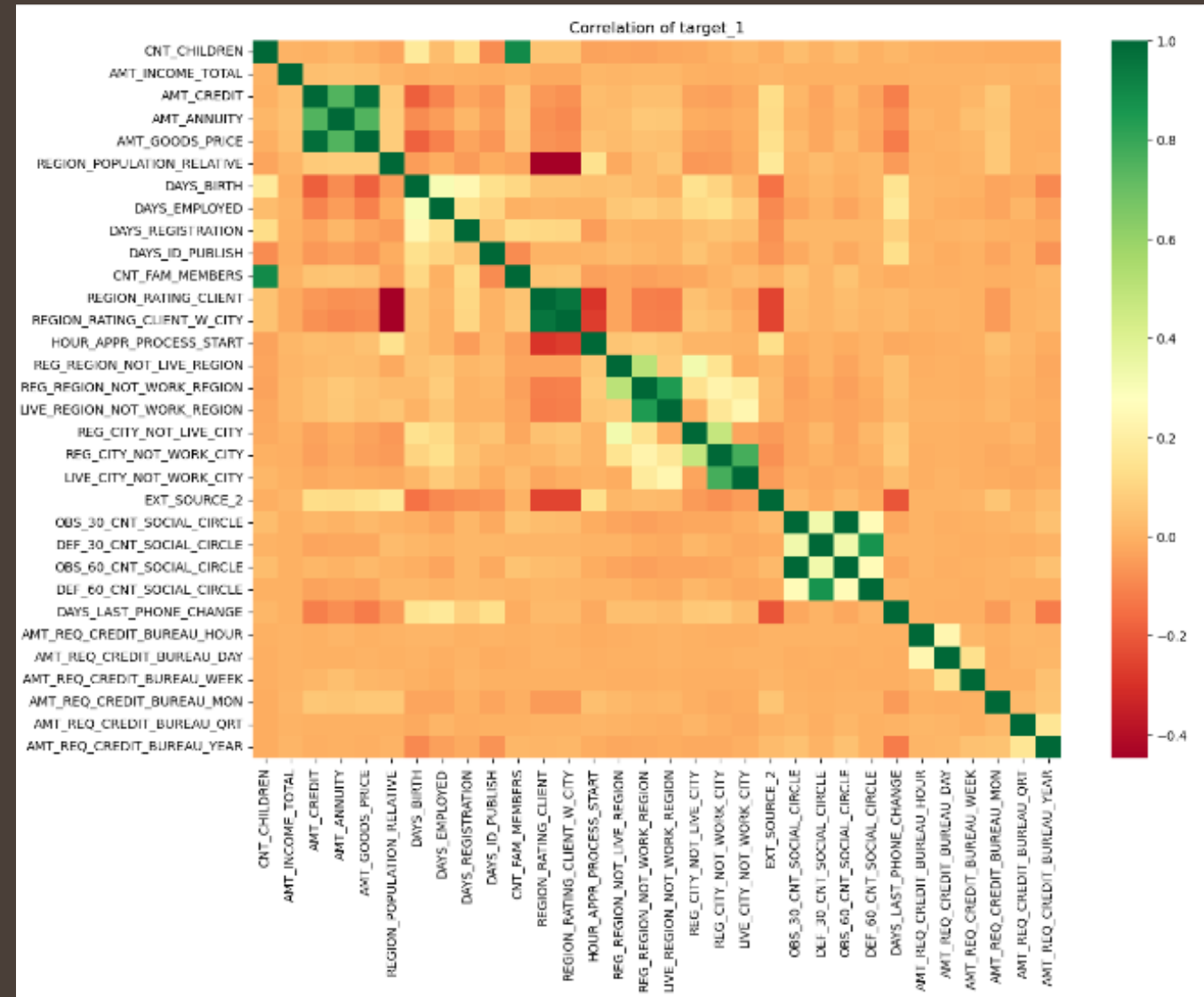
Correlation of Target_0

Conclusion:

- 1) Credit Amount is higher for low age group and vice versa
- 2) Credit Amount is higher in populated area
- 3) Income is higher in populated area
- 4) Credit amount is higher for those client who has less children and vice versa



Correlation of Target-1
Conclusions :: 1) The clients permanent address does not match with contract address are having less children and vice versa 2) The clients permanent address does not match with work address are having less children and vice versa



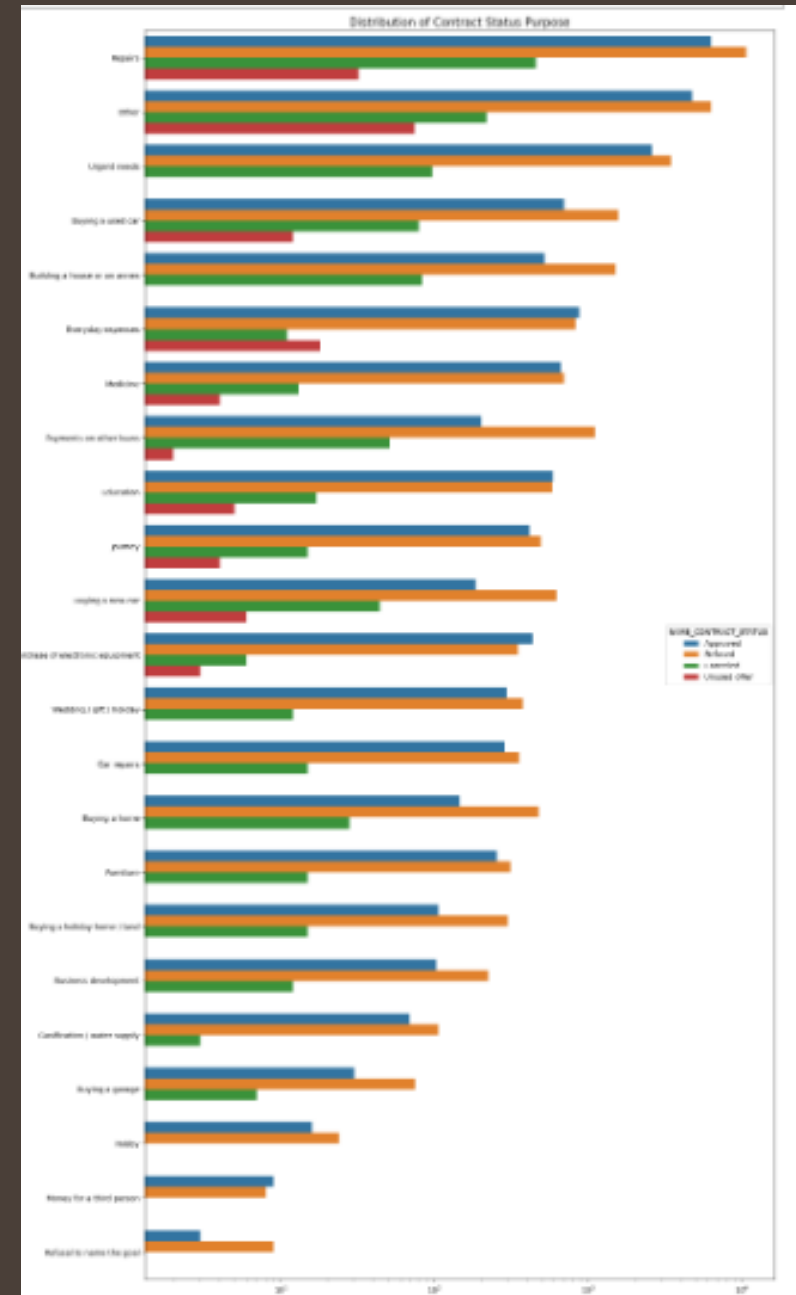


Merge the two
Dataset

Distribution of Contract Status with Purpose

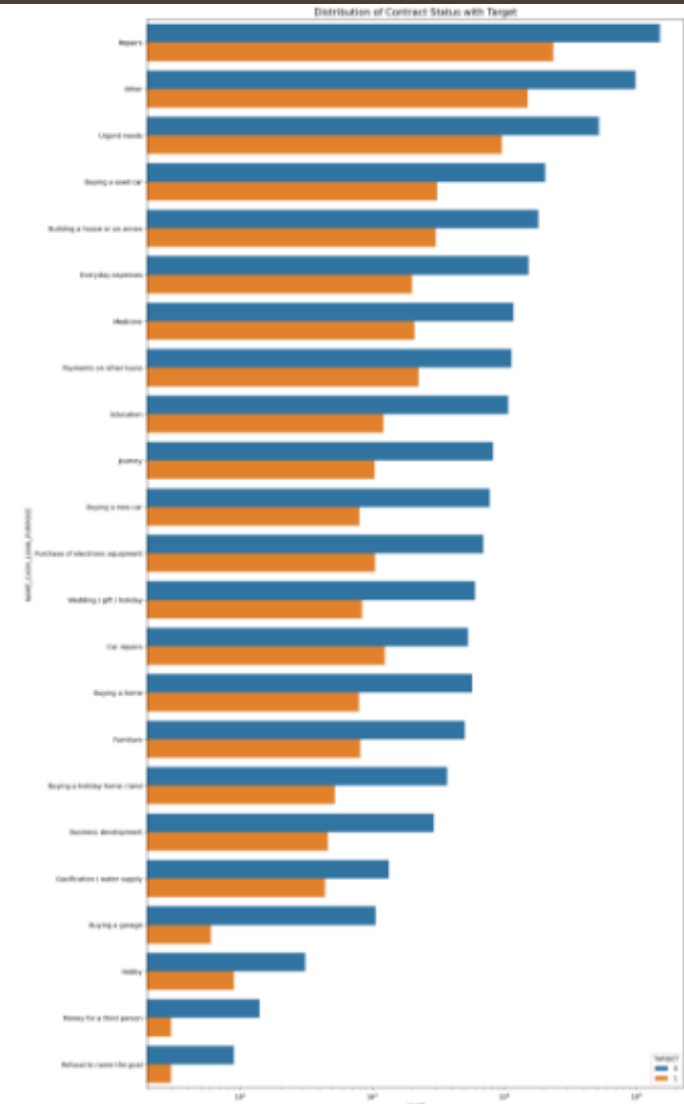
Conclusions ::

- 1) Most Rejections received for Loan Purpose 'Repairs'
- 2) For Education Purpose we have equal number of 'Approved' & 'Rejections'
- 3) Paying other loans & Buying a new car has significant higher rejection than approves



Distribution of Contract Status with Target

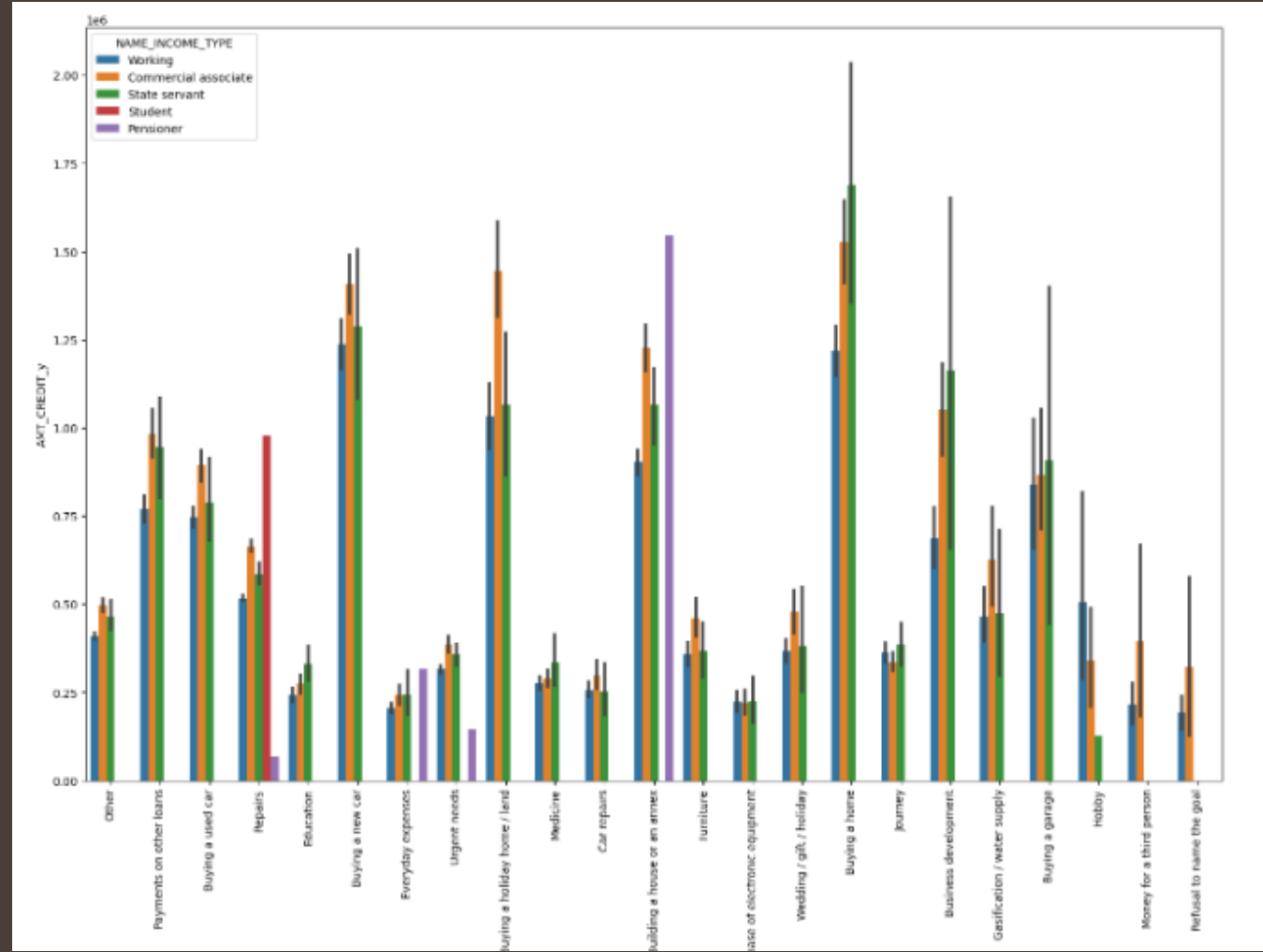
- 1) Loan Purpose Repairs are defaulters
- 2) 'Buying a garage', 'Buying a land', 'Buying a new car', 'Business Development' are places where loan payment is significant higher than facing difficulties



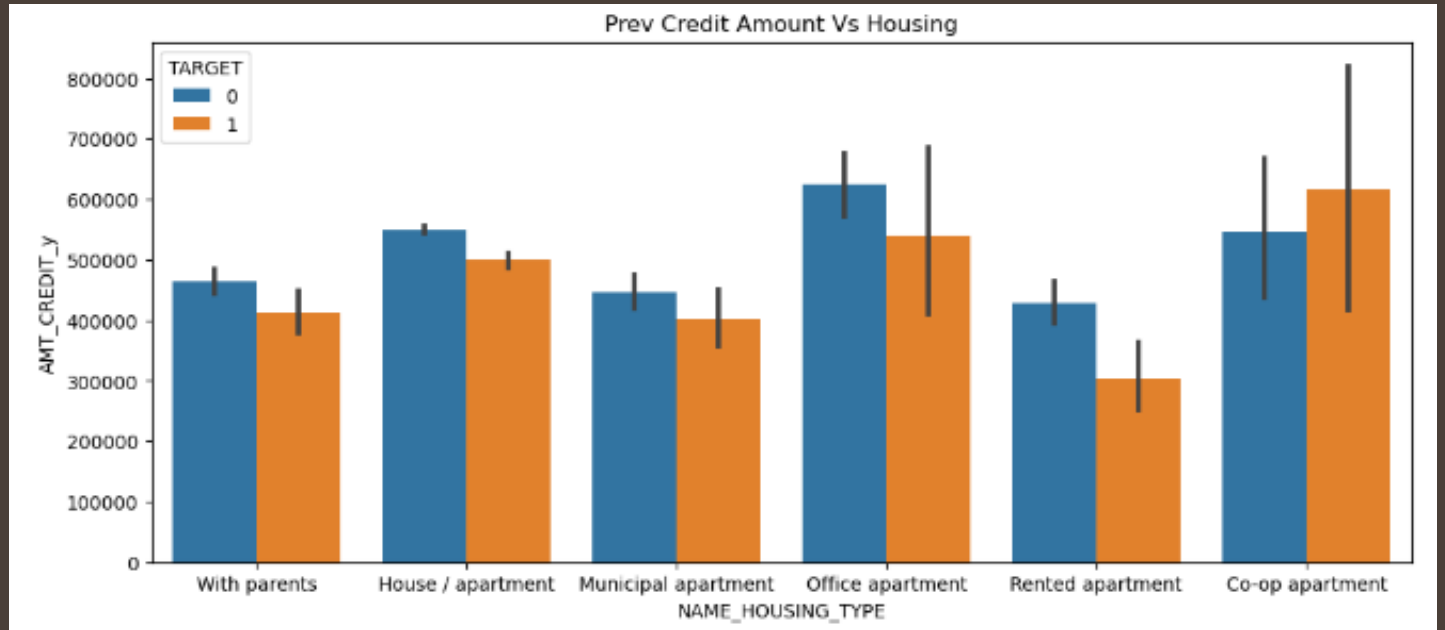
Bivariate Analysis on Merge Dataset



Previous Credit VS Loan Purpose



Previous Credit Amount VS Housing



Conclusion:

- 1) Banks should provide loans to Repairs & Other purposes
- 2) Working Female employees are the best target for Loans.
- 3) Banks should provide more loans more to Office Apartments or Co-op Housing apartments as there are less payment difficulties



THANKYOU

