Credit EDA_Assignment

~By Ravi Kishore Yerra

BUSINESS OBJECTIVES

- This case study aims to identify patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.
- The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

PROBLEM STATEMENT

- From the given data we have the information about the loan application at the time of applying for the loan. There are two types of risks associated with any type of loan requests:

1. (“Target = 0”) – All the applicants when the payment is paid on time.
2. (“Target = 1”) – All the applicants when the payment is not paid on time: he/she is likely to had late payments of the loan in our sample of dataset.

ANALYSIS

1. Data Understanding & Structure Check
2. Data Quality Check & Missing Values Check (Cleaning Process)
3. Data Imbalance Check
4. Splitting the DataFrame into 2 SubDataFrames
5. Univariate Analysis
6. Bivariate Analysis
7. Correlation using Heatmap
8. Merging application data with previous application data
9. Again Univariate & Bivariate Analysis and Correlation
10. Risks & Recommendations

Data Understanding & Structure Check

- Initially after importing the relevant libraries we check:

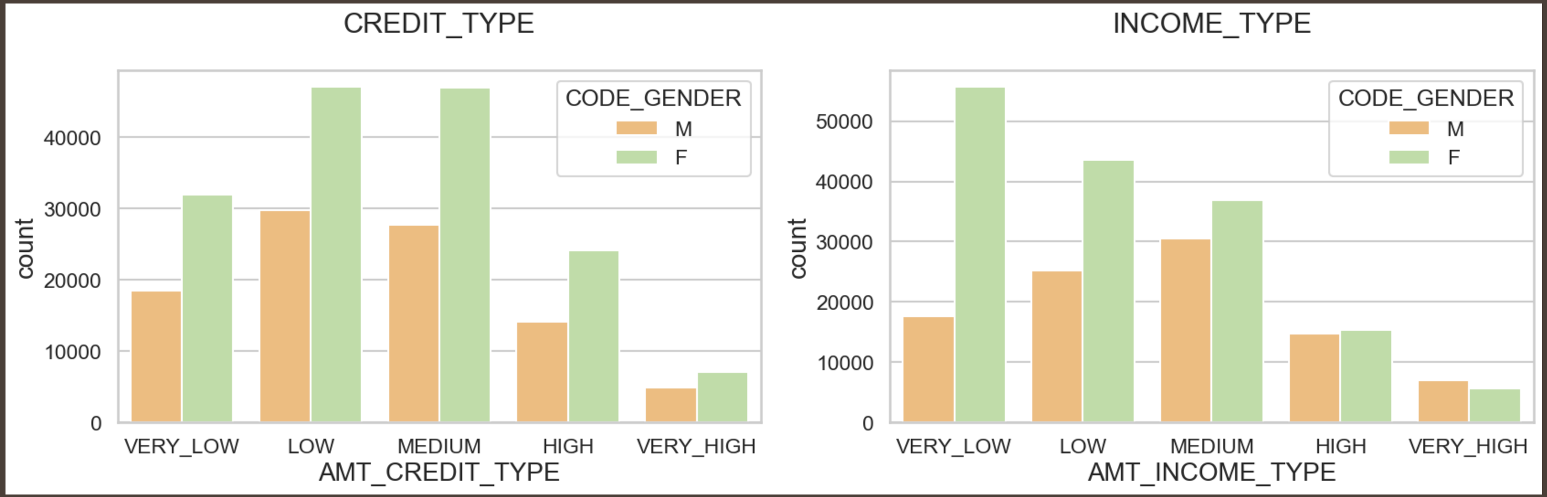
1. The shape of the dataset
2. Info & D-types
3. Describe of the application dataset

Data Cleaning

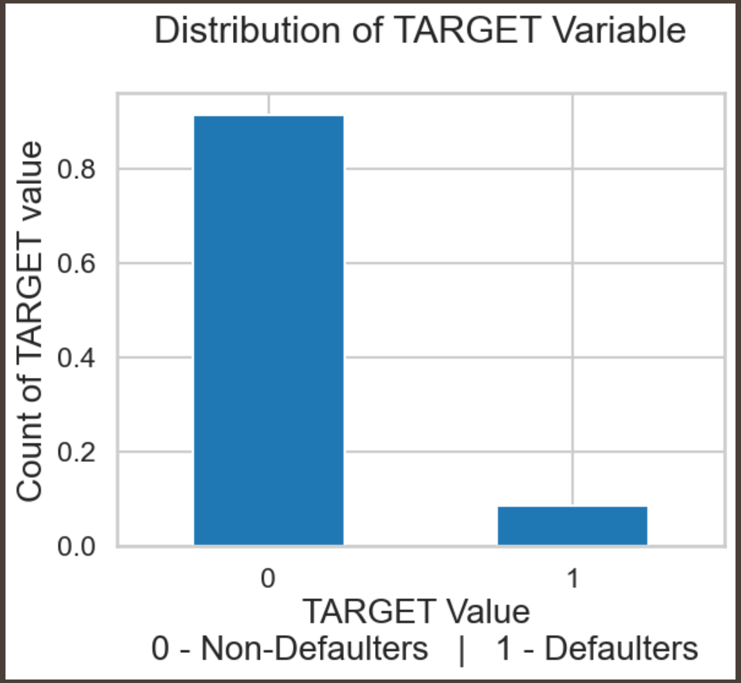
- Identifying the missing values and cleaning them.

1. Initially we listed the null values columns having more than 40% and we found several columns and we dropped them.
2. Remaining columns with the null values we imputed with mean/median /mode values as per the outliers analysis.
3. Outliers: We can see several columns imbedded with outliers namely: AMT_ANNUITY, AMT_GOODS_PRICE, OCCUPATION_TYPE

Plot for defaulters and non-defaulters on basis of Gender

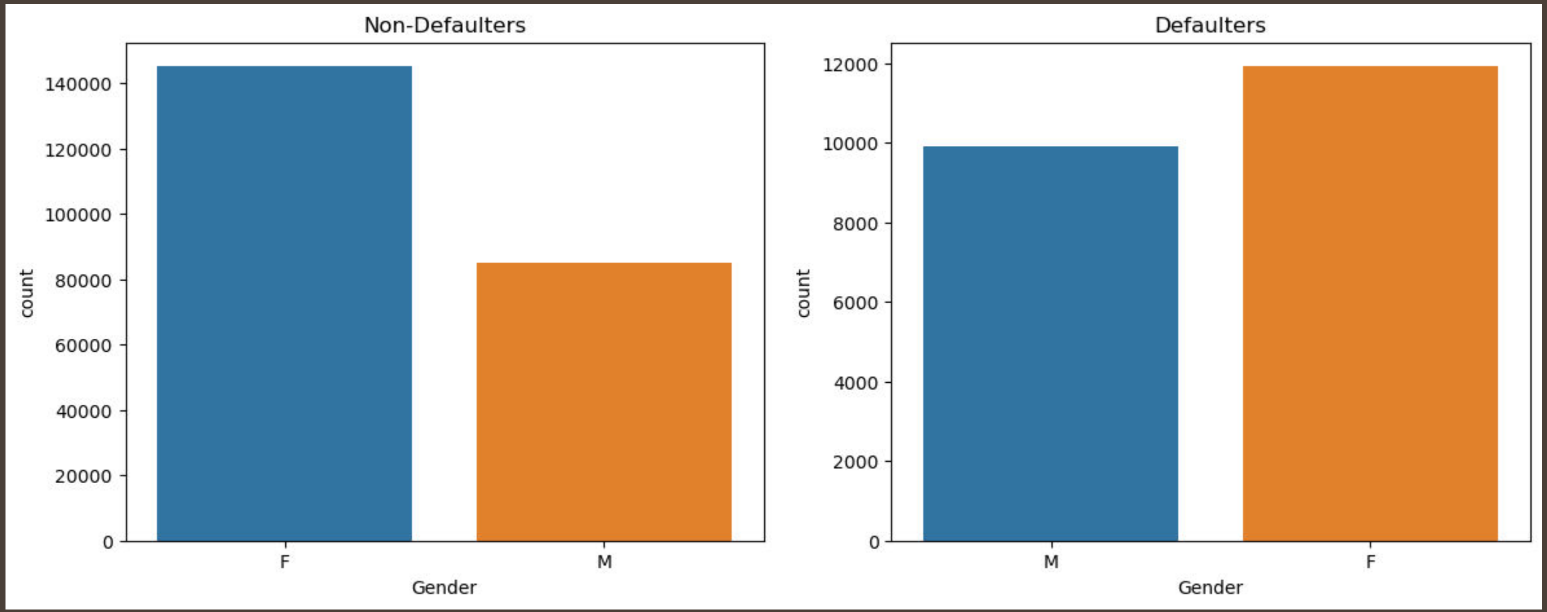


Imbalance Check

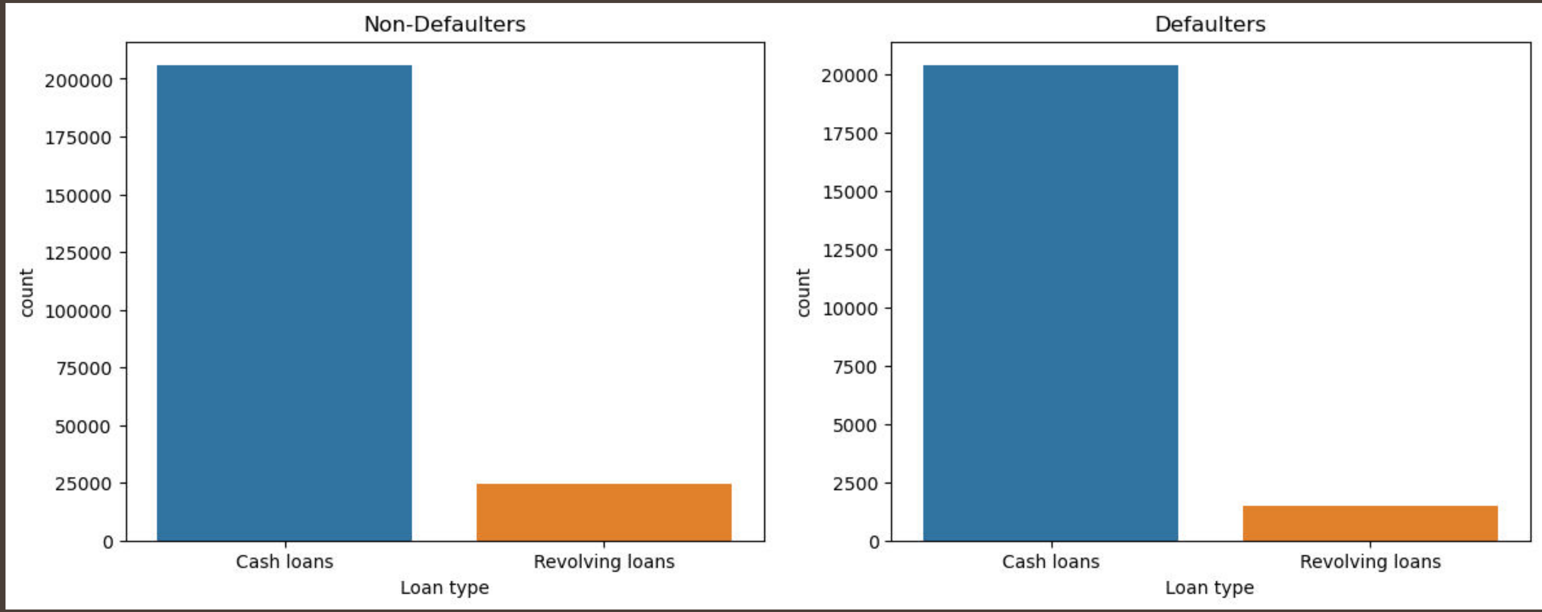


- We can see that the percentage defaulters is varied.
- Target value – (0) : Non – Defaulters
- Target value – (1) : Defaulters
- The ratio of imance is 10.55

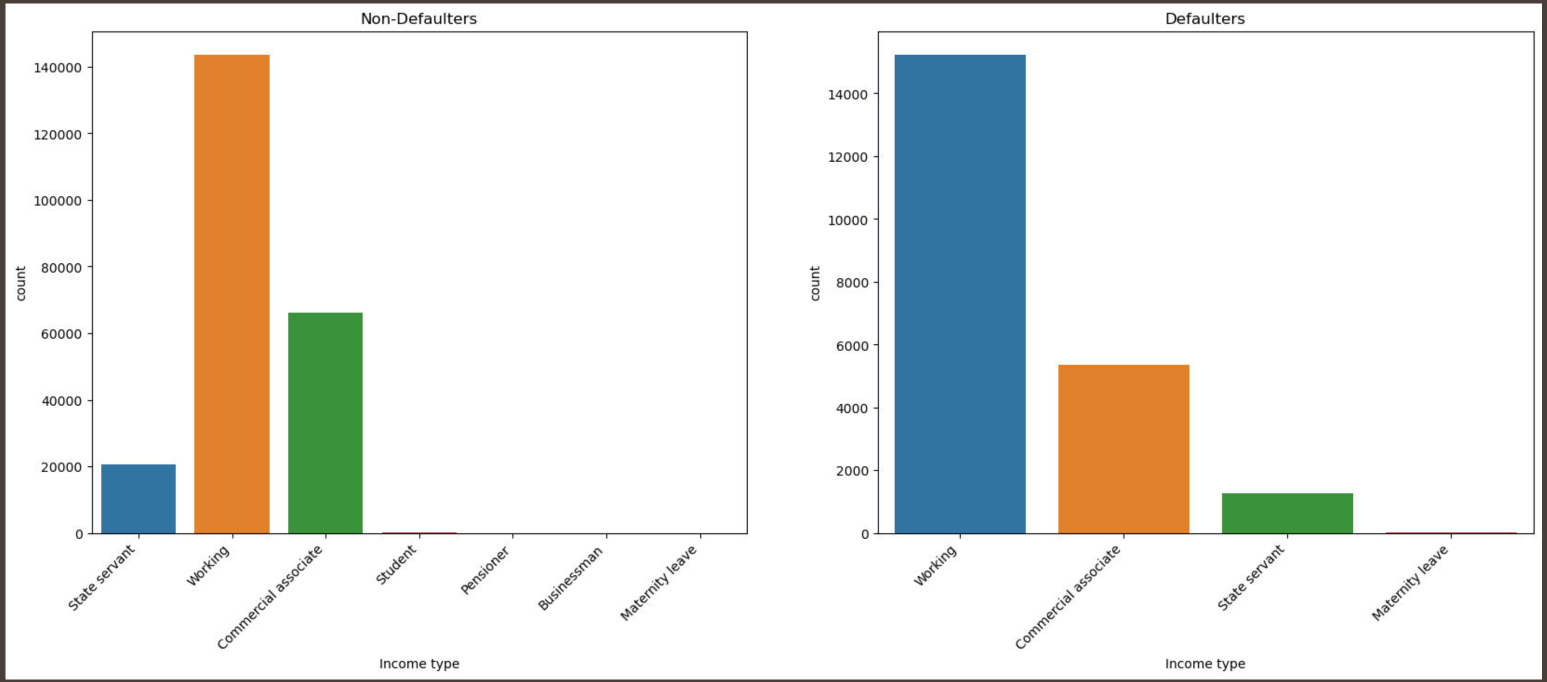
Plot for defaulters and non-defaulters on basis of gender



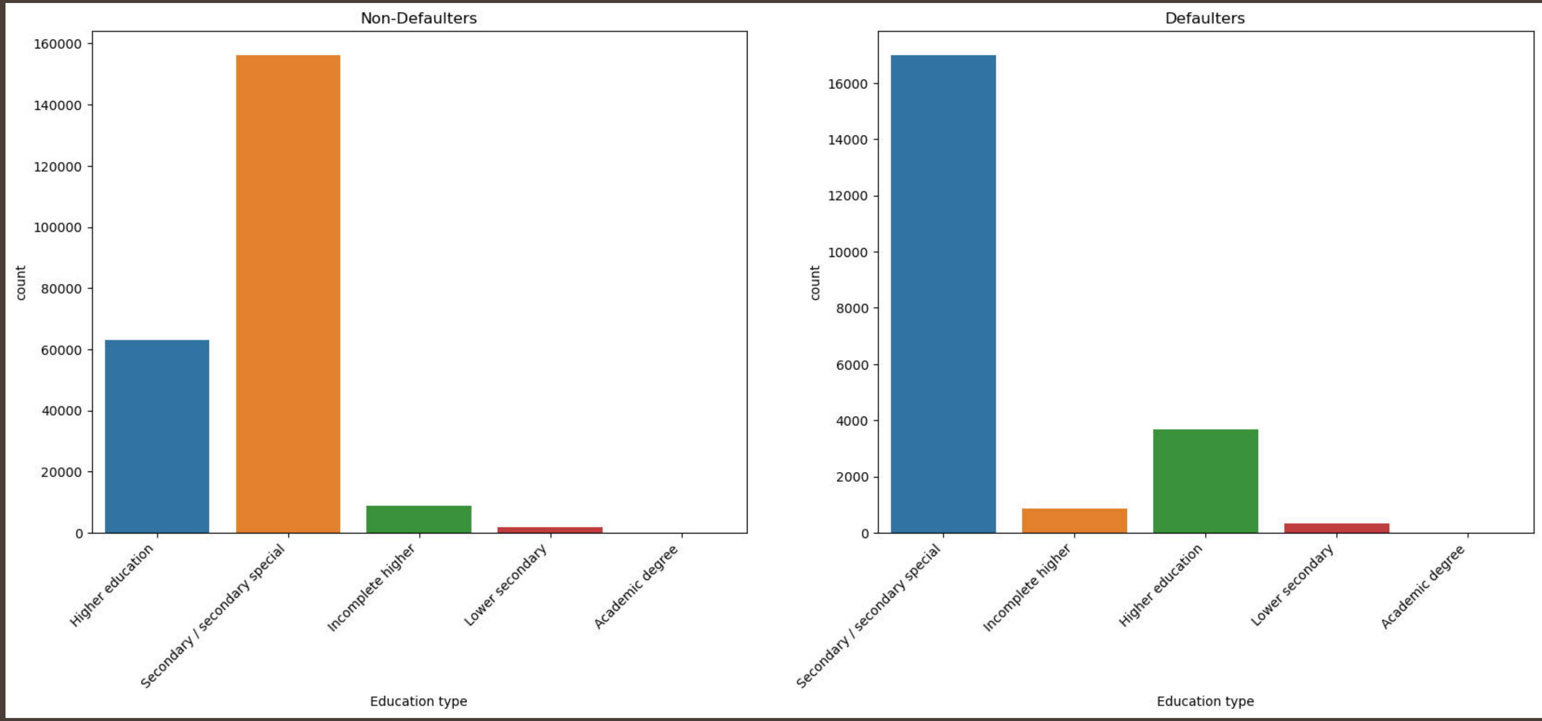
Plotting two plots for defaulters and non defaulters on basis of loan type



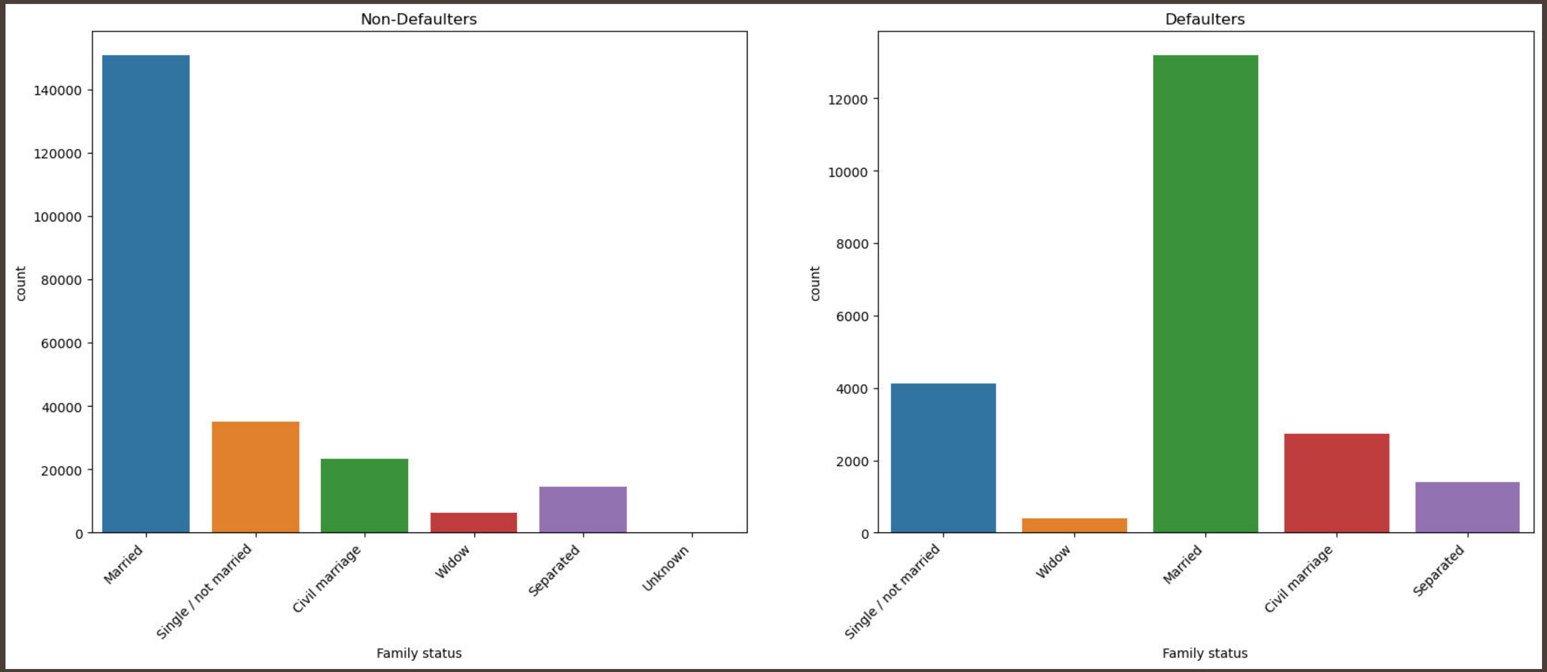
Plotting two plots for defaulters and non defaulters on income type



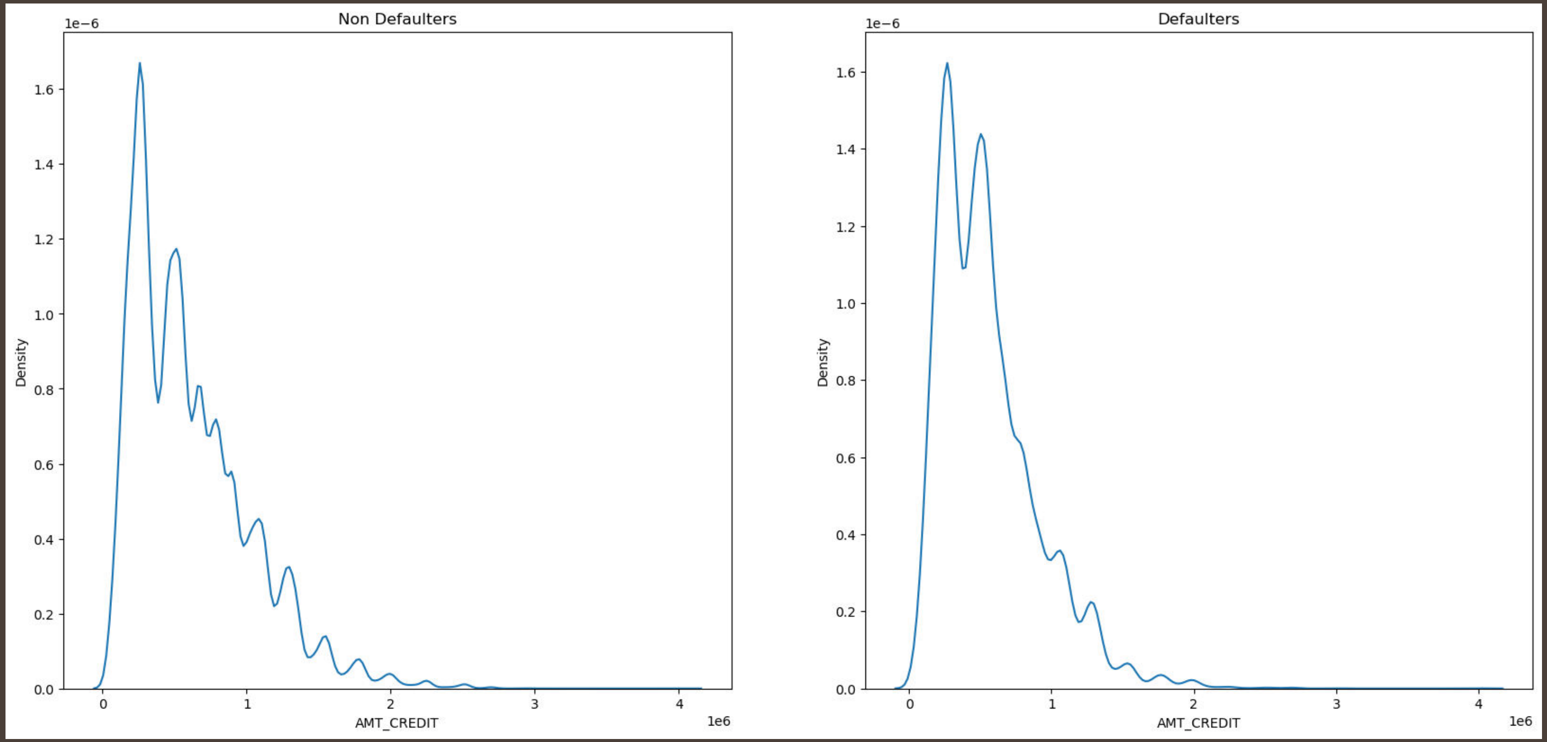
Plotting two plots for defaulters and non defaulters on education type



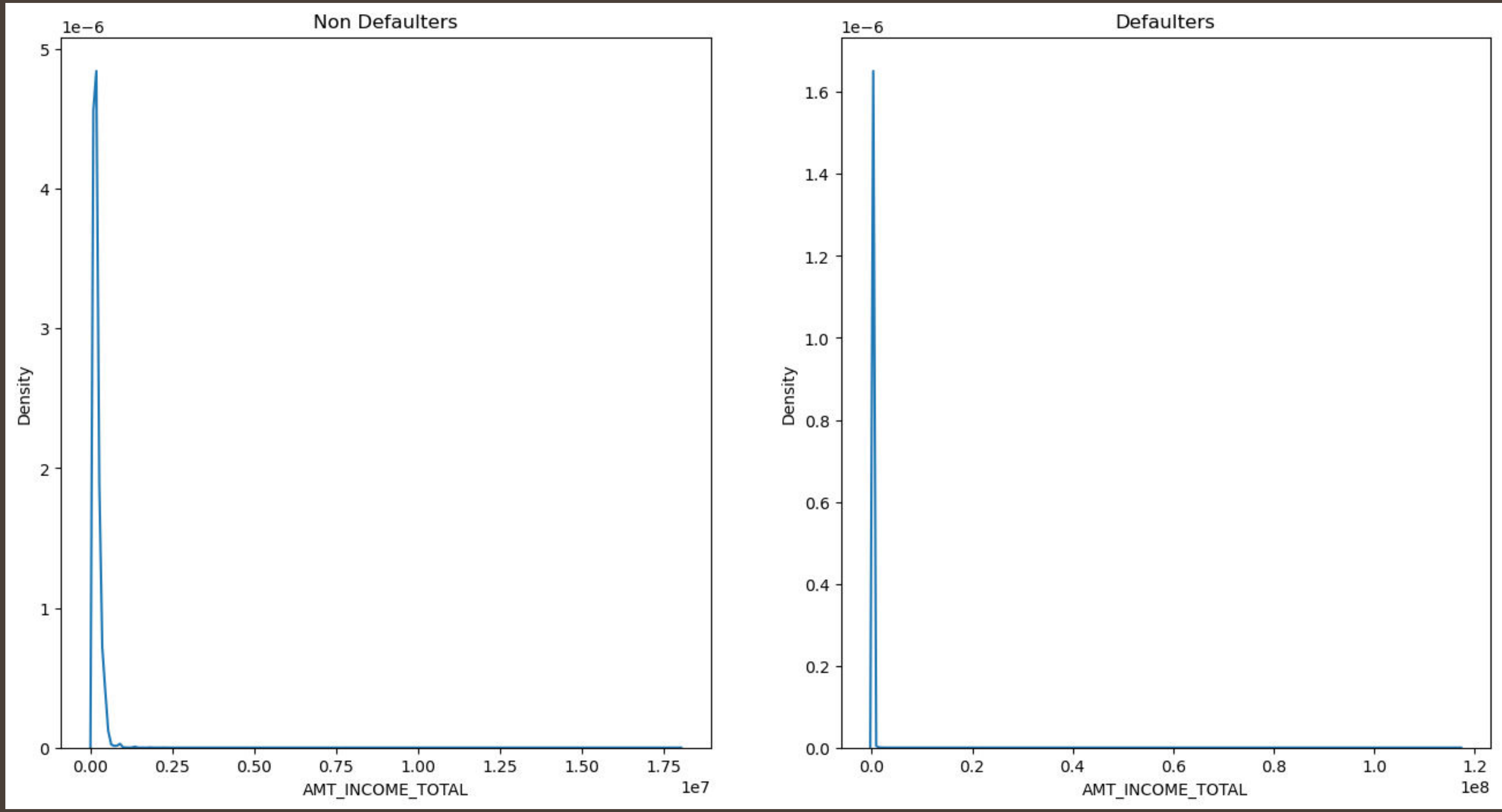
Plotting two plots for defaulters and non defaulters on family type



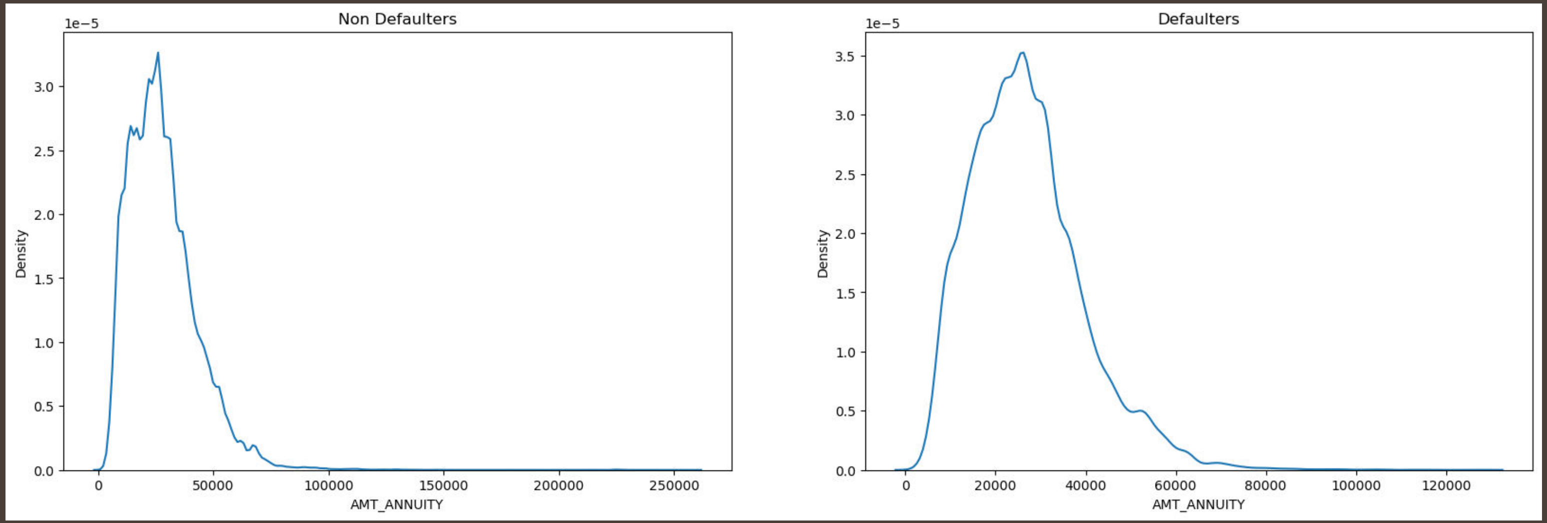
Plotting two plots for defaulters and non defaulters on Credit Amount



Plotting two plots for defaulters and non defaulters on Amount Income Total



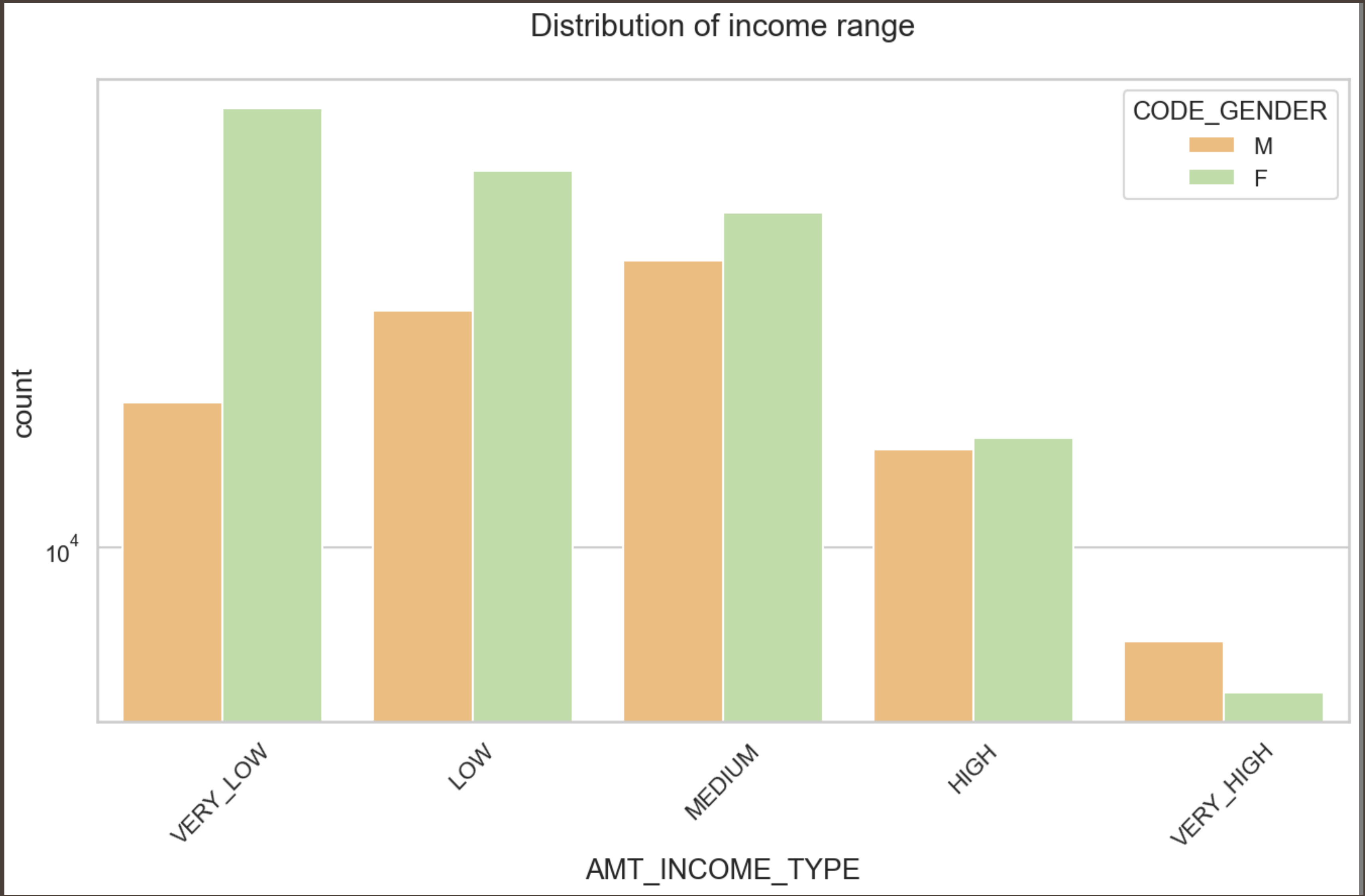
Plotting two plots for defaulters and non defaulters on Amount Annuity



Categorical Univariate Analysis

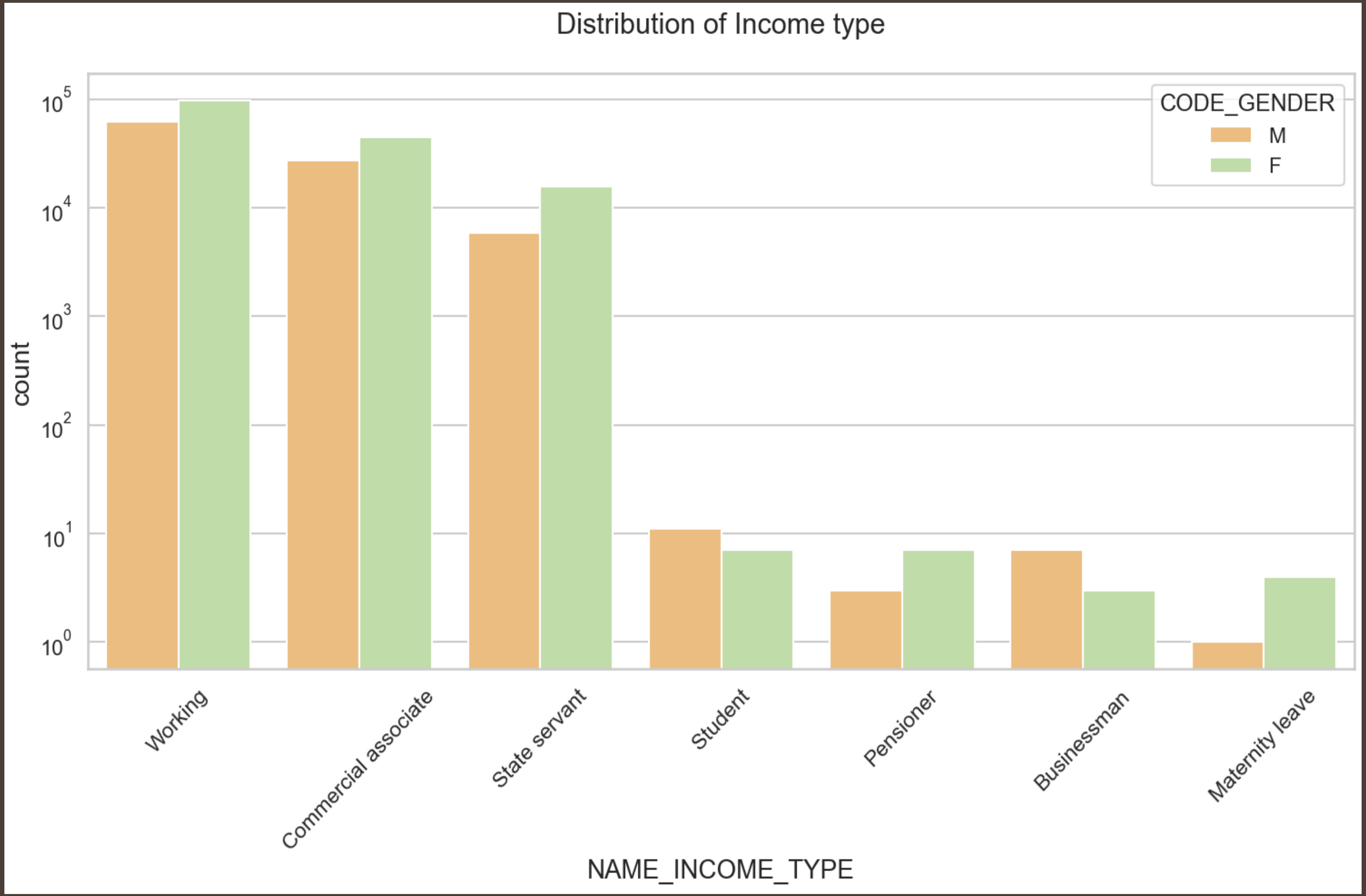
Distribution of Income Range for Target_o

- Female counts are higher than the male count.
- Income range with very-low is having more number of credits.
- This graph shows that females are more than male in having the credits for that range.
- Income range with very-high are having very less count in number.



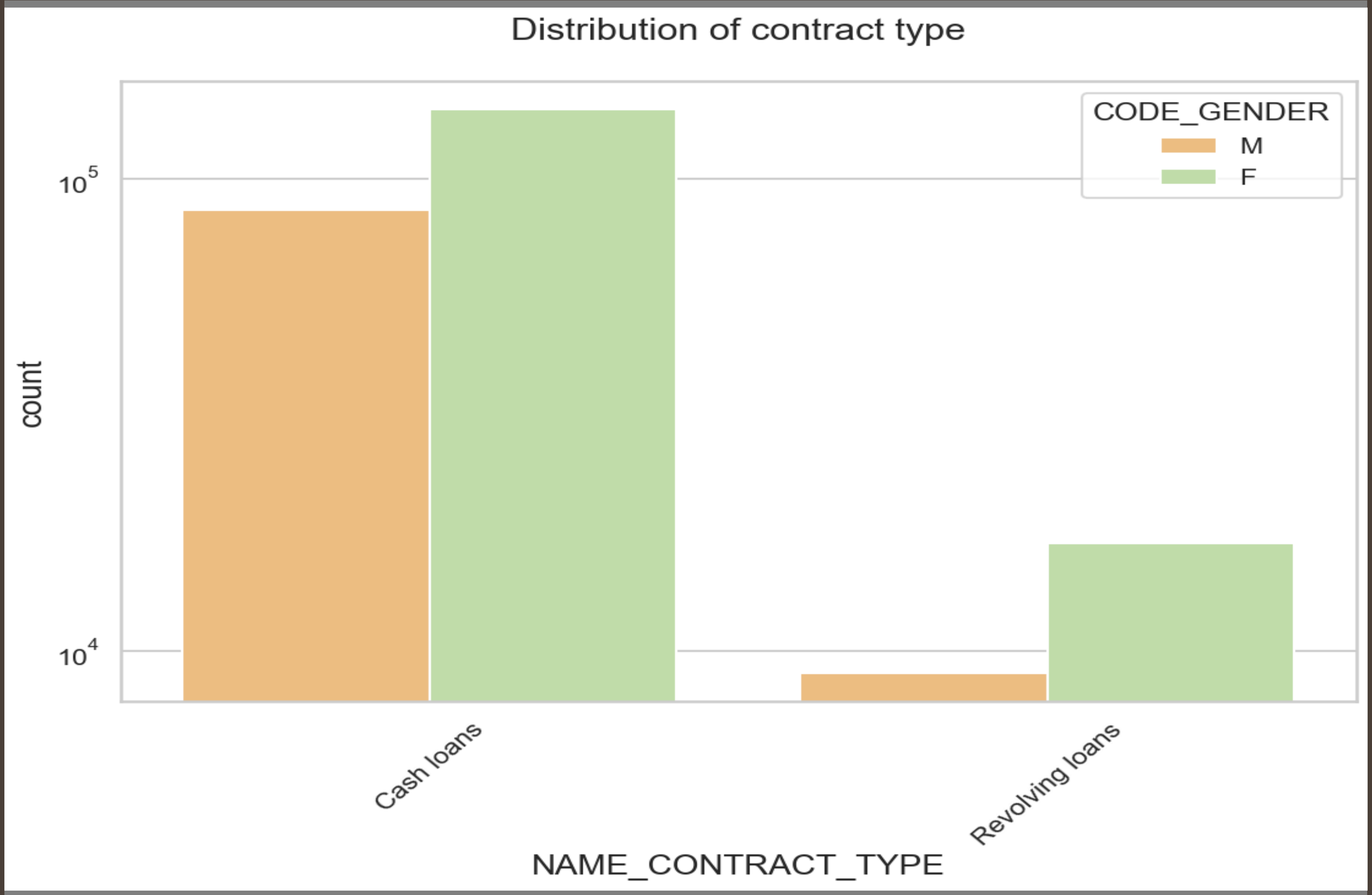
Distribution of Income Type for Target_o

- For income type working, commercial associate and state servant having the highest credits than the others.
- Females are having the more credits than males.
- Less number of credits for student, pensioner, businessman and maternity leave income types.



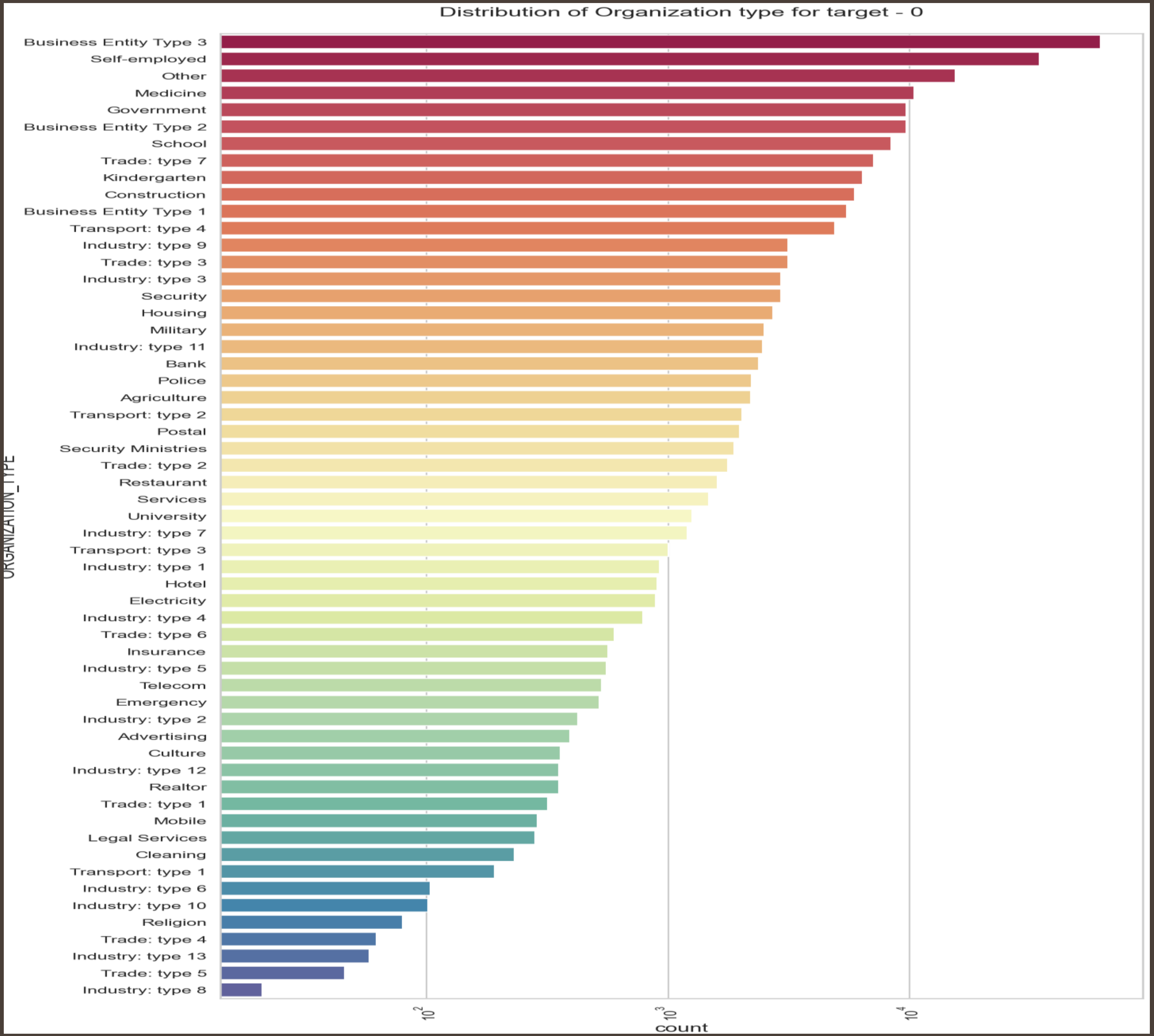
Distribution of Contract Type for Target_o

- Females are having high credits than male.
- Cash loans is having higher number of credits than Revolving loans for this contract type.



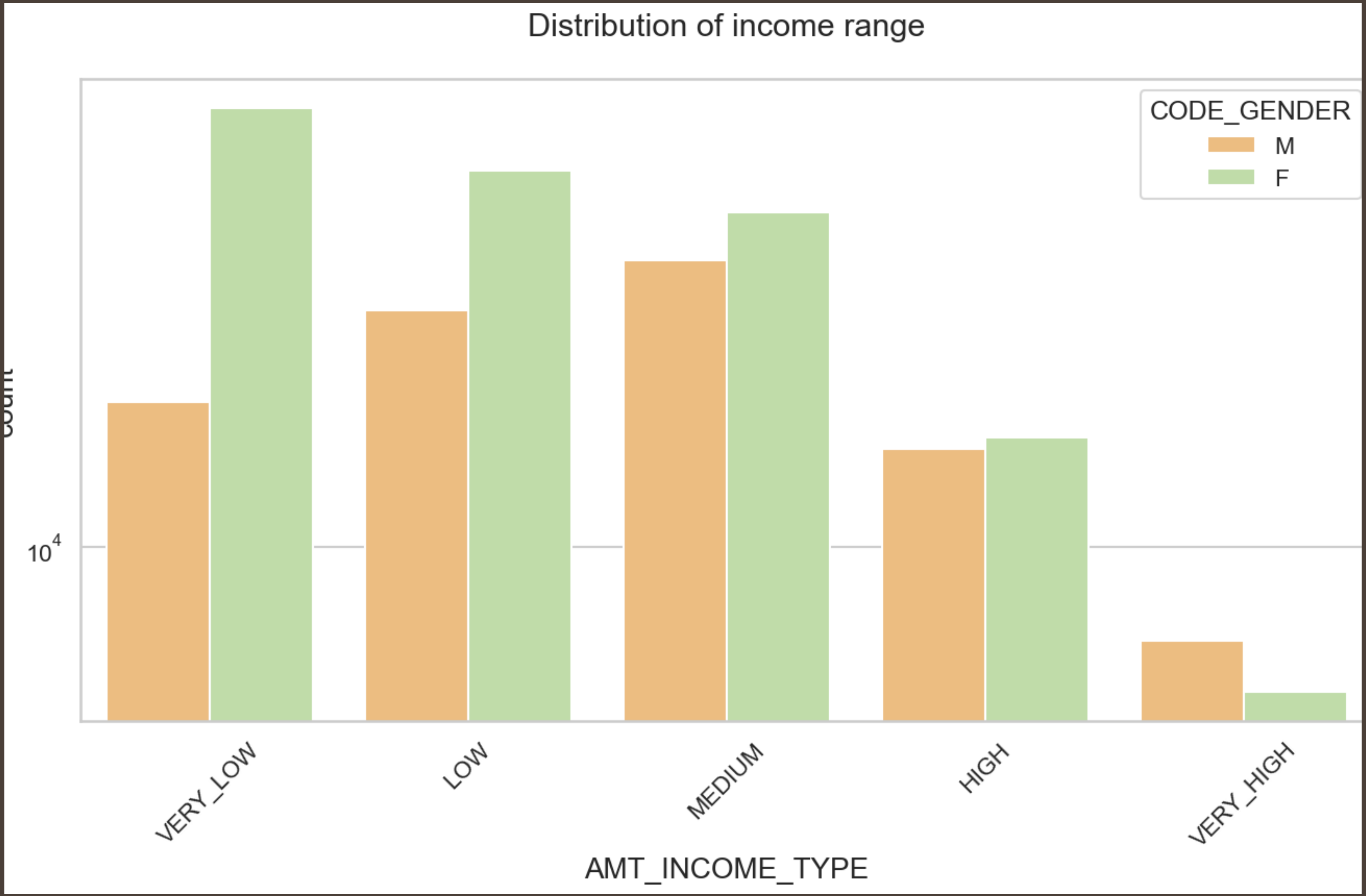
Distribution of Organization Type for Target_0

- The highest credits applied from most of the Business entity Type 3, Self employed, Other, Medicine and Government.
- The less credits are from industry type 8, type 6, type 10, religion and trade type 5, type 4.



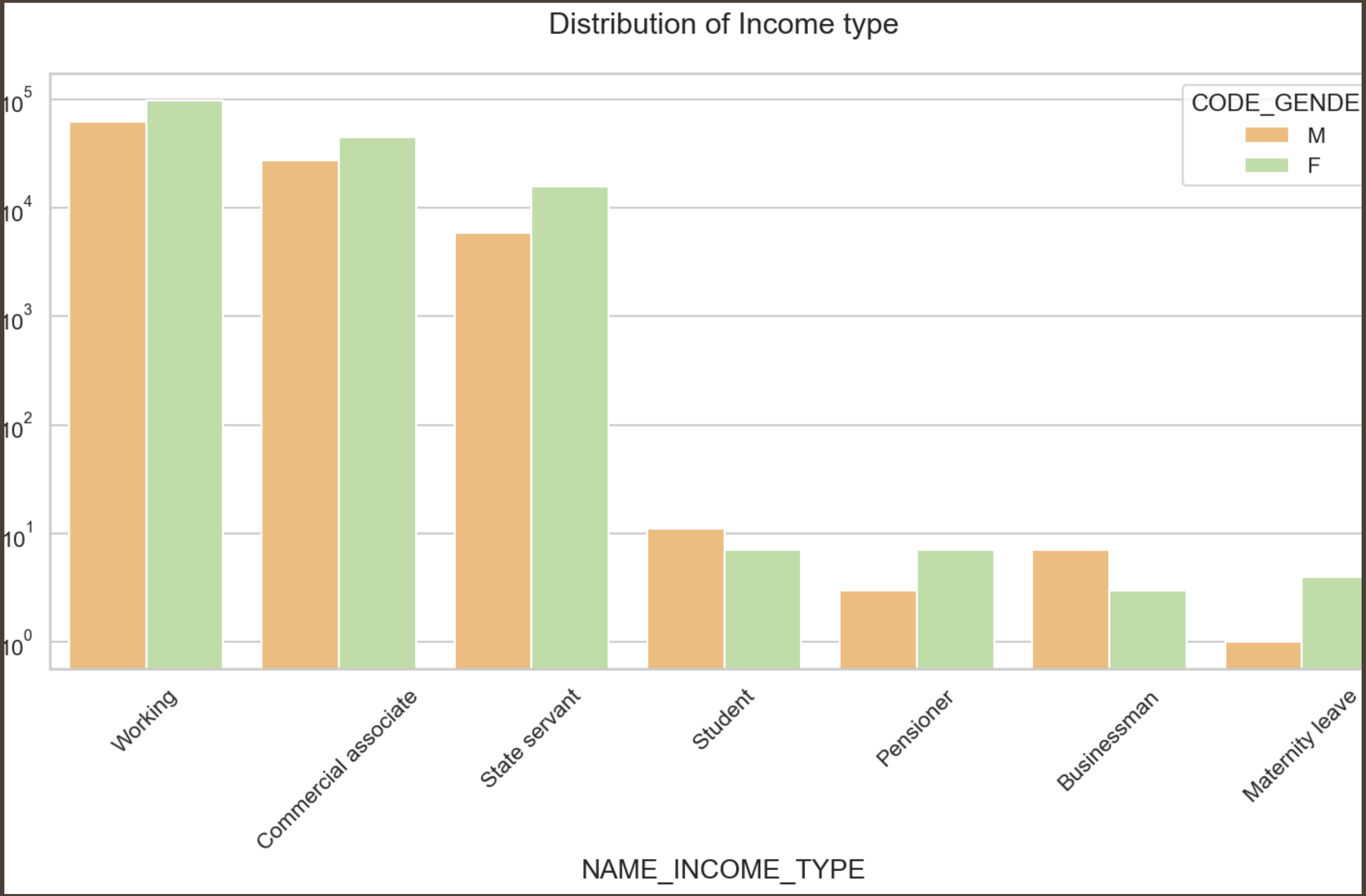
Distribution of Income Range for Target_1

- Income range with very-low are having higher number of credits.
- Very less count for the income type with very-high



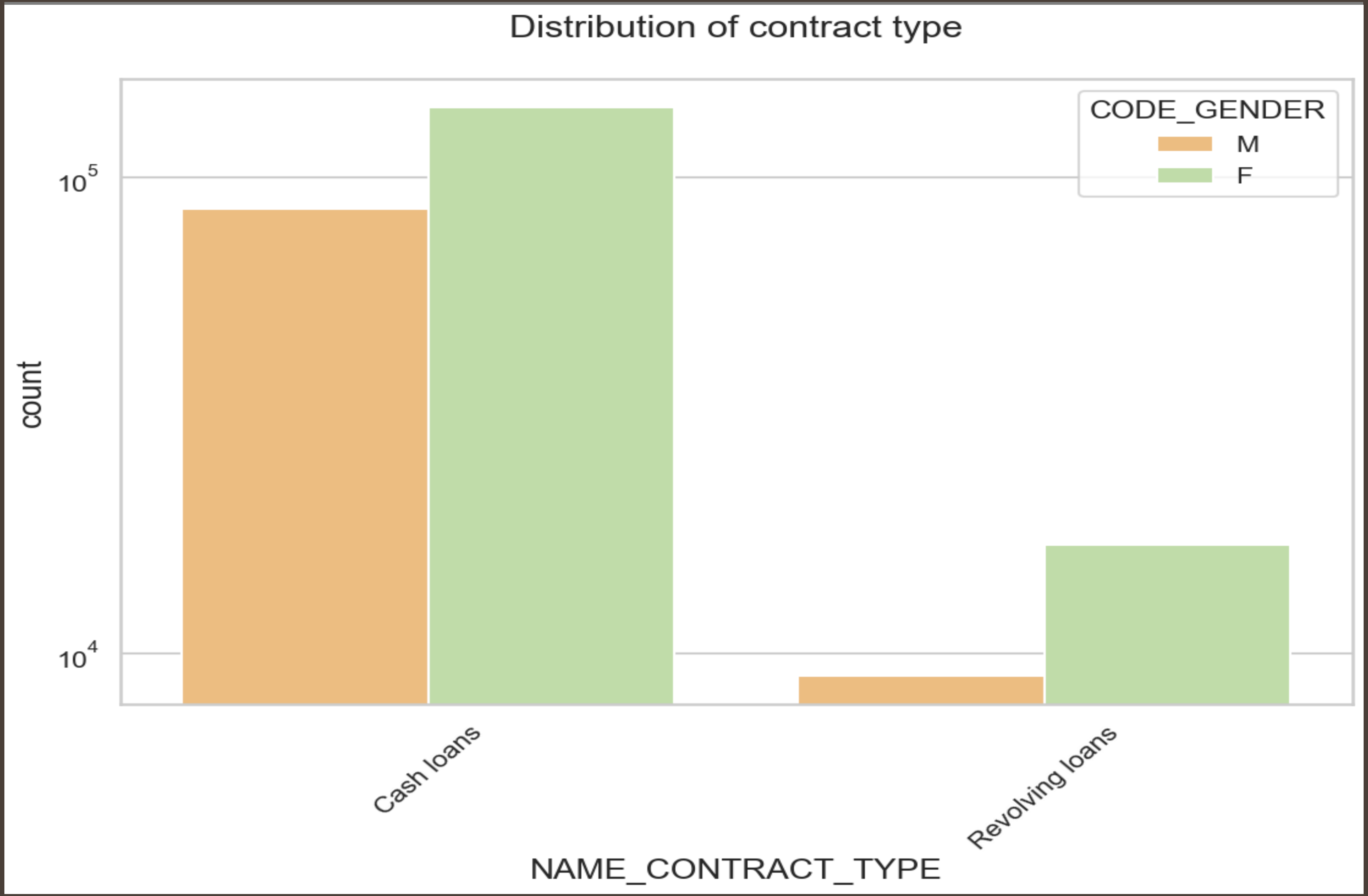
Distribution of Income Type for Target_1

- Income types namely working, commercial associate and state servant having the higher number of credits than other..
- Maternity leave having the less number of credits.
- Females are having higher number of credits than male.
- For the income types namely student, pensioner and businessman are not having any credits.



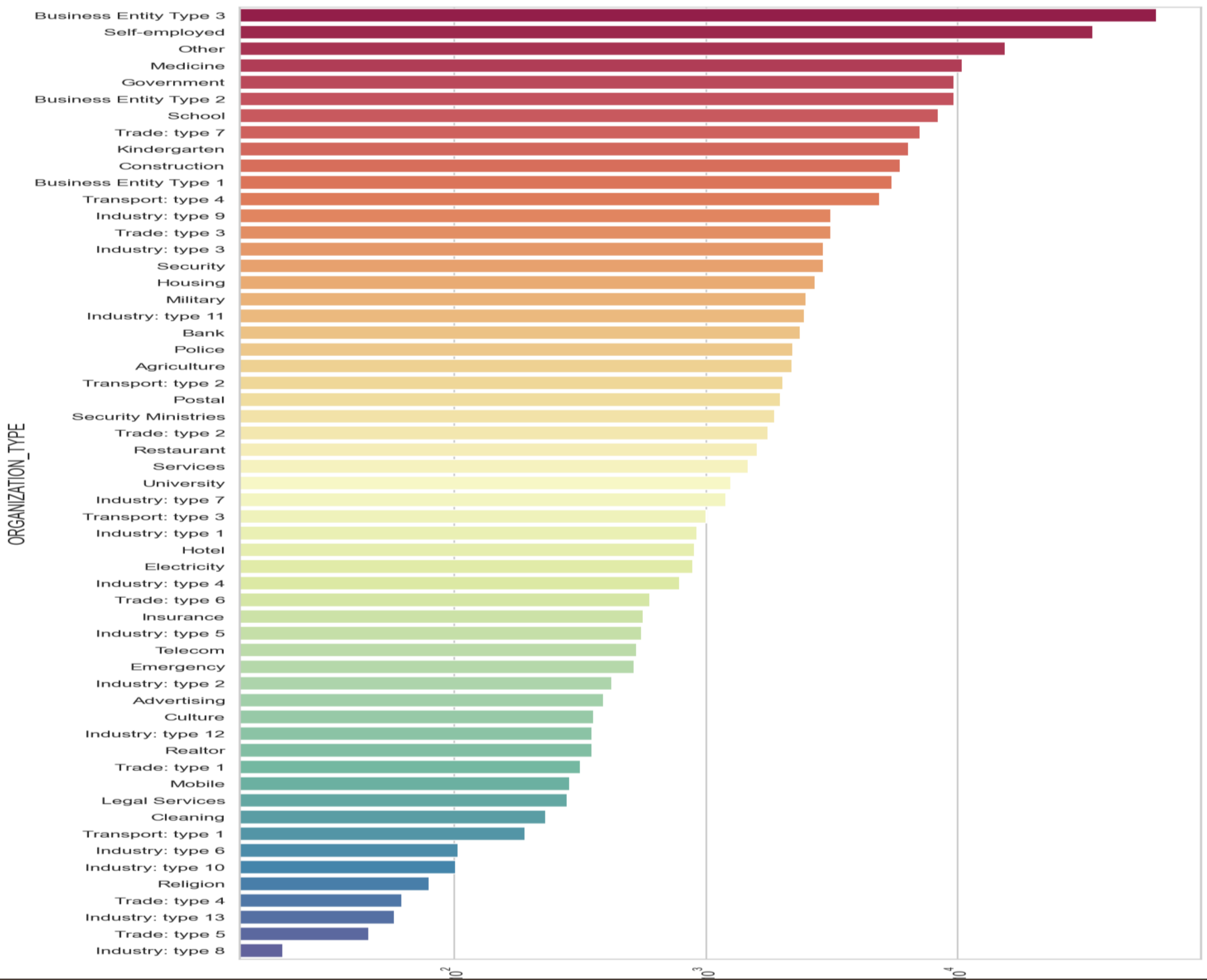
Distribution of Contract Type for Target_1

- Cash loans are having higher number of credits than the Revolving loans.
- There is only female Revolving loans in this type-1
- Females are having higher credits applied for.



Distribution of Organization Type for Target_1

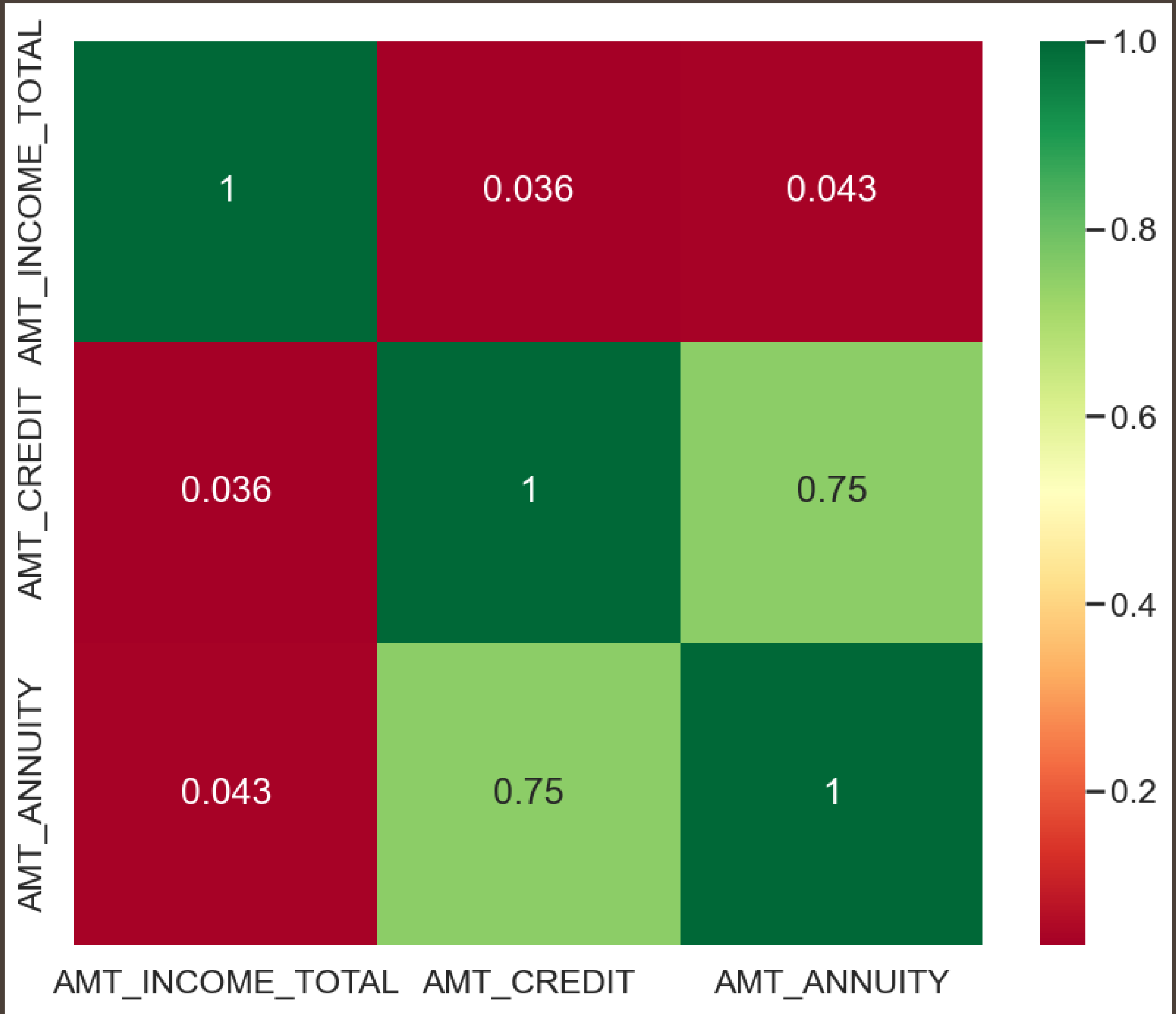
- The highest credits applied from most of the Business entity Type 3, Self employed, Other, Medicine and Government.
- The less credits are from industry type 8, type 6, type 10, religion and trade type 5, type 4.
- Same as target-0.



Correlation Matrix for Target-0 & Target-1

Correlation Matrix for Target_o

Correlation Matrix for Target_1

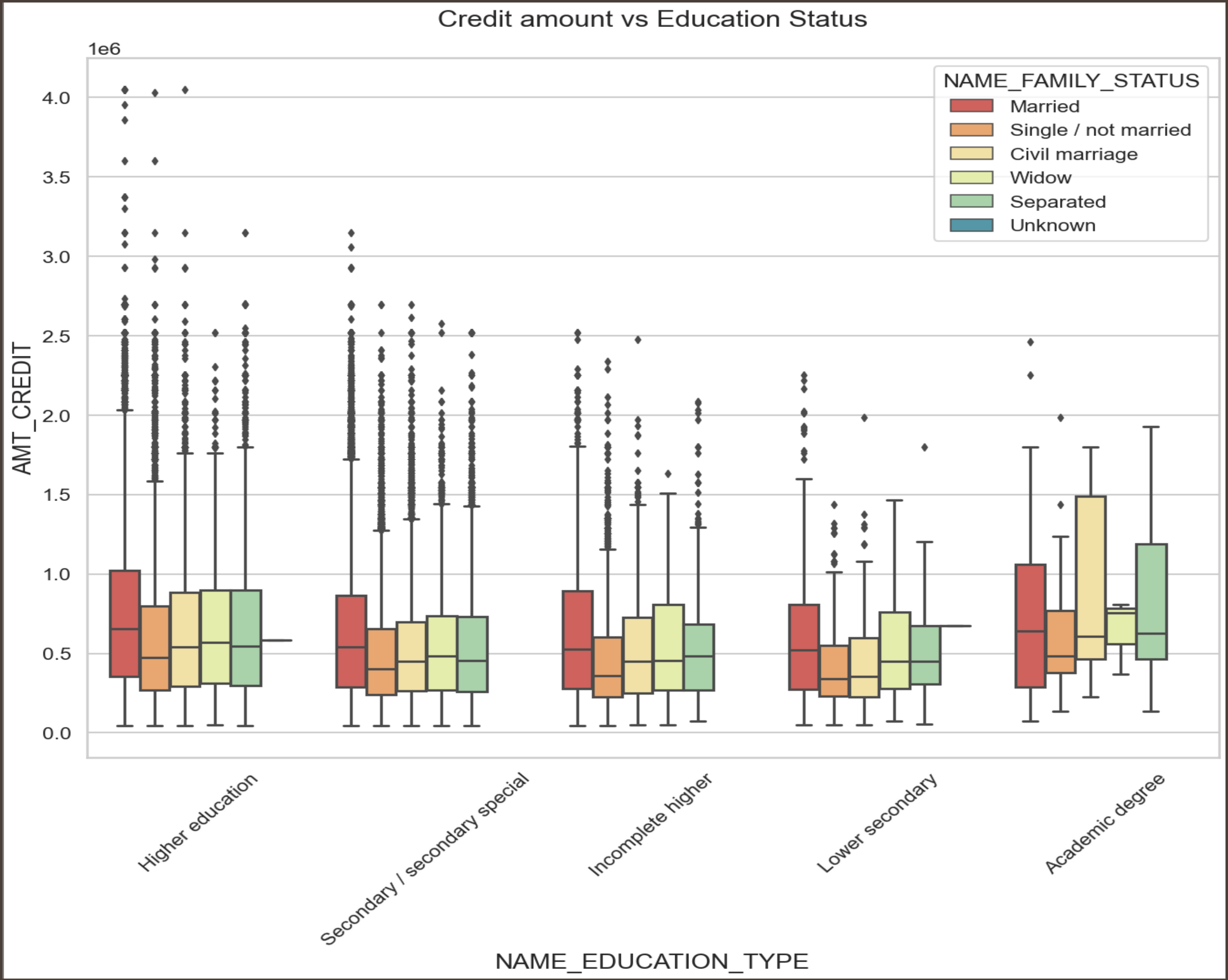


Bivariate Analysis

(Categorical to Numerical Values)

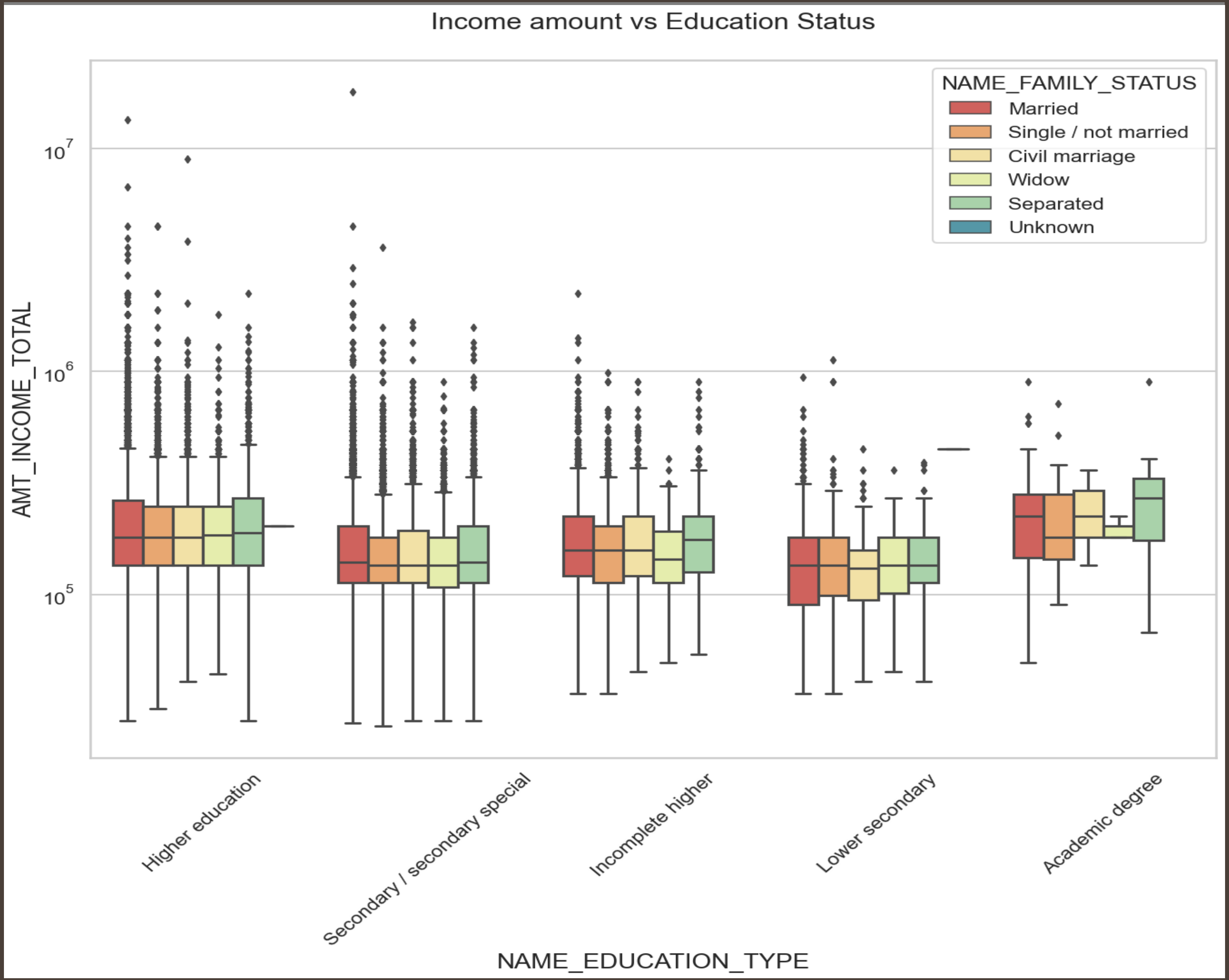
Credit Amount Vs Education Status

- Academic degree education for civil marriage, marriage and separated Family status are having higher number of credits.
- For the marriage, single and civil marriage with higher education are having more outliers.



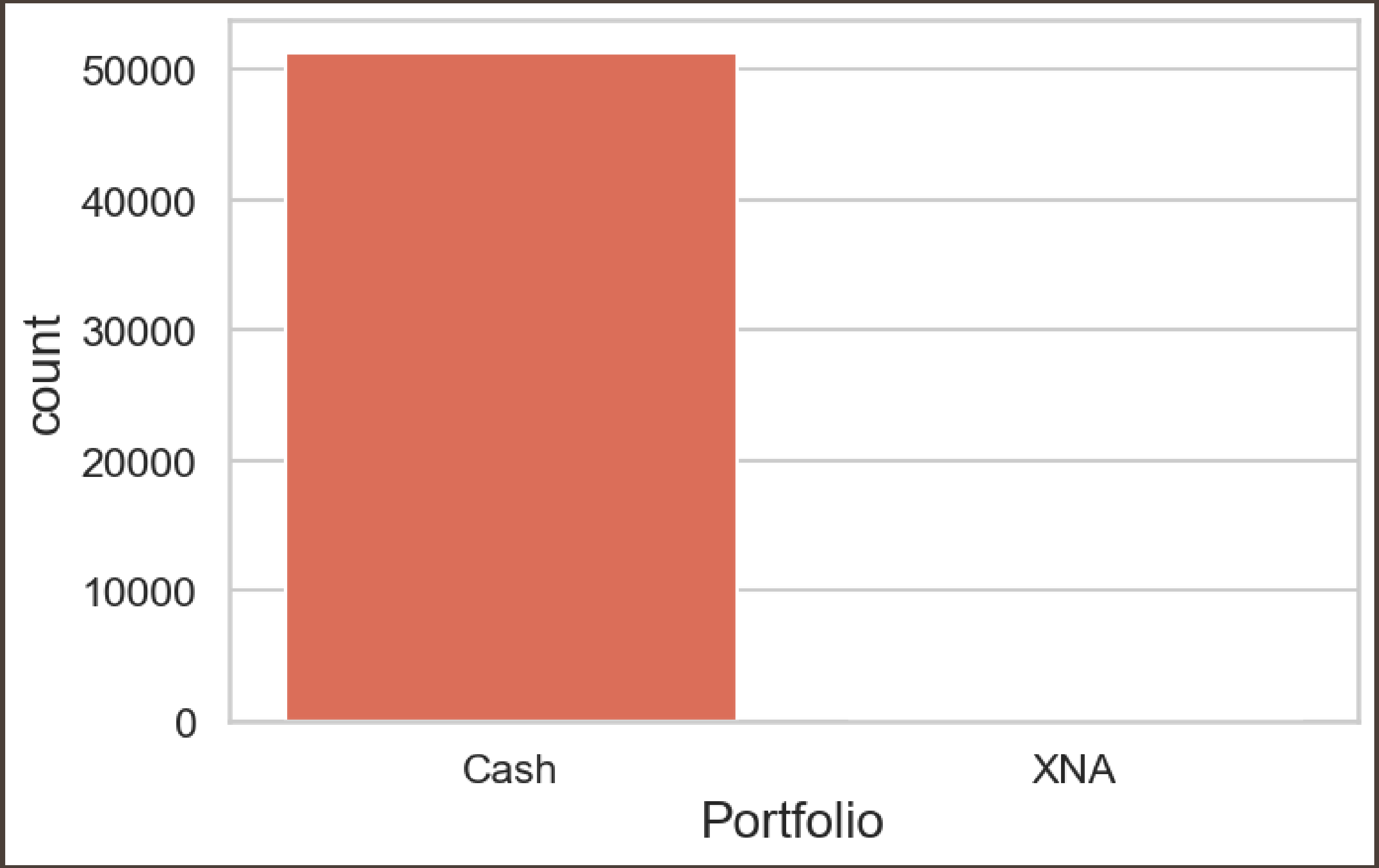
Income Amount Vs Education Status

- Less outliers are having for Academic degree education but they are having the income amount is inline higher than that Higher education.
- For Higher education the income amount mean is mostly equal with family status.

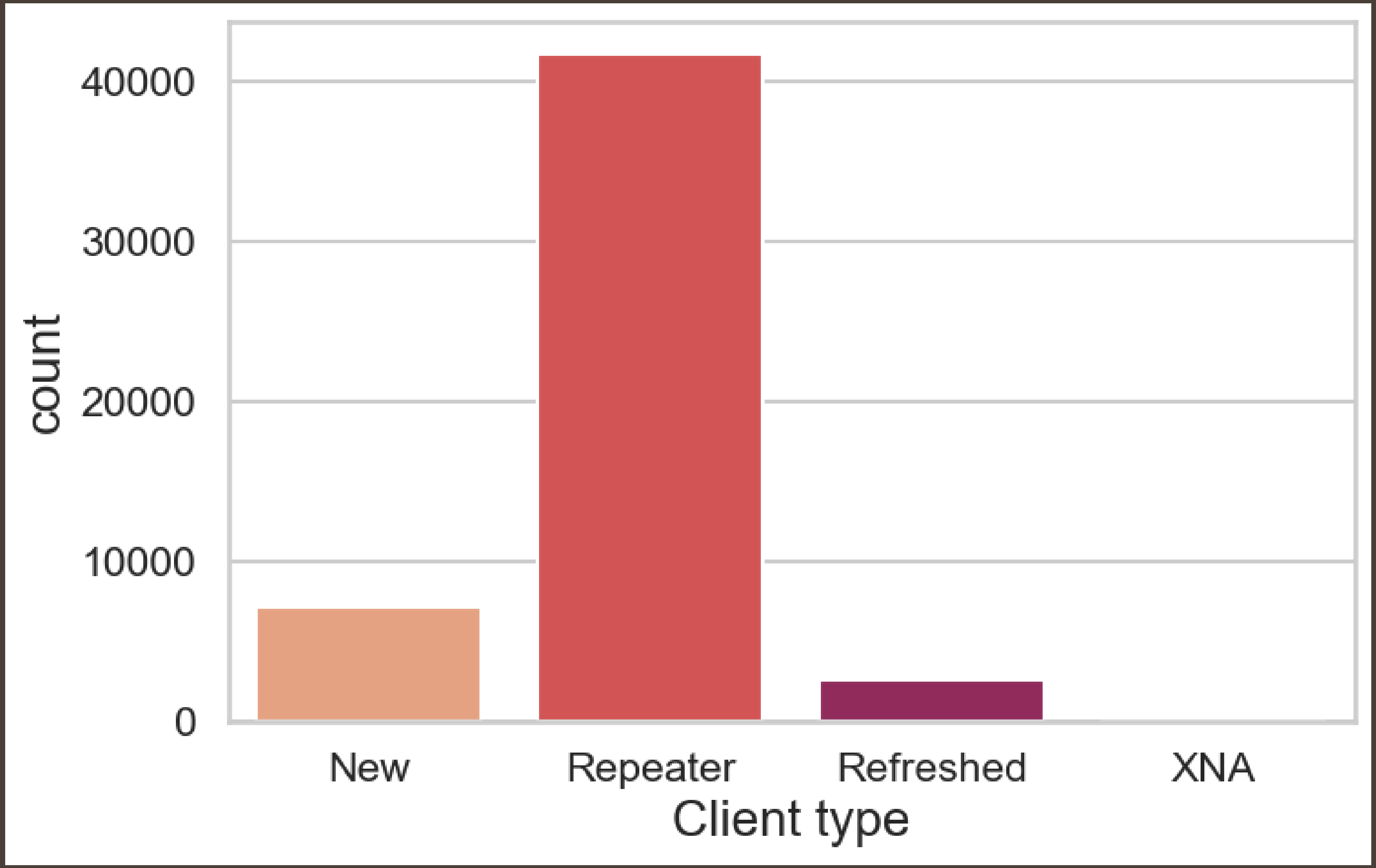


Previous Application Dataset Analysis

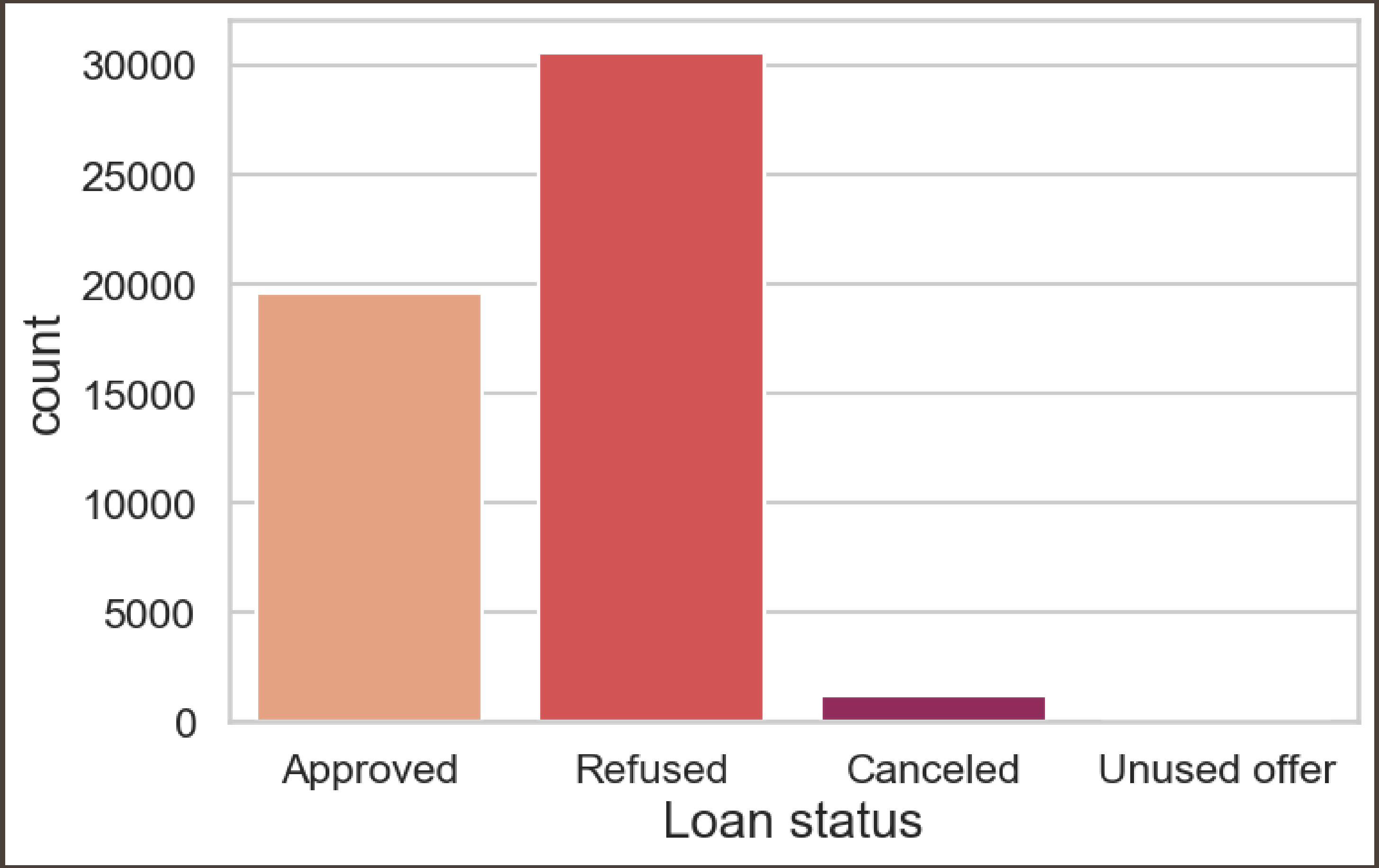
Previous application dataset - Portfolio



Previous application dataset – Client Type

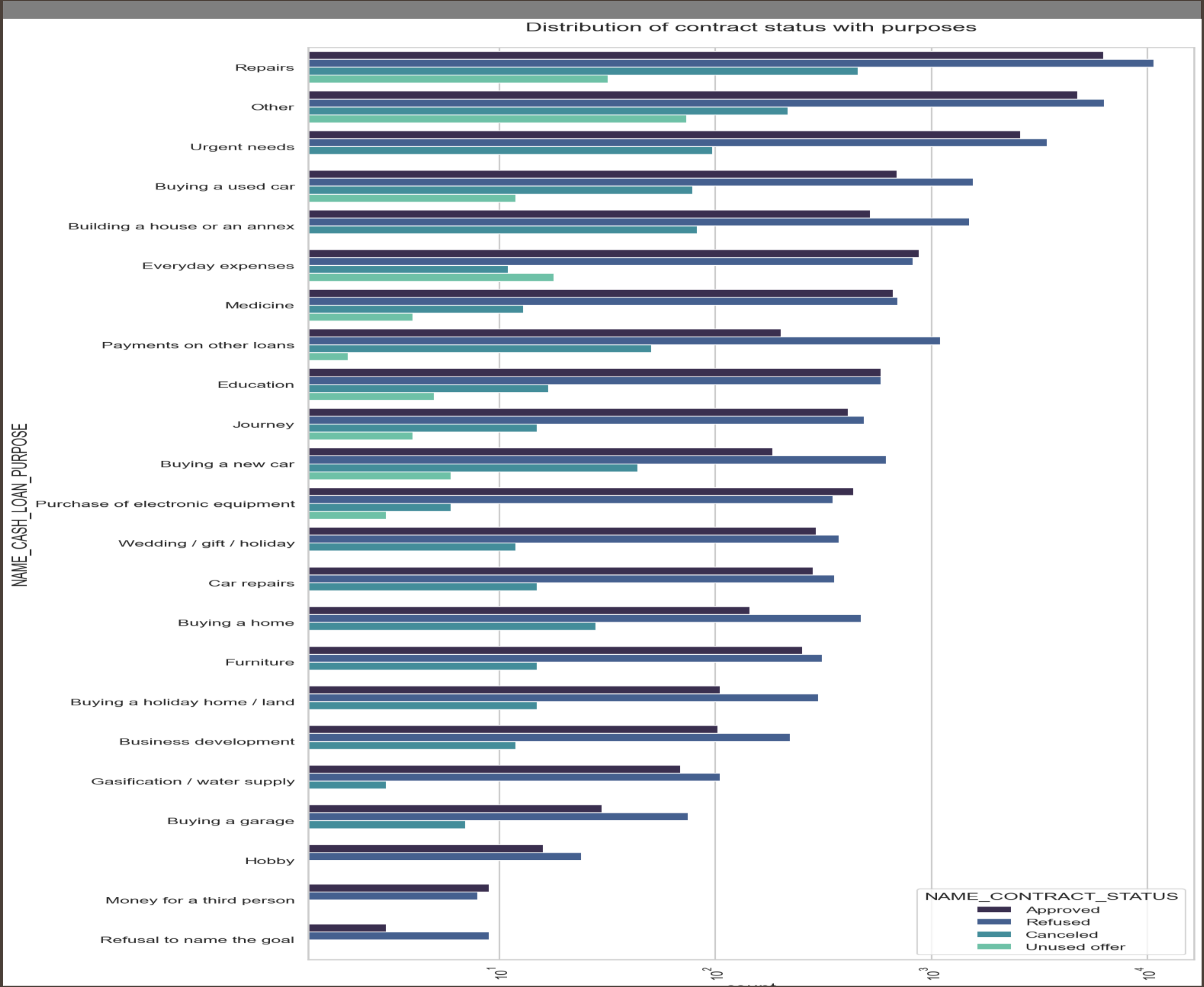


Previous application dataset – Lone Status



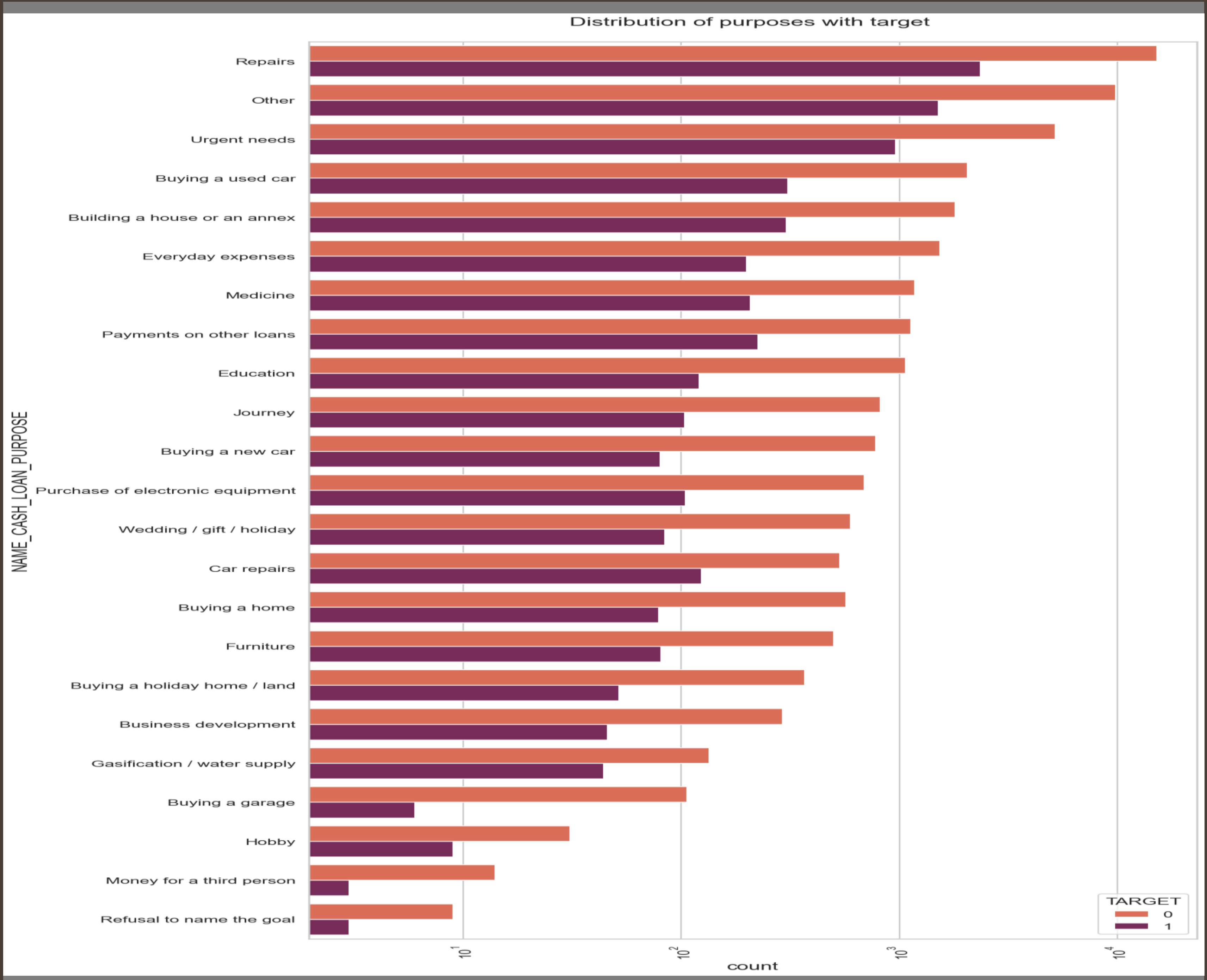
Distribution of contract status with purposes

- Repairs are having the most number of rejections.
- The education purposes the approvals and rejections are same in number.
- Paying other loans and buying a new car are having higher rejections than approvals.



Distribution of purposes with target

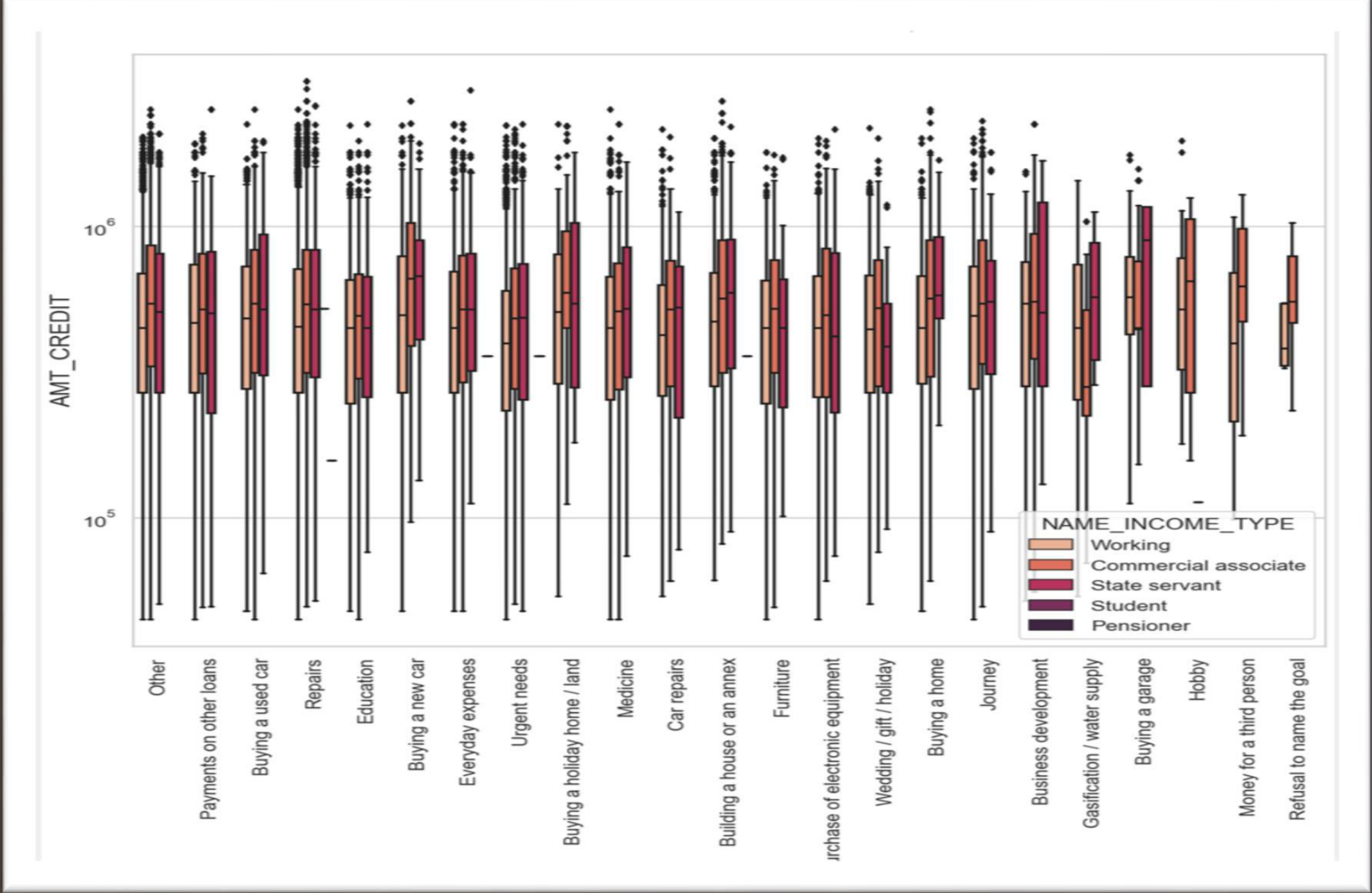
- Repairs are having more difficulties for repay the loan.
- Busing a garage, Business development, Buying land, Buying a new car and Education are the best in repaying the loans.



Previous Credit Type vs Loan Purpose

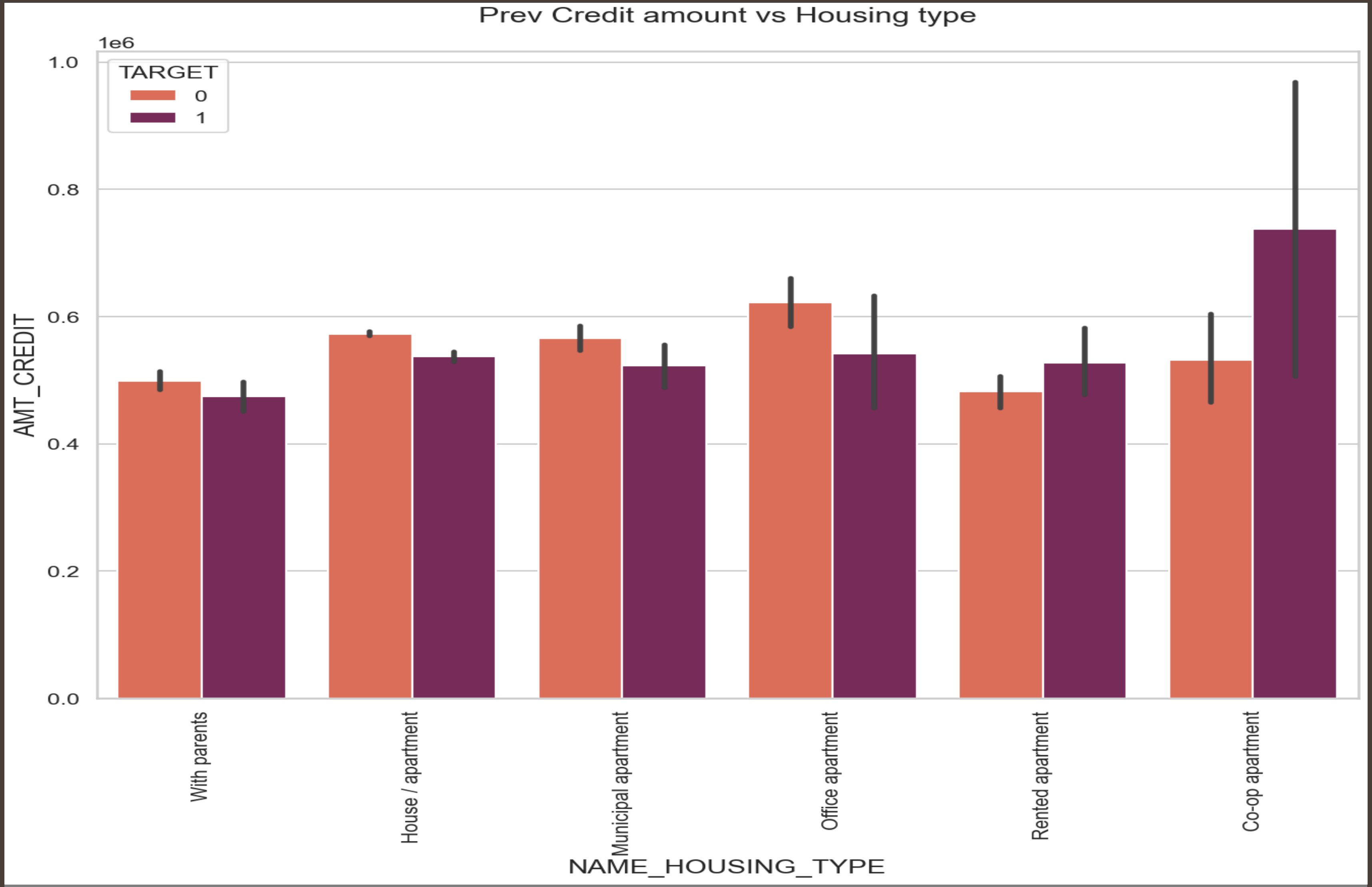
From the previous graph we can conclude

- The credit amount of Loan Purposes like Buying a home, a land, a new car & Building a house is higher.
- Income type of state servants have a significant amount of credit applied
- Money for 3rd person is having less credits applied for.

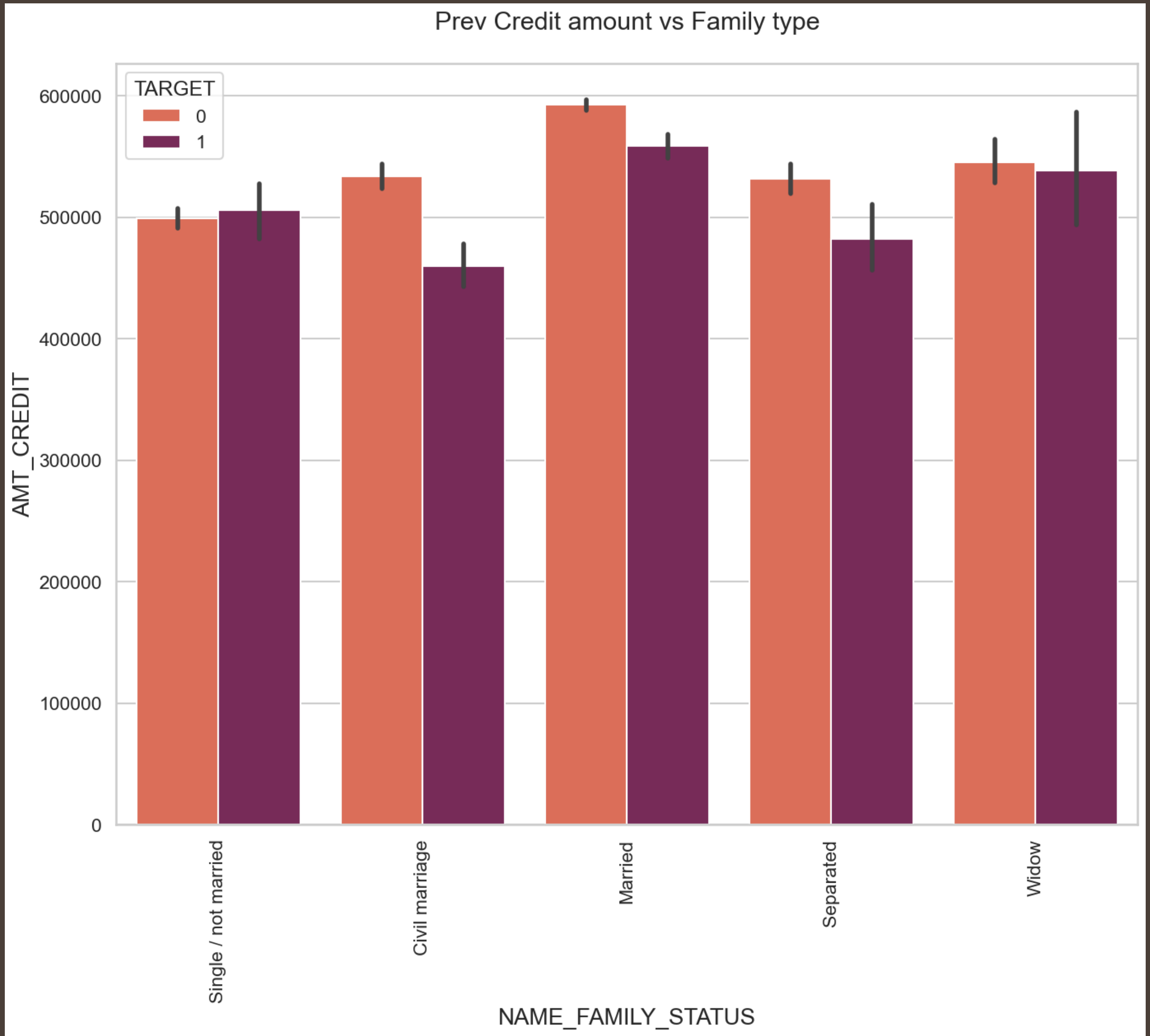


Previous Credit Type vs Housing Type

- For housing type, office apartment is having higher credit of T(0) is having higher credit of T(1).
- So, we can conclude that bank should avoid giving loans to the housing type of co-op apartments as they are having difficulties in payment.
- Banks should focus on housing or municipal apartments for successful payments.



Previous Credit Type vs Family Type



Final Analysis

- Finally we can conclude by seeing all the results:
- 1. Banks should focus on contract type like: Student, Pensioner and Businessman with Income type for successful payments.
- 2. Banks should focus less on Income type like: working Category as they having more unsuccessful payments.
- 3. Banks should focus more on Female applicants as we saw the graphs for defaulters and non-defaulters with CODE GENDER.
- 4. (Income type & Credit amount) and the no.of childern clients have is inversely proportional.
- 5. The Imbalance ratio is 10.55
- 6. Loan purpose with Repair is having more no.of unsuccessful payments.
- 7. As much as clients from housing type with parents as they are having least no.of unsuccessful payments.

Thanks.!