

CREDIT EDA CASE STUDY

Prepared By

Pritamkumar Suryavanshi

Abhishek N R

PROBLEM STATEMENT

Two types of risk's are associated with Bank's decision :

- If the applicant is likely to repay the loan , then not approving the loan results in loss of business to the company.
- If the applicant is not likely to repay the loan , i.e. he/she is likely to default , then approving the loan may lead to a financial loss for the company.

When a client applies for a loan, there are four types of decisions that could be taken by the client/company

- i. Approved
- ii. Cancelled
- iii. Refused
- iv. Unused Offer

Dataset consist of 2 files.

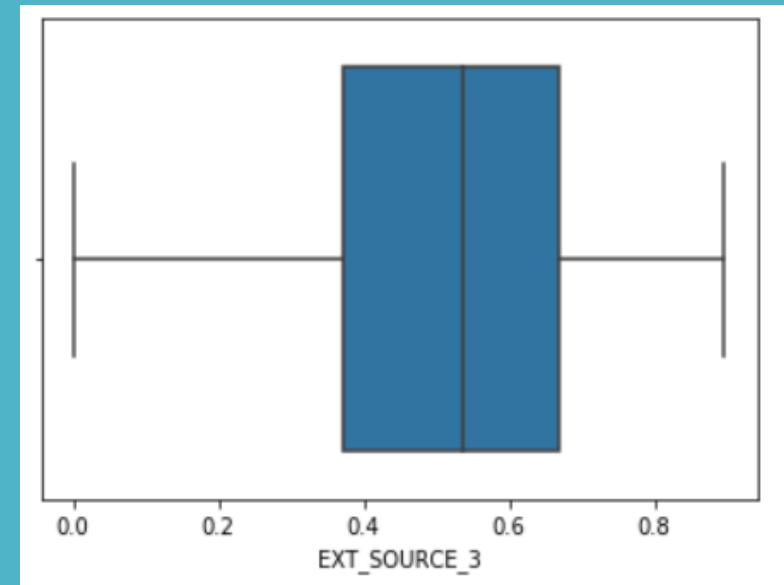
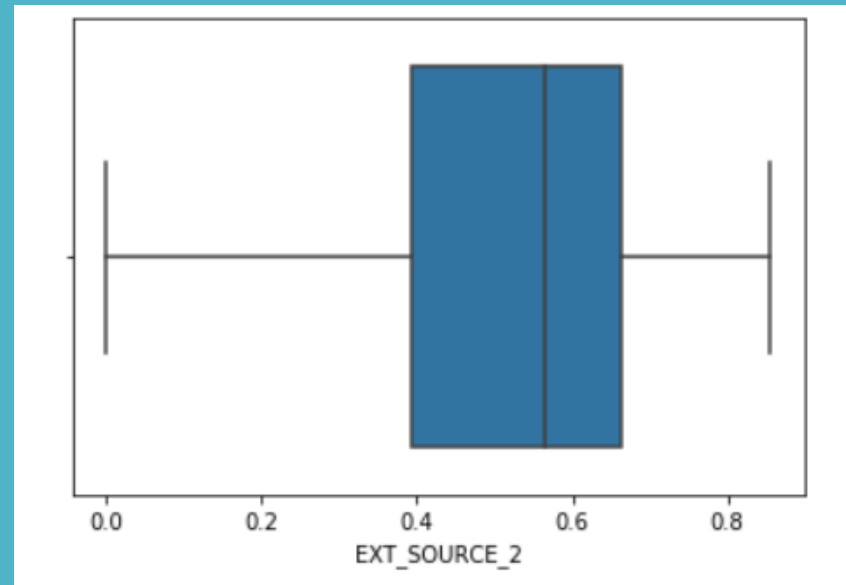
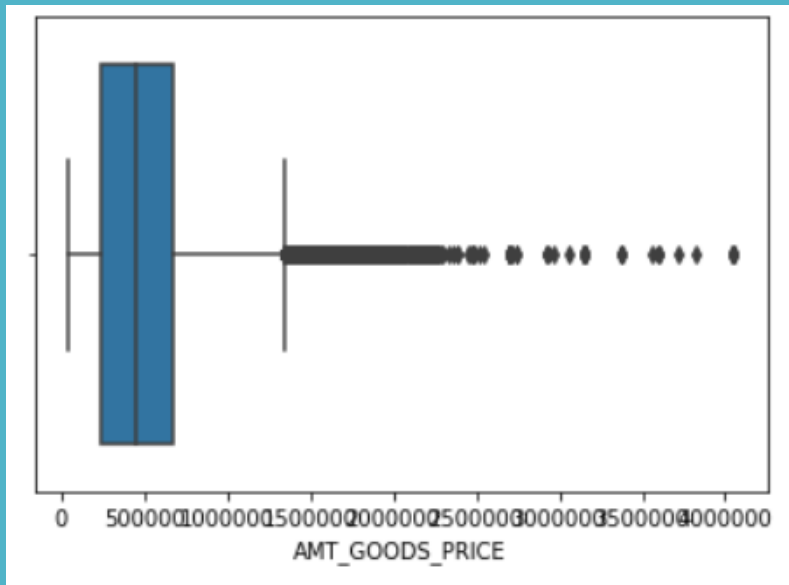
1. Application Data – contains client has payment difficulties.
2. Previous application – contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.

The company wants to understand the driving factors behind the loan default i.e. the variables which are strong indicators of default.

Outliers and Missing Values in Application Data

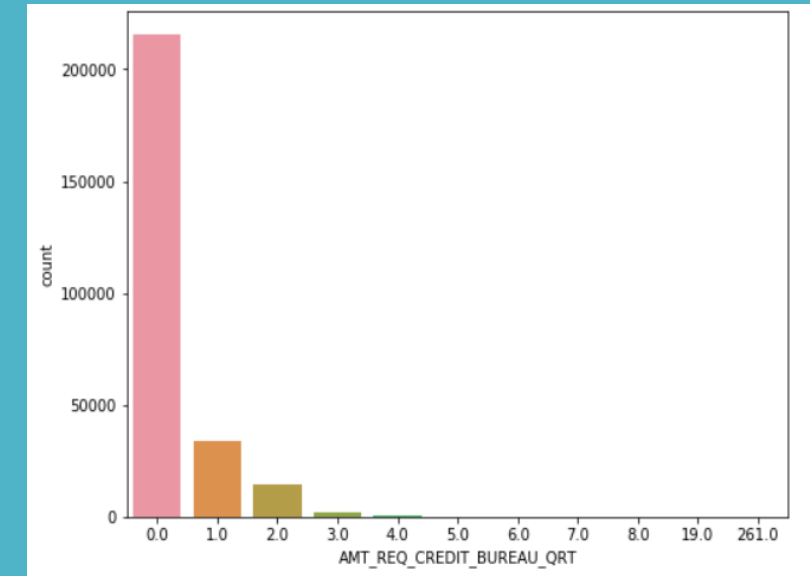
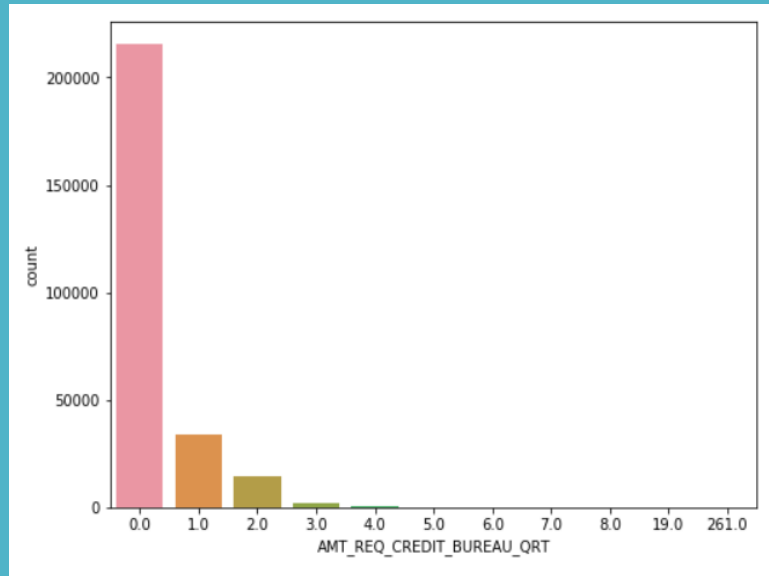
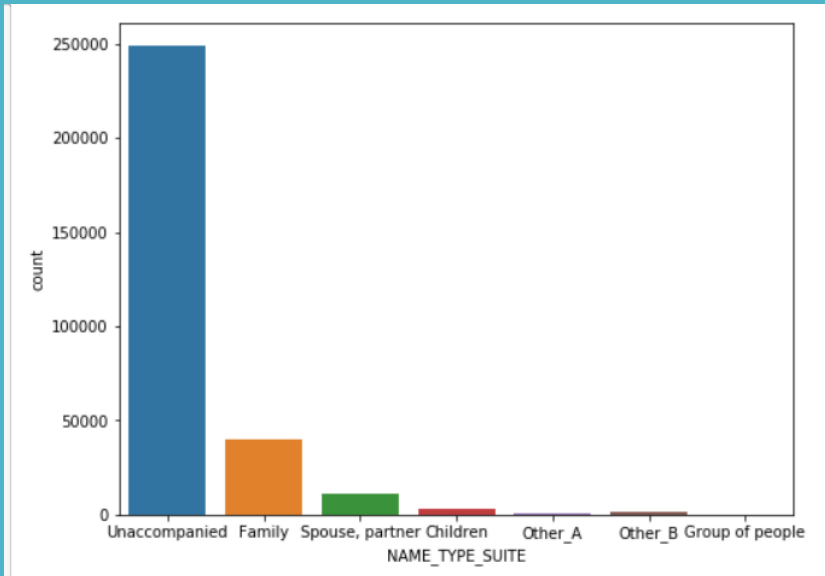
Outliers in Continuous Variables

- *Continuous Variable are* - AMT_GOODS_PRICE, EXT_SOURCE_2, EXT_SOURCE_3.
- Outliers are present in only AMT_GOODS_PRICE column and we can impute it using Median.



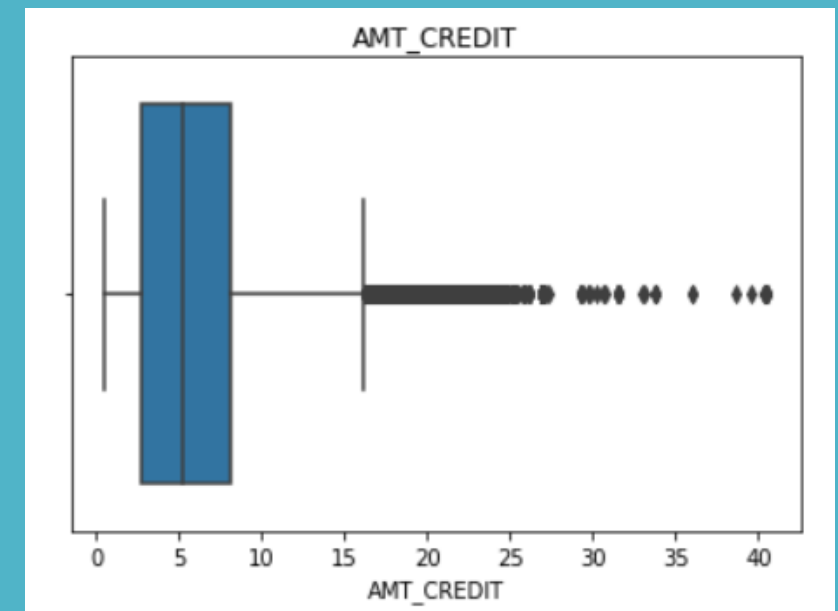
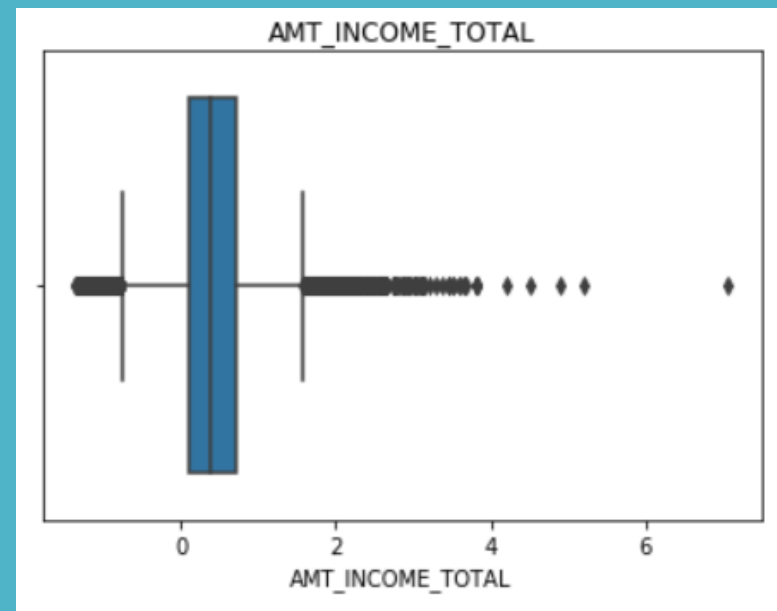
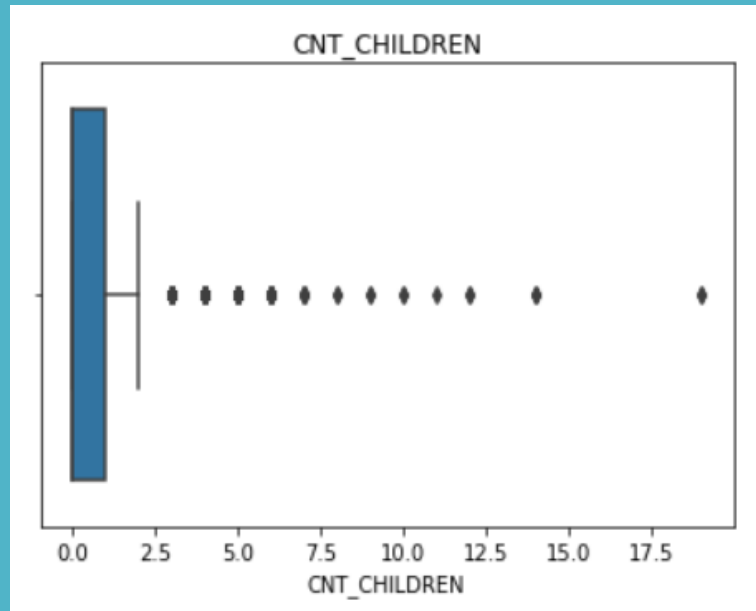
Missing Values in Categorical Variables

- **Categorical Variable are** - AMT_REQ_CREDIT_BUREAU_QRT , NAME_TYPE_SUITE, DEF_30_CNT_SOCIAL_CIRCLE.
- Impute missing values in categorical variables based on more no of occurrences.



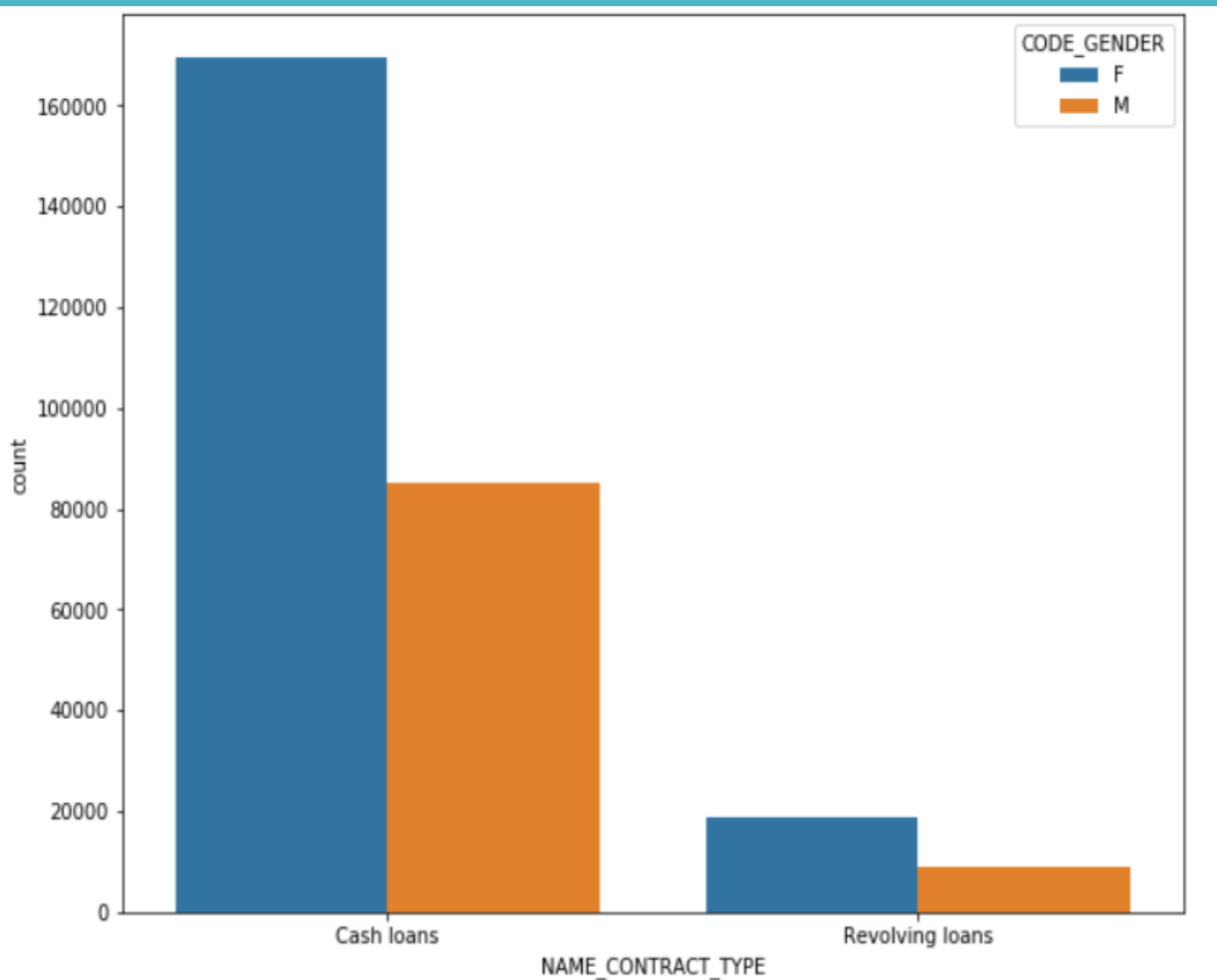
Outliers in Numerical Columns

- Outliers in Numerical Columns - CNT_CHILDREN , AMT_INCOME_TOTAL , AMT_CREDIT.
- Outliers are present in above numerical columns due to Max percentage value is greater than Mean , Mode and Median values.



Univariate Analysis in Categorical Variables for Target 0

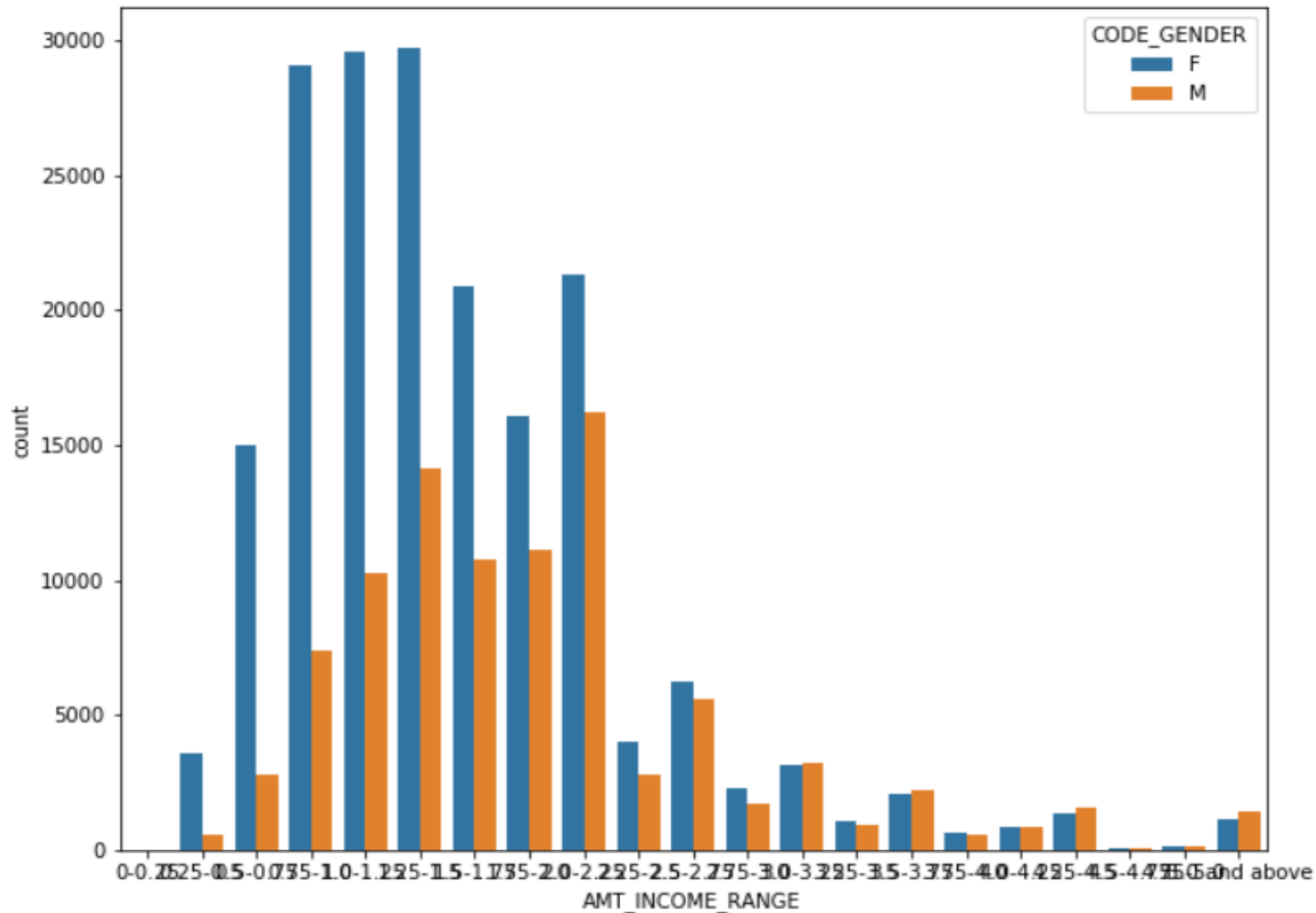
Observation Across Categorical Variables For TARGET0



Based on CONTRACT TYPE

We have concluded that

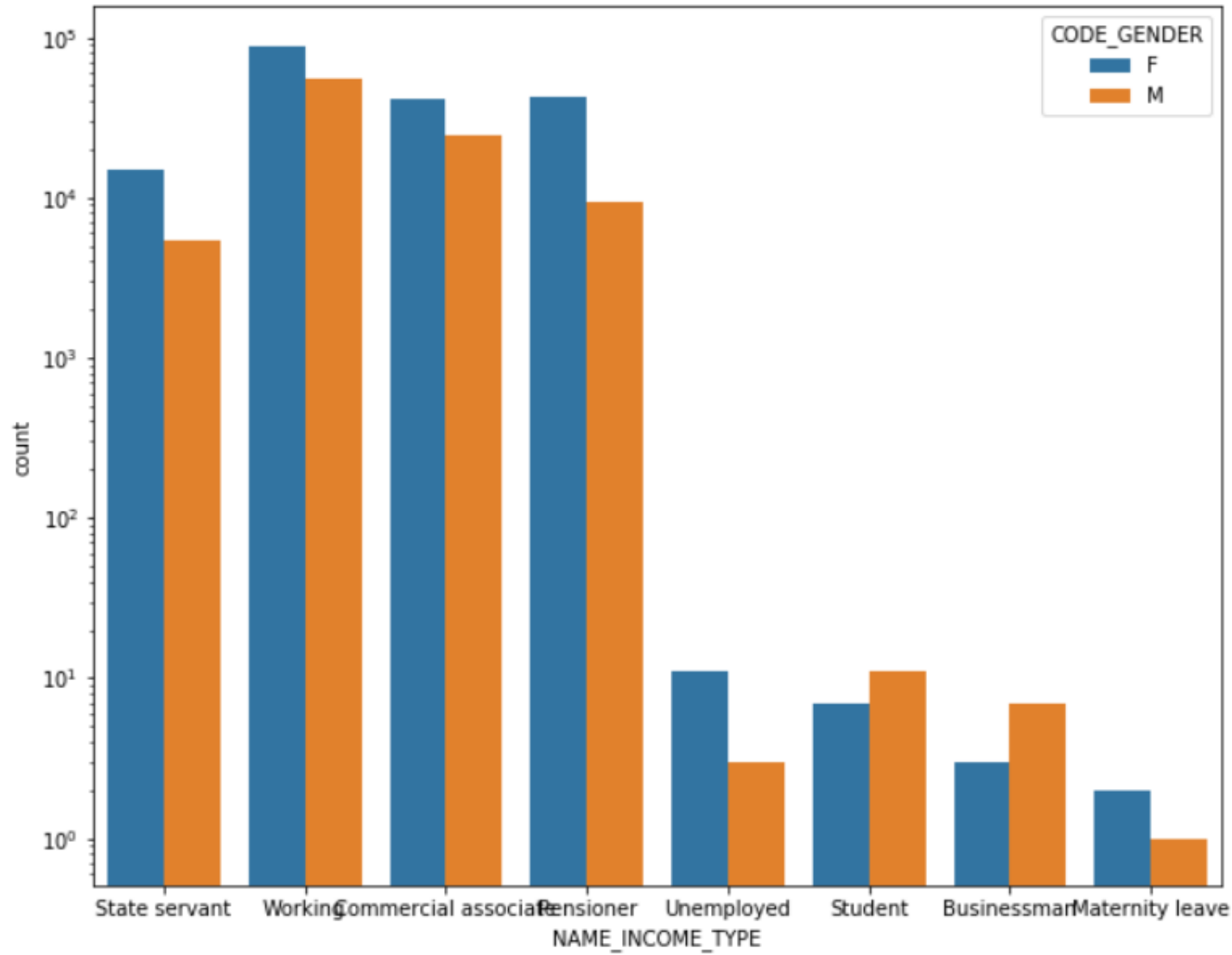
- Contract Type Cash Loans is having higher number of credits than Revolving Loans contract type.
- Credits are more in Female than Male.



Based on INCOME RANGE

We concluded that

- Females are leading credits range than Male.
- Income range is very high between 0.5 - 0.75 , 0.75 -1.0 , 1.0 - 1.25 , 1.25 - 1.5 and low in 4.5-4.75 range.



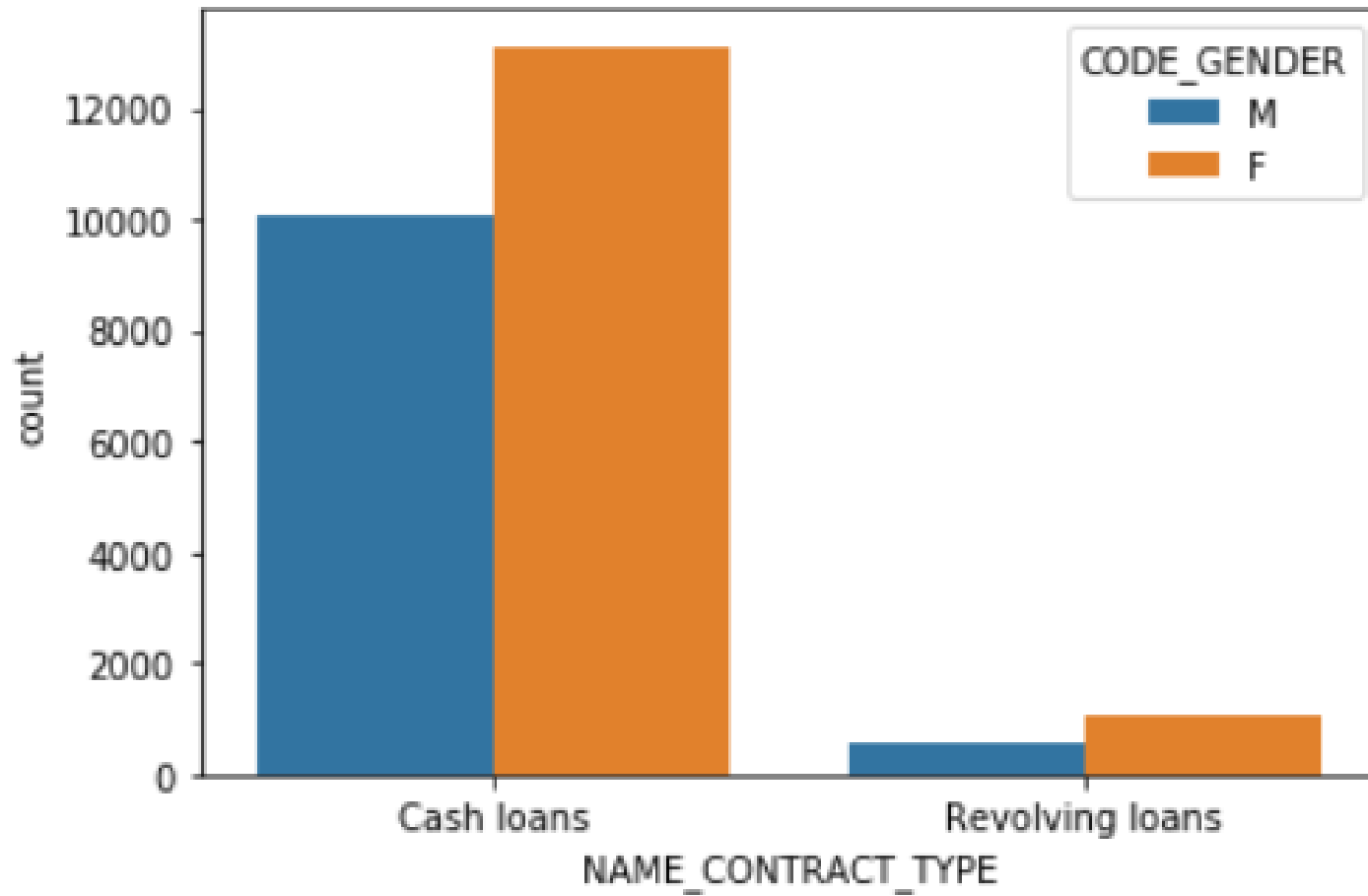
Based on INCOME TYPE

We concluded that

- For income type Working , Commercial associate , Pensioner are having more credits than other income type.
- For income type Student , Businessman and Maternity Leave income type has very no. of credits.

Univariate Analysis in Categorical Variables for Target1

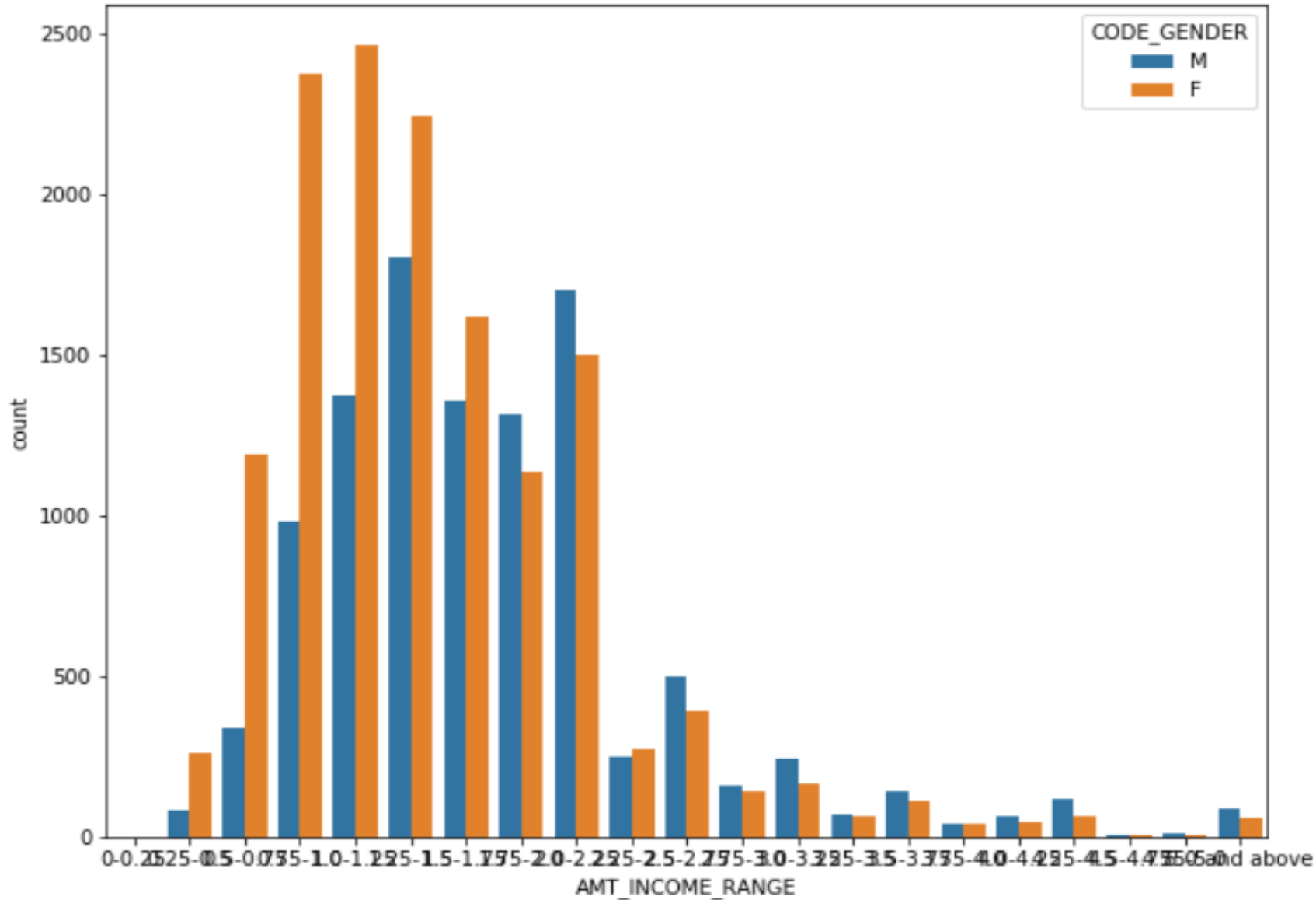
Observation Across Categorical Variables For TARGET1



Based on CONTRACT TYPE

We concluded that

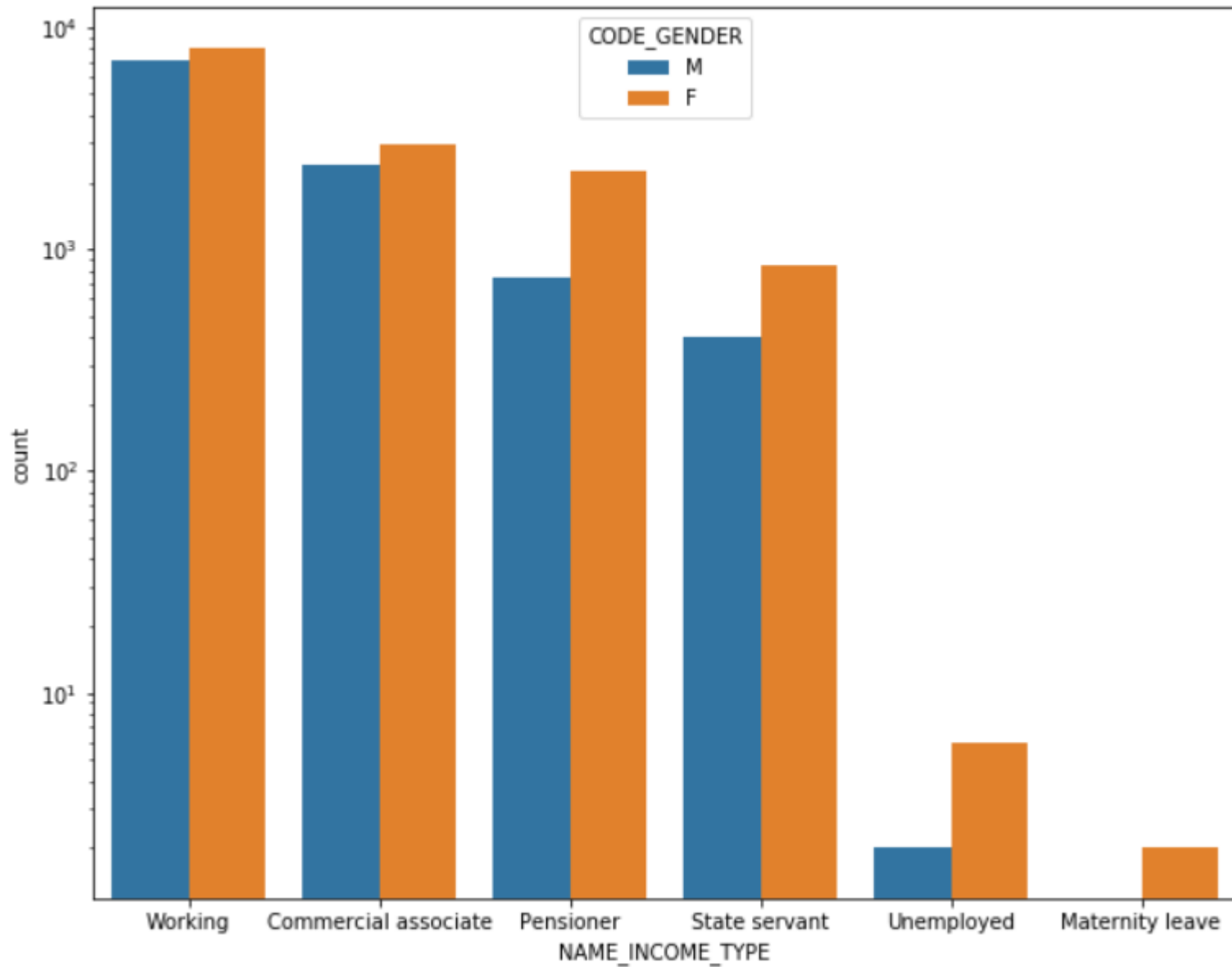
- Contract Type Cash Loans is having higher number of credits than Revolving loans contract type .
- Females are having more counts than Male in both Contract type.



Based on INCOME RANGE

We concluded that

- Income Range 0.75 - 1.0 , 1.0 - 1.25 , 1.25 - 1.5 are having high income range in Female as compared to Male.
- Very low income range is in 8.0 - 8.5 , 8.5 - 9.0.

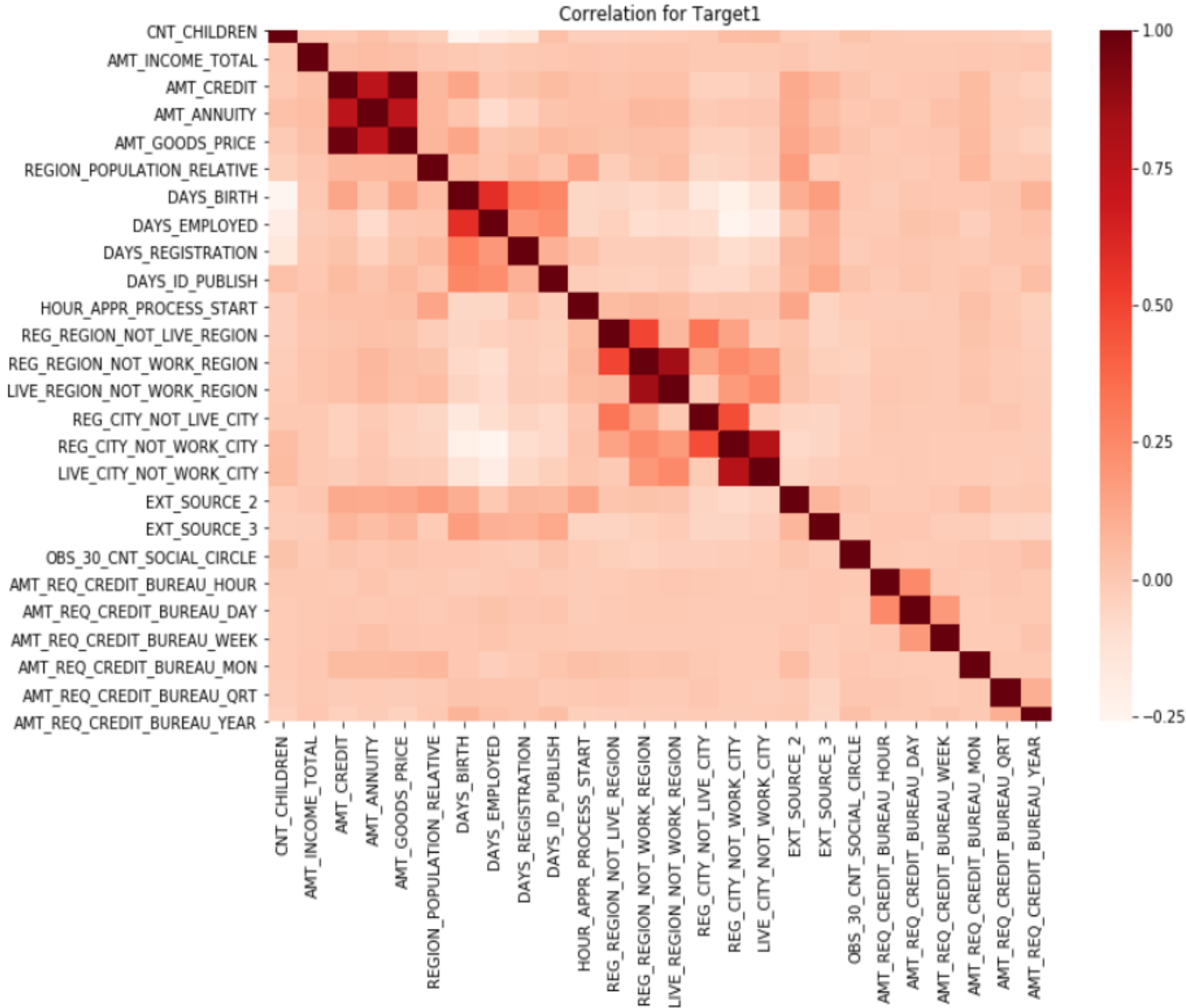


Based on INCOME TYPE

We concluded that

- For the income type Working , Commercial Associate , Pensioner and State Servant are in descending order of Income Type.
- Less income type is Maternity Leave.
- Females are leading in all income type than Male.

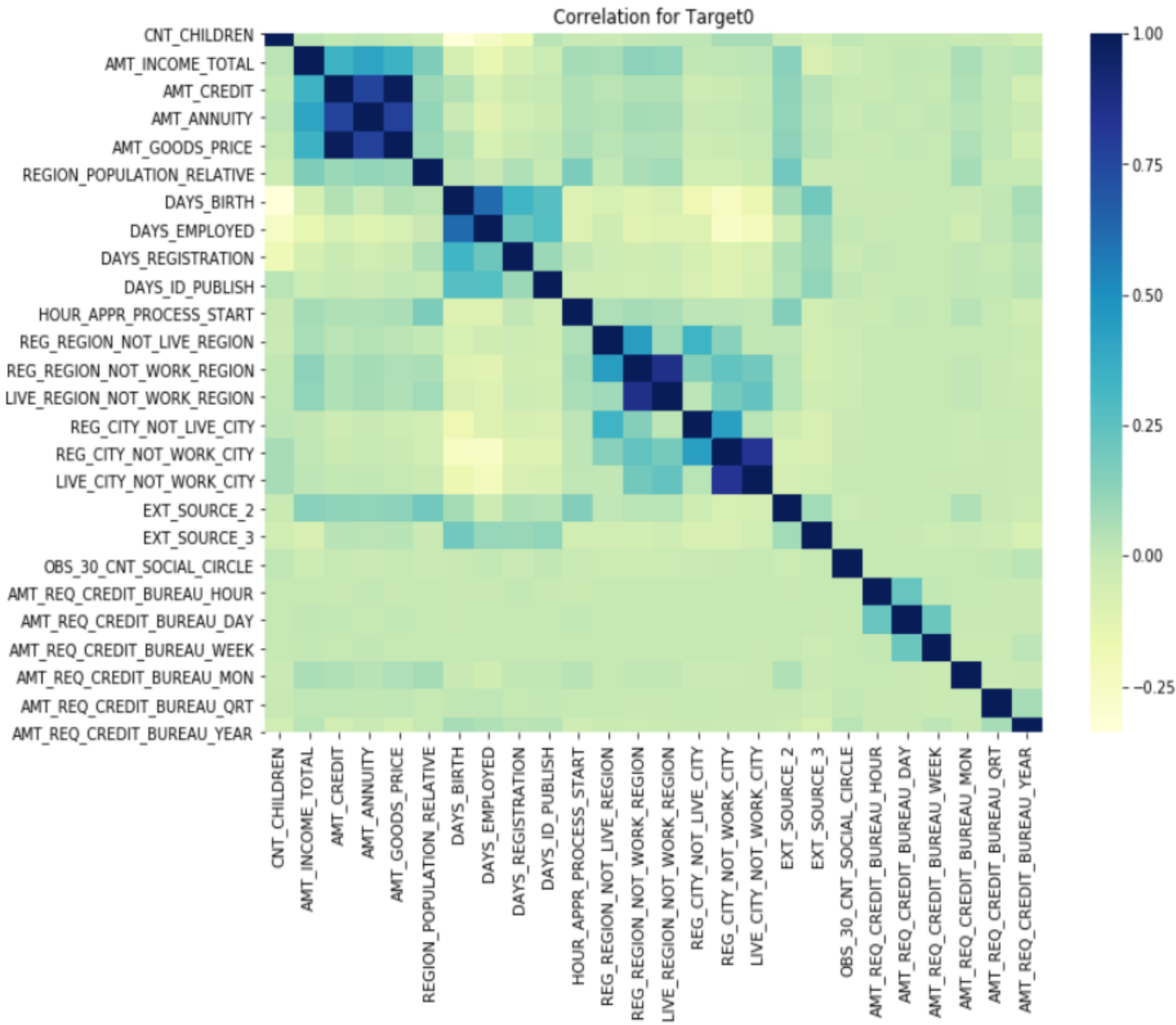
Correlation on Numerical Column



Correlation for Target 1

We conclude that

- AMT_GOODS_PRICE and AMT_CREDIT are strongly positively correlated.
- AMT_GOODS_PRICE and AMT_ANNUITY are strongly correlated.
- AMT_ANNUITY and AMT_CREDIT are strongly correlated.
- LIVE_REGION_NOT_WORK_REGION and REG_REGION_NOT_WORK_REGION are strongly correlated.
- LIVE_CITY_NOT_WORK_CITY and REG_CITY_NOT_WORK_CITY are strongly correlated.



Correlation for Target0

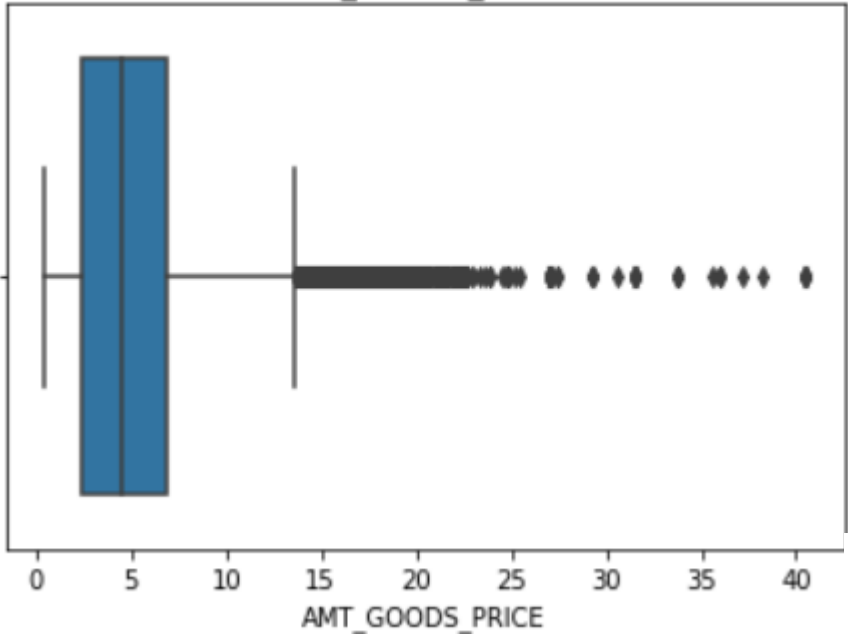
We concluded that

- AMT_GOODS_PRICE and AMT_CREDIT are strongly positively correlated.
- AMT_GOODS_PRICE and AMT_ANNUITY are strongly correlated.
- AMT_ANNUITY and AMT_CREDIT are strongly correlated.
- LIVE_REGION_NOT_WORK_REGION and REG_REGION_NOT_WORK_REGION are strongly correlated.
- LIVE_CITY_NOT_WORK_CITY and REG_CITY_NOT_WORK_CITY are strongly correlated.

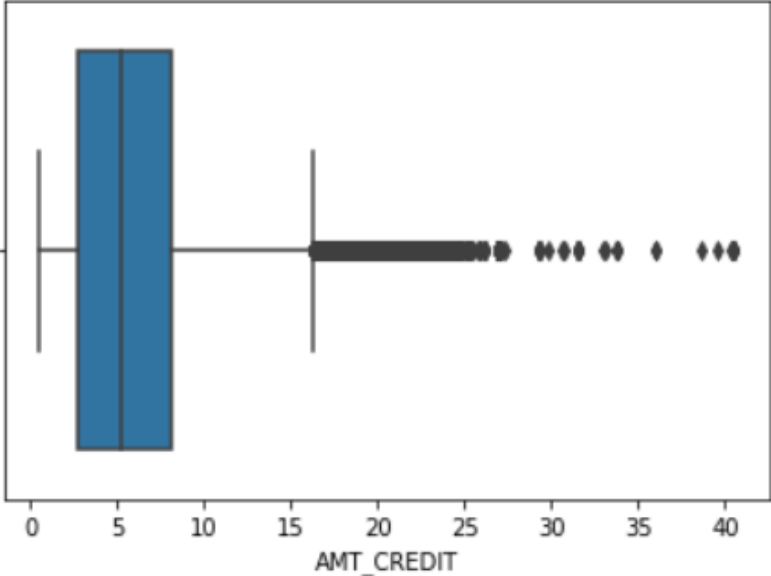
Univariate in Numerical Variables for Target0

Outliers present in AMT_GOODS_PRICE and AMT_CREDIT and no outliers in DAYS_BIRTH for TARGET0

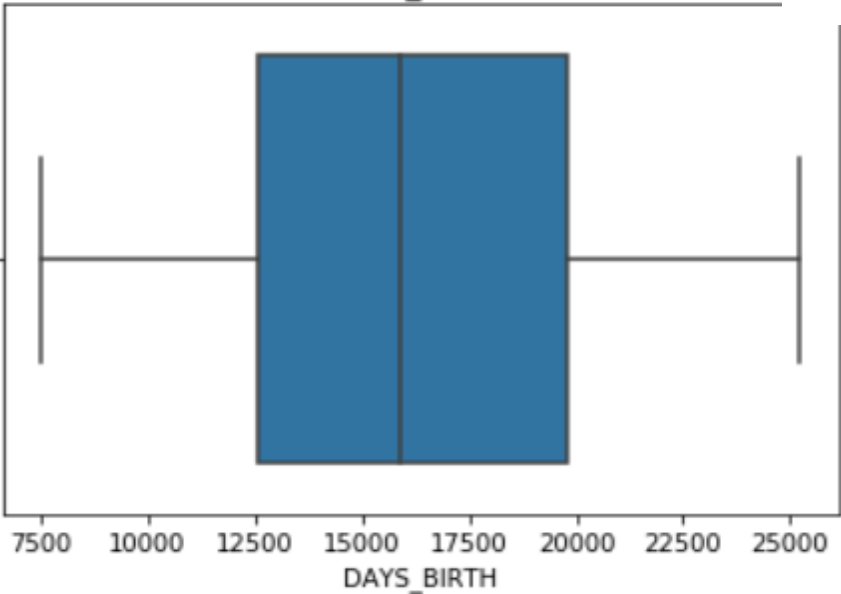
AMT_GOODS_PRICE



AMT_CREDIT

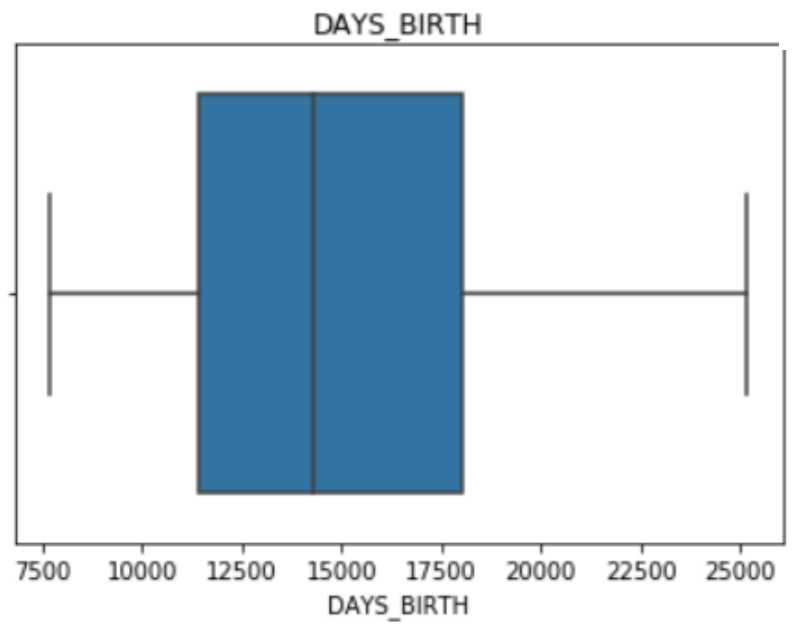
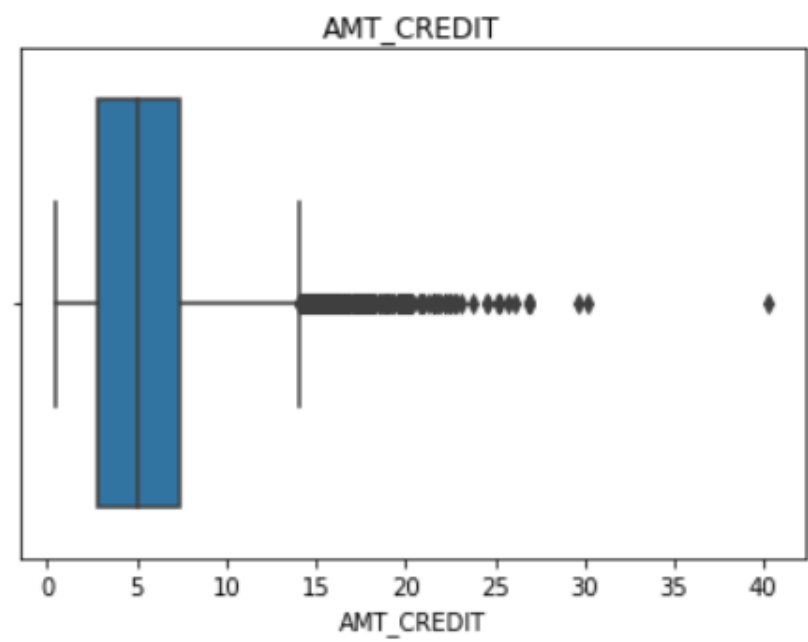
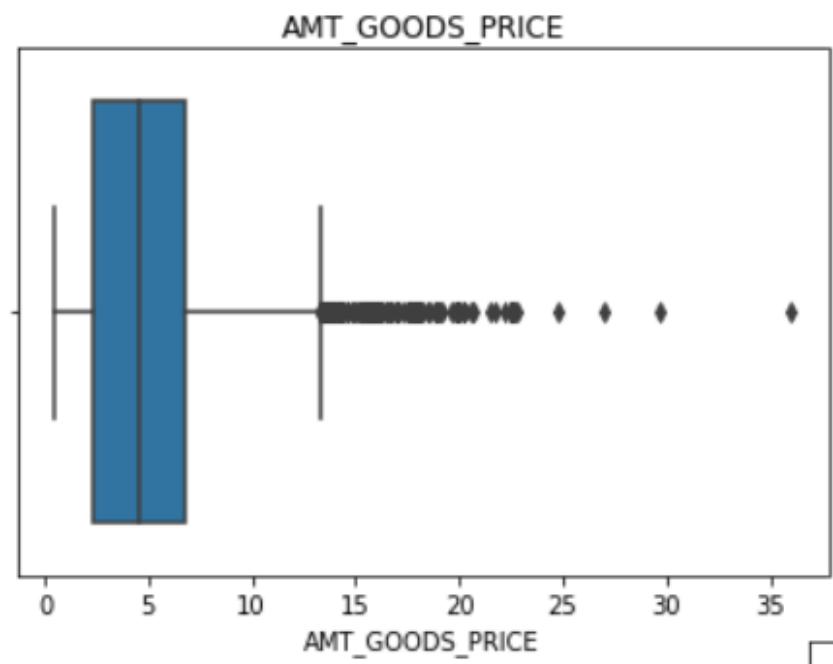


DAYS_BIRTH

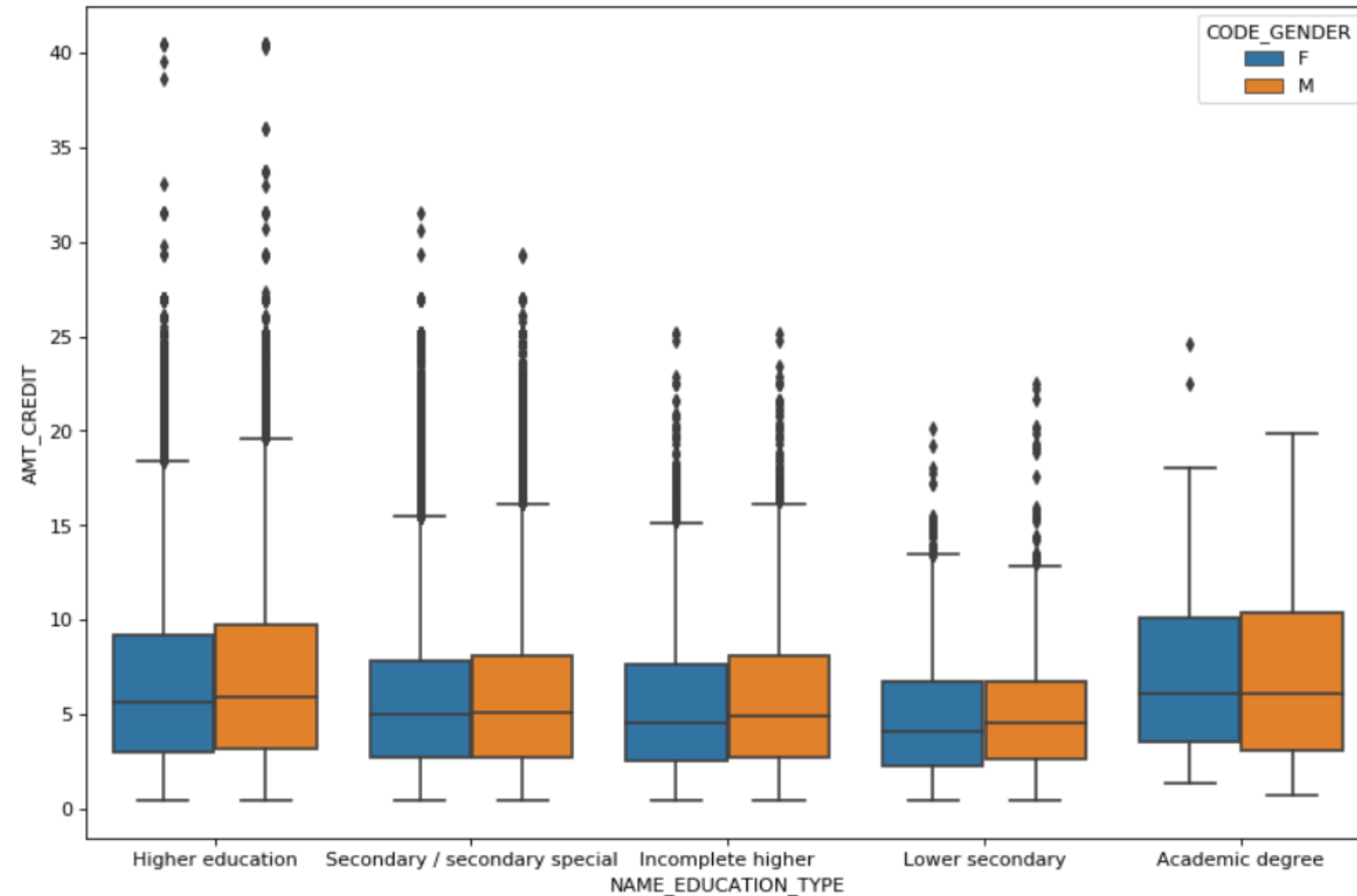


Univariate in Numerical Variables for Target1

Outliers present only in AMT_GOODS_PRICE and AMT_CREDIT and no outliers in DAYS_BIRTH for TARGET1

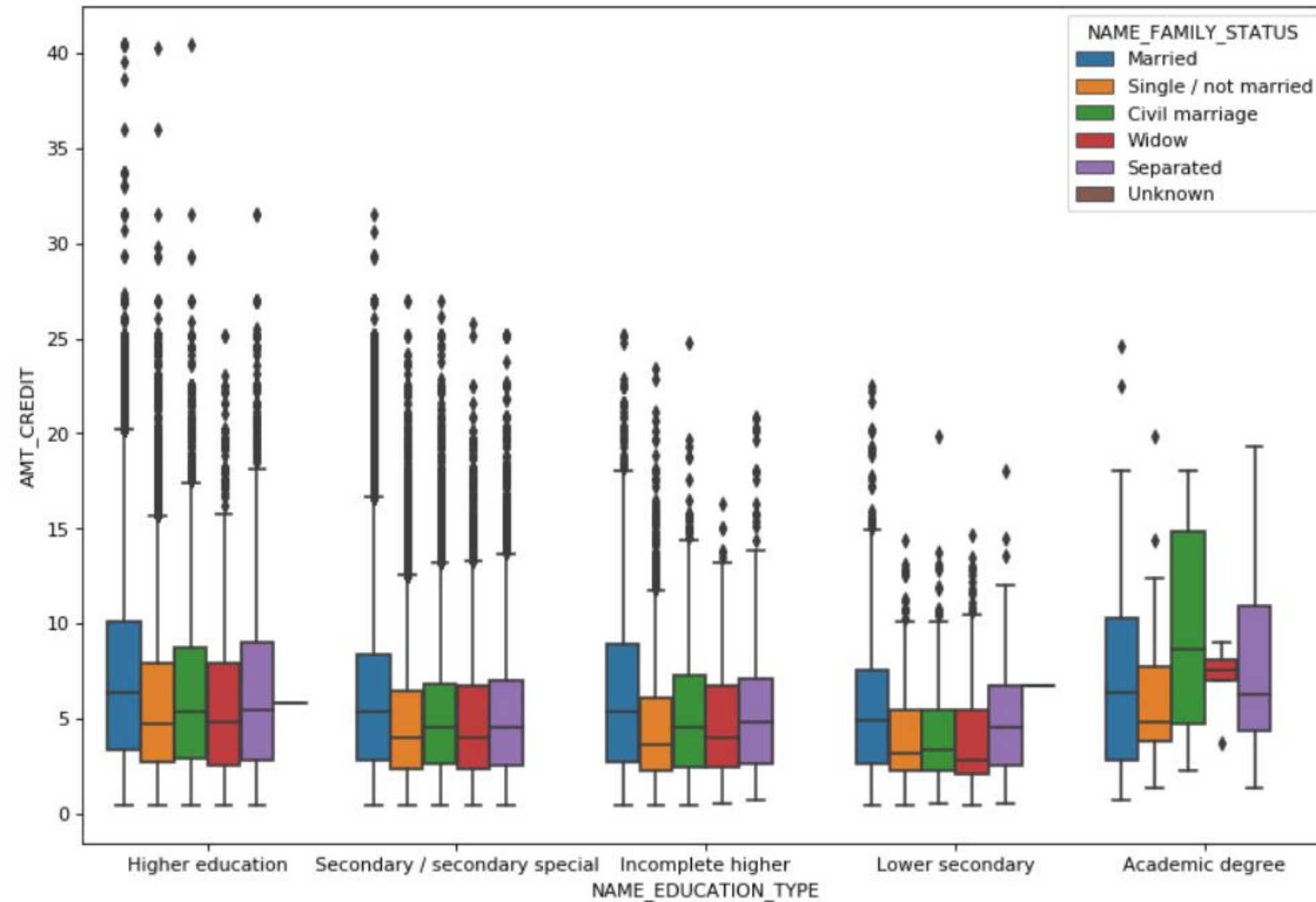


Bivariate Analysis for Numerical Variable in Target0



We concluded in Bivariate Analysis for Numerical variable in Target0 that

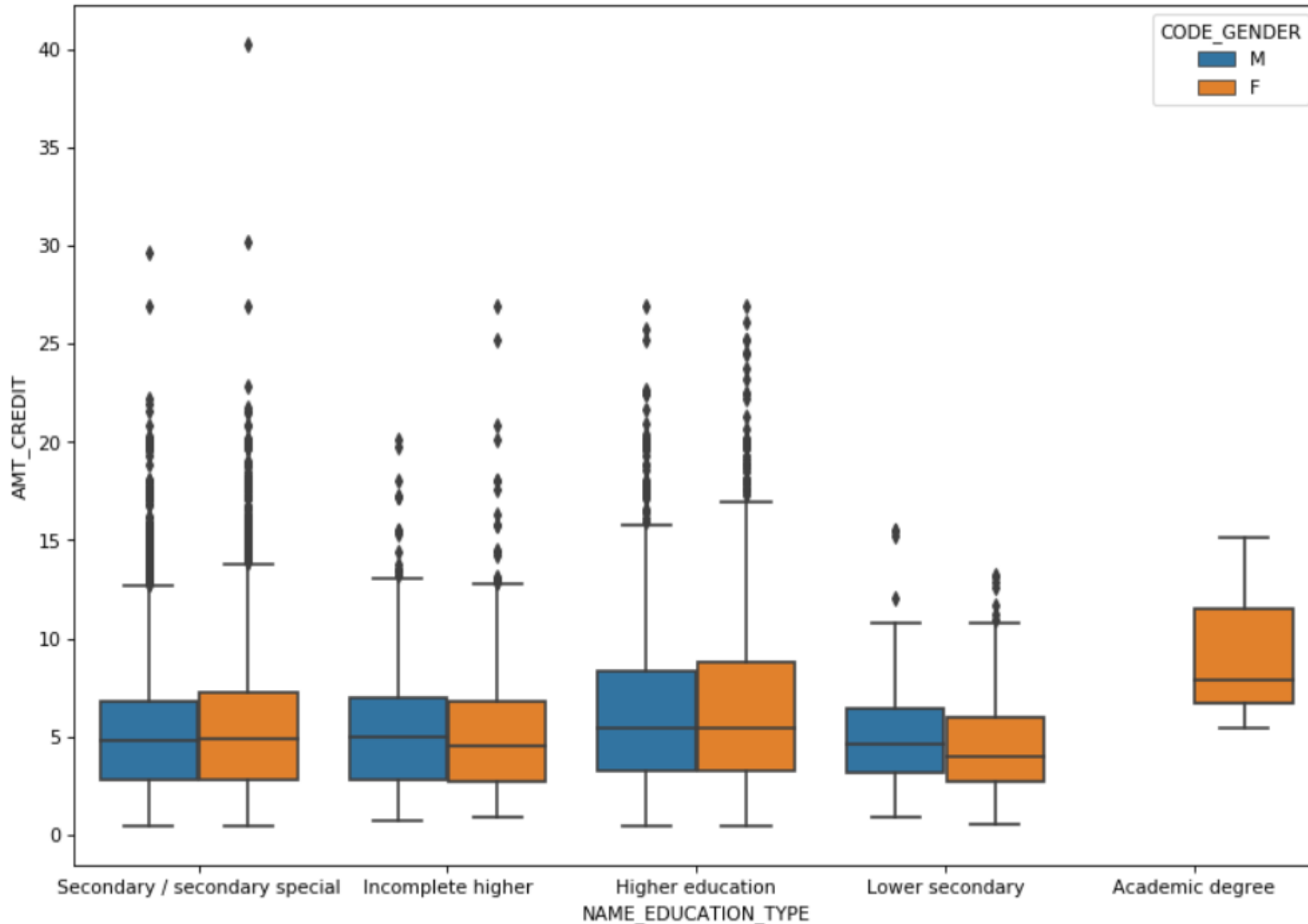
- Male and Female with Academic degree and Higher education are having more credits in Education type than others.
- Very less credits in Lower secondary education type.



We concluded in Bivariate Analysis for Numerical variable in Target0 that

- Married and Civil Marriage with academic degree are having more credits than others.
- Lower secondary education type has less credits in all family status.

Bivariate Analysis for Numerical Variable in Target1

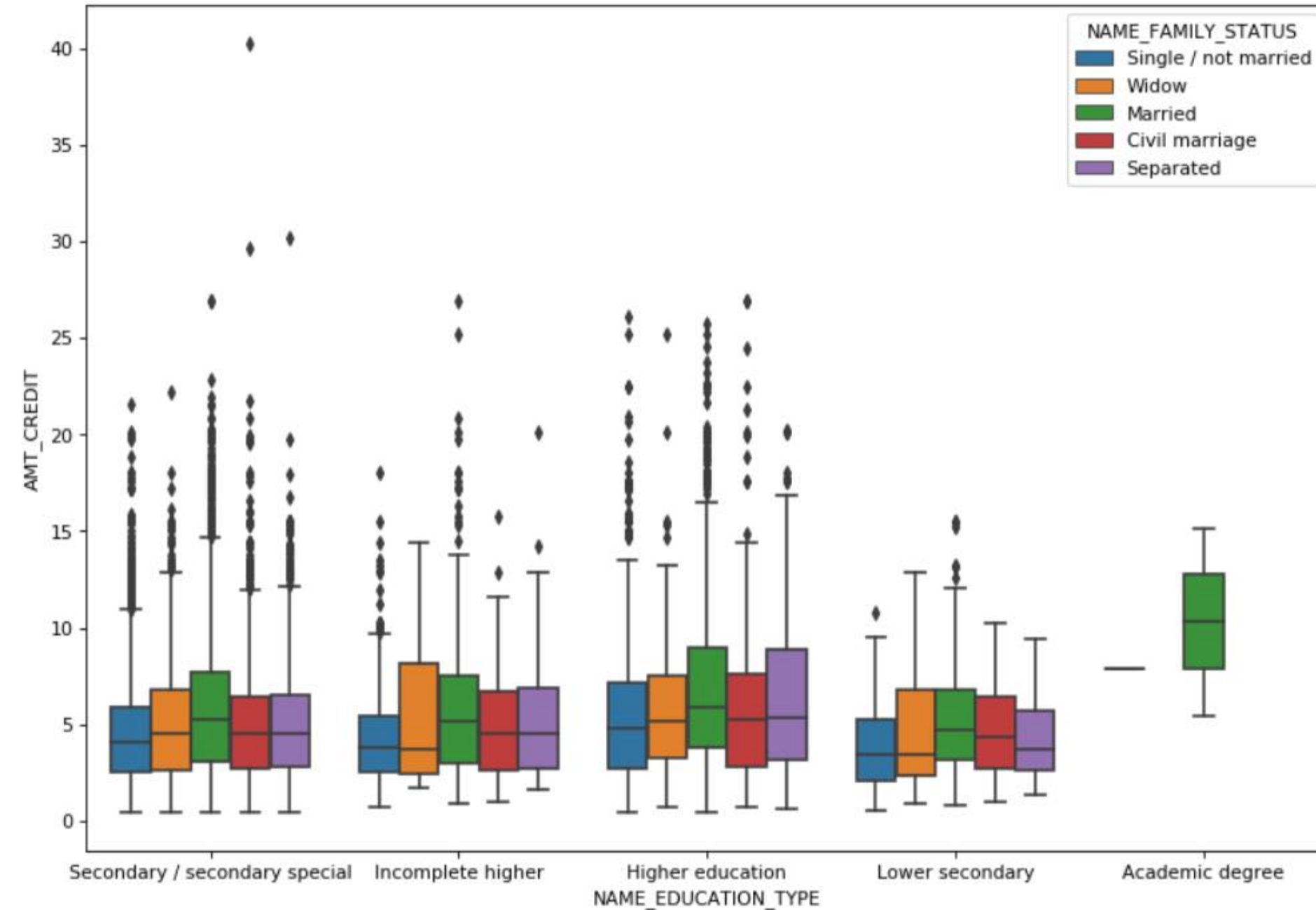


We concluded in Bivariate Analysis for Numerical variable in Target1 that

- Female with Academic degree are having more credits than Male/Female with other education type.
- Lower secondary education type has less credits for Male and Female gender.

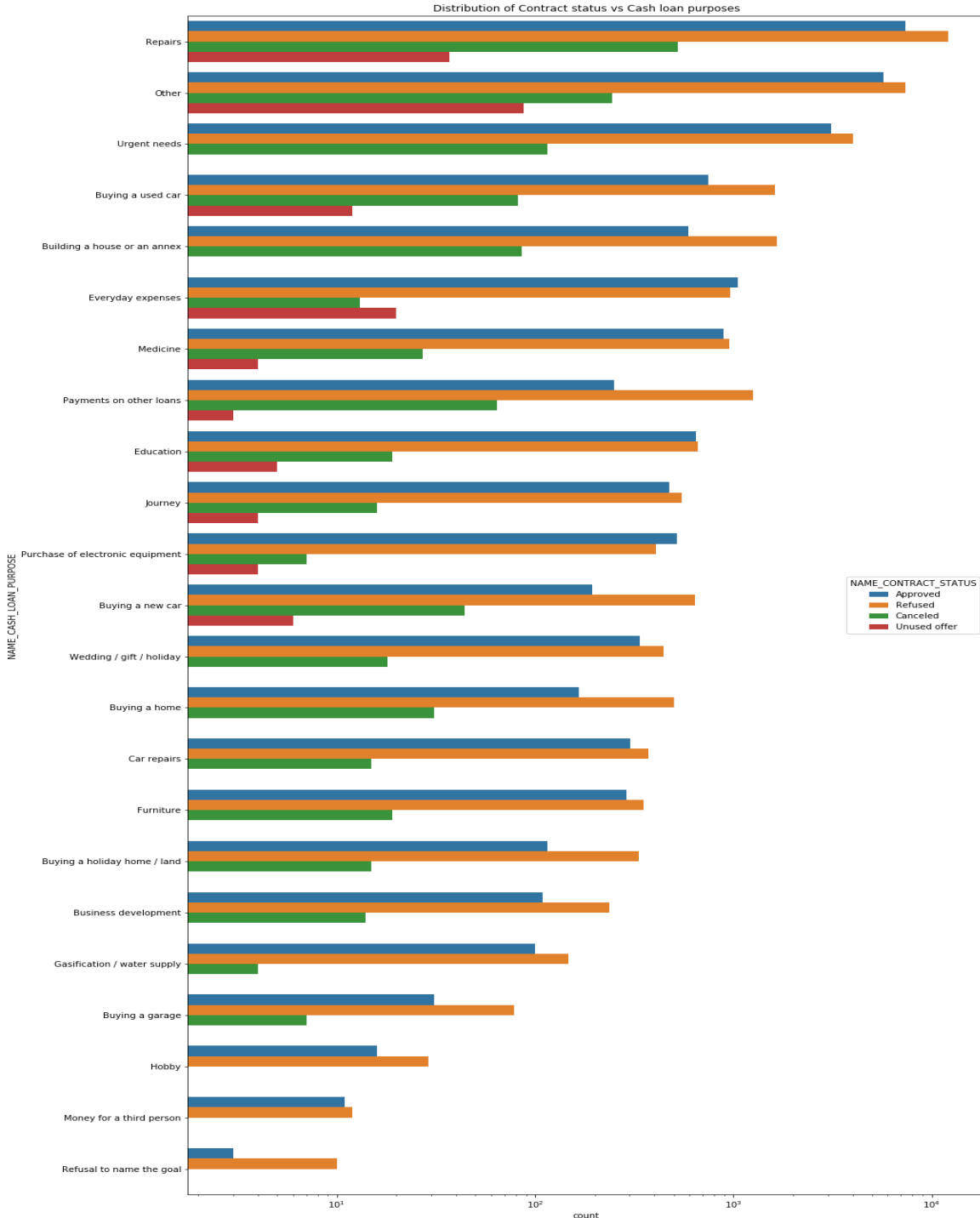
We concluded in Bivariate Analysis for Numerical variable in Target1 that

- Married family status with Academic education type are having higher credits than others.
- And After that married, separated are having more credits in Higher education and widow is having more credits than others in Incomplete higher.



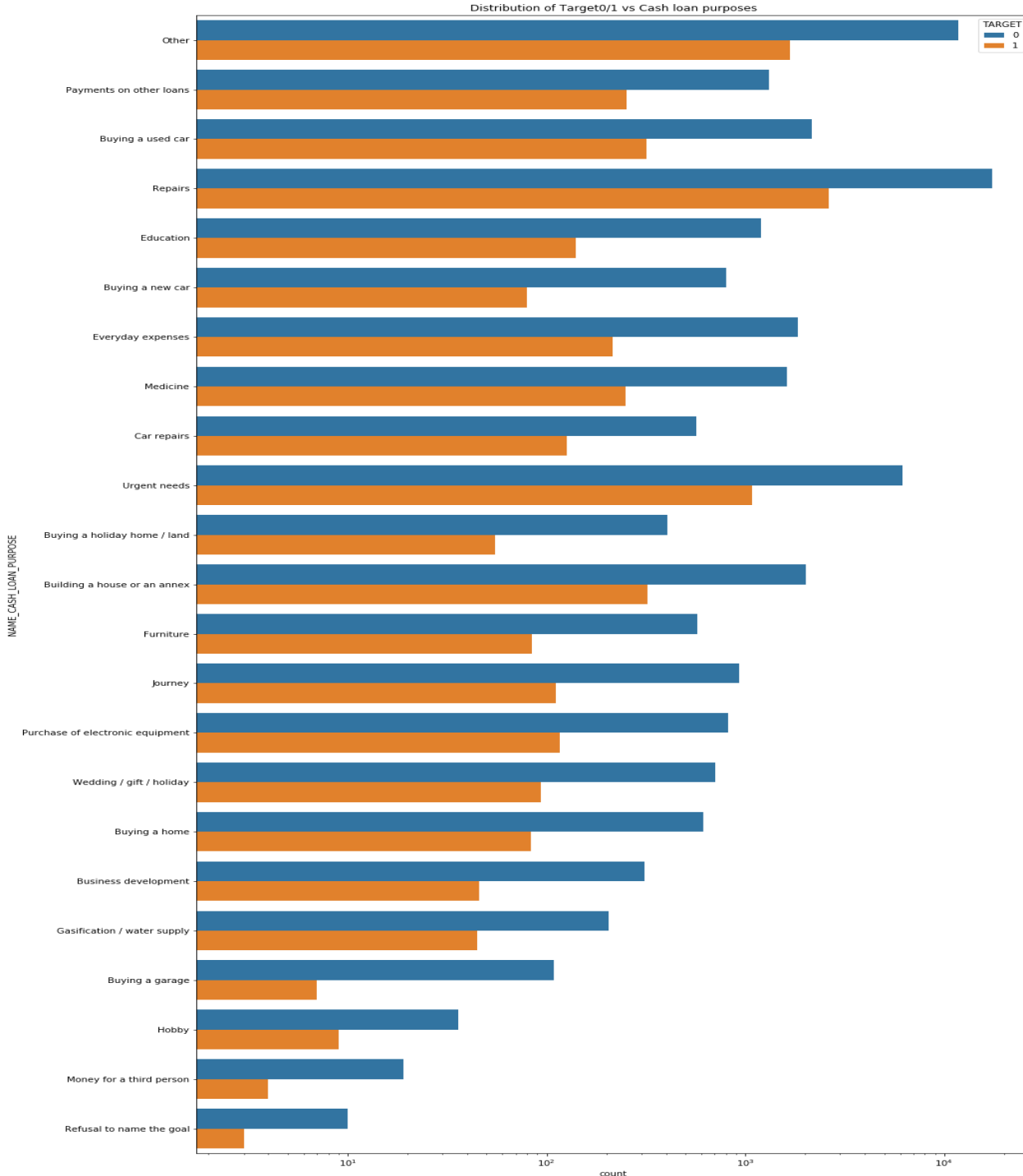
Univariate Analysis on Merge Dataset

- We merged Application data and Previous application dataset in new data frame called Merge_DF.



From this Univariate analysis on Merge Dataset –

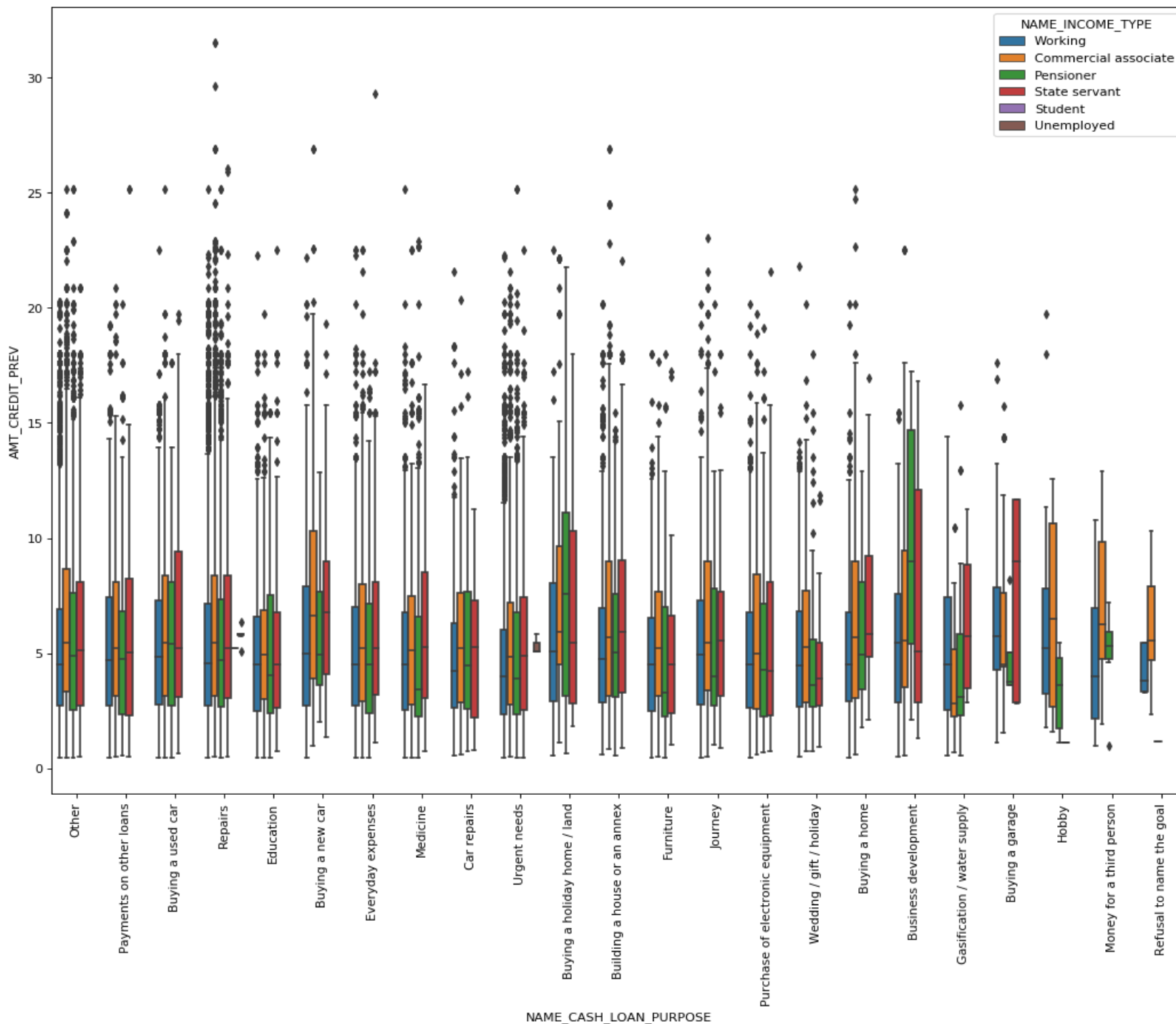
- Purpose 'Repairs' has more no. of loan rejection.
- And for Education purpose we have equal no of Approved and Refused loans.
- Everyday expenses are more approved than refused contract status.
- Customers are spending very less on cash loans type of Refusal to name the goal.



From the above Univariate analysis on Merge Dataset-

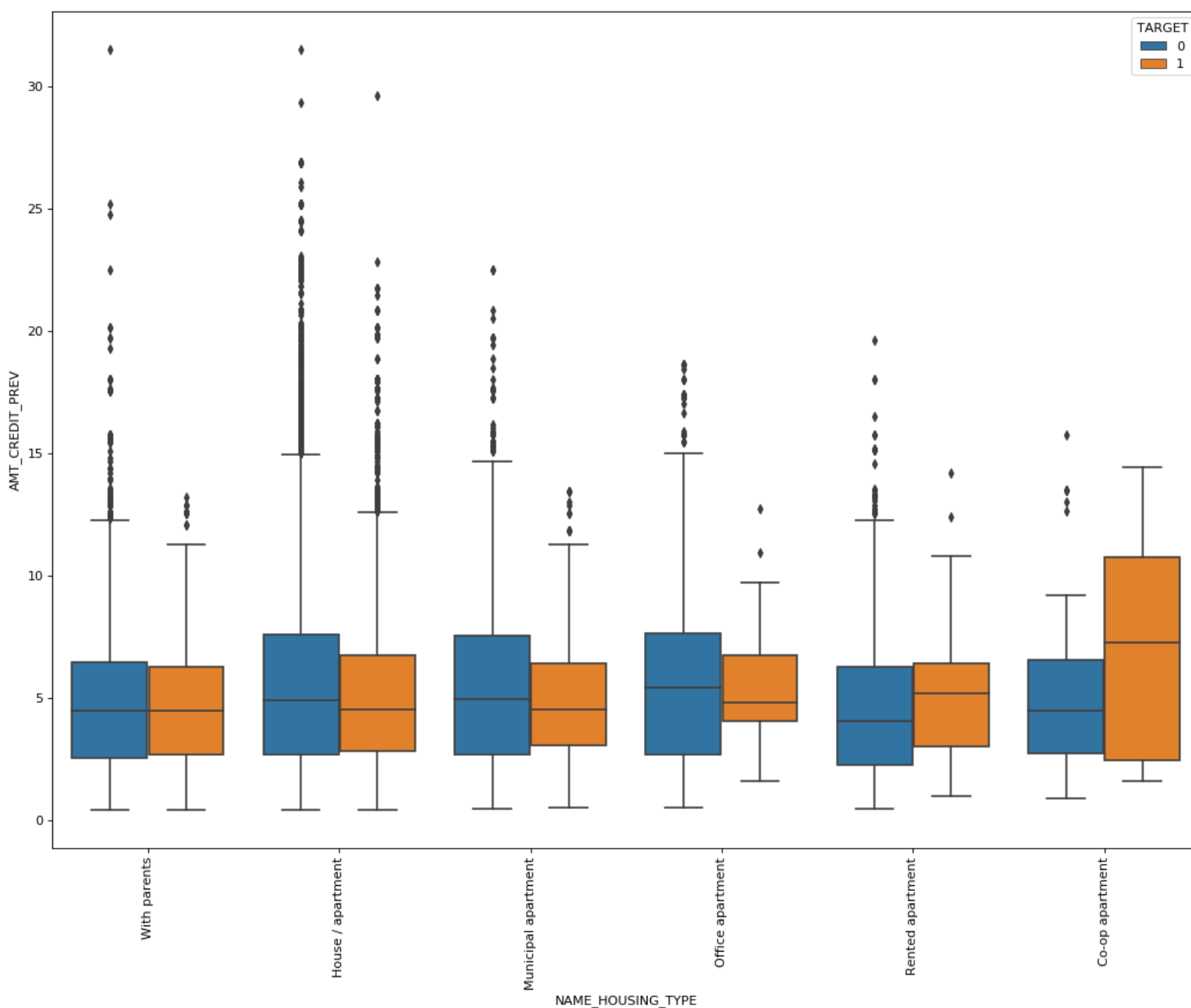
- Loan Purpose 'Repairs' are having more difficulties at the time of payment.
- And for education purpose we have equal no of approved and Refused loans.
- Purchase of electronic equipment's and everyday expenses has more approved rate than refused, cancelled or unused offers.
- Buying a new car and Refusal to name the goal are more likely to be refused more than approved.

Bivariate Analysis using Merge dataset



Bivariate analysis using Merge dataset , we concluded that

- Business development and Buying a holiday home or land are having more credits than other cash loan purpose.
- Very less credits on Refusal to name the goal cash loan purpose.
- Buying a new car , buying a used car , everyday expenses , urgent needs , Building a house or an annex and journey all these columns are having significant amount of credits.



Bivariate analysis using Merge dataset , we concluded that

- From the above bivariate analysis Co-op apartment are having more credit for Target1.
- House/apartment , Municipal apartment and office apartment are having more credits for Target0.
- Very less credits present in office apartment for Target1.
- And Co-op apartment has very less credits for Target0.

CONCLUSION

1. Bank should provide priority to Housing type With Parents. House/apartment , Municipal and Office apartment customers for loans as they are having more no. of successful transactions as compared to Co-op apartment.
2. Bank Should focus on cash loan purpose for Buying a Business development , holiday home or land and buying a new car customers as they have less no of unsuccessful transactions.
3. Banks should consider contract type Student , Pensioner and Commercial Associate with housing type other than Co-op apartment for more successful transactions.
4. Bank should provide less priority to Repairs and Urgent need in Cash loan purposes and Income Type - Working are having more no. of unsuccessful transactions .
5. Bank should consider giving loans to customers who are having academic degree or higher education.

Thank You...!