Credit EDA Assignment

Lasya Nallamalli
Batch ID: 5706

Data Cleaning

For both application_data.csv and previous_application.csv

- Dropped columns if the percentage of null values is greater than 30%.
- Dropped unwanted columns for data analysis.
- Dropped null value rows if the number is insignificant.
- Dropped null value rows if they can't be added by any external source or by imputing some standard methods.
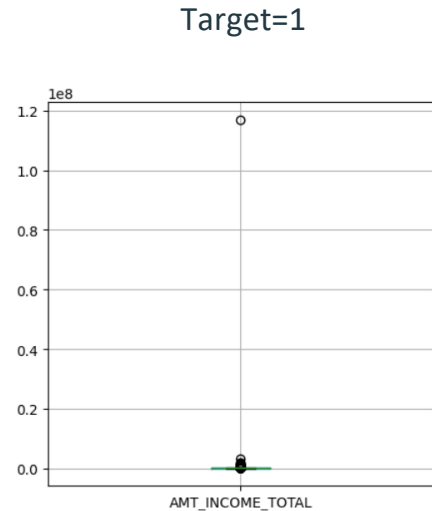
Imbalance ratio for target variable = 11.38:1

Analysing Columns

Univariate analysis on AMT_INCOME_TOTAL

- Few people have very high salary compared to others .
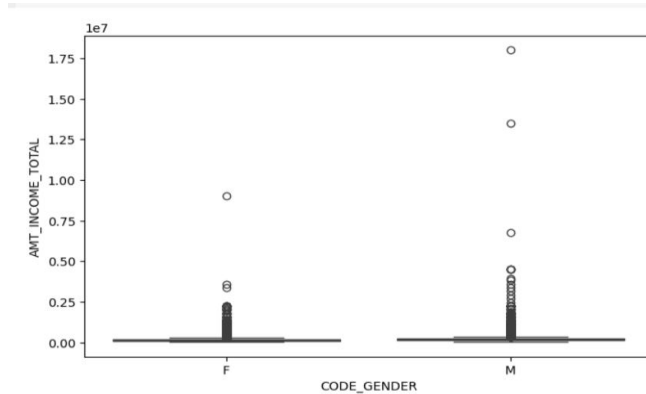- Considering these values in the analysis as they may be useful for getting insights
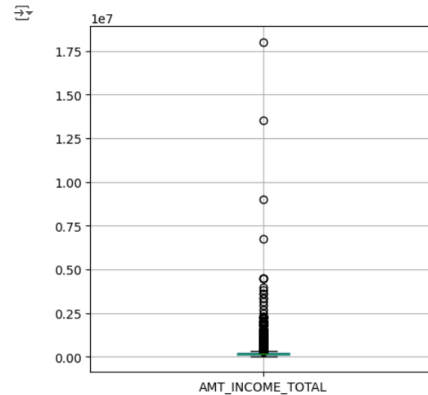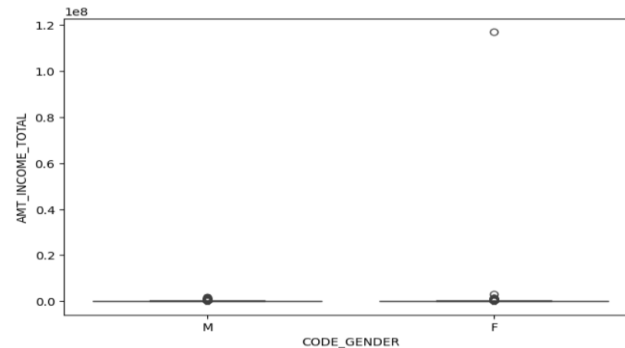

Target=0

Target=1

Bivariate analysis on AMT_INCOME_TOTAL x CODE_GENDER

- In target =1, almost everyone's income is nearby, (except for 1).
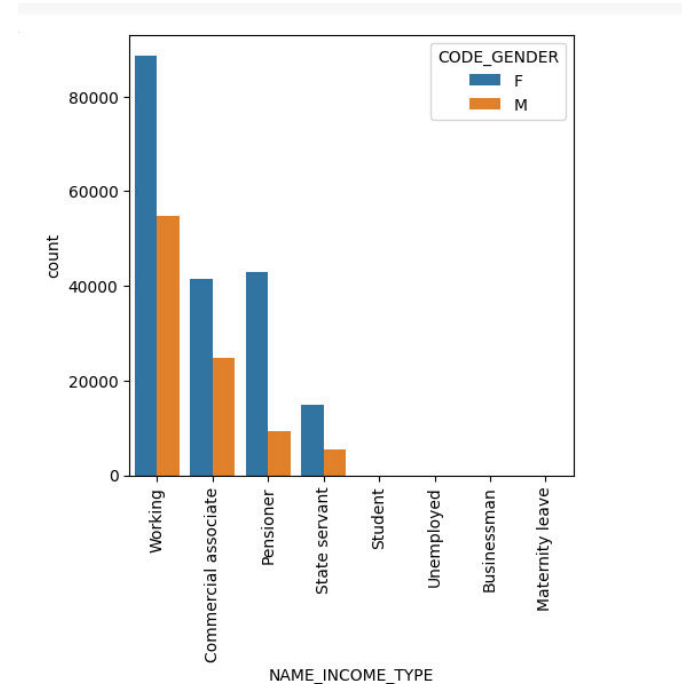- In target=0, there are few outliers in Income total for both male and female.

Target=0

Target=1

# Bivariate analysis on NAME_INCOME_TYPE x CODE_GENDER

Target= 0

- Income type Working, Commercial associate, Pensioner are more like to have no payment difficulties.
  - In these three income type, Female have no payment difficulties compared to men

# Bivariate analysis on NAME_CONTRACT_TYPE x CODE_GENDER

Target= 0

- Clients with contract type Cash loans are more likely to have no payment difficulties.
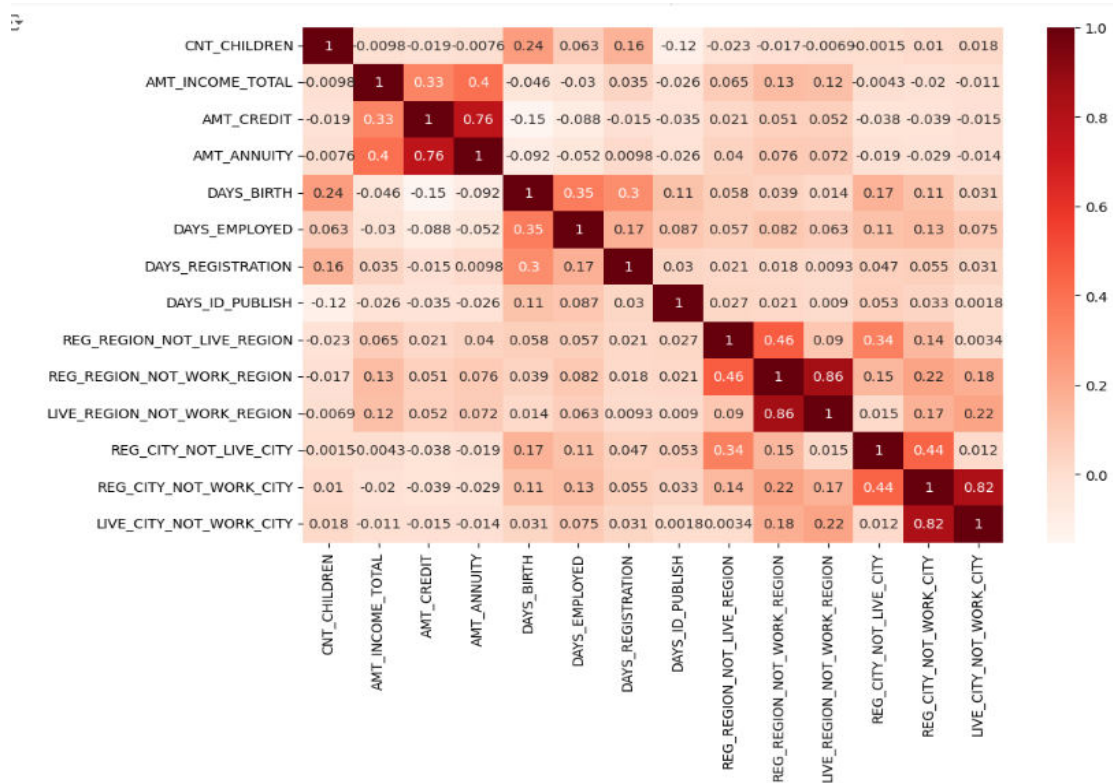 - In Cash loans, female are more likely to have no payment difficulties.
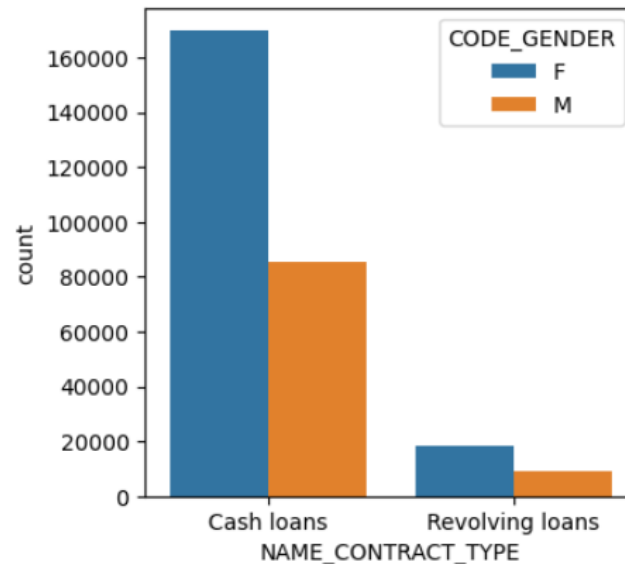
Univariate analysis on ORGANIZATION_TYPE

Target= 0

- ORGANIZATION_TYPE Business Entity Type 3 are more likely to apply for loan and also more likely to have no payment difficulties.
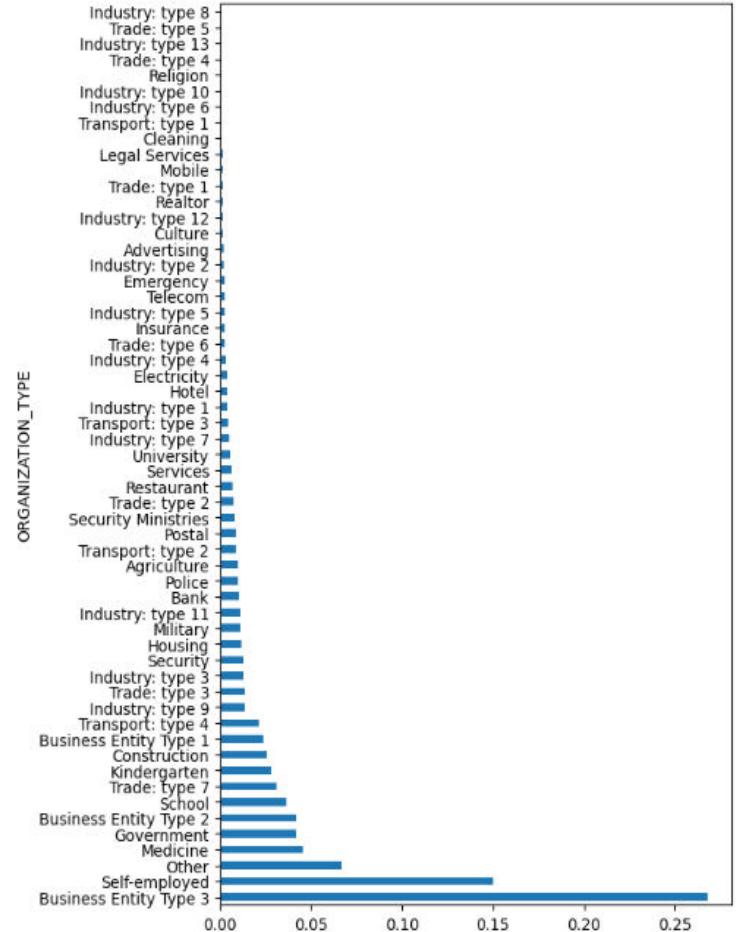
# Multivariate analysis on Target=0 records

# Multivariate analysis on Target=0 records Cont'd

- AMT_CREDIT is inversely proportional to DAYS_BIRTH, CNT_CHILDREN, DAYS_EMPLOYED, DAYS_REGISTRATION, DAYS_ID_PUBLISH
- AMT_CREDIT is highly correlated with AMT_ANNUIY.
- AMT_ANNUIY is directly proportional to AMT_INCOME_TOTAL.
- REG_REGION_NOT_LIVE_REGION is directly proportional to REG_REGION_NOT_WORK_REGION.
- LIVE_REGION_NOT_WORK_REGION is highly correlated to REG_REGION_NOT_WORK_REGION.
- LIVE_CITY_NOT_WORK_CITY is highly correlated to REG_CITY_NOT_WORK_CITY

Multivariate analysis on Target=1 records
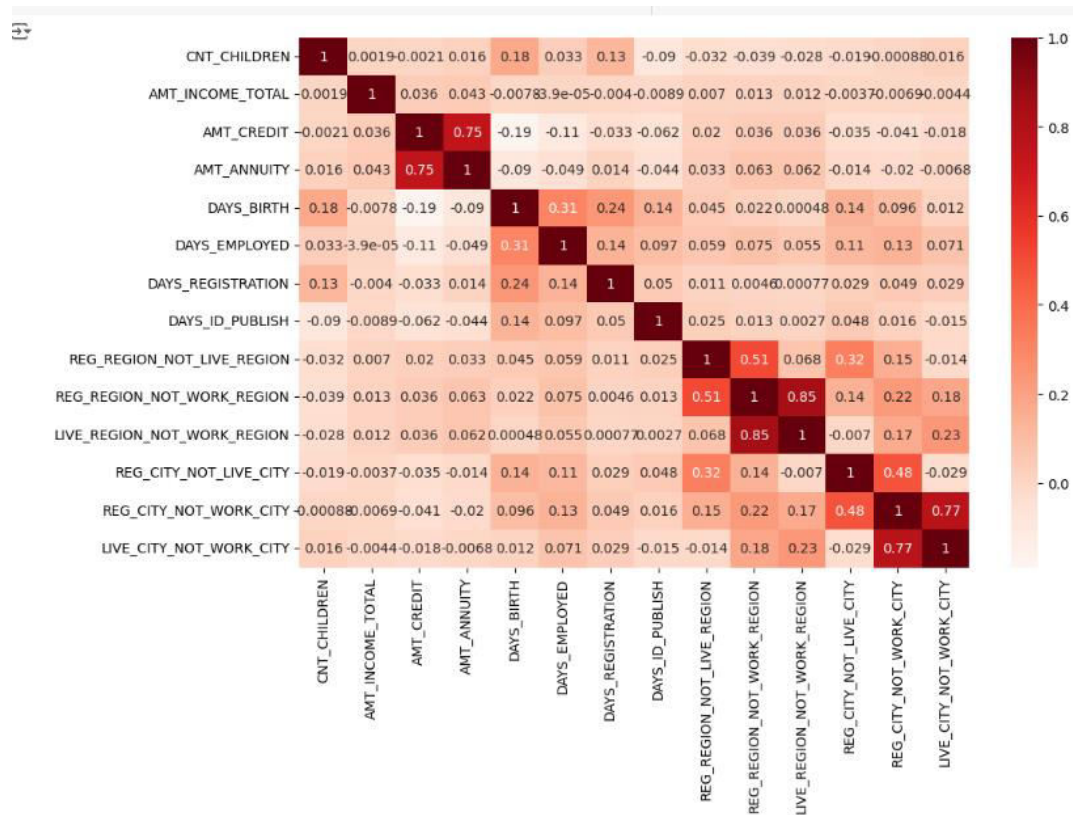
Multivariate analysis on AMT_CREDIT x NAME_EDUCATION_TYPE x NAME_FAMILY_STATUS

- Clients with Higher education and Married or Civil marriage or seperated has more AMT_CREDIT.
- Clients with Academic degree and Married or Civil marriage has more AMT_CREDIT.
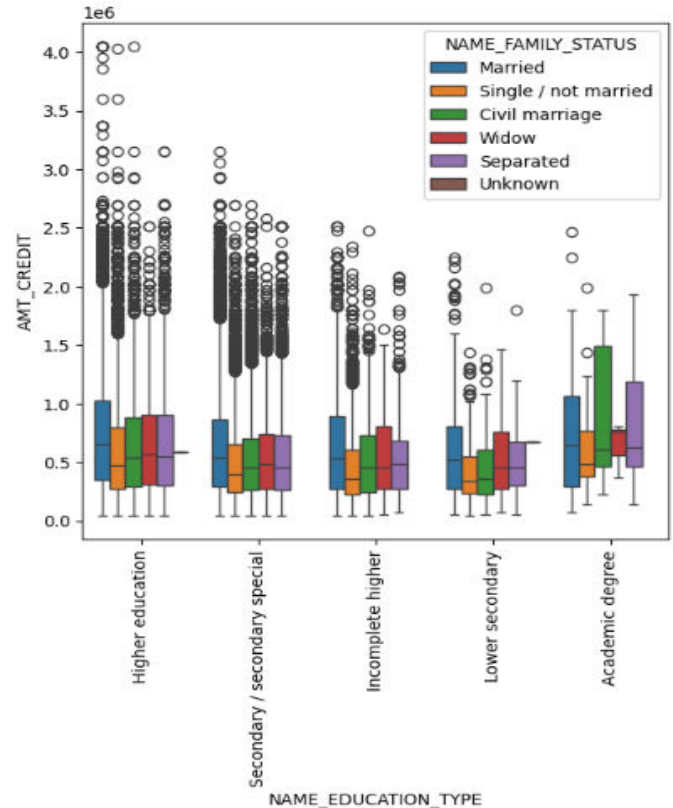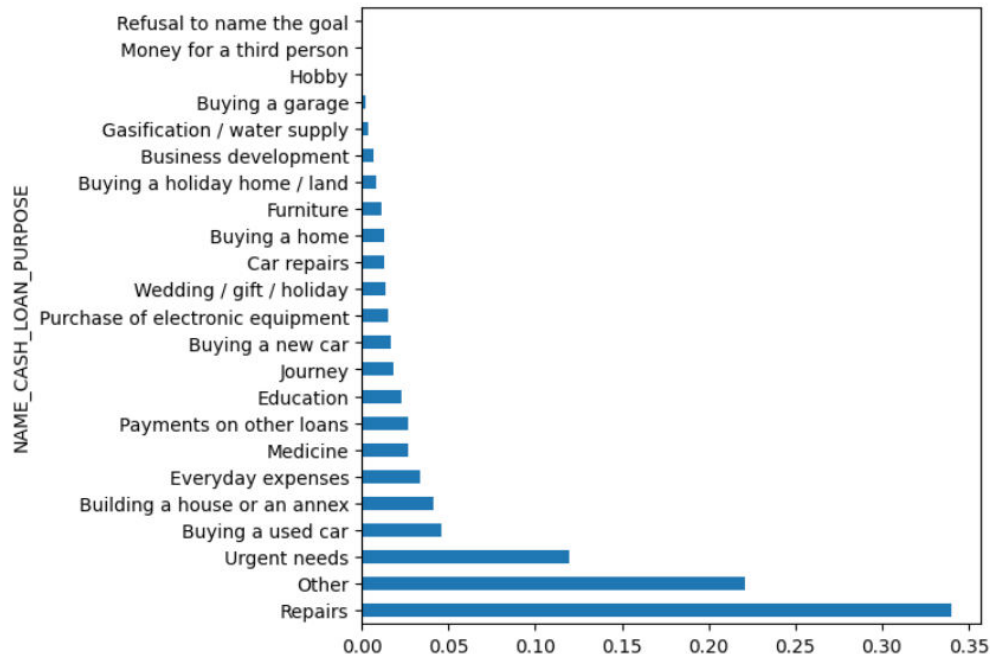
Merging 'application_data.csv' and 'previous_application.csv'

mergedf=pd.merge(left=appdf,right=prevdf,how='inner',on='SK_ID_CURR')

- Merging application_data.csv and previous_application.csv by using inner join method on the variable SK_ID_CURR'.
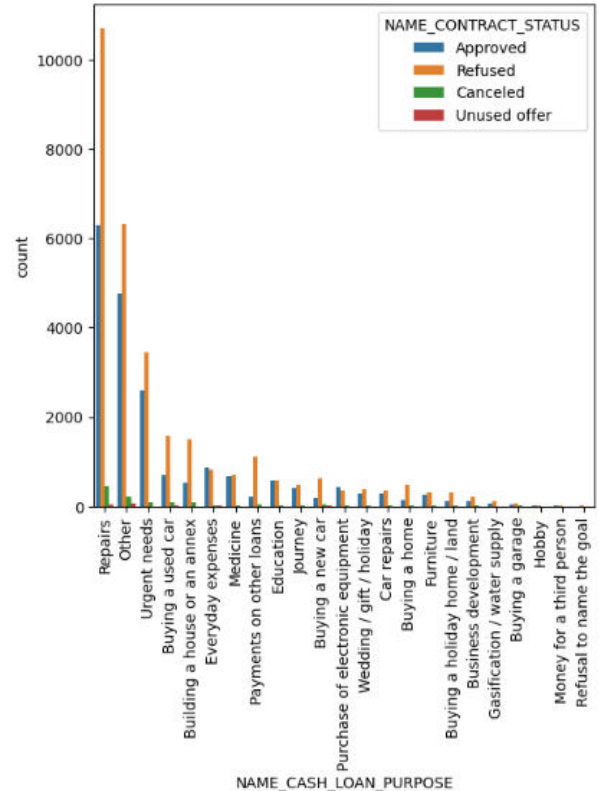
Univariate analysis on NAME_CASH_LOAN_PURPOSE

- Most of the clients taking loans for Repairs.

Bivariate analysis on NAME_CASH_LOAN_PURPOSE x NAME_CONTRACT_STATUS

- Education reason has equal chances of Approving and Refusing.
- Most applications got refused for the reason Repairs.
- Everyday expenses, Purchase of electronic equipment got more approvals than refusing.
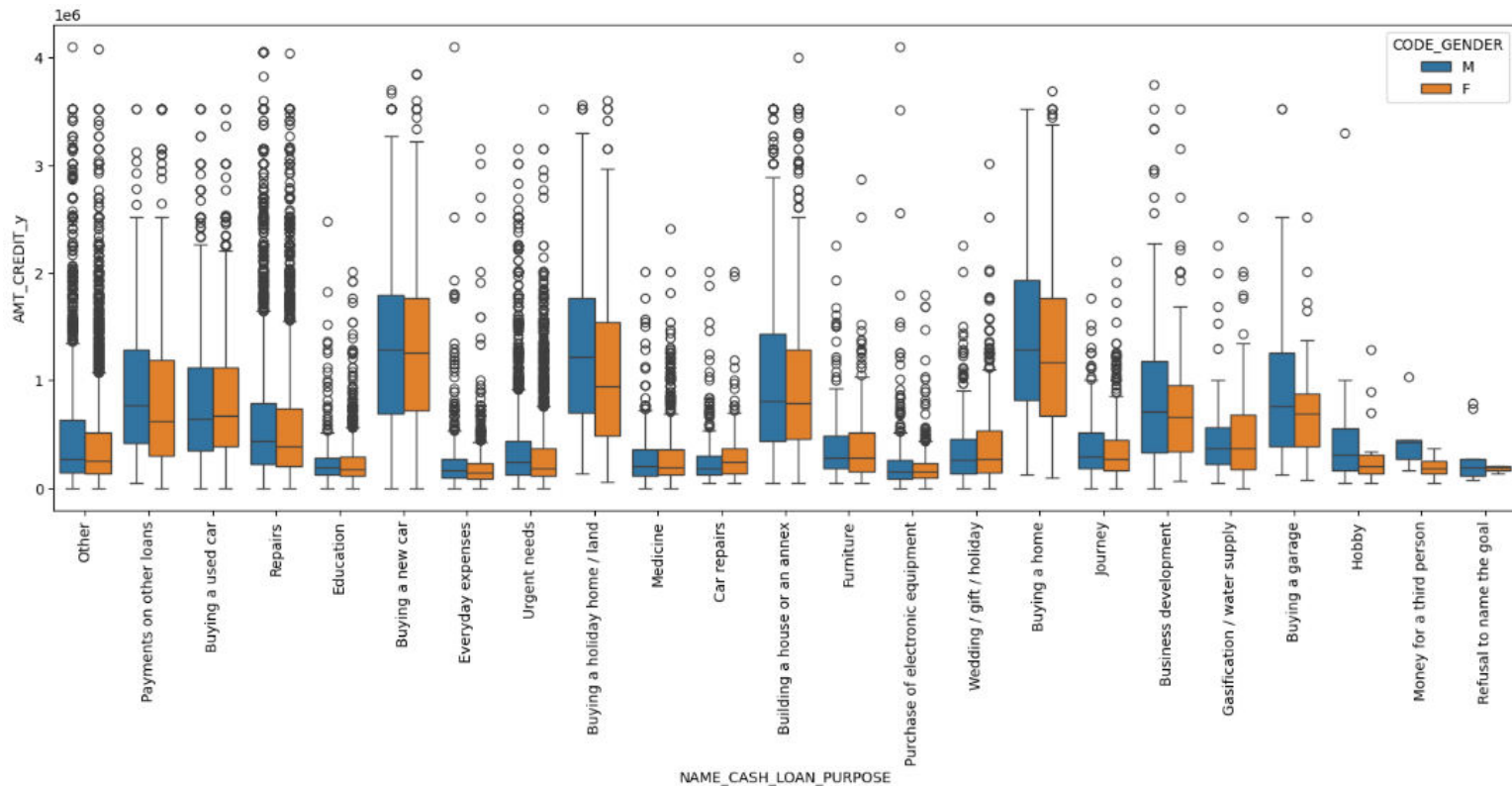
Multivariate analysis on NAME_CASH_LOAN_PURPOSE x NAME_INCOME_TYPE x AMT_CREDIT (in prev)
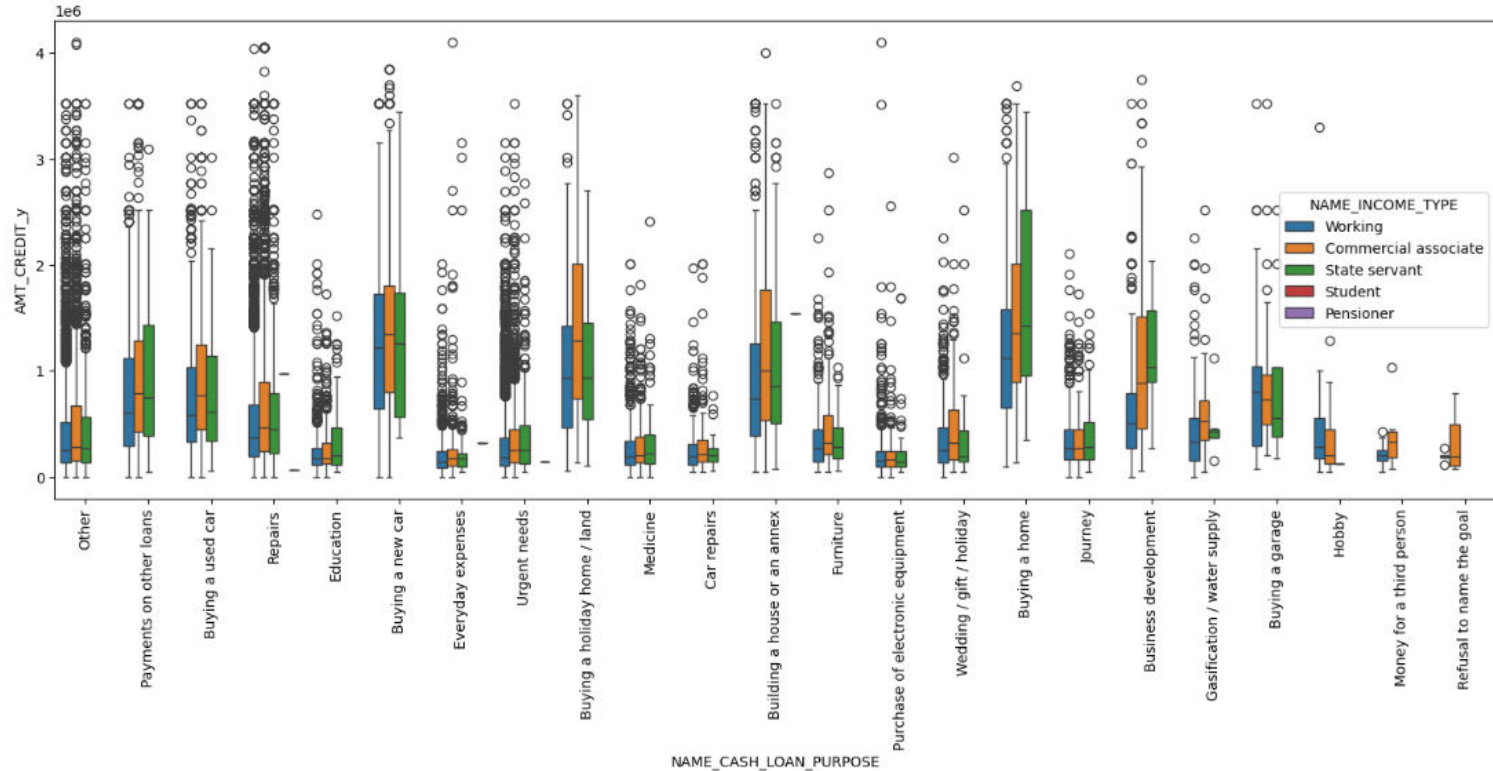
Multivariate analysis on NAME_CASH_LOAN_PURPOSE x NAME_INCOME_TYPE x AMT_CREDIT (in prev)

- Working clients have low credit amt.
- Commercial associate has high credit amt.
- Buying a new car, Buying a holiday home, Buying a house has high credit amt.
- Hobby, Money for a third person, Refusal to name the goal has low credit amt.
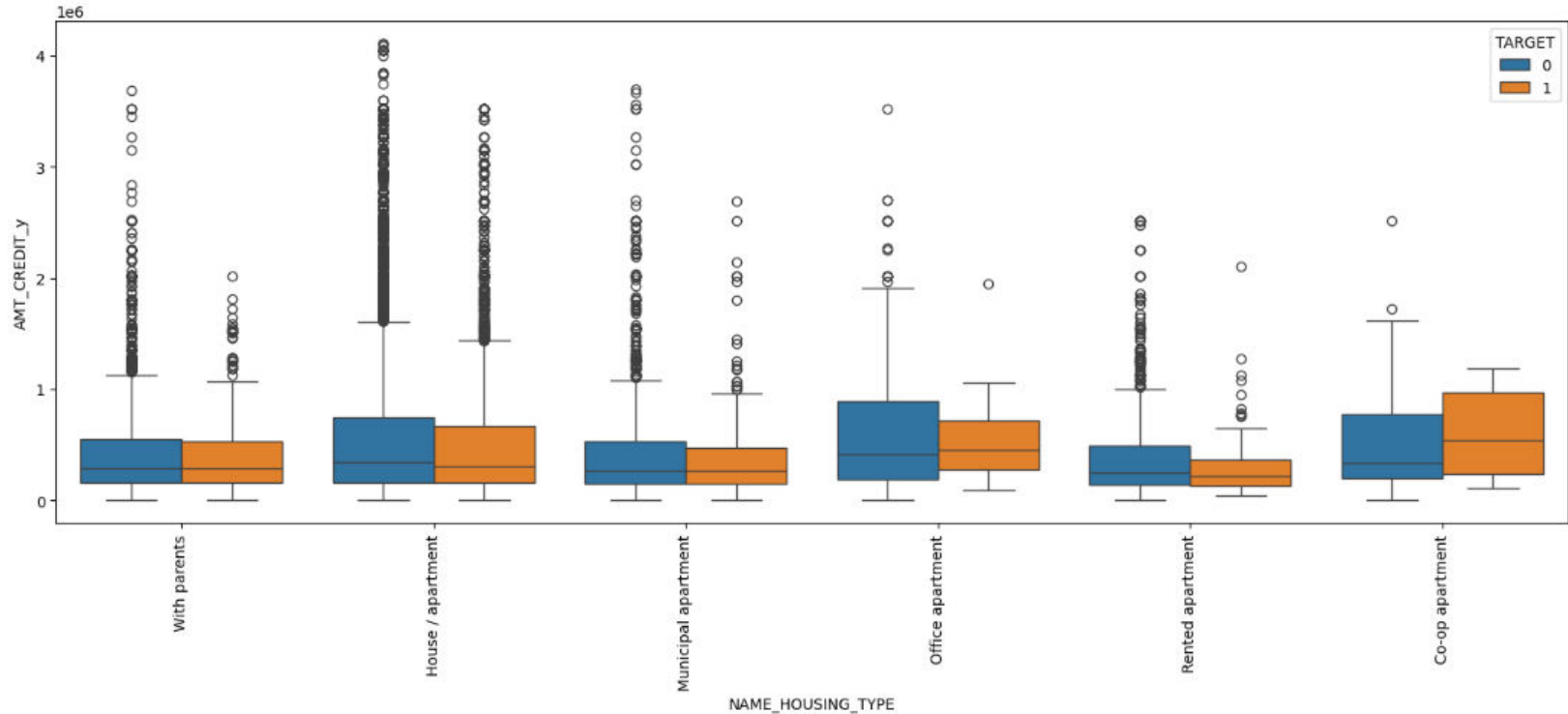
Multivariate analysis on NAME_CASH_LOAN_PURPOSE x CODE_GENDER x AMT_CREDIT (in prev)

Multivariate analysis on NAME_CASH_LOAN_PURPOSE x CODE_GENDER x AMT_CREDIT (in prev)

- Men has more credit amt than women.
- For car repairs, furniture women have more credit amt.

Multivariate analysis on NAME_HOUSING_TYPE x AMT_CREDIT (in prev) x TARGET

Multivariate analysis on NAME_HOUSING_TYPE x AMT_CREDIT (in prev) x TARGET

- House/apartment, office apartment, rented apartment have high credit amt for target = 0.
- Co-op apartment has less credit amt for target = 0.

Multivariate analysis on NAME_INCOME_TYPE x AMT_CREDIT (in prev) x TARGET
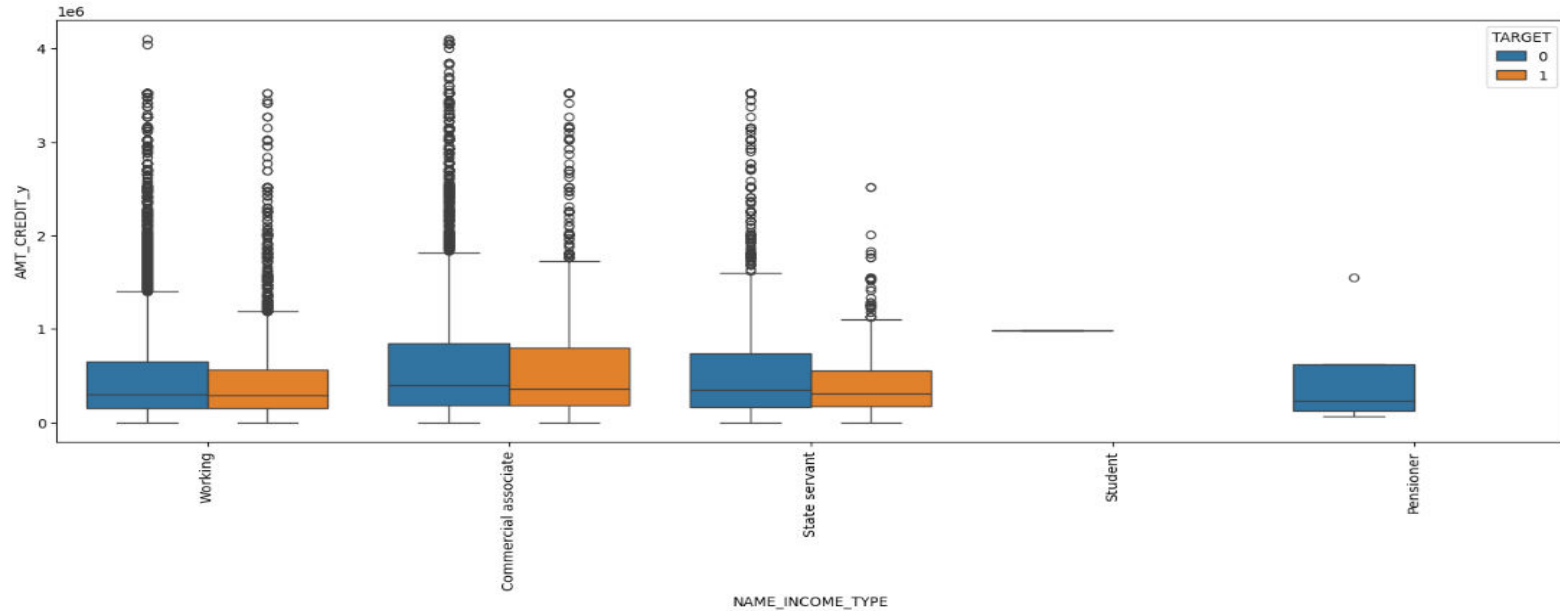
- Pensioner has very have high credit amt for target = 0 than target = 1 .

Conclusion

- Bank should focus on giving loans to Pensioners.
- Banks can try to avoid Co-op apartment clients and can focus on House/apartment, office apartment, rented apartment.
- Males have comparatively little high credit amt than females.
- Banks can avoid working clients and focus on Commercial associates.
- Banks can focus on clients with purpose Buying a new car, Buying a holiday home, Buying a house and avoid Hobby, Money for a third person, Refusal to name the goal.