# CASE STUDY

DECISION MAKING FOR BANK LOANS

Analysis of New Applications

- The analysis that are presented in the following slides is based on the data of new applications that have poured in for the loans and the inferences are drawn based on the plots of the same.

- There are multiple stages of analysis namely univariate , bivariate and which this document would take you through by providing meaningful insights regarding what the plots are all about.

- The inference numbering has been done based on the python notebook

Univariate analysis of Application dataset

**INCOME RANGE**

Inference 1.1

- From the above plot we can infer that the maximum number of clients without payment difficulties lie in the income range 1.2lac-1.5lac and immediate next income range is 2lac-.25 lac.

- Most clients with payment difficulties lie in the income range is 4.5lac-4.75lac
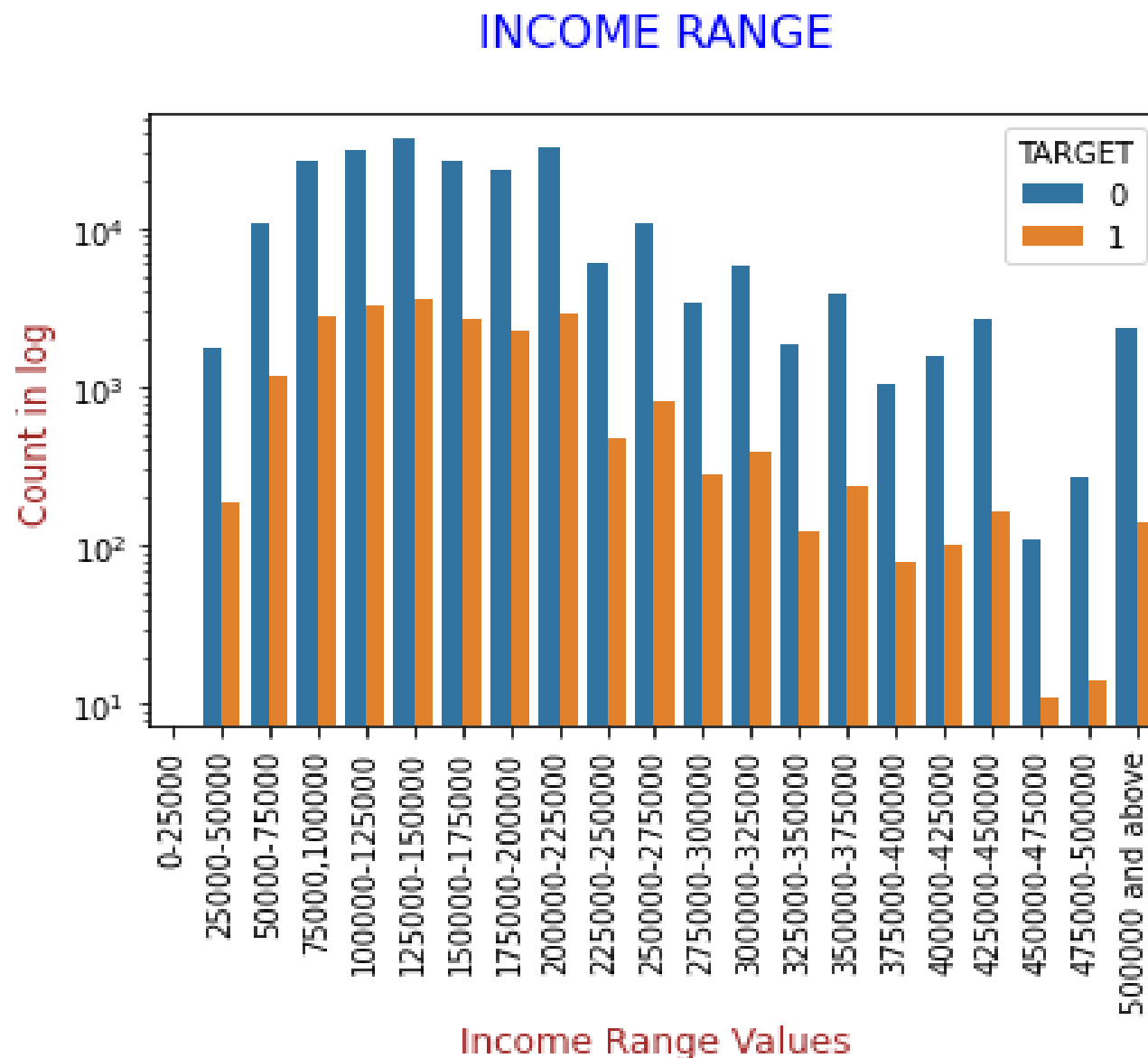
Inference 1.2

- It can be seen that both number of males and females is same for having payment difficulties

- It can also be seen that number of females is more than males for not having payment difficulties
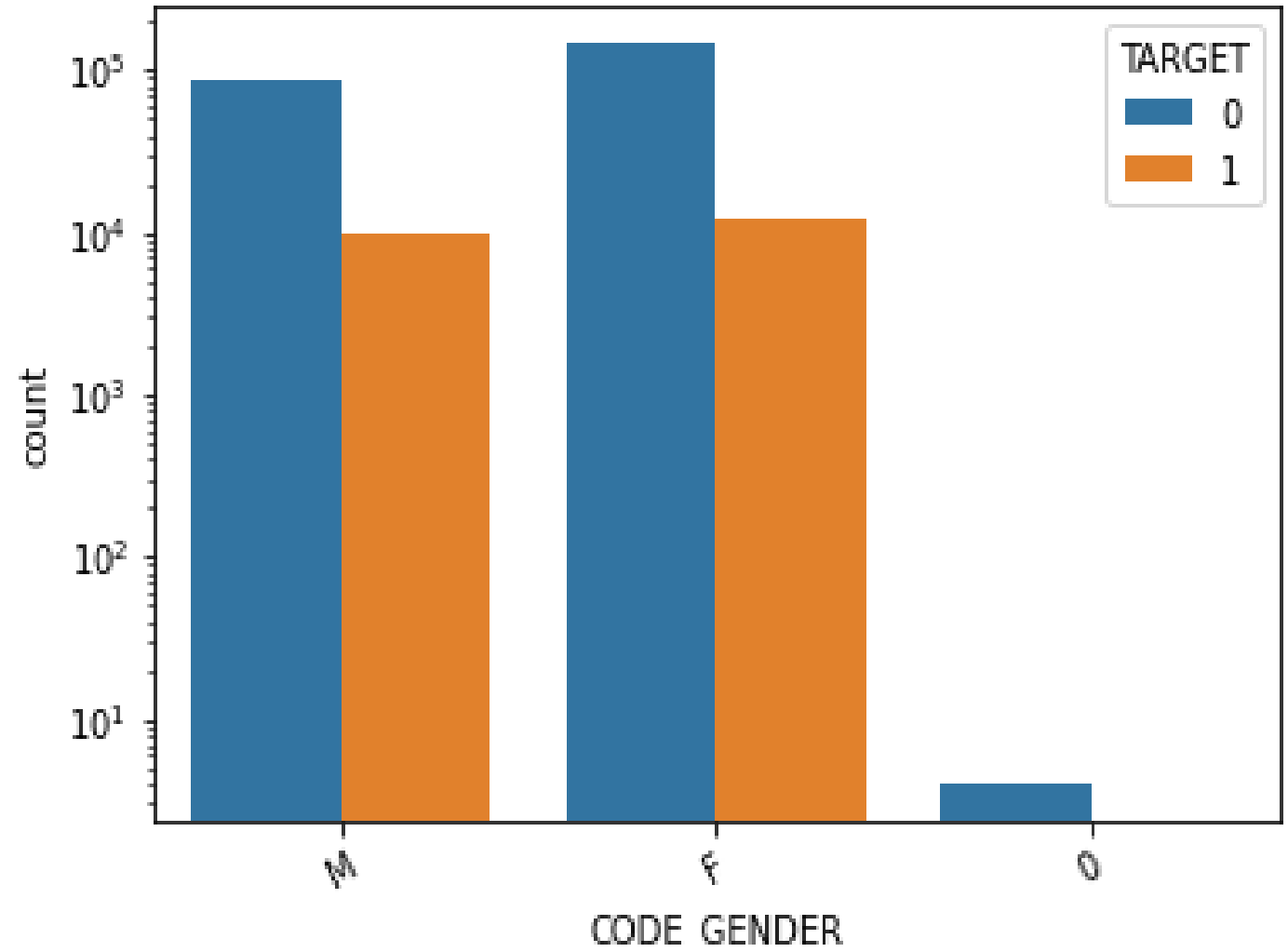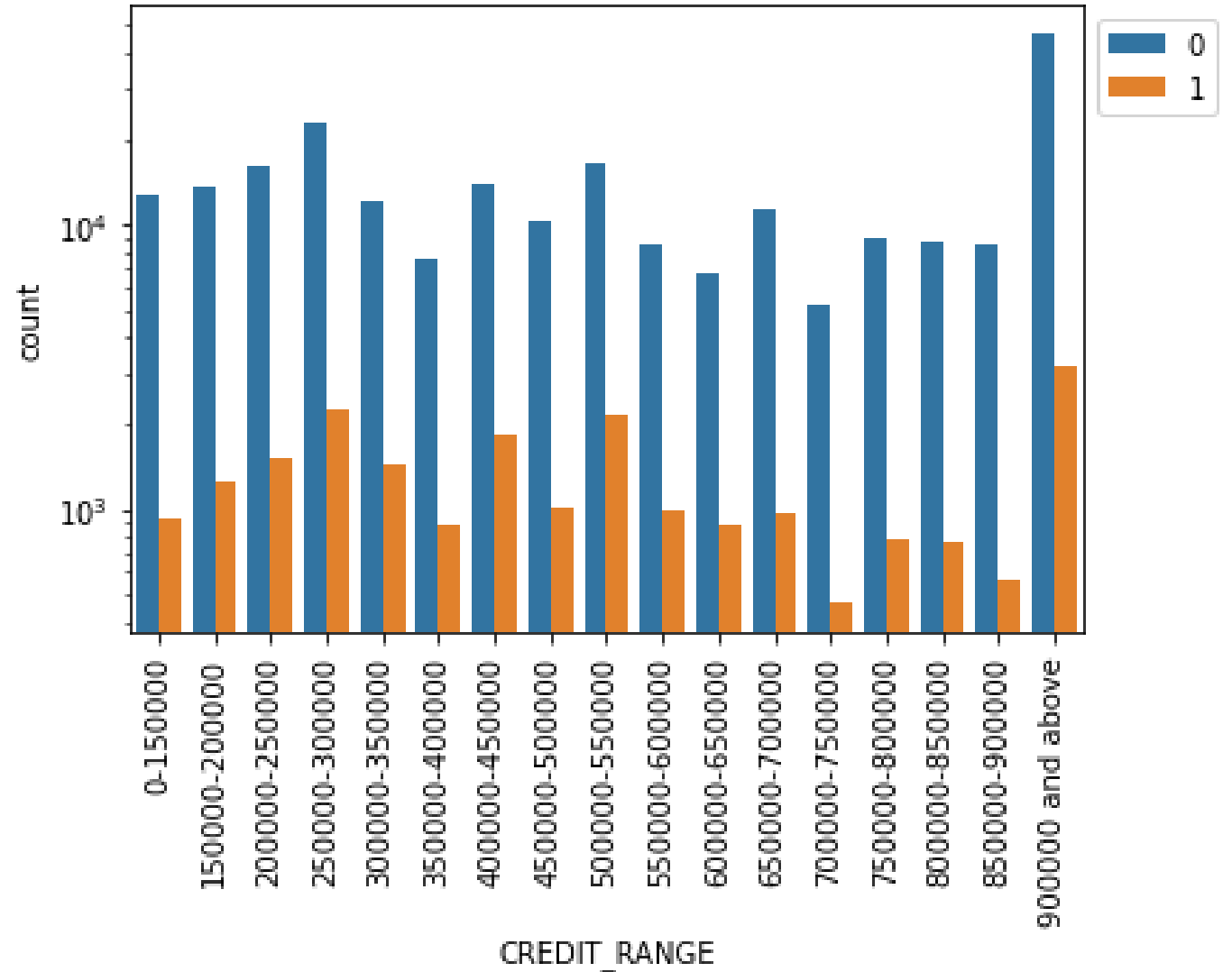
Inference 1.4

- It can be seen that Trade: type 4, organisation type have least count of payment difficulty clients.

- Most clients with no payment difficulties lie in organisation type named Business Entity Type 3.


Distribution of Organization type
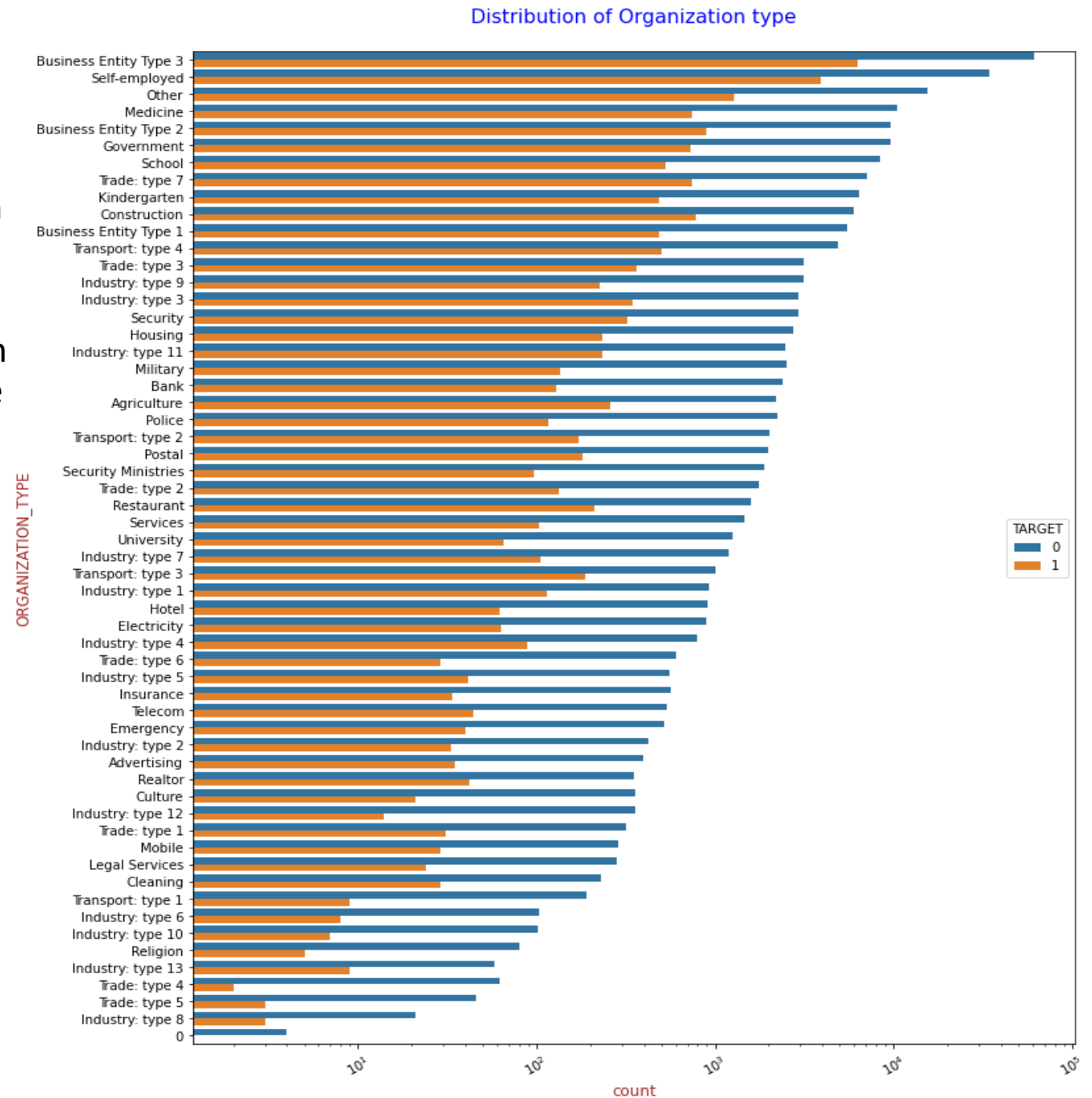
Inference 1.6

- Clients with credit range lying in 900000 and above are the ones who are capable of paying the loans back

- Least number of clients lying in income range 7lac- 7.5 lac are not capable of paying

Credit Range w.r.t Target

Bivariate Analysis of Application dataset

- **Inference 1.7**

- Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.

- Also, higher education of family status of 'marriage', 'single or not' and 'civil marriage' are having more outliers. Civil marriage for Academic degree is having most of the credits in the third quartile.

Credit amount vs Family Status
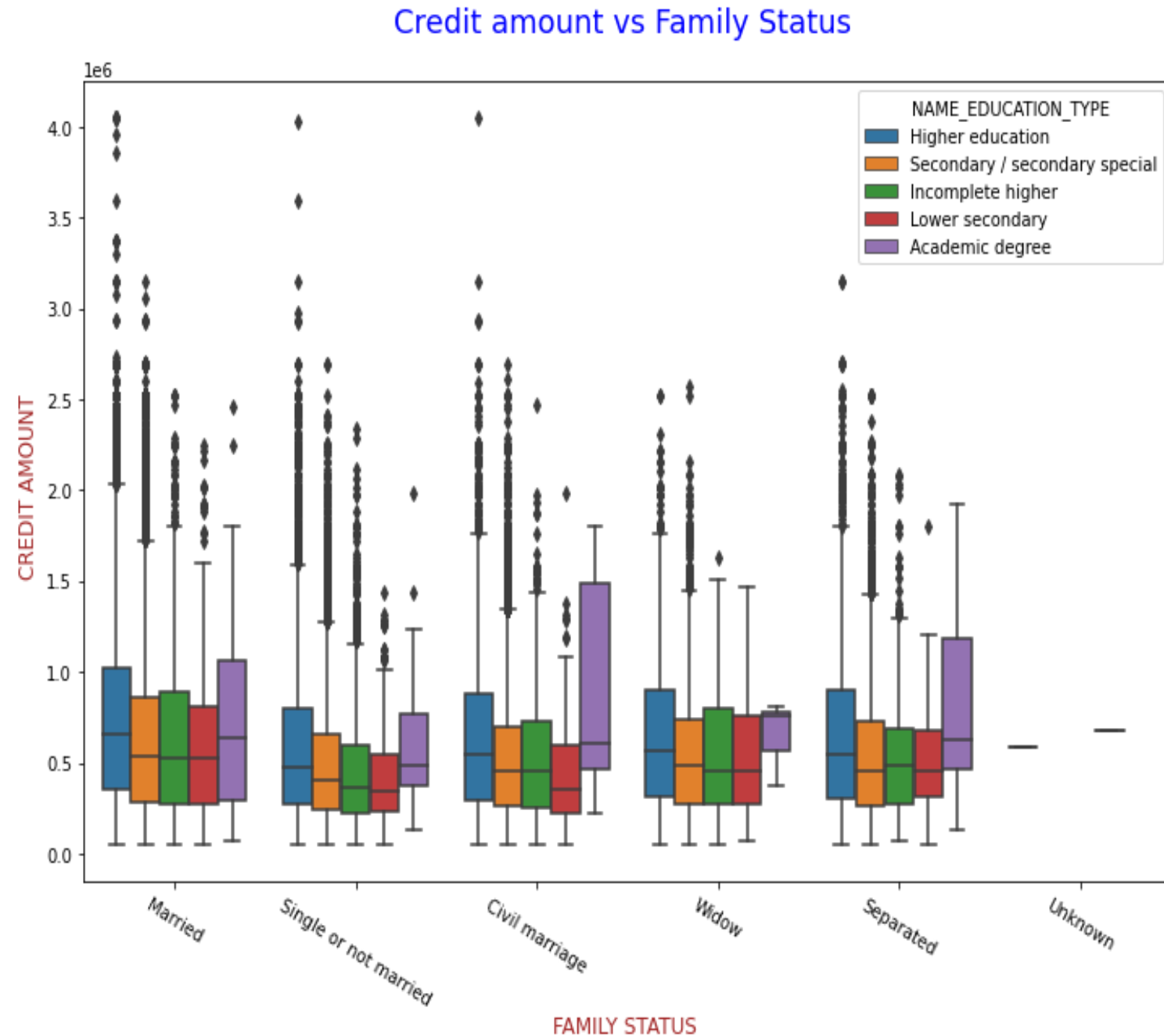
Inference 1.8

- Family status of 'civil marriage', 'marriage' and 'separated' of Higher education are having higher number of income than others.

- Also, higher education and secondary/second special education statuses with family status of 'marriage', 'single or not' and 'civil marriage' are having more outliers. Married for Higher education is having most of the incomes in the lower bound.

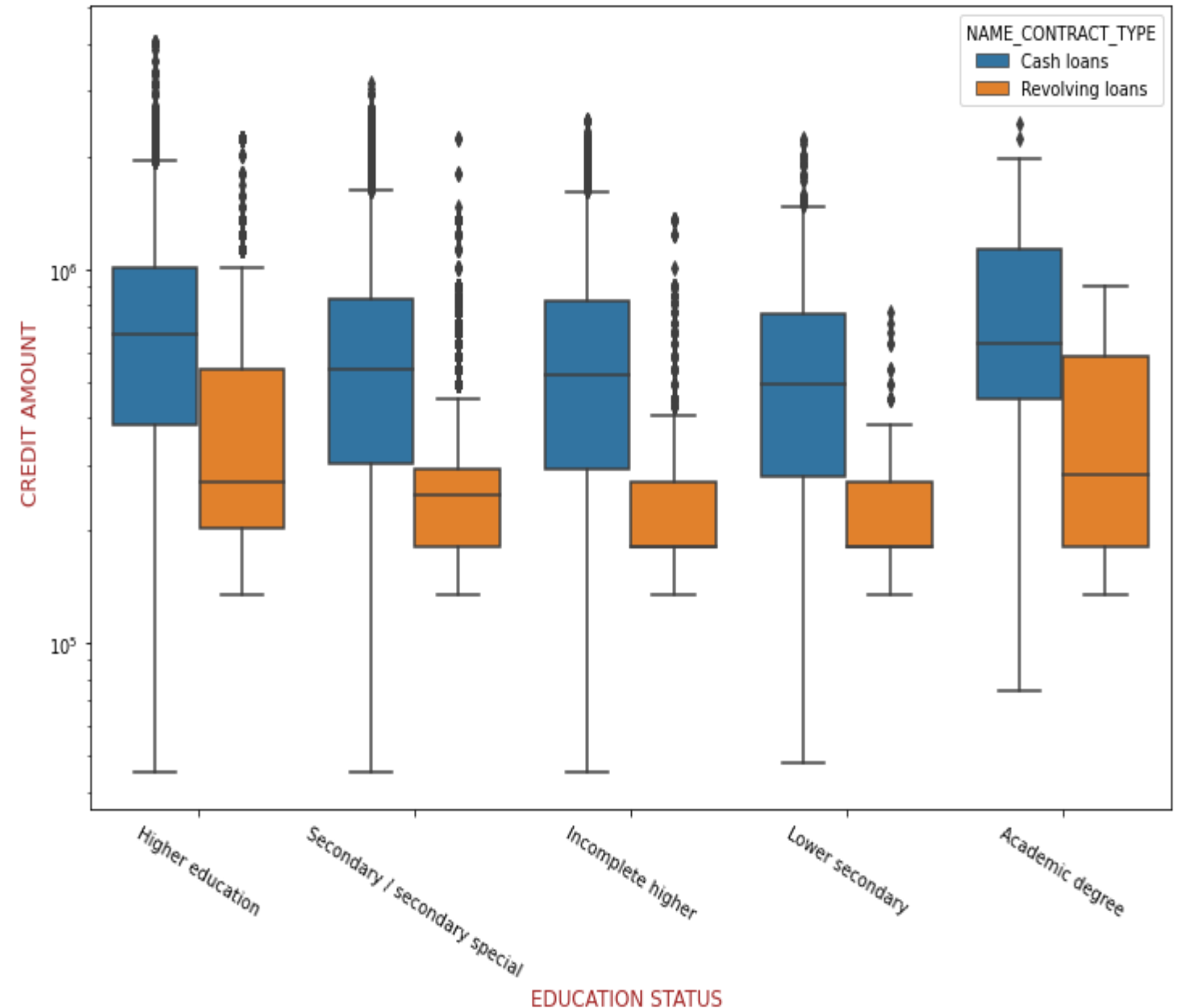
Total Income amount vs Family Status
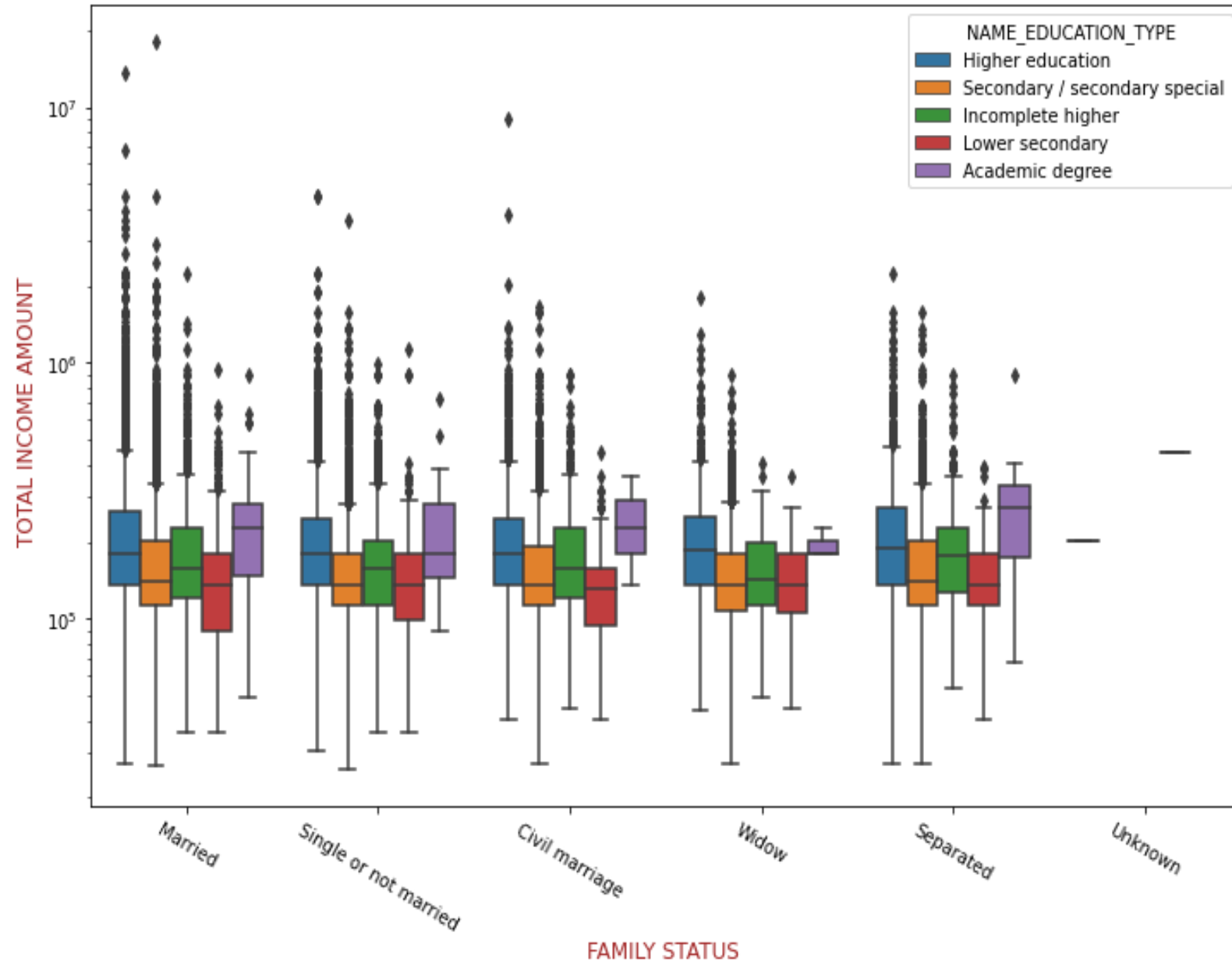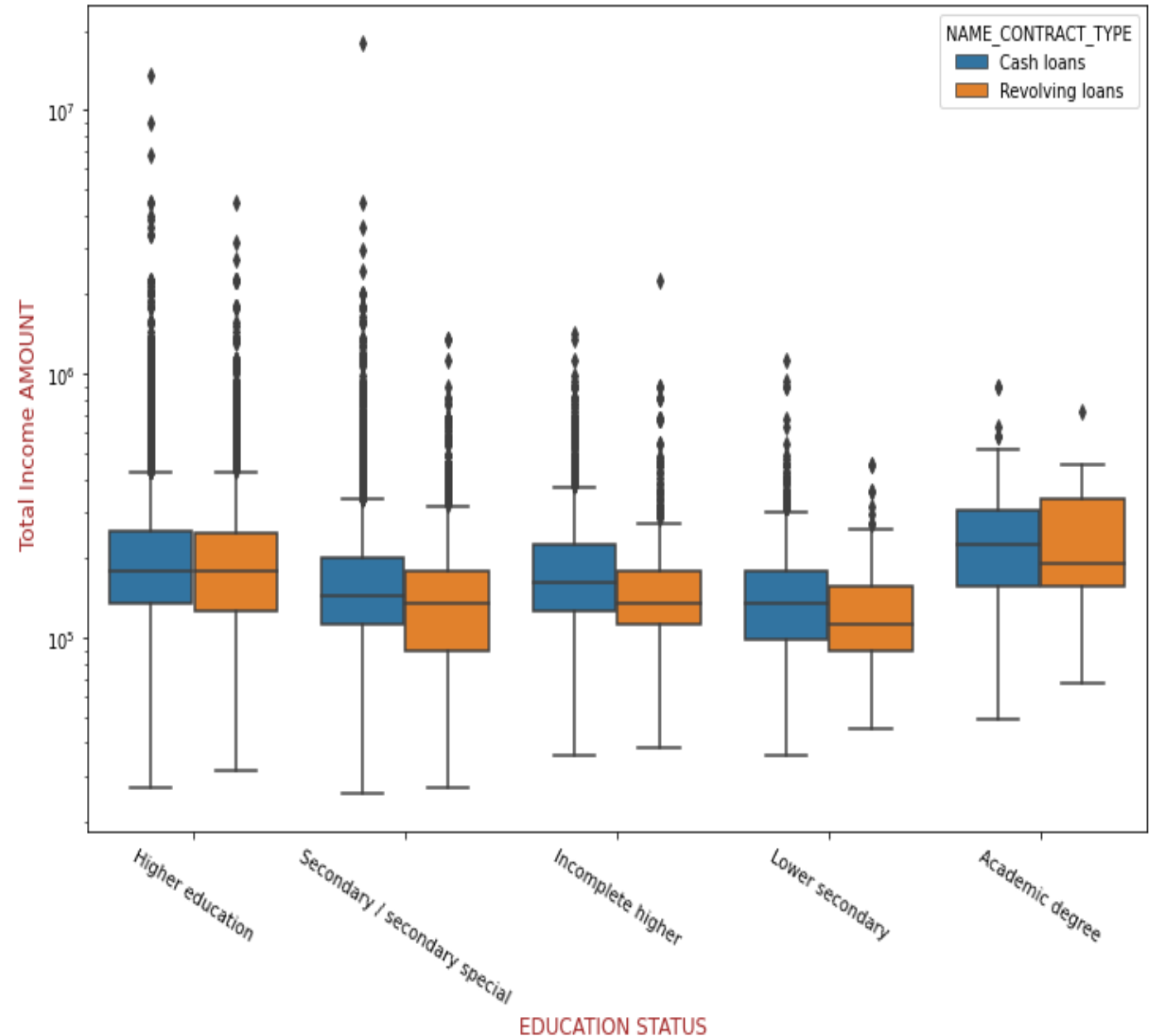
Inference 1.9

- Education status of higher education and secondary/secondary special have most clients for contract type cash loans

- Most number of clients applying for revolving loans are in education status Academic degree and higher education


Credit amount vs Education Status

Inference 1.10

• It can be seen that the clients with education status of Higher education are the maximum credit seekers with highest in terms of contract type of cash loans of contract type

• It can also be observed that the contract type revolving loans issued maximum credit amount holders education status Higher education

• The hightest credit amount in cash loans is given to a client with education level secondary/secondary special

• Basically education level or type is not playing much role as of who gets what amount of credit

Income amount vs Education Status

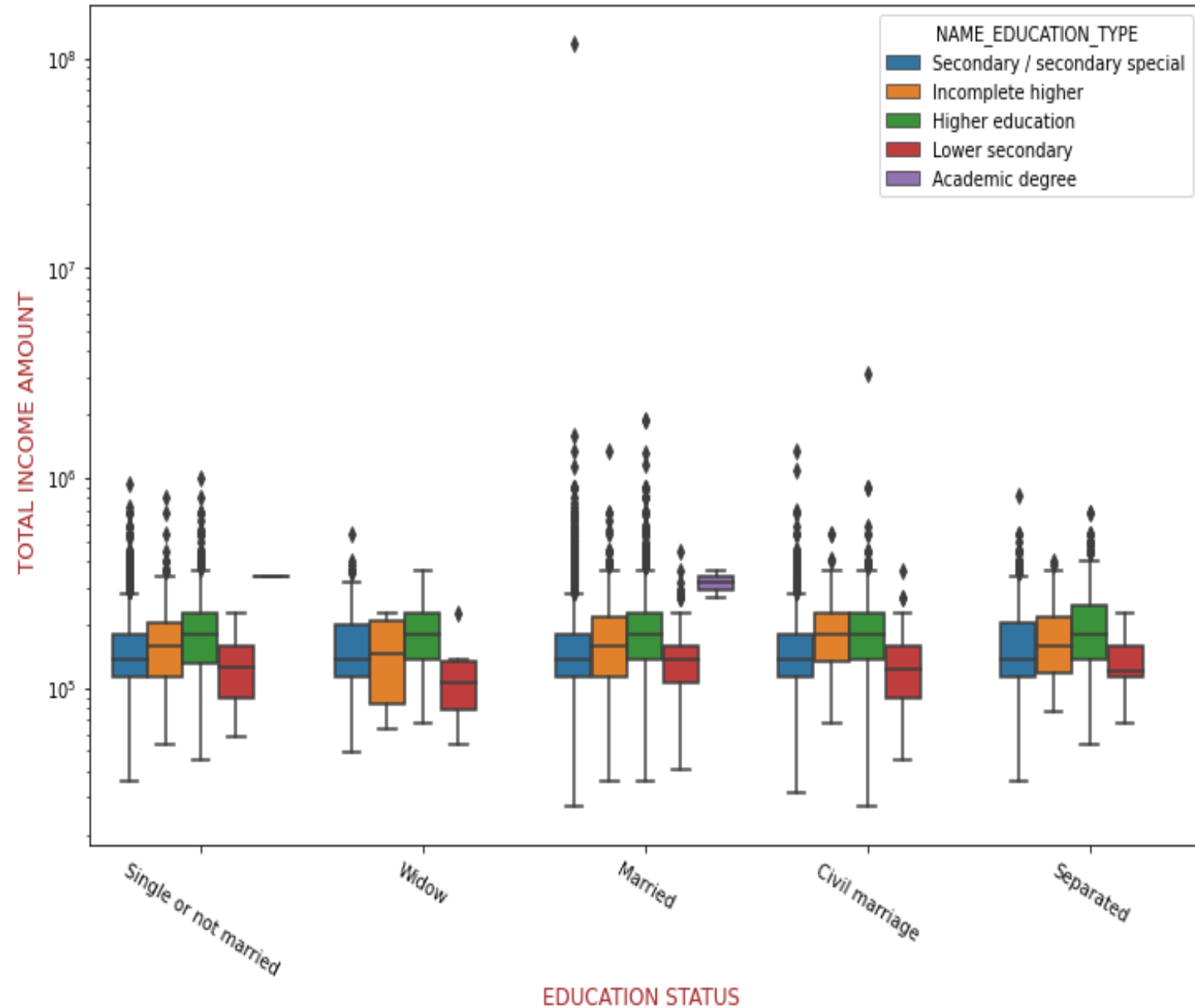# Inference 1.11

- Observations are Quite similar with Target0

- Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having less number of credits than others.

- Most of the outliers are from Education type 'Higher education' and 'Secondary'.

- Most number of all types of education as well as family lie in lower bound

Credit amount vs Education Status
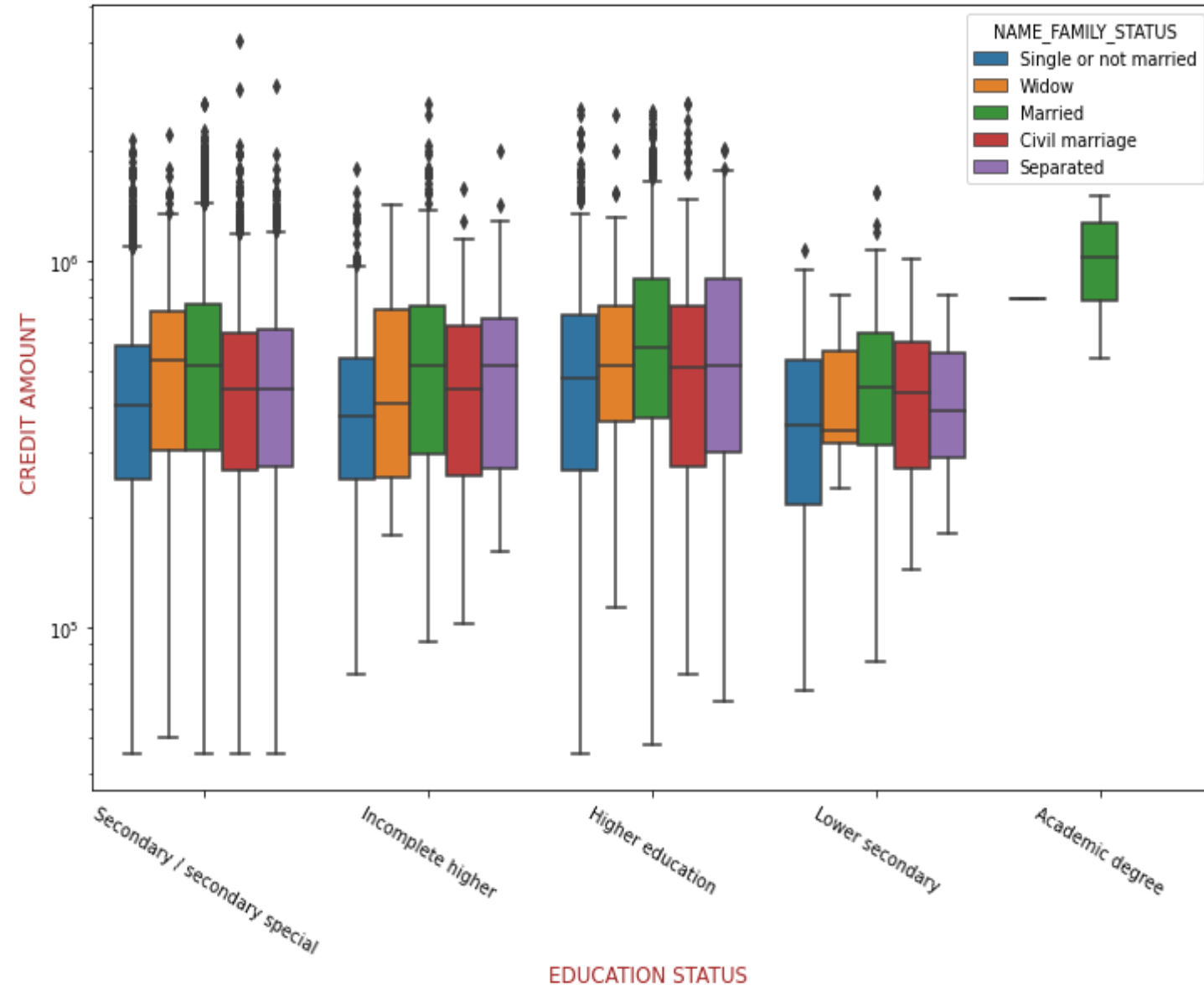
Inference 1.12

- This is also having some similarity with target0,

- Almost all Family types and education types have the income amount mostly equal.

- Least outlies are for Lower secondary and their income amount is also little lesser than that of all other education types.

- Academic degree are very less number of people in payments with difficult data frame named target1

Total Income amount vs Education Status

Inference 1.13

- Its similar to that of inference 1.10

- Even though people have difficulties with payments, loan application are still pouring in with similar credit amount demands from the clients
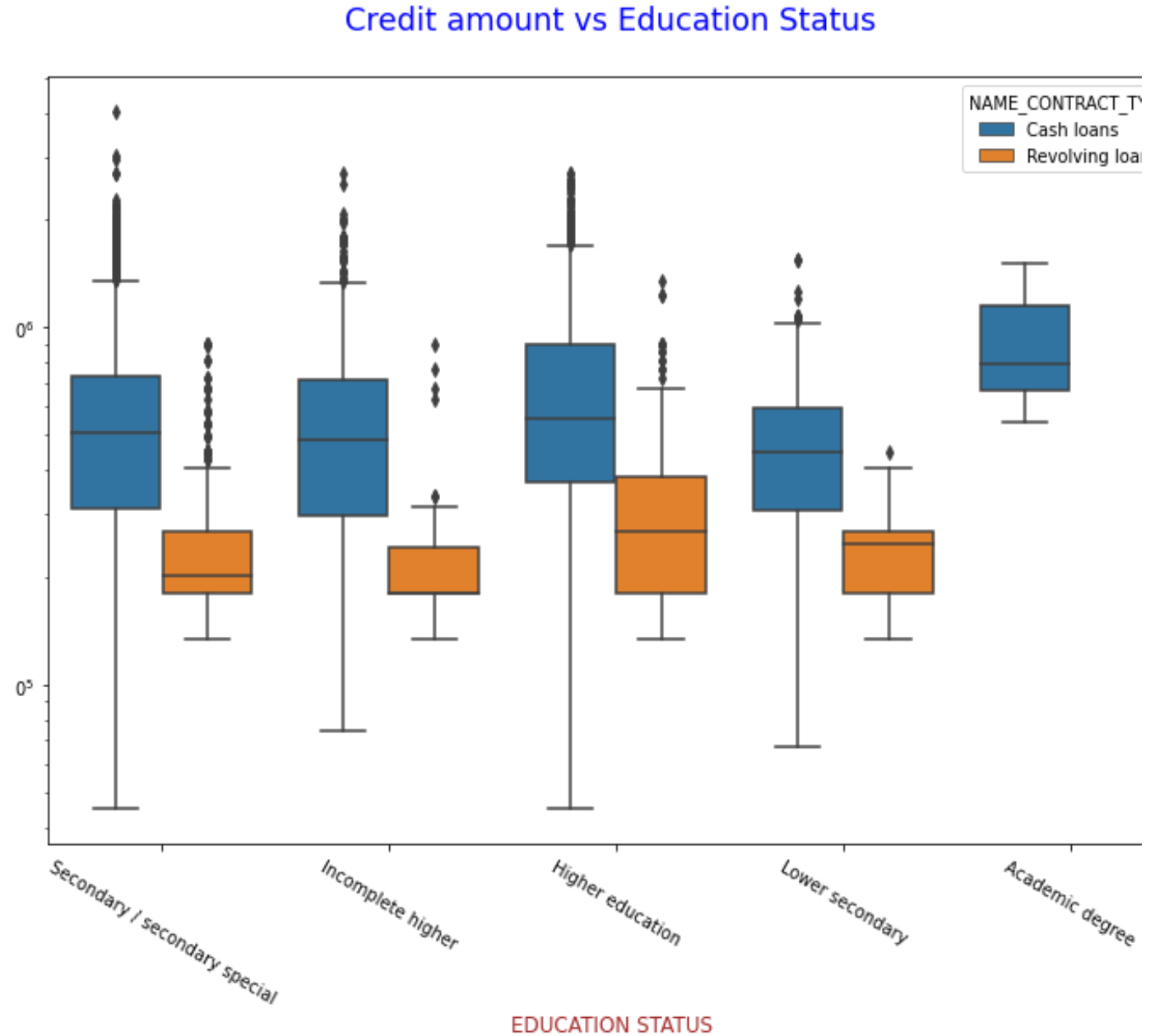
Credit amount vs Education Status

Inference 1.14

- It can be seen that the maximum income amount holders are with education levels secondary and higher education levels

- Its strange that, there are no revolving amount loans are demanded by clients with education level academic degree

Income amount vs Education Status
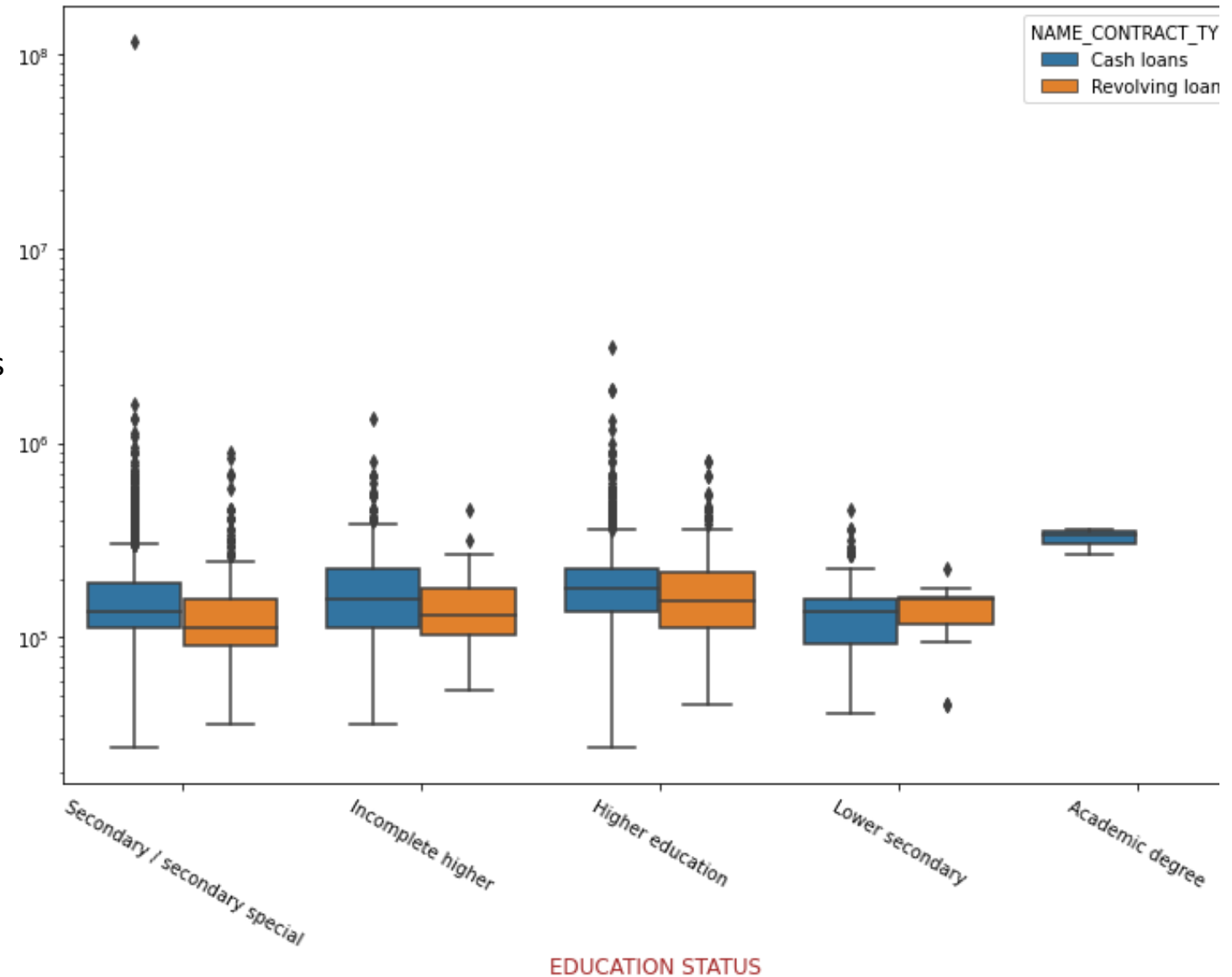
Multivariate Analysis on Application dataset

Inference 1.15

MOST RELEVANT CORRELATIONAL PAIRS with no difficulties of payment

• It is observed that the maximum correlation is between the variables listed below in pairs

- AMT_CREDIT to AMT_ANNUITY
- AMT_CREDIT to AMT_TOTAL_INCOME
- AMT_ANNUITY to AMT_INCOME_TOTAL
- REG_CITY_NOT_WORK_CITY to REG_CITY_NOT_LIVE_CITY
- DAYS_EMPLOYED to DAYS_BIRTH
- DAYS_BIRTH to DAYS_REGISTRATION

• Least or negative correlation is between the variables listed below in pairs

- DAYS_BIRTH to CNT_CHILDREN

CORRELATION TABLE FOR target0

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_PUBLISH	HOUR_APPR_PROCESS_START	LIVE_REGION_NOT_WORK_REGION	REG_CITY_NOT_LIVE_CITY	REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY
CNT_CHILDREN	1	-0.0095	-0.018	-0.0072	-0.03	-0.24	-0.061	-0.16	0.12	-0.033	-0.007	-0.0017	0.01	0.018
AMT_INCOME_TOTAL	-0.0095	1	0.33	0.4	0.17	0.046	0.031	-0.035	0.026	0.056	0.12	-0.0042	-0.02	-0.011
AMT_CREDIT	-0.018	0.33	1	0.76	0.1	0.15	0.089	0.015	0.035	0.04	0.052	-0.038	-0.038	-0.015
AMT_ANNUITY	-0.0072	0.4	0.76	1	0.12	0.092	0.053	-0.0099	0.026	0.034	0.072	-0.019	-0.029	-0.014
REGION_POPULATION_RELATIVE	-0.03	0.17	0.1	0.12	1	0.041	-0.0093	0.06	0.0039	0.17	0.092	-0.054	-0.049	-0.017
DAYS_BIRTH	-0.24	0.046	0.15	0.092	0.041	1	0.35	0.3	0.11	-0.049	-0.014	-0.17	-0.11	-0.031
DAYS_EMPLOYED	-0.061	0.031	0.089	0.053	-0.0093	0.35	1	0.17	0.083	-0.023	-0.063	-0.11	-0.13	-0.074
DAYS_REGISTRATION	-0.16	-0.035	0.015	-0.0099	0.06	0.3	0.17	1	0.029	0.032	-0.0091	-0.047	-0.055	-0.031
DAYS_ID_PUBLISH	0.12	0.026	0.035	0.026	0.0039	0.11	0.083	0.029	1	-0.0056	-0.0088	-0.053	-0.033	-0.0015
HOUR_APPR_PROCESS_START	-0.033	0.056	0.04	0.034	0.17	-0.049	-0.023	0.032	-0.0056	1	0.058	0.011	-0.0014	-0.0051
LIVE_REGION_NOT_WORK_REGION	-0.007	0.12	0.052	0.072	0.092	-0.014	-0.063	-0.0091	-0.0088	0.058	1	0.015	0.17	0.22
REG_CITY_NOT_LIVE_CITY	-0.0017	-0.0042	-0.038	-0.019	-0.054	-0.17	-0.11	-0.047	-0.053	0.011	0.015	1	0.44	0.012
REG_CITY_NOT_WORK_CITY	0.01	-0.02	-0.038	-0.029	-0.049	-0.11	-0.13	-0.055	-0.033	-0.0014	0.17	0.44	1	0.82
LIVE_CITY_NOT_WORK_CITY	0.018	-0.011	-0.015	-0.014	-0.017	-0.031	-0.074	-0.031	-0.0015	-0.0051	0.22	0.012	0.82	1

All columns of target0

Inference 1.16

•Almost all of the correlational pairs of variables are as same as in Inference 1.15
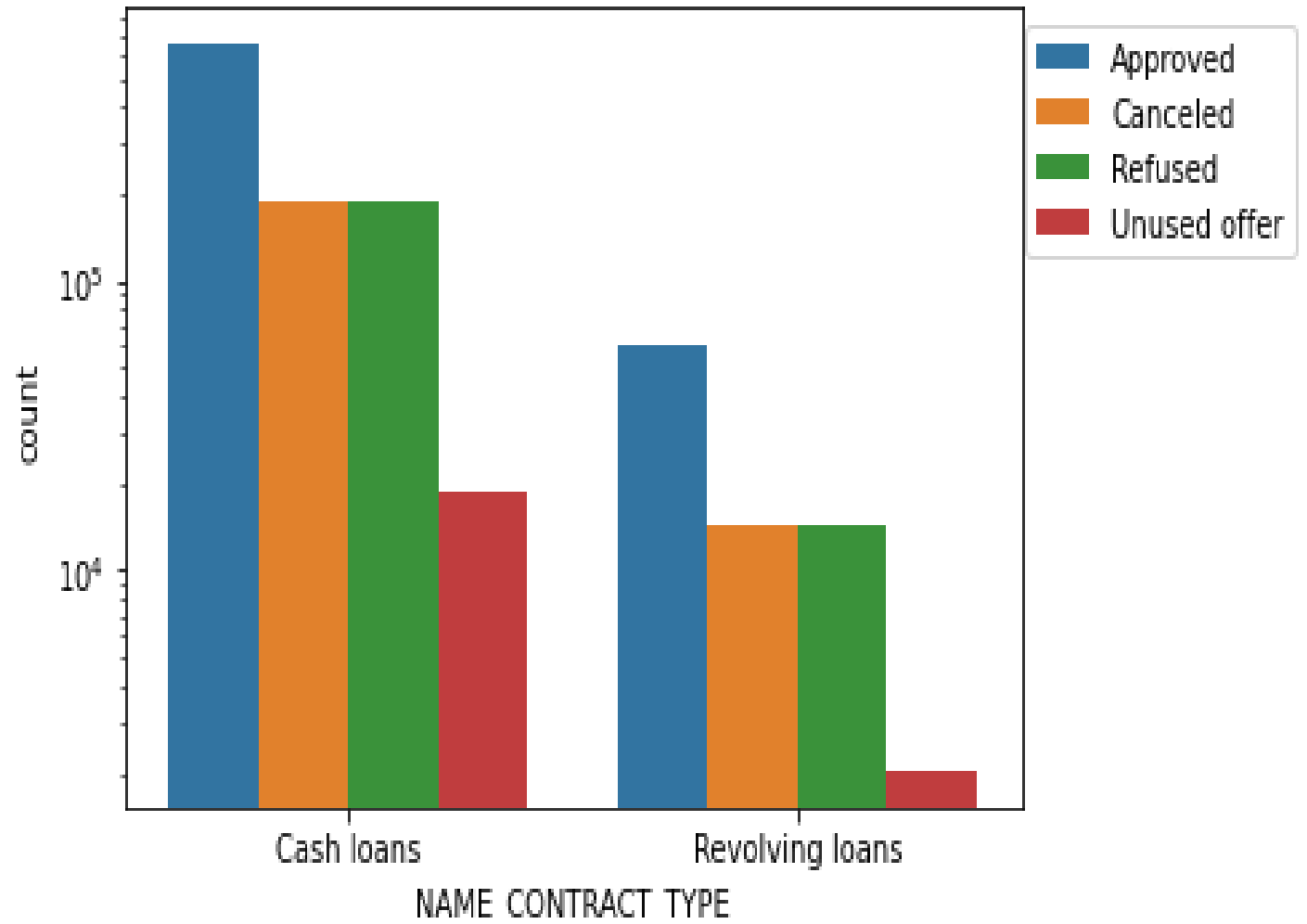

CORRELATION TABLE FOR target1
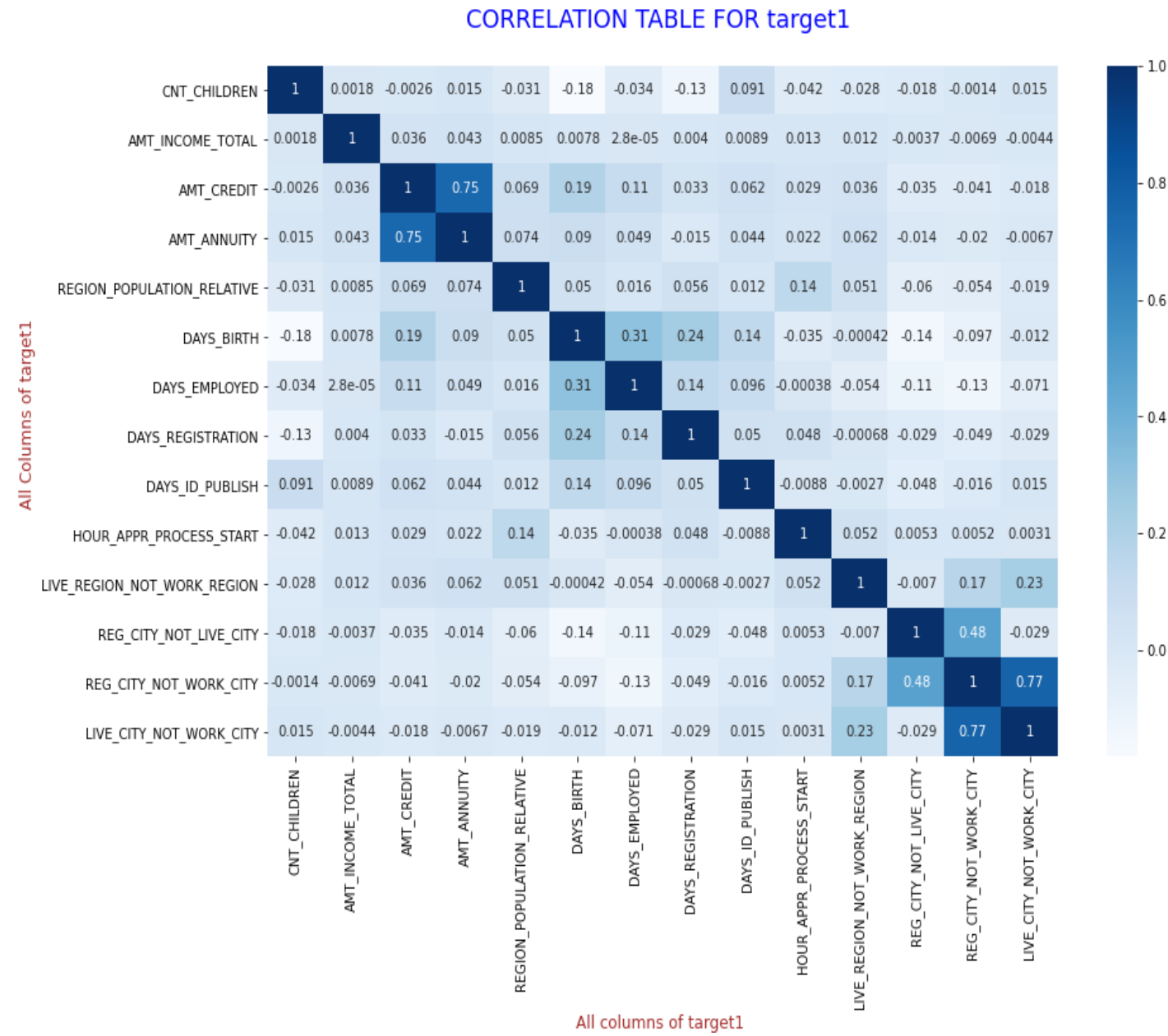
Univariate analysis on merged data frame

Inference 2.1

- Contract type cash loans are maximum in number where all kinds of contract statuses are more than contract type revolving loans
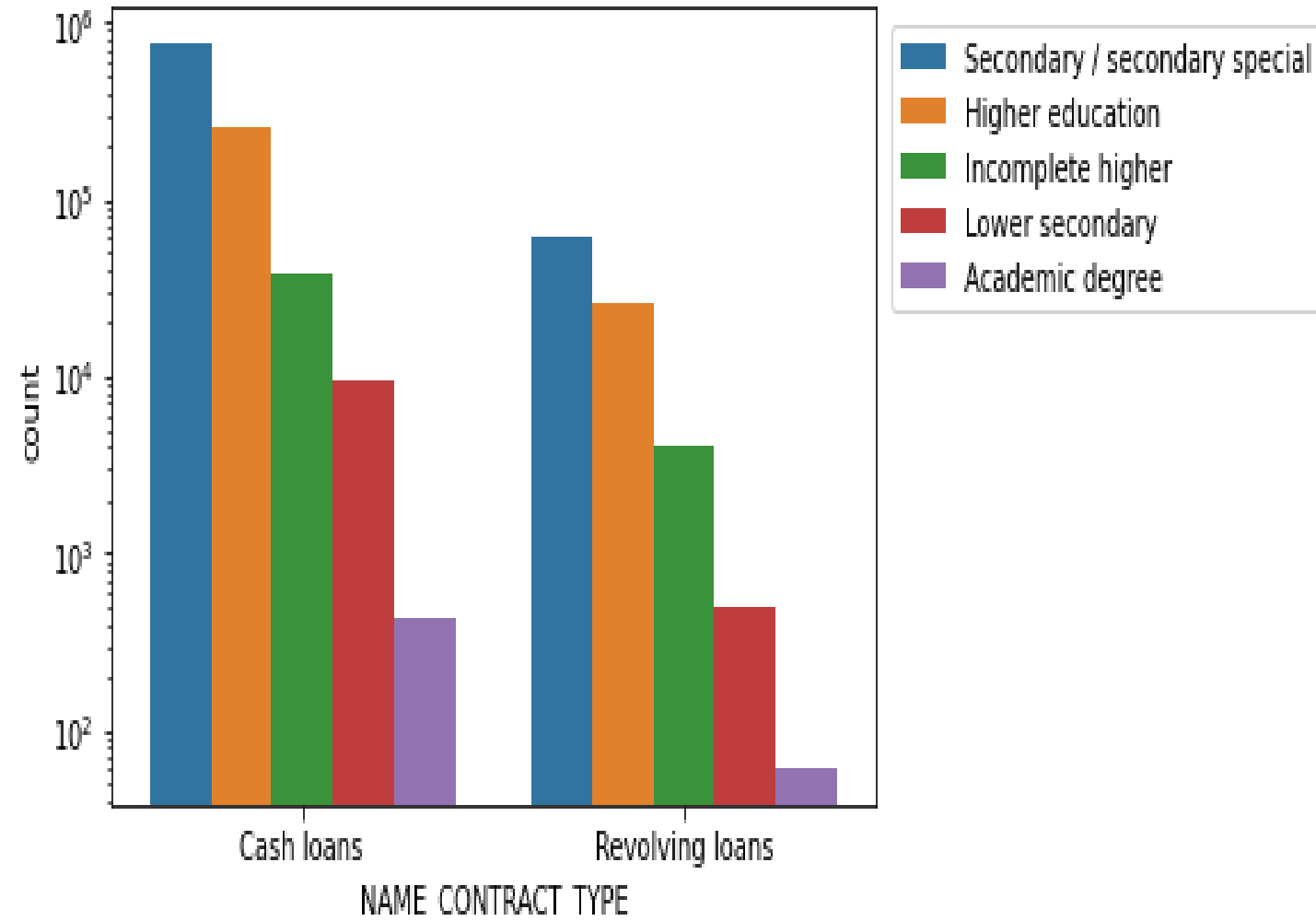

Count of Contract Types W.R.T Contract Status

Count of Contract Types W.R.T Education Type

Inference 2.2

- Most clients with all kinds of education types apply mostly for cash loans rather than revolving loans
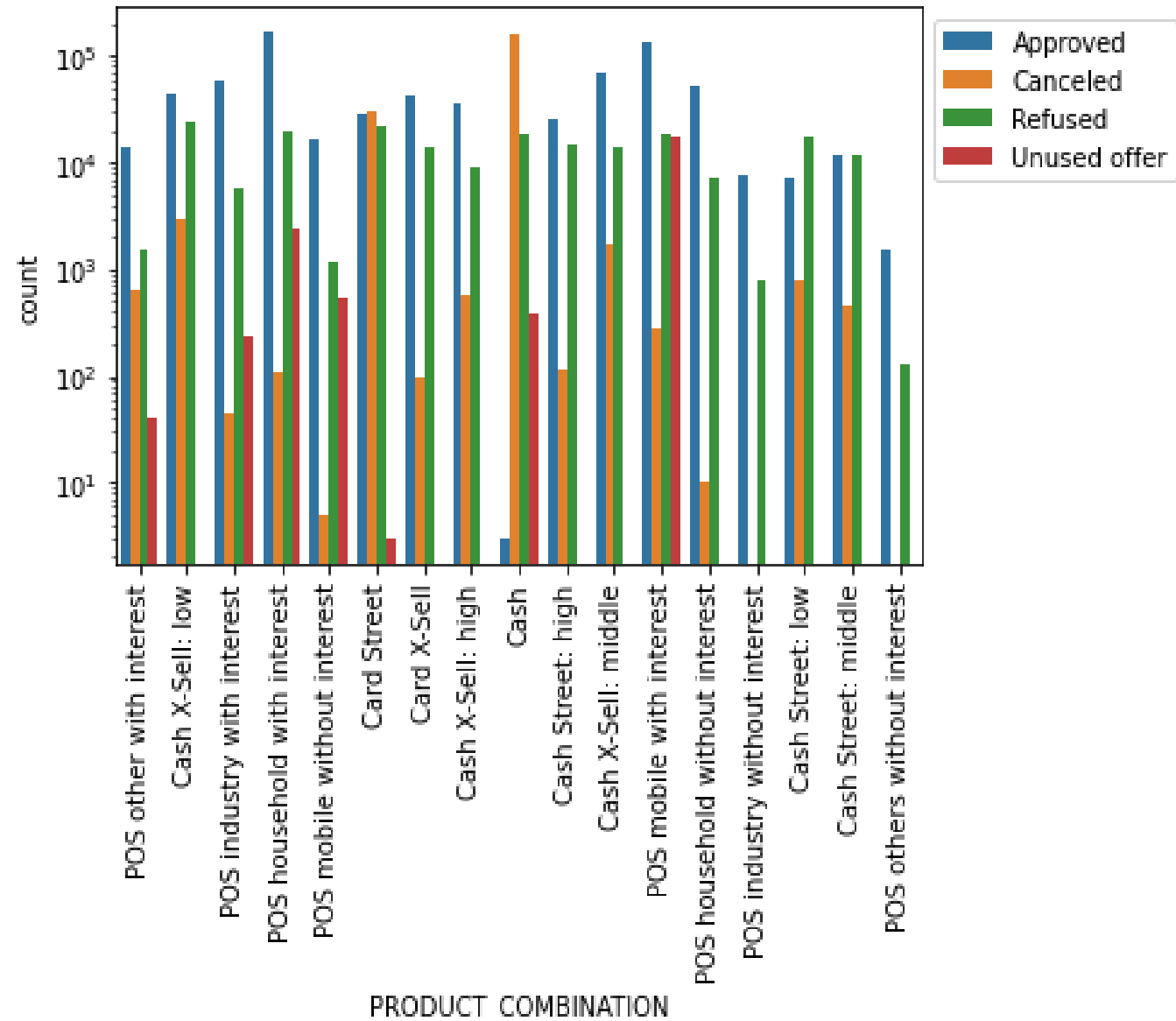
Inference 2.4

- Maximum number of contract types are of Product combination of POS household with interest followed by Cash and then followed by POS mobile with interest

- Least number of clients opt for product combination of POS others without interest
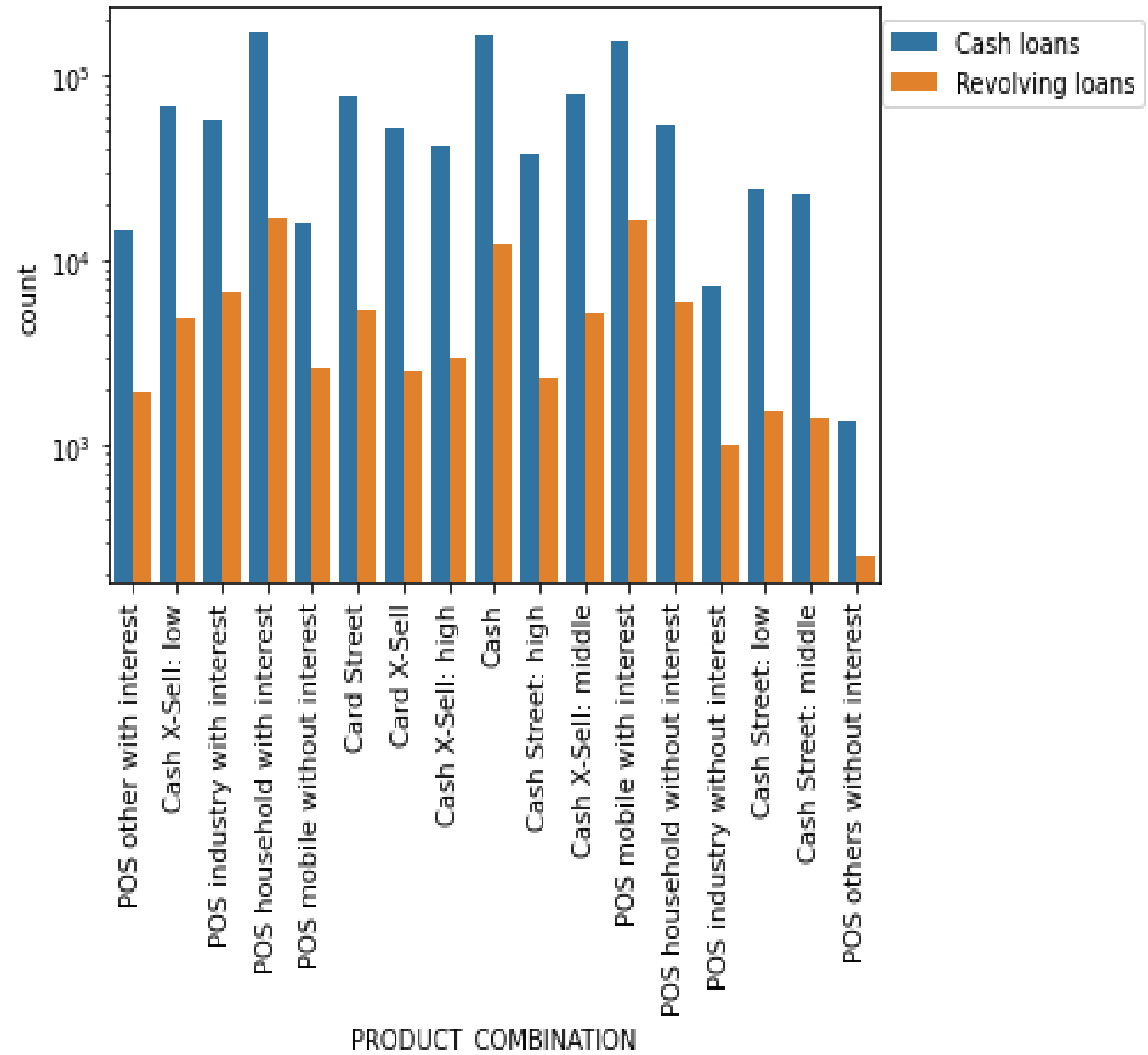
Count of Product Combinations W.R.T Contract Type

Inference 2.5

• Most canceled loans are of the product combination, Cash.

• Most refused loans are of product combination Cash X-Sell: low

• There are almost nil unused offers in product combinations listed below

  • Cash X-Sell: low
  • Card X-Sell
  • Card X-Sell: high
  • Cash Street: high
  • POS household with interest
  • Cash Street: middle
  • Cash X-Sell: middle
  • Cash Street: low

• Some product combinations are nil incase of unused offers as well as cancelled loans

  • POS industry without interest
  • POS others without interest

Count of Product Combinations W.R.T Contract Status

Distribution of Loan Purpose w.r.t Contract Status

Inference 2.6

- Most rejection of loans came from purpose 'repairs'.

- For education purposes we have equal number of approves

- Rejection for Paying other loans and buying a new car are having significant higher rejection than approves.
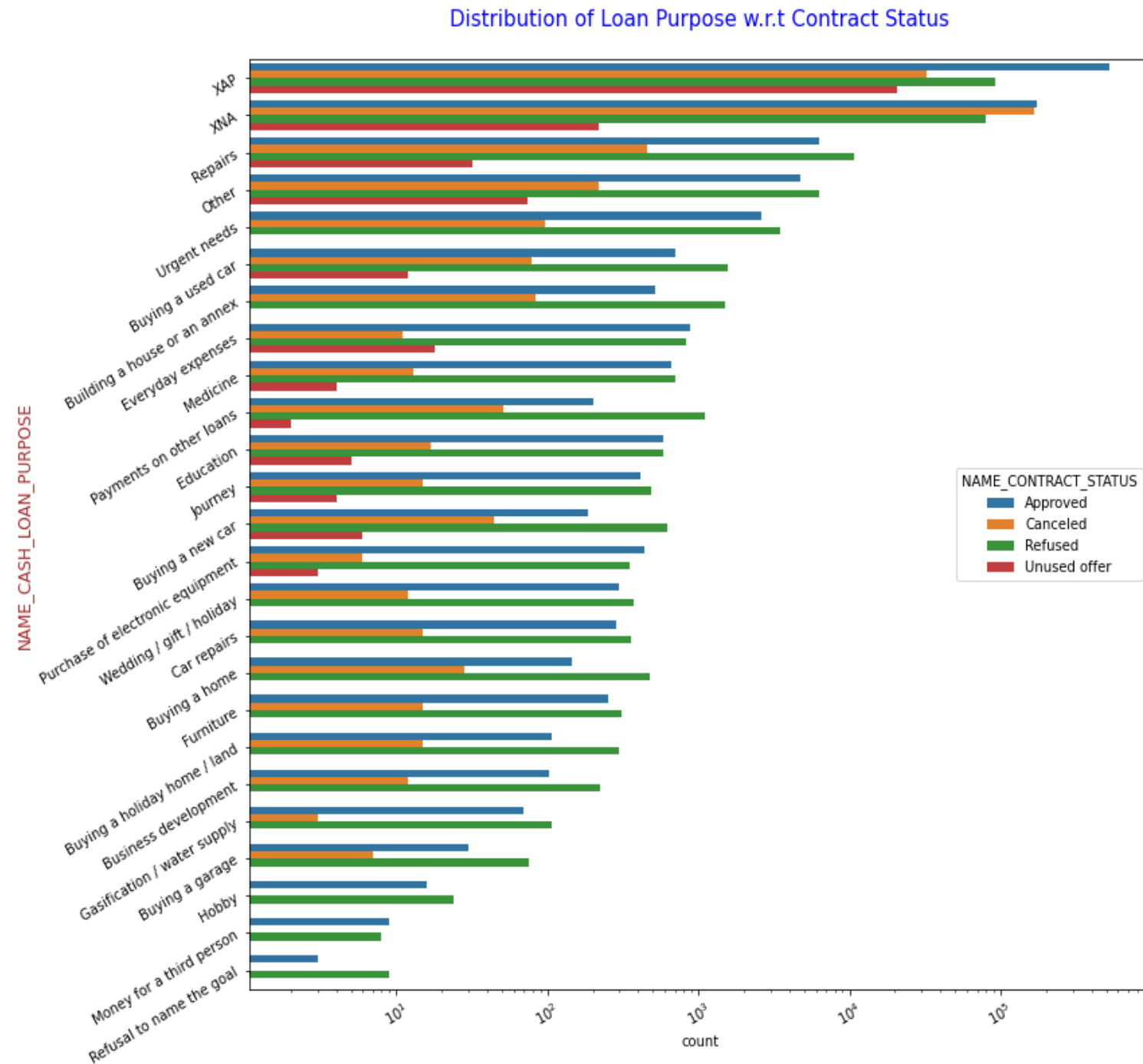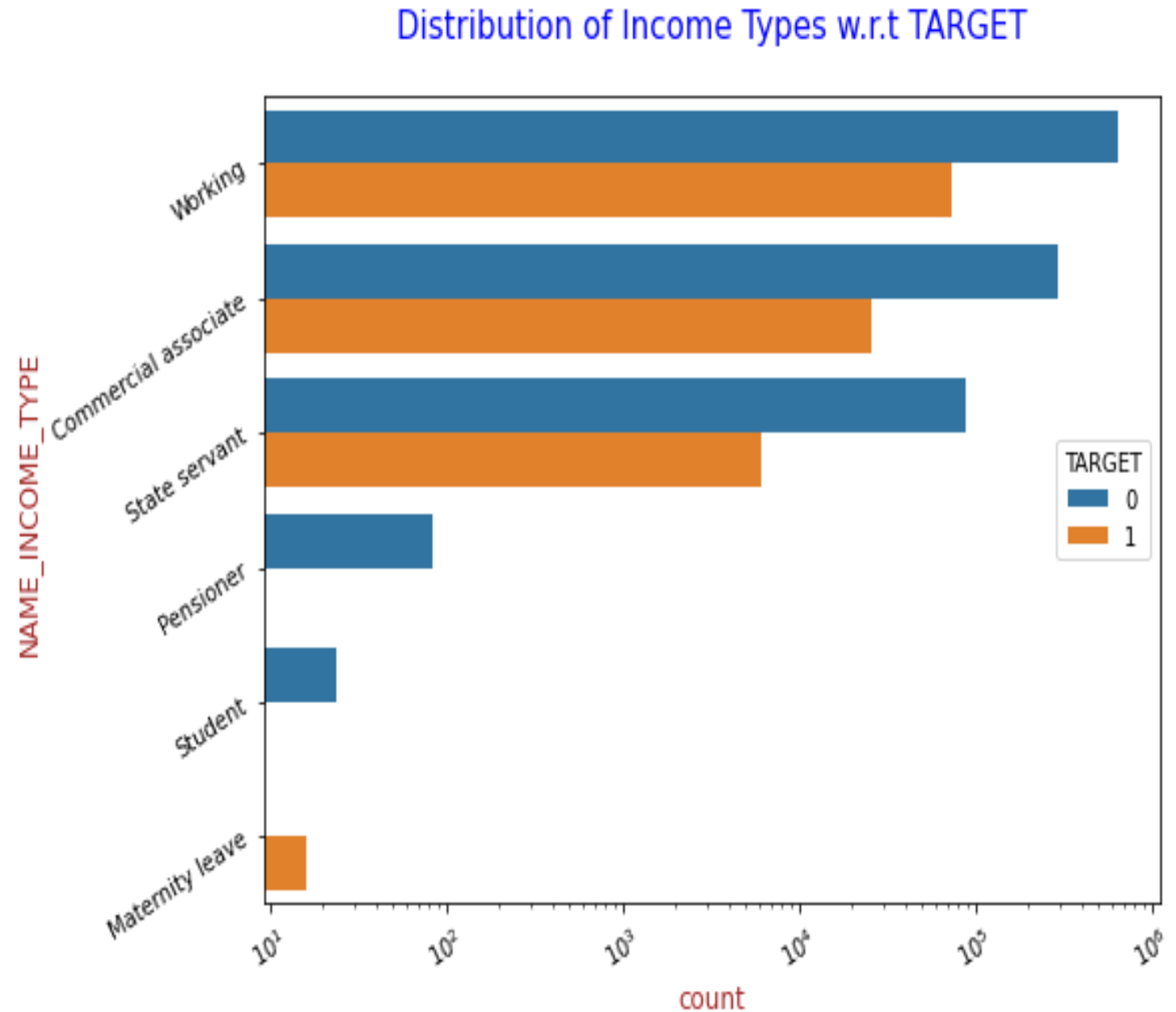
Inference 2.7

- There are no unused offers for students and pensioners

- The number of approved loans for state servants is almost equal to the refusal or canceled loans for Commercial associates

- Maximum unused offers is by working clients

Distribution of Income Types w.r.t Contract Status

Distribution of Income Types w.r.t TARGET
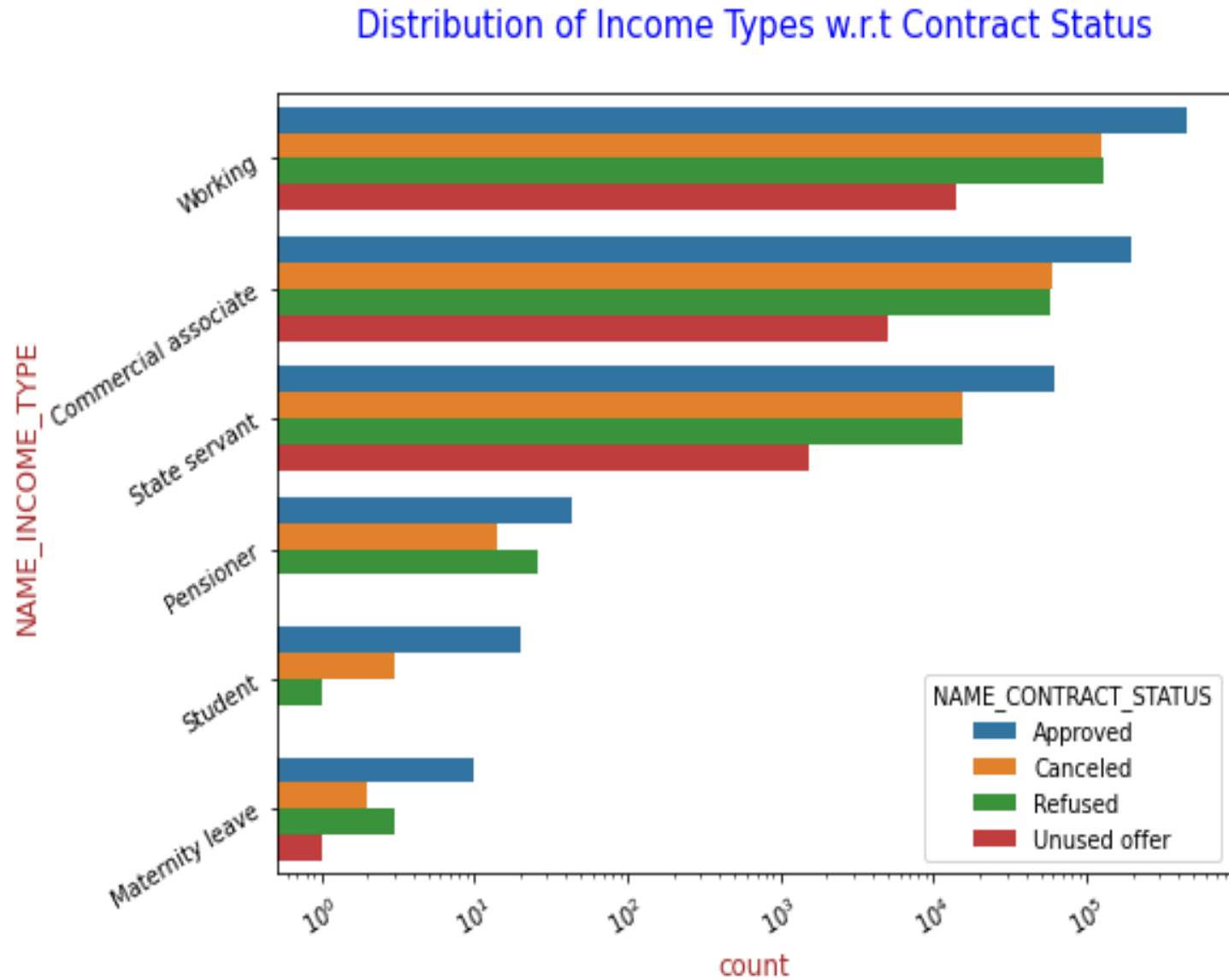
Inference 2.8

- There are no clients with payment difficulties who are students and pensioners

- There are no clients on maternity leave with no payment difficulties

Inference 2.9

- Loan purposes with 'Repairs' are facing more difficulties in payment.

- There are few places where loan payment is significantly higher than facing difficulties. They are 'Buying a garage', 'Business development', 'Buying land', 'Buying a new car' and 'Education'.

- Hence we can focus on these purposes for which the client is having minimal payment difficulties.
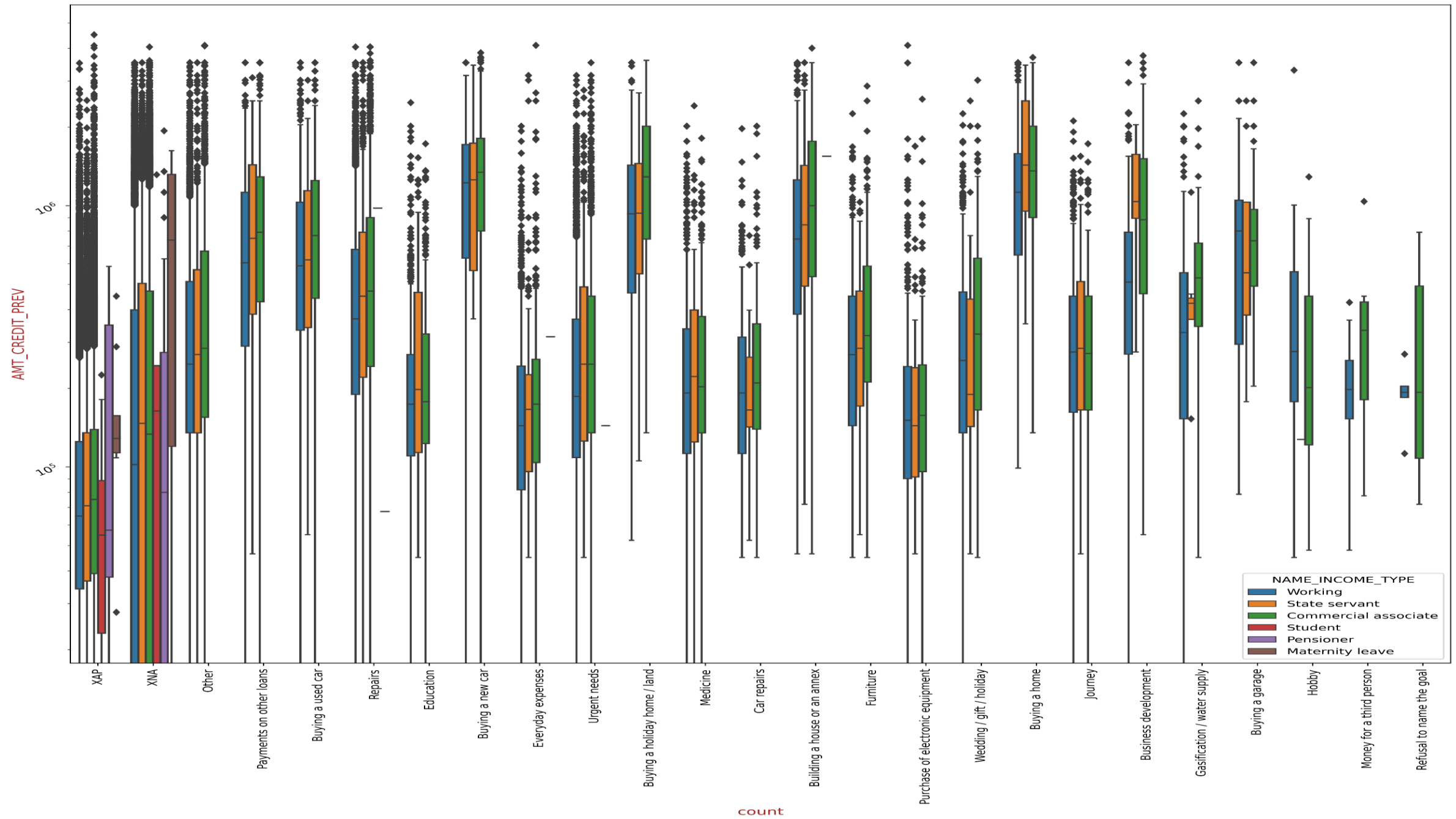

Distribution of Loan Purpose w.r.t TARGET
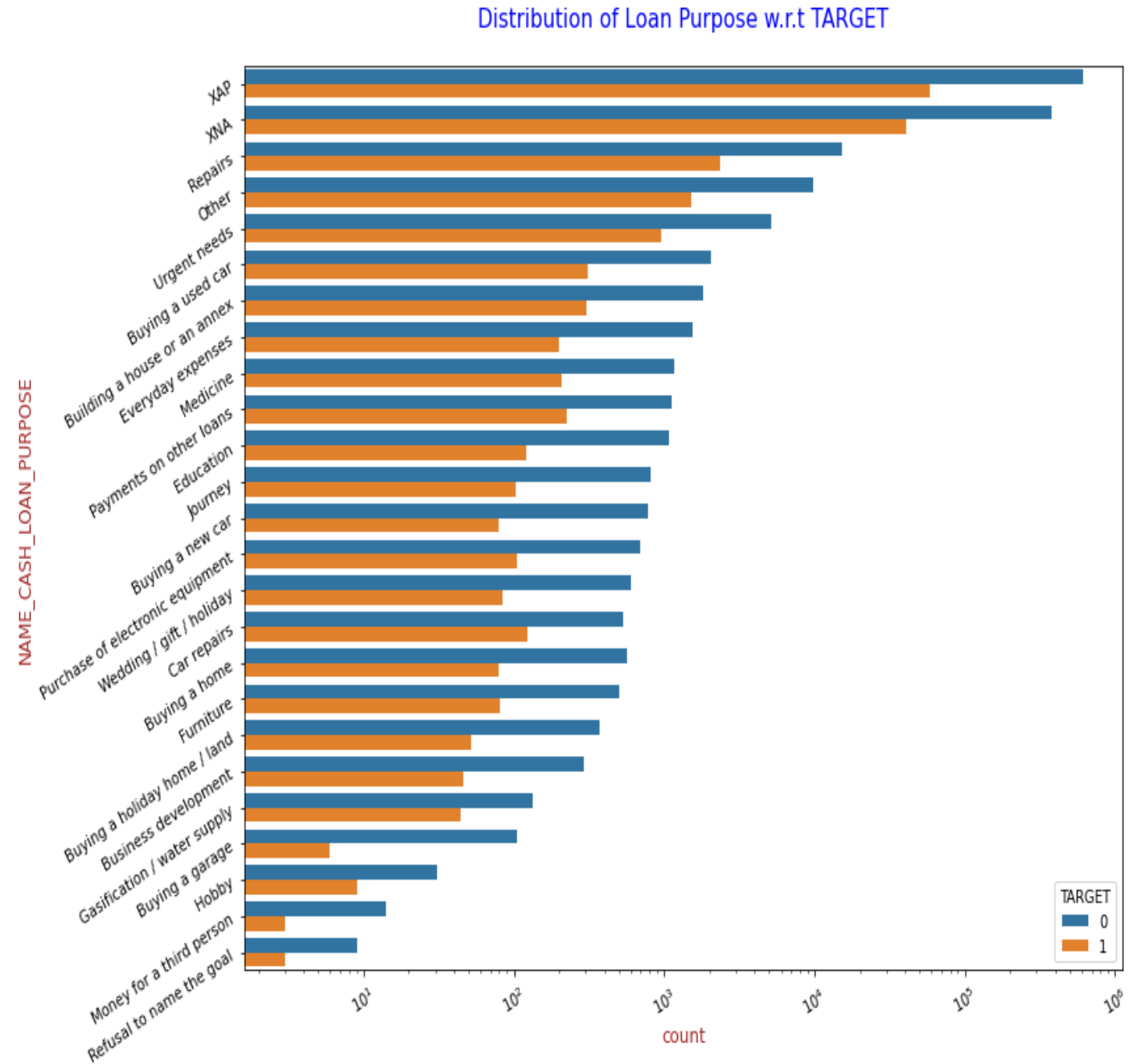
Inference 2.10 is for the plot in next slide

•The credit amount of Loan purposes like 'Buying a home', 'Buying a land', 'Buying a new car' and 'Building a house' is higher being a 'commercial associate' and even in general irrespective of income type.

•Income type of state servants having no amount of credit applied for purpose of Money for third person or a Hobby.

•Most of the commercial associates have refused to provide purpose of credit

Distribution of Loan Purpose w.r.t TARGET

Inference 2.11

• Here for Housing type, office apartment and co-op apartment are having higher credit of target 1.

• So, we can conclude that bank should avoid giving loans to the housing type of office apartment and co-op apartment as they are having difficulties in payment.

• Bank can focus mostly on housing type with parents or House or apartment or municipal apartment for successful payments.

Distribution of Amount Credited previously vs Housing Type w.r.t TARGET
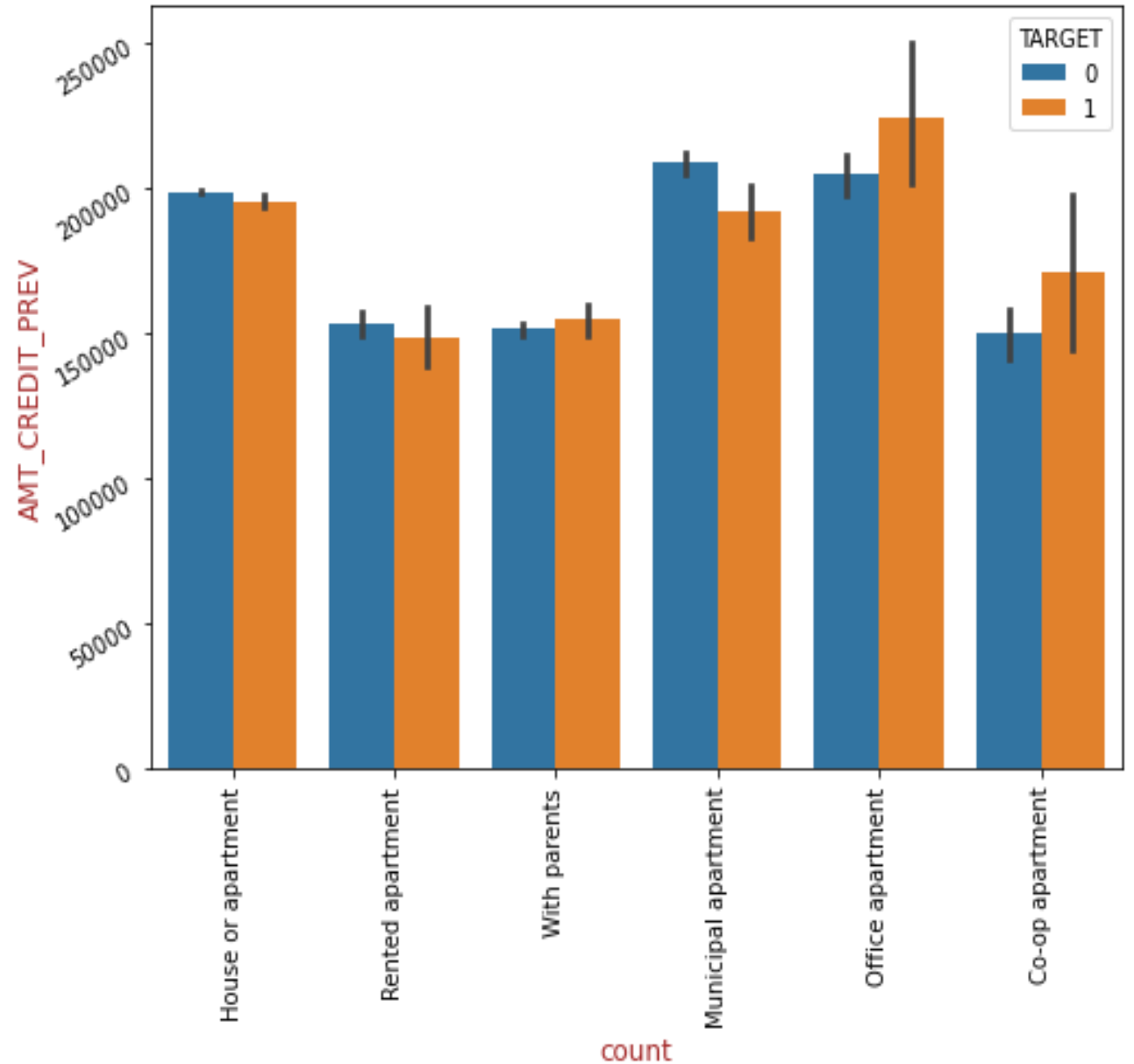
Distribution of Amount Credited previously vs Education Type w.r.t Contract Status

Inference 2.12

•It can be seen that number of refusals are for clients has nothing to do with their education levels

•Almost all education level clients have equal unused offers.
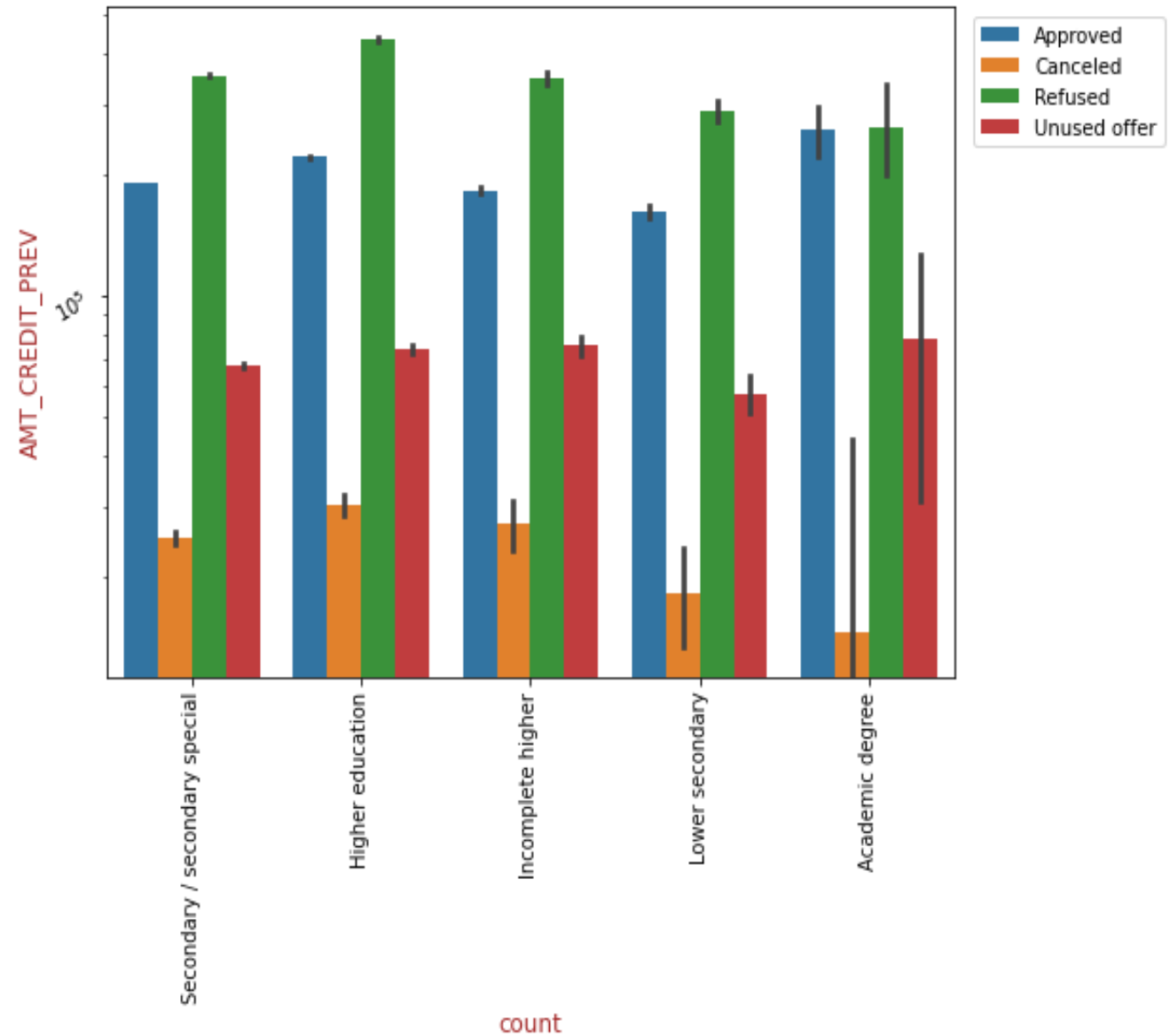
Inference of Multivariate Analysis of merged data frame

We can observe correlations of different variables plotted in the above graph and draw almost similar inference as we drew in Inference 1.15, with few extra variables

CORRELATION TABLE FOR mergedloandf1
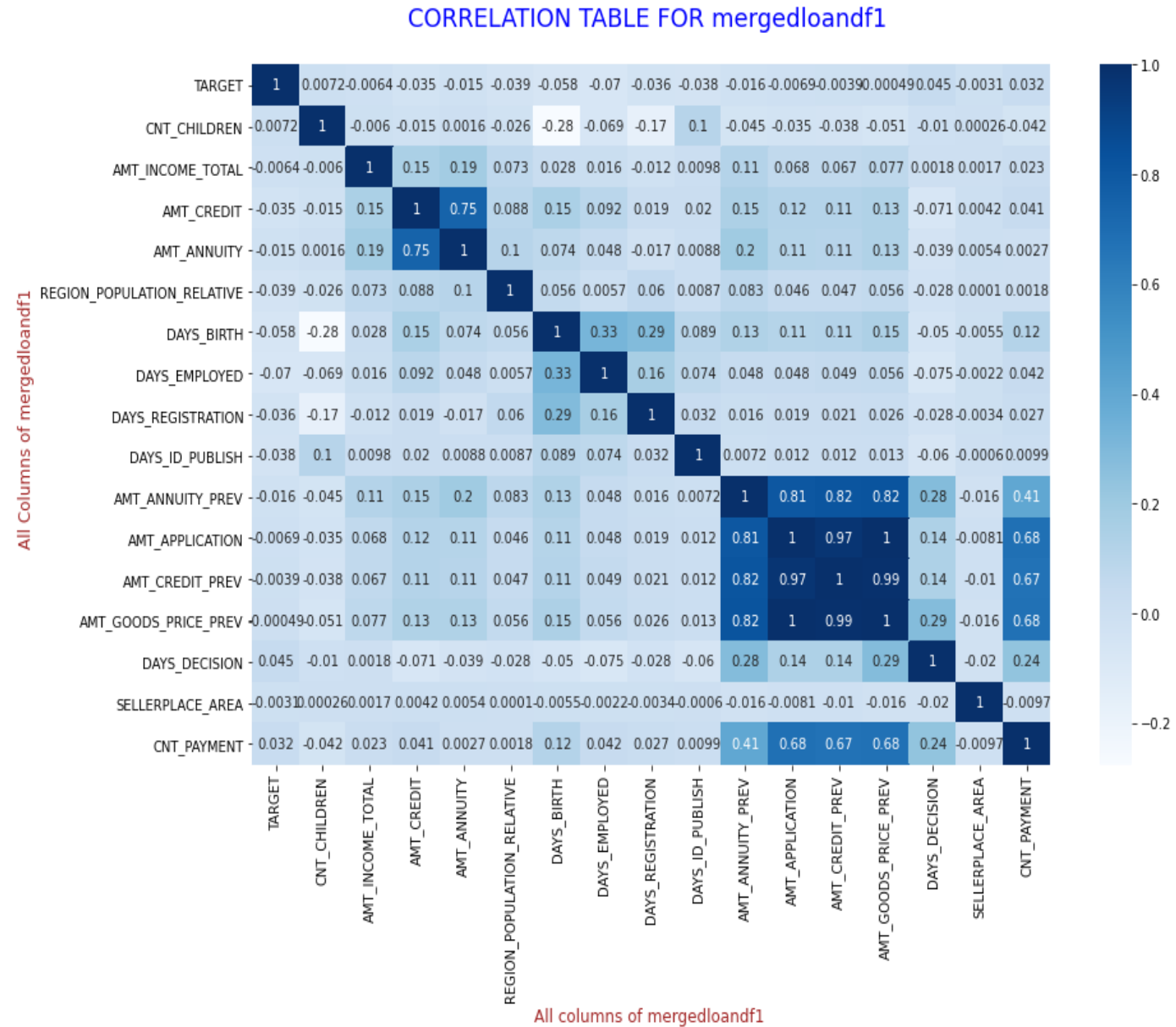
Final Inference

1.Banks should focus more on contract type 'Student' ,'pensioner' and 'Businessman' with housing 'type other than 'Co-op apartment' and 'office apartment' for successful payments.

2.Banks should focus less on income types maternity leave and working as they have most number of unsuccessful payments

3.In loan purpose 'Repairs':
1.   Although having higher number of rejection in loan purposes with 'Repairs' we can observe difficulties in payment.
2.   There are few places where loan payment difficulty is significantly high.
3.   Bank should continue to be cautious while giving loan for this purpose.

4.Bank can focus mostly on housing type with parents , House or apartment and municipal apartment with purpose of education, buying land, buying a garage, purchase of electronic equipment and some other purposes with target0 significantly more than target1 for successful payments.

5.Banks can offer more offers to clients who are students and pensioners as they take all offers and are more likely to pay back