

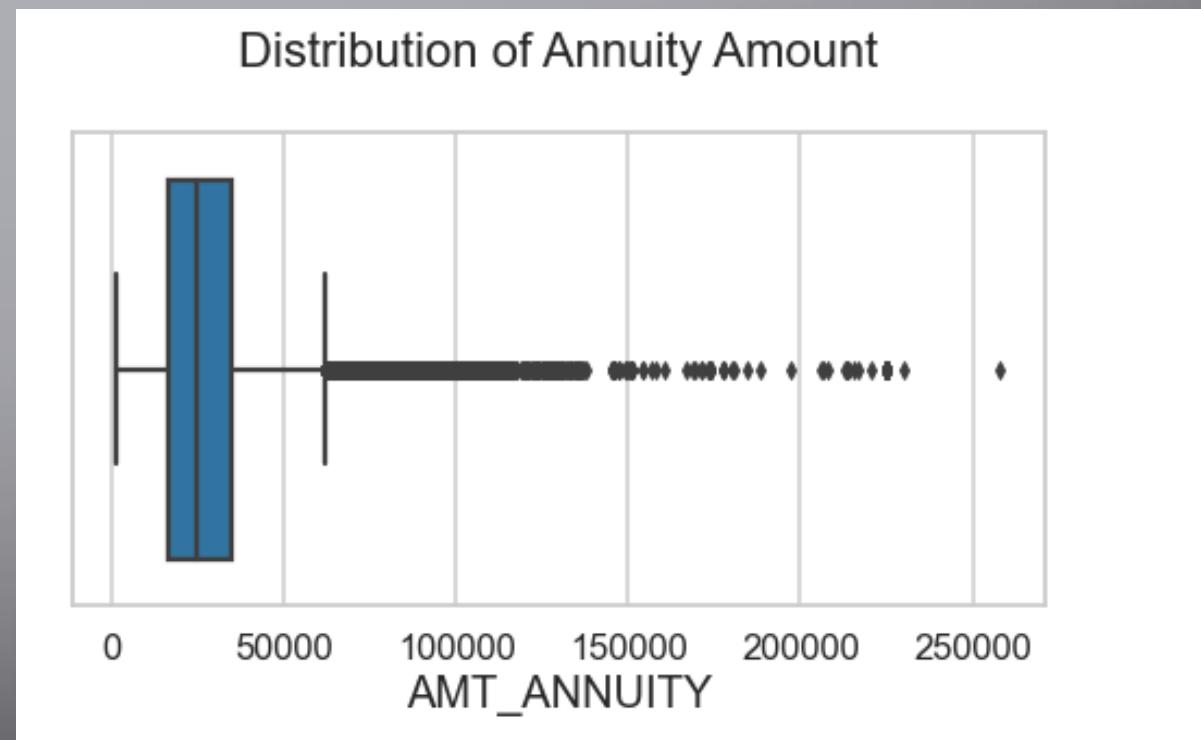
CREDIT EDA (BANK LOAN CASE STUDY)

Submitted By : Ashutosh
Lande

Boxplot for Annuity amount

Few points can be concluded from the graph.

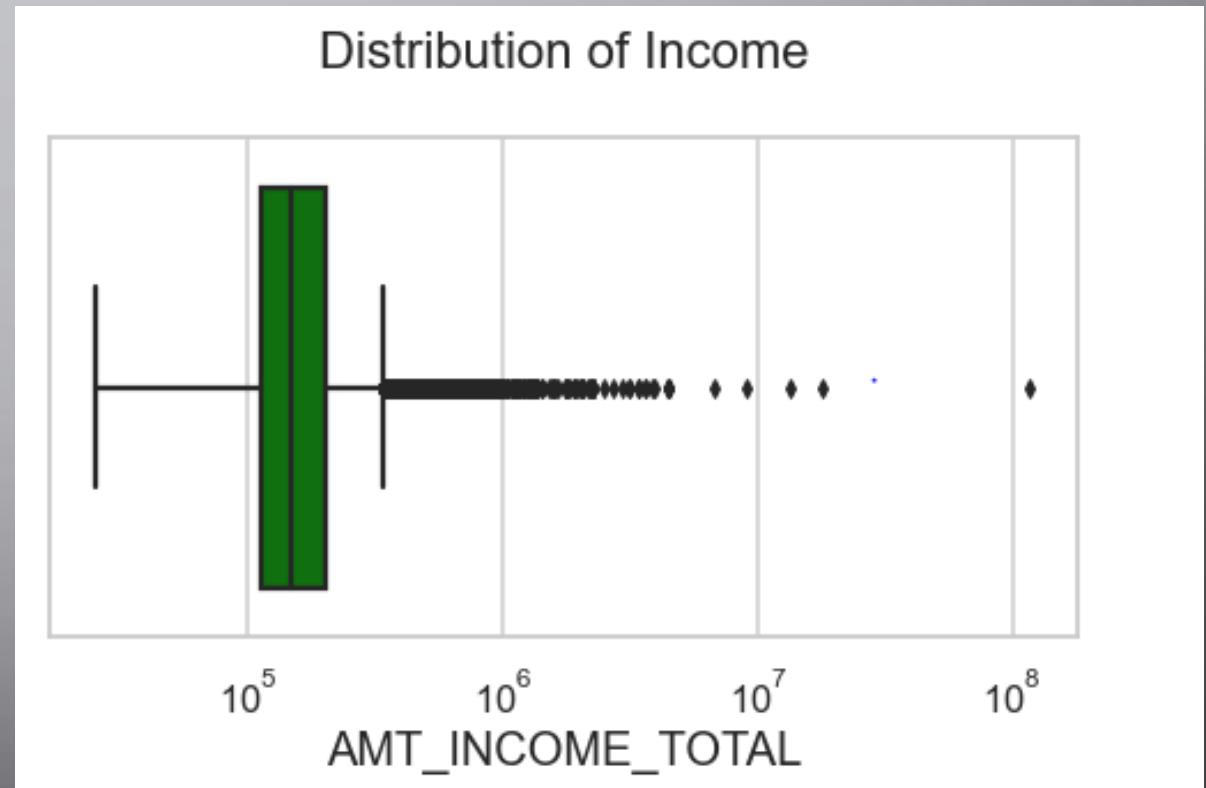
- Some outliers are noticed in income amount.
- After 250000 we can impute it with median.



Boxplot for Income amount

Few points can be concluded from the graph.

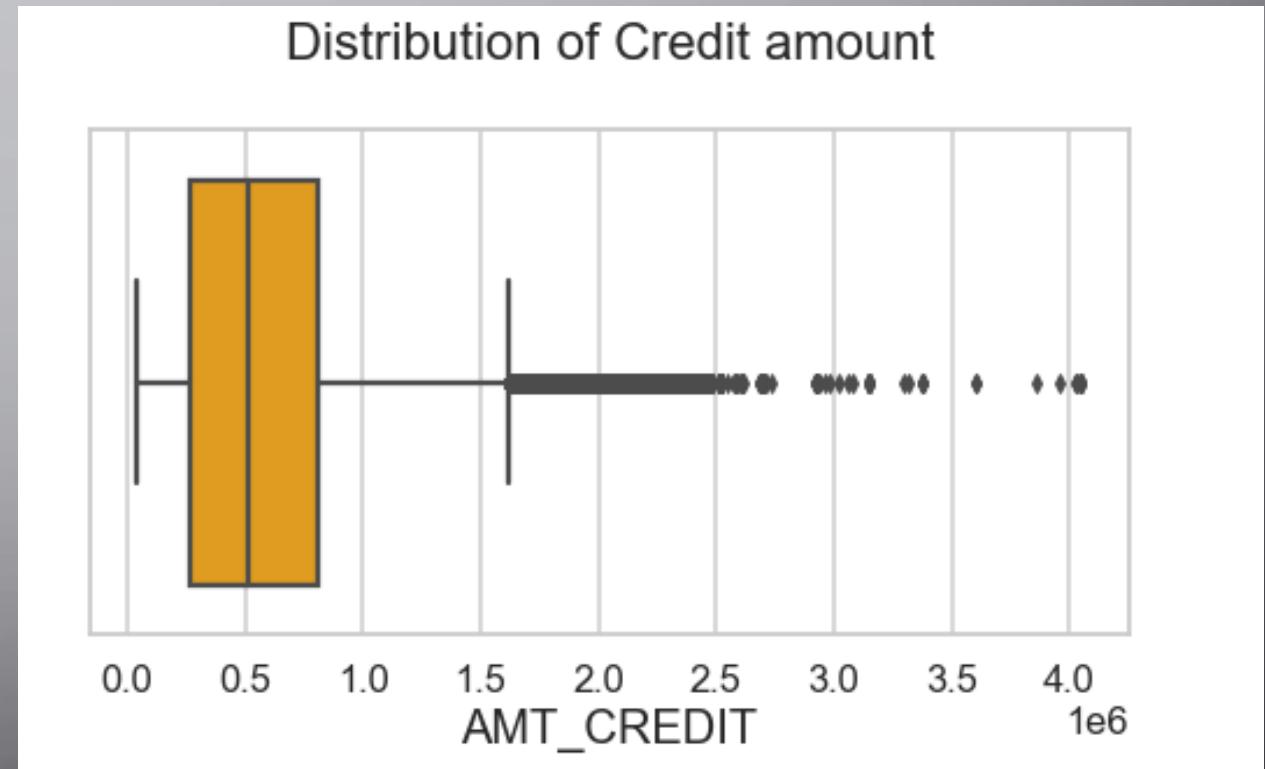
- Some outliers are noticed in credit amount.
- After $\log_{10} 8$ though most of the customer of income are present in first quartile.



Boxplot for Credit amount

Few points can be concluded from the graph.

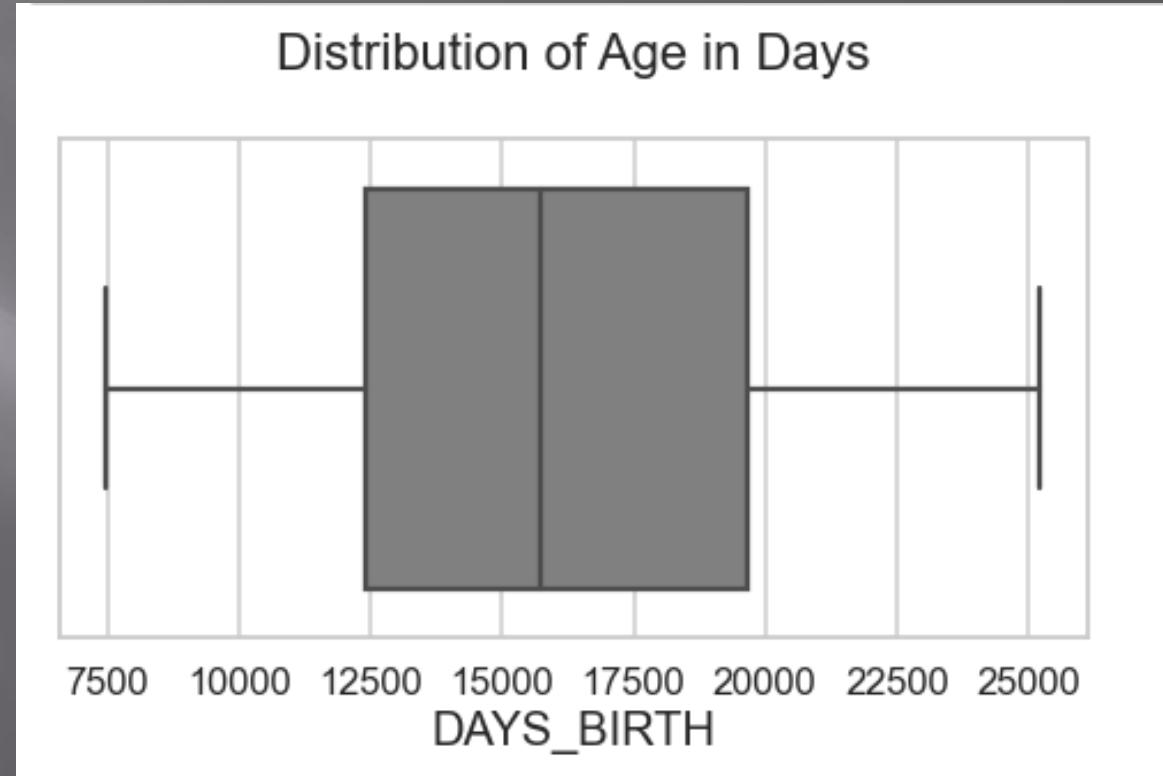
- Some outliers are noticed in annuity amount.
- Credit amt has outlier after 99th quintile at 4.05×10^6 amount credited also varies from person to person.



Boxplot for Day Birth

Few points can be concluded from the graph.

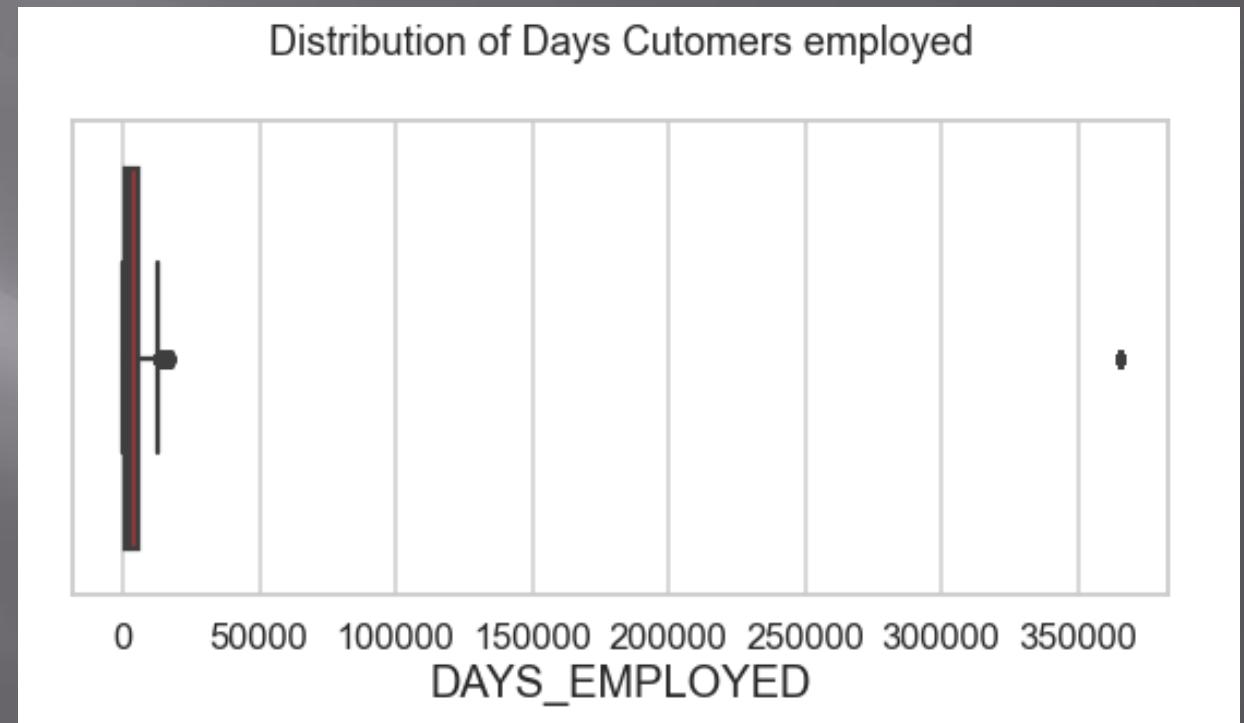
- No outliers are noticed in Day Birth.
- There is no as such difference between mean and median as applications received from the customers are of almost same age.



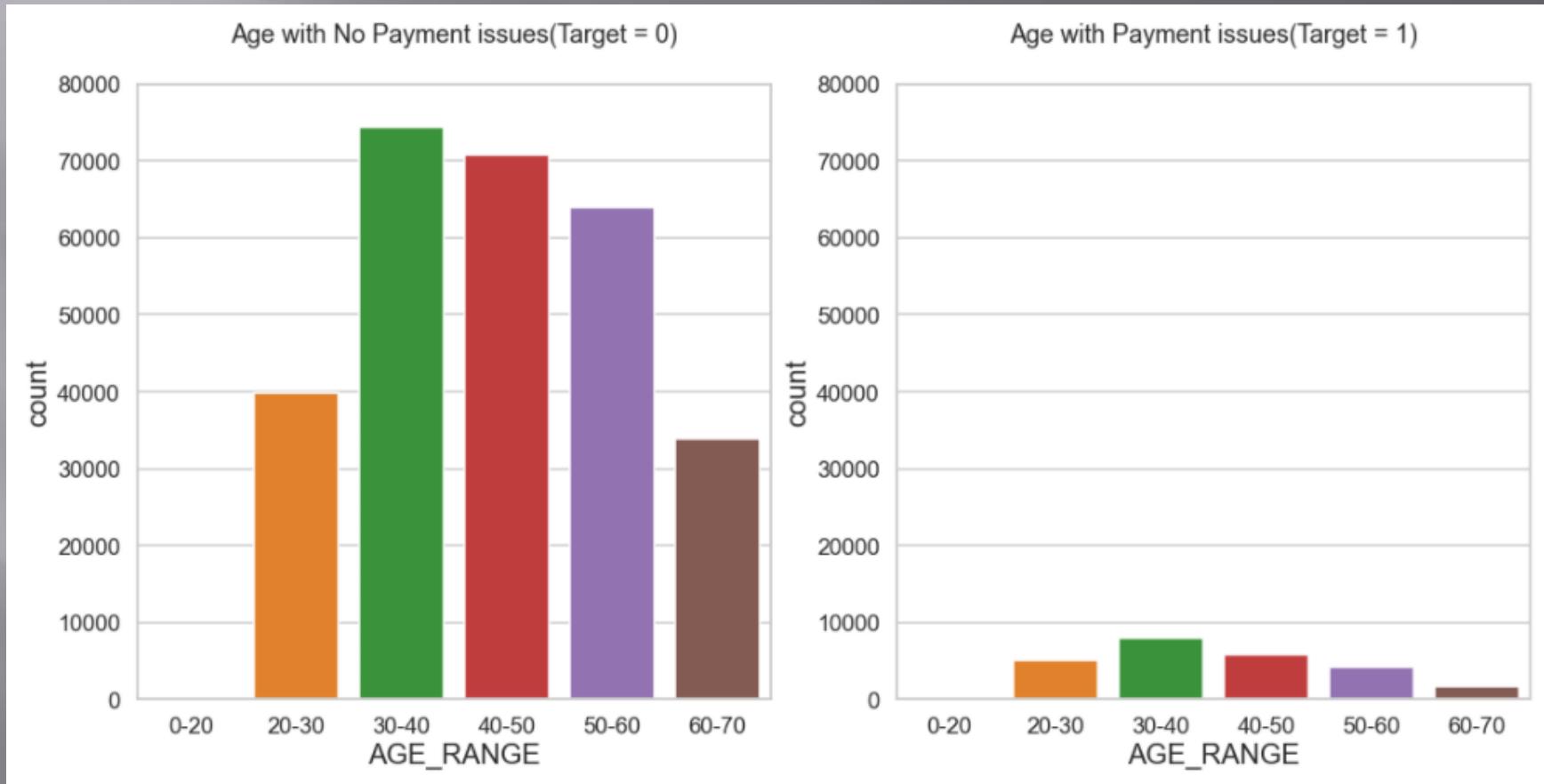
Boxplot for Day Customer Employed

Few points can be concluded from the graph.

- Some outliers are noticed in Days Employed.
- After approx 365000, Employment depends upon customer to customer
- may vary.

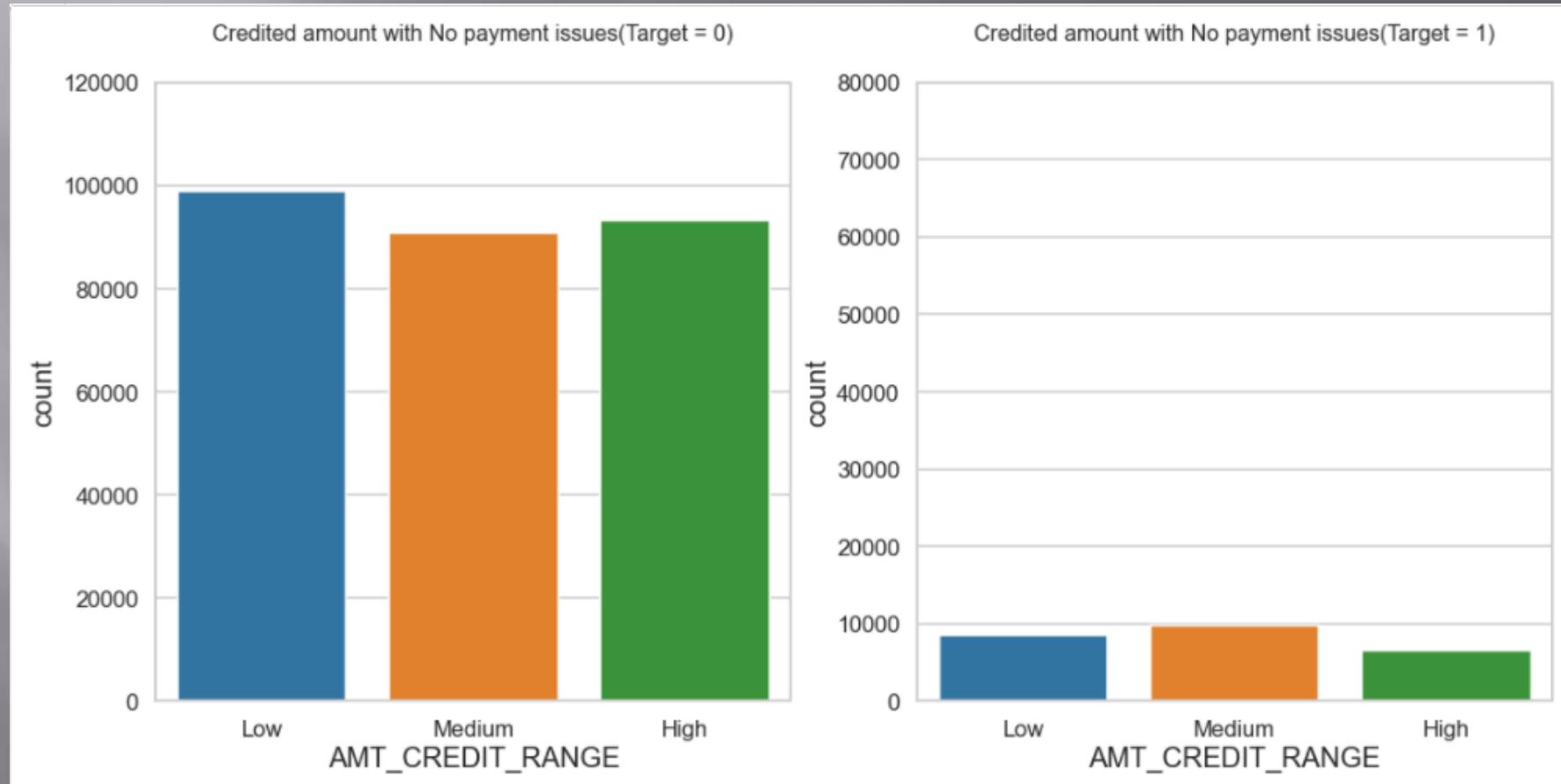


Analysis for Target=0 & Target=1 for Age



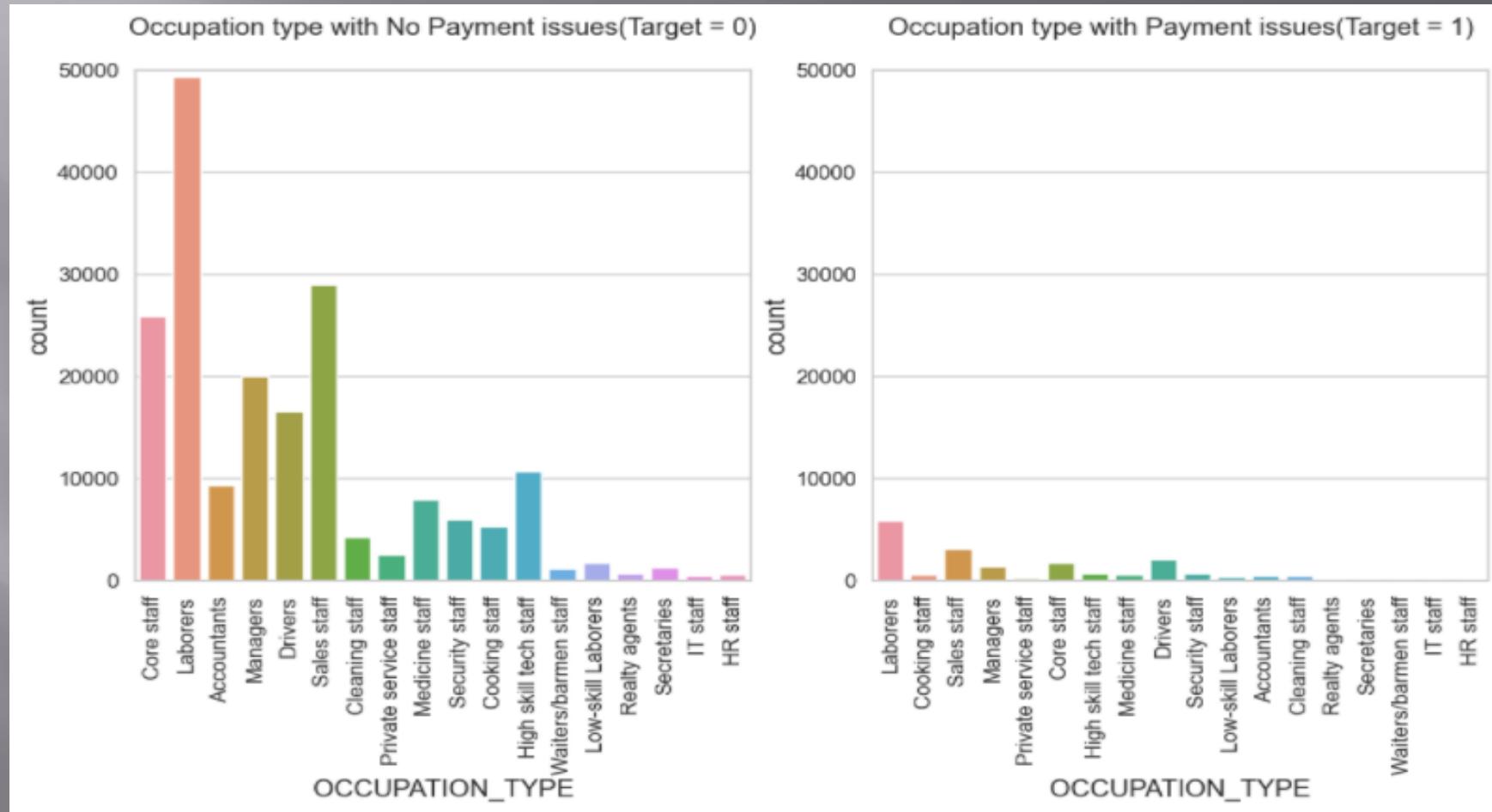
From plot we can observe 30-40 & 40-60 age group customers are giving payment on time can be consider for loan

Analysis for Target=0 & Target=1 of for Amount Credited



- From plot we can consider depending upon credit score if customer should consider for loan or not.

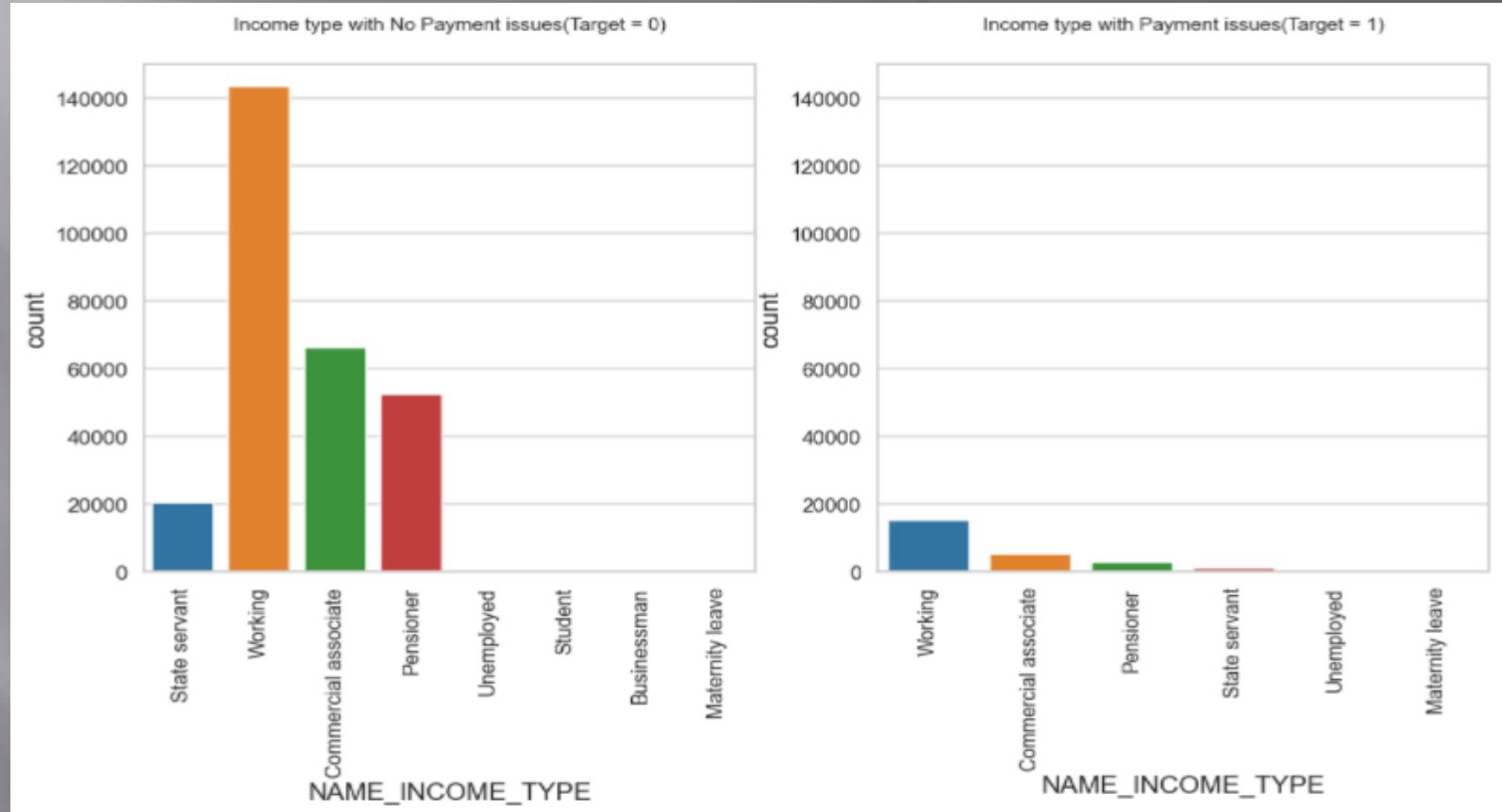
Categorical variable analysis Target=0 & Target=1 for OCCUPATION TYPE



From Plot we can clearly see that labourers are prompt with payment whereas HR Staff mostly delaying payment.

Categorical variable analysis for Target=0 & Target=1

Income Type



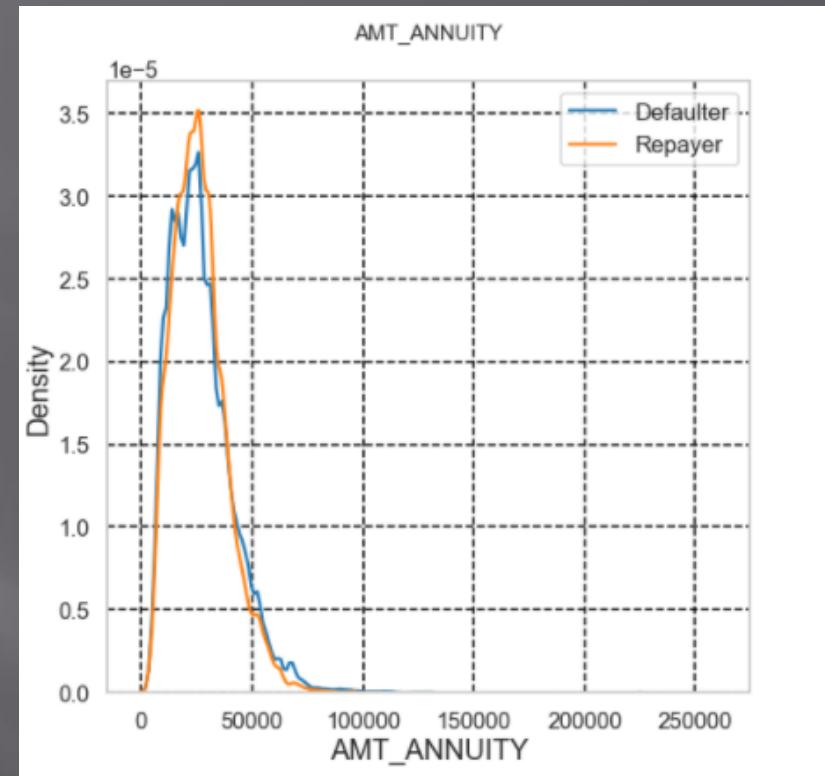
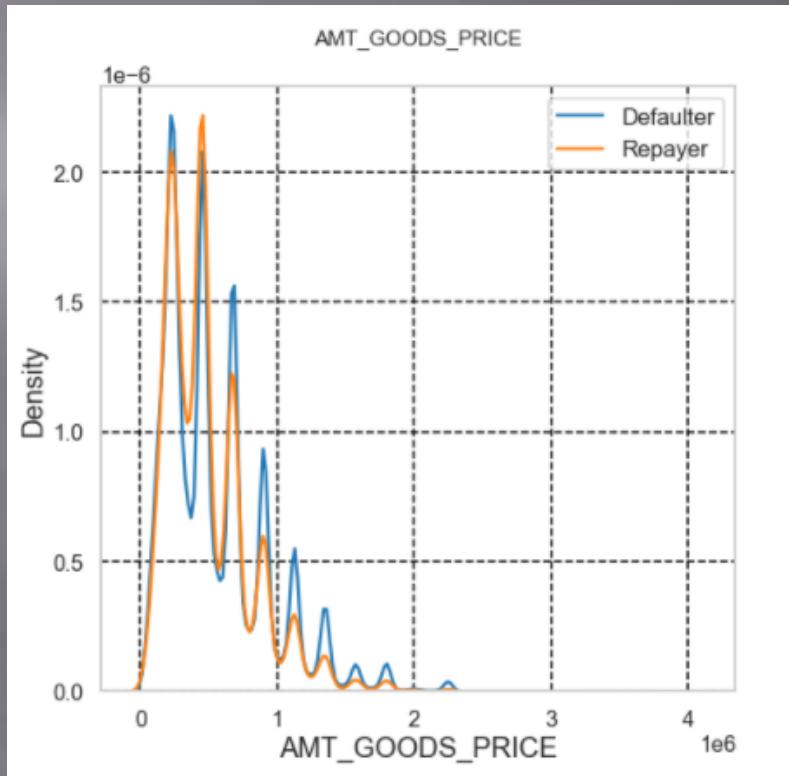
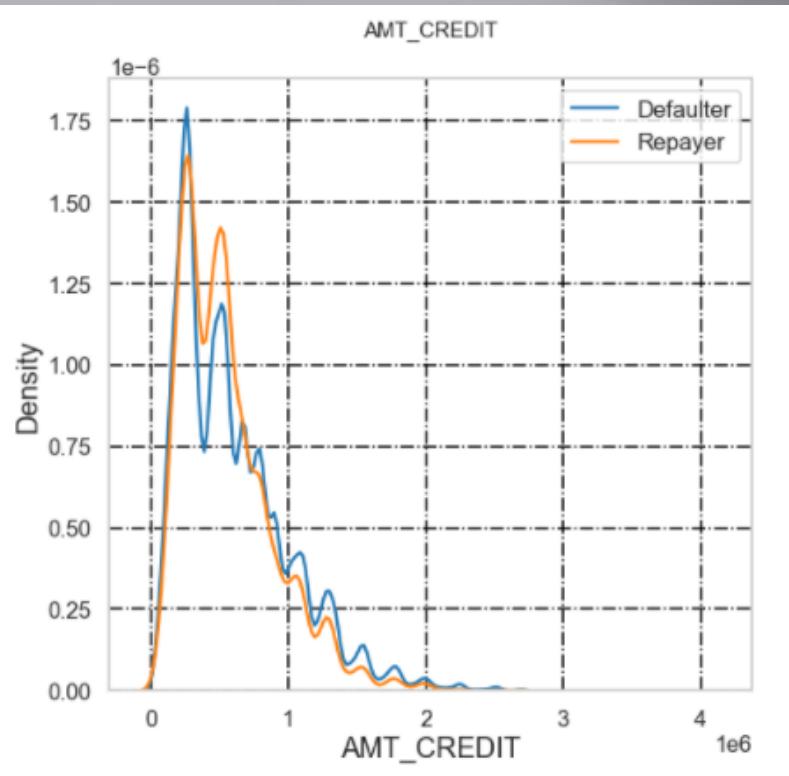
Points to be concluded from the above graph:-

For income type 'working', 'commercial associate', and 'State Servant' the number of credits are higher than others.

For this Females are having more number of credits than male.

Less number of credits for income type 'student' , 'pensioner' , 'Businessman' and 'Maternity leave'.

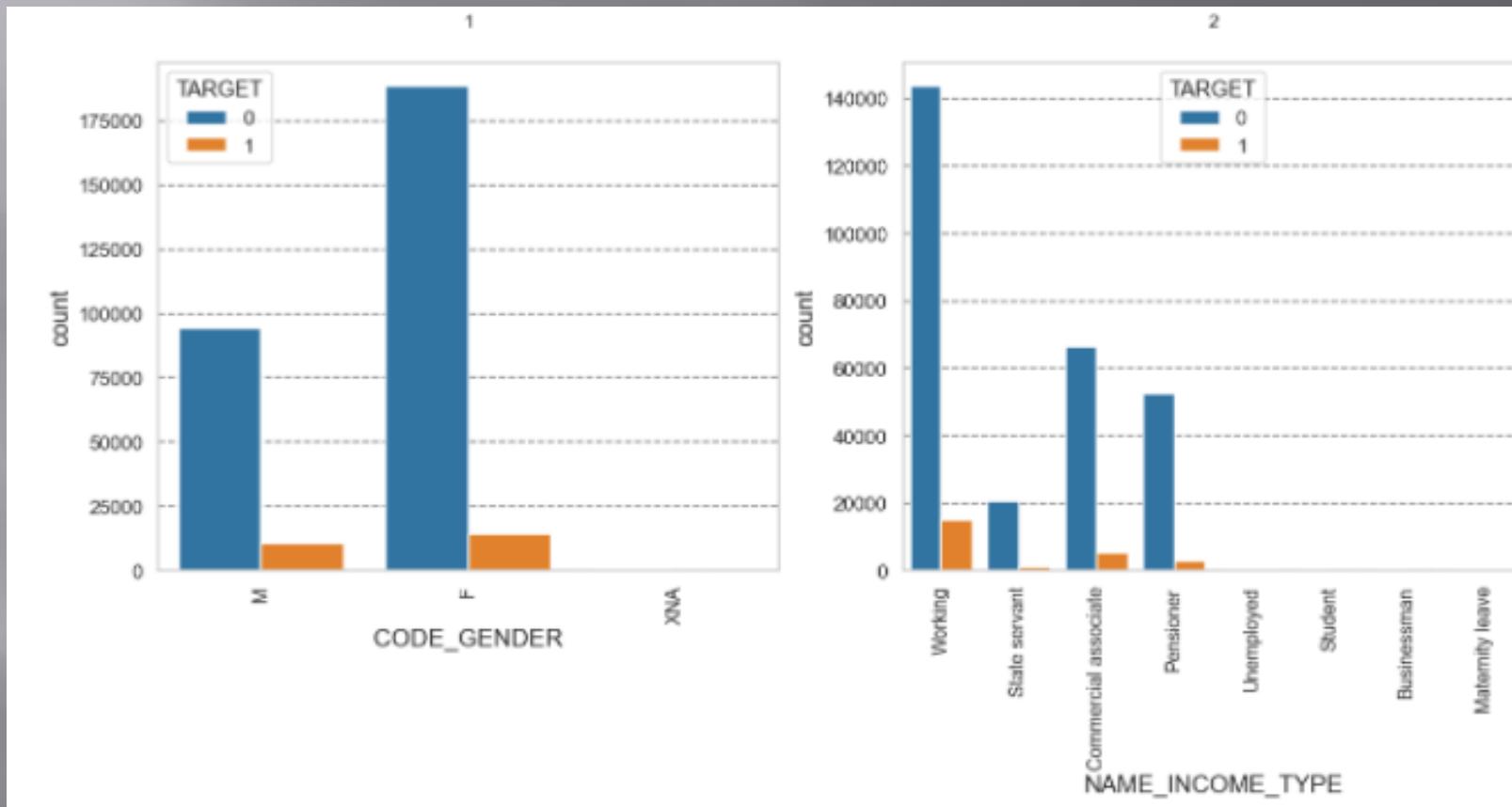
Analysis continuous column W.r.t Target column



From the plot we can clearly note following points:

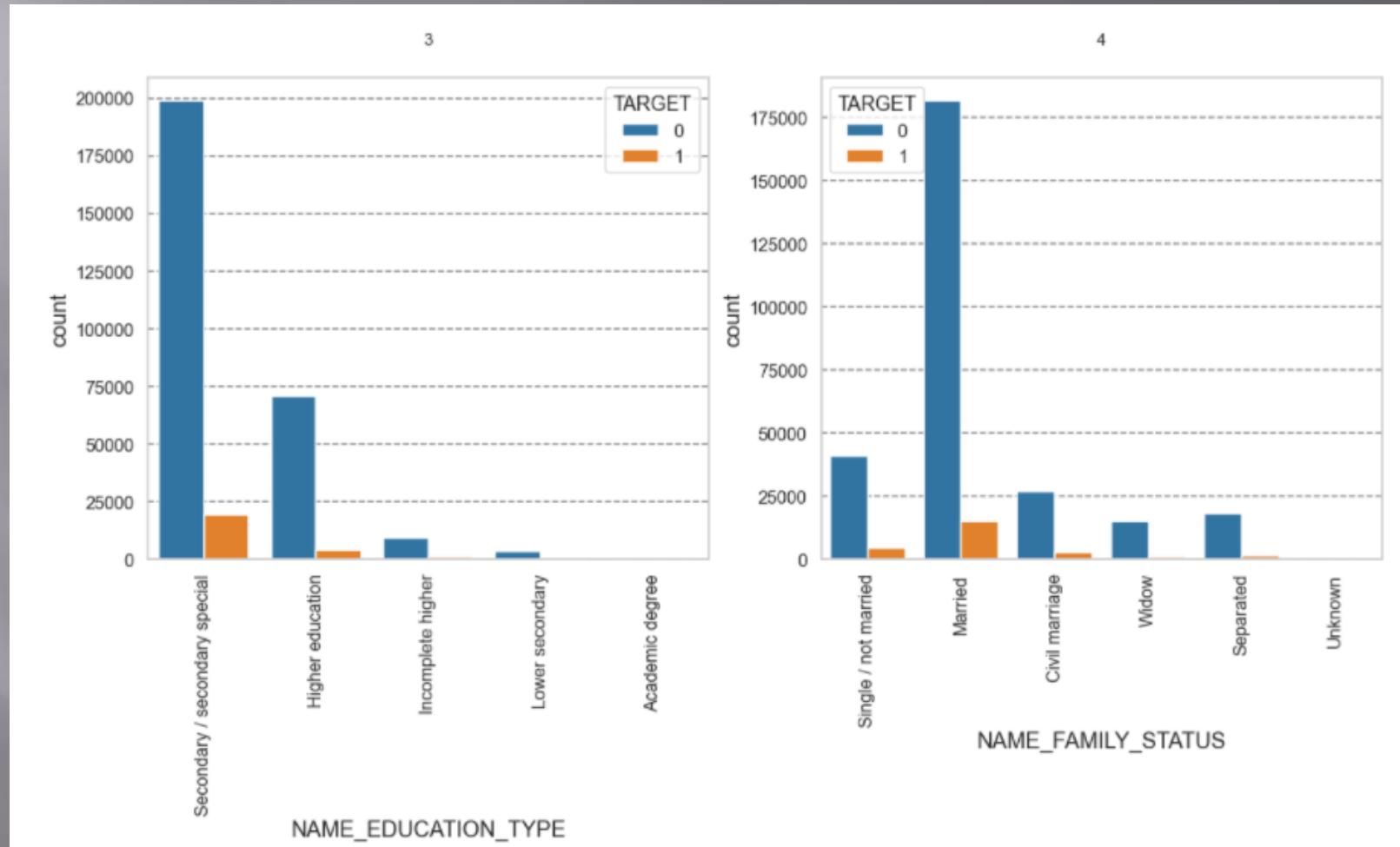
We can observe from graph that w.r.t to Amount of Credit ,Amount of good price and annuity Defaulter & Repay cases are changing Customer to Customer.

Categorical columns with respect to Target column



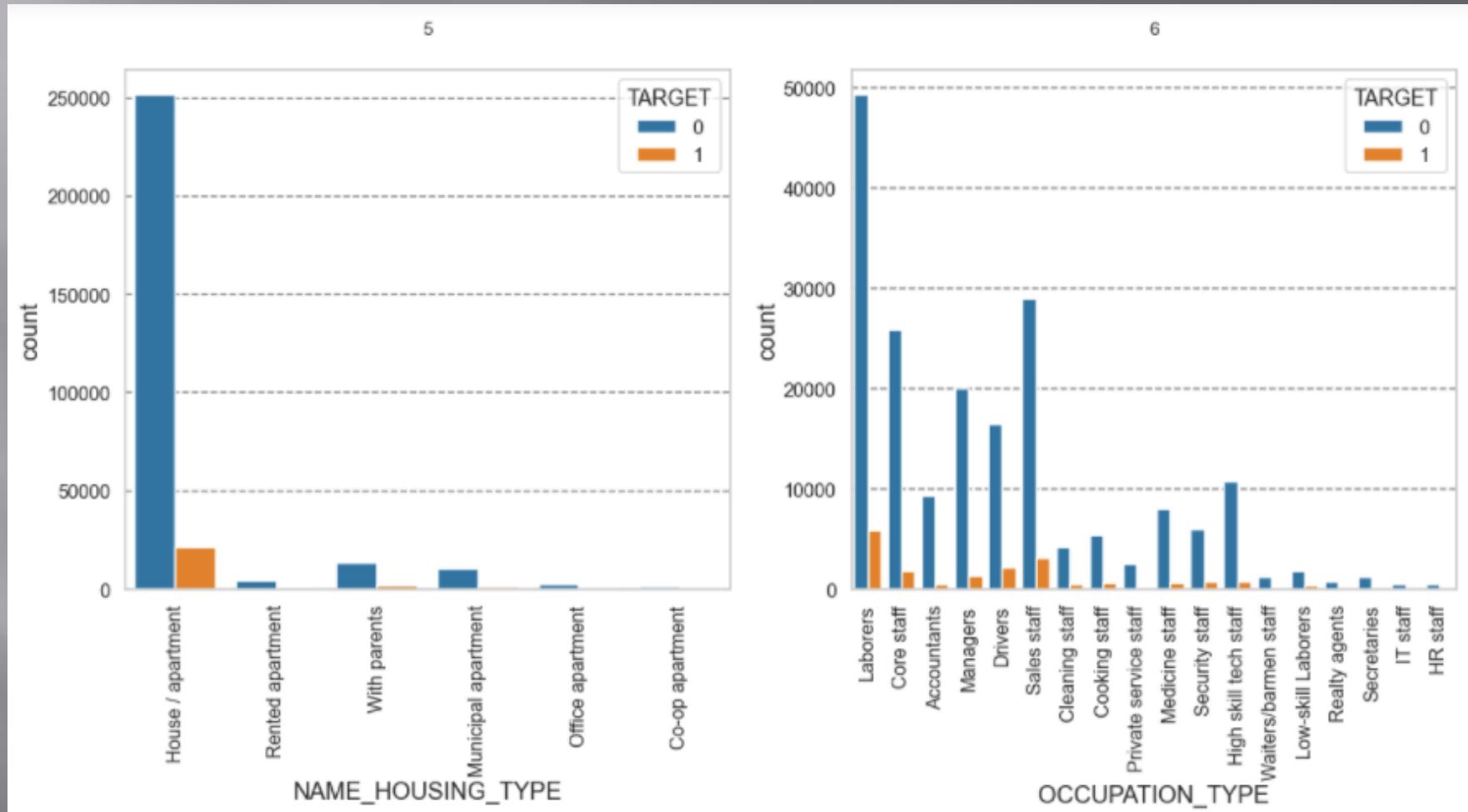
- From the plot we can clearly note following points:
 - 1) Female customers pay loan on time and banks can target more female customers for loan.
 - 2) Working/Employed customers can be consider for loans as there are higher percentage of payments on time.

Categorical columns with respect to Target column



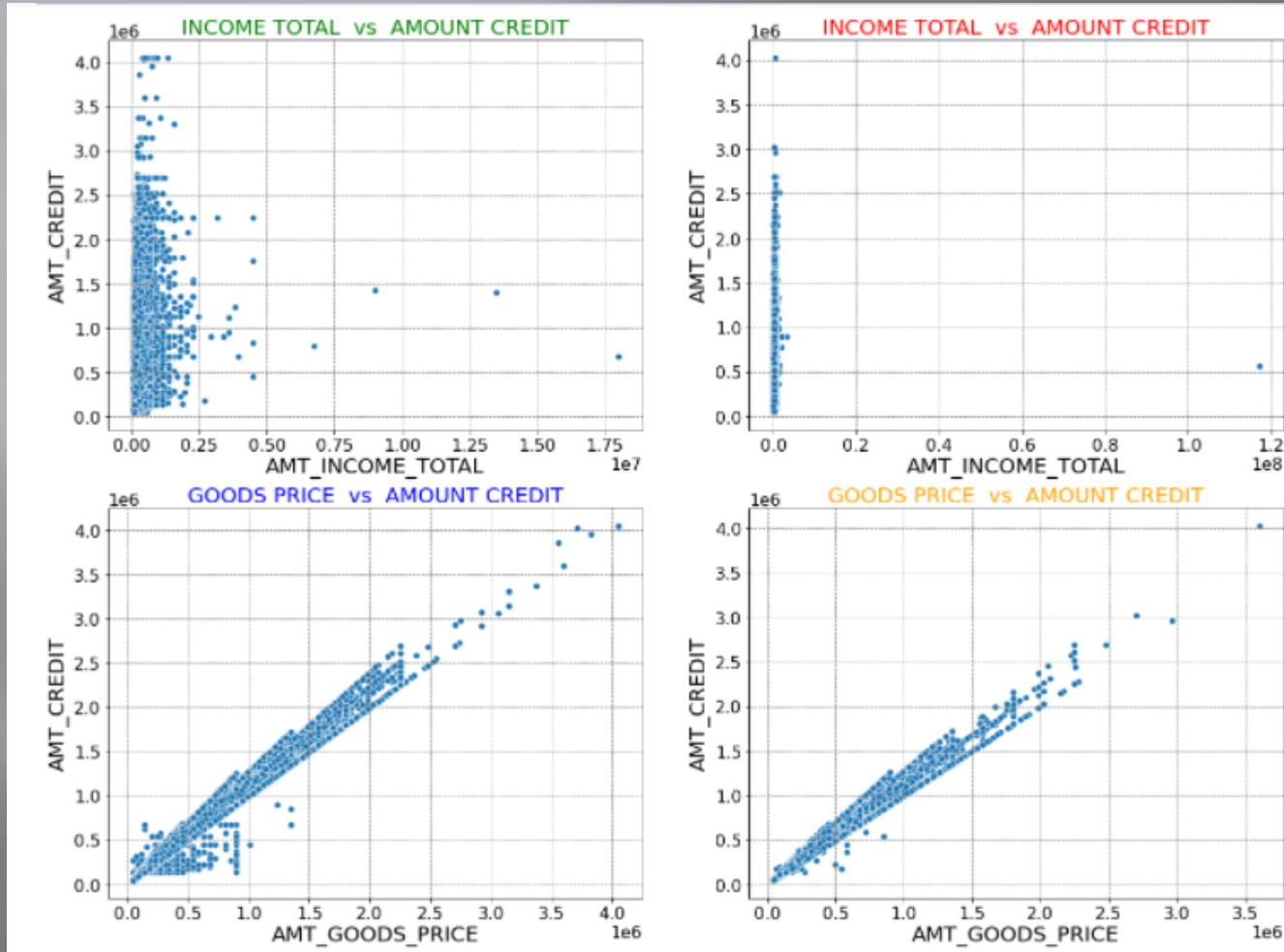
- 3) Customers having secondary education make payments on time compared to customers with academic degree.
- 4) Married customers have paid loan amount on time compared to widows.

Categorical columns with respect to Target column



- 5) Customers having House/apartment make payments on time compared to those living in other apartment.
- 6) Labourers have high payment percentage value hence can be considering Low-medium scale loan.

Plot for Numeric columns



- From plot:
- we can see that those who have paid loan within time will get good credit score as compared to other customer

Numeric variable analysis for Target=0 & Target=1



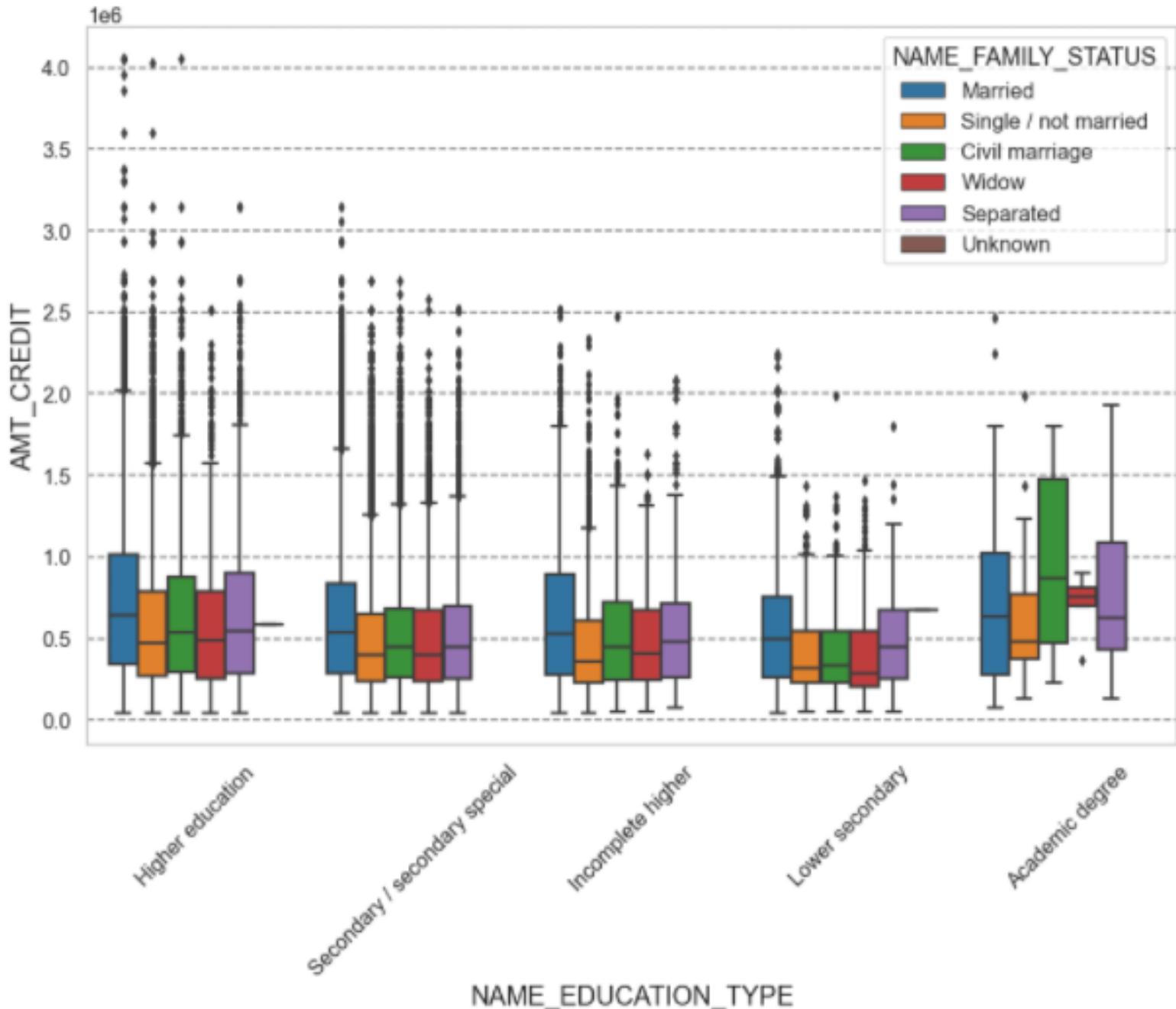
From graph we can observe even though Female with low income don't have any issues in loan payment on time.

Credit amount vs Education Status

Few points can be concluded from the graph.

- Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.
- Higher education of family status of 'marriage', 'single' 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 Education Status(Target=0)

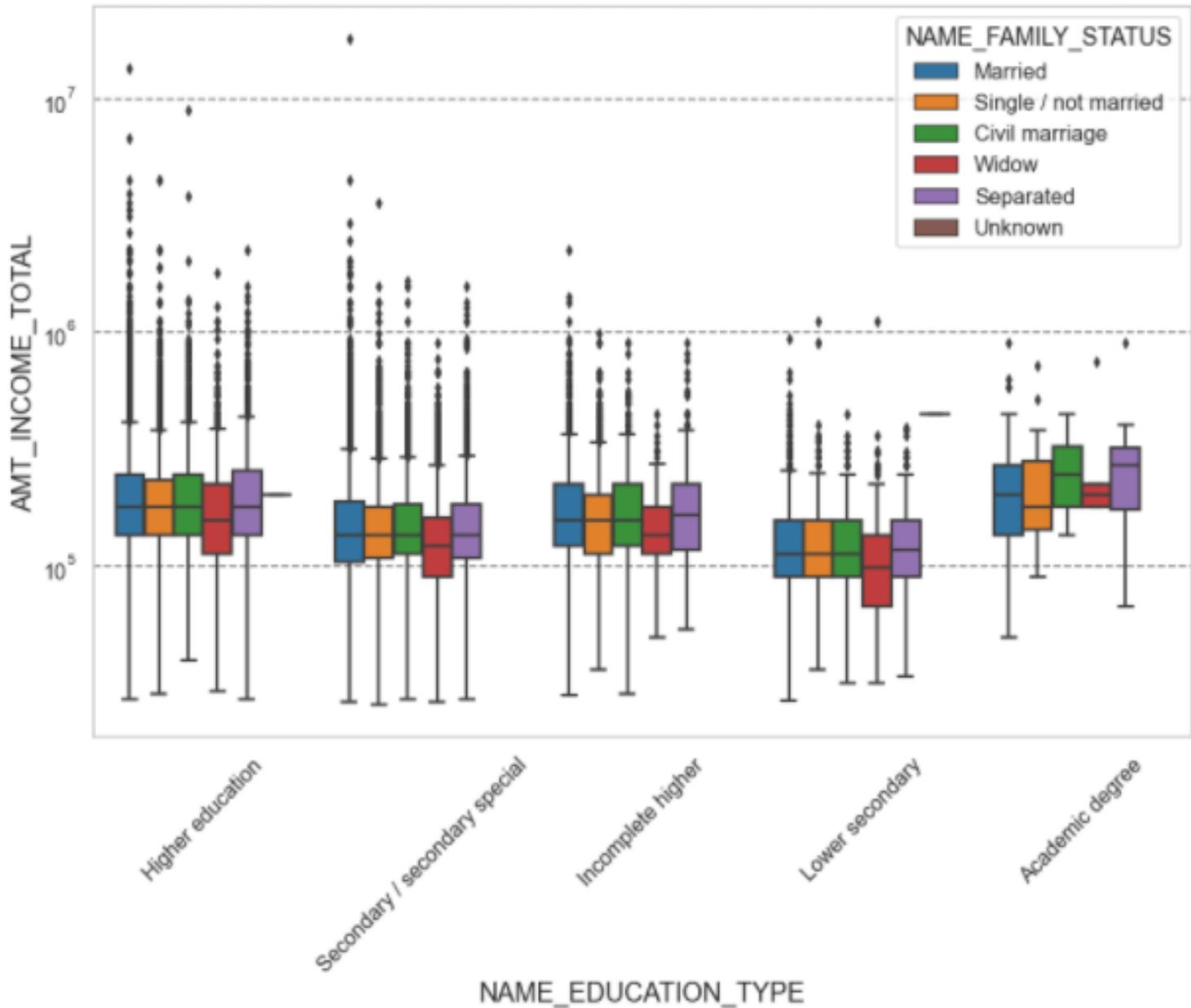


Income amount vs Education Status

Few points can be concluded from the graph.

- For Education type 'Higher education' the income amount mean is mostly equal with family status. It does contain many outliers.
- Less outlier are having for Academic degree but they are having the income amount is little higher than Higher education.
- Lower secondary of civil marriage family status are have less income amount than others.

Income amount vs Education Status(Target=0)

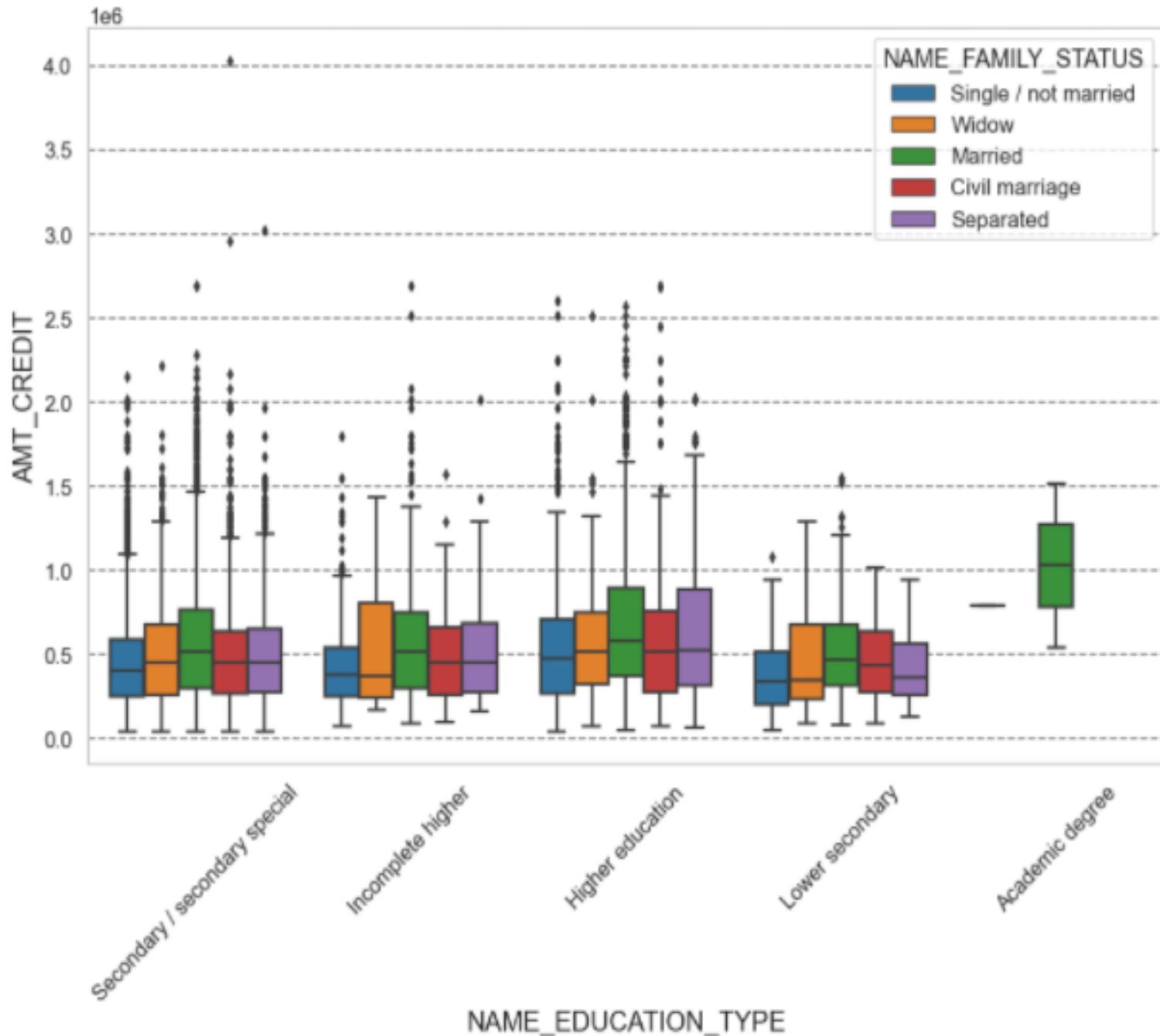


Credit amount vs Education Status

Few points can be concluded from the graph.

- Quite similar from Target 0, we can say that Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.
- Most of the outliers are from Education type 'Higher education' and 'Secondary'.
- Civil marriage for Academic degree is having most of the credits in the third quartile.

Credit amount vs Education Status(Target=1)

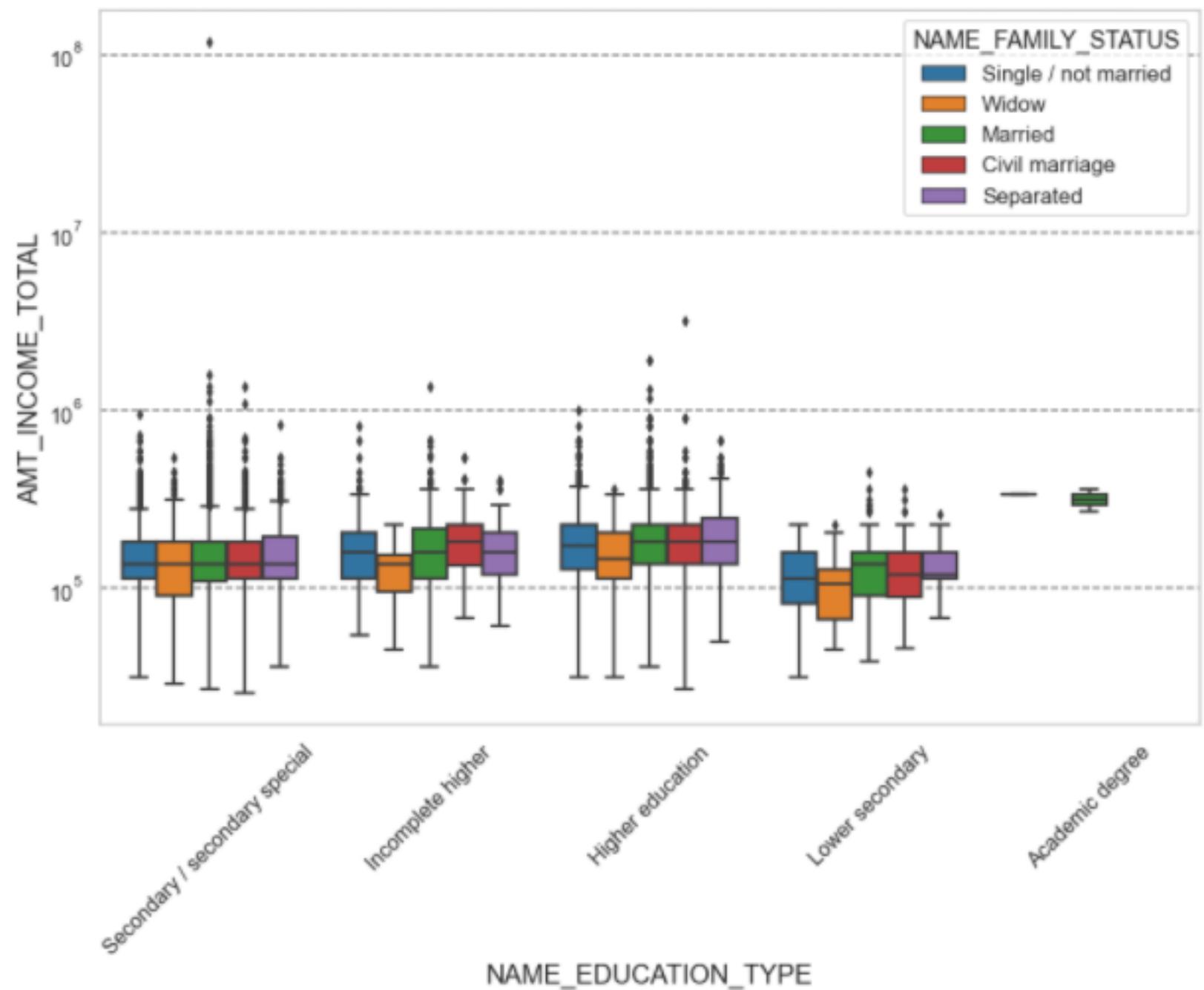


Income amount vs Education Status

Few points can be concluded from the graph.

- Have some similarity with Target0, From above boxplot for Education type 'Higher education' the income amount is mostly equal with family status.
- Less outlier are having for Academic degree but there income amount is little higher than Higher education.
- Lower secondary are have less income amount than others.

Income amount vs Education Status (Target=1)



Correlation Matrix for Target = 0&1 Res.

```
1 #Sorting the corr values Target = 0
```

```
2 corr_app_data.sort_values(by = 'Correlation', ascending = False).head(10)
```

	VAR1	VAR2	Correlation
398	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	1.00
148	AMT_GOODS_PRICE	AMT_CREDIT	0.99
423	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.86
149	AMT_GOODS_PRICE	AMT_ANNUITY	0.78
124	AMT_ANNUITY	AMT_CREDIT	0.77
224	DAYS_EMPLOYED	DAYS_BIRTH	0.63
123	AMT_ANNUITY	AMT_INCOME_TOTAL	0.42
147	AMT_GOODS_PRICE	AMT_INCOME_TOTAL	0.35
194	DAYS_BIRTH	CNT_CHILDREN	0.34
99	AMT_CREDIT	AMT_INCOME_TOTAL	0.34

```
16 #Doing same modification to find for correlation Target=1
```

```
17 corr_app_data.sort_values(by = 'Correlation', ascending = False).head(10)
```

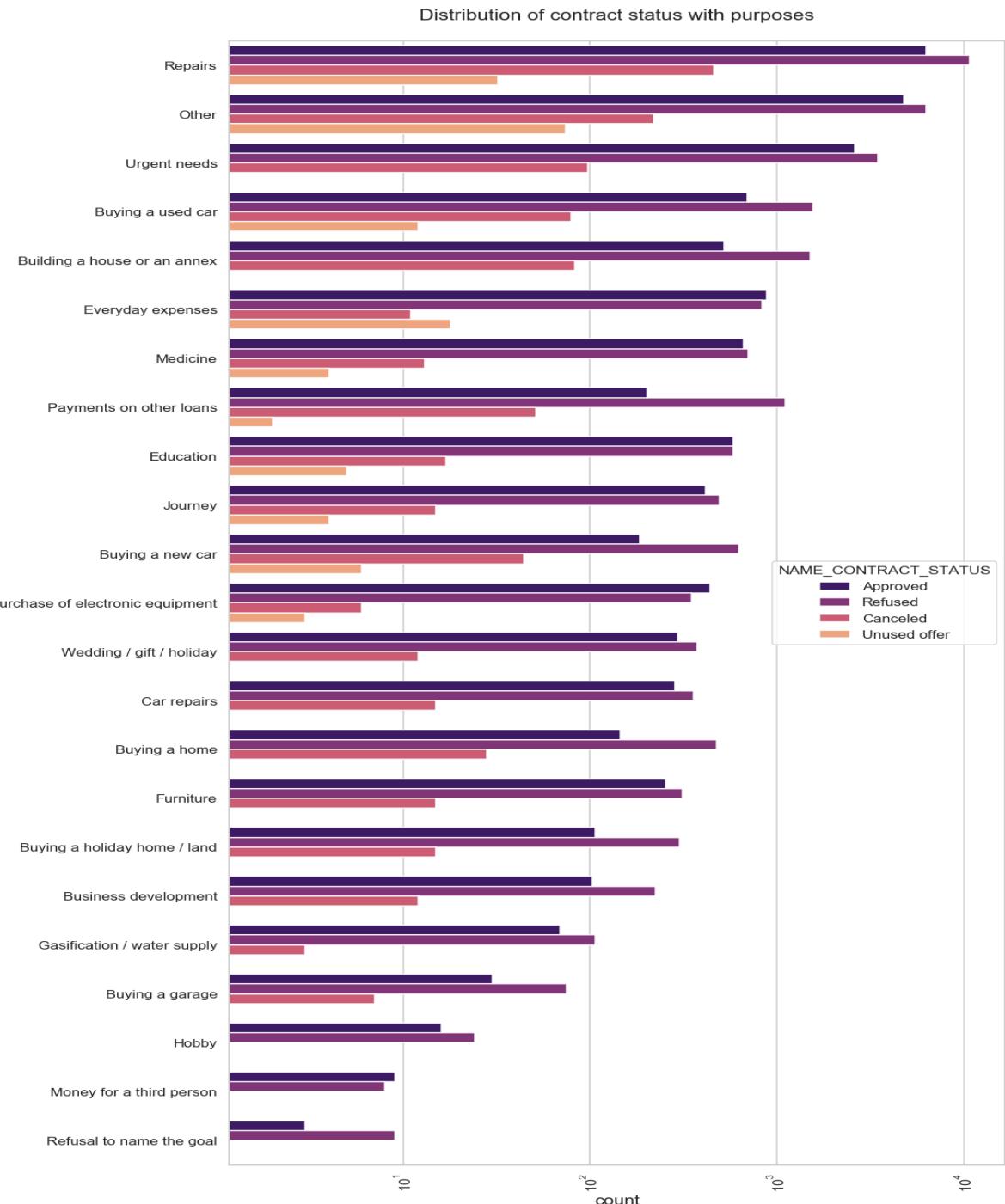
	VAR1	VAR2	Correlation
398	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	1.00
148	AMT_GOODS_PRICE	AMT_CREDIT	0.98
423	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.87
124	AMT_ANNUITY	AMT_CREDIT	0.75
149	AMT_GOODS_PRICE	AMT_ANNUITY	0.75
224	DAYS_EMPLOYED	DAYS_BIRTH	0.58
399	OBS_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.34
374	DEF_30_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.33
248	DAYS_REGISTRATION	DAYS_BIRTH	0.29
422	DEF_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.26

From the observation above we can say that for Target _1 Goods price and loan credit amount are most correlated next to social circle observations for different days. So the variables correlated in Target_0 and Target_1 are same with slightly varying correlation values.

Distribution of Contract status with purposes

Few points can be concluded from the graph.

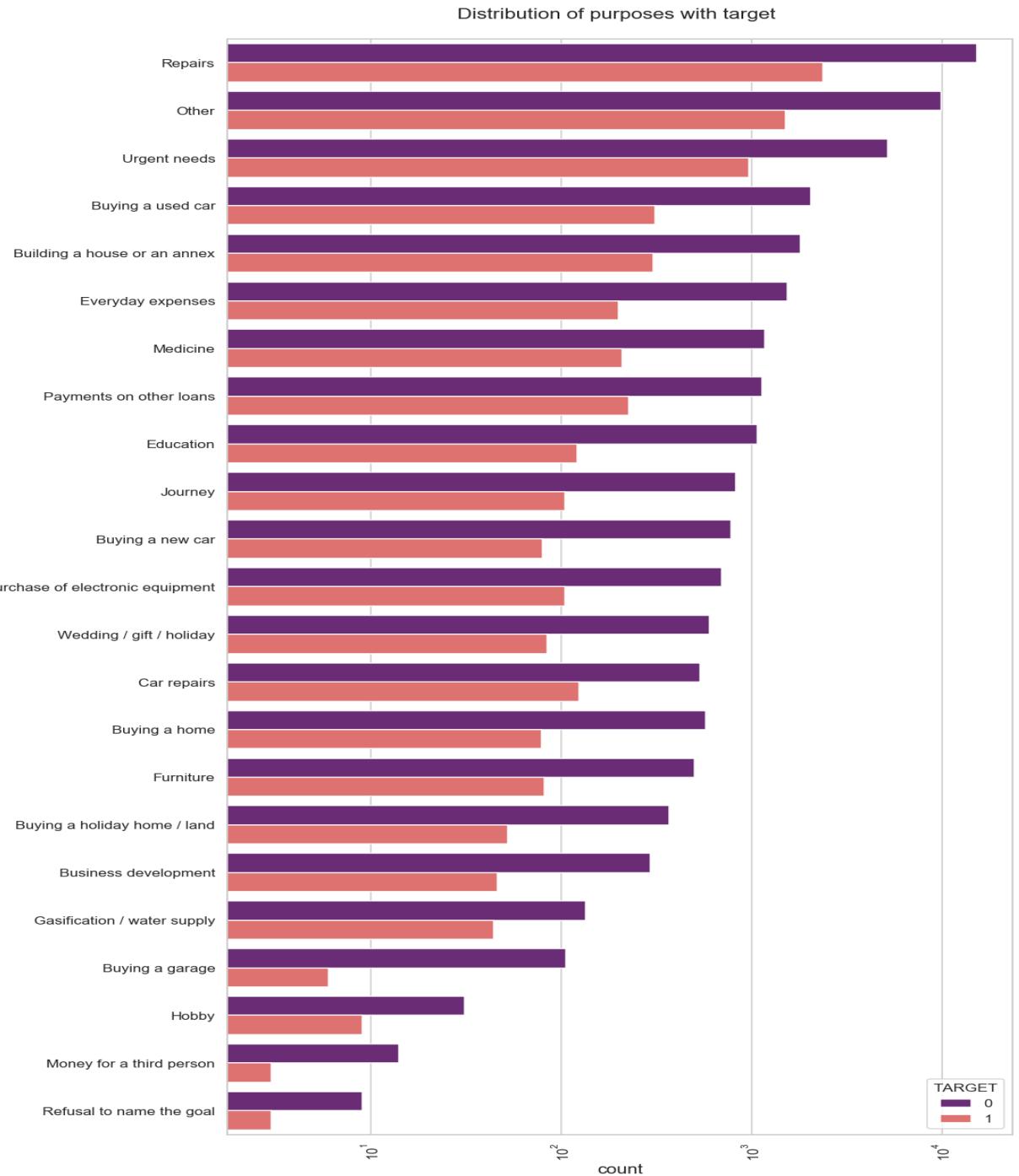
- ❑ Most rejection of loans came from purpose 'repairs'.
- ❑ For education purposes we have equal number of approves and rejection
- ❑ Paying other loans and buying a new car is having significant higher rejection than approves.



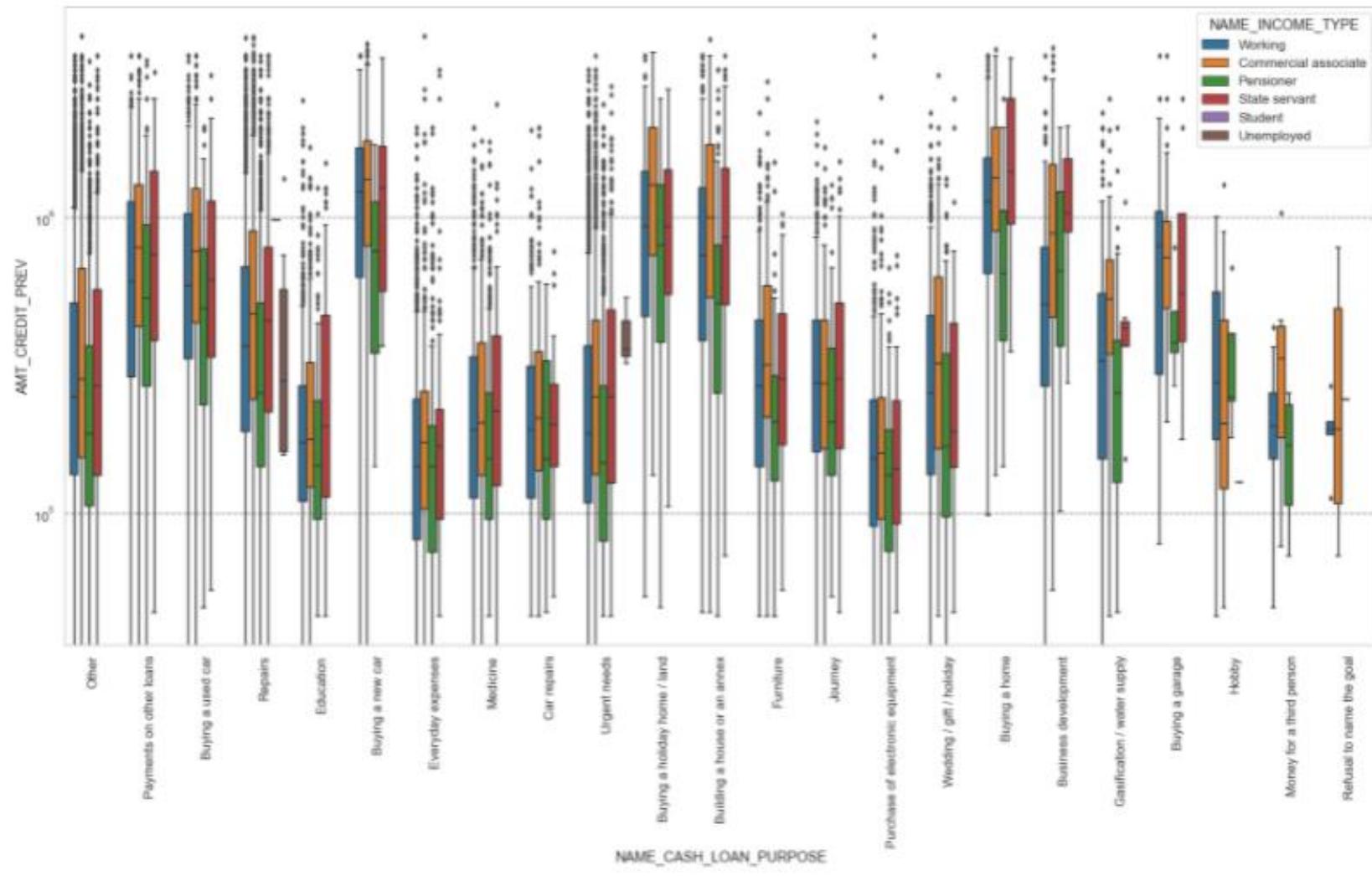
Distribution of Purposes with Target

Few points can be concluded from the graph.

- Loan purposes with 'Repairs' are facing more difficulties in payment on time.
- There are few places where loan payment is significant 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.



Prev Credit amount vs Loan Purpose



Prev Credit amount vs Loan Purpose

From the previous graph we can conclude the below points:

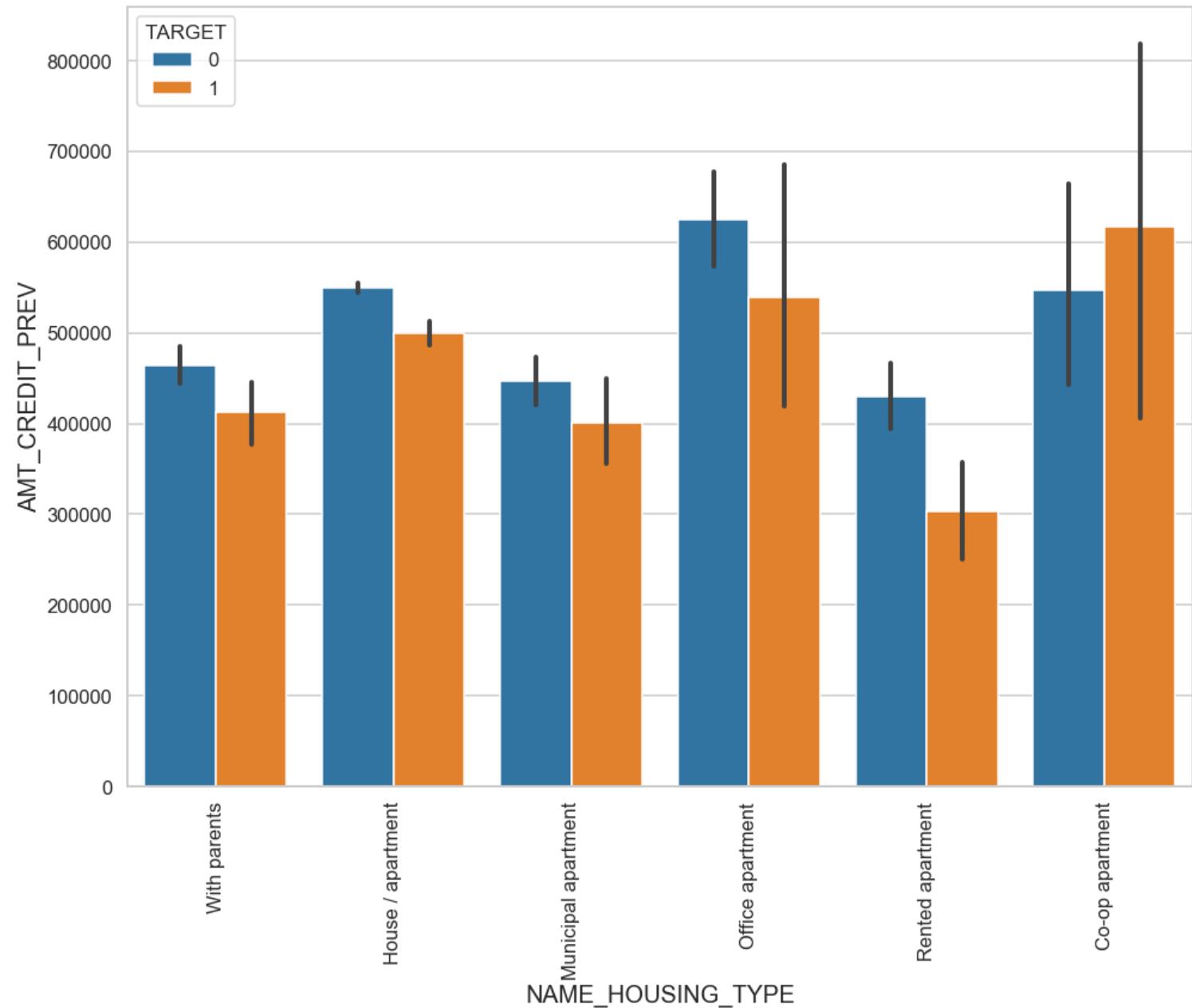
- The credit amount of Loan purposes like Buying a home
- Buying a land Buying a new car and Building a house is higher.
- Income type of state servants have a significant amount of credit applied
- Money for third person or a Hobby is having less credits applied for.

Prev Credit amount vs Housing type

Few points can be concluded from the graph.

- Here for Housing type, office apartment is having higher credit of target 0 and co-op apartment is having higher credit of target 1.
- So, we can conclude that bank should avoid giving loans to the housing type of co-op apartment as they are having difficulties in payment.
- Bank can focus mostly on housing type with parents or House\apartment or municipal apartment for successful payments.

Prev Credit amount vs Housing type



Conclusion

- Banks should focus more on contract type Student , pensioner and Businessman with housing type other than Co-op apartment for successful payments.
- Banks should focus less on income type Working as they are having most number of unsuccessful payments.
- Also with loan purpose Repair is having higher number of unsuccessful payments on time.
- Get as much as clients from housing type With parents as they are having least number of unsuccessful payments.