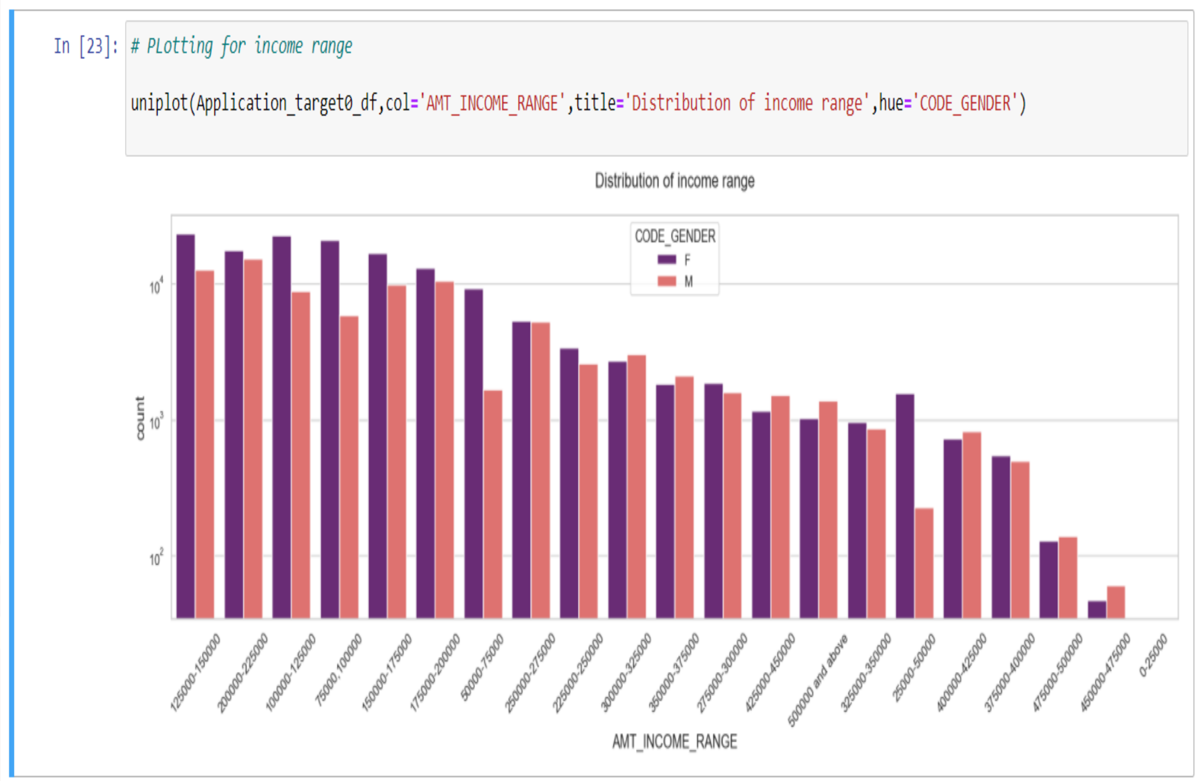
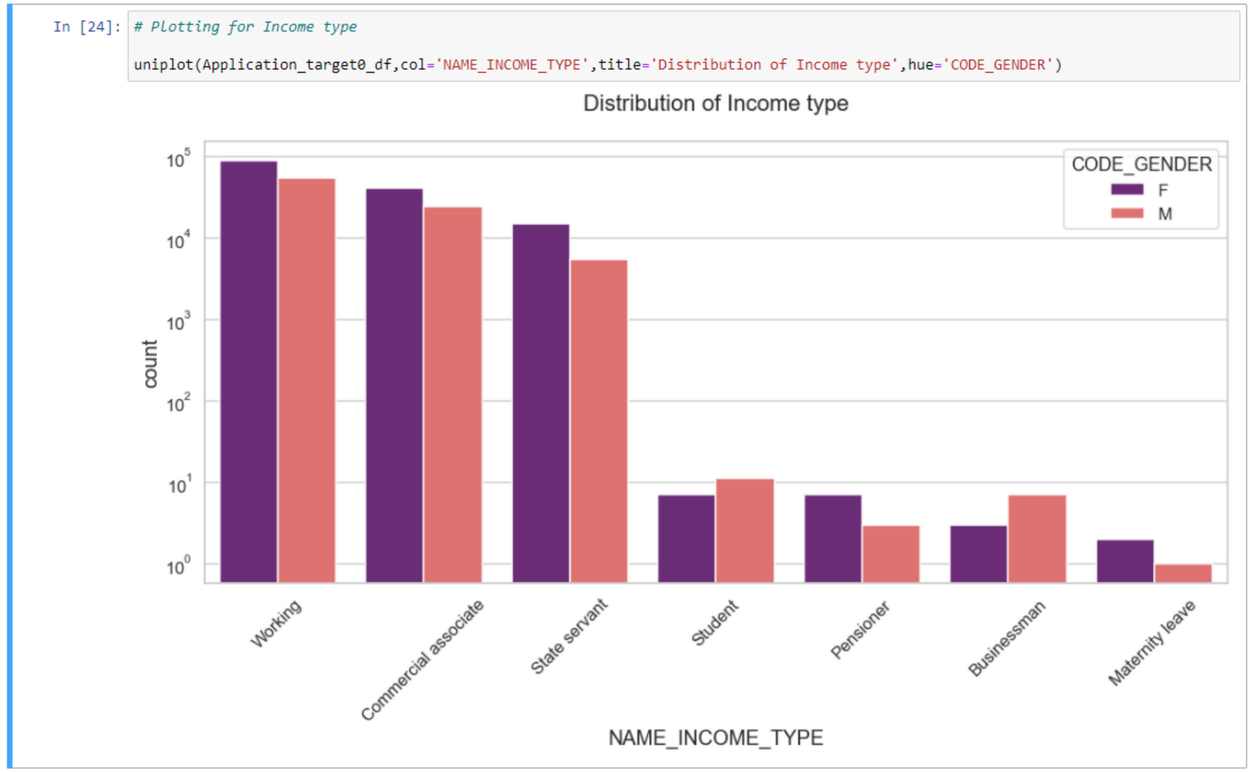
. P20 S#20%#. %Q2# QST0\

P££«a}vB vflufia a¡l %iua%#Qa…n¡a


In [23]:
# PLotting for income range
uniplot(Application_target0_df,col='AMT_INCOME_RANGE',title='Distribution of income range',hue='CODE_GENDER')
Distribution of income range
CODE_GENDER
F
M
count
AMT_INCOME_RANGE
125000-150000
200000-225000
100000-125000
75000,100000
150000-175000
175000-200000
50000-75000
250000-275000
225000-250000
300000-325000
350000-375000
275000-300000
425000-450000
500000 and above
325000-350000
25000-50000
400000-425000
375000-400000
475000-500000
450000-475000
0-25000

0 vfl‡fivi · ‡v£ ¡ ‡£ s‡:¡ j£ ~ n‡
S‰«n

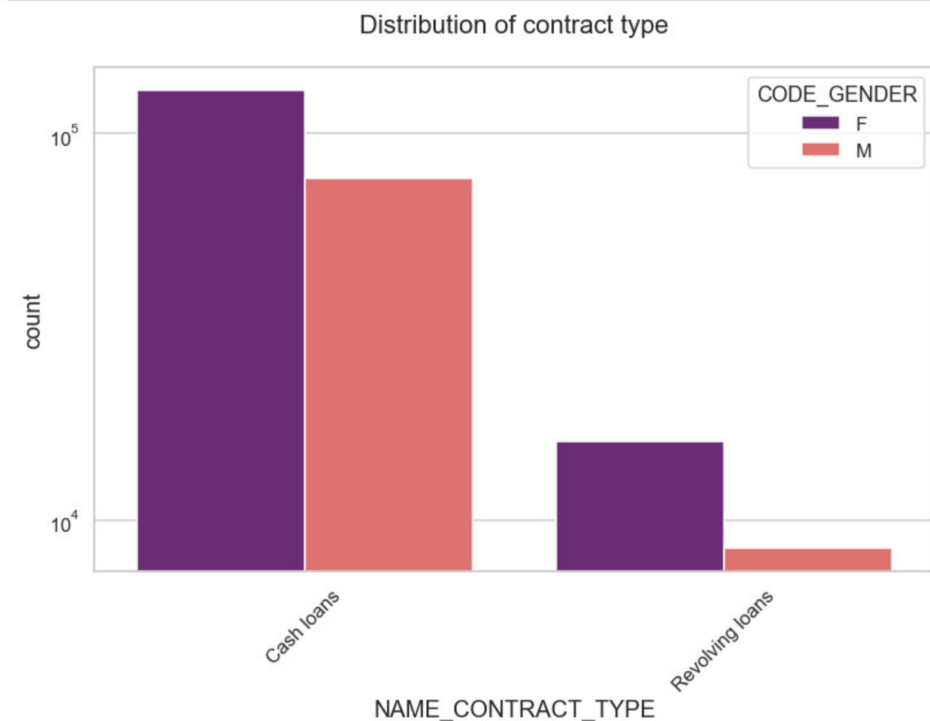
M£v¡ ‡fl‡£ ‡i n‡j £ ¡ j}· l n l ‡f£ ~ ‡un‡
tfia«u‡£ ¡ ‡un‡ivtu‡i˙ i
˙ ‡7£fi‡v¡ j£ ~ n‡‰«n‡fi› £fi|v¡ t ð‡
ğ£ ~ ~ nfijva}‡aflfl£jva‡nð‡a¡ l ‡Q‡a‡n‡
Qnfi”a¡ ‡ð‡un‡¡ · ~ infi‡£ s‡jfinl vfl‡afin‡
uvtunfi‡ua¡ ‡£ ‡unffli
˝ ‡7£fi‡uvfl‡7n~ a}nfl‡afin‡ua”v¡ t ‡~ £fin‡
¡ · ~ infi‡£ s‡jfinl vfl‡ua¡ ‡~ a}ni
˚ ‡Anflfl‡¡ · ~ infi‡£ s‡jfinl vfl‡£fi‡
v¡ j£ ~ n‡‰«n‡fl‡· l n¡ ‡ð‡›n¡ fl‡£ ¡ nfið‡
ñ · flv¡ nflfl~ a¡ ð‡a¡ l ‡B a‡nfi¡ v‰‡
}na”nð

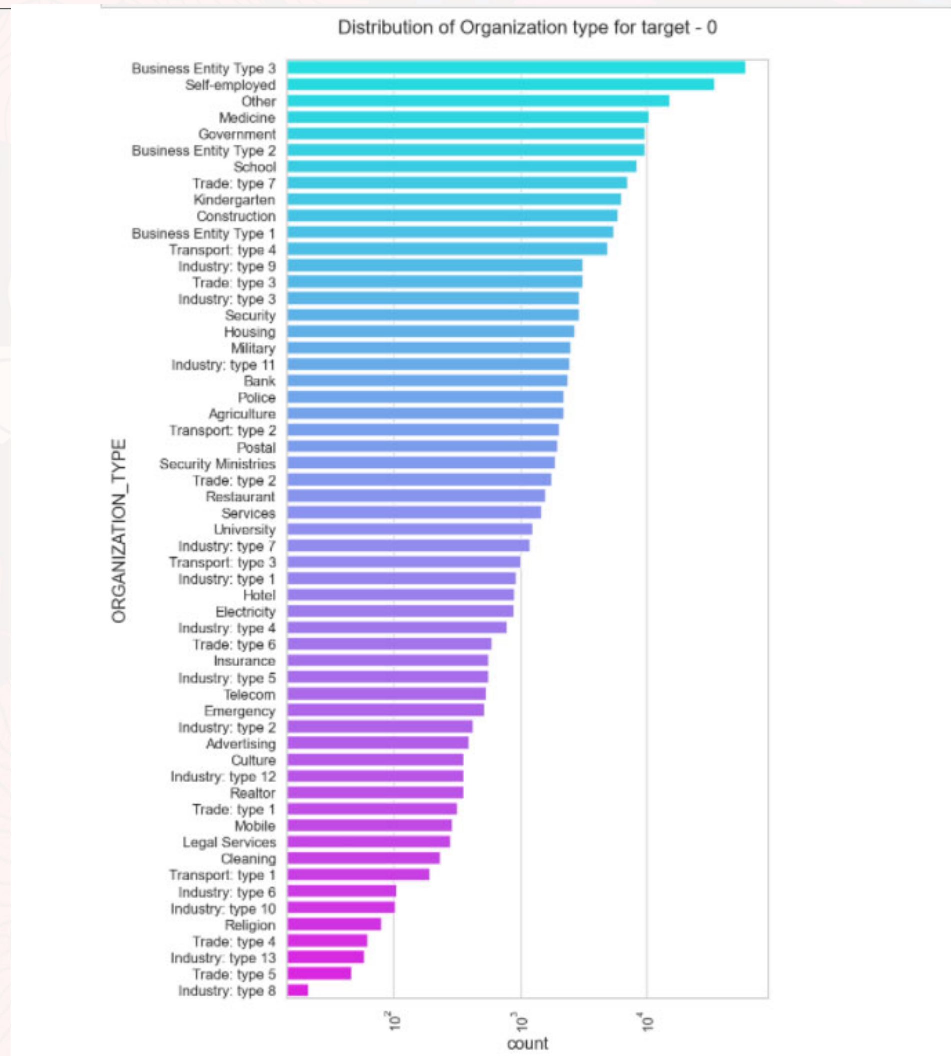


0 vfl#fiλi · ŧvÆ ¡#sÆfi#jÆ ¡#fiaj#
‰«n

M£v¡ #fl#£#i n#j£ ¡ j}· l nl #fi£ ~ #un#
tfia«u#£ ¡#un#ivtu ŧı
˙#7£fi#j£ ¡ #fiaj ##‰«n#jaflu}£a ¡ fl3#fl#
ua”v¡ t #ivtunfi#¡ · ~ infi#£s#jfnl vfl#
ŧua ¡ #Pn”£ }”v¡ t #}£a ¡ fl3#j£ ¡ #fiaj#
‰«nı
˝#7£fi#uvfl#a }fl£#7n~ a }n#fl#nalv¡ t #
s£fi#a« « }‰v¡ t #jfnl vflı

```
In [25]: # Plotting for Contract type
uniplot(Application_target0_df,col='NAME_CONTRACT_TYPE',title='Distribution of contract type',hue='CODE_GENDER')
```

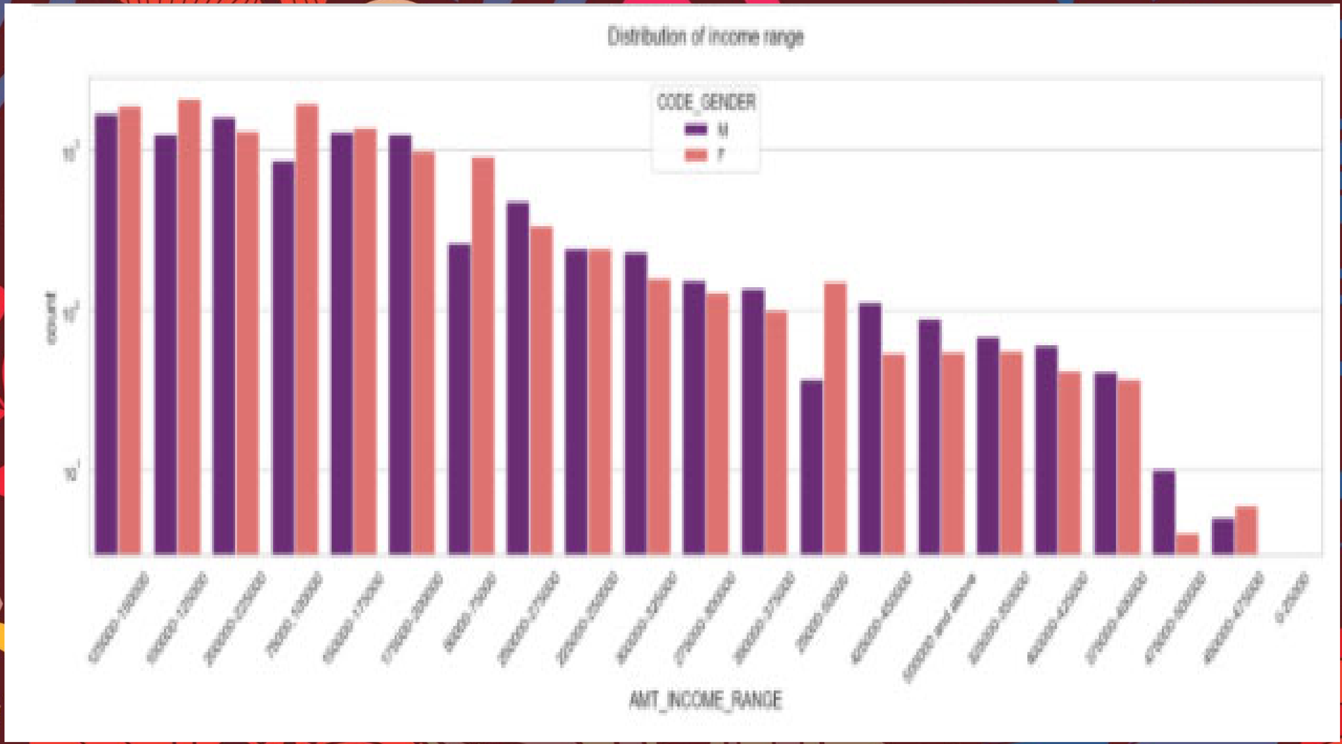




Distribution of Organization type for target - 0
Business Entity Type 3
Self-employed
Other
Medicine
Government
Business Entity Type 2
School
Trade: type 7
Kindergarten
Construction
Business Entity Type 1
Transport: type 4
Industry: type 9
Trade: type 3
Industry: type 3
Security
Housing
Military
Industry: type 11
Bank
Police
Agriculture
Transport: type 2
Postal
Security Ministries
Trade: type 2
Restaurant
Services
University
Industry: type 7
Transport: type 3
Industry: type 1
Hotel
Electricity
Industry: type 4
Trade: type 6
Insurance
Industry: type 5
Telecom
Emergency
Industry: type 2
Advertising
Culture
Industry: type 12
Realtor
Trade: type 1
Mobile
Legal Services
Cleaning
Transport: type 1
Industry: type 6
Industry: type 10
Religion
Trade: type 4
Industry: type 13
Trade: type 5
Industry: type 8
ORGANIZATION_TYPE
10^2
10^3
10^4
count

Points to be concluded from the above graph.

1. Male counts are higher than female.
2. Income range from 100000 to 200000 is having more number of credits.
3. This graph show that males are more than female in having credits for that range.
4. Very less count for income range 400000 and above.

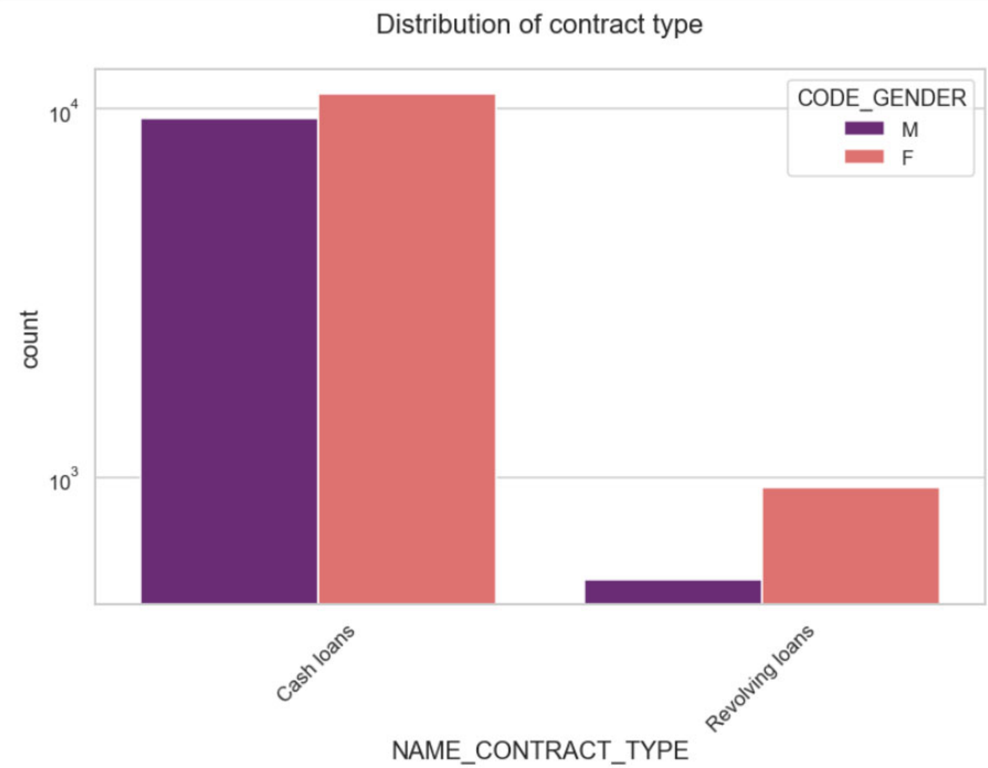


Points to be concluded from the above graph.

1. For income type ‘working’, ’commercial associate’, and ‘State Servant’ the number of credits are higher than other i.e. ‘Maternity leave.
2. For this Females are having more number of credits than male.
3. Less number of credits for income type ‘Maternity leave’.
4. For type 1: There is no income type for ‘student’ , ’pensioner’ and ‘Businessman’ which means they don’t do any late payments.

Points to be concluded from the above graph.
1. For contract type ‘cash loans’ is having higher number of credits than ‘Revolving loans’ contract type.
2. For this also Female is leading for applying credits.
3. For type 1 : there is only Female Revolving loans.

```
In [29]: # Plotting for Contract type

uniplot(Application_target1_df,col='NAME_CONTRACT_TYPE',title='Distribution of contract type',hue='CODE_GENDER')
```

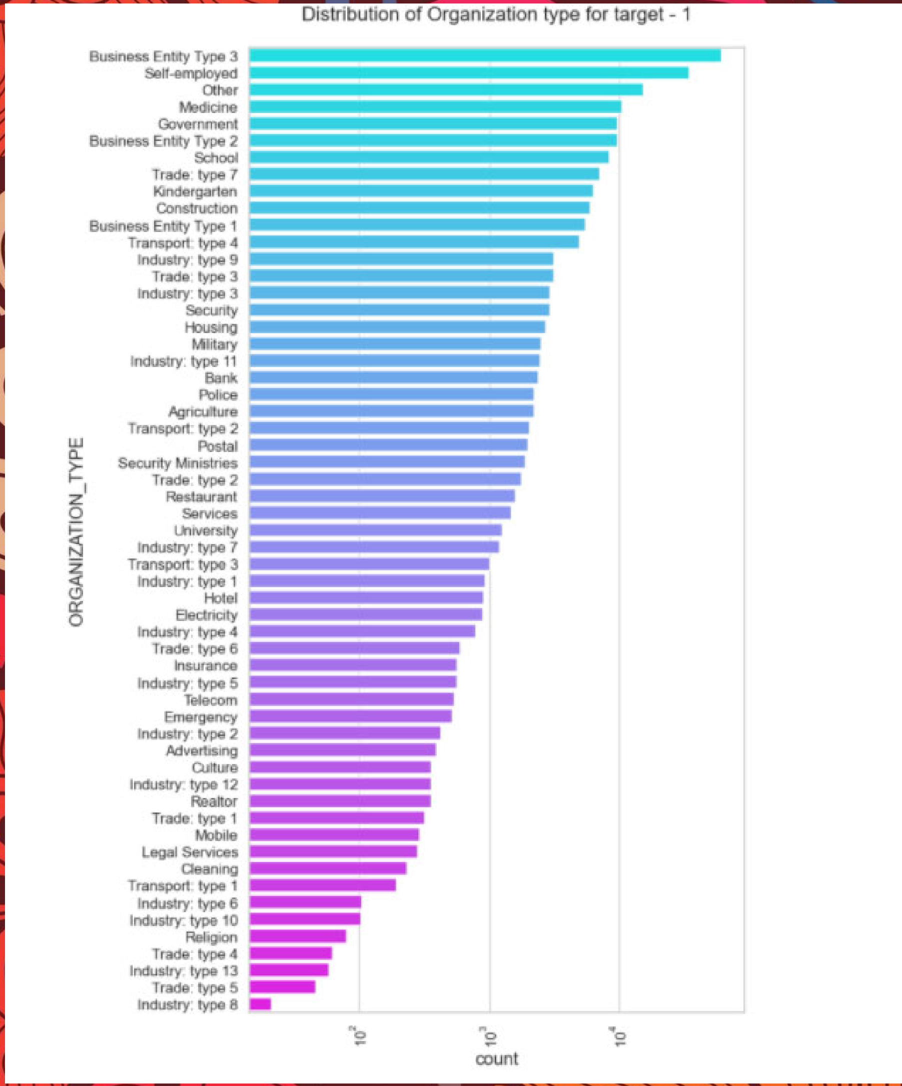


Points to be concluded from the above graph.

1. Clients which have applied for credits are from most of the organization type ‘Business entity Type 3’ , ‘Self employed’ , ‘Other’ , ‘Medicine’ and ‘Government’.

2. Less clients are from Industry type 8,type 6, type 10, religion and trade type 5, type 4.

3. Same as type 0 in distribution of organization type.



In [32]:

```
# Correlation for target 0

Application_target0
```

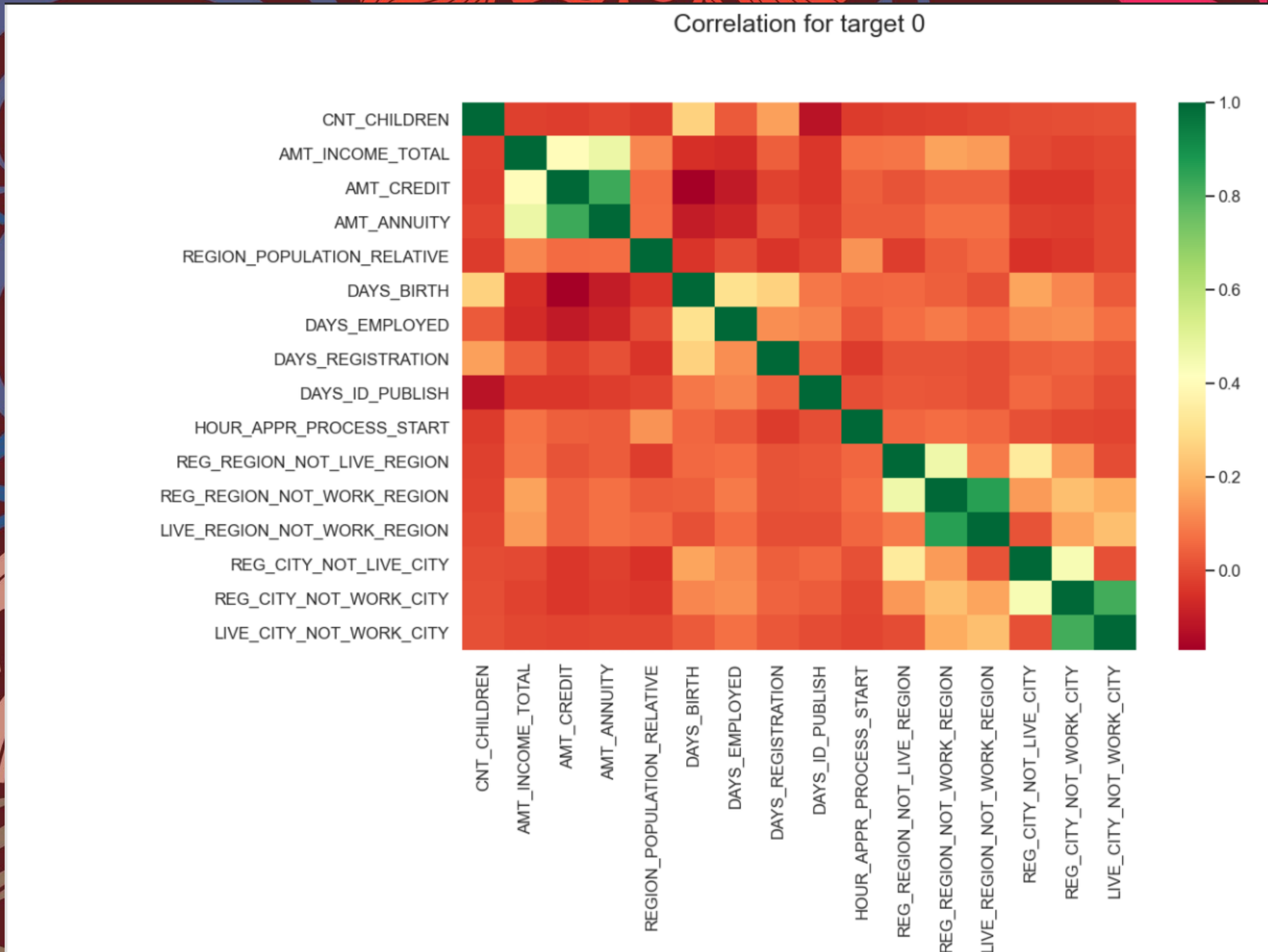
Out[32]:

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAY:
CNT_CHILDREN	1.000000	-0.021950	-0.023652	-0.010795	-0.030579	0.266534	
AMT_INCOME_TOTAL	-0.021950	1.000000	0.403876	0.472204	0.110074	-0.054666	
AMT_CREDIT	-0.023652	0.403876	1.000000	0.826689	0.060706	-0.169030	
AMT_ANNUITY	-0.010795	0.472204	0.826689	1.000000	0.064328	-0.100287	
REGION_POPULATION_RELATIVE	-0.030579	0.110074	0.060706	0.064328	1.000000	-0.041663	
DAYS_BIRTH	0.266534	-0.054666	-0.169030	-0.100287	-0.041663	1.000000	
DAYS_EMPLOYED	0.030948	-0.060868	-0.104251	-0.074643	0.000900	0.307787	
DAYS_REGISTRATION	0.155518	0.040559	-0.015318	0.010712	-0.042400	0.265449	
DAYS_ID_PUBLISH	-0.119164	-0.036702	-0.038197	-0.027354	-0.010299	0.083331	
HOUR_APPR_PROCESS_START	-0.030162	0.073503	0.036923	0.032953	0.133213	0.051299	
REG_REGION_NOT_LIVE_REGION	-0.022813	0.077634	0.015118	0.033435	-0.025292	0.058627	
REG_REGION_NOT_WORK_REGION	-0.015475	0.159962	0.041693	0.070841	0.032446	0.038104	
LIVE_REGION_NOT_WORK_REGION	-0.005576	0.148281	0.045175	0.069051	0.056814	0.012789	
REG_CITY_NOT_LIVE_CITY	0.002344	-0.001023	-0.040616	-0.019954	-0.049779	0.167477	
REG_CITY_NOT_WORK_CITY	0.007487	-0.013856	-0.037000	-0.024085	-0.034808	0.111539	
LIVE_CITY_NOT_WORK_CITY	0.013295	-0.004758	-0.011194	-0.008087	-0.007332	0.029007	

In [33]:
```python
# Correlation for target 1

Application_target1
```

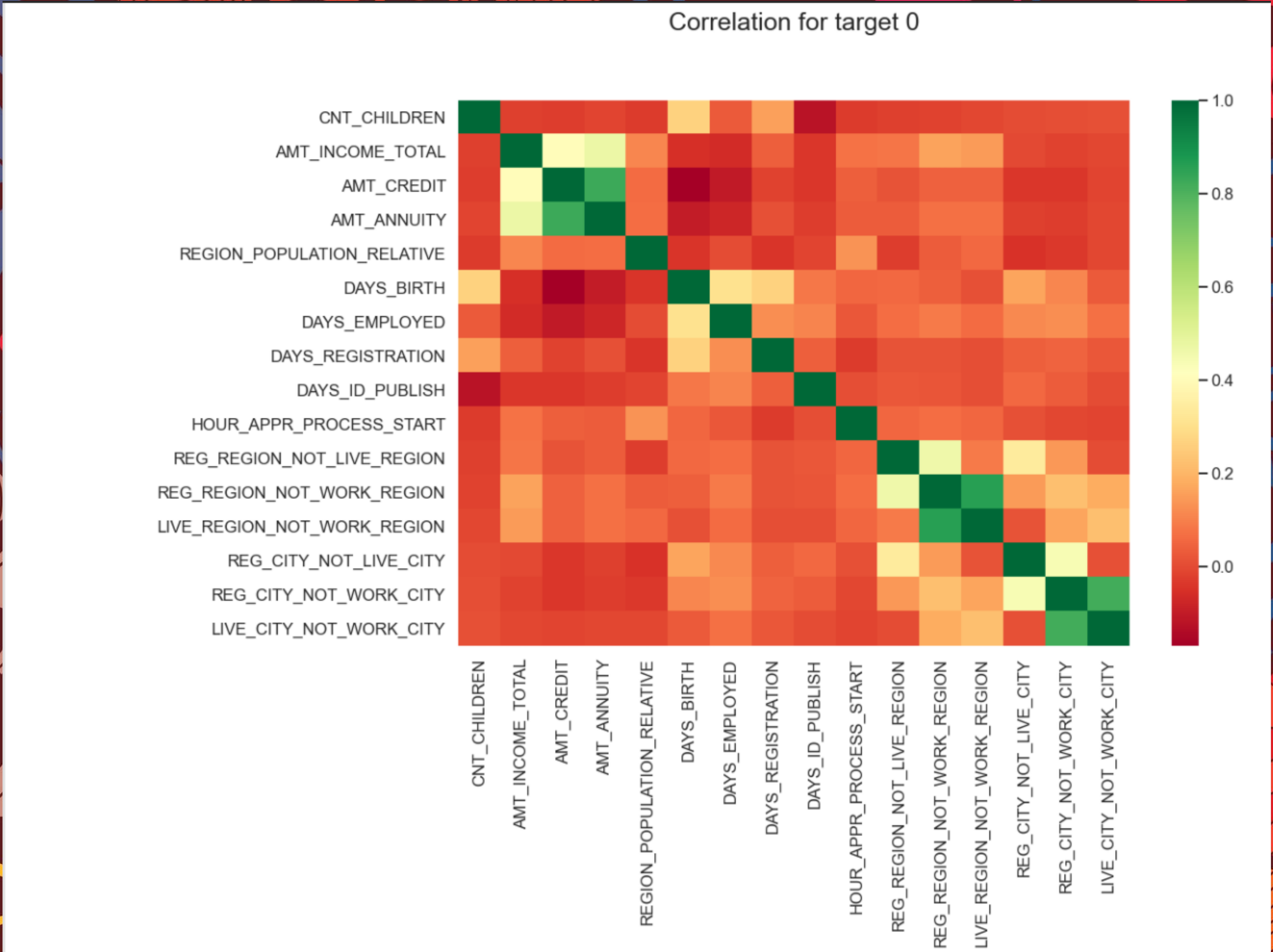
Out[33]:

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	REGION_POPULATION_RELATIVE	DAYS_BIRTH	DAY
CNT_CHILDREN	1.000000	-0.039123	0.000427	0.015133	-0.029682	0.175025	
AMT_INCOME_TOTAL	-0.039123	1.000000	0.364559	0.428947	0.058005	-0.103026	
AMT_CREDIT	0.000427	0.364559	1.000000	0.812093	0.043545	-0.200718	
AMT_ANNUITY	0.015133	0.428947	0.812093	1.000000	0.028666	-0.100200	
REGION_POPULATION_RELATIVE	-0.029682	0.058005	0.043545	0.028666	1.000000	-0.044444	
DAYS_BIRTH	0.175025	-0.103026	-0.200718	-0.100200	-0.044444	1.000000	
DAYS_EMPLOYED	0.006823	-0.053798	-0.107605	-0.060193	-0.015246	0.256870	
DAYS_REGISTRATION	0.110854	0.011378	-0.021973	0.019762	-0.033490	0.192350	
DAYS_ID_PUBLISH	-0.091042	-0.051113	-0.065143	-0.044128	-0.017779	0.146246	
HOUR_APPR_PROCESS_START	-0.040338	0.078779	0.024616	0.021129	0.109400	0.041994	
REG_REGION_NOT_LIVE_REGION	-0.035213	0.075615	0.015043	0.029646	-0.032702	0.046320	
REG_REGION_NOT_WORK_REGION	-0.040853	0.156374	0.032536	0.060363	-0.008160	0.022208	
LIVE_REGION_NOT_WORK_REGION	-0.027993	0.145982	0.034861	0.059724	0.012602	0.000356	
REG_CITY_NOT_LIVE_CITY	-0.016072	-0.003813	-0.030974	-0.011744	-0.057239	0.145884	
REG_CITY_NOT_WORK_CITY	-0.005444	-0.006241	-0.032882	-0.015938	-0.044761	0.096181	
LIVE_CITY_NOT_WORK_CITY	0.009557	0.004230	-0.012465	-0.003012	-0.014753	0.009633	


Correlation for target 0
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
REG_REGION_NOT_LIVE_REGION
REG_REGION_NOT_WORK_REGION
LIVE_REGION_NOT_WORK_REGION
REG_CITY_NOT_LIVE_CITY
REG_CITY_NOT_WORK_CITY
LIVE_CITY_NOT_WORK_CITY
1.0
0.8
0.6
0.4
0.2
0.0

As we can see from above correlation heatmap, There are number of observation we can point out

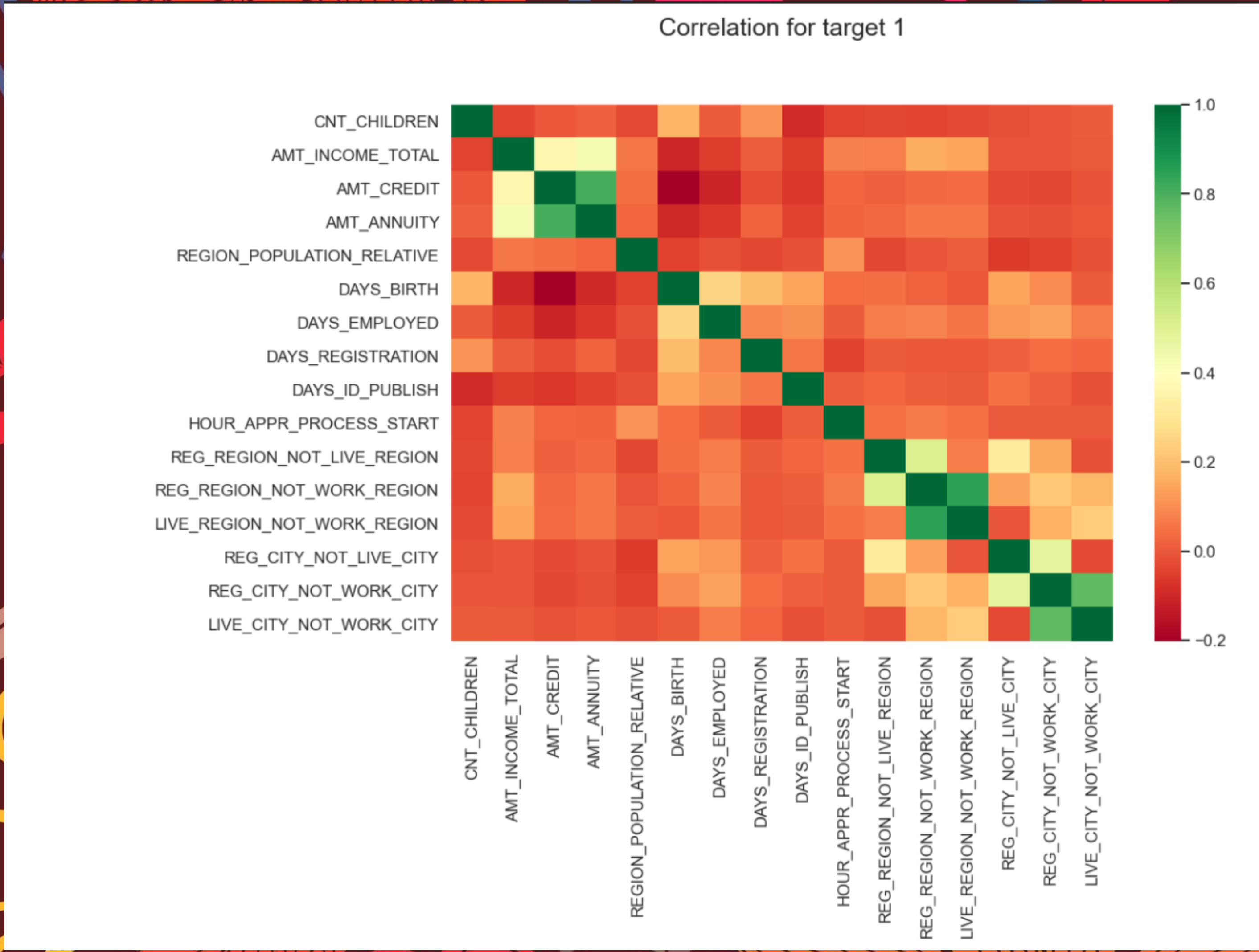
1. Credit amount is inversely proportional to the date of birth, which means Credit amount is higher for low age and vice-versa.
2. Credit amount is inversely proportional to the number of children client have, means Credit amount is higher for less children count client have and vice-versa.
3. Income amount is inversely proportional to the number of children client have, means more income for less children client have and vice-versa.
4. less children client have in densely populated area.
5. Credit amount is higher to densely populated area.
6. The income is also higher in densely populated area.

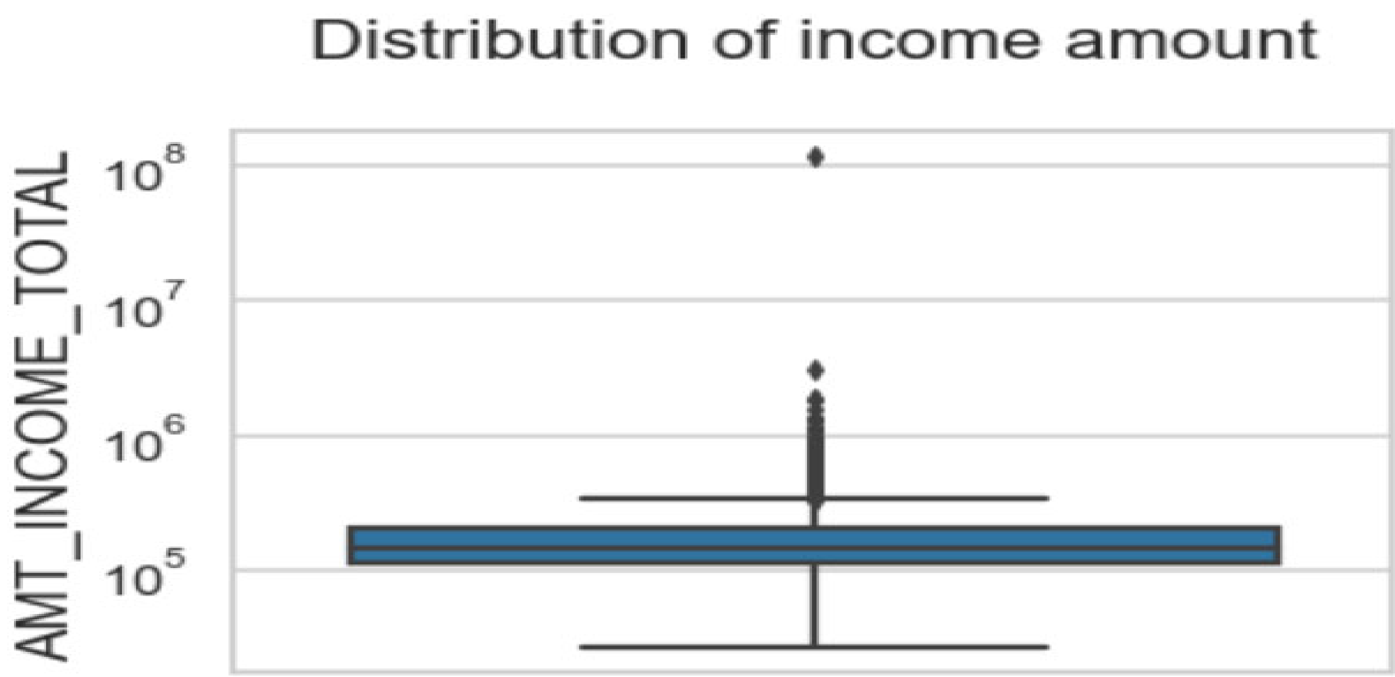


This heat map for Target 1 is also having quite a same observation just like Target 0. But for few points are different. They are listed below.

1. The client's permanent address does not match contact address are having less children and vice-versa
2. the client's permanent address does not match work address are having less children and vice-versa

Univariate analysis for variables



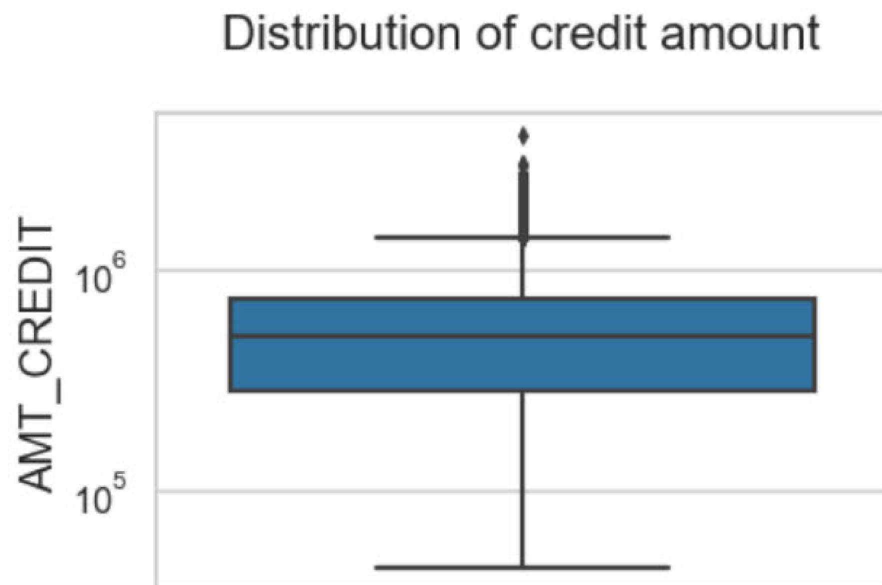


Few points can be concluded from the graph above.

1. Some outliers are noticed in income amount.

2. The third quartiles is very slim for income amount.

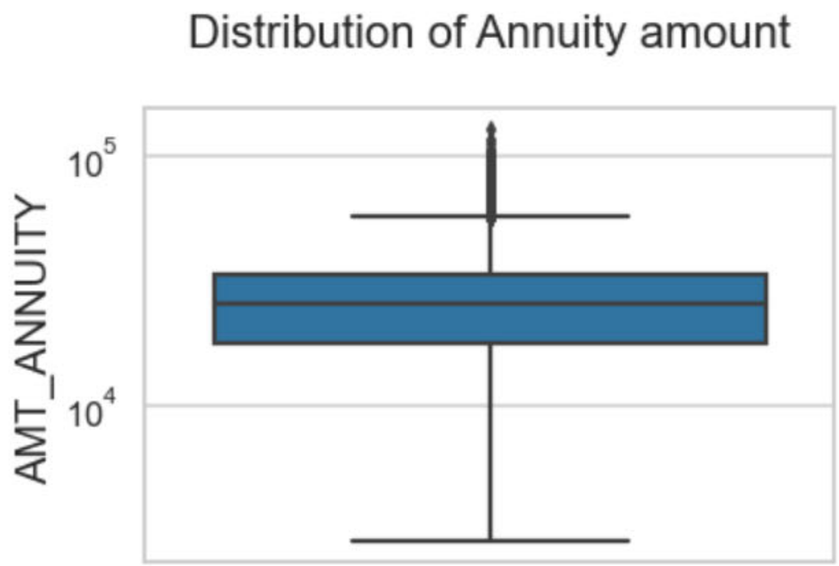
In [39]:
```
# Disrtibution of credit amount

univariate_numerical(data=Application_target0_df,col='AMT_CREDIT',title='Distribution of credit amount')
```



Few points can be concluded from the graph above.

1. Some outliers are noticed in credit amount.

In [40]:
```
# Distribution of anuuity amount

univariate_numerical(data=Application_target0_df,col='AMT_ANNUITY',title='Distribution of Annuity amount')
```



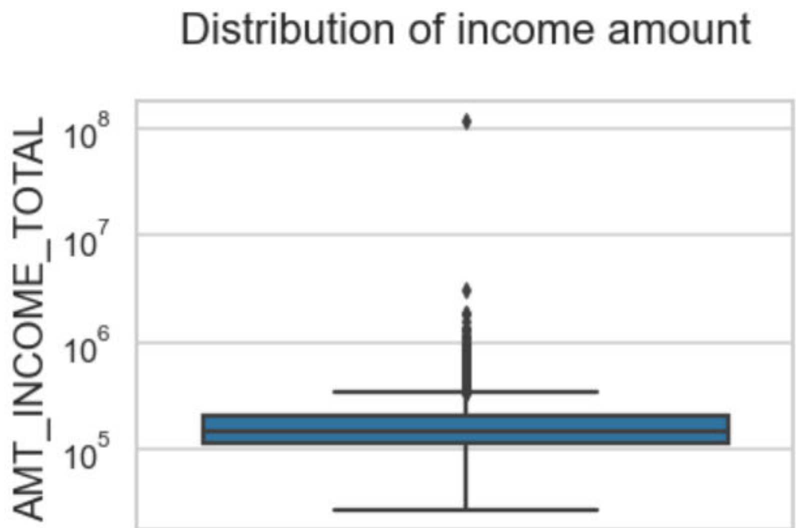
Few points can be concluded from the graph above.

1. Some outliers are noticed in annuity amount.

2. The first quartile is bigger than third quartile for annuity amount which means most of the annuity clients are from first quartile.

For Target 1 - Finding any outliers

In [41]:
```
# Distribution of income amount

univariate_numerical(data=Application_target1_df,col='AMT_INCOME_TOTAL',title='Distribution of income amount')
```



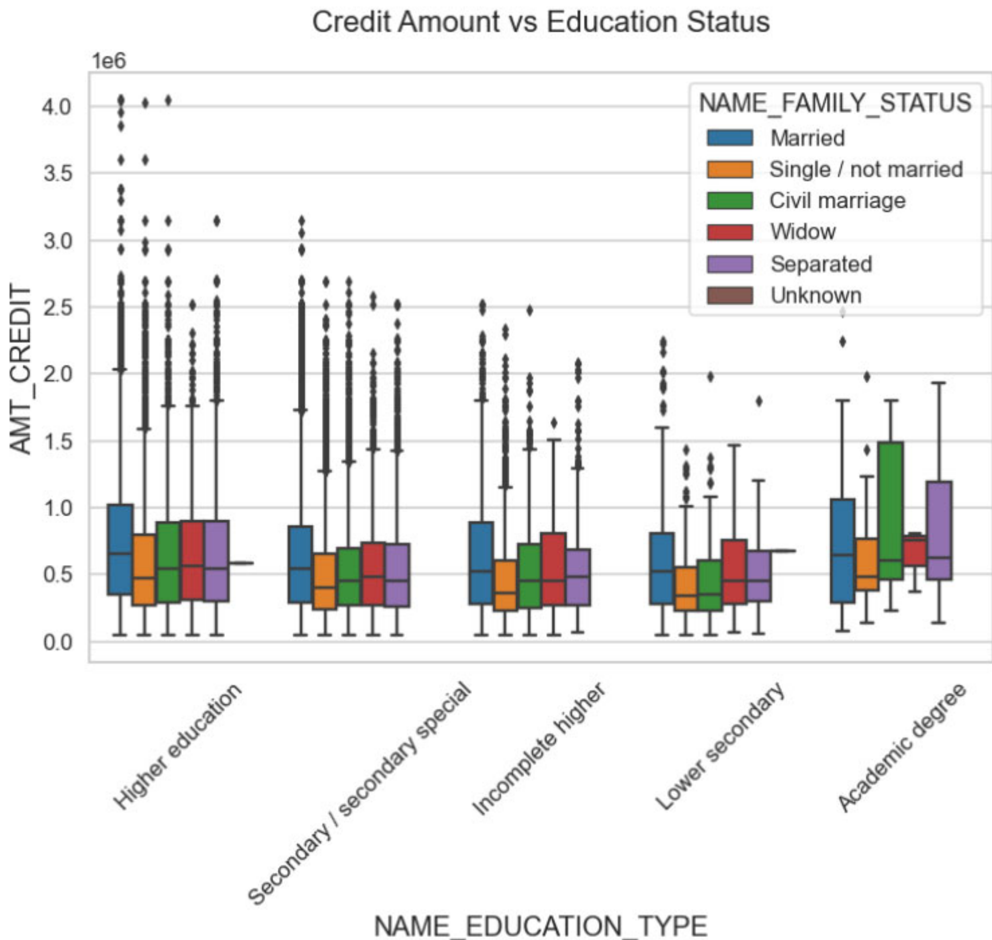
Few points can be concluded from the graph above.

1. Some outliers are noticed in income amount.

2. The third quartiles is very slim for income amount.

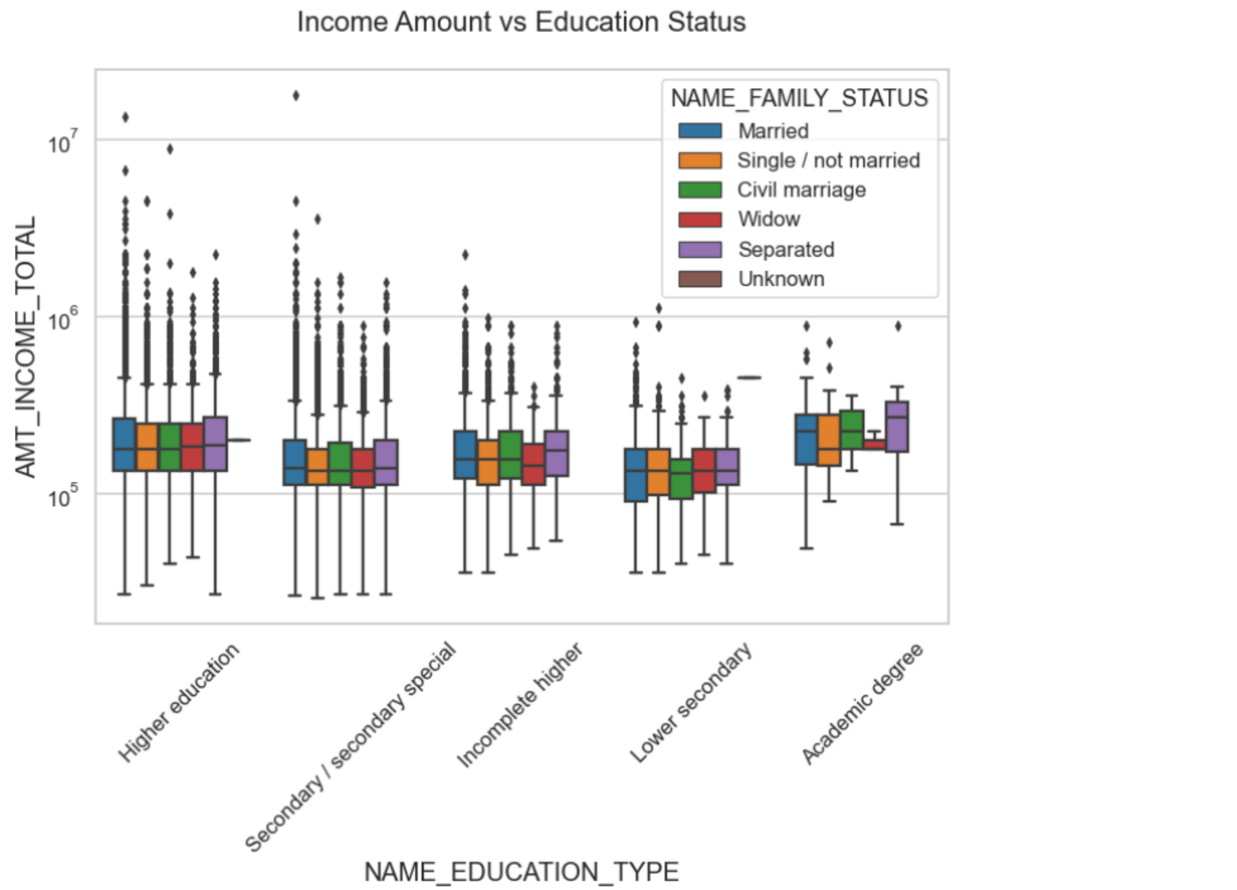
3. Most of the clients of income are present in first quartile.

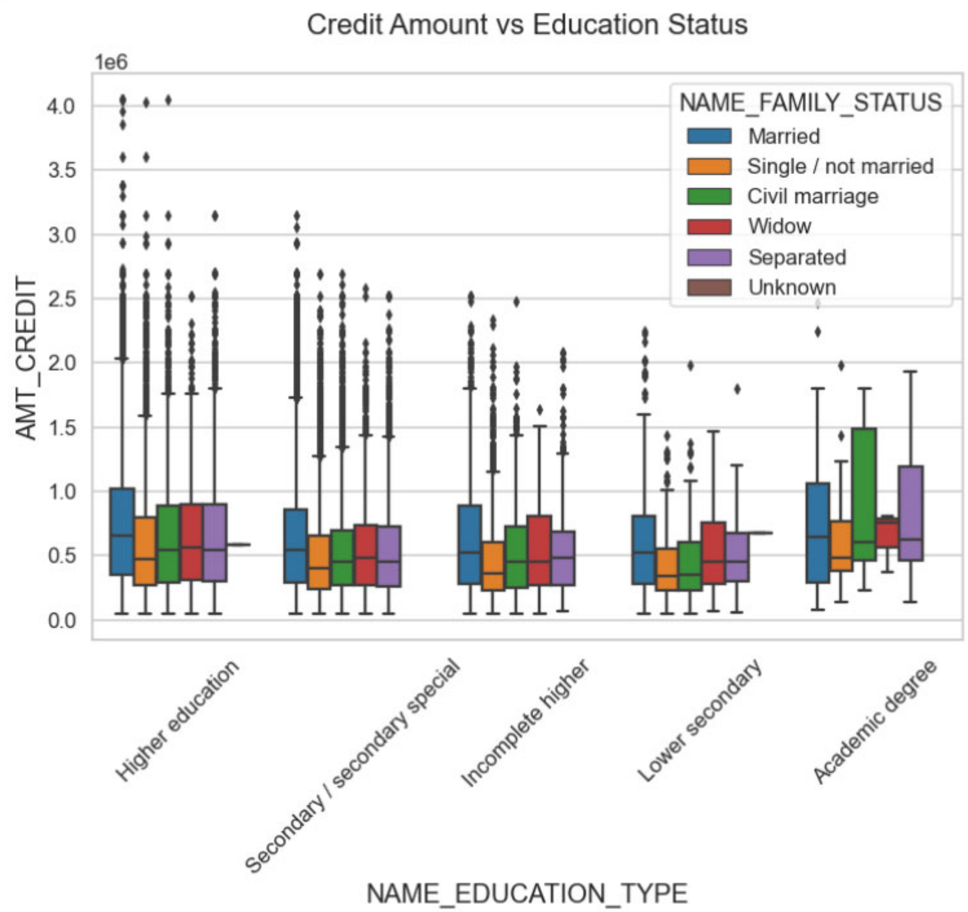
```
In [44]: # Box plotting for Credit amount

         plt.figure(figsize=(12,8))
         plt.xticks(rotation=45)
         sns.boxplot(data =Application_target0_df, x='NAME_EDUCATION_TYPE',y='AMT_CREDIT', hue ='NAME_FAMILY_STATUS',orient='v')
         plt.title('Credit Amount vs Education Status')
         plt.show()
```



In [45]:

```
# Box plotting for Income amount in logarithmic scale

plt.figure(figsize=(12,8))
plt.xticks(rotation=45)
plt.yscale('log')
sns.boxplot(data =Application_target0_df, x='NAME_EDUCATION_TYPE',y='AMT_INCOME_TOTAL', hue ='NAME_FAMILY_STATUS',orient='v')
plt.title('Income Amount vs Education Status')
plt.show()
```

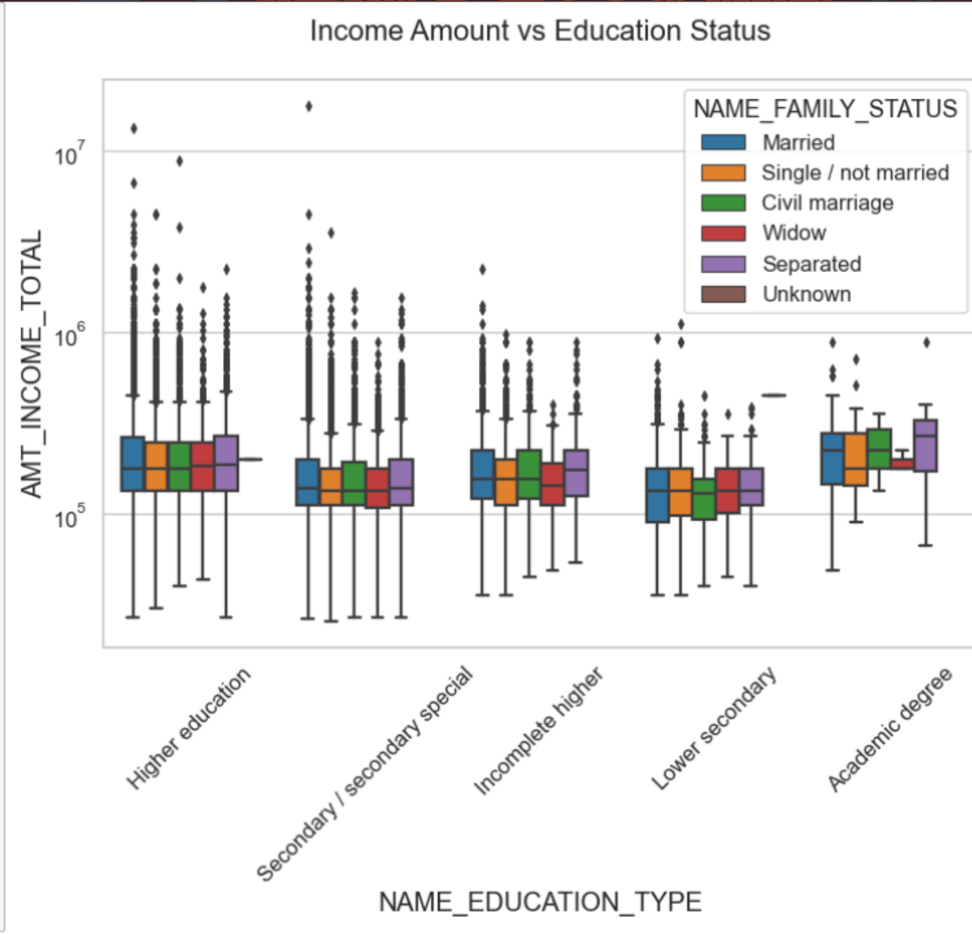




Quite similar with Target 0

From the above box plot 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.

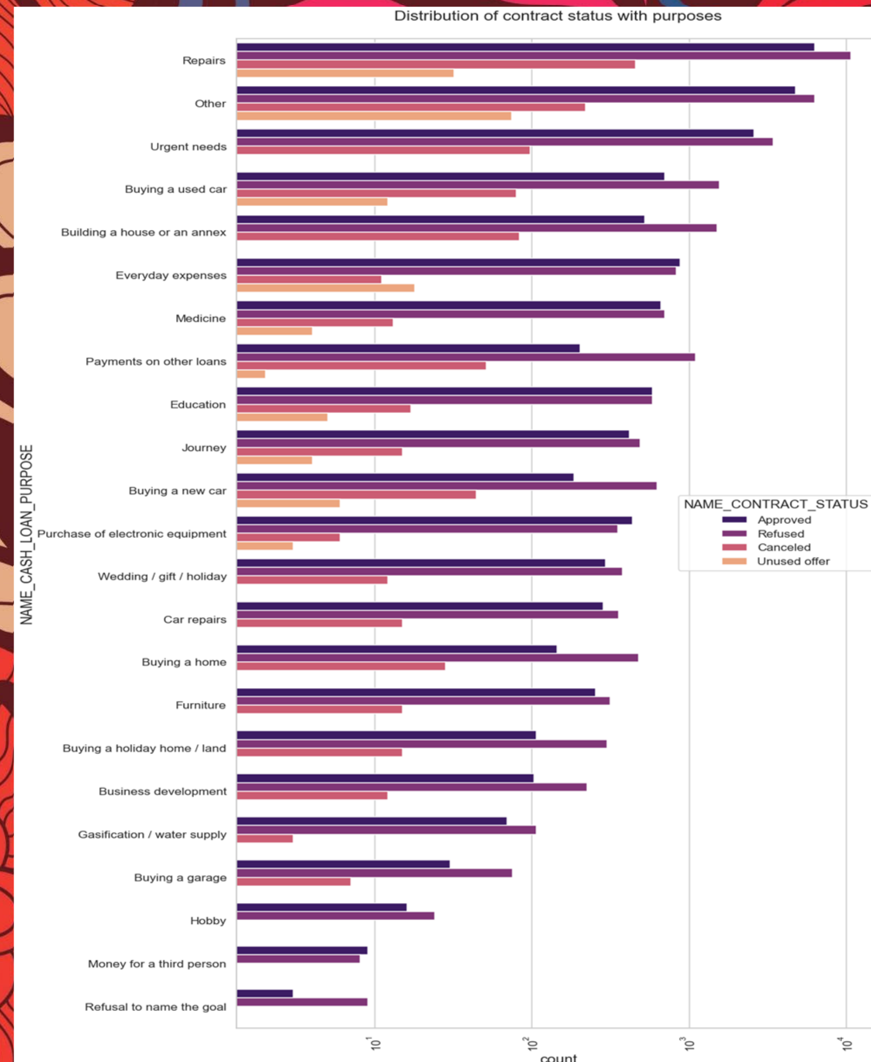


Have some similarity with Target0,

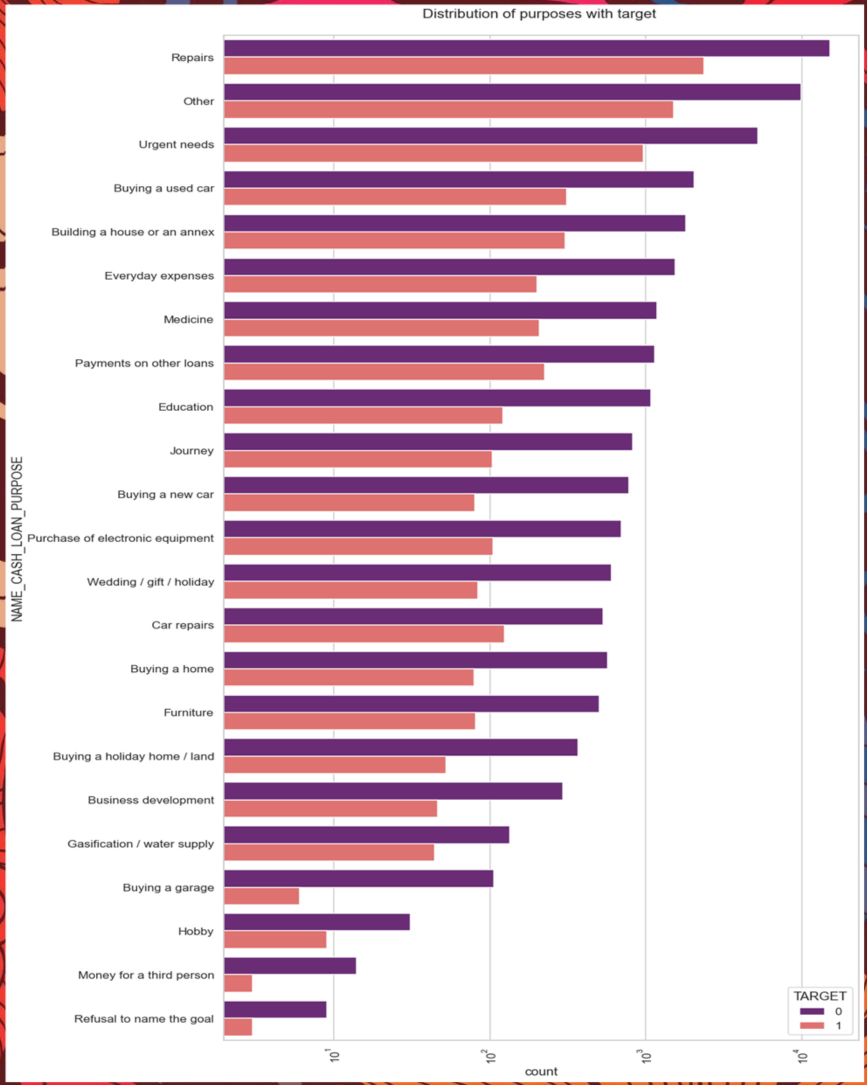
From above boxplot for Education type 'Higher education' the income amount is mostly equal with family status. Less outlier are having for Academic degree but there income amount is little higher that Higher education.

Lower secondary are have less income amount than others.

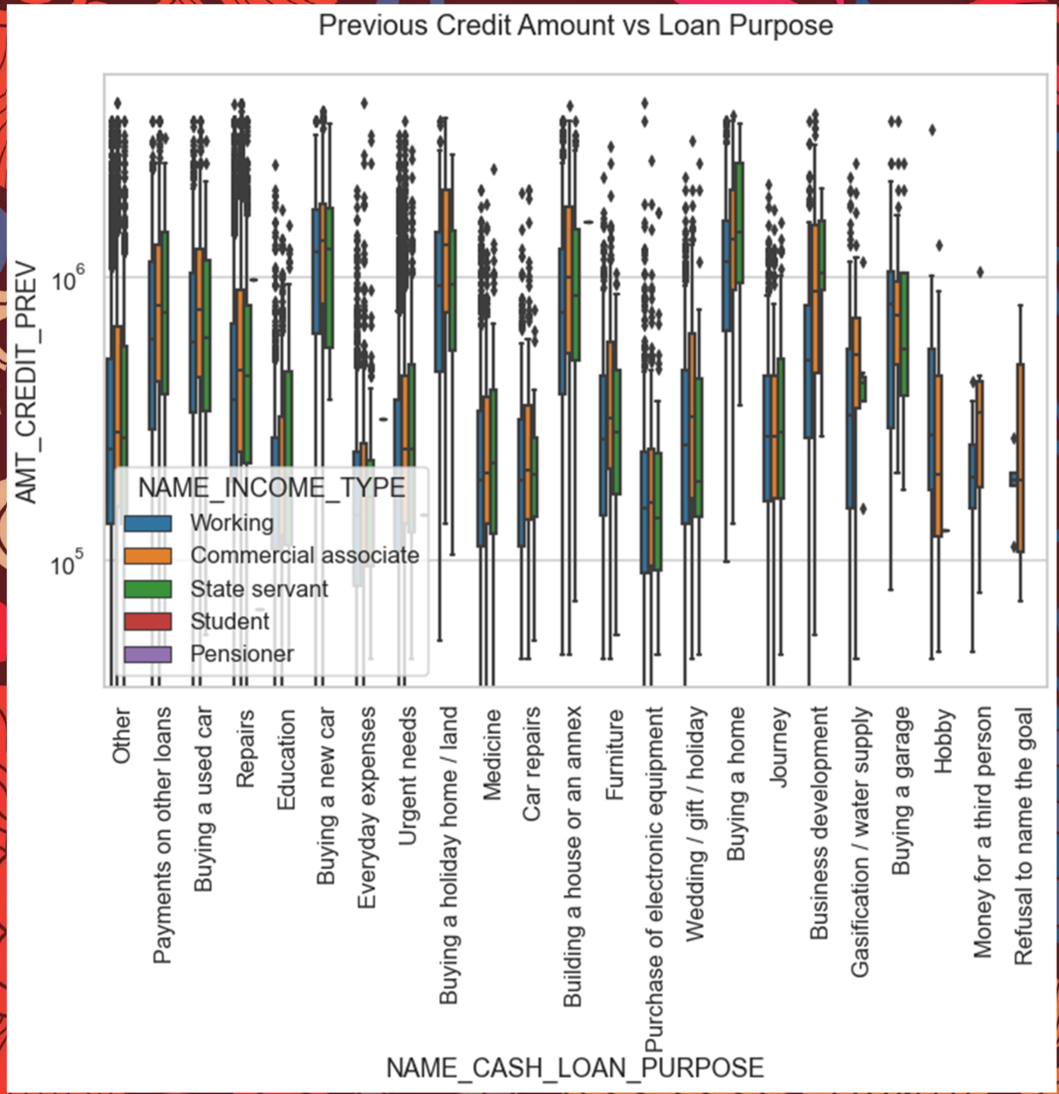
Points to be concluded from above plot:
1. Most rejection of loans came from purpose 'repairs'.
2. For education purposes we have equal number of approves and rejection
3. Payign other loans and buying a new car is having significant higher rejection than approves.



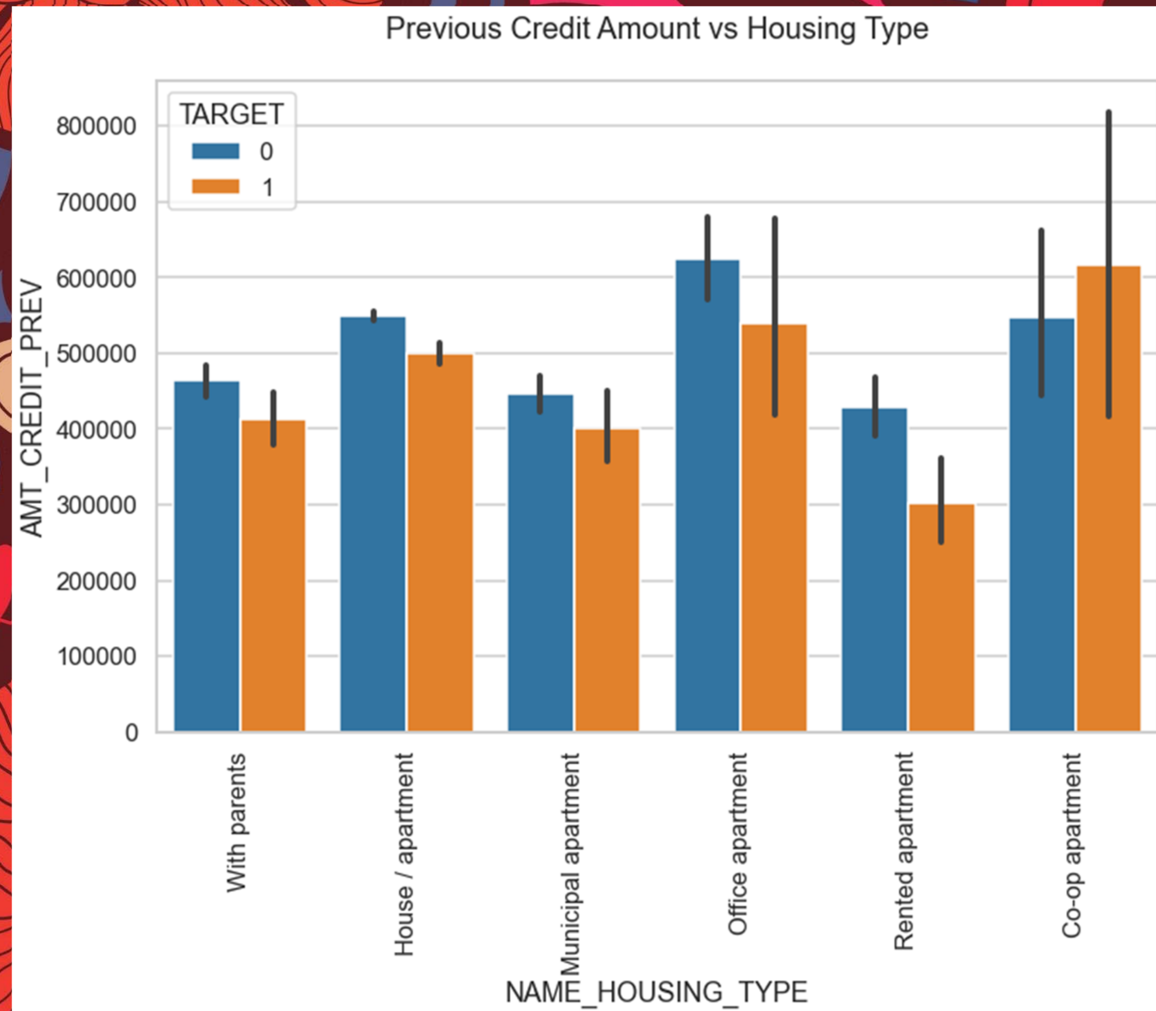
Few points we can conclude from abpve plot:
1. Loan purposes with 'Repairs' are facing more difficulites in payment on time.
2. There are few places where loan payment is significant higher than facing difficulties.
They are 'Buying a garage', 'Business developemt', 'Buying land','Buying a new car' and 'Education'
Hence we can focus on these purposes for which the client is having for minimal payment difficulties.



From the above we can conclude some points-
1. The credit amount of Loan purposes like 'Buying a home','Buying a land','Buying a new car' and'Building a house' is higher.
2. Income type of state servants have a significant amount of credit applied
3. Money for third person or a Hobby is having less credits applied for.



Here for Housing type, office appartment 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\appartment or miuncipal appartment for successful payments.



CONCLUSION

˙ı-a¡|fl#u£·}l#£j·fl#-£fin#£¡#j£¡#aj#‰«n#Q‡·ln¡#ð˜n¡flv£¡nfið#a¡l#ñ·flv¡nflfl~a¡ð#»vu# u£·flv¡t#‰«n#£‡unfi#ua¡#ñ£Œ«#a«afi#~n¡#ð#£fi#fl·jjnflfls·}#«a‰~n¡#flı

¨ı#-a¡|fl#u£·}l#£j·fl#}nflfl#£¡#v¡j£~n#‰«n#Z£fi|v¡tð#afl#un‰#afin#ua”v¡t#-£fl#¡·~infi#£s# ·¡fl·jjnflfls·}#«a‰~n¡#flı

˚ı#}fl£#»vu#}£a¡#«·fi«£fln#Pn«avið#V#ua”v¡t#uvtunfi#¡·~infi#£s#¡fl·jjnflfls·}#«a‰~n¡#fl#£¡# ‡v~nı

¸ı#3n#afl#-·ju#afl#j}n¡#fl#£fi£~#u£·flv¡t#‰«n#Z·vu#«afin¡#flð#afl#un‰#afin#ua”v¡t#}nafl# ¡·~infi#£s#¡fl·jjnflfls·}#«a‰~n¡#flı