



CREDIT EDA CASE STUDY

By 

SAMIR KHAN

BUSINESS PROBLEM

- Use EDA to analyse the patterns present in the data.

This will ensure that the applicants capable of repaying the loan are not rejected.

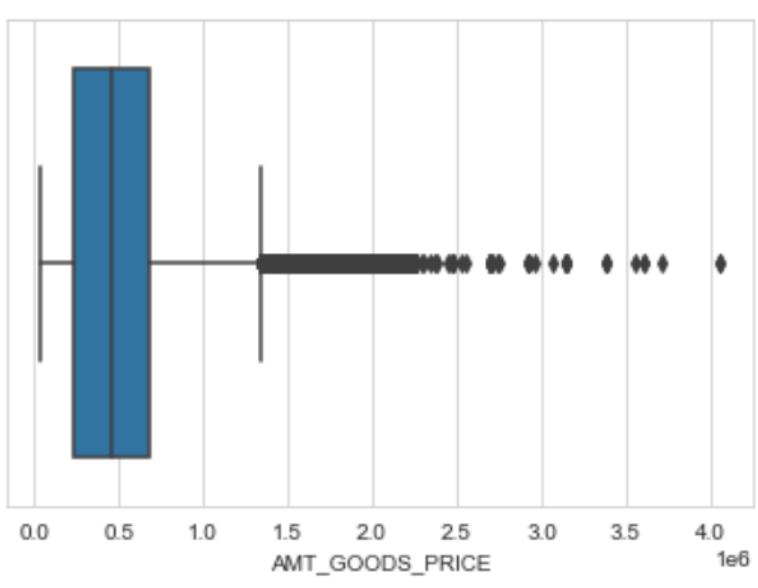
- When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

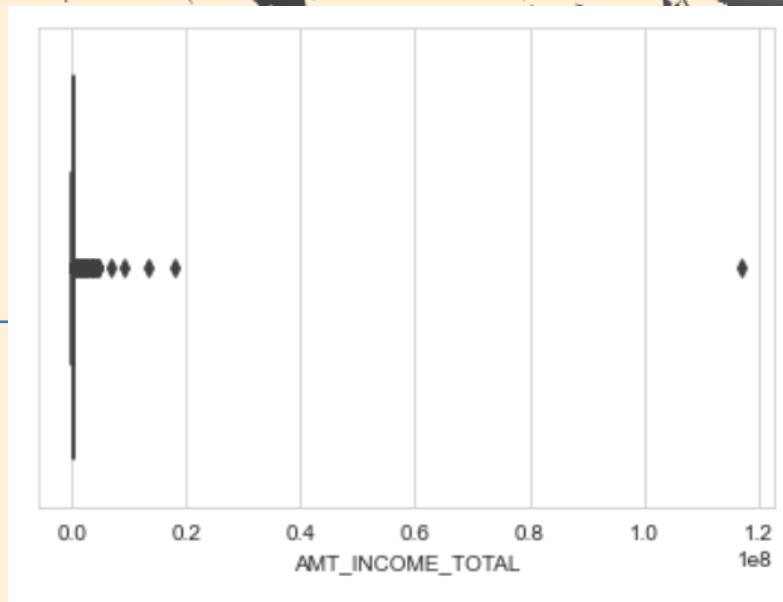
- If the applicant is not likely to repay the loan, then approving the loan may lead to a financial loss for the company.



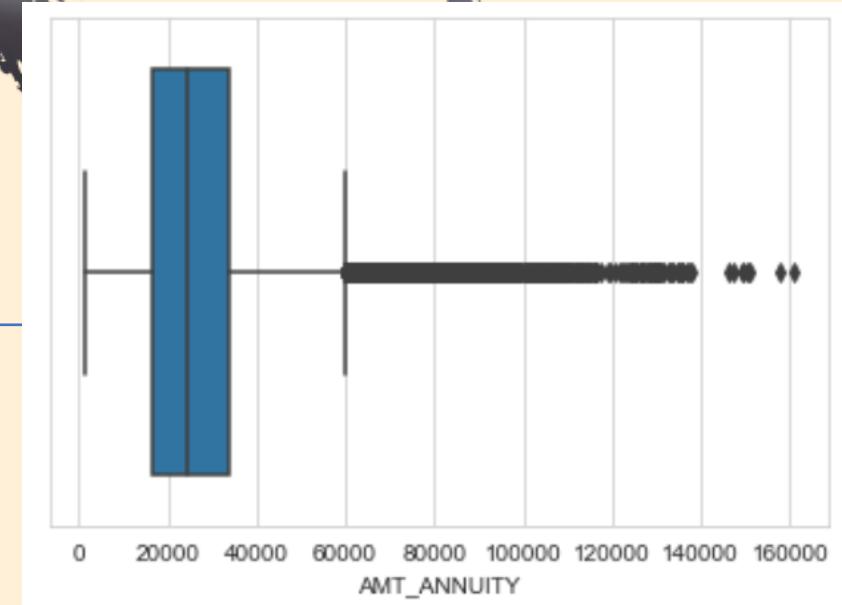
Outliers Found in Various Columns:



AMT_GOODS_PRICE



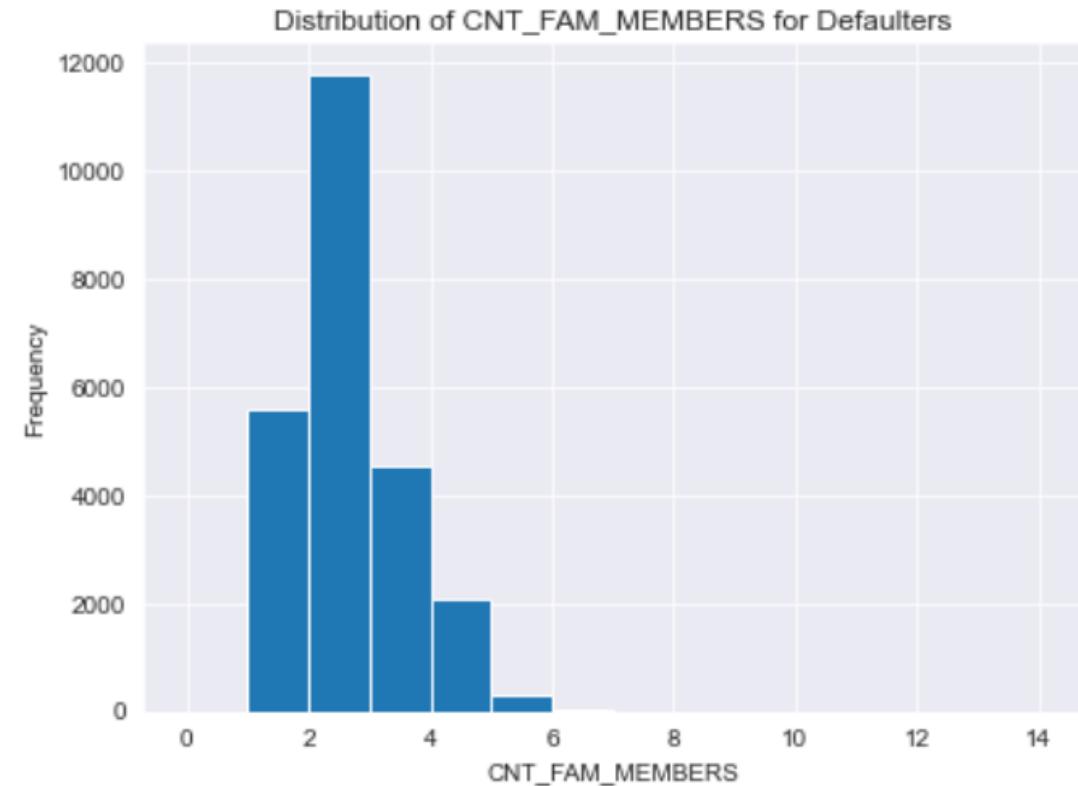
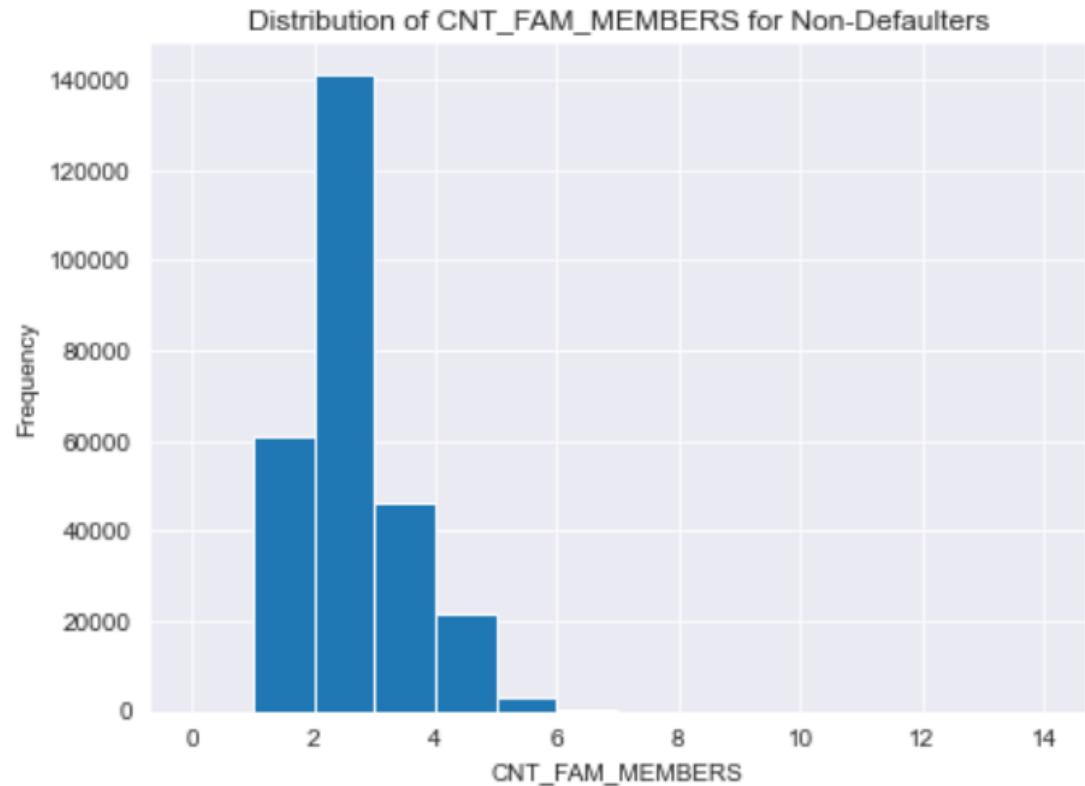
AMT_INCOME_TOTAL



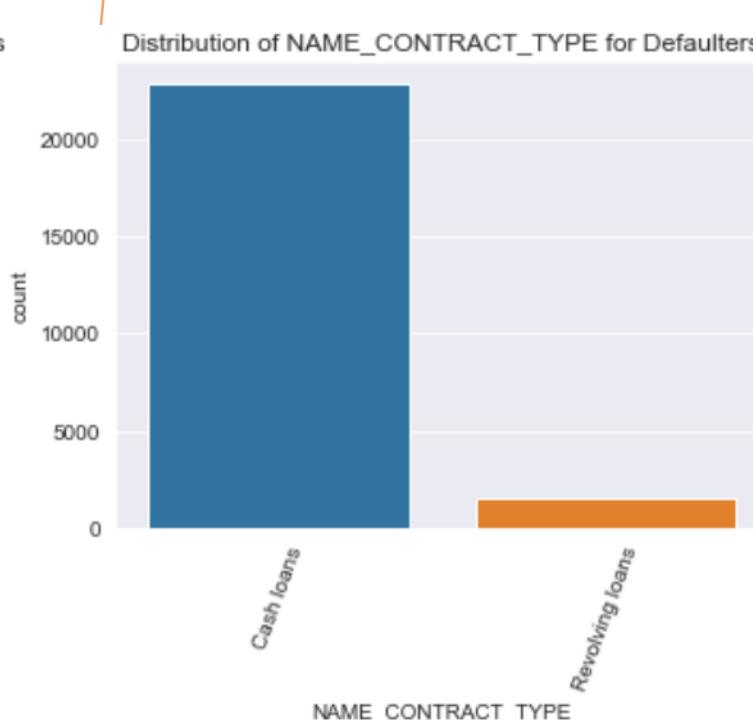
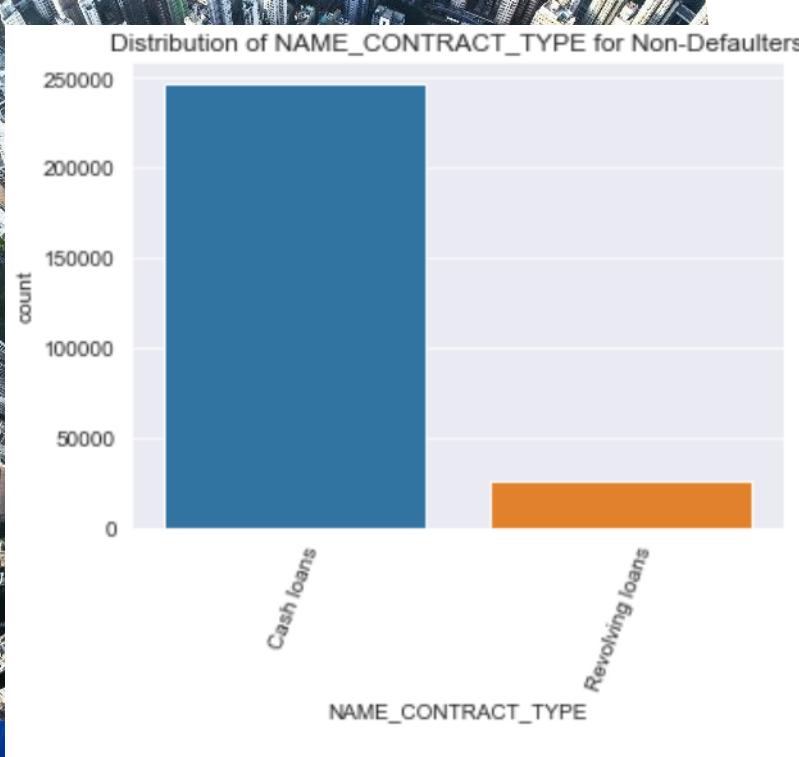
AMT_ANNUITY



Univariate Analysis

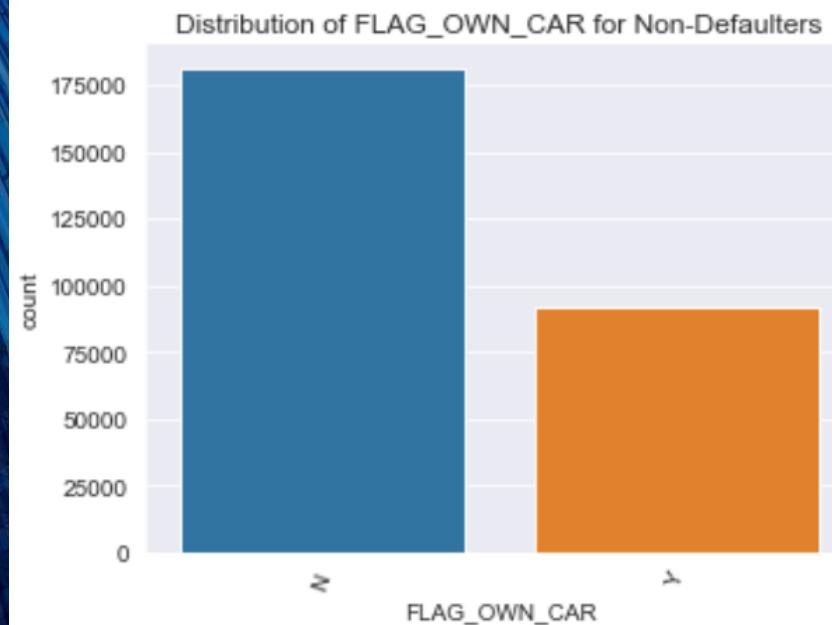


We can see that people with family member count of 3 take more loans.

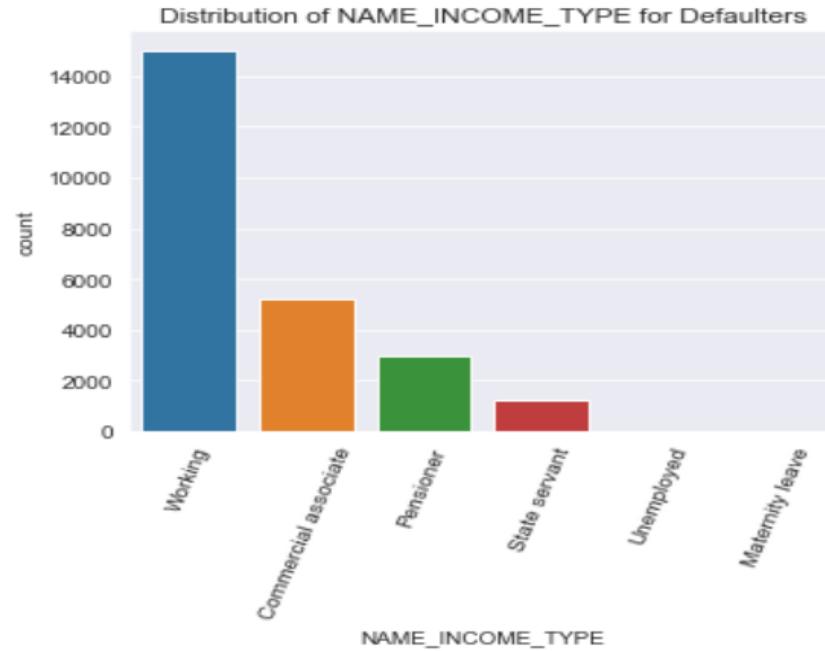
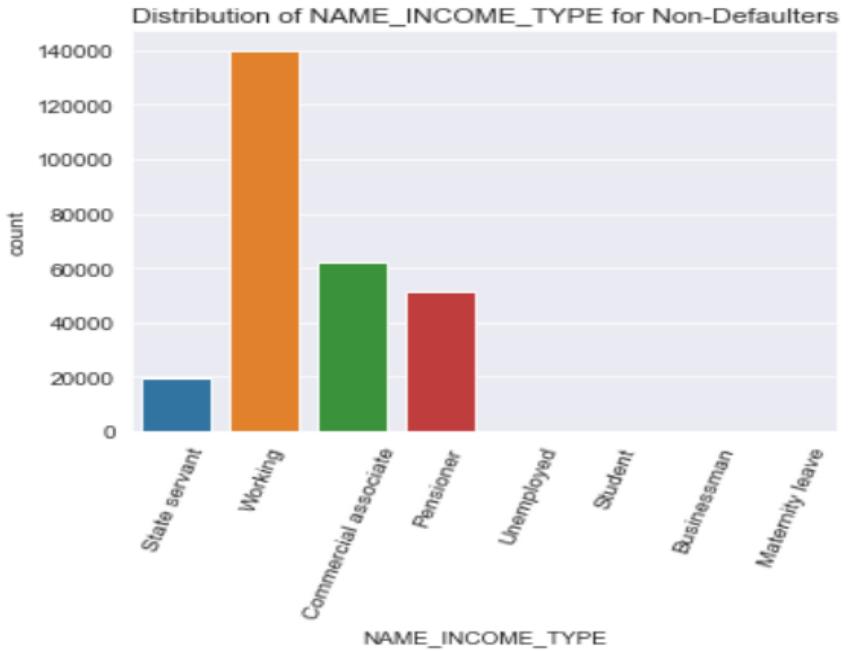


As there are more female who takes the loan, so their number in both the default and non-default have a similar trend. But the males constitute a higher proportion in the default section.

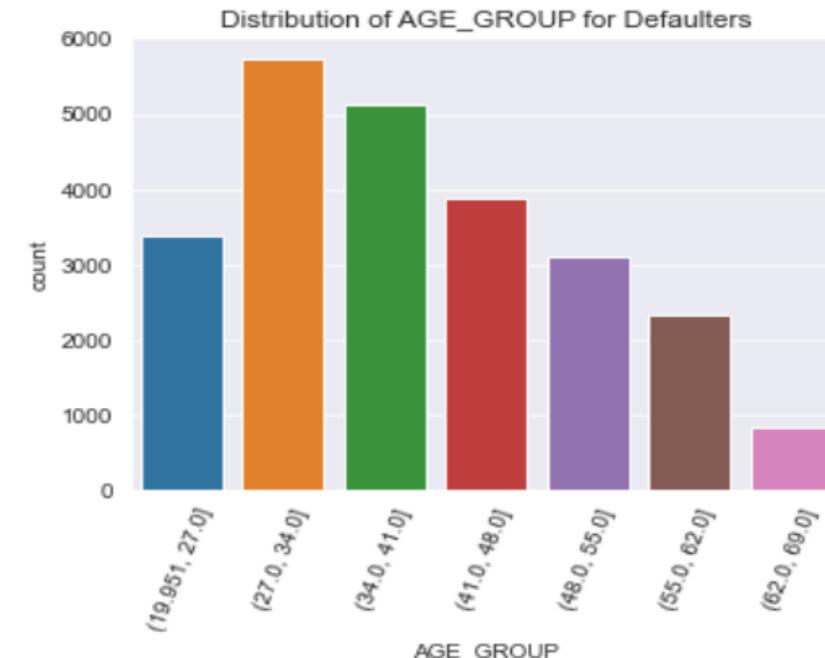
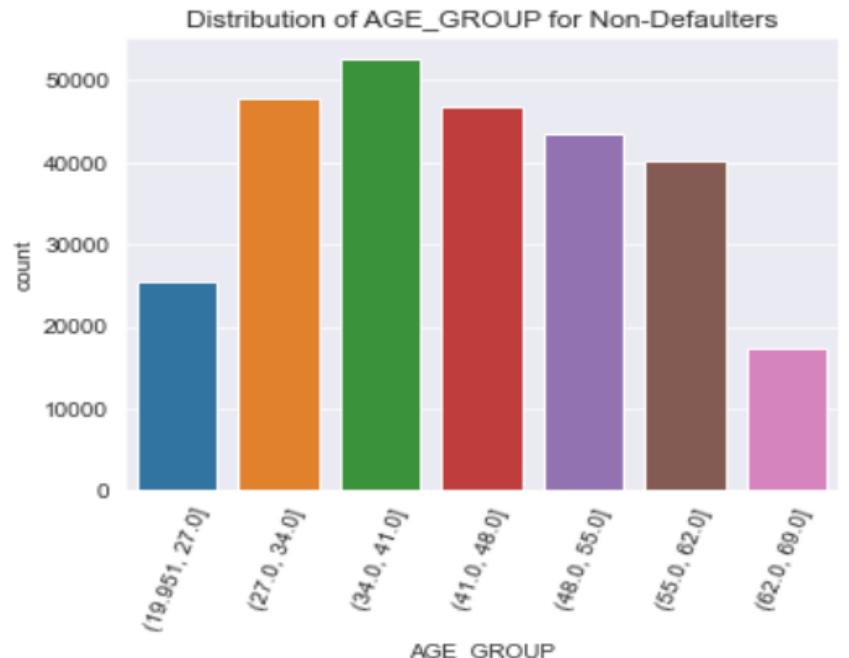
People tend to take cash loans much more than the revolving loans as sometimes the interest on the revolving loan is higher and the approved amount is lower than the traditional loans. And the revolving loan have a little higher proportion of non-defaulters.

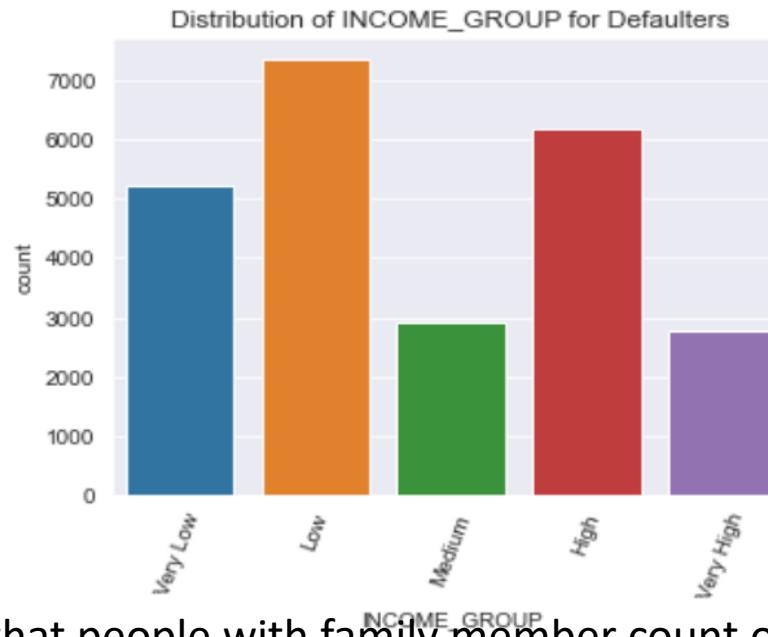
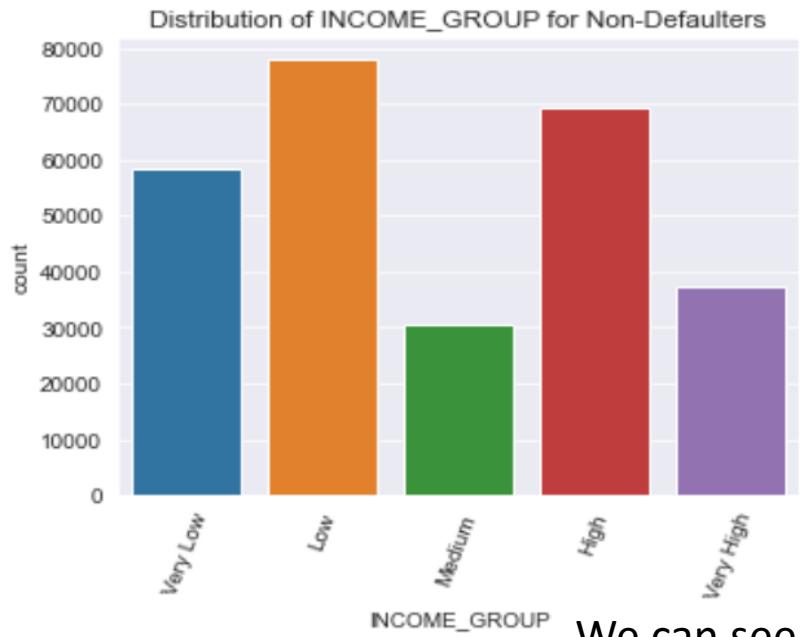


**Students and Businessmen
don't default at all. and
Working class |
professionals use to take
more loans comparatively
to other class of the
population.**

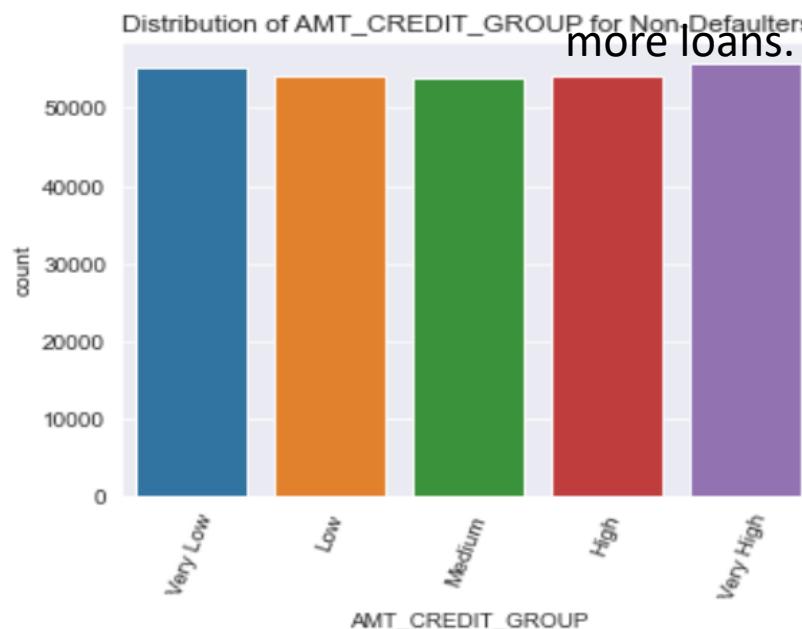


**As the age increases the
intake of loan decreases
and so accordingly the
non-defaulters also
increases from the age of
27. It might happen due to
the fact that, as age
increases, people's salary
increases and they are not
much dependant on loans.**

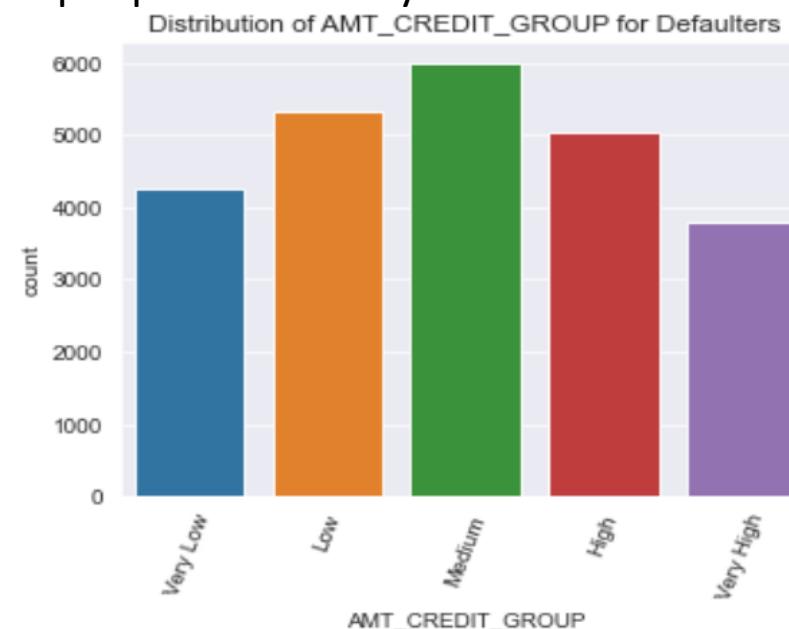




Medium income group tends to take less loans than other groups as the low tier groups might need the money for their some basic needs and the higher income group might need the loan amount for some of their other lavish needs.

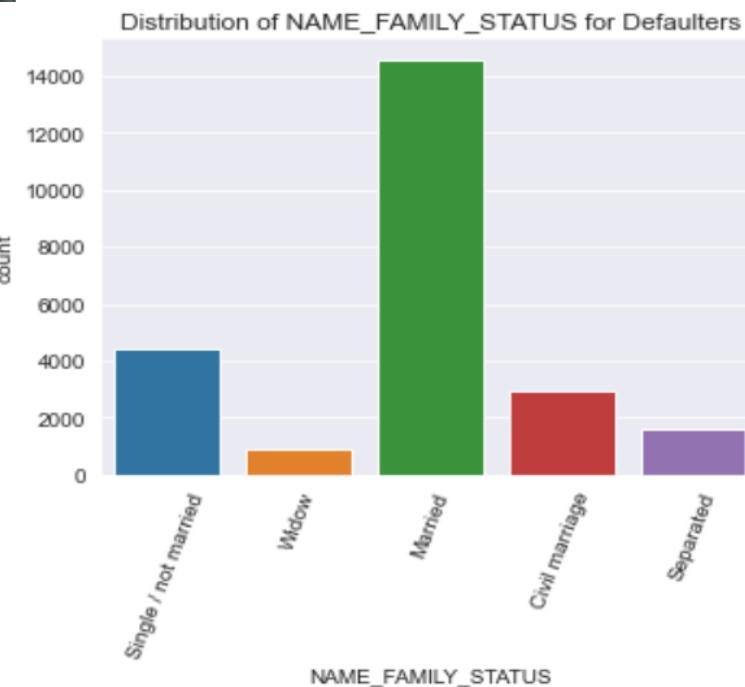
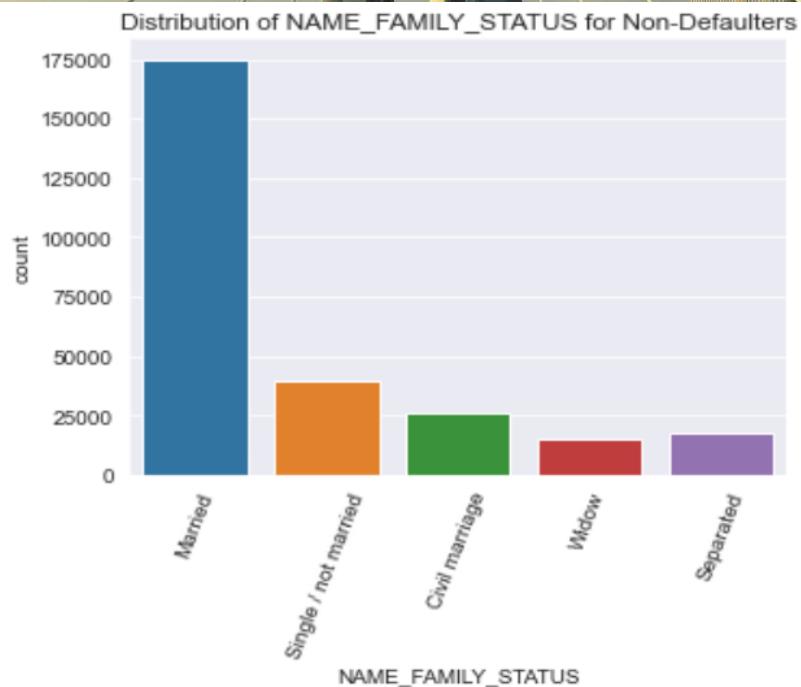
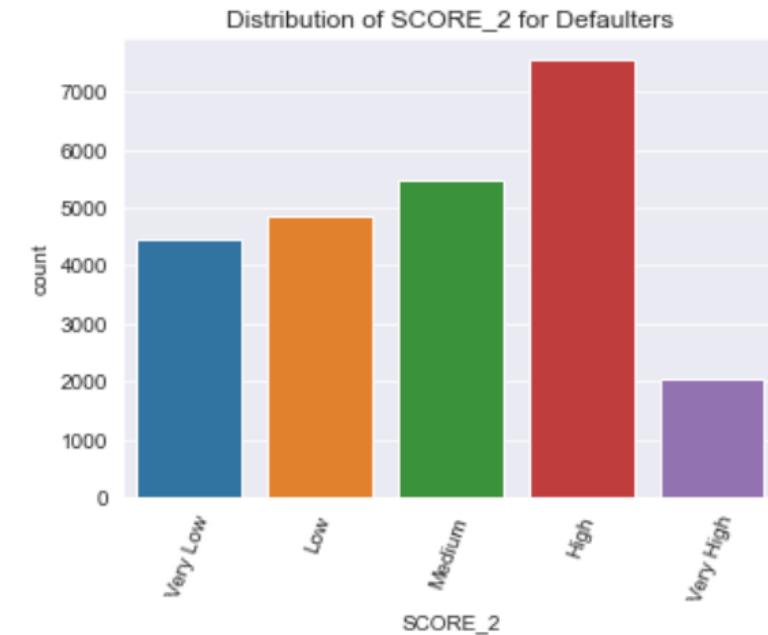
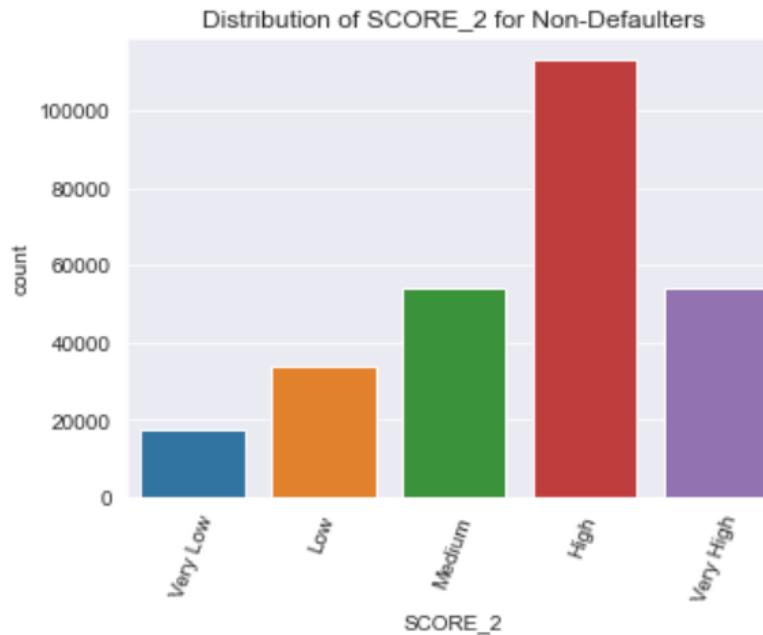


We can see that people with family member count of 3 take more loans.

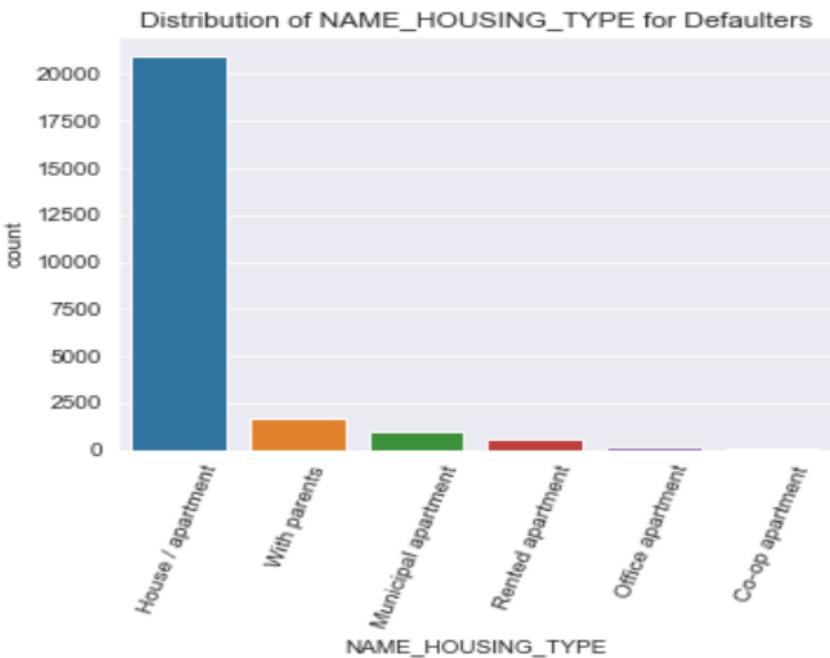
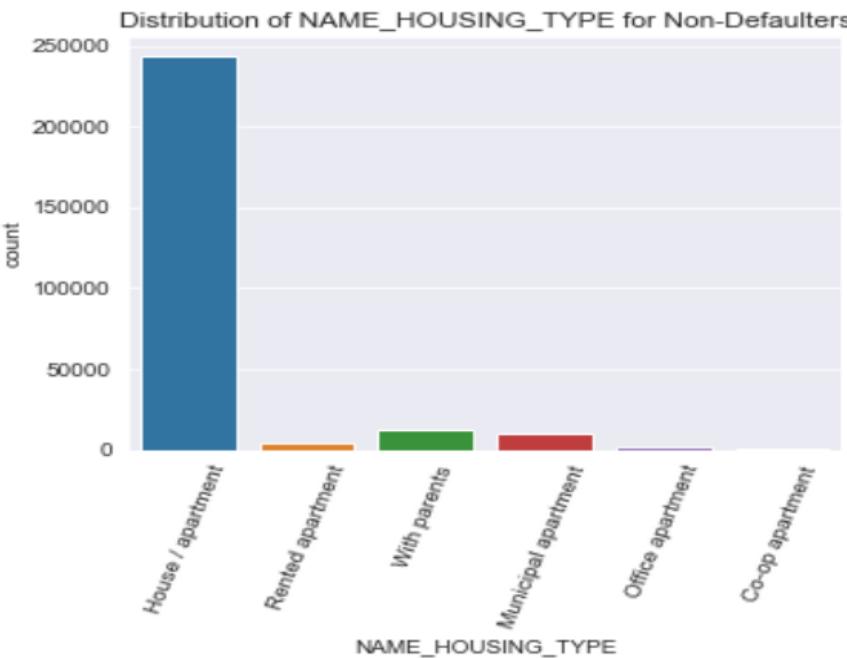


The Defaulter have a bell curve type structure and the medium credit group default very much often than the others.

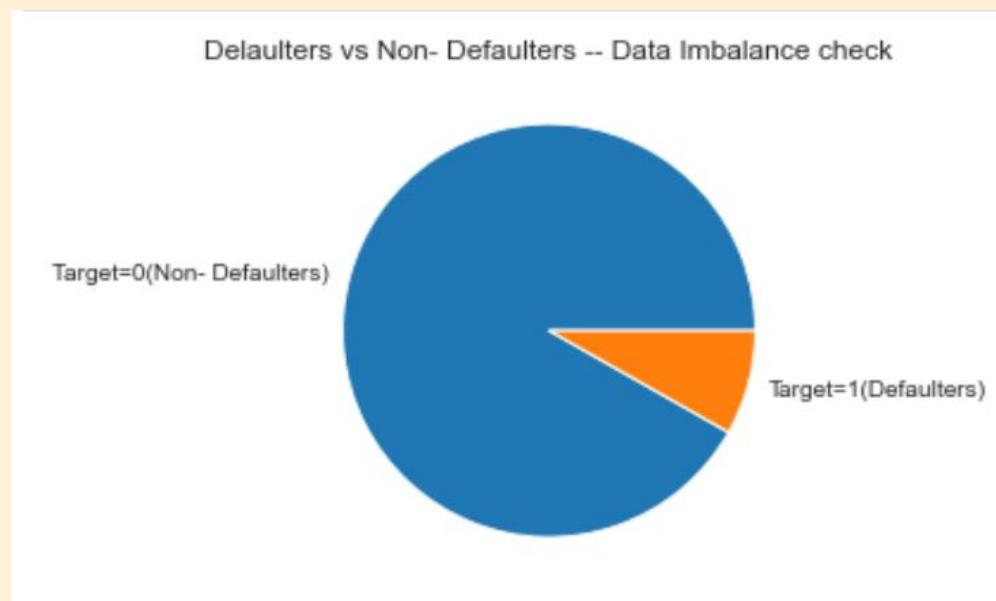
With the increase in the external score we can see there is a rise in the non defaulters. And there is a dip in the number of loan taken by the very high group from the high group.



The married people tend to take more loans than the others. Civil Marriage and Single people default little more in proportion. Widow people have lesser chance of default.

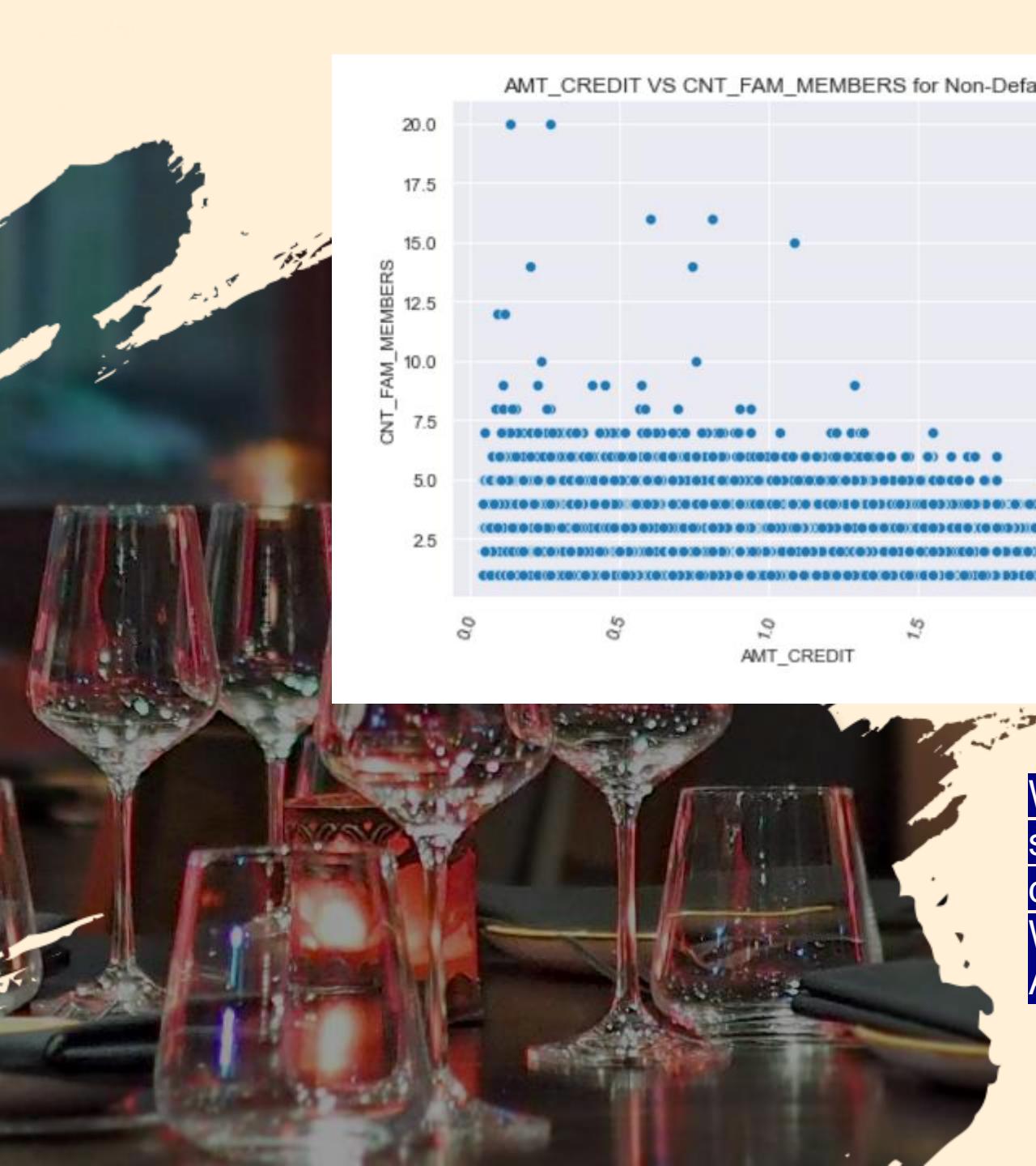


We can see that people with house or apartment takes more loans than other. People with parents default a little more, might be due to the extra living cost living with them.

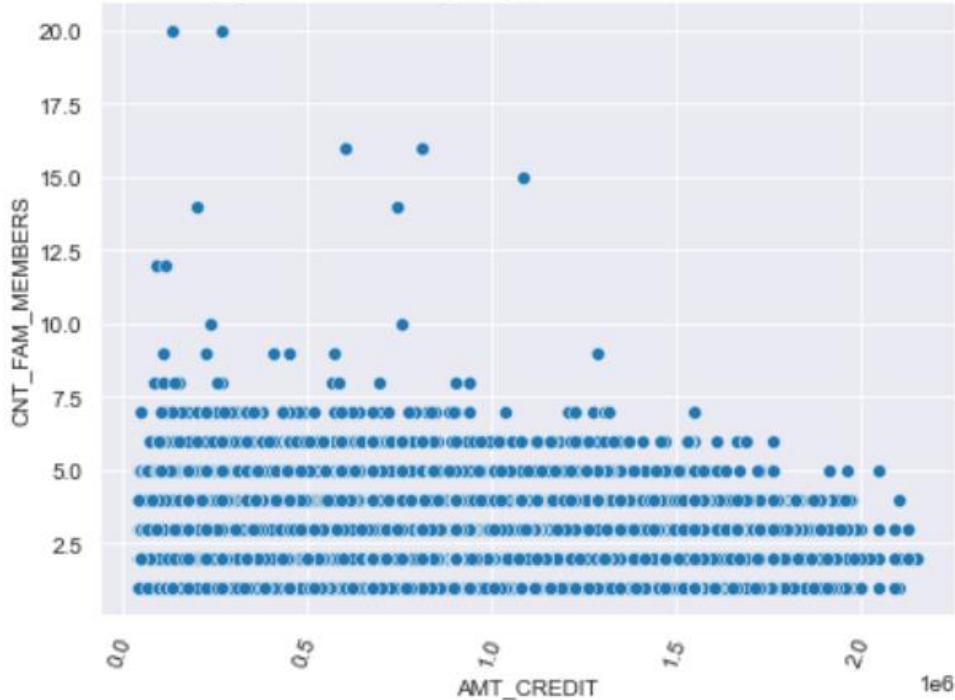




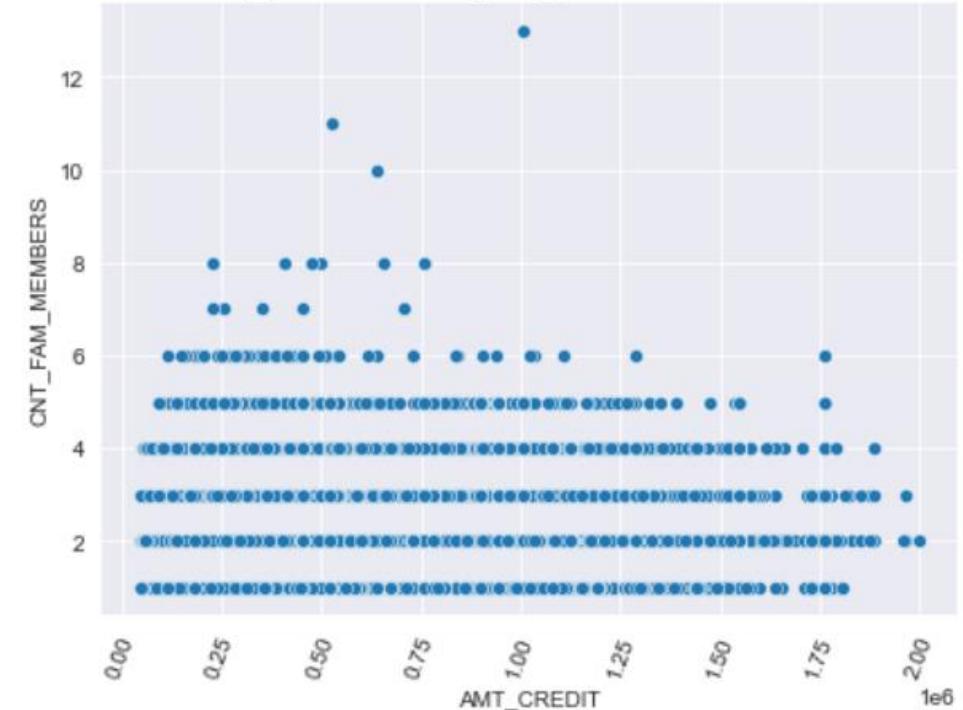
Bivariate Analysis



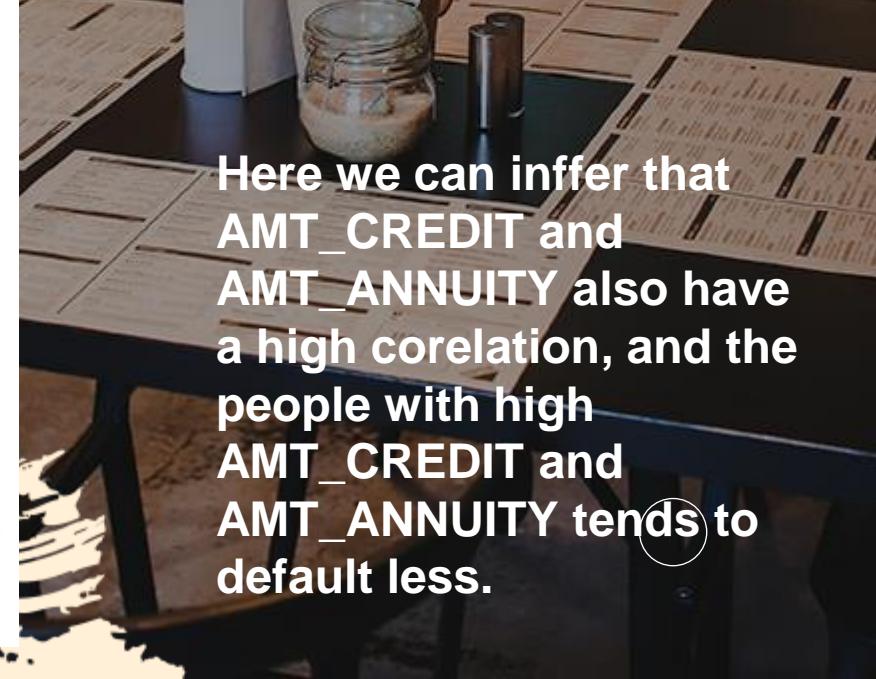
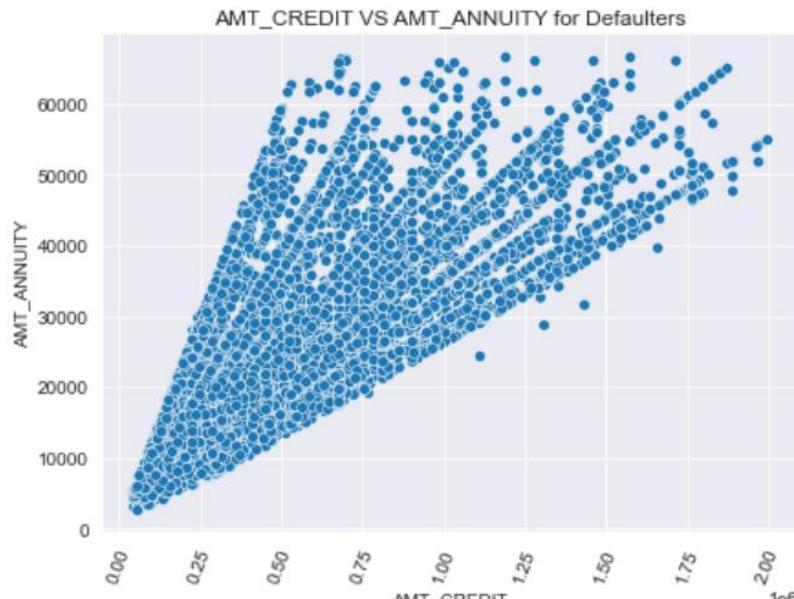
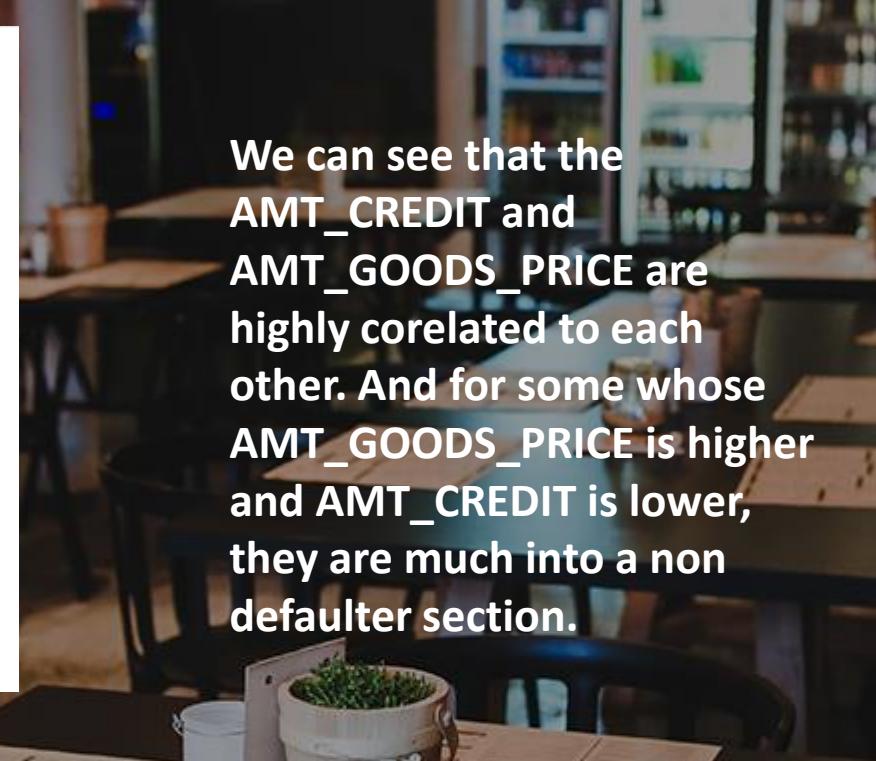
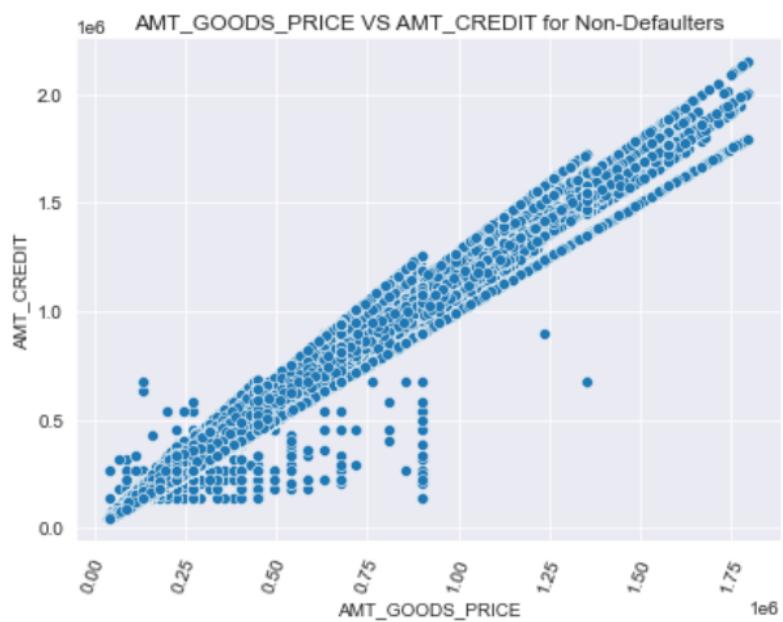
AMT_CREDIT VS CNT_FAM_MEMBERS for Non-Defaulters



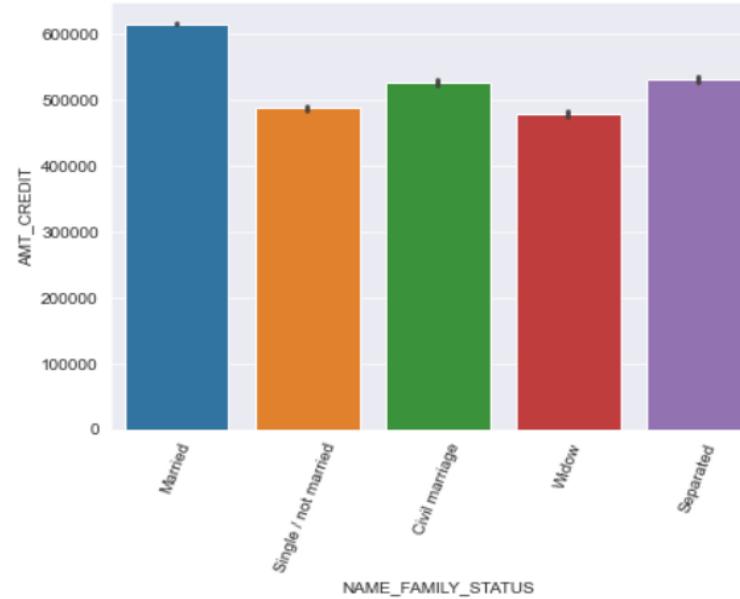
AMT_CREDIT VS CNT_FAM_MEMBERS for Defaulters



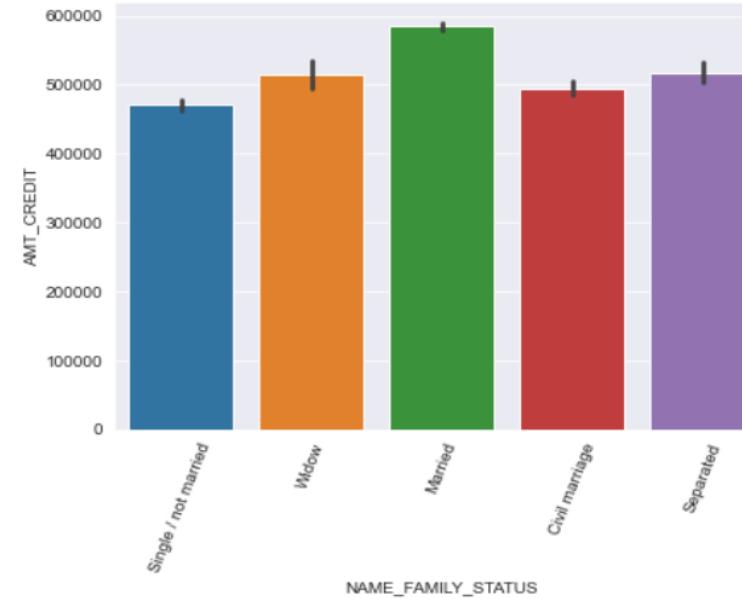
We can see that the density in the lower left corner is similar in both the case, so the people are equally likely to default if the family is small and the AMT_CREDIT is low. We can observe that larger families and people with larger AMT_CREDIT default less often.



NAME_FAMILY_STATUS VS AMT_CREDIT for Non-Defaulters

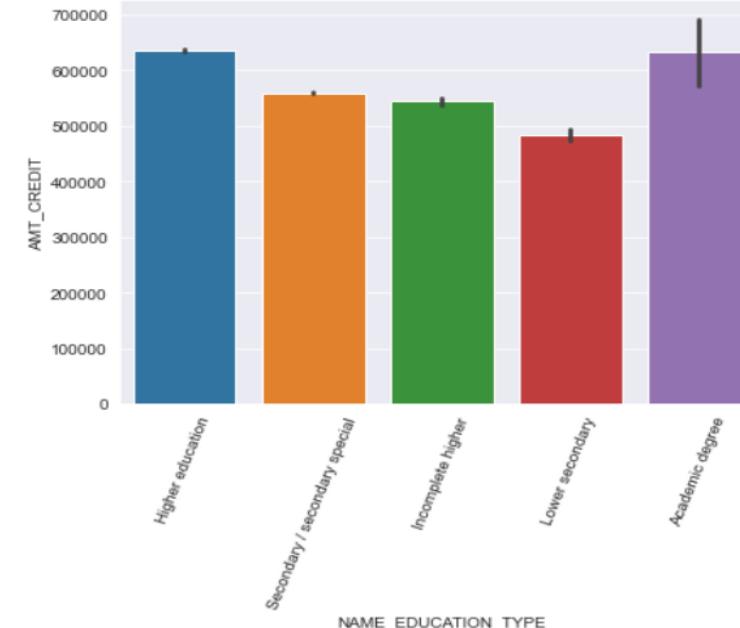


NAME_FAMILY_STATUS VS AMT_CREDIT for Defaulters

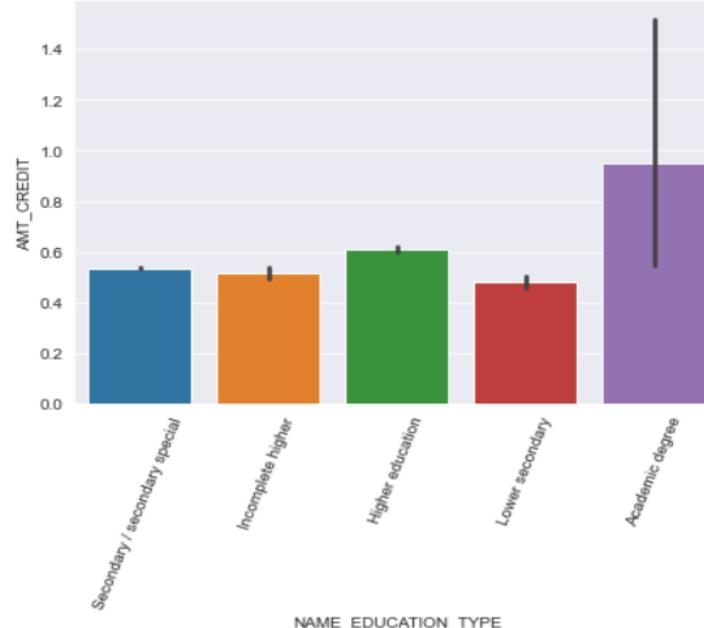


As we have seen before the married tends to high number of loans so they also take a higher amount of loan credit. And even if the Widow default less they take high amount of loan credit. And for Civil married and single people as they default more often it might be because they have a little higher loan credit amount.

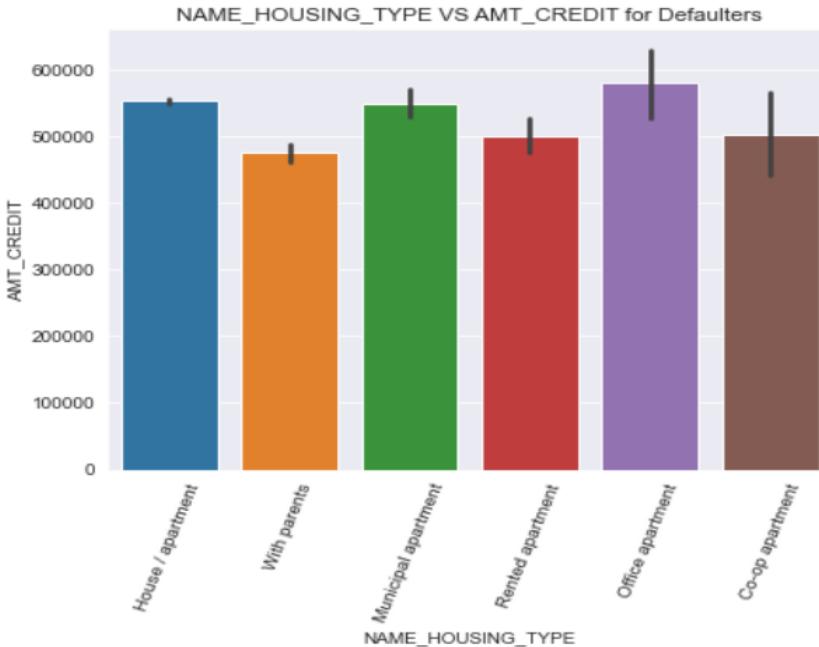
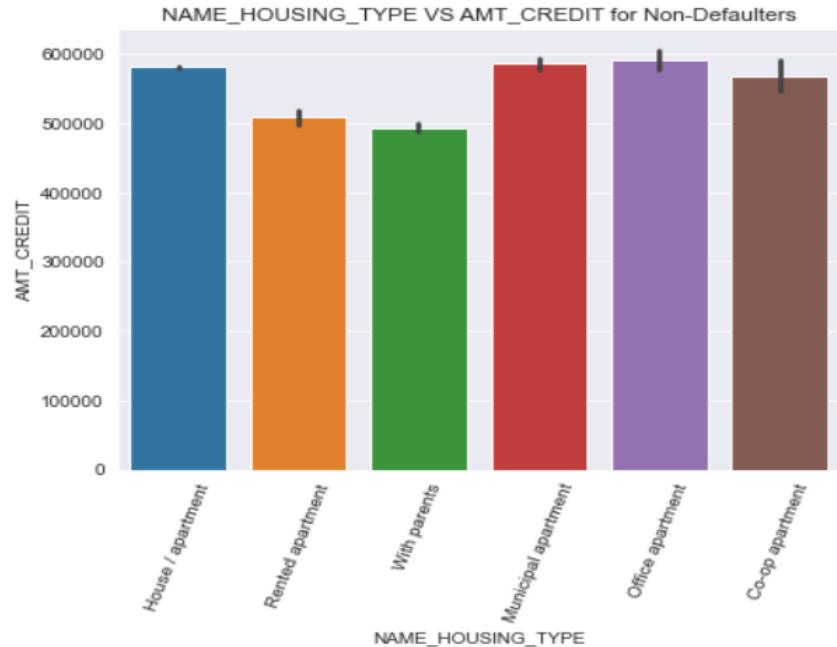
NAME_EDUCATION_TYPE VS AMT_CREDIT for Non-Defaulters



NAME_EDUCATION_TYPE VS AMT_CREDIT for Defaulters

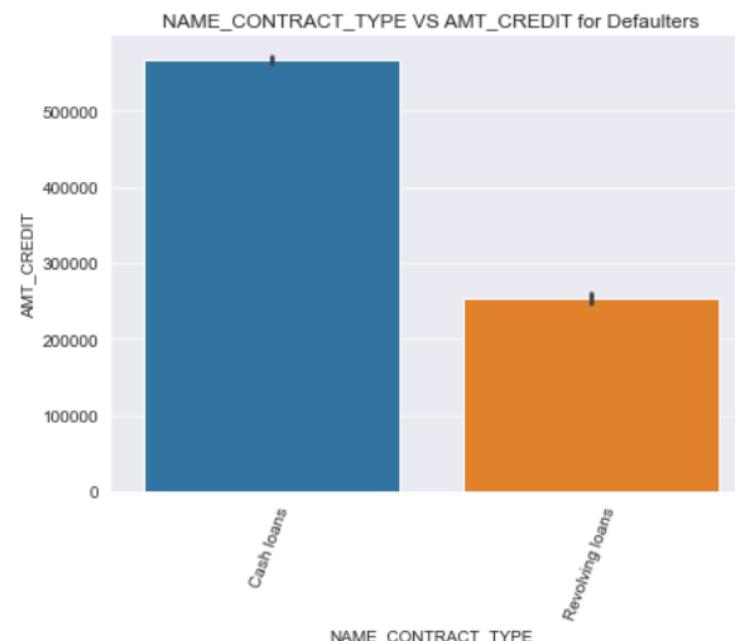
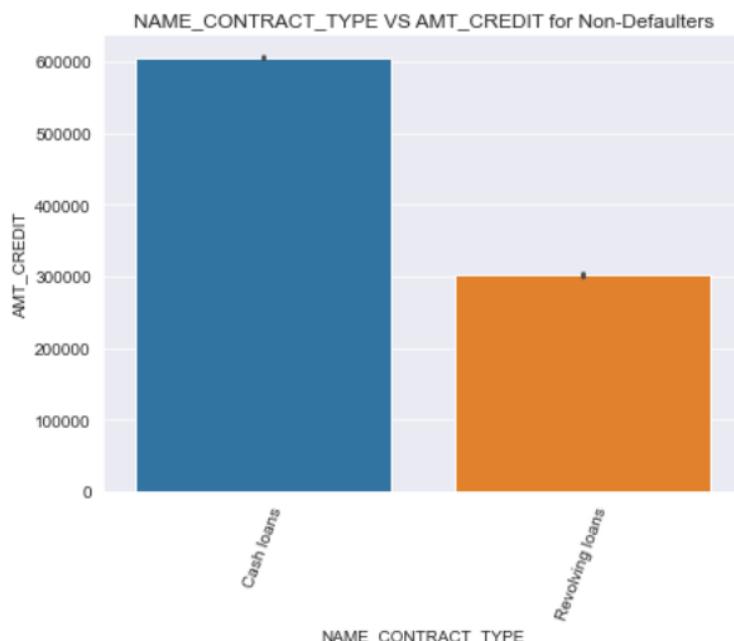


We can see over here that the lower amount credited by people under any education type is defaulting. And even if the higher education people take higher amount of credit they default less (from our previous inference).



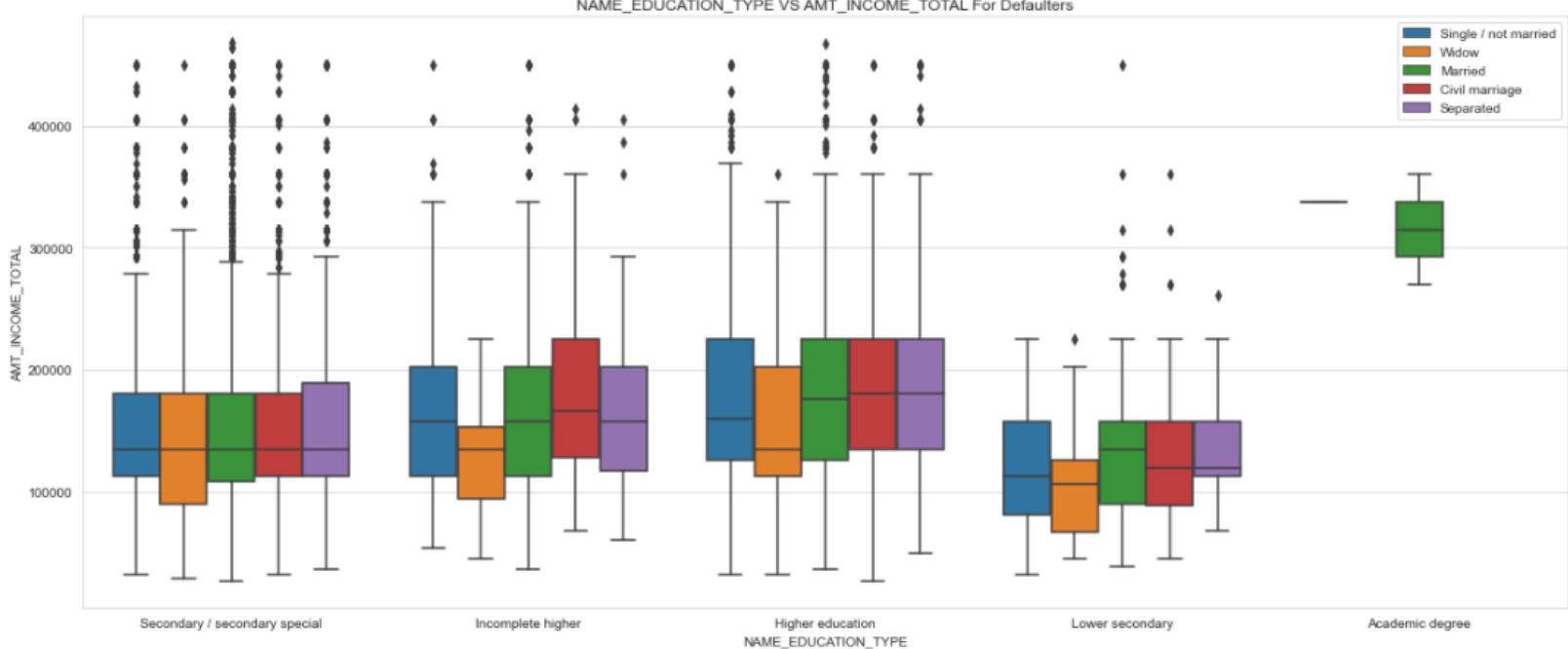
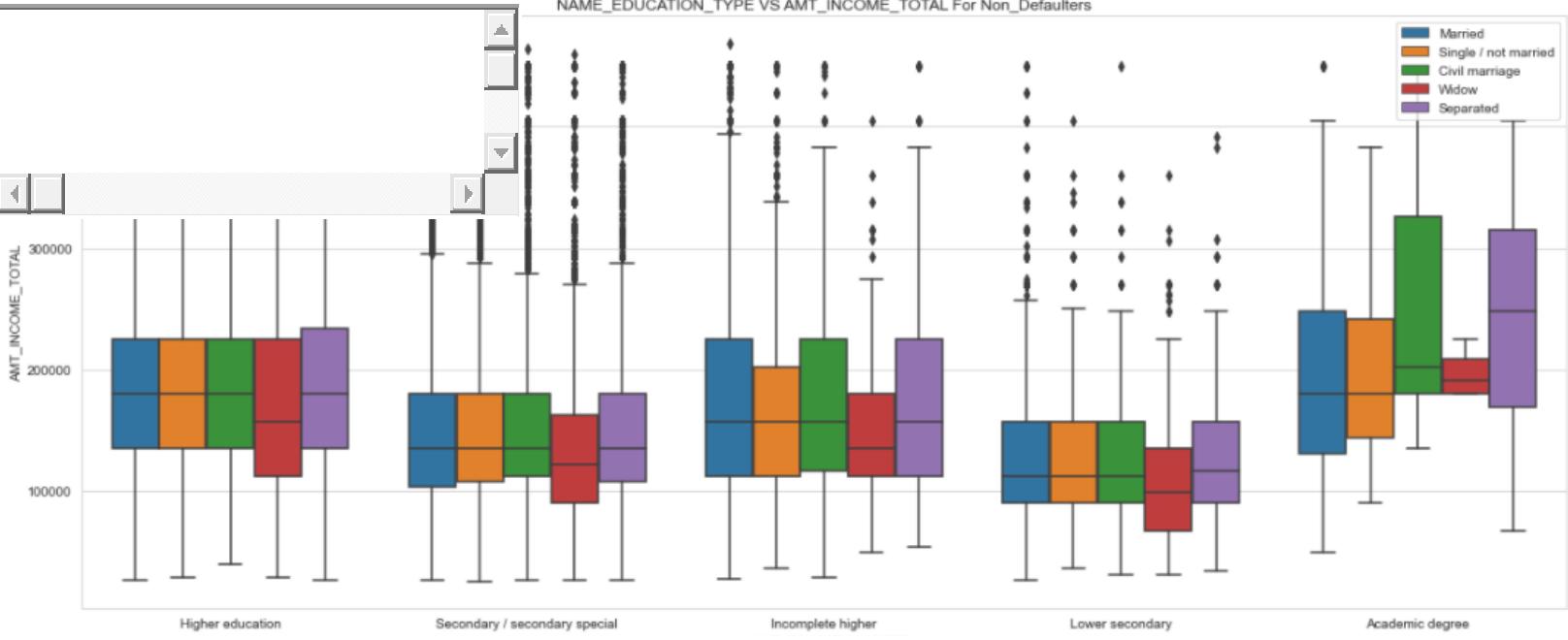
Population in Office apartment takes high credit amount as loan and here also the higher credit amount population default less than the comparatively lower credit amount population in each housing category.

As we know, in the revolving loans the amount approved for the loan or the amount credit is less. So from the previous and the plot above we can see that the revolving loan has a little lesser proportion of non-defaulters.



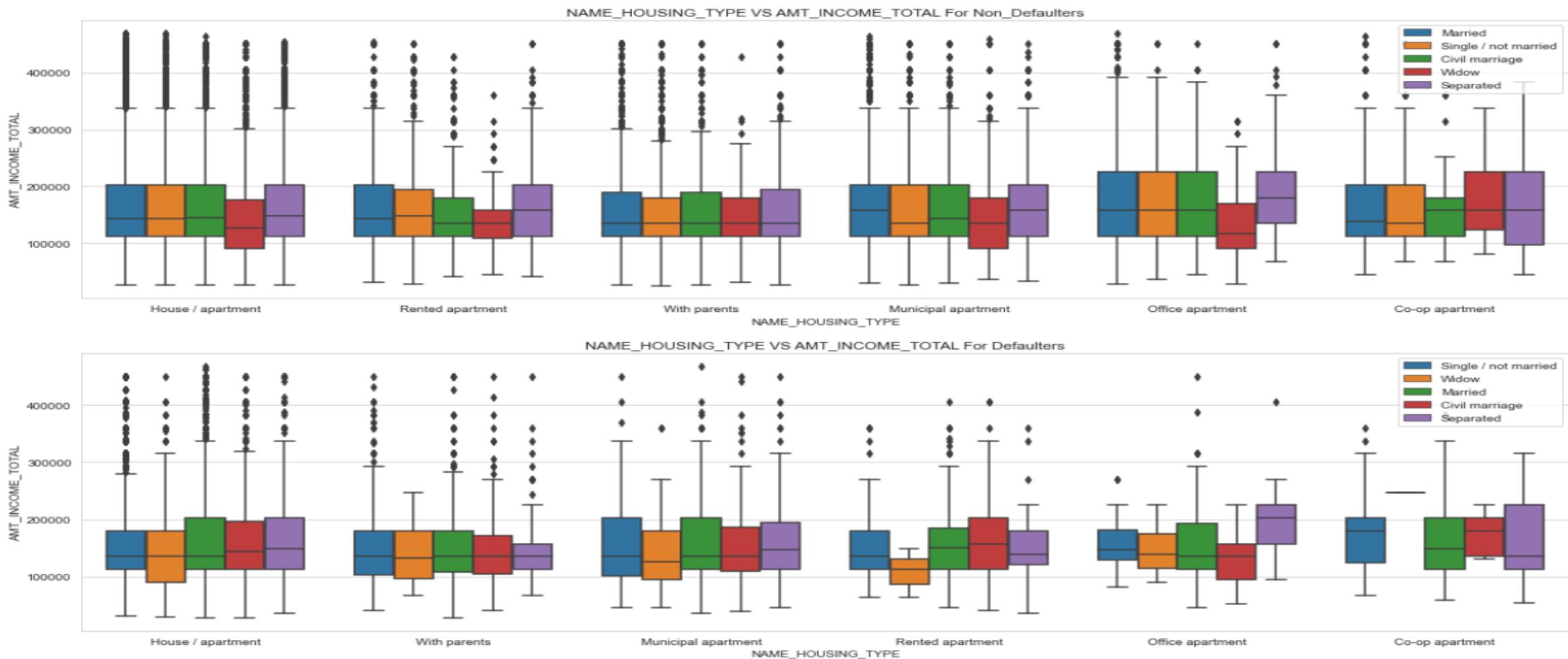


Multivariate Analysis

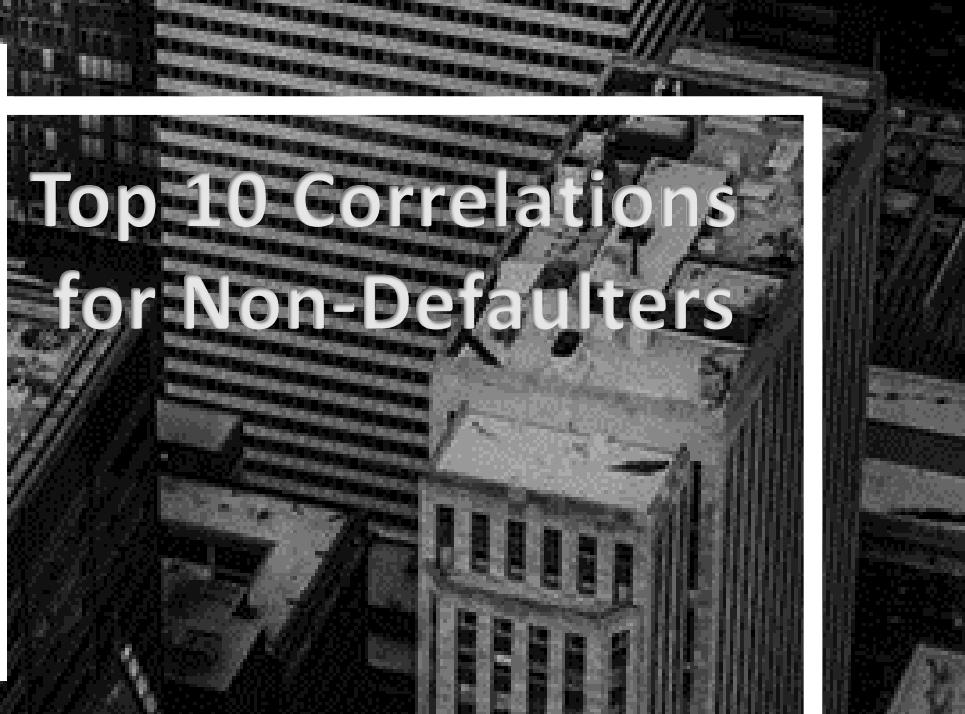


- As we seen before the Academic Degree population default less, so we got to know here that the married population in the academic degree population is the only section that defaults and the widows to a very negligible amount.
- We now confirmed here that the lower secondary people having less income use to default more. And the Widows not defaulting in the lower secondary section have a lesser median than the others in the group.
- The people having academic degree seem to be the perfect group for the loan as they do not default much.

- We can see that the high total income people irrespective of housing type and family status do not default much. And the median income of Widow who are non defaulting is on a lower side.
- Again Married population have a higher income range and widow population have a low income range and both of the combinations default less.
- We also can see the Office apartment section having a high annual income range in the non-defaulting section. Widows staying at a co-op apartment do not default at all.



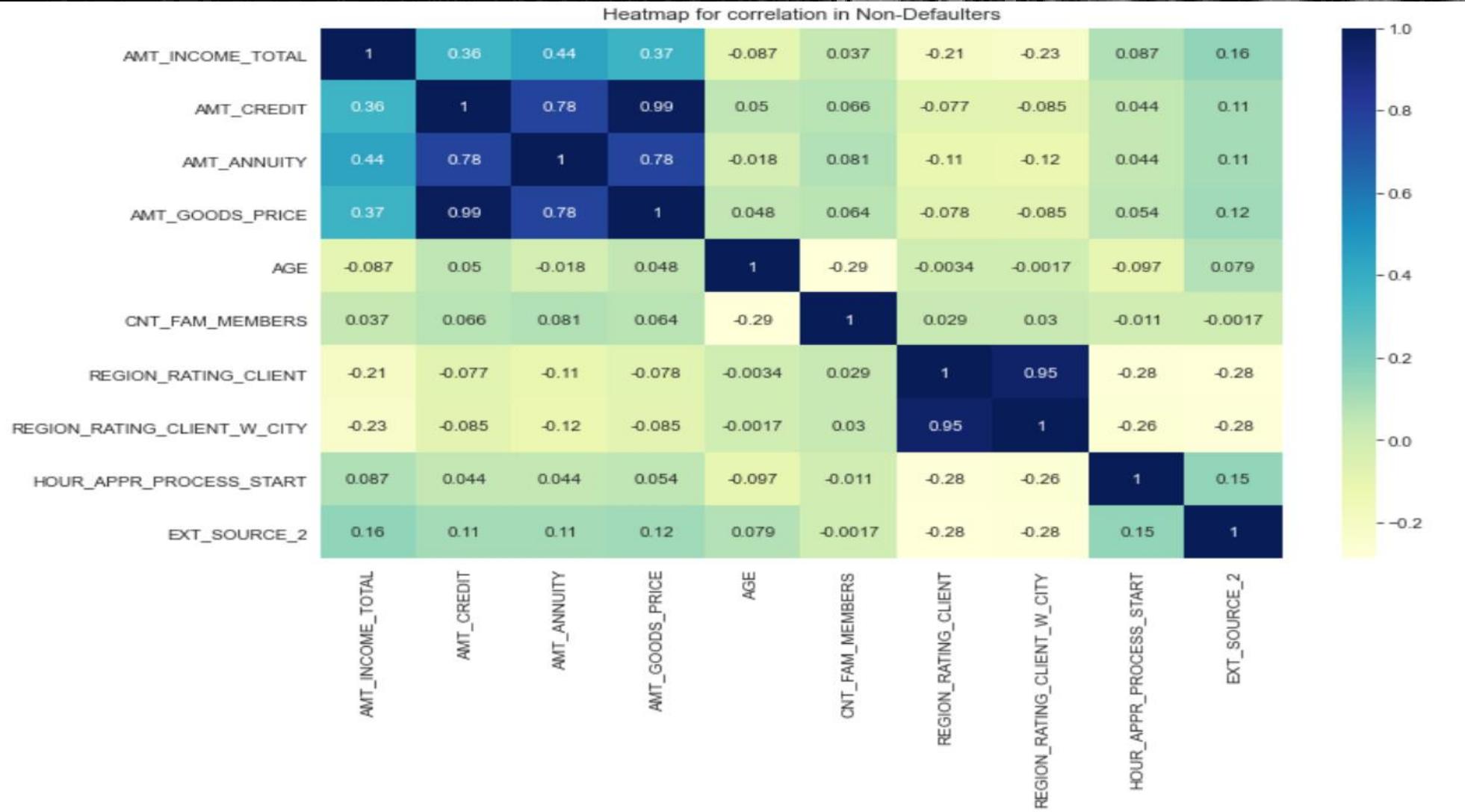
	VAR1	VAR2	Correlation	Abs_Correlation
31	AMT_GOODS_PRICE	AMT_CREDIT	0.981344	0.981344
76	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.957161	0.957161
21	AMT_ANNUITY	AMT_CREDIT	0.748863	0.748863
32	AMT_GOODS_PRICE	AMT_ANNUITY	0.746733	0.746733
20	AMT_ANNUITY	AMT_INCOME_TOTAL	0.406451	0.406451
30	AMT_GOODS_PRICE	AMT_INCOME_TOTAL	0.330146	0.330146
10	AMT_CREDIT	AMT_INCOME_TOTAL	0.329047	0.329047
86	HOUR_APPR_PROCESS_START	REGION_RATING_CLIENT	-0.294311	0.294311
87	HOUR_APPR_PROCESS_START	REGION_RATING_CLIENT_W_CITY	-0.276468	0.276468
96	EXT_SOURCE_2	REGION_RATING_CLIENT	-0.245854	0.245854



	VAR1	VAR2	Correlation	Abs_Correlation
31	AMT_GOODS_PRICE	AMT_CREDIT	0.985211	0.985211
76	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.950049	0.950049
21	AMT_ANNUITY	AMT_CREDIT	0.777327	0.777327
32	AMT_GOODS_PRICE	AMT_ANNUITY	0.777259	0.777259
20	AMT_ANNUITY	AMT_INCOME_TOTAL	0.440189	0.440189
30	AMT_GOODS_PRICE	AMT_INCOME_TOTAL	0.367429	0.367429
10	AMT_CREDIT	AMT_INCOME_TOTAL	0.363993	0.363993
54	CNT_FAM_MEMBERS	AGE	-0.287635	0.287635
96	EXT_SOURCE_2	REGION_RATING_CLIENT	-0.284568	0.284568
86	HOUR_APPR_PROCESS_START	REGION_RATING_CLIENT	-0.280732	0.280732

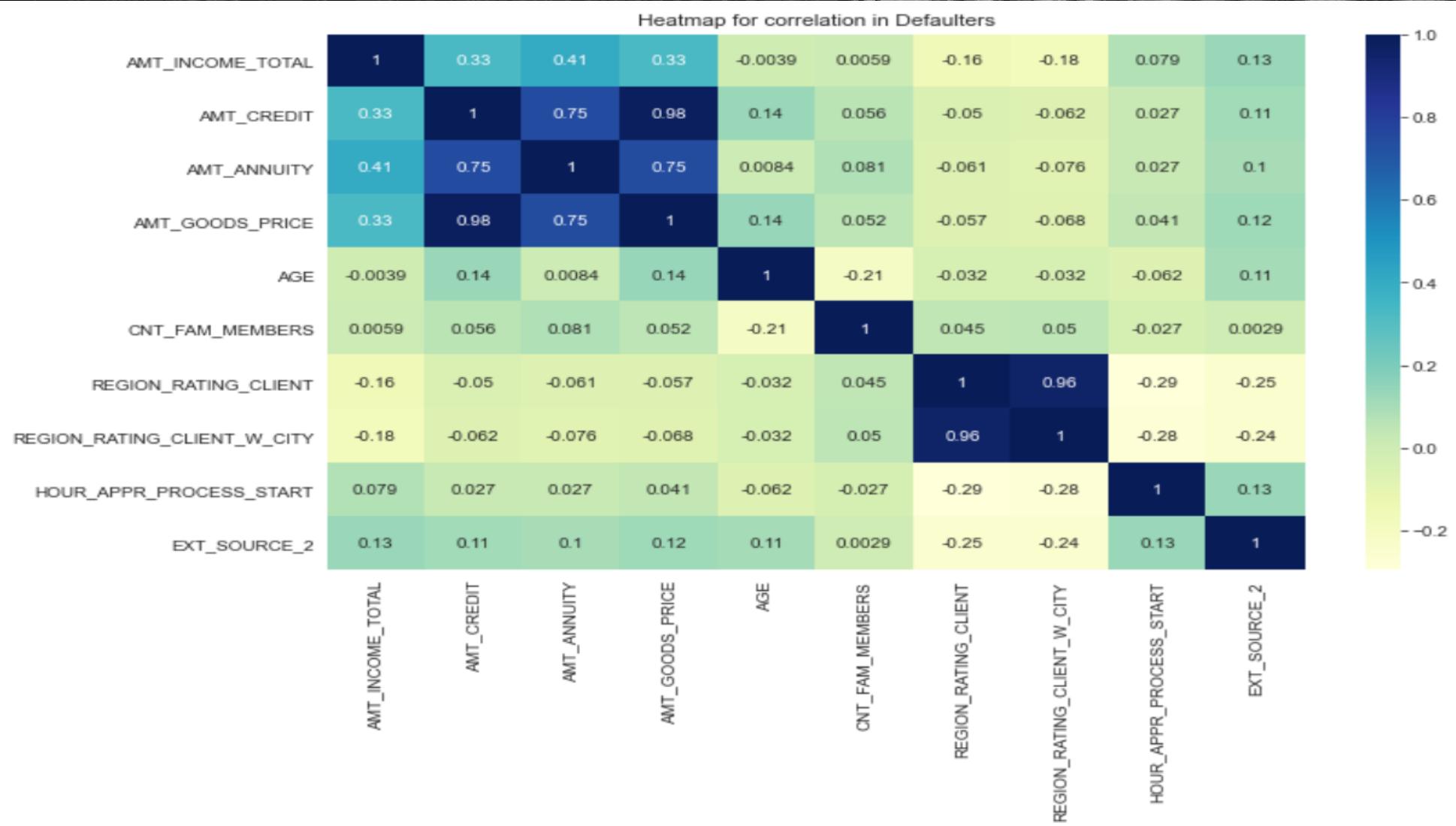
Heatmap for Non-Defaulters:

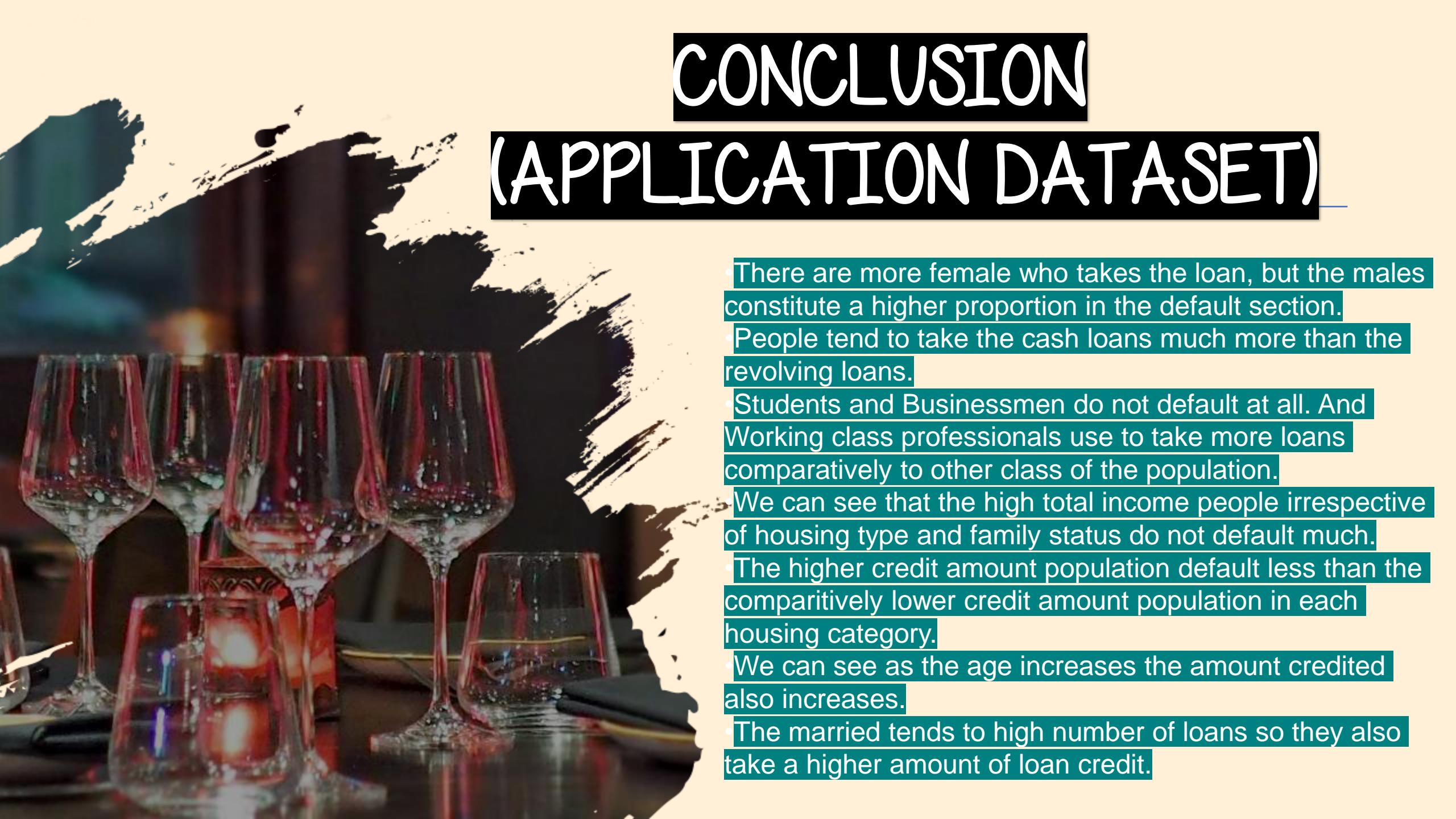
CORRELATION MATRIX



Heatmap for Defaulters:

CORRELATION MATRIX





CONCLUSION (APPLICATION DATASET)

- There are more female who takes the loan, but the males constitute a higher proportion in the default section.
- People tend to take the cash loans much more than the revolving loans.
- Students and Businessmen do not default at all. And Working class professionals use to take more loans comparatively to other class of the population.
- We can see that the high total income people irrespective of housing type and family status do not default much.
- The higher credit amount population default less than the comparatively lower credit amount population in each housing category.
- We can see as the age increases the amount credited also increases.
- The married tends to high number of loans so they also take a higher amount of loan credit.

CONCLUSION (PREVIOUS DATASET)

- We can infer that no Consumer loans got cancelled, the cash loans have similar kind of approval and cancel rate. Huge amount of consumer loans got approved.
- So we got to know that people prefer to pay through cash.
- There are very loyal customer to the bank who comes for the loan more number of times as the repeater number is very high.
- The loans with high credit amount got refused by the bank.
- The loans with high goods price was cancelled by the client at some point of time.





Thank You