CREDIT EDA : CASE STUDY

Presented By:

Niharika Girdhar

Nishant Kumar

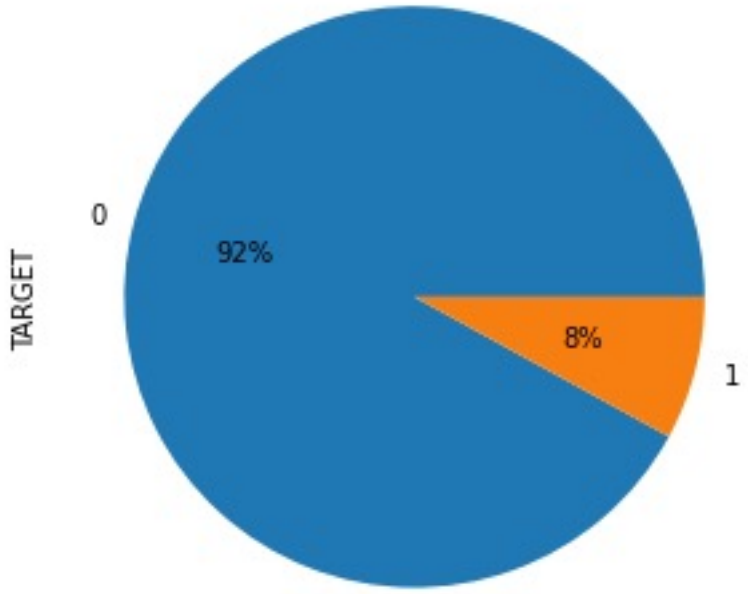
Problem Statement

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

1. If the applicant is likely to repay the loan, then not approving the loan will result in loss of bank or company.
2. If the applicant is not likely to repay the loan, he/she will default will result in financial loss to bank or a company.

TARGET VARIABLE

- Target = 0 means Customer is Non-Defaulter

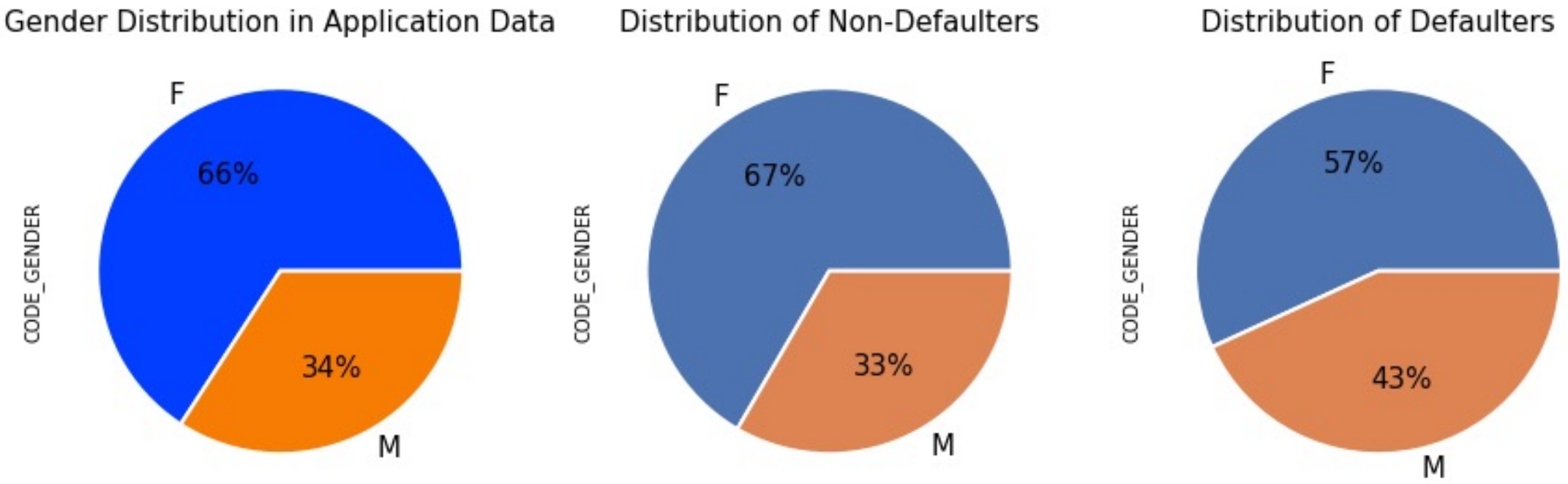


- Target = 1, means Customer is Defaulter

By this Pie chart : Its clearly visible that we have 92 % Non-Defaulters and 8% Defaulters

GENDER vs TARGET

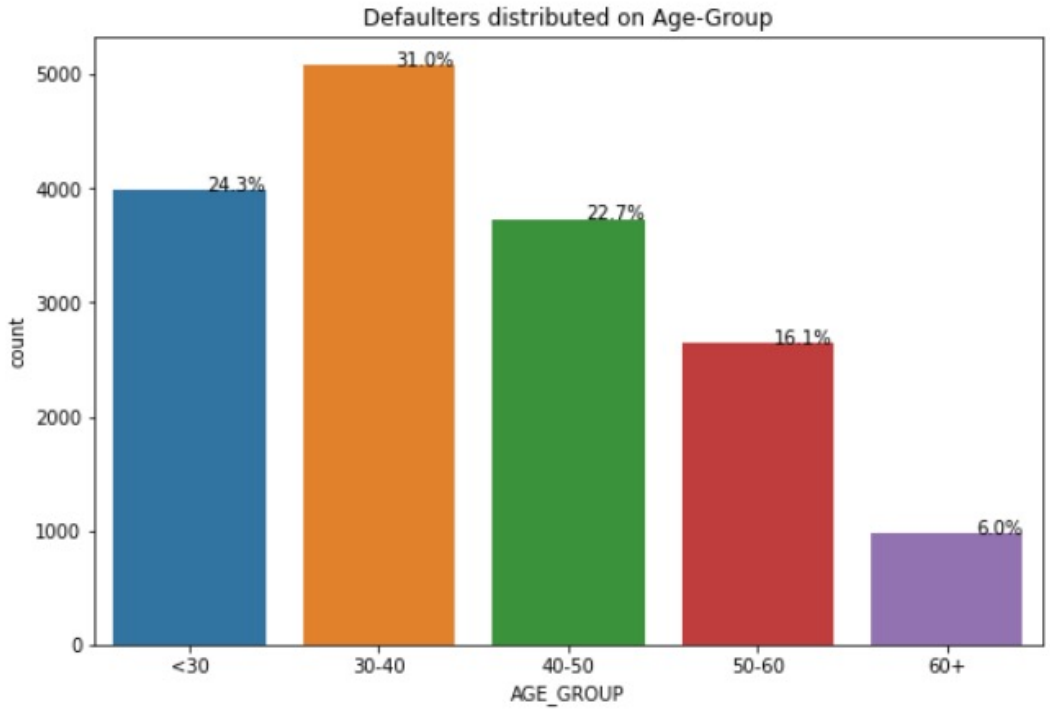
- Distribution of Non-Defaulters are almost same as original data frame.
- But Distribution of Defaulters are different. Percentage gap between genders is smaller. Only 1% difference is creating a big difference in case of Defaulters.

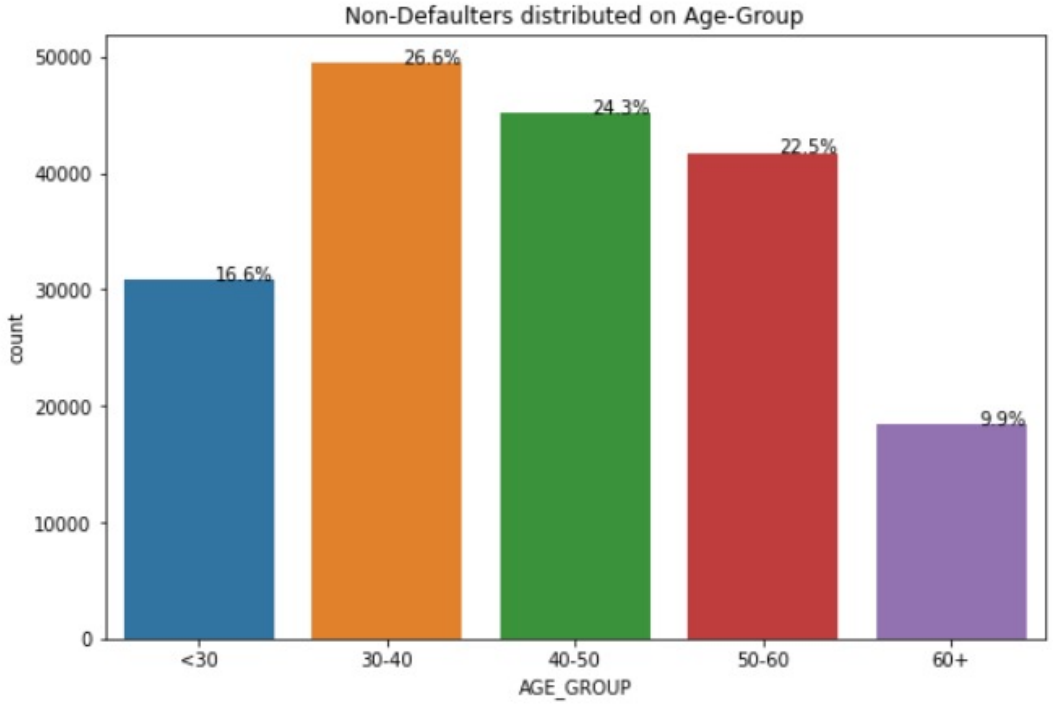


Original Data

AGE-GROUP vs TARGET

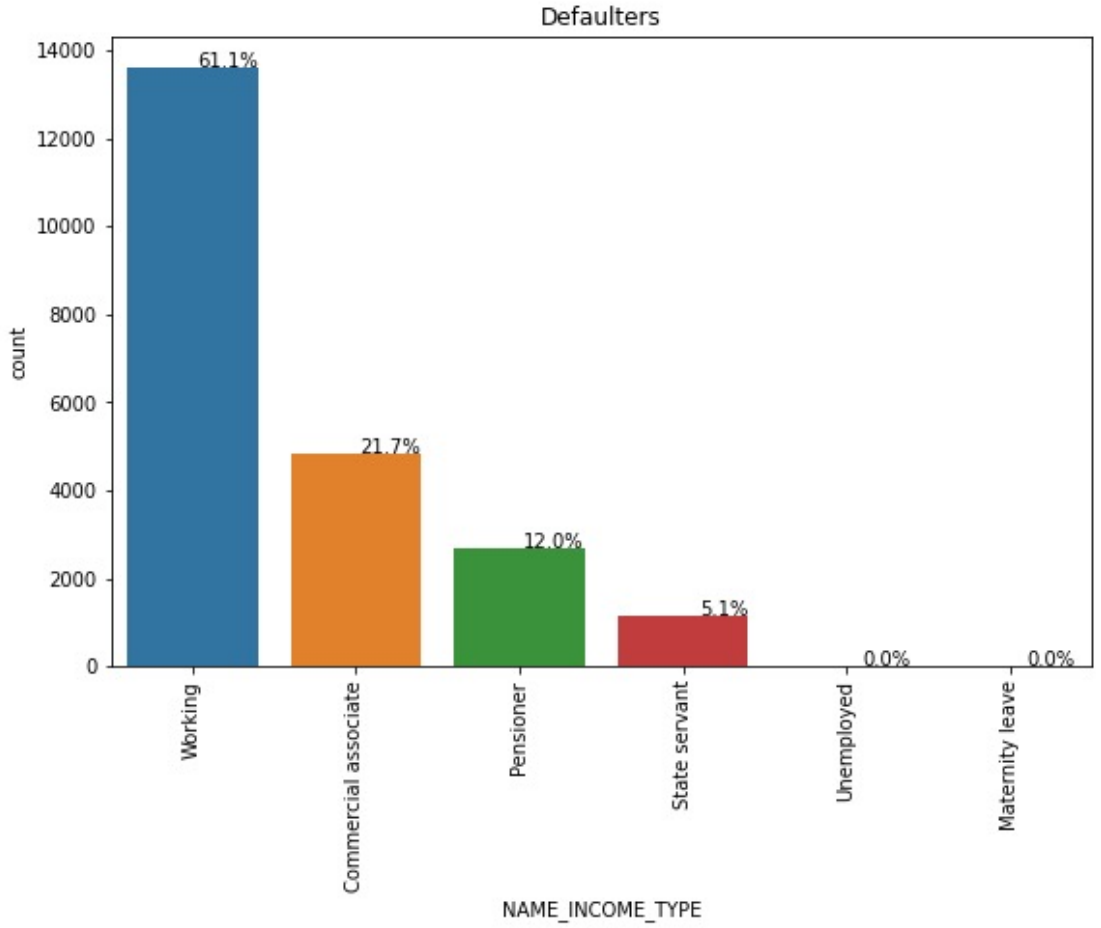
- Percentage of defaulters in age group of 30-40 years are more (31%) than non defaulters (26.6%).
- Defaulters' percentage is more than Non-Defaulters in every age group except 60+.
- Senior citizens (age group of 60+) are likely to default less. Banks can focus more on them for lending hassle free loans.

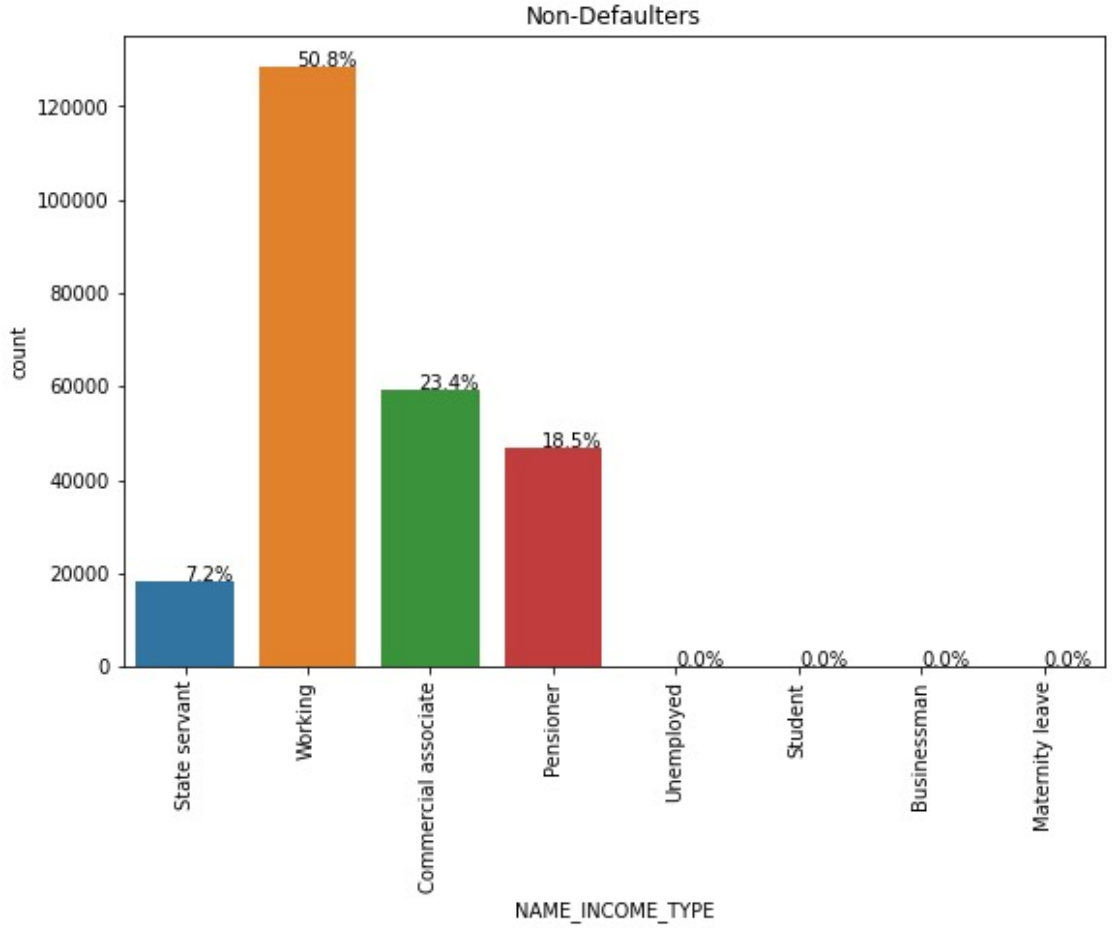




INCOME TYPE vs TARGET

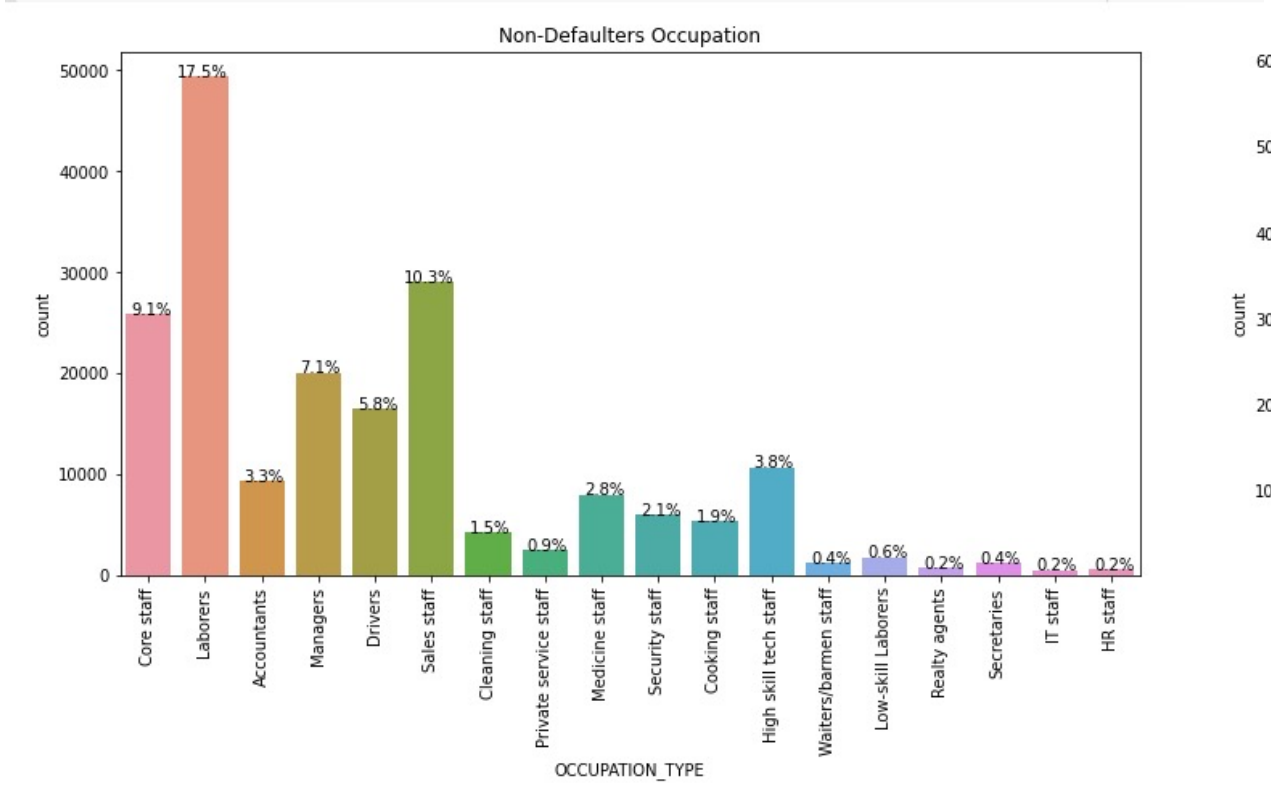
- Working people take more loans but their defaulters' percentage is also more.
- But Commercial Associate, Pensioner and State-servant are more on Non-Defaulters side.

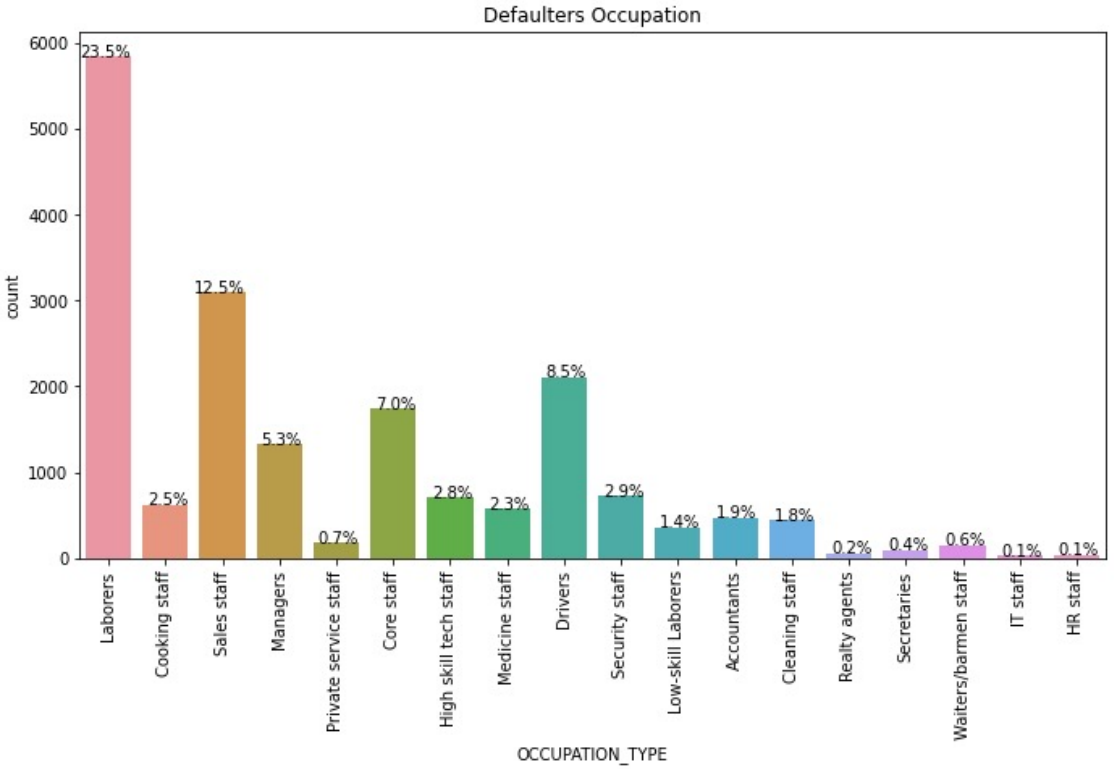




OCCUPATION TYPE vs TARGET

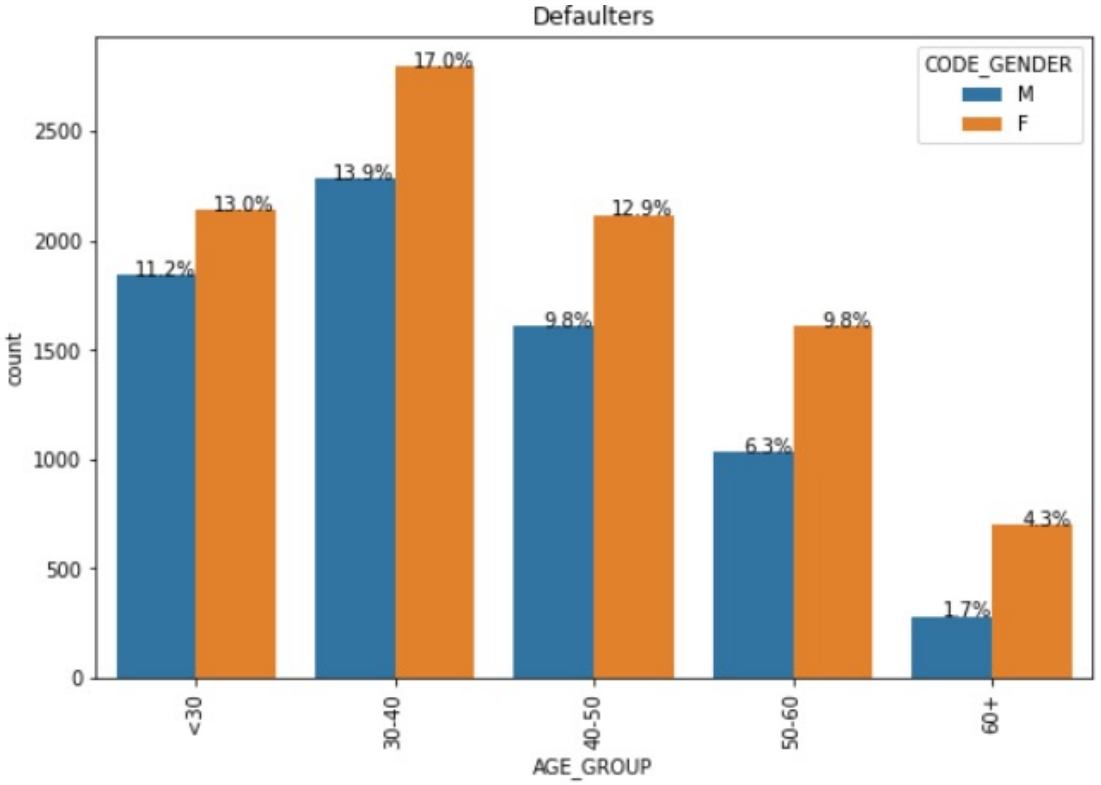
1. Applicants who are Laborers are most likely to be defaulters.
2. Managers and Core-Staff are more on Non-Defaulters side.
3. Highly skilled tech staffs are also more on Non-Defaulters side.
4. Banks must target on lending loans to skilled tech staffs and core staffs. For that they can view their income slab.

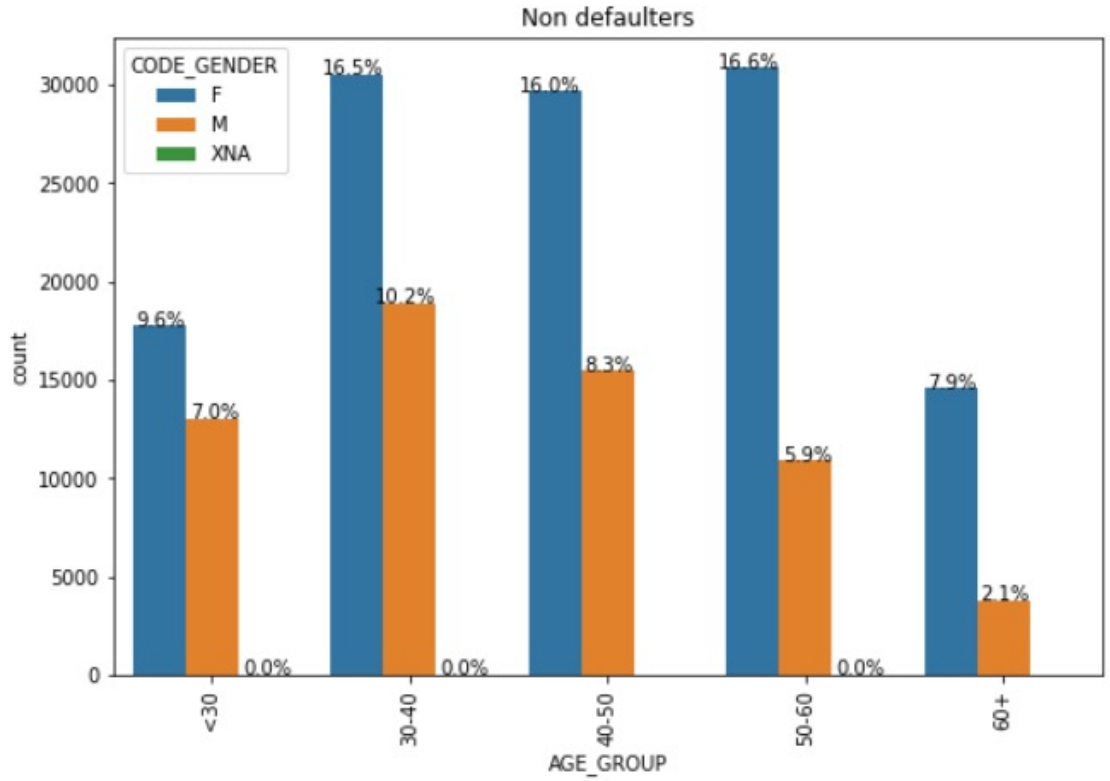




AGE-GROUP vs GENDER

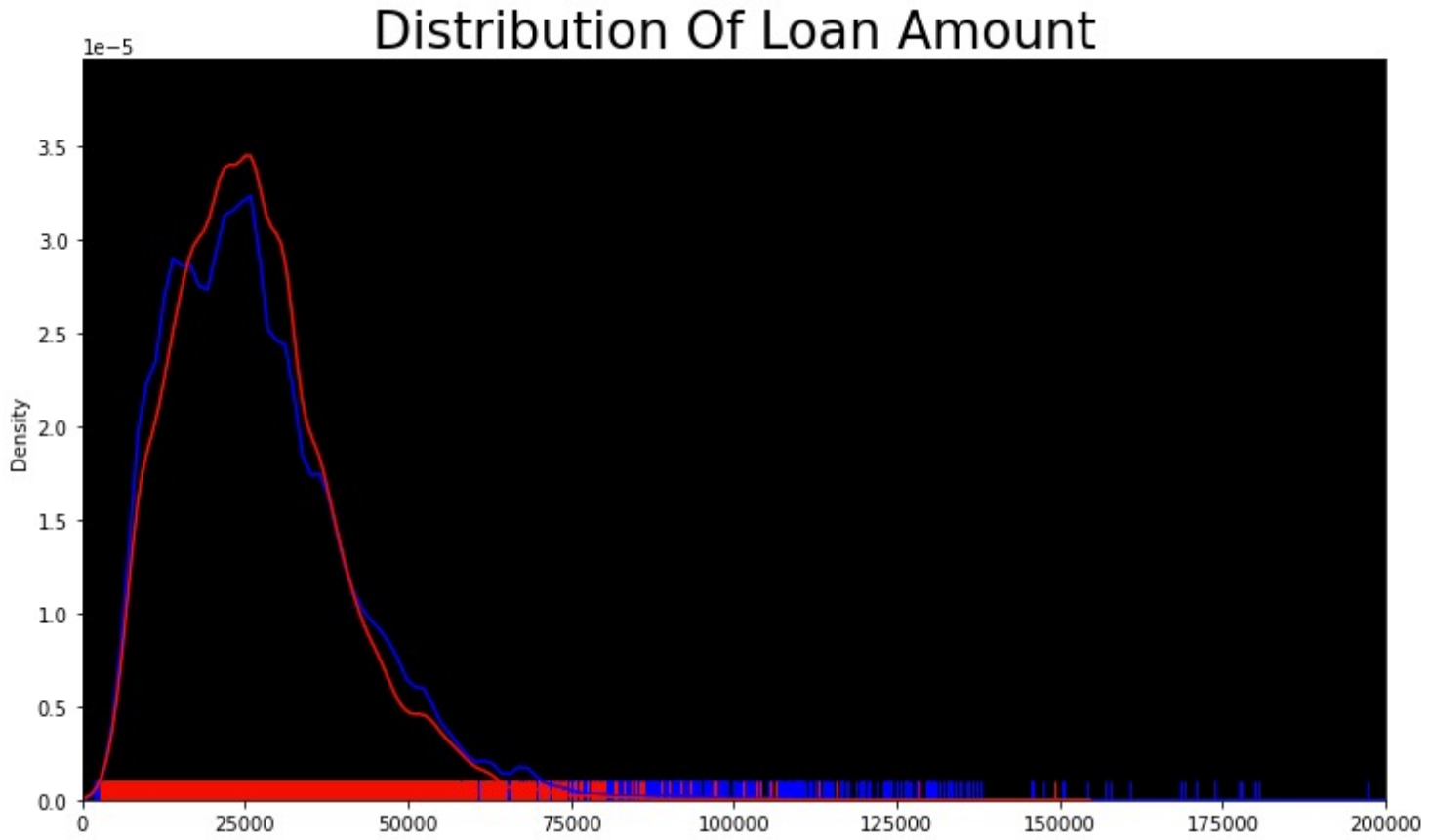
- Females in age group of 30-40 years are most likely to be Non-Defaulters and Defaulters both.
- Females above 40 years of age are more likely to repay the loans compared to Males.
- Males in age group of 30-40 years are more on Defaulters side.





ANNUITY AMOUNT vs TARGET

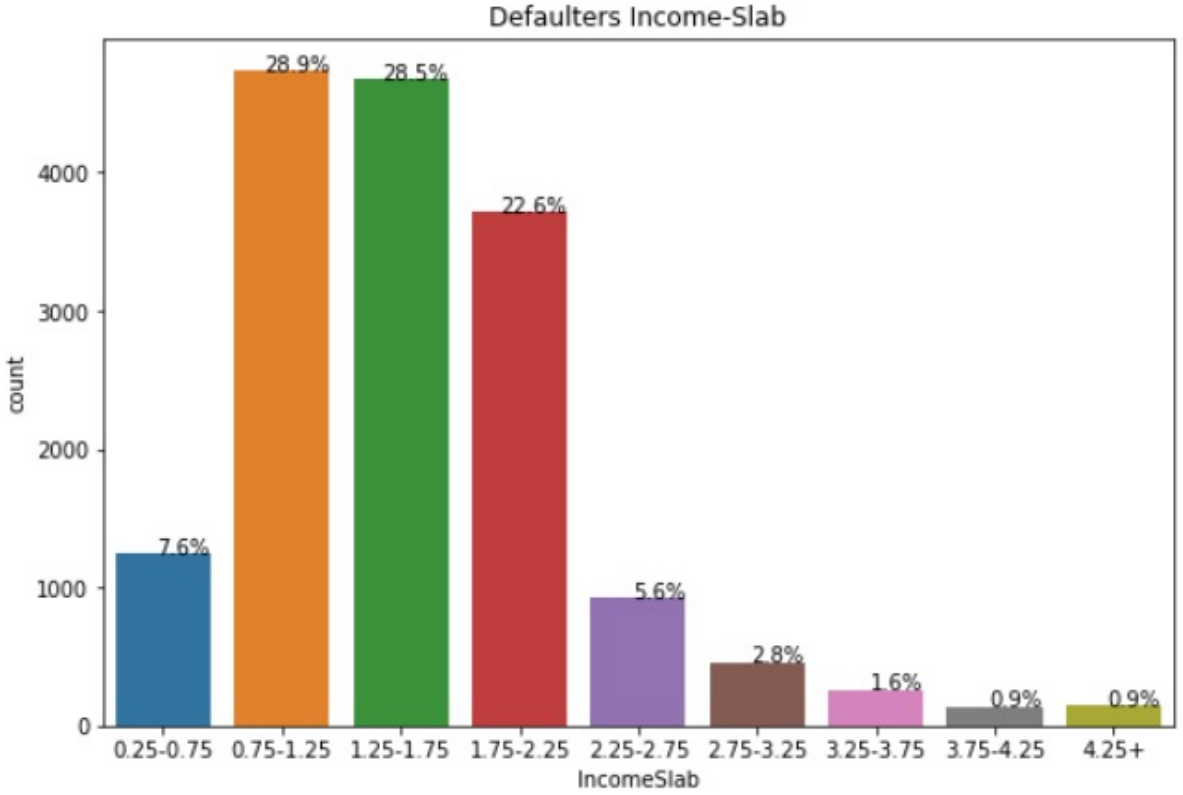
- Statistics says that - Higher loan amount applicants mostly are non-defaulters.

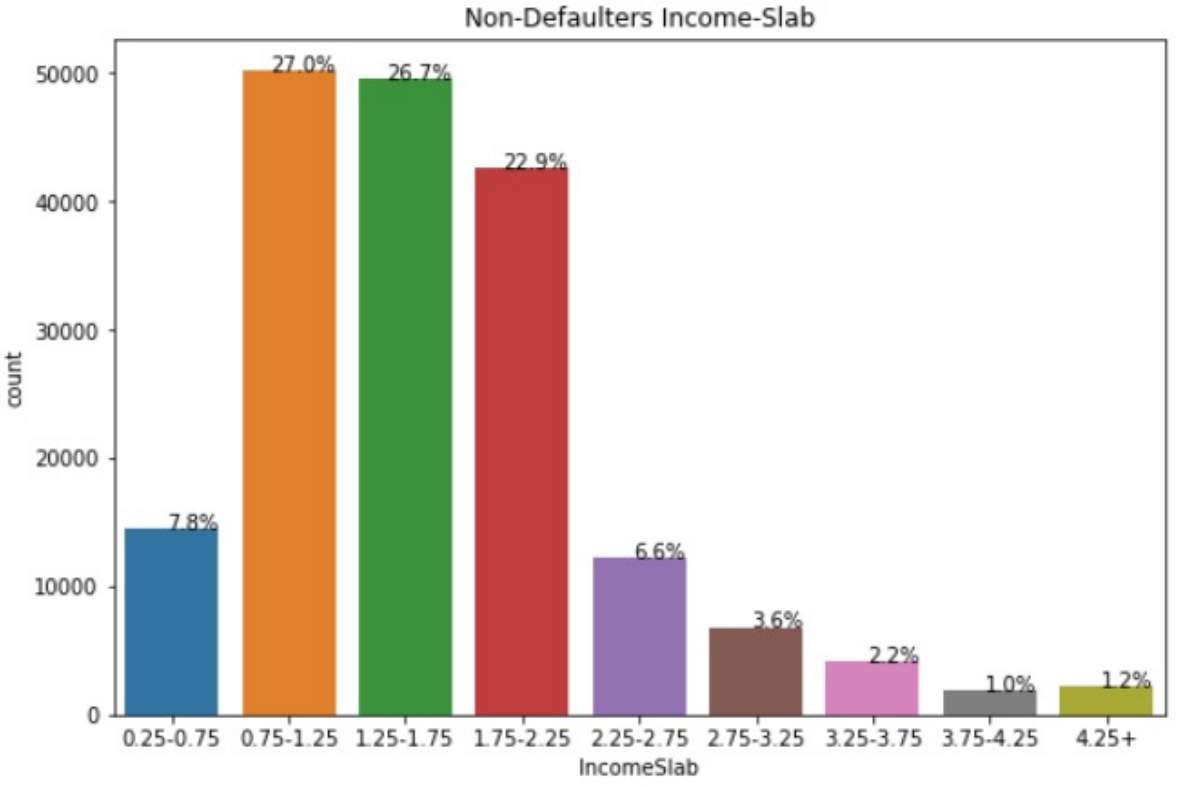


Red : Non-Defaulters
Blue : Defaulters

INCOME RANGE vs TARGET

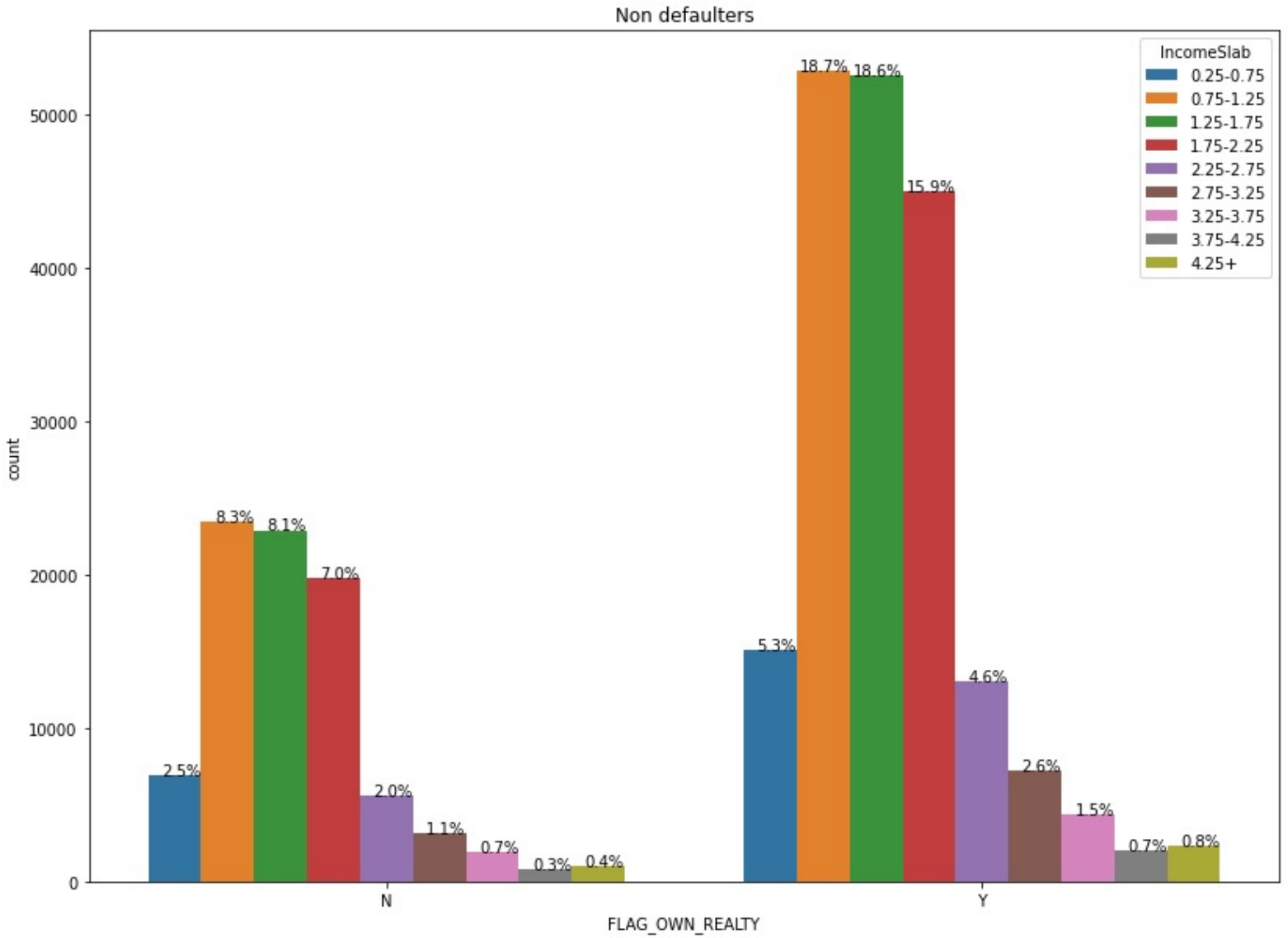
- Applicants having income slab of 0.75-1.25LPA are most likely to repay and default the loans.
- There is only 1.9% difference between two results so banks should target applicants having income slab between 0.75-1.25 LPA and above as per the observation.





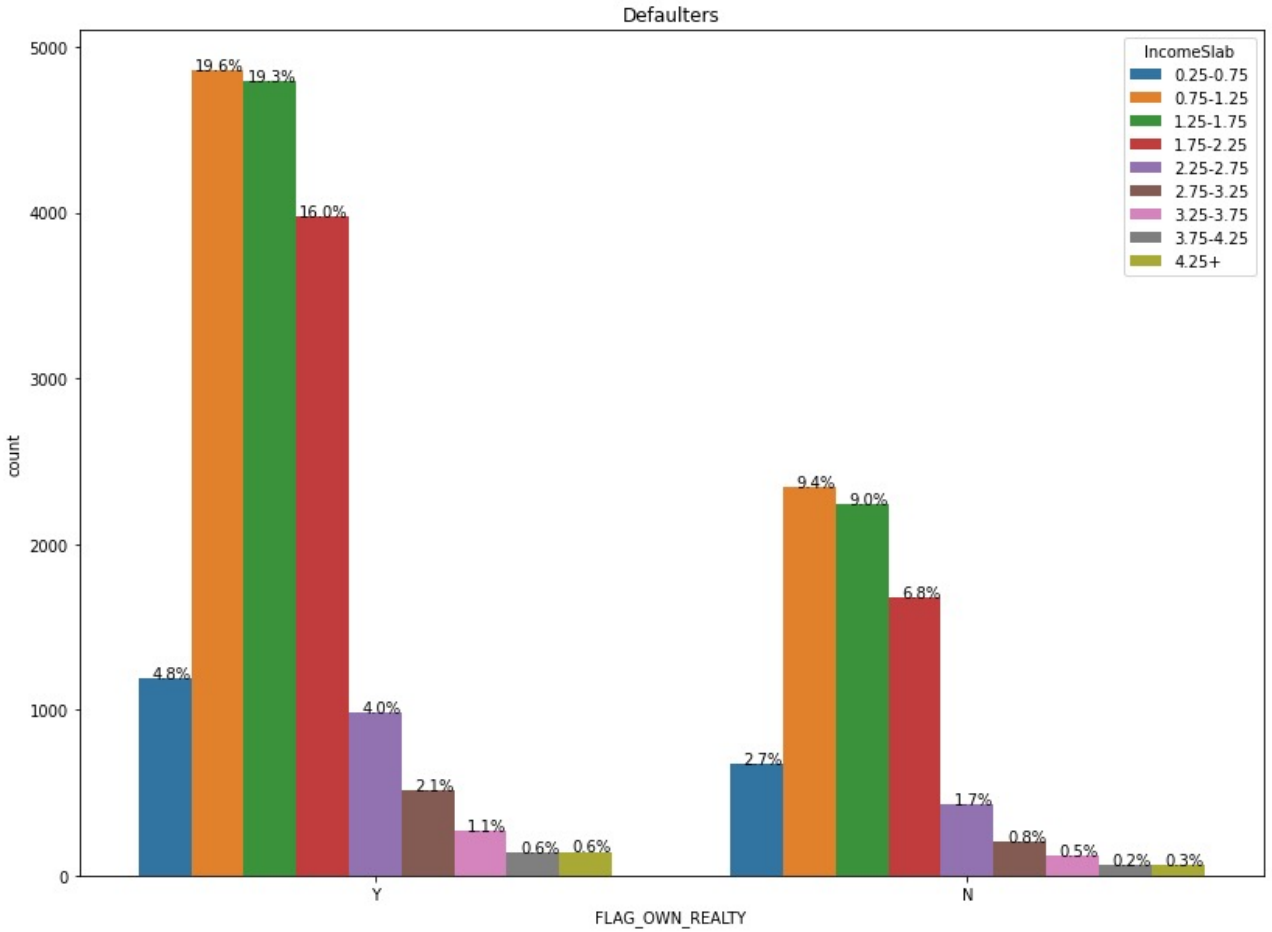
OWN HOUSE vs INCOME (NON-DEFAULTERS)

- Own house and Income Range from 0.75-1.25(75000 to 1 lakh 25000) have almost same percentage of Non-Defaulters.
- Customers with Income Range between 25000 and 75000 are more Non-Defaulters who have their own property.



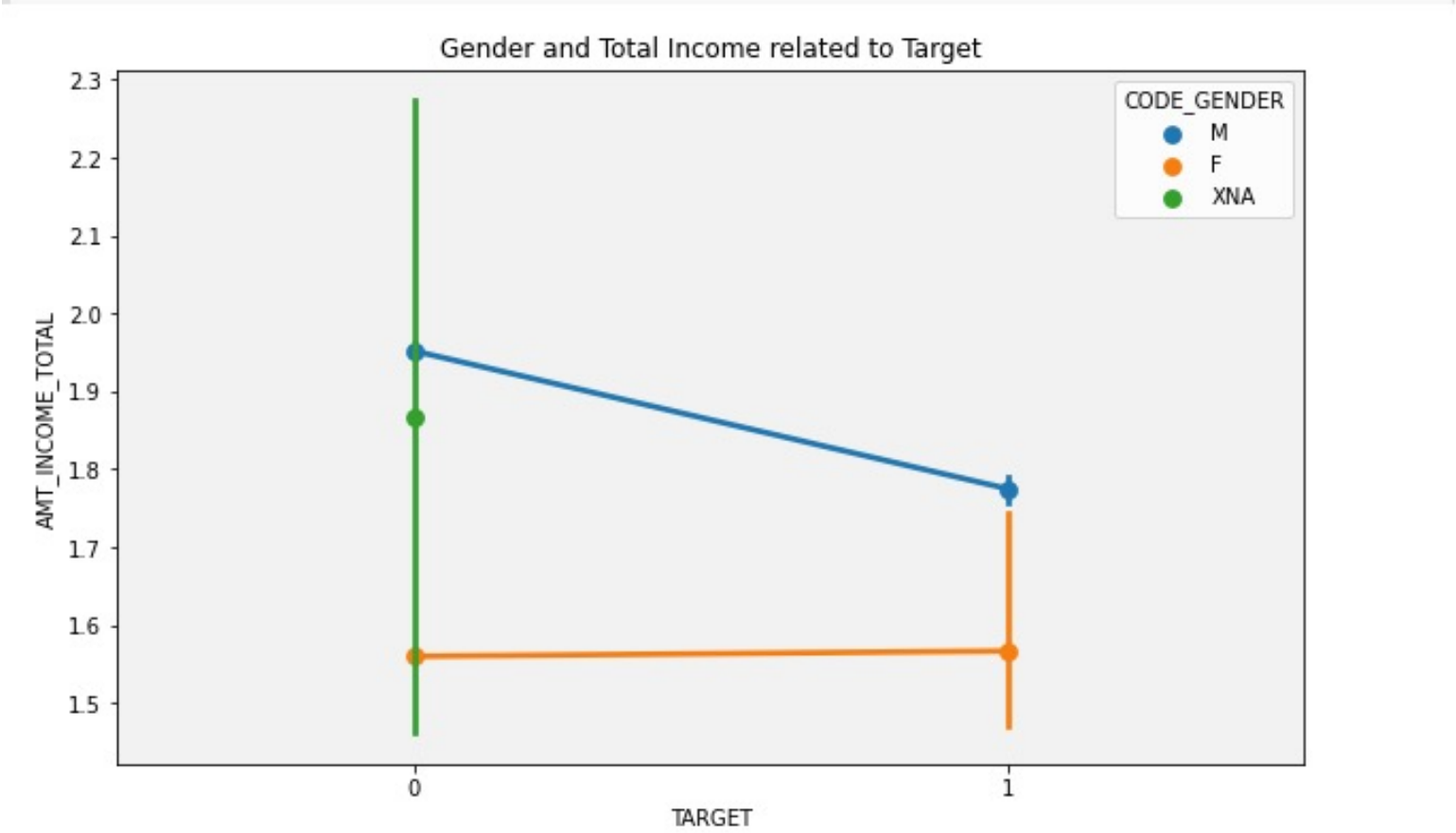
OWN HOUSE vs INCOME (DEFAULTERS)

On Defaulters side they have little bit more percentage in the case either they have their own house or not.



TOTAL INCOME, GENDER vs TARGET

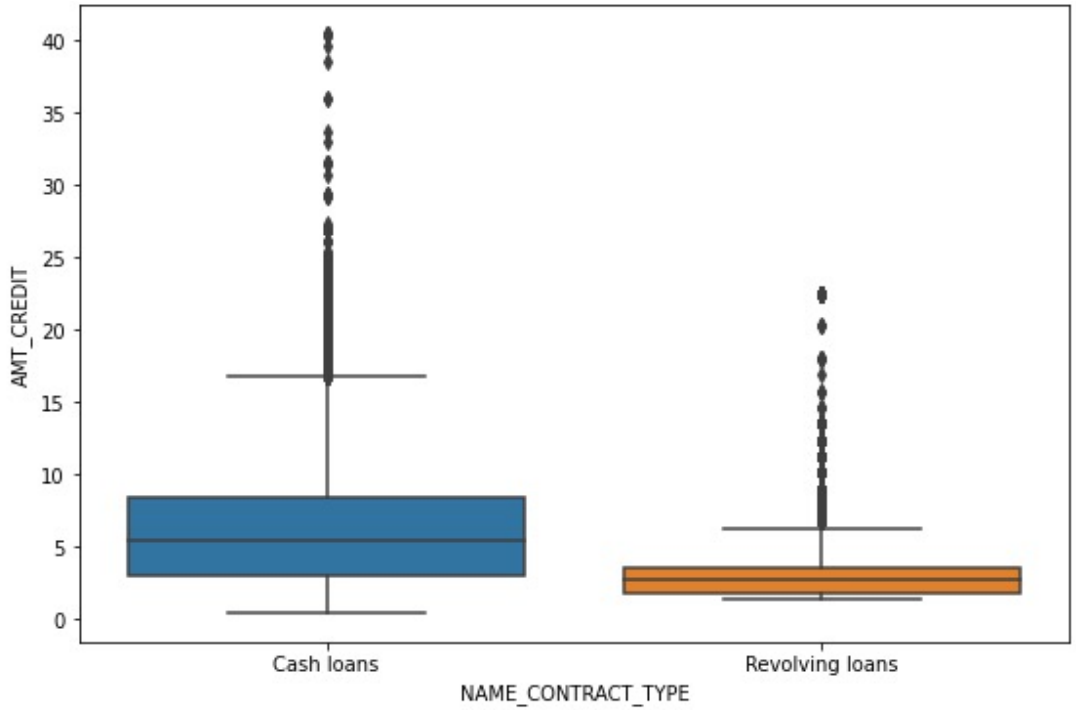
- Men: In the transition from Non-Defaulters to Defaulters, average income is decreasing.
- Women: In the transition from Non-Defaulters to Defaulters, average income remains same.



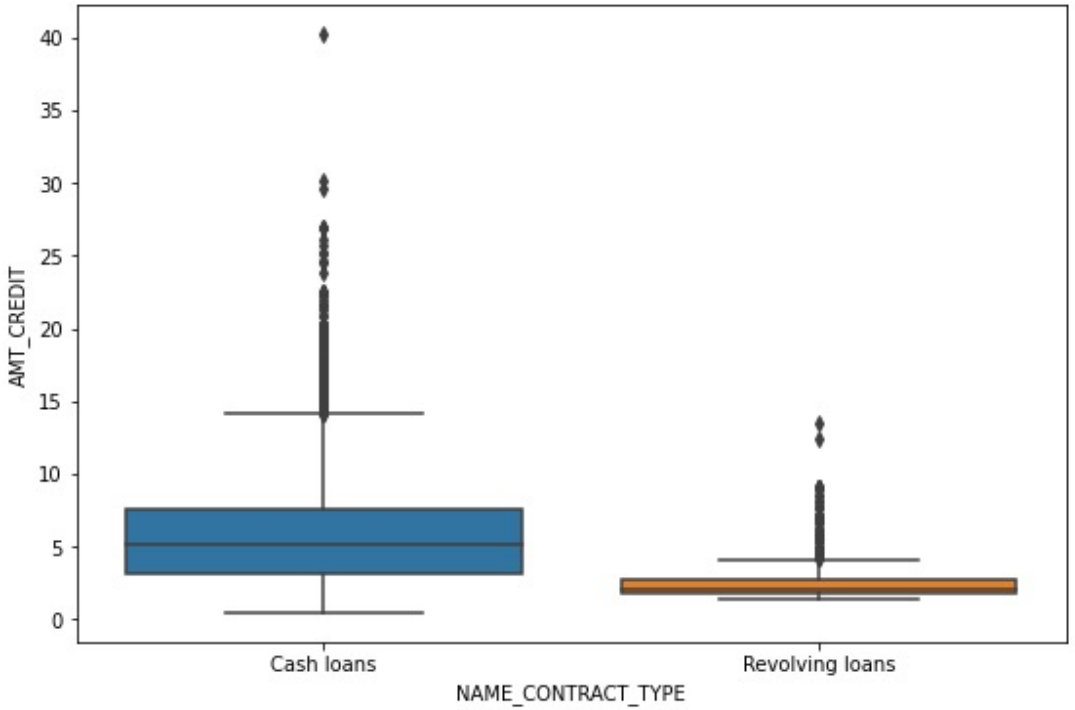
CREDIT AMOUNT vs CONTRACT TYPE

- Cash loans have more credit than revolving loans
- Cash loans and Revolving loans have similar credit amount in case of Defaulters or Non-Defaulters.
- Median for cash loans has same value for both defaulters and non-defaulters.
- Median for Revolving loans in case of defaulters is not exactly visible but for non-defaulters it is almost same as 25th percentile of cash loans.

Non-Defaulters



Defaulters

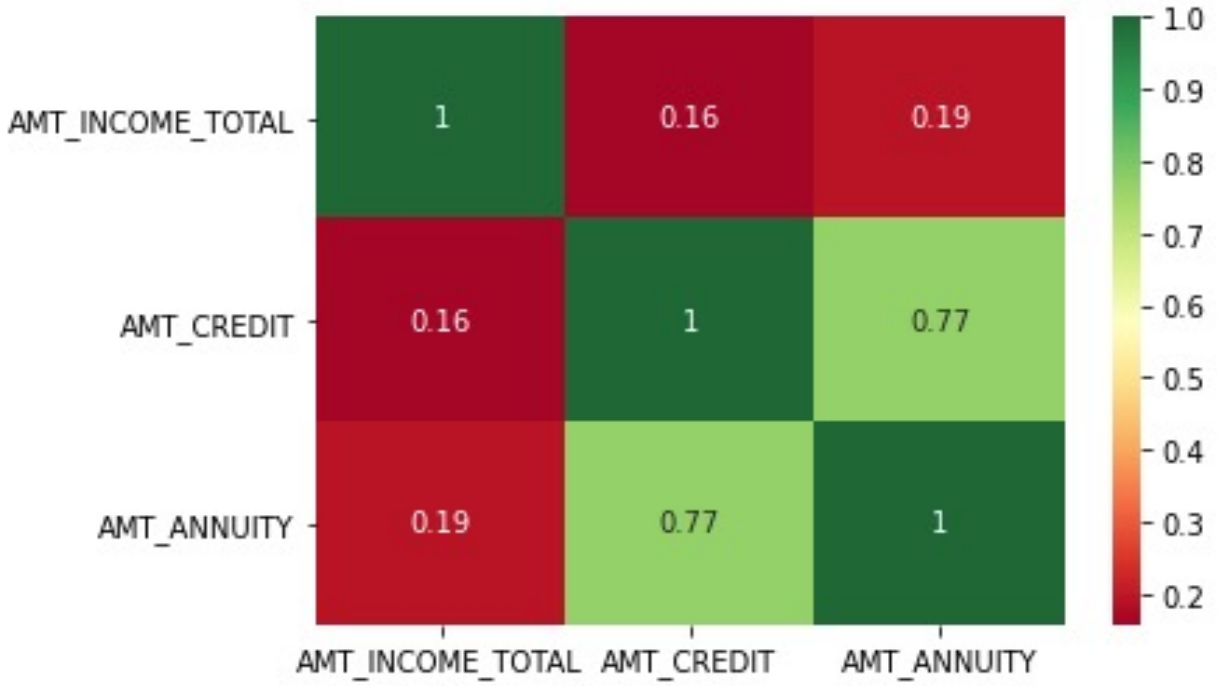


CORRELATION

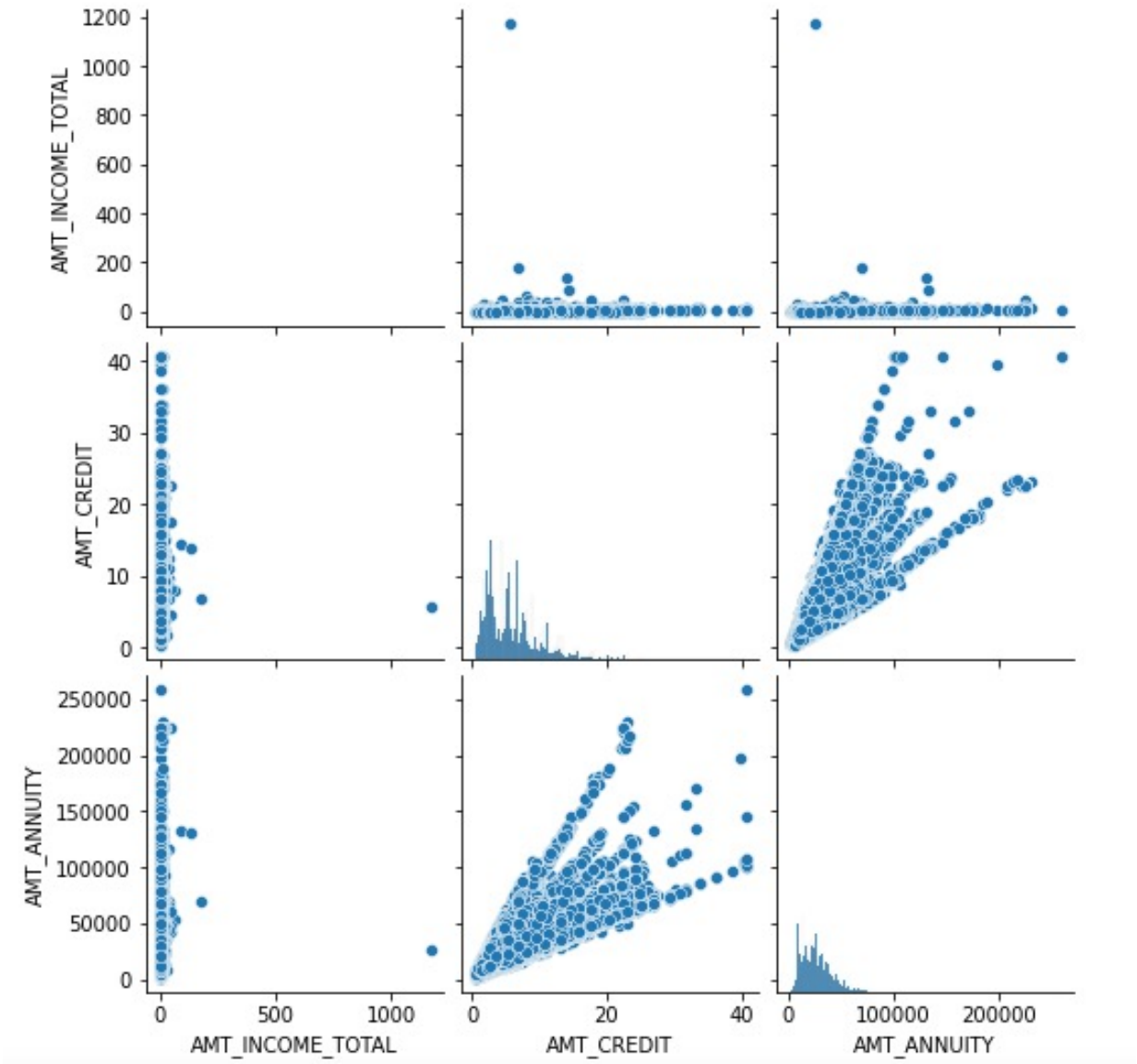
(Total Income, Credit Amount, Annuity Amount)

- Annuity amount is highly correlated with Credit amount, means they have linear relation and have positive correlation.

Heat Map (Total Income, Credit Amount, Annuity Amount)



This Pair plot is also showing the same correlation between Annuity amount and Credit amount.



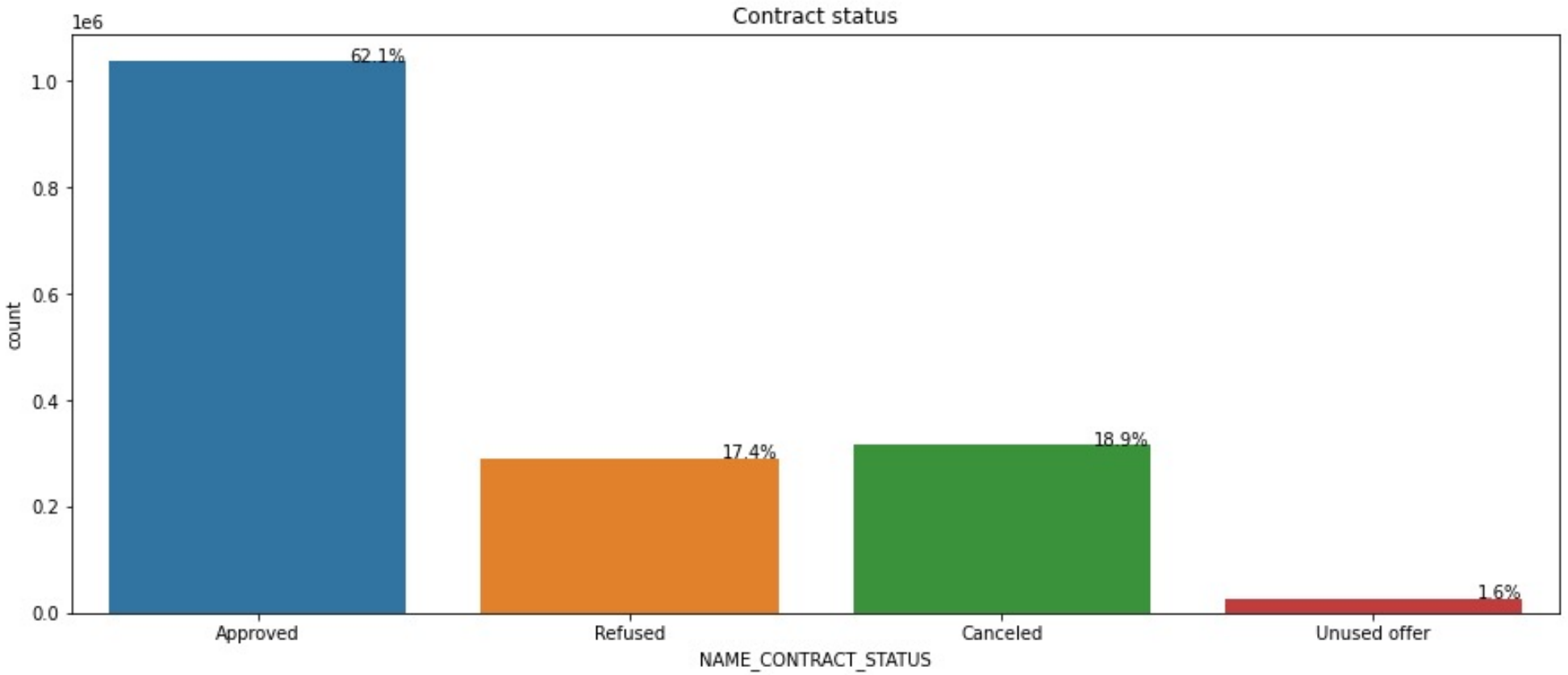
TOP 10 CORRELATED COLUMNS

- Client work contacts (FLAG_EMP_PHONE) and Years employed(YEARS_EMPLOYED) have high correlation.
- Goods price amount(AMT_GOODS_PRICE) and Credit amount(AMT_CREDIT) have also good correlation.

	Column1	Column2	Correlation	Abs_Correlation
2363	YEARS_EMPLOYED	FLAG_EMP_PHONE	-0.999753	0.999753
1848	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998490	0.998490
1646	FLOORSMAX_MEDI	FLOORSMAX_AVG	0.997034	0.997034
1596	YEARS_BEGINEXPLUATATION_MEDI	YEARS_BEGINEXPLUATATION_AVG	0.993825	0.993825
1648	FLOORSMAX_MEDI	FLOORSMAX_MODE	0.988237	0.988237
298	AMT_GOODS_PRICE	AMT_CREDIT	0.986968	0.986968
1548	FLOORSMAX_MODE	FLOORSMAX_AVG	0.985689	0.985689
1498	YEARS_BEGINEXPLUATATION_MODE	YEARS_BEGINEXPLUATATION_AVG	0.971893	0.971893
1598	YEARS_BEGINEXPLUATATION_MEDI	YEARS_BEGINEXPLUATATION_MODE	0.963539	0.963539
899	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.950842	0.950842

CONTRACT STATUS (PREVIOUS-APPLICATION DATA)

Around 62% applications are approved.

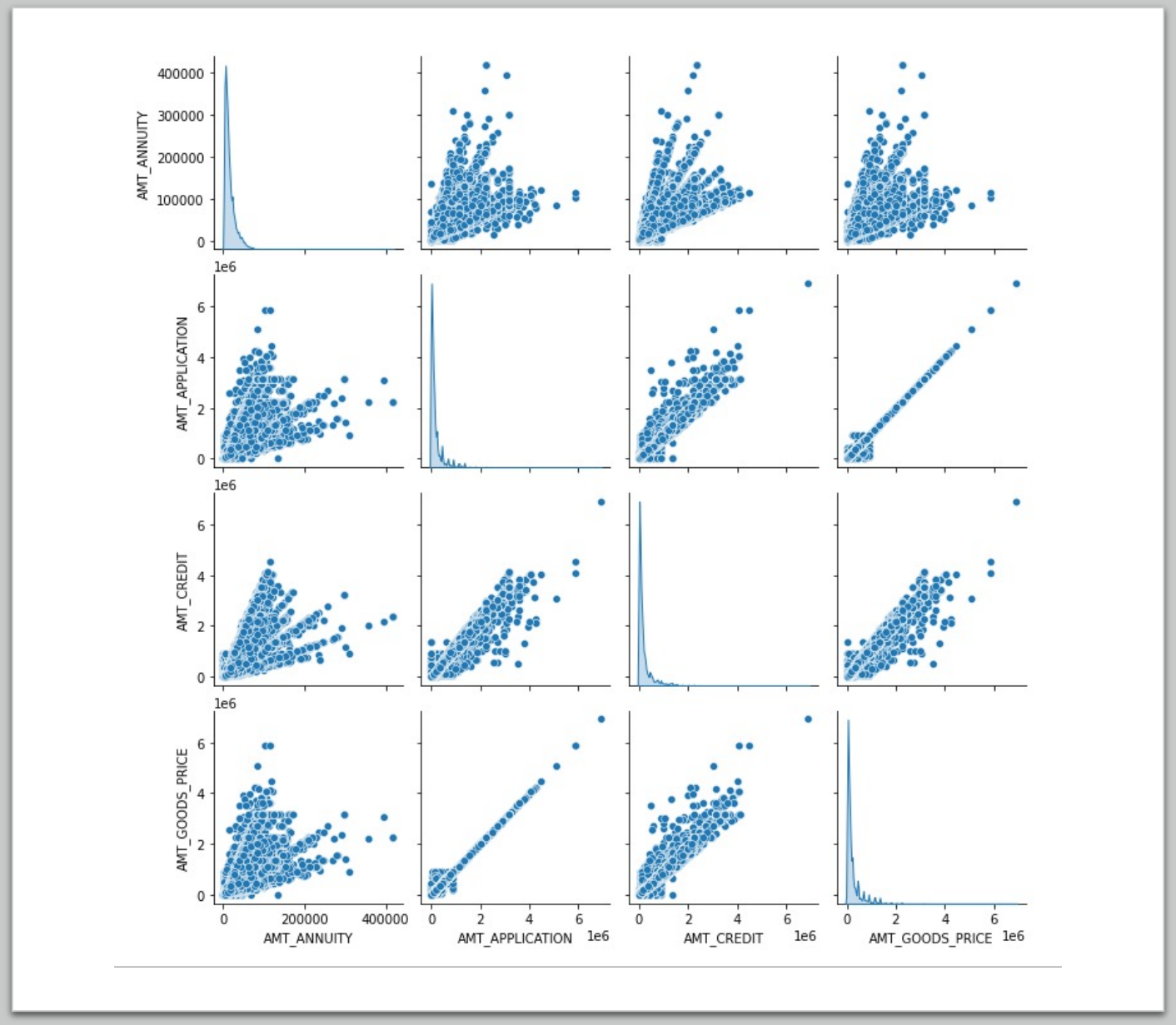


CORRELATION

(NUMERICAL COLUMNS OF PREVIOUS APPLICATION DATA)

Annuity of previous application has a very high and positive influence over: (Increase of annuity increases below factors)

- How much credit did client asked on the previous application
- Final credit amount on the previous application that was approved by the bank
- Goods price of good that client asked for on the previous application.

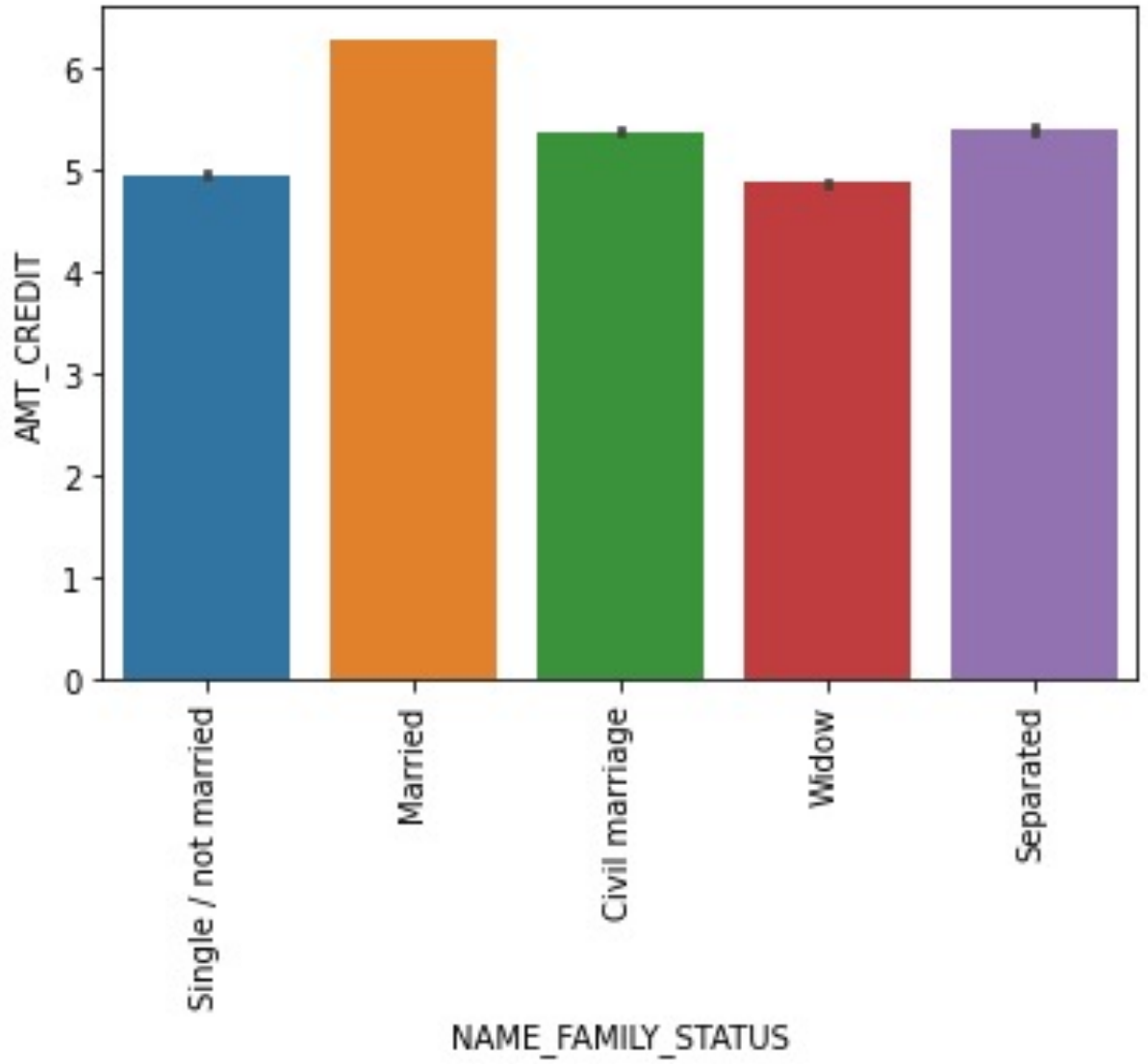


TOP 10 CORRELATED COLUMNS (PREVIOUS APPLICATION)

	Column1	Column2	Correlation	Abs_Correlation
88	AMT_GOODS_PRICE	AMT_APPLICATION	0.999892	0.999892
89	AMT_GOODS_PRICE	AMT_CREDIT	0.993186	0.993186
71	AMT_CREDIT	AMT_APPLICATION	0.976134	0.976134
269	DAYS_TERMINATION	DAYS_LAST_DUE	0.928864	0.928864
87	AMT_GOODS_PRICE	AMT_ANNUITY	0.822610	0.822610
70	AMT_CREDIT	AMT_ANNUITY	0.817688	0.817688
53	AMT_APPLICATION	AMT_ANNUITY	0.810793	0.810793
232	DAYS_LAST_DUE_1ST_VERSION	DAYS_FIRST_DRAWING	-0.803947	0.803947
173	CNT_PAYMENT	AMT_APPLICATION	0.680316	0.680316
174	CNT_PAYMENT	AMT_CREDIT	0.674317	0.674317

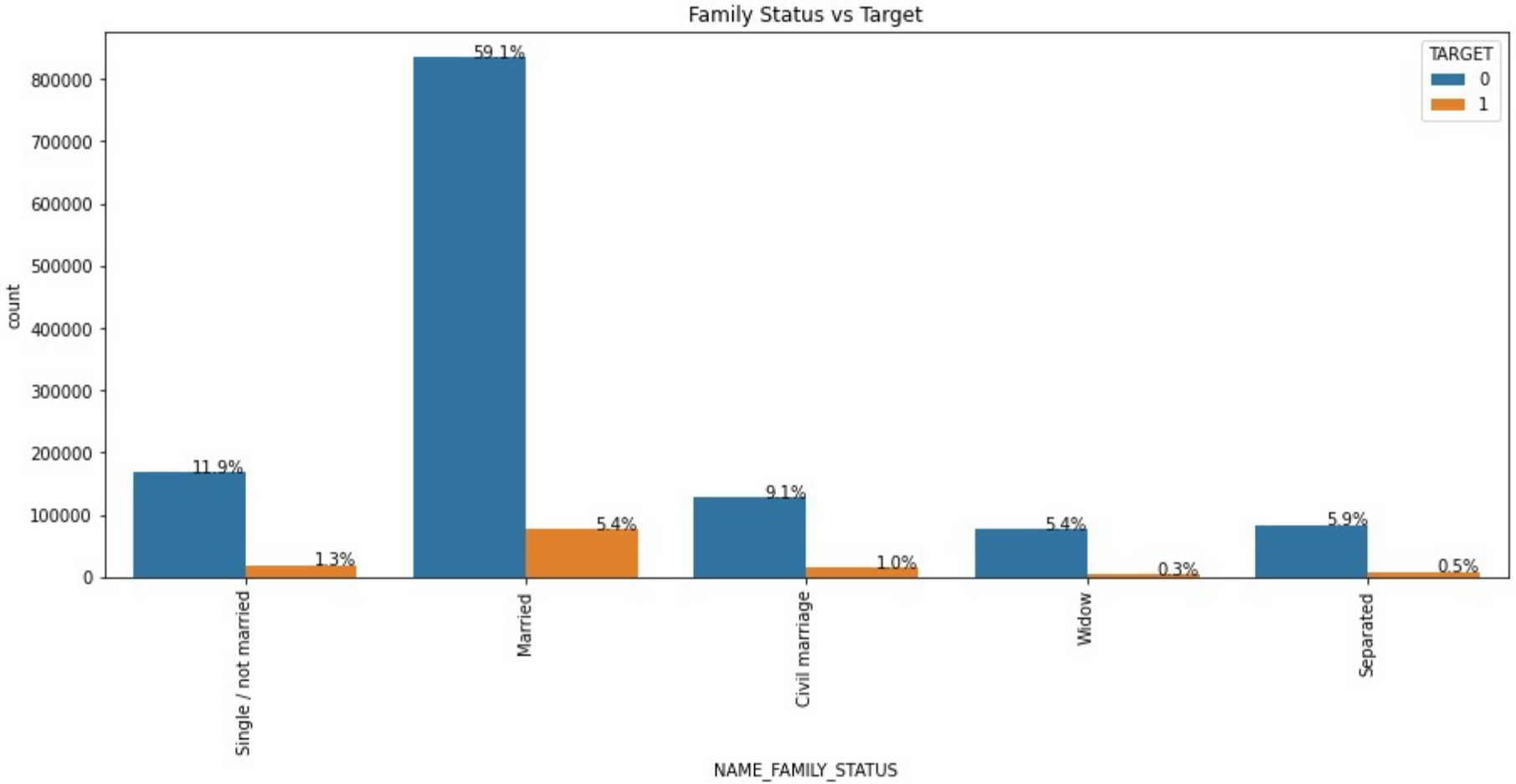
FAMILY STATUS vs CREDIT AMOUNT (Merged Data)

Married Applicants apply for large credit amount than others.



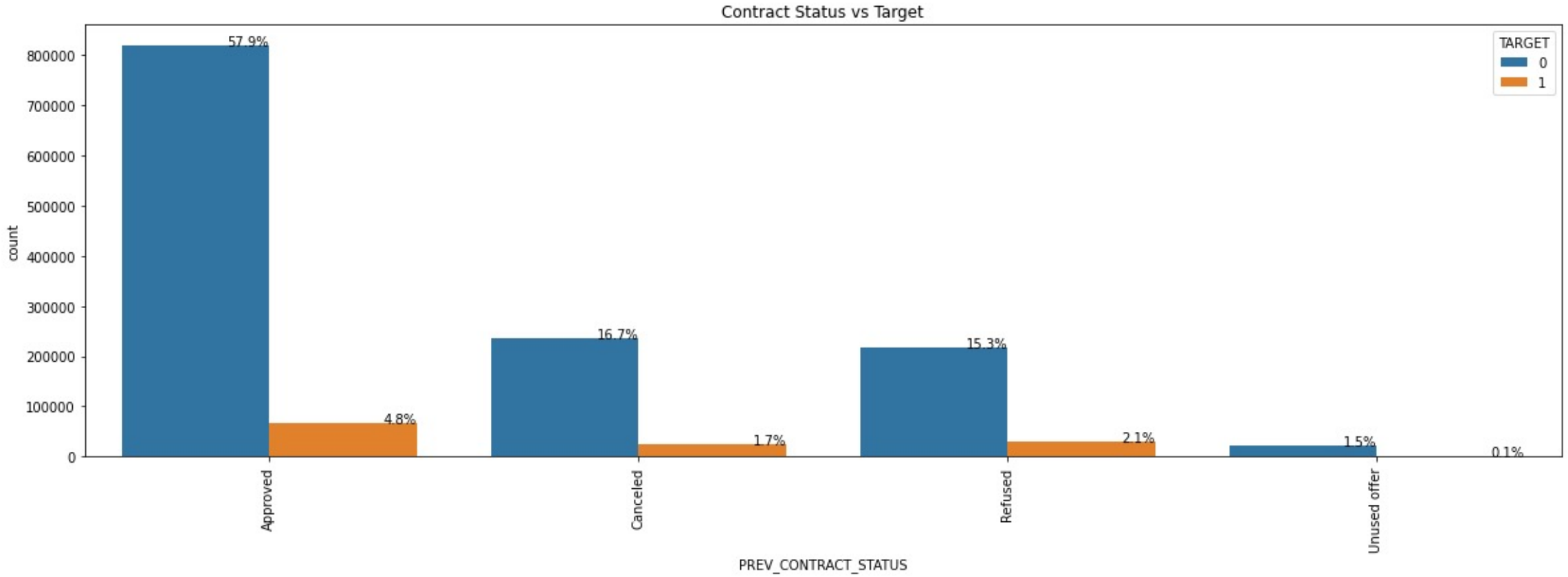
FAMILY STATUS vs TARGET

Married people are mostly non-defaulters



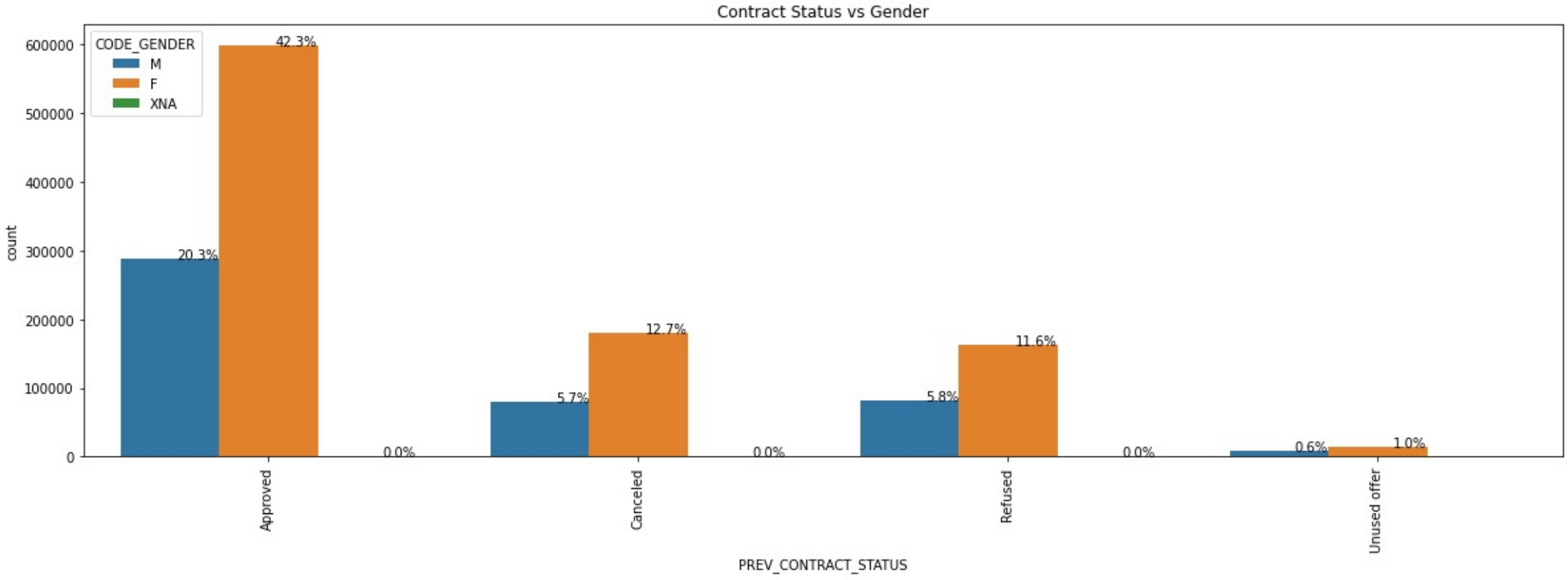
CONTRACT STATUS vs TARGET

We can see that the people whose loan application has approved, defaulted less often.



CONTRACT STATUS vs TARGET

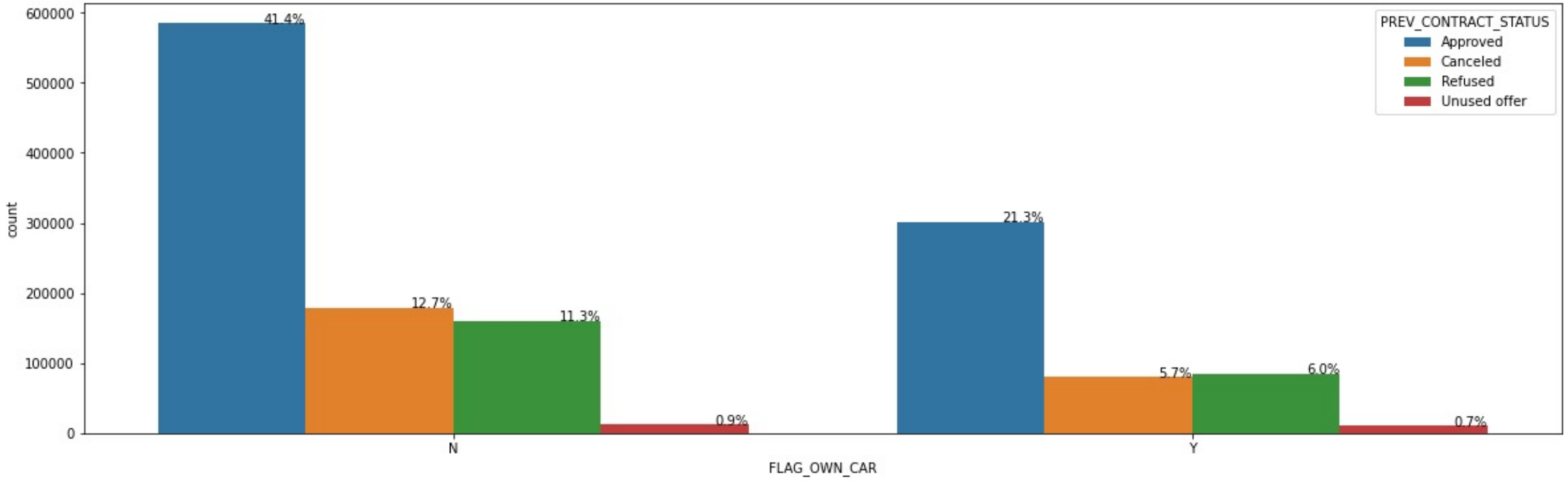
- Female have lesser chances of default compared to males. The bank adds more weightage to female while approving a loan amount.



PREVIOUS CONTRACT STATUS vs CAR OWNER FLAG

- We see that car ownership didn't affect application approval or rejection. But we saw earlier that the people who didn't have car had lesser chances of default.

CAR OWNER – CONTRACT STATUS



THANK YOU

Regards – Niharika Girdhar
Nishant Kumar