

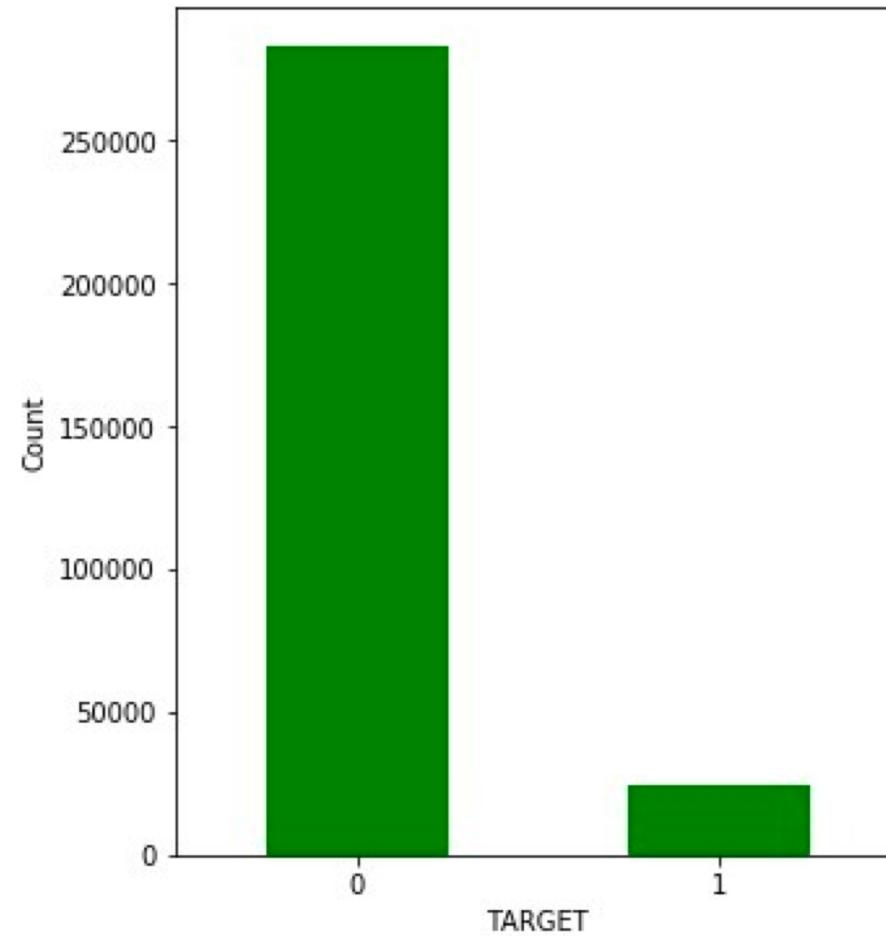
EDA

CASE STUDY

BY
KISHORE & SRINIVAS

DATA IMBALANCE:

- Target feature distribution is highly imbalanced.
- 'Target-0' - 91.93%
- 'Target-1' - 8.07%

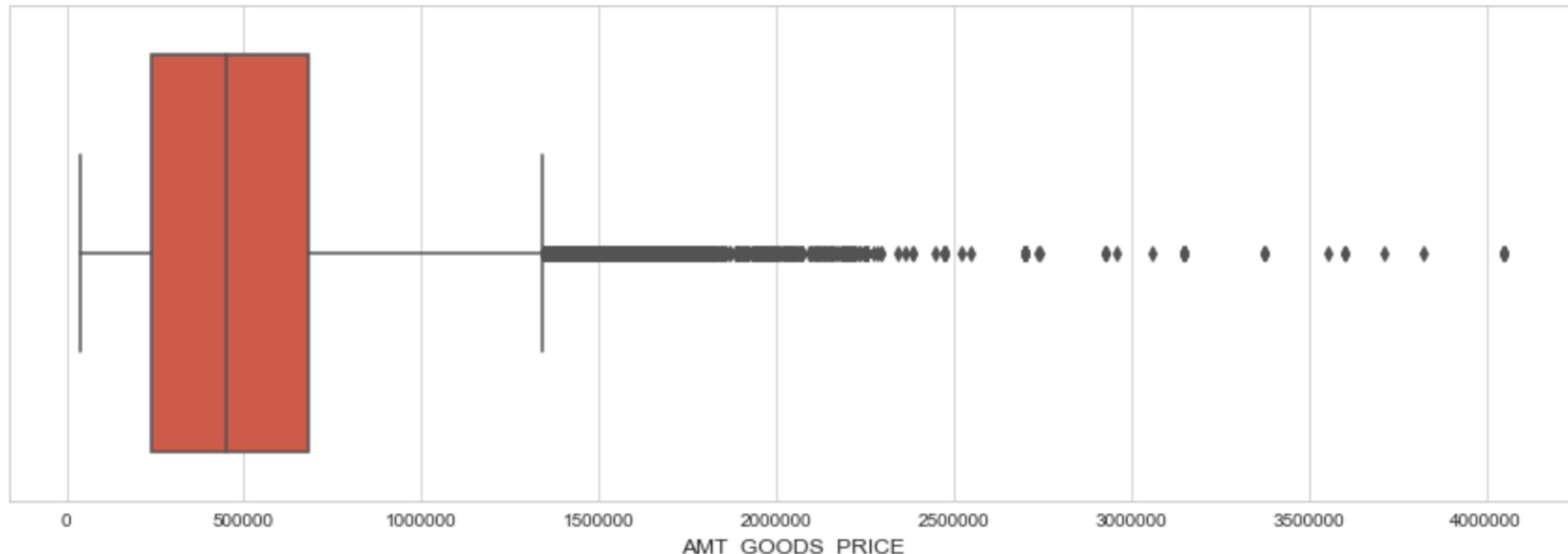


OUTLIERS:

(values from 95 to 100%)

Feature – Amount credited (**AMT_CREDIT:**)

- Here after 0.95 quantile value outliers are present so will remove the outliers from AMT_CREDIT.
- we should Remove Outliers

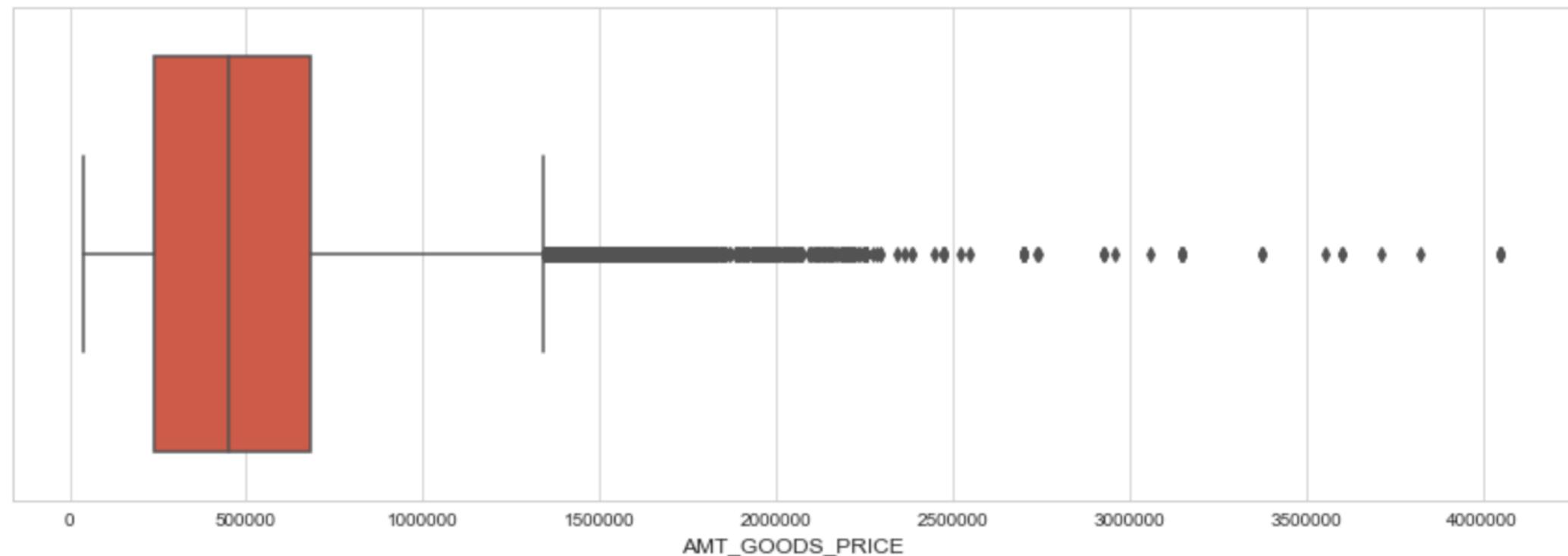


OUTLIERS:

(values from 99 to 100%)

Feature - Amount Income Total: (AMT_INCOME_TOTAL:)

- Income ranges from 25k to 300k. As per above after 0.99 quantile value outliers are present
- we should Remove Outliers

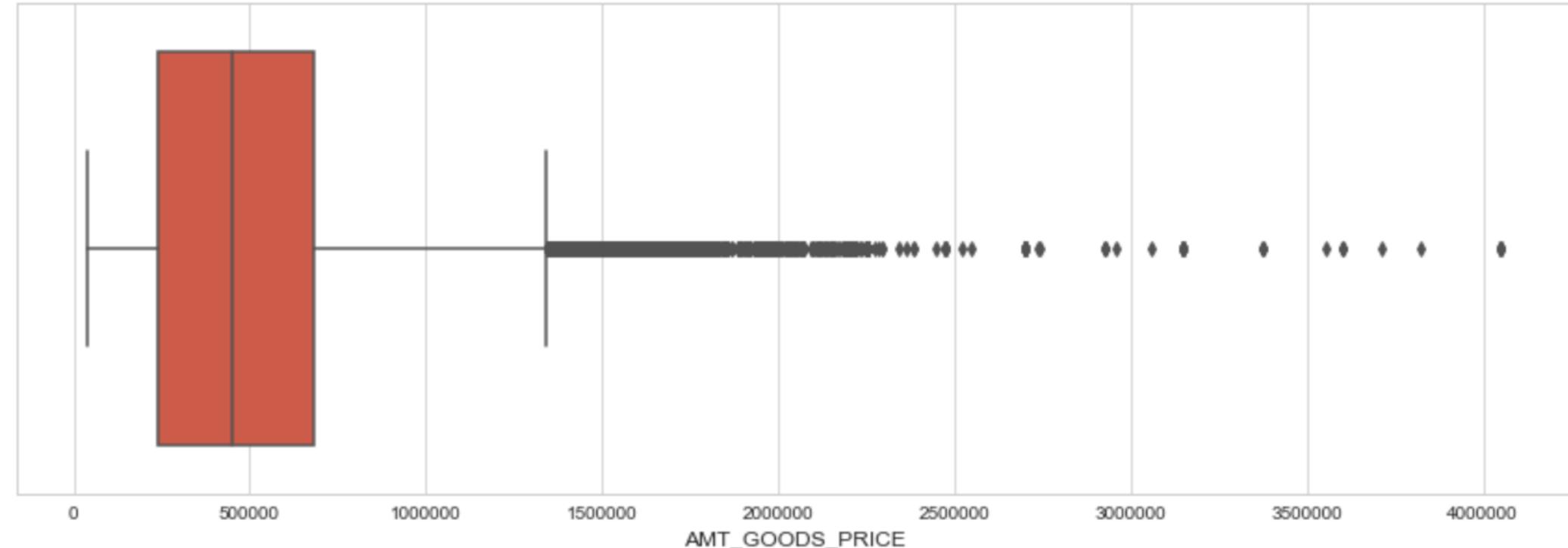


OUTLIERS:

>0.95%

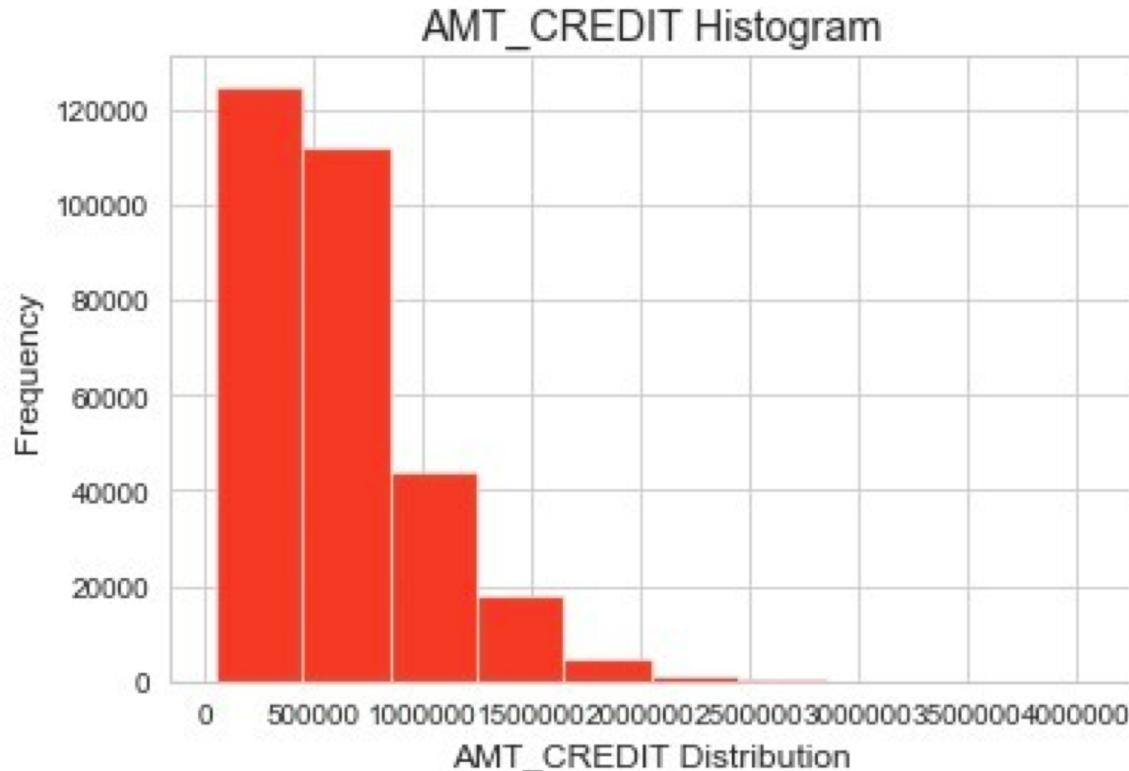
Feature - Amount Annuity - (AMT_ANNUITY:)

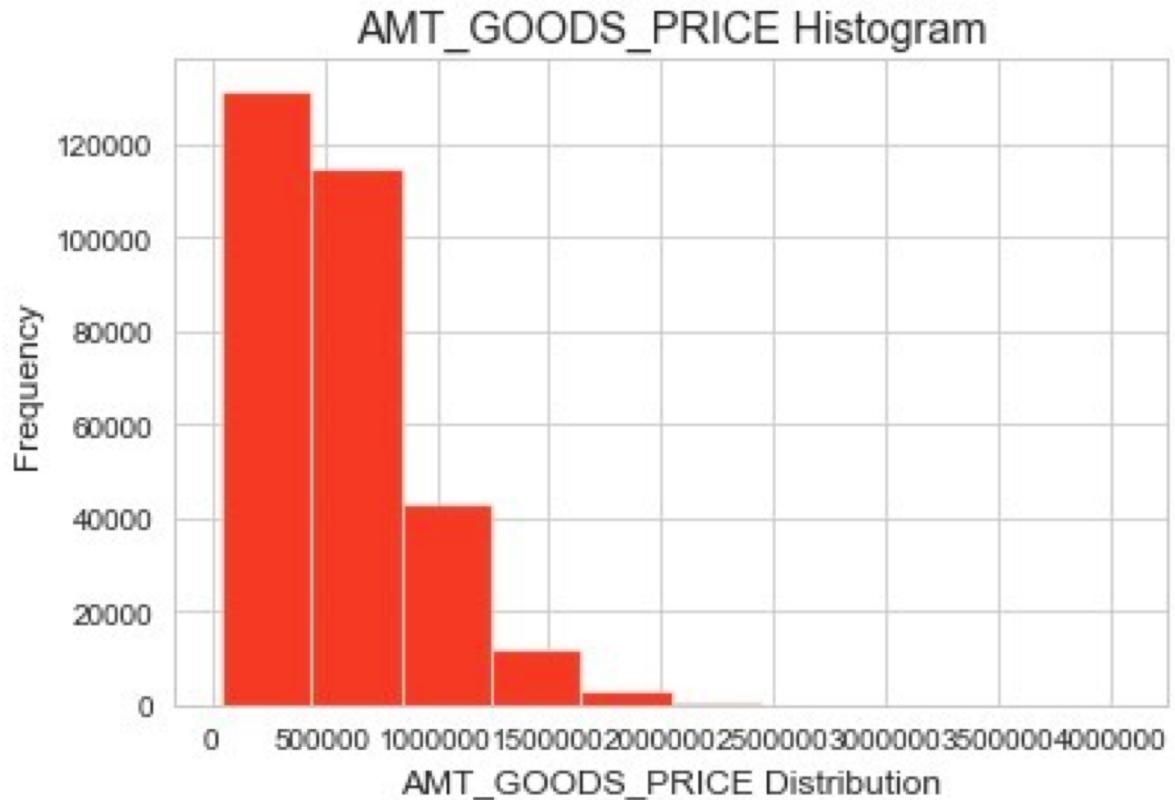
- Here so many outliers are there after 0.95 range so will remove the outlier present after 0.95 them from the column AMT_ANNUITY.
- we should Remove Outliers.



AMOUNT CREDIT HISTOGRAM:

- Amount which is credited for the clients in all the loans.
- We can see the distribution of that amount on x axis & frequency on y-axis

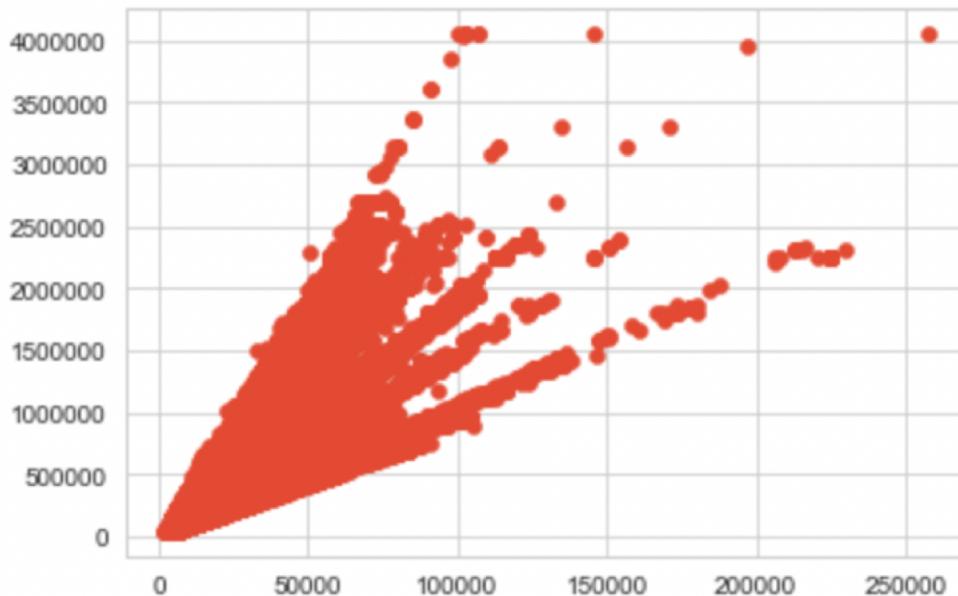




AMOUNT GOODS PRICE HISTOGRAM:

We have seen till now both the Histogram's where we have seen the distribution for “AMT_GOODS_PRICE” & “AMT_CREDIT”

Both plots are showing almost similar distribution.



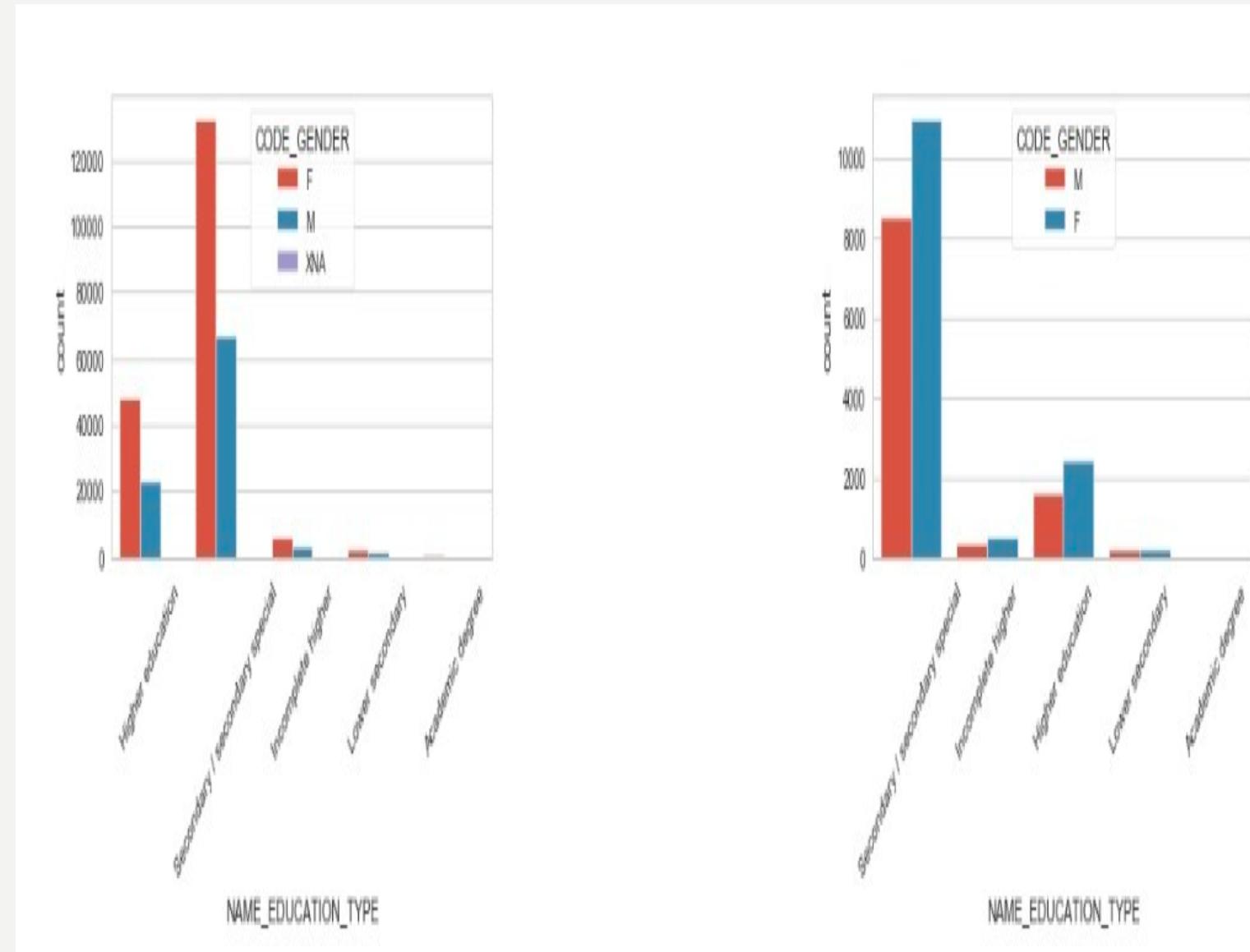
Inference :

Here range is from 0 to 200.

It shows if AMT_ANNUITY increases then AMT_CREDIT also increasing and also it having outliers

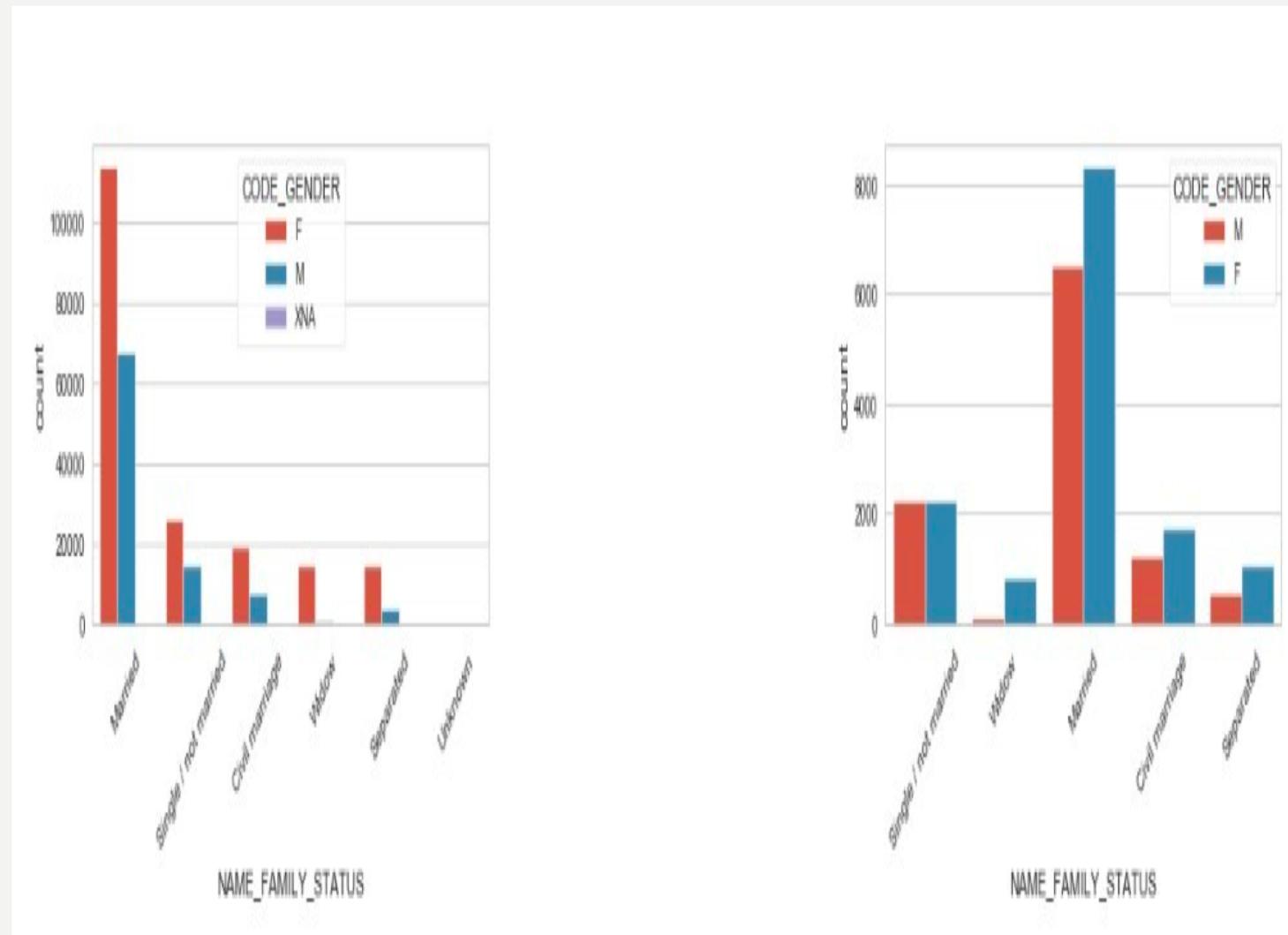
INSIGHTS FROM EDUCATION GRAPH:

- ANALYSIS FOR GENDER FEATURE:
- COMPARISON ON BOTH MALE & FEMALE w.r.to NAME_EDUCATION_TYP E in both Target – 0 & 1



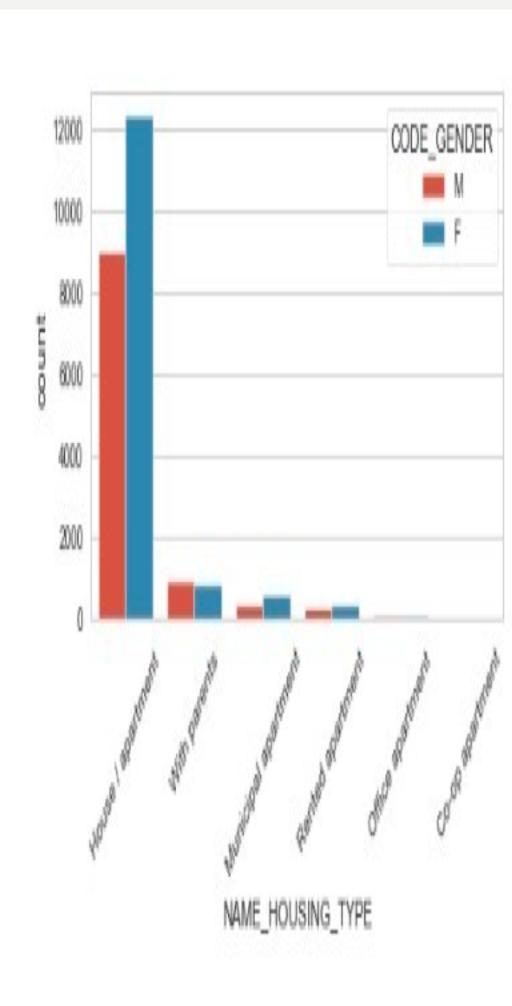
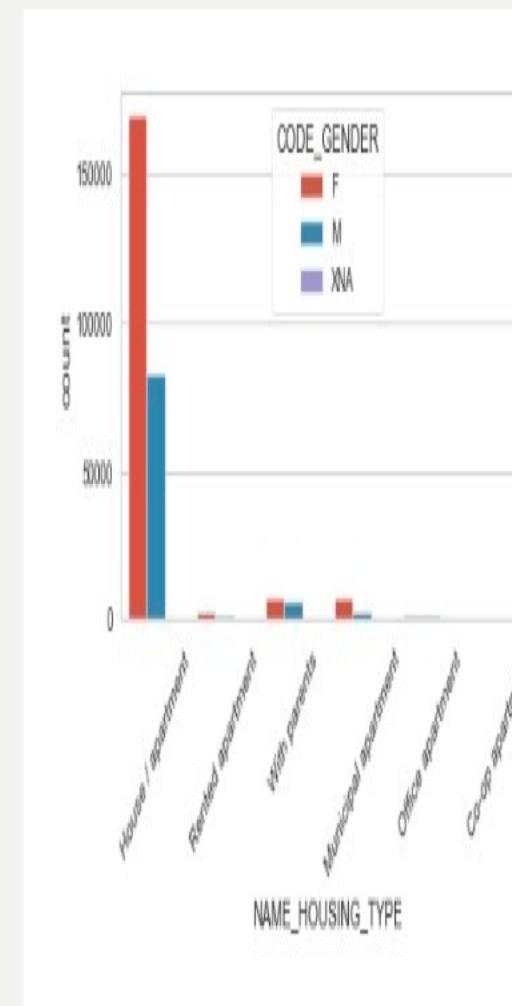
INSIGHTS FROM FAMILY GRAPH:

- ANALYSIS FOR GENDER FEATURE:
- COMPARISON ON BOTH MALE & FEMALE w.r.to NAME_FAMILY_STATUS in both Target – 0 & 1



INSIGHTS FROM FAMILY GRAPH:

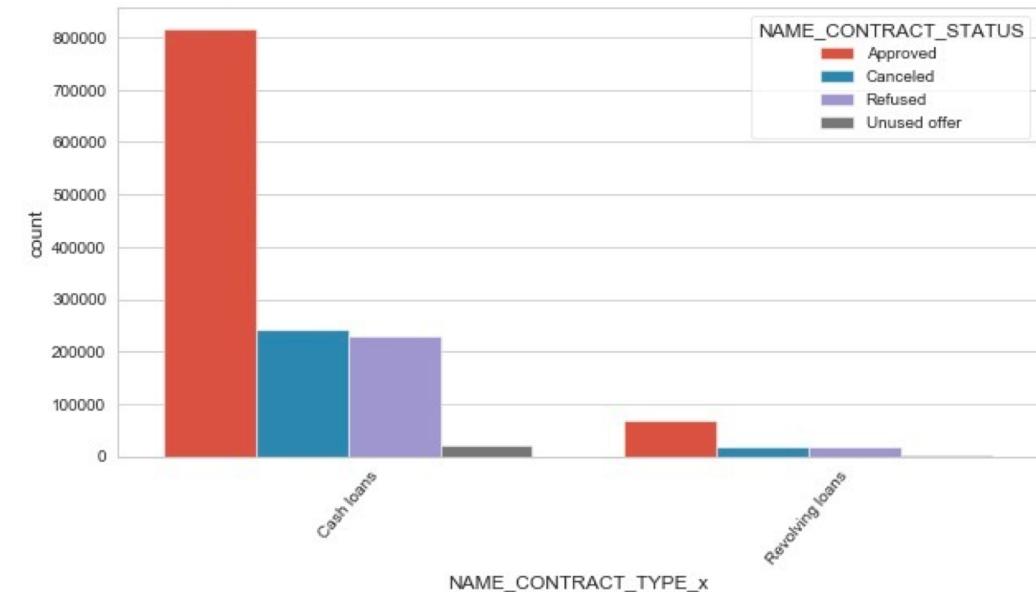
- ANALYSIS FOR GENDER FEATURE:
- COMPARISON ON BOTH MALE & FEMALE w.r.to NAME_HOUSING_TYPE in both Target – 0 & 1



CONTRACT TYPE

Inference :

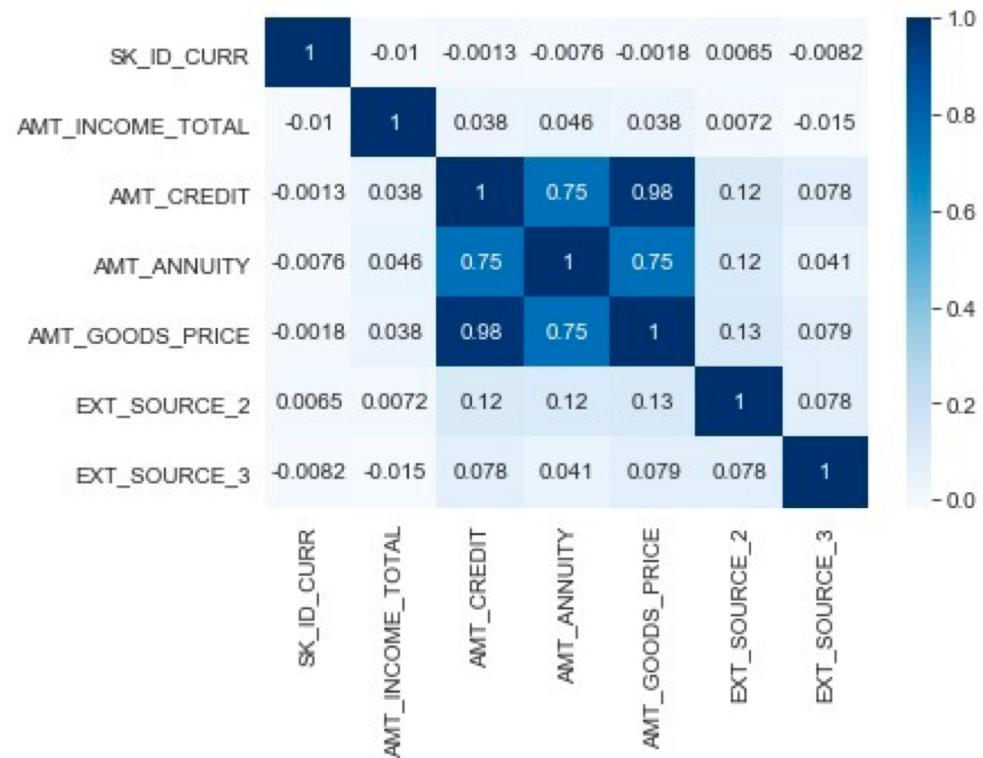
- When compared to the revolving loans cash loans are greater in percentages.



CORRELATION

TOP CORRELATED FEATURES ARE:

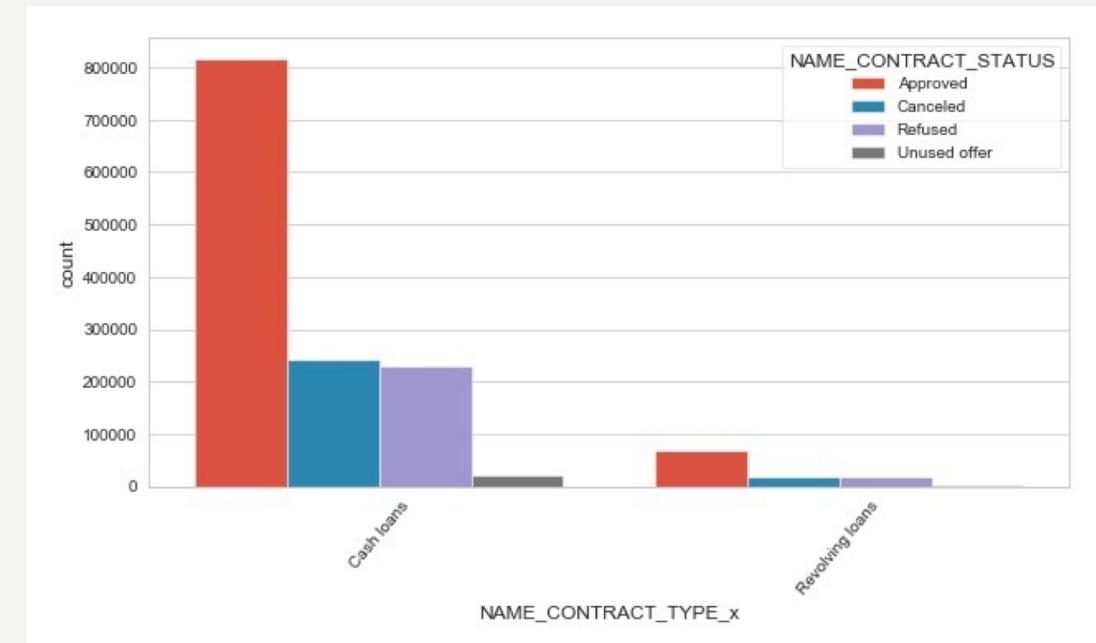
- I. AMOUNT_CREDIT VS AMOUNT GOODS PRICE
2. AMOUNT_ANNUITY VS AMOUNT_CREDIT



CONTRACT STATUS:

Inference :

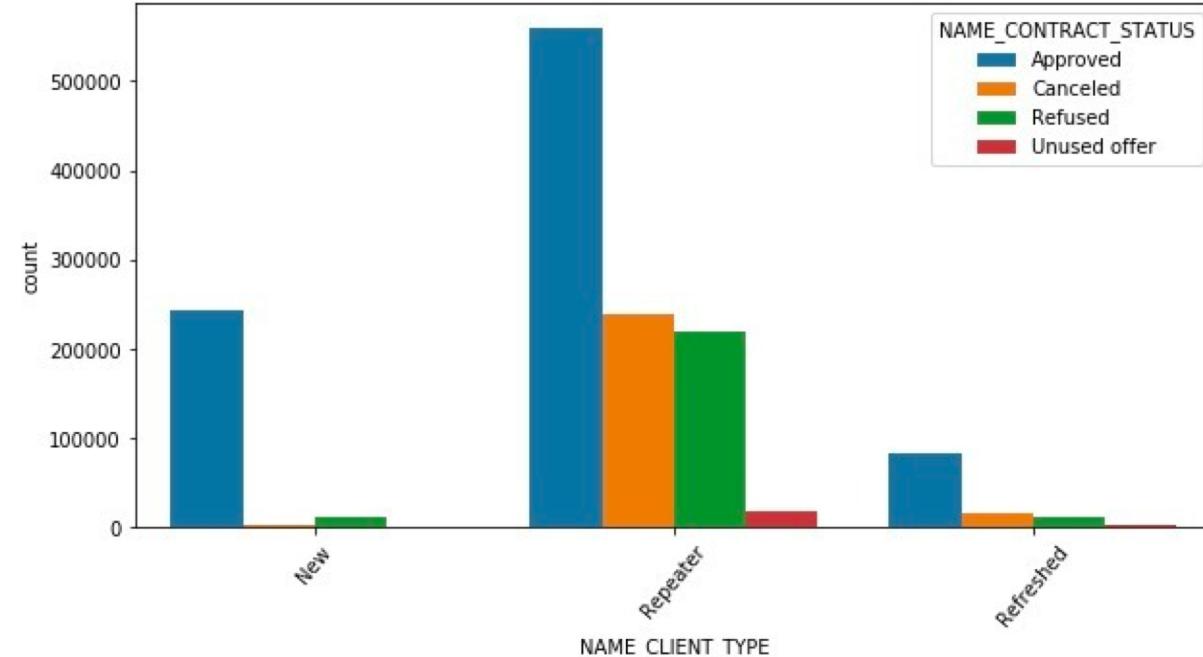
- we can see that around 85% of people's loan got accepted in previous application.

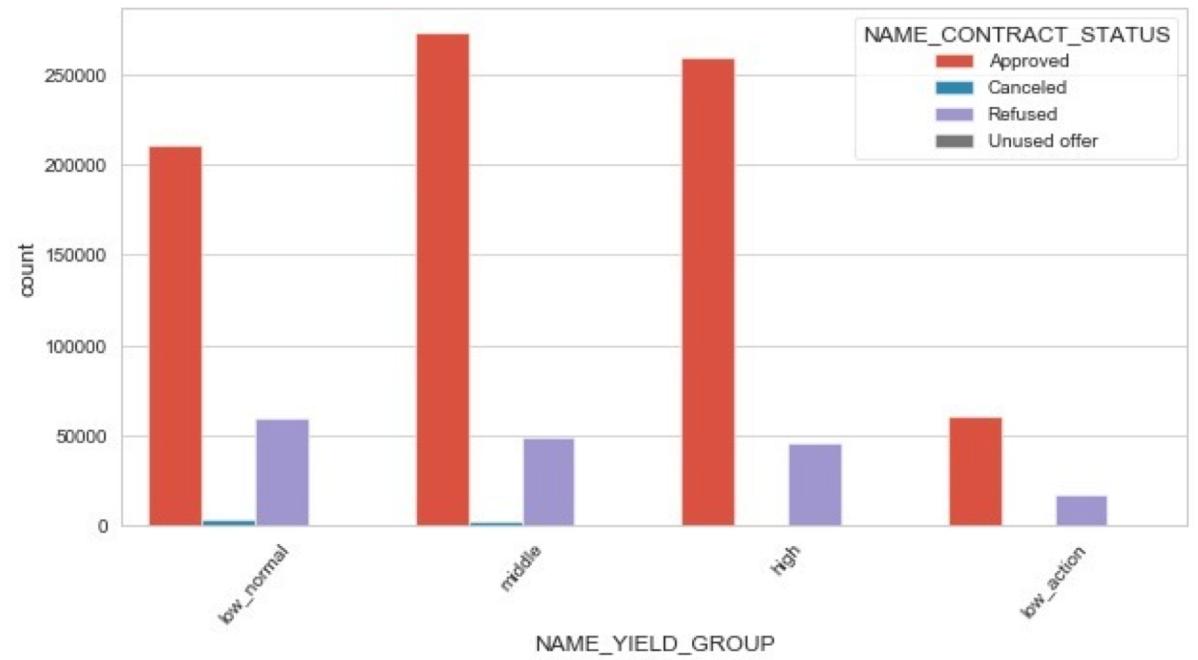


CLIENT TYPE:

Inference :

- High number of customers applied many times and got approved.*
- 22% of loans which are repeated
- New customers loans are approved around 25% for refreshed customers loans are approved below 10%.





YIELD GROUP:

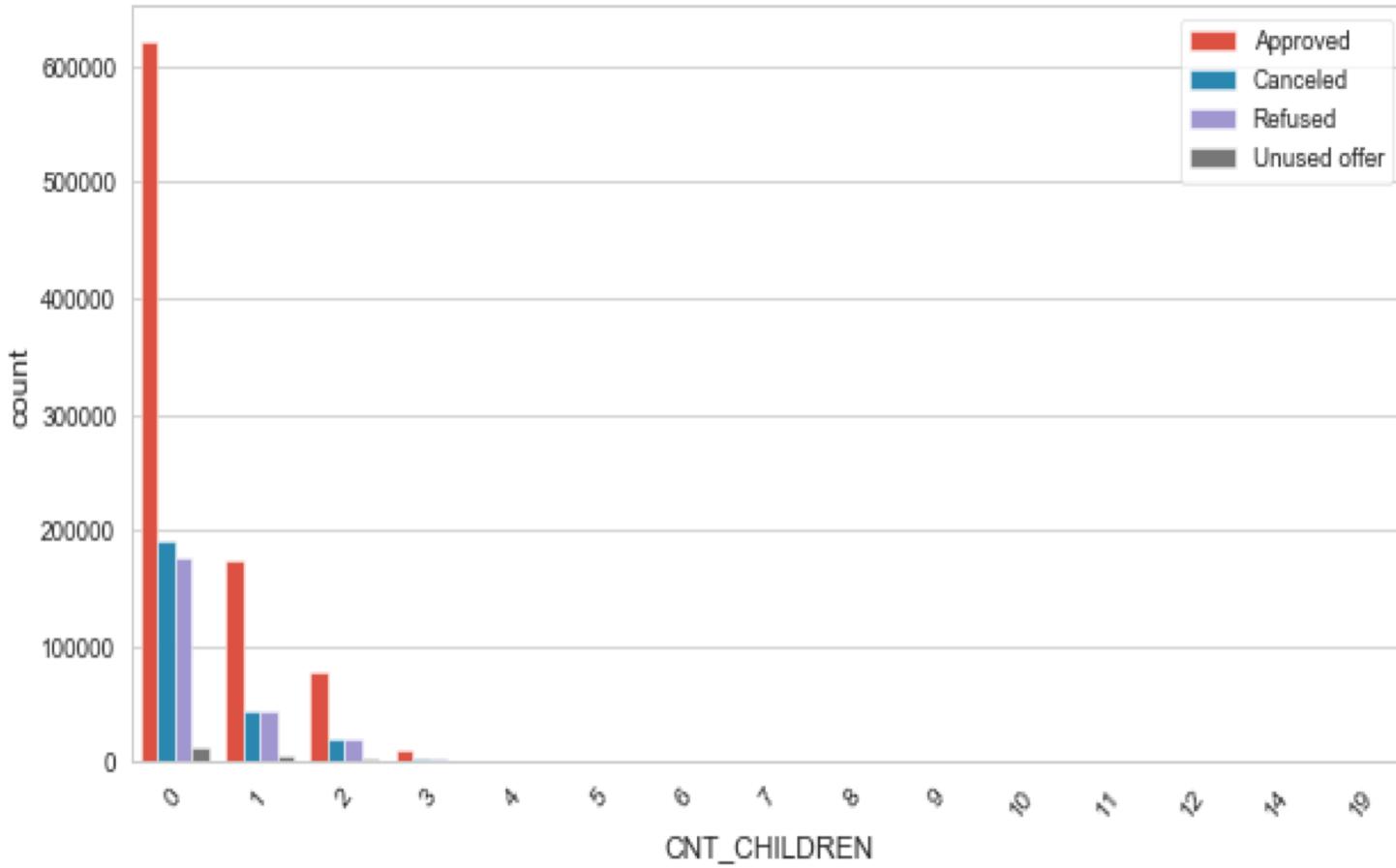
Inference :

Loan is approved for more
Middle

High people around 28%

Rejection is High for
low_action group people

Rejection is Moderate for
low_normal people



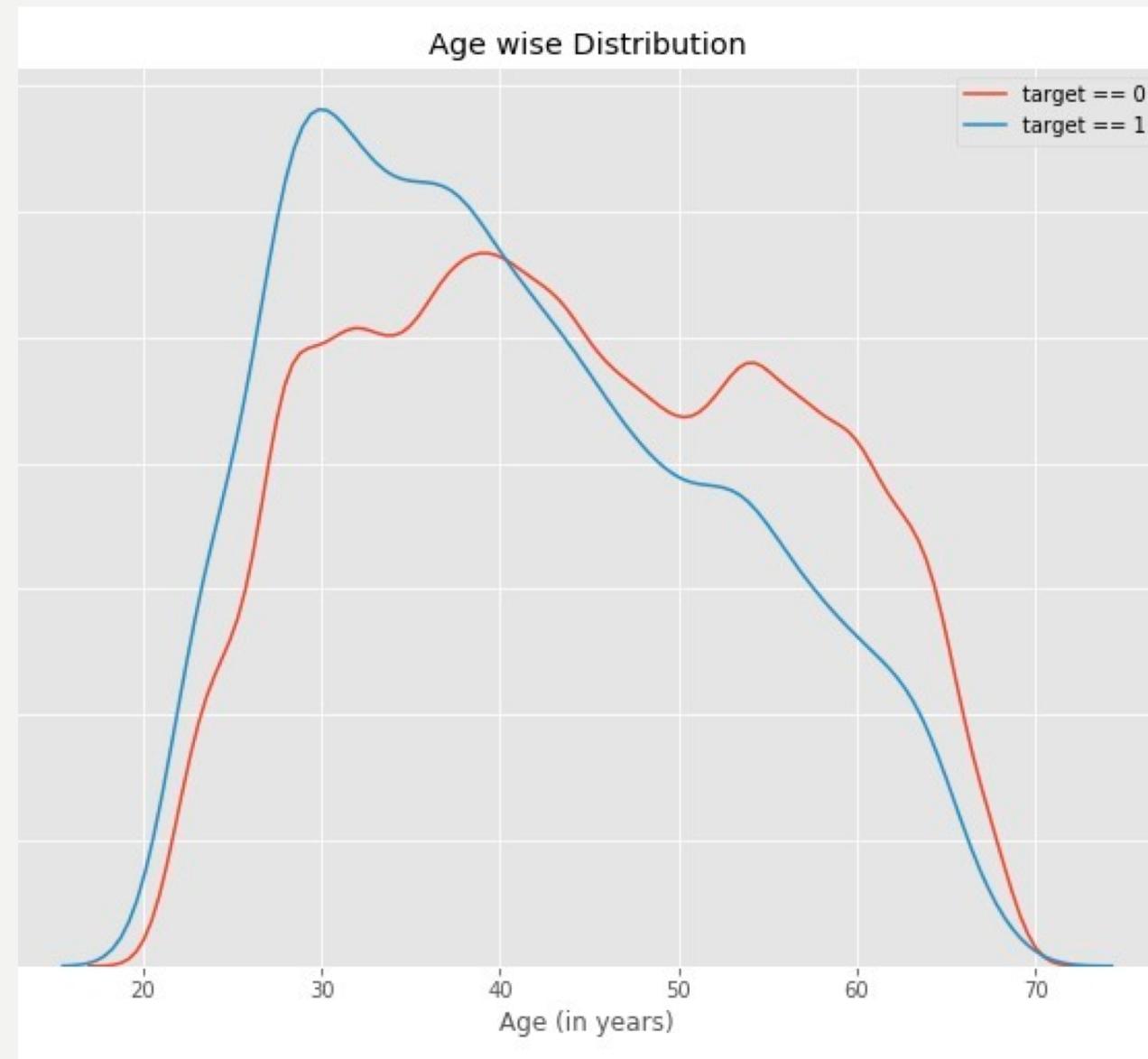
CNT CHILDREN:

Inference :

mostly the people who are having less
number of kids have applied loans

AGE DISTRIBUTION:

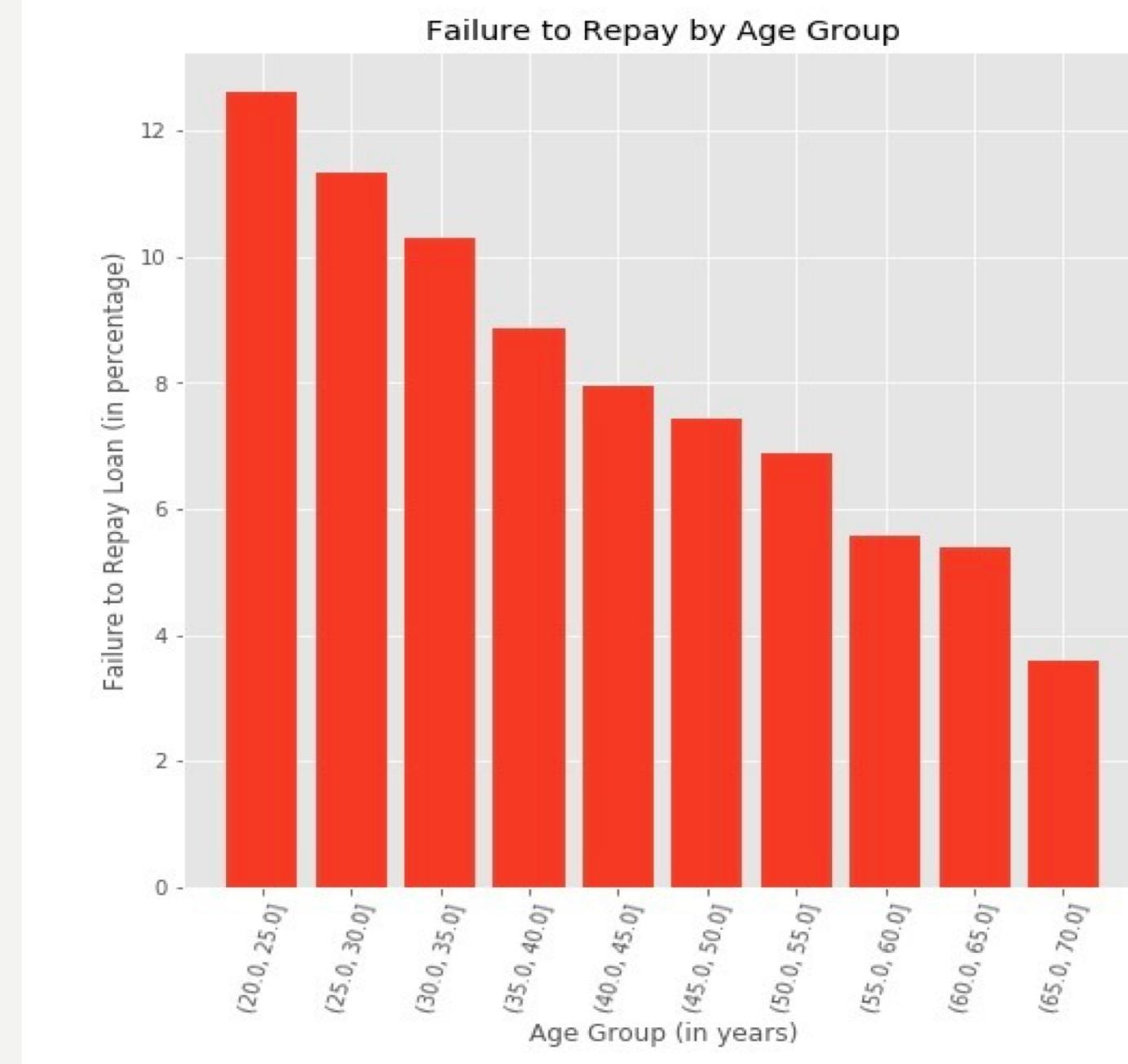
- Inference:
- Most defaulters lies on around the 28 age.
- Most on_time Loan payers are unevenly distributed



DIFFICULT TO REPAY LOAN BY AGE GROUP:

Inference:

- Most defaulters lies on in the 20-25 age group.





The end.

Thank You!