

Credit EDA Assignment

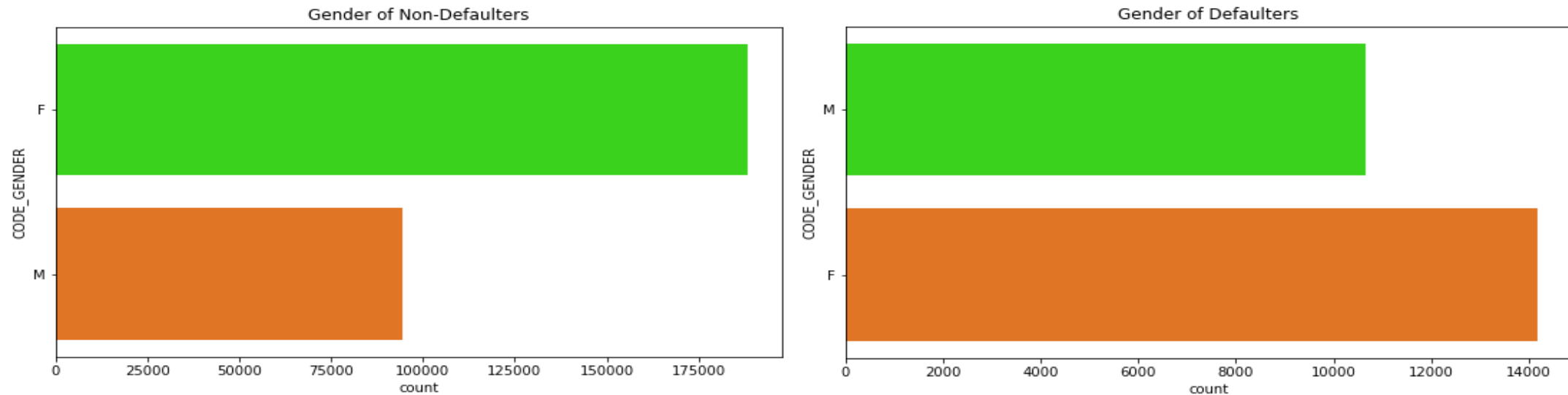
Problem Statement:

To identify the driving factors (or driver variables) behind loan default, i.e. The variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Univariate Analysis

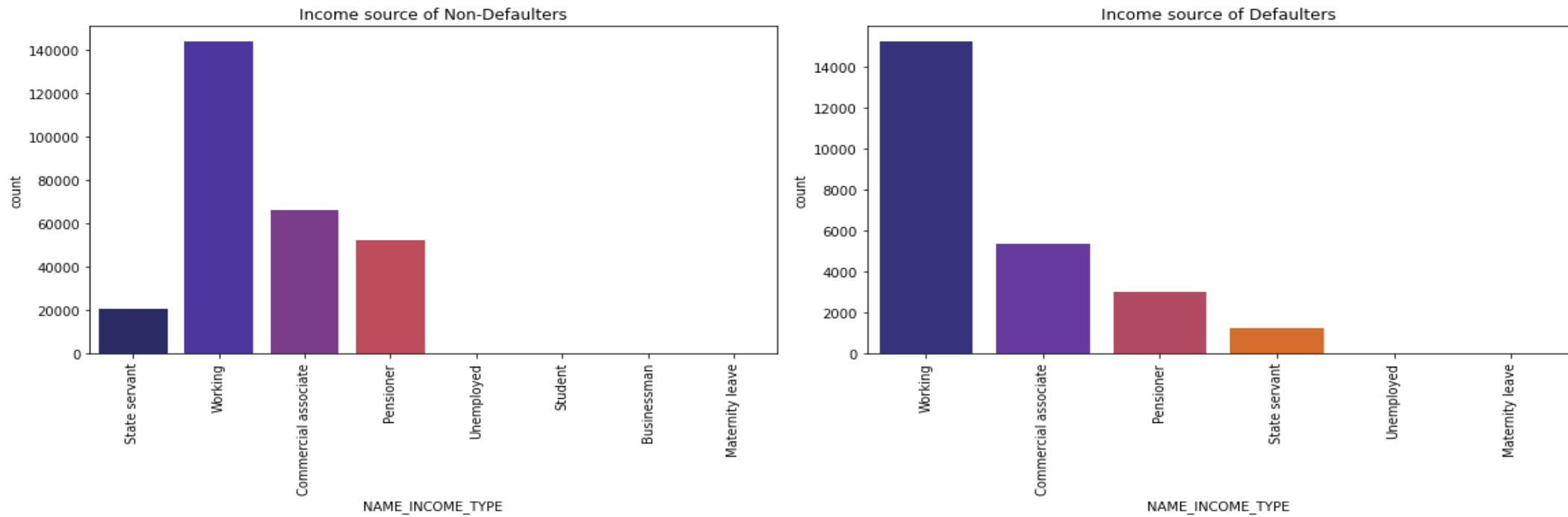
A. CATEGORICAL VARIABLES

1. Gender Wise



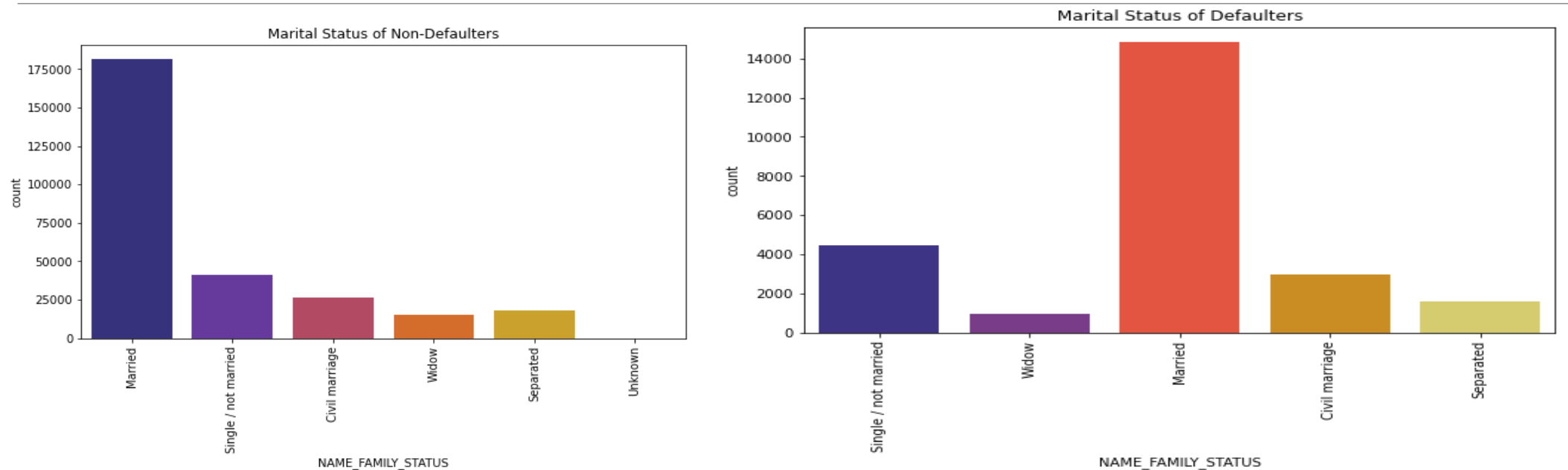
- The overall count of Female applicants is higher than Males, which translates into Higher number of Females in Default Case. But, in percentage Terms (wrt to total applications for that category) Males Default rate is Higher than those of Females.

2. Income Nature



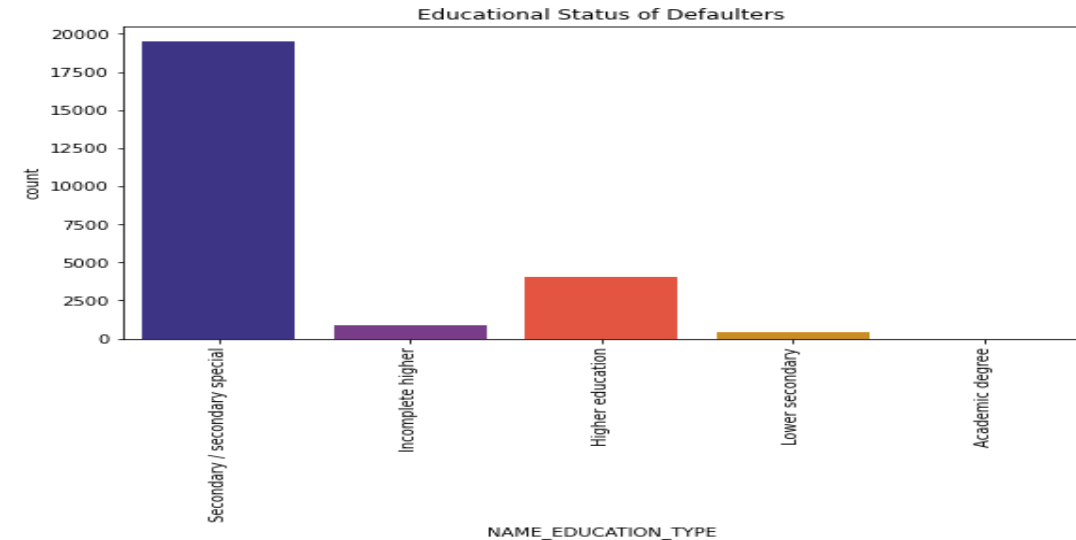
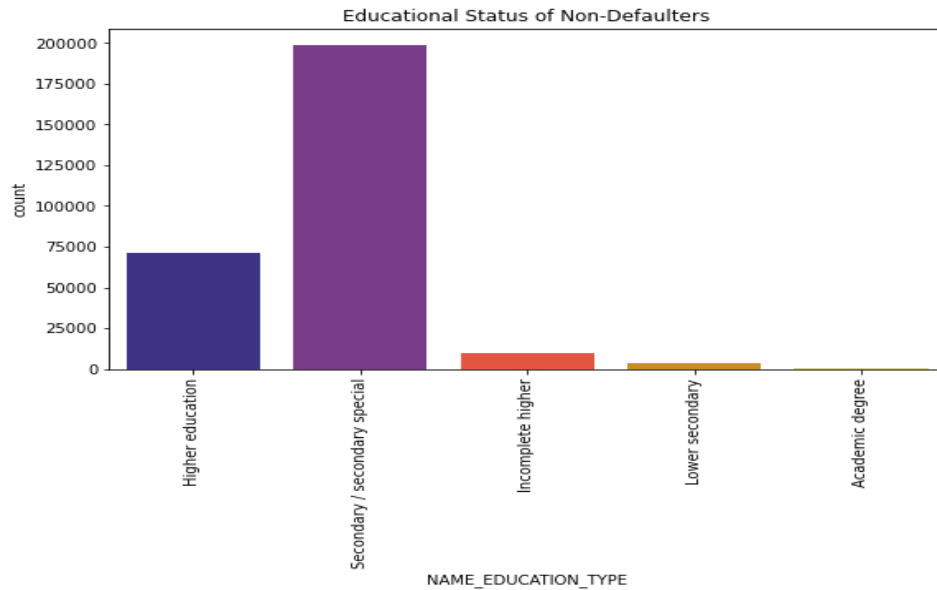
- Applicants belonging to “Working Class” forms the highest number, this subsequently translates into them being the highest number of Default Cases followed by Commercial associates.

3.Marital Status



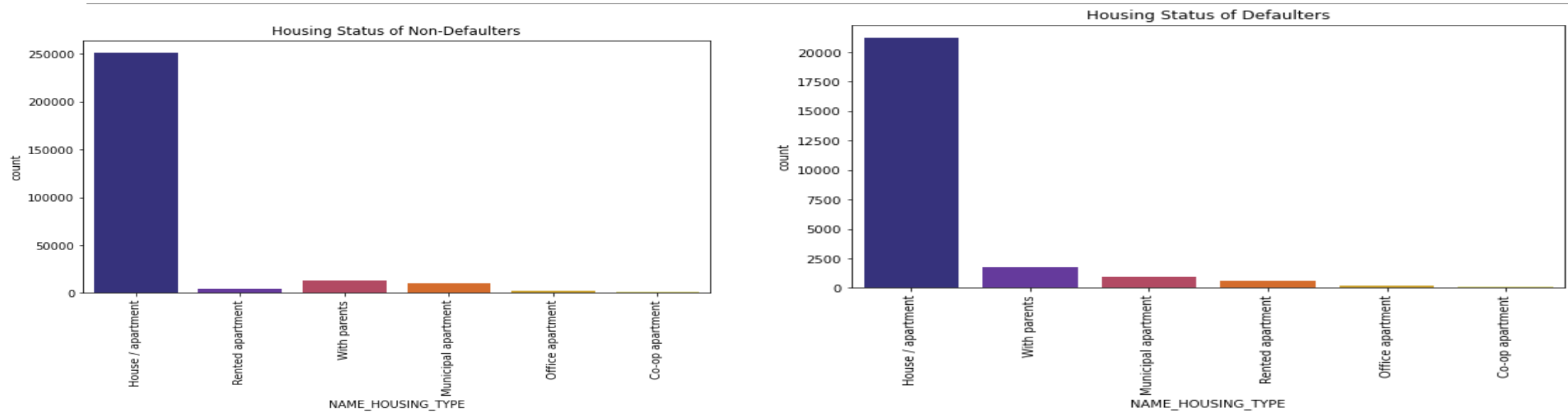
- Applicants from “Married” category form the highest number in case of both Default and Non Default criteria. Widows are least in number for both Default & Non- default Case.
- In percentage (wrt to total applications for that category) terms the ones belonging to “Civil Marriage” are highest in Default rate followed by those belonging to “Single” category.

4. Educational Status



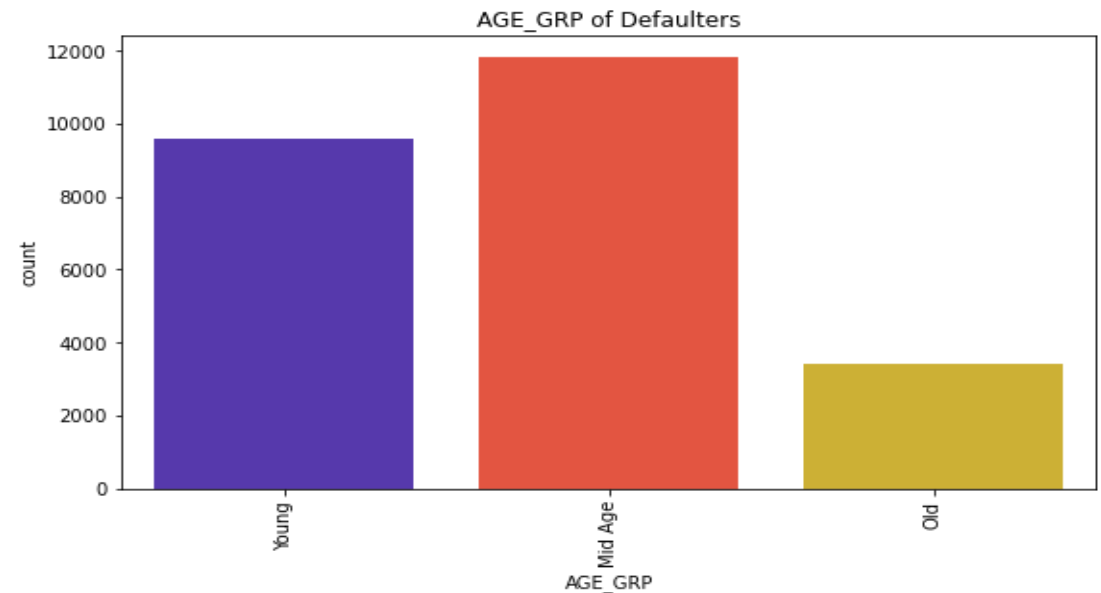
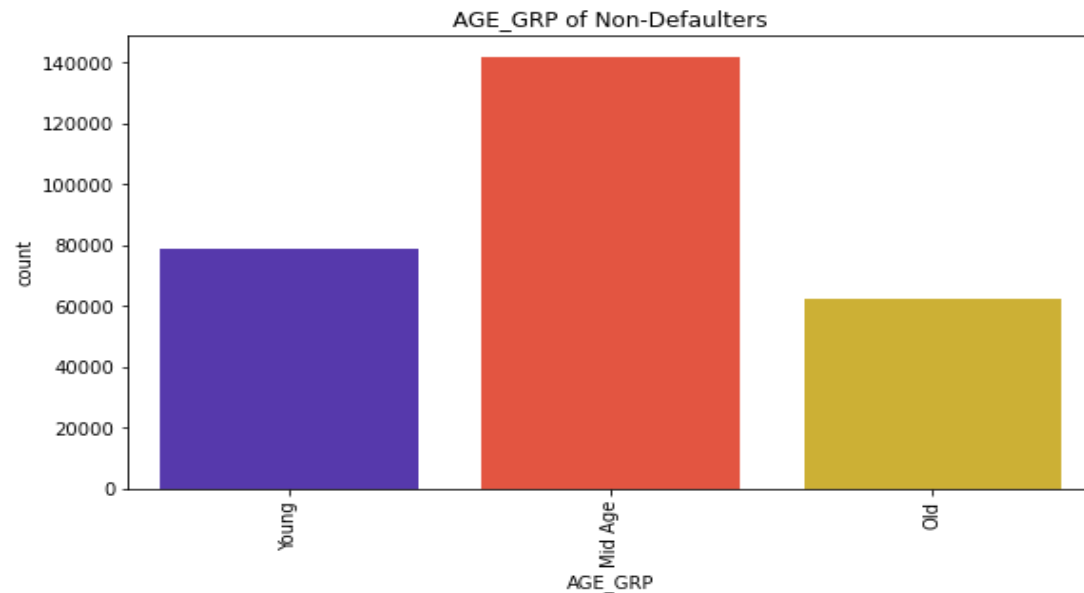
- Applicants belonging to group “**Lower Secondary**” in percentage wise (wrt to total applications for that category) terms forms the highest proportion in case of Default. they are followed by those belonging to “**Secondary/Secondary special**” category in case of Default.

5. Type of House Ownership



- Applicants with House ownership status as **“With Parents”** form the highest percentagewise (wrt to total applications in that category) default rate.

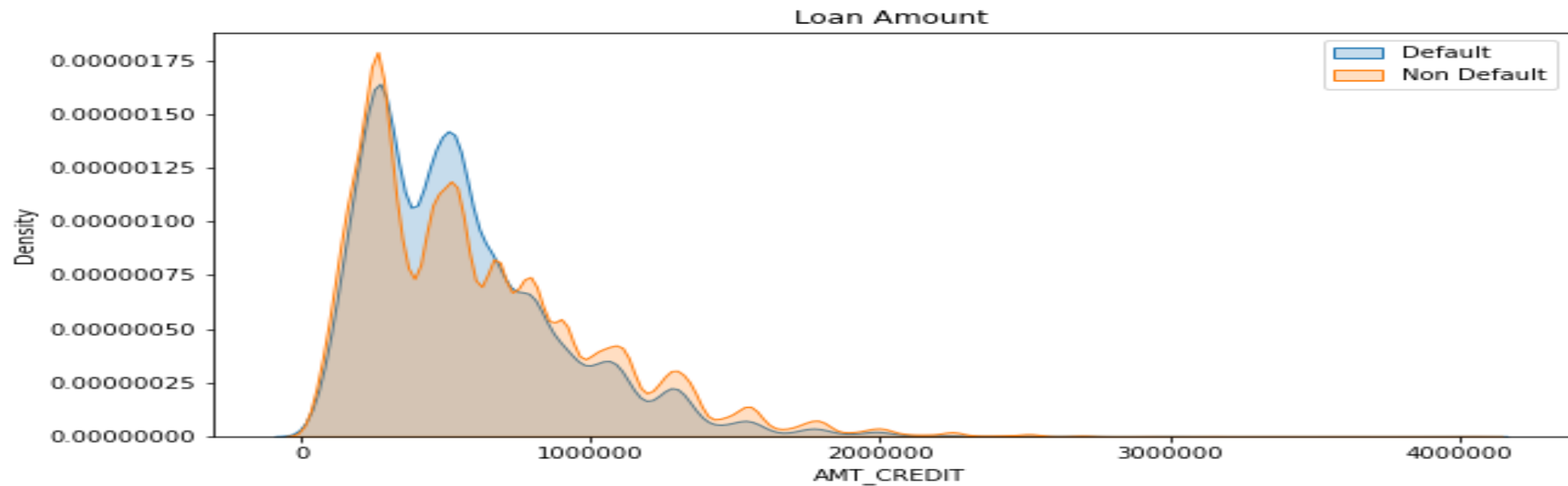
6. Age Group Wise



- Applicant's belonging to **“Young age group”** are the riskiest as proportion wise they form the highest percentage of default (w.r.t to no. of applications sourced in that age group)

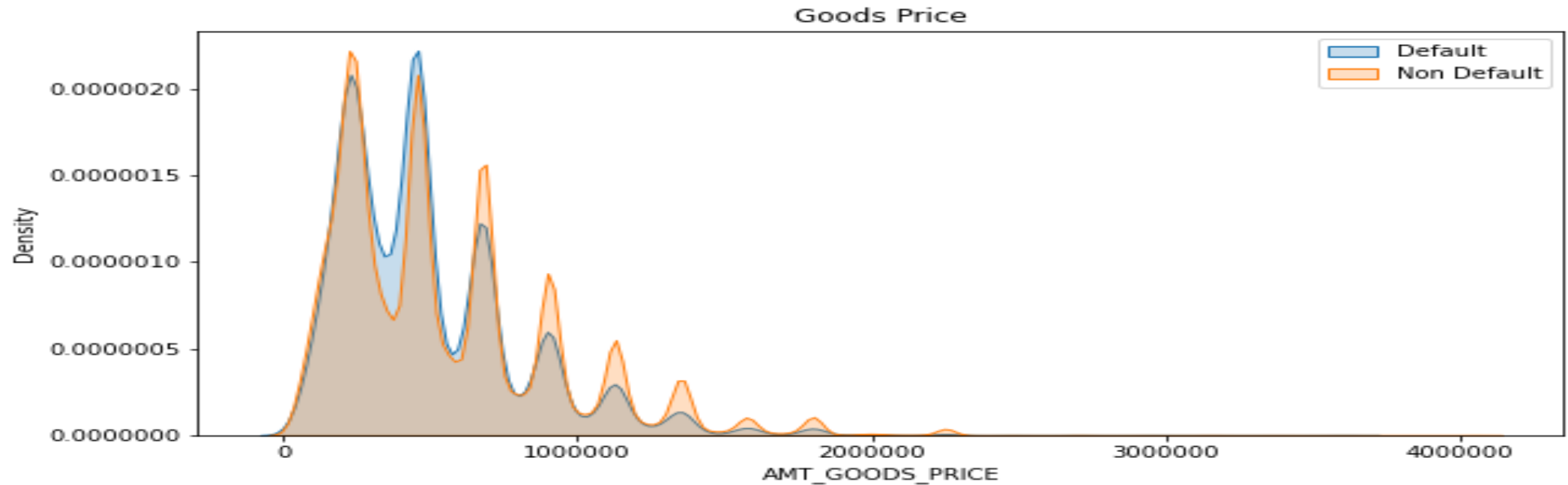
B. Univariate Analysis of continuous variables

1. LOAN AMOUNT



- It Appears that propensity to default is more towards lower range of Loan Amount.

2. Price of Goods

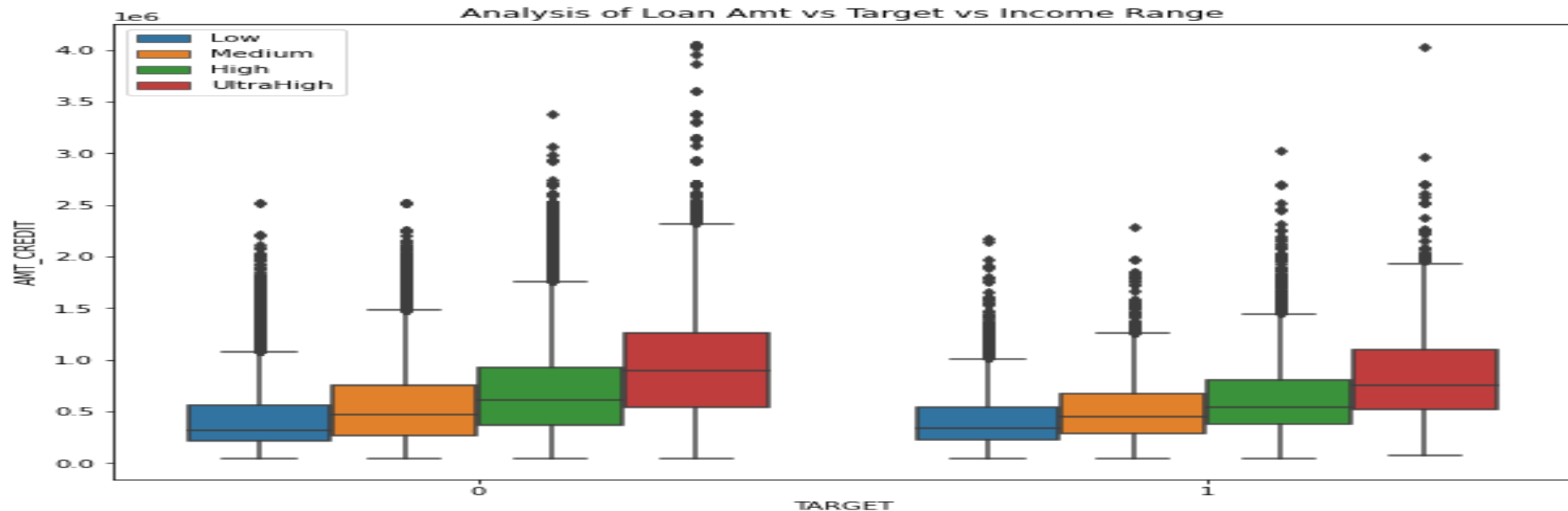


- The pattern of default to Non Default forms a continuous crest and troughs. Further, the default pattern goes in decline at high Goods Price showing people obtaining high amount of credit default less.

Segmented Analysis

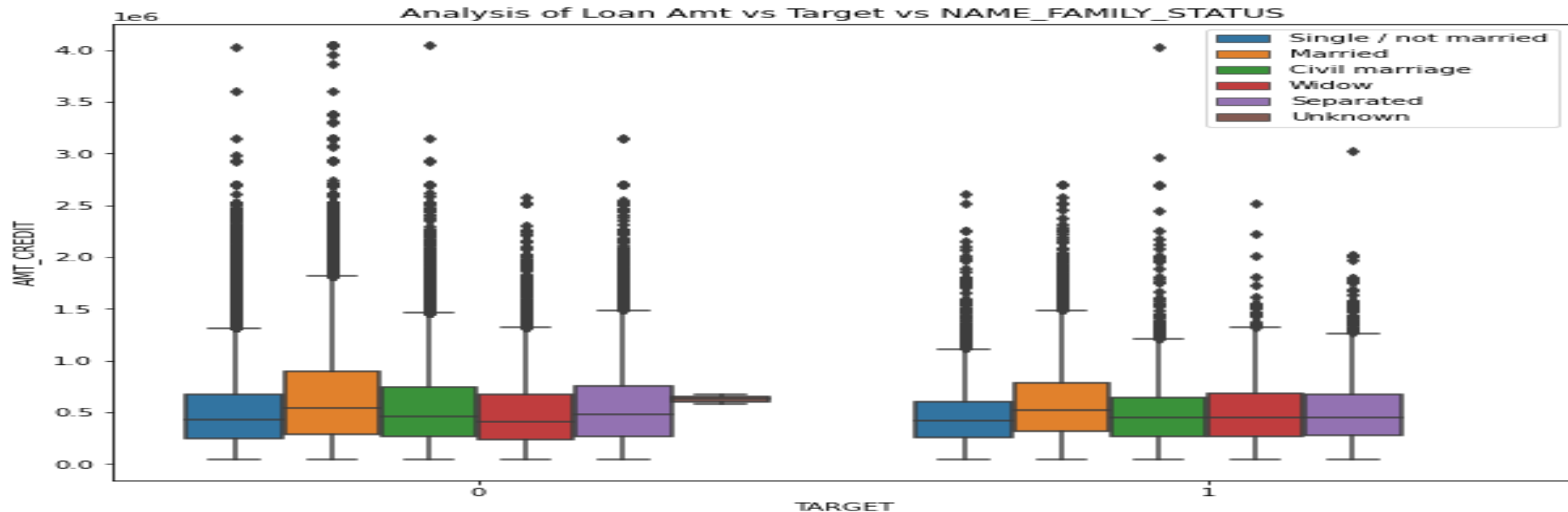
Bivariate Analysis

Loan Amount, Income Range, wrt Target Variable



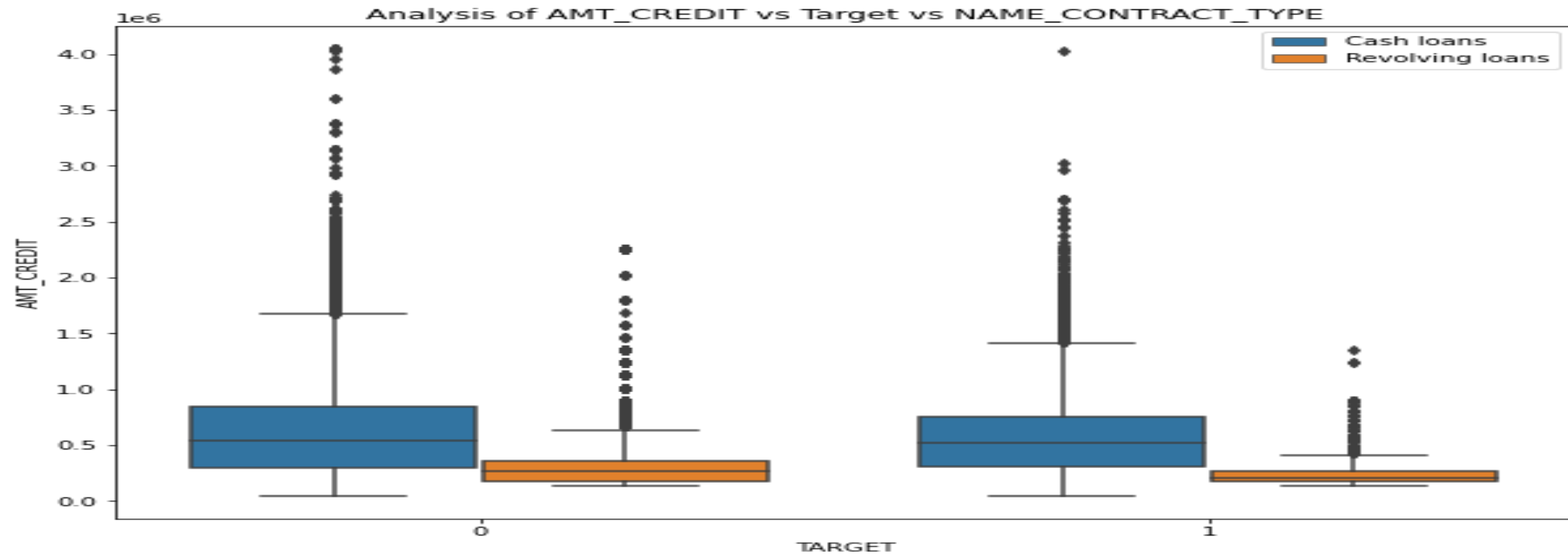
- As stated before the High Income group avail credit for Higher amount and have lesser default

2. Loan Amount, Family Status wrt Target Variable



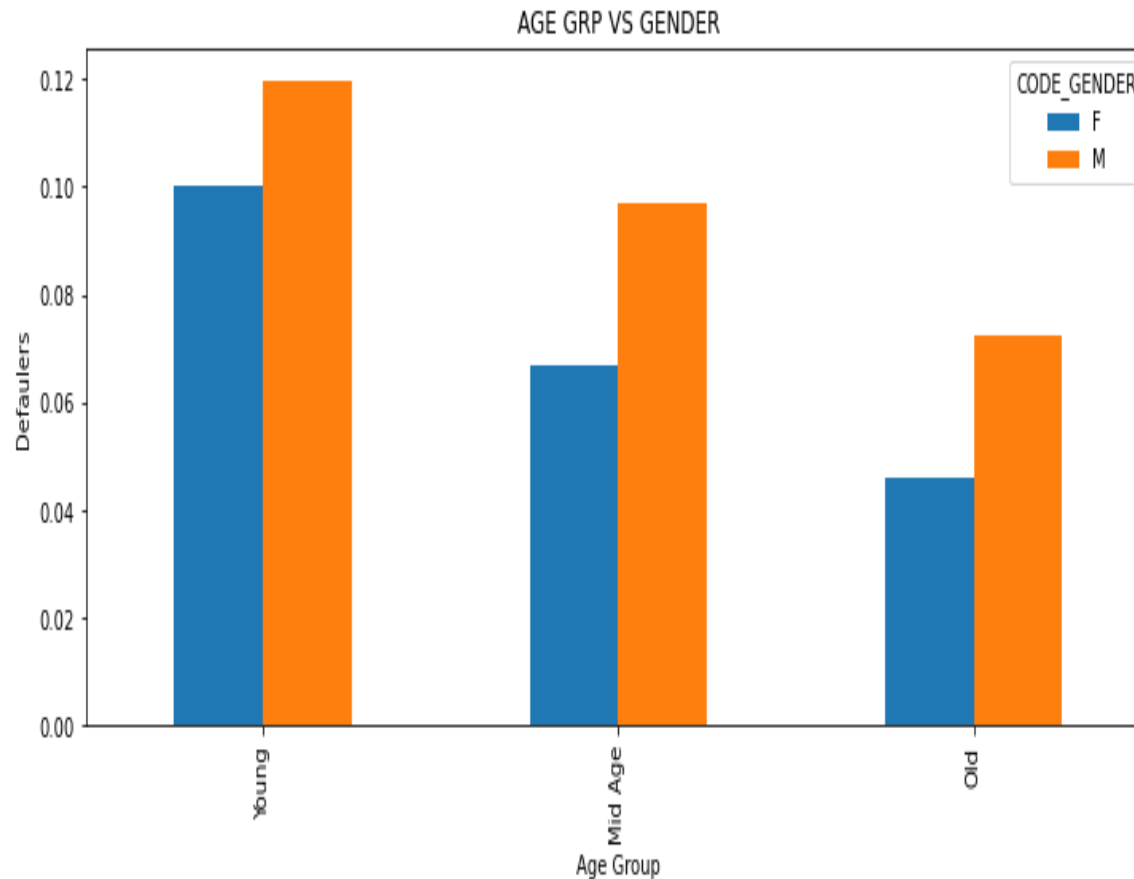
- In case of default applicants' belonging to "**Married**" category have higher median value for Loan Amt, whereas "Civil Marriage", "Widow" & "Separated" have same median Loan Amount (in case of Default). therefore, a cautious approach is to be followed while engaging with such clients.

3. Loan Amount, Type of Credit wrt Target



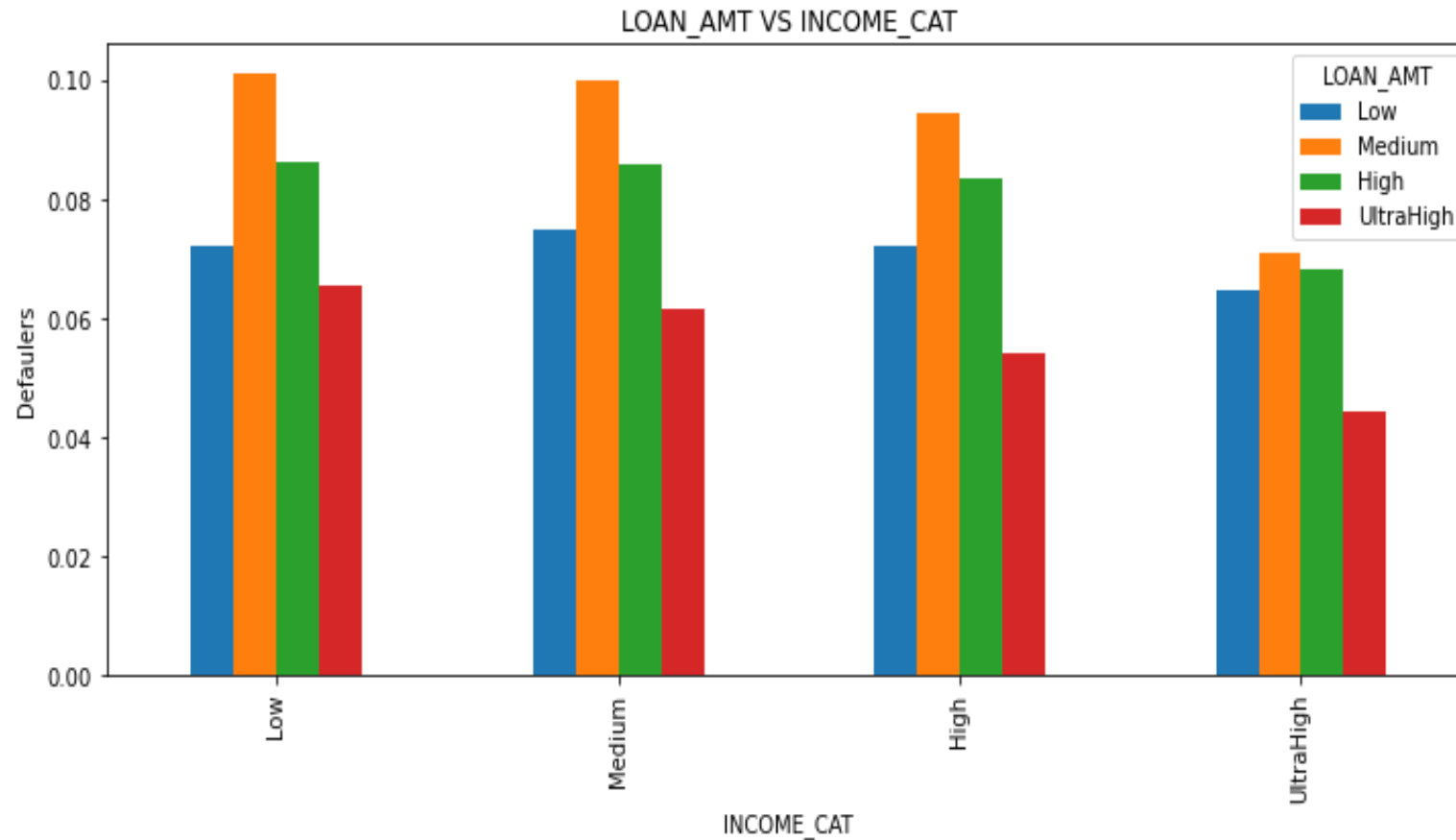
➤ Cash Loans Have Higher Default than Revolving Loans

4. Age Group, Gender wrt Default Category



- Young male clients are more in number under default category and are likely to have more default propensity

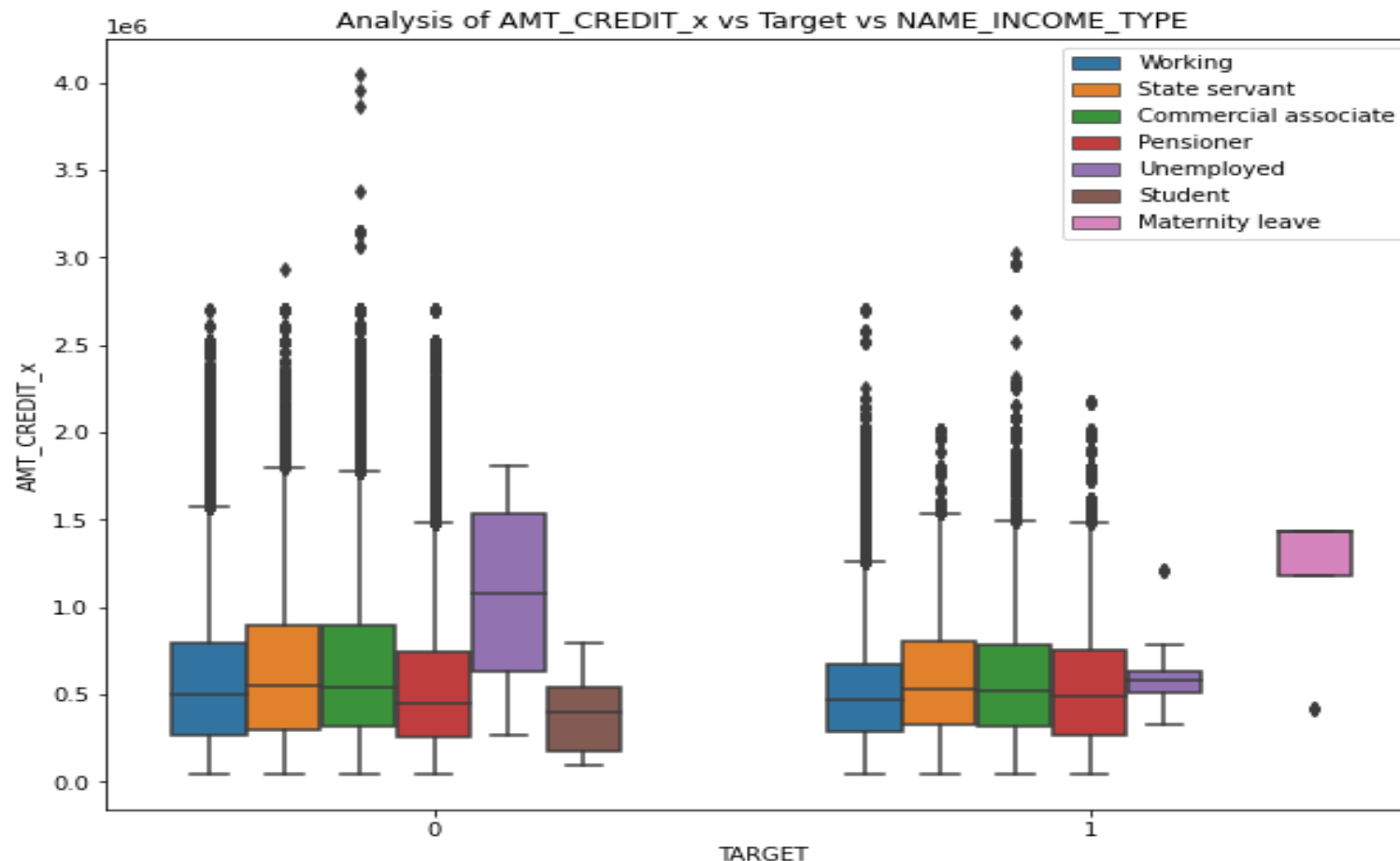
5. Loan Amount, Income Category wrt Defaulters



- In all income ranges applicants belonging to Medium loan Amt range i.e. AMT_CREDIT falling between percentile range 25% to 50% consists of highest no. of defaulters. followed by High Loan Amt i.e. percentile 50% to 90%. applicants falling in these ranges should be monitored.

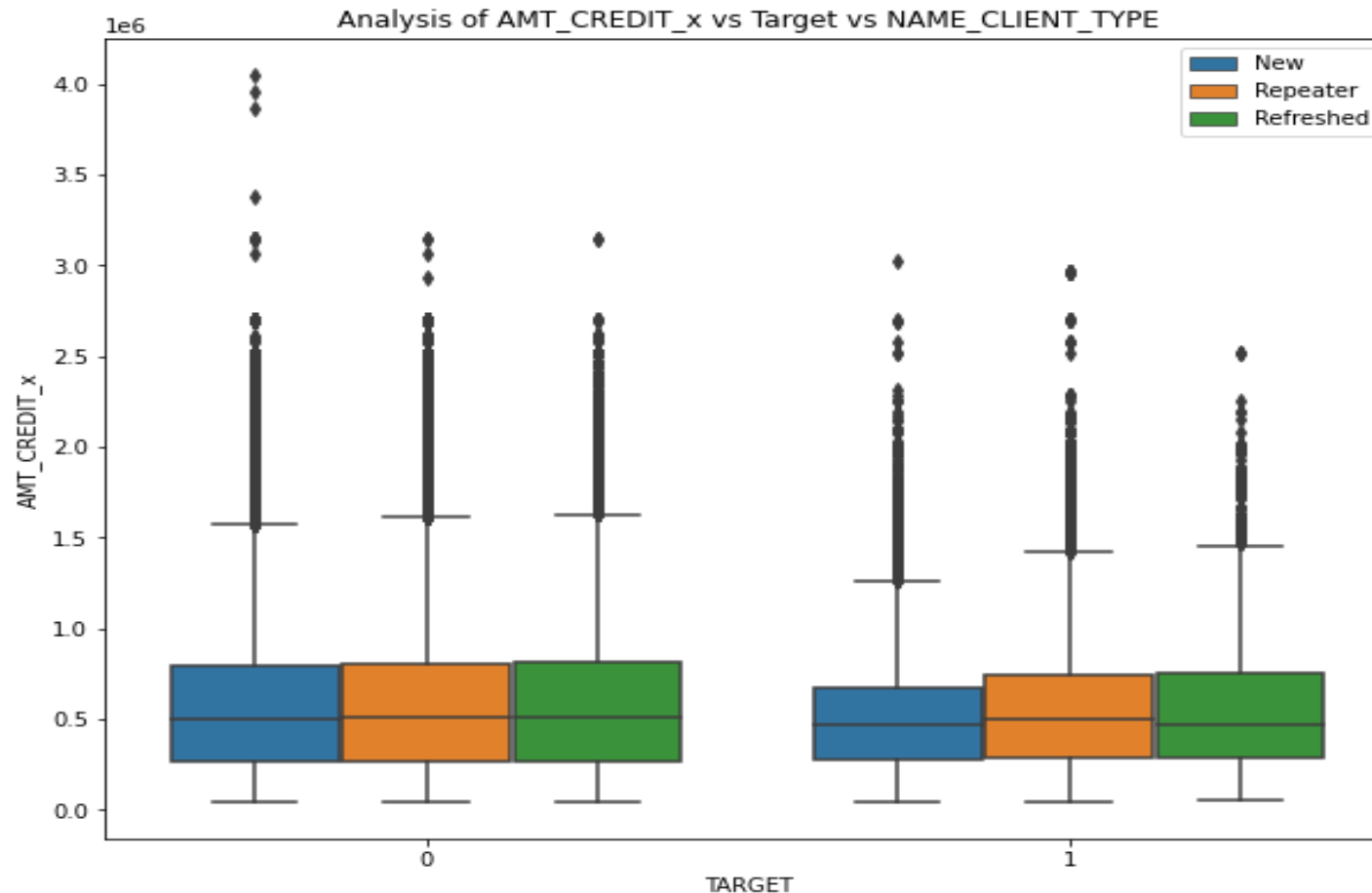
Merged Data Analysis

1. LOAN_AMT vs Target Vs Name_Income_type



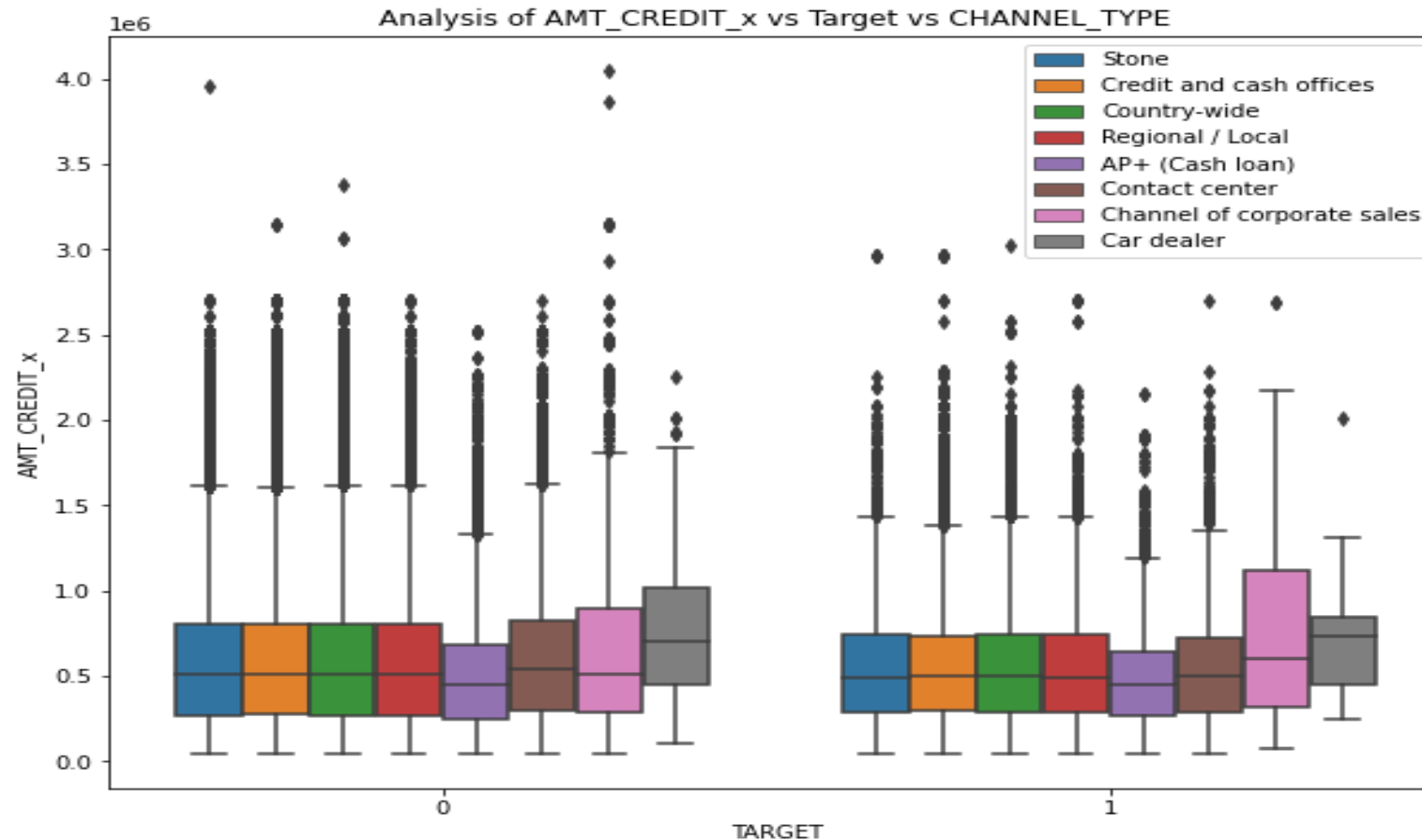
- Clients belonging to Income Cat "Maternity Leave" almost clearly fall in Default category and none in Non Default category. henceforth, while processing such applicants should be dealt cautiously.
- On the other hand "Unemployed" form the Highest range in Non-Default Category

2. AMT_CREDIT vs Target VS NAME_CLIENT_TYPE



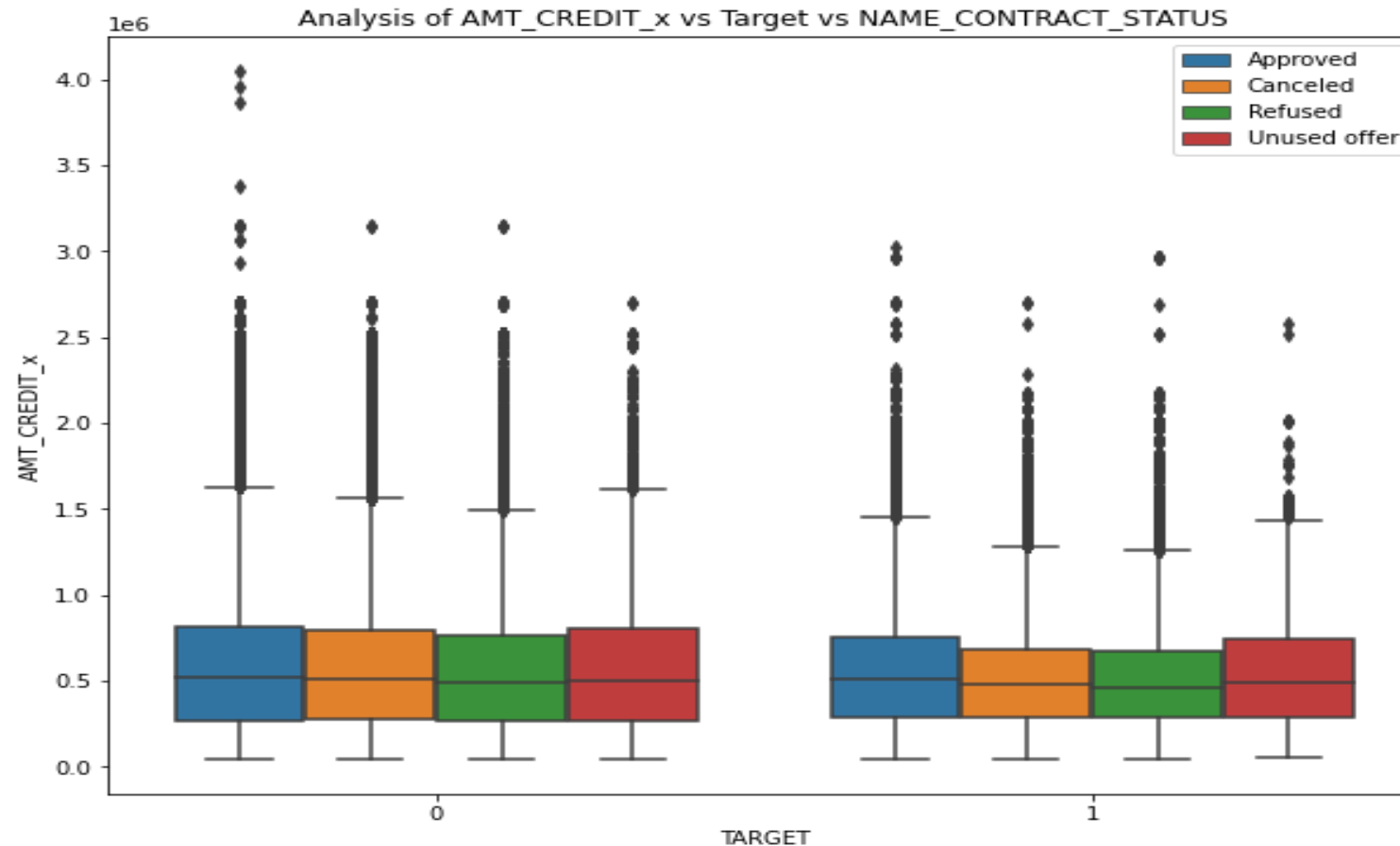
- the distribution in Non Default category is same for all client type, depicting that non default customers have similar propensity towards credit amount offtake irrespective of client type.
- For Default Refreshed and Repeater are more than New applicants.

AMT_CREDIT vs Target VS CHANNEL_TYPE



➤ the customers who were previously acquired through "Channel of Corporate Sales" are maximum in default. such require special monitoring.

AMT_CREDIT vs Target VS CHANNEL_TYPE



- Previous Applications "Contract status" has not much impact on the Default and Non Default status for current applications.