CREDIT EDA ASSIGNMENT

By Sharad Choudhury
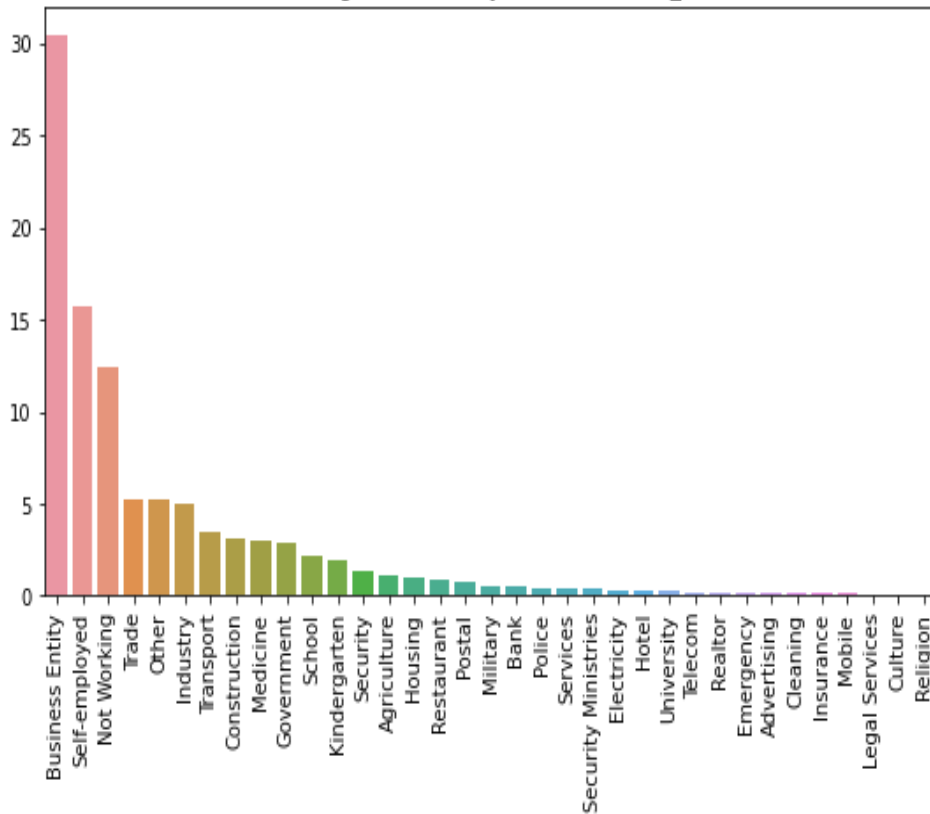
CURRENT APPLICATION DATA
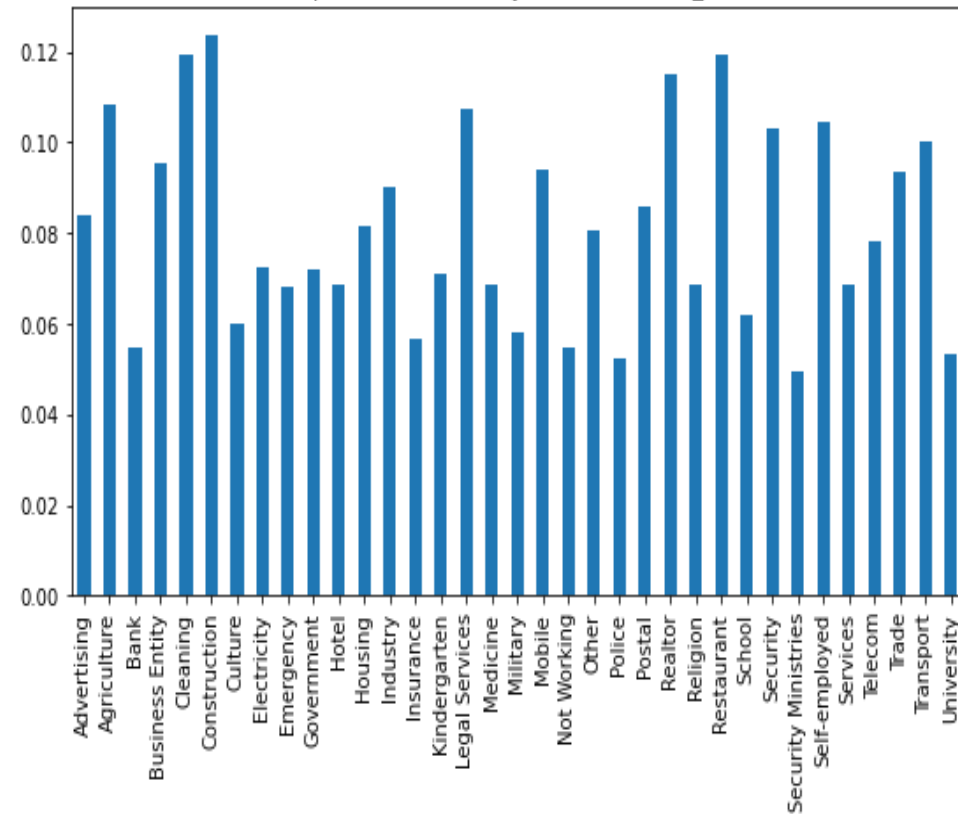
UNIVARIATE ANALYSIS
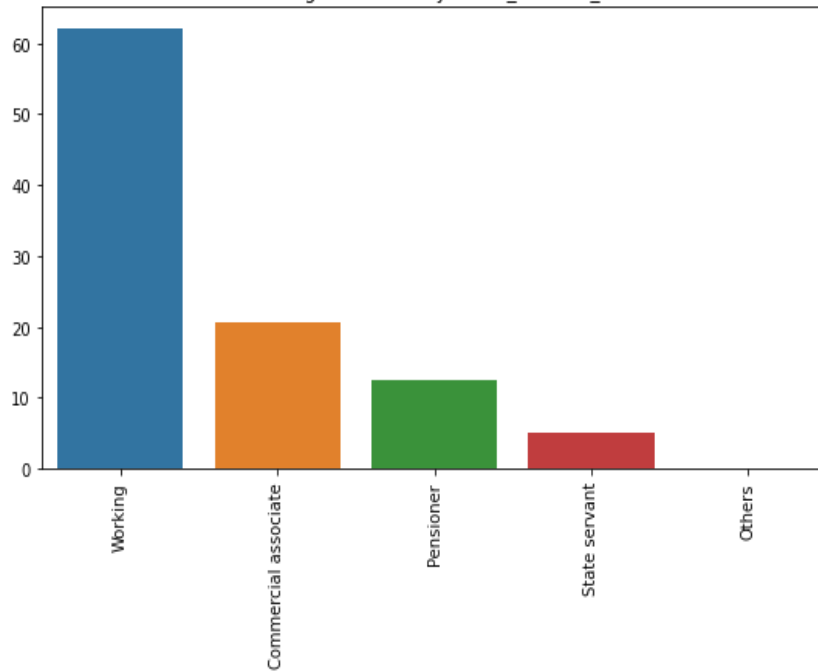### DATA OF DEFAULTERS BY PERCENTAGES AND PROPORTIONS OF EACH CATEGORY WITHIN COLUMNS
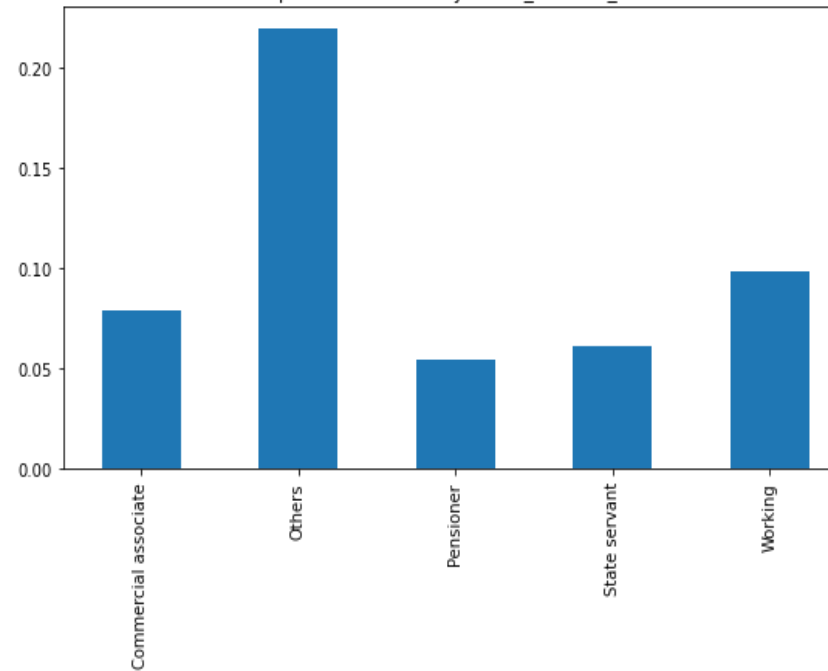

- Highest proportion of defaulters are from the Construction, Restaurant, Cleaning, Agriculture industries.

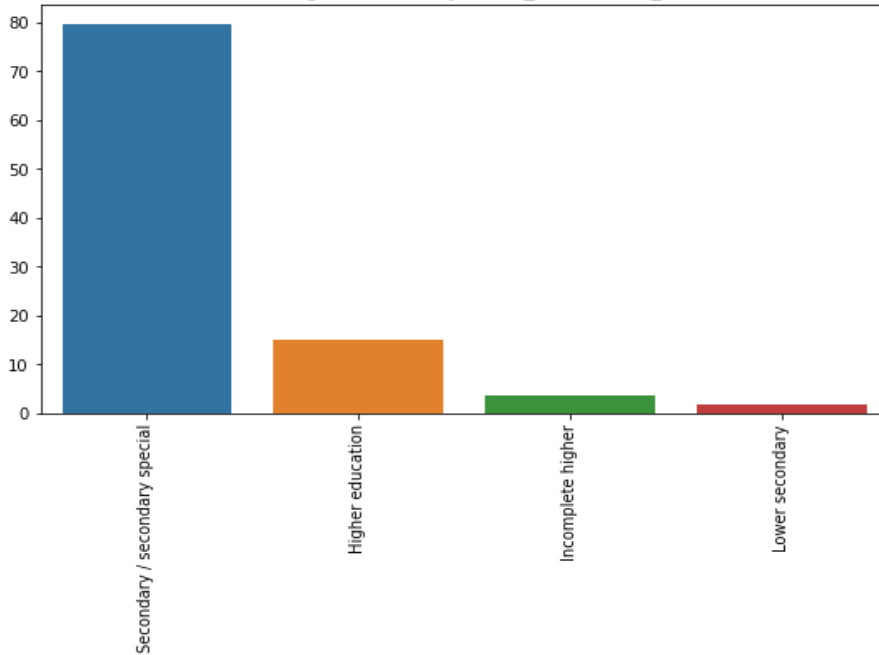- Lowest proportion is from University, Police, Security ministries and Bank industries.
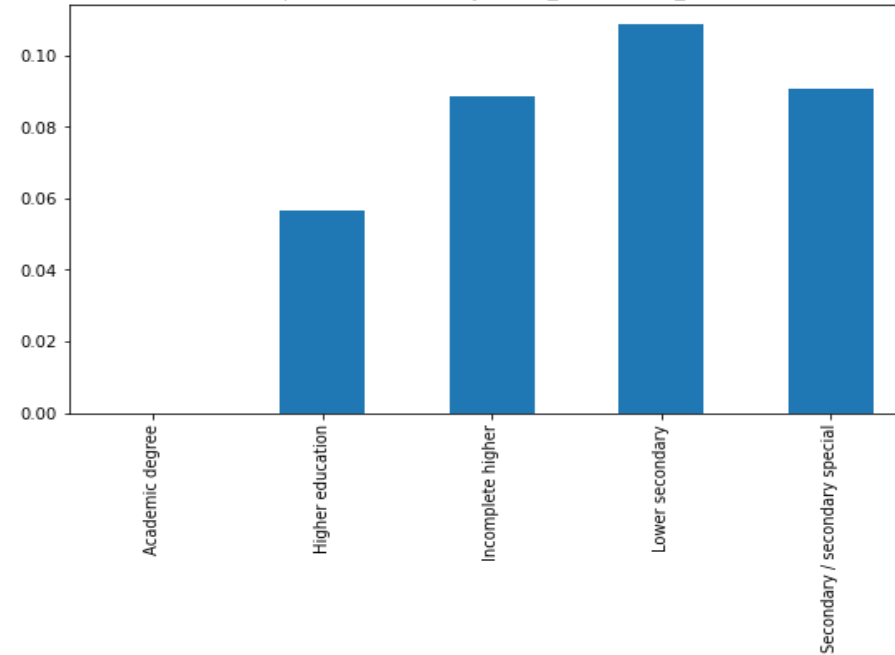
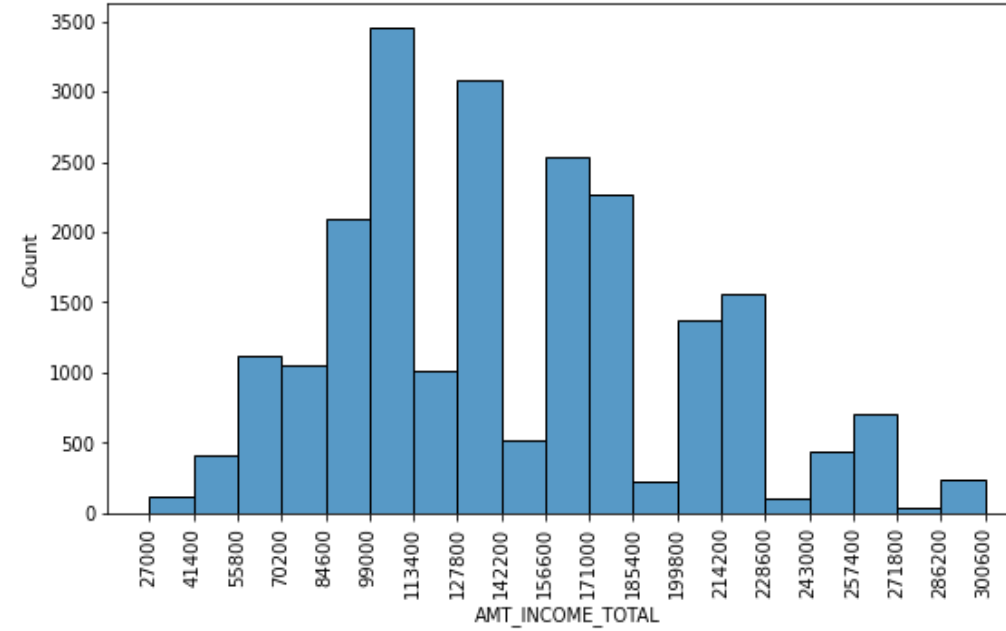- Most defaulters are working but higher proportions from Others (Student, Businessman, Unemployed).

- Pensioners have the lowest proportions.

- Most defaulters have Secondary level education, but highest proportion of defaulters are from Lower secondary level.
- The proportion is low for clients with Higher education.
- No defaulters were with an Academic degree.
- This means more educated people are less likely to default.

Quantile distribution of AMT_INCOME_TOTAL
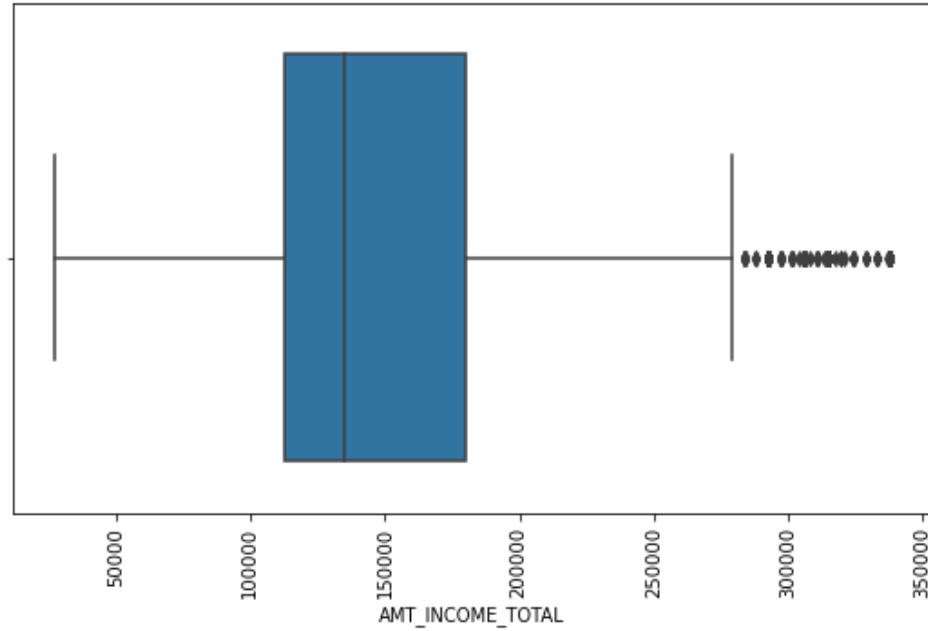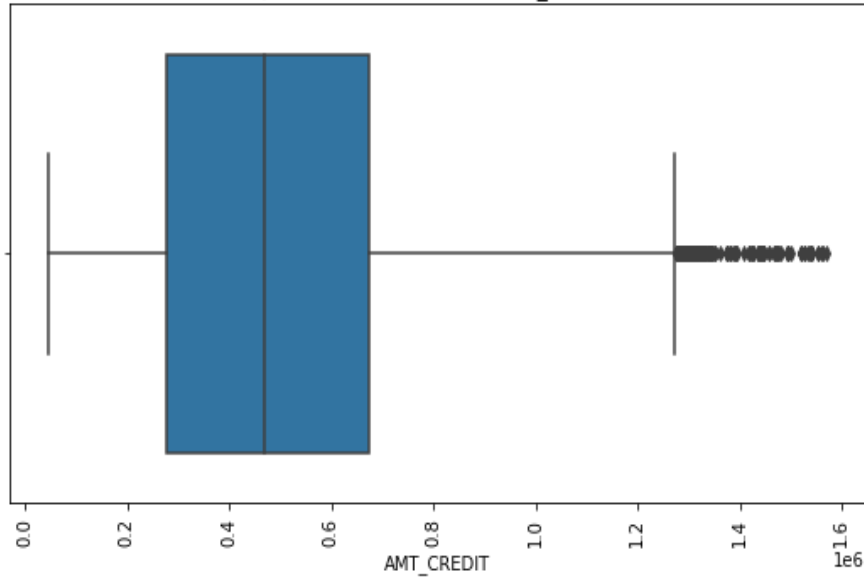
Frequency Distribution in AMT_INCOME_TOTAL

Percentage of clients by amt_income_groups

Proportion of clients by amt_income_groups

- 50% of defaulters earn below 1.3 lakh.

- Highest no of defaulters earn between 1 – 1.13 lakh
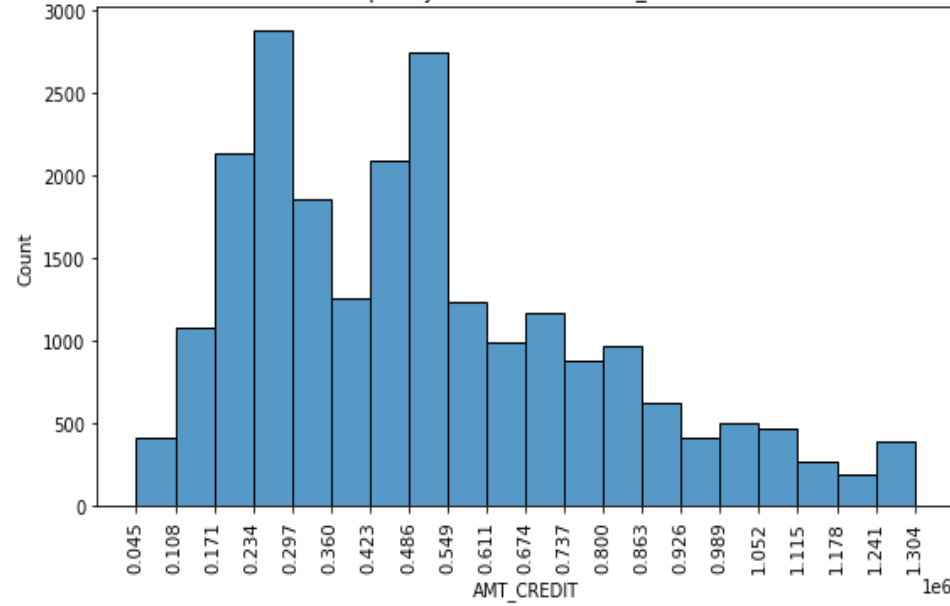
- Highest proportion of defaulters earn between 1-2 lakhs.
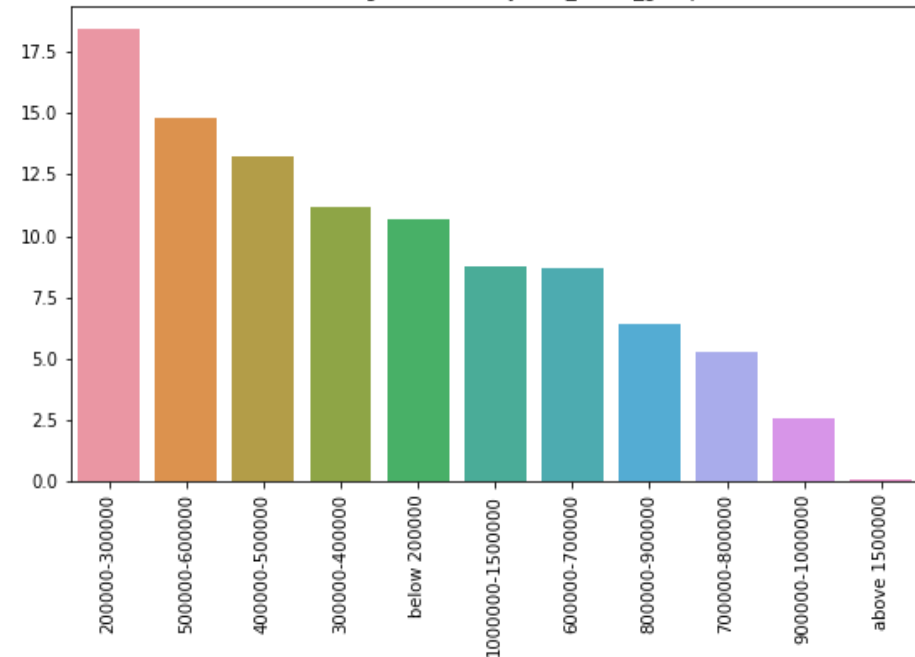
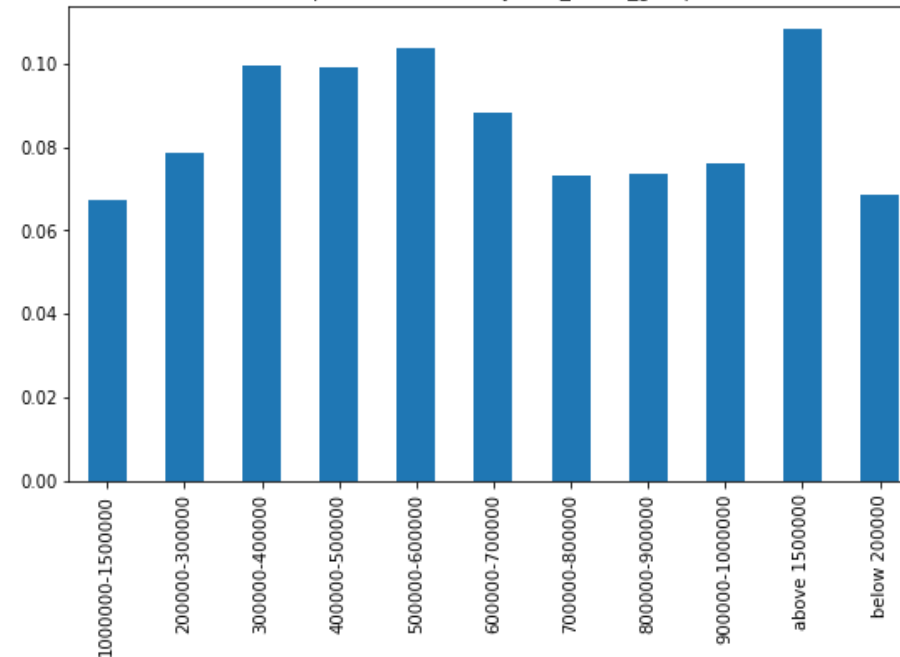- Clients with higher income less likely to default

- 50% of defaulters had credit amount below around 5 lakhs.

- Most defaulters had credit between 2-3 lakhs.

- But highest proportion of defaulters had Loan credit amount as 3-6 lakhs.

Percentage of clients by CODE_GENDER / Proportion of clients by CODE_GENDER / Percentage of clients by age_group / Proportion of clients by age_group
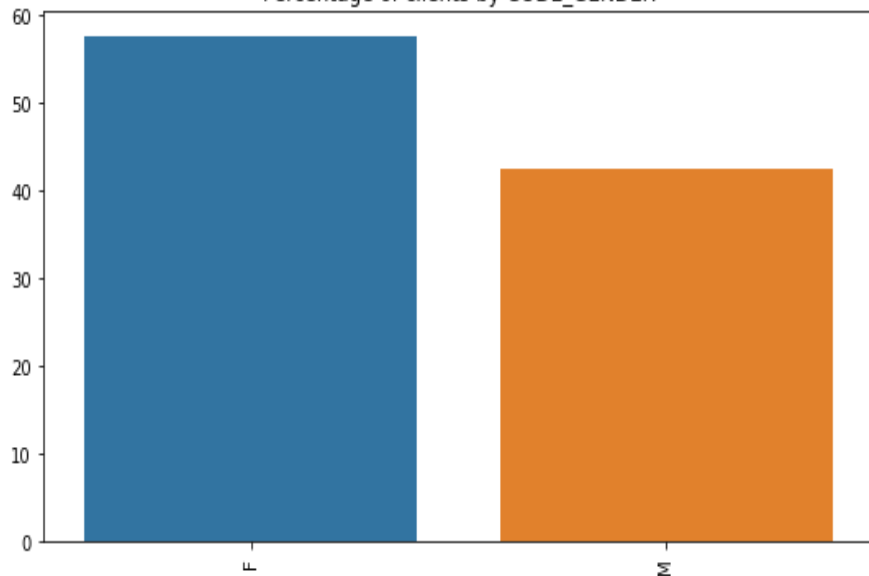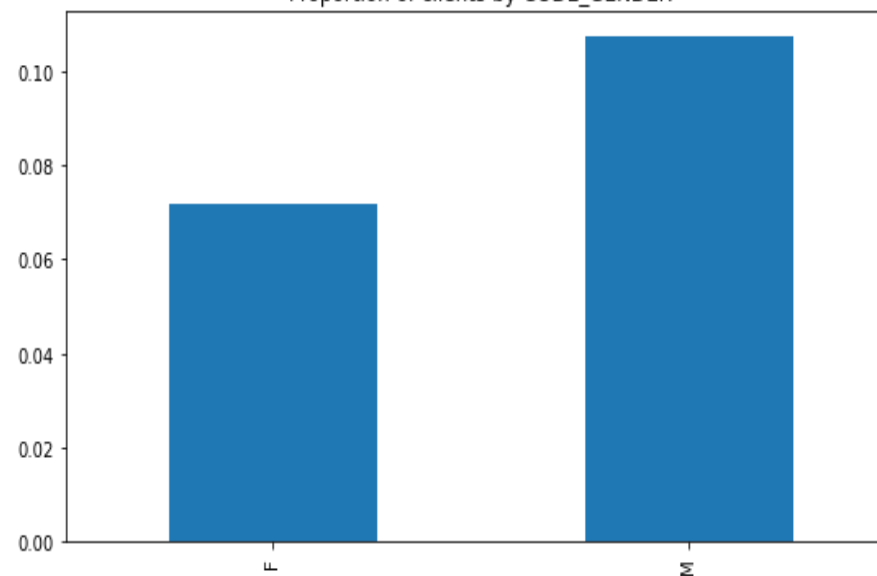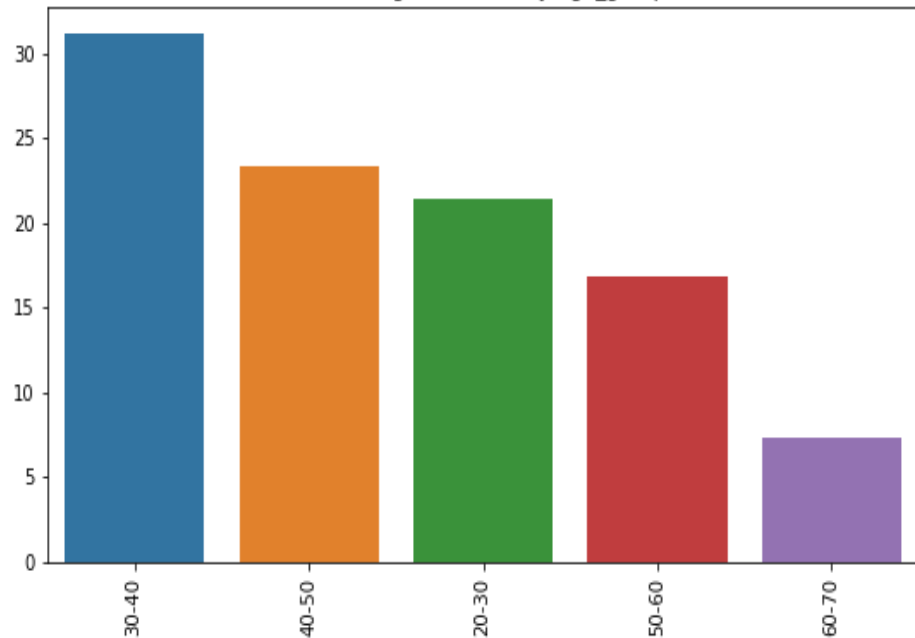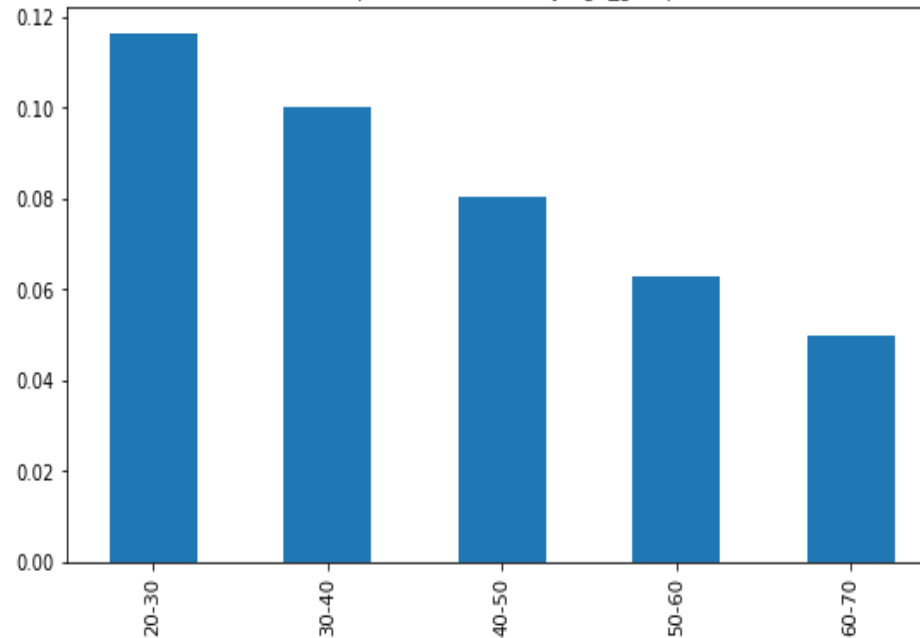
- **Most defaulters were females but highest proportion of defaulters are from the Male gender**

- 50% of defaulters are below 40 years of age.

- Most defaulters are in 30-40 age group but highest proportion of defaulters are from the 20-30 age group.

- We can see that people with higher ages are less likely to default.

DATA OF NON-DEFAULTERS BY PERCENTAGES AND PROPORTIONS OF EACH CATEGORY WITHIN COLUMNS
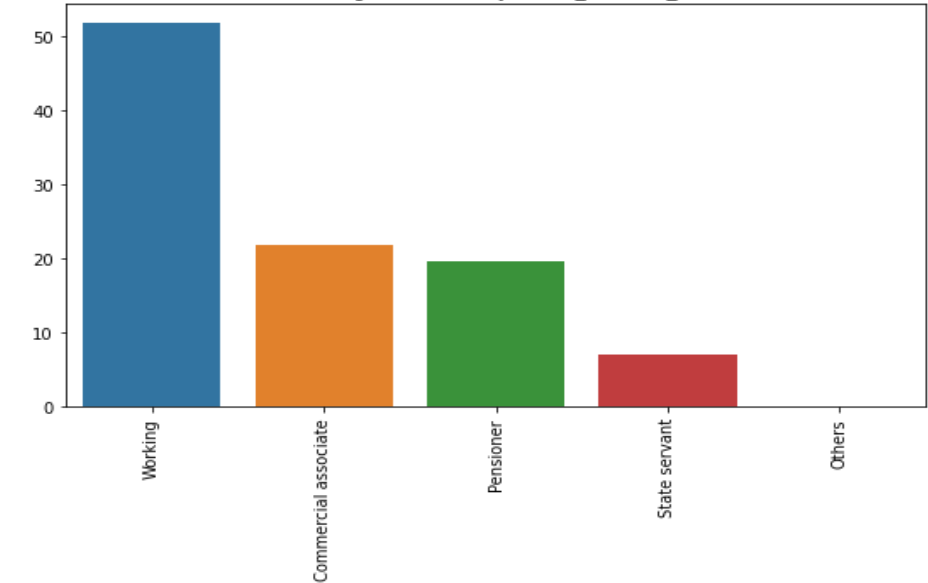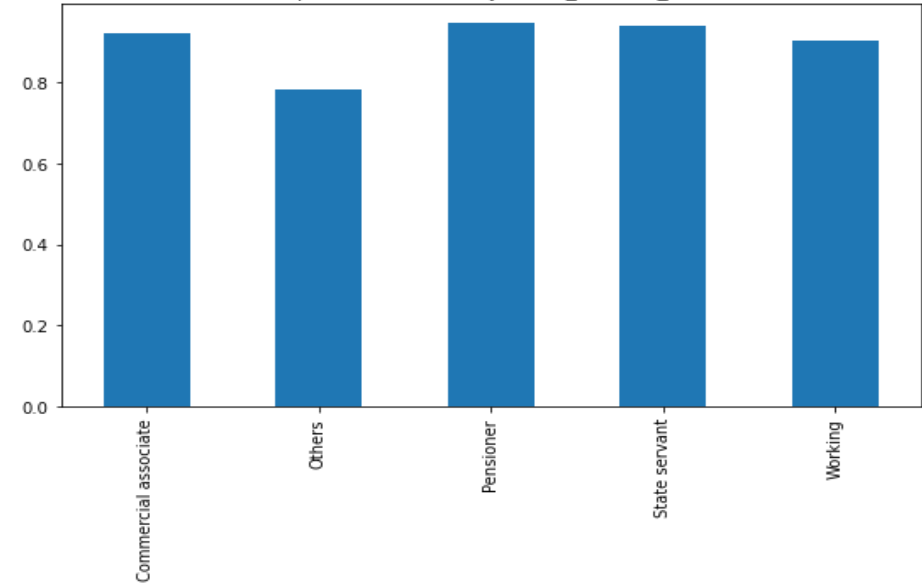
Pensioners have the highest proportions of Non-defaulters.
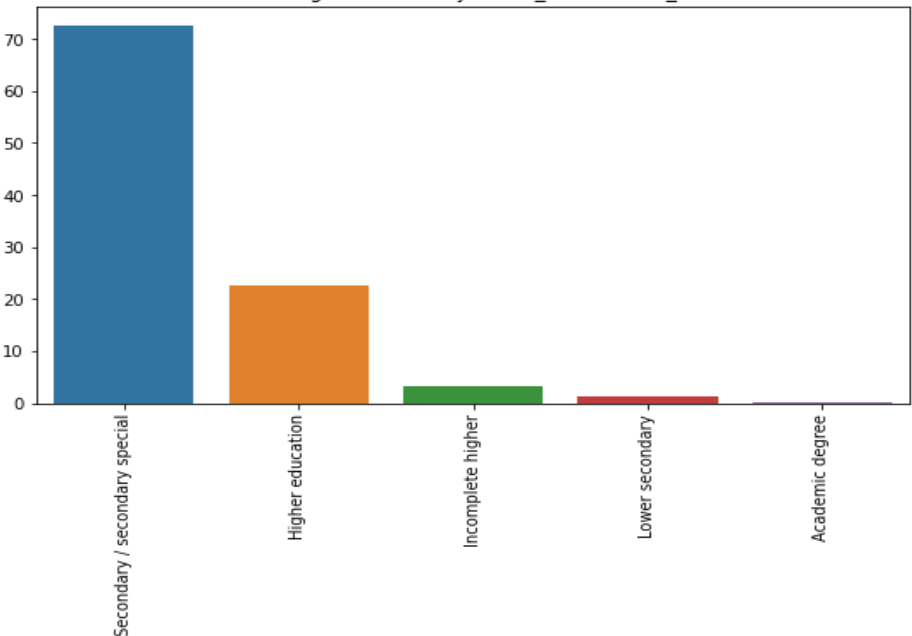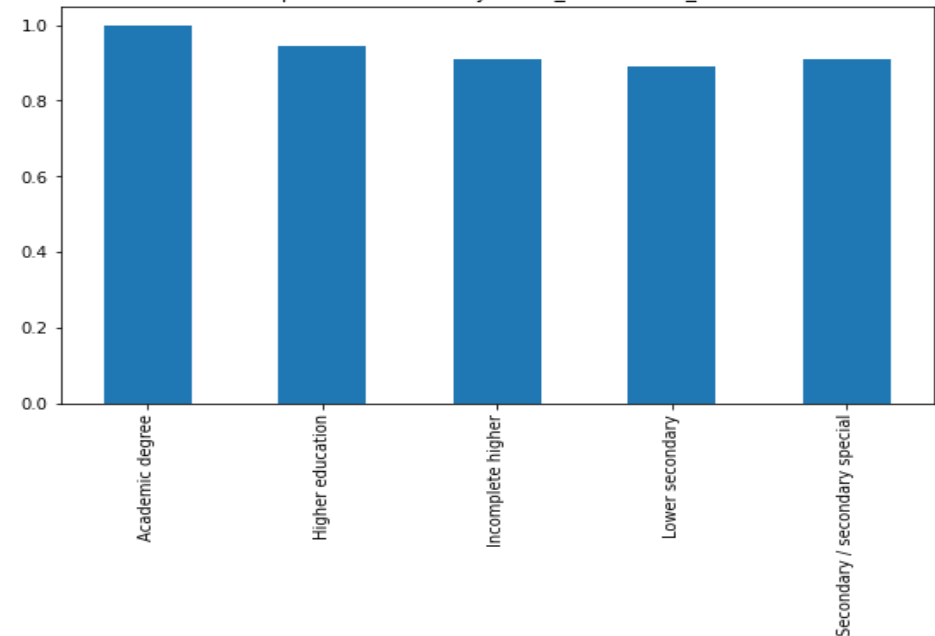
Academic degree holders have highest proportion of Non-defaulters.

BIVARIATE ANALYSIS
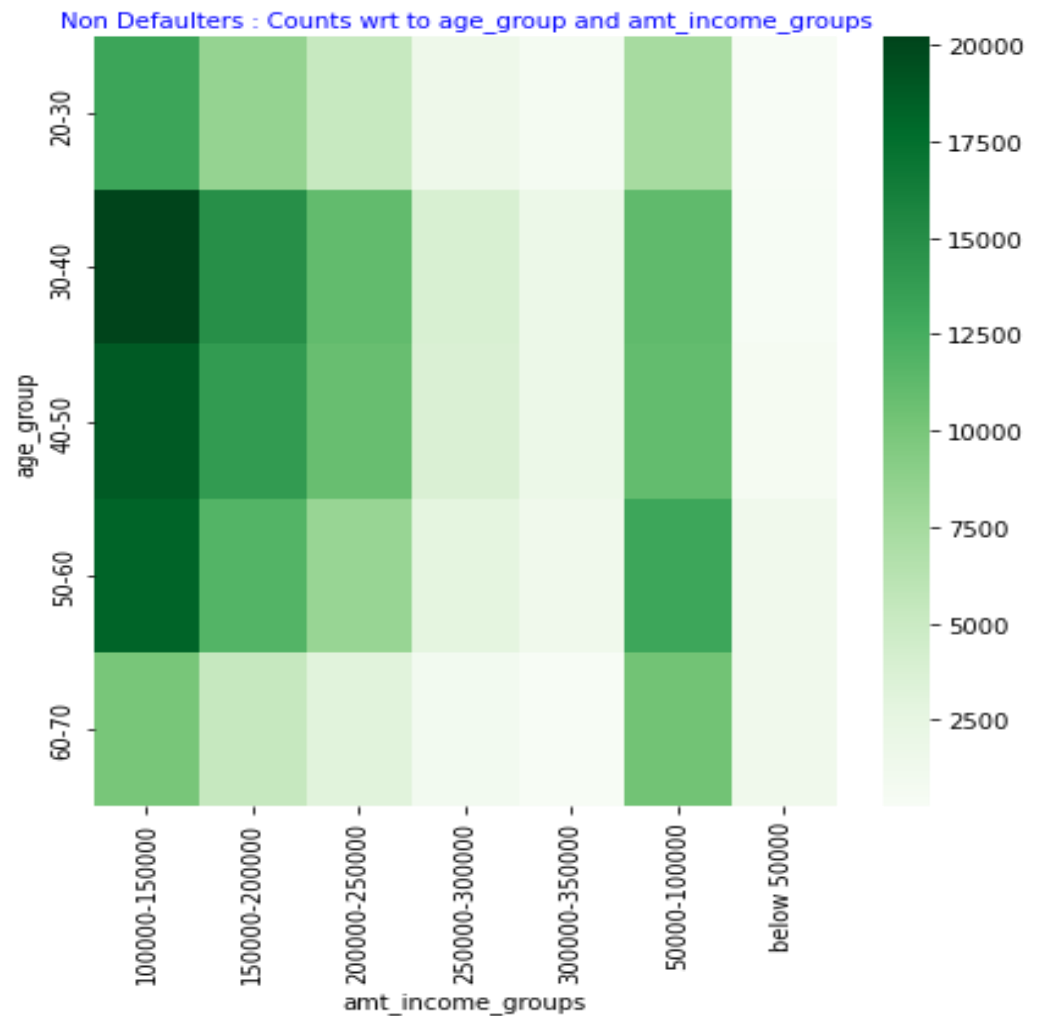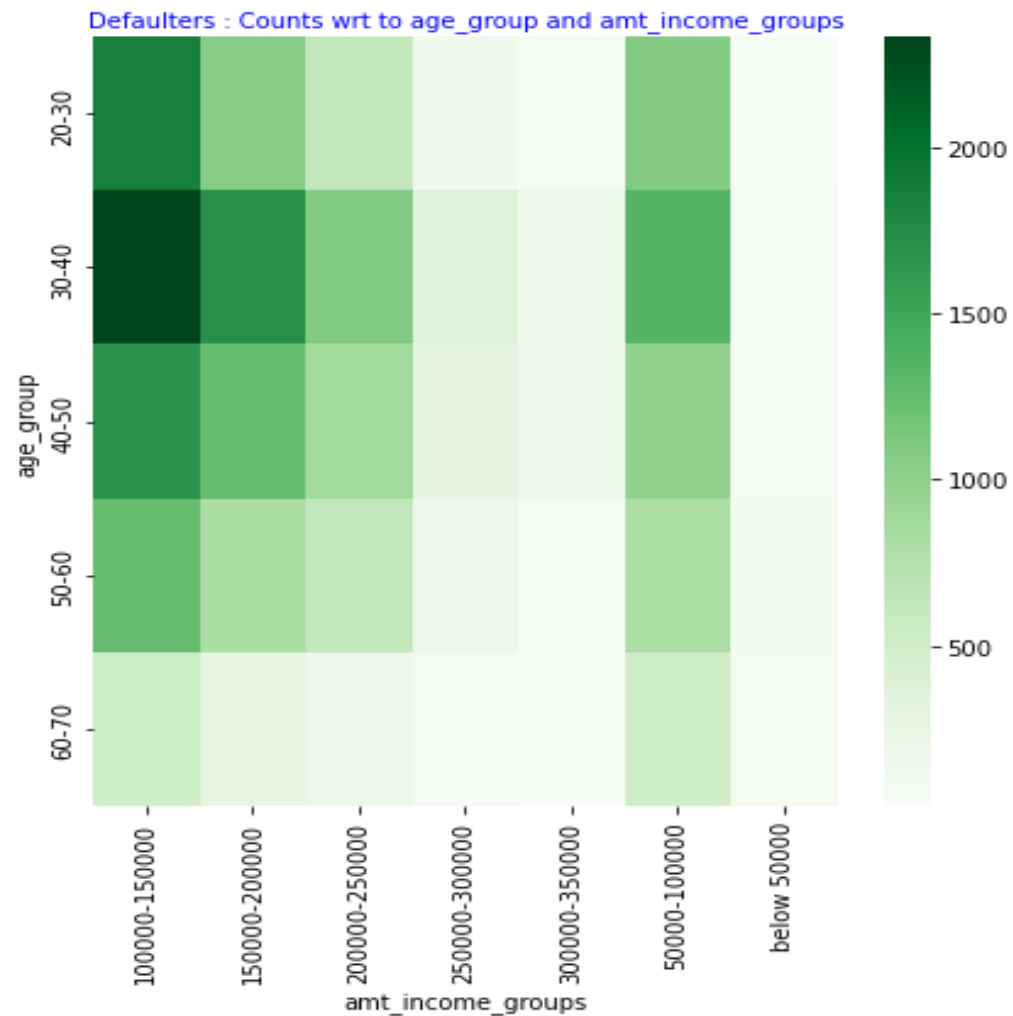
- Most defaulters with salary 1-1.5 lakh were 20-40 years of age.
- There are many Non defaulters in the age group 30-60 earning 1-1.5 lakhs

Defaulters : Counts wrt to amt_credit_groups and NAME_EDUCATION_TYPE

Non Defaulters : Counts wrt to amt_credit_groups and NAME_EDUCATION_TYPE
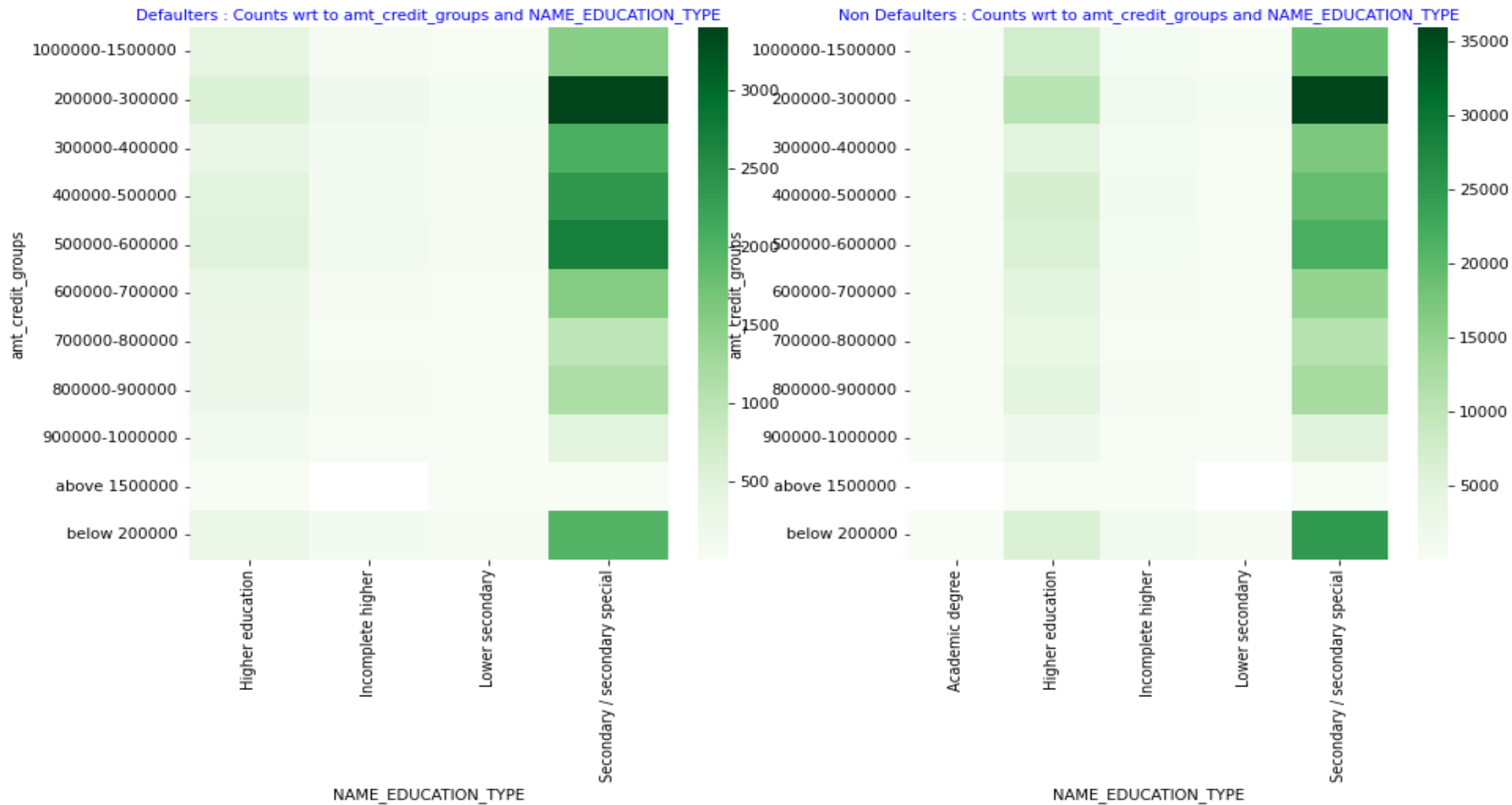
- Most defaulters had Credit between 2-6 lakhs and had secondary education

Defaulters:Distribution of AMT_INCOME_TOTAL for each NAME_EDUCATION_TYPE by CODE_GENDER
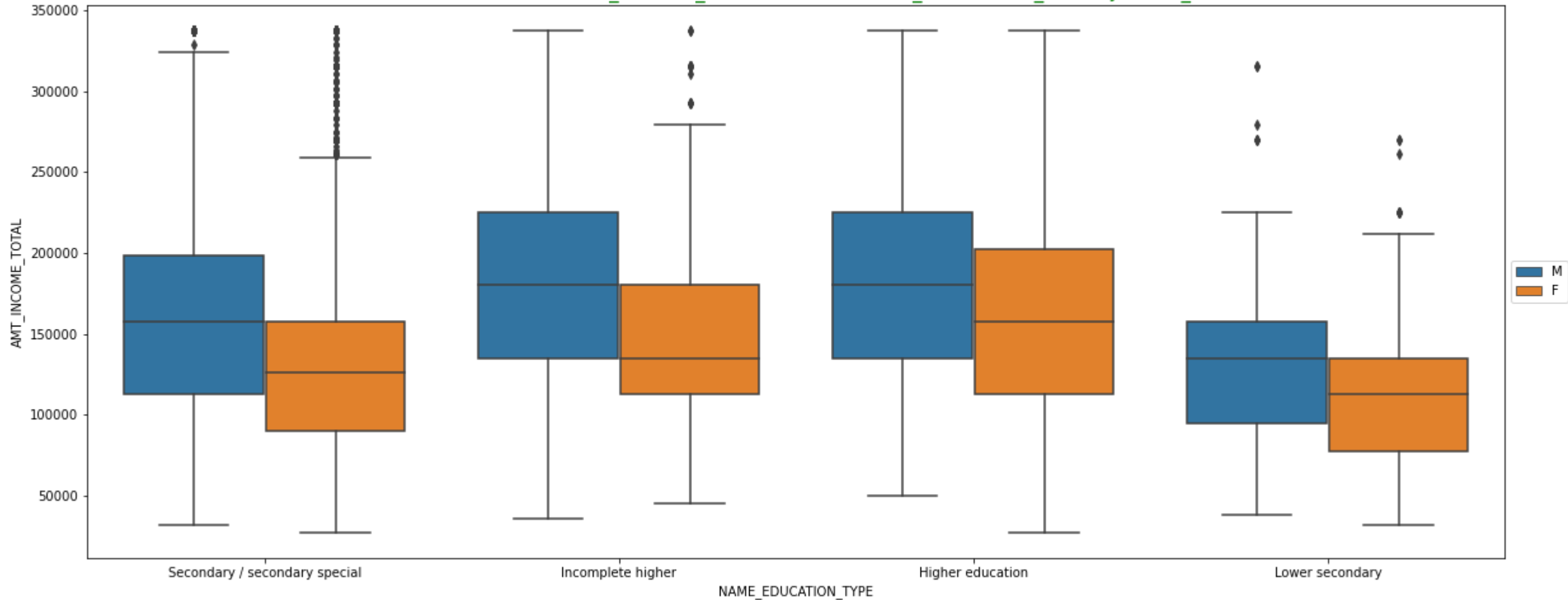
- Mostly males are having higher median salary across all education levels.

- None of those with Academic degrees were defaulters. All those with Academic degrees were non defaulters

- Median salary for higher education is higher.

- Among the defaulters with Lower secondary level education, most males and females are in lower income range.

Defaulters: Distribution of AMT_CREDIT for each amt_income_groups by NAME_EDUCATION_TYPE

Non defaulters: Distribution of AMT_CREDIT for each amt_income_groups by NAME_EDUCATION_TYPE
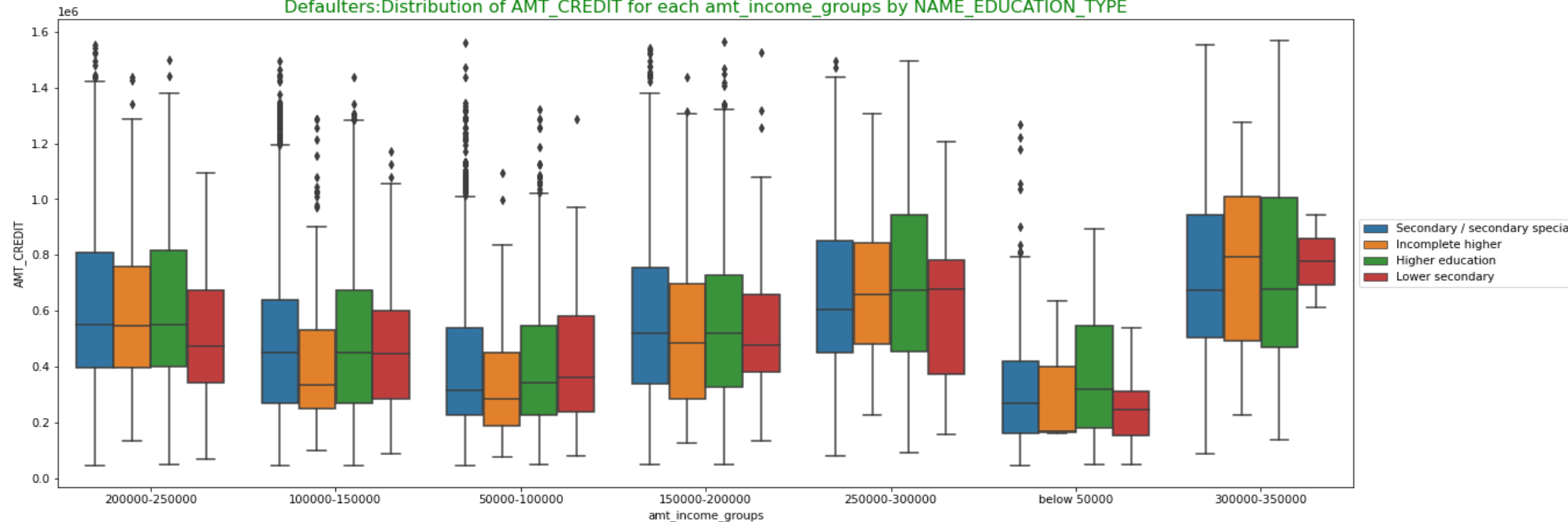
- Credit amount is generally higher for higher income groups.

- Non-Defaulters with Academic degree had the highest credits compared to other education levels.

- This means all people with academic degree were able to pay back loans even with high credits.

CORRELATION MAP

For Clients with payment difficulties
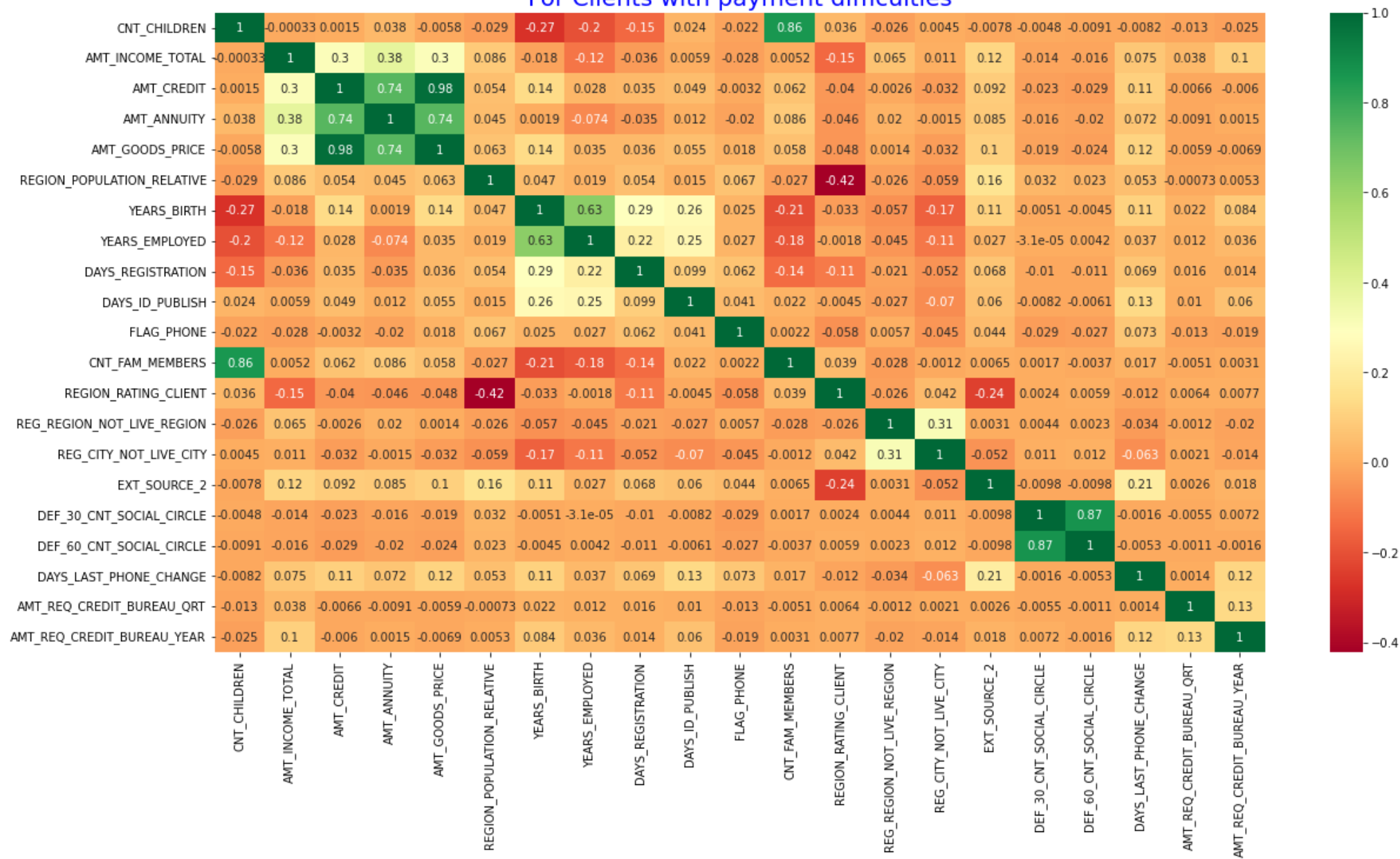
We can see that for clients with Payment difficulties:

- The region rating is inversely proportional to Income, credit, population, normalized external score. This means better regions have rating 1.

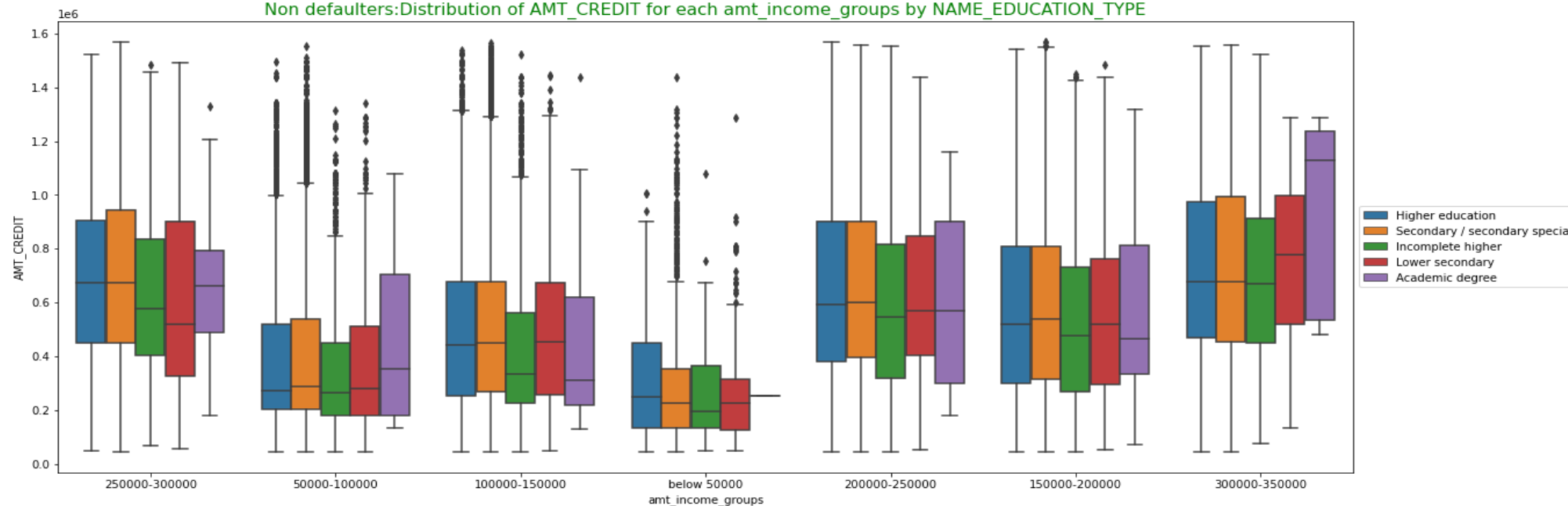- Higher the Income, higher the annuity, credit and goods price.

- Client permanent address does not match contact address for mostly those defaulters who had low age and years of employment.

- Defaulters that changed registration recently before the application had poor region ratings.

- Age and Years of employment is inversely proportional to count of children and family members. This means mostly young defaulters had more family members and children.

PREVIOUS APPLICATION DATA

Distribution (in %) of NAME_CONTRACT_TYPE_prev for each category in NAME_CONTRACT_STATUS

Distribution (in %) of NAME_CONTRACT_STATUS for each category in age_group
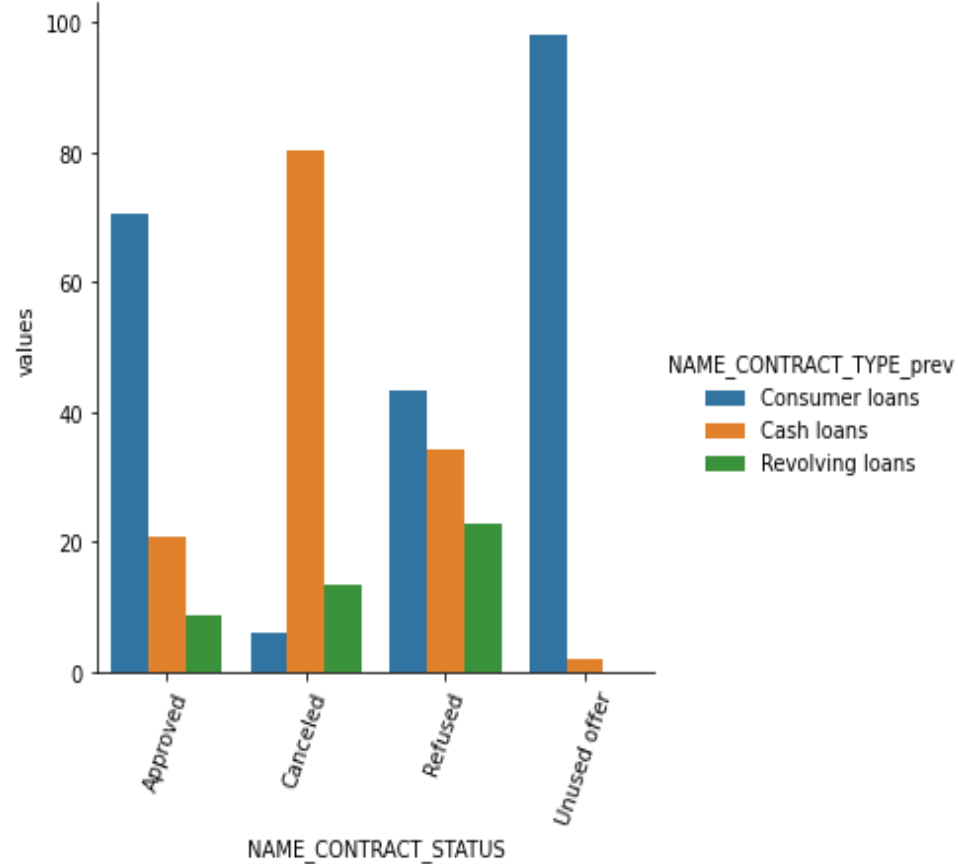
- Most of the loans that were approved were Consumer loans
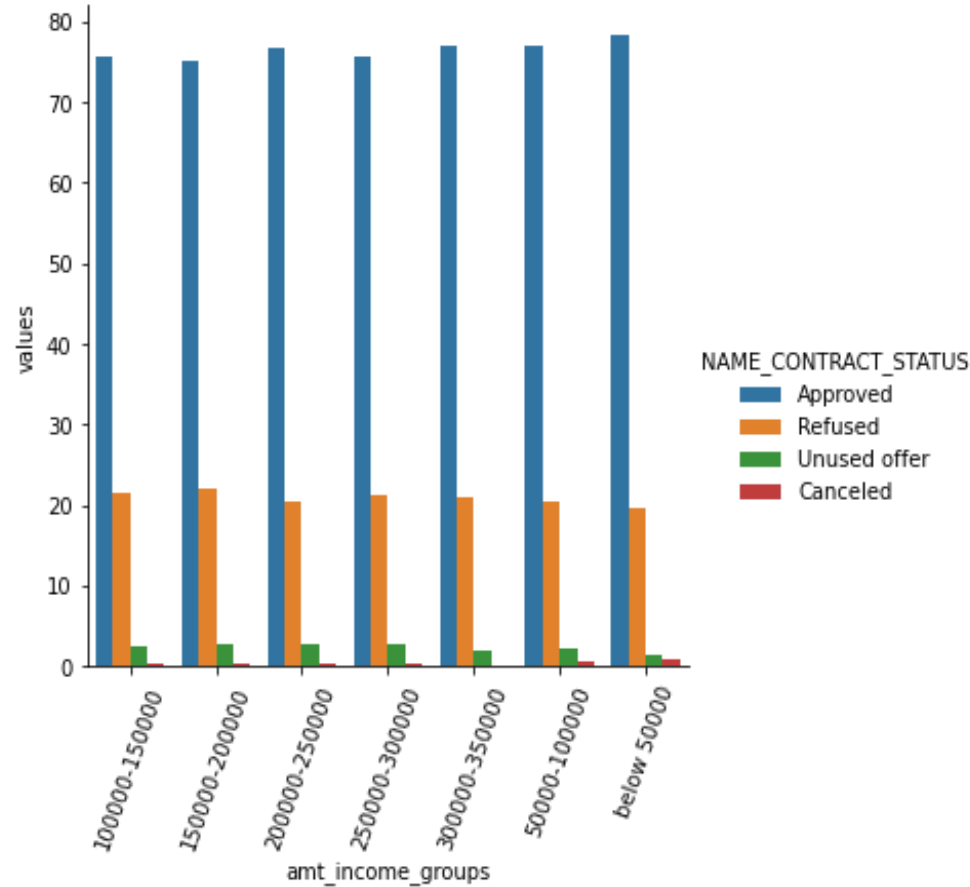- Most loans that were cancelled were cash loans followed by revolving loans
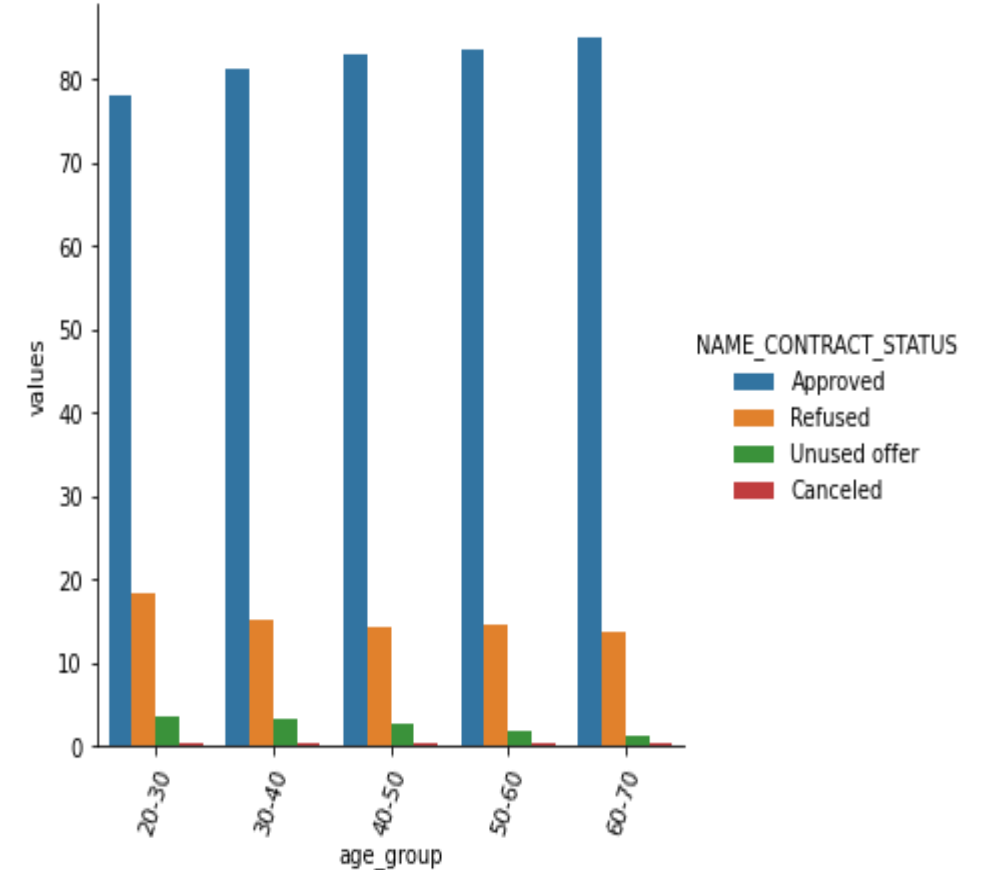
- Clients with age 20-30 have lowest no of Approved contracts and highest no. of refused contracts.
- People with age 60-70 had least unused offers

ANALYSIS OF CONTRACT STATUS FOR DEFAULTERS

Distribution (in %) of NAME_CONTRACT_STATUS for each category in amt_income_groups

Distribution (in %) of NAME_CONTRACT_STATUS for each category in amt_credit_groups
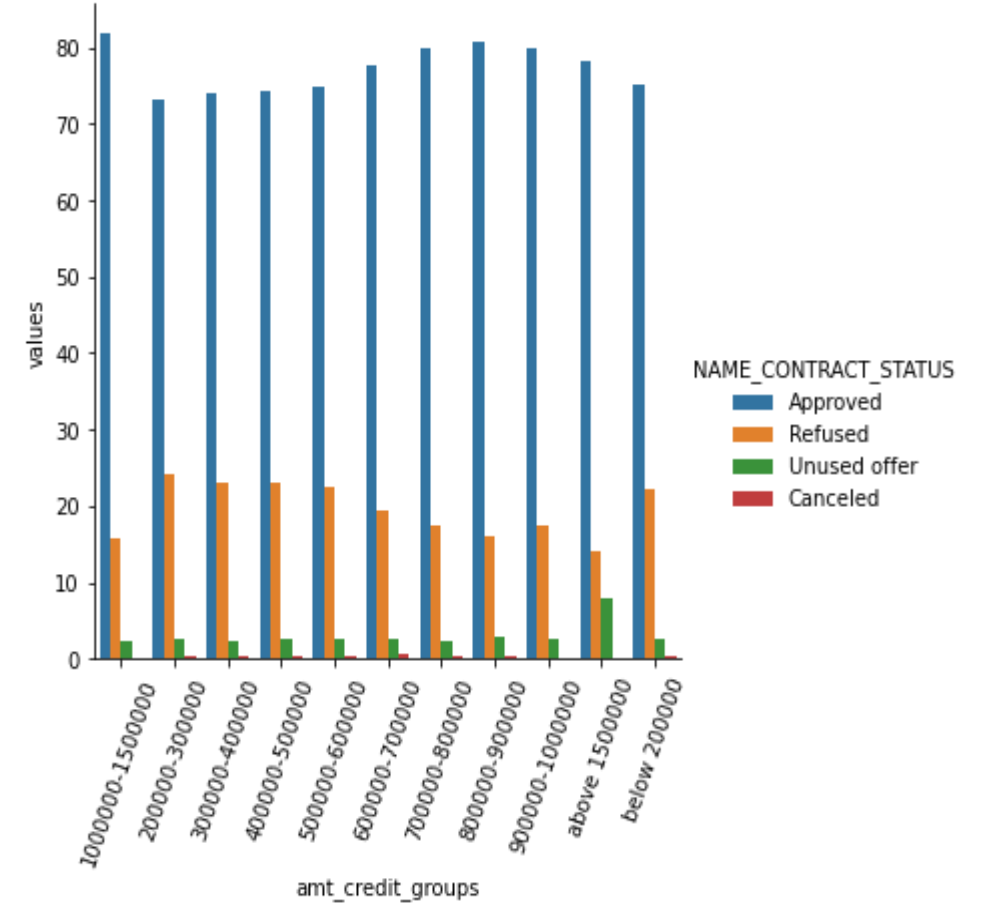

- Defaulters earning below 50k had highest cancelled offers percentage compared to other income groups

- Defaulters with credits between 1-1.5 lakhs had highest % of approved offers.
- Defaulters with credit > 15 lakhs had highest % of unused offers. This means extremely high credits were mostly unused.

Defaulters:Distribution of AMT_INCOME_TOTAL for each age_group by NAME_CONTRACT_STATUS
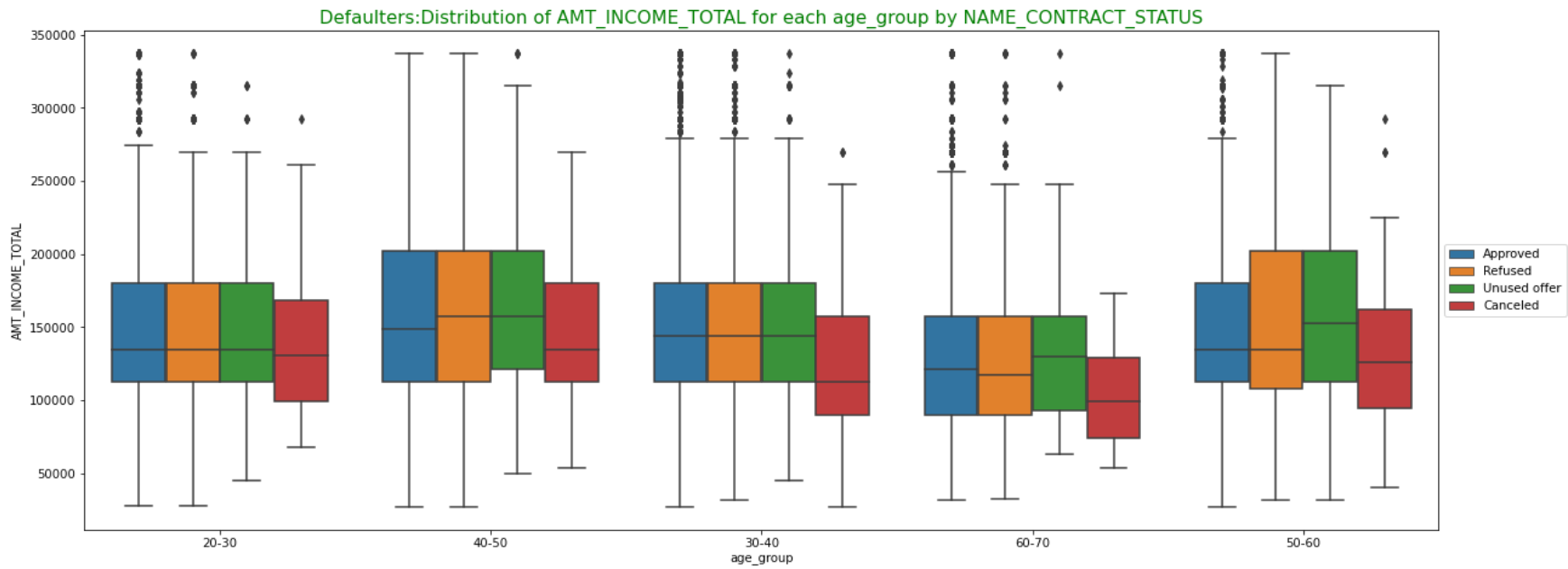
- The loans were mostly cancelled for clients with lower median income.

- For defaulters in age 50-70, the median income was higher of those who did not use the offer.

Defaulters:Distribution of AMT_CREDIT_prev for each amt_income_groups by NAME_CONTRACT_STATUS
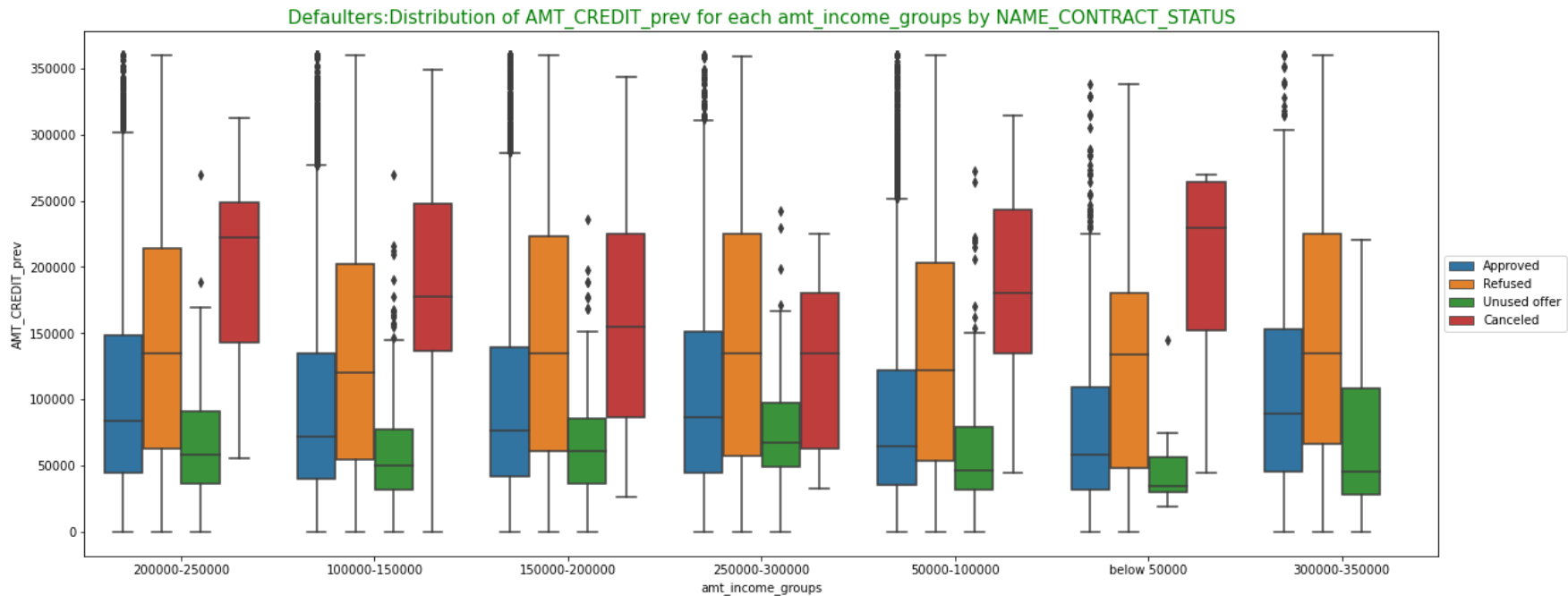
- No loans were cancelled previously for defaulters that earned 3-3.5 lakhs.

- The median credit amount of cancelled loans is higher than median credit of the other loans.

Defaulters:Distribution of AMT_APPLICATION for each NAME_EDUCATION_TYPE by NAME_CONTRACT_STATUS

Non defaulters:Distribution of AMT_APPLICATION for each NAME_EDUCATION_TYPE by NAME_CONTRACT_STATUS
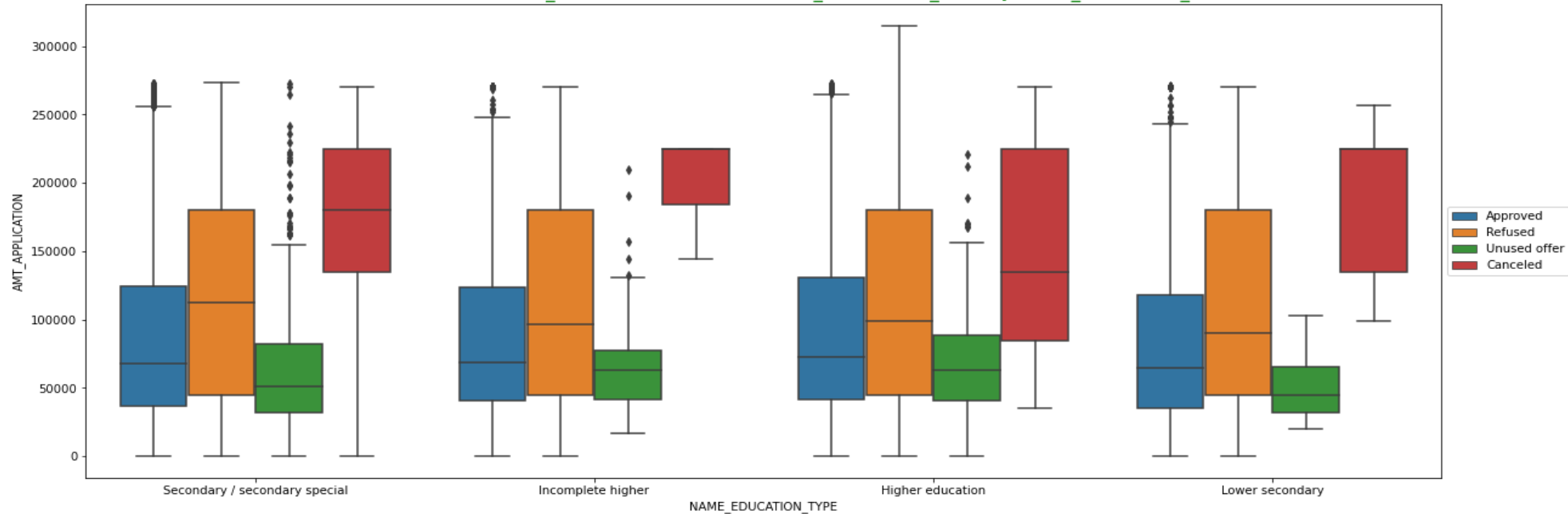
- The median application amount was highest for those defaulters with Lower secondary level education and Incomplete higher education and these loans were cancelled.
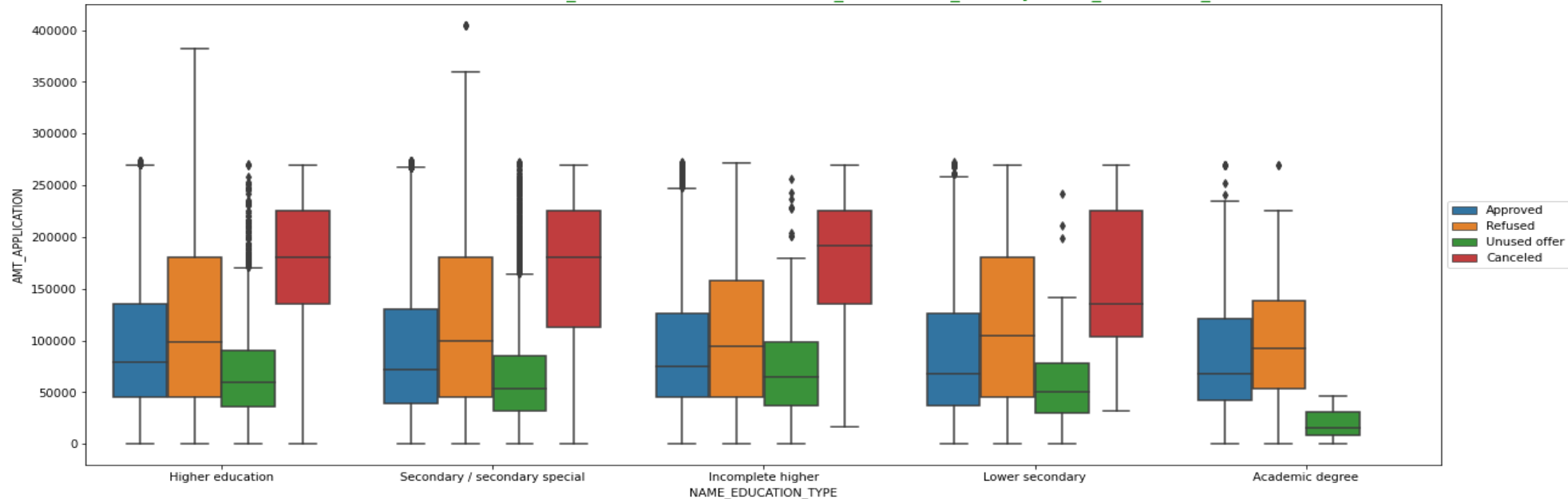
- Approved loan amounts were mostly between 40k to 130k.

- No loans were cancelled for clients with academic degree and none of them were defaulters

CONCLUSIONS

- Males are more likely to default

- Clients with lower income and lower secondary education more likely to default.

- Clients with Academic degree least likely to default.

- Pensioners are also least likely to default.

- Clients with high credit amounts above 15 lakhs mostly don't use the offer.