



Credit EDA Case Study

- Project report - upGrad & IITB

- Vishal Garg



Scope of Project

- This assignment aims to extract relevant inputs from banking data.
- It will help banks make informed decisions and reduce financial risks.
- The analysis will focus on two sets of data –
- ***“Applications data” with current loan applicant information and default status.***
- ***“Previous data” with details of past loan applicants.***

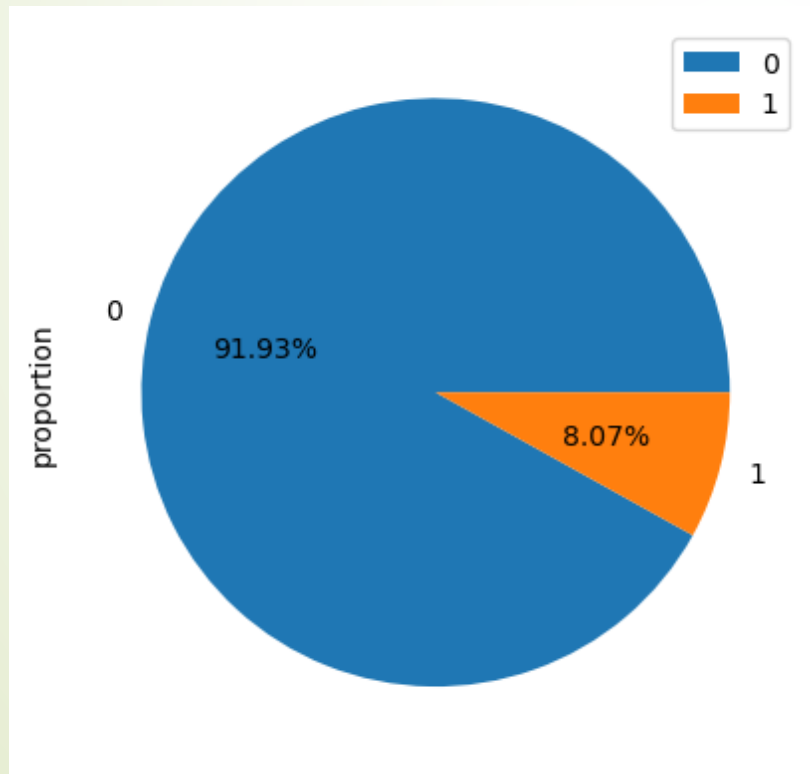


Steps Involved in Study

- **Data Sourcing & analysis of data structure.**
- **Data Cleaning**
 - *Dealing with “miss” & “null” values.*
 - *Examining of anomalies/outliers.*
 - *Binning important values.*
- **Data Analysis via Univariate, Bivariate & Multivariate analysis.**
- **Cocusion.**

Analysis of data via Univariate, Bivariate & Multivariate methodology.

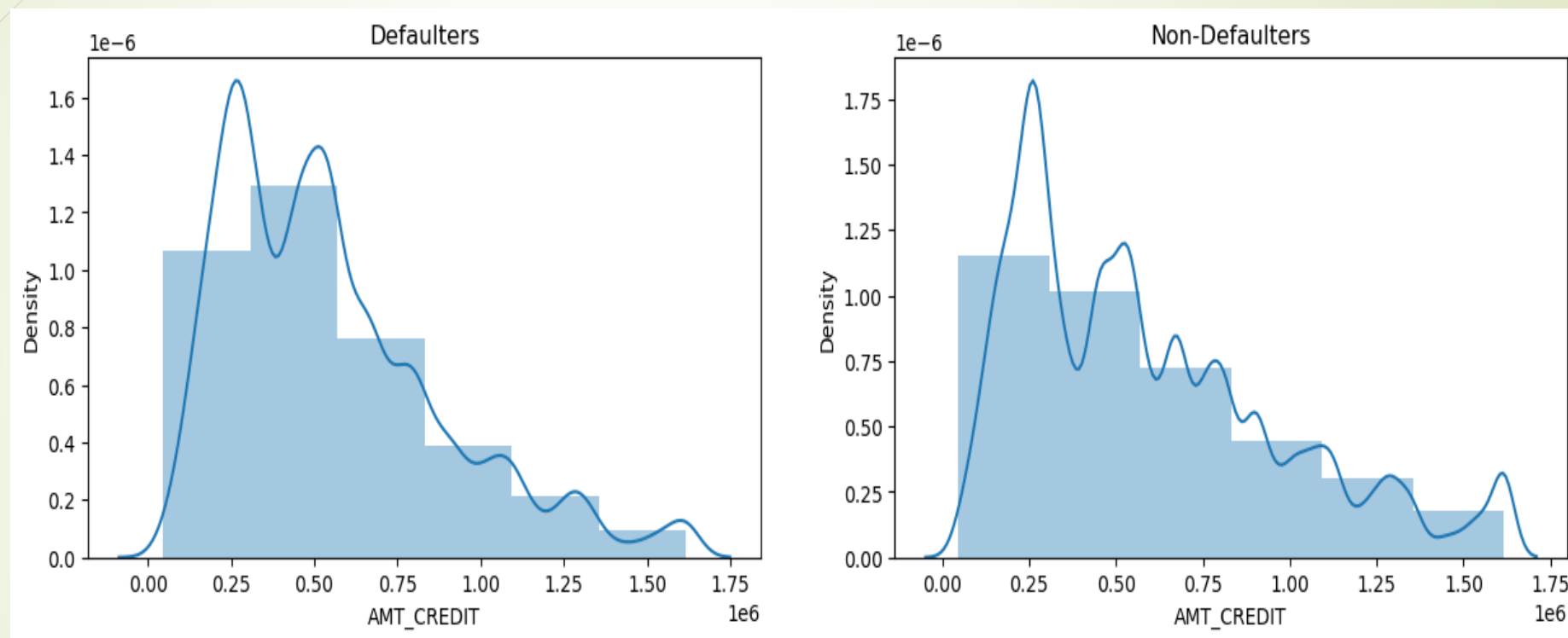
- First, we will bifurcate the data into **Defaulters and Non-defaulters**, post that conduct Univariate, Bivariate, and Multi-variate analysis simultaneously.



- Non-defaulters represented by blue section concluding at 91.93%.
- Defaulters represented by the Yellow section concluding at 8.07%.

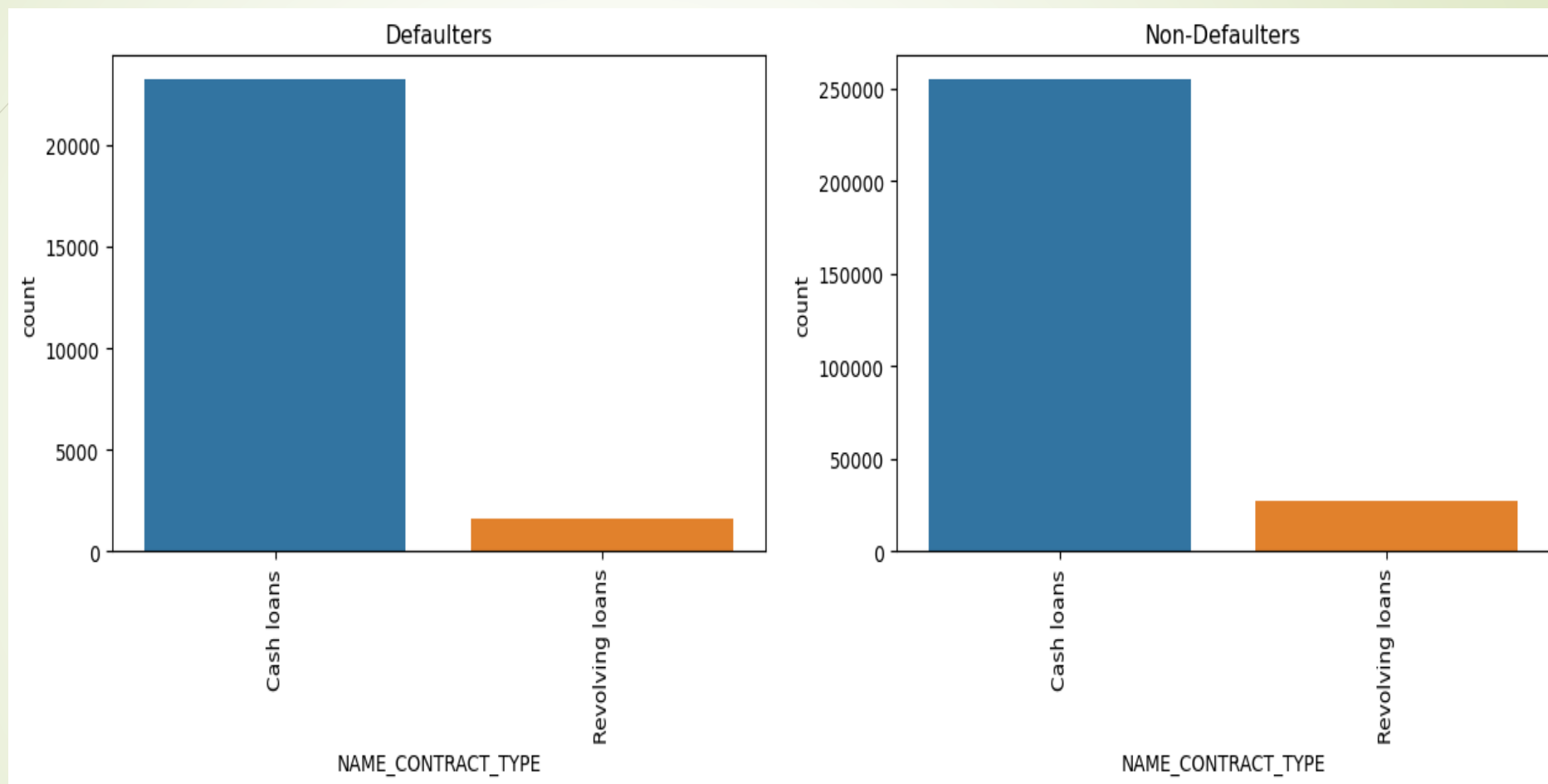
Univariate Analysis

- Amount Credit Segment for both Defaulters & Non-defaulters



- In the non-defaulters category, a majority of individuals borrow around 3 lakhs, with fewer people taking larger loans.
- Defaulters, on the other hand, tend to have taken up to 5 lakh loans.

➤ Name Contact type segment for both Defaulters & Non-defaulters



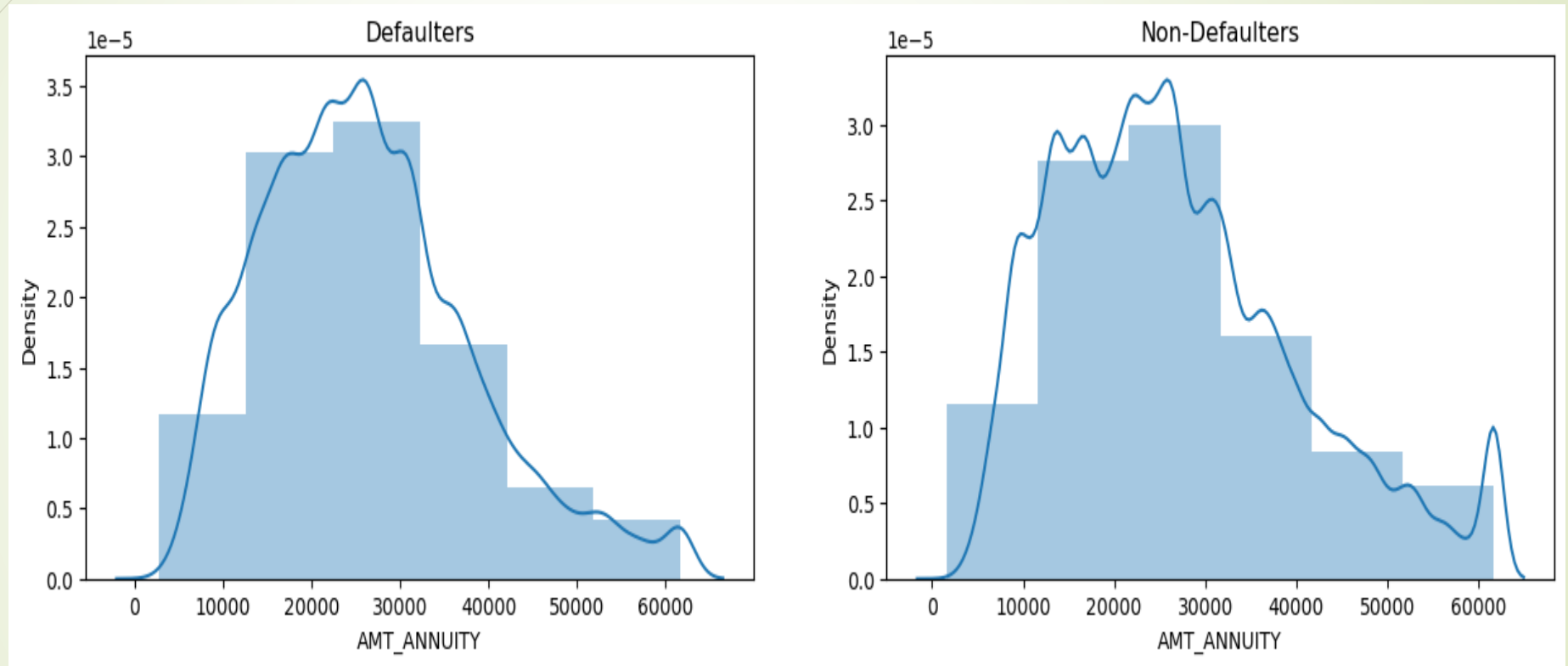
- The majority of defaulters and non-defaulters have obtained cash loans, while only a small number have taken revolving loans.

➤ Code Gender segment for both Defaulters & Non-defaulters



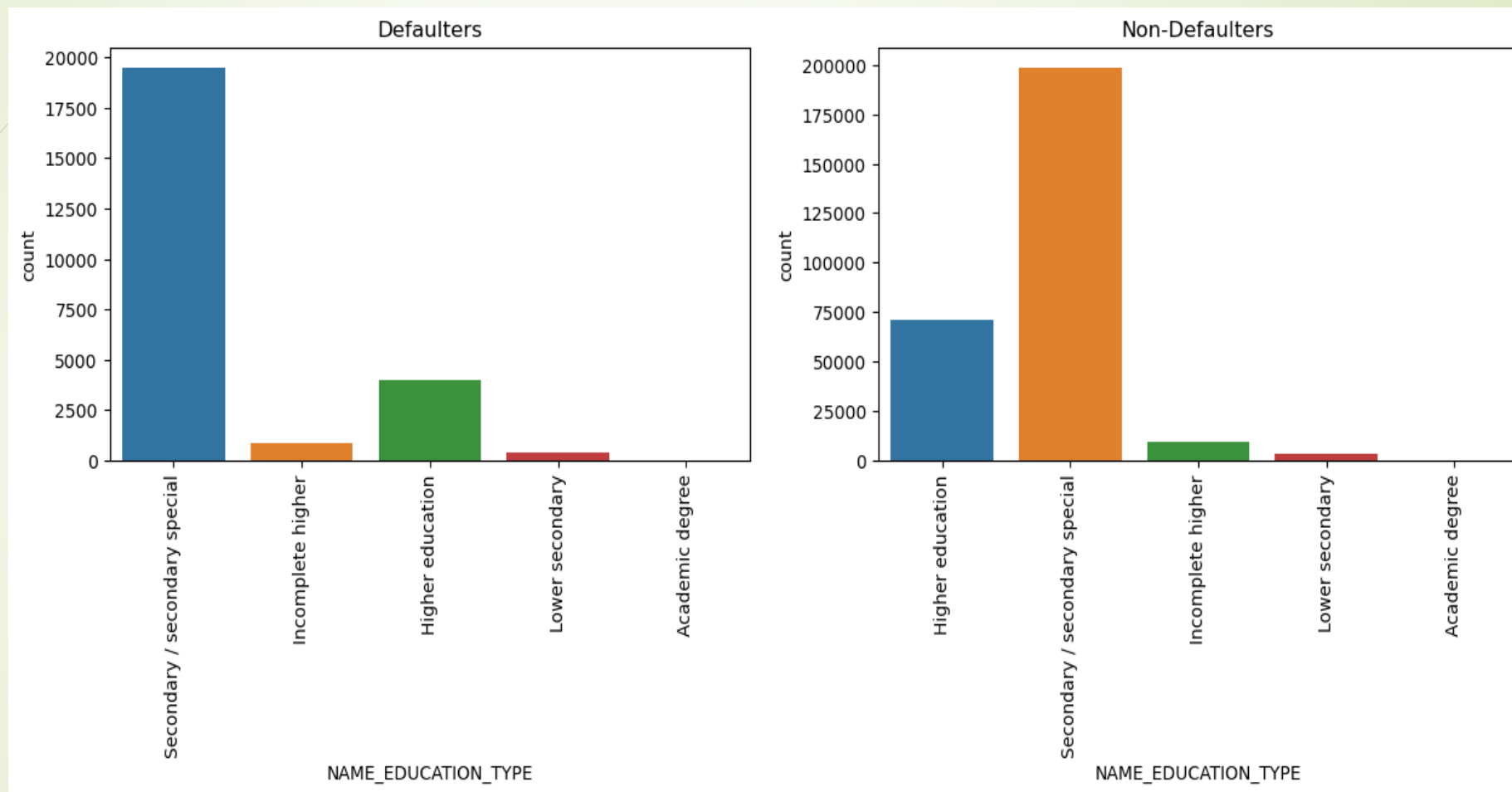
- The number of females is highest among both those who have defaulted and those who have not defaulted.

➤ Amount Annuity segment for both Defaulters & Non-defaulters



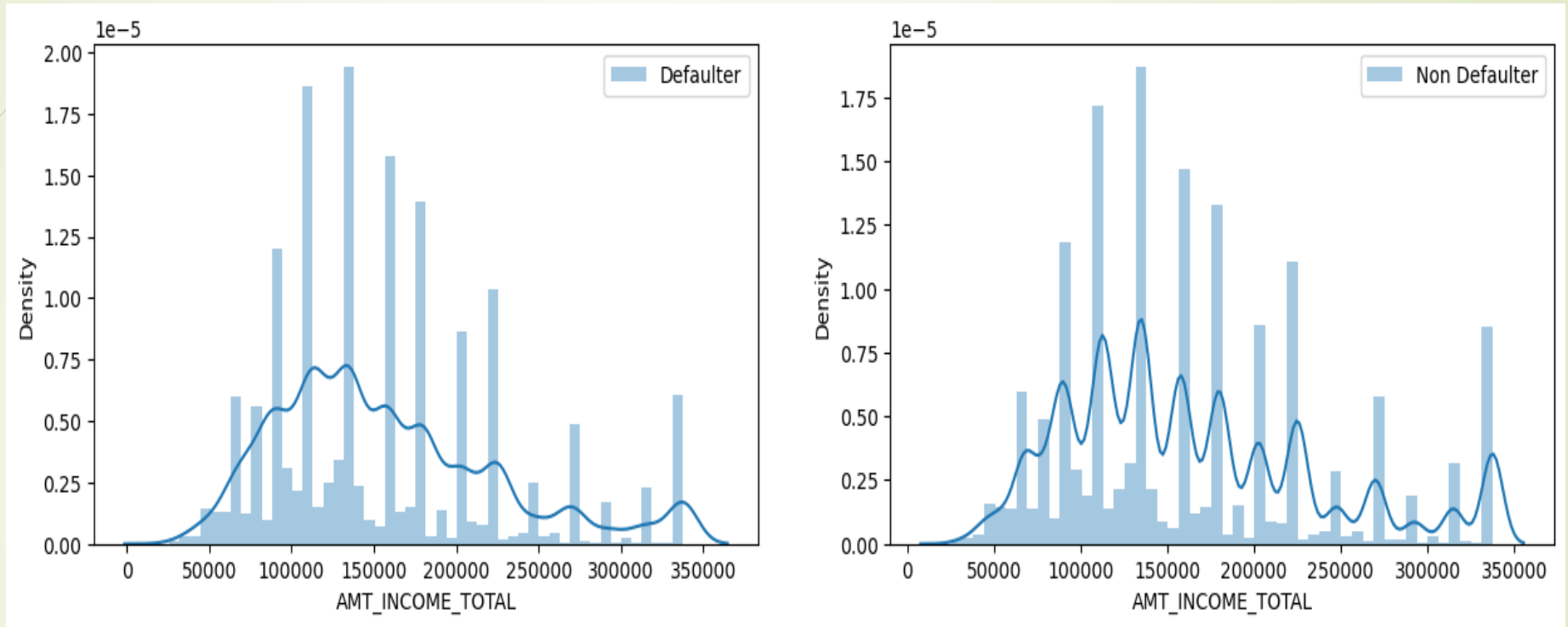
➤ A nearly identical pattern can be observed in both situations

➤ Name Education Type segment for both Defaulters & Non-defaulters



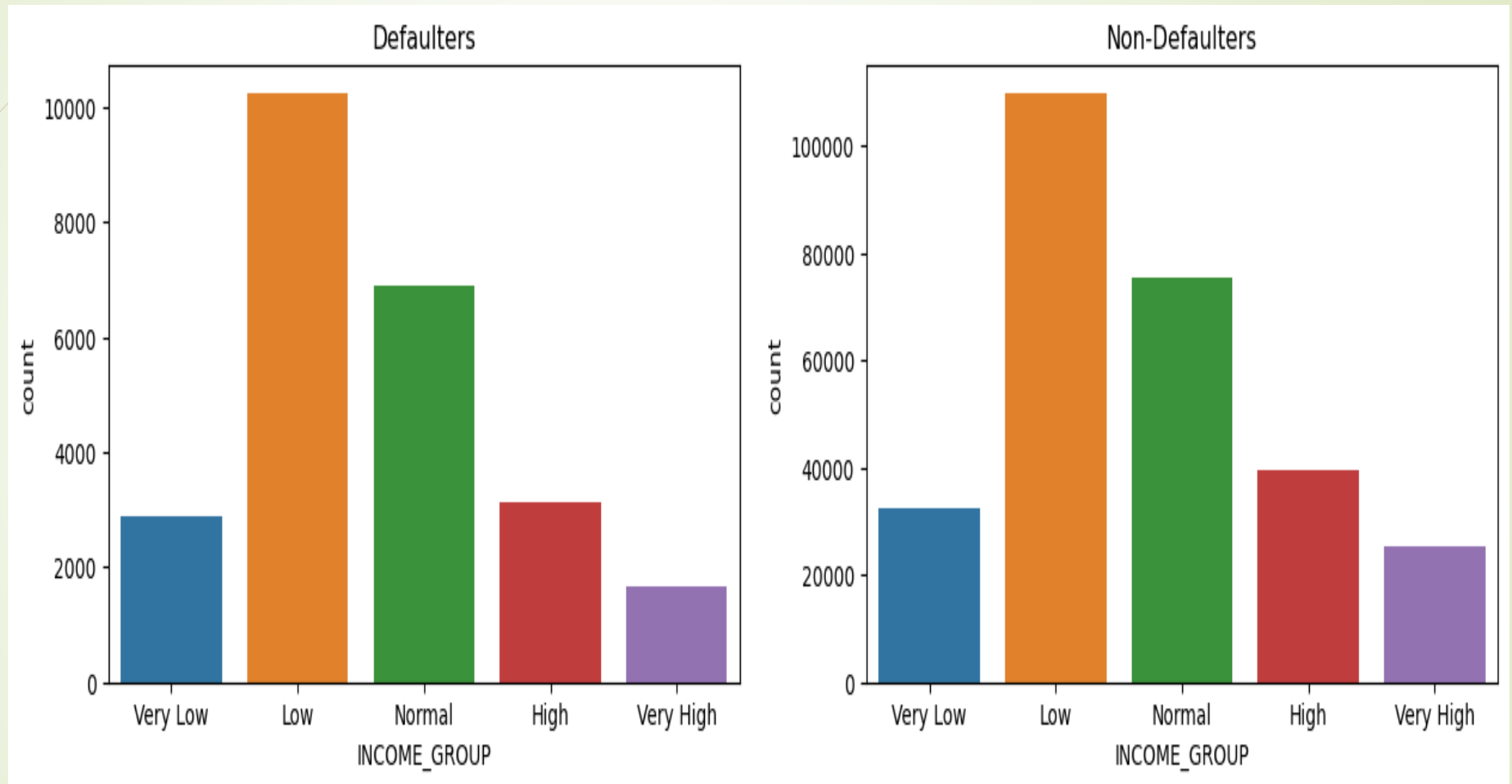
- Individuals with secondary/secondary special education make up the largest portion of both defaulters and non-defaulters, with those holding higher education coming next in line

➤ Amount Income Total segment for both Defaulters & Non-defaulters



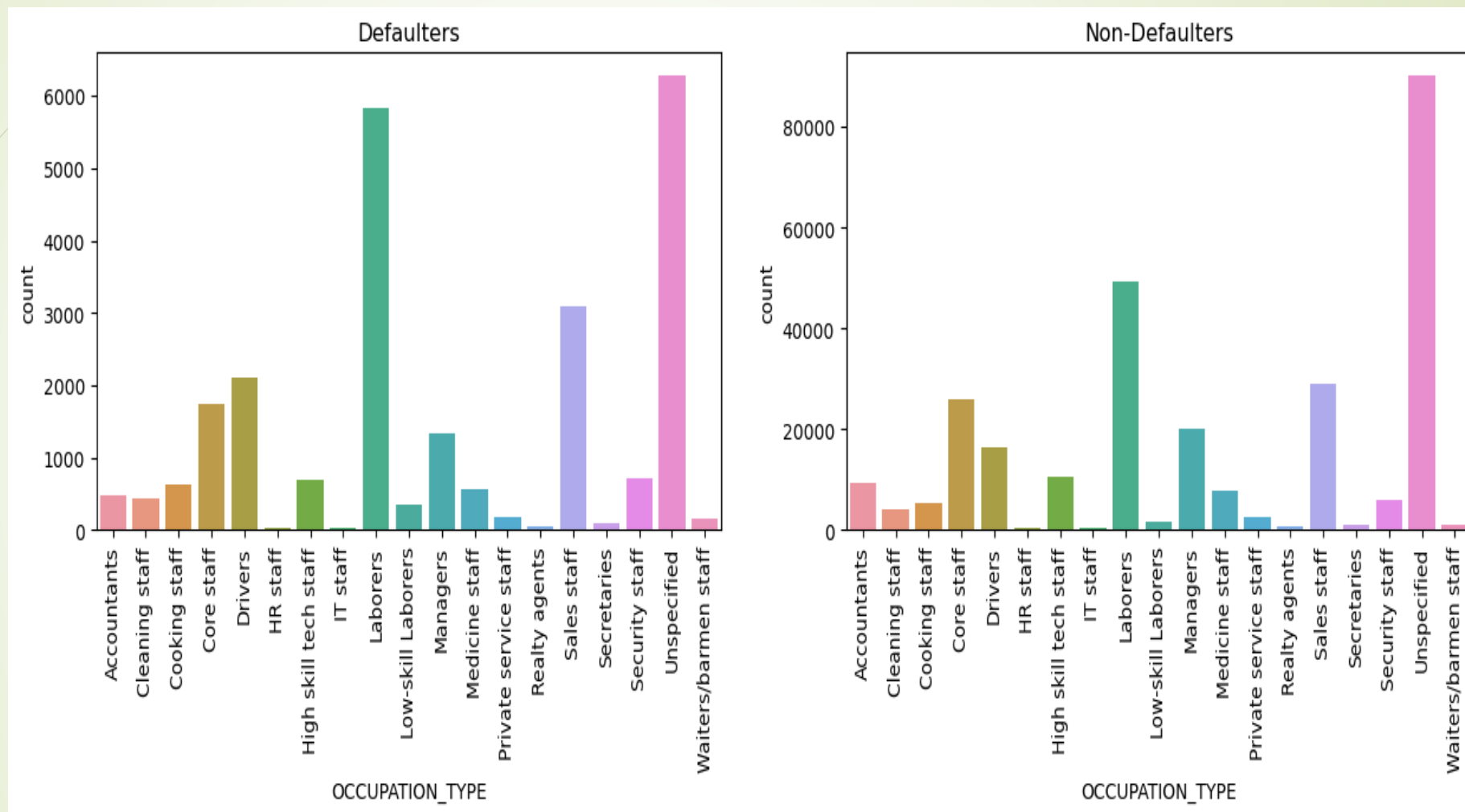
- A regular pattern of high and low behaviour is observed in individuals who do not default.
- Defaulters with incomes between 1L-1.5L outnumber those with incomes exceeding 2.5L, indicating that lower-income individuals are more likely to default compared to higher-income individuals

➤ Income group segment for both Defaulters & Non-defaulters



➤ The above graphs indicate that individuals with lower incomes have the highest numbers of defaulters and non-defaulters

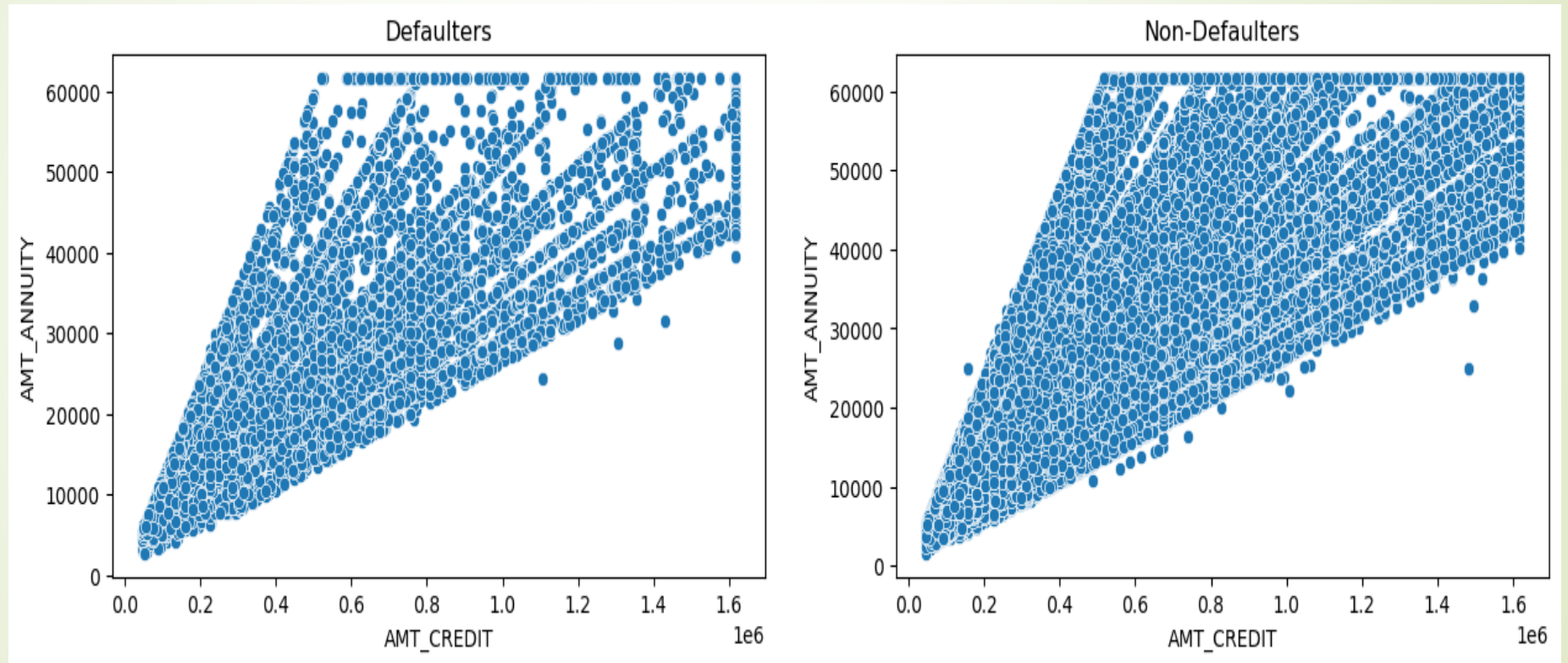
Income group segment for both Defaulters & Non-defaulters



➤ Laborers appear to be at the highest level in both defaulters and non-defaulters

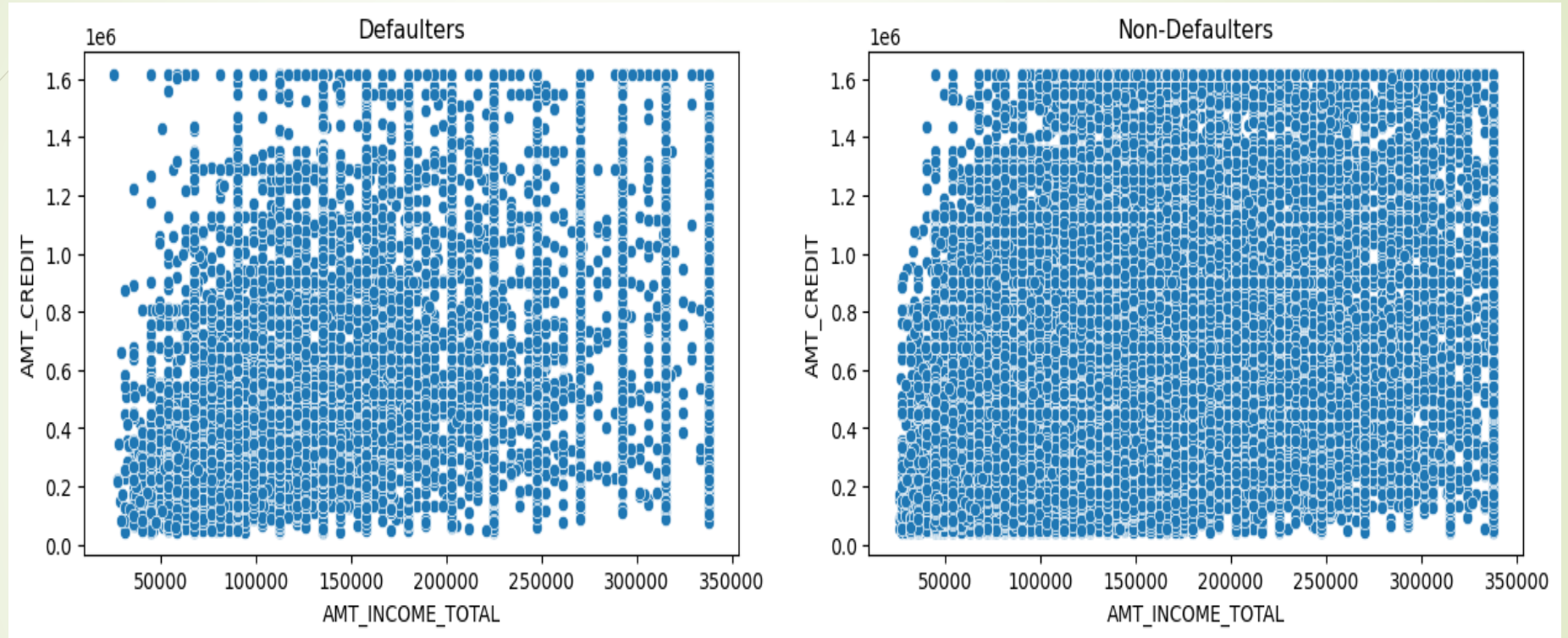
Bivariate Analysis

- Correlation between Amount credited vs. Amount Annuity.



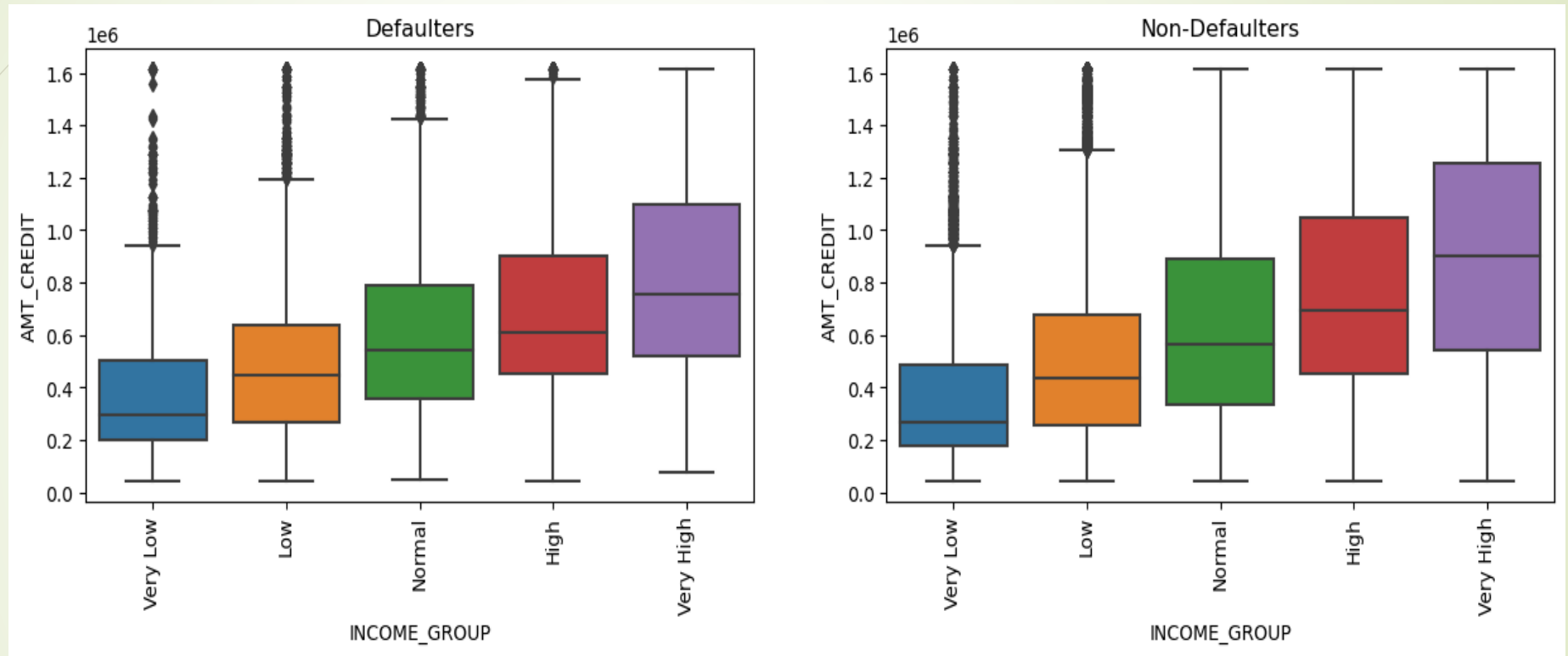
- A standard relationship appears to exist between the annuity amount and credit amount.

➤ **Correlation between Amount credited vs. Amount Income Total.**



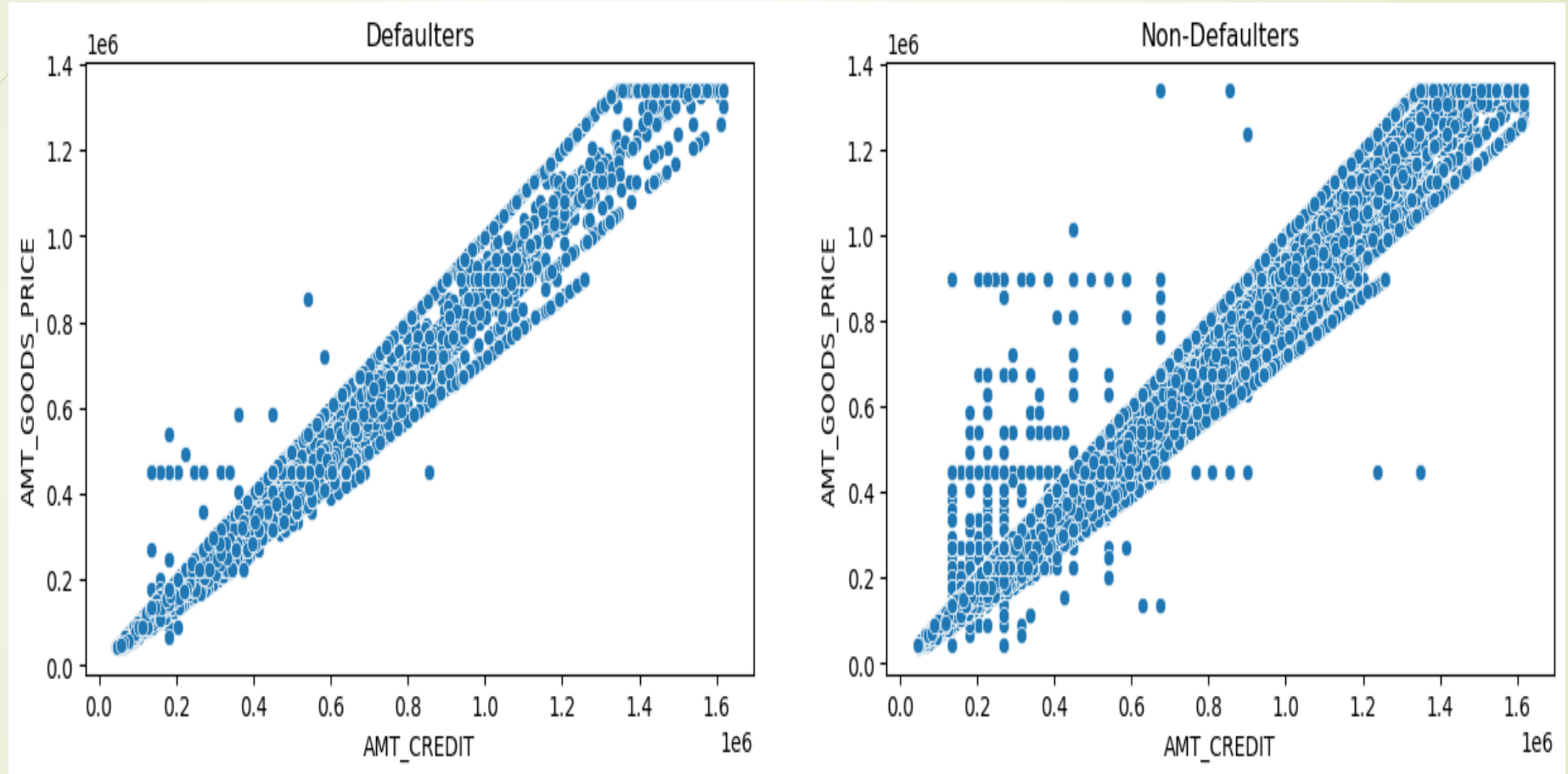
- The relationship between client income and loan amount does not show a strong correlation for defaulters and non-defaulters. While higher-income individuals are expected to take larger loans and repay them easily, those with lower incomes also take large loans for luxurious needs. This can be observed in the defaulter's graph where points are not closely packed as loan and credit amounts increase

Correlation between Amount credited vs. Income Group.



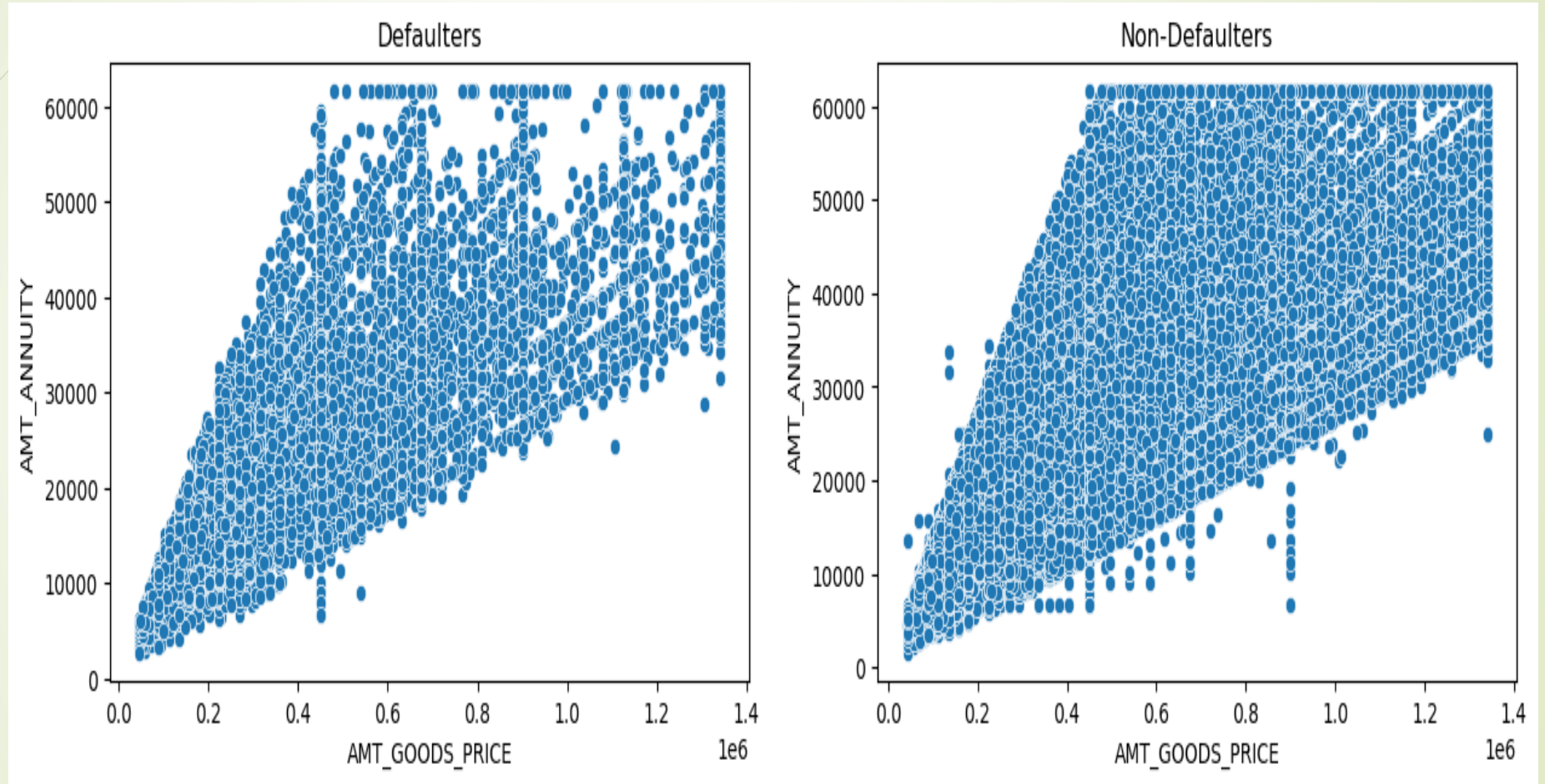
- There is a typical pattern where, as income rises, the median loan amount also increases
- Both graphs show a similar pattern. Rich defaulters typically borrow a median loan amount of 7L, while non-defaulters in the same income bracket borrow around 10L. This indicates that the bank trusts non-defaulters more and is more willing to extend credit to them

➤ Correlation between Amount credited vs. Amount Goods Price.



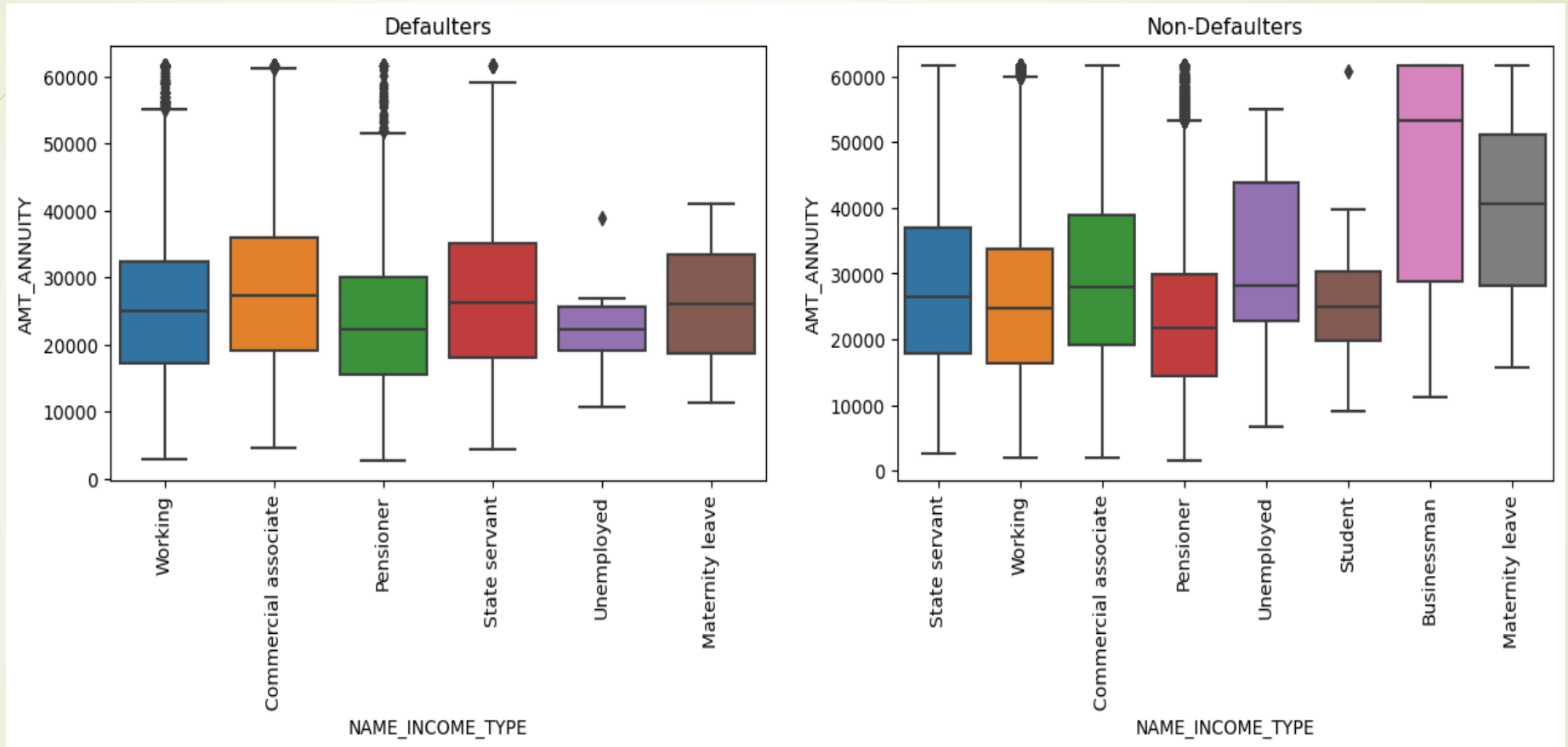
- The data clearly shows a strong correlation between credit amount and price of goods for defaulters and non-defaulters, as evidenced by a correlation coefficient of 0.98. This suggests that higher prices of goods result in larger loan amounts.

➤ **Correlation between Amount Goods Price Vs Amount Annuity.**



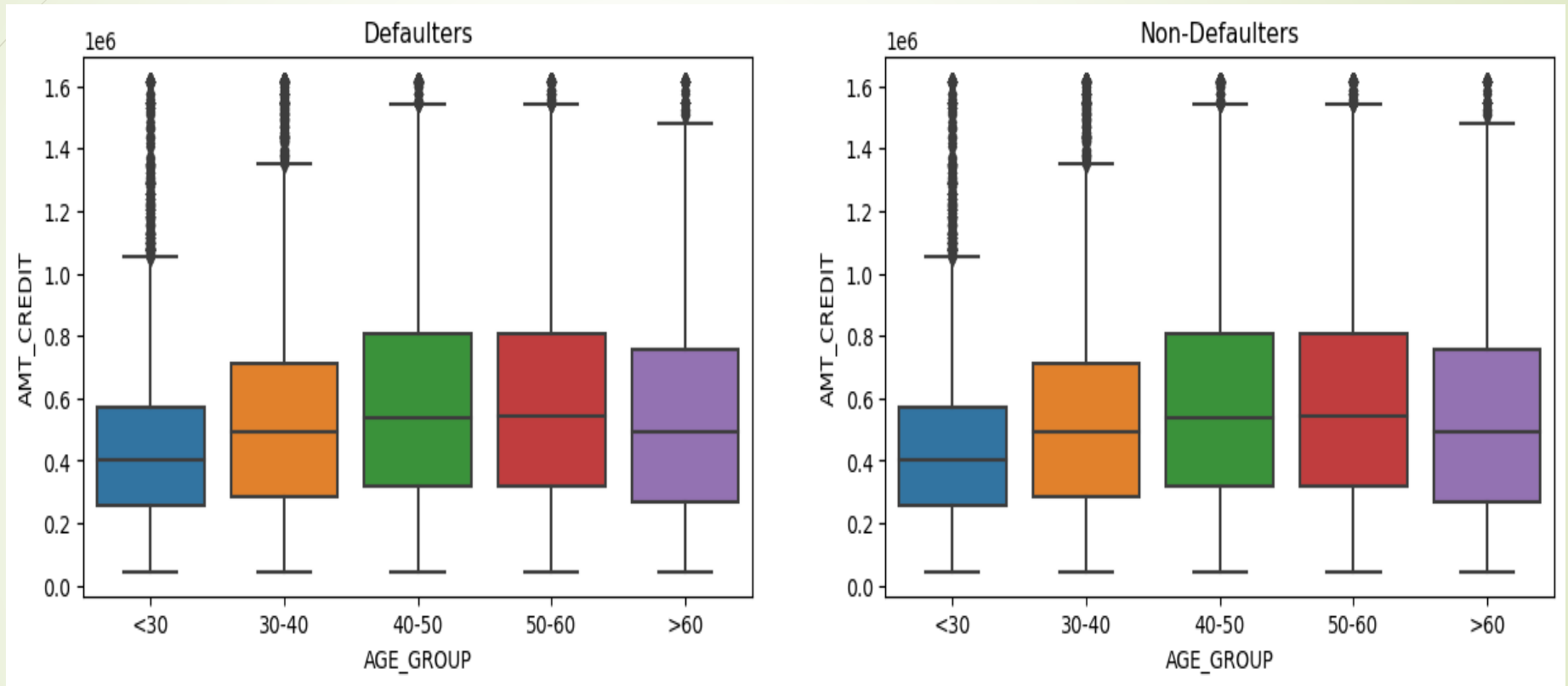
➤ There appears to be a typical connection between the cost of items and the level of annuity, although the relationship between them is not certain

Correlation between Amount Annuity vs. Name Income Type.



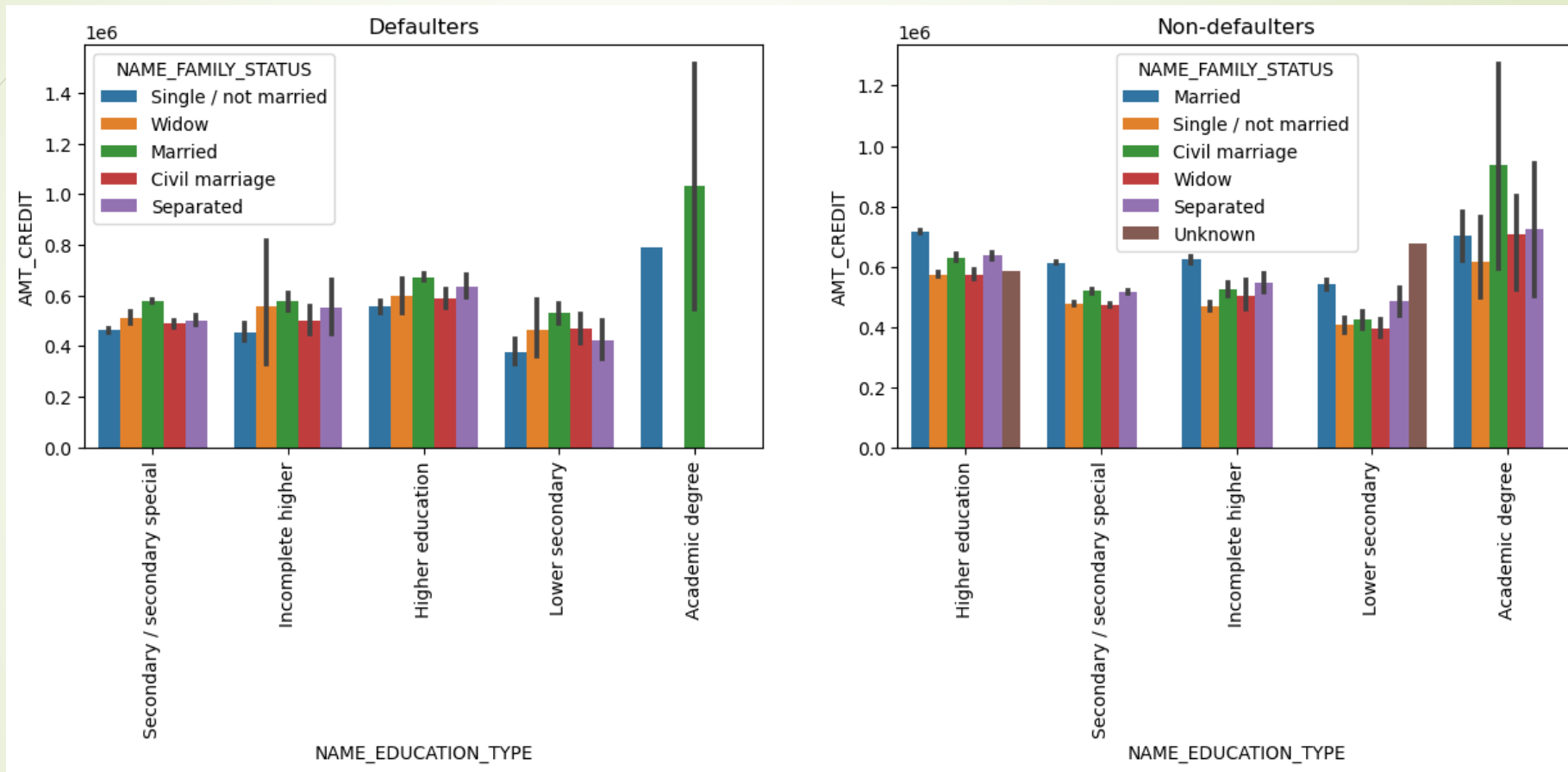
- Among those who did not default on their payments, business owners receive the highest median annuity sum, with individuals on maternity leave coming in second
- In defaulters, everyone has a median annuity between 20,000 - 30,000 with no businessmen or students falling in this range

Correlation between Age Group Vs Amount Credit.



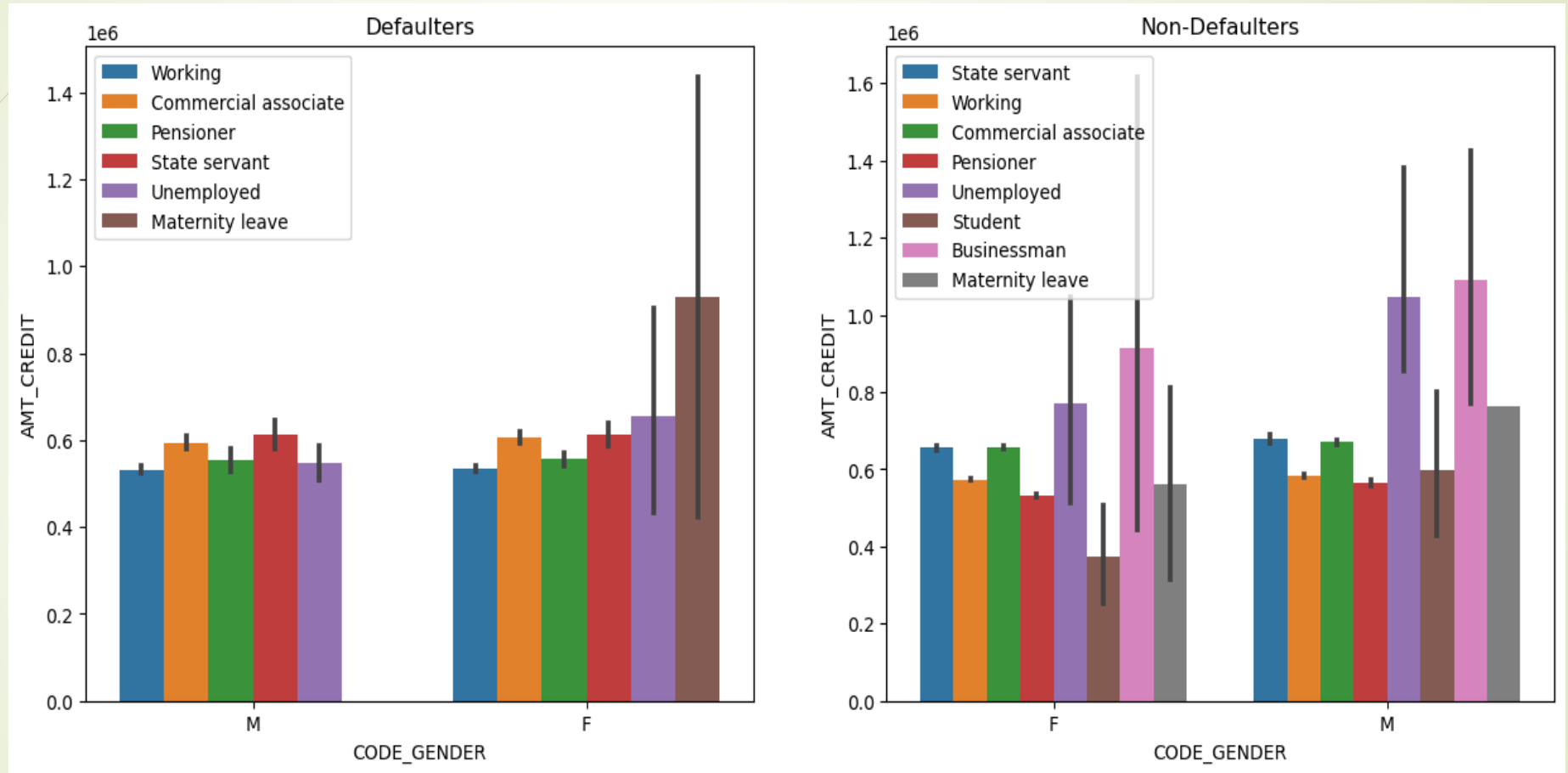
- The age group with the highest credit utilization is 40-50, followed by 50-60. Defaulters under 30 have low credit amounts and struggle with loan repayment, possibly due to unstable employment

Correlation between Name Education Type Vs Amount Credit Vs Name Family Status.



- Defaulters include all married individuals with an academic degree who have not repaid their loans. Non-defaulters, except those with academic degrees, have taken the highest amount of credit

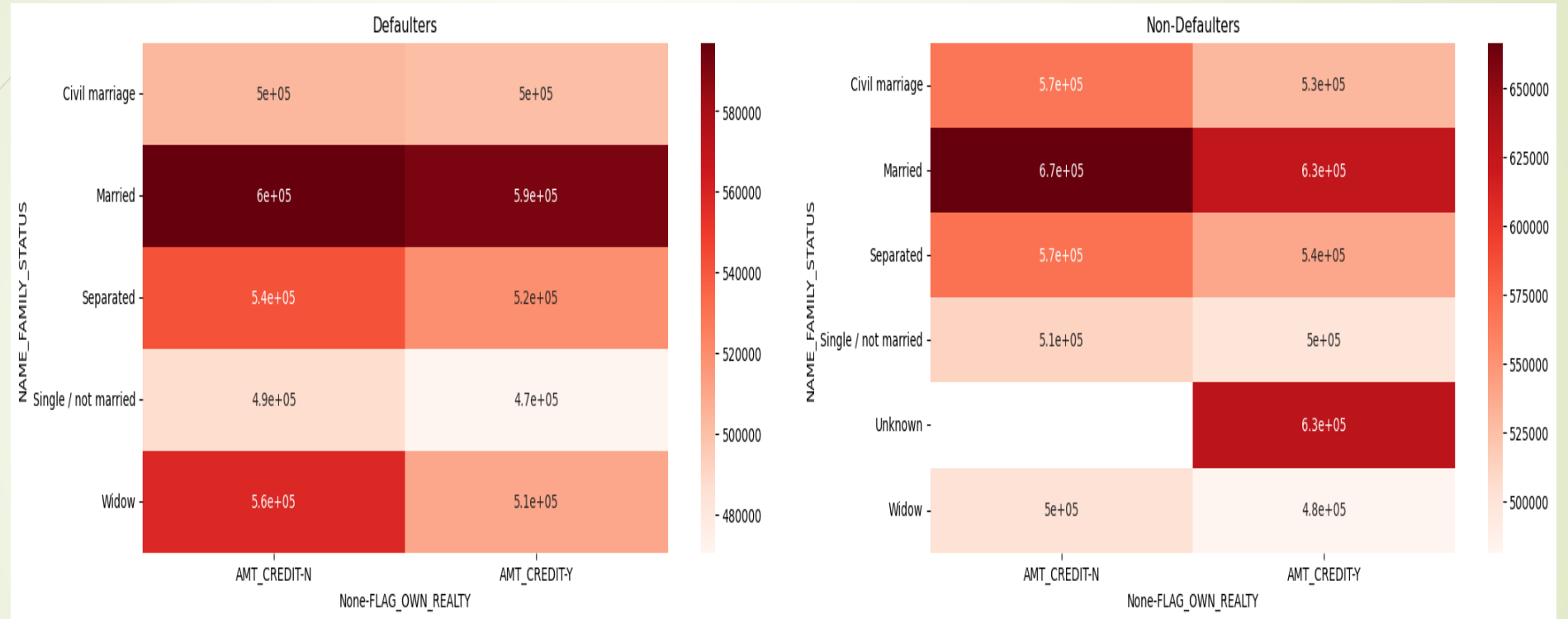
Correlation between Code Gender Vs Amount Credit Vs Name Income Type.



- Pregnant women tend to borrow more and struggle to pay back, while those who borrow less repay successfully. No students or businessmen have failed to repay loans. Surprisingly, unemployed men and women have successfully repaid loans, even after borrowing large amounts

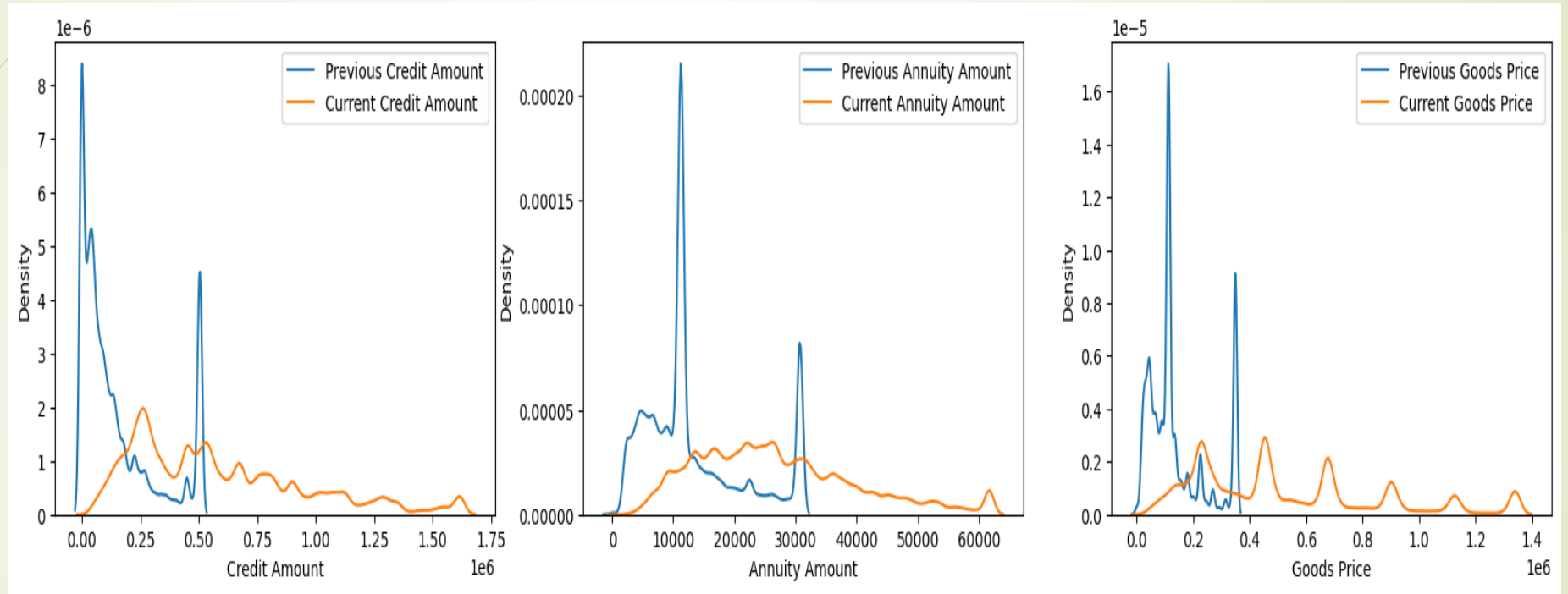
Multivariate Analysis

Correlation between Name family status Vs Flag Own Realty Vs Amount Credited.



- Widows without property have taken large loans and struggle to pay them back, although some have successfully repaid.
- Married individuals without property often take out substantial loans
- Individuals in civil marriages who default on loans have similar credit amounts to those who do not default, but they still struggle to repay

➤ Correlation between Credited Amount Vs Annuity Amount Vs Goods Price.



- A comparable pattern can be observed in the pricing chart of past applications, while in new applications, most individuals bought goods with amounts close to 3L, 5L, 7L, specifically in odd figures
- The earlier candidates had a highest credit score of 1L-2L, but a sharp decline was seen in the credit range of 2L-4L followed by an increase at 5L. The present candidates appear to be on a typical trend



Final Summary

- Based on the provided analysis of the Credit EDA (Exploratory Data Analysis) assignment, here are five significant insights and conclusions:

1. Income and Default Risk:

1.1 Lower-income individuals, especially those earning between 1L-1.5L, exhibit a higher likelihood of defaulting on loans compared to higher-income groups.

2. Loan and Annuity Relationship:

2.1 There's a consistent relationship between the annuity amount and credit amount, suggesting that loan size influences annuity payments.

3. Age and Credit Utilization:

3.1 Individuals aged 40-50 and 50-60 have the highest credit utilization, while younger defaulters under 30 demonstrate lower credit amounts, possibly due to unstable employment.



4. Gender, Occupation, and Repayment:

4.1 Pregnant women tend to borrow more and face challenges in loan repayment, while unemployed individuals, both male and female, have successfully repaid loans despite borrowing large amounts.

5. Family Status and Loan Repayment:

5.1 Widows without property and married individuals without property struggle to repay loans, indicating financial strain post-loss of a spouse or within certain marital dynamics, while civil marriages show similar credit amounts between defaulters and non-defaulters.

- These insights highlight the intricate relationship between socioeconomic factors, personal circumstances, and borrowing behaviour, influencing loan repayment outcomes