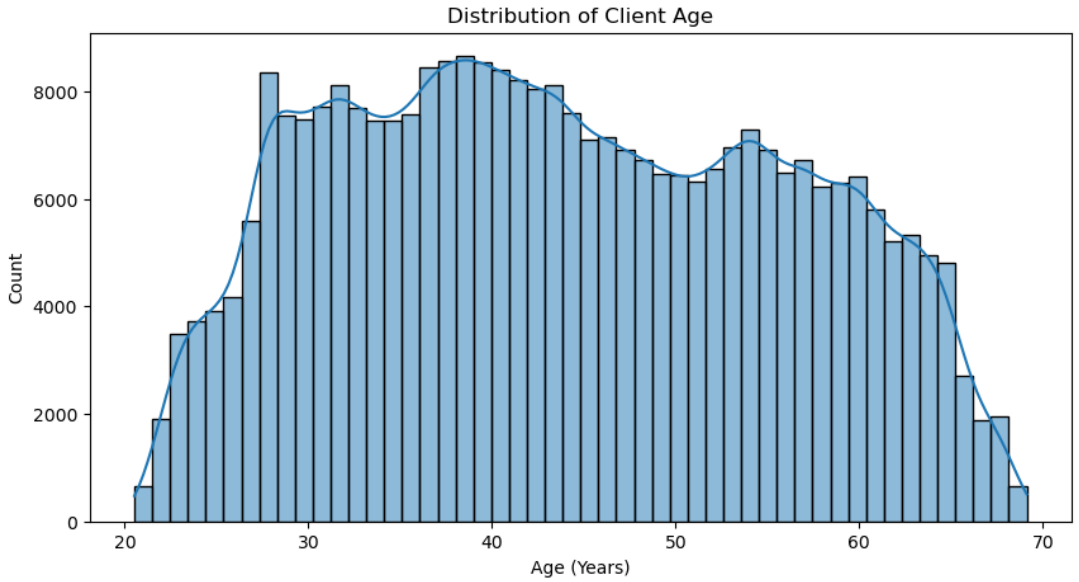
Credit EDA Case Study

By Thang Truong

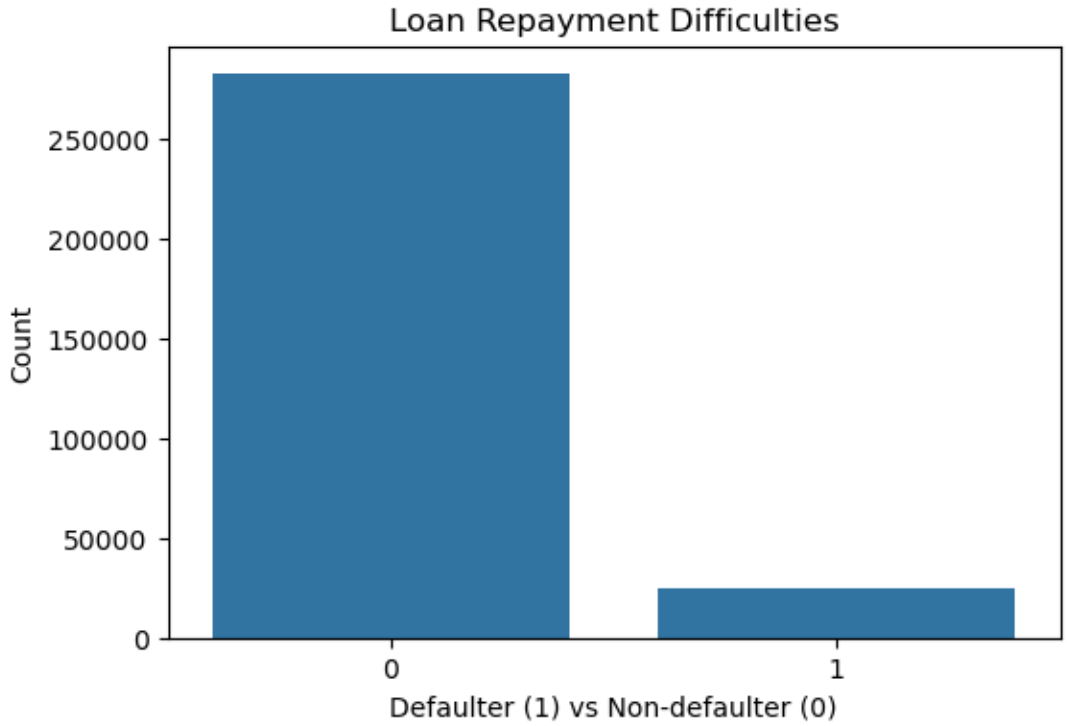
Distribution of Client Age



Conclusions from the graph:

- Most clients fall within the age range of 30 to 40 years old.
- Risk Assessment:
  - Younger Clients: Often associated with higher risk due to less stable income and shorter credit history.
  - Older Clients: Tend to be more financially stable with established credit histories.

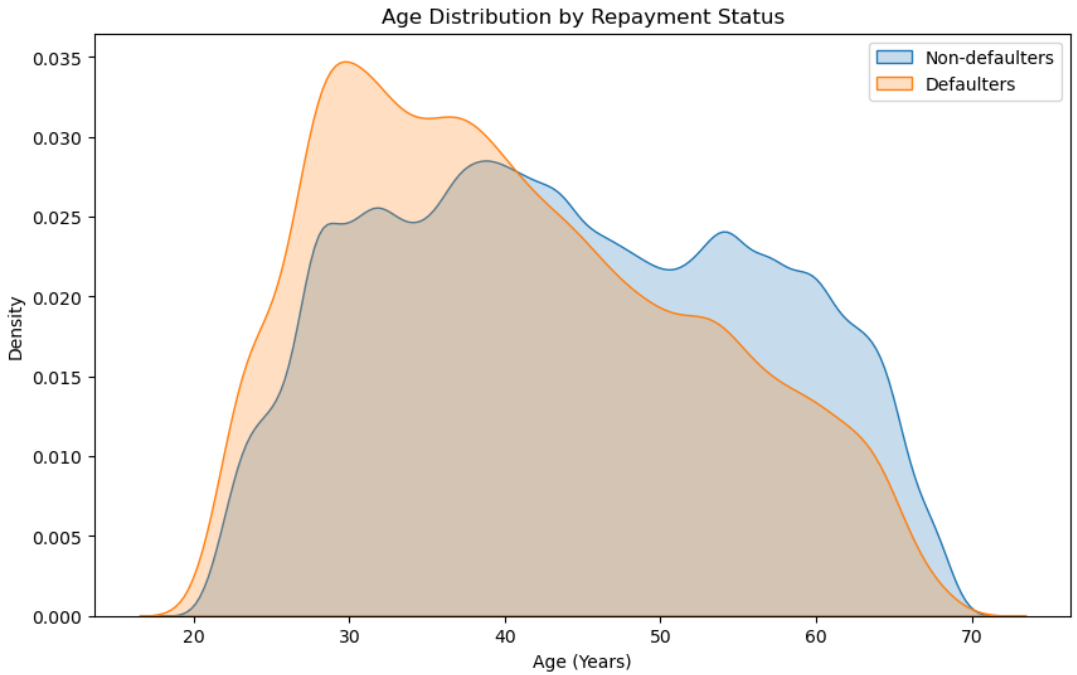
Loan Repayment Difficulties



Conclusions from the graph:

- The majority of loan applicants are successfully repaying their loans, with fewer experiencing repayment difficulties.

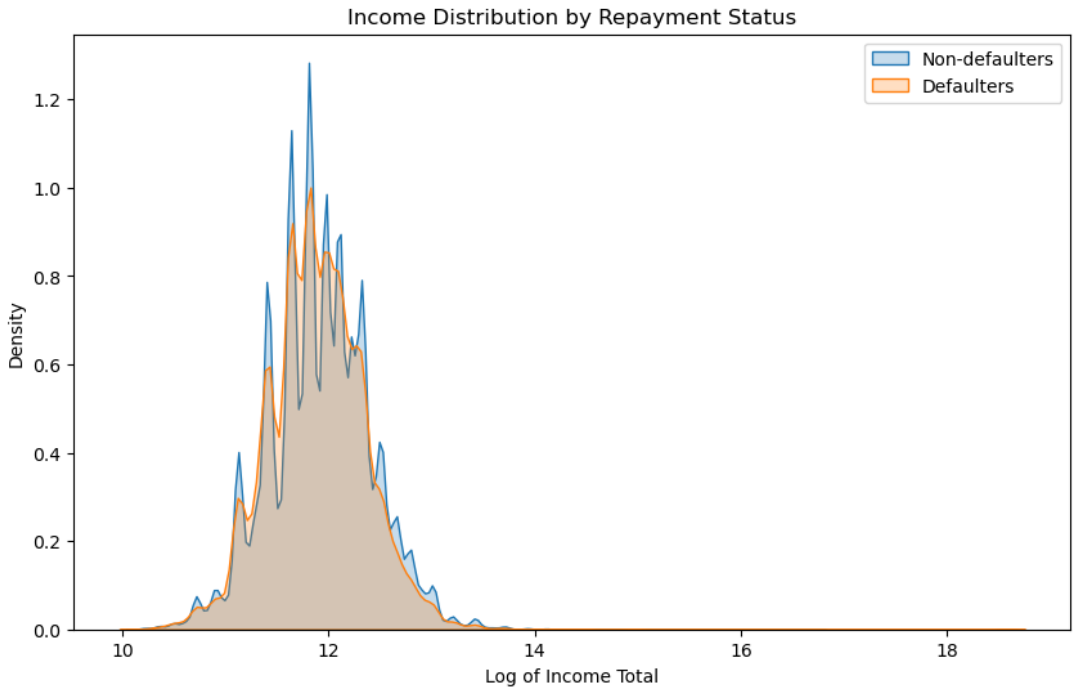
Age Distribution by Repayment Status



Conclusions from the graph:

- Young customers (under 30 years old) face more challenges in loan repayment, possibly due to less stable income or lack of experience in financial management.
- The majority of our customers fall within the 30 to 40-year-old age range, a group with a better repayment rate.

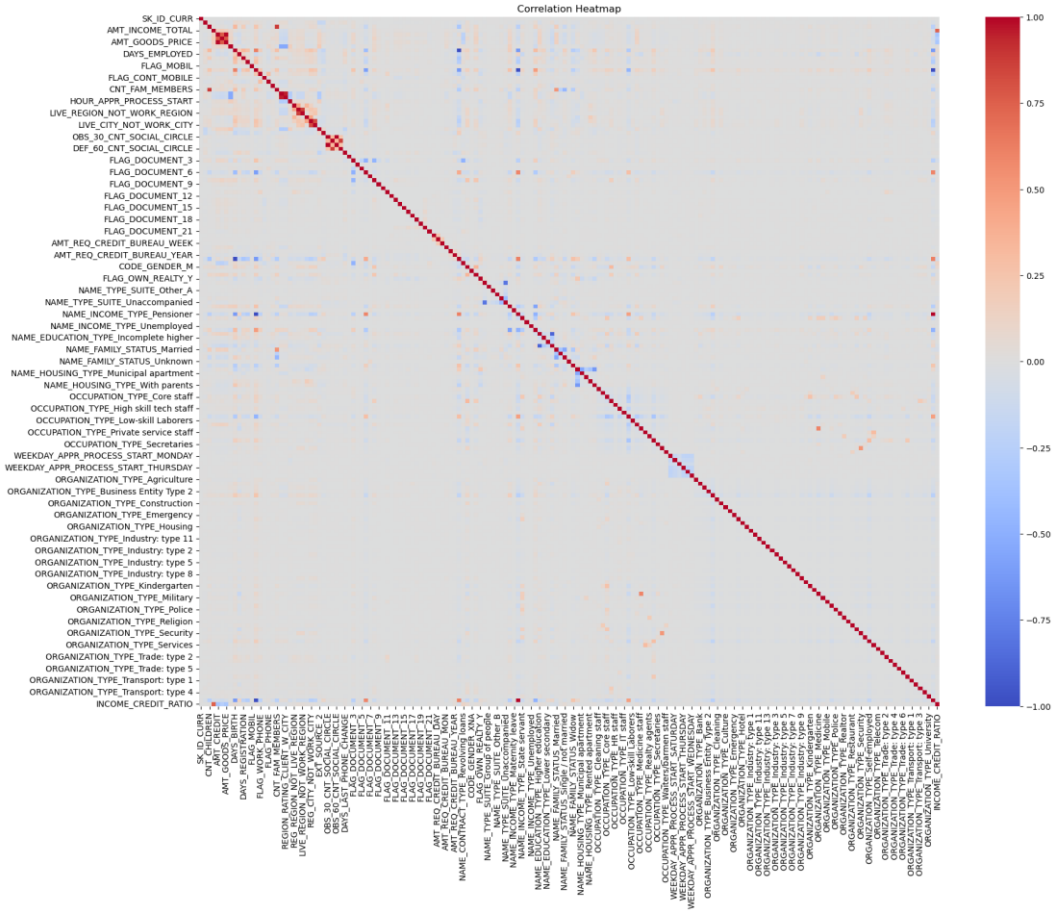
Income Distribution by Repayment Status



Conclusions from the graph:

- In the graph, you likely observed a strong negative correlation between income and default status.
- Essentially, as individuals' incomes increase, their probability of defaulting decreases.

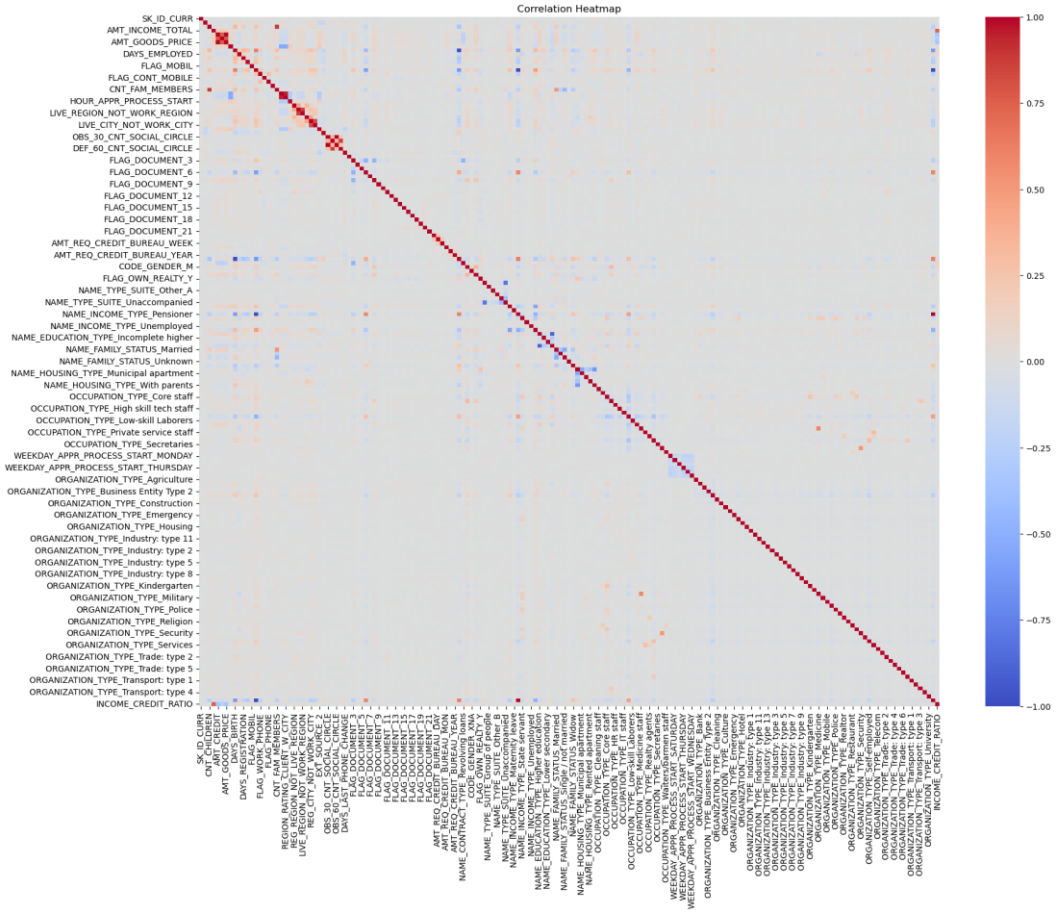
Correlation Heatmap



Conclusions from the graph:

- Income Level:
  - If AMT_INCOME_TOTAL has a strong negative correlation with TARGET, it confirms that higher-income individuals are less likely to default—a critical insight for risk assessment.
- Age and Years Employed:
  - Variables like DAYS_BIRTH and DAYS_EMPLOYED might be highly correlated, as older applicants typically have longer employment histories.

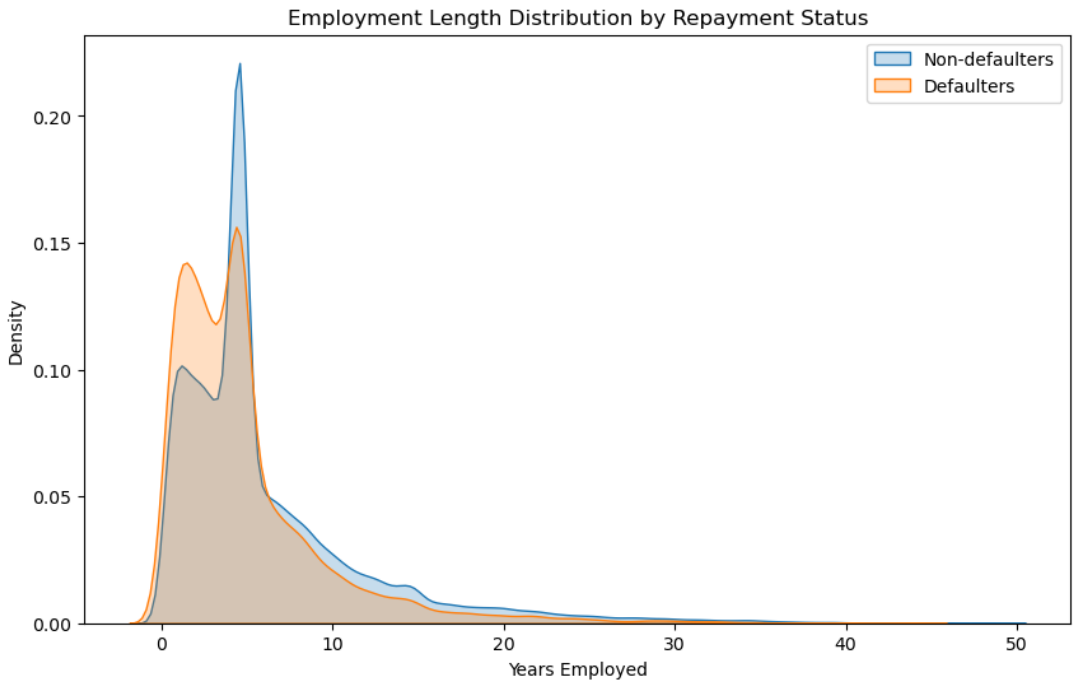
Correlation Heatmap



Conclusions from the graph:

- Education Level:
  - If NAME_EDUCATION_TYPE shows a strong correlation, it suggests education significantly impacts default risk, perhaps more than anticipated.
- Credit Bureau Scores:
  - Variables related to credit history might cluster together, emphasizing the importance of past credit behavior.
- Employment Type:
  - If NAME_INCOME_TYPE correlates with default, certain employment categories may require closer scrutiny.

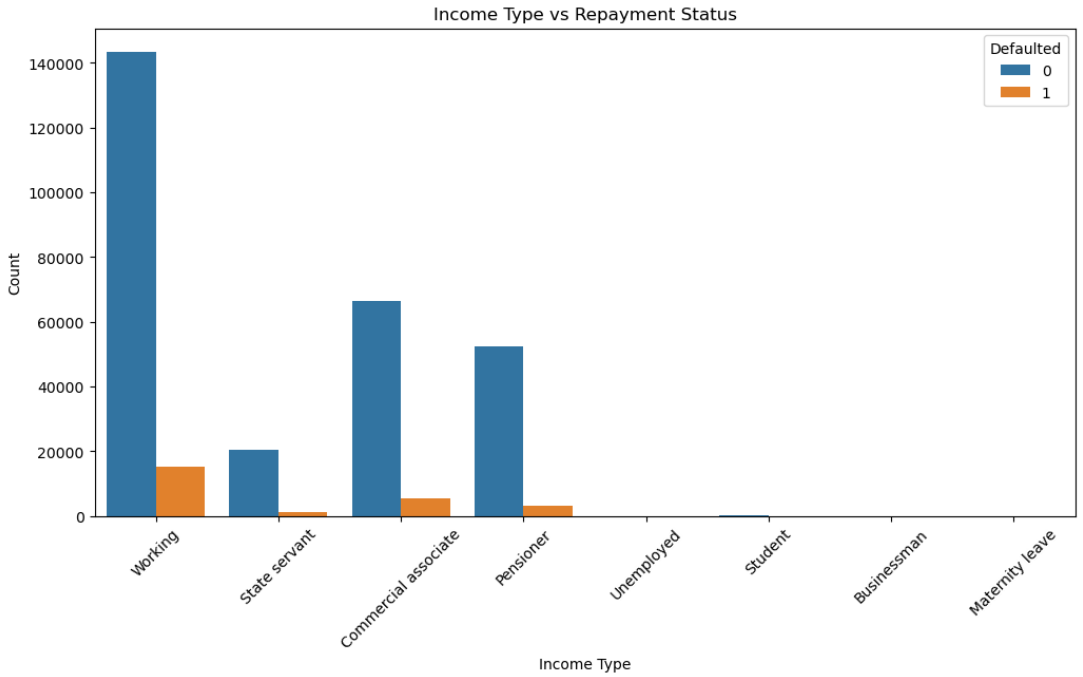
Employment Length Distribution by Repayment Status



Conclusions from the graph:

- The clients with longer employment histories are more likely to repay loans successfully.
- The clients with shorter employment durations are more prone to default.

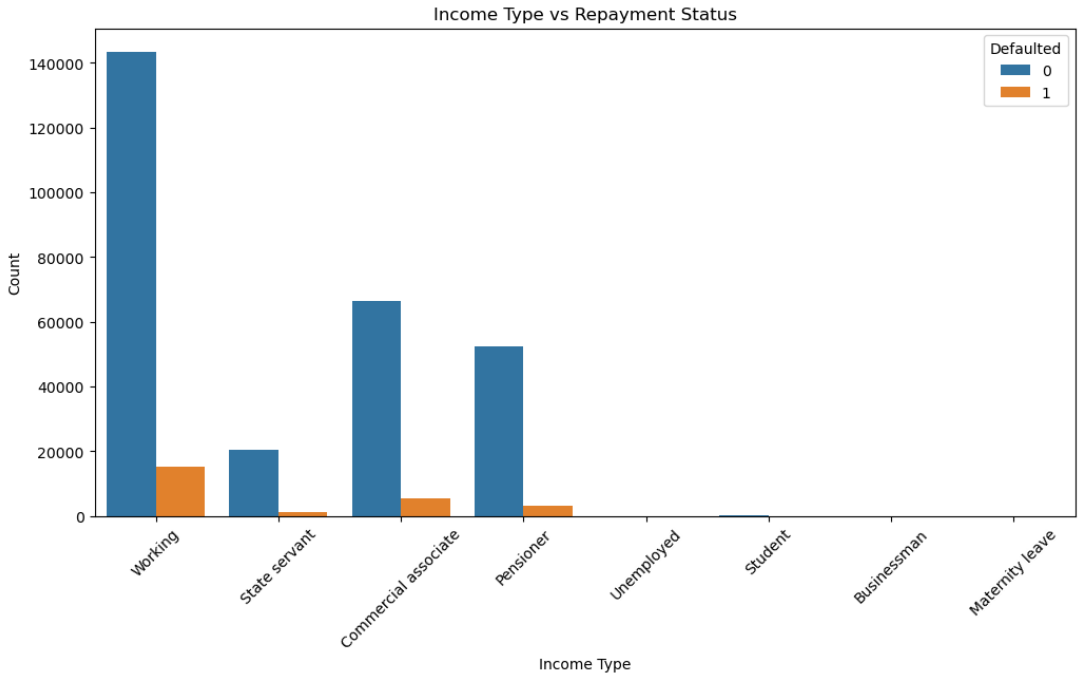
Income Type vs Repayment Status



Conclusions from the graph:

- The categories Unemployed, Businessman, and Maternity leave either do not need to borrow money or are unable to borrow money.
- Students have a low borrowing capacity but are all capable of repaying their loans.

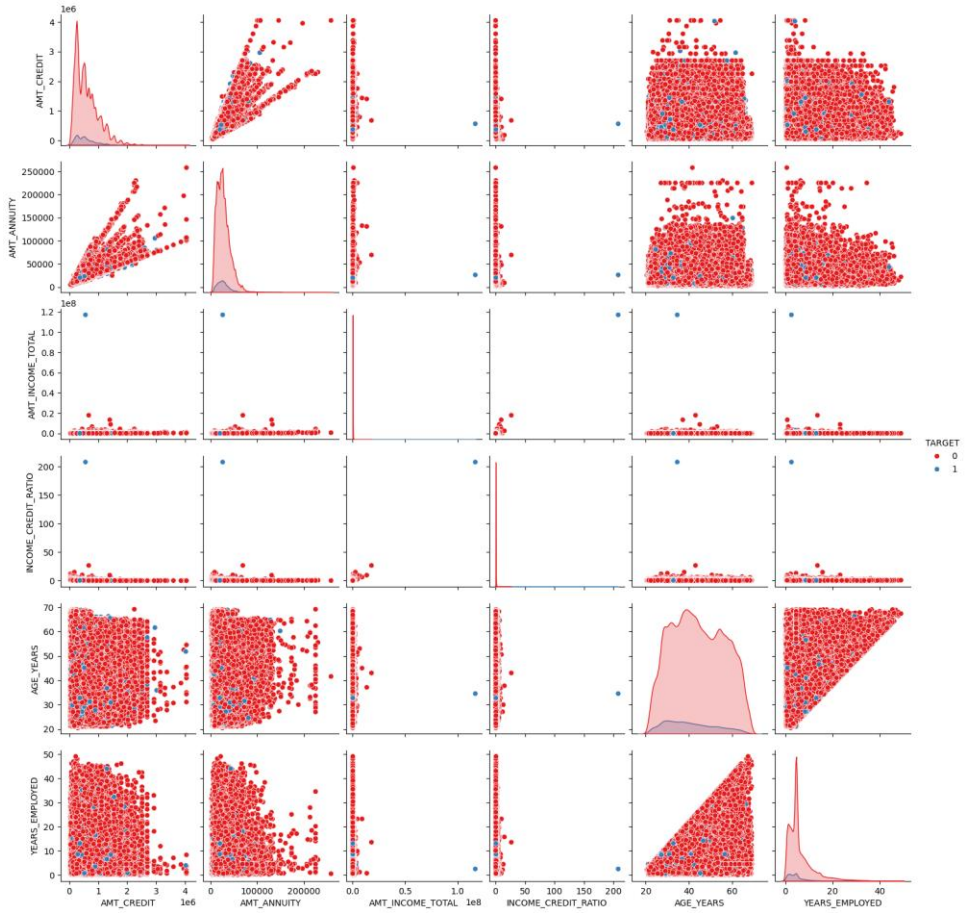
Income Type vs Repayment Status



Conclusions from the graph:

- Among the remaining four categories, State servants have the lowest borrowing amounts, and their default rate is very low (nearly 0%).
- For the other three categories (Working, Commercial associate, and Pensioner), those in the Working category borrow the most, but the default rate for all three categories is around 5%.

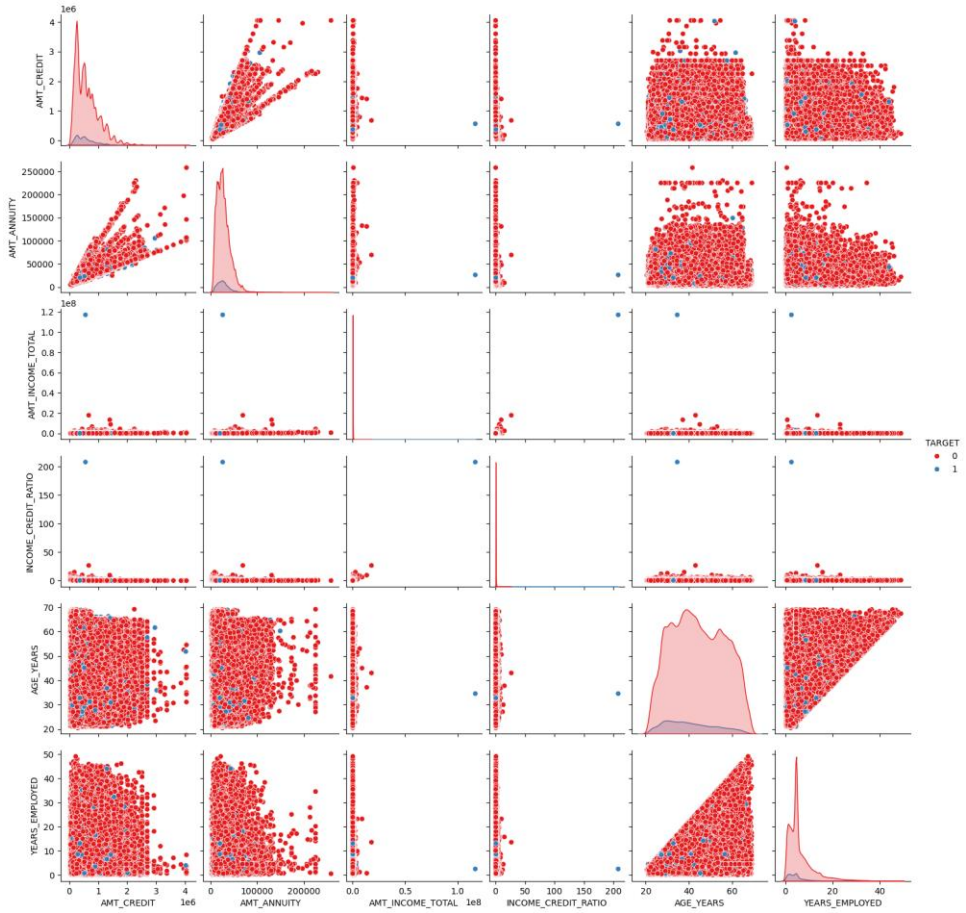
Multivariate Analysis



Conclusions from the graph:

- Income and Credit Amount
  - Non-defaulted applicants (TARGET = 0) tend to have higher incomes and higher credit amounts, whereas defaulted applicants (TARGET = 1) cluster around lower income levels.
  - Higher income levels are associated with larger loan amounts and a lower risk of default. This supports the idea that income is a strong predictor of loan repayment capability.

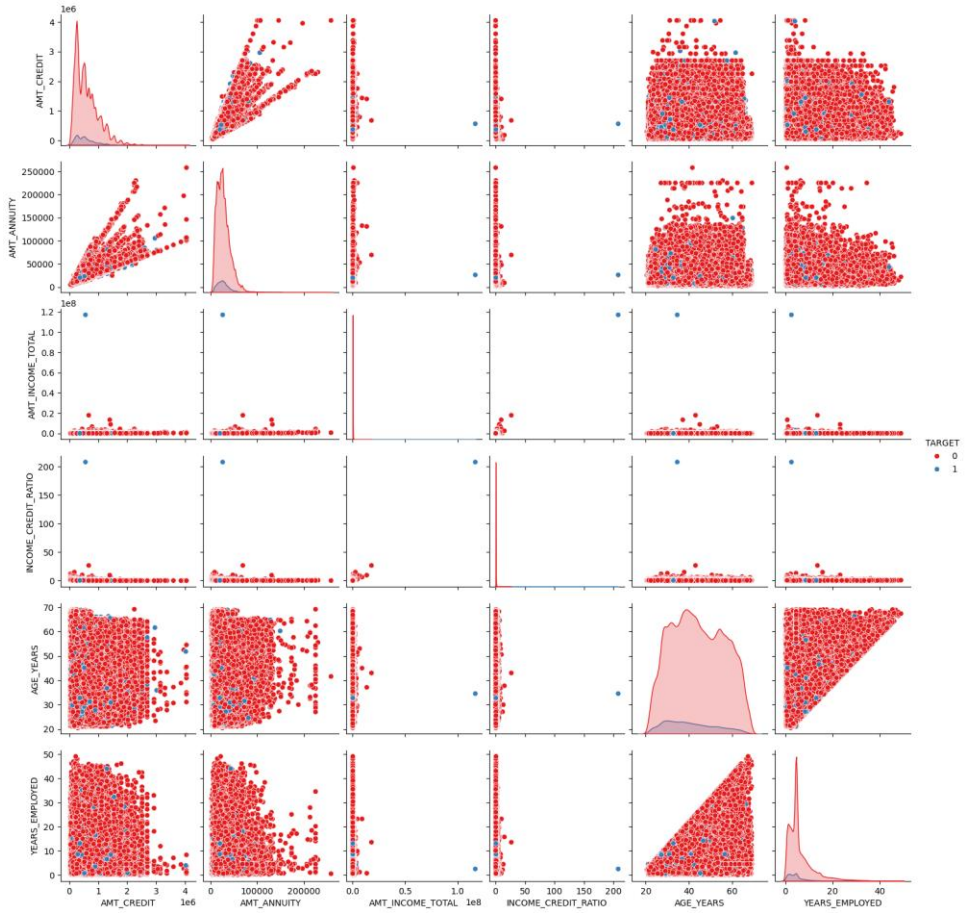
Multivariate Analysis



Conclusions from the graph:

- Income and Annuity Amount
  - There's a positive correlation between income and annuity amounts for both defaulted and non-defaulted applicants. However, defaulted applicants often have a higher annuity relative to their income.
  - Applicants who default may be over-leveraged, taking on repayment amounts that are high relative to their income. This indicates that the debt burden (as represented by the annuity) is a critical factor in default risk.

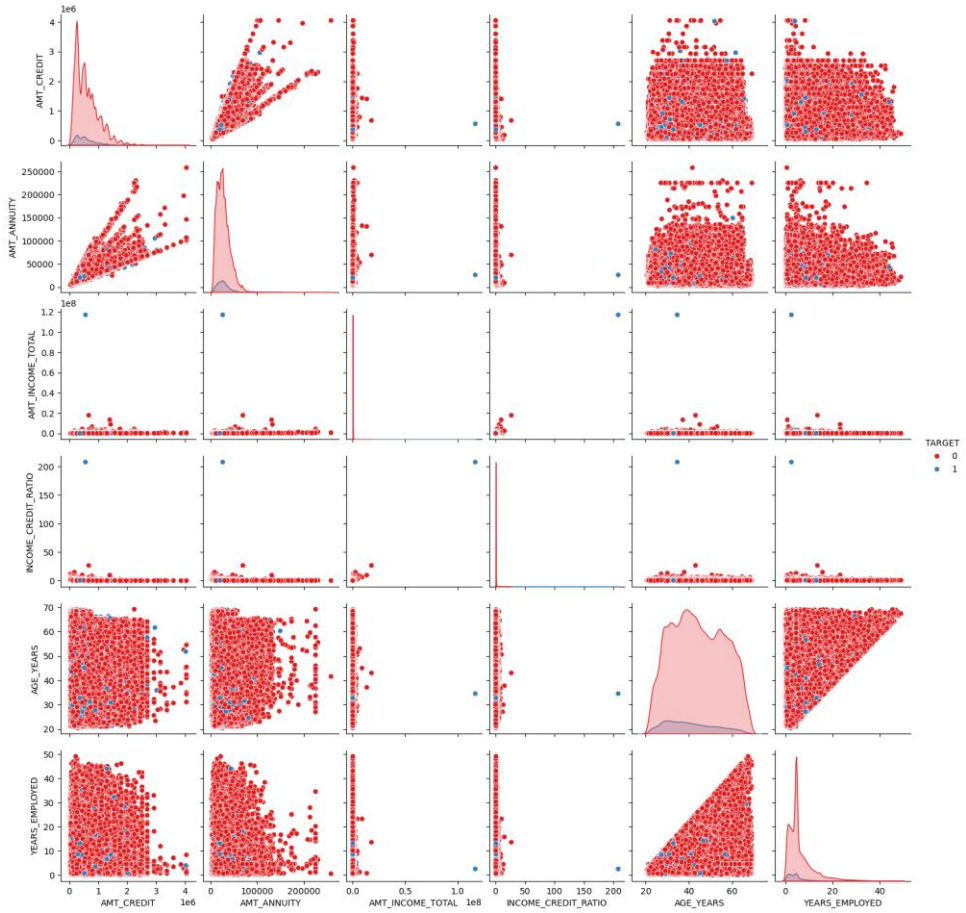
Multivariate Analysis



Conclusions from the graph:

- Income Credit Ratio
  - Defaulted applicants have a lower INCOME_CREDIT_RATIO, meaning their income is low relative to their credit amount.
  - A lower income-to-credit ratio increases the likelihood of default, emphasizing the importance of ensuring that loan amounts are proportionate to the applicant's income.

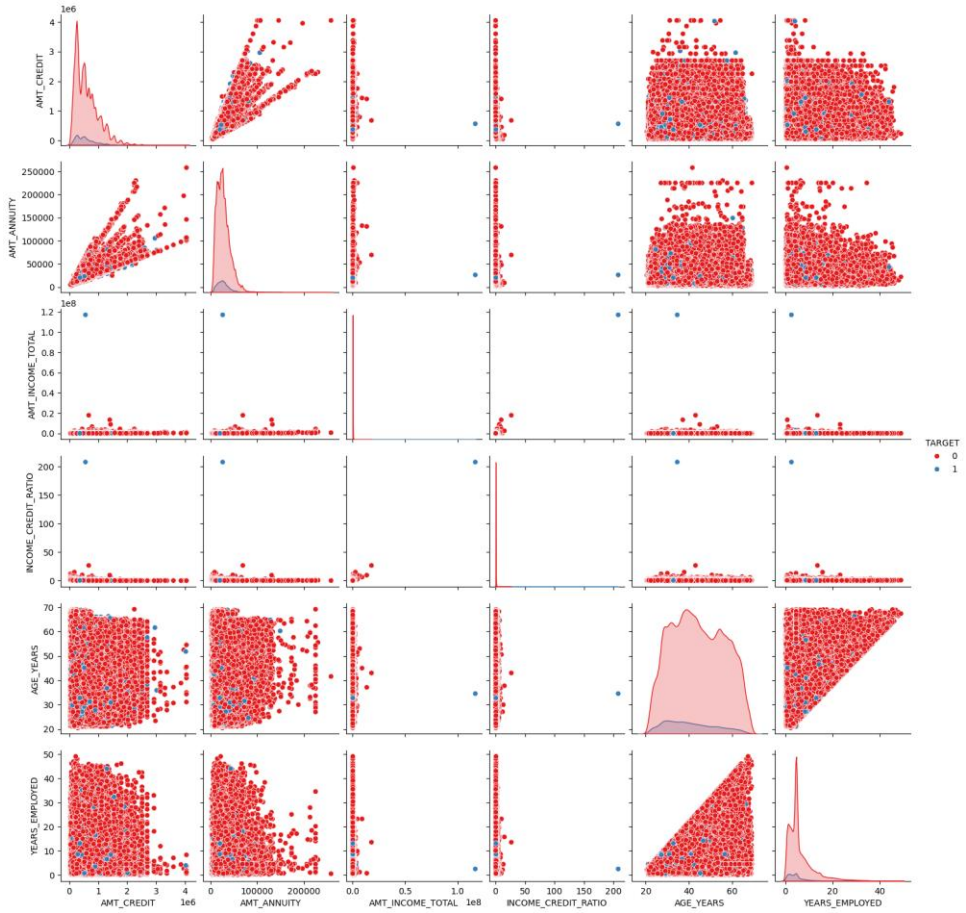
Multivariate Analysis



Conclusions from the graph:

- Age vs. Default Status
  - Younger applicants have a higher concentration in the defaulted group, while older applicants are more represented in the non-defaulted group.
  - Age appears to be inversely related to default risk; with increasing age, the probability of default decreases. This could be due to factors like financial stability improving with age.

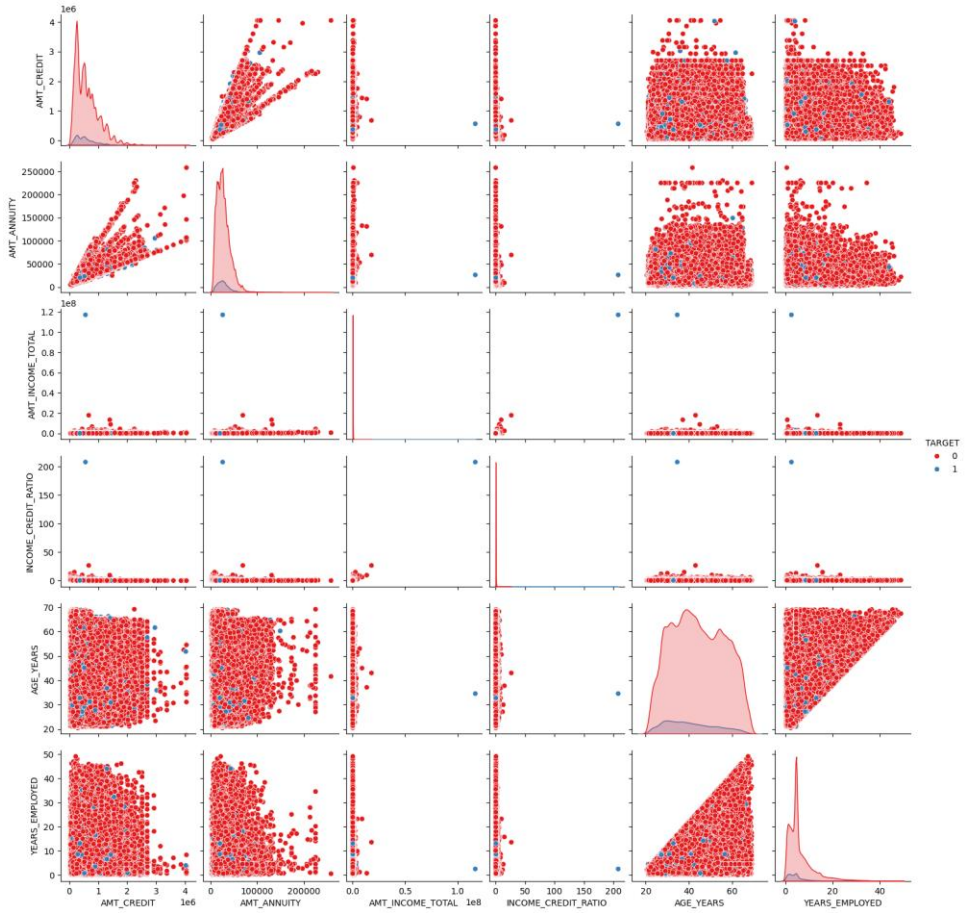
Multivariate Analysis



Conclusions from the graph:

- Years Employed
  - Applicants with more years of employment tend to default less. Defaulted applicants often have shorter employment histories.
  - Employment stability is a strong indicator of creditworthiness. Longer employment tenure may reflect steady income and job security, reducing default risk.

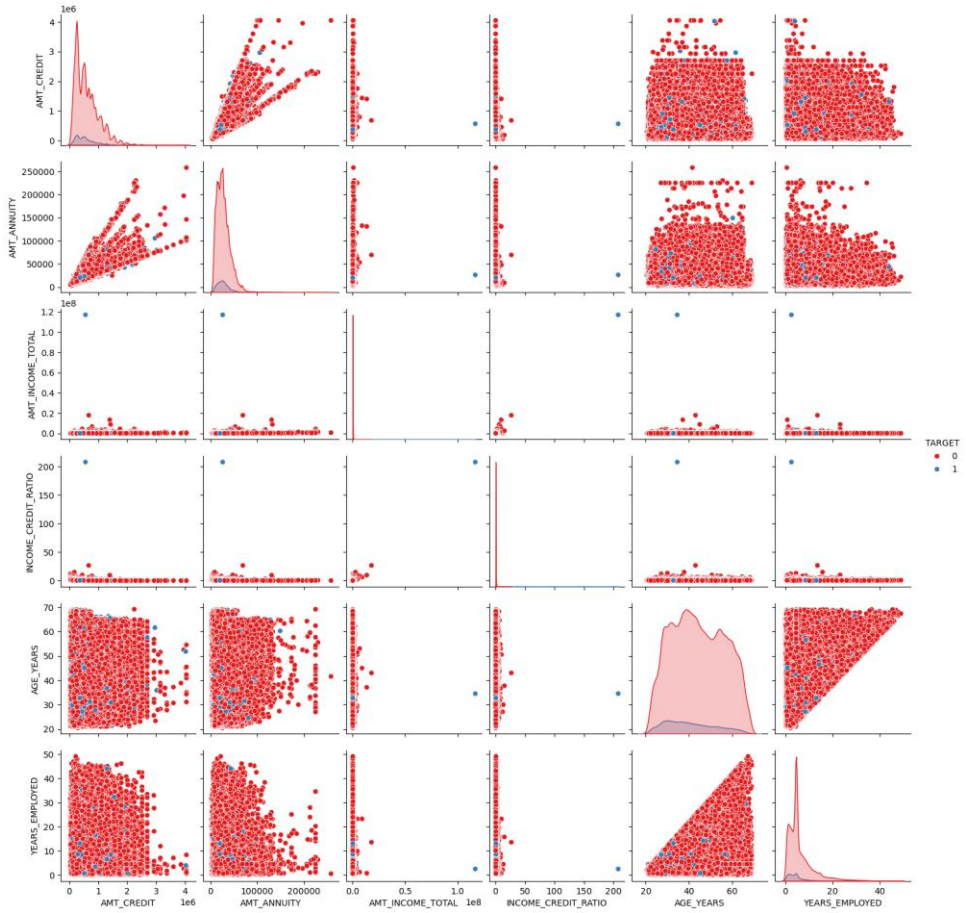
Multivariate Analysis



Conclusions from the graph:

- Credit Amount and Annuity
  - There's a strong positive correlation between the credit amount and the annuity for both groups, which is expected since larger loans require higher annual repayments.
  - While this relationship is logical, it's important to ensure that the annuity is set appropriately relative to the applicant's income to minimize default risk.

Multivariate Analysis



Conclusions from the graph:

- Interactions Between Features
  - Combination of Low Income and High Credit Amount: Applicants who have lower incomes but are granted high credit amounts are predominantly in the defaulted group.
  - Income and Years Employed: Those with both higher income and longer employment are less likely to default, suggesting a compounded effect of these two factors on default risk.

Insights and Recommendations

- Age Factor: Younger clients might have higher default rates.
- Income Level: Lower income levels could be associated with higher defaults.
- Employment Length: Clients with shorter employment history may be riskier.
- Income Type: Certain income types might be more prone to defaults.