

A wooden pencil with a dark eraser and a sharpened lead tip lies diagonally across the frame. It rests on a document featuring a line graph with a grid. The graph has a vertical axis with labels '100' and '50' and a horizontal axis with labels '1993' and '1998'. The background is a soft-focus view of the document and pencil.

EDA Case Study

Credit Risk Analysis

Gaurav Singh Bisht

Problem Statement

- ▶ When the company receives a loan application, the company has to decide for loan approval based on the applicants profile. Two types of risks associated with any loan request: -
 - ❖ If the application is likely to repay the loans, then not approving the loan results in a loss of business to the company.
 - ❖ If the application is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

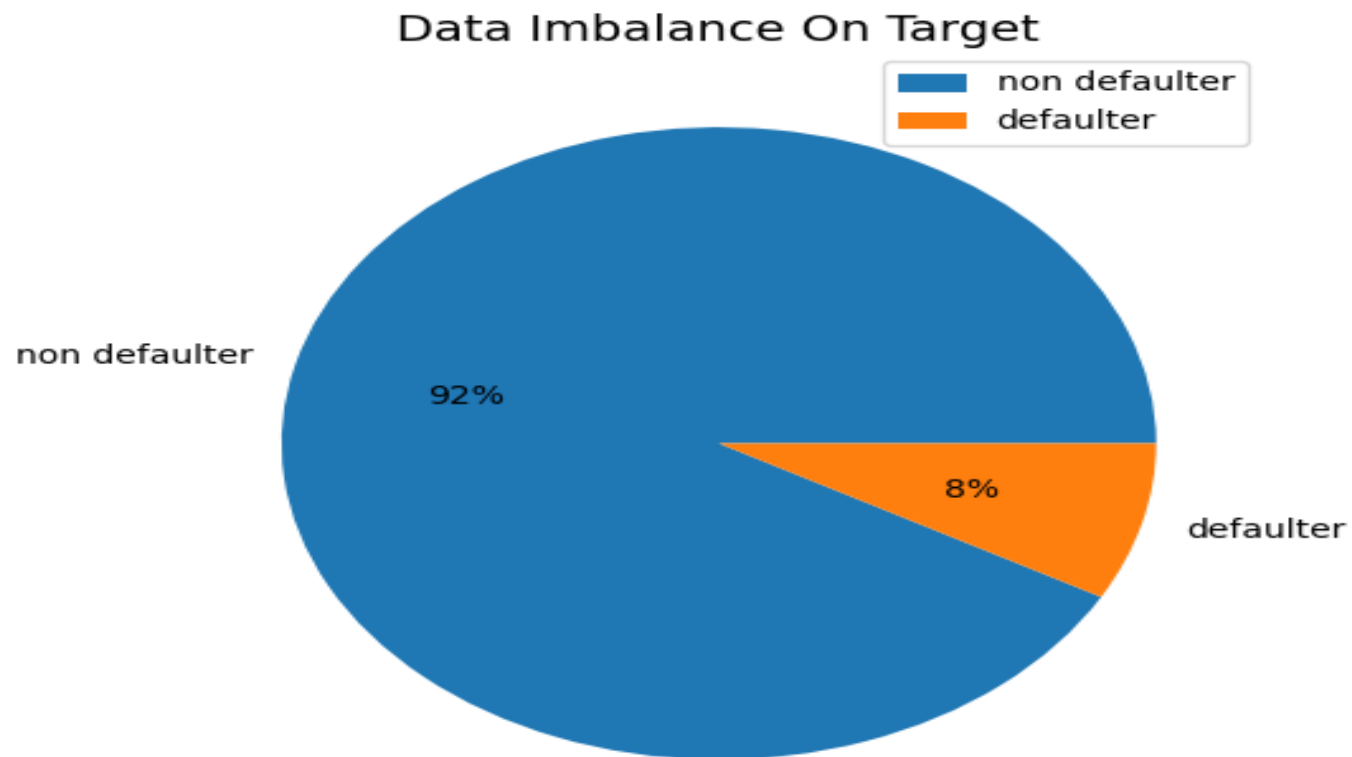
Analysis of the data set has been done in Jupyter Notebook using Python

Step Involved: -

- Understanding the domain/ variables.
- Import/ Load the data.
- Check the structure/ metadata.
- Merging the application data with previous application data
- Missing value check.
- Performing the Univariate Analysis.
- Performing the Bivariate Analysis.

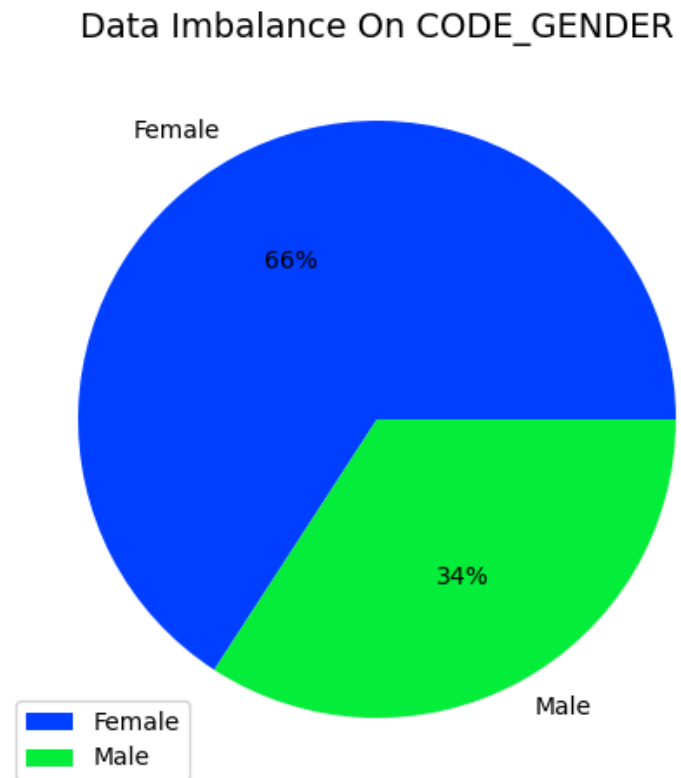
Proportion of Defaulter in the Data set

- Total Defaulters are approx. 8% of the total.



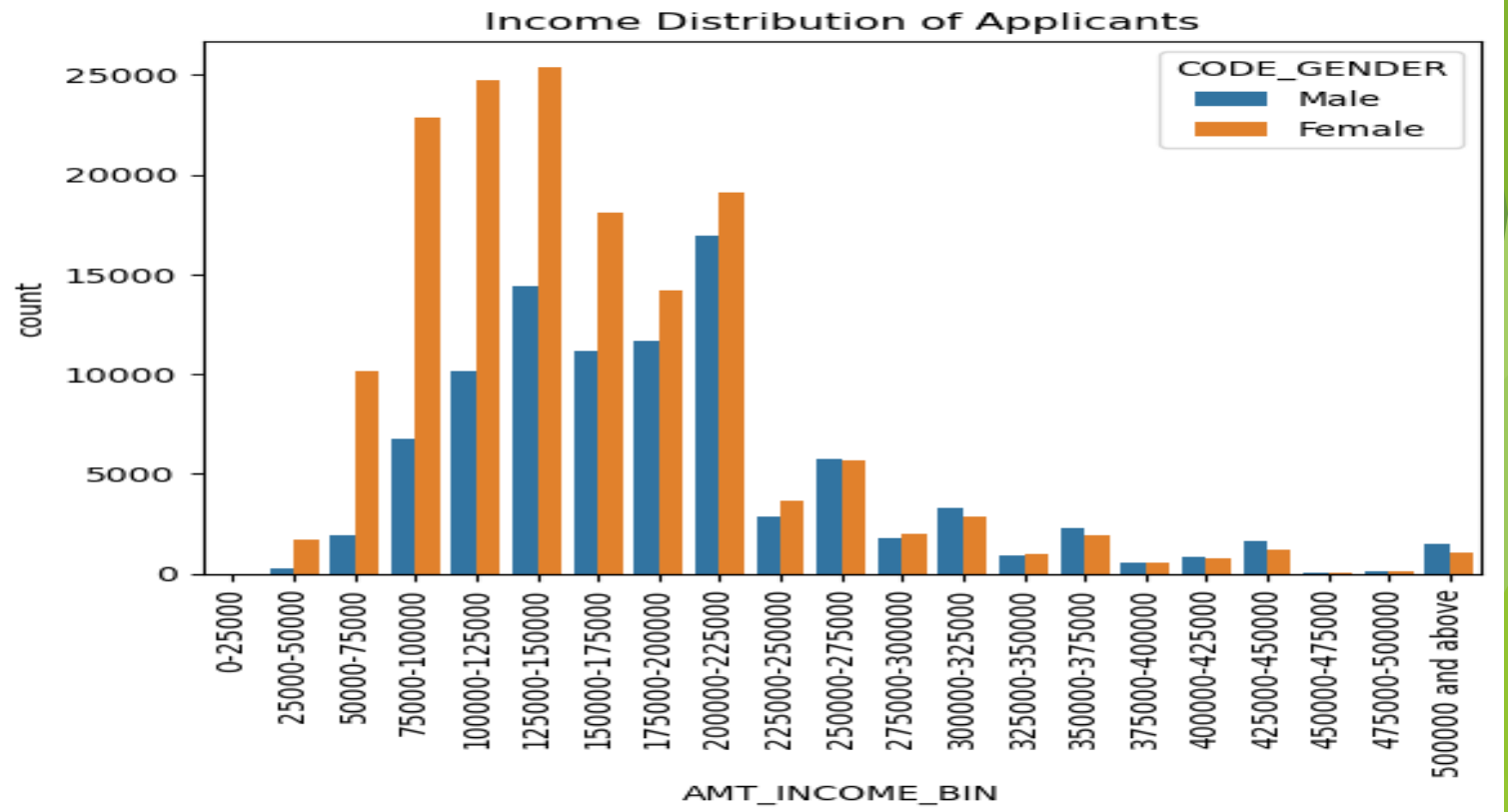
Gender Imbalance in the data set

- From the Total applicant
Approx. 65% are Females.

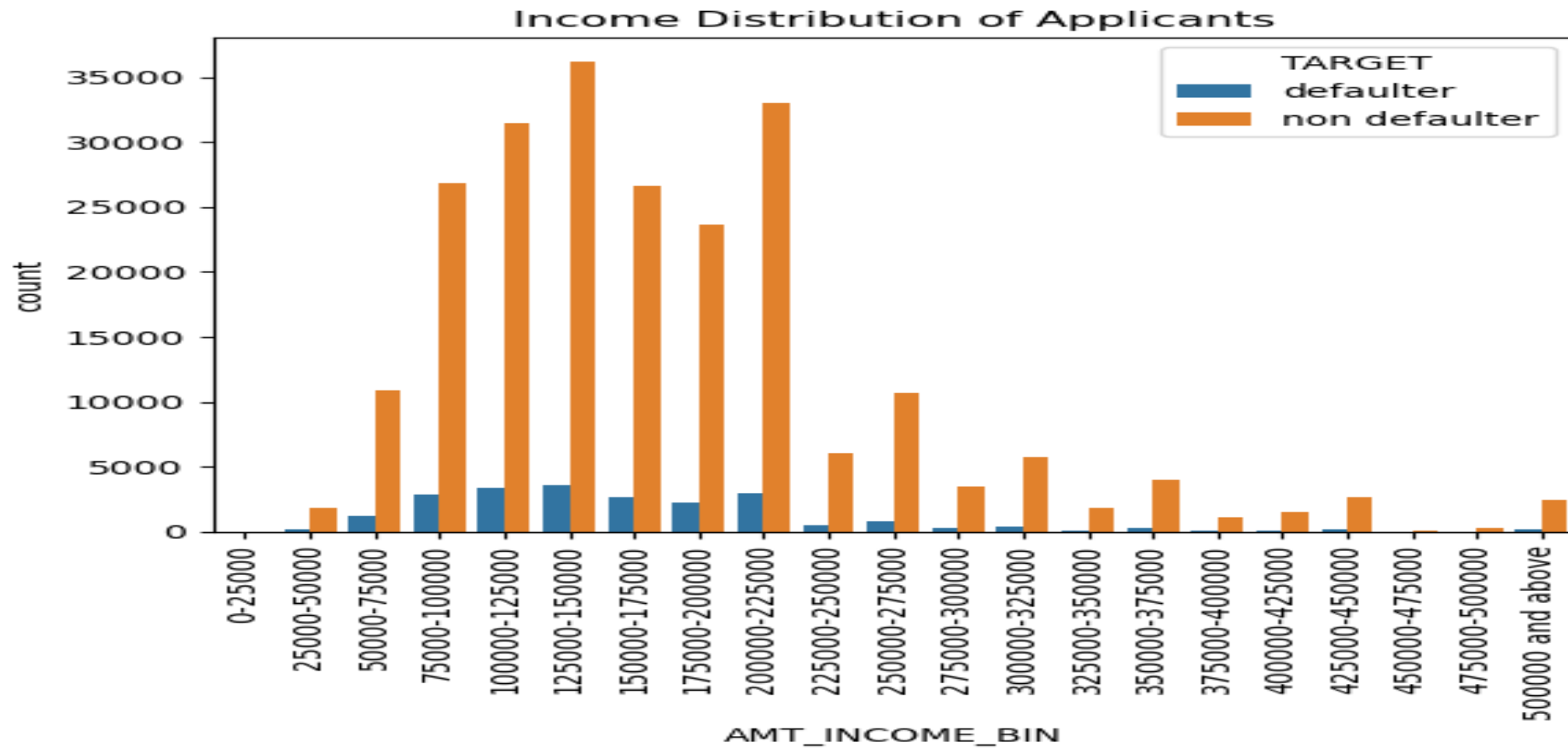


Income Distribution of Applicants

- ❑ Females have more incomes than the males
- ❑ The Difference in incomes decrease as the salary increases
- ❑ Females Applicants are more than Males

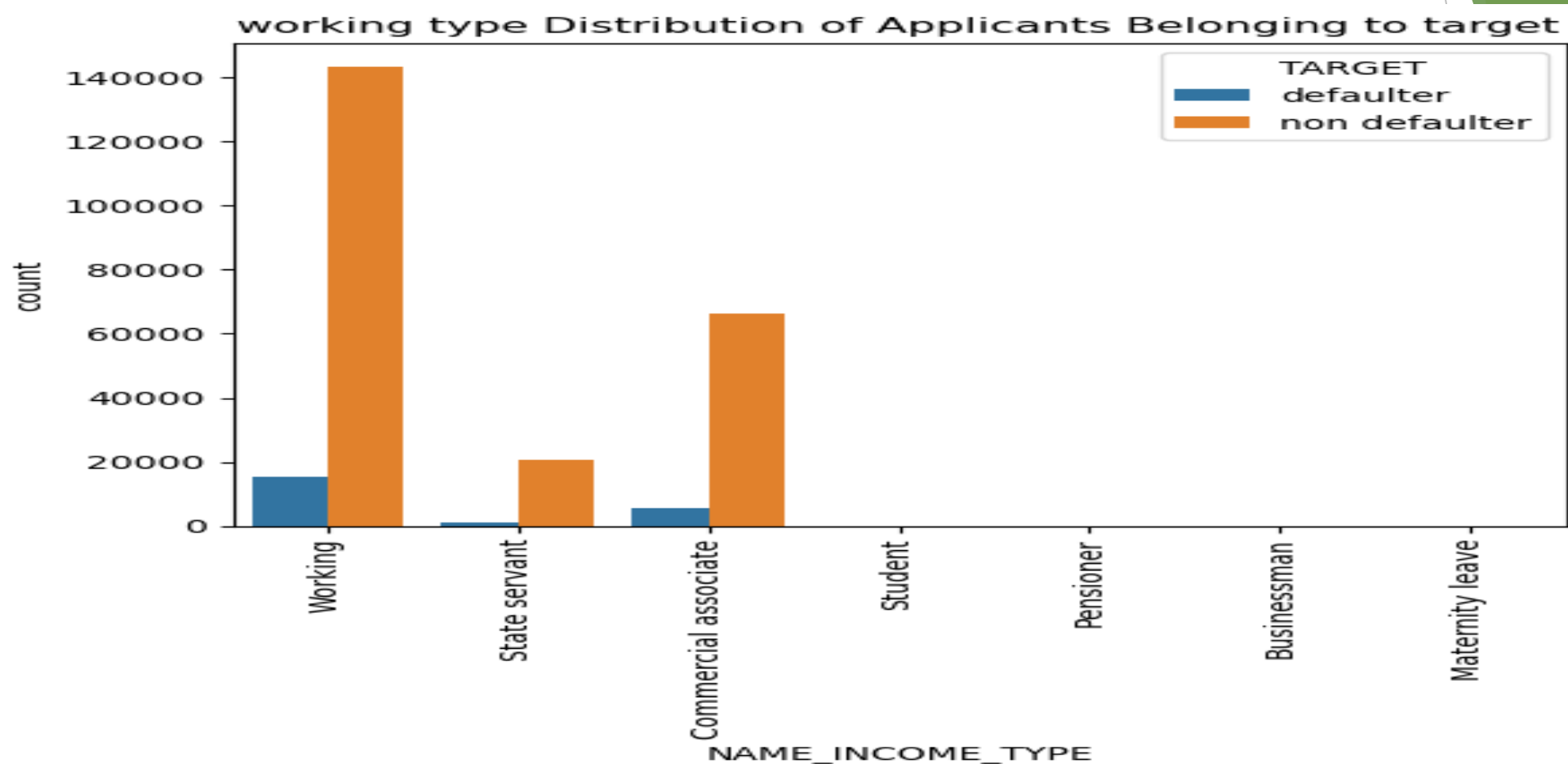


Income distribution of Applicants



- ❖ Income range between 50000 and 225000 has majority of applicant
- ❖ As the salary is more than 250000, the number of the applicants also decreases

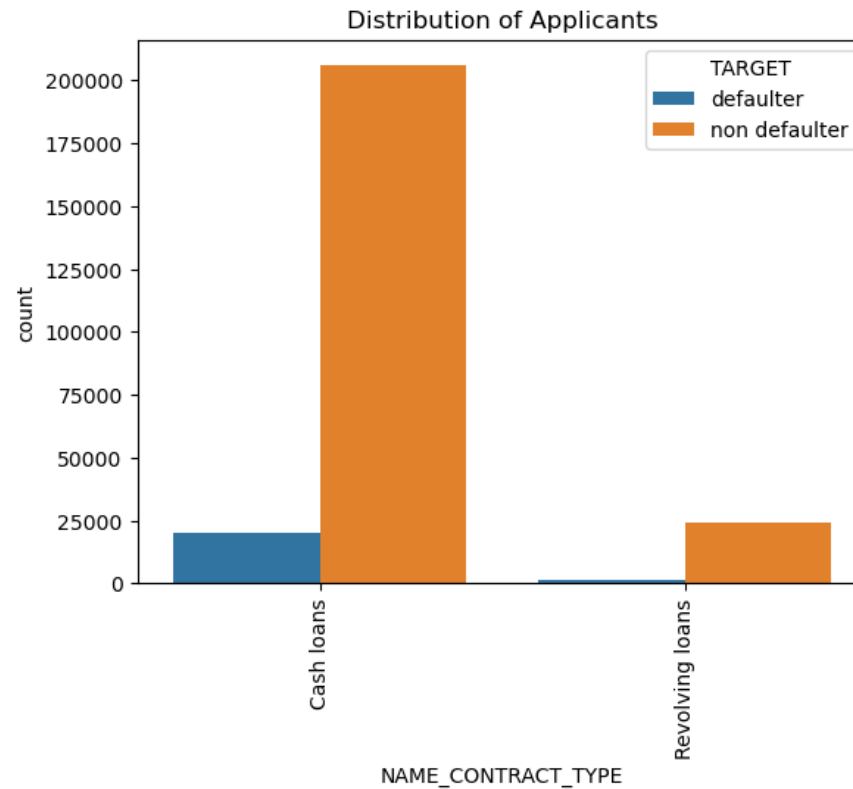
Proportion of Defaulter by Profession



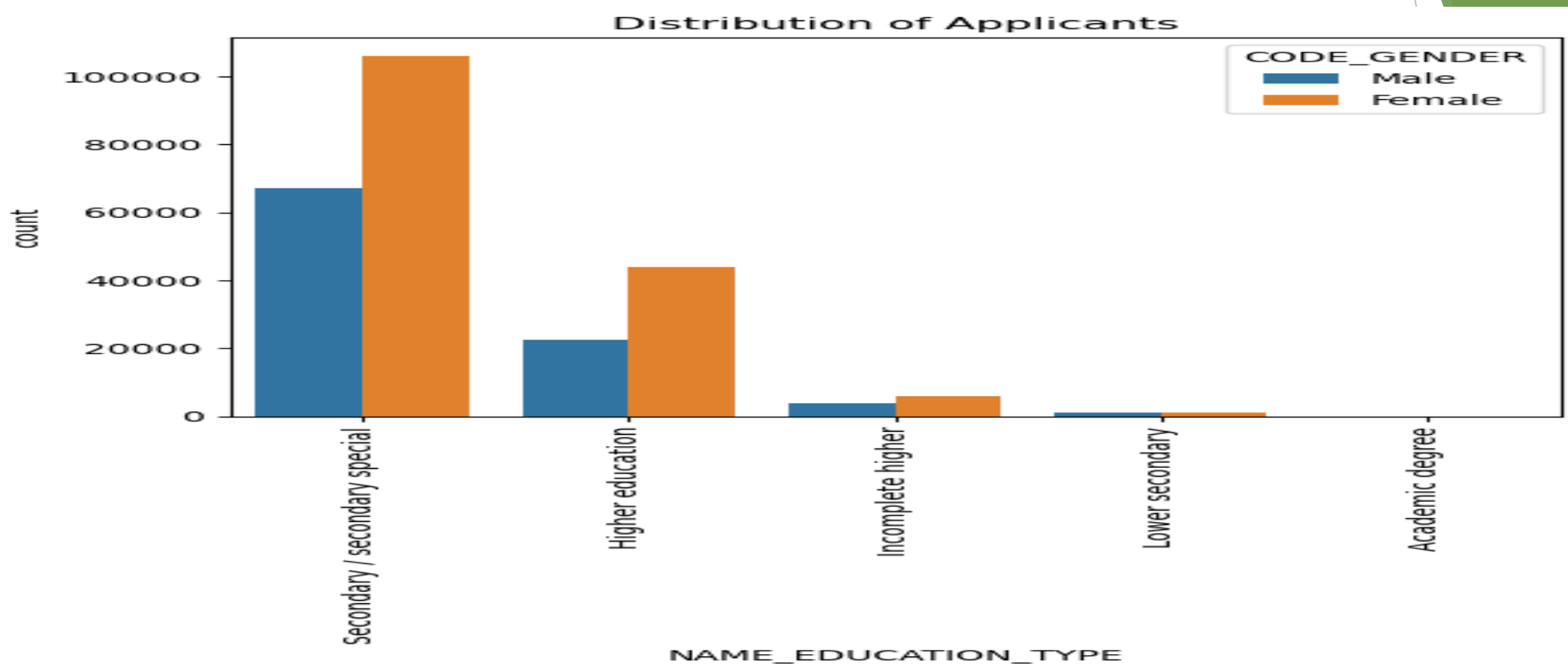
- ▶ 1) Working class applies for the most number of loans applicants and the number of defaulter is also max.
- ▶ 2) Unemployed, Students, Businessman and Maternity leaves has minimum number of applicants

Proportion of Defaulter by contract type

- ▶ 1) Cash Loan contracts have higher number of credit than revolving loans.
- ▶ 2) Number of Defaulters is least in Revolving loans

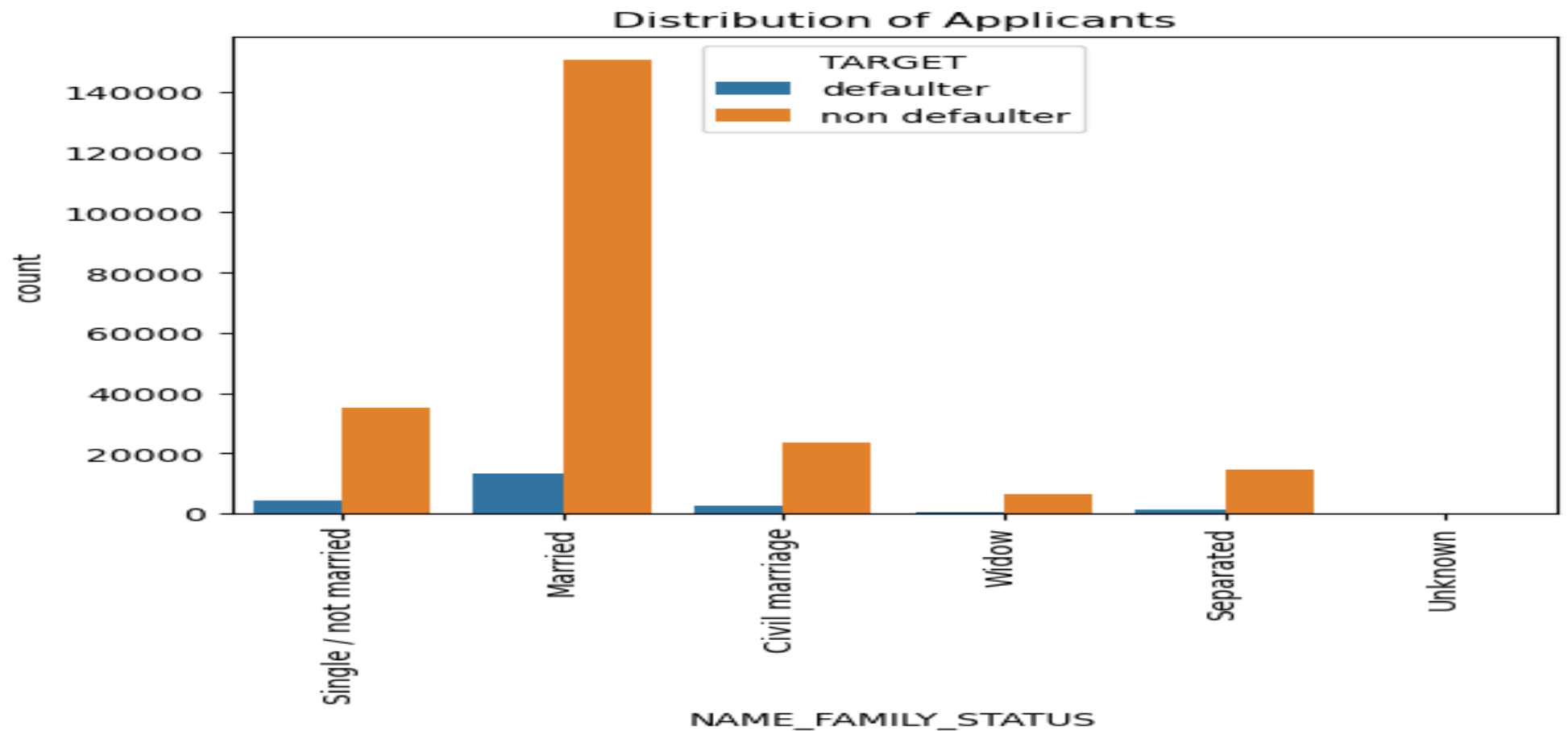


Proportion of Defaulter by education type



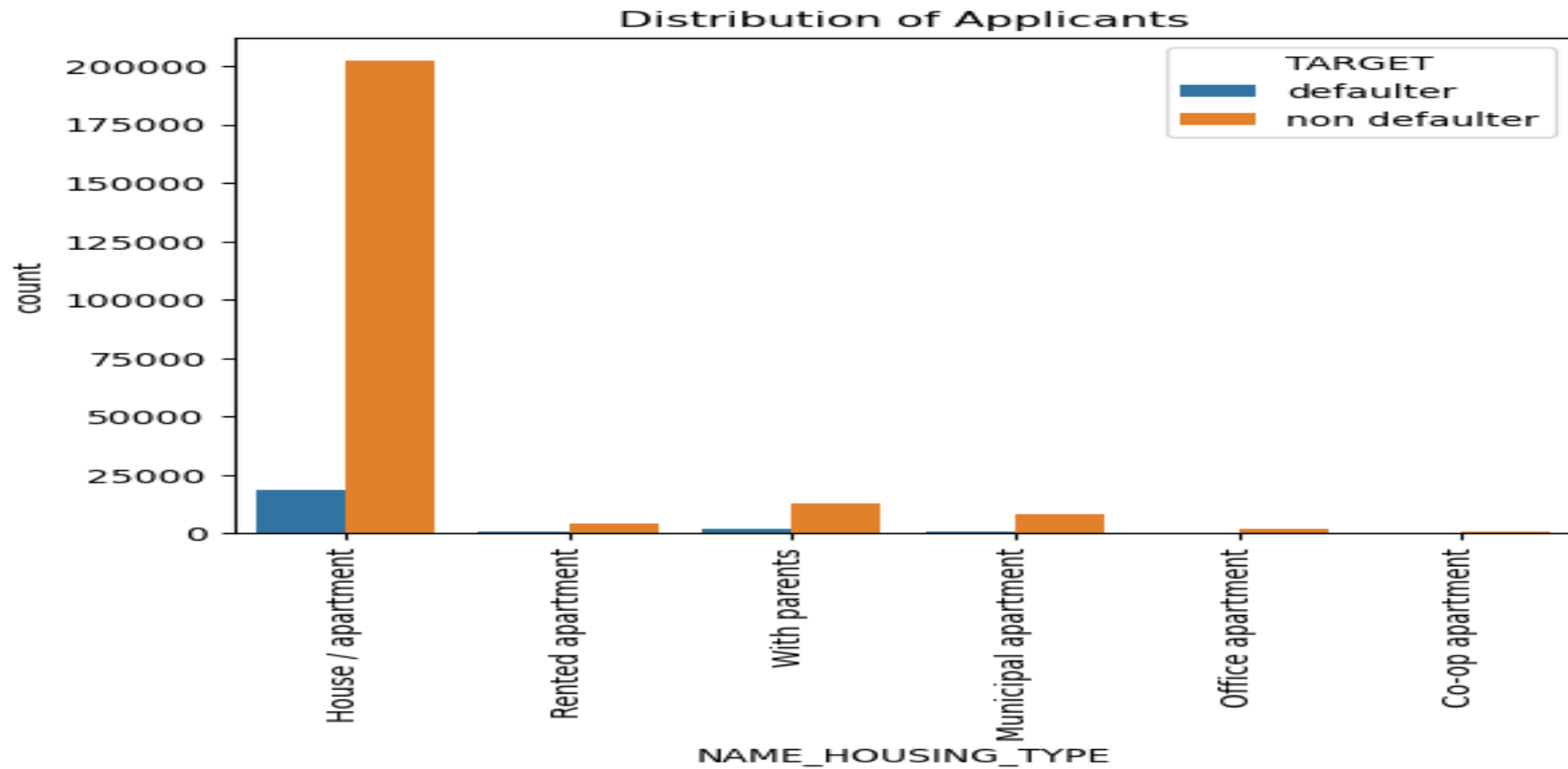
- ▶ 1) Number of females applicants are more than males
- ▶ 2) females from secondary education are max in the applicants

Proportion of Defaulter by family status

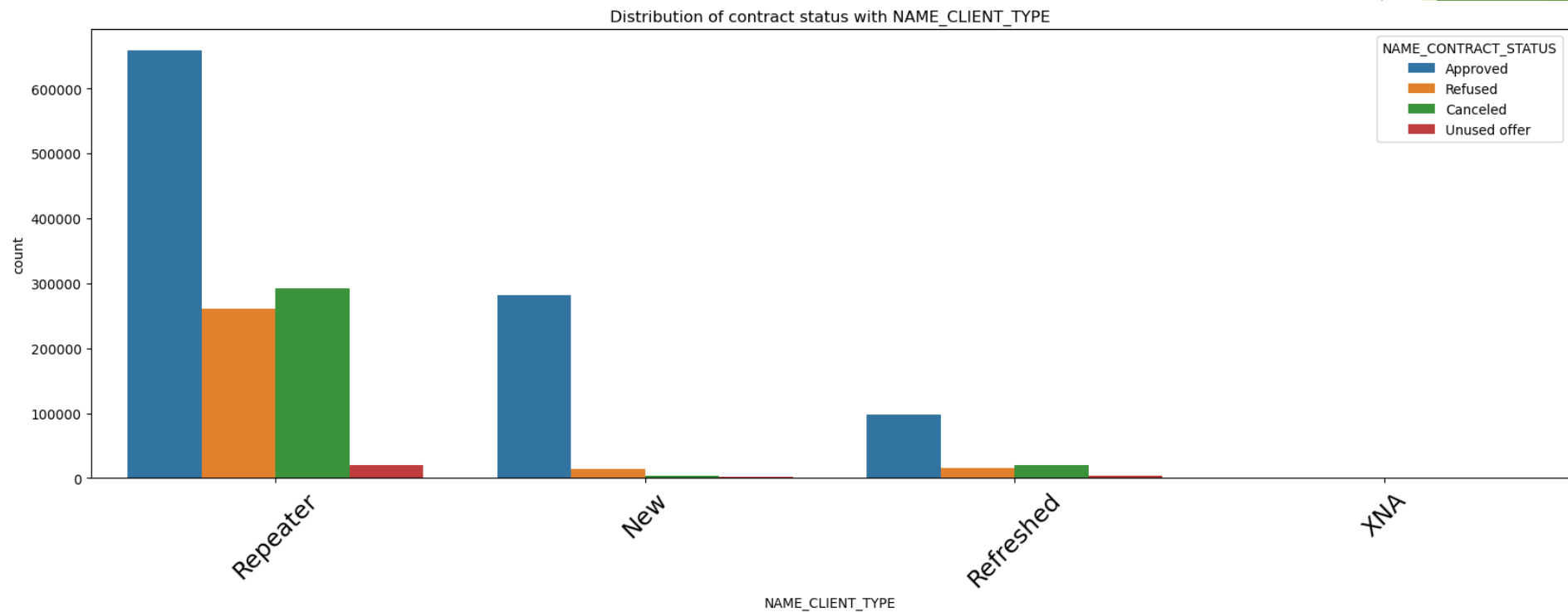


- 1) Married people are most frequent to apply for the loans

Proportion of Defaulter by Housing Type

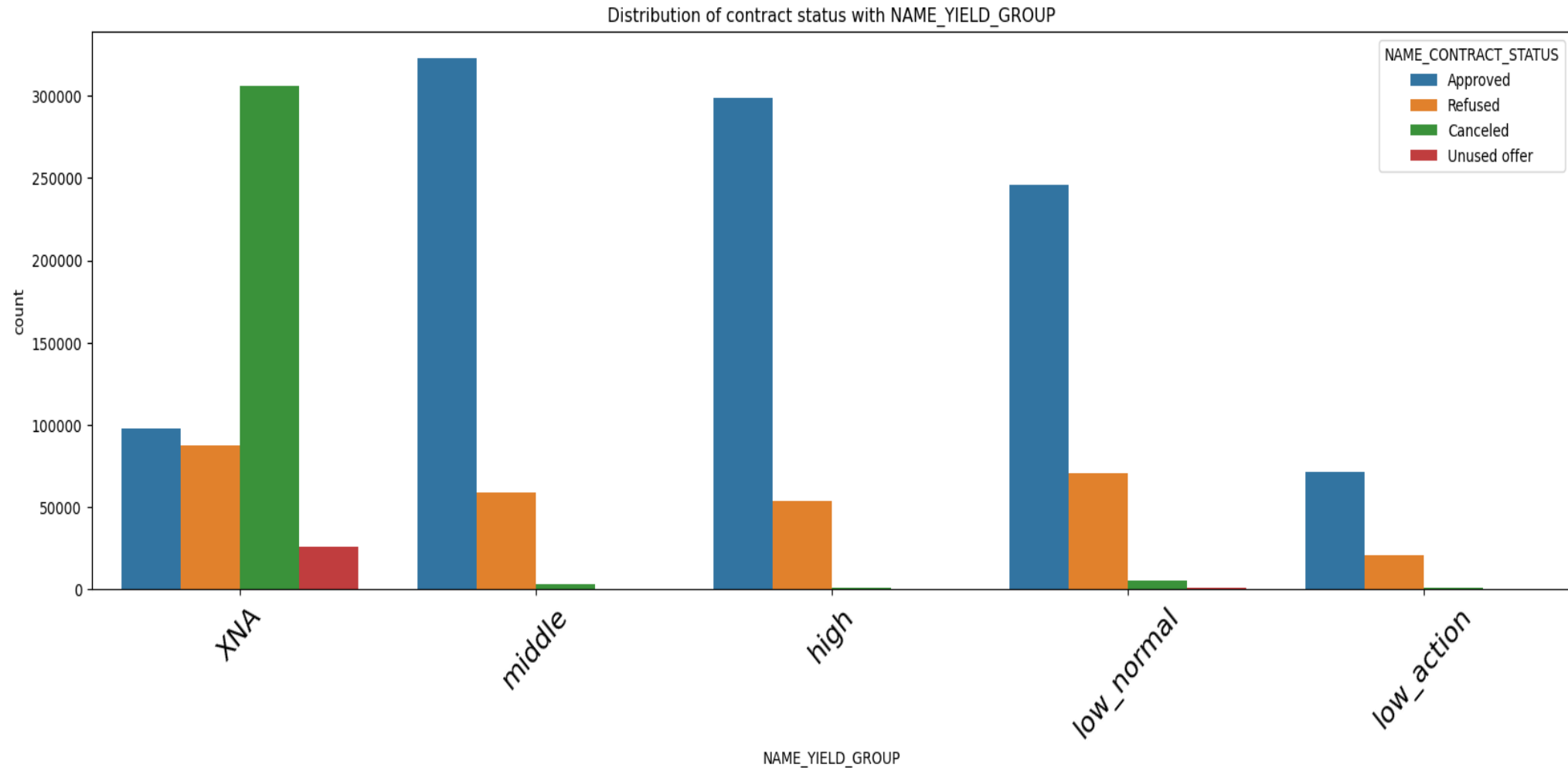


Distribution of contract status with NAME_CLIENT_TYPE



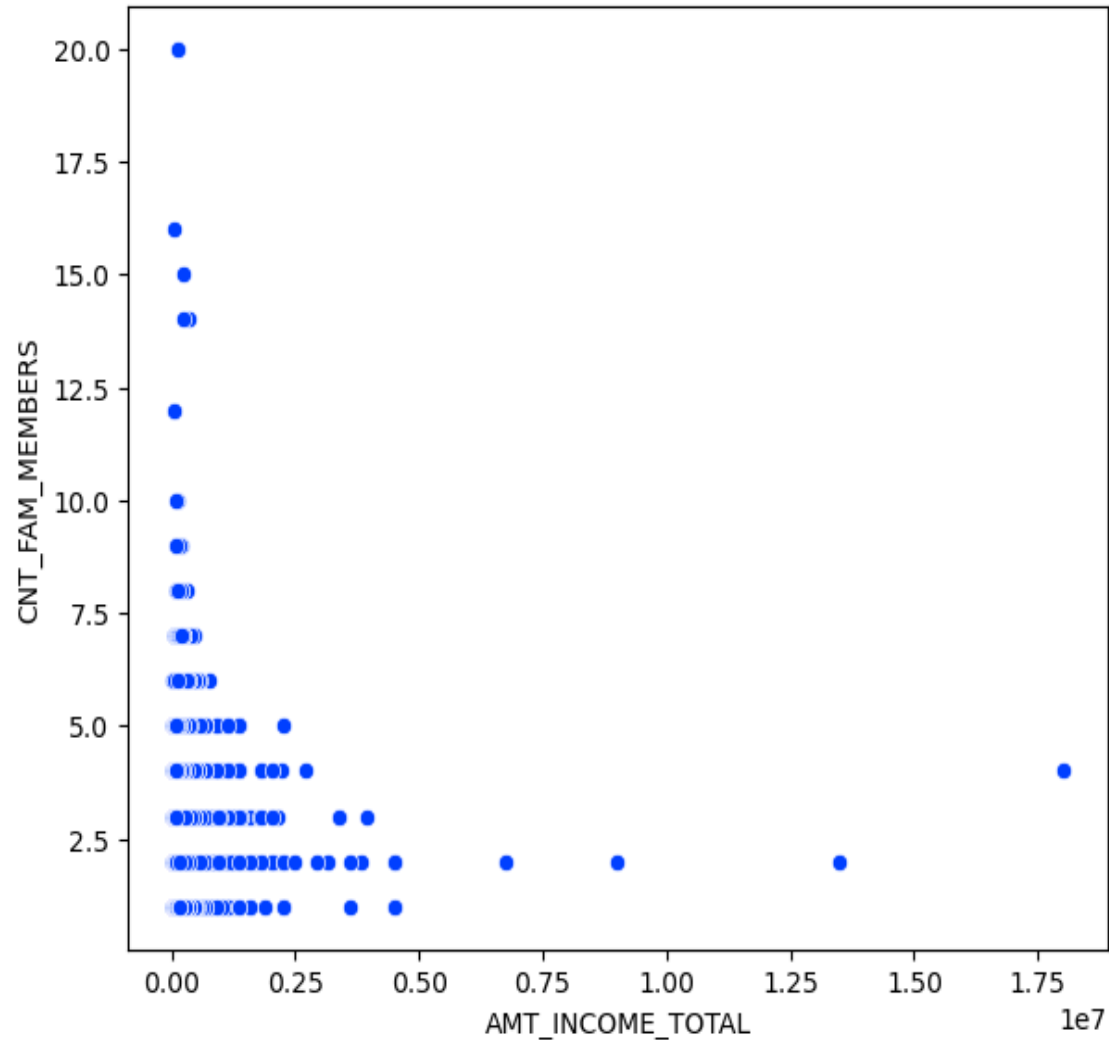
- ▶ 1) Lenders has approved the loans of repeater applicants mostly
- ▶ 2) Lenders has rejected only application of new client

Distribution of contract status with NAME_YIELD_GROUP

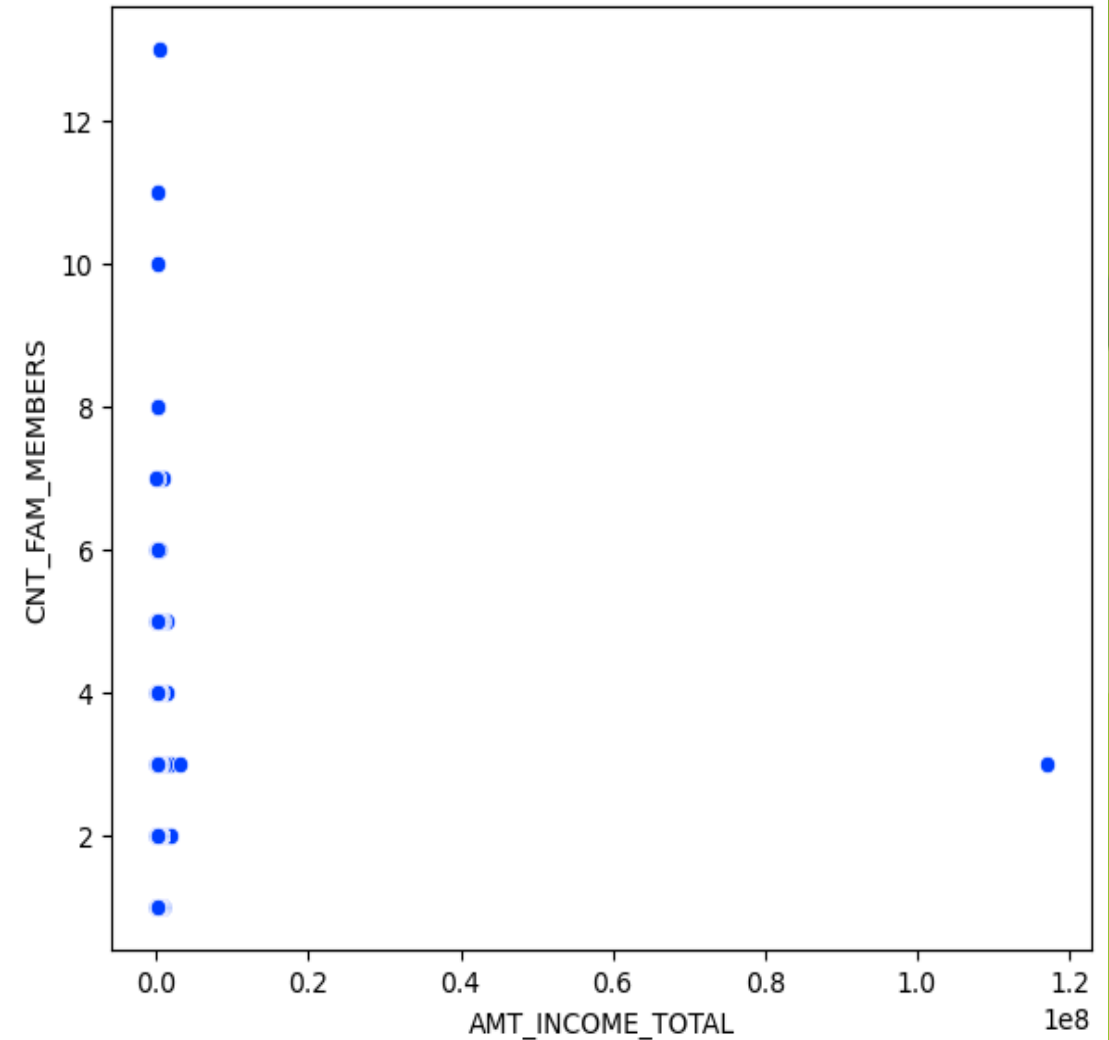


Correlation - Income and family members

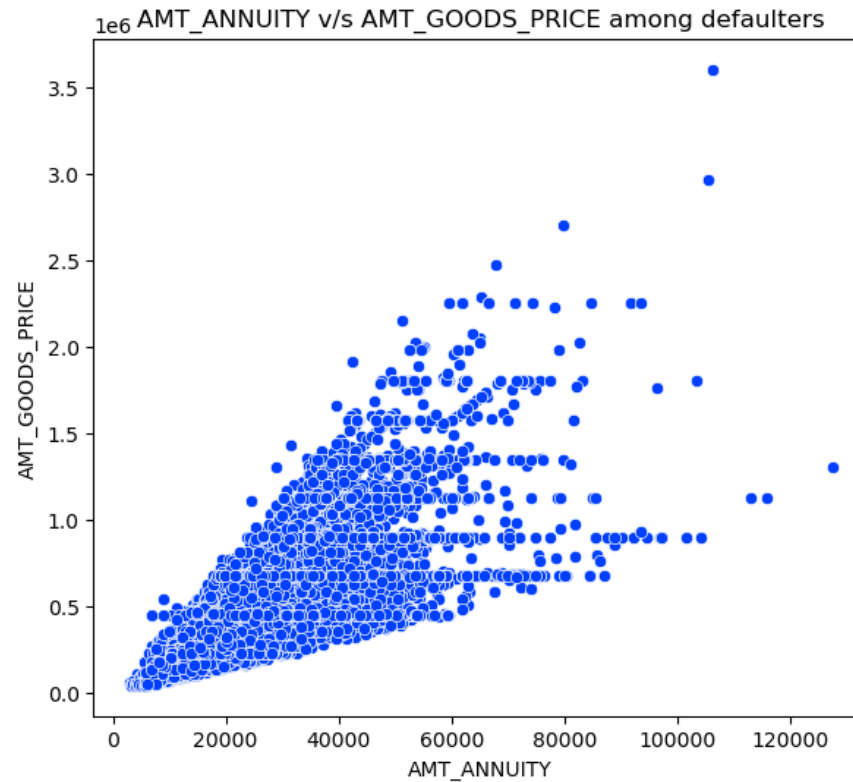
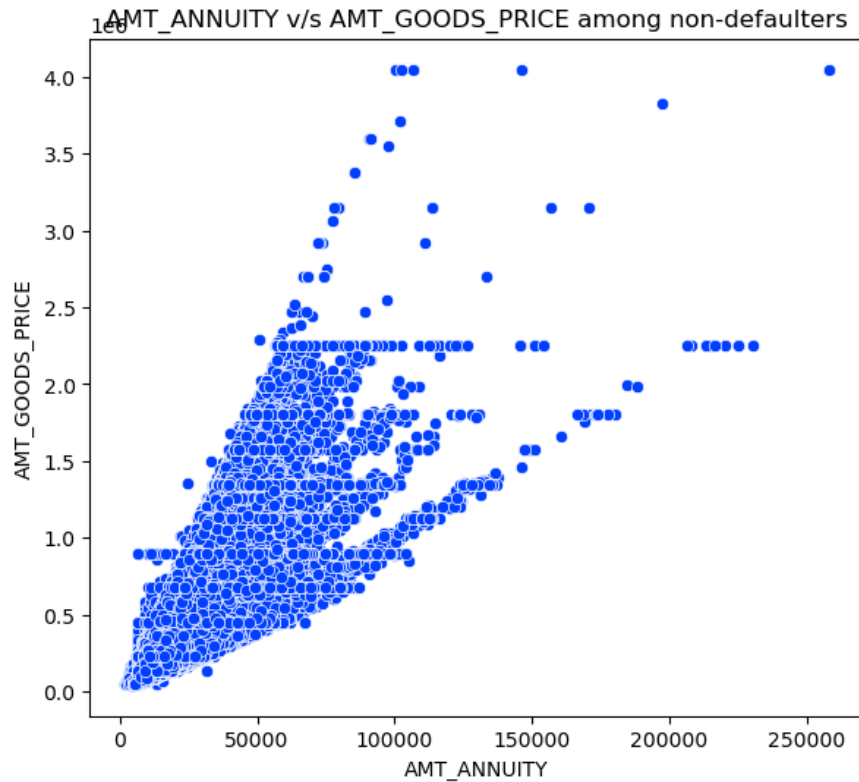
Income v/s family members among non-defaulters



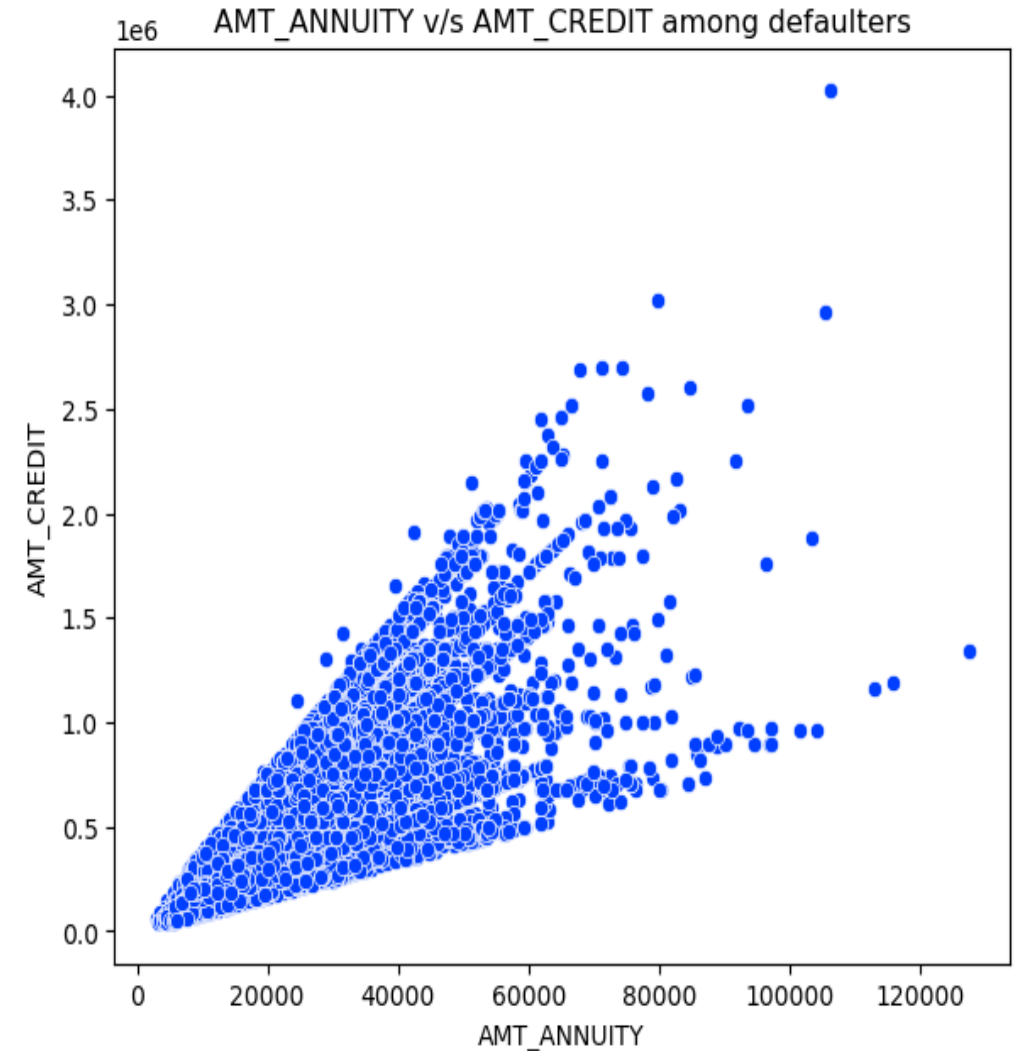
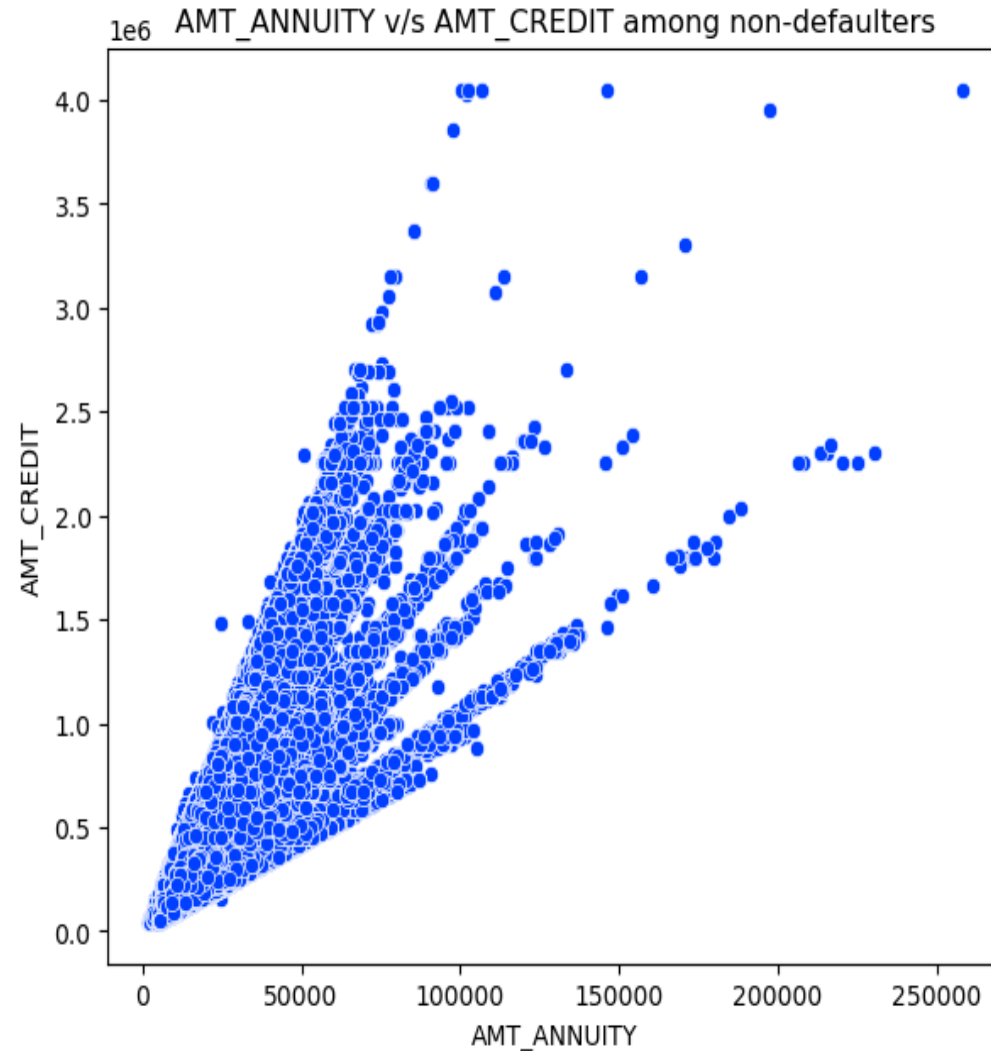
Income v/s family members among defaulters



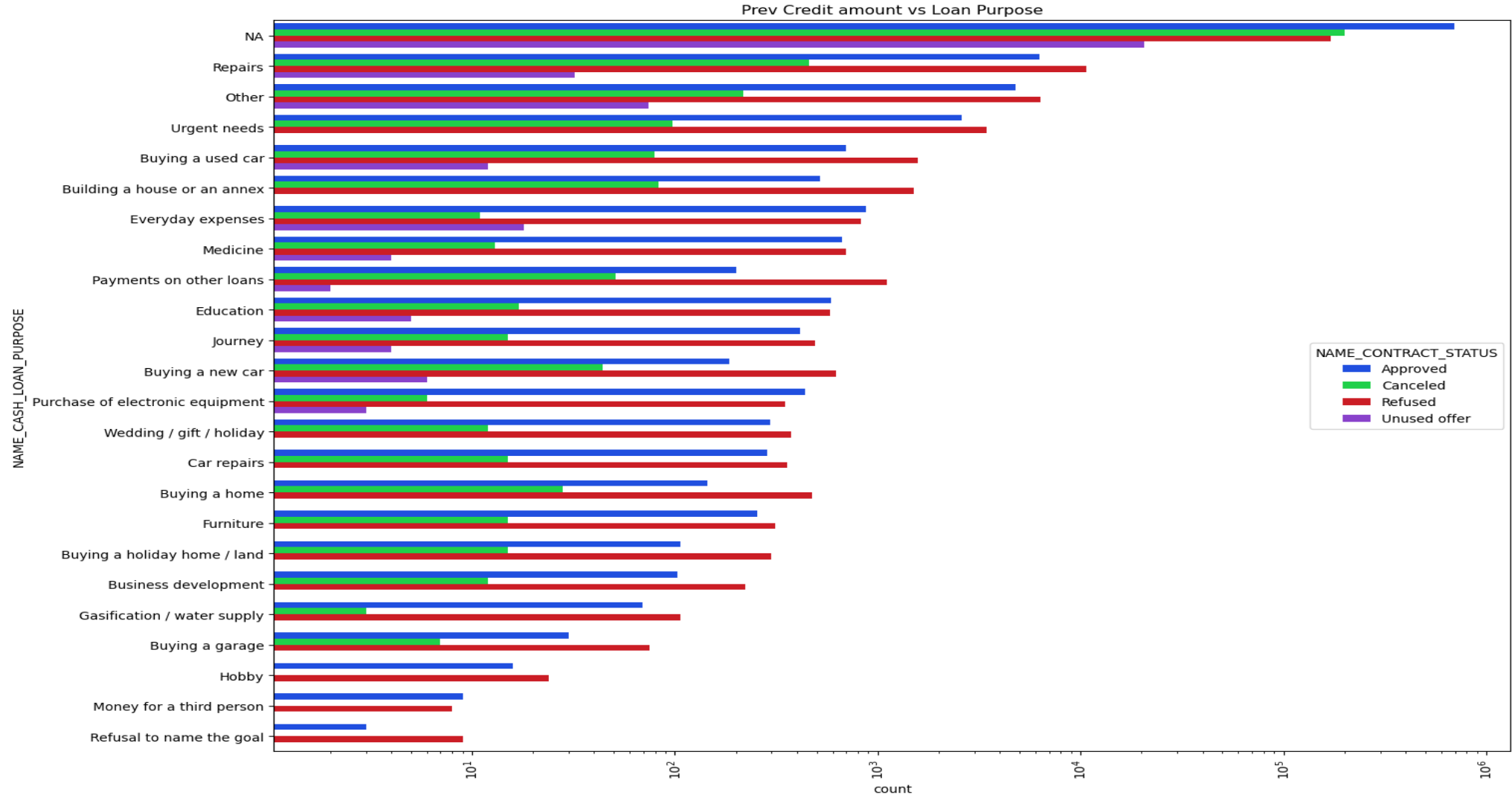
Correlation - AMT ANNUITY and AMT GOODS PRICE



Correlation - AMT ANNUITY and AMT CREDIT



Correlation - NAME CASH LOAN PURPOSE and NAME CONTRACT STATUS



Recommendation

- ❖ Considering rejection loans to the applicants with Academic degree.
- ❖ State servants have lowest level of defaulters, prefer loans to this type of applicants.
- ❖ More cash loans go into default so bank should give loans to revolving loans
- ❖ Single people default more than the married one.
- ❖ Loans previously refused or cancelled have higher default rate.
- ❖ Female applicants with higher education has less chance of defaults.
- ❖ Any applicants whose loans are approved previously has less chance of defaults