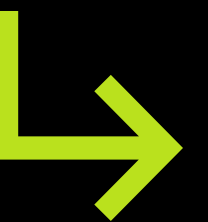


EDA Credit Assignment

Aastha Gautam



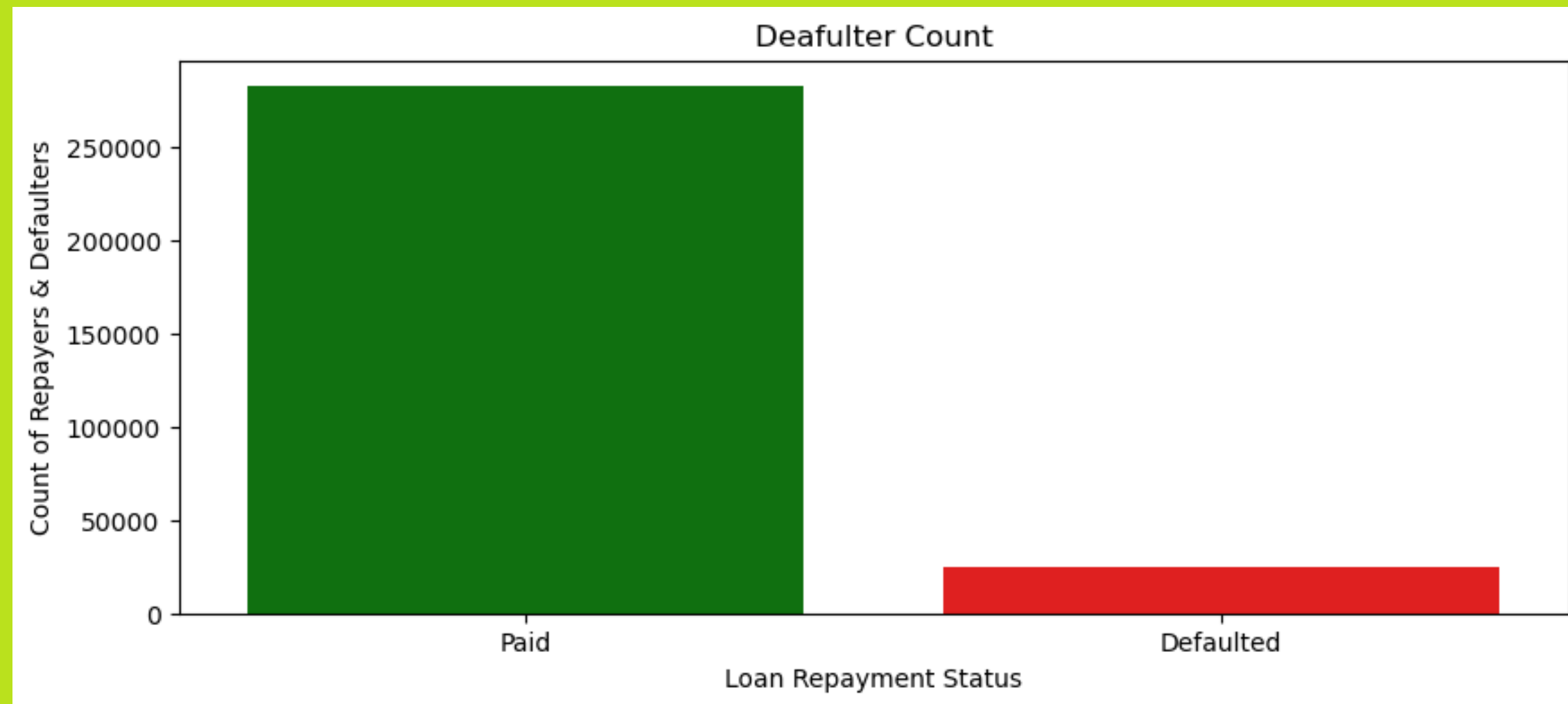
Problem Statement

Find the common pattern in the defaulters at the bank and help the bank take better decisions regarding selection of people being given the loan to.

Assumptions:

- Some of the columns with null/missing values have been imputed with the mode/median/mean depending on the outlier analysis done in the python notebook.
- Columns with more than 40% null values have been ignored for the analysis as they will not be able to give a conclusive result because of the missing data.
- Correlation matrix has been made in the beginning to understand the correlation of some of the columns with the "TARGET" variable. On the basis of the result, the columns have been ignored yet again for the analysis.

IMBALANCE



Ratios of imbalance in percentage with respect to Repayer and Defaulter datas are: 91.93 and 8.07
Ratios of imbalance in relative with respect to Repayer and Defaulter datas is 11.39 : 1 (approx)

Binning Insights

20% of the loans were taken for goods ranging between Rs. 2 lakh and Rs. 3 lakh.

About 24% of the credits given ranged between Rs. 1.5 lakh and Rs. 3 lakh.

About 24% of the annuity amount recieved were more than Rs. 35000

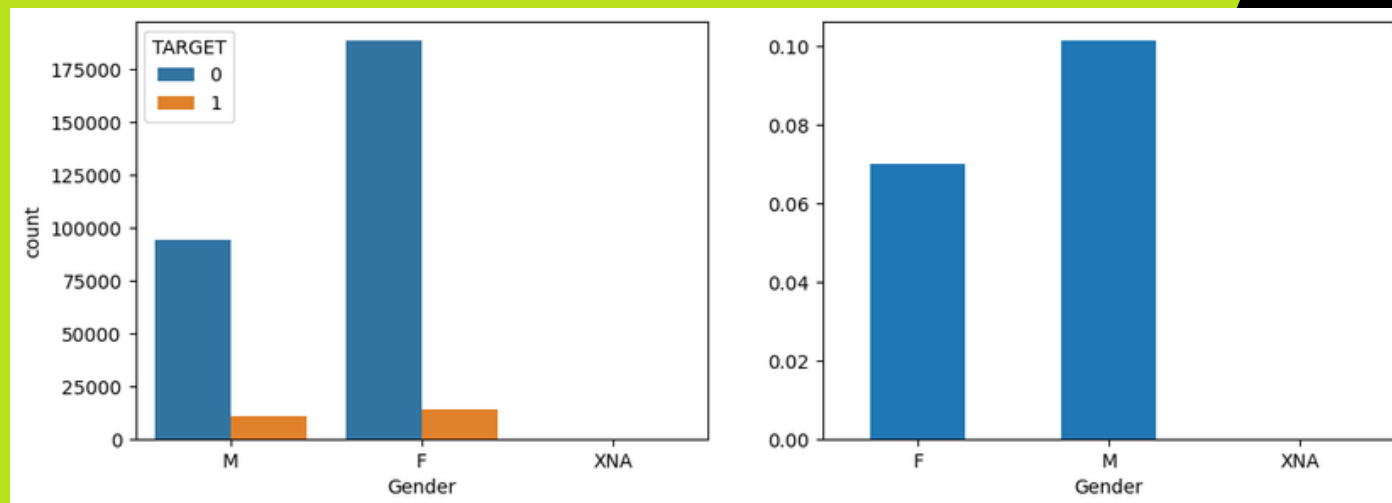
About 30% of the loans taken were taken by people with total income ranging between Rs. 1 Lakh and Rs. 1.5 Lakh

About 25% of the clients have been employed for more than 16 years. However, an entry of 365243 days (>100 years) is an impossible event and is not useful for the analysis.

CATEGORICAL

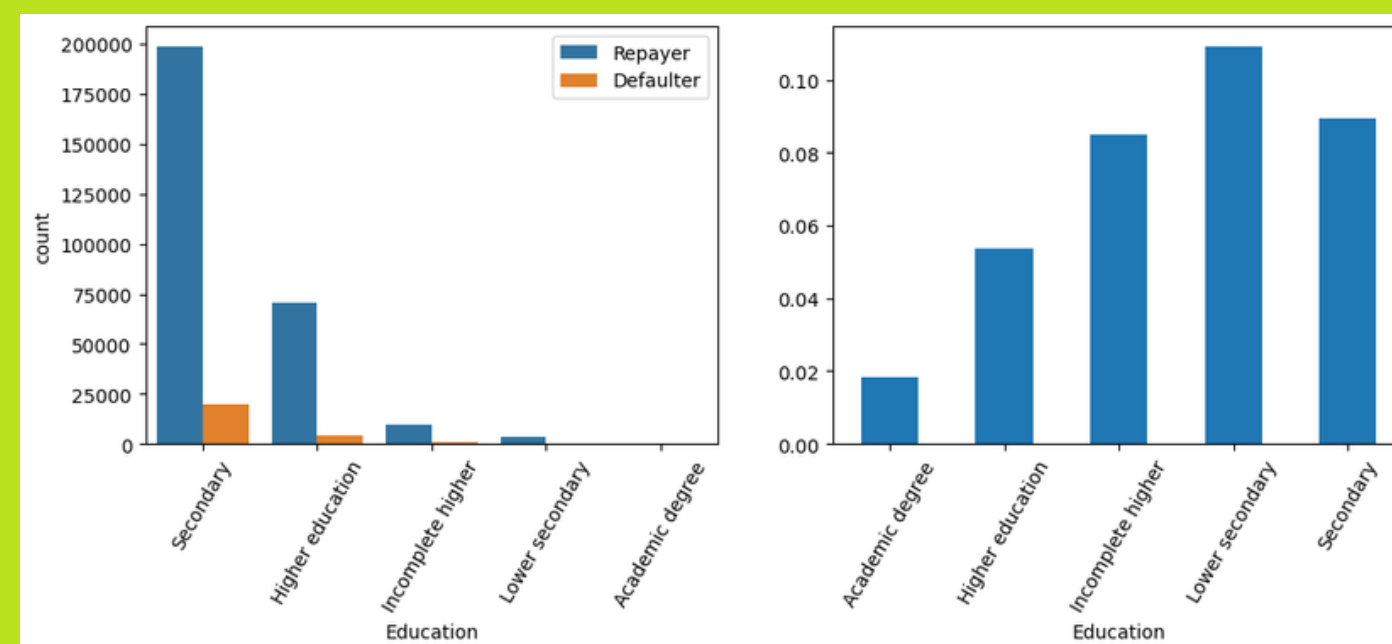
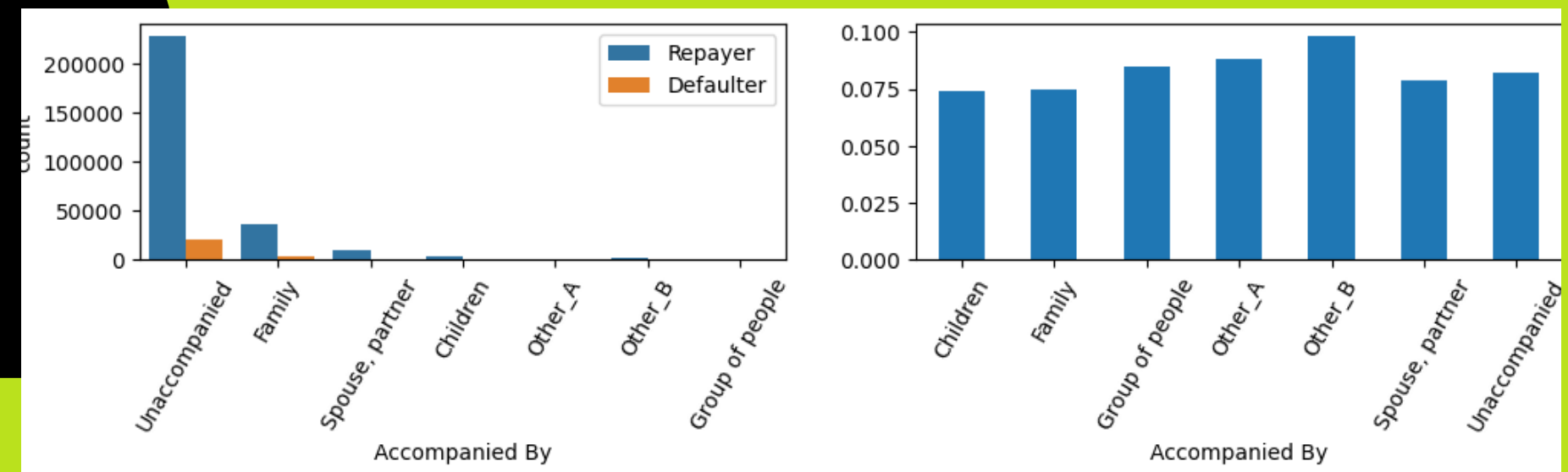
UNIVARIATE

ANALYSIS

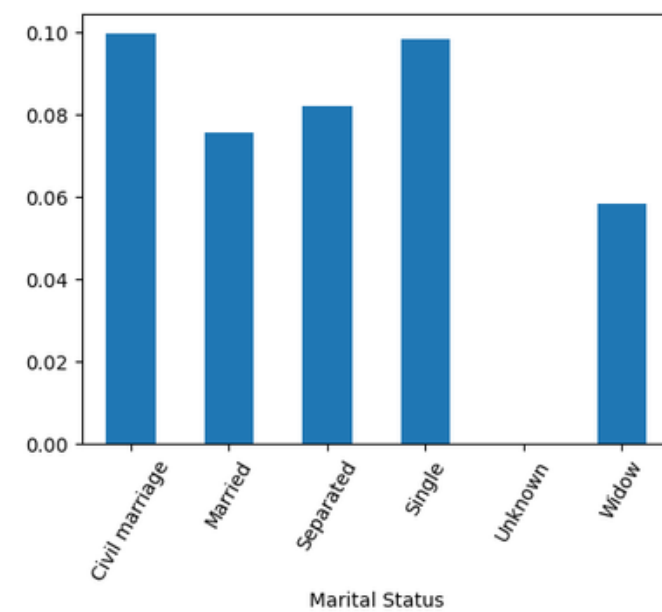
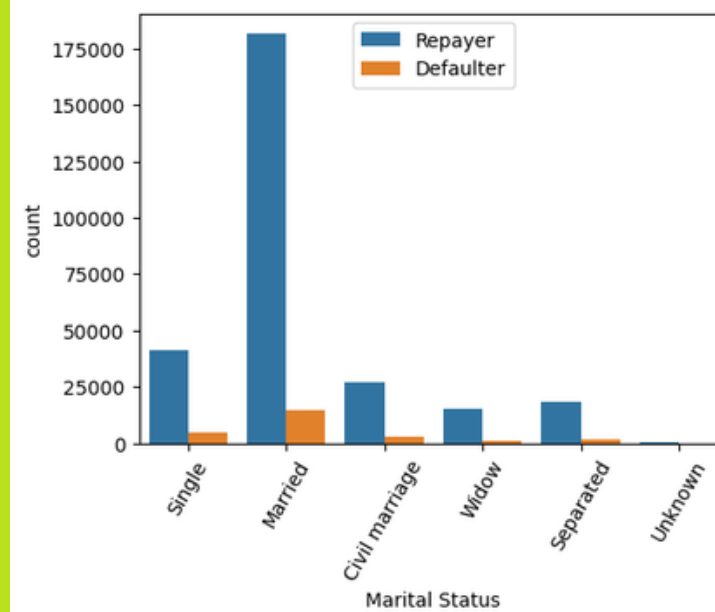


Even though females have taken more loans, the males have defaulted more in paying back the loans.

People who were unaccompanied while taking loans have taken the most number of loans and, of them, only 8% (approx) have defaulted.

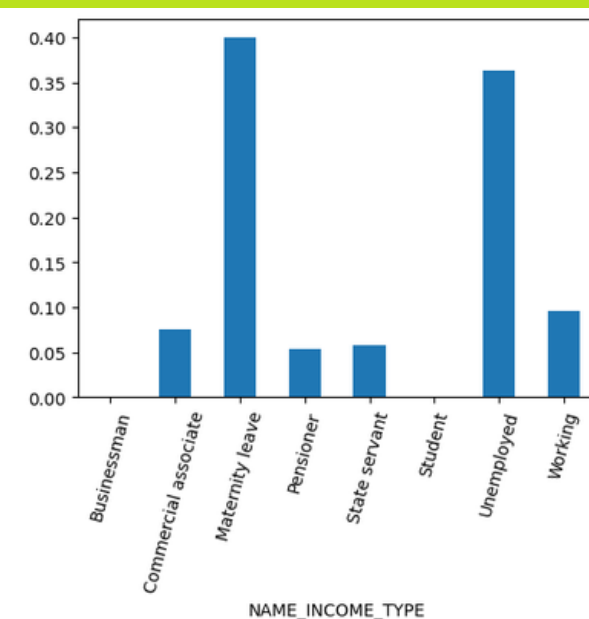
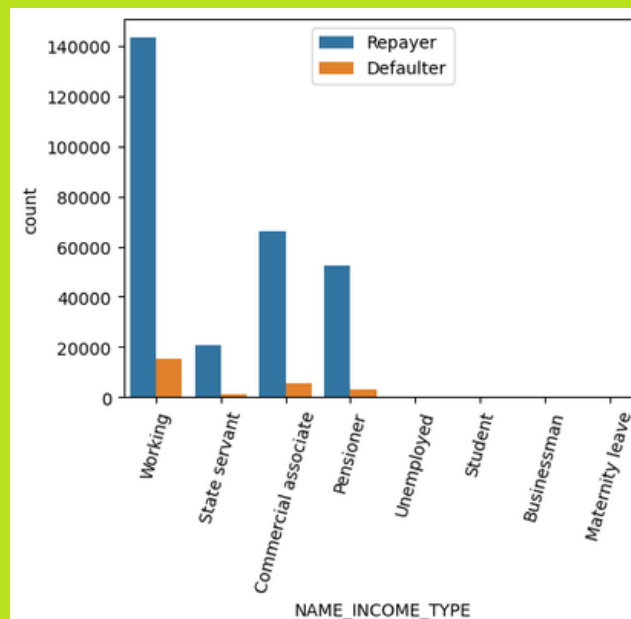
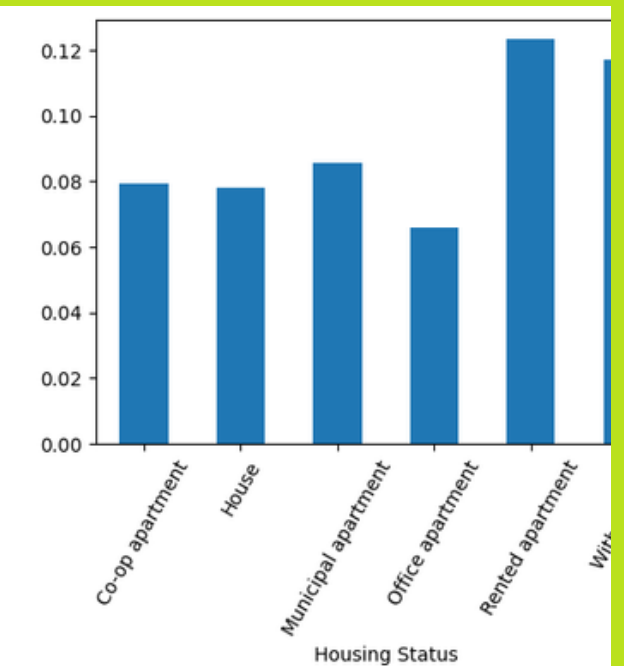
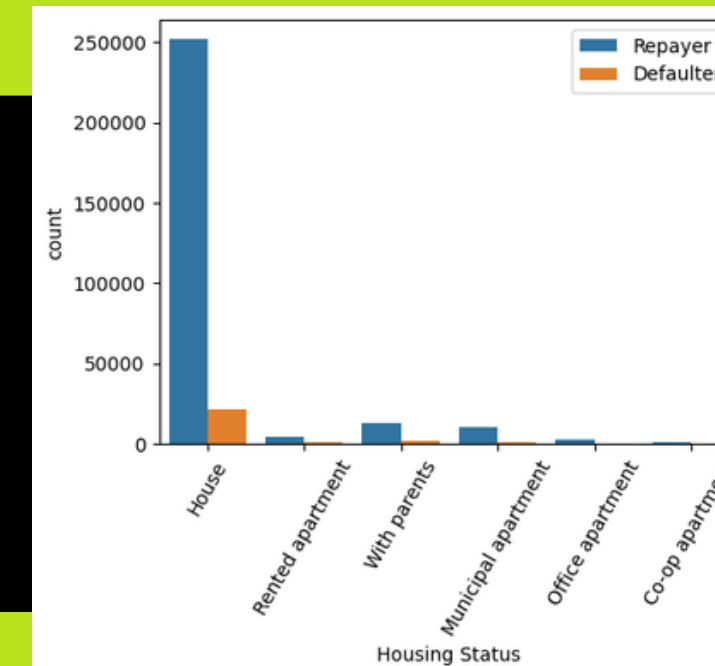


Major clients have a Secondary education degree, and, the Lower Secondary Category, even though is low in number, has defaulted the most during repayment.



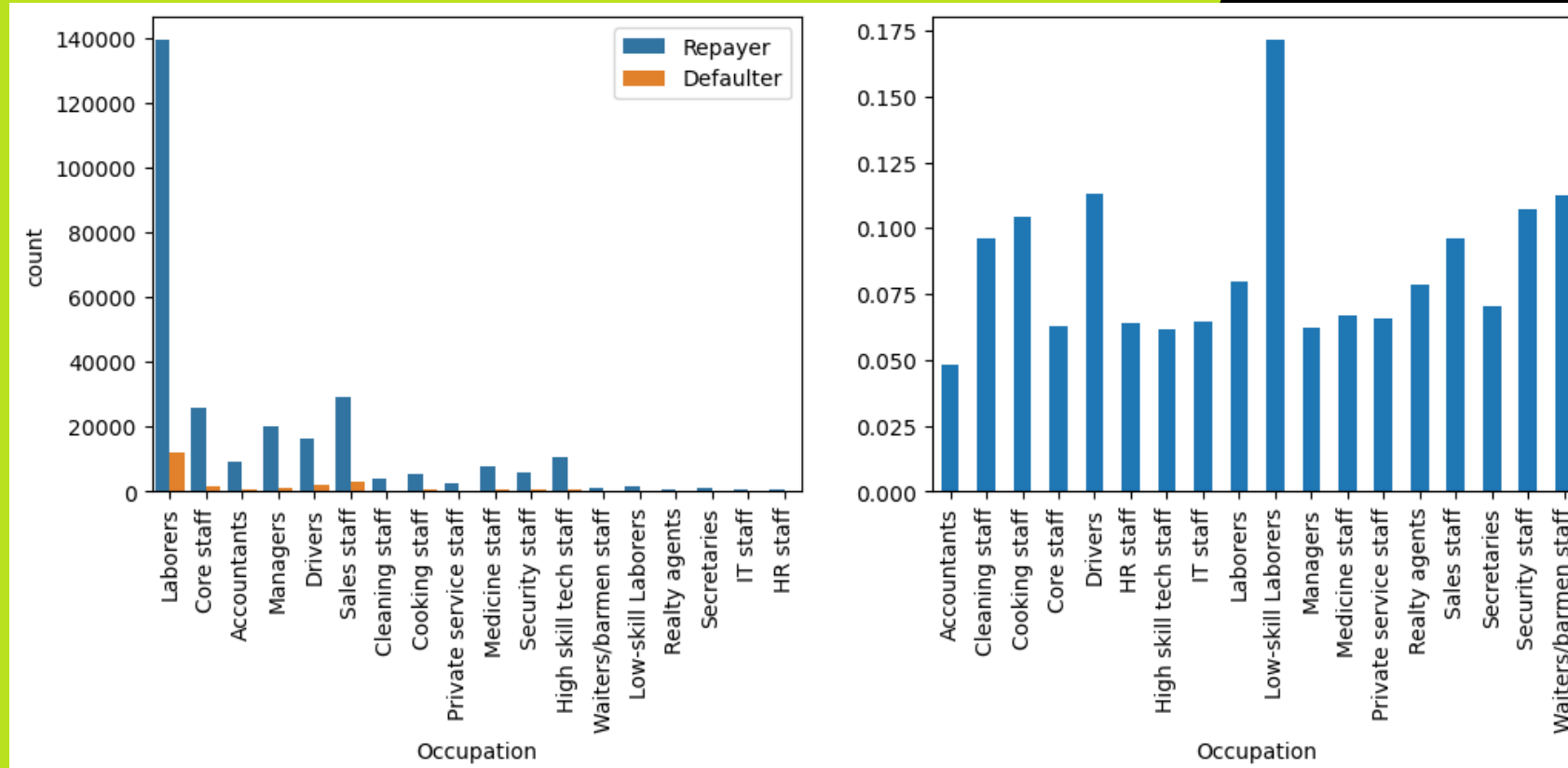
Married couples have taken the most loans followed by Singles. Civil marriage couples and Singles(~10%) have defaulted the most while repayment.

Clients with their own real estate have taken the most loans. People living in rented apartments (~12%) or with their parents have defaulted the most.

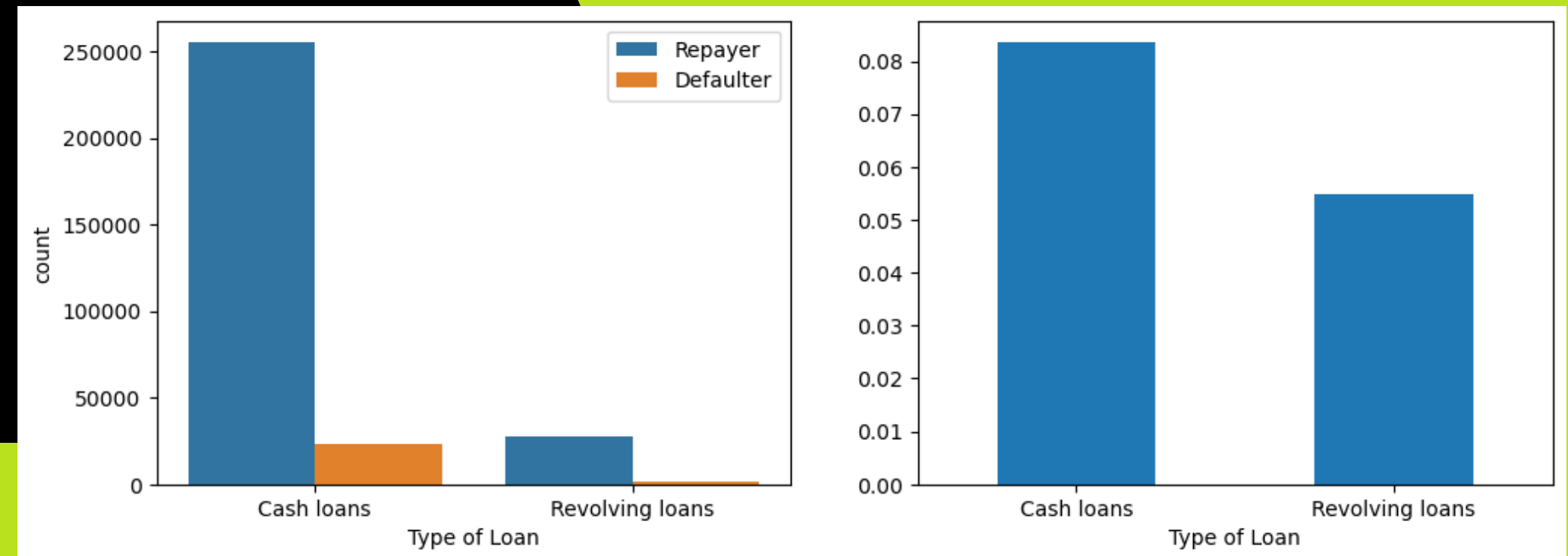


It is the working class which takes the most loans, while it is the people on maternity leave (~40%) who default the most while repayment. *Students and Businessmen (though less in number) are the safest when it comes to giving a loan.*

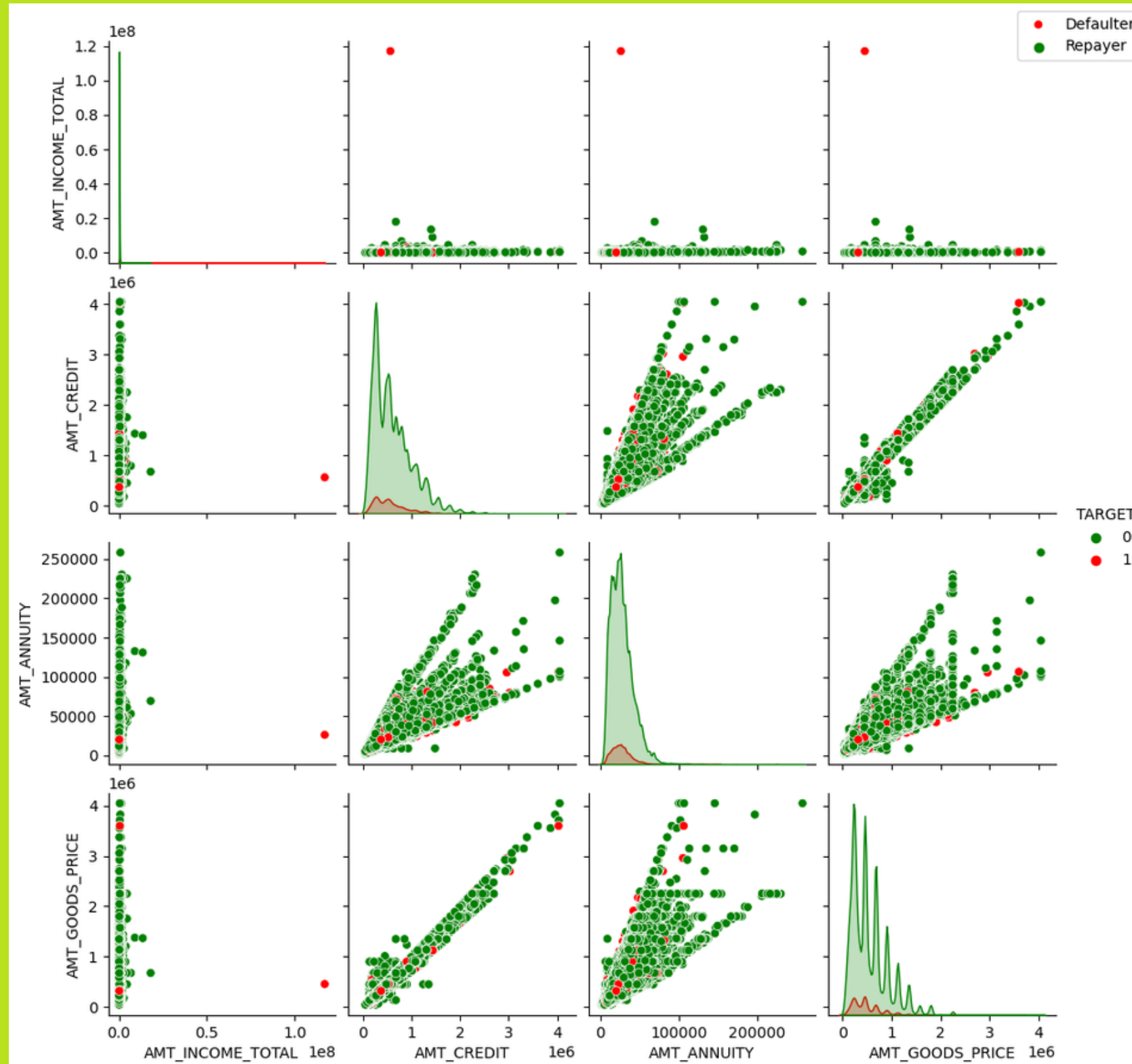
Most of the loans are taken by Laborers, followed by Sales staff. IT staff take the lowest amount of loans. The group with highest percent of not repaid loans are Low-skill Laborers (above 17%), followed by Drivers and Waiters/barmen staff.



Only 10% of all loans are revolving loans, however when compared to how frequently they are repaid, a higher percentage of revolving loans are not paid back.

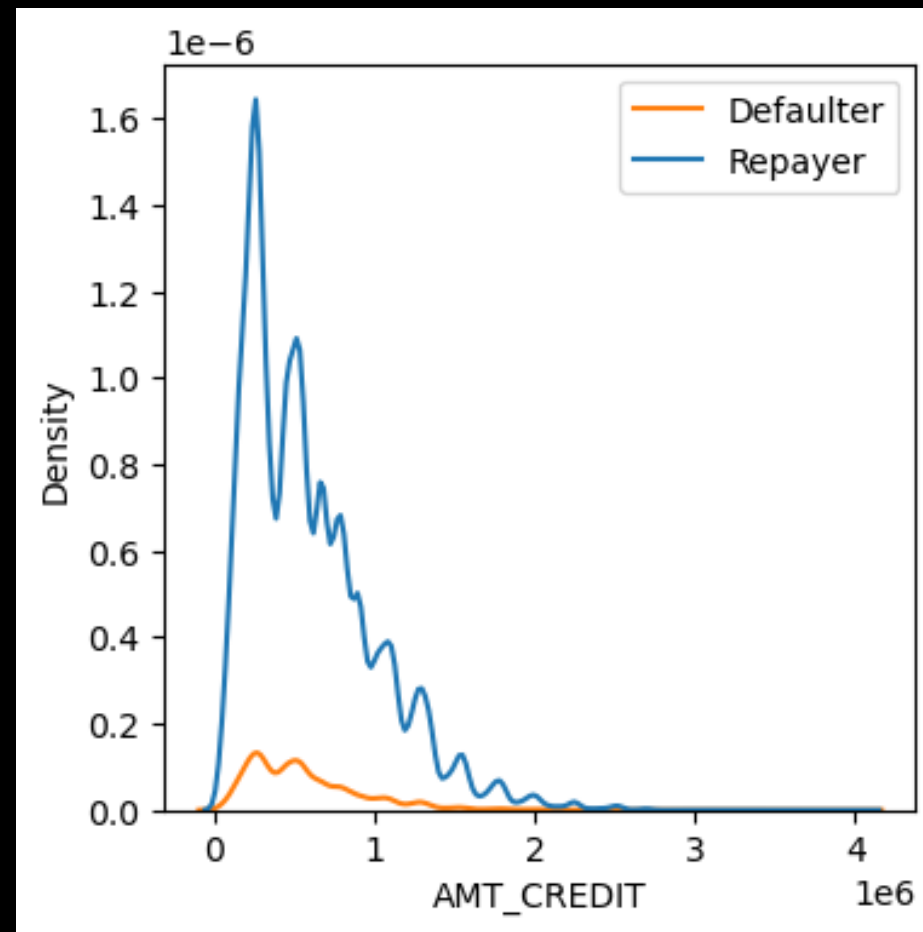


BI-VARIATE ANALYSIS



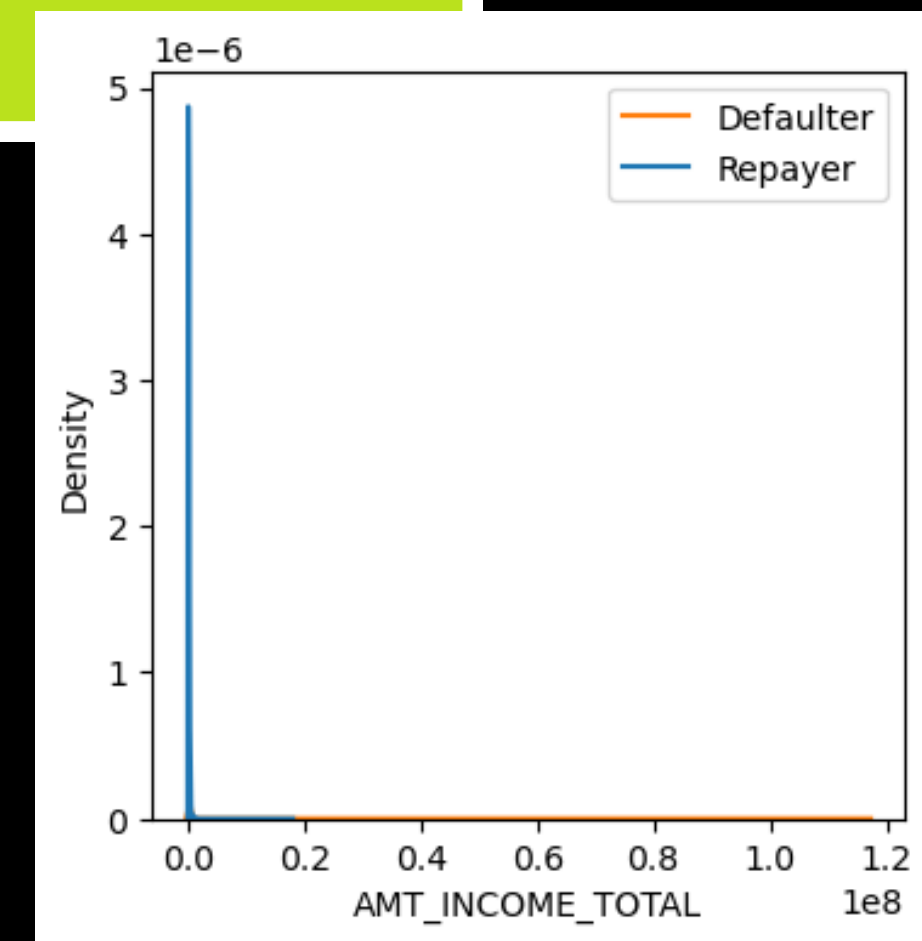
- **AMT_CREDIT** and **AMT_GOODS_PRICE** are highly associated as based on the scatterplot where most of the data are aggregated in form of a line.
- When **AMT_ANNUITY** > 15000 and **AMT_GOODS_PRICE** > 30 Lakhs, there is a decreased possibility of defaulters.
- For **AMT_CREDIT** > 30 Lakhs defaulters are quite rare.

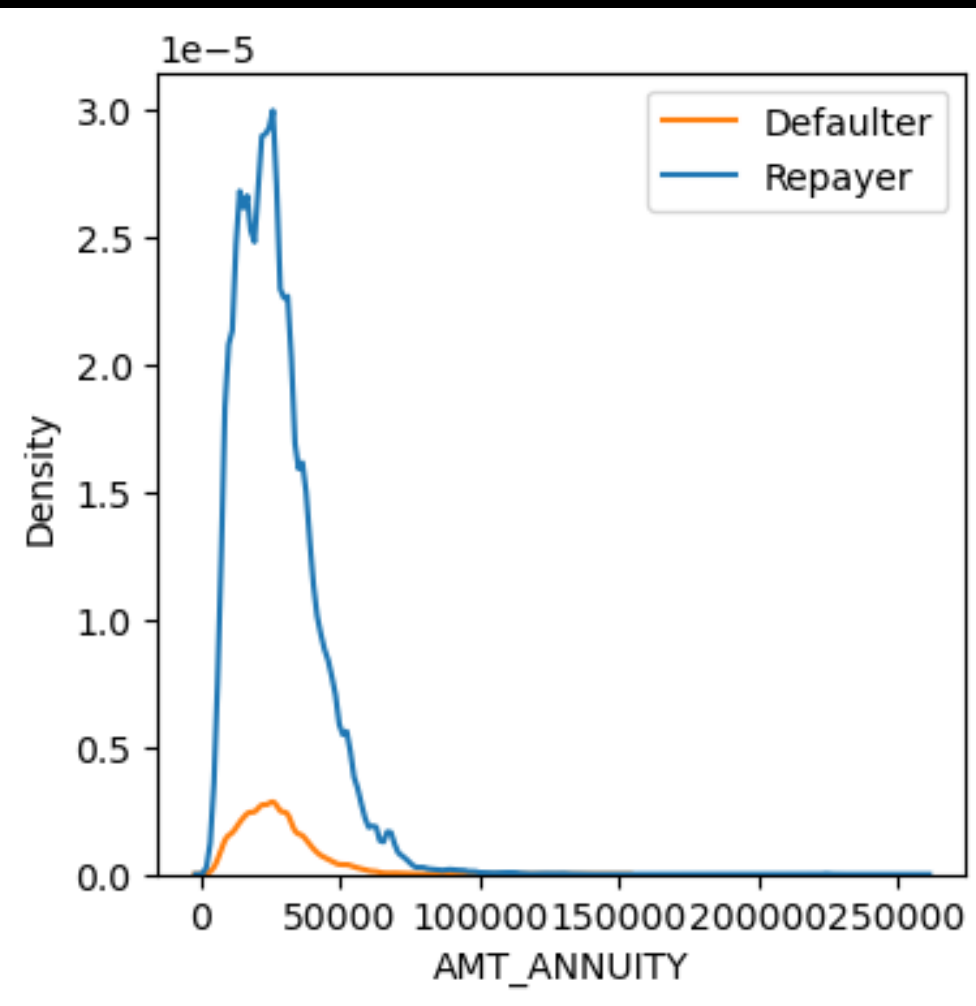
NUMERICAL UNIVARIATE ANALYSIS



**Most of the loans
were given for the
credit
amount ranging
from
0 to 10 lakhs.**

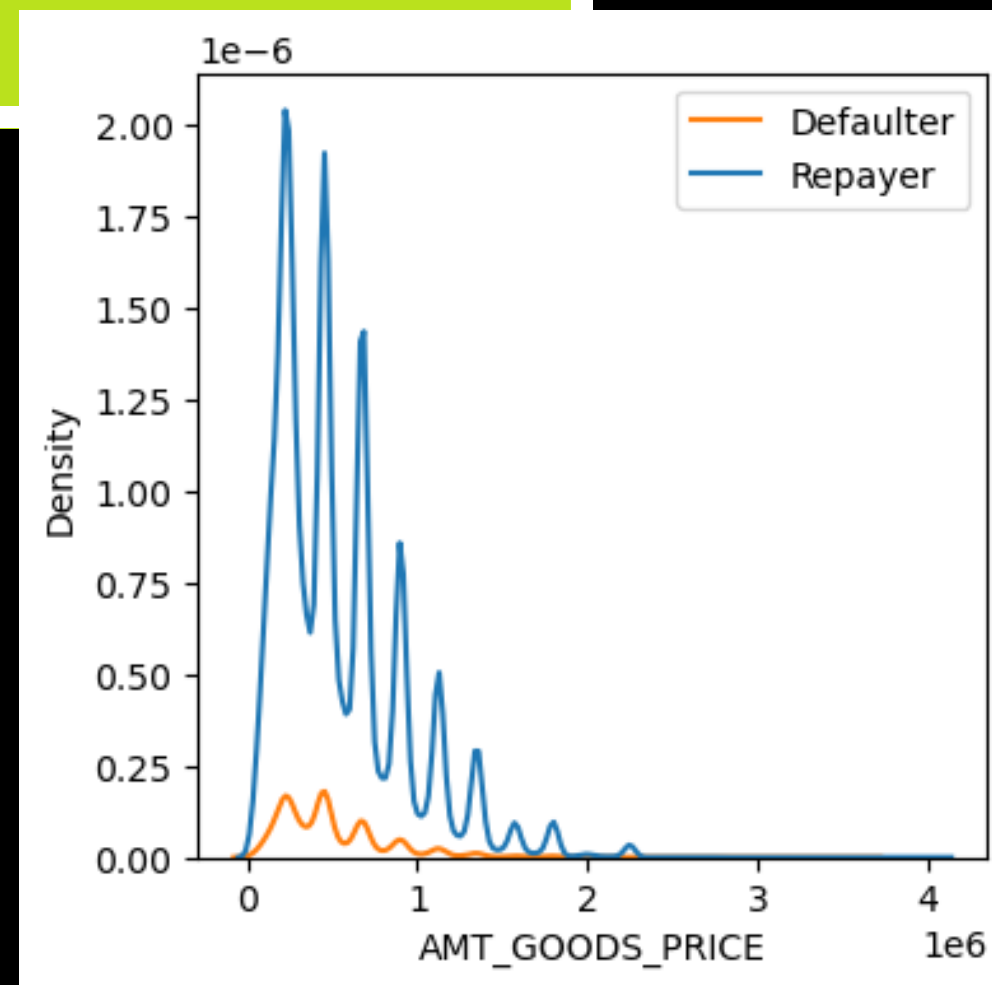
**Most of the loans
applications were
given by people
earning from
0 to Rs. 10 Lakhs**



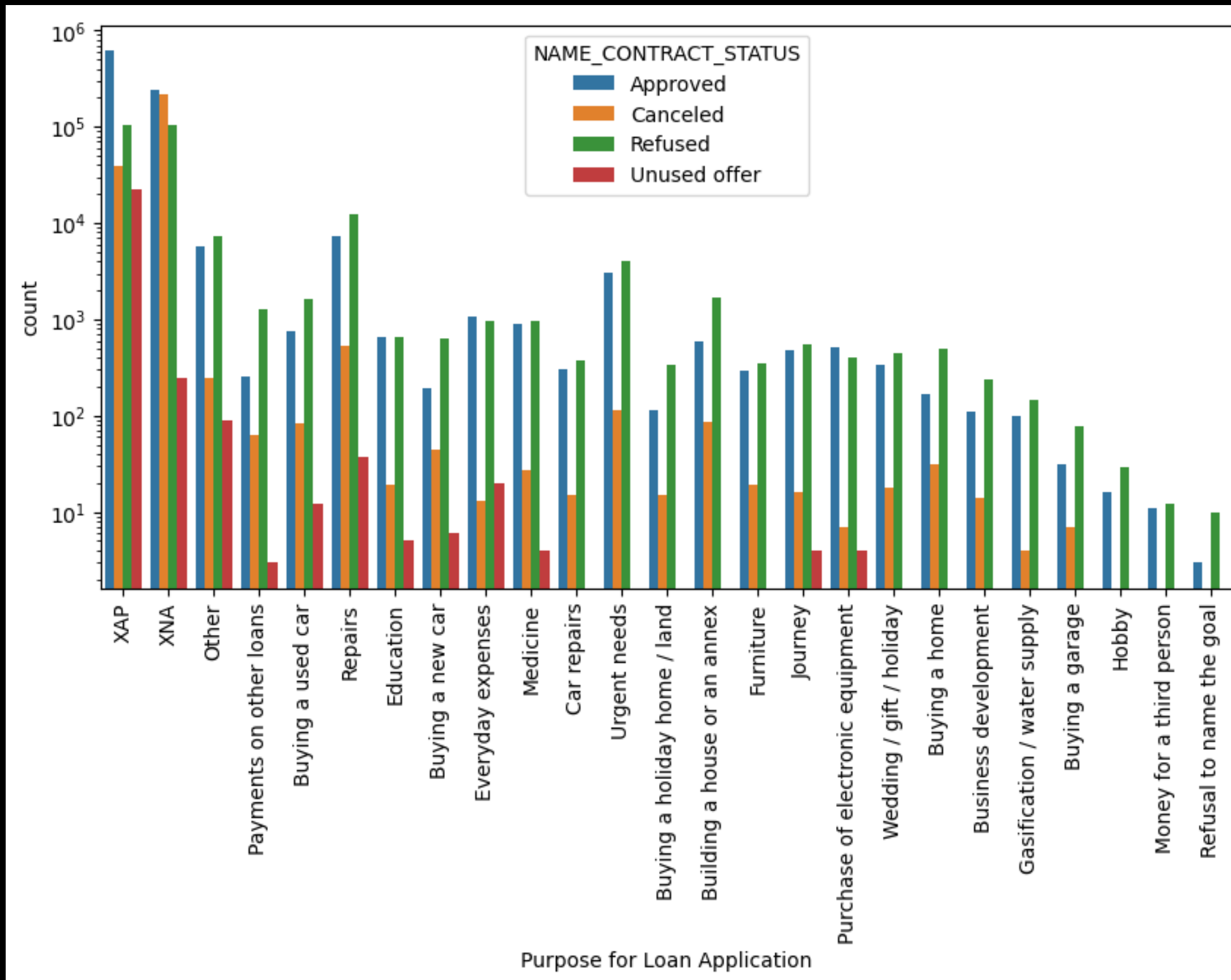


Most of the customers are paying at an annuity from 0 to Rs. 50,000.

Most of the loans were taken for goods ranging upto Rs. 10 Lakh.

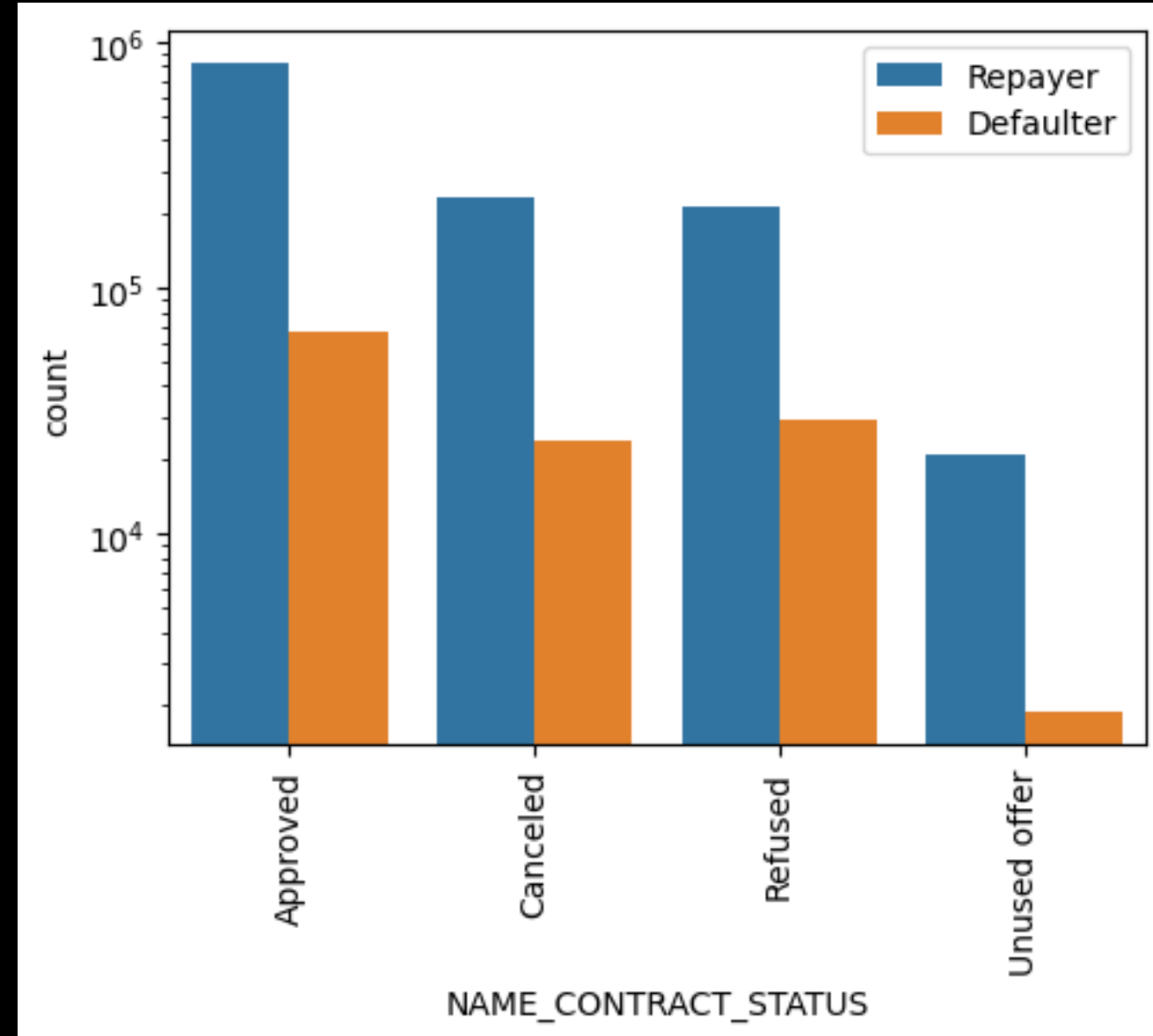


MERGED DATA ANALYSIS

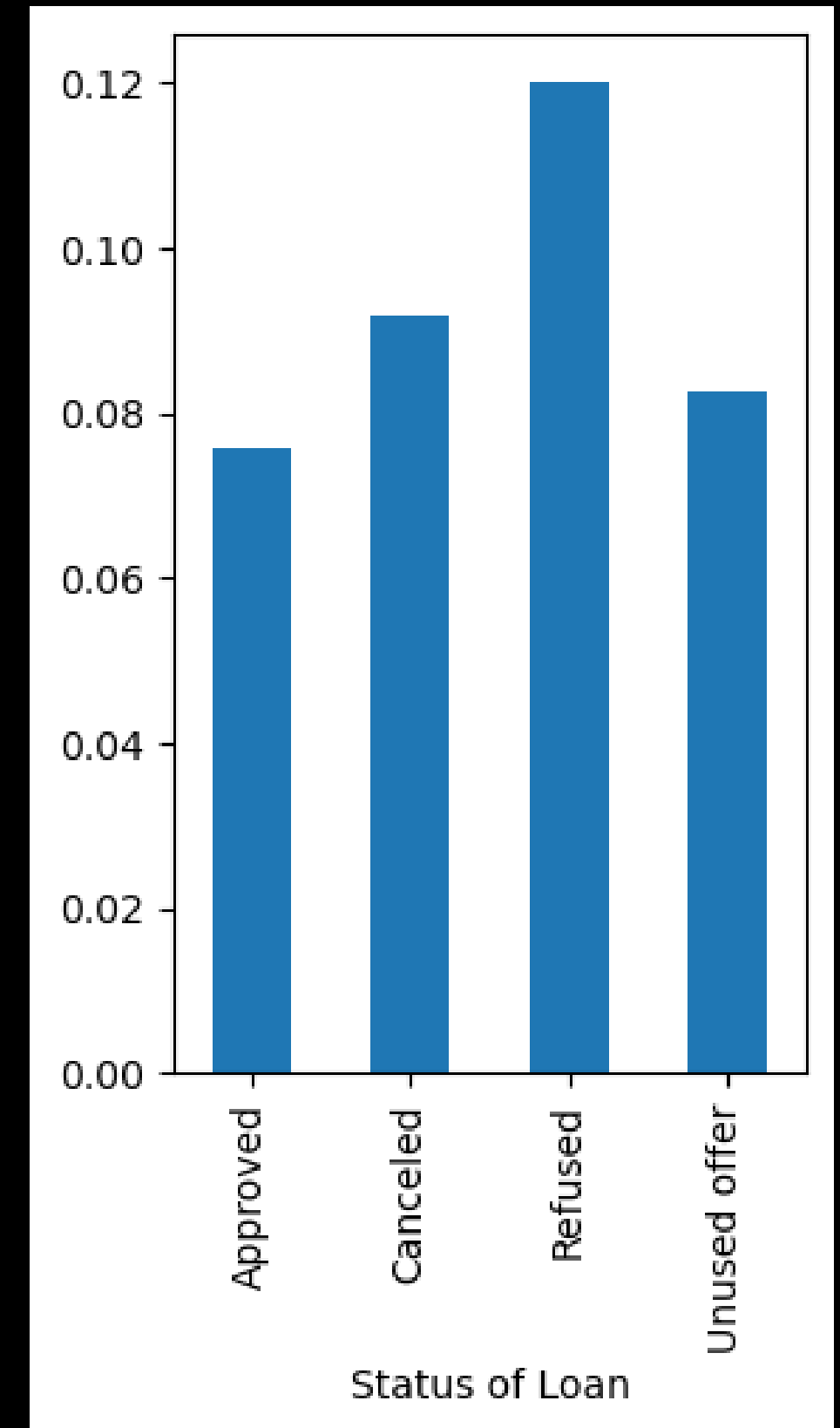


Loans taken out for repairs appear to have the greatest default rates.

Many applications with purposes like repairs or other have been turned down by banks or clients. This demonstrates that banks view lending for purpose repair as high risk and either reject such loans or offer them at extremely high interest rates that the clients cannot afford.



- About 90% of the previously cancelled client have actually repayed the loan. Revisiting the interest rates would increase business opportunity for these clients.
- About 88% of the clients who have been previously refused a loan have payed back the loan in current case.



Recommendations

- Bank should give focus on providing the cash loan than revolving loan as cash loan are less likely to default.
- Female loan has less default rate. So, the bank should give a little bit priority to females.
- Those clients who do not have any accompany should be focused.
- Safest segmentation of employment are workers, commercial associates and pensioners.
- Client who have the higher education should be given more loans.
- Married clients are safer than unmarried.
- People having house/apartment are safer to provide loans.
- Low-skill laborers and drivers should be given less priority as they have high probability of making defaults.

- **People having income less than 1 million and taking loans near to 1 million have higher chance of defaults. So, should not be given focus.**
- **Client having annuity less than 100K are safer side for the bank.**
- **80-90% of customers who were previously cancelled/ refused are repayers. So. banks have to reverify the those applications.**

TO SUMMARISE THE RECOMMENDATIONS:

- **NAME_EDUCATION_TYPE:** Academic degree has less defaults.
 - **NAME_INCOME_TYPE:** Student and Businessmen have no defaults.
 - **AMT_INCOME_TOTAL:** Applicant with Income more than 700,000 are less likely to default
 - **NAME_CASH_LOAN_PURPOSE:** Loans bought for Hobby, Buying garage are being repayed mostly.
-
- **CODE_GENDER:** Men are at relatively higher default rate
 - **NAME_FAMILY_STATUS :** People who have civil marriage or who are single default a lot.
 - **OCCUPATION_TYPE:** Avoid Low-skill Laborers, Drivers and Waiters/barmen staff, Security staff, Laborers and Cooking staff as the default rate is huge.
 - **AMT_GOODS_PRICE:** When the credit amount goes beyond 3M, there is an increase in defaulters.