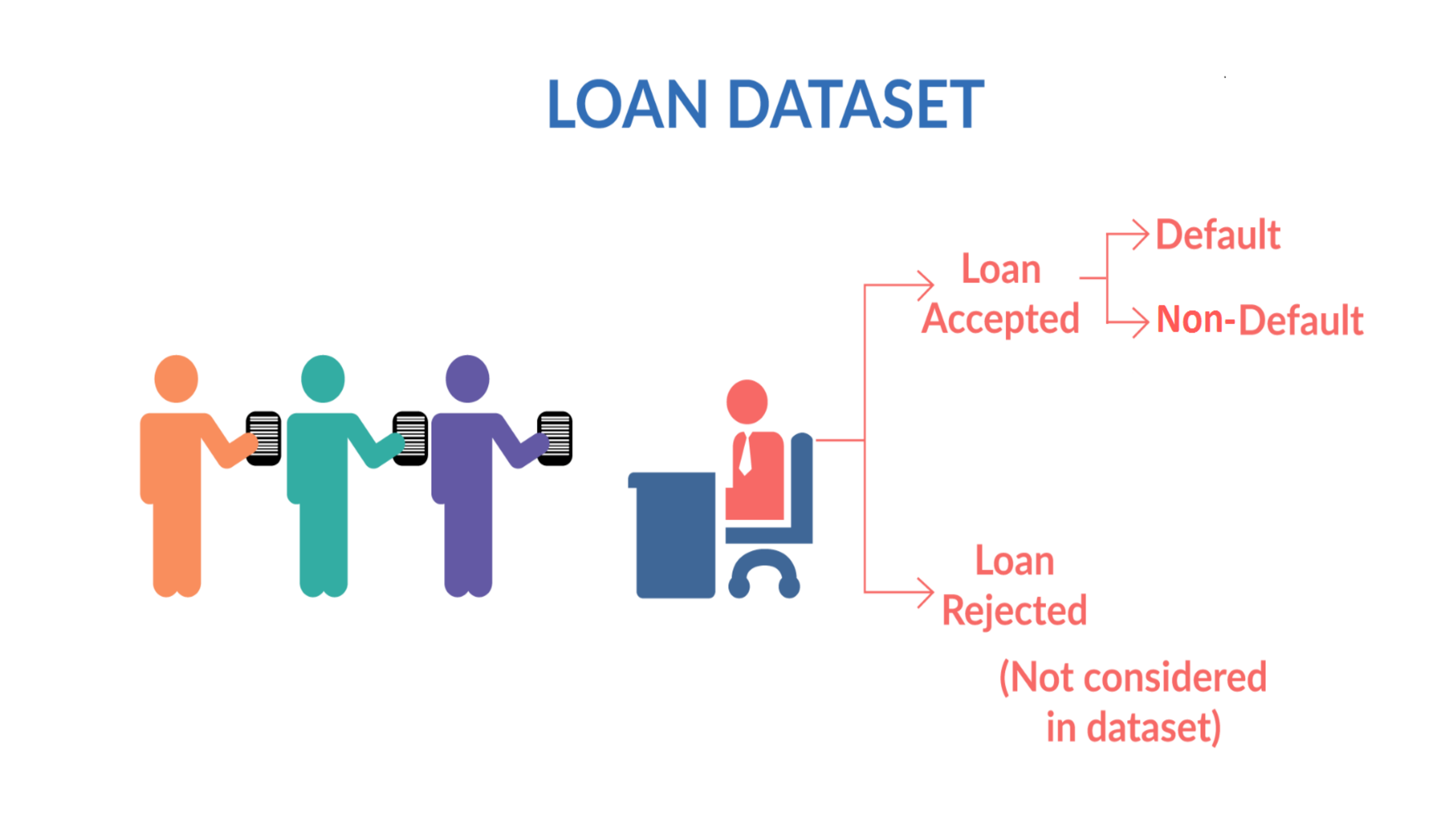
Introduction

The loan providing companies find it difficult to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it to their advantage by becoming a defaulter.

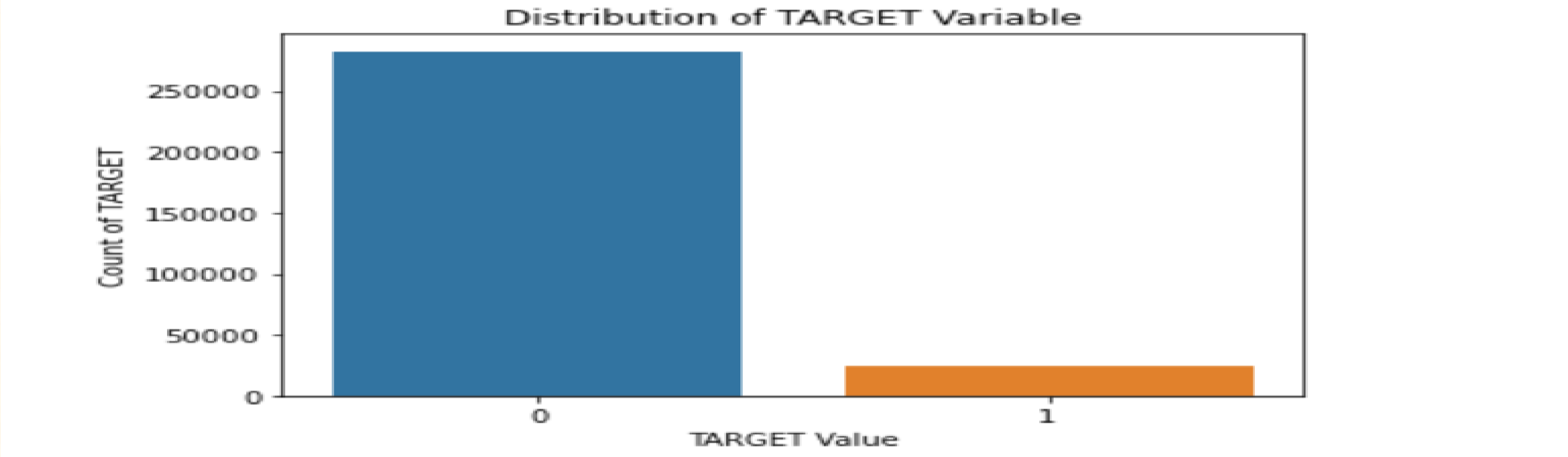
This case study aims to identify patterns which indicate if a client has difficulty paying their installments, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected.

In other words, the Bank wants to understand the driving factors behind loan default, i.e. the variables which are strong indicators of default. The Bank can utilize this analysis for its portfolio and risk assessment.

EXPLORATORY DATA ANALYSIS ON BANK LOAN DATA

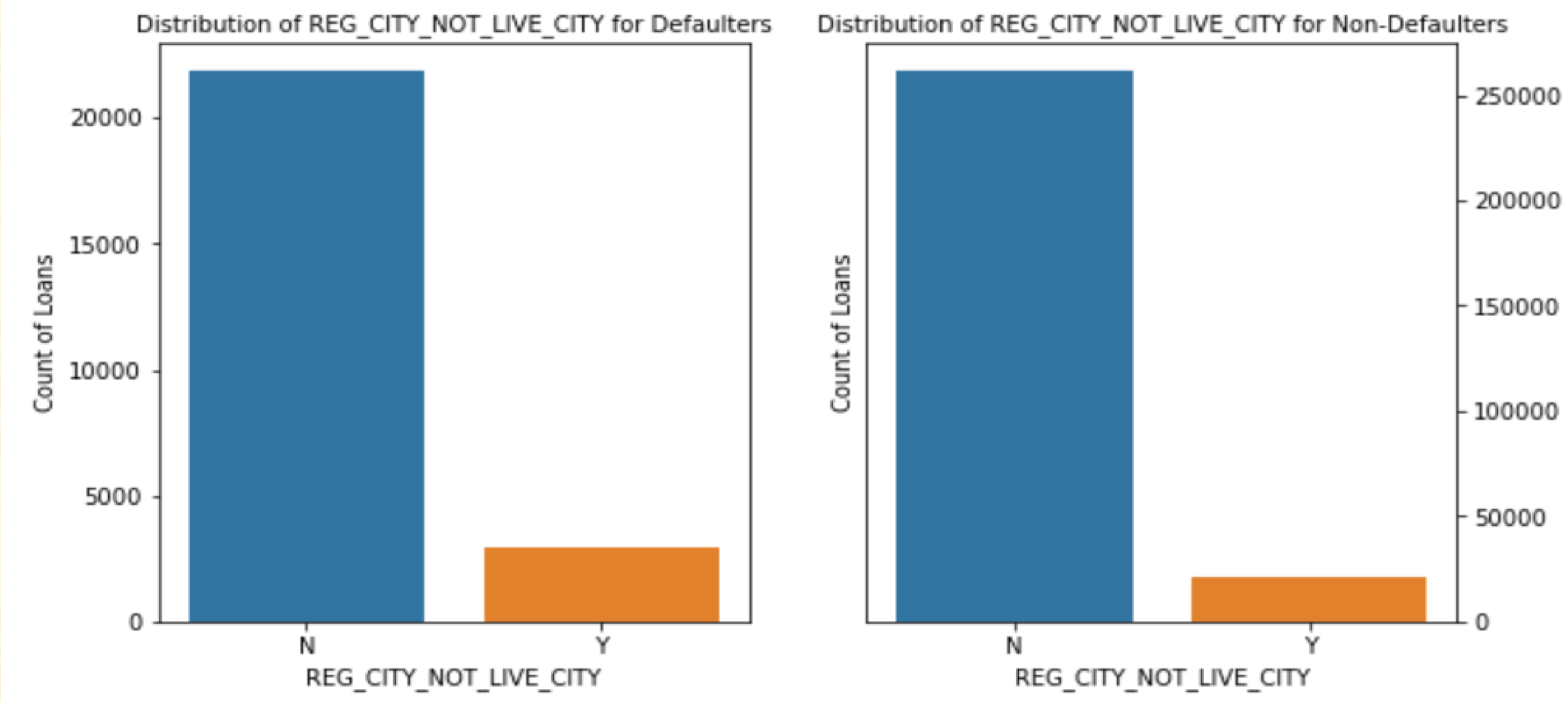

LOAN DATASET
Loan Accepted
Default
Non-Default
Loan Rejected
(Not considered in dataset)

By performing EDA on the datasets provided, we found out the total loan distribution among the defaulter and the non-defaulters. We have visualized the data using different plotting methods for better insights.

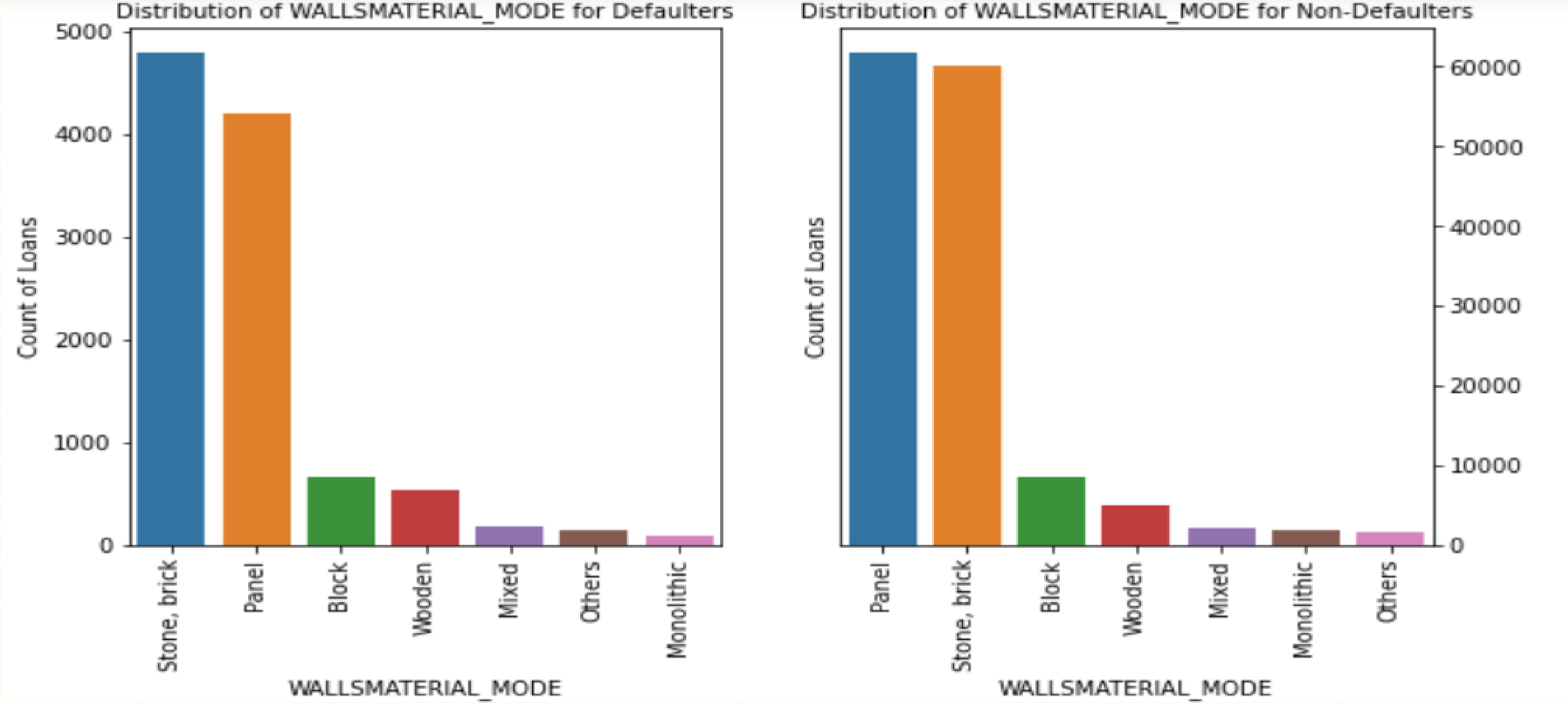


From this plot, we see this is an imbalanced data set. There are far more loans that were repaid on time than loans that were not repaid.
More than 250000 loans were repaid, less than 50000 loans were not repaid.

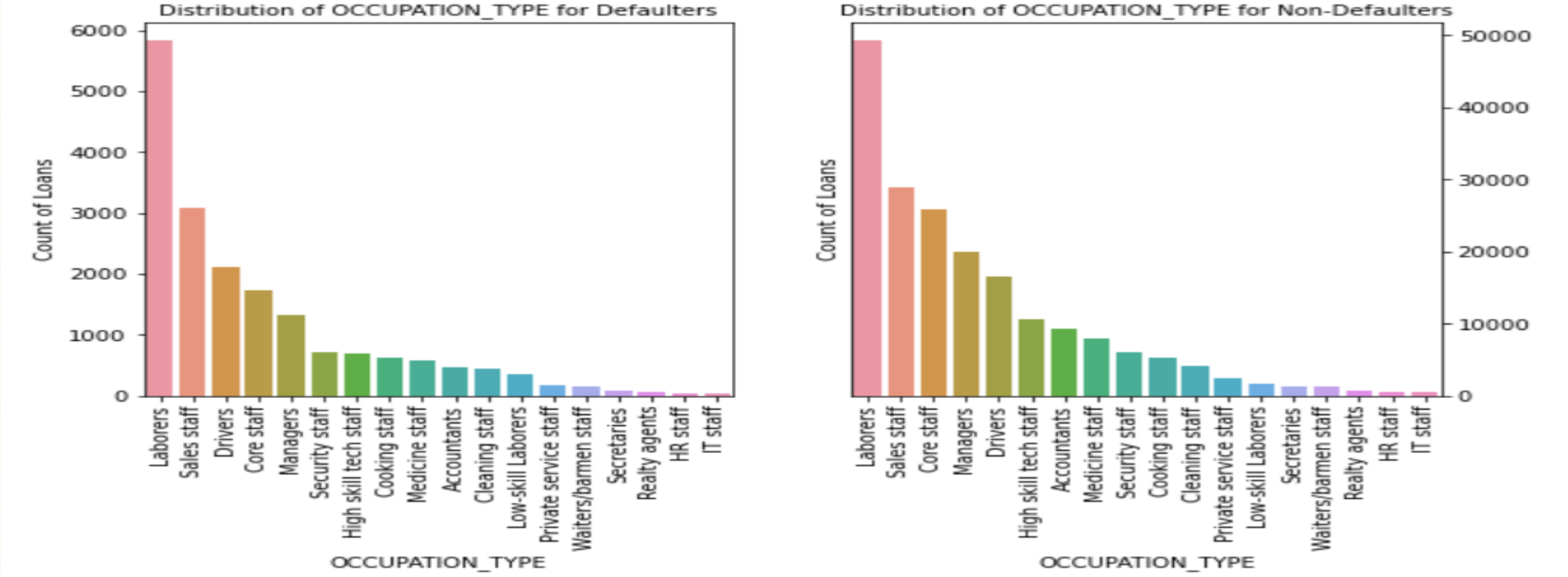
We observed that the ratio of people whose registration city is not same as the live city or work city is higher in case of the defaulters.



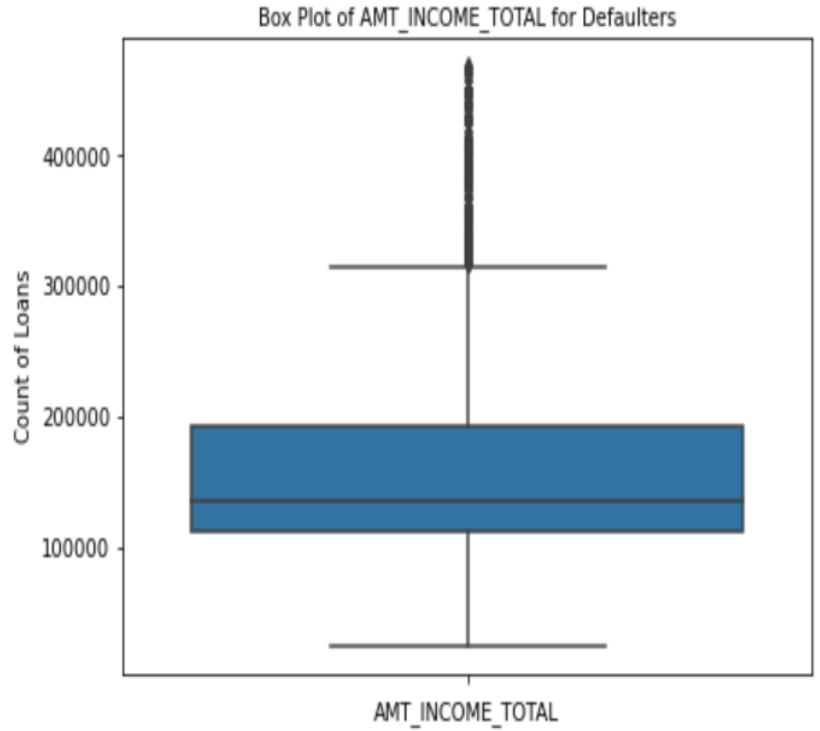
This chart tells us that most defaulters have houses made up of stone and brick while most non-defaulters have houses made up of panel.

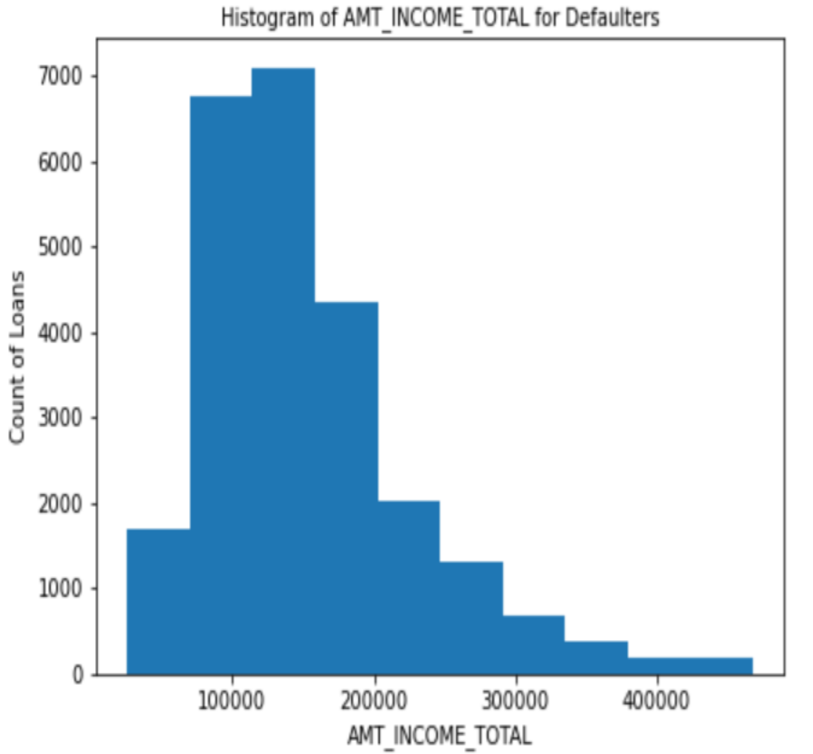


Realty agents, HR staff and IT staff occupation people are less likely to be defaulters. They can be seen as a potential customers.

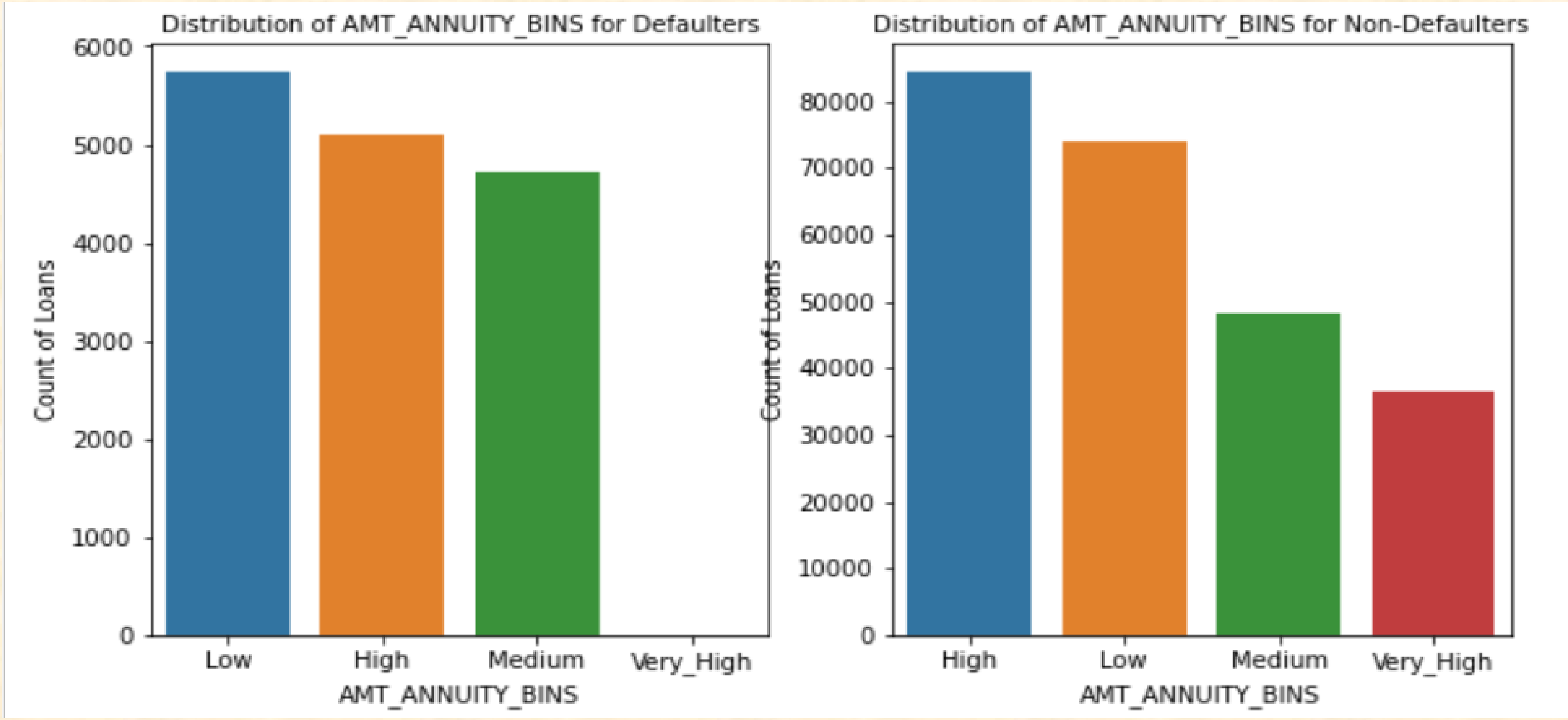


Below box plot and histogram tells us that most people with payment difficulties have incomes in the lower range between 100000 to 200000.

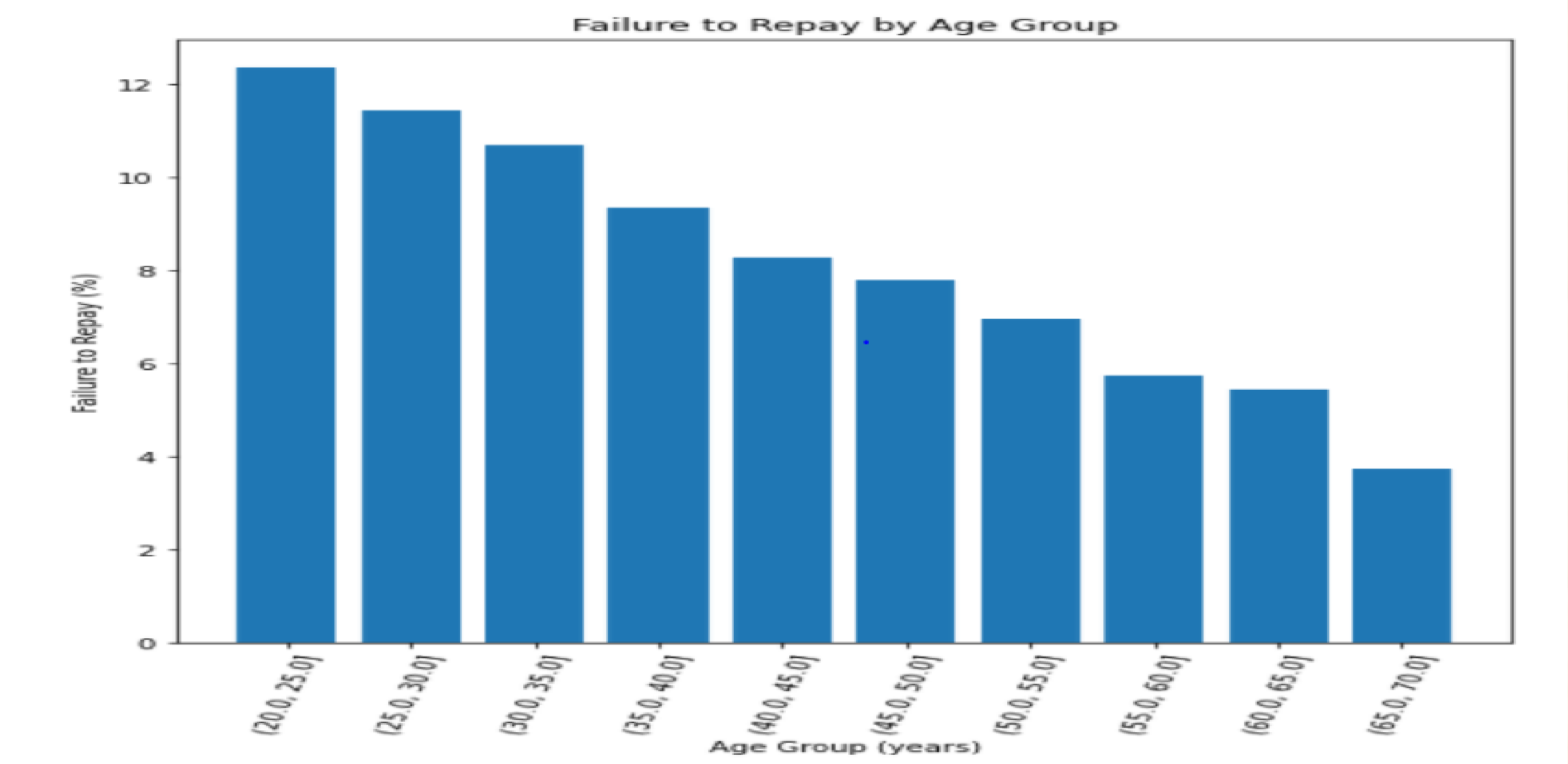




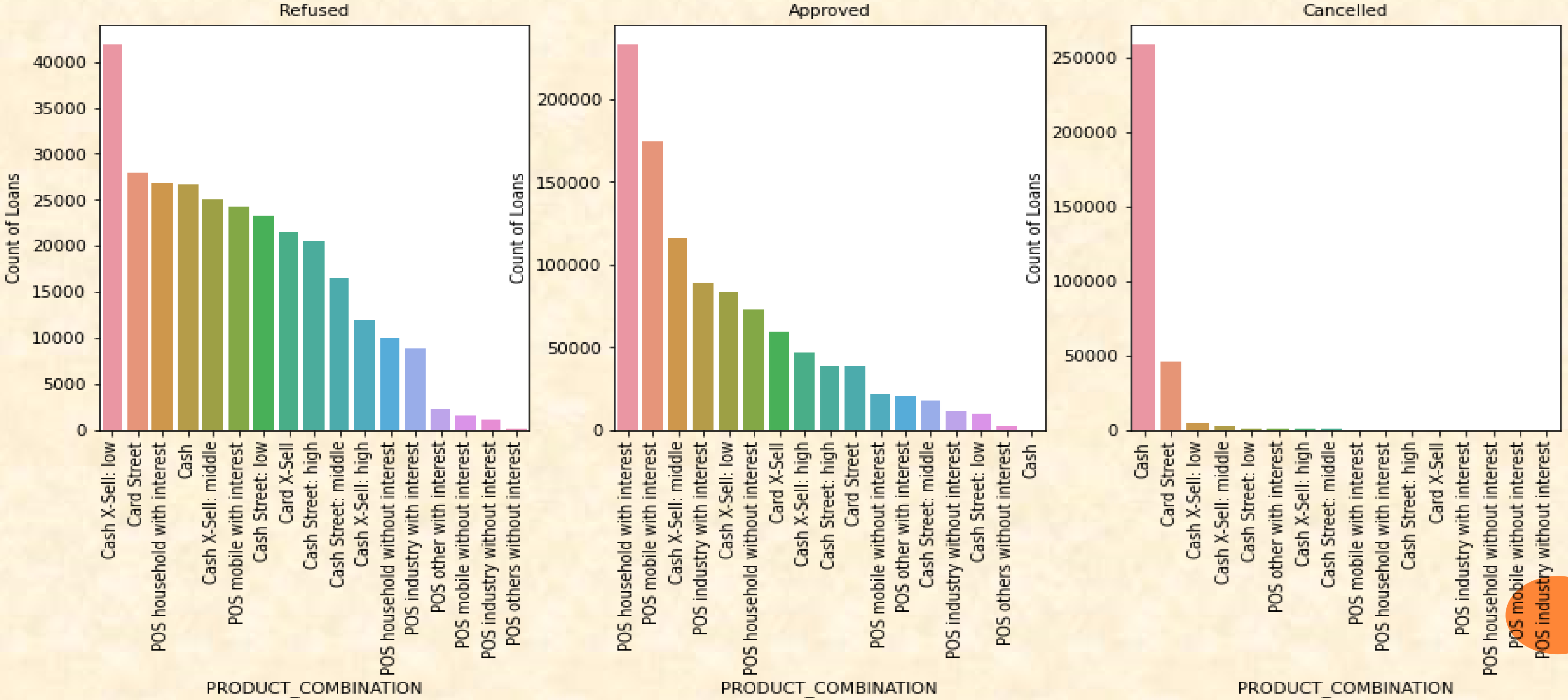
Maximum number of defaulters have low annuity values , whereas maximum number of non-defaulters have high annuity values.



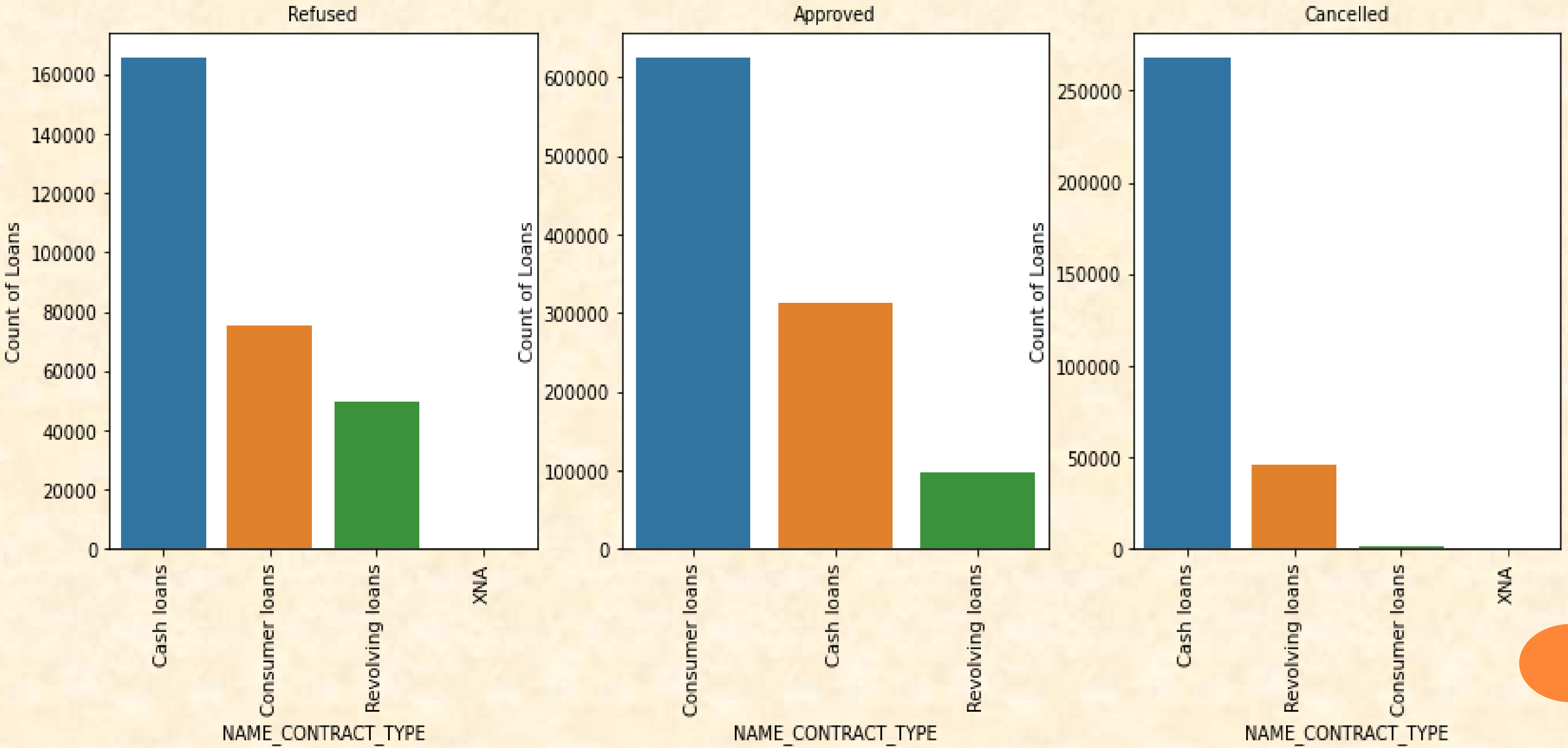
Maximum people who fails to repay the loan is in age group 20-25.



We observed that most number of loans were approved for POS household with interest. Most number of refused loans were of Cash X-sell: Low Production combination.



Cash loans were mostly cancelled or refused, whereas consumer loans were the ones which mostly got approved.




Thank you