

CREDIT EDA CASE STUDY

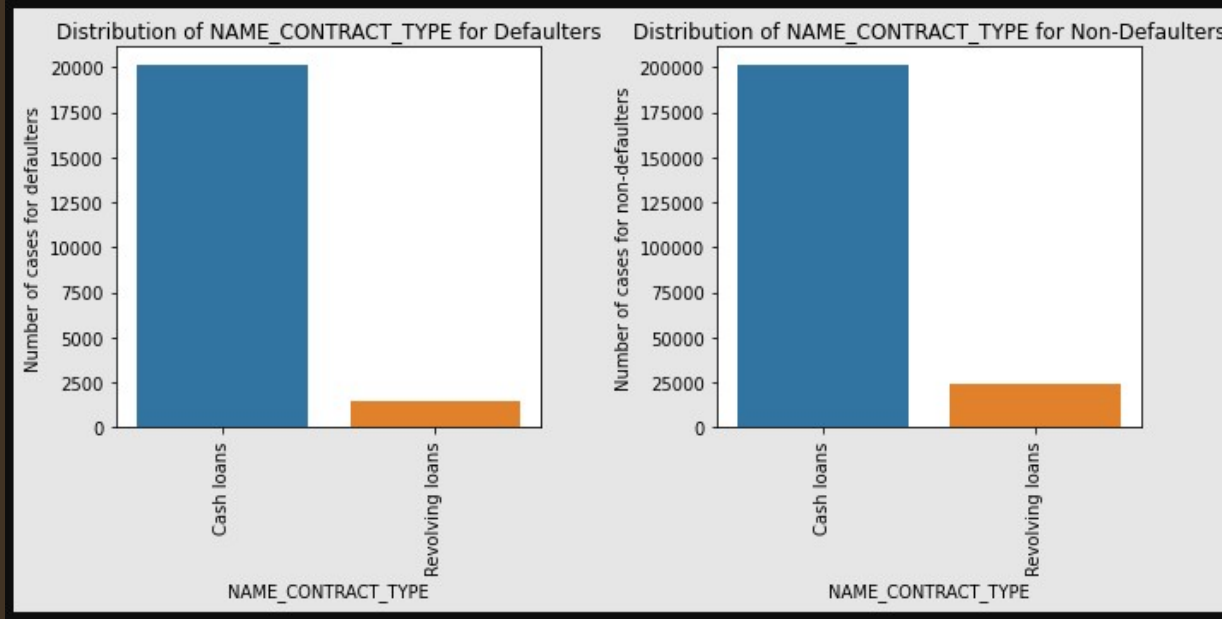


Presentation by Bharat Gupta

Uni Variate Analysis of Different Variables



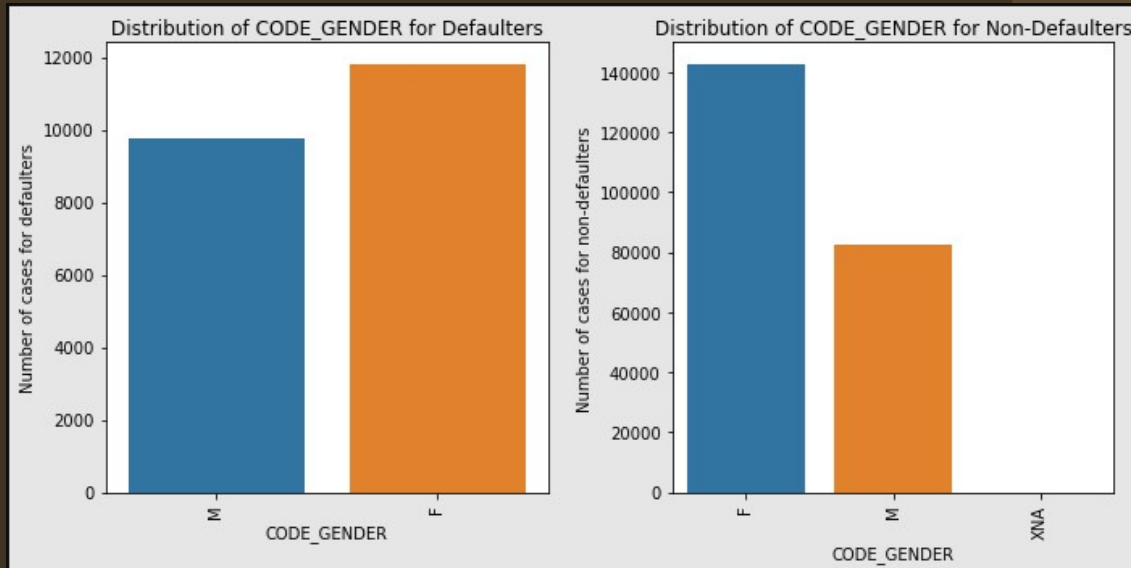
Distribution of Contract Type



Inferences:

- Revolving Loans are less likely to be defaulted
- It may be because of higher flexibility of repayment in revolving loans than cash loans
- Flexibility in repayment, choosing tenure and re-borrowing may leads to lesser defaults.

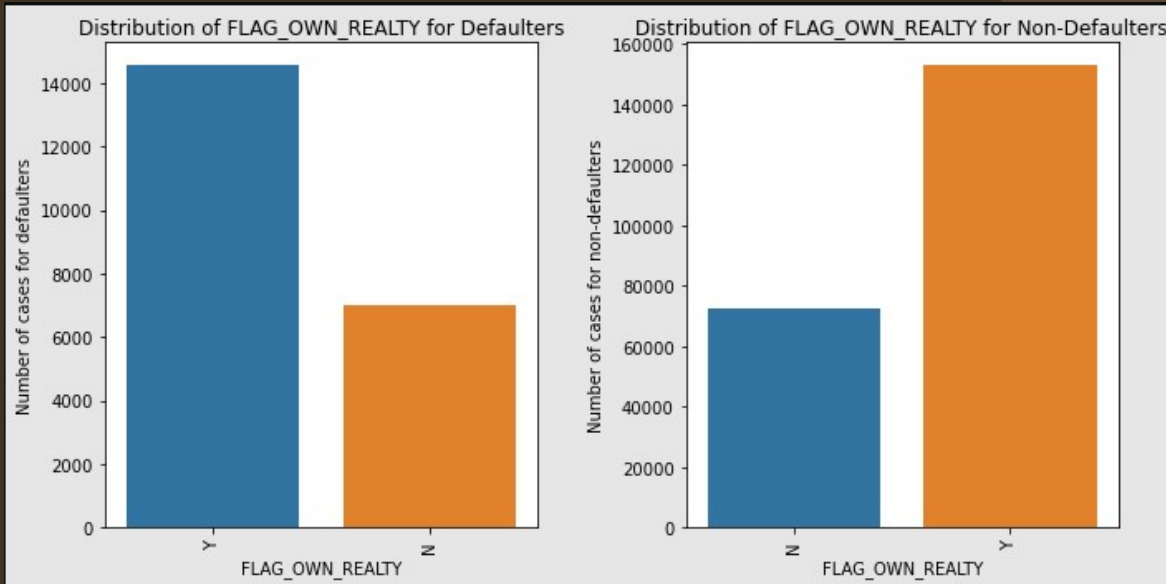
Distribution of Age



Inferences:

- Total number of loans taken by the females is more than males
- Proportionate default rate for Males is more than Females even though total no. of loans taken is more for Females.

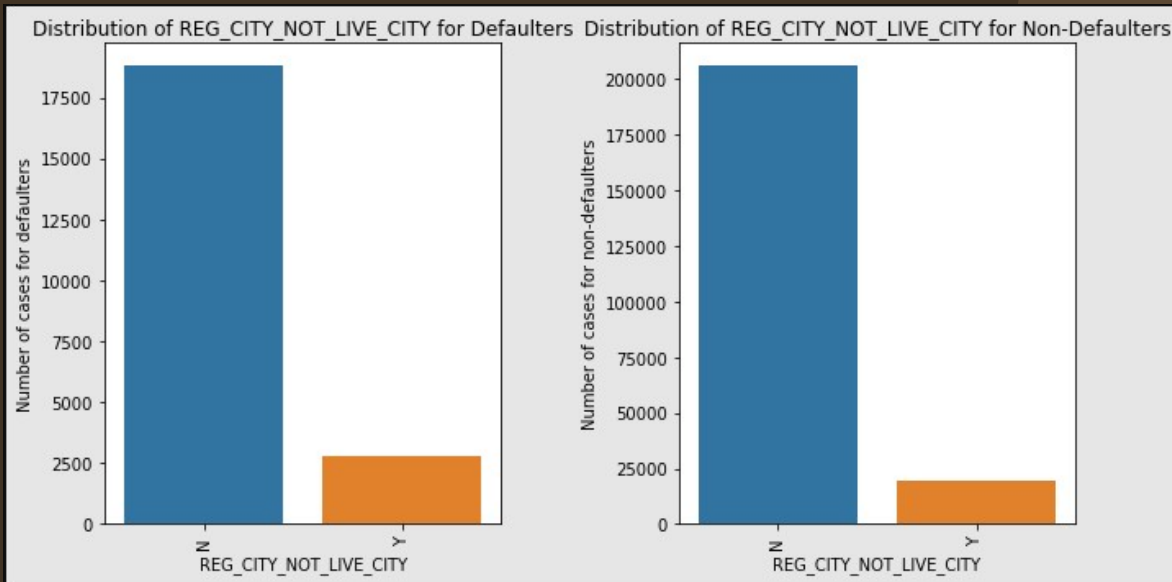
Distribution of Having Own House/Flat



Inferences:

- No. of people who has own house/flat has applied for loan more than who do not have own house or flat
- The percentage of defaulters who has No-Own house is slightly more than who has Own house

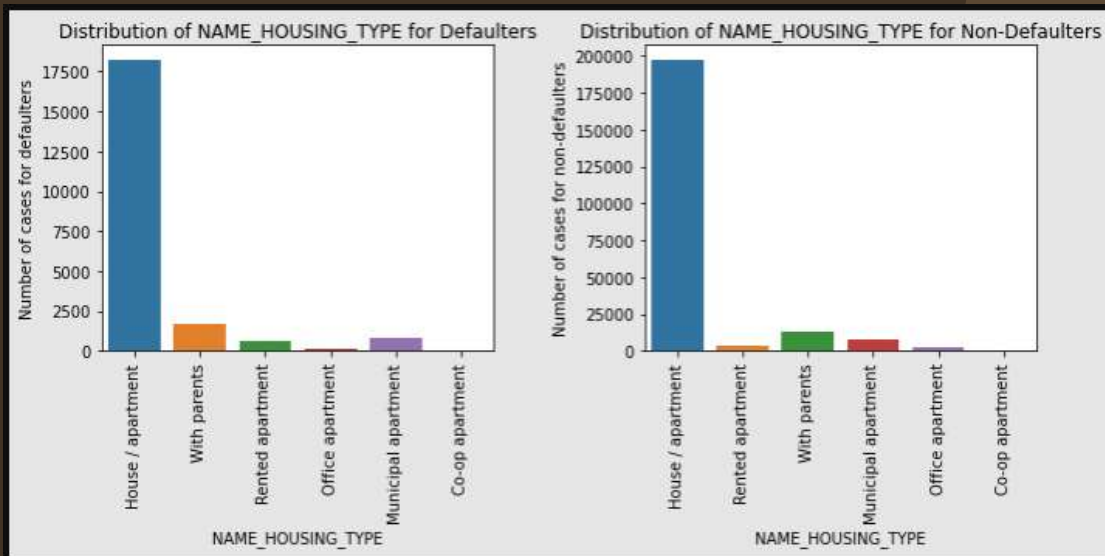
Distribution of Registered City and Live City



Inferences:

- No. of people whose registered city while applying for the loan is not same as live city are more than whose registered city and live city are same
- Similarly, the persons whose registered city and live city are not same are having more likely to default

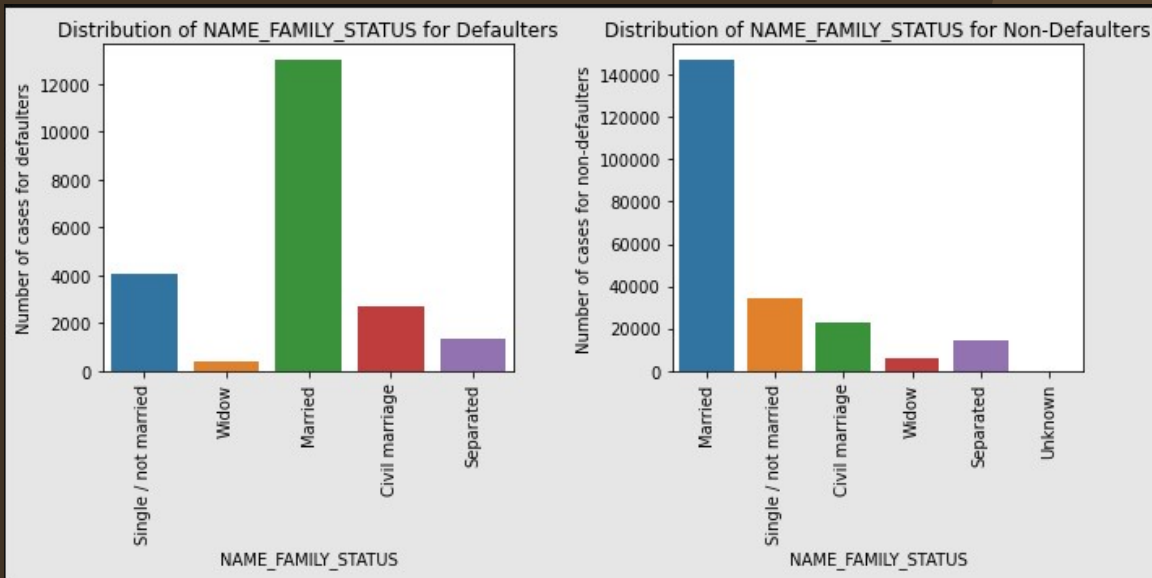
Distribution of Housing Type



Inferences:

- Most of the people who has applied for loan are residing in House/Apartment
- Ratio of people who are living with parents are having more default rate

Distribution of Family Status



Inferences:

- Married people are more likely to avail loans which may be because more responsibilities after marriage or applicant is the main income earner in the family
- No. of married people availing loans is more and having more payment difficulties as well – which may be because of unforeseen family difficulties or having children which may leads to higher financial burden of school fees, medical bills..etc..

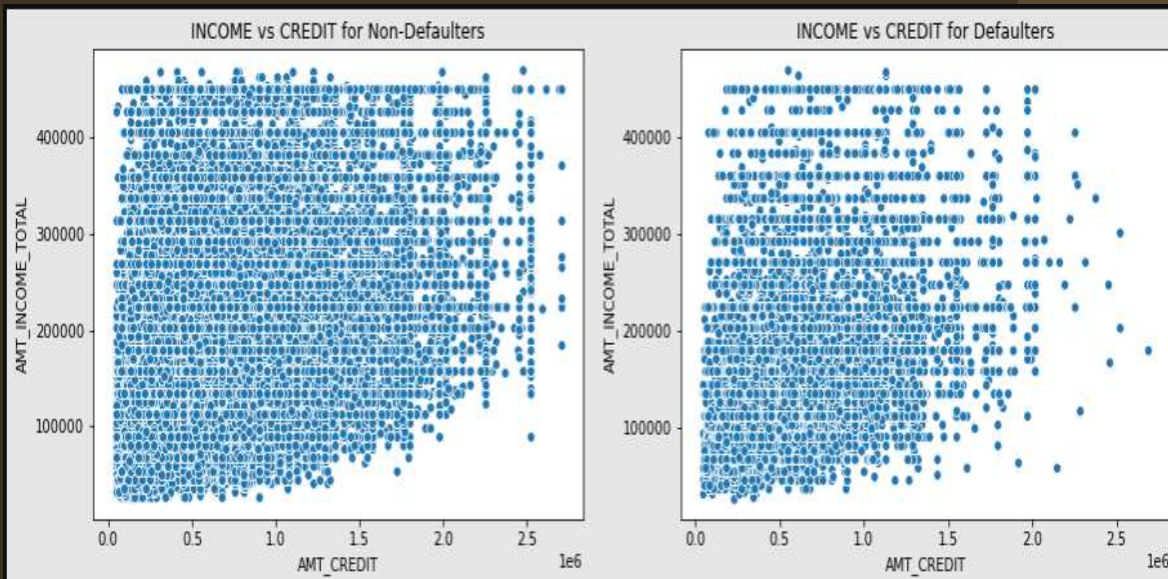
Bi-Variate Analysis



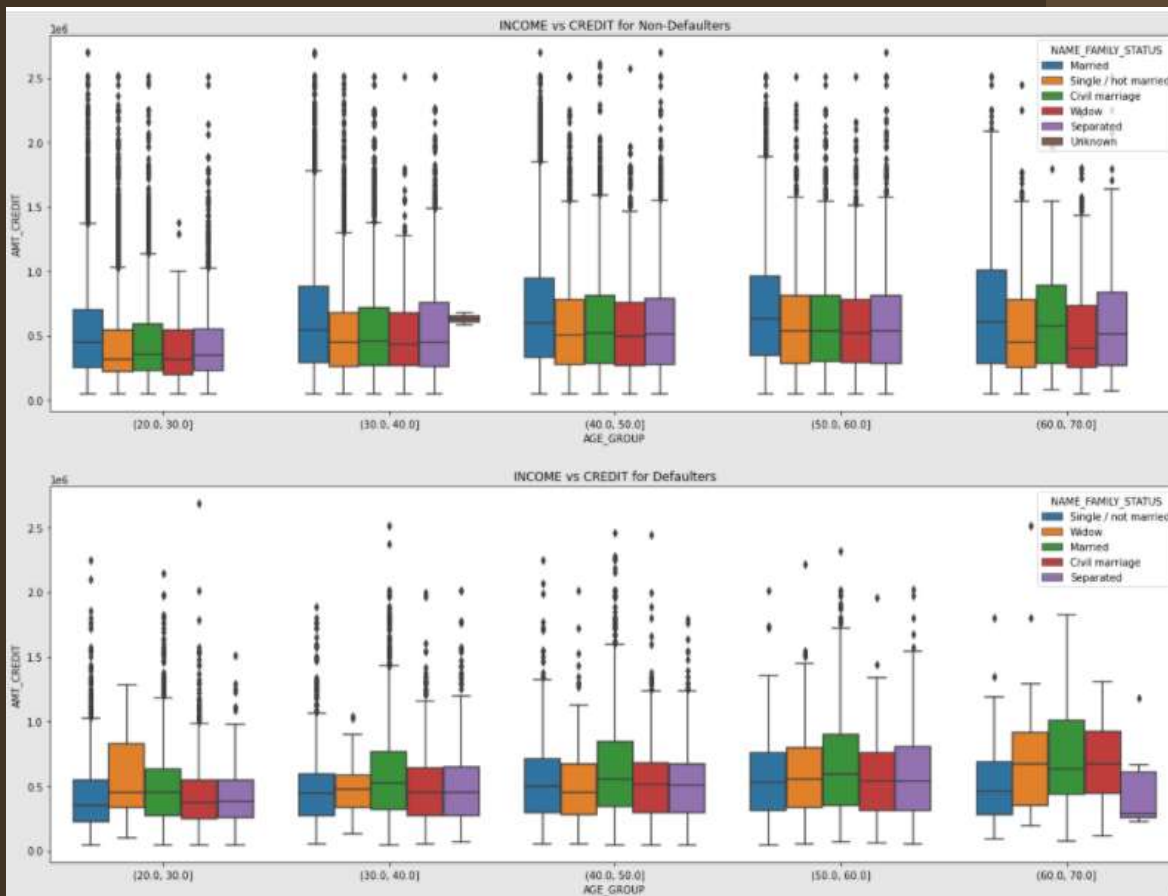
Income Vs Credit Bivariate

Inferences:

- When the income above 3lacs, lower the density of defaults – Higher the income lower the defaults
- AMT_CREDIT is less for income below 1lac – This is because of lower loan eligibility for lesser salaries
- Some of the loans are defaulted even though the income is above 3lacs and amt_credit is low

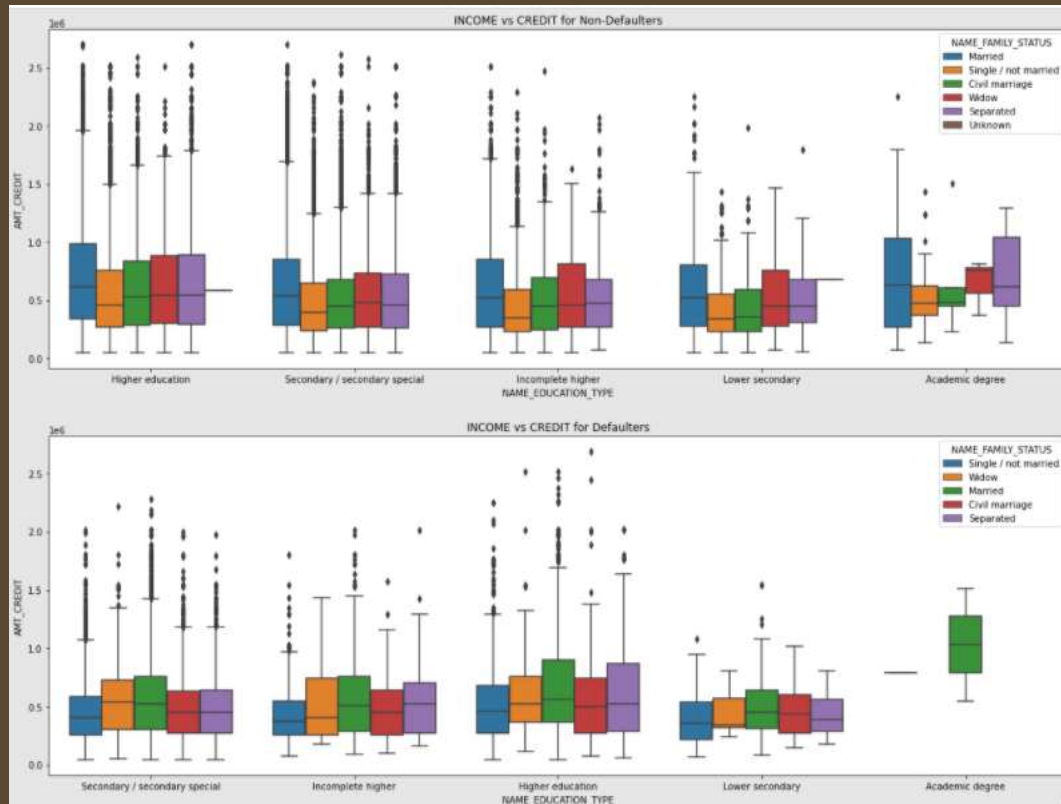


Credit Amount vs Education



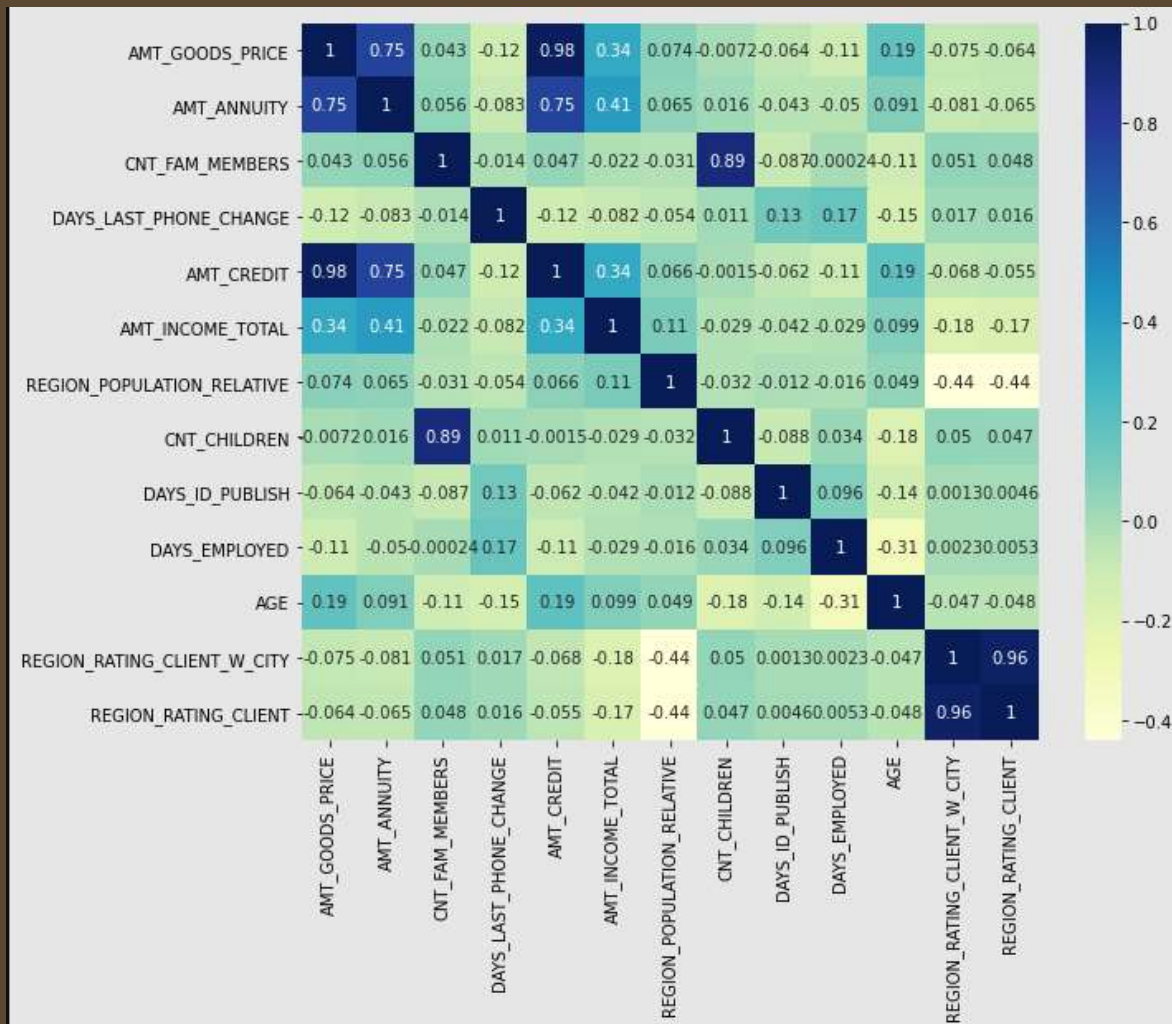
- Can see some extreme outlier in age group of 20-30 which may be because of good salary of client or a typographical error
- Married People in almost all the age group are in greater number to apply for loans than others.
- Married People in age group 60-70 are proportionally more defaulters than others.

Credit Amount vs Education



- Married People are proportionally higher non defaulters for all education levels.
- Married People with academic degree and high amount credit are more defaulters than others.

Correlation of selected variables for Defaulter



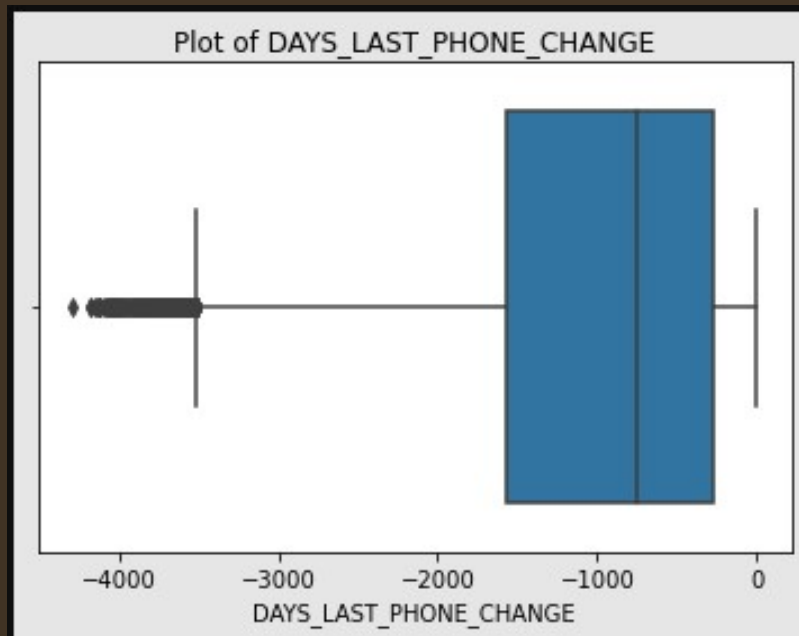
Inferences:

- Rating of the region where client lives and rating of the city where client lives in are directly proportional
- Goods price and Credit Amount are positively correlated, which means higher the consumer durable good value leads to higher the loan sanctioned
- Count of family members and Count of children are positively correlated
- Age of the customer and experience of the defaulters are negatively correlated, which may be because of getting job/income lately and leading to the payment difficulties due to bills and expenditures.

Analysis of Different Variables (Distribution & Outliers)



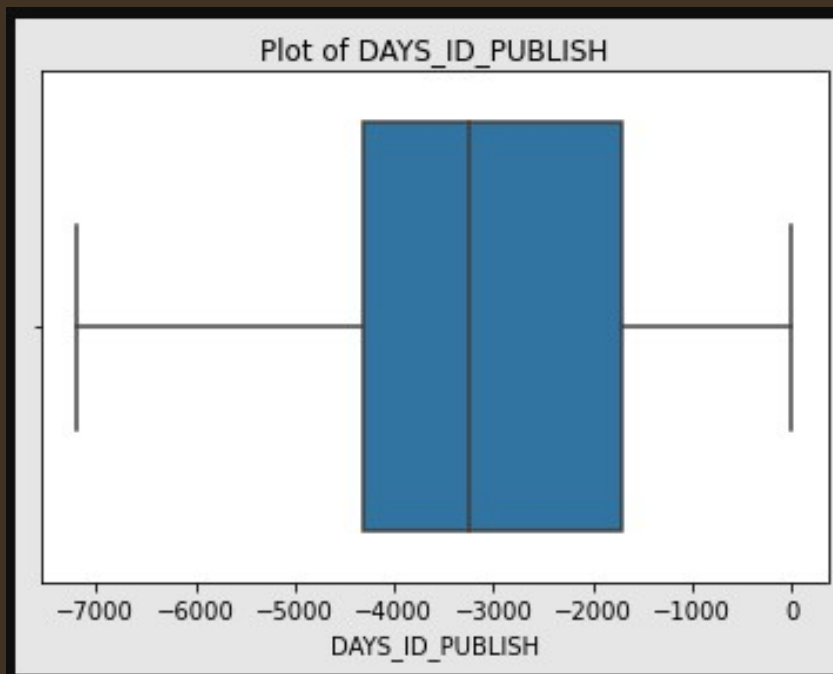
Box Plot for How many days before the application did client change phone



Inferences:

- There are clients who changed their mobile number a no. of days of loan application
- Change of mobile number before the loan application is a bad indicator and warns about reaching the client in case of default

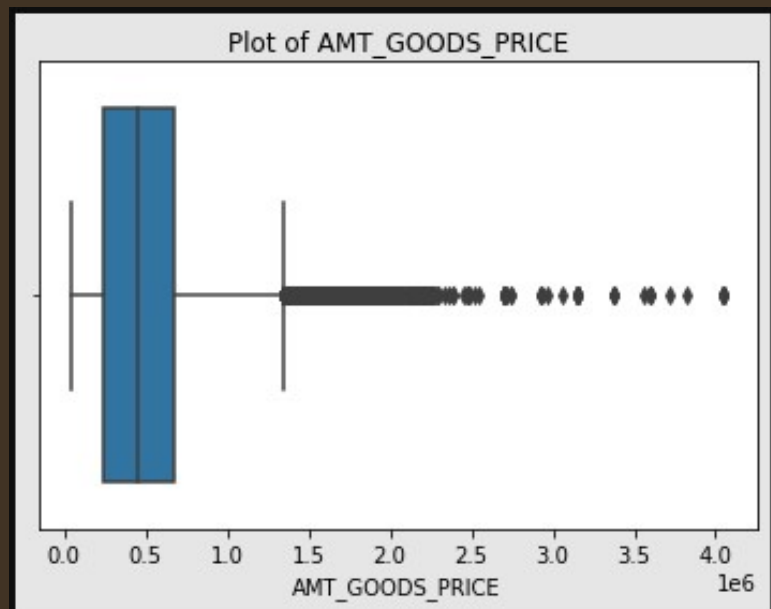
Box Plot for Distribution of ID change Days



Inferences:

- Credit history is generally linked to the KYC or ID documents of the client
- A recent modification or change in Identification documents may be because of the derogatory in previous loans or because the client is much younger and availed required ID documents recently

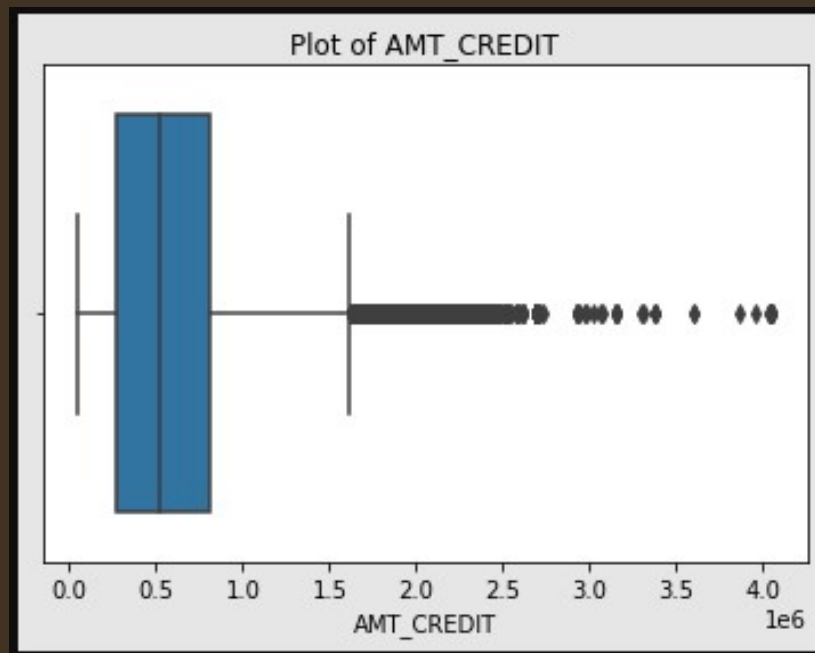
Box Plot for Goods Price



Inferences:

- There are some extreme outliers in goods price, possibilities are:
- Client applied loan for buying luxurious consumer durable goods
- Or
- There is a chance of selecting other loan types as consumer durable loan (Manual data entry errors)

Box Plot for Credit Amount



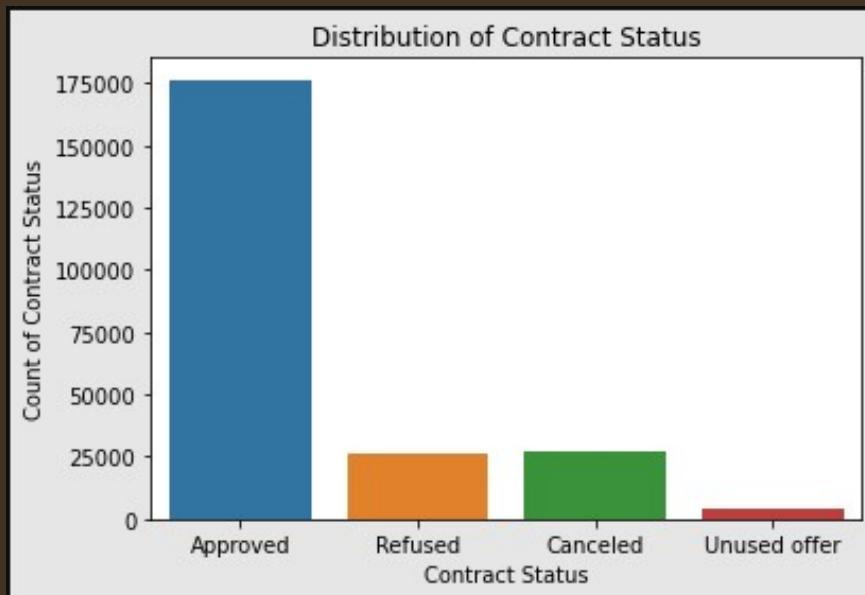
Inferences:

- Some outliers are clearly observable in credit amount.
- The first quartile is larger than third quartile for credit amount which means most of the credit amounts are present in the first quartile.

Analysis of Different Variables After Merging Previous Data

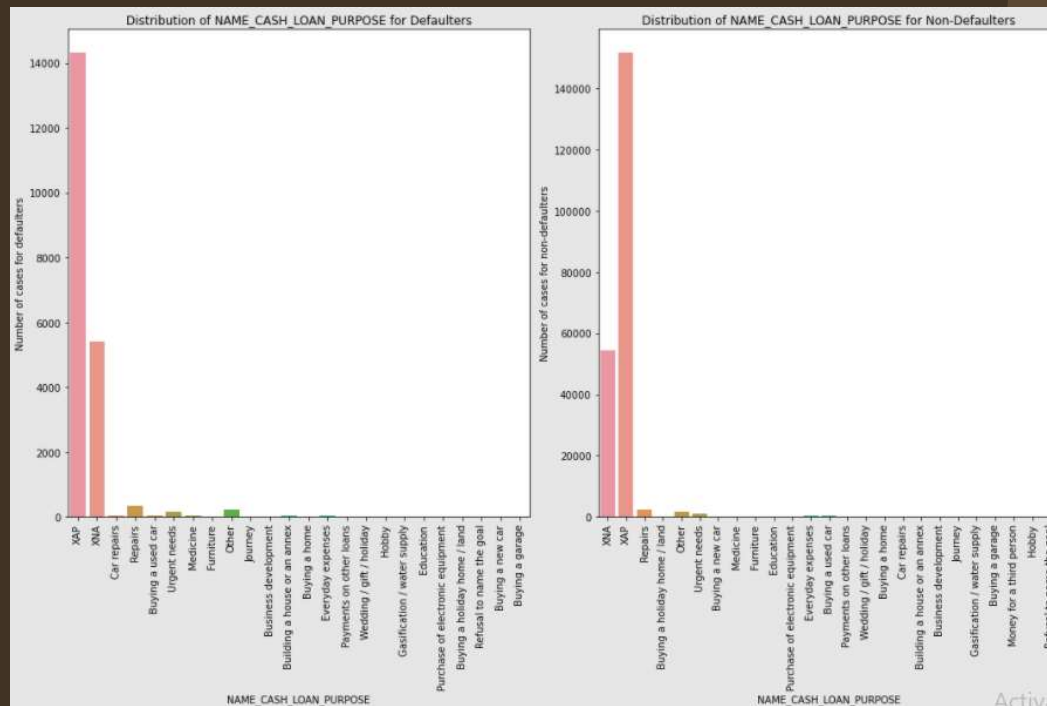


Status of Loan



- The counts of 'Approved' status is high when compared to the 'Cancelled' and 'Unused Offer'
- Cancelled loans are less which indicates the rate of interest or terms offered by the bank/lender are mostly acceptable by clients
- Refused loans are less which indicates bank/lender may not be rejecting the eligible client/applicant of the loan
- Unused offer could be reduced further by identifying factors which are leading the client to cancel the loan process

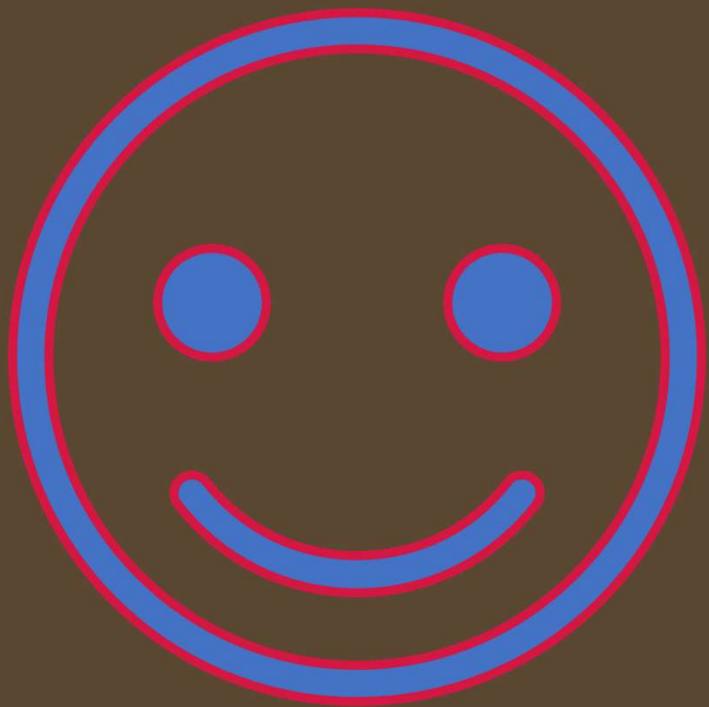
Analysis of Purpose of Loan



- We can observe the maximum portion of data is not available for loan purpose
- We can also observe that there are more payment difficulties for the purpose of loan 'Repairs'
- Then bank/lender should have the correct data about purpose of loan, because a loan which is required for gambling or betting or some un-conventional investment practices may lead to higher losses to client and will create payment difficulties

Conclusion:

- Lender/bank should focus on Income Type - Working as they are having most no. of unsuccessful payments
- Higher number of unsuccessful payments when the purpose of loan is "Repairs"
- Population living in Rented apartments and those living with parents have higher default rate
- Living in rental house cause payment of more bills towards rent and maintenance expenses may lead to less cash in hand to repay the loan
- Living with parents is having higher default rate which may be because of the medical conditions of elder people and payments towards medicine or hospitalization expenses is more



Thank You