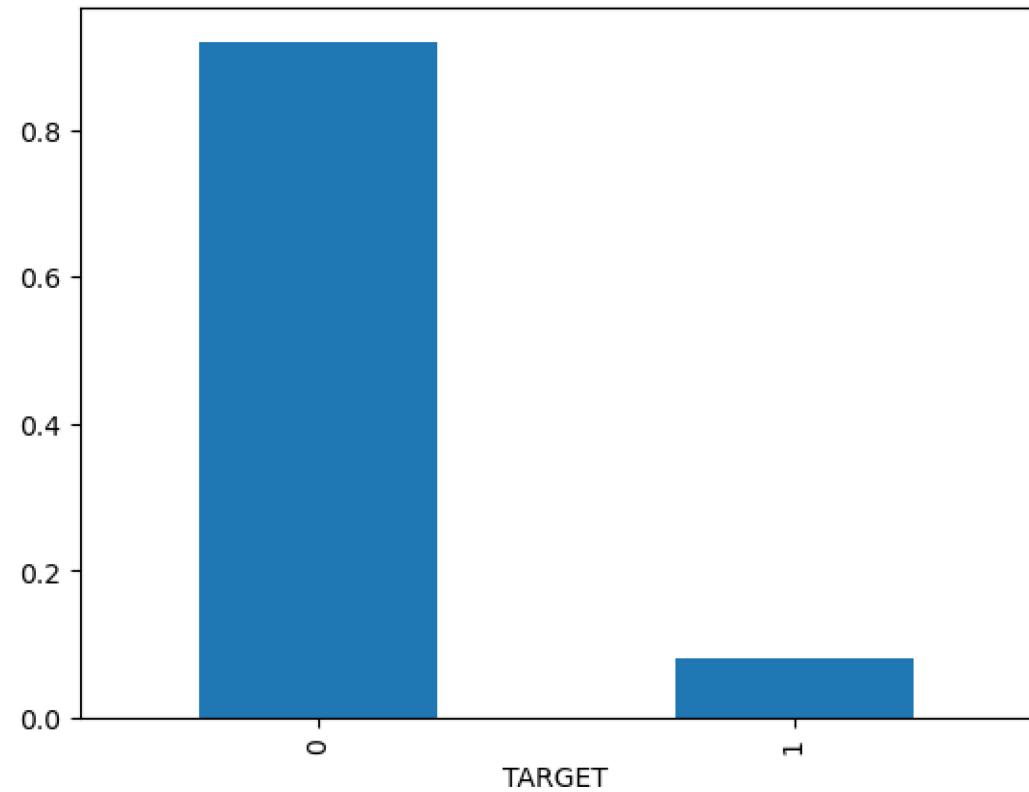


Credit EDA Case Study

By Manasa Sambaraju

Application Data

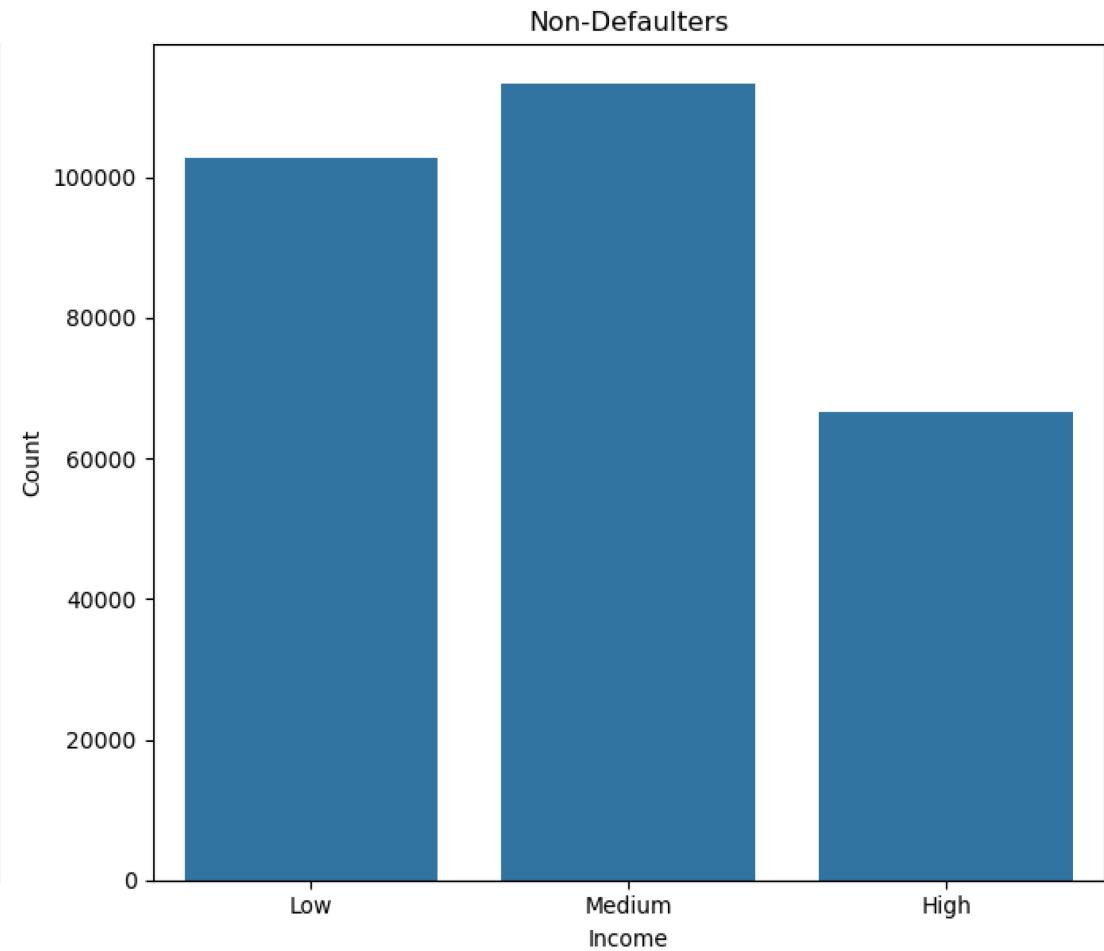
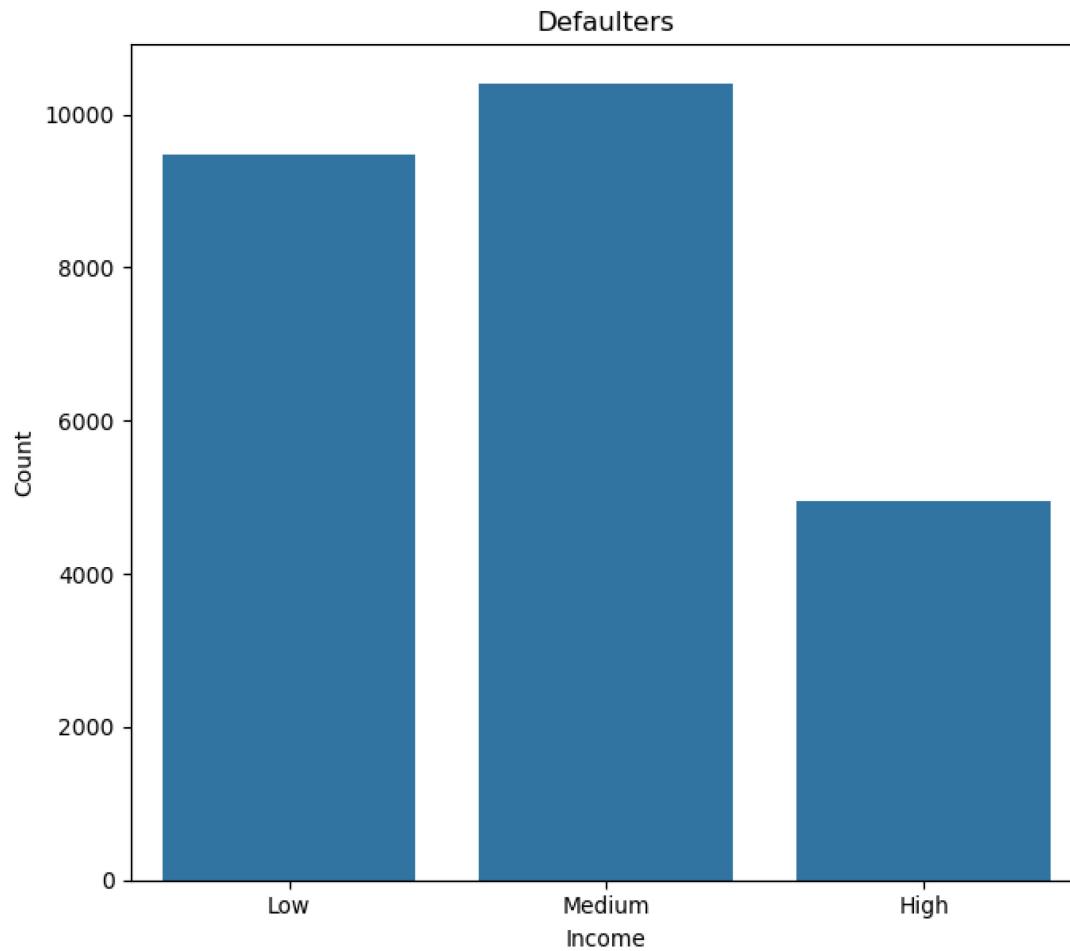


- 91% of data has the data with clients no payment difficulties and 9% of data with payment difficulties.

Key takeaways from Univariate and Segmented Univariate Analysis of Defaulters and Non Defaulters

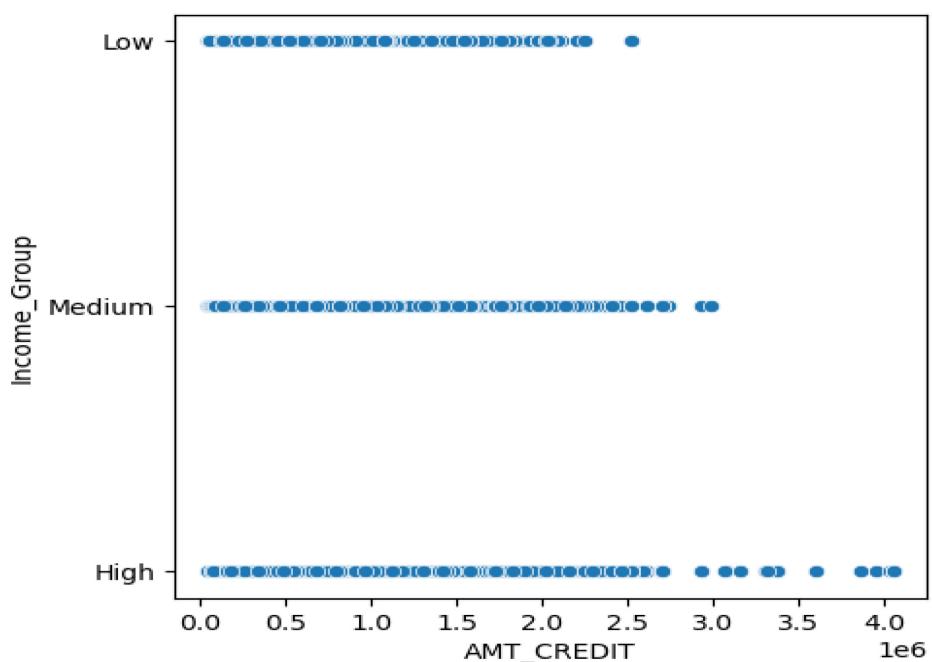
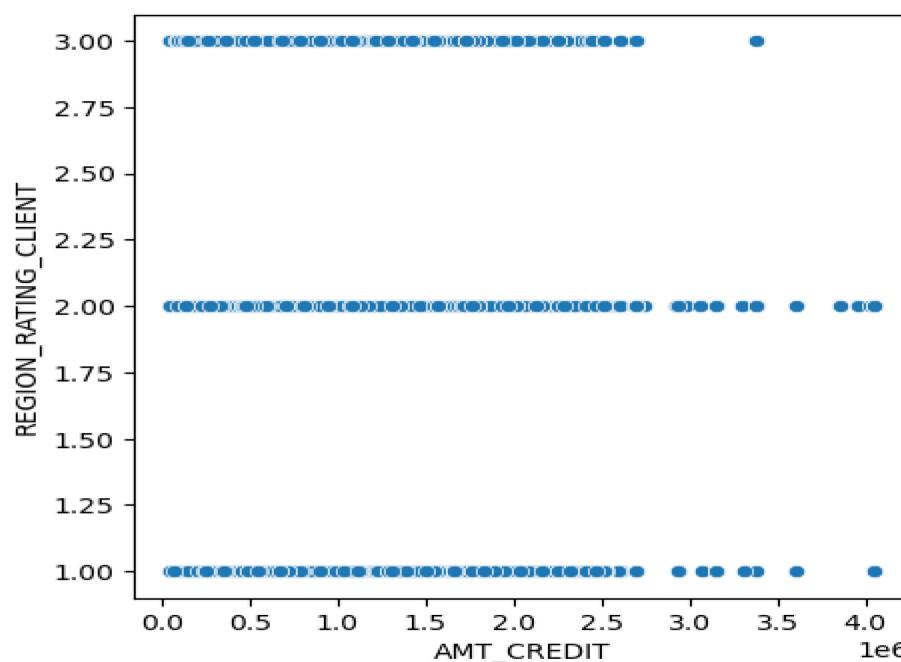
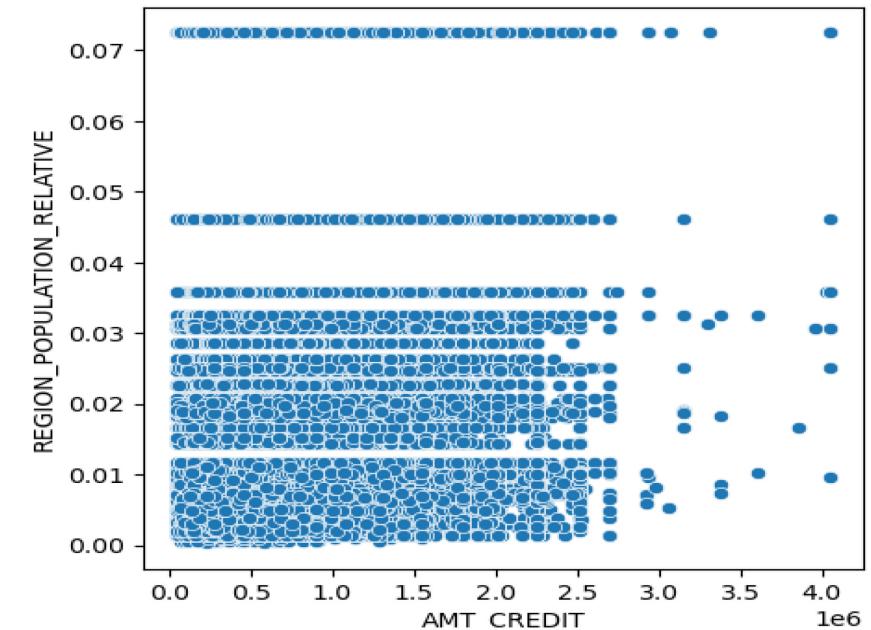
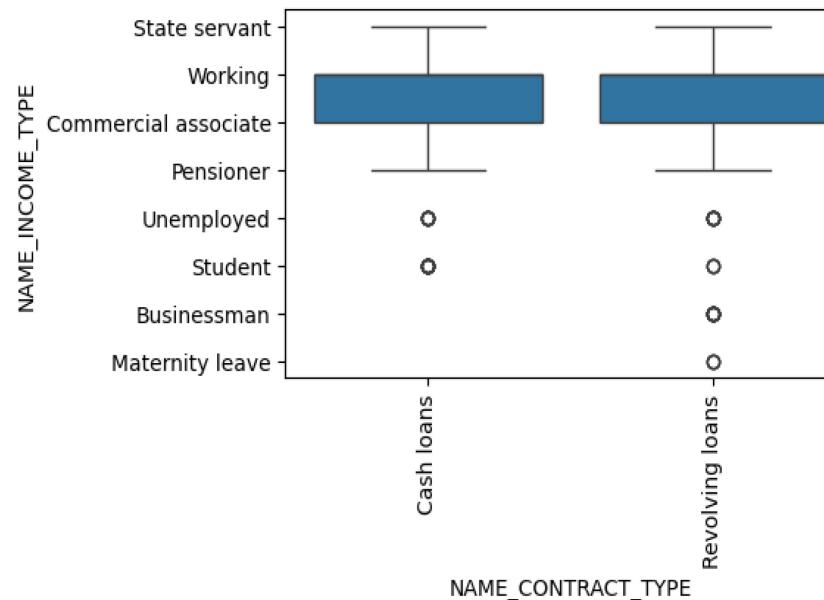
- Clients are more interested in taking Cash Loans than Revolving Loans.
- Clients who are Married, with no children are the one to take more loans than other categories.
- Clients who are Middle aged(30-60yrs),with medium income range(130000-202500 per annum) and with 1-5 yrs of Work Experience are the one to take more loans than other categories.
- Clients who are Females are the one to take more loans than Males
- Clients who own a House/Apartment are more likely to take loans than others.

Clients who are in medium income range are the one to take more loans and also the leading one in both Defaulters and Non Defaulters slightly beating the low income range



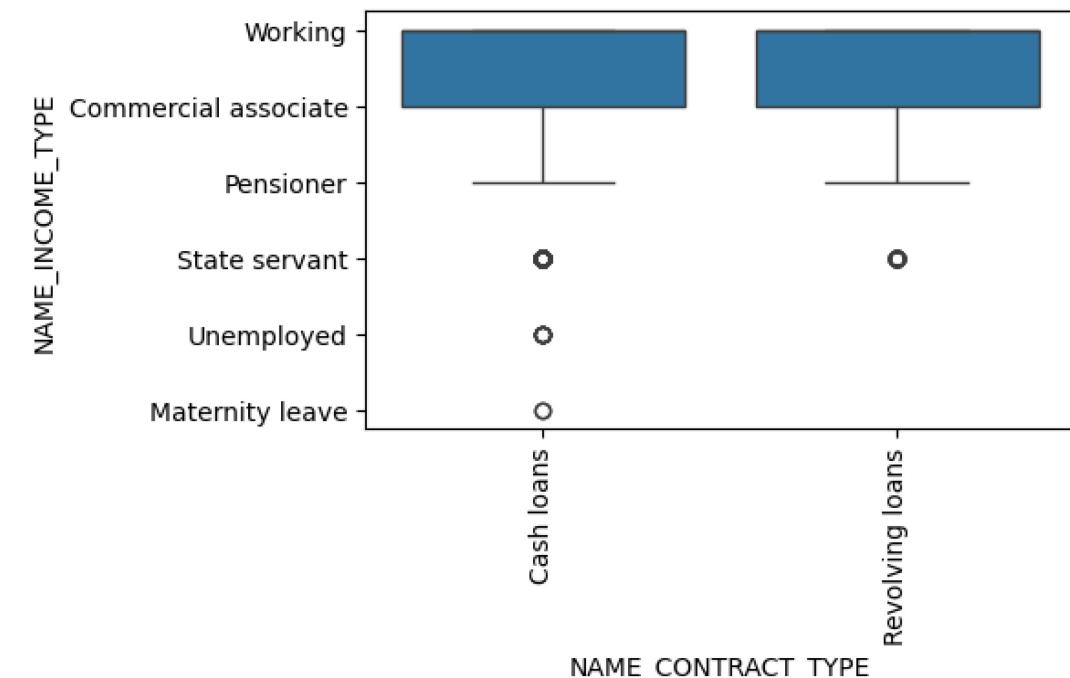
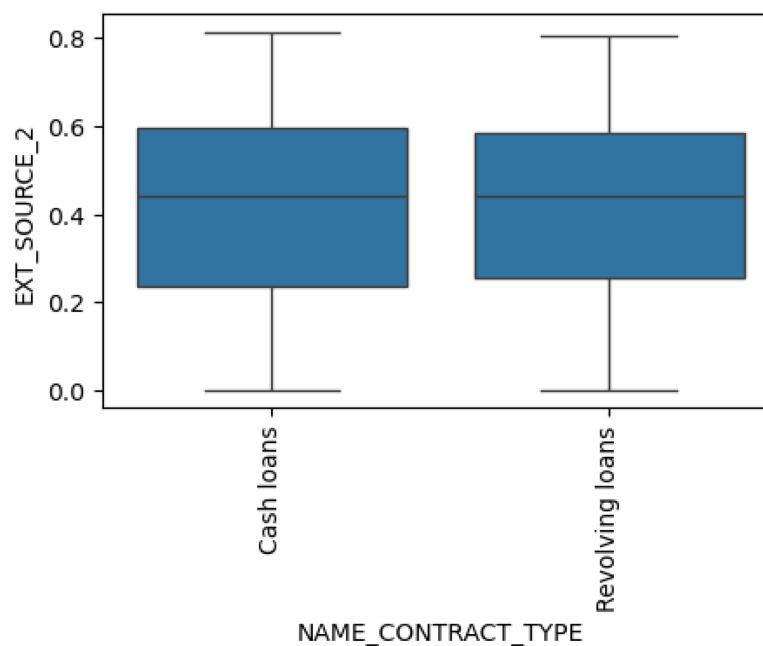
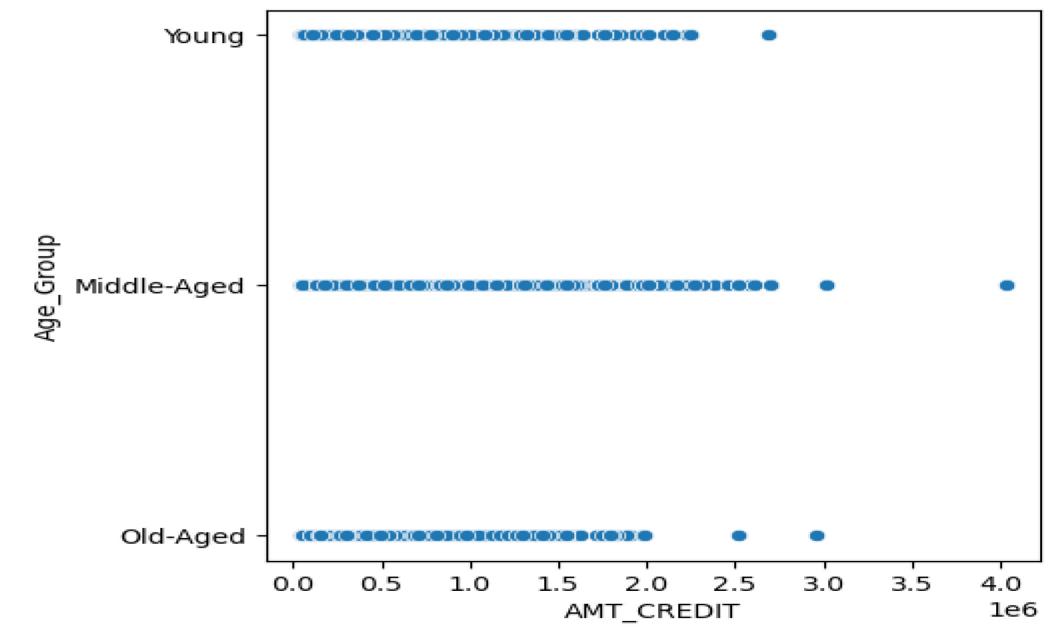
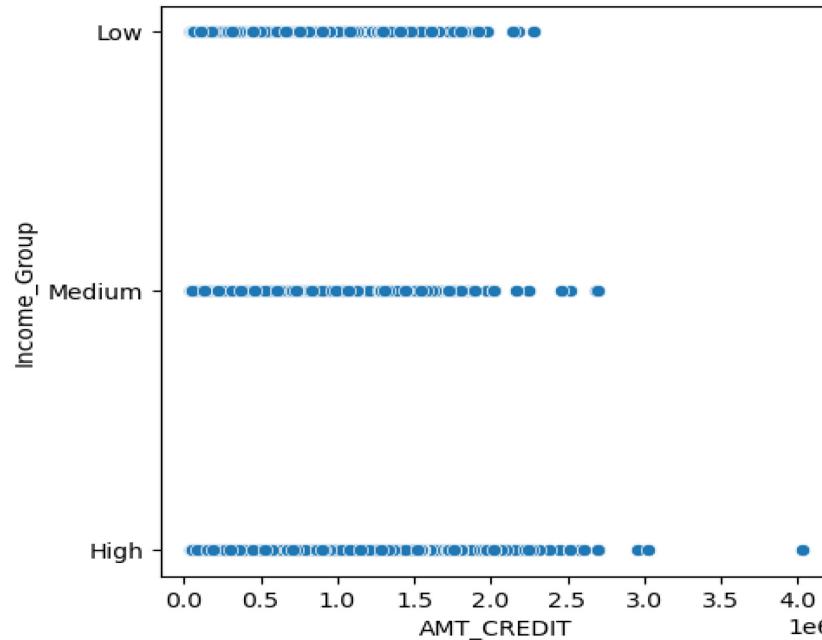
Key Takeaways from Bivariate Analysis of Non Defaulters

- Regional population relative with index 0.1-0.3 have take more loans than other
- Regional Rating and Regional rating of Client wrt city of rating 2 are the areas with more number of loans
- Clients with higher external source score have opted for more loan amount
- Higher Income Group has more amount Credited than lower and middle income group
- Cash Loans and Revolving loans are mostly spread across State Servant, Working , Commercial Associates and Pensioners.
- Revolving loans have less credit amount compared to cash loans
- Business Man tend to take more Credit Amount of loan than other classes
- Clients with higher education and academic degree tend to take more Credit Amount of loan



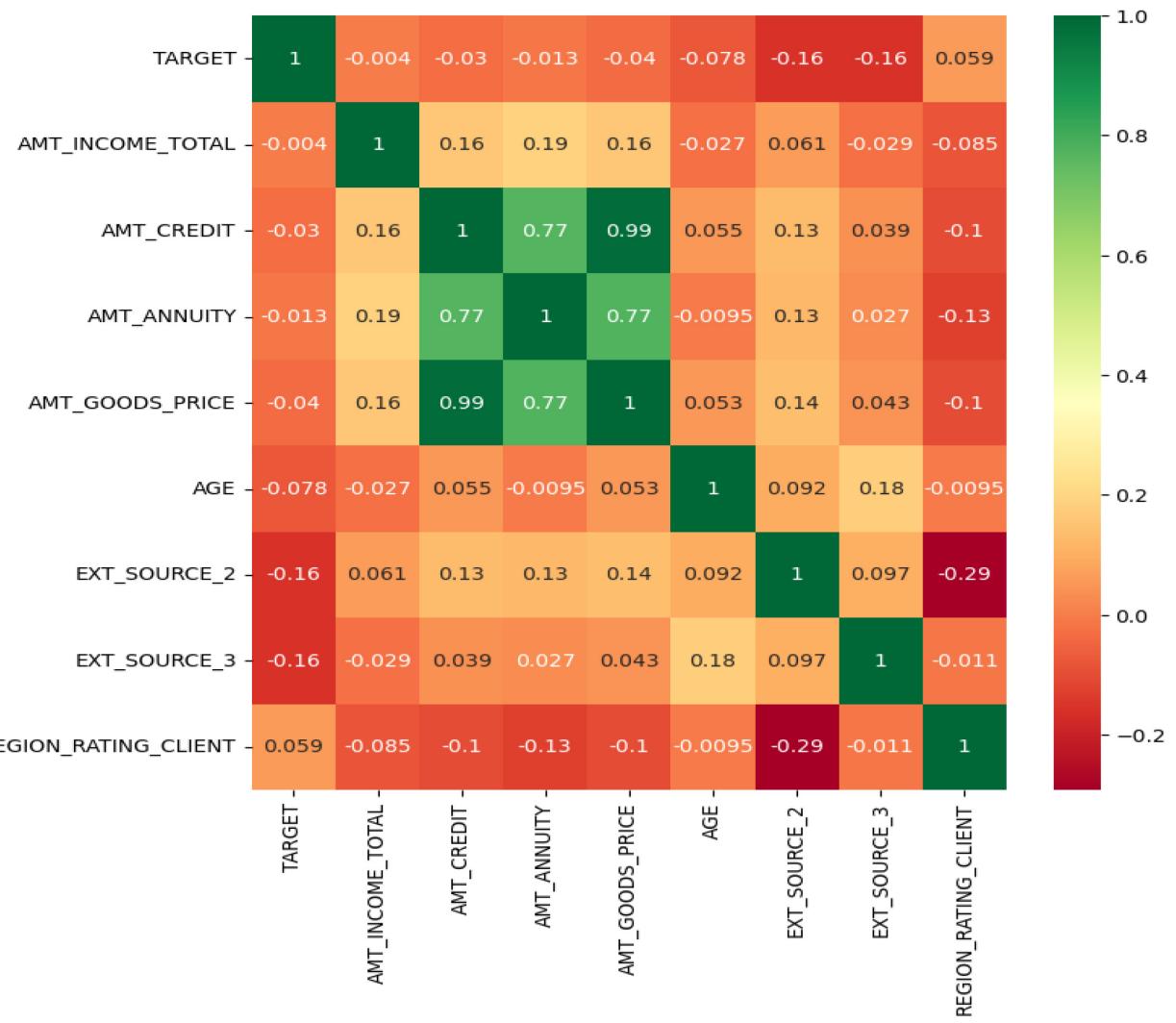
Key Takeaways from Bivariate Analysis of Defaulters

- Clients with Higher Income are more likely to be defaulters with payment difficulties
- Clients of middle aged group are more likely to be defaulters
- Working and commercial associates and some pensioners tend to have payment difficulties for both cash and revolving loans
- Mostly Married clients have payment difficulties for Cash loans where as range of married , widowers, single people tend to have difficulties for Revolving loans
- Clients with external source score with 0.3-0.6 IQR tend to have payment difficulties
- Most of Male and Female customers opted for Cash loans have difficulties in paying with few of them in revolving loans



Multivariate Analysis

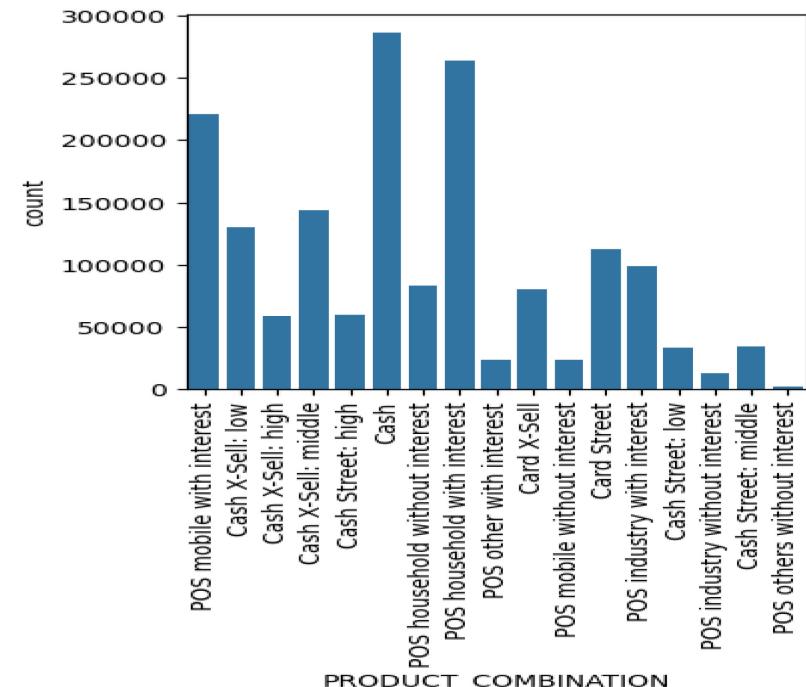
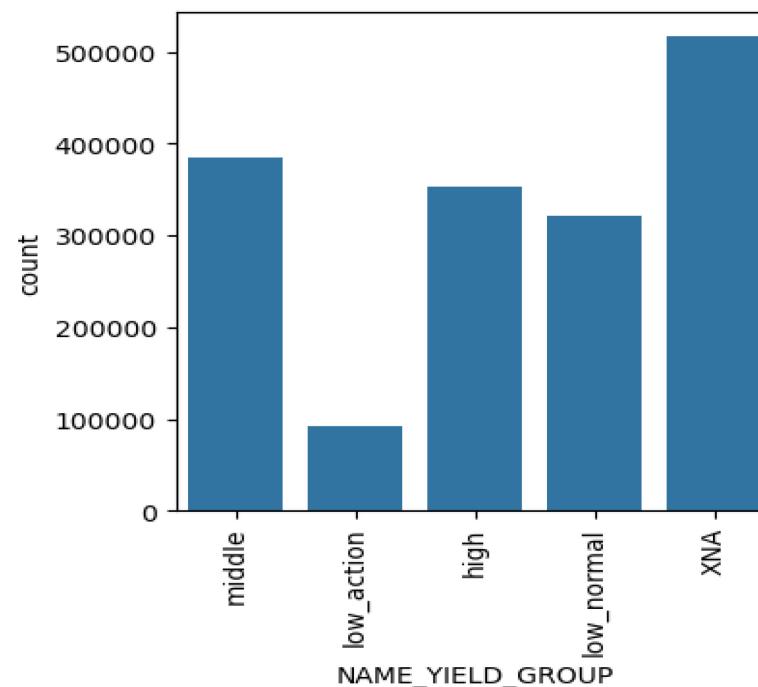
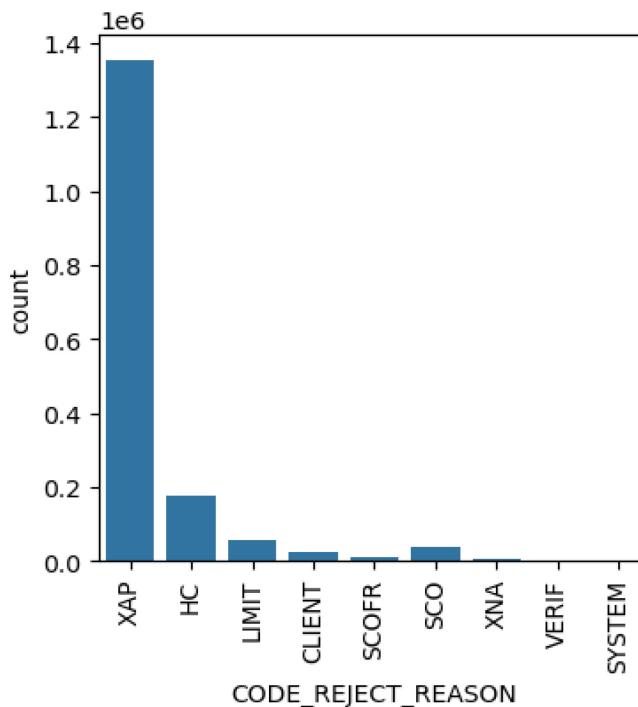
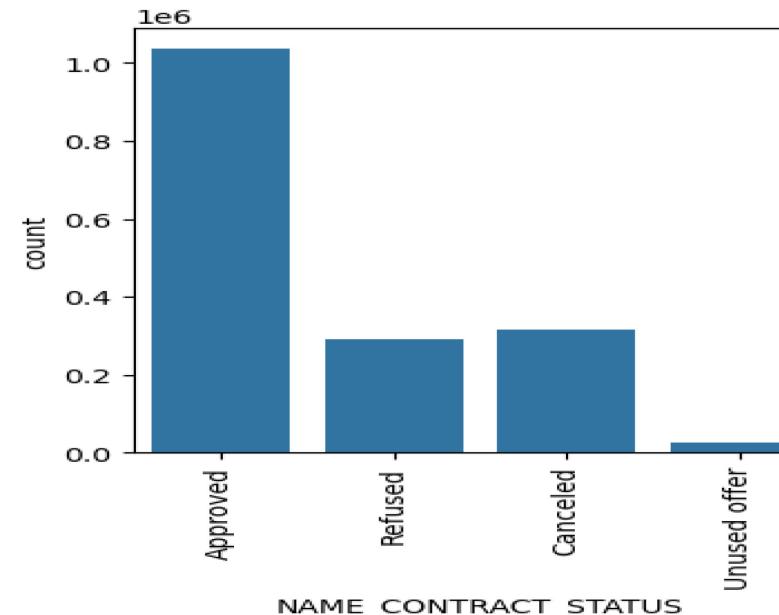
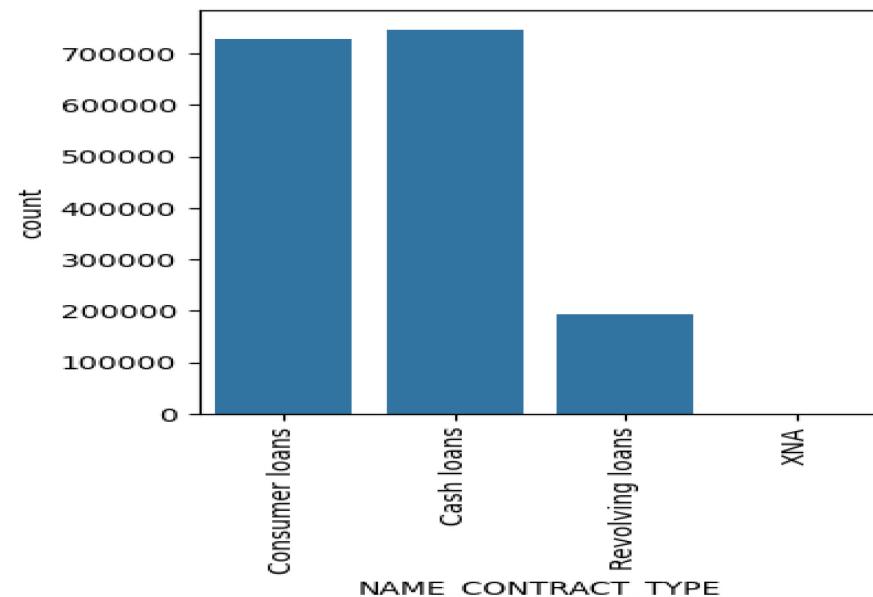
- Credit amount and Goods Price have strong Correlation (0.99)
- Credit Amount and Annuity have good correlation (0.77)
- Goods Price and Annuity have good correlation (0.77)
- Regional Rating and External Source score have highest negative correlation(-0.29)



Previous Application Data

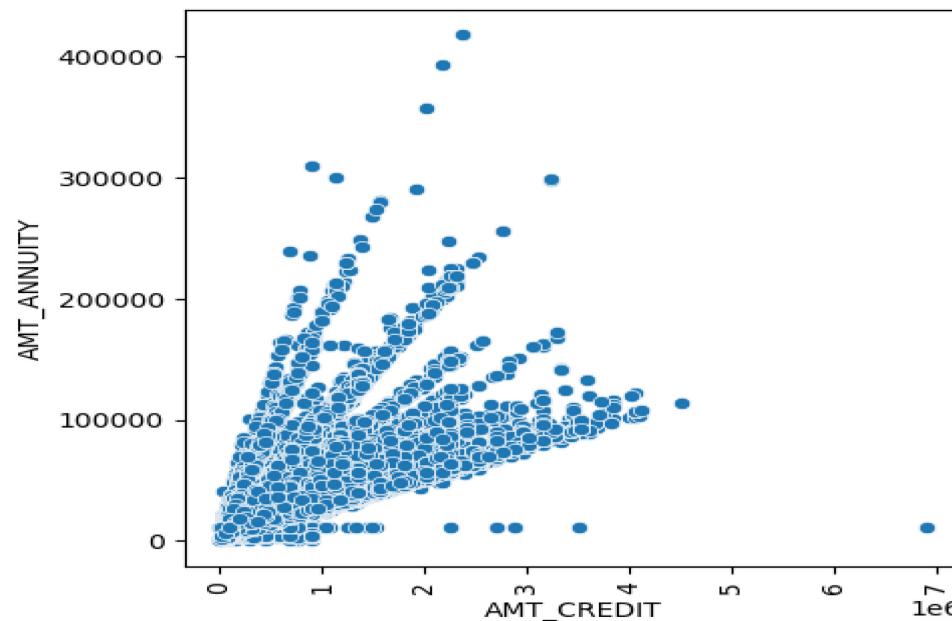
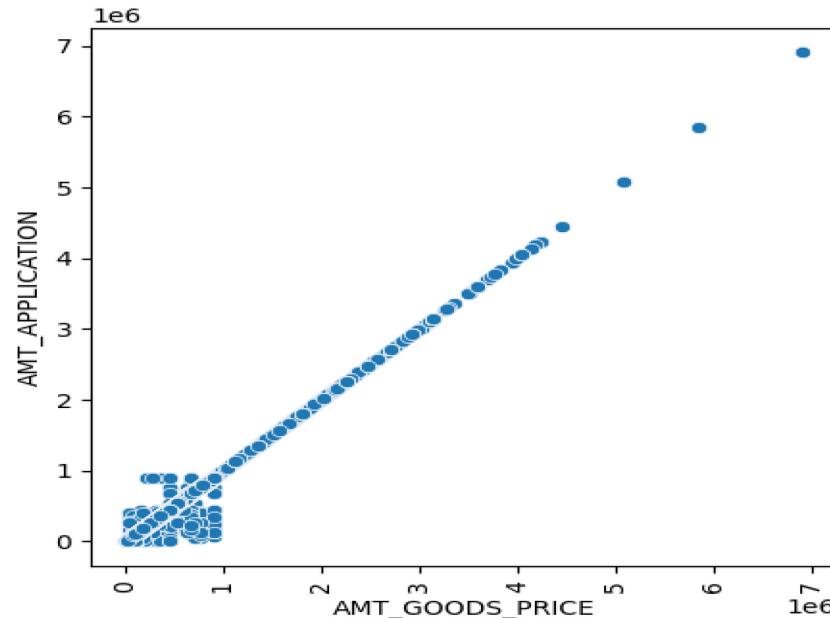
Key observations from Univariate Analysis of Previous Application Data

- Consumer Loans and Cash loans take a major part in previous application than revolving loans
- Most of the loans were approved with a few in Refused and Cancelled status while a very minute % of loan was unused offer
- Major reason for rejection of loan is HC-High Credit followed by Limit and Verification issue.
- Cash Loans and POS loans for Household were the maximum loans taken by the previous customers
- Interest Rate was mostly in middle range closely followed by high interest rates for Previous Applications



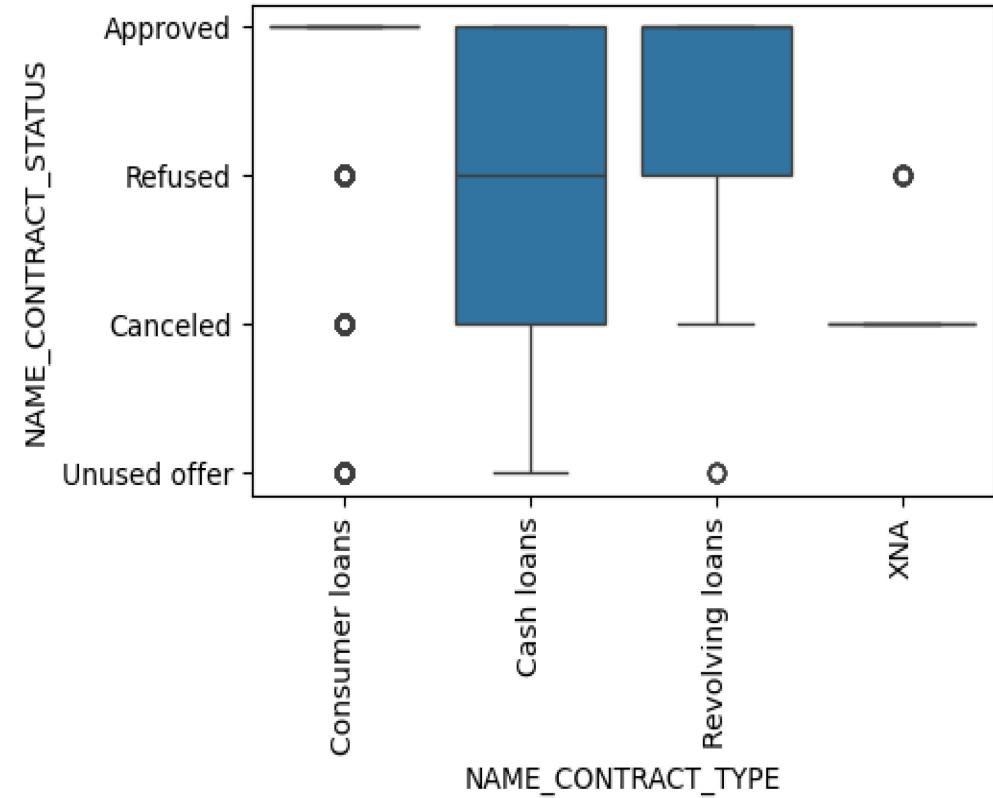
Bivariate Analysis of Numerical Columns

- Annuity amount and Credit Amount shows significant linear relation which implies most of the clients have received the amount quoted in Application
- Application amount and Amount of Goods Price also has a linear relation stating with increase of goods price, loan amount also increases

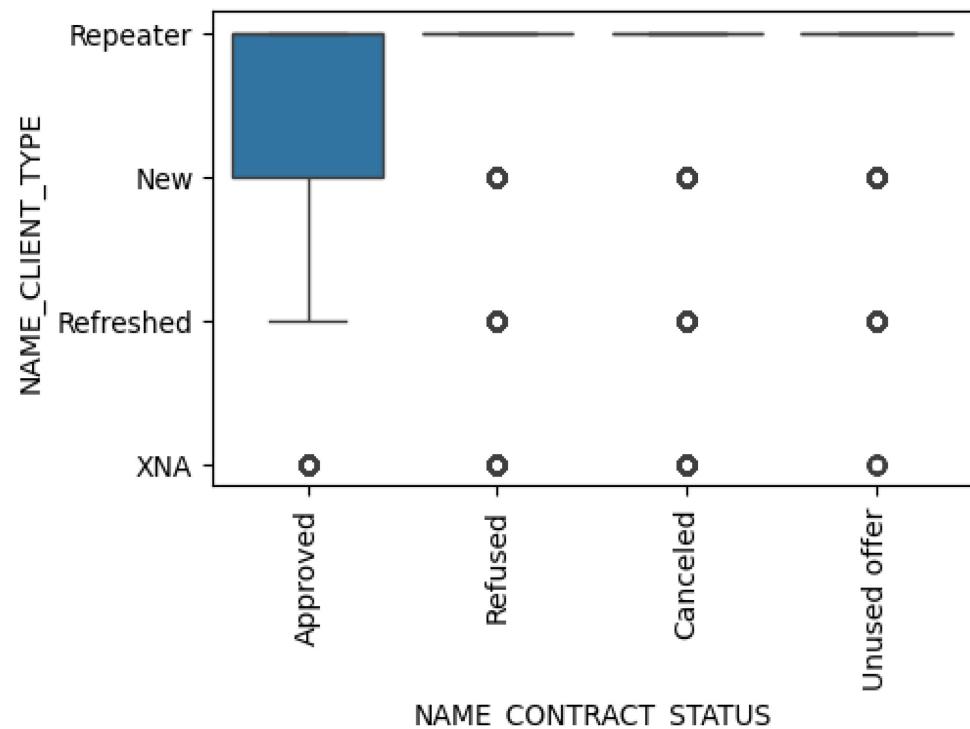


Bivariate Analysis of Categorical Columns

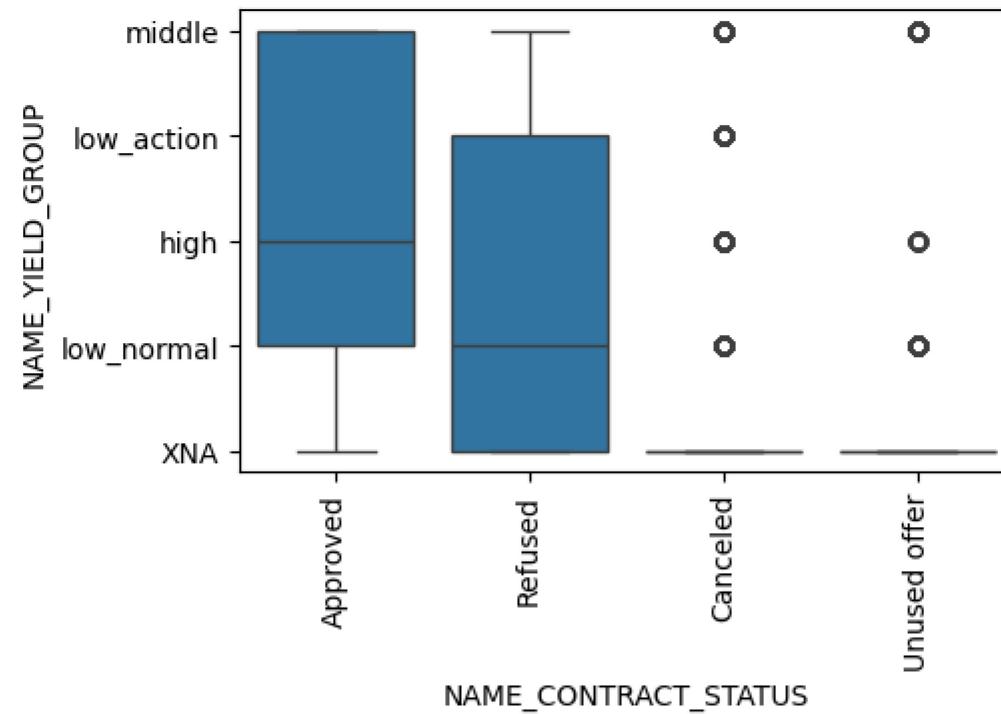
- Most of the Consumer Loans are approved with a very few Refused or Cancelled, whereas for Cash Loans IQR ranges from Approved, Refused and Cancelled which implies chances of Cash Loans Refused /Cancelled is equal to that of approval



- Loans were mostly rejected for the Repeated customers.

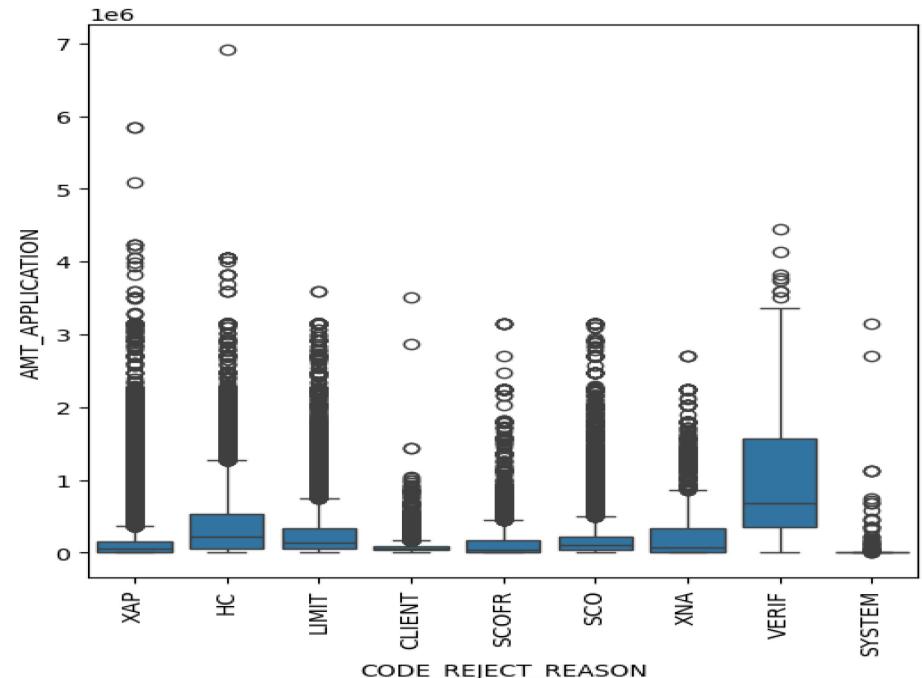
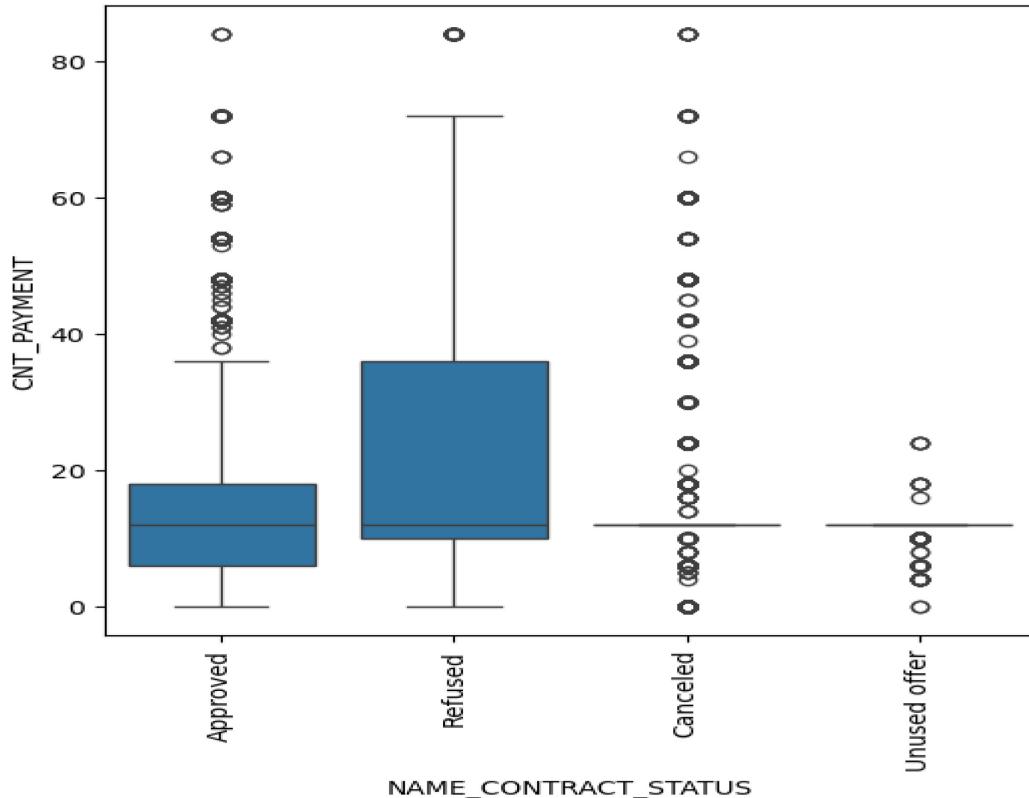


- Loans were mostly rejected for high and low-normal interest rates.

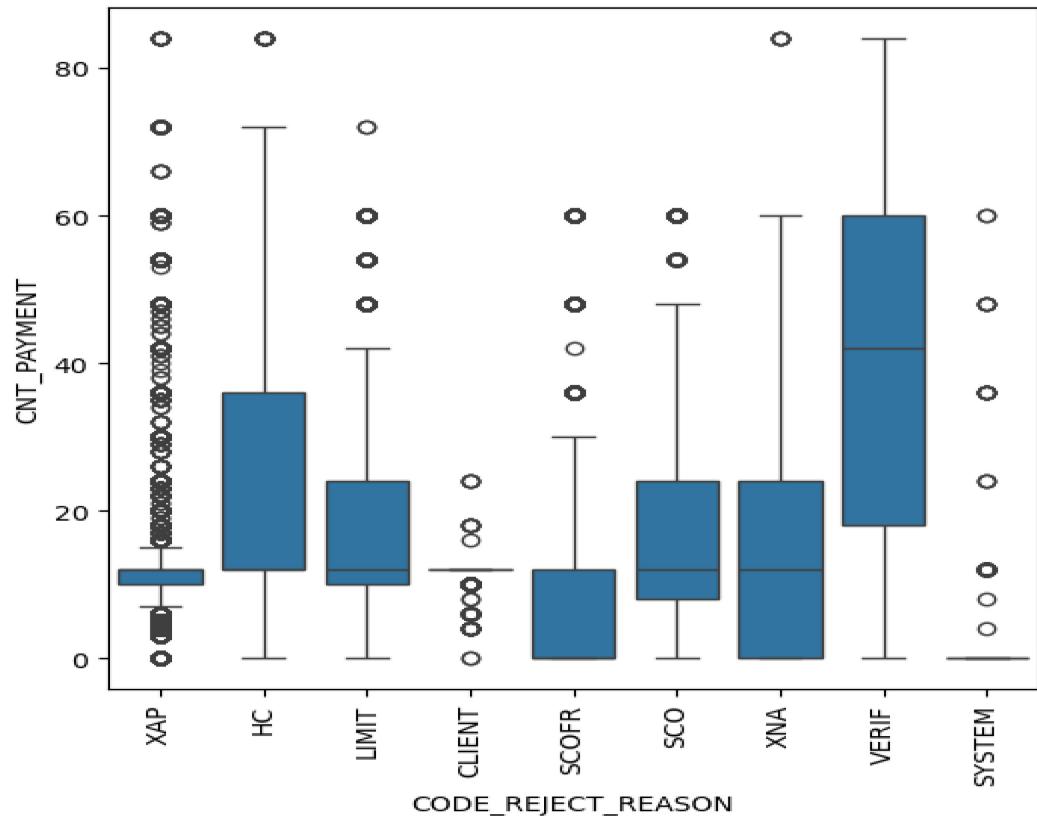


Bivariate Analysis of Numerical vs Categorical columns

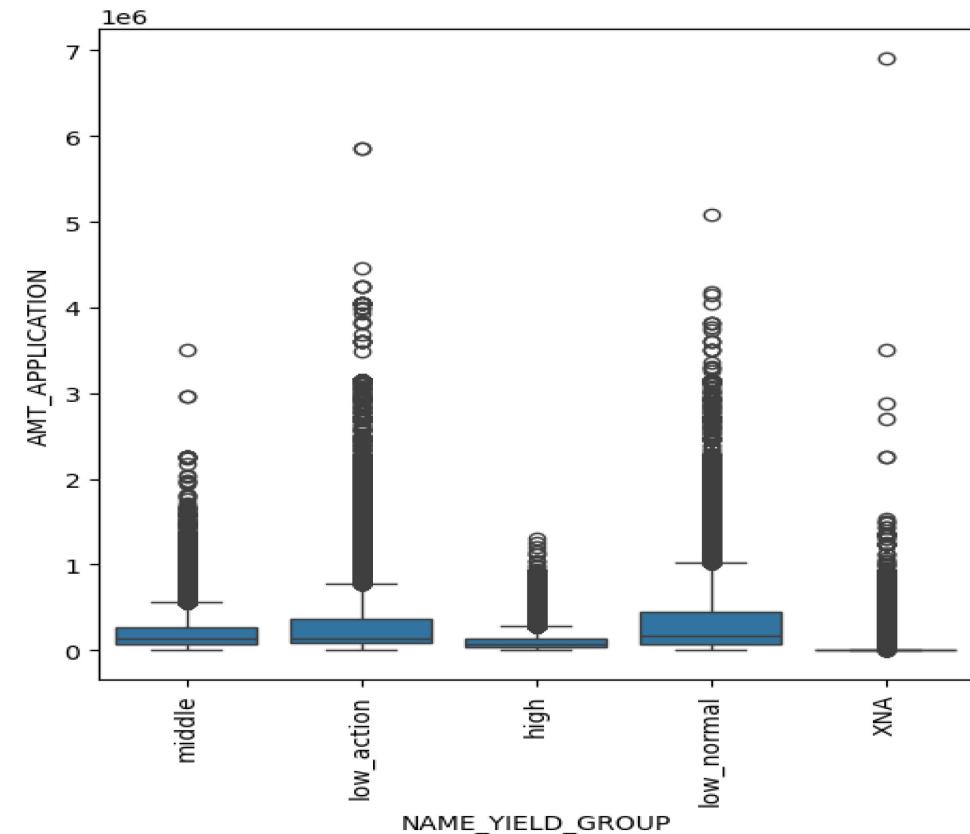
- We can see as the term of payment of previous application is increased chances of loan getting refused is high
- Loans Refused with Verification issues have considerably High amount loan in Application. i.e. they may not be able to meet the standard for loan repayment



- Loans refused with Verification and High Credit reasons have also high count of Repayment days for Previous application. i.e. They might have faced issues in repaying the previous loan.



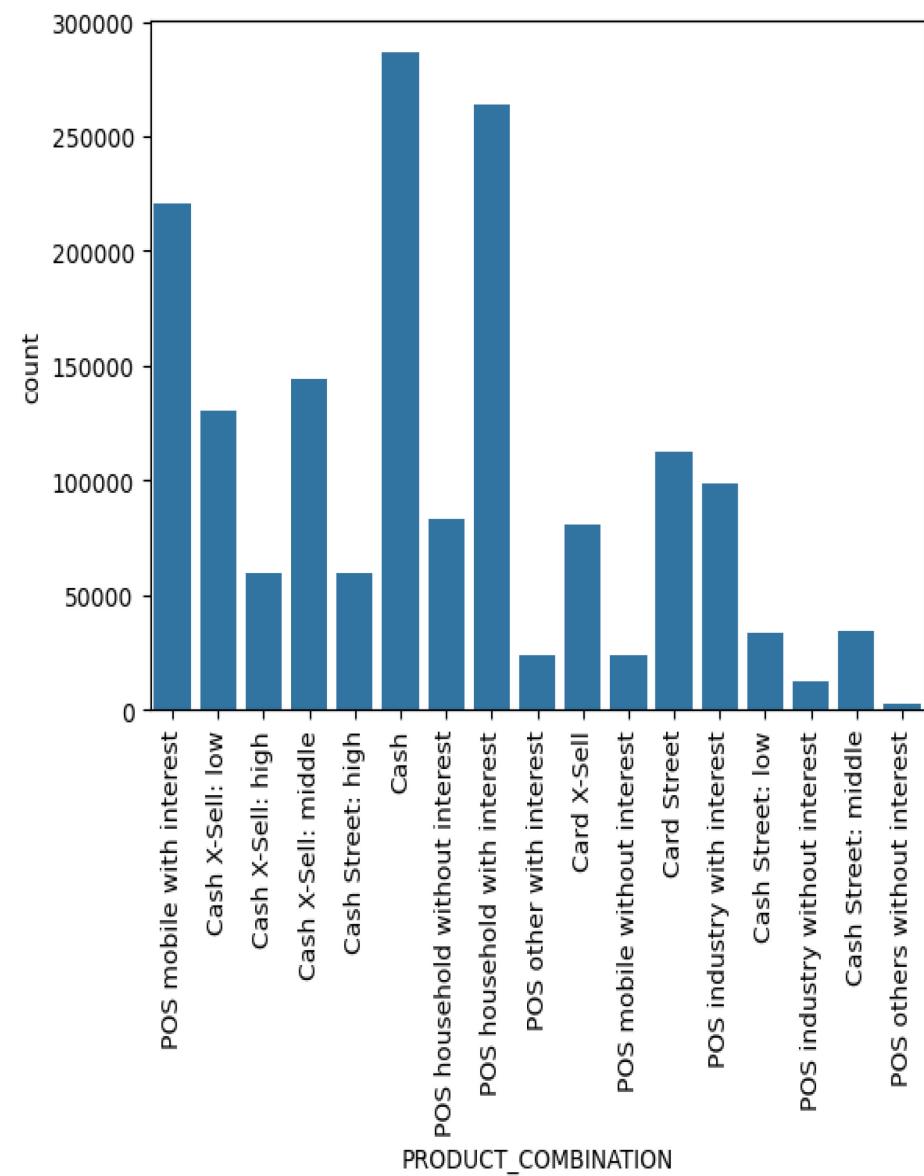
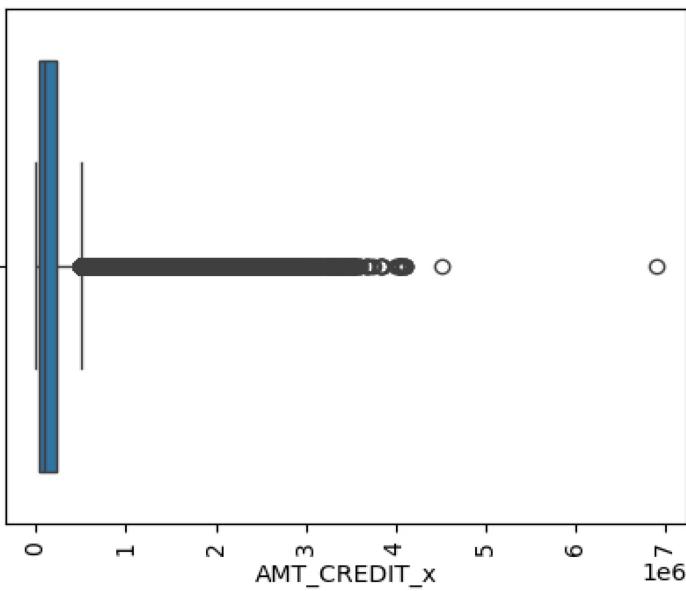
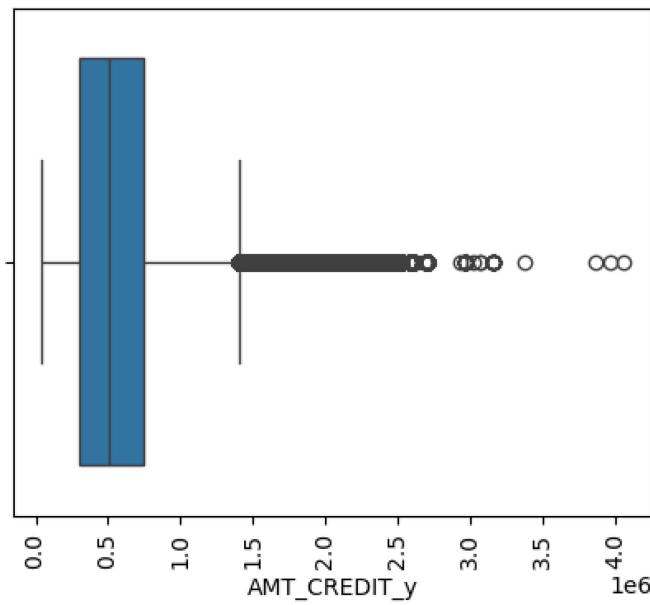
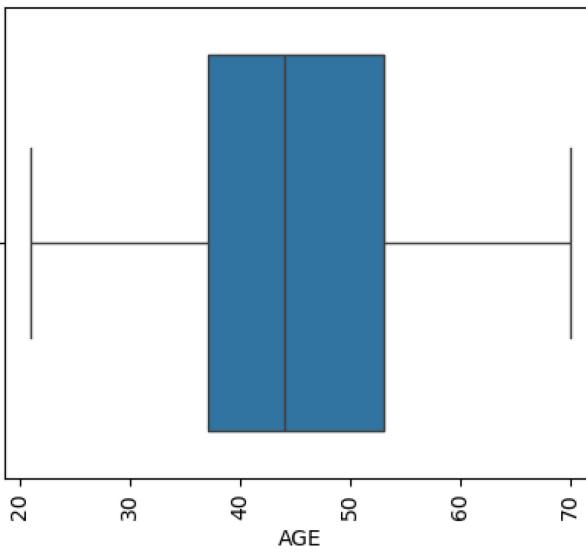
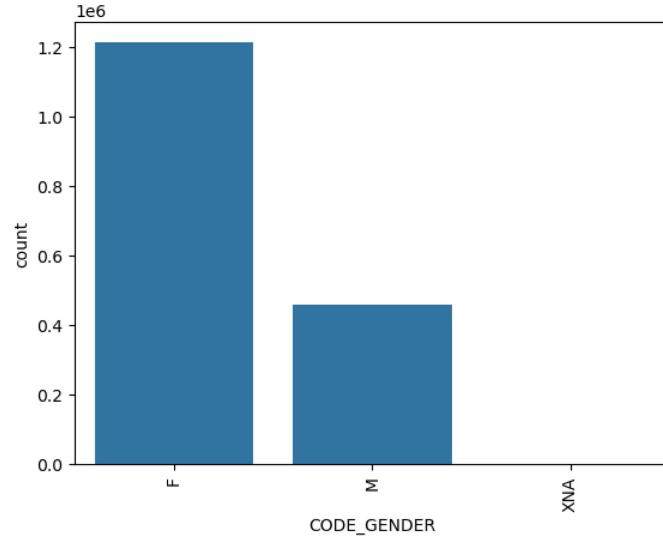
- We can see that when the Interest rate is high the amount applied for loan is significantly low than the interest rate ranging with low-normal, low-action and middle



Analysis of Combined Data Set

Key observations from Univariate Analysis

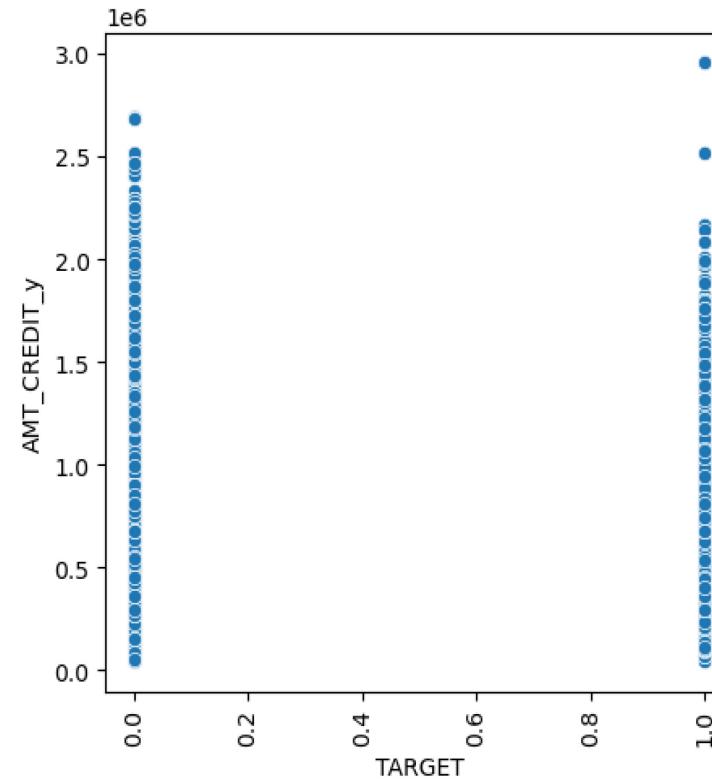
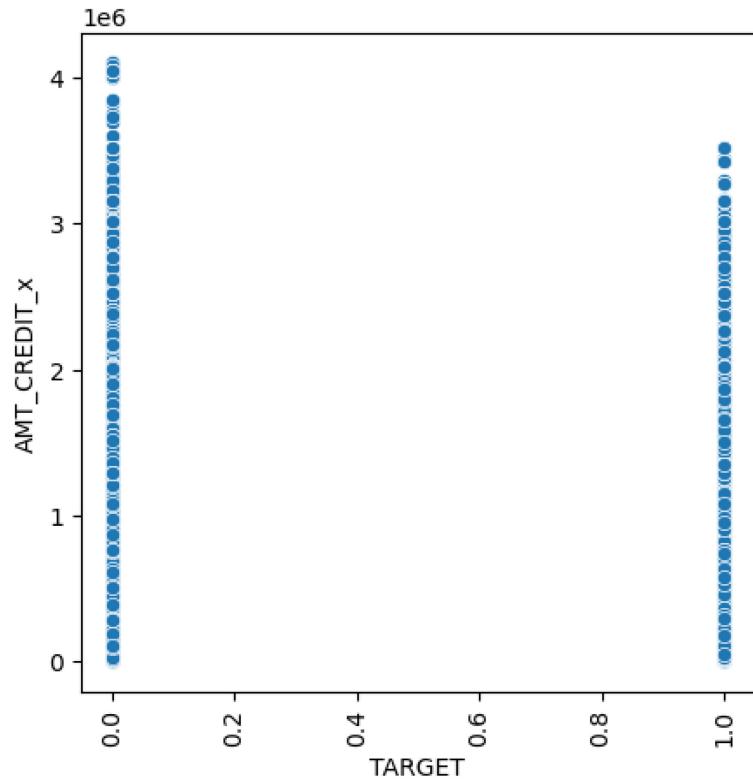
- Amount Annuity and Amount Credit are less for Previous Application compared to Current Application
- Loans are mostly from Mid aged clients, clients who are married and clients who are working
- Females are taking more loans compared to males
- Consumer and Cash loans were majorly taken for Previous application where as more Cash loans are taken for Current application
- POS Mobile loans and POS House loans with Interest take precedence in type of loans opted next to Cash loans



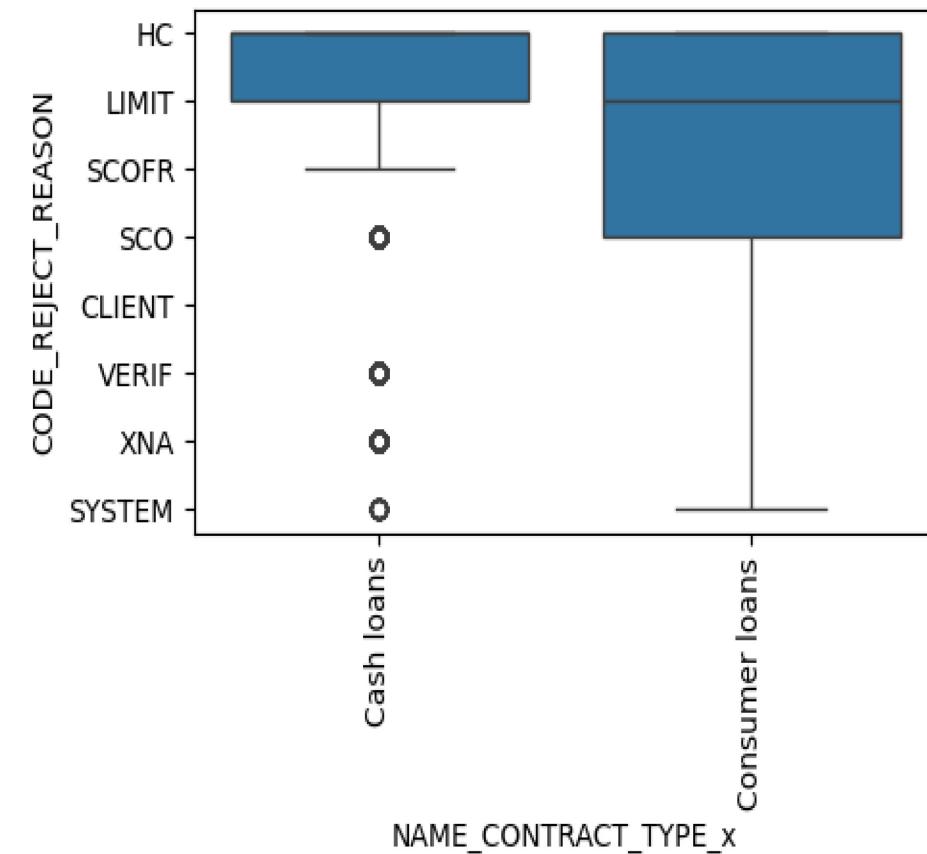
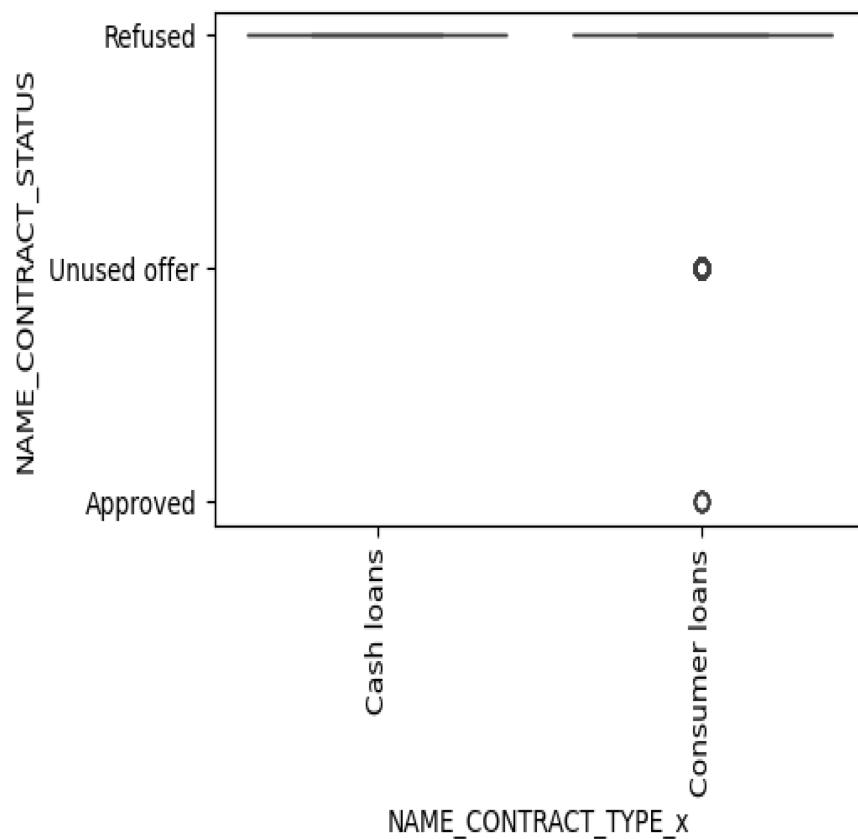
BIVARIATE ANALYSIS

Target vs Credit amount in Previous and Current Application

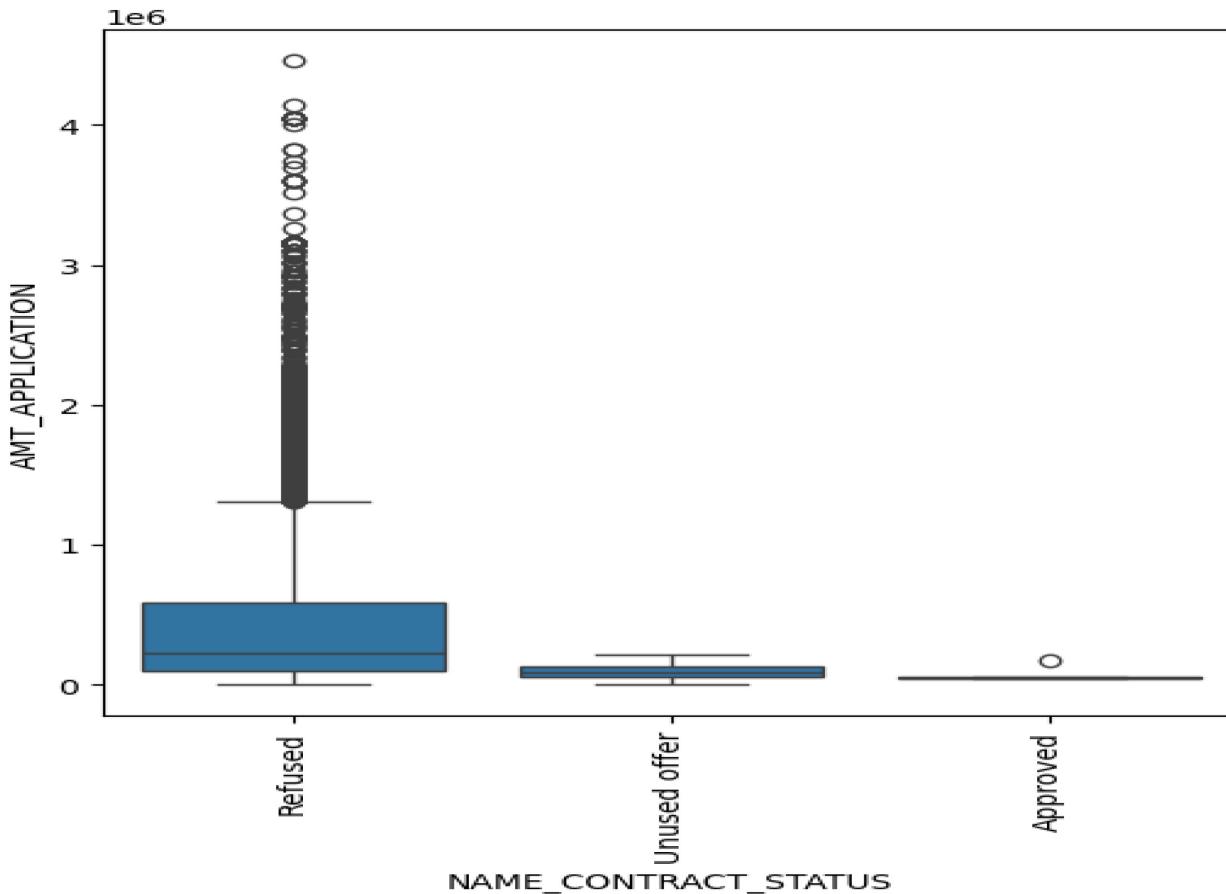
Most of clients who have applied for loan previously continue to have payment difficulties .



As we can see most of the Cash loans are refused and Cash loans are refused due to High Credit and limit reasons mostly.

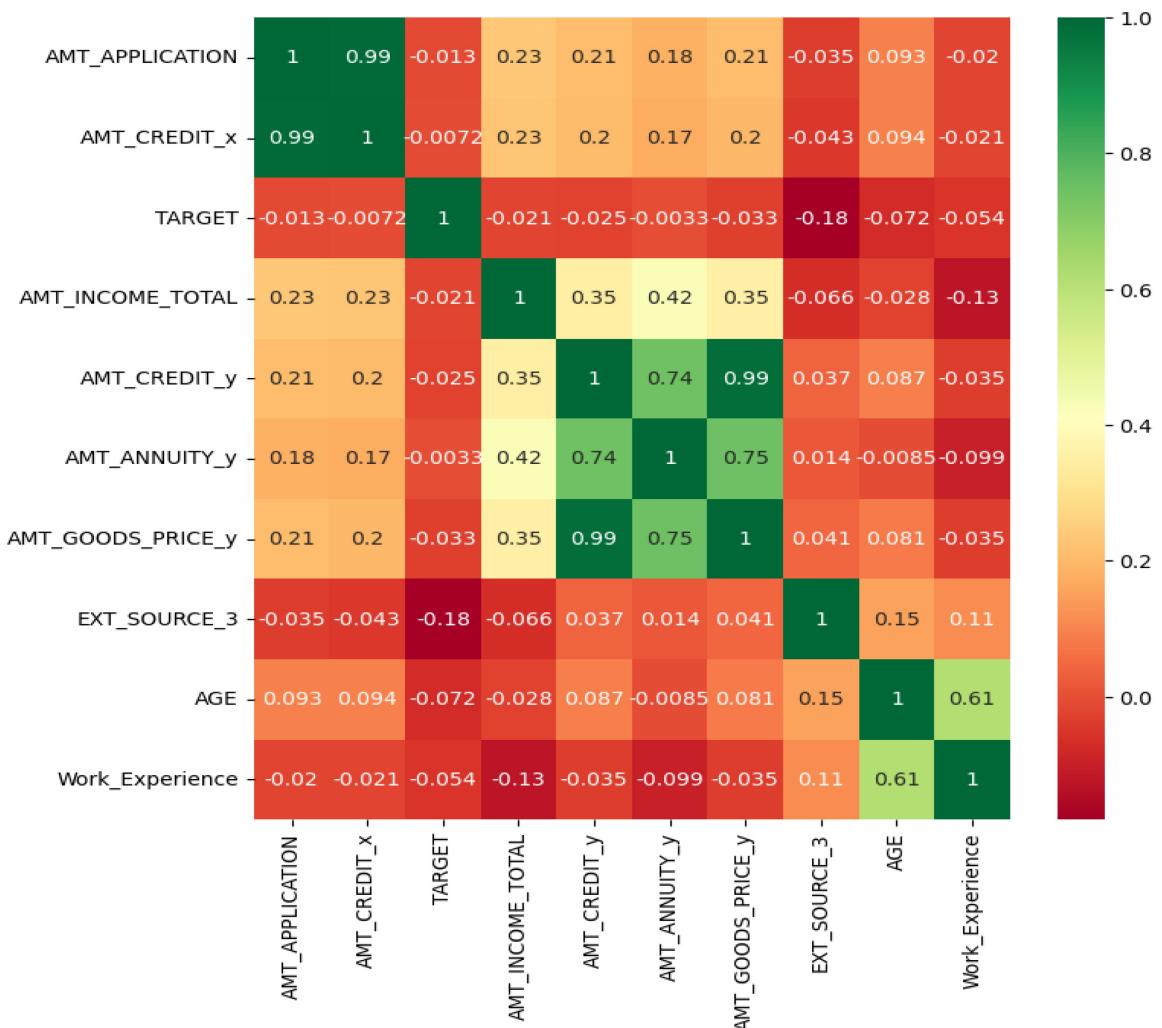


We can see loans which are Refused have very High application amount.



Multivariate Analysis

- Application amount and amount credit have Highest correlation
- Age and work experience have good correlation
- Annuity amount and Credit amount have good correlation
- External source info and Target have highest Negative Correlation
- Work Experience and Total income have negative correlation



Key Points from EDA of Credit

- **High Loan Application Amounts:** Clients applying for large loan amounts that do not align with their income levels are often at a higher risk of loan refusal.
- **High Credit Amounts:** Cash loans involving significantly high credit amounts are more likely to be declined.
- **Low-Risk Approvals:** It is generally safer to approve loans for clients with lower credit amounts, moderate interest rates, and reasonable payoff limits.
- **History of Loan Rejections or Payment Issues:** Clients who have experienced loan rejections or payment difficulties in the past are likely to exhibit similar patterns in new applications, categorizing them as high-risk for loan approval.
- **Potential Customers:** Female clients, married clients, middle-aged individuals, and those belonging to the working class are promising candidates for loan applications.