

# Credit EDA Case Study

Darshan Patel  
Manjari Joshi



# Problem Statement

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.

## Methodology to solve the case study

- Load dataset file using pandas function `read_csv()`
- Perform Data Quality Check i.e. checking for the datatype , shape and understanding of the data
- Checking Null values and drop the columns having >50% of missing data recommending imputation for columns having around 13% missing data.
- Checking whether the data is imbalanced or not. On that basis diving the data into two dataframe, Loan Defaulters and Non-Defaulters.
- Converting the continuous data to categorical by binning it for better analysis.
- Performing Univariate analysis for **Categorical variables and Continuous variables**
- Finding **Correlation between Continuous variables**
- Performing Bivariate analysis for **Continuous-Continuous variables, Categorical-Categorical variables and Continuous -Categorical variables**

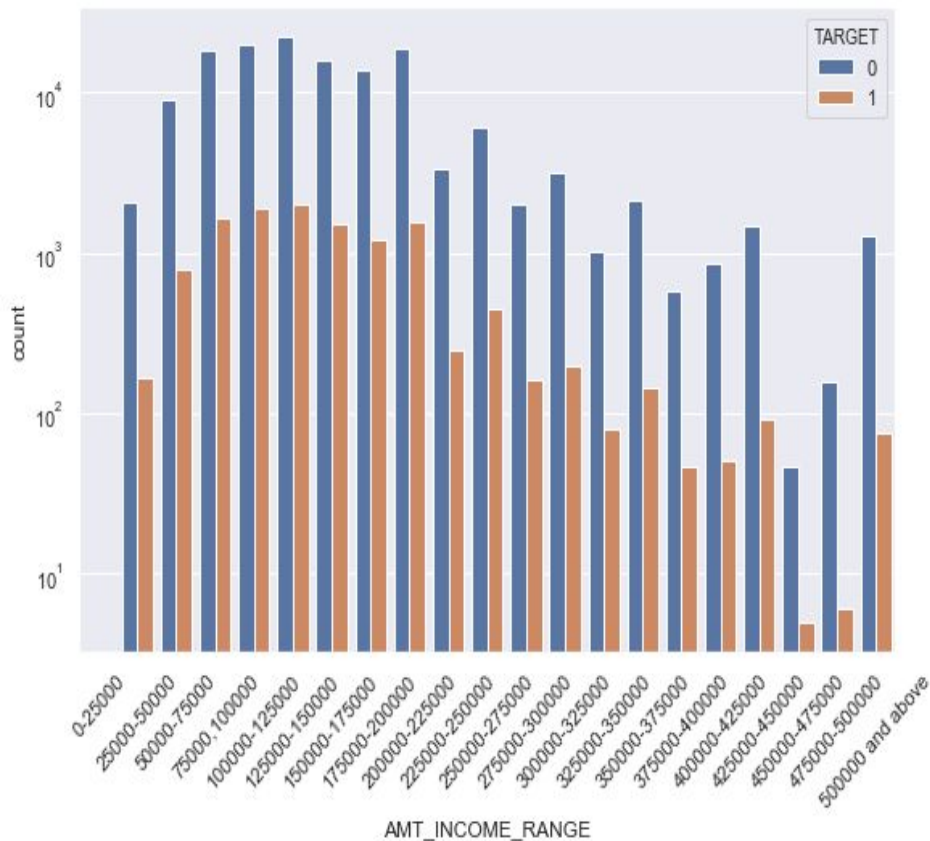
# Univariate Analysis

## Impact of Income on loan payment

From this graph one can infer that the applicant having 100000-125000 of income are having higher number of loan and can have higher chance of defaulting a loan.

Contradictory to that the income segment of 450000-475000 are having less chances of defaulting a loan.

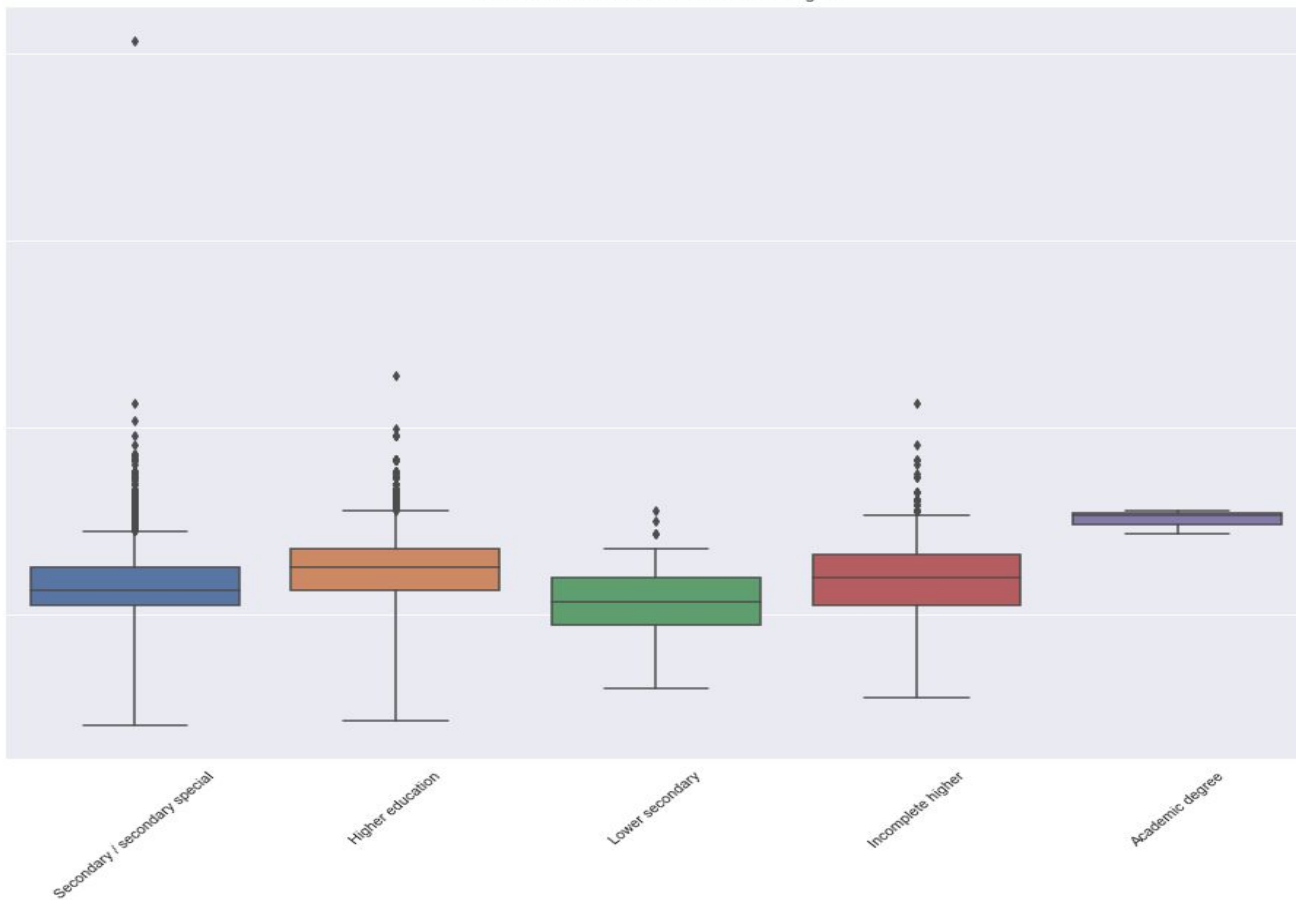
From this we get to know that people with less income but higher credit will default on their loan



# Bivariate analysis

## Credit amount vs Education Type

Income amount vs Education Status for Target 1

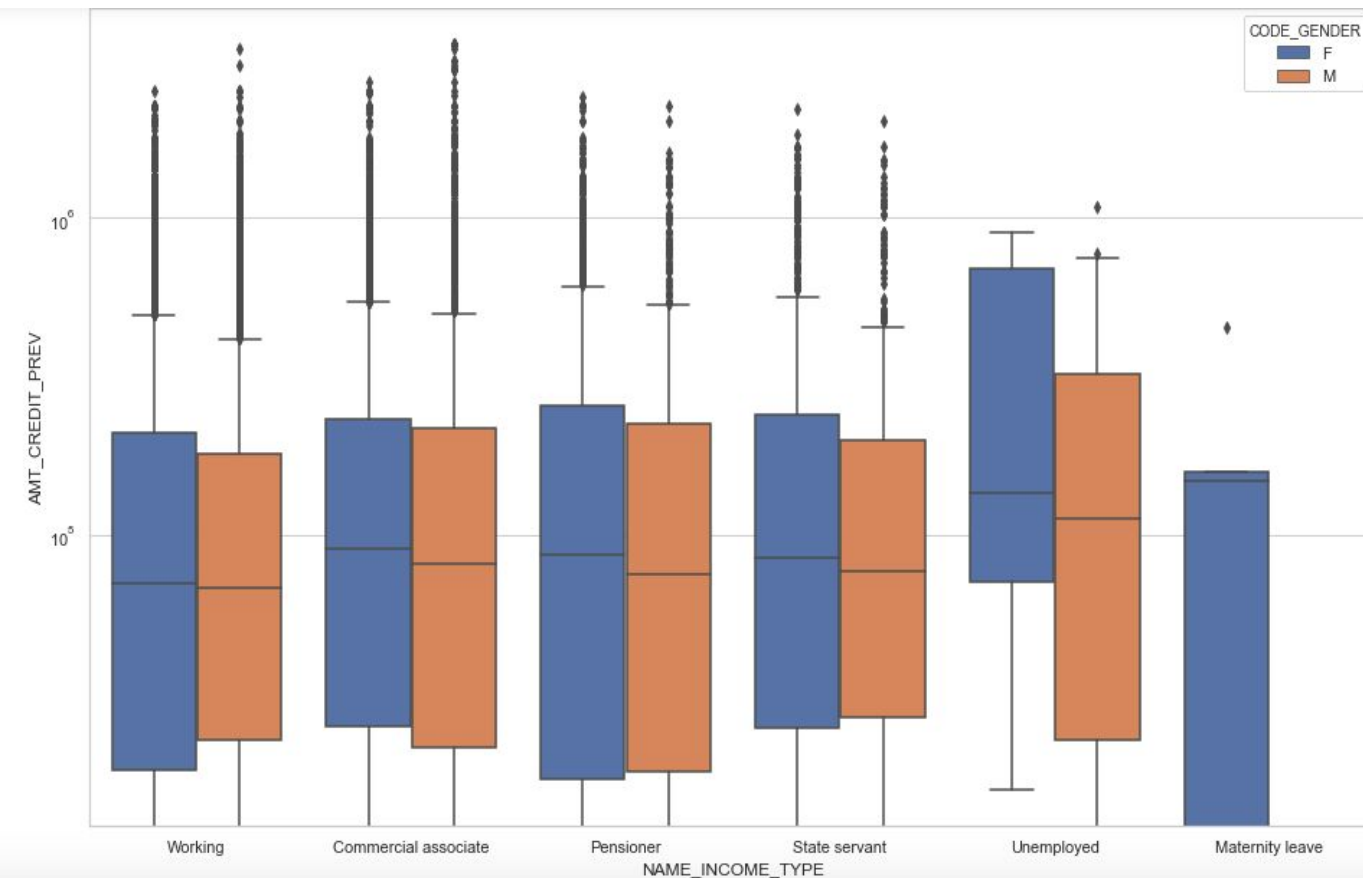


- There are no outlier for Academic degree and there income amount is also little higher than other education type.
- Education type 'Secondary/special education has lot of outliers.
- Lower Secondary has less amount of total income whereas incomplete education has more total income from lower secondary.

This tells us higher the education degree an application have less are the chance of an applicant to default on loan repayment

# Bivariate analysis

## Gender w.r.t. Prev amount Credited & Income Type

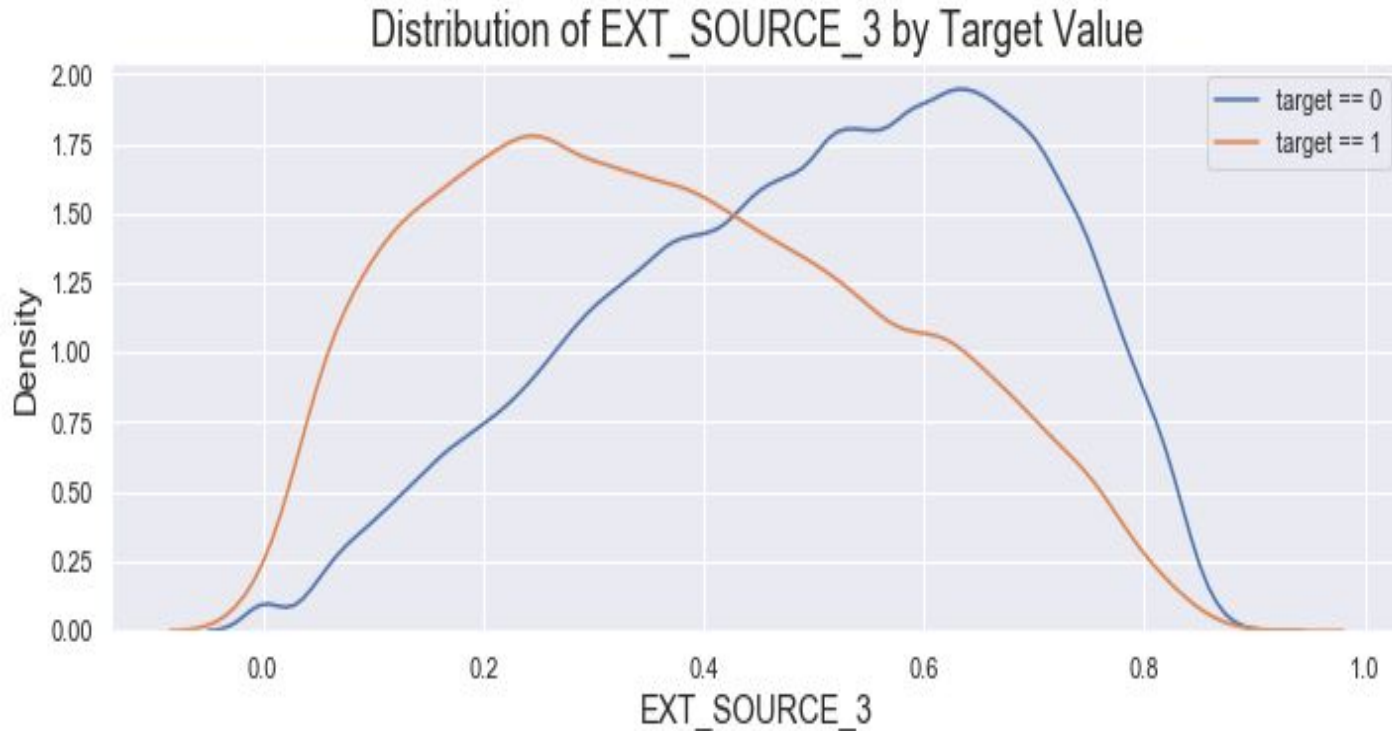


Maternity category- FEMALE is significantly higher problem in repayment.

Unemployed Female category also shows more defaulting rate.

Also, Commercial Associate and Working has the many outliers.

# Impact of EXT Score on Loan Repayment

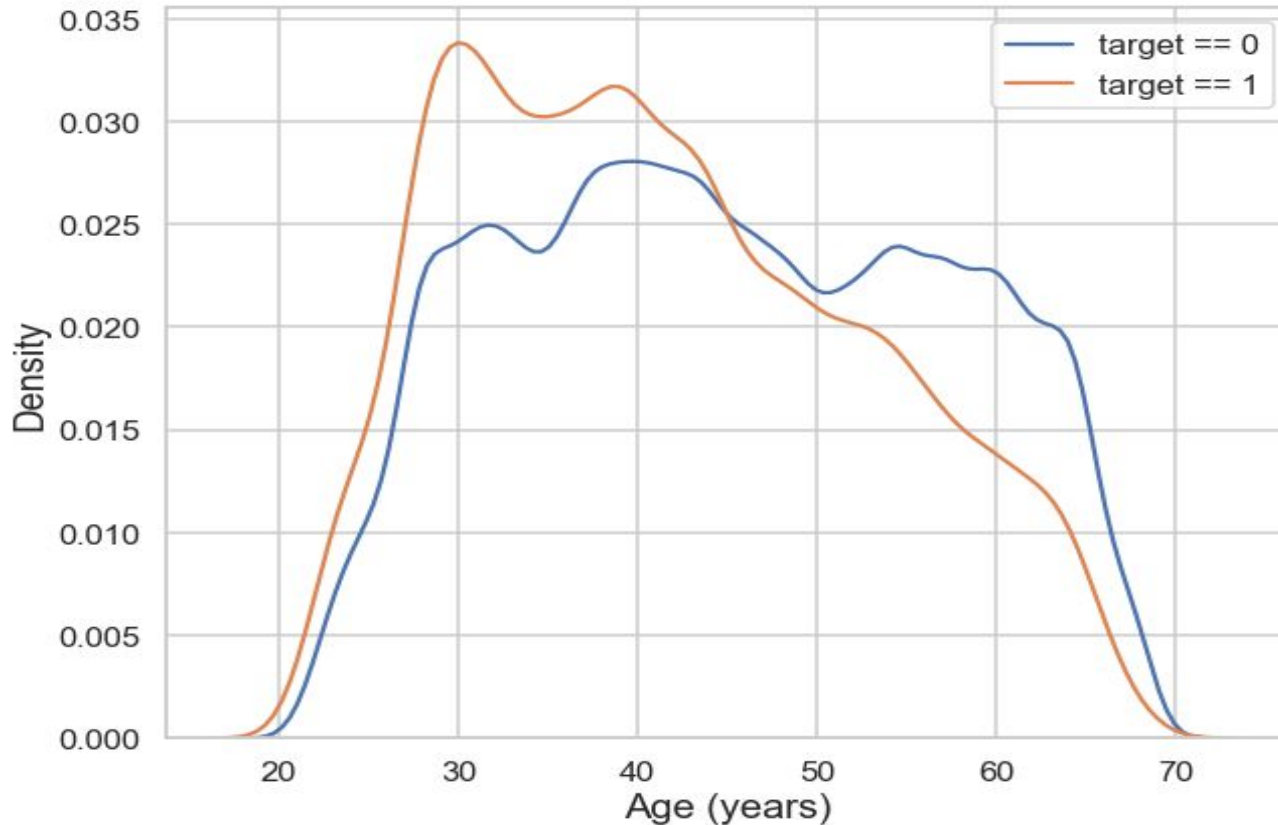


\* EXT\_SOURCE\_3 displays the greatest difference between the values of the target.

\* It can be clearly observed that with small value of EXT\_SOURCE\_3 higher is the defaulter rate. And as the EXT\_SOURCE\_3 value increases the rate of client becoming defaulter decreases.

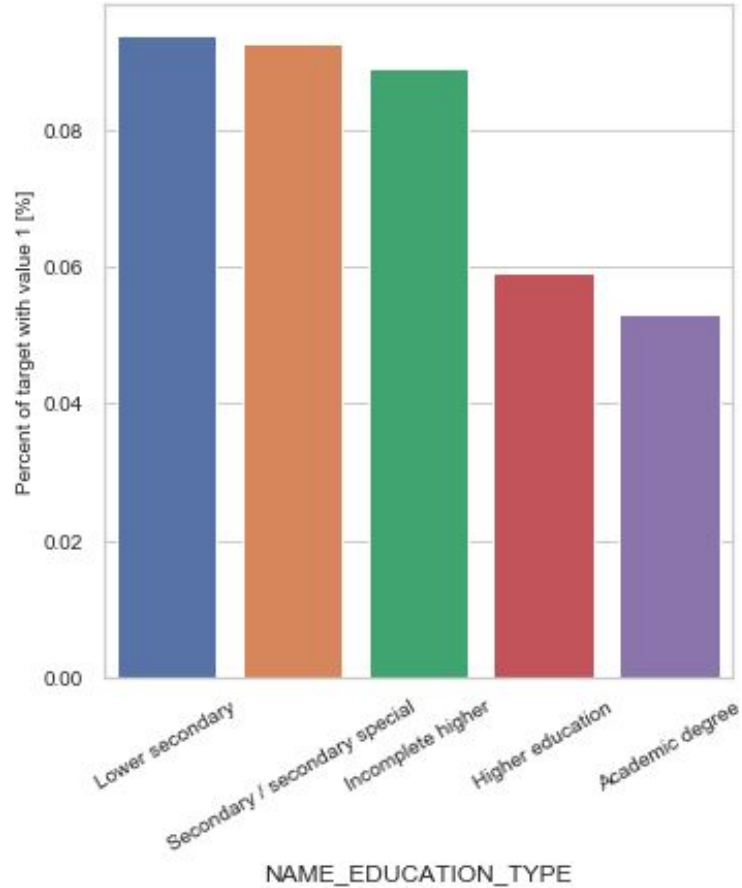
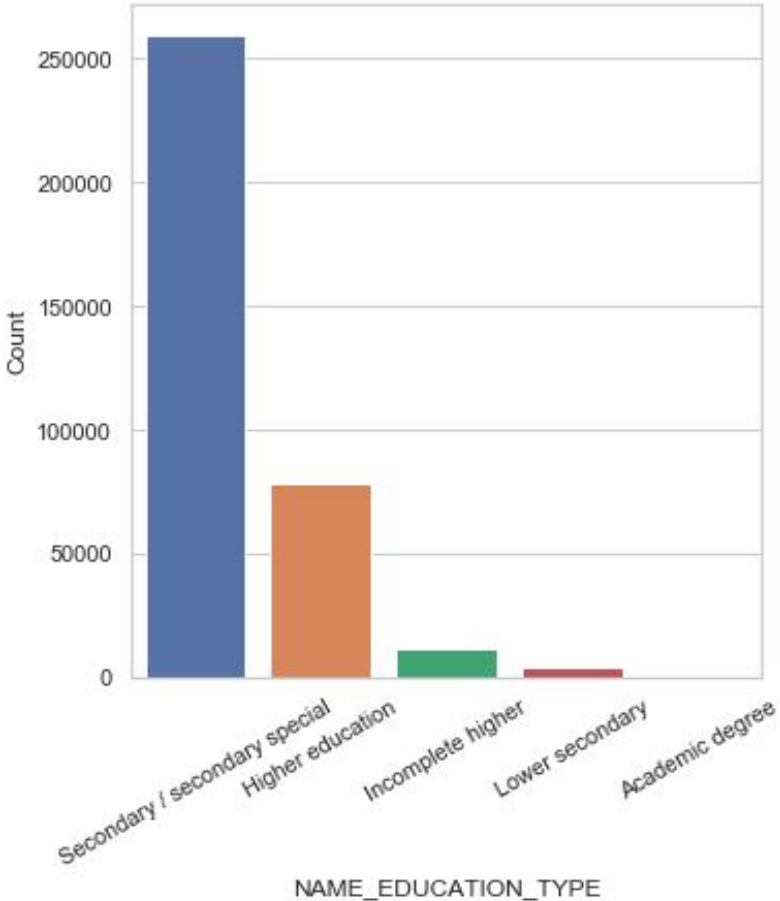
# Impact of Age on Loan Repayment

Distribution of Ages



Age is having negative linear relationship with the target means as the client gets older they tend to repay their loans on time more often. However, younger clients tend to default the most.

# Impact of Education on Loan Repayment

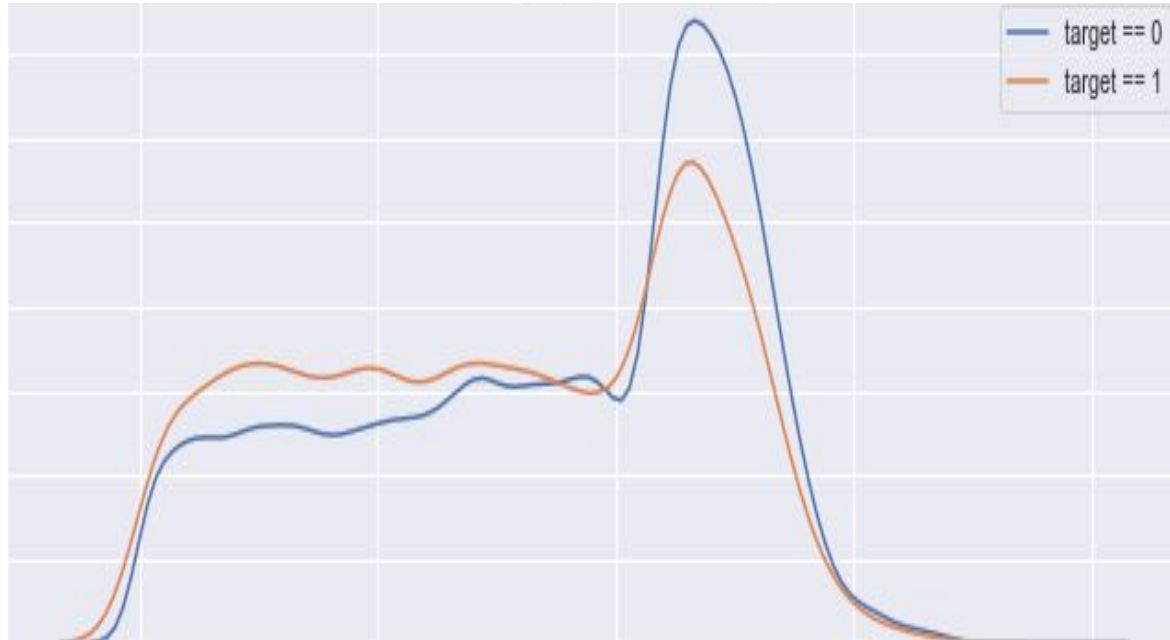
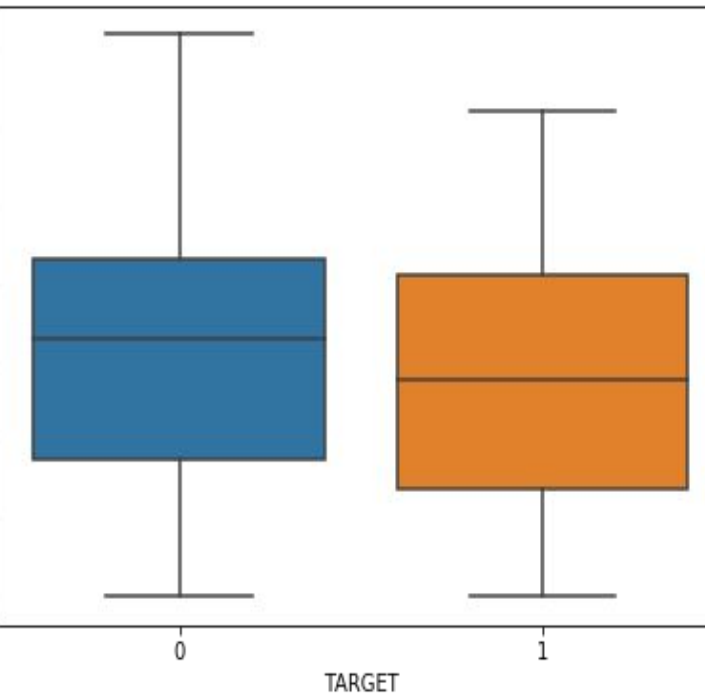


\* Majority of the clients have Secondary / secondary special education, followed by clients with Higher education. Only a very small number having an academic degree.

\* It can be clearly observed that the Lower secondary category, although rare, have the largest rate of not returning the loan. Whereas people with Academic degree have the least defaulter rate.

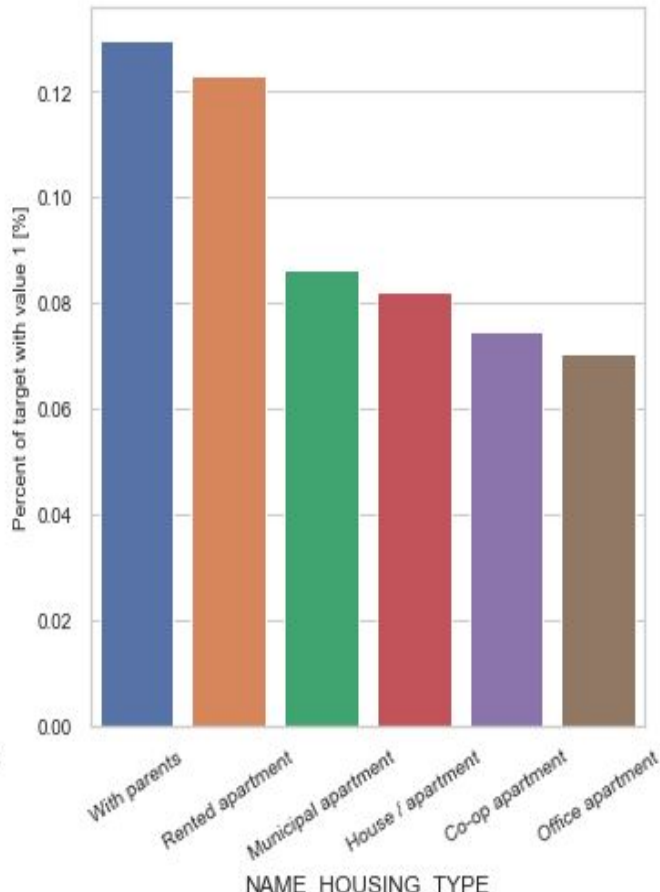
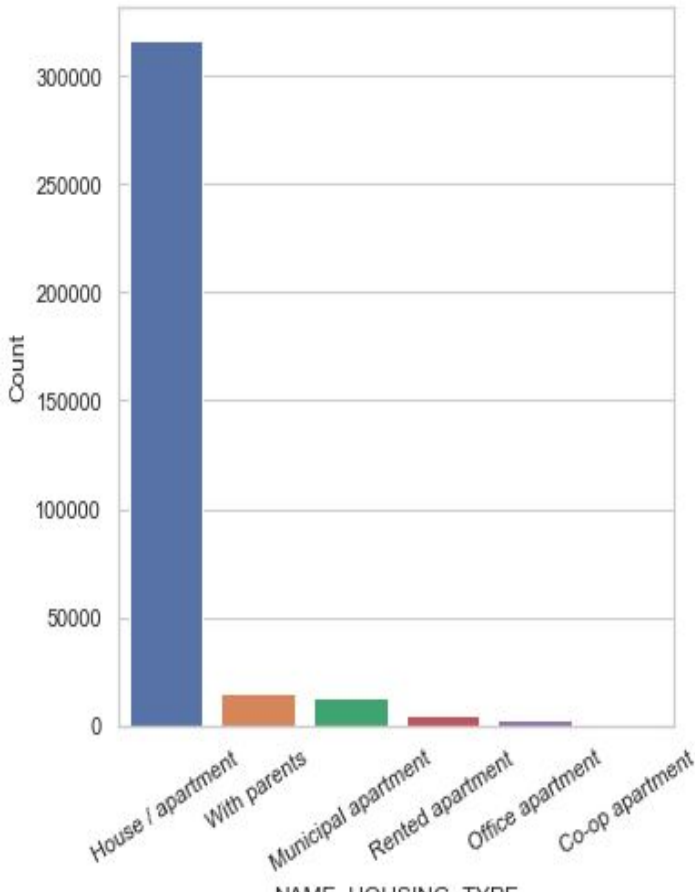


# Impact of DAYS\_ID\_PUBLISH on Repayment



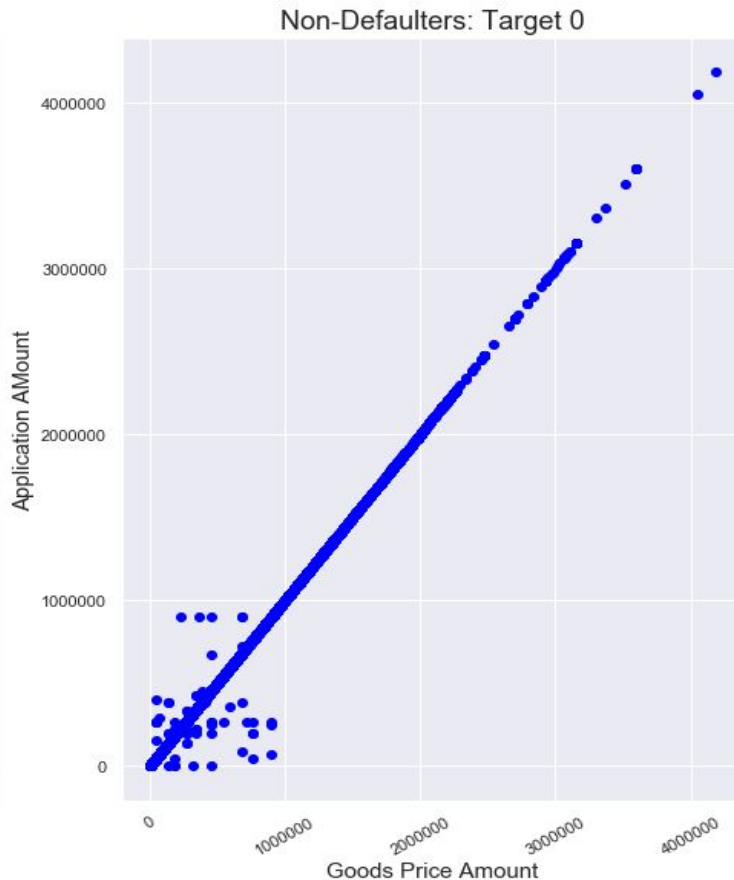
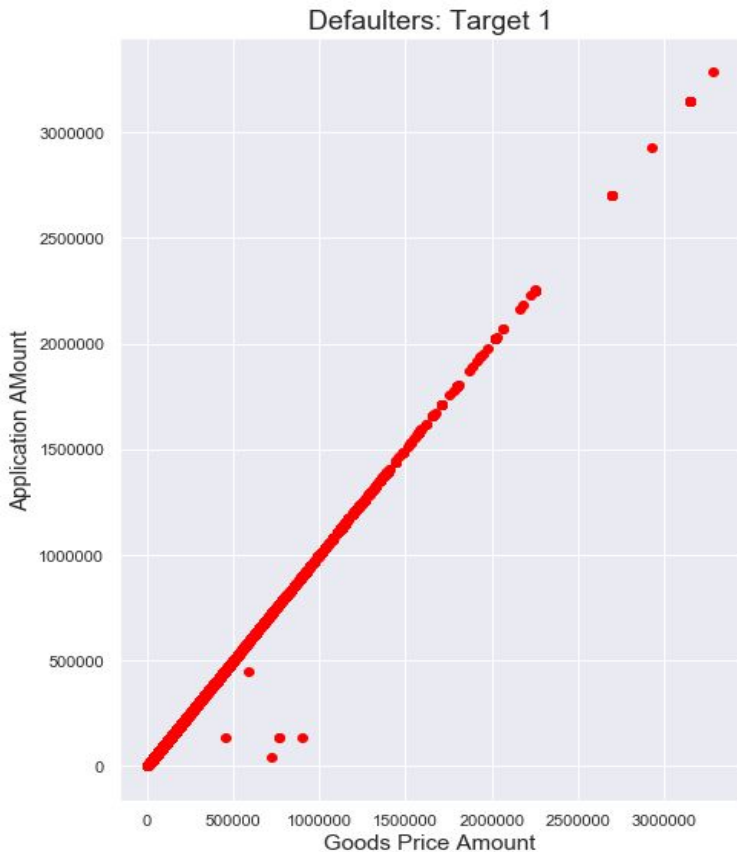
Although the data is imbalanced but we can clearly observe that clients who have recently changed their identity documents before applying for a loan have a higher rate of defaulting to repay the loan amount.

# Impact of Housing type on Loan Repayment



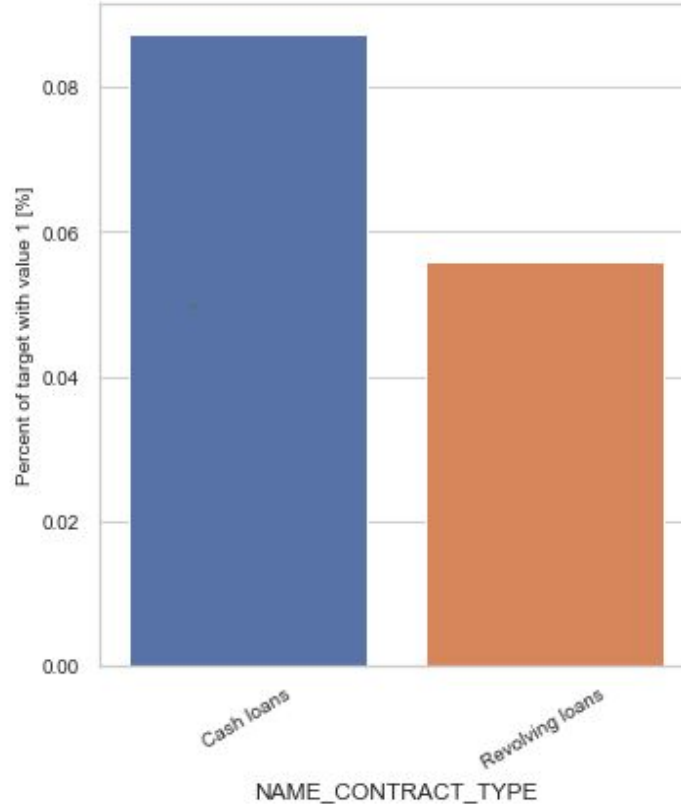
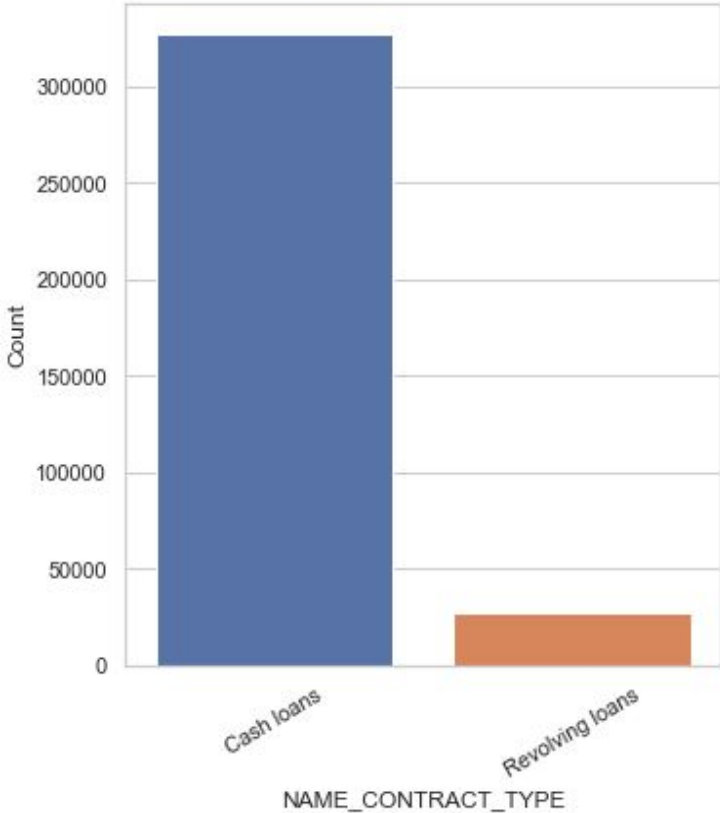
Although House/apartment are the maximum clients who applied for loan. But Clients who are staying with the Parents or are staying in Rented Apartment tend to show the highest percentage defaulting. Rate.

# Goods Price & Application Amount



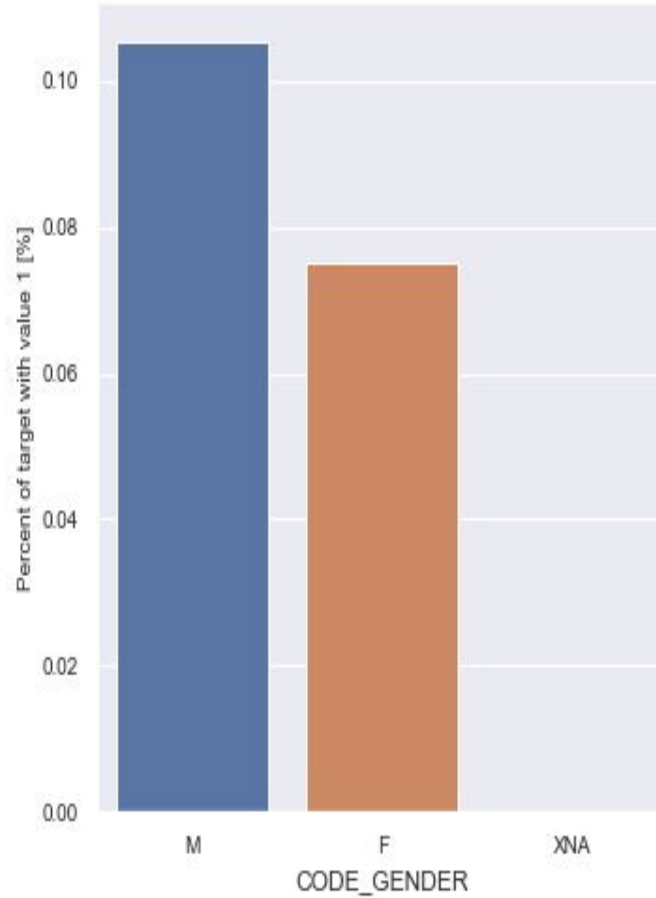
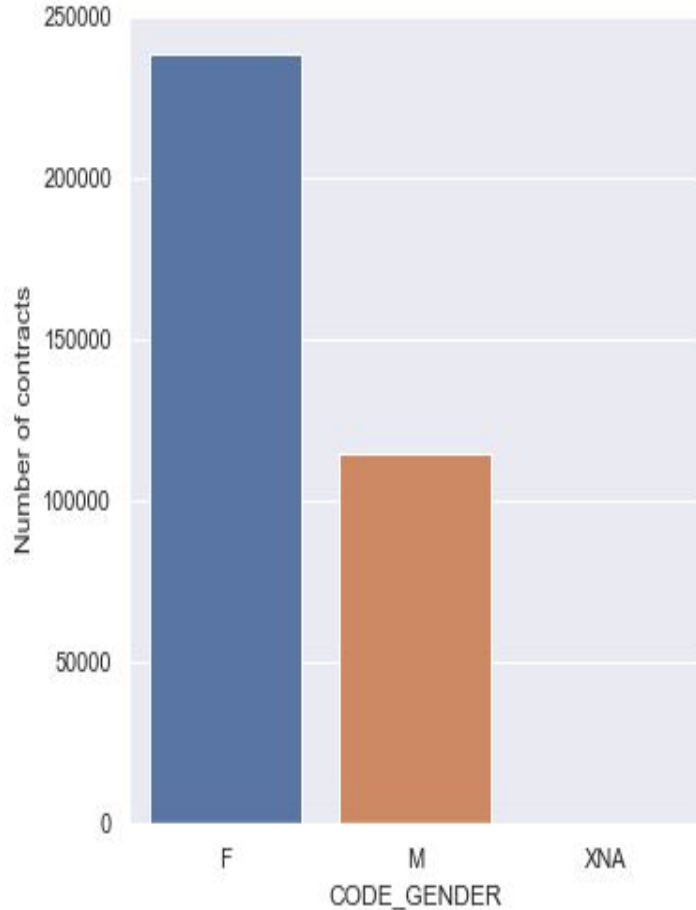
Goods price amount is positively correlated with the Application amount, meaning the application amount i.e. amount asked by client for the loan depends on the amount of goods for which client has applied. If the goods price is high then client will apply for higher amount of loan or vice versa.

# Impact of Contract type on Loan Repayment



It can be observed that clients who took Cash loans tend to have high defaulting rate compared to the clients who opted for Revolving loan type. So, bank should prefer giving Revolving Loans to the clients.

# Impact of Gender on Loan Repayment



Although the number of Female applying for loan is more but it can be clearly observed that percent of defaulters are more in Male than Female.

# Conclusion/Recommendation:

1. Banks should look into where is the applicant based and verify the permanent living address, work address and temporary address this could eliminate the red flag of whether the applicant can default on this loan or not.
2. Through our analysis of different variables of both the data set. We recommend the bank to look at variables such as Housing, Amount of income, Education and work type as a primary point of elevation of the loan application.
3. From our analysis, we have found an applicant who has Rented Apartments tended to default on their loan repayment. From education, the prospective applicant who has secondary education tend to default on their loan repayments.
5. The bank should consider the Income of a loan applicant and can check with the amount of credit the applicant has. This factor will tell the bank whether the applicant is able to repay the loan amount or not.
6. Ext Source displays the greatest difference between the values of the target. It can be clearly observed from the graph that with a small value of Ext Source higher is the defaulter rate. And as the Ext Source value increases the rate of the client becoming defaulter decreases so the bank should also check the Ext source.
7. Bank can also focus on the cash loan purpose as it will help to determine whether the application will be default on the loan repayment or not.
8. Bank can look the pervious loan application can check it the applicant has good repo with banks in terms of its transactions and how loyal is with the bank and how long he has been maintaining the account with the bank.