Bank Loan Risk Analysis

Credit EDA Assignment

-Shyam Sunder Balaji

Problem Statement

When a bank receives a loan application, two broad types of risks are associated with the bank's decision for approving/disapproving the loan:

1. The loan is approved but the customer regularly defaults on the monthly payments
2. The client has a financially stable background but the loan is disapproved

Both these cases can incur massive losses to the bank. Therefore, this assignment aims to identify the patterns and factors which strongly indicate a tendency to default.

Assumptions and Approach

- Merge application_data.csv and previous_application.csv to ease the data cleaning process

- Remove columns with over 40% null values to avoid poor analysis

- Assuming ‘XAP’ and ‘XNA’ values are null, replace them with numpy.nan throughout the dataset

- Choose possible indicators of default and separate them as continuous or categorical

Assumptions and Approach

- For categorical variables, impute the null values with the mode
- Handle columns with negative values in the continuous variables
- Cap outliers at a threshold to avoid poor analysis
- Analyze the TARGET with respect to each of the chosen variables

Data Analysis

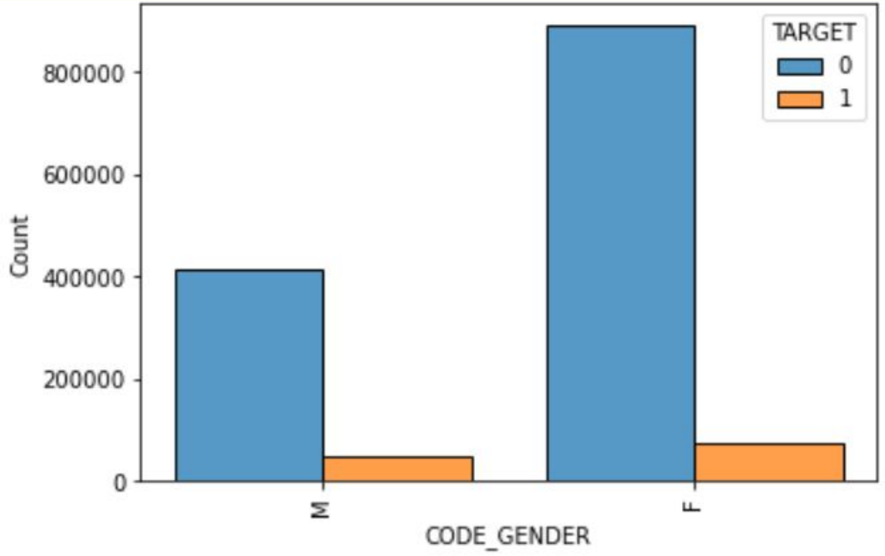
Balance of Target Data

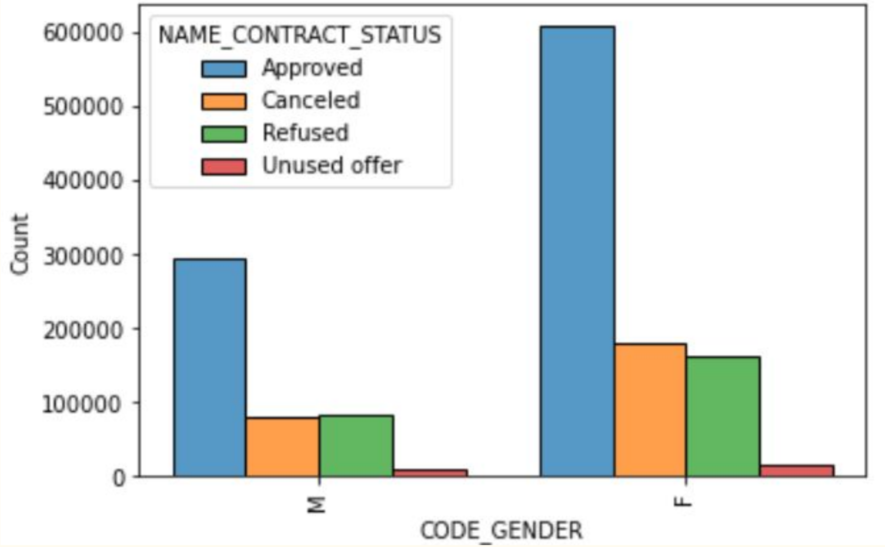
```
In [15]: df3['TARGET'].value_counts(normalize=True)

Out[15]: 0    0.913758
         1    0.086242
         Name: TARGET, dtype: float64
```

- A clear imbalance is observed in the data with only ~8.6% of the customers being defaulters

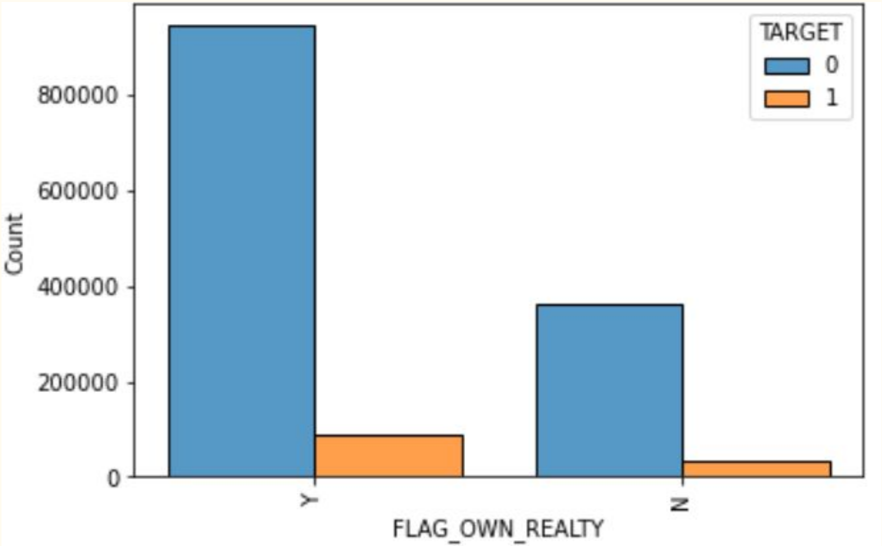
Gender-based Distribution

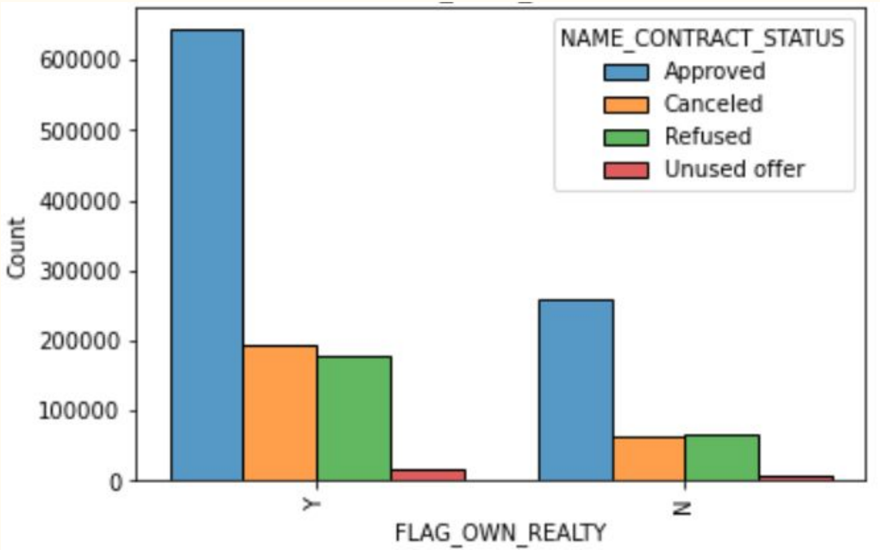




- Females are the dominant customer base
- Despite the huge difference in number, default by female customers seems to disproportionately low

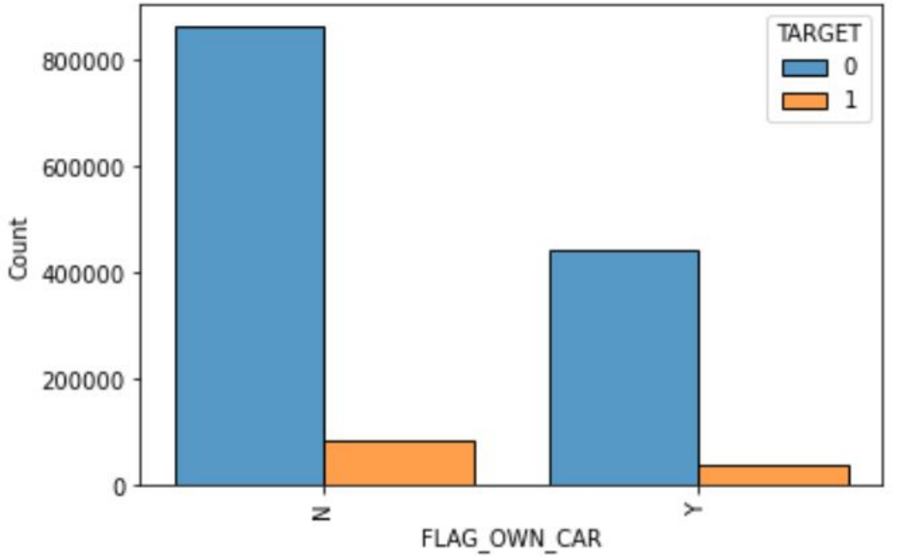
Realty Owners

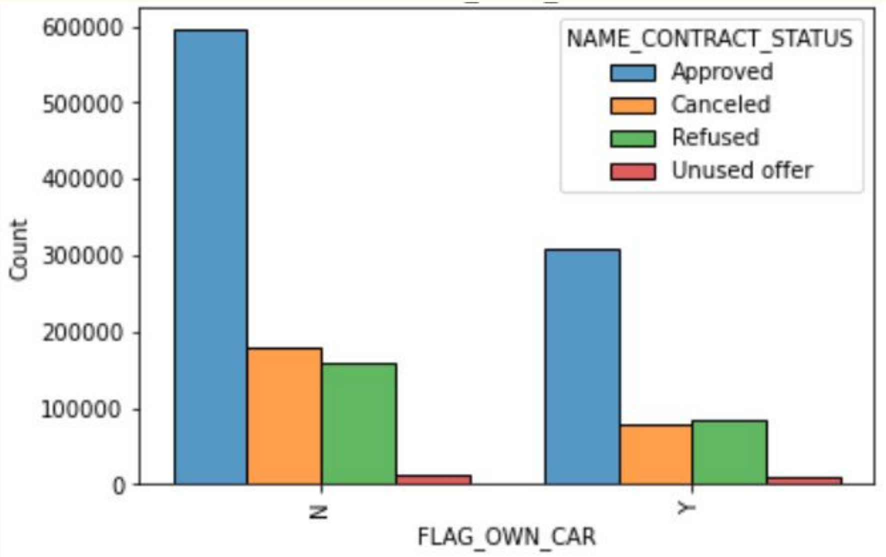




- Only a small proportion of default is seen among realty owners compared to non-owners
- Owning realty can imply that the customer does not have to pay rents and/or has an additional source of income through rents

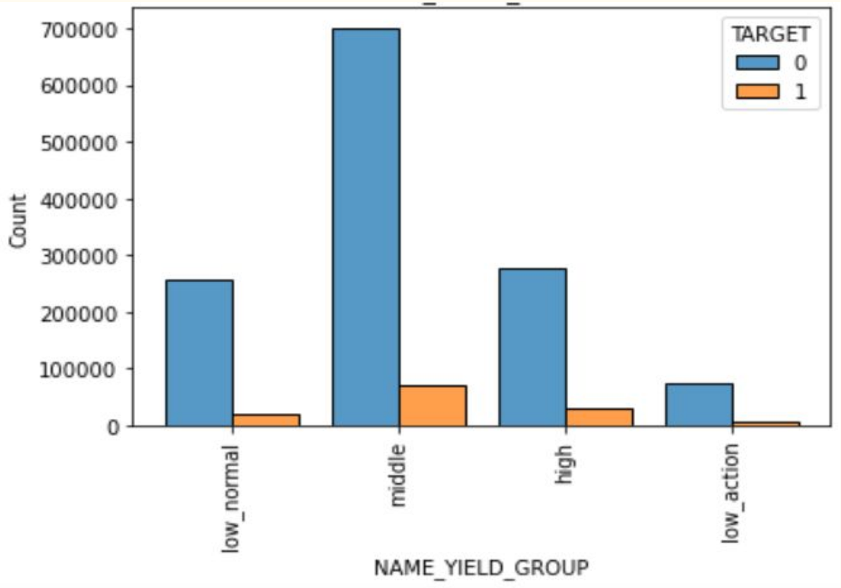
Car Owners

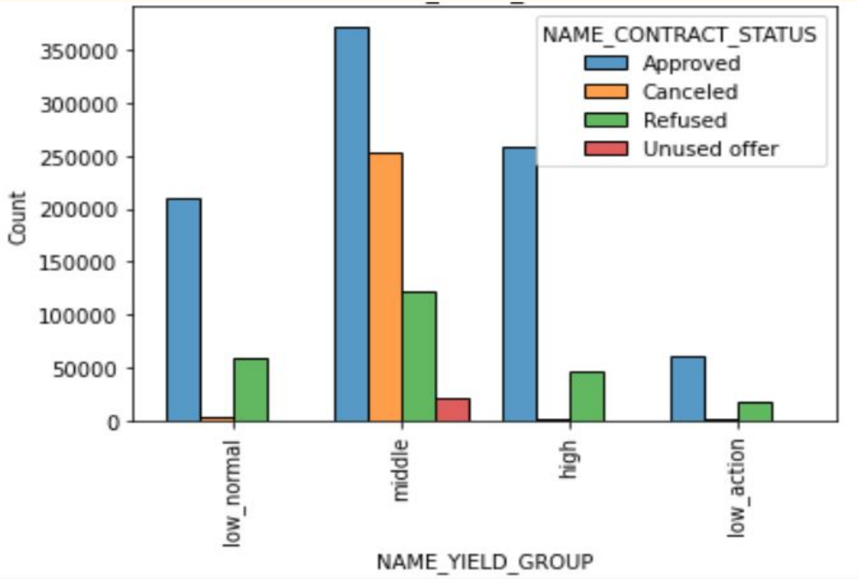




- Car owners’ default rate seems to be higher/similar in proportion compared to non-owners
- Owning a vehicle might incur the customer extra costs, such as fuel, repair, or insurance

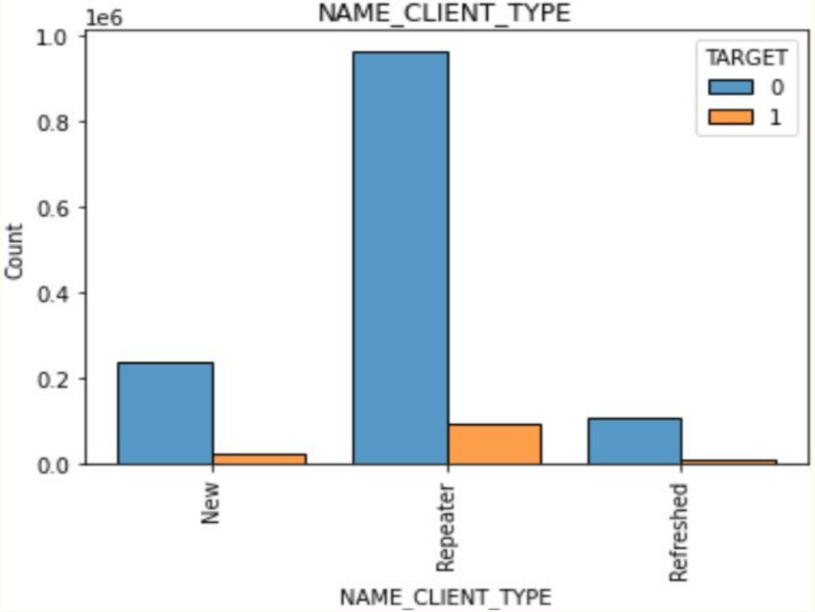
Interest Rate Grouping

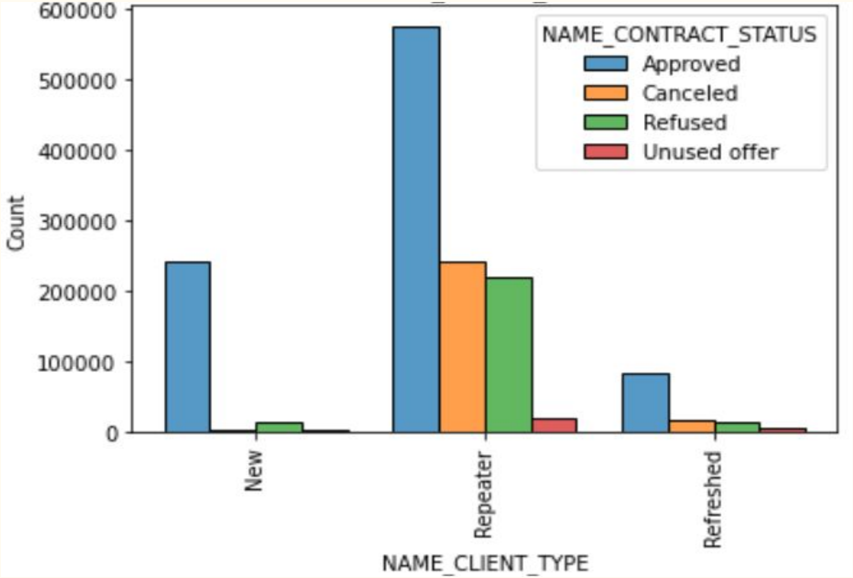




- Middle range interest rate is the most popular
- But it also incurs losses through a high cancellation rate and higher proportion of default

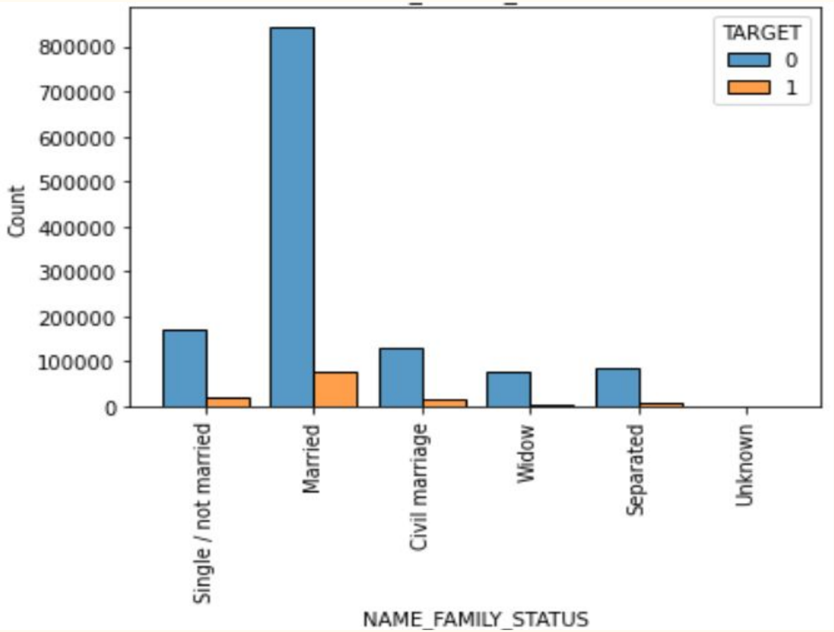
Client Type

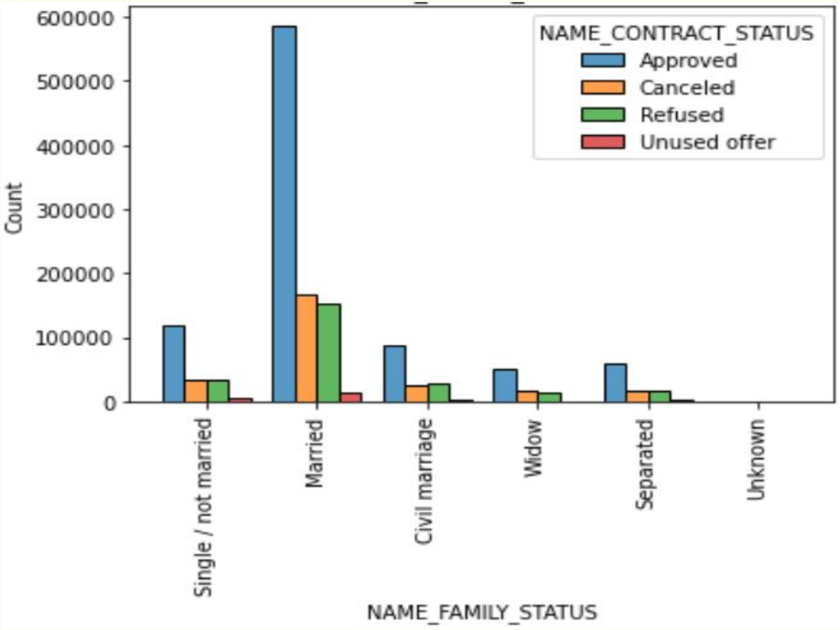




- A large number of customers are repeat customers
- They also boast of a higher default as well as cancellation rate; new applications could be for the repayment of a previous loan

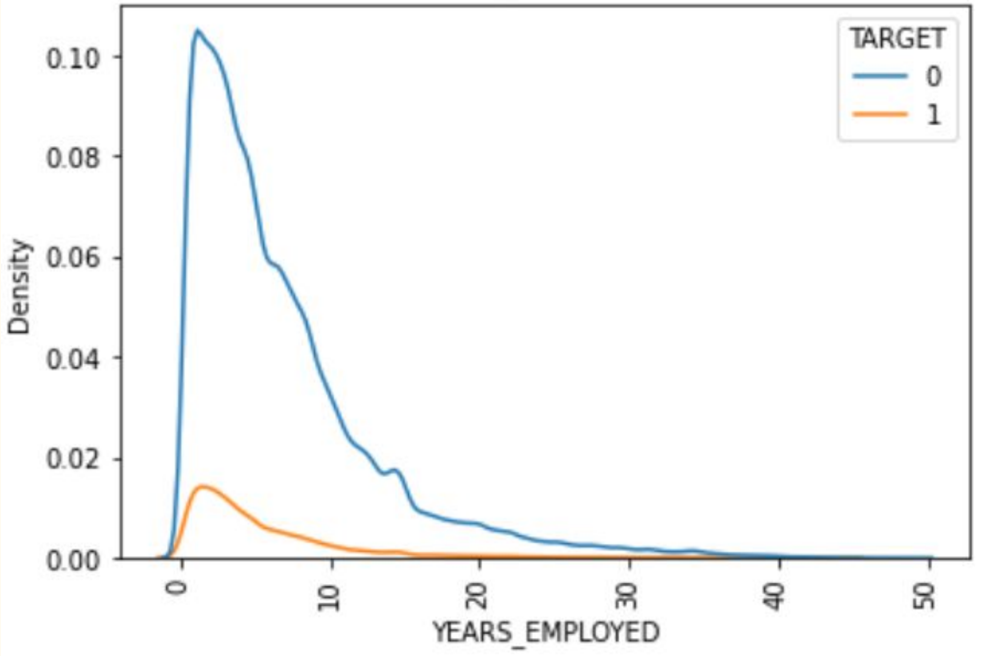
Family Status





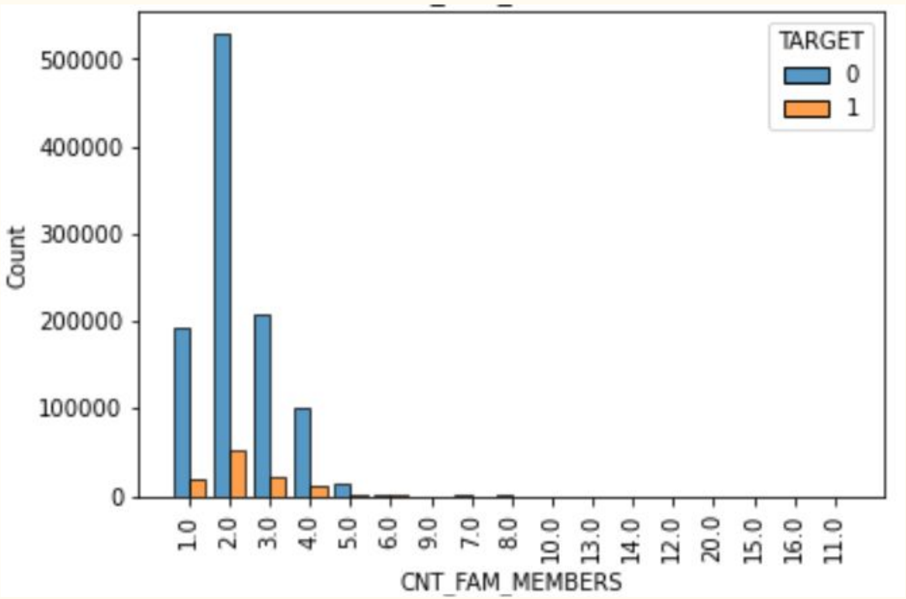
- Majority of the customers are married; families often require loans of different kinds
- Default rates are low as marriage suggests the possibility of two working members in the family

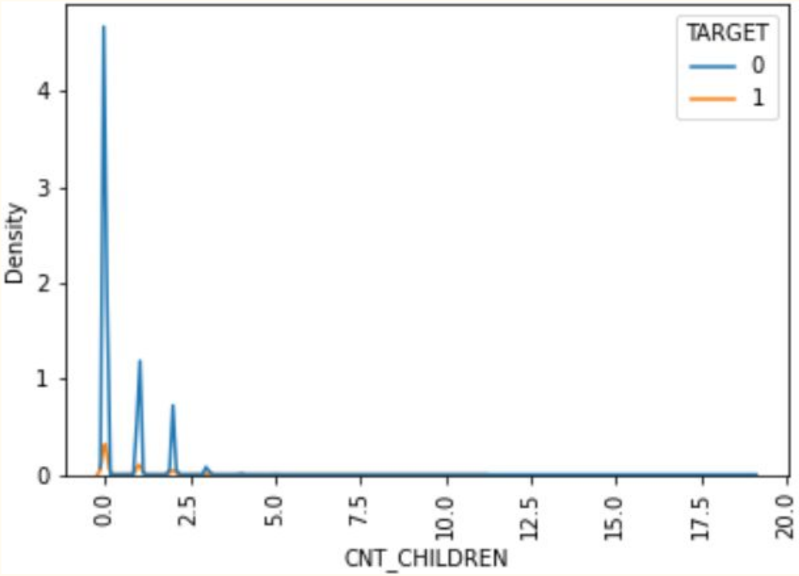
Years Employed



- Customers employed not long before the submission of the loan application seem to have a high chance

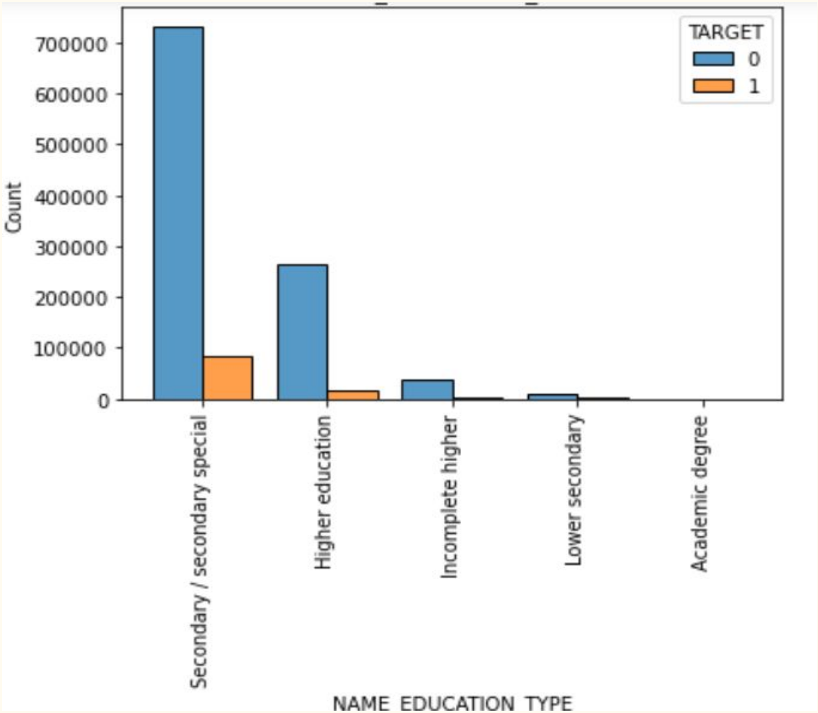
Family Members

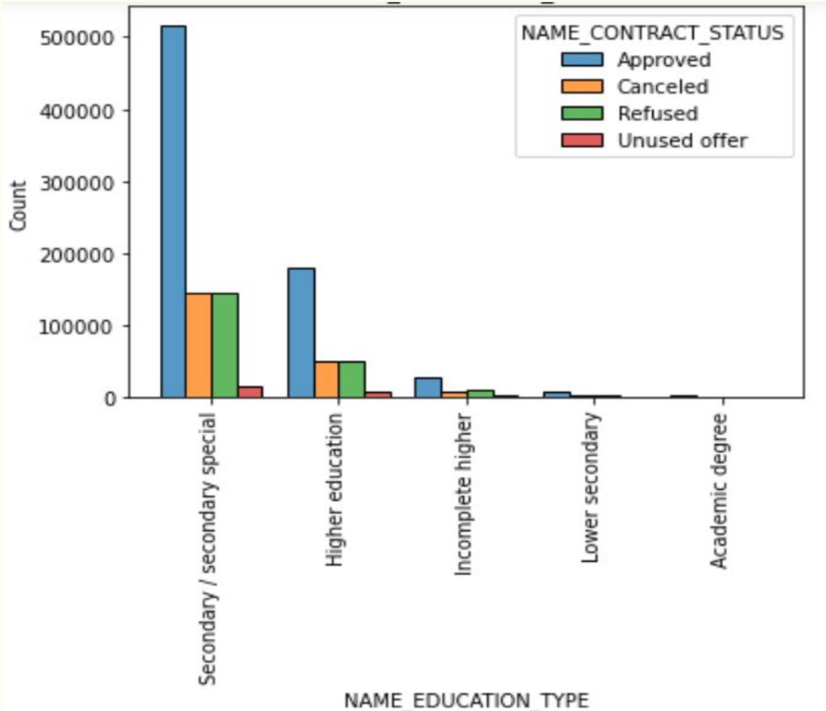




- Customers with fewer family member have lesser expenses to attend to and have a lower chance of defaulting

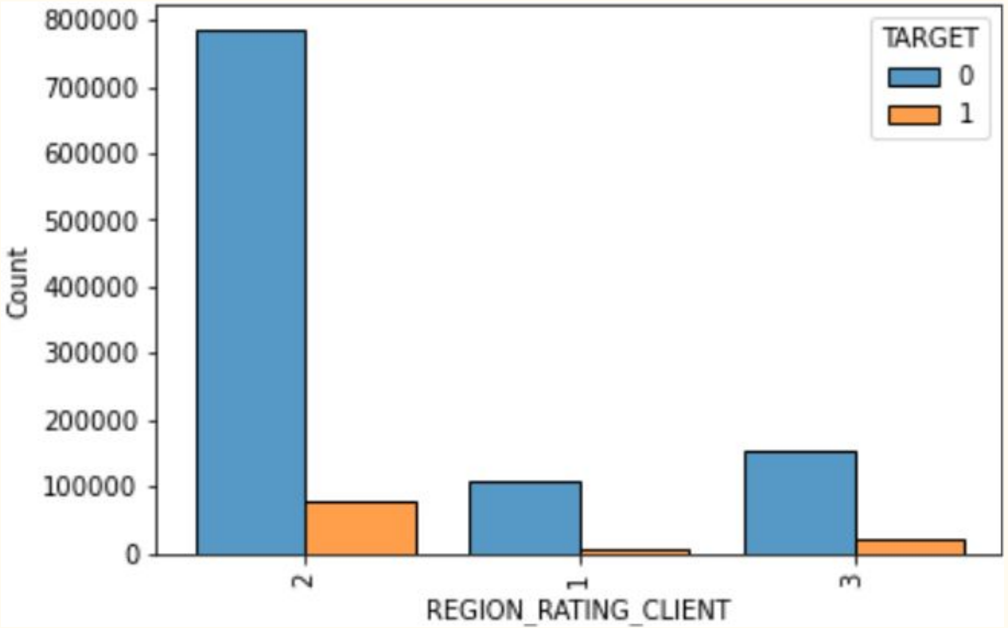
Education





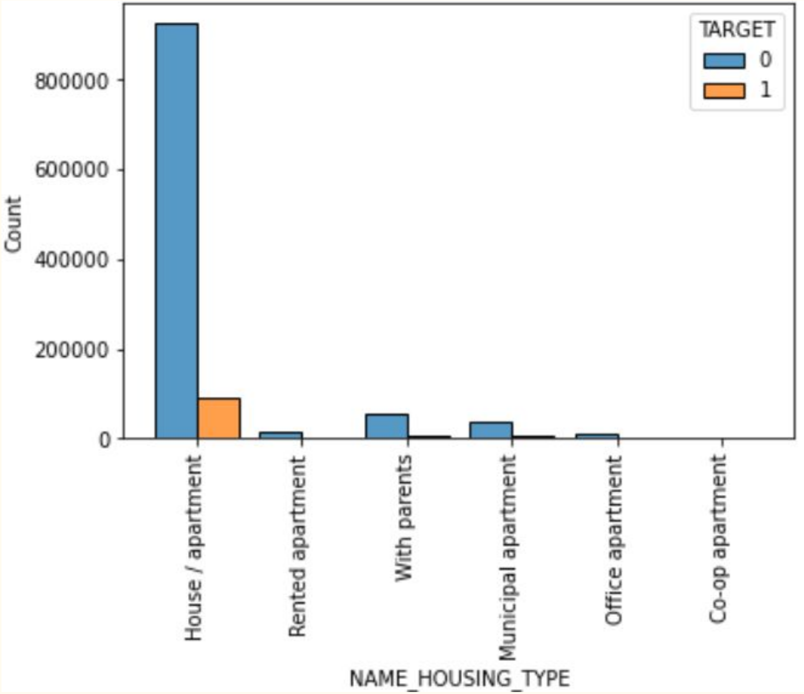
- Customers with at least secondary education seem to require the most loans
- The higher the education level, the lower the default rate

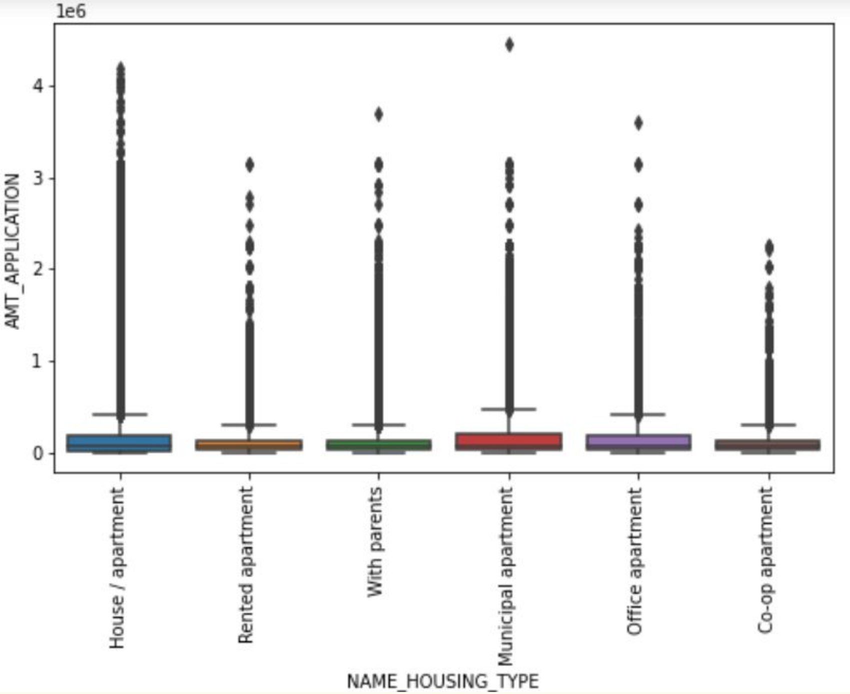
Region Rating



- Clients living in tier 2 regions require the most loans and have a disproportionately low default rate

Housing Type





- People who live in apartments or separate houses have a high loan requirement and a comparatively low default rate

Conclusions and Recommendations

- Female customers tend to default less frequently compared to males (better targets for granting risk free loans)
- Realty (and home) owners and car owners seem to default less and more, respectively; former would incur no rent based costs and the latter would incur vehicle maintenance costs
- Middle range interest rates are the most popular with low default rates but also see a high cancellation rate, perhaps due to competition with other banks
- Repeat customers require the most loans but also tend to default more; more caution is required in granting them loans
- Customers with fewer family members and children are less likely to default due to fewer monthly costs
- Customers who have at least primary education and who have had a job for a number of years prior to loan application tend to default less
- Customers in tier 2 regions would not be very well off, and hence, would require more loans, but would be financially stable enough to default less frequently compared to tier 3 region customers

Thank You!