

Credit EDA Analysis

Exploratory data analysis for loan application process



Objective

Lending companies find it challenging to make loans to consumers that have a poor or non-existent credit history. As a consequence, some consumers take advantage of the situation by defaulting on their payments.

By identifying trends, this case study may determine whether to refuse a loan, reduce the loan amount, or lend (to riskier applicants) at a higher interest rate. This will prevent customers who can repay the loan from being denied. This case study aims to identify such applications using EDA.

In other words, the corporation needs to know the characteristics that strongly indicate loan default. This information may be used for portfolio and risk evaluation.

Data - Loan Applications

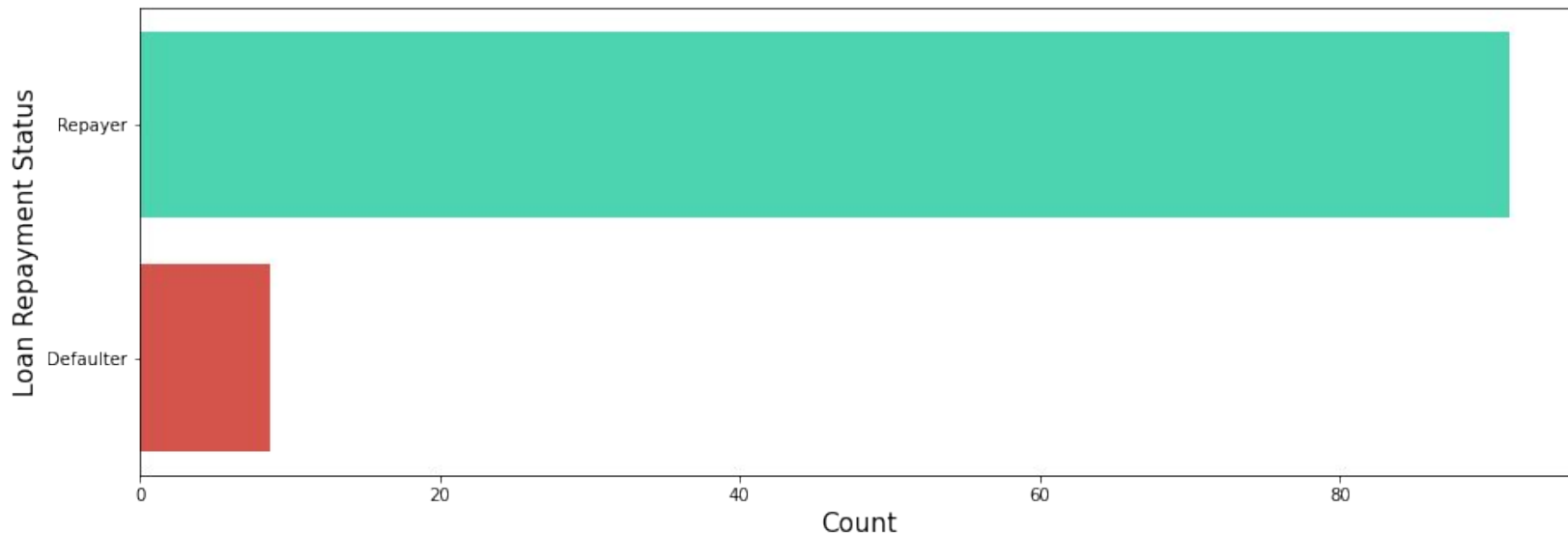
Loan Applications: The information below belongs to the loan application at the time of application. It has two scenarios:

1. The client having payment difficulties: he/she was more than X days late on at least one of the first Y payments of the loan in our sample.
2. Other cases: When the payment is made on time.

The dataset has over 30 lakh entries for over 100 different attributes.

EDA - Loan Applications

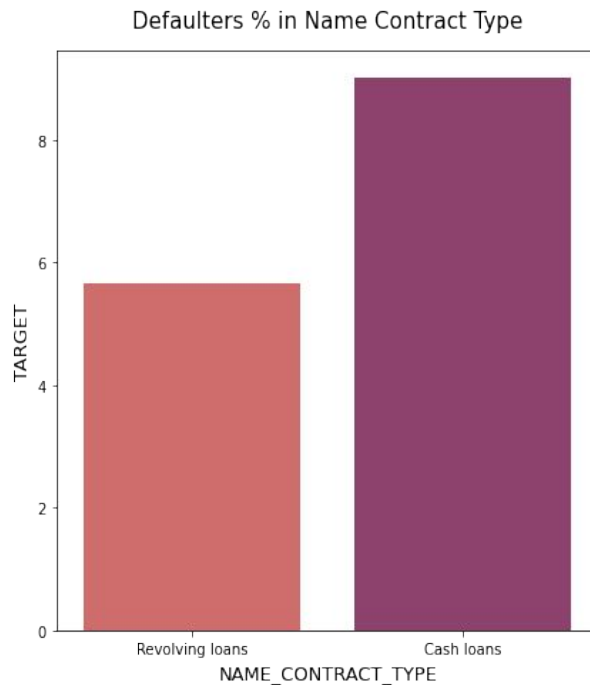
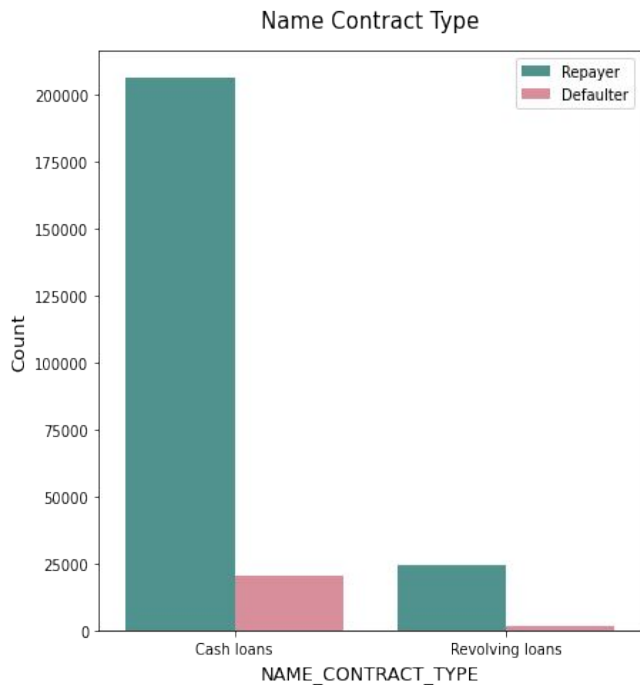
Imbalance Plot (Repayer Vs Defaulter)



Observation: Over 90% of the applicants had repaid the loan, while the remaining were defaulters.

Univariate Analysis

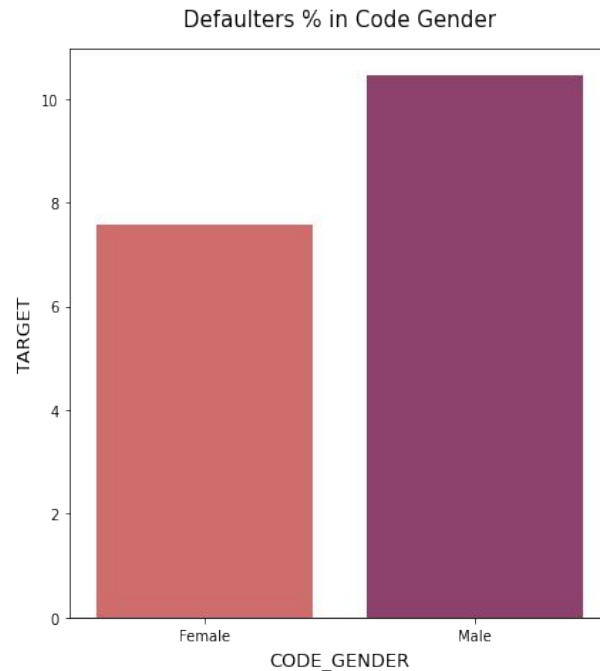
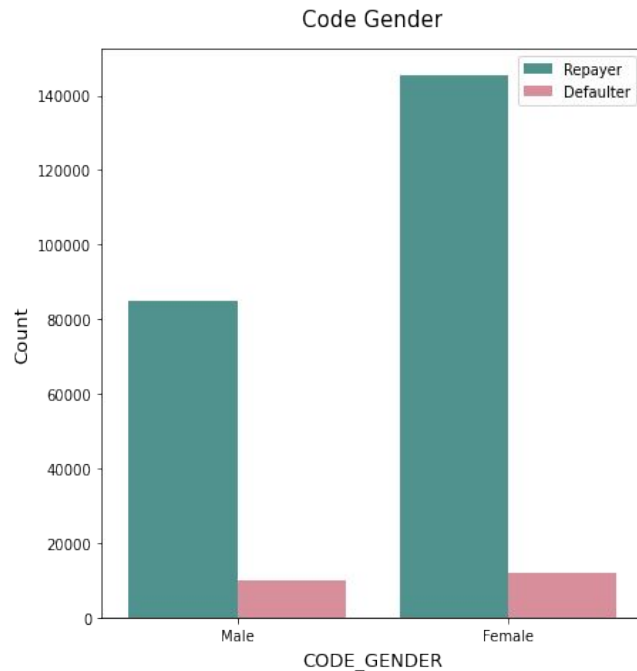
Contract Type and Loan Status



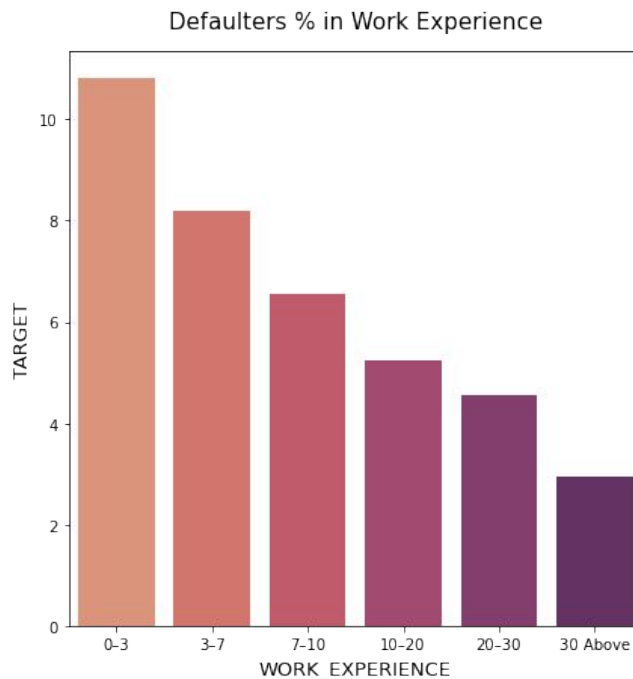
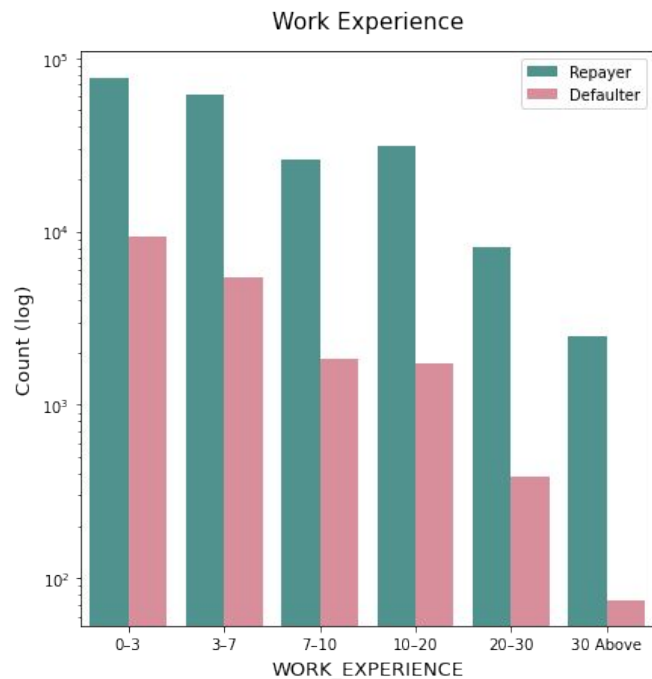
- Revolving loans account for just a modest percentage of overall loans.
- Approximately 8-9% of cash loan applicants and 6% of revolving loan applicants default.

Gender and Loan Status

- Female customers outnumber male clients by almost two to one.
- According to the proportion of defaulted loans, males have a 10% chance of not returning their obligations, while women have just below 7% chance.



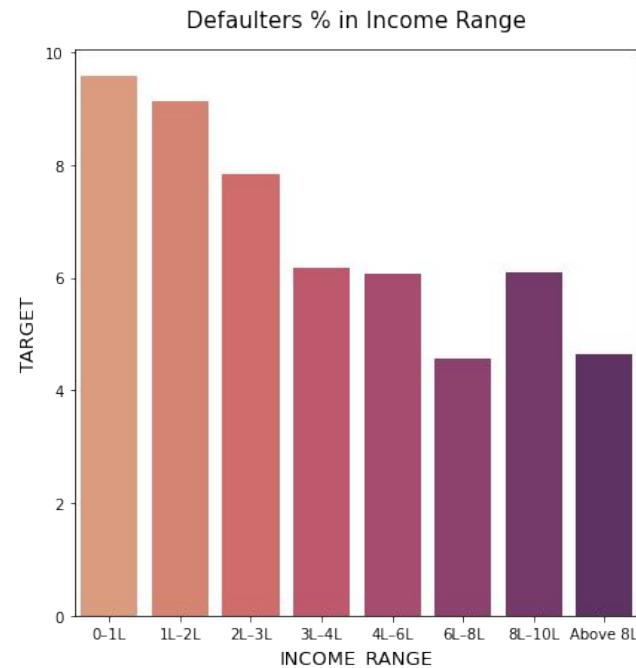
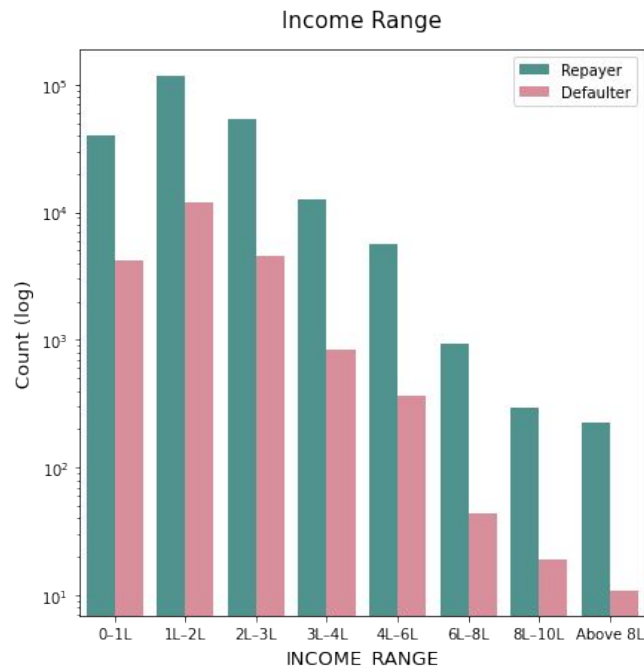
Work Experience and Loan Status



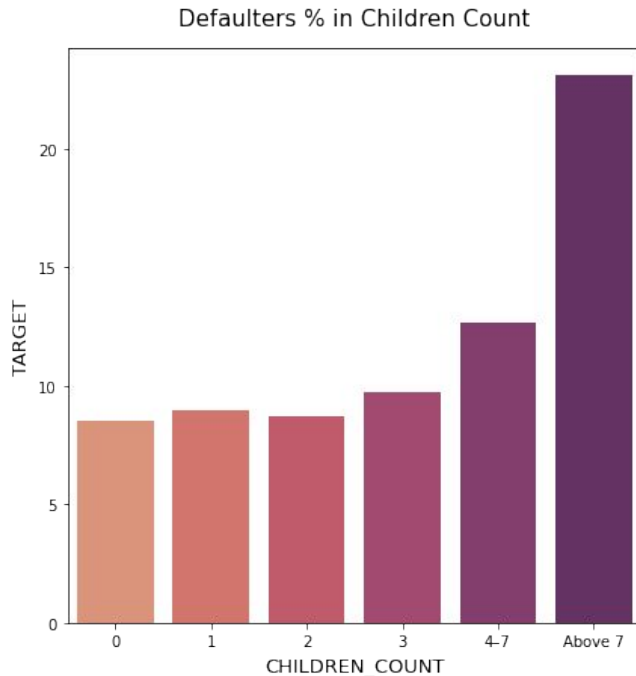
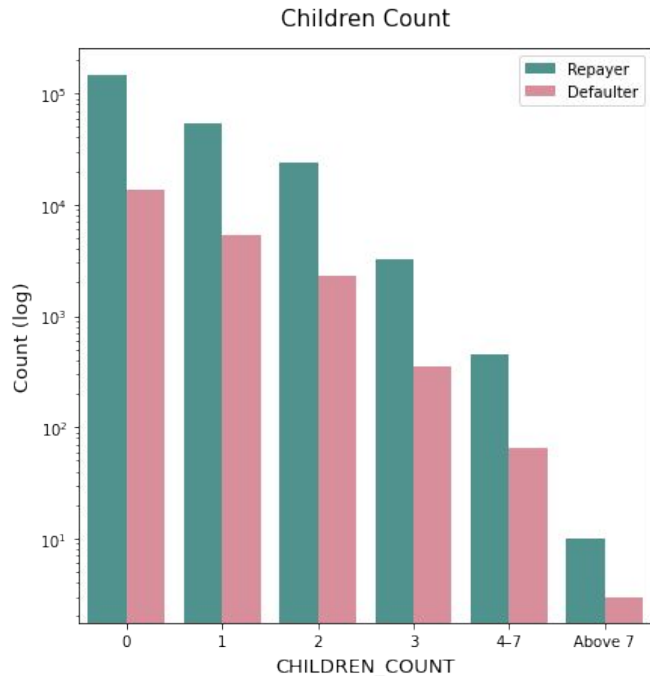
- Clearly, the majority of loan applicants have little or no employment experience.
- The defaulters percentage chart clearly shows a downward trend in the number of defaulters associated with increasing years of work experience, with the 0-3 year experience group having the highest defaulters.

Income Range and Loan Status

- Evidently, the majority of loan applicants earn less than 2 lakh.
- Overall, the defaulters % chart indicates a declining trend in the number of defaulters as income increases, with applicants earning less than 3 lakhs defaulting more often



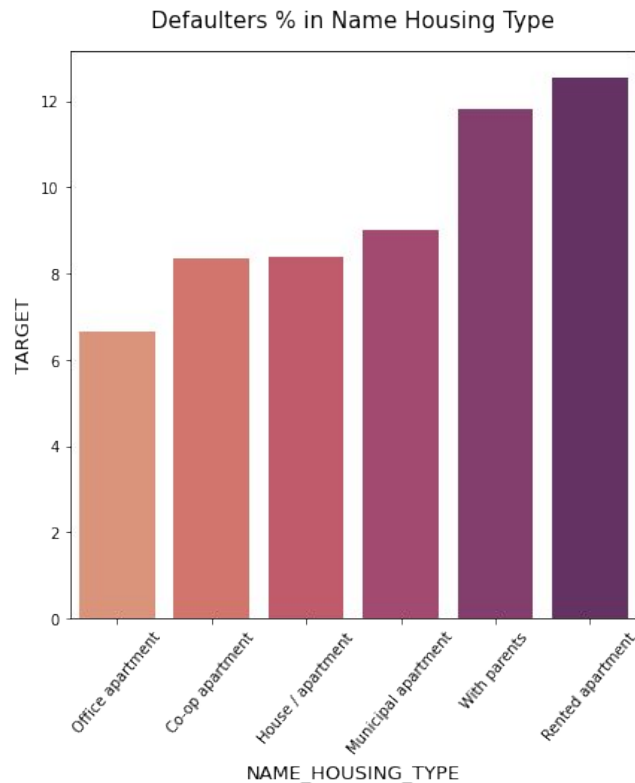
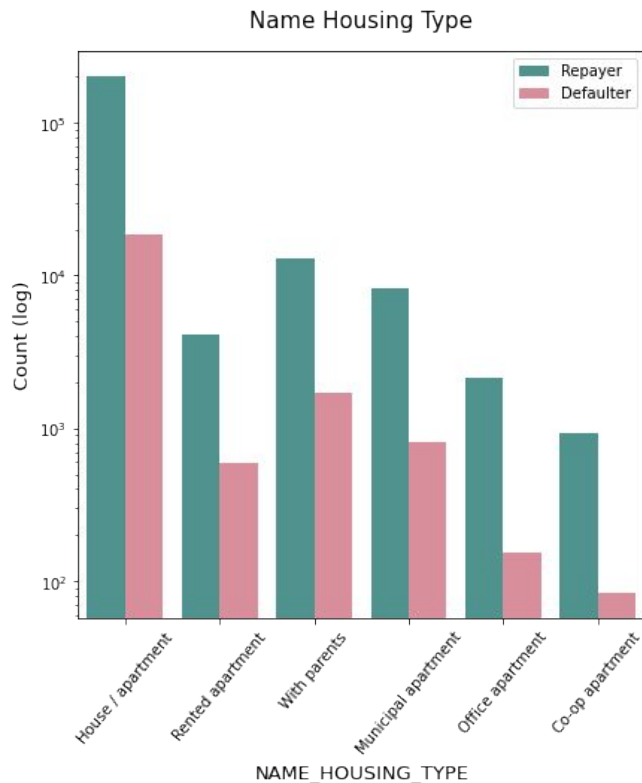
Work Experience and Loan Status



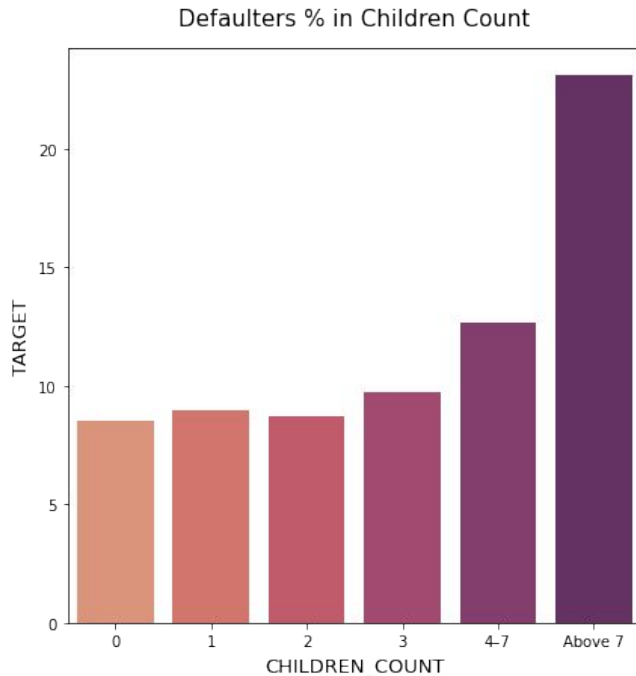
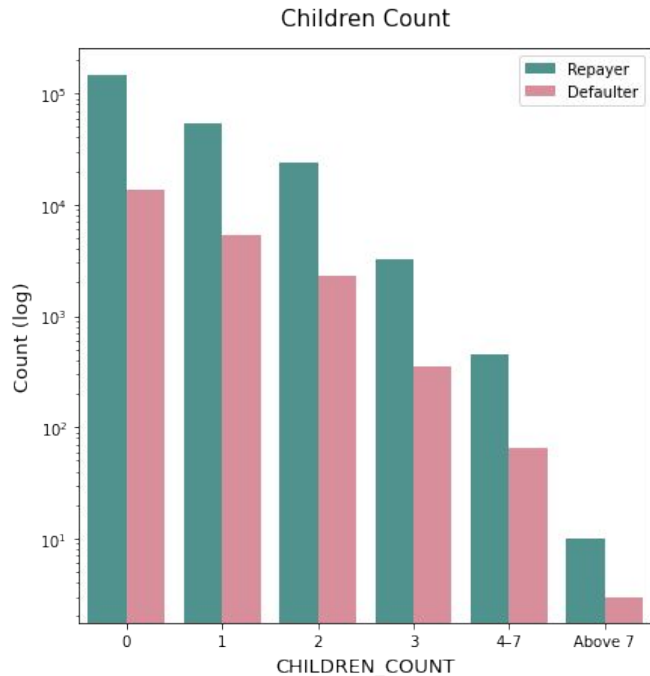
- Applicants with 1-3 children account for approximately 10% of defaulters.
- Applicants with no children are more likely to repay the loan, whereas those with more than seven children are more likely to default.

Housing Type and Loan Status

- The vast majority of individuals live in a house or apartment.
- Those who live in office apartments have the lowest default rate.
- Applicants who live with their parents and in leased flats have a greater chance of defaulting (roughly 12%).



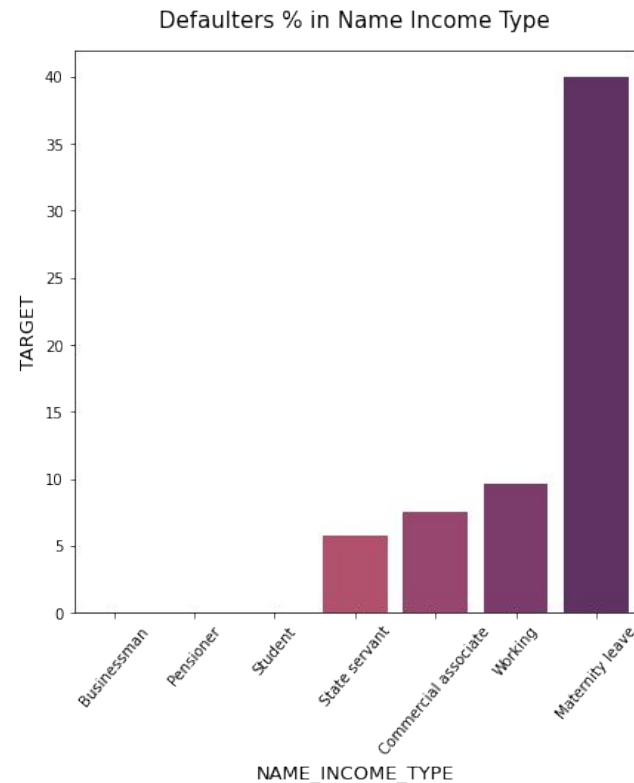
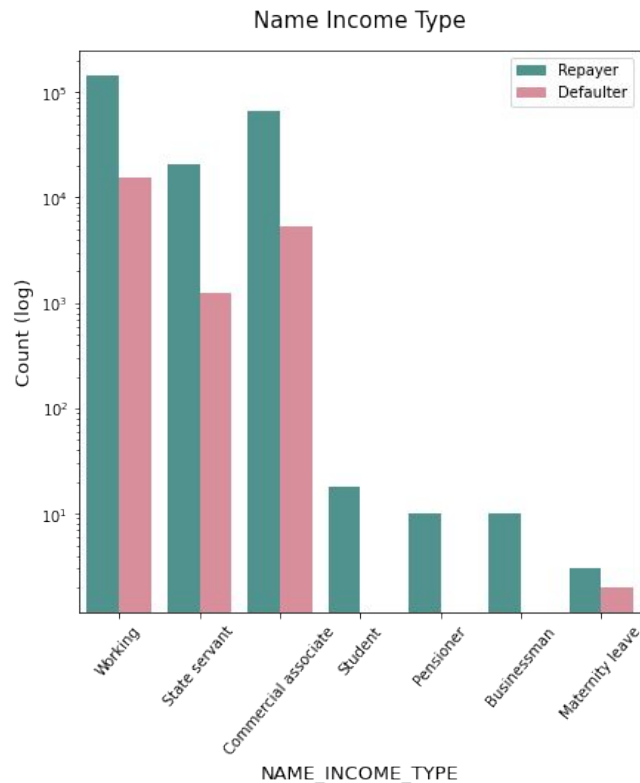
Education and Loan Status



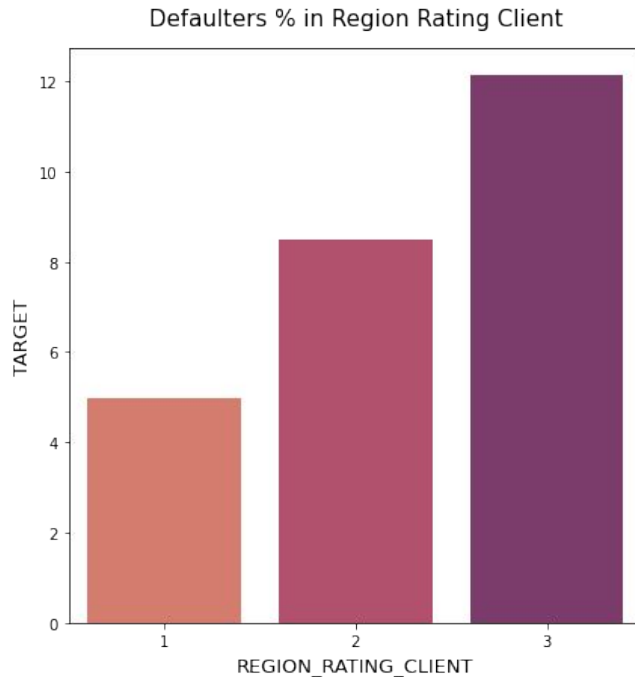
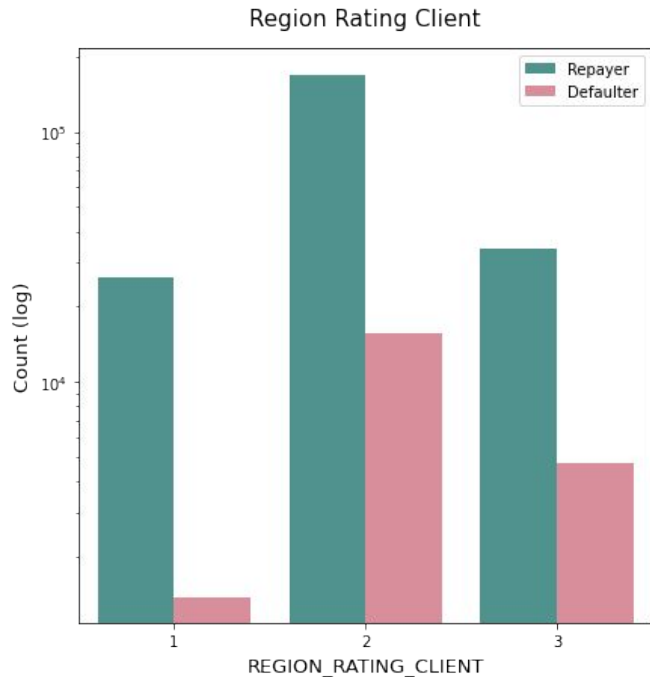
- Clients with secondary education outnumber those with higher education, on the other hand, very few have an academic degree.
- Lower secondary education has the greatest probability of defaulting at roughly 11%, whilst those with an academic degree are the least likely to default.

Work Type and Loan Status

- The majority of loan applicants have a working income, followed by a commercial associate, and a state employee.
- Maternity leave applicants had the highest defaulting rate of 40%.
- Despite their smaller numbers, students and businessmen do not have a default record. The two most secure loan kinds.

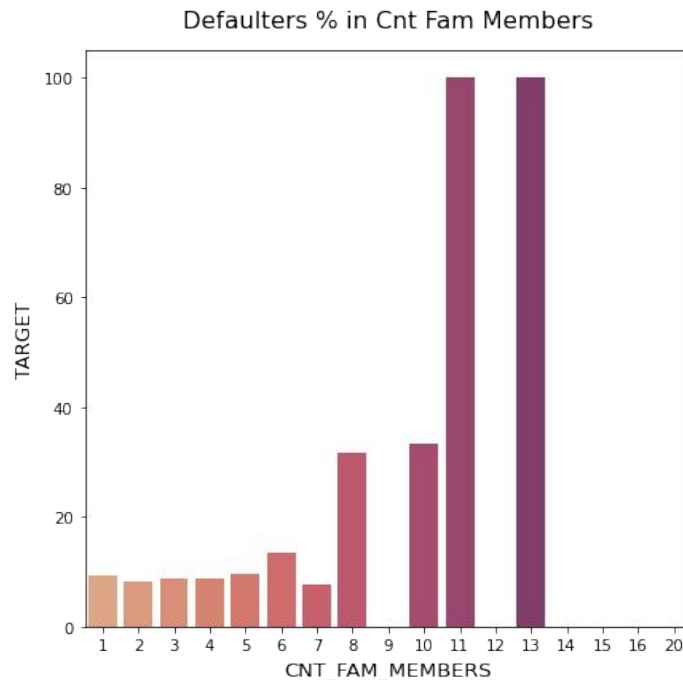
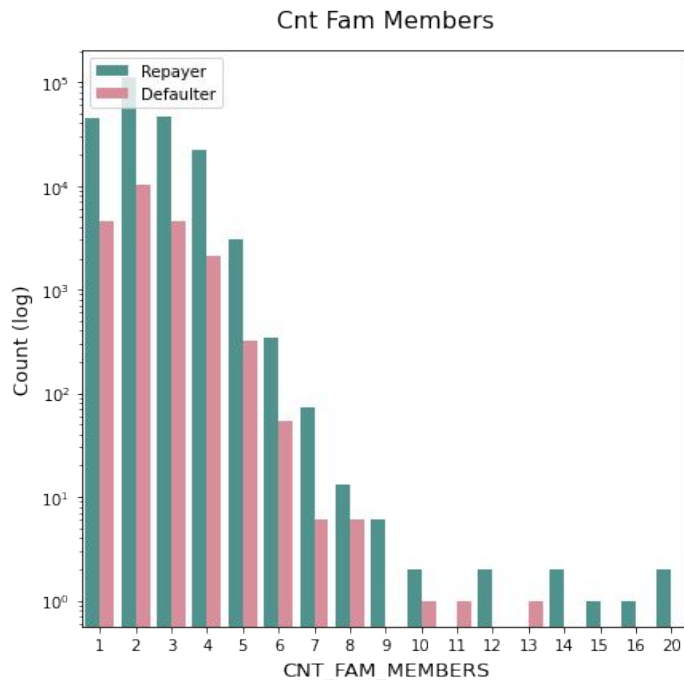


Education and Loan Status



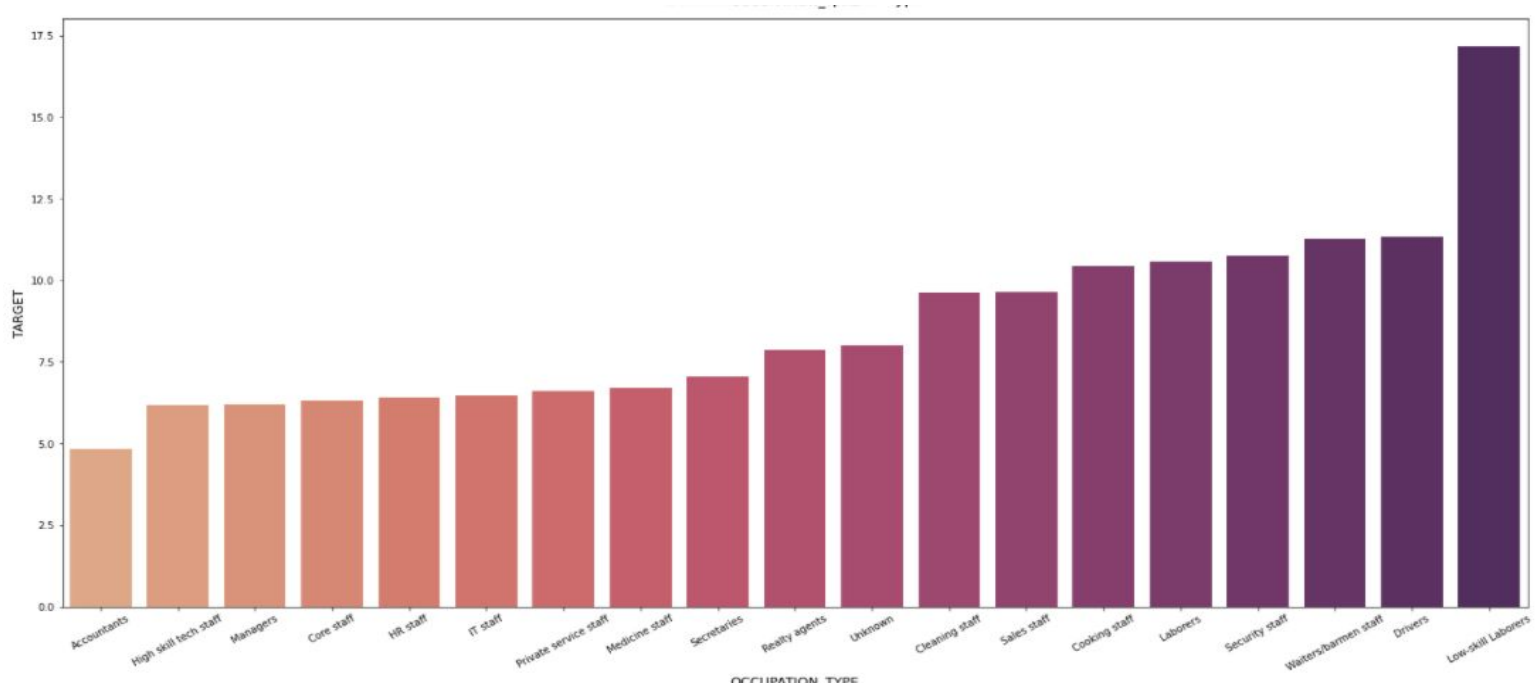
- The majority of applicants live in a Region with a Rating 2 location.
- The region with the greatest default rate is Region Rating 3(11%).
- Clients residing in Region Rating 1 has the lowest likelihood of defaulting, making loan approval safer.

Family members count and Loan Status



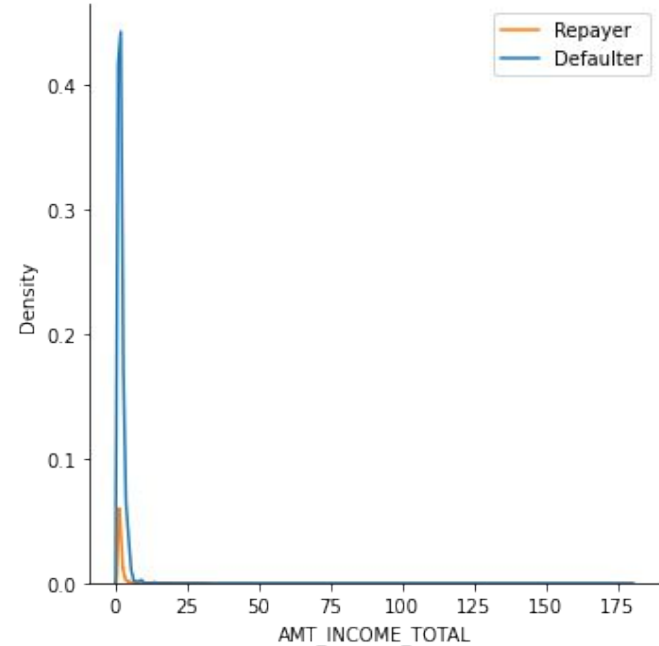
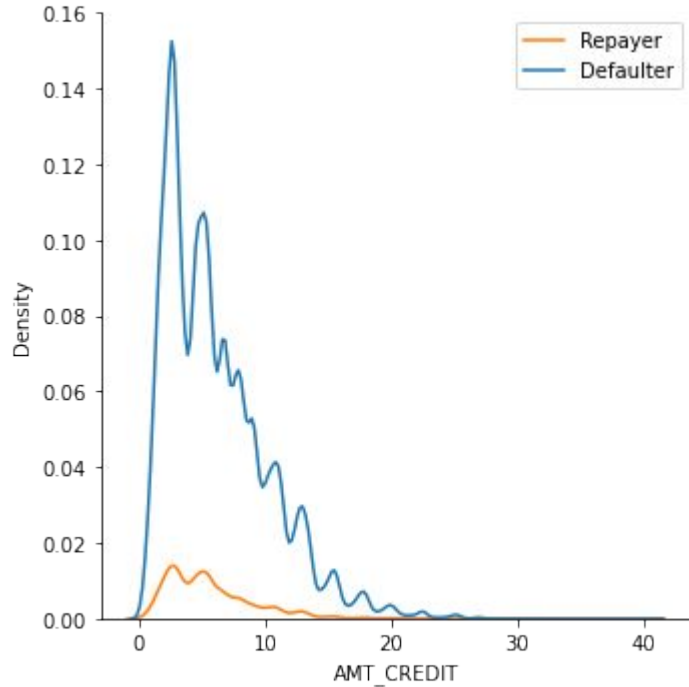
- Family members follow the same pattern as children in that having more family members raises the probability of defaulting.

Occupation Type and Defaulters Percentage



- Low-skill labourers had the largest percentage of defaulters (almost 17%), followed by drivers and waiters/bartenders, security personnel, labourers, and cooks.

Income and Amount credit on Loan Status

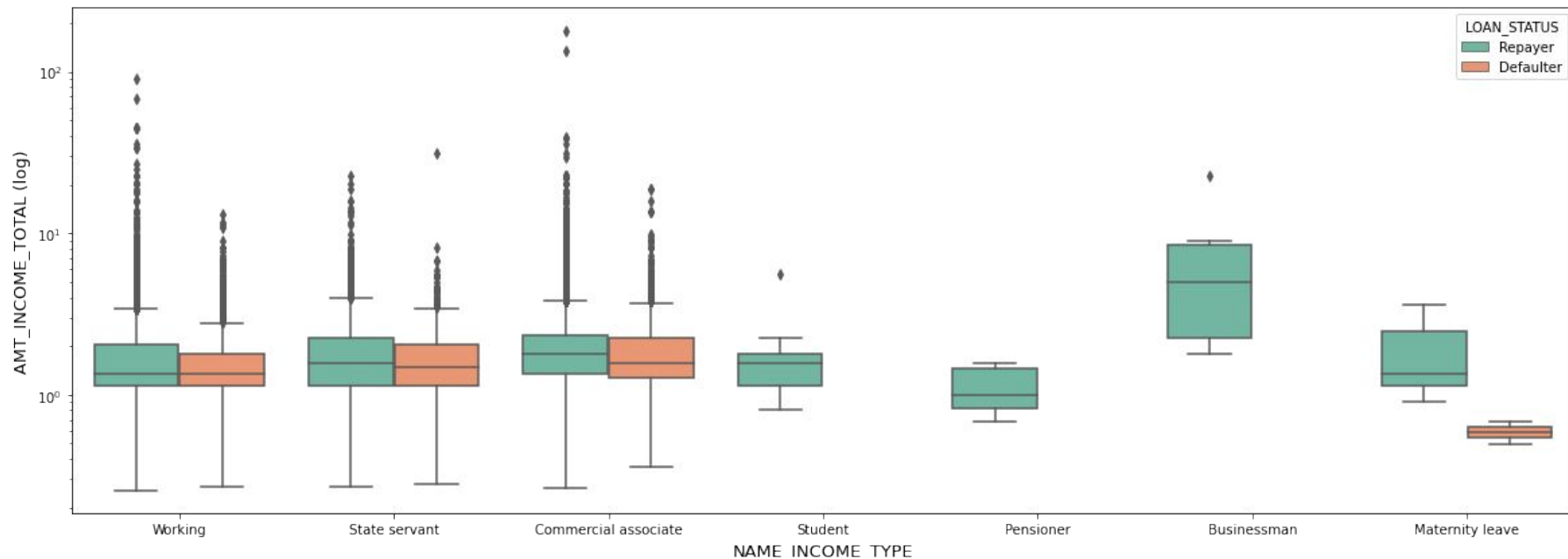


Since the repayers and defaulters distributions overlap in all of the plots, we cannot make a conclusion based just on any of the above four continuous variables.

Bivariate Analysis

Income type and Total income on Loan Status

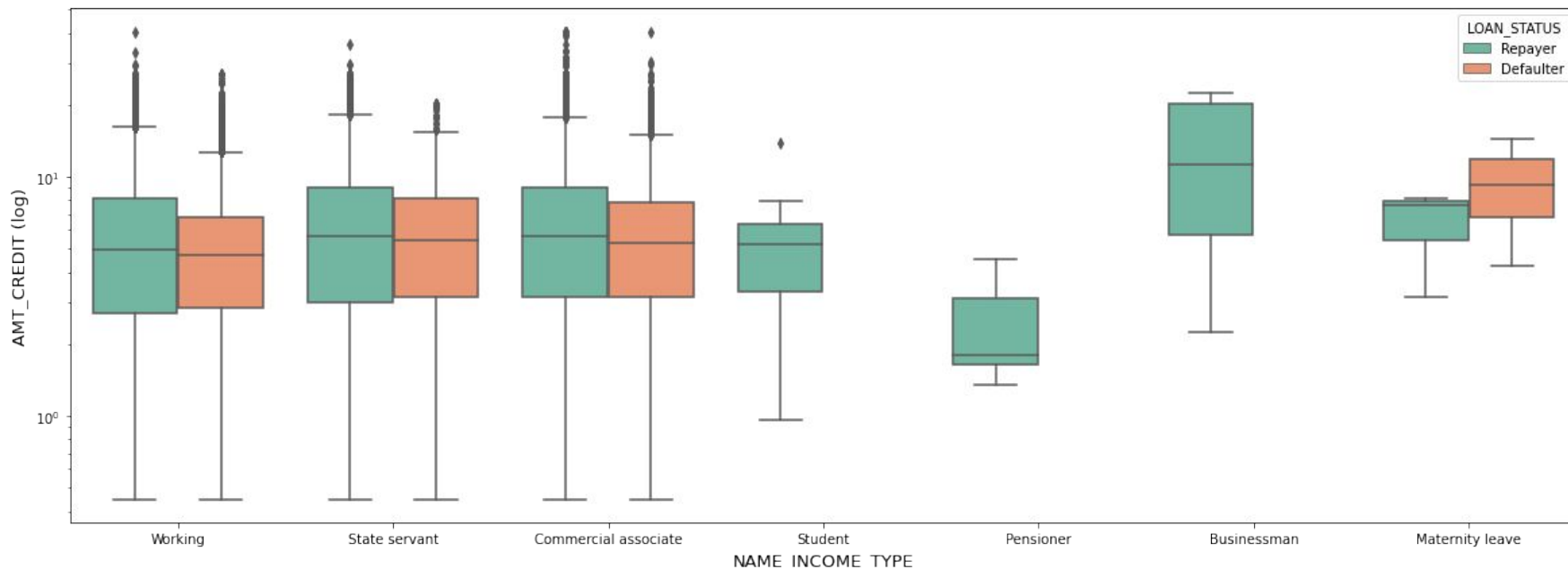
Distribution of Amt Income Total for Repayed and Defaulter categorised by Name Income Type



Clients earning less and on maternity leave seem to be more likely to default on their loans, while business professionals and students had nearly no defaults regardless of income type.

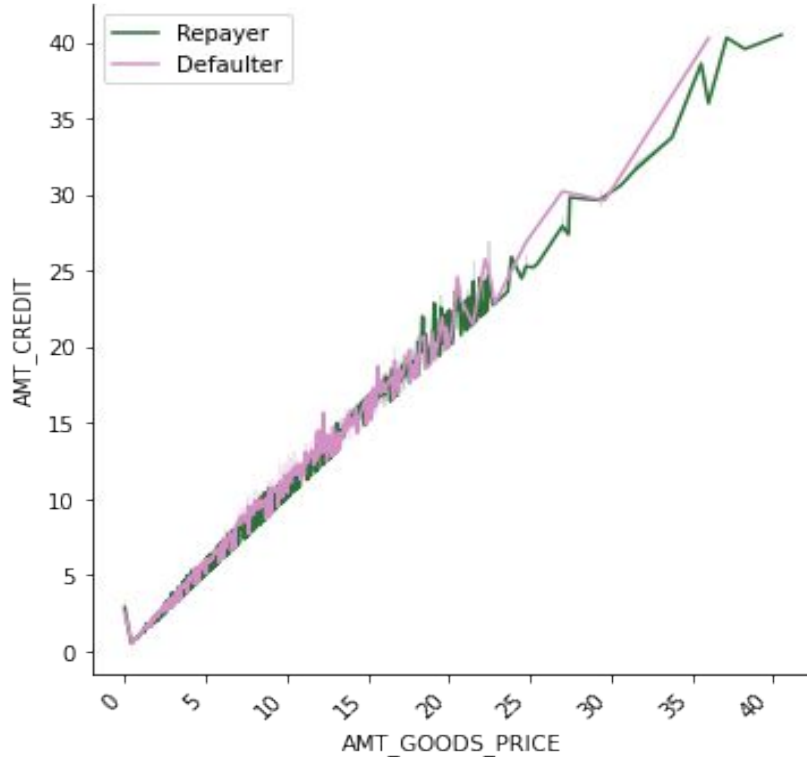
Income Type and Amount Credit on Loan Status

Distribution of Amt Credit for Repayed and Defaulter categorised by Name Income Type



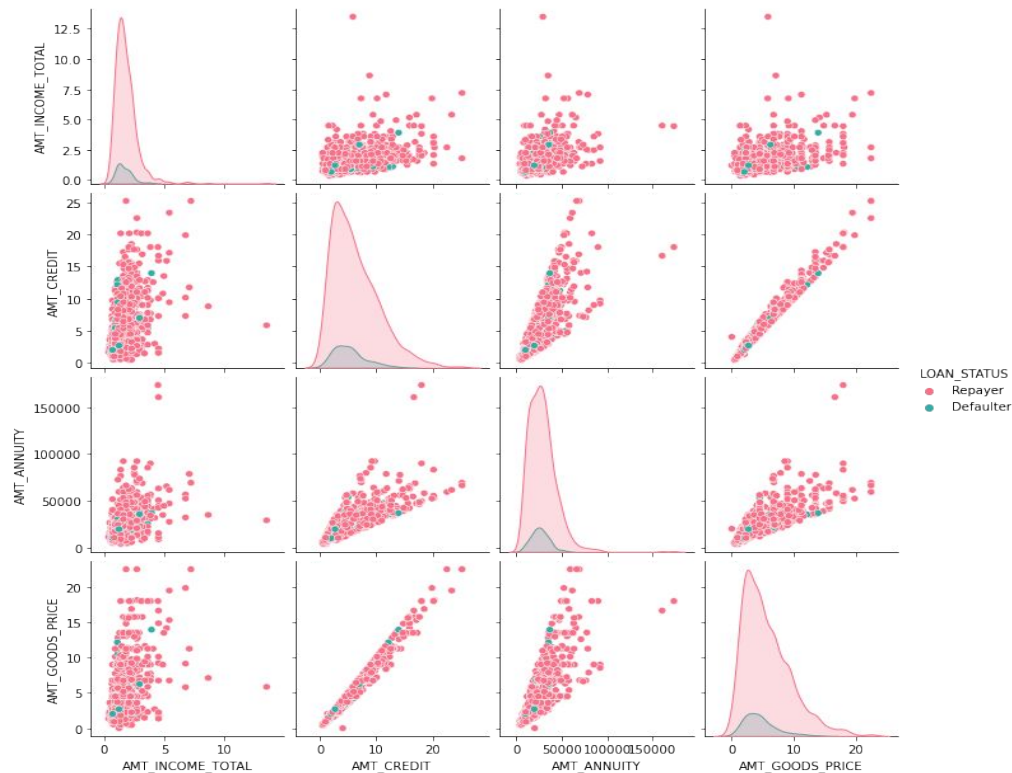
Clients with larger loan credit and on maternity leave seem to be more likely to default on their loans, while business professionals and students had nearly no defaults regardless of income type.

Goods Price and Credit on Loan Status



When the loan amount exceeds 30 lakhs, the number of defaulters increases.

Analysis of Amount Variables



- There is a lower possibility of defaulters when the Annuity Amount > 15K and the Good Price Amount > 20 Lakhs.
- There is a lower possibility of defaulters when the Annuity Amount > 15K and the Good Price Amount > 20 Lakhs.
- For Amount Credit > 20 Lakhs, there are relatively few defaulters.

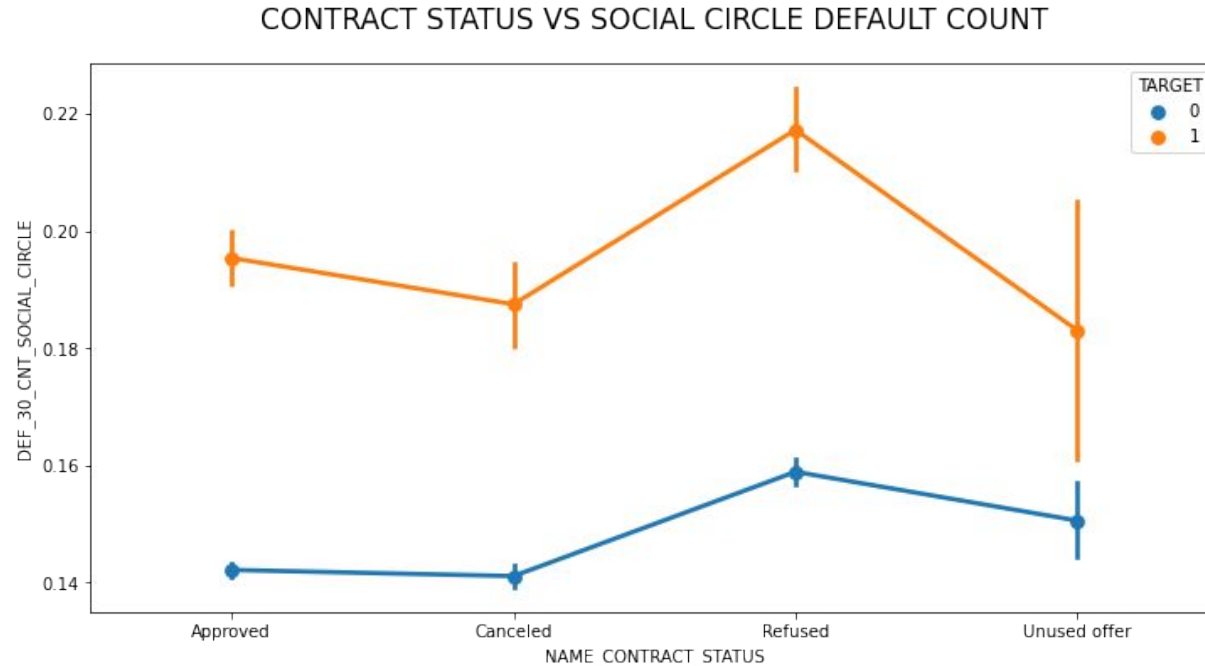
Data - Previous Applications

Previous Applications: When a customer requests for a loan, the client/company has four options:

- Approved: The Company has approved loan Application.
- Cancelled: The client cancelled the application sometime during approval.
- Refused: The company had rejected the loan.
- Unused offer: Loan has been cancelled by the client but on different stages of the process.

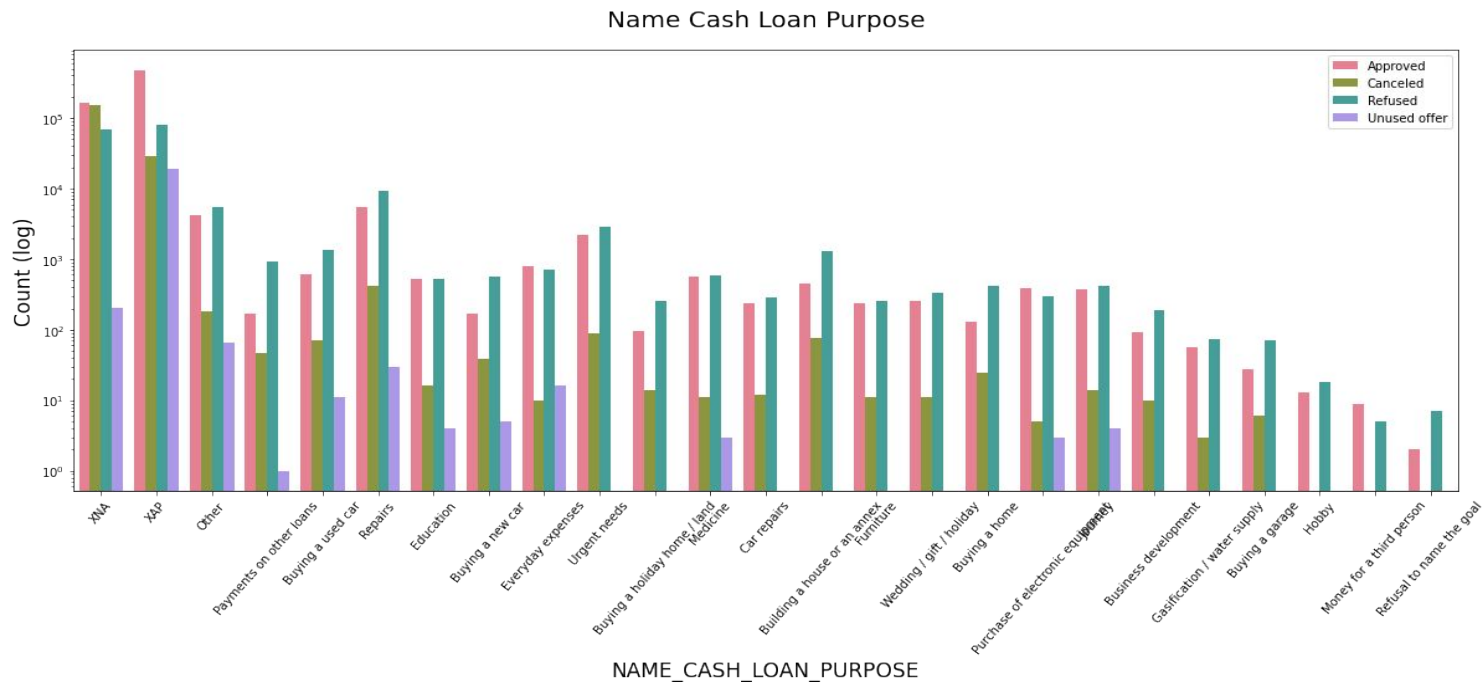
Multivariate Analysis

Social Status Vs Social Circle Default Count



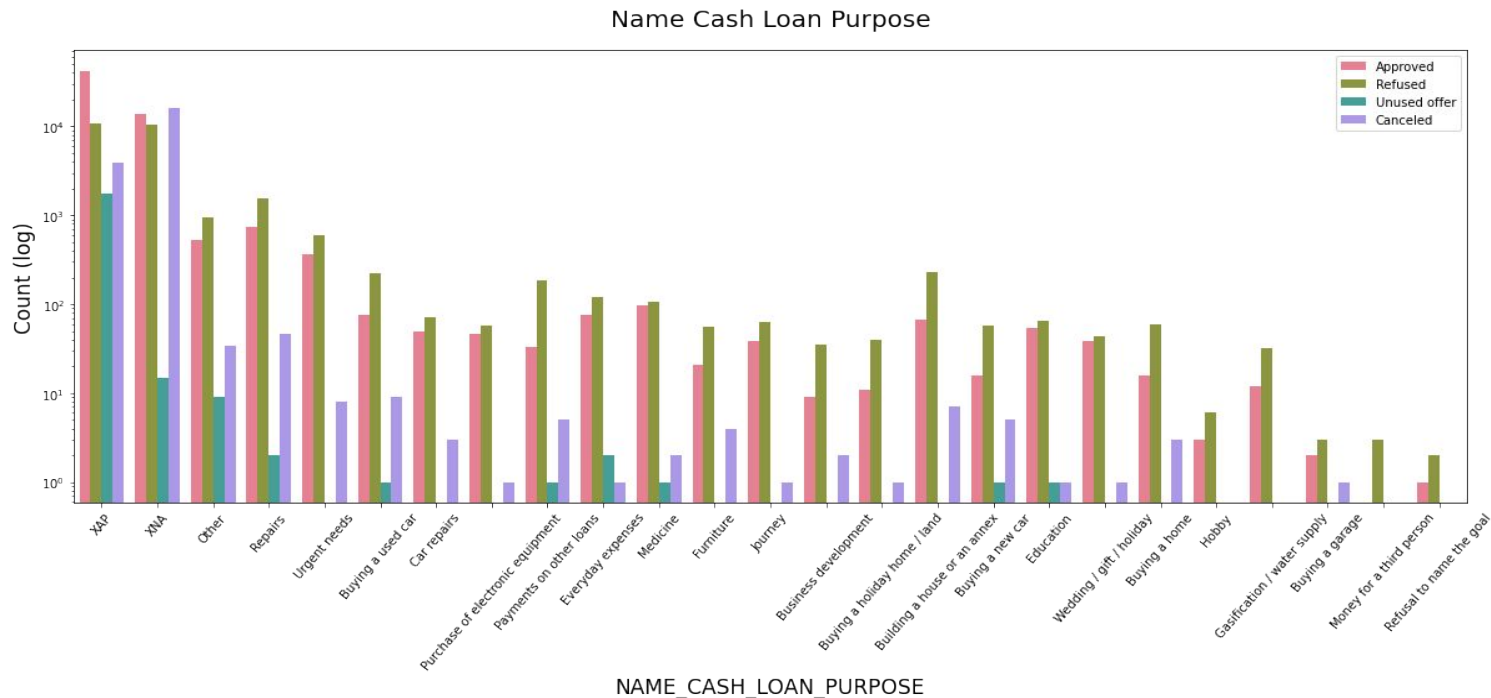
Clients with a DEFAULT 30 COUNT SOCIAL CIRCLE score of 0.18 or above are more likely to default, hence analysing the client's social circle might aid in loan disbursement.

Loan Purpose Vs Contract Status for Payers



There are a larger number of unknown loan reasons, and loans obtained for the purpose of repairs seem to have the greatest default rate.

Loan Purpose Vs Contract Status for Defaulters



A large percentage of applications for Repair or Other have been denied by banks or declined by clients. Furthermore, they are either refused or the bank gives a loan with a high interest rate that the consumers cannot afford, and they decline the loan.

Repayment status analysis for business loss



		Counts		Percentage
NAME_CONTRACT_STATUS	TARGET			
Approved	0	657609		91.86
	1	58295		8.14
Canceled	0	181475		89.95
	1	20282		10.05
Refused	0	175572		87.11
	1	25979		12.89
Unused offer	0	19069		91.48
	1	1776		8.52

- 90% of the previously terminated clients have paid back their loans.
- 88% of consumers who were previously rejected a loan have paid it back.

Results

Following is the analysis with relevant elements and classification based on which the bank may determine a client's repayment ability

Factors that influence whether or not an applicant is likely to repay

1. REGION_RATING_CLIENT: Rating 1 is the safest.
2. NAME_EDUCATION_TYPE: Academic degree has fewer defaults.
3. NAME_INCOME_TYPE: Businessmen and Students have little or no defaults.
4. DAYS_EMPLOYED: Applicants with more than 40 years of expertise have a default rate of less than 1%.
5. AMT_INCOME_TOTAL: Clients earning more over 7 lakhs have a lower risk of default.
6. ORGANIZATION_TYPE: Applicants belonging to Industry Types 4 and 5 have defaulted at a rate of less than 3%.
7. AMT_CREDIT: Applicants with loan amounts less than Rs. 30 lakhs have the lowest default rate.

Factors that influence whether or not an applicant is likely to default

1. DAYS_EMPLOYED: Individuals with fewer than five years of job experience have a significant default rate.
2. CODE_GENDER: Men default at a larger rate than women.
3. NAME_EDUCATION_TYPE: Individuals with a secondary or lower secondary education are more likely to default
4. NAME_FAMILY_STATUS: Individuals who are single or had civil marriages often default.
5. NAME_INCOME_TYPE: People who are unemployed or on maternity leave often default.
6. CNT_CHILDREN: Clients with 7 or more children are substantially more likely to default.
7. REGION_RATING_CLIENT: Residents of Rating 3 locales have the greatest default rates.
8. AMT_GOODS_PRICE: When the loan amount exceeds 3 lakhs, the number of defaulters increases.
9. AMT_INCOME_TOTAL: Individuals earning less than two lakhs are more prone to default.
10. OCCUPATION_TYPE: The default rate for low-skilled labourers, drivers, and waiters/bartenders, as well as security personnel, labourers, and cooks, is quite high.

Suggestions

- Ninety percent of the previously cancelled customers have actually paid back the loan in full and on schedule. Keep track of the reasons for the cancellation so that the bank may better identify and negotiate conditions with clients who want to pay back in the future.
- Almost Ninety percent of the customers who were previously turned down for a loan by a bank have now become prepaid customers. Documenting the reasons for denial might help to minimise company losses, and these customers may be approached again for more loans.
- A large proportion of loan applications come from individuals who live in rented flats and live with their parents, therefore extending the loan would lessen the damage if any of them defaulted.

Thank You