CREDIT EDA CASE STUDY

PROBLEM STATEMENT

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specialises in lending various types of loans to urban customers. You have to use EDA to analyse the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

The data given below contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:

- The client with payment difficulties: he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample.
- All other cases: All other cases when the payment is paid on time.

When a client applies for a loan, there are four types of decisions that could be taken by the client/company):

1. **Approved:** The Company has approved loan Application
2. **Cancelled:** The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.
3. **Refused:** The company had rejected the loan (because the client does not meet their requirements etc.).
4. **Unused offer:** Loan has been cancelled by the client but on different stages of the process.

In this case study, you will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

BUSINESS OBJECTIVE

- 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.
- This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.
- In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.
- The company can utilise this knowledge for its portfolio and risk assessment.

DATASETS

The datasets used for this analysis are:

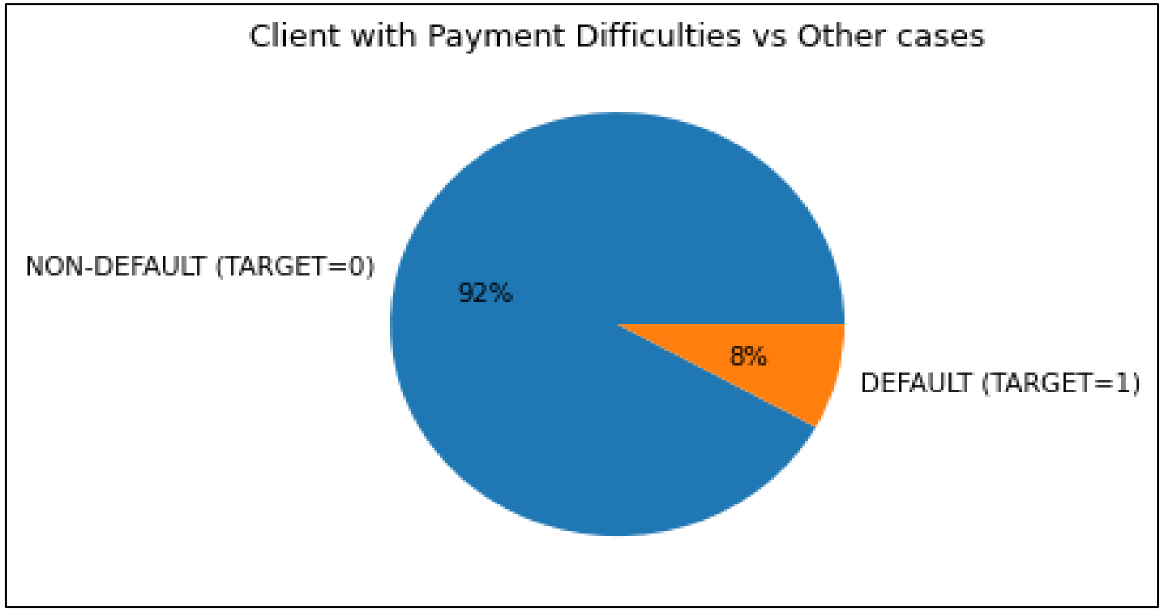
1. 'application_data.csv' contains all the information of the client at the time of application.
The data is about whether a **client has payment difficulties.**

2. 'previous_application.csv' contains information about the client's previous loan data. It contains the data whether the previous application had been **Approved, Cancelled, Refused or Unused offer.**

DATA IMBALANCE

On examining the data and performing the pre-checks on the application_data dataset we observed data imbalance and calculated the Data Imbalance Ratio where:

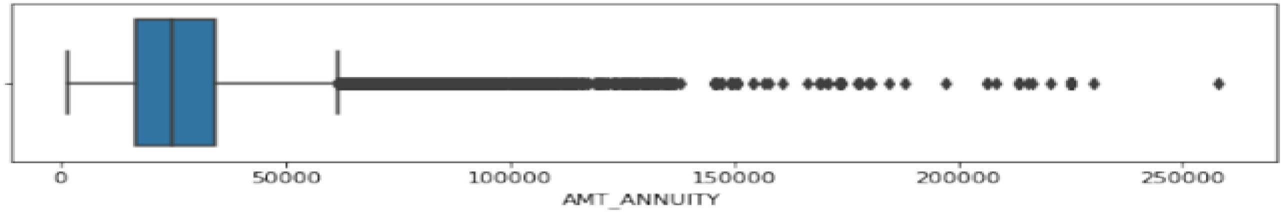
- ❖ Target = 0 refers to Client without payment difficulties. (Non-defaulters)
- ❖ Target = 1 refers to Client with payment difficulties. (Defaulters)

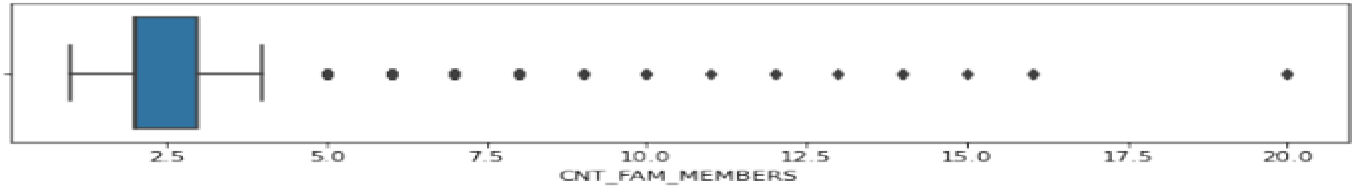


OUTLIERS TREATMENT

Many of the continuous variable columns of the dataset contained outliers and they have been replaced either by mean or median of the column.

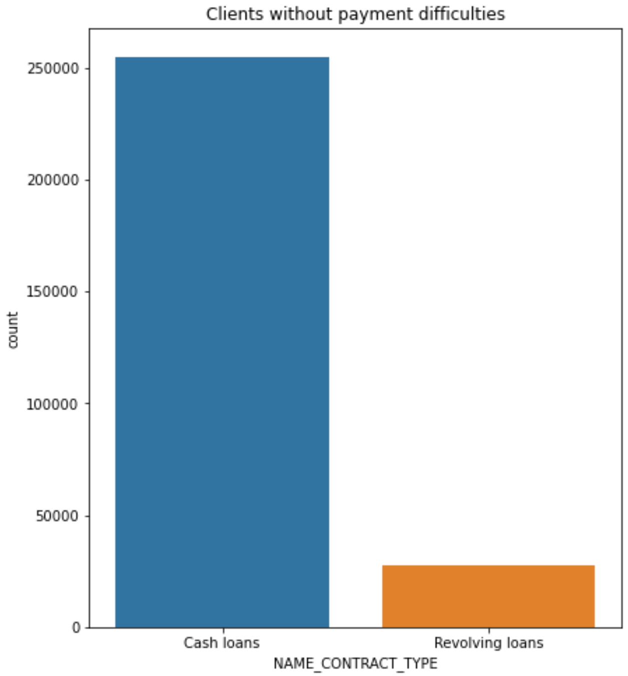
Few of the examples of outliers are as depicted below:

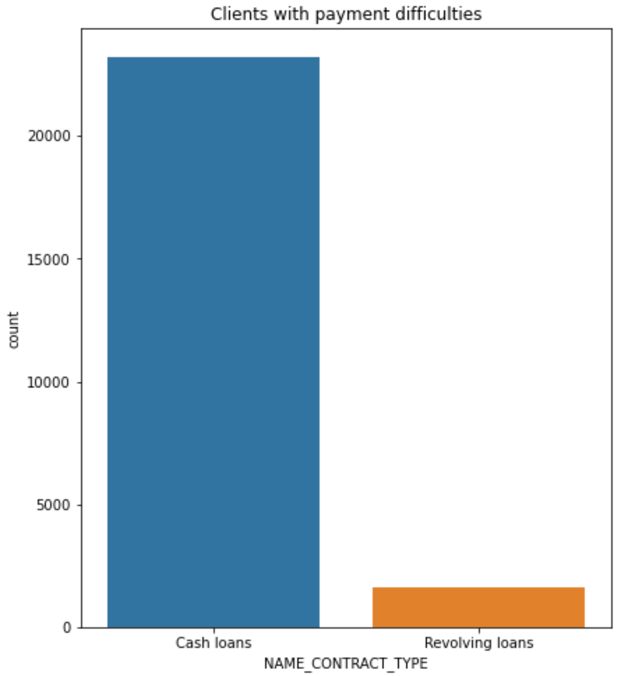




ANALYSIS OF THE APPLICATION DATA

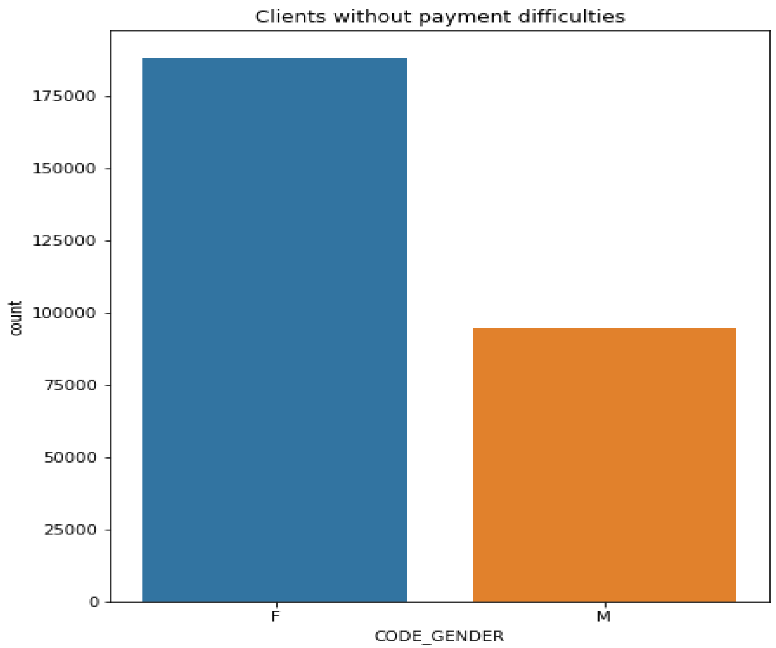
NAME_CONTRACT_TYPE ANALYSIS WITH RESPECT TO TARGET VARIABLE:

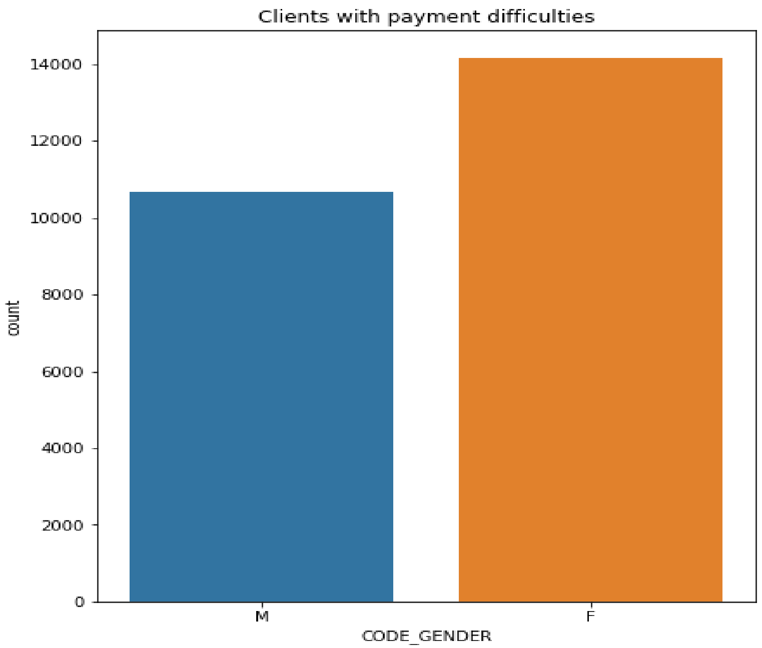




From the analysis of the Contract type with respect to Target we can observe that most of the applicants apply Cash Loans compared to Revolving Loans.

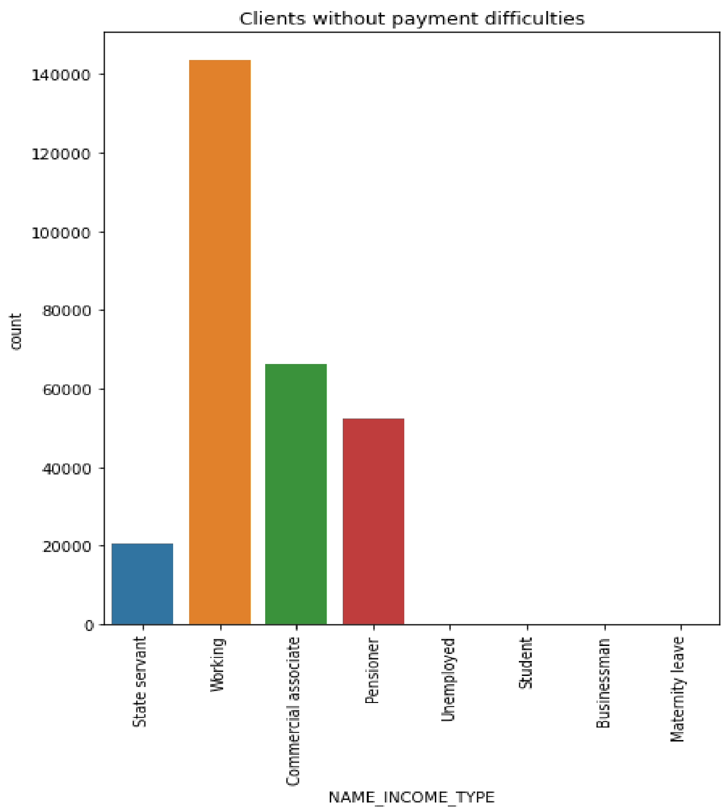
GENDER COLUMN WITH RESPECT TO TARGET VARIABLE:

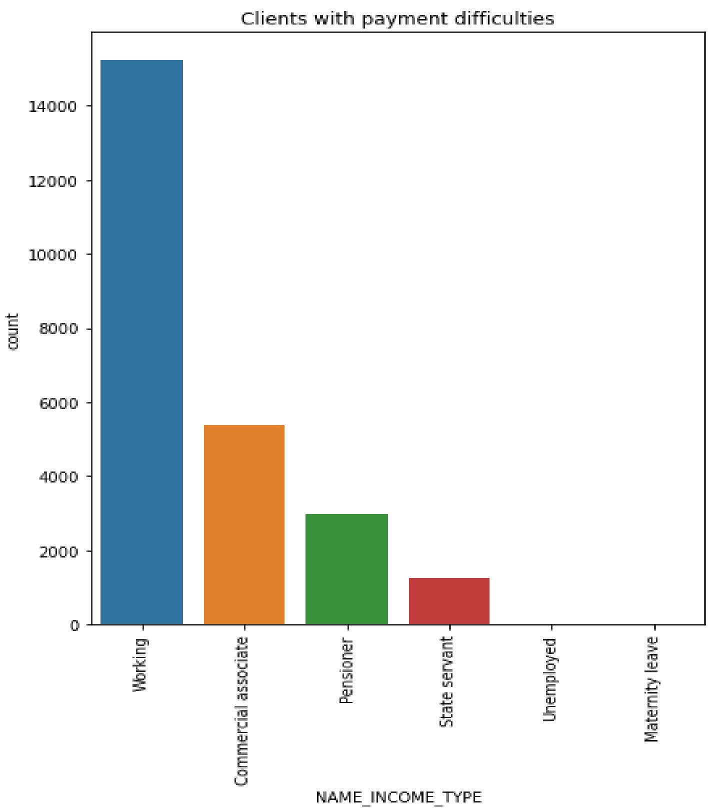




From the Gender plotting we observe that the Female clients applying loans are of higher number than Male Clients.

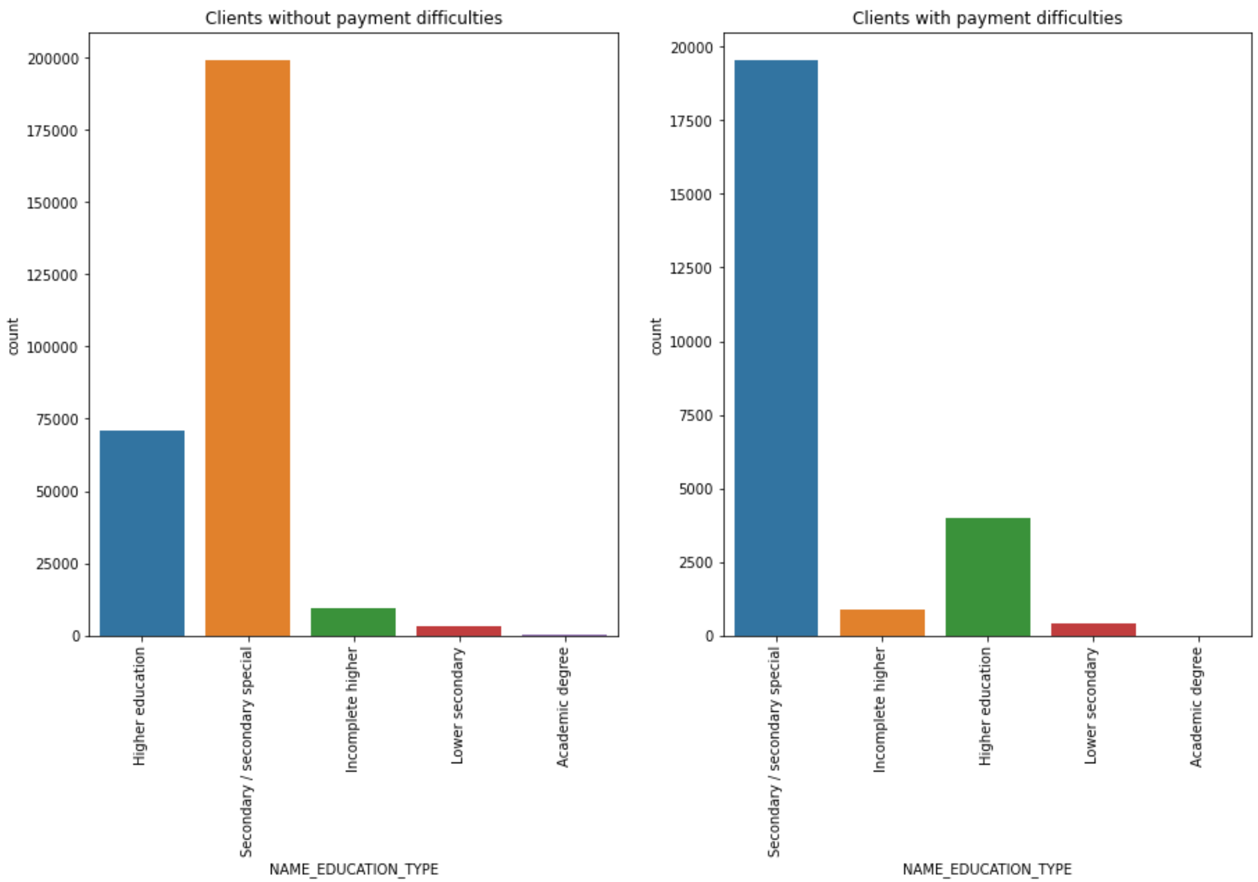
INCOME TYPE ANALYSIS OF THE DATA:





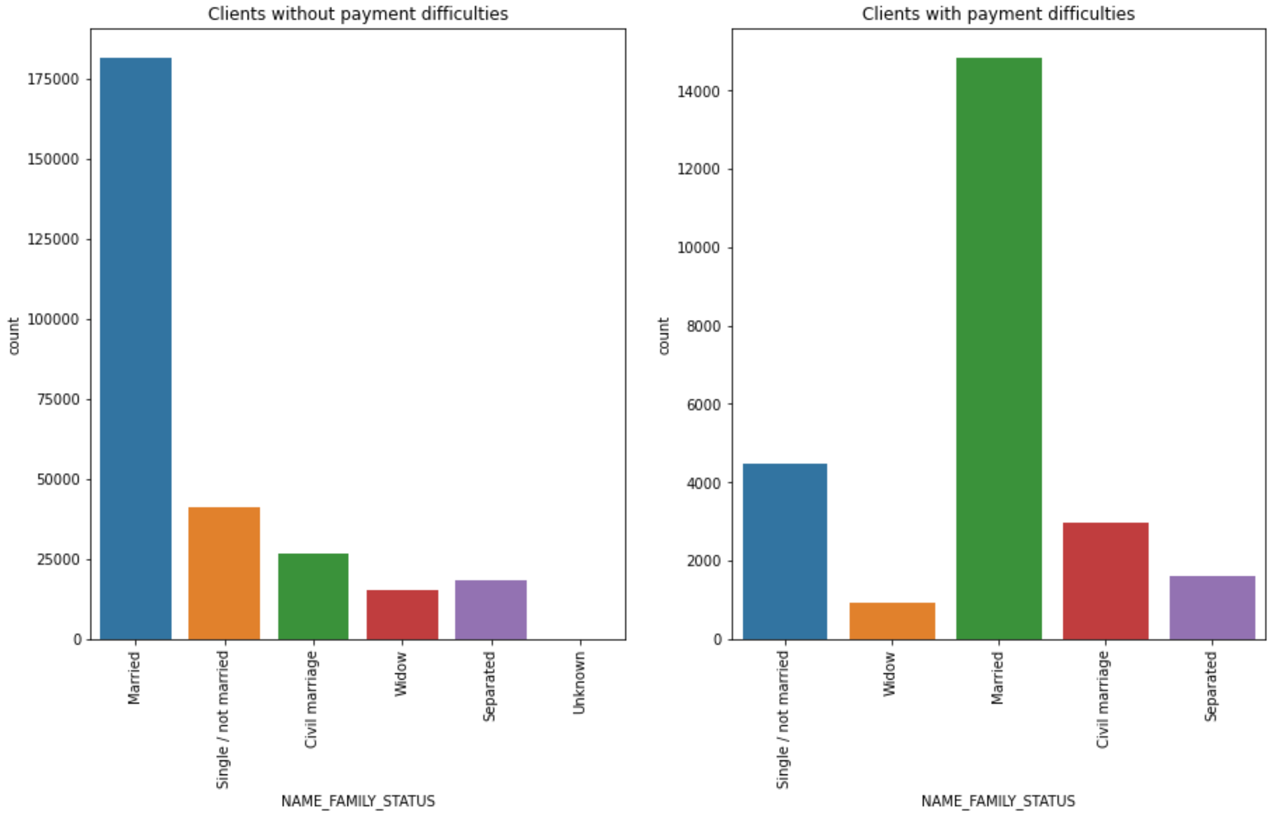
Income type Analysis proves that Working people take more loans but are also most likely to default.

ANALYSIS USING EDUCATION LEVEL OF APPLICANTS:



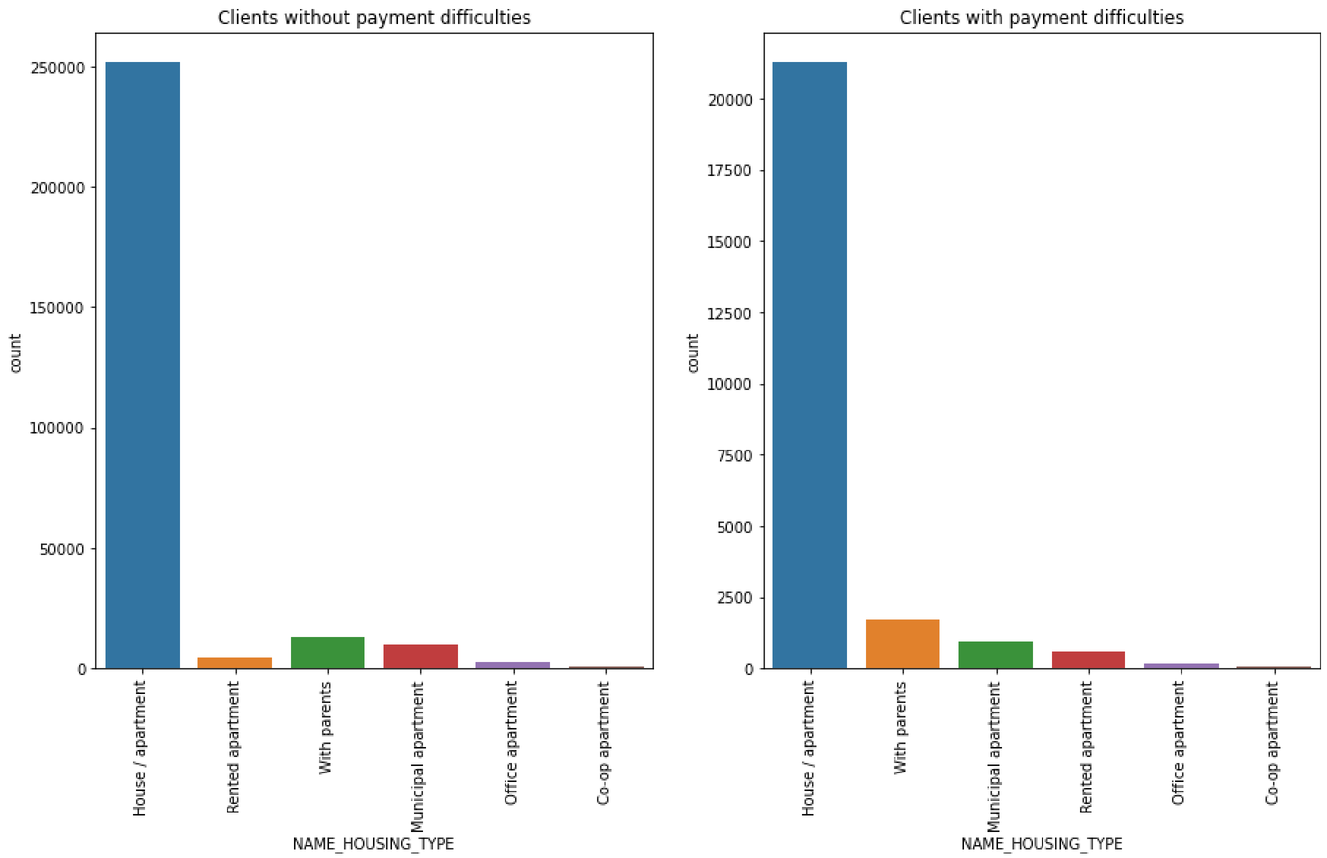
The clients with Secondary / Secondary special education level are most prone to default the loan. Higher Education clients can be targeted for loans as they have less difficulties in repaying the loan.

FAMILY STATUS OF THE APPLICANTS:



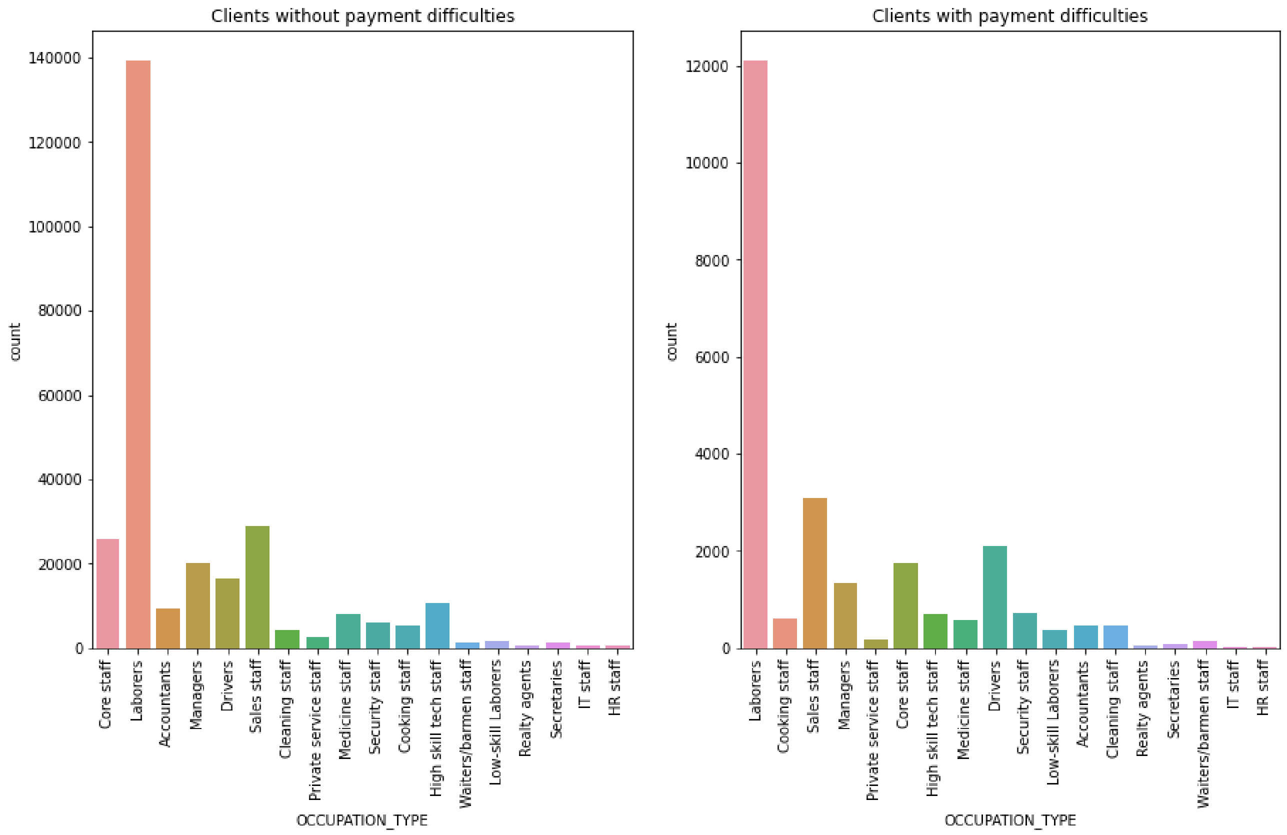
More number of Married clients apply for loans as well as face difficulties in repaying the same. Bank can target Single clients as they are more likely to repay the loans.

HOUSING TYPE OF THE CLIENT:



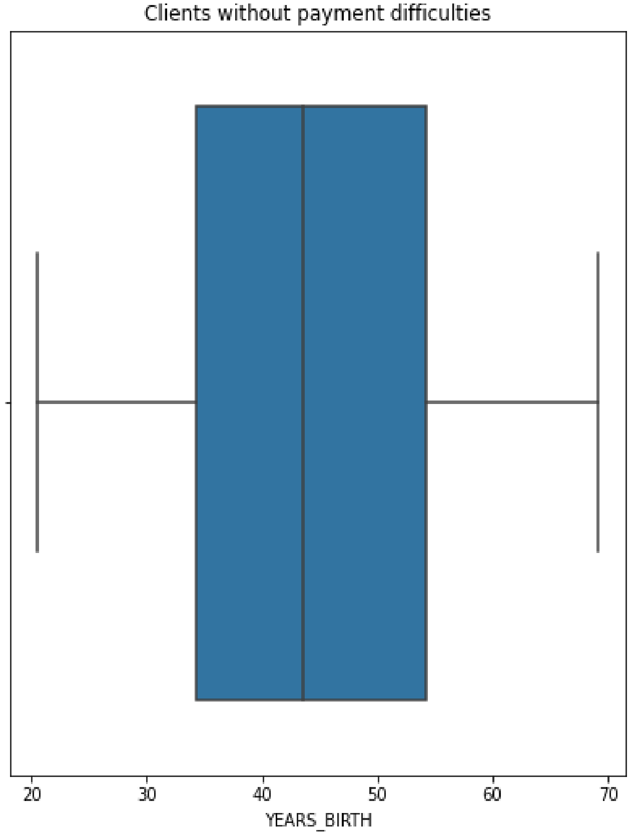
Clients living in a House / Apartment have more chances of taking a loan and also are more prone to defaulting the loan. Bank can target the clients living With Parents who have a greater chance of repaying.

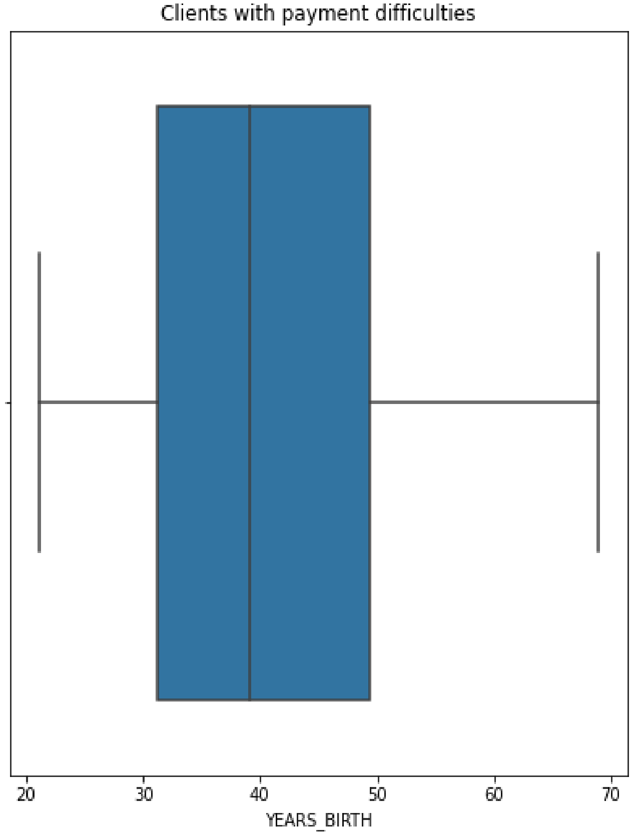
OCCUPATION TYPE OF THE APPLICANTS:



Labourers category take more loans and also have difficulties in repaying them. Core staff can be targeted for loans as they have less difficulties in repaying.

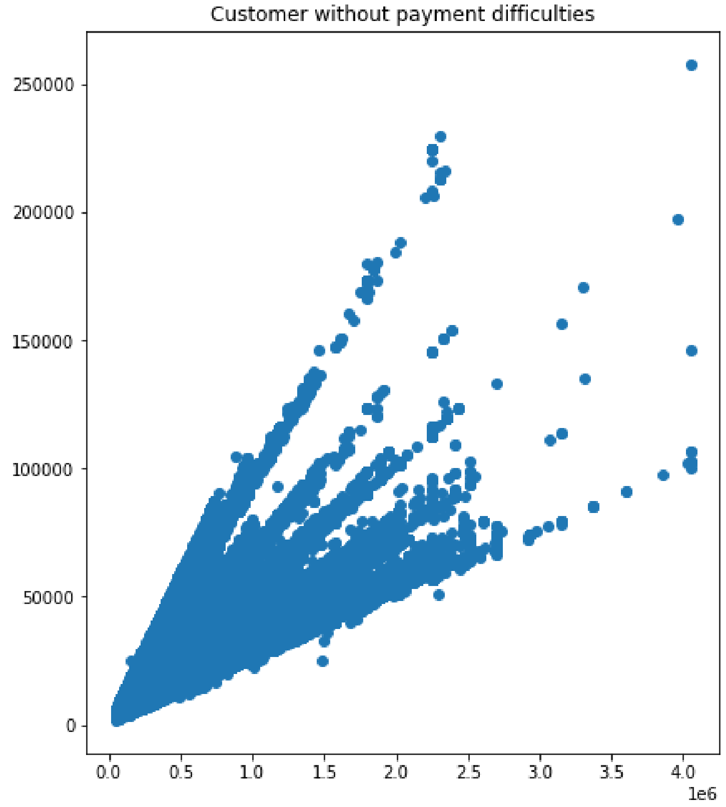
AGE GROUP ANALYSIS OF THE APPLICANTS:

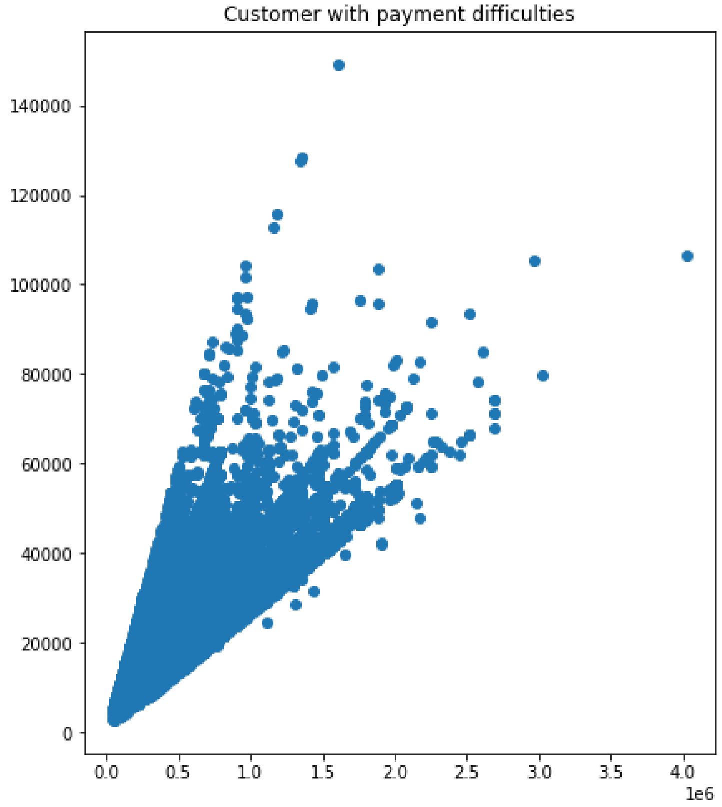




Clients between the age group of 30 and 50 are the ones who have payment difficulties.

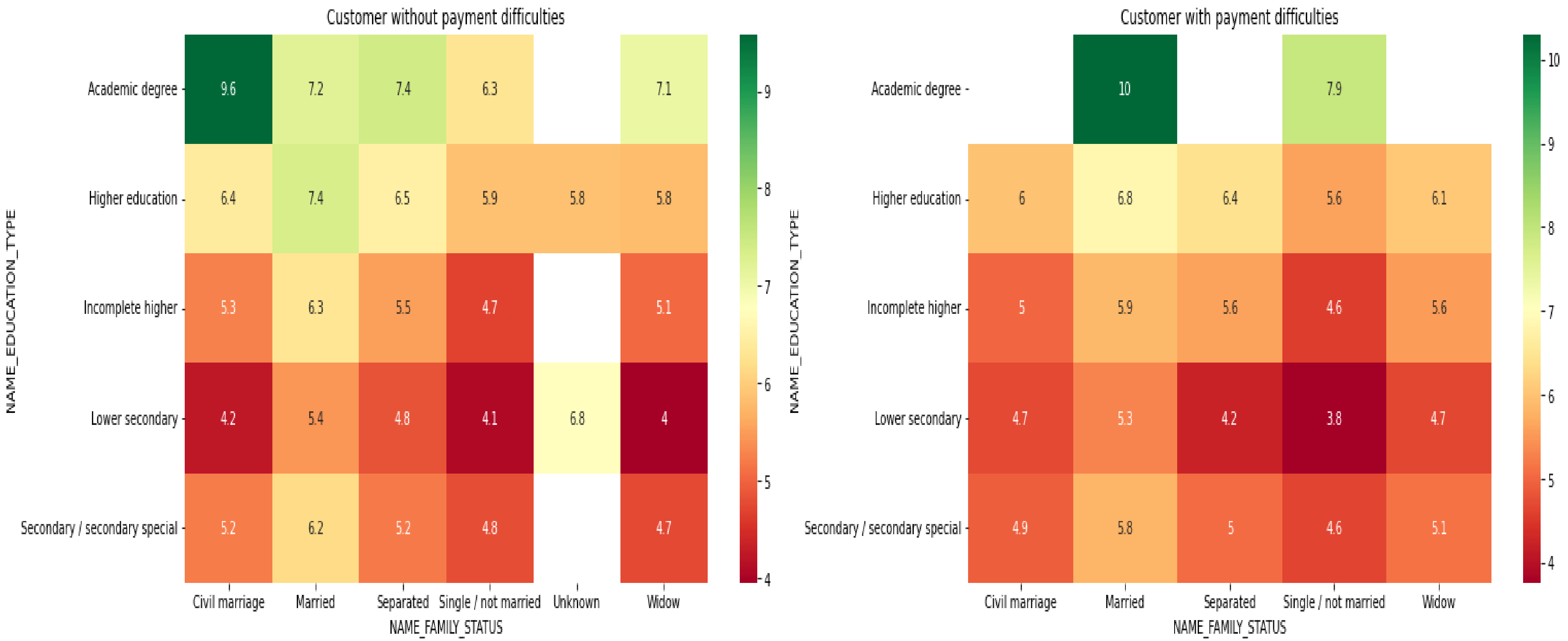
CREDIT WITH RESPECT TO GOODS PRICE:





As per the above scatter plotting, the credit amount of the loan increases in correlation with the price of the consumer goods.

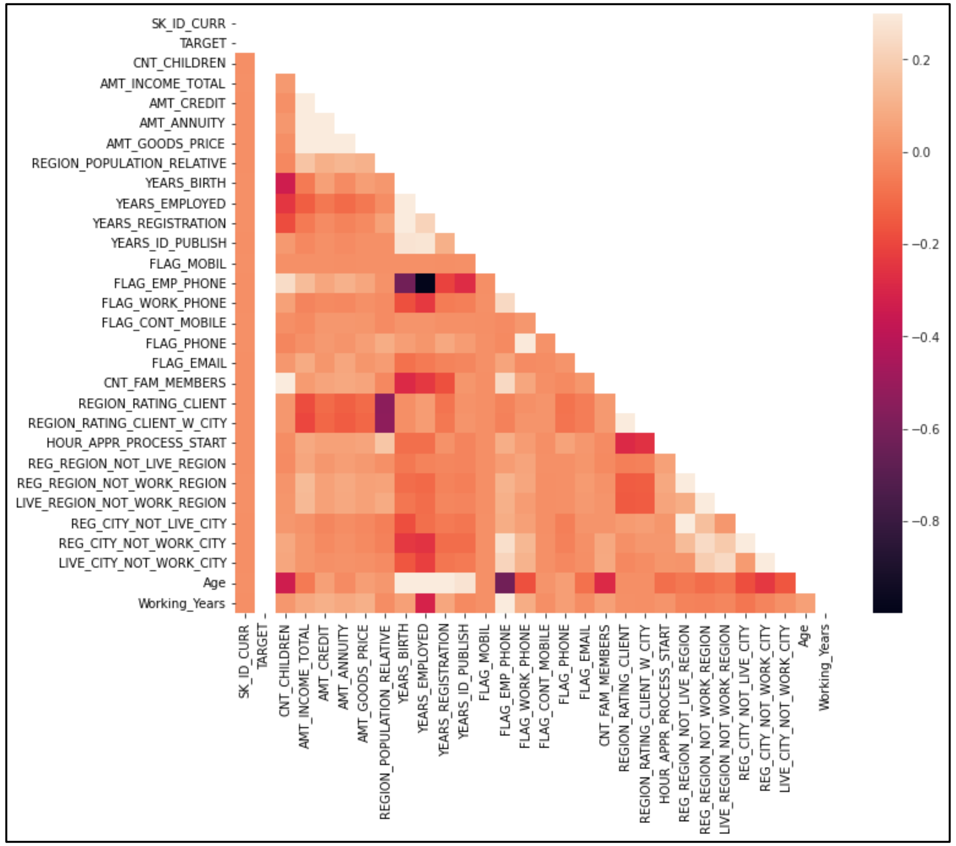
EDUCATION vs FAMILY STATUS vs CREDIT AMOUNT:



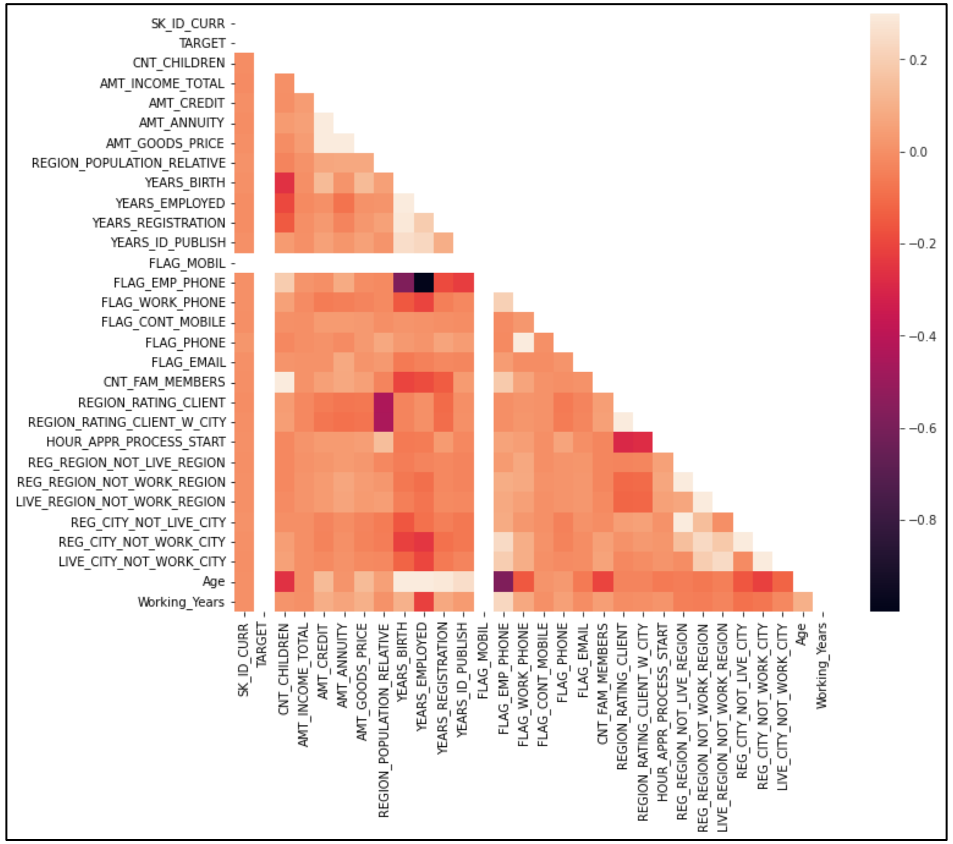
Married clients who have an Academic Degree are more likely to apply for a higher credit amount.

CORRELATION FOR TARGET VARIABLE

NON-DEFAULTERS:

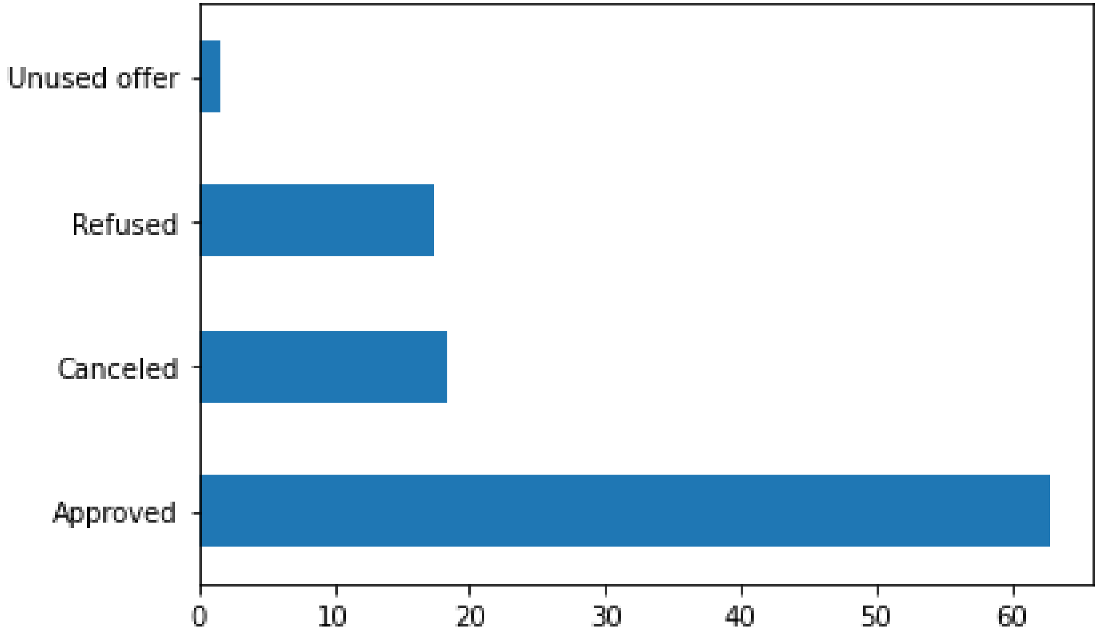


DEFAULTERS:



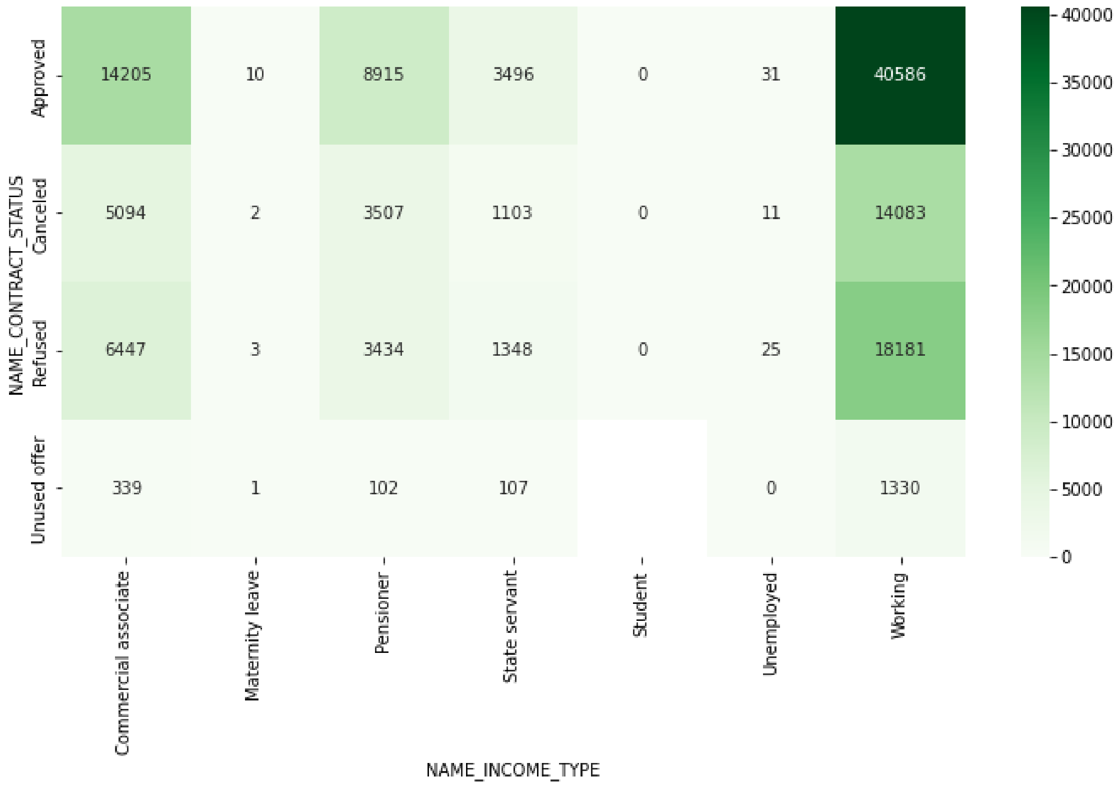
ANALYSIS OF THE PREVIOUS APPLICATION

The variables in consideration are: Approved, Cancelled, Refused and Unused offer.



The above bar graph shows that more than 60% of the loans are Approved in the previous application. The Cancelled and Refused loans are more of the same percentage and the Unused Offer are very low in number.

CONTRACT STATUS vs INCOME TYPE vs TARGET:



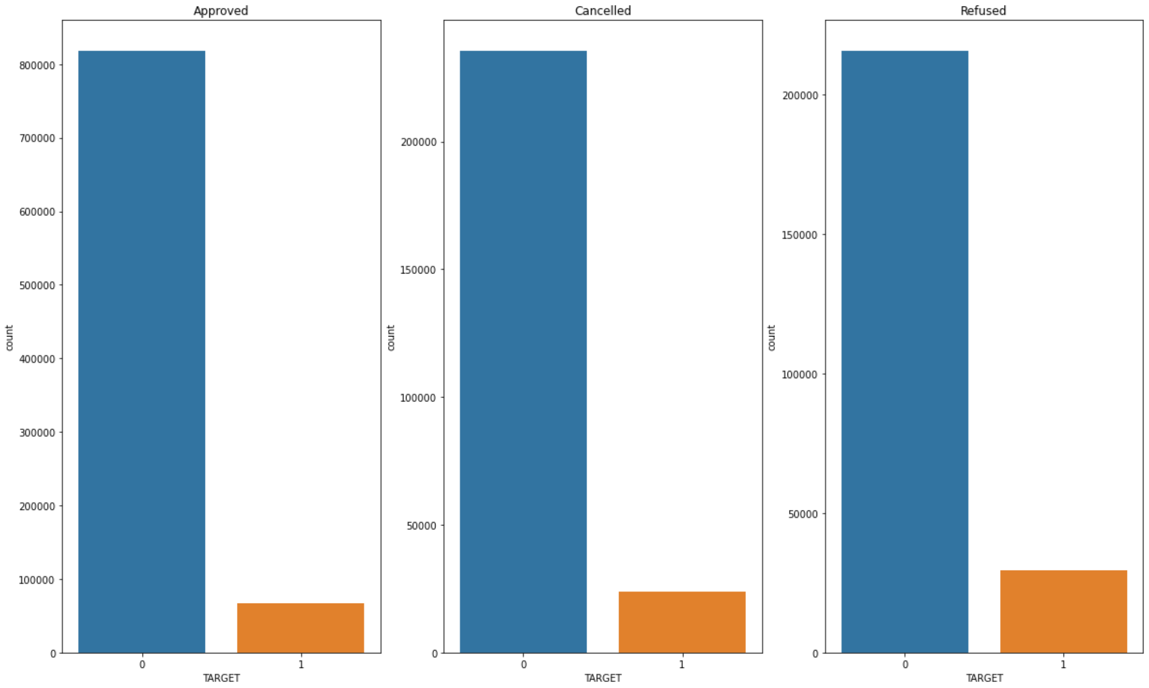
The highest number of defaults have been from the Working category with Approved loans compared to the other categories. Moreover 18,000+ working people have been refused of the loan in the previous application.

CONTRACT STATUS vs EDUCATION TYPE vs TARGET:

Applicants of the Secondary / Secondary Special education type are more likely to default even after their loans being Approved in the previous application. Around 23,000 applicants were refused and for around 19,000 applicants, loans were cancelled in the previous application.

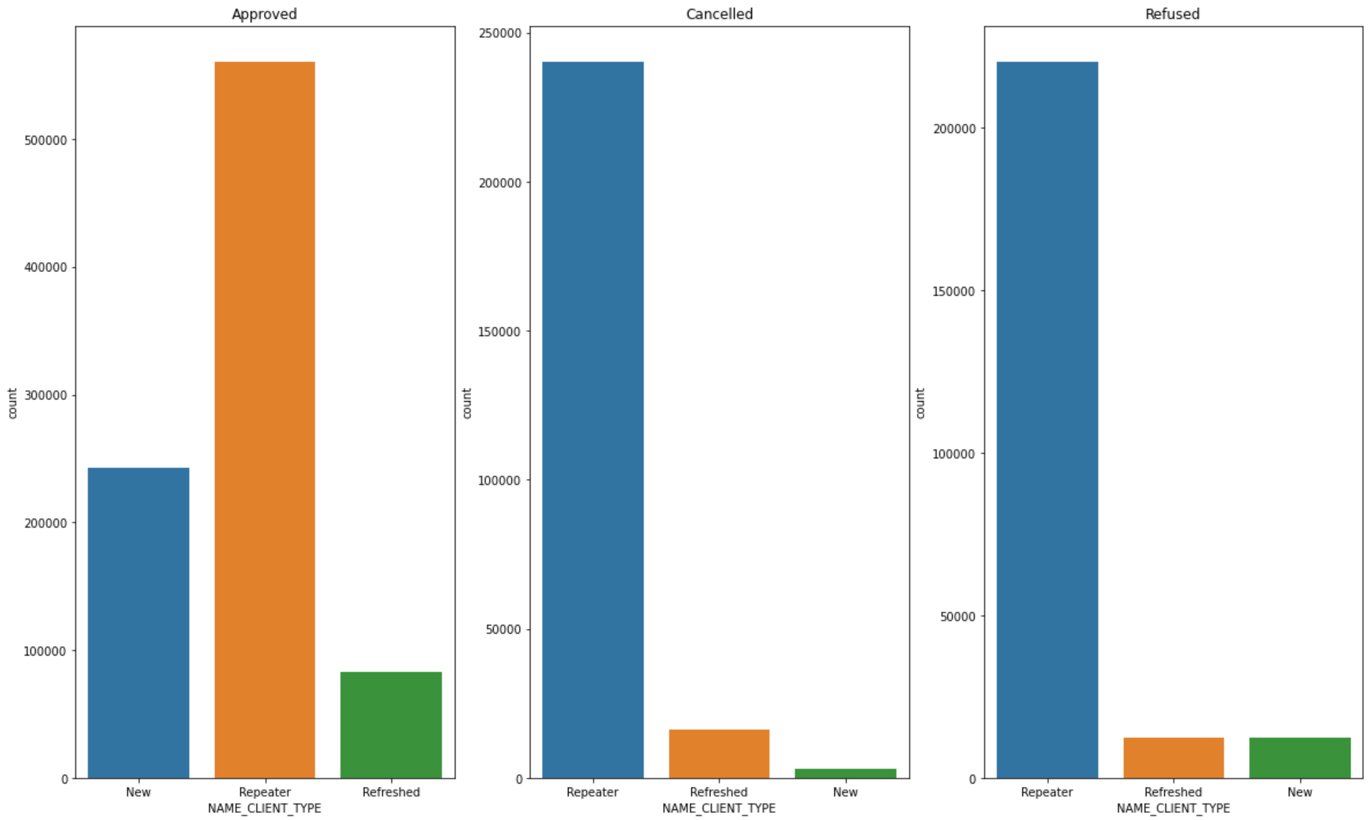
ANALYSIS AFTER MERGING THE DATAFRAMES

CONTRACT STATUS COMPARISON WITH TARGET:



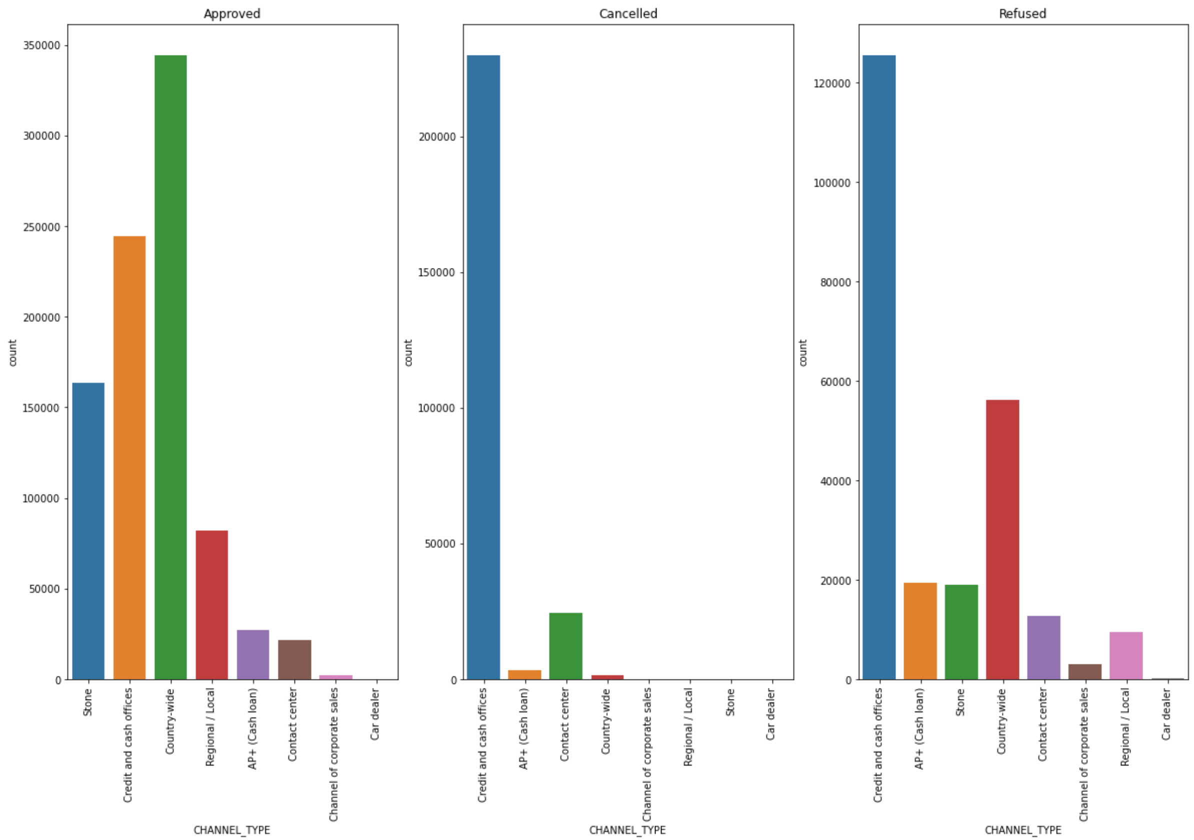
Cancelled and Refused category in the previous application has more number of defaulters.

NAME_CLIENT_TYPE COMPARISON:



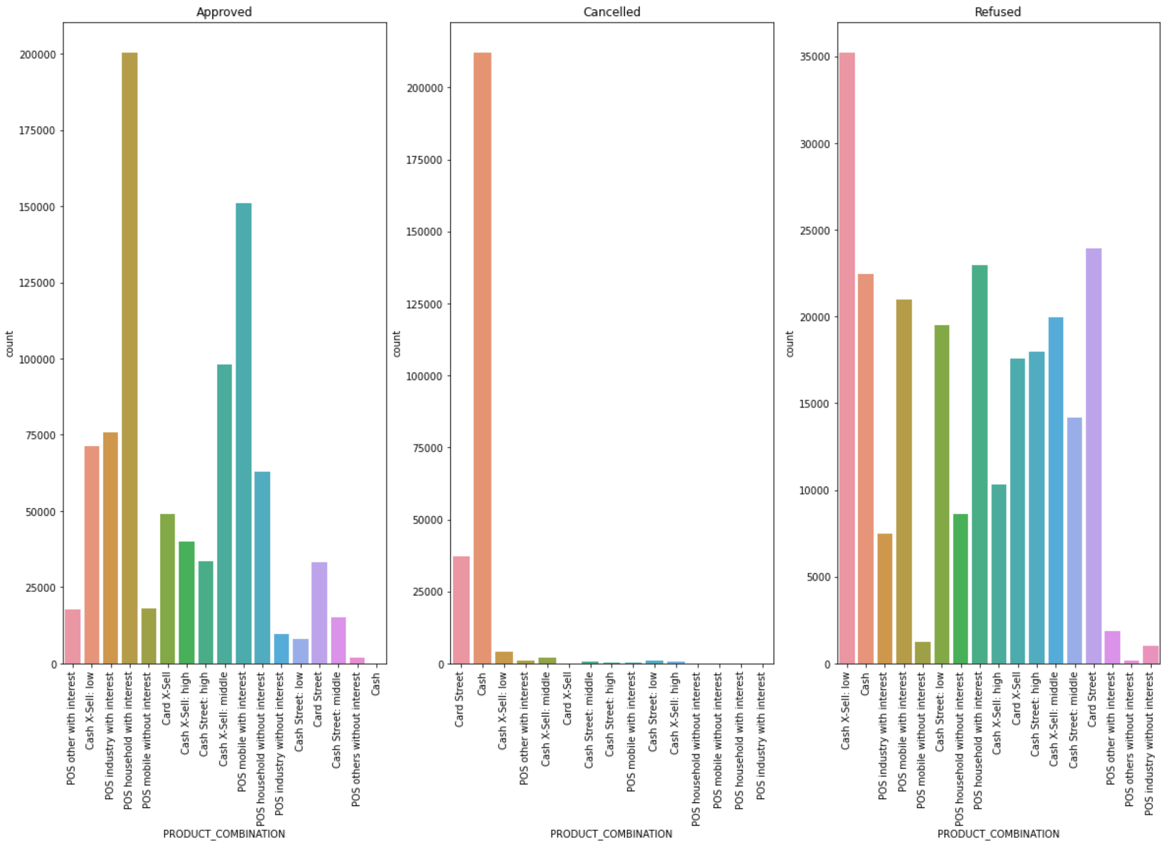
Repeater type client has more number of loan applications in the previous application. More number of New clients are also getting their loans Approved.

CHANNEL TYPE COMPARISONS:



More number of loans are approved in the Country-Wide channel type. Through Cash and cash offices channel more number of loans are Refused and Cancelled.

PRODUCT_COMPARISON ANALYSIS:



POS household with interest category sees more loans Approved, and POS mobile with interest follows not far behind. Cash category witnesses more Cancellations. Cash X-Sell:low category loans are mostly refused.

CONCLUSIONS

- Working people take more loans but also default in large numbers, so it is not wise for the bank to target them for loans.
- Secondary / Secondary special educated applicants are also more prone to default. Higher Education clients can be targeted for loans as they have less difficulties in repaying the loan.
- People living in House/Apartments take more loans but are also prone to default the loan. So people living with parents and in Co-op apartments are more likely to pay back the loan.
- Bank should focus more on Country-wide channel type as it sees more number of Approved loans. Whereas, Credit and cash offices channel type sees more number of Cancelled and Refused loans.
- Cash category in Product Combination sees more Cancelled loans.
- Bank can target Married clients with Academic Degree as they have a larger chance of applying for loans.
- Labourers category take more loans and also have difficulties in repaying them. Core staff can be targeted for loans as they have less difficulties in repaying.
- The age group of 30-50 should be less targeted as they have higher chances of defaulting.