



EDA CREDIT ASSIGNMENT

By

Dr.Nandini Arvind



Introduction

- This case study aims to give you an idea of applying EDA in a real business scenario.
- In this case study, apart from applying the techniques ,we will develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

Business Objectives

- 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.

Business Understanding

- 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 to their advantage by becoming a defaulter.
- Suppose you work for a consumer finance company which specializes in lending various types of loans to urban customers. You have to use EDA to analyze 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.

Aim:

- Find the driving factors behind loan default.
The variables which are strong indicators of default.
- Insights should explain why the variable is important for differentiating the clients with payment difficulties with all other cases.
- The company can utilize this knowledge for its portfolio and risk assessment.

THE DATASETS CONTAIN

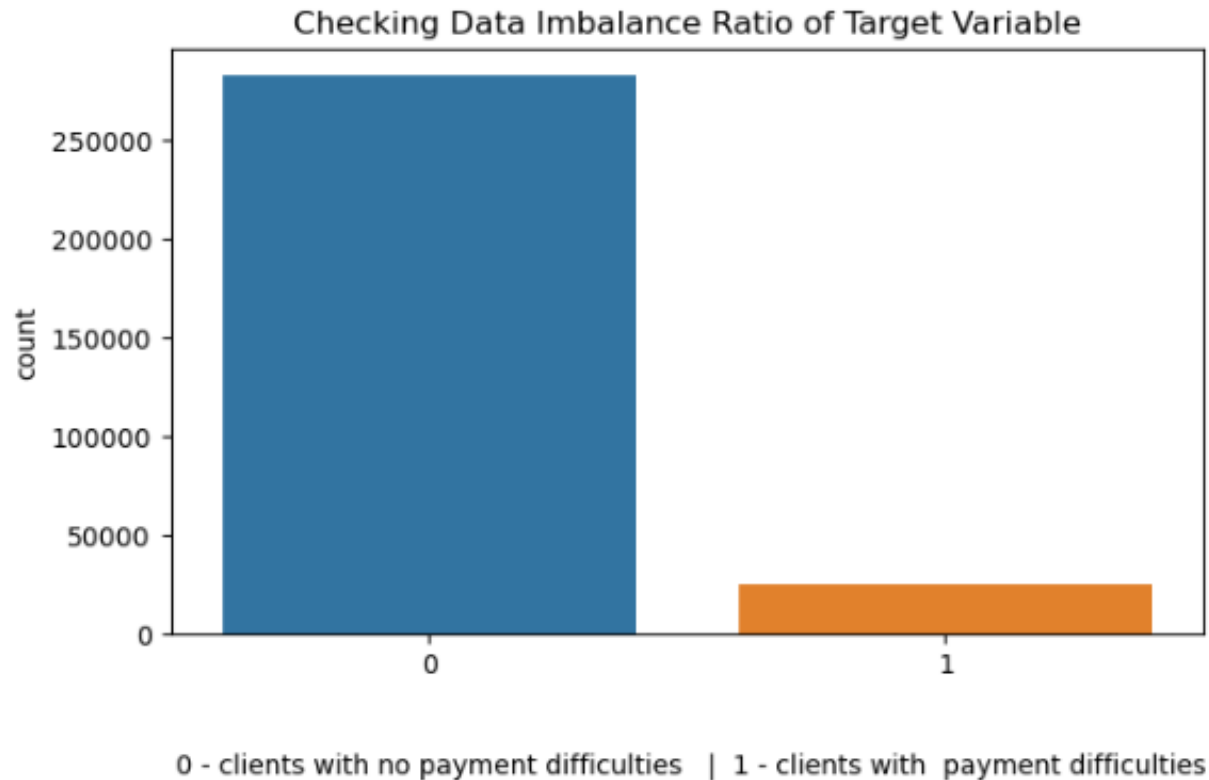
1. All the information of the client at the time of application.
The data is about whether a client has payment difficulties.
2. Contains information about the client's previous loan data.
It contains the data on whether the previous application had been Approved, Cancelled, Refused or Unused offer.

Approach

- Importing the dataset provided
- Data cleaning and pre processing steps
 - Inspecting rows, columns, basic information ,type of the column
 - Finding null values and its percentage,
 - Deleting columns with null values > 40%
 - Dealing with outliers and suggested possible treatment/correction required
 - Finding statistical summary.
 - Making standardization of values at required places
- Exploratory Data Analysis
 - Perform univariate, bivariate,multi variate analysis. Based on numerical/continuous and categorical variables.
 - Plotting various graphs and visualizations to analyze , derive insights from the dataset.
 - Present key summary statistics, such as mean, median, mode for important variables.
- Concluding with the main insights and recommendations.
- Find the top 10 correlation for the **Client with payment difficulties and** and derive insight if there.

The clients without payment difficulties are grouped as target 0
the clients with payment difficulties are grouped as target 1

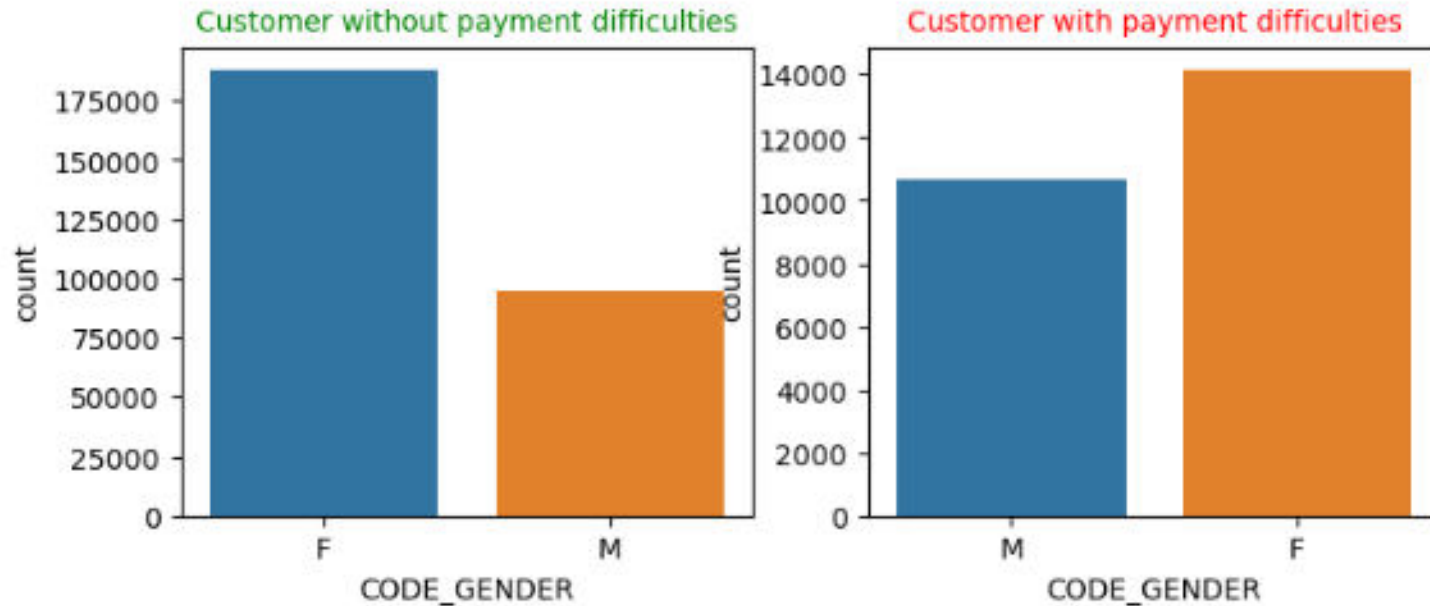
Checking data imbalance of target variable



Ratio of data imbalance: 11.387150050352467

Which means 1 in every 11 applicant has payment difficulty.

Target group Vs Gender of the client



Insight:

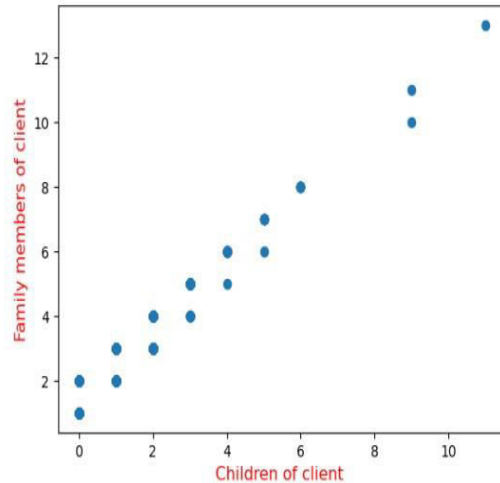
- Female customers are more in number in both the targets.
- And hence the payment difficulties is more in females relatively.
- The defaulters ratio is more in females than non Defaulters.

Suggestions/plans/changes that can be done from bank side:

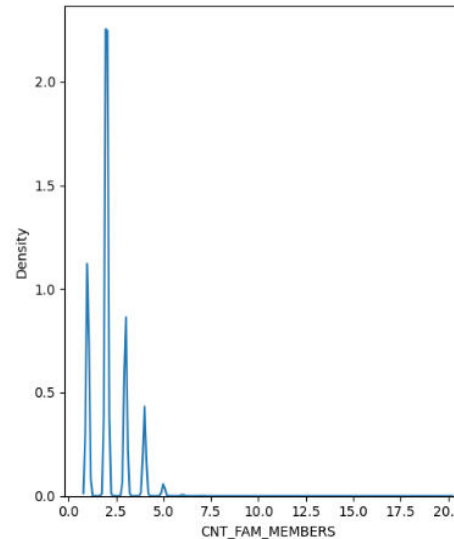
- Take necessary measure in approving loans to capable females and focus on successful repayments.

Target group Vs number of members in family of the client

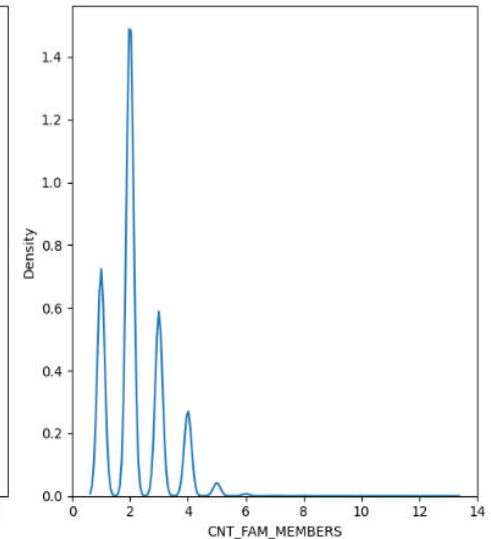
Children vs Family members of client with payment difficulties



CNT_FAM_MEMBERS for Customer without payment difficulties



CNT_FAM_MEMBERS for Customer with payment difficulties



Insight:

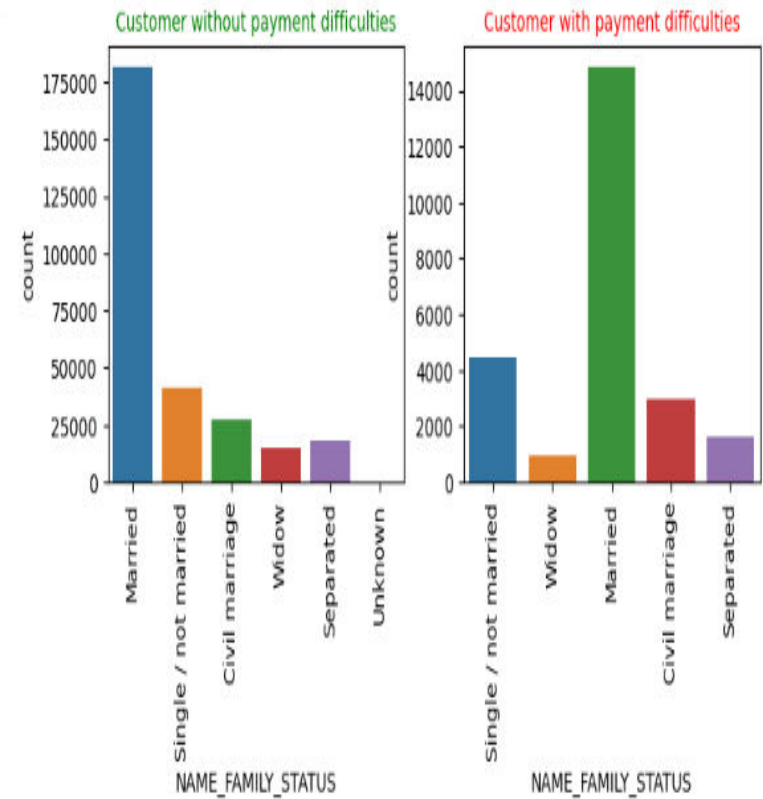
- High positive correlation with "causation." in children and family members.
- Because of increased children the family size increased.
- No much significant correlation in family members and payment difficulty.
- Hence family members count, children count is not a great indicator.

Target group Vs Education, Family status of the client



Insight:

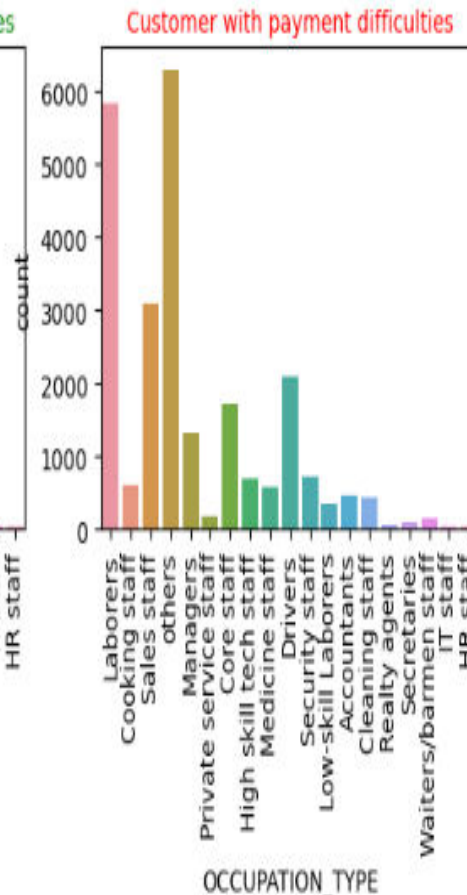
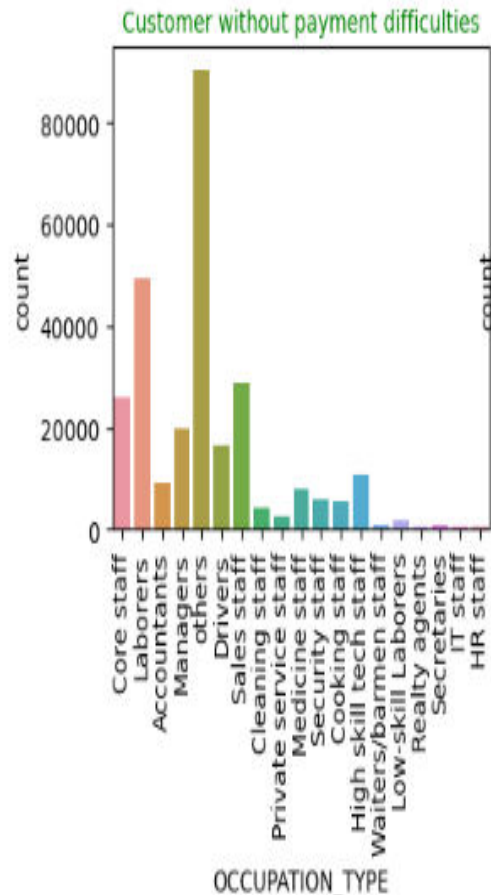
- Secondary / secondary special has repayment defaults .
- Next defaulter is Higher education group comparatively.
- Academic degree is least in number in both the targets.



Insight:

- Married people prefer loans mostly because of their requirement and tend to be more defaulter.

Target Vs Occupation, Income type of the client



Insights :

- Most loan clients are laborers: True because they have low income relatively.
- Least loan clients are IT staffs.

Defaulters :

- High among laborers, sales staffs, drivers and core staffs respectively.
- Unknown group: occupation not mentioned; high defaulters too.

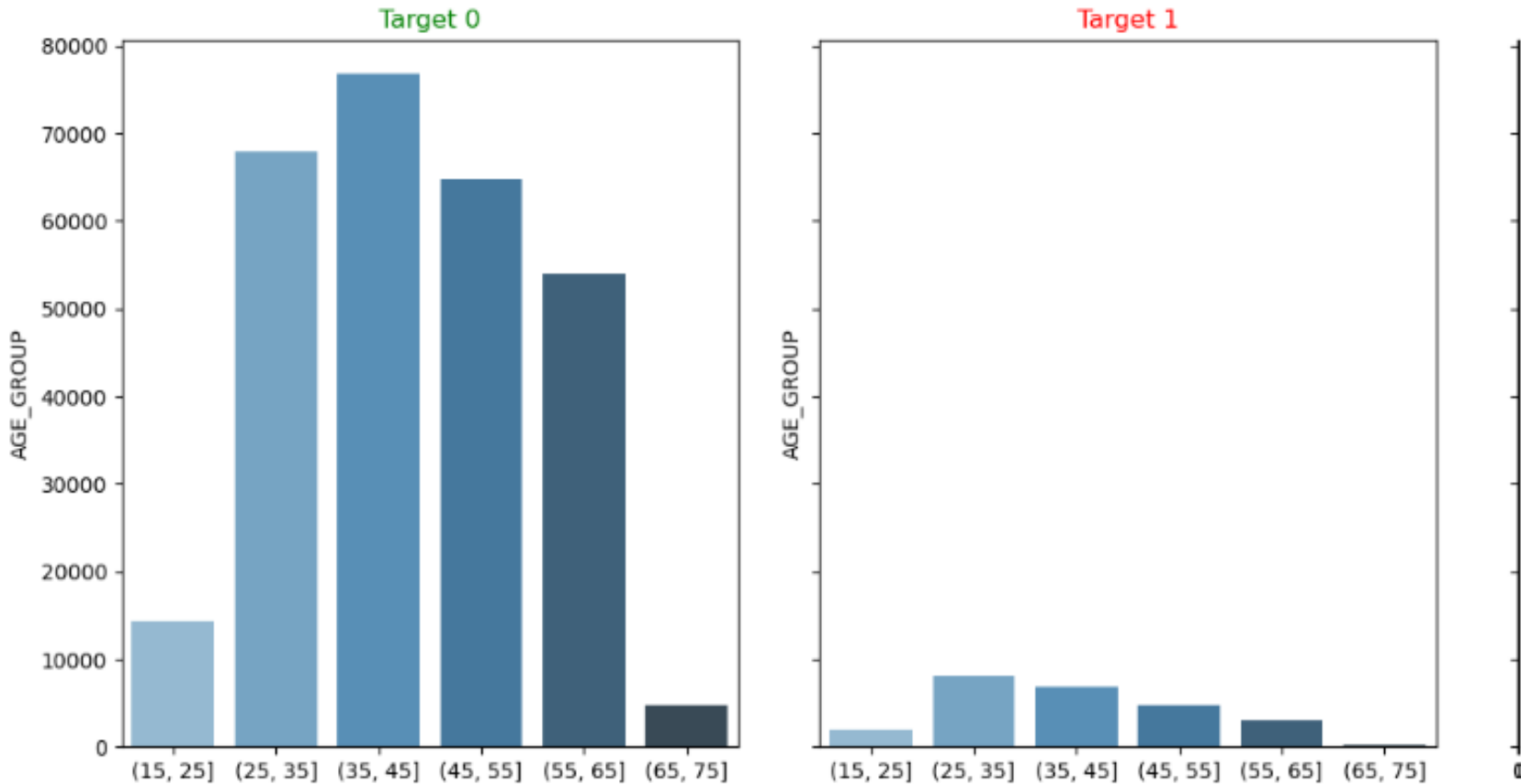
Insight:

- Working people is higher in count in both target.
- Hence Working people are defaulting more.



Target group Vs Age group of the client

Graph for : AGE_GROUP

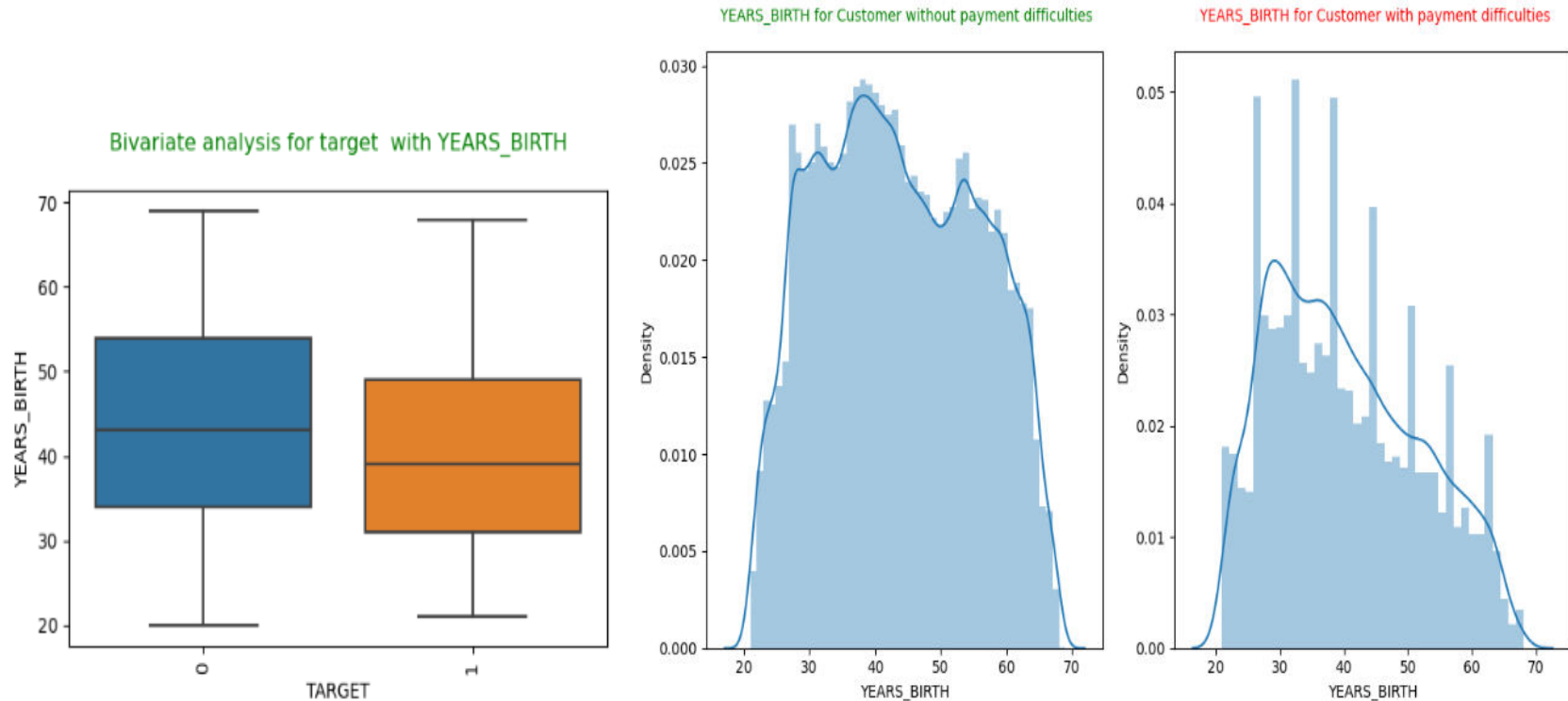


Insight:

- AGE group :

- [35-45] years people have taken more loans and more in target 0.
- Young age 25-35 years face difficulty in repayment.

Target group Vs Age of the client

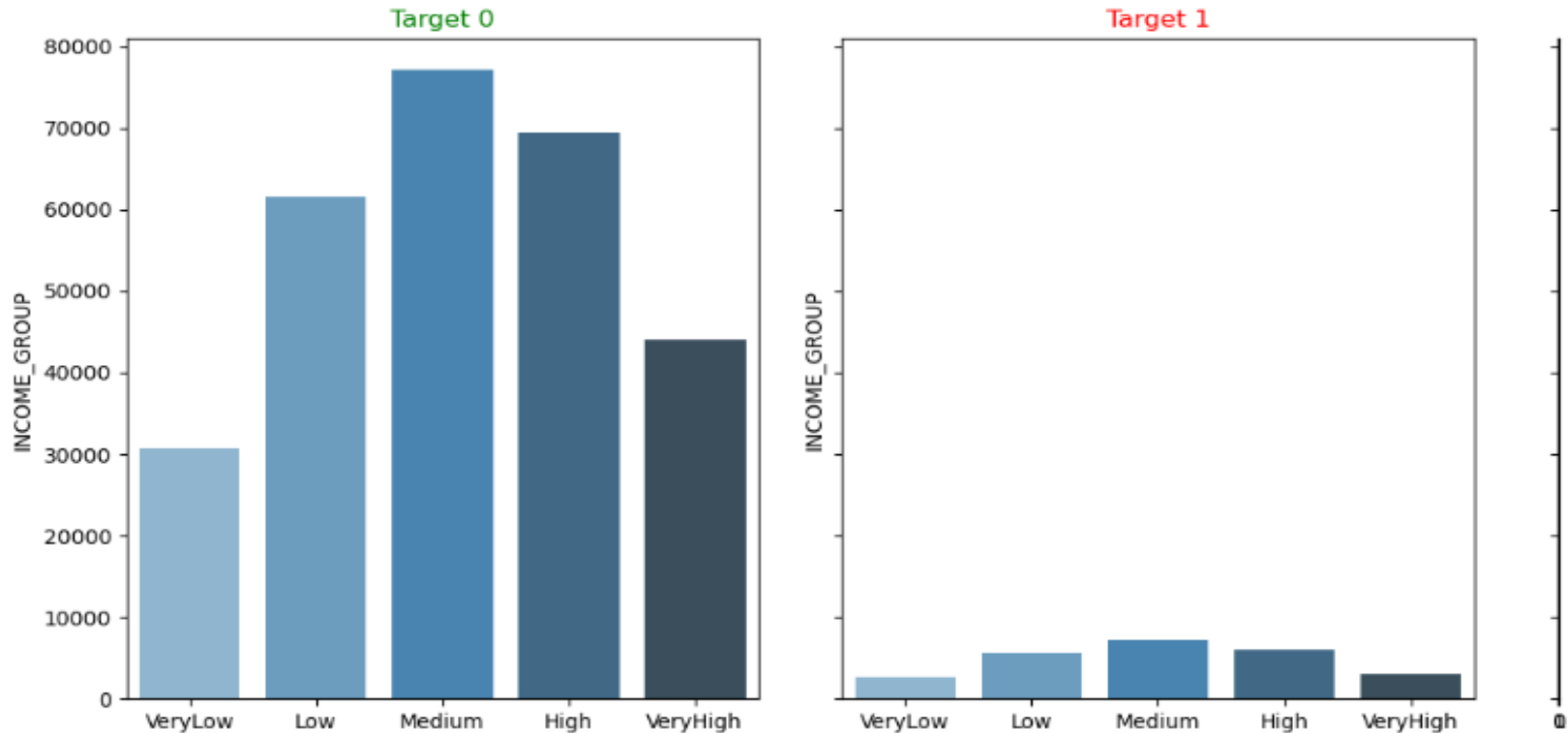


Insight:

- Repayment difficulties are seen in younger population relatively. with age range 31-49 years.
- Old age people are repaying better.
- Mean and median are comparatively higher in repayment group.
- Age is a good indicator for successful repayments.

Target group Vs Income Group of the client

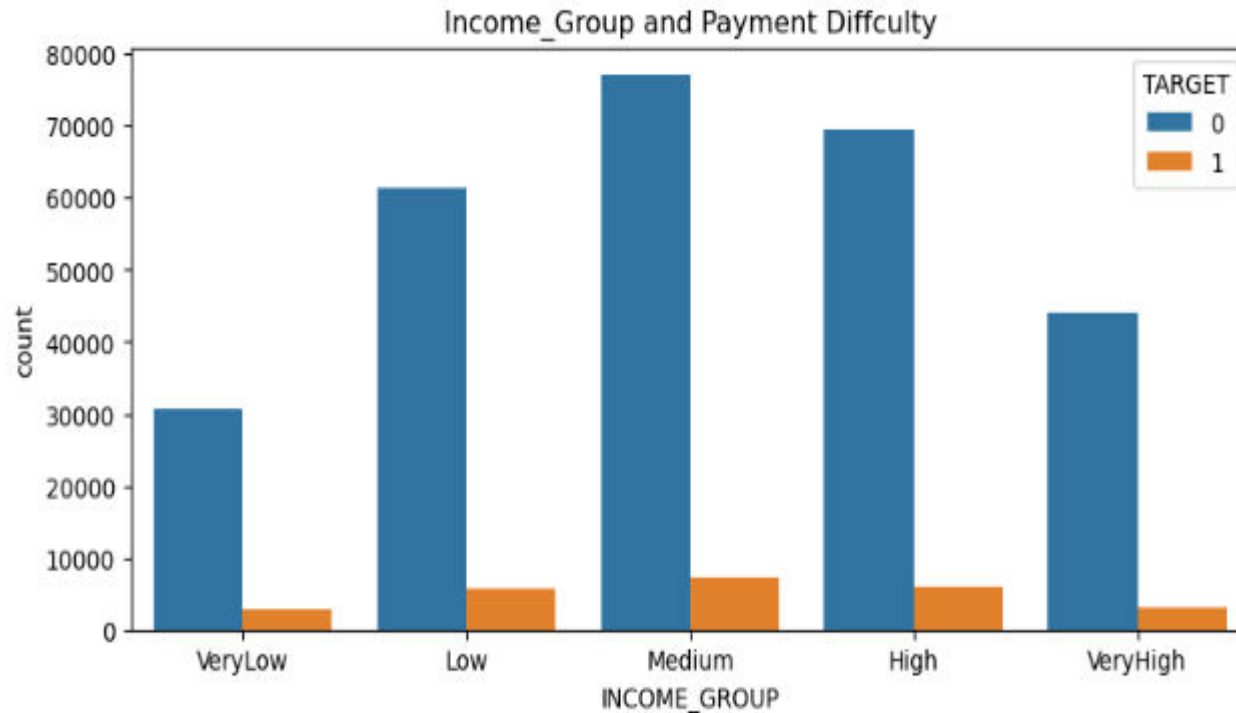
Graph for : INCOME_GROUP



Insight:

- Medium income group are more in number in both the Target groups.
- Medium income people group is the largest group applying for loans.
- Low income people tend to be defaulter. It is possible mostly.
- Difficulty in repayment is seen in higher salary group also.

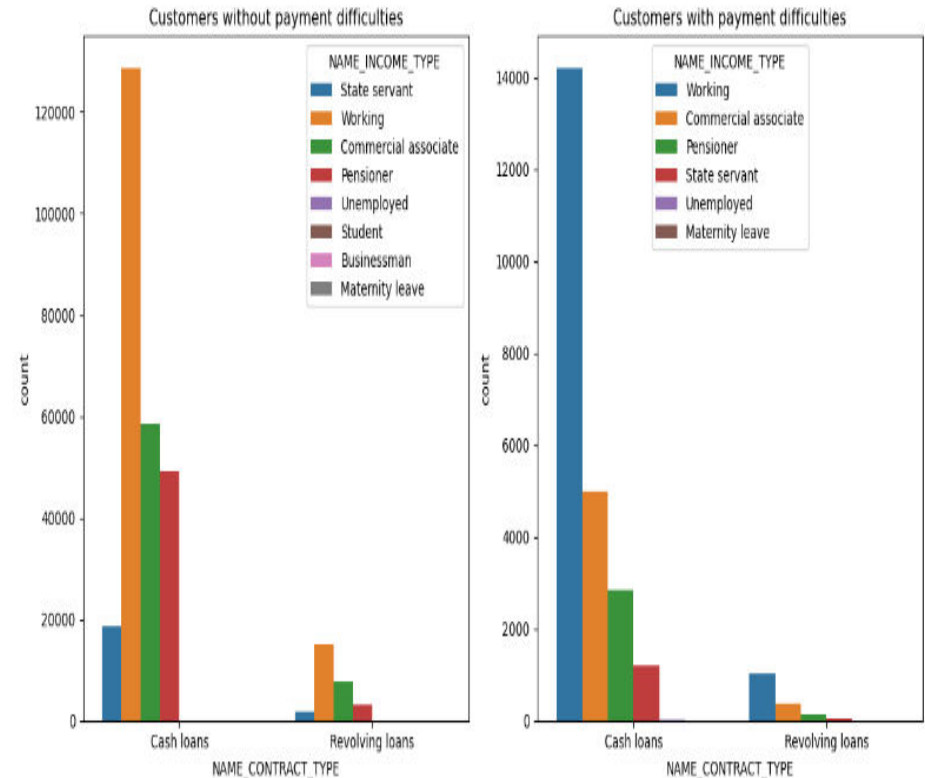
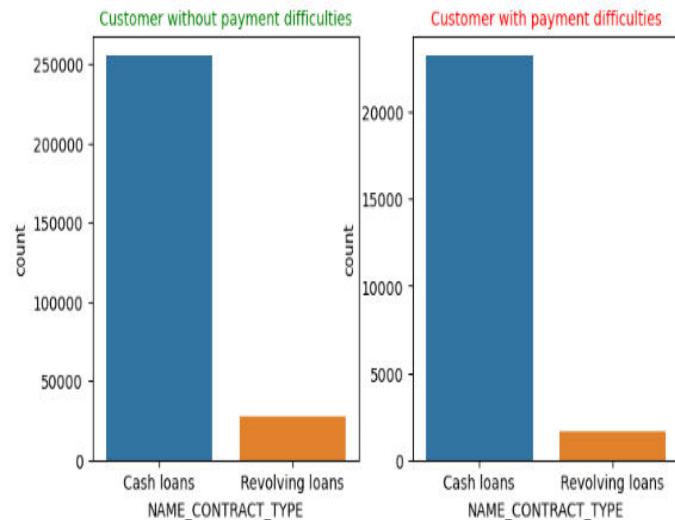
Comparison of income group in target 0 and 1



Suggestions:

- Medium income group are a good target to give loan as their repayment rate is also higher relatively.
- Higher salary group should be focused more to repay and make them to be nil defaulter.

Target group Vs Contract type, Income type of the client



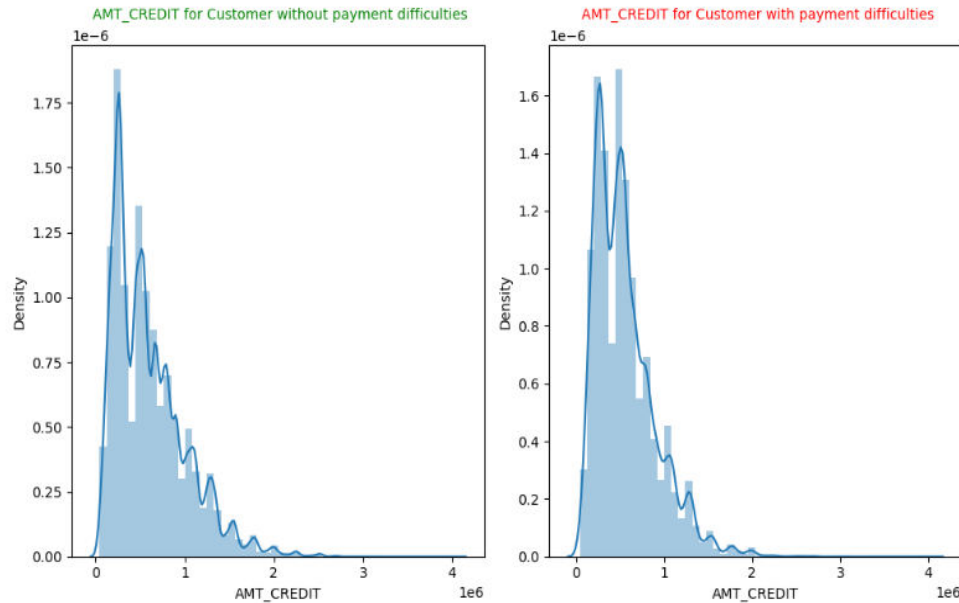
Insight:

- Cash loans is preferred in both targets than revolving loans.
- Repayment difficulty is more seen in cash loans.

Insight :

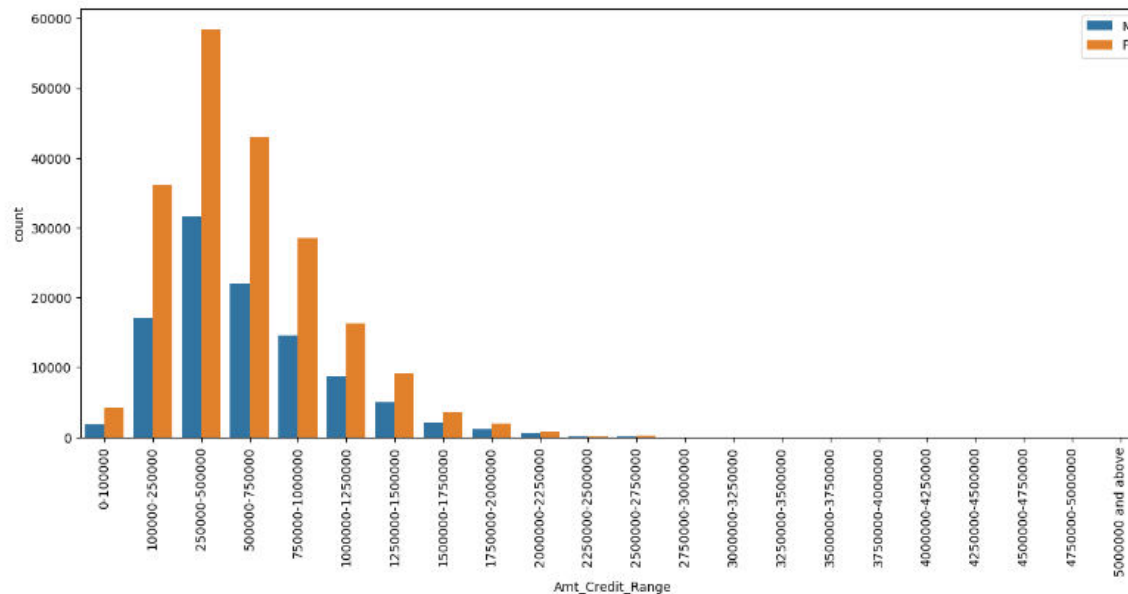
- Most taken are cash loans especially from working group.
- People without facing payment difficulties have revolving loans more than other case.
- No much difficulty in repaying revolving loans

Target group Vs Credit Loan Amount in Gender



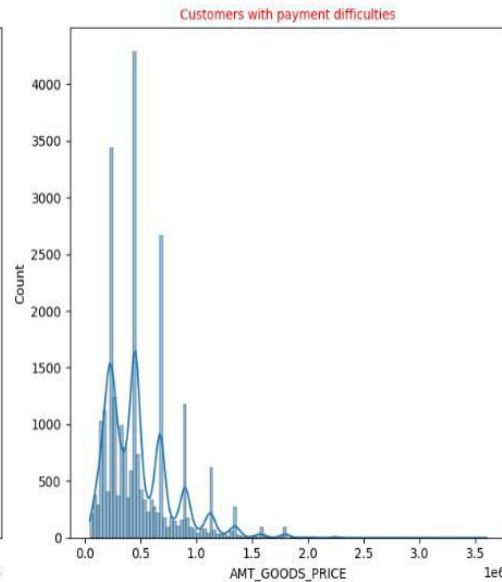
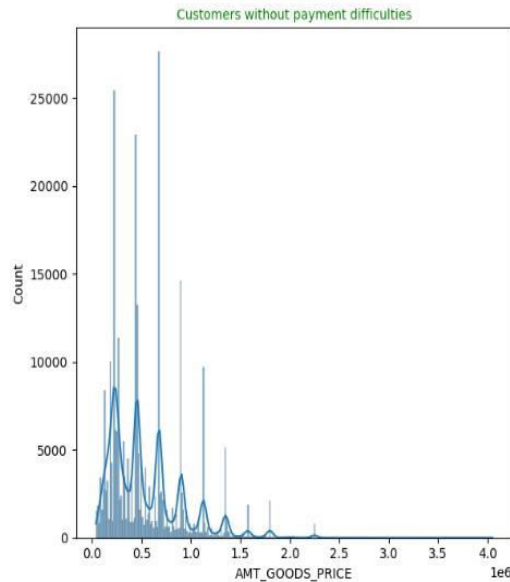
Insight:

- As the loan credit amount increases, the repayment is also improved.



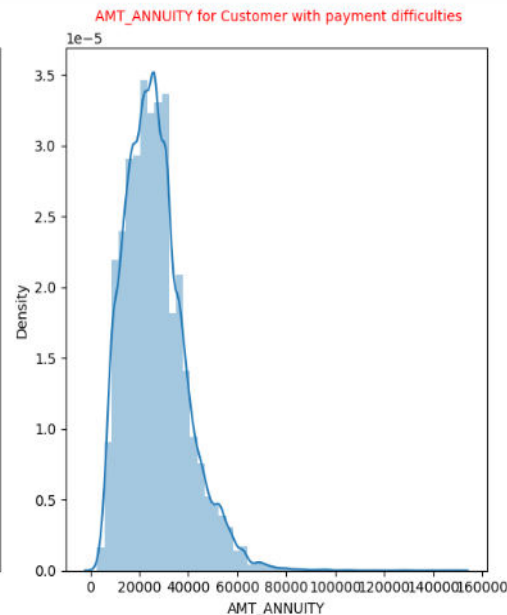
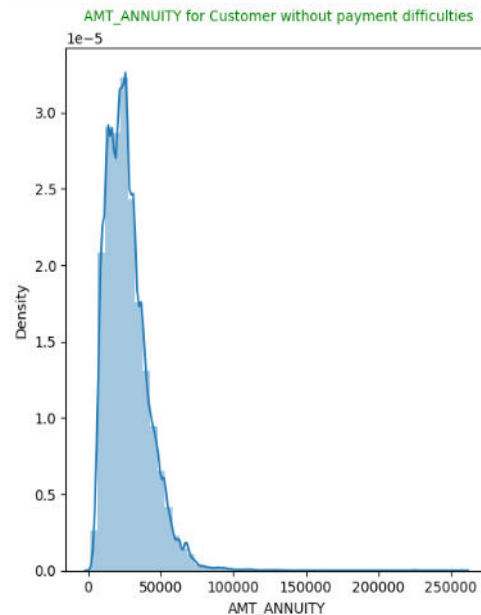
Credit amount loan is high among females than males.

Target group Vs Goods Price, Annuity of Loan Amount



Insight:

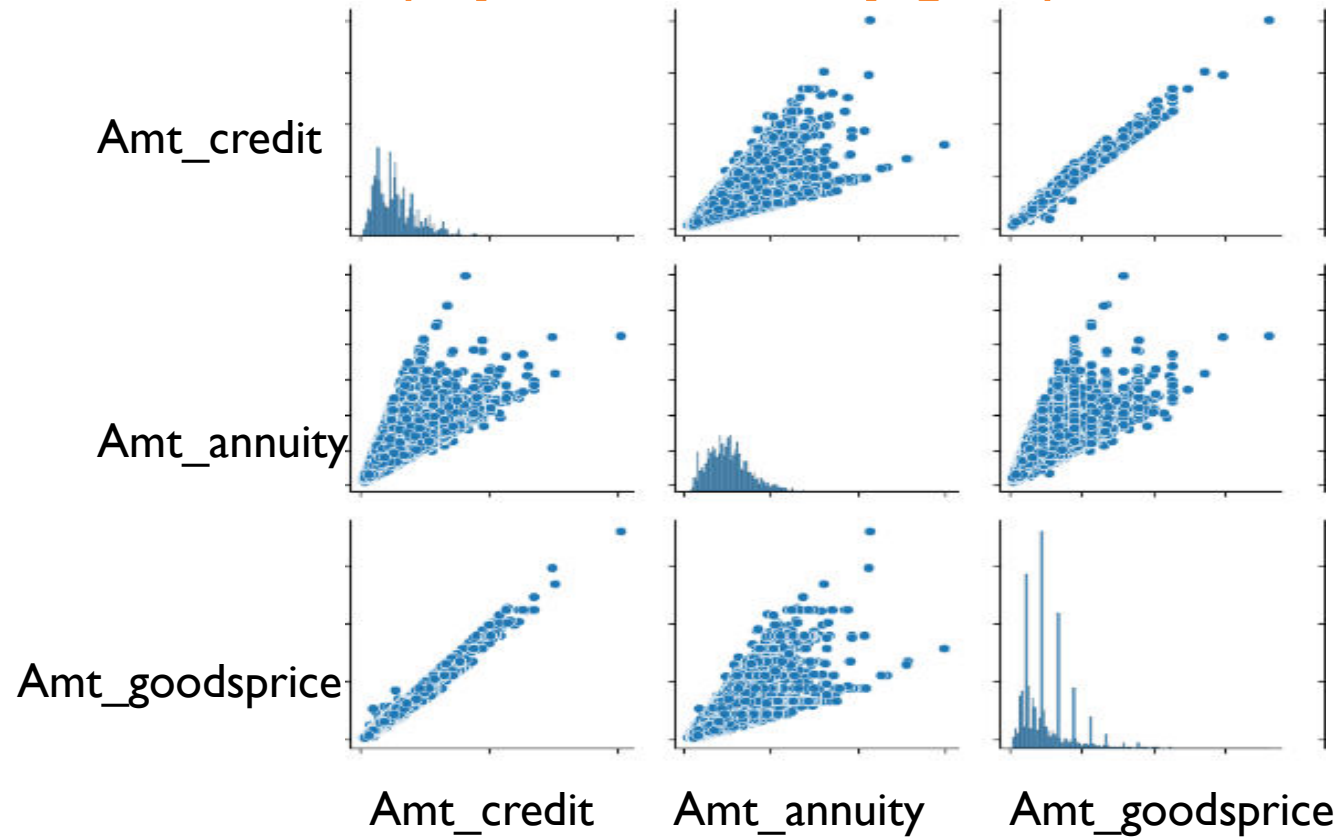
- As goods price increases, credit loan amount given also increases and hence the repayment is also difficult.
- so goods price is directly related to repayment.



Insight:

- As the loan amount annuity increases, the repayment is also increased.

Comparison of how Amount credit , Annuity, Goods price contribute in payment difficulty group



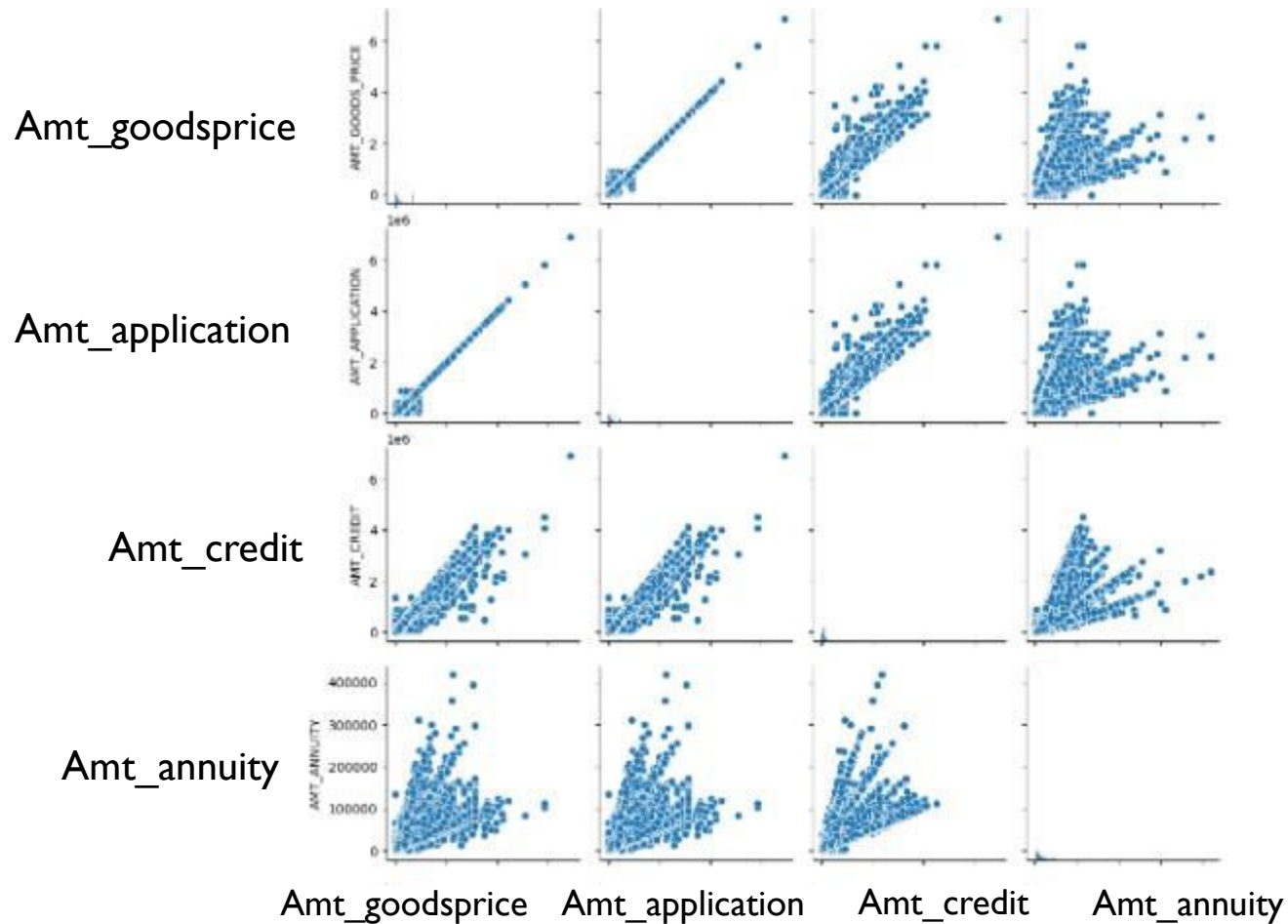
-- POSITIVE CORRELATION:

- More credit amount of the loan the more money is required to pay back periodically and obviously increased difficulty in repaying.
- The more the price of the goods, the more is the credit amount of the loan given, and more to repay periodically.

• AND HENCE LEADING TO MORE LOAN DEFAULTERS.

- People without payment difficulties take more credit for the annuity

Comparison of how Amount credit , Application, Goods price contribute in payment difficulty group

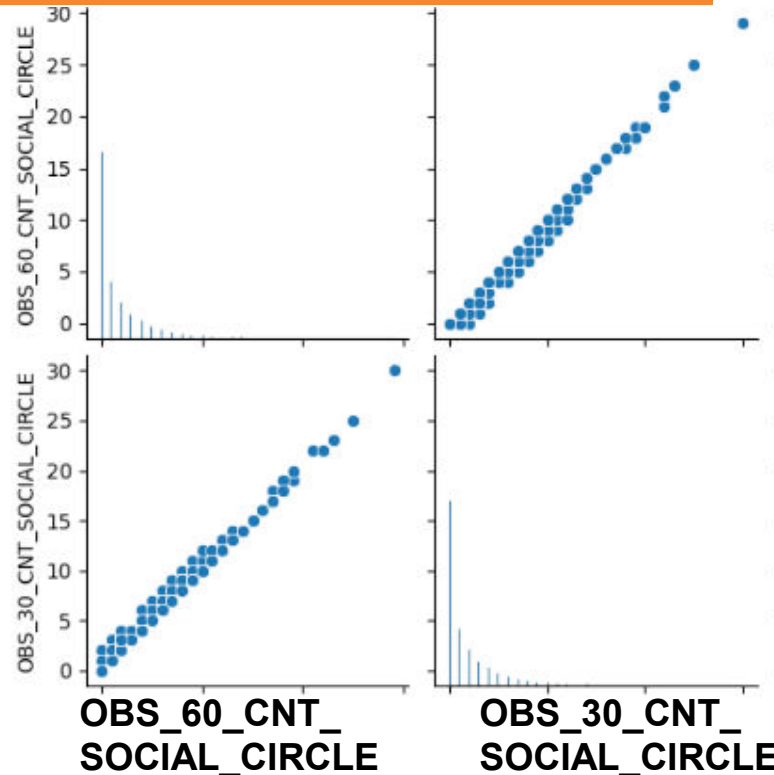


- POSITIVE CORRELATION: steep slope relation between goods price and application.
- Higher the value of good purchased ,there will be need of loan and hence the correlation.

OBS 60 CNT SOCIAL CIRCLE Vs OBS 30 CNT SOCIAL CIRCLE OF THE CLIENTS VIN TARGET 1 GROUP:

OBS_60_CNT_
SOCIAL_CIRCLE

OBS_30_CNT_
SOCIAL_CIRCLE

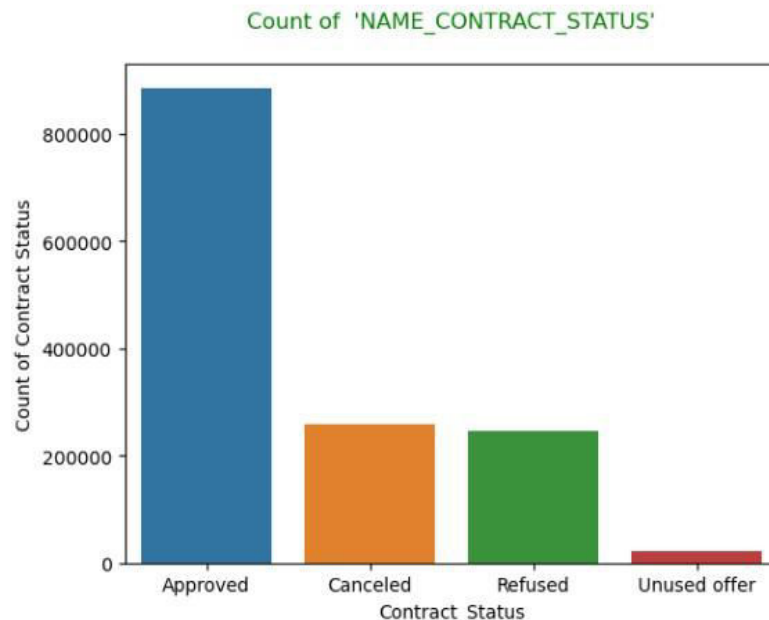


Refers to How many observation of client's social surroundings with observable 30/60 DPD (days past due) default

-- HIGH POSITIVE CORRELATION:

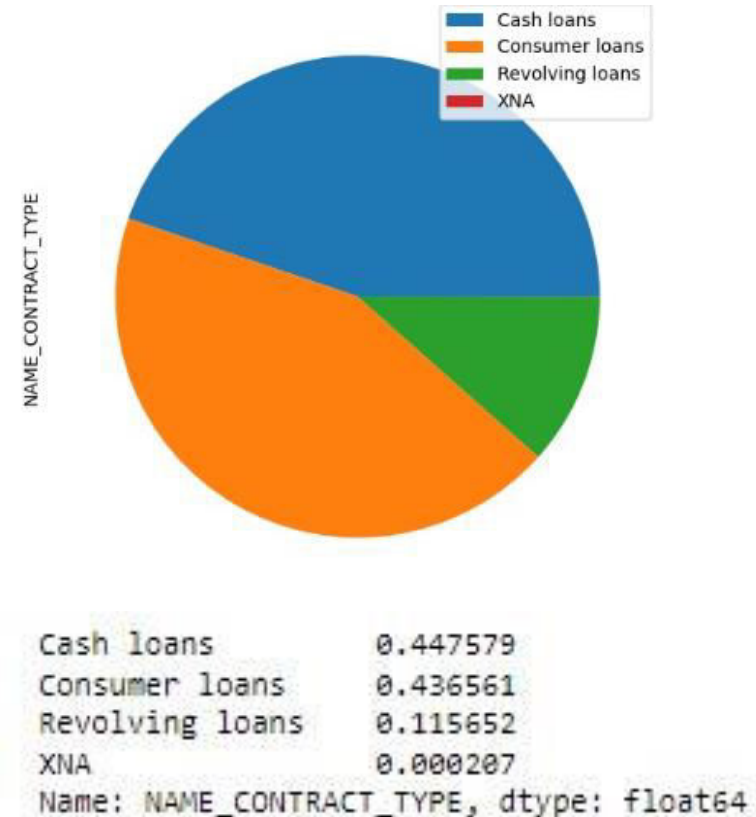
--It is a good indicator /variable to approve or reject loans.

Contract status and Contract Type



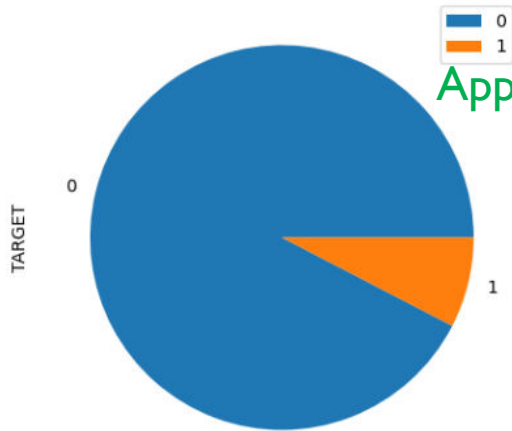
```
Approved      62.679378
Canceled      18.351900
Refused       17.357984
Unused offer   1.610737
Name: NAME_CONTRACT_STATUS, dtype: float64
```

- Data Imbalance noted.
- Approved loans are 62% and cancelled are 18%.
- Refused and unused are 17% and 1.6%

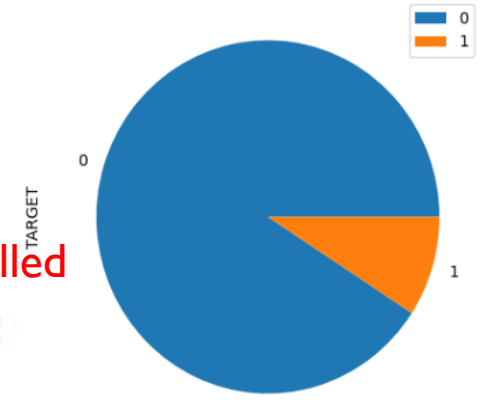


- There is a new type of loan in the previous application dataset called consumer loans which constitutes around 43% and it is second highest.
- The highest being cash loans (44%)

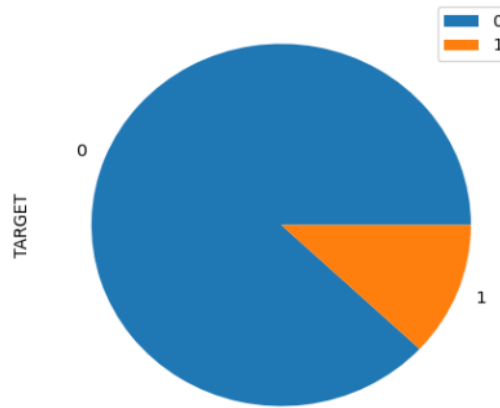
CONTRACT STATUS OF THE LOAN - FOR TARGET 0 & 1



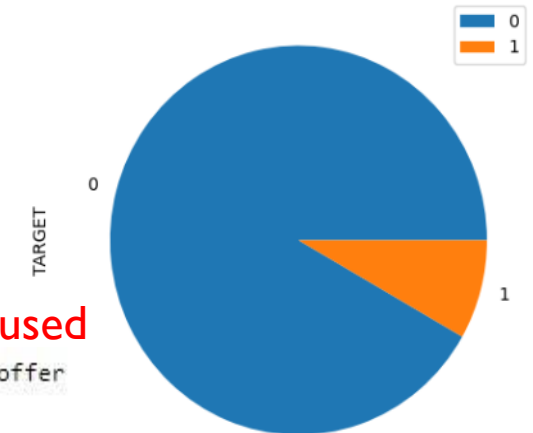
Target 0 and 1 for : Approved
0 0.924113
1 0.075887
Name: TARGET, dtype: float64



Target 0 and 1 for : Canceled
0 0.908264
1 0.091736
Name: TARGET, dtype: float64



Target 0 and 1 for : Refused
0 0.880036
1 0.119964
Name: TARGET, dtype: float64



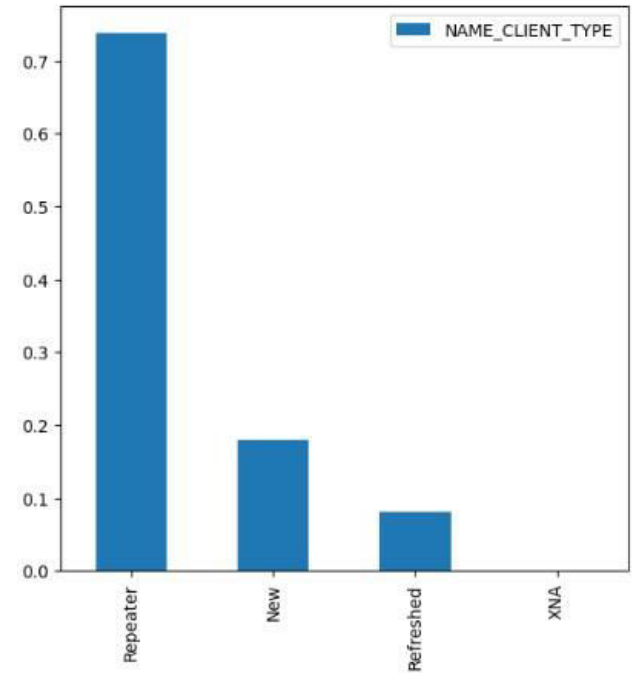
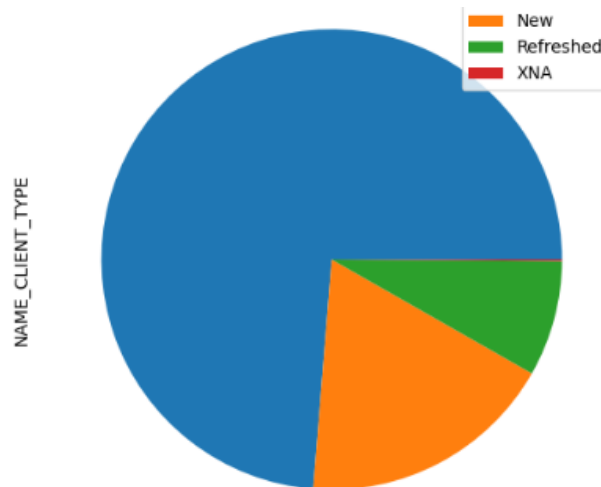
Target 0 and 1 for : Unused offer
0 0.917483
1 0.082517
Name: TARGET, dtype: float64

Payment difficulties: (target 1)

- Defaulters: For Approved cases (7.5%), Cancelled cases (9%), refused (11%) & unused (8%) rate are also higher than the approved rate in target 1.

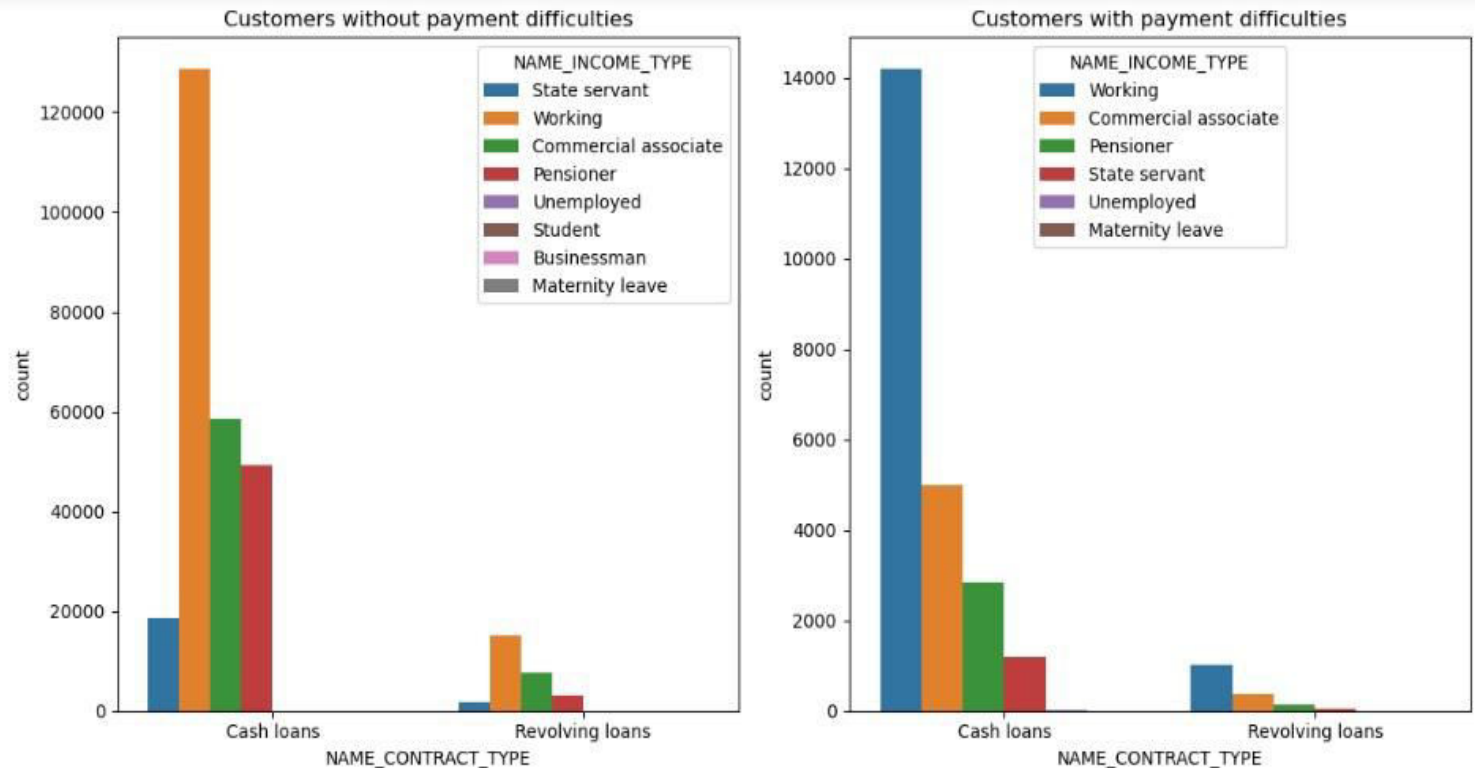
Contract Status Vs Client type for Target 0 1

```
Repeater    0.737188  
New         0.180434  
Refreshed   0.081217  
XNA         0.001162  
Name: NAME_CLIENT_TYPE, dtype: float64
```



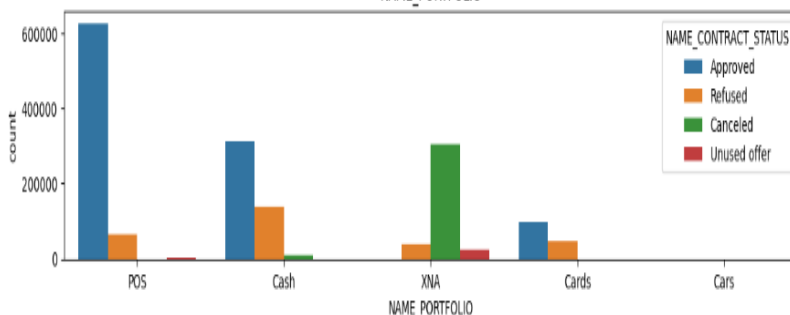
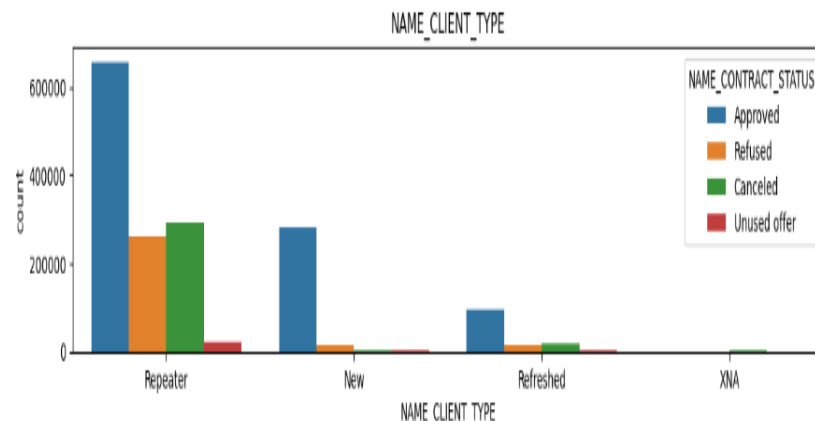
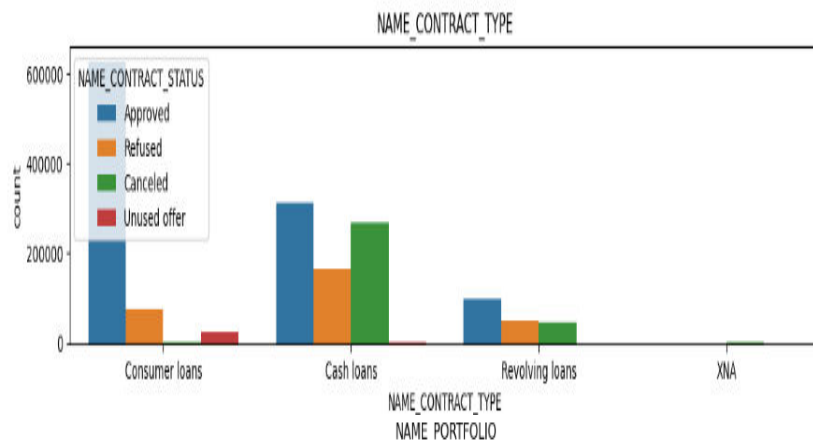
- Most clients are repeaters with 73%
- And new clients are 18%

Comparison of Contract status , Income type on target 0 & 1



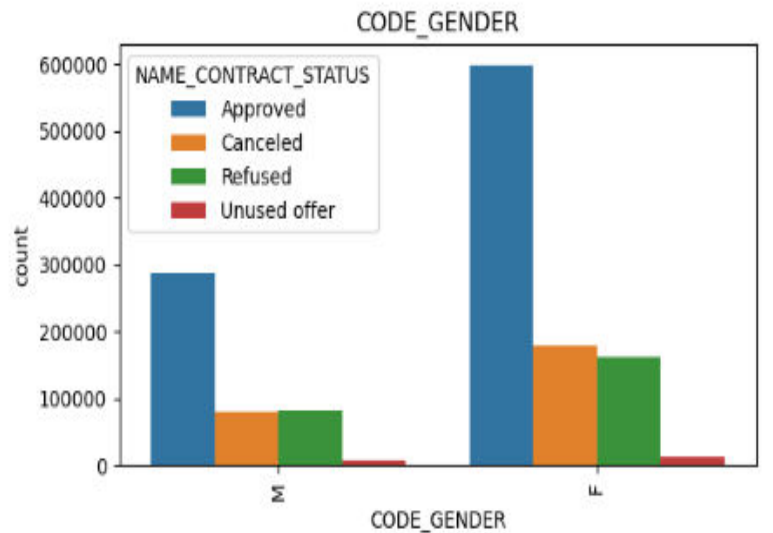
- Cash loans is majorly taken by Working group in both targets than revolving loans.
- And Working group face repayment difficulty in cash loans.
- Pensioners have moderately applied for cash loans and have good repayment rate.(good target)
- Revolving loans are also taken mostly by working group and have better repayment, though the ratio is very low in comparison with cash loans.

Comparison of contract type , client type with contract status

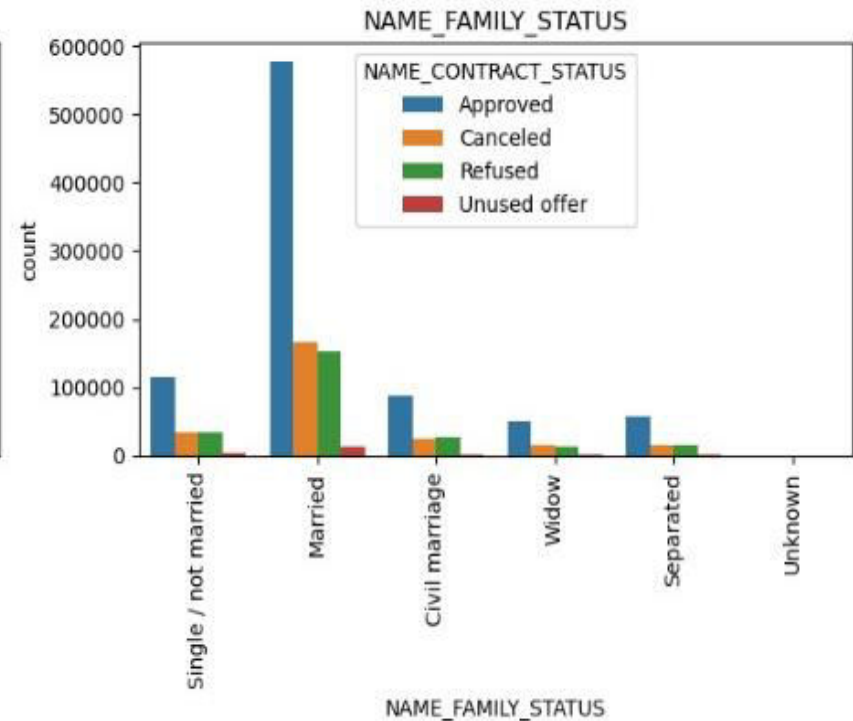


- **Approved category:**
 - High for consumer loan types , repeaters and POS kind of transaction loans . Least for revolving loans, cards portfolio.
- **Refusal category:**
 - Cash loans and consumer loans refused most and for repeaters.
 - Refusal rate is low in new clients.
- **Cancelled category:**
 - High for repeaters and for cash loans.
- **Unused offer:**
 - Seen in consumer loan types who are mostly repeaters.

Contract status - Gender, Family status of the Client

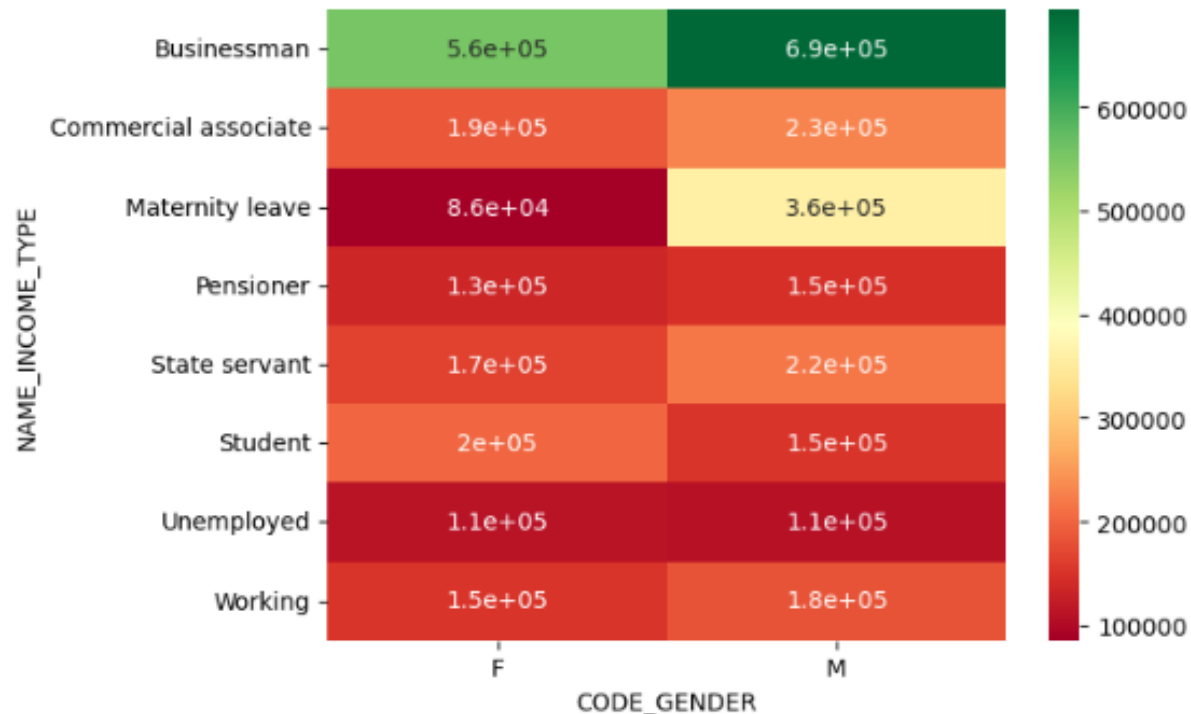


- More loans applied and approved for females.
- Cancellation, rejection, refused rate also higher among females.



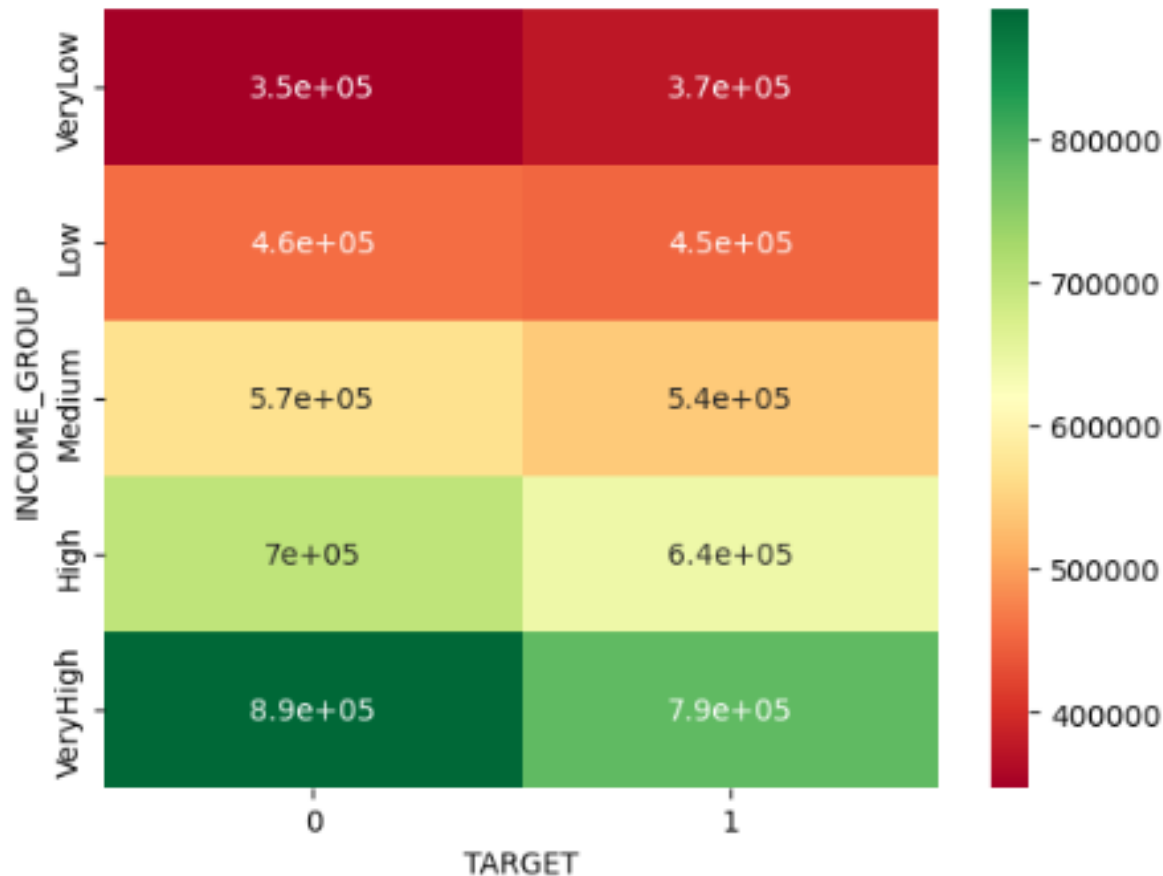
- More loans applied and approved for married.
- Refusal and unused rate also comparatively higher in married people.

Comparison of Income type, Gender, Total Income With target group



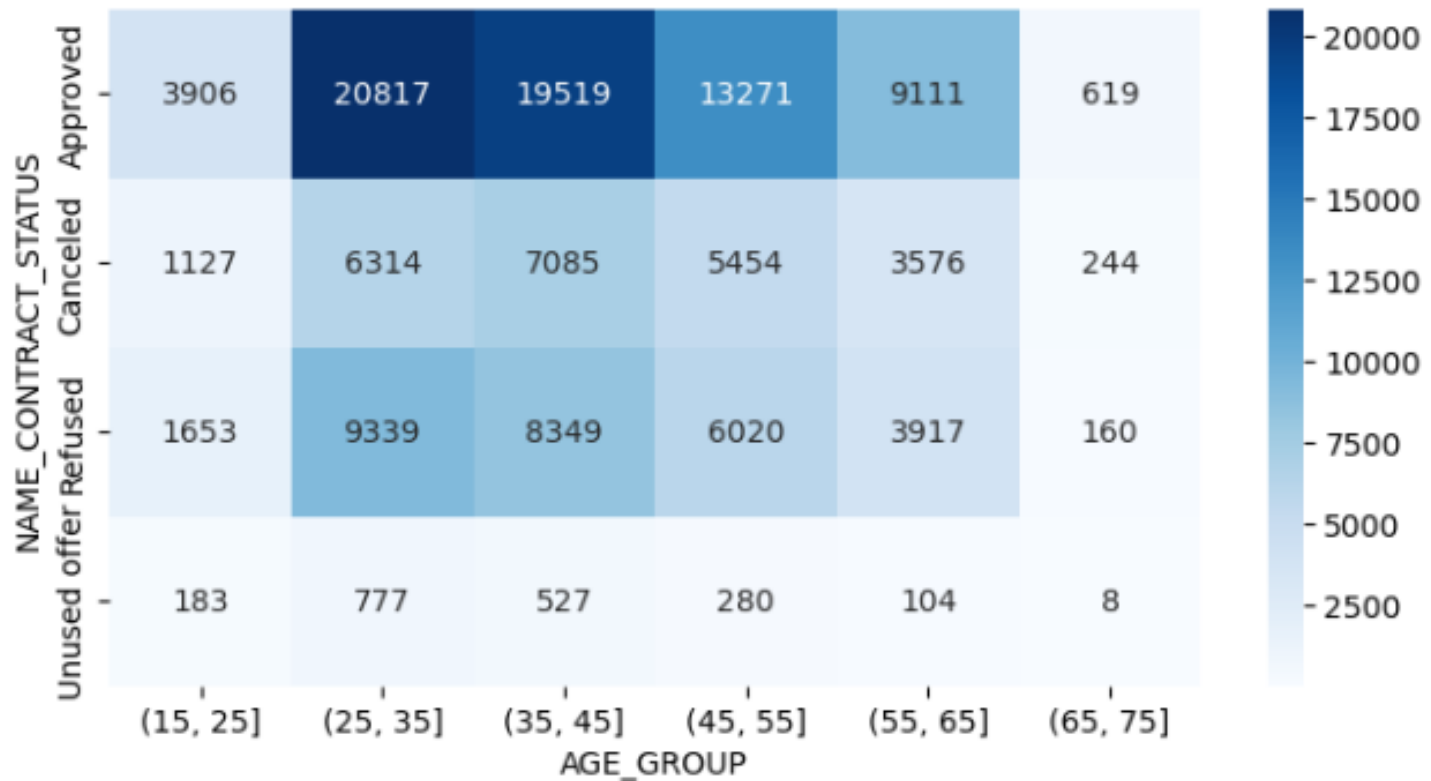
- More income seen in businessman, commercial associate, and hence more repayment possible if proper protocol and rules taken by bank.
- Least income for maternity leave females and hence repayment from them will also be difficult.
- Least repayment from unemployed group also observed.

Comparison of Income group ,Target group with Credit loan amount



- High income group have higher Credit amount of the loan. since their loan sanctioned would be at higher amount.
- Medium income group have more in number
- in both in Target groups.

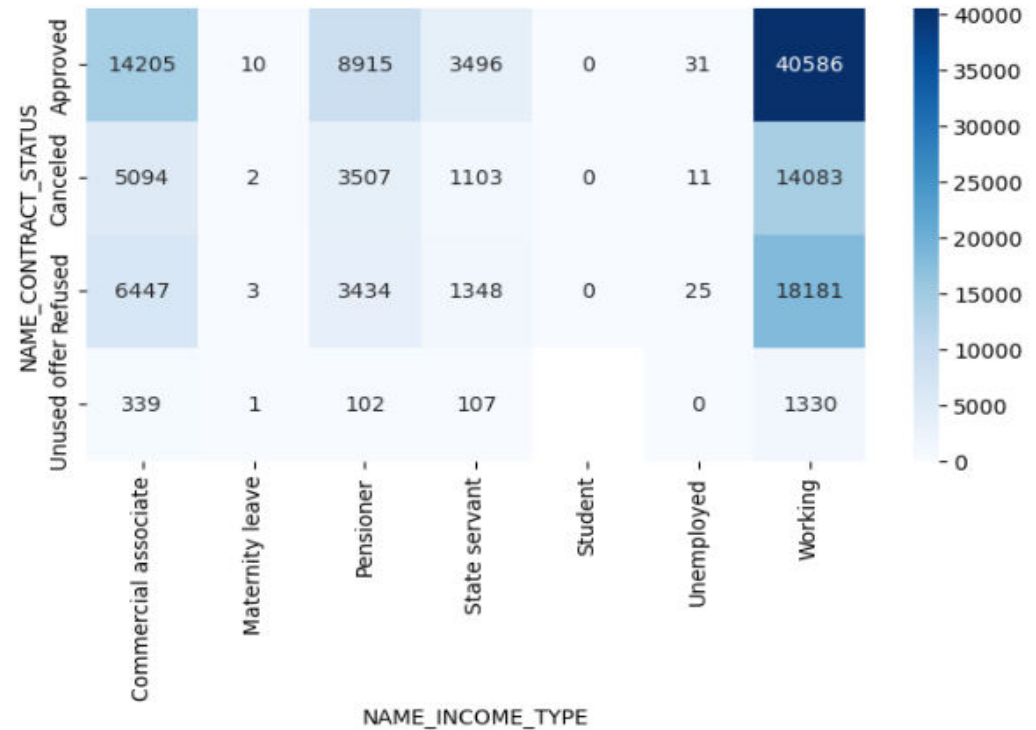
Comparison of Contract Status of the Loan , Age group on Target 1



- More Approved and defaults in young age group [25-35] years.
- Elder people [>55years] repay better.
- A good amount of loan refused & unused for younger population.

Contract status, Income type on target 0 & 1

maternity leave
females,
Students are poor
target for loan
approval.



High loan taken by:

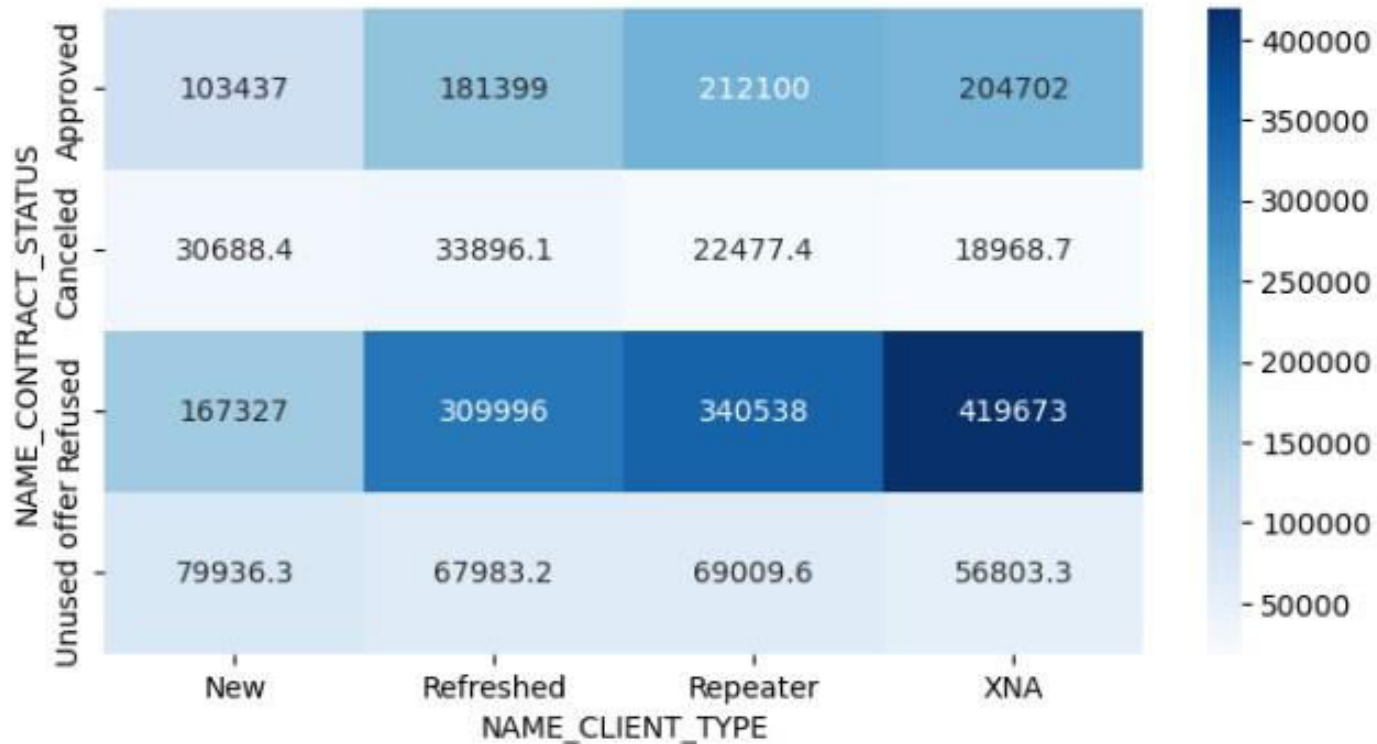
- First: working people and defaulters also more in working group.
- Second : commercial associate and pensioners.

Low loan taken by:

- maternity leave and nil for students.

-- low and nil preference to loans by maternity leave people and students who cannot afford to repay.

Contract status , Client type on target 0 & 1



- Most loans are approved for repeaters; least for new clients.
- Cancelled more for refreshed category.
- Refusal rate is highest for repeaters.
- Unused rate is relatively higher for all categories.

Top correlations between Target 0 and 1

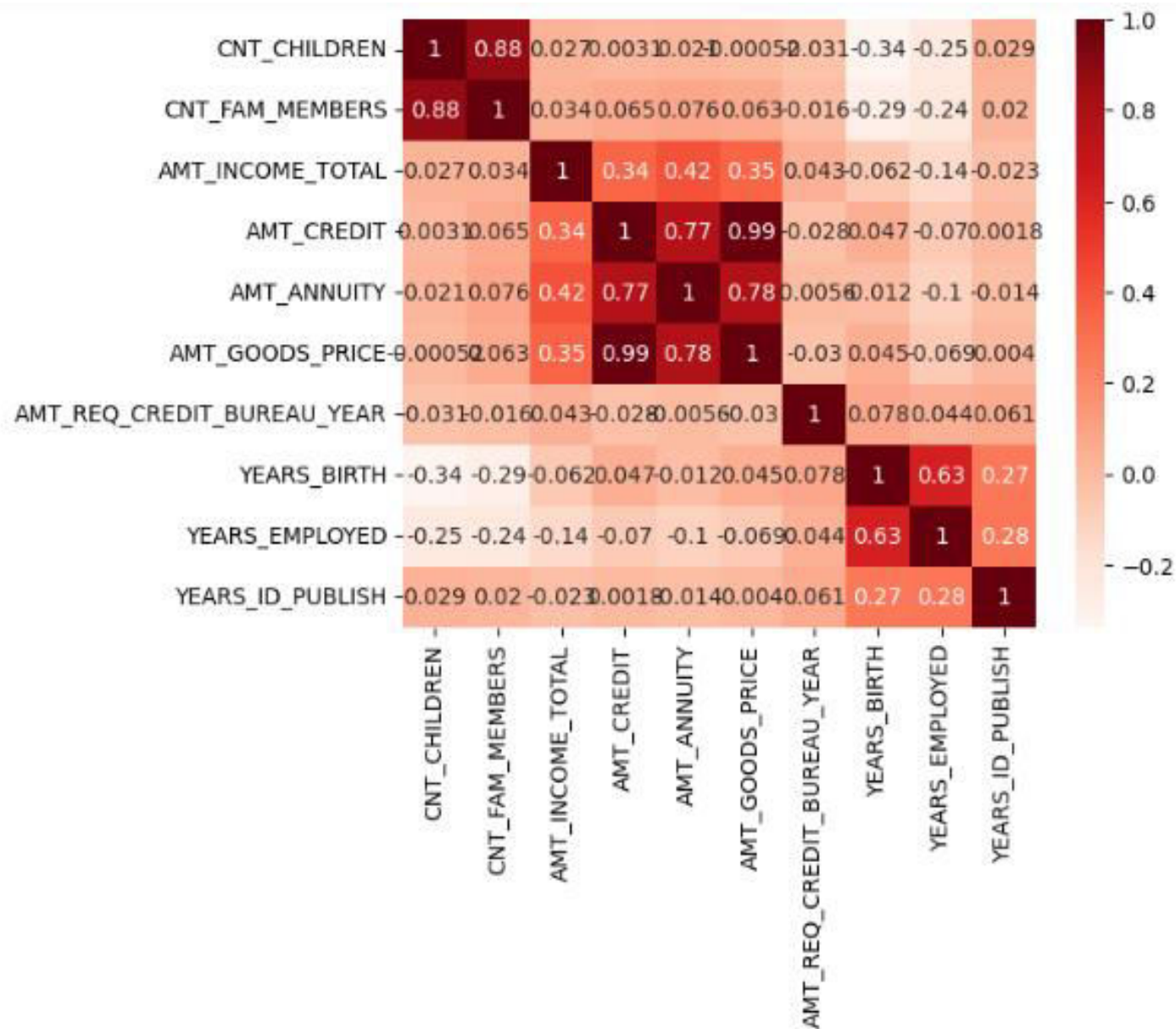
Target 0

	var_1	var_2	correlation	corr_abs
508	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998508	0.998508
178	AMT_GOODS_PRICE	AMT_CREDIT	0.987250	0.987250
234	CNT_FAM_MEMBERS	CNT_CHILDREN	0.878571	0.878571
538	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.859332	0.859332
359	LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.830381	0.830381
179	AMT_GOODS_PRICE	AMT_ANNUITY	0.776686	0.776686
149	AMT_ANNUITY	AMT_CREDIT	0.771309	0.771309
779	YEARS_EMPLOYED	YEARS_BIRTH	0.626028	0.626028
329	REG_CITY_NOT_WORK_CITY	REG_CITY_NOT_LIVE_CITY	0.435514	0.435514
148	AMT_ANNUITY	AMT_INCOME_TOTAL	0.418953	0.418953

Target 1

	var_1	var_2	correlation	corr_abs
508	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998269	0.998269
178	AMT_GOODS_PRICE	AMT_CREDIT	0.983103	0.983103
234	CNT_FAM_MEMBERS	CNT_CHILDREN	0.885484	0.885484
538	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.868994	0.868994
359	LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.778540	0.778540
179	AMT_GOODS_PRICE	AMT_ANNUITY	0.752699	0.752699
149	AMT_ANNUITY	AMT_CREDIT	0.752195	0.752195
779	YEARS_EMPLOYED	YEARS_BIRTH	0.582443	0.582443
329	REG_CITY_NOT_WORK_CITY	REG_CITY_NOT_LIVE_CITY	0.472052	0.472052
509	OBS_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.337181	0.337181

The correlation matrix heat map for target 1



Positive correlation seen between

- Positive correlation between loan amount, amount annuity, goods price, loan application:
 - As the loan amount annuity increases, the more money is required to pay back periodically, the repayment is also difficult.
 - As the loan credit amount increases, the repayment is also difficult.
 - As goods price increases, credit loan amount given also increases and hence the repayment is also difficult.
 - So goods price is directly related to repayment.
- More the children, more the family members.
- The number of observations or instances in the dataset where clients have a social surroundings variable indicating a 30/60 days past due default that is observable is also highly positively correlated with payment difficulty group.
- The number of observations or instances in the dataset where clients have a social surroundings variable indicating a default on 30 /60 days past due is also highly positively correlated with payment difficulty group.
- As age increases the years employed also increase.

Insights for factors that contribute for difficulties in repayment:

More seen in :

- Females (57%) , Married people (59.81%), younger population 31-49 years.
- Cash loans (93.53%), 'Repair'work category , Medium income people .
- Working group (61%) , laborers (23%) , Education wise: Secondary / secondary special (78.6%) , unemployed ,contrastingly among businessmen group

Difficulty in repayment with previous application seen:

High Refusal of loans: Cash loans ,consumer loans ,for repeaters, for younger population.

High Cancellation of loans: For Cash loans, repeaters, 35-45 years.

High unused loan offers: For Consumer loan types who are mostly repeaters. for younger population.

Recommendations: Changes that can be done/ suggestions from bank side:

Bank can take:

- **Careful measures** to overcome , **the unused and rejected loans** for repeaters ,workers ,especially in 35-45 years (since they are major in count)to get high repayment yield.

the repayment issues by **approving loans** by different strategy especially to capable females, widow/seperated / single, staffs, driver; and focus on successful repayments.

- **Increase the loan** offers to new people.

- **Check why refusal** -- bank should create a different strategy to analyze to give loan. It shouldn't be a business loss for the bank. May be the rejected clients have the capacity to repay well.

- **Unused:** Appropriate awareness to the clients in accepting and benefitting out of the loans especially 25-35 years

- **Create loan limits ,Release the loan amount in installments** to avoid loss especially for working group and young population.

- **Motivate elder population** like pensioners ,etc to avail loan as their repayment rate is very high.

Thank you