

CREDIT EDA CASE STUDY

Laxmi Sahu

Sushant Yadav

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. 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.

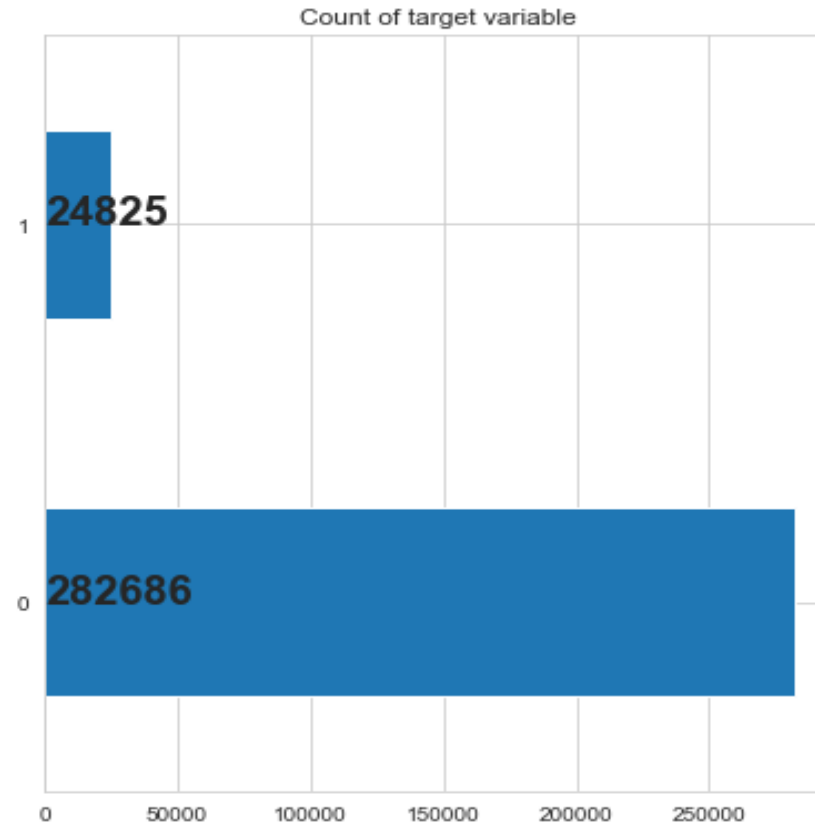
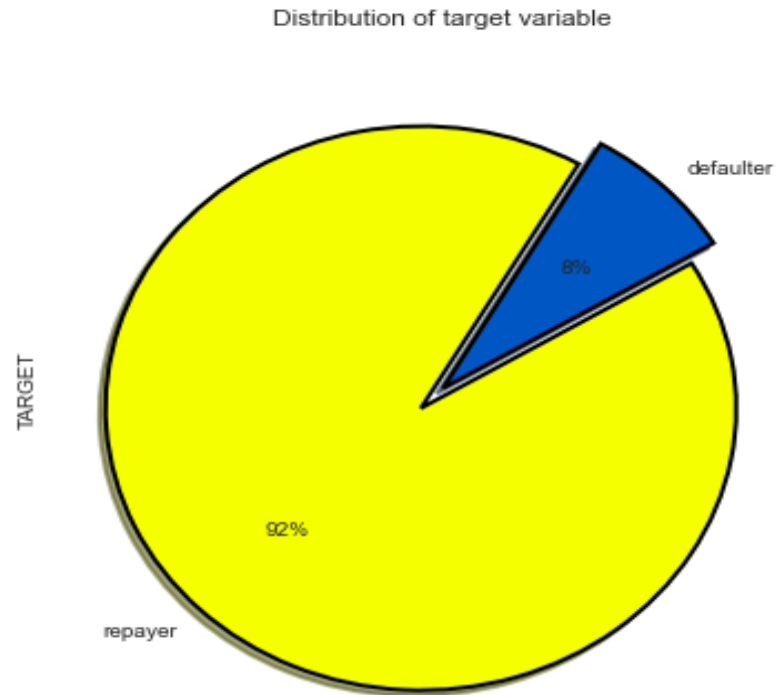
SOLUTION APPROACH

We have used the below approach for deriving the insights :

- The required libraries needed for data cleansing and visualisation are imported.
- We have done the data cleansing for columns wherever necessary and dropped the columns with majority of data as NA. Outliers are identified and handled wherever possible. Data imbalance is checked.
- Created new columns as per the requirements
- Univariate/Bivariate Analysis of the relevant Categorical/numerical is done and insights are derived
- Current and Previous application data is done to derive insights based on bank Approval loan status

DATA IMBALANCE

DATA IMBALANCE

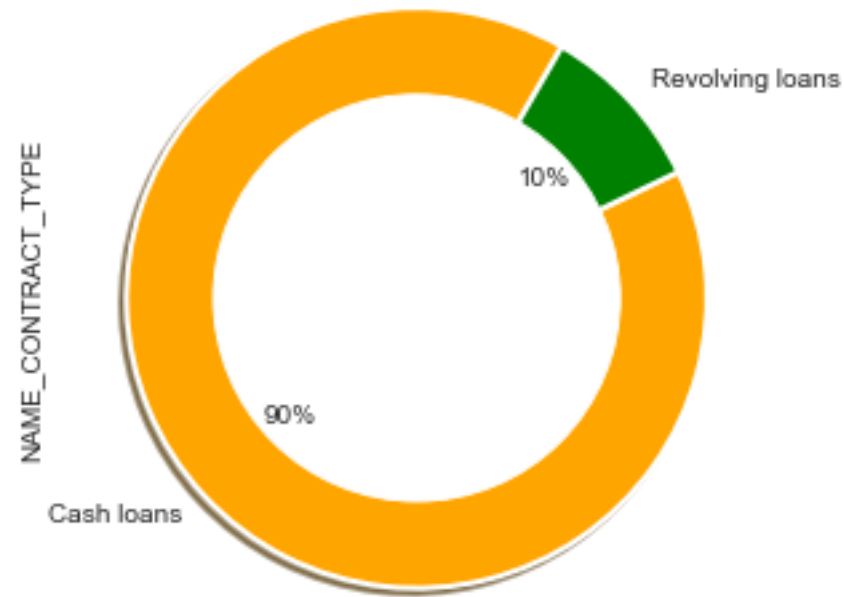


Inference:- There are 8% out of total client population has problem in repaying loan.

UNIVARIATE ANALYSIS

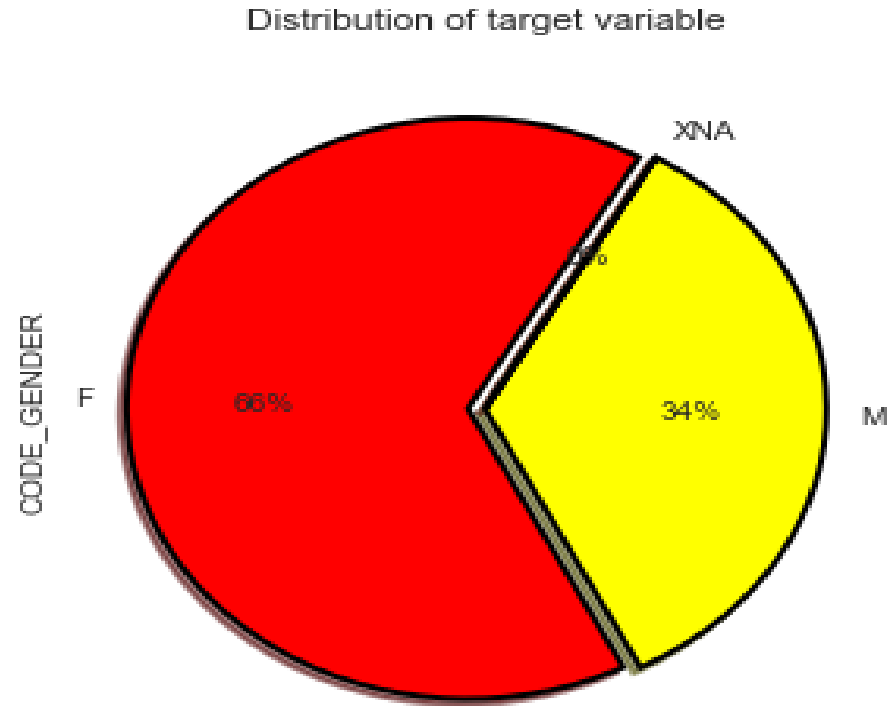
DISTRIBUTION OF CONTRACT TYPE - CURRENT APPLICATION

distribution of contract types in current_application



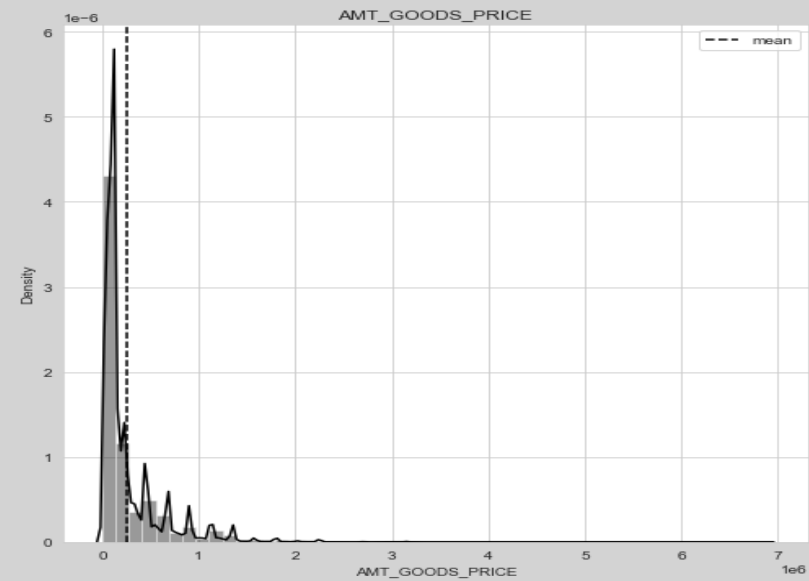
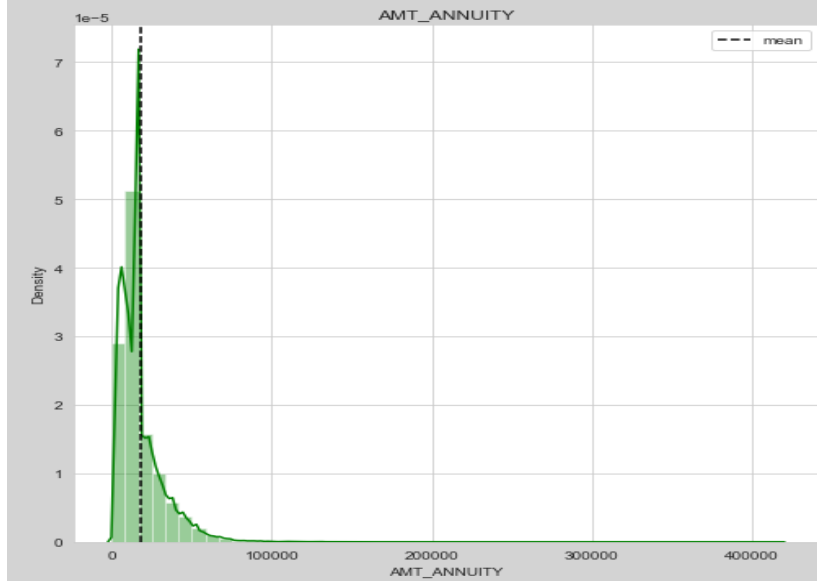
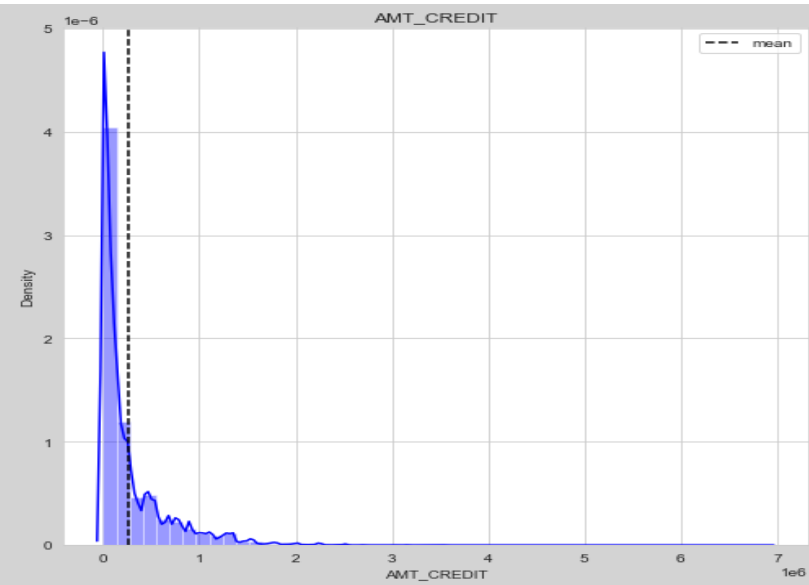
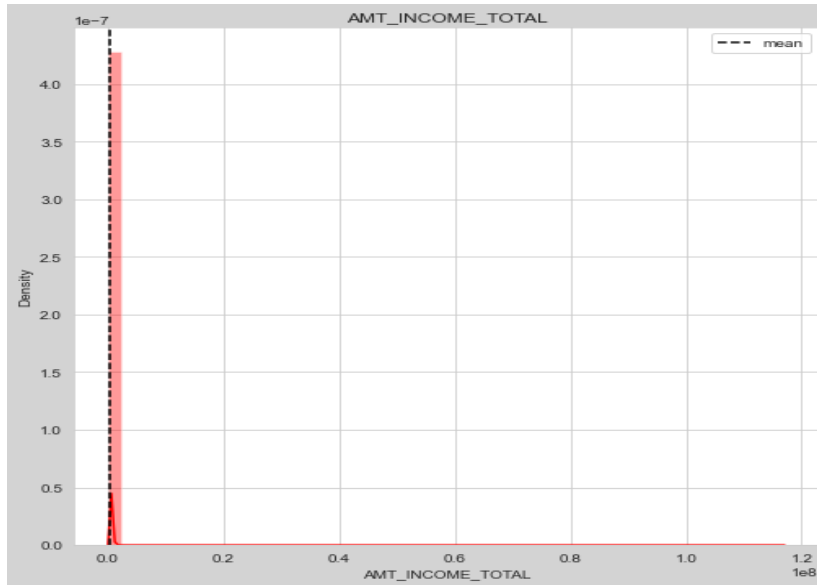
INFERENCE:- 2 types of contracts are there, out of which 90% are cash loans.

GENDER DISTRIBUTION IN CURRENT APPLICATION

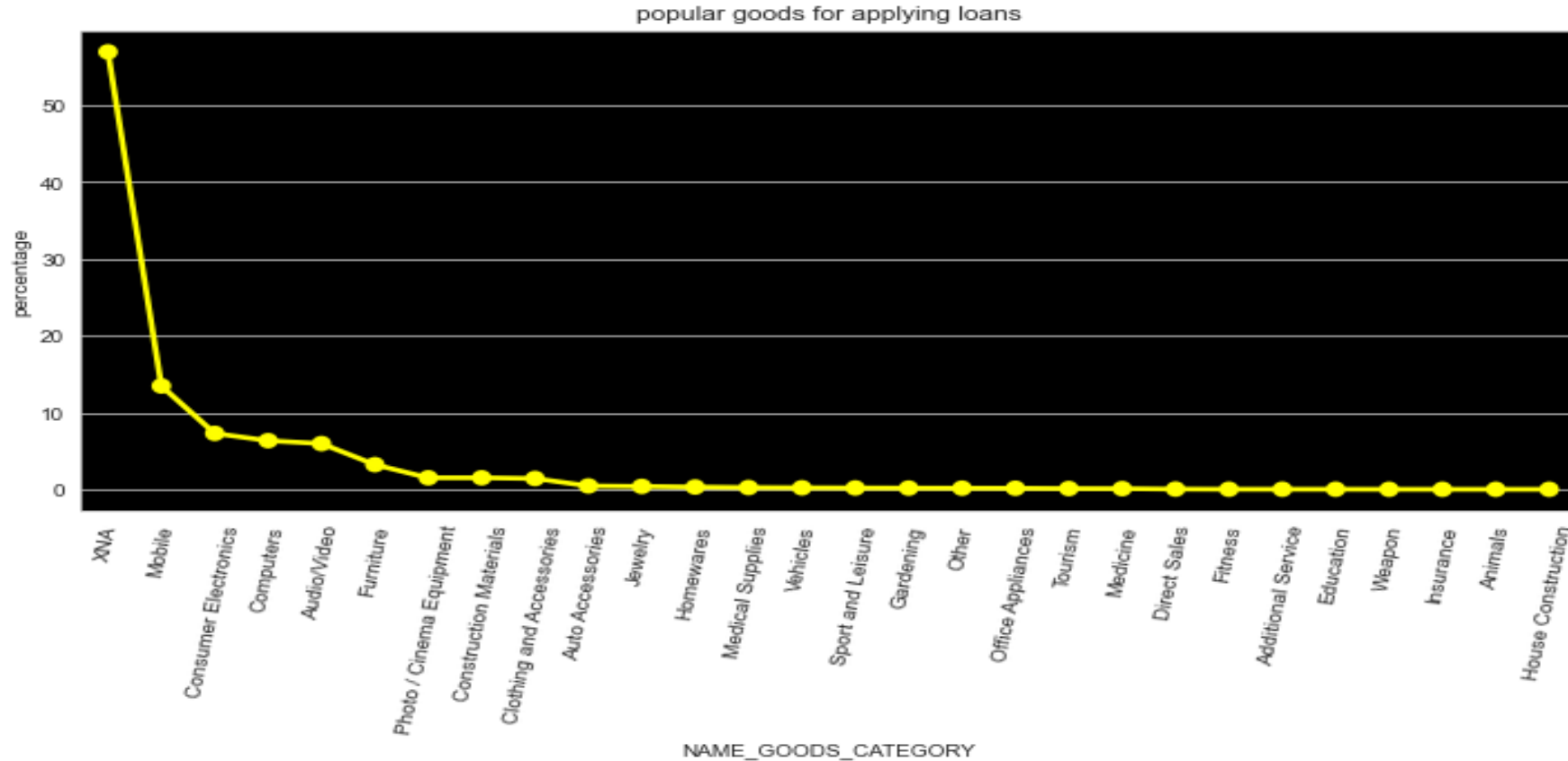


INFERENCES:- Female clients form 66% part of current application data.

DISTRIBUTION OF INCOME



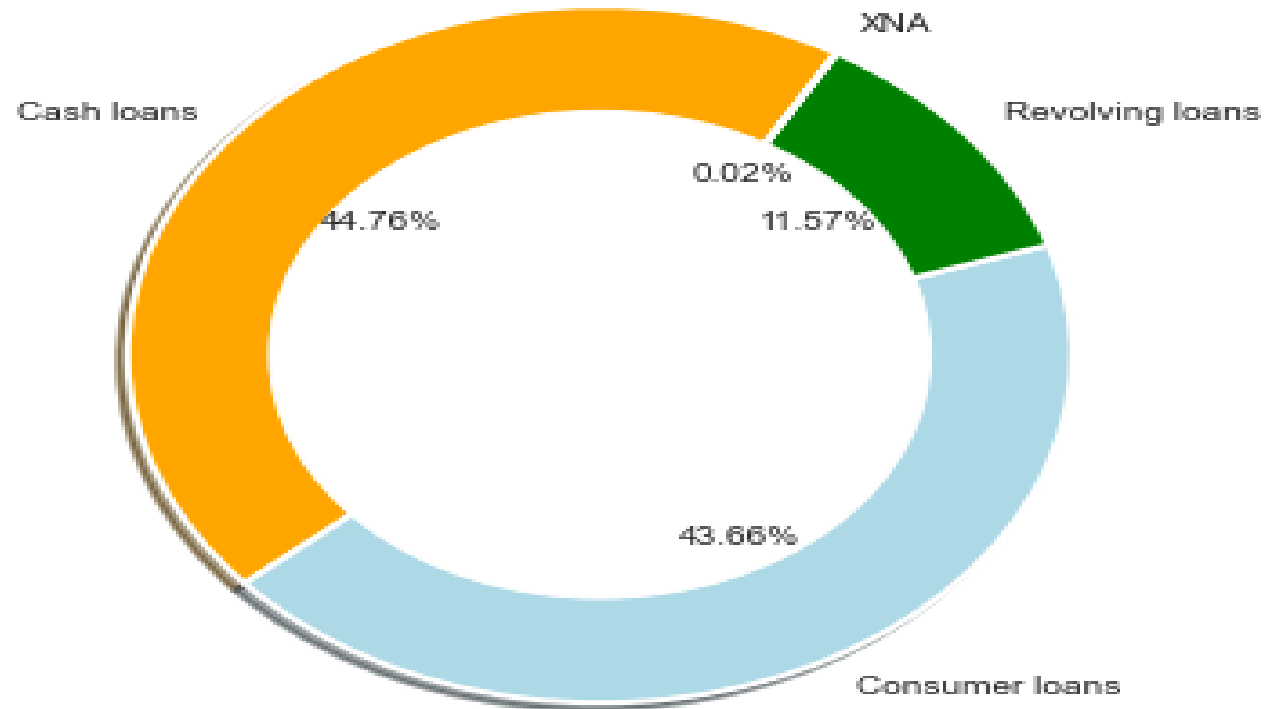
POPULAR GOODS FOR APPLYING LOANS



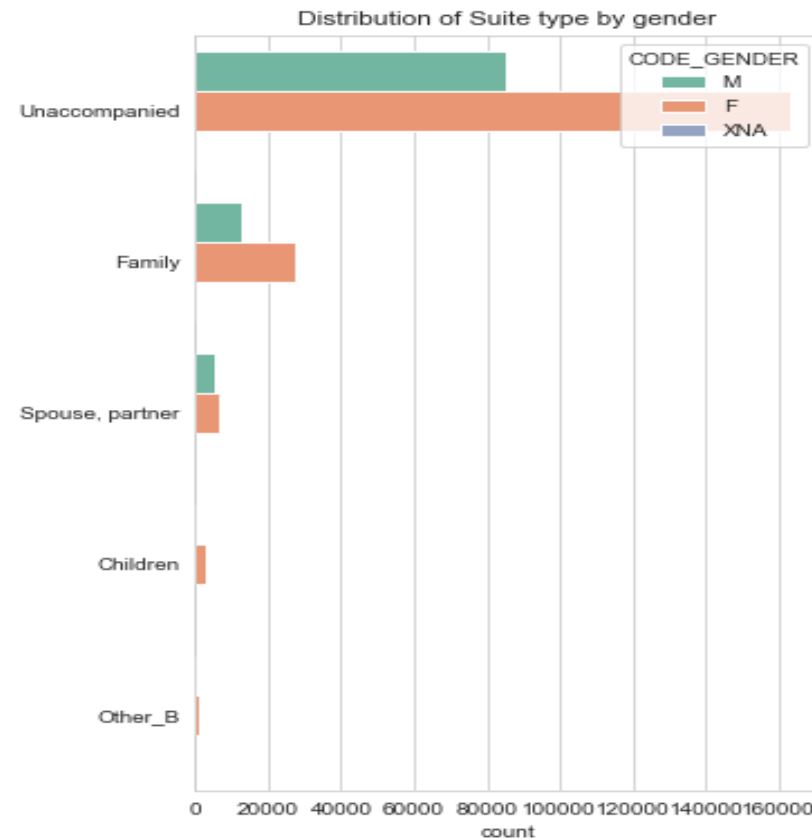
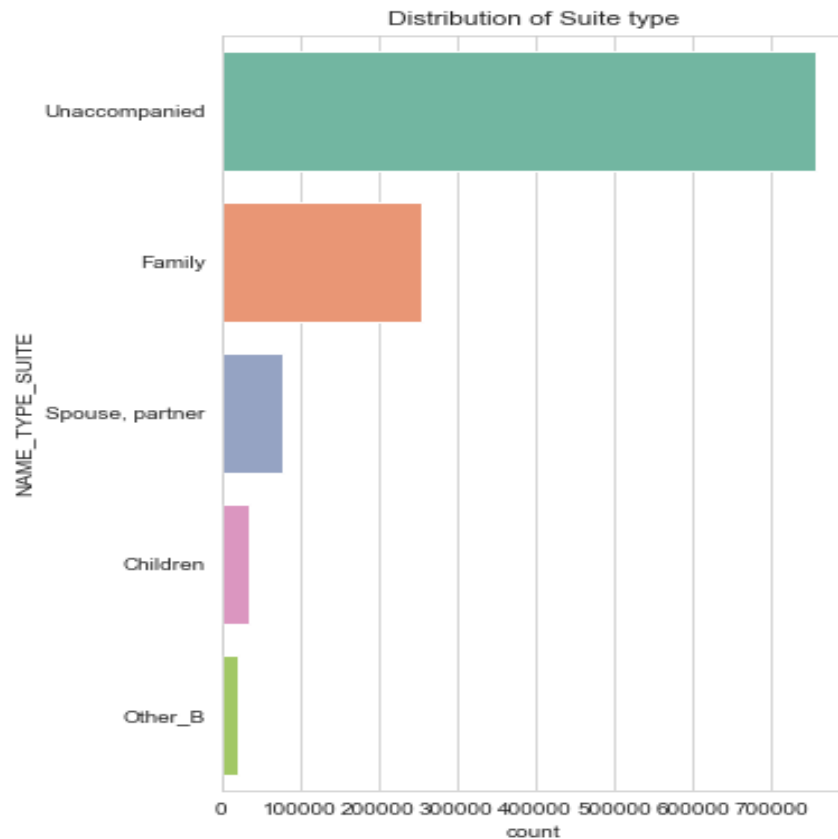
Inference:- Mobiles ,Computers and consumer electronics are popular goods for applying loans

DISTRIBUTION IN CONTRACT TYPE - PREVIOUS APPLICATION

distribution of contract types in previous_application



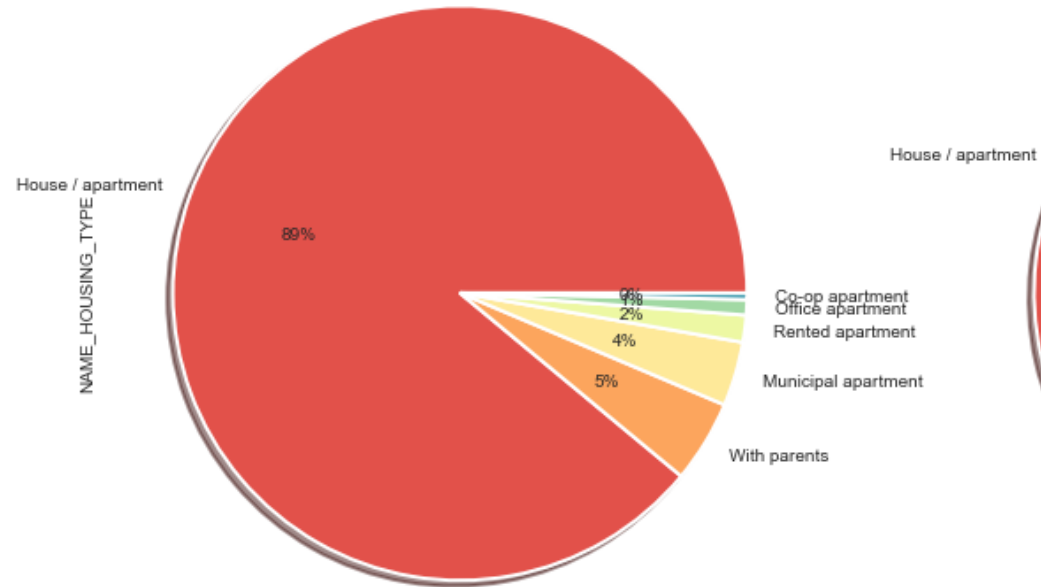
DISTRIBUTION OF SUITE TYPE



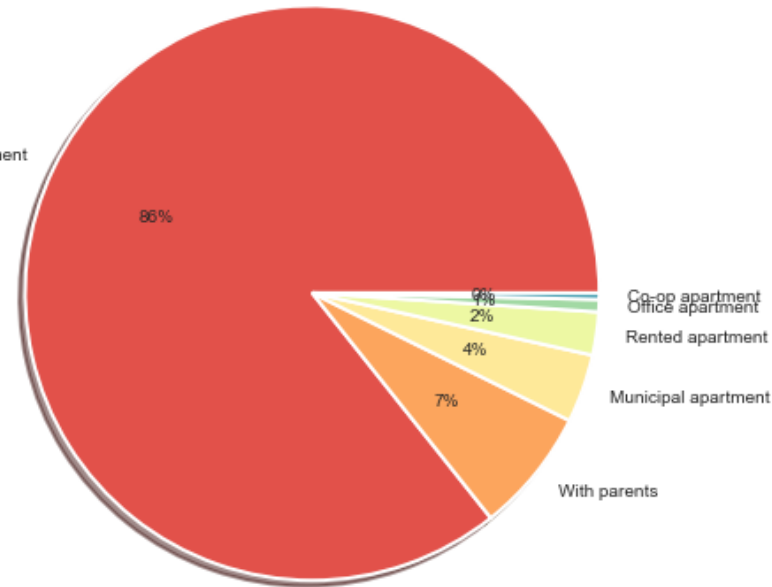
INFERENCE:-Mostly clients unaccompanied when they were applying for the loan. Out of the total unaccompanied clients, female clients were more in number.

DISTRIBUTION OF HOUSING TYPE BY LOAN REPAYMENT STATUS

Distribution of housing type for Repayer



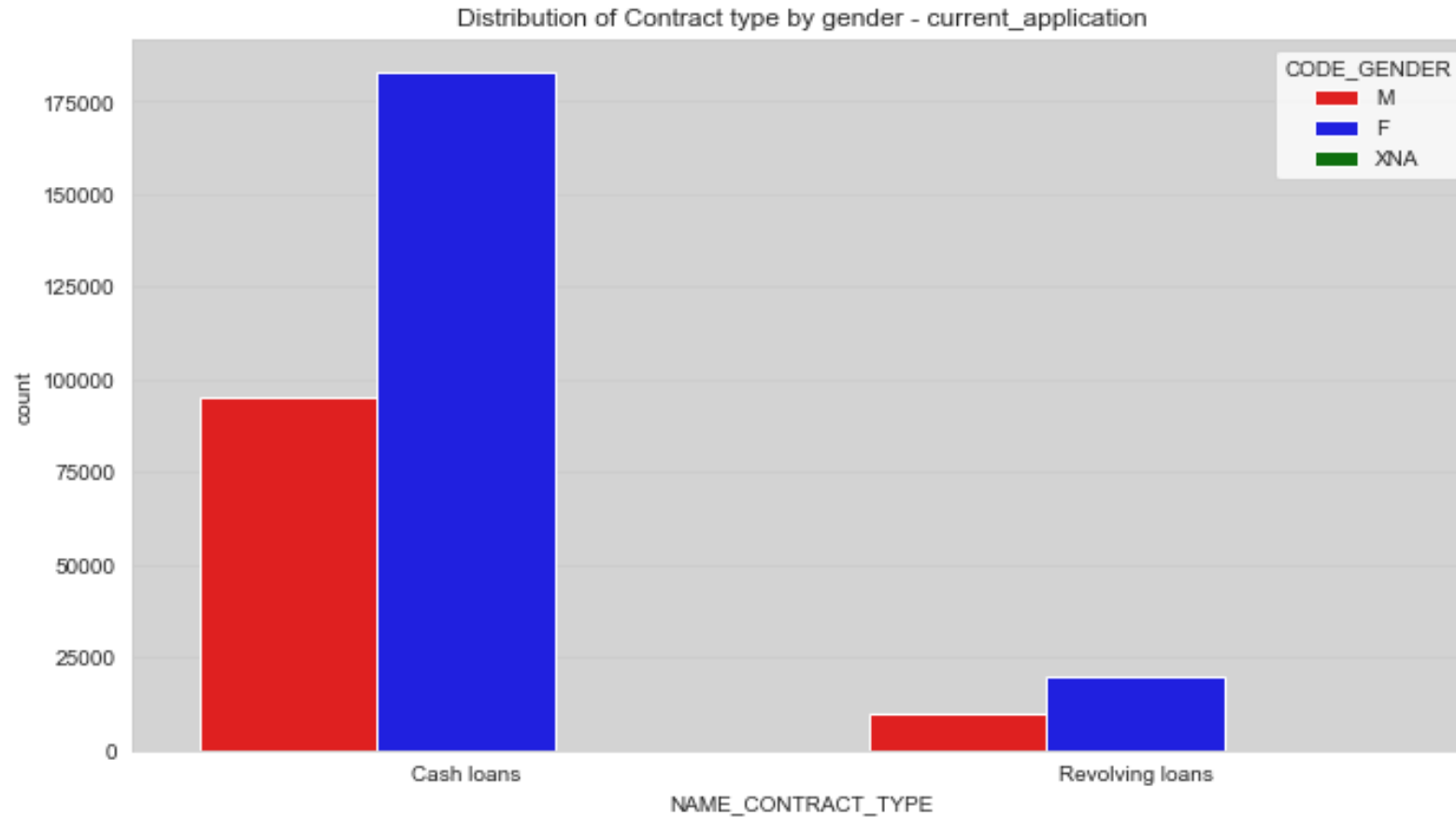
Distribution of housing type for Defaulters



INFERENCE:- Defaulters are more from the clients who are living with parents than repayers.

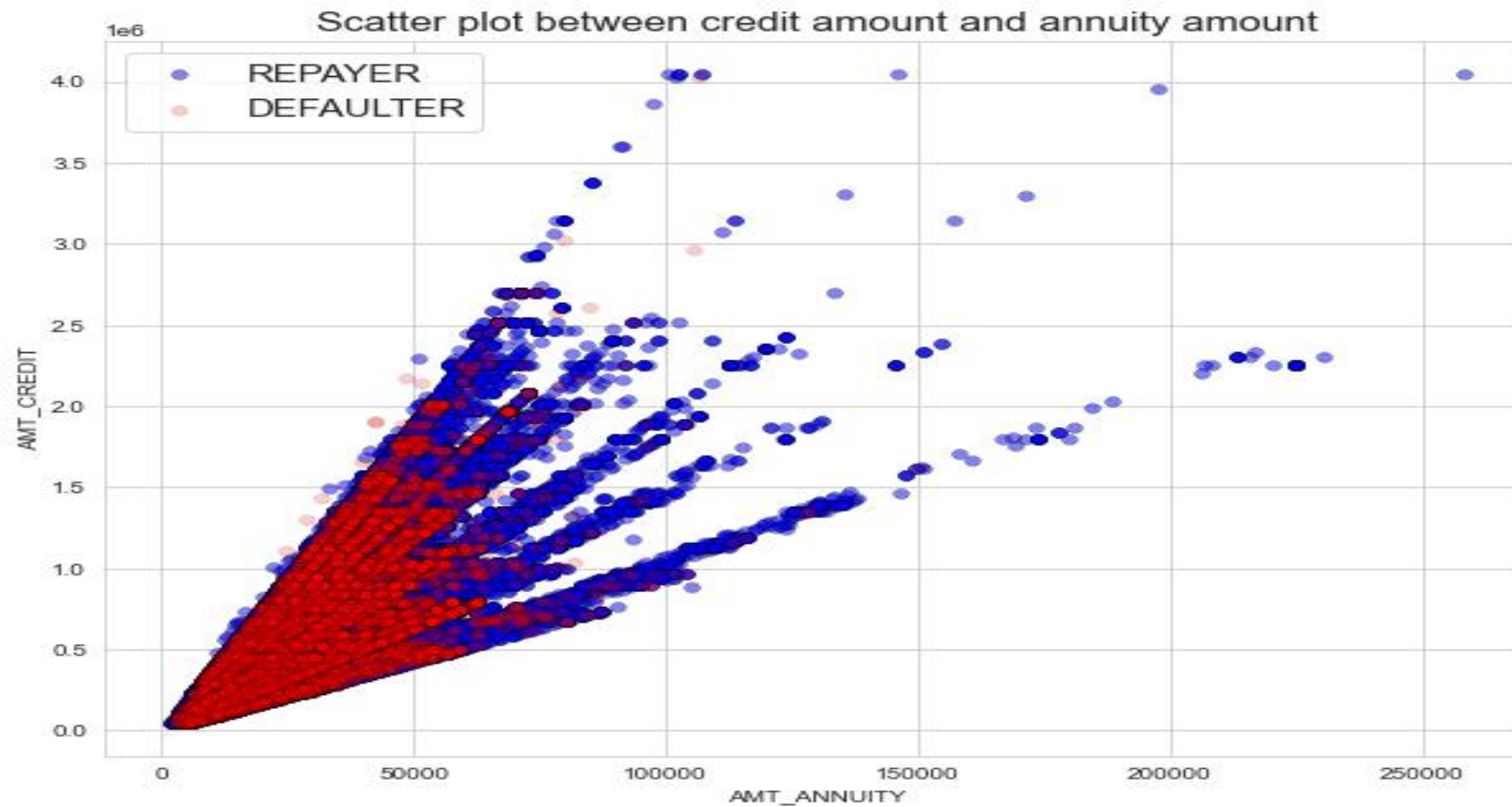
BIVARIATE ANALYSIS

DISTRIBUTION OF CONTRACT TYPE BY GENDER

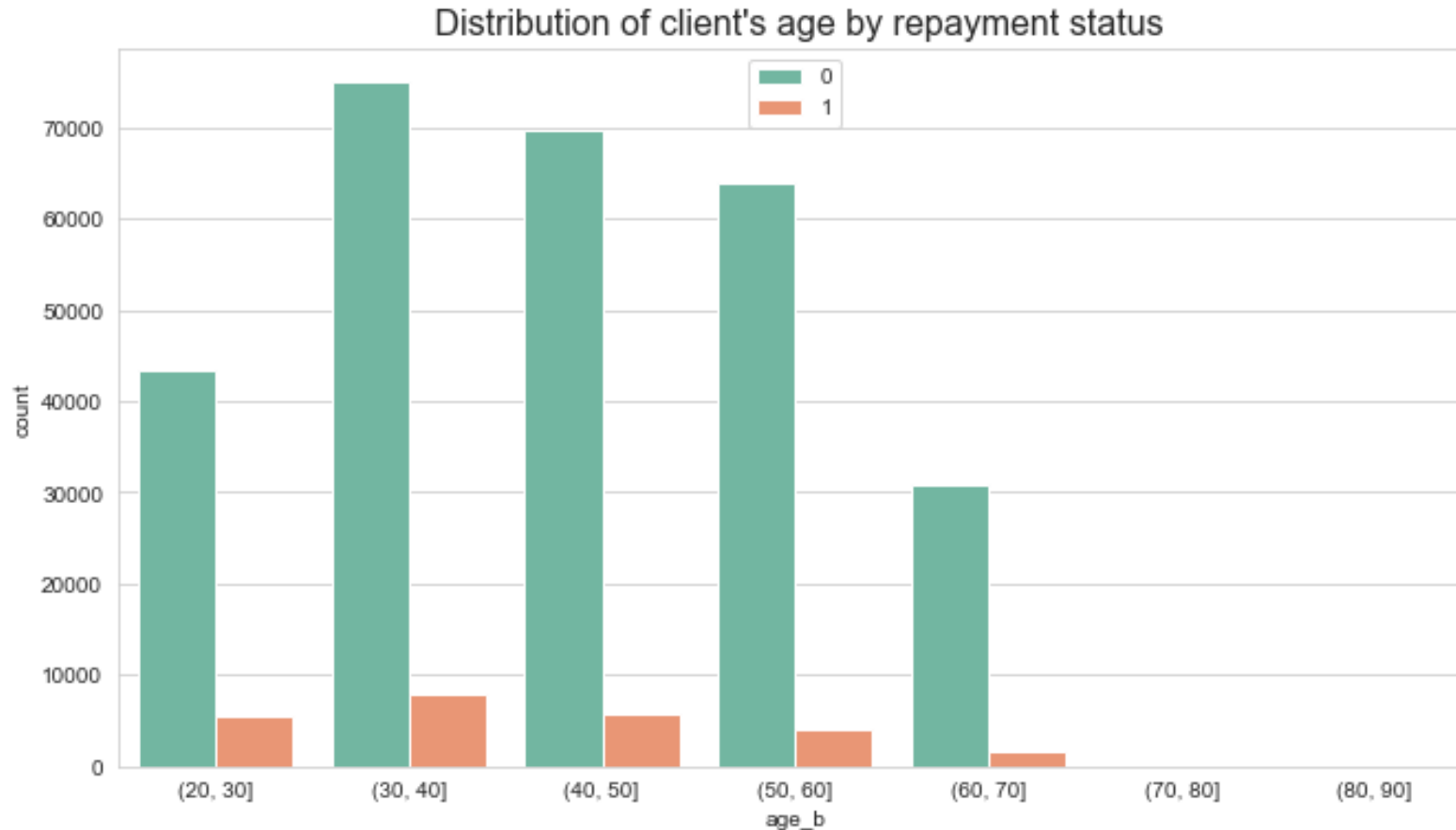


Inference:- Both genders prefer cash over revolving loans. Female clients are more than male clients.

SCATTER PLOT BETWEEN CREDIT AMOUNT AND ANNUITY AMOUNT



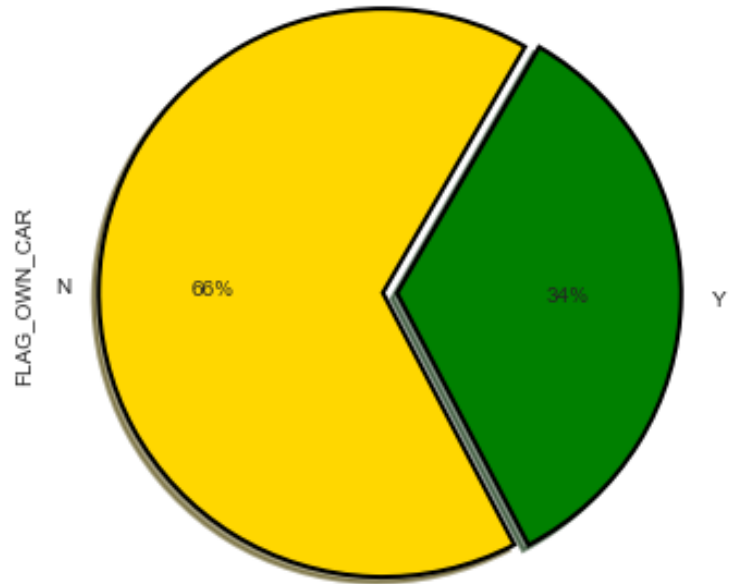
DISTRIBUTION OF CLIENT'S AGE BY REPAYMENTS STATUS



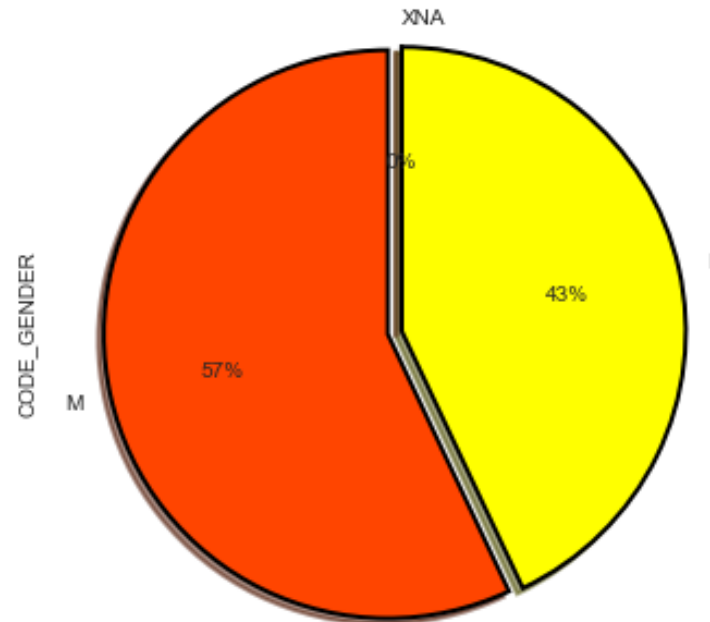
Inference:- Client's of age group 30-40 years get the loan in maximum numbers, also they are the one who are become major part of defaulters.

DISTRIBUTION OF CLIENT OWNING A CAR OR NOT BY GENDER

distribution of client owning a car



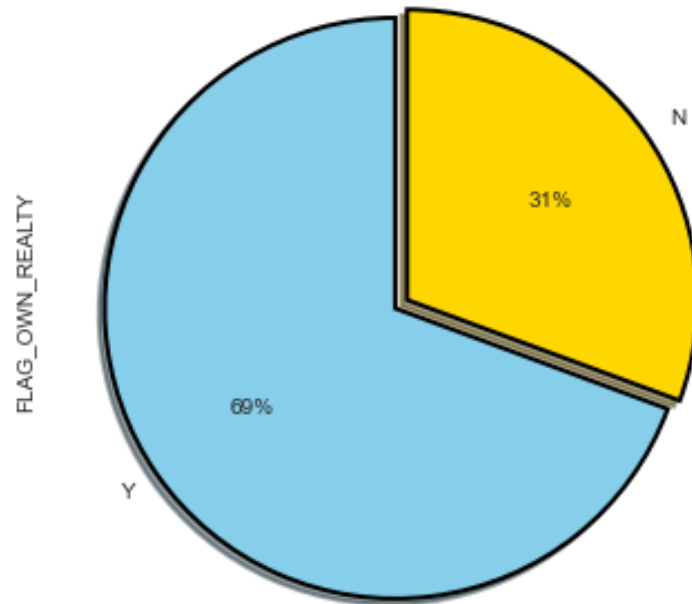
distribution of client owning a car by gender



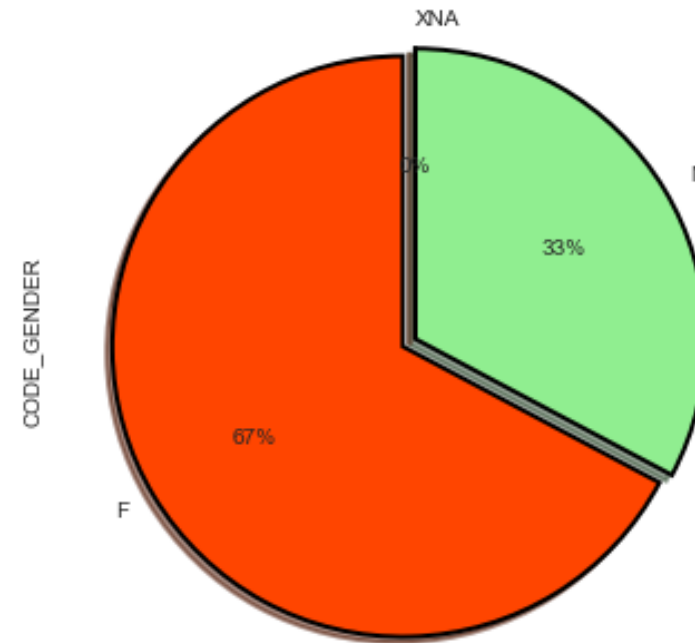
Inference:- Only 34% out of total clients own a car and out of this 34% clients, 57% are males and 43% females.

DISTRIBUTION OF CLIENT OWNING A FLAT BY GENDER

Distribution of client owns a house or flat

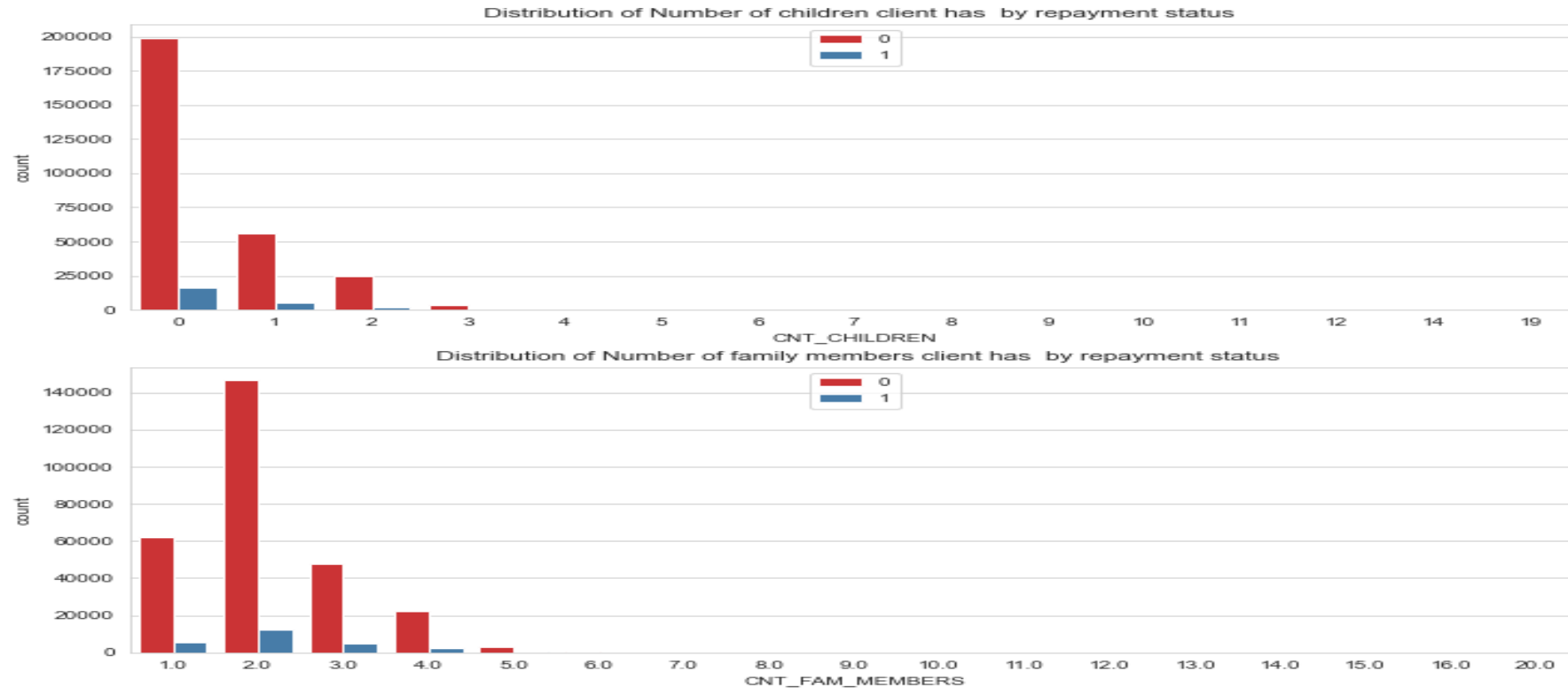


Distribution of client owning a house or flat by gender



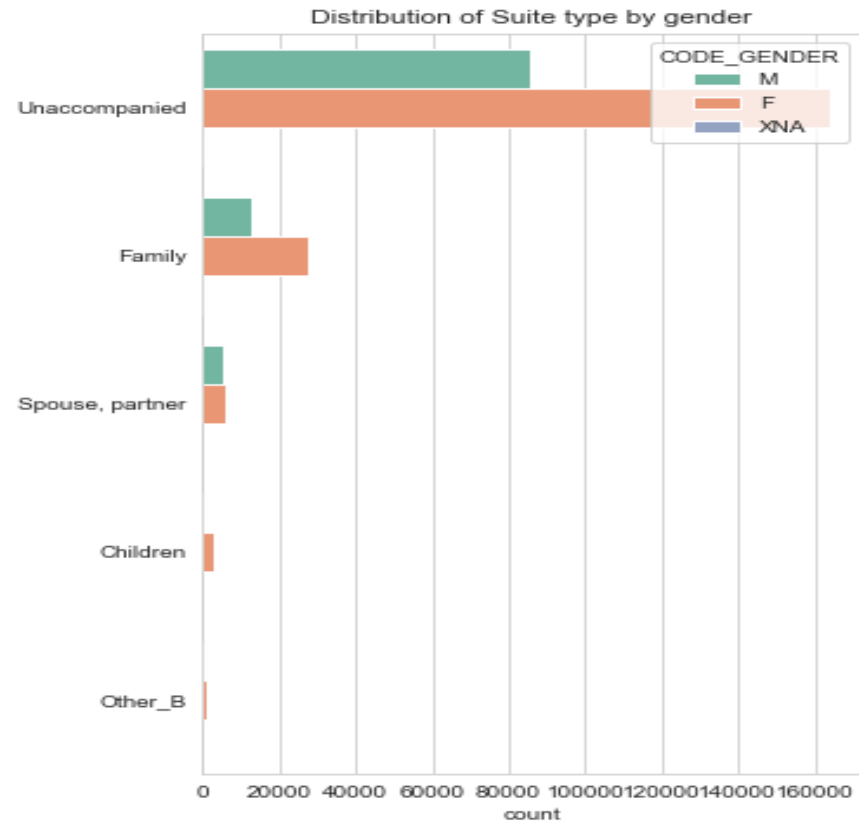
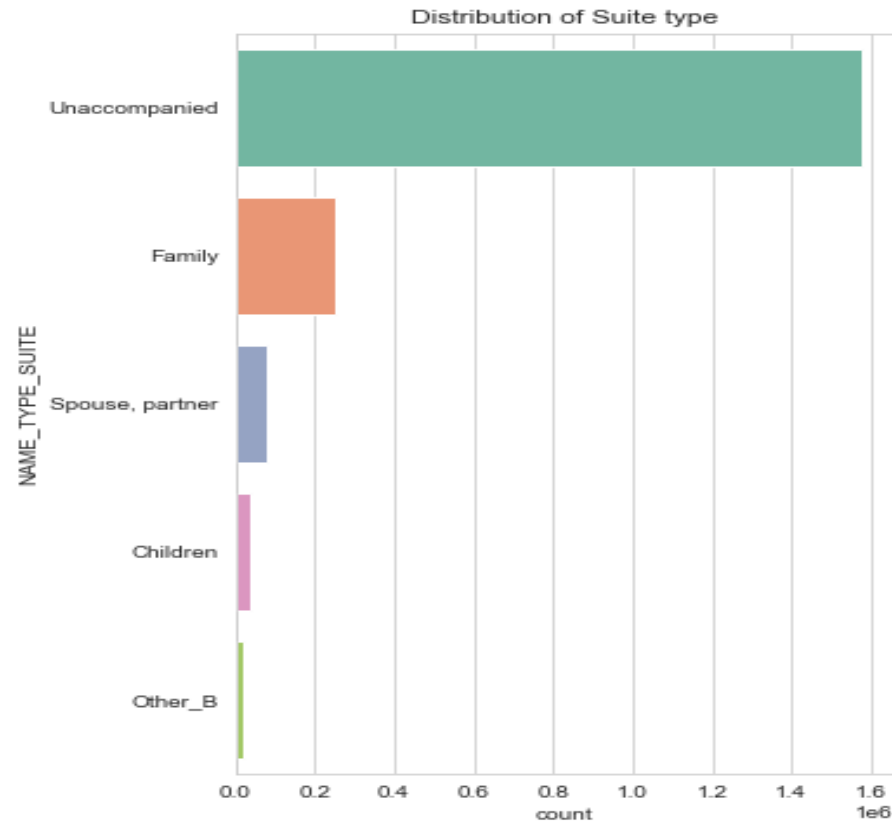
Inference:- only 31% out of total clients own house or flat and in these 31% clients 67% are female owners while 33% are male owners.

DISTRIBUTION OF NUMBER OF CHILDREN AND FAMILY MEMBERS OF CLIENT BY REPAYMENT STATUS



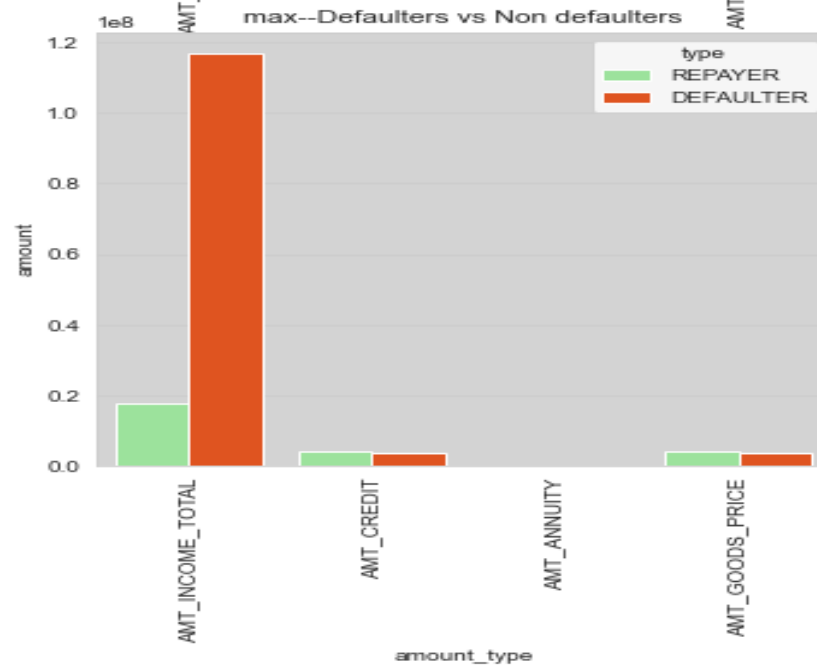
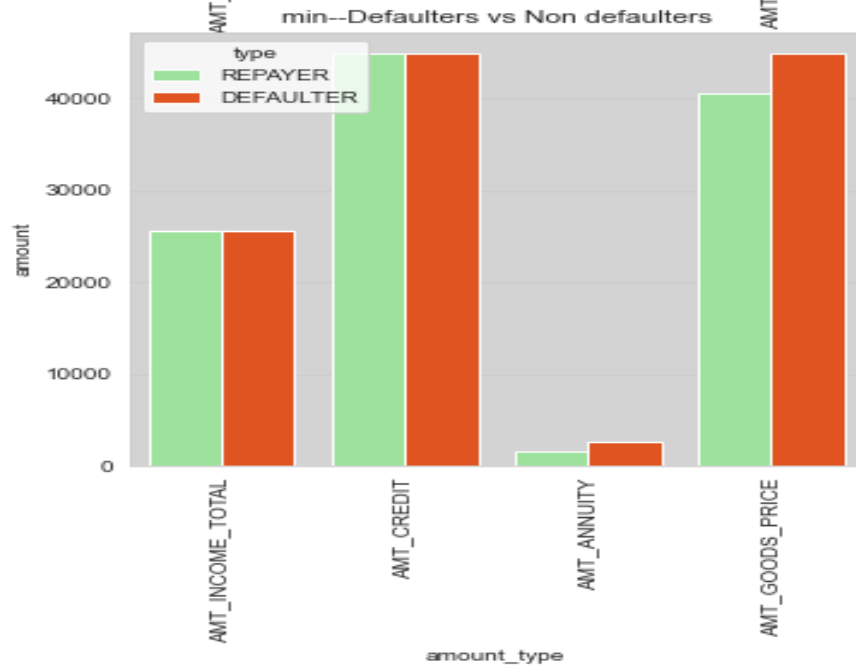
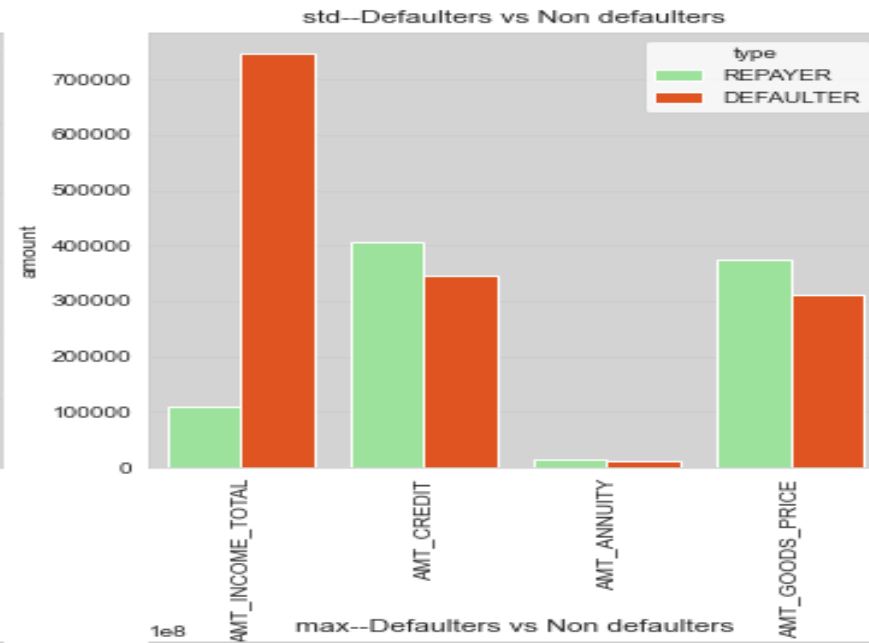
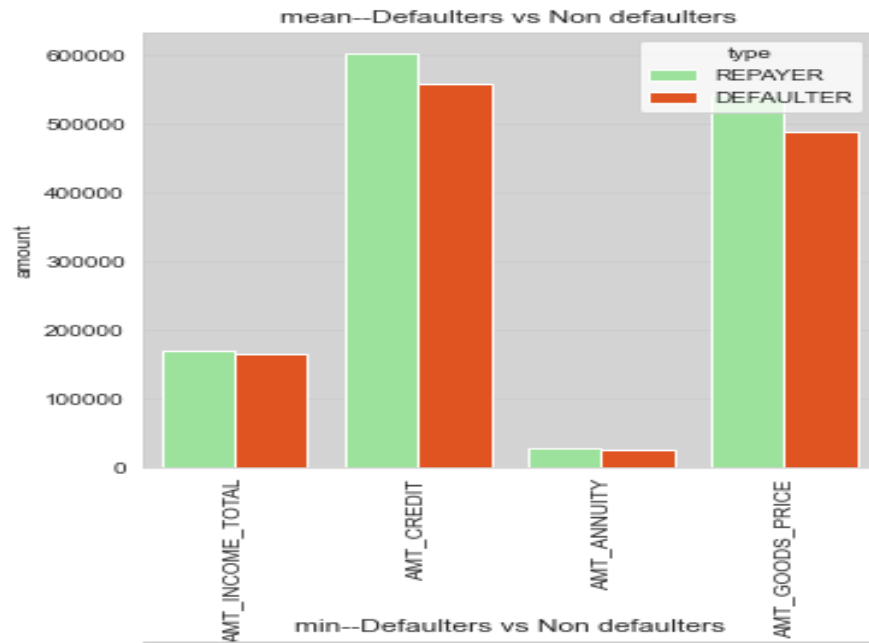
Inference:- Most of the clients having no children whether they are defaulter or not. But majority of clients are having 2 member in the family who are getting loan, also these are the one who make large chunk of defaulter as well.

DISTRIBUTION OF SUITE TYPE



INFERENCE:- Mostly clients unaccompanied when they were applying for the loan. Out of the total unaccompanied clients, female clients were more in number.

COMPARING SUMMARY STATISTICS BETWEEN DEFAULTER AND NON- DEFAULTER FOR LOAN AMOUNT



Inference

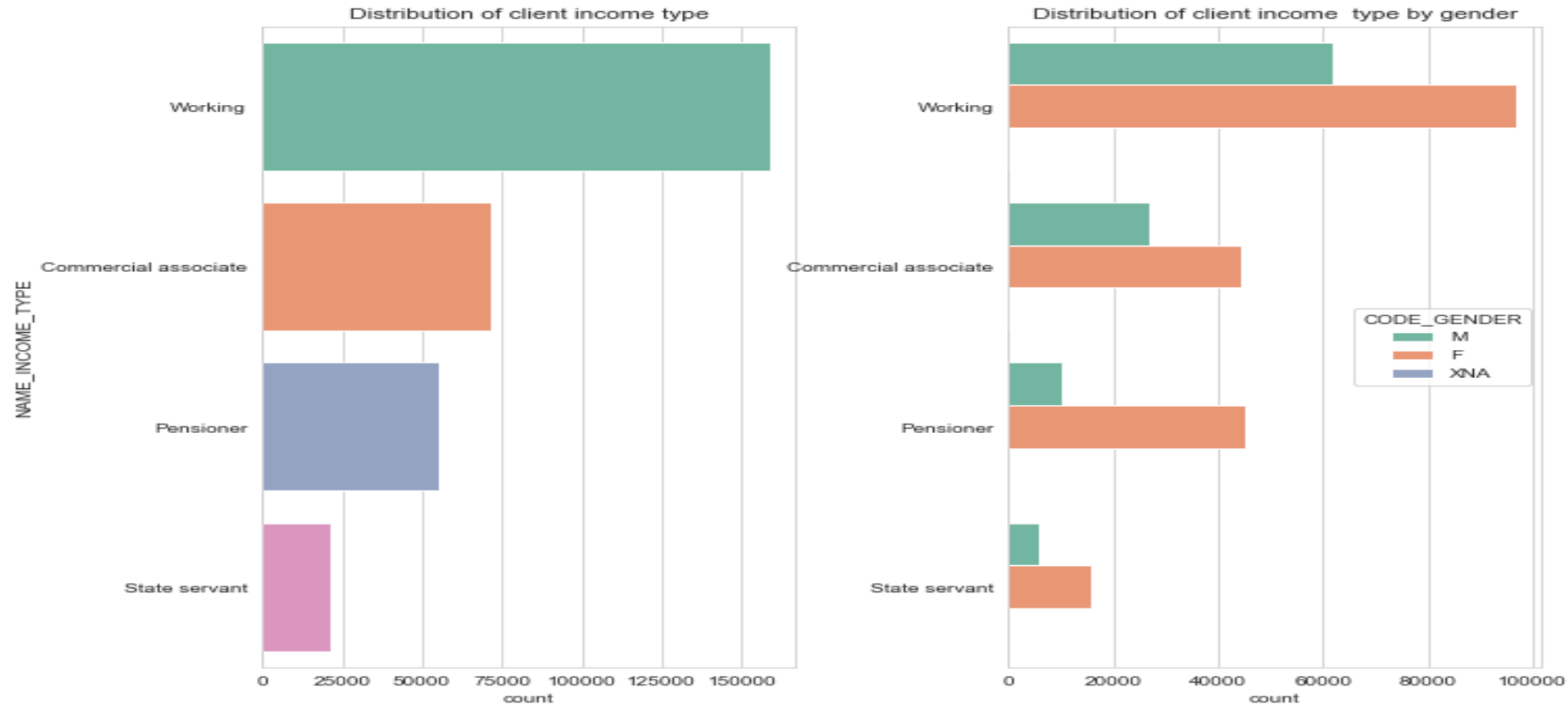
Income of client :- 1 . Average income of clients who default and who do not are almost same.

2 . Standard deviation in income of client who default is very high compared to who do not default that, means defaulter divert his/her income into many other things as well.

3 . Clients who default also has maximum income earnings than who repay the loan.

Credit amount of the loan ,Loan annuity, Amount goods price :- 1 . Statistics between credit amounts, Loan annuity and Amount goods price given to clients who default and who don't are almost similar.

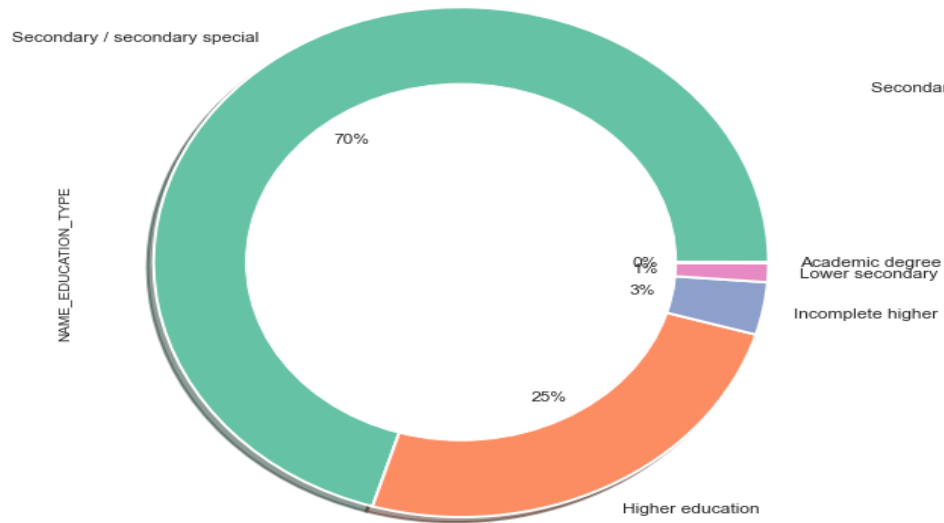
DISTRIBUTION OF CLIENT INCOME TYPE



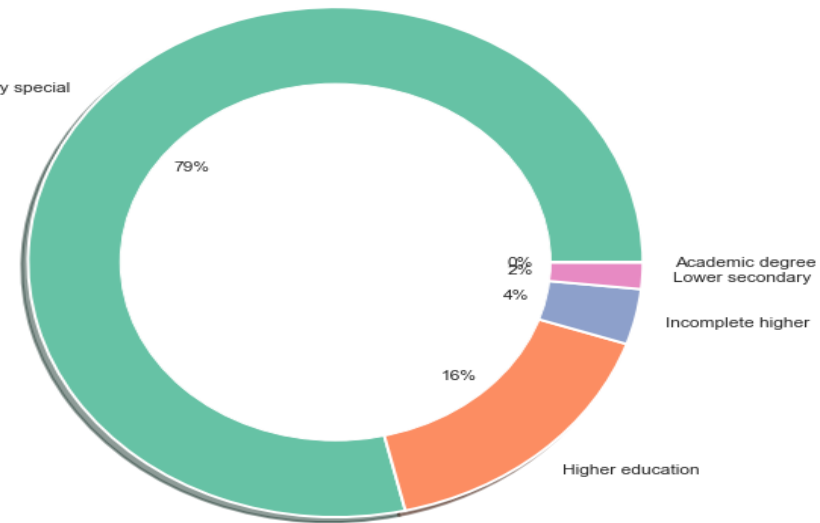
INFERENCE:- clients applied for loan are highest from working class than commercial associate and others. also female clients are more in number than males in every income type.

DISTRIBUTION OF EDUCATION TYPE BY LOAN REPAYMENT STATUS

Distribution of Education type for Repayers



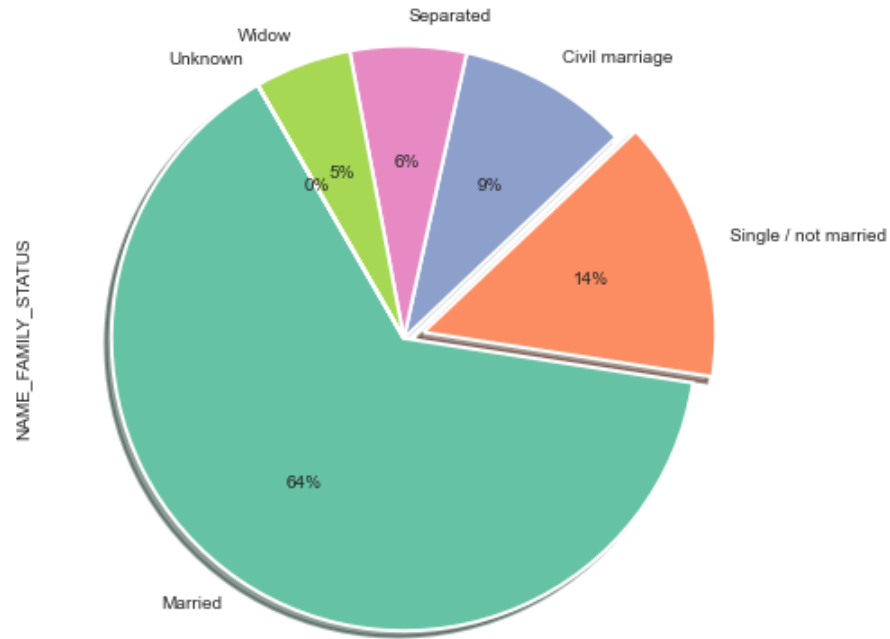
Distribution of Education type for Defaulters



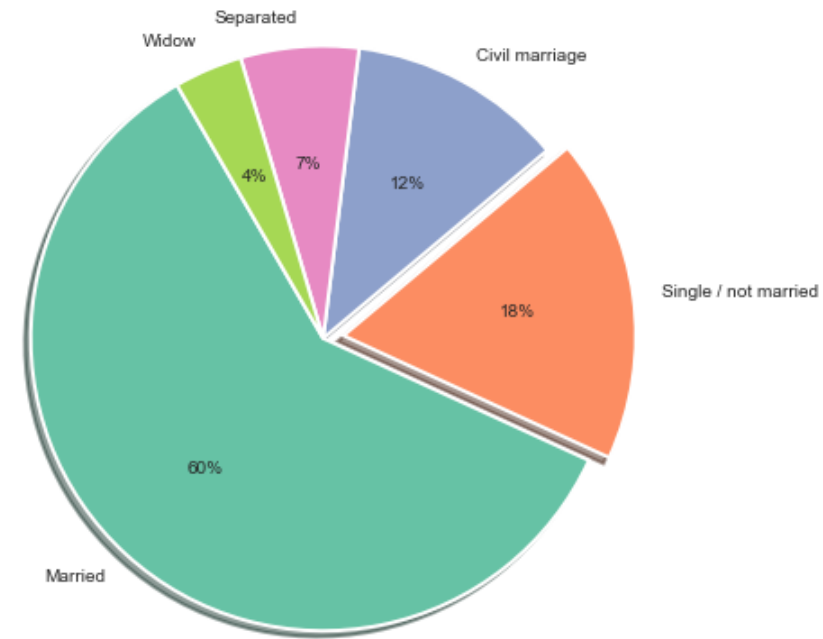
INFERENCE:- Clients who default have proportionally 9% less higher education compared to clients who do not default.

DISTRIBUTION OF FAMILY STATUS TYPE BY LOAN REPAYMENT STATUS

Distribution of Family status for Repayers

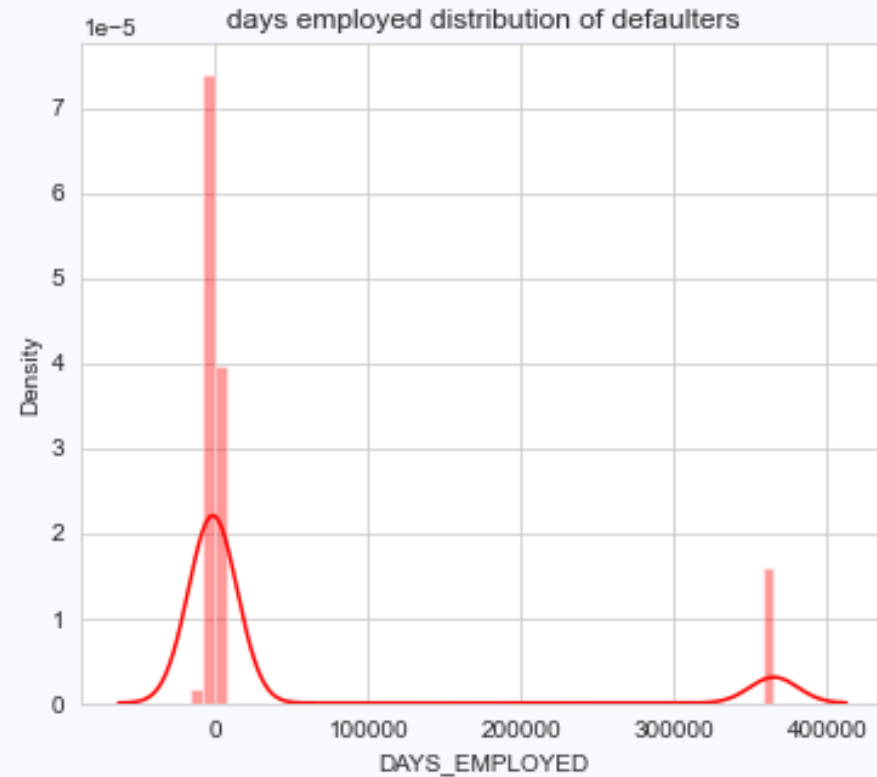
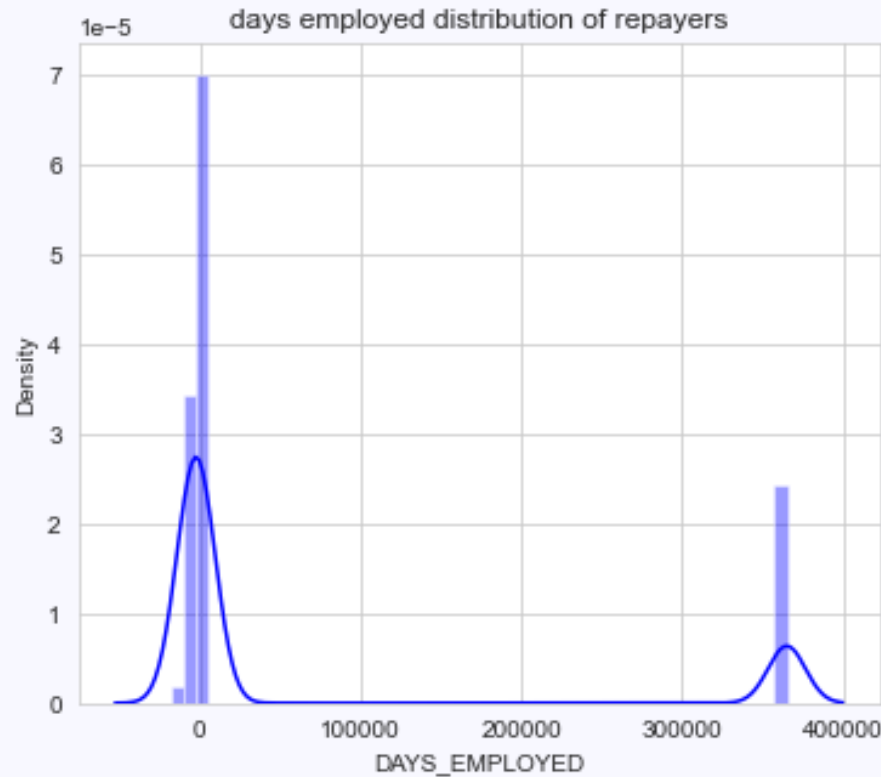


Distribution of Family status for Defaulters



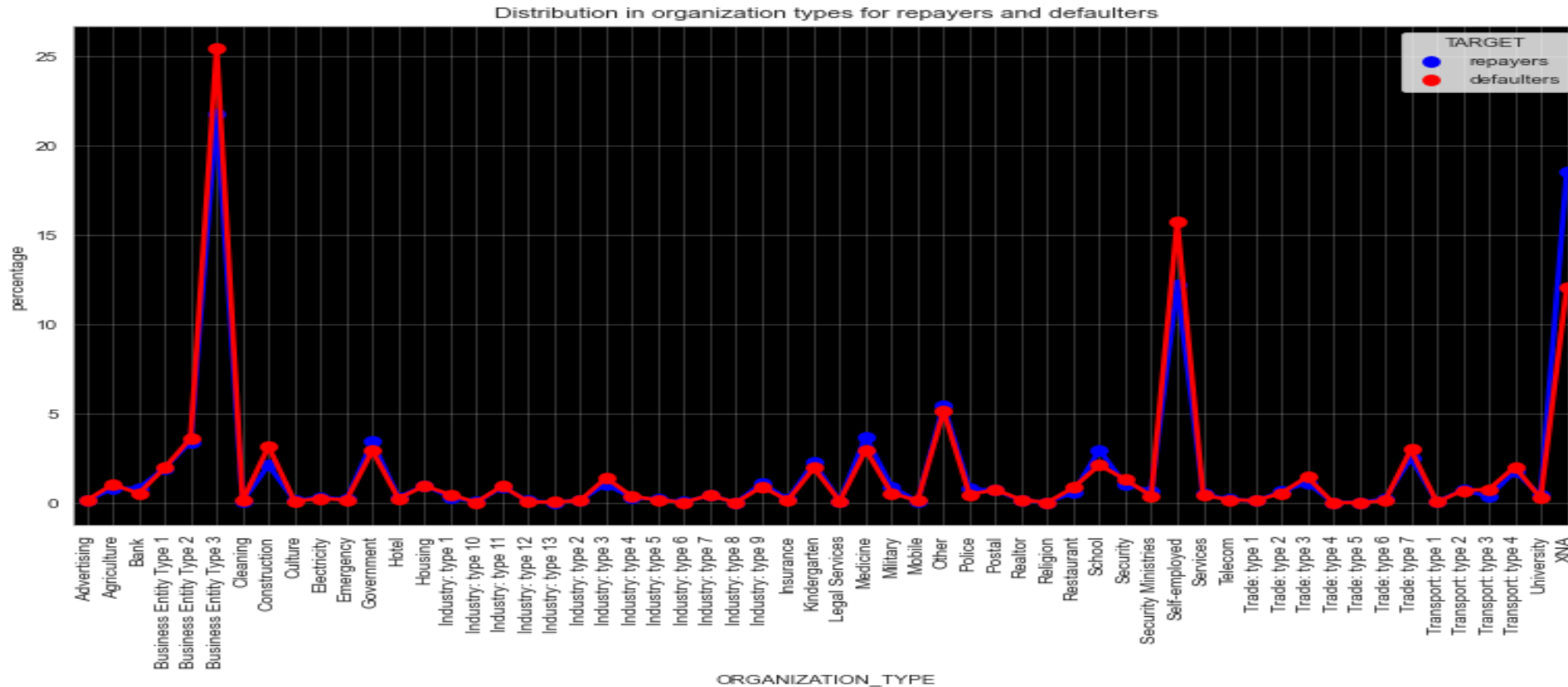
INFERENCE:-Percentage of single people and Civil marriage are more in defaulters than others.

DISTRIBUTION OF DAYS EMPLOYED FOR TARGET VARIABLE



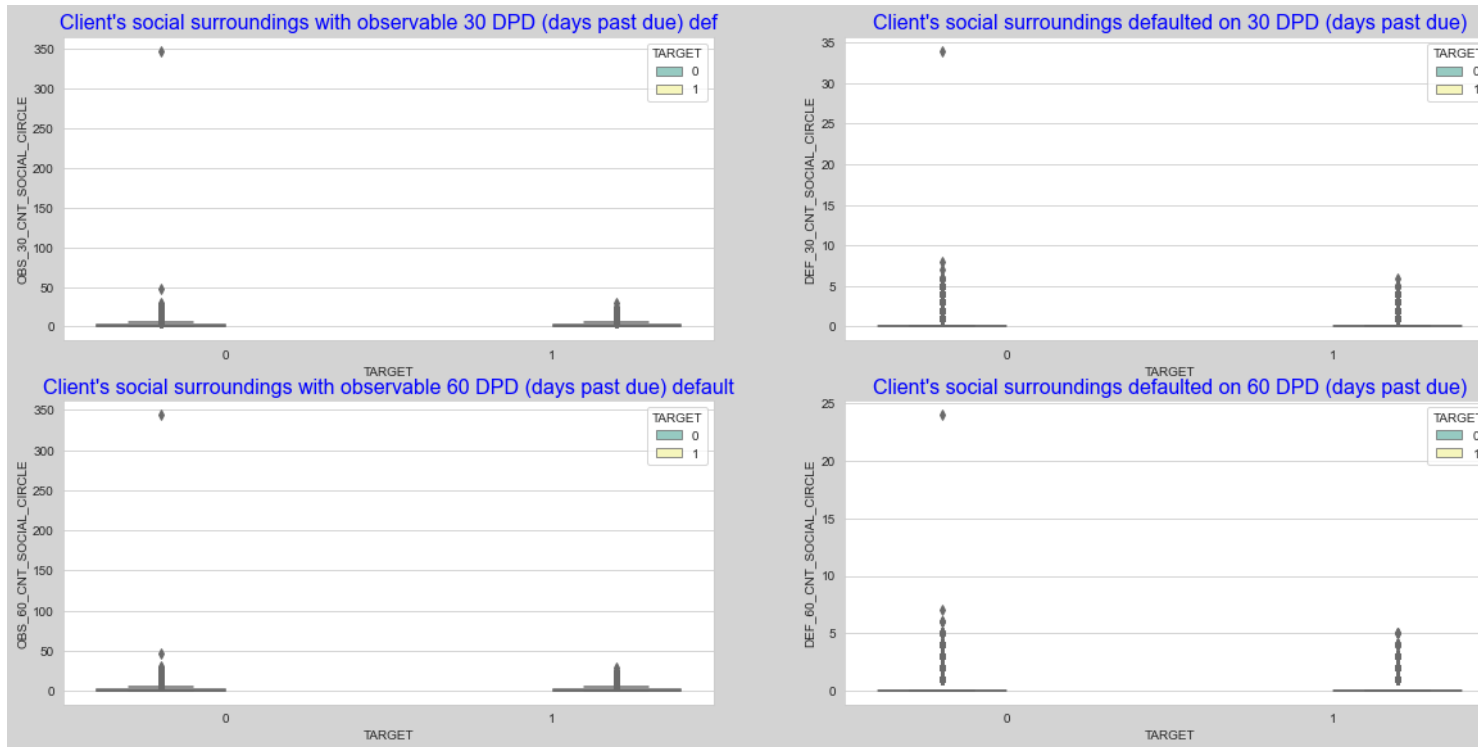
INFERENCE:- Highest number of clients are from the first 5 years of their career and second highest are from the last years of their career.

DISTRIBUTION IN ORGANIZATION TYPES FOR REPAYERS AND DEFAULTERS



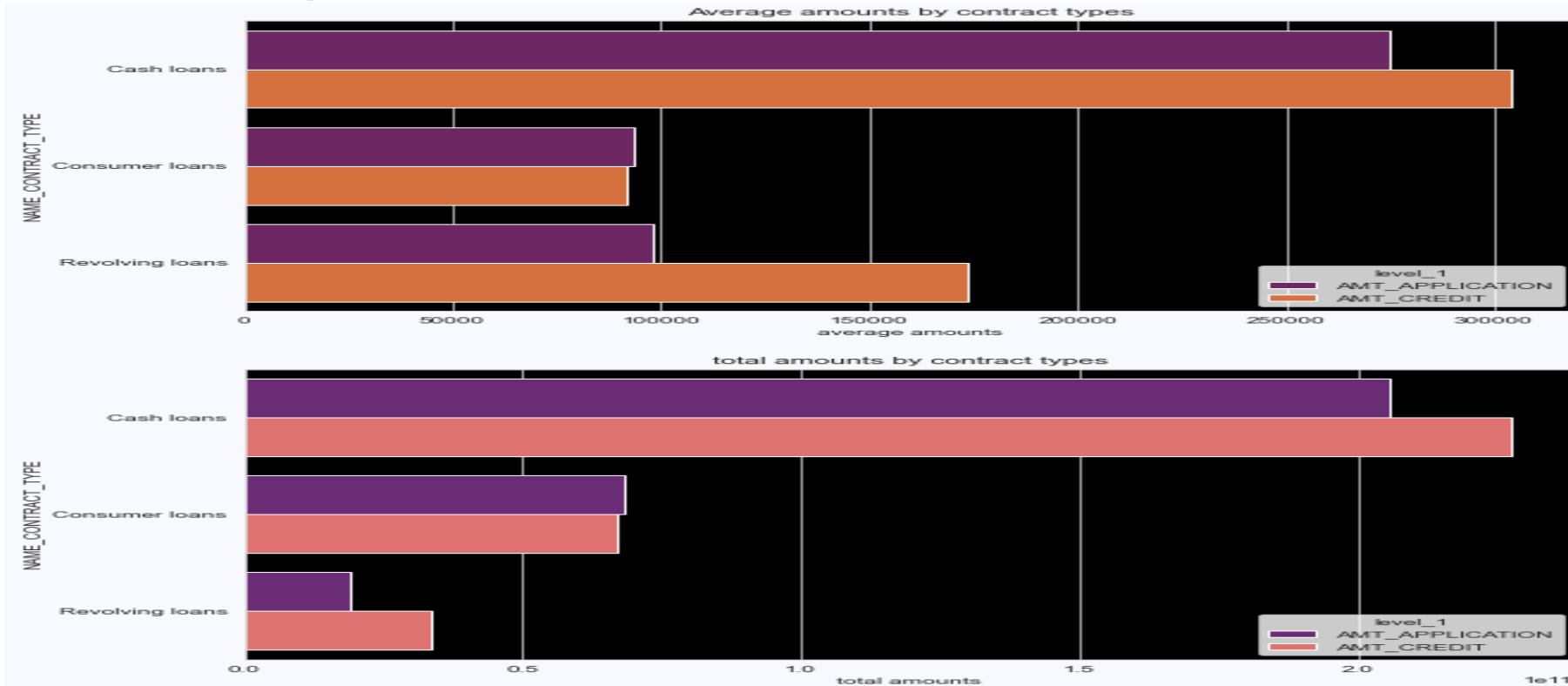
INFERENCE:- Business Entity Type 3 organization grab highest percentage of loans after that self-employed and then others.

DISTRIBUTION CLIENT'S SOCIAL SURROUNDINGS WITH OBSERVED AND DEFAULTED 30 DPD (DAYS PAST DUE)



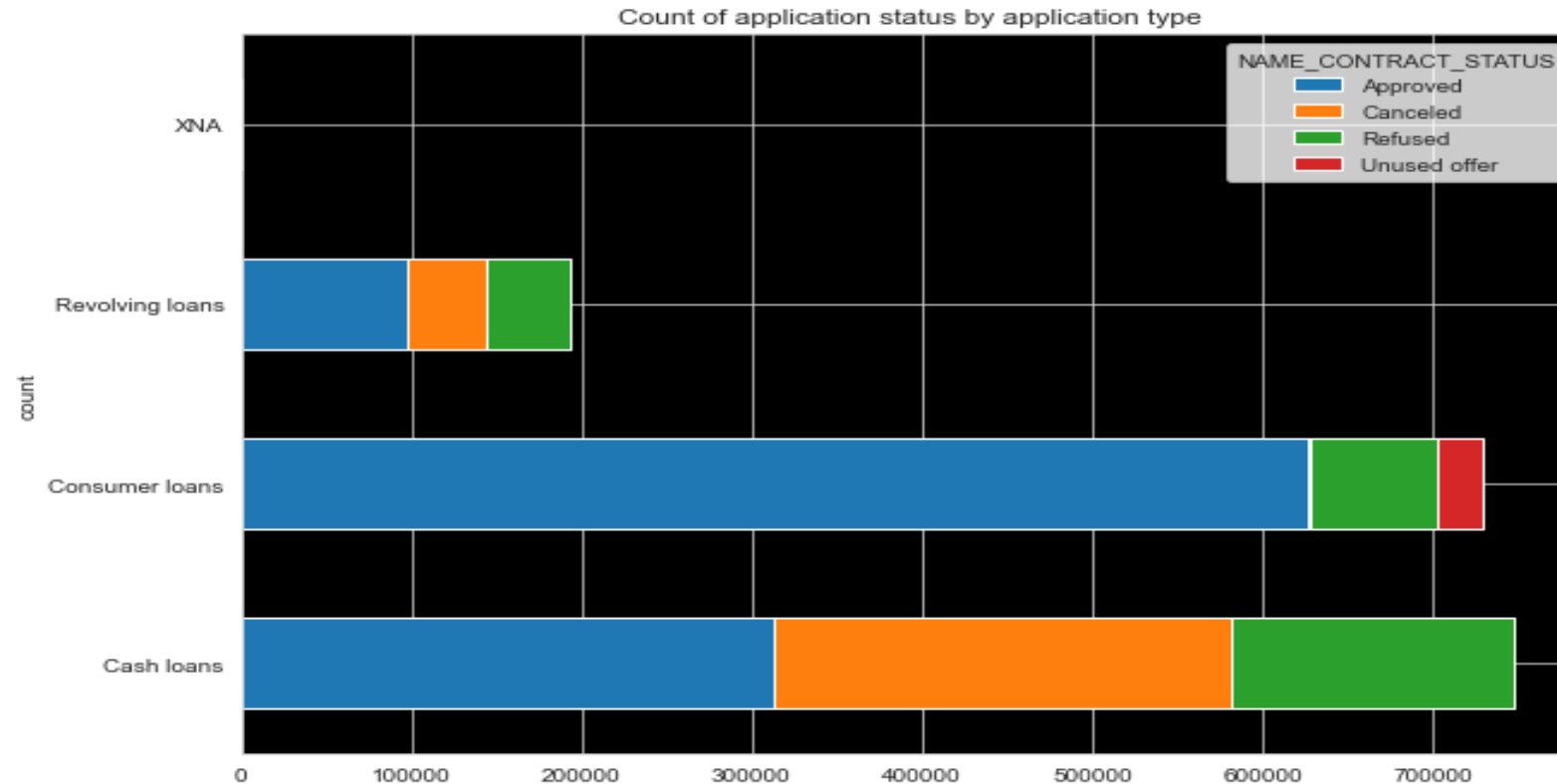
INFERENCE:- People after becoming defaulter tend to get less social than they were earlier used to.

TOTAL AND AVERAGE AMOUNTS APPLIED AND CREDITED IN PREVIOUS APPLICATIONS



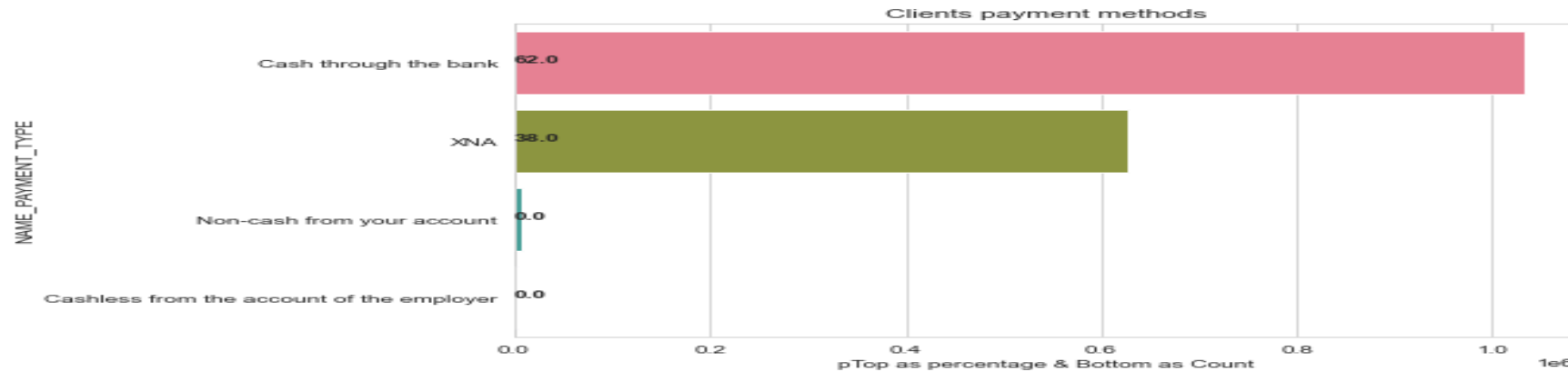
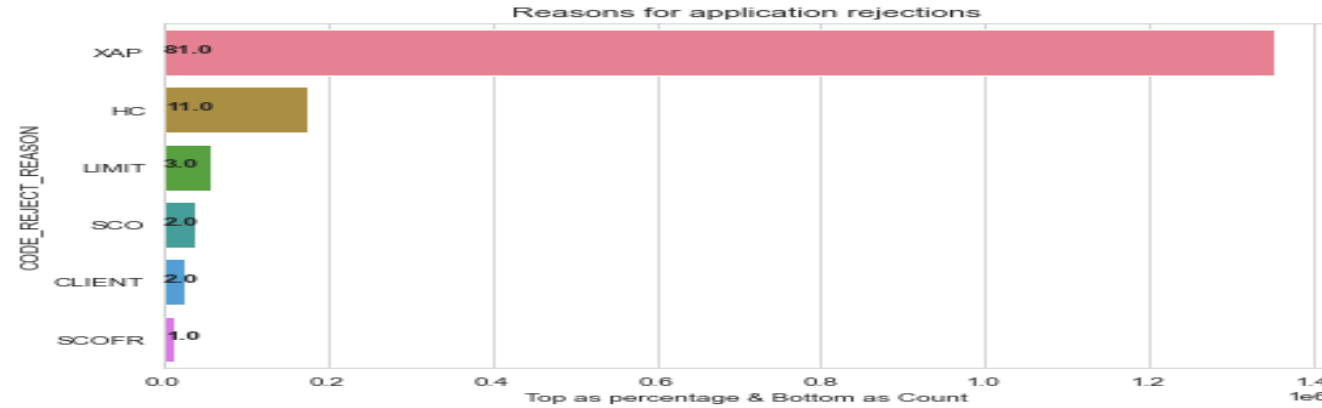
INFERENCE:- Total as well as Average Amount of credit is more than the Amount of application in Cash loan contracts and revolving loan contracts while they are less in consumer loan contracts.

COUNT OF APPLICATION STATUS BY APPLICATION TYPE



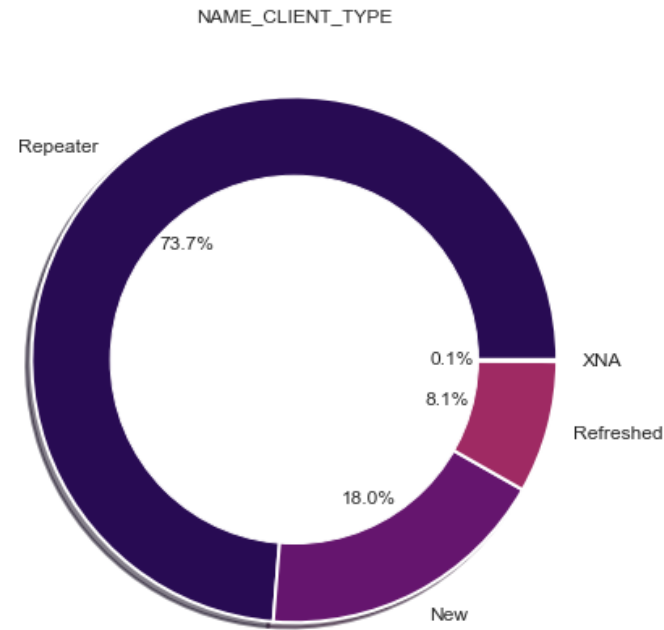
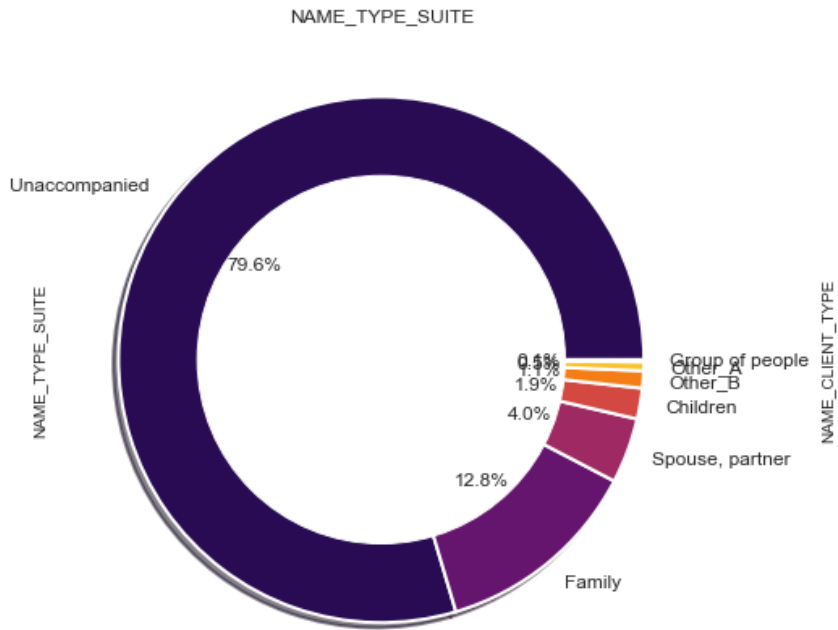
Inference:- Consumer loan applications are most approved loans and cash loans are most cancelled and refused loans.

CLIENT PAYMENT METHODS & REASONS FOR APPLICATION REJECTIONS



**Inference:- Around 81% of rejected applications the reason is XAP.
62% of chose to pay through cash by bank for previous applications.**

DISTRIBUTION IN CLIENT SUITE TYPE & CLIENT TYPE



Inference:- About 60% clients are un-accompained when applying for loans. 73% clients are old clients

MULTIVARIATE ANALYSIS

AVERAGE INCOME, CREDIT, ANNUITY & GOODS PRICE BY GENDER

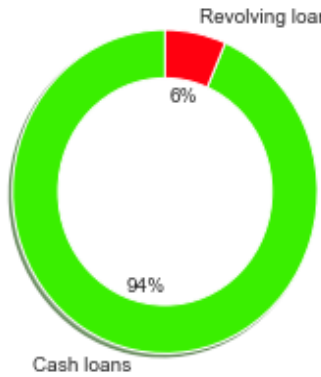


Inference:-

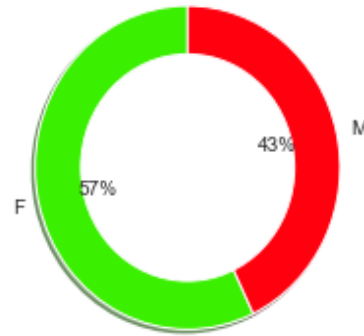
1. Average income is highest for male and least for female.
2. Amount of credit, annuity and Amount of goods price all are again highest for male but least for XNA.

***DISTRIBUTION OF CONTRACT TYPE,
GENDER, OWN CAR, OWN HOUSE WITH
RESPECT TO REPAYMENT
STATUS(TARGET VARIABLE)***

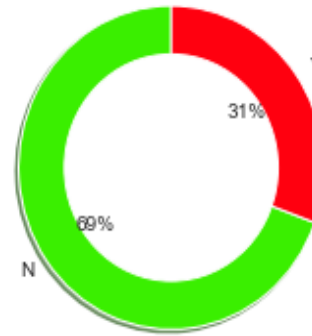
NAME_CONTRACT_TYPE-Defaulter



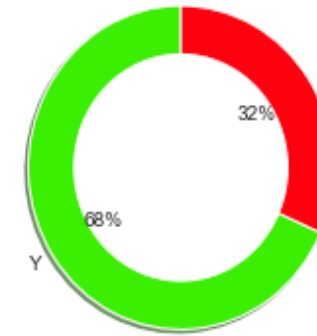
CODE_GENDER-Defaulter



FLAG_OWN_CAR-Defaulter



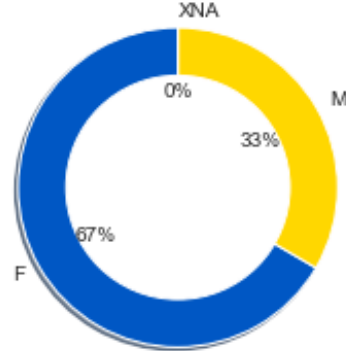
FLAG_OWN_REALTY-Defaulter



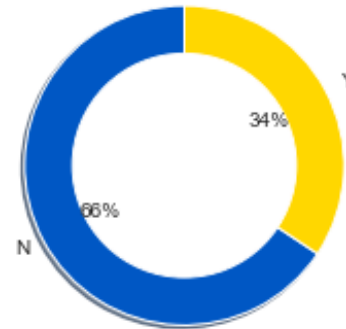
NAME_CONTRACT_TYPE-Repayer



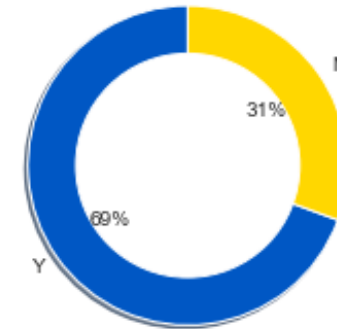
CODE_GENDER-Repayer



FLAG_OWN_CAR-Repayer

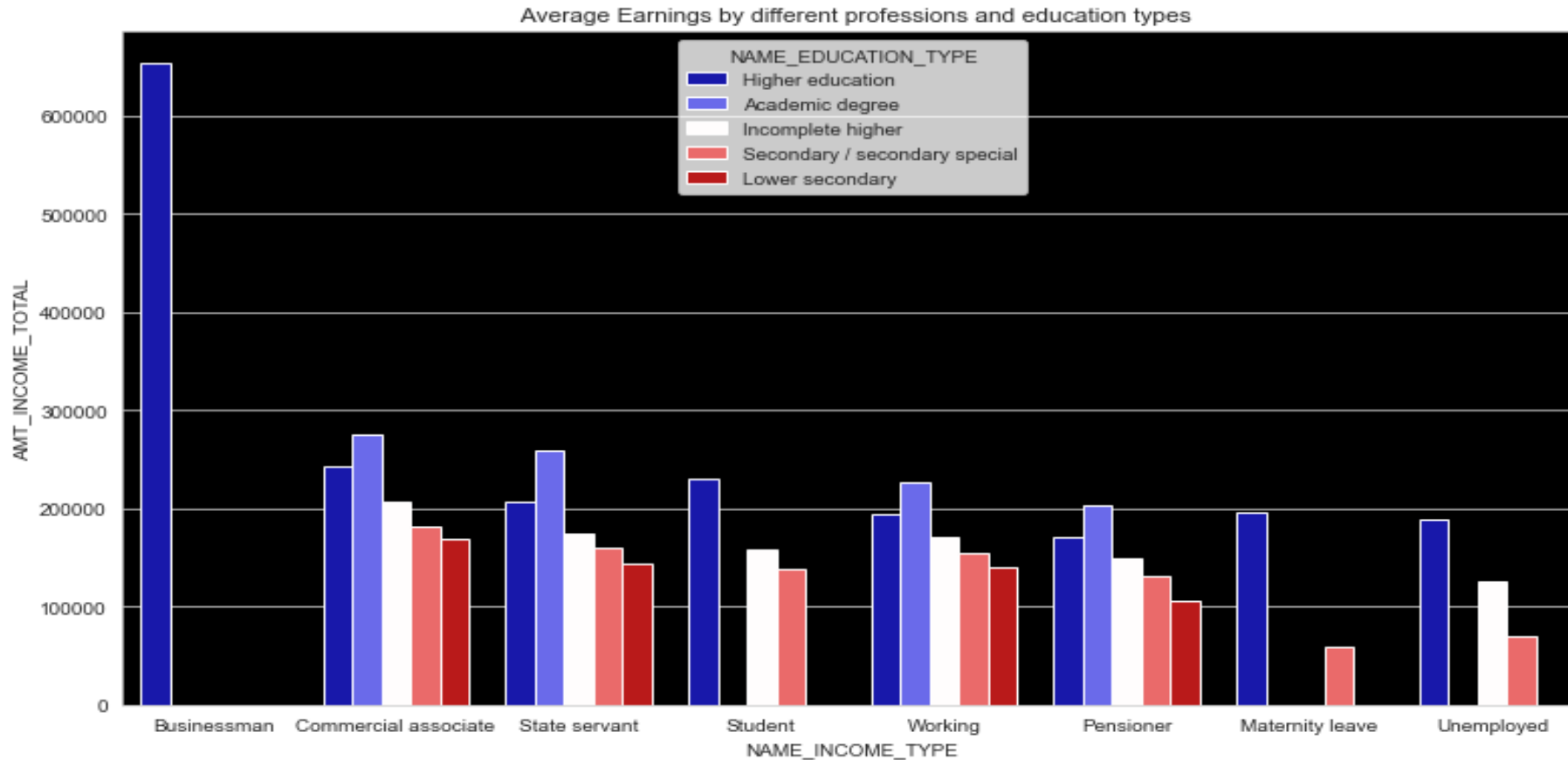


FLAG_OWN_REALTY-Repayer



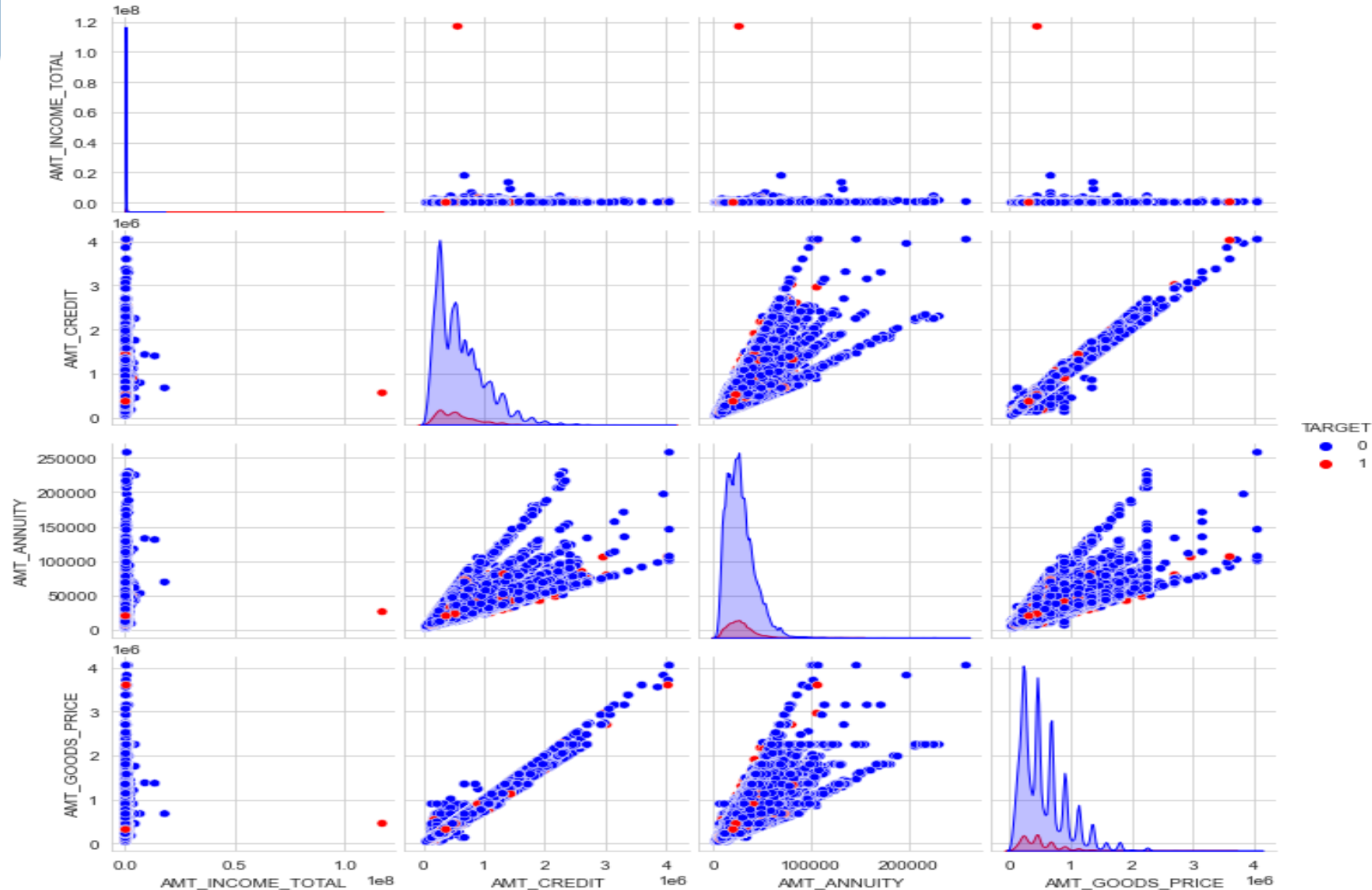
- Inference:-**
1. Percentage of revolving loans are less in defaulters than repayers.
 2. Percentage of males is 10% more in defaulters category than repayers.
 3. Percentage of car owners is less in defaulter than repayers.
 4. Percentage of house owners is less in defaulter than repayers.

AVERAGE EARNINGS BY DIFFERENT PROFESSIONS AND EDUCATION TYPES



INFERENCE:- Businessman clients are having higher education only, while in other categories of income type clients with academic degrees are more. But Students mainly apply for loan for higher education.

PAIR PLOT BETWEEN AMOUNT VARIABLE



CORRELATION BETWEEN VARIABLES

CONCLUSION

- ❑ Banks should focus on small family sized clients, as there will be less diversion of funds and more chances of repayment.
- ❑ Banks should be more careful while approving loans for business entities, self-employed, industry and trade sector, as their percentage of default is higher than repayment. on the other hand sectors like medicine, government, schools etc have higher percentage of repayment.
- ❑ Banks need to avoid income type of 'Working' clients as they have high percentage of paying difficulties. Instead focus on Commercial associate, pensioner and State servant.
- ❑ Banks should focus more on education type 'Higher education' and avoid Secondary/secondary special, incomplete higher or lower secondary as they face paying difficulties.
- ❑ People with high income status have more deviation in income and also are maximum defaulters than the corresponding low and middle income clients.

THANK YOU