# CREDIT EDA ASSIGNMENT

**SAKSHI RAI** 

Signature

### **OBJECTIVE**

Developing a risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money lending to customer.

To Identify pattern of loan denied, lending loan at high interest rate(risky applicants) and reducing loan amount for risky clients.

Identifying loan defaulter by EDA to ensure client capable of loan are not rejected.

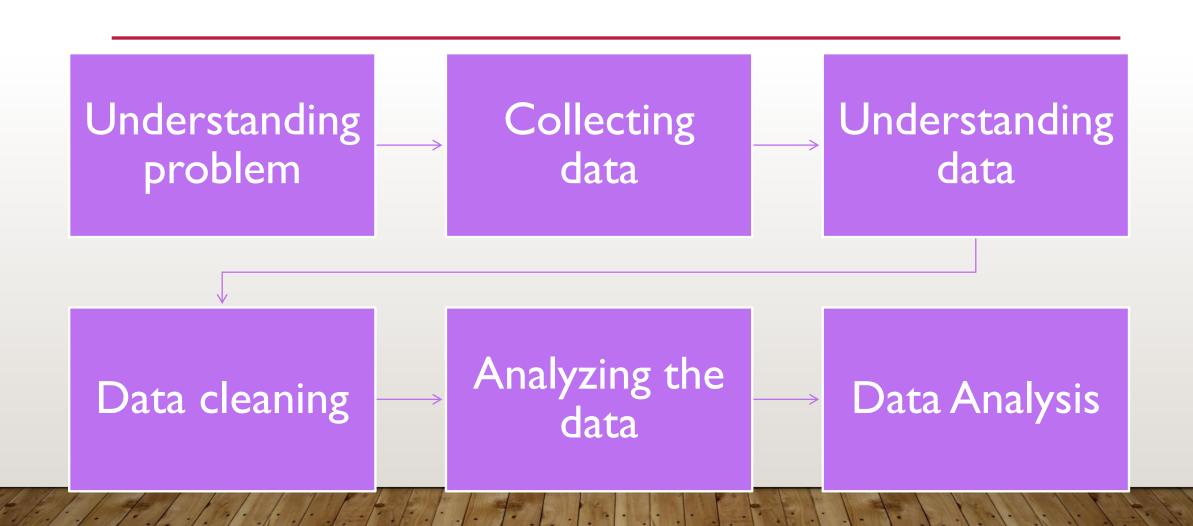
#### PROBLEM STATEMENT

Understanding the factors for loan defaulter criteria.

Variables influencing the criteria.

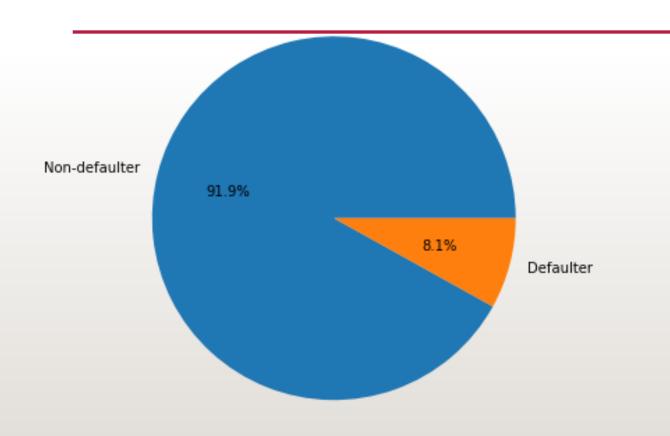
Hypothesis:If client is able to repay loan than disapproval may lead to loss for bank, incase of defaulter approval of loan will result loss to bank.

#### **APPROACH USED:**



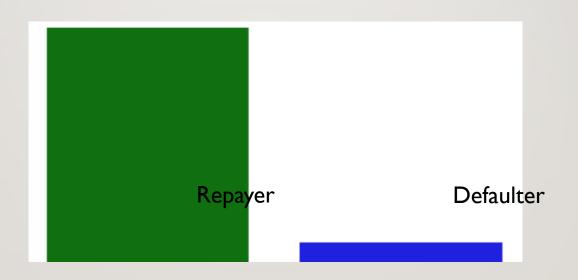
#### Target Variable

#### DATA INSPECTION



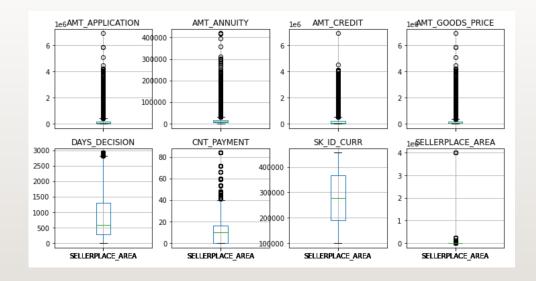
Data Imbalance was found among Repayer and defaulter

## RATION OF IMBALANCE



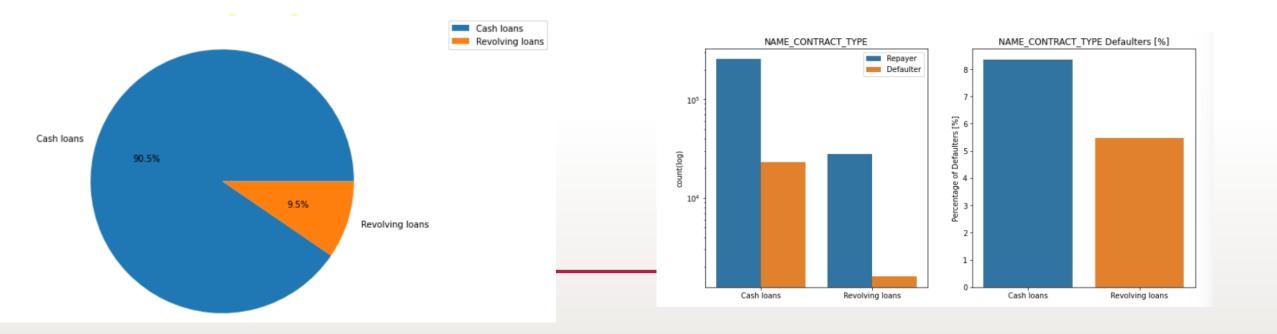
Ratio of repayer and defaulter 11:39:1

#### **OUTLIERS**



- AMT\_ANNUITY, AMT\_APPLICATION, AMT\_CREDIT, AMT\_GOODS\_PRICE, SELLERPLACE\_AREA has more number of outliers.
- CNT\_PAYMENT, DAYS\_DECISION have some outlier values.
- SK\_ID\_CURR had no outlier.

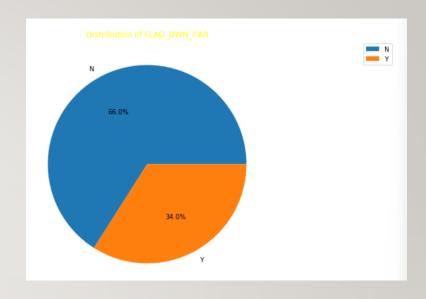
# UNIVRATE ANALYSIS

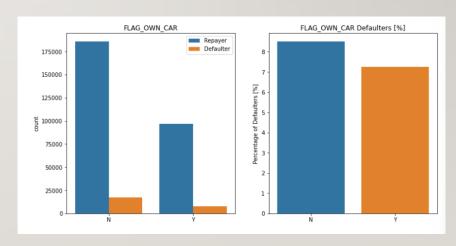


#### NAME CONTRACT TYPE

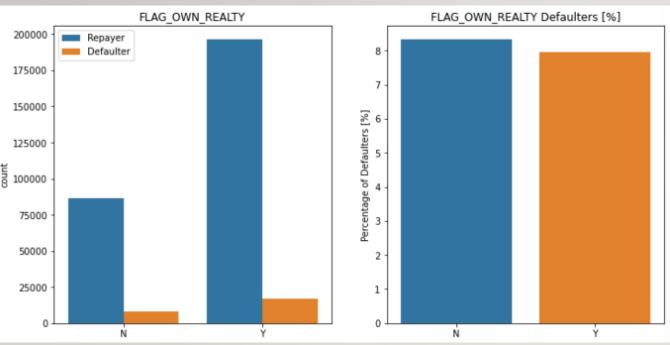
#### FLAG OWN CAR

CUSTOMERS OWNING A CAR ARE HALF IN NUMBER OF THE CUSTOMER WHO DON'T OWN A CAR BUT BASED ON THE PERCENTAGE OF DEFAULT, THERE IS NO CORRELATION BETWEEN OWNING A CAR AND LOAN REPAYMENT AS IN BOTH CASES THE DEFAULT PERCENTAGE IS ALMOST SAME.



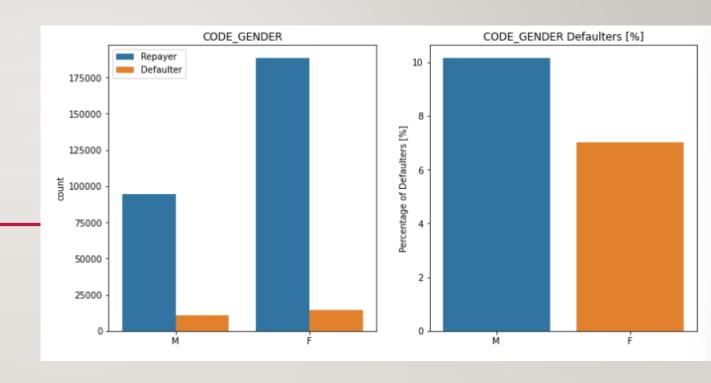


## FLAG\_OWN\_REALTY



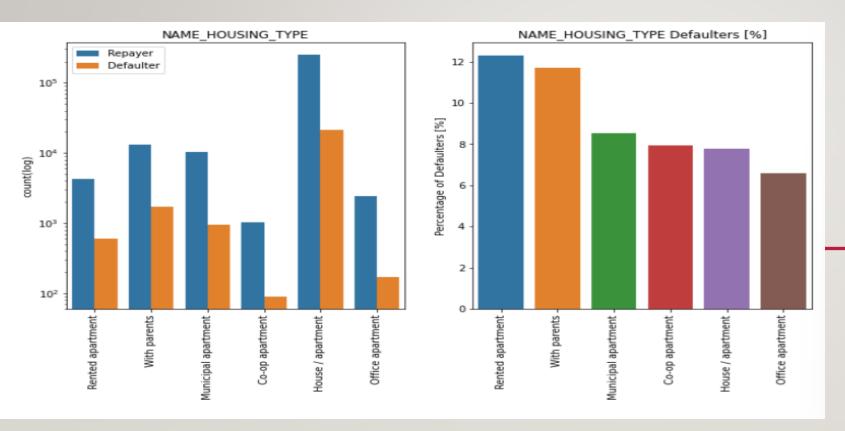
THE CUSTOMER OWNING REAL ESTATE ARE MORE THAN DOUBLE OF THE ONES THAT DON'T OWN. BUT THE DEFAULTING RATE OF BOTH CATEGORIES ARE AROUND THE SAME . HENCE, NO CORRELATION BETWEEN OWNING A REALITY AND DEFAULTING THE LOAN.

## CODE\_GENDER



FEMALE CUSTOMERS ARE MORE THAN MALE CUSTOMERS, BUT IT SEEMS MALE LIST IS MORE AS DEFAULTER THEN FEMALE. SO BASED ON IT MALES GIVES MORE CHANCE OF BEING A DEFAULTER I.E NOT REPAYING THE LOAN.

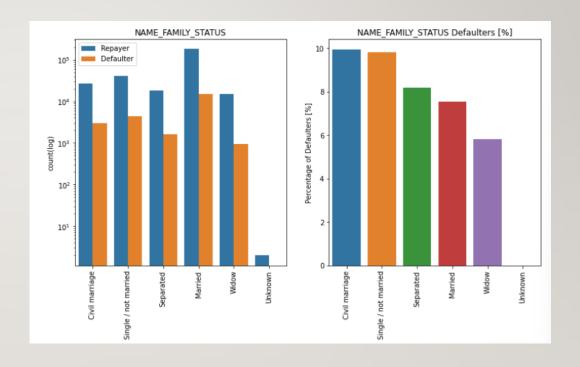
## HOUSING\_TYPE



THERE ARE MORE NO. OF PEOPLE LIVING IN HOUSE/APARTMENT, WHILE OFFICE APARTMENT HAS LOWEST PEOPLE LIVING. BUT IN DEFAULTERS LIST NO. OF PEOPLE LIVING IN RENTED APARTMENTS AND WITH PARENTS HAVE HIGH CHANCE AS DEFAULTERS.

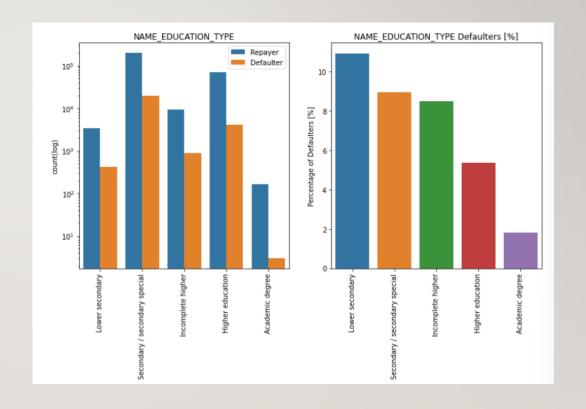
#### NAME\_FAMILY\_STATUS

PEOPLE WHO ARE MARRIED AND SINGLE/ NOT MARRIED HAVE HIGHER RATE OF LOANING, BUT CIVIL MARRIAGE PEOPLE SEEMS TO HAVE HIGHER DEFAULTING RATE(10%), LOWEST RATE GOES TO WIDOWS(6%), EXCEPTION UNKNOWN.

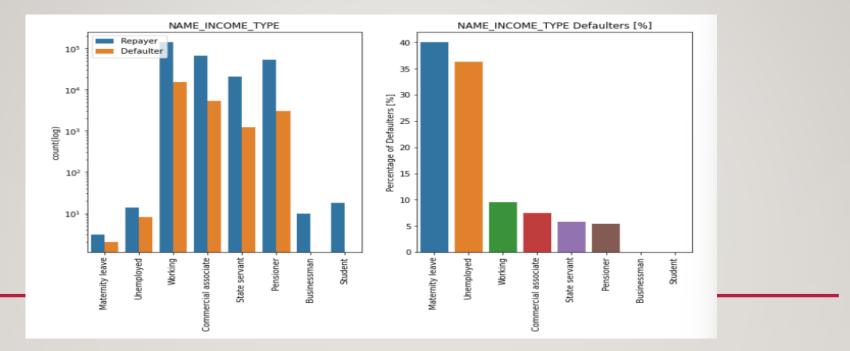


#### NAME\_EDUCATION\_TYPE

SECONDARY/ SECONDARY SPECIAL
EDUCATION FOLLOWED BY CUSTOMERS
WITH HIGHER EDUCATION HAVE MOJORITY
OF CUSTOMERS, LOWER SECONDARY HAVE
HIGHER RATE OF BEING DEFAULTER.

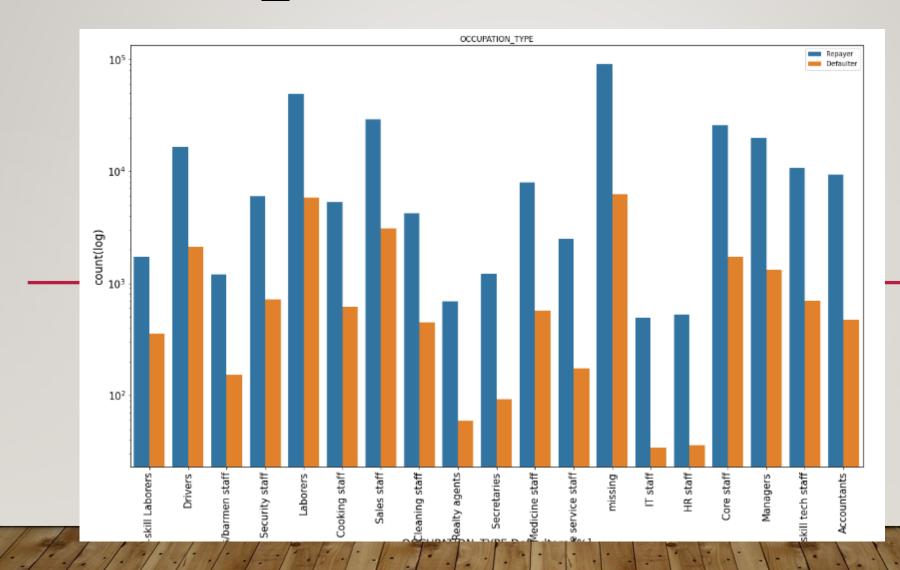


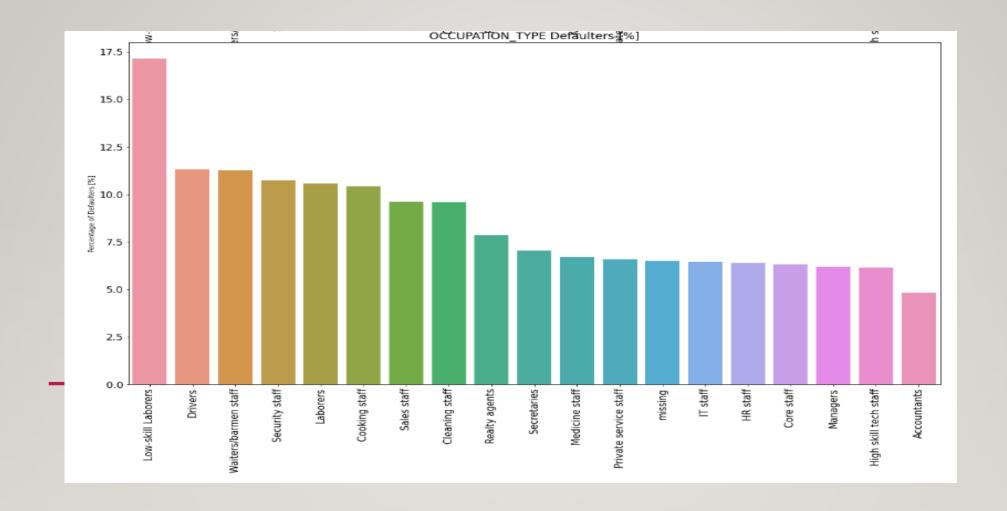
## NAME\_INCOME\_TYPE



Income type as Working, followed by Commercial associate, Pensioner and State servant have more no of applicant. the type of income Maternity leave no of applicants have almost made 40% ratio of the defaulters, followed by Unemployed (37%). The rest of types of incomes are under the average of 10% for not returning loans. Student and Businessmen, though less in numbers do not have any default record. Thus these two category are safest for providing loan.

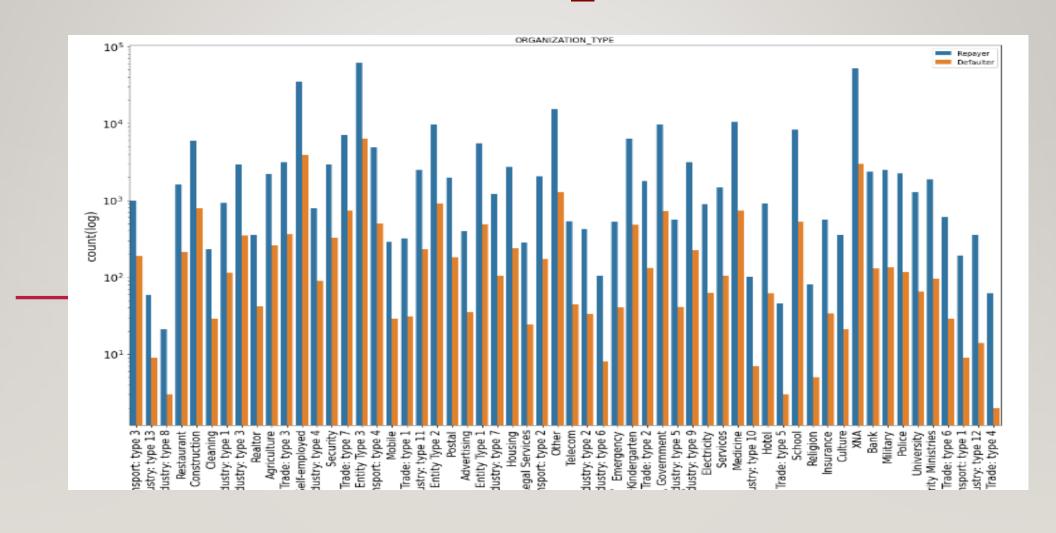
## OCCUPATION\_TYPE

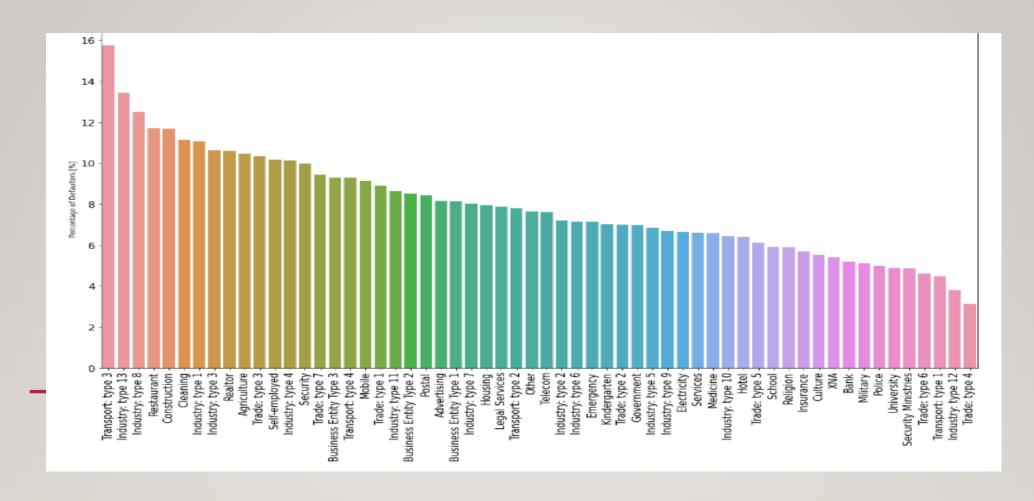




No. of the loans are taken by people by whose occuption is 'missing'is the most in no. Lowest amount of loans have been taken by IT stafs. low-skill laborers have highest percent of not repaid loans.

### ORGANIZATION\_TYPE

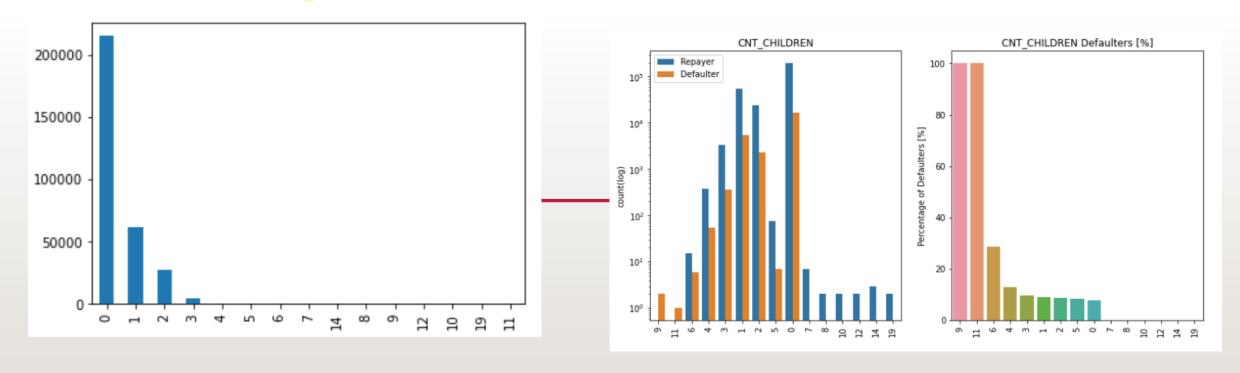




Business Entity type 3 have most no. Organizations with highest percent of loans not repaid are Transport: type 3 (around 16%), Industry: type 13 (13.5%), Industry: type 8 (around 12.5%) and Restaurant (less than 12%). Organization type information is unavailable(XNA)for high no of applicants. Trade Type 4, Industry type 12 can be seen having lesser defaulter.

### ## CNT\_CHILDREN

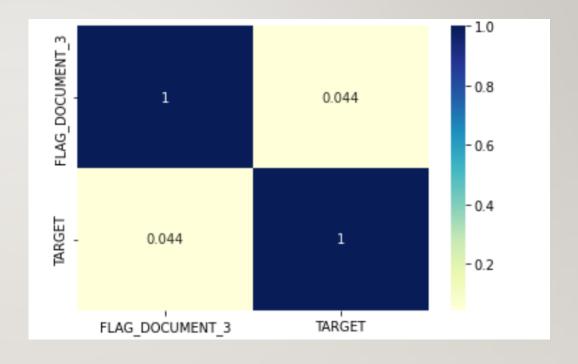
#### Distribution of CNT\_CHILDREN Variable



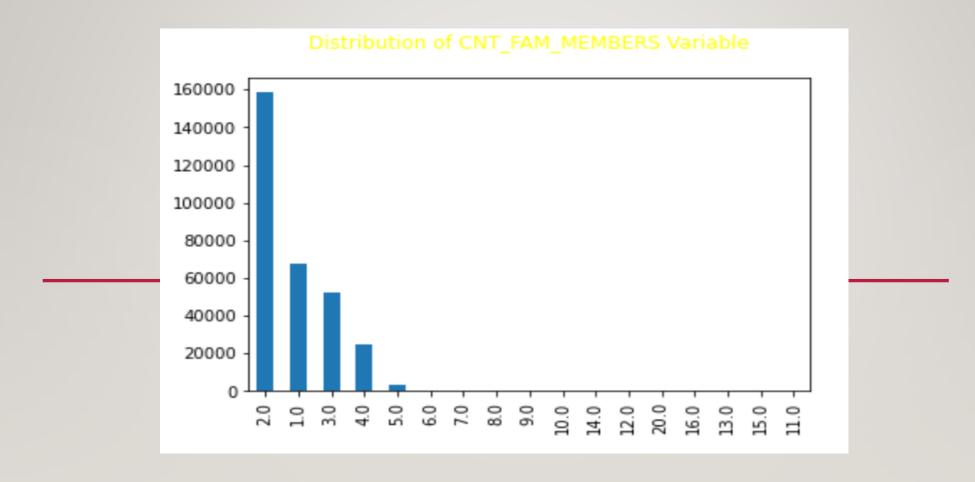
We see there are higher no. of people who don't have children, only some of them are having 3 children. People having children more than 6 on the pltfig. show higher chances of being a defaulter, i.e (

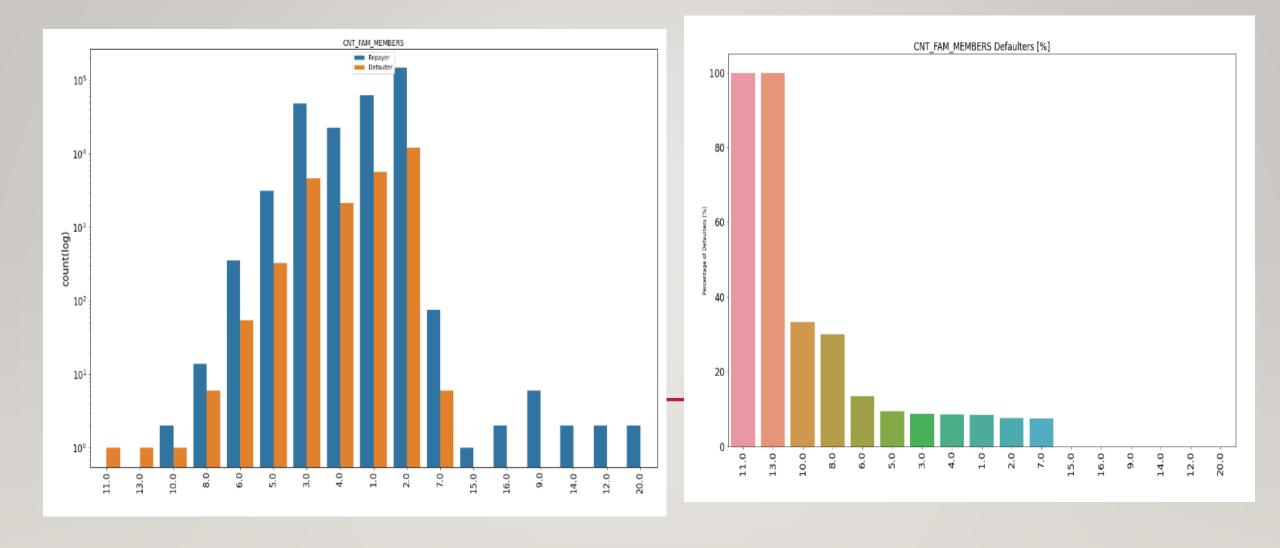
#### FLAG\_DOCUMENT\_3

WE SEE, REPAYERS AND DEFAULTERS IN TERMS OF SUBMITTING DOCUMENT 3 DOES NOT SHOW ANY CORRELATION, THOUGH THE CUSTOMERS WHO HAVE SUBMITTED THE DOCUMENTS ALSO SEEM AS A SLIGHT DEFALUTER; THEN WHO HAVE NOT SUBMITTED THE DOCUMENTS.



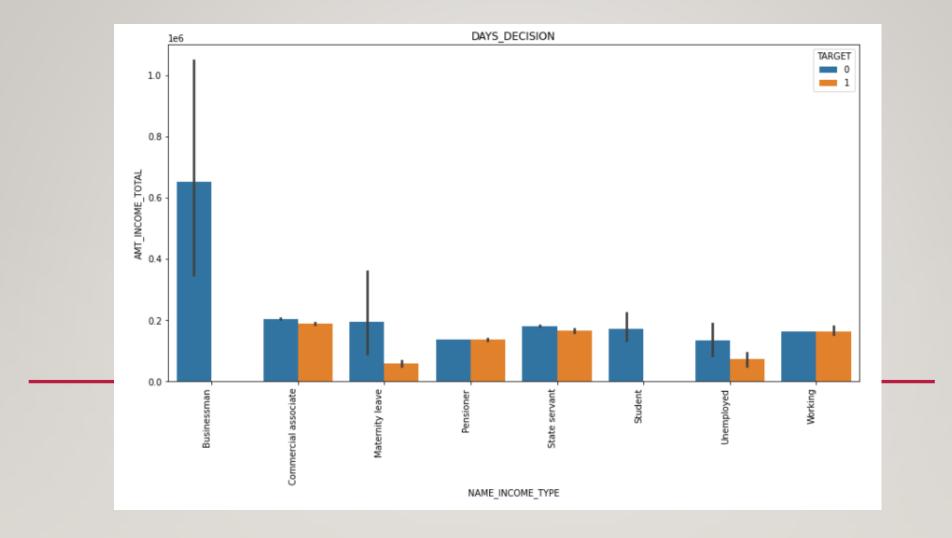
## NO. OF FAMILY MEMBER OF CLIENT DISTRIBUTION



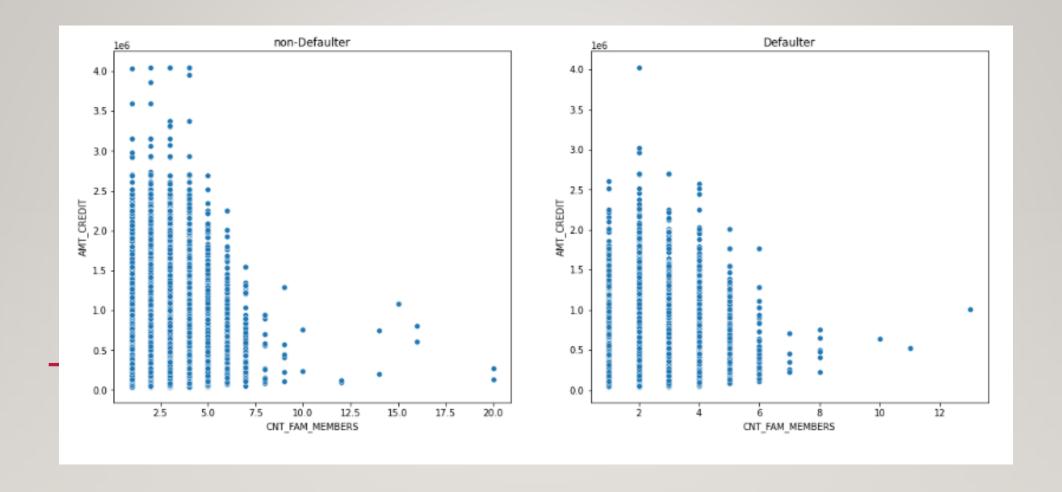


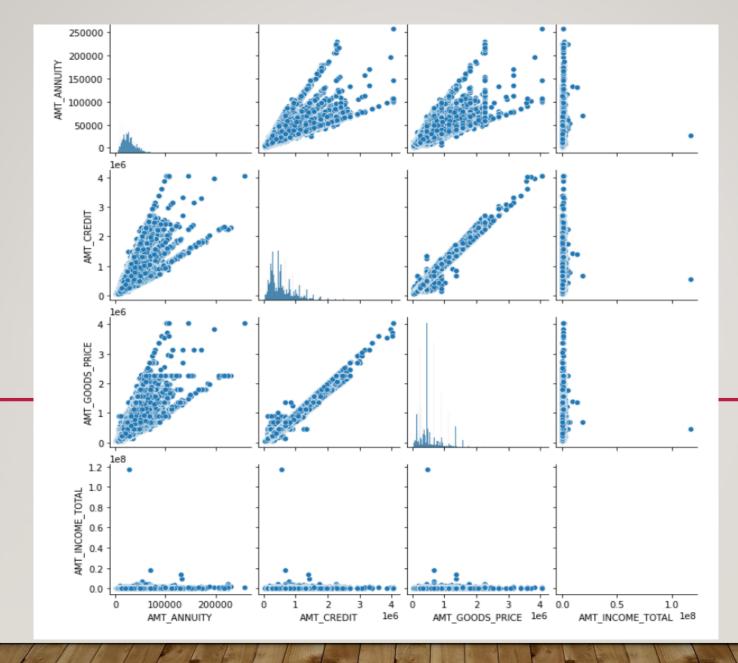
Here, family members increases the risk of being defaulter as they follow the same trend of children having more no of people in family.

## BI/MULTIVARIATE CATEGORICAL ANALYSIS

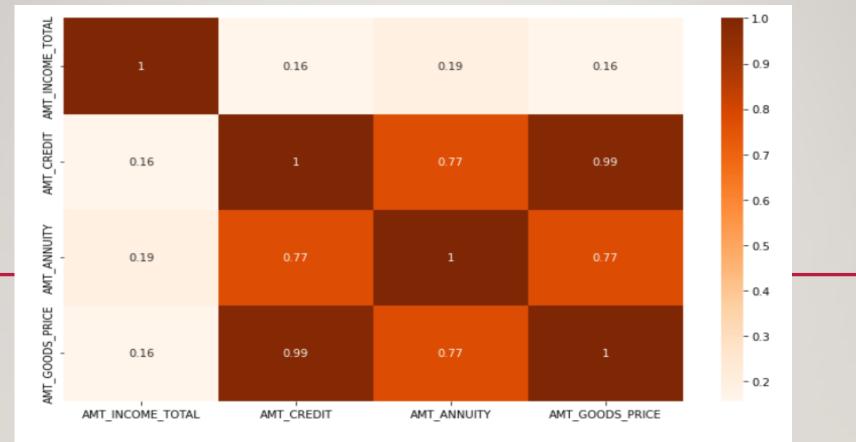


Analysis CNT\_FAM\_MEMBERS and AMT\_CREDIT Its effect on loan repayment status

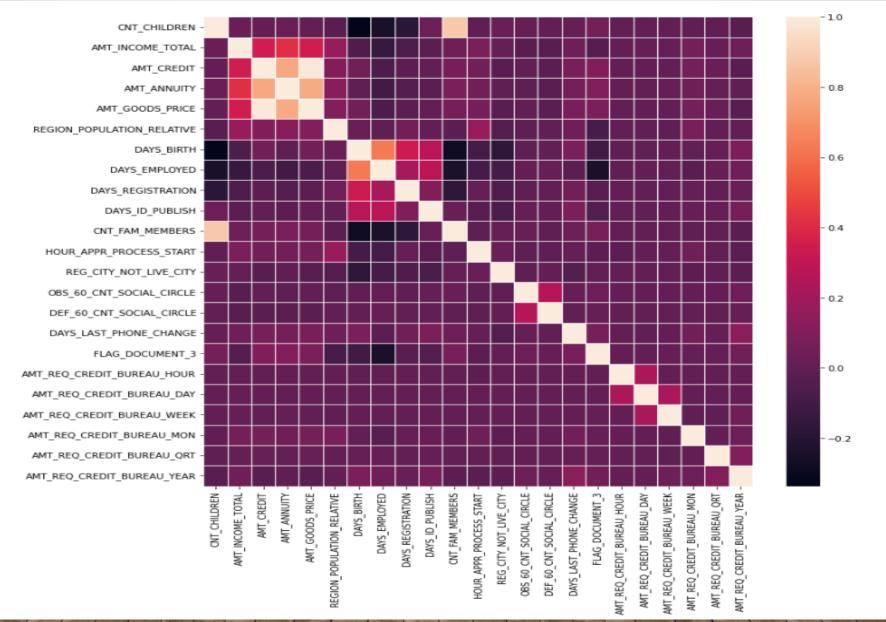




#### **CORRELATION BETWEEN VARIABLES**

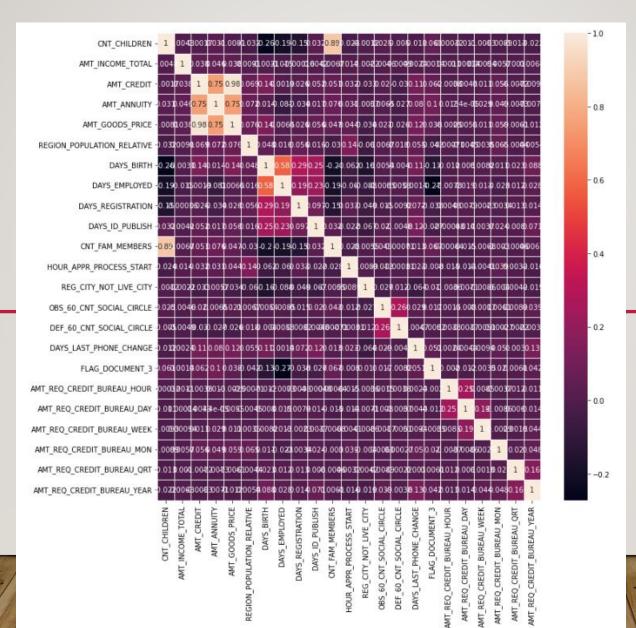


#### **NUMERIC VARIABLES ANALYSIS**



Correlating factors of Credit amount is highly correlated with amount of goods price, loan annuity, total income. Repayers show high coorelation in ho. of days employed.

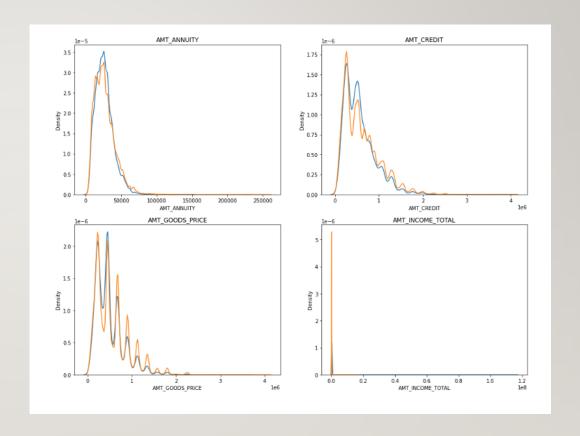
#### FETCHING THE CORRELATION OF TOP 10 FOR THE DEFAULTERS DATA:



- HIGH CORRELATION BETWEEN CREDIT AMOUNT AND AMOUNT OF GOODS PRICE WHICH SEEMS SAME AS REPAYERS.
- LOAN ANNUITY CORRELATION WITH CREDIT AMOUNT HAS SLIGHT REDUCTION IN DEFAULTERS(0.75) WHILE COMPARING TO REPAYERS(0.77)
- REPAYERS HAVE HIGH CORRELATION IN NUMBER OF DAYS EMPLOYED(0.62) WHILE COMPARING TO DEFAULTERS(0.58).
- THERE IS A SEVERE DROP IN THE CORRELATION BETWEEN TOTAL INCOME OF THE CLIENT AND THE CREDIT AMOUNT(0.038) AMONGST DEFAULTERS, WHEREAS IT IS 0.342 AMONG REPAYERS.
- DAYS\_BIRTH AND NUMBER OF CHILDREN CORRELATION HAS REDUCED TO 0.259 IN DEFAULTERS
  WHEN COMPARED TO 0.337 IN REPAYERS.
- THERE IS A SLIGHT INCREASE IN DEFAULTED TO OBSERVED COUNT IN SOCIAL CIRCLE AMONG DEFAULTERS(0.264) WHEN COMPARED TO R REPAYERS(0.254)

## NUMERICAL UNIVARIATE ANALYSIS

WE SEE CREDIT LOAN IS MOSTLY LESS
 THEN 10 LACS, MOST OF THE PEOPLE
 PAY ANNUITY BELOW 50000 FOR THE
 CREDIT LOAN, MOST NO OF THE LOANS
 ARE GIVEN FOR GOODS PRICE ARE LESS



#### **RE-PAYER APPLICANT: I.E APPLICATION MAY BE APPROVED**

- NAME\_EDUCATION\_TYPE: Academic degree has less defaults.
- NAME\_INCOME\_TYPE: Student and Businessmen have no defaults.
- REGION\_RATING\_CLIENT: applicants living in rating I are safe borrowers.
- CNT\_CHILDREN: APPLICANTS with zero to two children tend to repay to loans
- ORGANIZATION\_TYPE: applicants with trade type 4 and 5 and industry type 8 have defaulted less than 3%.

# DEFAULTER APPLICANTS; I.E APPLICATION CAN BE REJECTED TO PREVENT BUSINESS LOSS TO COMPANY

- NAME\_EDUCATION\_TYPE: appliacnts with lower secondary, secondary education and incomplete higher education have higher default rate.
- CNT\_CHILDREN: Applicants who have more then 6 children have a very high default rate and hence higher interest should be imposed on their loans.
- AMT\_CREDIT: Applicants getting loan for 300-600k tend to default more than others, having higher interest specifically for this credit range would be better.
- CNT\_FAM\_MENBERS: applicant who have more than 11 family members higher default rate for application rejection.
- OCCUPTION\_TYPE: applicants who have low-skill laborer, driver and barmen staff default rate
  is huge.
- REGION\_RATING\_CLIENT: applicants who live in areas with region rating as 3 has highest defaults.
- ORGANIZATION\_TYPE: transport type 3 , industry type 13 people have high defaulter rate and should be avoied to be approved for loan.

#### CONCLUSION

 The analysis of the data set shows there are some allocate criteria of the clients from which bank can measure the risk analytics of the defaulters who would be abel to repay the loan or be considered as defaulters.