CREDIT EDA CASE STUDY

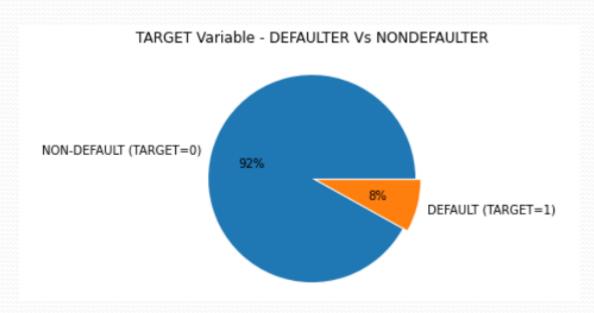Presented by

Surbhi Mathur

There are two type of risk:

- 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.

Steps:

- Start by importing the 'application_data.csv'.
- Check the structure of the data(Normal routine check)
- Data Quality Check and Missing values
- Analysis
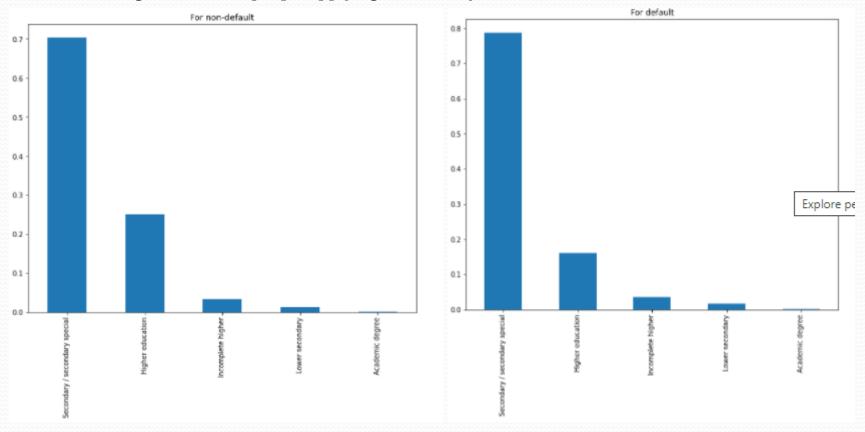- Read "Previous Application" data.
- Final words
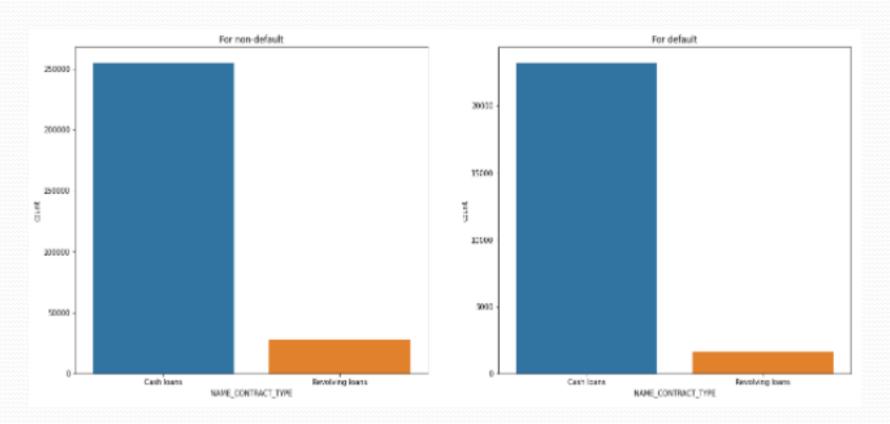
Imbalance in Target

From the total
8% are defaulters

TARGET Variable - DEFAULTER Vs NONDEFAULTER

NON-DEFAULT (TARGET=0) 92%

8% DEFAULT (TARGET=1)

EDUCATION

- From the plot Above, we can conclude that in both case secondary/special educated people are applying loans in high in number.
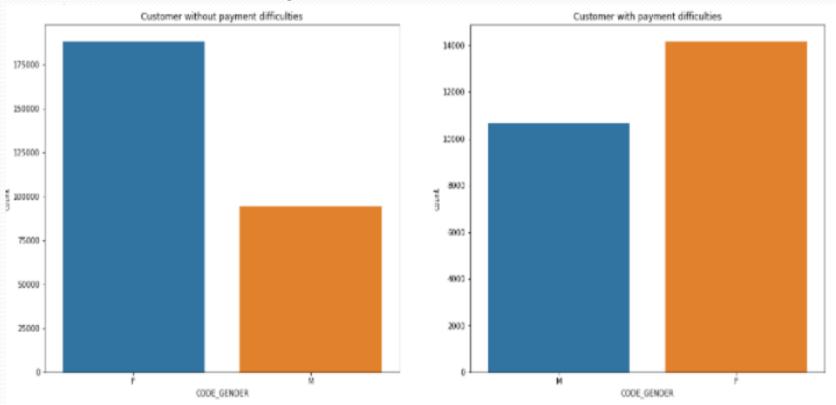- Academic degree educated people applying loan in very less number.

CONTRACT PRODUCT TYPE

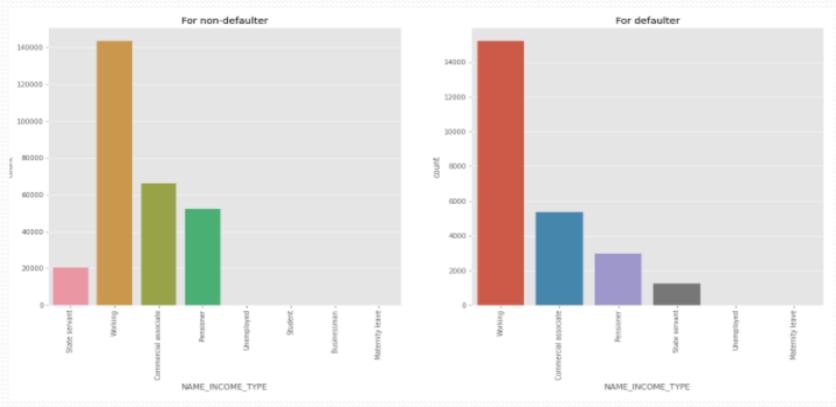- Cash loans are preferred by both Loan Payment Difficulties and Loan-Non Payment Difficulties

GENDER

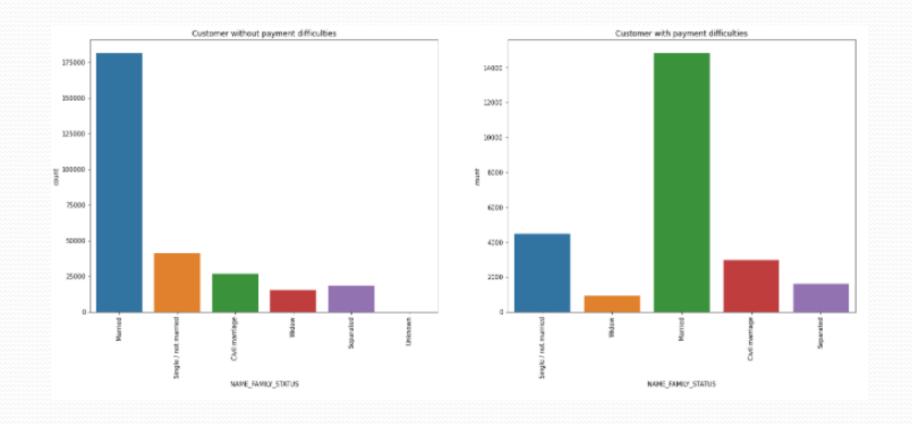- An increase in the percentage in Male Payment Difficulties from Non-defaulter to

INCOME SOURCE

- Decrease the payment difficulties who are Pensioner & increase the percentage of payment difficulties who are working,
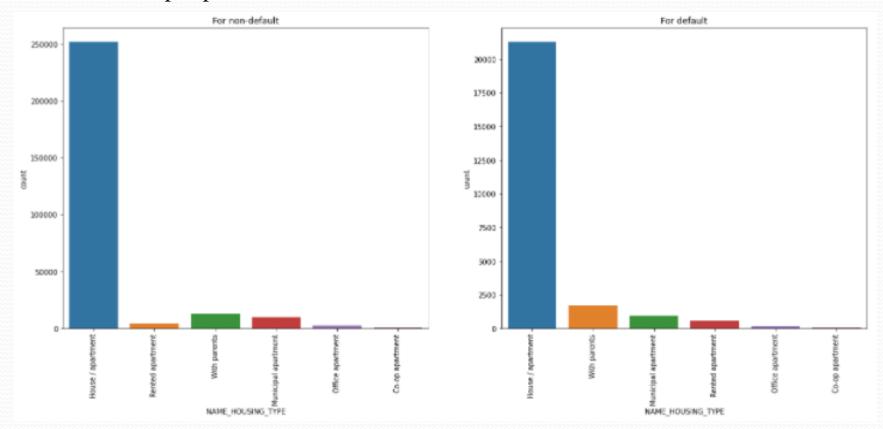
FAMILY STATUS

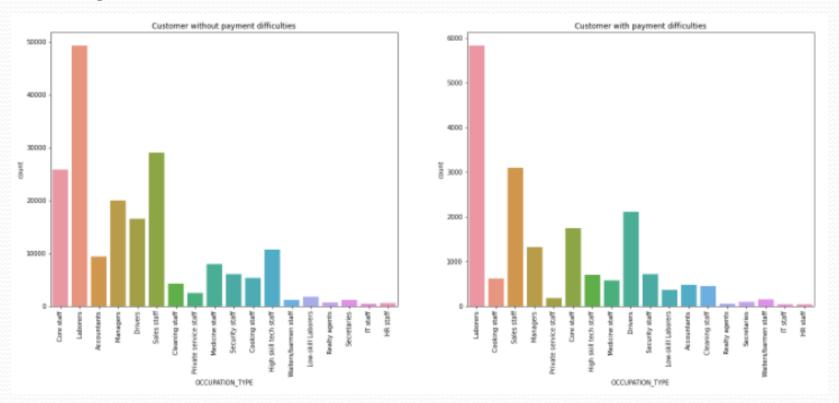- The married people tend to take more Loan as compared to other categories.

HOUSING

- There exists people who have own house
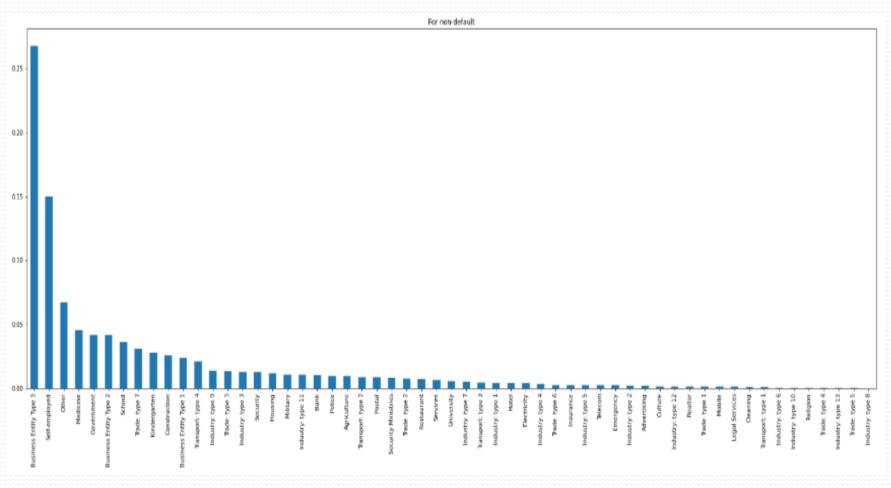
OCCUPATION TYPE

- Labourers having more difficulties to paying loan and IT staff having least difficulties to paying the loan, compare both case
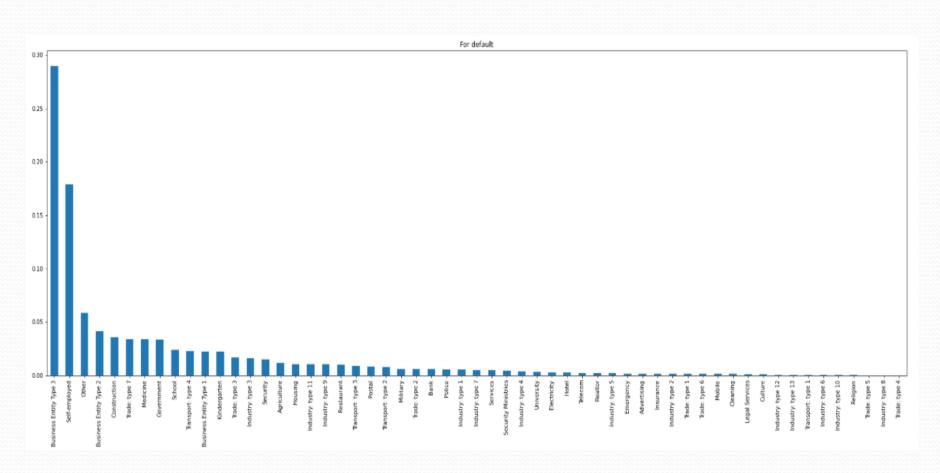
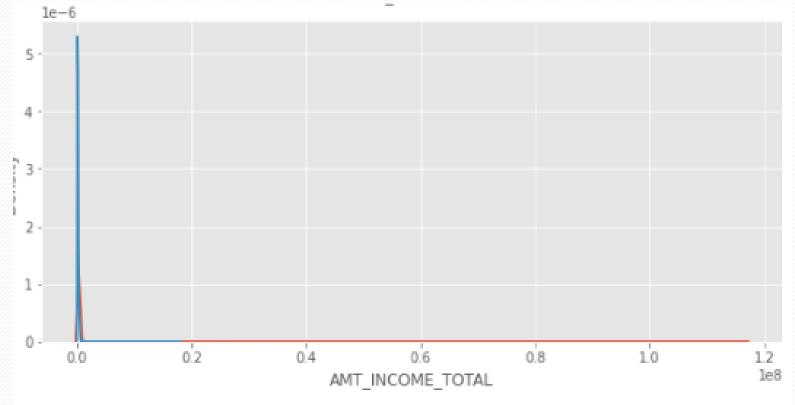# ORGNIGATION TYPE

Non Defaulter

Defaulter

For default

Inferences on Organization type

- Clients which have applied for credits are from most of the organization type 'Business entity

- In both case

- Less clients are from Industry type 8,type 6, type 10, religion and trade type 5, type 4.
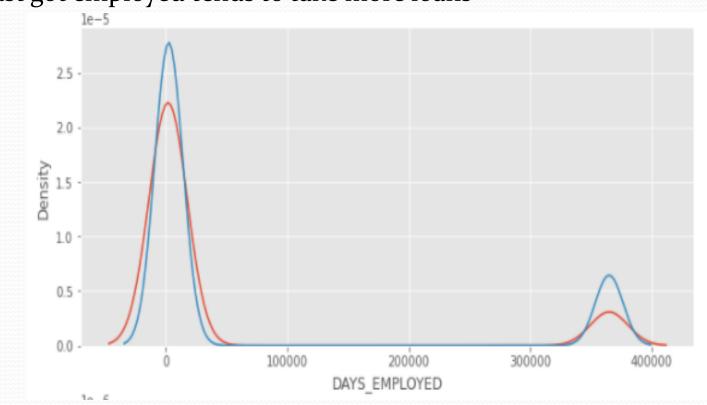
CONTINUOUS VARIABLE

AMOUNT OF INCOME

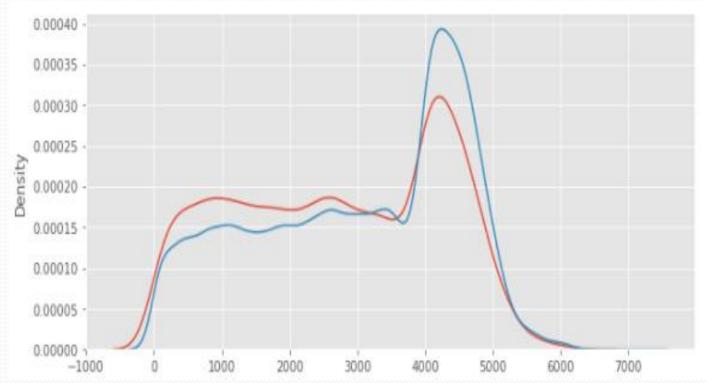- People with lower total income are more likely to default

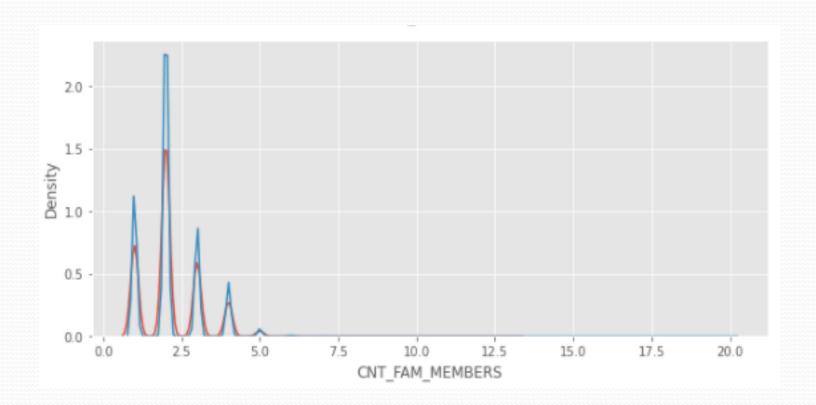# DAY OF EMPLOYEMENT

- Just got employed tends to take more loans

DAYS OF ID PUBLISHED

- Id(s) got published between 4000 days and 5000 days ago tend to take more loans

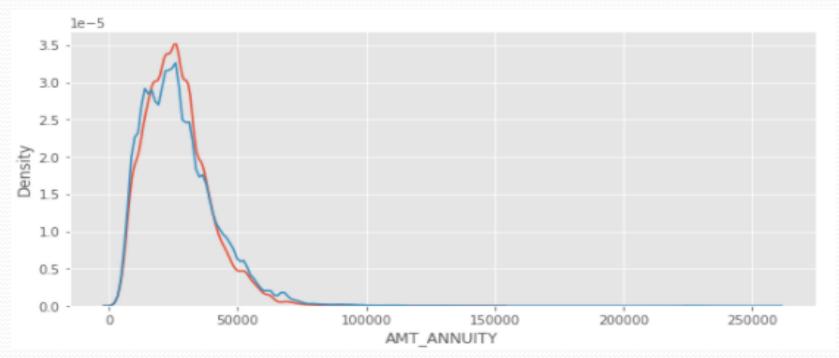# NUMBER OF FAMILY MEMBERS

- Nuclear family tends to take more loans

LOAN ANNUITY

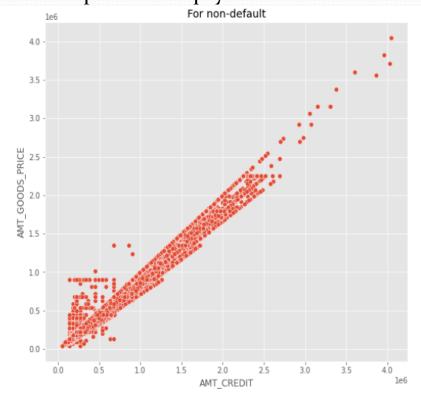- low amount annuity has high number of loans
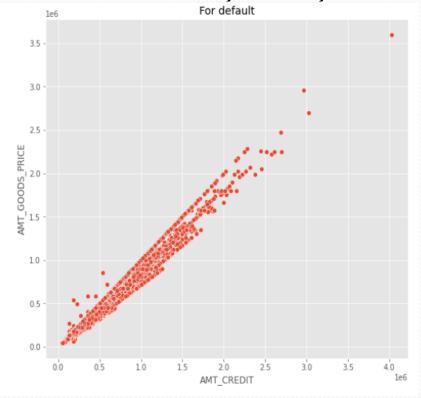
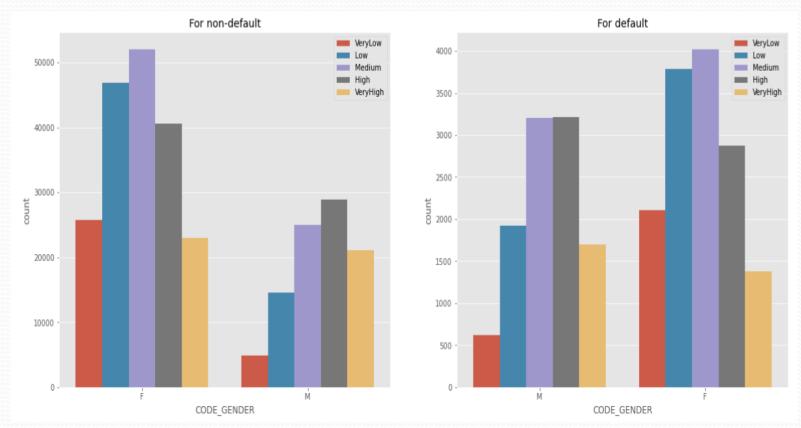# BIVARIATE ANALYSIS

CREDIT AMOUNT WITH RESPECT ANNUALTY

- People without payment difficulties take more credit for the annuity that they have

AMOUNT INCOME RANGE WITH RESPECT GENDER

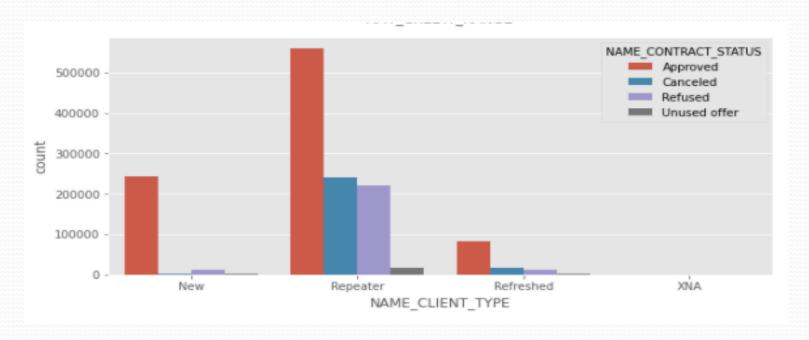- Female counts are higher than male.
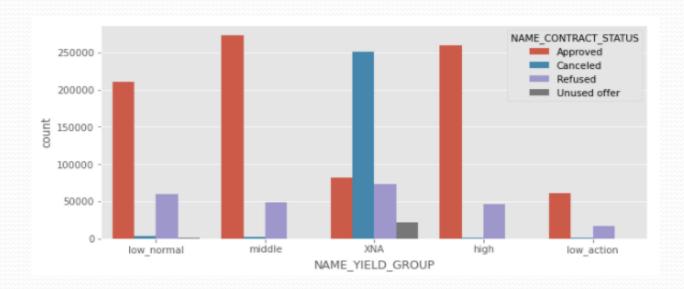
MERGE
CURRENT APPLICATION DATA
&
PREVIOUS APPLICATION DATA

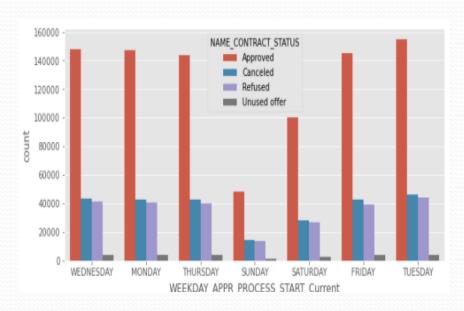## TYPE OF CLIENT

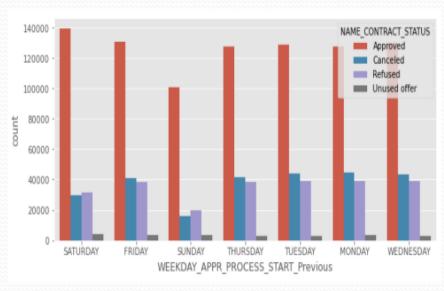- Repeater has highest number of approved loans.

GROUPED INTEREST RATE

- Middle NAME_YIELD_GROUP has highest approval.
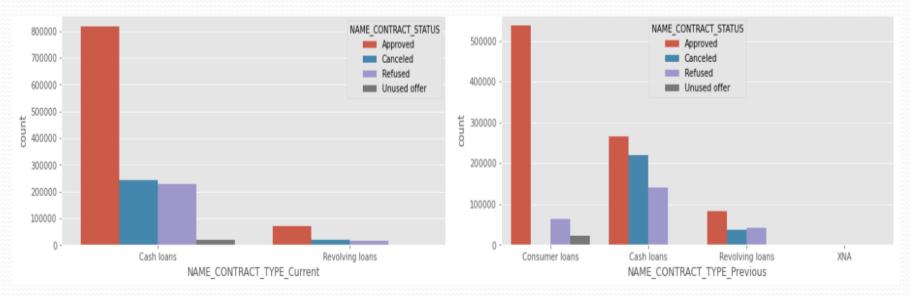
DAY OF THE WEEK CLIENT APPLY LOAN

- In previous application Saturday has the highest approval rate.
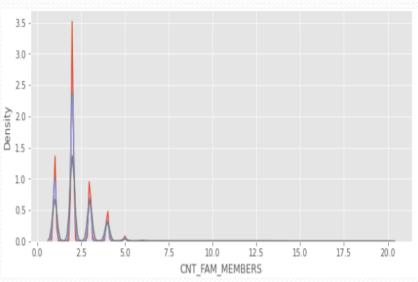 but in current application it is Tuesday.
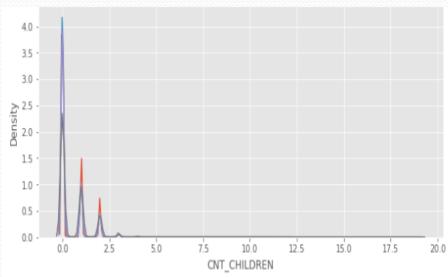
CONTRACT PRODUCT TYPE

- Previously bank was providing Cash, Revolving and Consumer loans.
- Number of consumer loans were highest previously and now highest number is Cash loans.

NUMBER OF CHILDREN & FAMILY MEMBER

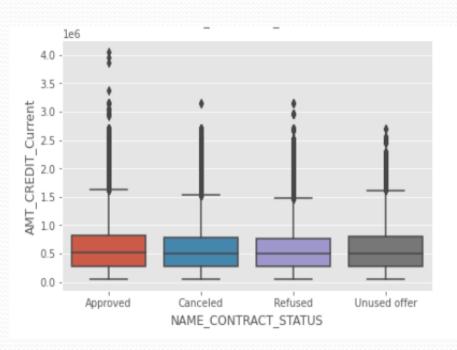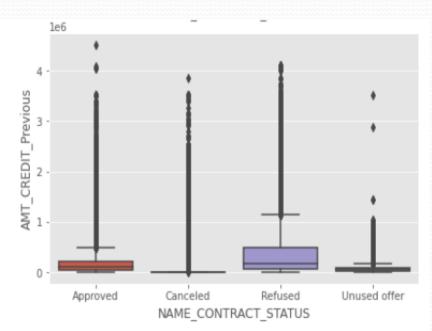- Nuclear family tends to take more loans
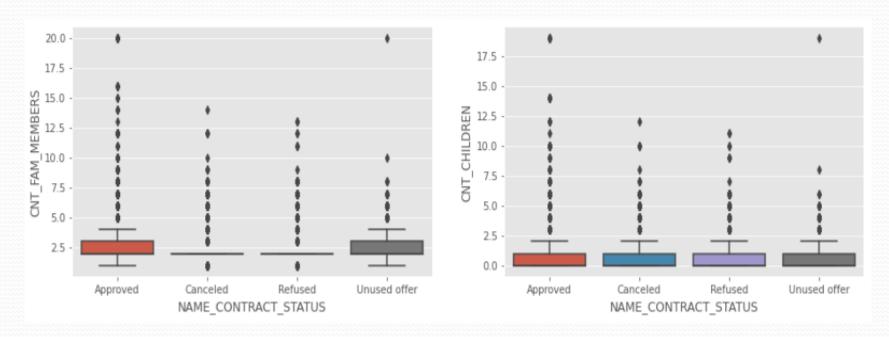
CREDIT AMOUNT WITH RESPECT TO CONTRACT STATUS

- AMT_CREDIT_Previous has highest refused cases and AMT_CREDIT_Current is similar for all 4 cases.

FAMILY MEMBER WITH RESPECT TO CONTRACT STATUS

- Nuclear family(2-3 people in family) get highest approval.

CONCLUSION

- The percentage of defaulters is 8%.
- The bank lends more to females.
- Banks should focus less on income type 'Working' as they are having most number of unsuccessful payments.
- Banks should focus more on contract type 'Student','pensioner' and 'Businessman' with housing 'type other than 'Co-op apartment' for successful payments.

THANK YOU