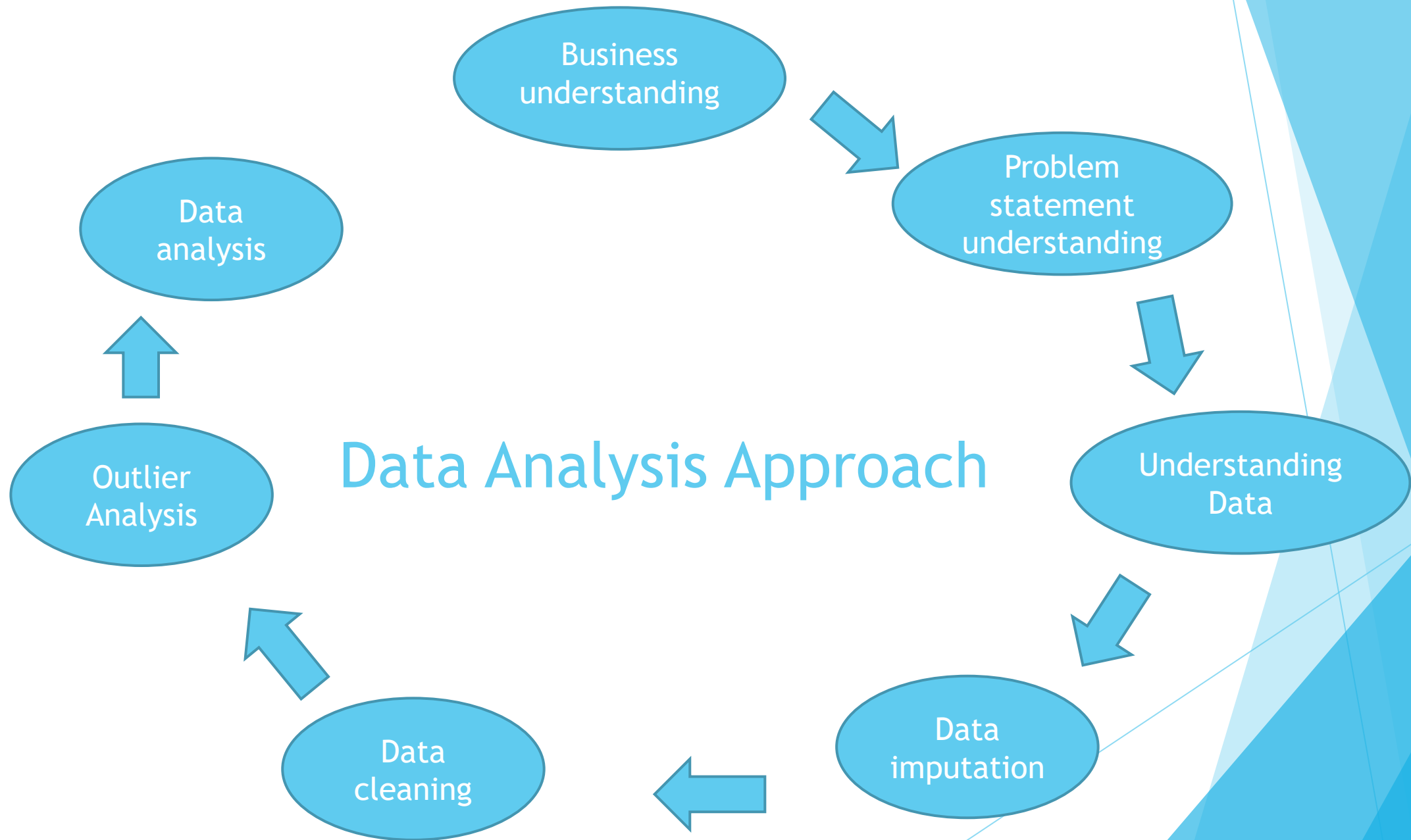


Credit EDA Assignment

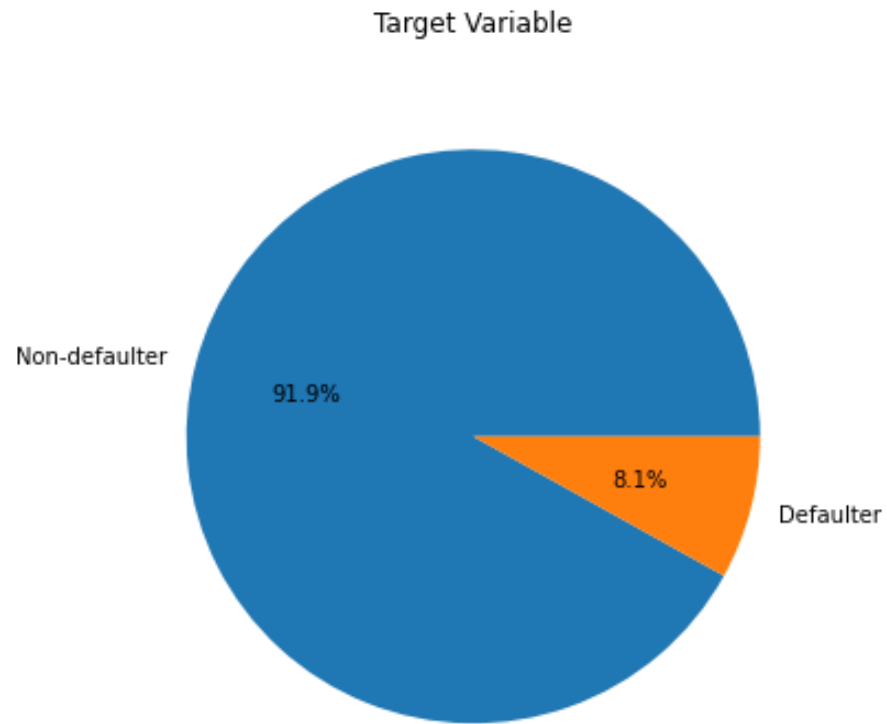
VANANEE RAVIKUMAR

Problem Statement

- ▶ Two type 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 company wants to understand the driving factors behind loan default i.e. the variables which are strong indicators of default.

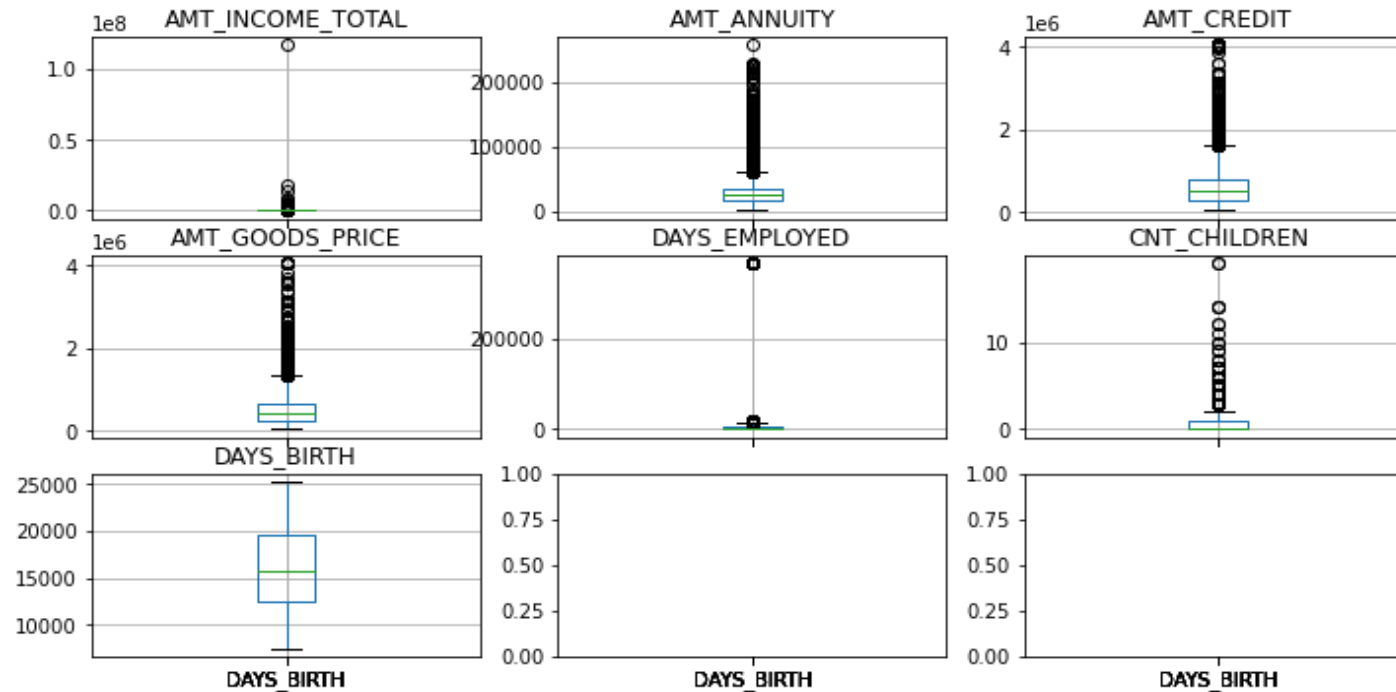


Data Imbalance



Ratio of imbalance for Non-defaulter and defaulter : 11.387150050352469 :1

Outlier Analysis



Results

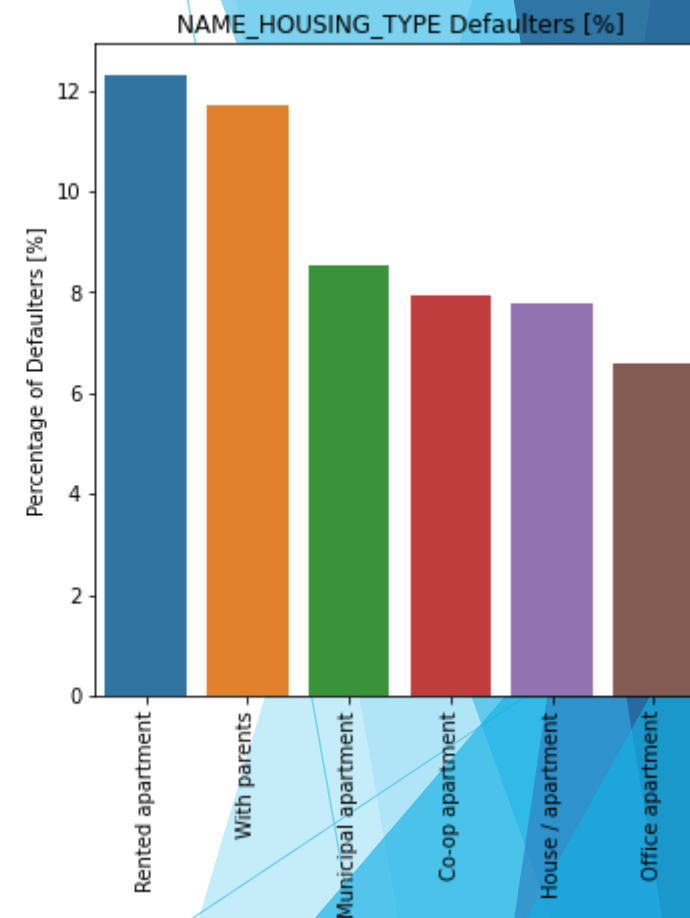
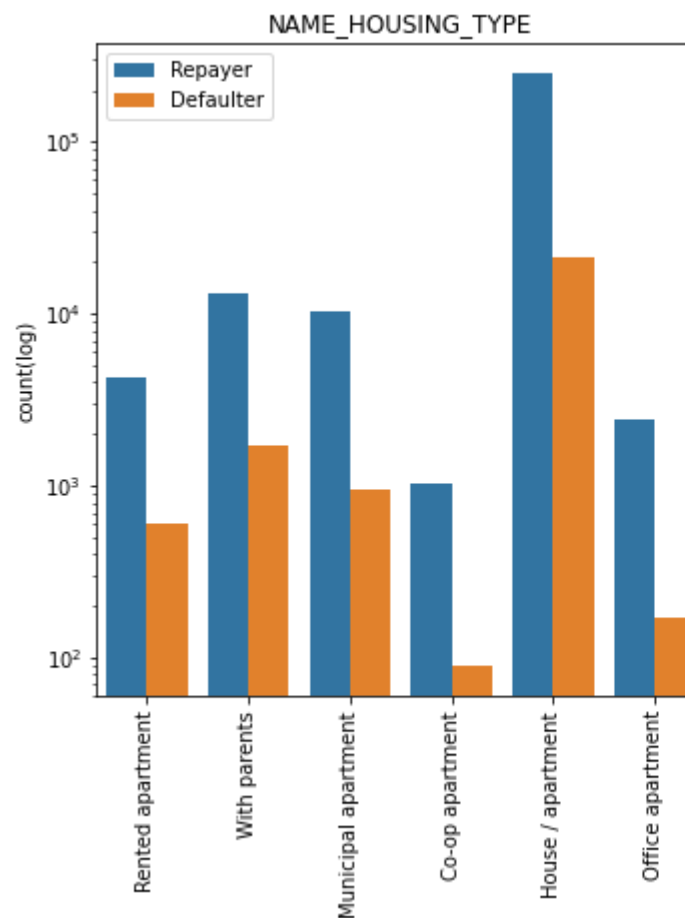
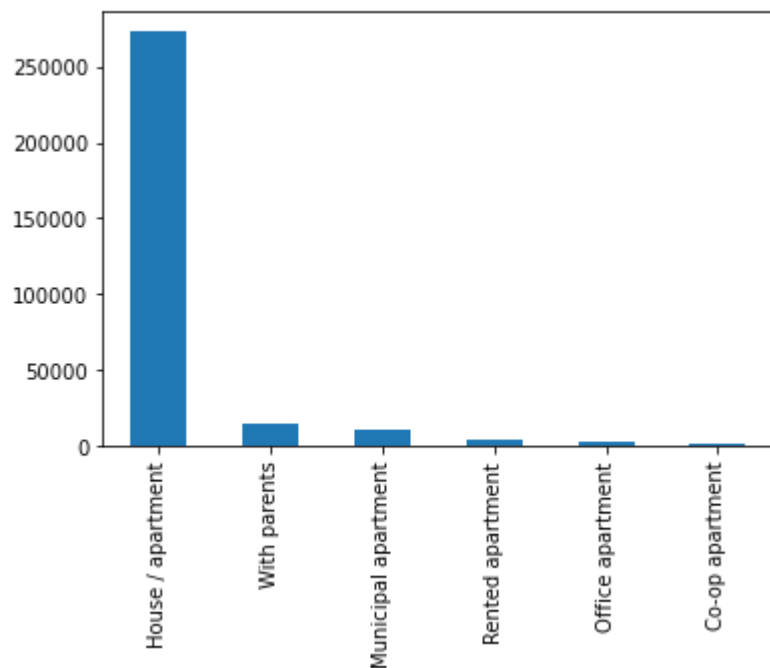
- AMT_ANNUITY, AMT_CREDIT, AMT_GOODS_PRICE ,CNT_CHILDREN have outliers.
- AMT_INCOME_TOTAL has more number of outlier which indicate some customer have high income compare to others.
- DAYS_BIRTH has no outlier.
- DAYS_EMPLOYED has outlier around 900 years which is impossible and hence this has to be incorrect entry.

Univariate



HOUSING TYPE

Distribution of NAME_HOUSING_TYPE Variable

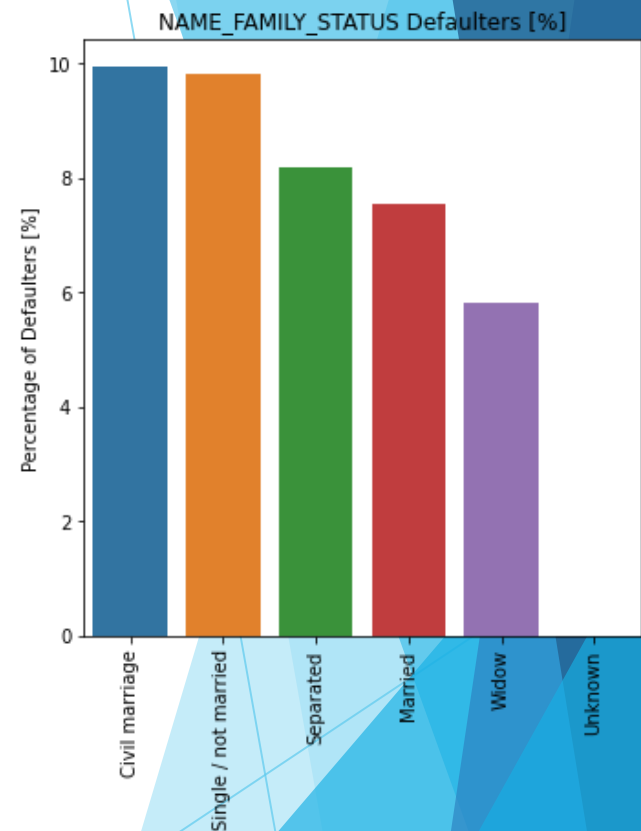
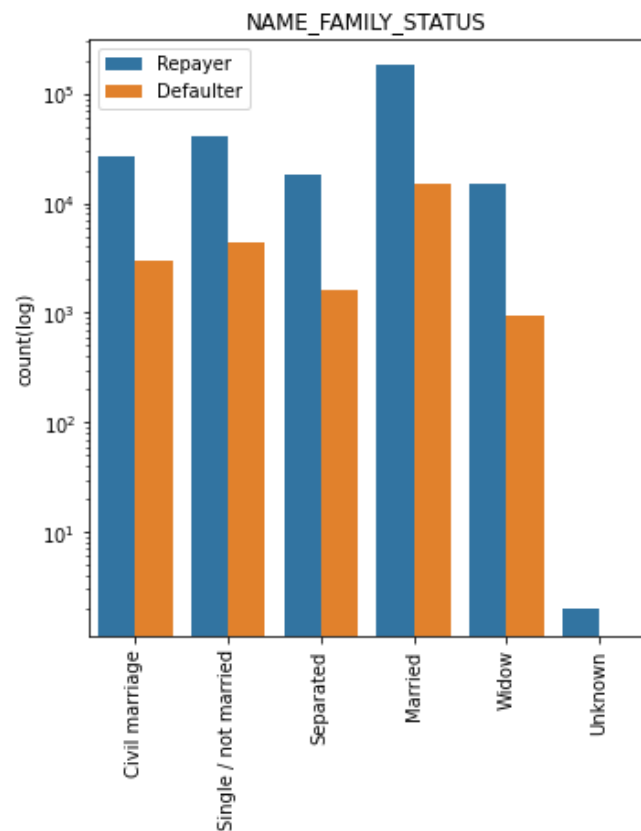
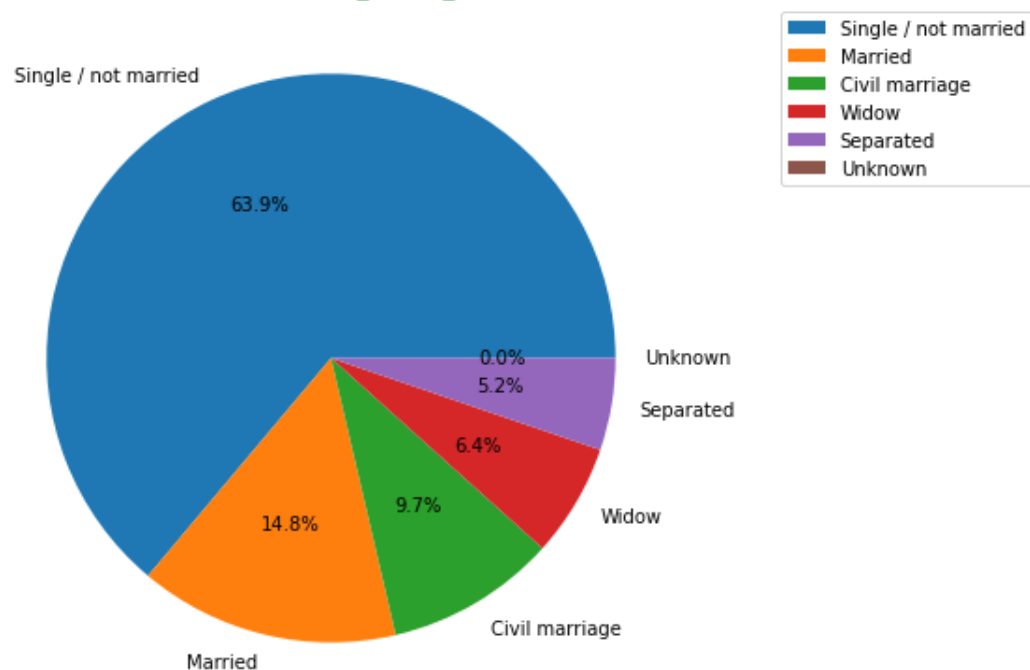


Inference:

- mostly people live in house/apartment
- office apartment have lowest default rate
- rented apartment and with parents have higher defaulting

FAMILY STATUS

Distribution of NAME_FAMILY_STATUS

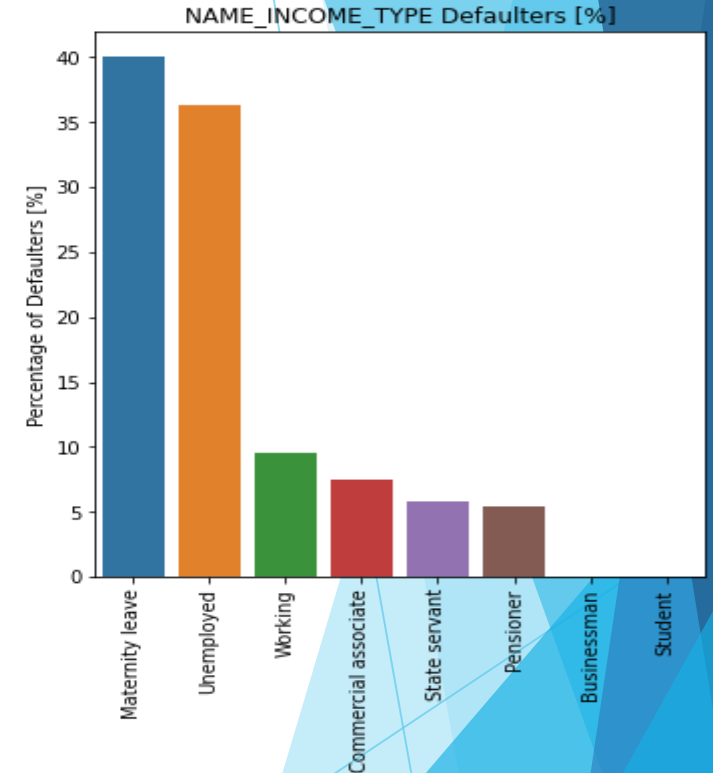
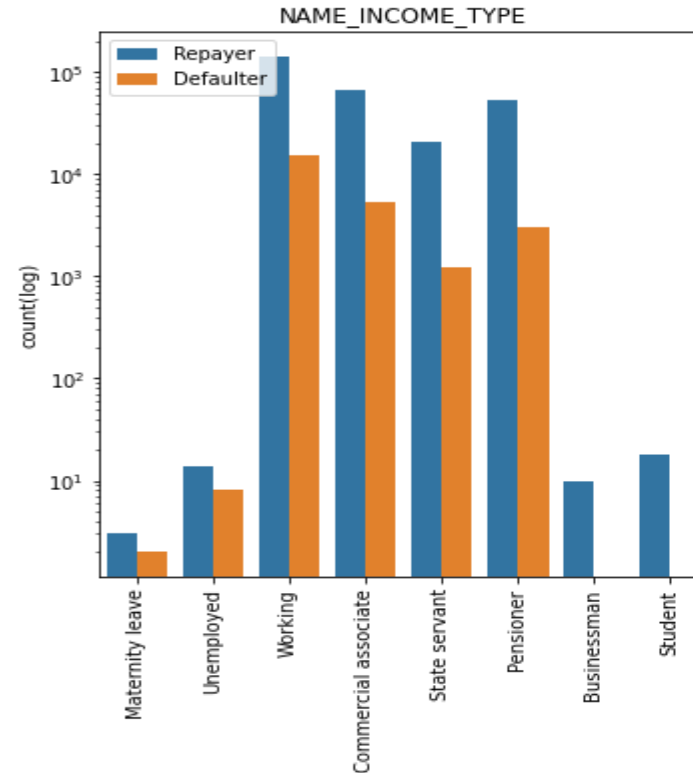
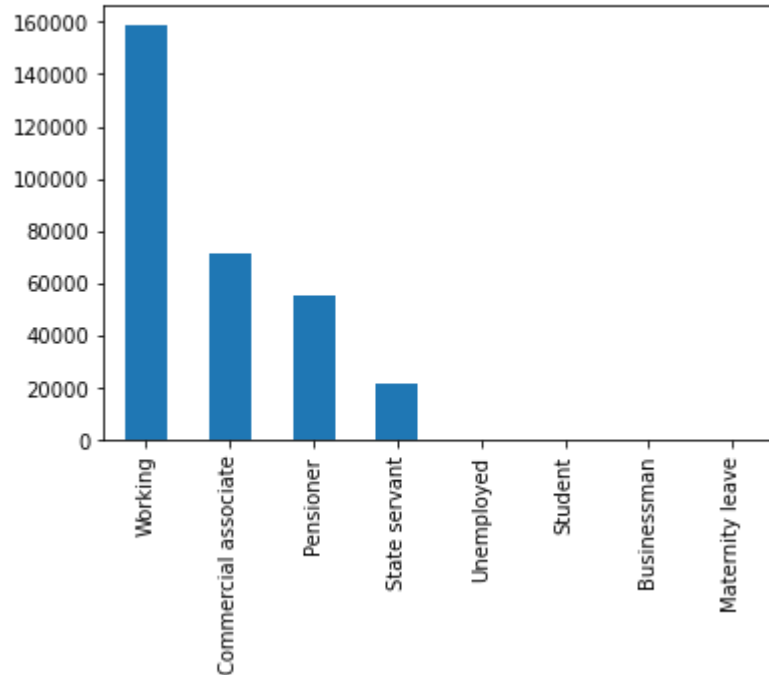


Inference:

- most of the people who have taken loan are married, followed by single/not married
- civil marriage has highest percentage of not repayment
- widow has lowest percentage of not repayment

INCOME TYPE

Distribution of NAME_INCOME_TYPE Variable

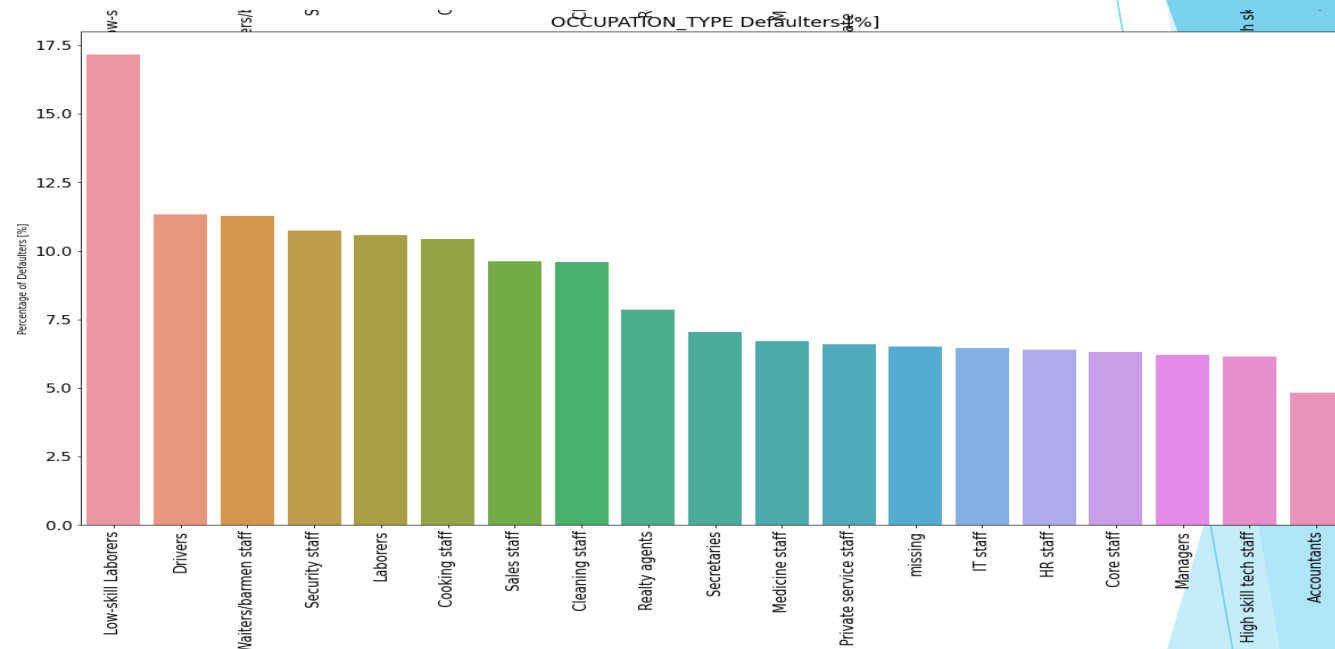
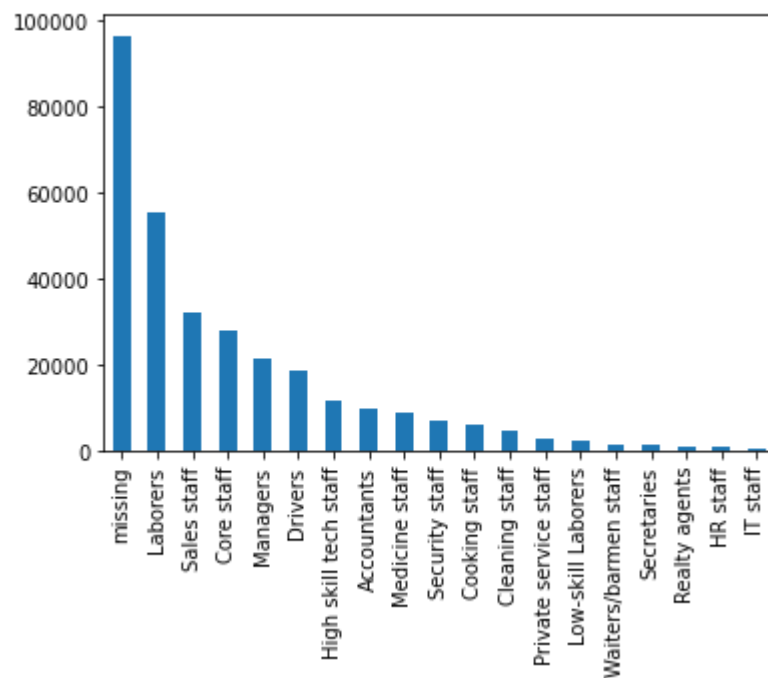


Inference:

- Most of clients for loans have income type as followed by commercial associate, Pensioner and state servant.
- The applicants with the type of income maternity leave have almost make 40% ratio of the defaulters.
- Student and bussiness , though less in numbers do not have any default record.

OCCUPATION TYPE

Distribution of OCCUPATION_TYPE Variable

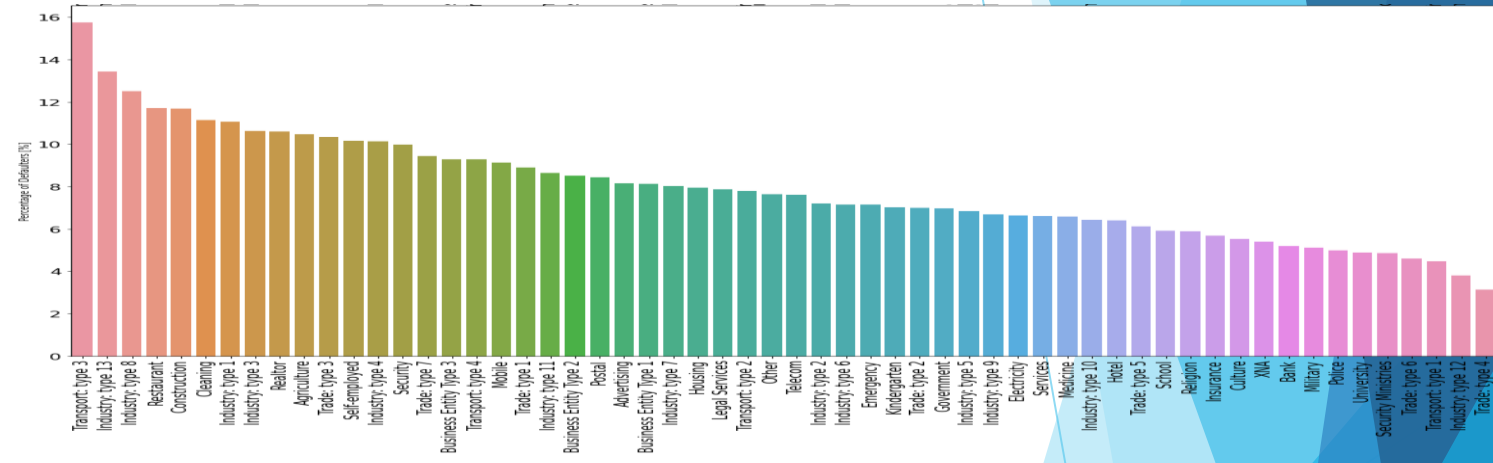
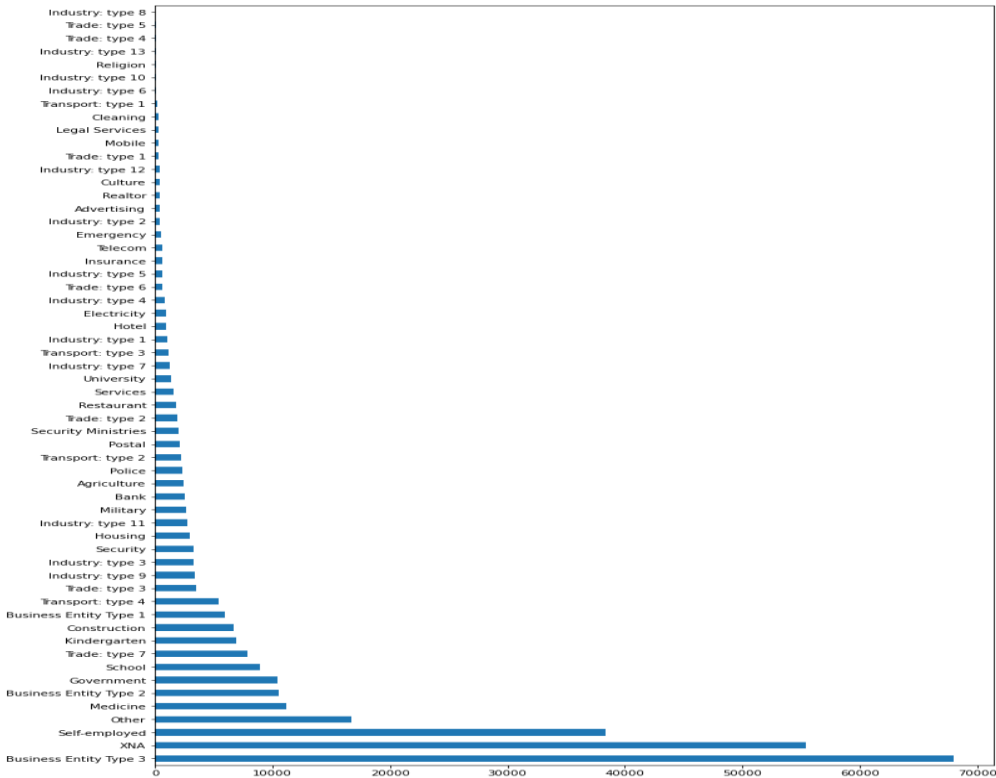


Inference:

- Most of the loans are taken by people whose occupation is 'missing'.
- IT staff take the lowest amount of loans.
- The highest percent of not repaid loans are from low-skill laborers.

ORGANIZATION TYPE

Distribution of ORGANIZATION_TYPE Variable

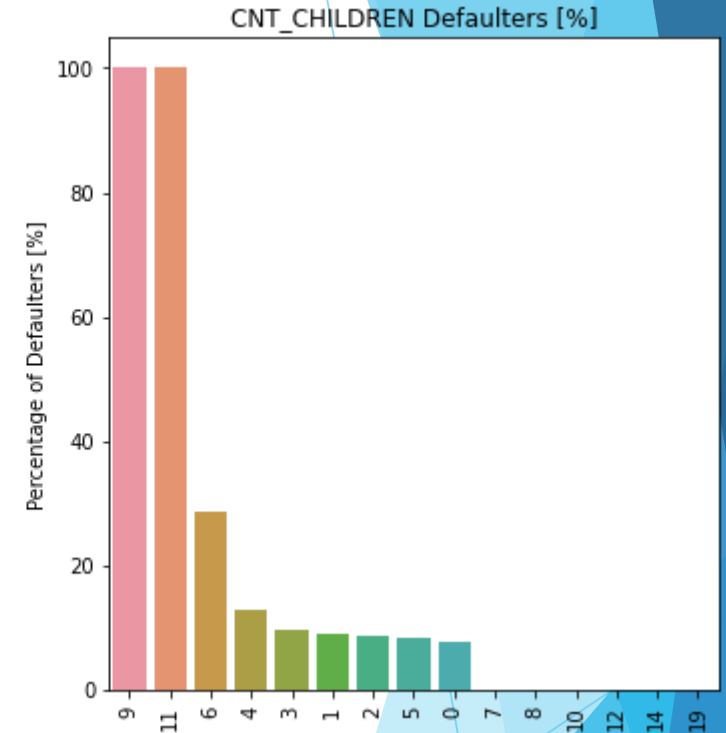
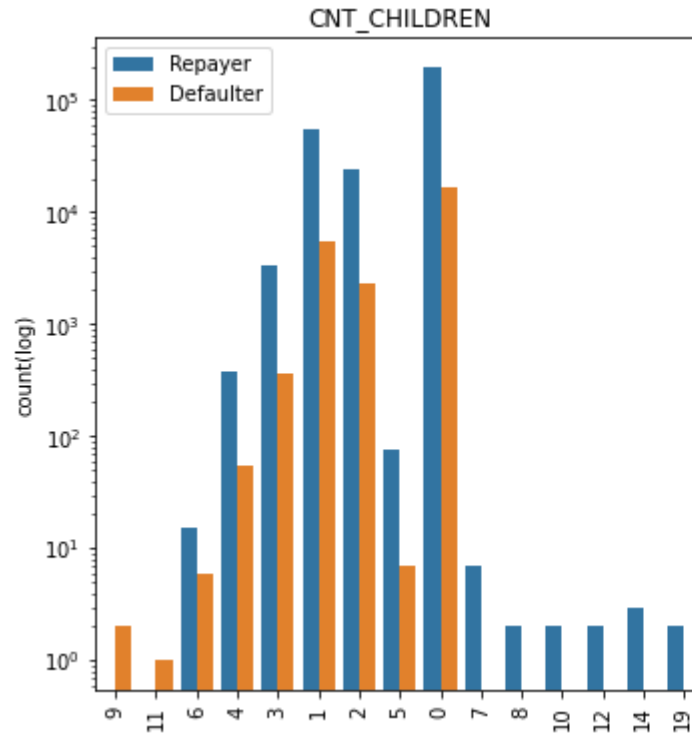
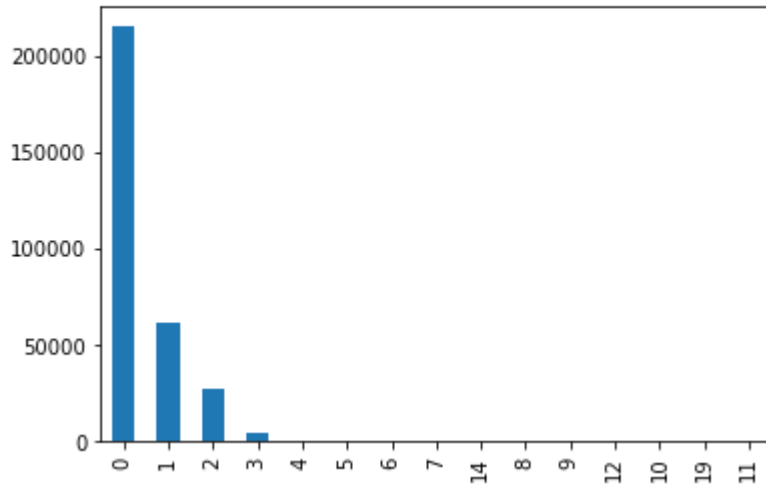


Inference:

- most of the clients for loan in bussiness Entity type 3
- Highest percent of loan not repaid are transport type 3
- For a very high number of clients is unavaiable(XNA)
- trade type 4 is lesser defaulters

Count of children

Distribution of CNT_CHILDREN Variable

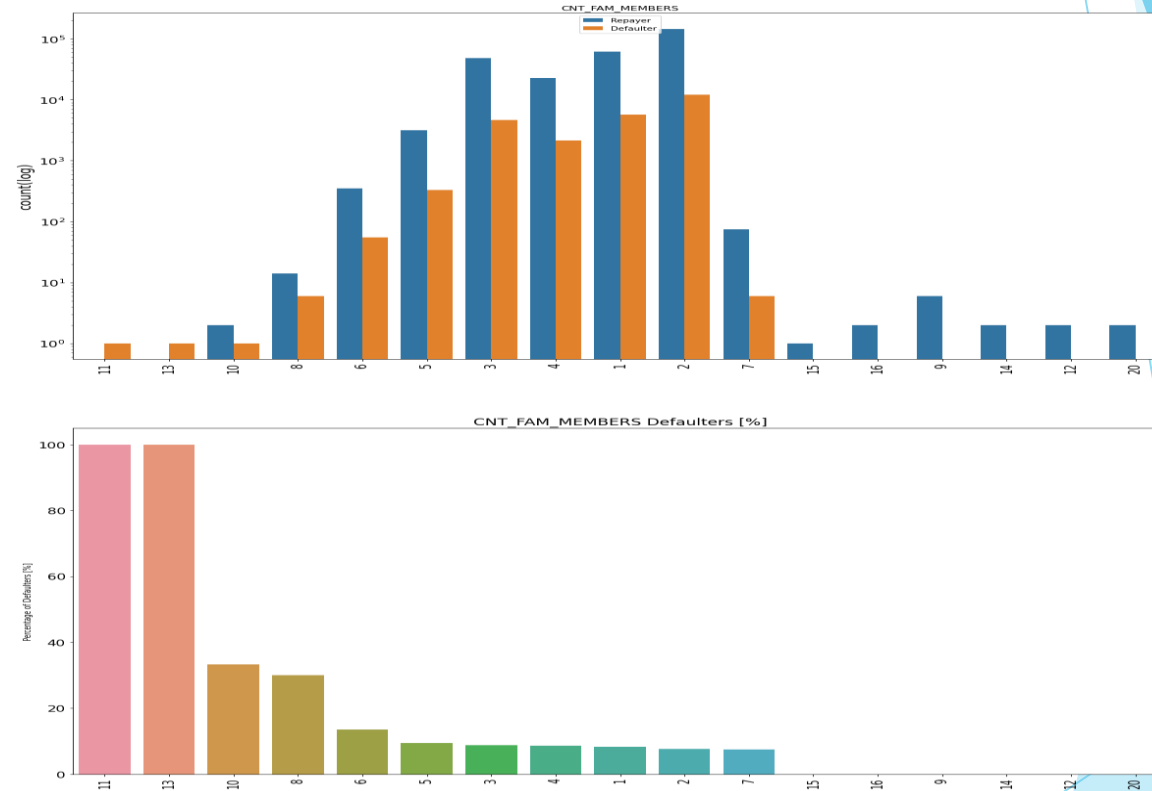
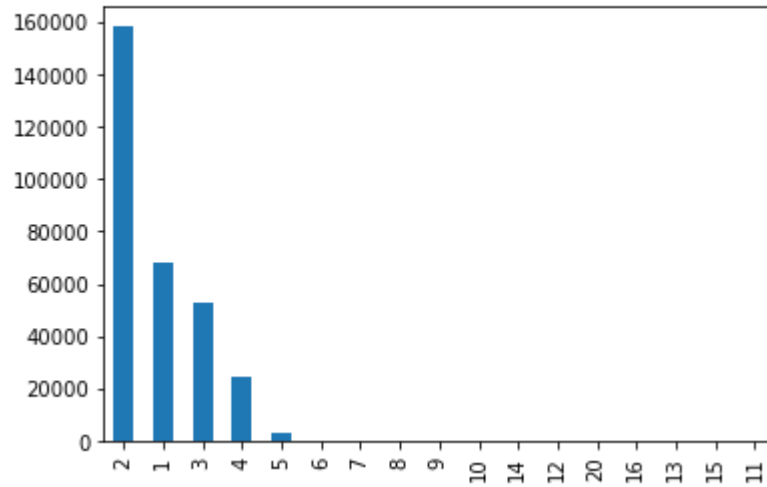


Inference:

- most of the clients do not have children
- very few client have more than 3 children
- client who have child count 9 and 11 showing 100% default rate

Count of Family members

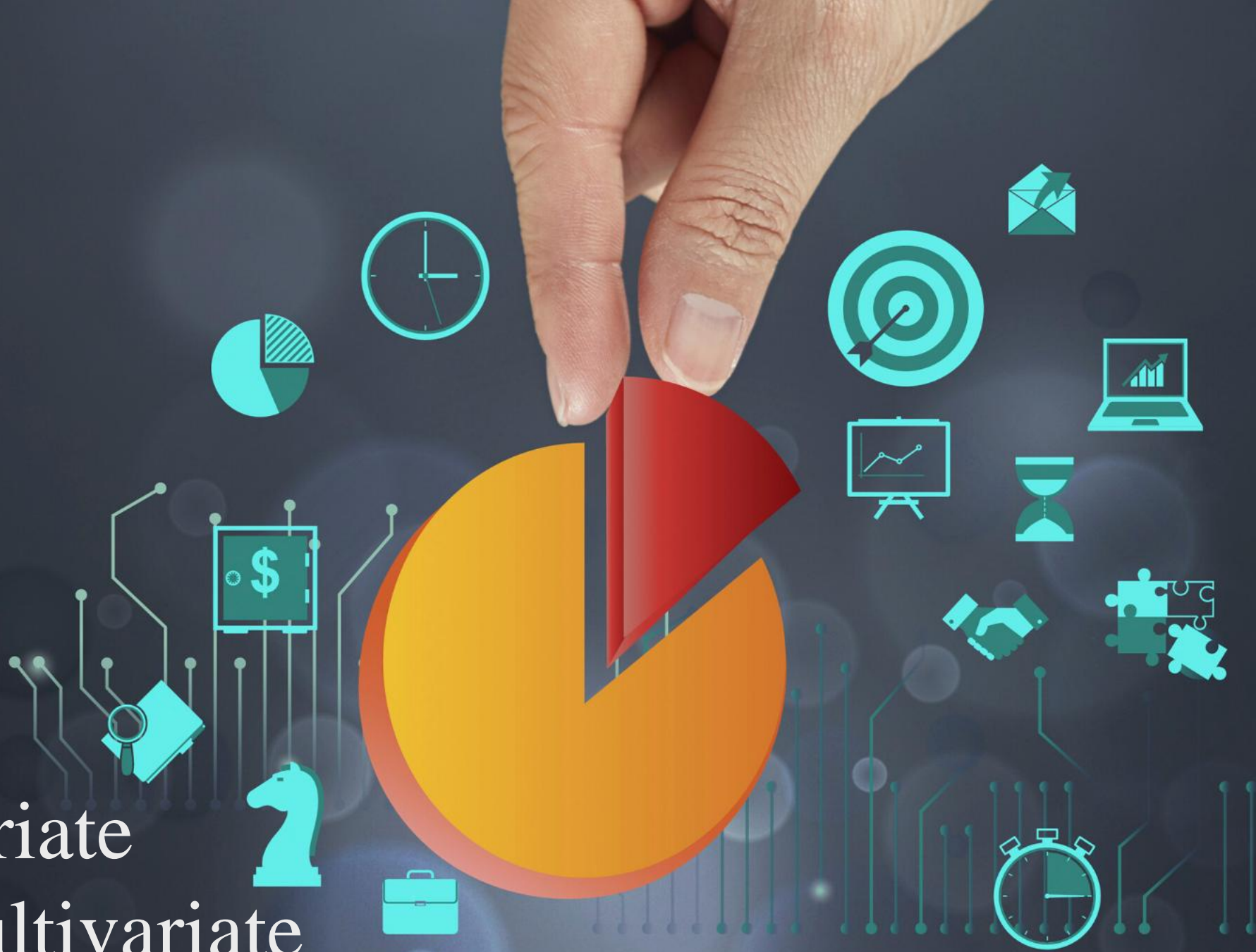
Distribution of CNT_FAM_MEMBERS Variable



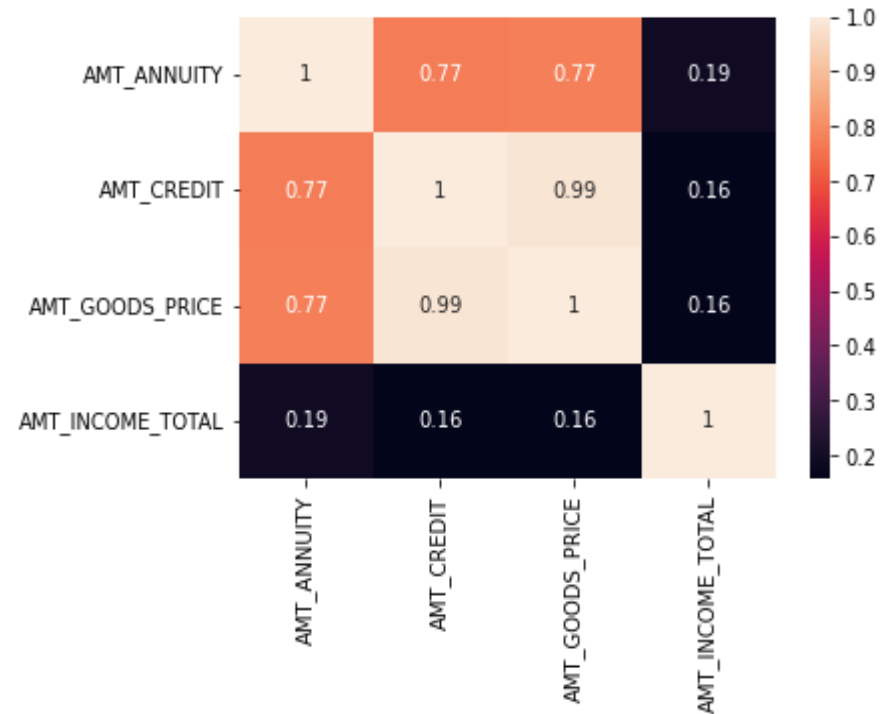
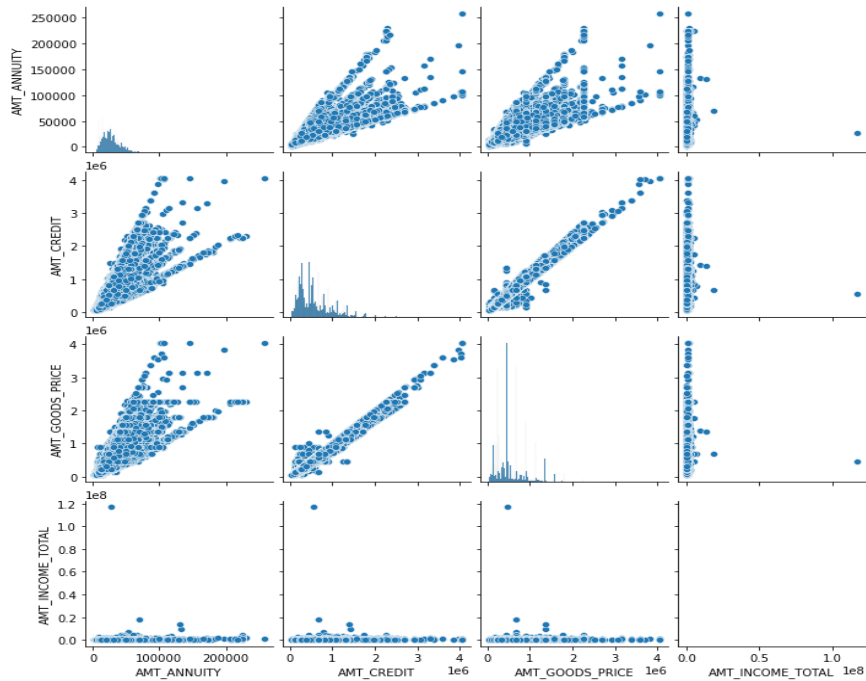
Inference:

- most of the clients do not have children
- very few client have more than 3 children
- client who have child count 9 and 11 showing 100% default rate

Bivariate & Multivariate



AMT_ANNUITY, AMT_CREDIT, AMT_GOODS_PRICE, AMT_INCOME_TOTAL



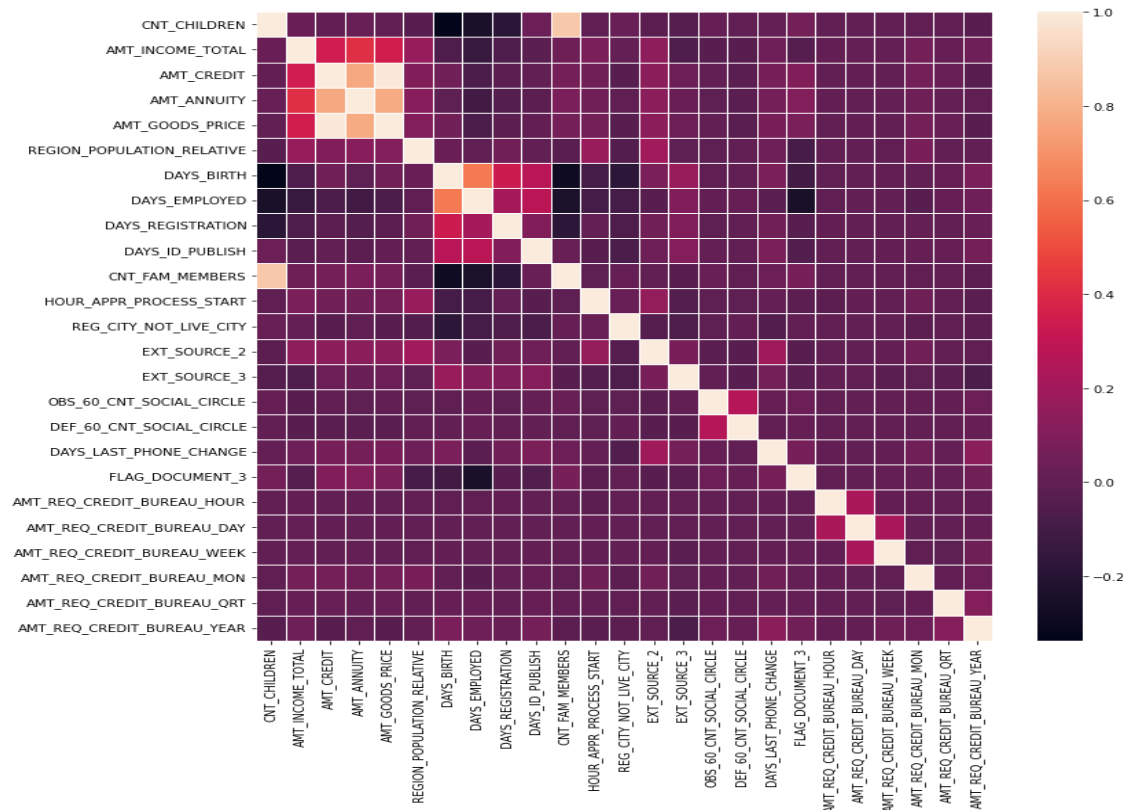
Inference

- high correlation between AMT_CREDIT and AMT_GOODS_PRICE
- clients owning goods of high value can take loans higher amounts

Multivariate Analysis (Numerical column)

Inference:

- Credit amount is highly correlated with amount of goods price is same as repayers
- But the loan annuity correlation with credit amount has slightly reduced in defaulters when compared to repayers
- these is a severe drop in correlation between total income of the client and credit amount
- days_birth and number of children correlation has reduced in defaulters when compared to repayers
- there is a slight increase in defaulted to observed count in social among defaulters when compared to repayers



Summary



► **Applicant to be non-defaulter, hence application can be approved**

- CNT_CHILDREN: APPLICANTS with zero to two children tend to repay to loans
- NAME_EDUCATION_TYPE : Academic degree has less defaults.
- NAME_INCOME_TYPE: students and business men have no defaults.
- ORGANIZATION_TYPE: applicants with trade type 4 and 5 and industry type 8 have defaulted less than 3%.
- AMT_INCOME_TOTAL: applicants with more than 700000 are less likely to default
- REGION_RATING_CLIENT: applicants who live in rating 1 are safe borrowers.

► **Applicant to be defaulter , hence application can be rejected:**

- CNT_CHILDREN: applicants who have children equal to or more than 9 default 100% and hence their application can be rejected.
- CNT_FAM_MEMBERS: applicant who have more than 11 family members higher default rate and their application rejected.
- NAME_EDUCATION_TYPE: applicants with lower secondary , secondary education and incomplete higher education have higher default rate.
- NAME_FAMILY_STATUS : Applicants in civil marriage or who are single have higher default rate.
- NAME_INCOME_TYPE: applicants who are either at maternity leave or unemployed have higher default rate.
- OCCUPATION_TYPE: applicants who are low-skill laborer, driver and barmen staff default rate is huge
- ORGANIZATION_TYPE: transport type 3 , industry type 13 people have high defaulting rate and thus should be avoided to be approved for loan
- REGION_RATING_CLIENT: applicants who live in areas with region rating as 3 has highest defaults.

Thank You