Banking Risk Analysis

EDA Assignment

STATEMENT

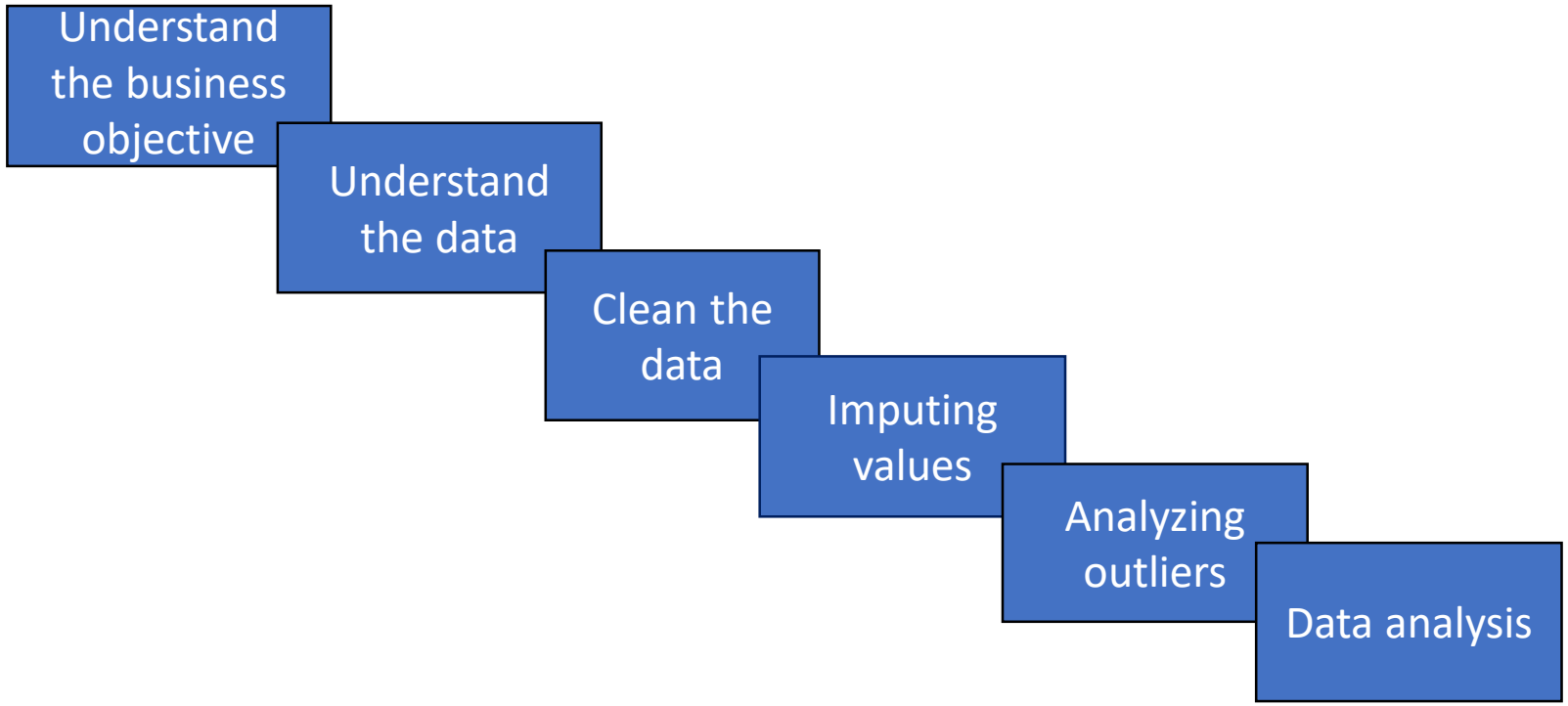
The loan providing company provide different kinds of loans to urban consumers. When customer apply for a loan, based on her\his profile, company has to make decision for loan approval. In this , there are two kind of risks with the company's decision

If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

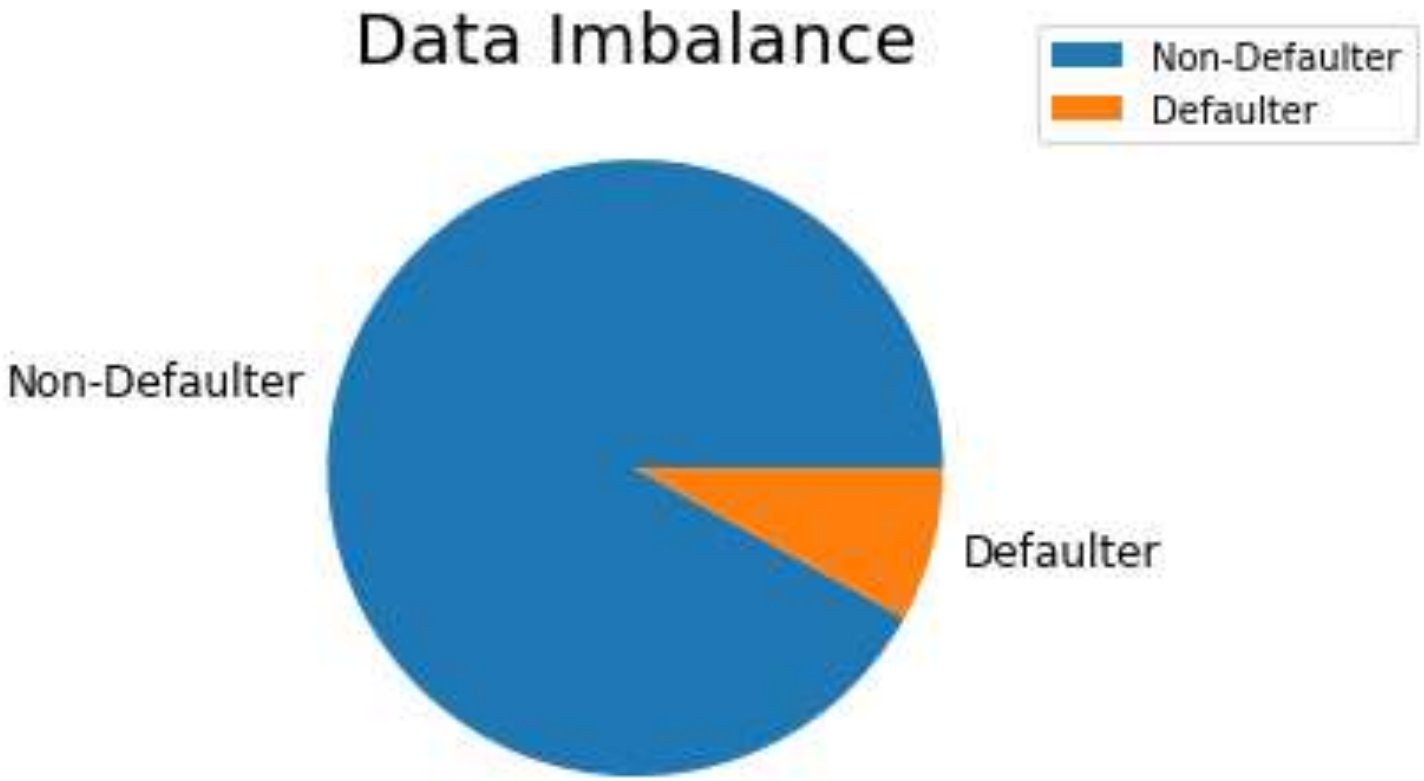
If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Analysis Approach

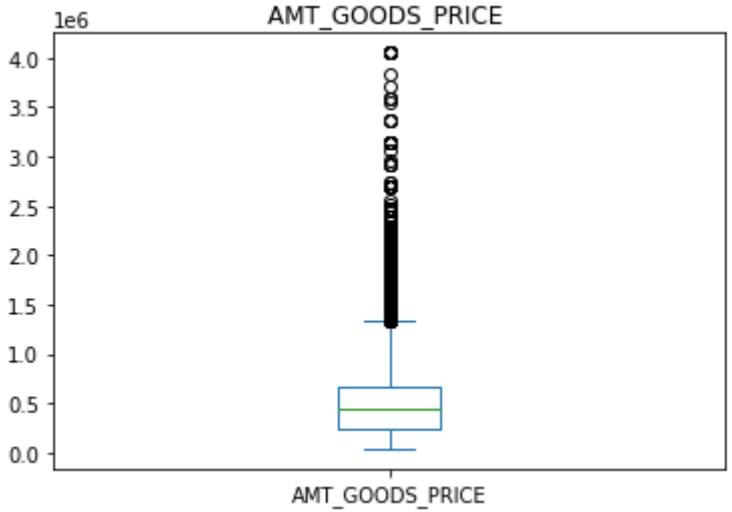


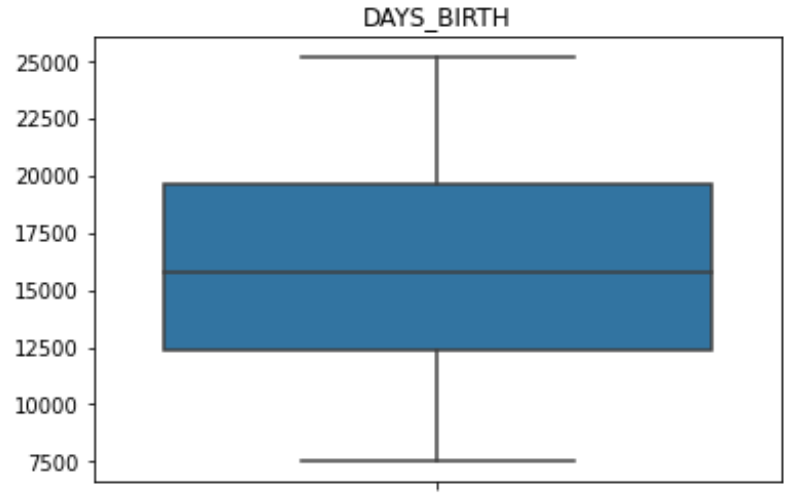
Data Imbalance

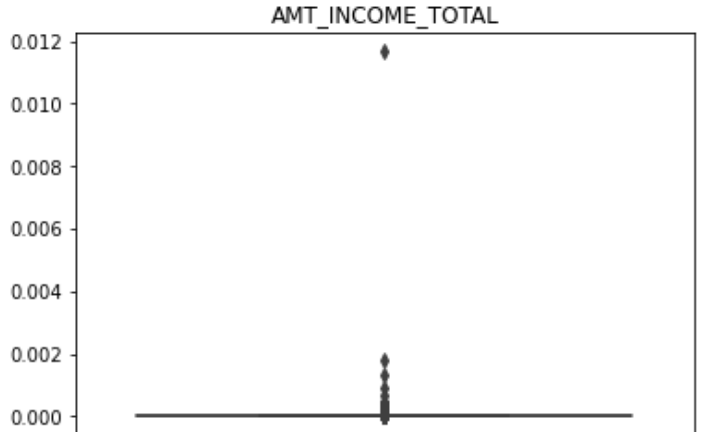


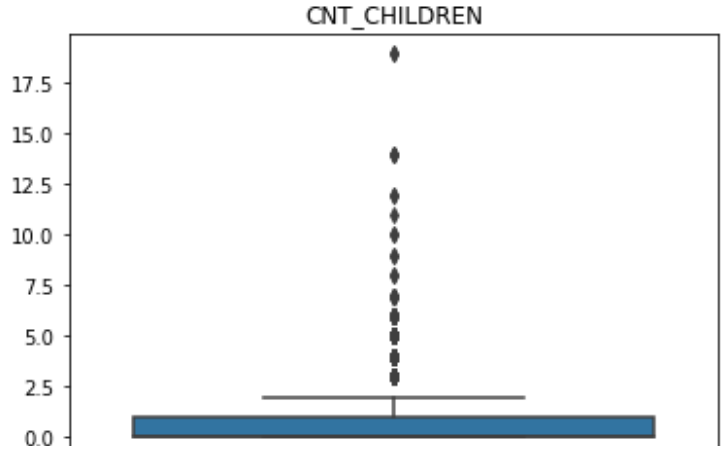
- **data is imbalanced, number of non defaulters applicants more than the number of defaulter applicants**
- **ratio of non defaulter applicants to the defaulters applicants is 11.391:1**

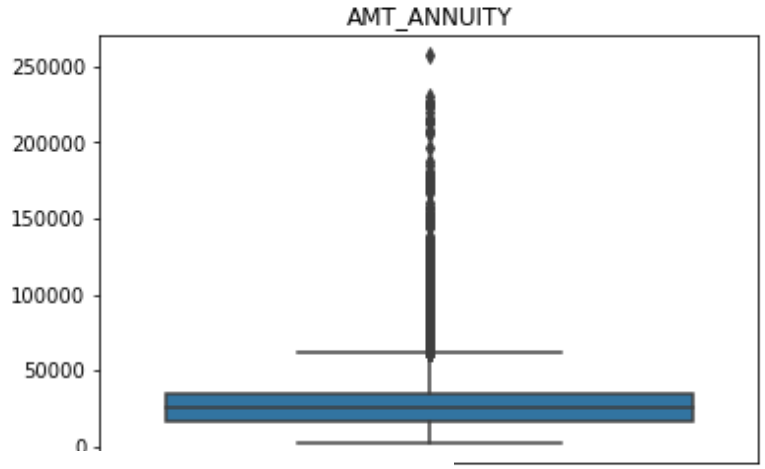
Outlier in Application Data





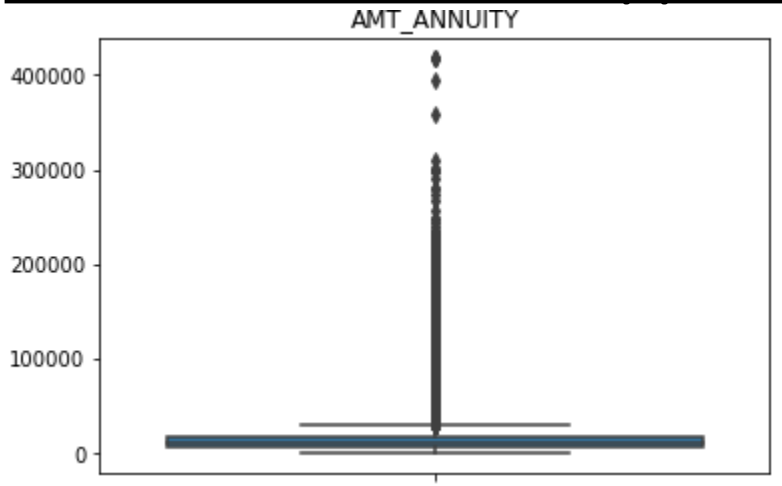


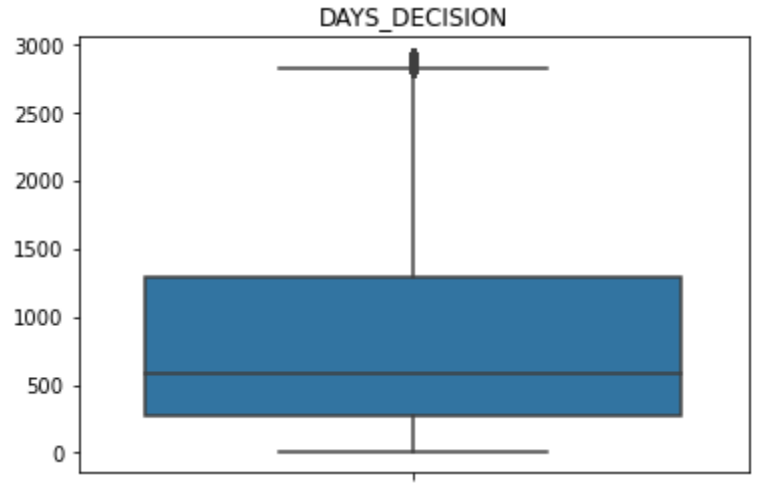


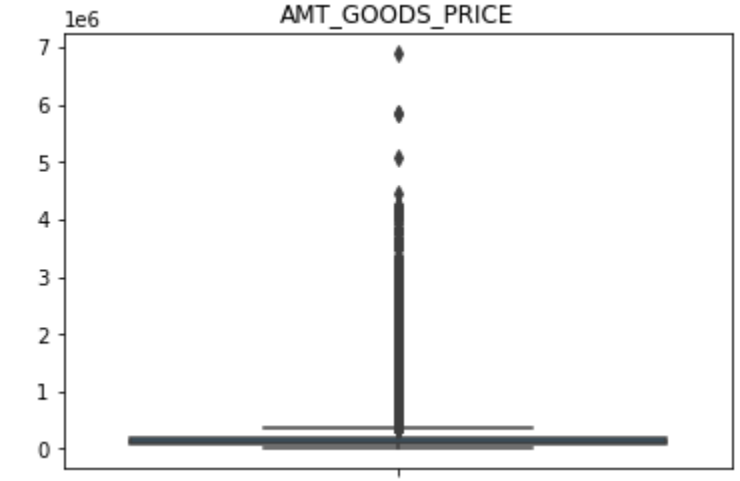


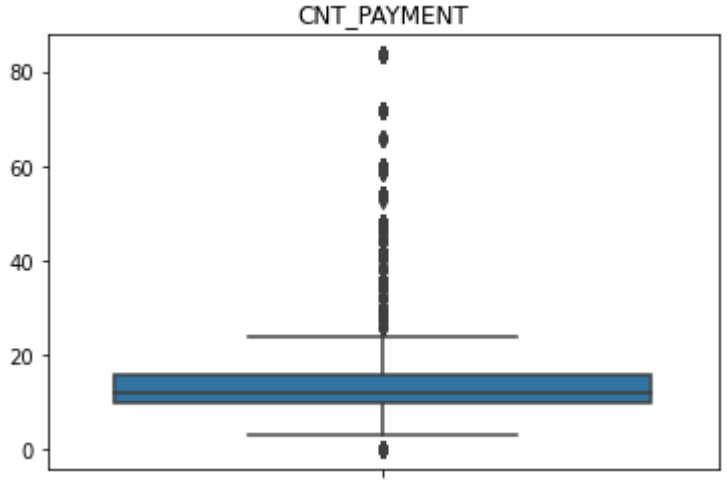
1.AMT_ANNUITY has large no. of outliers
2. AMT_TOTAL_INCOME and CNT_CHILDREN has some no. of outliers
3. DAYS_BIRTH has no outliers which means values are in the reliable
4. In AMT_TOTAL_INCOME, some applicants have very high income then other
5. DAYS_EMPLOYED has one outlier having very large value (greater than 350000 day = 958.91 years), which is impossible. we conclude that there is incorrect entry in data

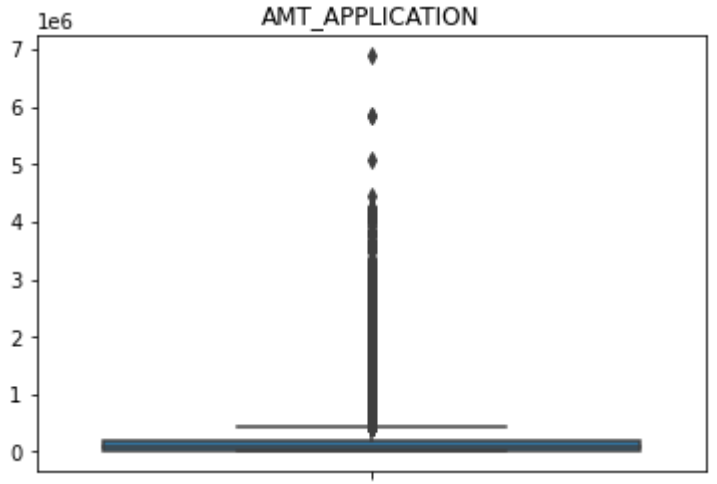
Outliers in Pervious Applications data

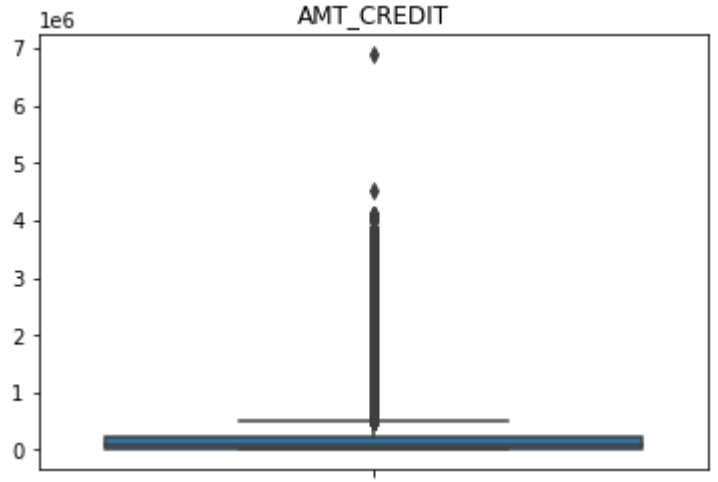


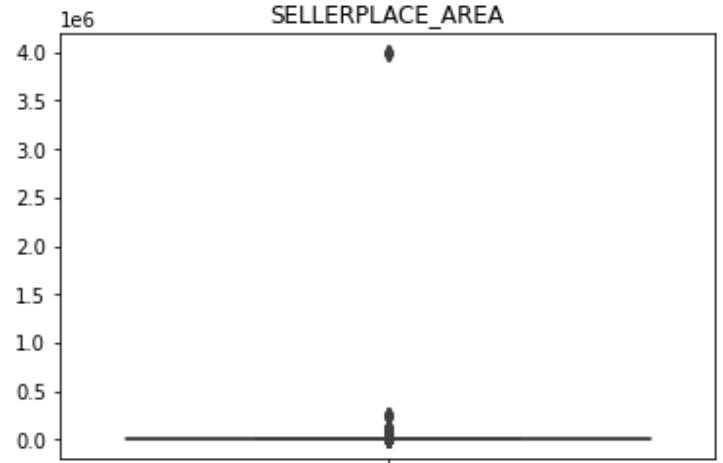






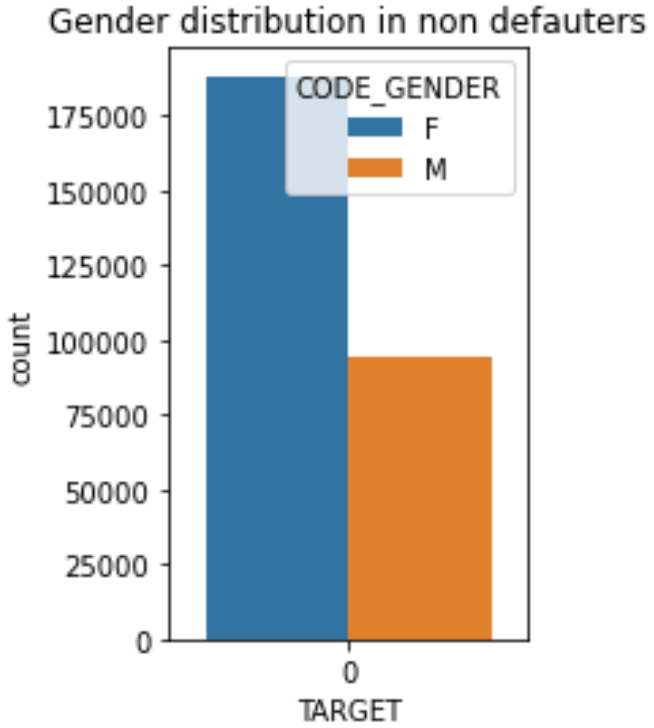


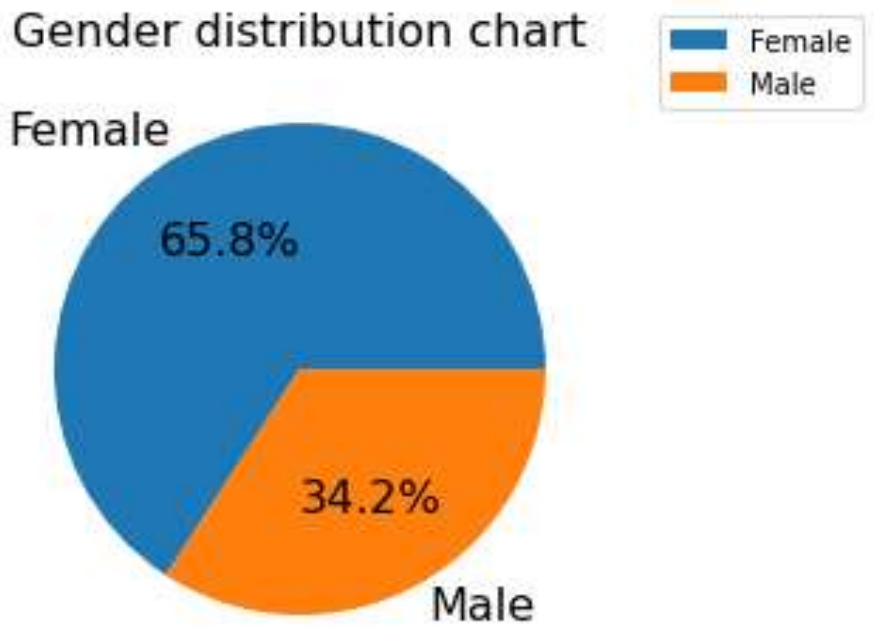




1. AMT_AANUITY,AMT_CREDIT,AMT_APPLICATION, AMT_GOODS_PRICE has large no. of outliers
2. SELLERPLACEAREA has few outliers out them one applicatn has huge area
3. CNT_PAYMENT has some outliers
4. DAYS_DECISION has few outliers and from this box plot we conclude that in previous sessions, some decision took long time

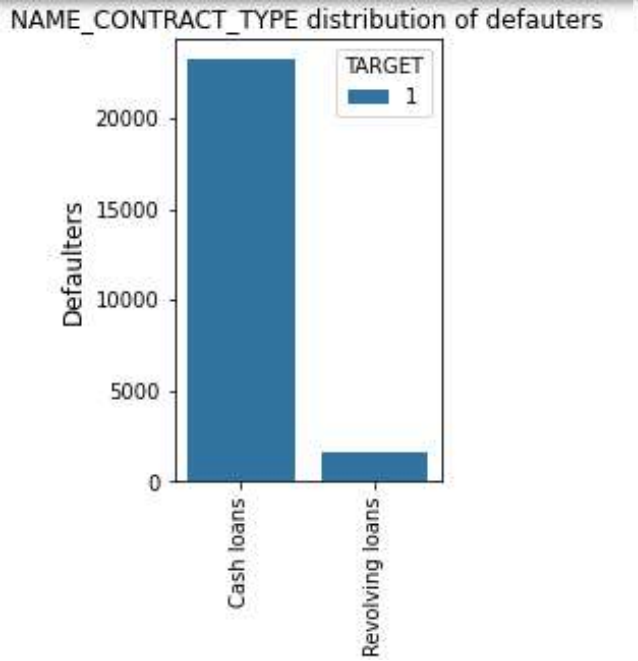
Gender Distributions

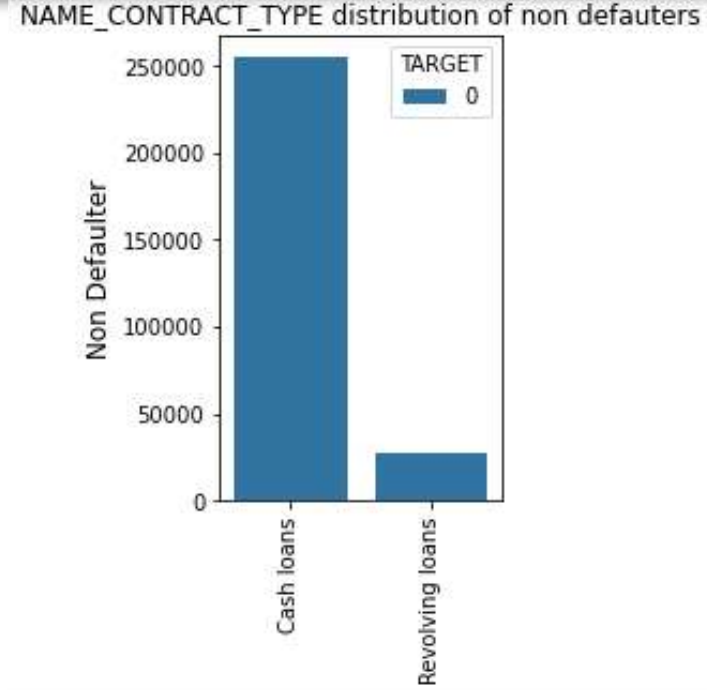


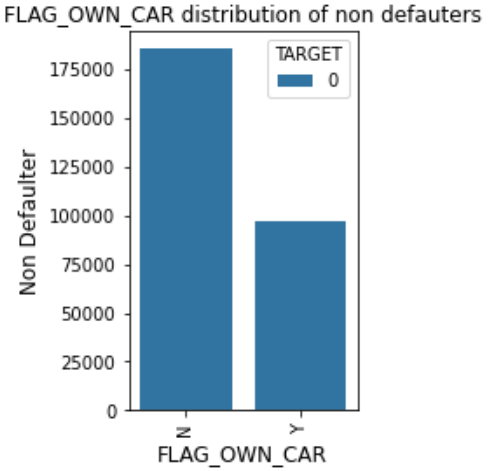


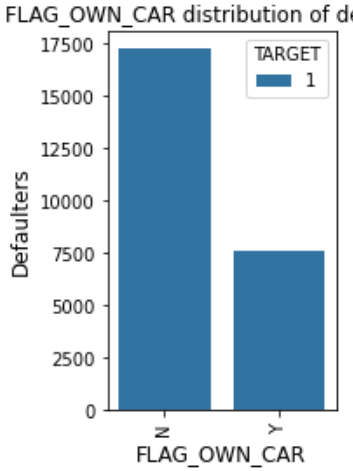
1 From pie chart, percentage of female and male applicants is 65.8% and 34.2% respectively

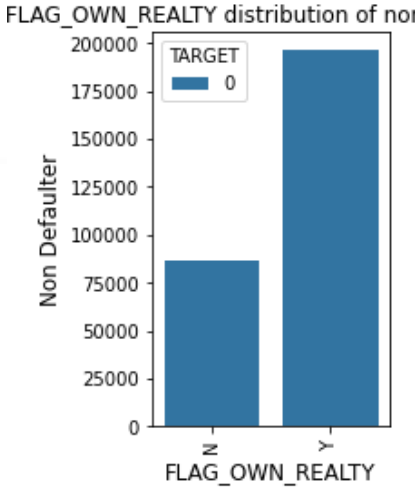
2. From bar graph 1 and 2 we conclude that on the basis of gender, company does not make any decisions

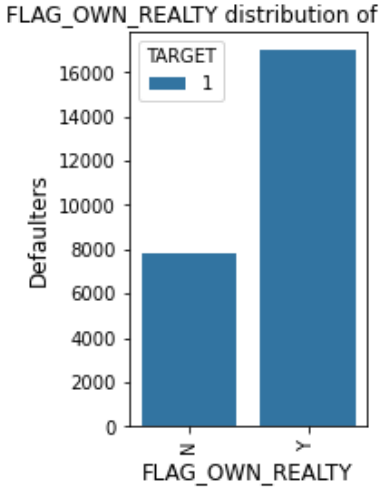


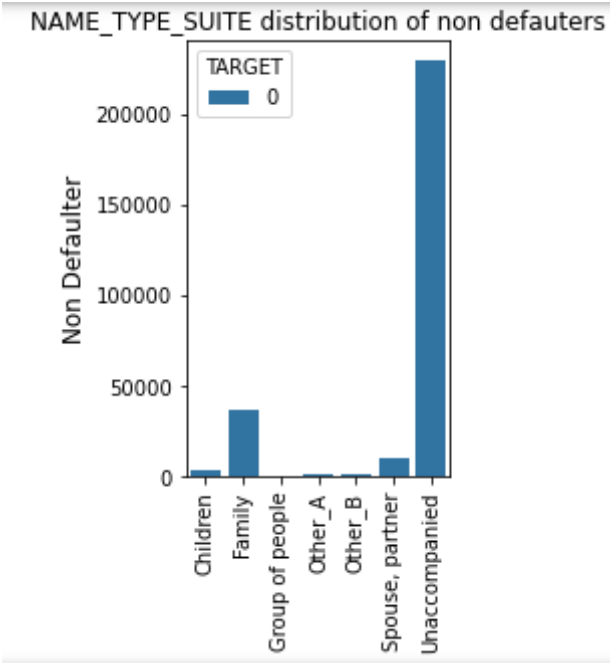


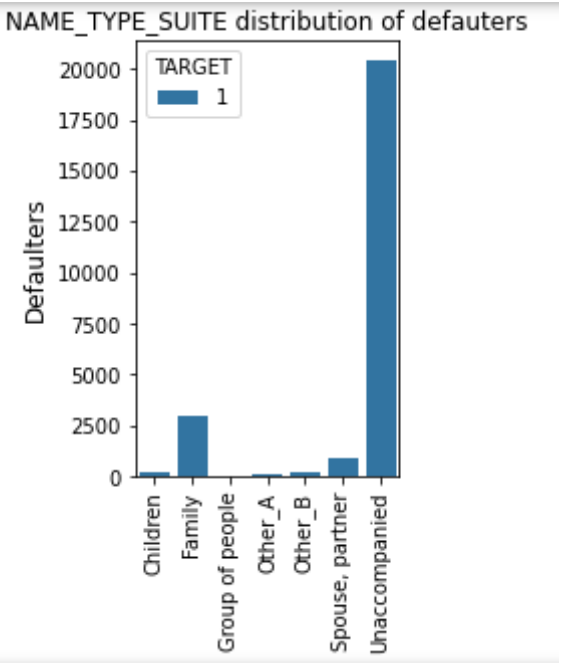






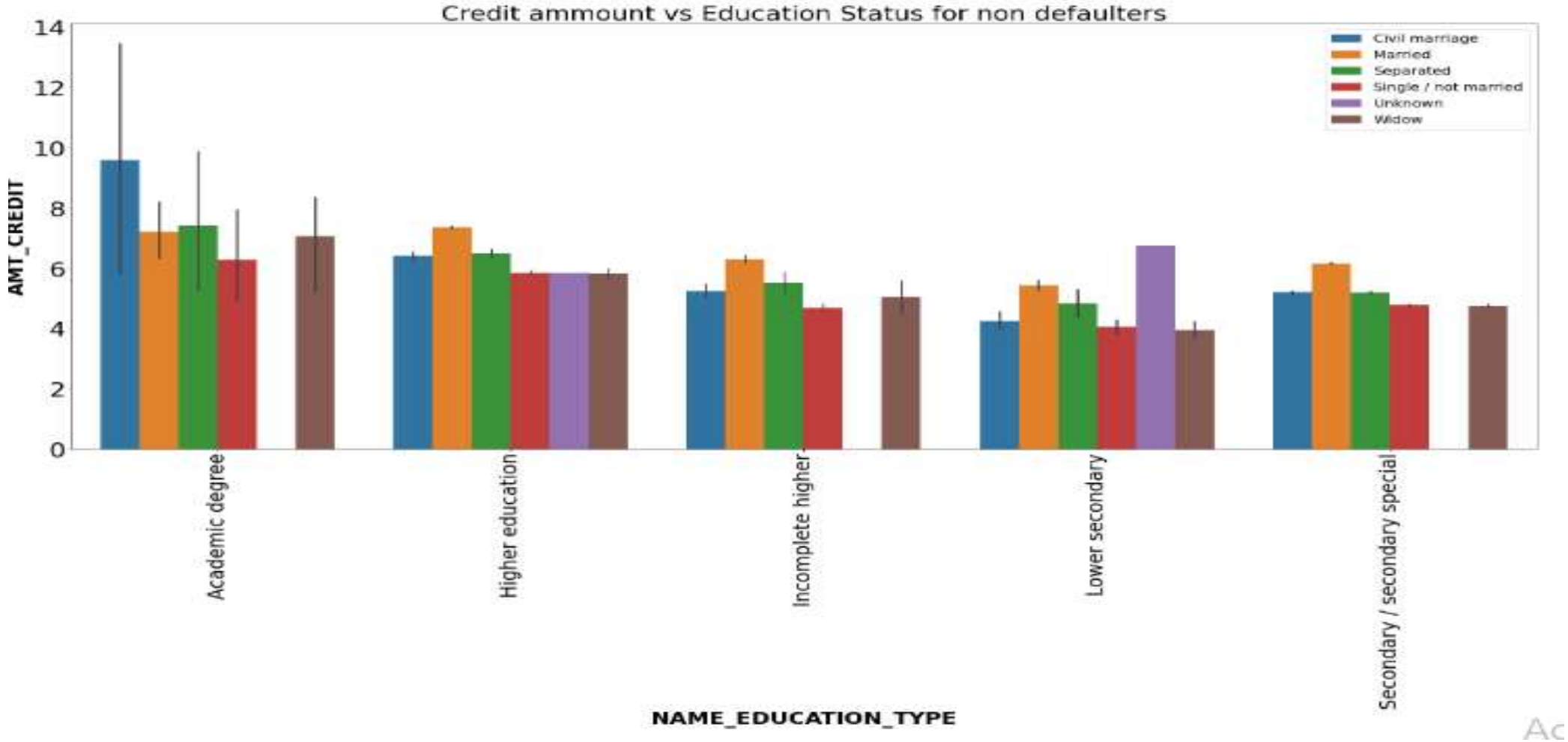






From all above bar graphs it is clear that company should not take any decision on the basis of type of suite, flag own reality, contract type and flag own car

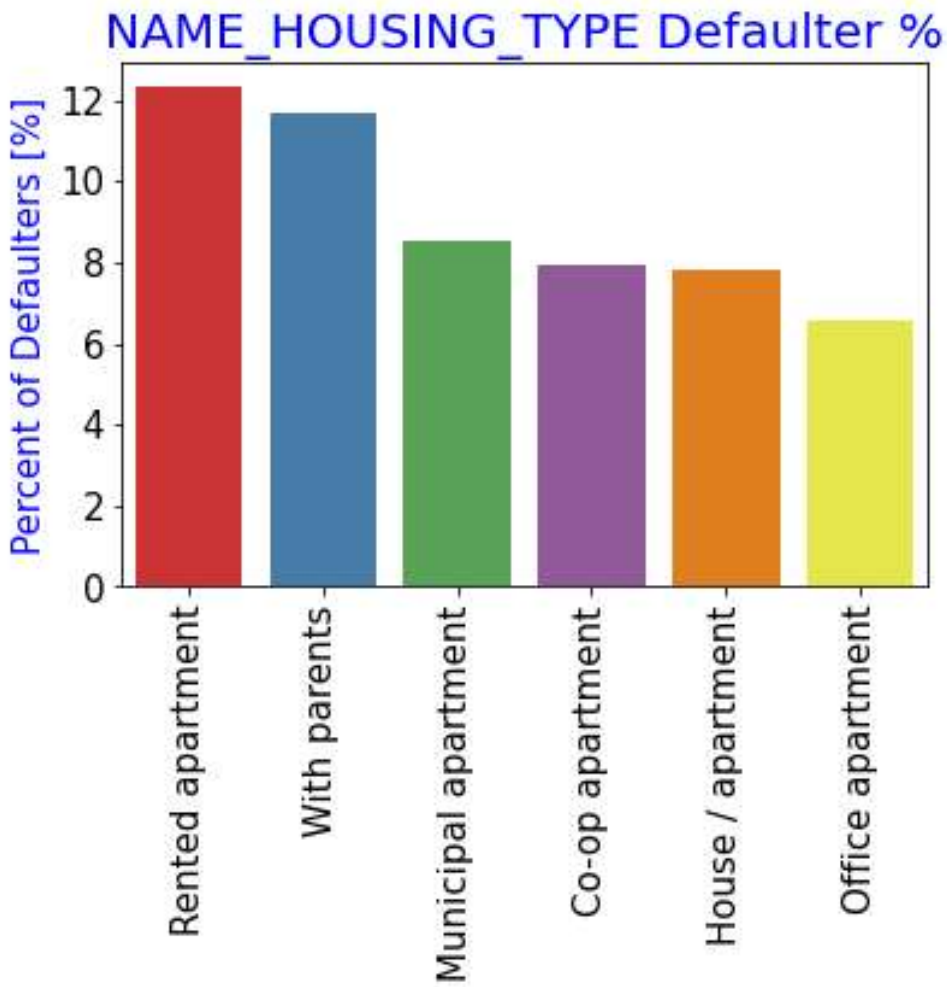
Credit amount versus education status

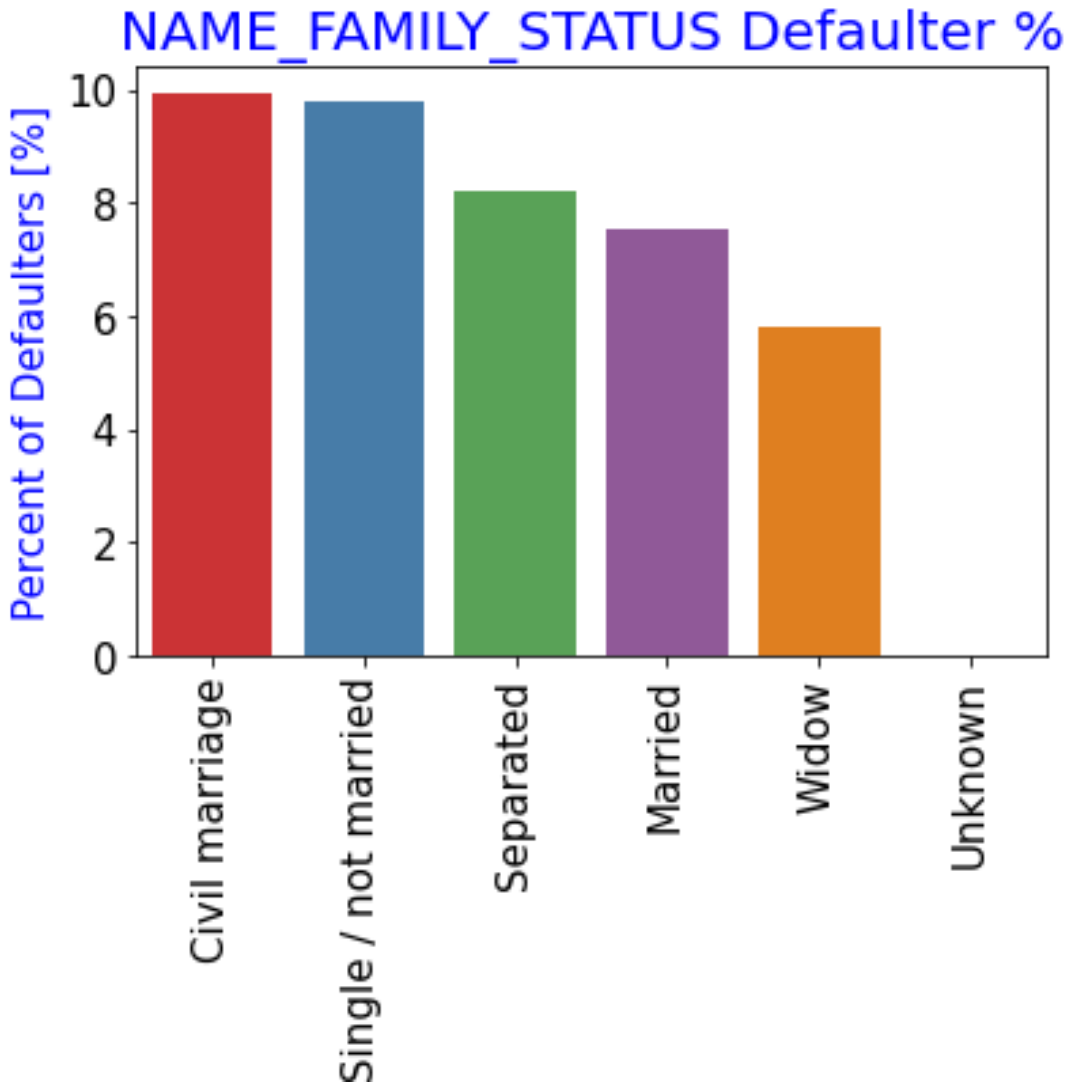


Housing type

- **CONCLUSION**

Applicants have rented apartment and those who live with parents are more likely to be defaulters



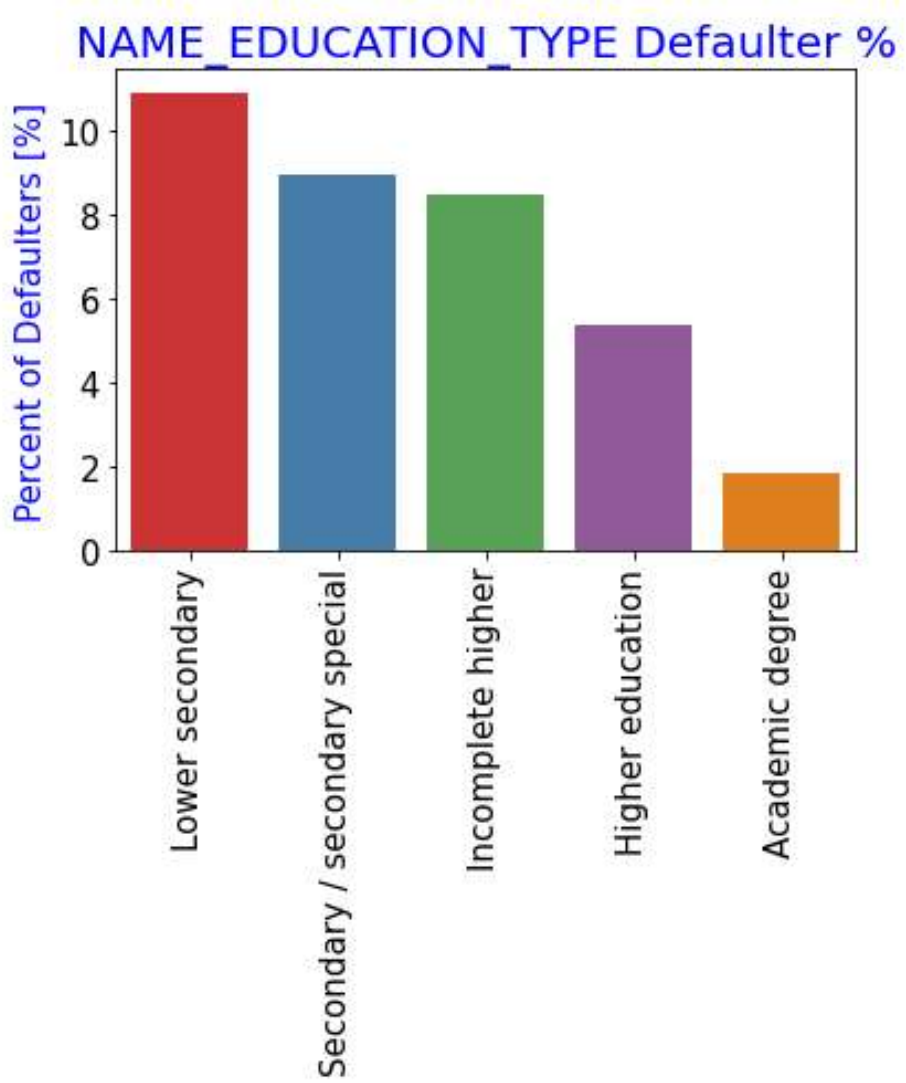


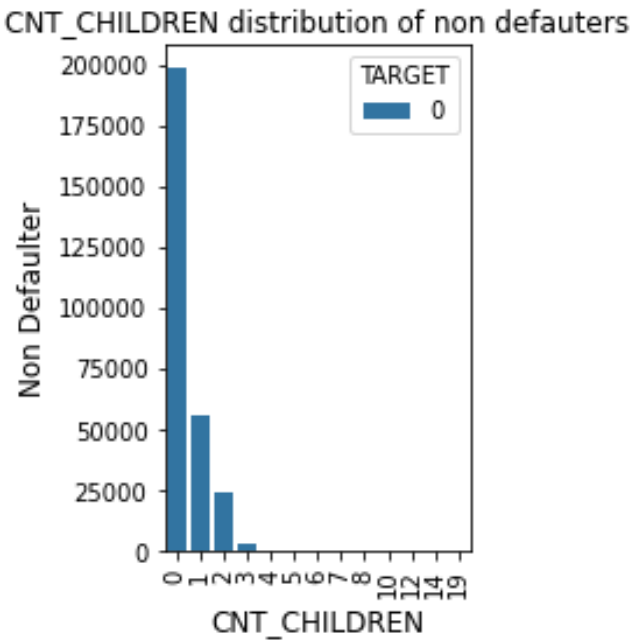
Family Status

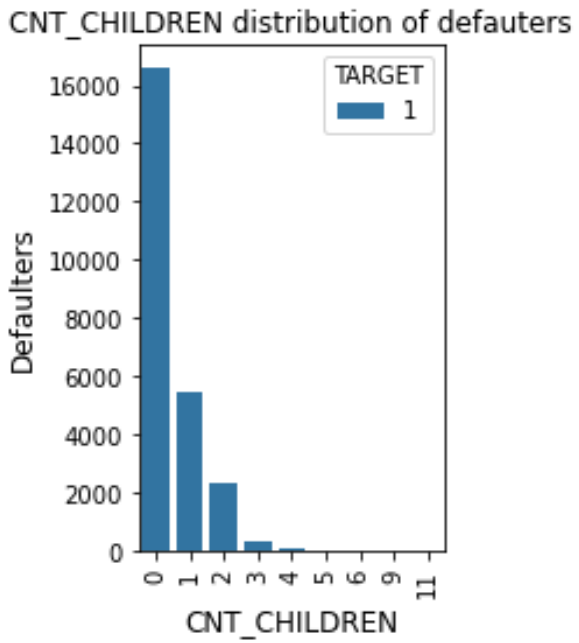
- **INSIGHTS:-**
- Civil married applicants has highest percentage to be defaulters ie. Approx10%
- Single applicants also have high chances to be defaulters
- Where as in comparision widow applicants have less chance to be defaulters

Education type

- From this barplot , we conclude that applicants of lower secondary educations have highest percentage of defaulters and followed by applicants of secondary/secondary special and incomplete higher education and least percentage defaulters are of academic degree holder





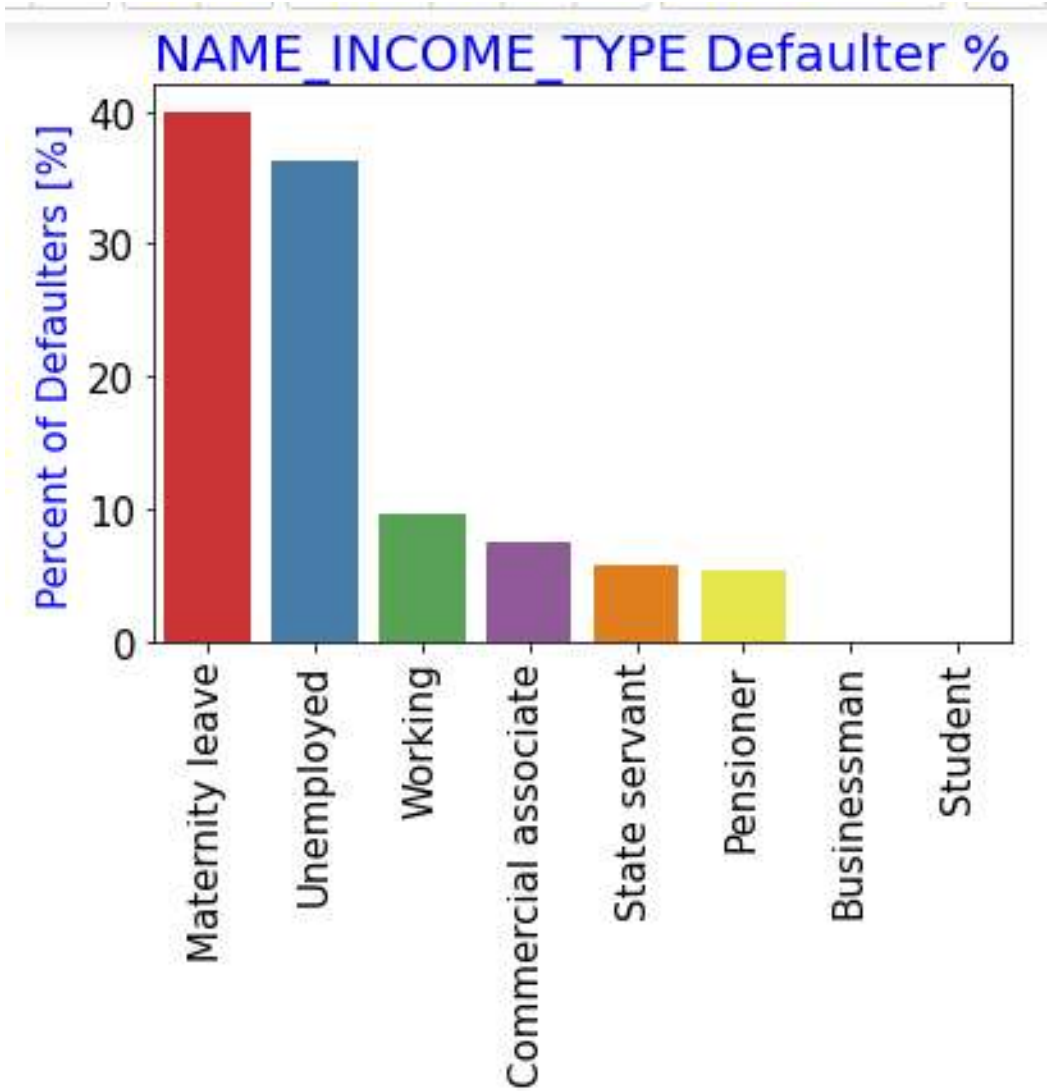


CNT_CHILDREN

- From these graphs , we conclude that applicant having 1-3 children are likely to be defaulter and non defaulter but applicant having four children have more chances to be defaulters

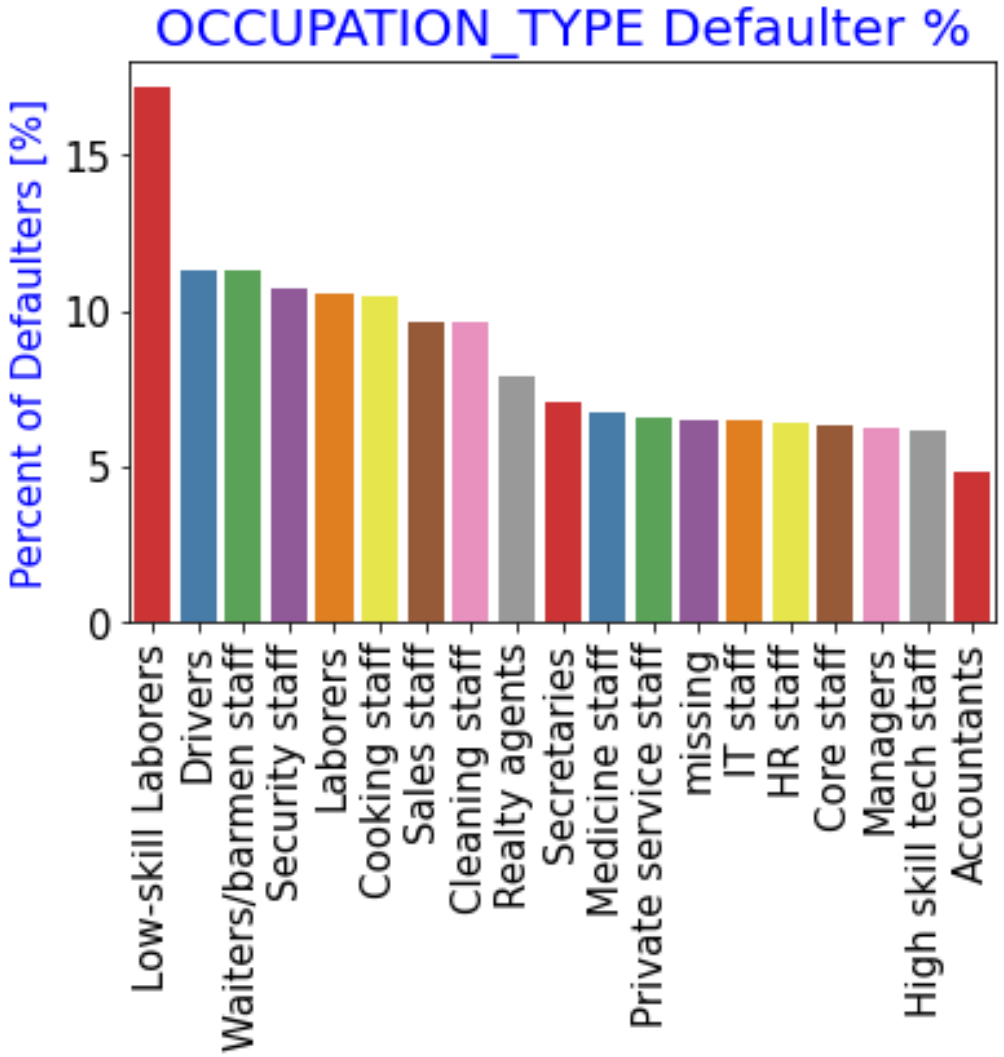
Income Type Defaulters

- Applicants on maternity leaves has high percentage of defaulters i.e. approx 40%
- Unemployed applicants are on second number having defaulter percentage approx. 37%

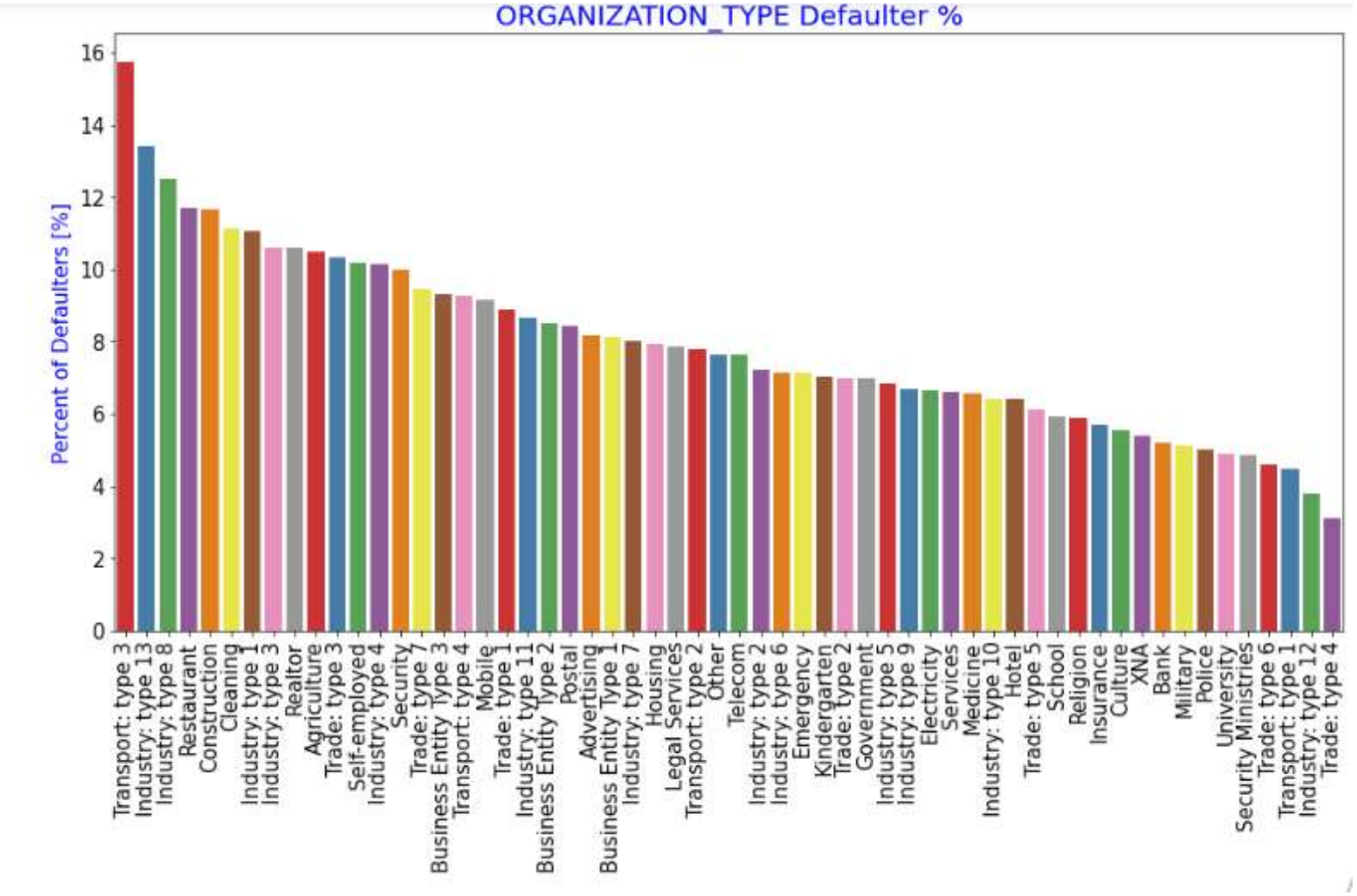


Occupation type

- Low skill laborers have maximum percentage of defaulters
- Drivers, waiter/barmen staff has almost same percentage
- Accountants have least defaulters percentage

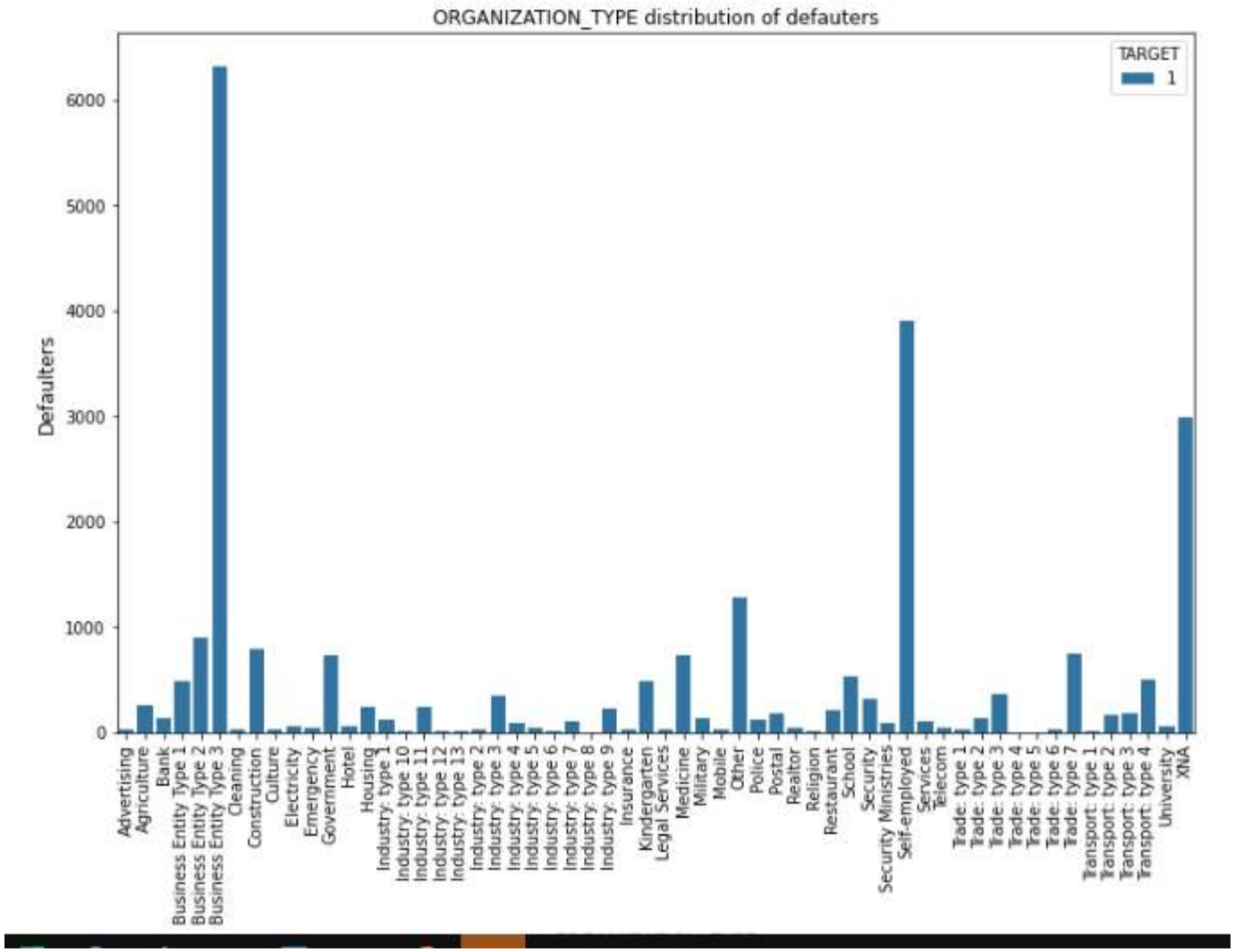


Organization type



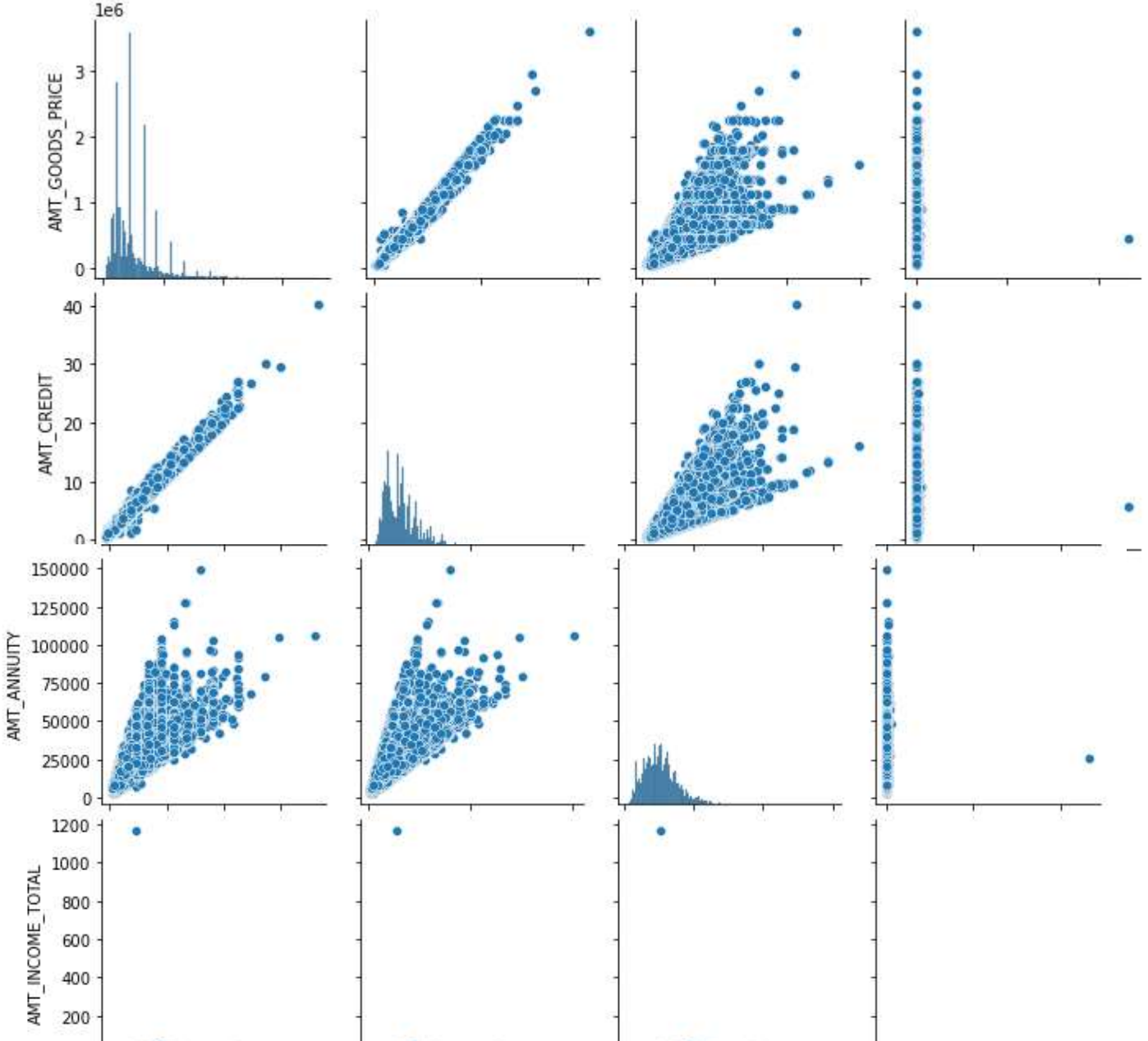
1. Huge no. of defaulters are from transport:type3 organizations (approx16%)
2. Least no. of defaulters are from industry:type12 and trade:type4

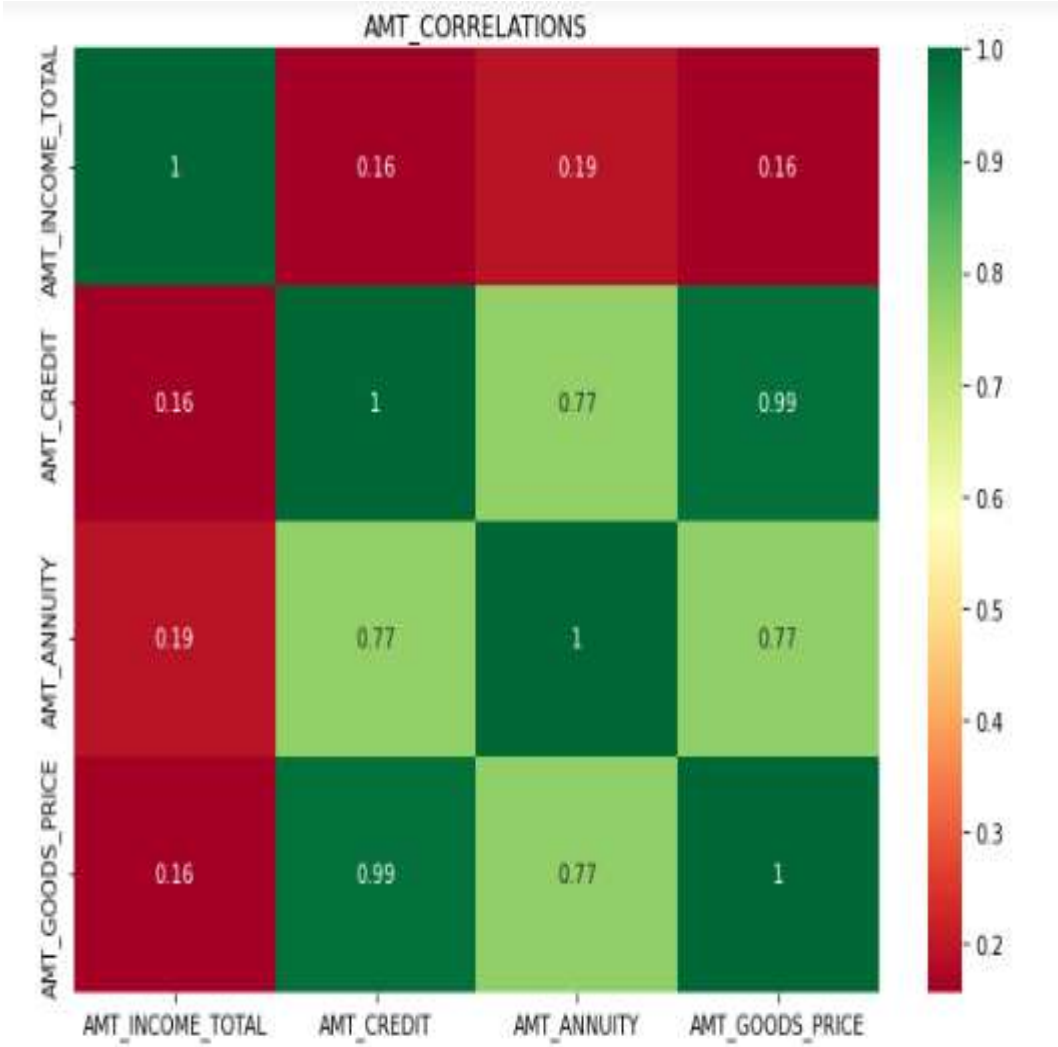
Organization type



1. maximum no. of applicants from business entity type3
2. Self employed applicants are on second and xna are on third in this list

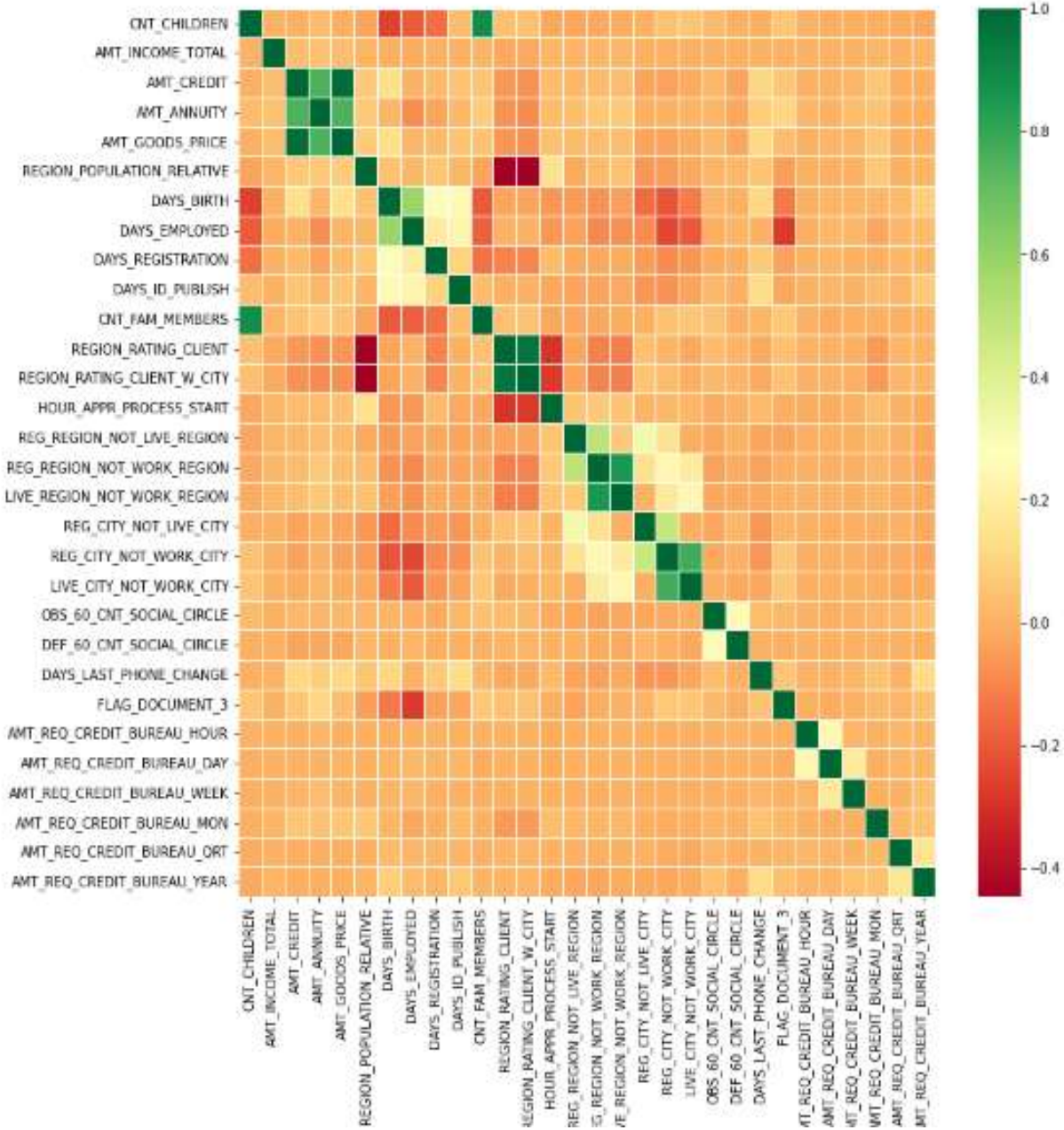
AMT CORRELATIONS





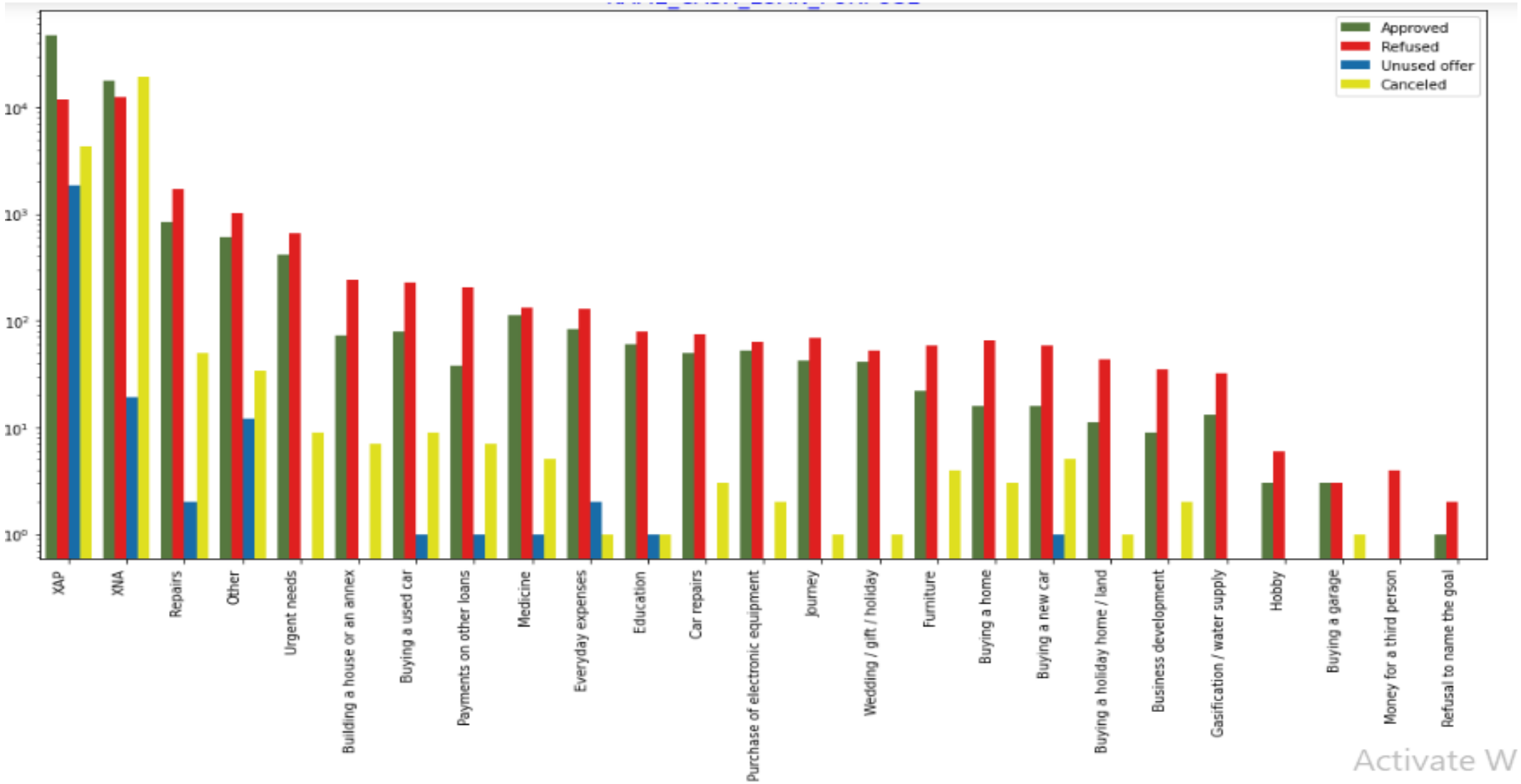
From Heat map and pair plot , its is very clear that there is high correlation between AMT_GOODS and AMT_CREDIT means those have high goods price also take loan of high amount

Mutivatriate correlations for defaulter



From this multivariate heatmap, we found that there is high correlation between AMT_ANNUITY, AMT_CREDIT and AMT_GOODS PRICES

Defaulters vs cash loan purpose



Submitted by:- Shreya