Bank Loan and its Status Analysis

BY-

NAYANTARA SINGH

Table of Contents

Aim of Study

Ensure that the consumers capable of repaying the loan are not rejected.

Identification of such applicants using EDA

The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.

Problem Statement

Two types 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.

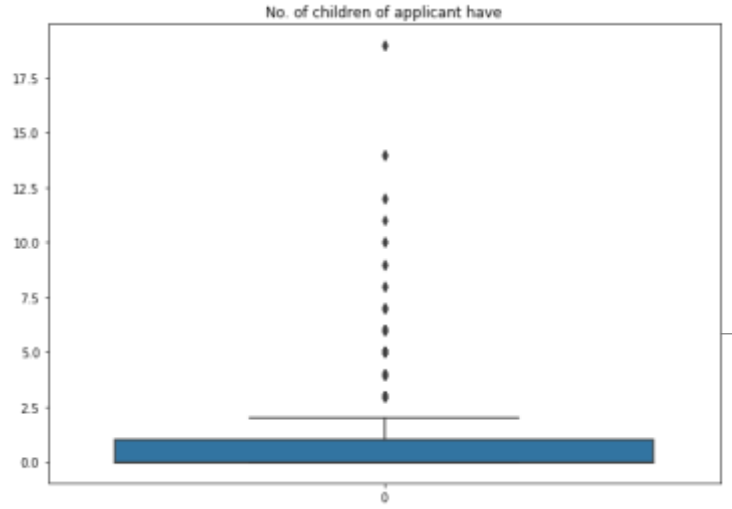
SK_ID_CURR	TARGET	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELATIVE	FLAG_EMP_PHONE	CNT_FAM_MEMBERS	REGION_RATING_CLIENT_W_CITY	TOTAL_REQ_CREDIT_YEAR	FLAG_DOC_NOTSUBMITTED	AGE	YEARS_OF_EMPLOYMENT	
count	306203.000000	306203.000000	306203.000000	3.062030e+05	3.062030e+05	306203.000000	3.062030e+05	306203.000000	306203.000000	306203.000000	306203.000000	306203.000000	306203.000000	306203.000000	306203.000000
mean	278165.940905	0.080845	0.417021	1.687823e+05	5.988010e+05	27122.210470	5.379491e+05	0.020865	0.819760	2.152794	2.031652	2.148914	0.928675	43.948838	185.738974
std	102787.005389	0.272597	0.722113	2.375221e+05	4.019625e+05	14490.897429	3.689205e+05	0.013830	0.384388	0.910595	0.502790	2.291216	0.342580	11.960925	382.271564
min	100002.000000	0.000000	0.000000	2.565000e+04	4.500000e+04	1615.500000	4.050000e+04	0.000290	0.000000	1.000000	1.000000	0.000000	0.000000	21.000000	0.000000
25%	189133.500000	0.000000	0.000000	1.125000e+05	2.700000e+05	16551.000000	2.385000e+05	0.010006	1.000000	2.000000	2.000000	0.000000	1.000000	34.000000	3.000000
50%	278187.000000	0.000000	0.000000	1.476000e+05	5.135310e+05	24930.000000	4.500000e+05	0.018850	1.000000	2.000000	2.000000	2.000000	1.000000	43.000000	6.000000
75%	367127.500000	0.000000	1.000000	2.025000e+05	8.086500e+05	34596.000000	6.795000e+05	0.028663	1.000000	3.000000	2.000000	3.000000	1.000000	54.000000	16.000000
max	456255.000000	1.000000	19.000000	1.170000e+08	4.050000e+06	258025.500000	4.050000e+06	0.072508	1.000000	20.000000	3.000000	262.000000	4.000000	69.000000	1001.000000

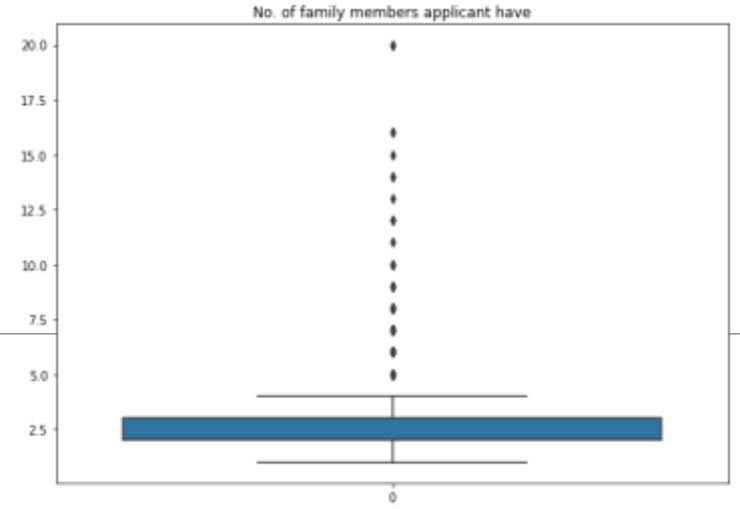
Inferences from the table above

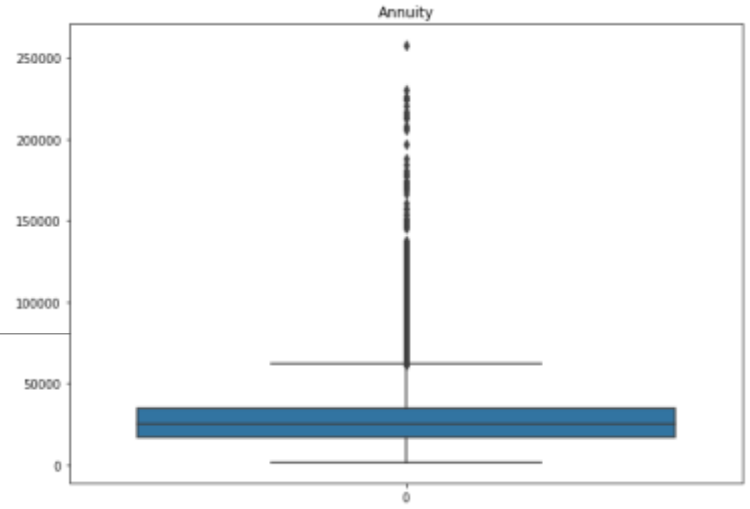
- Years of Employment has max of 1001 year which is not possible (hence outlier)
- Maximum age of people for loan is 69 years
- 75% of people have 1 child or no child hence most people having more than 5 to 6 and beyond can be treated as extreme outliers
- 75% of people have family members upto 3 and max is 20 hence it can be also seen as an extreme case or outlier
- 50% of the people have got credit enquiries of approx 0 to 3 times in a year before applying for loan

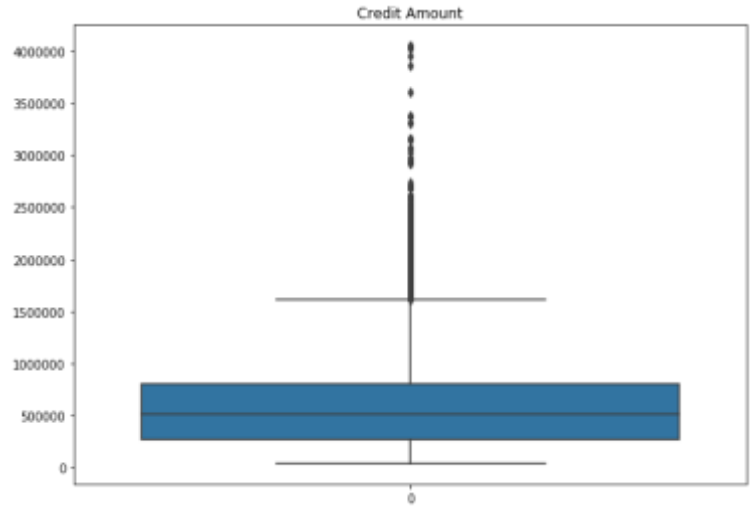
Other inferences

- On an average most people have not submitted just 1 document out of 21 documents (can imply that document submission might not be a factor for defaults)
- We can see that amount income and amount credit has really large values hence we can use binning for them
- Can use binning for age for better visualisation and grouping







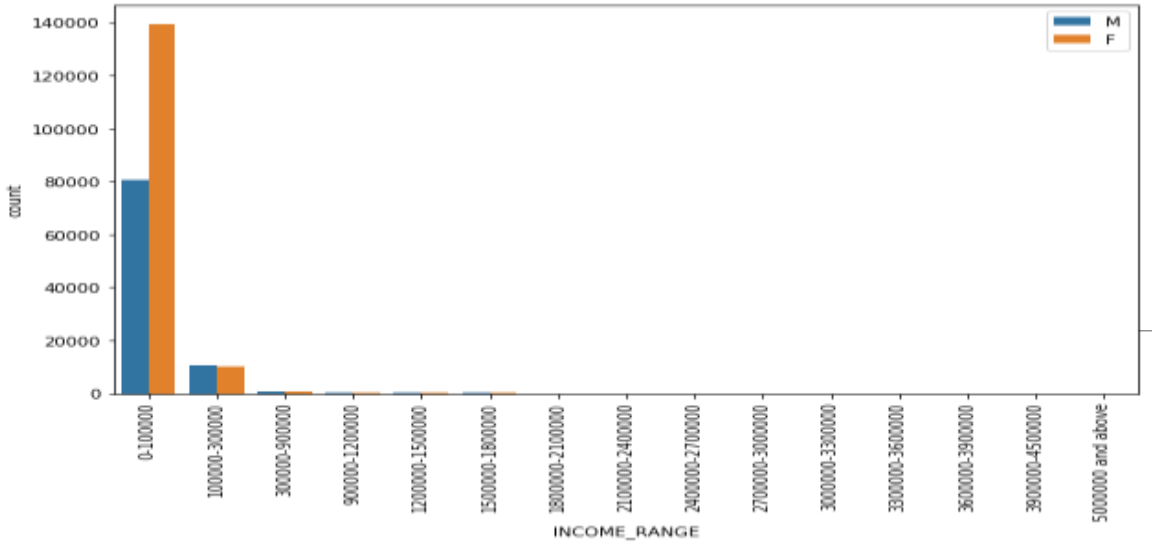


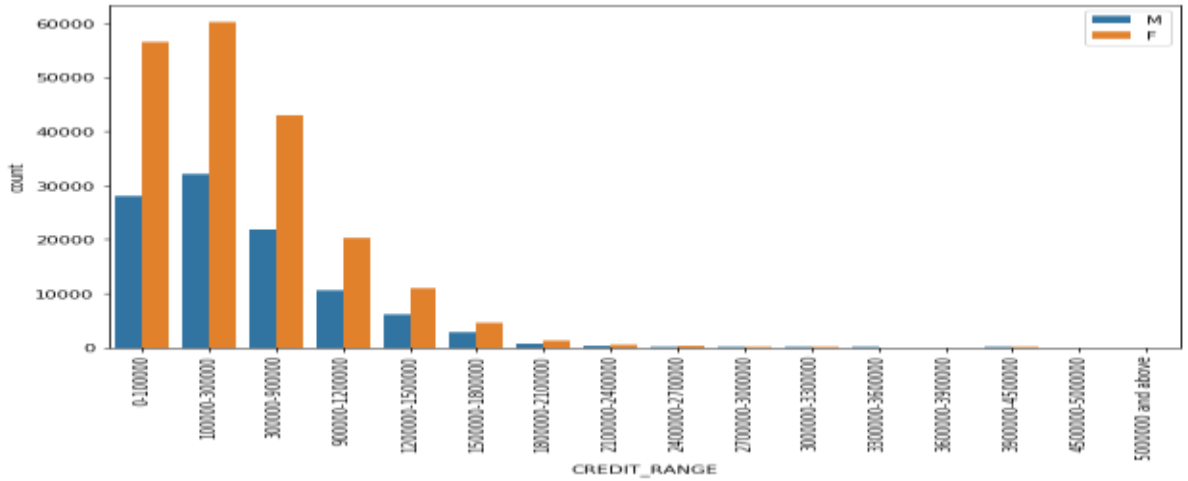
- The 99th percentile of no. of children is 3 i.e 99% of people have 3 children or below

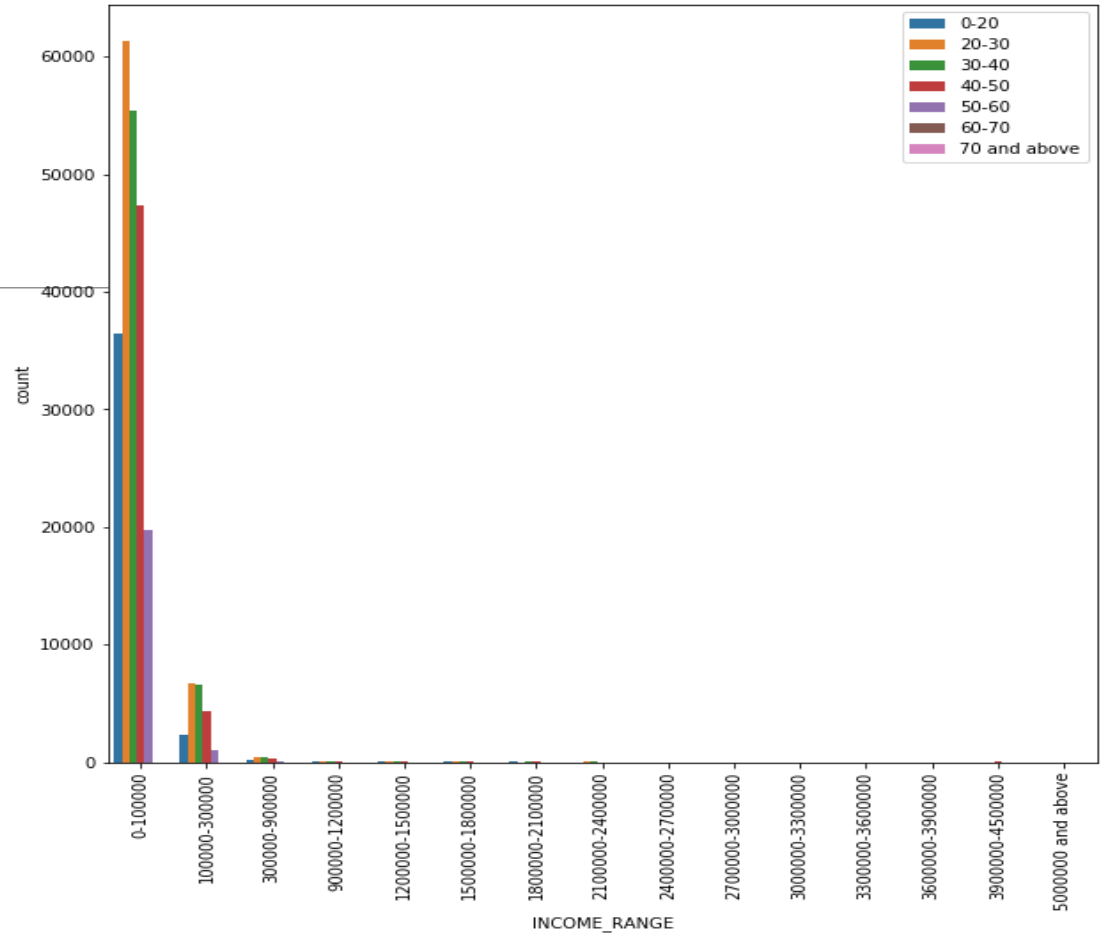
- The 99th percentile of family members is 5 i.e 99% of people have family members 5 or below

- We can see that there is vast difference between percentiles of 75th and above for annuity and credit amount as well

- To treat such outliers we can always cap the maximum to 99th percentile for the last 1% o rows



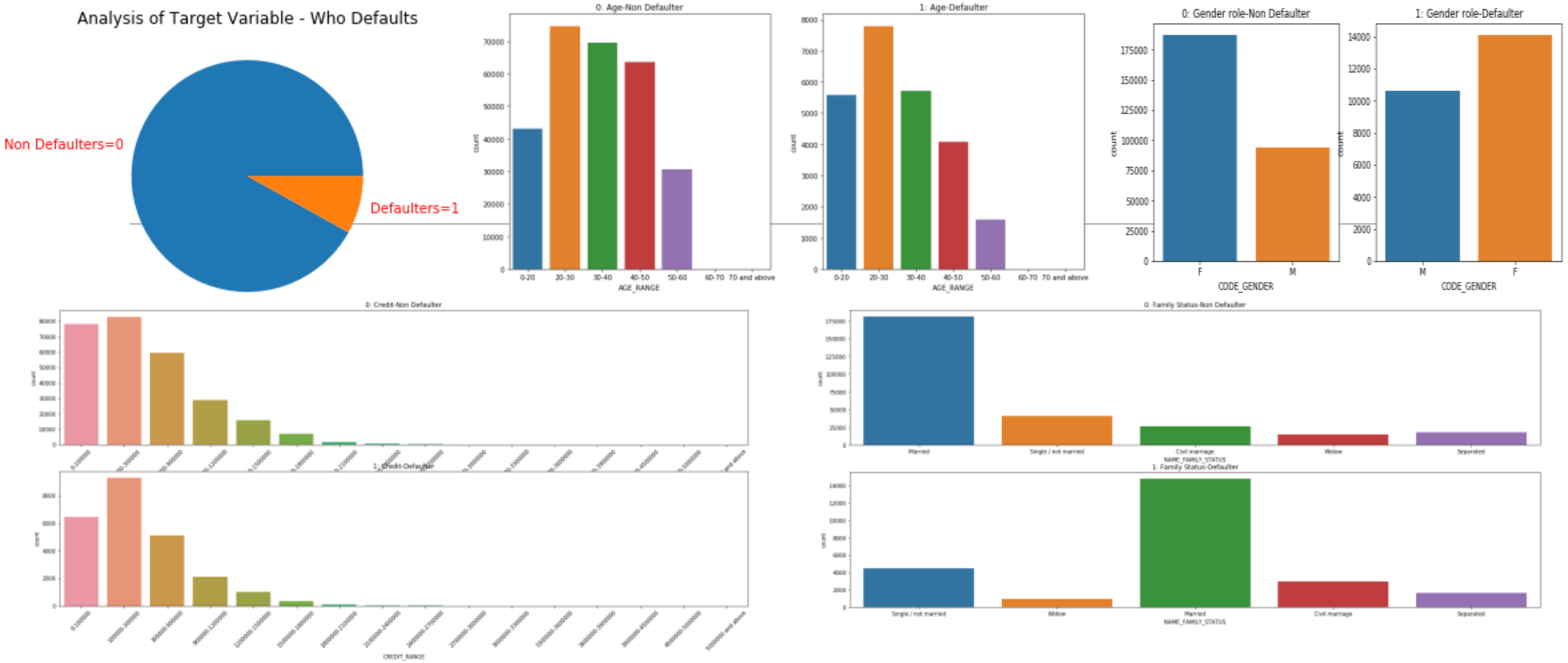




- females have more income that males for lower income range and the distribution gets even for both gender in higher income groups

-females also got more credit amount than males

-for less income group 20-30 have highest count followed by 30-40(with age income rises too)



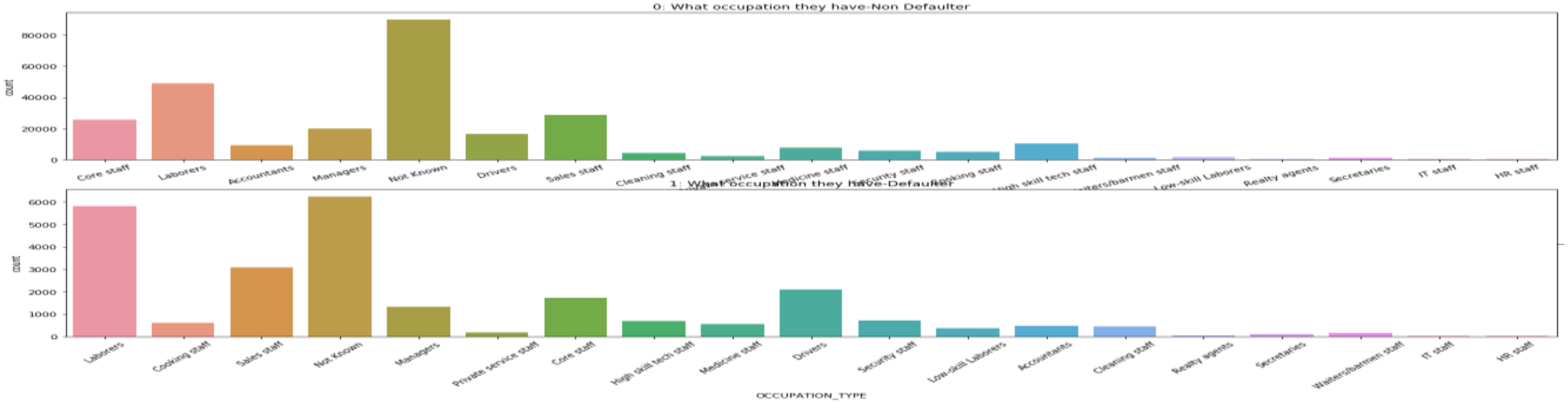
- The target variable is highly imbalanced

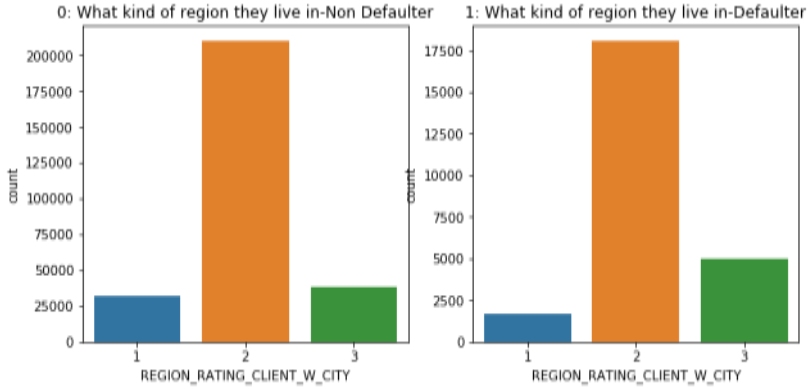
-20-30 age group have maximum defaulters

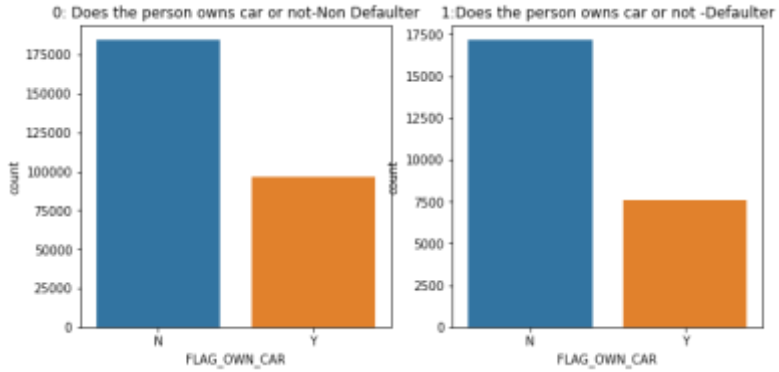
- females defaulters and non defaulters are both more than males so we can say as this can be because of high acceptance of loans of females

- married people are more defaulters and non defaulters can be because they might have high approvals

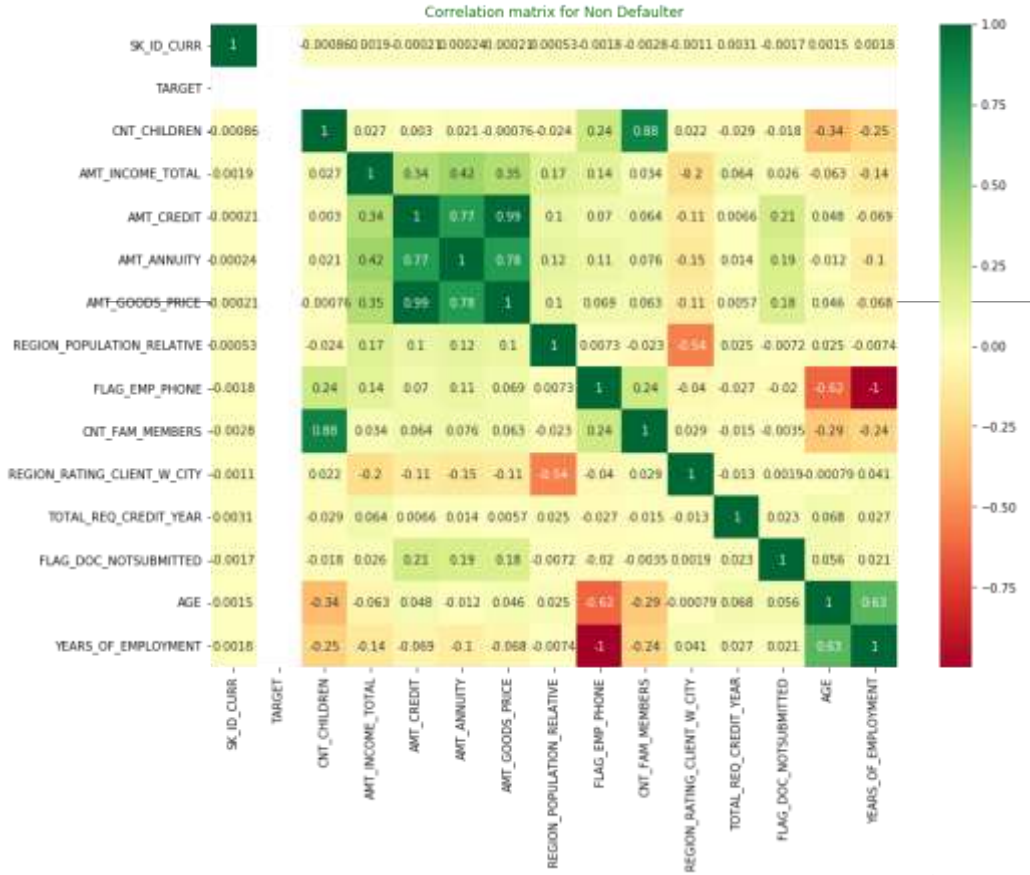
-credit defaulters and non defaulters maximum lie in ranges of 3 lacs or so

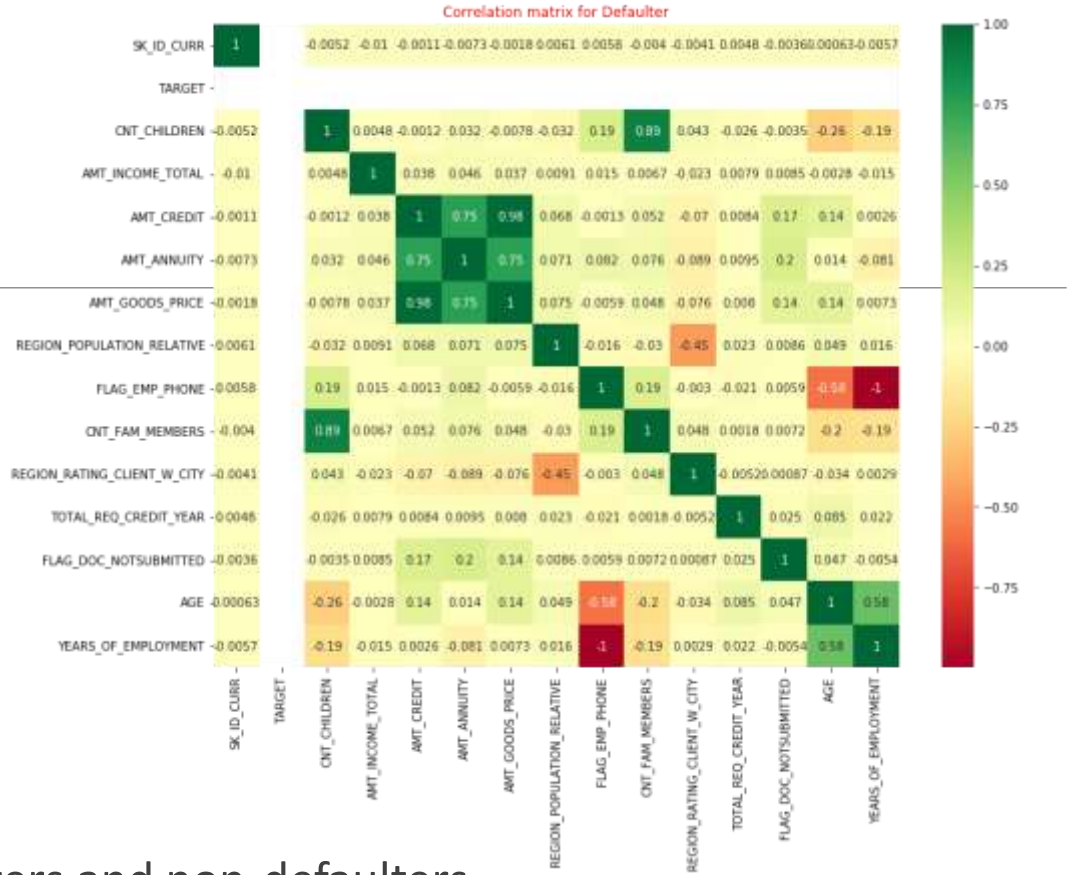






- Labourers default the most in loans followed by sales staff and driver
- Region 2 might have people taking maximum loans
- More people who don’t own a car take loan

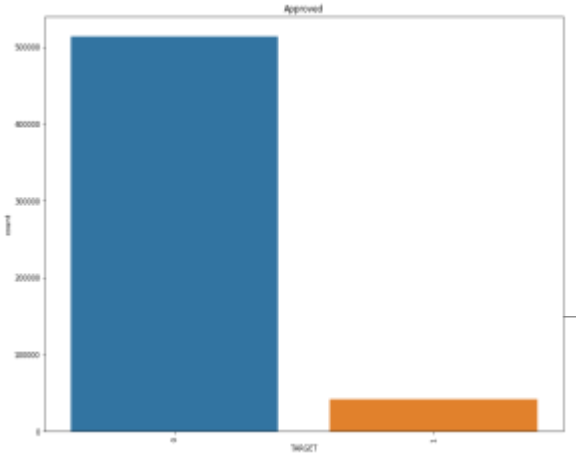


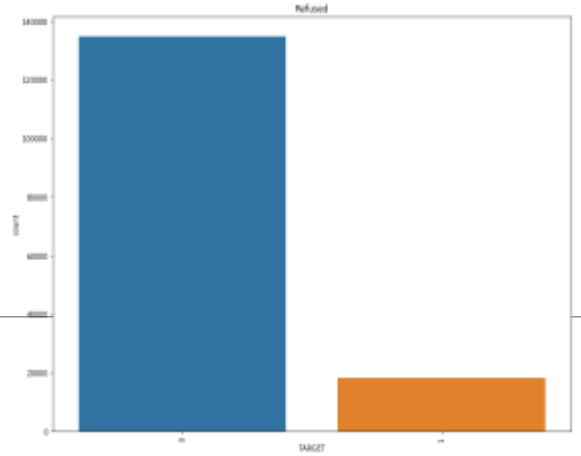


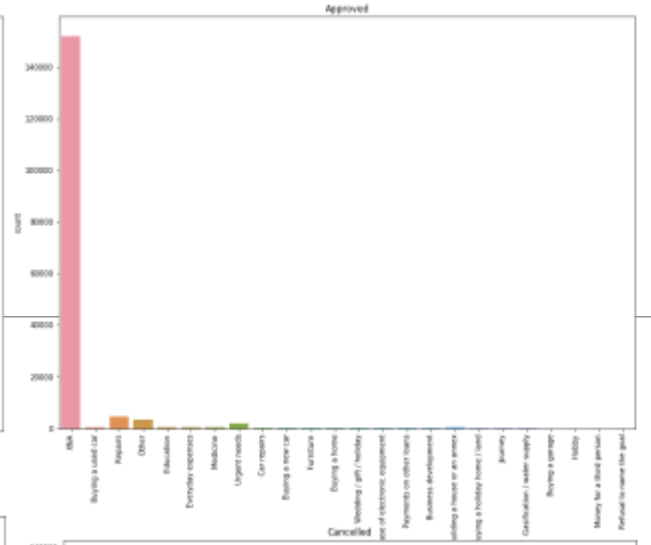
-Correlation Matrix are moreover same for defaulters and non-defaulters

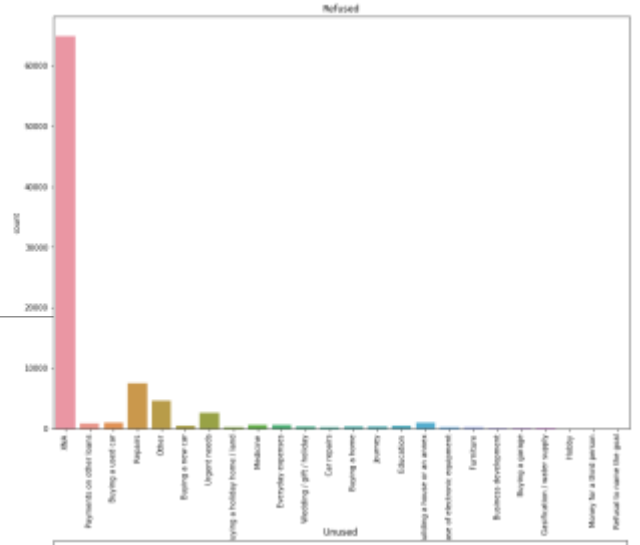
- Years of employment have perfect negative correlation with giving work phone number i.e. elder people are retired or at high position(age high) people share no. less

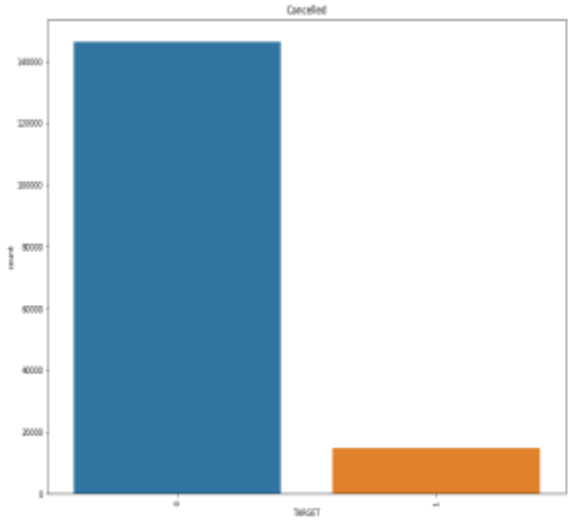
- high correlation between family members and children- maybe most family have children as family members as well

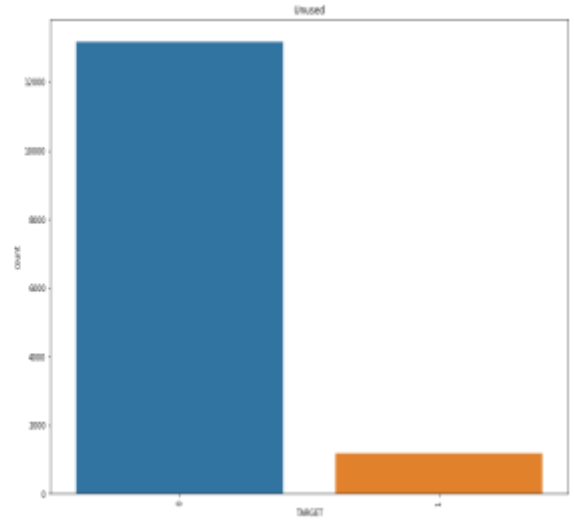


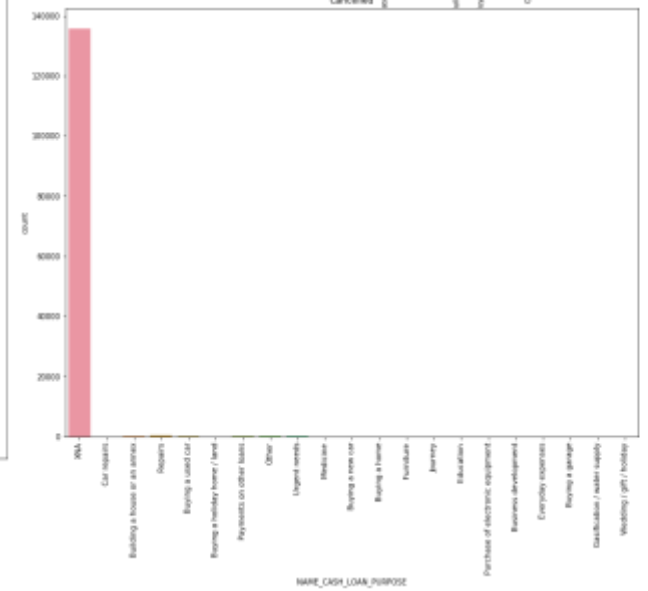


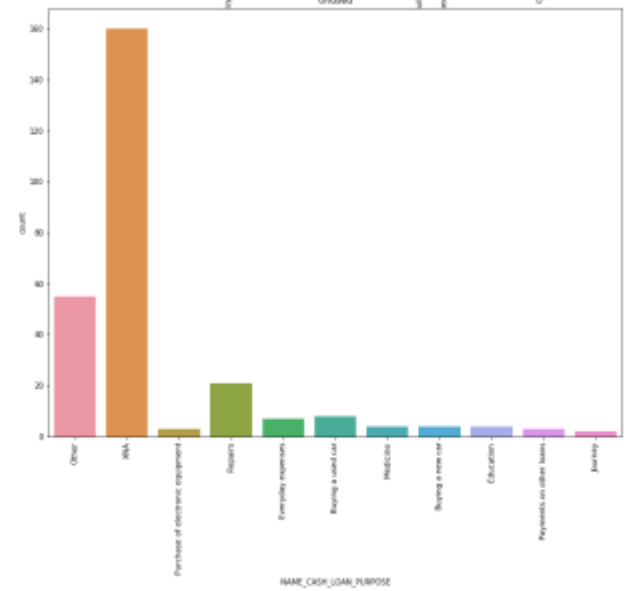




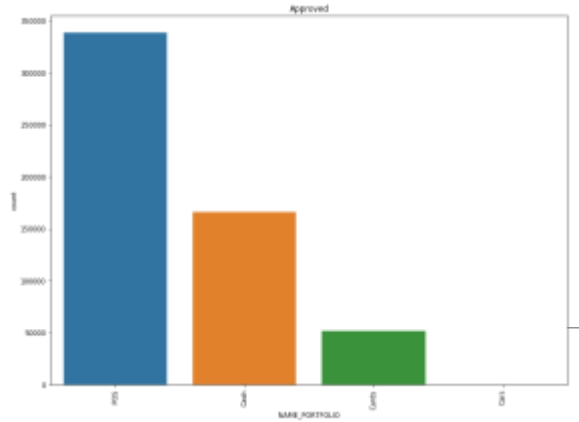


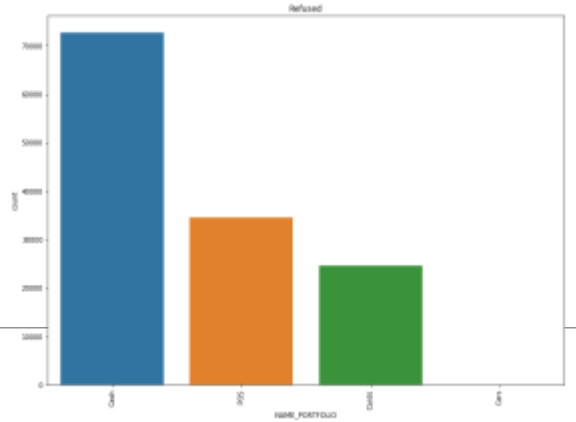


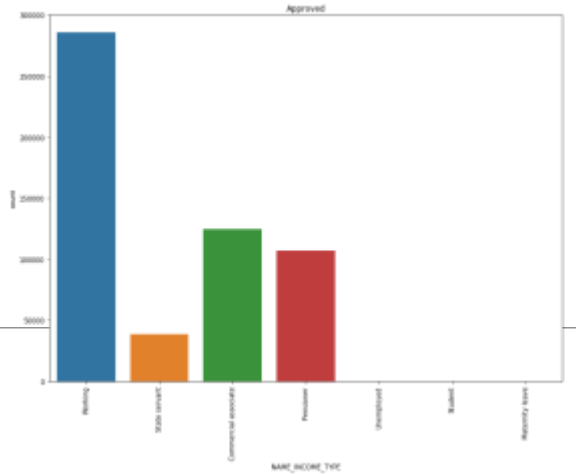


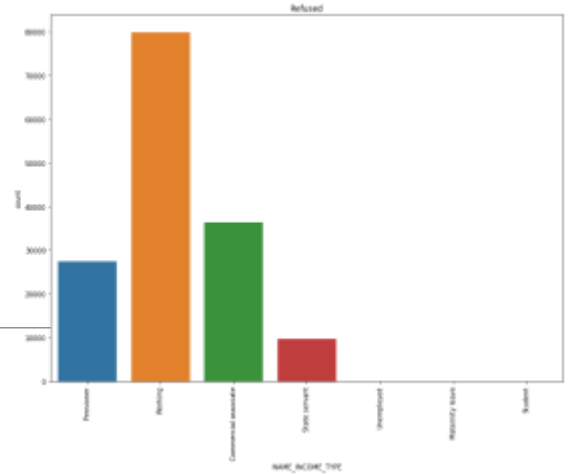


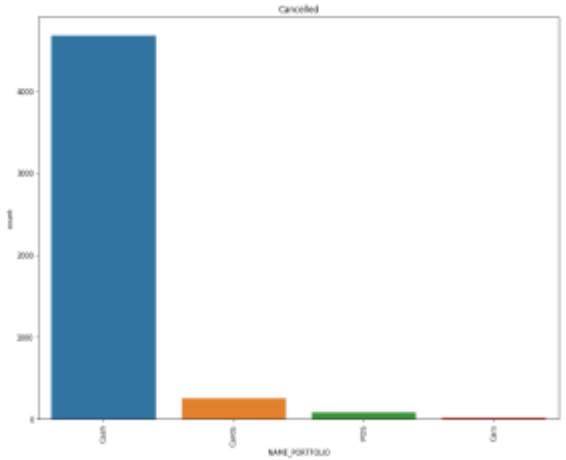
- Defaulters in refused load status is comparatively higher than rest
- Most approved loans are for repairs

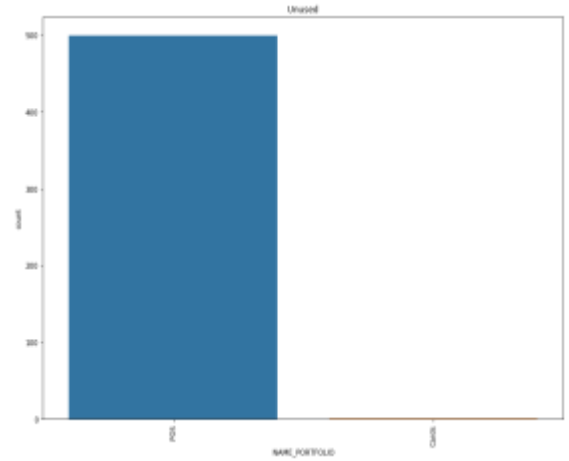


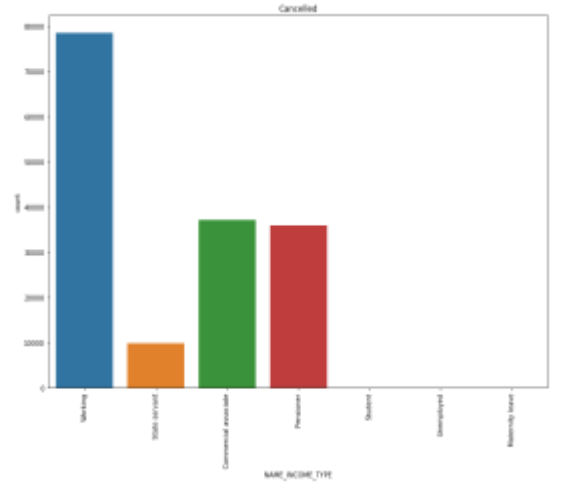


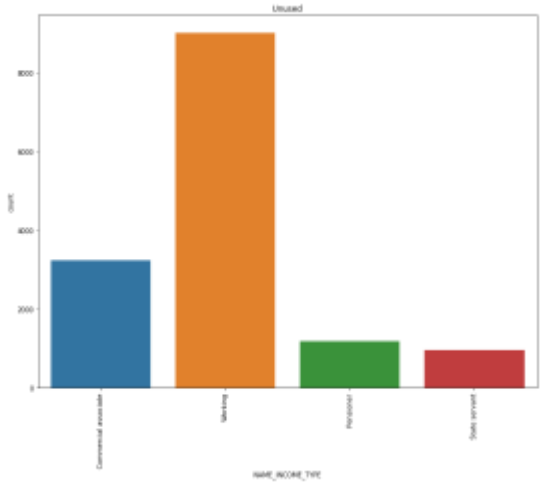












- Most approved loans are for pos and unused as well

-Working class have most approved loans and unused as well

Conclusion and Recommendations

- The proportion of defaulters is 8.0729% - The bank hence would have less NPA

- The bank lends more loans to females, the proportion of females in defaults is same as compared to males

- People with higher education tend to loan paybacks – hence a good customer base for Bank

- 'Repair' Purpose have a higher chance of loan repayment, but they also have highest chance of being a defaulter. So, bank should be more cautious in paying the loan for 'Repair’