

GROUP CASE STUDY 1:

ANALYZE PATTERNS IN LOAN APPLICATIONS DATA AND UNDERSTAND HOW CONSUMER ATTRIBUTES AND LOAN ATTRIBUTES INFLUENCE THE TENDENCY OF DEFAULT.

ABHINAV SHARMA

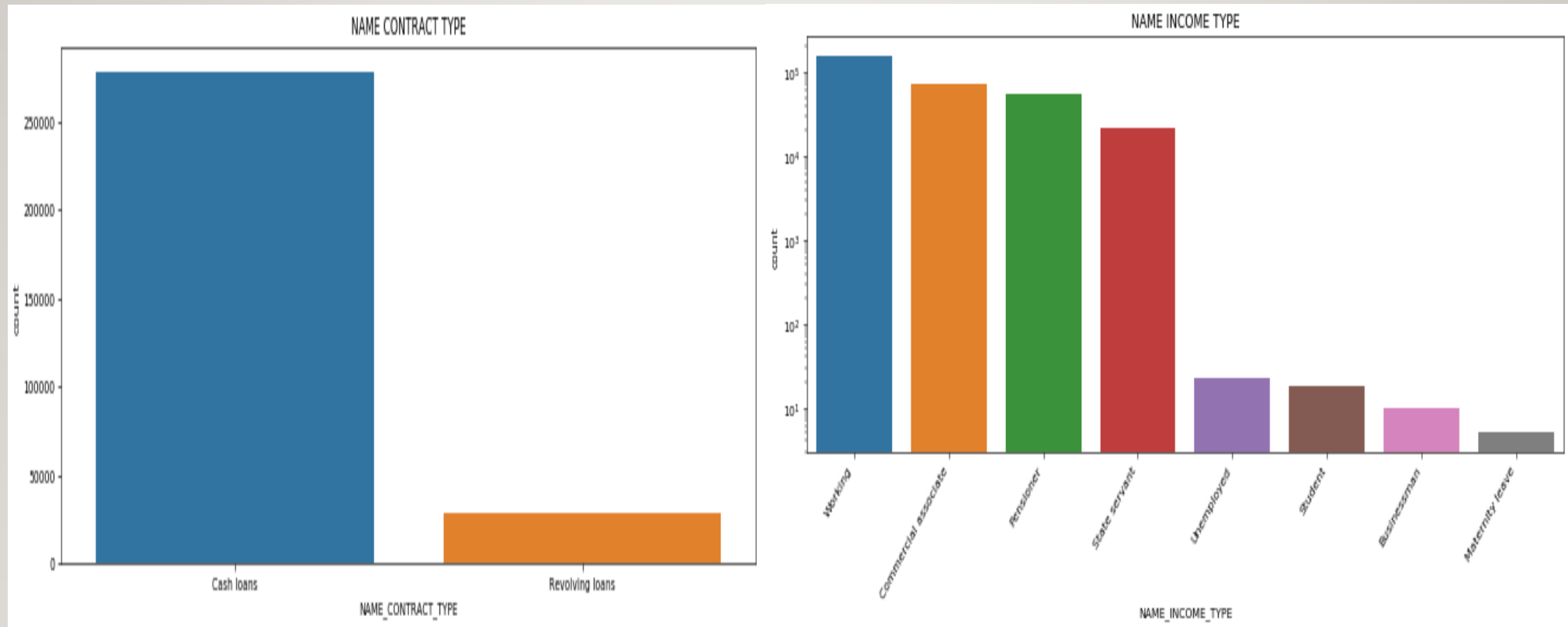
SHASHANK TIKKU

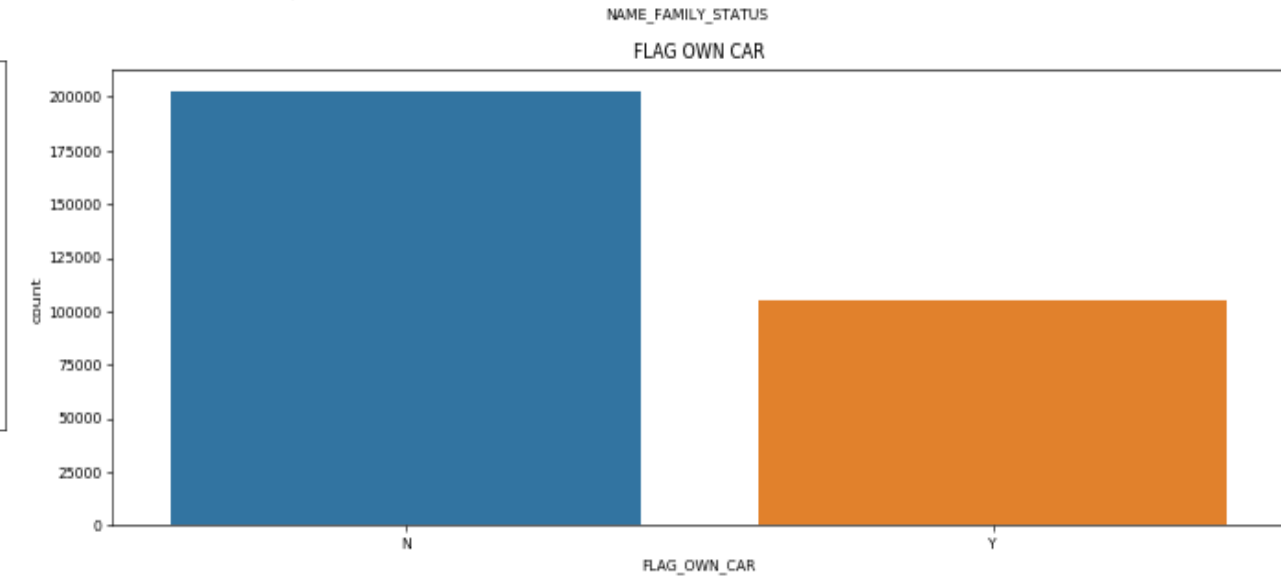
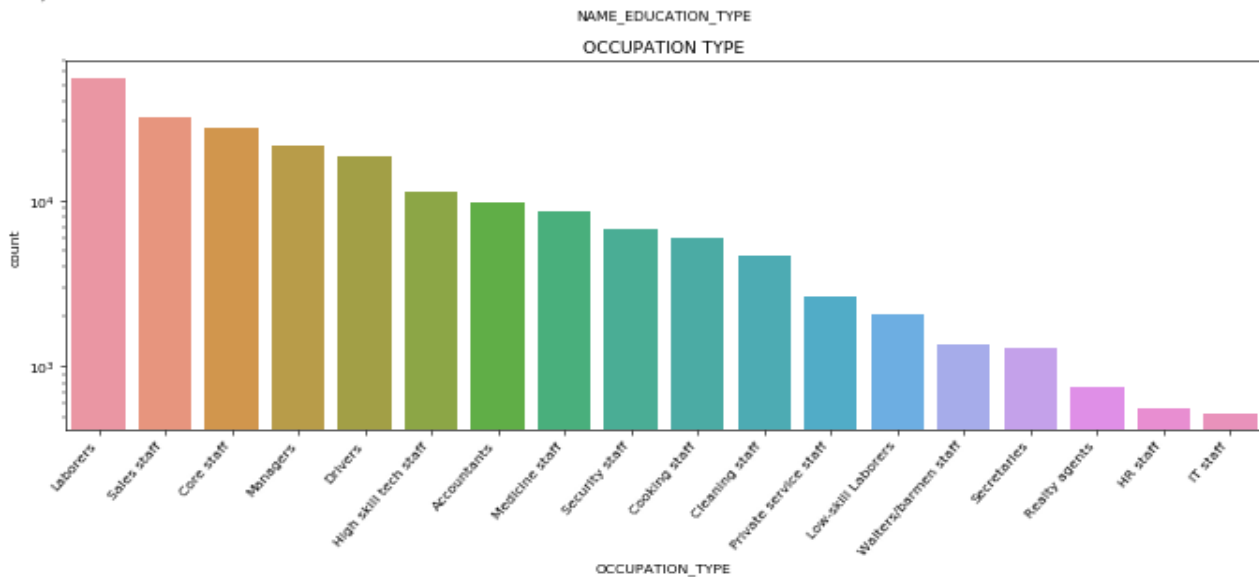
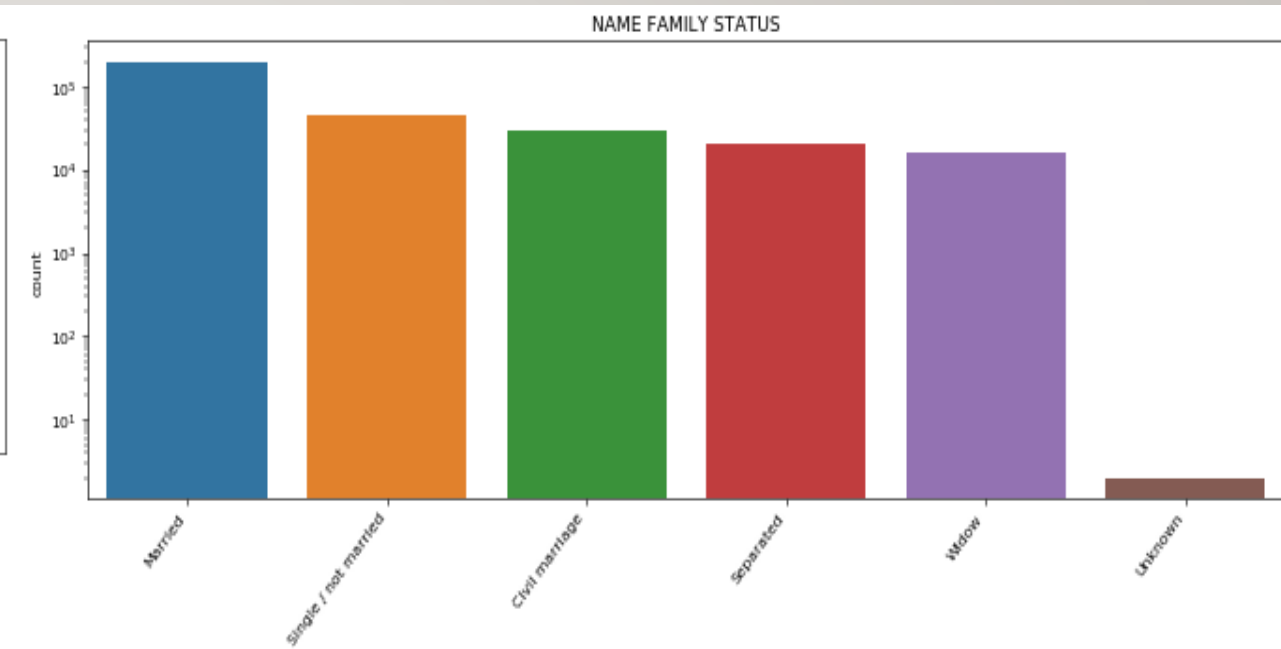
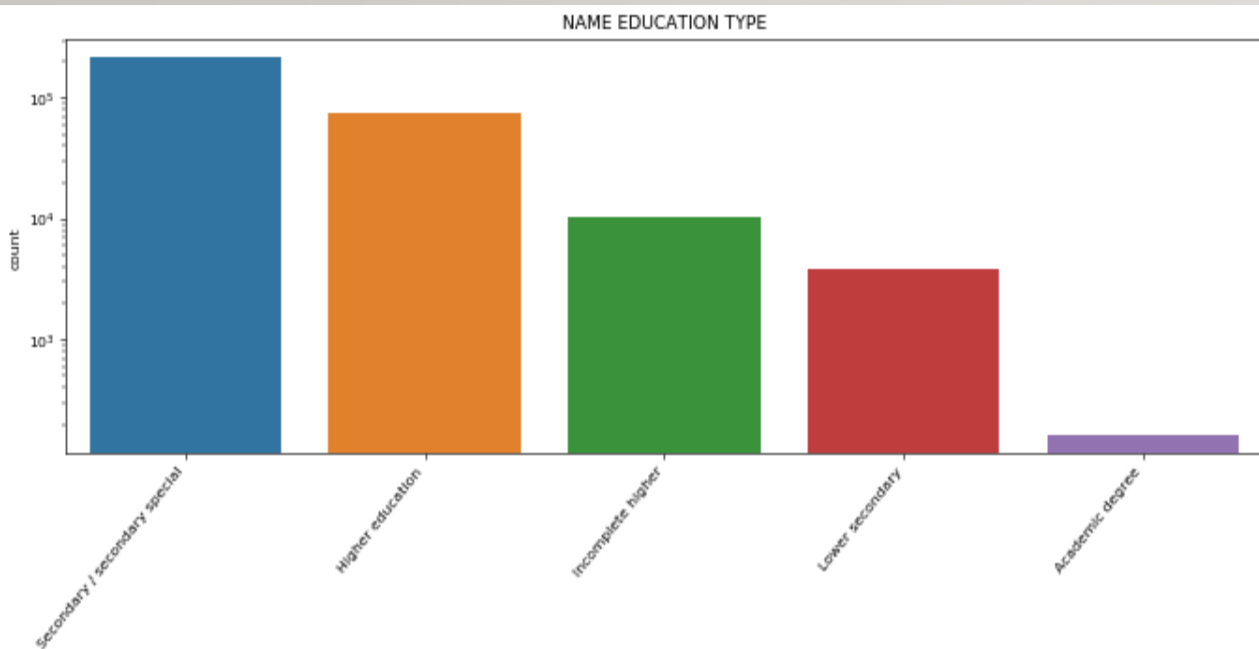


APPLICATION DATASET

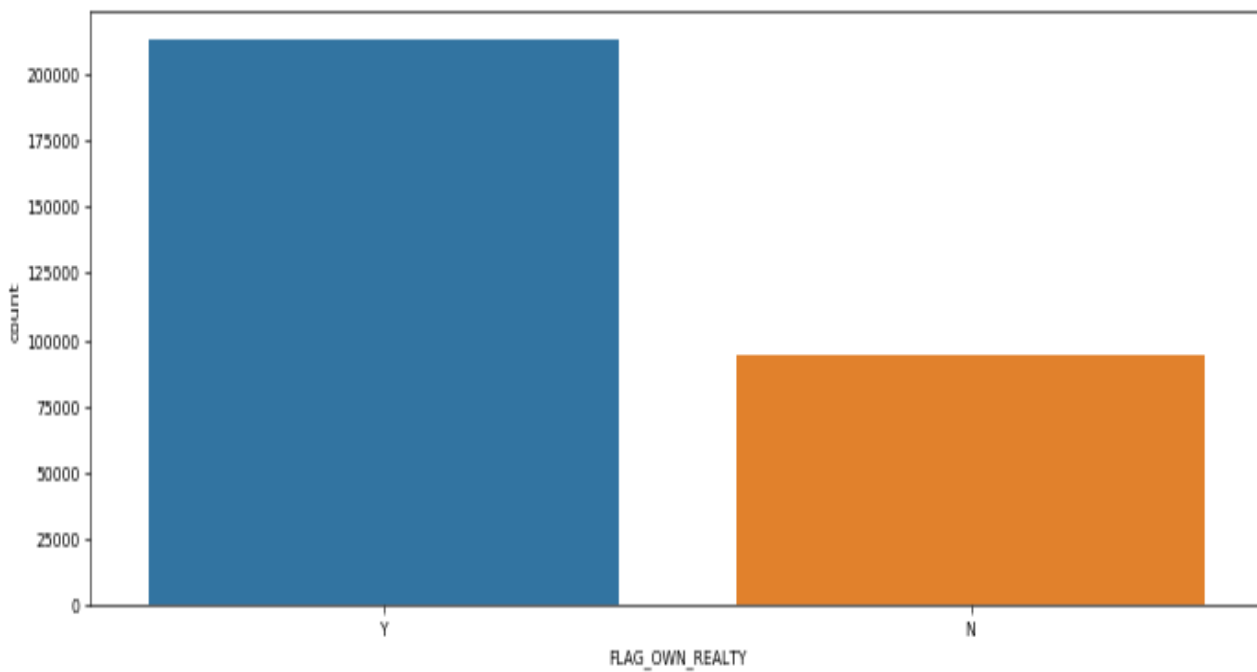
SOURCE: `application_data.csv`

UNIVARIATE ANALYSIS: CATEGORICAL VARIABLES

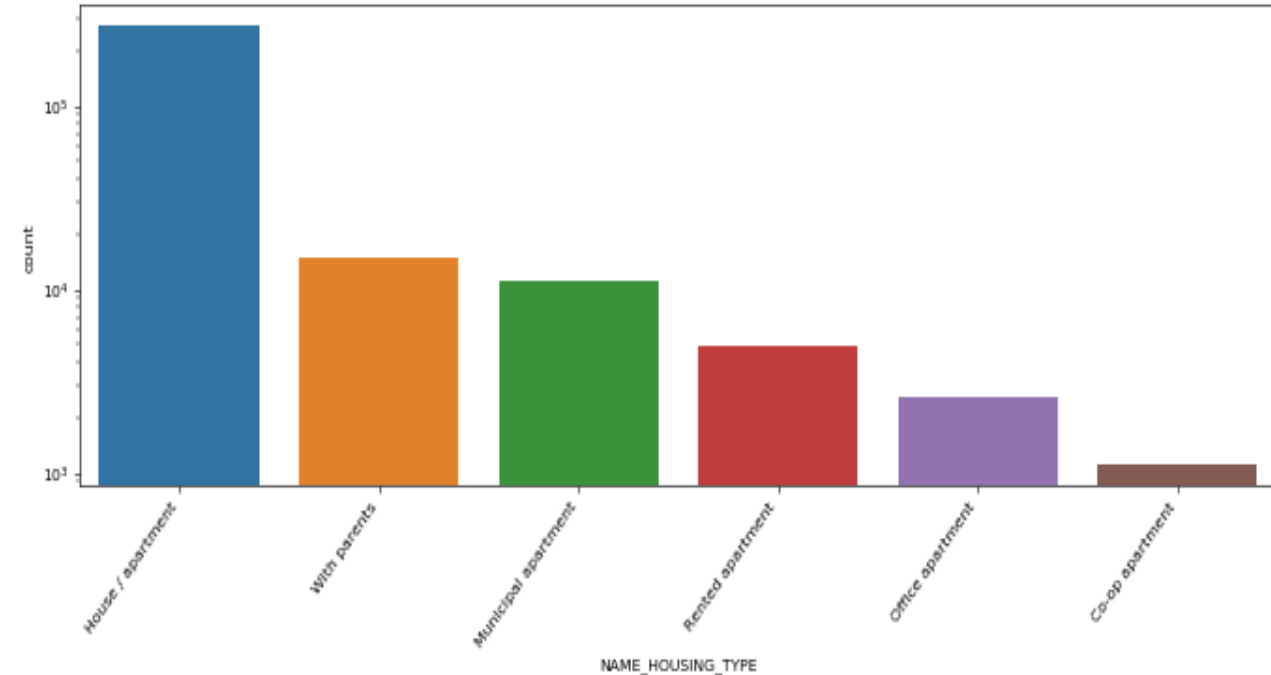




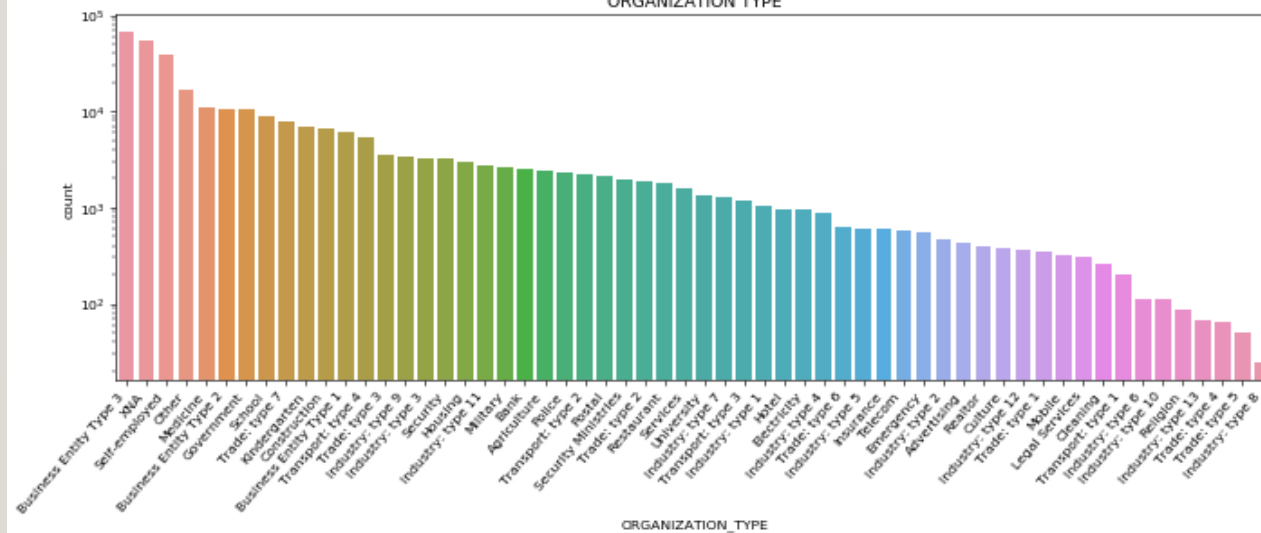
FLAG OWN REALTY



NAME HOUSING TYPE



ORGANIZATION TYPE

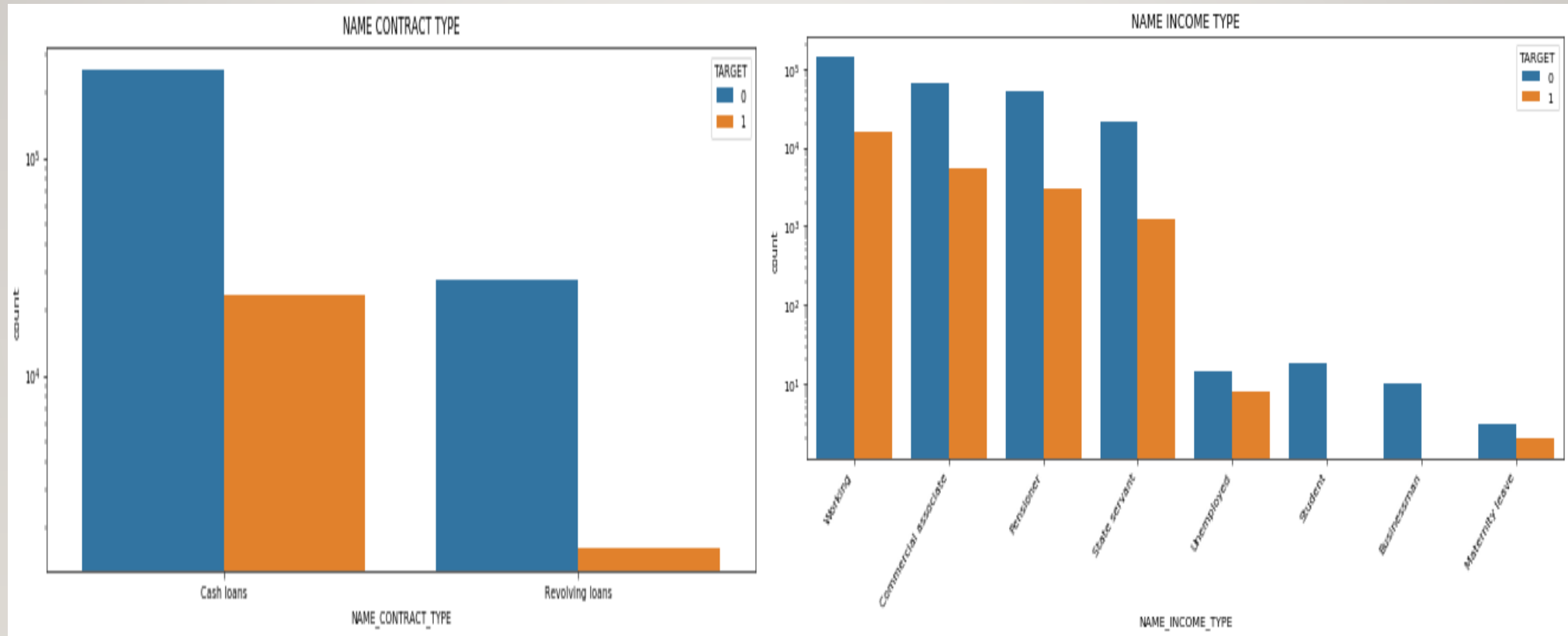


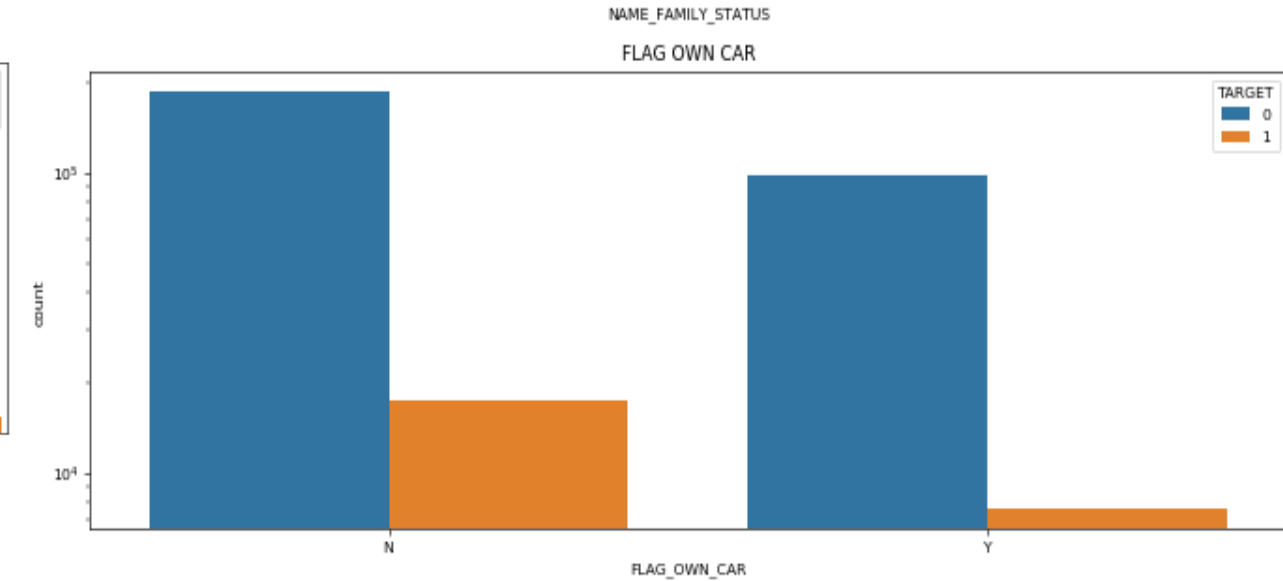
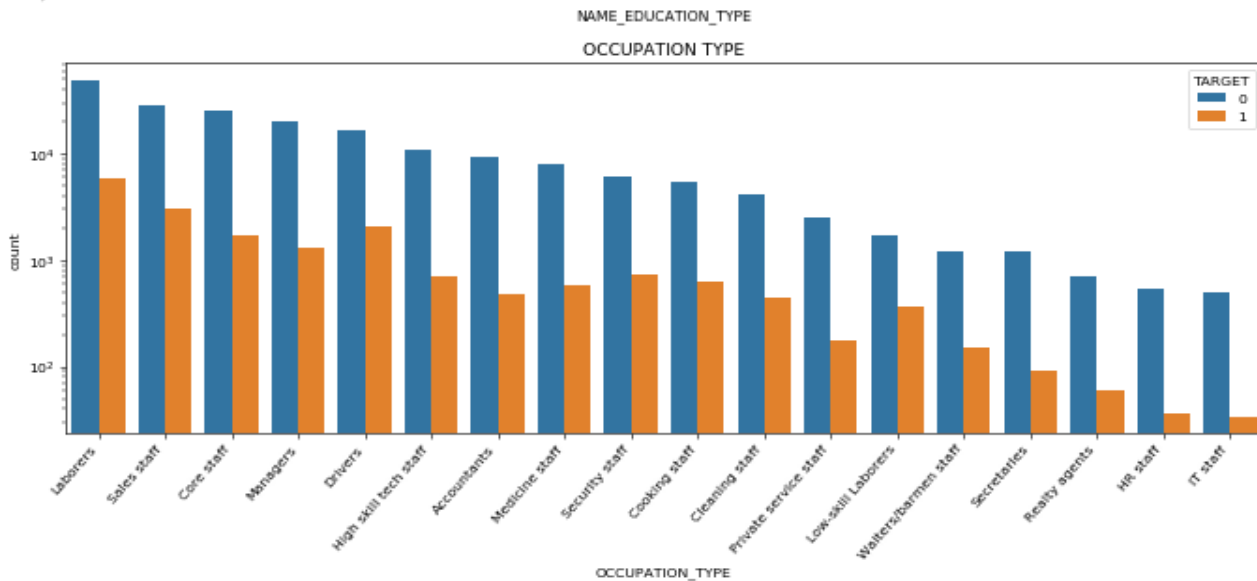
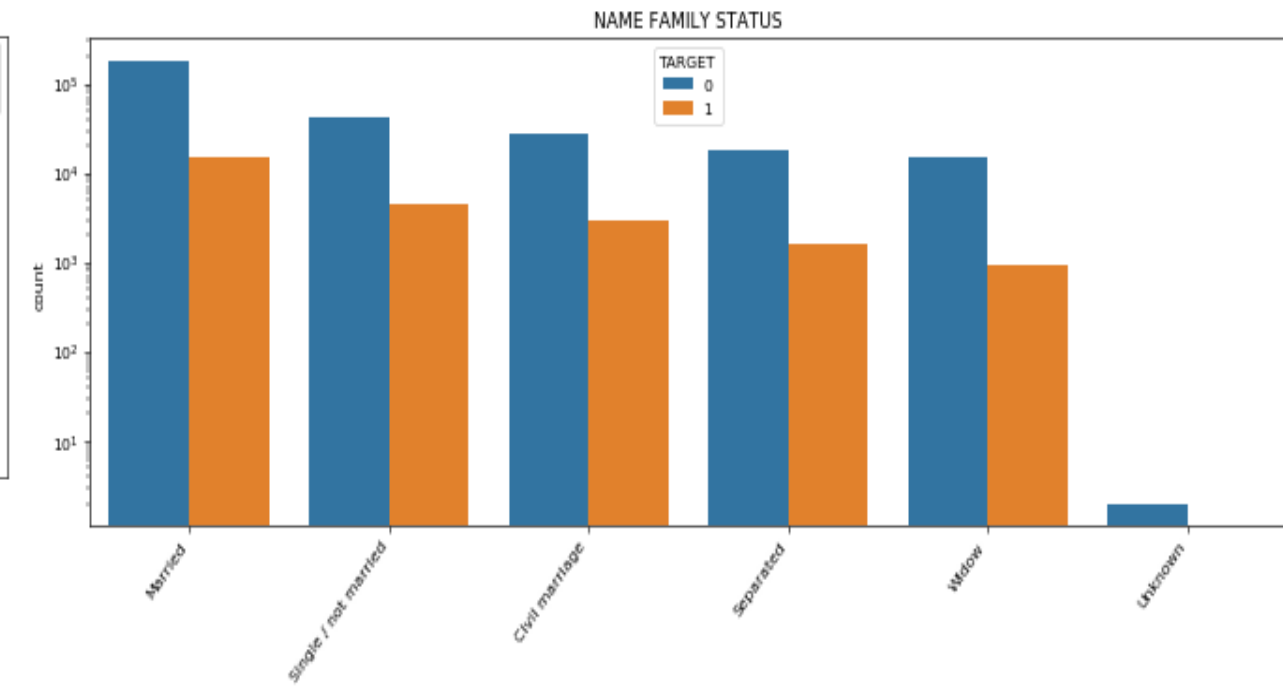
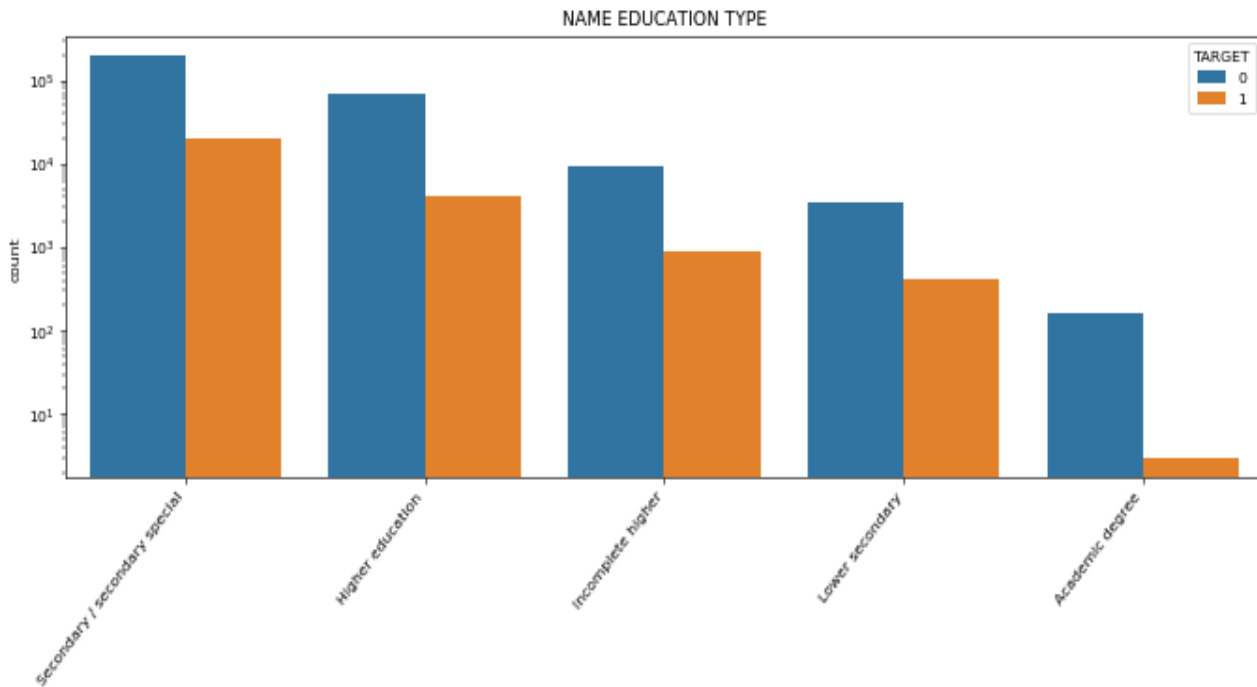
INSIGHTS

- Cash Loans are taken more than Revolving ones as they don't have any credit limit.
- Under Income Type, people who are Working, Commercial Associates, Pensioners and State Servants have taken the highest number of loans.
- People with secondary/higher secondary education background have taken the highest number of loans – may be for repaying their education loans.
- Married people took more loans; more liabilities, so more responsibilities.
- Laborers, Core staff, Sales Staff, Managers and Drivers took the highest number of loans which reflects the working class taking more loans. Surprisingly, the high skill technical staff took more loans than the low skilled laborers which can be seen for further skill enhancement.
- The ones owning a car have taken fewer loans than those who don't whereas the ones owning property have taken more loans because:
 1. We get income tax rebates on housing loans.
 2. Getting a loan showing house papers is much easier because you have an appreciating asset compared to depreciating assets like a car.

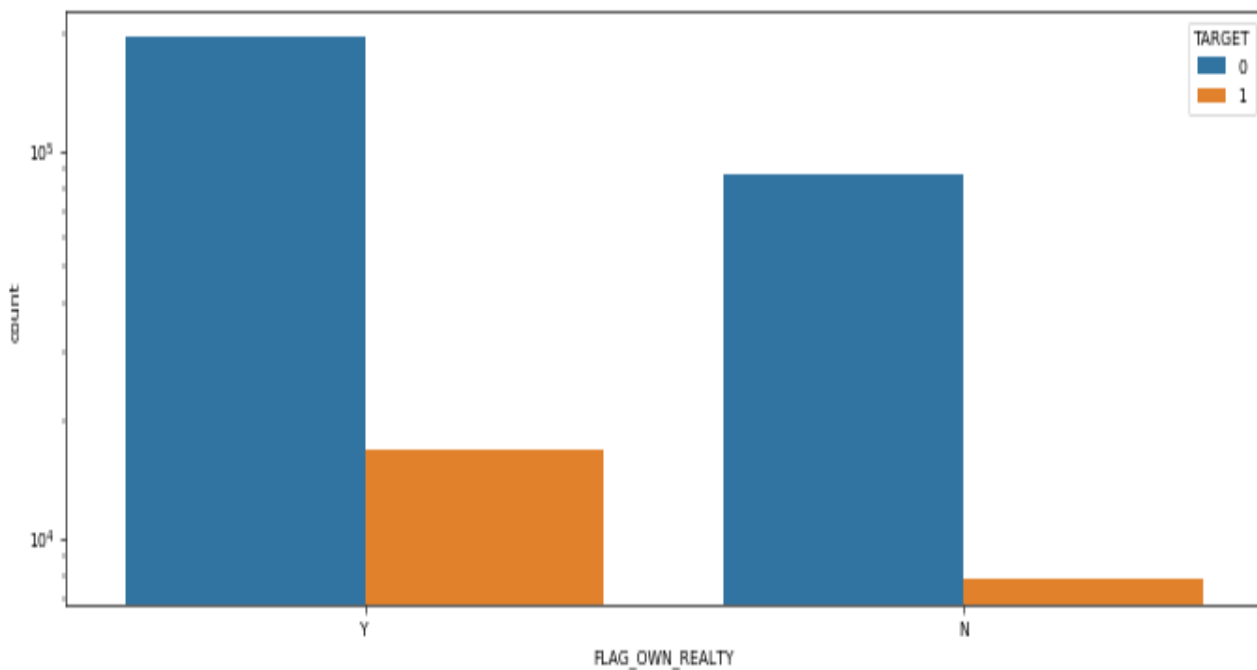
BIVARIATE ANALYSIS w.r.t TARGET:

CATEGORICAL VARIABLES

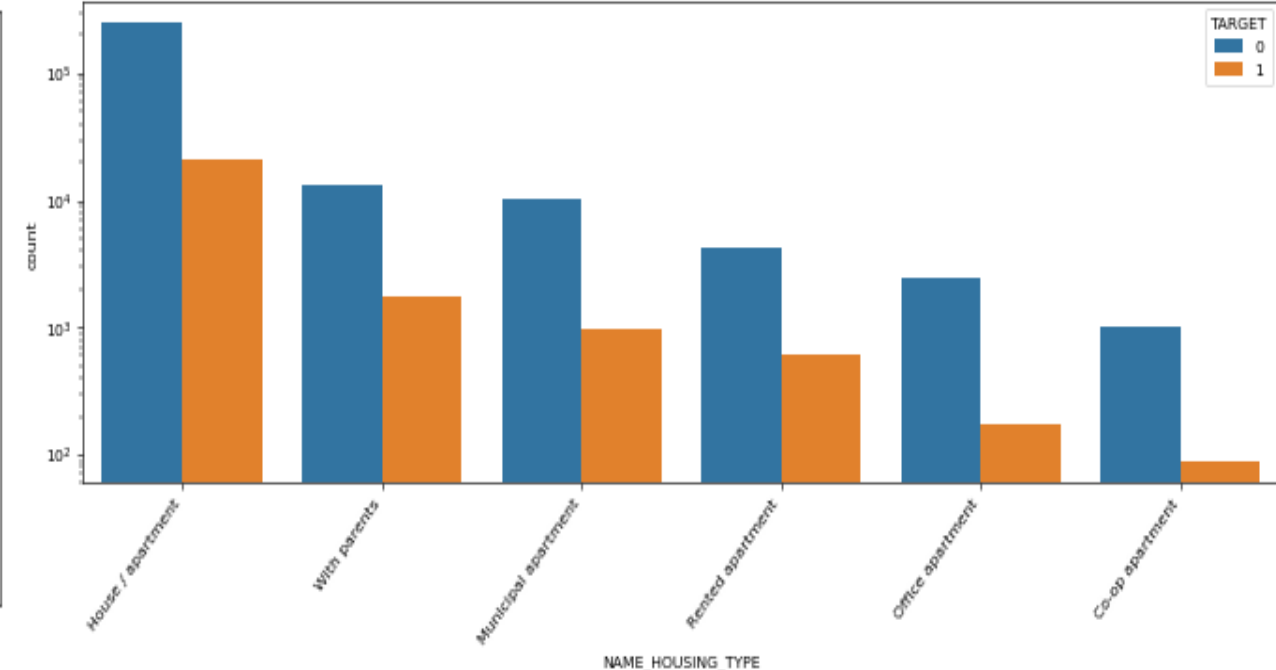




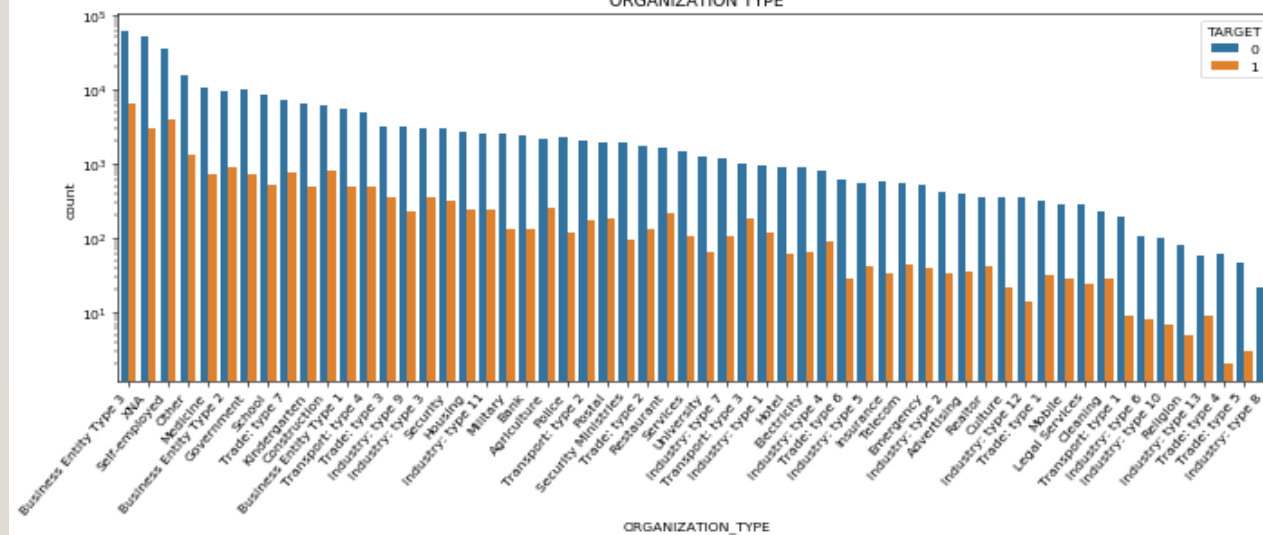
FLAG_OWN_REALTY



NAME_HOUSING_TYPE



ORGANIZATION_TYPE

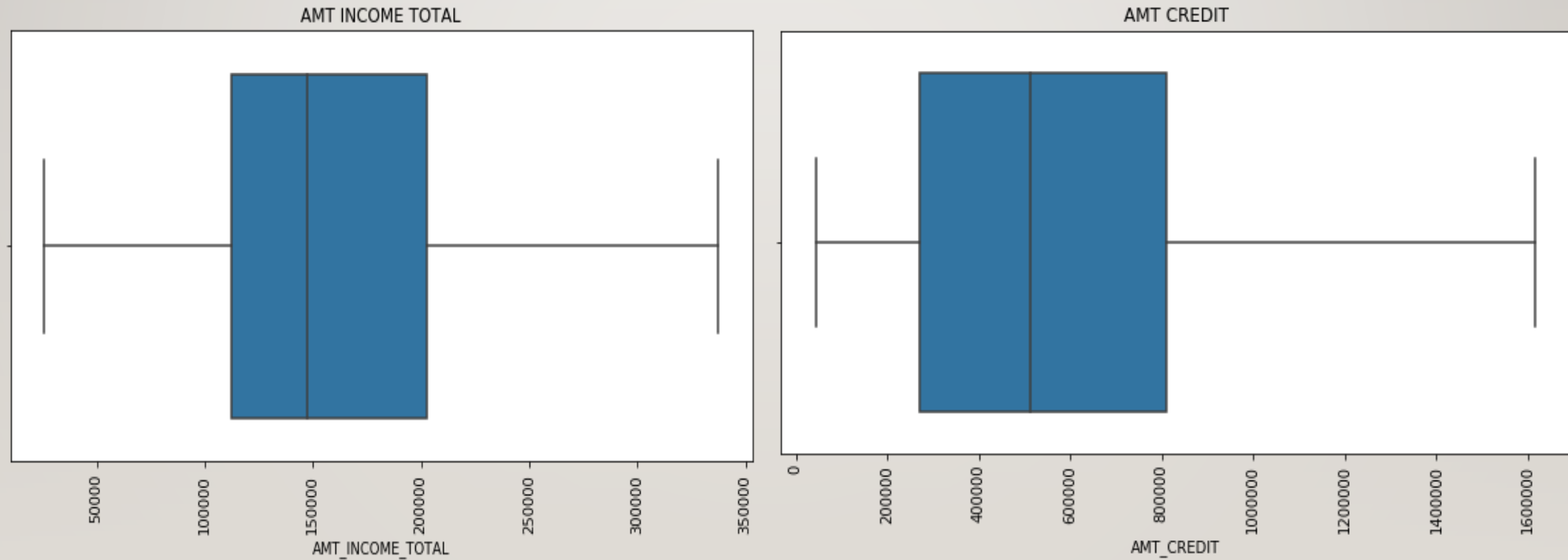


INSIGHTS

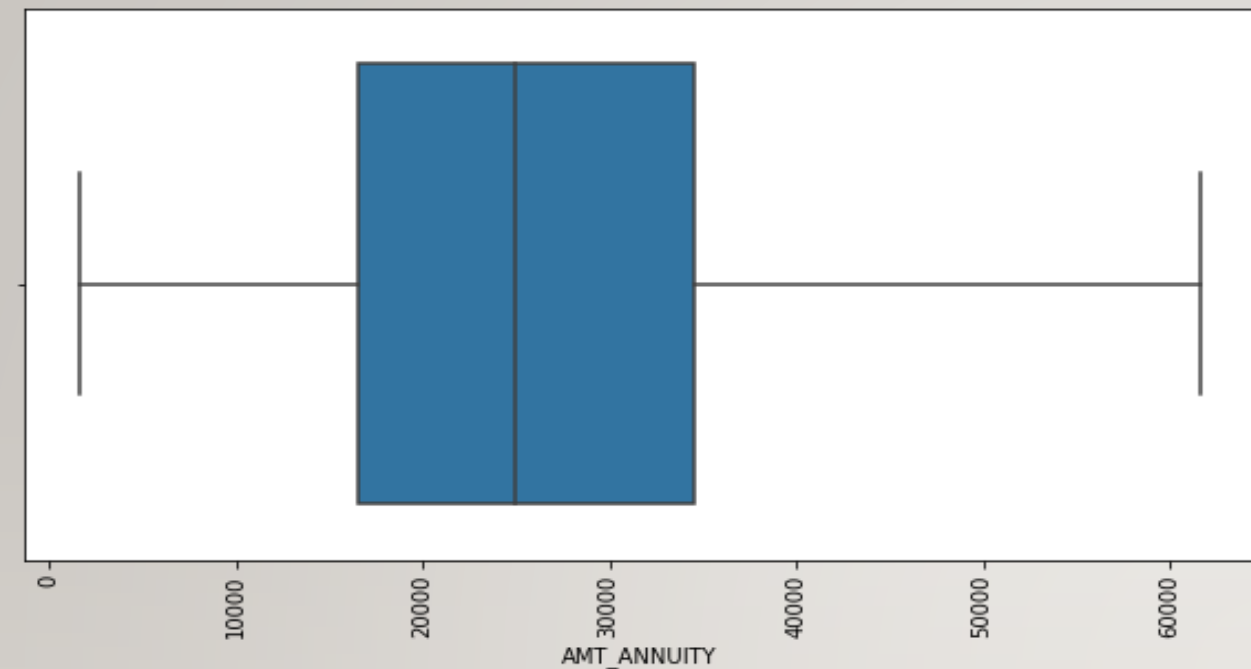
- Defaulters follow the same trend as income types like working professionals, Commercial associates, Pensioners and then State Servants. People with maternity leaves have also taken considerable loans because of medical expenses.
- Persons with Secondary and Higher secondary education background are the most frequent defaulters. The same holds true for Married people.
- Laborers, Sales Staff, Drivers, Security Staff are the most frequent defaulters - surprisingly cooking staff and low skilled laborers have also shown a spike in defaulting.
- Conclusion: Laborers (working on fields) and Sales Staff come with higher education backgrounds but still end up being defaulters as they apply for loans showing their house papers to get income tax rebates and credit loans but are not able to repay it as they are married and have various liabilities. This in turn traps them in a vicious circle of loan from which they cannot escape.

UNIVARIATE ANALYSIS: CONTINUOUS VARIABLES

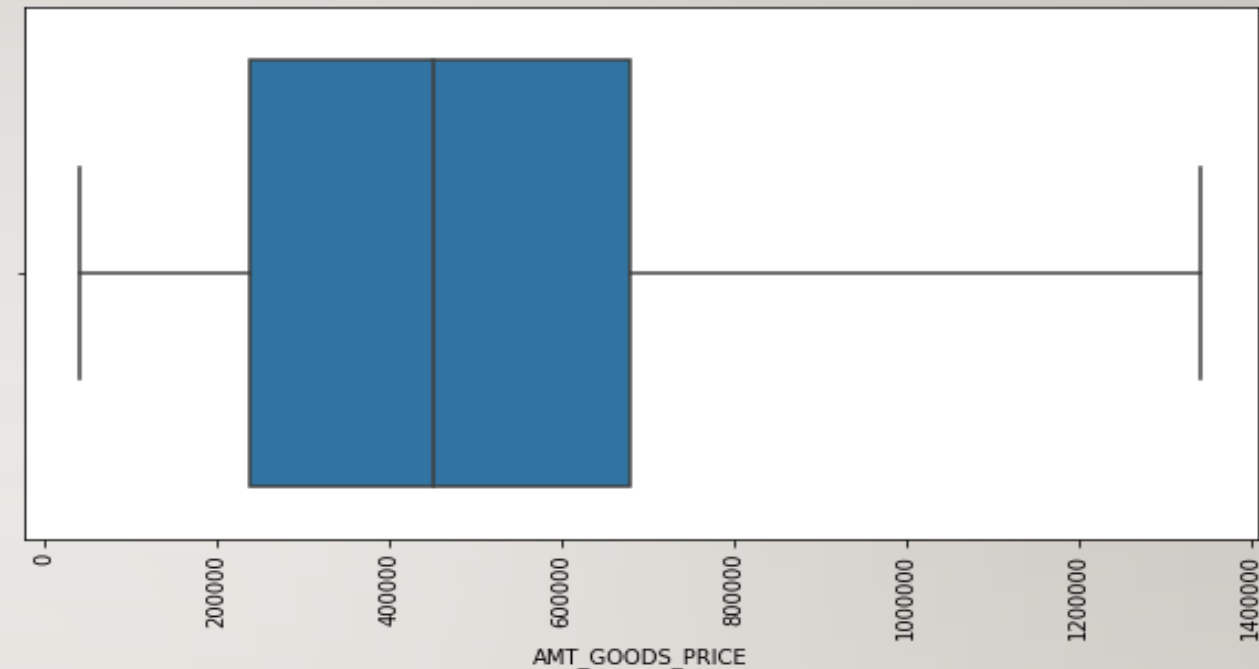
Using boxplots after capping outliers via the IQR method.



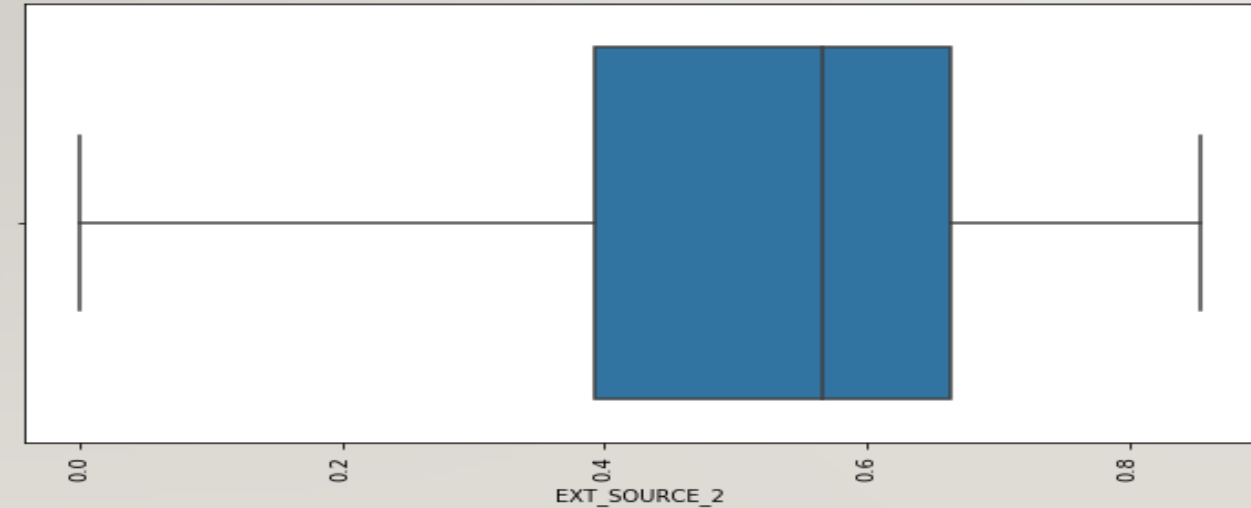
AMT ANNUITY



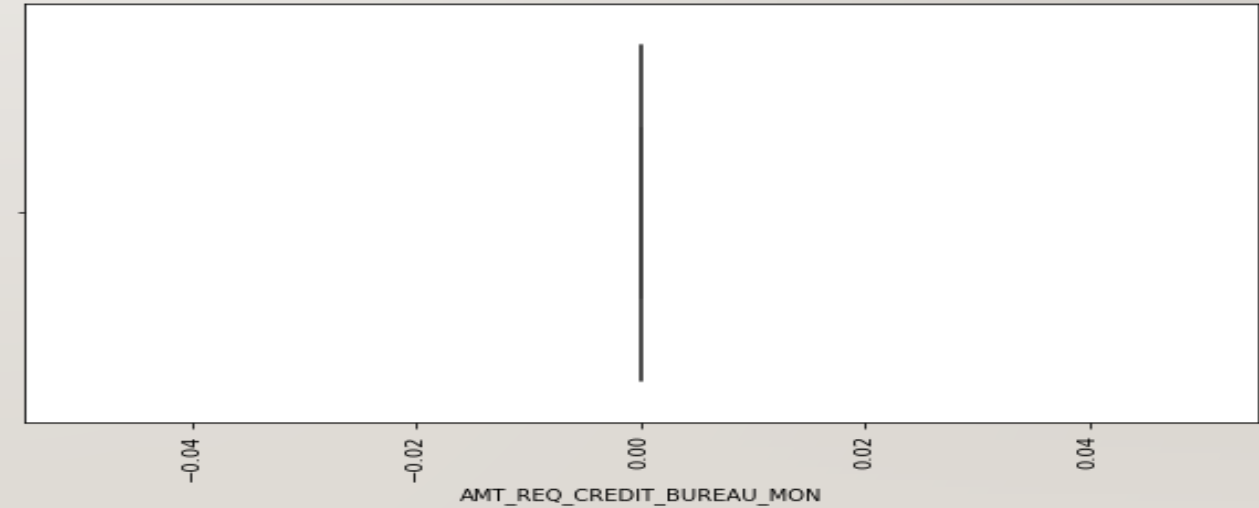
AMT GOODS PRICE



EXT_SOURCE_2

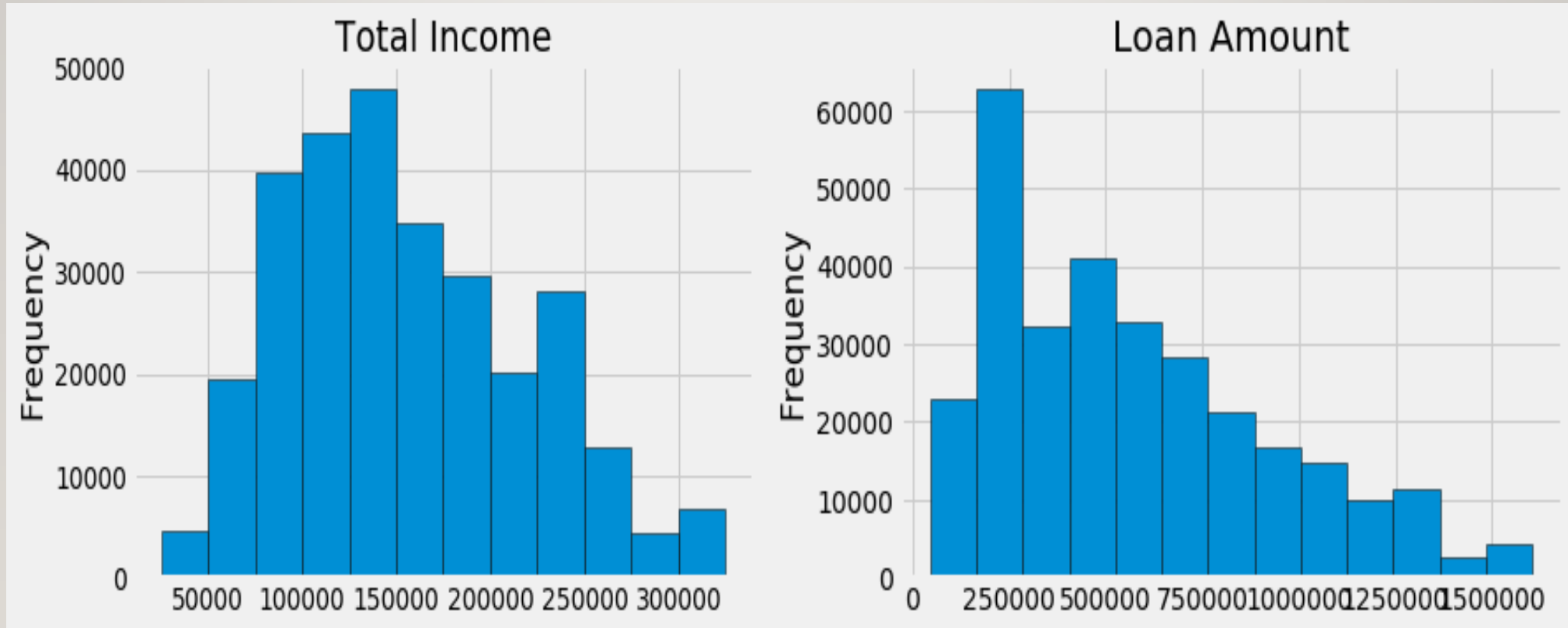


AMT_REQ_CREDIT_BUREAU_MON

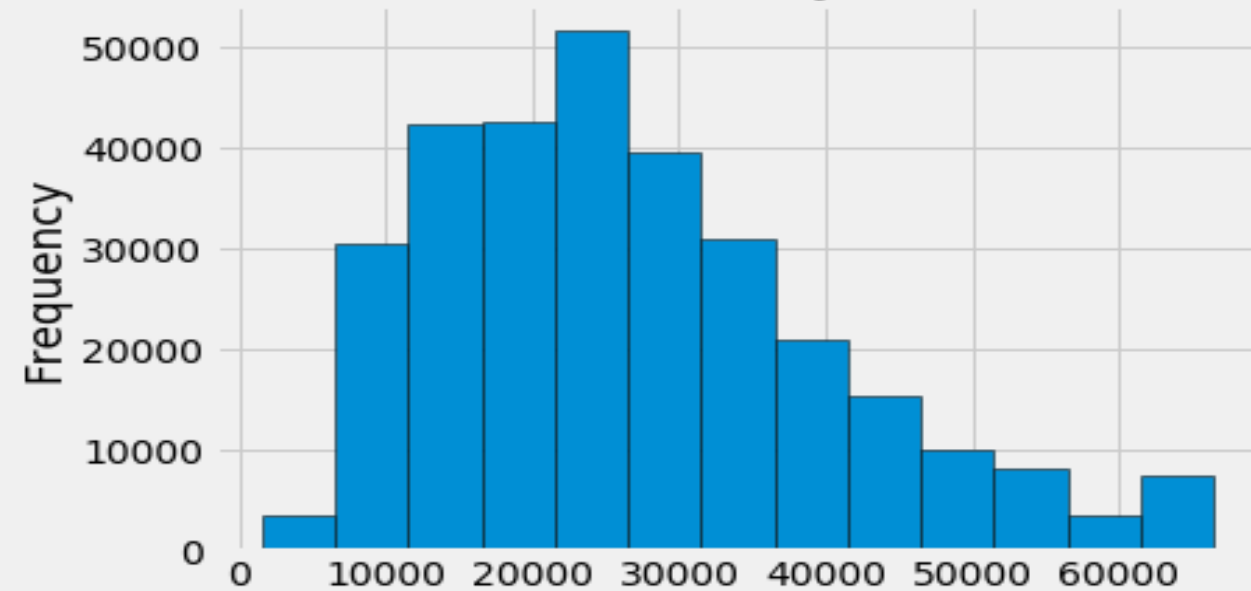


SEGMENTED UNIVARIATE ANALYSIS:

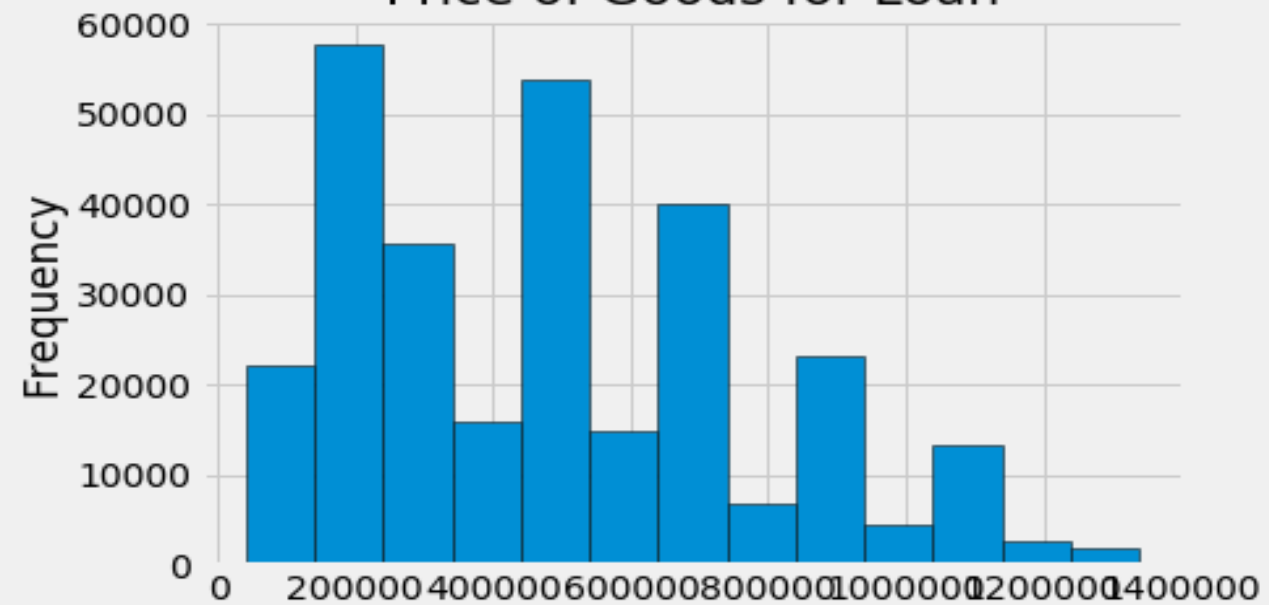
CONTINUOUS VARIABLES



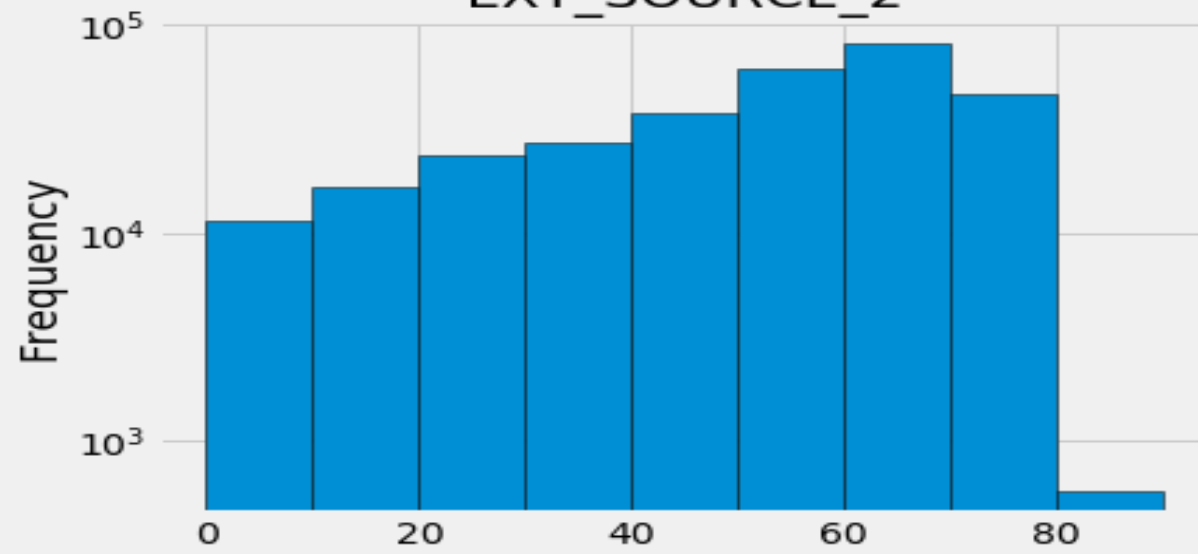
Annuity



Price of Goods for Loan



EXT_SOURCE_2

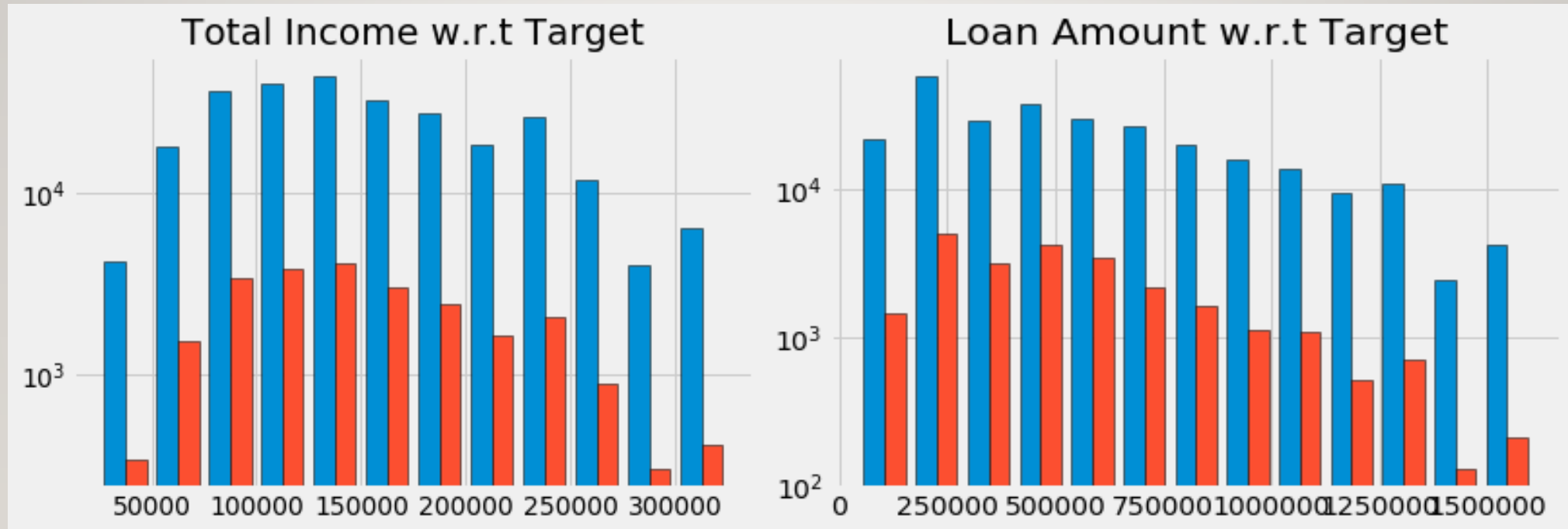


INSIGHTS

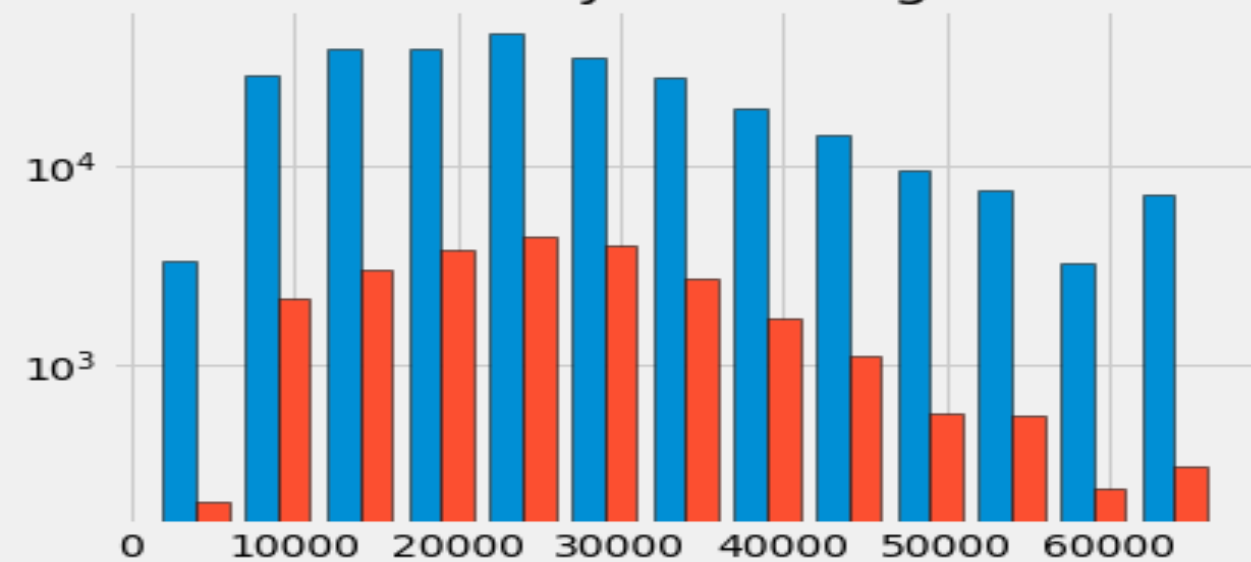
- People with total annual income between 100,000-150,000 have applied for the greatest number of loans.
- Loan amounts between 250,000-500,000 have been applied for the most.
- Annuity between 20,000-30,000 is observed having the highest number of loan applications.
- The price of goods for which loans have been applied showed a discrete trend and goods with prices close to 200,000, 500,000, 750,000 have the highest number of loans meaning people are applying for all sorts of loans i.e. low as well as high value loans.

BIVARIATE ANALYSIS w.r.t TARGET:

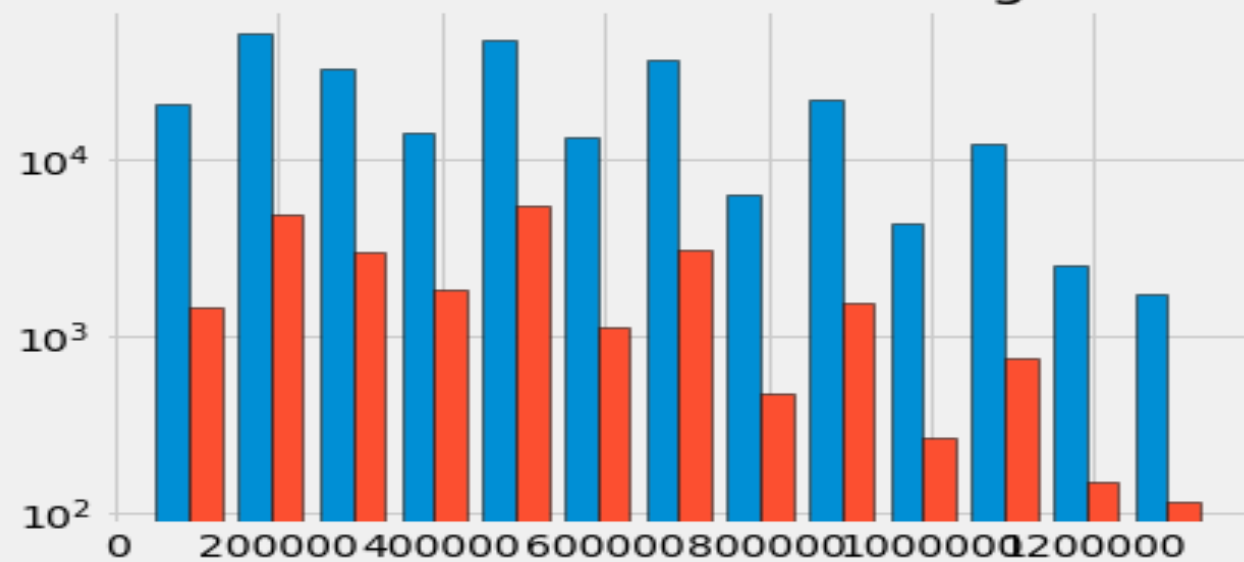
CONTINUOUS VARIABLES



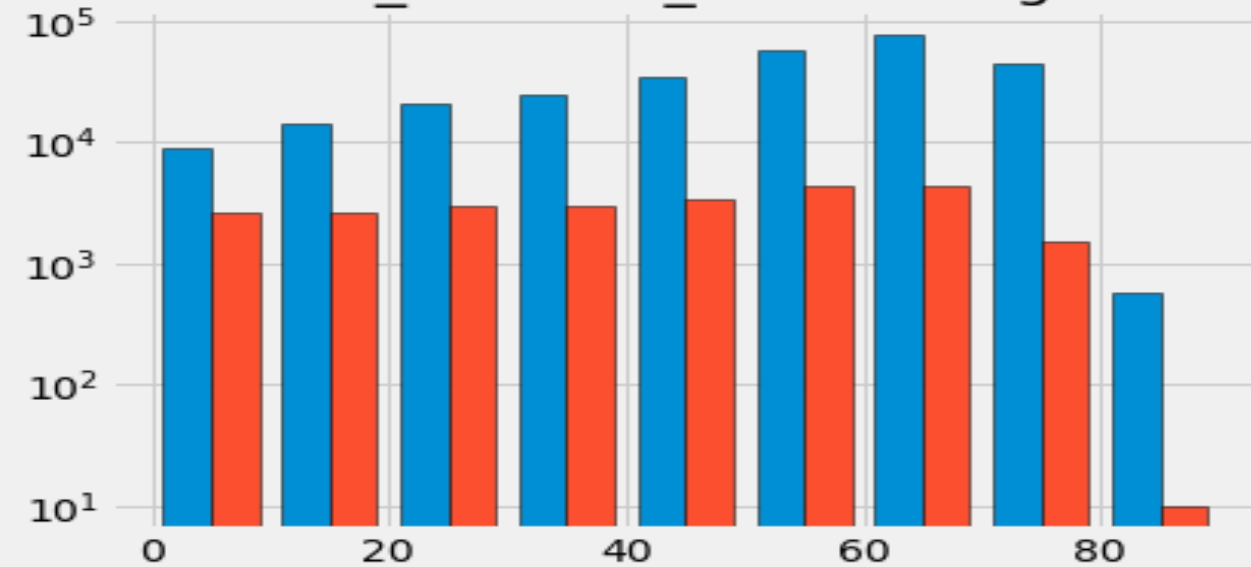
Annuity w.r.t Target



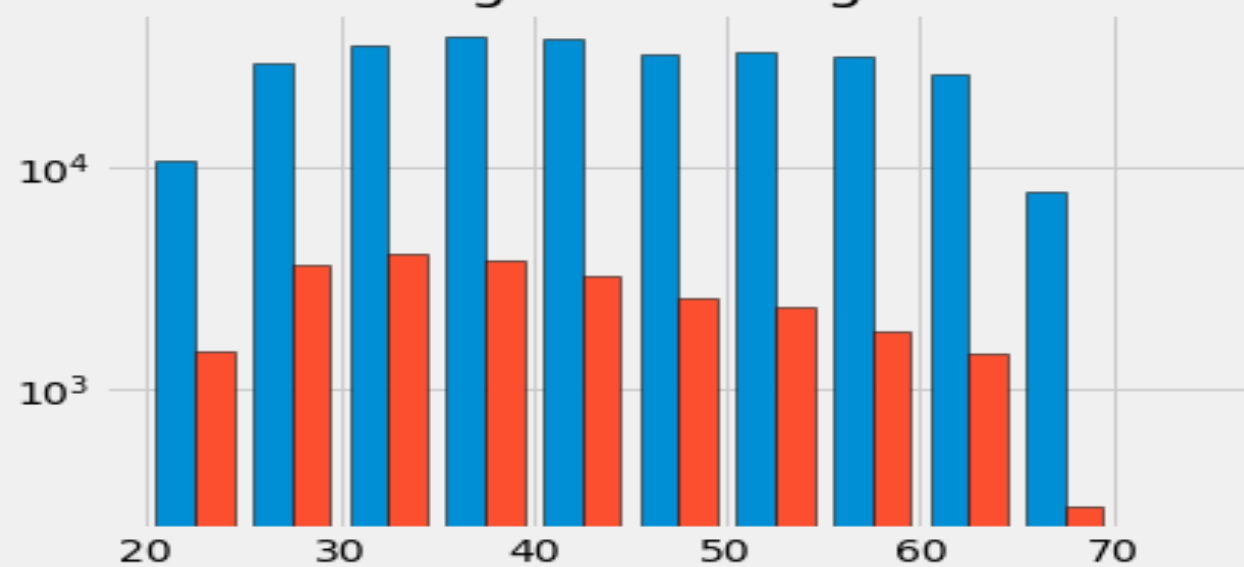
Price of Goods w.r.t Target



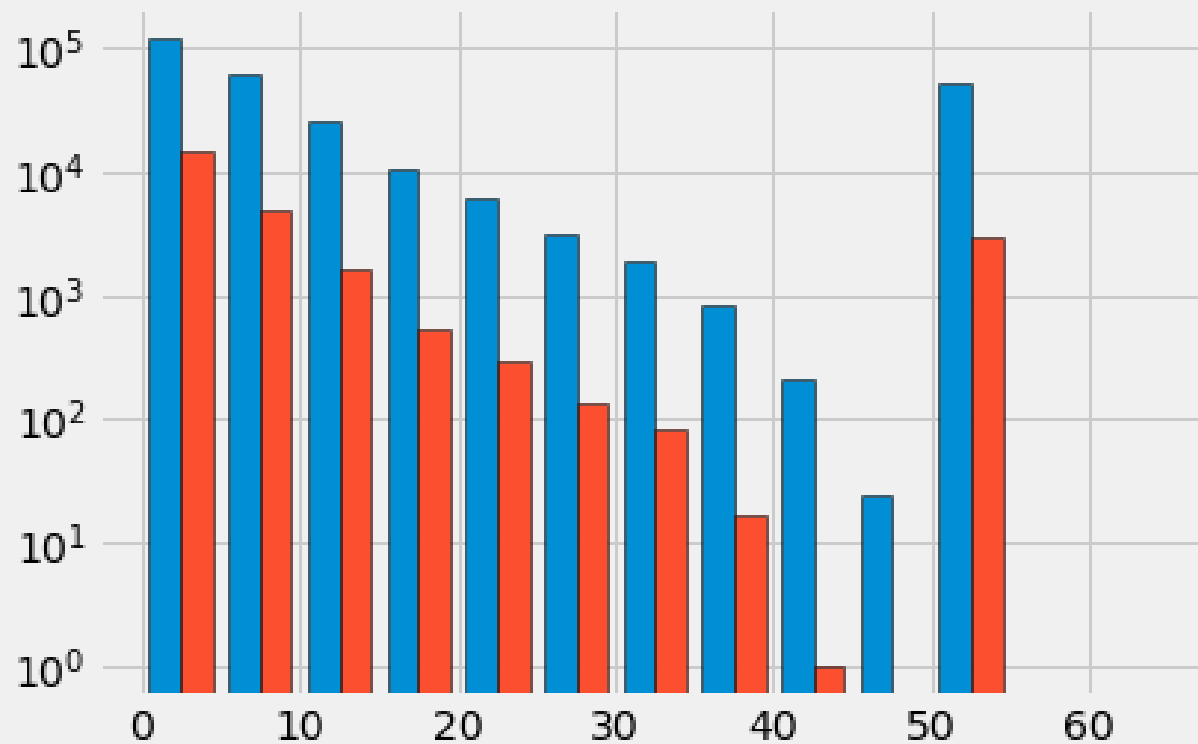
EXT_SOURCE_2 w.r.t Target



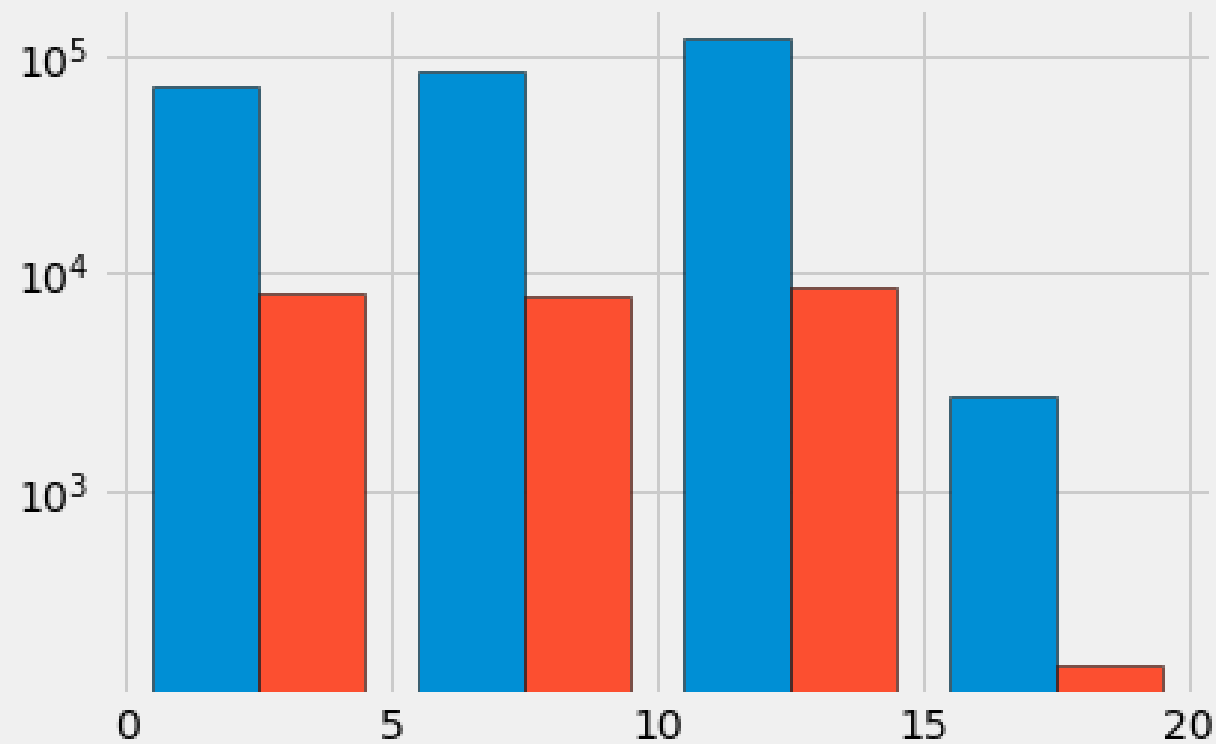
Age w.r.t Target



Years Employed w.r.t Target



Years Since ID Change w.r.t Target

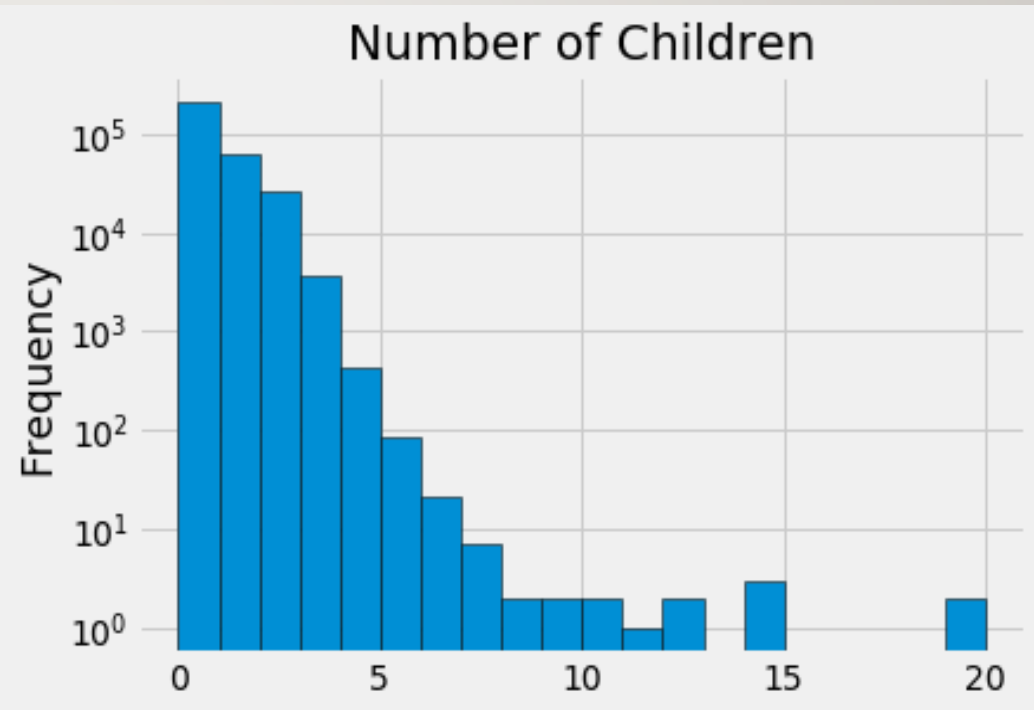
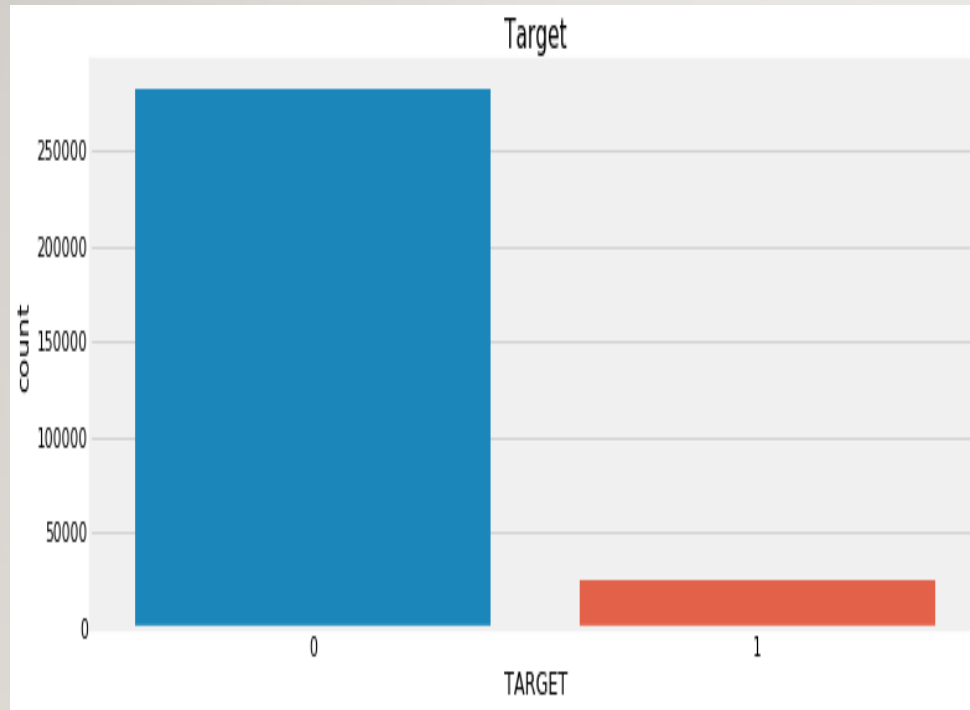


INSIGHTS

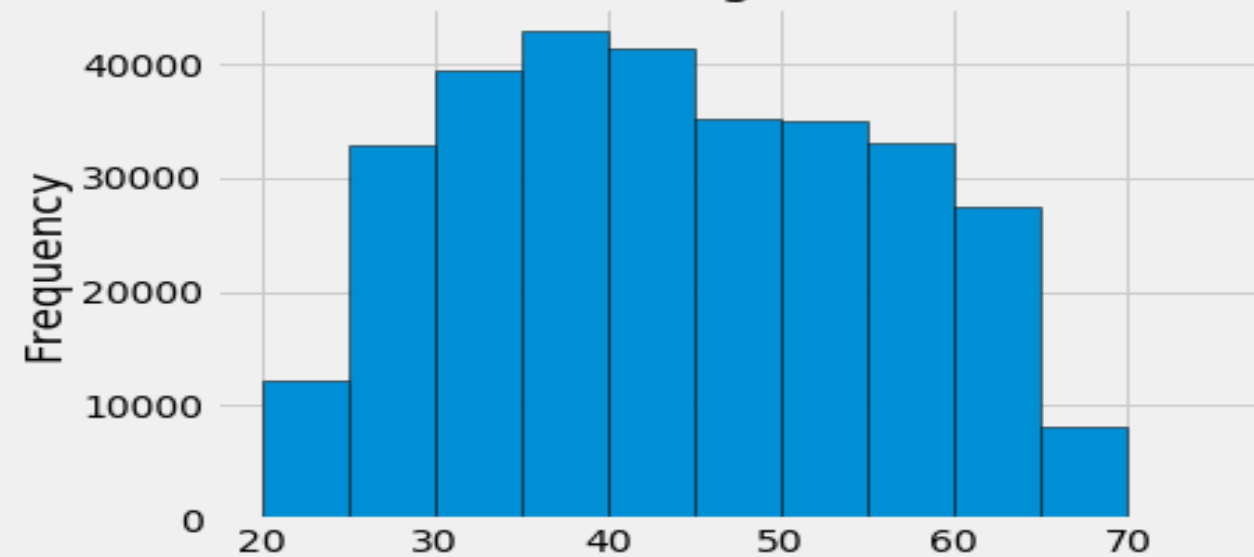
- As the income increases from 100,000 to 150,000 per annum , the number of defaulters also increase because these applicants become eligible for loan but repaying capability is still less.
- As the loan amount ranges from 250,000 to 500,000 the number of defaulters increase and then decrease stating mostly the applicants who apply for car loans and are unable to pay it as they belong to the low-income salary group.
- People between the age of 35-45 tend to be the biggest defaulters as we've mentioned earlier that they want to settle down with their families. repaying their housing loans or car loans.
- People with more family members are defaulting less and those with more family members are defaulting more .

UNIVARIATE ANALYSIS:

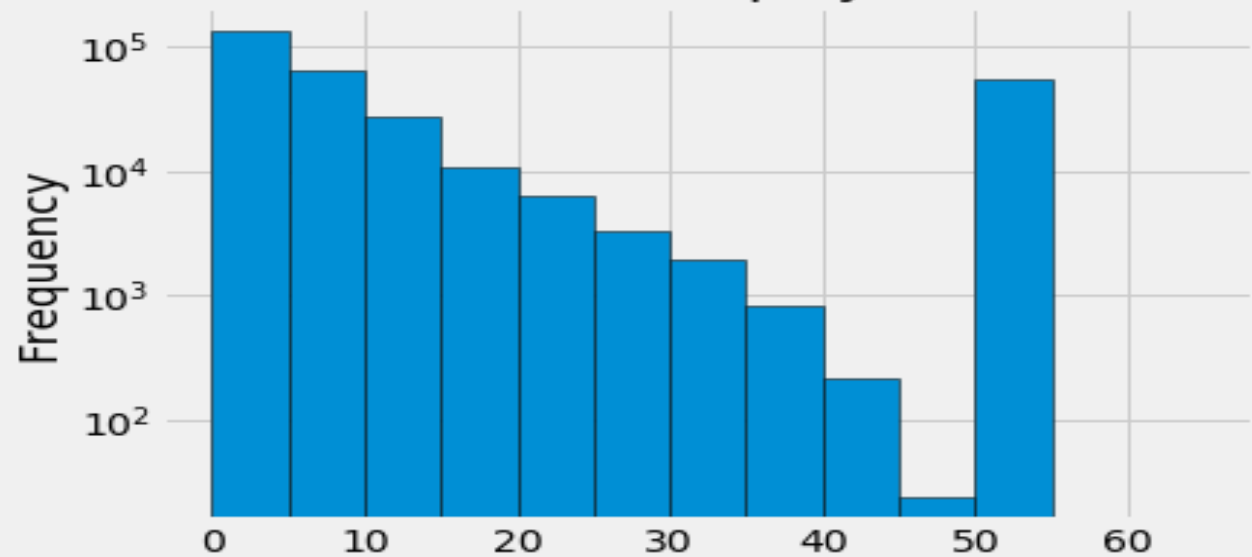
DISCRETE VARIABLES



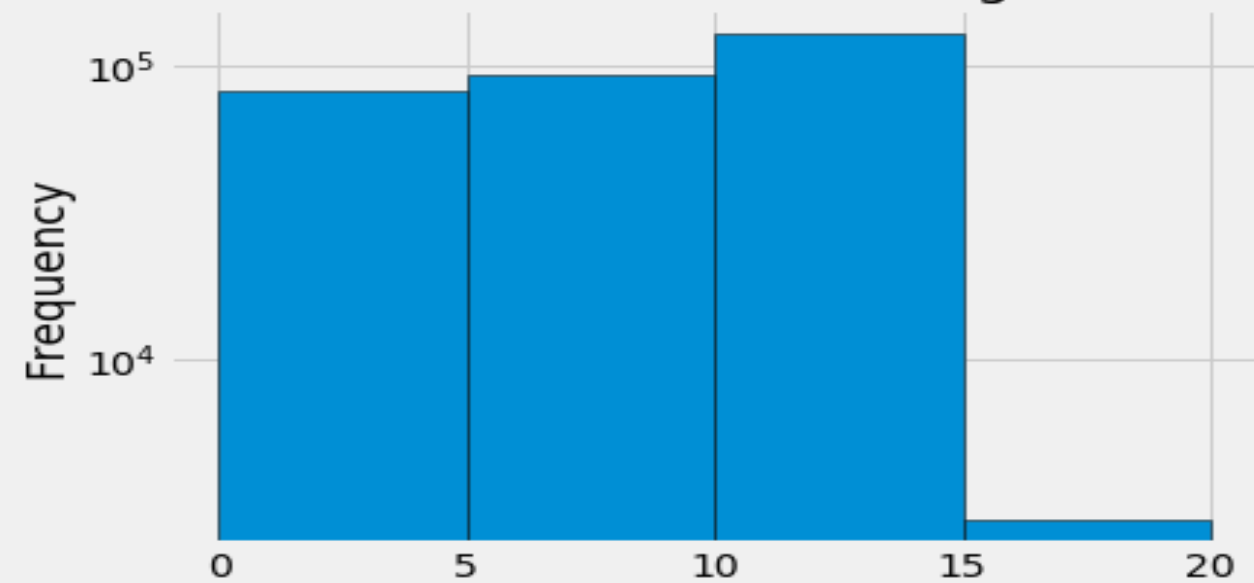
Age



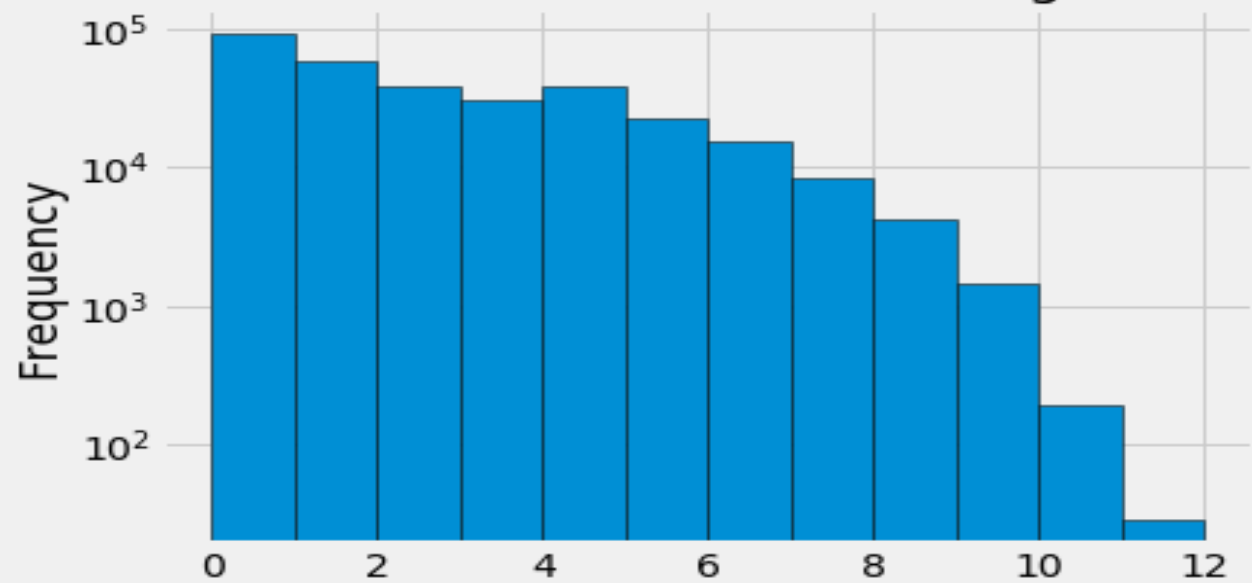
Years Employed

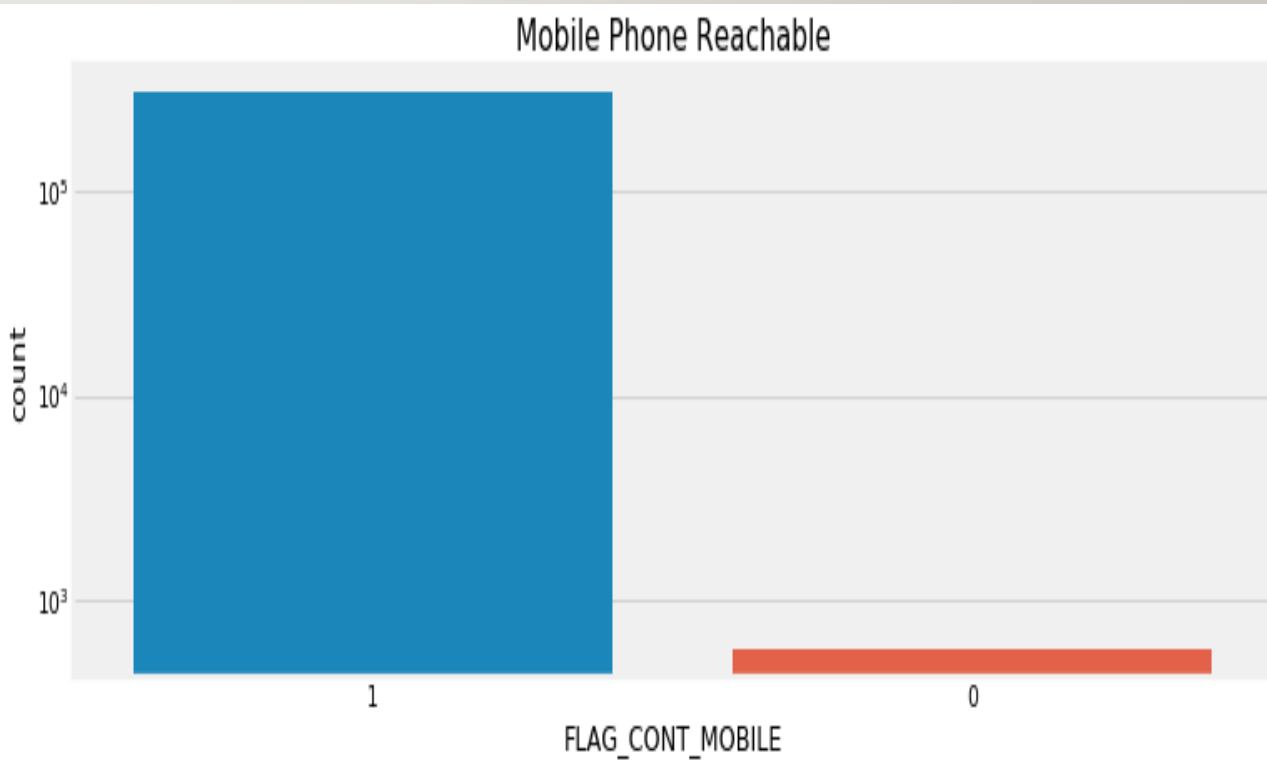
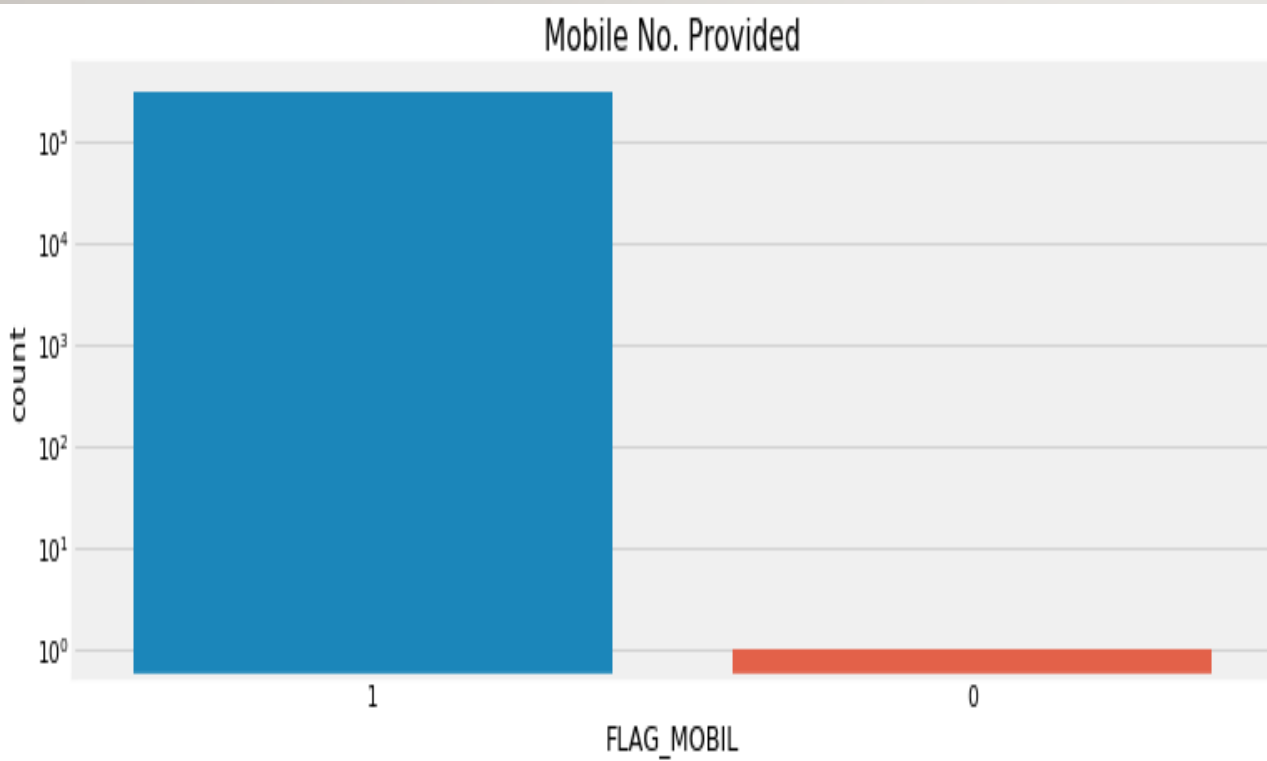


Years Since ID Change

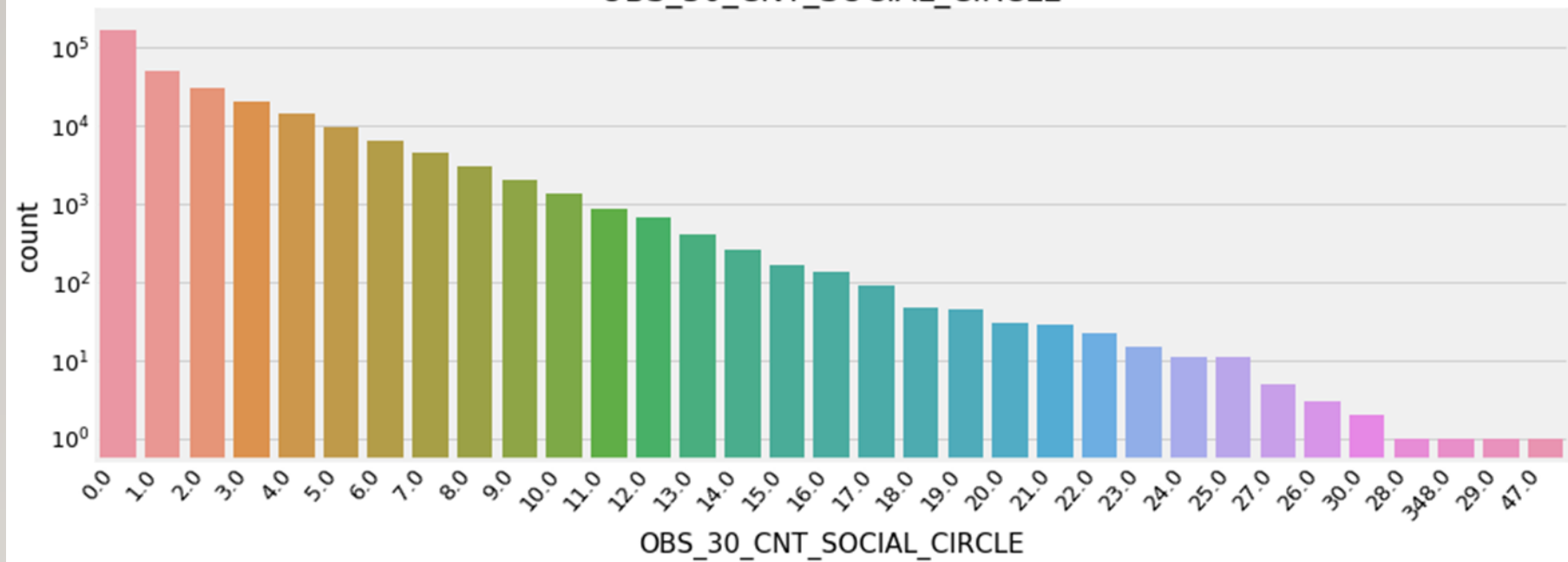


Years Since Phone Change

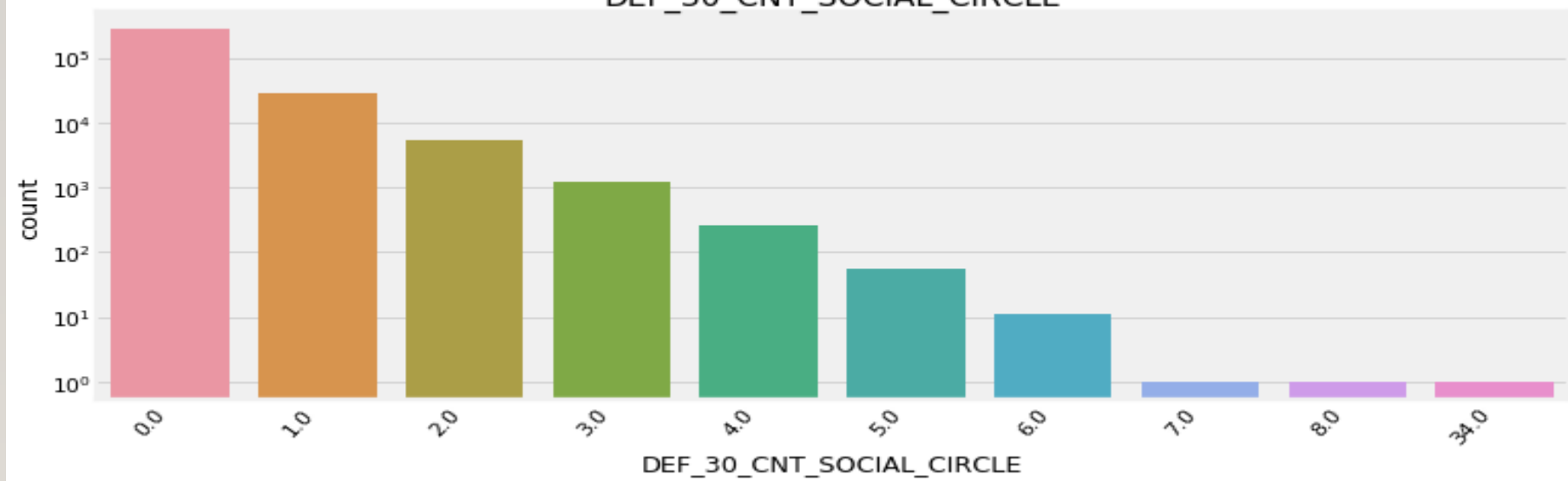


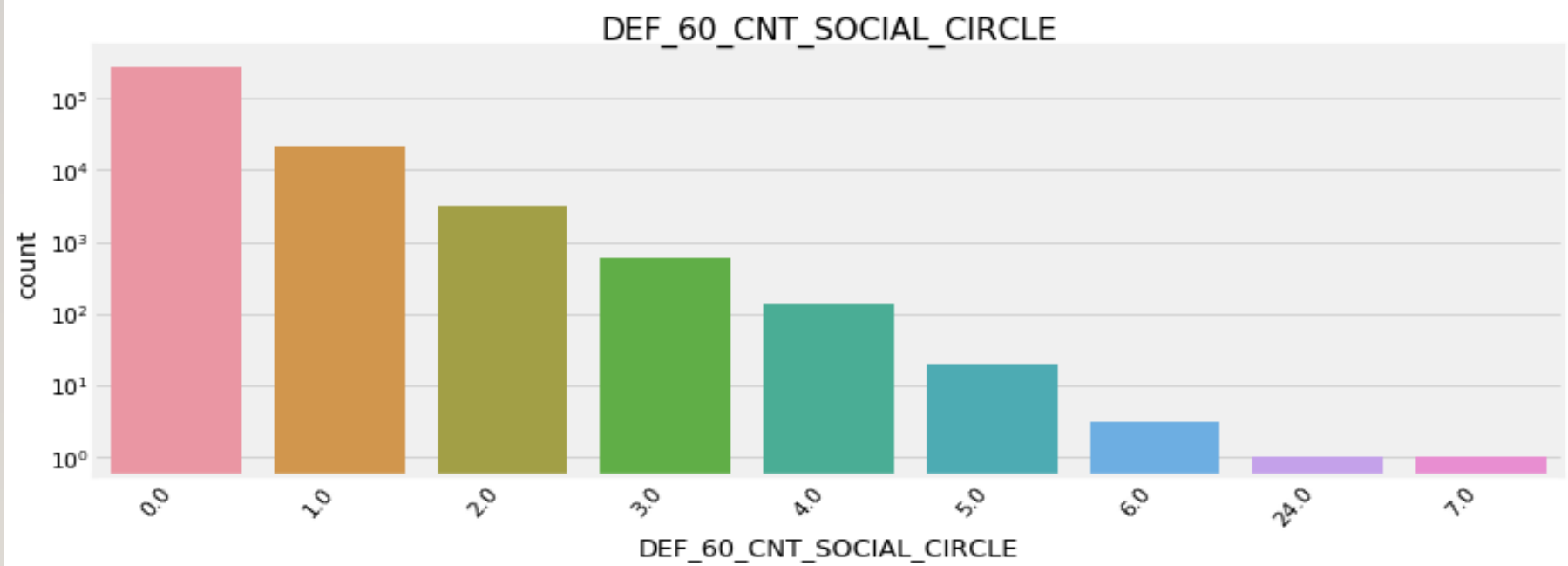
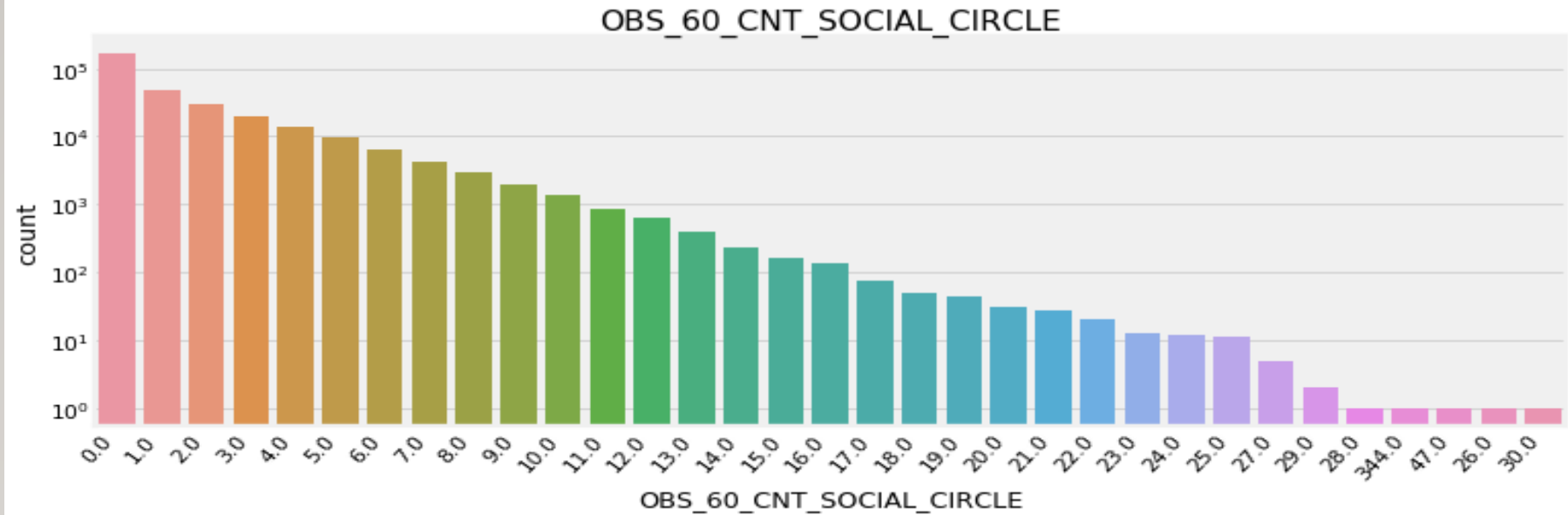


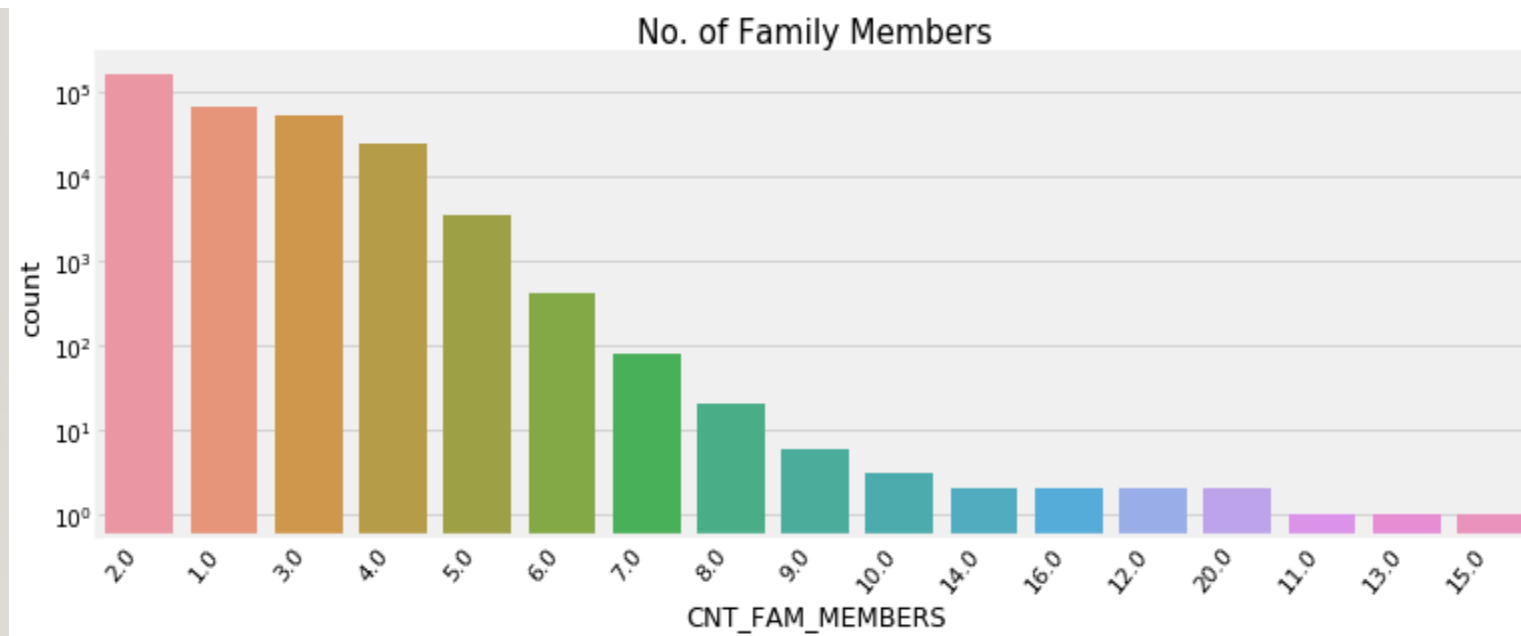
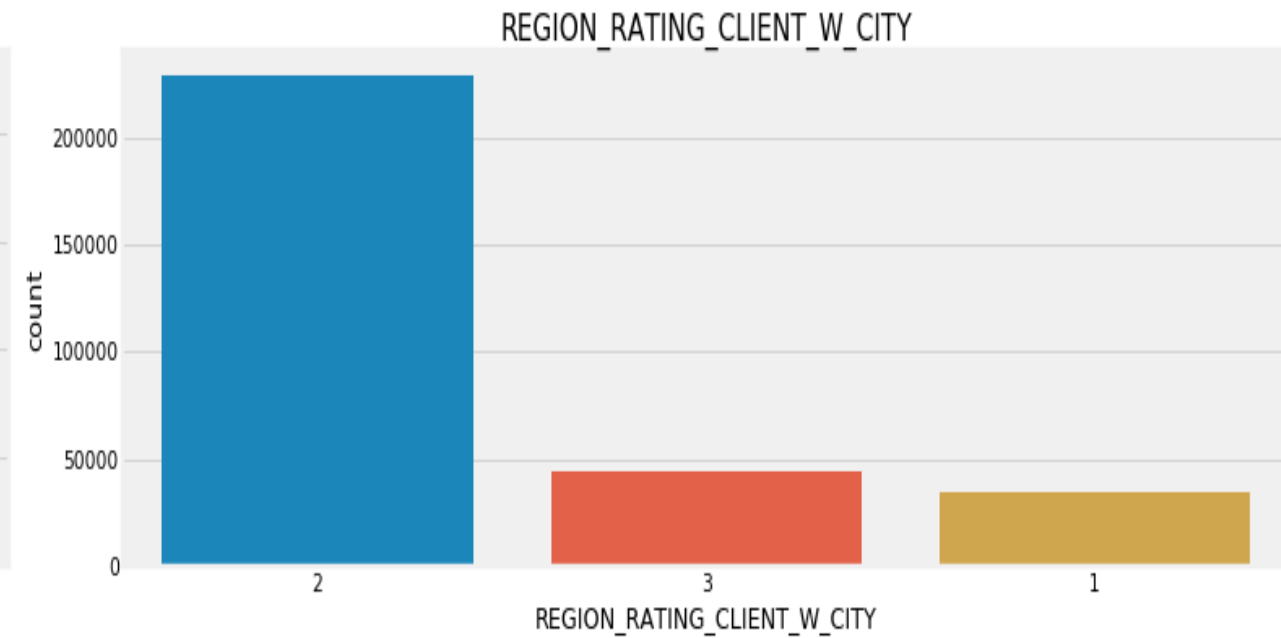
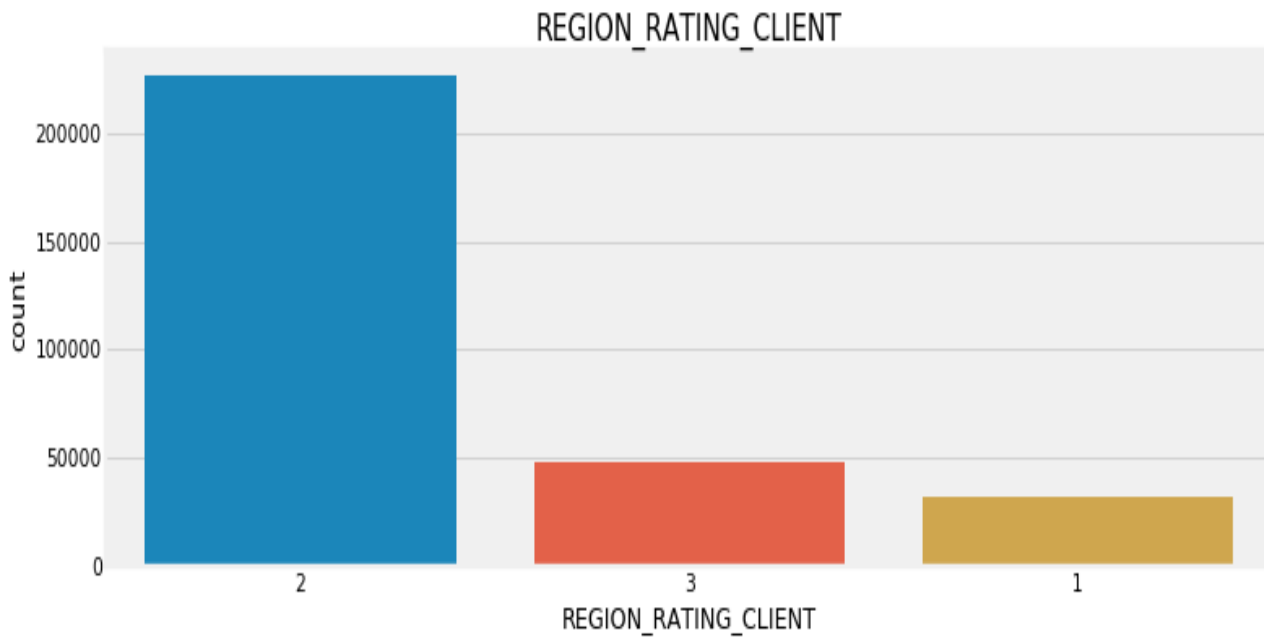
OBS_30_CNT_SOCIAL_CIRCLE



DEF_30_CNT_SOCIAL_CIRCLE

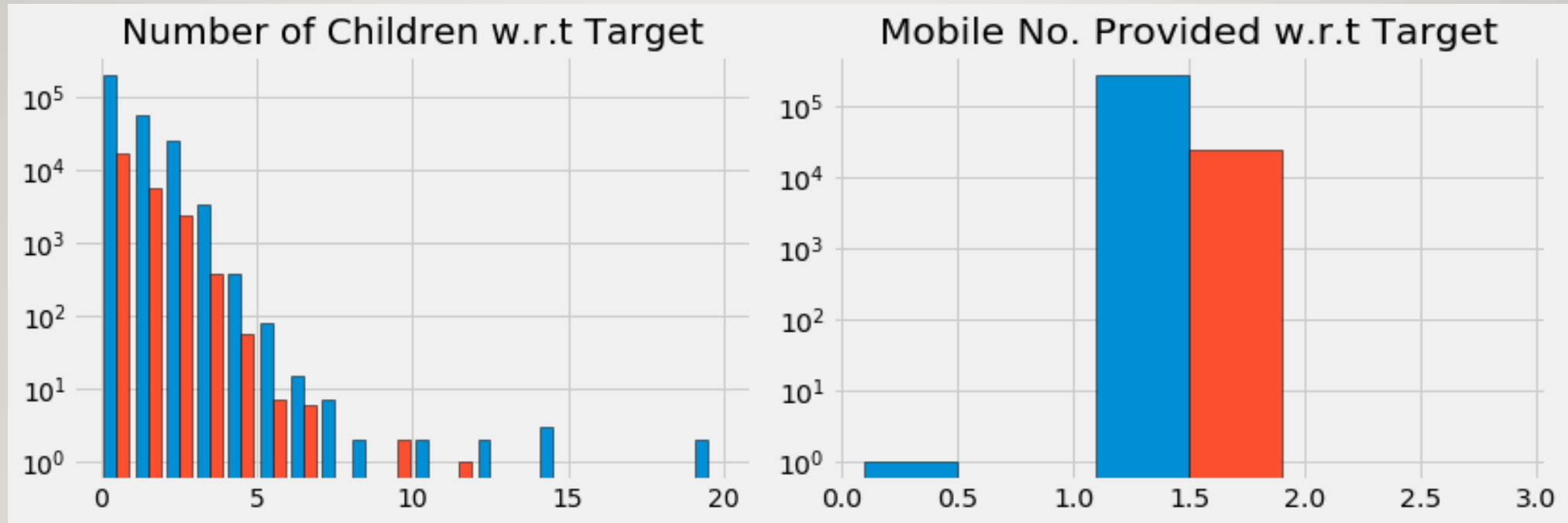




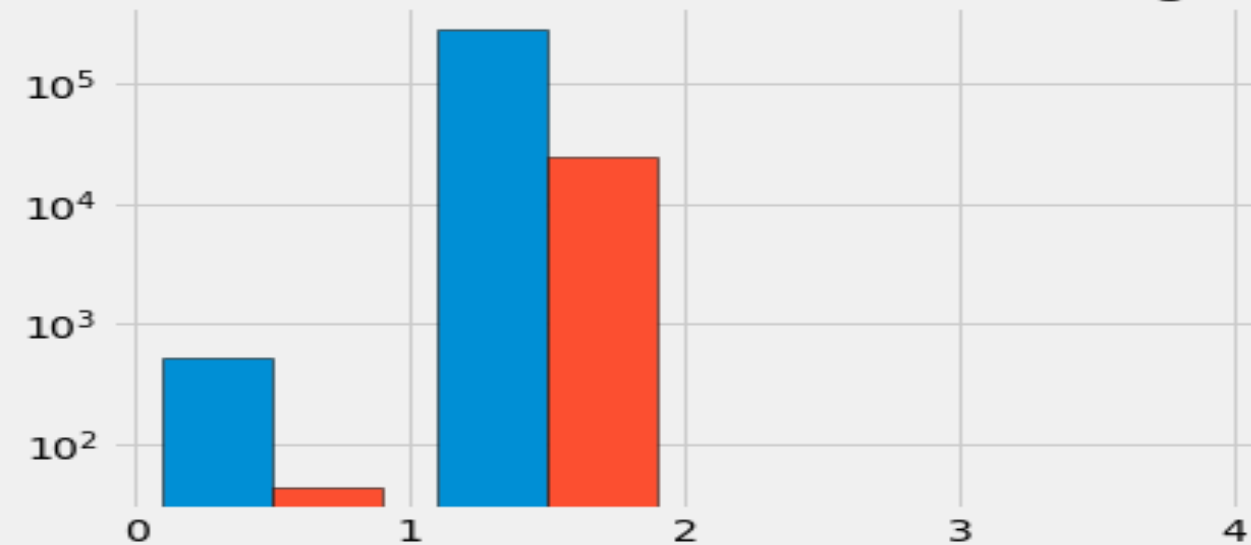


BIVARIATE ANALYSIS w.r.t TARGET:

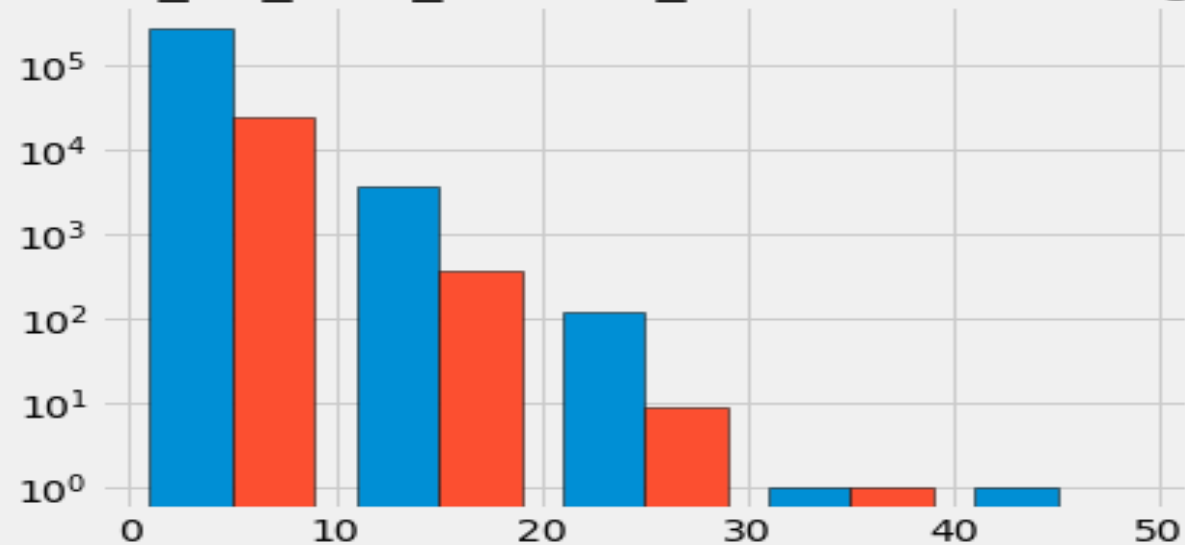
DISCRETE VARIABLES



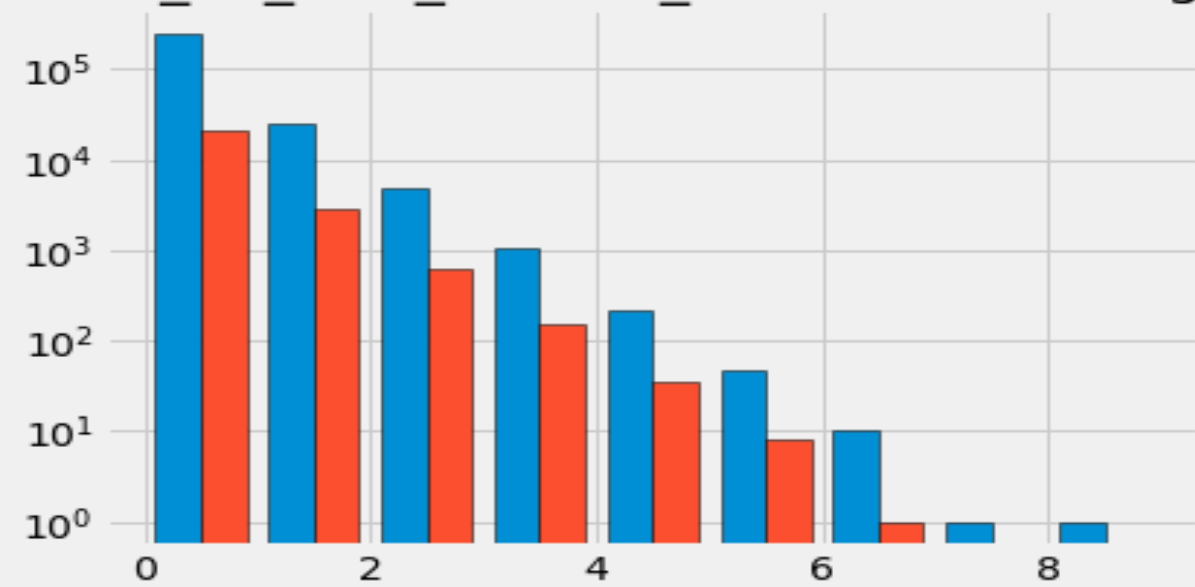
Mobile Phone Reachable w.r.t Target



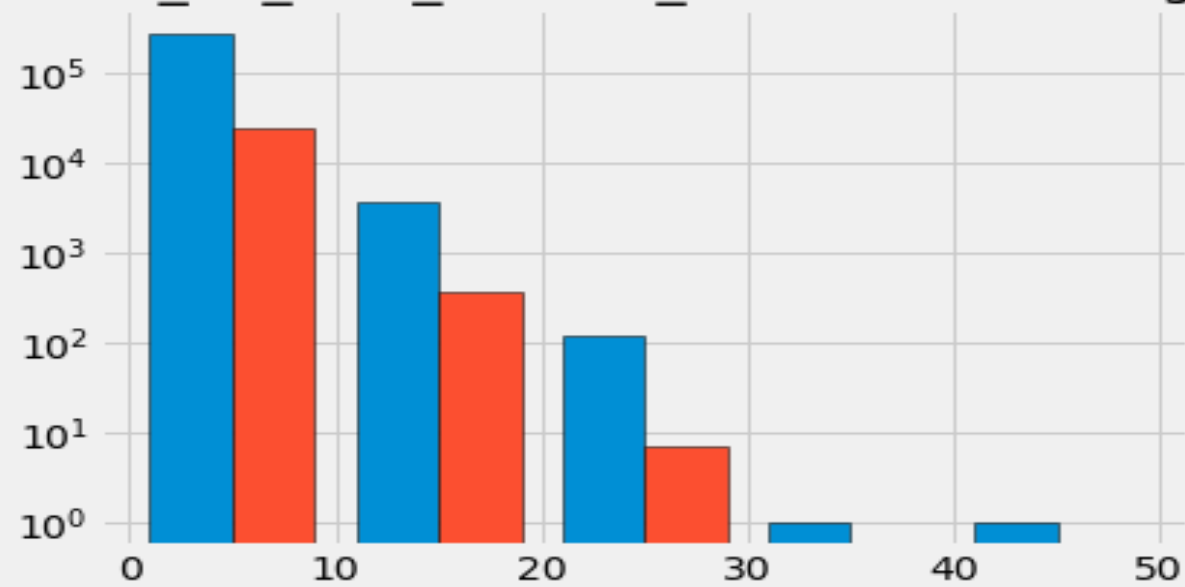
OBS_30_CNT_SOCIAL_CIRCLE w.r.t Target



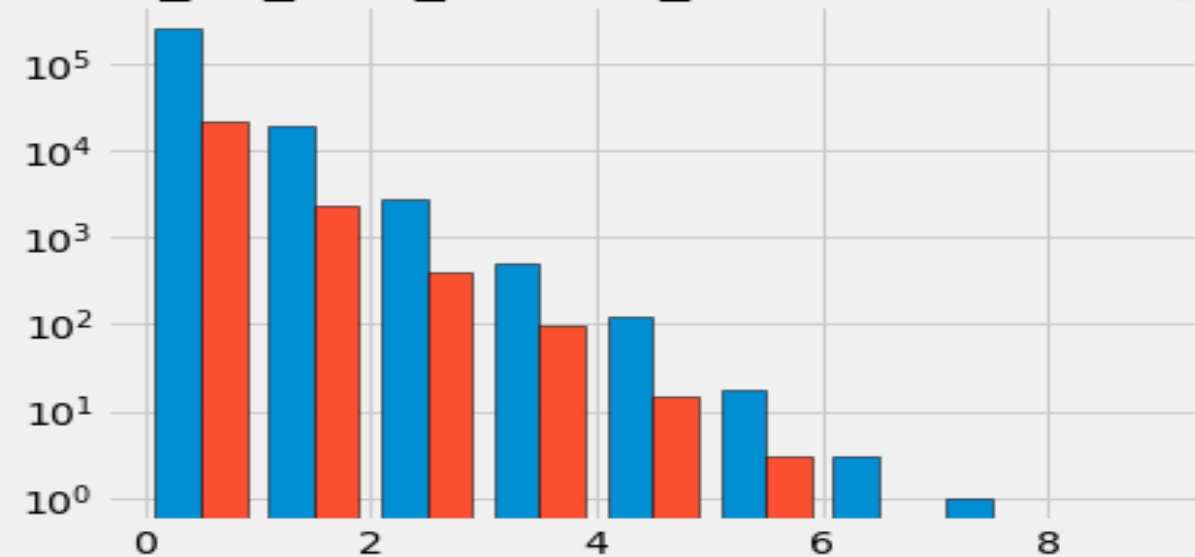
DEF_30_CNT_SOCIAL_CIRCLE w.r.t Target



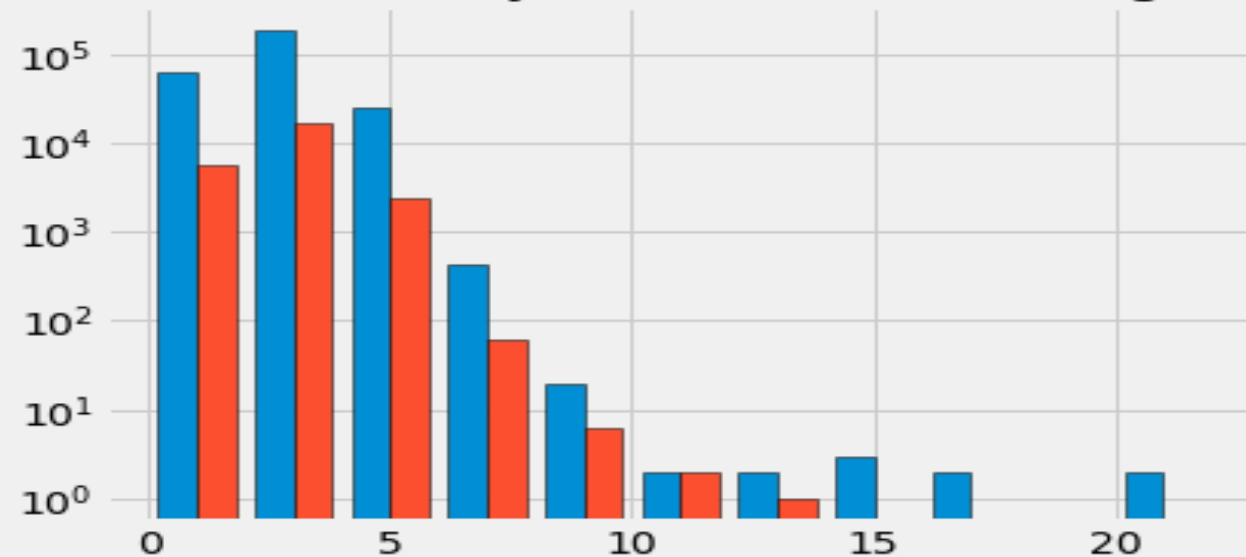
OBS_60_CNT_SOCIAL_CIRCLE w.r.t Target



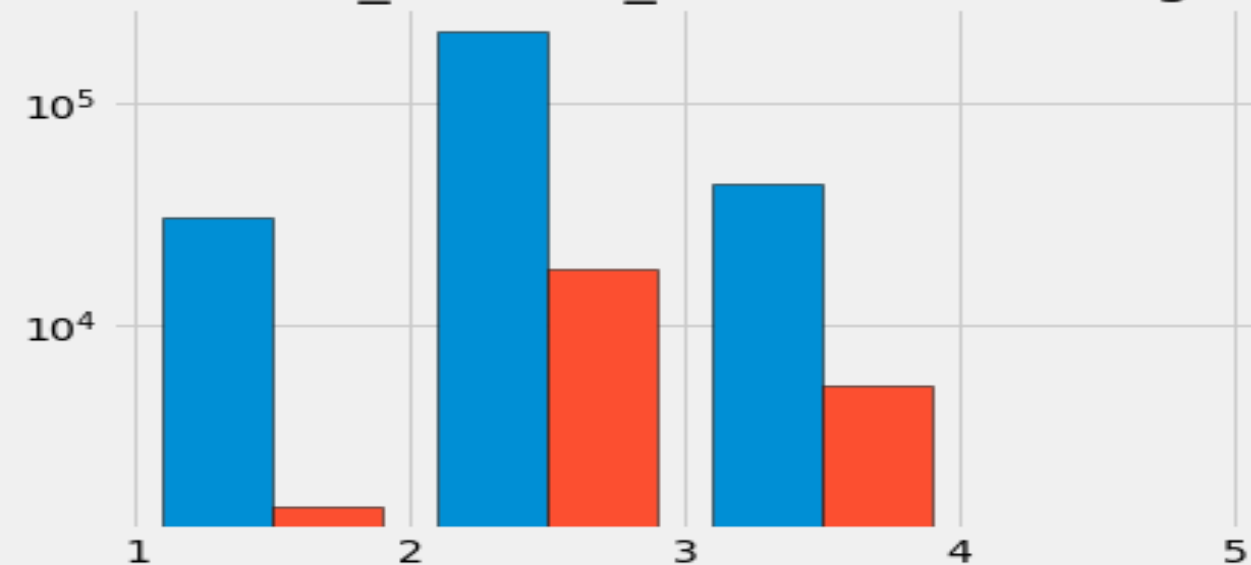
DEF_60_CNT_SOCIAL_CIRCLE w.r.t Target



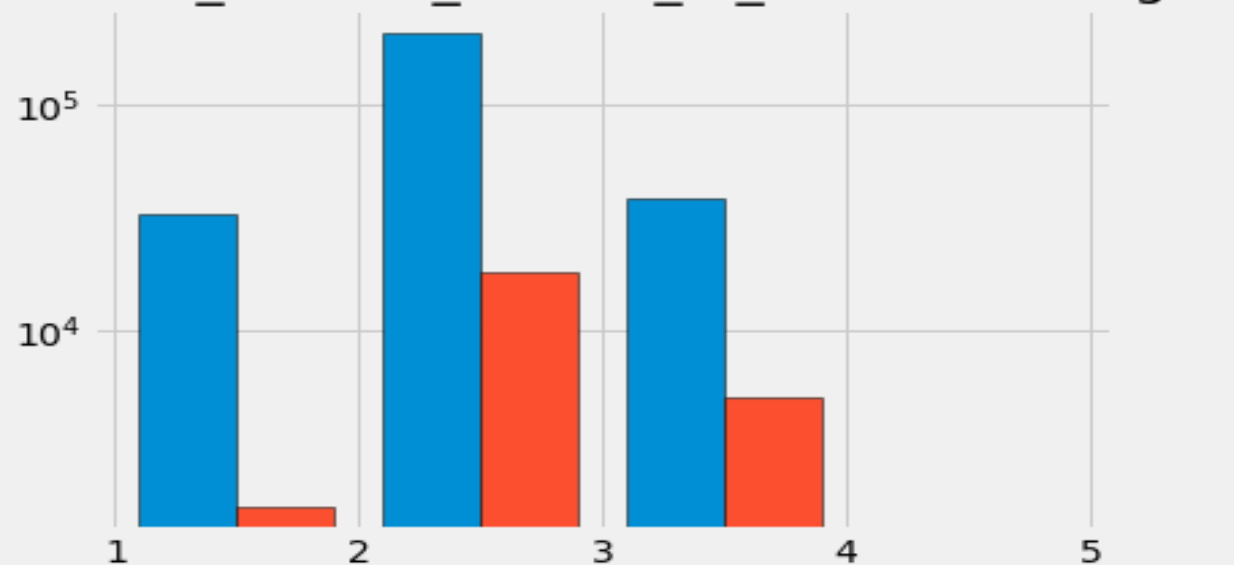
No. of Family Members w.r.t Target



REGION_RATING_CLIENT w.r.t Target



REGION_RATING_CLIENT_W_CITY w.r.t Target



INSIGHTS

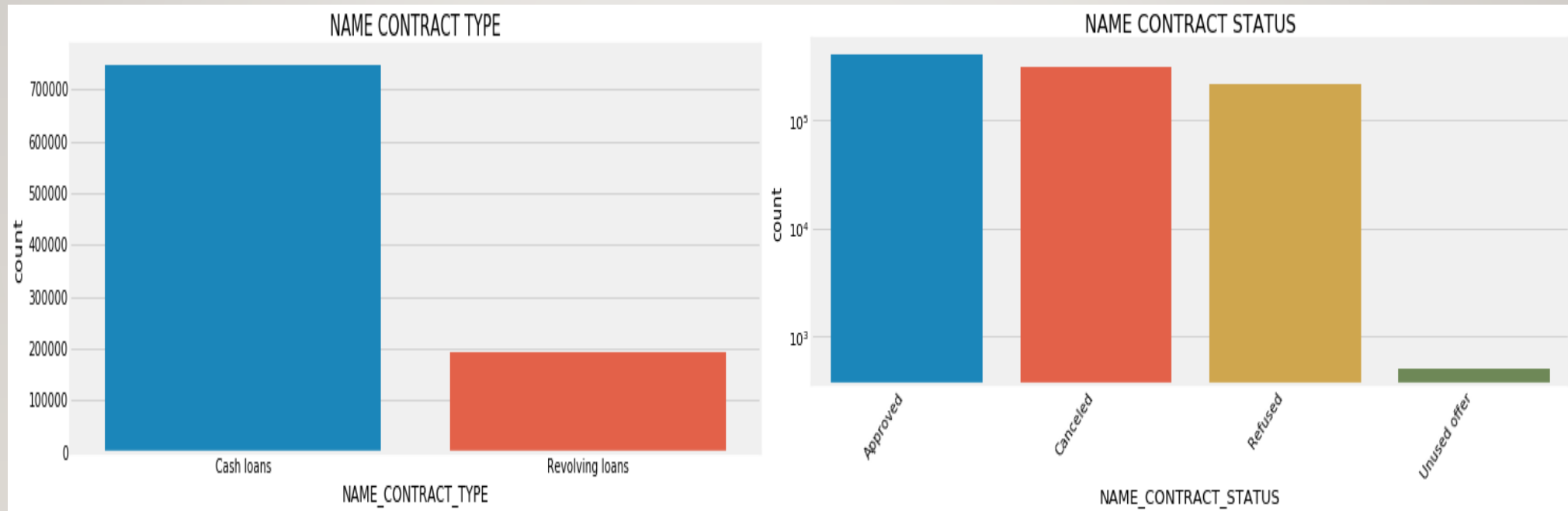
- As per the analysis, people are having fewer children and the count is less than 5 . Their family condition is unable to support more children as most of them are workers and sales staff.
- People between the age of 35-45 are the most frequent loan appliers as they want to settle down with their families, repaying their housing loans or car loans. The trend doesn't fall much after 50 as well because the Pensioners have responsibility as parents to get their children into good schools and colleges with decent fees and hence the defaulters rate increase.
- As per the trends, it is clear that the applicants don't change their ID frequently; on the other hand, applicants change their phone numbers very frequently which puts the bank at risk and signals them to verify their documents even more strictly as ID generation process was much easier long back and not much effort has been made to replace all the government IDs. Aadhar is into effect but not everyone has the relevant IDs.
- Also the applicants provided their phone numbers and were reachable too but still counted as the majority of defaulters which shows that they have fooled the banks by giving correct numbers initially but changing their numbers afterwards.
- As the number of family members are increasing, the family standards increase as more people contribute towards the family.
- Clients with their city rated as 2 took most of the loans.

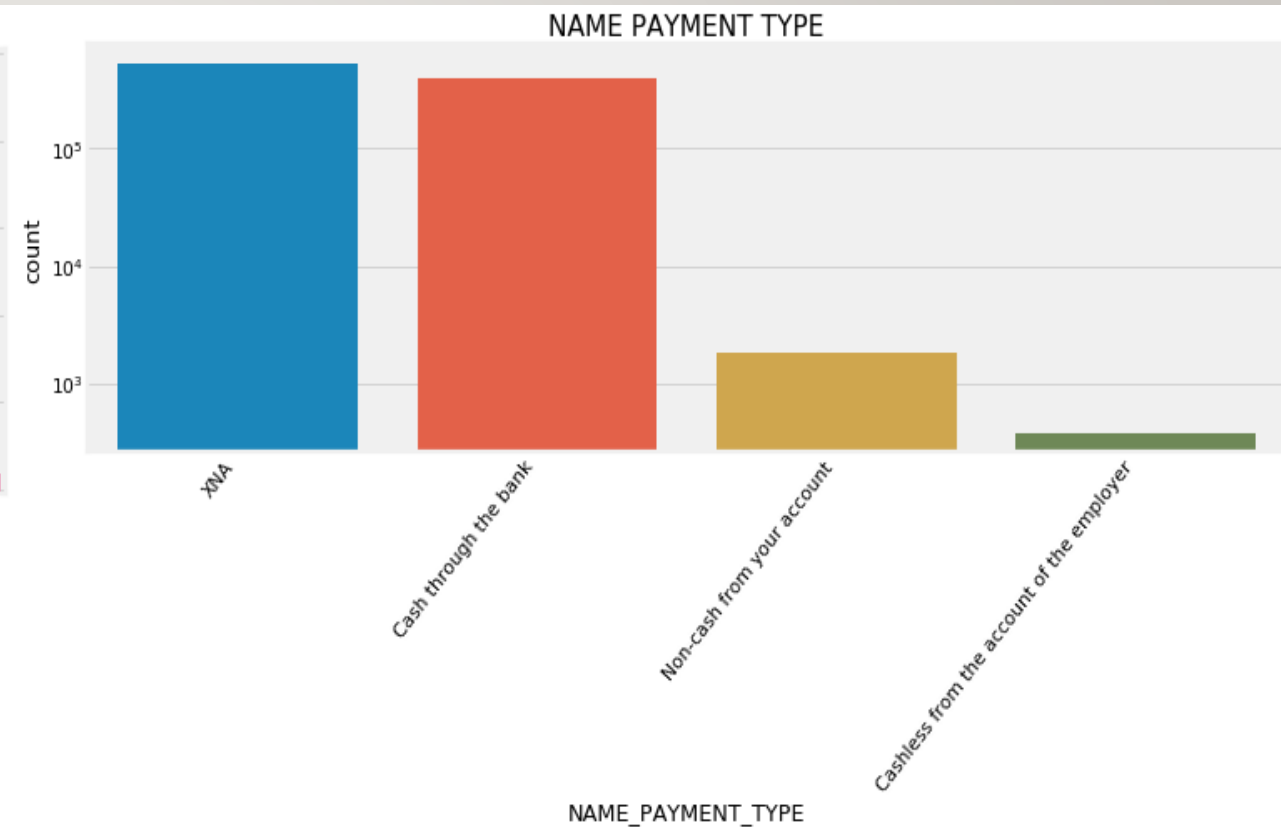
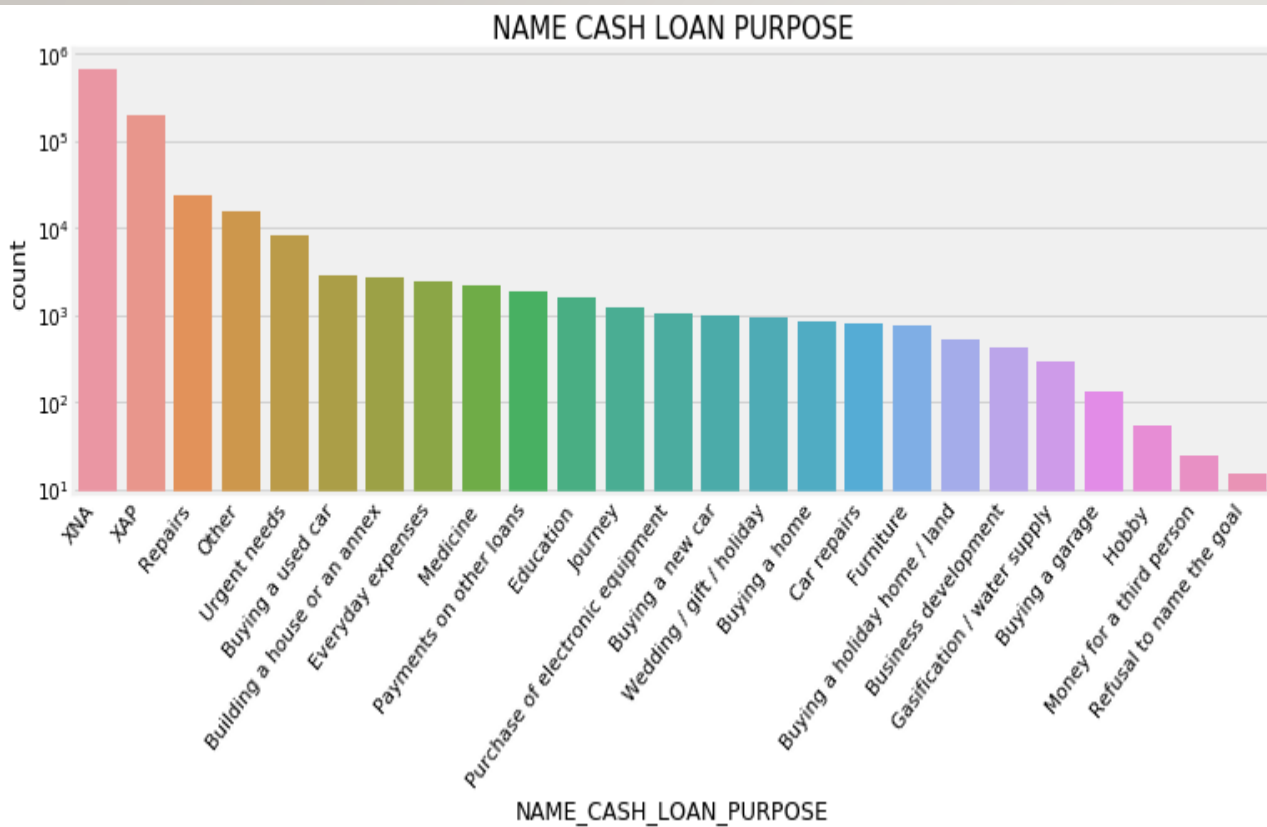
PREVIOUS APPLICATION DATASET

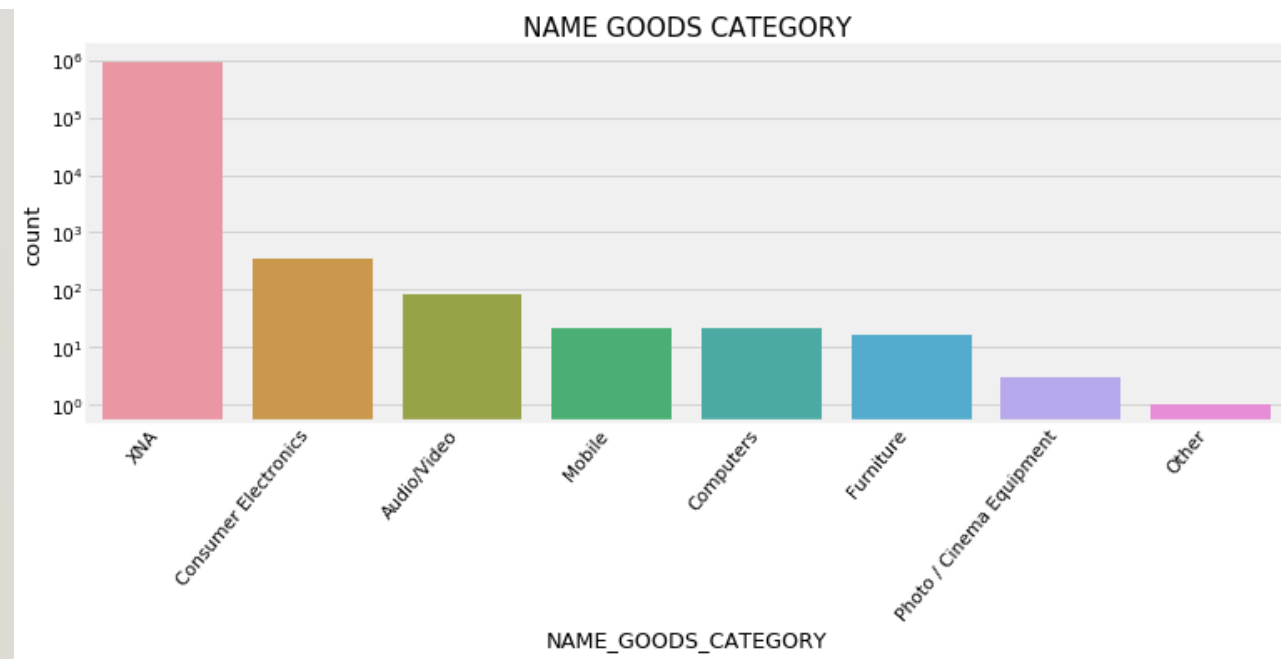
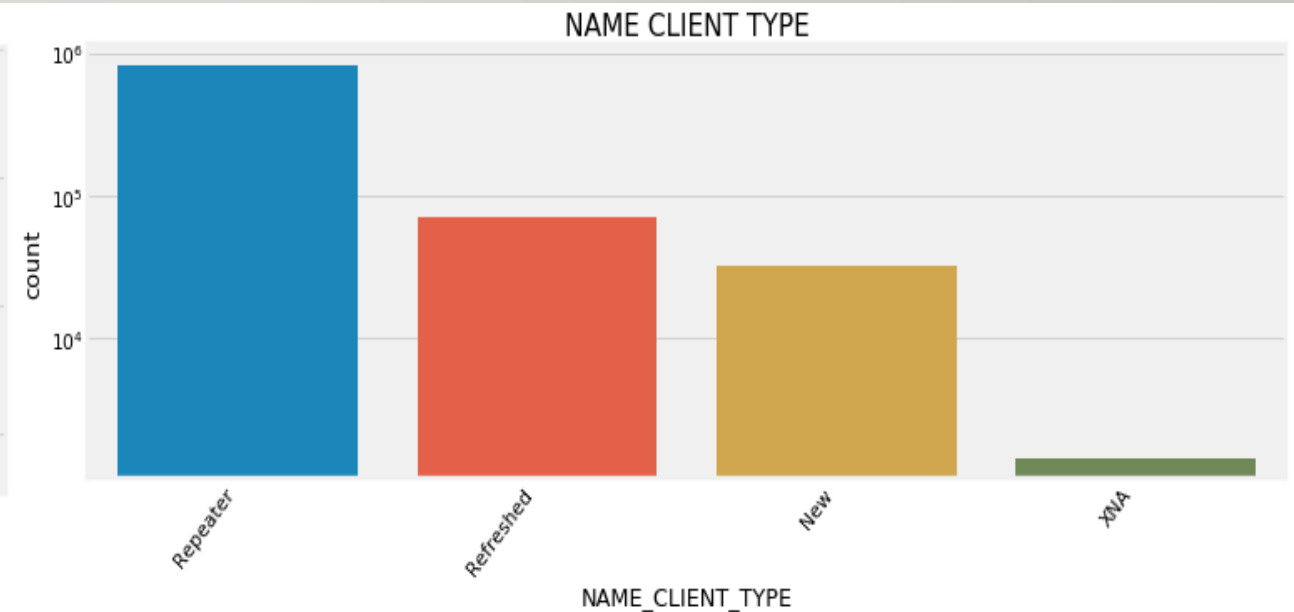
Source: [previous_application.csv](#)

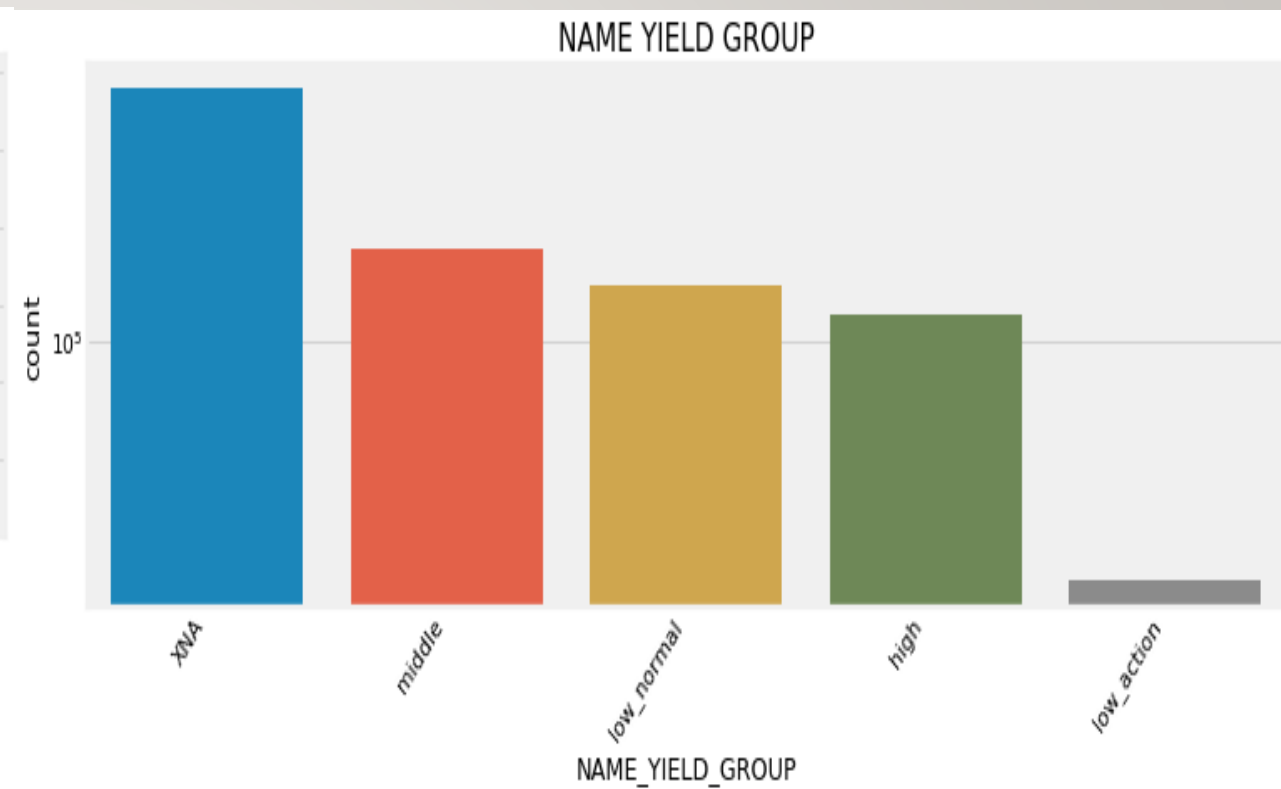
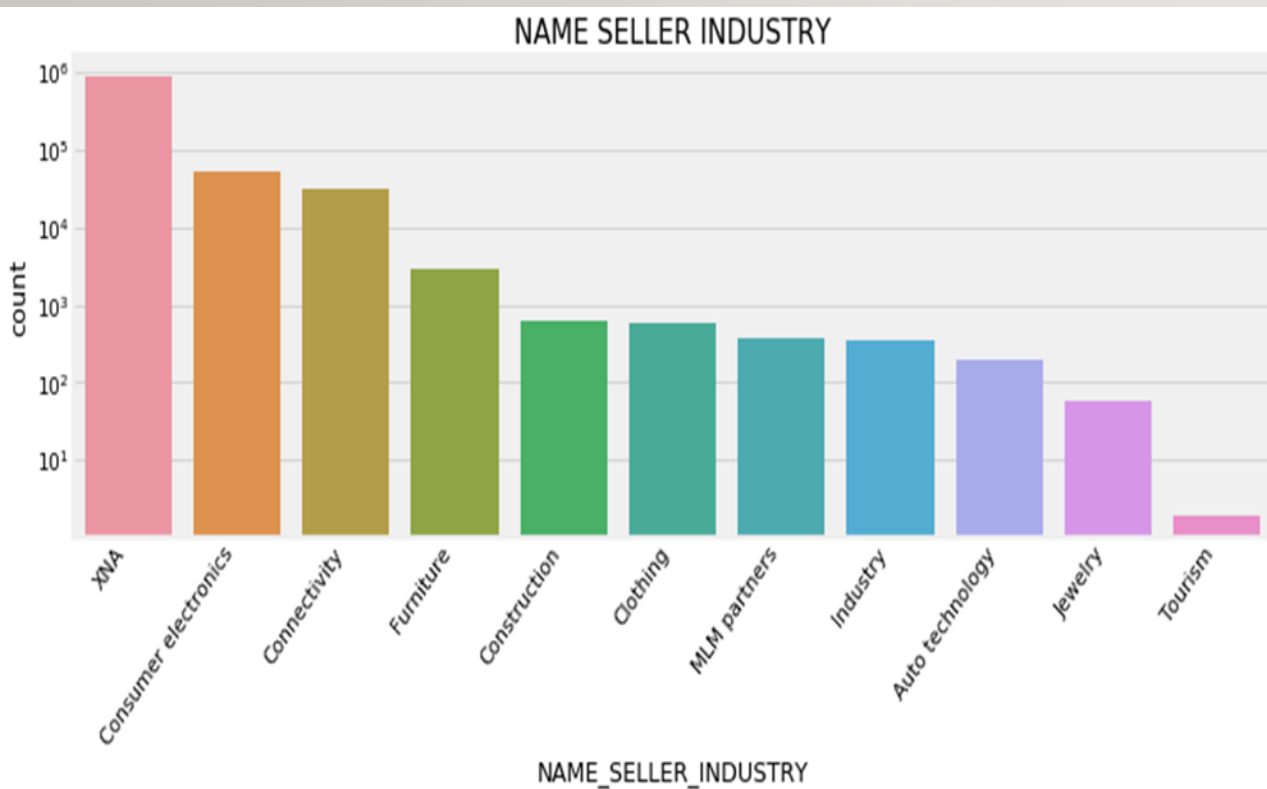
UNIVARIATE ANALYSIS:

CATEGORICAL VARIABLES







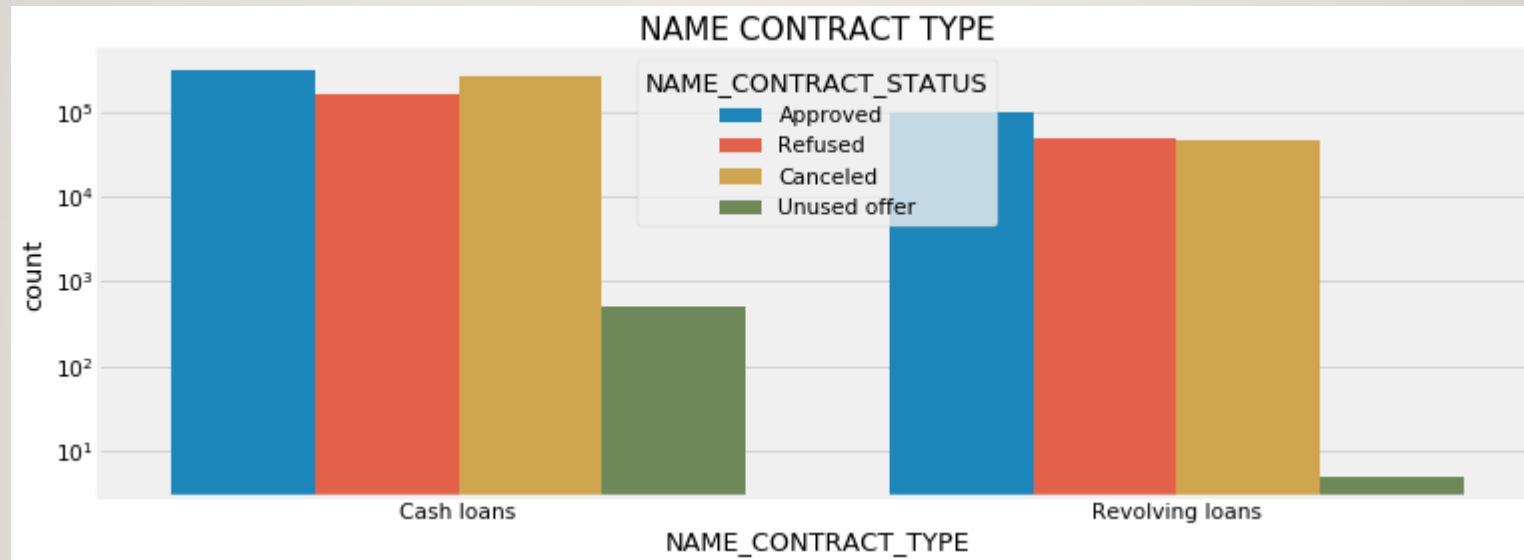


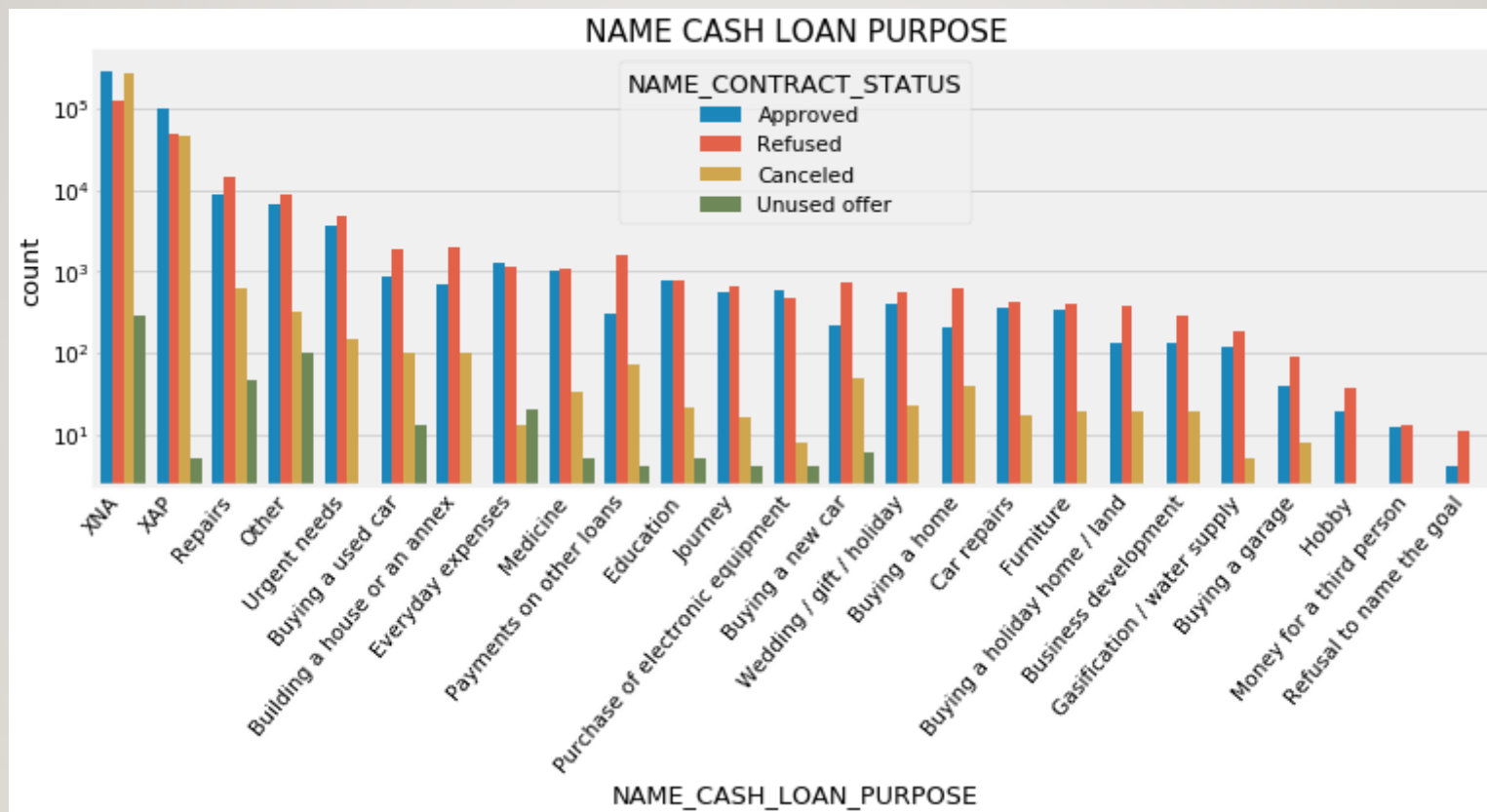
INSIGHTS

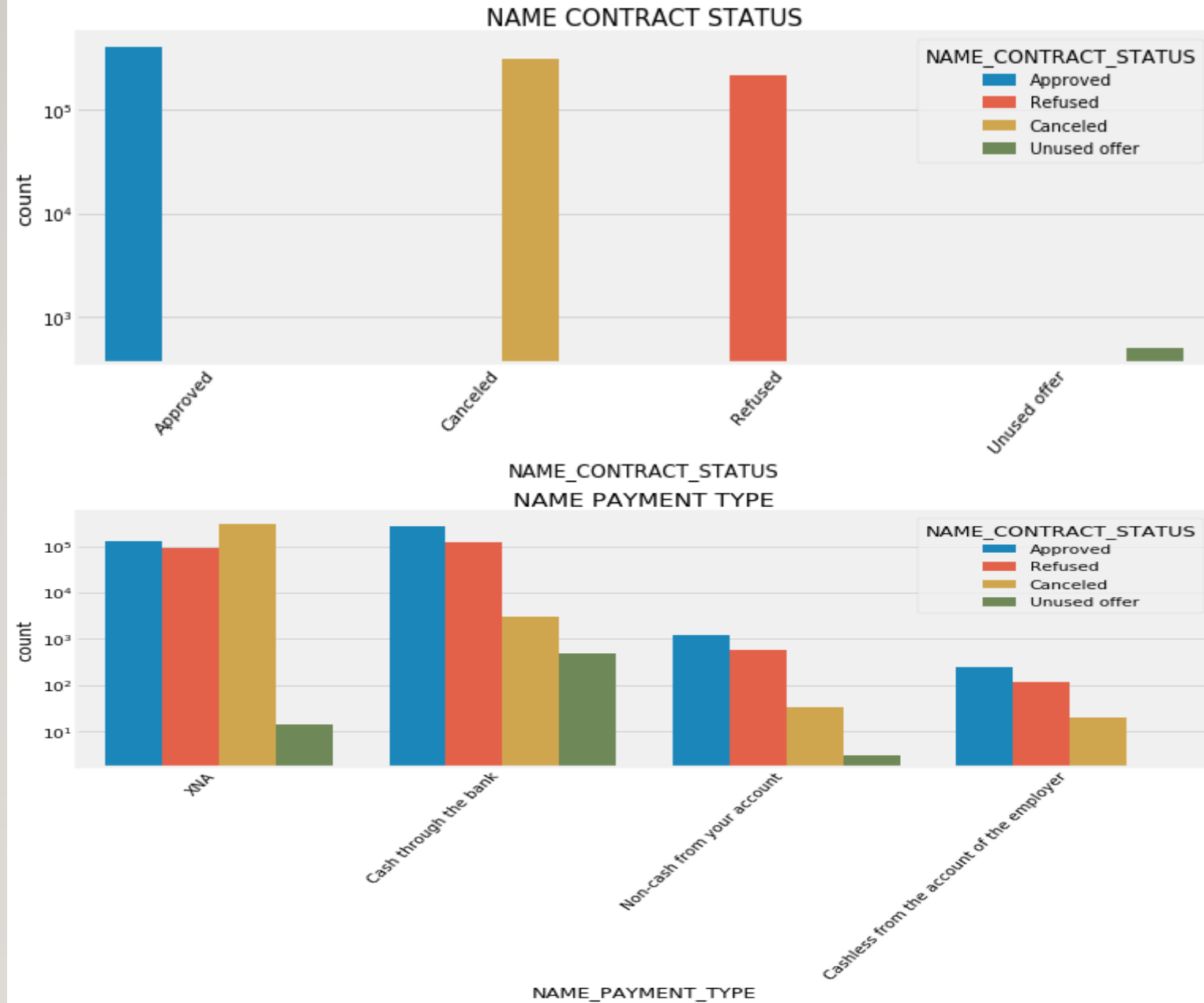
- As per the form provided by the applicants, most of the individuals have not provided their purpose of Cash Loan. The ones who have provided their purpose have loans taken for Repairs and Urgent needs. Less salaried people have taken more loans to buy used cars rather than buying a new one.
- The percentage of loans getting approved were slightly higher than the ones getting cancelled and refused which is a model that may vary from bank to bank.
- Most loan repayments were through cash.
- Repeaters showed the highest trends by taking the same loans again or may be for some other purpose.
- Grouped interest rates of middle category with highest loan applications signify more consumer goods and housing loans.

BIVARIATE ANALYSIS w.r.t CONTRACT STATUS:

CATEGORICAL VARIABLES



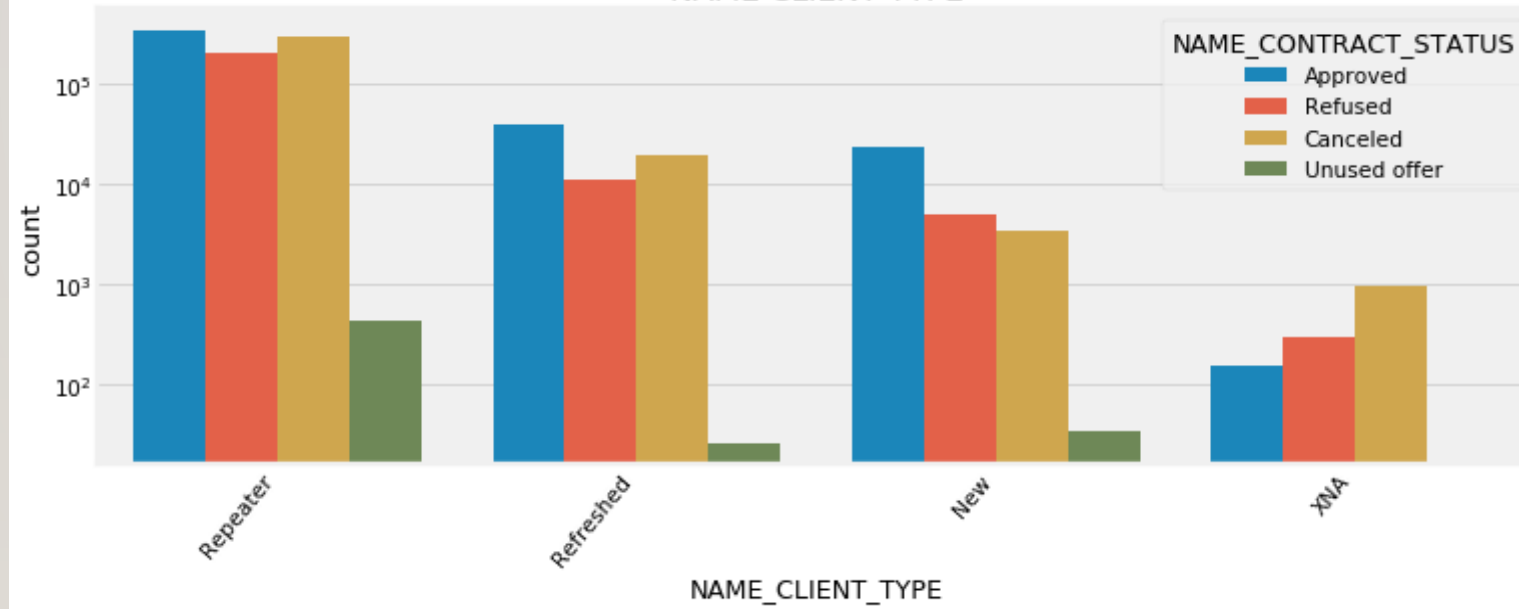




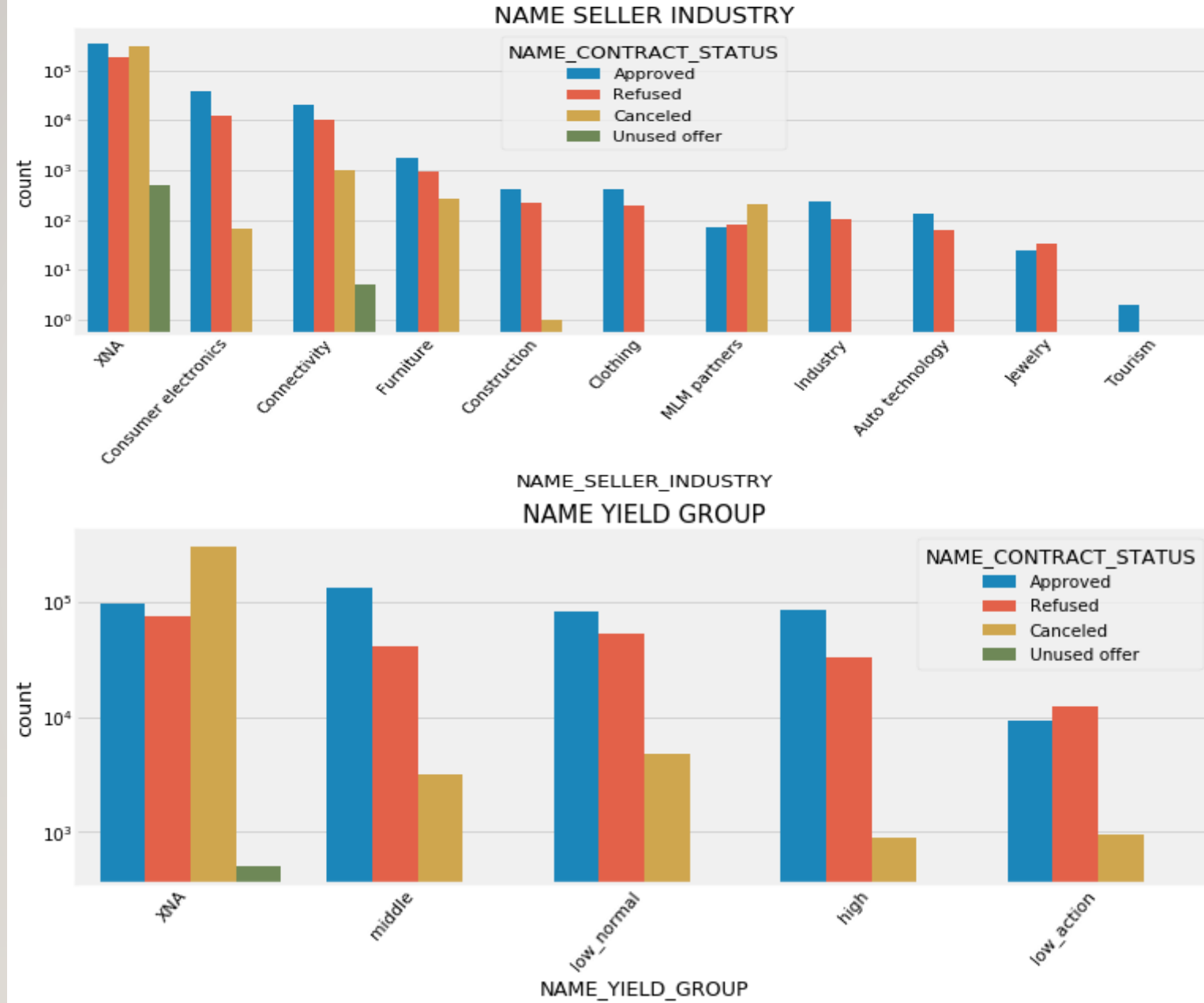
CODE REJECT REASON



NAME CLIENT TYPE







INSIGHTS

- Taking loans for auto repairs and purchase of automobiles have more chances of loans getting refused from the bank whereas buying a loan for everyday needs will get the loan cleared easily.
- Payments made via cash through the bank always led to loans getting accepted easily.
- Repeaters got their loans approved easily compared to the refreshed and the new ones but also showed more applications cancelled, which shows the change of mind. On the contrary, new applicants got their loans approved easily and had fewer cancellations as banks spend a lot of their budget promoting themselves for everyday loans via advertisements.
- The loans for consumer products like electronics were easily granted as the loan amount was small and with salaried individuals its easier to repay the loan in max intervals of 6 months.
- Loan approval was much easier for middle interest loans and hard for low interest personal loans as can be seen from the trends showing a high number of applications refused.

MERGING CURRENT AND PREVIOUS APPLICATION DATASETS

CREATED PIVOT TABLE FROM MERGED DATAFRAME :
CORRELATION MATRIX

		AGE	AMT_ANNUITY_x	AMT_ANNUITY_y	AMT_APPLICATION	AMT_CREDIT_x	AMT_CREDIT_y	AMT_GOODS_PRICE_x	AMT_GOODS_PRICE_y	AMT_INCOME_TOTAL	CNT_CHILDREN	CNT_FAM_MEMBERS	DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	EXT_SOURCE_2	OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	YEARS_EMPLOYED	YEARS_SINCE_ID_CHANGE	YEARS_SINCE_PHONE_CHANGE
AGE		1	-0.012	0.054	0.094	0.057	0.08	0.057	0.084	-0.078	-0.33	-0.28	-0.00021	-0.0019	0.092	-0.0074	-0.007	0.67	0.27	0.083
AMT_ANNUITY_x		-0.012	1	0.31	0.19	0.79	0.19	0.8	0.25	0.49	0.023	0.08	-0.022	-0.024	0.13	-0.01	-0.0097	-0.09	-0.014	0.068
AMT_ANNUITY_y		0.054	0.31	1	0.72	0.26	0.72	0.27	0.79	0.36	-0.0012	0.049	-0.022	-0.023	0.12	-0.008	-0.0077	-0.036	-0.0045	0.098
AMT_APPLICATION		0.094	0.19	0.72	1	0.21	0.96	0.22	0.85	0.24	-0.02	0.028	-0.0069	-0.011	0.085	0.015	0.015	0.013	0.02	0.12
AMT_CREDIT_x		0.057	0.79	0.26	0.21	1	0.22	0.99	0.27	0.41	0.0022	0.065	-0.021	-0.023	0.13	0.0013	0.0014	-0.038	0.0063	0.079
AMT_CREDIT_y		0.08	0.19	0.72	0.96	0.22	1	0.22	0.85	0.25	-0.017	0.033	-0.0064	-0.011	0.089	0.014	0.014	-0.0061	0.016	0.14
AMT_GOODS_PRICE_x		0.057	0.8	0.27	0.22	0.99	0.22	1	0.28	0.41	-0.0017	0.063	-0.022	-0.024	0.14	0.0017	0.0017	-0.035	0.0087	0.083
AMT_GOODS_PRICE_y		0.084	0.25	0.79	0.85	0.27	0.85	0.28	1	0.32	-0.013	0.047	-0.018	-0.021	0.11	0.0084	0.0086	-0.0055	0.0098	0.13
AMT_INCOME_TOTAL		-0.078	0.49	0.36	0.24	0.41	0.25	0.41	0.32	1	0.029	0.037	-0.032	-0.032	0.17	-0.026	-0.026	-0.16	-0.028	0.064
CNT_CHILDREN		-0.33	0.023	-0.0012	-0.02	0.0022	-0.017	-0.0017	-0.013	0.029	1	0.88	-0.0013	-0.0019	-0.018	0.016	0.015	-0.25	0.028	0.0059
CNT_FAM_MEMBERS		-0.28	0.08	0.049	0.028	0.065	0.033	0.063	0.047	0.037	0.88	1	-0.0027	-0.0047	-0.0018	0.025	0.025	-0.23	0.021	0.027
DEF_30_CNT_SOCIAL_CIRCLE		-0.00021	-0.022	-0.022	-0.0069	-0.021	-0.0064	-0.022	-0.018	-0.032	-0.0013	-0.0027	1	0.86	-0.03	0.33	0.33	0.013	-0.0027	-0.00042
DEF_60_CNT_SOCIAL_CIRCLE		-0.0019	-0.024	-0.023	-0.011	-0.023	-0.011	-0.024	-0.021	-0.032	-0.0019	-0.0047	0.86	1	-0.032	0.25	0.26	0.01	-0.0043	-0.002
EXT_SOURCE_2		0.092	0.13	0.12	0.085	0.13	0.089	0.14	0.11	0.17	-0.018	-0.0018	-0.03	-0.032	1	-0.019	-0.019	0.0065	0.051	0.2
OBS_30_CNT_SOCIAL_CIRCLE		-0.0074	-0.01	-0.008	0.015	0.0013	0.014	0.0017	0.0084	-0.026	0.016	0.025	0.33	0.25	-0.019	1	1	0.006	0.012	0.015
OBS_60_CNT_SOCIAL_CIRCLE		-0.007	-0.0097	-0.0077	0.015	0.0014	0.014	0.0017	0.0086	-0.026	0.015	0.025	0.33	0.26	-0.019	1	1	0.0062	0.013	0.015
YEARS_EMPLOYED		0.67	-0.09	-0.036	0.013	-0.038	-0.0061	-0.035	-0.0055	-0.16	-0.25	-0.23	0.013	0.01	0.0065	0.006	0.0062	1	0.29	0.018
YEARS_SINCE_ID_CHANGE		0.27	-0.014	-0.0045	0.02	0.0063	0.016	0.0087	0.0098	-0.028	0.028	0.021	-0.0027	-0.0043	0.051	0.012	0.013	0.29	1	0.089
YEARS_SINCE_PHONE_CHANGE		0.083	0.068	0.098	0.12	0.079	0.14	0.083	0.13	0.064	0.0059	0.027	-0.00042	-0.002	0.2	0.015	0.015	0.018	0.089	1



CORRELATION MATRIX

w.r.t TARGET=0



CORRELATION MATRIX

w.r.t TARGET=1

AGE_AMT_ANNUITY_x_AMT_ANNUITY_y_AMT_APPLICATION_AMT_CREDIT_x_AMT_CREDIT_y_AMT_GOODS_PRICE_x_AMT_GOODS_PRICE_y_AMT_INCOME_TOTAL_CNT_CHILDREN_CNT_FAM_MEMBERS_DEF_30_CNT_SOCIAL_CIRCLE_DEF_60_CNT_SOCIAL_CIRCLE_EXT_SOURCE_2_OBS_30_CNT_SOCIAL_CIRCLE_OBS_60_CNT_SOCIAL_CIRCLE_YEARS_EMPLOYED_YEARS_SINCE_ID_CHANGE_YEARS_SINCE_PHONE_CHANGE																				
AGE	1	0.013	0.095	0.15	0.14	0.14	0.14	0.14	-0.0027	-0.26	-0.2	-0.0049	-0.004	0.11	0.0052	0.0054	0.62	0.25		0.11
AMT_ANNUITY_x	0.013	1	0.26	0.17	0.76	0.16	0.76	0.2	0.44	0.032	0.077	-0.022	-0.028	0.12	0.0059	0.007	-0.069	0.015		0.081
AMT_ANNUITY_y	0.095	0.26	1	0.73	0.23	0.75	0.24	0.81	0.33	-0.0018	0.048	-0.0099	-0.017	0.11	0.017	0.017	-0.0082	0.016		0.14
AMT_APPLICATION	0.15	0.17	0.73	1	0.22	0.96	0.23	0.85	0.22	-0.014	0.037	-0.0071	-0.018	0.096	0.041	0.041	0.048	0.048		0.16
AMT_CREDIT_x	0.14	0.76	0.23	0.22	1	0.22	0.98	0.25	0.36	-0.001	0.053	-0.026	-0.031	0.12	0.02	0.021	0.031	0.052		0.11
AMT_CREDIT_y	0.14	0.16	0.75	0.96	0.22	1	0.23	0.85	0.23	-0.014	0.04	-0.0013	-0.011	0.1	0.043	0.043	0.038	0.053		0.18
AMT_GOODS_PRICE_x	0.14	0.76	0.24	0.23	0.98	0.23	1	0.26	0.36	-0.0072	0.05	-0.023	-0.027	0.13	0.021	0.021	0.037	0.057		0.12
AMT_GOODS_PRICE_y	0.14	0.2	0.81	0.85	0.25	0.85	0.26	1	0.29	-0.016	0.045	-0.015	-0.027	0.11	0.037	0.037	0.044	0.039		0.17
AMT_INCOME_TOTAL	-0.0027	0.44	0.33	0.22	0.36	0.23	0.36	0.29	1	-0.0025	0.0054	-0.024	-0.025	0.15	-0.0048	-0.0042	-0.11	0.004		0.086
CNT_CHILDREN	-0.26	0.032	-0.0018	-0.014	-0.001	-0.014	-0.0072	-0.016	-0.0025	1	0.89	0.0014	-0.0051	-0.012	0.026	0.025	-0.19	0.032		-0.012
CNT_FAM_MEMBERS	-0.2	0.077	0.048	0.037	0.053	0.04	0.05	0.045	0.0054	0.89	1	0.0071	-0.00078	0.0019	0.044	0.043	-0.18	0.032		0.013
DEF_30_CNT_SOCIAL_CIRCLE	-0.0049	-0.022	-0.0099	-0.0071	-0.026	-0.0013	-0.023	-0.015	-0.024	0.0014	0.0071	1	0.87	-0.01	0.33	0.34	0.0029	-0.0076		-0.00023
DEF_60_CNT_SOCIAL_CIRCLE	-0.004	-0.028	-0.017	-0.018	-0.031	-0.011	-0.027	-0.027	-0.025	-0.0051	-0.00078	0.87	1	-0.011	0.26	0.26	0.0051	-0.0049		-0.0046
EXT_SOURCE_2	0.11	0.12	0.11	0.096	0.12	0.1	0.13	0.11	0.15	-0.012	0.0019	-0.01	-0.011	1	0.018	0.018	0.025	0.06		0.21
OBS_30_CNT_SOCIAL_CIRCLE	0.0052	0.0059	0.017	0.041	0.02	0.043	0.021	0.037	-0.0048	0.026	0.044	0.33	0.26	0.018	1	1	6.10E-05	0.02		0.029
OBS_60_CNT_SOCIAL_CIRCLE	0.0054	0.007	0.017	0.041	0.021	0.043	0.021	0.037	-0.0042	0.025	0.043	0.34	0.26	0.018	1	1	-0.00023	0.02		0.029
YEARS_EMPLOYED	0.62	-0.069	-0.0082	0.048	0.031	0.038	0.037	0.044	-0.11	-0.19	-0.18	0.0029	0.0051	0.025	6.10E-05	-0.00023	1	0.24		0.044
YEARS_SINCE_ID_CHANGE	0.25	0.015	0.016	0.048	0.052	0.053	0.057	0.039	0.004	0.032	0.032	-0.0076	-0.0049	0.06	0.02	0.02	0.24	1		0.12
YEARS_SINCE_PHONE_CHANGE	0.11	0.081	0.14	0.16	0.11	0.18	0.12	0.17	0.086	-0.012	0.013	-0.00023	-0.0046	0.21	0.029	0.029	0.044	0.12		1

CONCLUSION

Suggestion

- The company should make use of Skip – Trace tools which credit information bureaus provide allowing users to mine their repositories to identify whether the applicant has provided multiple contact details or not as proof to identify the defaulters in the small ticket personal loan segment.