EDA Credit Assignment

BANK LOAN CREDIT

PROBLEM STATEMENT

The Banks and other financial institutions lends the loans to the consumers, to lend the money they need to keep risk analysis approach on the basis of applicants profile

Bank cannot approve all the loan applications, there may be some applicants, for repaying of loan may difficult, this may lead to financial loss to the bank

Bank cannot reject all the loan applications, there may have applicants who can repay the loan, if bank rejects loan application, there will be business loss for the bank

For the risk assessment there are certain variable which are direct indicator to identify the risk in lending the money to the applicants.

Assumptions :- For cleansing the null values from both the data set, taken the 40% criteria for cleansing the data

Overall Approach

Understanding Problem statement

Understanding Datasets

Data information analysis

Data Cleansing

Data imputation

Computing Outliers

Univariate Analysis

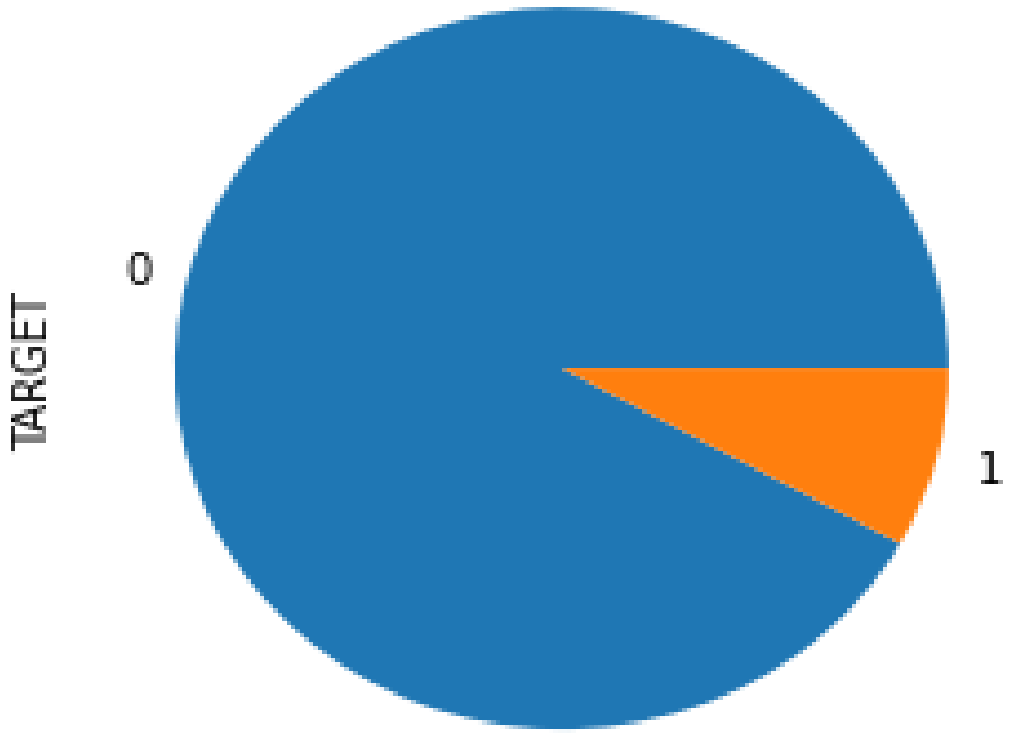
Bivariate Analysis

Multivariate Analysis

Merging Two Dataset

Merged Dataset Analysis

Data Imbalance Analysis



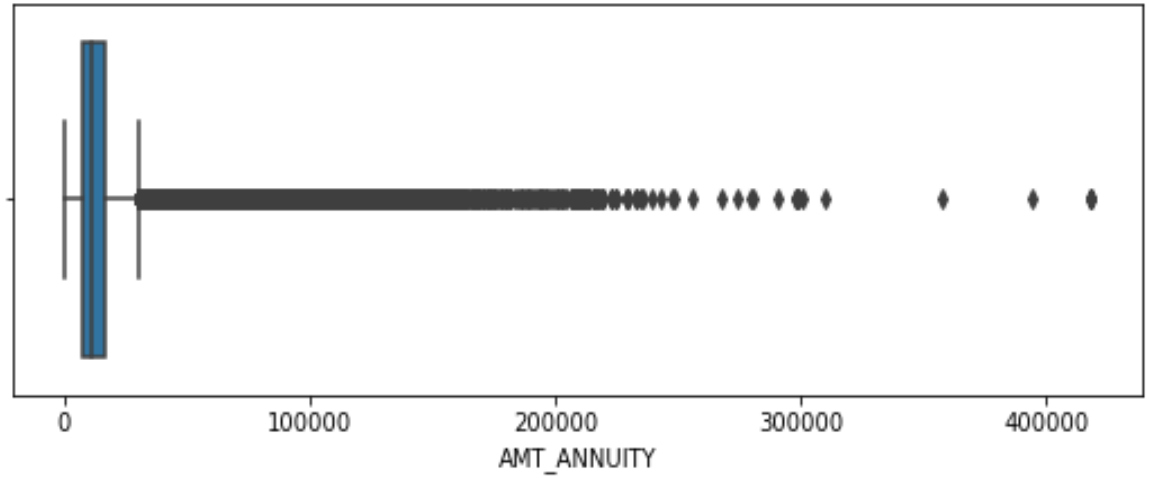
There is imbalance in TARGET variable 0 represents the Re-payers and 1 represents the Defaulters

0 = 91.93%

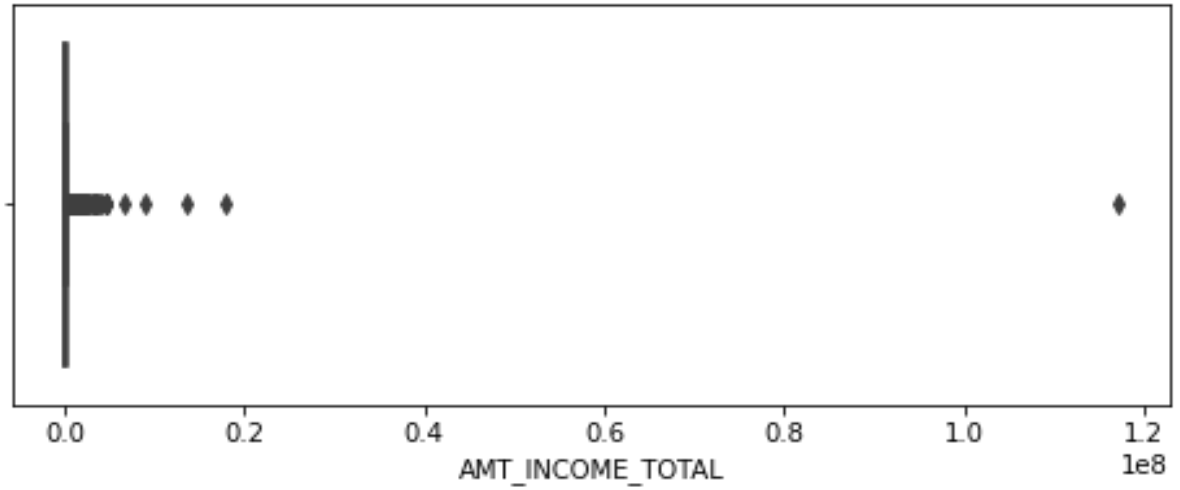
1 = 8.07%

Outliers Analysis

AMT_ANUUITY VARIABLE HAS SOME NUMBERS OF OUTLIERS

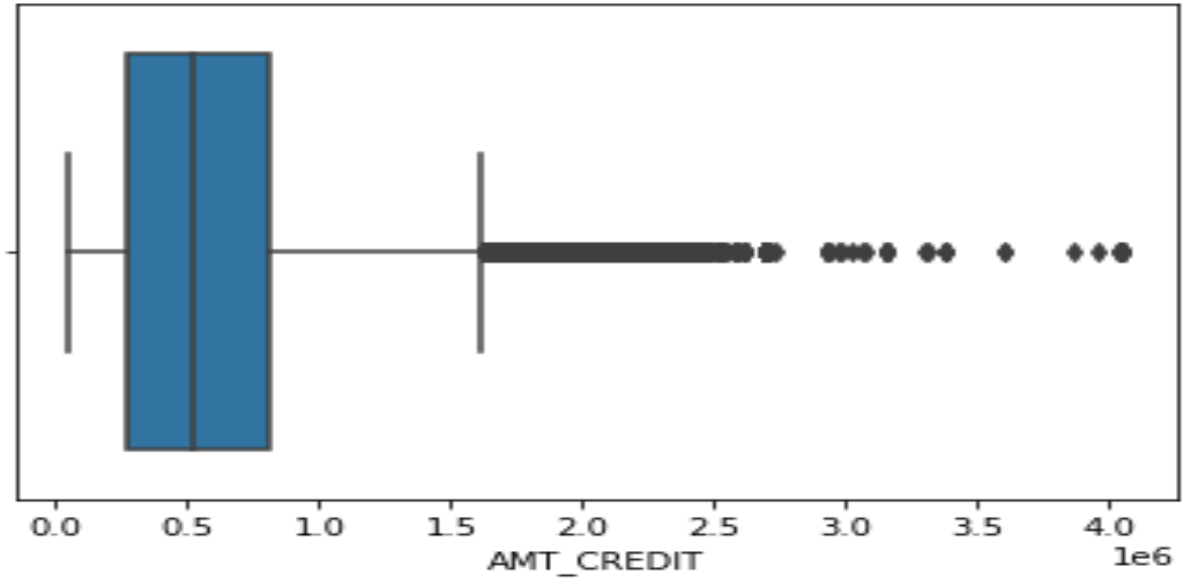


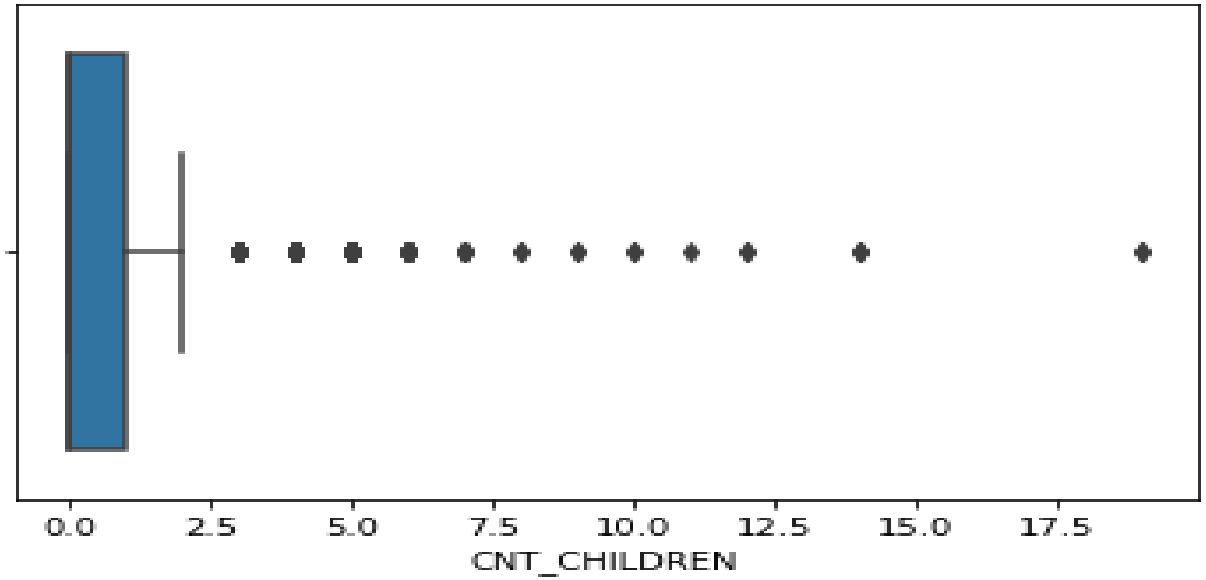
NUMBER OF OUTLIERS ARE MORE IN AMT_INCOME_TOTAL VARIABLE WHICH INDICATES FEW APPLICANTS HAVE THE HIGH INCOME THEN THE OTHERS



AMT_CREDIT VARIABLE HAS SOME OUTLIERS

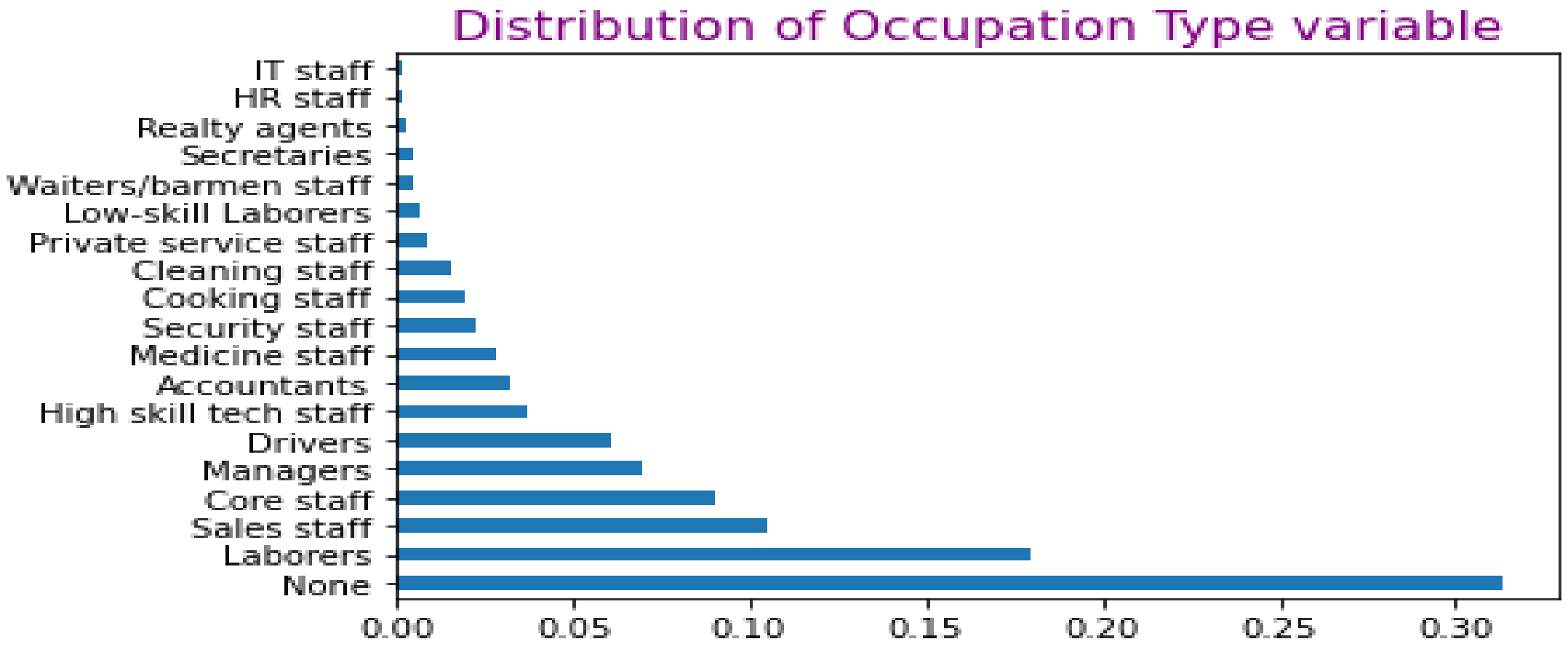
CNT_CHILDREN HAS SOME OUTLIERS



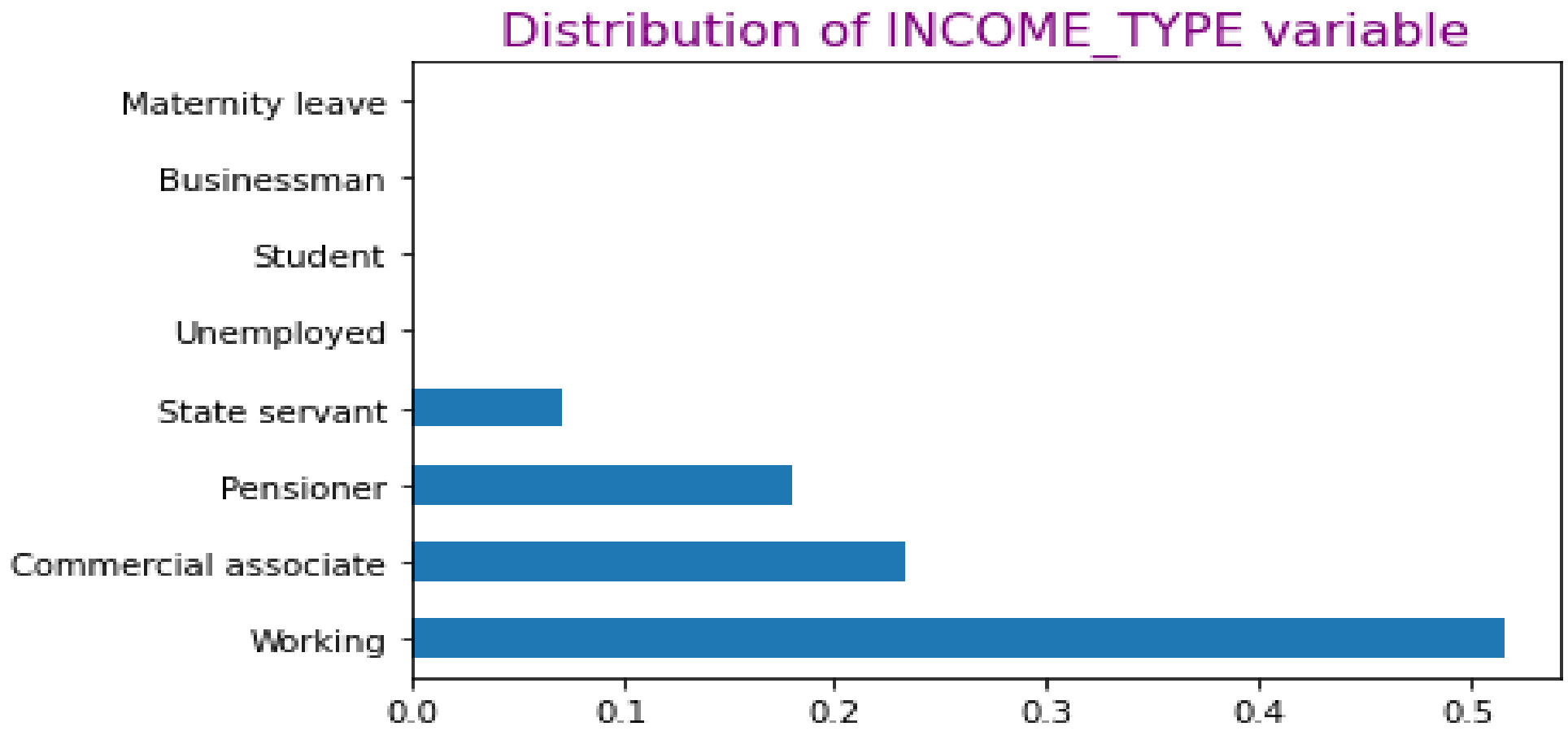


Univariate

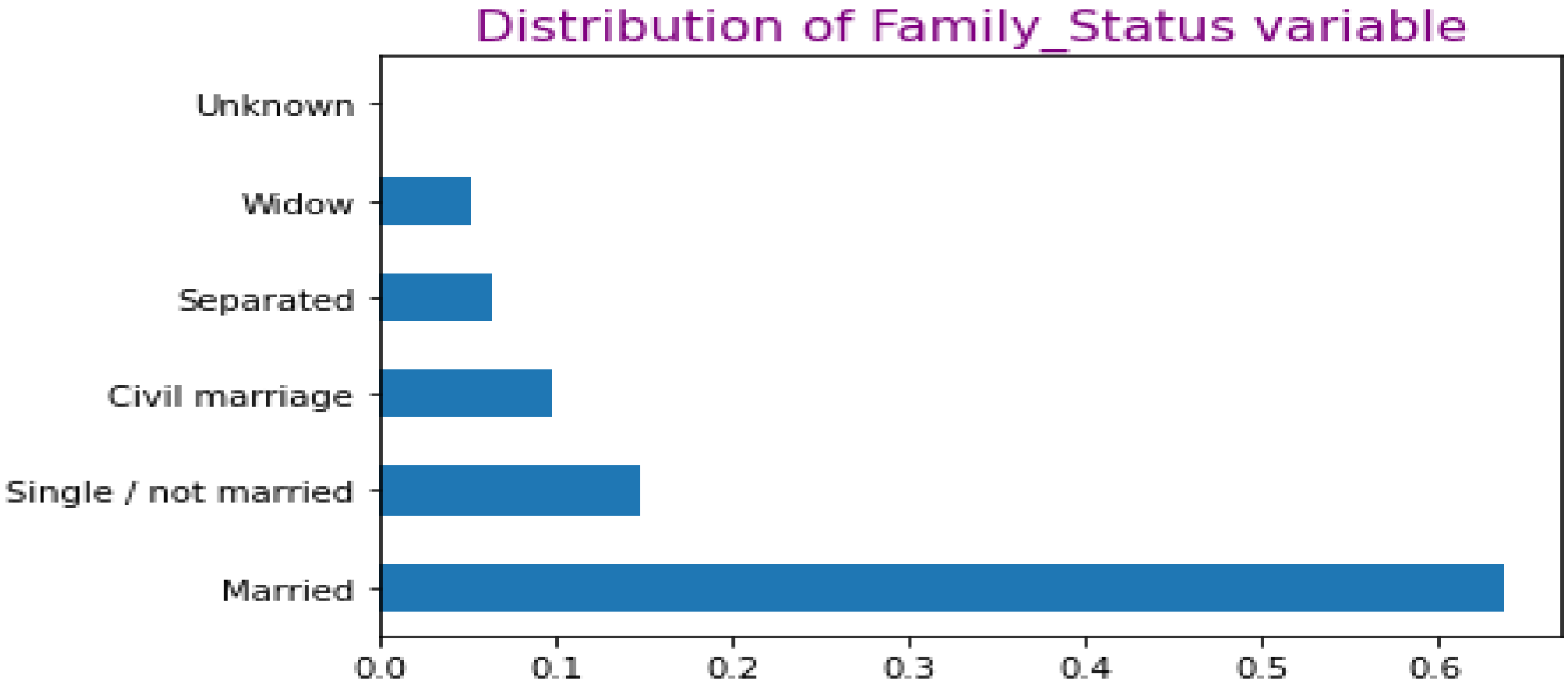
Most of the loan application is received by the people where the category is missing



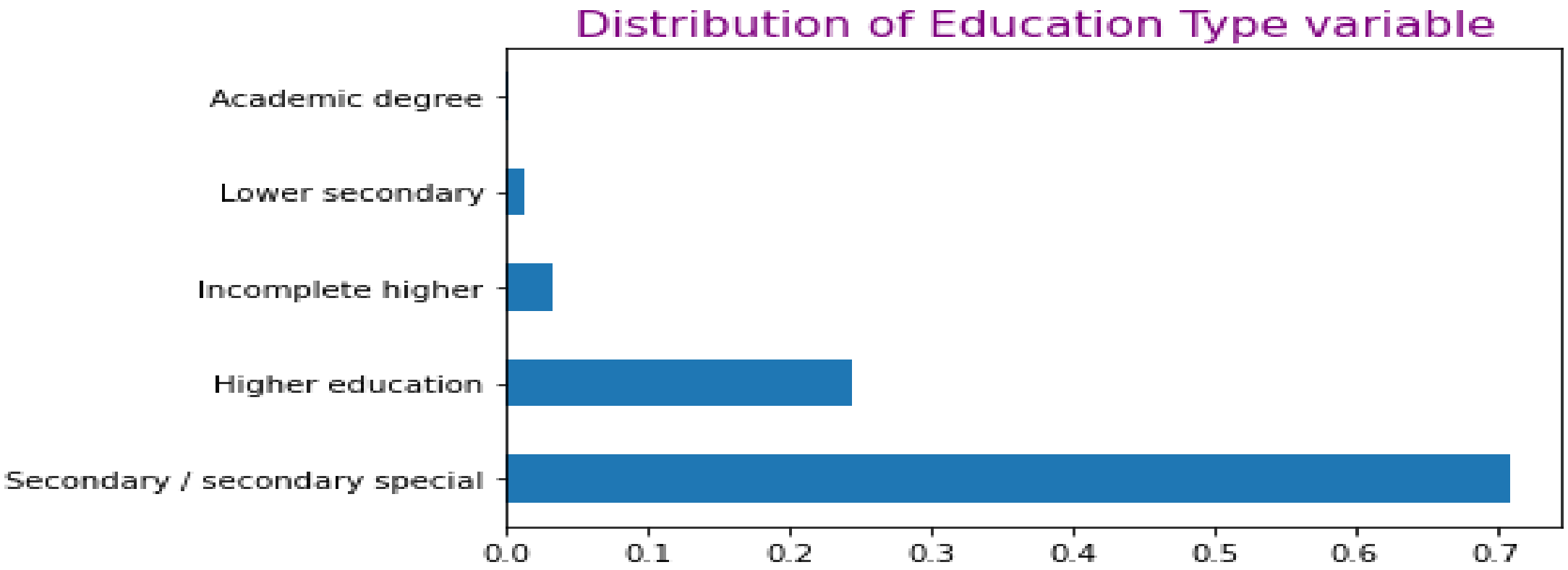
Observation : Here most loan application is received by the working category and least from the maternity leave and businessman category



Loan applicants are more who are in Family and least of widow category

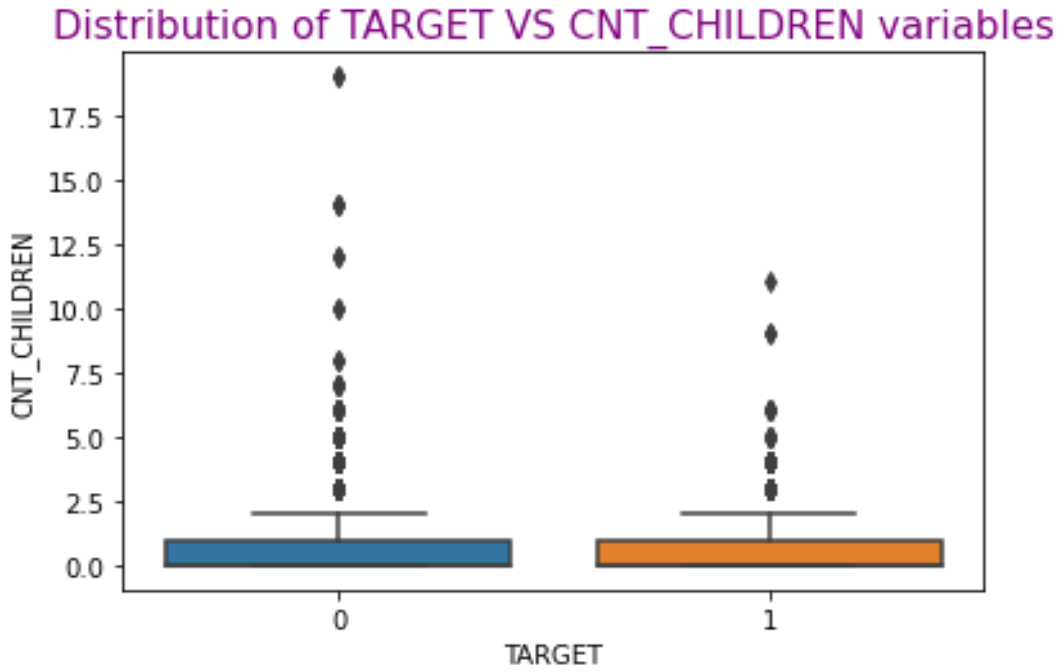


Application received more from Secondary/secondary special category then the other Education type

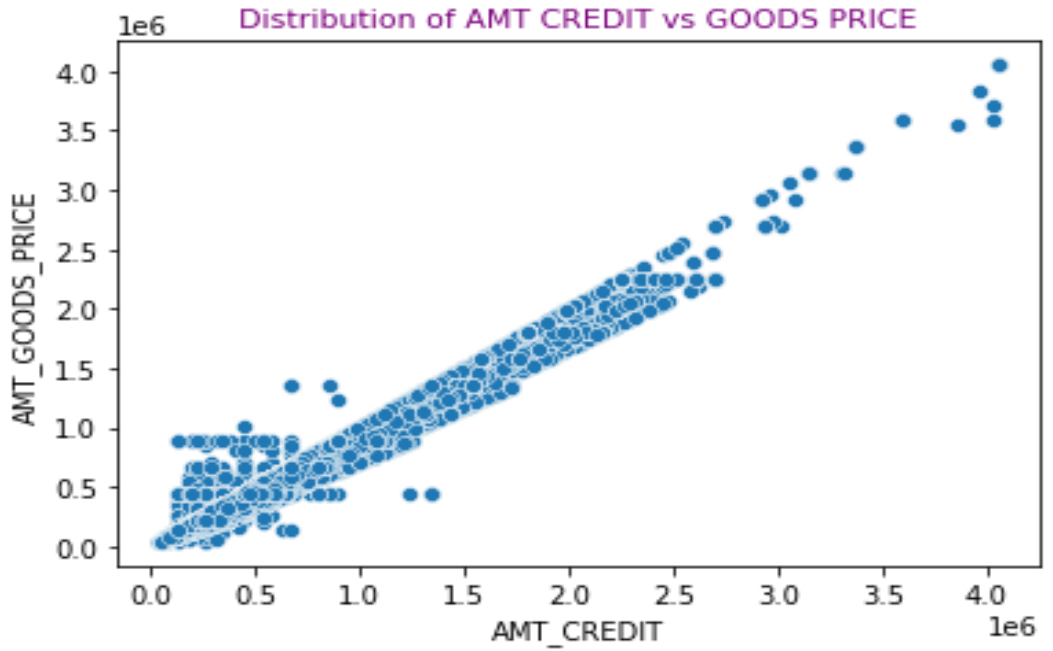


Bivariate & Multivariate

OBSERVATION: DEFAULTERS ARE MORE IN NO CHILDREN COUNT



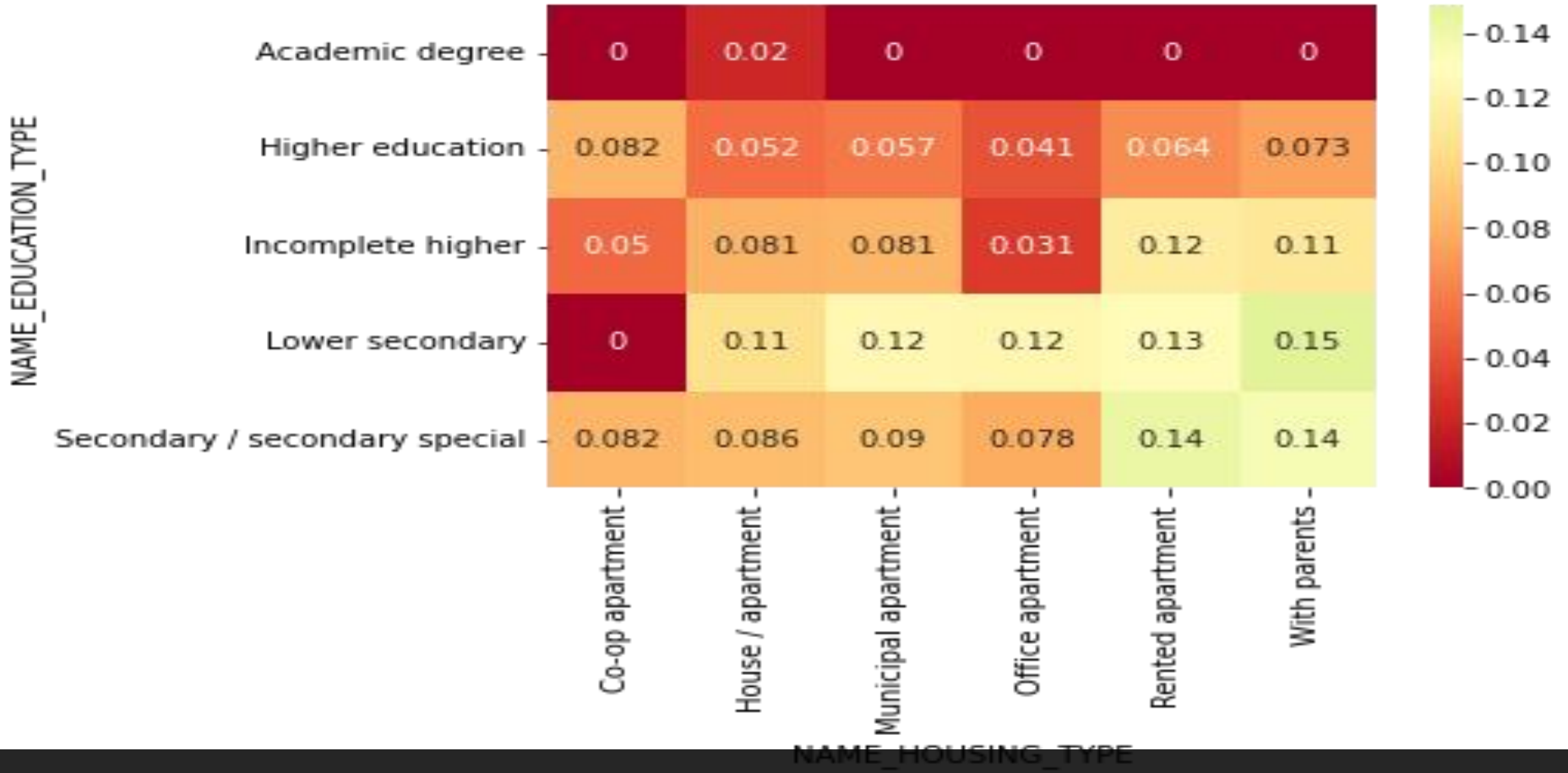
OBSERVATION : MORE THE GOODS PRICE MORE THE CREDIT , POSITIVE RELATION



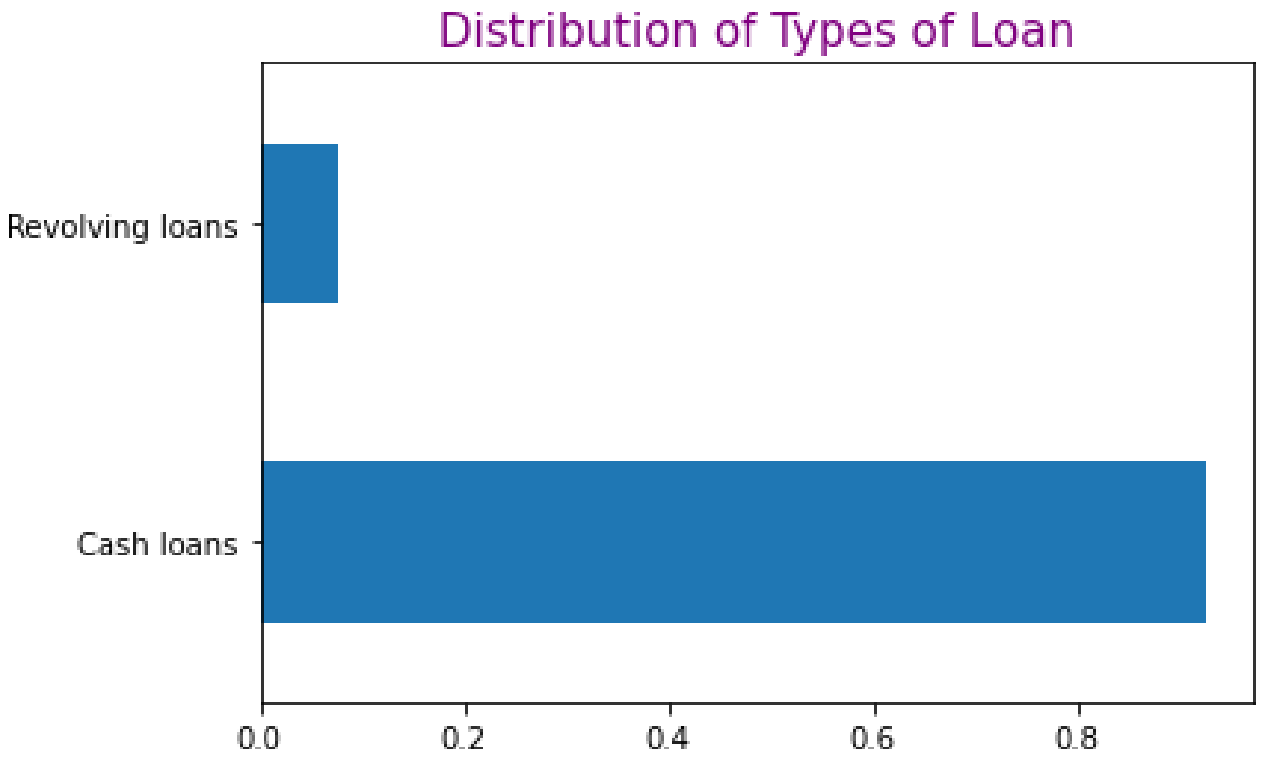
Pivot table – Application data for making the Heatmap

NAME_HOUSING_TYPE NAME_EDUCATION_TYPE	Co-op apartment	House / apartment	Municipal apartment	Office apartment	Rented apartment	With parents
Academic degree	0.000000	0.020408	0.000000	0.000000	0.000000	0.000000
Higher education	0.081800	0.052062	0.057464	0.041199	0.063970	0.073232
Incomplete higher	0.050000	0.081281	0.081433	0.03092	0.115385	0.110132
Lower secondary	0.000000	0.105115	0.122744	0.121951	0.126582	0.148649
Secondary / secondary special	0.081996	0.086323	0.089675	0.078209	0.142356	0.135323

Heatmap of NAME_EDUCATION_TYPE vs NAME_HOUSING_TYPE vs TARGET variables



Merging Data sets



After merging, it can be seen application of Cash Loan more then the Revolving Loans

Percentage of Applicant type

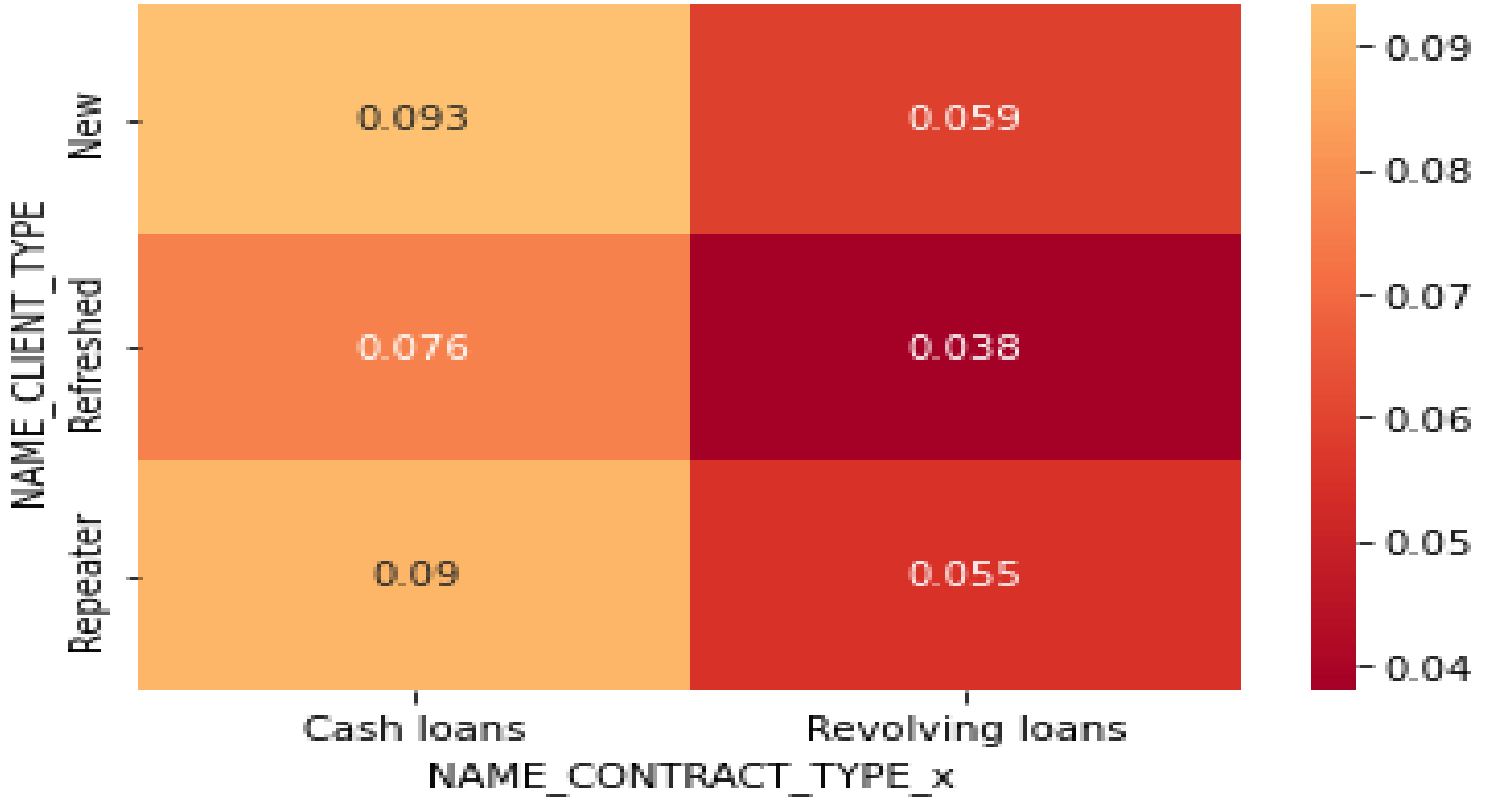
Loan Type	Cash loans	Revolving loans
NAME_CLIENT_TYPE		
New	0.093238	0.059055
Refreshed	0.076321	0.038029
Repeater	0.089531	0.055217

Repeater = 73.51%

New = 18.36%

Refreshed = 8.13%

Heatmap Client Type vs Loan Type vs Target



Summary

- Academic Degree has less defaulters
- Applicants with zero to two children tend to repay the loan
- Cash loan applicants are more then the revolving loan type
- Applicants with high Income are less likely to be default