

EDA Bank Credit Case Study

BY:
ABHISHEK JAIN

NEETHU MARIA TOM

Problem Statement

Loan providing company wants to understand the driving factors behind loan default, i.e. the variables which are strong indicators of default. 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.

Analysis Approach

Analysis approach focuses on identifying key factors/variables that could possibly influence a loan applicant to be a defaulter or not. These insights can help bank to approve or reject applicants.

This analysis also aims to identify top correlations between the available data variables.

Steps followed for analysis:

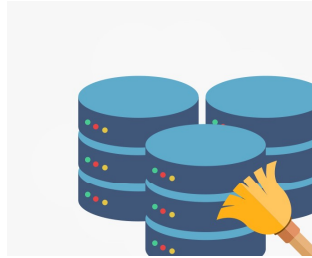
- Null values and outliers were identified , handled and sufficient data cleaning was performed.
- Created new columns for easy analysis when needed.
- Found out imbalance in the data.
- Derived insights after conducting univariate, bivariate and multivariate analysis.

Steps carried out in EDA for Credit Case Study

Data Understanding



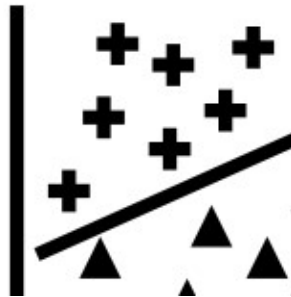
Data Cleaning



Data Analysis(Univariate/Bivariate)



Correlation Analysis



Insights & Conclusions

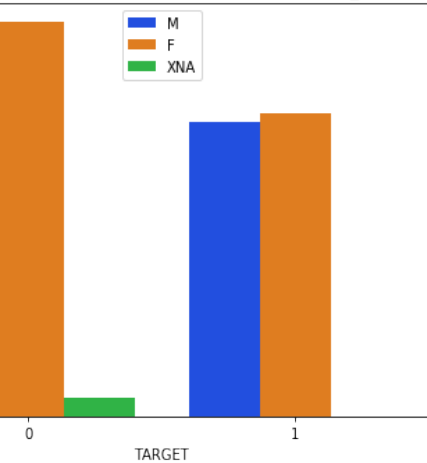


Data Understanding

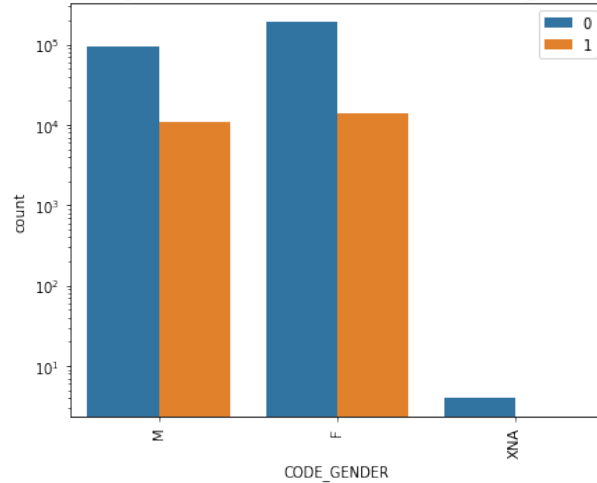
Applicant Demographics	Loan related information
AMT_INCOME_TOTAL	AMT_CREDIT
CODE_GENDER	
NAME_FAMILY_STATUS	AMT_ANNIUTY
NAME_EDUCATION_TYPE	
REG_CITY_NOT_LIVE_CITY	AMT_GOODS_PRICE
ORGANIZATION_TYPE	
NAME_INCOME_TYPE	NAME_CONTRACT_TYPE
OCCUPATION_TYPE	

Association of variable gender with TARGET

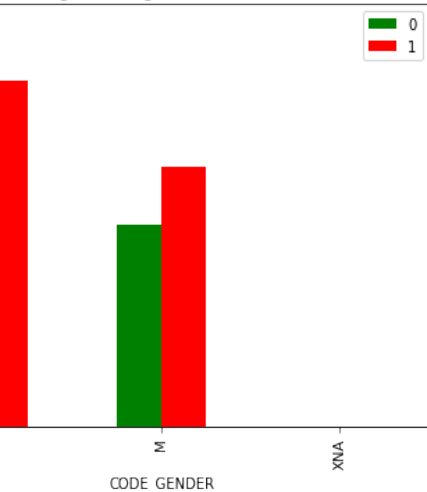
count based on TARGET values for CODE_GENDER



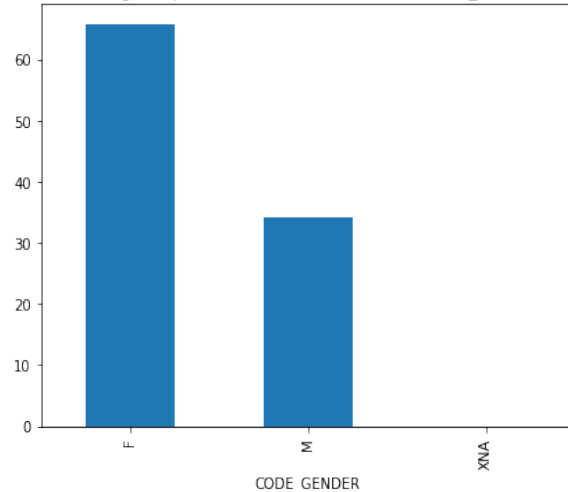
Plotting total count based on TARGET values for CODE_GENDER



percentage of categories based on TARGET values



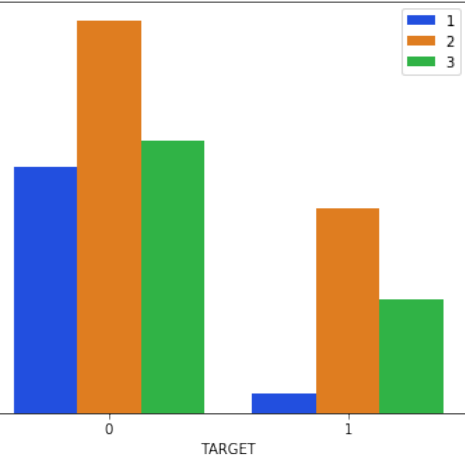
Plotting composition data for the column: CODE_GENDER



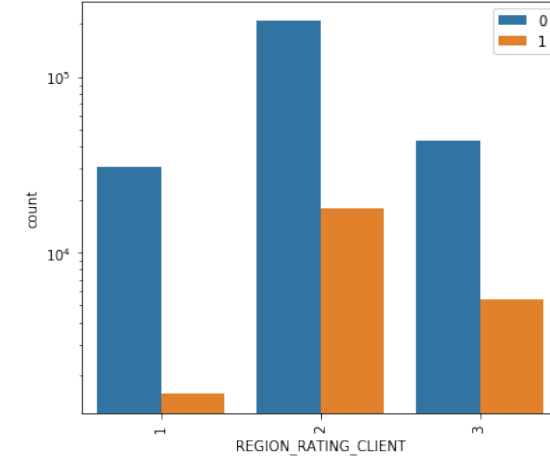
There are nearly twice female applicants than male applicants. Even though there are defaulters from both genders proportional to the gender counts, male applicants tend to default more than female applicants. The count of defaulters for male and females are same.

Association of variable REGION_RATING_CLIENT with TARGET.

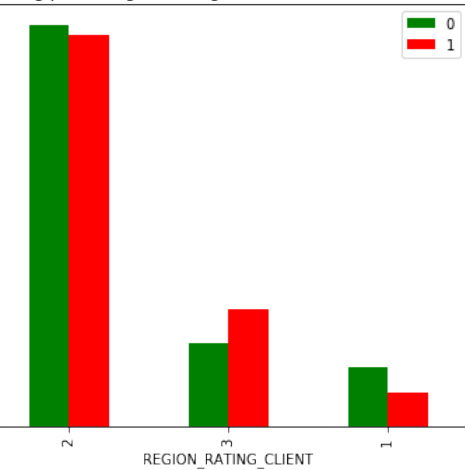
Plotting total count based on TARGET values for REGION_RATING_CLIENT



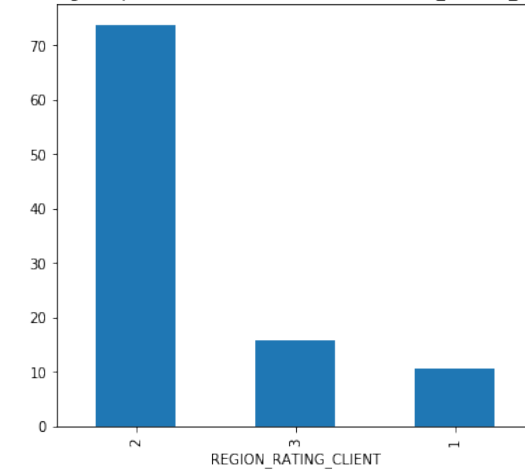
Plotting total count based on TARGET values for REGION_RATING_CLIENT



Plotting percentage of categories based on TARGET values



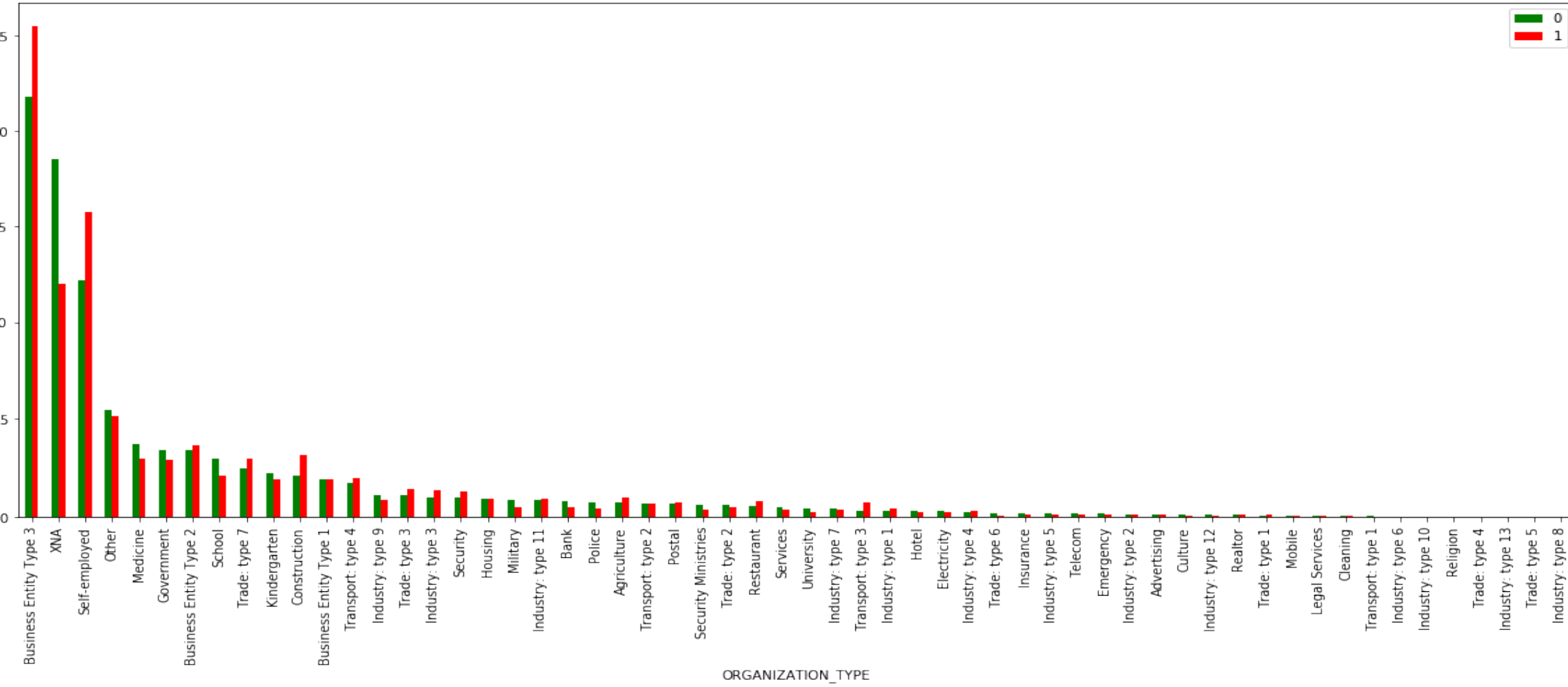
Plotting composition data for the column: REGION_RATING_CLIENT



Clients from Region2 seems to be availing and getting more loans.
People from region with rating 3 have high default percentage.
Less number of people default from region rated 1.

Association of variable ORGANIZATION_TYPE with TARGET.

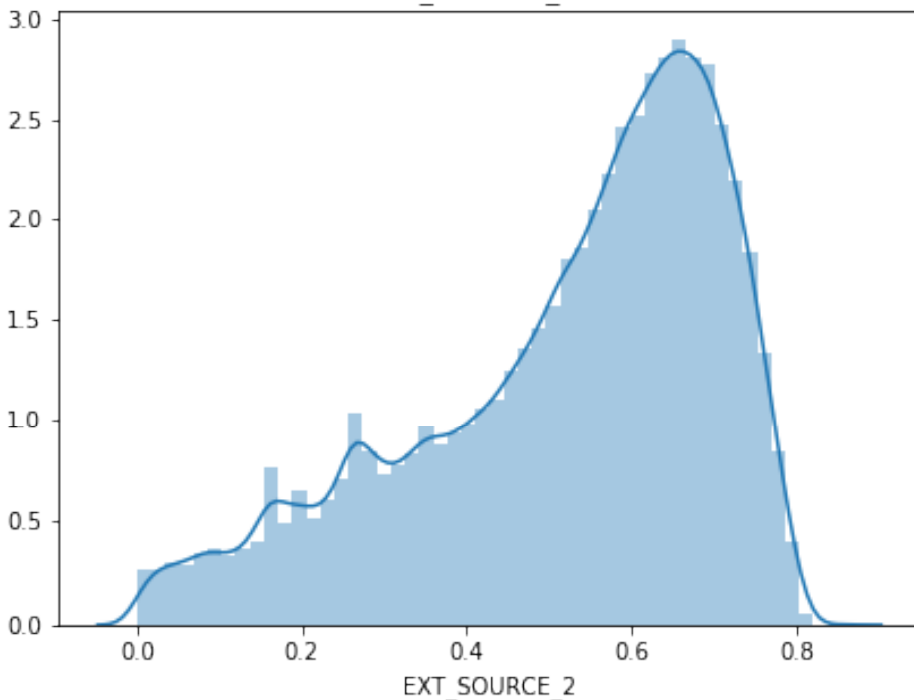
Plotting percentage of categories based on TARGET values



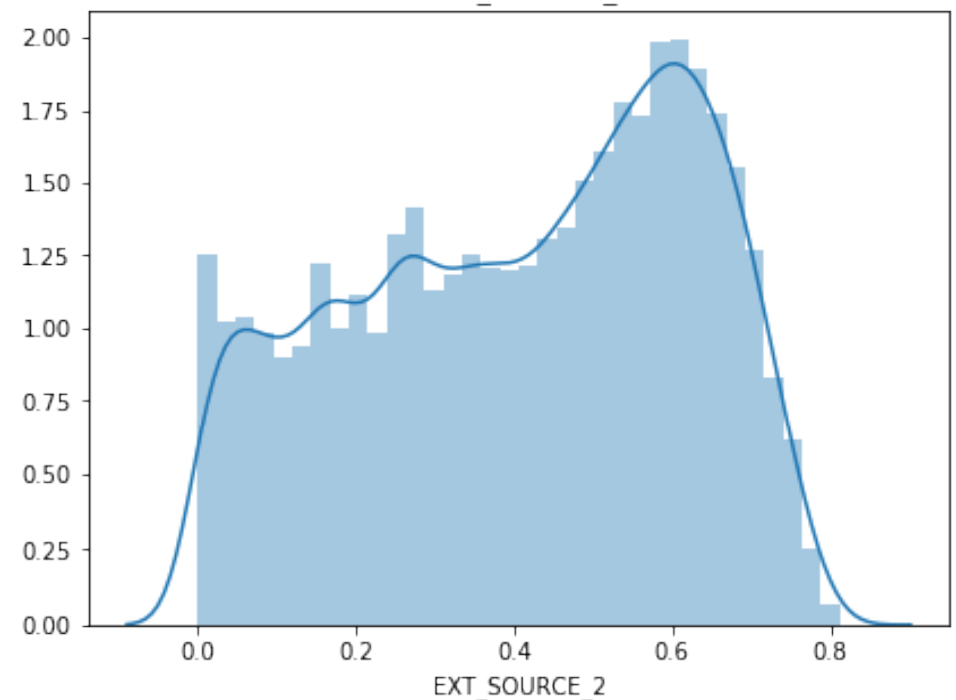
From the ORGANIZATION_TYPE column, Business Entity Type 3 and Self-employed tend to be having high default percentage. Even though applicants from construction organization are less, it also has considerable defaulted

Association of variable EXT_SOURCE_2 with TARGET.

Distribution of EXT_SOURCE_2 for non defaulter



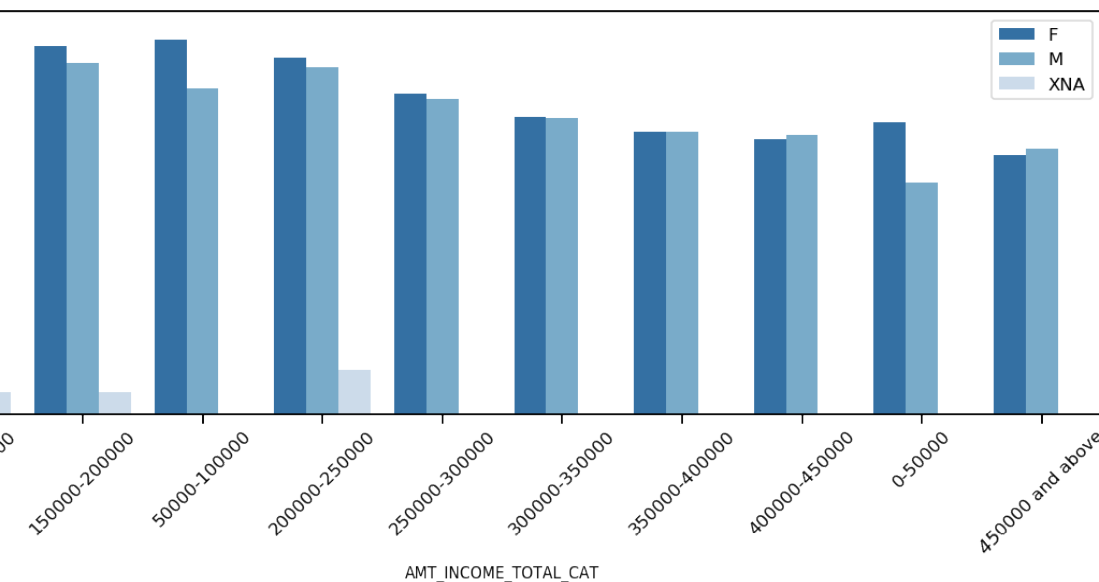
Distribution of EXT_SOURCE_2 for defaulter



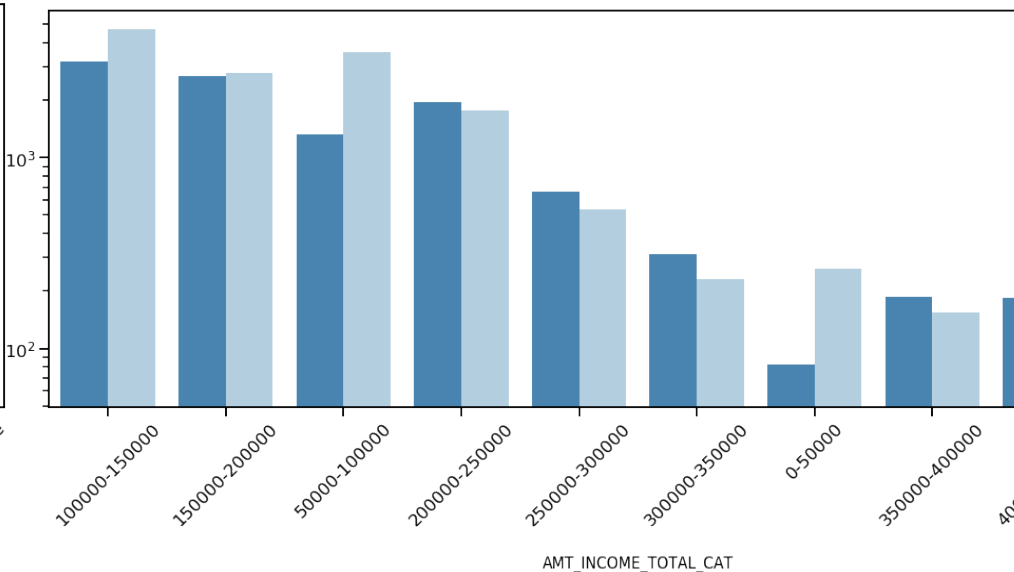
In case of EXT_SOURCE_2 score non-defaulter's distribution, number of applicants with score between 0 and 0.3 are high compared to defaulters. This means that applicants with score < 0.3 has higher chance to default.

Income distribution for Defaulters and Non-Defaulters

Distribution of total income for Non-Defaulters



Distribution of total income for Defaulters



Defaulters:

For Defaulters, Female counts are higher than male.

Income range from 100000 to 150000 is having higher income.

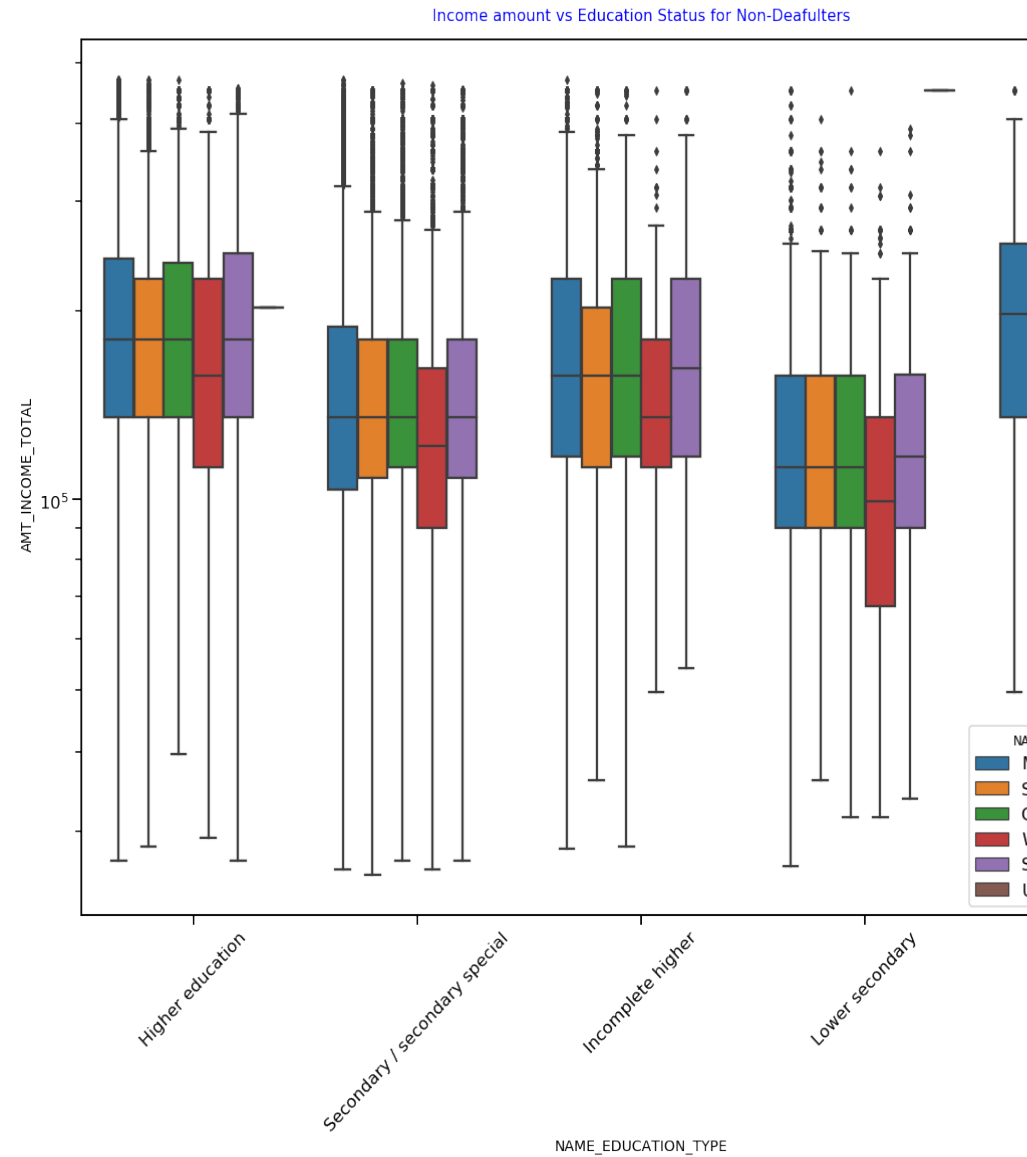
Applicants have more income in almost all ranges than male applicants.

Defaulters:

- For Defaulters, Female counts are nearly equal for most categories as compared to male.
- Income range from 100000 to 150000 is having a great
- Lowest counts in the income ranges for males and females above 450000.

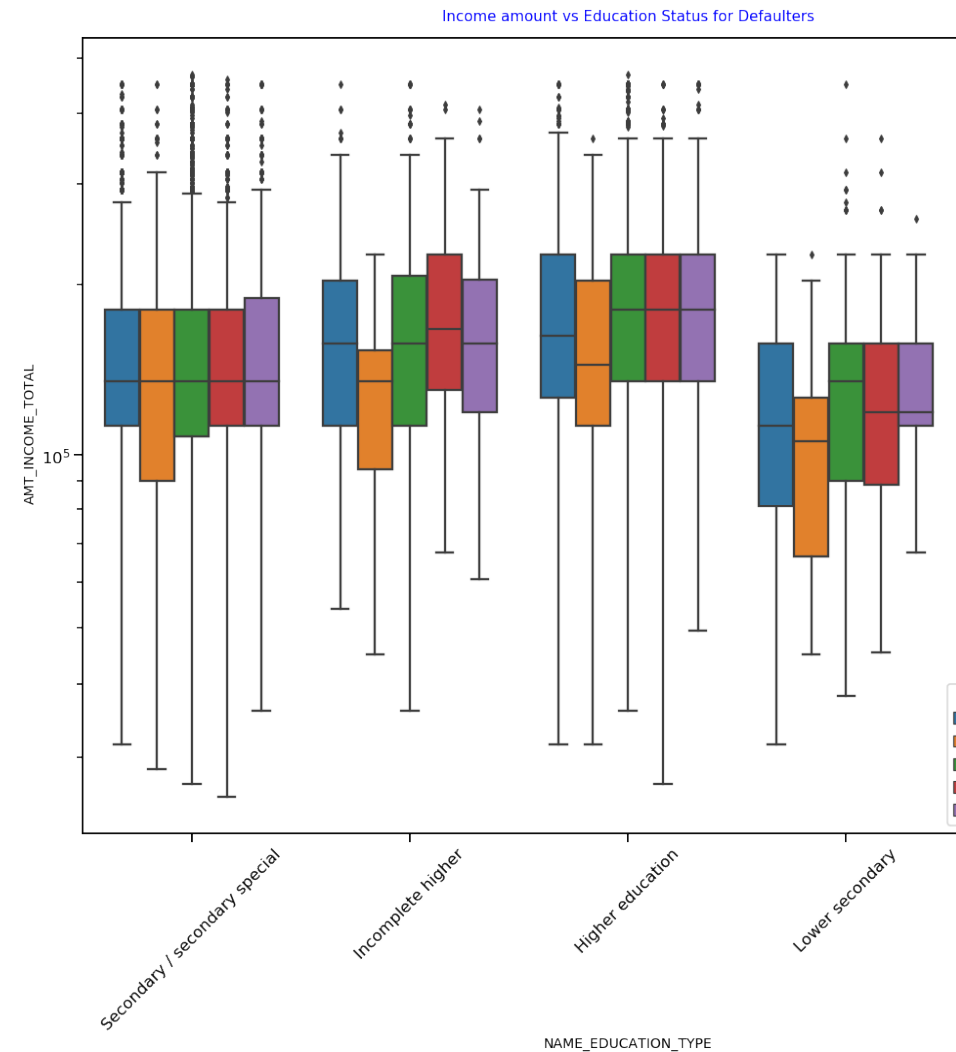
Analysis for income vs education status for non-defaulters

1. Least number of outliers in income are present for applicants with Academic degree & the median income is mostly higher than that of all other education categories.
2. When compared to others, people with Lower secondary education and widow status have less income amount.
3. Except for Academic degree holders, in all other education categories, median income is almost same for all family statuses with widow as an exception.



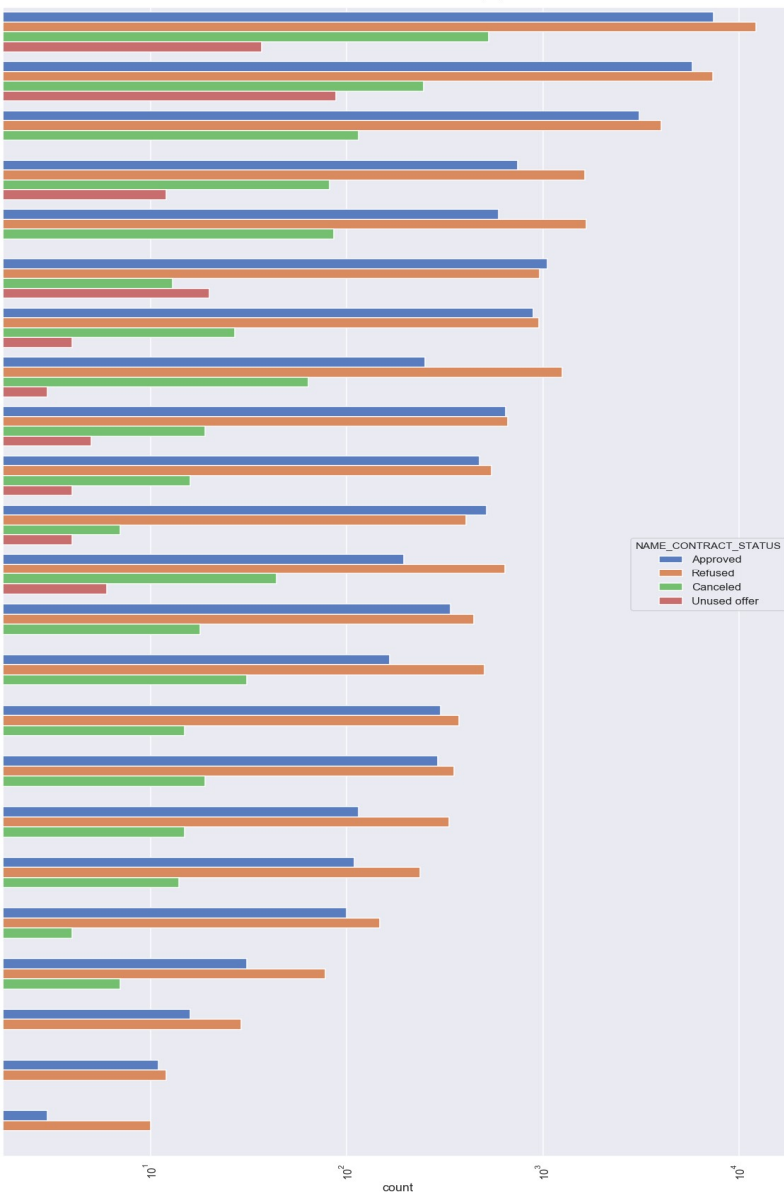
Analysis for income vs education status for defaulters

1. People with Academic degree tend to default less. Even then **married** people with academic category defaults.
2. The income amount is higher for Academic degree than Higher education with minimum number of outliers.
3. Lower secondary have less income amount than others.
4. More number of outliers are present for Married status people who are secondary/secondary special qualified.



Distribution of Loan Purpose hued to contract status

Distribution of contract status with loan purposes



Inferences:

- In almost all categories of loan purpose, most of the loans were rejected by the bank except for education, everyday expense, purchase of electronic equipment.
- Mostly applicants has cancelled a loan applied for buying something, urgent needs and repairs.
- There are equal number of approvals & refusals in “Education” loan requests
- When the applicant refuses to name the goal of taking loan, bank mostly rejects it.
- Around 10000 requests for “Repair” type loans and “Payment on other loans” have been refused & holds highest rejections.

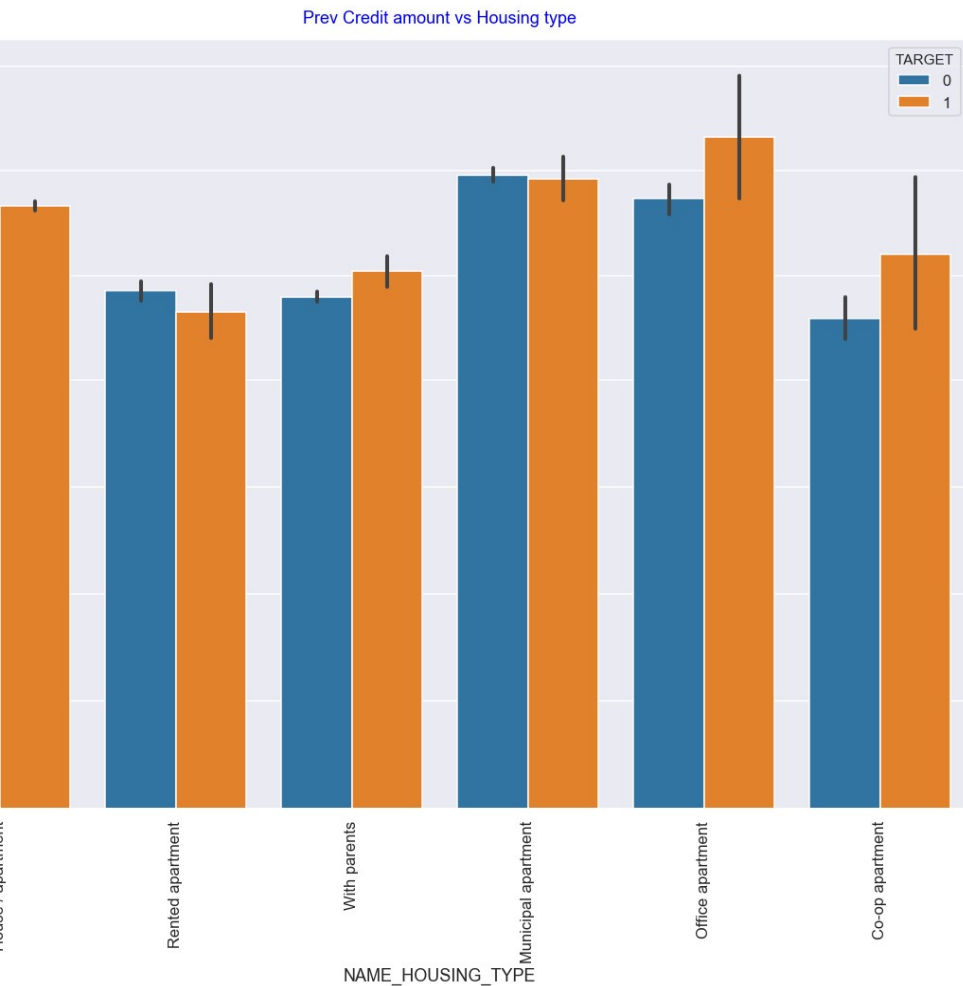
Relation between loan status and income type

Loans contract status shows that mostly status are approved for working, commercial associate and pensioner. Rejections are also higher in case of working, commercial associate and pensioner. Approvals are more than that of rejections.

People from Students, unemployed and maternity leave categories apply least number of loans.

NAME_INCOME_TYPE	Commercial associate	2e+05	6e+04	5.8e+04	5.1e+03
	Maternity leave	10	2	3	1
	Pensioner	1.7e+05	5.8e+04	4.4e+04	1.9e+03
	State servant	6.2e+04	1.6e+04	1.6e+04	1.5e+03
	Student	20	3	1	
	Unemployed	68	16	38	1
	Working	4.6e+05	1.3e+05	1.3e+05	1.4e+04
		Approved	Canceled	Refused	Unused offer
		NAME_CONTRACT_STATUS			

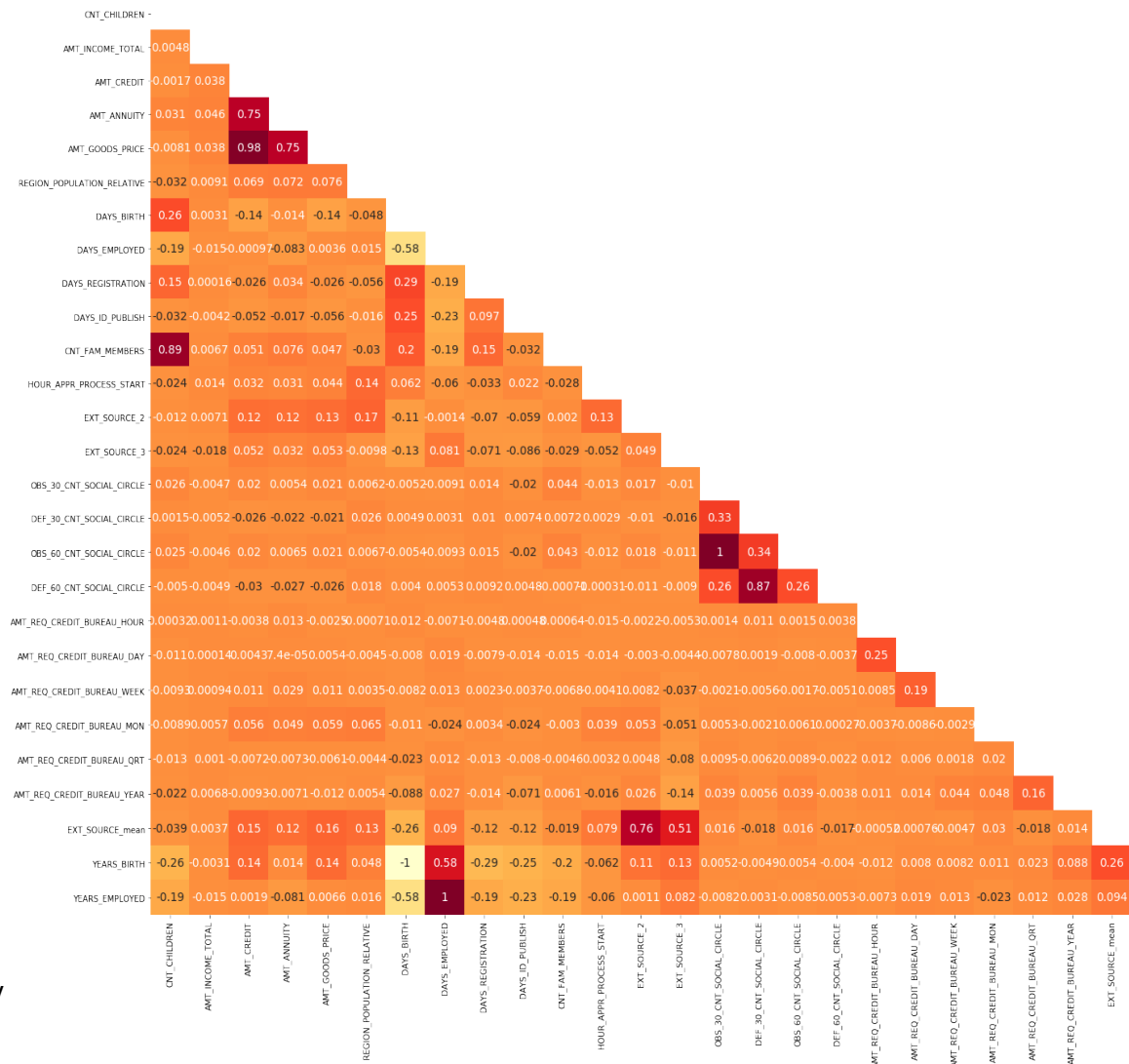
relation between housing type and amount credited with the target variable



1. Bank must be careful when giving credit to housing types, with parents, office apartment and co-op apartment. Among those categories default applicants has high credit than non-defaulters.
2. Mean credit amount for House/apartment, municipal apartment for both defaulters and non defaulters are the similar.

Correlation between variables leading to Loan Defaulter & Non-Defaulter:

en AMT_CREDIT,AMT_ANNUITY, AMT_GOODS_PRICE had correlation greater than 0.3 with AMT_INCOME_TOTAL in case of non-defaulters, it is less than 0.04 in case of defaulters. This could mean that for defaulters even if AMT_INCOME_TOTAL increases, there will not be a corresponding increase in AMT_CREDIT,AMT_ANNUITY,AMT_GOODS_PRICE. Similar to non-defaulter's correlation, in case of defaulters AMT_CREDIT has high correlation with AMT_ANNUITY and AMT_GOODS_PRICE. In case of correlation between DAYS_EMPLOYED and DAYS_BIRTH, it is less for defaulters compared to non-defaulters. EXT_SOURCE_mean has slight correlation with AMT_CREDIT,AMT_ANNUITY,AMT_GOODS_PRICE, AMT_INCOME_TOTAL, REGION_POPULATION_RELATIVE and DAYS_BIRTH for both defaulters and non-defaulters. OBS_30_CNT_SOCIAL_CIRCLE and OBS_60_CNT_SOCIAL_CIRCLE has a correlation coefficient of 1, which could mean that both variables contain exactly the same data or it could mean that within 60 days of duration the data is not changing for a same social circle.



Correlation between variables

s:

From the above correlation we can say that Final credit amount on the previous application(AMT_CREDIT_P) has a high correlation with AMT_APPLICATION , for the month in which client initially applied.

As the count of children , less is the years employed.

For consumer loans, higher the price of the goods given as collateral, higher the price for loan, higher is the loan credited given to the consumer.

As SK_ID_PREV is employed has a negative correlation between AMT_APPLICATION process start from previous data. AMT_APPLICATION process happens when the client is employed for more days.

As the Number of enquiries to Credit Bureau about the loan in the day year, higher is the days took for decision.

TARGET	1	0.02	-0.0022	-0.024	-0.0074	-0.033	-0.035	0.075	-0.048	0.043	0.051	0.015	-0.022	-0.15	-0.17	0.015	0.032	0.014	0.029	0.00022	0.004	-0.00069	-0.015	-0.0052	0.0021	-0.22	-0.075	-0.05	0.002	-0.0056	-0.0024	-0.028	0.04	
CNT_CHILDREN	0.02	1	0.012	0.0028	0.029	-0.001	-0.022	0.36	-0.25	0.19	-0.013	0.88	-4.1e-05	-0.023	-0.042	0.018	0.00026	0.017	-0.0014	-0.00011	-0.002	0.0032	-0.01	-0.013	-0.056	-0.075	-0.36	-0.25	-0.0028	-0.034	-0.035	0.017	-0.013	
AMT_INCOME_TOTAL	-0.0022	0.012	1	0.17	0.21	0.17	0.077	0.026	-0.067	0.028	0.014	0.014	0.04	0.065	-0.039	-0.0099	-0.013	-0.0098	-0.013	0.0025	0.0046	0.0056	0.03	0.0088	0.027	0.031	-0.026	-0.067	0.00034	0.071	0.071	0.043	0.0039	
AMT_CREDIT_A	-0.024	0.0028	0.17	1	0.76	0.99	0.086	-0.048	-0.061	0.0094	0.0092	0.063	0.047	0.12	0.036	0.0071	-0.018	0.0071	-0.022	-0.0027	0.0053	0.00021	0.059	0.023	-0.032	0.13	0.048	-0.059	-0.00043	0.12	0.12	0.047	-0.06	
AMT_ANNUITY	-0.0074	0.029	0.21	0.76	1	0.76	0.1	0.029	-0.11	0.046	0.03	0.084	0.051	0.11	0.014	-0.0051	-0.02	-0.0047	-0.022	0.0024	0.0021	0.011	0.034	0.0023	-0.011	0.1	-0.029	-0.11	0.00084	0.11	0.11	0.06	-0.035	
AMT_GOODS_PRICE	-0.033	-0.001	0.17	0.99	0.76	1	0.088	-0.046	-0.059	0.012	0.0064	0.061	0.057	0.12	0.04	0.0088	-0.018	0.0087	-0.022	-0.0018	0.006	0.00068	0.061	0.023	-0.034	0.14	0.046	-0.056	-0.00025	0.12	0.12	0.053	-0.061	
REGION_POPULATION_RELATIVE	-0.035	-0.022	0.077	0.086	0.1	0.088	1	-0.04	0.00082	-0.054	-0.0065	-0.021	0.16	0.19	-0.0066	-0.006	0.0071	-0.0054	0.0044	-0.0027	-0.000390	0.0018	0.075	0.001	0.0065	0.14	0.04	0.000980	0.0088	0.045	0.045	0.15	-0.026	
DAYS_BIRTH	0.075	0.36	0.026	-0.048	0.029	-0.046	-0.04	1	-0.63	0.33	0.26	0.33	0.086	-0.09	-0.17	0.015	0.001	0.015	0.0032	0.0051	0.0017	0.00028	0.0023	-0.021	-0.12	-0.27	-1	-0.63	-0.0016	-0.08	-0.078	0.12	0.025	
DAYS_EMPLOYED	-0.048	-0.25	-0.067	-0.061	-0.11	-0.059	0.00082	-0.63	1	-0.21	-0.28	-0.25	-0.088	-0.016	0.11	0.0022	0.016	0.0024	0.014	-0.0046	-0.0024	-0.00073	-0.035	0.02	0.08	0.1	0.63	1	-0.001	0.0095	0.0038	-0.11	0.015	
DAYS_REGISTRATION	0.043	0.19	0.028	0.0094	0.046	0.012	-0.054	0.33	-0.21	1	0.1	0.18	-0.011	-0.063	-0.098	0.014	0.0071	0.014	0.0064	-0.0048	-0.0018	-0.0023	-0.0083	-0.032	-0.038	-0.13	-0.33	-0.21	-0.00068	-0.013	-0.013	0.00026	0.022	
DAYS_ID_PUBLISH	0.051	-0.013	0.014	0.0092	0.03	0.0064	-0.0065	0.26	-0.28	0.1	1	0.0025	0.031	-0.043	-0.11	-0.012	0.0012	-0.012	0.0026	0.00072	-0.0035	-0.0055	-0.016	-0.017	-0.076	-0.11	-0.26	-0.28	0.0017	-0.012	-0.01	0.041	0.049	
CNT_FAM_MEMBERS	0.015	0.88	0.014	0.063	0.084	0.061	-0.021	0.33	-0.25	0.18	0.0025	1	-0.0037	-0.0099	-0.036	0.031	-0.00037	0.03	-0.0042	6.1e-05	0.0022	0.0044	-0.0079	-0.009	-0.043	-0.054	-0.33	-0.25	-0.0027	-0.0058	-0.0052	0.0083	-0.018	
HOURL_APPR_PROCESS_START_A	-0.022	-4.1e-05	0.04	0.047	0.051	0.057	0.16	0.086	-0.088	-0.011	0.031	-0.0037	1	0.15	-0.042	-0.0083	-0.0055	-0.008	-0.0081	-0.012	0.0019	0.00092	0.044	0.076	-0.01	0.082	-0.086	-0.088	-0.00047	0.0084	0.0079	0.39	-0.0032	
EXT_SOURCE_2	-0.15	-0.023	0.065	0.12	0.11	0.12	0.19	-0.09	-0.016	-0.063	-0.043	-0.0099	0.15	1	0.096	-0.019	-0.026	-0.019	-0.028	-0.0054	0.00056	0.0025	0.055	0.00095	-0.0037	0.74	0.09	-0.014	-0.0018	0.049	0.047	0.16	-0.037	
EXT_SOURCE_3	-0.17	-0.042	-0.039	0.036	0.014	0.04	-0.0066	-0.17	0.11	-0.098	-0.11	-0.036	-0.042	0.096	1	-0.0085	-0.033	-0.0087	-0.03	-0.0022	-0.0081	-0.025	-0.019	-0.031	-0.089	0.6	0.17	0.11	-0.0077	-0.0091	-0.013	-0.029	-0.093	
OBS_30_CNT_SOCIAL_CIRCLE	0.015	0.018	-0.0099	0.0071	-0.0051	0.0088	-0.006	0.015	0.0022	0.014	-0.012	0.031	-0.0083	-0.019	-0.0085	1	0.3	1	0.23	7.3e-05	-0.0024	-0.0014	0.0011	0.0055	0.032	-0.021	-0.015	0.0023	0.00036	0.0078	0.078	-0.015	-0.017	
DEF_30_CNT_SOCIAL_CIRCLE	0.032	0.00026	-0.013	-0.018	-0.02	-0.018	0.0071	0.001	0.016	0.0071	0.0012	-0.00037	-0.0055	-0.026	-0.033	0.3	1	0.3	0.86	4.4e-05	-0.0026	-0.0038	-0.0017	-0.0019	0.012	-0.039	-0.001	0.016	-0.00041	-0.0015	-0.00051	-0.0084	-0.0048	
OBS_60_CNT_SOCIAL_CIRCLE	0.014	0.017	-0.0098	0.0071	-0.0047	0.0087	-0.0054	0.015	0.0024	0.014	-0.012	0.03	-0.008	-0.019	-0.0087	1	0.3	1	0.23	4.9e-05	-0.0026	-0.0013	0.0011	0.005	0.033	-0.021	-0.015	0.0025	0.00033	0.008	0.008	-0.015	-0.018	
DEF_60_CNT_SOCIAL_CIRCLE	0.029	-0.0014	-0.013	-0.022	-0.022	-0.022	0.0044	0.0032	0.014	0.0064	0.0026	-0.0042	-0.0081	-0.028	-0.03	0.23	0.86	0.23	1	-0.0012	-0.0033	-0.0042	-0.0032	-7.8e-05	0.011	-0.04	-0.0032	0.014	-0.00089	0.0037	-0.003	-0.0097	-0.0006	
AMT_REQ_CREDIT_BUREAU_HOUR	0.000220	0.00011	0.00025	-0.00027	0.00024	-0.00018	-0.00027	0.00051	-0.00046	-0.00048	0.00072	6.1e-05	-0.012	-0.0054	-0.0022	7.3e-05	4.4e-05	4.9e-05	-0.0012	1	0.21	0.0076	0.0026	0.003	0.0077	-0.0052	-0.0051	-0.00046	0.00021	0.0014	0.0013	-0.008	-0.0029	
AMT_REQ_CREDIT_BUREAU_DAY	0.004	-0.002	0.0046	0.0053	0.0021	0.006	-0.00039	0.0017	-0.0024	-0.0018	-0.0035	-0.0022	0.0019	-0.00056	-0.0081	-0.0024	-0.0026	-0.0026	-0.0033	0.21	1	0.19	-0.0031	0.0017	0.0086	-0.005	-0.0017	-0.0023	-0.00089	0.0036	0.0033	0.00066	0.0072	
AMT_REQ_CREDIT_BUREAU_WEEK	-0.00069	0.0032	0.0056	0.00021	0.011	0.000680	0.00180	0.00280	0.00073	-0.0023	-0.0055	0.0044	0.00092	0.0025	-0.025	-0.0014	-0.0038	-0.0013	-0.0042	0.0076	0.19	1	-0.011	-0.0056	0.037	-0.0088	-0.000280	0.000670	0.000120	0.000330	0.00066	0.0014	0.0078	
AMT_REQ_CREDIT_BUREAU_MON	-0.015	-0.01	0.03	0.059	0.034	0.061	0.075	0.0023	-0.035	-0.0083	-0.016	-0.0079	0.044	0.055	-0.019	0.0011	-0.0017	0.0011	-0.0032	0.0026	-0.0031	-0.011	1	0.007	0.021	0.03	-0.0023	-0.034	0.00041	0.025	0.026	0.045	-0.031	
AMT_REQ_CREDIT_BUREAU_QRT	-0.0052	-0.013	0.0088	0.023	0.0023	0.023	0.001	-0.021	0.02	-0.0032	-0.017	-0.009	0.0076	0.00095	-0.031	0.0055	-0.0019	0.005	-7.8e-05	0.003	0.0017	-0.0056	0.007	1	0.11	-0.011	0.021	0.02	0.0022	0.0045	0.0024	-0.0017	0.068	
AMT_REQ_CREDIT_BUREAU_YEAR	0.0021	-0.056	0.027	-0.032	-0.011	-0.034	0.0065	-0.12	0.08	-0.038	-0.076	-0.043	-0.01	-0.0037	-0.089	0.032	0.012	0.033	0.011	0.0077	0.0086	0.037	0.021	1	0.11	1	-0.03	0.12	0.08	0.0094	0.068	0.075	-0.054	0.13
EXT_SOURCE_mean	-0.22	-0.075	0.031	0.13	0.1	0.14	0.14	-0.27	0.1	-0.13	-0.11	-0.054	0.082	0.74	0.6	-0.021	-0.039	-0.021	-0.04	-0.0052	-0.005	-0.0088	0.03	-0.011	-0.03	1	0.27	0.1	-0.0054	0.048	0.045	0.089	-0.084	
YEARS_BIRTH	-0.075	-0.36	-0.026	0.048	-0.029	0.046	0.04	-1	0.63	-0.33	-0.26	-0.33	-0.086	0.09	0.17	-0.015	-0.001	-0.015	-0.0032	-0.0051	-0.0017	-0.00028	-0.0023	0.021	0.12	0.27	1	0.63	0.0016	0.08	0.078	-0.12	-0.025	
YEARS_EMPLOYED	-0.05	-0.25	-0.067	-0.059	-0.11	-0.056	0.00098	-0.63	1	-0.21	-0.28	-0.25	-0.088	-0.014	0.11	0.0023	0.016	0.0025	0.014	-0.0046	-0.0023	-0.00067	-0.034	0.02	0.08	0.1	0.63	1	0.00097	0.011	0.005	-0.11	0.013	
SK_ID_PREV	0.002	-0.00028	0.000340	0.000430	0.000840	0.000250	0.00088	-0.0016	0.001	-0.00068	0.0017	-0.0027	-0.00047	-0.0018	-0.0077	0.00036	-0.000410	0.000330	0.000890	0.000210	-0.000890	0.00120	0.00041	0.0022	0.0094	-0.0054	0.0016	0.00097	1	0.0037	0.0041	-0.0026	0.018	
AMT_APPLICATION	-0.0056	-0.034	0.071	0.12	0.11	0.12	0.045	-0.08	0.0095	-0.013	-0.012	-0.0058	0.0084	0.049	-0.0091	0.0078	-0.0015	0.008	-0.0037	0.0014	0.0036	0.00033	0.025	0.0045	0.068	0.048	0.08	0.011	0.0037	1	0.96	-0.015	0.14	
AMT_CREDIT_P	-0.0024	-0.035	0.071	0.12	0.11	0.12	0.045	-0.078	0.0038	-0.013	-0.01	-0.0052	0.0079	0.047	-0.013	0.0078	-0.00051	0.008	-0.003	0.0013	0.0033	0.00066	0.026	0.0024	0.075	0.045	0.078	0.005	0.0041	0.98	1	-0.022	0.14	
HOURL_APPR_PROCESS_START_P	-0.028	0.017	0.043	0.047	0.06	0.053	0.15	0.12	-0.11	0.00026	0.041	0.0083	0.39	0.16	-0.029	-0.015	-0.0084	-0.015	-0.0097	-0.008	0.00066	0.0014	0.045	-0.0017	-0.054	0.089	-0.12	-0.11	-0.0026	-0.015	-0.022	1	-0.04	
DAYS_DECISION	0.04	-0.013	0.0039	-0.06																														

Conclusion:

- As per our observation, loans repayment is most likely to be done without any difficulties for men and women who are of following income range categories: 100000-150000.
- Business Entity Type3 has equal chances of being defaulters and non-defaulters.
- Working officials, commercial associates and pensioners are the potential customers a bank can give loans to.
- People with academic degree has high income compared to other educational background and they tend to default less. Even then **married** people within academic degree category can become defaulters.