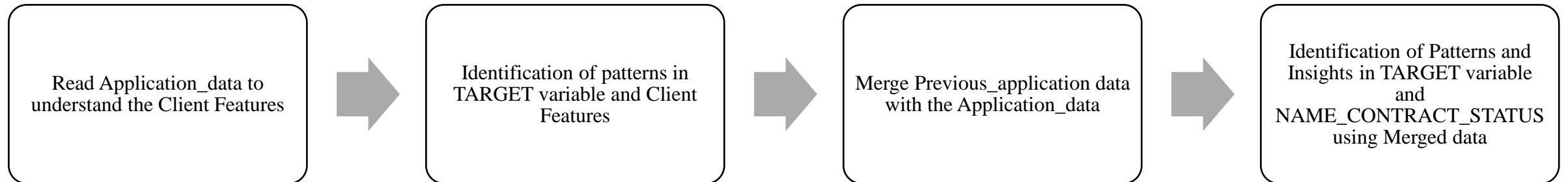


Credit EDA Case Study

Should you get the loan? Will you pay the loan on time?

Problem Statement: Identify Driving factors behind Loan Default

Approach:



CLIENT DEMOGRAPHICS

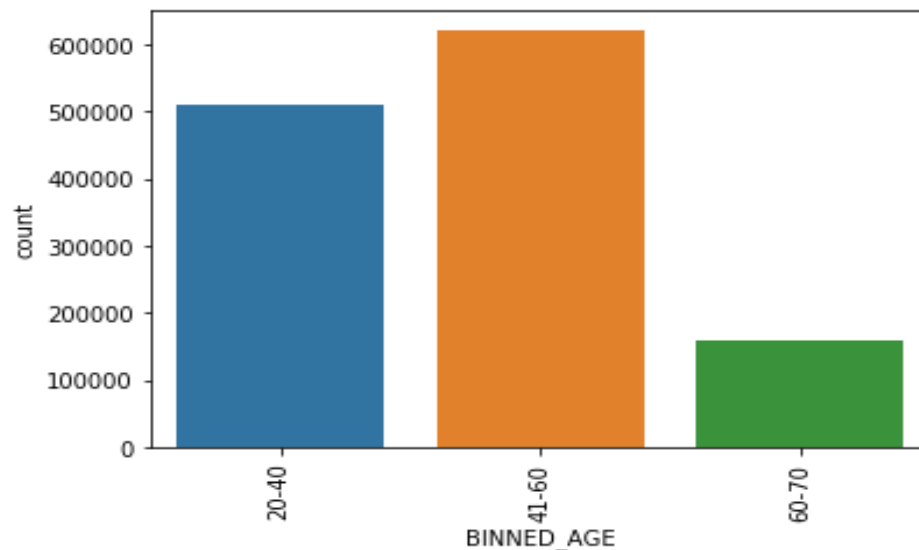
AGE GROUPS

41-60 age groups people were mostly non-defaulters. 20-40 age group people were majorly defaulters

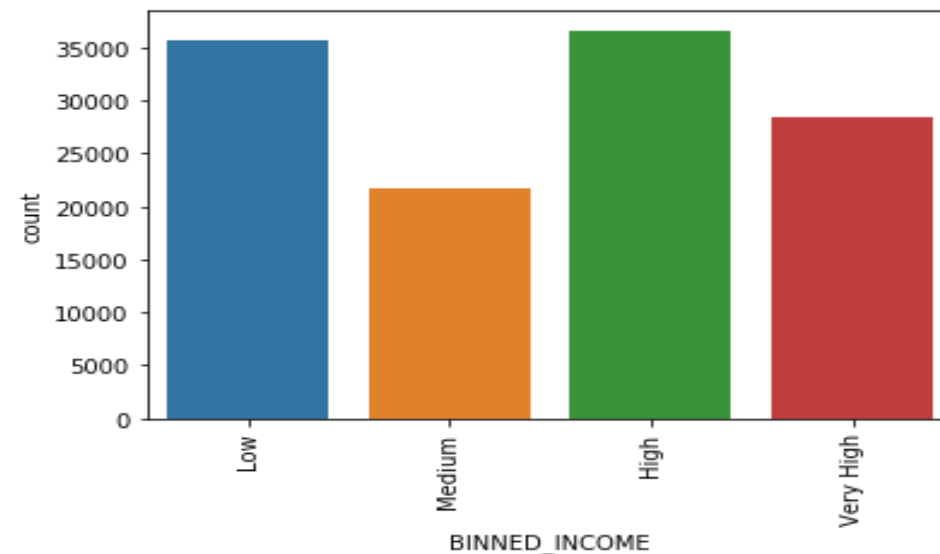
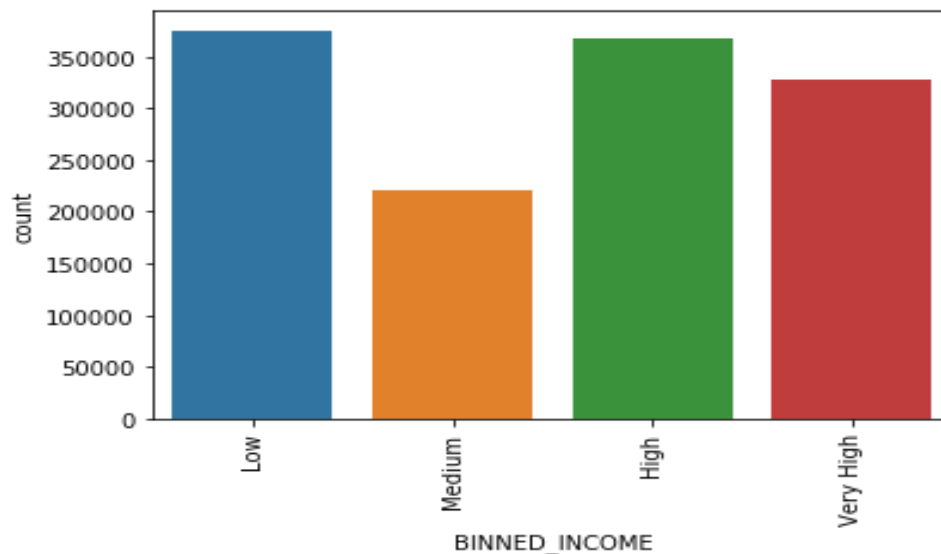
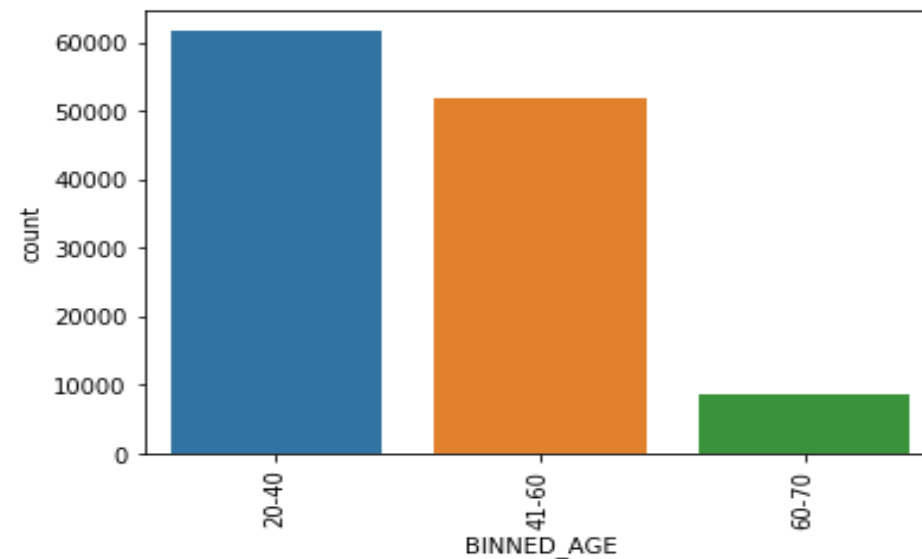
INCOME GROUPS

Risk follows as:
LOW>HIGH>VERY HIGH>MEDIUM

TARGET 0: NON DEFAULTERS



TARGET 1: DEFAULTERS



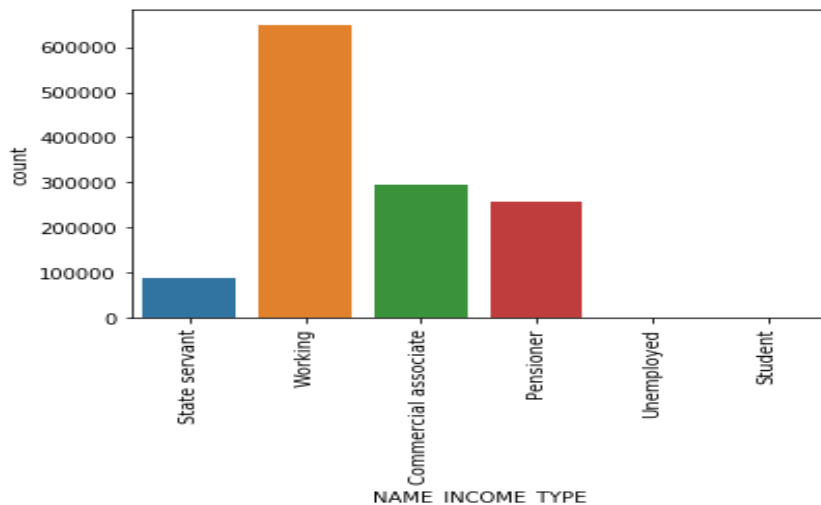
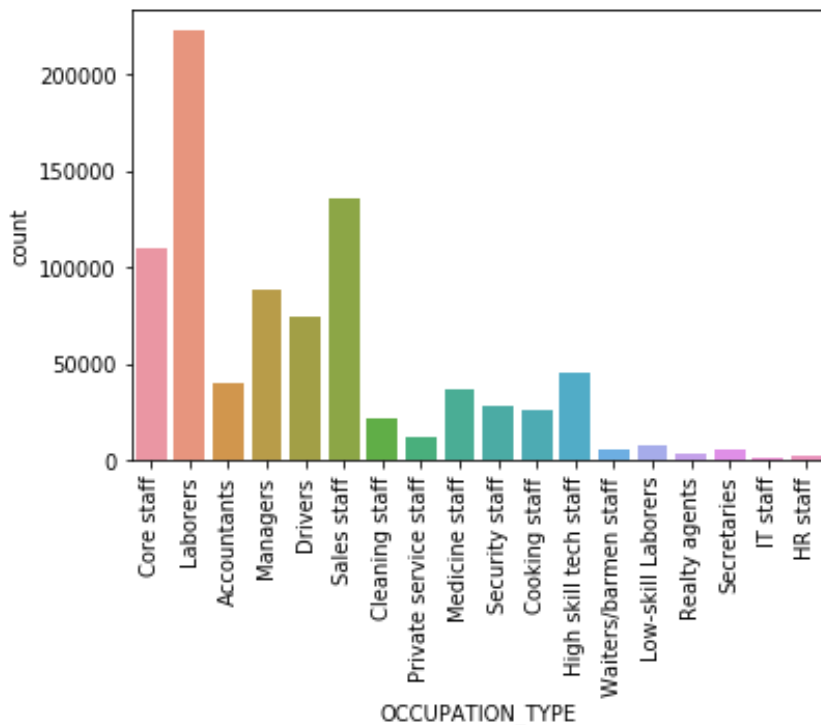
OCCUPATION TYPE

Labourers are
majorly
defaulters
followed by
Sales staff
and Drivers

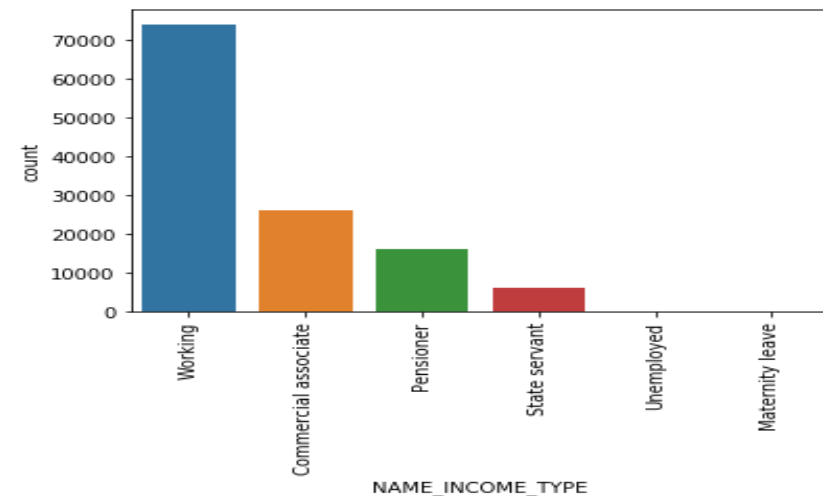
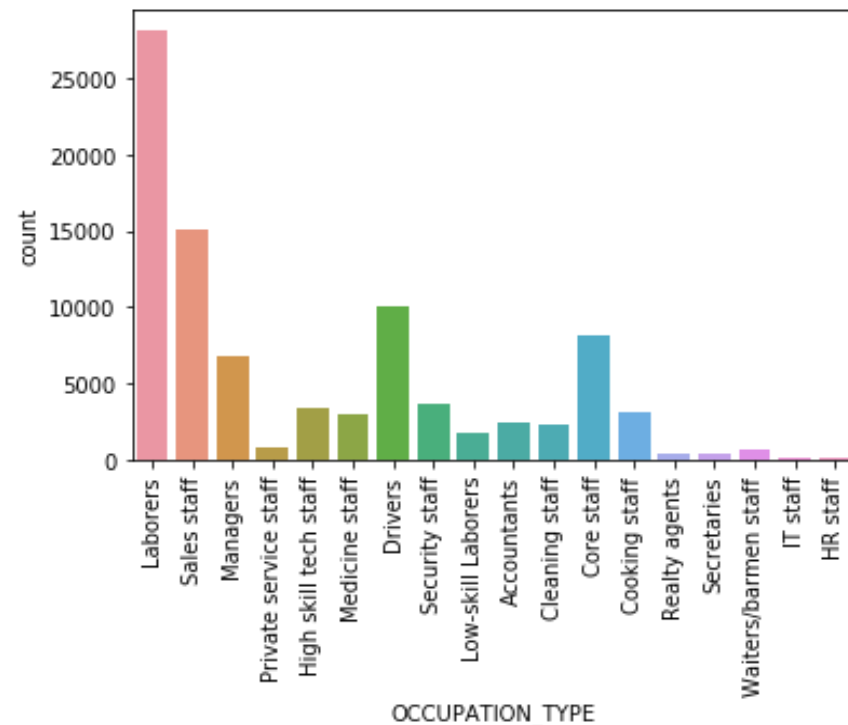
INCOME TYPE

Commercial
associates
default on
payment.
State
Servants have
very less
defaulters

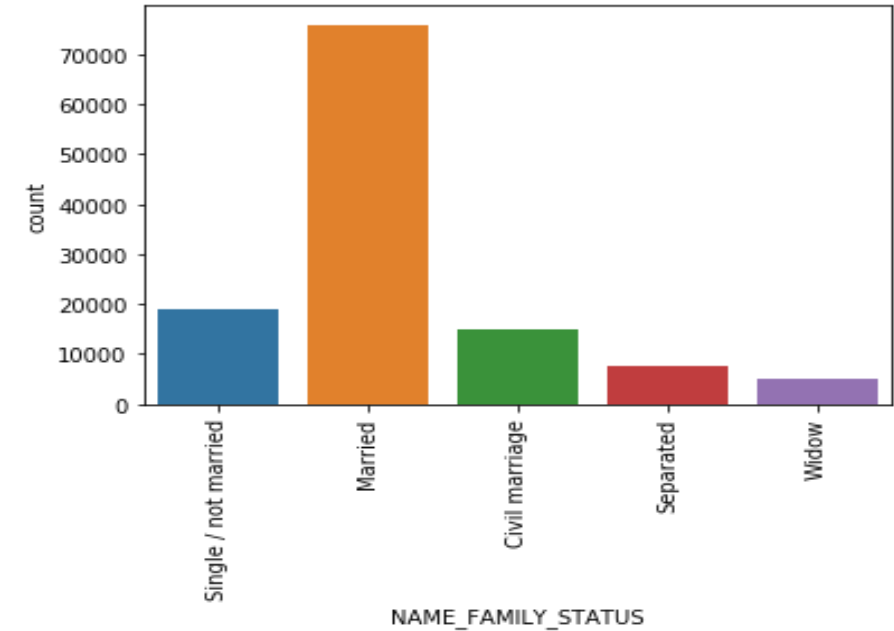
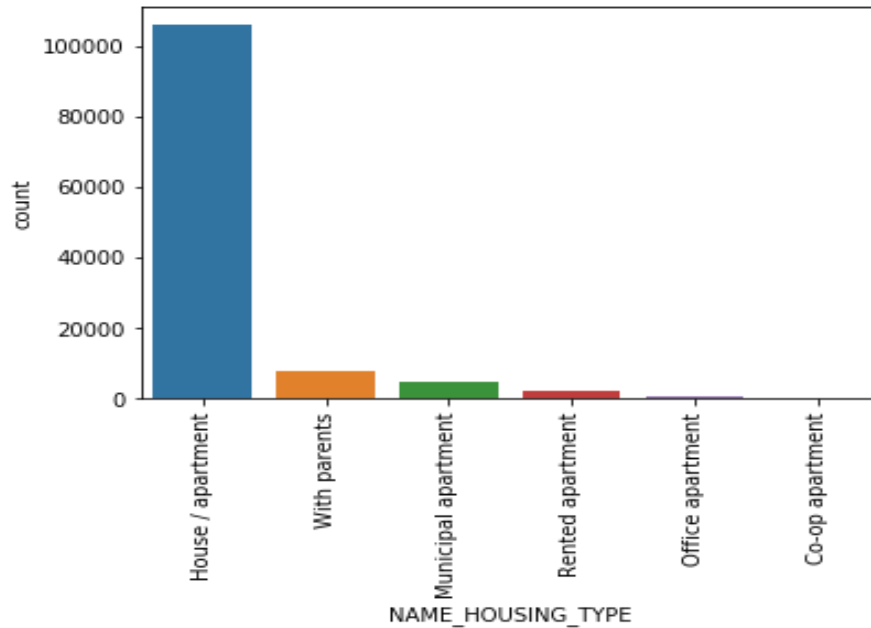
TARGET 0: NON DEFAULTERS



TARGET 1: DEFAULTERS



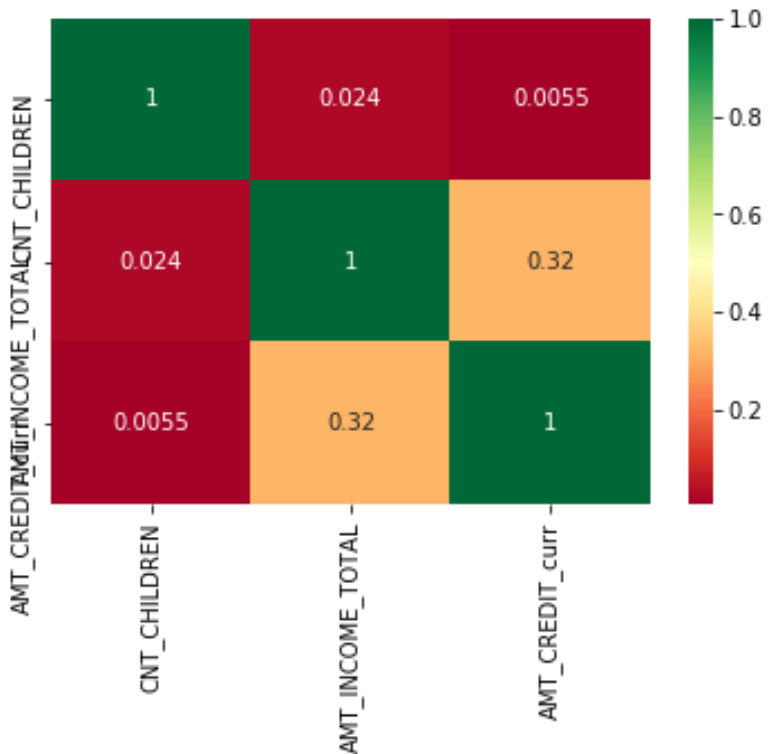
Save Business loss as well as financial loss



The pattern of HOUSING TYPE and FAMILY STATUS do not show major difference for Defaulters and Non- Defaulters. Hence, they are equally threat to business loss as well as financial loss.

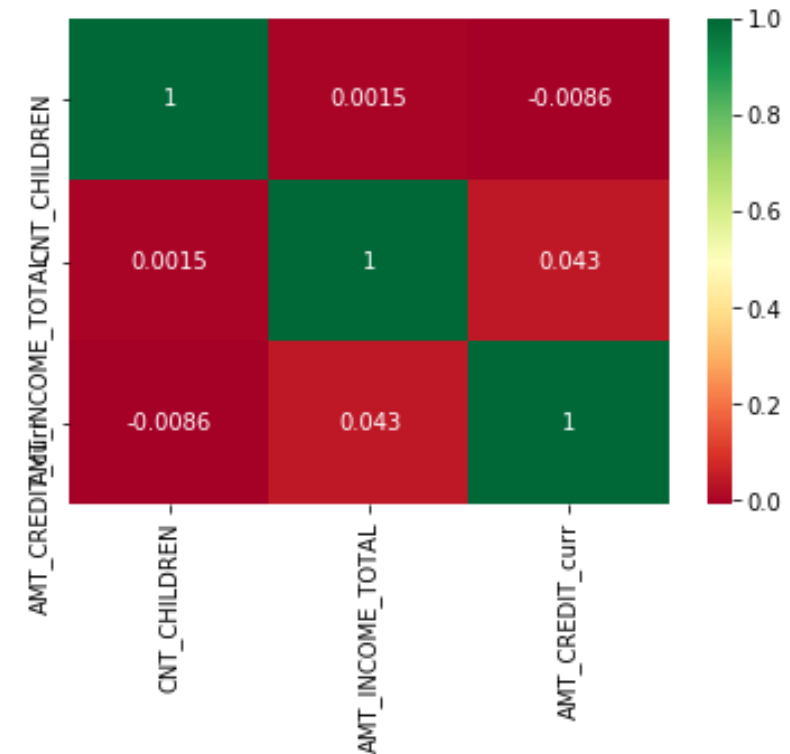
Correlation between Income, Credit and Children

TARGET 0: NON DEFAULTERS



INCOME and CREDIT AMOUNT are highly correlated in Non Defaulters. Which means people having high income were correctly given high credit limits and people with low incomes were given low credit limits. Hence, the result is Non Default in payments.

TARGET 1: DEFAULTERS



INCOME and CREDIT AMOUNT are not correlated in Defaulters. Which means that people with low incomes were given high credit limits. Which resulted in Default.

Top 10 Correlations for Defaulters and Non-Defaulters

TARGET 0: NON DEFAULTERS

	VAR1	VAR2	Correlation
100	AMT_GOODS_PRICE	AMT_CREDIT	0.912207
194	CNT_FAM_MEMBERS	CNT_CHILDREN	0.878571
84	AMT_ANNUITY	AMT_CREDIT	0.737964
101	AMT_GOODS_PRICE	AMT_ANNUITY	0.695551
152	YEARS_EMPLOYED	YEARS_BIRTH	0.625827
168	YEARS_REGISTRATION	YEARS_BIRTH	0.333009
67	AMT_CREDIT	AMT_INCOME_TOTAL	0.318169
83	AMT_ANNUITY	AMT_INCOME_TOTAL	0.300013
185	YEARS_ID_PUBLISH	YEARS_EMPLOYED	0.276254
184	YEARS_ID_PUBLISH	YEARS_BIRTH	0.270395

TARGET 1: DEFAULTERS

	VAR1	VAR2	Correlation
100	AMT_GOODS_PRICE	AMT_CREDIT	0.943037
194	CNT_FAM_MEMBERS	CNT_CHILDREN	0.885484
84	AMT_ANNUITY	AMT_CREDIT	0.730797
101	AMT_GOODS_PRICE	AMT_ANNUITY	0.702081
152	YEARS_EMPLOYED	YEARS_BIRTH	0.581771
168	YEARS_REGISTRATION	YEARS_BIRTH	0.288672
184	YEARS_ID_PUBLISH	YEARS_BIRTH	0.252795
185	YEARS_ID_PUBLISH	YEARS_EMPLOYED	0.229150
169	YEARS_REGISTRATION	YEARS_EMPLOYED	0.192537
248	EXT_SOURCE_3	YEARS_BIRTH	0.171801

The top 5 correlated variables are same for both defaulters and non defaulters. However, the CREDIT AMOUNT and INCOME is not related in DEFAULTERS. Hence, they have defaulted. The ANNUITY and INCOME shall also be correlated, so that client is able to pay back using his salary. In defaulters, this is not the case. Hence, the annuity might be high than their income. Hence, the default.

LOAN FEATURES

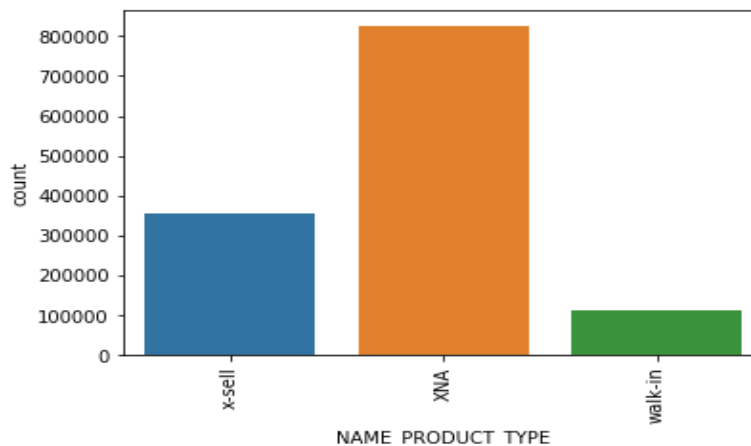
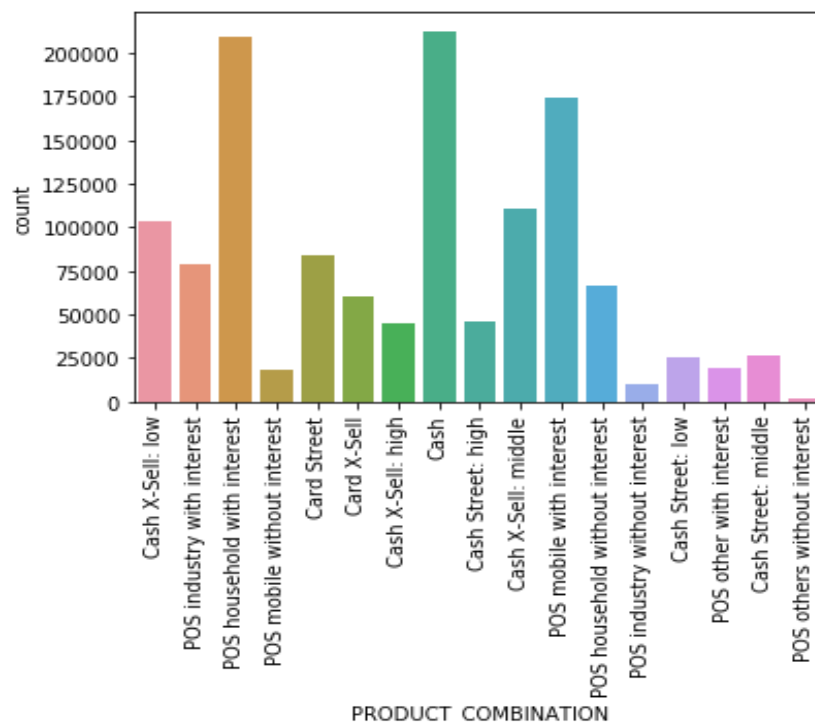
PRODUCT COMBINATION

Various product combinations can be seen resulting in default and not default.

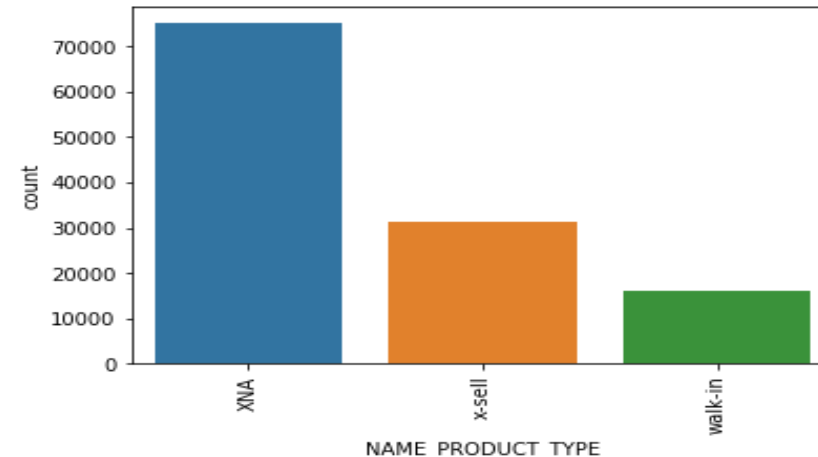
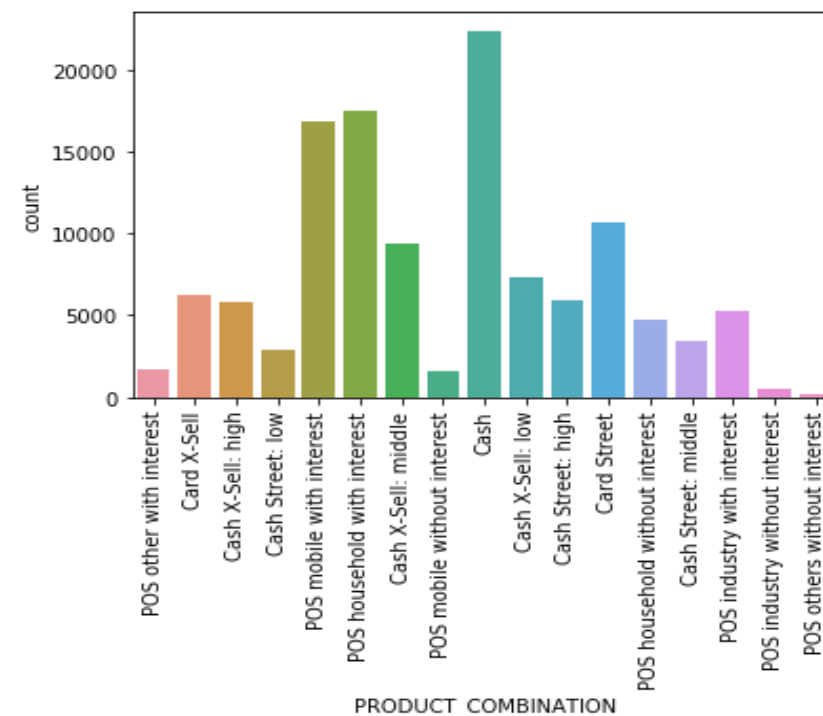
PRODUCT TYPE

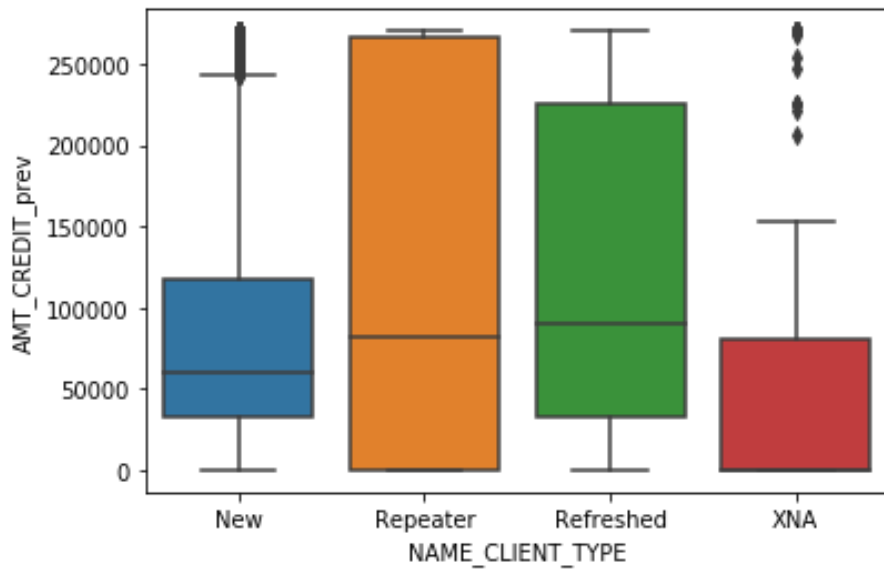
Of the people who defaulted, x-sell people were more. Which means, that direct selling scheme is wrong.

TARGET 0: NON DEFAULTERS

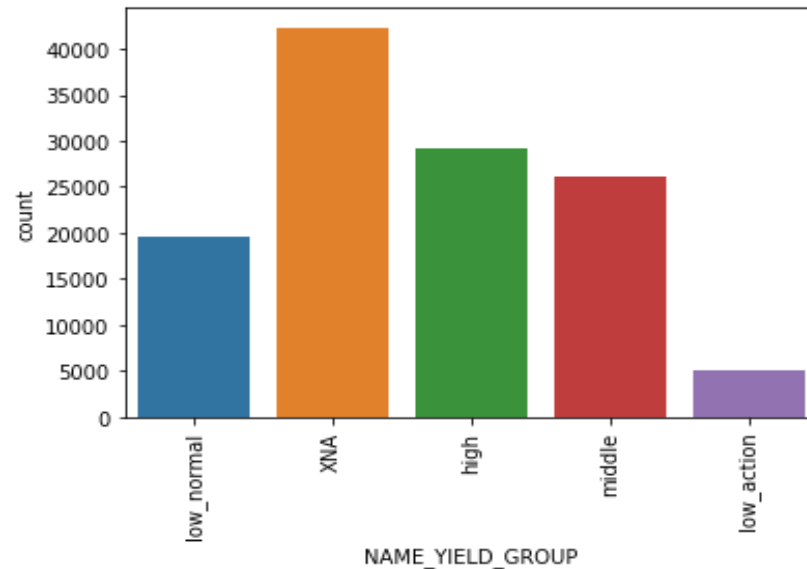


TARGET 1: DEFAULTERS

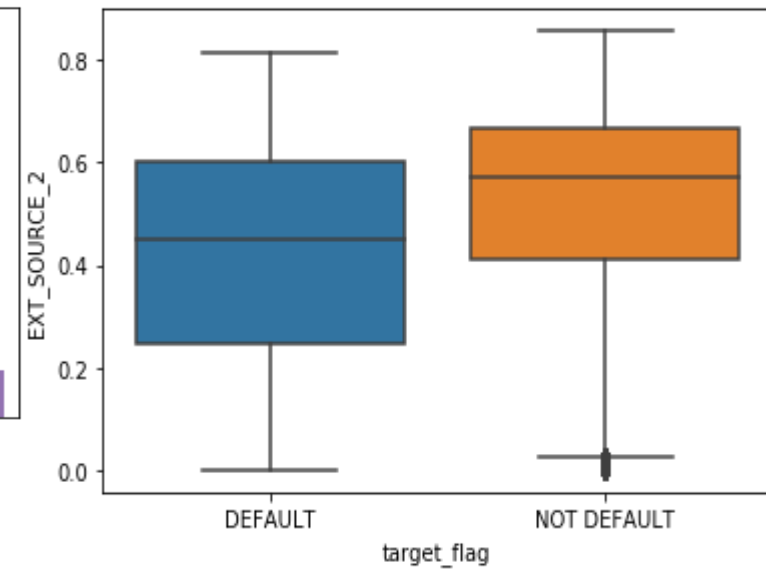




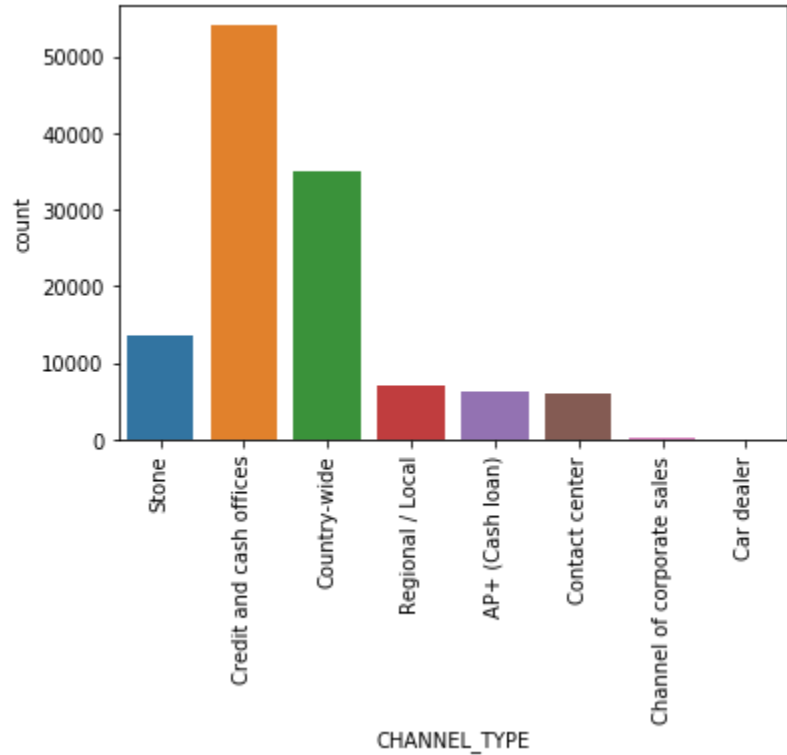
This is the data for Defaulters. The Repeated clients were the defaulters majorly.



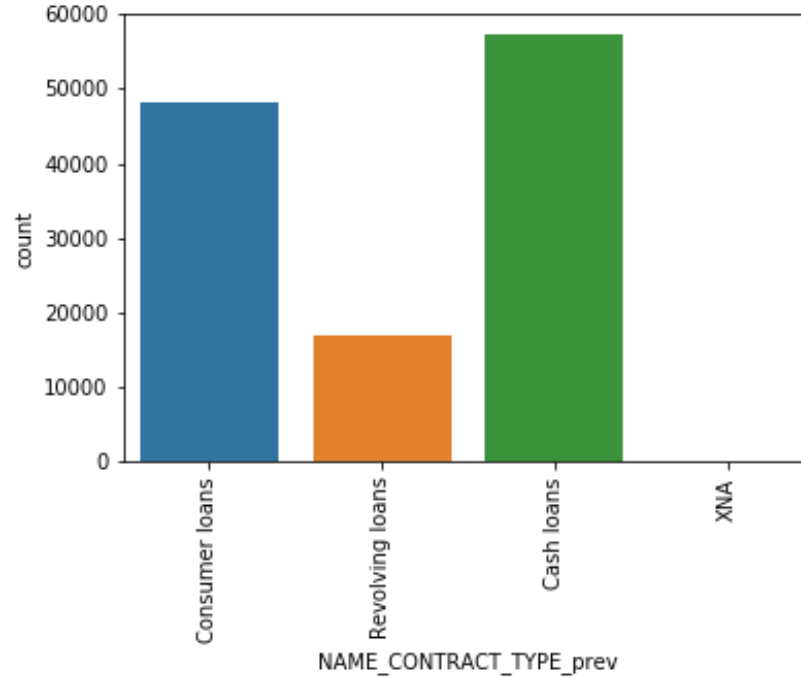
This is the data for Defaulters. The people who had High interest rates on their credit amount were the defaulters majorly.



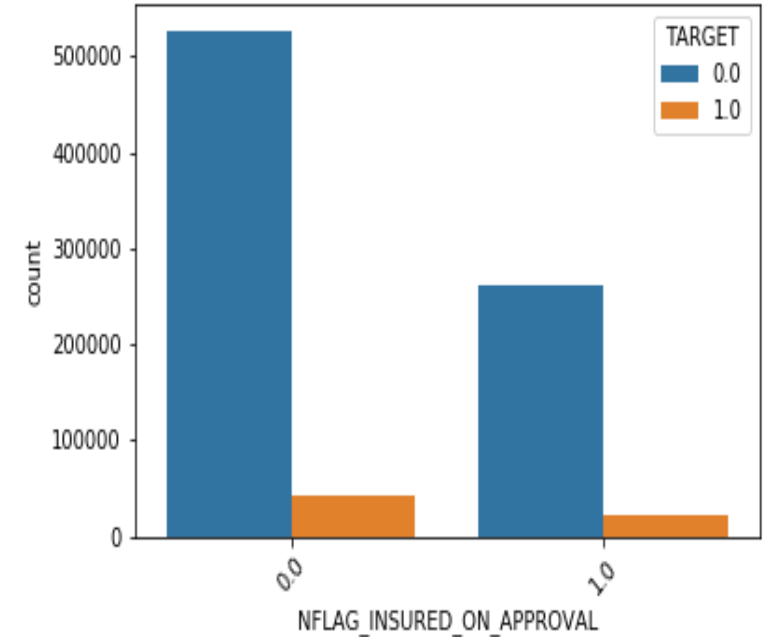
This is the data for Approved clients. Same clients when given loan based on EXT_SOURCE_2 did not default. Hence, EXT_SOURCE_2 is a reliable external data.



This is the data for Defaulters. The Credit and Cash offices generated the majority of defaulters.

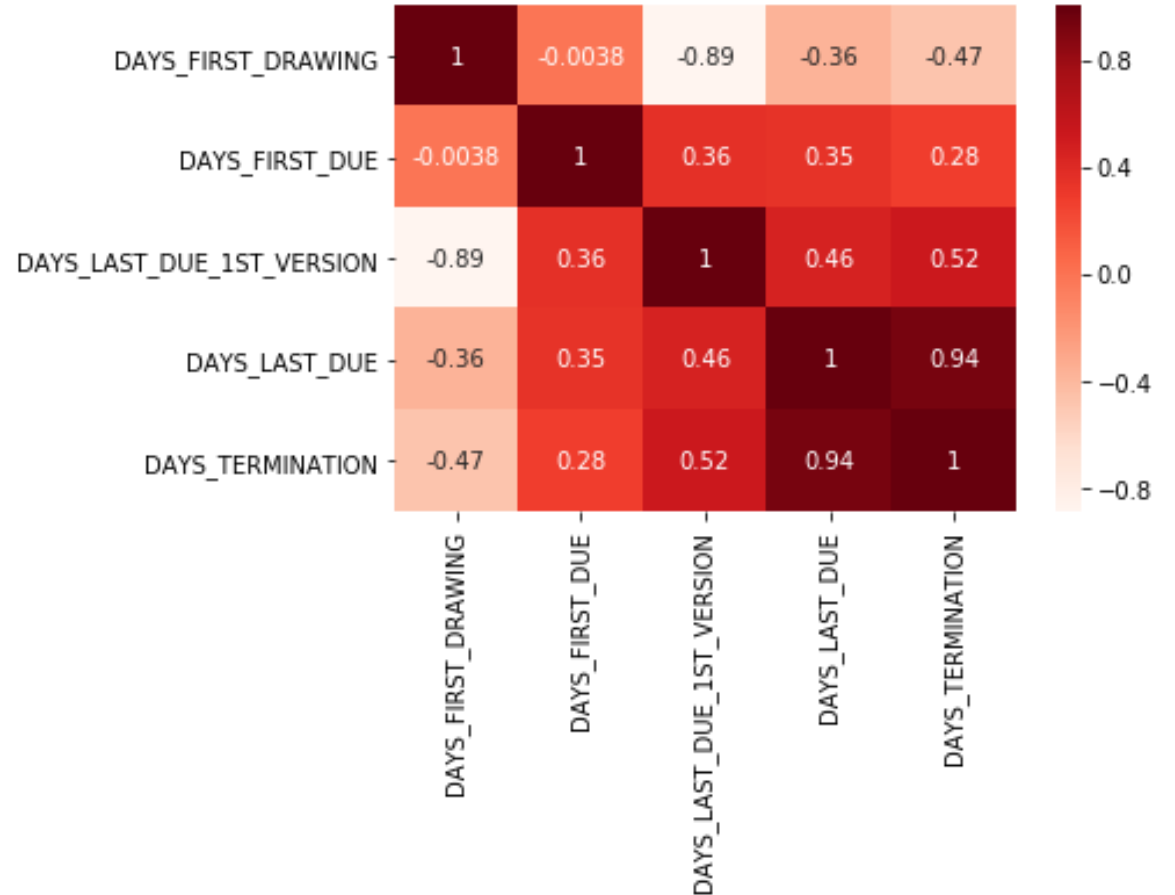
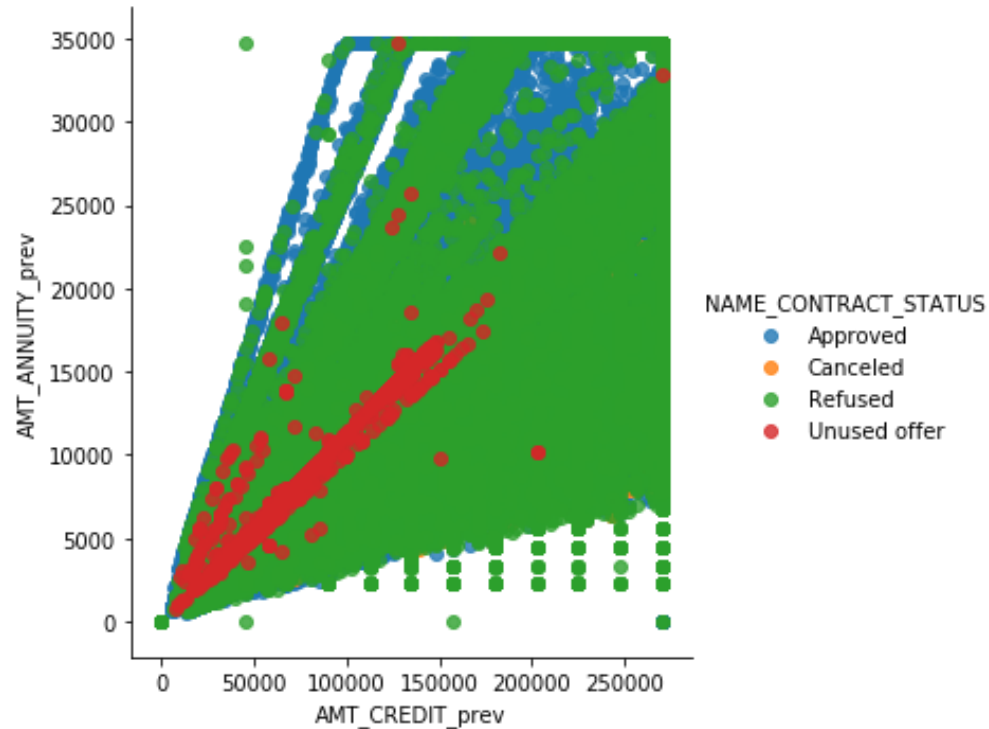


This is the data for Defaulters. The clients make default payments on Cash Loans.



This is the data for applications which were approved. Of them, the clients which were issued insurance resulted in less default payments.

Previous Application Scheme

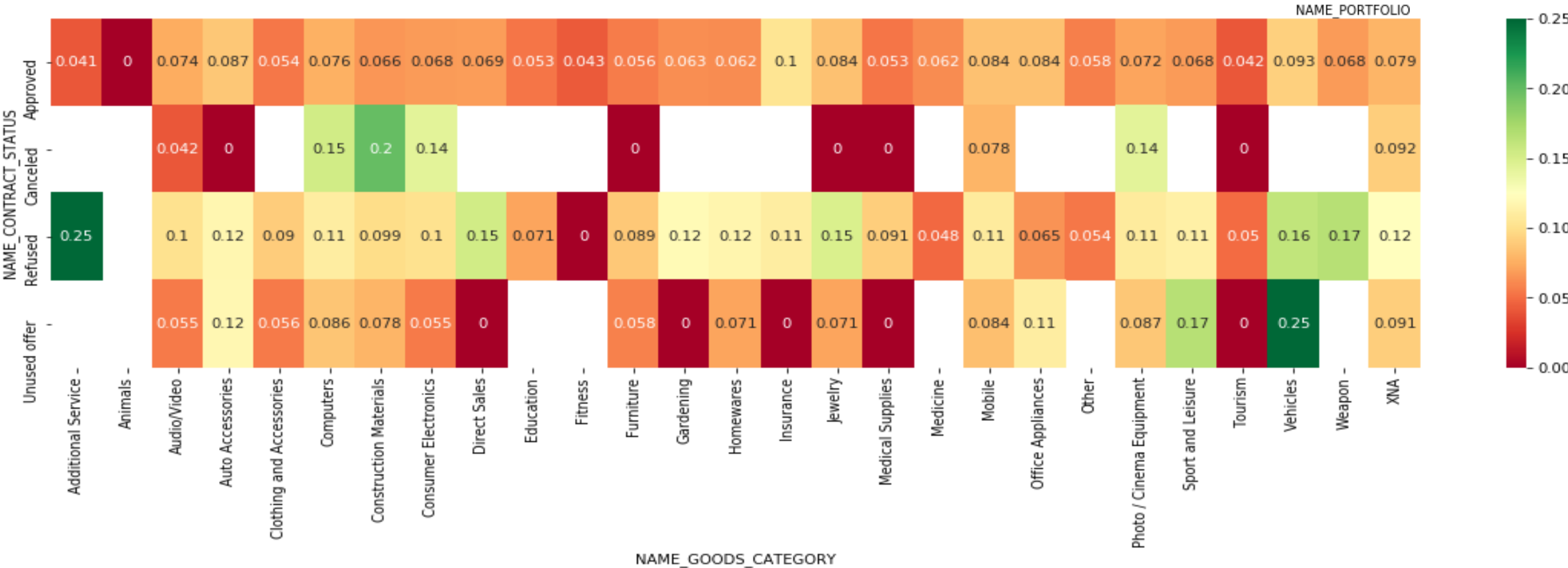
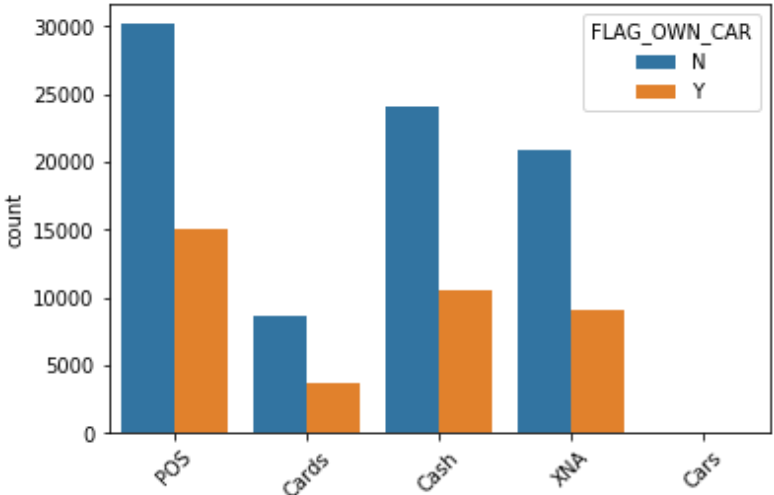


As we see, people with high credit amount and high annuity were approved. Also, people with high annuity but low credit were also approved. The result of income and credit correlation suggests that previously income, credit and annuity were not considered together. Second graph is for Defaulters based on Previous Application. The DAYS_FIRST_DRAWING is negatively correlated to DAYS_FIRST_DUE. It should be Positively correlated for people to be Non Defaulters. This is the lag in Previous Application scheme.

New Application Scheme: SUGGESTED

This is data for defaulters.

As for the repeated customers, we have data of the portfolio for which they took loan, and currently we have data that they own a car or not. Similarly, previous and current applications of repeated customers must be checked in order to reduce defaults. Cancelling application for Auto Accessories, Furniture, Medical Supplies clients is wrong. They are non-defaulters by current application data.



CONCLUSIONS

- Age groups 20-40 should be scanned thoroughly. Most defaulters belong to this age group.
- Labourers should be checked for giving loans.
- INCOME and CREDIT AMOUNT are correlated. Higher the income, higher shall the credit limit be.
- INCOME, CREDIT and ANNUITY go hand-in-hand. Lesser the income, lesser the credit amount, lesser the annuity to be paid and vice-versa.
- PRODUCT TYPE should be considered. Of the people who defaulted, x-sell were more. This could be because the direct selling scheme applied was wrong.
- Cash and credit offices generated the maximum defaulters.

RECOMMENDATIONS

- Age of the client: Middle aged and above 50 years old client would be able to repay. Target them by directly selling. Other age groups should be considered into the Walk-in selling.
- Labourers should be given the loan at reduced amount of credit. This saves the business loss as well as cater to a lot of people because of high population of Labourers. Same goes for Sales staff and drivers.
- Commercial associates should be given loan at High Interest rates as they default on payment. But, to reduce Business loss, they should not be rejected.
- Income, credit and annuity amounts should be related. If the income of a client is less, reduce the Loan amount for them.
- The Interest rates could also be reduced for these people.
- A direct selling scheme should be designed to reduce reaching out to people who are potential defaulters.
- While acquiring clients from Cash and Credit offices, stricter screening has to be done.

RECOMMENDATIONS: Strong Variables

- Age
- Occupation type
- Income type
- Income
- Credit amount
- Annuity amount
- Product type
- Goods category
- Portfolio of loan
- Days drawing and Days due
- Amount credit previously
- Channel type
- Insurance approval previously
- Client type
- External source of data
- Product combination