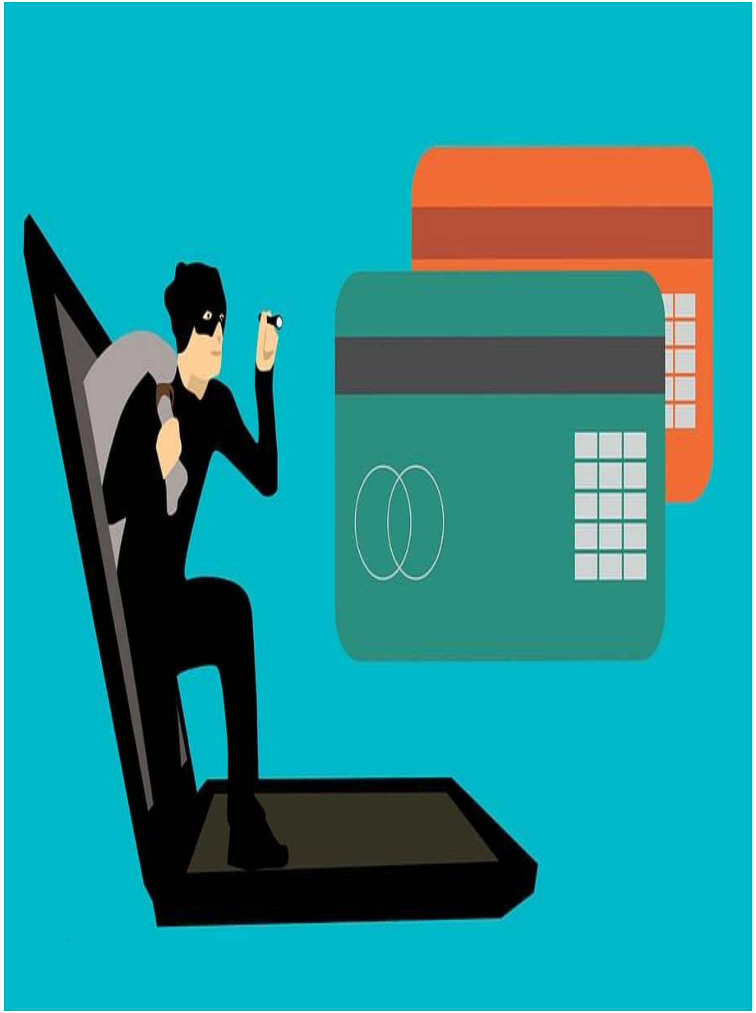
Credit EDA Case Study

Tanay Dhupkar & Vishakha Chavan

Credit EDA Case Study

Problem statement

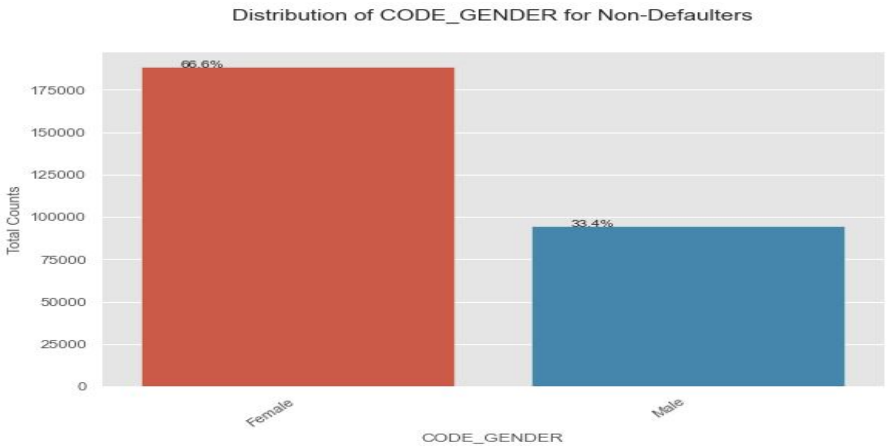
- This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.
- This will ensure that the consumers capable of repaying the loan are not rejected.
- Identification of such applicants using EDA is the aim of this case study.

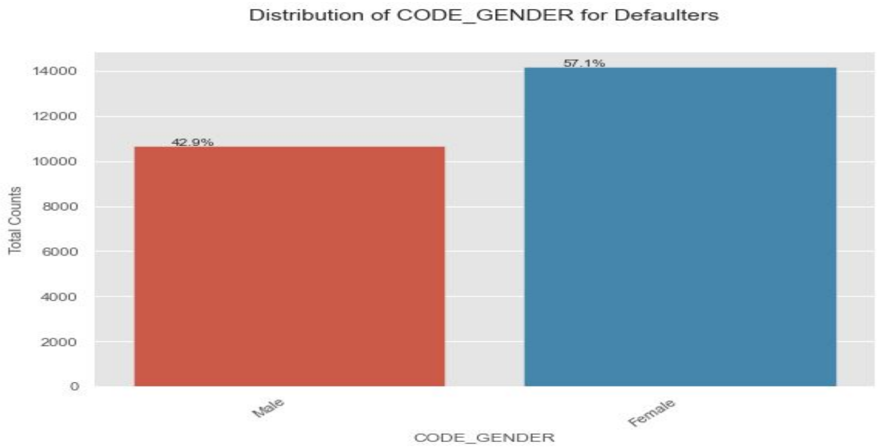


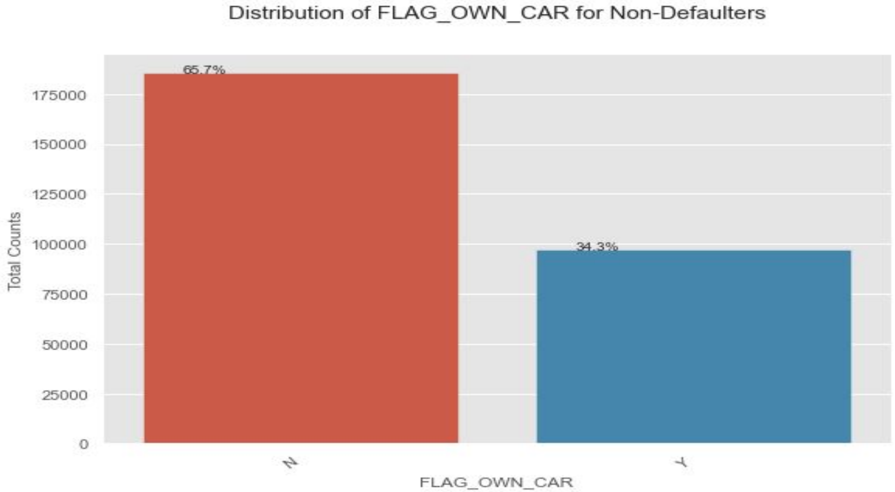
Exploratory Data Analysis

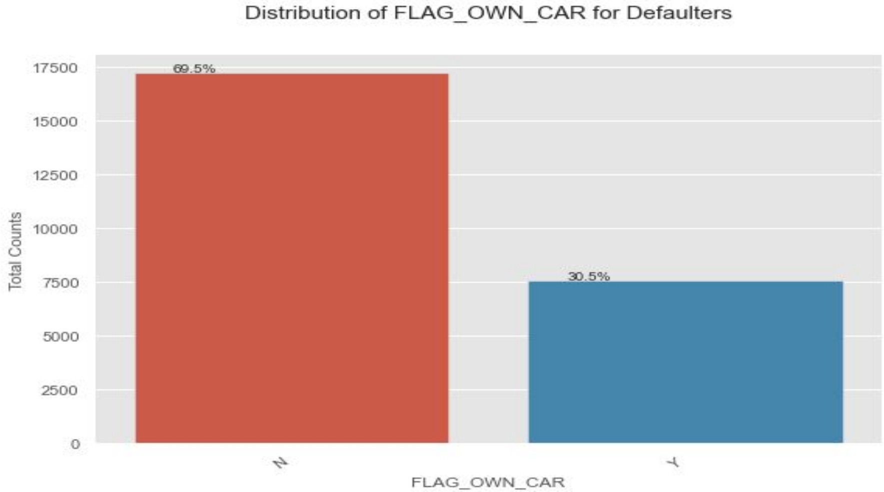
1) Here we are dealing with two huge datasets namely new_application and previous_application.
2) Also we be viewing the data keeping in mind the two types: Defaulters(target 1) and Non-Defaulters(target 0).

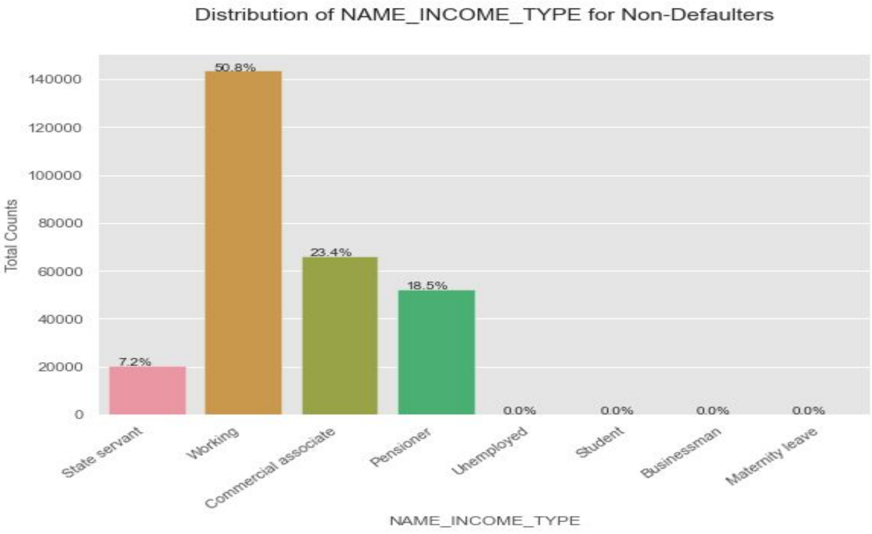
Customers falling under Category: 1(Defaulters/Having payment difficulties) is about 8 percent and Customers falling under Category : 0(Non-Defaulters) is about 92 percent

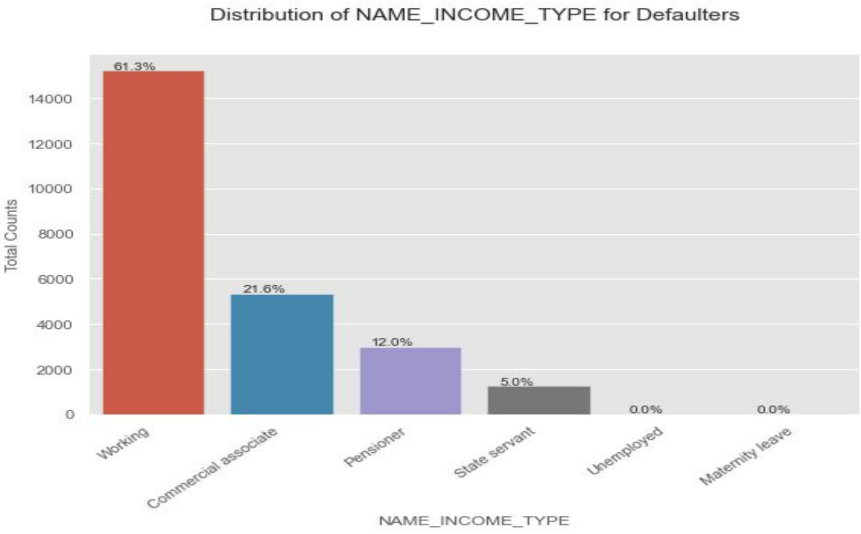

Distribution of CODE_GENDER for Non-Defaulters
66.6%
33.4%
Total Counts
0
25000
50000
75000
100000
125000
150000
175000
Female
Male
CODE_GENDER

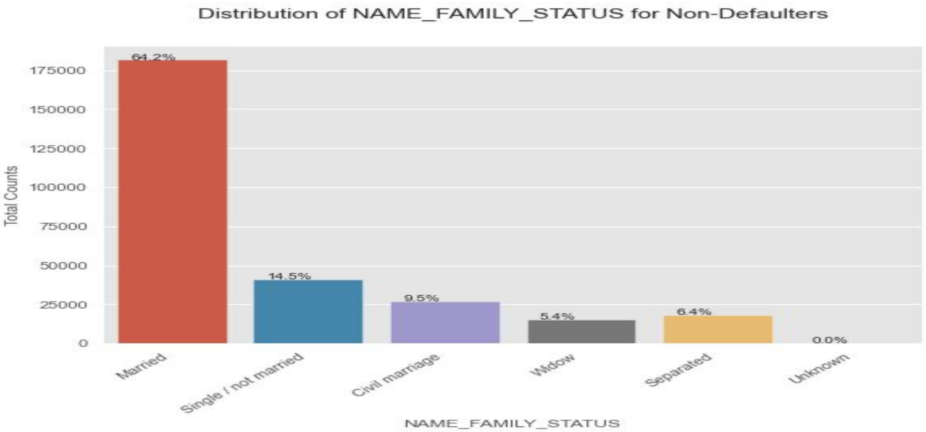

Distribution of CODE_GENDER for Defaulters
42.9%
57.1%
Total Counts
0
2000
4000
6000
8000
10000
12000
14000
Male
Female
CODE_GENDER

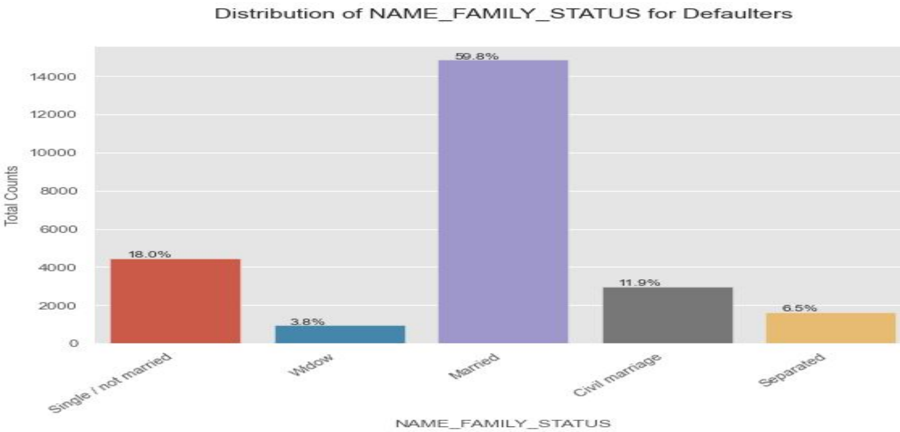

Distribution of FLAG_OWN_CAR for Non-Defaulters
65.7%
34.3%
Total Counts
0
25000
50000
75000
100000
125000
150000
175000
N
Y
FLAG_OWN_CAR

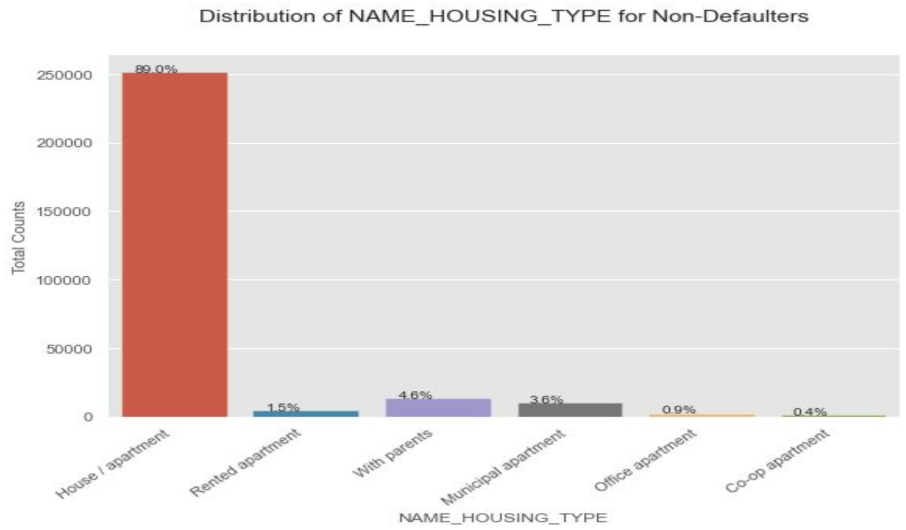

Distribution of FLAG_OWN_CAR for Defaulters
69.5%
30.5%
Total Counts
0
2500
5000
7500
10000
12500
15000
17500
N
Y
FLAG_OWN_CAR

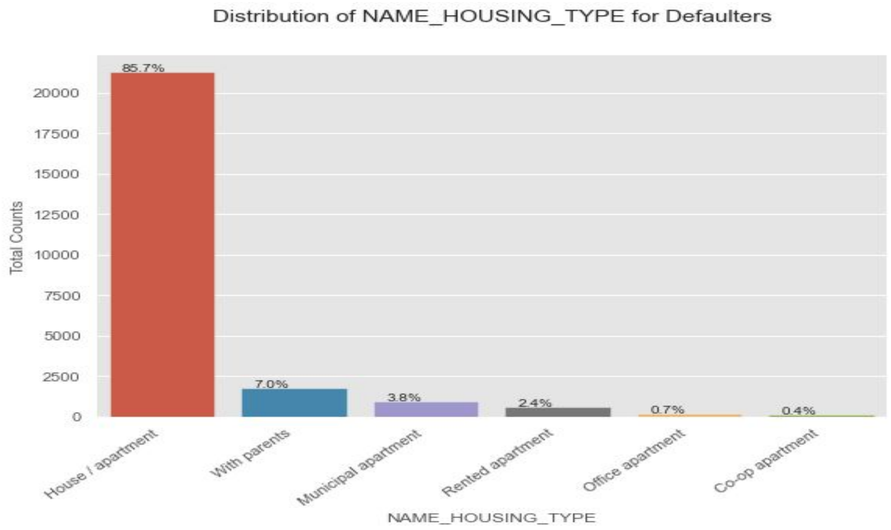

Distribution of NAME_INCOME_TYPE for Non-Defaulters
Total Counts
NAME_INCOME_TYPE
State servant
Working
Commercial associate
Pensioner
Unemployed
Student
Businessman
Maternity leave
7.2%
50.8%
23.4%
18.5%
0.0%
0.0%
0.0%
0.0%
0
20000
40000
60000
80000
100000
120000
140000

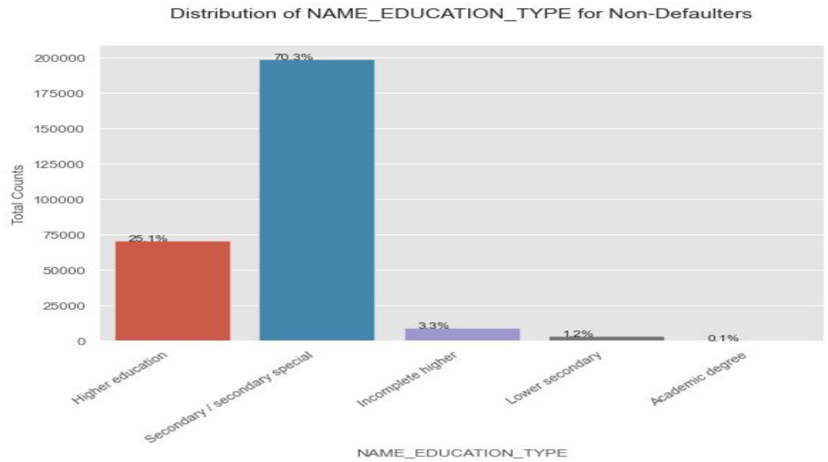

Distribution of NAME_INCOME_TYPE for Defaulters
Total Counts
NAME_INCOME_TYPE
Working
Commercial associate
Pensioner
State servant
Unemployed
Maternity leave
61.3%
21.6%
12.0%
5.0%
0.0%
0.0%
0
2000
4000
6000
8000
10000
12000
14000

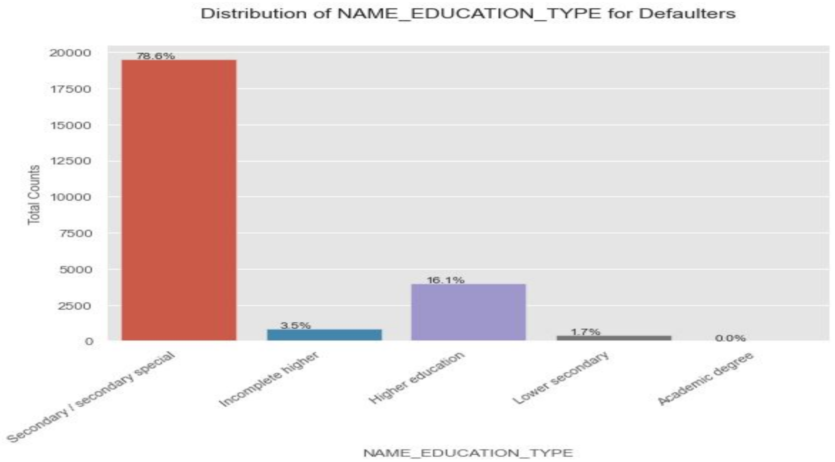

Distribution of NAME_FAMILY_STATUS for Non-Defaulters
Total Counts
NAME_FAMILY_STATUS
Married
Single / not married
Civil marriage
Widow
Separated
Unknown
64.2%
14.5%
9.5%
5.4%
6.4%
0.0%
0
25000
50000
75000
100000
125000
150000
175000

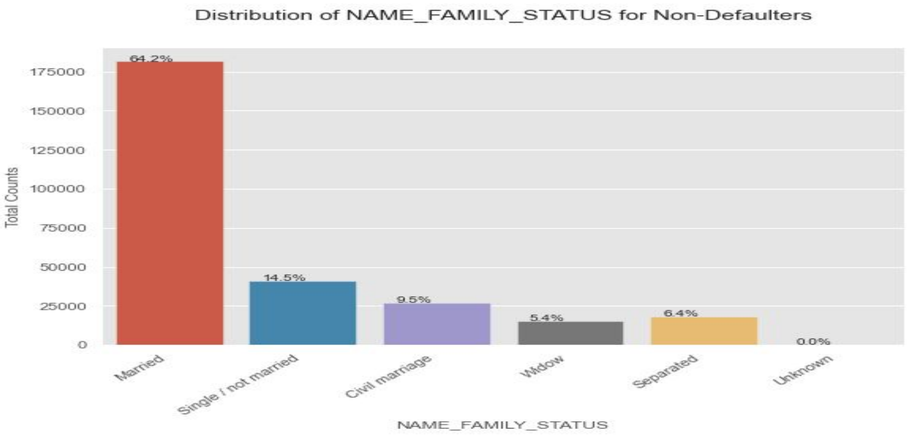

Distribution of NAME_FAMILY_STATUS for Defaulters
Total Counts
NAME_FAMILY_STATUS
Single / not married
Widow
Married
Civil marriage
Separated
18.0%
3.8%
59.8%
11.9%
6.5%
0
2000
4000
6000
8000
10000
12000
14000

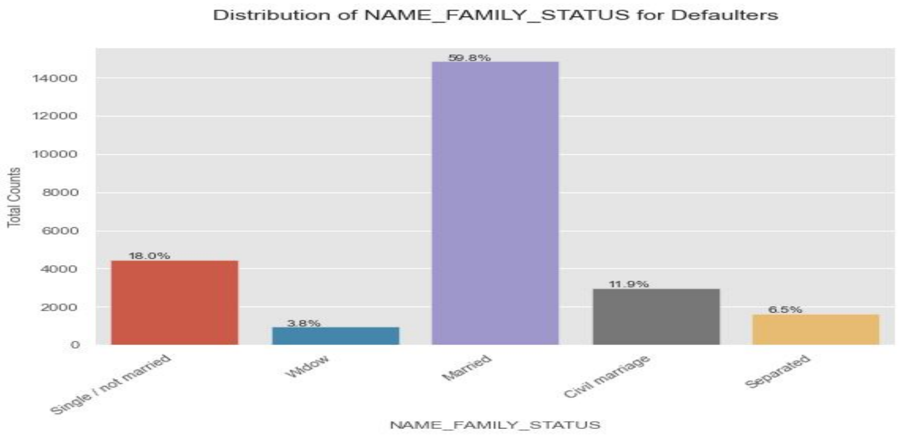

Distribution of NAME_HOUSING_TYPE for Non-Defaulters
Total Counts
0
50000
100000
150000
200000
250000
89.0%
1.5%
4.6%
3.6%
0.9%
0.4%
House / apartment
Rented apartment
With parents
Municipal apartment
Office apartment
Co-op apartment
NAME_HOUSING_TYPE


Distribution of NAME_HOUSING_TYPE for Defaulters
Total Counts
0
2500
5000
7500
10000
12500
15000
17500
20000
85.7%
7.0%
3.8%
2.4%
0.7%
0.4%
House / apartment
With parents
Municipal apartment
Rented apartment
Office apartment
Co-op apartment
NAME_HOUSING_TYPE


Distribution of NAME_EDUCATION_TYPE for Non-Defaulters
Total Counts
0
25000
50000
75000
100000
125000
150000
175000
200000
25.1%
70.3%
3.3%
1.2%
0.1%
Higher education
Secondary / secondary special
Incomplete higher
Lower secondary
Academic degree
NAME_EDUCATION_TYPE


Distribution of NAME_EDUCATION_TYPE for Defaulters
Total Counts
0
2500
5000
7500
10000
12500
15000
17500
20000
78.6%
3.5%
16.1%
1.7%
0.0%
Secondary / secondary special
Incomplete higher
Higher education
Lower secondary
Academic degree
NAME_EDUCATION_TYPE





Conclusion: 1) Females apply for loans for often than males.

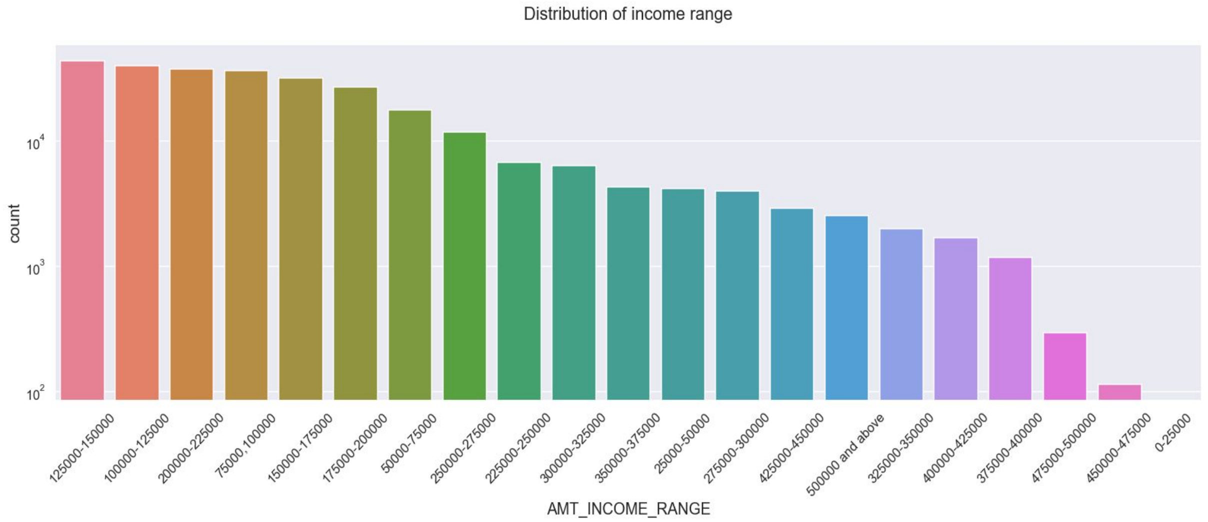
2) People with cars default more often.

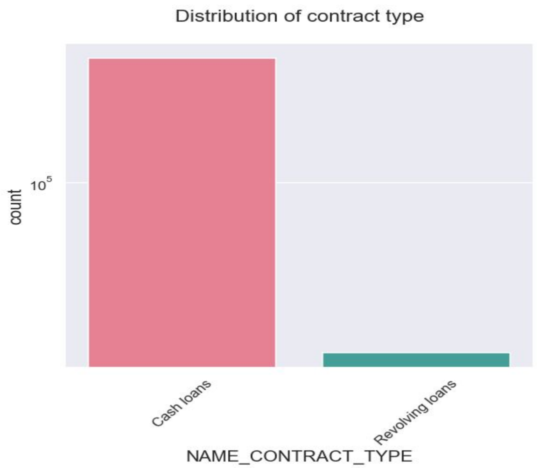
3) Students default very less while working class people default comparatively more as compared to Others.

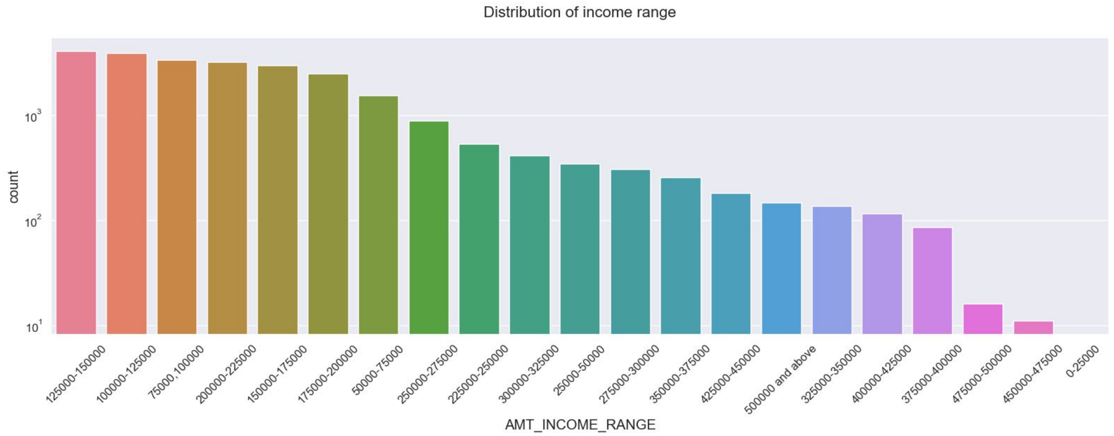
4) Married people and people with property apply for loans more often.

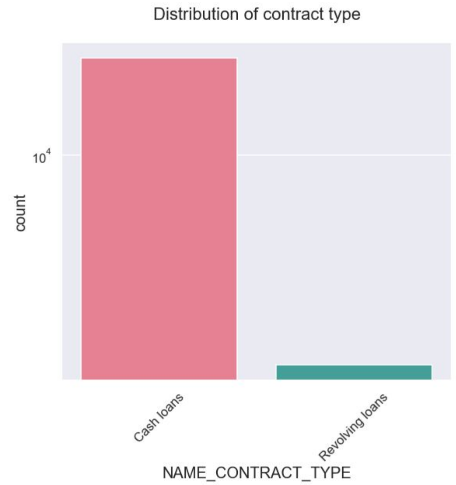
5) Almost all Education categories are equally likely to default except for higher educated ones.

6) Majority of middle class people apply for loans and contribute significantly to defaulters and non-defaulters.


Distribution of income range
count
10^4
10^3
10^2
125000-150000
100000-125000
200000-225000
75000,100000
150000-175000
175000-200000
50000-75000
250000-275000
225000-250000
300000-325000
350000-375000
25000-50000
275000-300000
425000-450000
500000 and above
325000-350000
400000-425000
375000-400000
475000-500000
450000-475000
0-25000
AMT_INCOME_RANGE

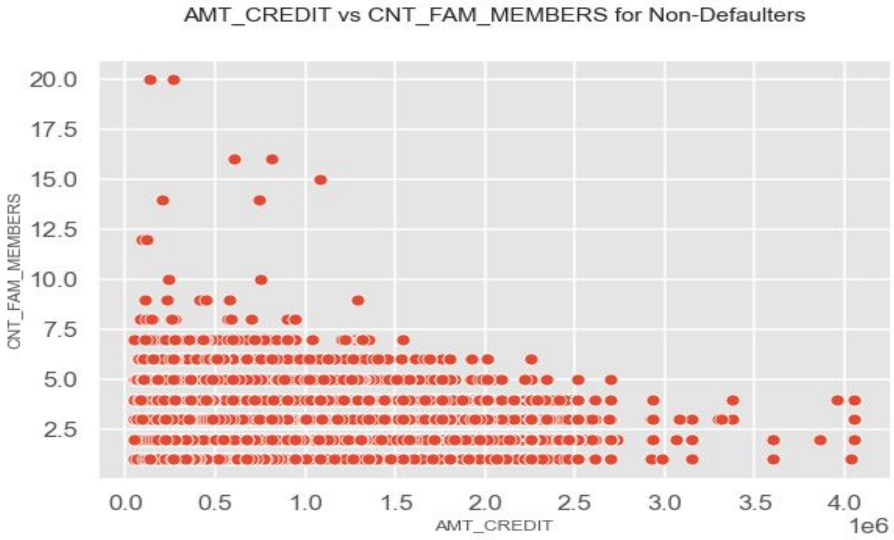
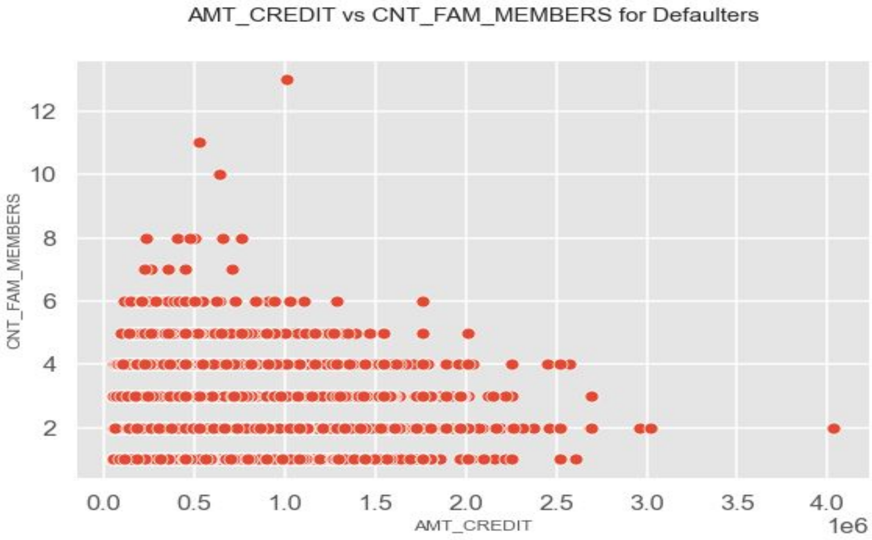
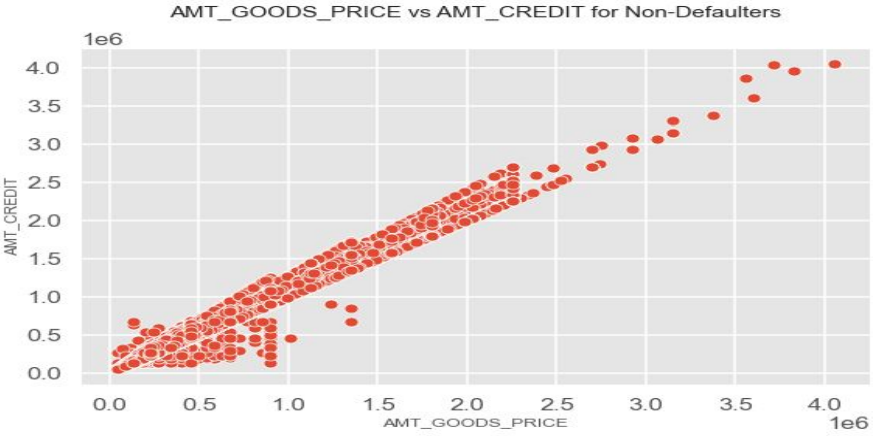
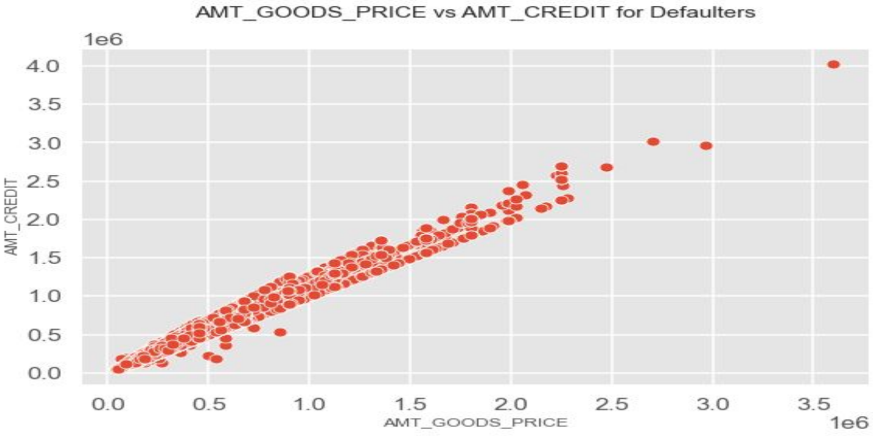

Distribution of contract type
count
10^5
Cash loans
Revolving loans
NAME_CONTRACT_TYPE

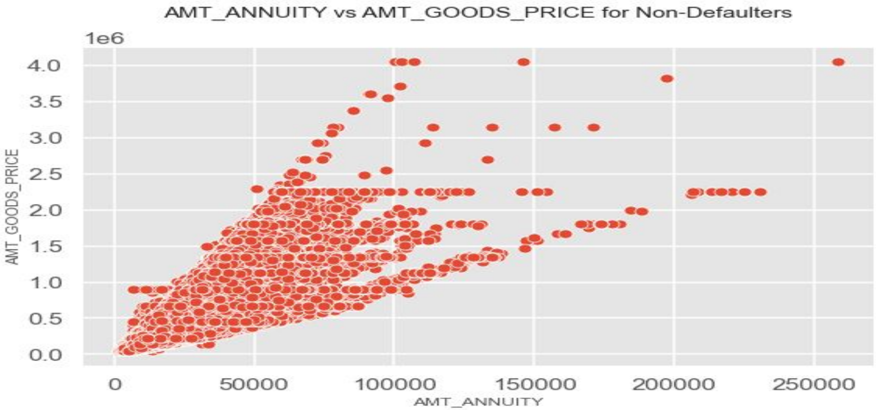

Distribution of income range
count
10^3
10^2
10^1
125000-150000
100000-125000
75000,100000
200000-225000
150000-175000
175000-200000
50000-75000
250000-275000
225000-250000
300000-325000
25000-50000
275000-300000
350000-375000
425000-450000
500000 and above
325000-350000
400000-425000
375000-400000
475000-500000
450000-475000
0-25000
AMT_INCOME_RANGE

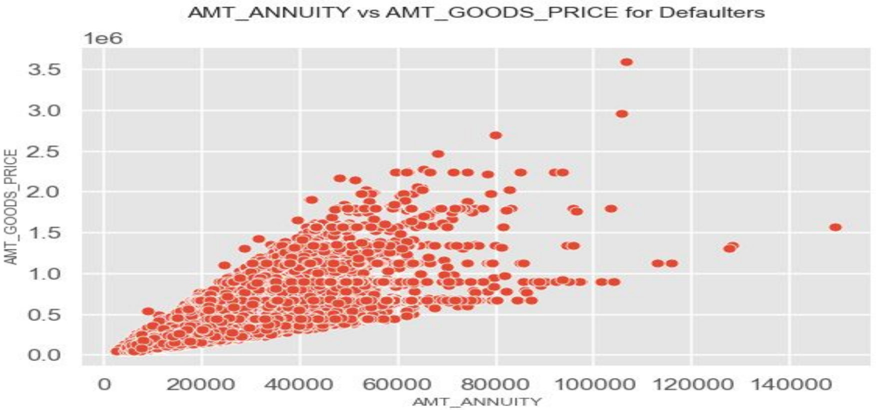

Distribution of contract type
count
10^4
Cash loans
Revolving loans
NAME_CONTRACT_TYPE

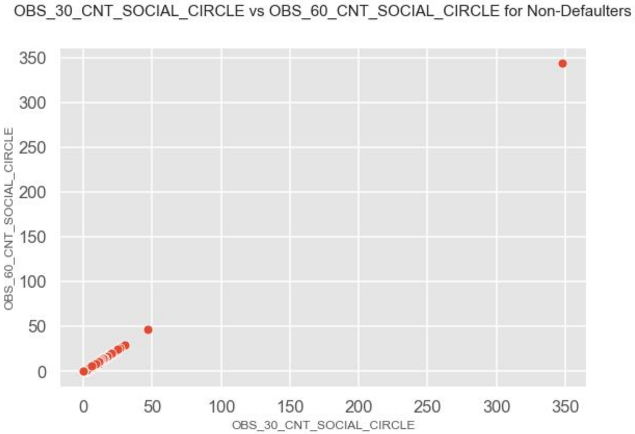
Conclusions:

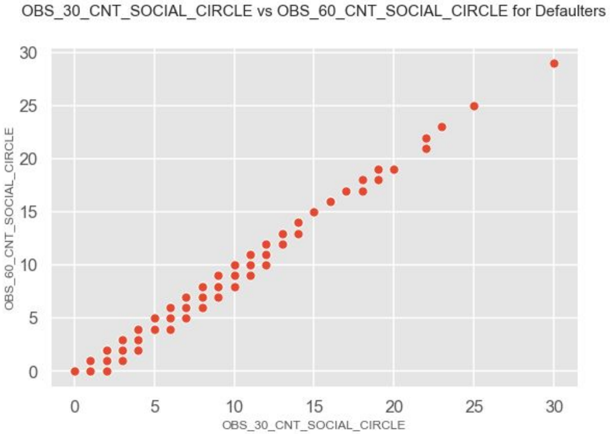
1) Income ranges between 100k-200k have more credits.
2) ‘Working’, ‘commercial associate’, ‘state servant’ have higher credits over others.
3) Less credits for ‘student’, ‘pensioner’, ‘businessmen’, ‘maternity leave’ categories.
4) For contract type ‘cash loans’ have higher credits over ‘revolving loans’.


AMT_CREDIT vs CNT_FAM_MEMBERS for Non-Defaulters
CNT_FAM_MEMBERS
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
1e6
AMT_CREDIT

AMT_CREDIT vs CNT_FAM_MEMBERS for Defaulters
CNT_FAM_MEMBERS
12
10
8
6
4
2
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
1e6
AMT_CREDIT

AMT_GOODS_PRICE vs AMT_CREDIT for Non-Defaulters
1e6
AMT_CREDIT
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
1e6
AMT_GOODS_PRICE

AMT_GOODS_PRICE vs AMT_CREDIT for Defaulters
1e6
AMT_CREDIT
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1e6
AMT_GOODS_PRICE

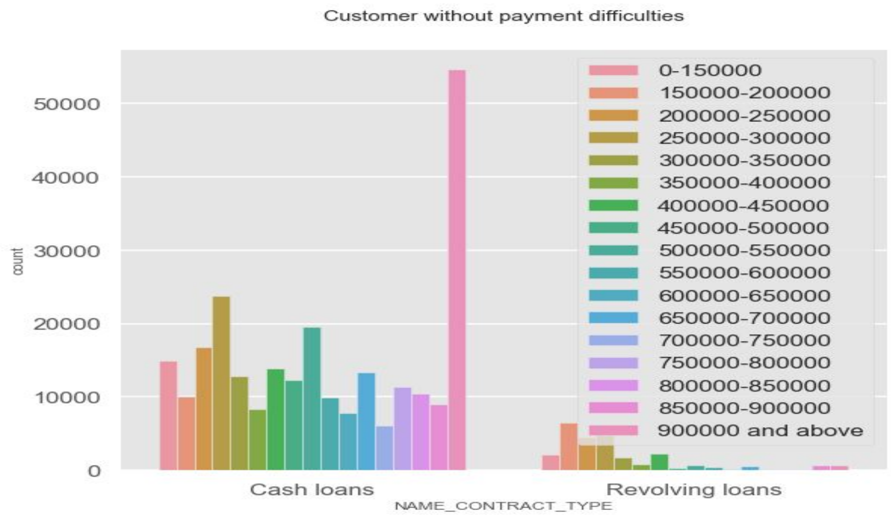


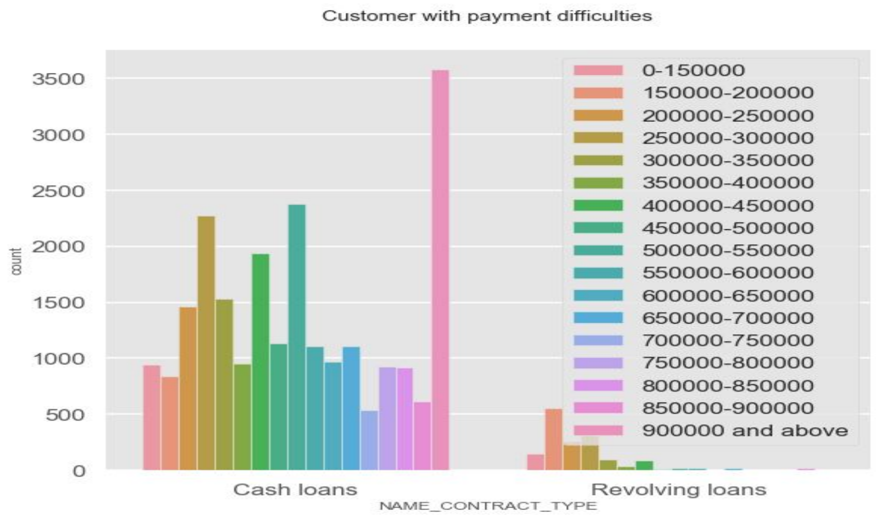


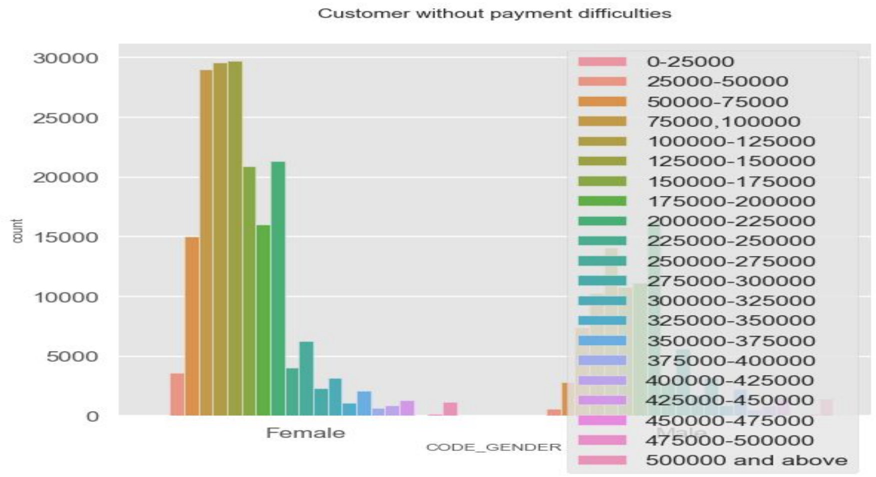


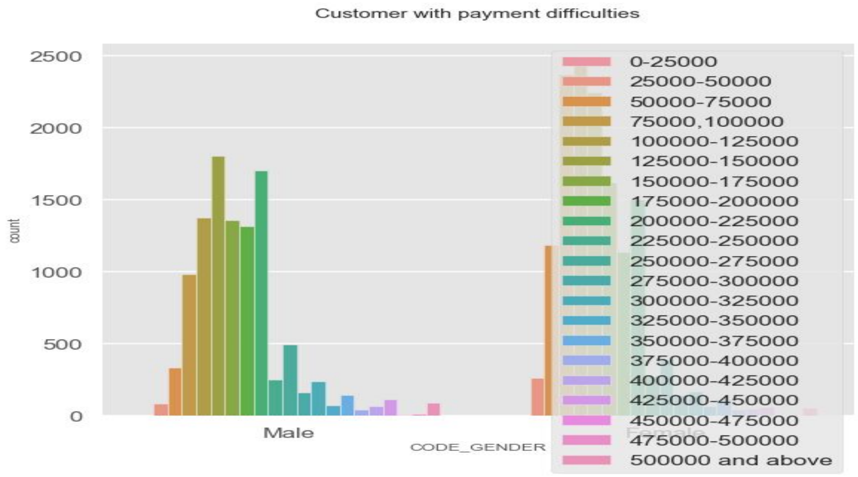


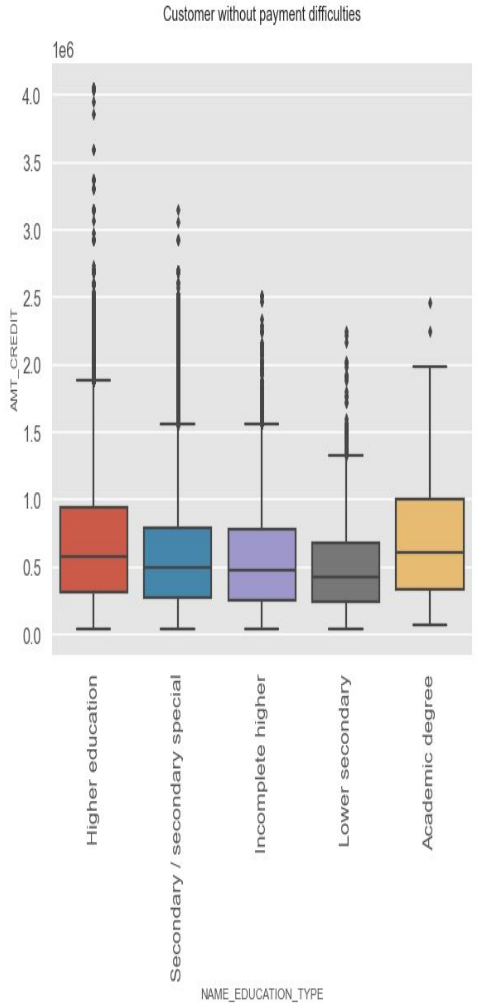
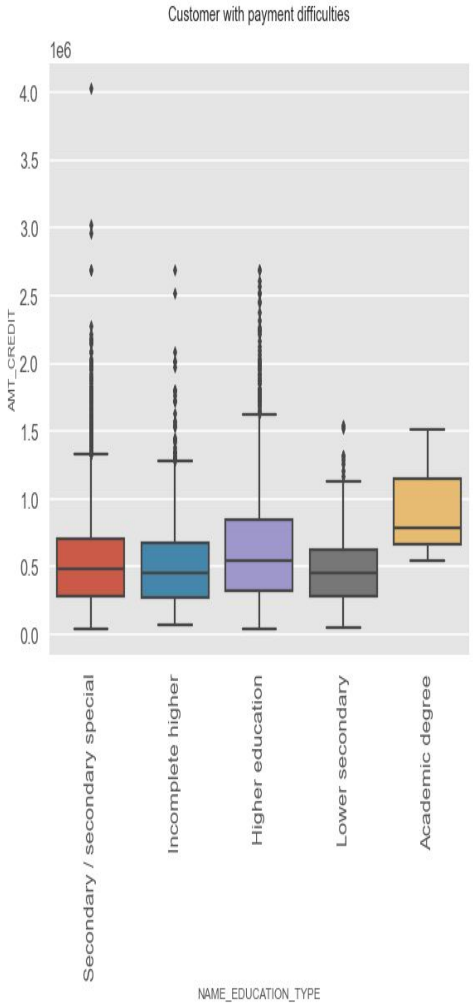
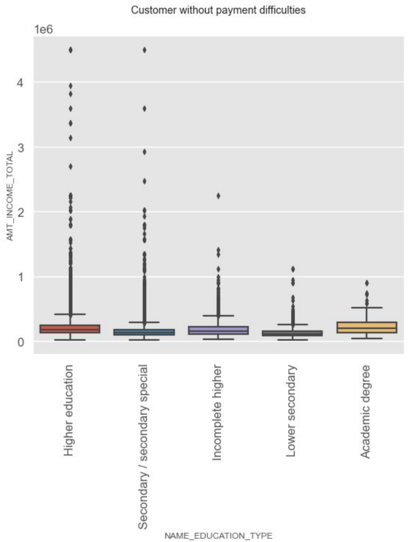
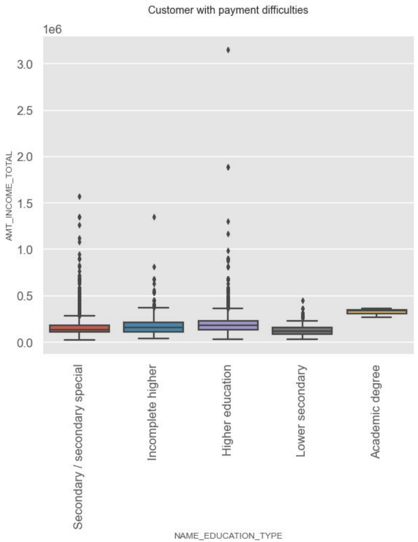
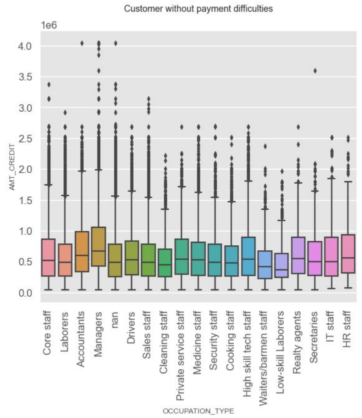
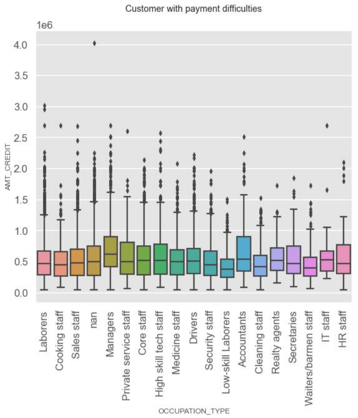
Thus, people with larger families and AMT_CREDIT default less. Also higher amount installments don't have much defaulters.


Customer without payment difficulties
count
0
10000
20000
30000
40000
50000
Cash loans
Revolving loans
NAME_CONTRACT_TYPE
0-150000
150000-200000
200000-250000
250000-300000
300000-350000
350000-400000
400000-450000
450000-500000
500000-550000
550000-600000
600000-650000
650000-700000
700000-750000
750000-800000
800000-850000
850000-900000
900000 and above


Customer with payment difficulties
count
0
500
1000
1500
2000
2500
3000
3500
Cash loans
Revolving loans
NAME_CONTRACT_TYPE
0-150000
150000-200000
200000-250000
250000-300000
300000-350000
350000-400000
400000-450000
450000-500000
500000-550000
550000-600000
600000-650000
650000-700000
700000-750000
750000-800000
800000-850000
850000-900000
900000 and above

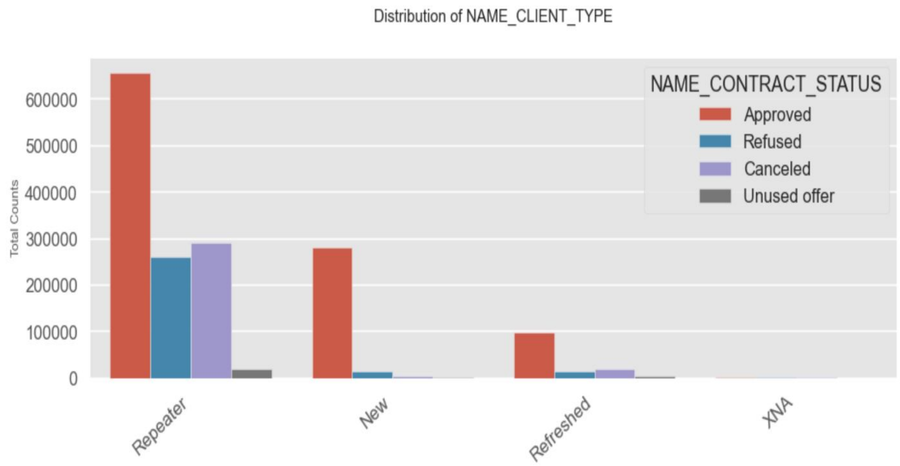

Customer without payment difficulties
count
0
5000
10000
15000
20000
25000
30000
Female
CODE_GENDER
0-25000
25000-50000
50000-75000
75000,100000
100000-125000
125000-150000
150000-175000
175000-200000
200000-225000
225000-250000
250000-275000
275000-300000
300000-325000
325000-350000
350000-375000
375000-400000
400000-425000
425000-450000
450000-475000
475000-500000
500000 and above


Customer with payment difficulties
count
0
500
1000
1500
2000
2500
Male
CODE_GENDER
0-25000
25000-50000
50000-75000
75000,100000
100000-125000
125000-150000
150000-175000
175000-200000
200000-225000
225000-250000
250000-275000
275000-300000
300000-325000
325000-350000
350000-375000
375000-400000
400000-425000
425000-450000
450000-475000
475000-500000
500000 and above

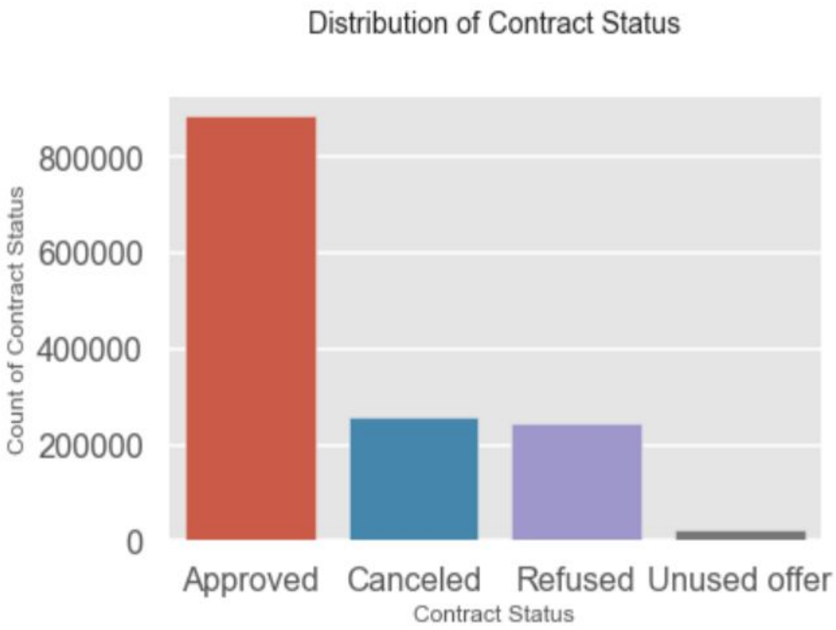

Customer without payment difficulties
1e6
AMT_CREDIT
Higher education
Secondary / secondary special
Incomplete higher
Lower secondary
Academic degree
NAME_EDUCATION_TYPE

Customer with payment difficulties
1e6
AMT_CREDIT
Secondary / secondary special
Incomplete higher
Higher education
Lower secondary
Academic degree
NAME_EDUCATION_TYPE

Customer without payment difficulties
1e6
AMT_INCOME_TOTAL
Higher education
Secondary / secondary special
Incomplete higher
Lower secondary
Academic degree
NAME_EDUCATION_TYPE

Customer with payment difficulties
1e6
AMT_INCOME_TOTAL
Secondary / secondary special
Incomplete higher
Higher education
Lower secondary
Academic degree
NAME_EDUCATION_TYPE

Customer without payment difficulties
1e6
AMT_CREDIT
Core staff
Laborers
Accountants
Managers
nan
Drivers
Sales staff
Cleaning staff
Private service staff
Medicine staff
Security staff
Cooking staff
High skill tech staff
Waiters/barmen staff
Low-skill Laborers
Realty agents
Secretaries
IT staff
HR staff
OCCUPATION_TYPE

Customer with payment difficulties
1e6
AMT_CREDIT
Laborers
Cooking staff
Sales staff
nan
Managers
Private service staff
Core staff
High skill tech staff
Medicine staff
Drivers
Security staff
Low-skill Laborers
Accountants
Cleaning staff
Realty agents
Secretaries
Waiters/barmen staff
IT staff
HR staff
OCCUPATION_TYPE

Conclusion:

1) Range of customers without payment of Academic degree is higher.
2) Customers without payment have more outliers.
3) Range of customers without payment more than with payment.

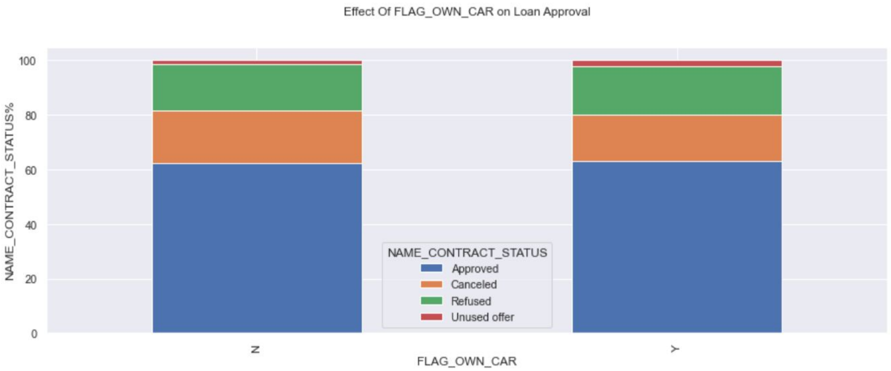


- Most of the loan applications are from repeat customers, out of the total applications 70% of customers are repeaters.
- They also get refused most often.

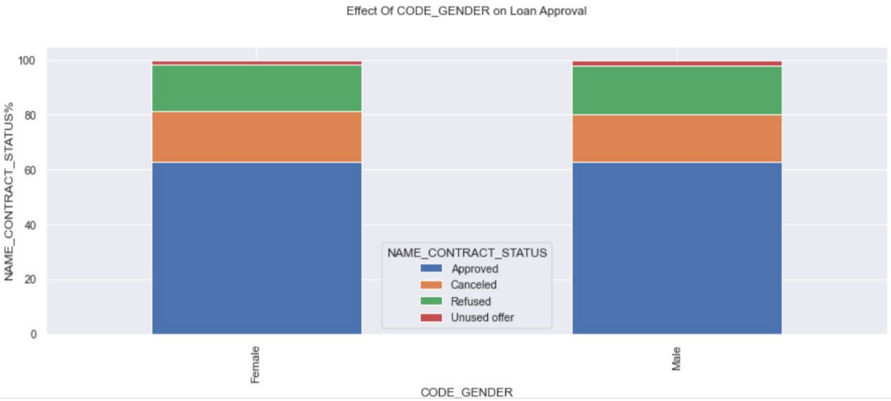



```
Approved        62.680190
Canceled        18.352119
Refused         17.356891
Unused offer     1.610799
Name: NAME_CONTRACT_STATUS, dtype: float64
```

Here 62% contract are approved while 18% are cancelled.

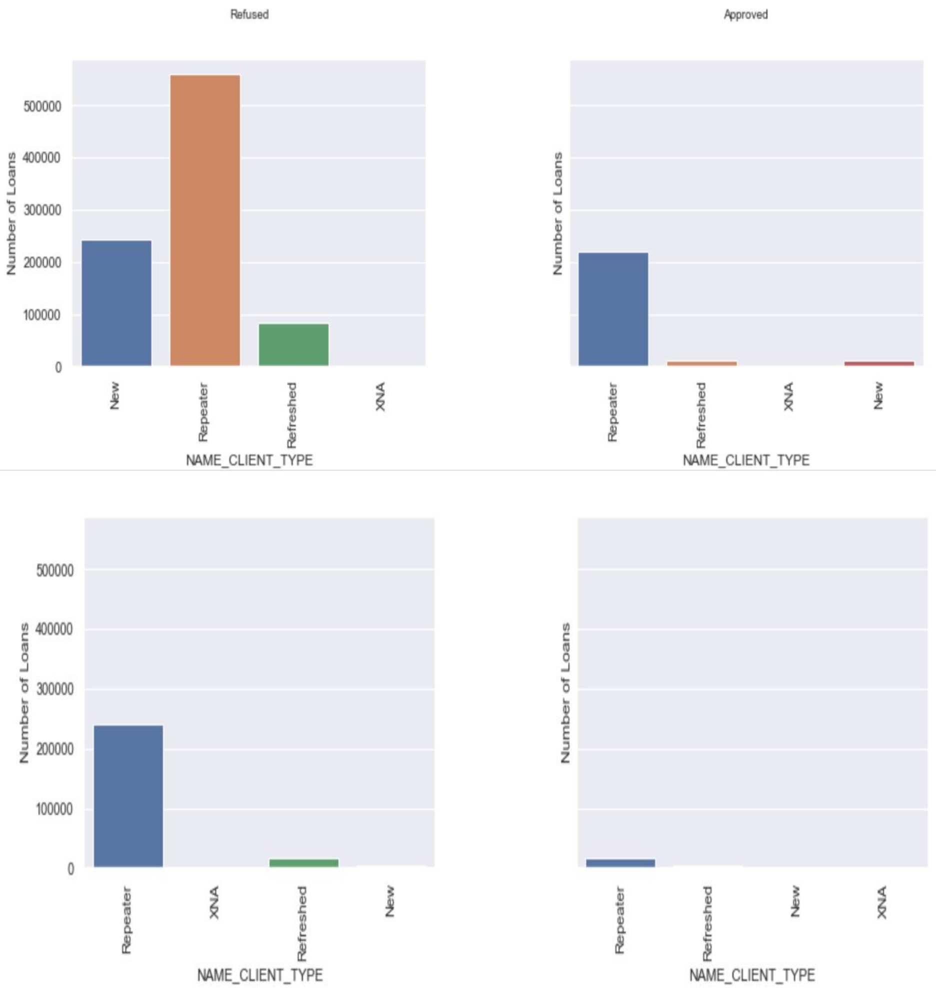


- We see that car ownership doesn't have any effect on application approval or rejection.
- But we saw earlier that the people who has a car has lesser chances of default.
- The bank can add more weightage to car ownership while approving a loan amount

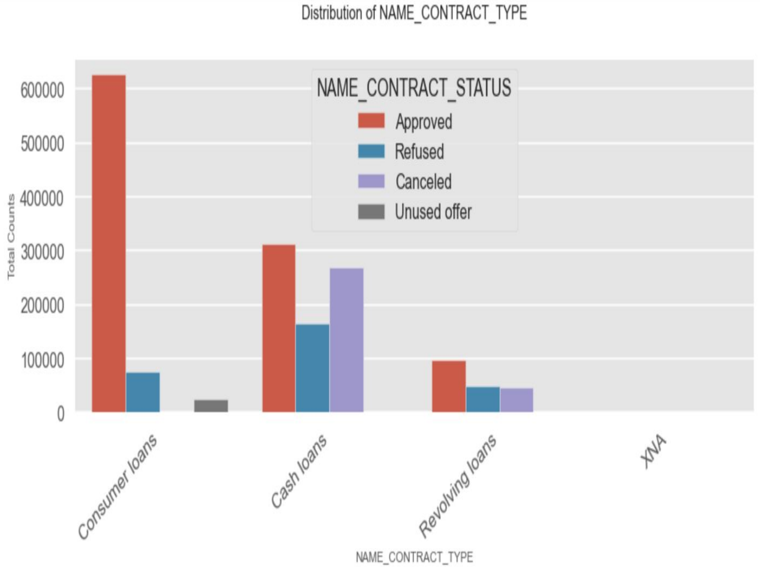


- We see that code gender doesn't have any effect on application approval or rejection.
- But we saw earlier that female have lesser chances of default compared to males.
- The bank can add more weightage to female while approving a loan amount.

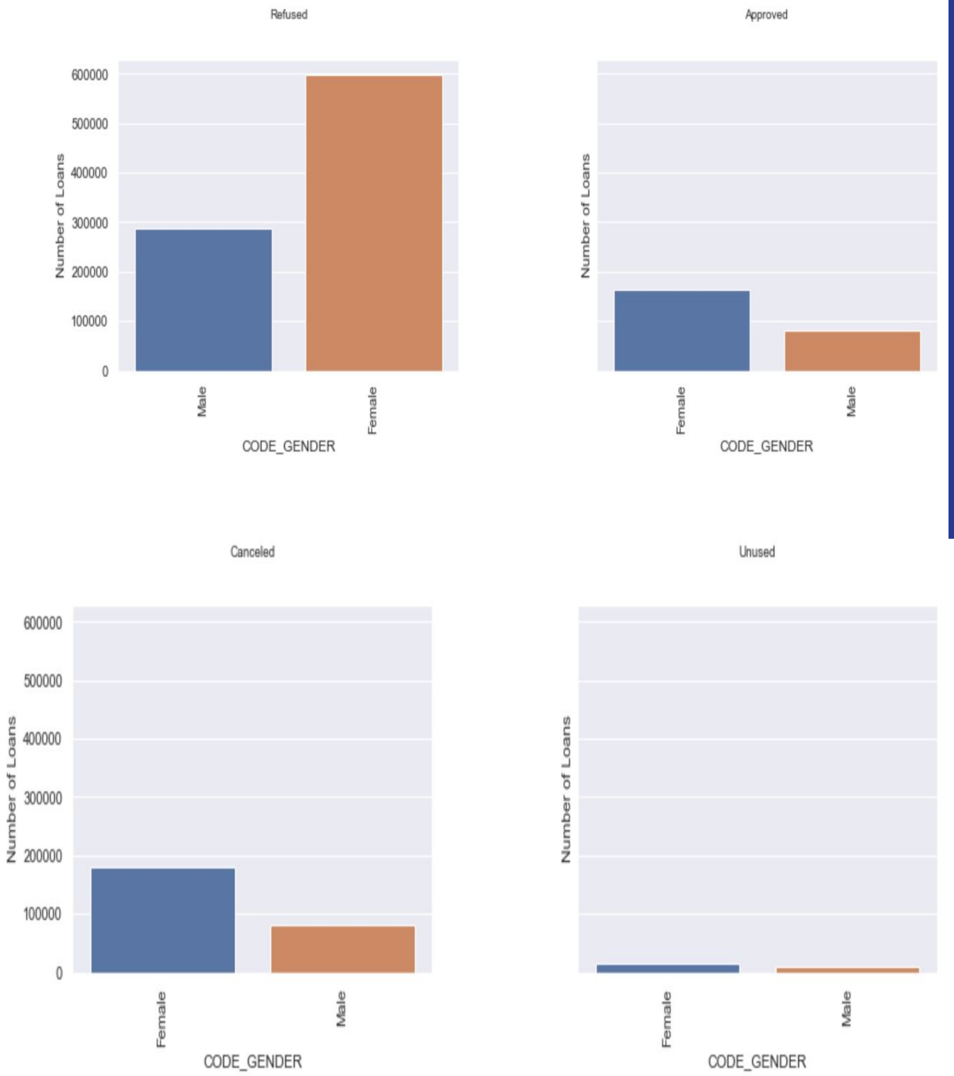
We can see that the people who were approved for a loan earlier, defaulted less often where as people who were refused a loan earlier have higher chances of defaulting.



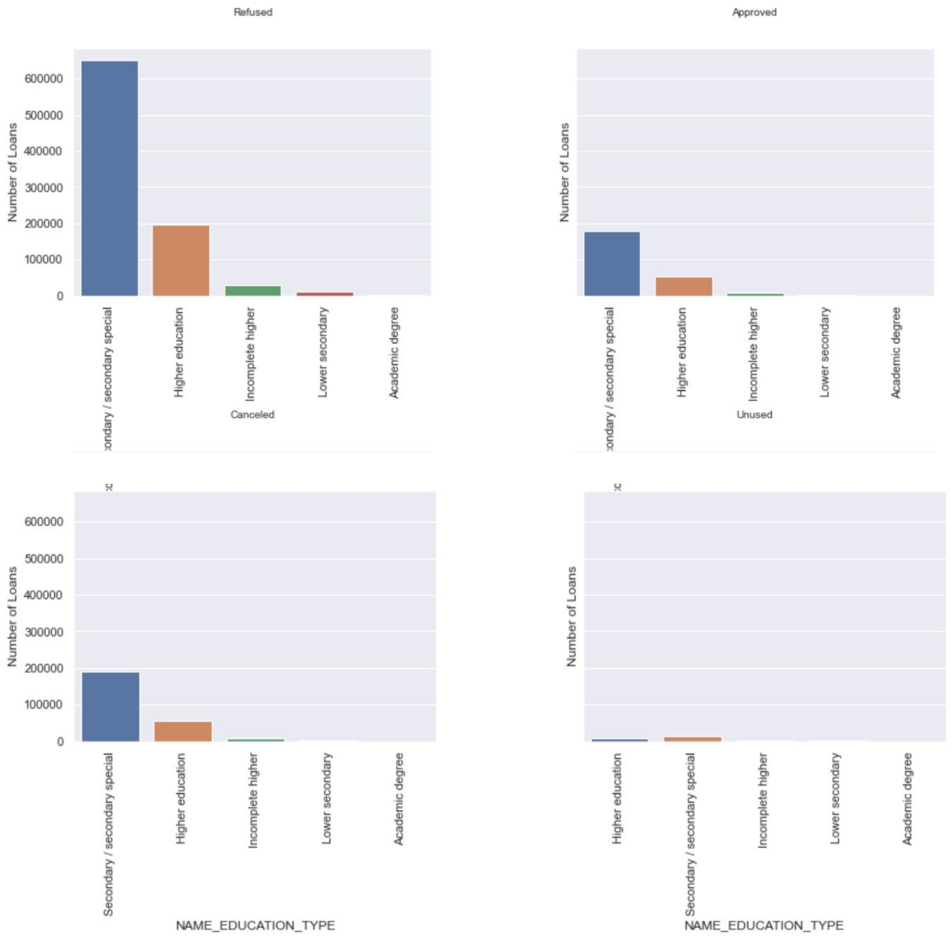
Here we can see that the Repeater is getting more Refused but also we can see that the it also getting more approved and even that it is getting more canceled and more usused.



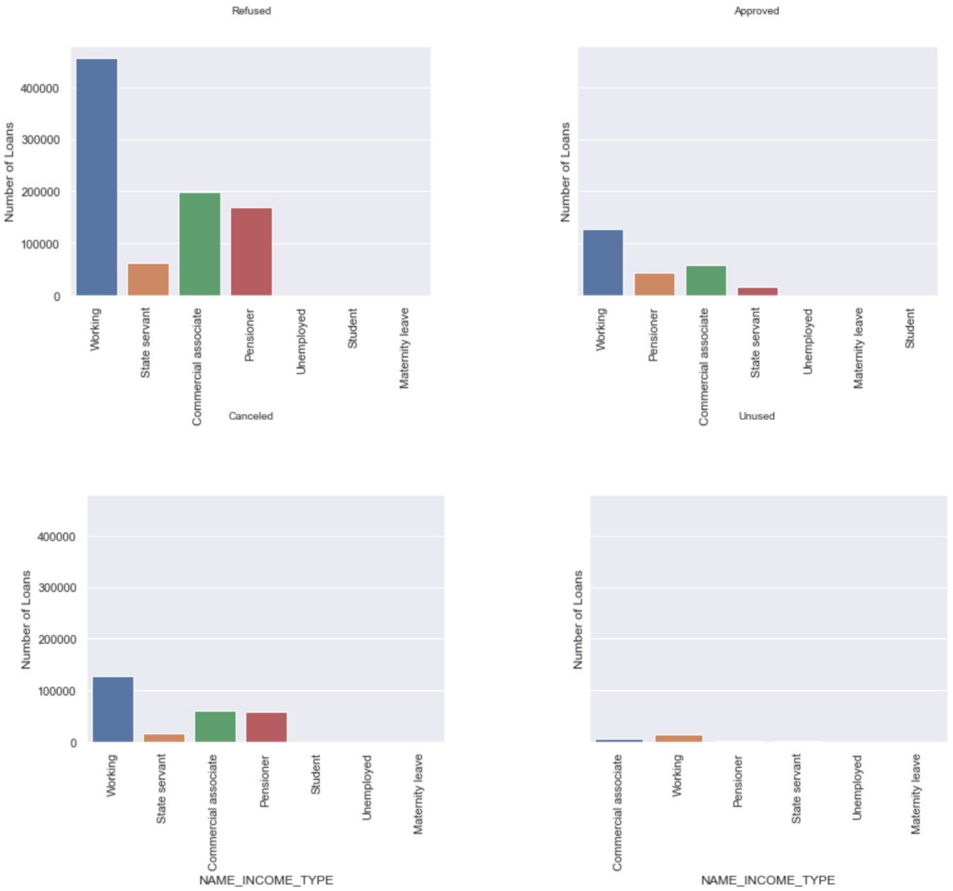
- Most of the applications are for 'Cash loan' and 'Consumer loan'.
- Although the cash loans are refused more often than others.



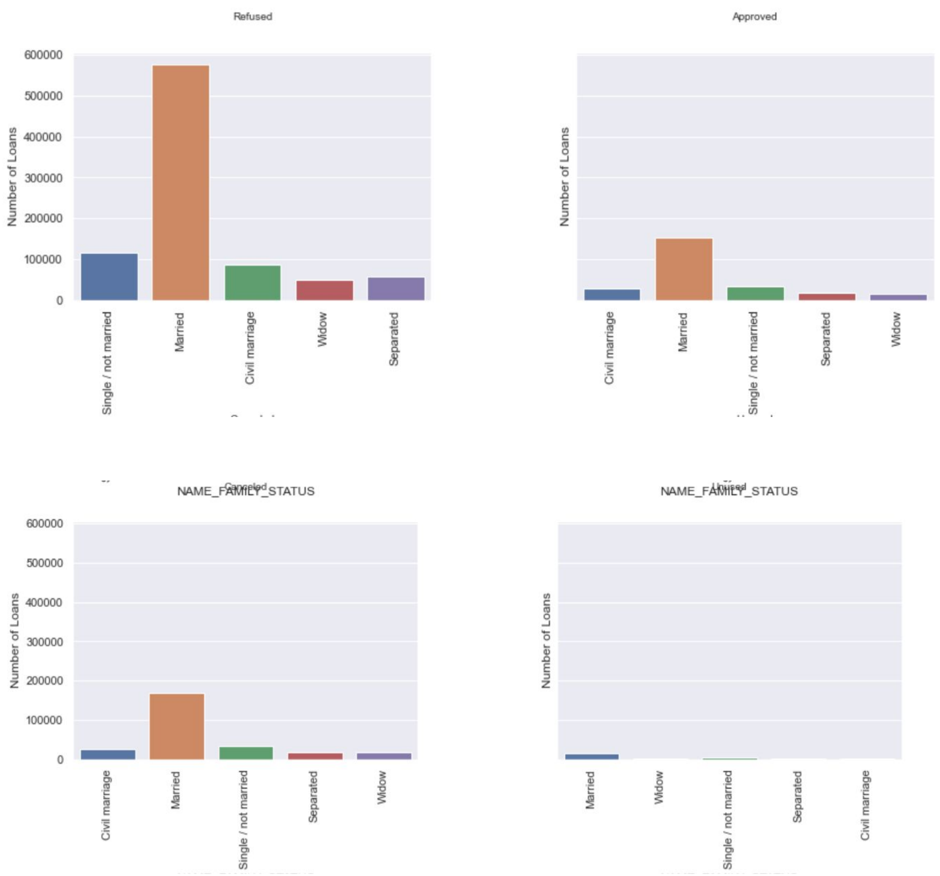
Here we can see that Female is getting more Refused more approved more canceled more unused but in case of male it is having average in every category



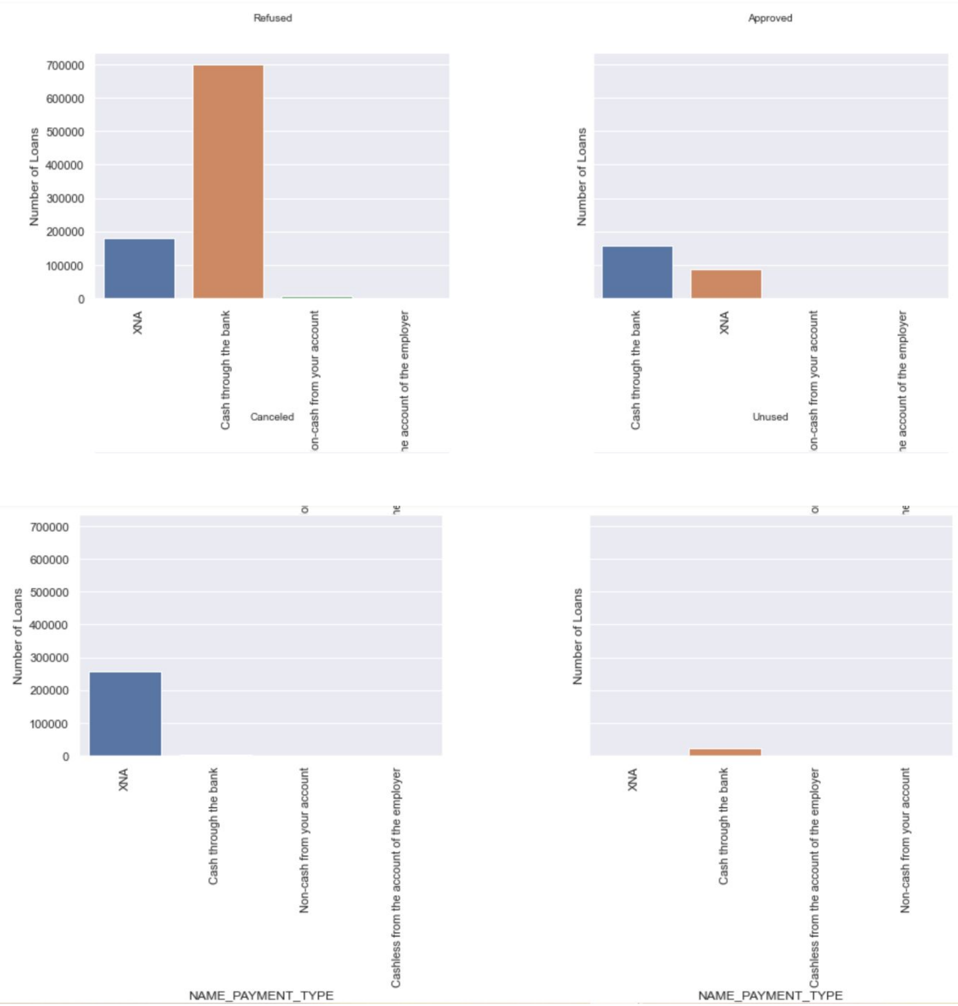
we can see that Secondary/ Secondary special is more effective in every case



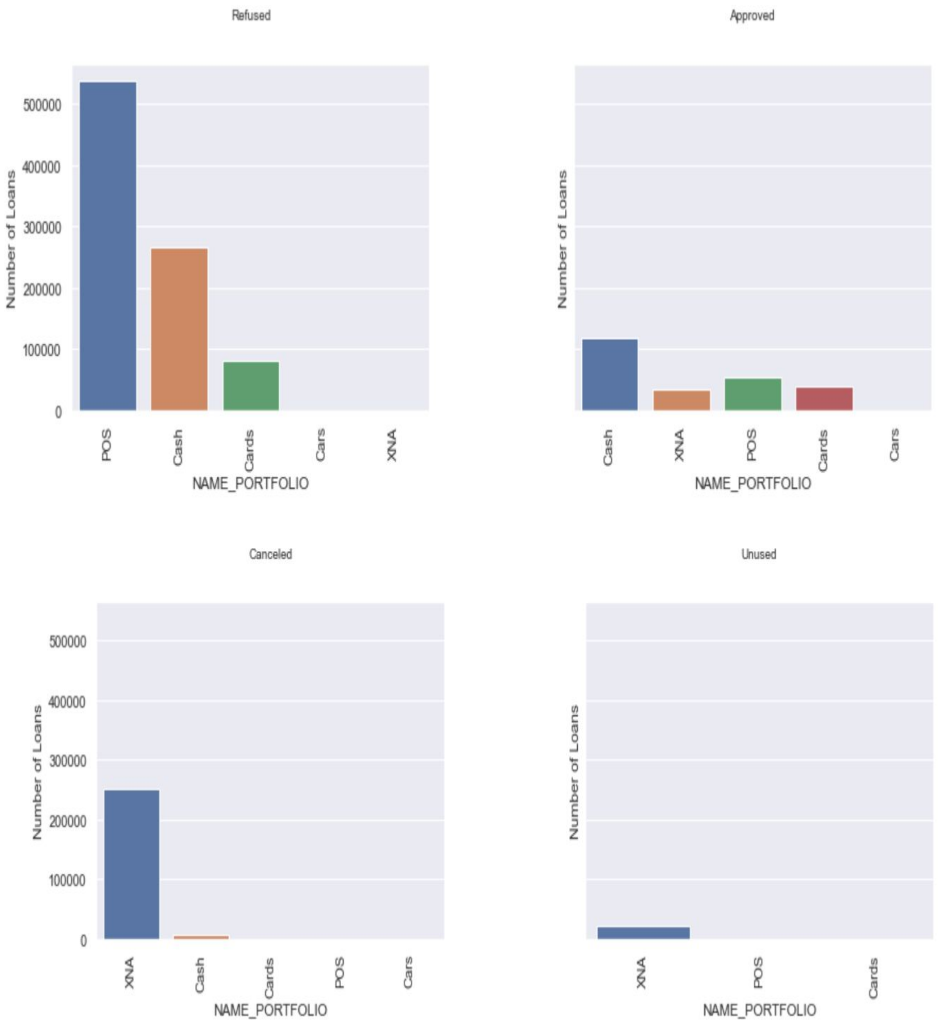
- Here we can see that the working type people are applying more loans as compare to others
- Also Commercial associates people are taking more loans.



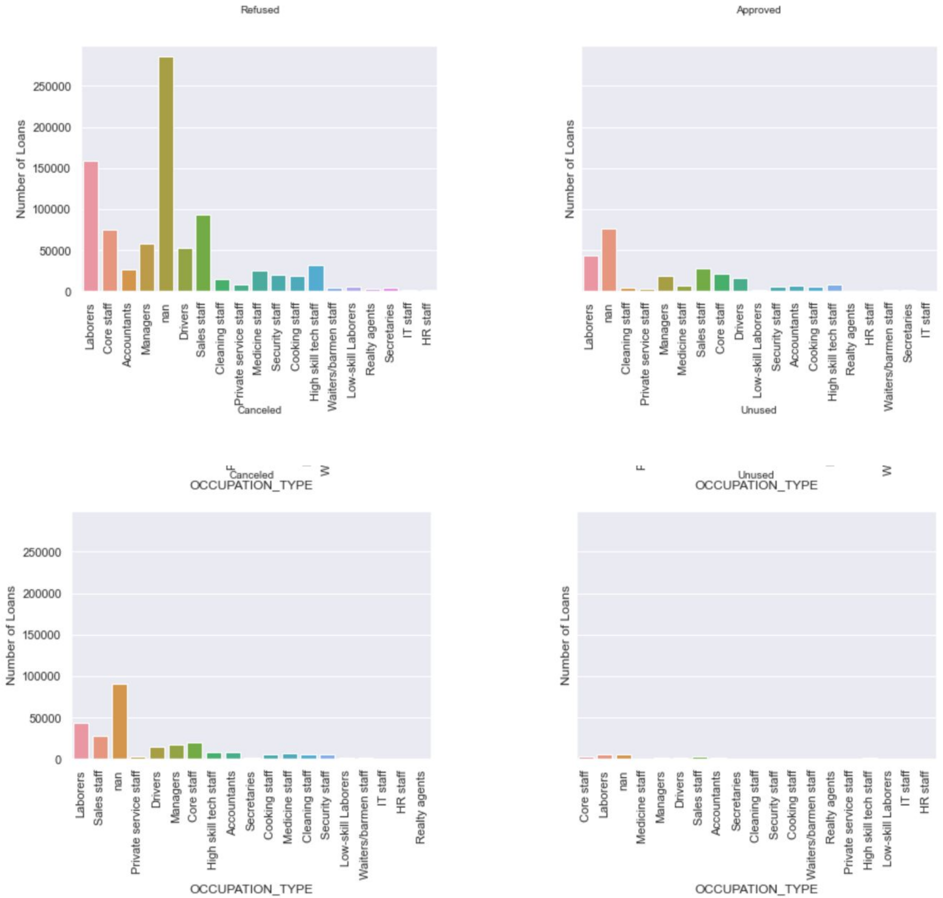
Here we can see that the Married people are applying and taking loans more than the others.



Here we can see that the people are taking more loan in format of cash through the bank.



Here most approved loan were through POS and Most refused loans were in cash.



- Here laborers are getting most refused and most approved loans.
- And also Sales staff is also getting the second most refused and approved loans.

Recommended groups to target:

1. Clients working as state servants.
2. Old people with any income.
3. Clients with high income.
4. Female clients preferably old.
5. Clients whose previous loans are approved.
6. Clients who have unused loan status previously.

Risky clients:

1. Male clients with civil marriage.
2. Previously refused loan status group.
3. Lower secondary educated clients.