EDA CASE STUDY

LOAN DEFAULTER CASE STUDY

PRESENTED BY-CR LIKITH KUMAR

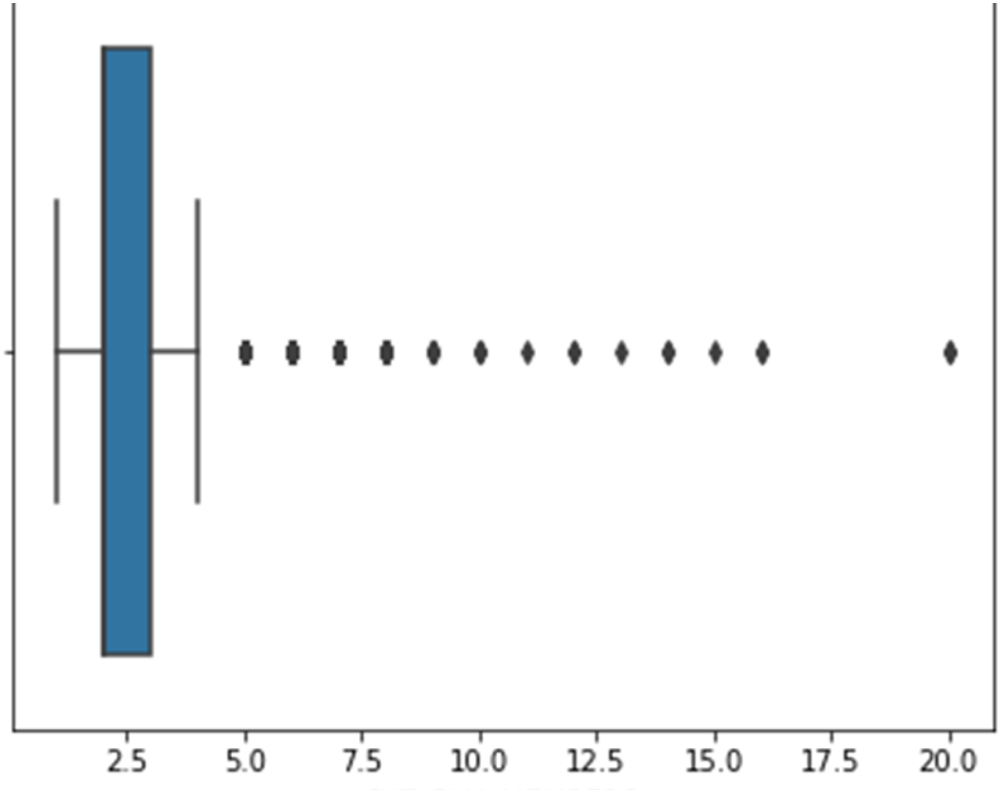
Objective

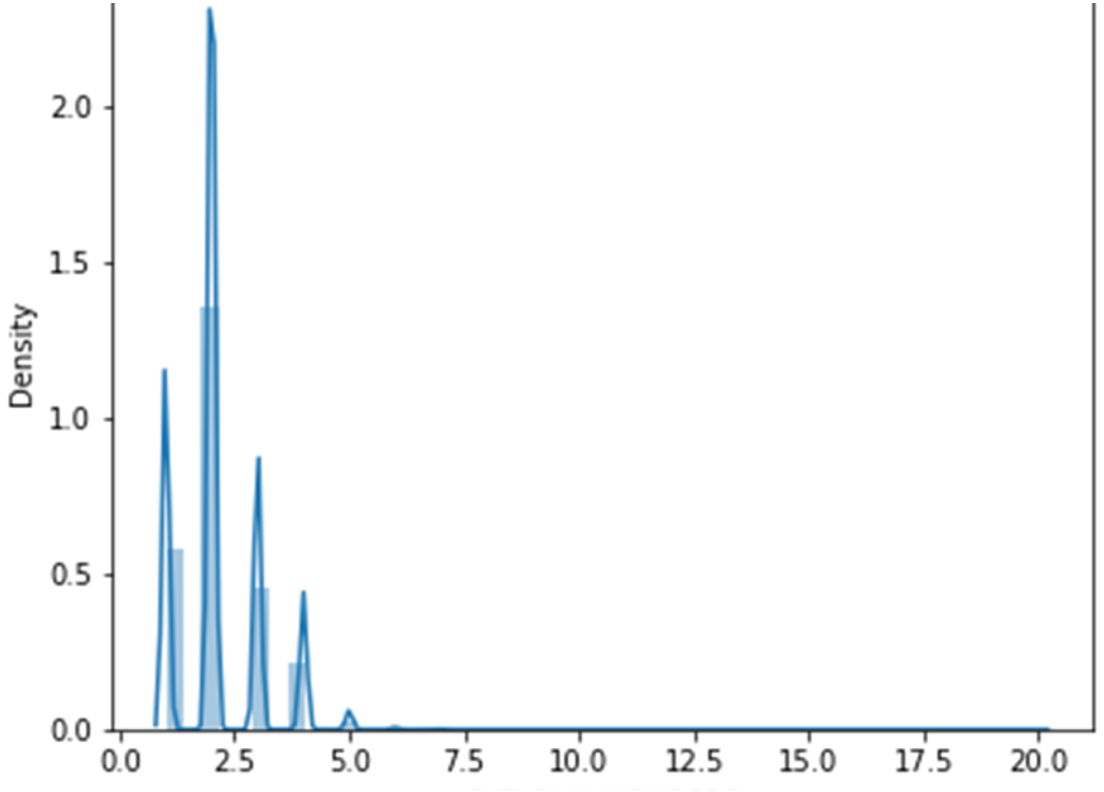
- This case study aims to give an idea of applying EDA in a real business scenario. And this case study identify the understanding of risk in banking and financial sectors.
- Find out applicant who are facing payment difficulties at present or have history of becoming loan defaulter so that Banks should not approve loan application request from these customers as it can lead to a business loss.

Steps

- Data understanding
- Quality data check of datasets
- Data cleaning
- Finding the missing values in data
- Checking outliers
- Check for Data Imbalance and find Ratio of Imbalance
- Univariate , bivariate analysis and correlation of columns
- Merging of application data with previous application data
- Conclusion

Count Of Family Members





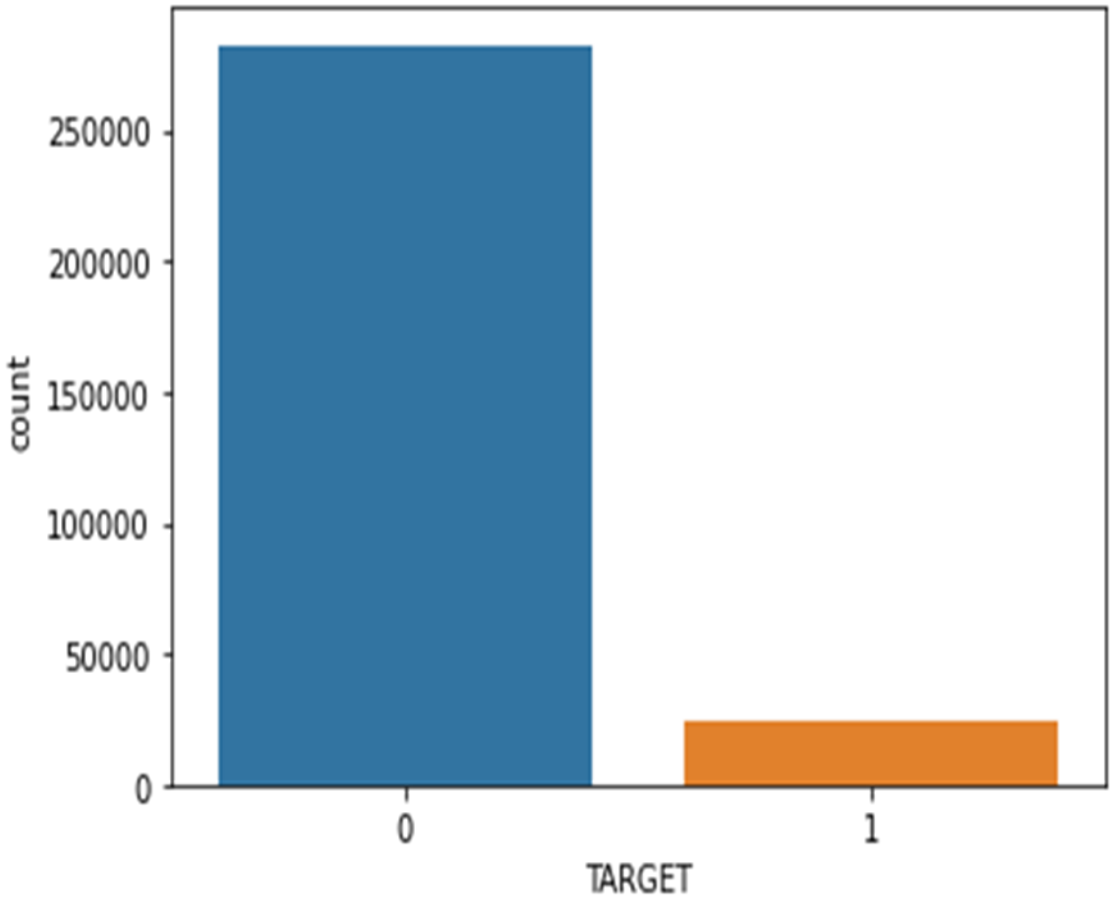
Univariate Analysis

- This univariate analysis shows the difference between payment difficulty (defaulters) and non-payment difficulty(non – defaulters) banking sector who has taken the loan from bank.

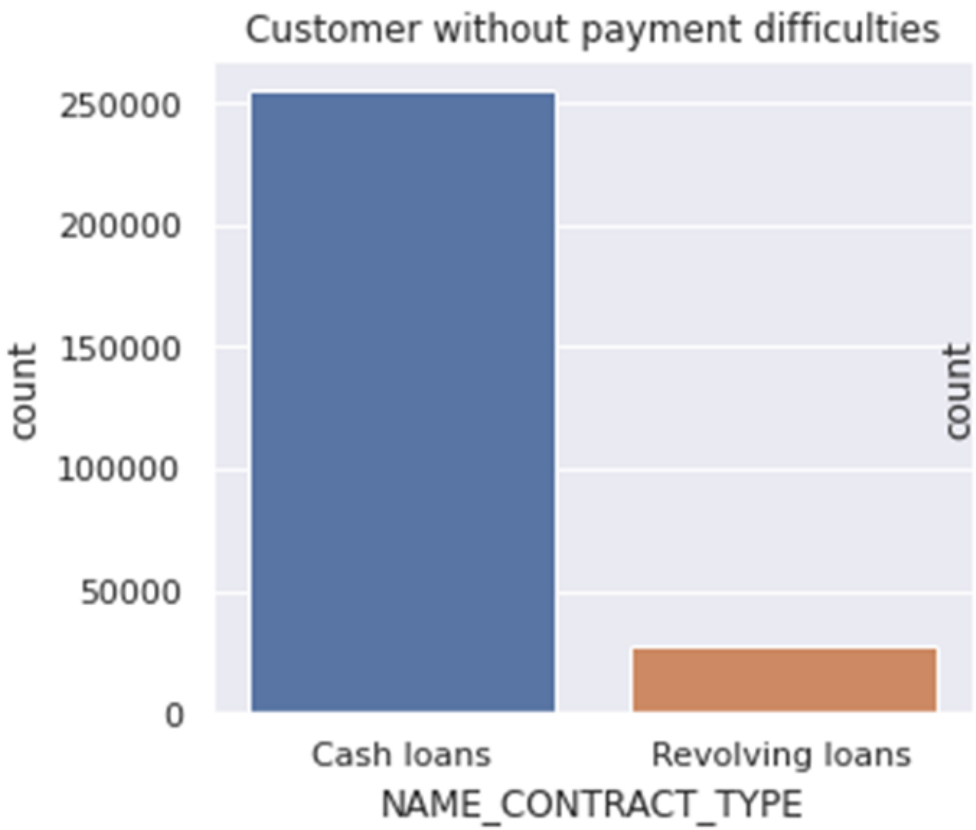
TARGET

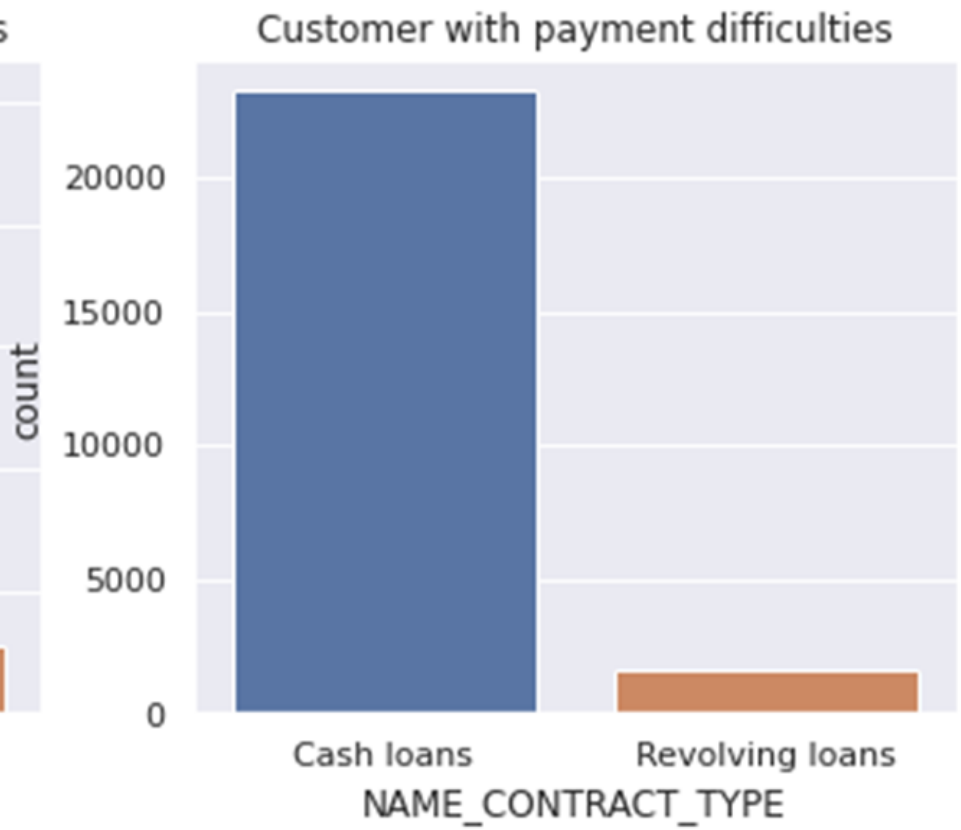
Data Imbalance Ratio is 11.4

So we can say that for almost every 11 non-defaulter there is 1 defaulter who is likely not paying the loan

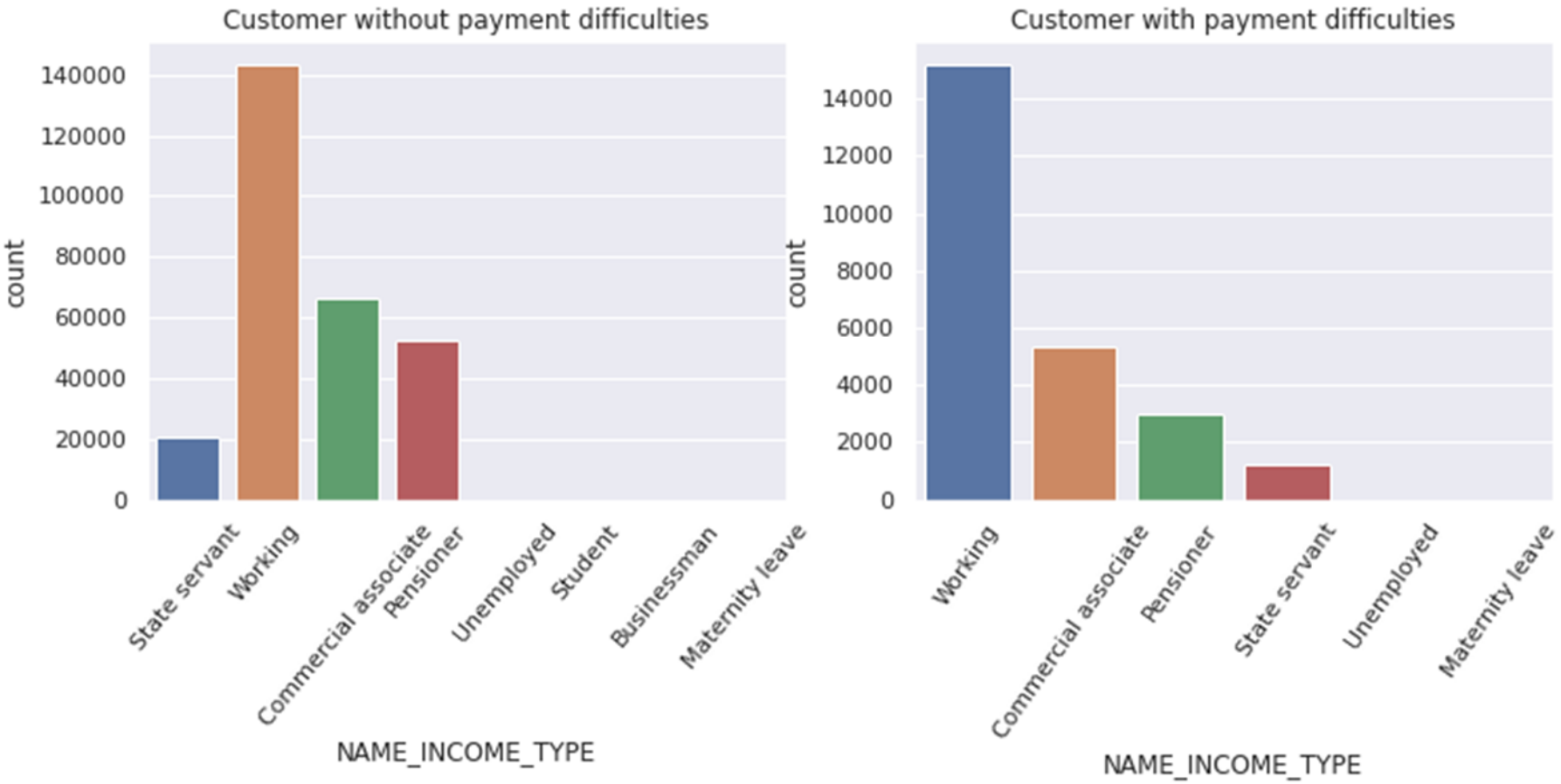


CONTRACT TYPE

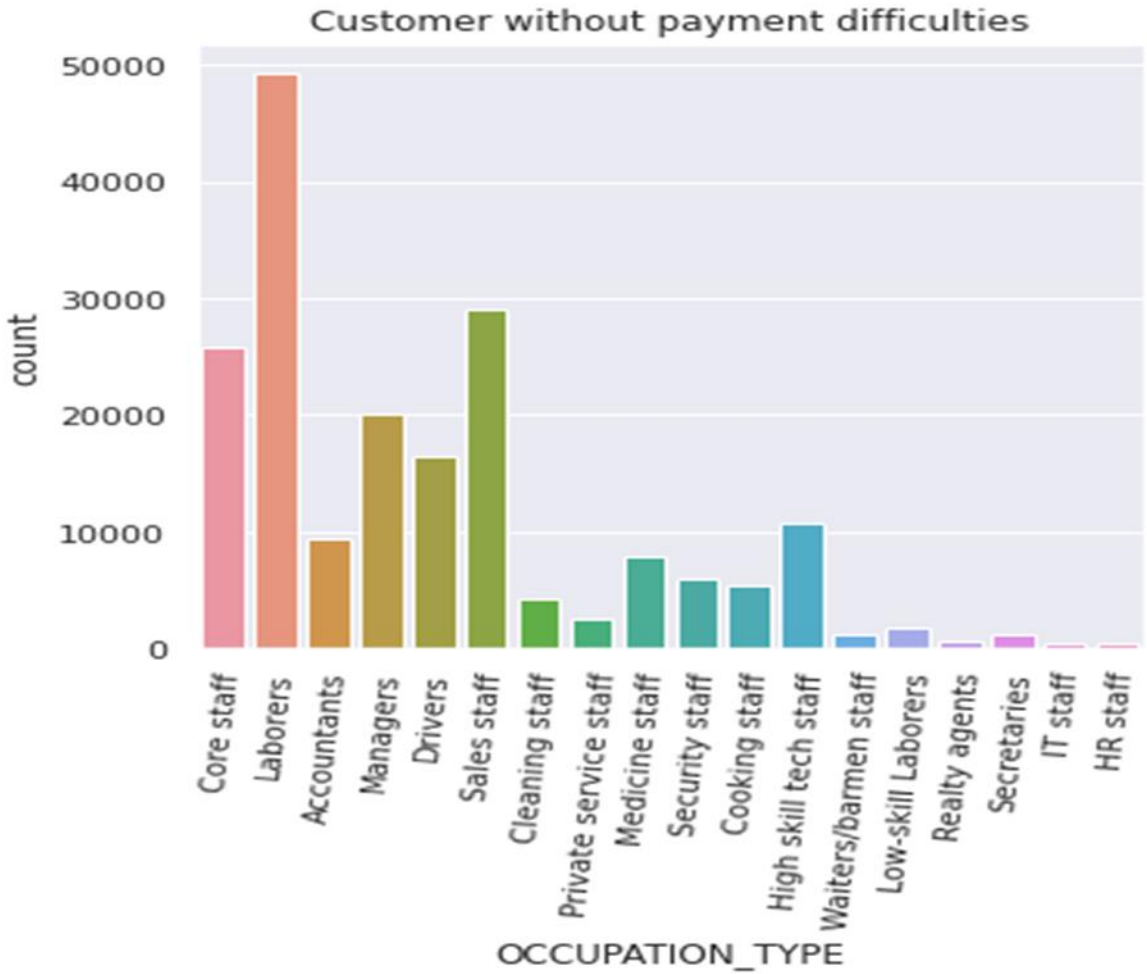


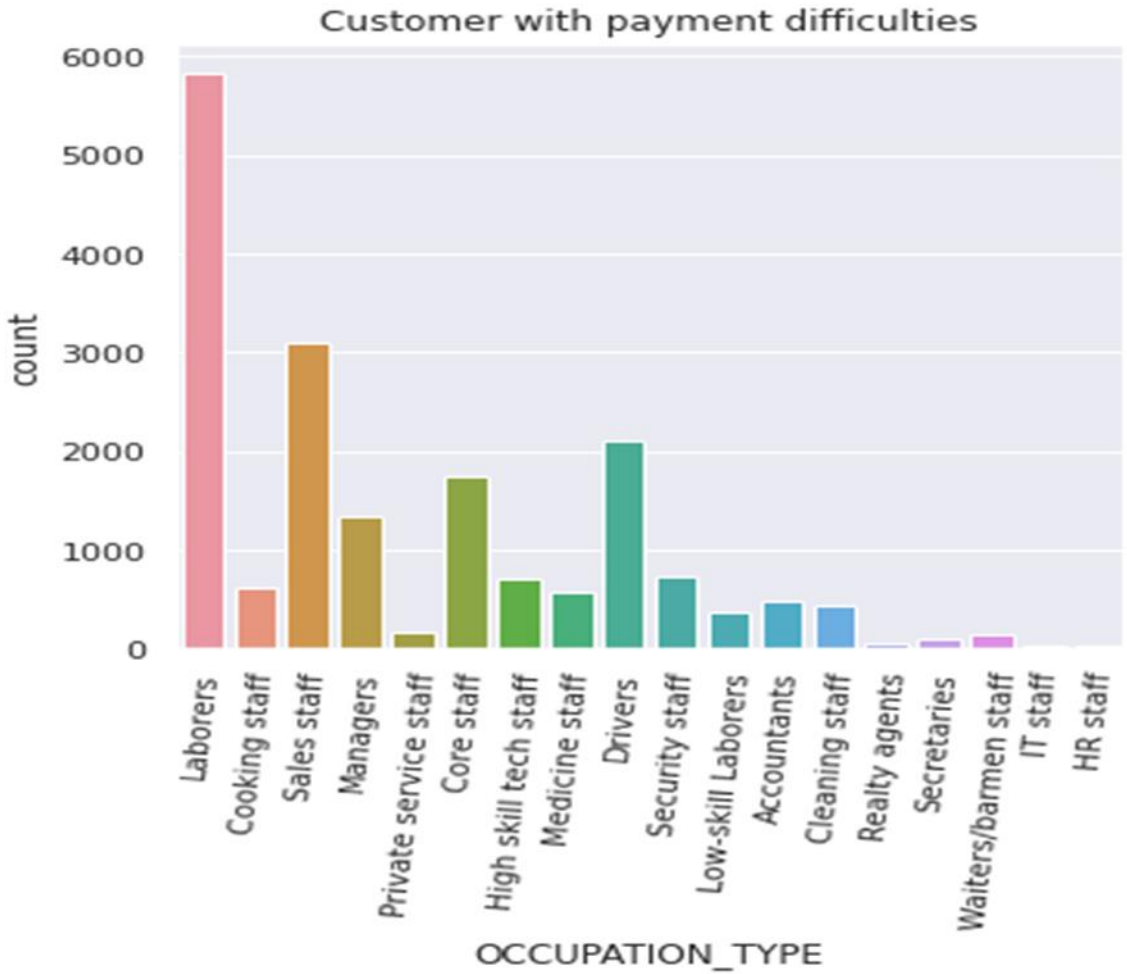


Income Type

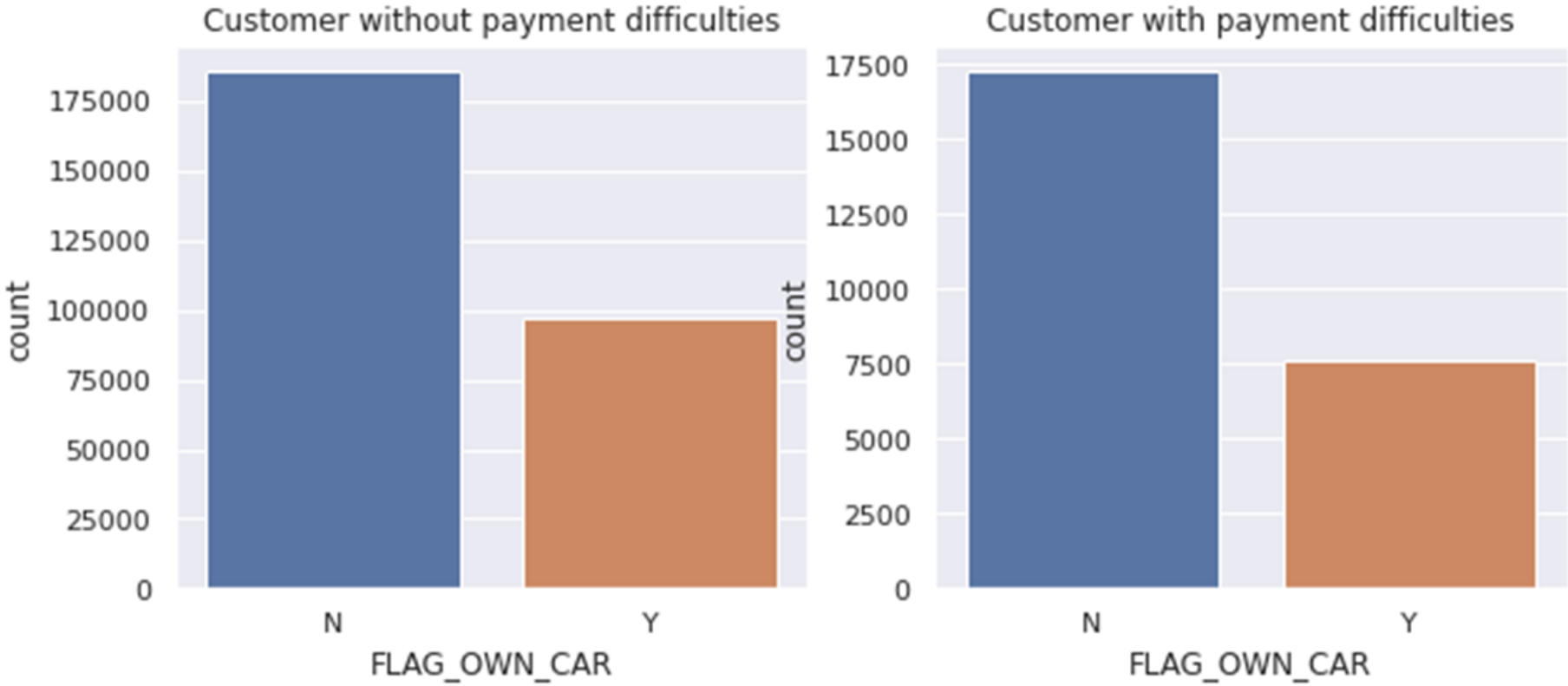

Customer without payment difficulties
count
140000
120000
100000
80000
60000
40000
20000
0
State servant
Working
Commercial associate
Pensioner
Unemployed
Student
Businessman
Maternity leave
NAME_INCOME_TYPE
Customer with payment difficulties
count
14000
12000
10000
8000
6000
4000
2000
0
Working
Commercial associate
Pensioner
State servant
Unemployed
Maternity leave
NAME_INCOME_TYPE

Occupation Type

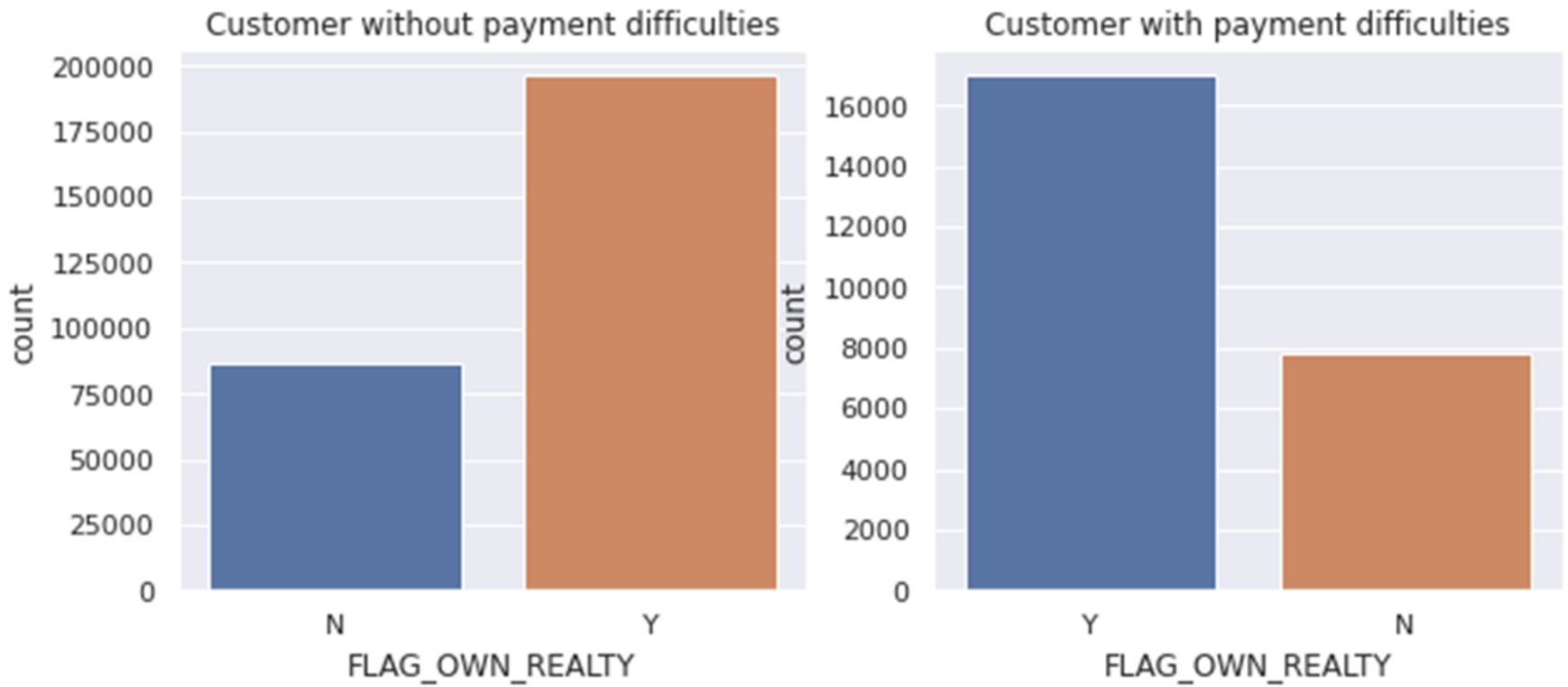




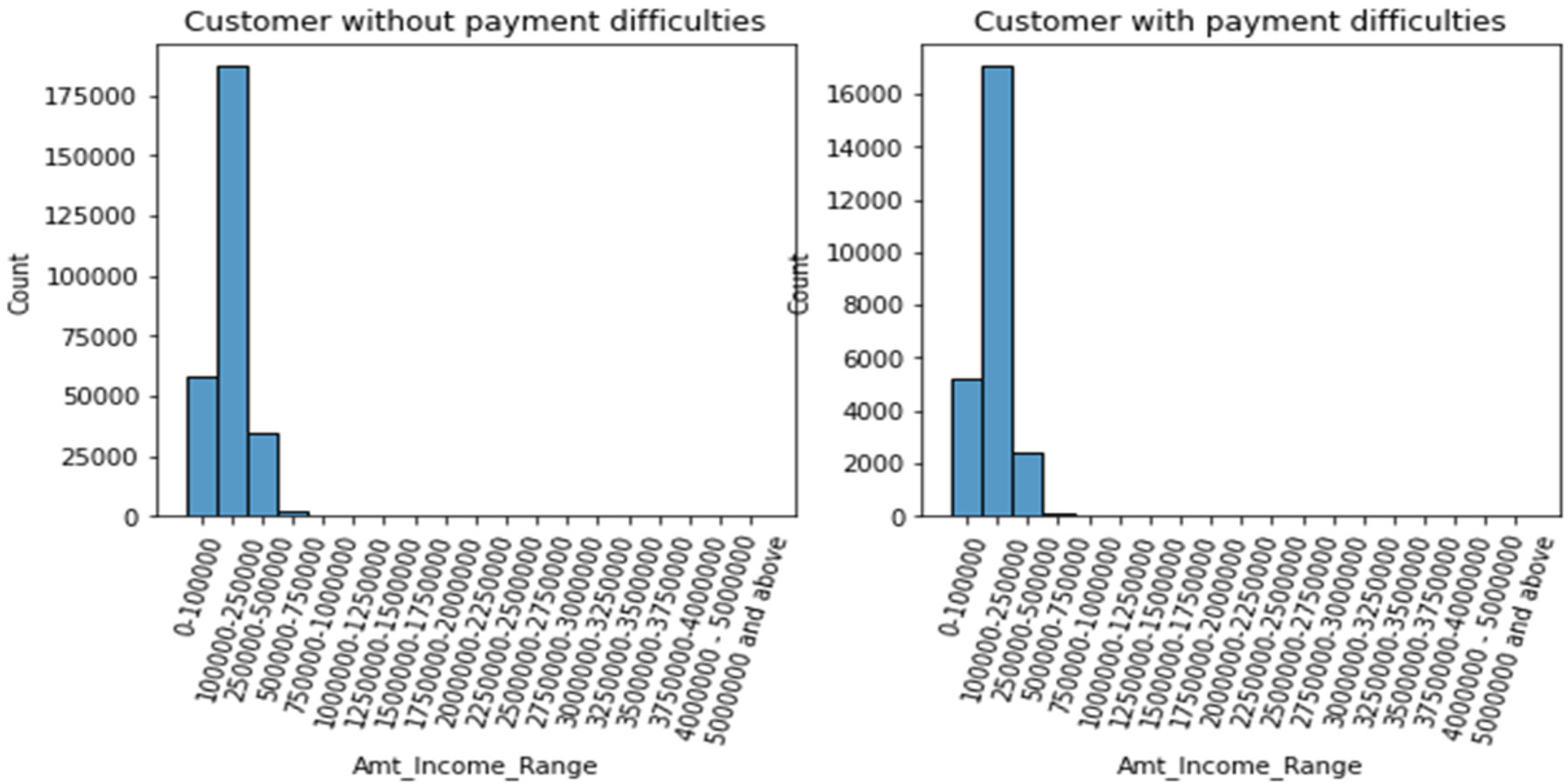
Owns A Car


Customer without payment difficulties
count
175000
150000
125000
100000
75000
50000
25000
0
N
Y
FLAG_OWN_CAR
Customer with payment difficulties
count
17500
15000
12500
10000
7500
5000
2500
0
N
Y
FLAG_OWN_CAR

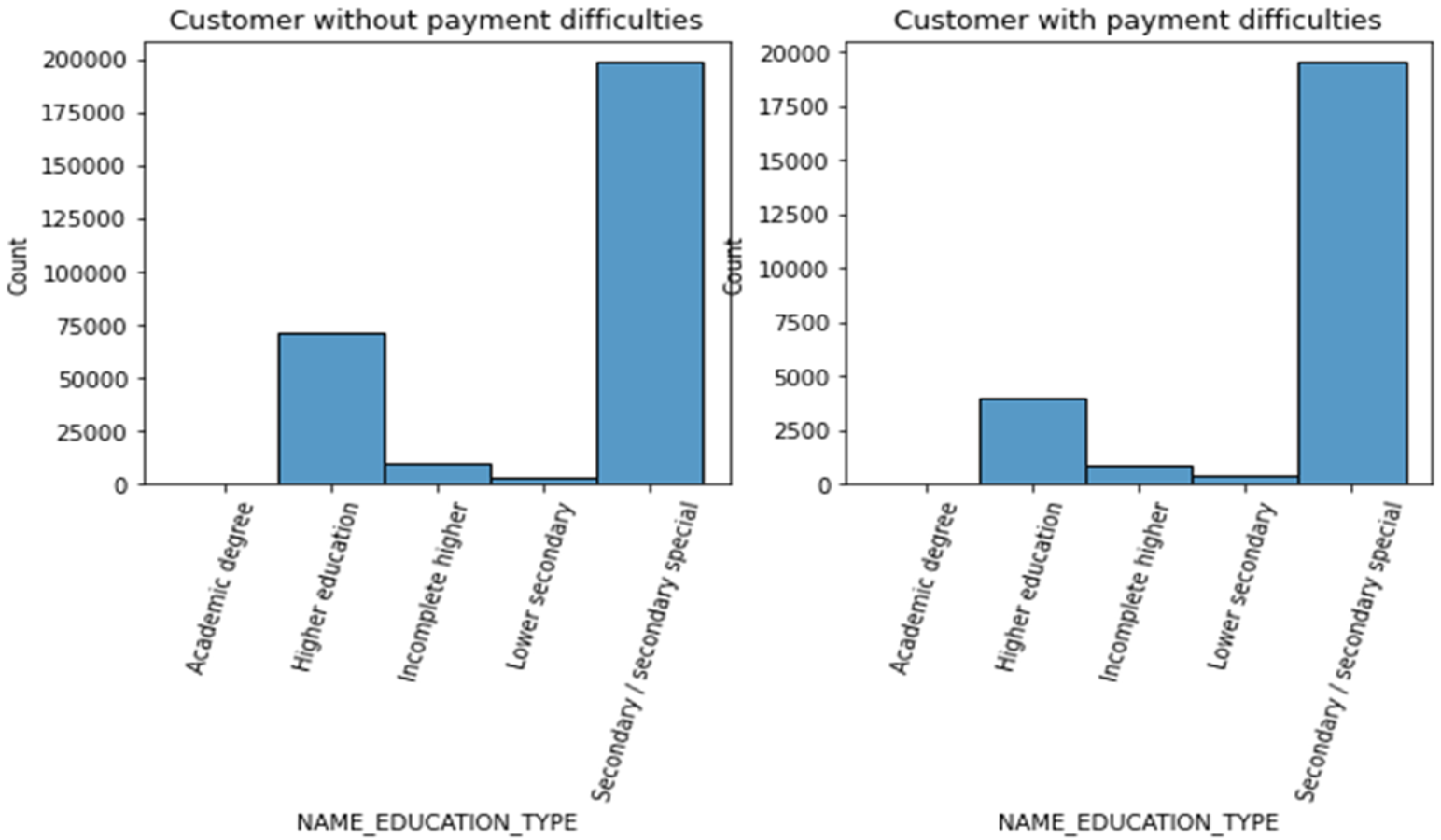
Owning A Realty



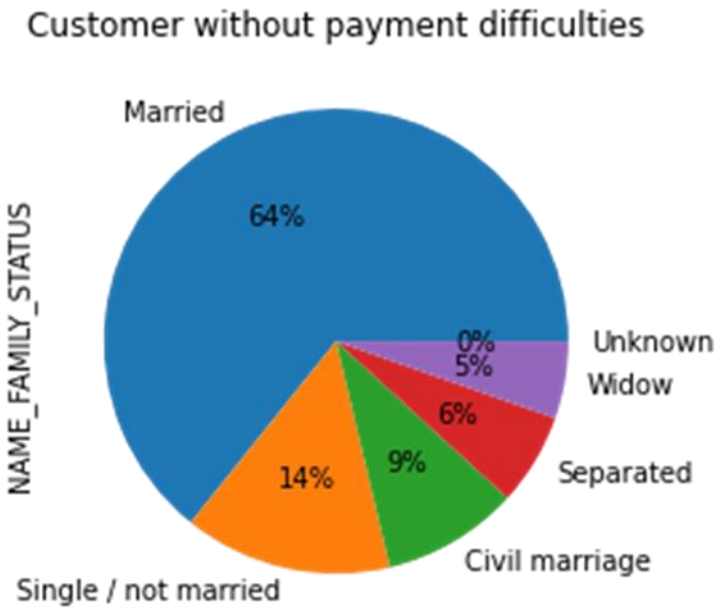
Income Range

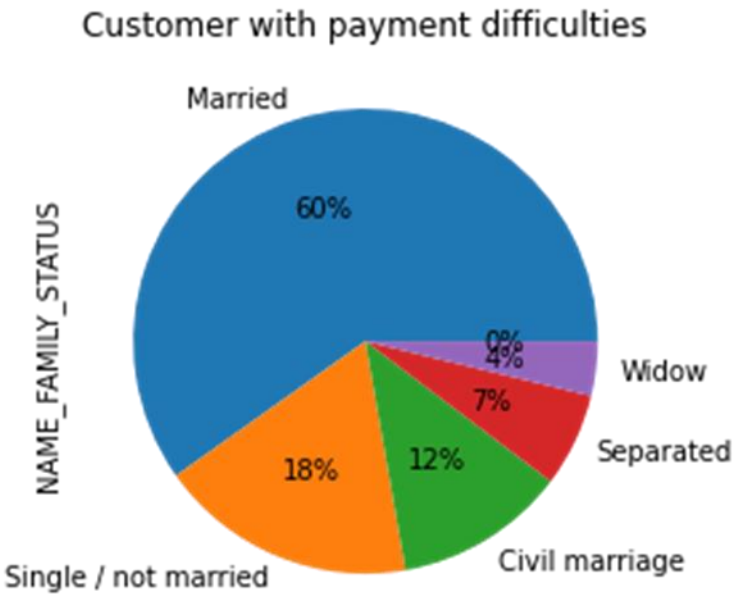


Education Type

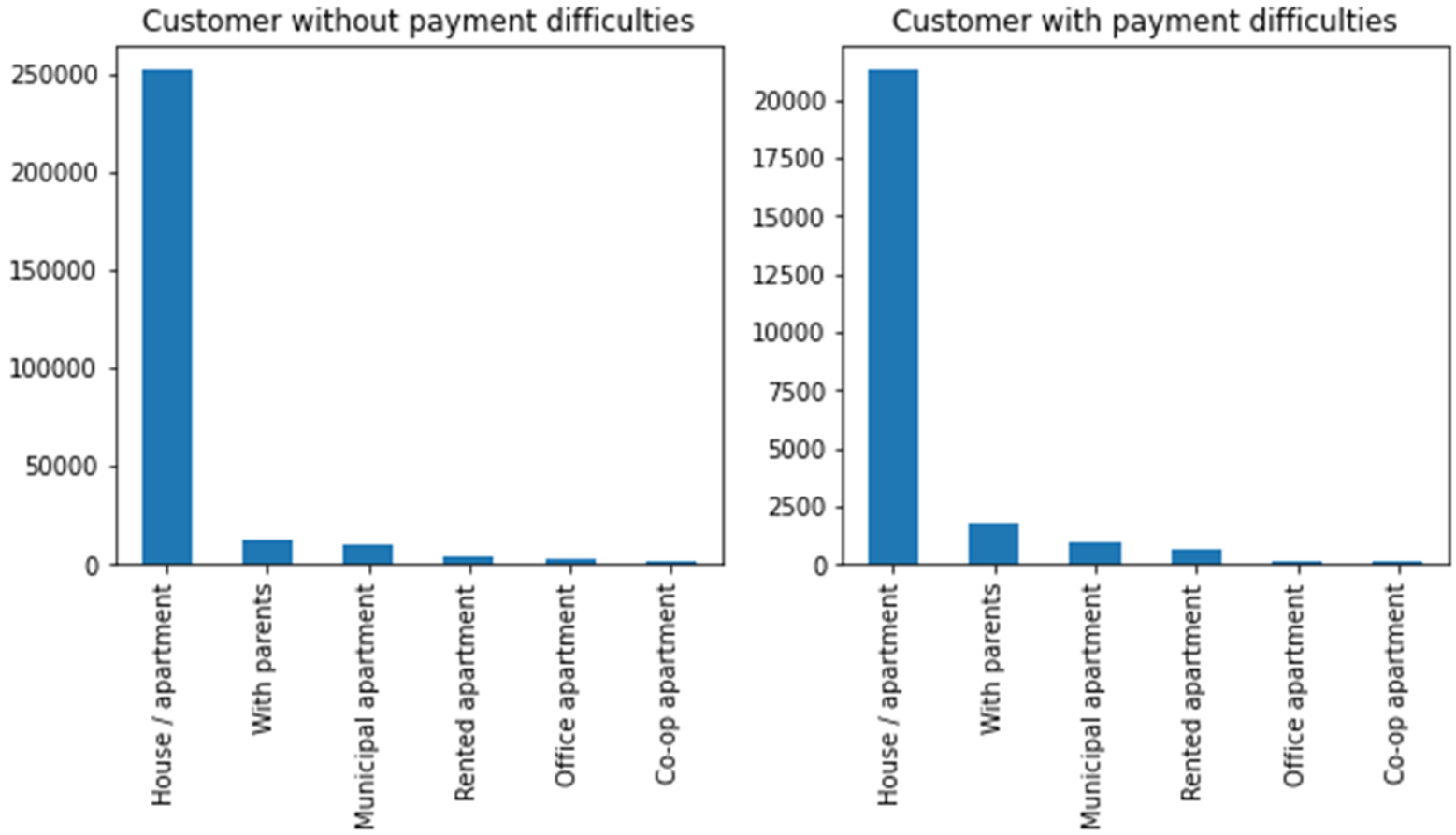


Family Status

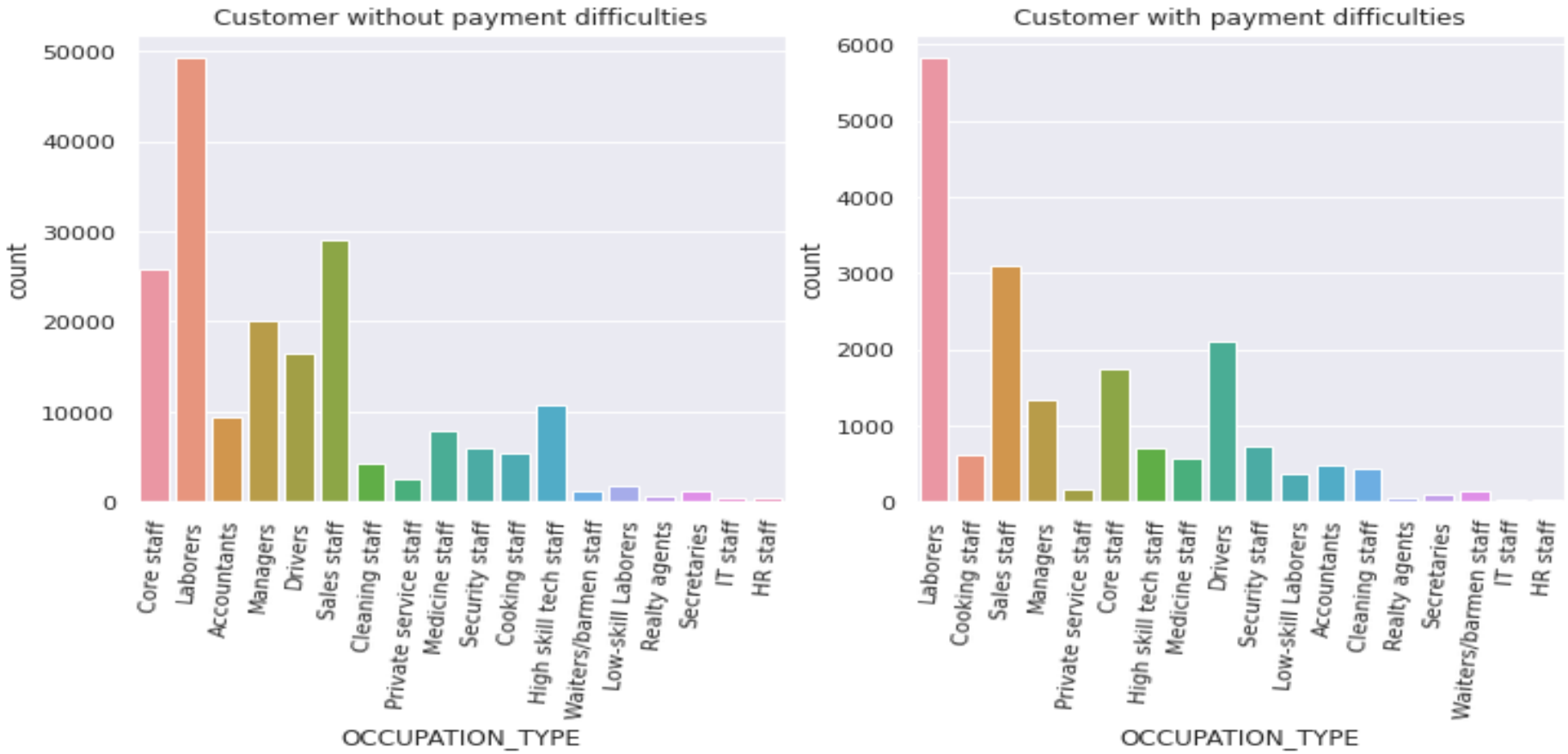




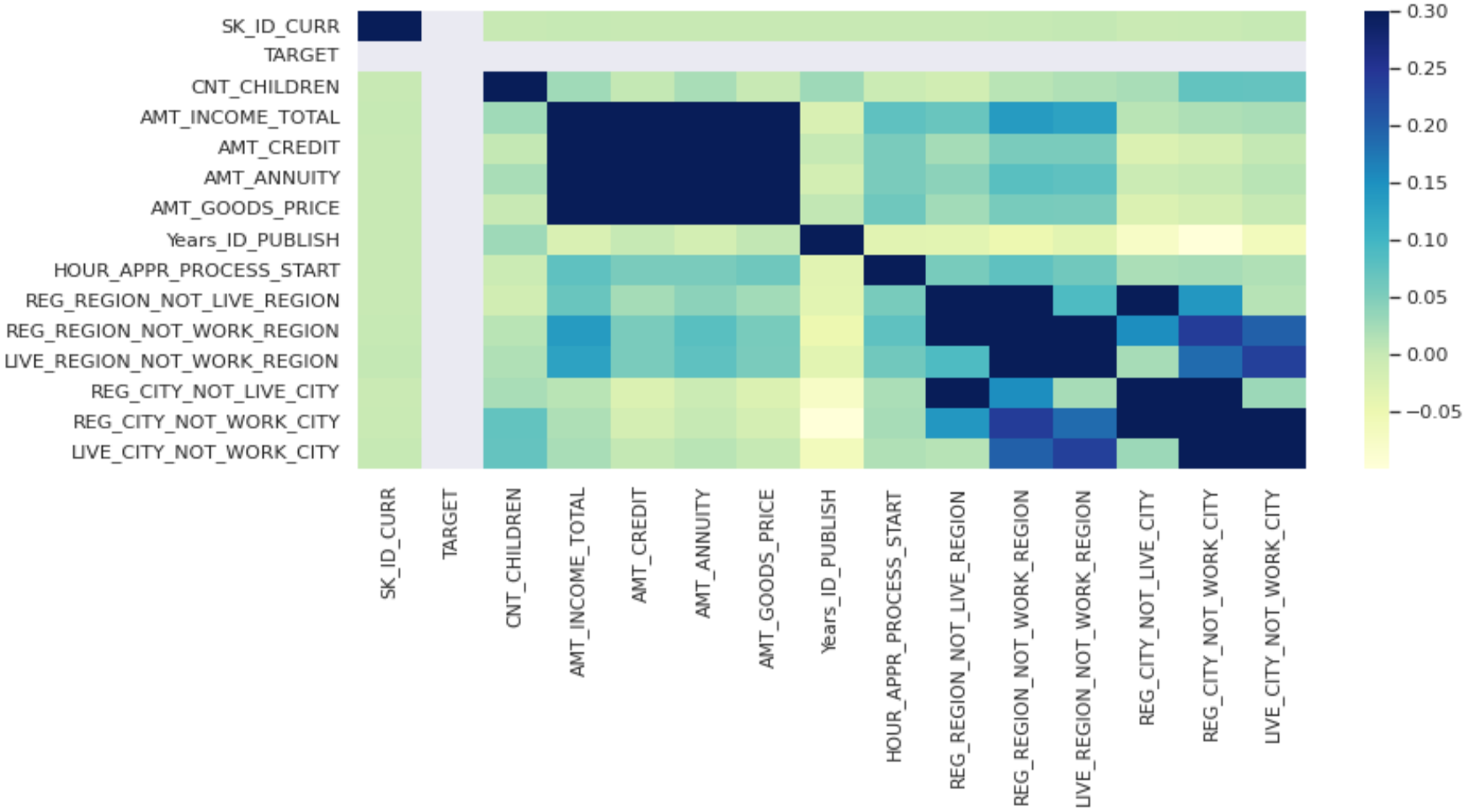
House Type


Customer without payment difficulties
250000
200000
150000
100000
50000
0
House / apartment
With parents
Municipal apartment
Rented apartment
Office apartment
Co-op apartment
Customer with payment difficulties
20000
17500
15000
12500
10000
7500
5000
2500
0
House / apartment
With parents
Municipal apartment
Rented apartment
Office apartment
Co-op apartment

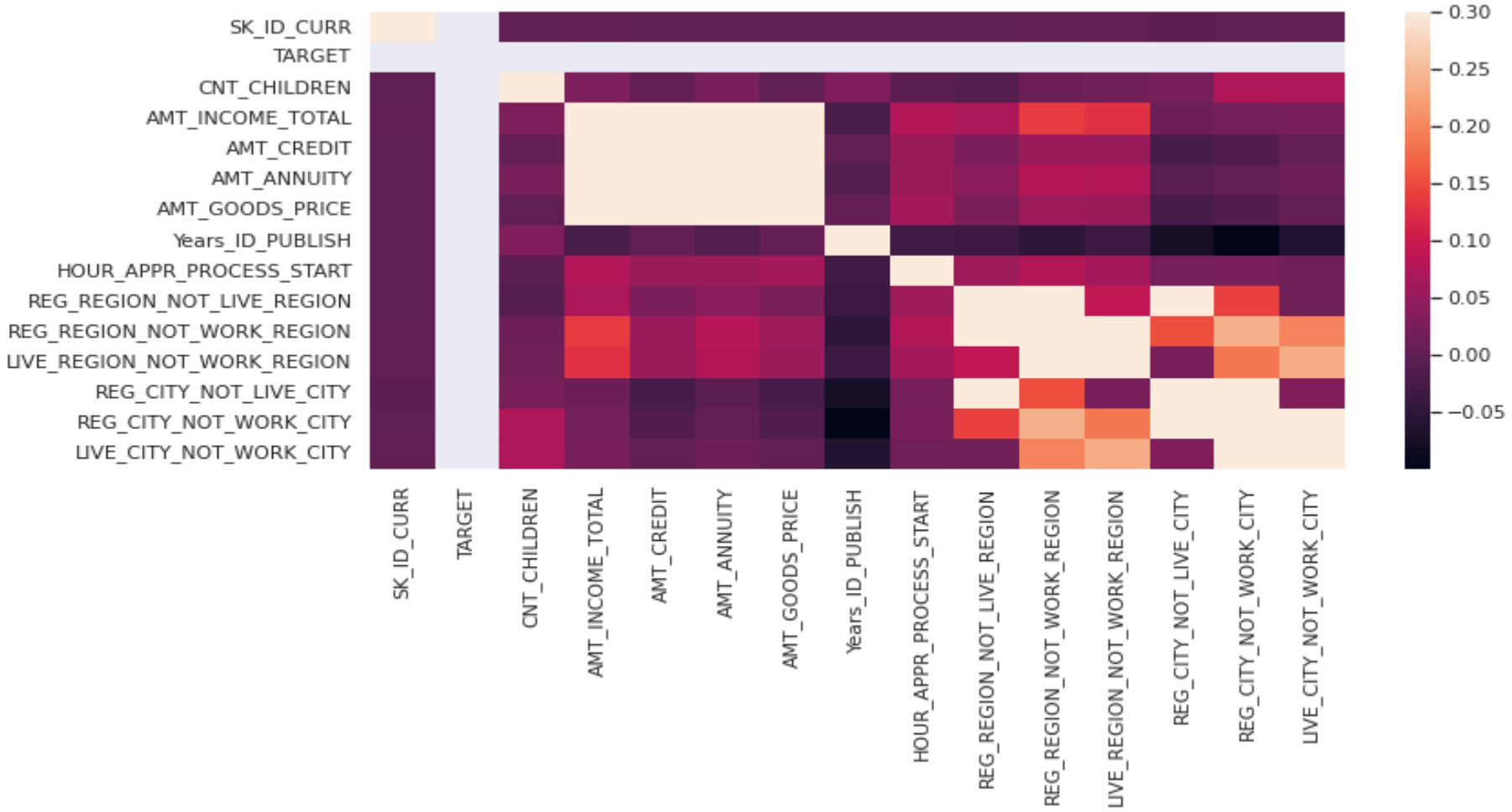
Occupation Type



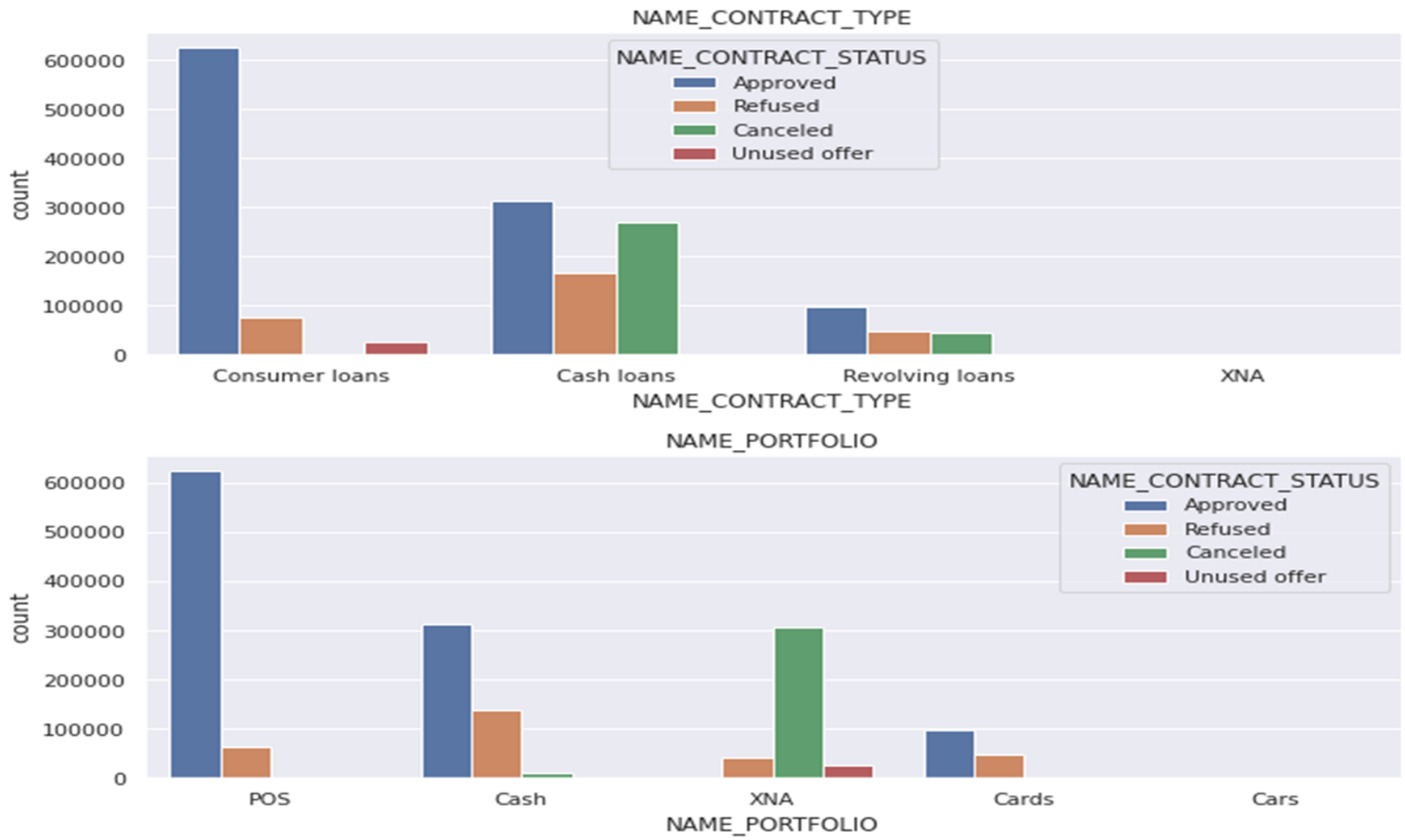
Heat map for defaulters data



Heat map for non defaulters data



NAME_CONTRACT_TYPE and NAME_PORTFOLIO

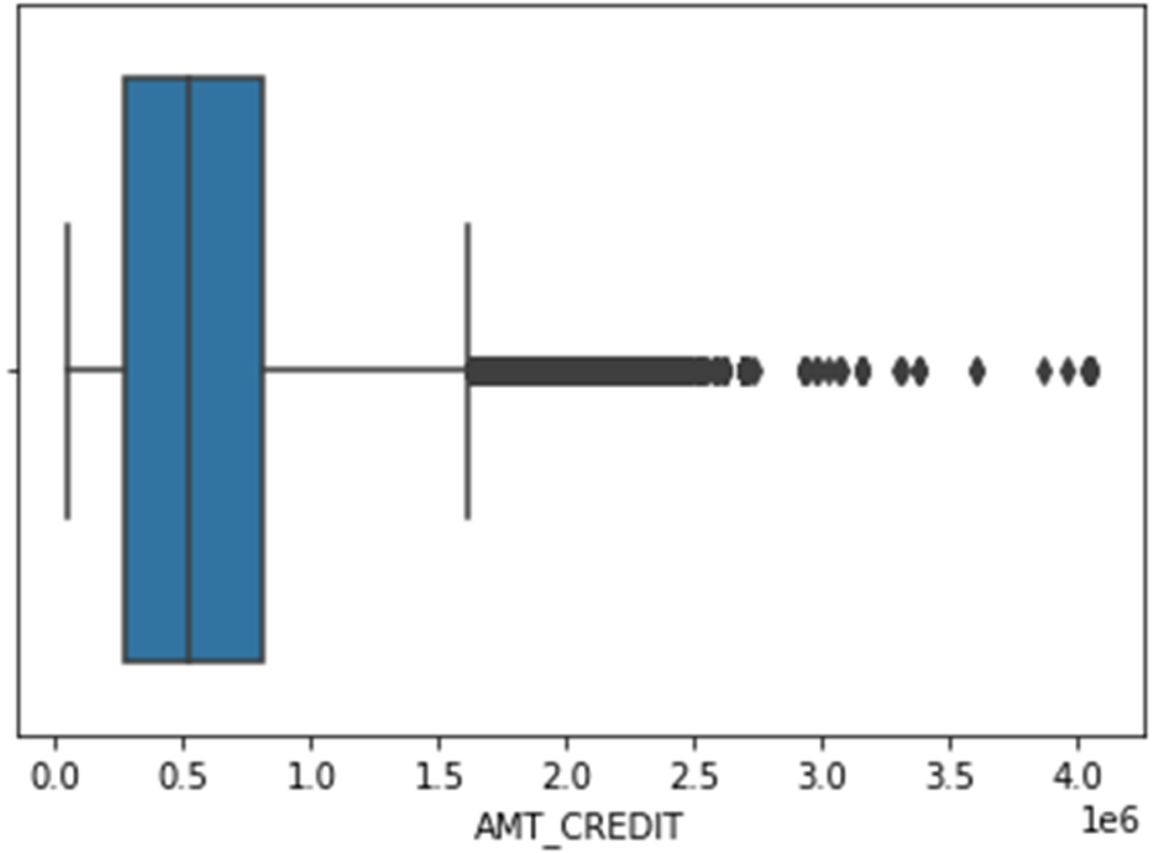


Client Type


NAME_CLIENT_TYPE
NAME_CONTRACT_STATUS
Approved
Refused
Canceled
Unused offer
count
600000
500000
400000
300000
200000
100000
0
Repeater
New
Refreshed
XNA
NAME_CLIENT_TYPE

Amt_Credit

The amt credit variable describes based on the missing values of the given data on case study



Conclusion

- For income type: ‘working’, ’commercial associate’, and ‘State Servant’ the number of Loans are higher than others.
- Less number of loans for income type ‘student’ ,’pensioner’, ‘Businessman’ and ‘Maternity leave’.
- On Gender: Females are having more number of Loans than male.
- For contract type ‘cash loans’ is having higher number of credits than ‘Revolving loans’ contract type.
- ‘Laborers’ have major number of loans in both defaulters and non defaulters

- Majority customers does not have a own car in both the cases.
- Majority customers owns a house or flat in both the cases.
- Customers of house/apartment category are the one who takes loan majorly.
- Most client take loan for secondary education and the default rate in secondary education is much high.
- Laborers and different categories of staffs mostly take the loan, but the managers and the high skilled tech staffs are reliable.

The End