



# CREDIT EDA ANALYSIS

ARJUN P

DS C57 JUNE2023



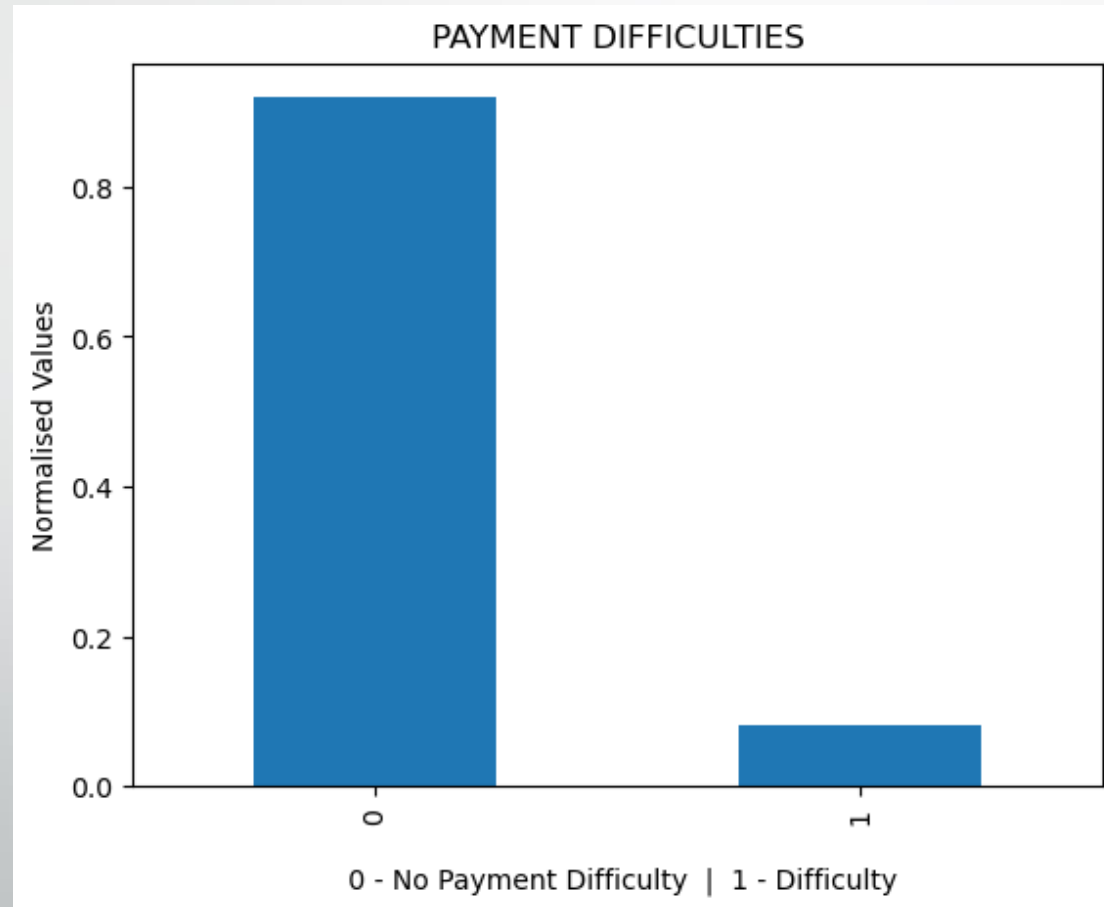
# APPLICATION DATA



# UNIVARIATE ANALYSIS

# CATEGORICAL NOMINAL

## *PAYMENT DIFFICULTIES v/s NORMALISED VALUES*

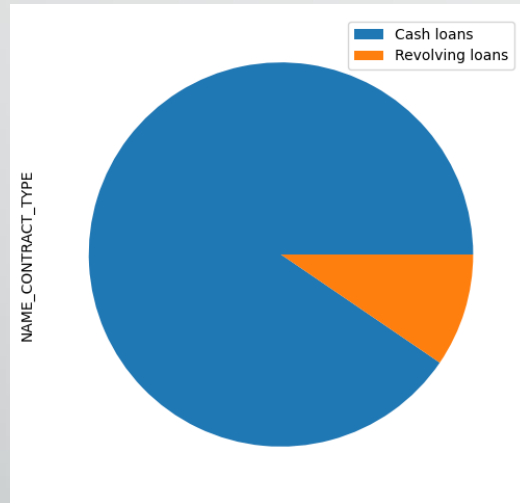




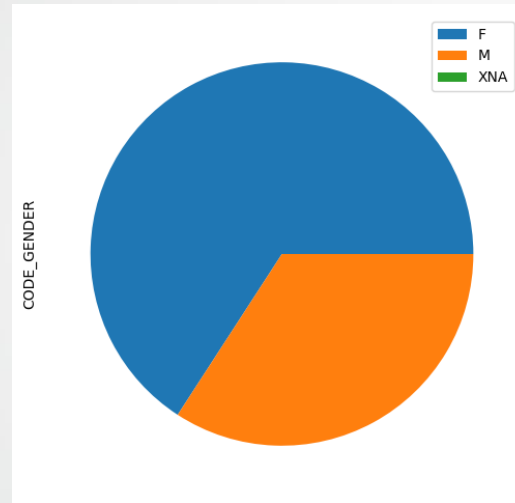
# INFERENCE

- MOST OF THE PEOPLE DO NOT FACING ANY PAYMENT DIFFICULTIES

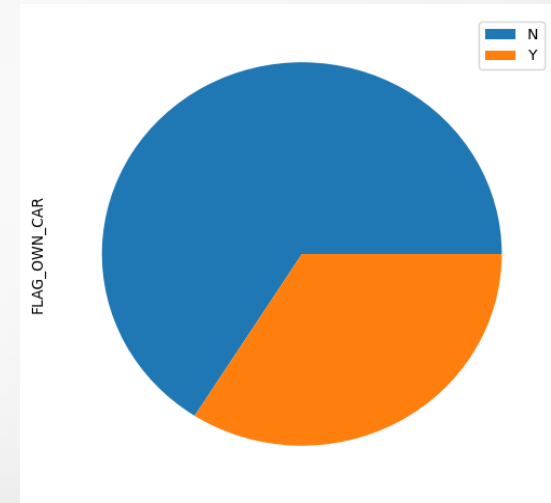
# object type variable and their values (Using pie chart)



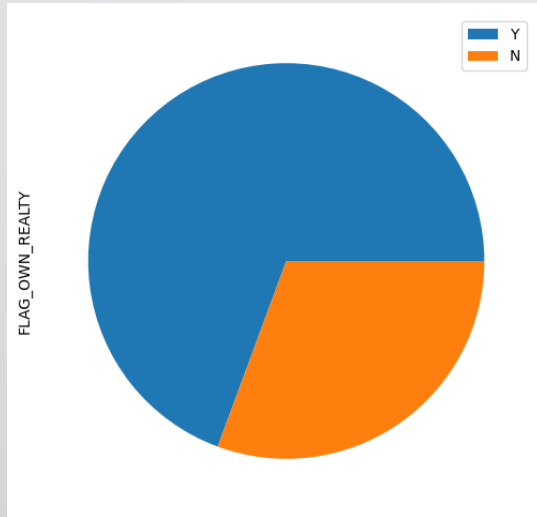
Cash loans 0.904787  
Revolving loans 0.095213



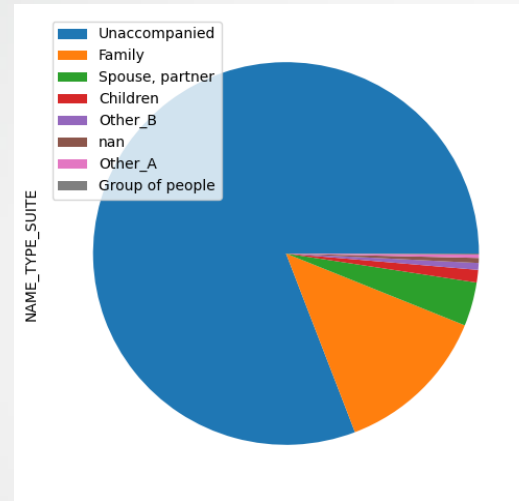
F 0.658344  
M 0.341643



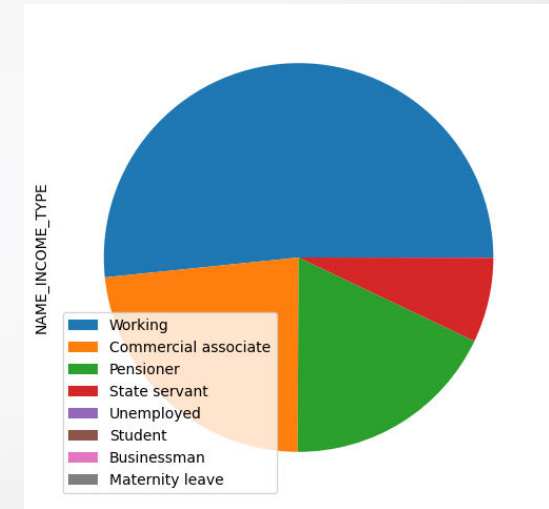
N 0.659892  
Y 0.340108



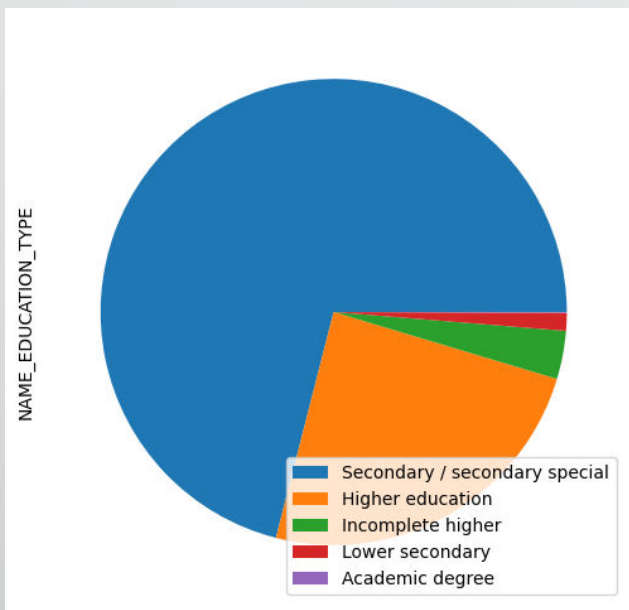
Y 0.693673  
N 0.306327



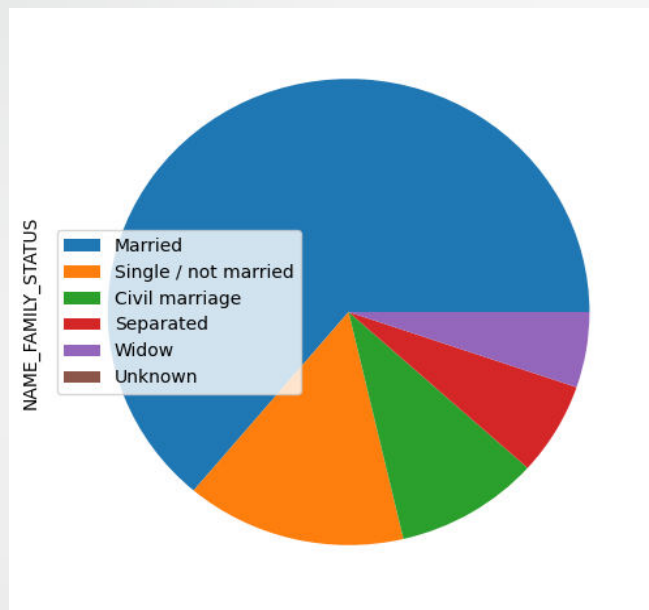
Unaccompanied 0.808186  
Family 0.130561  
Spouse, partner 0.036974  
Children 0.010624  
Other\_B 0.005756  
NaN 0.004201  
Other\_A 0.002816  
Group of people 0.000881



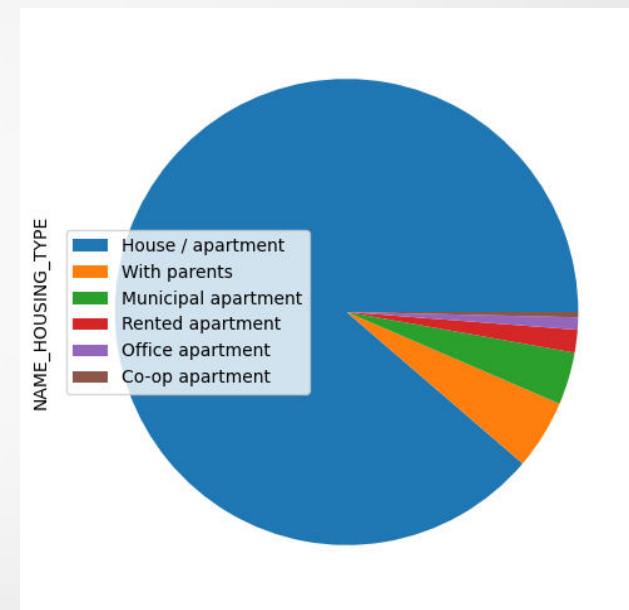
Working 0.516320  
Commercial associate 0.232892  
Pensioner 0.180033  
State servant 0.070576  
Unemployed 0.000072  
Student 0.000059  
Businessman 0.000033  
Maternity leave 0.000016



Secondary 0.710189  
 Higher education 0.243448  
 Incomplete higher 0.033420  
 Lower secondary 0.012409  
 Academic degree 0.000533

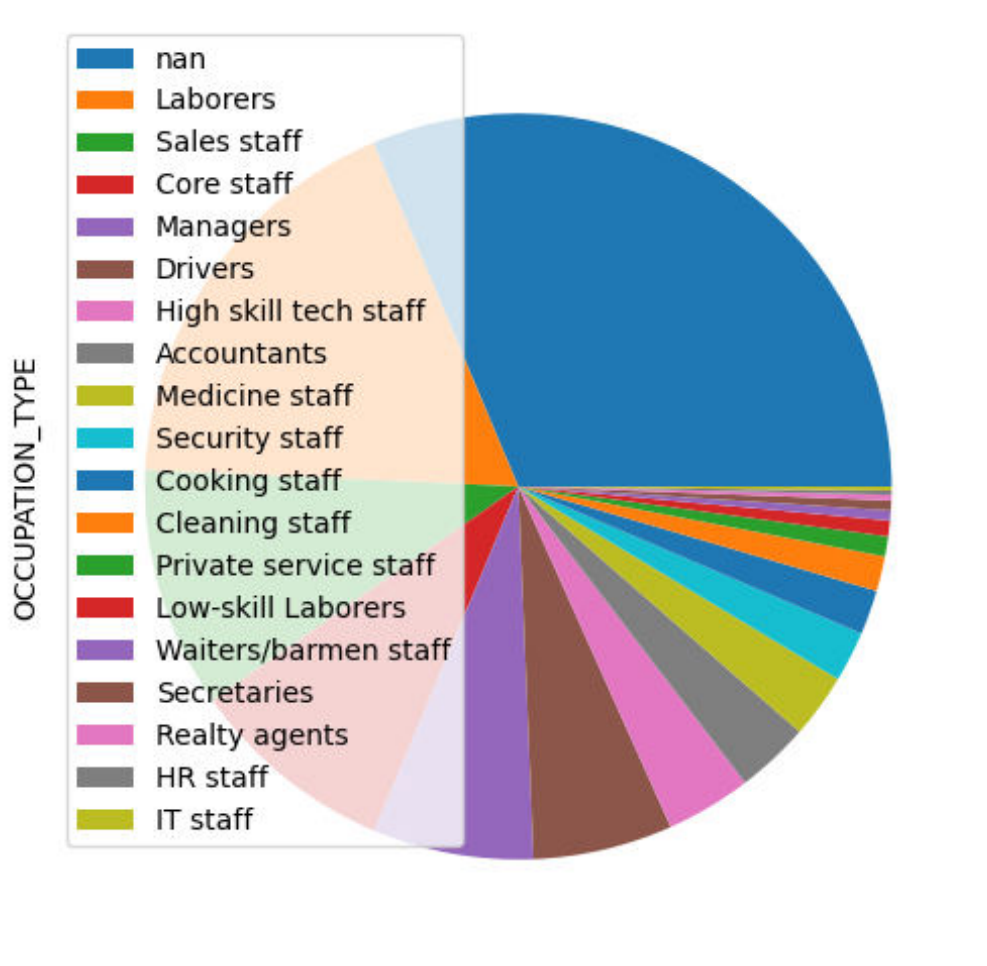


Married 0.638780  
 Single / not married 0.147780  
 Civil marriage 0.096826  
 Separated 0.064290  
 Widow 0.052317  
 Unknown 0.000007



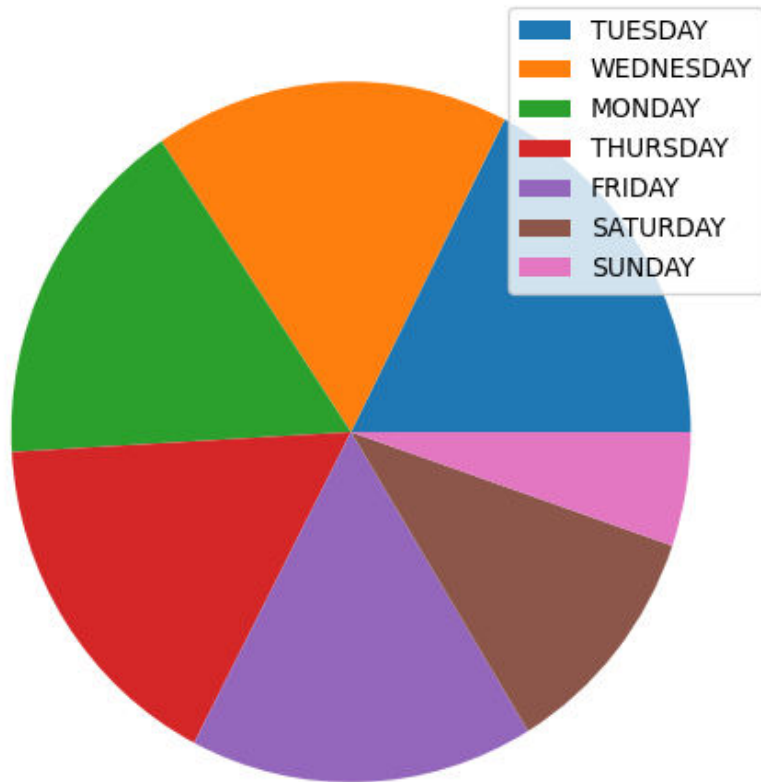
House / apartment 0.887344  
 With parents 0.048258  
 Municipal apartment 0.036366  
 Rented apartment 0.015873  
 Office apartment 0.008510  
 Co-op apartment 0.003649



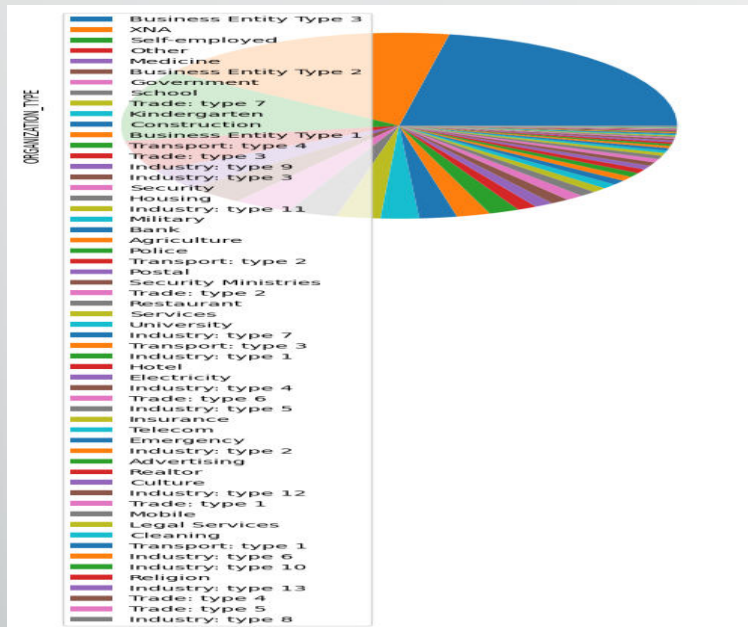


NaN 0.313455  
Laborers 0.179460  
Sales staff 0.104393  
Core staff 0.089655  
Managers 0.069497  
Drivers 0.060495  
High skill tech staff 0.037007  
Accountants 0.031911  
Medicine staff 0.027762  
Security staff 0.021856  
Cooking staff 0.019336  
Cleaning staff 0.015131  
Private service staff 0.008624  
Low-skill Laborers 0.006806  
Waiters/barmen staff 0.004384  
Secretaries 0.004244  
Realty agents 0.002442  
HR staff 0.001831  
IT staff 0.001711

WEEKDAY\_APPR\_PROCESS\_START



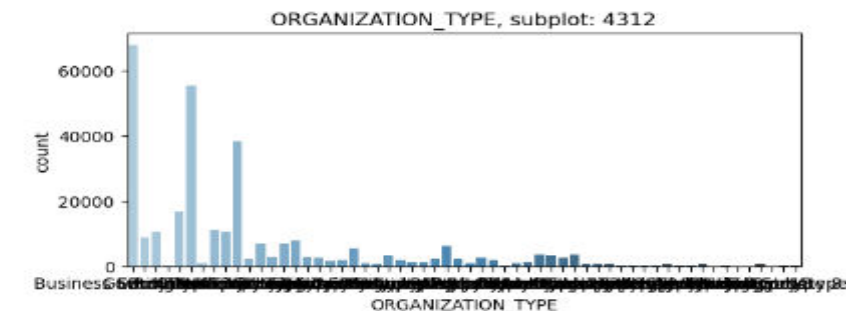
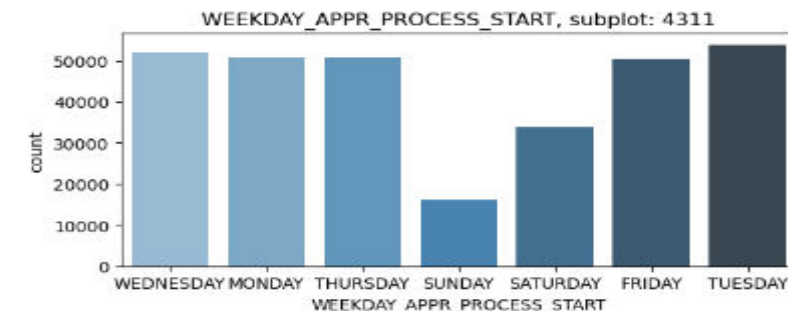
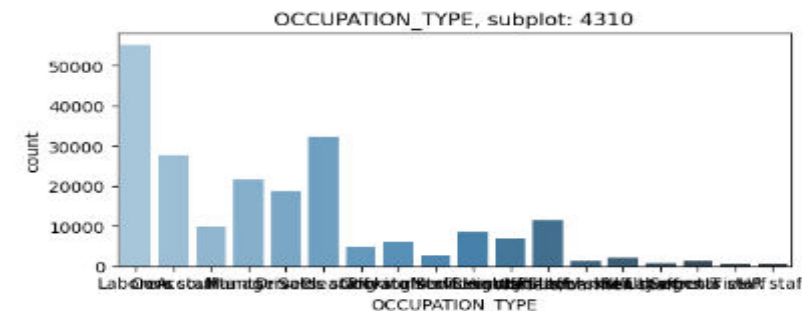
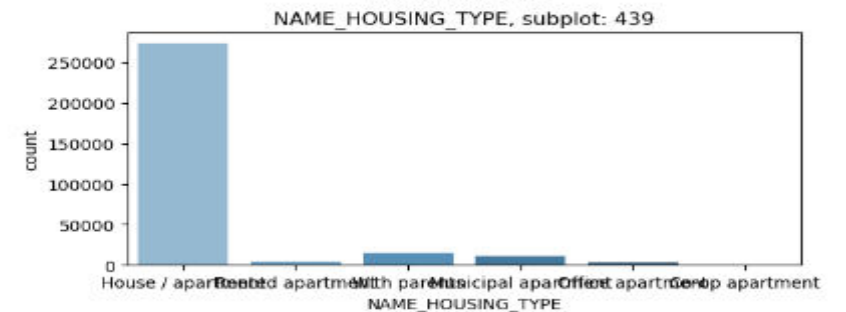
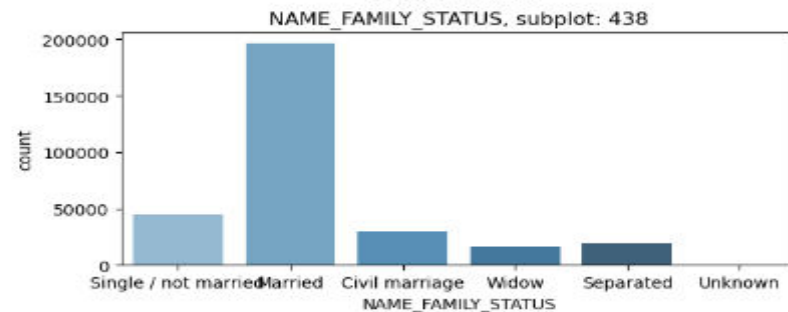
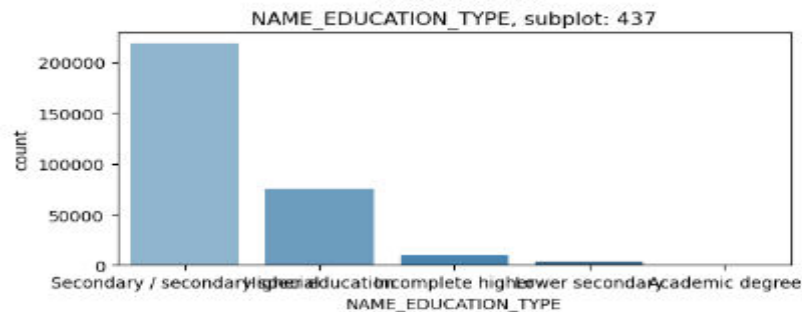
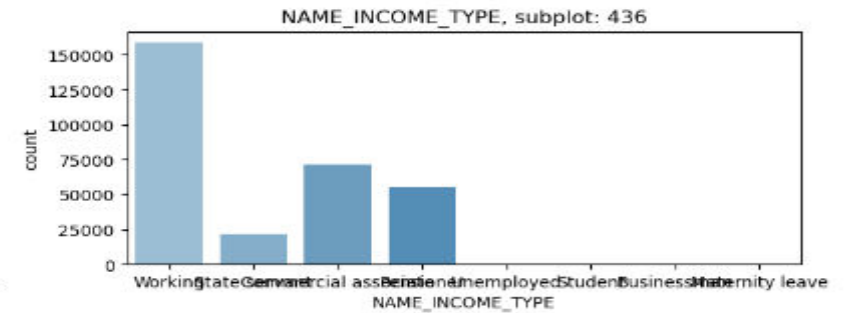
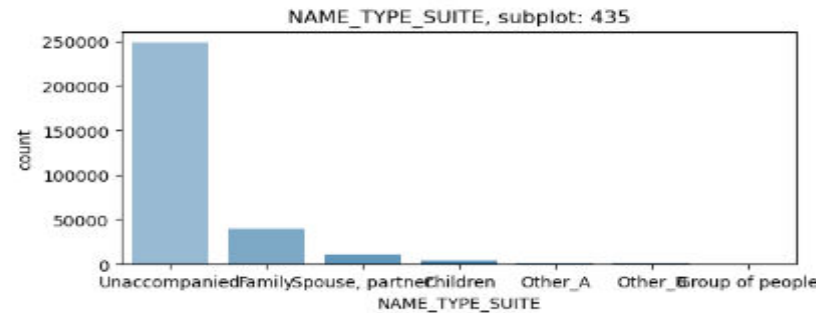
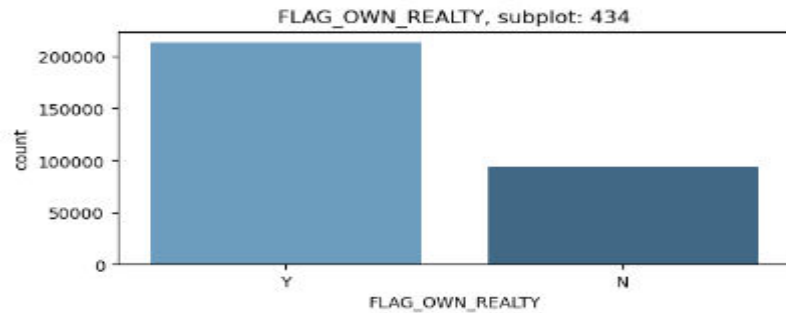
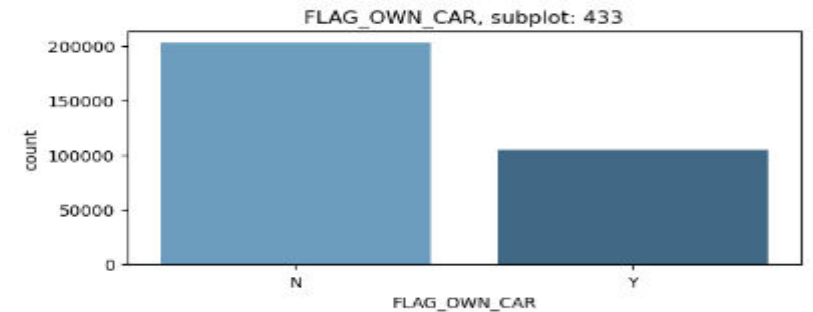
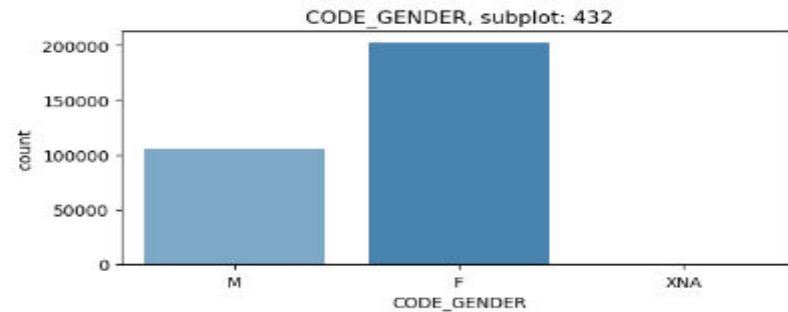
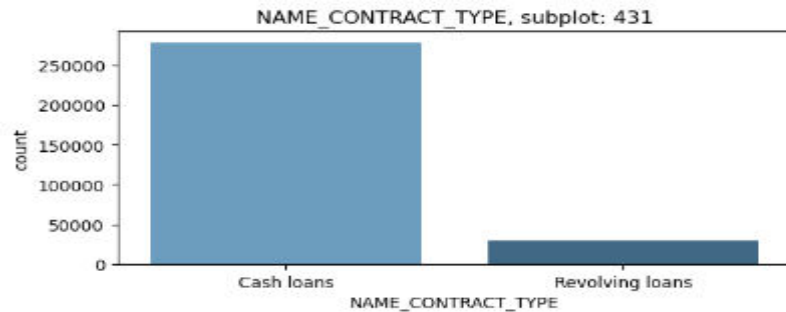
TUESDAY 0.175282  
WEDNESDAY 0.168885  
MONDAY 0.164918  
THURSDAY 0.164518  
FRIDAY 0.163695  
SATURDAY 0.110084  
SUNDAY 0.052619



Industry: type 4 0.002852 Trade: type 6 0.002052 Industry: type 5  
 0.001948 Insurance 0.001941 Telecom 0.001876 Emergency 0.001821  
 Industry: type 2 0.001489 Advertising 0.001395 Realtor 0.001288 Culture  
 0.001232 Industry: type 12 0.001200 Trade: type 1 0.001132 Mobile 0.001031  
 Legal Services 0.000992 Cleaning 0.000845 Transport: type 1 0.000654  
 Industry: type 6 0.000364 Industry: type 10 0.000354 Religion 0.000276  
 Industry: type 13 0.000218 Trade: type 4 0.000208 Trade: type 5 0.000159  
 Industry: type 8 0.000078

Business Entity Type 3 0.221104  
 XNA 0.180072 Self-employed 0.124913 Other 0.054252  
 Medicine 0.036399 Business Entity Type 2 0.034317  
 Government 0.033833 School 0.028919  
 Trade: type 7 0.025466 Kindergarten 0.022373  
 Construction 0.021856 Business Entity Type 1 0.019459  
 Transport: type 4 0.017554 Trade: type 3 0.011356  
 Industry: type 9 0.010952 Industry: type 3 0.010660  
 Security 0.010559 Housing 0.009619  
 Industry: type 11 0.008793 Military 0.008566 Bank 0.008153  
 Agriculture 0.007980 Police 0.007613  
 Transport: type 2 0.007167 Postal 0.007014  
 Security Ministries 0.006419 Trade: type 2 0.006179  
 Restaurant 0.005889 Services 0.005122  
 University 0.004315 Industry: type 7 0.004250  
 Transport: type 3 0.003860 Industry: type 1 0.003379  
 Hotel 0.003141 Electricity 0.003089

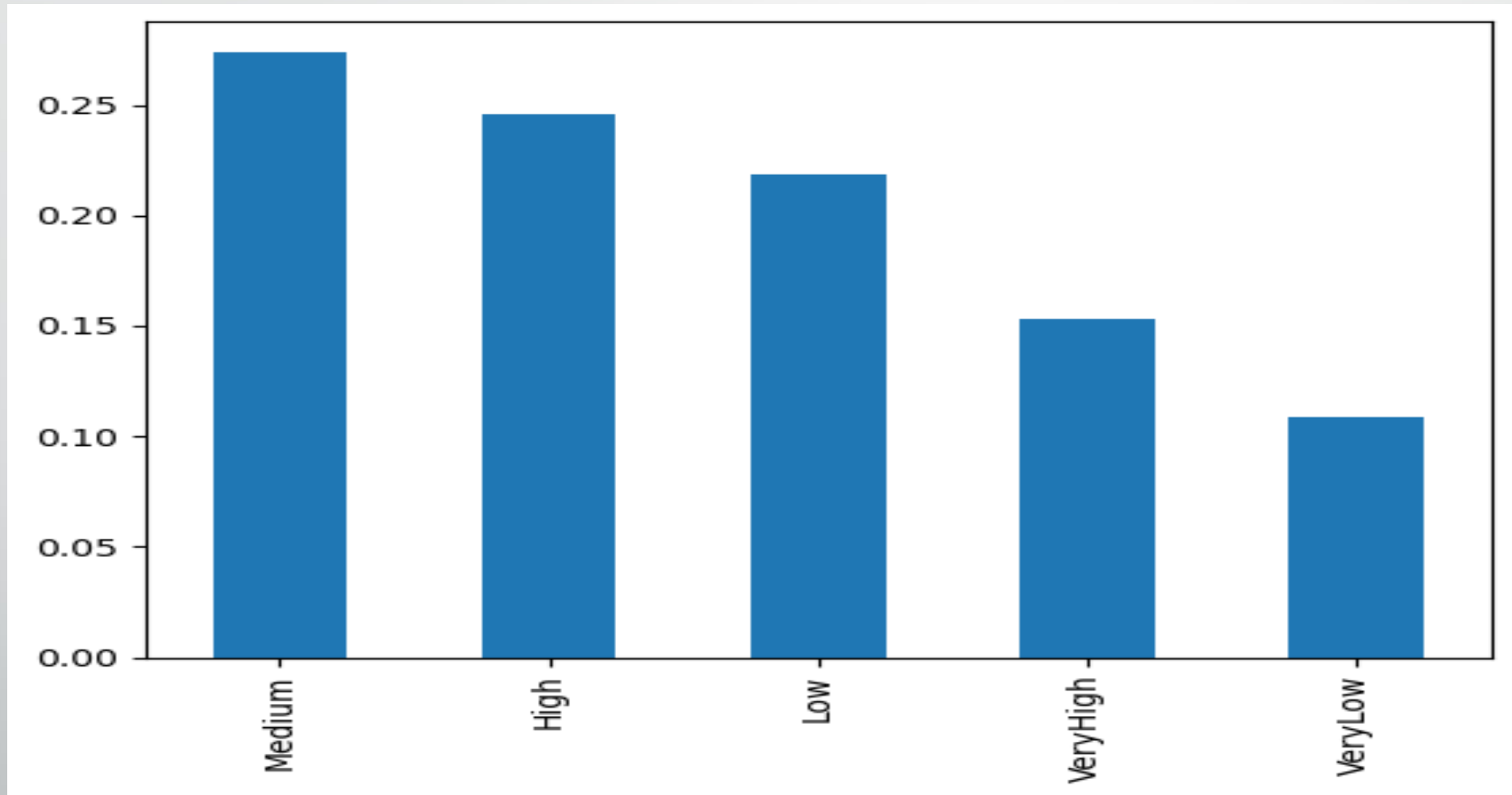
# object type variable and their values (Using subplot)



# INFERENCE

- Cash loans offered are more than revolving loans, at 90%
- 65% Females have taken loans in comparison to 34% male.
- 65% applicant don't own cars
- 69% applicants own living quarters
- 81% applicants came accompanied for loan application
- While most applicants are working class, 18% are pensioners
- 71% have secondary education
- 63% are married
- 31% have not mentioned their occupation type

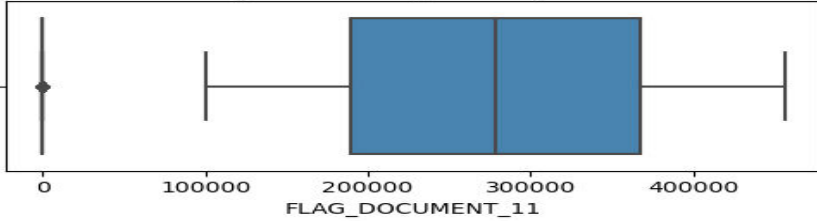
# Numerical Variable



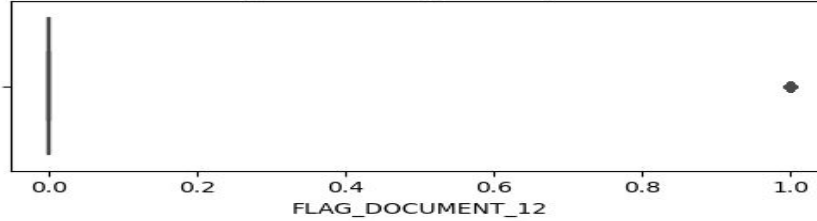
Medium Income group is the largest Group applying for loans.

# Data Correction and Standardization (INT Data-Analysis by boxplots)

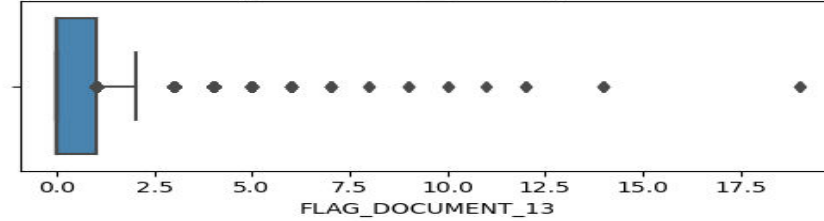
FLAG\_DOCUMENT\_11, subplot: 531



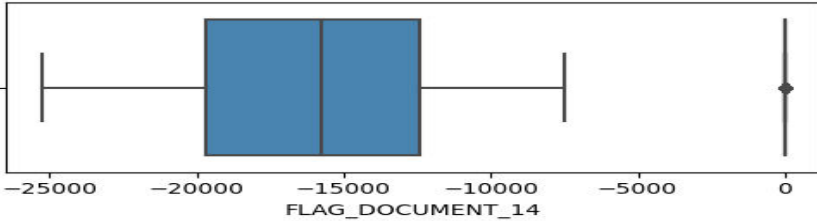
FLAG\_DOCUMENT\_12, subplot: 532



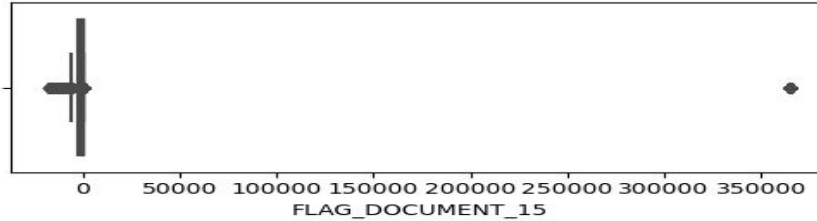
FLAG\_DOCUMENT\_13, subplot: 533



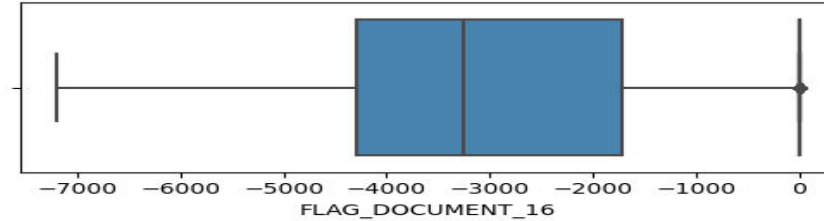
FLAG\_DOCUMENT\_14, subplot: 534



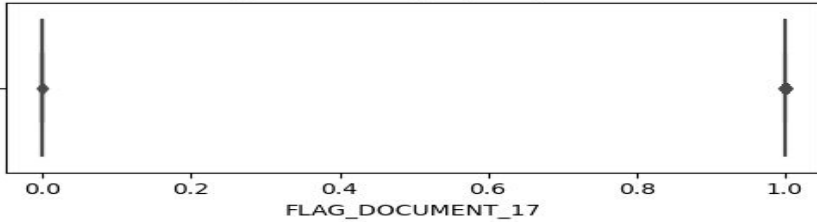
FLAG\_DOCUMENT\_15, subplot: 535



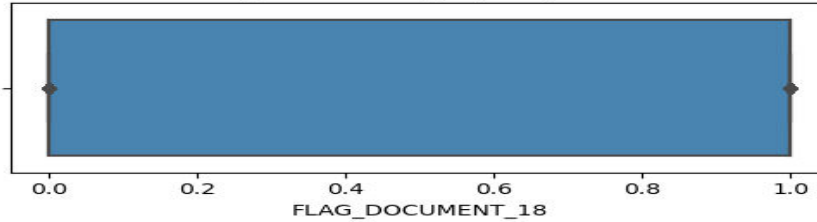
FLAG\_DOCUMENT\_16, subplot: 536



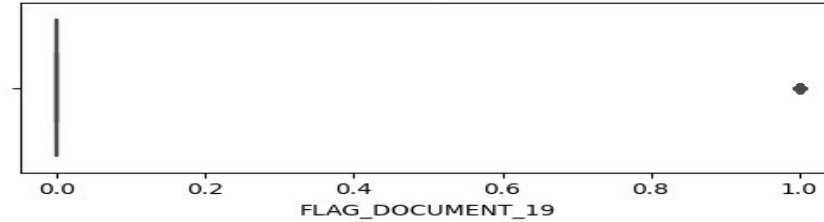
FLAG\_DOCUMENT\_17, subplot: 537



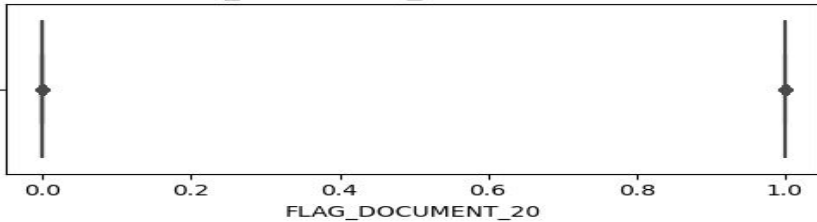
FLAG\_DOCUMENT\_18, subplot: 538



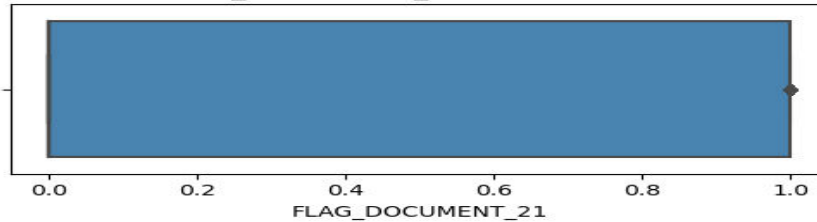
FLAG\_DOCUMENT\_19, subplot: 539



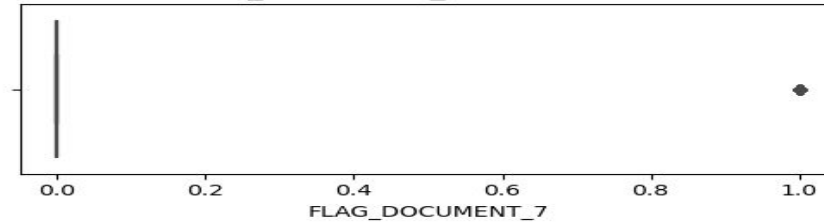
FLAG\_DOCUMENT\_20, subplot: 5310



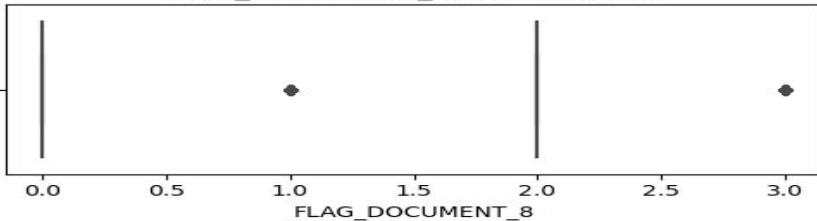
FLAG\_DOCUMENT\_21, subplot: 5311



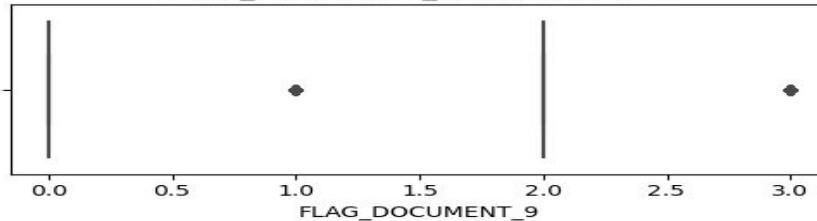
FLAG\_DOCUMENT\_7, subplot: 5312



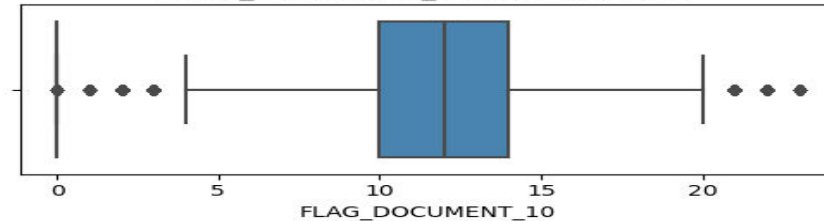
FLAG\_DOCUMENT\_8, subplot: 5313



FLAG\_DOCUMENT\_9, subplot: 5314

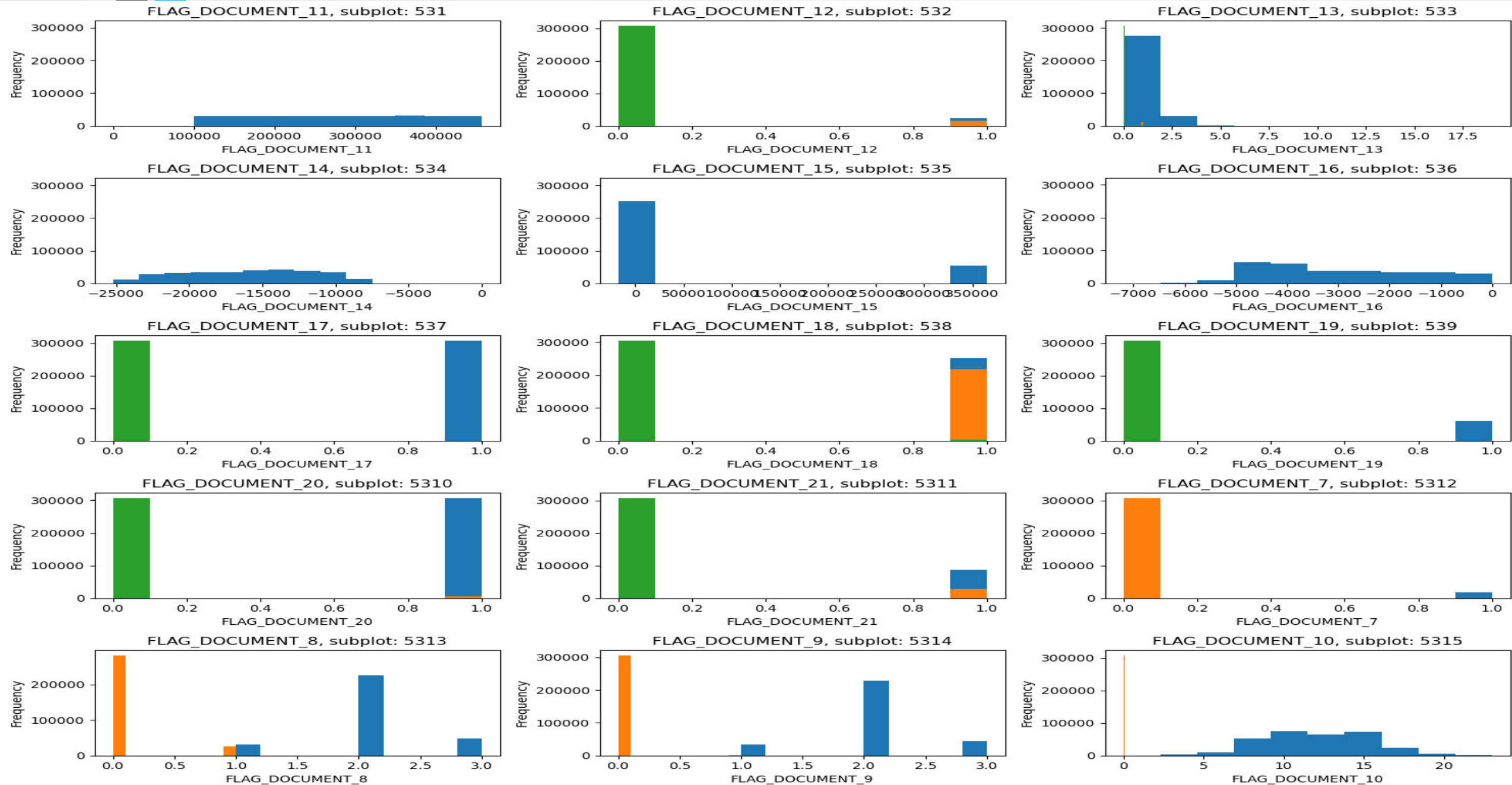


FLAG\_DOCUMENT\_10, subplot: 5315



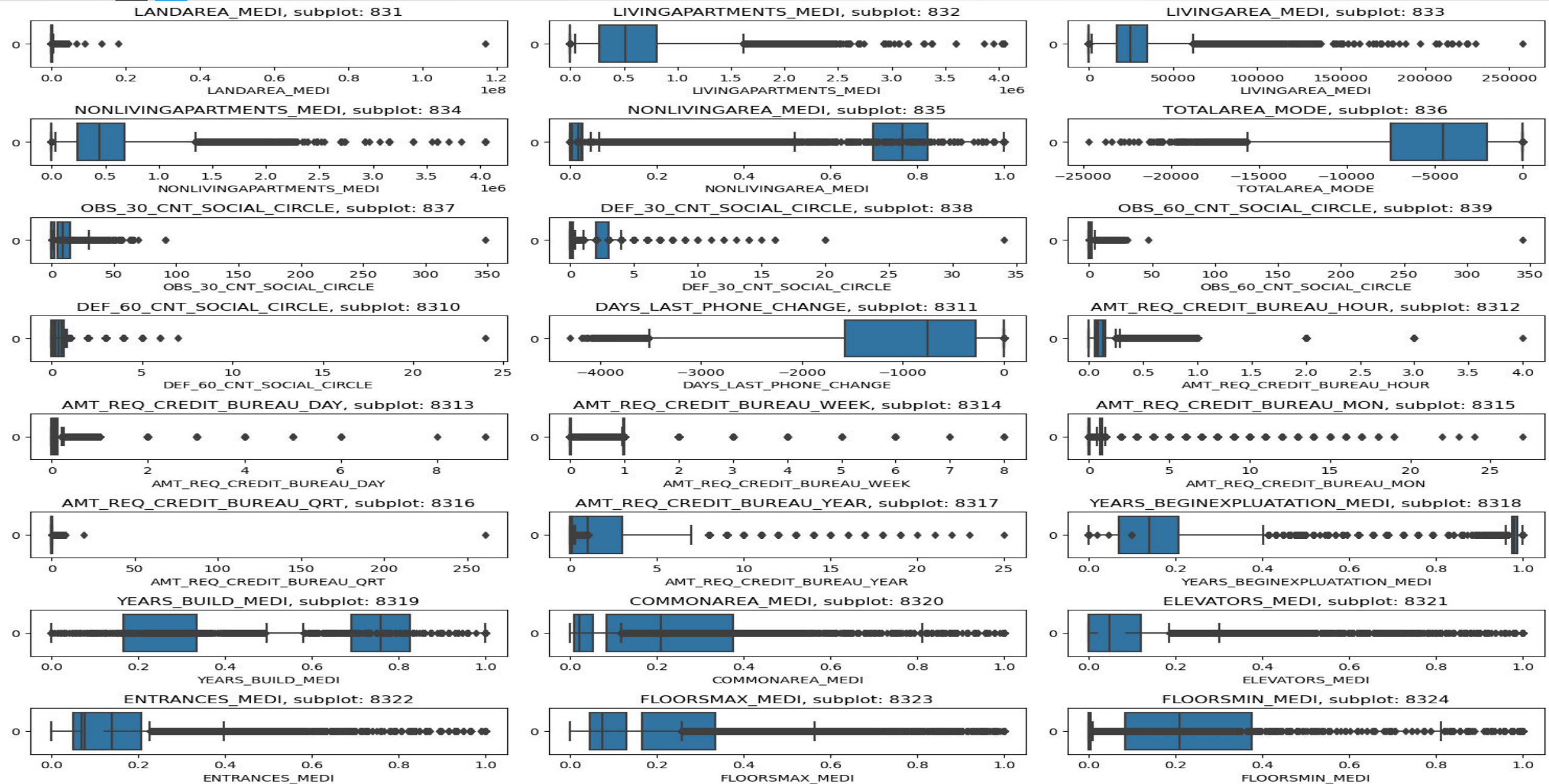


# INT Data-Histogram-Analysis by histogram

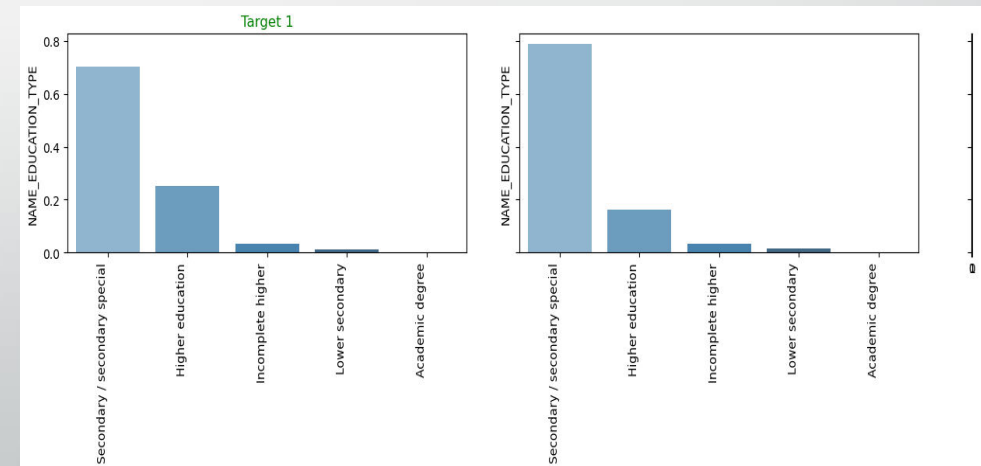
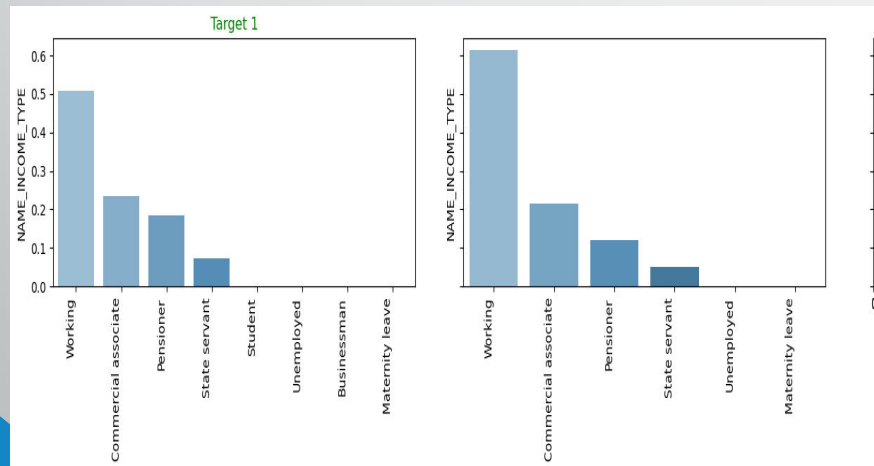
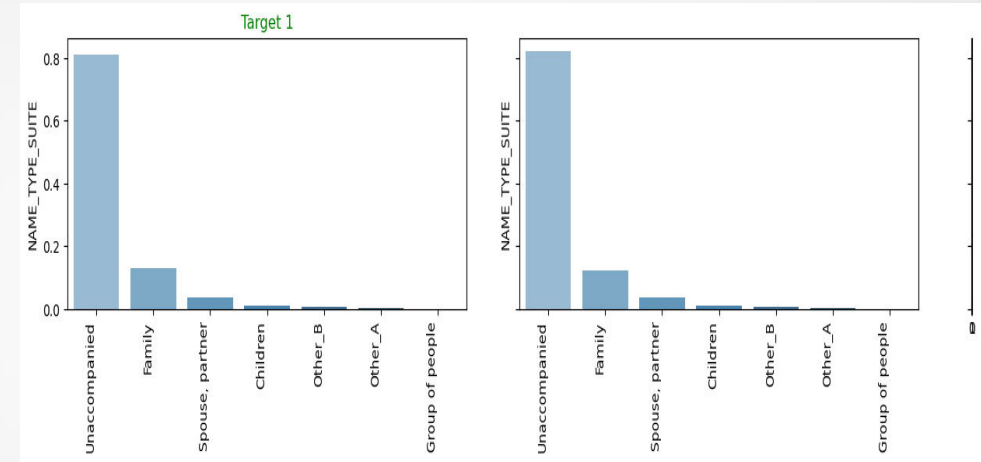
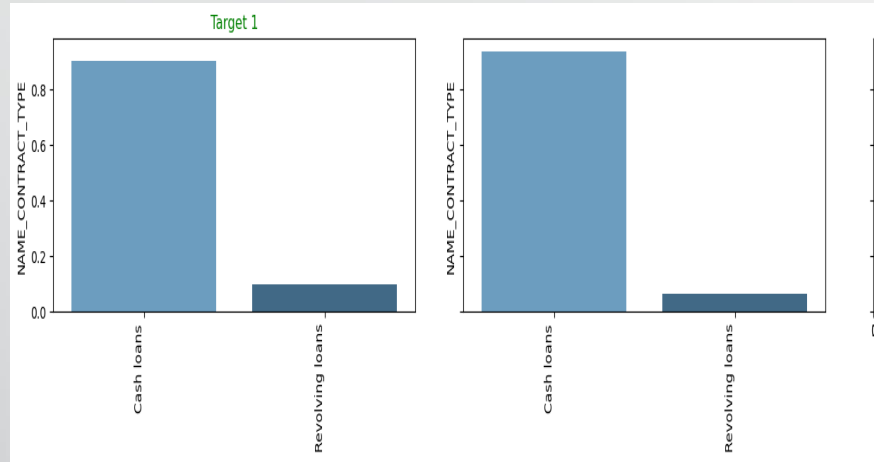


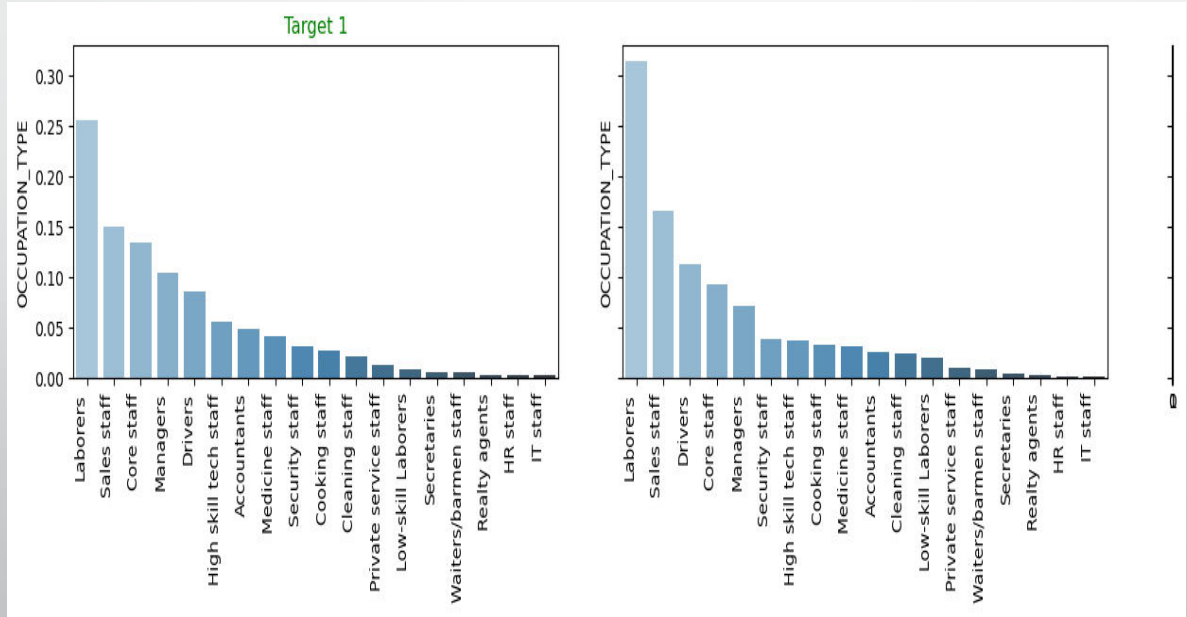
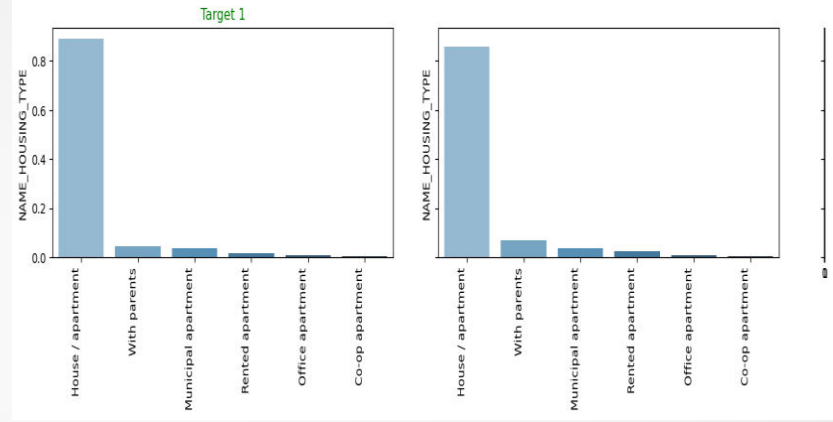
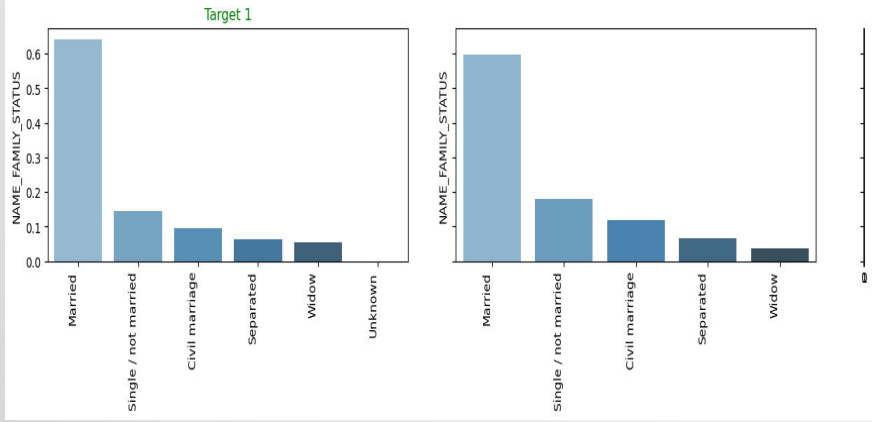


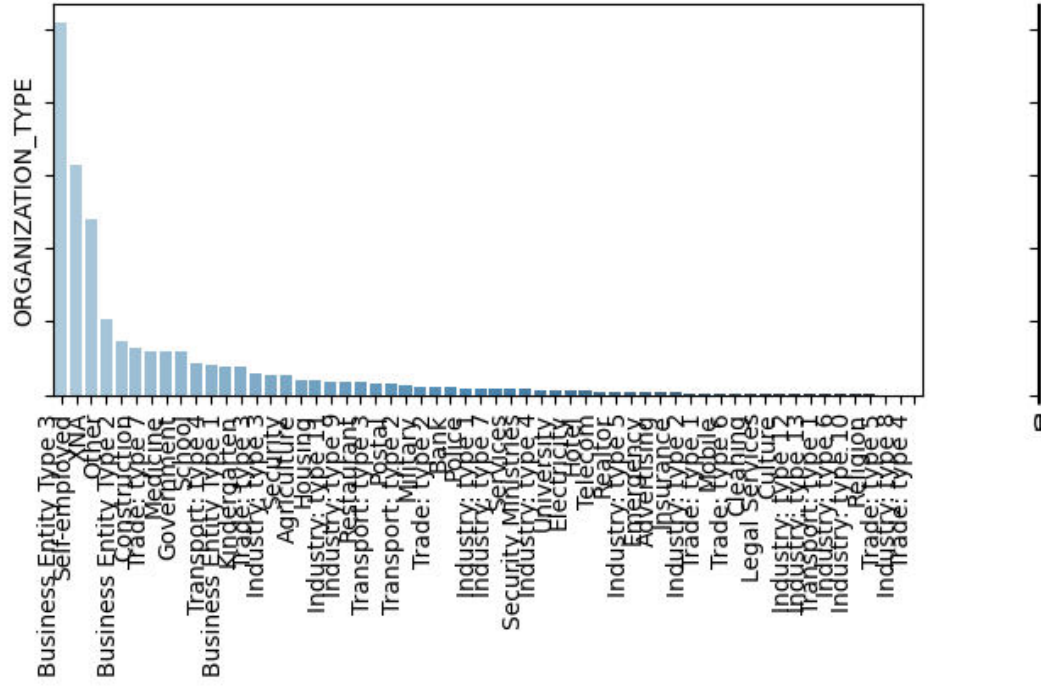
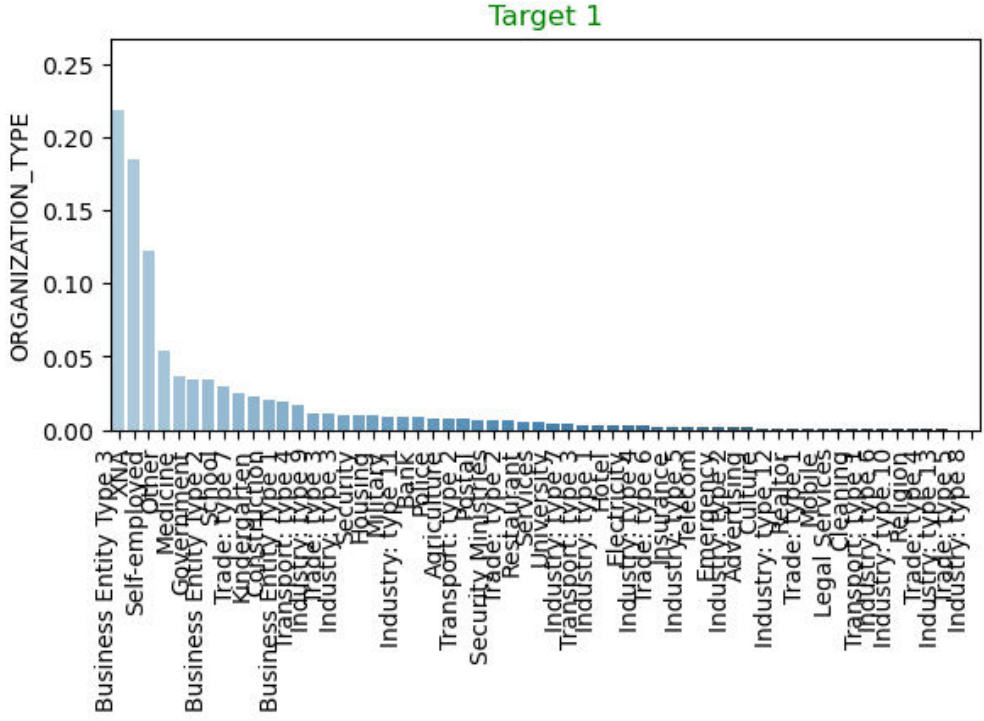
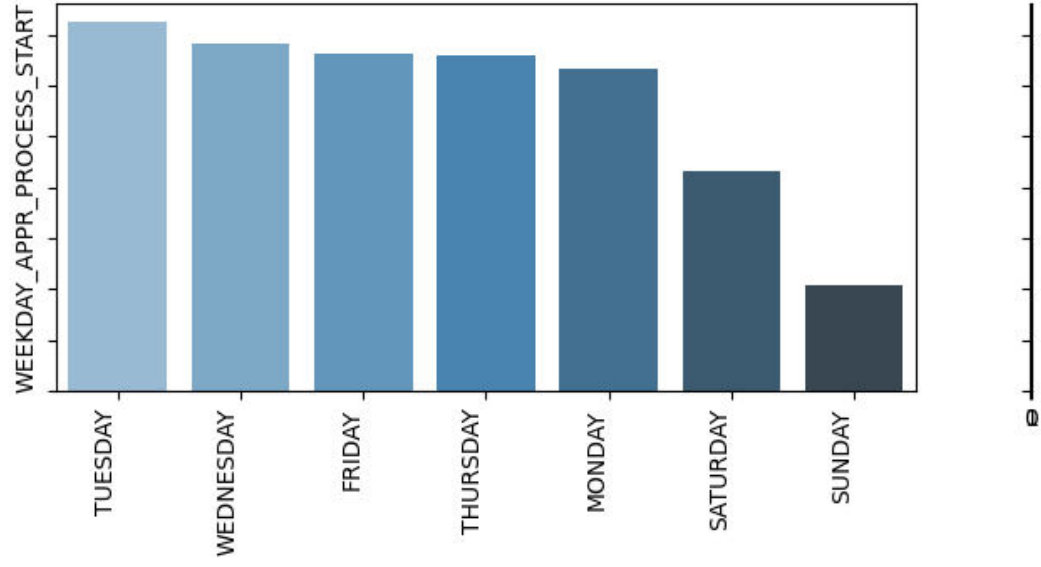
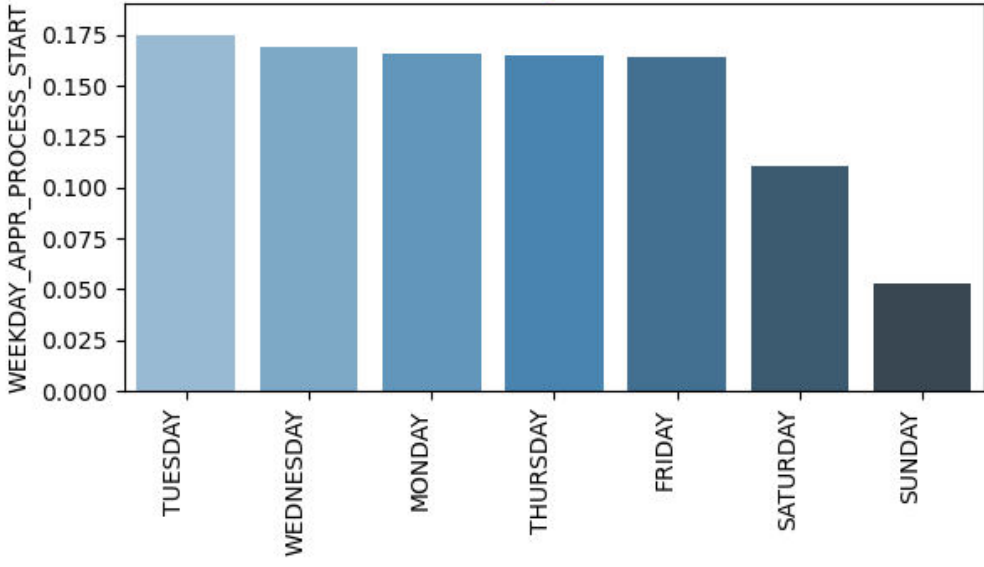
# float data

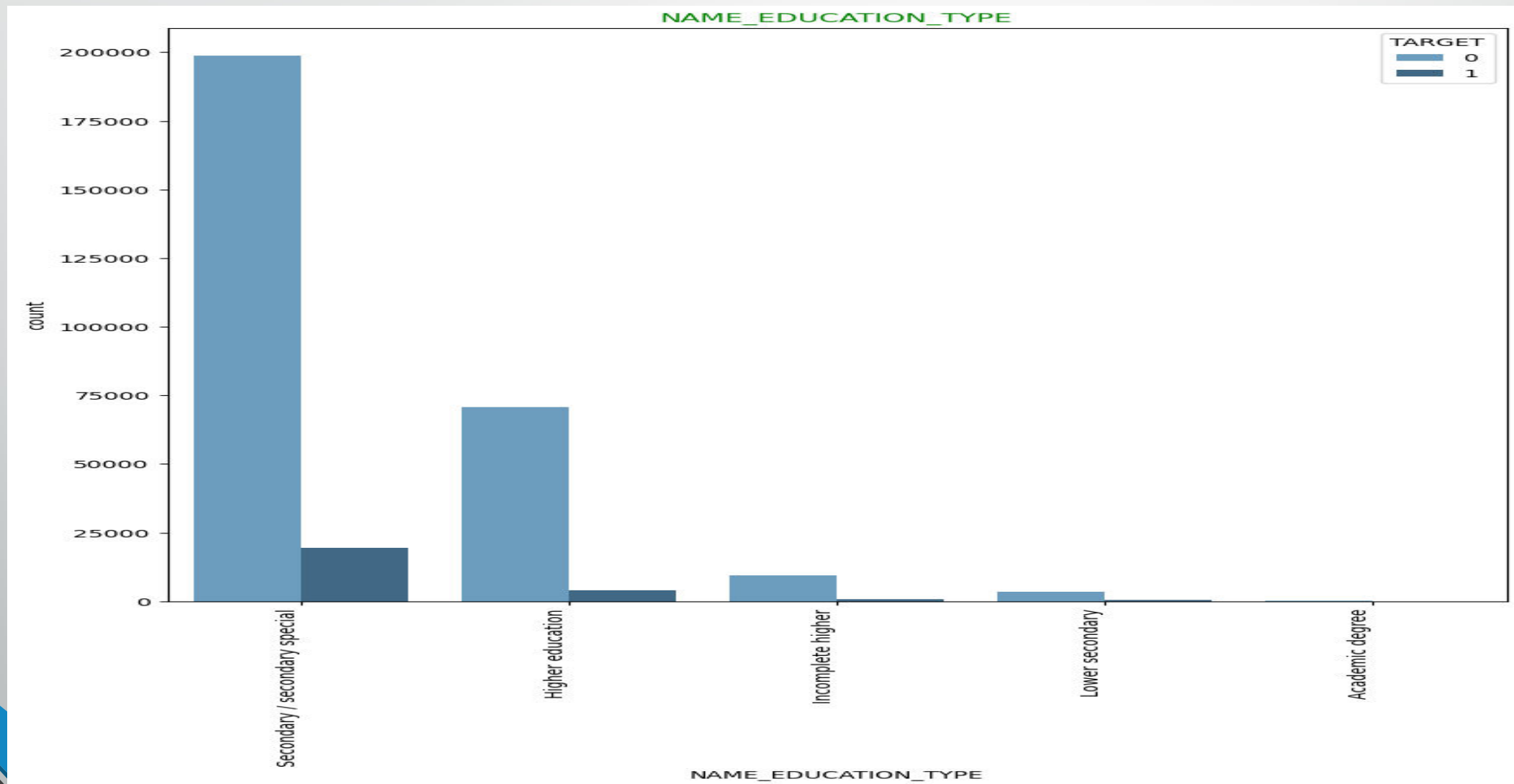


# Univariate analysis on categorical nominal to analyse both data frame











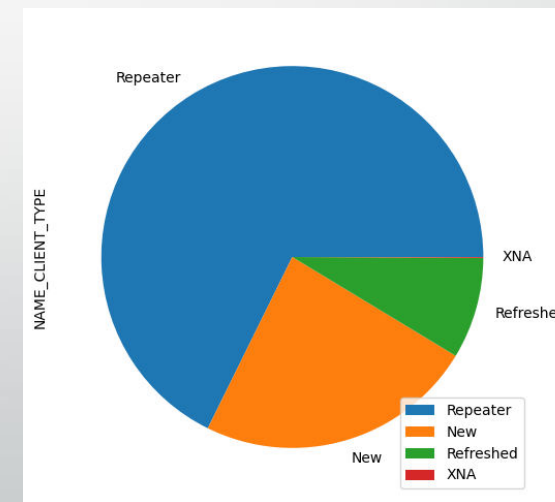
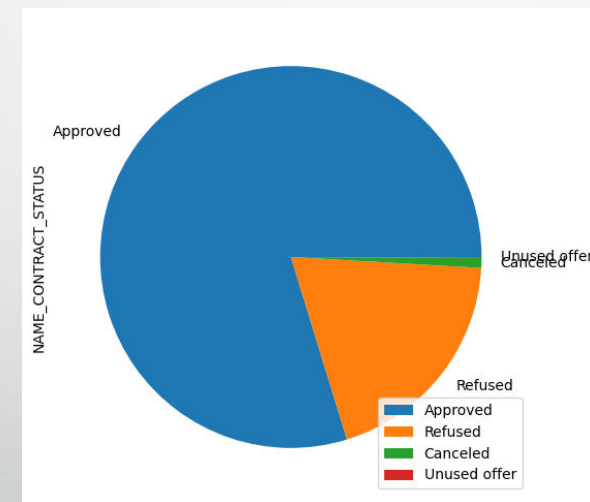
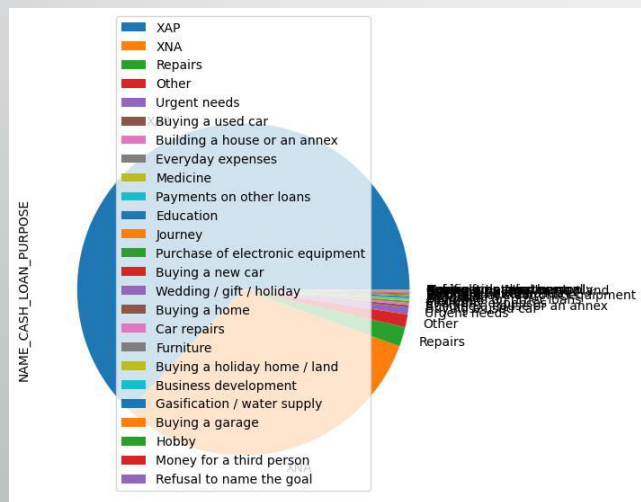
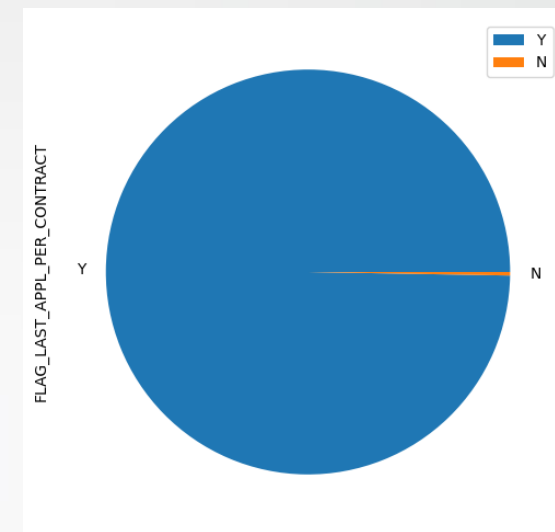
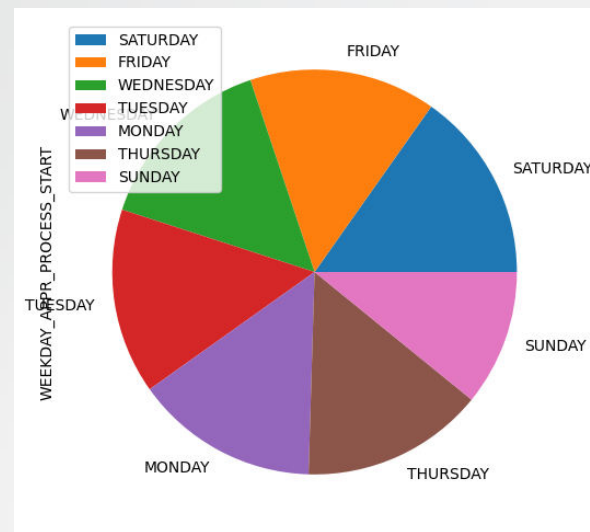
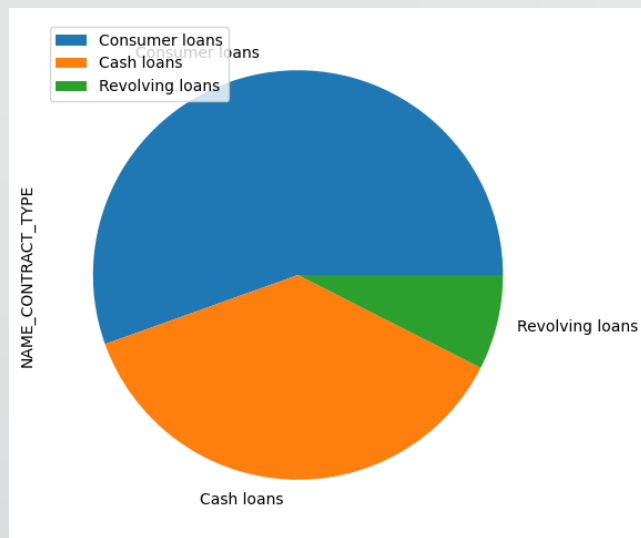
PREVIOUS DATA



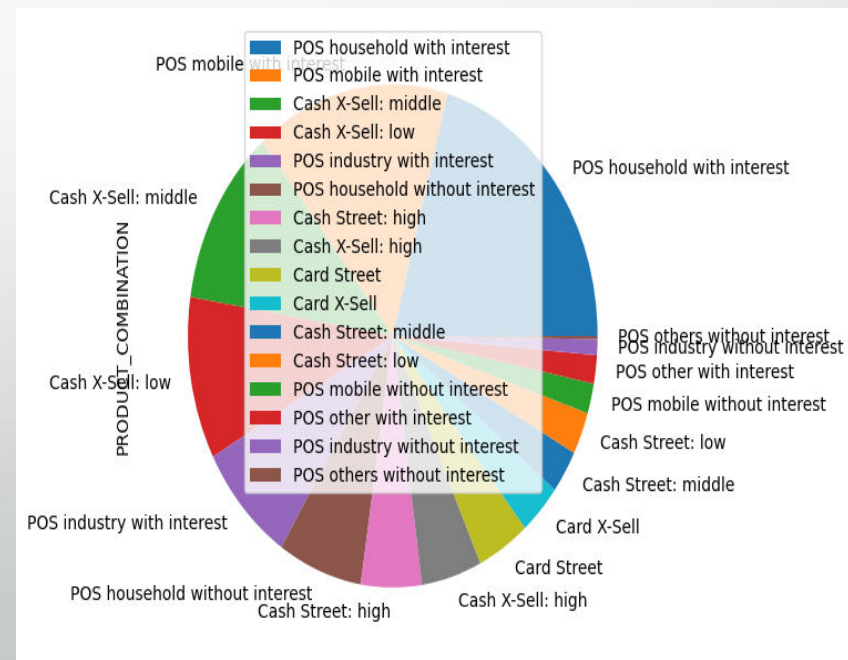
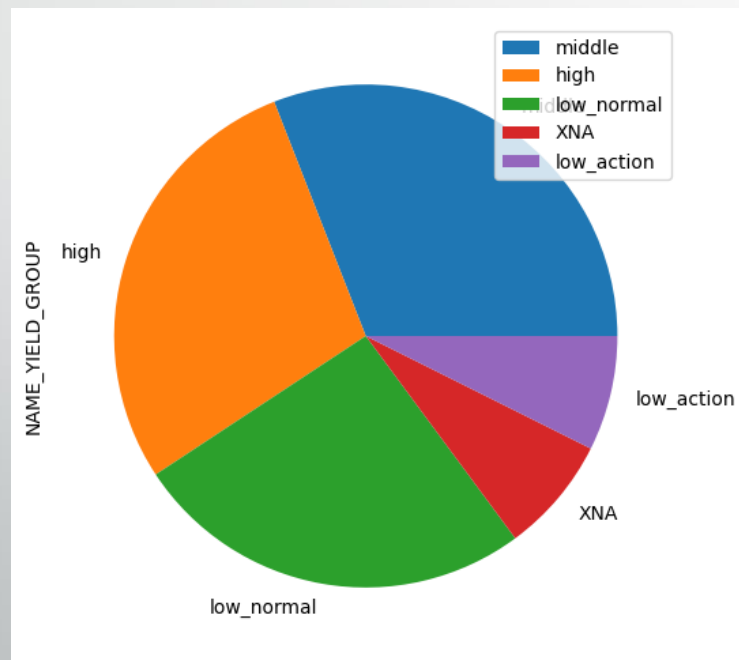
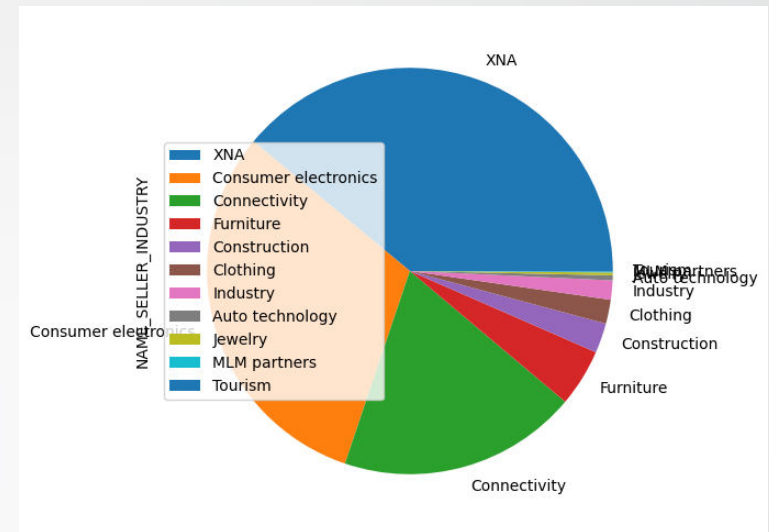
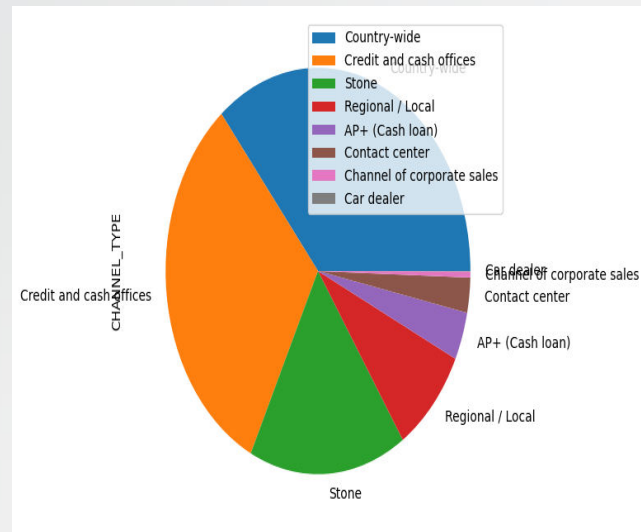
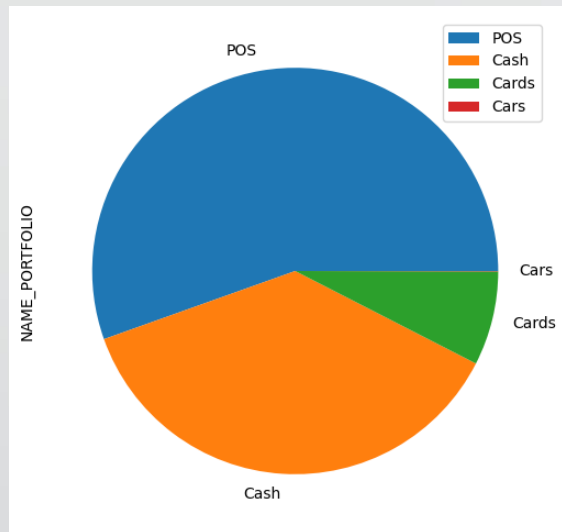
# UNIVARIATE ANALYSIS

**Categorical Nominal**



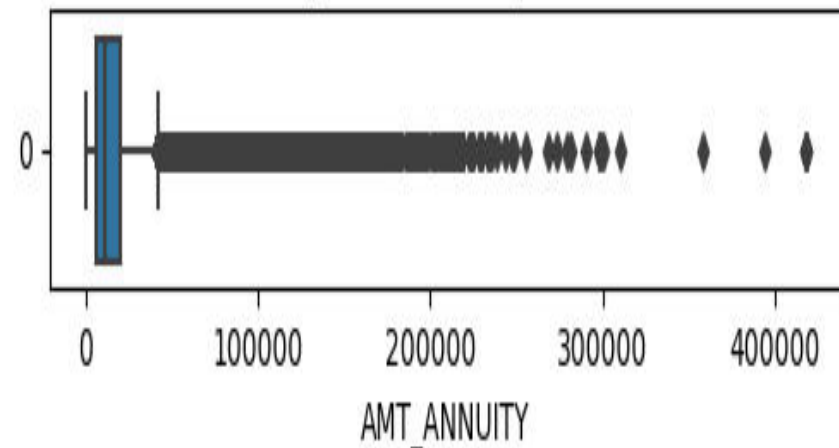




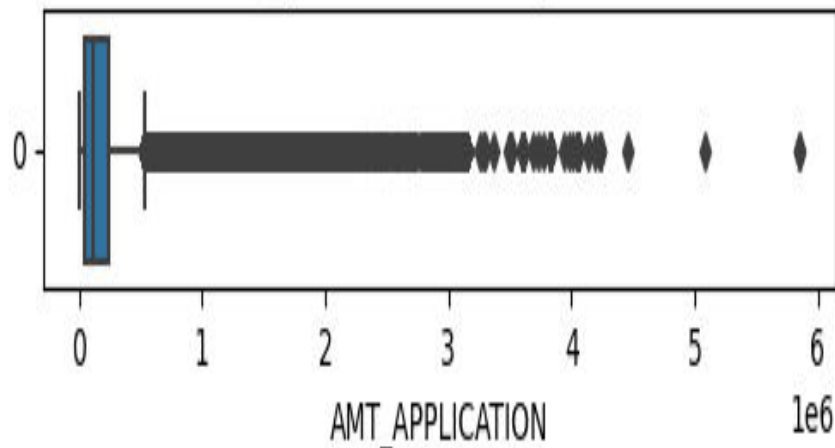


# float data

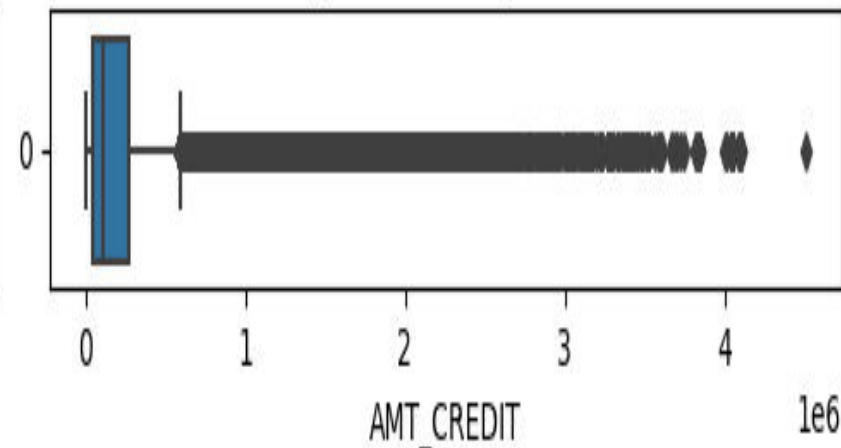
AMT\_ANNUIITY, subplot: 831



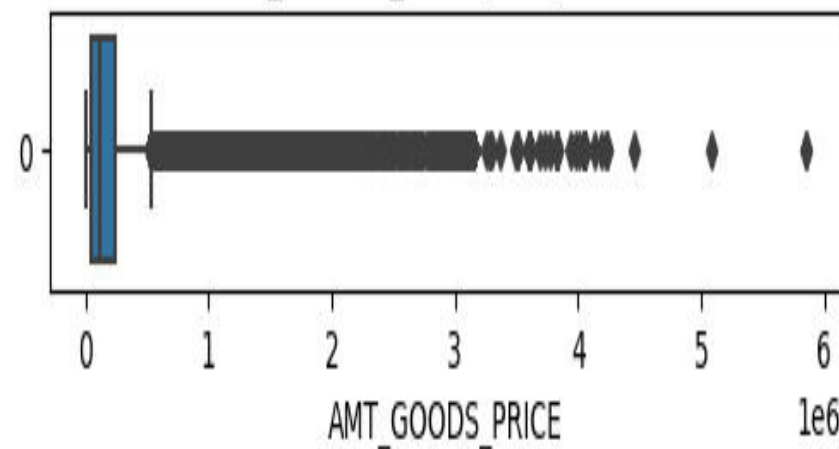
AMT\_APPLICATION, subplot: 832



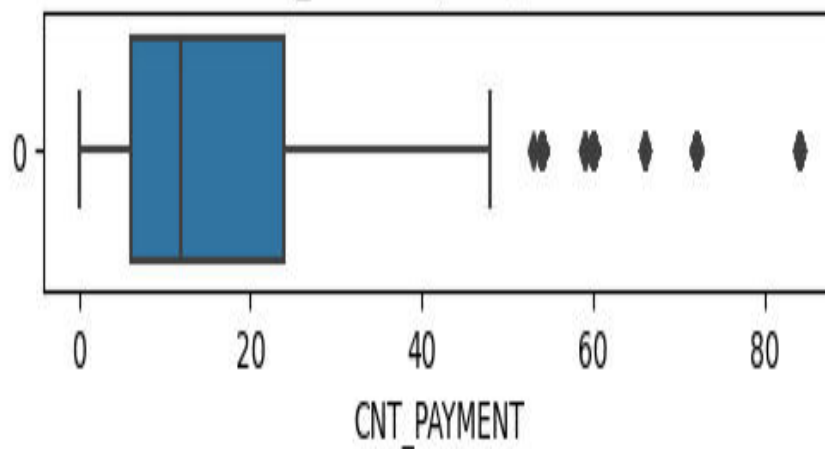
AMT\_CREDIT, subplot: 833



AMT\_GOODS\_PRICE, subplot: 834

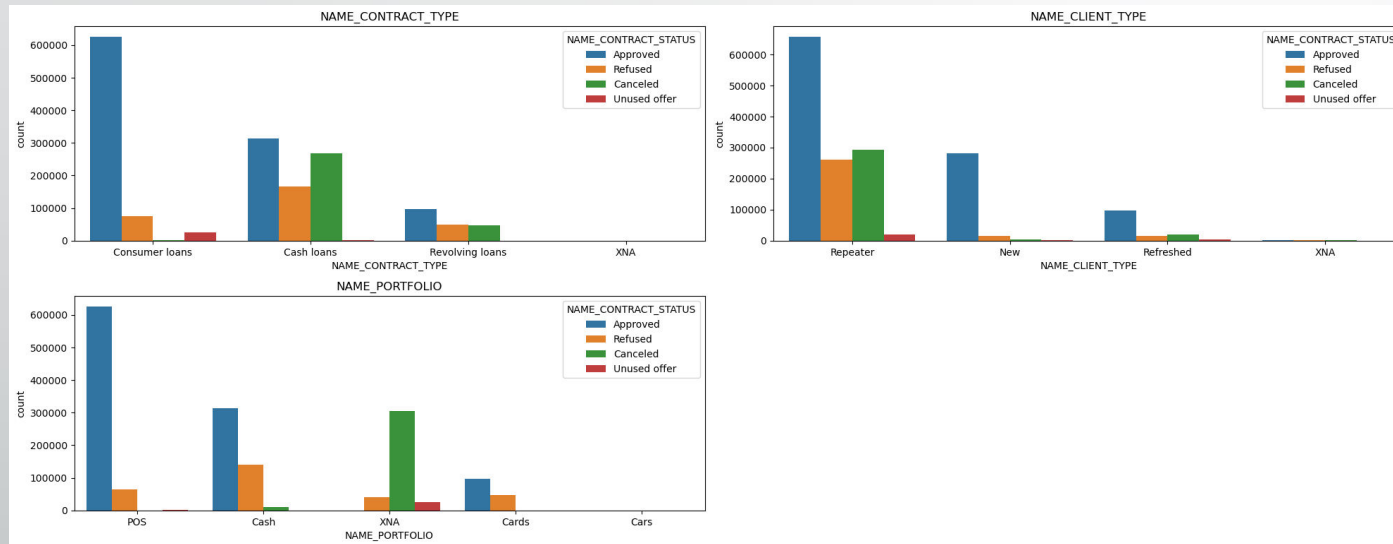


CNT\_PAYMENT, subplot: 835



# Bivariate Analysis

- Categorical & categorical variable

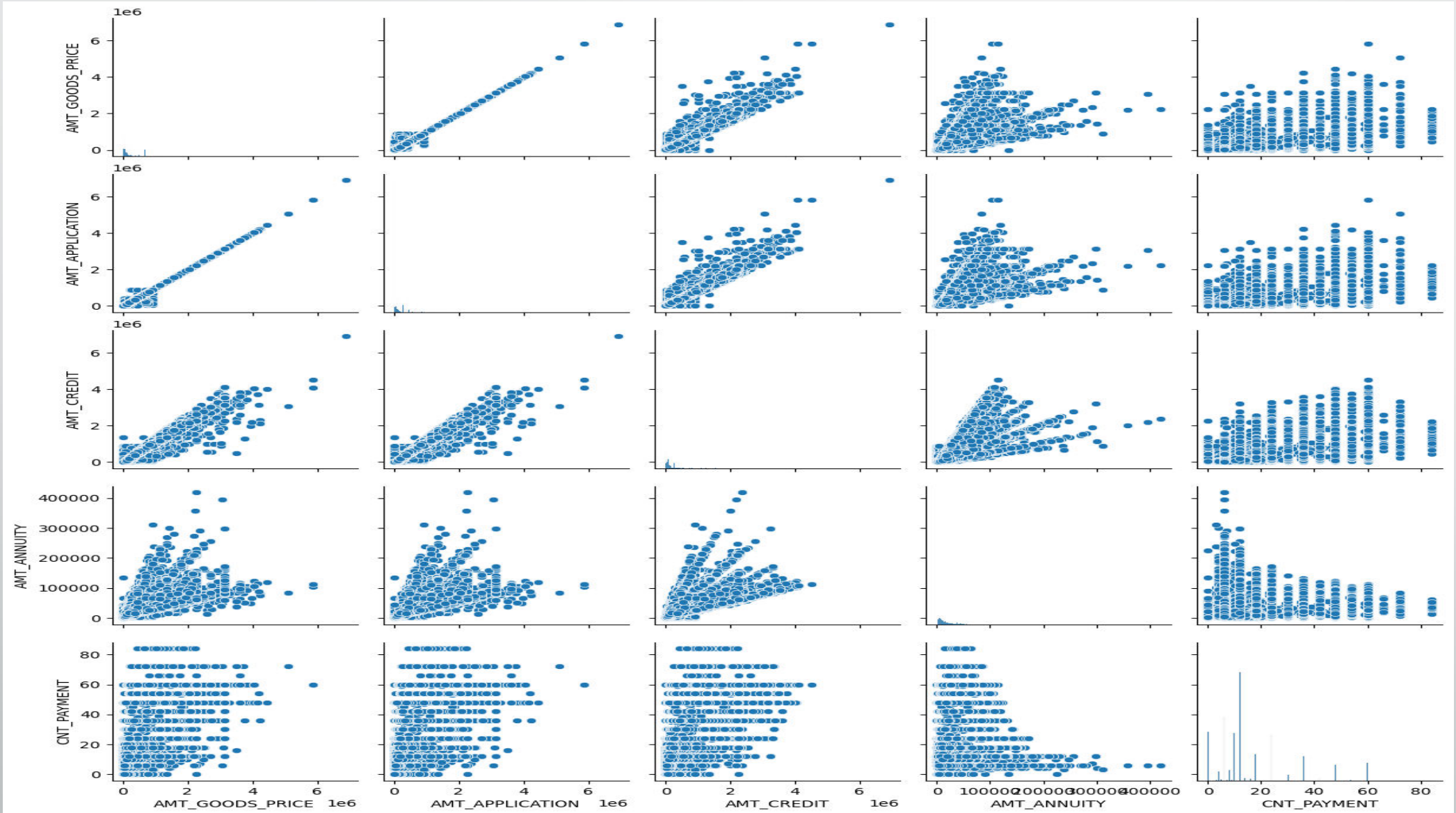


In approved category, consumer loan has largest no of applicants.

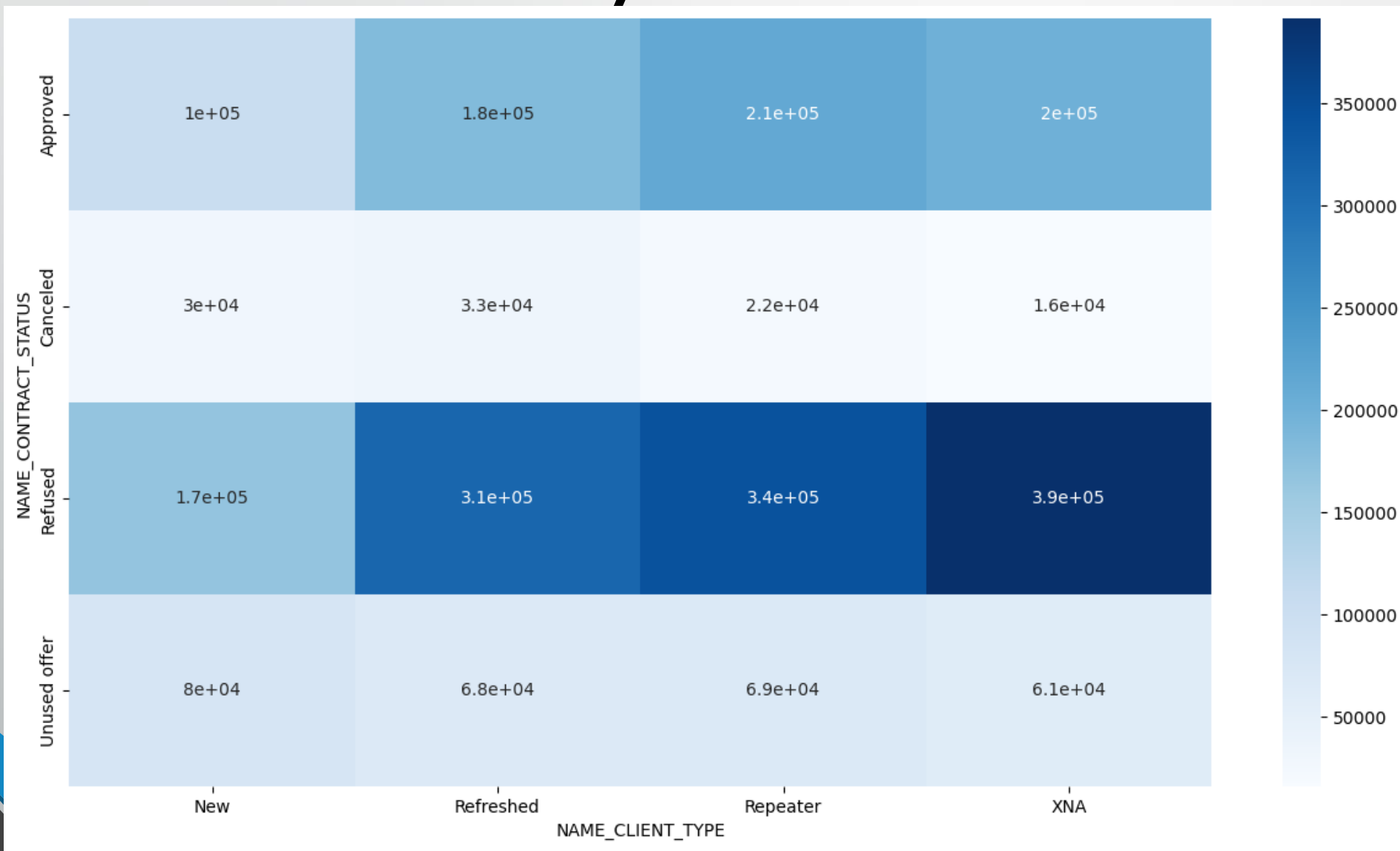
There seem to be no cancelled loans in cash loan category than consumer loan

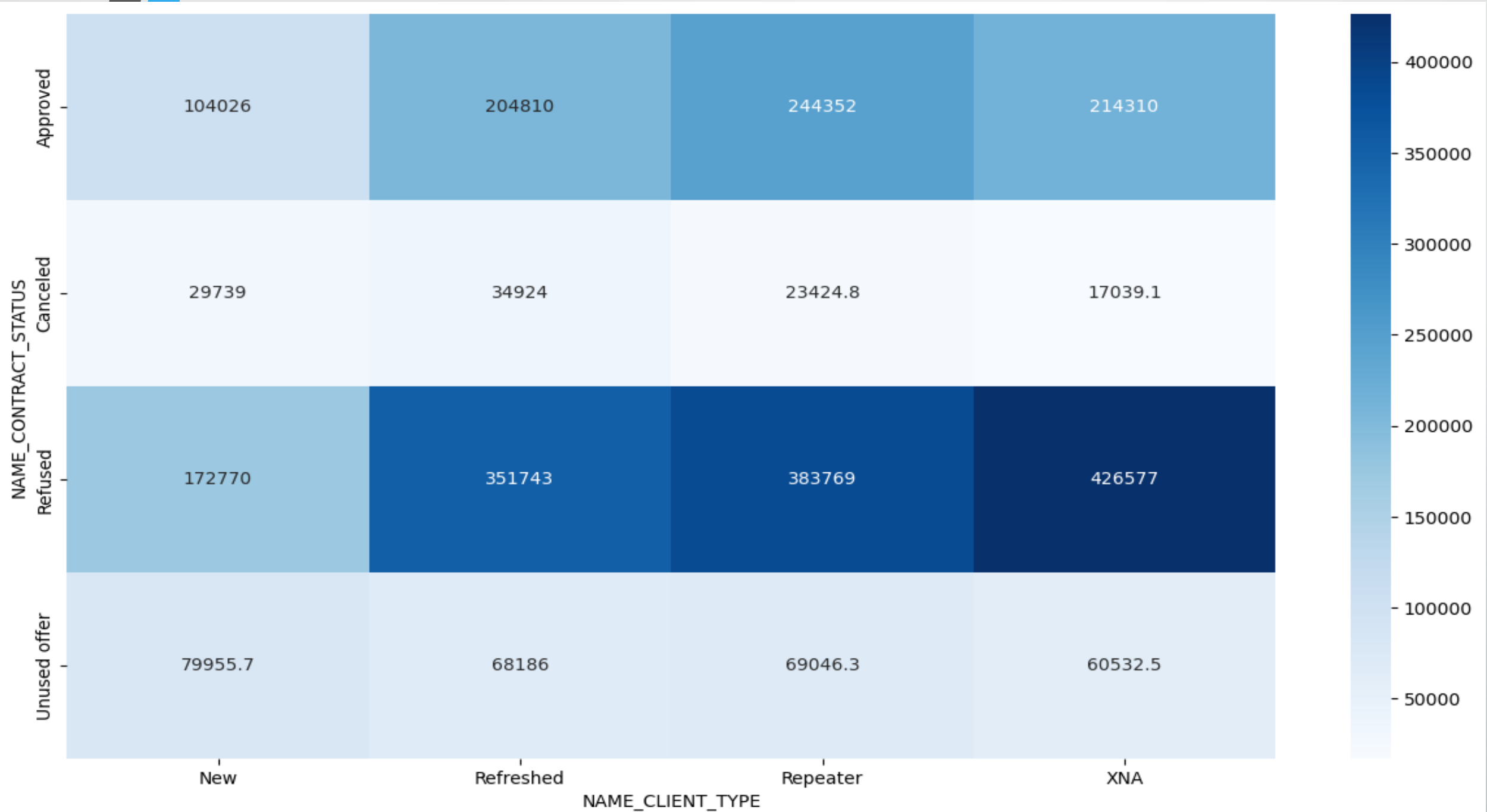
More cash loans have been refused than consumer loans.

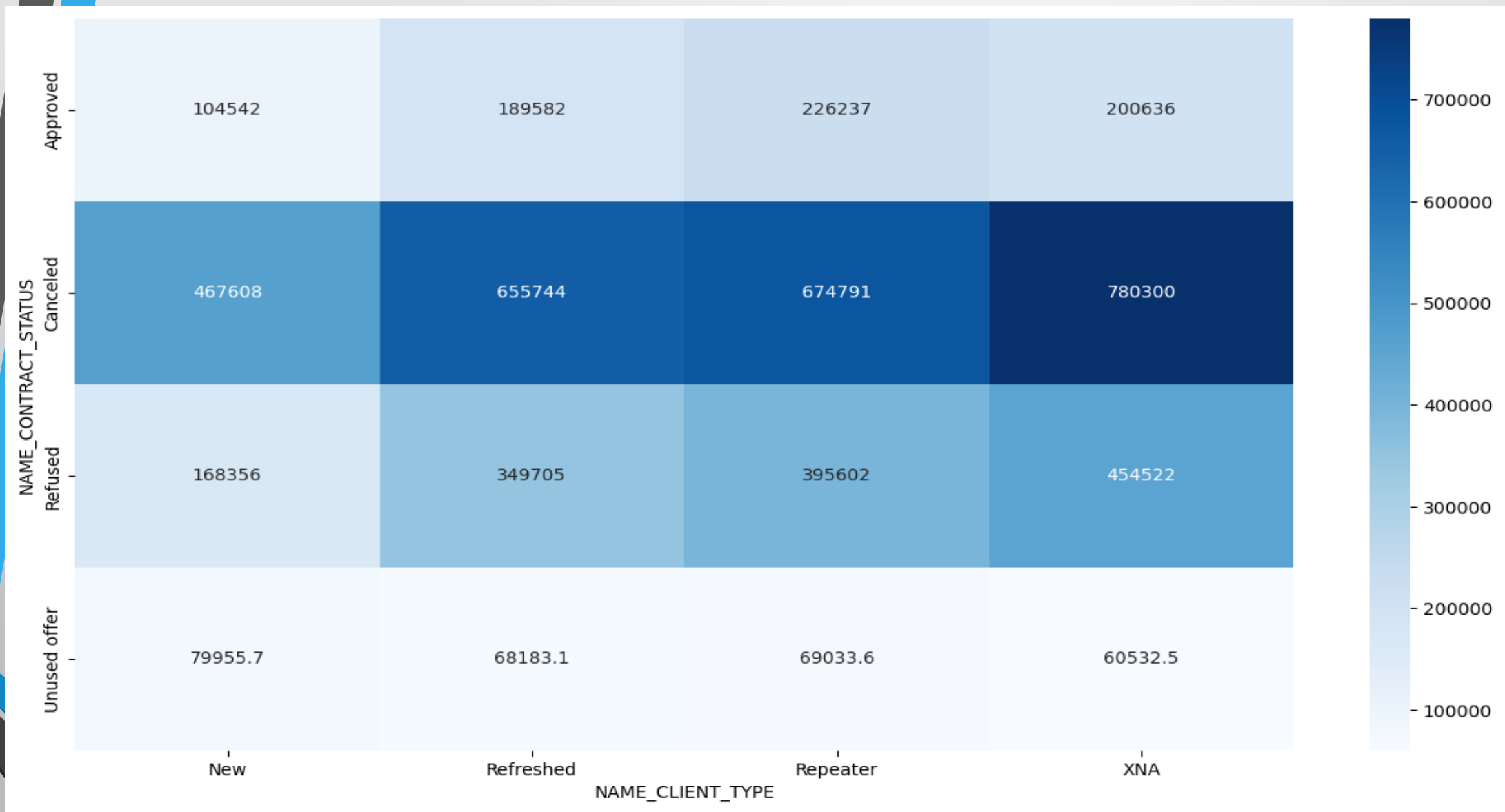
# Top Correlations



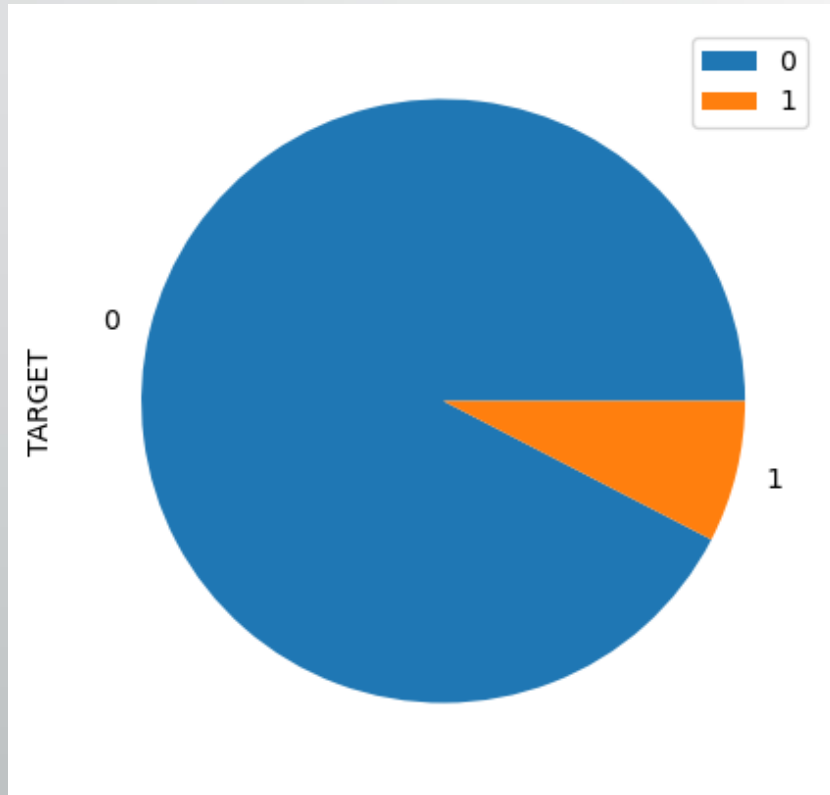
# Multivariate Analysis





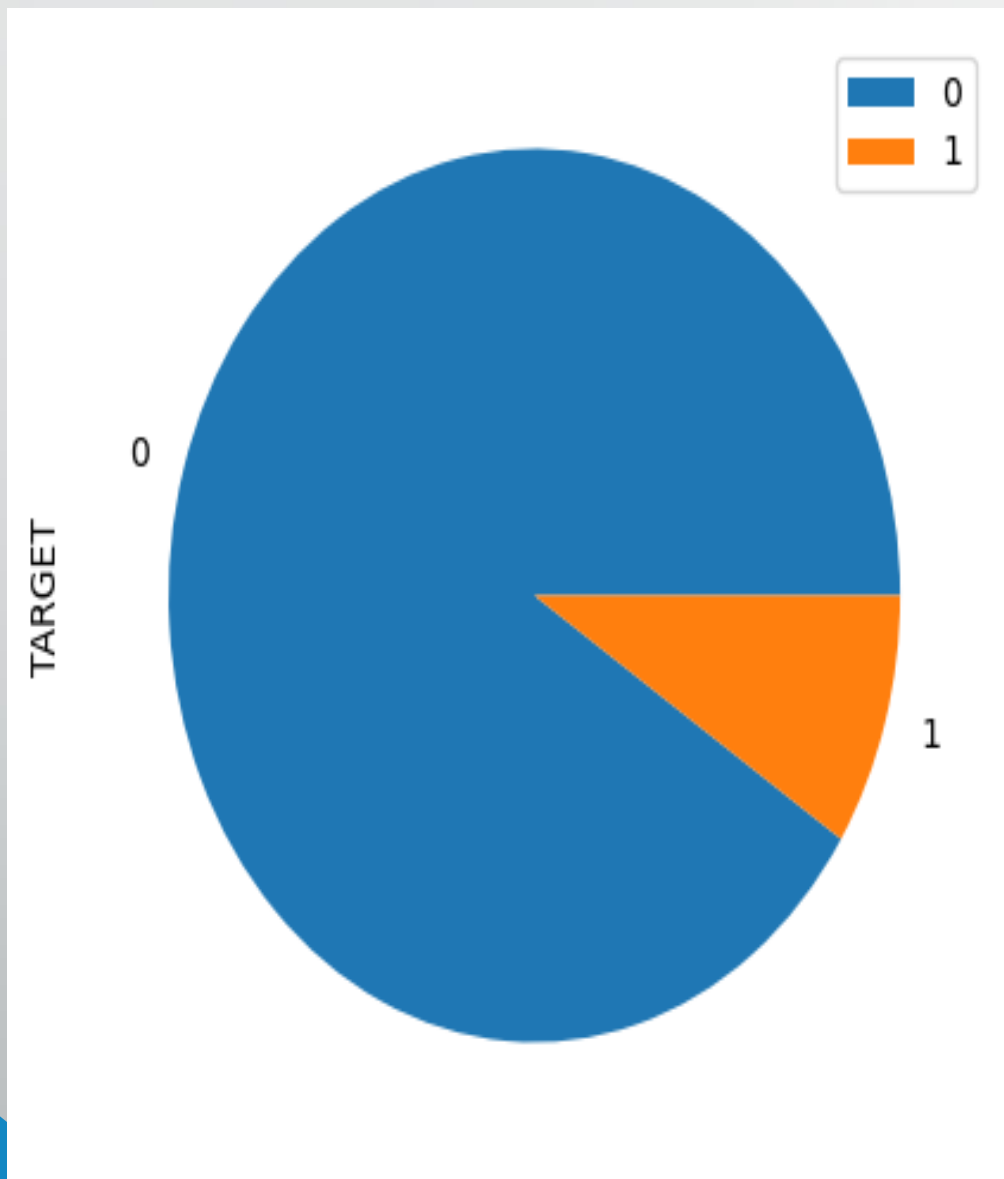


# Merged data frame analysis



Target 0 and 1 for: Approved  
Target 0- 0.924113  
Target 1- 0.075887

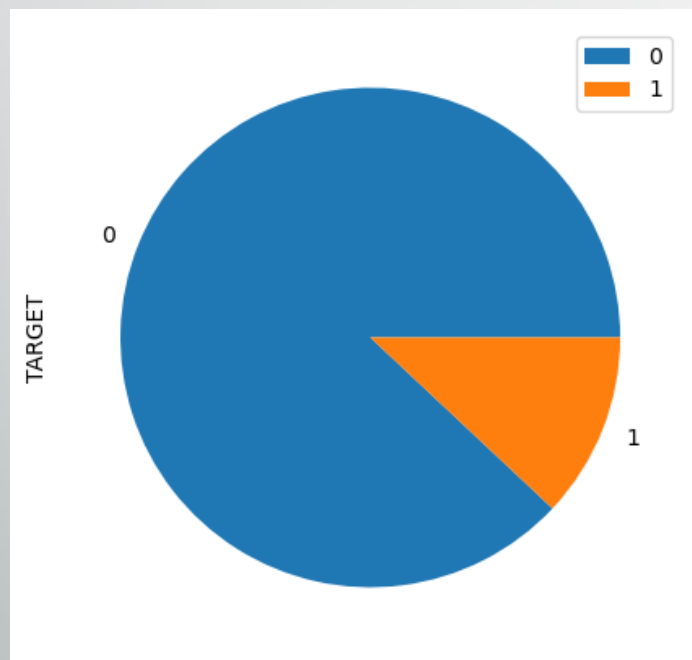




Target 0 and 1 for:  
Canceled

Target 0- 0.908264

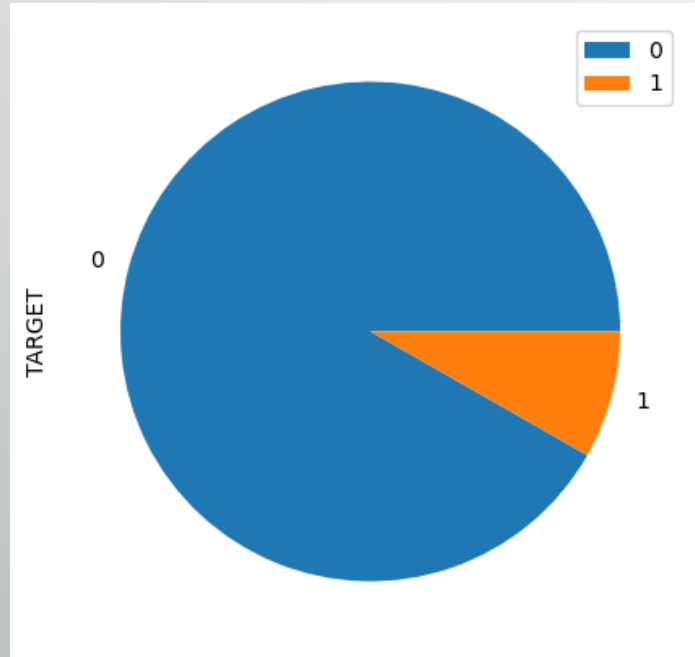
Target 1- 0.091736



Target 0 and 1 for: Refused

Target 0 0.880036

Target 1 0.119964

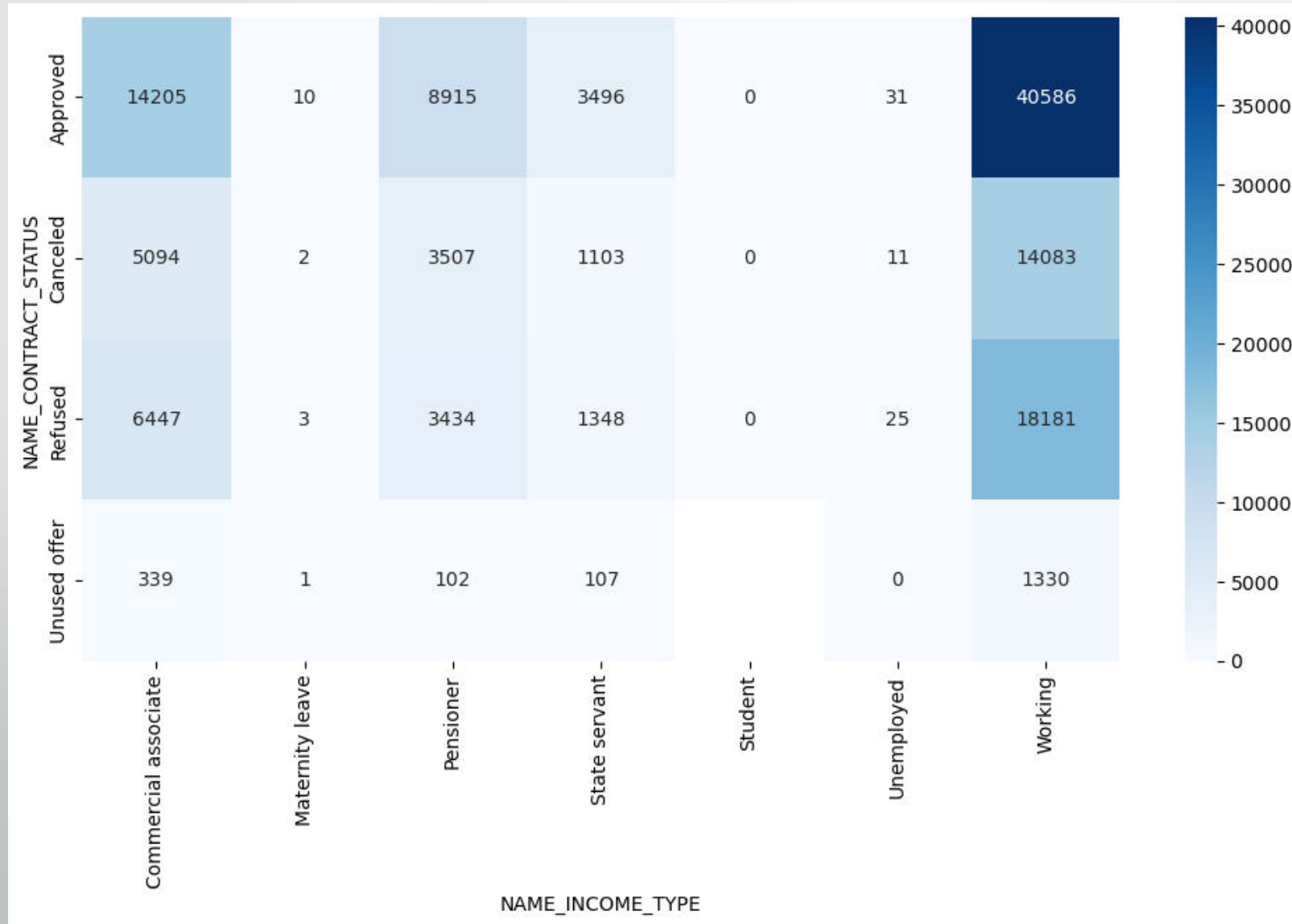


Target 0 and 1 for: Unused offer

Target 0 0.917483

Target 1 0.082517

## NAME\_CONTRACT\_STATUS, NAME\_INCOME\_TYPE, aggregating on Target



NAME\_CONTRACT\_STATUS, NAME\_INCOME\_TYPE, aggregating on Target

