CREDIT EDA CASE STUDY

SUJITH M TOM
SHARIE R NATH

AGENDA

- Give an idea of applying EDA in a real business scenario.
- Develop a basic understanding of risk analytics in banking and financial services
- Understand how data is used to minimise the risk of losing money while lending to customers

Business Understanding

When the company receives a loan application, the company has to decide for loan approval based on the applicant’s profile. 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.

TARGET VARIABLE

The dataset contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:

- **The client with payment difficulties:** he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample,
- **All other cases:** All other cases when the payment is paid on time.

TYPES OF DECISION

When a client applies for a loan, there are four types of decisions that could be taken by the client/company

- **Approved:** The Company has approved loan Application
- **Cancelled:** The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.
- **Refused:** The company had rejected the loan (because the client does not meet their requirements etc.).
- **Unused offer:** Loan has been cancelled by the client but on different stages of the process.

PURPOSE OF CASE STUDY

- Using EDA techniques analyse the dataset to understand how consumer attributes and loan attributes influence the tendency of default

DATA UNDERSTANDING

- *'application_data.csv'* contains all the information of the client at the time of application.
  The data is about whether a **client has payment difficulties.**

- *'previous_application.csv'* contains information about the client's previous loan data. It contains the data whether the previous application had been **Approved, Cancelled, Refused or Unused offer.**

PROBLEM STATEMENT

- Identify patterns which indicate if a client has difficulty paying their installments
- Ensuring consumers capable of repaying the loan are not rejected
- Find out client variables and loan variables that are high indicators of defaulting

Exploratory Data Analysis Approach

The dataset is being analysed using following steps:

- Data Sourcing
- Data Cleansing
- Derived Metrics
- Univariate Analysis
- Segmented Univariate Analysis
- Bivariate Analysis
- Correlation Analysis
- Arriving at Insights

Data Sourcing

- *'application_data.csv' and 'previous_application.csv' datasets are meged on current application ID*
- *The combined dataset is used for analysis*
- *The columns are divided into numerical and catrgorical for ease of analysis*

Data Cleansing

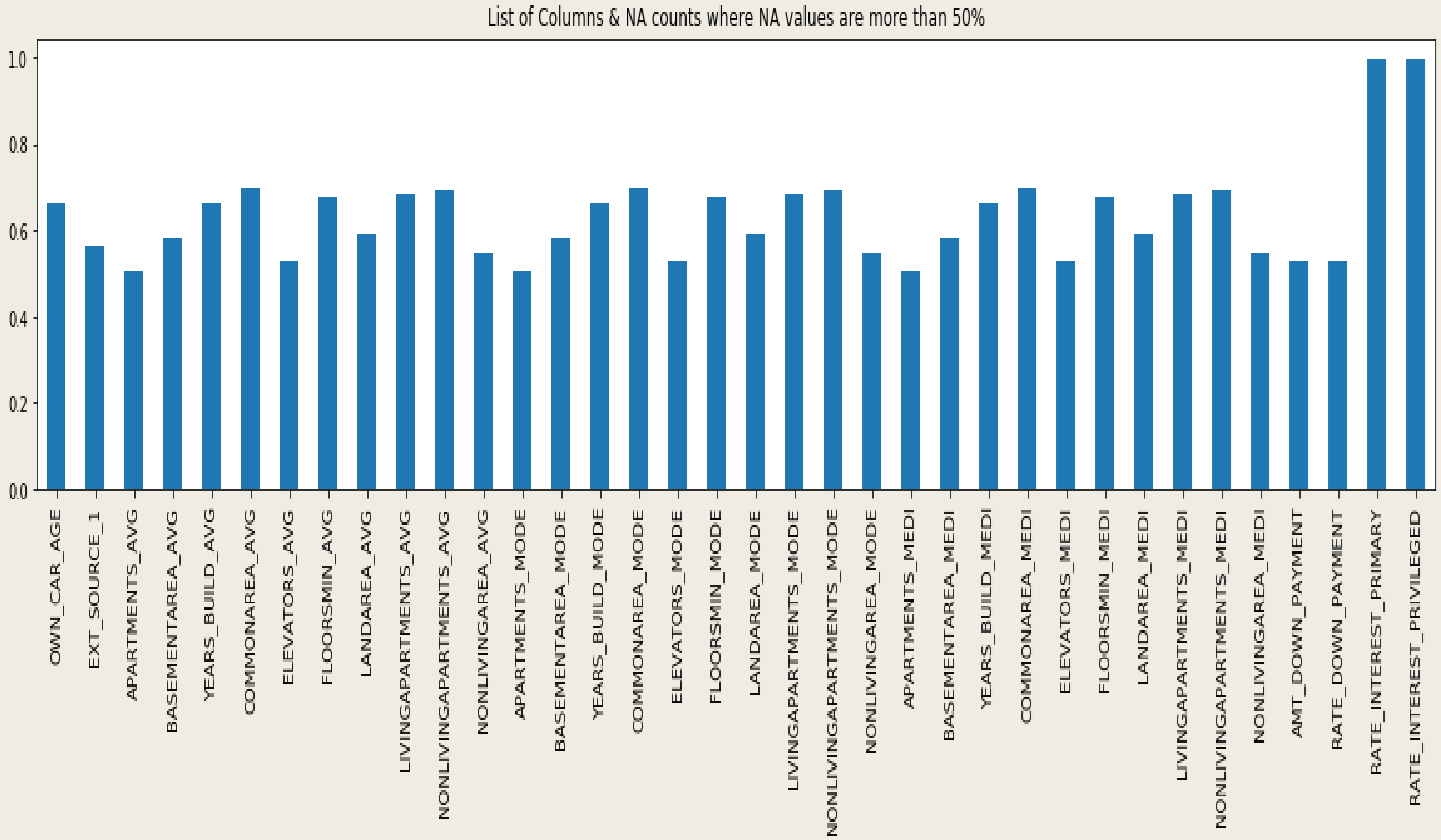
- Fixing Rows and Columns
- Dealing with missing values
- Detecting Outliers
- Finding Data Imbalance

Fixing Rows and Columns

- Duplicate Rows and Columns are removed
- Certain columns are renamed
- Datatype of certain columns changed to category
- Columns segregated based on datatype as numerical,object and categorical

Dealing with missing values

- Null values in categorical columns replaced with appropriate values
- Numerical Columns with more than 50% missing values identified
- Insignificant columns are removed
- RATE_INTEREST_PRIMARY and RATE_INTEREST_PRIVILEGED removed


List of Columns & NA counts where NA values are more than 50%
1.0
0.8
0.6
0.4
0.2
0.0
OWN_CAR_AGE
EXT_SOURCE_1
APARTMENTS_AVG
BASEMENTAREA_AVG
YEARS_BUILD_AVG
COMMONAREA_AVG
ELEVATORS_AVG
FLOORSMIN_AVG
LANDAREA_AVG
LIVINGAPARTMENTS_AVG
NONLIVINGAPARTMENTS_AVG
NONLIVINGAREA_AVG
APARTMENTS_MODE
BASEMENTAREA_MODE
YEARS_BUILD_MODE
COMMONAREA_MODE
ELEVATORS_MODE
FLOORSMIN_MODE
LANDAREA_MODE
LIVINGAPARTMENTS_MODE
NONLIVINGAPARTMENTS_MODE
NONLIVINGAREA_MODE
APARTMENTS_MEDI
BASEMENTAREA_MEDI
YEARS_BUILD_MEDI
COMMONAREA_MEDI
ELEVATORS_MEDI
FLOORSMIN_MEDI
LANDAREA_MEDI
LIVINGAPARTMENTS_MEDI
NONLIVINGAPARTMENTS_MEDI
NONLIVINGAREA_MEDI
AMT_DOWN_PAYMENT
RATE_DOWN_PAYMENT
RATE_INTEREST_PRIMARY
RATE_INTEREST_PRIVILEGED

Detecting Outliers

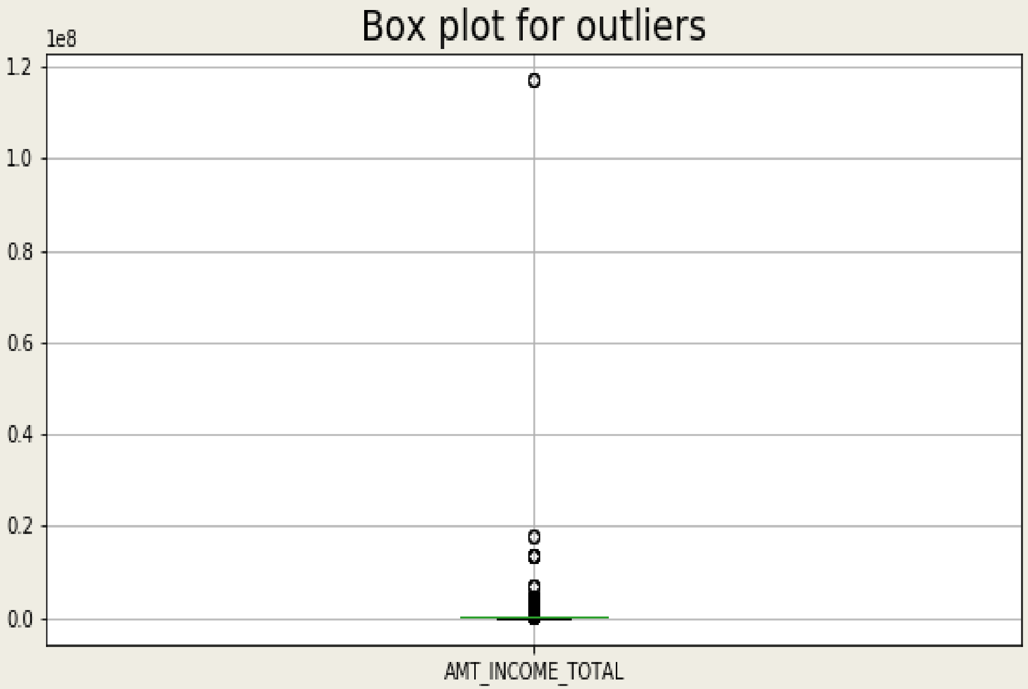
- Numerical columns are analysed for identifying outliers using **Box plots**
- Outlier Datapoints are detected using **IQR method**
- Datapoints that are beyond 1.5 times Inter Quartile Regions are considered as outliers
- Rows with Significant Outliers for particular variables are dropped

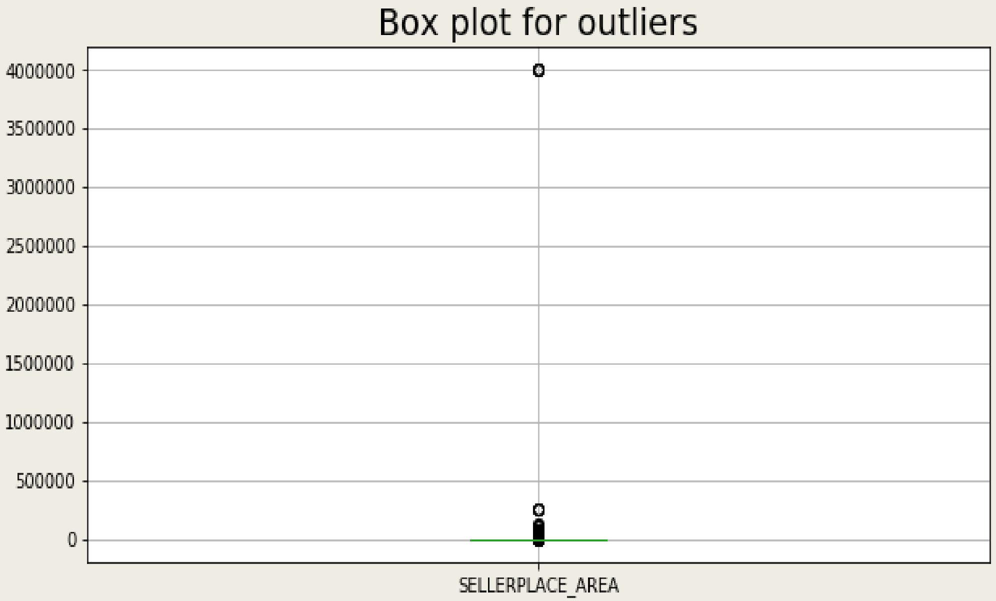
Detected and Removed Outliers

SELLERPLACE_AREA :[4000000]

AMT_INCOME_TOTAL: [117000000.0, 18000090.0, 13500000.0]

Box Plots for outliers





Pseudo Outliers

Highly Significant outliers were found as follows:

DAYS_EMPLOYED:[365243]

DAYS_FIRST_DUE :[365243.0]

DAYS_LAST_DUE_1ST_VERSION:[365243.0]

DAYS_LAST_DUE :[365243.0]

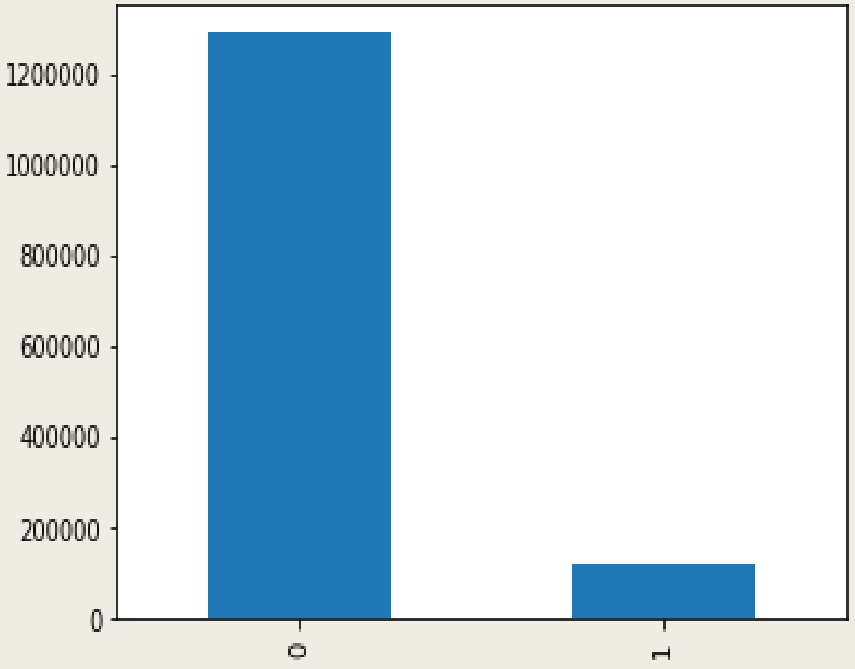
DAYS_TERMINATION :[365243.0]

Since the value is spread over many rows, it is assumed as indication of unemployment or other category

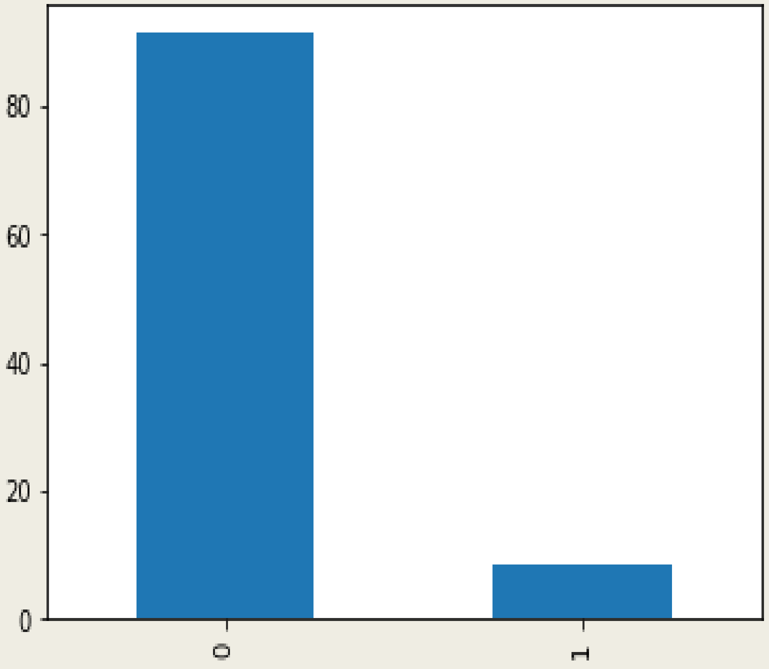
Data Imbalance

- Dataset is segmented into Repayers with **Target=0** and Defaulters with **Target=1**
- Countplot of Repayers and Defaulters indicates DATA IMBALANCE
- Count of Repayers = 1291326
- Count of Defaulters =122357
- Ratio of Data Imbalance is 1291326 :122357 ie. **10.554:1**

TARGET VARIABLE -ABSOLUTE VALUE PLOT

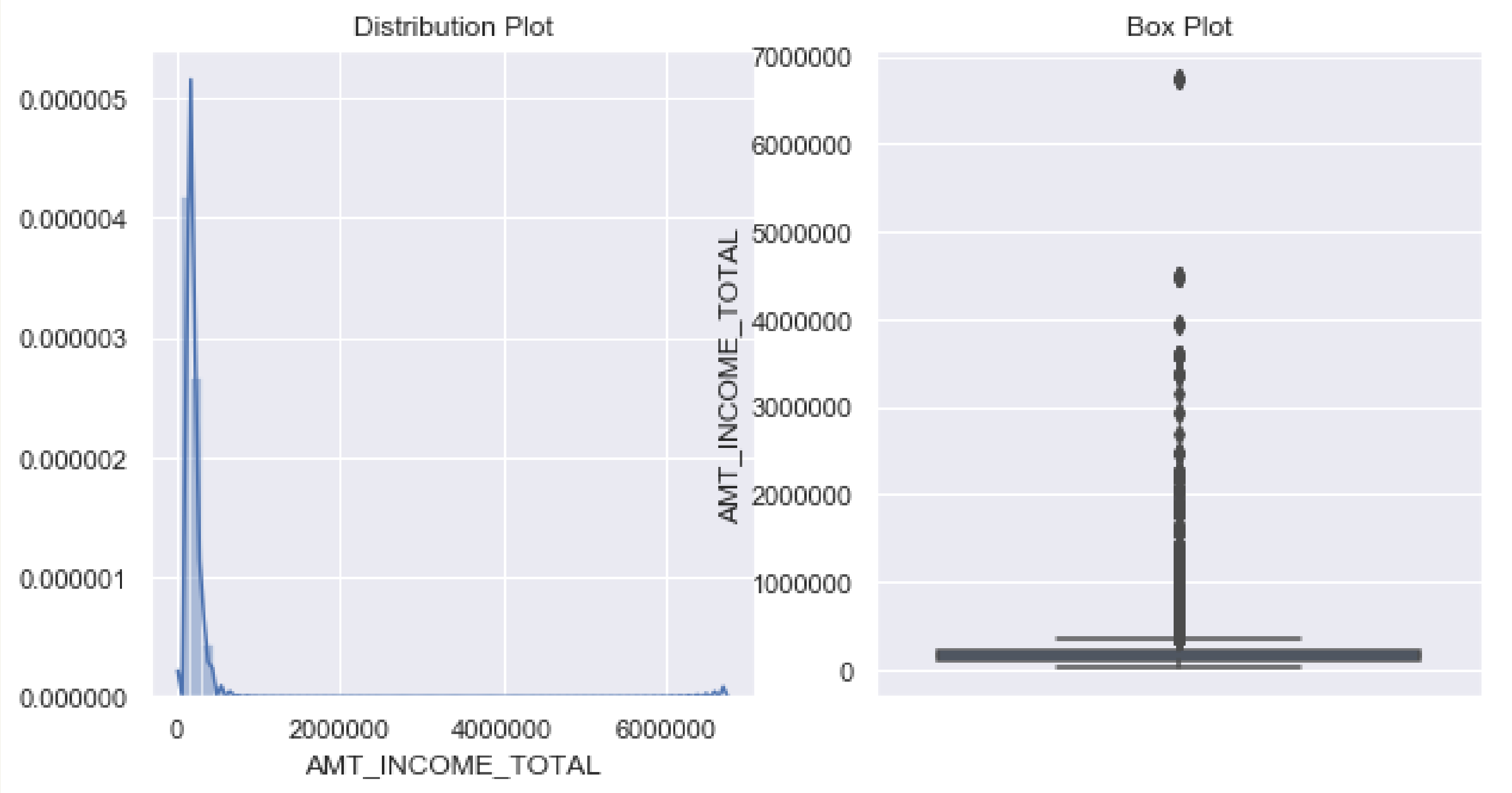


TARGRT VARIABLE- % of Total Plot

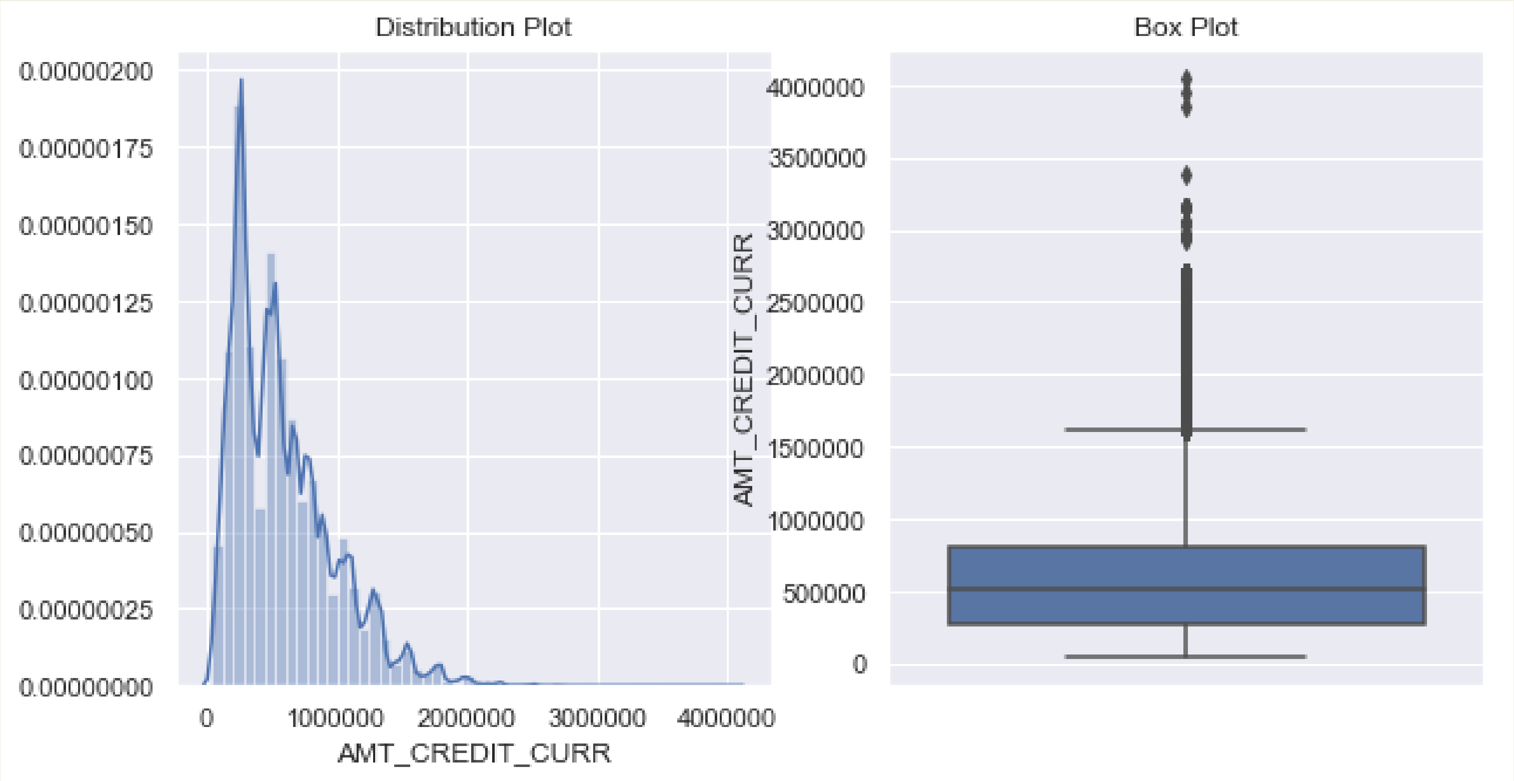


Univariate Analysis

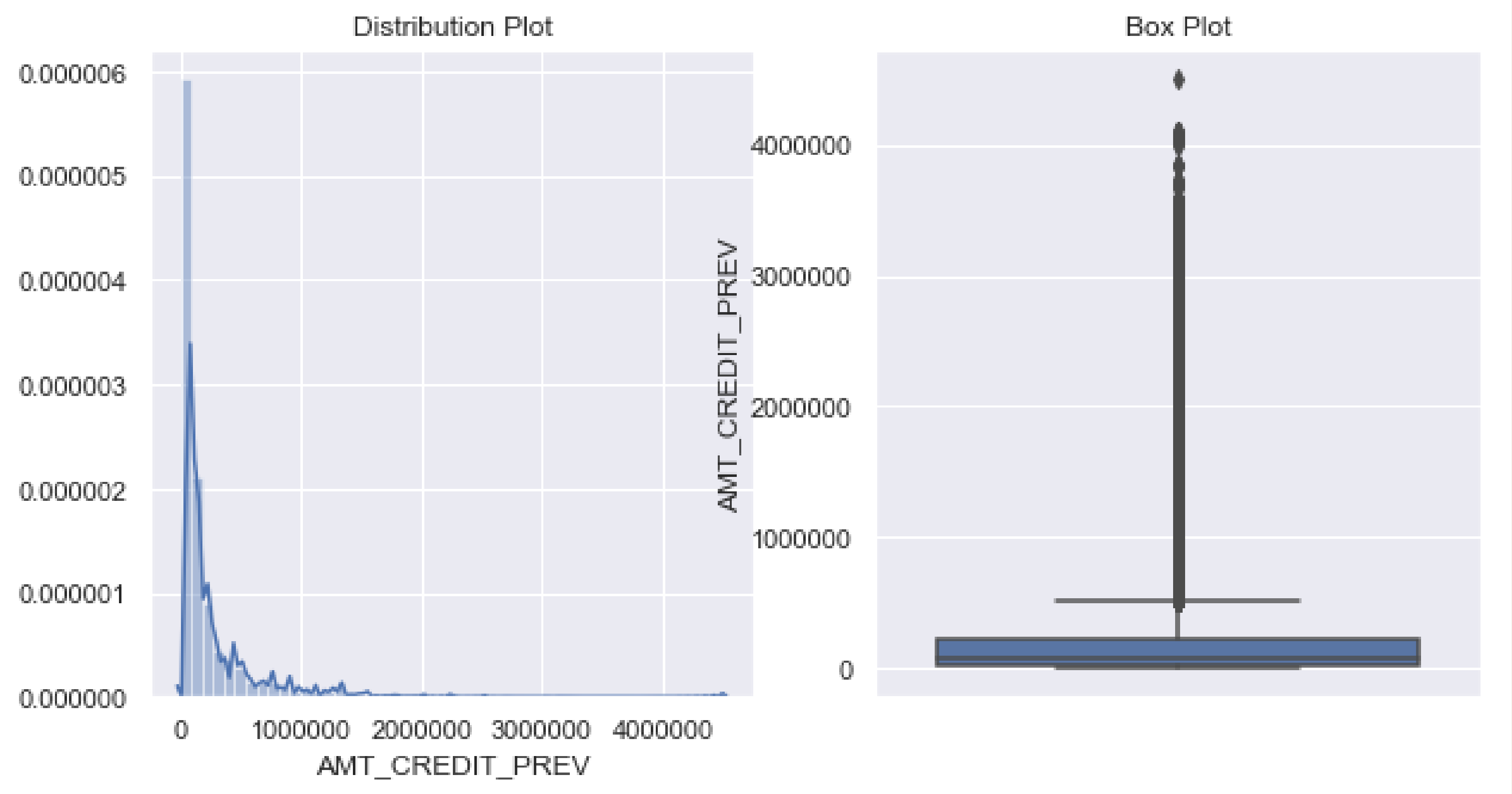
TOTAL INCOME



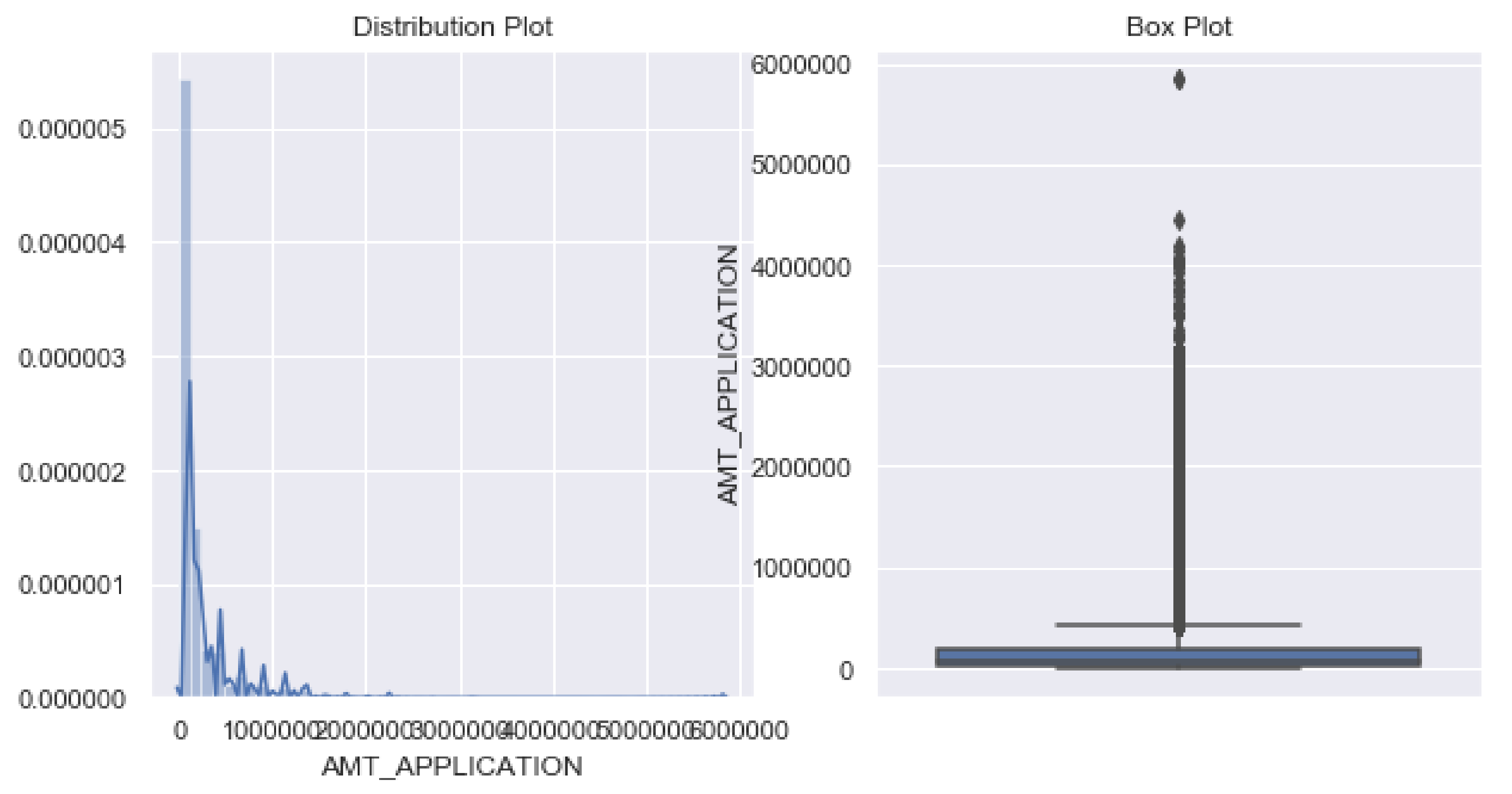
CURRENT CREDIT AMOUNT



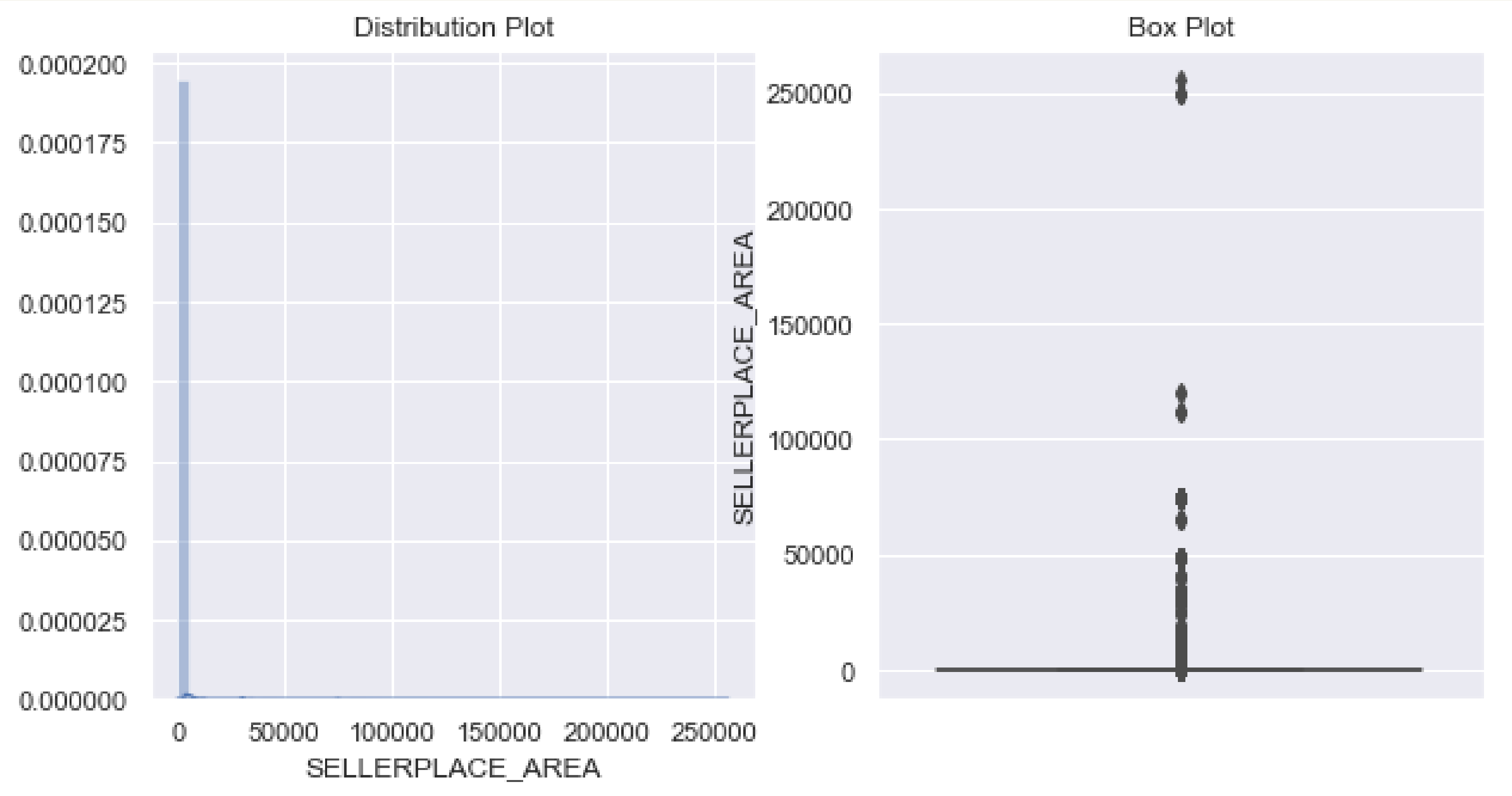
PREVIOUS CREDIT AMOUNT



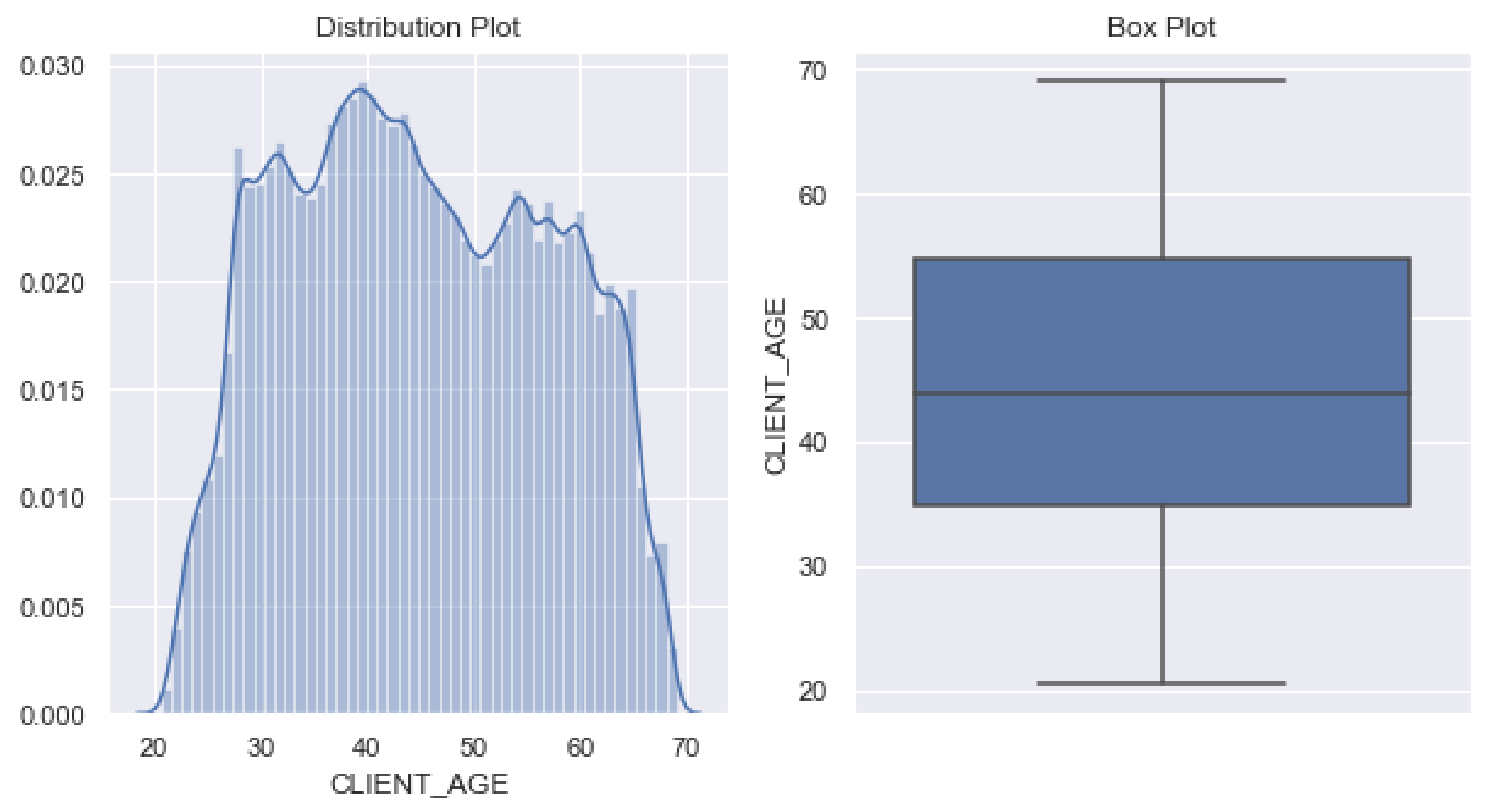
APPLICATION AMOUNT



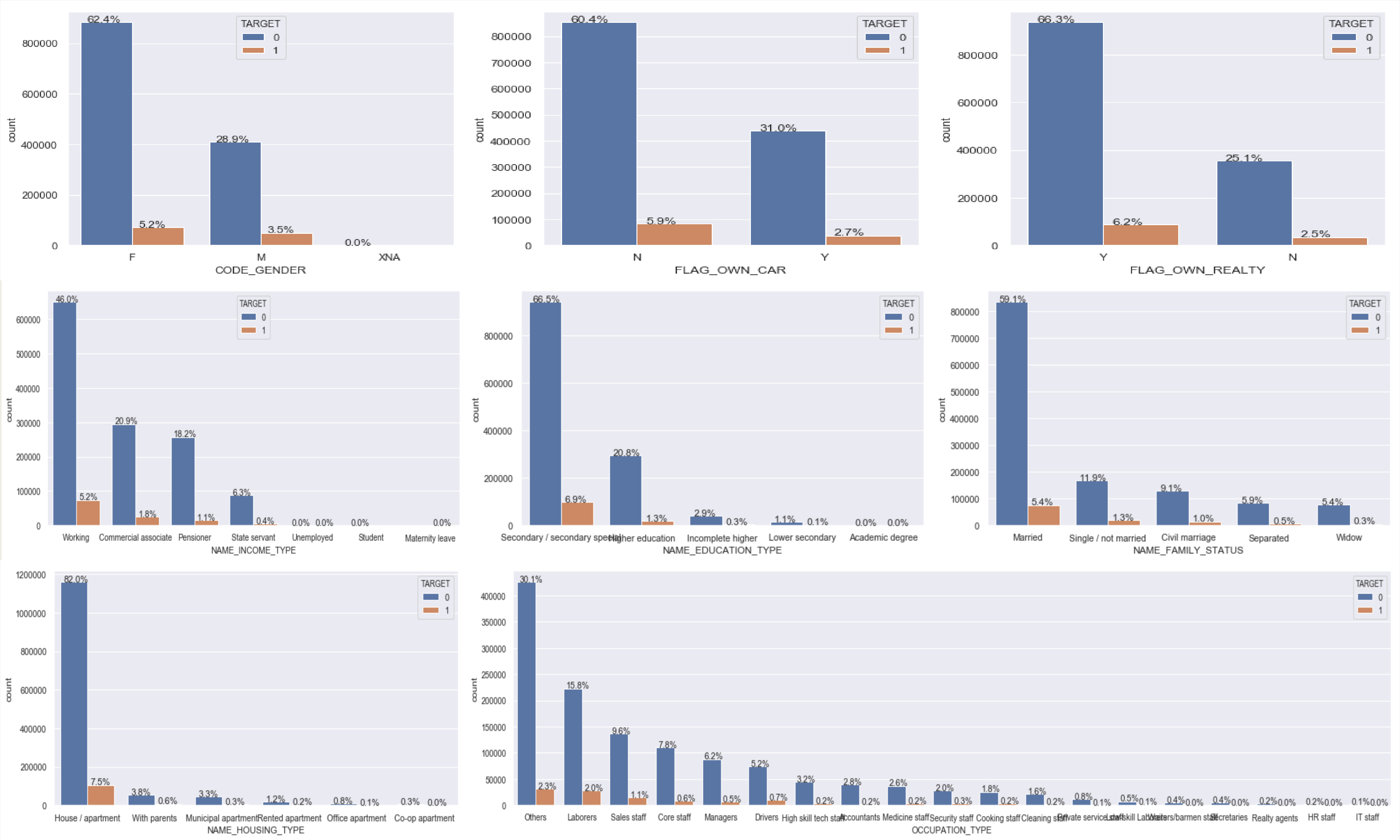
SELLER PLACE AREA

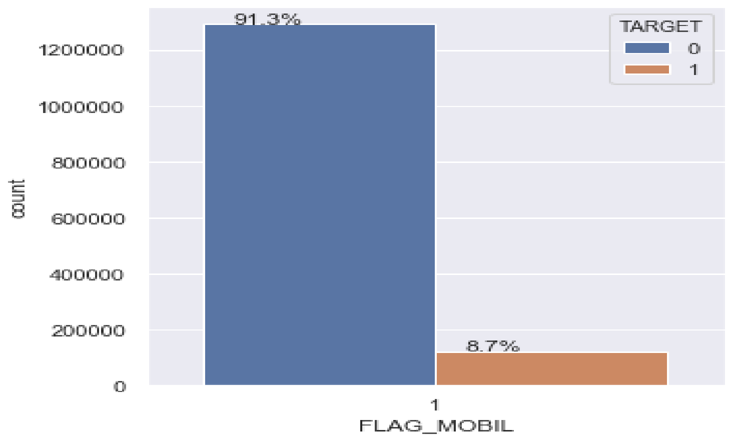


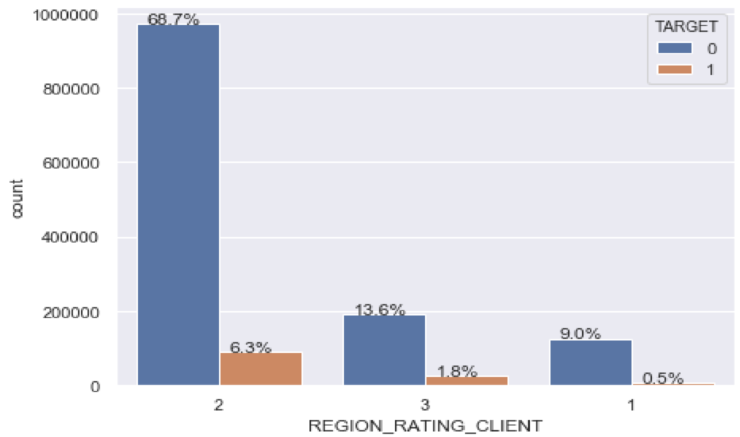
CLIENT AGE

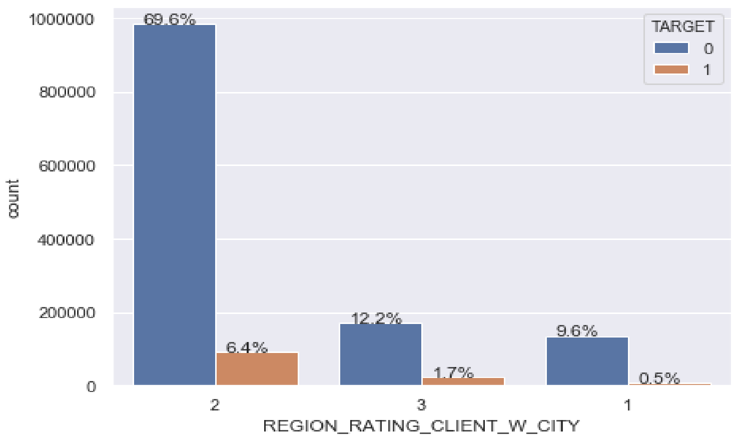

Distribution Plot
0.030
0.025
0.020
0.015
0.010
0.005
0.000
20
30
40
50
60
70
CLIENT_AGE
Box Plot
70
60
50
40
30
20
CLIENT_AGE

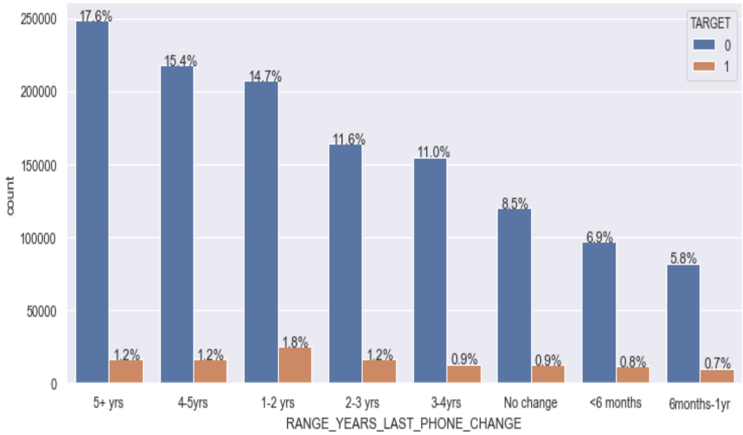
Segmented Univariate Analysis

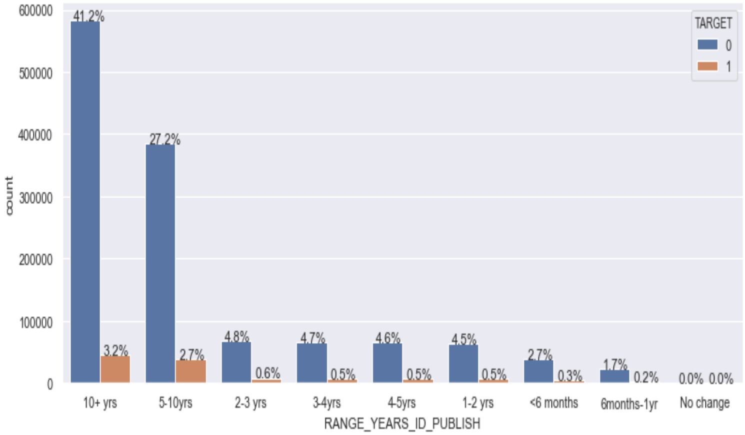

count
CODE_GENDER
F
M
XNA
62.4%
5.2%
28.9%
3.5%
0.0%
TARGET
0
1
0
200000
400000
600000
800000
FLAG_OWN_CAR
N
Y
60.4%
5.9%
31.0%
2.7%
0
100000
200000
300000
400000
500000
600000
700000
800000
FLAG_OWN_REALTY
Y
N
66.3%
6.2%
25.1%
2.5%
NAME_INCOME_TYPE
Working
Commercial associate
Pensioner
State servant
Unemployed
Student
Maternity leave
46.0%
5.2%
20.9%
1.8%
18.2%
1.1%
6.3%
0.4%
0.0%
0.0%
0.0%
0.0%
NAME_EDUCATION_TYPE
Secondary / secondary special
Higher education
Incomplete higher
Lower secondary
Academic degree
66.5%
6.9%
20.8%
1.3%
2.9%
0.3%
1.1%
0.1%
0.0%
0.0%
NAME_FAMILY_STATUS
Married
Single / not married
Civil marriage
Separated
Widow
59.1%
5.4%
11.9%
1.3%
9.1%
1.0%
5.9%
0.5%
5.4%
0.3%
NAME_HOUSING_TYPE
House / apartment
With parents
Municipal apartment
Rented apartment
Office apartment
Co-op apartment
82.0%
7.5%
3.8%
0.6%
3.3%
0.3%
1.2%
0.2%
0.8%
0.1%
0.3%
0.0%
1200000
1000000
OCCUPATION_TYPE
Others
Laborers
Sales staff
Core staff
Managers
Drivers
High skill tech staff
Accountants
Medicine staff
Security staff
Cooking staff
Cleaning staff
Private service staff
Low-skill Laborers
Waiters/barmen staff
Secretaries
Realty agents
HR staff
IT staff
30.1%
2.3%
15.8%
2.0%
9.6%
1.1%
7.8%
0.6%
6.2%
0.5%
5.2%
0.7%
3.2%
0.2%
2.8%
0.2%
2.6%
0.2%
2.0%
0.3%
1.8%
0.2%
1.6%
0.2%
0.8%
0.1%
0.5%
0.1%
0.4%
0.0%
0.4%
0.0%
0.2%
0.0%
0.2%
0.0%
0.1%
0.0%
50000
100000
150000
200000
250000
300000
350000
400000

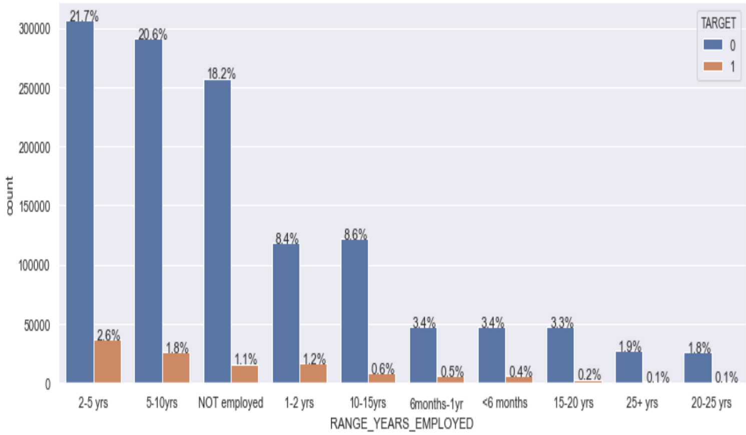

count
0
200000
400000
600000
800000
1000000
1200000
91.3%
8.7%
TARGET
0
1
1
FLAG_MOBIL

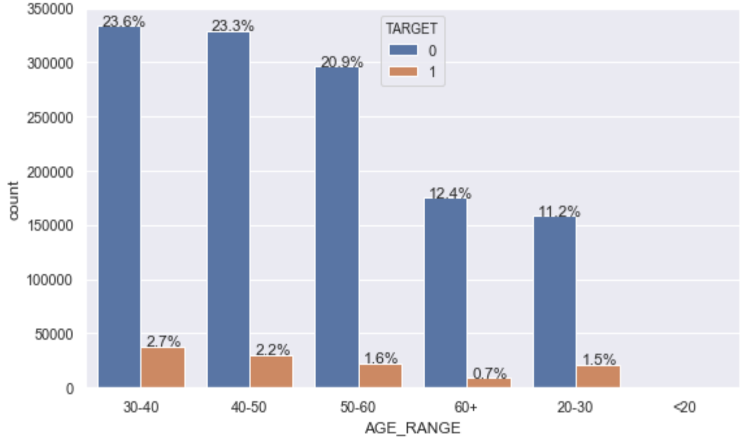

count
0
200000
400000
600000
800000
1000000
68.7%
6.3%
13.6%
1.8%
9.0%
0.5%
TARGET
0
1
2
3
1
REGION_RATING_CLIENT

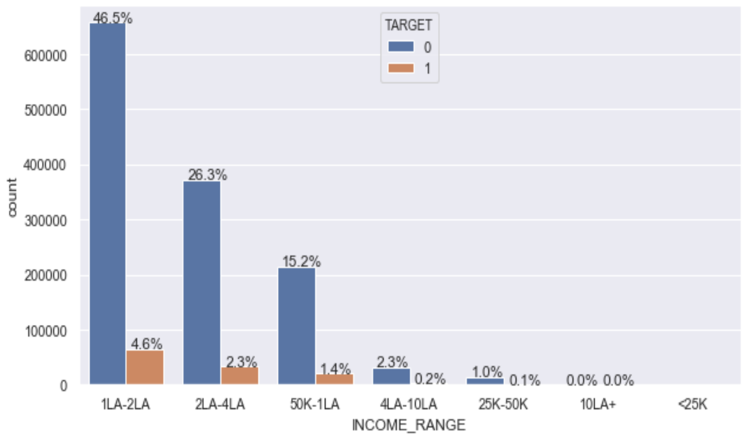

count
0
200000
400000
600000
800000
1000000
69.6%
6.4%
12.2%
1.7%
9.6%
0.5%
TARGET
0
1
2
3
1
REGION_RATING_CLIENT_W_CITY

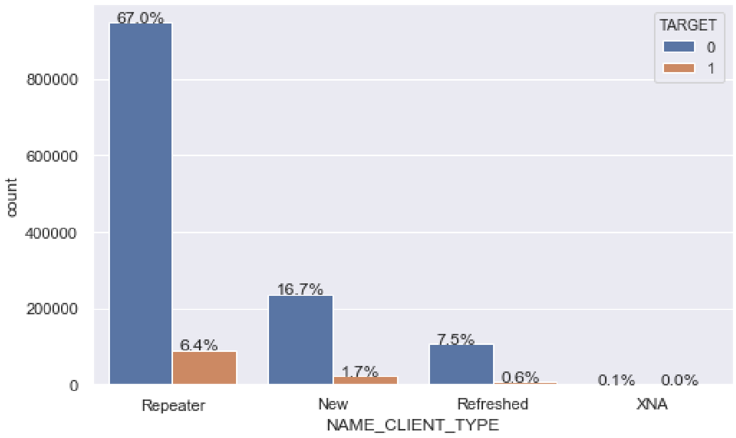

count
0
50000
100000
150000
200000
250000
17.6%
1.2%
15.4%
1.2%
14.7%
1.8%
11.6%
1.2%
11.0%
0.9%
8.5%
0.9%
6.9%
0.8%
5.8%
0.7%
TARGET
0
1
5+ yrs
4-5yrs
1-2 yrs
2-3 yrs
3-4yrs
No change
<6 months
6months-1yr
RANGE_YEARS_LAST_PHONE_CHANGE

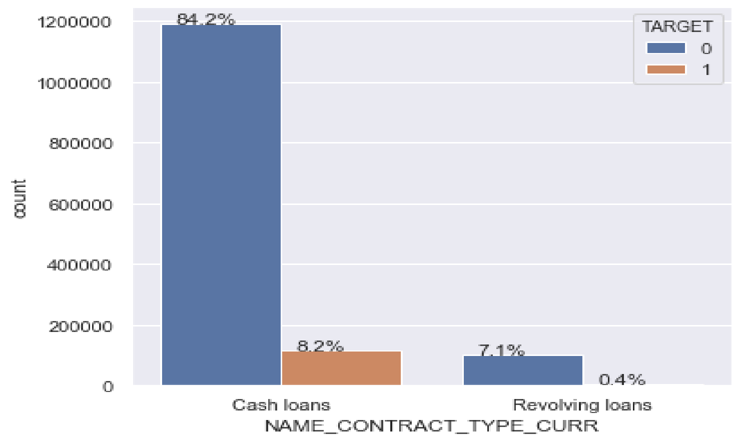

count
0
100000
200000
300000
400000
500000
600000
41.2%
3.2%
27.2%
2.7%
4.8%
0.6%
4.7%
0.5%
4.6%
0.5%
4.5%
0.5%
2.7%
0.3%
1.7%
0.2%
0.0%
0.0%
TARGET
0
1
10+ yrs
5-10yrs
2-3 yrs
3-4yrs
4-5yrs
1-2 yrs
<6 months
6months-1yr
No change
RANGE_YEARS_ID_PUBLISH

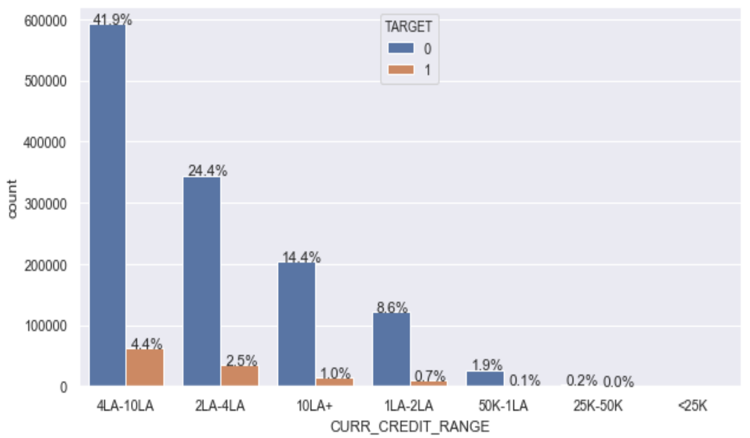

count
0
50000
100000
150000
200000
250000
300000
21.7%
2.6%
20.6%
1.8%
18.2%
1.1%
8.4%
1.2%
8.6%
0.6%
3.4%
0.5%
3.4%
0.4%
3.3%
0.2%
1.9%
0.1%
1.8%
0.1%
TARGET
0
1
2-5 yrs
5-10yrs
NOT employed
1-2 yrs
10-15yrs
6months-1yr
<6 months
15-20 yrs
25+ yrs
20-25 yrs
RANGE_YEARS_EMPLOYED

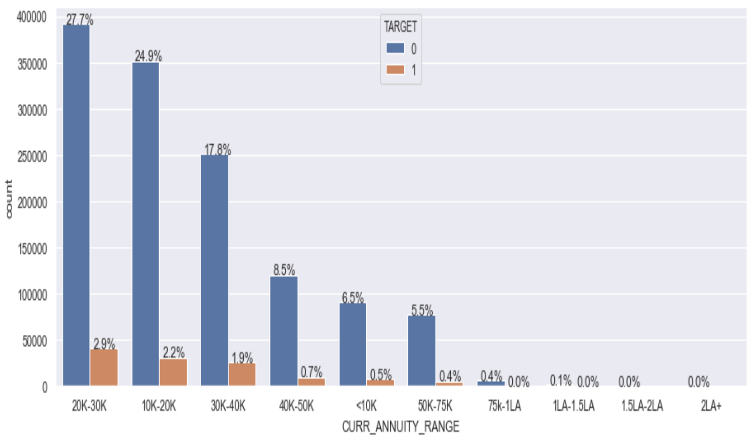

count
0
50000
100000
150000
200000
250000
300000
350000
23.6%
2.7%
23.3%
2.2%
20.9%
1.6%
12.4%
0.7%
11.2%
1.5%
TARGET
0
1
30-40
40-50
50-60
60+
20-30
<20
AGE_RANGE

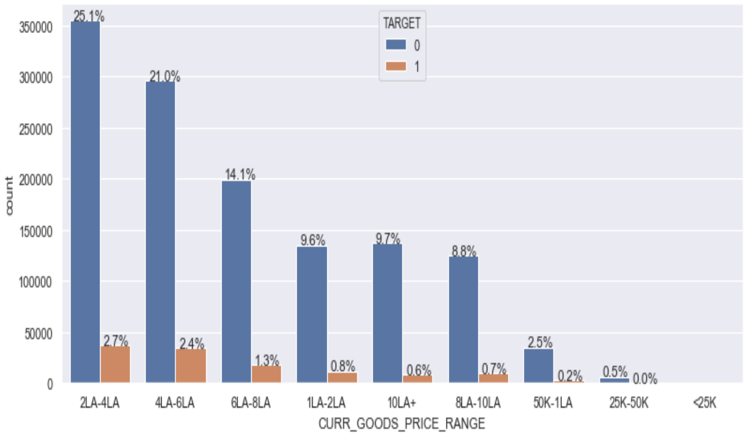

count
0
100000
200000
300000
400000
500000
600000
46.5%
4.6%
26.3%
2.3%
15.2%
1.4%
2.3%
0.2%
1.0%
0.1%
0.0%
0.0%
TARGET
0
1
1LA-2LA
2LA-4LA
50K-1LA
4LA-10LA
25K-50K
10LA+
<25K
INCOME_RANGE

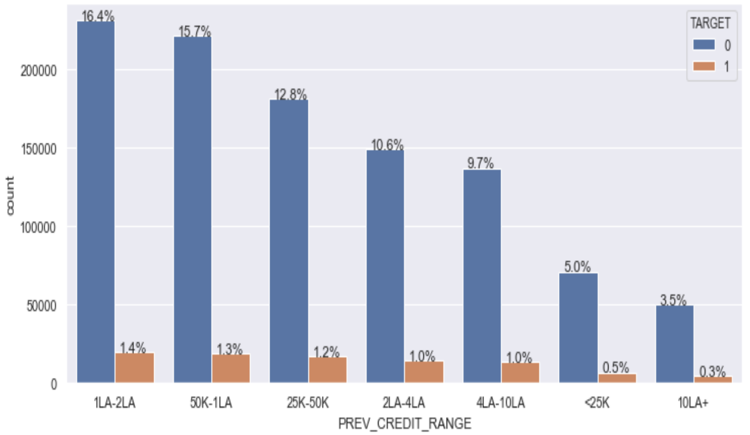

count
0
200000
400000
600000
800000
67.0%
6.4%
16.7%
1.7%
7.5%
0.6%
0.1%
0.0%
TARGET
0
1
Repeater
New
Refreshed
XNA
NAME_CLIENT_TYPE

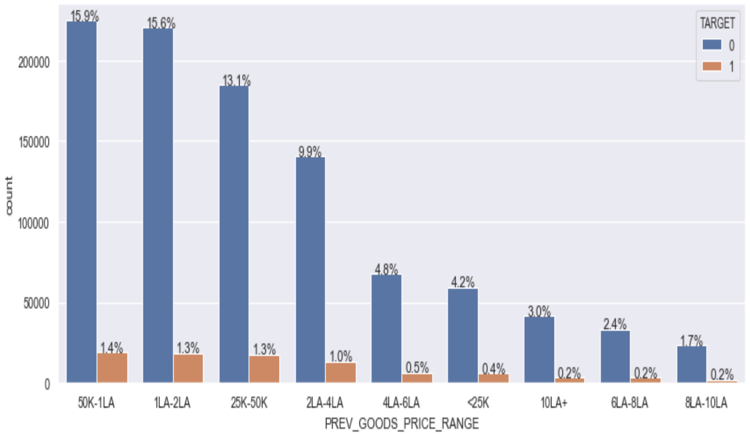

count
1200000
1000000
800000
600000
400000
200000
0
84.2%
8.2%
7.1%
0.4%
TARGET
0
1
Cash loans
Revolving loans
NAME_CONTRACT_TYPE_CURR

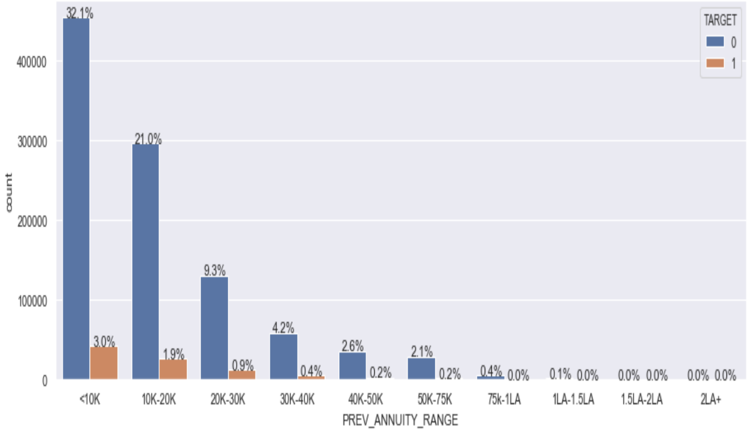

count
600000
500000
400000
300000
200000
100000
0
41.9%
4.4%
24.4%
2.5%
14.4%
1.0%
8.6%
0.7%
1.9%
0.1%
0.2%
0.0%
TARGET
0
1
4LA-10LA
2LA-4LA
10LA+
1LA-2LA
50K-1LA
25K-50K
<25K
CURR_CREDIT_RANGE

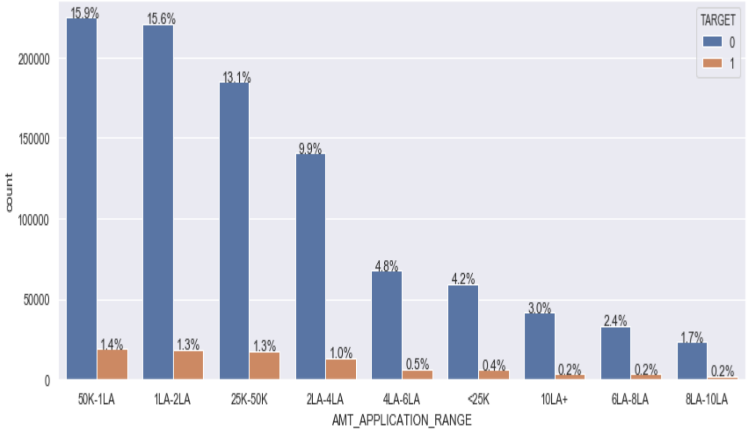

count
400000
350000
300000
250000
200000
150000
100000
50000
0
27.7%
2.9%
24.9%
2.2%
17.8%
1.9%
8.5%
0.7%
6.5%
0.5%
5.5%
0.4%
0.4%
0.0%
0.1%
0.0%
0.0%
0.0%
TARGET
0
1
20K-30K
10K-20K
30K-40K
40K-50K
<10K
50K-75K
75k-1LA
1LA-1.5LA
1.5LA-2LA
2LA+
CURR_ANNUITY_RANGE

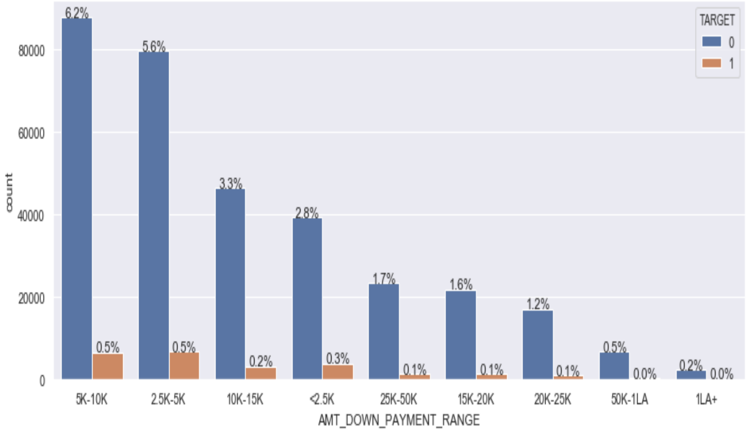

count
350000
300000
250000
200000
150000
100000
50000
0
25.1%
2.7%
21.0%
2.4%
14.1%
1.3%
9.6%
0.8%
9.7%
0.6%
8.8%
0.7%
2.5%
0.2%
0.5%
0.0%
TARGET
0
1
2LA-4LA
4LA-6LA
6LA-8LA
1LA-2LA
10LA+
8LA-10LA
50K-1LA
25K-50K
<25K
CURR_GOODS_PRICE_RANGE

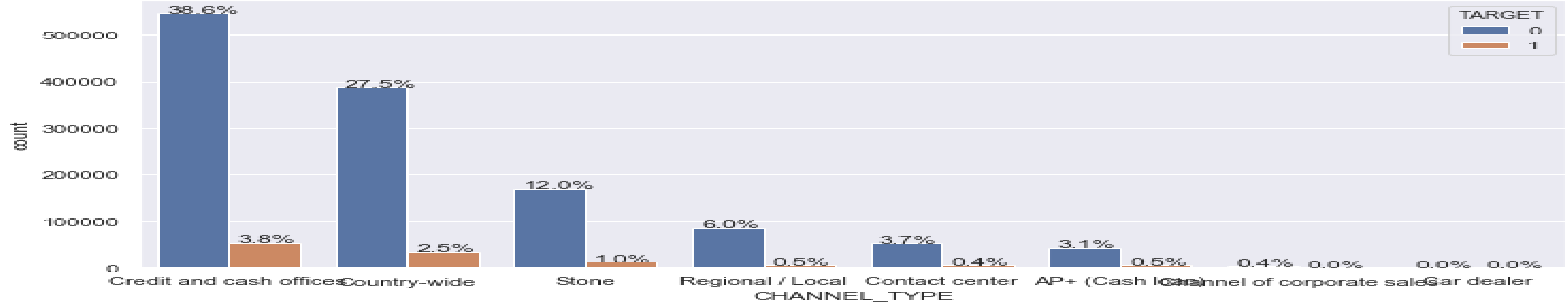

count
200000
150000
100000
50000
0
16.4%
1.4%
15.7%
1.3%
12.8%
1.2%
10.6%
1.0%
9.7%
1.0%
5.0%
0.5%
3.5%
0.3%
TARGET
0
1
1LA-2LA
50K-1LA
25K-50K
2LA-4LA
4LA-10LA
<25K
10LA+
PREV_CREDIT_RANGE

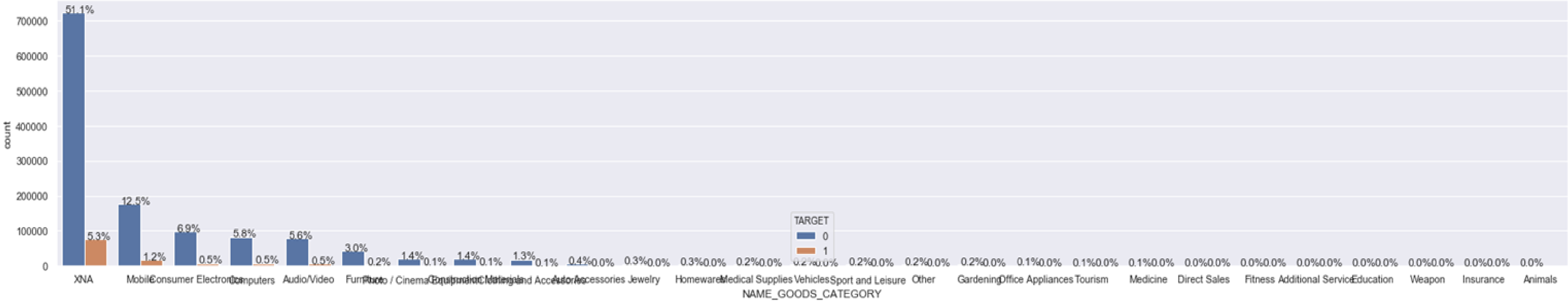

count
200000
150000
100000
50000
0
15.9%
1.4%
15.6%
1.3%
13.1%
1.3%
9.9%
1.0%
4.8%
0.5%
4.2%
0.4%
3.0%
0.2%
2.4%
0.2%
1.7%
0.2%
TARGET
0
1
50K-1LA
1LA-2LA
25K-50K
2LA-4LA
4LA-6LA
<25K
10LA+
6LA-8LA
8LA-10LA
PREV_GOODS_PRICE_RANGE

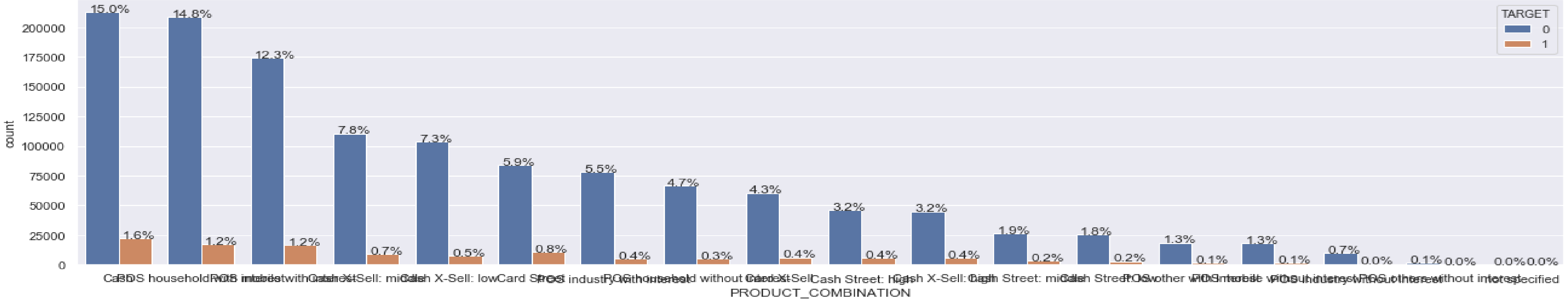

count
400000
300000
200000
100000
0
32.1%
3.0%
21.0%
1.9%
9.3%
0.9%
4.2%
0.4%
2.6%
0.2%
2.1%
0.2%
0.4%
0.0%
0.1%
0.0%
0.0%
0.0%
0.0%
0.0%
TARGET
0
1
<10K
10K-20K
20K-30K
30K-40K
40K-50K
50K-75K
75k-1LA
1LA-1.5LA
1.5LA-2LA
2LA+
PREV_ANNUITY_RANGE

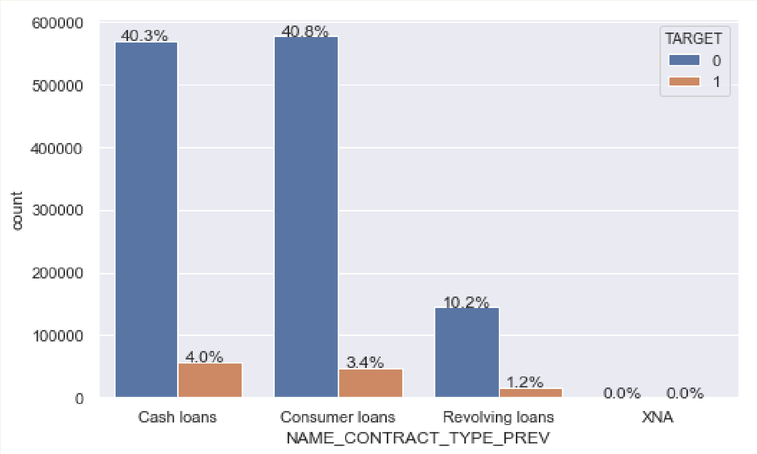

count
200000
150000
100000
50000
0
15.9%
1.4%
15.6%
1.3%
13.1%
1.3%
9.9%
1.0%
4.8%
0.5%
4.2%
0.4%
3.0%
0.2%
2.4%
0.2%
1.7%
0.2%
TARGET
0
1
50K-1LA
1LA-2LA
25K-50K
2LA-4LA
4LA-6LA
<25K
10LA+
6LA-8LA
8LA-10LA
AMT_APPLICATION_RANGE


count
80000
60000
40000
20000
0
6.2%
0.5%
5.6%
0.5%
3.3%
0.2%
2.8%
0.3%
1.7%
0.1%
1.6%
0.1%
1.2%
0.1%
0.5%
0.0%
0.2%
0.0%
TARGET
0
1
5K-10K
2.5K-5K
10K-15K
<2.5K
25K-50K
15K-20K
20K-25K
50K-1LA
1LA+
AMT_DOWN_PAYMENT_RANGE

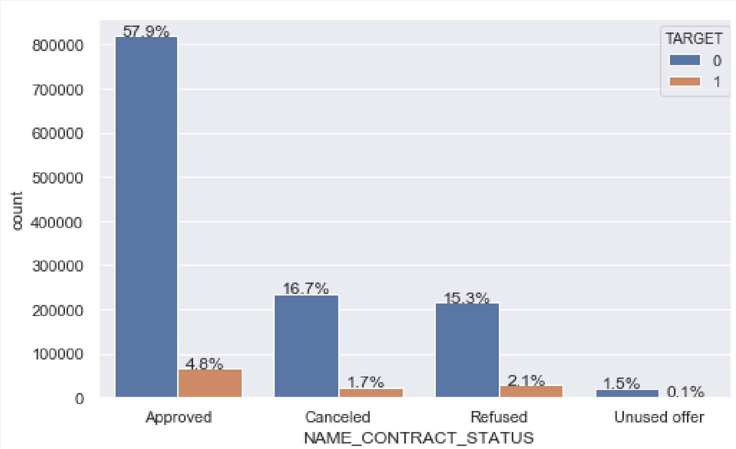

TARGET
0
1
count
500000
400000
300000
200000
100000
0
38.6%
3.8%
27.5%
2.5%
12.0%
1.0%
6.0%
0.5%
3.7%
0.4%
3.1%
0.5%
0.4%
0.0%
0.0%
0.0%
Credit and cash offices
Country-wide
Stone
Regional / Local
Contact center
AP+ (Cash loan)
Channel of corporate sales
Car dealer
CHANNEL_TYPE


count
700000
600000
500000
400000
300000
200000
100000
0
TARGET
0
1
51.1%
5.3%
12.5%
1.2%
6.9%
0.5%
5.8%
0.5%
5.6%
0.5%
3.0%
0.2%
1.4%
0.1%
1.4%
0.1%
1.3%
0.1%
0.4%0.0%
0.3%0.0%
0.3%0.0%
0.2%0.0%
0.2%0.0%
0.2%0.0%
0.2%0.0%
0.2%0.0%
0.1%0.0%
0.1%0.0%
0.1%0.0%
0.0%0.0%
0.0%0.0%
0.0%0.0%
0.0%0.0%
0.0%0.0%
0.0%0.0%
0.0%
XNA
Mobile
Consumer Electronics
Computers
Audio/Video
Furniture
Photo / Cinema Equipment
Construction Materials
Clothing and Accessories
Auto Accessories
Jewelry
Homewares
Medical Supplies
Vehicles
Sport and Leisure
Other
Gardening
Office Appliances
Tourism
Medicine
Direct Sales
Fitness
Additional Service
Education
Weapon
Insurance
Animals
NAME_GOODS_CATEGORY

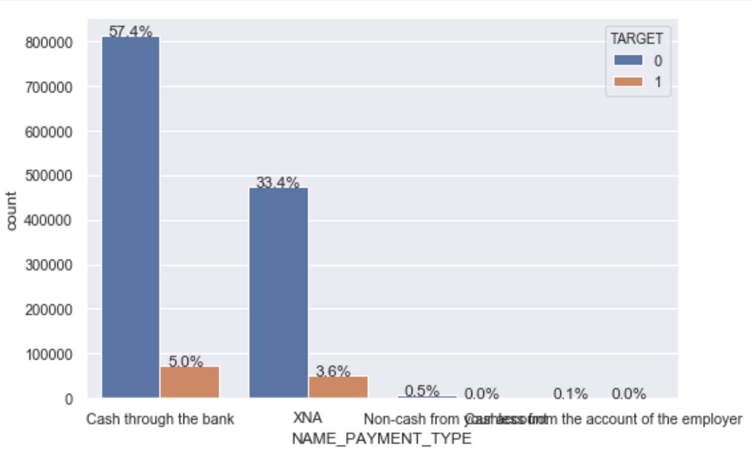

count
200000
175000
150000
125000
100000
75000
50000
25000
0
TARGET
0
1
15.0%
1.6%
14.8%
1.2%
12.3%
1.2%
7.8%
0.7%
7.3%
0.5%
5.9%
0.8%
5.5%
0.4%
4.7%
0.3%
4.3%
0.4%
3.2%
0.4%
3.2%
0.4%
1.9%
0.2%
1.8%
0.2%
1.3%
0.1%
1.3%
0.1%
0.7%
0.0%
0.1%
0.0%
0.0%0.0%
Cash
POS household with interest
POS mobile with interest
Cash X-Sell: middle
Cash X-Sell: low
Card Street
POS industry with interest
POS household without interest
Card X-Sell
Cash Street: high
Cash X-Sell: high
Cash Street: middle
Cash Street: low
POS other with interest
POS mobile without interest
POS industry without interest
POS others without interest
not specified
PRODUCT_COMBINATION



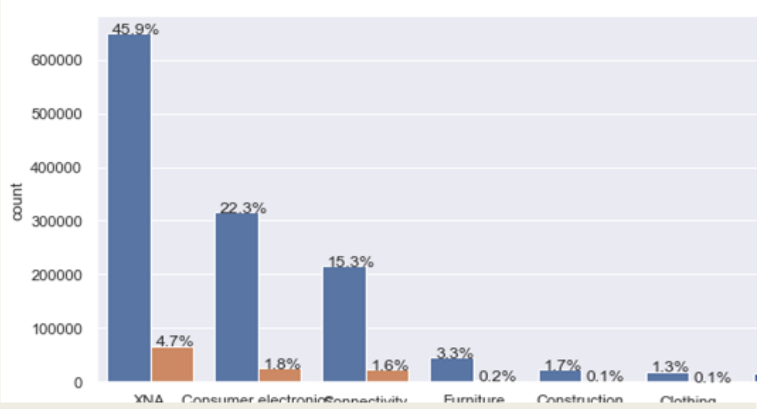
TARGET
0
1
count
600000
500000
400000
300000
200000
100000
0
40.3%
4.0%
40.8%
3.4%
10.2%
1.2%
0.0%
0.0%
Cash loans
Consumer loans
Revolving loans
XNA
NAME_CONTRACT_TYPE_PREV



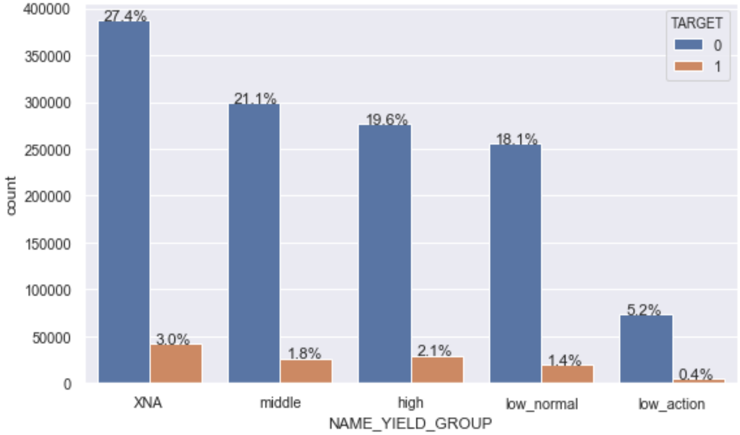
TARGET
0
1
count
800000
700000
600000
500000
400000
300000
200000
100000
0
57.9%
4.8%
16.7%
1.7%
15.3%
2.1%
1.5%
0.1%
Approved
Canceled
Refused
Unused offer
NAME_CONTRACT_STATUS



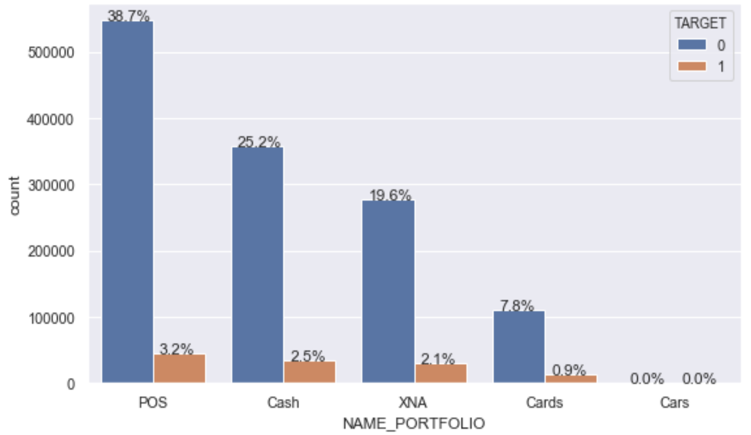
TARGET
0
1
count
800000
700000
600000
500000
400000
300000
200000
100000
0
57.4%
5.0%
33.4%
3.6%
0.5%
0.0%
0.1%
0.0%
Cash through the bank
XNA
Non-cash from your account
Cashless from the account of the employer
NAME_PAYMENT_TYPE



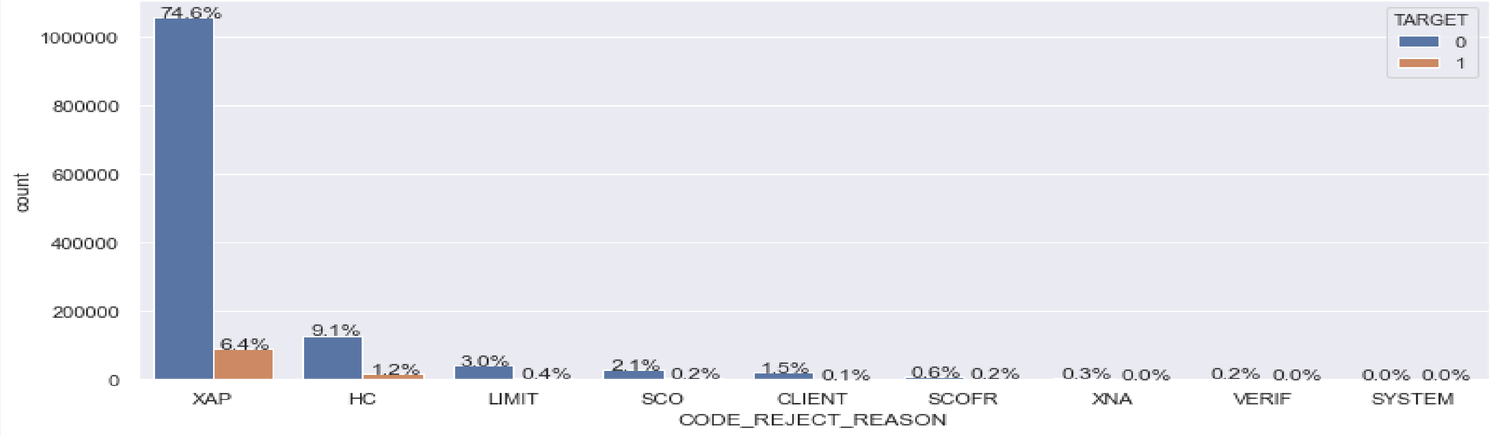
count
600000
500000
400000
300000
200000
100000
0
45.9%
4.7%
22.3%
1.8%
15.3%
1.6%
3.3%
0.2%
1.7%
0.1%
1.3%
0.1%
XNA
Consumer electronics
Connectivity
Furniture
Construction
Clothing



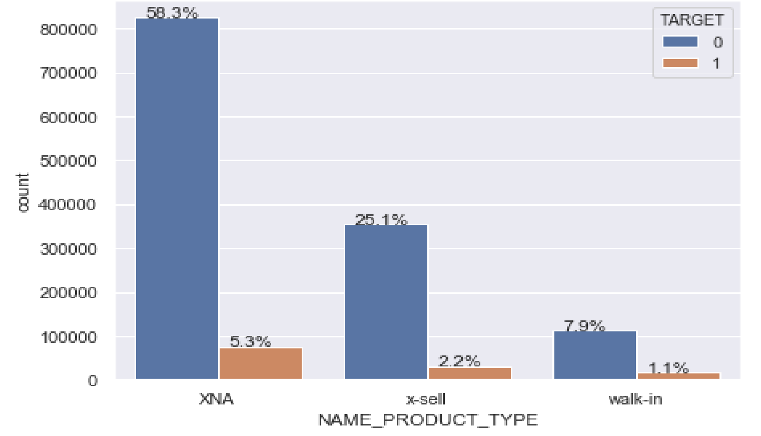
TARGET
0
1
count
400000
350000
300000
250000
200000
150000
100000
50000
0
27.4%
3.0%
21.1%
1.8%
19.6%
2.1%
18.1%
1.4%
5.2%
0.4%
XNA
middle
high
low_normal
low_action
NAME_YIELD_GROUP



TARGET
0
1
count
500000
400000
300000
200000
100000
0
38.7%
3.2%
25.2%
2.5%
19.6%
2.1%
7.8%
0.9%
0.0%
0.0%
POS
Cash
XNA
Cards
Cars
NAME_PORTFOLIO



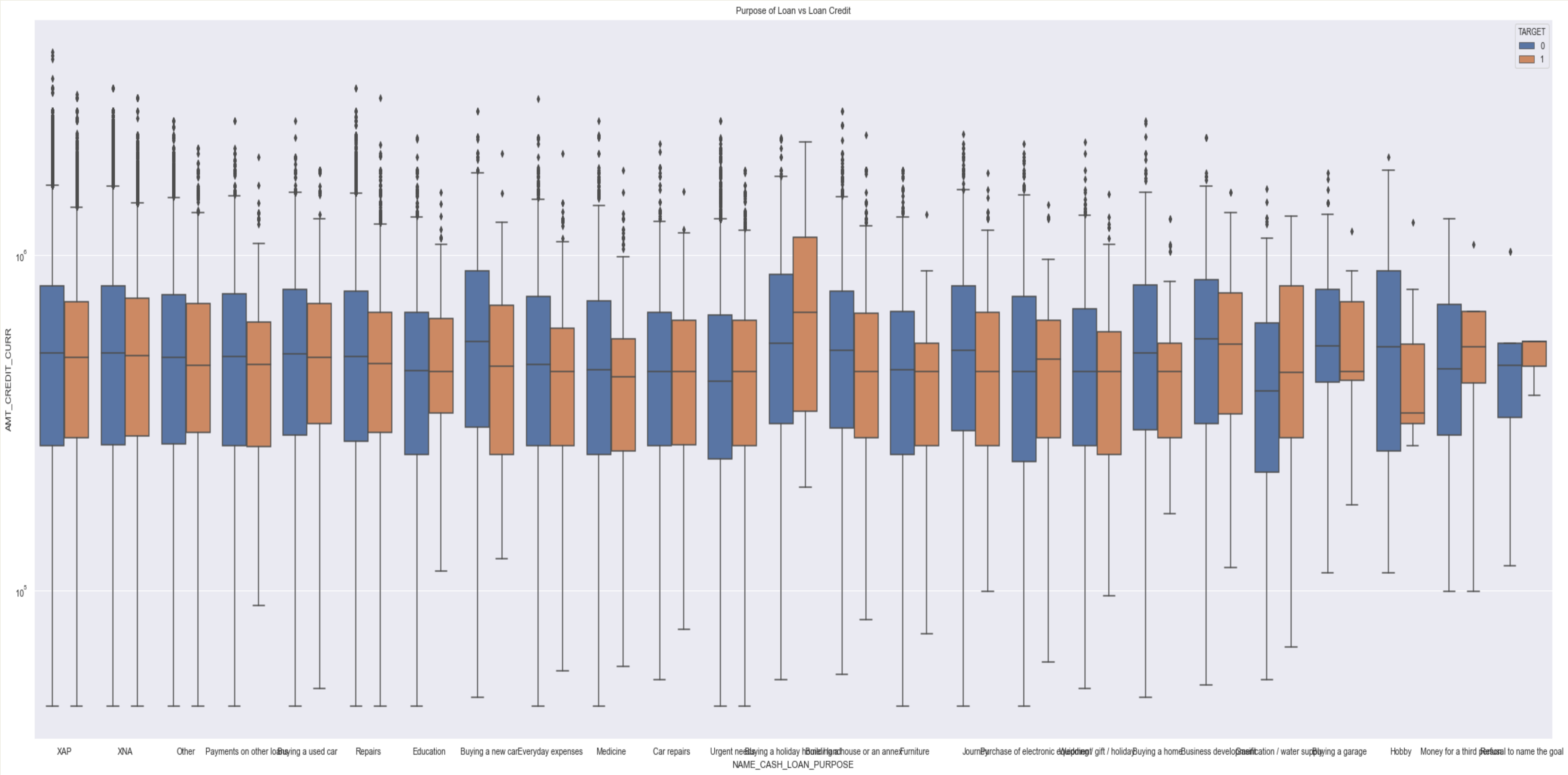
TARGET
0
1
count
1000000
800000
600000
400000
200000
0
74.6%
6.4%
9.1%
1.2%
3.0%
0.4%
2.1%
0.2%
1.5%
0.1%
0.6%
0.2%
0.3%
0.0%
0.2%
0.0%
0.0%
0.0%
XAP
HC
LIMIT
SCO
CLIENT
SCOFR
XNA
VERIF
SYSTEM
CODE_REJECT_REASON



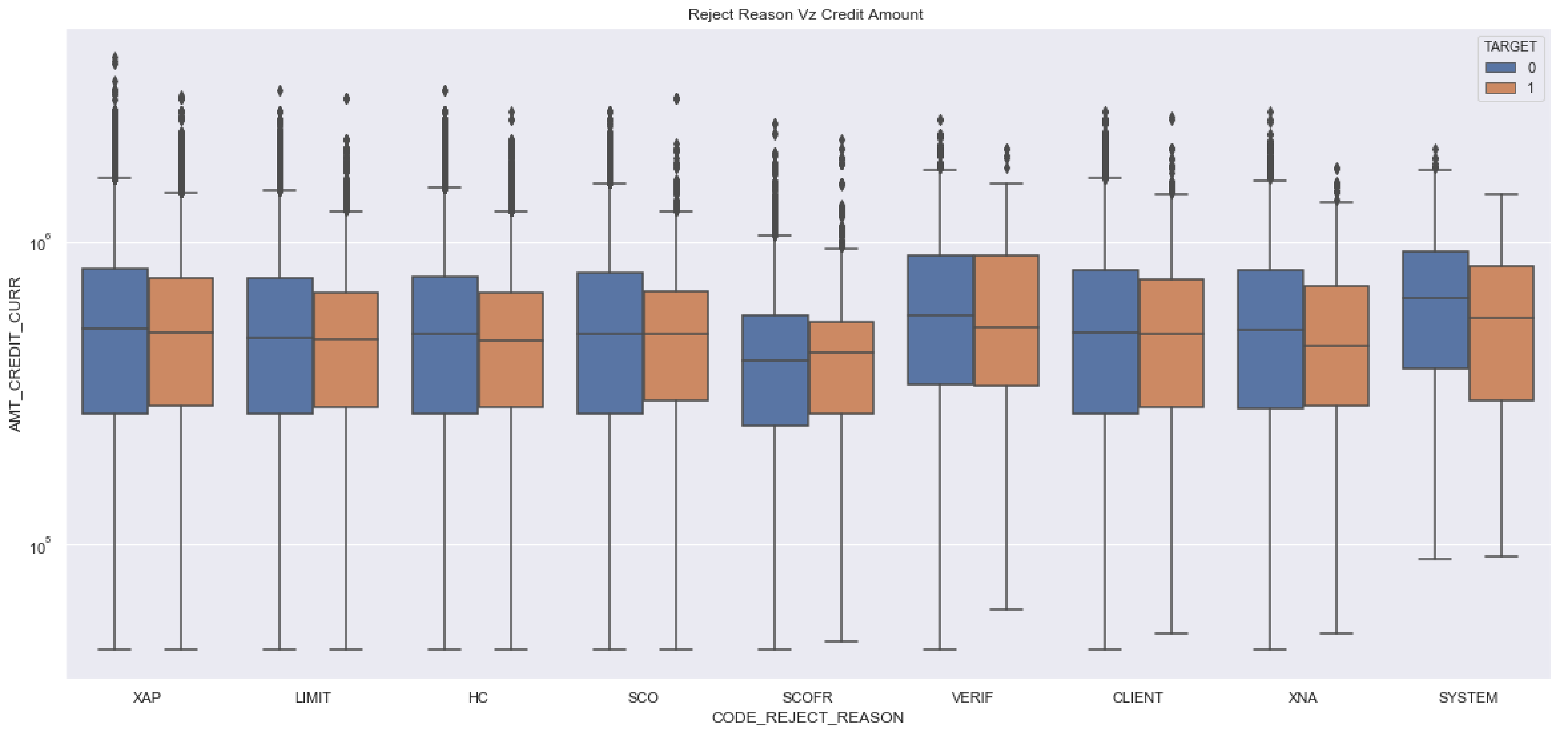
TARGET
0
1
count
800000
700000
600000
500000
400000
300000
200000
100000
0
58.3%
5.3%
25.1%
2.2%
7.9%
1.1%
XNA
x-sell
walk-in
NAME_PRODUCT_TYPE

Bivariate Analysis

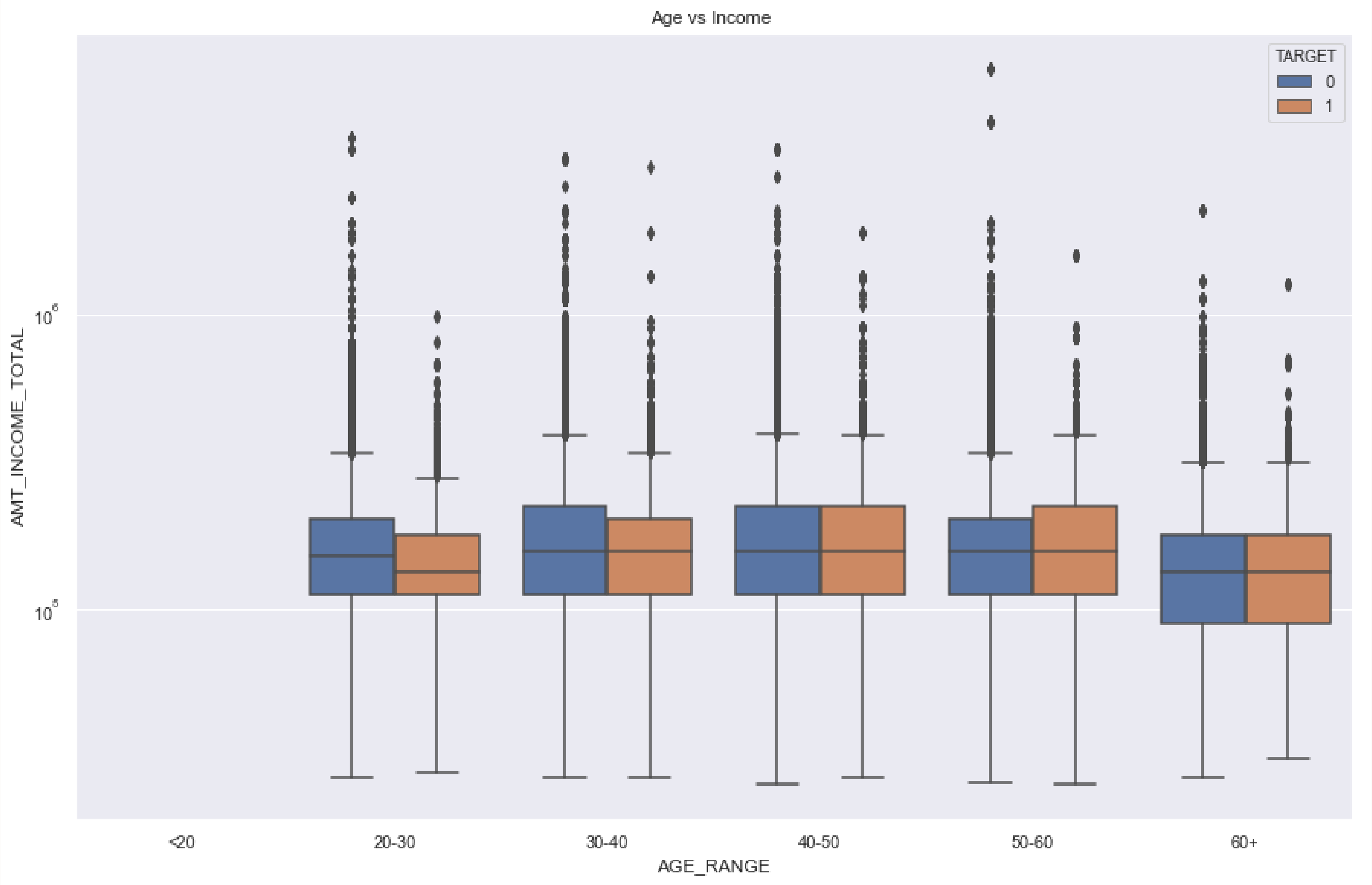
Purpose of Loan vs Loan Credit



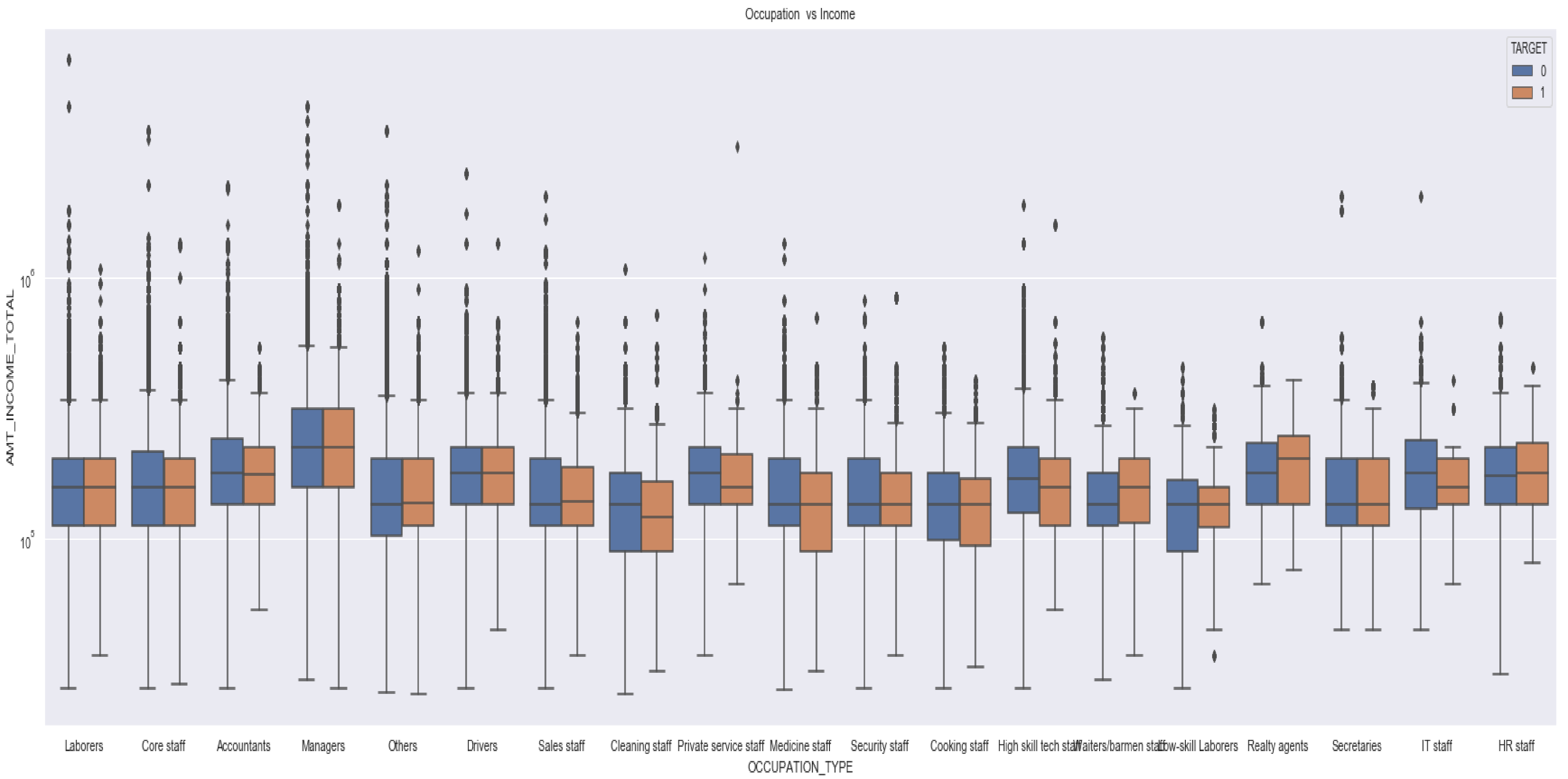
REJECT REASON VZ CREDIT AMOUNT



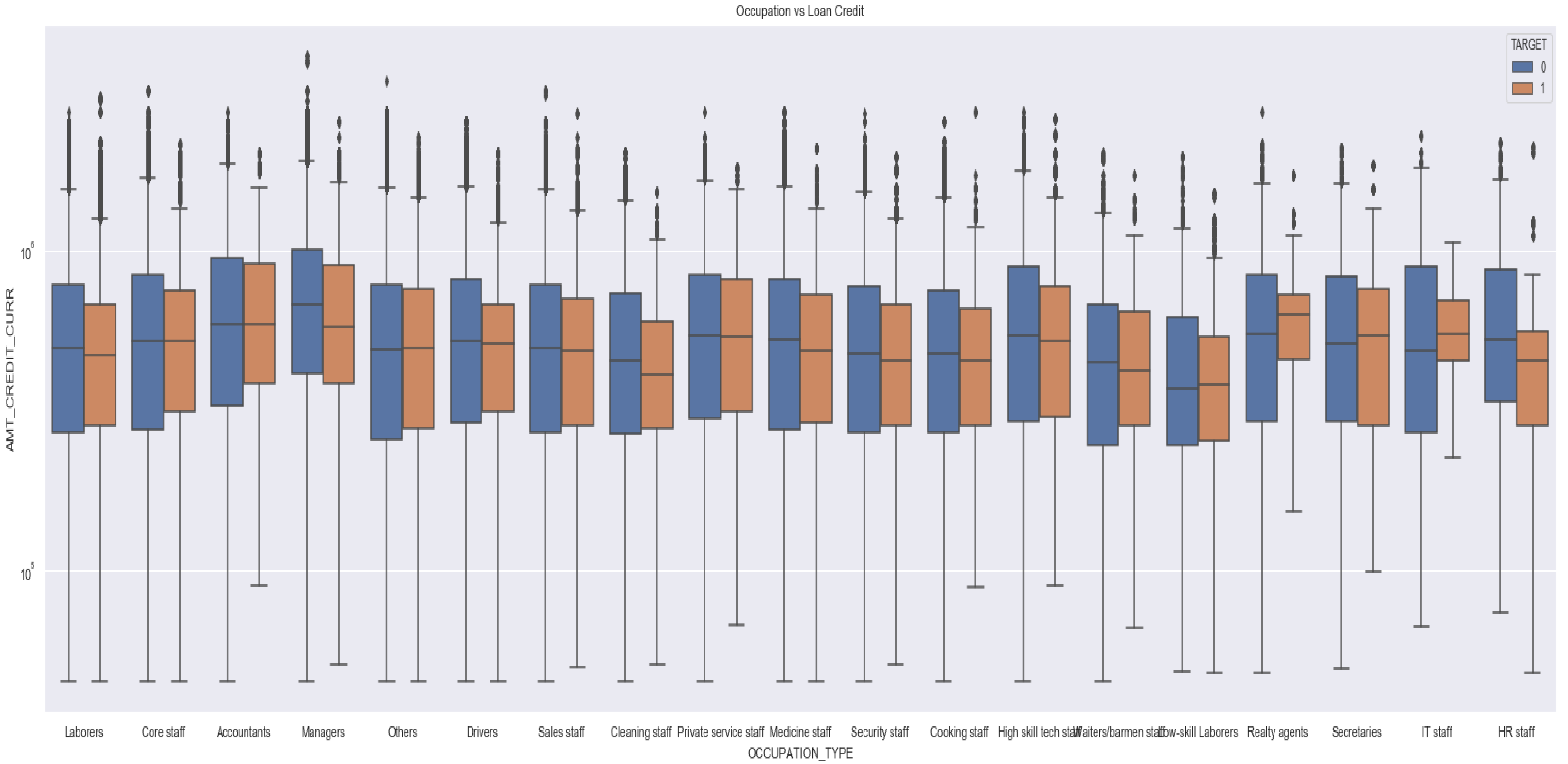
AGE VZ INCOME AMOUNT



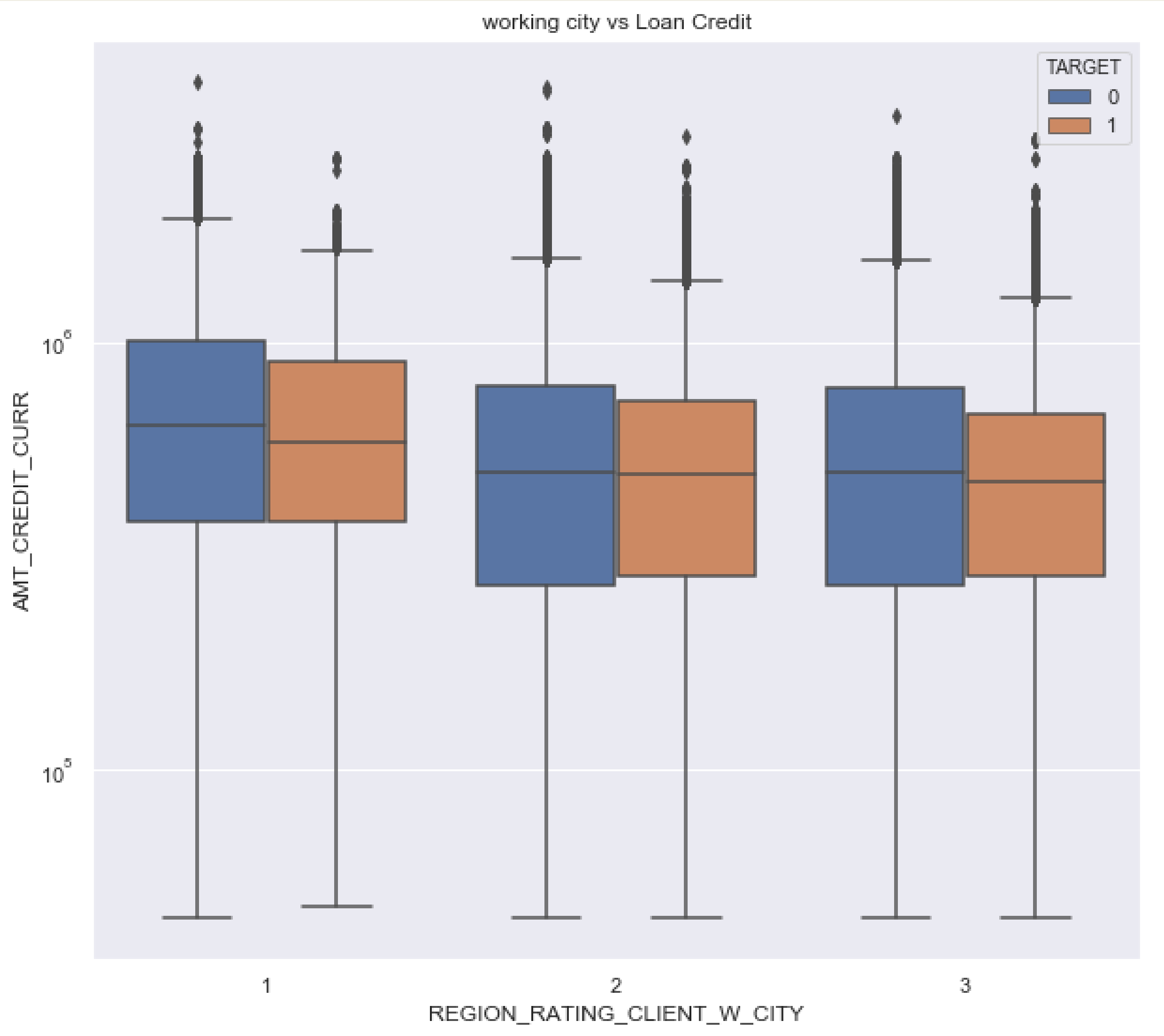
OCCUPATION VZ INCOME AMOUNT



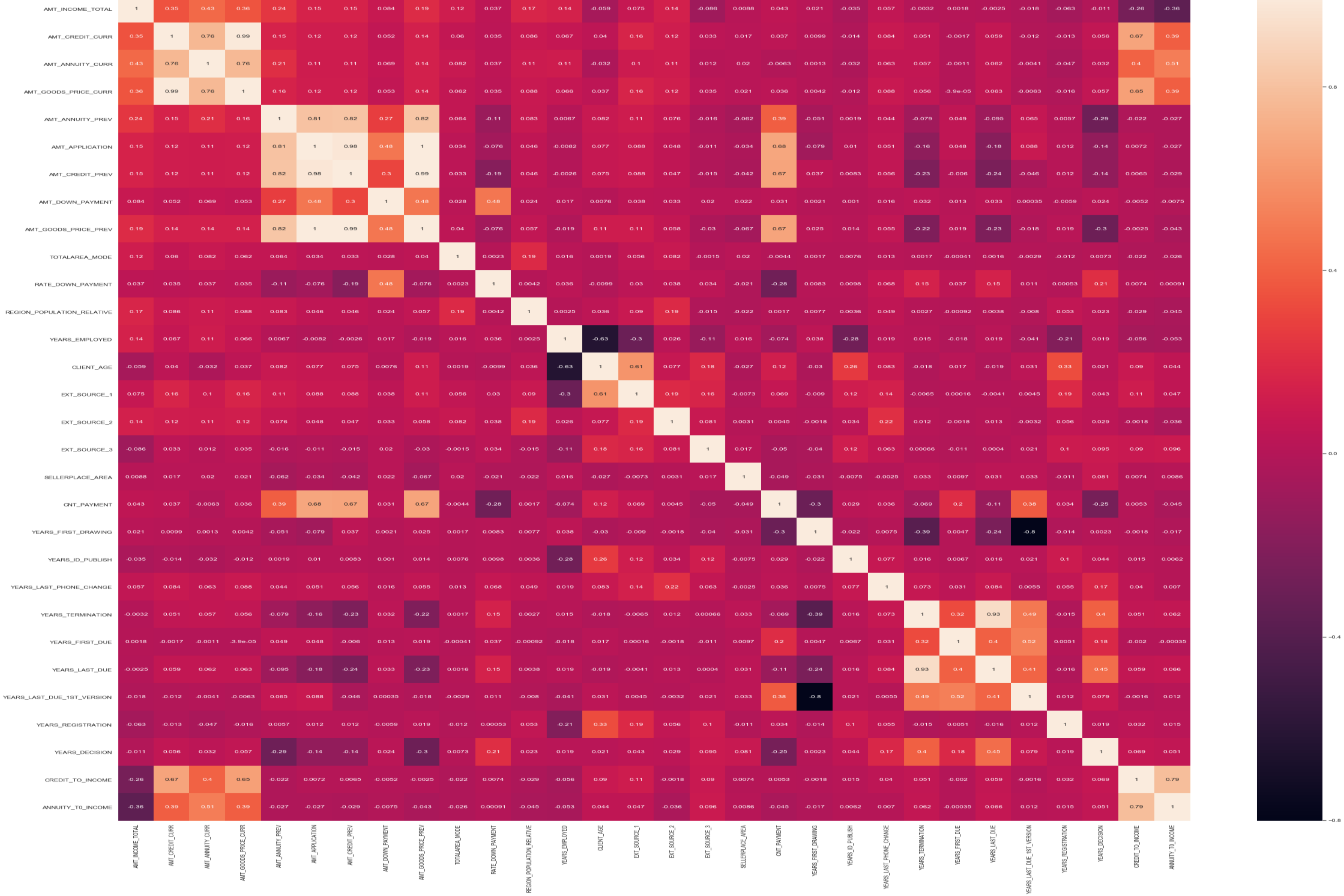
OCCUPATION VZ LOAN AMOUNT



WORKING CITY VZ LOAN AMOUNT



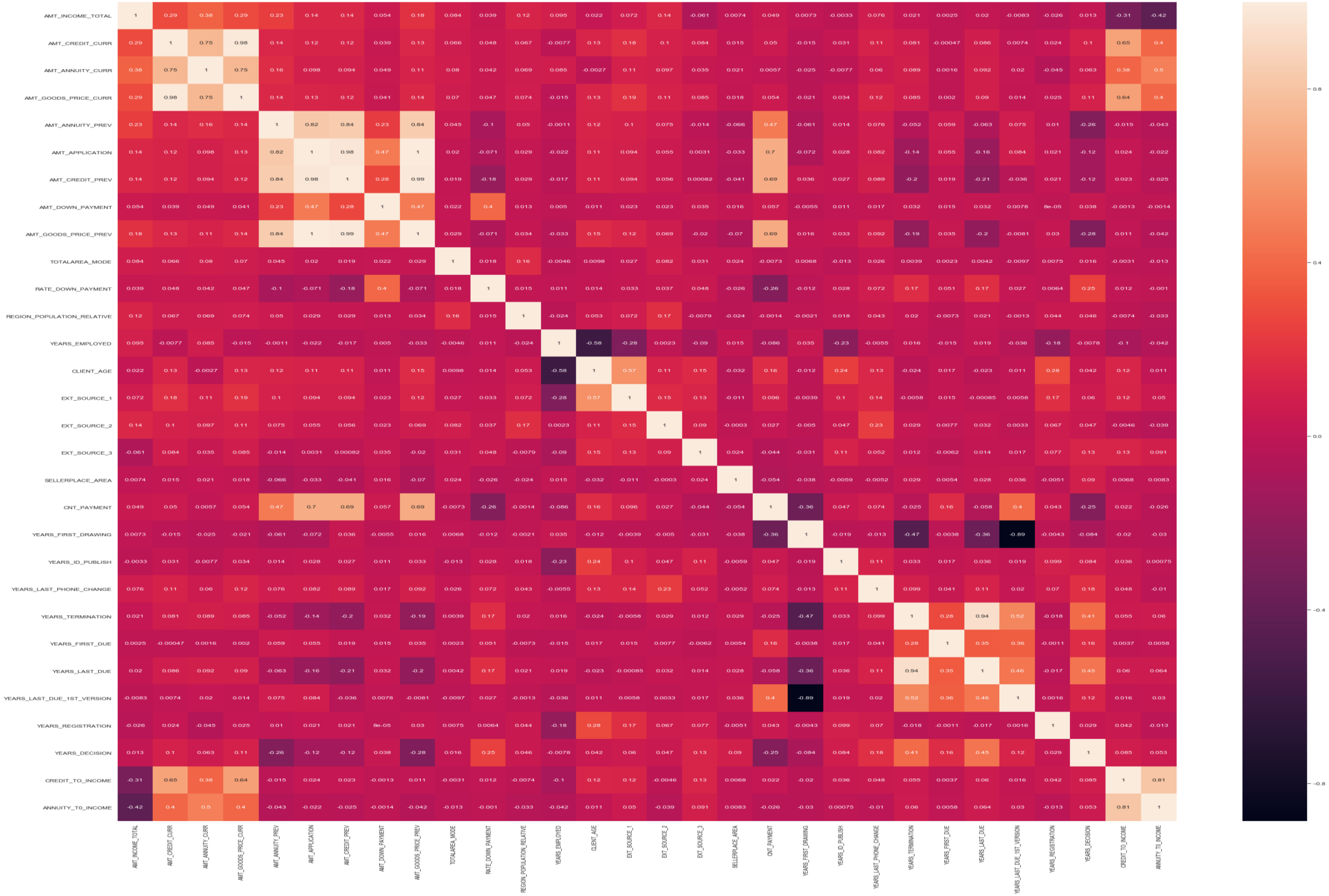
REPAYERS CORRELATION MATRIX- HEATMAP


AMT_INCOME_TOTAL 1 0.35 0.43 0.36 0.24 0.15 0.15 0.084 0.19 0.12 0.037 0.17 0.14 -0.059 0.075 0.14 -0.086 0.0088 0.043 0.021 -0.035 0.057 -0.0032 0.0018 -0.0025 -0.018 -0.063 -0.011 -0.26 -0.36
AMT_CREDIT_CURR 0.35 1 0.76 0.99 0.15 0.12 0.12 0.052 0.14 0.06 0.035 0.086 0.067 0.04 0.16 0.12 0.033 0.017 0.037 0.0099 -0.014 0.084 0.051 -0.0017 0.059 -0.012 -0.013 0.056 0.67 0.39
AMT_ANNUITY_CURR 0.43 0.76 1 0.76 0.21 0.11 0.11 0.069 0.14 0.082 0.037 0.11 0.11 -0.032 0.1 0.11 0.012 0.02 -0.0063 0.0013 -0.032 0.063 0.057 -0.0011 0.062 -0.0041 -0.047 0.032 0.4 0.51
AMT_GOODS_PRICE_CURR 0.36 0.99 0.76 1 0.16 0.12 0.12 0.053 0.14 0.062 0.035 0.088 0.066 0.037 0.16 0.12 0.035 0.021 0.036 0.0042 -0.012 0.088 0.056 -3.9e-05 0.063 -0.0063 -0.016 0.057 0.65 0.39
AMT_ANNUITY_PREV 0.24 0.15 0.21 0.16 1 0.81 0.82 0.27 0.82 0.064 -0.11 0.083 0.0067 0.082 0.11 0.076 -0.016 -0.062 0.39 -0.051 0.0019 0.044 -0.079 0.049 -0.095 0.065 0.0057 -0.29 -0.022 -0.027
AMT_APPLICATION 0.15 0.12 0.11 0.12 0.81 1 0.98 0.48 1 0.034 -0.076 0.046 -0.0082 0.077 0.088 0.048 -0.011 -0.034 0.68 -0.079 0.01 0.051 -0.16 0.048 -0.18 0.088 0.012 -0.14 0.0072 -0.027
AMT_CREDIT_PREV 0.15 0.12 0.11 0.12 0.82 0.98 1 0.3 0.99 0.033 -0.19 0.046 -0.0026 0.075 0.088 0.047 -0.015 -0.042 0.67 0.037 0.0083 0.056 -0.23 -0.006 -0.24 -0.046 0.012 -0.14 0.0065 -0.029
AMT_DOWN_PAYMENT 0.084 0.052 0.069 0.053 0.27 0.48 0.3 1 0.48 0.028 0.48 0.024 0.017 0.0076 0.038 0.033 0.02 0.022 0.031 0.0021 0.001 0.016 0.032 0.013 0.033 0.00035 -0.0059 0.024 -0.0052 -0.0075
AMT_GOODS_PRICE_PREV 0.19 0.14 0.14 0.14 0.82 1 0.99 0.48 1 0.04 -0.076 0.057 -0.019 0.11 0.11 0.058 -0.03 -0.067 0.67 0.025 0.014 0.055 -0.22 0.019 -0.23 -0.018 0.019 -0.3 -0.0025 -0.043
TOTALAREA_MODE 0.12 0.06 0.082 0.062 0.064 0.034 0.033 0.028 0.04 1 0.0023 0.19 0.016 0.0019 0.056 0.082 -0.0015 0.02 -0.0044 0.0017 0.0076 0.013 0.0017 -0.00041 0.0016 -0.0029 -0.012 0.0073 -0.022 -0.026
RATE_DOWN_PAYMENT 0.037 0.035 0.037 0.035 -0.11 -0.076 -0.19 0.48 -0.076 0.0023 1 0.0042 0.036 -0.0099 0.03 0.038 0.034 -0.021 -0.28 0.0083 0.0098 0.068 0.15 0.037 0.15 0.011 0.00053 0.21 0.0074 0.00091
REGION_POPULATION_RELATIVE 0.17 0.086 0.11 0.088 0.083 0.046 0.046 0.024 0.057 0.19 0.0042 1 0.0025 0.036 0.09 0.19 -0.015 -0.022 0.0017 0.0077 0.0036 0.049 0.0027 -0.00092 0.0038 -0.008 0.053 0.023 -0.029 -0.045
YEARS_EMPLOYED 0.14 0.067 0.11 0.066 0.0067 -0.0082 -0.0026 0.017 -0.019 0.016 0.036 0.0025 1 -0.63 -0.3 0.026 -0.11 0.016 -0.074 0.038 -0.28 0.019 0.015 -0.018 0.019 -0.041 -0.21 0.019 -0.056 -0.053
CLIENT_AGE -0.059 0.04 -0.032 0.037 0.082 0.077 0.075 0.0076 0.11 0.0019 -0.0099 0.036 -0.63 1 0.61 0.077 0.18 -0.027 0.12 -0.03 0.26 0.083 -0.018 0.017 -0.019 0.031 0.33 0.021 0.09 0.044
EXT_SOURCE_1 0.075 0.16 0.1 0.16 0.11 0.088 0.088 0.038 0.11 0.056 0.03 0.09 -0.3 0.61 1 0.19 0.16 -0.0073 0.069 -0.009 0.12 0.14 -0.0065 0.00016 -0.0041 0.0045 0.19 0.043 0.11 0.047
EXT_SOURCE_2 0.14 0.12 0.11 0.12 0.076 0.048 0.047 0.033 0.058 0.082 0.038 0.19 0.026 0.077 0.19 1 0.081 0.0031 0.0045 -0.0018 0.034 0.22 0.012 -0.0018 0.013 -0.0032 0.056 0.029 -0.0018 -0.036
EXT_SOURCE_3 -0.086 0.033 0.012 0.035 -0.016 -0.011 -0.015 0.02 -0.03 -0.0015 0.034 -0.015 -0.11 0.18 0.16 0.081 1 0.017 -0.05 -0.04 0.12 0.063 0.00066 -0.011 0.0004 0.021 0.1 0.095 0.09 0.096
SELLERPLACE_AREA 0.0088 0.017 0.02 0.021 -0.062 -0.034 -0.042 0.022 -0.067 0.02 -0.021 -0.022 0.016 -0.027 -0.0073 0.0031 0.017 1 -0.049 -0.031 -0.0075 -0.0025 0.033 0.0097 0.031 0.033 -0.011 0.081 0.0074 0.0086
CNT_PAYMENT 0.043 0.037 -0.0063 0.036 0.39 0.68 0.67 0.031 0.67 -0.0044 -0.28 0.0017 -0.074 0.12 0.069 0.0045 -0.05 -0.049 1 -0.3 0.029 0.036 -0.069 0.2 -0.11 0.38 0.034 -0.25 0.0053 -0.045
YEARS_FIRST_DRAWING 0.021 0.0099 0.0013 0.0042 -0.051 -0.079 0.037 0.0021 0.025 0.0017 0.0083 0.0077 0.038 -0.03 -0.009 -0.0018 -0.04 -0.031 -0.3 1 -0.022 0.0075 -0.39 0.0047 -0.24 -0.8 -0.014 0.0023 -0.0018 -0.017
YEARS_ID_PUBLISH -0.035 -0.014 -0.032 -0.012 0.0019 0.01 0.0083 0.001 0.014 0.0076 0.0098 0.0036 -0.28 0.26 0.12 0.034 0.12 -0.0075 0.029 -0.022 1 0.077 0.016 0.0067 0.016 0.021 0.1 0.044 0.015 0.0062
YEARS_LAST_PHONE_CHANGE 0.057 0.084 0.063 0.088 0.044 0.051 0.056 0.016 0.055 0.013 0.068 0.049 0.019 0.083 0.14 0.22 0.063 -0.0025 0.036 0.0075 0.077 1 0.073 0.031 0.084 0.0055 0.055 0.17 0.04 0.007
YEARS_TERMINATION -0.0032 0.051 0.057 0.056 -0.079 -0.16 -0.23 0.032 -0.22 0.0017 0.15 0.0027 0.015 -0.018 -0.0065 0.012 0.00066 0.033 -0.069 -0.39 0.016 0.073 1 0.32 0.93 0.49 -0.015 0.4 0.051 0.062
YEARS_FIRST_DUE 0.0018 -0.0017 -0.0011 -3.9e-05 0.049 0.048 -0.006 0.013 0.019 -0.00041 0.037 -0.00092 -0.018 0.017 0.00016 -0.0018 -0.011 0.0097 0.2 0.0047 0.0067 0.031 0.32 1 0.4 0.52 0.0051 0.18 -0.002 -0.00035
YEARS_LAST_DUE -0.0025 0.059 0.062 0.063 -0.095 -0.18 -0.24 0.033 -0.23 0.0016 0.15 0.0038 0.019 -0.019 -0.0041 0.013 0.0004 0.031 -0.11 -0.24 0.016 0.084 0.93 0.4 1 0.41 -0.016 0.45 0.059 0.066
YEARS_LAST_DUE_1ST_VERSION -0.018 -0.012 -0.0041 -0.0063 0.065 0.088 -0.046 0.00035 -0.018 -0.0029 0.011 -0.008 -0.041 0.031 0.0045 -0.0032 0.021 0.033 0.38 -0.8 0.021 0.0055 0.49 0.52 0.41 1 0.012 0.079 -0.0016 0.012
YEARS_REGISTRATION -0.063 -0.013 -0.047 -0.016 0.0057 0.012 0.012 -0.0059 0.019 -0.012 0.00053 0.053 -0.21 0.33 0.19 0.056 0.1 -0.011 0.034 -0.014 0.1 0.055 -0.015 0.0051 -0.016 0.012 1 0.019 0.032 0.015
YEARS_DECISION -0.011 0.056 0.032 0.057 -0.29 -0.14 -0.14 0.024 -0.3 0.0073 0.21 0.023 0.019 0.021 0.043 0.029 0.095 0.081 -0.25 0.0023 0.044 0.17 0.4 0.18 0.45 0.079 0.019 1 0.069 0.051
CREDIT_TO_INCOME -0.26 0.67 0.4 0.65 -0.022 0.0072 0.0065 -0.0052 -0.0025 -0.022 0.0074 -0.029 -0.056 0.09 0.11 -0.0018 0.09 0.0074 0.0053 -0.0018 0.015 0.04 0.051 -0.002 0.059 -0.0016 0.032 0.069 1 0.79
ANNUITY_T0_INCOME -0.36 0.39 0.51 0.39 -0.027 -0.027 -0.029 -0.0075 -0.043 -0.026 0.00091 -0.045 -0.053 0.044 0.047 -0.036 0.096 0.0086 -0.045 -0.017 0.0062 0.007 0.062 -0.00035 0.066 0.012 0.015 0.051 0.79 1
AMT_INCOME_TOTAL
AMT_CREDIT_CURR
AMT_ANNUITY_CURR
AMT_GOODS_PRICE_CURR
AMT_ANNUITY_PREV
AMT_APPLICATION
AMT_CREDIT_PREV
AMT_DOWN_PAYMENT
AMT_GOODS_PRICE_PREV
TOTALAREA_MODE
RATE_DOWN_PAYMENT
REGION_POPULATION_RELATIVE
YEARS_EMPLOYED
CLIENT_AGE
EXT_SOURCE_1
EXT_SOURCE_2
EXT_SOURCE_3
SELLERPLACE_AREA
CNT_PAYMENT
YEARS_FIRST_DRAWING
YEARS_ID_PUBLISH
YEARS_LAST_PHONE_CHANGE
YEARS_TERMINATION
YEARS_FIRST_DUE
YEARS_LAST_DUE
YEARS_LAST_DUE_1ST_VERSION
YEARS_REGISTRATION
YEARS_DECISION
CREDIT_TO_INCOME
ANNUITY_T0_INCOME
0.8
0.4
0.0
-0.4
-0.8

TOP 10 correlation for REPAYERS

1	AMT_CREDIT_CURR	AMT_GOODS_PRICE_CURR	0.9998
2	AMT_APPLICATION	AMT_GOODS_PRICE_PREV	0.9926
3	YEARS_TERMINATION	YEARS_LAST_DUE	0.9902
4	AMT_CREDIT_PREV	AMT_GOODS_PRICE_PREV	0.9902
5	AMT_APPLICATION	AMT_CREDIT_PREV	0.989
6	AMT_ANNUITY_PREV	AMT_GOODS_PRICE_PREV	0.964
7	AMT_ANNUITY_PREV	AMT_CREDIT_PREV	0.9584
8	AMT_ANNUITY_PREV	AMT_APPLICATION	0.9583
9	CREDIT_TO_INCOME	ANNUITY_T0_INCOME	0.9449
10	AMT_ANNUITY_CURR	AMT_GOODS_PRICE_CURR	0.944

DEFAULTERS CORRELATION MATRIX-HEATMAP


AMT_INCOME_TOTAL 1 0.29 0.38 0.29 0.23 0.14 0.14 0.054 0.18 0.084 0.039 0.12 0.095 0.022 0.072 0.14 -0.061 0.0074 0.049 0.0073 -0.0033 0.076 0.021 0.0025 0.02 -0.0083 -0.026 0.013 -0.31 -0.42
AMT_CREDIT_CURR 0.29 1 0.75 0.98 0.14 0.12 0.12 0.039 0.13 0.066 0.048 0.067 -0.0077 0.13 0.18 0.1 0.084 0.015 0.05 -0.015 0.031 0.11 0.081 -0.00047 0.086 0.0074 0.024 0.1 0.65 0.4
AMT_ANNUITY_CURR 0.38 0.75 1 0.75 0.16 0.098 0.094 0.049 0.11 0.08 0.042 0.069 0.085 -0.0027 0.11 0.097 0.035 0.021 0.0057 -0.025 -0.0077 0.06 0.089 0.0016 0.092 0.02 -0.045 0.063 0.38 0.5
AMT_GOODS_PRICE_CURR 0.29 0.98 0.75 1 0.14 0.13 0.12 0.041 0.14 0.07 0.047 0.074 -0.015 0.13 0.19 0.11 0.085 0.018 0.054 -0.021 0.034 0.12 0.085 0.002 0.09 0.014 0.025 0.11 0.64 0.4
AMT_ANNUITY_PREV 0.23 0.14 0.16 0.14 1 0.82 0.84 0.23 0.84 0.045 -0.1 0.05 -0.0011 0.12 0.1 0.075 -0.014 -0.066 0.47 -0.061 0.014 0.076 -0.052 0.059 -0.063 0.075 0.01 -0.26 -0.015 -0.043
AMT_APPLICATION 0.14 0.12 0.098 0.13 0.82 1 0.98 0.47 1 0.02 -0.071 0.029 -0.022 0.11 0.094 0.055 0.0031 -0.033 0.7 -0.072 0.028 0.082 -0.14 0.055 -0.16 0.084 0.021 -0.12 0.024 -0.022
AMT_CREDIT_PREV 0.14 0.12 0.094 0.12 0.84 0.98 1 0.28 0.99 0.019 -0.18 0.029 -0.017 0.11 0.094 0.056 0.00082 -0.041 0.69 0.036 0.027 0.089 -0.2 0.019 -0.21 -0.036 0.021 -0.12 0.023 -0.025
AMT_DOWN_PAYMENT 0.054 0.039 0.049 0.041 0.23 0.47 0.28 1 0.47 0.022 0.4 0.013 0.005 0.011 0.023 0.023 0.035 0.016 0.057 -0.0055 0.011 0.017 0.032 0.015 0.032 0.0078 8e-05 0.038 -0.0013 -0.0014
AMT_GOODS_PRICE_PREV 0.18 0.13 0.11 0.14 0.84 1 0.99 0.47 1 0.029 -0.071 0.034 -0.033 0.15 0.12 0.069 -0.02 -0.07 0.69 0.016 0.033 0.092 -0.19 0.035 -0.2 -0.0081 0.03 -0.28 0.011 -0.042
TOTALAREA_MODE 0.084 0.066 0.08 0.07 0.045 0.02 0.019 0.022 0.029 1 0.018 0.16 -0.0046 0.0098 0.027 0.082 0.031 0.024 -0.0073 0.0068 -0.013 0.026 0.0039 0.0023 0.0042 -0.0097 0.0075 0.016 -0.0031 -0.013
RATE_DOWN_PAYMENT 0.039 0.048 0.042 0.047 -0.1 -0.071 -0.18 0.4 -0.071 0.018 1 0.015 0.011 0.014 0.033 0.037 0.048 -0.026 -0.26 -0.012 0.028 0.072 0.17 0.051 0.17 0.027 0.0064 0.25 0.012 -0.001
REGION_POPULATION_RELATIVE 0.12 0.067 0.069 0.074 0.05 0.029 0.029 0.013 0.034 0.16 0.015 1 -0.024 0.053 0.072 0.17 -0.0079 -0.024 -0.0014 -0.0021 0.018 0.043 0.02 -0.0073 0.021 -0.0013 0.044 0.046 -0.0074 -0.033
YEARS_EMPLOYED 0.095 -0.0077 0.085 -0.015 -0.0011 -0.022 -0.017 0.005 -0.033 -0.0046 0.011 -0.024 1 -0.58 -0.28 0.0023 -0.09 0.015 -0.086 0.035 -0.23 -0.0055 0.016 -0.015 0.019 -0.036 -0.18 -0.0078 -0.1 -0.042
CLIENT_AGE 0.022 0.13 -0.0027 0.13 0.12 0.11 0.11 0.011 0.15 0.0098 0.014 0.053 -0.58 1 0.57 0.11 0.15 -0.032 0.16 -0.012 0.24 0.13 -0.024 0.017 -0.023 0.011 0.28 0.042 0.12 0.011
EXT_SOURCE_1 0.072 0.18 0.11 0.19 0.1 0.094 0.094 0.023 0.12 0.027 0.033 0.072 -0.28 0.57 1 0.15 0.13 -0.011 0.096 -0.0039 0.1 0.14 -0.0058 0.015 -0.00085 0.0058 0.17 0.06 0.12 0.05
EXT_SOURCE_2 0.14 0.1 0.097 0.11 0.075 0.055 0.056 0.023 0.069 0.082 0.037 0.17 0.0023 0.11 0.15 1 0.09 -0.0003 0.027 -0.005 0.047 0.23 0.029 0.0077 0.032 0.0033 0.067 0.047 -0.0046 -0.039
EXT_SOURCE_3 -0.061 0.084 0.035 0.085 -0.014 0.0031 0.00082 0.035 -0.02 0.031 0.048 -0.0079 -0.09 0.15 0.13 0.09 1 0.024 -0.044 -0.031 0.11 0.052 0.012 -0.0062 0.014 0.017 0.077 0.13 0.13 0.091
SELLERPLACE_AREA 0.0074 0.015 0.021 0.018 -0.066 -0.033 -0.041 0.016 -0.07 0.024 -0.026 -0.024 0.015 -0.032 -0.011 -0.0003 0.024 1 -0.054 -0.038 -0.0059 -0.0052 0.029 0.0054 0.028 0.036 -0.0051 0.09 0.0068 0.0083
CNT_PAYMENT 0.049 0.05 0.0057 0.054 0.47 0.7 0.69 0.057 0.69 -0.0073 -0.26 -0.0014 -0.086 0.16 0.096 0.027 -0.044 -0.054 1 -0.36 0.047 0.074 -0.025 0.16 -0.058 0.4 0.043 -0.25 0.022 -0.026
YEARS_FIRST_DRAWING 0.0073 -0.015 -0.025 -0.021 -0.061 -0.072 0.036 -0.0055 0.016 0.0068 -0.012 -0.0021 0.035 -0.012 -0.0039 -0.005 -0.031 -0.038 -0.36 1 -0.019 -0.013 -0.47 -0.0038 -0.36 -0.89 -0.0043 -0.084 -0.02 -0.03
YEARS_ID_PUBLISH -0.0033 0.031 -0.0077 0.034 0.014 0.028 0.027 0.011 0.033 -0.013 0.028 0.018 -0.23 0.24 0.1 0.047 0.11 -0.0059 0.047 -0.019 1 0.11 0.033 0.017 0.036 0.019 0.099 0.084 0.036 0.00075
YEARS_LAST_PHONE_CHANGE 0.076 0.11 0.06 0.12 0.076 0.082 0.089 0.017 0.092 0.026 0.072 0.043 -0.0055 0.13 0.14 0.23 0.052 -0.0052 0.074 -0.013 0.11 1 0.099 0.041 0.11 0.02 0.07 0.18 0.048 -0.01
YEARS_TERMINATION 0.021 0.081 0.089 0.085 -0.052 -0.14 -0.2 0.032 -0.19 0.0039 0.17 0.02 0.016 -0.024 -0.0058 0.029 0.012 0.029 -0.025 -0.47 0.033 0.099 1 0.28 0.94 0.52 -0.018 0.41 0.055 0.06
YEARS_FIRST_DUE 0.0025 -0.00047 0.0016 0.002 0.059 0.055 0.019 0.015 0.035 0.0023 0.051 -0.0073 -0.015 0.017 0.015 0.0077 -0.0062 0.0054 0.16 -0.0038 0.017 0.041 0.28 1 0.35 0.36 -0.0011 0.16 0.0037 0.0058
YEARS_LAST_DUE 0.02 0.086 0.092 0.09 -0.063 -0.16 -0.21 0.032 -0.2 0.0042 0.17 0.021 0.019 -0.023 -0.00085 0.032 0.014 0.028 -0.058 -0.36 0.036 0.11 0.94 0.35 1 0.46 -0.017 0.45 0.06 0.064
YEARS_LAST_DUE_1ST_VERSION -0.0083 0.0074 0.02 0.014 0.075 0.084 -0.036 0.0078 -0.0081 -0.0097 0.027 -0.0013 -0.036 0.011 0.0058 0.0033 0.017 0.036 0.4 -0.89 0.019 0.02 0.52 0.36 0.46 1 0.0016 0.12 0.016 0.03
YEARS_REGISTRATION -0.026 0.024 -0.045 0.025 0.01 0.021 0.021 8e-05 0.03 0.0075 0.0064 0.044 -0.18 0.28 0.17 0.067 0.077 -0.0051 0.043 -0.0043 0.099 0.07 -0.018 -0.0011 -0.017 0.0016 1 0.029 0.042 -0.013
YEARS_DECISION 0.013 0.1 0.063 0.11 -0.26 -0.12 -0.12 0.038 -0.28 0.016 0.25 0.046 -0.0078 0.042 0.06 0.047 0.13 0.09 -0.25 -0.084 0.084 0.18 0.41 0.16 0.45 0.12 0.029 1 0.085 0.053
CREDIT_TO_INCOME -0.31 0.65 0.38 0.64 -0.015 0.024 0.023 -0.0013 0.011 -0.0031 0.012 -0.0074 -0.1 0.12 0.12 -0.0046 0.13 0.0068 0.022 -0.02 0.036 0.048 0.055 0.0037 0.06 0.016 0.042 0.085 1 0.81
ANNUITY_TO_INCOME -0.42 0.4 0.5 0.4 -0.043 -0.022 -0.025 -0.0014 -0.042 -0.013 -0.001 -0.033 -0.042 0.011 0.05 -0.039 0.091 0.0083 -0.026 -0.03 0.00075 -0.01 0.06 0.0058 0.064 0.03 -0.013 0.053 0.81 1
AMT_INCOME_TOTAL
AMT_CREDIT_CURR
AMT_ANNUITY_CURR
AMT_GOODS_PRICE_CURR
AMT_ANNUITY_PREV
AMT_APPLICATION
AMT_CREDIT_PREV
AMT_DOWN_PAYMENT
AMT_GOODS_PRICE_PREV
TOTALAREA_MODE
RATE_DOWN_PAYMENT
REGION_POPULATION_RELATIVE
YEARS_EMPLOYED
CLIENT_AGE
EXT_SOURCE_1
EXT_SOURCE_2
EXT_SOURCE_3
SELLERPLACE_AREA
CNT_PAYMENT
YEARS_FIRST_DRAWING
YEARS_ID_PUBLISH
YEARS_LAST_PHONE_CHANGE
YEARS_TERMINATION
YEARS_FIRST_DUE
YEARS_LAST_DUE
YEARS_LAST_DUE_1ST_VERSION
YEARS_REGISTRATION
YEARS_DECISION
CREDIT_TO_INCOME
ANNUITY_TO_INCOME
0.8
0.4
0.0
−0.4
−0.8

TOP 10 correlation for DEFAULTERS

1	AMT_CREDIT_CURR	AMT_GOODS_PRICE_CURR	0.9998
2	YEARS_TERMINATION	YEARS_LAST_DUE	0.9942
3	AMT_APPLICATION	AMT_GOODS_PRICE_PREV	0.993
4	AMT_APPLICATION	AMT_CREDIT_PREV	0.989
5	AMT_CREDIT_PREV	AMT_GOODS_PRICE_PREV	0.989
6	AMT_ANNUITY_PREV	AMT_GOODS_PRICE_PREV	0.969
7	YEARS_FIRST_DRAWING	YEARS_LAST_DUE_1ST_VERSION	0.9688
8	AMT_ANNUITY_PREV	AMT_CREDIT_PREV	0.9665
9	AMT_ANNUITY_PREV	AMT_APPLICATION	0.9646
10	CREDIT_TO_INCOME	ANNUITY_T0_INCOME	0.9508

Business Insights

- Majority Loanees are more in Region_Rating_Client=2
- Majority Loanees are more in Region_Rating_Client_W_City=2
- No one loanee without mobile number
- Majority Loanees both repayers and defaulters are married and working with secondary education
- Majority of the loanee are employed within 2-5 years
- Majority of loanees are in the age range 30-40
- Defaulters are very less in the age range of 60+
- Majority of loanees in the range of income 1LA-2LA
- Majority of loanees are of organization type -Busniess Entity
- Majority of loanees are getting credit range 4LA-10LA and Annuity range 20k-30K
- Majority of loanees have current good price in the range 2la-6la
- Majority of loanees application amount is between 50k-4LA
- Majority of loanees Down payment in the range 2,5k-10K

- Majority of the repayers last phone change is 5+ years before application
- Majority of the defaulters last phone change is 1-2 years before application
- Defaulters are more in Region_Rating_Client_W_city=2
- Defaulters occupation is majorly Others and Laborers
- Defaulters are very less in the age range of 60+
- Deafulting high in Cash Loans and Approved (Contract Status),and cash through bank payment gtype
- Defaulting tendency more in Client type : Repeaters
- Defaulers are very less in IT SECTOR
- Defaulters are more on high yield group than middle and lower
- Defaulters are more in Loan Credited for the purpose: Buying a Holiday Home
- Loan with SCOFR reject reason defaulted less
- Main code reject reason is XAP then, HC

THANK YOU