EDA CASE STUDY
Credit Risk Analysis

By Abhishek Kunte - DS C52 Jan 2023

Problem Statement

- There are some risk associated with loan request. These are as follows:

  1) If the applicant is not likely to repay the loan that means he/she is likely to default, then approving the loan may lead to a financial loss for the bank.

  2) If the applicant is likely to repay the loan, then not approving the loan application results in loss of revenue to the company.

To find out solution to overcome this, we have analysed the data in Jupyter Notebook.

Methodology

- After loading the data sets, first we cleaned it for better analysis.
- Found out Out co-relation and imbalance between data.
- Did univariate and bivariate analysis on each dataset independently.
- After that merged the data set.
- Repeated univariate and bivariate analysis on merged dataset.

Dropping the columns which having more missing values

In [10]:
```
##get the columns having missing value percentage more then 22 to a list
missing_percentage_pre_app_col=missing_percentage_pre_app[missing_percentage_pre_app>=22]
missing_percentage_pre_app_col
```

Out[10]:
```
AMT_ANNUITY                 22.286665
AMT_DOWN_PAYMENT            53.636480
AMT_GOODS_PRICE             23.081773
RATE_DOWN_PAYMENT           53.636480
RATE_INTEREST_PRIMARY       99.643698
RATE_INTEREST_PRIVILEGED    99.643698
NAME_TYPE_SUITE             49.119754
CNT_PAYMENT                 22.286366
DAYS_FIRST_DRAWING          40.298129
DAYS_FIRST_DUE              40.298129
DAYS_LAST_DUE_1ST_VERSION   40.298129
DAYS_LAST_DUE               40.298129
DAYS_TERMINATION            40.298129
NFLAG_INSURED_ON_APPROVAL   40.298129
dtype: float64
```

In [11]:
```
#drop columns with percentage of missing value >22
pre_app=pre_app.drop(columns=missing_percentage_pre_app_col.index)

##validating the shape after dropping the columns
pre_app.shape
```

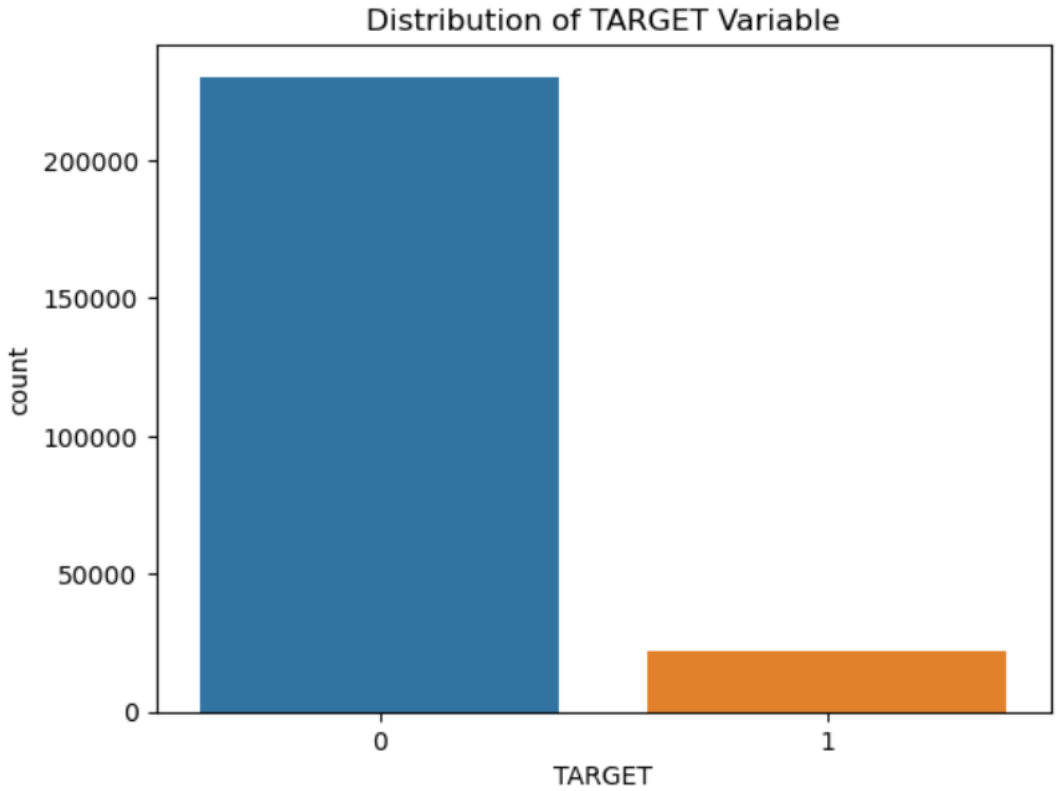
Out[11]:
```
(1670214, 23)
```

In [81]:
```
#drop columns with percentage of missing value >40
app_data_1=app_data.drop(columns=app_data_missing_percentage_col.index)
```

In [82]:
```
##validating the shape after dropping the columns
app_data_1.shape
```

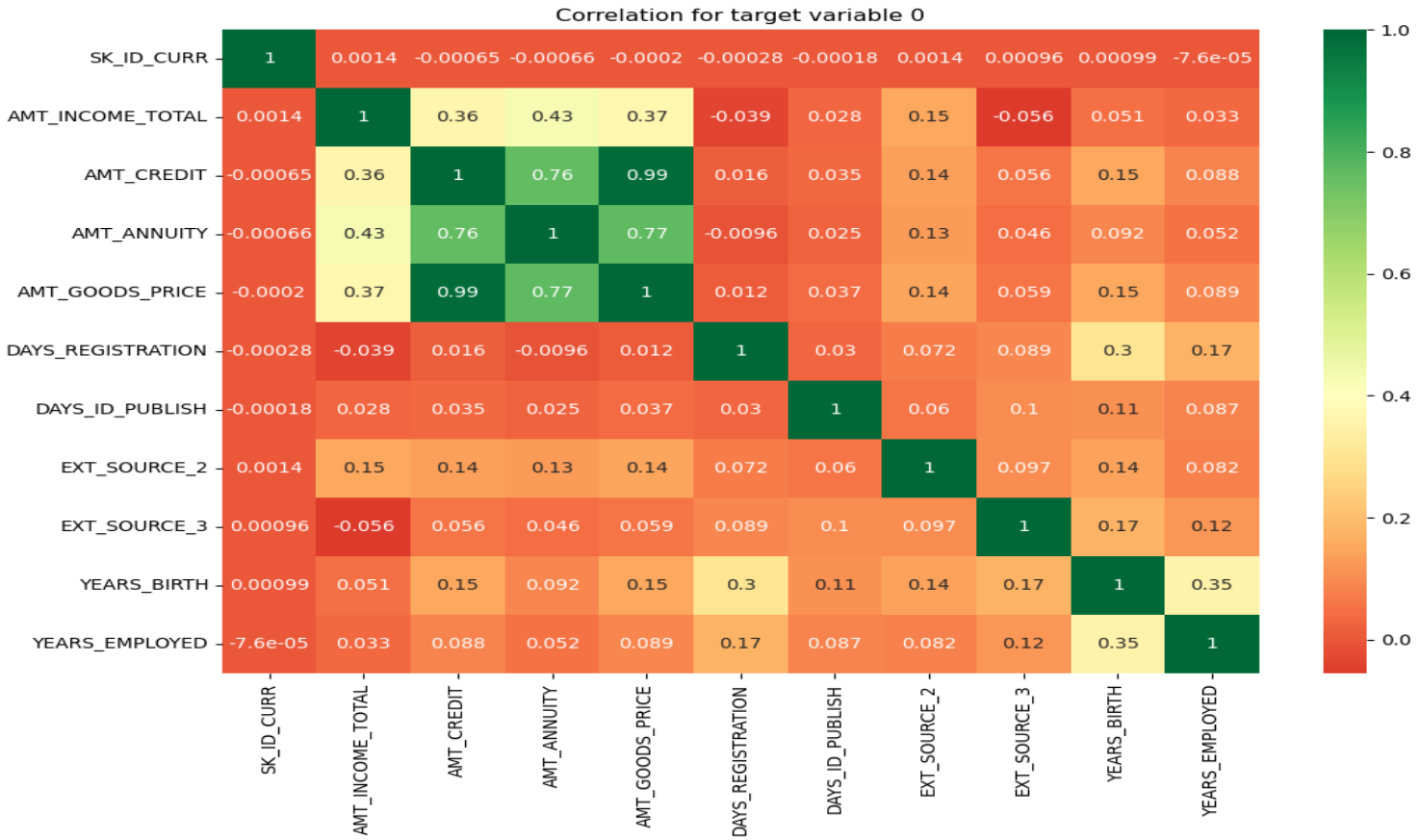
Out[82]:
```
(307511, 73)
```

Finding Imbalance



1) Percentage of customer being defaulter is imbalanced as compare to not being defaulter
2) The gap between the perecentage of taget values are significant

Correlation For Target Variable 0

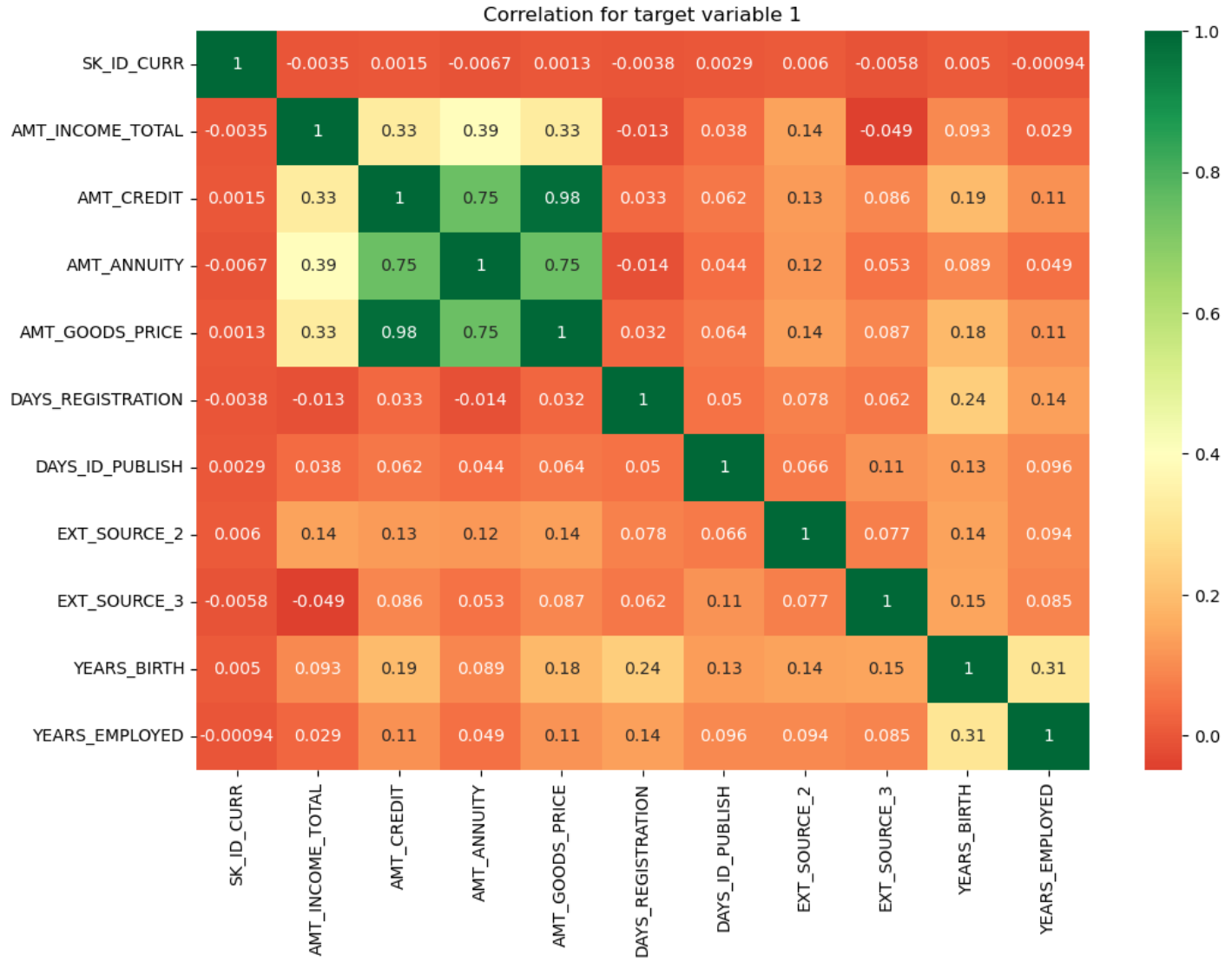


1) AMT_GOODS_PRICE and AMT_CREDIT have highest correlation around 0.99

2) AMT_GOODS_PRICE and AMT_ANNUITY is also showing better correlation

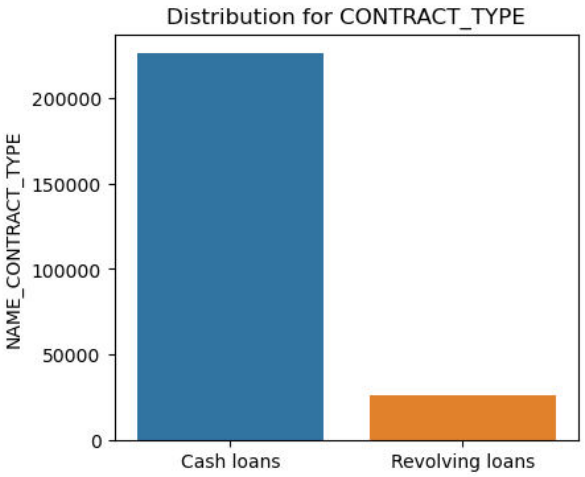
3) Days Registration is having negative correlation with AMT_INCOME_TOTAL

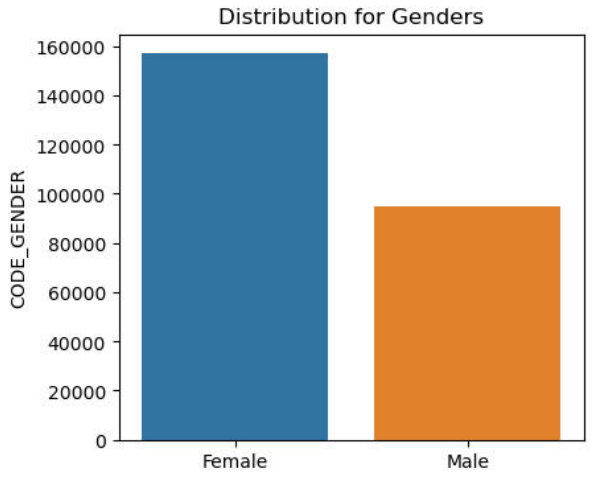
Correlation For Target Variable 1

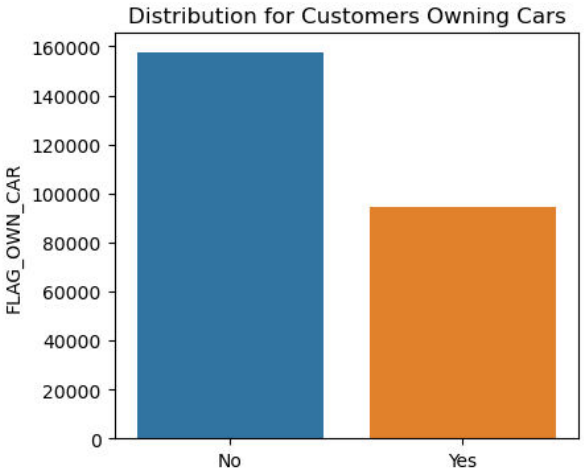


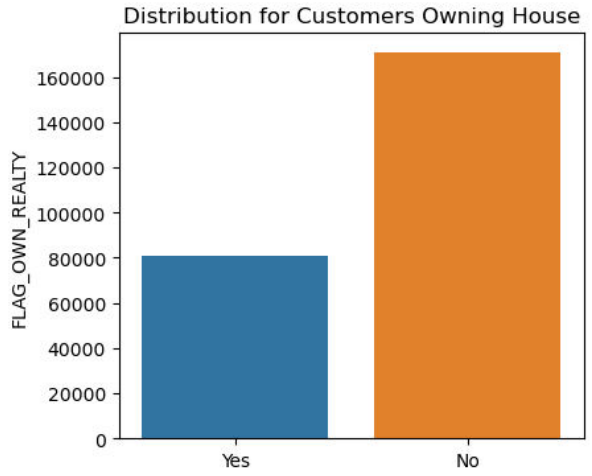
AMT_CREDIT and AMT_GOODS_PRICE have heighest correlation around 0.98

Univariate Analysis









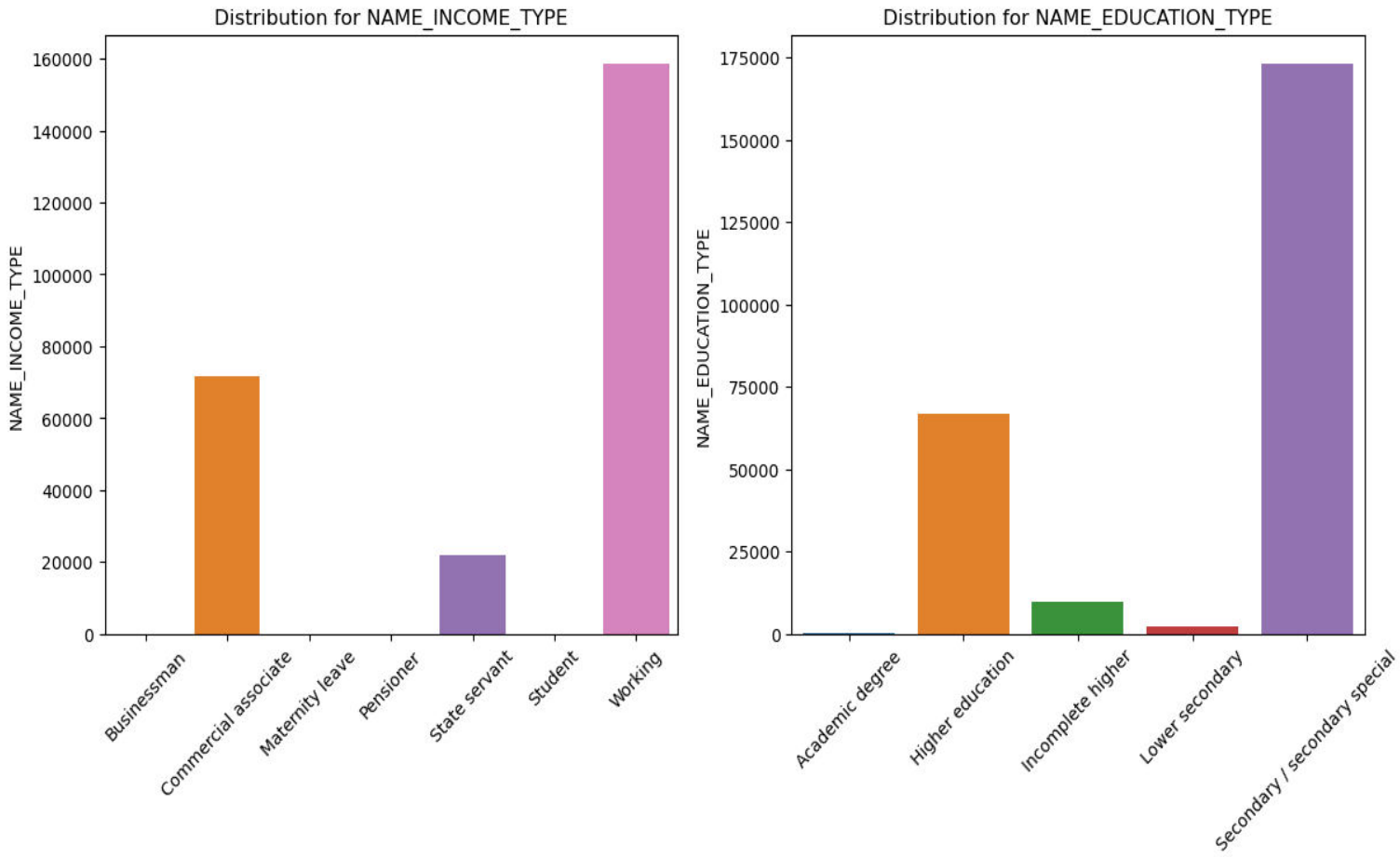
1) Number of Cash loan is more than Revolving loans it means more clients are opting for cash loans

2) Number of female clients is more than male clients

3)Clients who owns a car is less than clients who does not owns a car

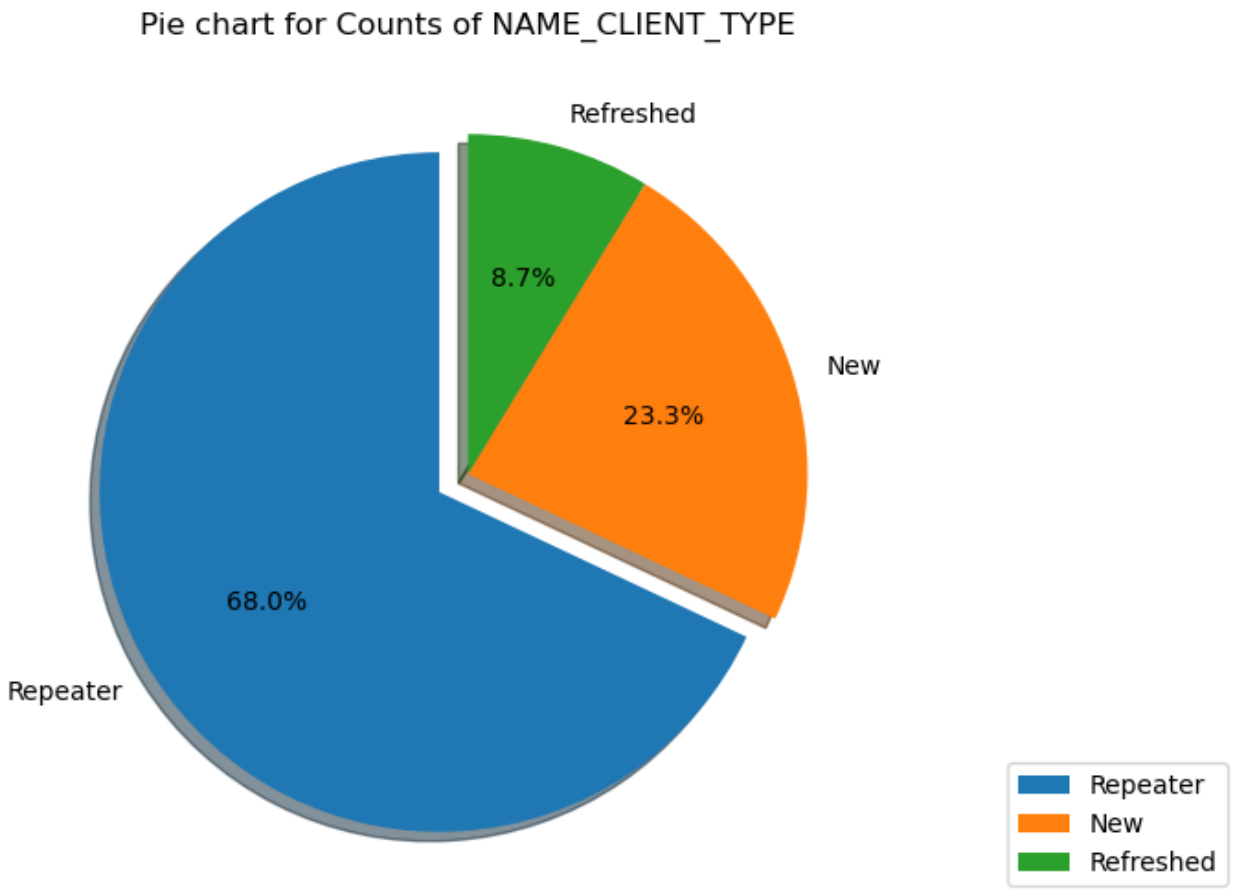
4) Clients who owns reality is around 50 % less than clients who does not owns a reality

Univariate Analysis



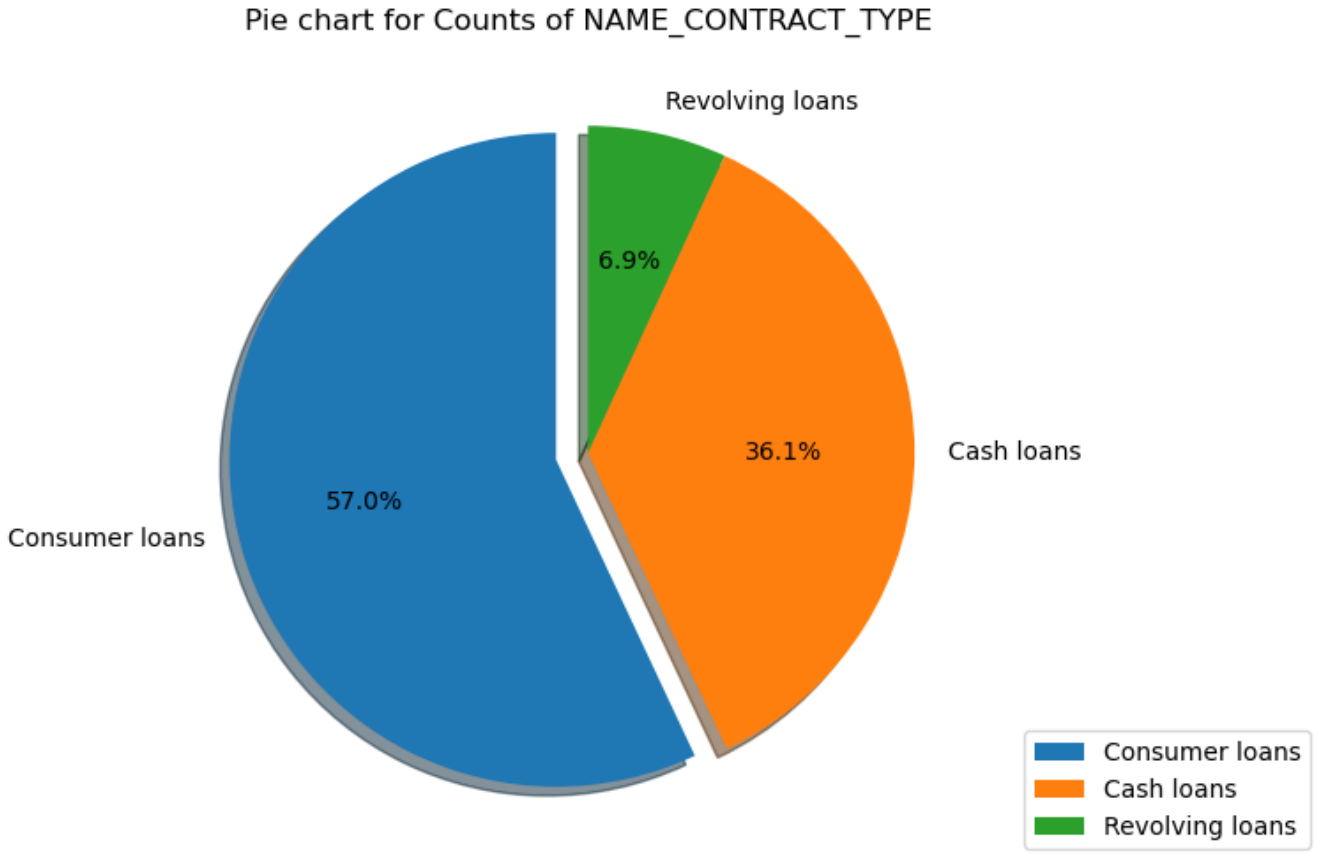
People with secondary/ special, higher education likely to make less default

Univariate Analysis

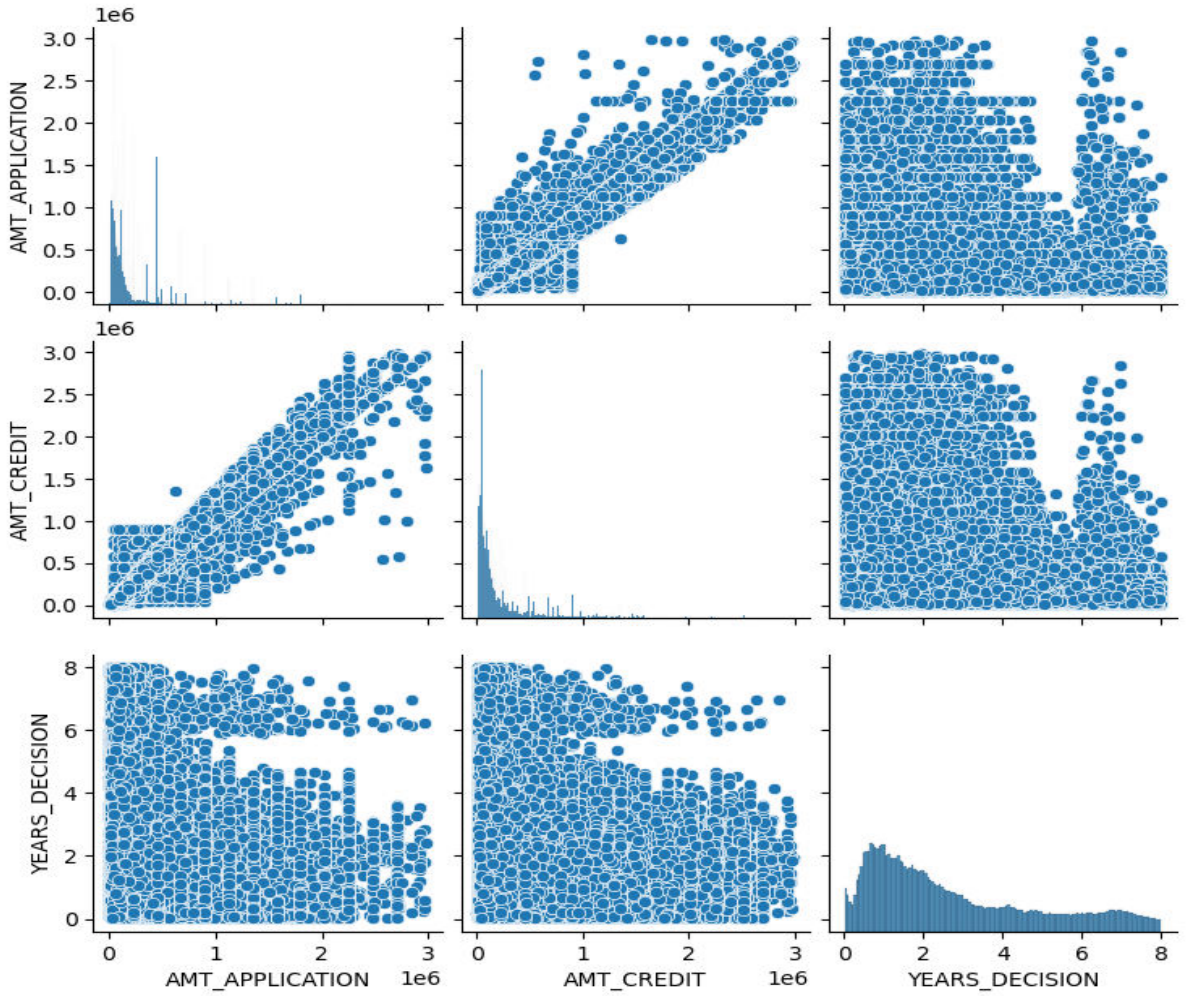


Most of the applicants are repeater

Univariate Analysis

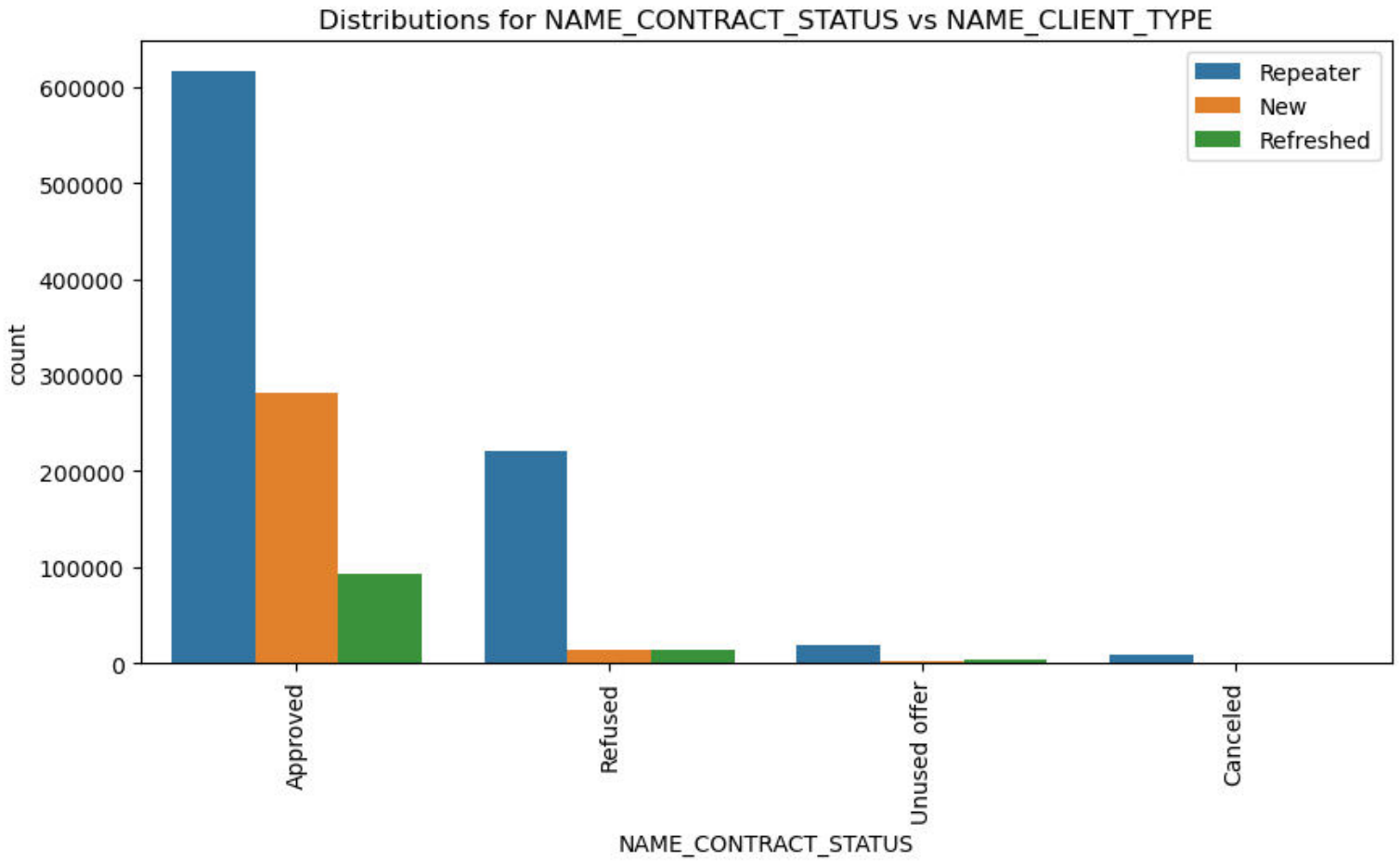


Bivariate Analysis



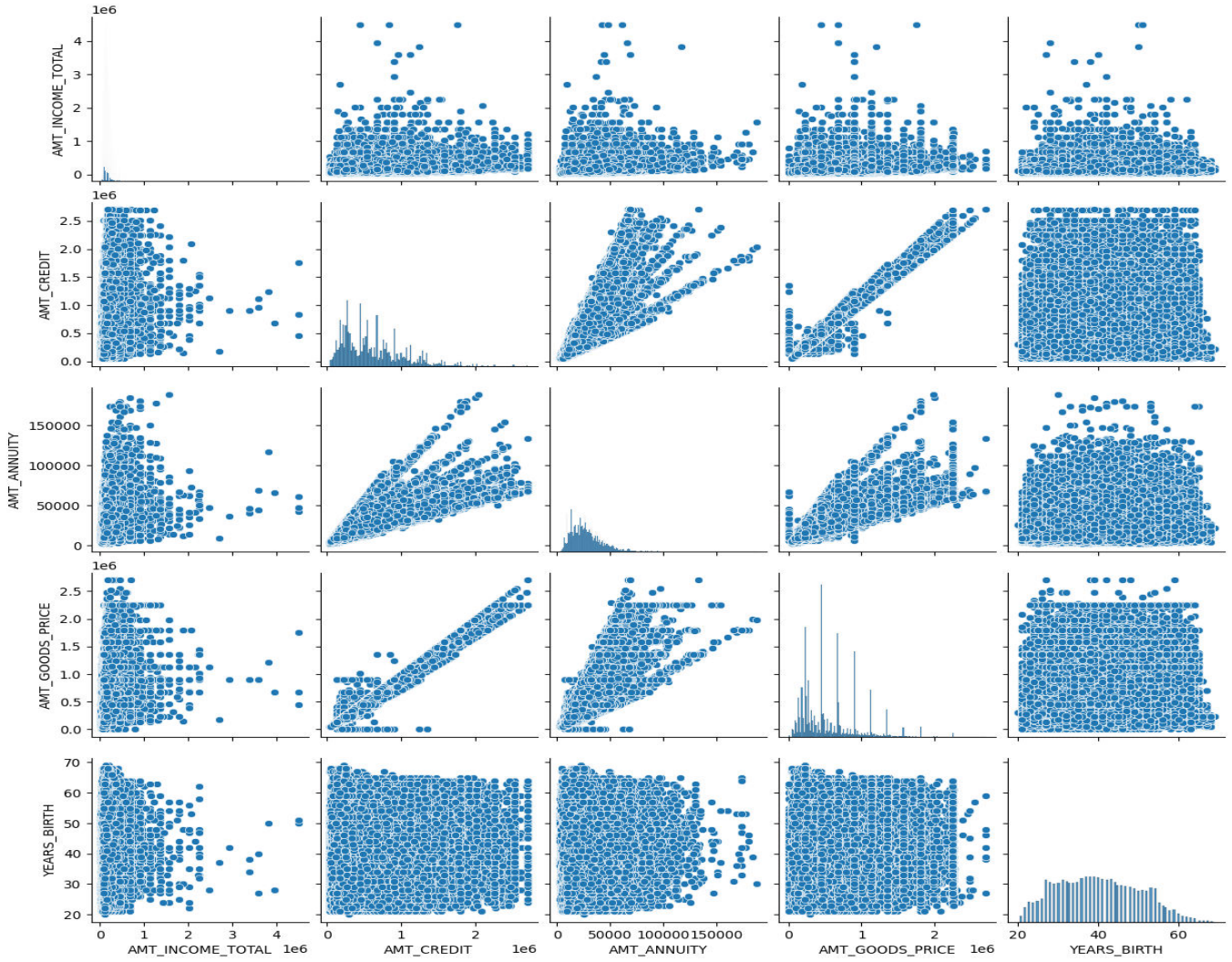
There is linear relationship between AMT_APPLICATION & AMT_CREDIT

Bivariate Analysis



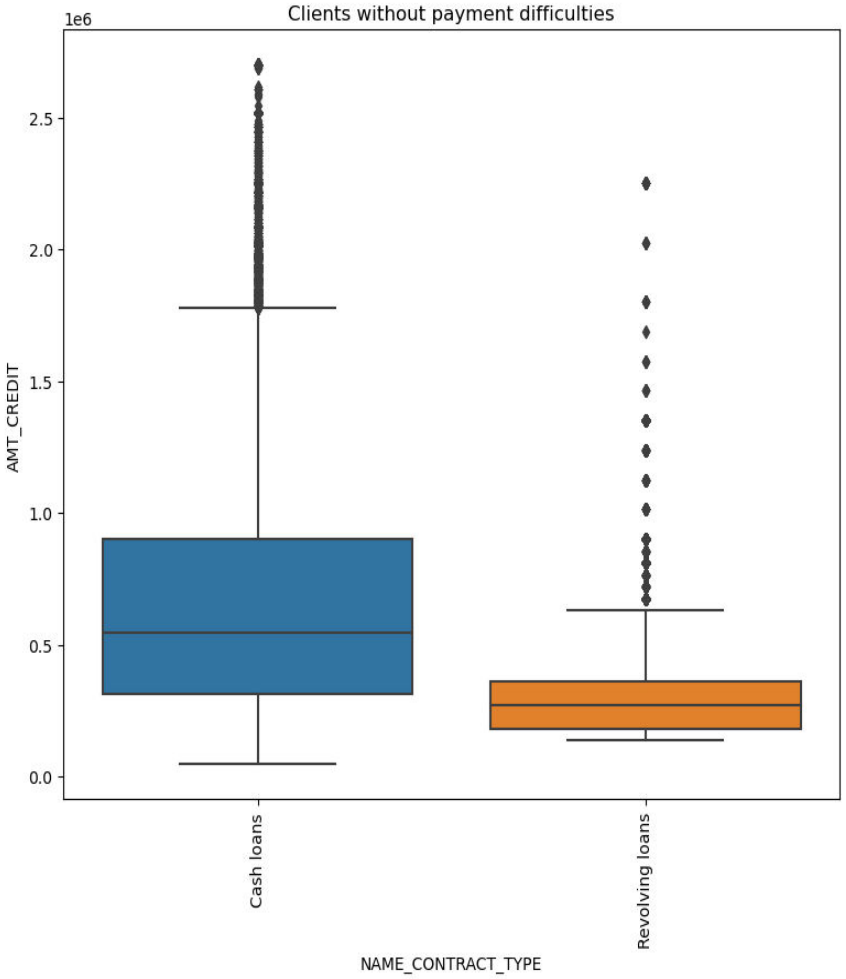
Approval rate for repeater applicants is higher than any other category

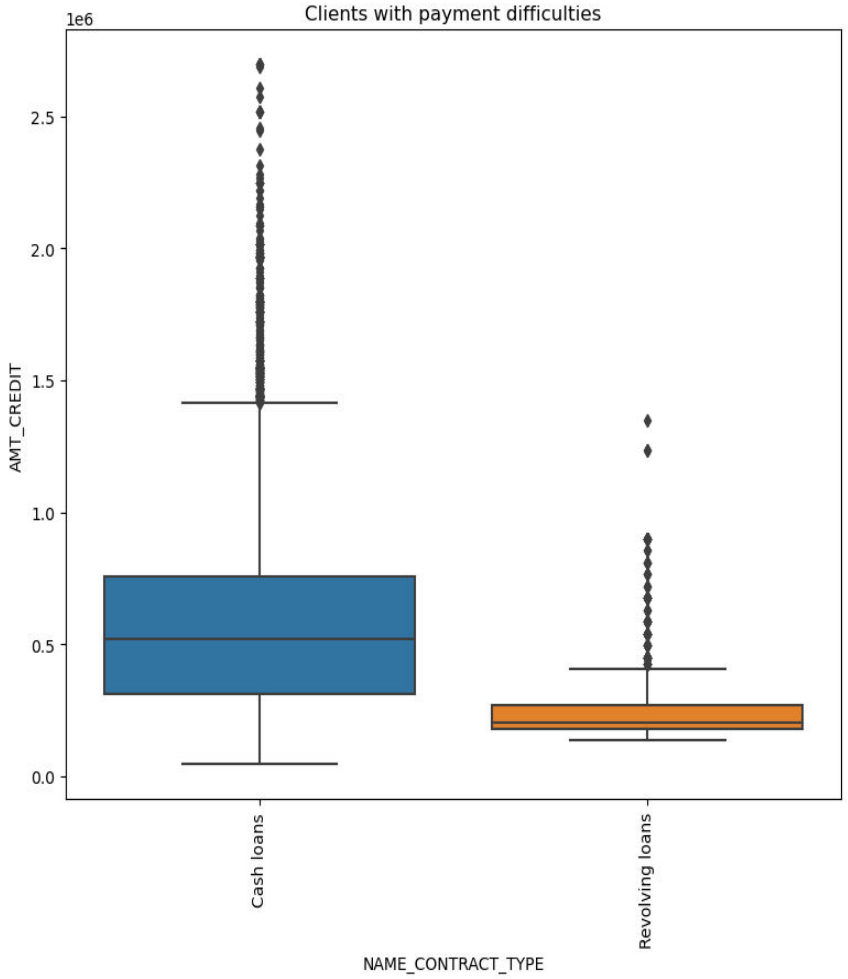
Bivariate Analysis



Linear correlation present between AMT_GOODS_PRICE vs AMT_CREDIT , AMT_GOODS_PRICE vs AMT_ANNUITY

Bivariate Analysis





Applicants more approached towards cash loans than revolving loans.

Conclusion

- Occupation type such as IT staff , HR staff , Managers etc have lower difficulty in repayment when observed in comparison to drivers, Cleaning staff ,laborers
- To increases the chances of repaying the loan provided by the companies should less focus on the clients having income type as Working
- In general more number of Cash loans are provided by bank in comparison to Revolving loan post chances to get defaulted increases
- People with higher education and older people default less
- Higher amount of income, less default
- Loans which previously refused or cancelled, high chance of default
- Giving loan to married people rather than single or widow if safer