CREDIT EDA CASE STUDY

BY:
KRUSHIL RAMANI

BUSINESS OBJECTIVE

- This case study aims to identify patterns, which indicate if a client has difficulty paying their instalments, which may be used for taking actions such as denying the loan, reducing the amount of loan, leading at the 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.

- The company wants to understand the driving factors behind the loan default, i.e.The variables which are strong indicators of the default. The company utilize this knowledge for its portfolio and risk assessment.

DATA UNDERSTAND

- The data which we have analysed contains the information about the loan application at the time of applying for the loan.It contains to type of The scenarios:

1. The client with payment difficulties: He/She Had late payment more than x days onAt least one of the first Y Instalment of the loan in our sample.(In our analysis,It is mentioned as target =1)
2. All Other cases When the payment is paid on time.(You know our analysis,It is mentioned as target=0)

STEPS

1. Dataunderstanding and data gathering
2. Check for data quality and data clening
3. Univariate analysis
4. Bivariate analysis
5. Insights

DATA ANALYSIS

- On New application dataset.

```
In [4]: #read dataset
        na_data = pd.read_csv('application_data.csv')
```

```
In [49]: na_data.TARGET.value_counts()
```

```
Out[49]: TARGET
         0    229841
         1     21790
         Name: count, dtype: int64
```

- there is a imbalance in target variable, so we need to saparate the target variable in two dataframe.

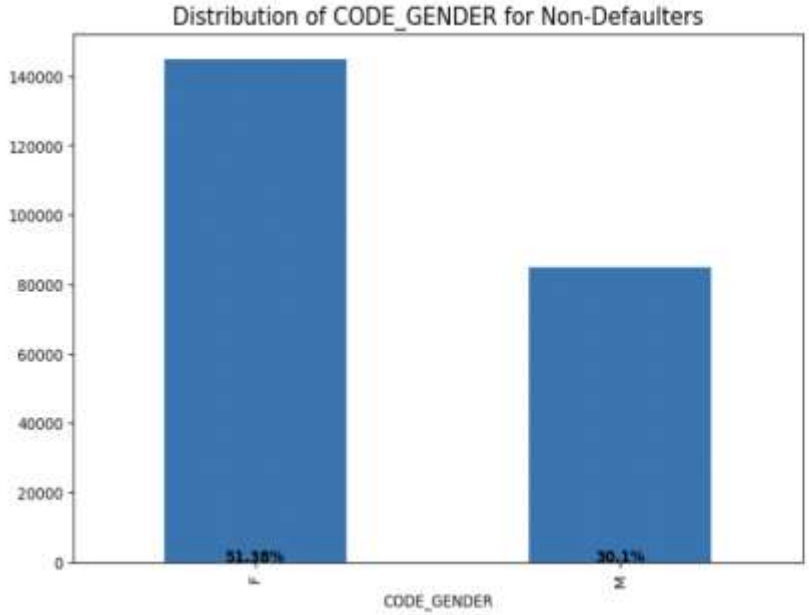
```
In [50]: good_client = na_data[na_data.TARGET == 0]
         bad_client = na_data[na_data.TARGET == 1]
```

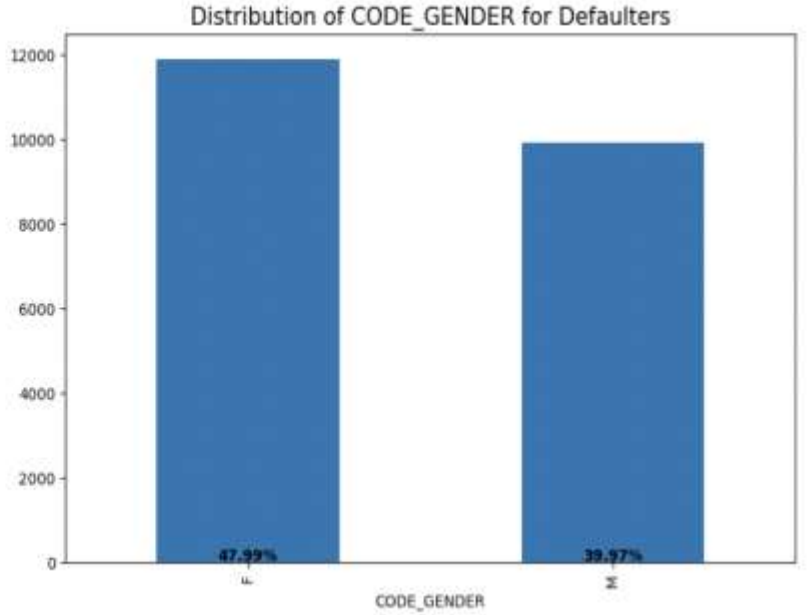
- Here good_client are the clients who do not have difficulty in payment of loans and bad_client are the clients with difficulty in payment of loan.

UNIVARIATE ANALYSIS

- On Categorical features

1. Gender :

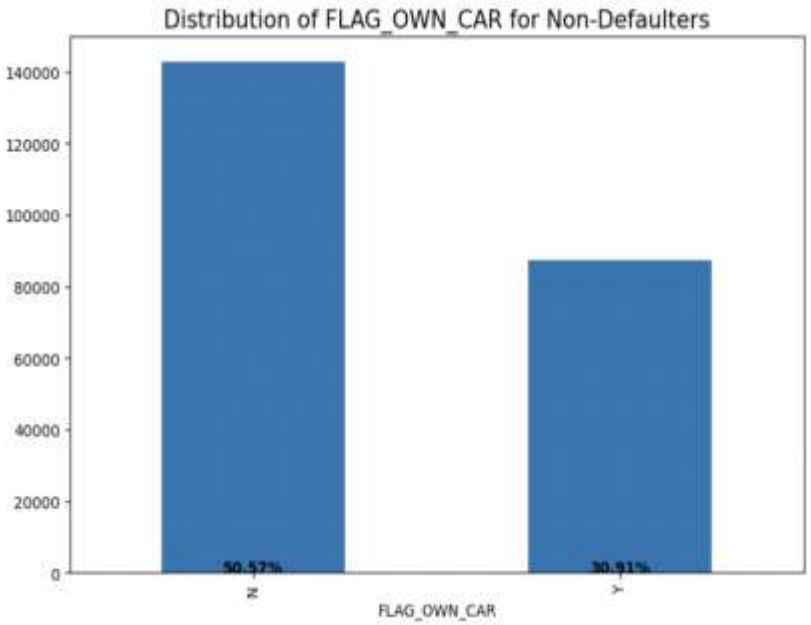


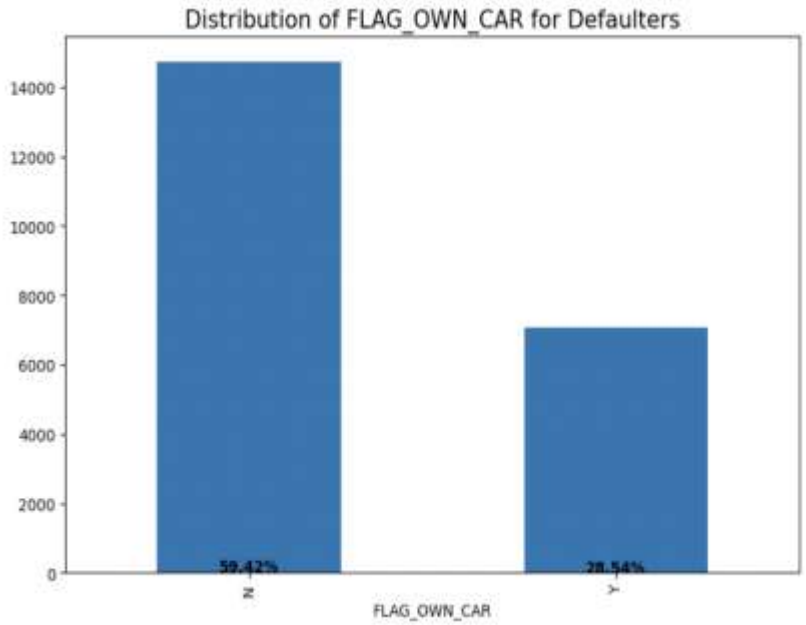


- We can see that Female contribute 51.38% to the non-defaulters while 47.99% to the defaulters.
- We see more female applying for loans than males and hence the more number of female defaulters as well.But the rate of defaulting of FEMALE is much lower compared to their MALE counterparts.

UNIVARIATE ANALYSIS

2. FLAG_OWN_CAR :

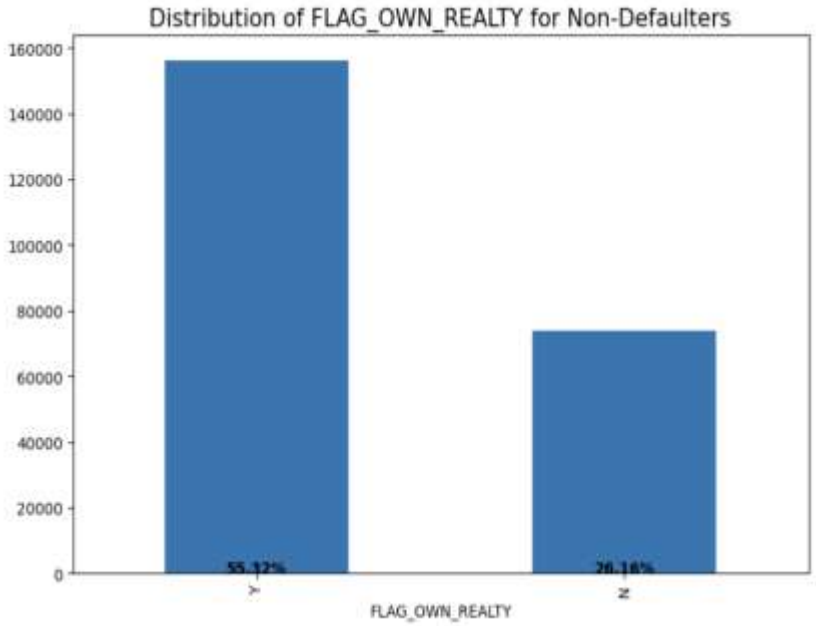


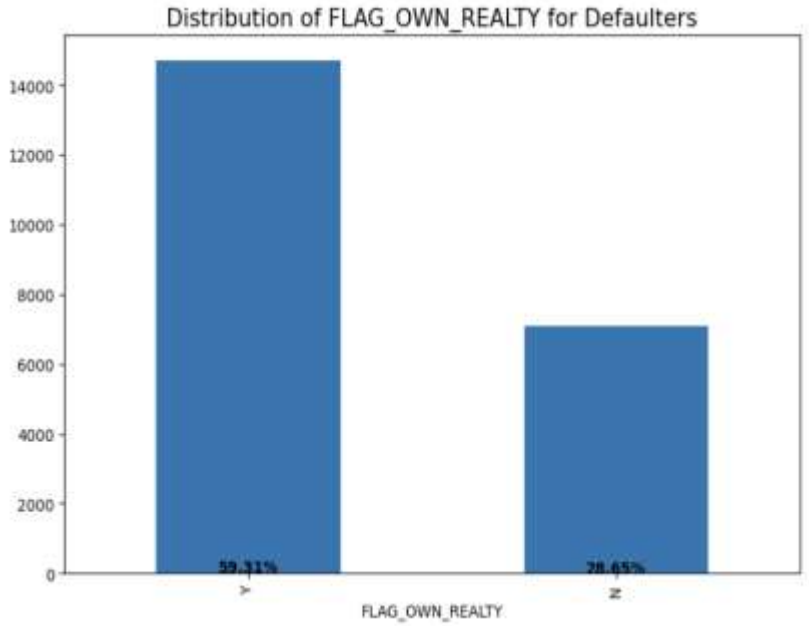


- clients with cars/vehicles are less in the population. we can see that clients who own car/vehical have less defaults compare to clients who do not own car/vehical.

UNIVARIATE ANALYSIS

3. FLAG_OWN_REALTY :

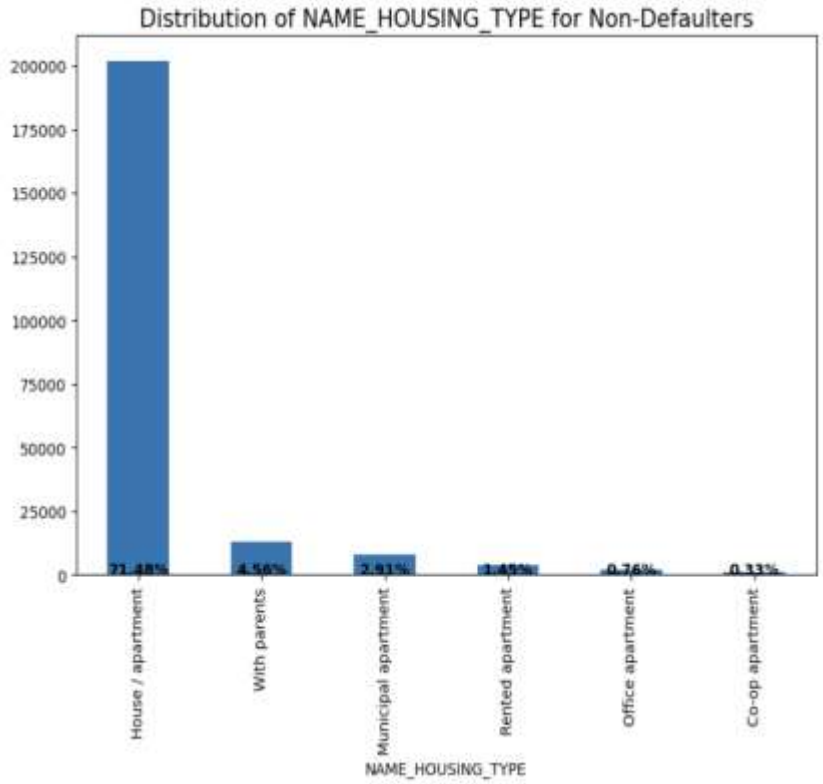


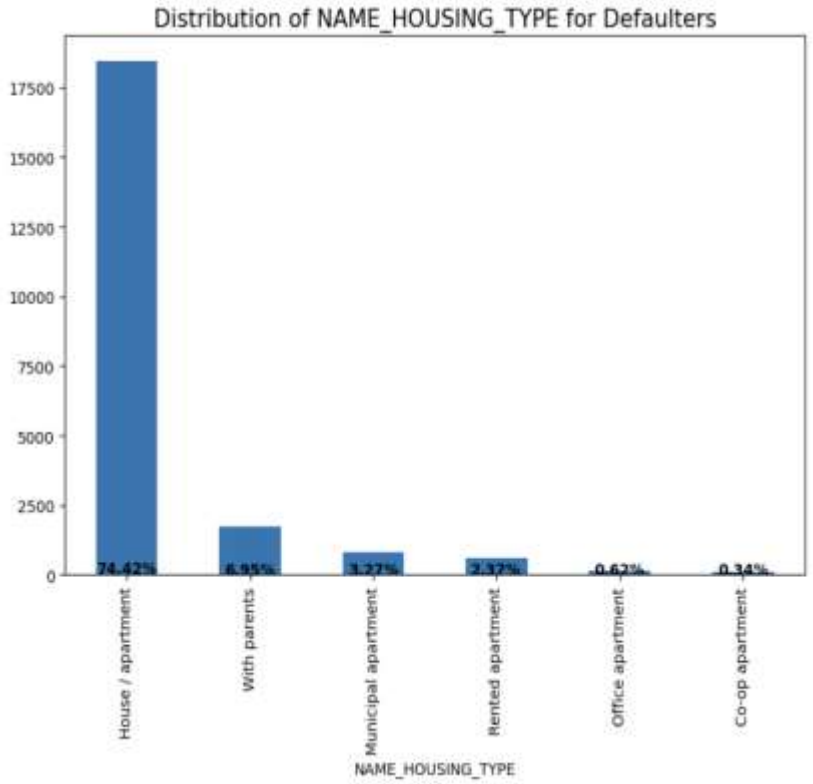


- From above observation we can see that clients having property are more than the clients with out property in application for loans. so we can say that clients having property are more likely to default than the clients with out property.

UNIVARIATE ANALYSIS

4. NAME_HOUSING_TYPE :

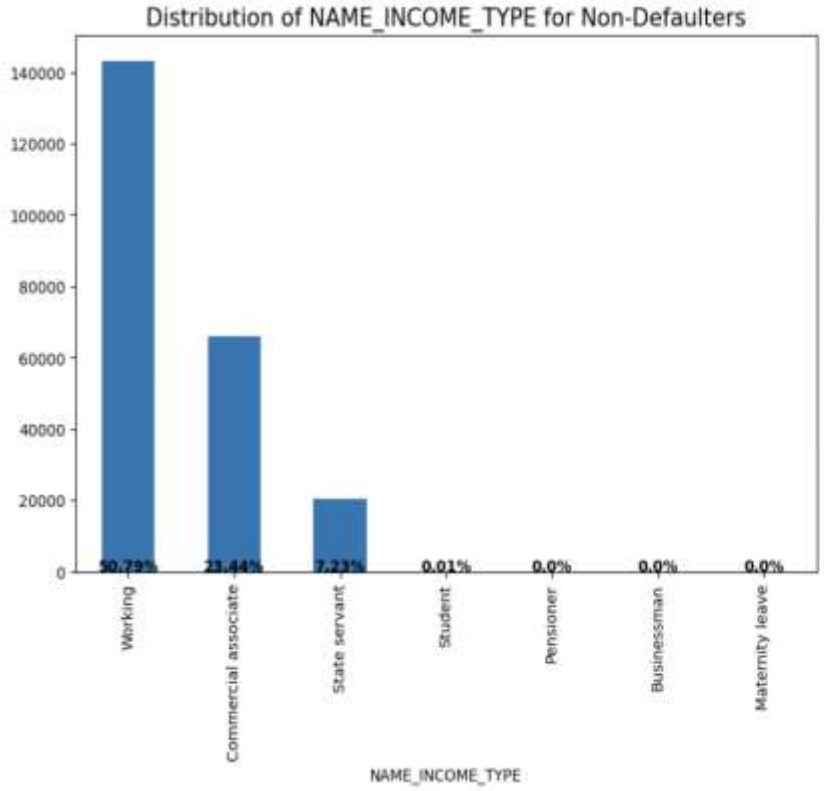


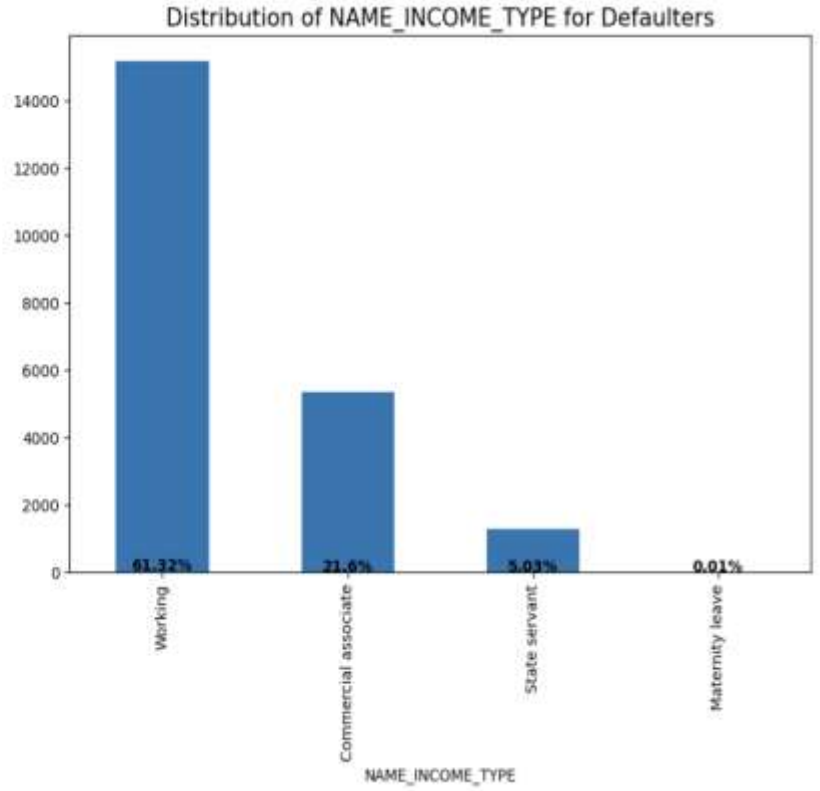


- house/apartment are majority loan seeker, and highest loan defaulters.

UNIVARIATE ANALYSIS

5. NAME_INCOME_TYPE :

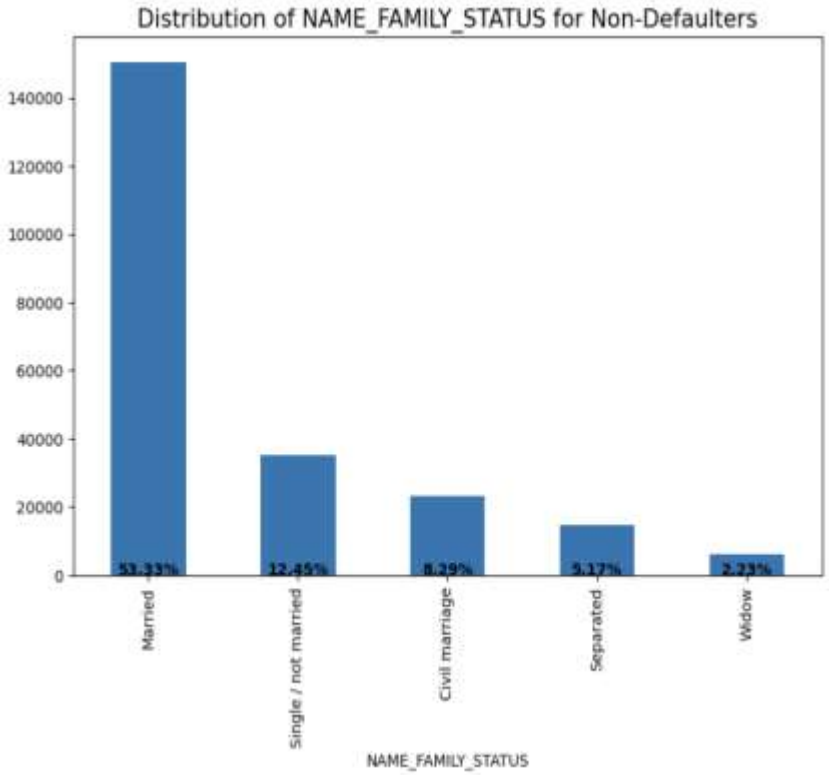


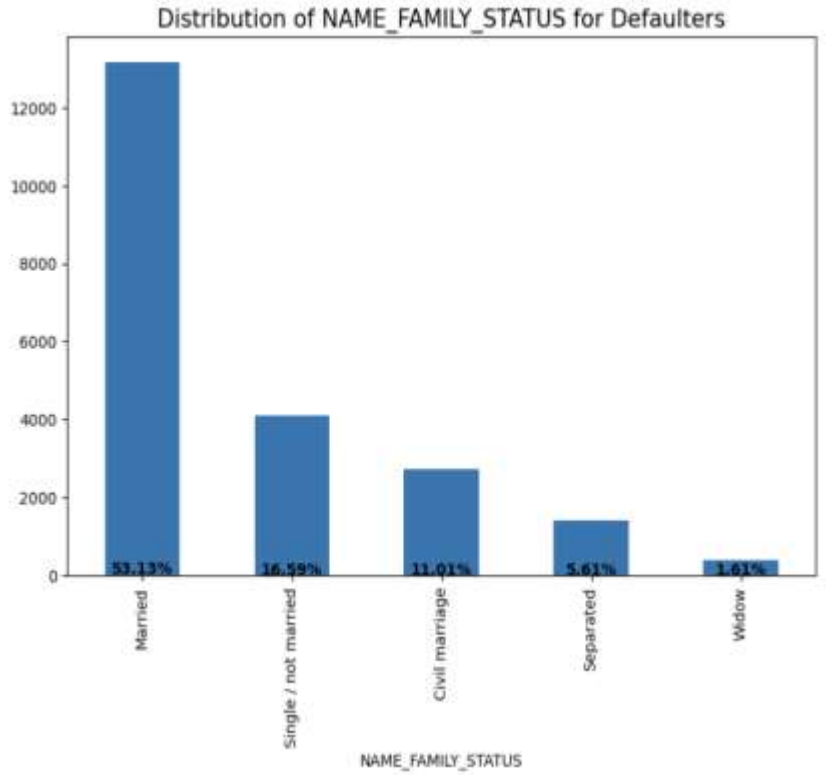


- clients who are working have more defaults than clients who are pensioner and others.

UNIVARIATE ANALYSIS

6. NAME_FAMILY_STATUS:

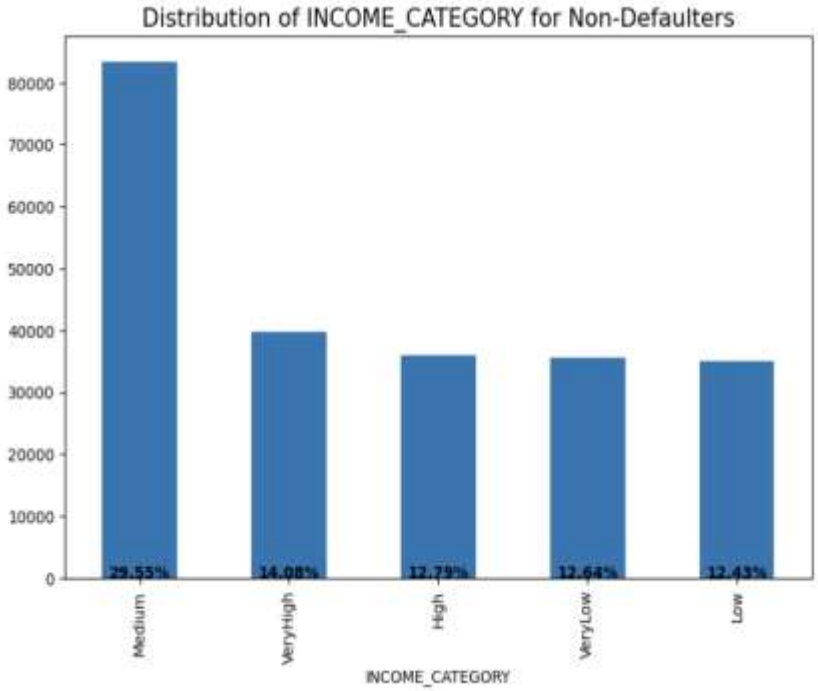


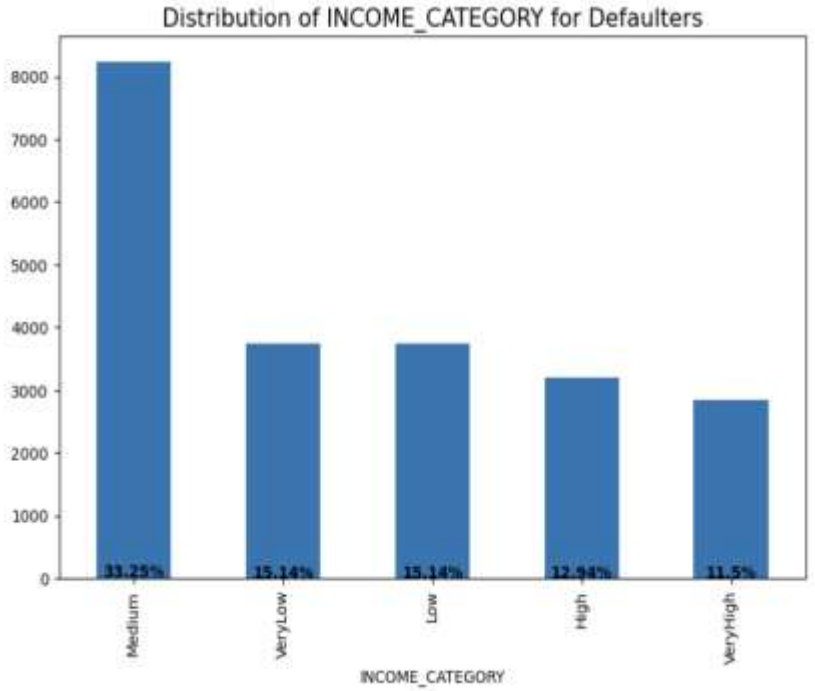


- maried clients have more defaults than other clients.

UNIVARIATE ANALYSIS

7. INCOME_CATEGORY:

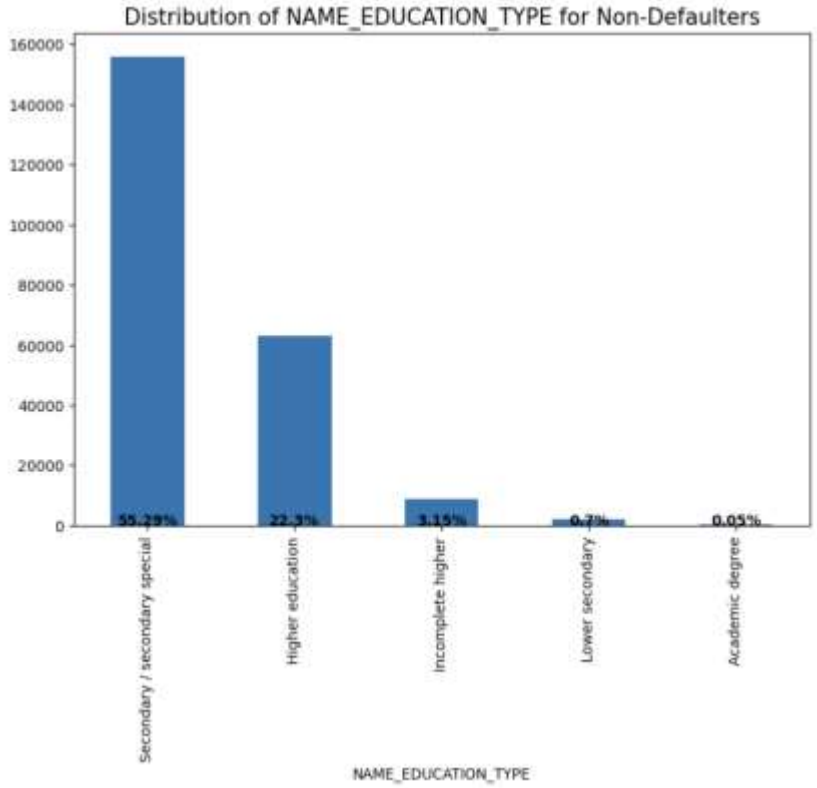


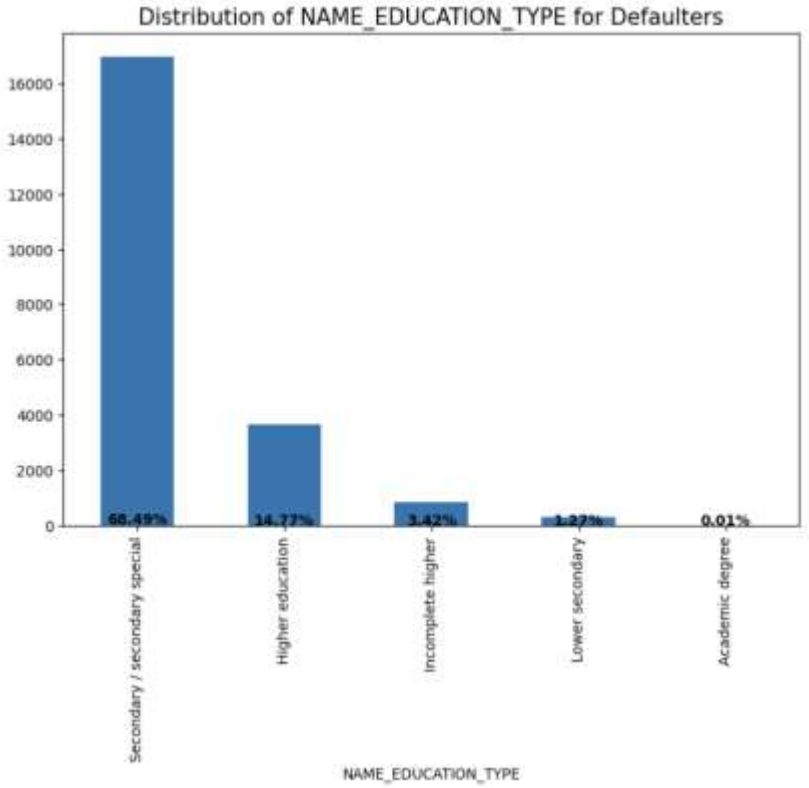


- Medium income clients have highest loan defaults. Client with very high income has less defaults.

UNIVARIATE ANALYSIS

8. NAME_EDUCATION_TYPE:



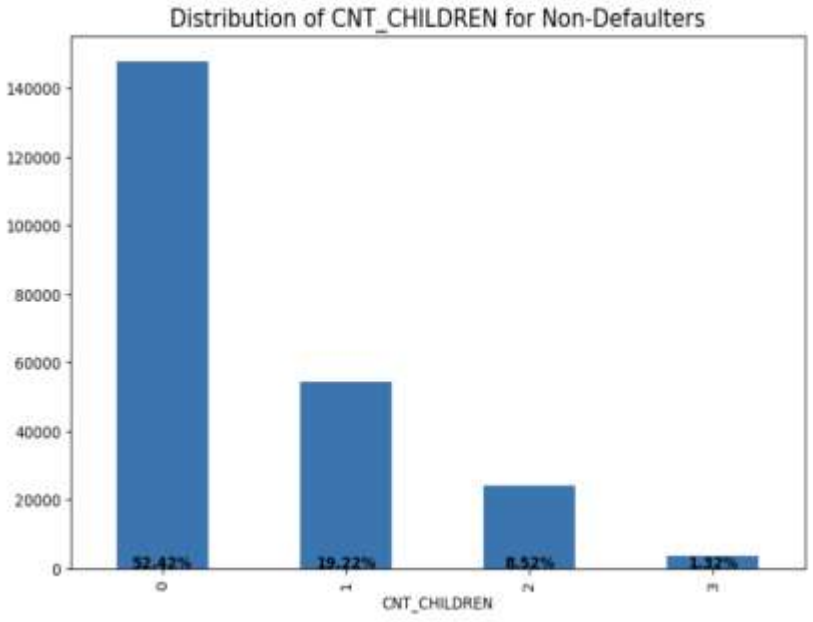


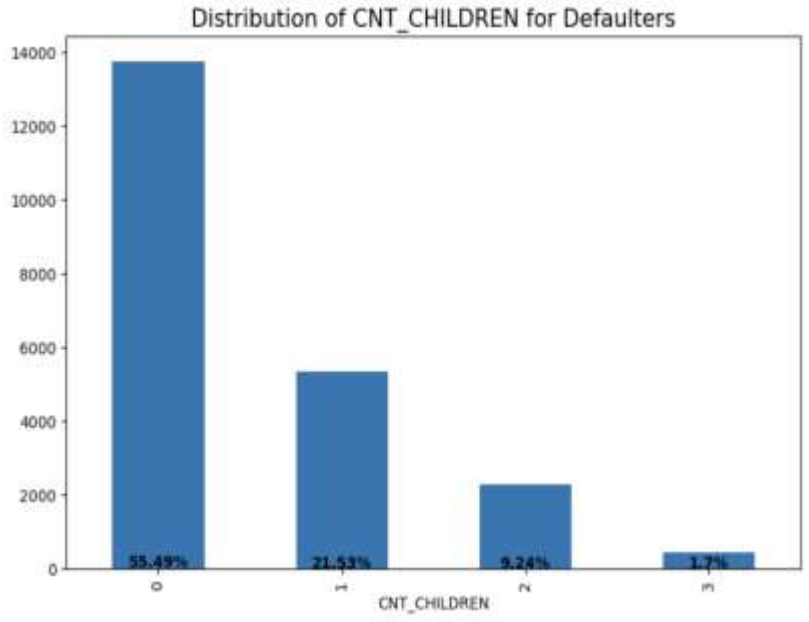
- clients with secondary educaton have more defaults than others.Client with higher education has less defaults.

UNIVARIATE ANALYSIS

- On Continous features

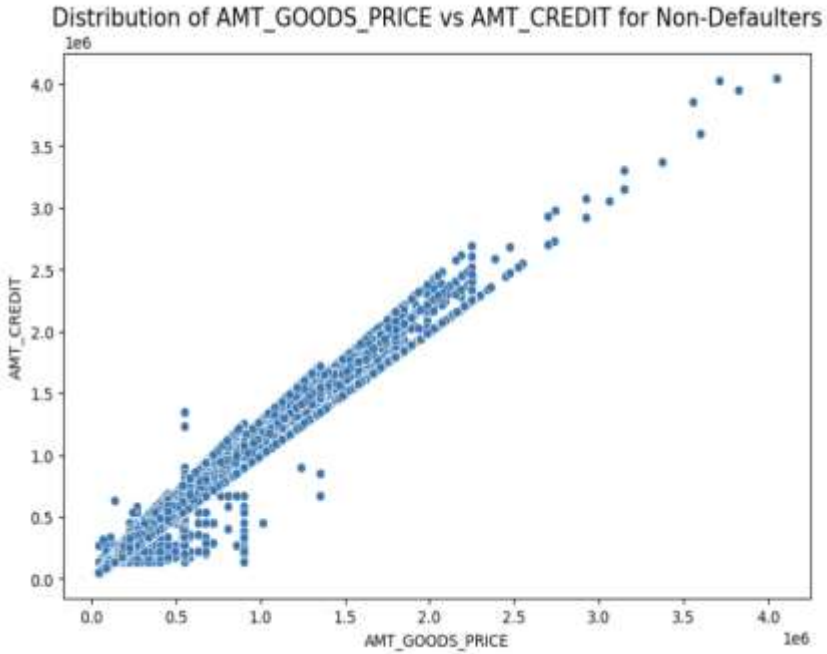
1. CNT_CHILDREN:

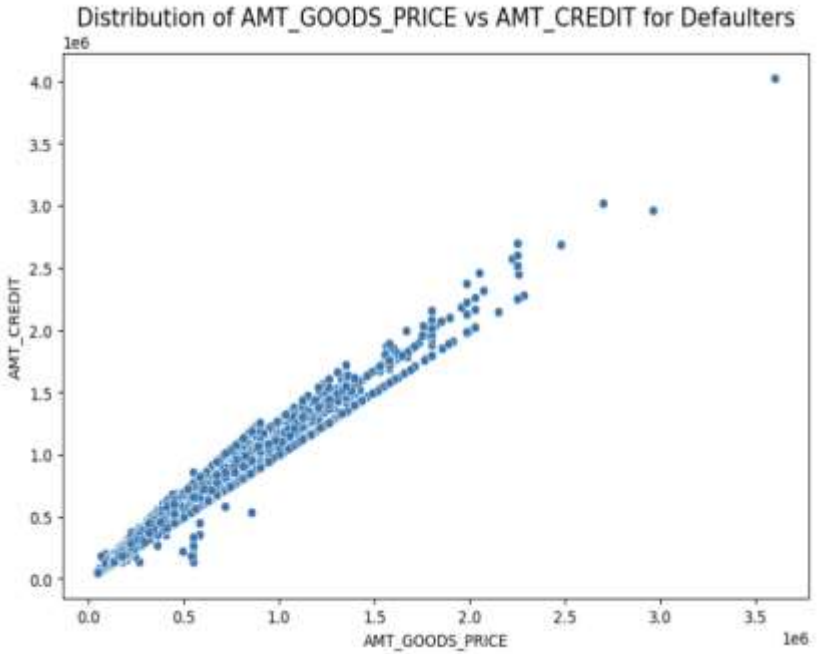




BIVARIATE ANALYSIS

1. AMT_GOODS_PRICE and AMT_CREDIT:

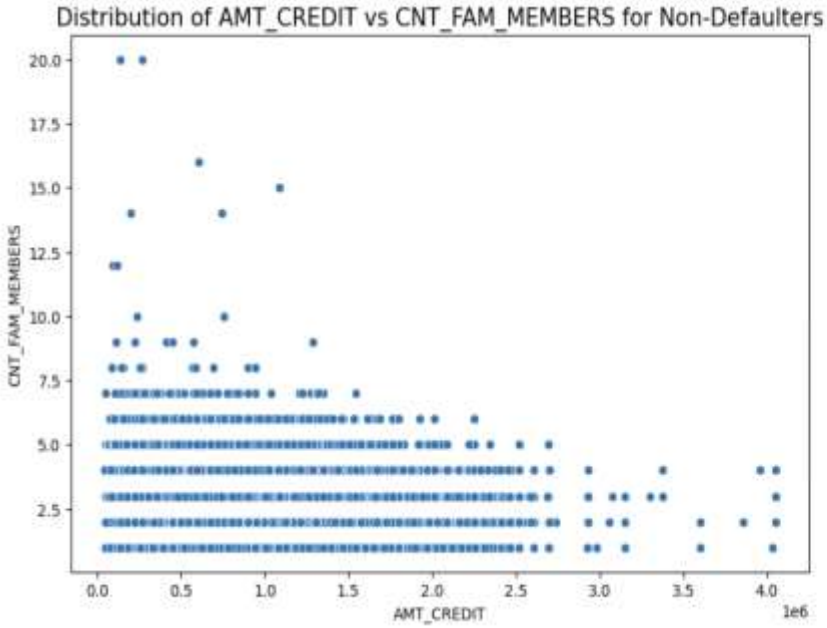


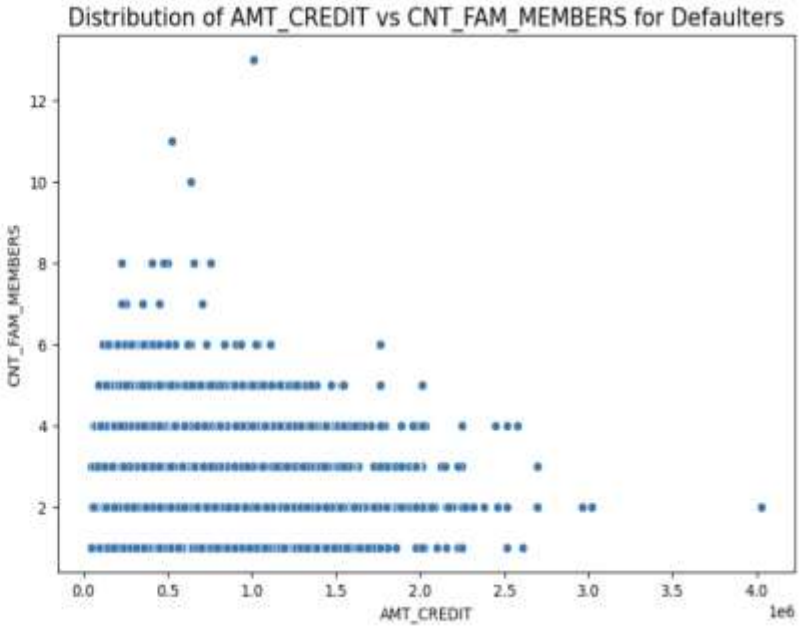


- From above graph we can see that the credit loan amount increases as the price of good increases.
- We can observe the client with high amount good price and high amount credit default less.

BIVARIATE ANALYSIS

2. AMT_CREDIT and CNT_FAM_MEMBERS:

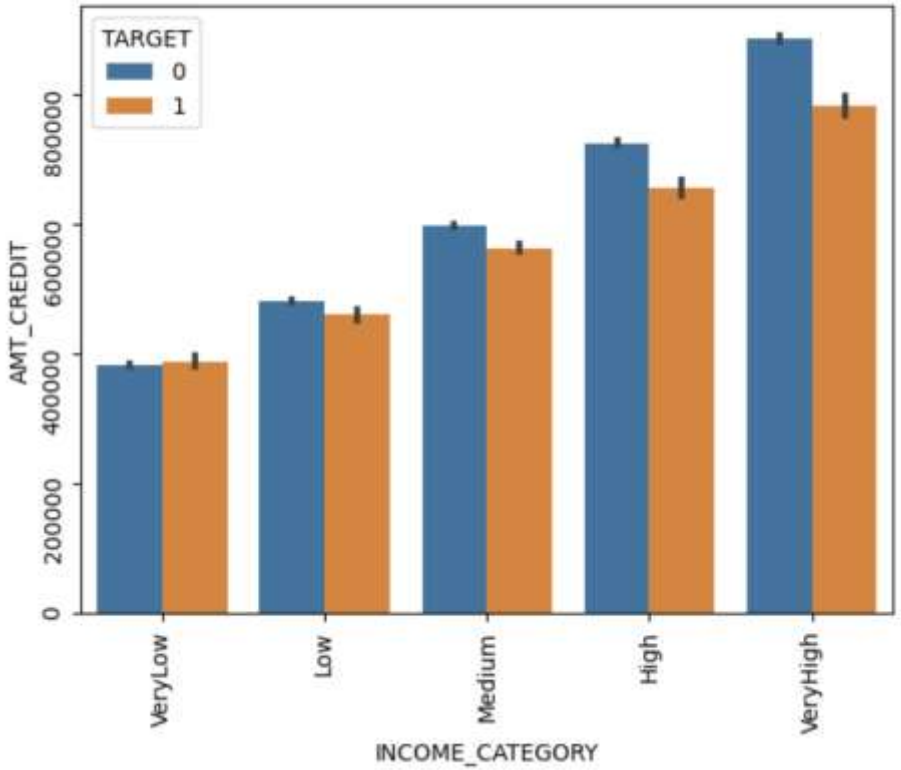




- From above observation the people are likely to default if the family is small and the AMT_CREDIT is low.

BIVARIATE ANALYSIS

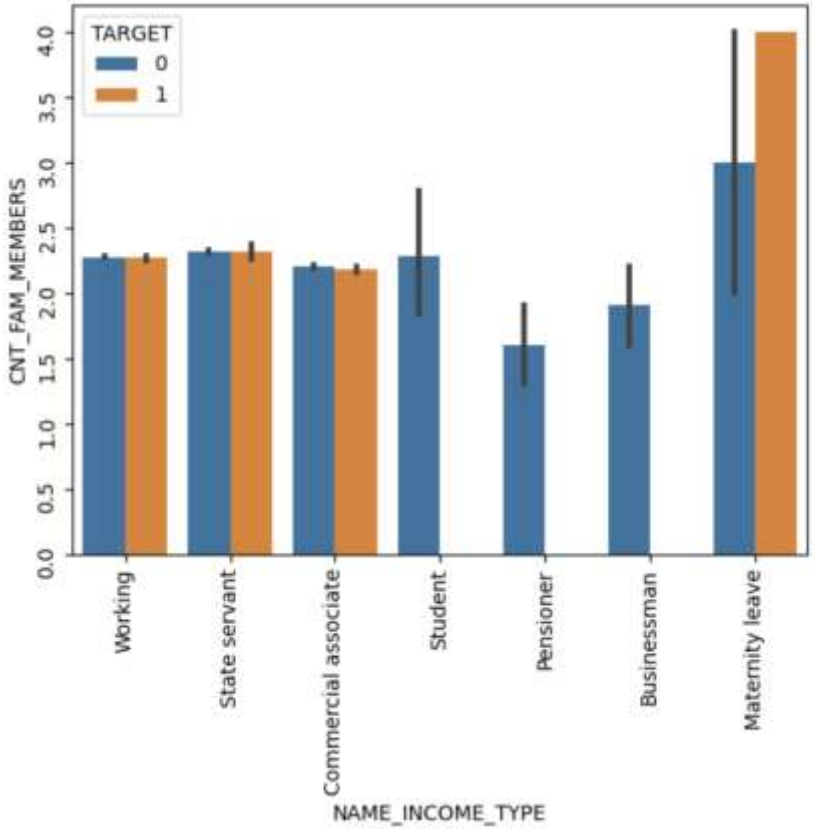
3. INCOME_CATEGORY and AMT_CREDIT:



- From above graph we can see that amount of loan credit increase as clients income increase, thus client with very low income has more chance to default.

BIVARIATE ANALYSIS

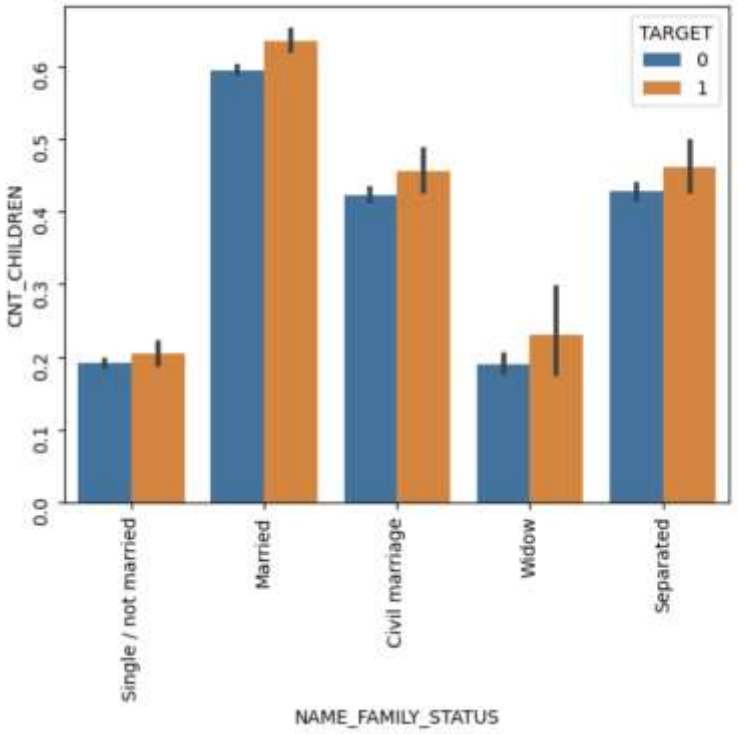
4. NAME_INCOME_TYPE and CNT_FAM_MEMBERS:



- Clients who getting income via maternity leave tends to be more defaulter when they have more family members.

BIVARIATE ANALYSIS

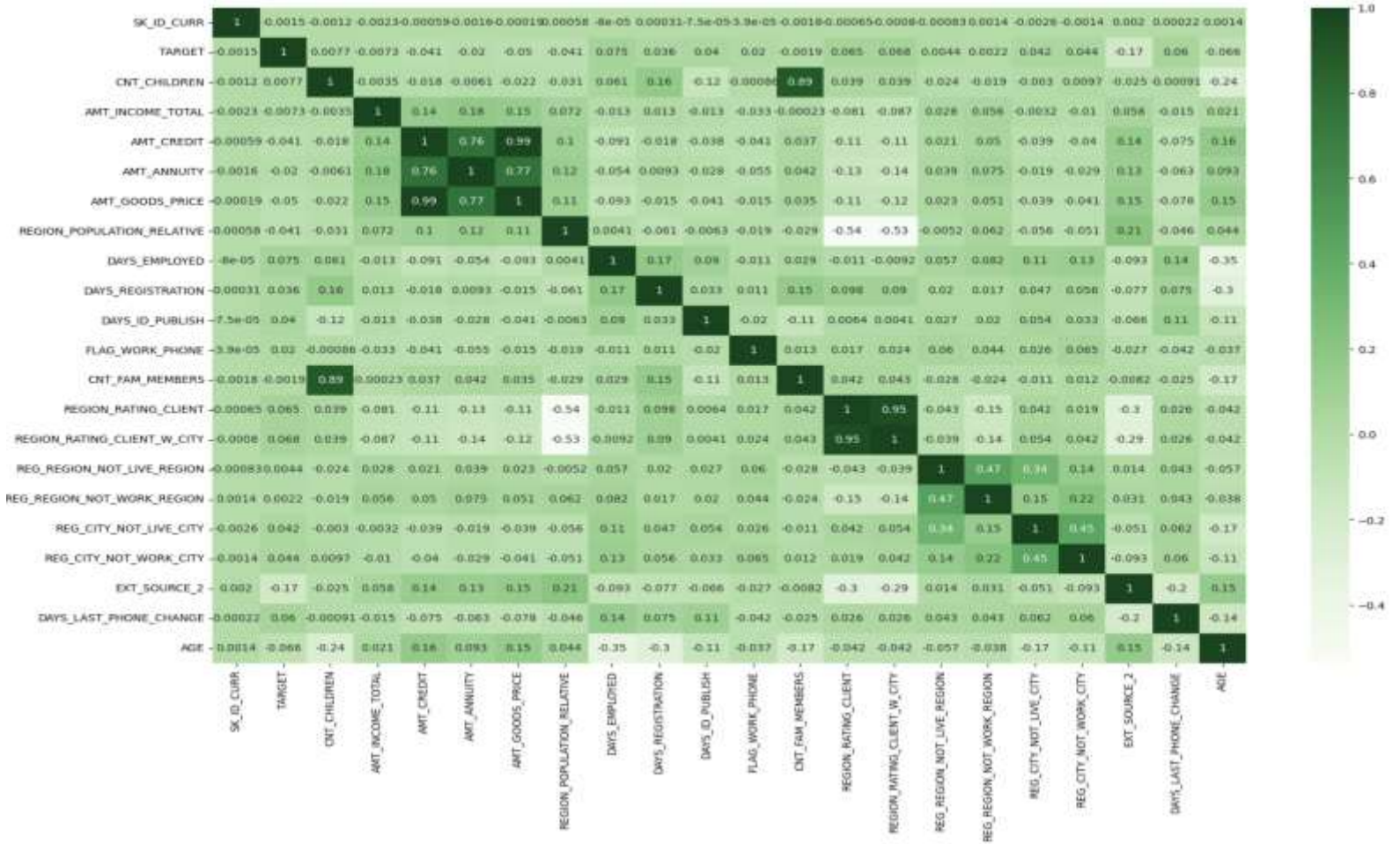
5. NAME_FAMILY_STATUS and CNT_CHILDREN:



- Clients who are married and has more children, chances to be a defaulter in high.

CORRELATION MATRIX

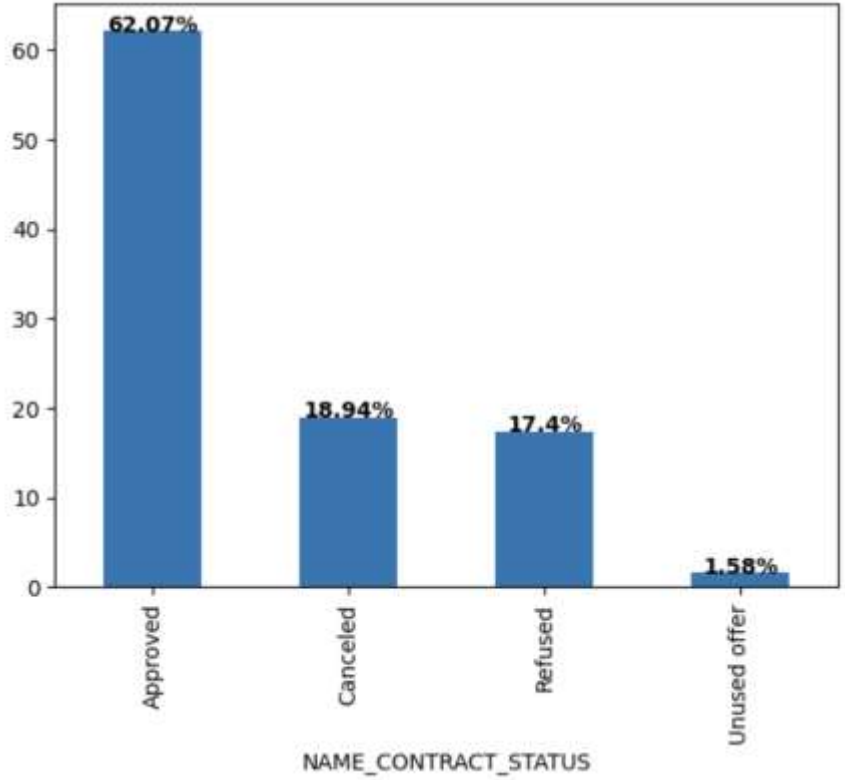
- Correlation matrix of new application dataset.



DATA ANALYSIS ON PREVIOUS APP

- UNIVARIATE ANALYSIS.

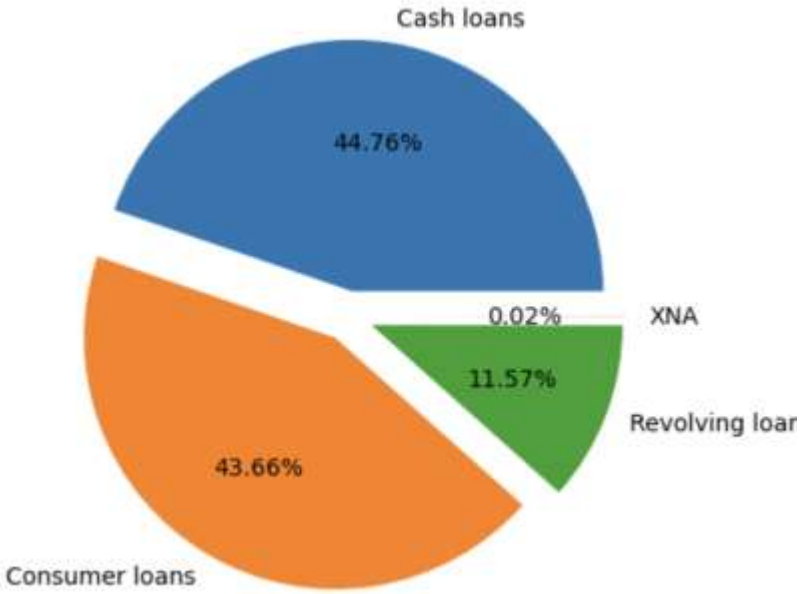
1. NAME_CONTRACT_STATUS:

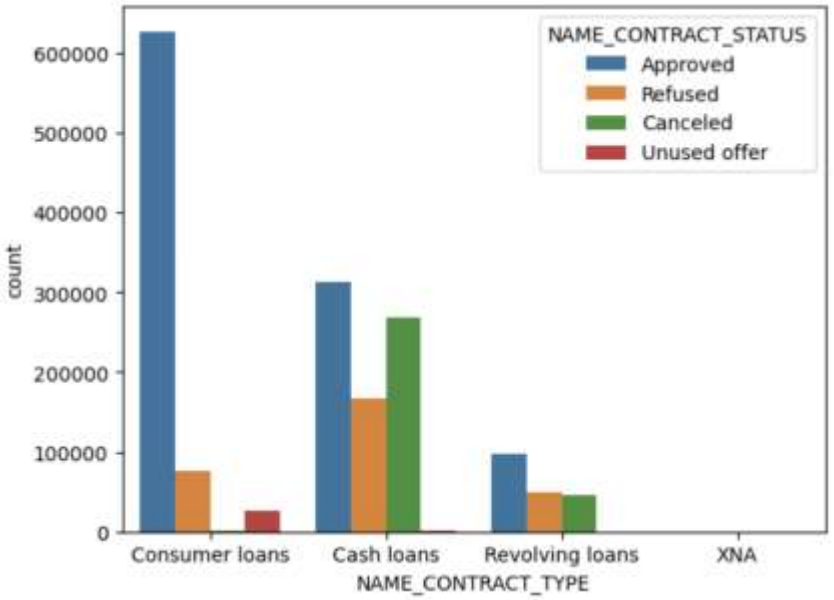


- 62.07% loan applications were approved.

UNIVARIATE ANALYSIS

2. NAME_CONTRACT_TYPE:

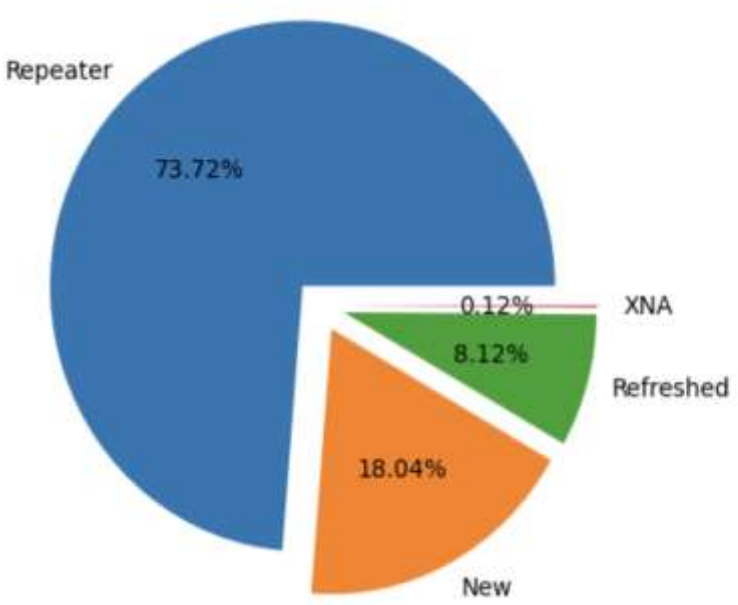


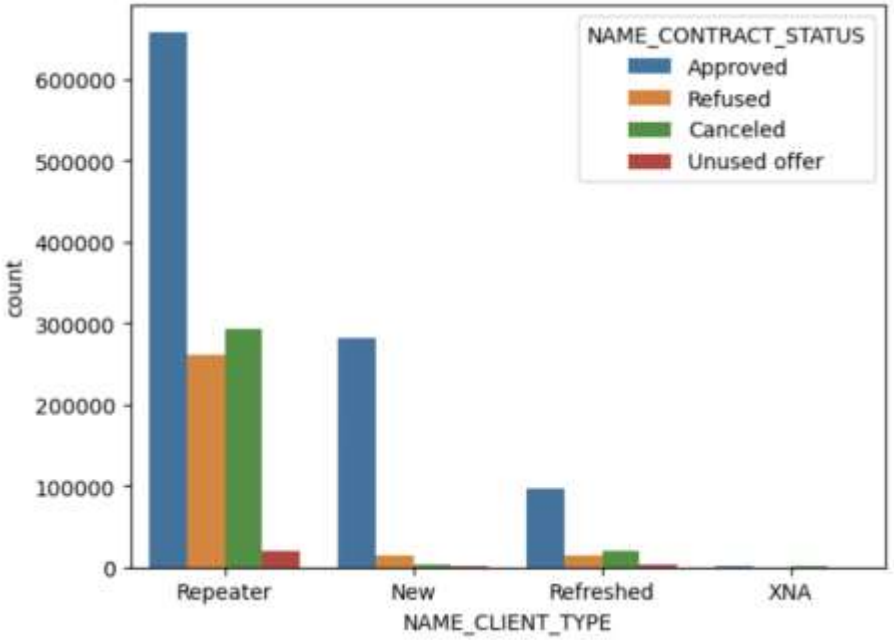


- Almost same application for cash loans and consumer loans.
- More consumer loans were approved.

UNIVARIATE ANALYSIS

3. NAME_CLIENT_TYPE :

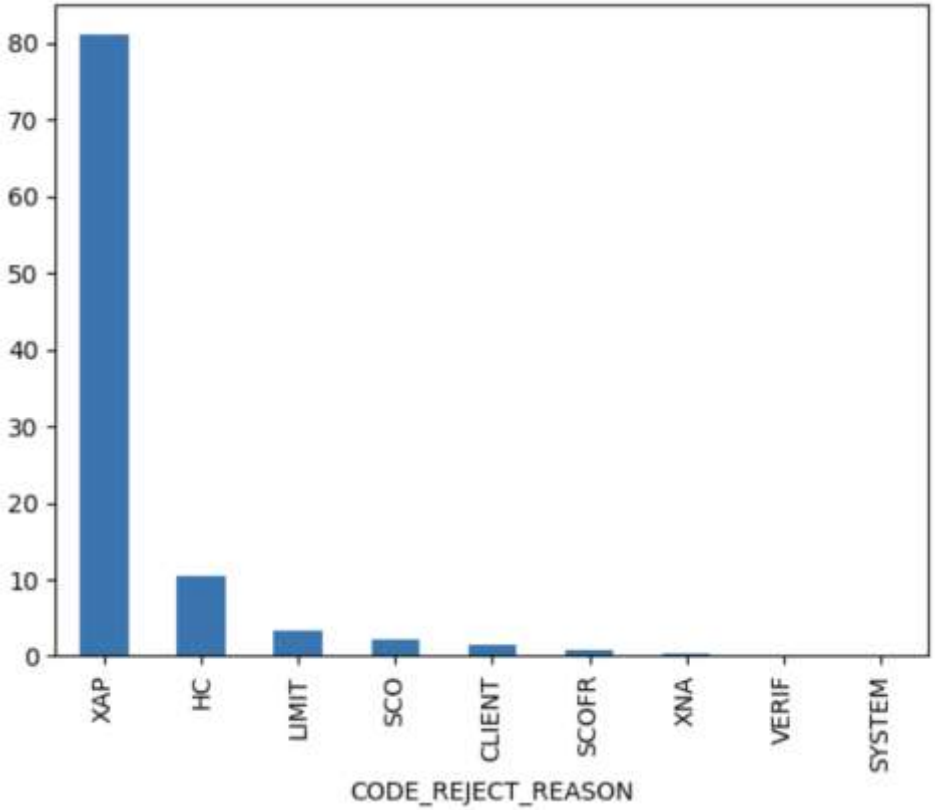




- Most of the clients were repeater and 18% clients were new.
- New applicants were less, but the acceptance rate is higher.

UNIVARIATE ANALYSIS

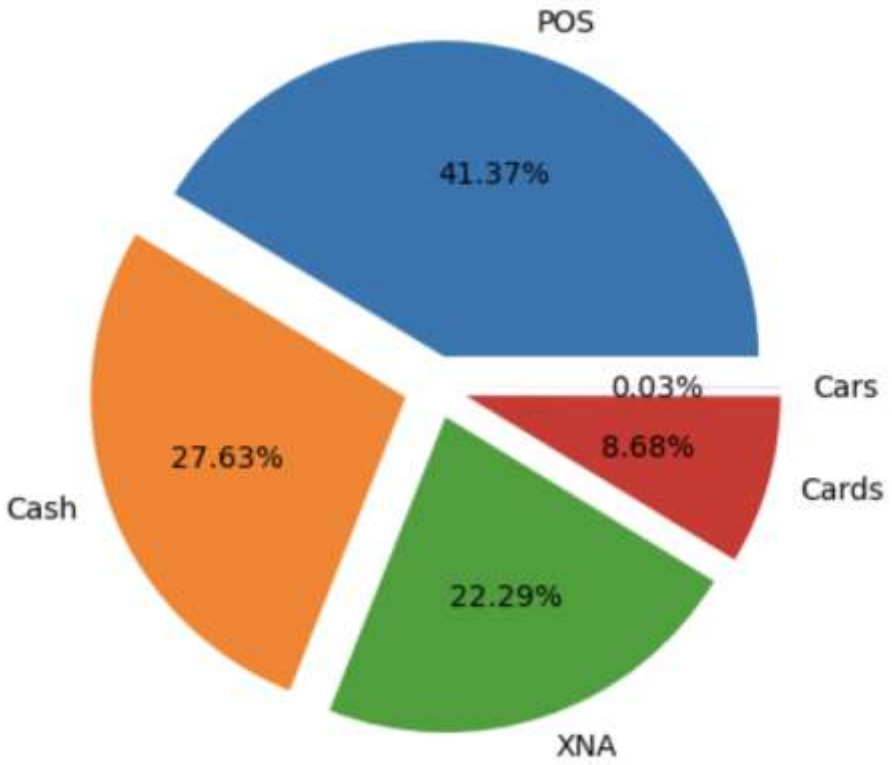
4. CODE_REJECT_REASON :



- Main reason for the Loan to get rejected is not recorded(XAP (81.01%)).

UNIVARIATE ANALYSIS

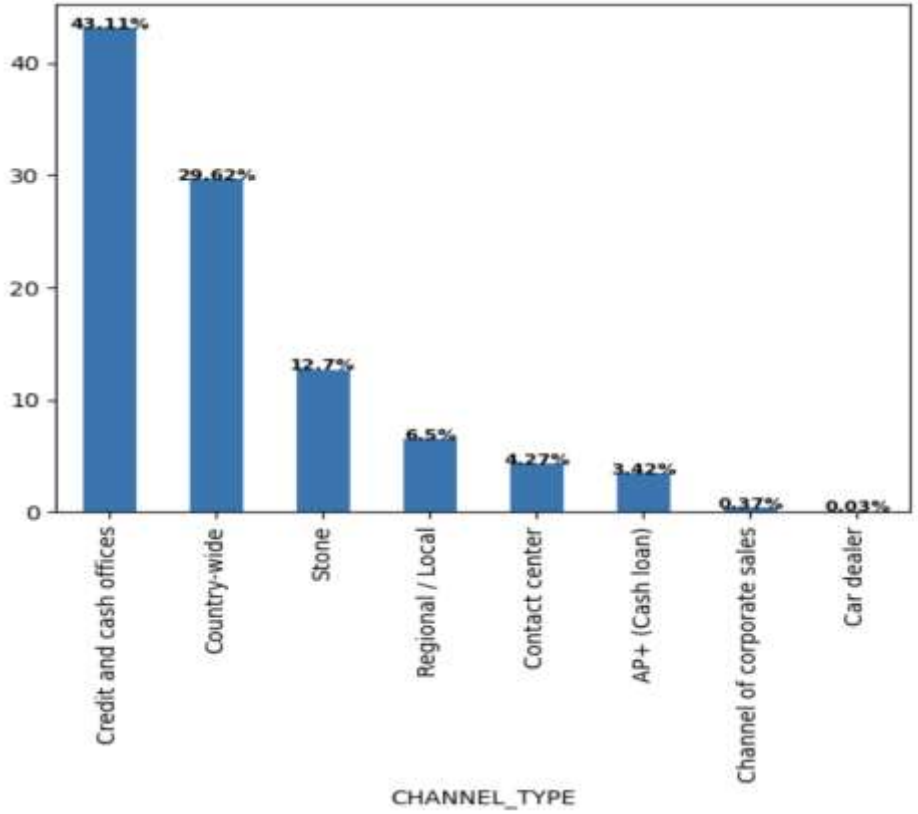
5. NAME_PORTFOLIO :



- Most of the previous applications were for POS(41.37%) and followed by Cash(27.63%).

UNIVARIATE ANALYSIS

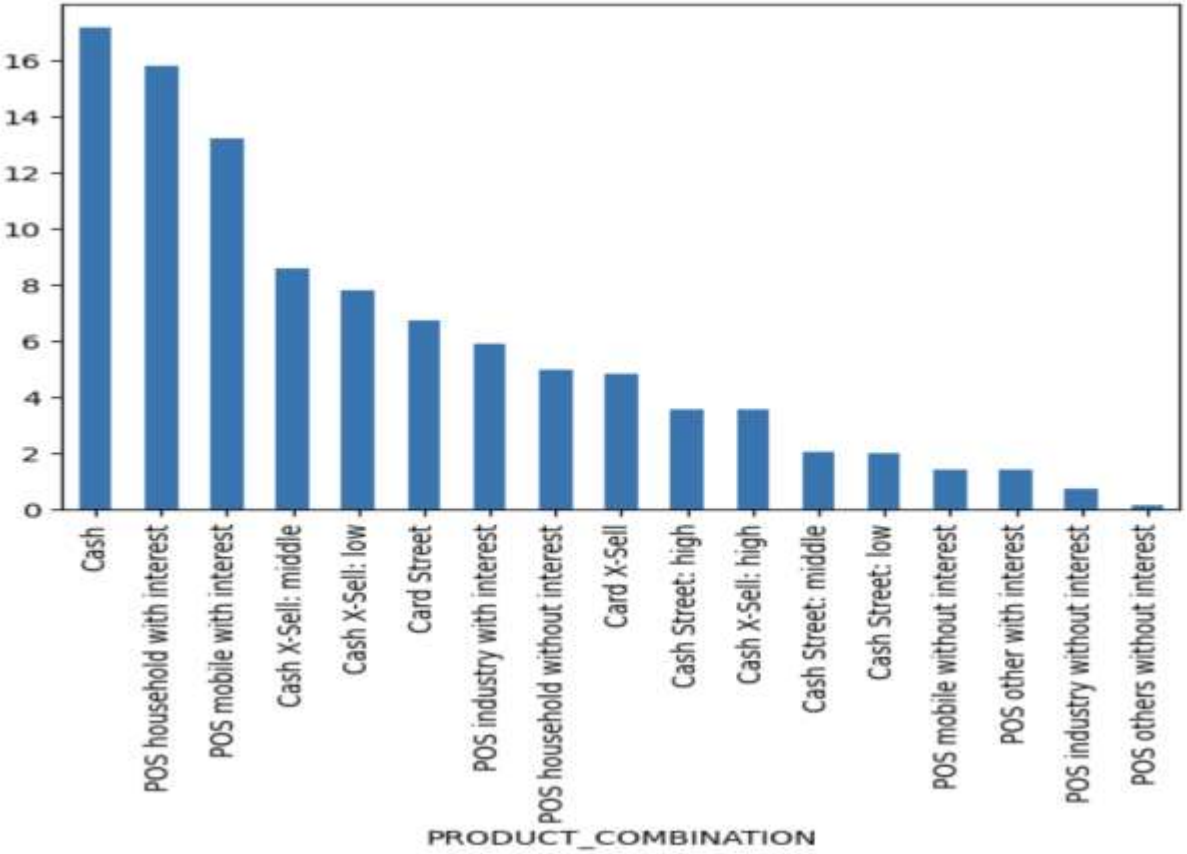
6. CHANNEL_TYPE :



- Most of the clients were acquired from credit and cash offices(43.11%) and followed by country-wide.

UNIVARIATE ANALYSIS

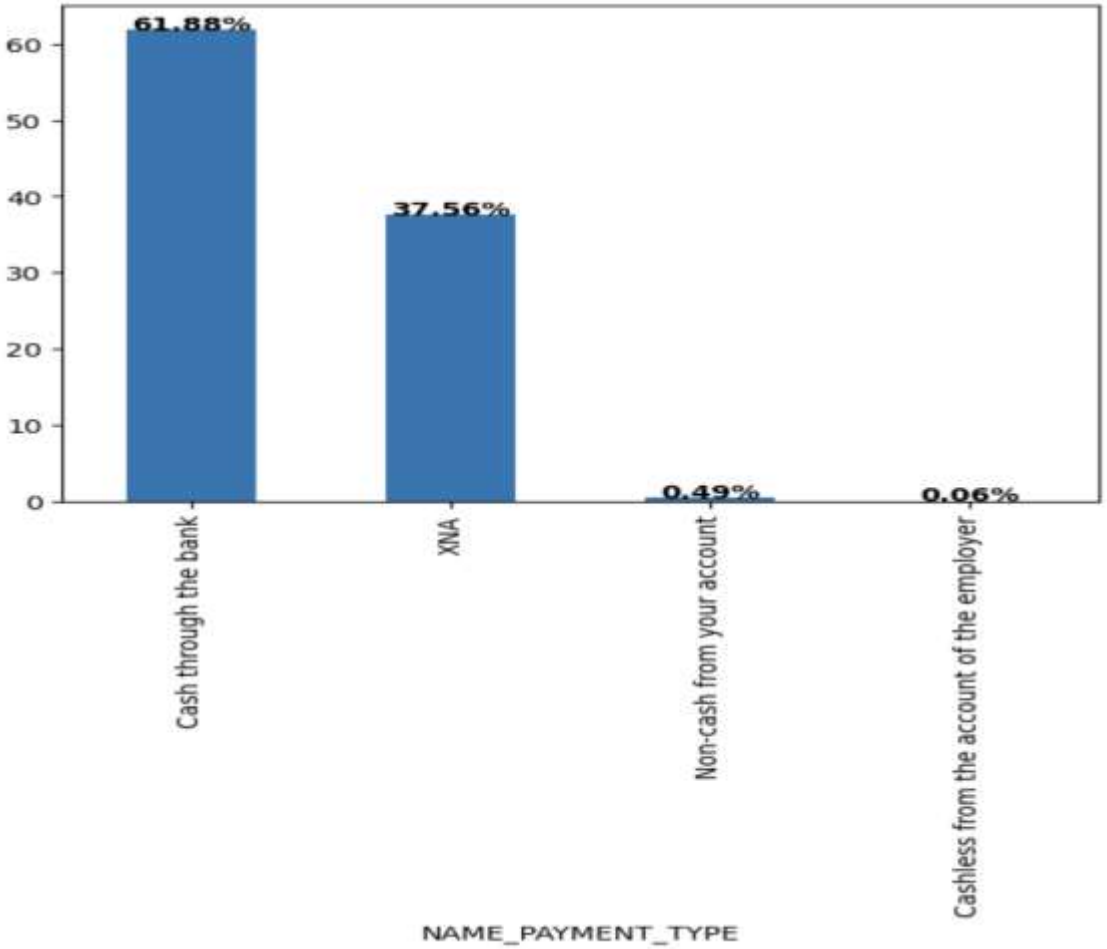
7. PRODUCT_COMBINATION :



- highest product combination is cash followed by POS household with interest and POS mobile with interest.

UNIVARIATE ANALYSIS

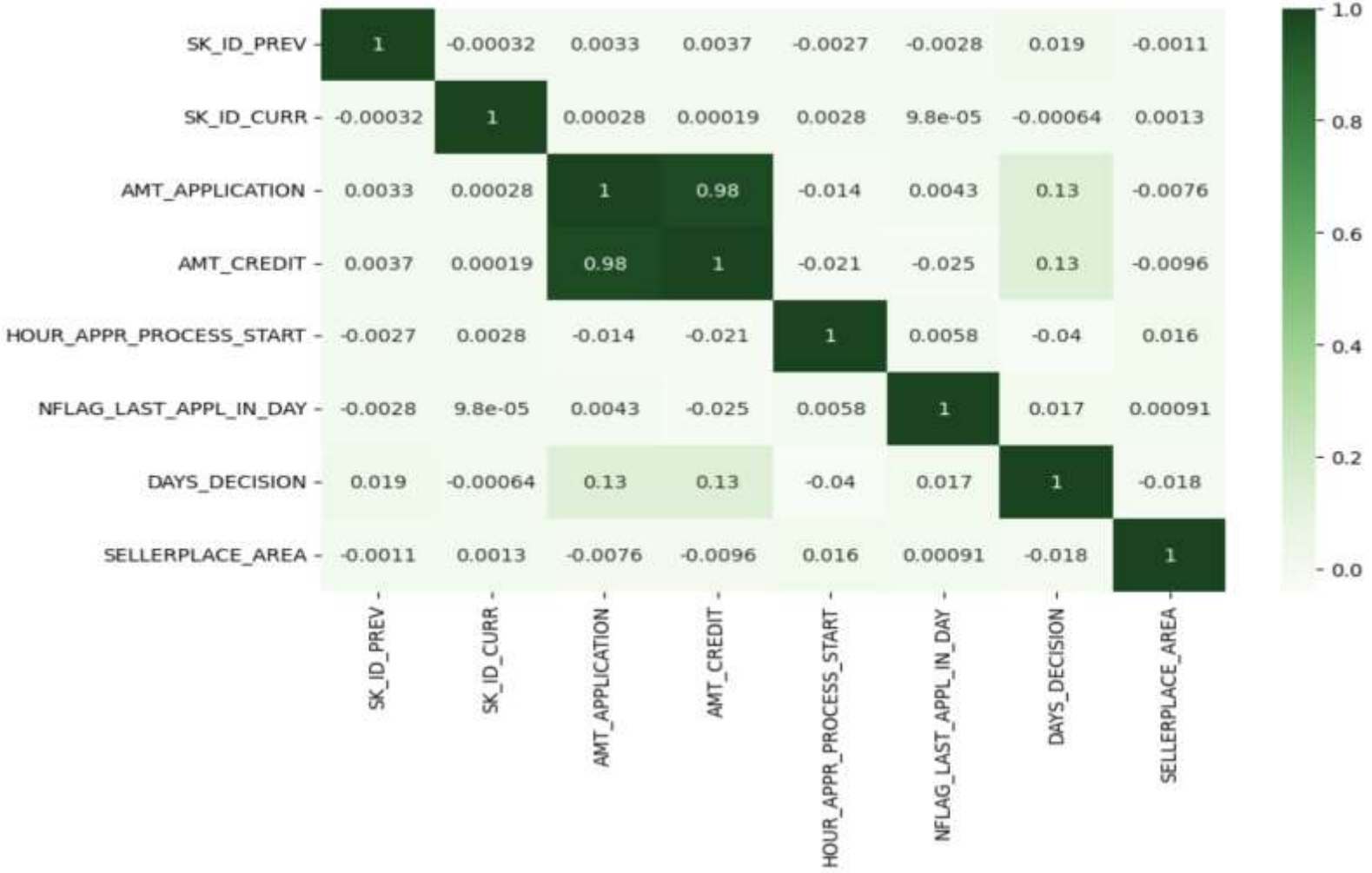
8. NAME_PAYMENT_TYPE :



- Most of the clients chose to pay cash through bank.

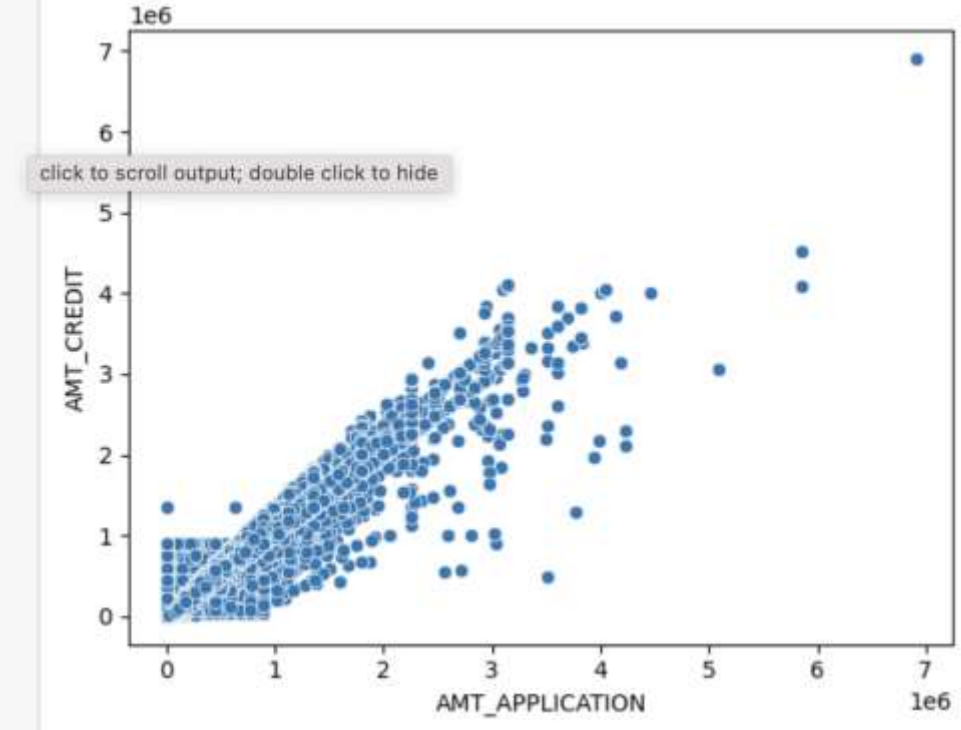
CORRELATION MATRIX

- Correlation matrix of previous application dataset.



CORRELATION MATRIX

- AMT_APPLICATION and AMT_CREDIT are highly correlated.



INSIGHTS

New Application dataset:

- Bank should give females more weitage when considering loans.
- Bank should be careful providing loan to client with regional rating 3.
- Client with more children has high chances of default.
- A lower income is also indicative of problem in payment.
- Client with secondary education and lower secondary have more issues in payment.
- Client Leaving which parents and prevented house are more likely to be default.

Previous Application dataset:

- Most of the customer choose the option and payment through bank.
- Many customers are repeaters.
- New applicants are less, but their acceptance rate is higher.
- Rate of acceptance of consumer loan are better than cash loans.
- The credit given to customer is proposal to goods price.
- Most of the clients were acquired from credit and cash offices.

INSIGHTS

- From allover data analysis.

- Bank should give out more revolving loans as client with revolving loans are less likely to default.
- Females have less default than mens.
- Consumer loans are most likely to approved by bank and It has less chances of default.
- There is a high correlation between goods, price and amount credited in both data set.