CREDIT EDA CASE STUDY

SUBMITTED BY RENUKA DESHMUKH AND MELITA VAS

INTRODUCTION

- The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specialises in lending various types of loans to urban customers. You have to use EDA to analyse the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.
- 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:

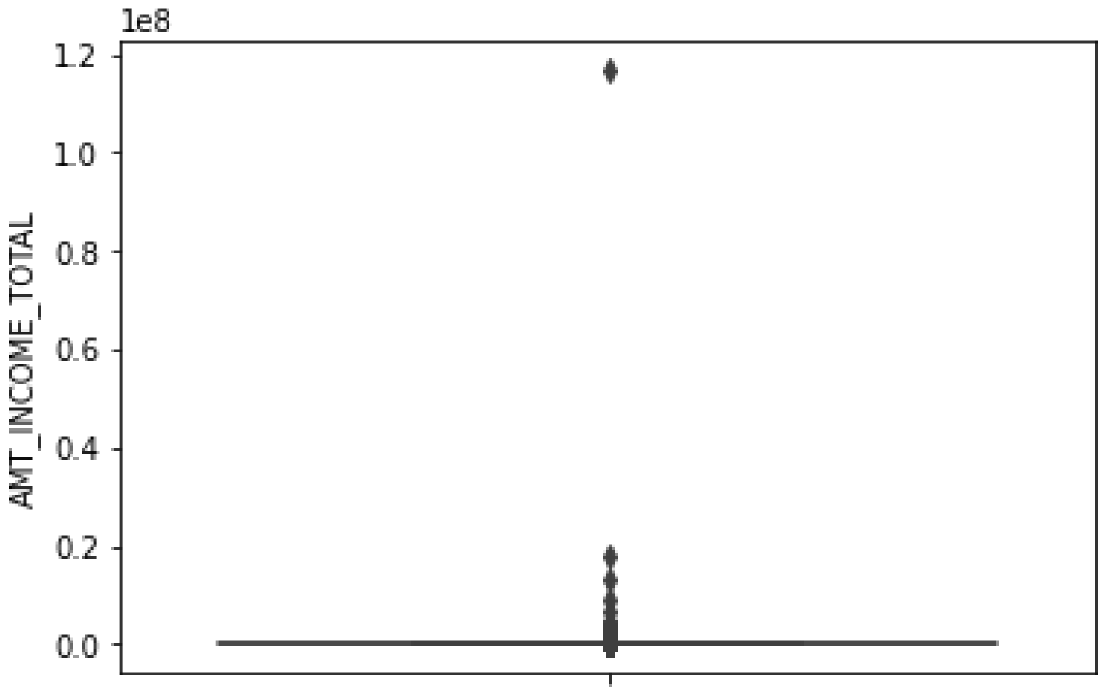
PROBLEM SOLVING APPROACH

- Cleaning the data and checking the data types
- Checking the imbalance between defaulters and non defaulters
- Univariate Analysis
- Bivariate Analysis
- Conclusions

OUTLIERS

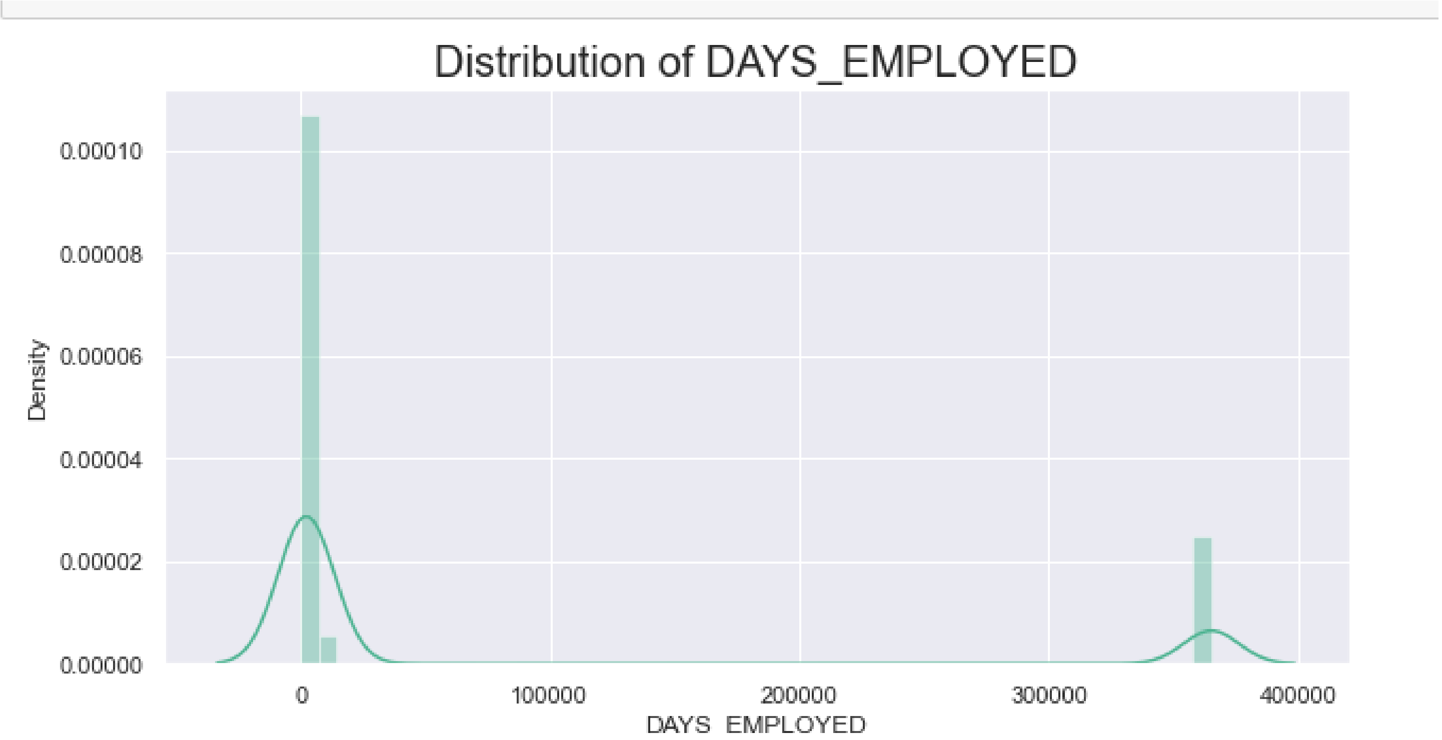
- Here in the column 'AMT_INCOME_TOTAL' which tells us the income of the client .We observed a value around 120M which is surely an outlier

```
]: sns.boxplot(y=df['AMT_INCOME_TOTAL'])
   plt.show()
```



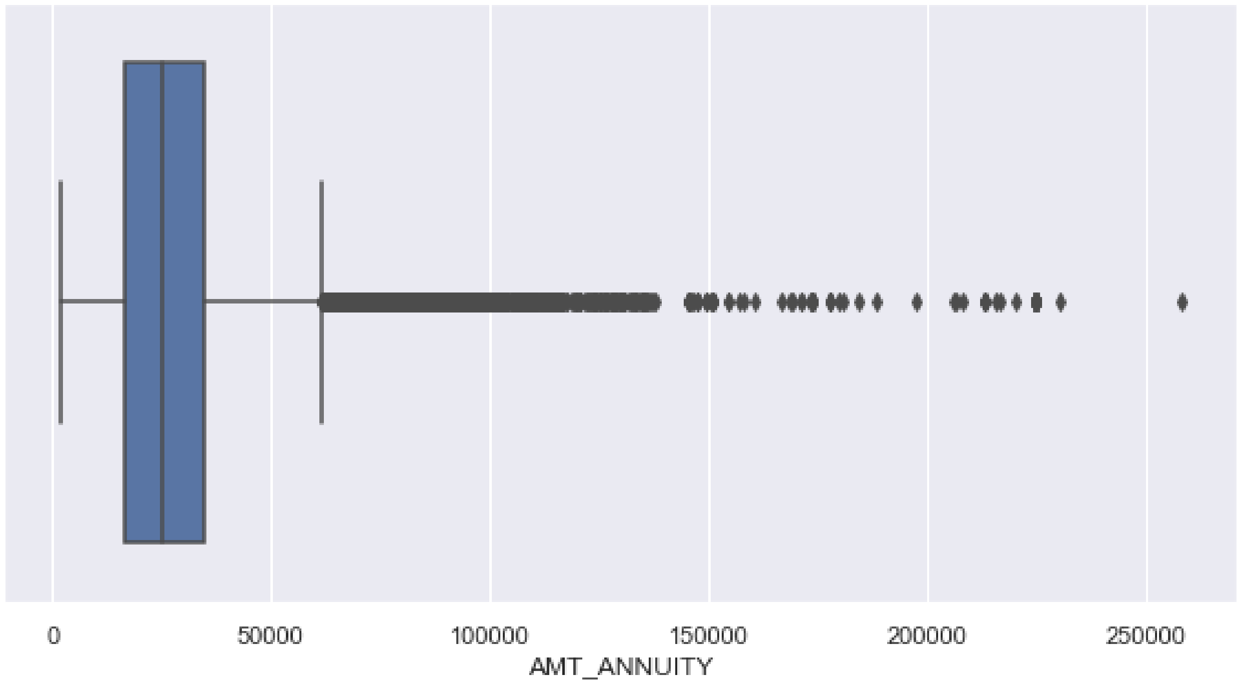
OUTLIERS

- We observe a value which is greater than 20000 which is surely an outlier because 250000/365 will be around 54 years. Considering that a person started working at age of 21, the person will be 21+54 will turn out to be 75 years



OUTLIERS

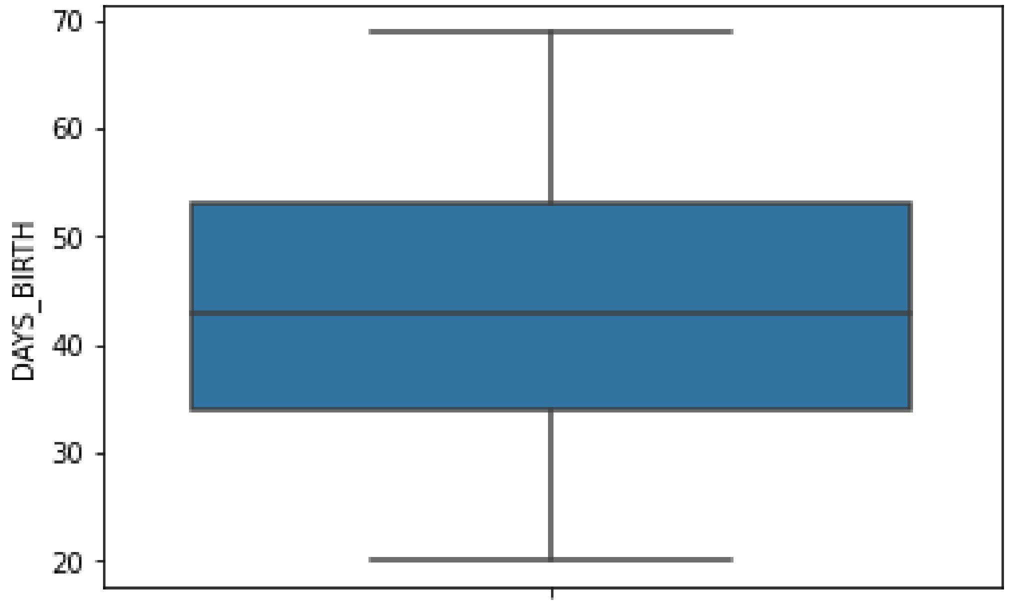
- The value greater than 250000 is an outlier



OUTLIERS

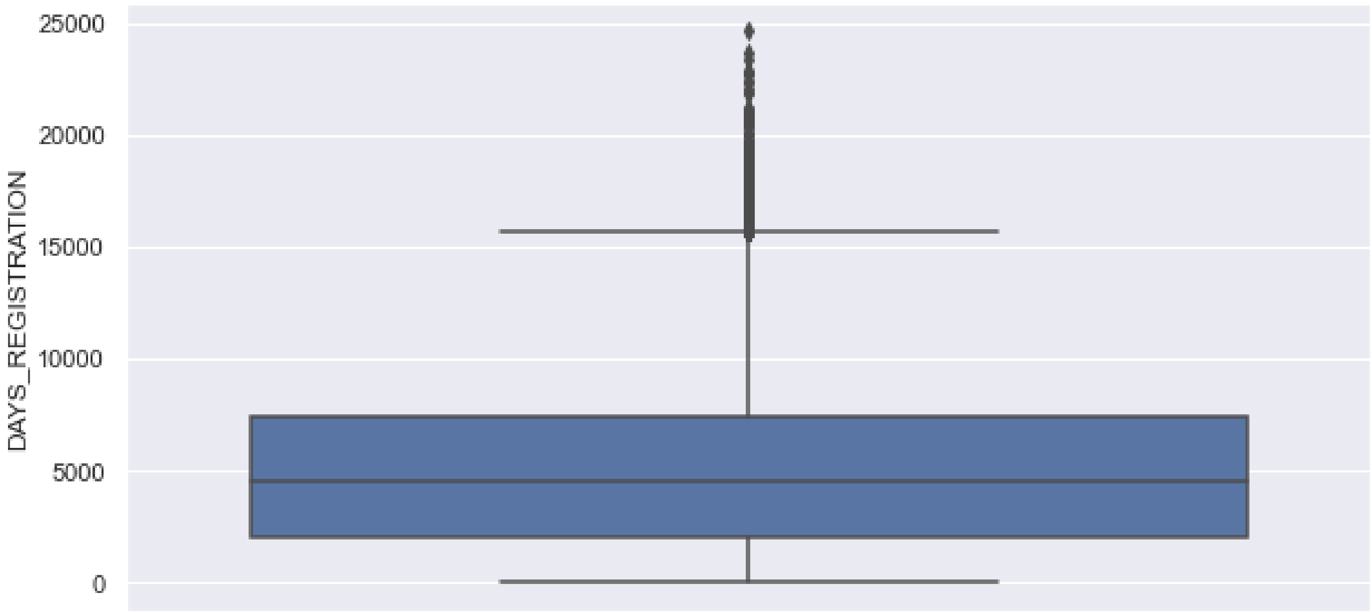
- No outliers in this case

```
sns.boxplot(y=df['DAYS_BIRTH'])
plt.show()
```



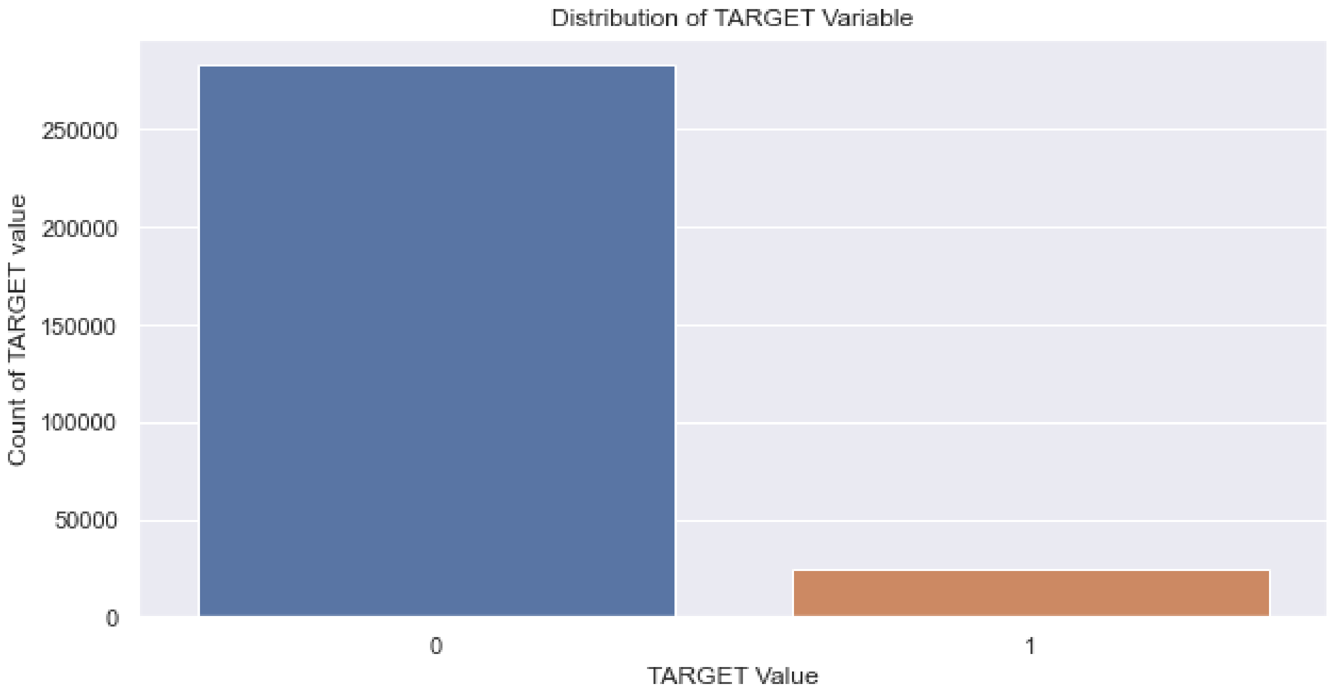
OUTLIER

- No Outlier as such



IMBALANCE BETWEEN TARGET VARAIBLE

- The data is highly disbalance as the count of defaulters is less than 50000 and other 250000



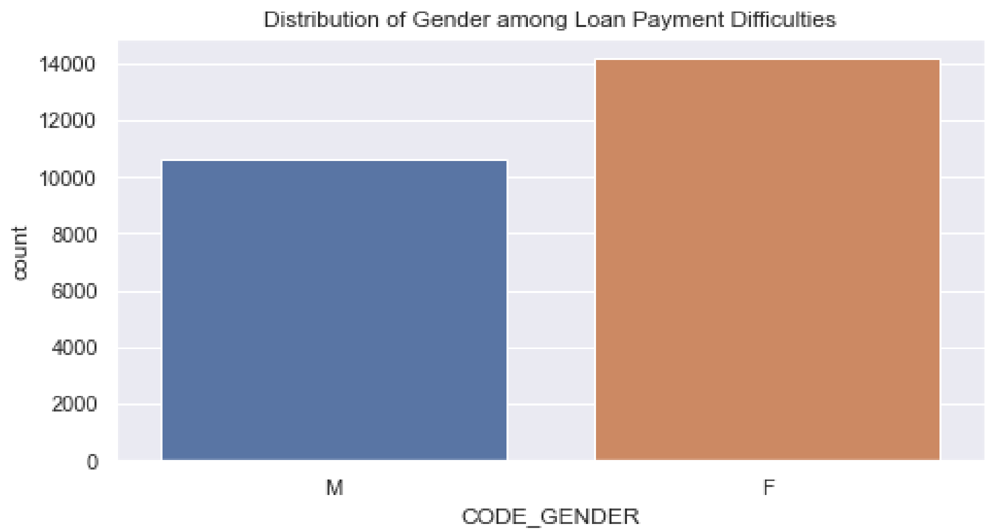
UNIVARIATE CATEGORICAL ANALYSIS

FOR LOAN PAYMENT DIFFICULTIES

- Female count is much more than male in case of defaulters

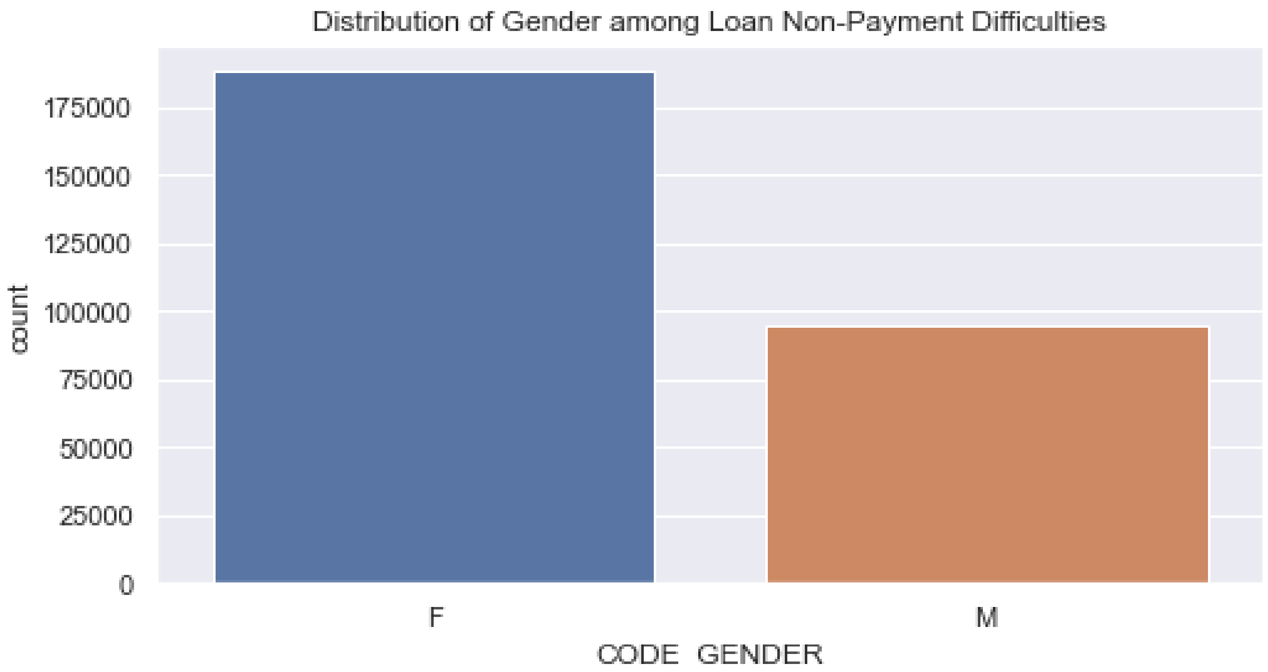
```
In [240]: #Checking the gender distribution among the defaulters
plt.figure(figsize=(8,4))
sns.countplot(x=df_1['CODE_GENDER'])
plt.title("Distribution of Gender among Loan Payment Difficulties")

Out[240]: Text(0.5, 1.0, 'Distribution of Gender among Loan Payment Difficulties')
```



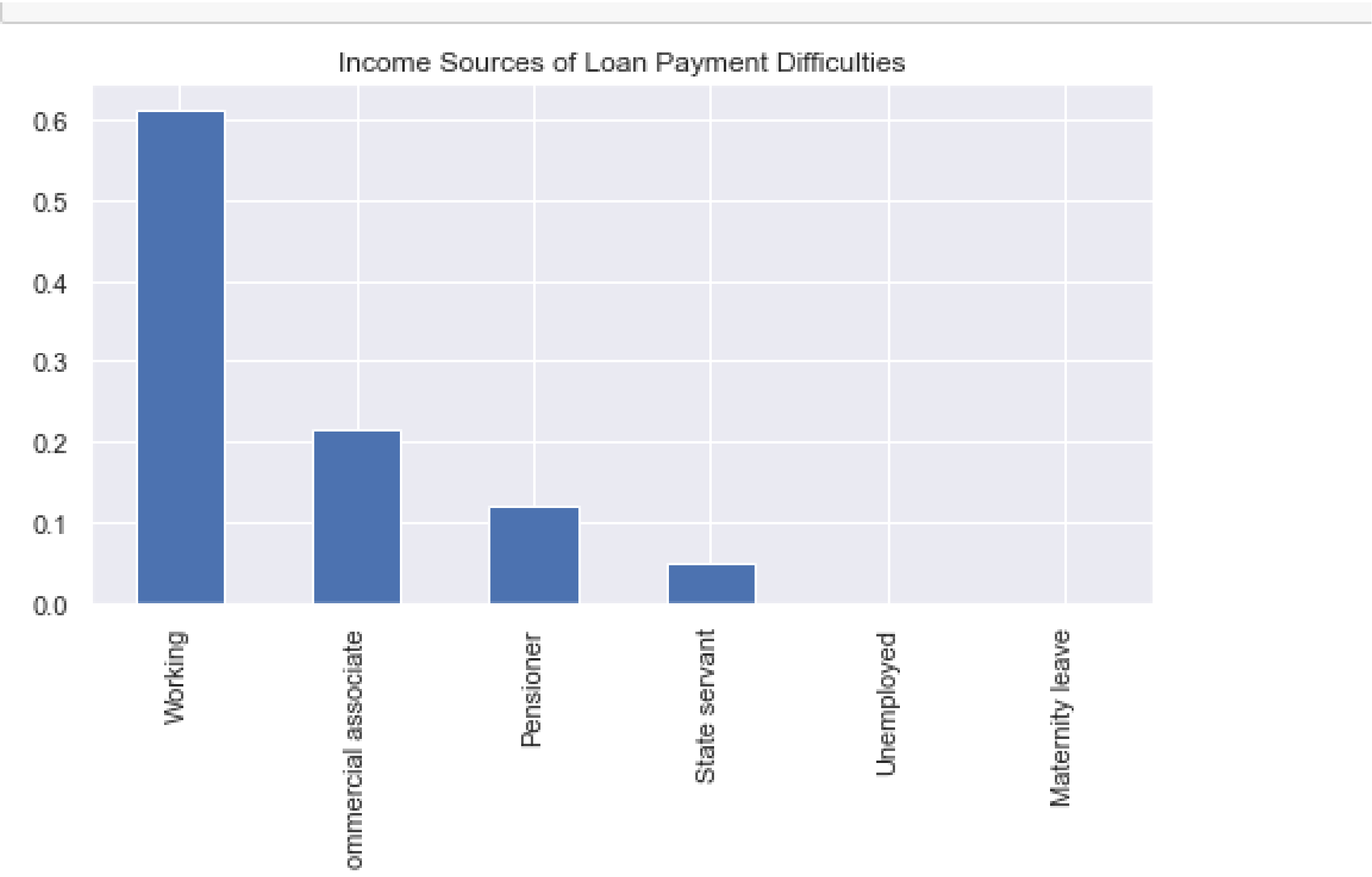
FOR LOAN NON-PAYMENT DIFFICULTIES

- Comparing the Payment Difficulties and Non-Payment Difficulties based on Gender we observe that Females are the majority in both the cases although there is increase in the percentage in Male Payment Difficulties from Non-Payment Difficulties



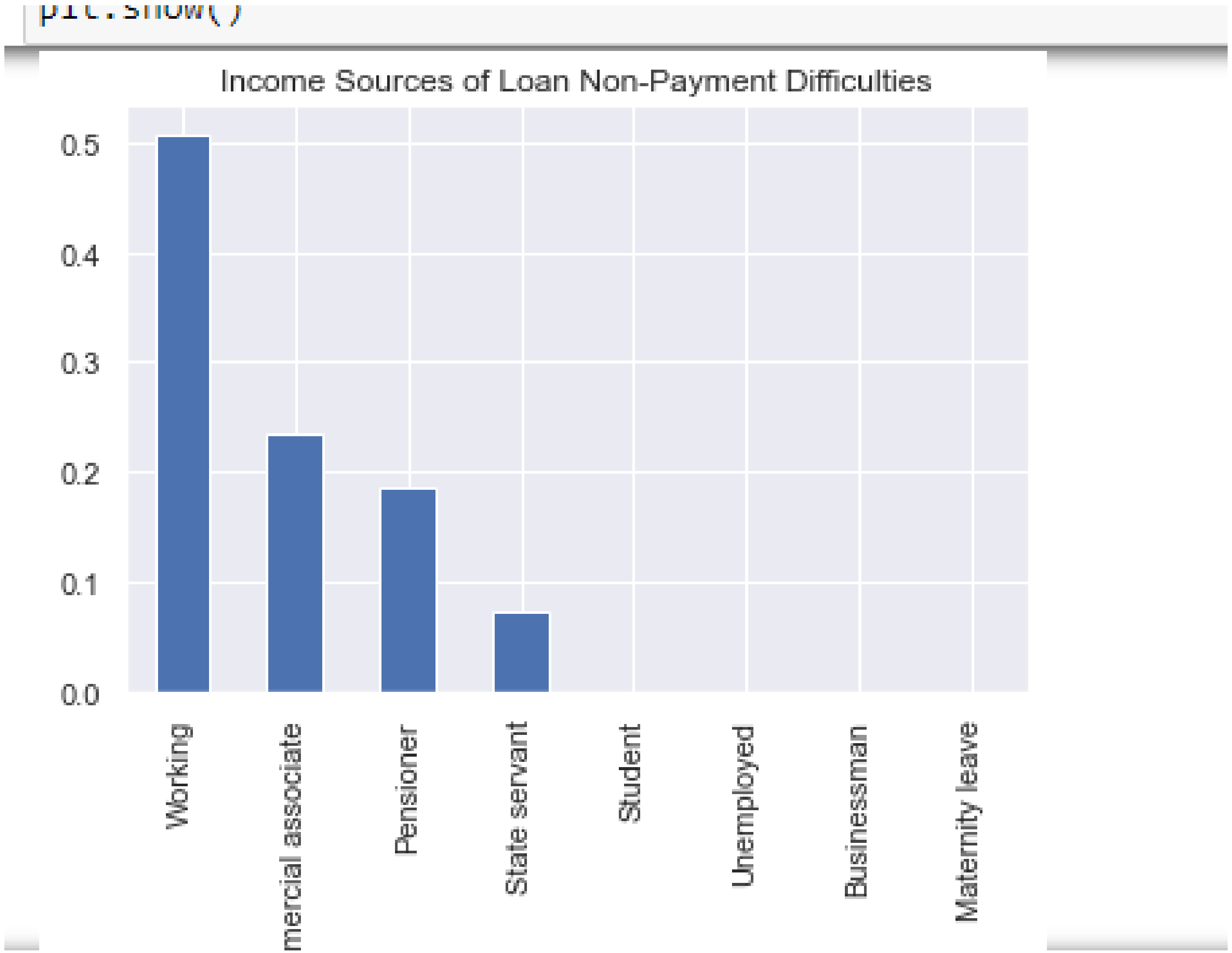
FOR LOAN PAYMENT DIFFICULTIES

- Working professionals is higher as compared to other categories



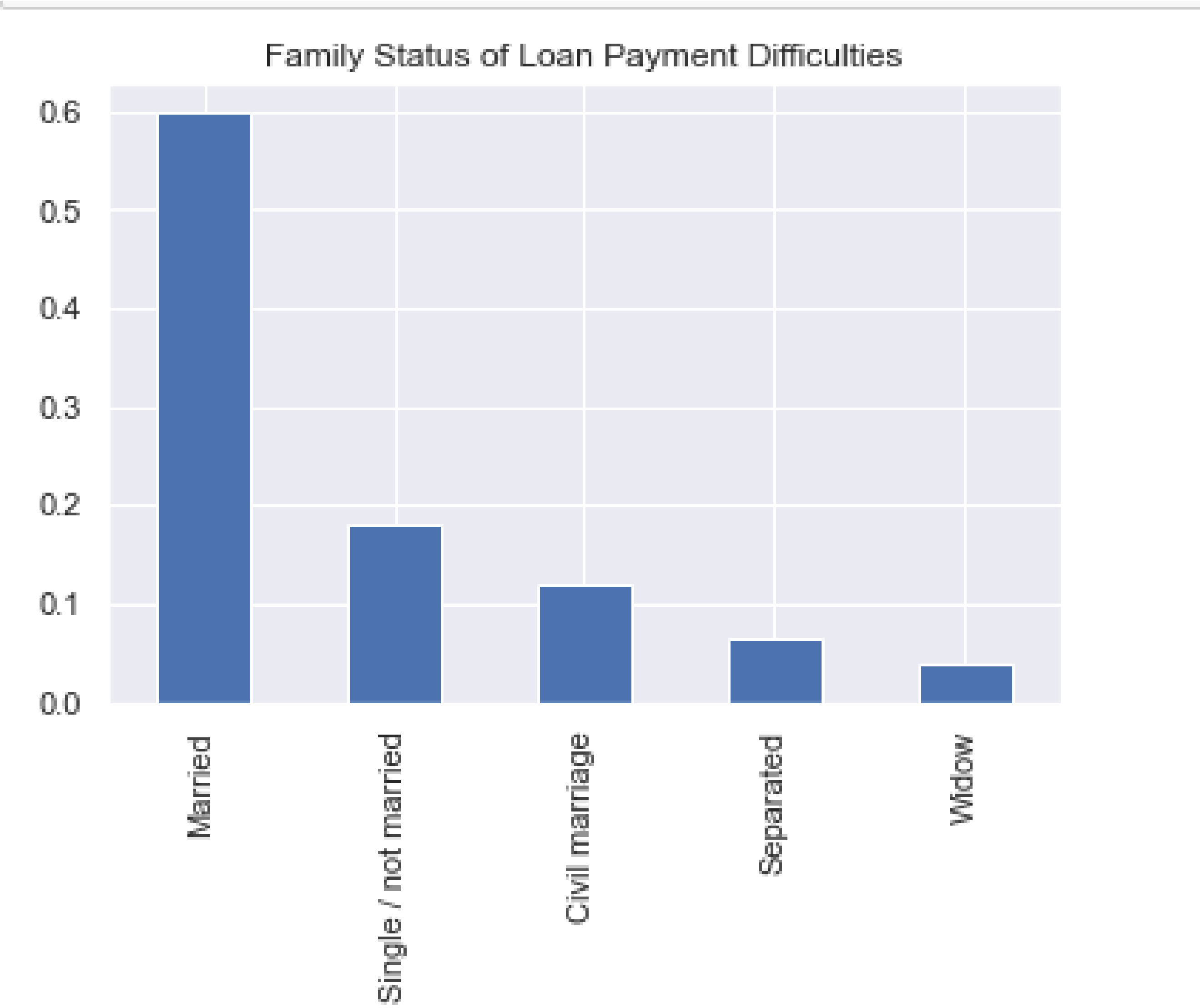
FOR LOAN NON-PAYMENT DIFFICULTIES

- Working Professionals are prominent in both Loan Payment Difficulties and Loan Non-Payment Difficulties. While there is decrease in percentage of Pensioner in Loan Payment difficulties



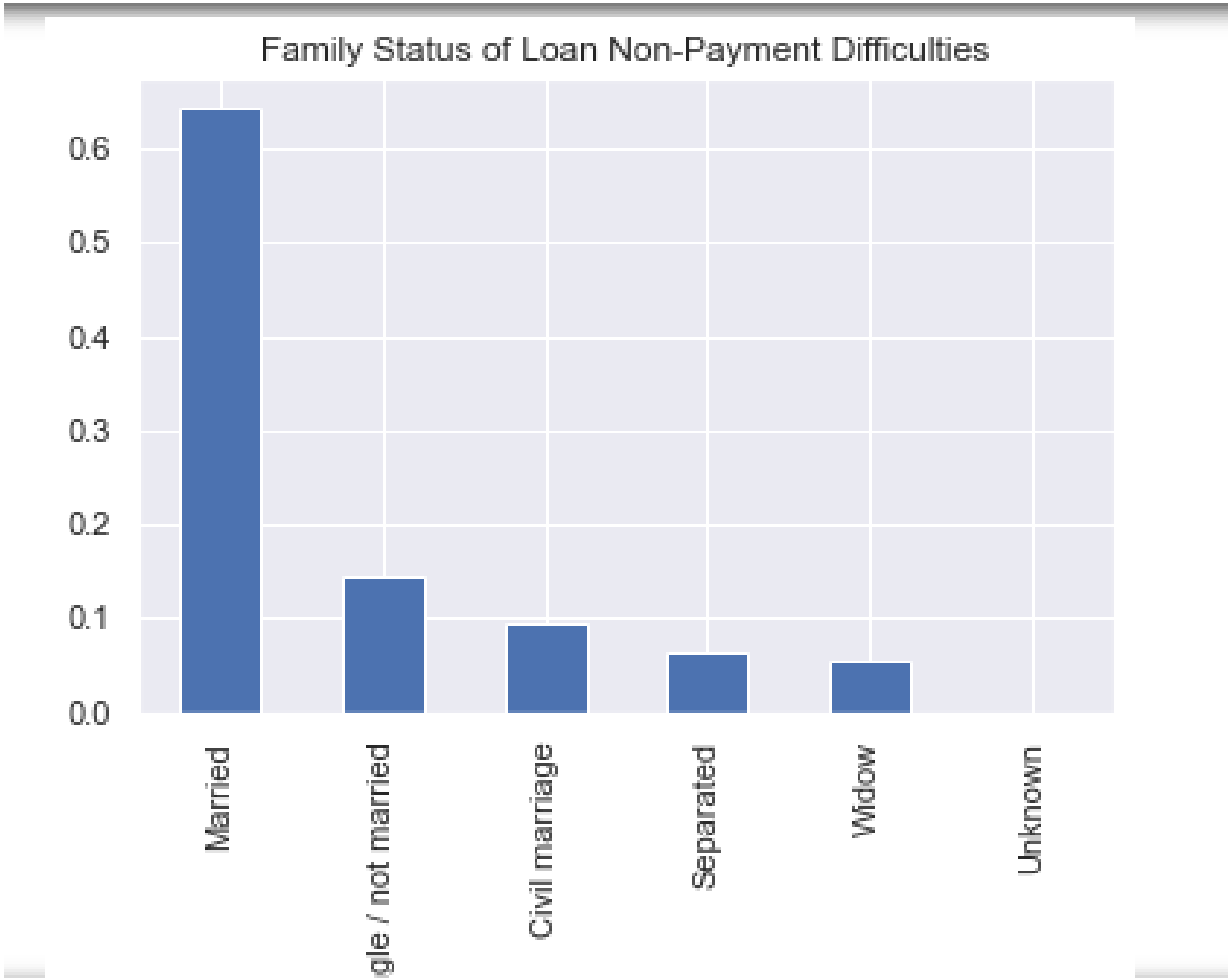
FOR LOAN PAYMENT DIFFICULTIES

- Married Clients are higher in case of defaulters



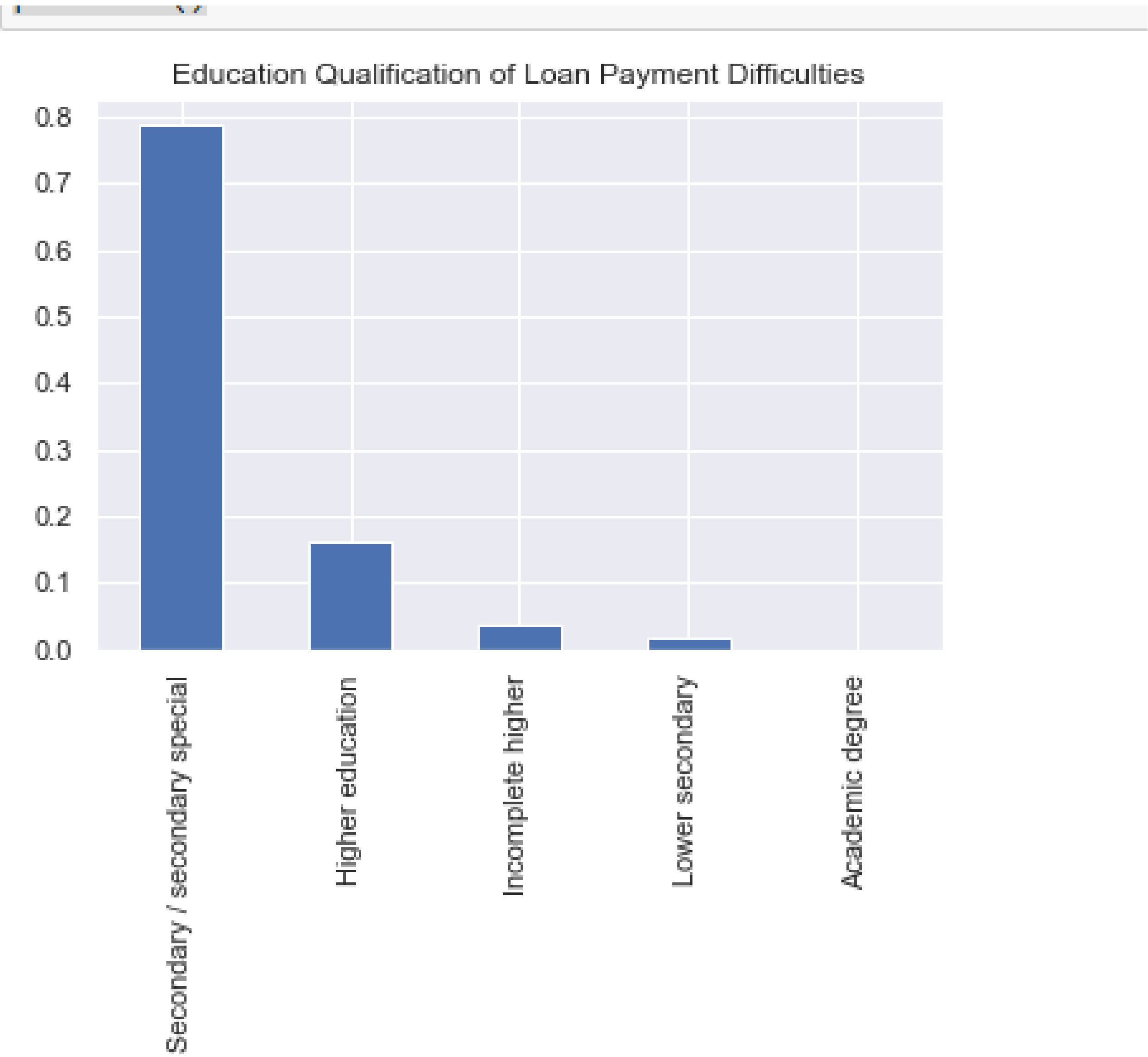
FOR LOAN NON-PAYMENT DIFFICULTIES

- We observe a decrease in the percentage of married and widowed with Loan Payment Difficulties and an increase in the percentage of single and civil married with Loan Payment Difficulties when compared with the percentage of both Loan Payment Difficulties and Loan Non-Payment Difficulties



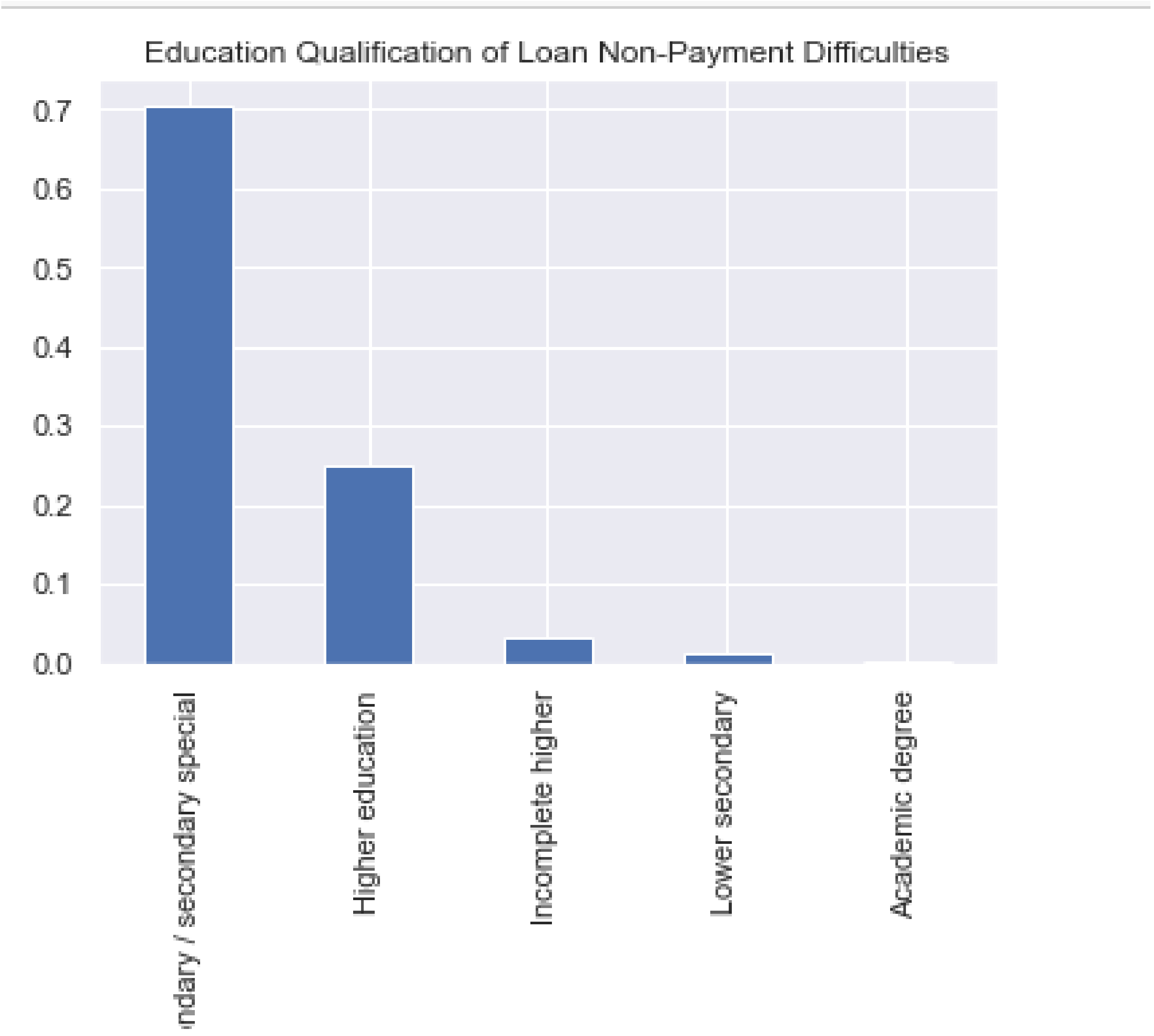
FOR LOAN PAYMENT DIFFICULTIES

- Academic degree are none in the graph that the higher the education the higher the degree and pays loan amount

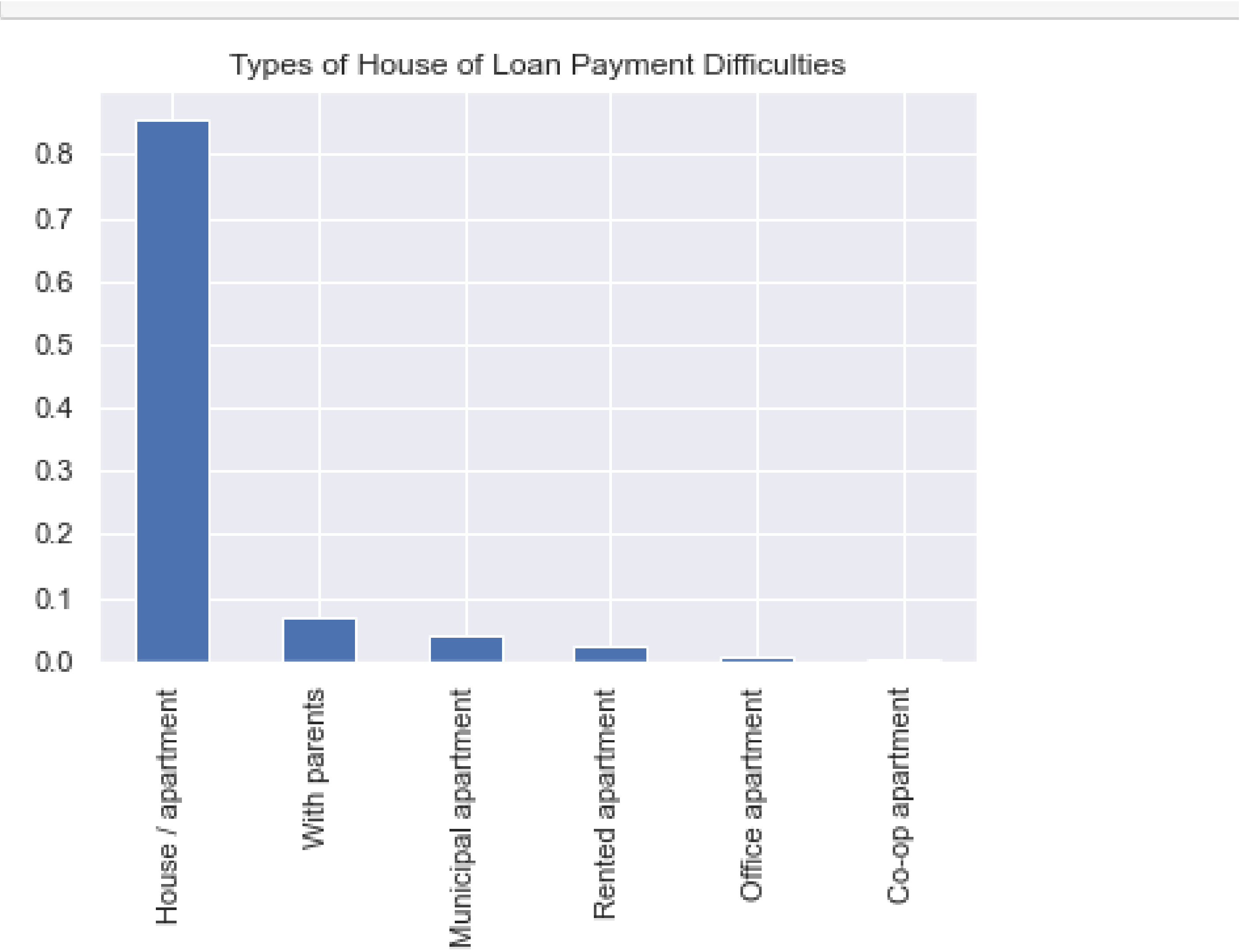


FOR LOAN NON-PAYMENT DIFFICULTIES

- We observe that there is decrease in number of secondary and secondary special in Loan Non-Payment Difficulties as compared to Loan Payment Difficulties

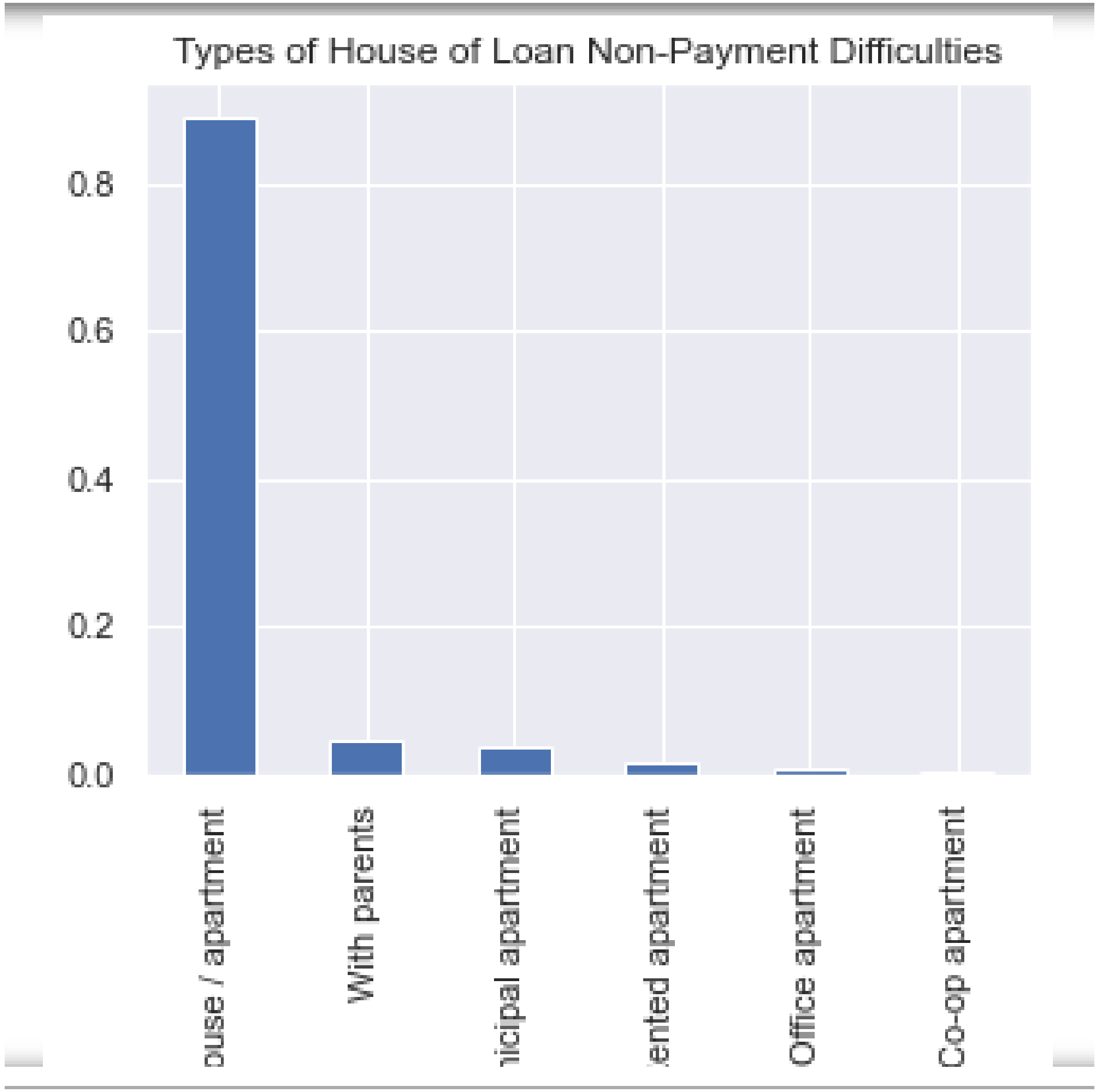


FOR LOAN PAYMENT DIFFICULTIES



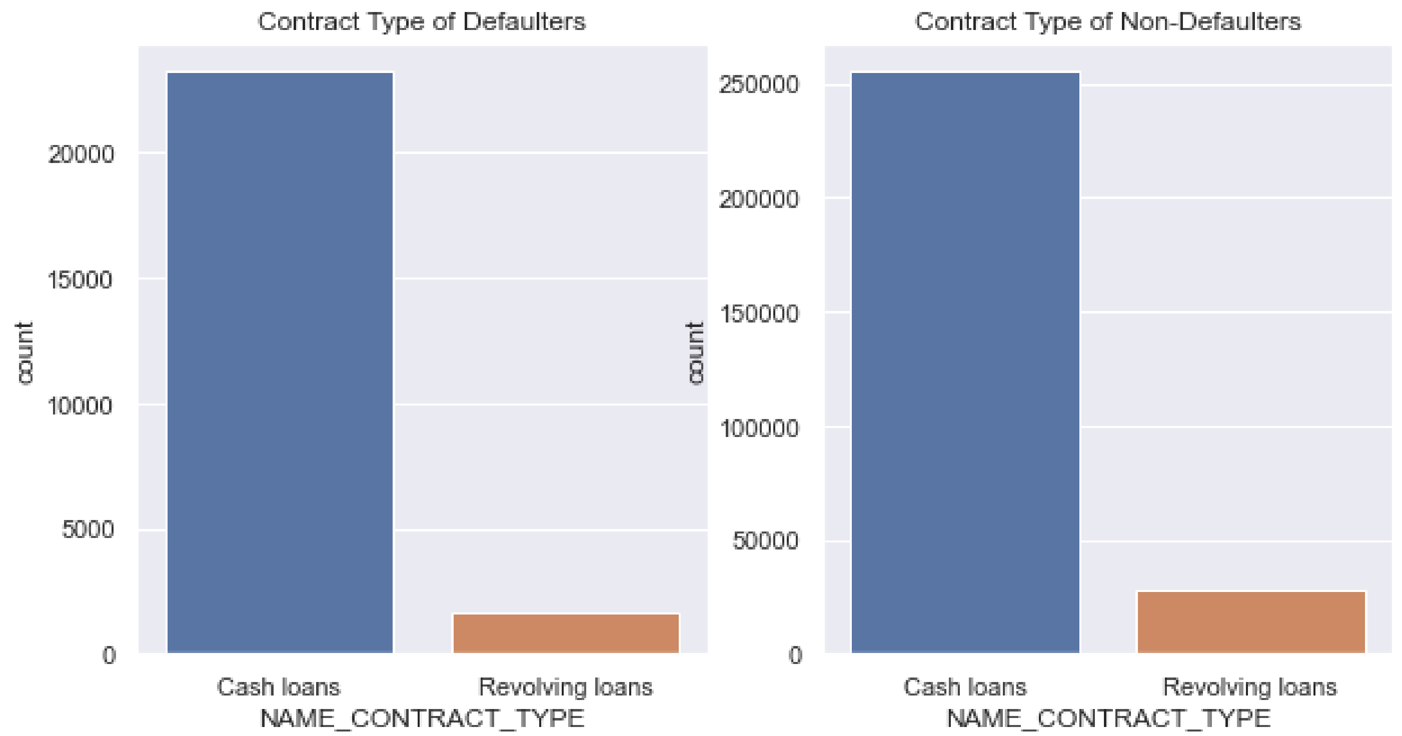
FOR LOAN NON-PAYMENT DIFFICULTIES

We observe that an increase in the percentage of Payment Difficulties who live with their parents when compared to the Percentage of Payment Difficulties and Non-Payment Difficulties

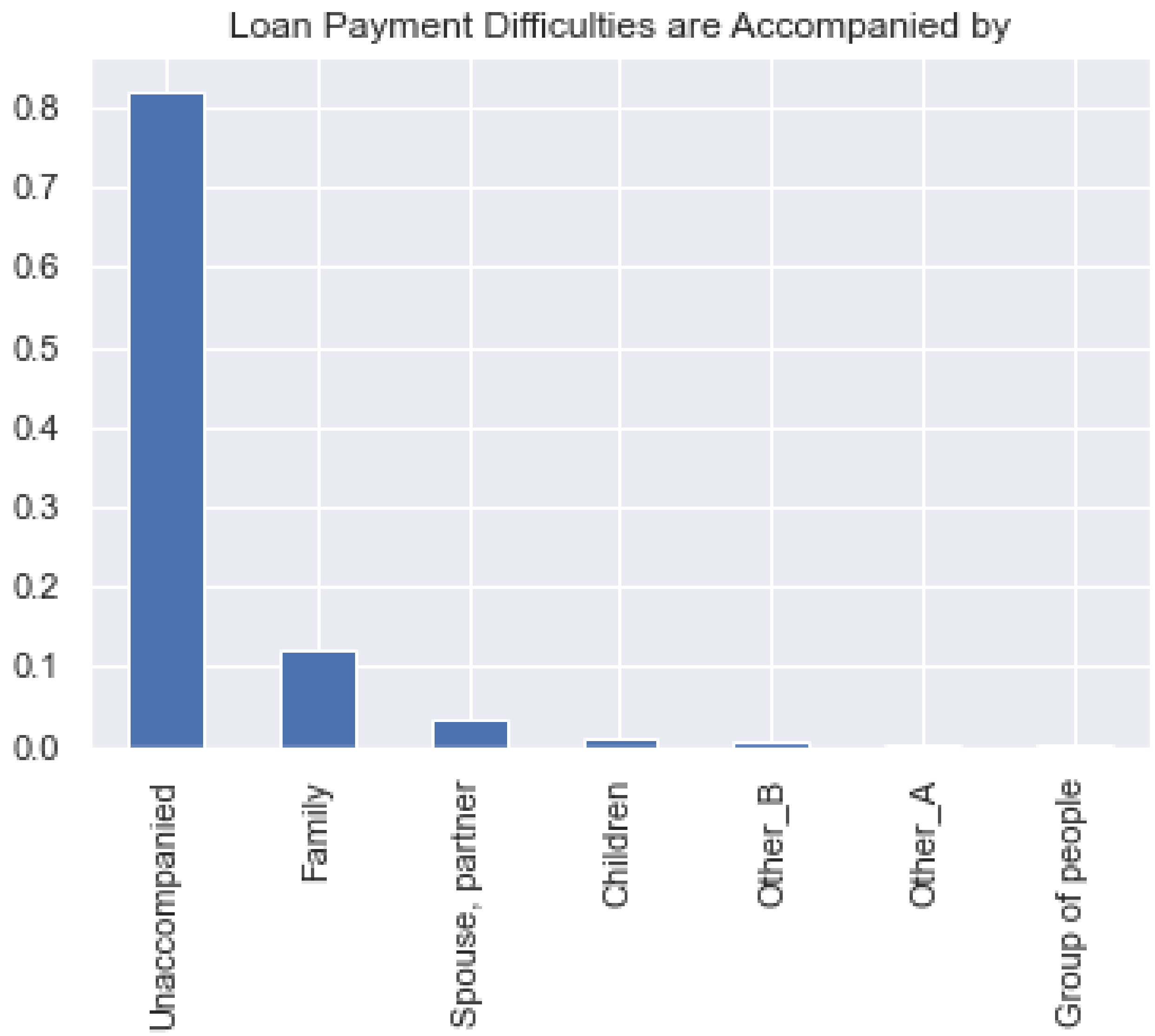


CONTRACT TYPE FOR BOTH DEFAULTERS AND OTHERS

- We observe there is increase in both the Cash and Revolving loans in loan Non-Payment Difficulties



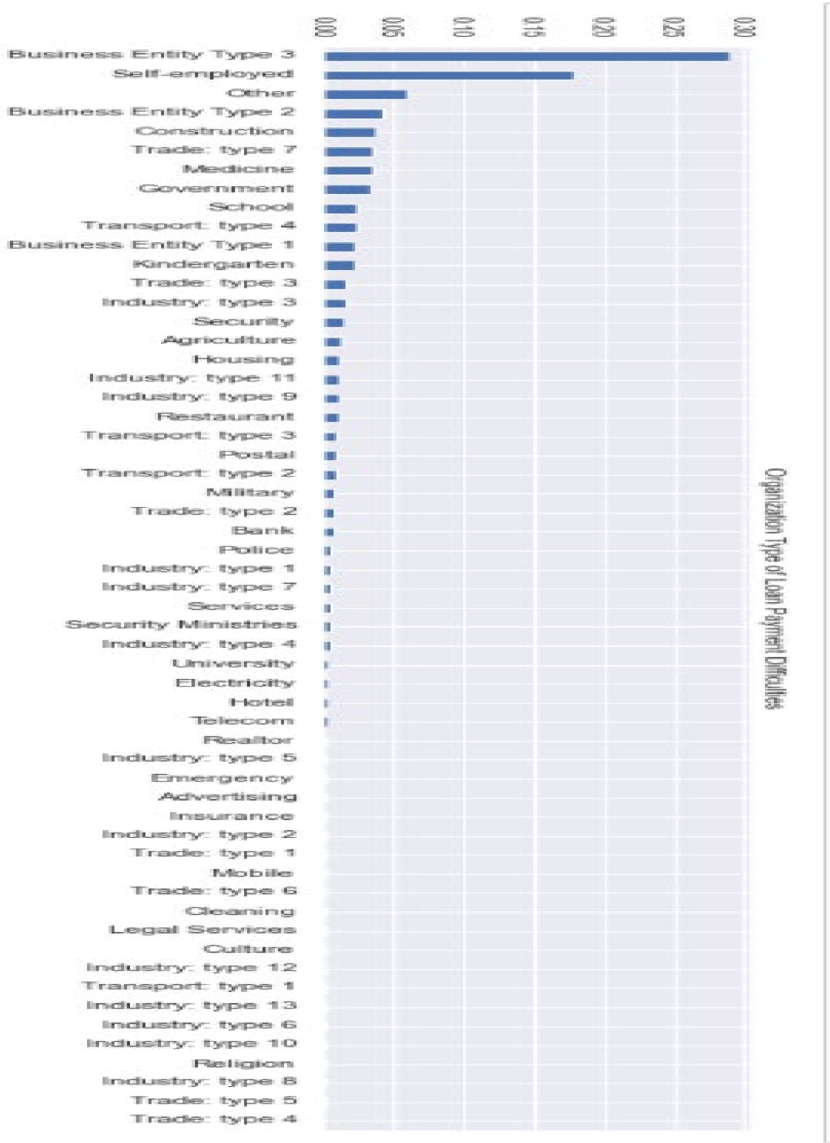
FOR LOAN PAYMENT DIFFICULTIES



FOR LOAN NON-PAYMENT DIFFICULTIES

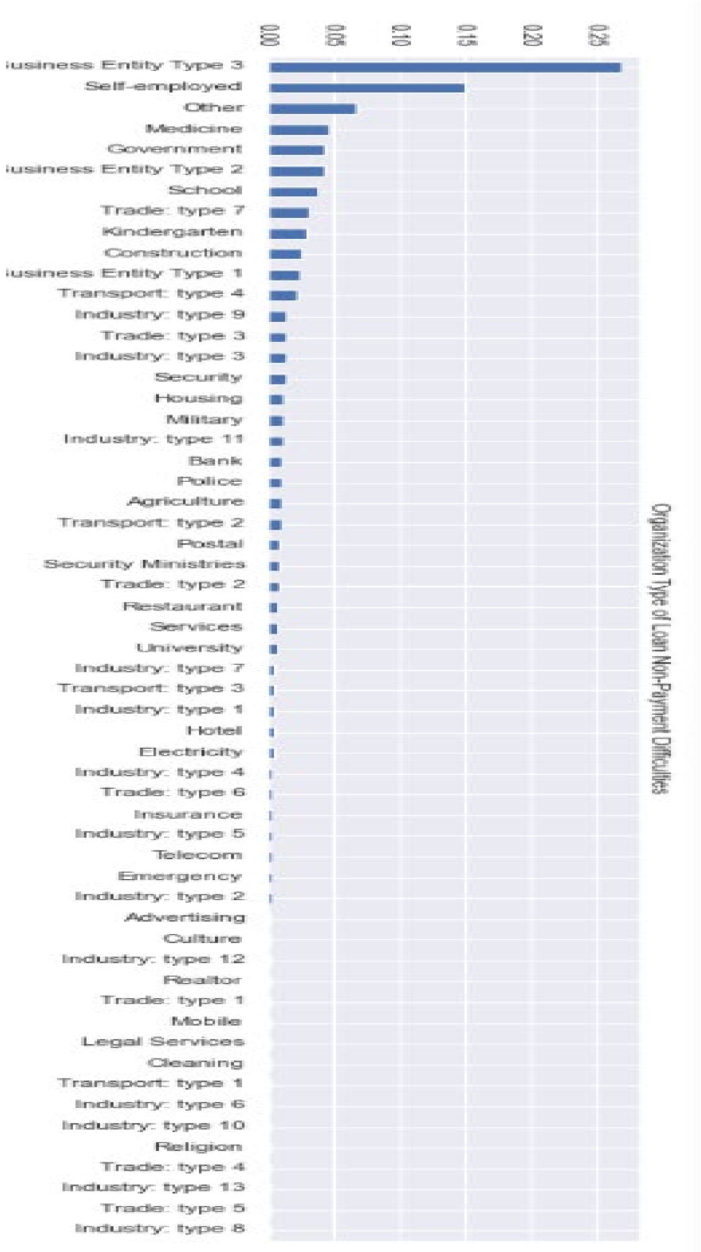
- We do see any major observations

FOR LOAN PAYMENT DIFFICULTIES

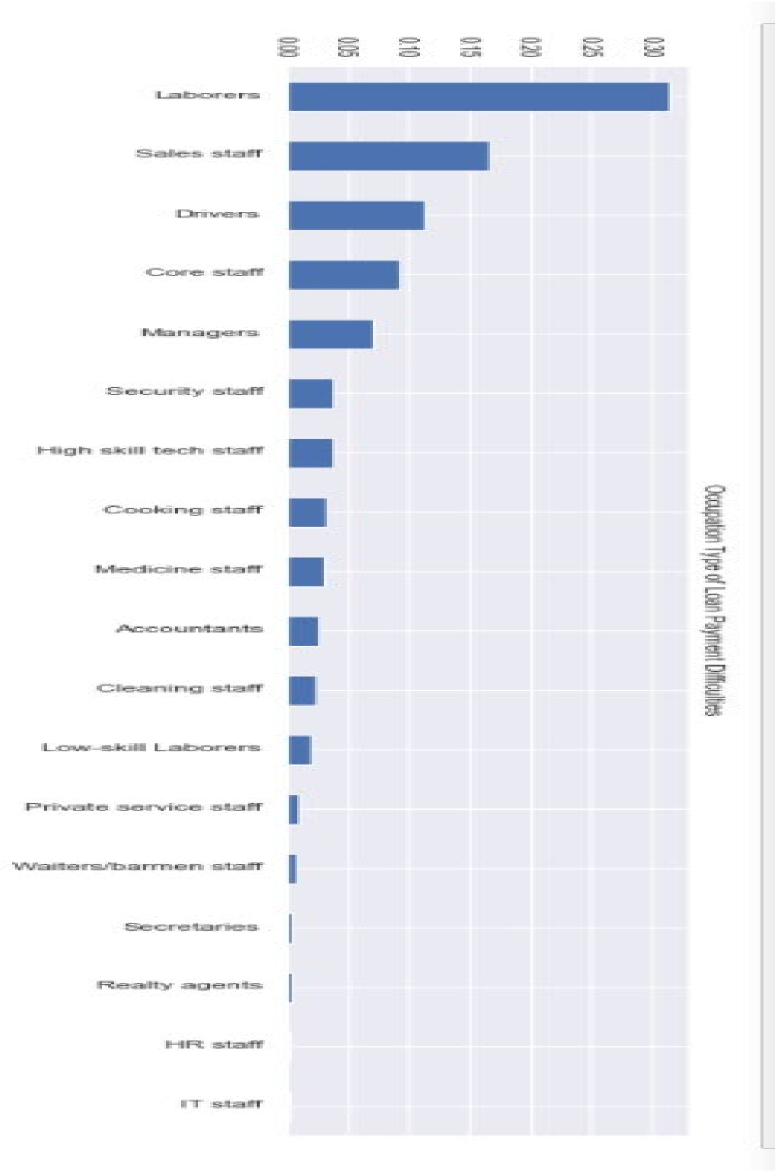


FOR LOAN NON-PAYMENT DIFFICULTIES

- Self Employed and Business Entity Type percentage is more in Defaulters as compared to Loan Non-Payment Difficulties

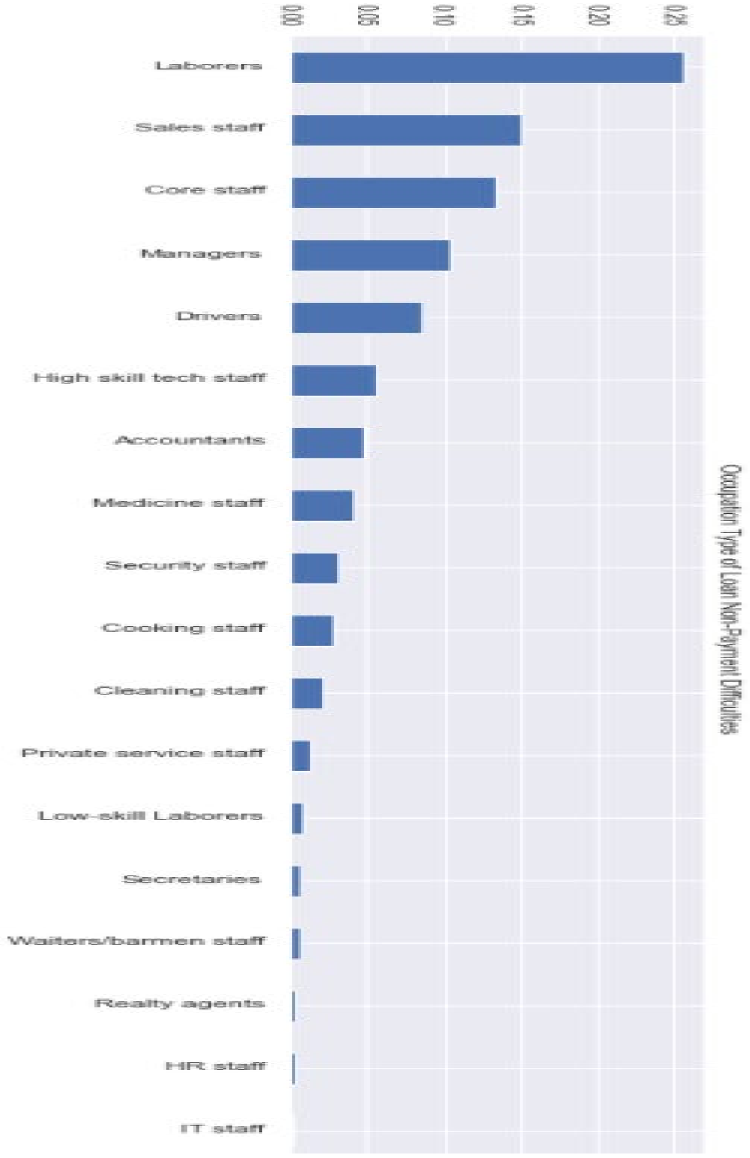


OCCUPATION TYPE FOR LOAN PAYMENT DIFFICULTIES

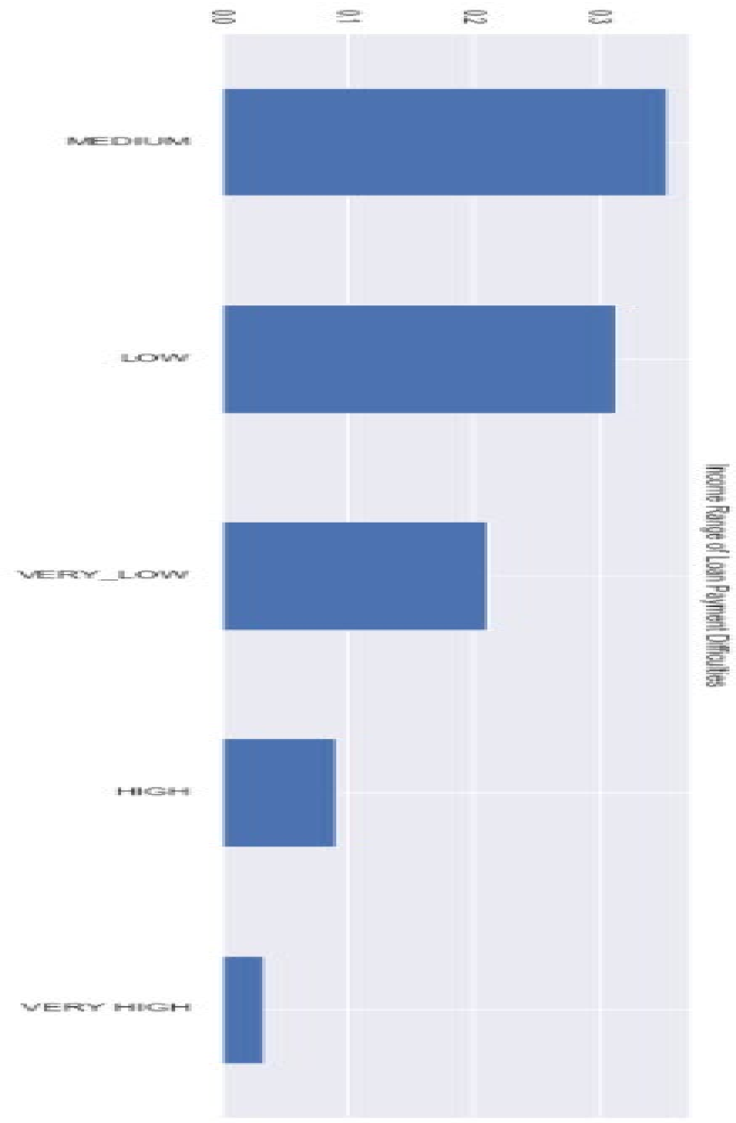


OCCUPATION TYPE FOR LOAN NON PAYMENT DIFFICULTIES

- We observe laborers are high in both the cases but laborers percentage is increased in case of defaulters

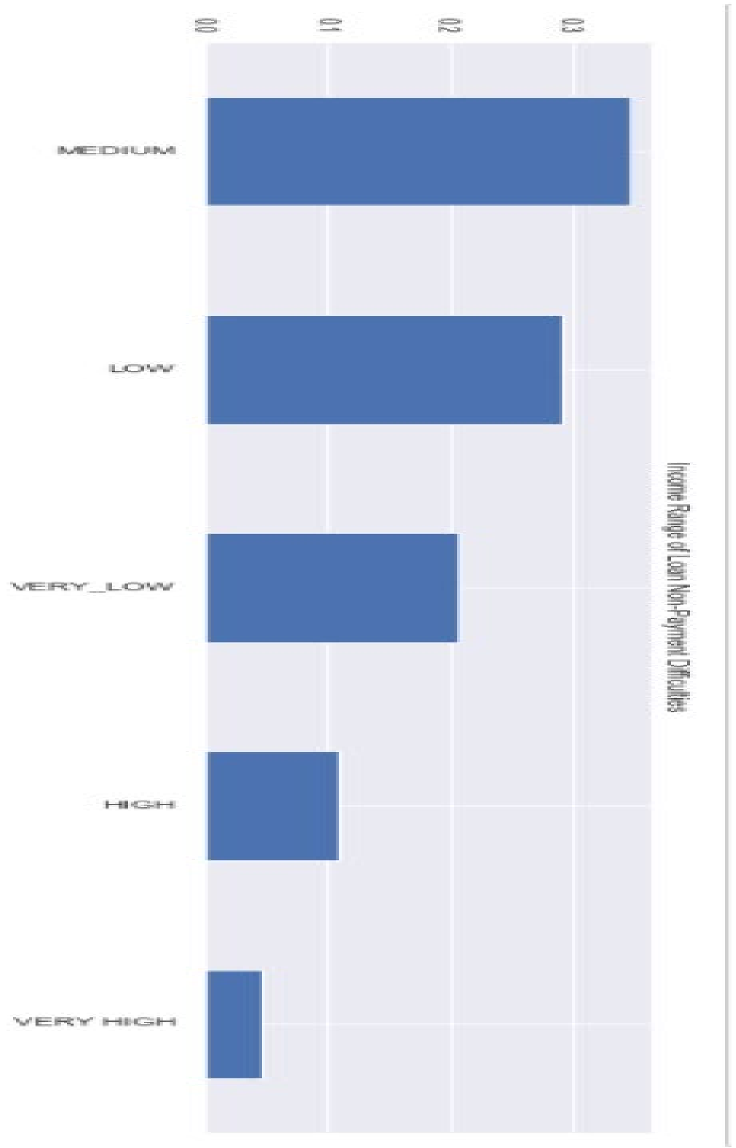


INCOME RANGE FOR LOAN PAYMENT DIFFICULTIES



INCOME RANGE FOR LOAN NON-PAYMENT DIFFICULTIES

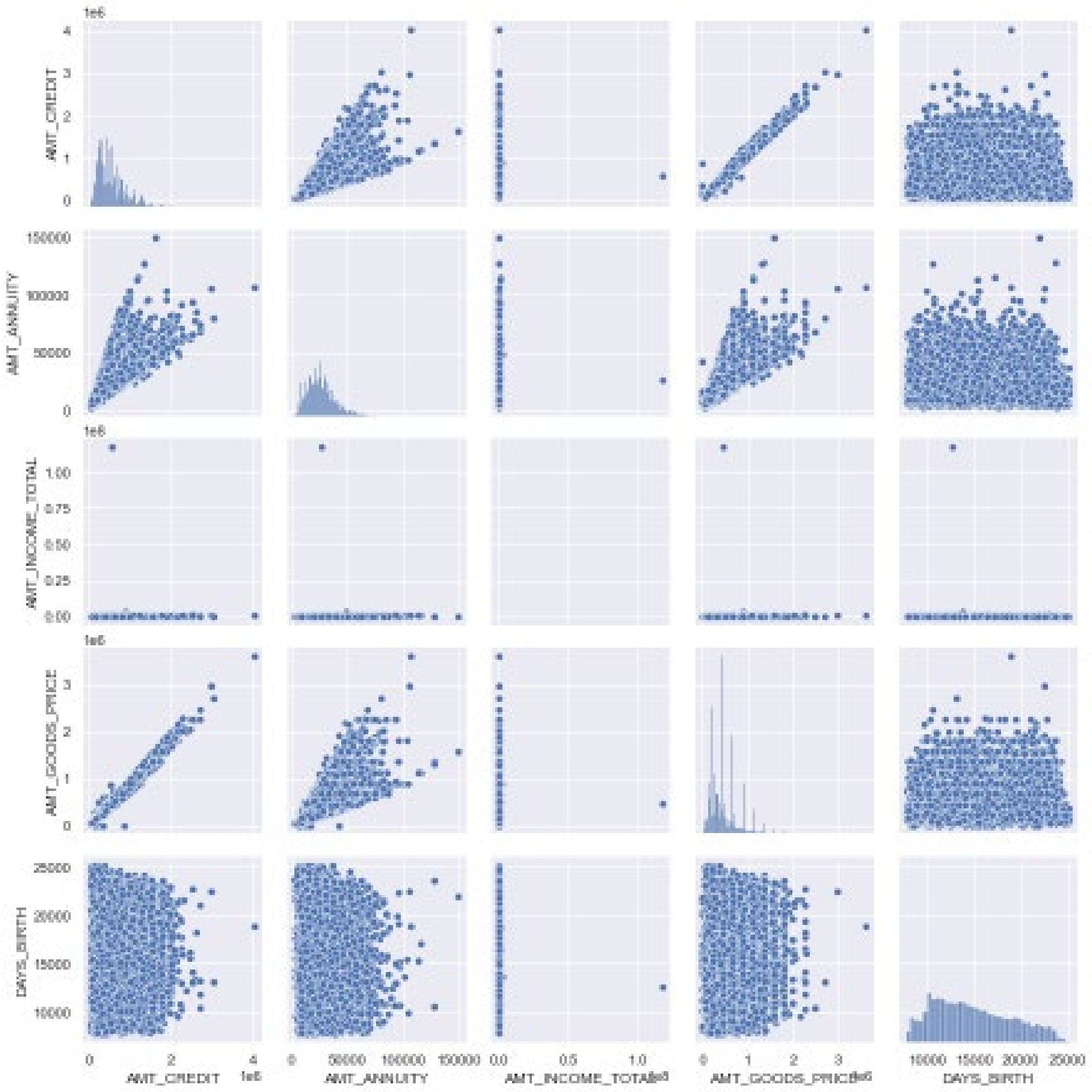
- We observe an increase in the percentage of Loan Payment Difficulties whose income is low when compared with the percentages of Payment Difficulties and loan Non-Payment Difficulties



BIVARIATE ANALYSIS

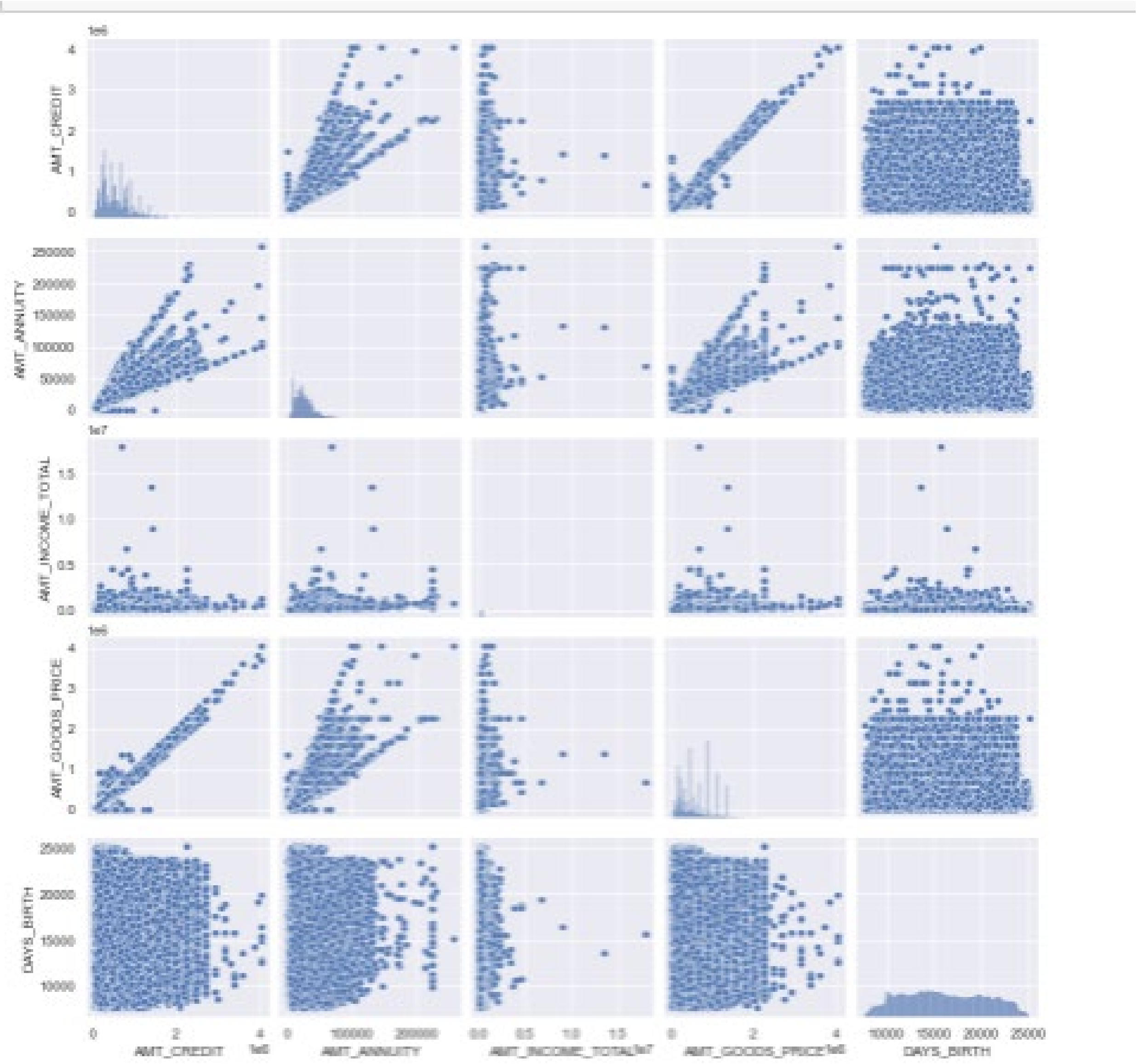
LOAN PAYMENT DIFFICULTIES

- Columns Amt_credit and Amt_Goods_Price are linear to each other
- AMT_CREDIT and AMT_ANNUITY are also linear



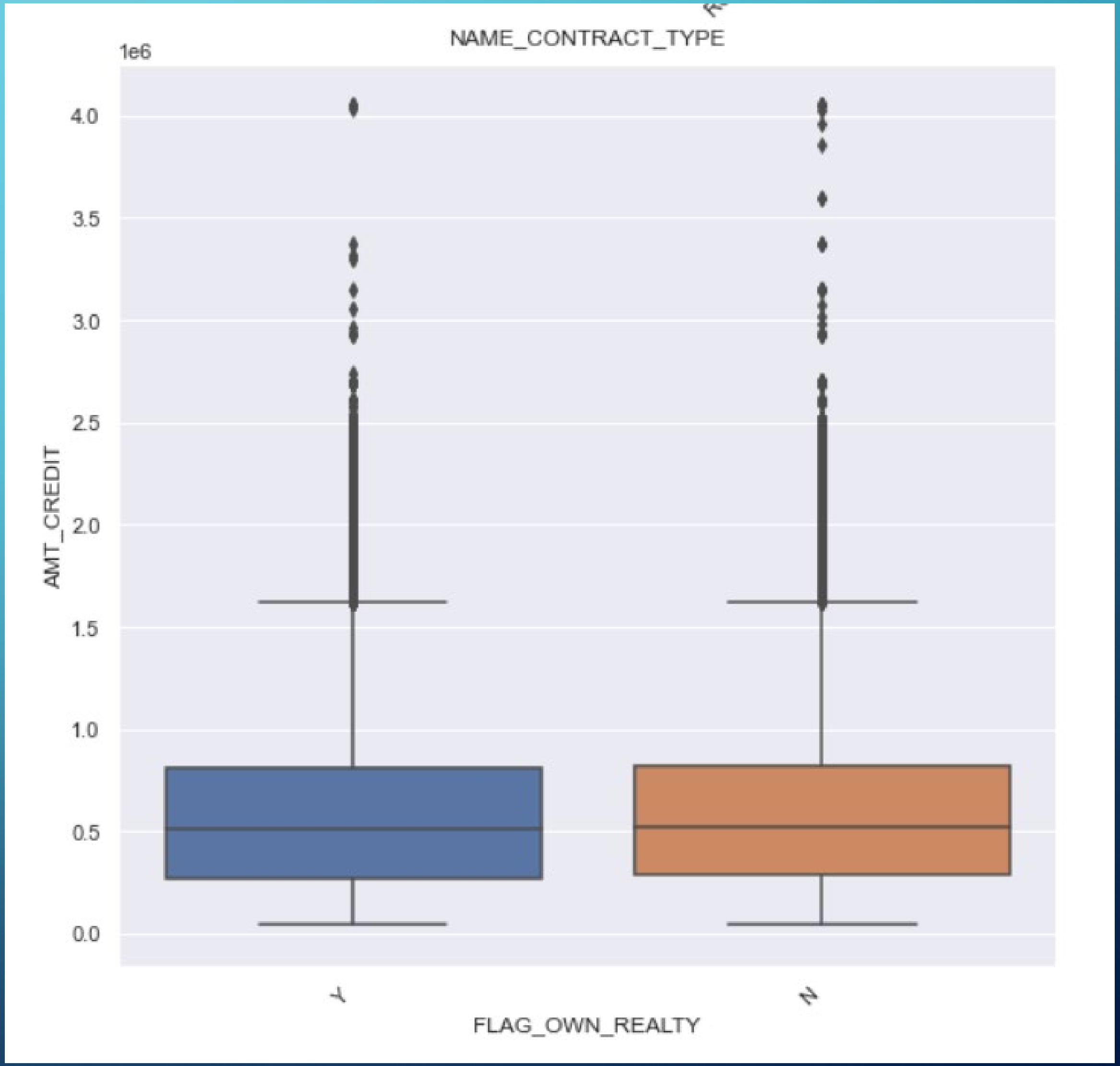
LOAN NON-PAYMENT DIFFICULTIES

- Columns Amt_credit and Amt_Goods_Price are linear to each other
- AMT_CREDIT and AMT_ANNUITY are also linear

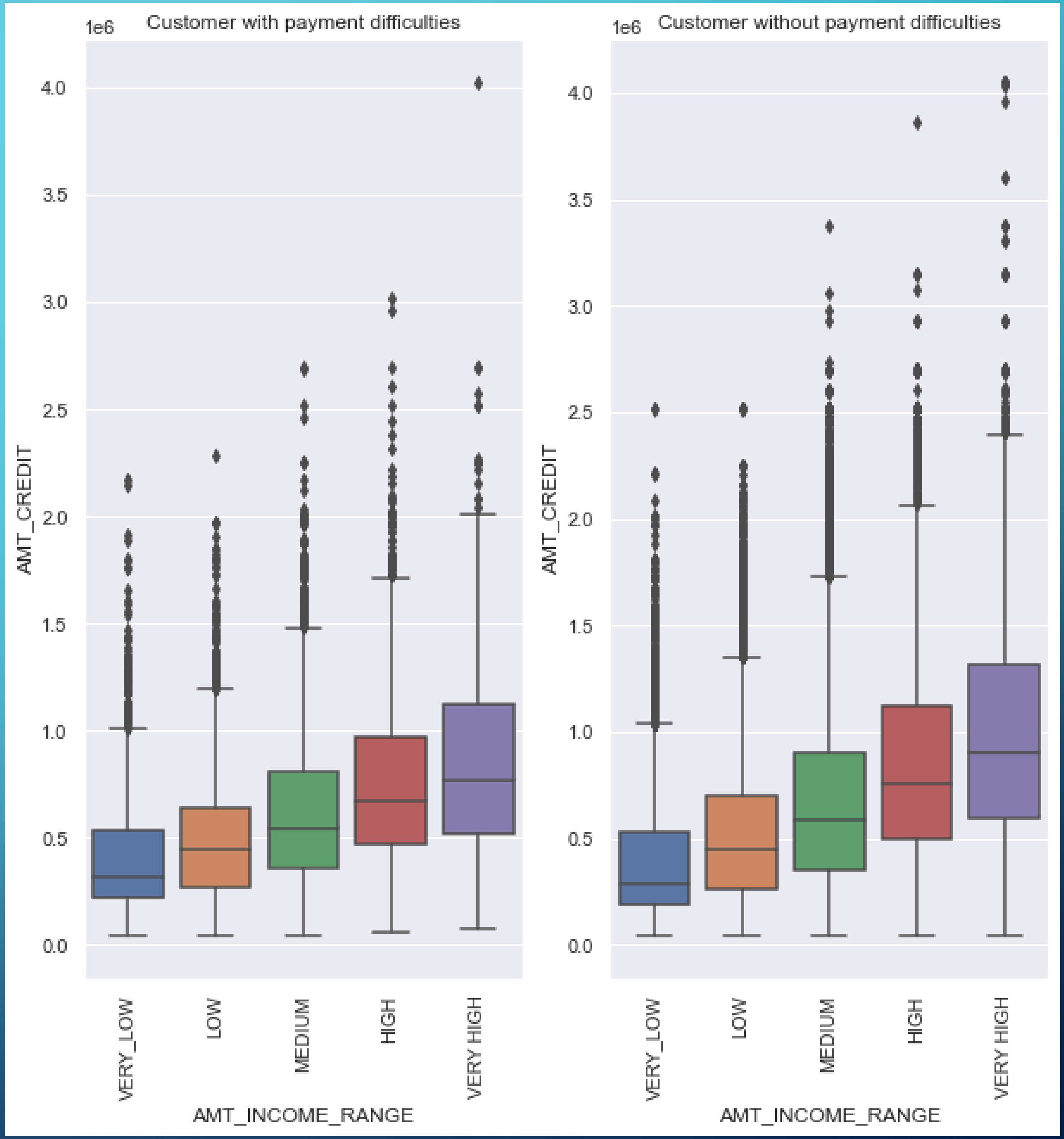


LOAN SANCTIONS - ANALYSIS

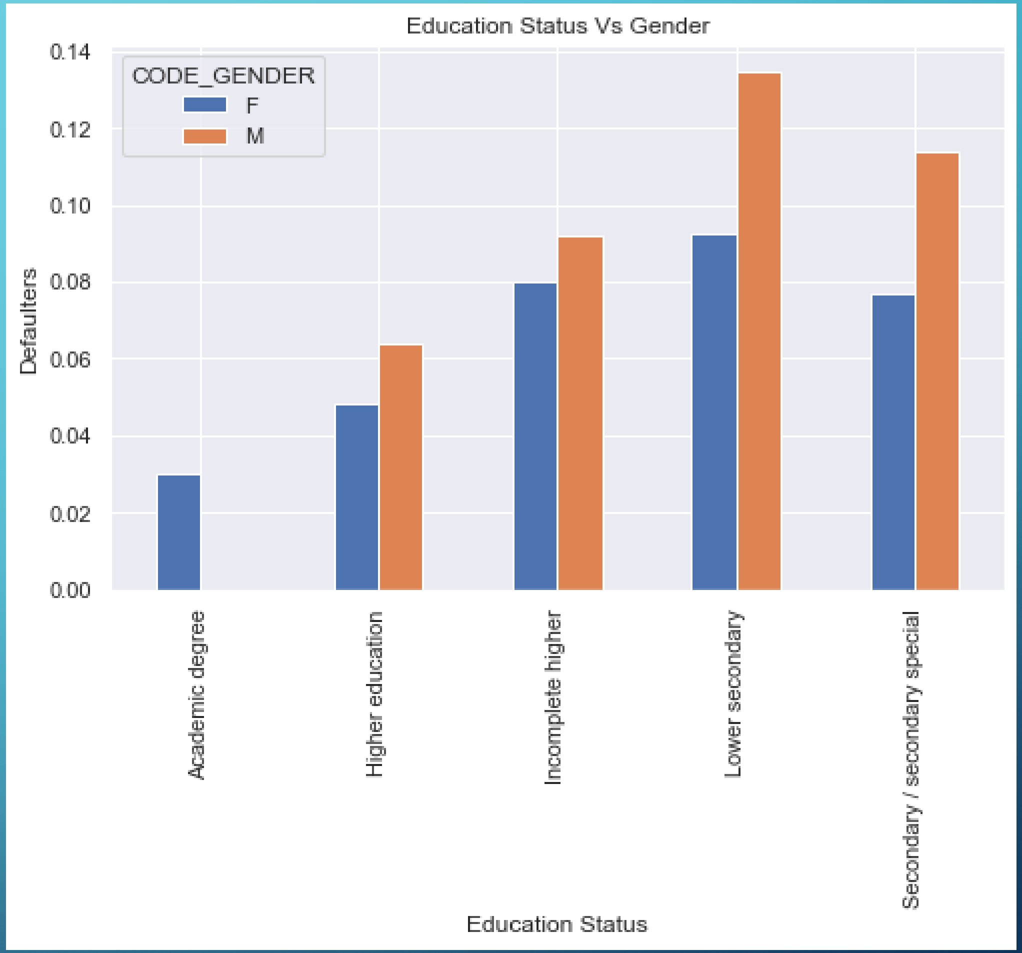
Married people got more number of loans



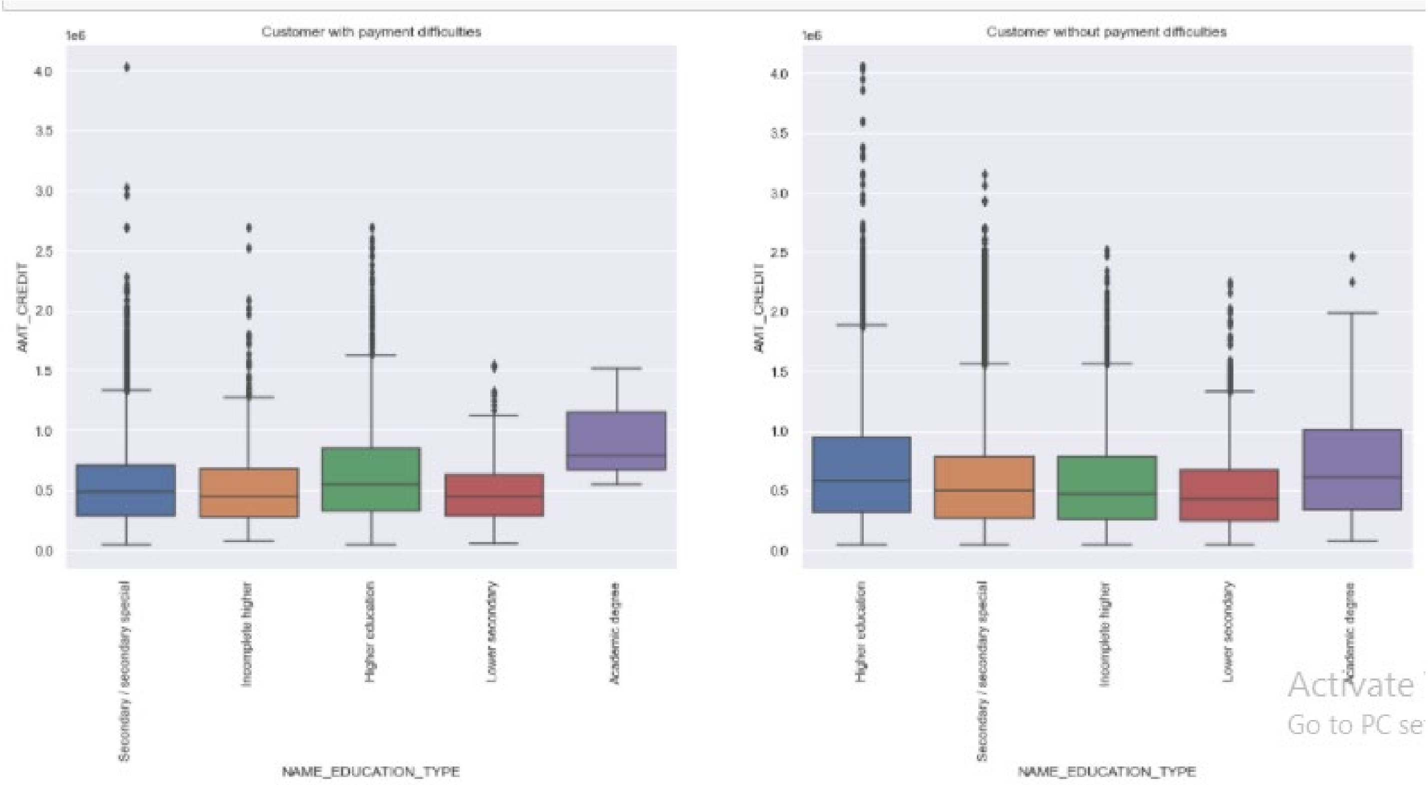
Customers in high income range are credited large amount and there percentage is high as defaulters



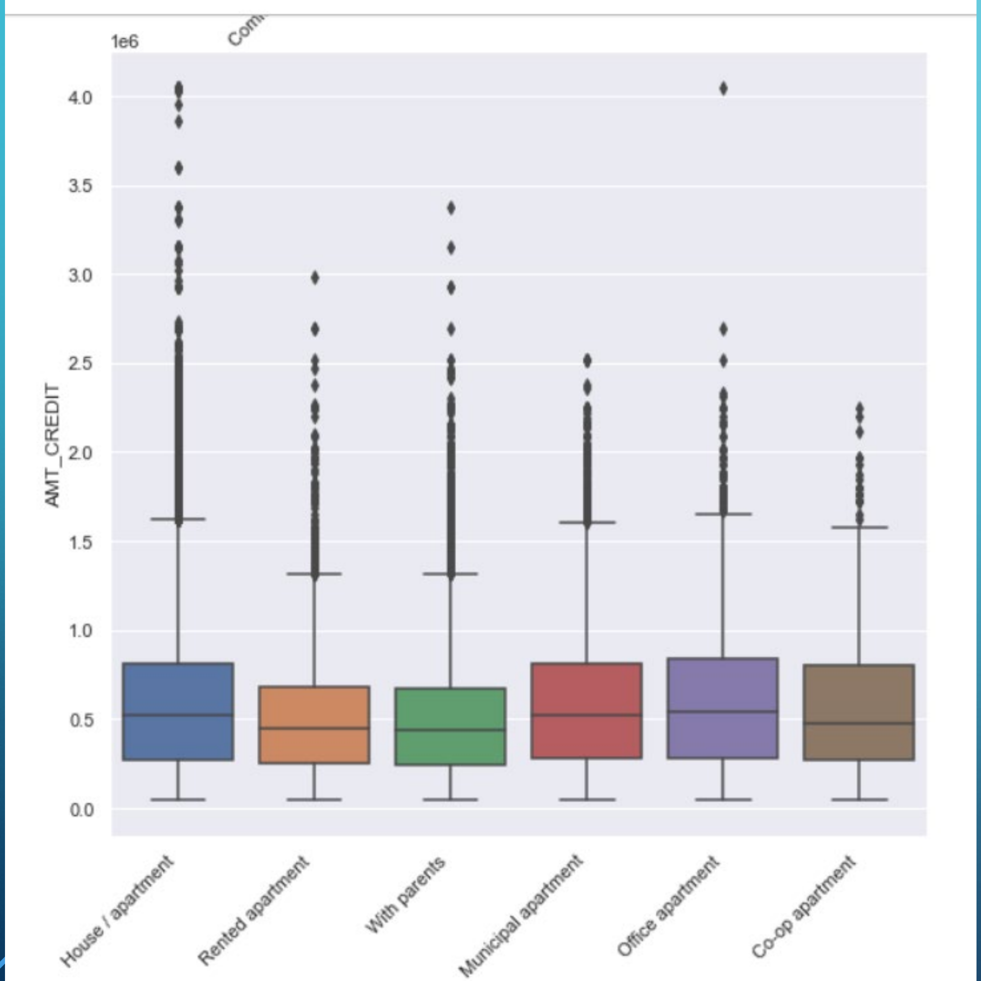
Male with lower secondary education are more defaulted followed by Secondary/secondary special education.

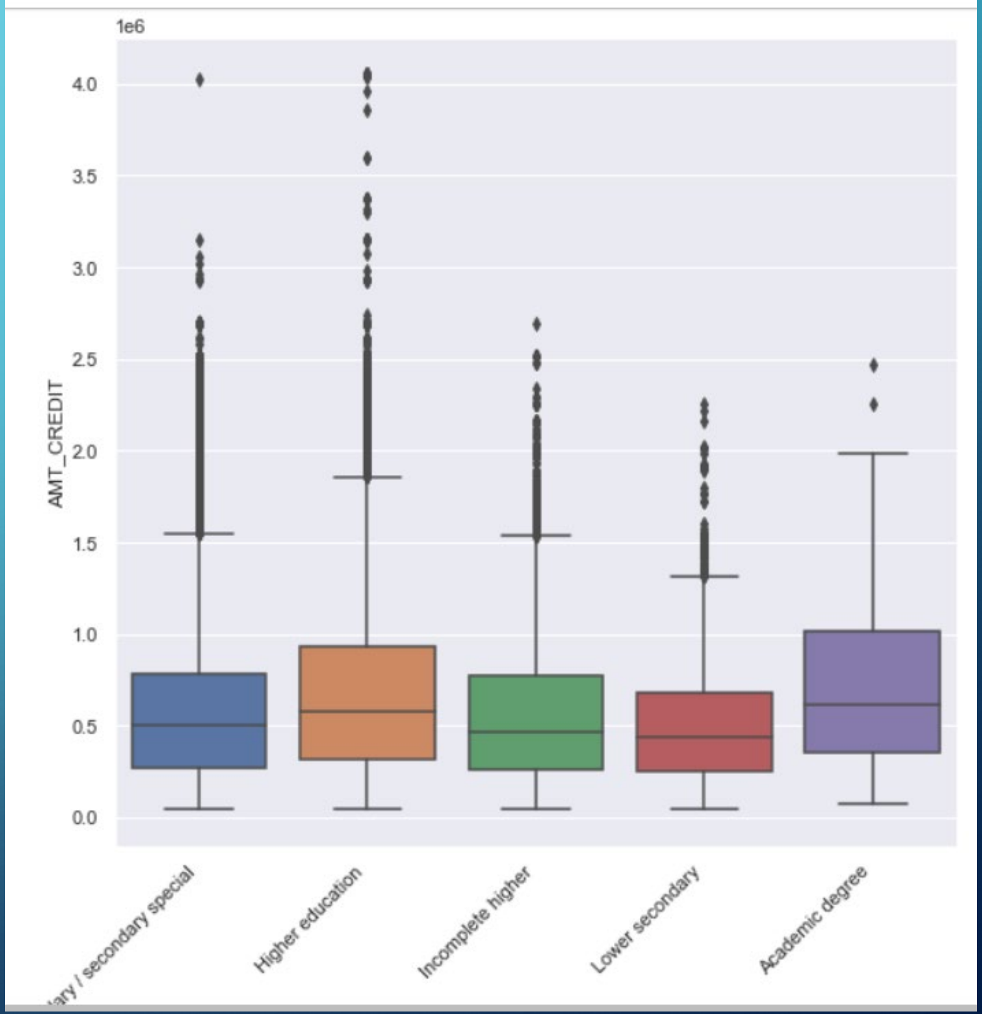


Customers without payment difficulties of Academic degree is higher than the customer of with payment difficulties. And rest of the Education type is almost same for both the cases



Clients who are living in municipal apartment, got more number of loans Higher education got more loans

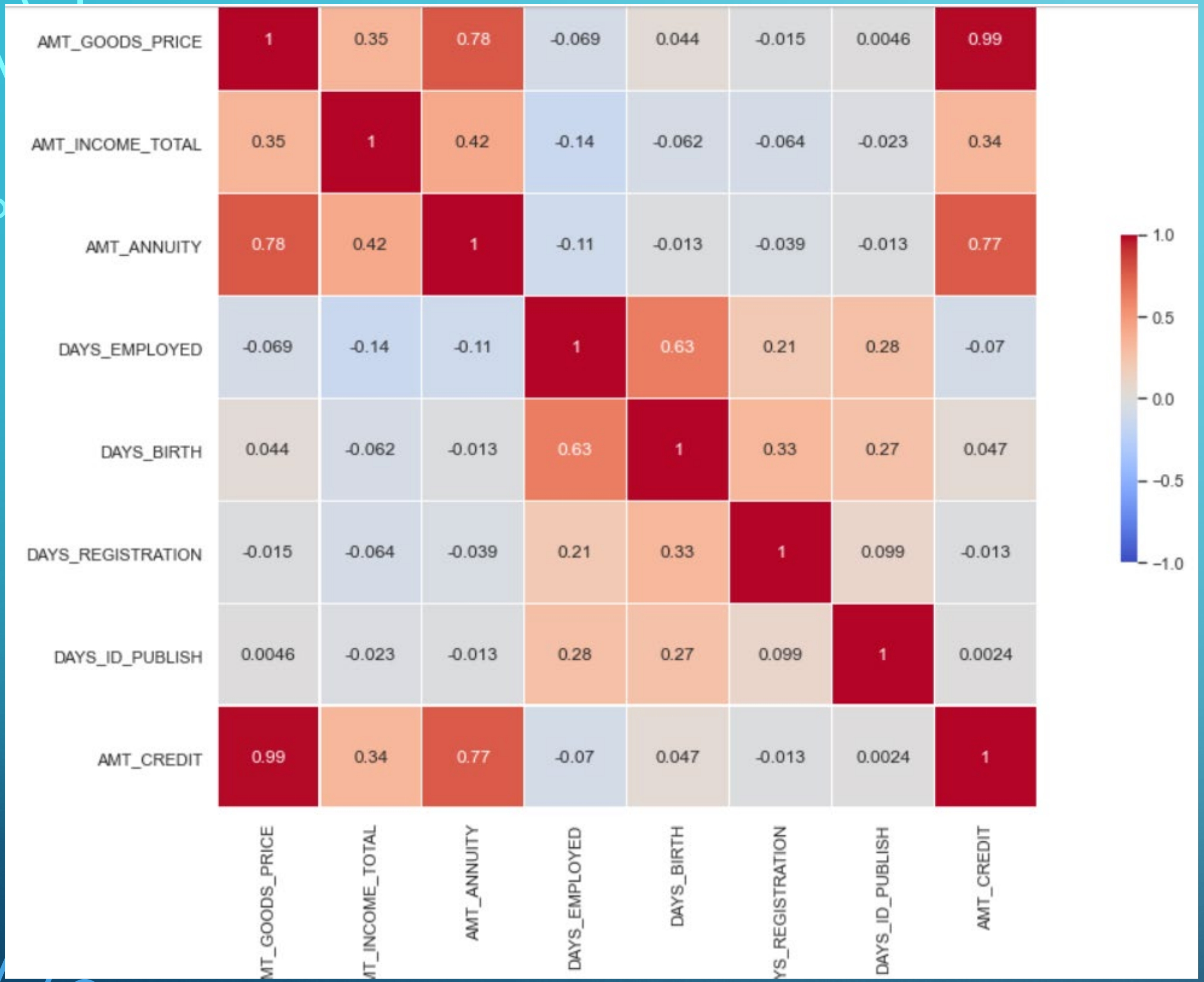
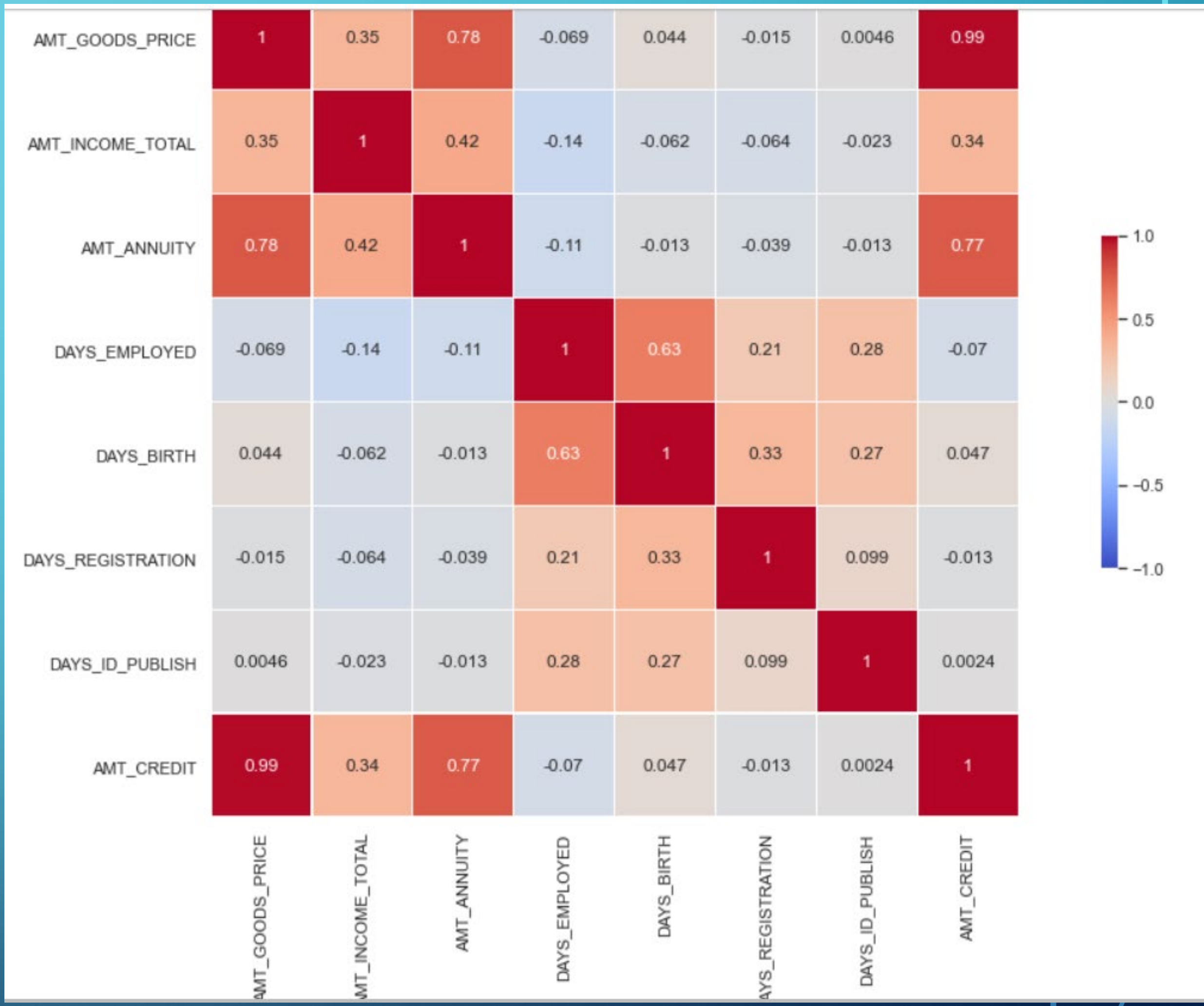




CORRELATION

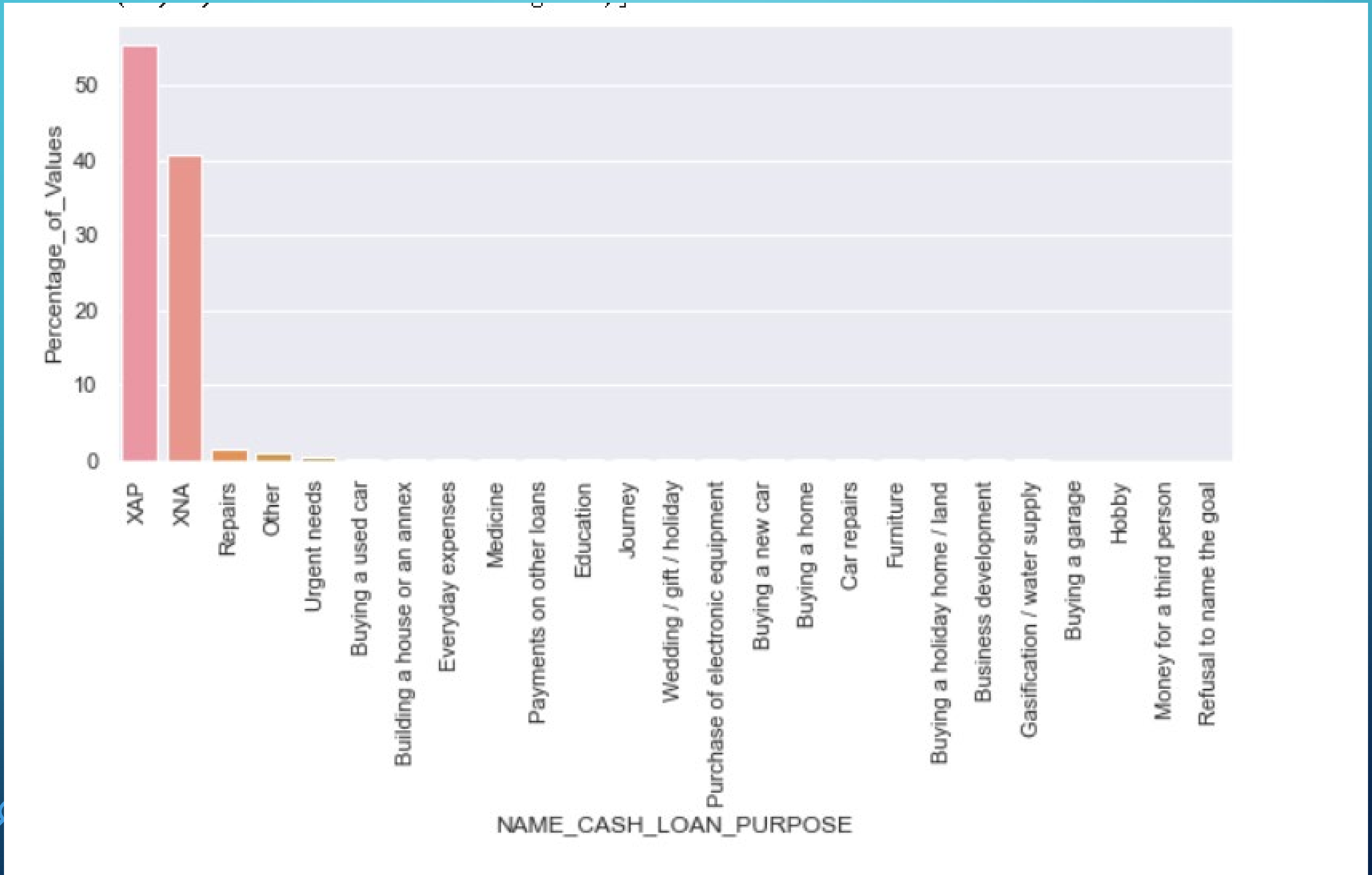
High correlation between credit amount and goods price.

There appears to be some deviancies in the correlation of Loan-Payment Difficulties and Loan- Non Payment Difficulties such as credit amount v/s income.

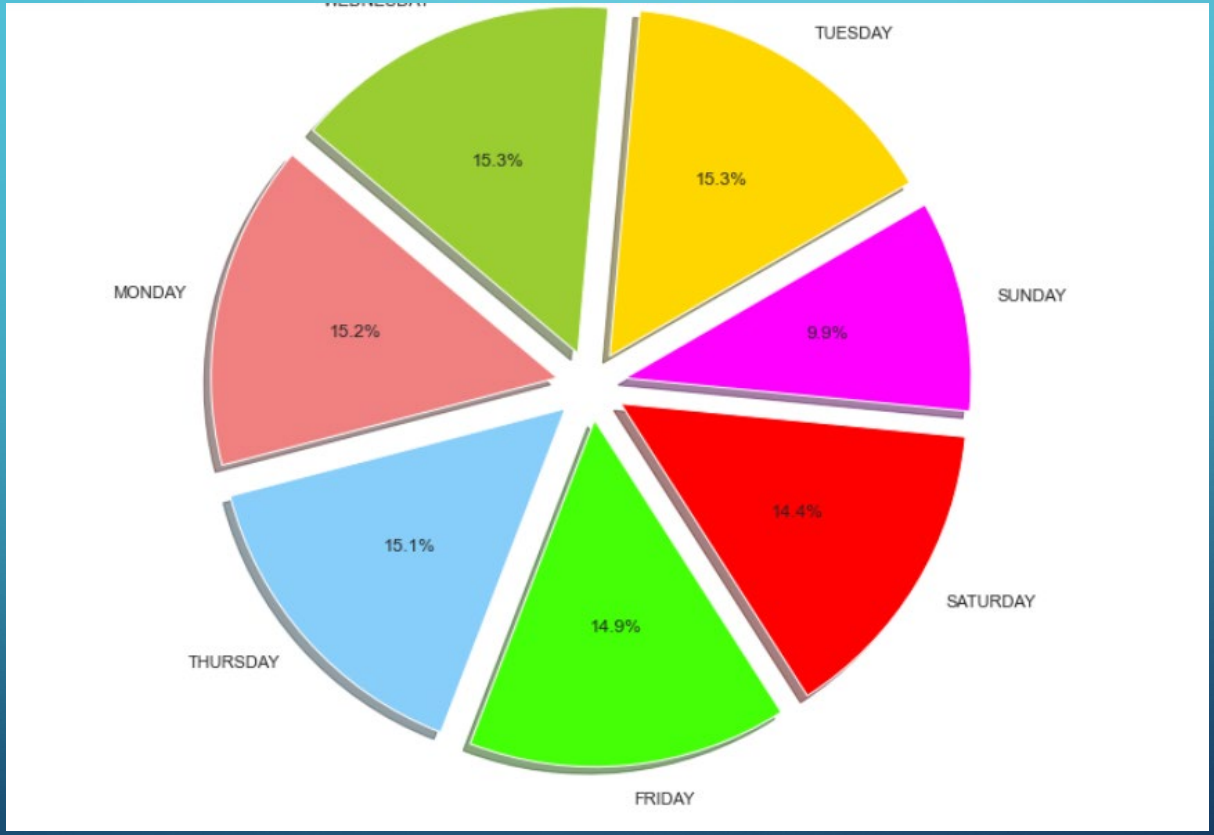

AMT_GOODS_PRICE 1 0.35 0.78 -0.069 0.044 -0.015 0.0046 0.99
AMT_INCOME_TOTAL 0.35 1 0.42 -0.14 -0.062 -0.064 -0.023 0.34
AMT_ANNUITY 0.78 0.42 1 -0.11 -0.013 -0.039 -0.013 0.77
DAYS_EMPLOYED -0.069 -0.14 -0.11 1 0.63 0.21 0.28 -0.07
DAYS_BIRTH 0.044 -0.062 -0.013 0.63 1 0.33 0.27 0.047
DAYS_REGISTRATION -0.015 -0.064 -0.039 0.21 0.33 1 0.099 -0.013
DAYS_ID_PUBLISH 0.0046 -0.023 -0.013 0.28 0.27 0.099 1 0.0024
AMT_CREDIT 0.99 0.34 0.77 -0.07 0.047 -0.013 0.0024 1
AMT_ANNUITY
DAYS_EMPLOYED
DAYS_BIRTH
DAYS_ID_PUBLISH
AMT_CREDIT
1.0
0.5
0.0
-0.5
-1.0

AMT_GOODS_PRICE 1 0.35 0.78 -0.069 0.044 -0.015 0.0046 0.99
AMT_INCOME_TOTAL 0.35 1 0.42 -0.14 -0.062 -0.064 -0.023 0.34
AMT_ANNUITY 0.78 0.42 1 -0.11 -0.013 -0.039 -0.013 0.77
DAYS_EMPLOYED -0.069 -0.14 -0.11 1 0.63 0.21 0.28 -0.07
DAYS_BIRTH 0.044 -0.062 -0.013 0.63 1 0.33 0.27 0.047
DAYS_REGISTRATION -0.015 -0.064 -0.039 0.21 0.33 1 0.099 -0.013
DAYS_ID_PUBLISH 0.0046 -0.023 -0.013 0.28 0.27 0.099 1 0.0024
AMT_CREDIT 0.99 0.34 0.77 -0.07 0.047 -0.013 0.0024 1
AMT_ANNUITY
DAYS_EMPLOYED
DAYS_BIRTH
DAYS_ID_PUBLISH
AMT_CREDIT
1.0
0.5
0.0
-0.5
-1.0

INSIGHTS

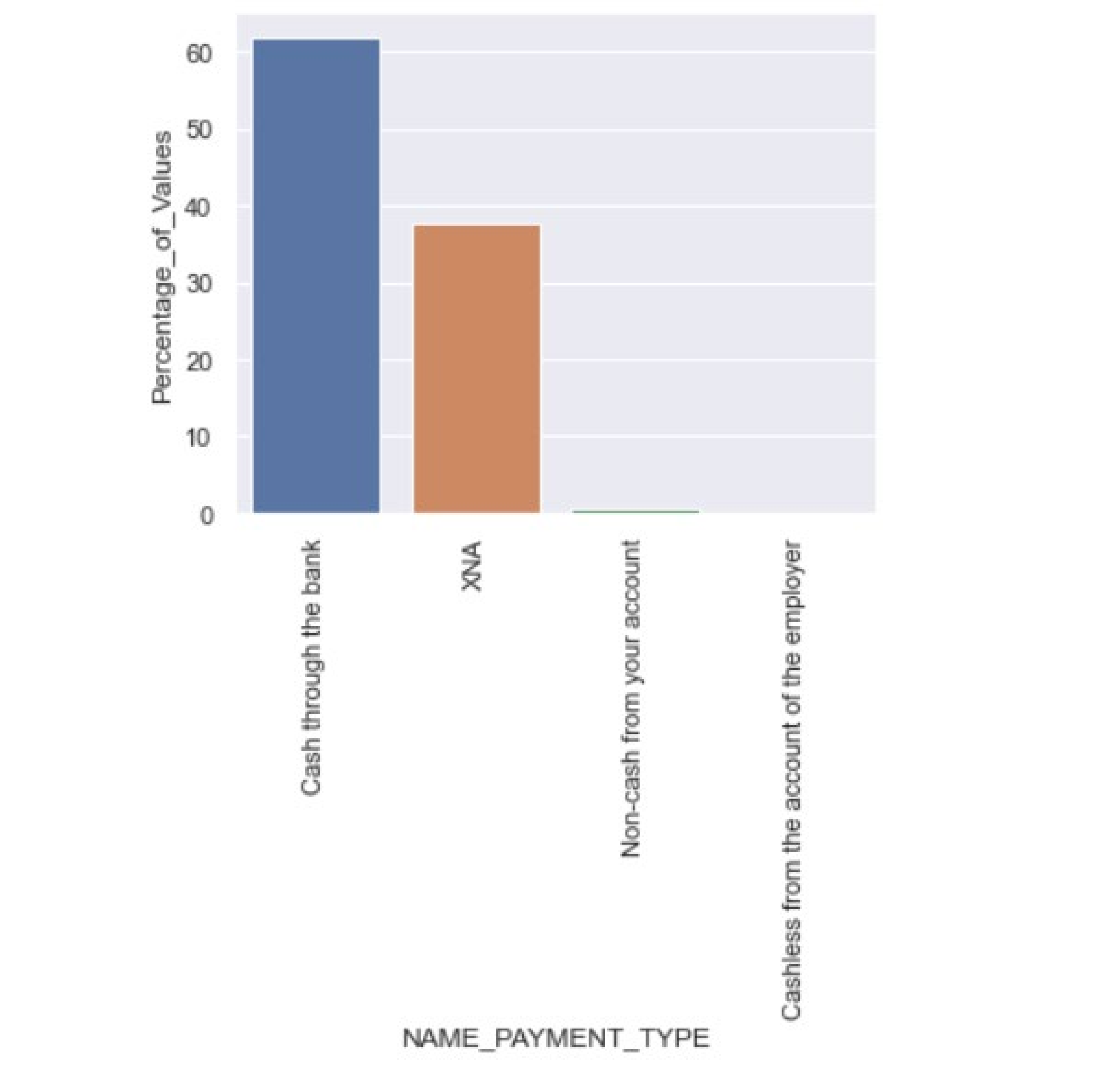
MOST LOAN PURPOSE WAS NOT RECORDED. XAP AND XNA VALUES ARE HIGHEST.



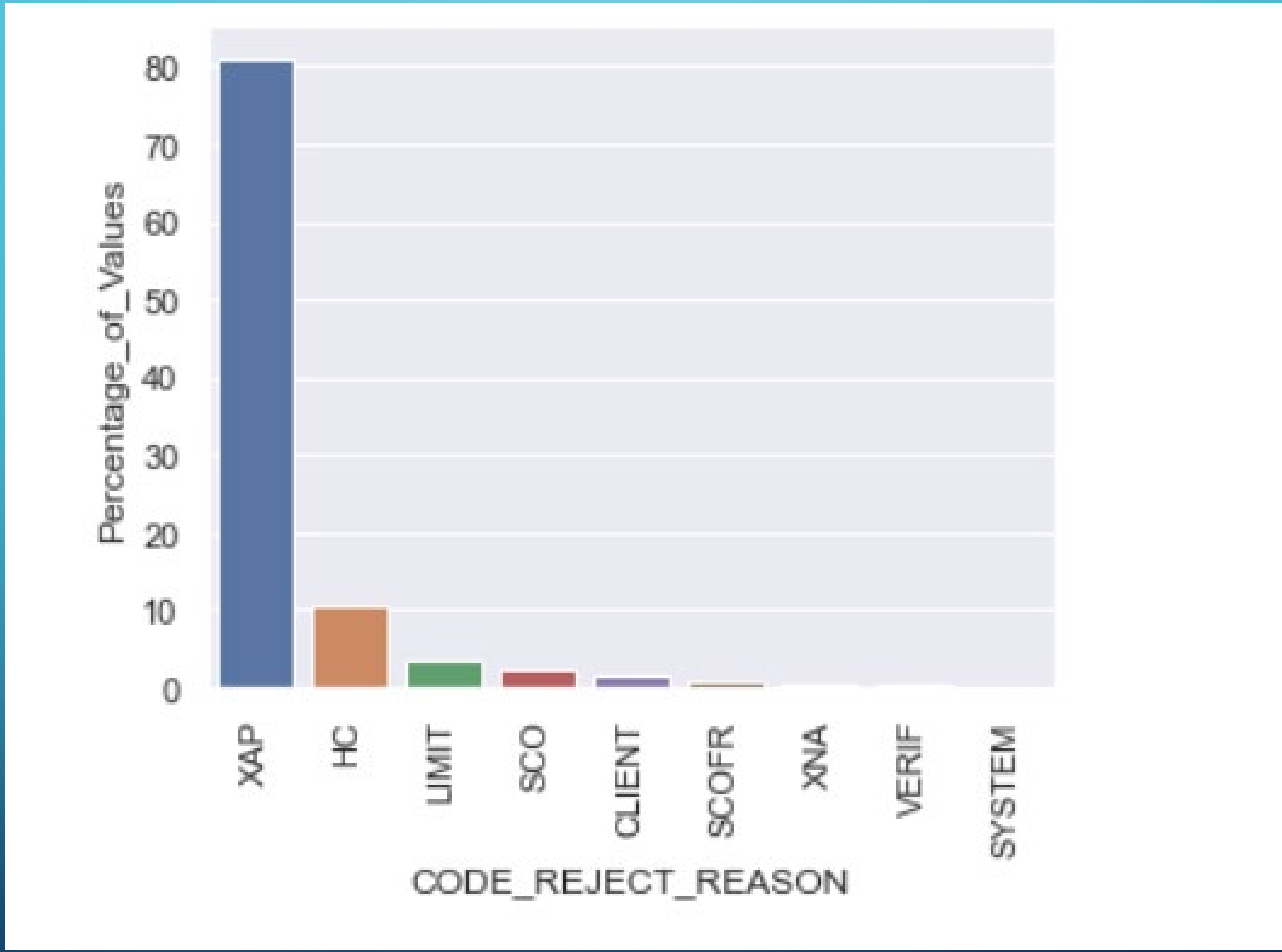
MOST OF THE CLIENTS HAVE OPTED TO APPLY LOAN ON TUESDAY RATHER THAN WEEKENDS



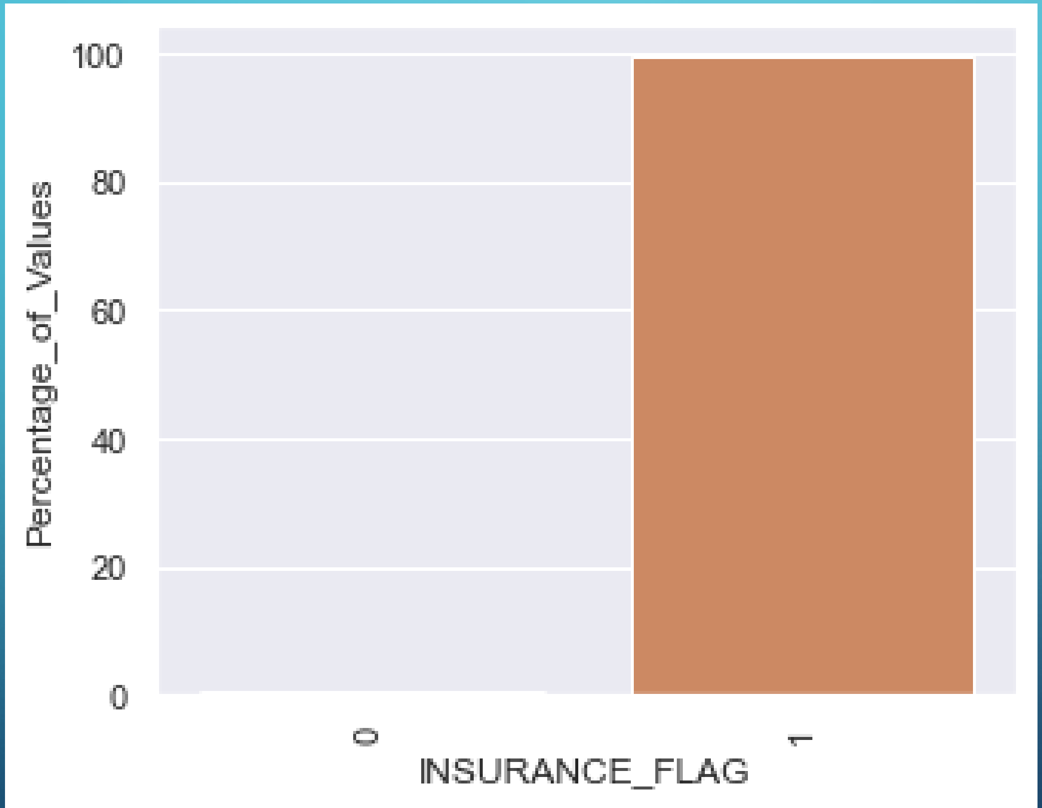
MOST PEOPLE PREFERRED CASH(62.44%) AS THE MODE OF PAYMENT



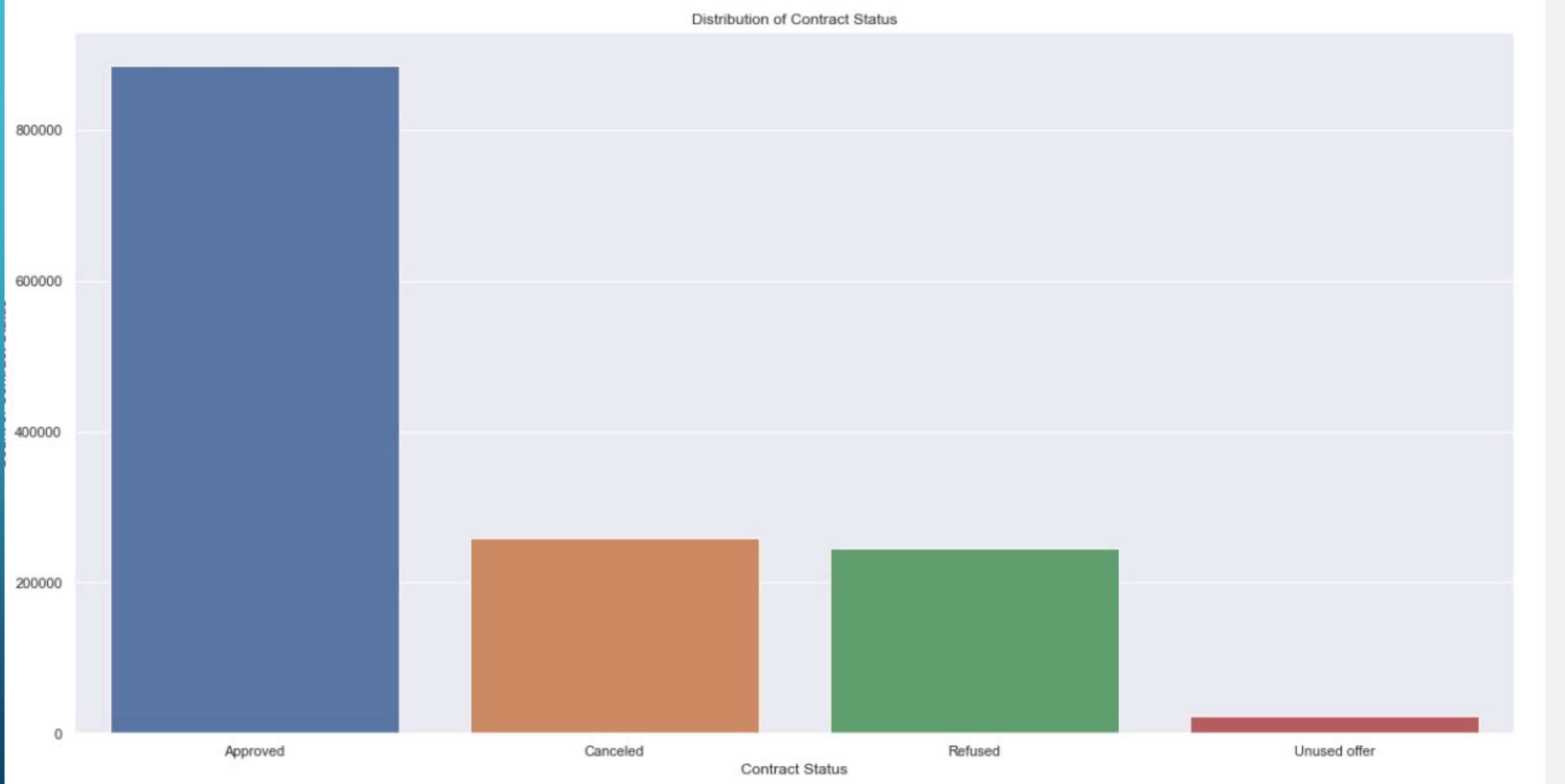
PRIMARY REASON FOR THE LOAN TO GET REJECTED IS NOT RECORDED(XAP (81%)) FOLLOWED BY HC.



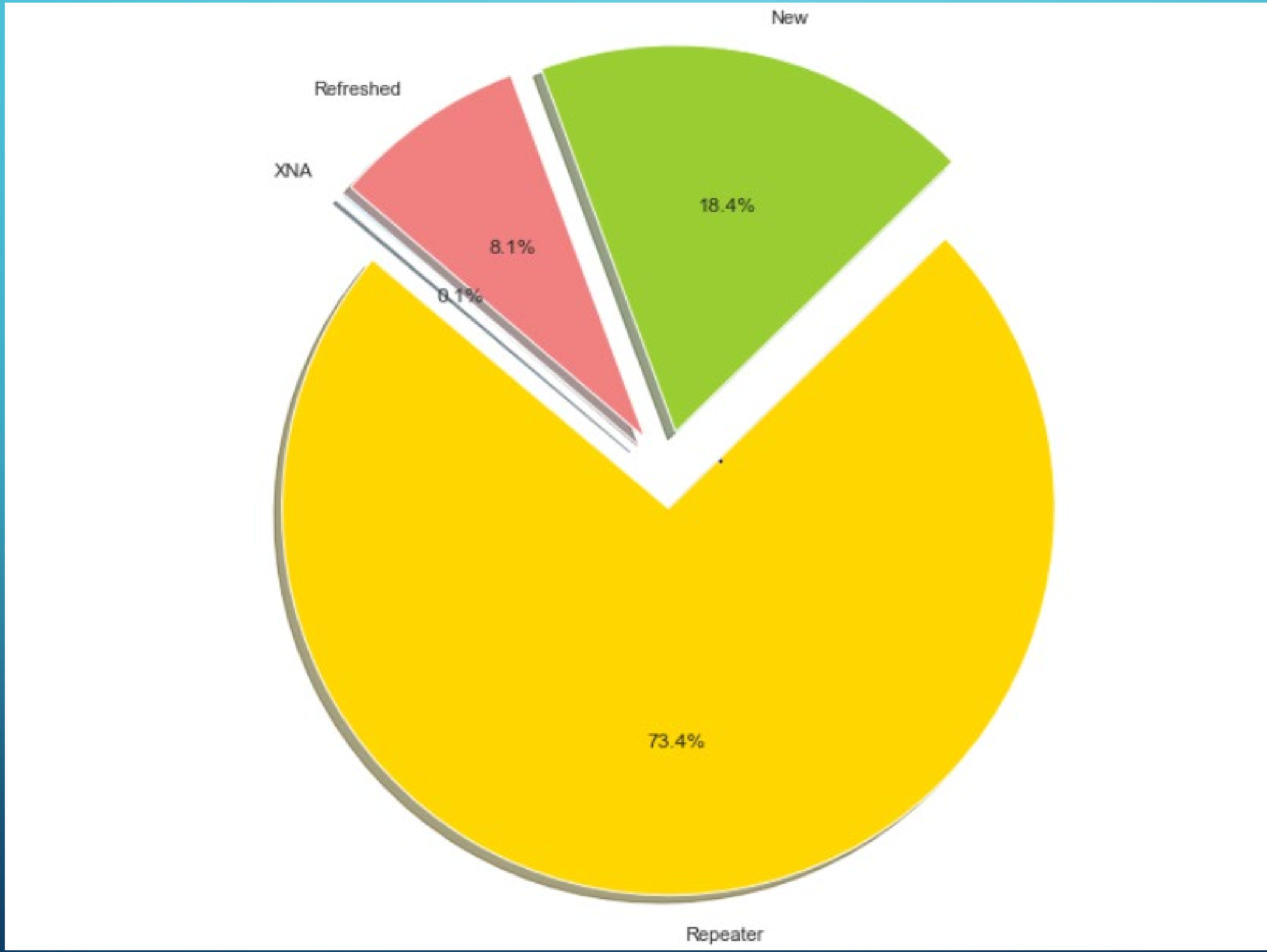
INSURANCE - FOR MOST CLIENTS IT WAS THE LAST APPLICATION OF THE DAY.



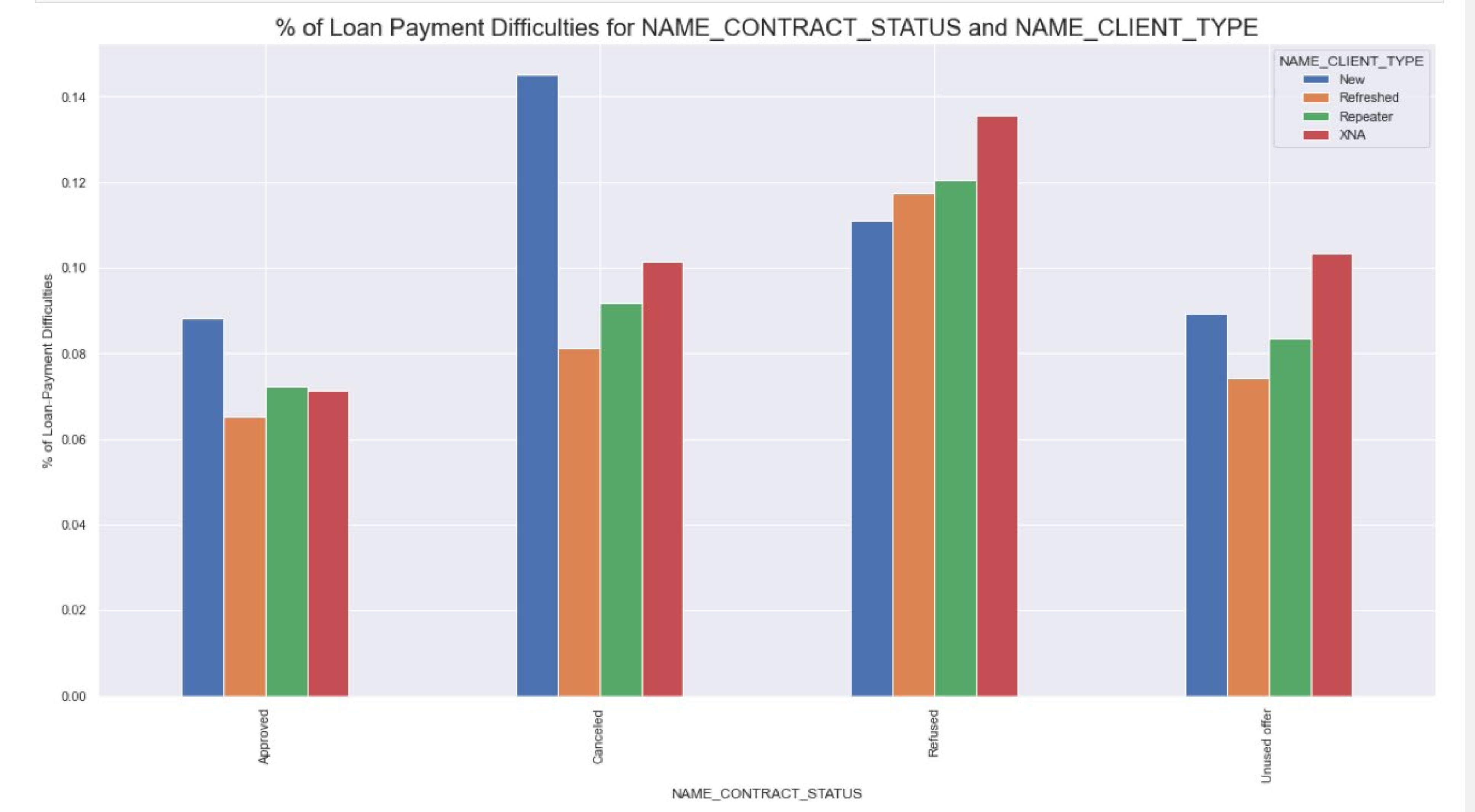
DISTRIBUTION OF CONTRACT STATUS



73.2% APPLICANTS ARE REPEATERS. ONLY, 18.5% ARE NEW CLIENTS.



CLIENT WHO WHERE 'NEW' AND HAD 'CANCELLED' PREVIOUS APPLICATION TEND TO HAVE MORE % OF LOAN-PAYMENT DIFFICULTIES IN CURRENT APPLICATION



CONCLUSIONS

LOANS PREVIOUSLY REFUSED OR CANCELLED – HIGHER DEFAULT RATE

LONGER EMPLOYMENT HISTORY – LESS DEFAULT

HIGH AMOUNT LOANS,HIGHER INCOME – LESS DEFAULTS

SINGLE PEOPLE DEFAULT , MARRIED PEOPLE ARE SAFE

PEOPLE WITH HIGHER EDUCATION , OLDER PEOPLE DEFAULT LESS

BANK SHOULD GIVE MORE REVOLVING LOANS , MORE CASH LOANS GO INTO DEFAULT

BANK LEND MORE TO FEMALES