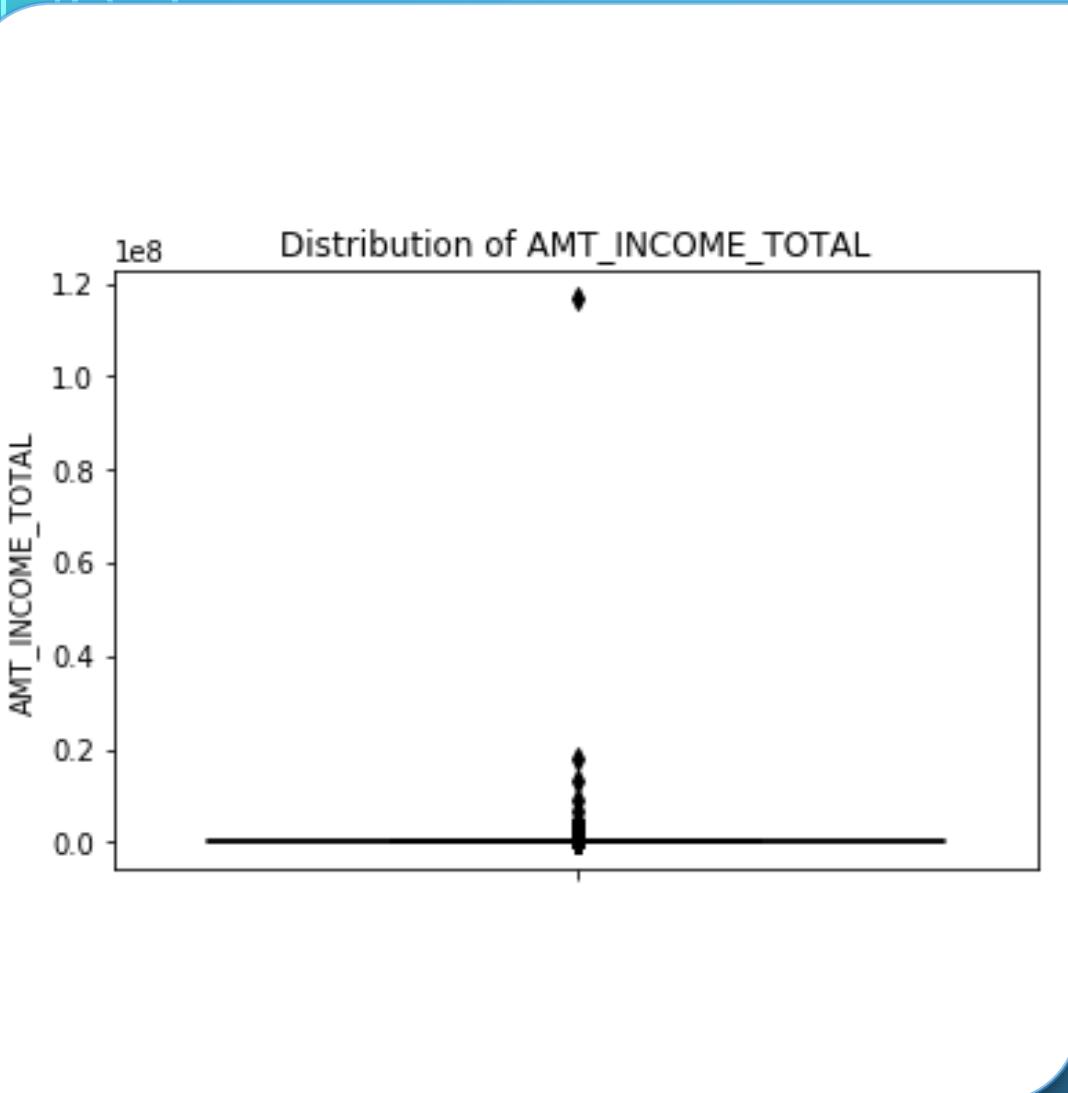
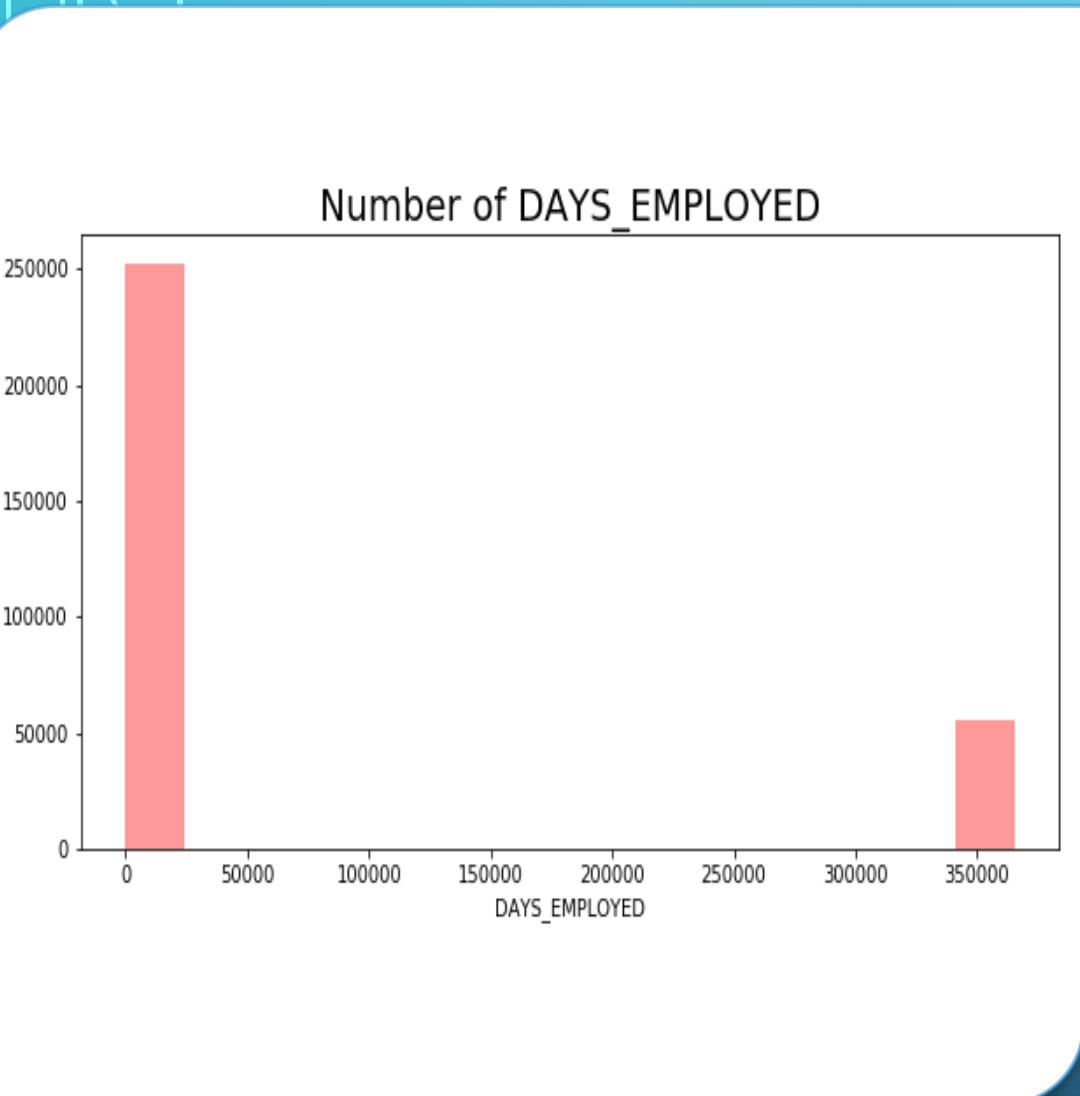


CREDIT EDA CASE STUDY

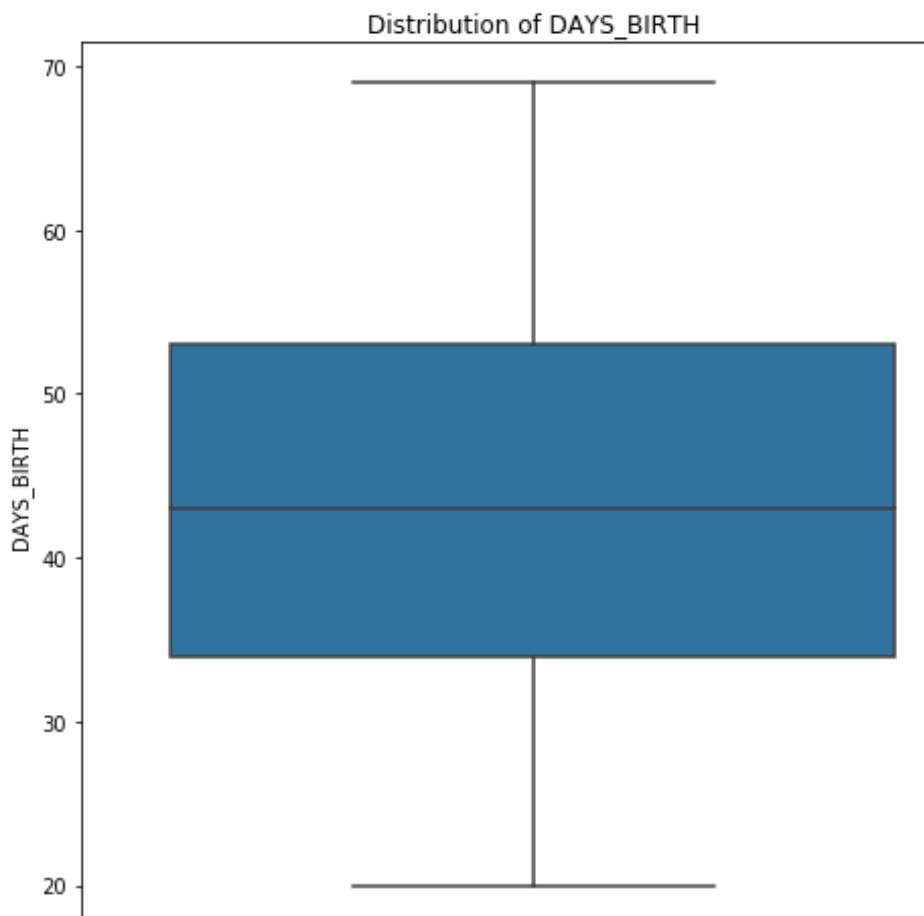
BY- MILAN MITHAL & UTKARSH MEHTA



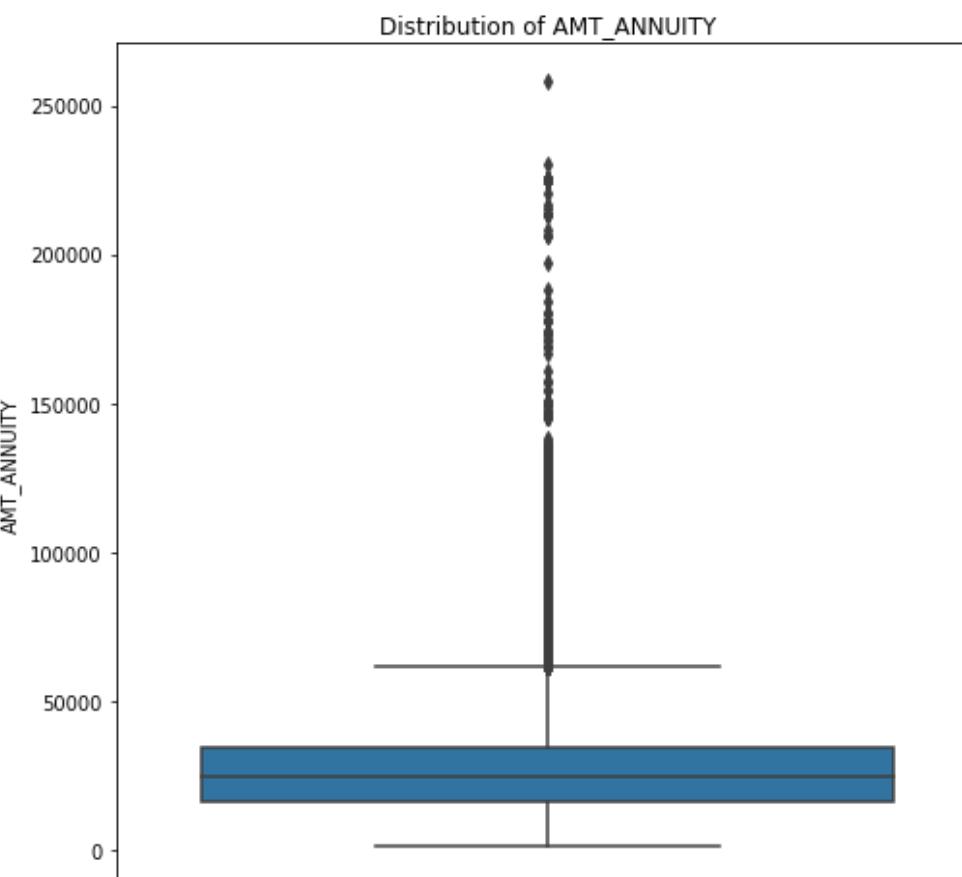
THE COLUMN AMT_INCOME_TOTAL WHICH TELLS US THE INCOME OF THE CLIENT. WE OBSERVE A VALUE VERY LARGE WHICH IS SURELY AN OUTLIER.



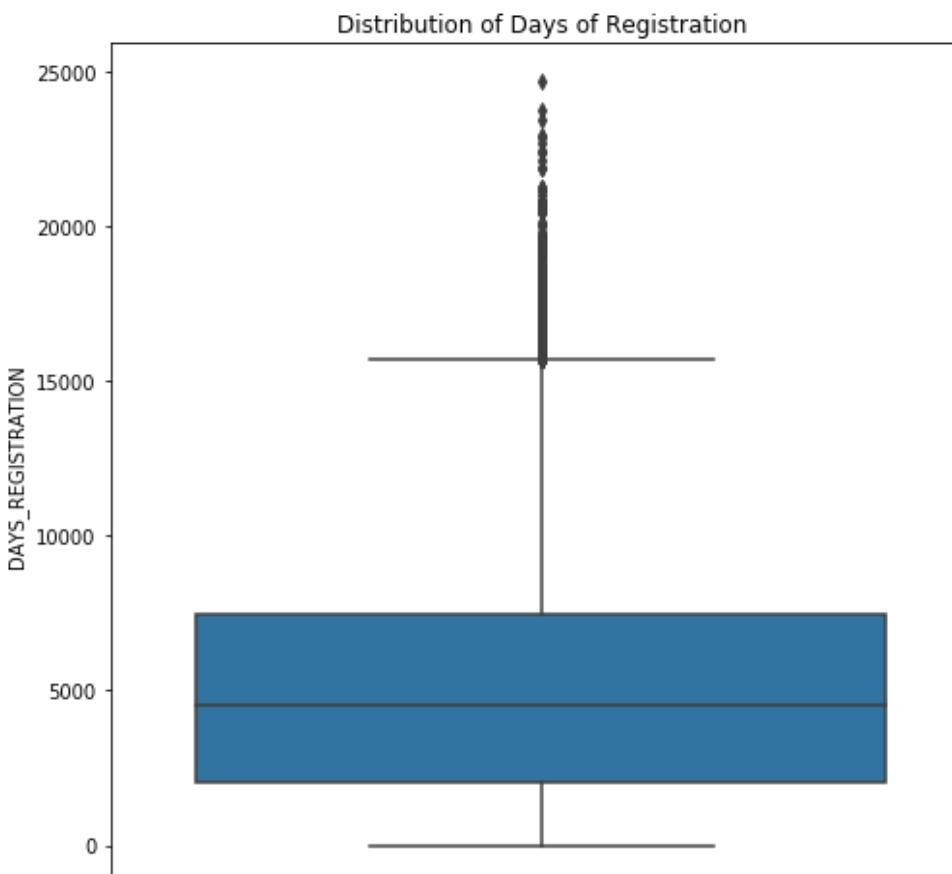
IN **DAYS_EMPLOYED** WHICH TELLS HOW MANY DAYS BEFORE THE APPLICATION THE PERSON STARTED HIS JOB. WE OBSERVE A VALUE WHICH IS GREATER THAT 20,000 WHICH IS SURELY AN OUTLIER BECAUSE $350000/365$ WILL BE AROUND **958 YEARS**. CONSIDERING THAT A PERSON STARTED WORKING AT AGE OF 21, THE PERSON WILL BE $21+958$ WILL TURN OUT TO BE 979 YEARS OLD.



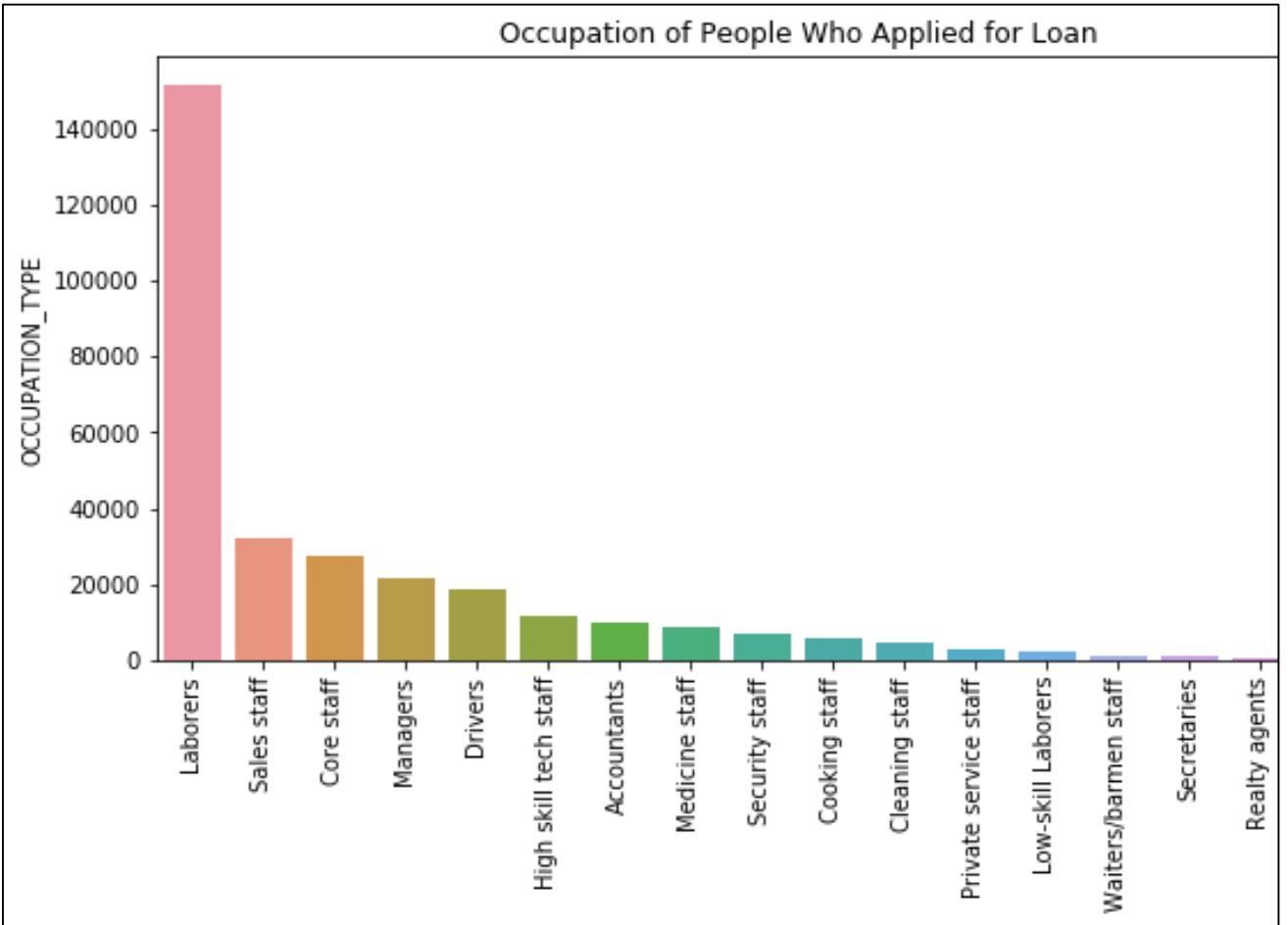
WE FIND NO OUTLIERS FOR **DAYS_BIRTH** COLUMN. WE CAN OBSERVE THAT MOST OF OUR CUSTOMERS RANGE BETWEEN 35-54 AGE GROUP WHICH IS OF COURSE A POSSIBILITY BECAUSE THIS DATA IS FOR WORKING PROFESSIONALS. ALSO , OUR TOTAL DATA LIES BETWEEN 20-70 YEARS , SO NOTHING IS WRONG WITH THIS COLUMN.



AMT_ANNUITY WHICH TELLS THE LOAN ANNUITY. WE OBSERVE A VALUE WHICH IS GREATER THAN 250000 WHICH LOOKS LIKE AN OUTLIER.

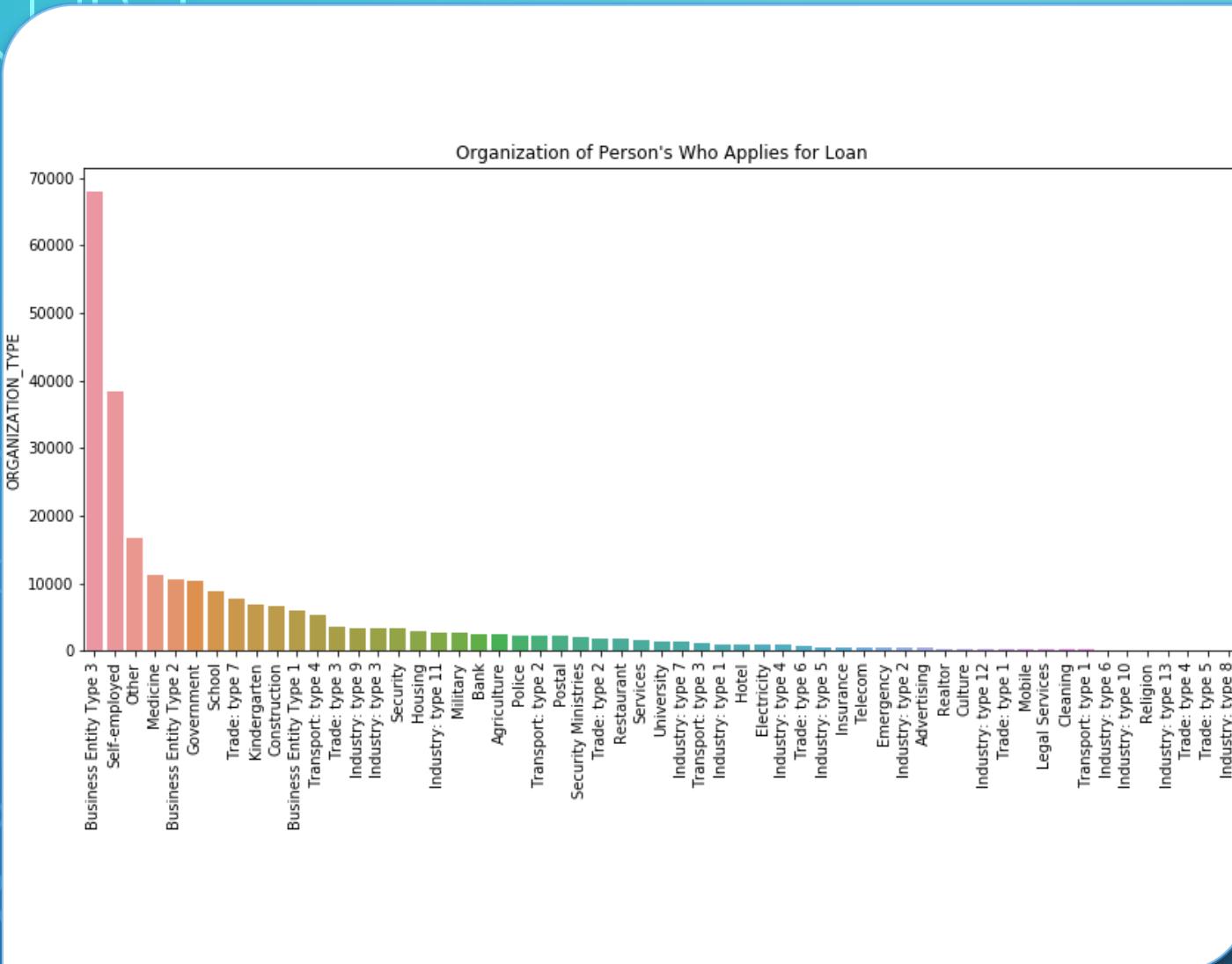


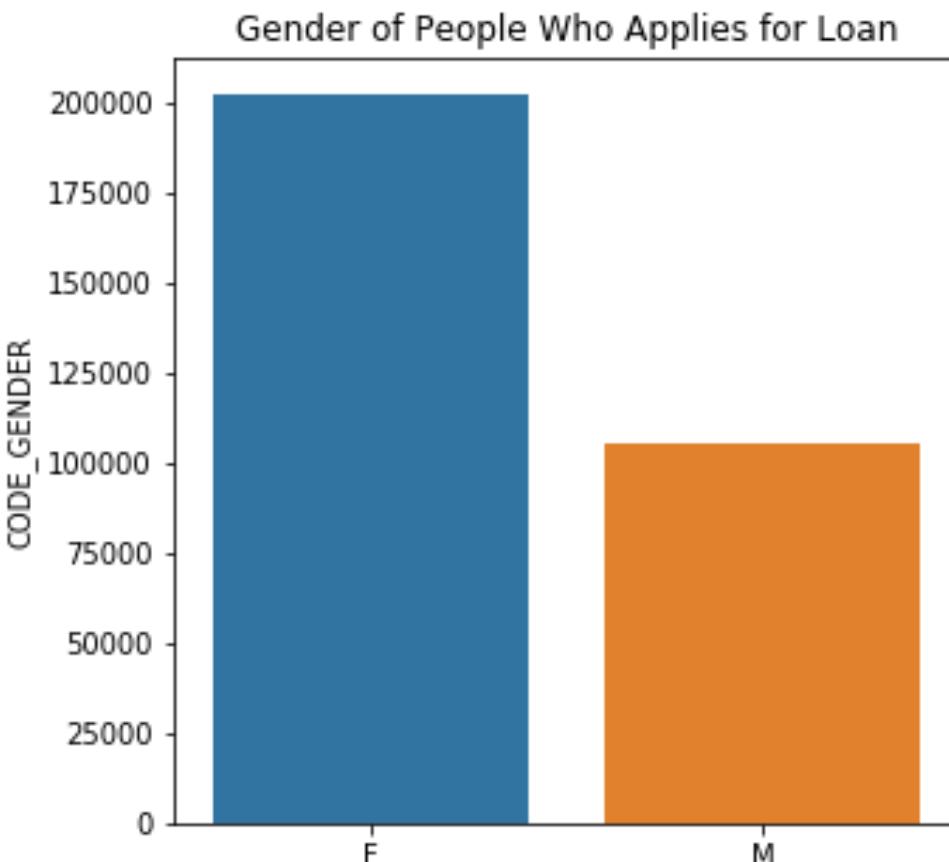
DAYs_REGISTRATION WHICH TELLS HOW MANY DAYS BEFORE THE APPLICATION DID CLIENT CHANGE HIS REGISTRATION. WE OBSERVE VALUES WHICH ARE GREATER THAN 24000 WHICH ARE OUTLIERS. BECAUSE 24000 DAYS IS ALMOST 66 YEARS ($24000/365$) AND CLIENT CAN NOT HAVE HIS PREVIOUS REGISTRATION 66 YEARS BACK.



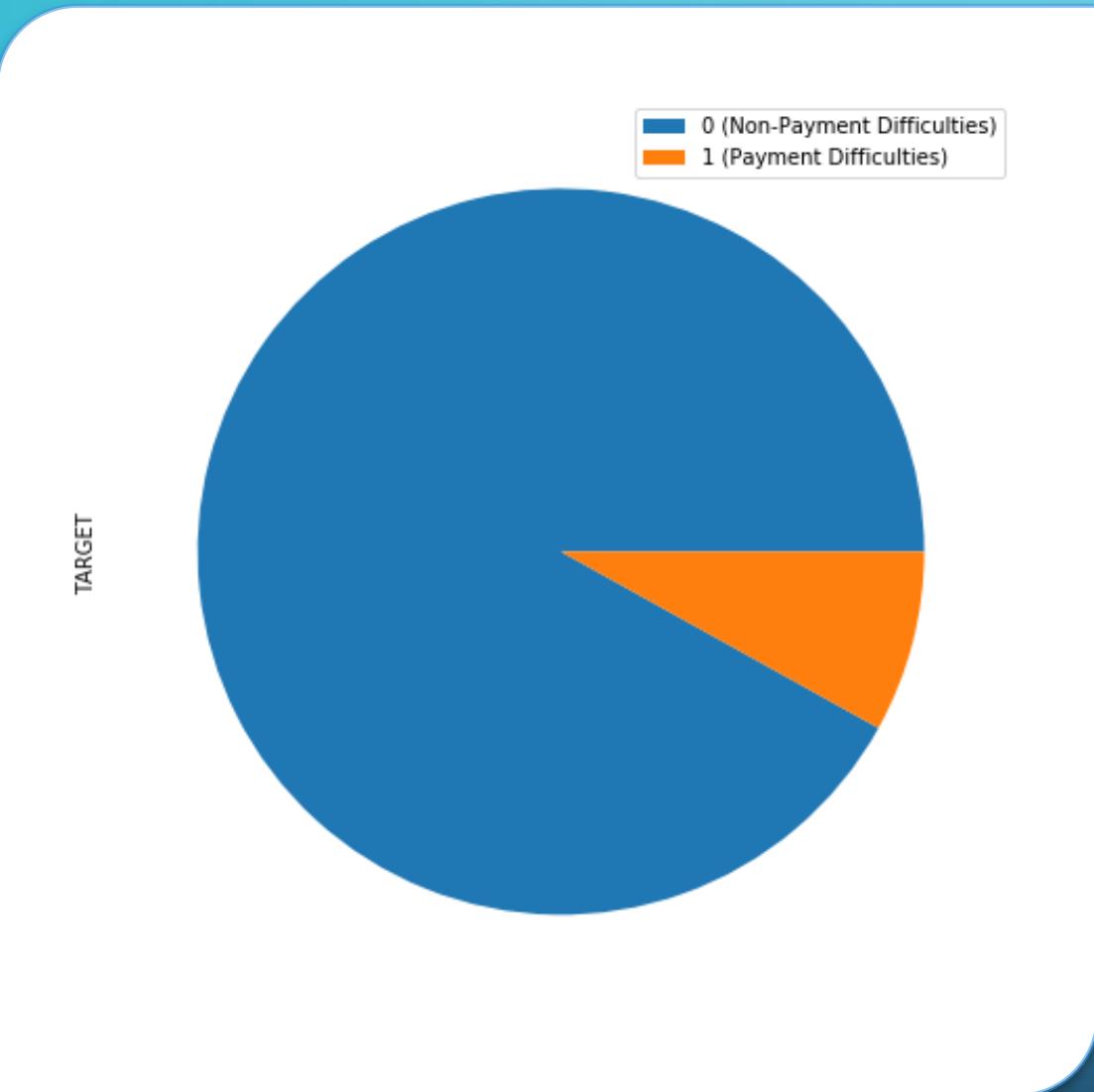
LABOUR, SALES STAFF, CORE STAFF, MANAGERS, DRIVERS OCCUPATION TYPE HAVE APPLIED FOR LOAN IN MORE NUMBER COMPARING TO OTHER OCCUPATION TYPE.

BUSINESS TYPE 3 ORGANIZATION TYPE HAVE APPLIED FOR LOAN IN LARGE NUMBER.





WE CAN SEE THAT **FEMALE** APPLICANTS HAVE APPLIED FOR LOAN MORE THAN MALE APPLICANTS.



WE HAVE SEEN IMBALANCE IS HIGHER BETWEEN TARGET VARIABLES. SO WE DIVIDE THIS DATASET INTO TWO DIFFERENT DATASETS BASED UPON **TARGET** VALUE.

- 0 = Loan-Non-Payment Difficulties
- 1 = Loan Payment Difficulties

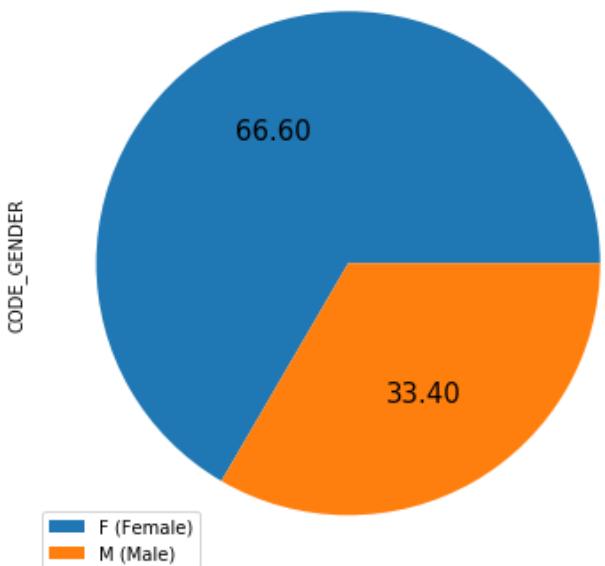


UNIVARIATE ANALYSIS OF CATEGORICAL VARIABLES

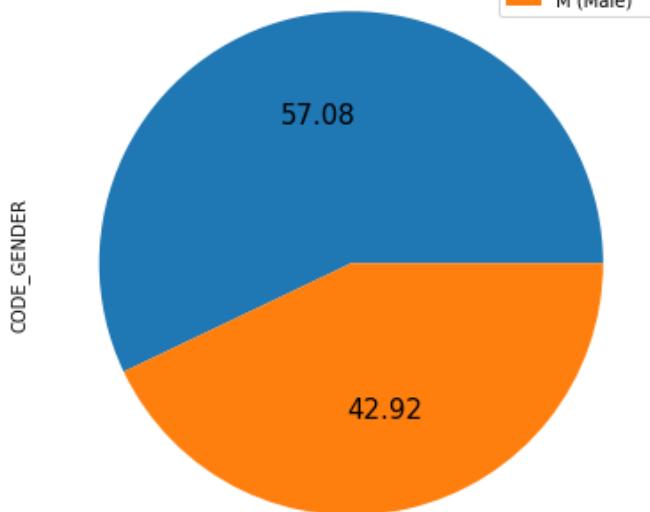
GENDER DISTRIBUTION

Comparing both Non-Payment difficulties & Payment difficulties based on **gender**, we observe that **Females are at majority in both the cases**. We also observe that there is an **increase in the percentage in male payment difficulties from non-payment difficulties**.

Gender Distribution of Loan-Non Payment Difficulties



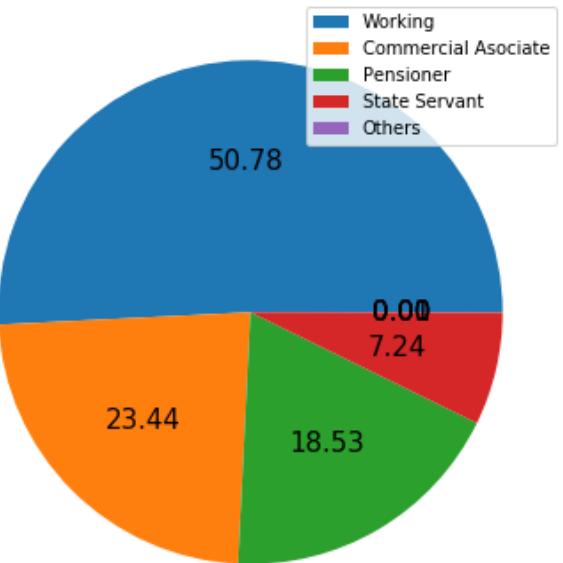
Gender Distribution of Loan Payment Difficulties



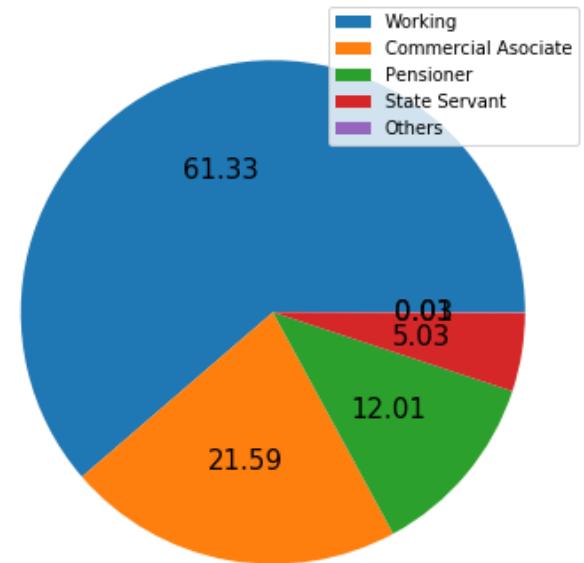
INCOME SOURCES

We observe a **Decrease** in the **Percentage of Payment Difficulties** who are **Pensioners** and an **increase** in the **Percentage of Payment Difficulties** who are **working** when compared the percentages of both payment difficulties and non-payment difficulties.

Income sources of Loan-Non Payment Difficulties



Income sources of Loan Payment Difficulties

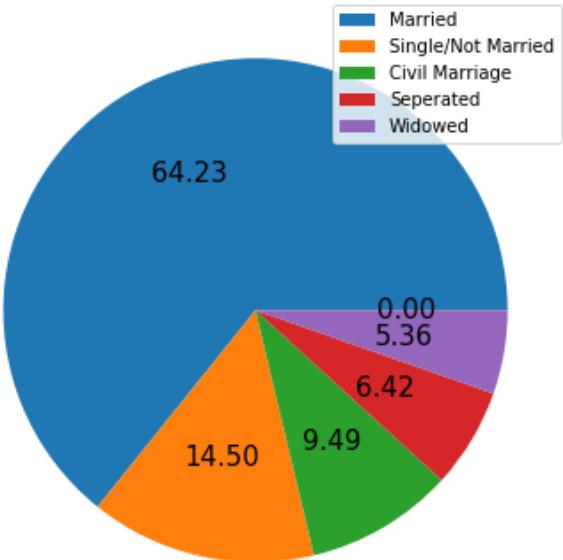


FAMILY STATUS

We observe that there is 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.

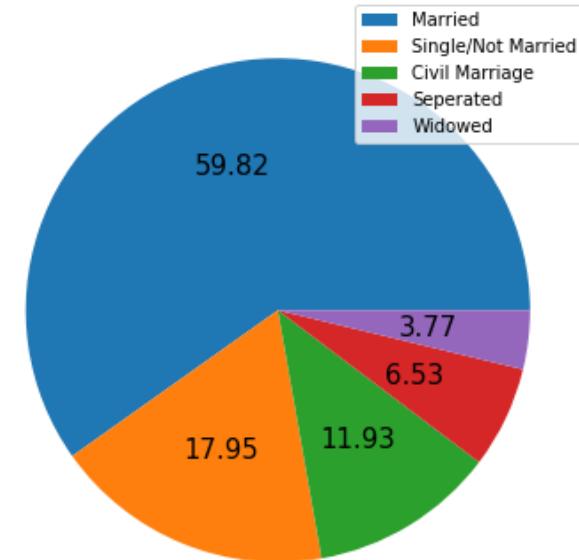
Family Status of Loan-Non Payment Difficulties

NAME_FAMILY_STATUS



Family Status of Loan Payment Difficulties

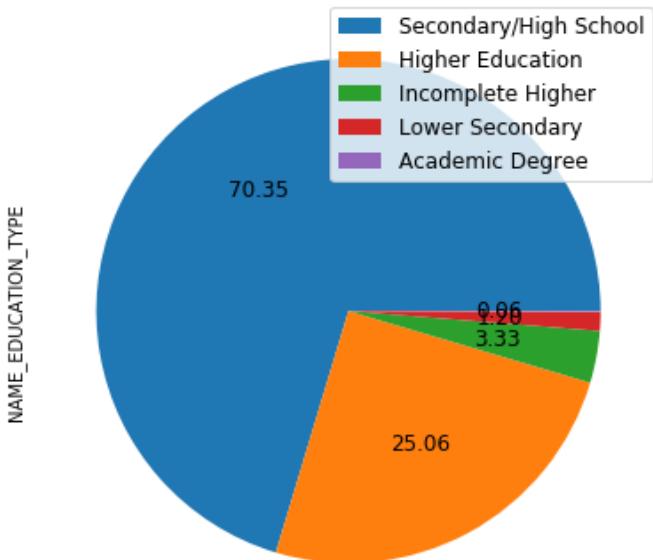
NAME_FAMILY_STATUS



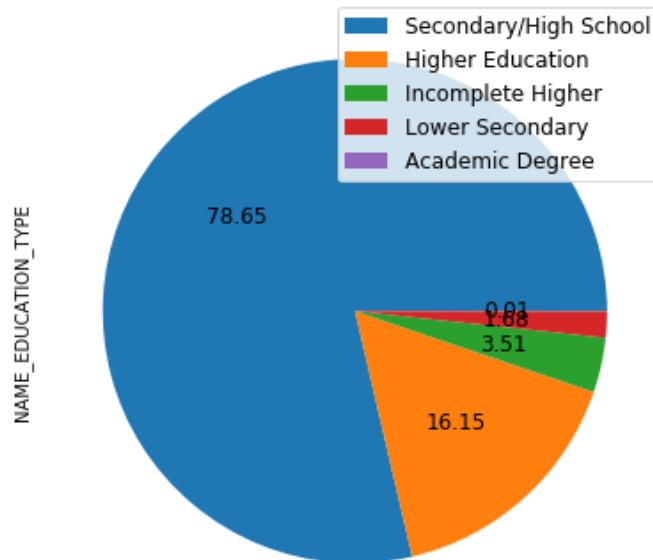
EDUCATION

We observe an **increase** in percentage of loan payment difficulties whose educational qualifications are secondary and a **decrease** in the percentage of loan payment difficulties who have completed higher education when compared the percentages of loan payment difficulties and loan non-payment difficulties.

Education Status of Loan-Non Payment Difficulties



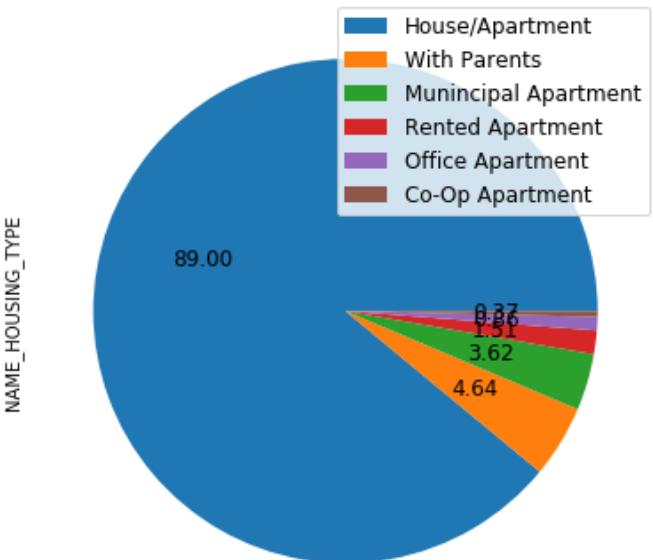
Education Status of Loan Payment Difficulties



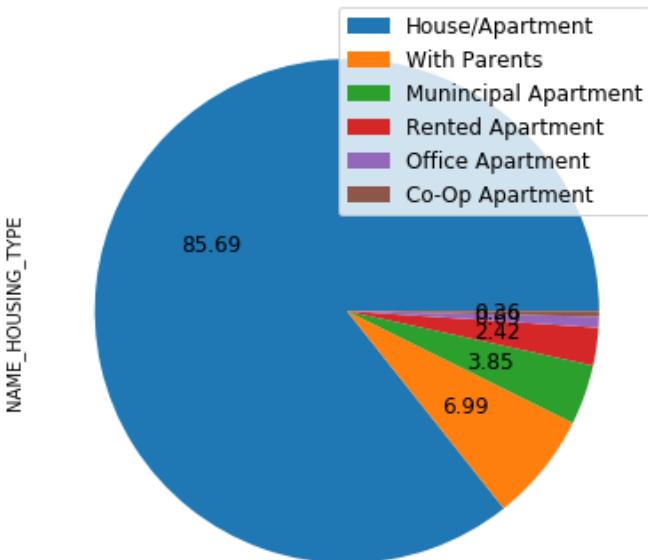
HOUSING

We observe an **increase** in the percentage of Payment Difficulties who live with their parents when compared to the percentages of Payment Difficulties and non-Payment Difficulties.

Education Status of Loan-Non Payment Difficulties



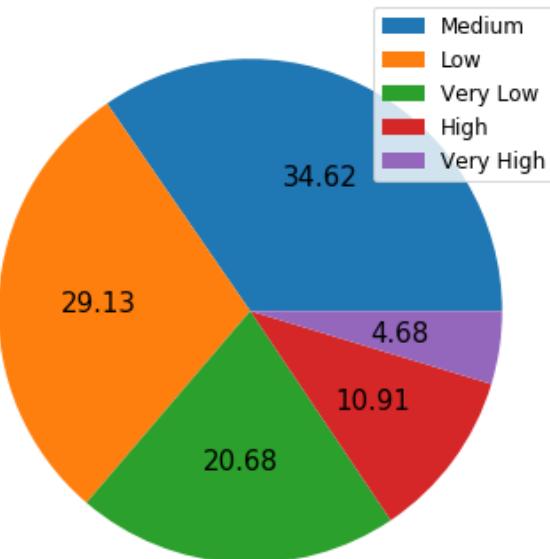
Education Status of Loan Payment Difficulties



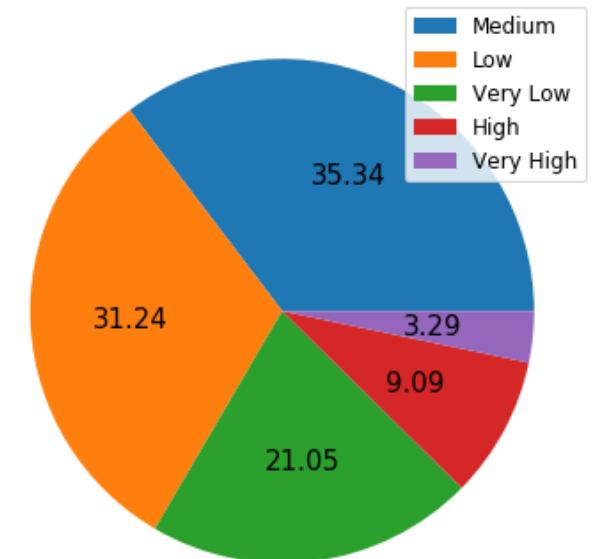
INCOME RANGE

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.

Income Range of Loan-Non Payment Difficulties



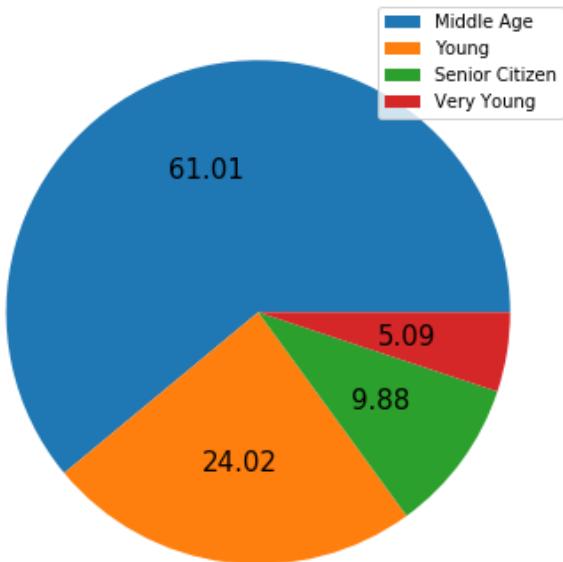
Income Range of Loan Payment Difficulties



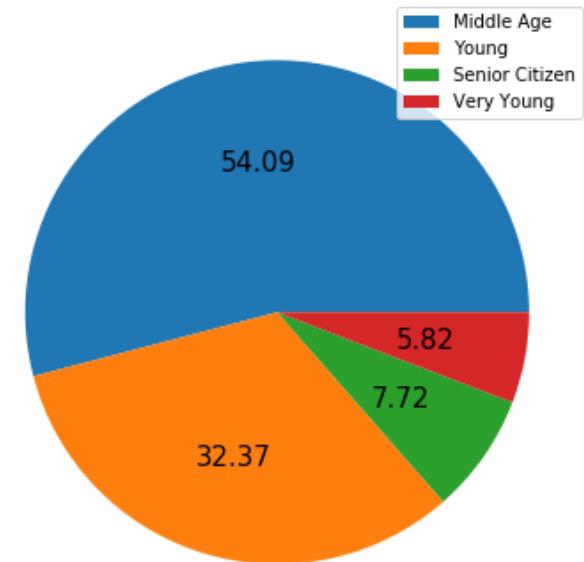
AGE

We observe that there is an increase in the percentage of Loan Payment Difficulties who are young when compared to the percentages Loan-Non-Payment Difficulties.

Age of Loan-Non Payment Difficulties



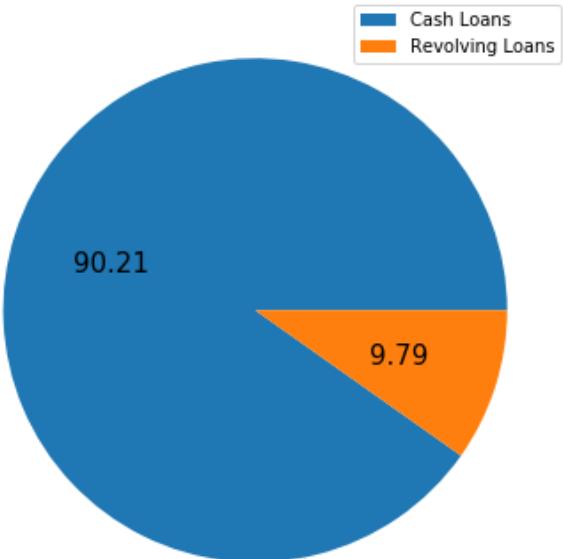
Age of Loan Payment Difficulties



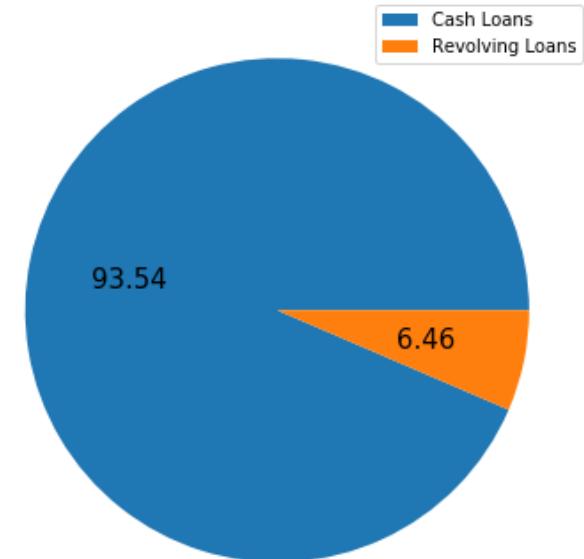
TYPE OF LOAN

We observed that **cash loans** are preferable than revolving loans by both Payment Difficulty & Non-Payment Difficulty applicants.

Loan Contract Type of Loan-Non Payment Difficulties



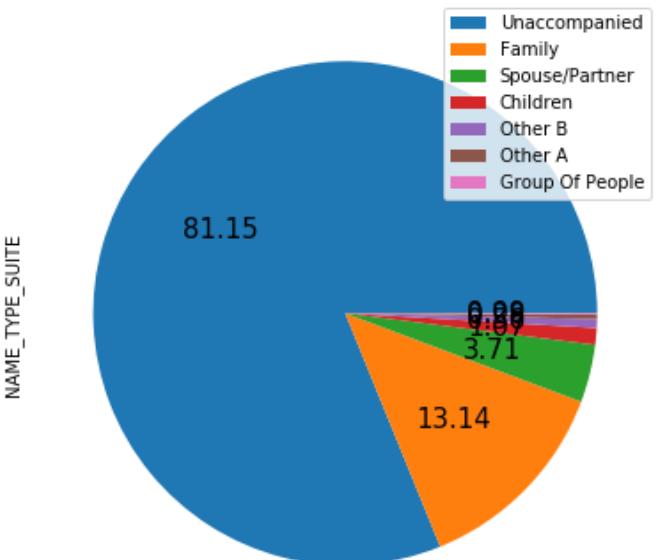
Loan Contract Type of Loan Payment Difficulties



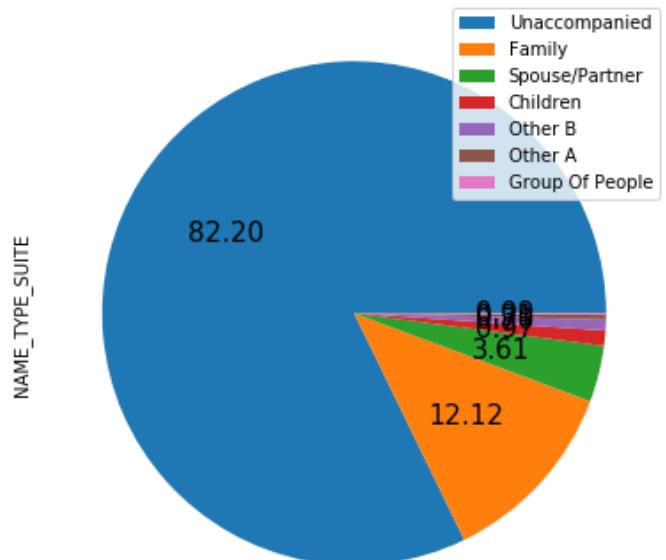
LOAN ACCOMPANIED BY

We do not observe any major changes here.

Loan-Non Payment Difficulties Are Accompanied By

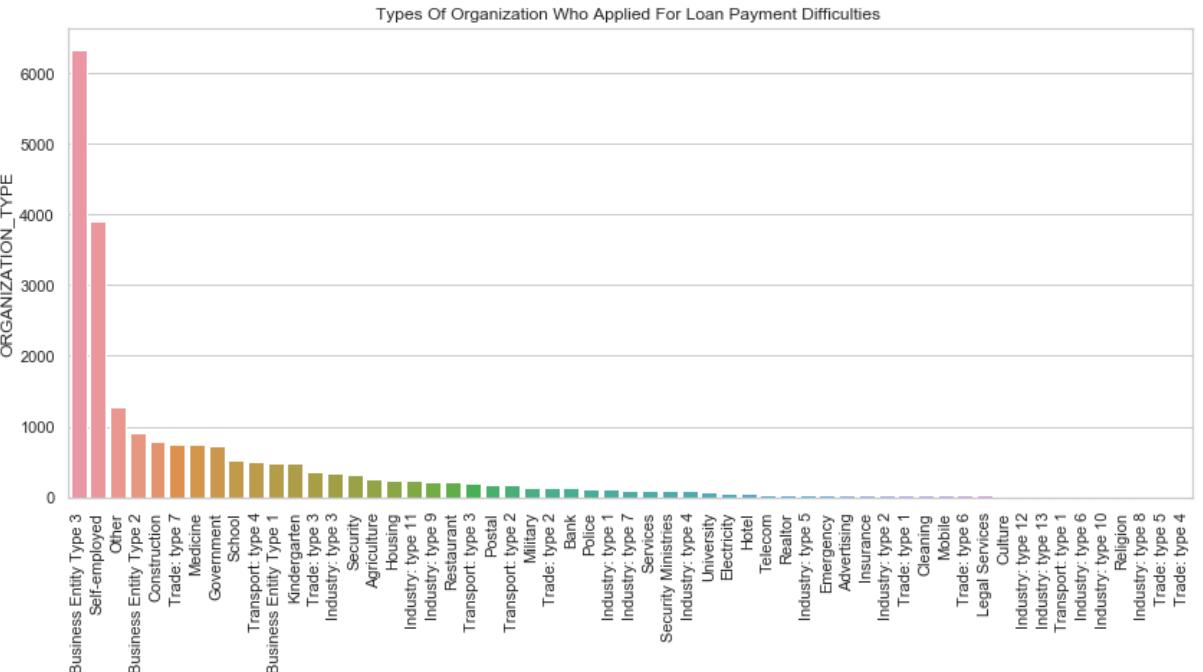
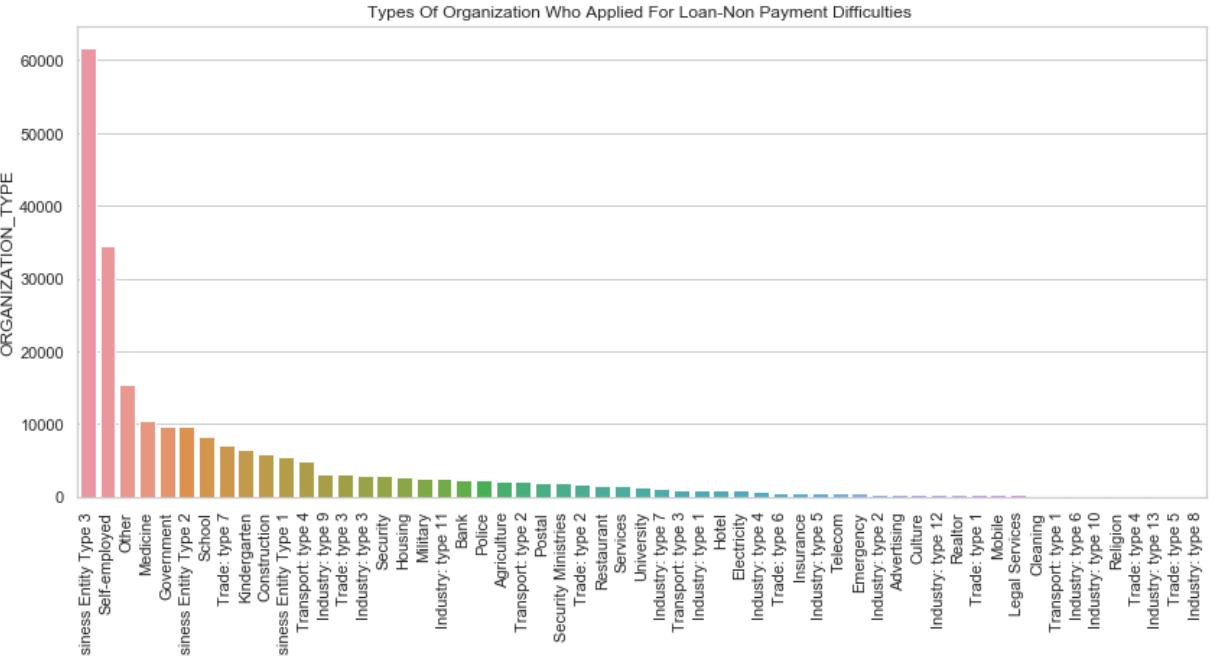


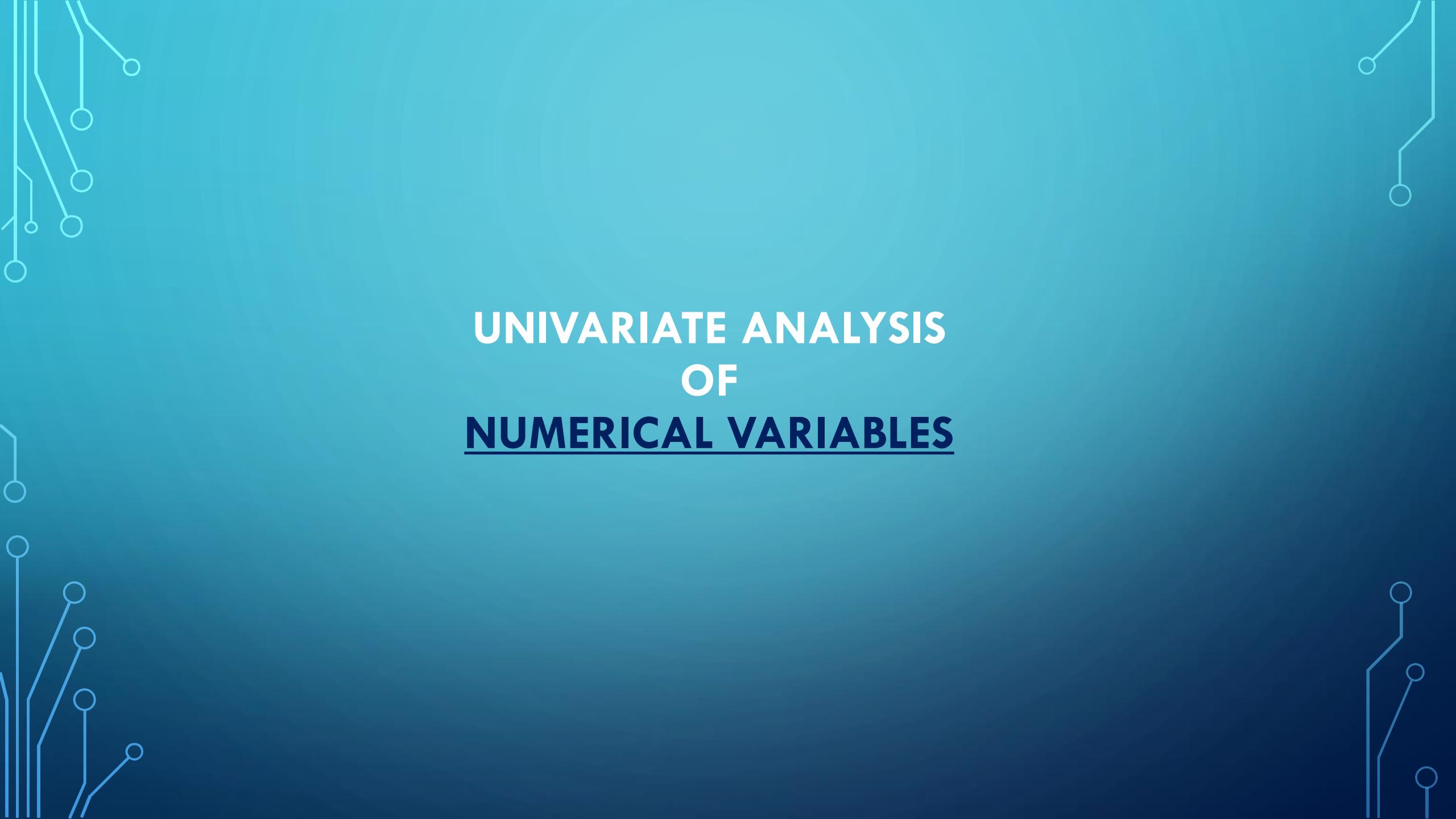
Loan Payment Difficulties Are Accompanied By



ORGANIZATION TYPE

We do not observe any major changes here.



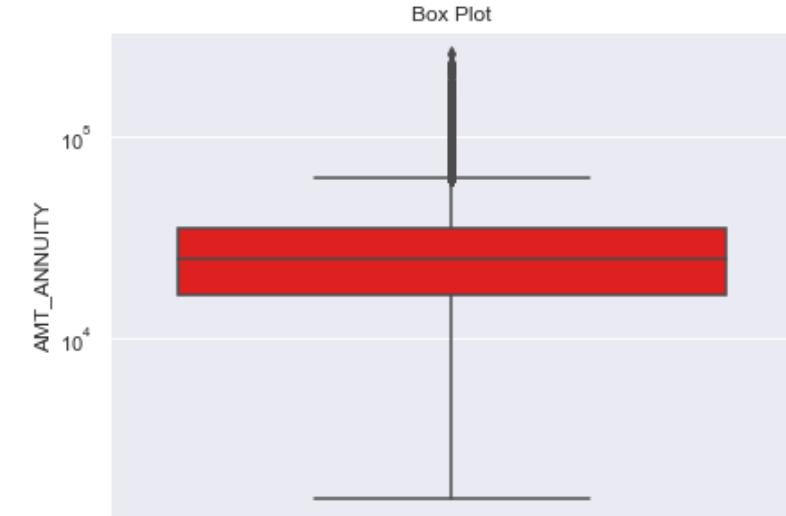
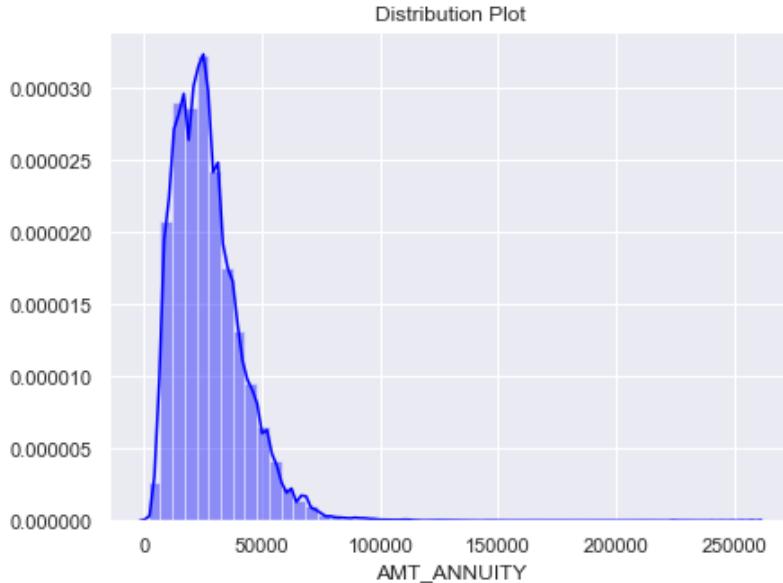


UNIVARIATE ANALYSIS OF NUMERICAL VARIABLES

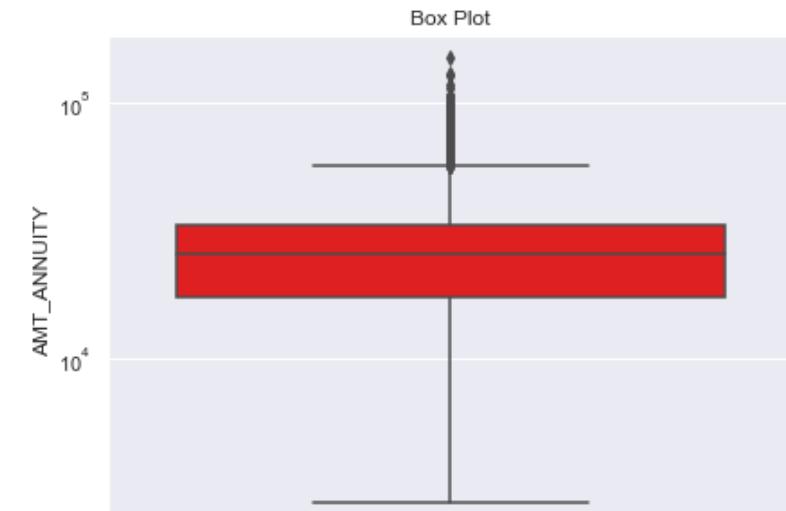
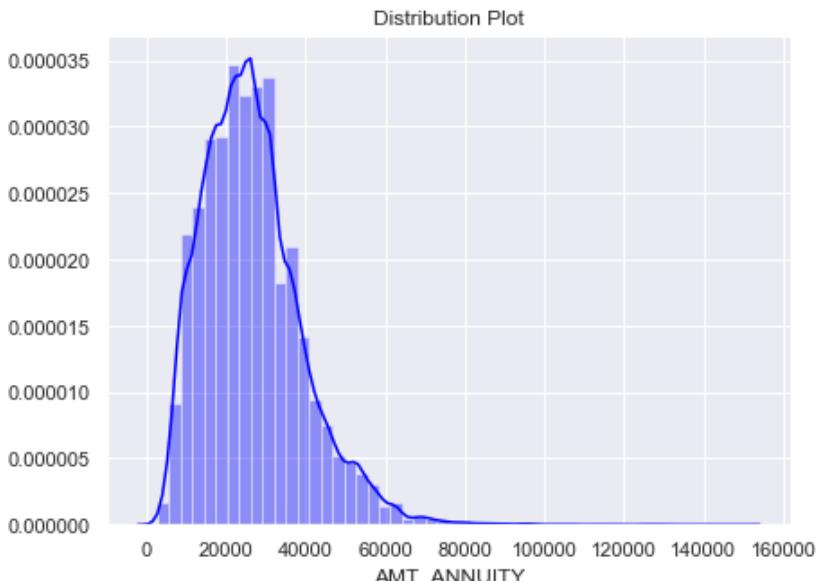
AMOUNT ANNUNITY

We can observe **some outliers** and the **first quartile is bigger than third quartile for annuity amount** which means most of the annuity clients are from first quartile.

Distribution and Box plot for Amount Annuity for Loan Non-Payment Difficulties



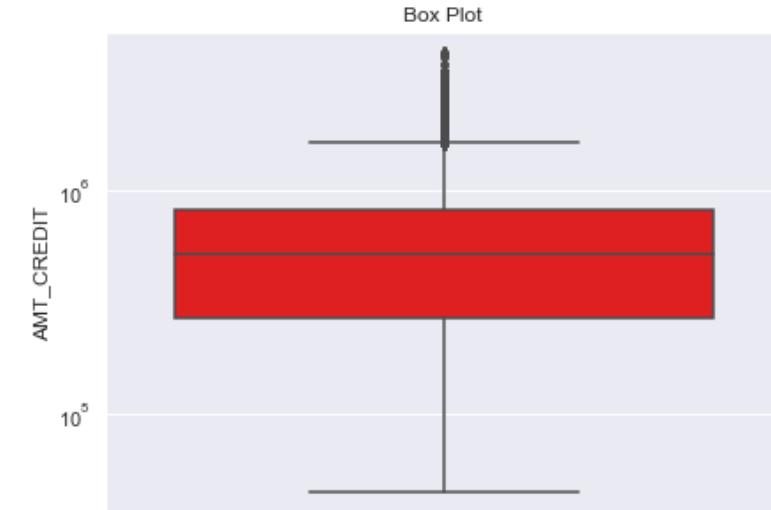
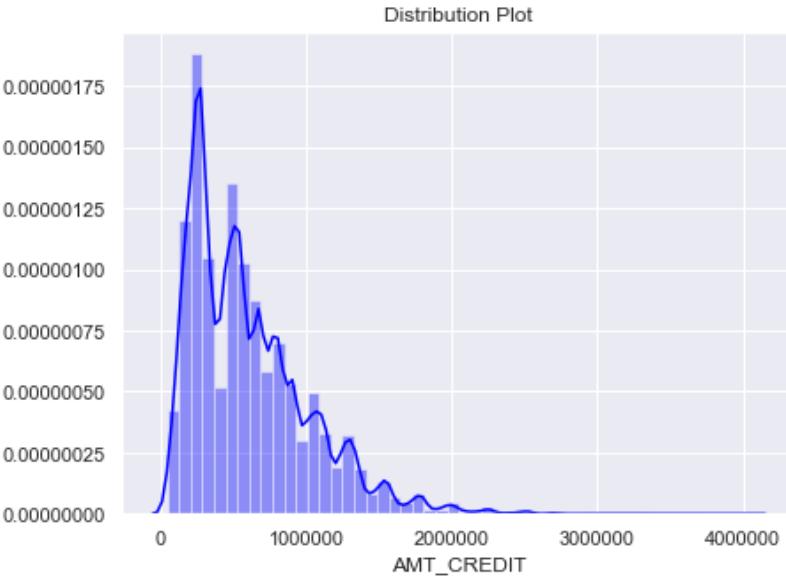
Distribution and Box plot for Amount Annuity for Loan Payment Difficulties



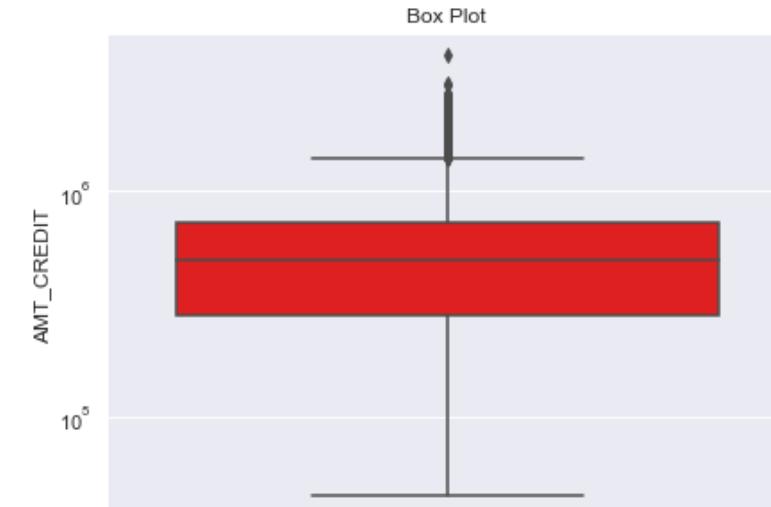
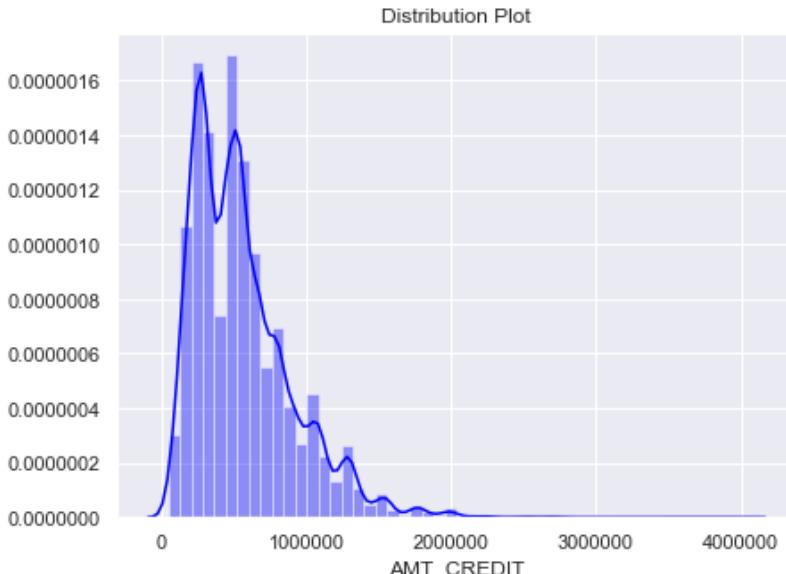
CREDIT AMOUNT

We can observe **some outliers** and the first quartile is bigger than third quartile for annuity amount which means most of the annuity clients are from first quartile. The distribution curve does not appear to be normal or bell curve.

Distribution and Box plot for Credit Amount for Loan Non-Payment Difficulties



Distribution and Box plot for Credit Amount Loan Payment Difficulties



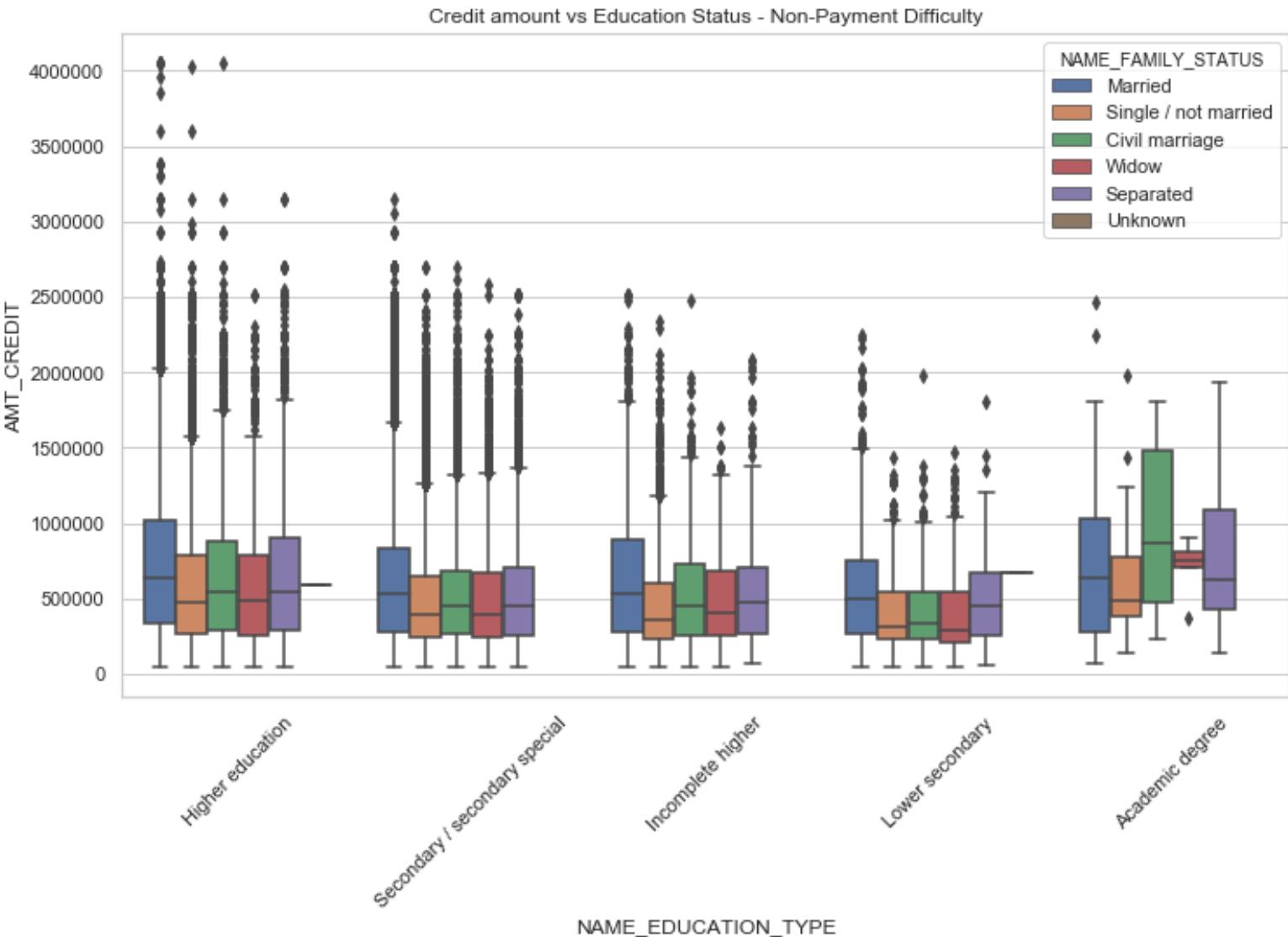
BIVARIATE ANALYSIS

BIVARIATE ANALYSIS OF CATEGORICAL VS NUMERICAL VALUES

CREDIT AMOUNT VS EDUCATION STATUS

TARGET 0 – NON-PAYMENT DIFFICULTY

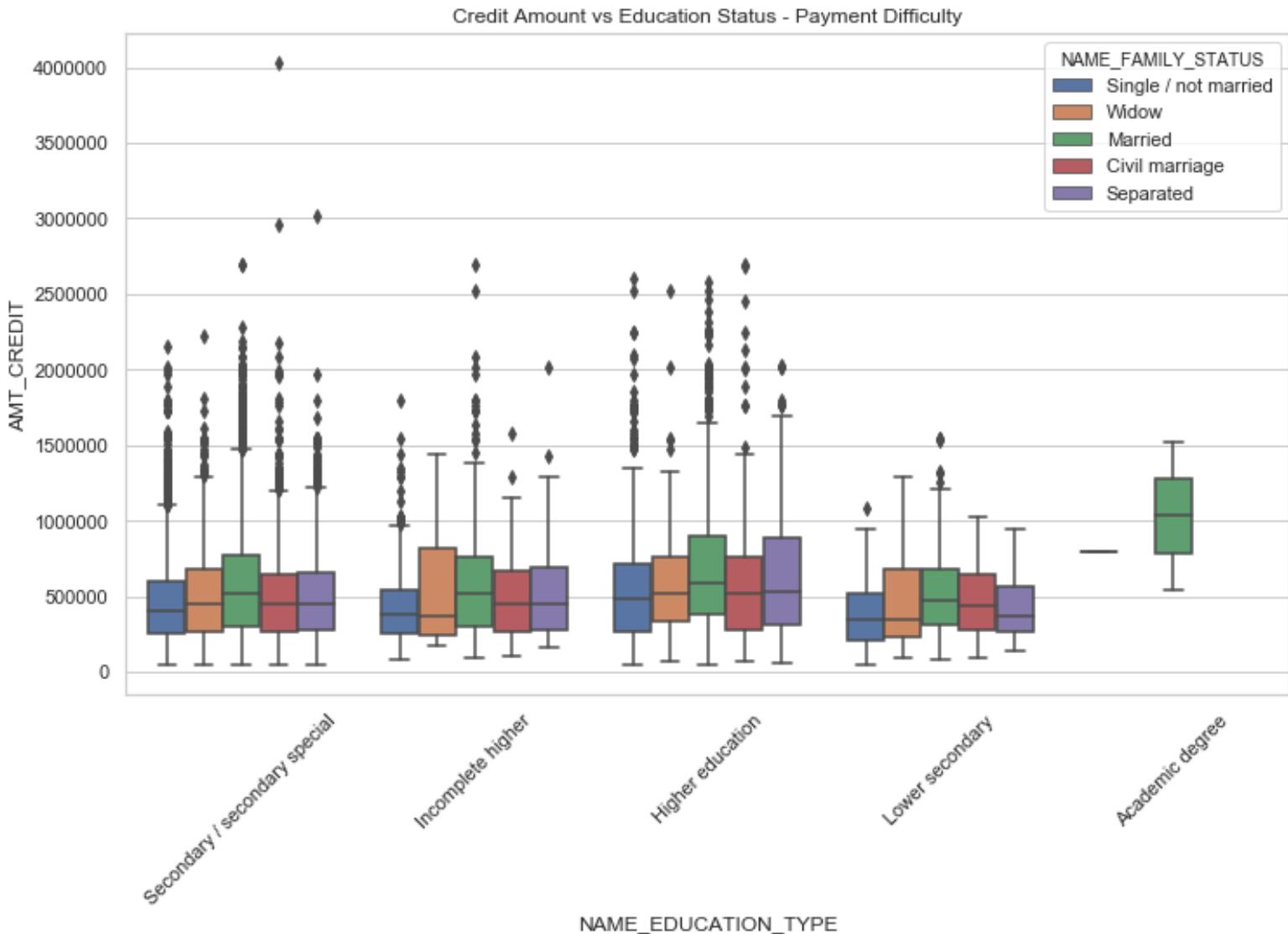
From the above box plot we can conclude that **Family status of civil marriage, marriage and separated of Academic degree education** are having **higher number of credits** than others. Also, **higher education of family status of marriage, single and civil marriage** are having more outliers. Civil marriage for Academic degree is having most of the credits in the third quartile.



CREDIT AMOUNT VS EDUCATION STATUS

TARGET 1 – PAYMENT DIFFICULTY

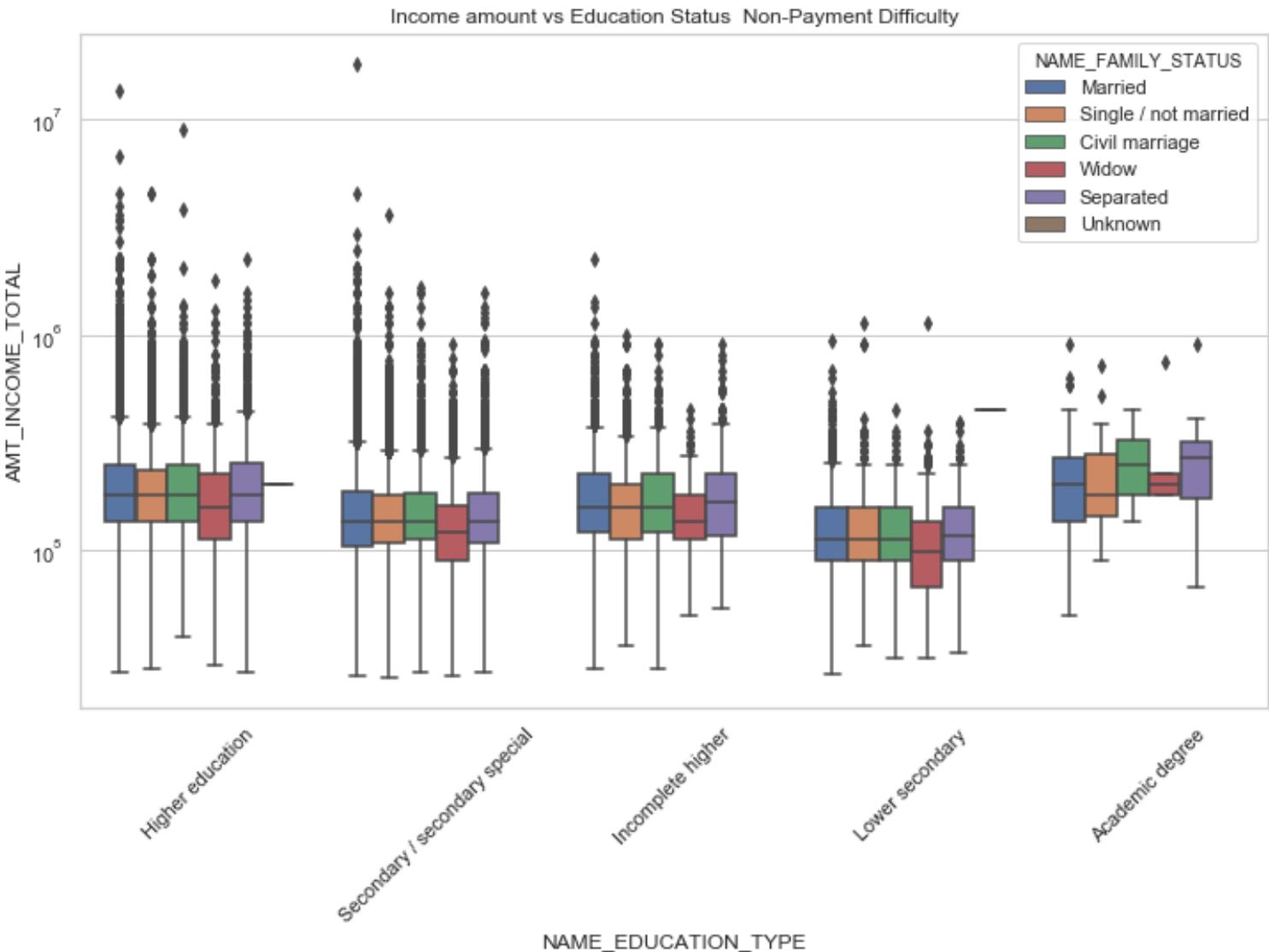
We can say that Family status of **civil marriage**, **marriage** and **separated** of Academic degree education are having **higher number of credits** than others. Most of the outliers are from Education type **Higher education** and **Secondary**. Civil marriage for Academic degree is having **most of the credits in the third quartile**.



INCOME AMOUNT VS EDUCATION STATUS

TARGET 0 – NON-PAYMENT DIFFICULTY

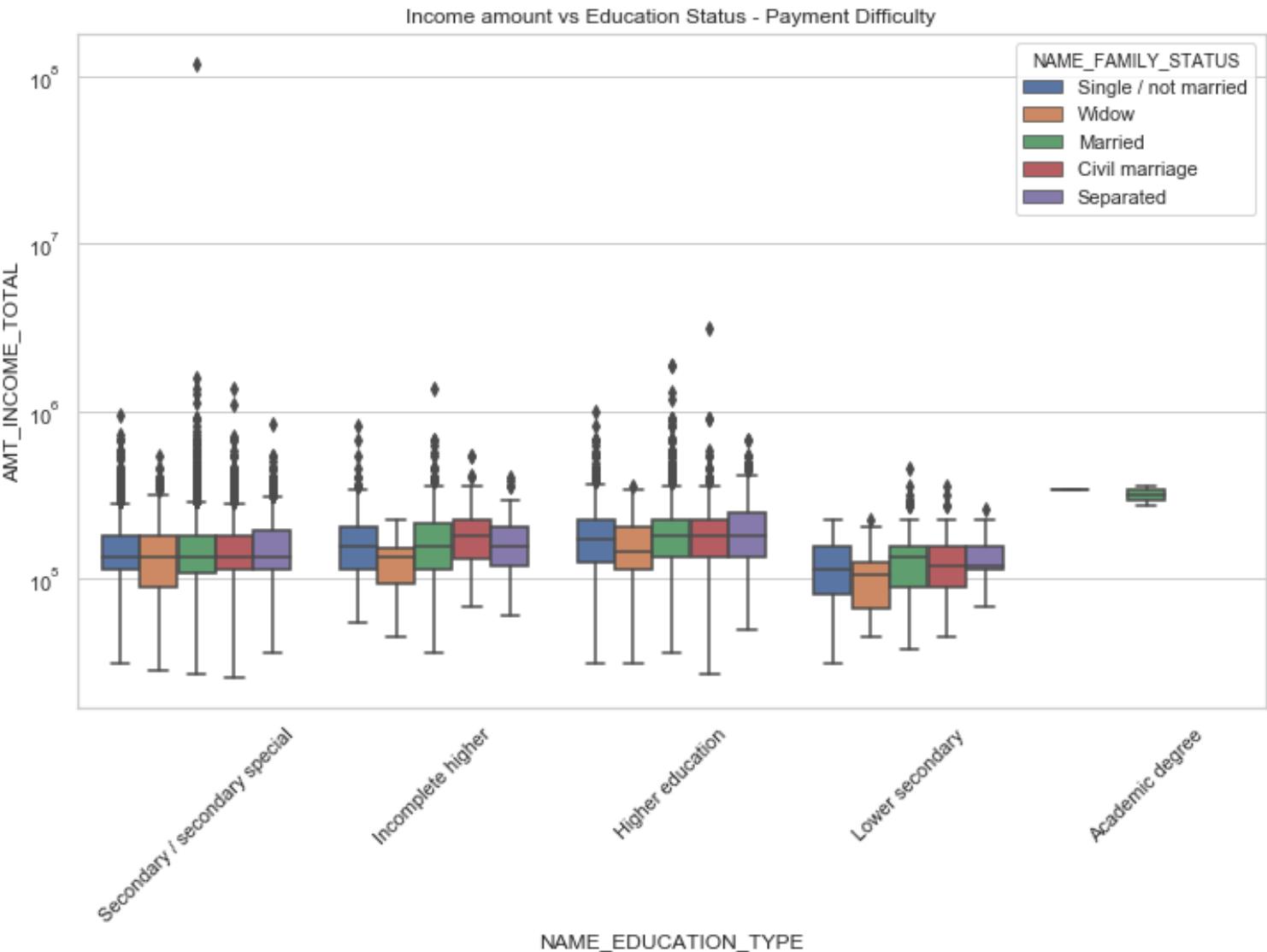
From above boxplot for Education type **Higher education** the income amount is mostly **equal** with family status. It does contain many outliers. **Less outliers for Academic degree** but there income amount is little higher than **Higher education**. Lower secondary of civil marriage family status are having less income amount than others.



INCOME AMOUNT VS EDUCATION STATUS

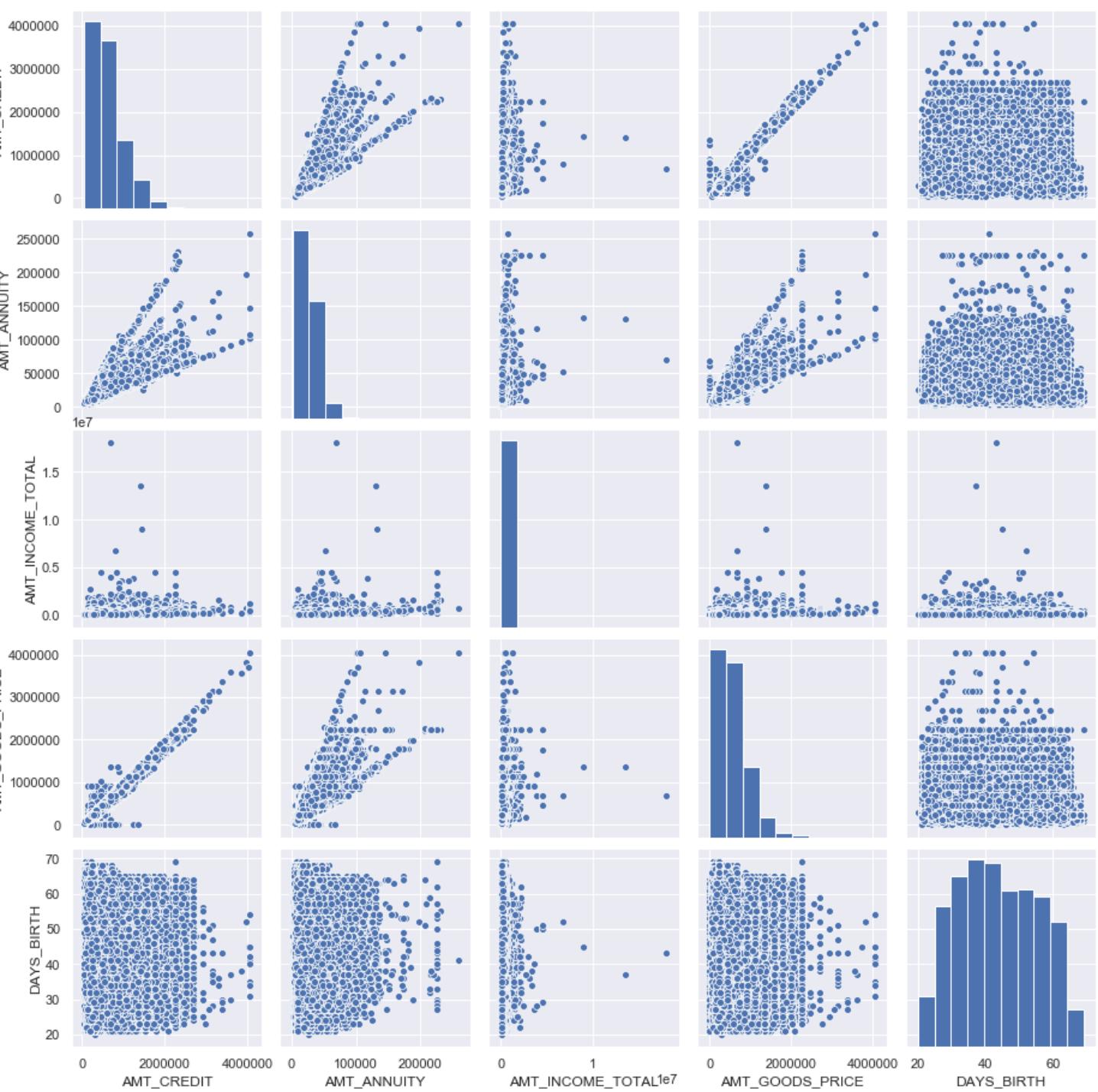
TARGET 1 – PAYMENT DIFFICULTY

From above boxplot for Education type '**Higher education**' the income amount is mostly **equal with family status**. Less outliers for Academic degree but their income amount is little higher than Higher education. **Lower secondary also have less income amount than others.**



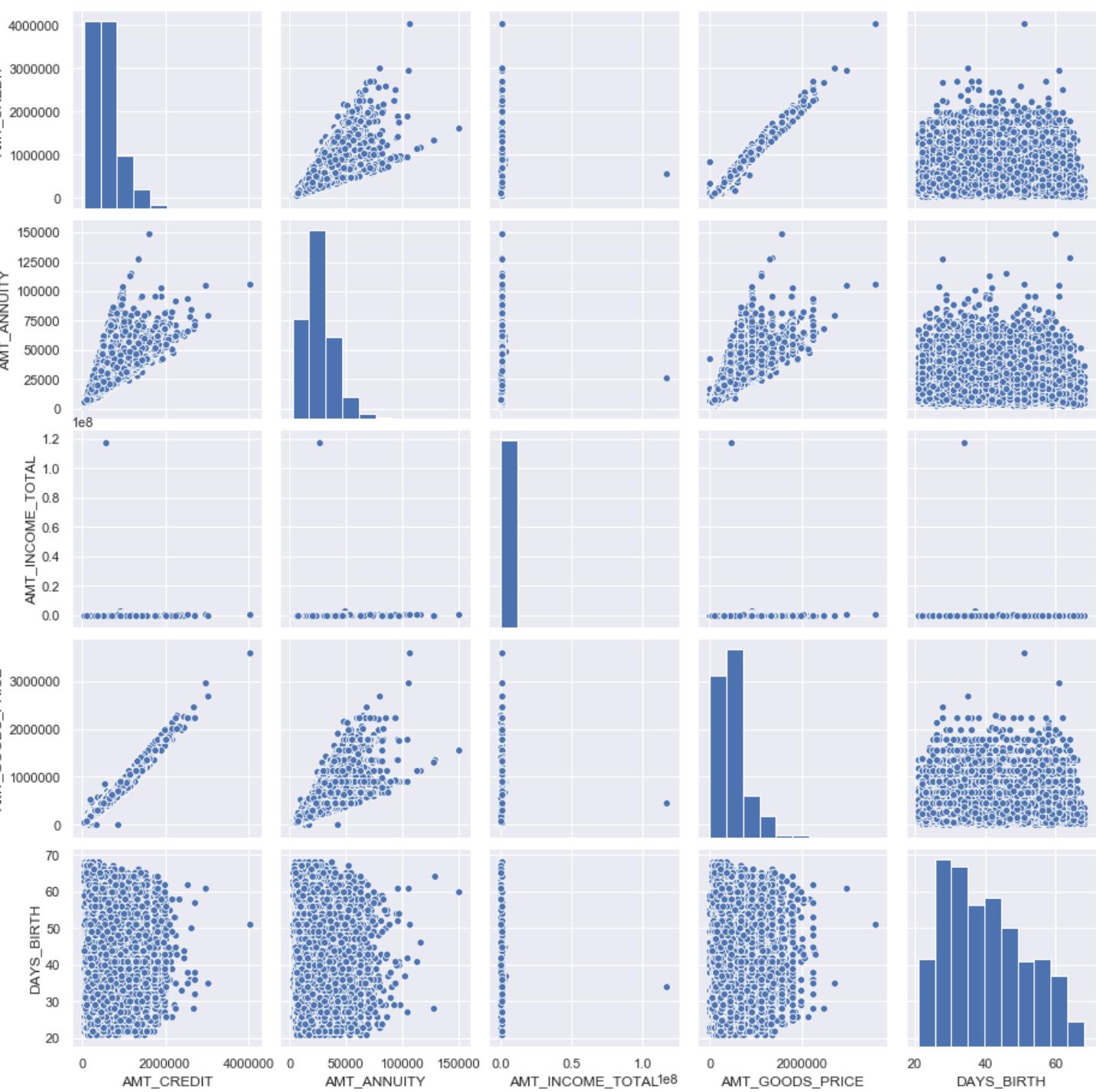
PAIR PLOT FOR TARGET 0

(Loan Non-payment difficulties)



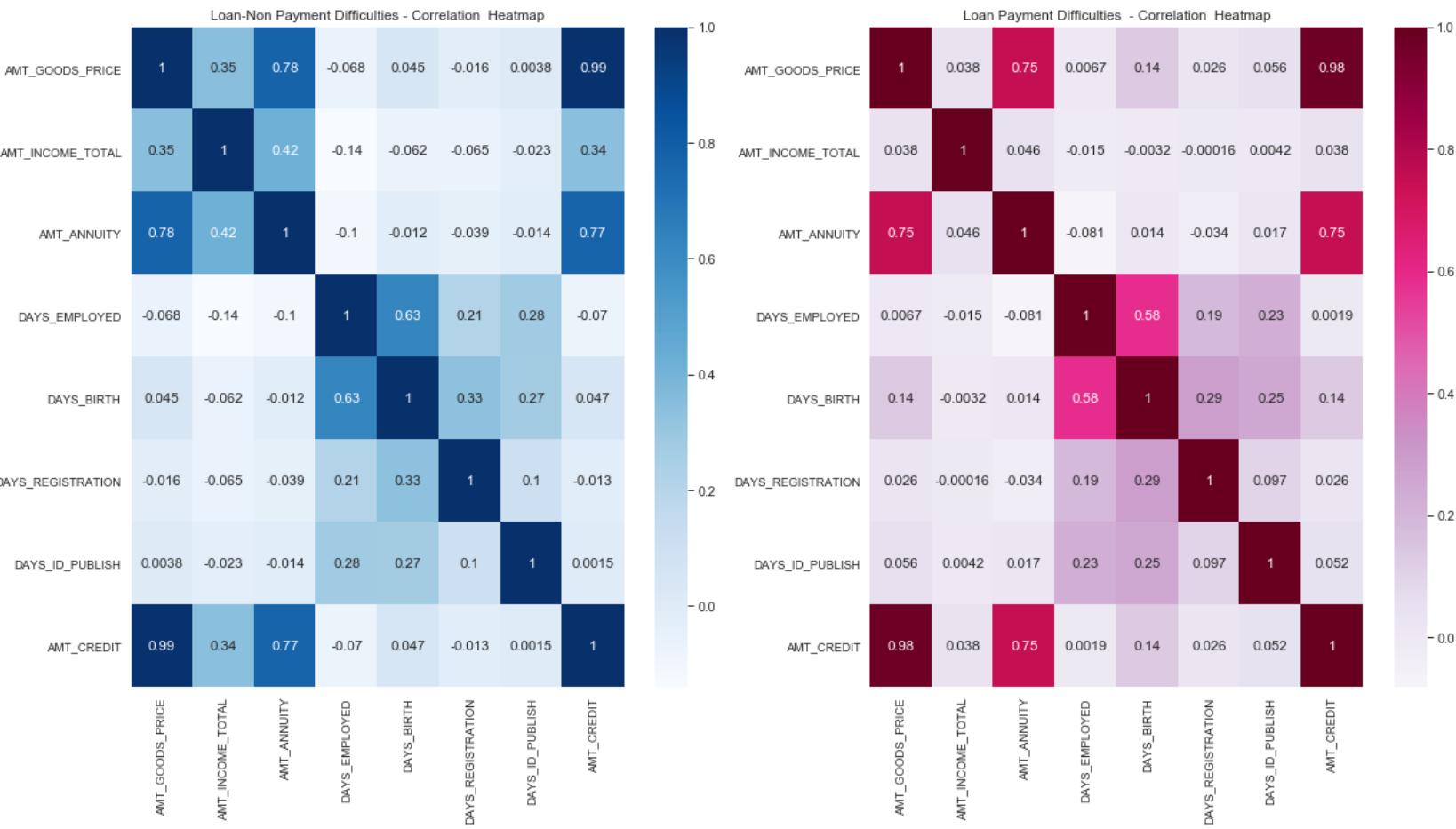
PAIR PLOT FOR TARGET 1

(Loan payment difficulties)



CORRELATION HEATMAP

We observe that there is a 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.



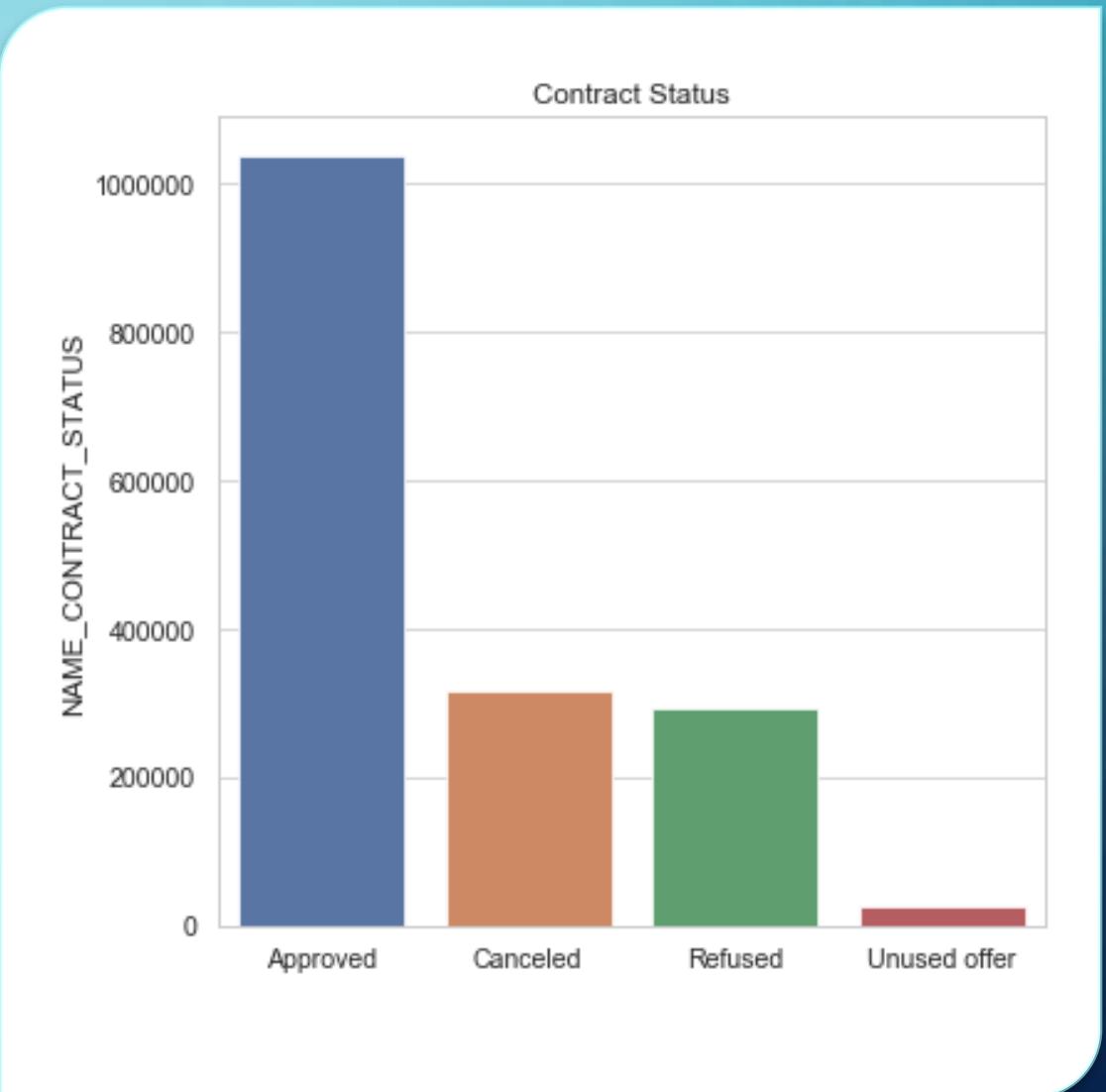


DATA ANALYSIS OF PREVIOUS APPLICATION

UNIVARIATE ANALYSIS

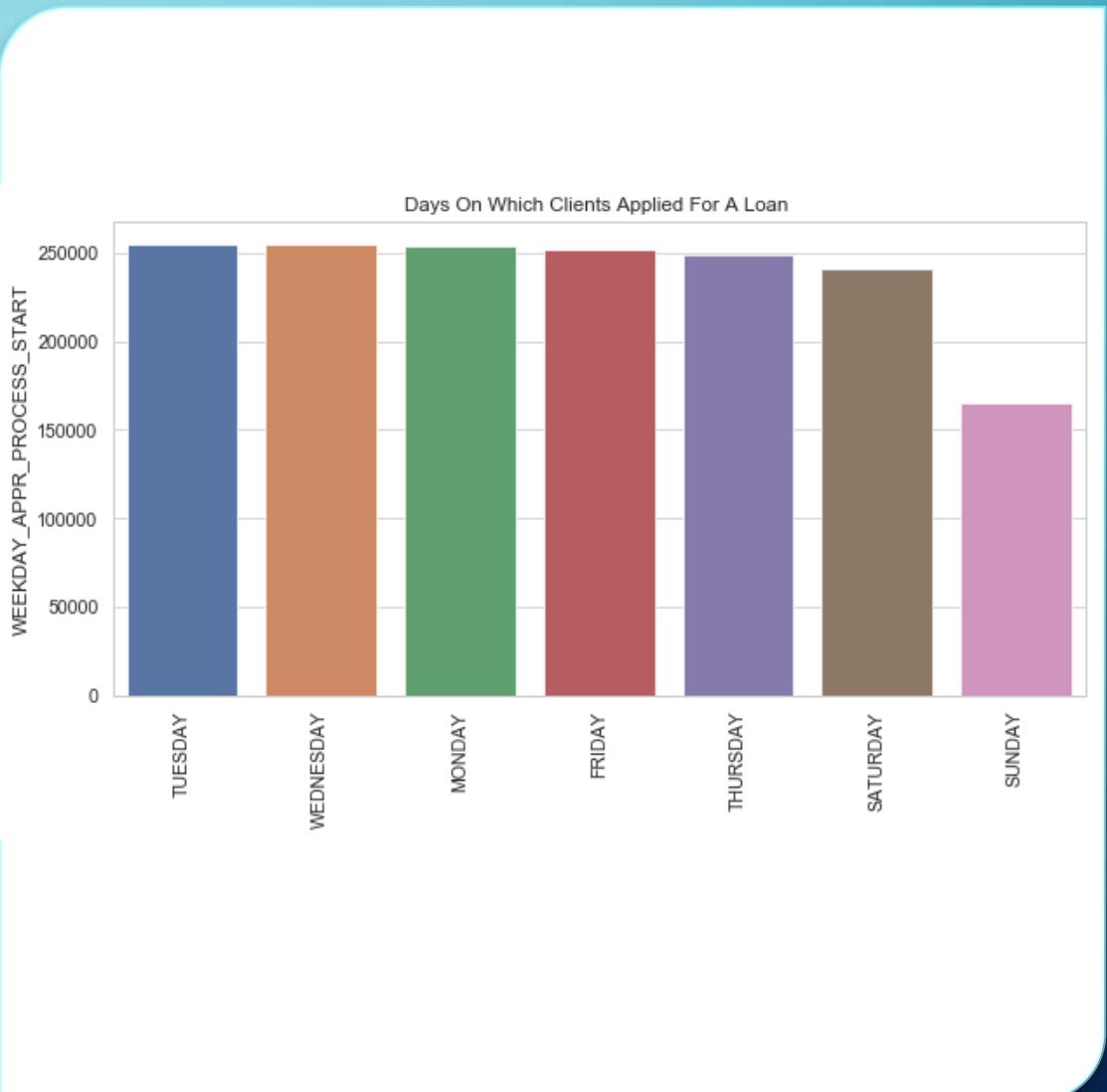
CONTRACT STATUS

We observe that a **majority** of contracts have been passed and a minority of contracts were left unused by the customers.



DAYS ON WHICH CLIENTS APPLIED FOR A LOAN

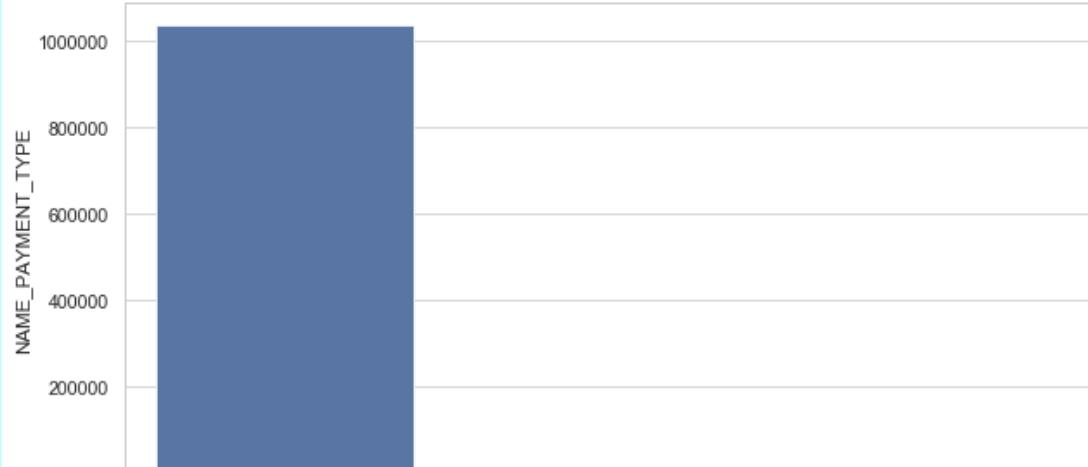
We can observe that **less loans are applied on the weekends.**



PAYMENT TYPE USED BY CLIENTS

99% of the clients chose to **pay**
cash through bank.

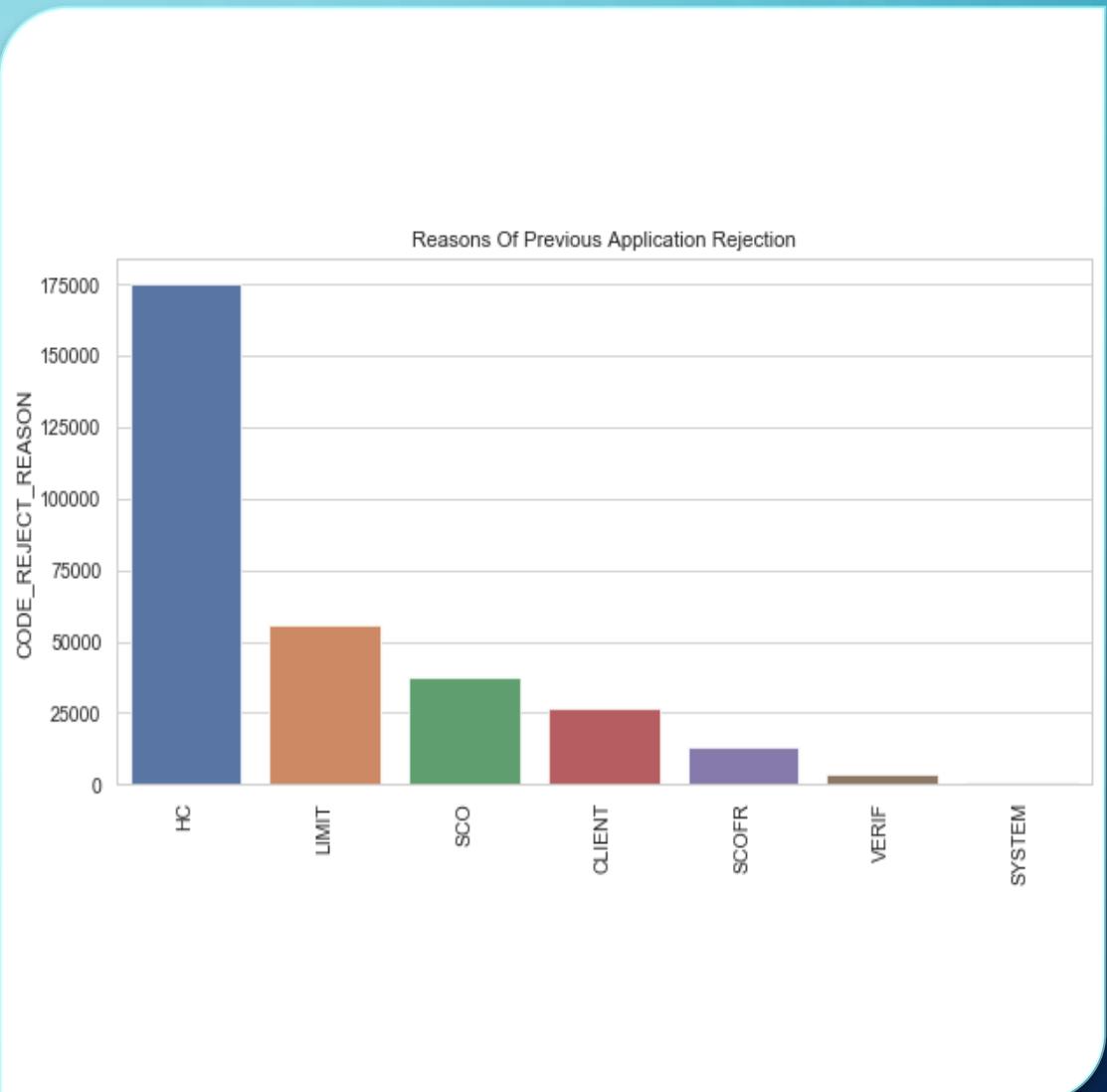
Payment Type Used By Clients



Cashless from the account of the employer

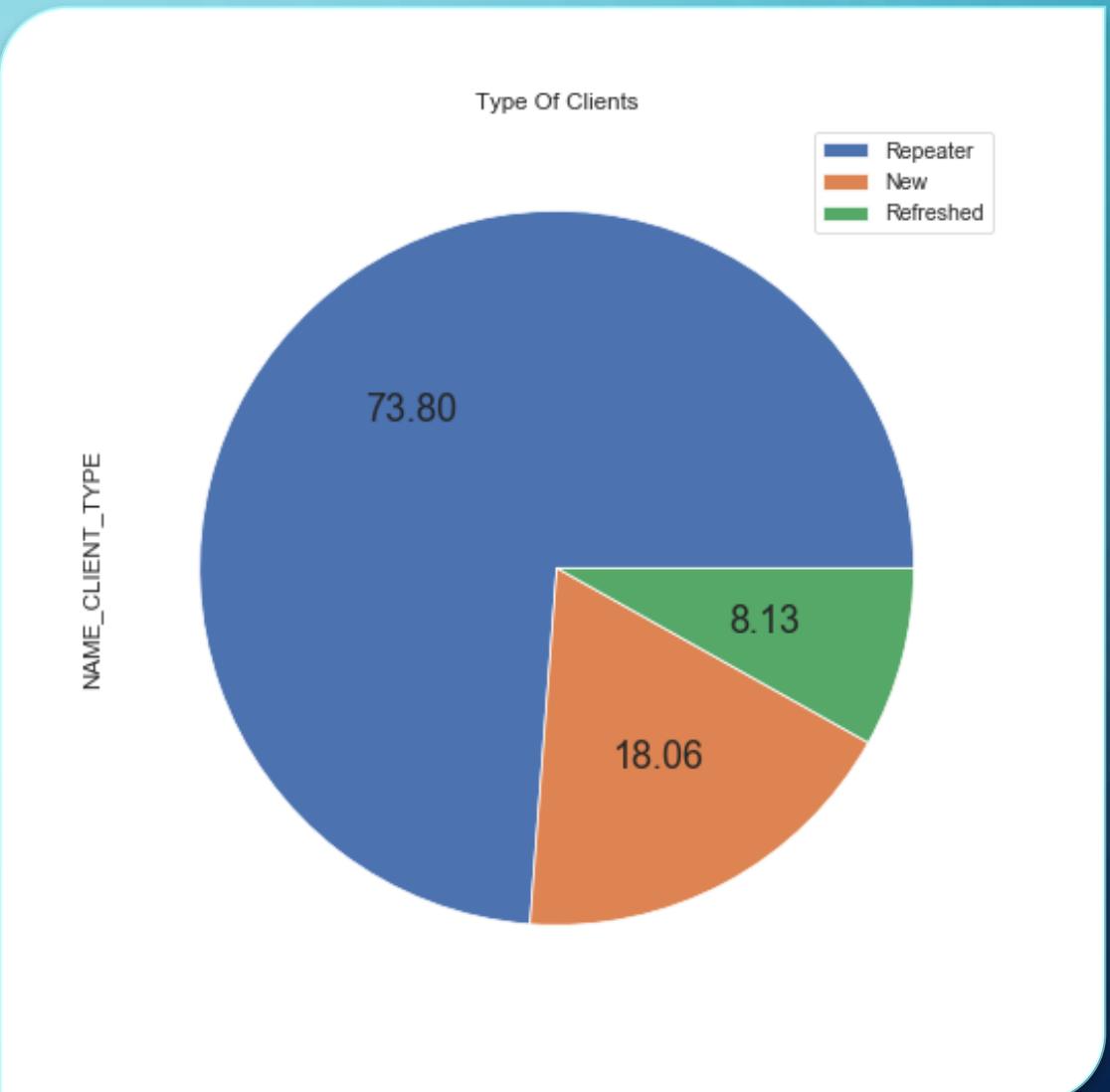
REASONS OF PREVIOUS APPLICATION REJECTION

We observe that HC is the reason majority of applications got rejected.



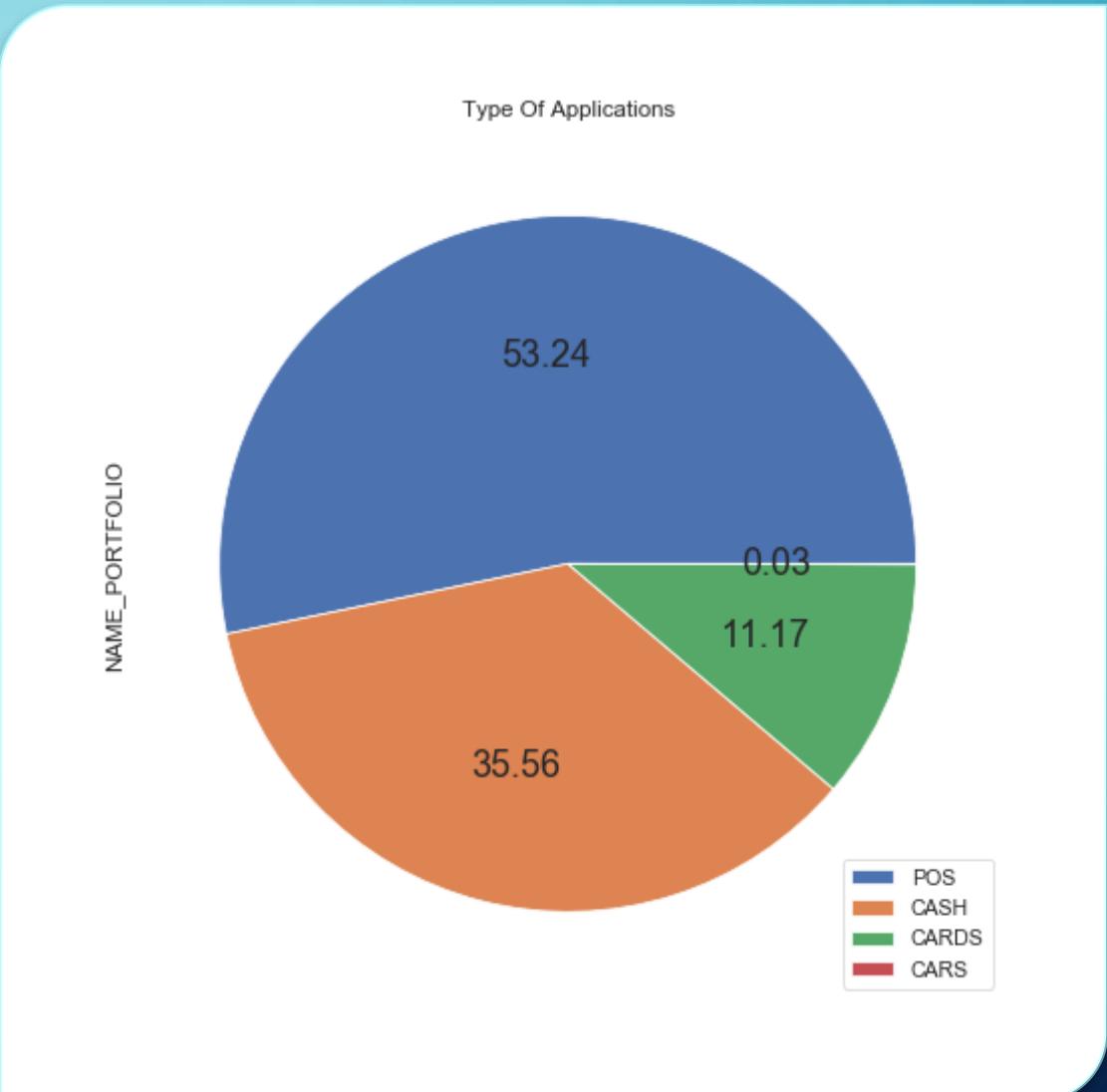
TYPES OF CLIENTS

We observe that majority of clients are **repeaters**.



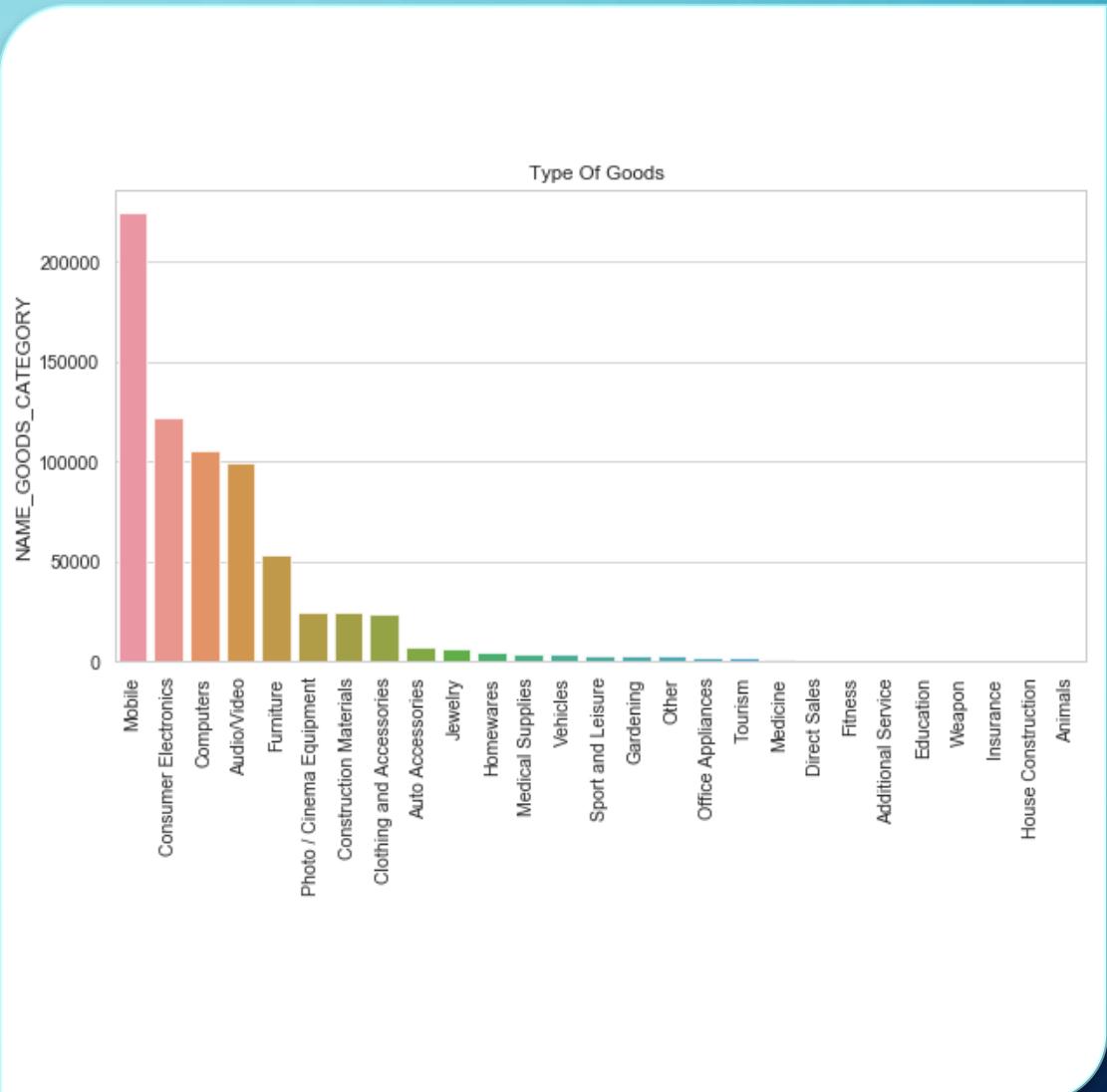
TYPE OF APPLICATIONS

We observe that majority of previous application is for POS and a good amount of it is for cash.



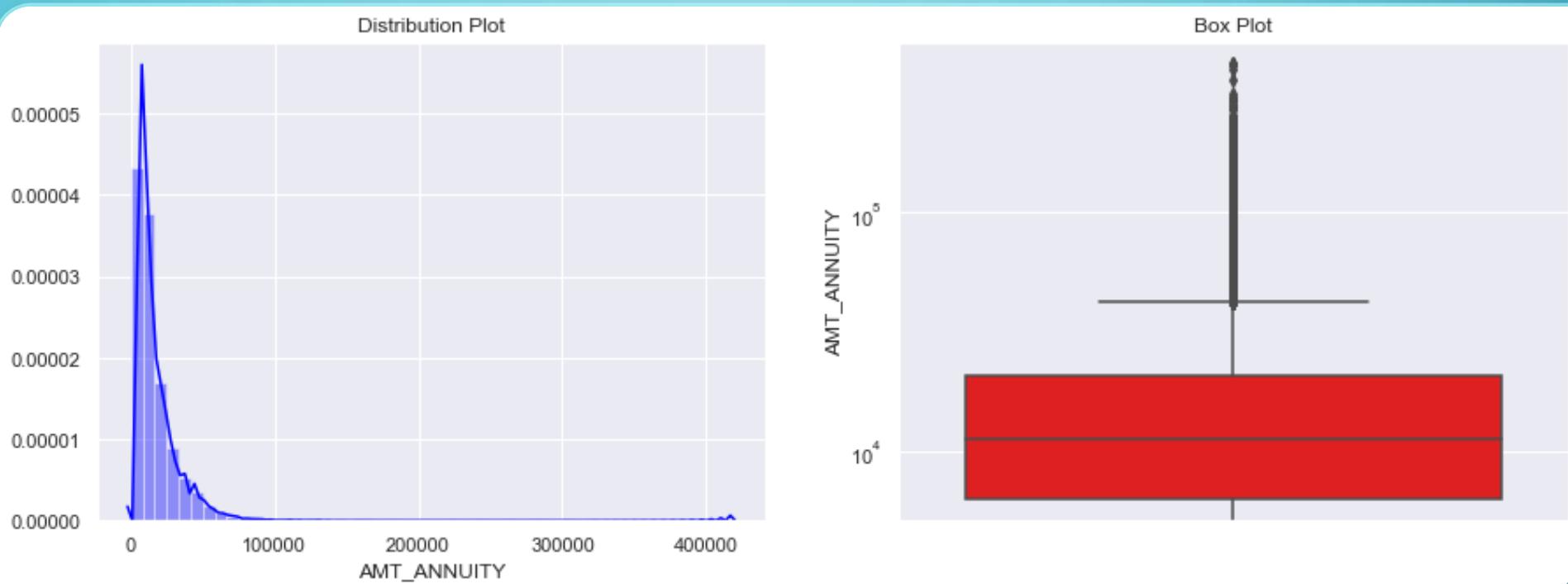
TYPE OF GOODS

We observe that most loans are for mobiles, consumer electronics, computers and furniture's.



UNIVARIATE ANALYSIS FOR NUMERICAL VALUES

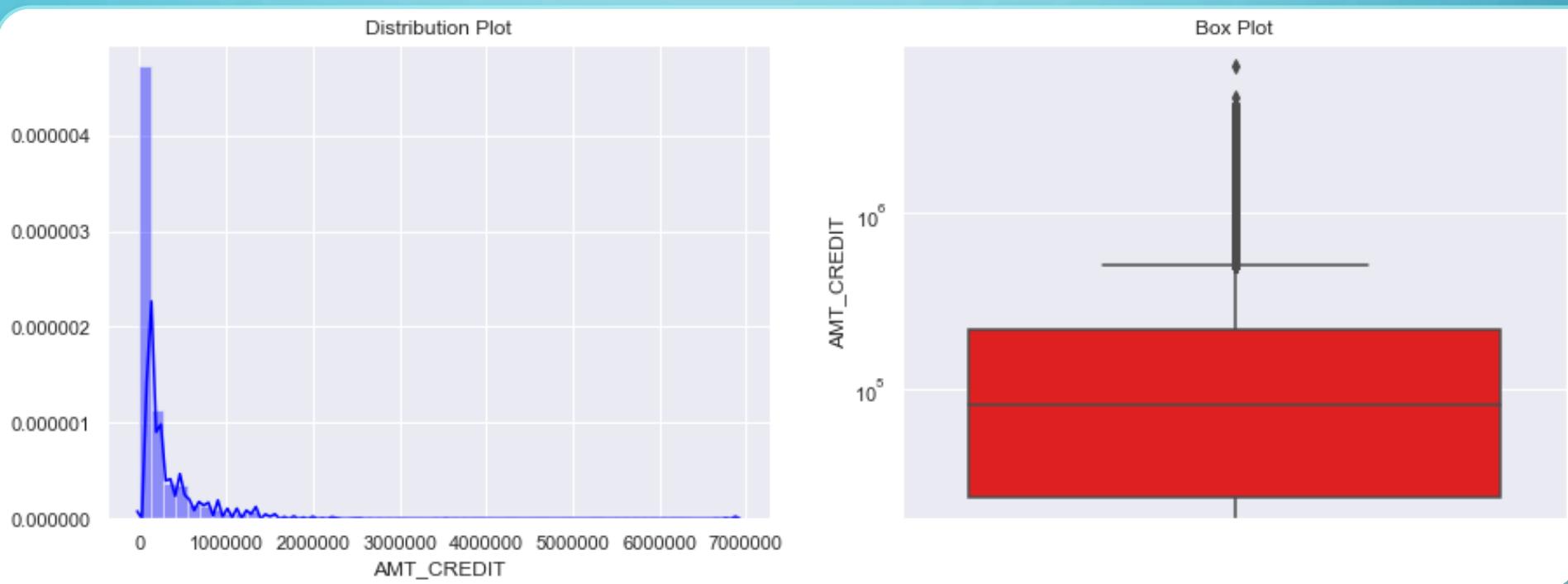
AMOUNT ANNUNITY



We observe that there are **some outliers**, and the curve is not normal or a bell curve.

UNIVARIATE ANALYSIS FOR NUMERICAL VALUES

AMOUNT CREDIT

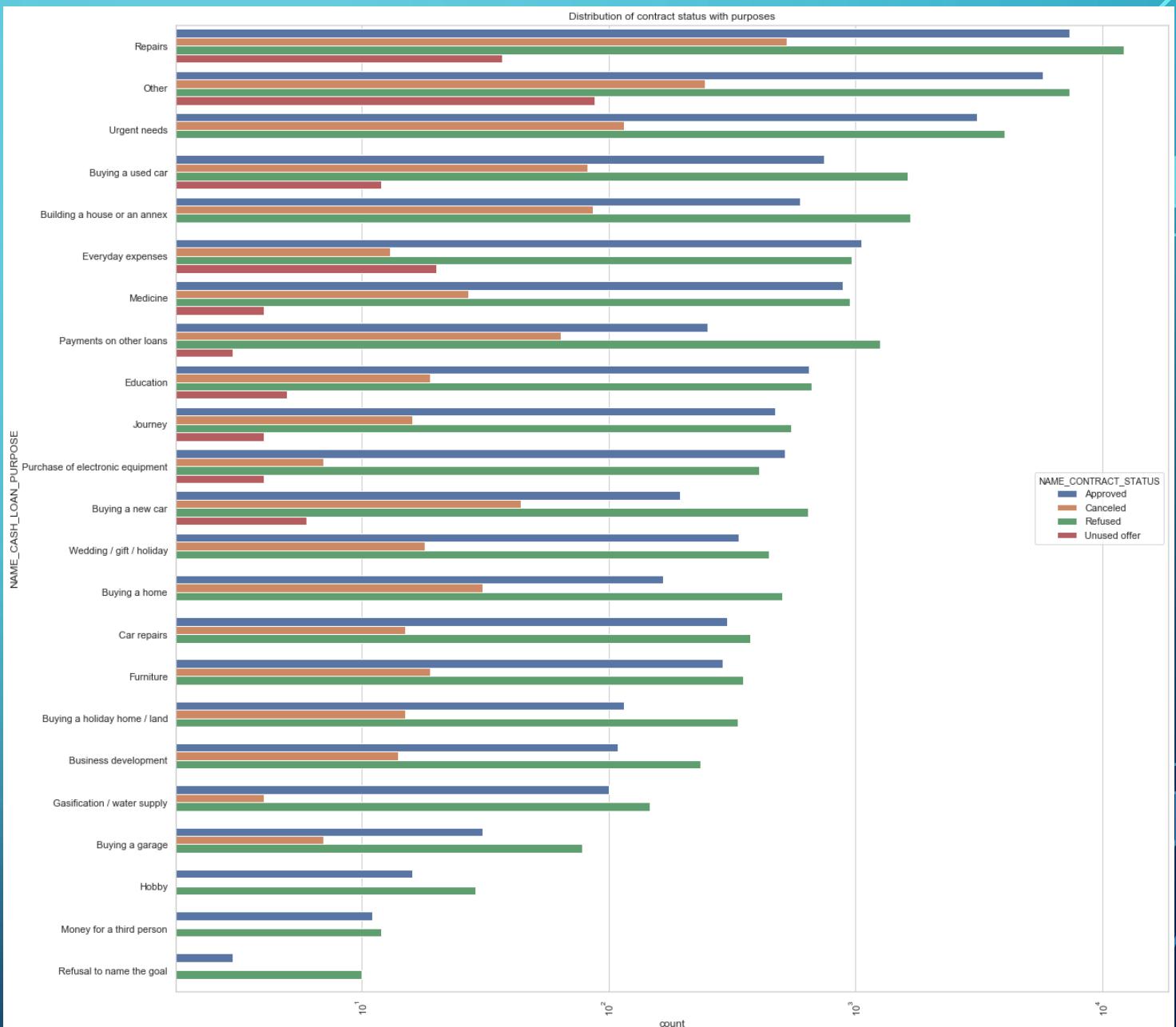


We observe that there are **some outliers**, and the curve is not normal or a bell curve.

DISTRIBUTION OF CONTRACT STATUS WITH PURPOSES

OBSERVATIONS:

1. Most loans rejection came from purpose Repairs.
2. For education purposes we have equal number of approves and rejection.
3. Paying other loans and buying a new car is having significant higher rejection than approves.

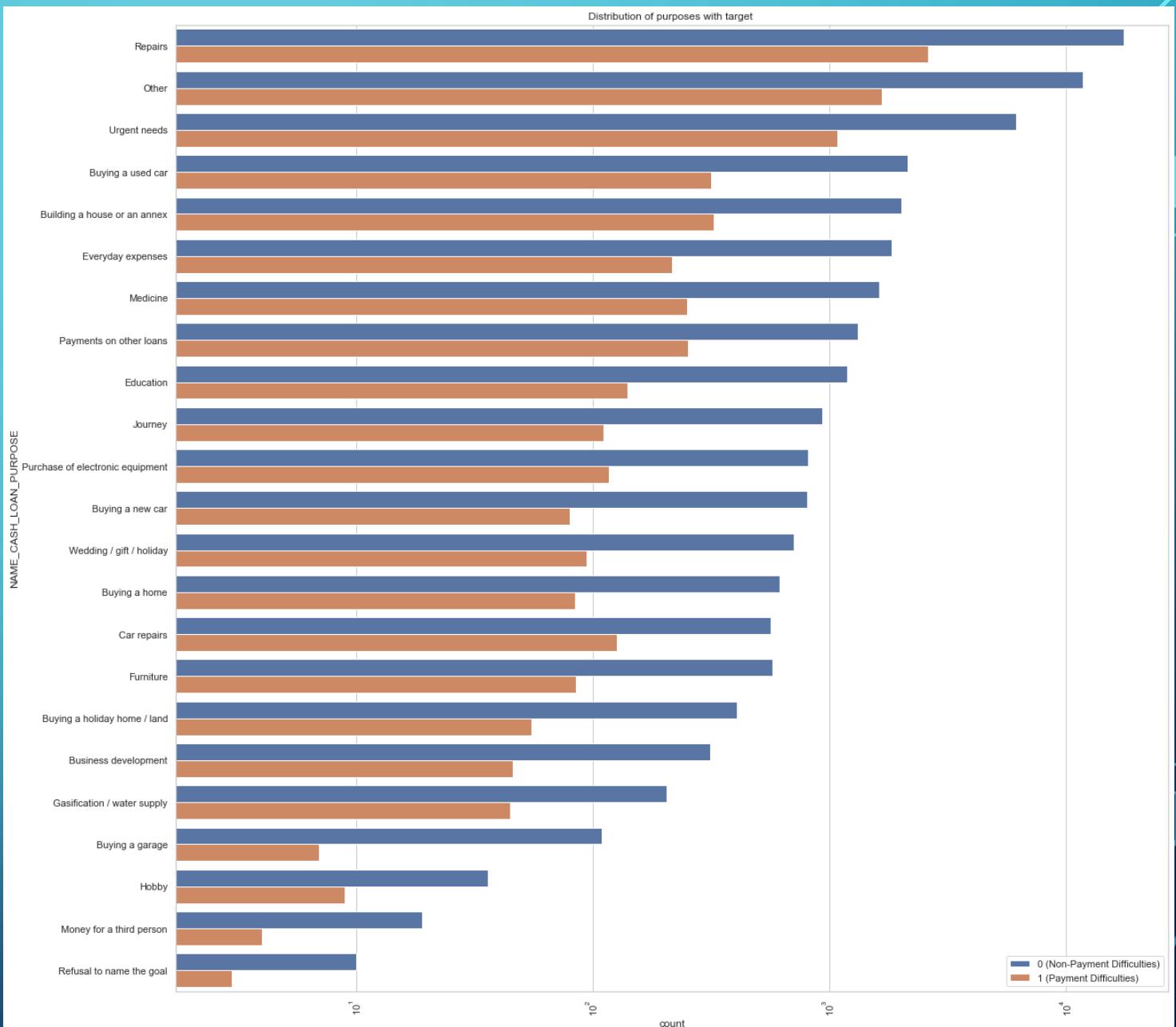


DISTRIBUTION OF PURPOSES WITH TARGET

OBSERVATIONS:

1. Loan purposes with **Repairs** are facing more difficulties in payment on time.
2. There are few places where loan payment is significantly higher than facing difficulties. They are '**Buying a garage**', '**Business development**', '**Buying land**', '**Buying a new car**' and '**Education**'.

Hence, we can focus on these areas for which the client is having minimal payment difficulties.



INSIGHTS DRAWN FROM THE CASE-STUDY

APPLICATION DATA

1. Banks should focus more on contract types such as `Pensioner` , `Businessman` , `Student` .
2. Banks should focus more on housing types other than `Co-op apartment ` for successful payments.
3. Banks need to focus less on income type `Working` , as they have the greatest number of unsuccessful payments.
4. Clients with housing type `With Parents 'have least number of unsuccessful payments so banks can focus more on them.
5. Count of `Maternity Leave` in `NAME_INCOME_TYPE` is very less while it has maximum percentage of payment difficulties. This means that they are `Loan Defaulters` and we should avoid giving loans to them.
6. Count of `Low skilled Laborers` in `OCCUPATION_TYPE` is very less and it also has maximum percentage of payment difficulties. Hence, such clients become Loan Defaulters and banks need to be cautious with such clients.
7. Count of `Lower Secondary` in `NAME_EDUCATION_TYPE` is very less while it has maximum percentage of payment difficulties. Hence, client with education type as `Lower Secondary` are the driving factors for Loan Defaulters and banks should avoid such clients so they remain profitable.

INSIGHTS DRAWN FROM THE CASE-STUDY

PREVIOUS APPLICATION DATA

1. Count of 'Refusal to name the goal' in 'NAME_CASH_LOAN_PURPOSE` is comparatively lesser while it has maximum percentage of payment difficulties. Therefore, clients who have 'Refused to name the goal` for cash loan in previous application are the driving factors for Loan Defaulters.
2. Count of 'Refused` in 'NAME_CONTRACT_STATUS` is lesser while it has maximum percentage of payment difficulties. Hence, client with contract status as 'Refused` in previous application are the driving factors for Loan Defaulters.
3. Clients with 'Revolving loans` and with 'Refused` previous application tend to have more payment difficulties in current application. Since the count of both 'Revolving loans` and 'Refused` is comparatively less , therefore clients with 'Revolving Loans` and 'Refused` from previous application are driving factors for Loan Defaulters.

THANK YOU