Presentation

# Credit Exploratory Data Analysis: Case Study – Assignment

By

Deeksha Sharma

# PROBLEM STATEMENT

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.

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, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicant's using EDA is the aim of this case study

# What are the DATASETS PROVIDED FOR ANALYSIS?

There are Two major Datasets provided which are mentioned below:

1. Application Data

2. Previous Application Data

These file provided in “Comma Separated Values” Format.(.csv)

Another file provided with Column Descriptions for defining and understanding each columns contribution.

Prerequisites:

1. Programming Language: Python.

2. Platform: Jupyter Notebook

3.Libraries: Pandas, Numpy, Matplotlib, Seaborn, Itertools and some Warnings

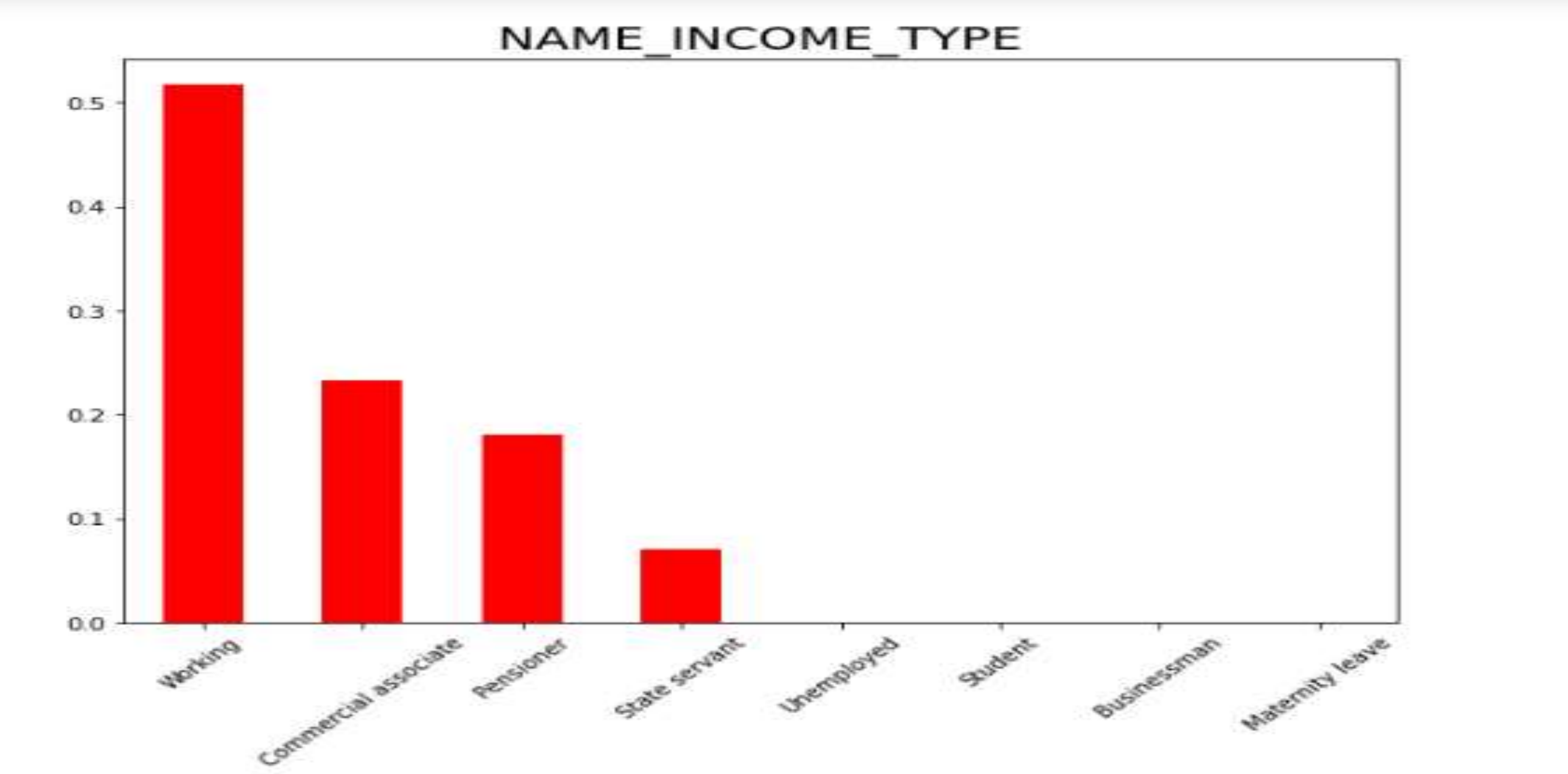
# Assigning Variables:

| Description | Assigned Variable |
|---|---|
| Data Set - 1: "Application_data.csv" | app_df |
| Data Set - 2: "previous_application.csv" | Pre_app_df |
| For Null values defined as | null |
| Merging the two files | New_left_prev |
| Loan Non-Payment Difficulties | target0 |
| Loan Payment Difficulties | target1 |
| | |

Note: There are many other Variables used in data analysis process those variables are mentioned in Jupyter Notebook

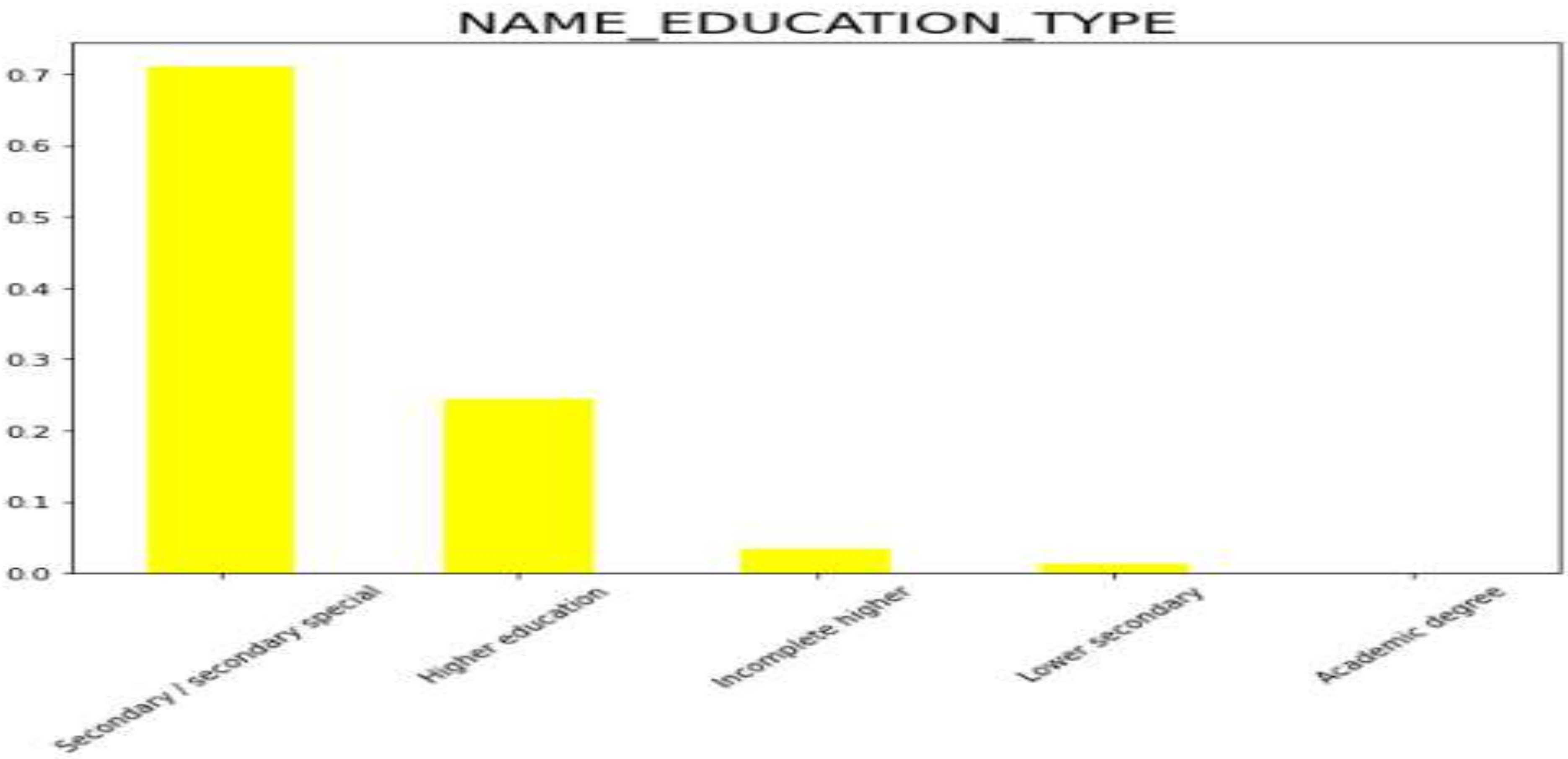
**NAME_INCOME_TYPE:**

Here, the graph indicates a decrease in the percentage of payment difficulties who are pensioner and an increase in the percentage of payment difficulties who are working when compared the percentage of both payment difficulties and non-payment difficulties

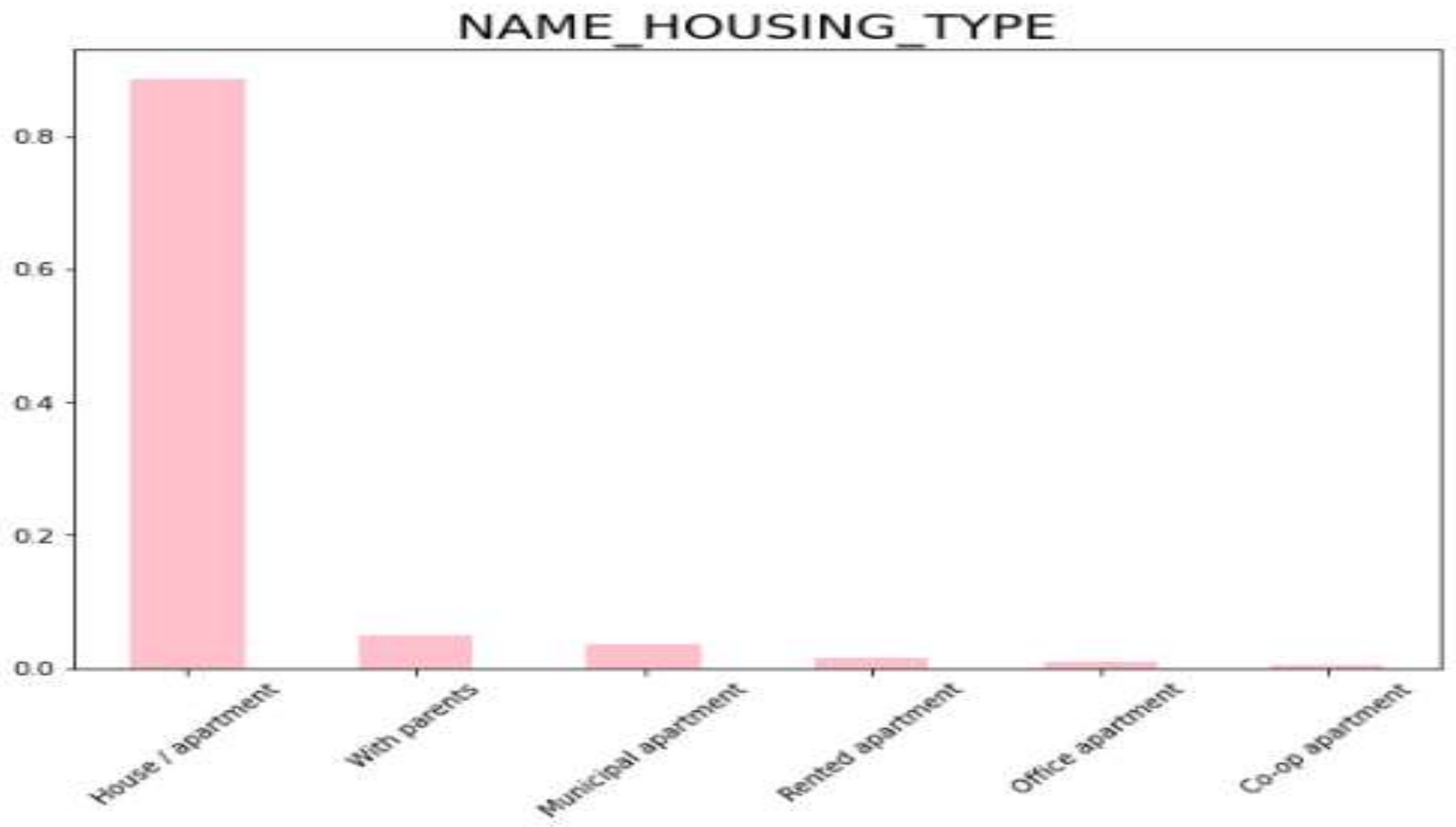


**NAME_EDUCATION_TYPE:**

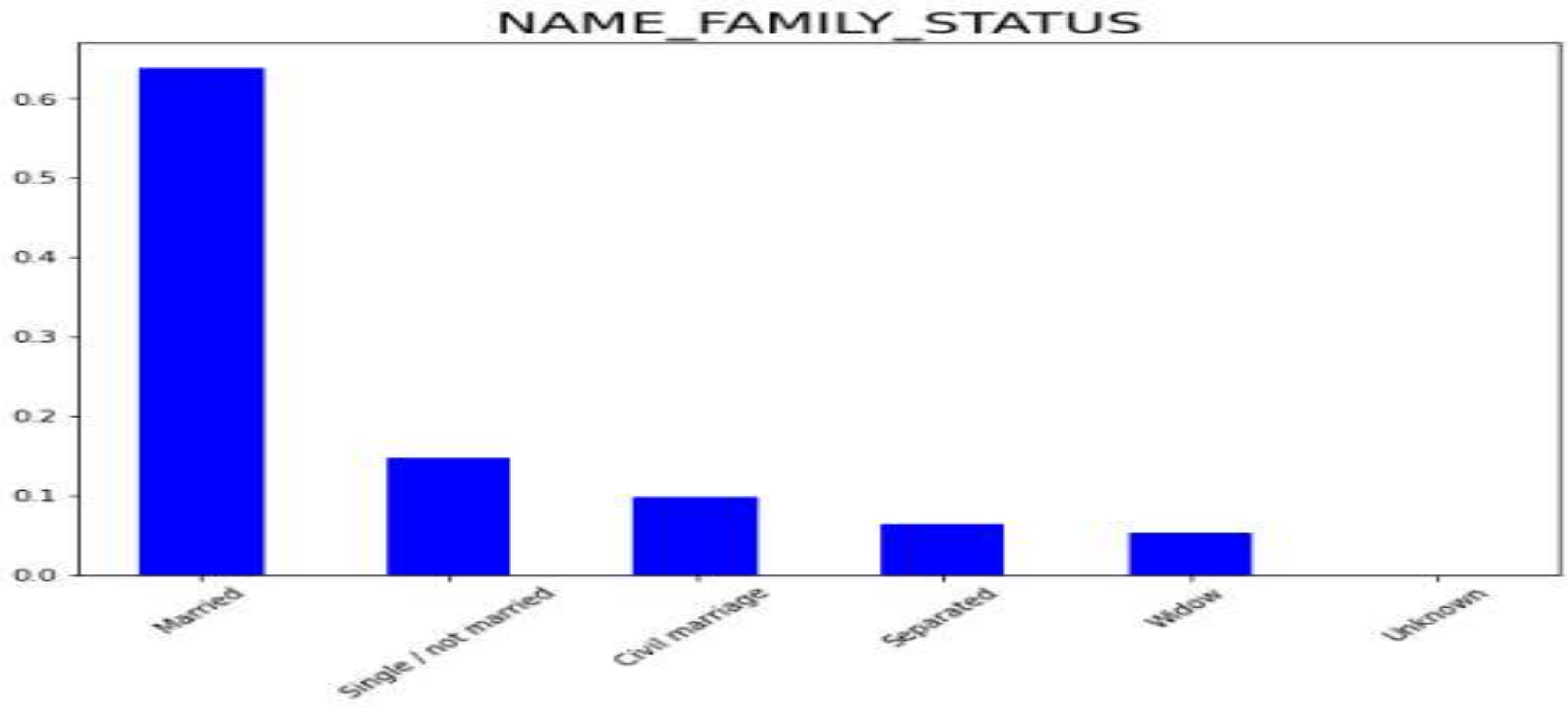
Here, the graph indicates an increase in percentage of loan payment difficulties whose educational qualifications are secondary/secondary special and a decrease in the percentage of loan payment difficulties who have completed higher education when compared with the percentages of loan payment difficulties and loan non-payment difficulties.



**On the basis of housing type:**
Here, the graph indicates the percentage of payment difficulties who live with their parents when compared to the percentages of payment difficulties and non-payment difficulties.

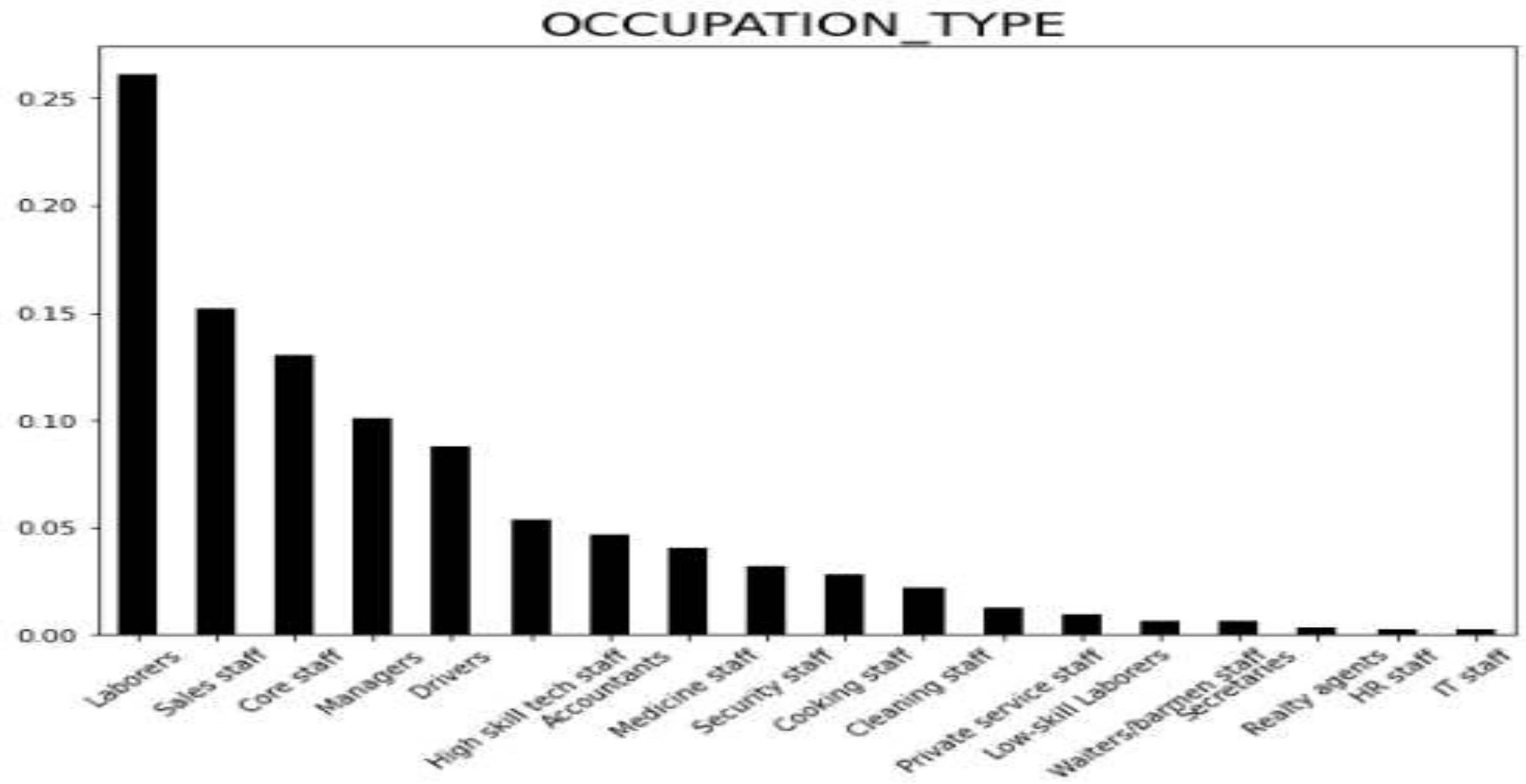


**NAME_FAMILY_STATUS:**
Here, the graph indicates a decrease in the percentage of married and widowed with loan payment difficulties and an increase in the percentage of single and civil married with loan payment difficulties when compared with the percentages of both loan payment difficulties and loan non-payment difficulties.

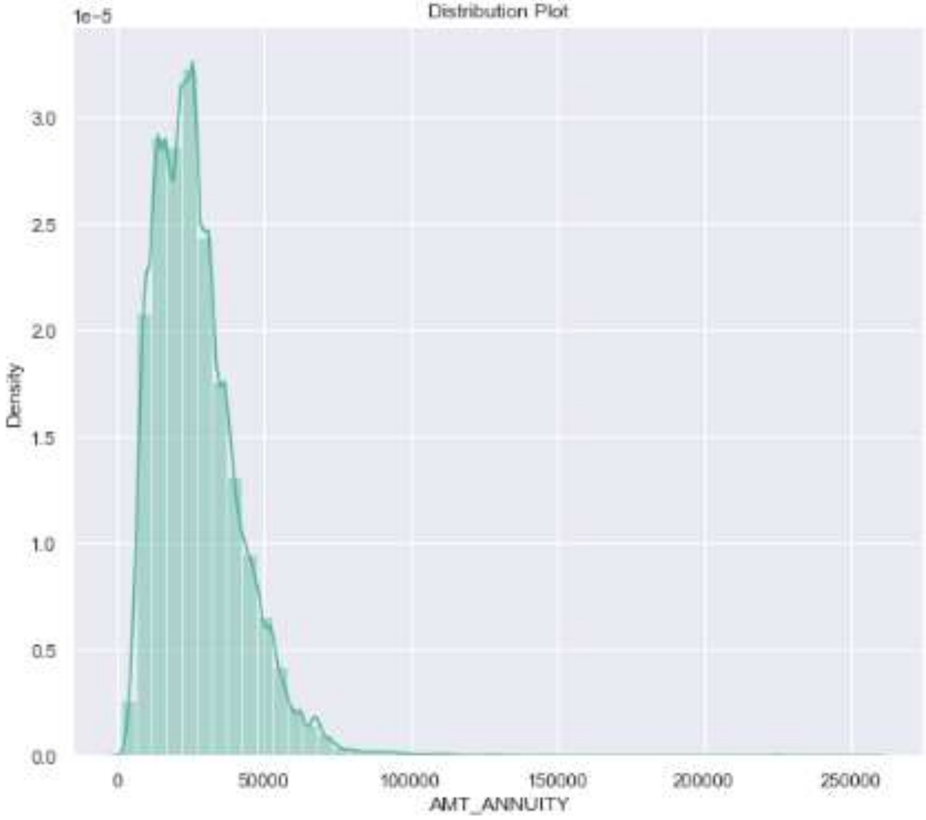


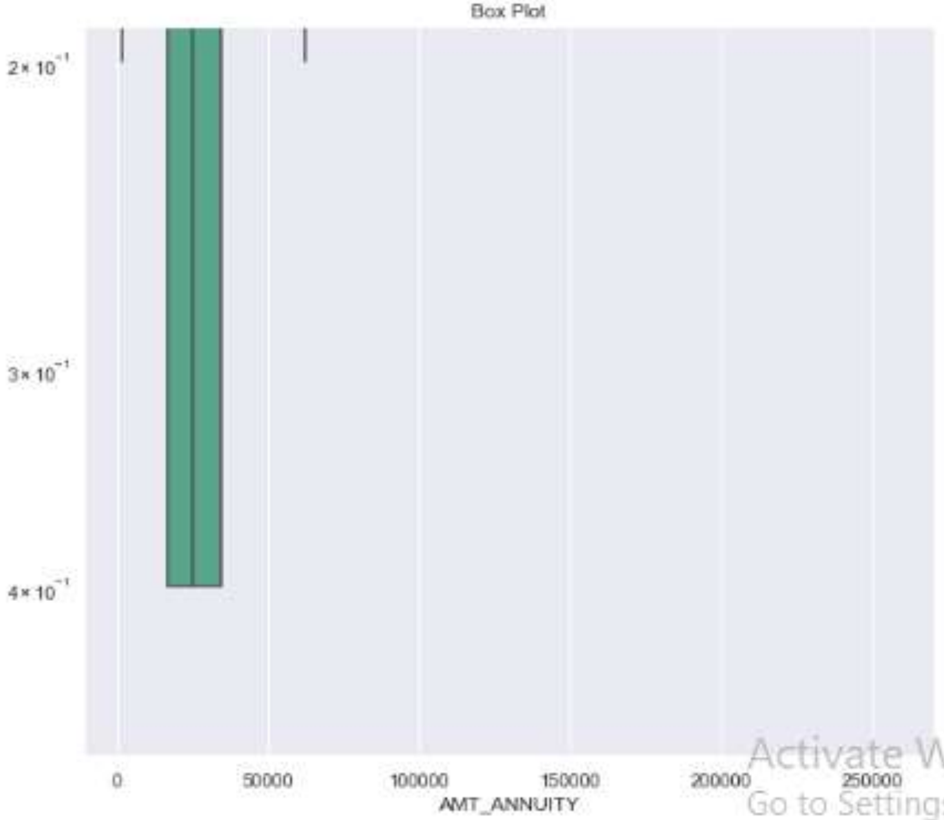
**Occupation_type:**

On the basis of occupation there is no major difficulties of payment

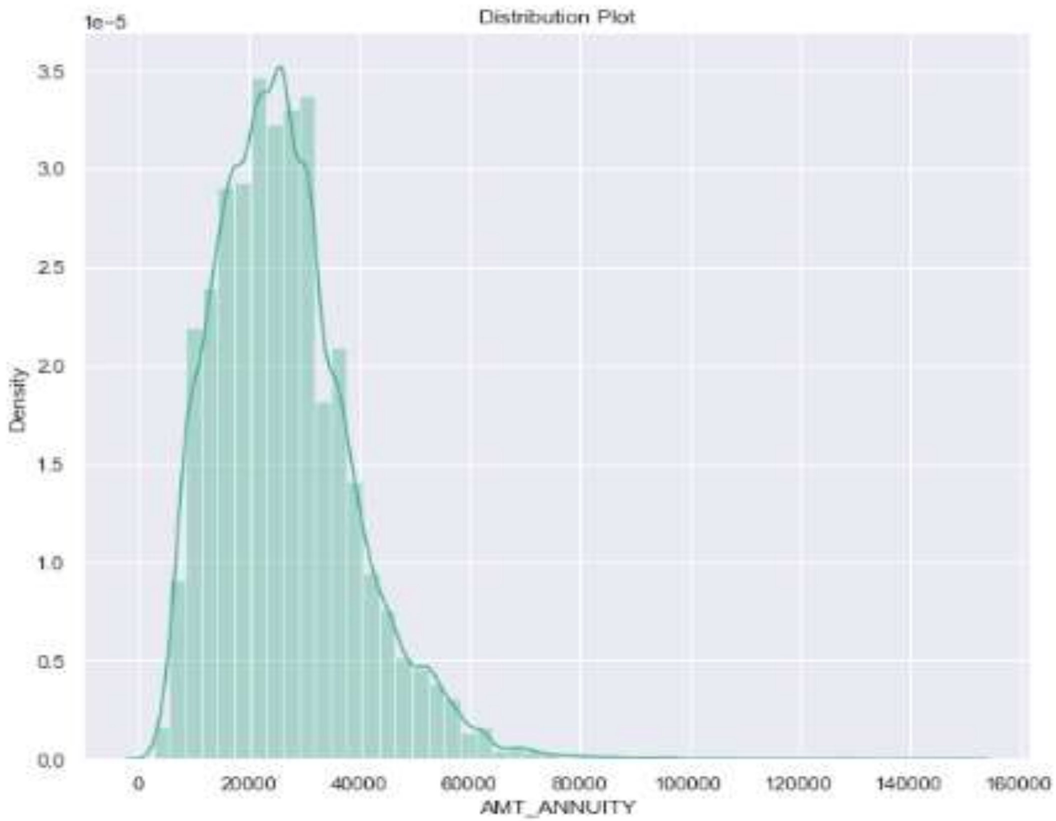


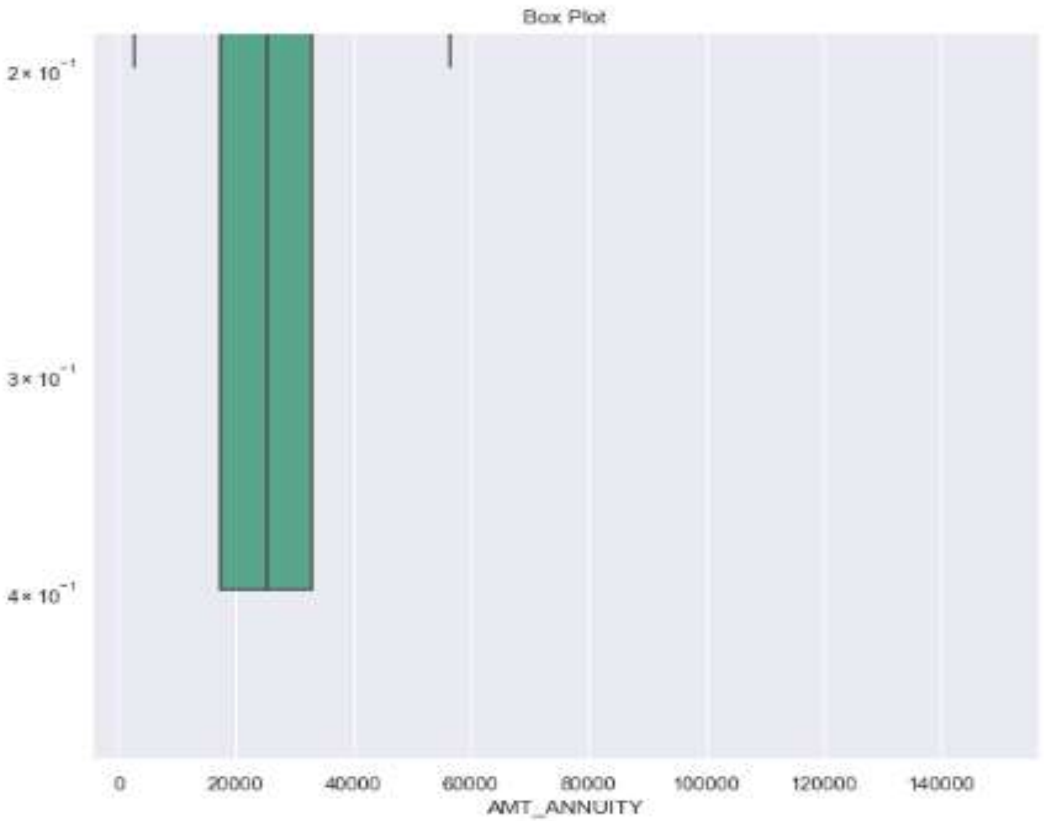
# Loan annuity for non-payment difficulties



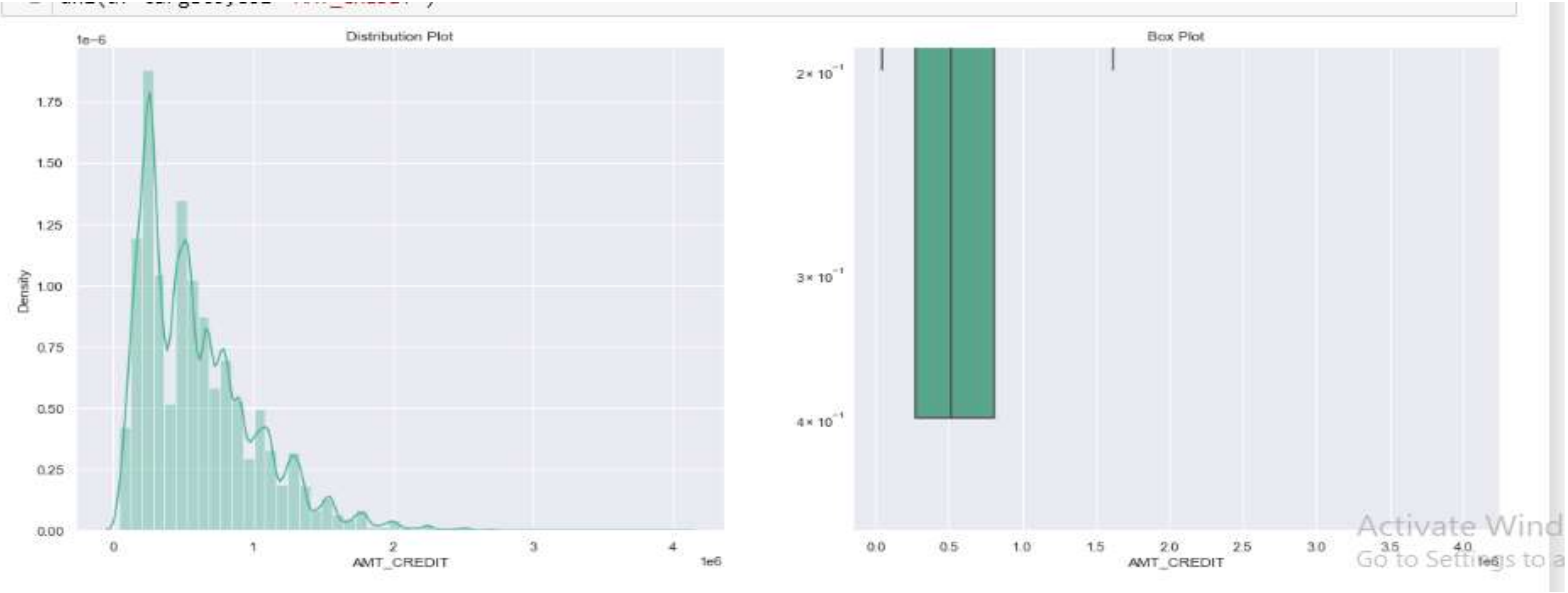
# Amt_annuity For loan-payment difficulties
We can observe some outliers and the first quartile is bigger than third quartile for anuuity amount which means most of the annuity clients are form first quartile.



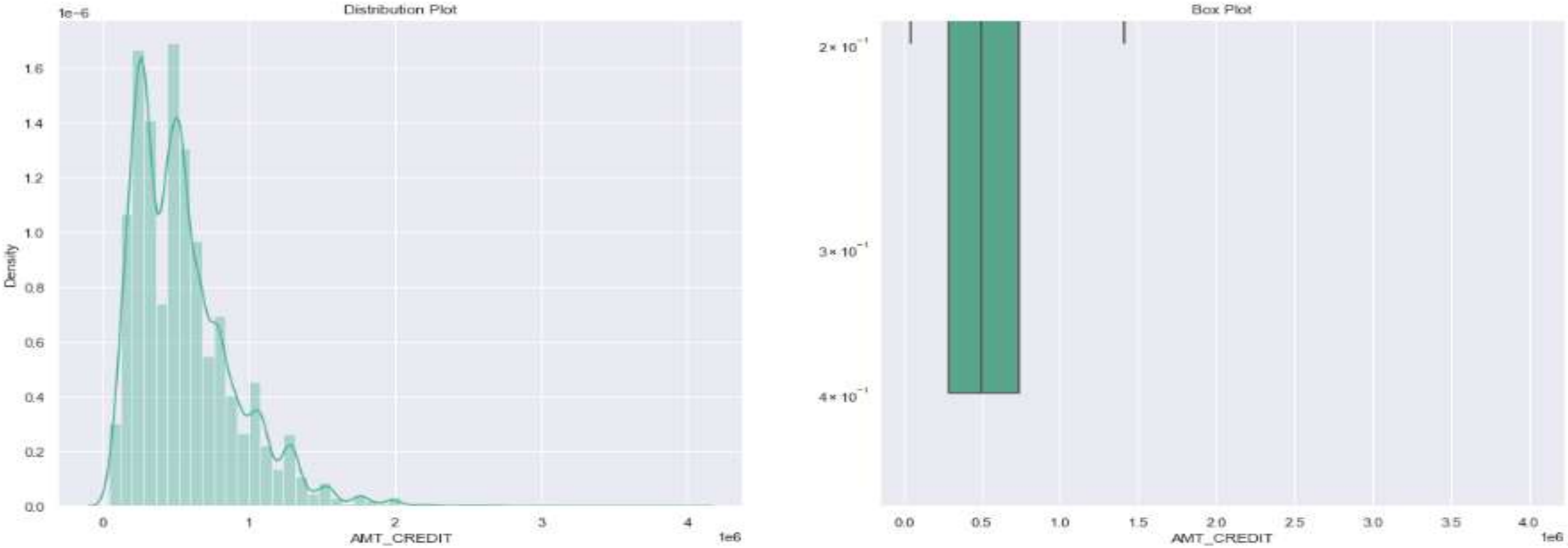

We can observe some outliers and the first quartile is bigger than third quartile for annuity amount which means most of the annuity clients are from first quartile.

# Credit amount for non-payment difficulties



## Credit amount for payment difficulties

We can observe some outliers and the first quartile is biggr 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




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.

**Heatmap for loan-payment difficulties**

We observe that there is a high correlation between credit amount and goods price. There to be some deviancies in the correlation of loan-payment difficulties and loan- non payment difficulties such as credit amount v/s income.

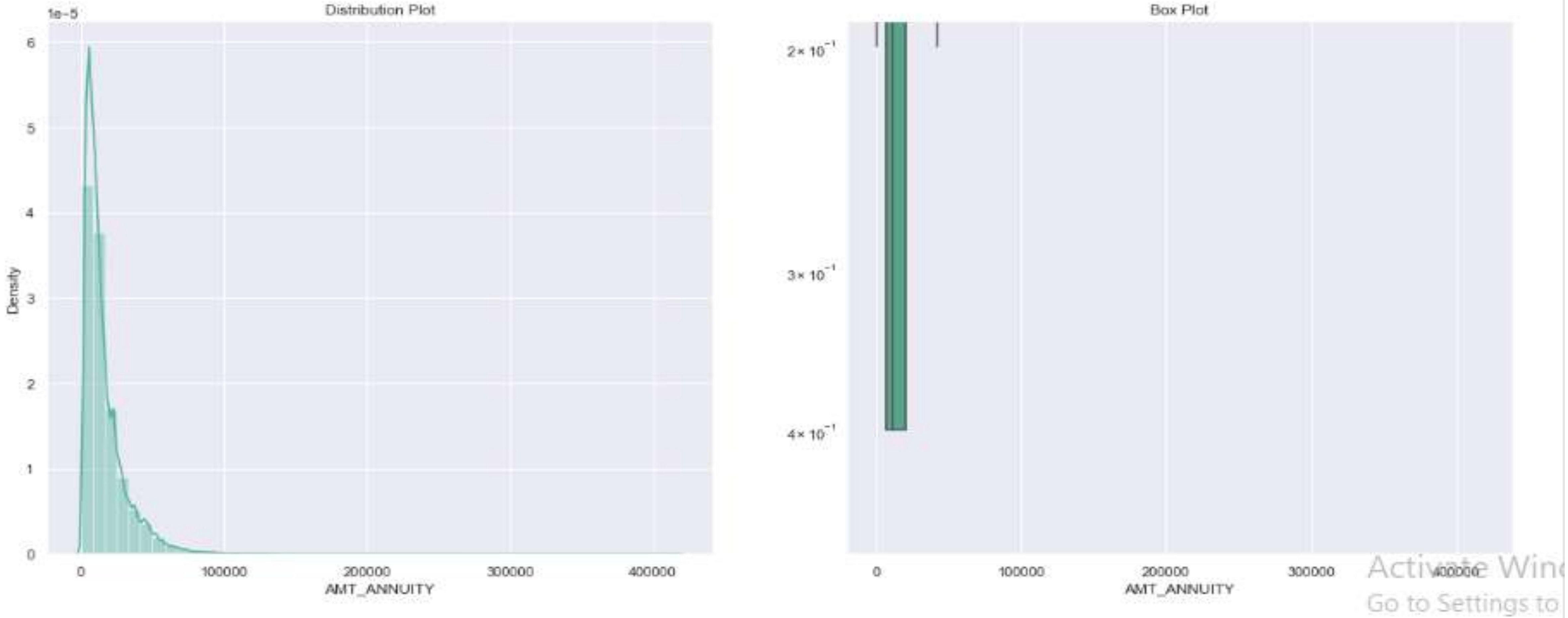
.]:

| | AMT_GOODS_PRICE | AMT_INCOME_TOTAL | AMT_ANNUITY | DAYS_EMPLOYED | DAYS_BIRTH | DAYS_REGISTRATION | DAYS_ID_PUBLISH |
|---|---|---|---|---|---|---|---|
| AMT_GOODS_PRICE | 1.000000 | 0.327484 | 0.752699 | 0.006640 | 0.135602 | 0.025682 | 0.056089 |
| AMT_INCOME_TOTAL | 0.327484 | 1.000000 | 0.398260 | -0.107740 | 0.002508 | -0.038519 | 0.003676 |
| AMT_ANNUITY | 0.752699 | 0.398260 | 1.000000 | -0.081208 | 0.014028 | -0.034279 | 0.016768 |
| DAYS_EMPLOYED | 0.006640 | -0.107740 | -0.081208 | 1.000000 | 0.582438 | 0.192468 | 0.229102 |
| DAYS_BIRTH | 0.135602 | 0.002508 | 0.014028 | 0.582438 | 1.000000 | 0.289135 | 0.252272 |
| DAYS_REGISTRATION | 0.025682 | -0.038519 | -0.034279 | 0.192468 | 0.289135 | 1.000000 | 0.096818 |
| DAYS_ID_PUBLISH | 0.056089 | 0.003676 | 0.016768 | 0.229102 | 0.252272 | 0.096818 | 1.000000 |
| AMT_CREDIT | 0.983103 | 0.325386 | 0.752195 | 0.001930 | 0.135072 | 0.025854 | 0.052329 |

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.

**Univariate analysis for numerical columns**

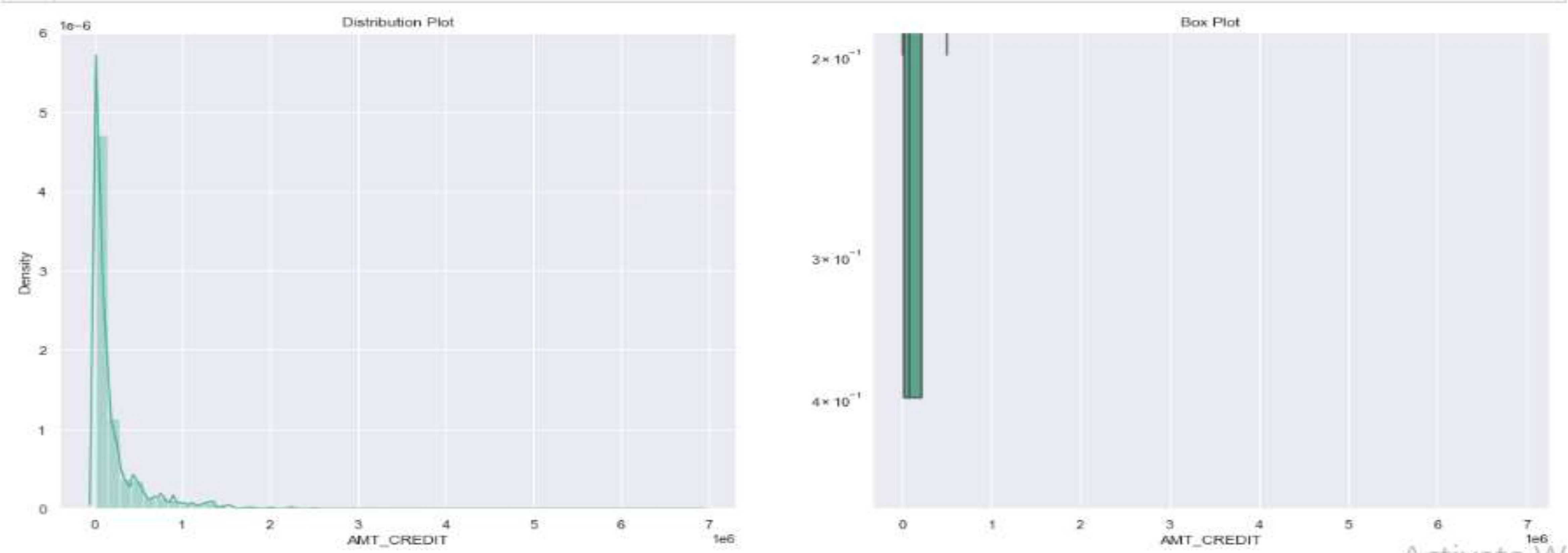
We observe that there are some outliers and the curve is not normal or a bell curve




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

## Univariate AMT_CREDIT

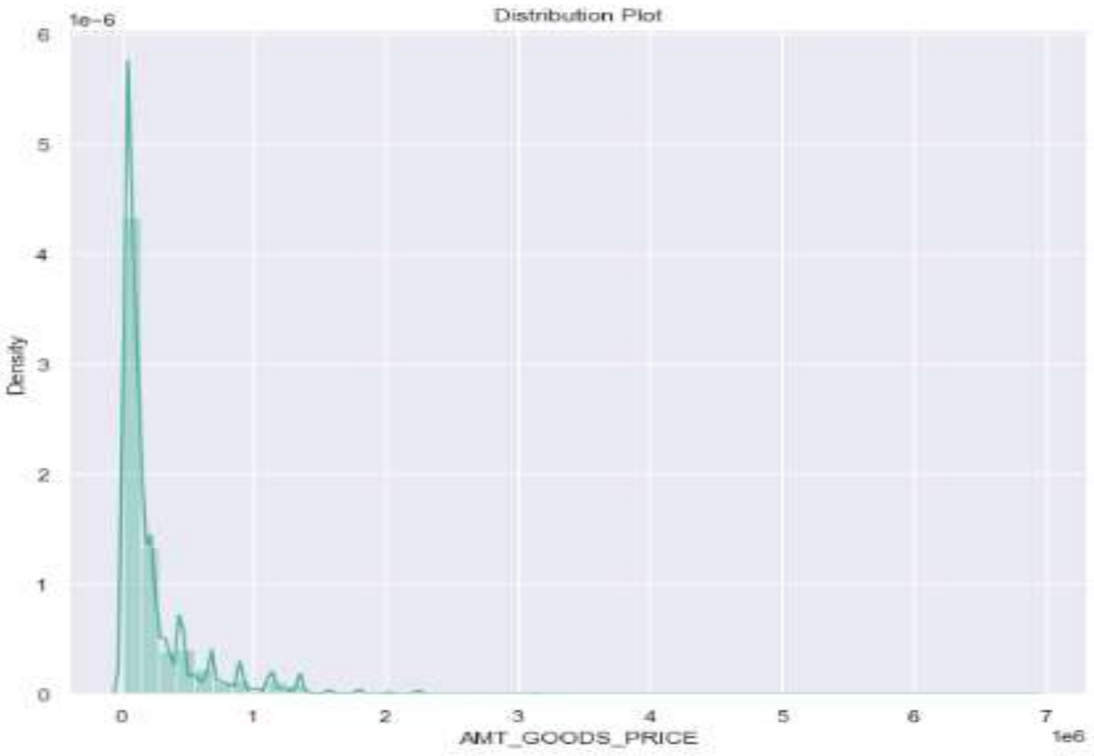
**We observe that there are some outliers and the curve is not normal or a bell curve.**

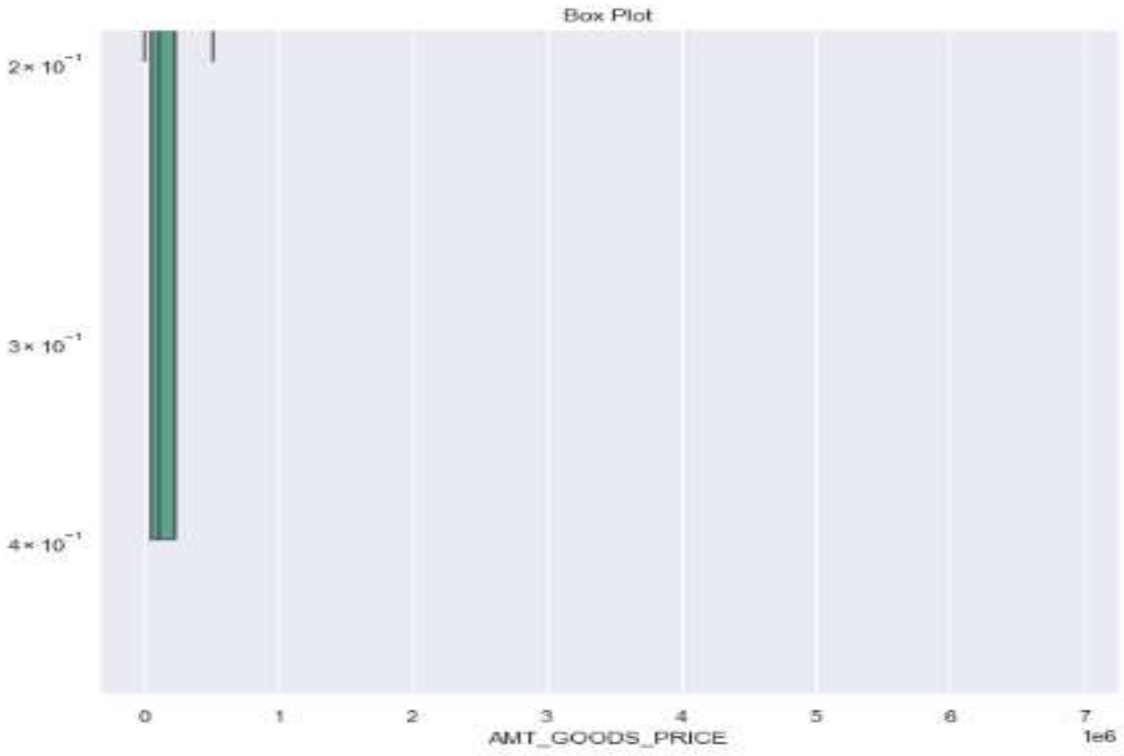



We observe that there are some outliers and the curve is not normal or a bell curve.

**AMT_GOODS_PRICE**

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

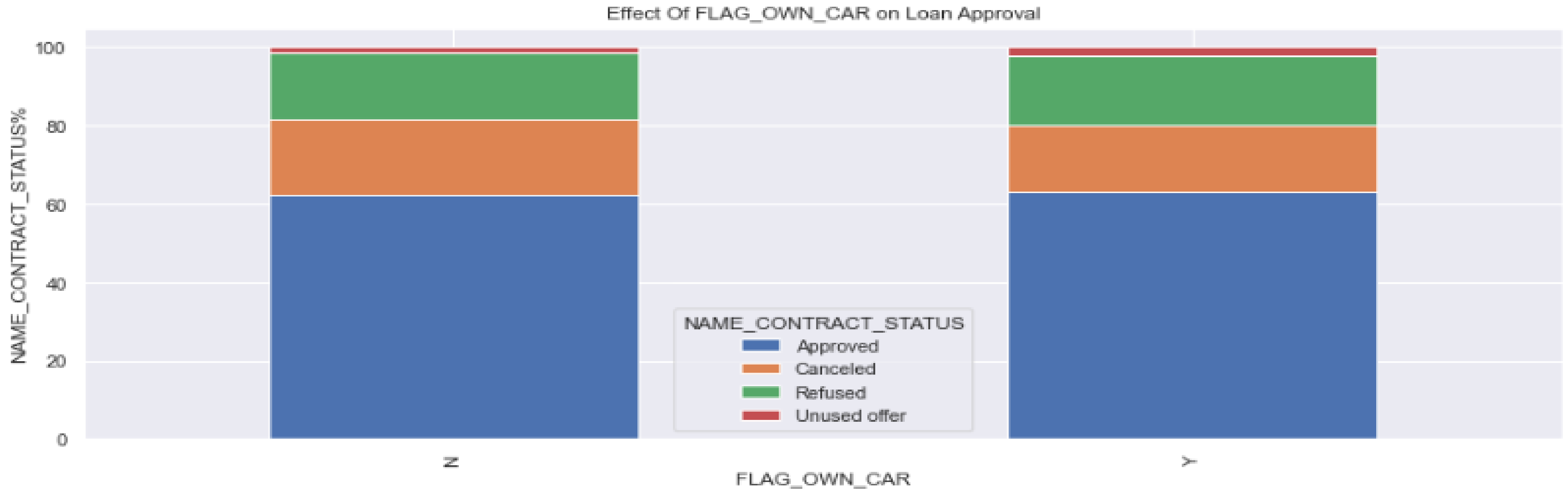



We observe that there are some outliers and the curve .is not normal or a bell curve.

# EFFECT OF FLAG_OWN_CAR on Loan Approval

We see that car ownership doesn't have any effect on application approval or rejection. But we saw earlier that the people who has a car has lesser chances of default.




We see taht car ownership doesn't have any effect on application approval or rejection. But we saw earlier that the people who has a car has lesser chances of default. The bank can add more weightage to car ownership while approving a loan amount.

# EFFECT OF CODE_GENDER ON LOAN APPROVALE

We see that code gender doesn’t have any effect on application or rejection. But we saw earlier that female have lesser chances of default compared to males.


We see that code gender doesn't have any effect on application approval or rejection. But we saw earlier that female have lesser chances of default compared to males. The bank can add more weightage to female while approving a loan amount.

***Target variable (0- Non Defaulter 1 - Defaulter )***

We can see that the people who were approved for a loan earlier, defaulted less often where as people who were refused a loan earlier have higher chances of defaulting.

# THANK YOU☺