

EDA – Credit case study

By – Sanket Badadal

Problem Statement

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 specializes in lending various types of loans to urban customers. You must use EDA to analyze 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 must decide for loan approval based on the applicant's profile. 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.

The data given below contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:

- **The client with payment difficulties:** he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample,
- **All other cases:** All other cases when the payment is paid on time.

When a client applies for a loan, there are four types of decisions that could be taken by the client/company):

Approved: The Company has approved loan Application

Cancelled: The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.

Refused: The company had rejected the loan (because the client does not meet their requirements etc.).

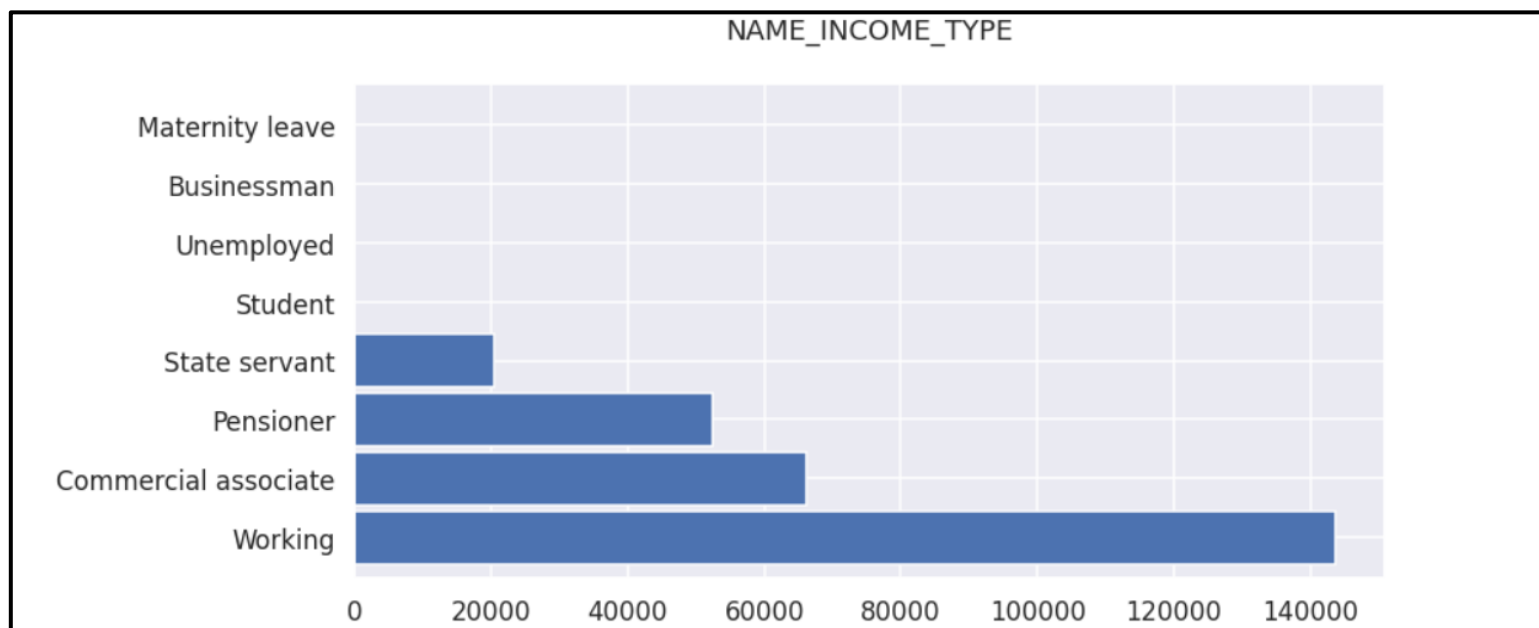
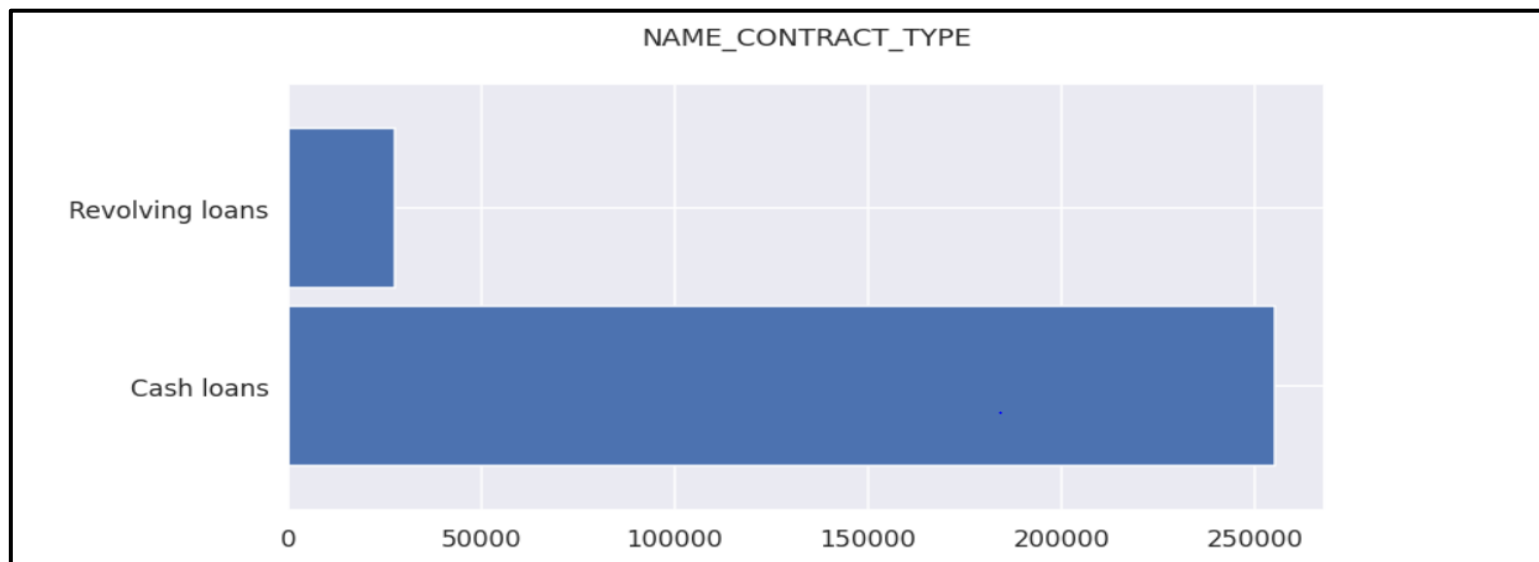
Unused offer: Loan has been cancelled by the client but on different stages of the process.

**Categorical univariate analysis for target = 0
(Customer with no payment difficulties)**

Distribution of NAME_CONTRACT_TYPE PE and NAME_INCOME_TYPE

Points to be concluded from
graphs to the right

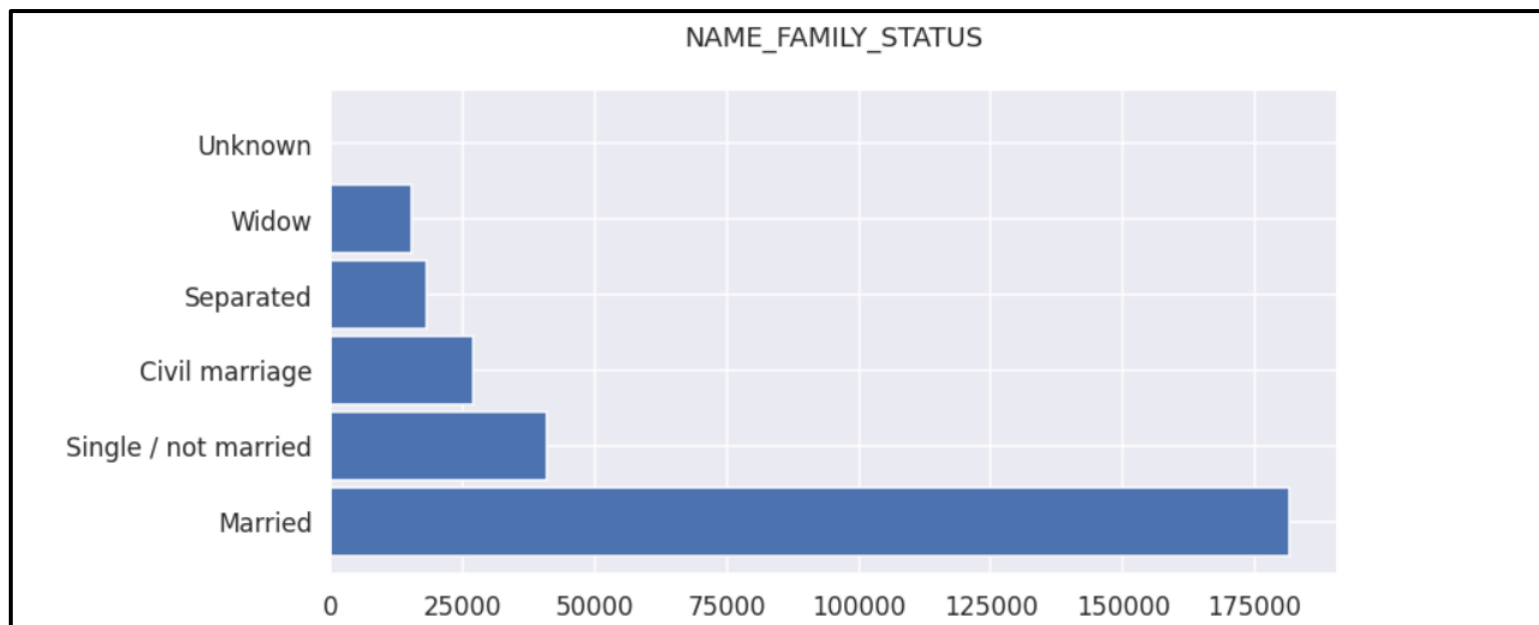
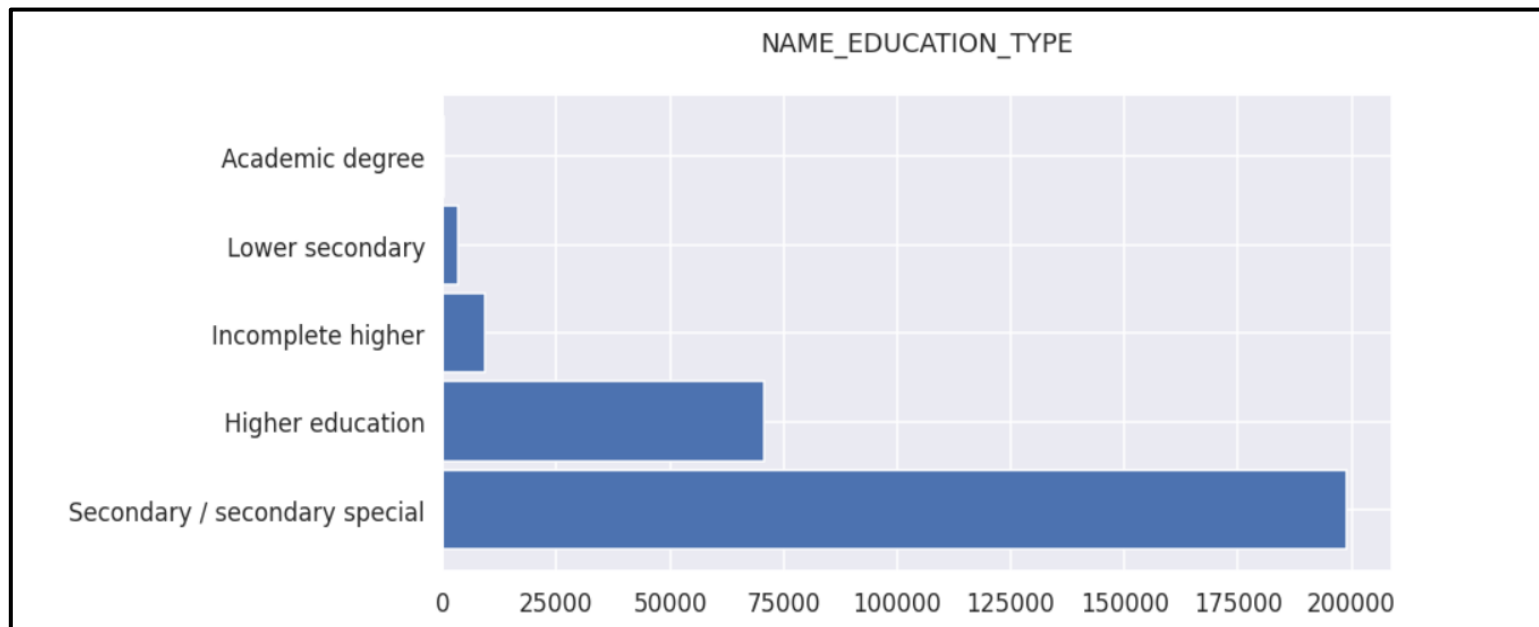
- For contract type 'Cash loans' is having higher number of credits than 'Revolving loans' type
- For income type 'working', 'commercial associate', and 'Pensioner' the number of credits are higher than others.



Distribution of NAME_EDUCATION_TYPE and NAME_FAMILY_STATUS

Points to be concluded from
graphs to the right

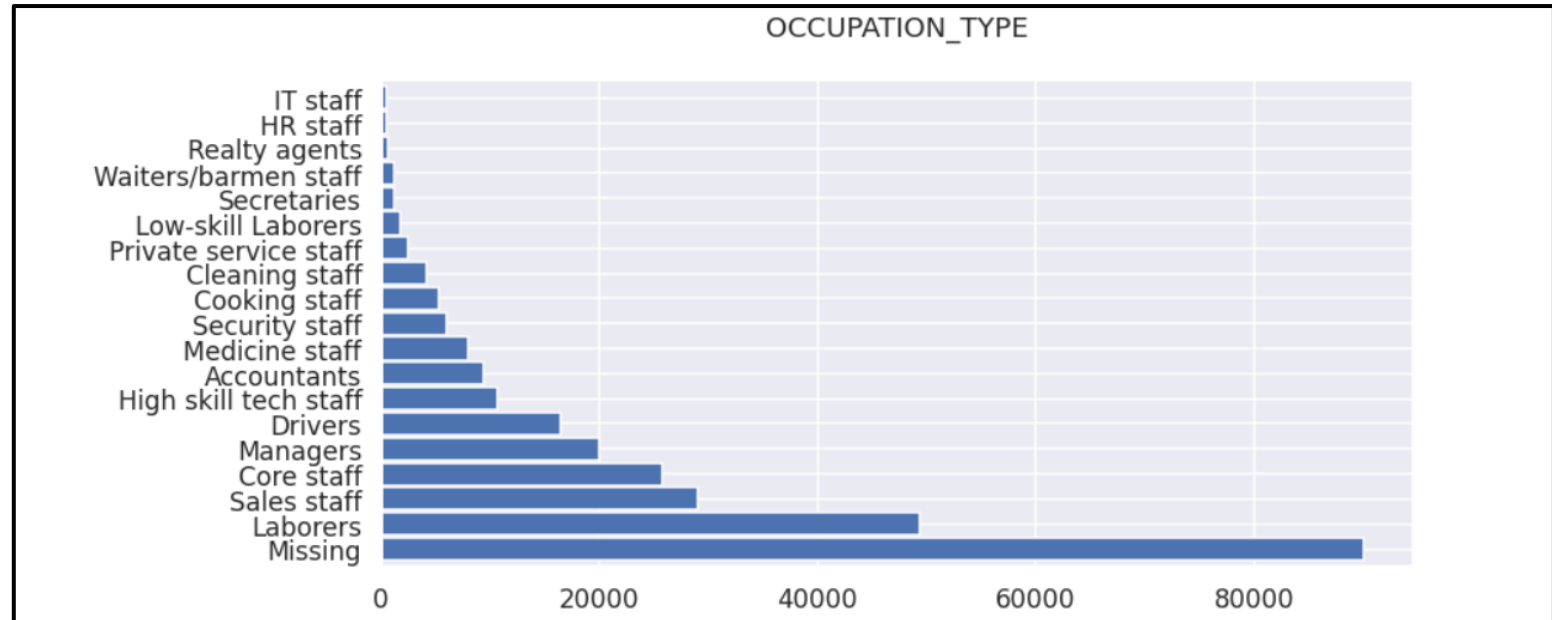
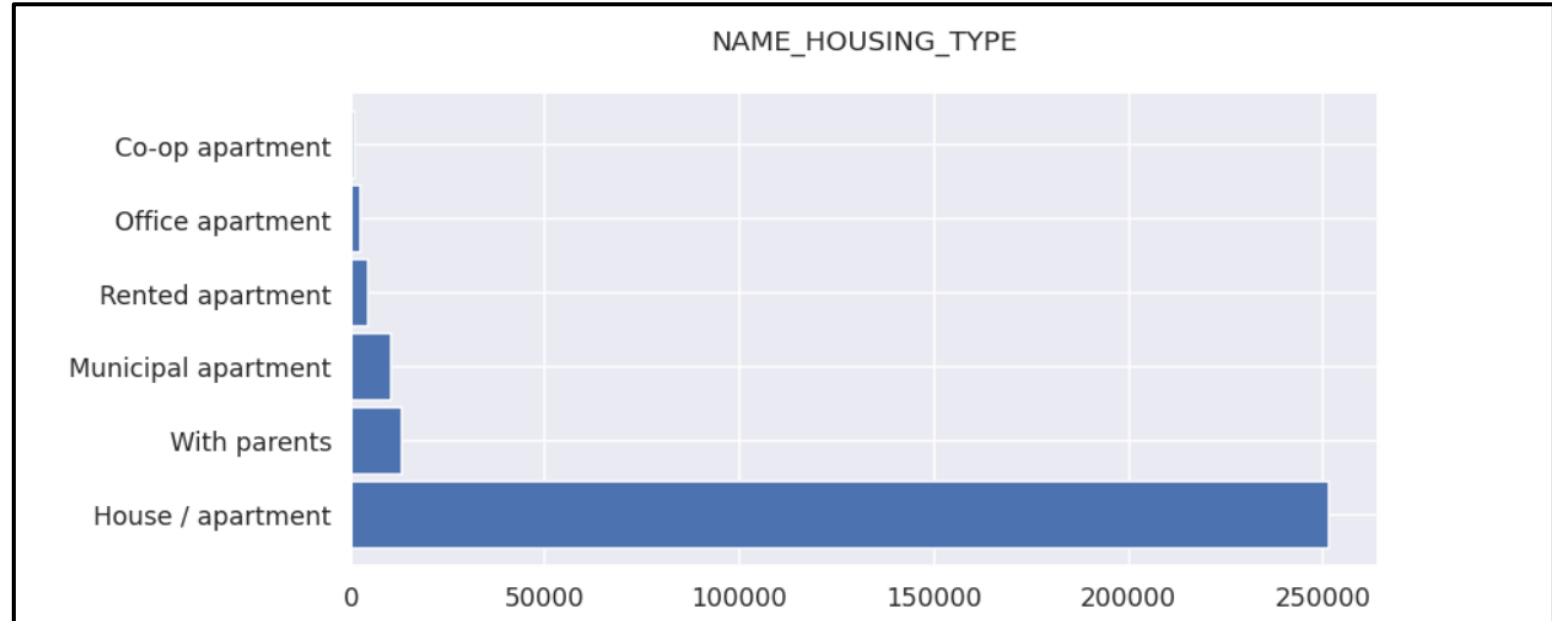
- **Secondary/Secondary special and Higher education are having a greater number of credit**
- **Married people are having a greater number of credit and widow is having lesser number of credit**



Distribution of NAME_HOUSING_TYPE and OCCUPATION_TYPE

Points to be concluded from
graphs to the right

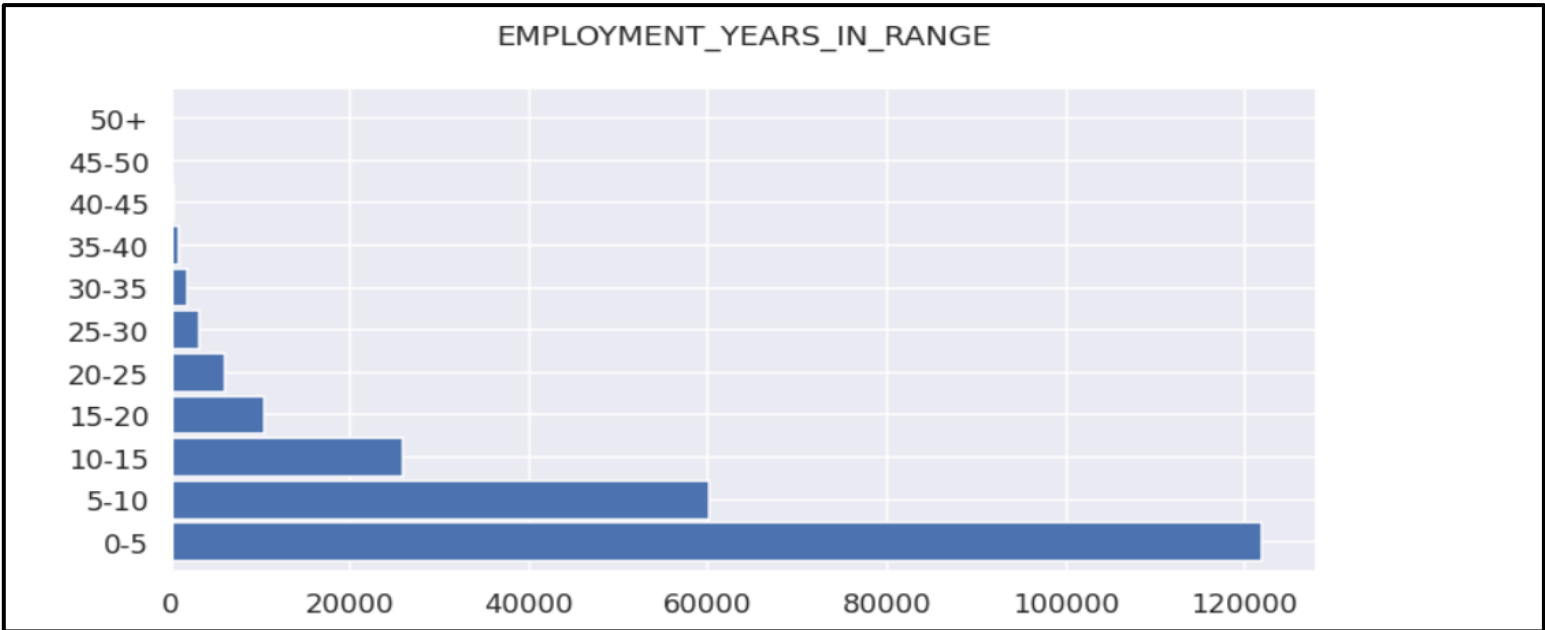
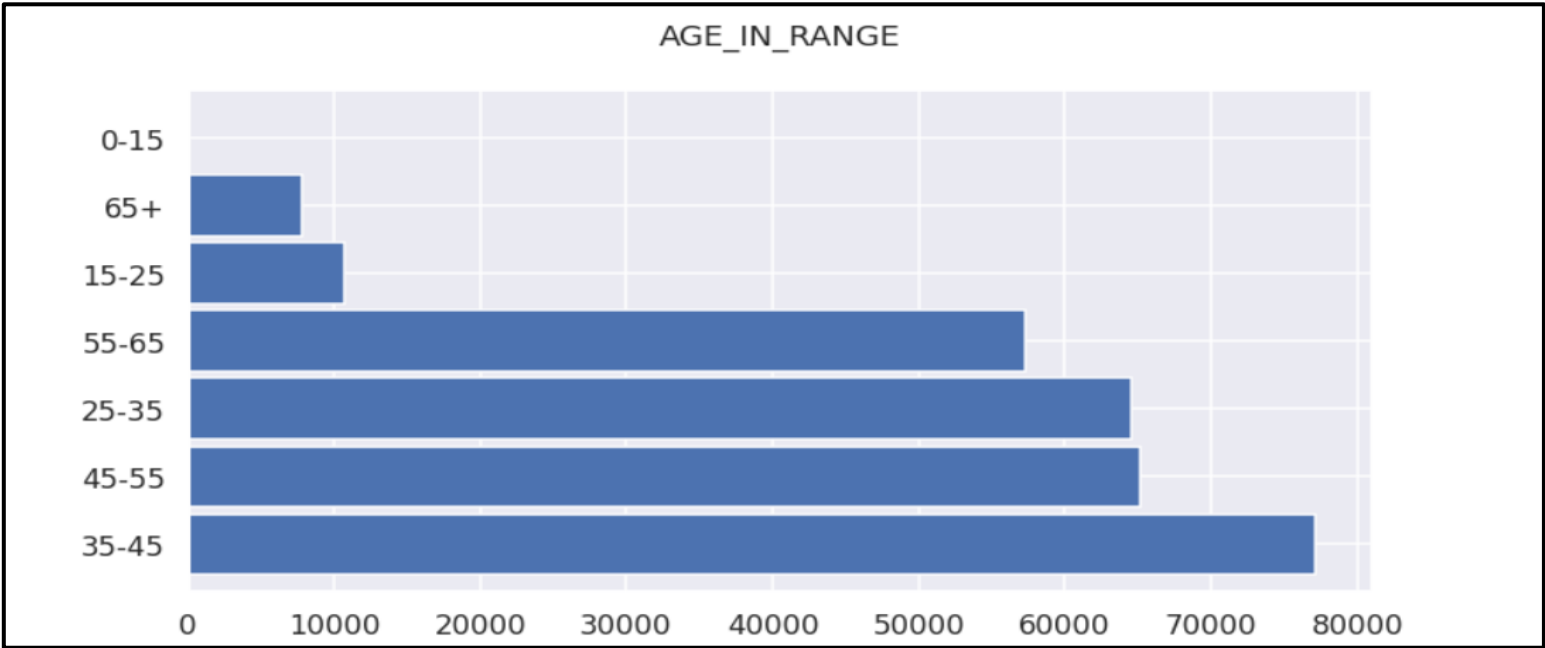
- People with own House/apartment are having a greater number of credit
- Laborers and sales staff are having a greater number of credit



Distribution of AGE_IN_YEARS and EMPLOYMENT_YEARS

Points to be concluded from
graphs to the right

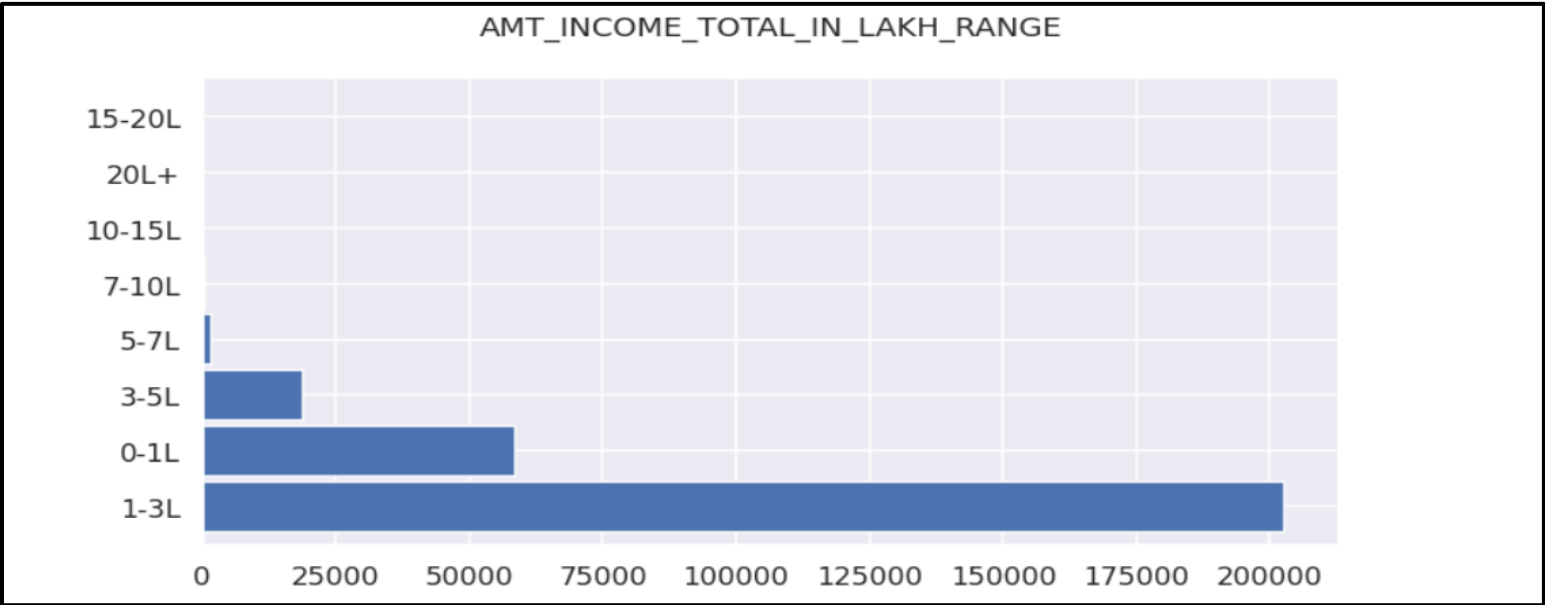
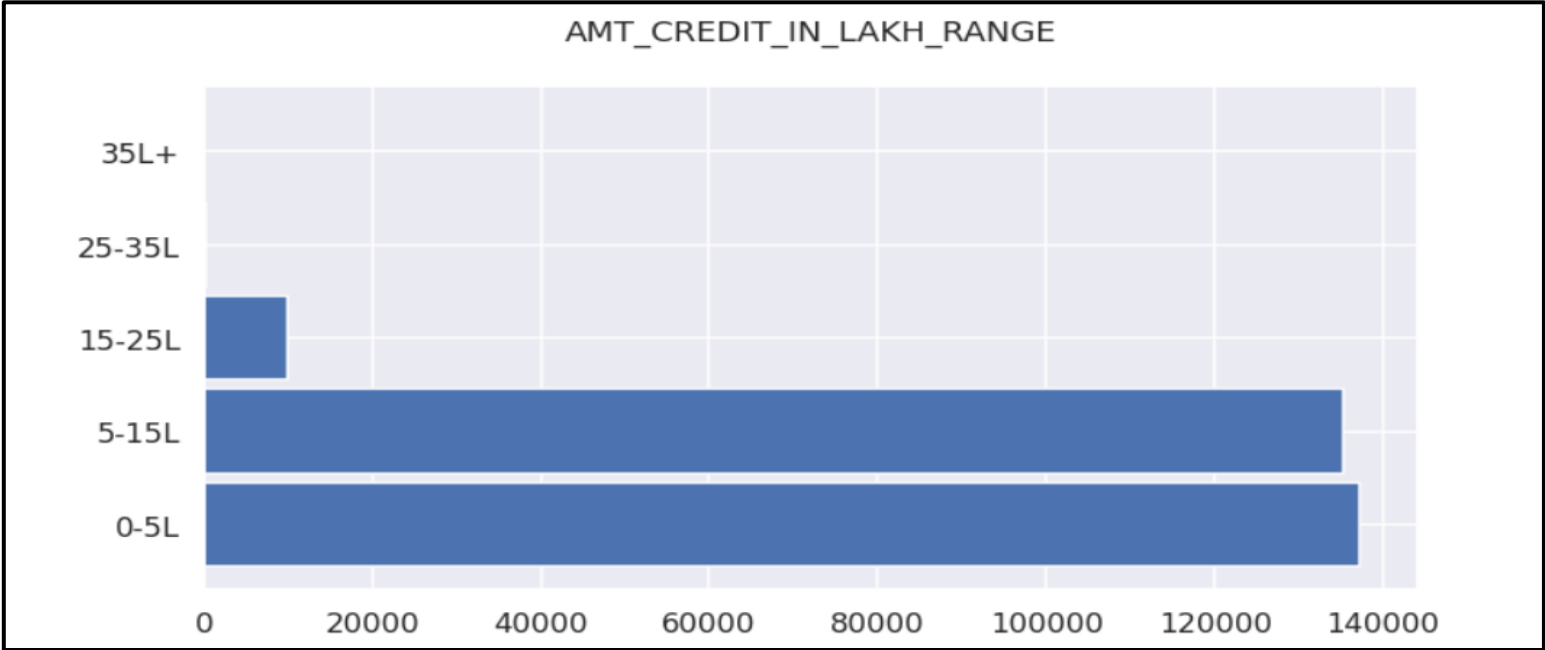
- People with age range 35-45 are having a greater number of credit
- People with employment years 0-5 are having a greater number of credit



Distribution of AMT_CREDIT and AMT_INCOME

Points to be concluded from
graphs to the right

- People with credit amount from 0 to 15 L are having a greater number of credit
- People with income 1-3 L are having a greater number of credit

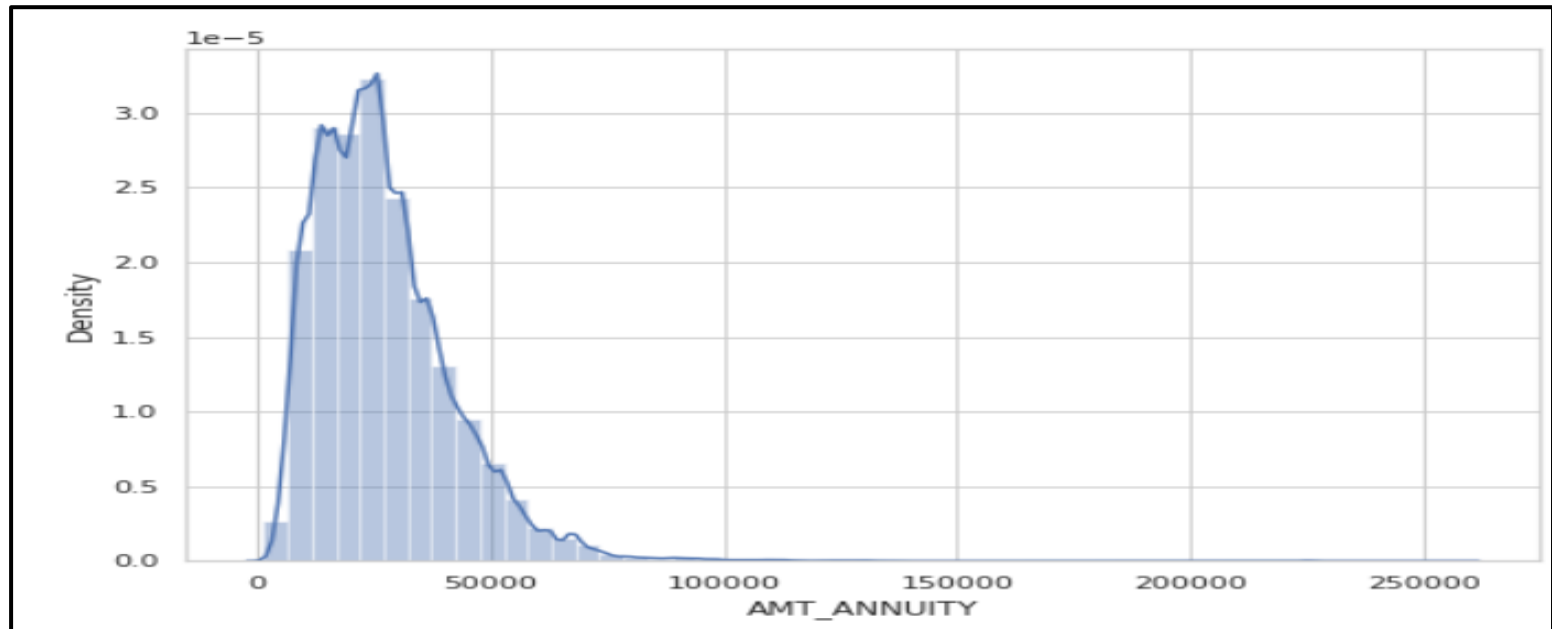
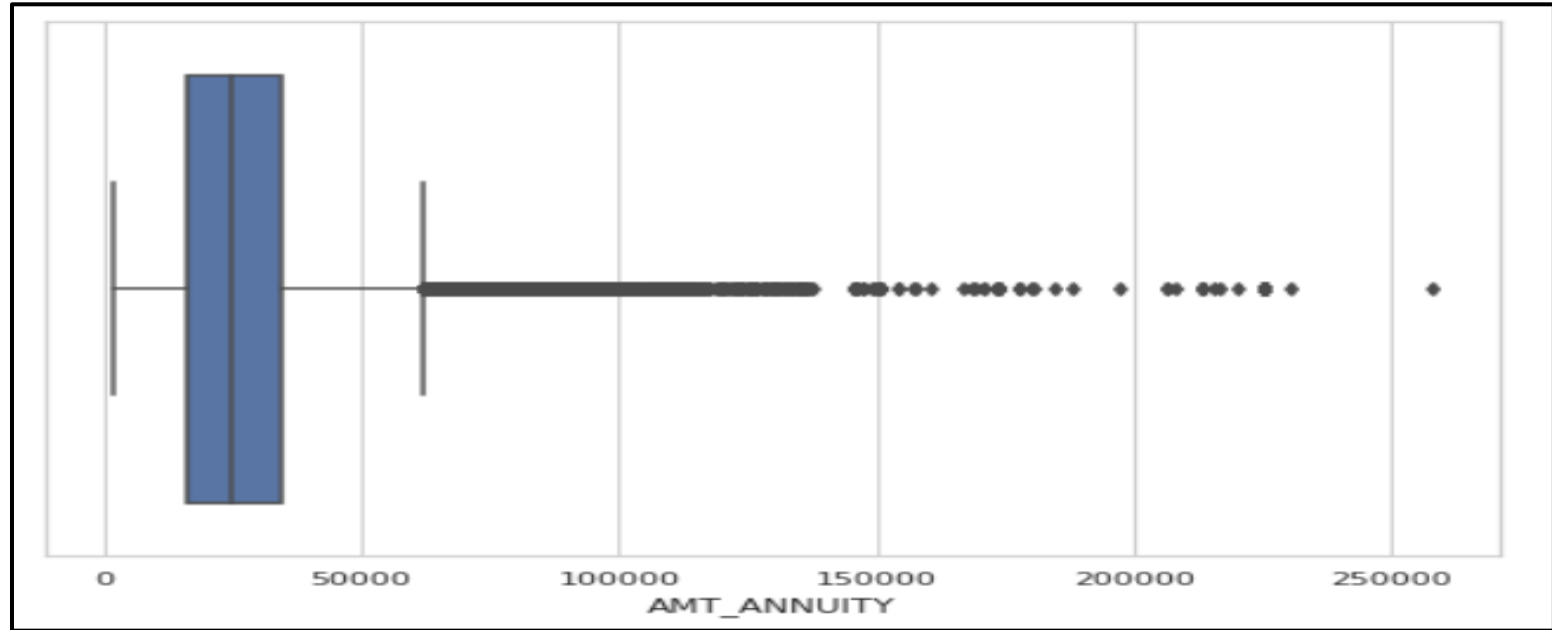


**Numerical univariate analysis for target = 0
(Customer with no payment difficulties)**

Distribution of Annuity Amount

Points to be concluded from graphs to the right

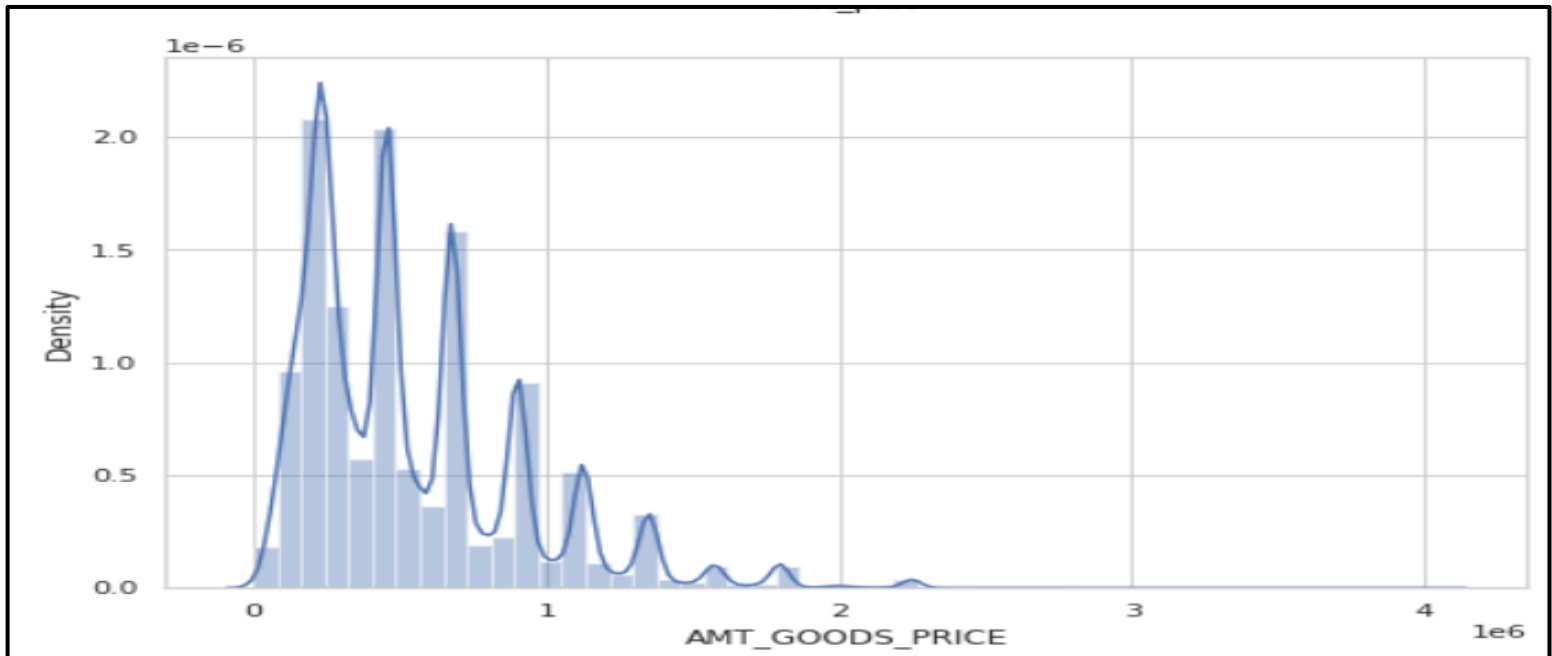
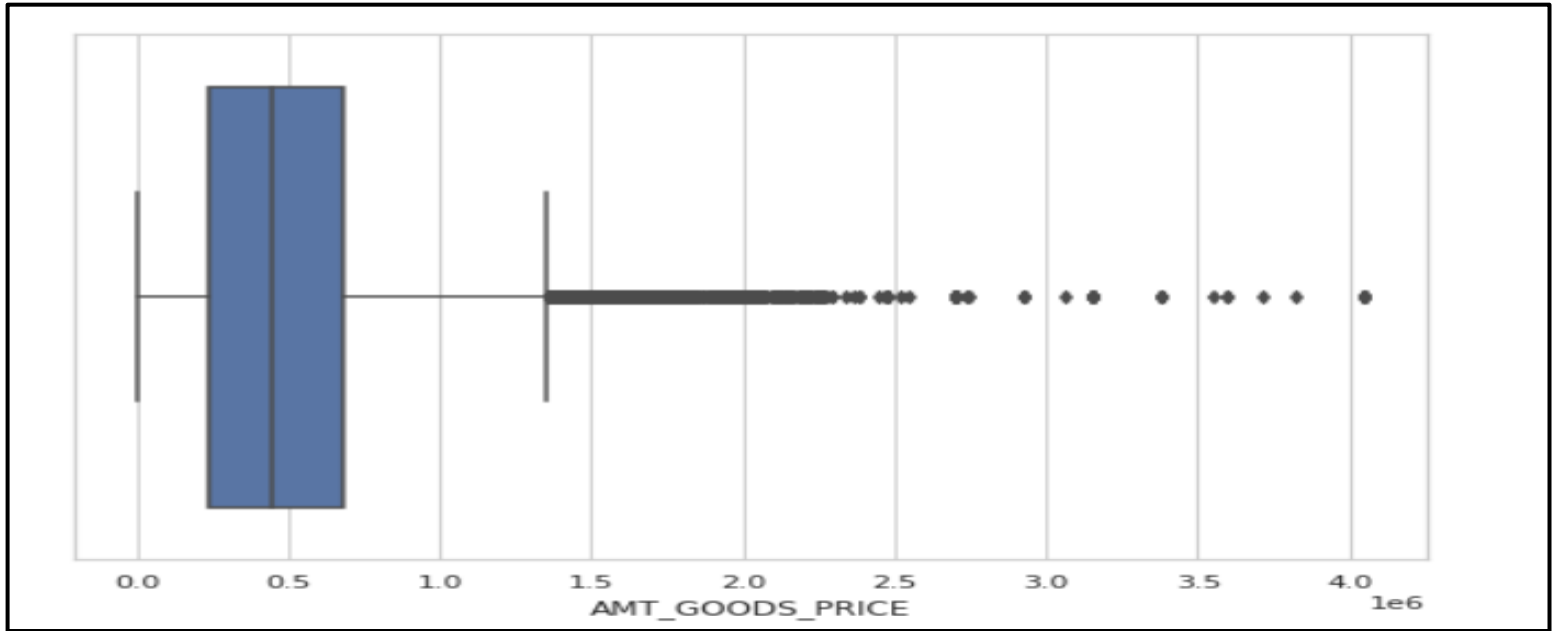
- Some Outliers are present in the annuity amount
- Maximum values are lies from 0 to 50,000 annuity amount and 50th percentile is close to 25,000 annuity amount



Distribution of Goods price Amount

Points to be concluded from graphs to the right

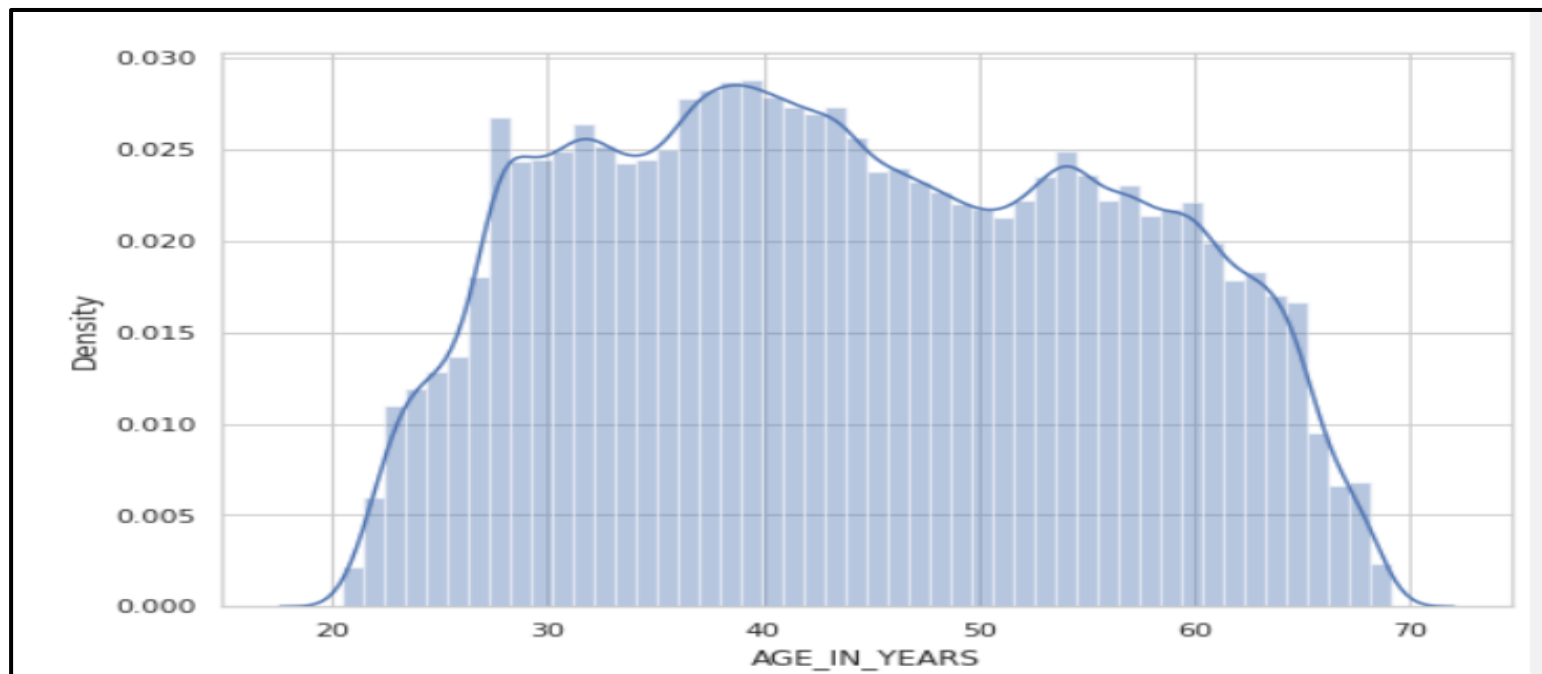
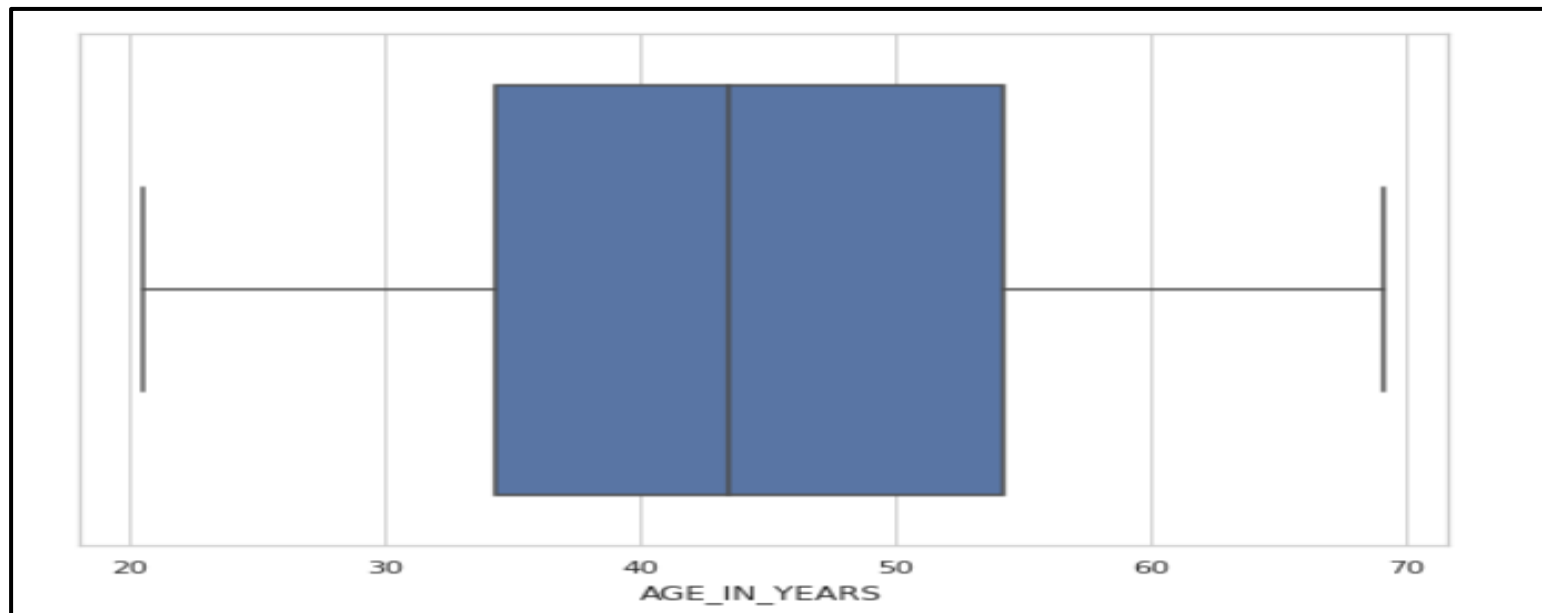
- Some Outliers are present in the goods price amount
- Maximum values lies from 0 to 10L approx. and first quartile is lesser than third quartile



Distribution of Age in years

Points to be concluded from graphs to the right

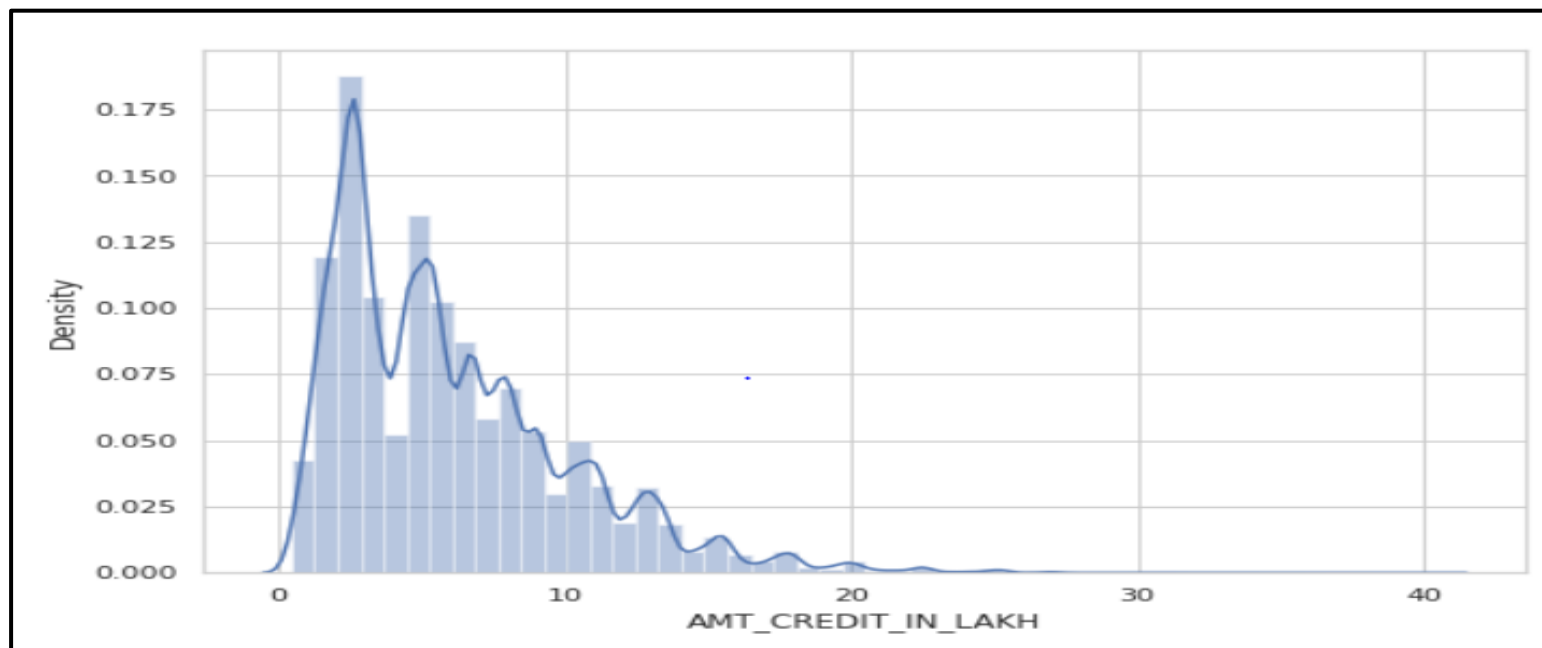
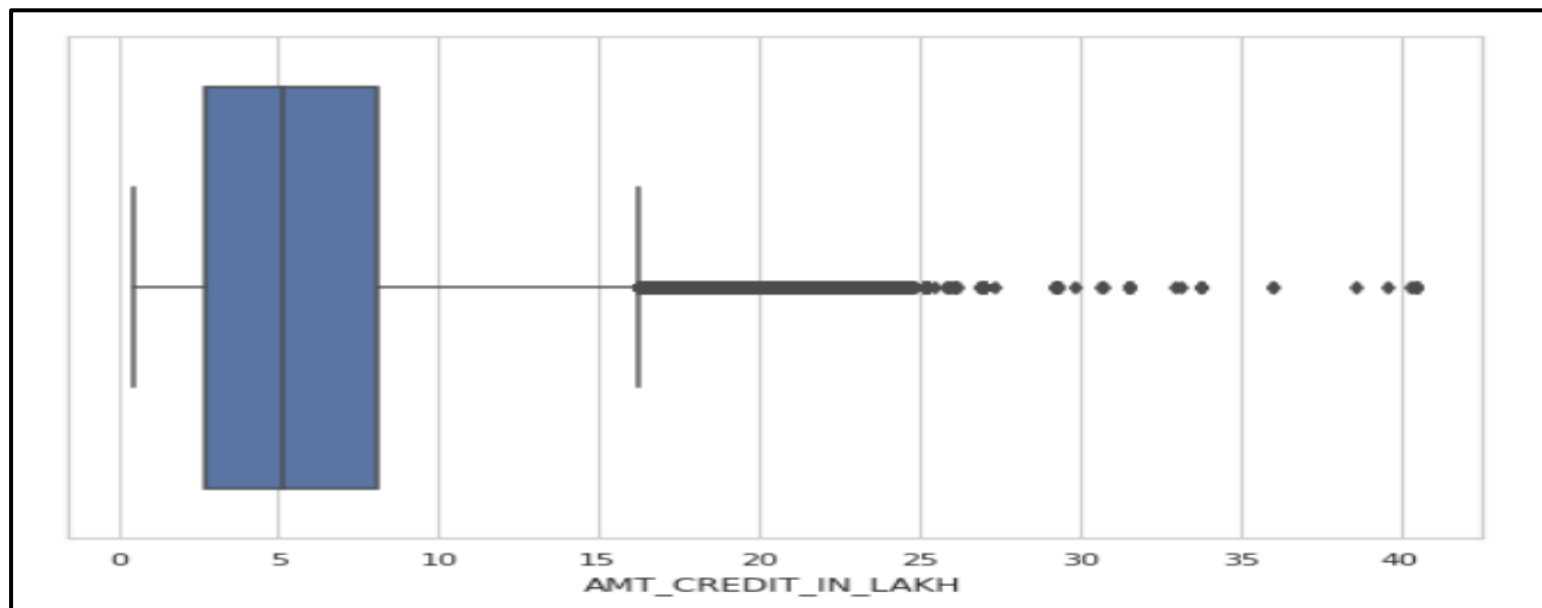
- Maximum values lies from 35 to 55 age in years
- Peoples which are 40 age are having maximum number of credits



Distribution of Credit Amount

Points to be concluded from graphs to the right

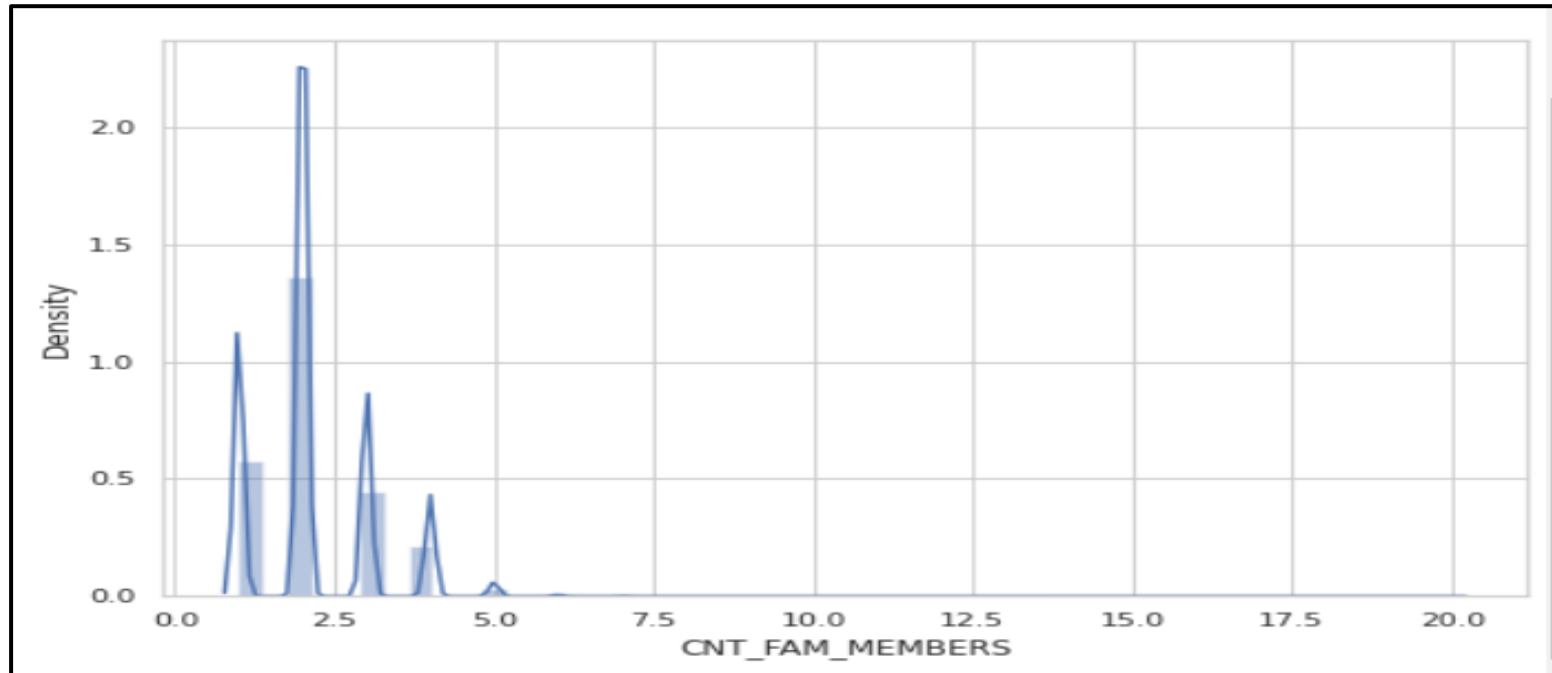
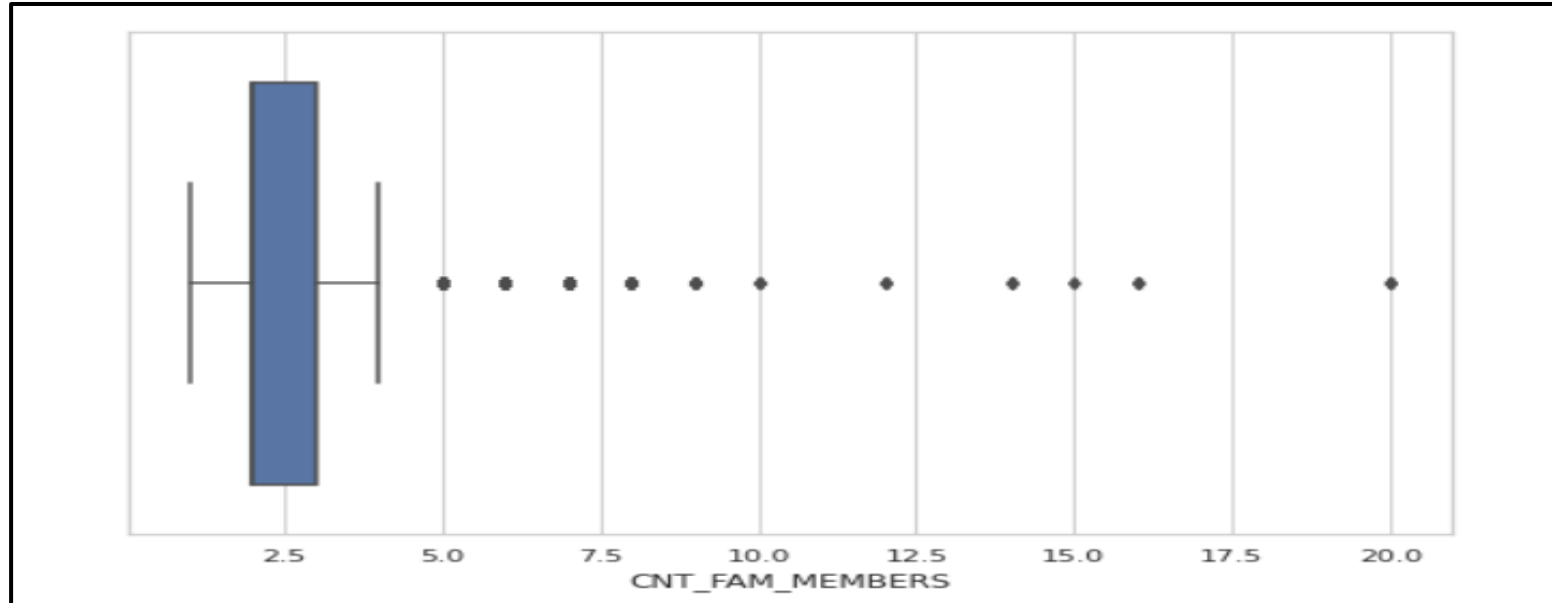
- Maximum values lies from 0 to 10 L credit amount
- Some outliers are present in credit amount



Distribution of Customer Family Members

Points to be concluded from graphs to the right

- Maximum values lies from 1 to 4
- Some outliers are present in Family members
- Most of the customers who are having higher credit having 2 Family members

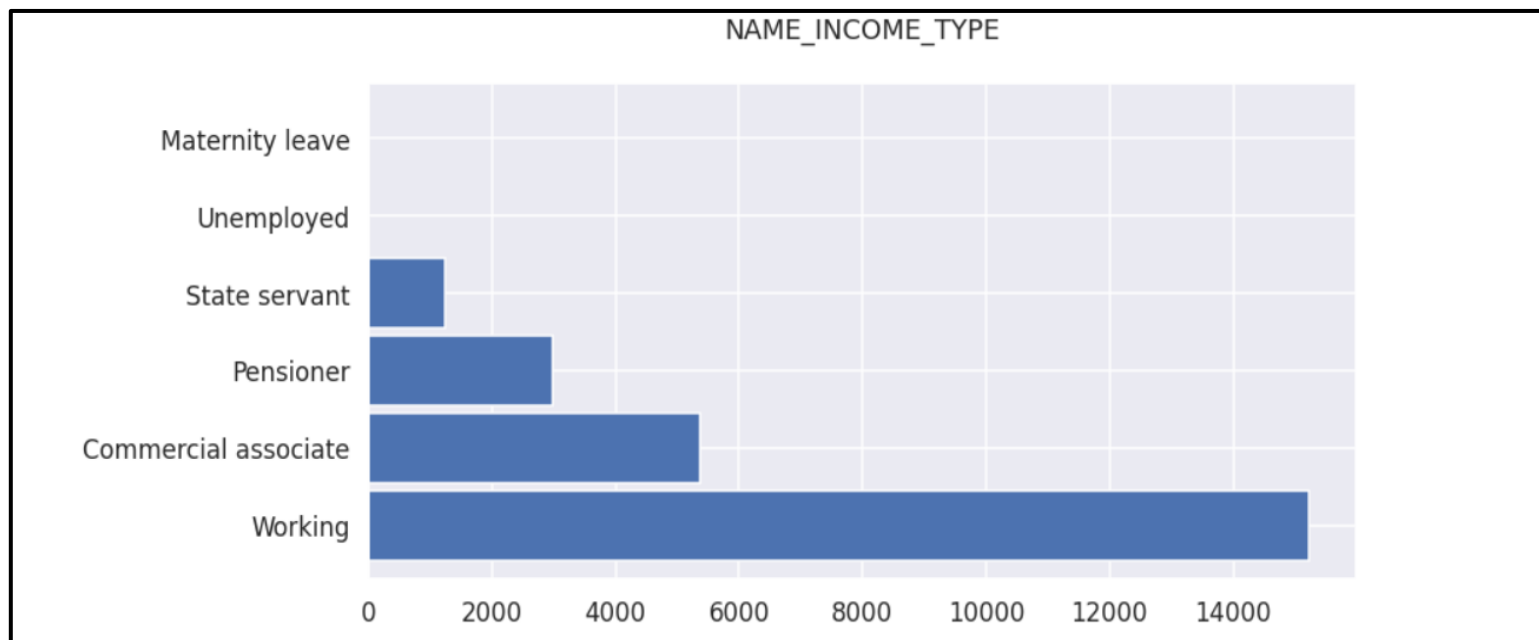
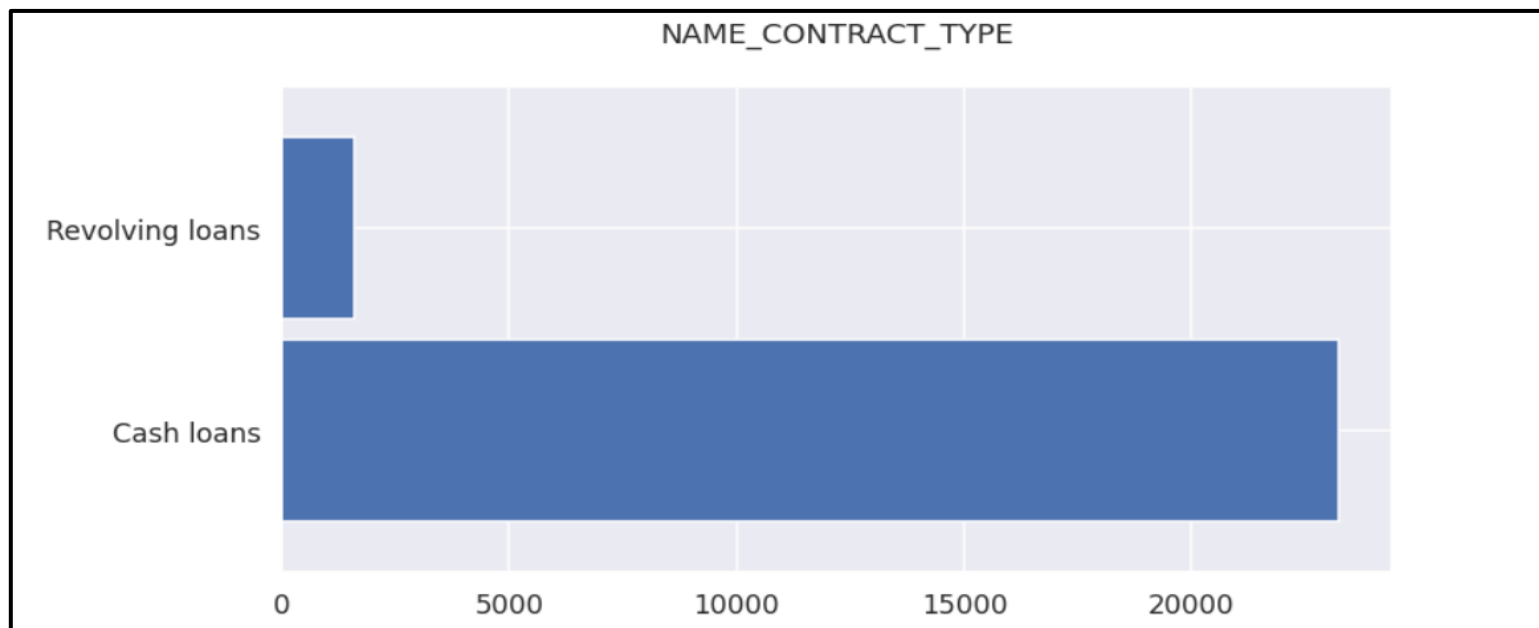


**Categorical univariate analysis for target = 1
(Customer with payment difficulties)**

Distribution of NAME_CONTRACT_TYPE and NAME_INCOME_TYPE

Points to be concluded from
graphs to the right

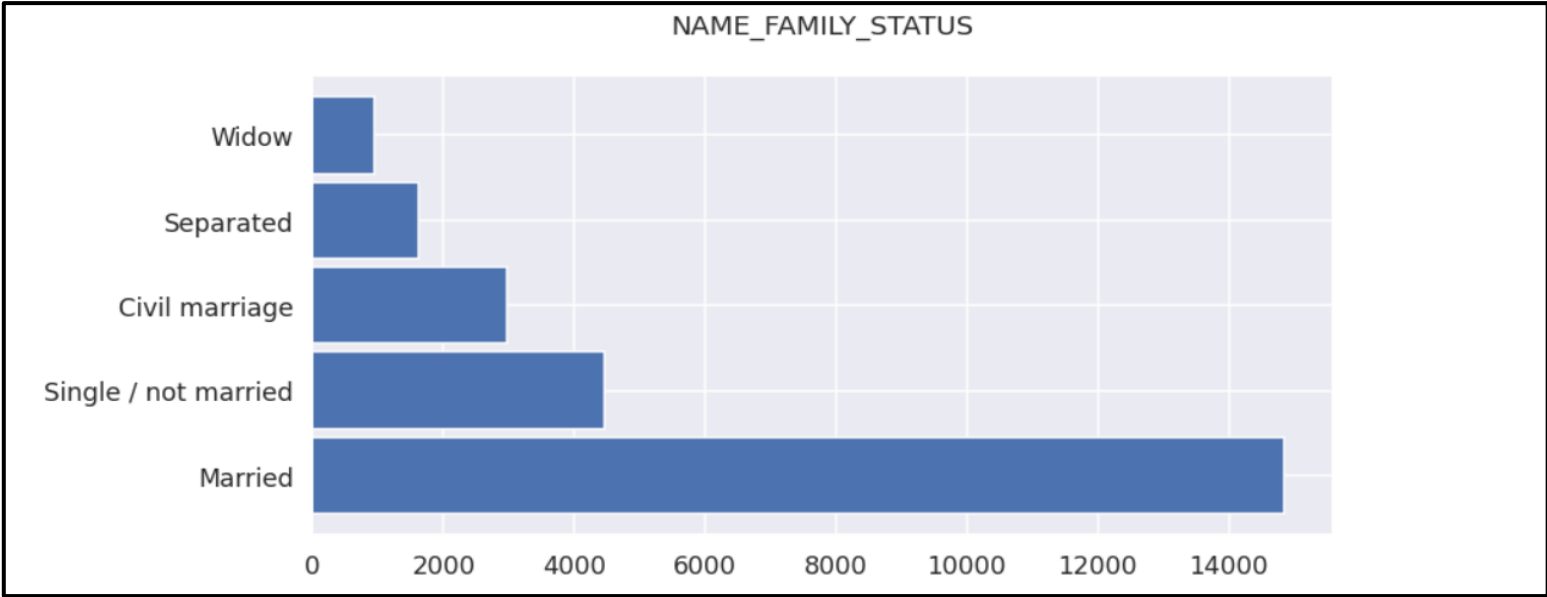
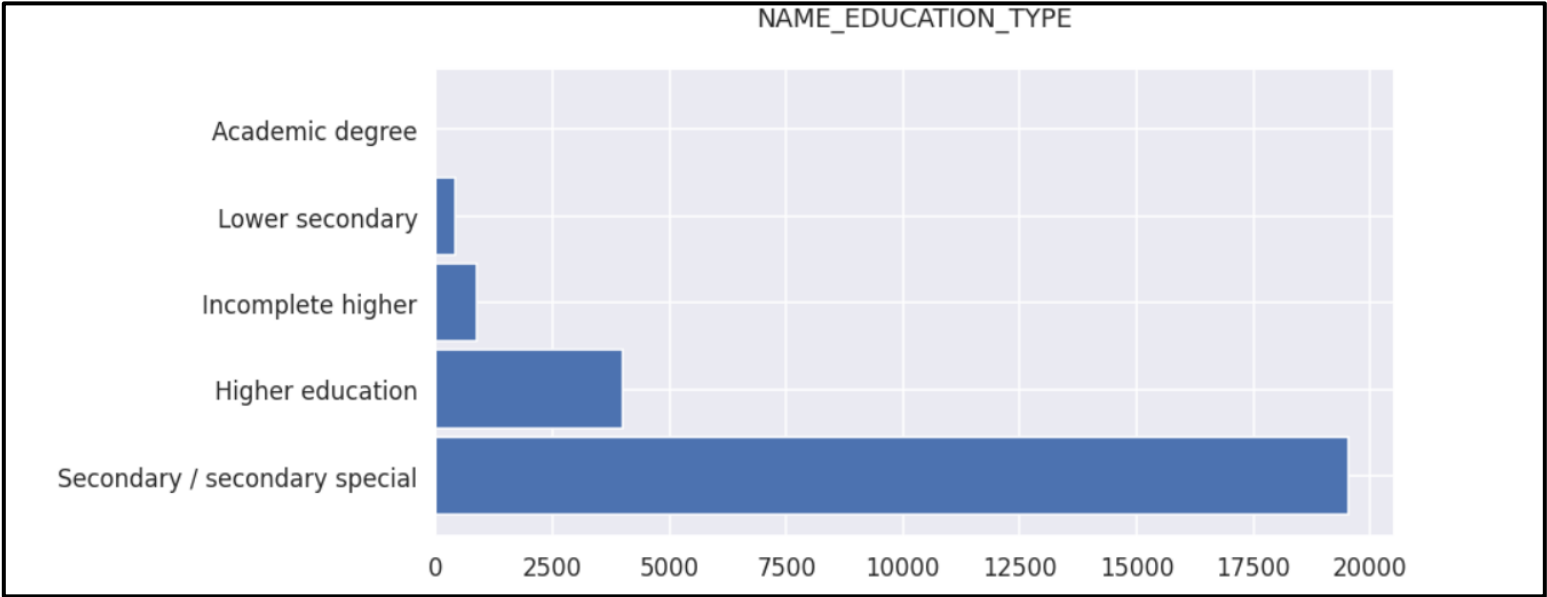
- For contract type 'Cash loans' is having higher number of credits than 'Revolving loans' type
- For income type 'working', 'commercial associate', and 'Pensioner' the number of credits are higher than others.



Distribution of NAME_EDUCATION_TYPE and NAME_FAMILY_STATUS

Points to be concluded from
graphs to the right

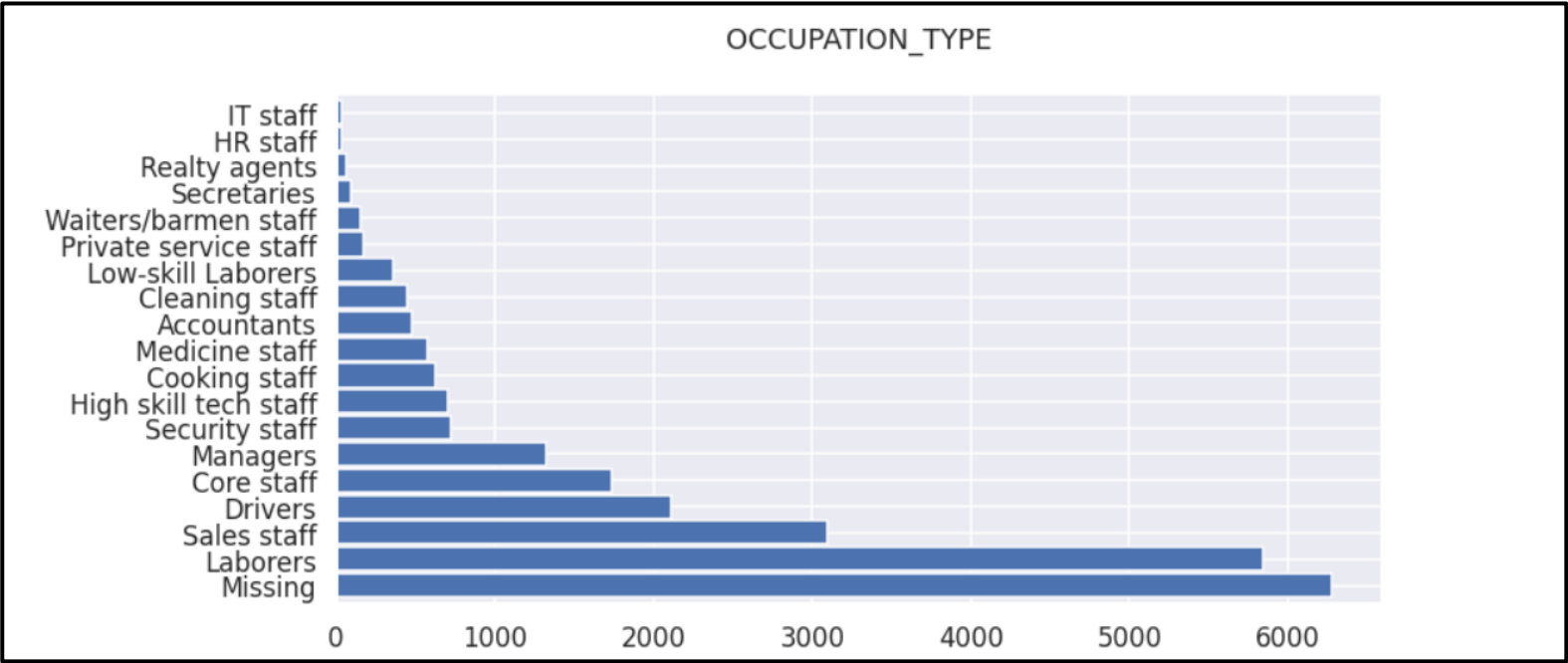
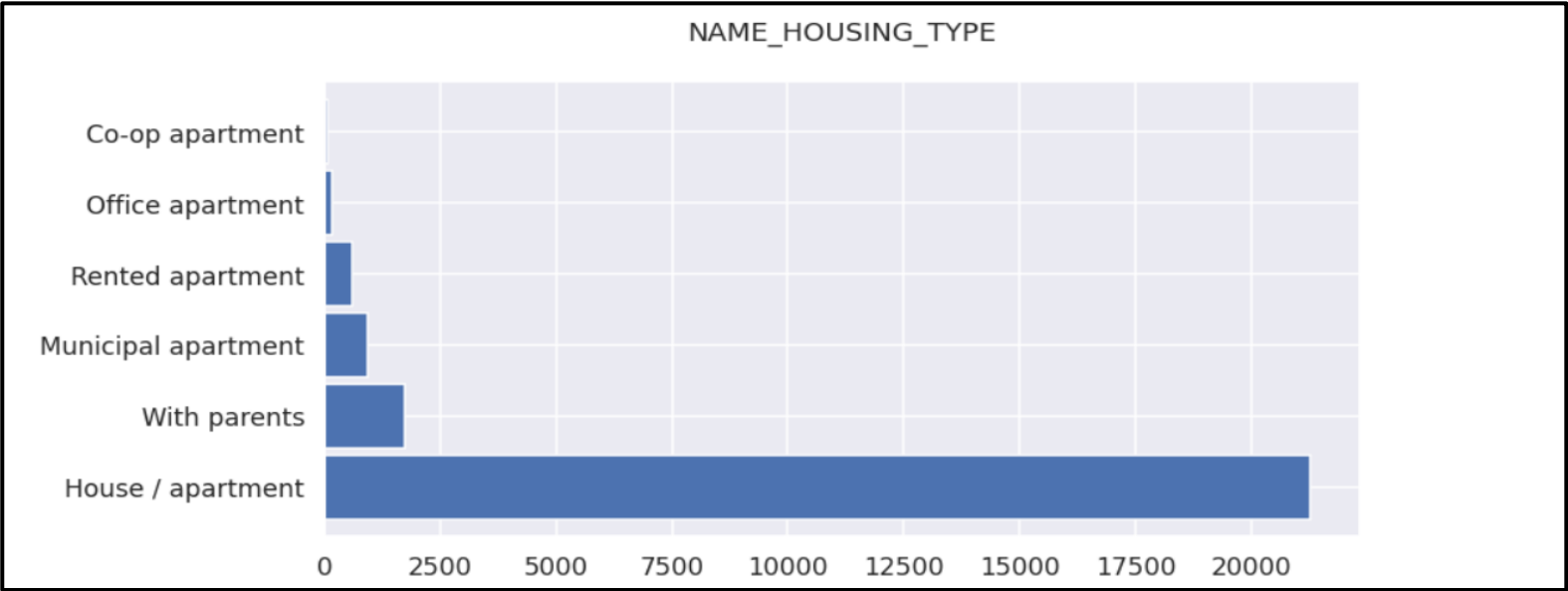
- People with
Secondary/Secondary
special and Higher
education are having a
greater number of credit
- Married people are having
a greater number of credit
and widow is having lesser
number of credit



Distribution of NAME_HOUSING_TYPE and OCCUPATION_TYPE

Points to be concluded from
graphs to the right

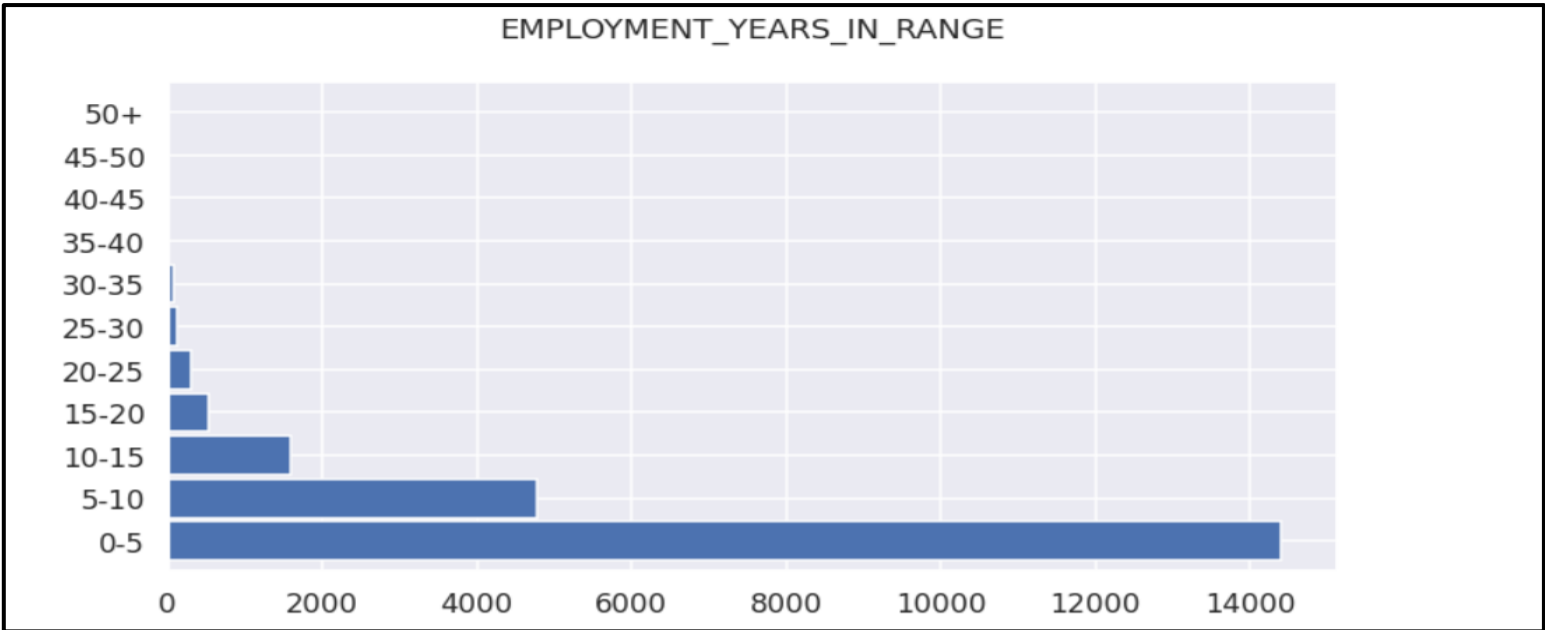
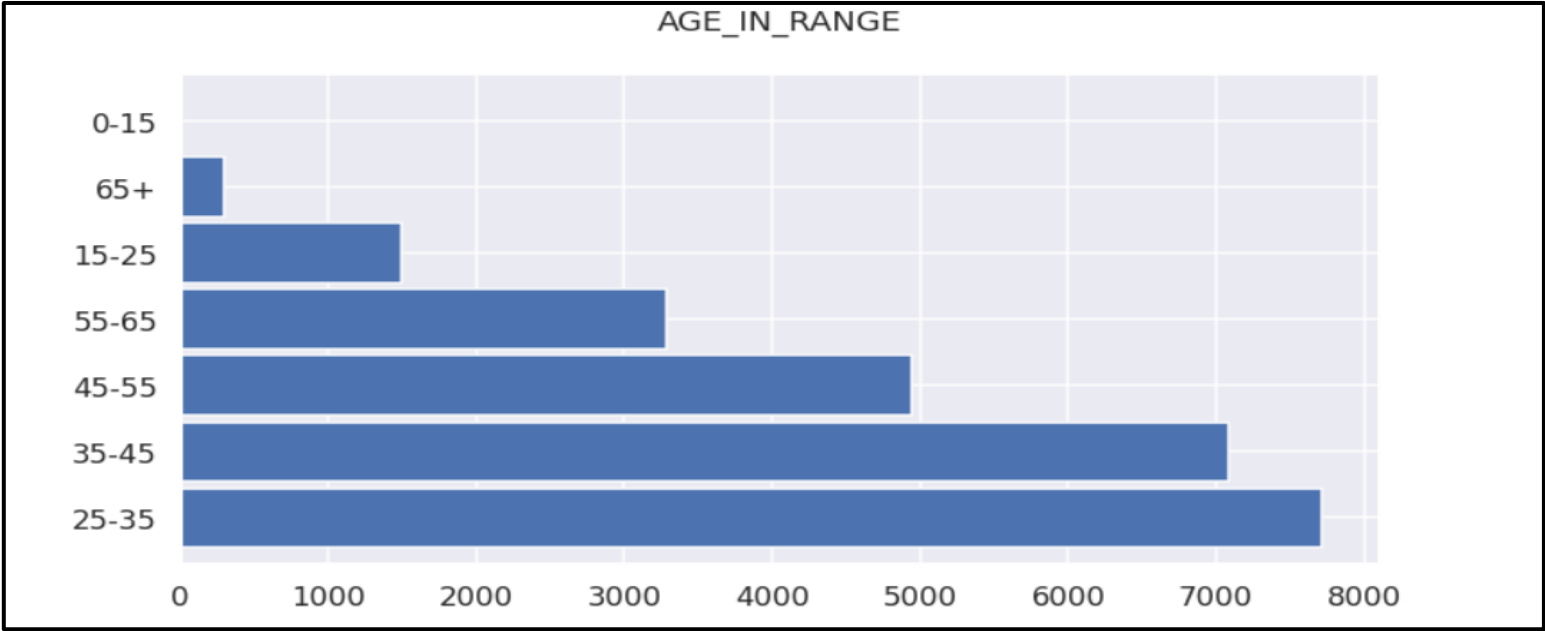
- People with own House/apartment are having a greater number of credit
- Laborers and sales staff are having a greater number of credit



Distribution of AGE_IN_YEARS and EMPLOYMENT_YEARS

Points to be concluded from
graphs to the right

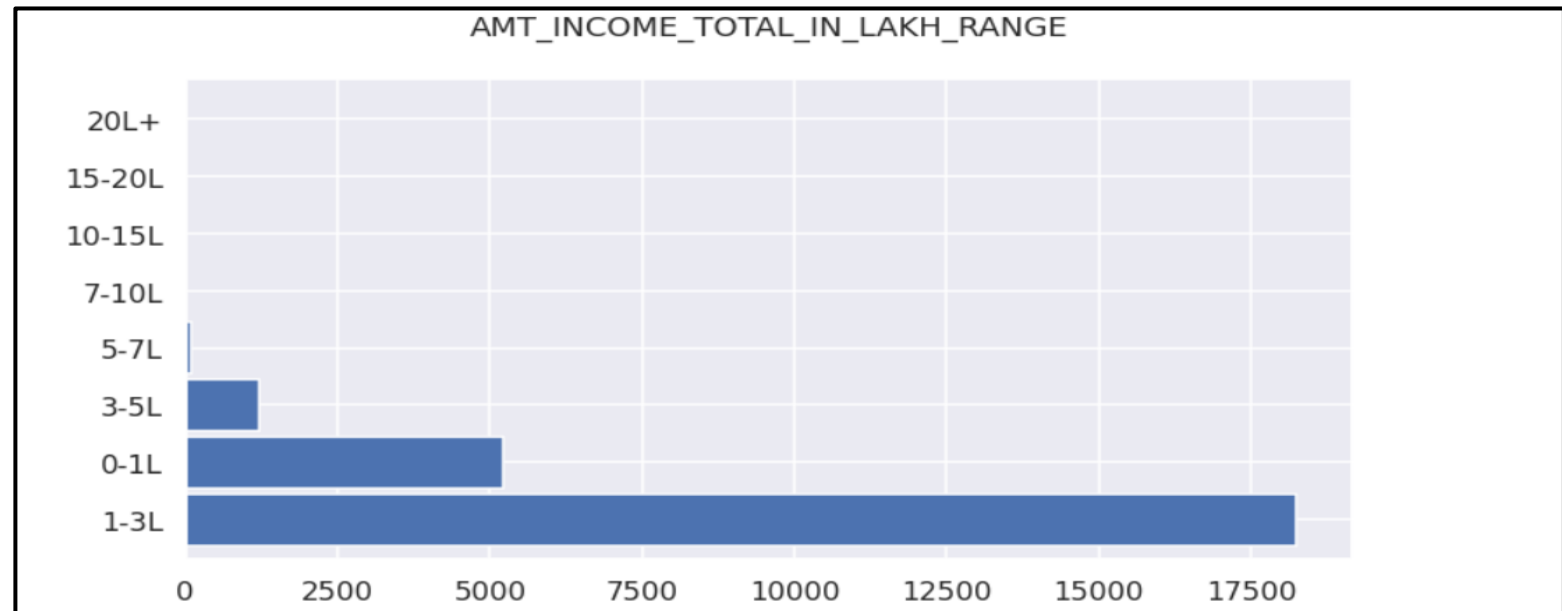
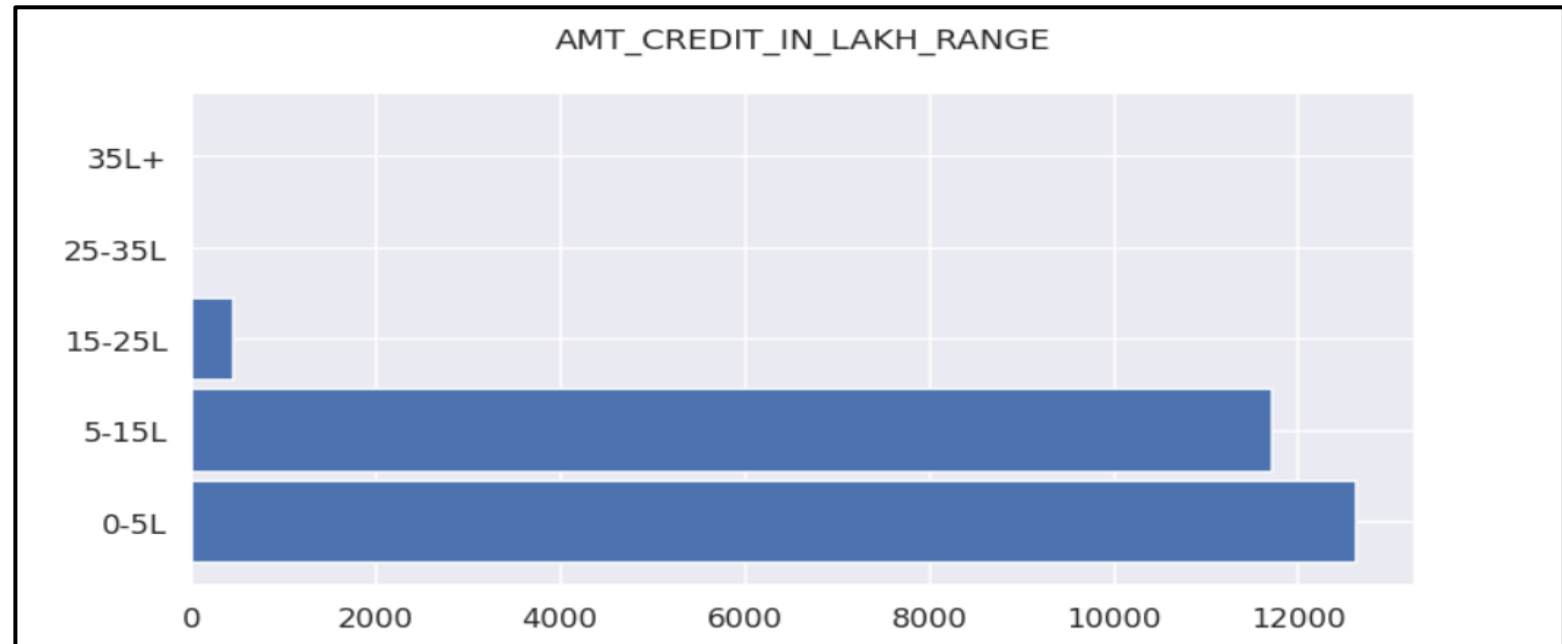
- People with age range 25-35 are having a greater number of credit
- People with employment years 0-5 are having a greater number of credit



Distribution of AMT_CREDIT and AMT_INCOME

Points to be concluded from
graphs to the right

- People with credit amount from 0 to 15 L are having a greater number of credit
- People with income 1-3 L are having a greater number of credit

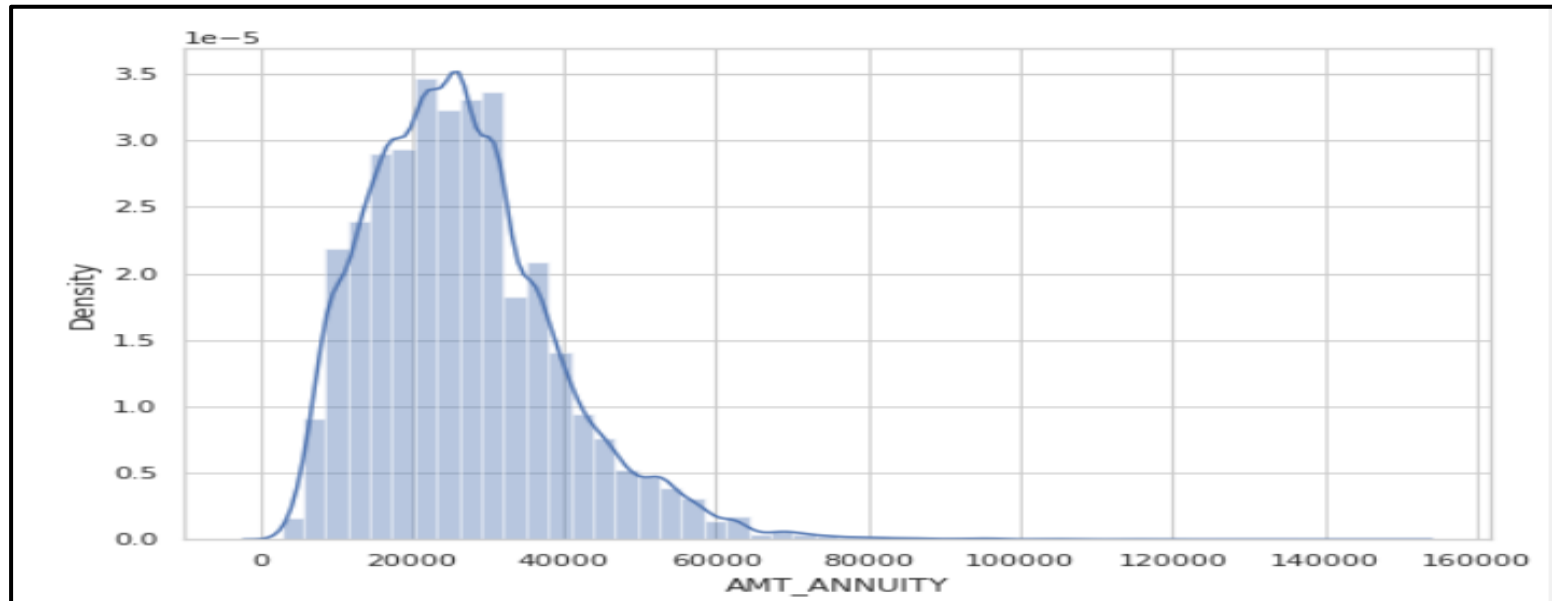
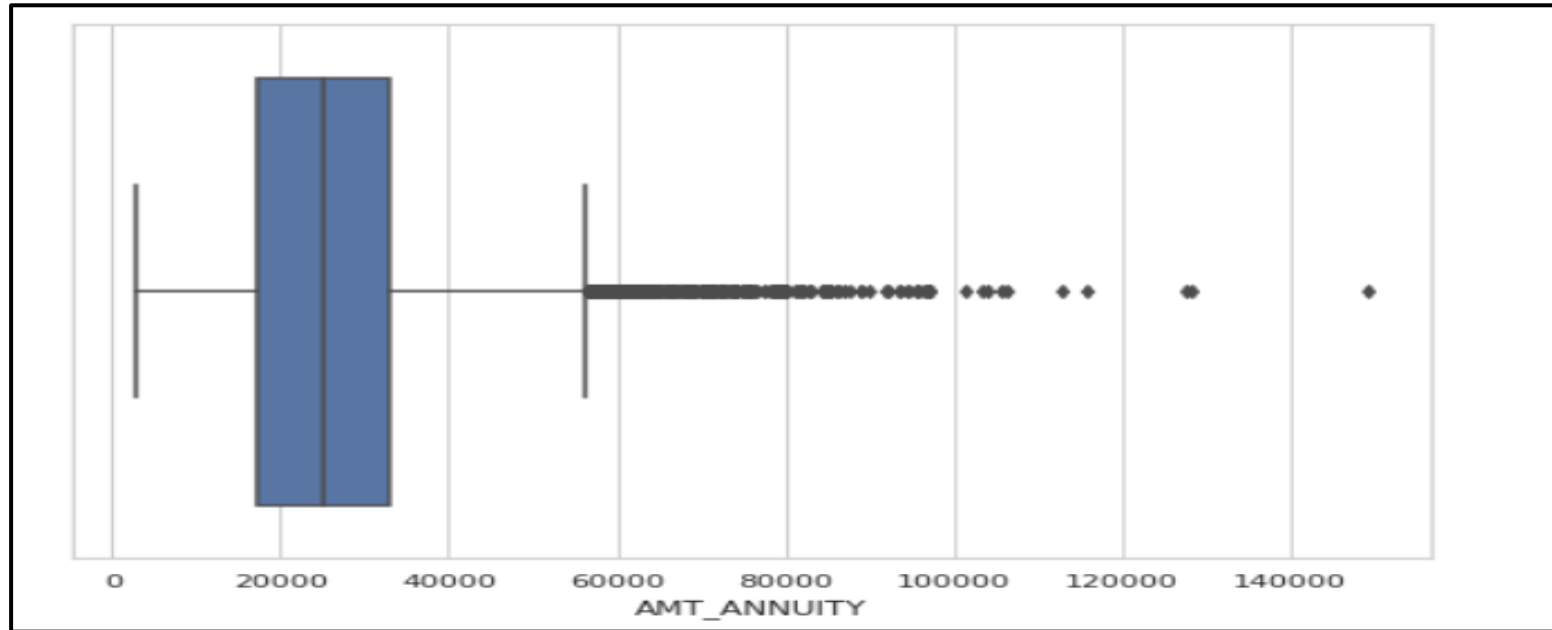


**Numerical univariate analysis for target = 1
(Customer with payment difficulties)**

Distribution of Annuity Amount

Points to be concluded from graphs to the right

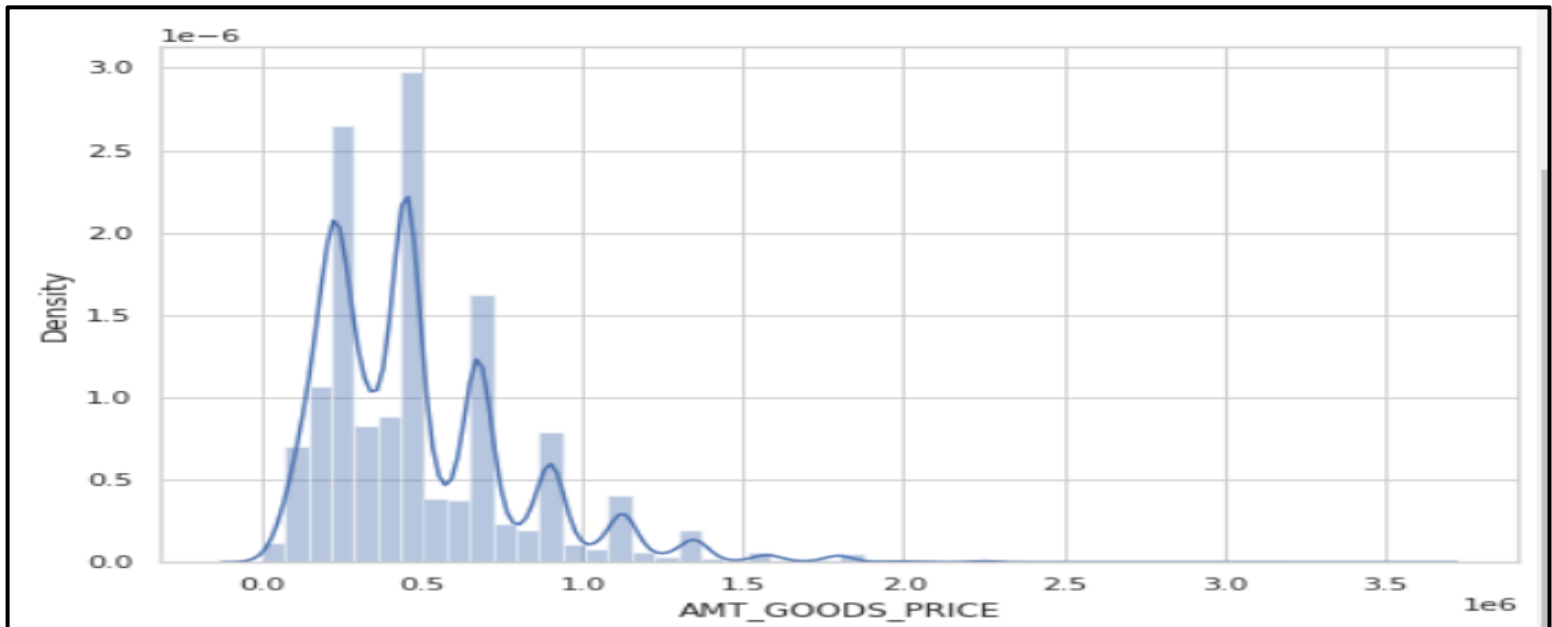
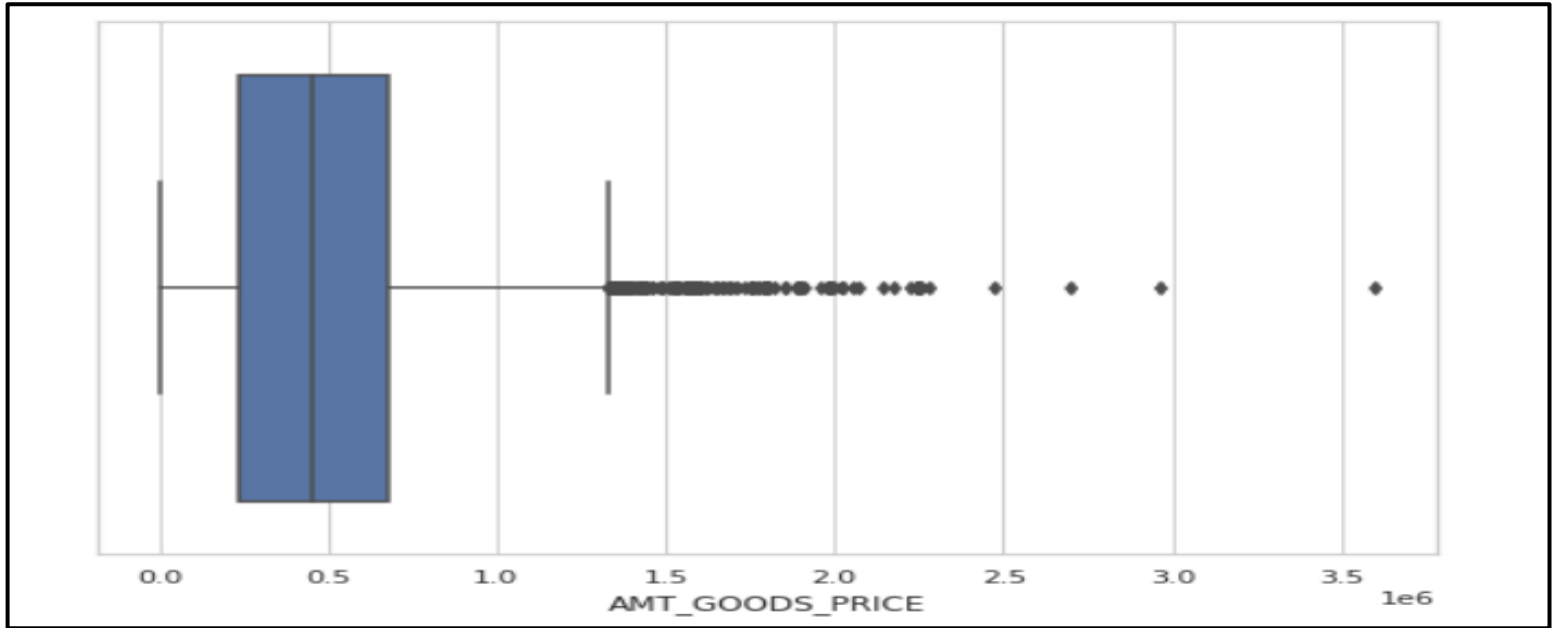
- Some Outliers are present in the annuity amount
- Maximum values are lies from 0 to 50,000 annuity amount and 50th percentile is close to 25,000 annuity amount



Distribution of Goods price Amount

Points to be concluded from graphs to the right

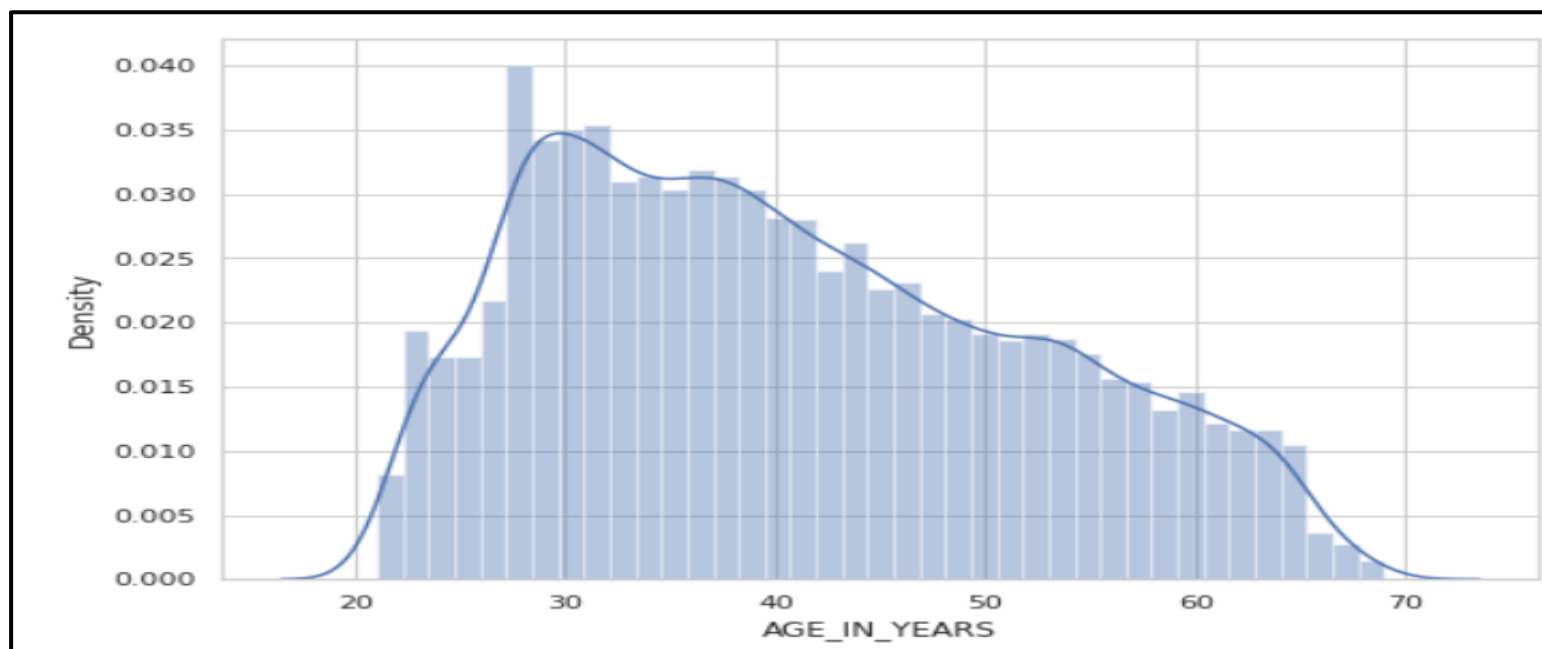
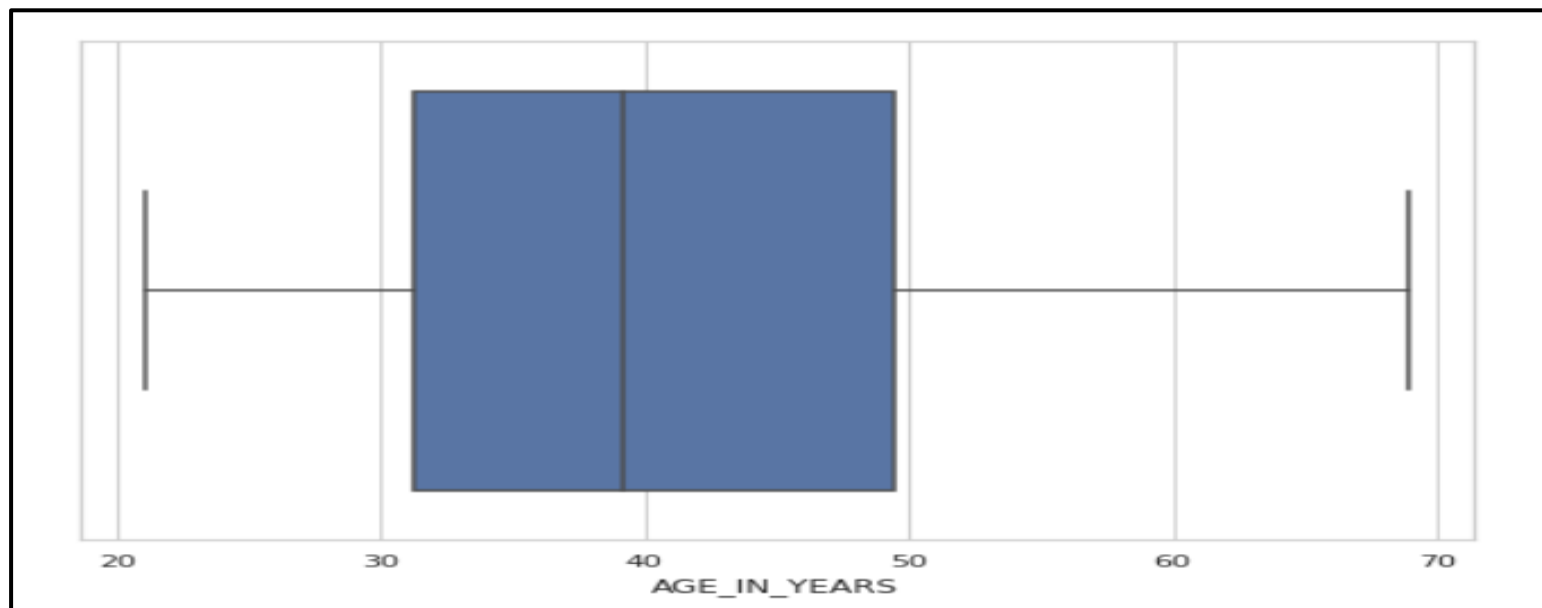
- Some Outliers are present in the goods price amount
- Maximum values lies from 0 to 10L approx. and first quartile is lesser than third quartile



Distribution of Age in years

Points to be concluded from graphs to the right

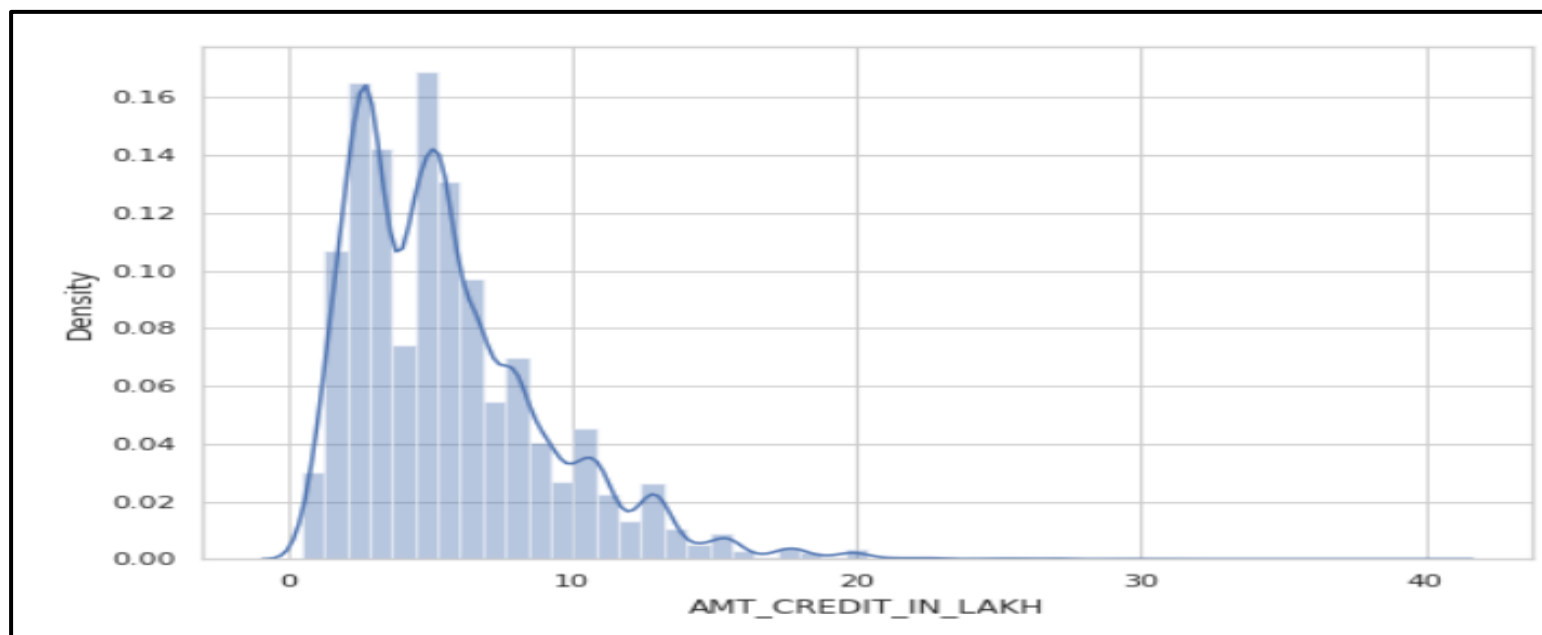
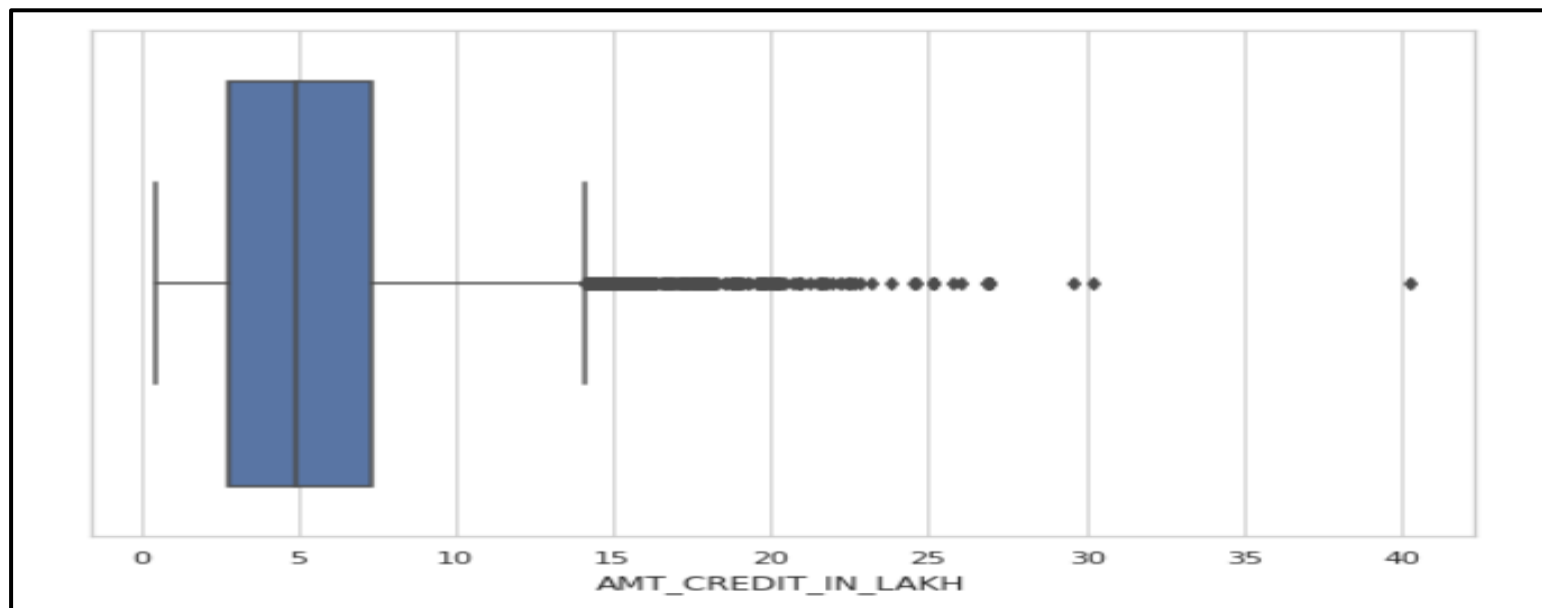
- Maximum values lies from 30 to 50 age in years
- Peoples which are 30 age are having maximum number of credits



Distribution of Credit Amount

Points to be concluded from graphs to the right

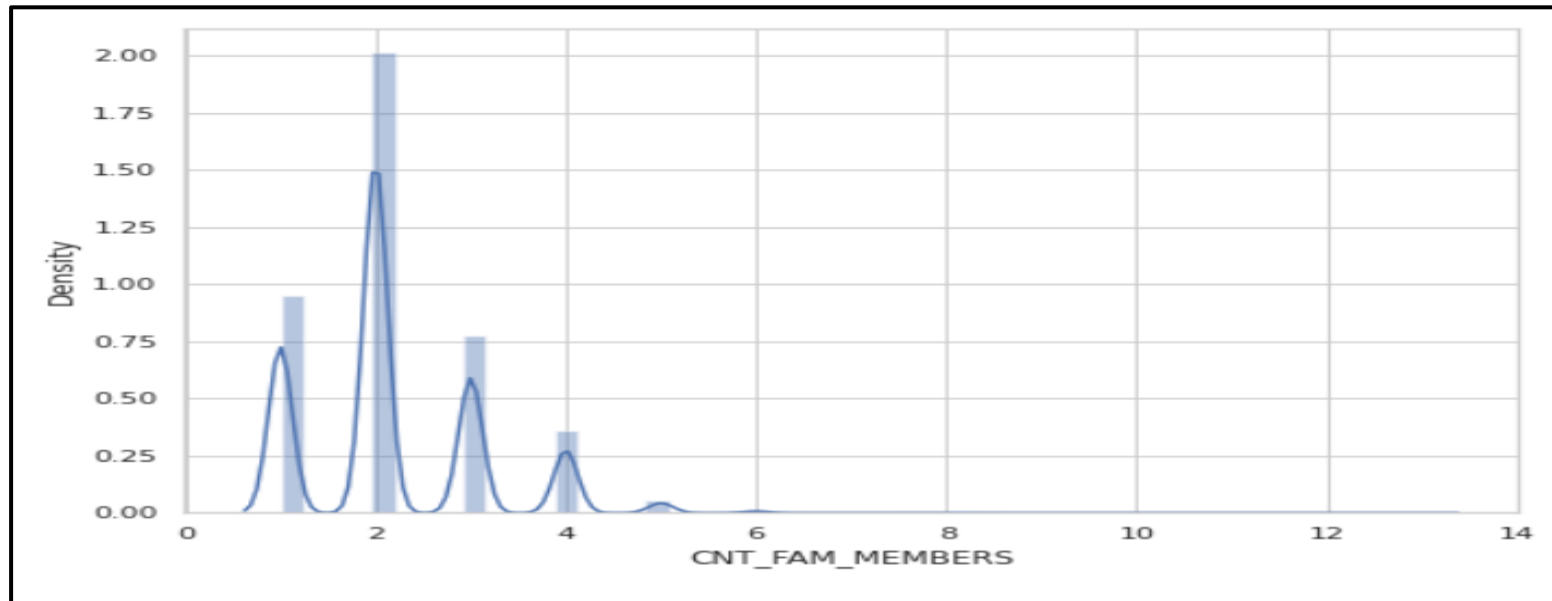
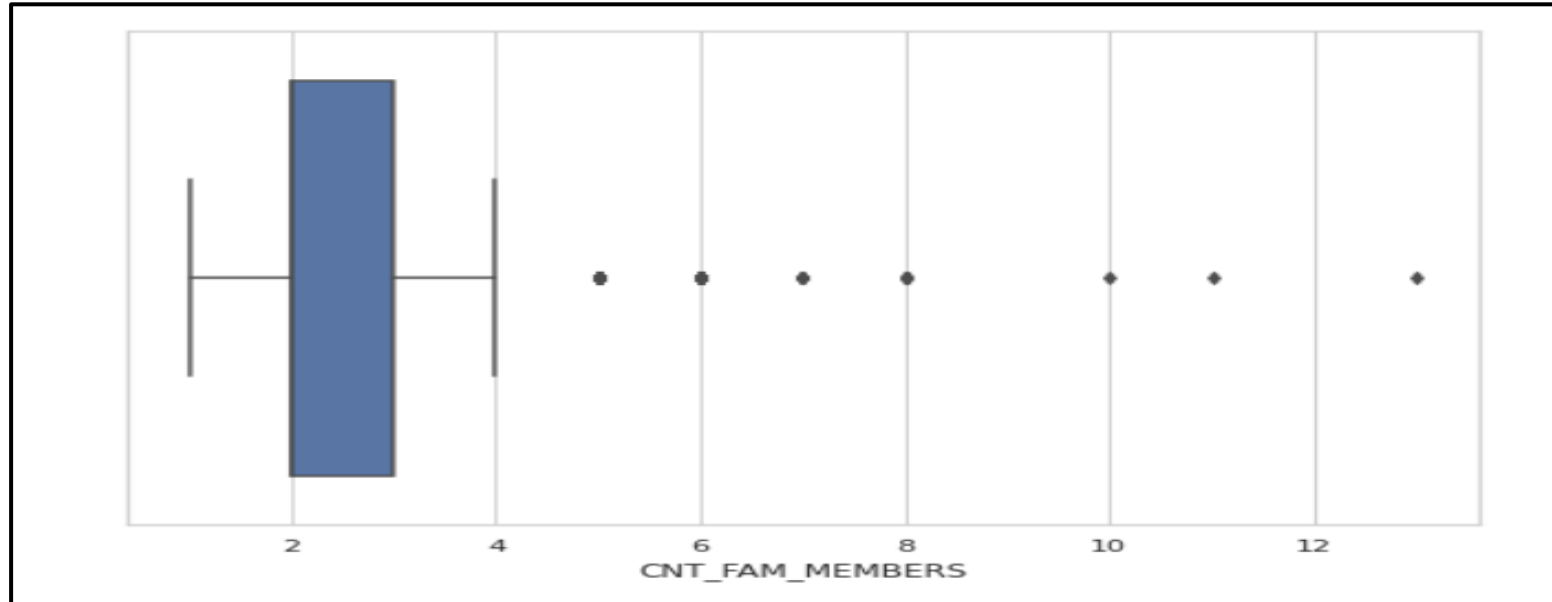
- Maximum values lies from 0 to 10 L credit amount
- Some outliers are present in credit amount



Distribution of Customer Family Members

Points to be concluded from graphs to the right

- Maximum values lies from 1 to 4
- Some outliers are present in Family members
- Most of the customers who are having higher credit having 2 Family members

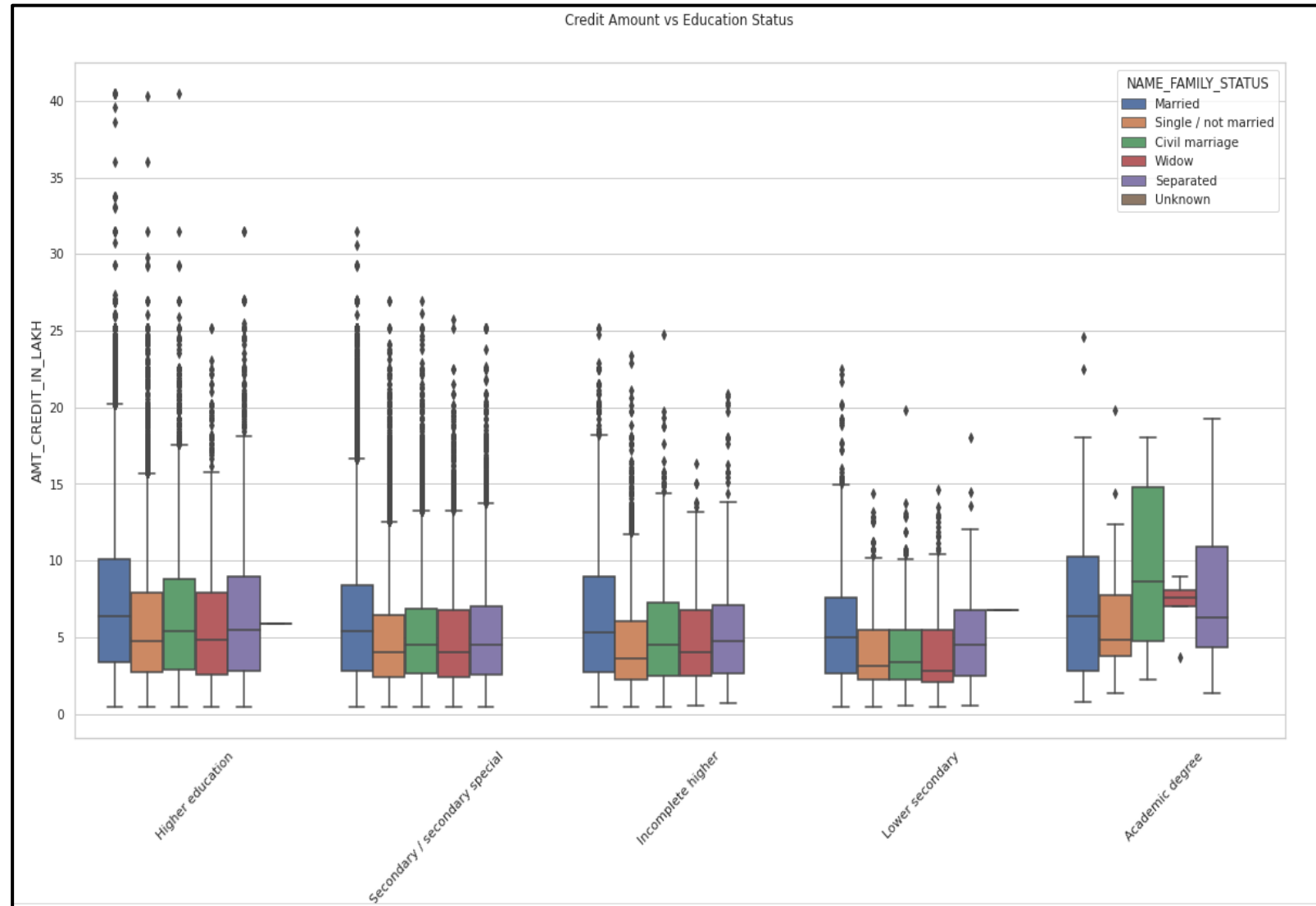


**Numerical bivariate analysis for target = 0
(Customer with no payment difficulties)**

Credit Amount Vs Education Status

Points to be concluded from graphs to the right

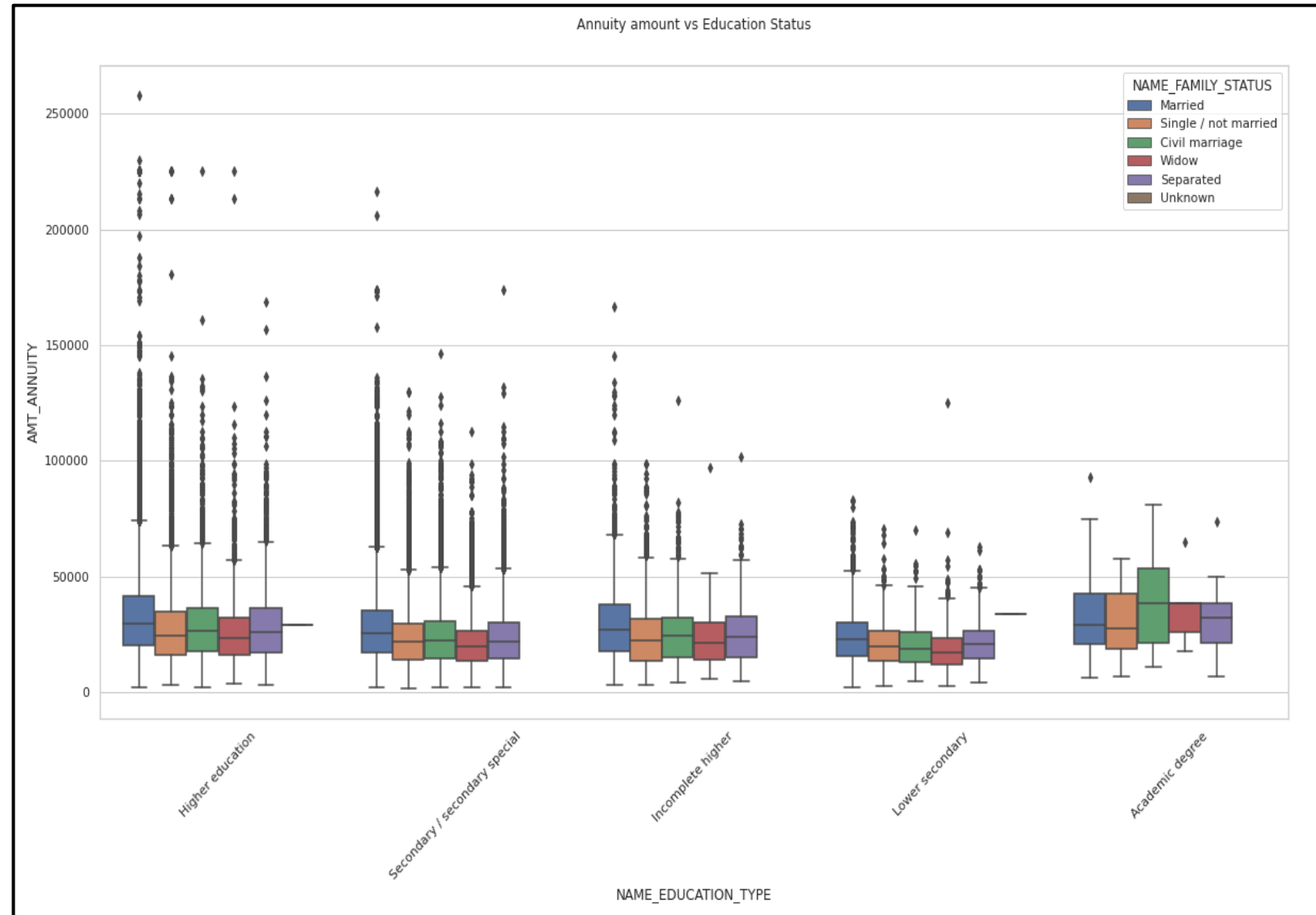
- Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.
- Higher education of family status of 'marriage', 'single' and 'civil marriage' are having more outliers.
- Civil marriage for Academic degree is having most of the credits in the third quartile.



Annuity Amount Vs Education Status

Points to be concluded from graphs to the right

- For Education type 'Lower Secondary' the annuity amount mean is mostly equal with family status. It contains less outliers.
- Less outlier are having for Academic degree, but they are having the annuity amount is little higher than Lower secondary.
- Higher Education is having more outliers

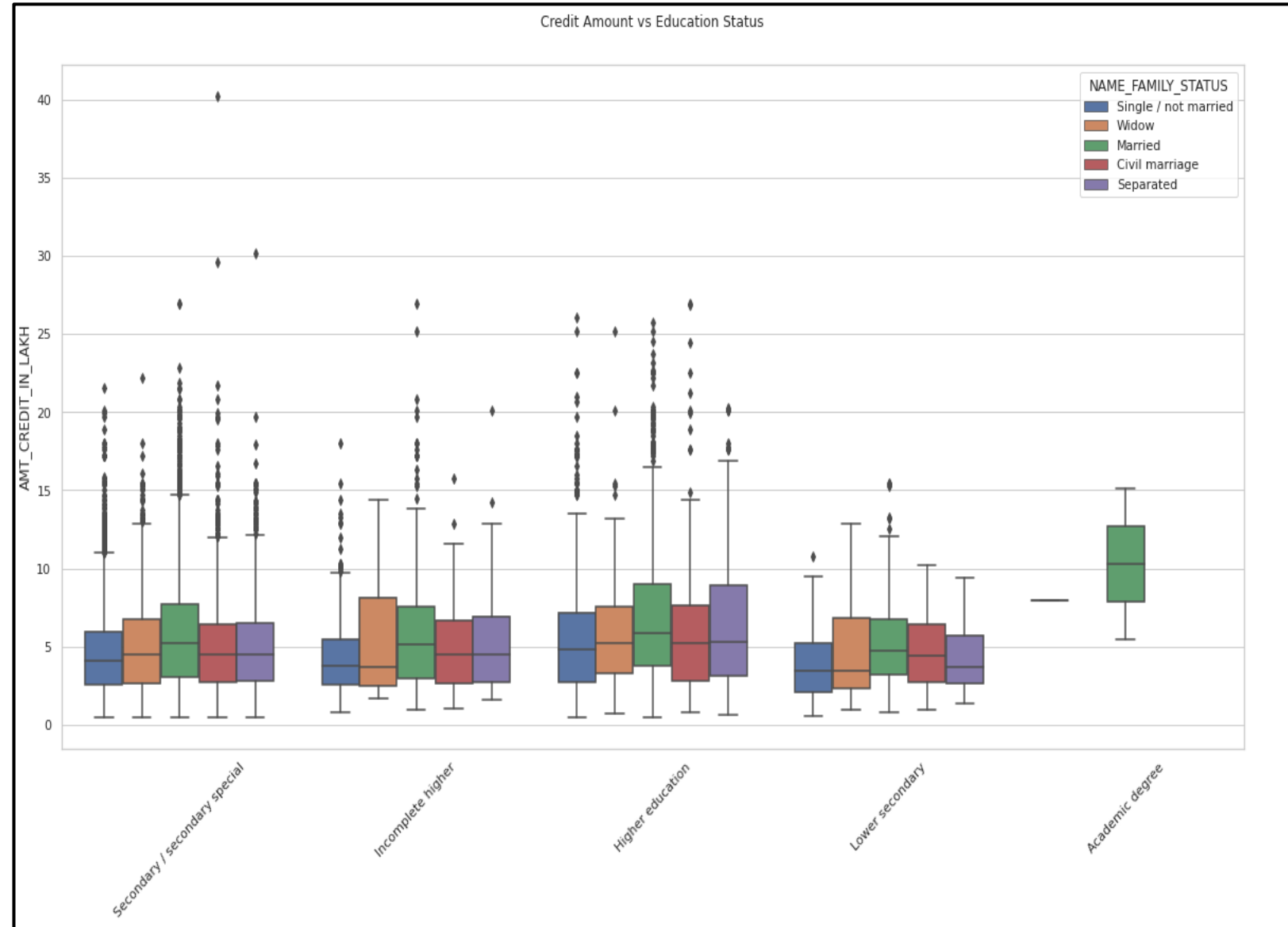


**Numerical bivariate analysis for target = 1
(Customer with payment difficulties)**

Credit Amount Vs Education Status

Points to be concluded from graphs to the right

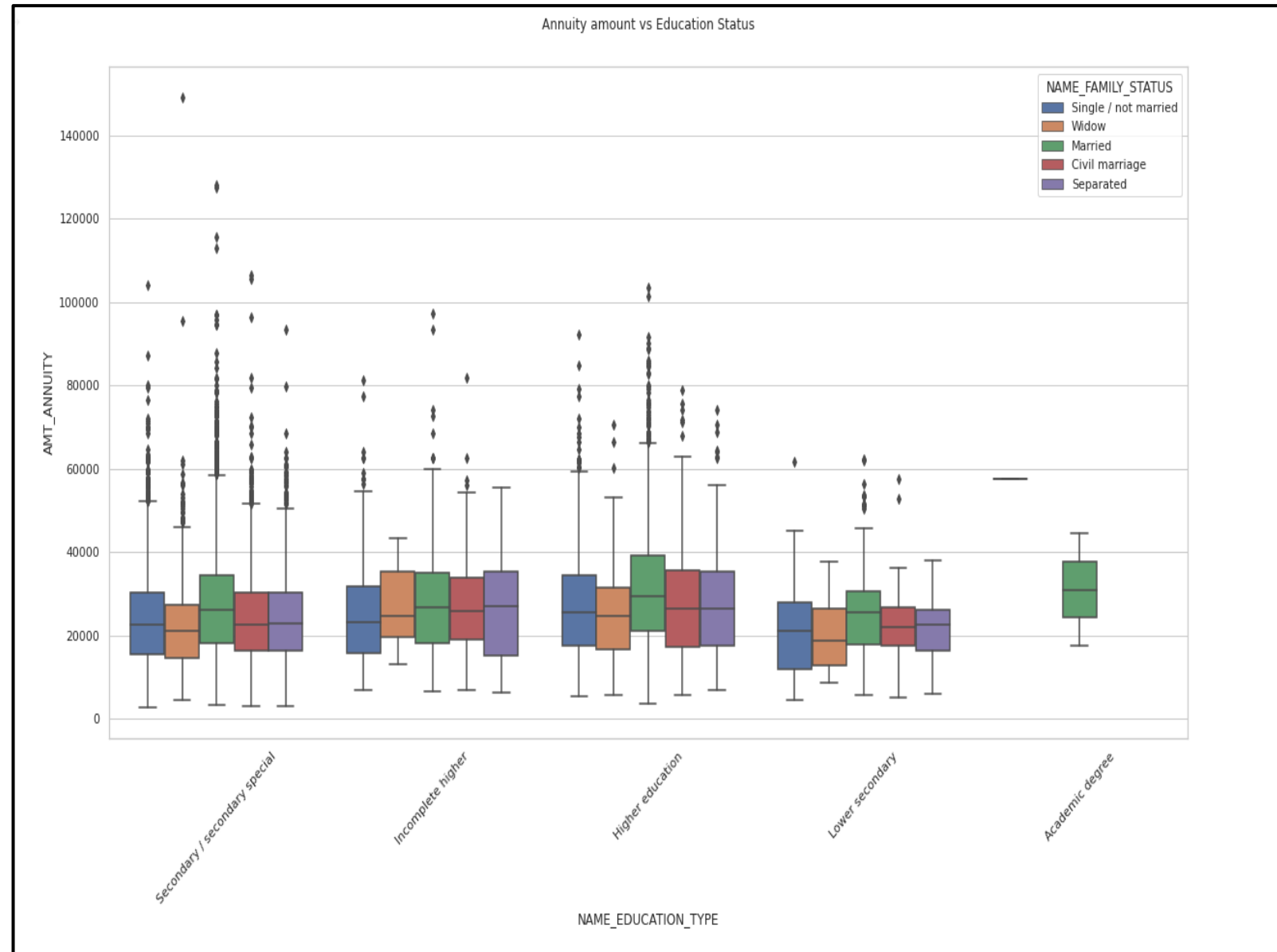
- Family status of 'marriage' of Academic degree education are having higher number of credits than others.
- Secondary/ secondary special with family status of 'Civil Marriage' are having more outliers.



Annuity Amount Vs Education Status

Points to be concluded from graphs to the right

- For Education type 'Lower Secondary' contains less outliers.
- Less outlier are having for Academic degree, but they are having the annuity amount is little higher than Lower secondary.
- Secondary/Secondary special are having more outliers



Distribution of NAME_CONTRACT_TYPE and NAME_CONTRACT_STATUS

Points to be concluded from
graphs to the right

- For contract type 'Cash loans' and 'Consumer loans' are having higher number of credits than 'Revolving loans' type
- For Contract status 'Approved' are having higher number of credits than others.

