Exploratory Data Analysis
CREDIT EDA CASE STUDY

BY

PRASHANTH REDDY PAGIDALA

Objectives

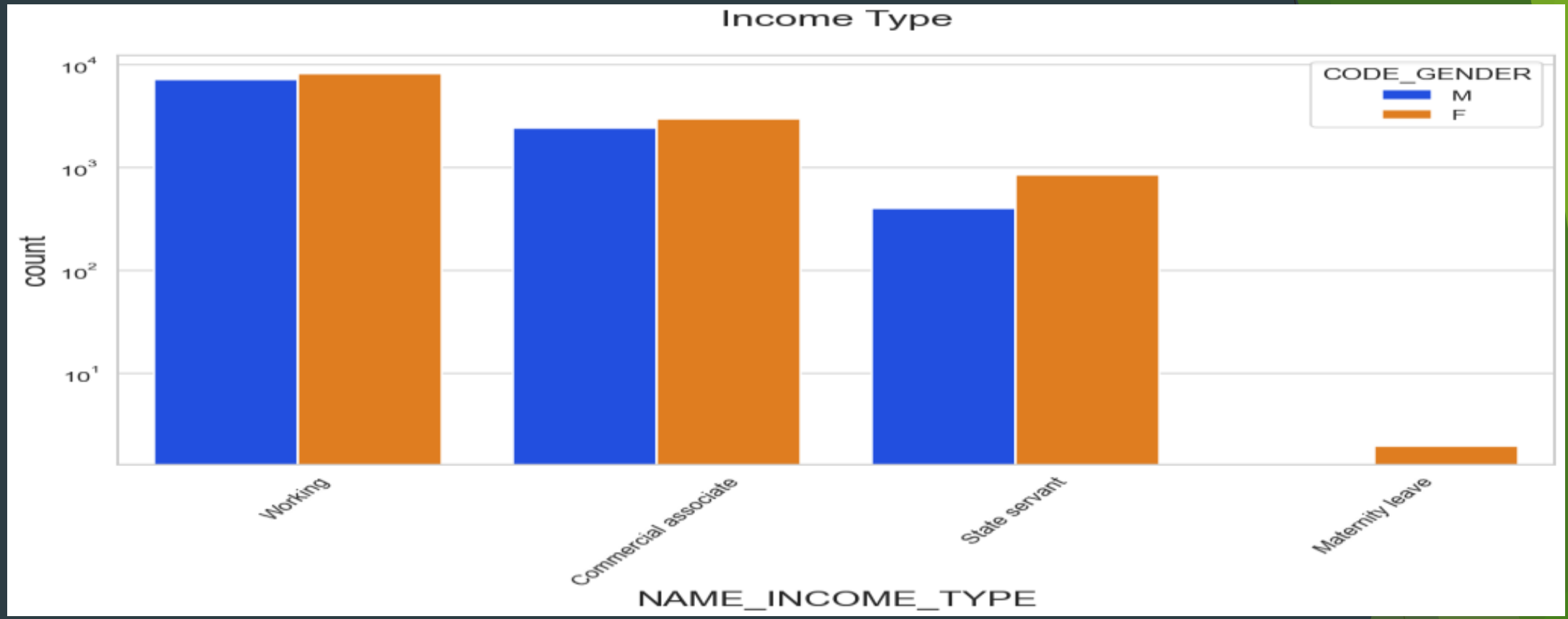
Business Understanding as follows

1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company and vice versa
2. Risk assessment
3. Problem statement and the analysis approach
4. Identify the missing data and use appropriate method to deal with it.

Univariate Analysis on Categorical Variables in the applicant data of the category – Clients with no payments difficulties

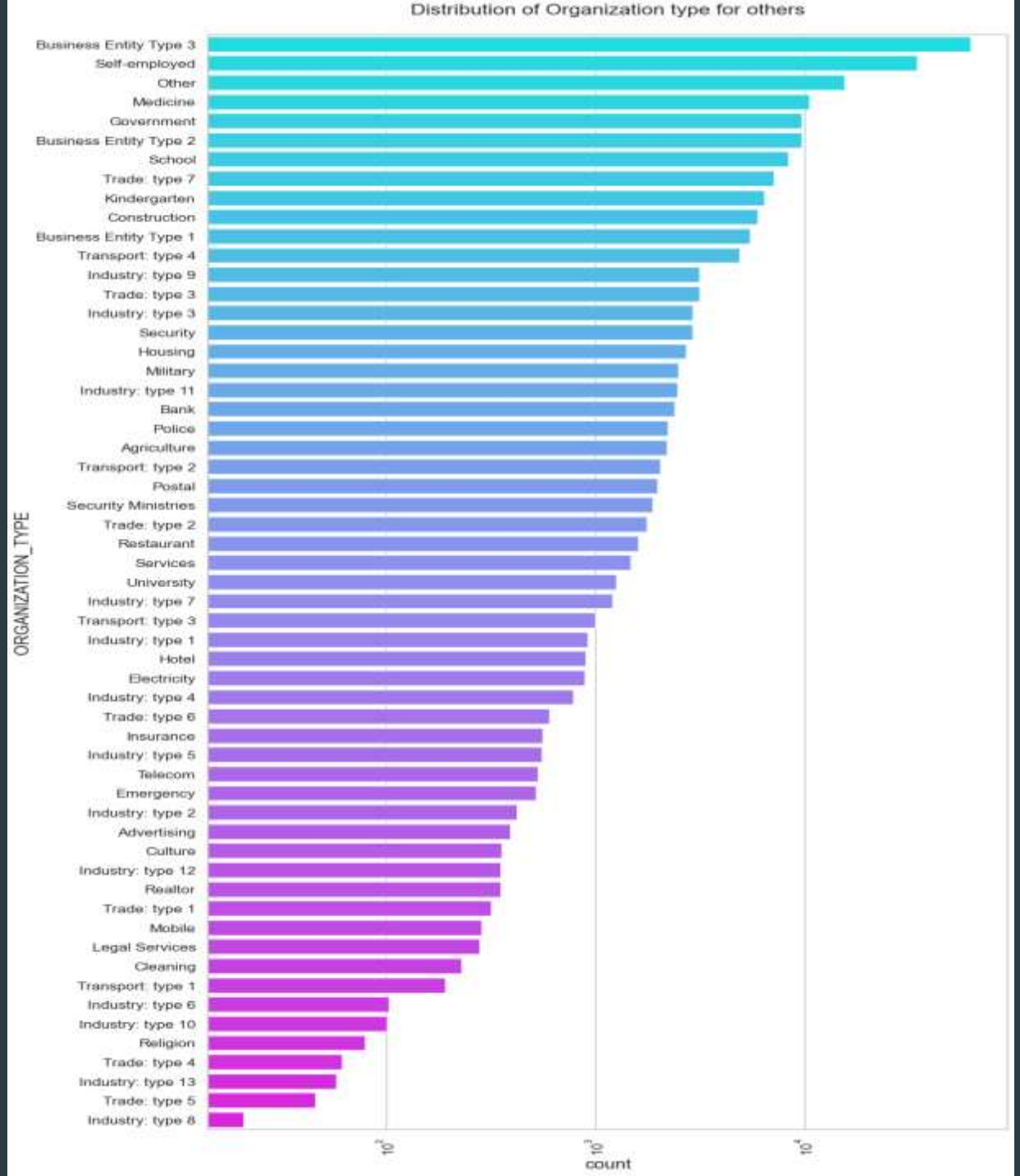
Insights:

1. Female count is in majority in terms of clients with no payment difficulties, in majority of the ranges
2. Income range 100000 - 200000 has highest number of credits.
3. There are very less credits in the range >450000 and above compared to other ranges



Insights:

- For the income type Working, Commercial Associate, and State Servant, the number of credits are higher than others.
- In terms of Income type also, Females are having higher number of credits than male.
- Less number of credits are of income type Maternity leave.

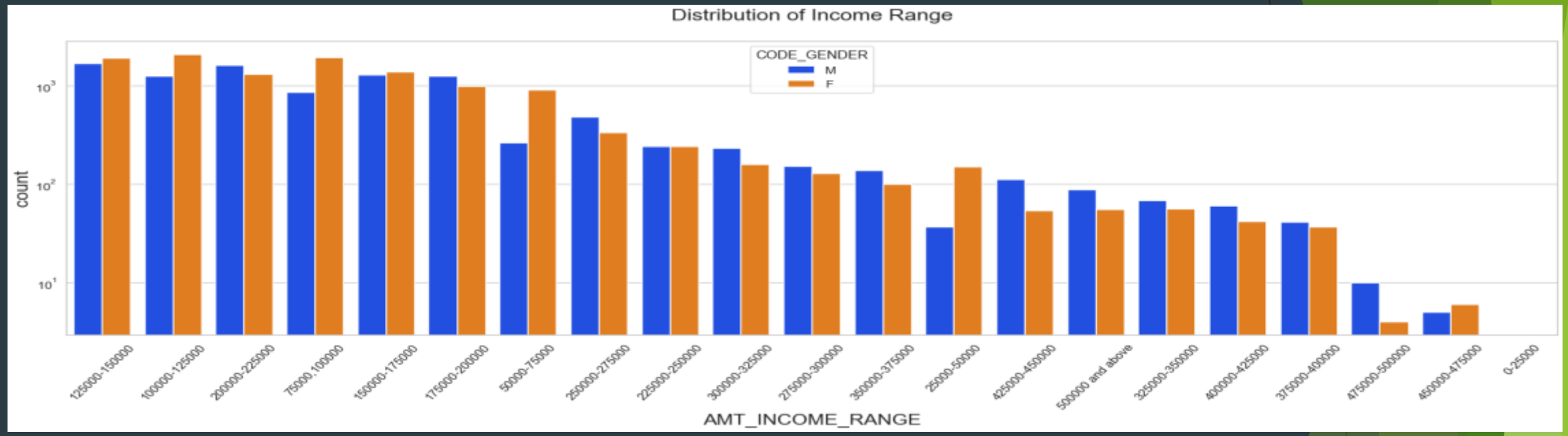


Insights:

- Clients who applied for credits are mostly from Business entity Type 3, Self employed, Other , Medicine and Government.

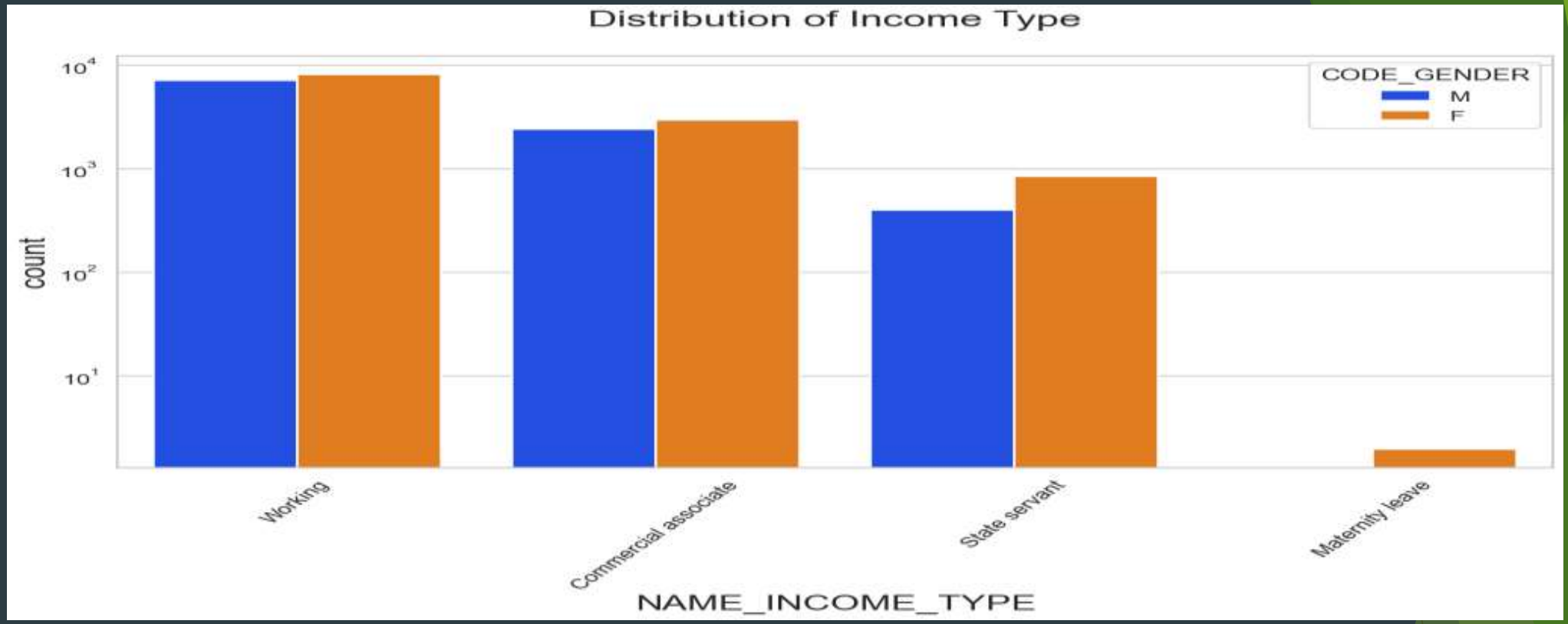
- Less number of clients are from Industry type 8,type 6, type 10, religion and trade type 5, type 4.

Univariate Analysis on Categorical Variables in the applicant data of the category – Clients with payments difficulties



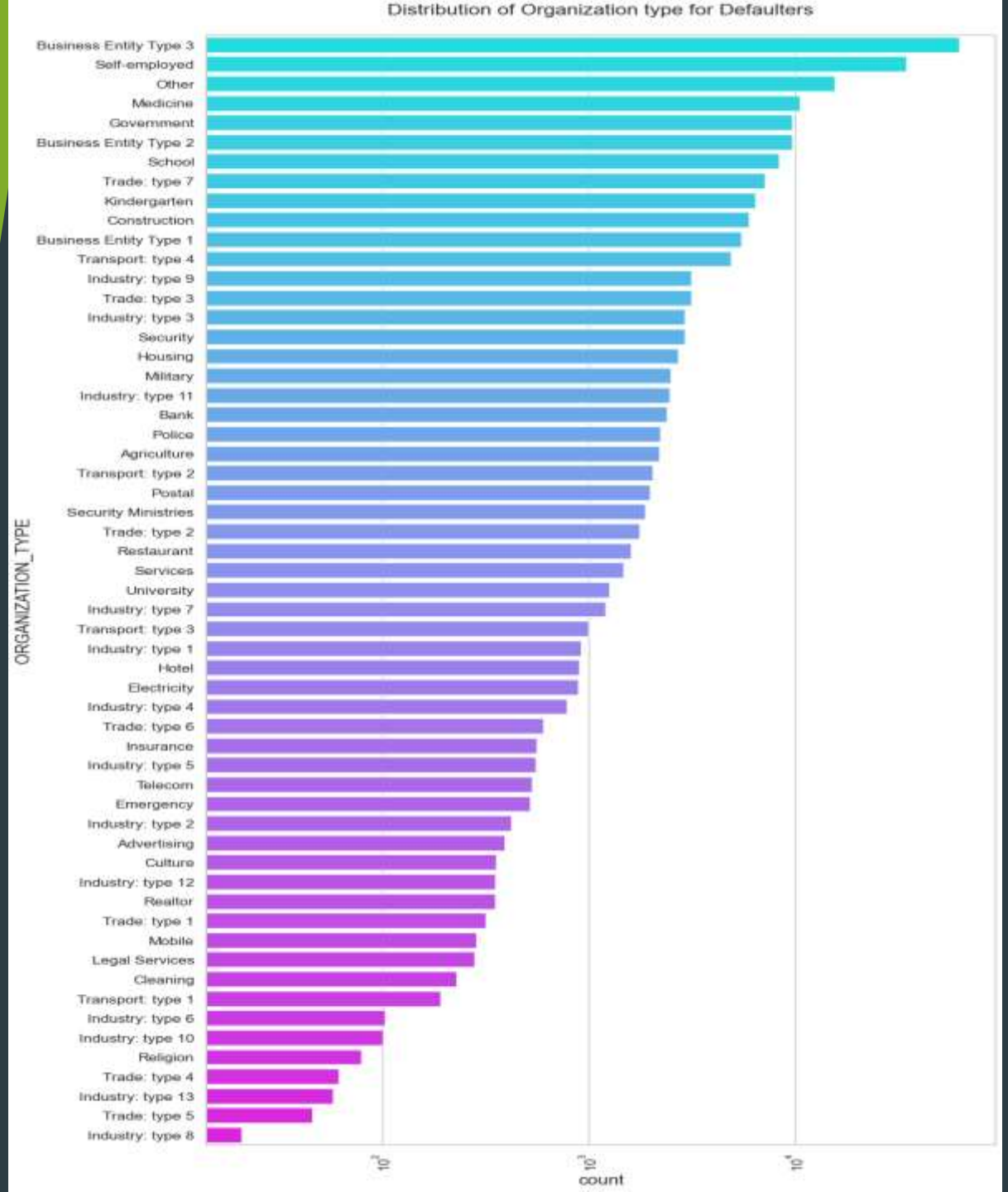
Insights:

- Count of males is higher than female.
- Income range 100000 - 200000 has more number of credits.



Insights:

- For working, Commercial Associate, and State Servant, the number of credits are higher than other i.e. Maternity leave.
- Females have more number of credits than males.
- There are no other income types as observed in clients with payments difficulties , like student, pensioner, etc

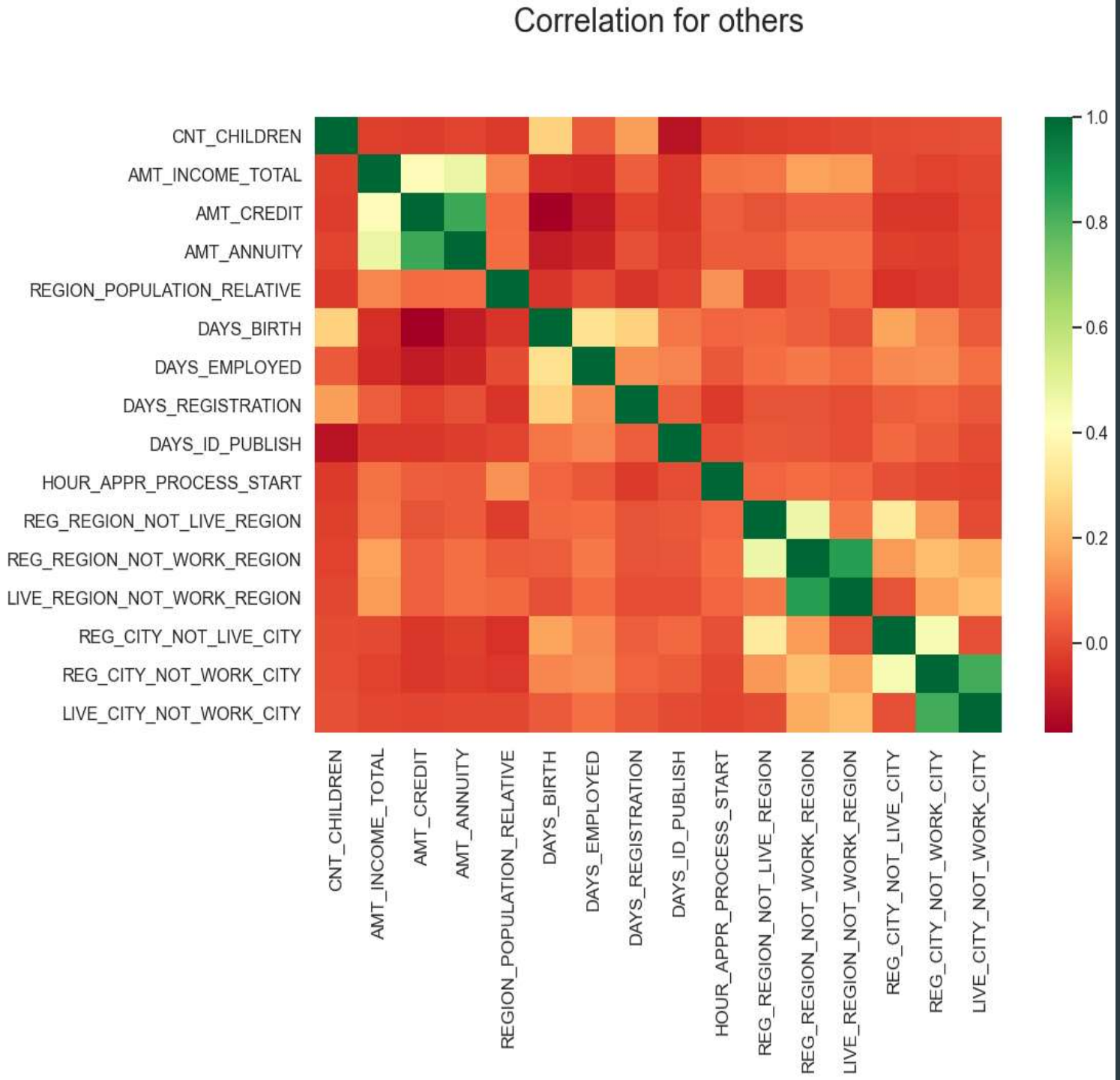


Insights

Clients who have applied for credits are mostly of Business entity Type 3 , Self employed , Other , Medicine and Government.

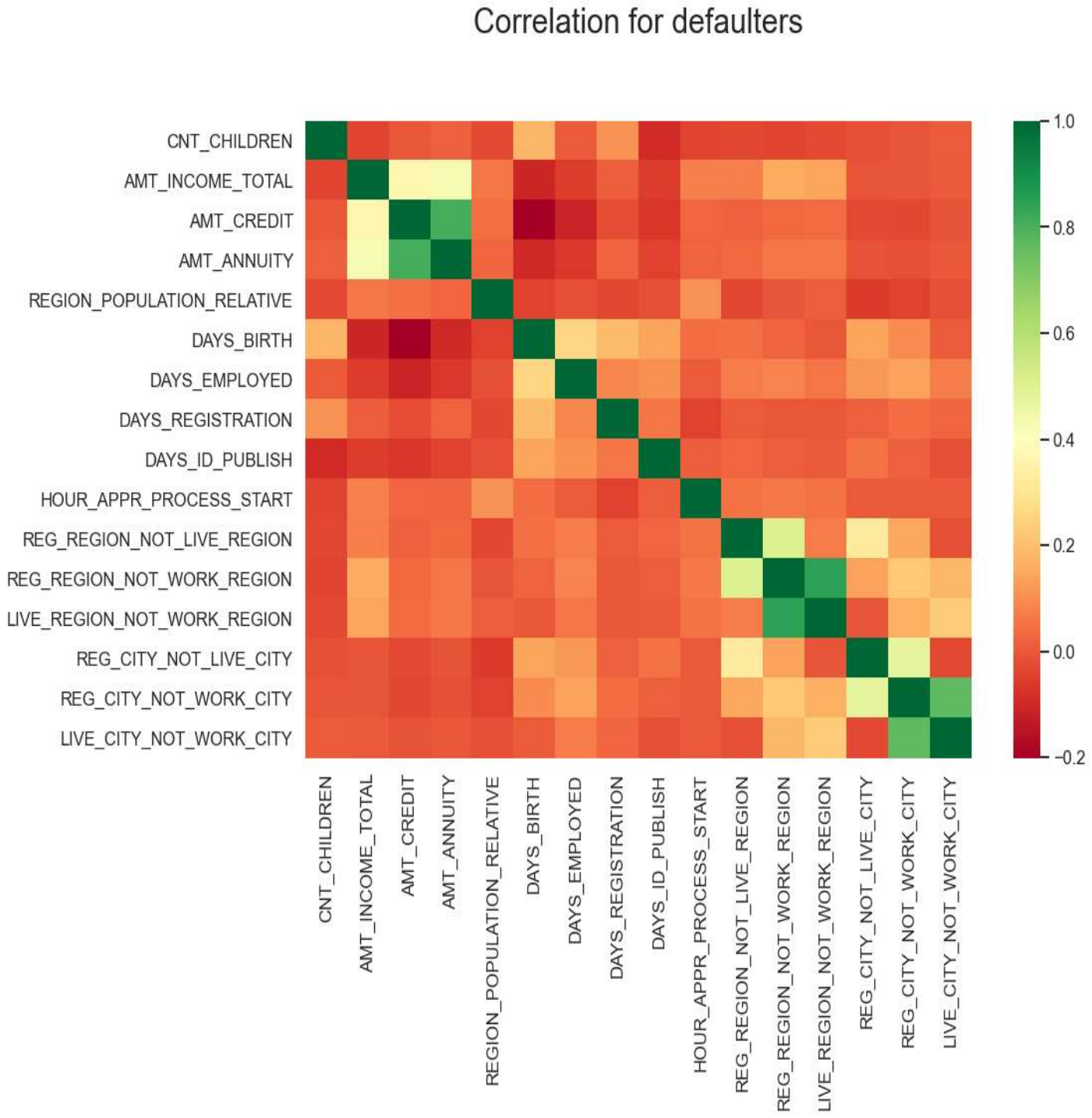
Correlation between

Clients with payments difficulties and Clients with no payments difficulties



Insights:

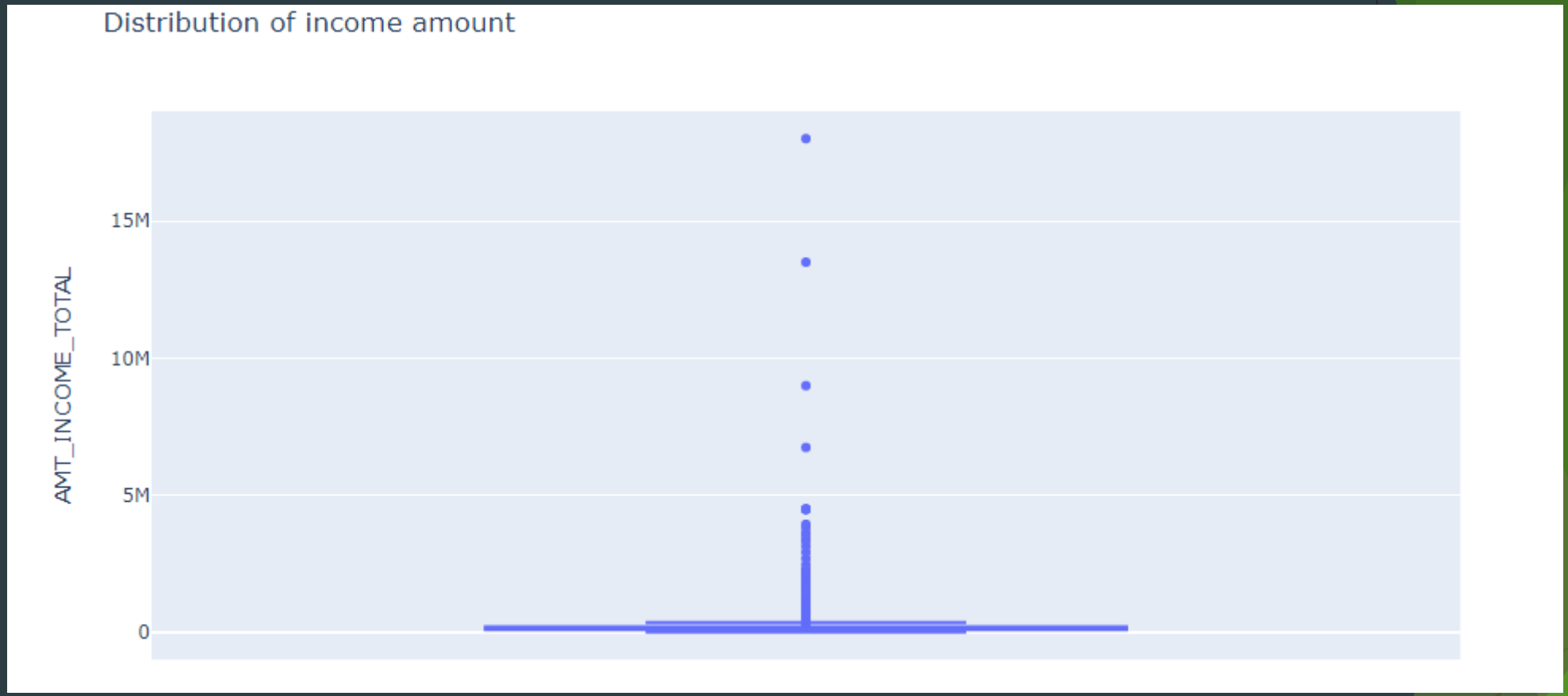
- Credit amount and Income are higher in densely populated area.
- The credit amount is inversely proportional to the date of birth, which means Credit amount is higher for low age and vice-versa.
- The credit amount is inversely proportional to the number of children client has, means Credit amount is higher for clients with less children and vice-versa.
- Income amount is inversely proportional to the number of children client have, means high income for clients with less children and vice-versa.



Insights

- Defaulters also has almost same observations as in others.
- For Clients who have less children, their permanent address does not match contact address and vice-versa.

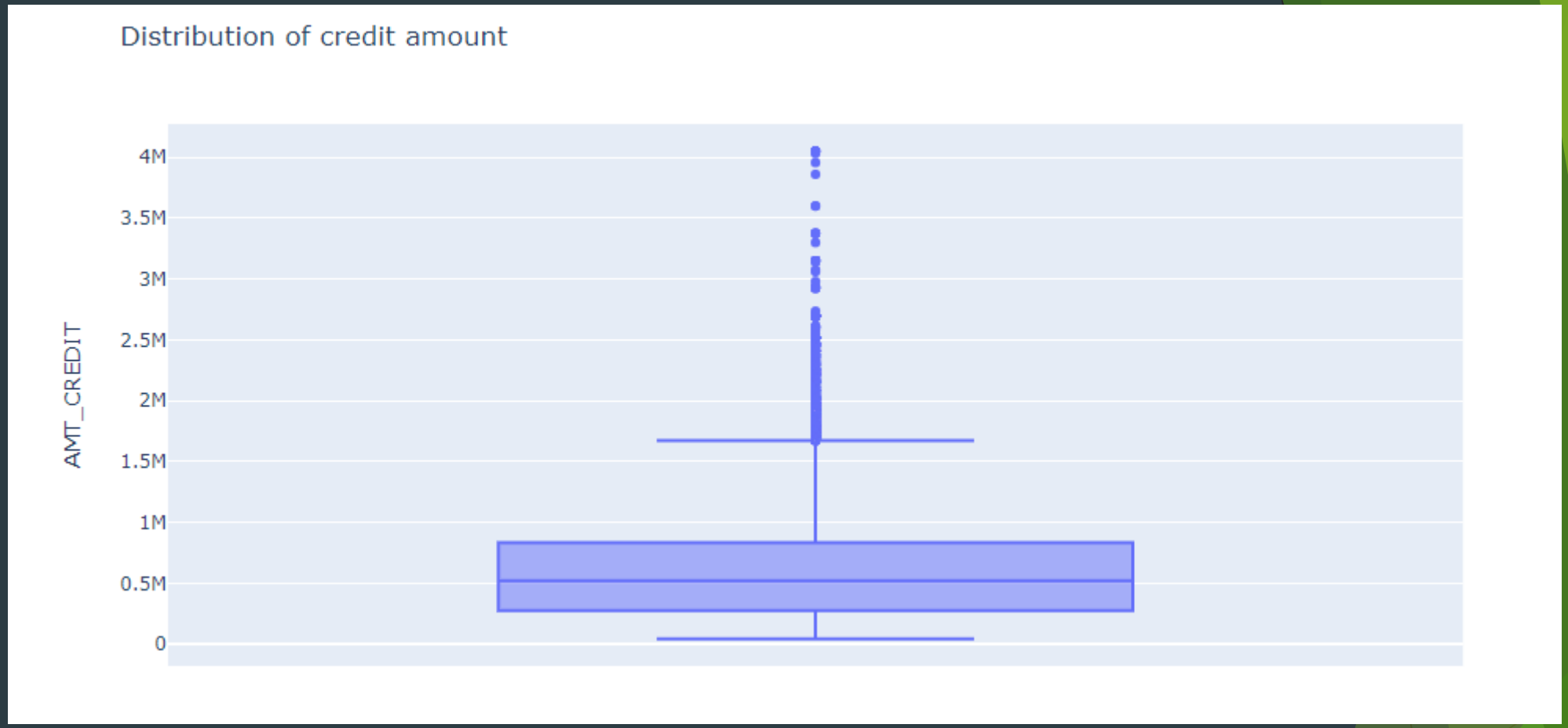
Univariate Analysis for Numerical Variables for Others



Insights:

There are outliers in total income.

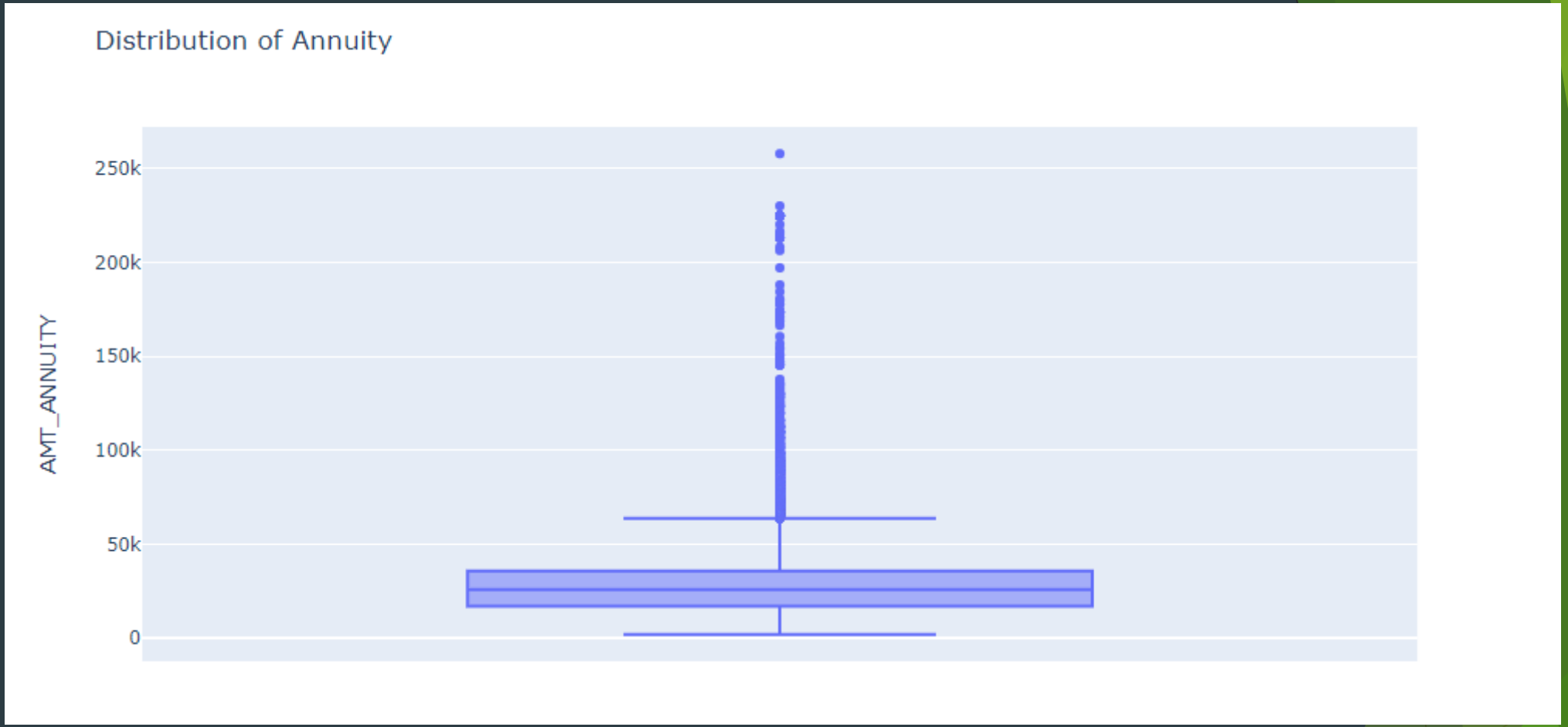
The third quartiles is very slim for income amount.



Insights:

There are outliers here too.

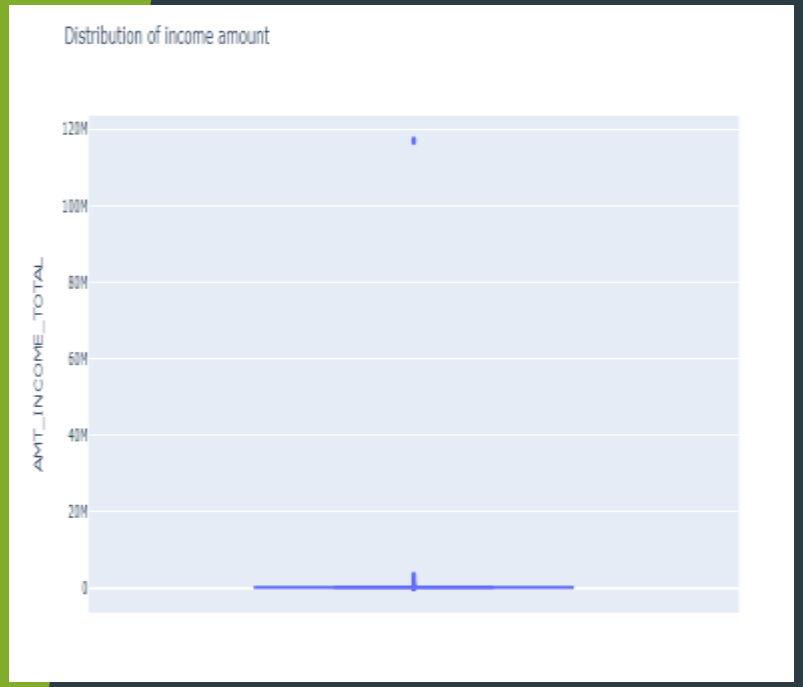
The first quartile is bigger than the third quartile, which indicates most of the credits of clients are present in the first quartile.



Insights:

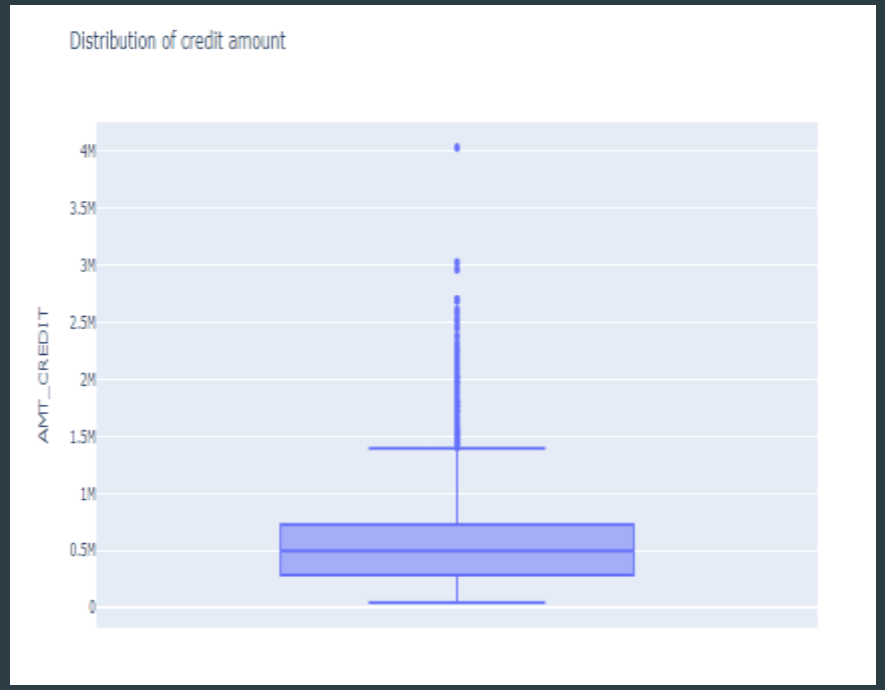
- There are outliers.
- Most of the clients are from Q1

Univariate Analysis for Numerical Variables for Defaulters



Insights:

- Few outliers can be noticed
- Q3 is slim and Q1 has most the clients



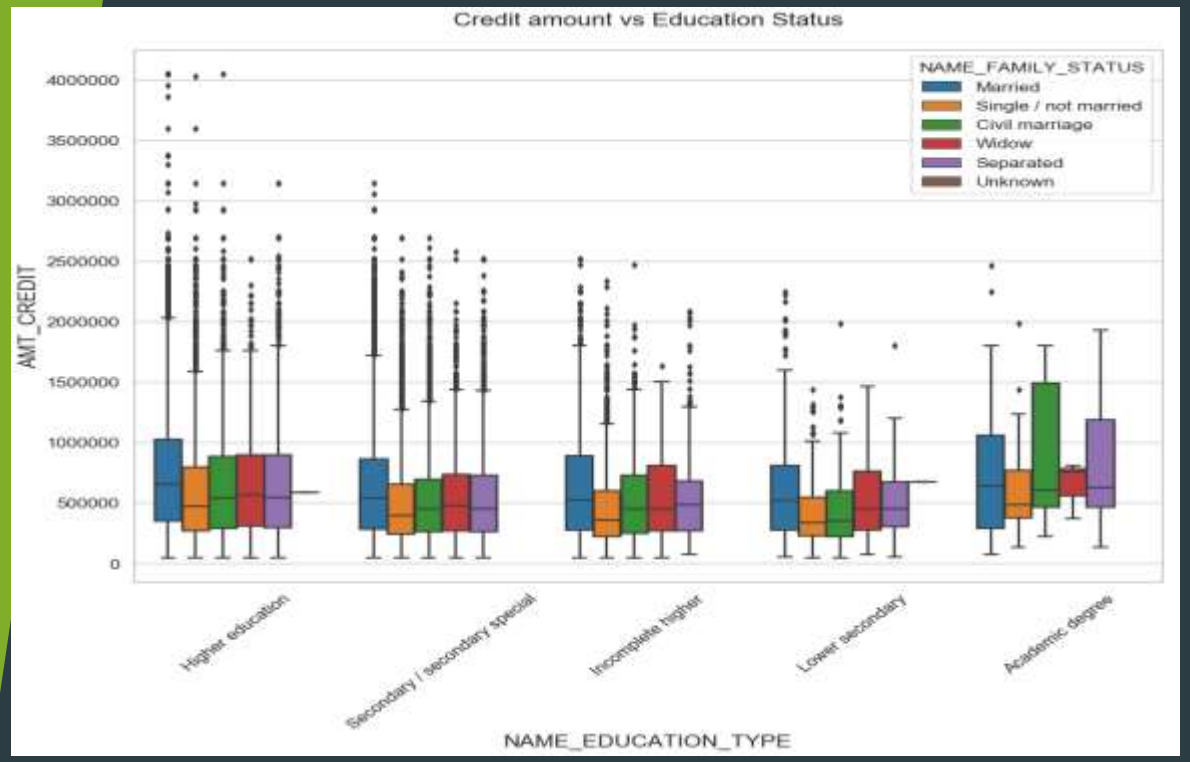
Insights:

- Few outliers
- Q1 is huge than Q3 i.e., it has more no of credits of clients

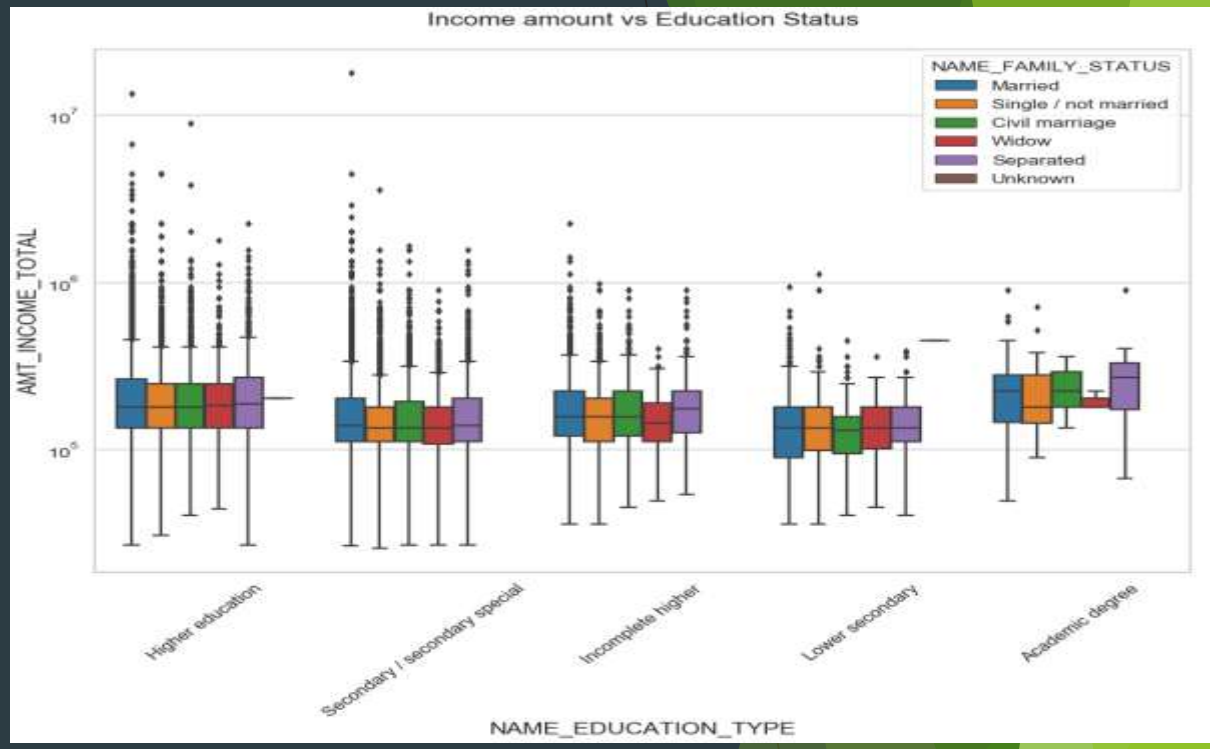
Insights:

- Few outliers
- Q1 is huge than Q3 i.e., it has more no of credits of clients

Bivariate Analysis for others

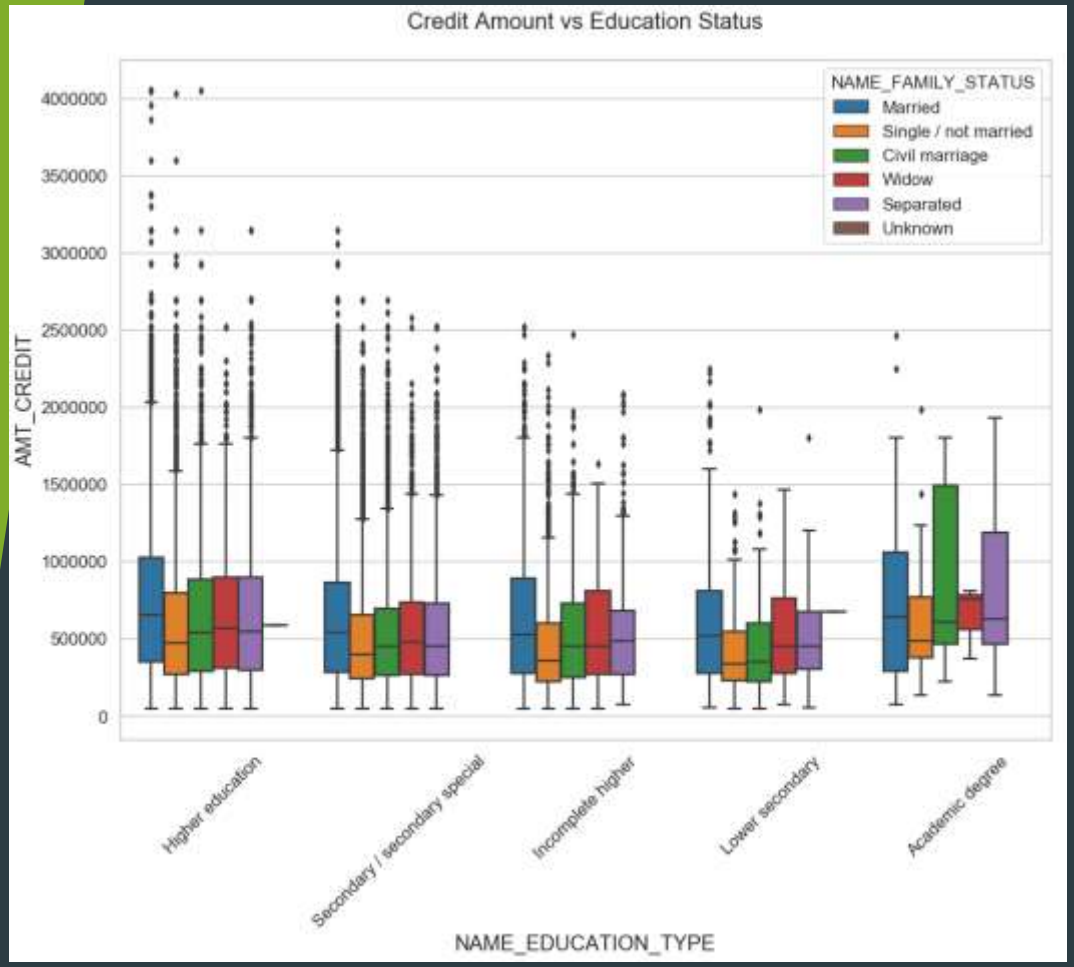


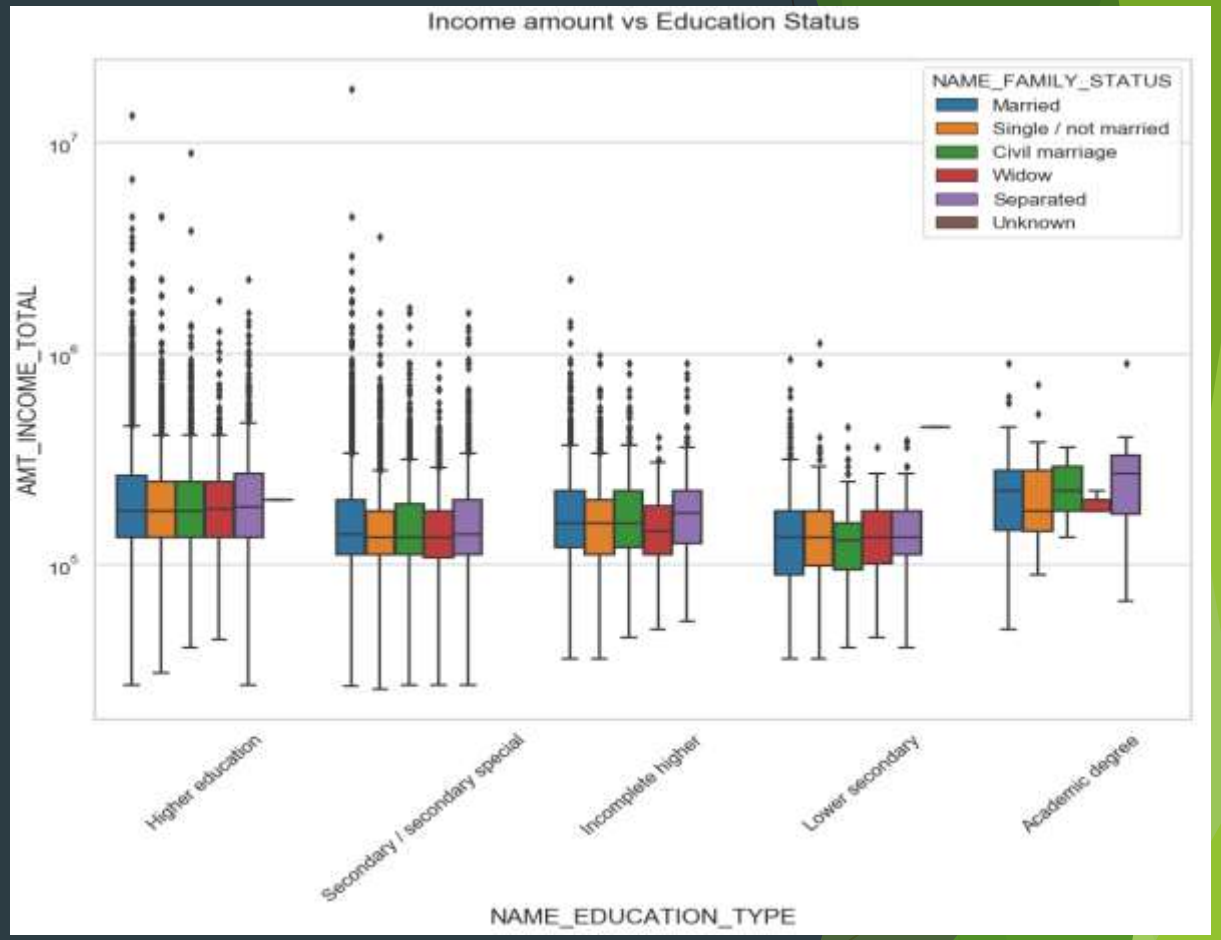
- we can say that Family status of civil marriage, separated and 'marriage.
- Academic degree education have more number of credits than others.
- Also, higher education of family status of marriage, single and civil marriage have more outliers.
- Civil marriage for Academic degree has most of the credits in the Q3.



- From above figure, its seen that for Higher education, the income amount is mostly equal with family status.
- with many outliers.
- Less outlier are for Academic degree but, the income amount is little higher than it is for Higher education.
- Lower secondary of civil marriage group have less income amount than that of others.

Bivariate Analysis for Defaulters





In both cases the insights are similar to as in others

CONCLUSION

- The contract type ‘Student’ ,’pensioner’ and ‘Businessman’ with housing ‘type other than ‘Co-op apartment’ have successful payments. Therefore, these may not be the loan defaulters
- However, clients with income type of ‘Working’ have maximum number of unsuccessful payments. Therefore they can be loan defaulters.
- Also, clients with purpose as ‘Repair’ have more number of unsuccessful payments on time. Therefore they can be loan defaulters.
- Clients with housing type as ‘With parents’ have least number of unsuccessful payments. These can be the safest category to approve loans.

Thank You