# CREDIT EDA CASE STUDY

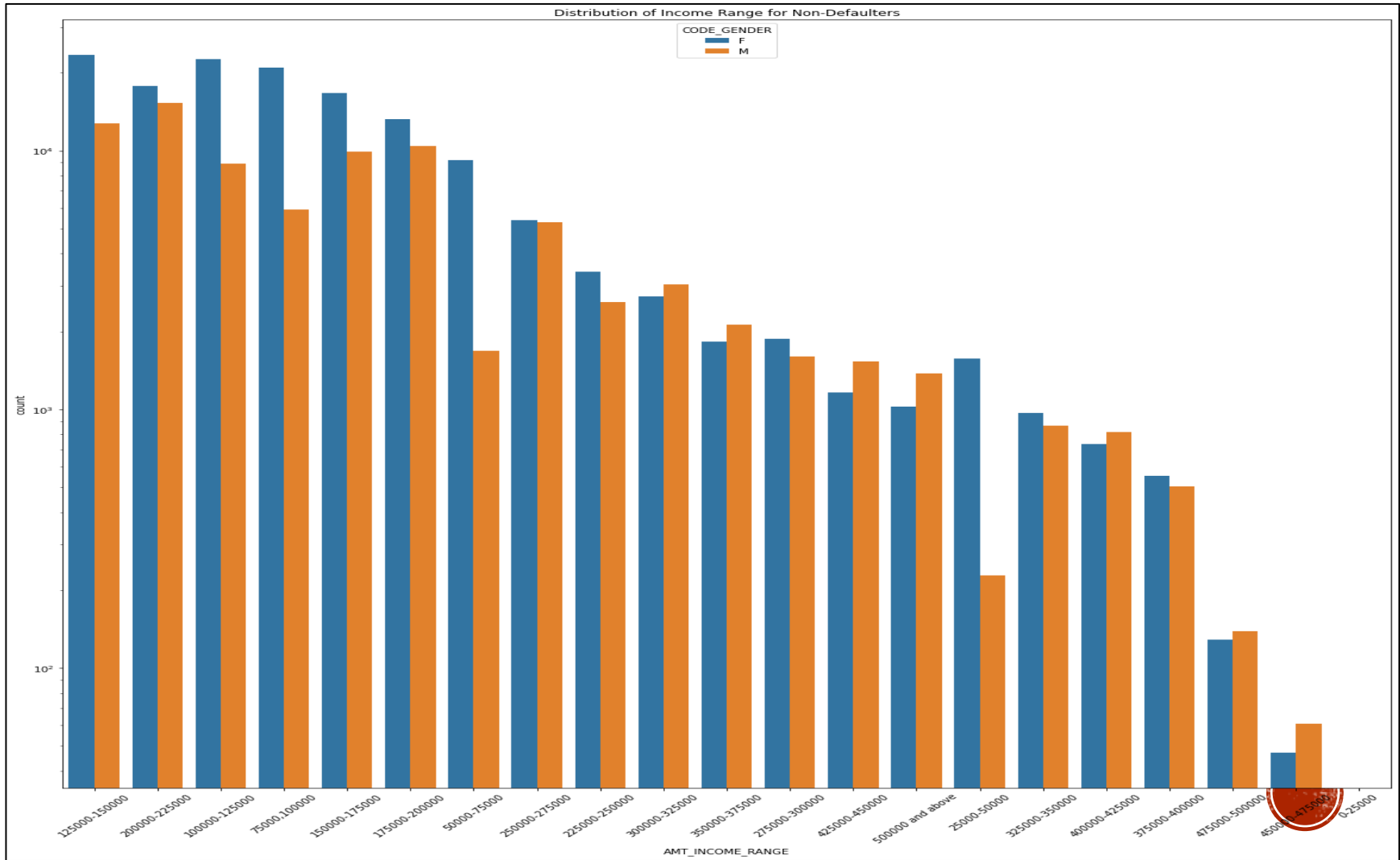
By Chinmaya Prusty & Baibaswat Bose

# CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET_0 (NON-DEFAULTERS)

## Distribution of Income Range for Non Defaulters

**Points to be concluded from the graph :**

- Female counts are higher than male.
- Income range from 100000 to 200000 is having more number of credits.
- Very less count for income range 400000 and above.

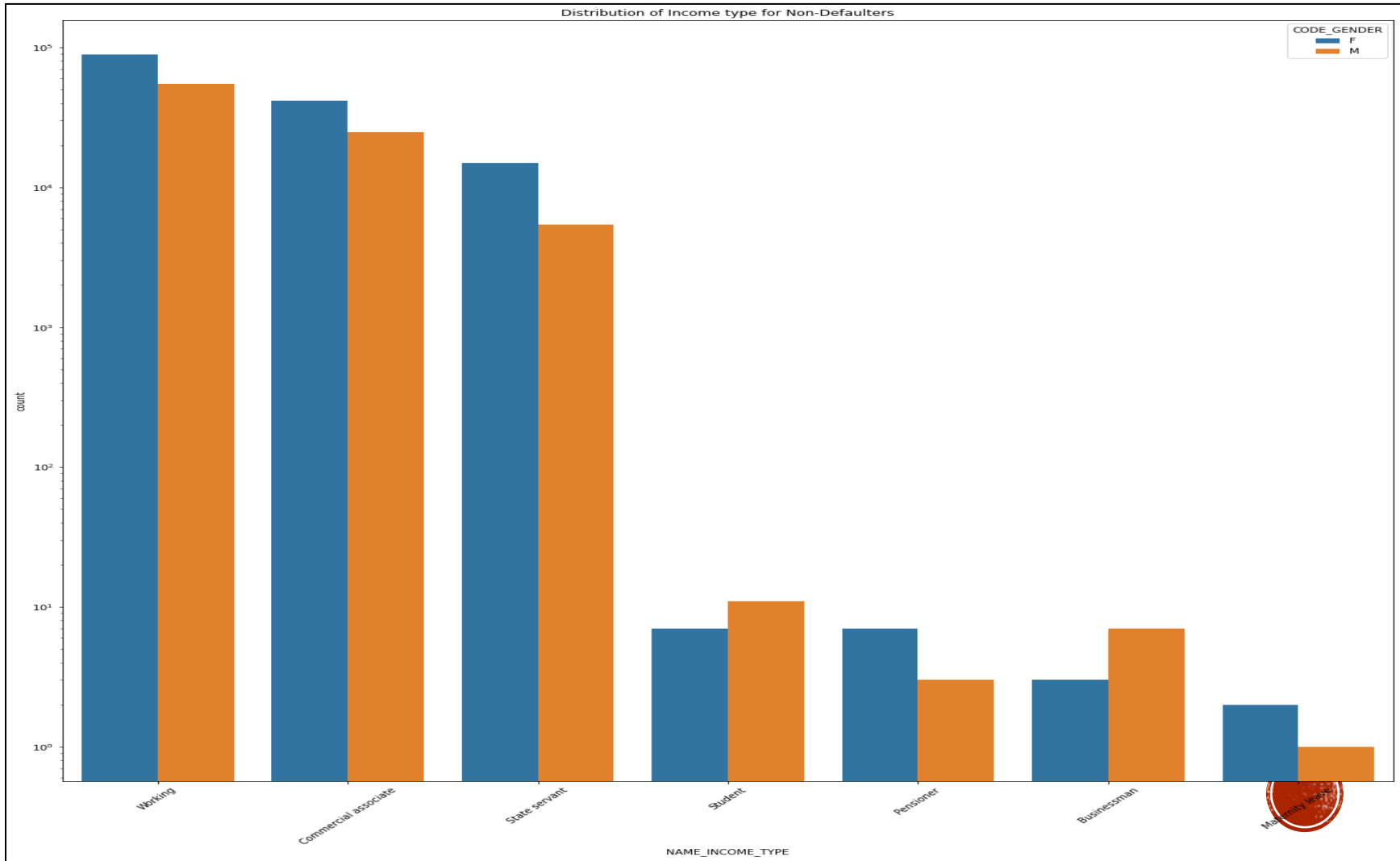


# CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET_0 (NON-DEFAULTERS)

## Distribution of Income Type for Non Defaulters

**Points to be concluded from the graph :**

- Type ‘working’, ’commercial associate’, and ‘State Servant’ the number of credits are higher than others.
- For this Females are having more number of credits than male.

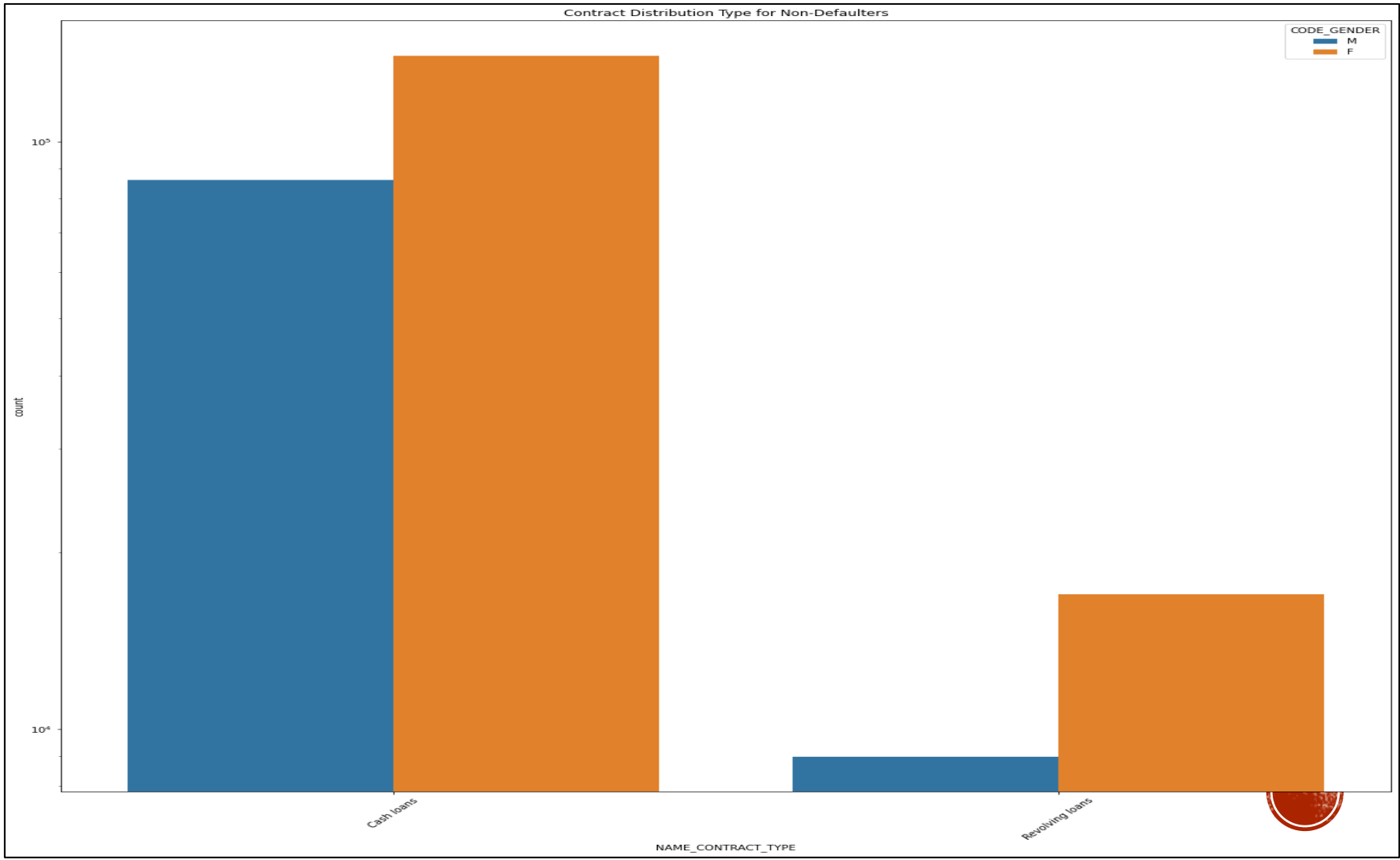


# CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET_0 (NON-DEFAULTERS)

## Contract Distribution Type for Non- Defaulters

**Points to be concluded from the graph :**

- 'Cash loans' is having higher number of credits than 'Revolving loans' contract type.
- For this also Female is having more credits.

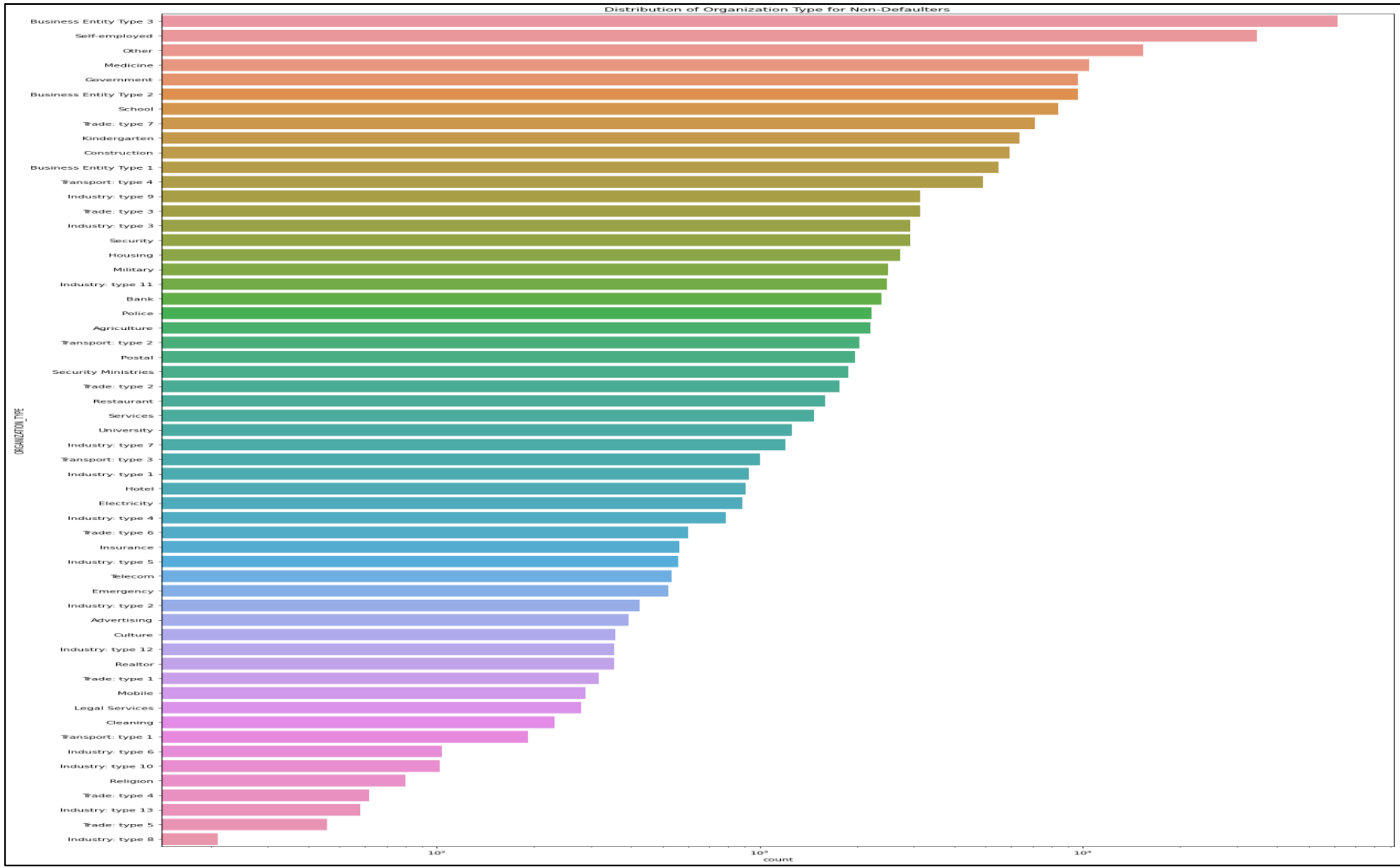


# CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET_0 (NON-DEFAULTERS)

## Distribution of Organization Type for Non- Defaulters

**Points to be concluded from the graph :**

- Clients which have applied for credits are from most of the organization type ‘Business entity Type 3’ , ‘Self employed’ , ‘Other’ , as compared to rest of the Organization Types

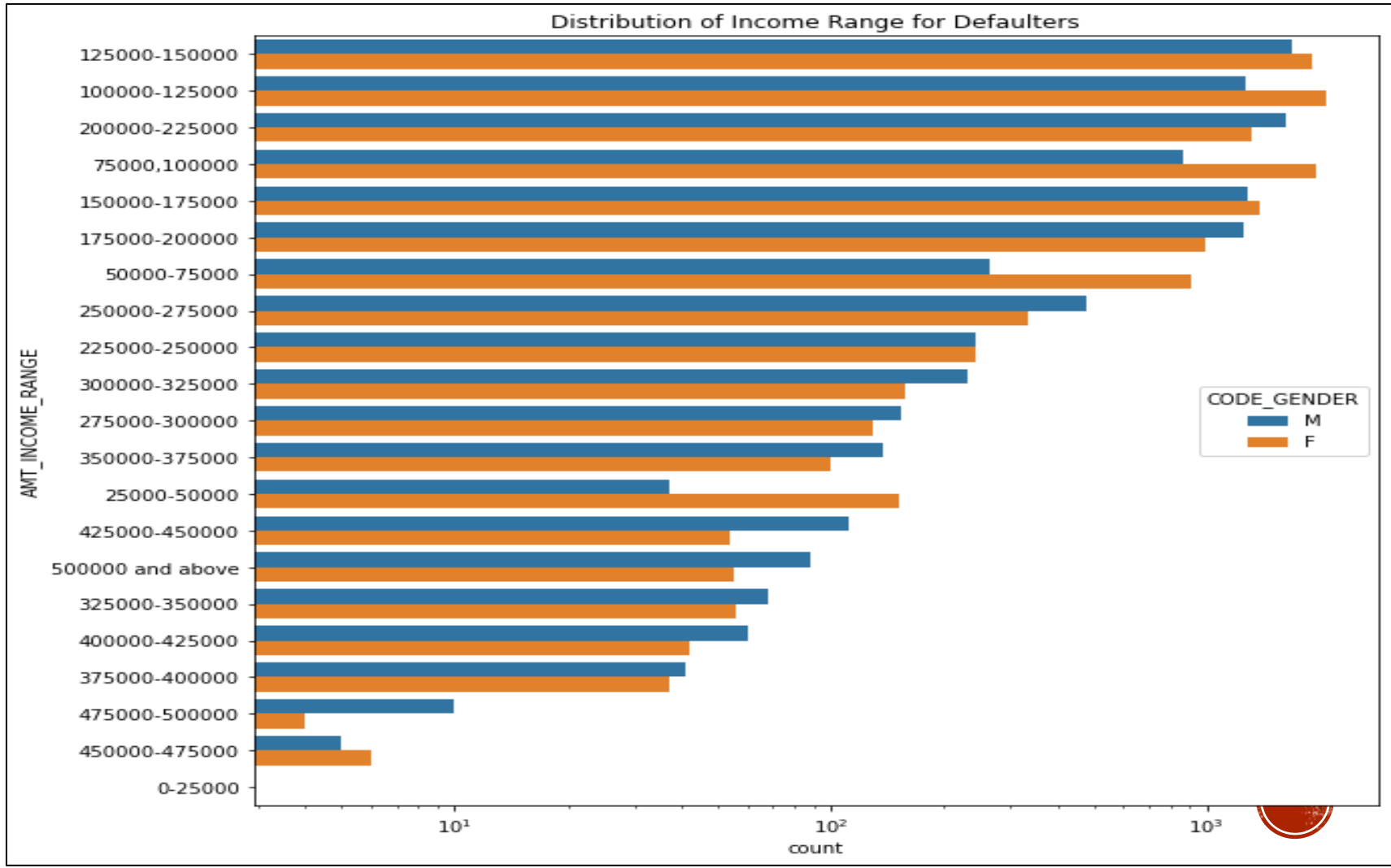


# CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET_1 (DEFAULTERS)

## Distribution of Income Range for Defaulters

**Points to be concluded from the graph :**

- Male counts are higher than female.
- This graph show that males are more than female in having credits for that range.
- Income range from 100000 to 200000 is having more number of credits.

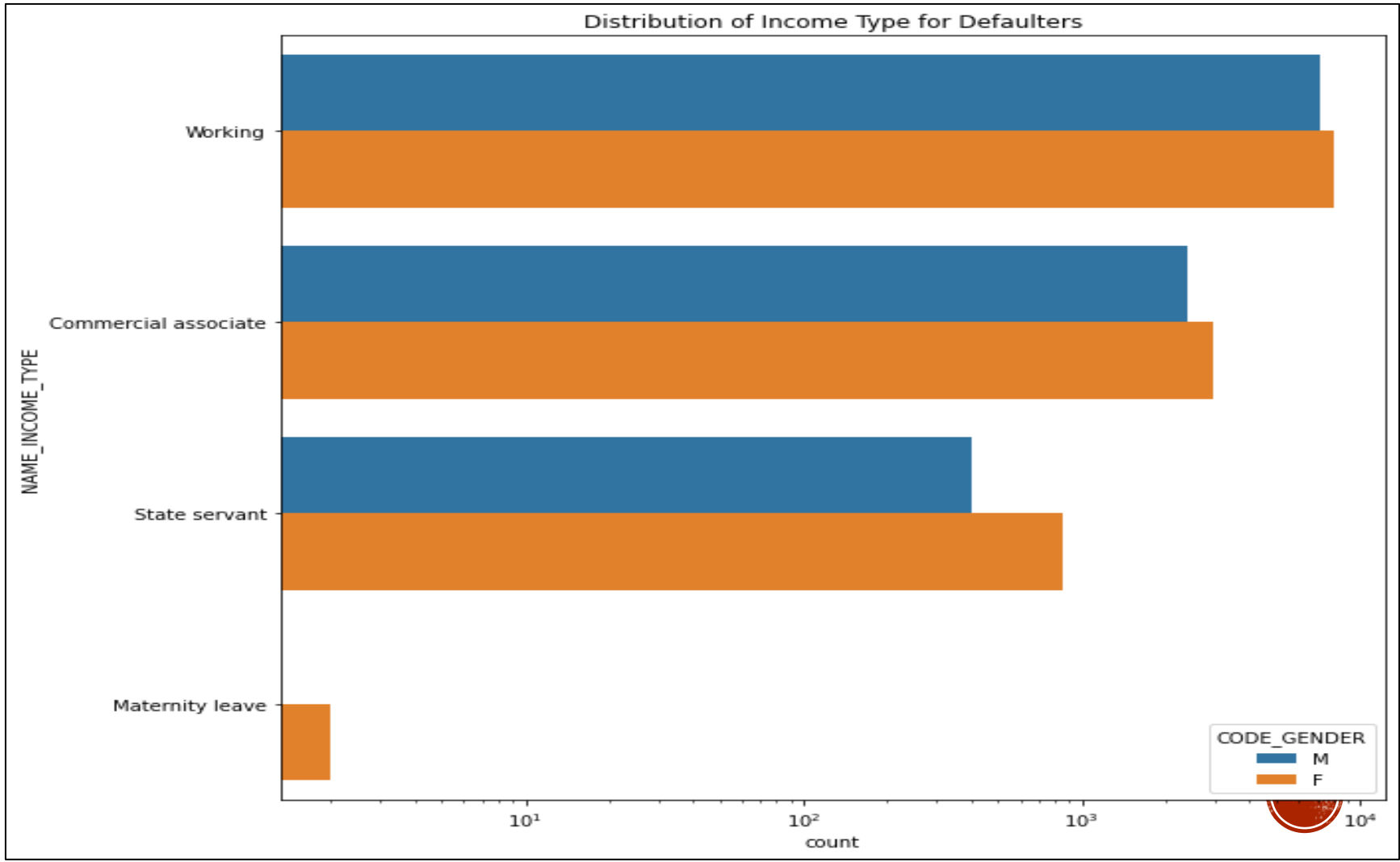


# CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET_1 (DEFAULTERS)

## Distribution of Income Type for Defaulters

**Points to be concluded from the graph :**

- For income type ‘working’, ’commercial associate’, and ‘State Servant’ the number of credits are higher than other i.e. ‘Maternity leave.
- For type 1: There is no income type for ‘student’ , ’pensioner’ and ‘Businessman’ which means they don’t do any late payments.
- For this Females are having more number of credits than male.
- Less number of credits for income type ‘Maternity leave’.

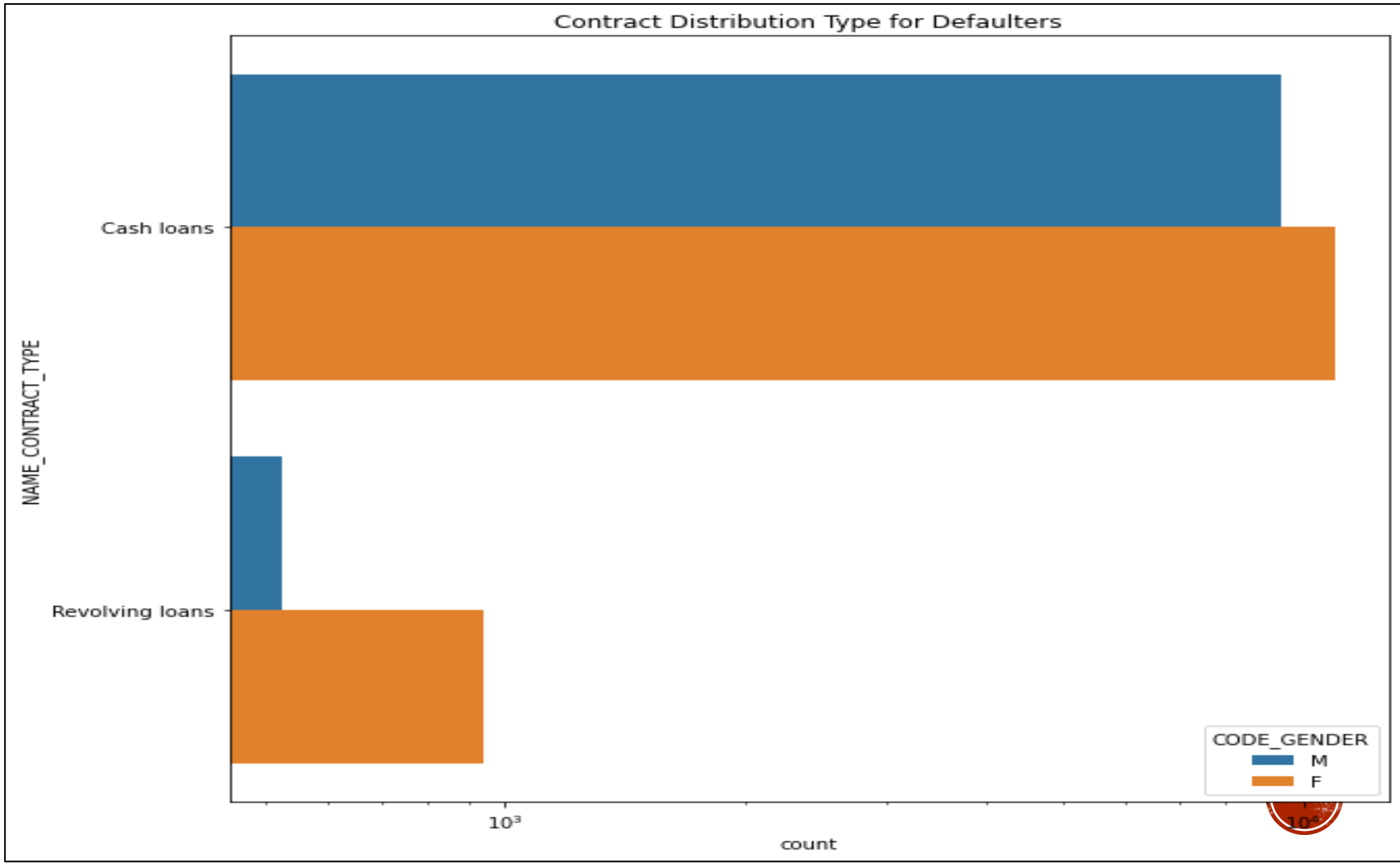


# CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET_1 (DEFAULTERS)

## Contract Distribution Type for Defaulters

**Points to be concluded from the graph :**

- For contract type ‘cash loans’ is having higher number of credits than ‘Revolving loans’ contract type.
- For this also Female is leading for applying credits.

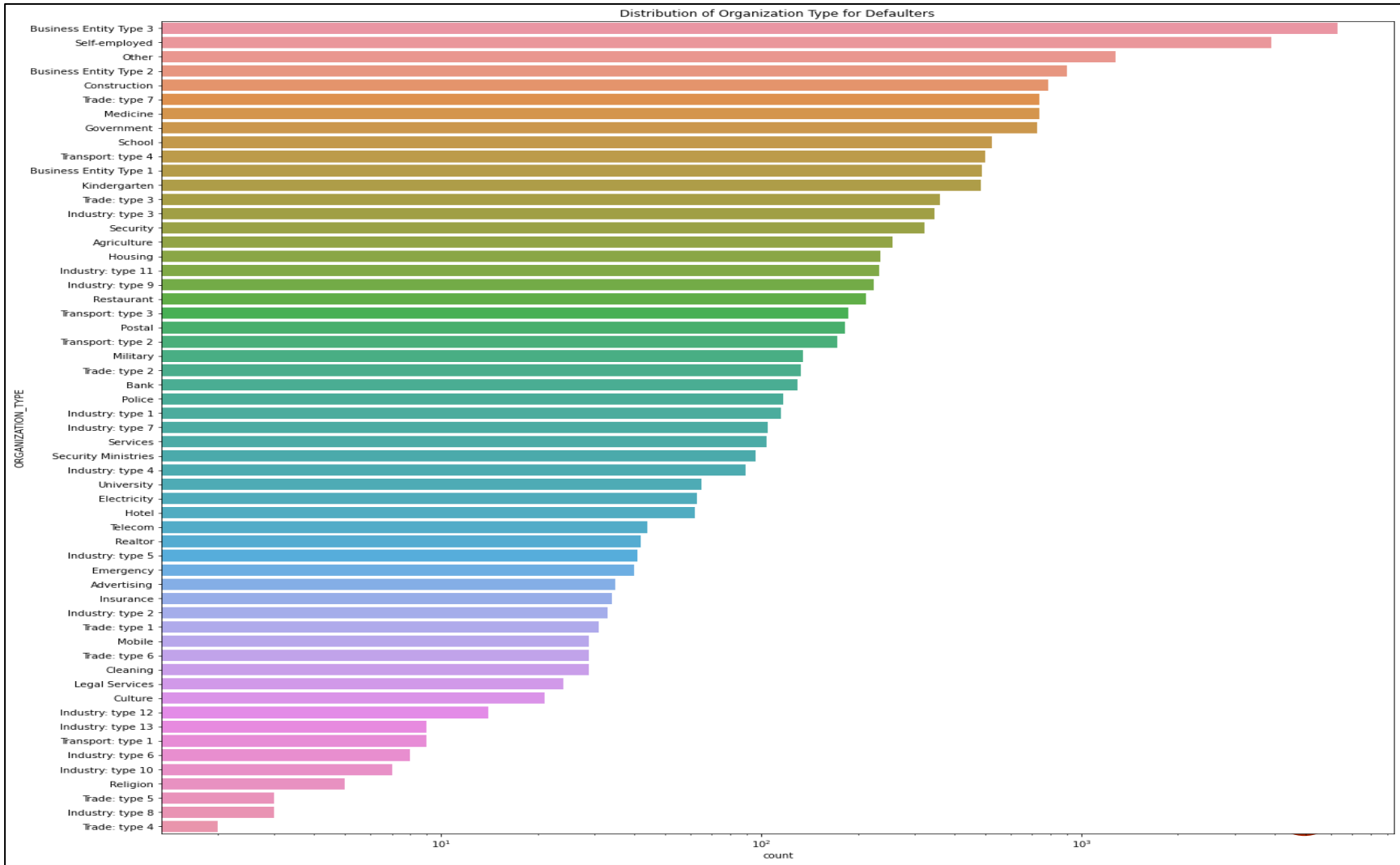


# CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET_1 (DEFAULTERS)

## Distribution of Organization Type for Defaulters

**Points to be concluded from the graph :**

- Clients which have applied for credits are from most of the organization type ‘Business entity Type 3’ , ‘Self employed’ , ‘Other’ , ‘Medicine’ and ‘Government’.
- Less clients are from Industry type 8,type 6, type 10, religion and trade type 5, type 4.

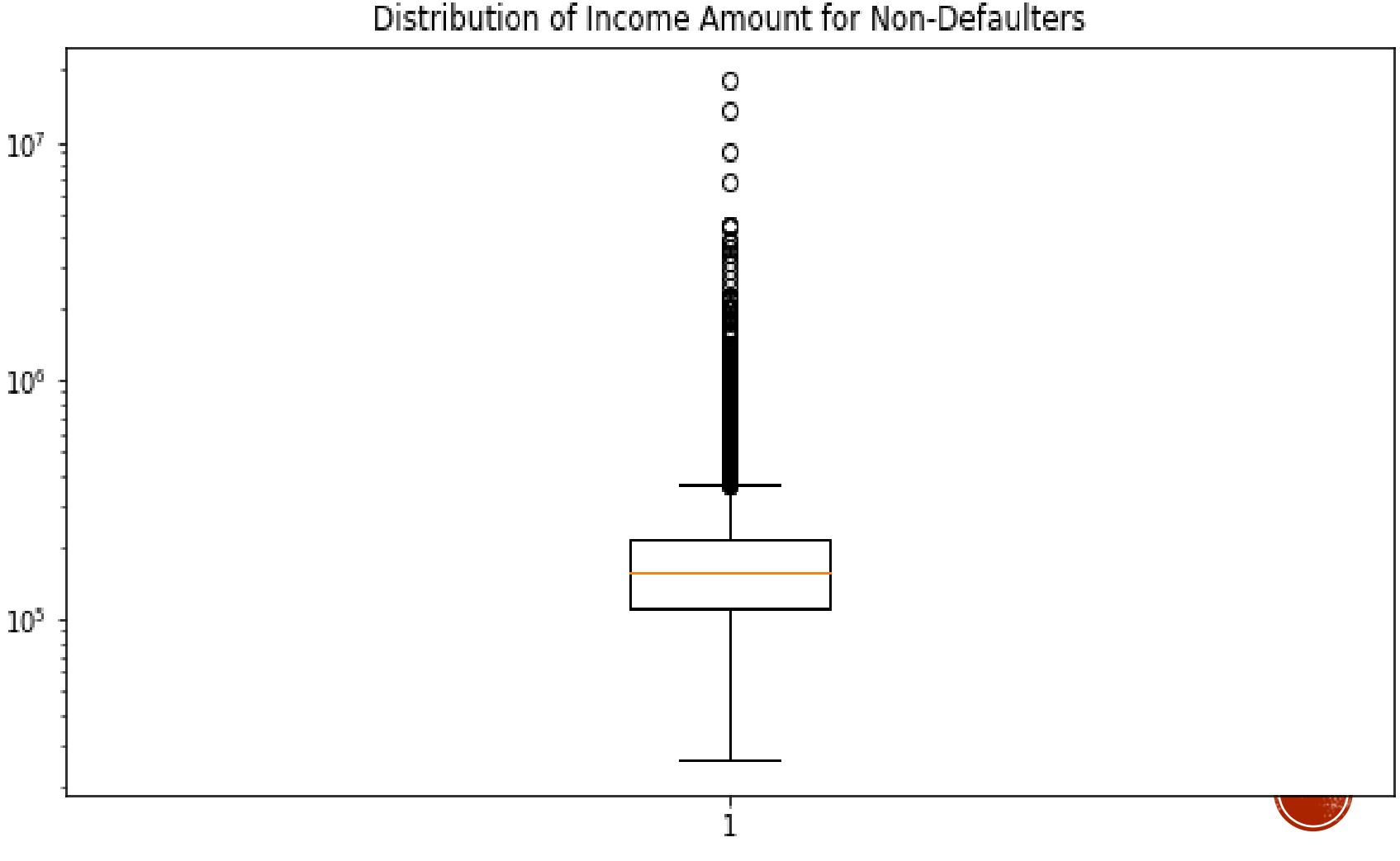


# UNIVARIATE ANALYSIS FOR VARIABLES ON TARGET_0 (NON-DEFAULTERS)

## Boxplot of Income Amount for Non-Defaulters

**Points to be concluded from the graph :**

- Outliers are noticed in income amount.
- The third quartiles is very slim for income amount.

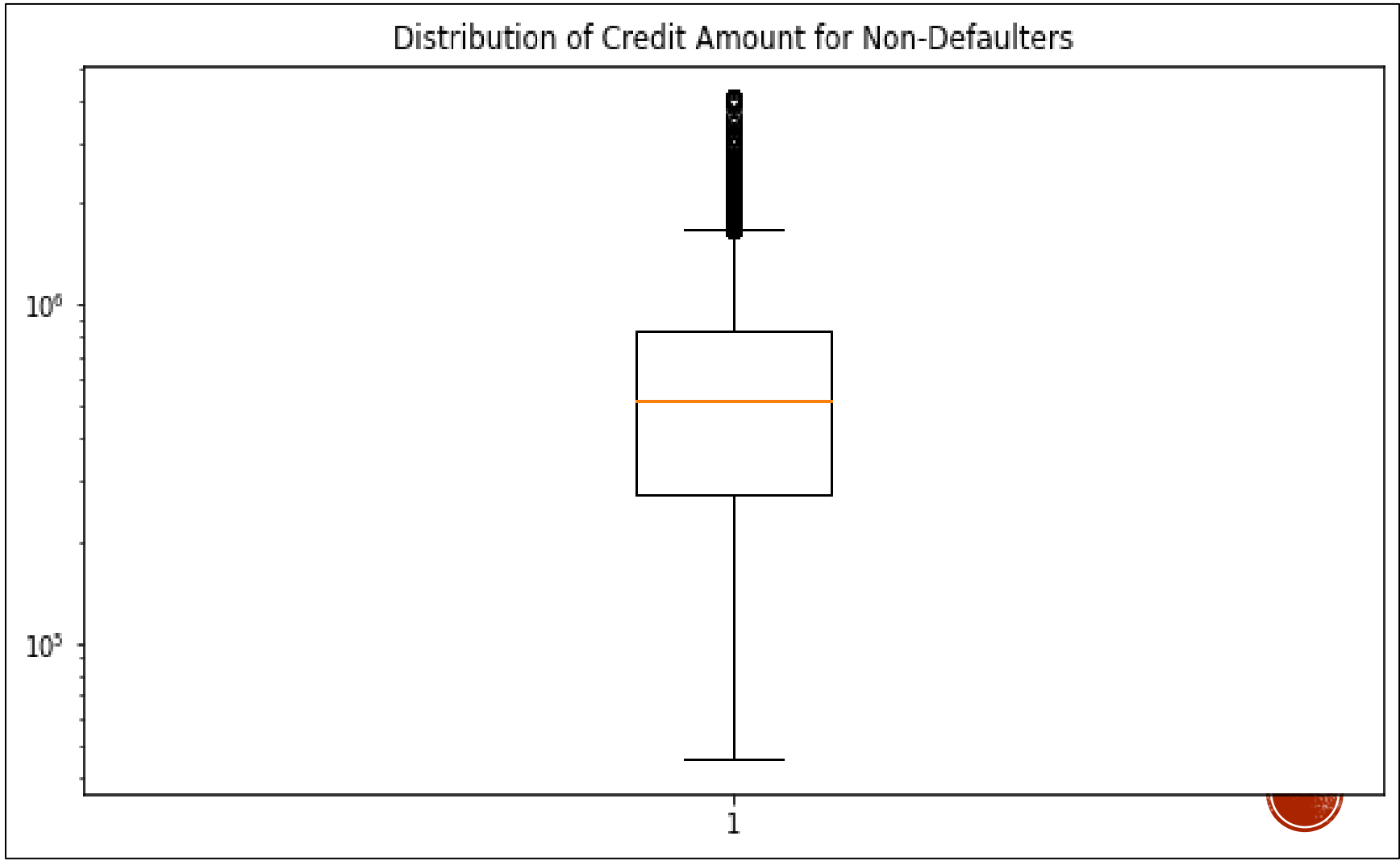


# UNIVARIATE ANALYSIS FOR VARIABLES ON TARGET_0 (NON-DEFAULTERS)

## Boxplot of Credit Amount for Non-Defaulters

**Points to be concluded from the graph :**

- Outliers are noticed in credit amount.
- The first quartile is bigger than third quartile for credit amount which means most of the credits of clients are present in the first quartile.

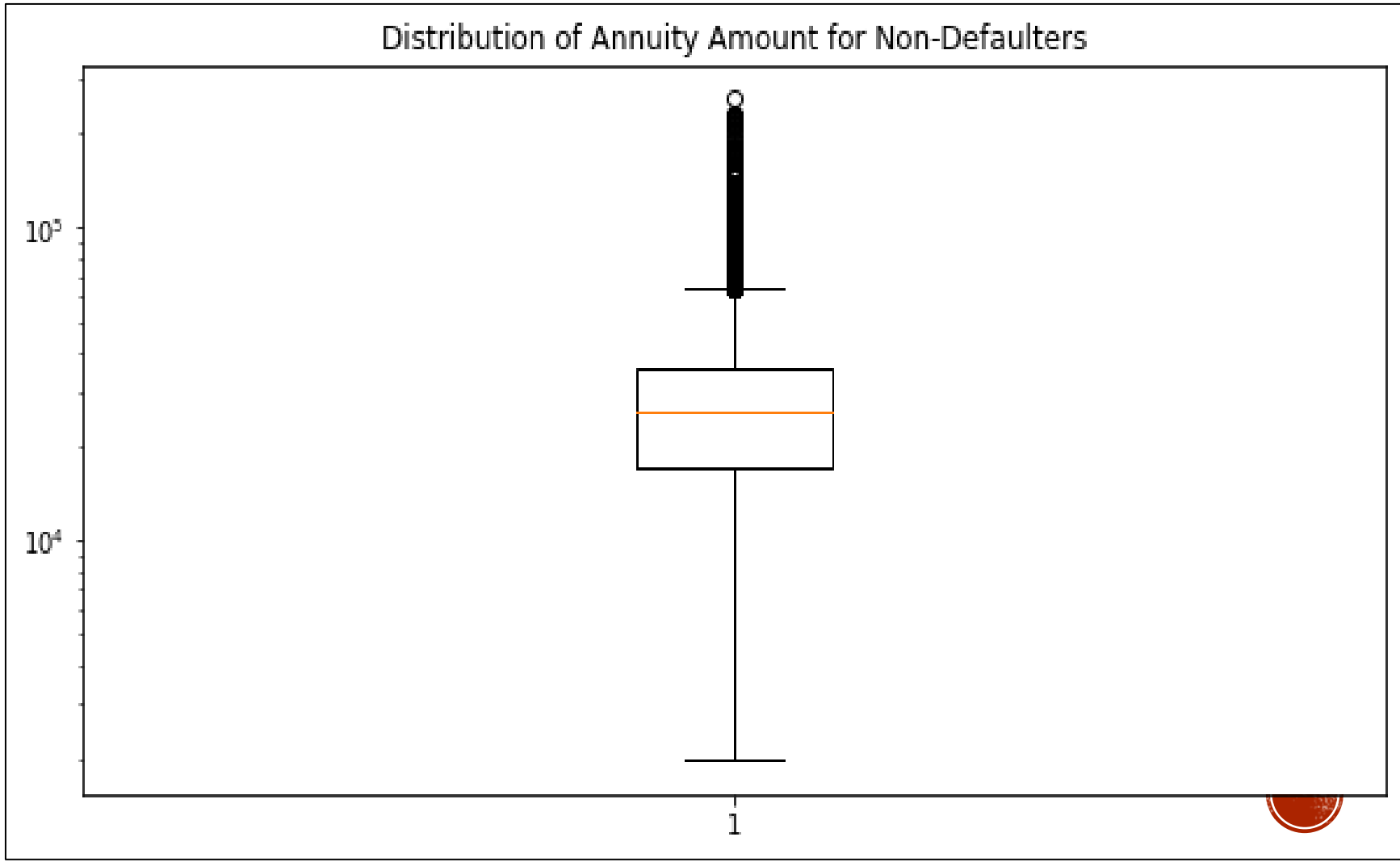


# UNIVARIATE ANALYSIS FOR VARIABLES ON TARGET_0 (NON-DEFAULTERS)

## Boxplot of Annuity Amount for Non-Defaulters

**Points to be concluded from the graph :**

- Outliers are noticed in annuity amount.
- The first quartile is bigger than third quartile for annuity amount which means most of the annuity clients are from first quartile

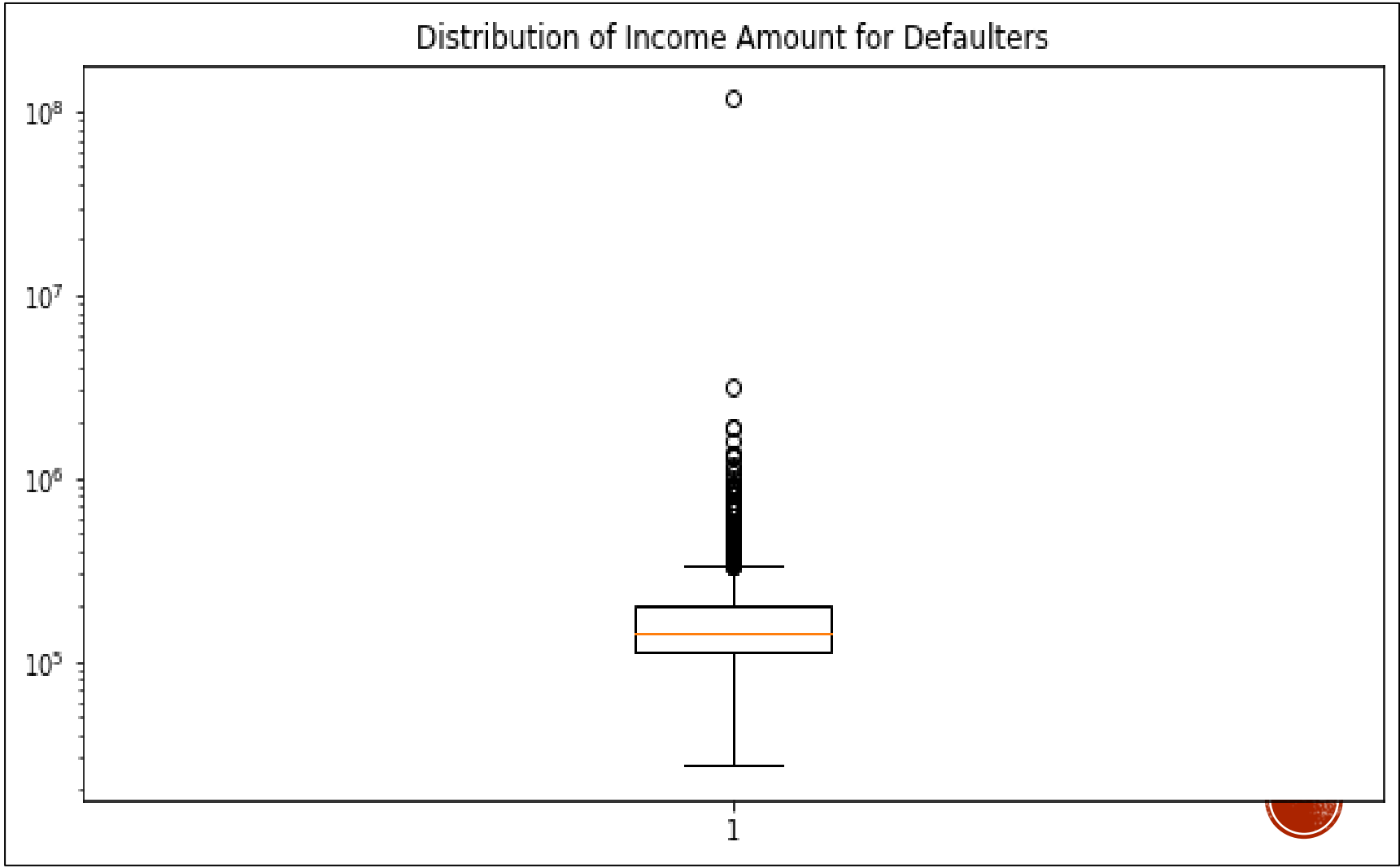


# UNIVARIATE ANALYSIS FOR VARIABLES TARGET_1 (DEFAULTERS)

## Boxplot of Income Amount for Defaulters

**Points to be concluded from the graph :**

- Outliers are noticed in income amount.
- The third quartiles is very slim for income amount.
- Most of the clients of income are present in first quartile.

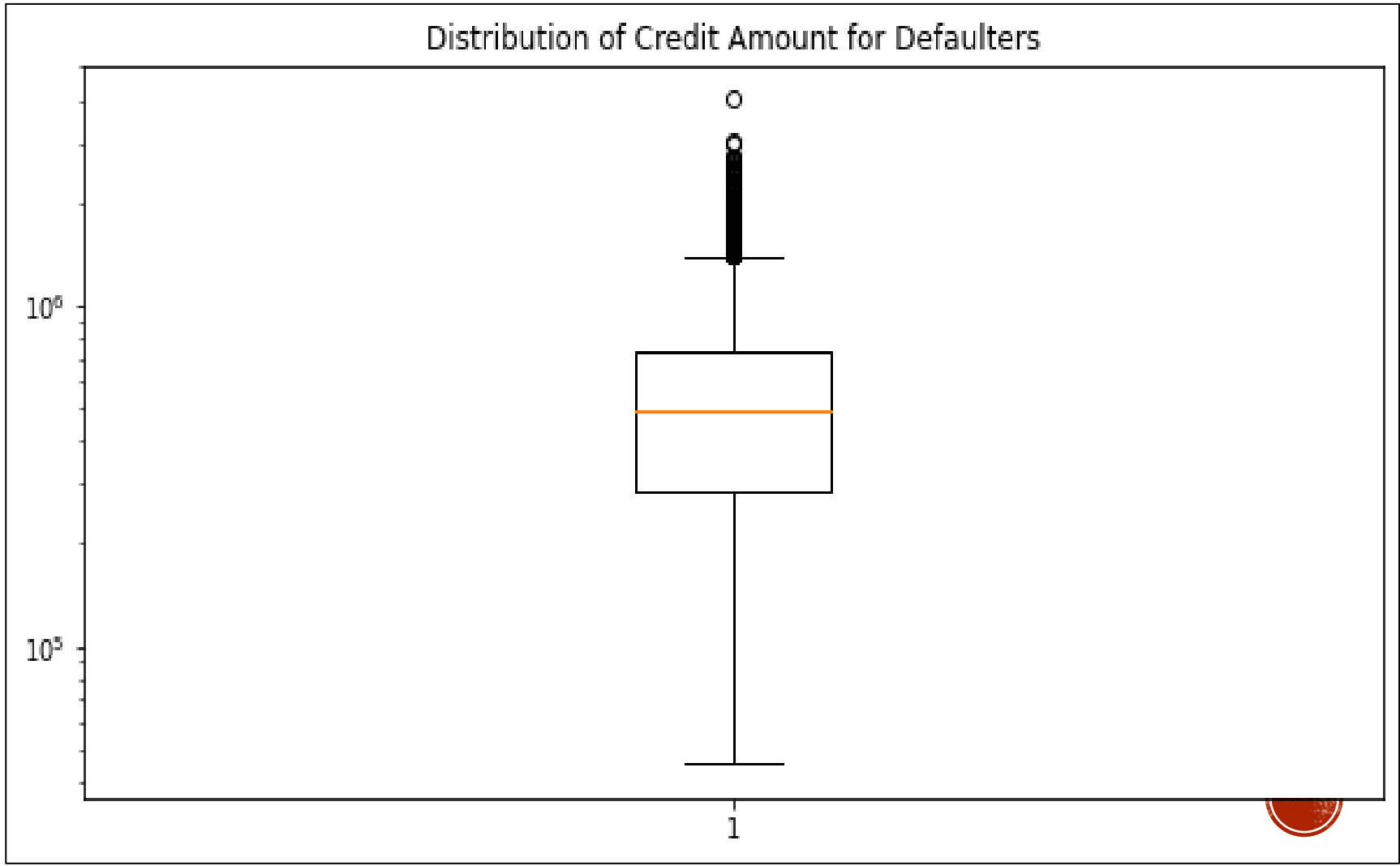


# UNIVARIATE ANALYSIS FOR VARIABLES TARGET_1 (DEFAULTERS)

## Boxplot of Credit Amount for Defaulters

**Points to be concluded from the graph :**

- Some outliers are noticed in credit amount.
- The first quartile is bigger than third quartile for credit amount which means most of the credits of clients are present in the first quartile.

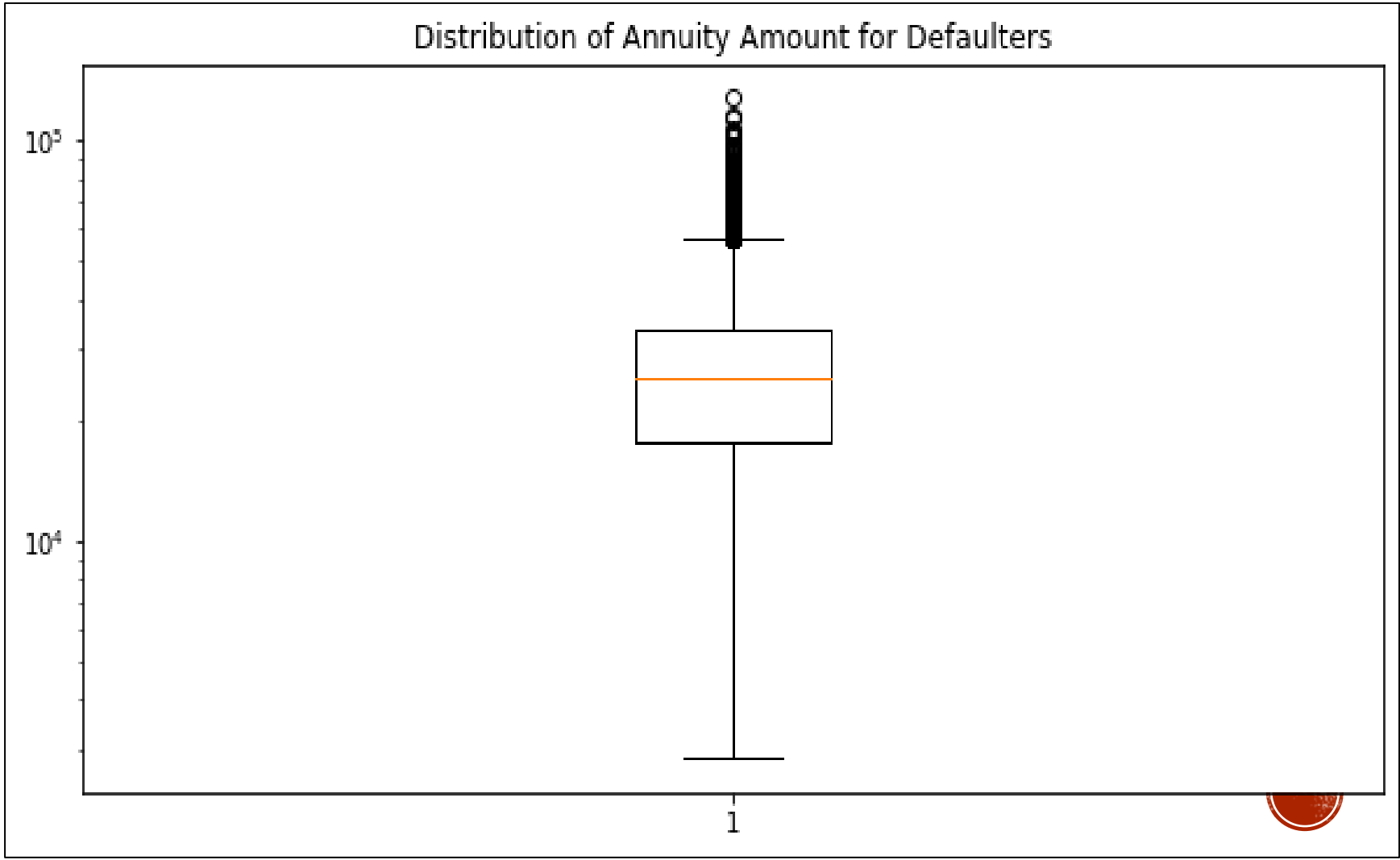


# UNIVARIATE ANALYSIS FOR VARIABLES TARGET_1 (DEFAULTERS)

## Boxplot of Annuity Amount for Defaulters

**Points to be concluded from the graph :**

- Some outliers are noticed in annuity amount.
- The first quartile is bigger than third quartile for annuity amount which means most of the annuity clients are from first quartile.

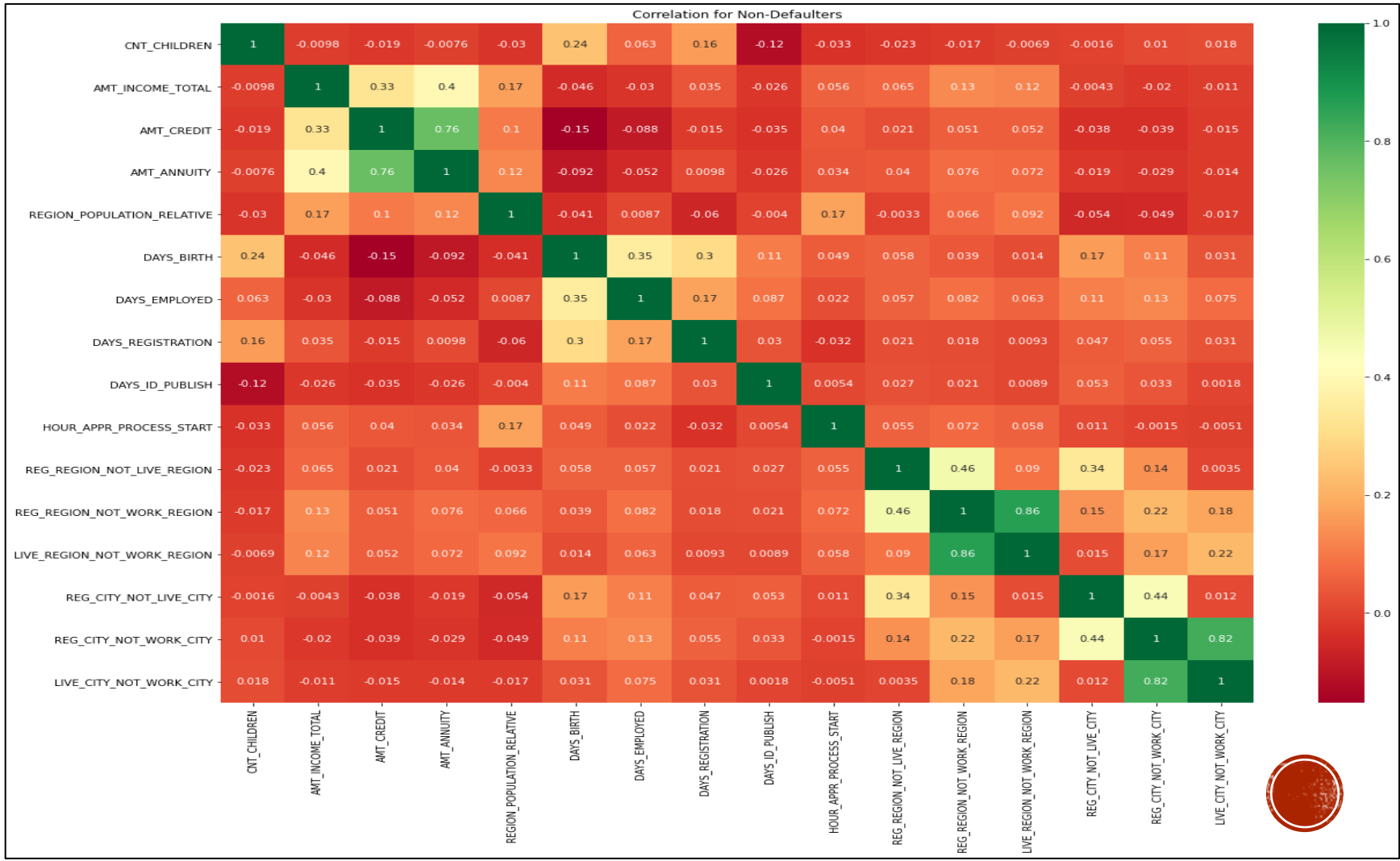


# CORRELATION OF TARGET_0 (NON-DEFAULTERS)

## Heat Map for Non-Defaulters

**Points to be concluded from the graph :**

- Credit amount is inversely proportional to the date of birth, which means Credit amount is higher for low age and vice-versa.
- Credit amount is inversely proportional to the number of children client have, means Credit amount is higher for less children count client have and vice-versa.
- Income amount is inversely proportional to the number of children client have, means more income for less children client have and vice-versa.
- Less children client have in densely populated area.
- Credit amount is higher to densely populated area.
- The income is also higher in densely populated area.



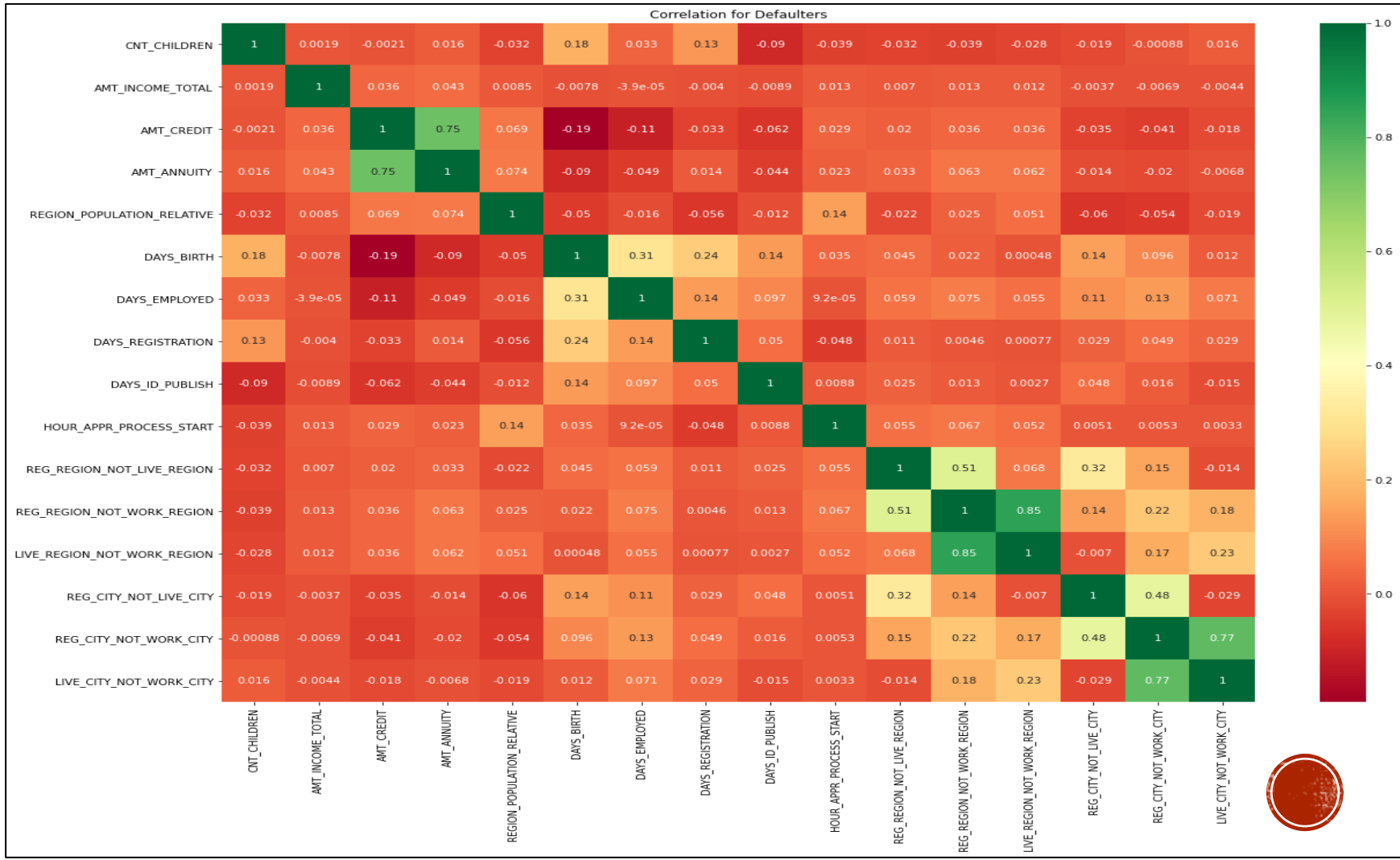
# CORRELATION OF TARGET_0 (DEFAULTERS)

## Heat Map for Defaulters

**Points to be concluded from the graph :**

This heat map for Target 1 is also having quite a same observation just like Target 0. But for few points are different. They are listed below.

- The client's permanent address does not match contact address are having less children and vice-versa
- The client's permanent address does not match work address are having less children and vice-versa

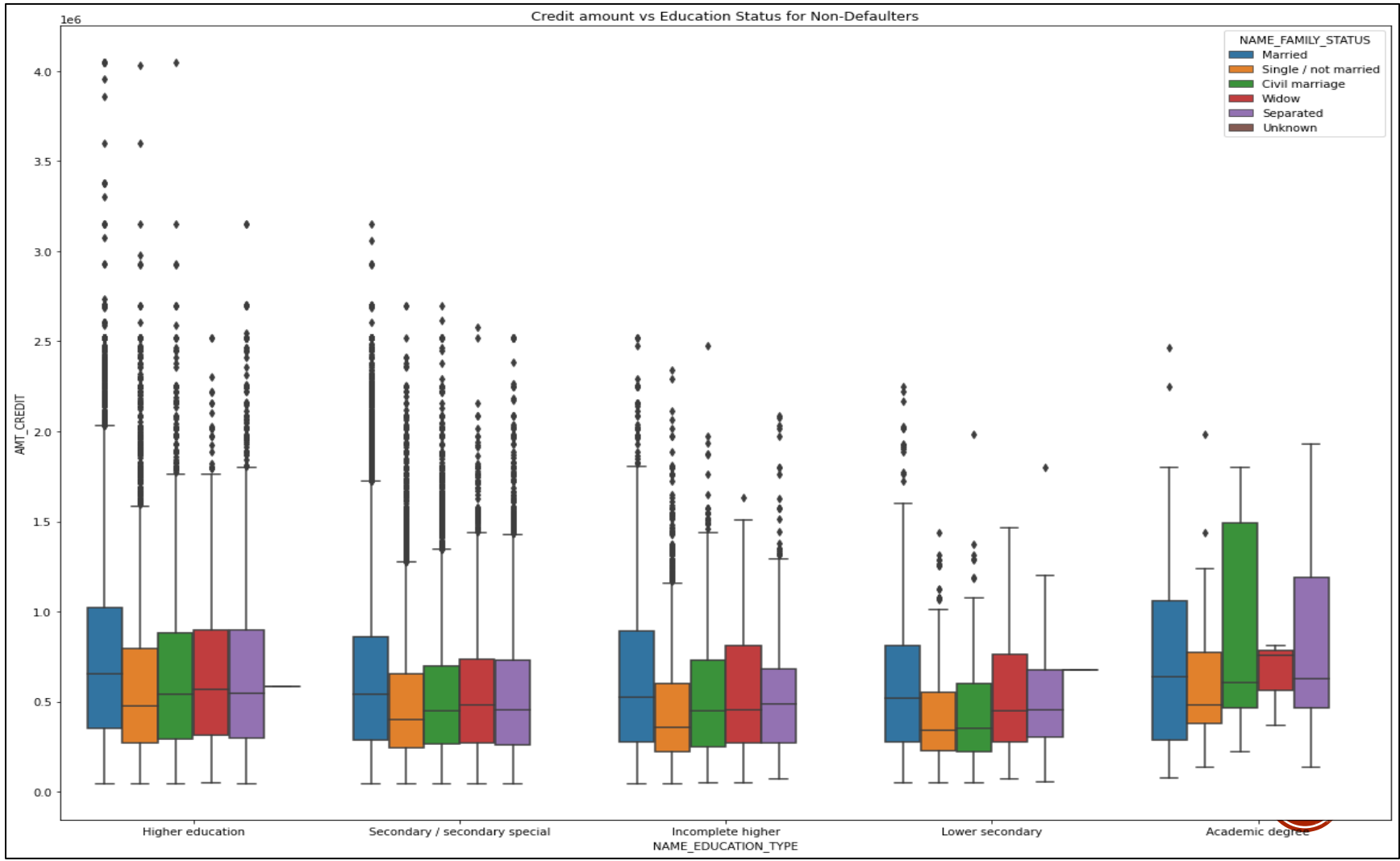


# BIVARIATE ANALYSIS FOR TYPE_0 (NON-DEFAULTERS)

## Credit Amount Vs Education Status for Non – Defaulters

**Points to be concluded from the graph :**

- 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.

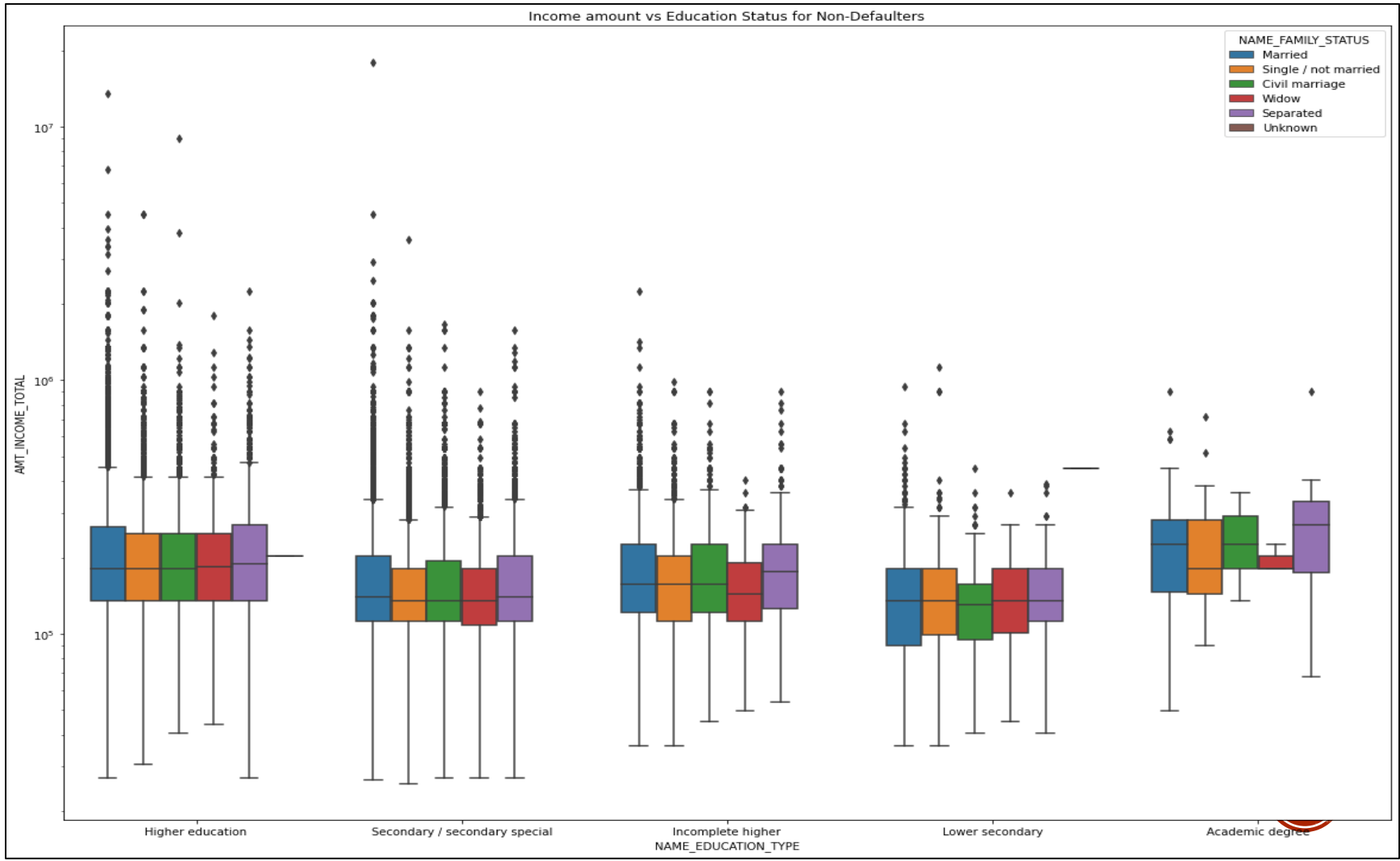


# BIVARIATE ANALYSIS FOR TYPE_0 (NON-DEFAULTERS)

## Income Amount Vs Education Status for Non – Defaulters

**Points to be concluded from the graph :**

- Education type 'Higher education' the income amount mean is mostly equal with family status. It does contain many outliers.
- Less outlier are having for Academic degree but they are having the income amount is little higher that Higher education.
- Lower secondary of civil marriage family status are have less income amount than others

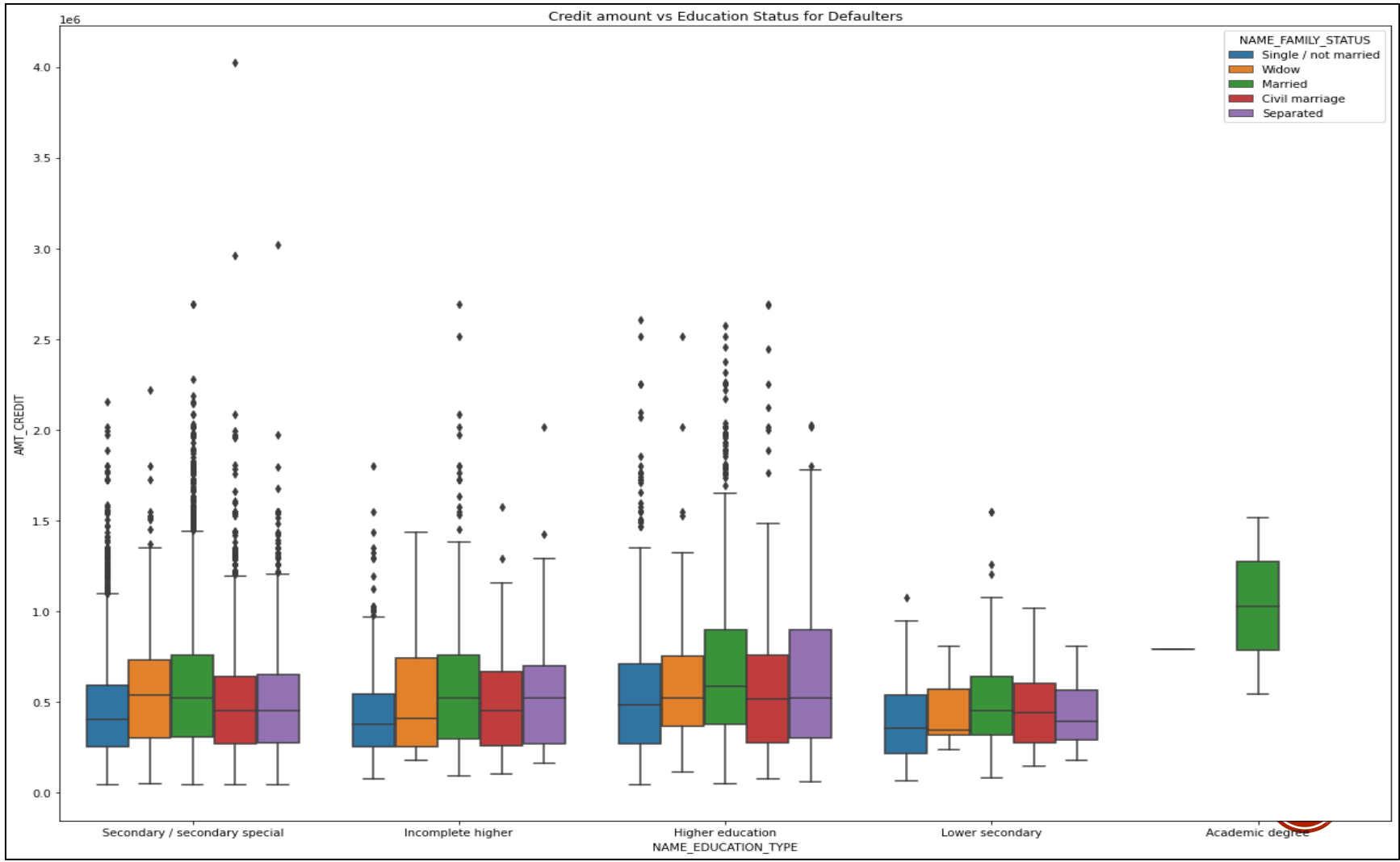


# BIVARIATE ANALYSIS FOR TYPE_1 (DEFAULTERS)

## Credit Amount Vs Education Status for Defaulters

**Points to be concluded from the graph :**

- Median of Academic Degree have higher credit amount than the rest.
- Outliers are present in the rest except Academic Degree..

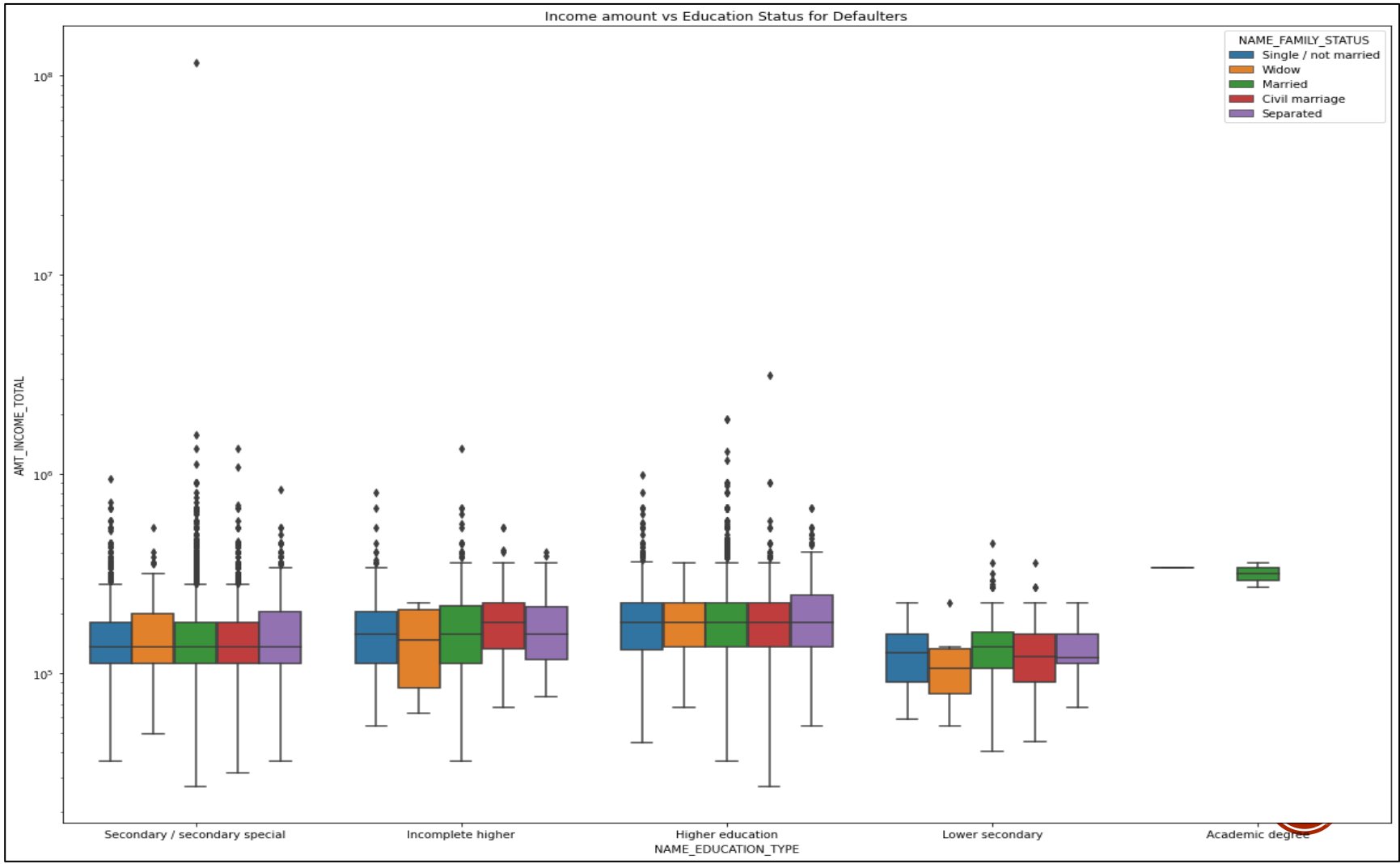


# BIVARIATE ANALYSIS FOR TYPE_1 (DEFAULTERS)

## Income Amount Vs Education Status for Defaulters

**Points to be concluded from the graph :**

- Academic Degree have Higher income amount in median compared to others.
- Outliers are present in the rest except Academic Degree.

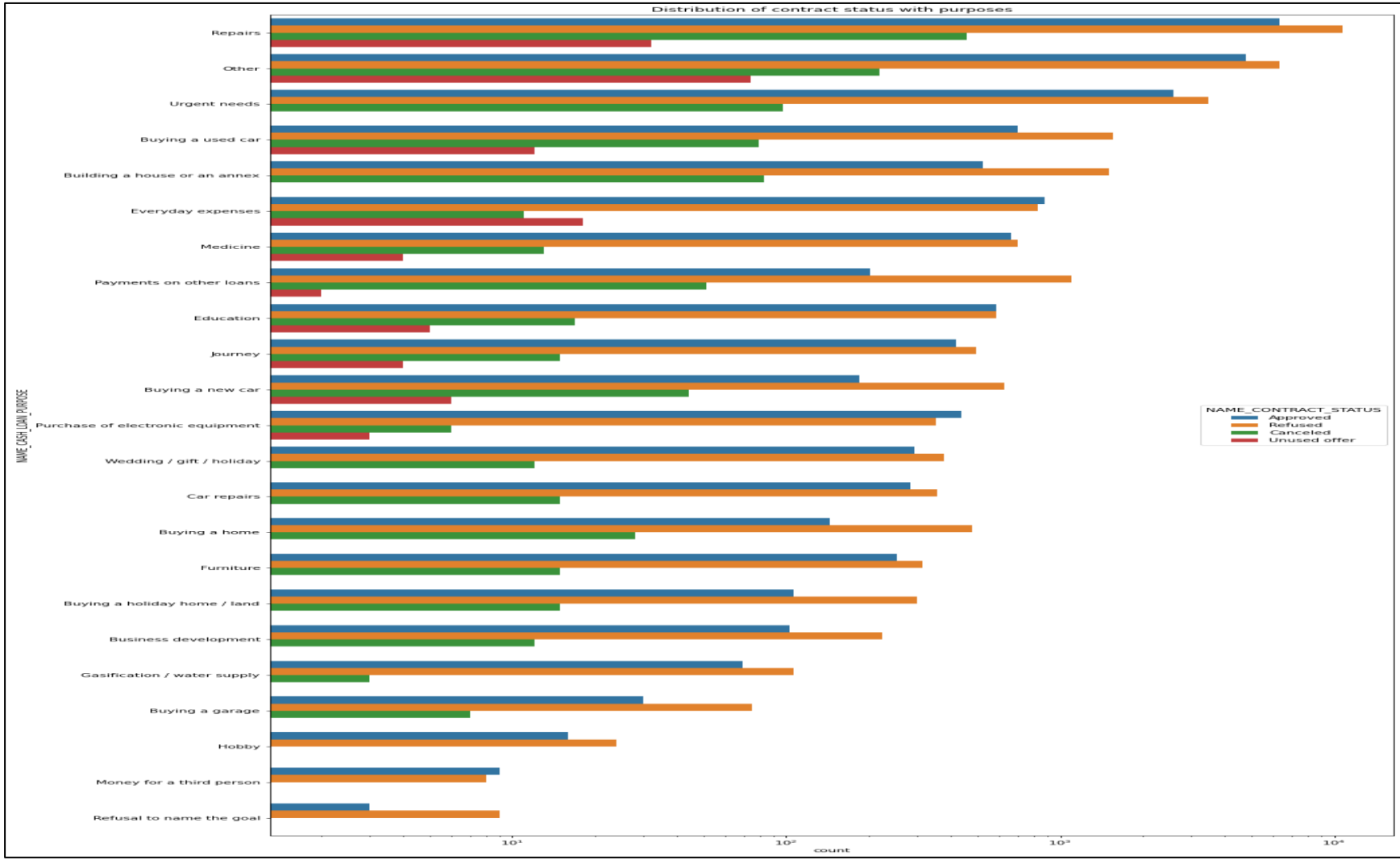


# UNIVARIATE ANALYSIS ON MERGING WITH PREVIOUS APPLICATION DATASET

## Distribution of Contract Status with Purpose

**Points to be concluded from the graph :**

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

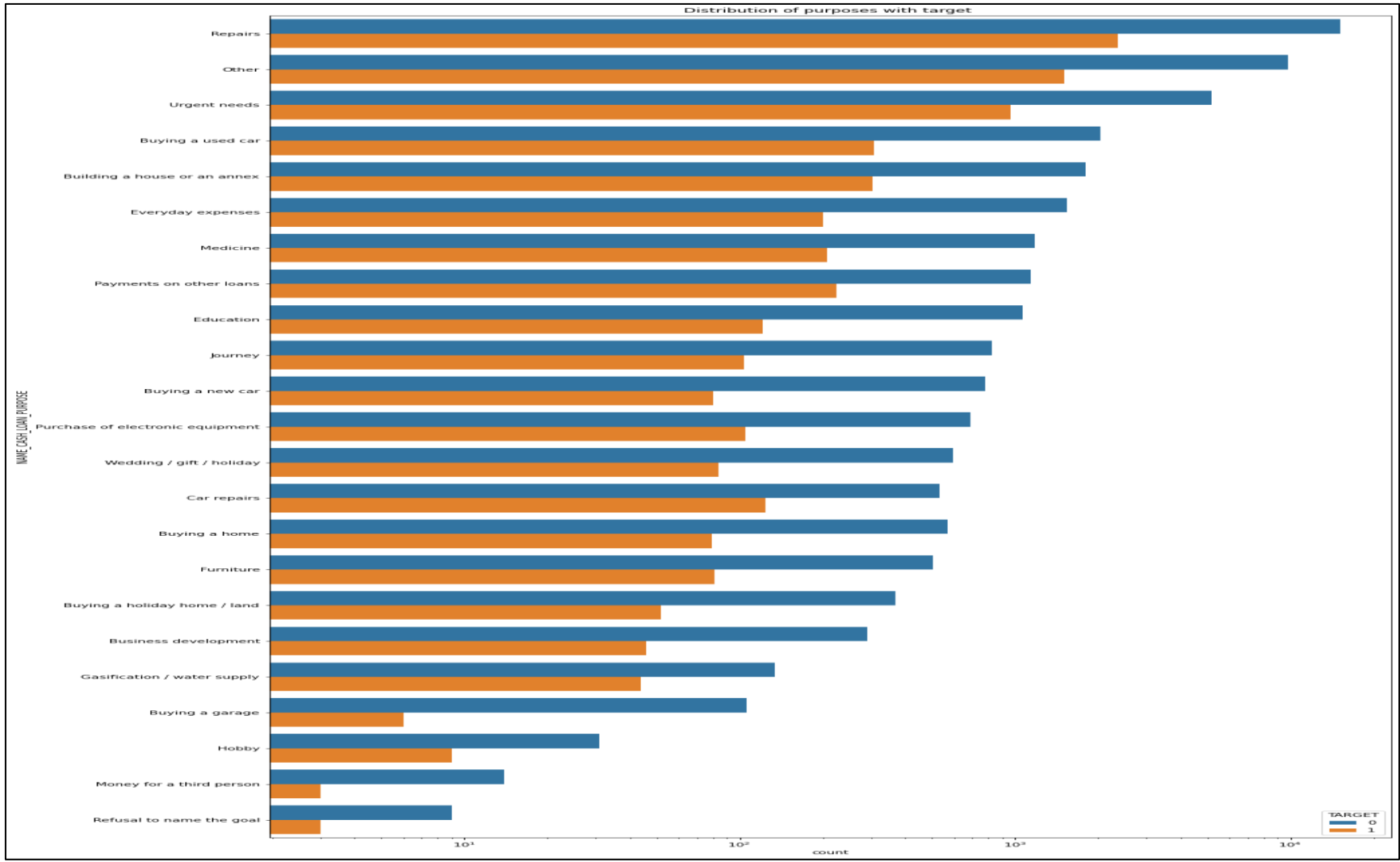


# UNIVARIATE ANALYSIS ON MERGING WITH PREVIOUS APPLICATION DATASET

## Distribution of Purposes with Target

**Points to be concluded from the graph :**

- Loan purposes with 'Repairs' are facing more difficulties in payment on time.
- There are few places where loan payment is significant 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 purposes for which the client is having for minimal payment difficulties.

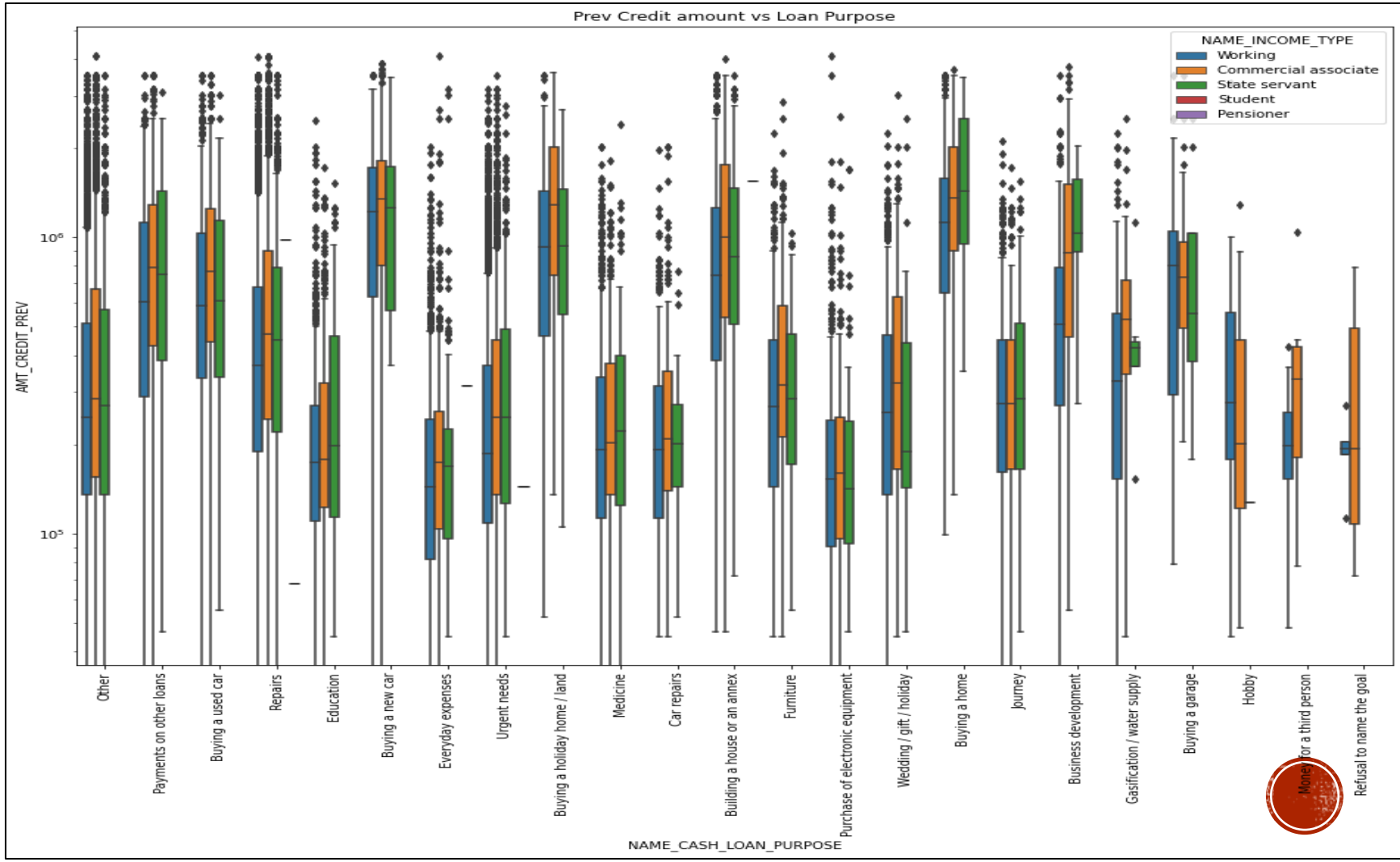


# BIVARIATE ANALYSIS ON MERGING WITH PREVIOUS APPLICATION DATASET

## Previous Credit Amount Vs Loan Purpose

**Points to be concluded from the graph :**

- The credit amount of Loan purposes like 'Buying a home', 'Buying a land', 'Buying a new car' and 'Building a house' is higher.
- Income type of state servants have a significant amount of credit applied
- Money for third person or a Hobby is having less credits applied for.

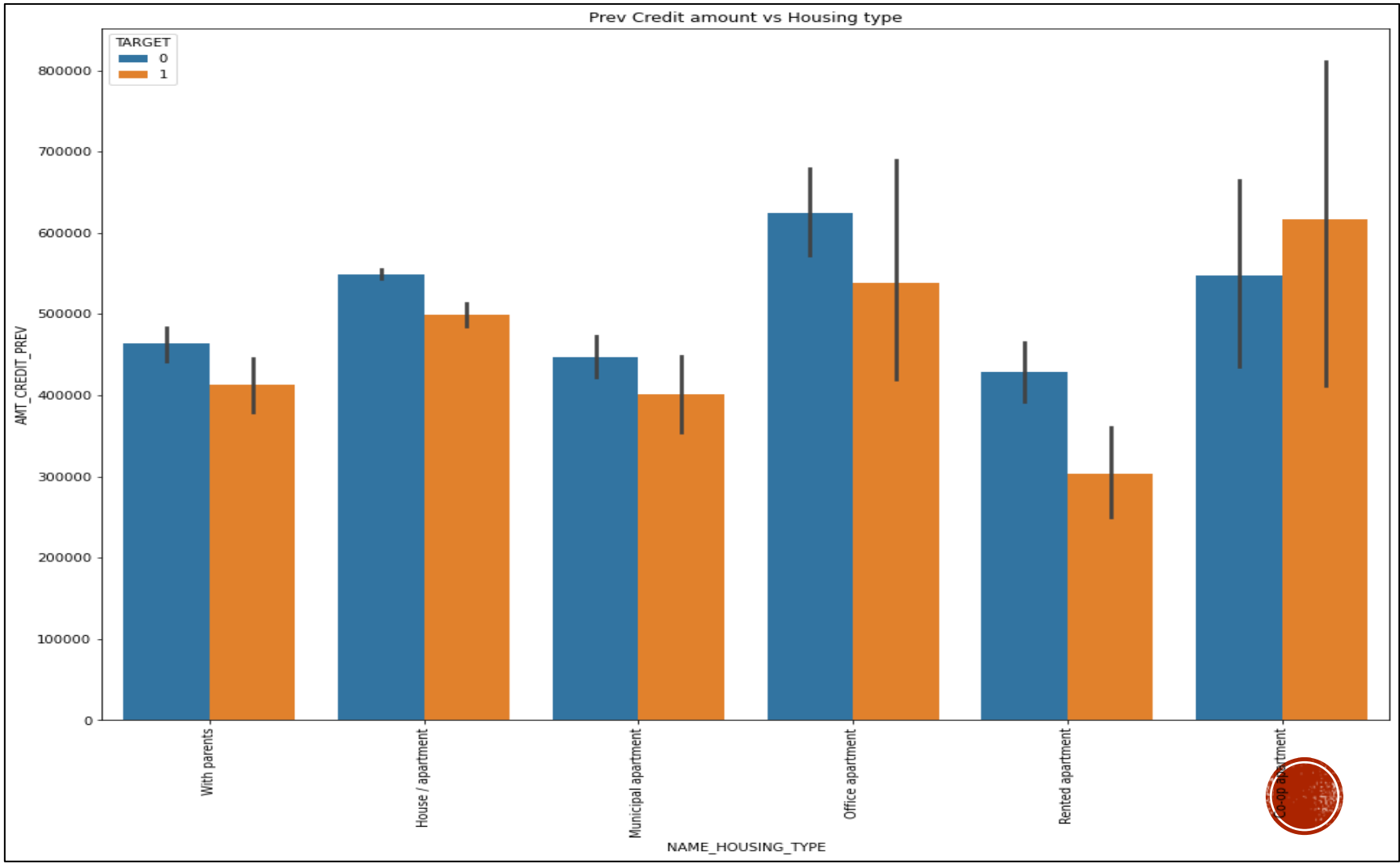


# BIVARIATE ANALYSIS ON MERGING WITH PREVIOUS APPLICATION DATASET

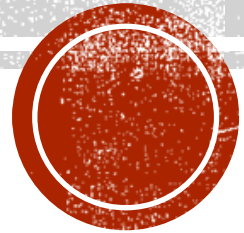
## Previous Credit Amount Vs Housing Type

**Points to be concluded from the graph :**

- Here for Housing type, office apartment is having higher credit of target 0 and co-op apartment is having higher credit of target 1.
- So, we can conclude that bank should avoid giving loans to the housing type of co-op apartment as they are having difficulties in payment.
- Bank can focus mostly on housing type with parents or House\apartment or municipal apartment for successful payments.



# CONCLUSION



- Banks should focus more on contract type ‘Student’ ,’pensioner’ and ‘Businessman’ with housing ‘type other than ‘Co-op apartment’ for successful payments.
- Get as much as clients from housing type ‘With parents’ as they are having least number of unsuccessful payments.
- Banks should focus less on income type ‘Working’ as they are having most number of unsuccessful payments.