Credit Exploratory Data Analysis

Presented by: Anwish Kumar(DSC57)

Objective

Credit risk analysis assists the company in making informed decisions regarding loan approval by evaluating the applicant's profile. This process helps prevent potential business loss for the company by mitigating the risk of financial losses.

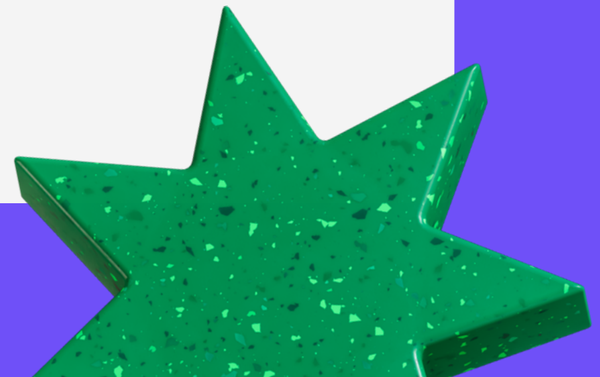
Problem Statement

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 to 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 have to 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 has to 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 at different stages of the process.

Business Understanding

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 applicants using EDA is the aim of this case study
In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i e the variables which are strong indicators of default The company can utilize this knowledge for its portfolio and risk assessment.



DATASETS

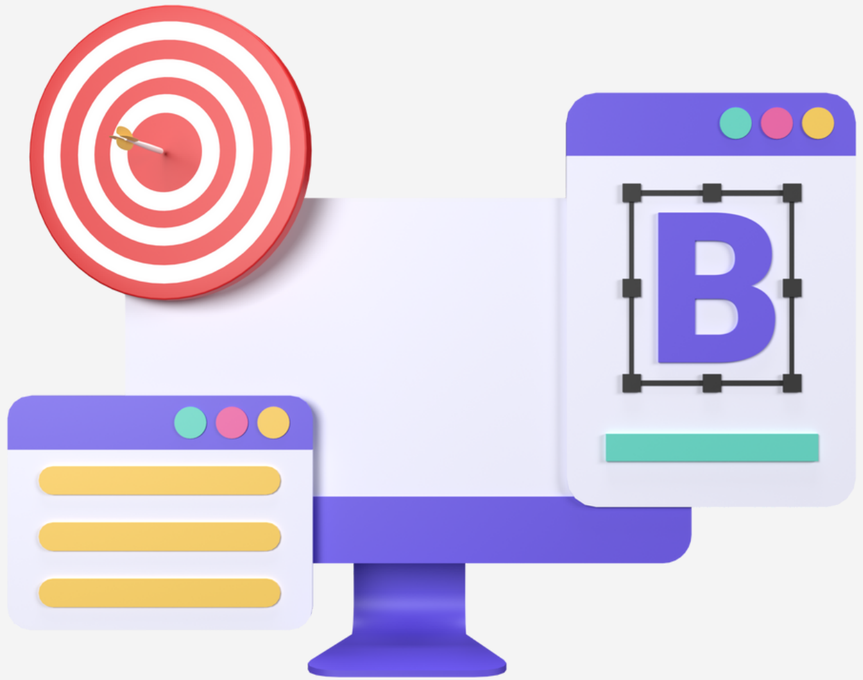
This dataset has 3 files as explained below:

1. 'application_data.csv' contains all the information of the client at the time of application.
The data is about whether a client has payment difficulties.

2. 'previous_application.csv' contains information about the client's previous loan data. It contains the data on whether the previous application had been Approved, Cancelled, Refused or Unused offer.

3. 'columns_description.csv' is data dictionary which describes the meaning of the variables.

Steps Involved

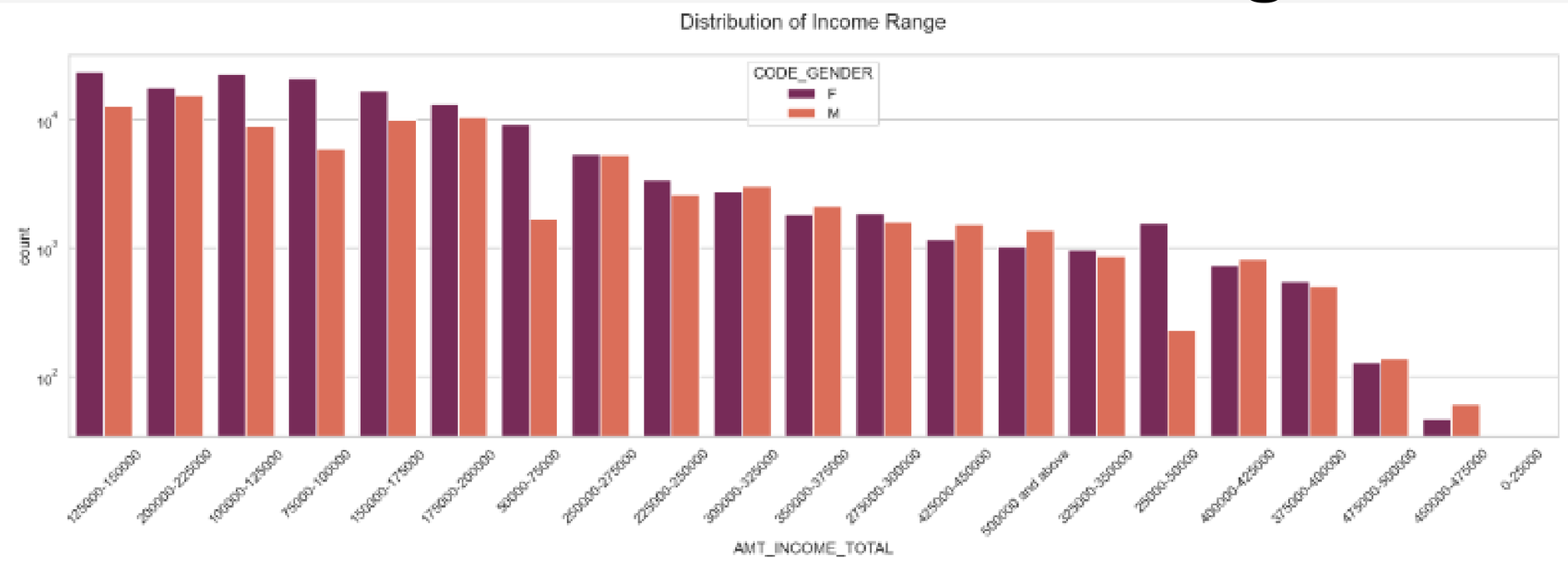


- Understanding the domain
- Importing the Data
- Check the structure of the data.
- Missing Value check and correction
- Outlier check
- Performing Univariate Analysis
- Performing Segmented Univariate Analysis
- Preforming Bivariate Analysis.

Data Visualization

UNIVARIATE ANALYSIS FOR TARGET O

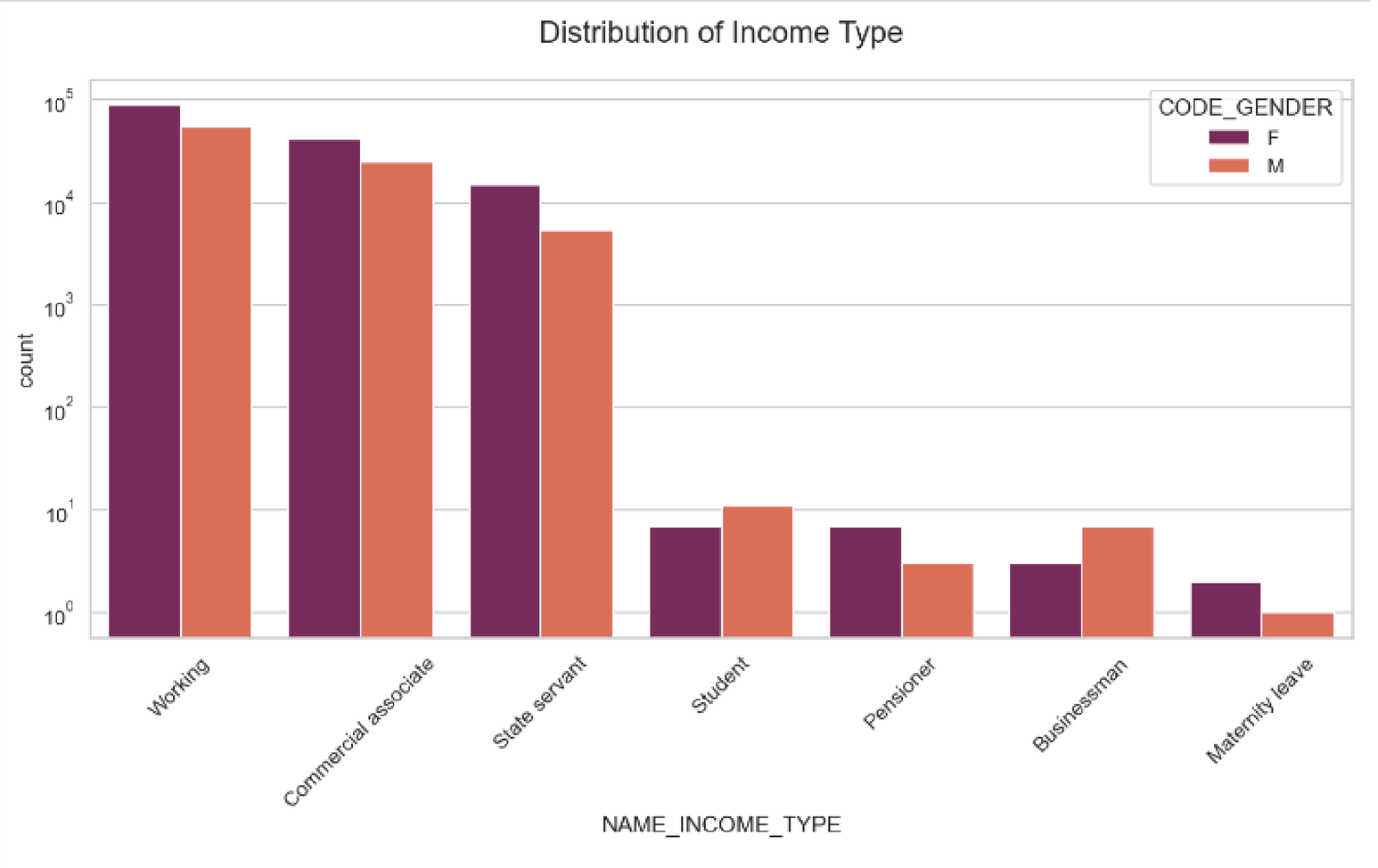
(Distribution of Income Range)



POINTS TO BE CONCLUDED FROM THE GRAPH:

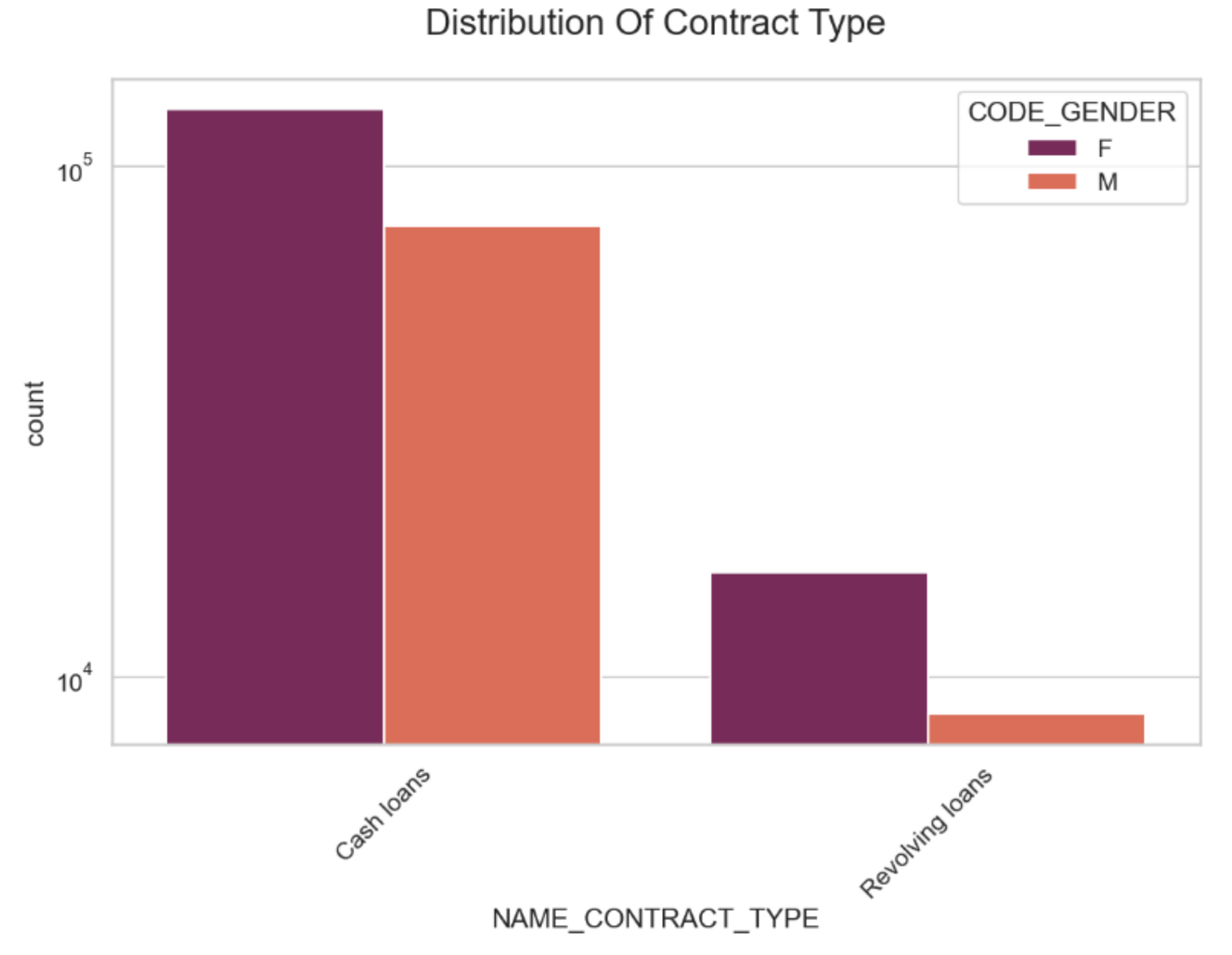
- **The number of females is greater than that of males.**
- **There is a higher frequency of credits within the income range of 100,000 to 200,000.**
- **The graph indicates a higher proportion of females obtaining credits within that specified income range.**
- **The count of individuals with credits is notably lower for income ranges of 400,000 and above.**

Distribution of Income Type



Points to be concluded from the graph:

- **For income 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.**
- **Less number of credits for income type ‘student’ ,’pensioner’, ‘Businessman’ and ‘Maternity leave’.**



Distribution for contract type

Points to be concluded from the graph on the left:

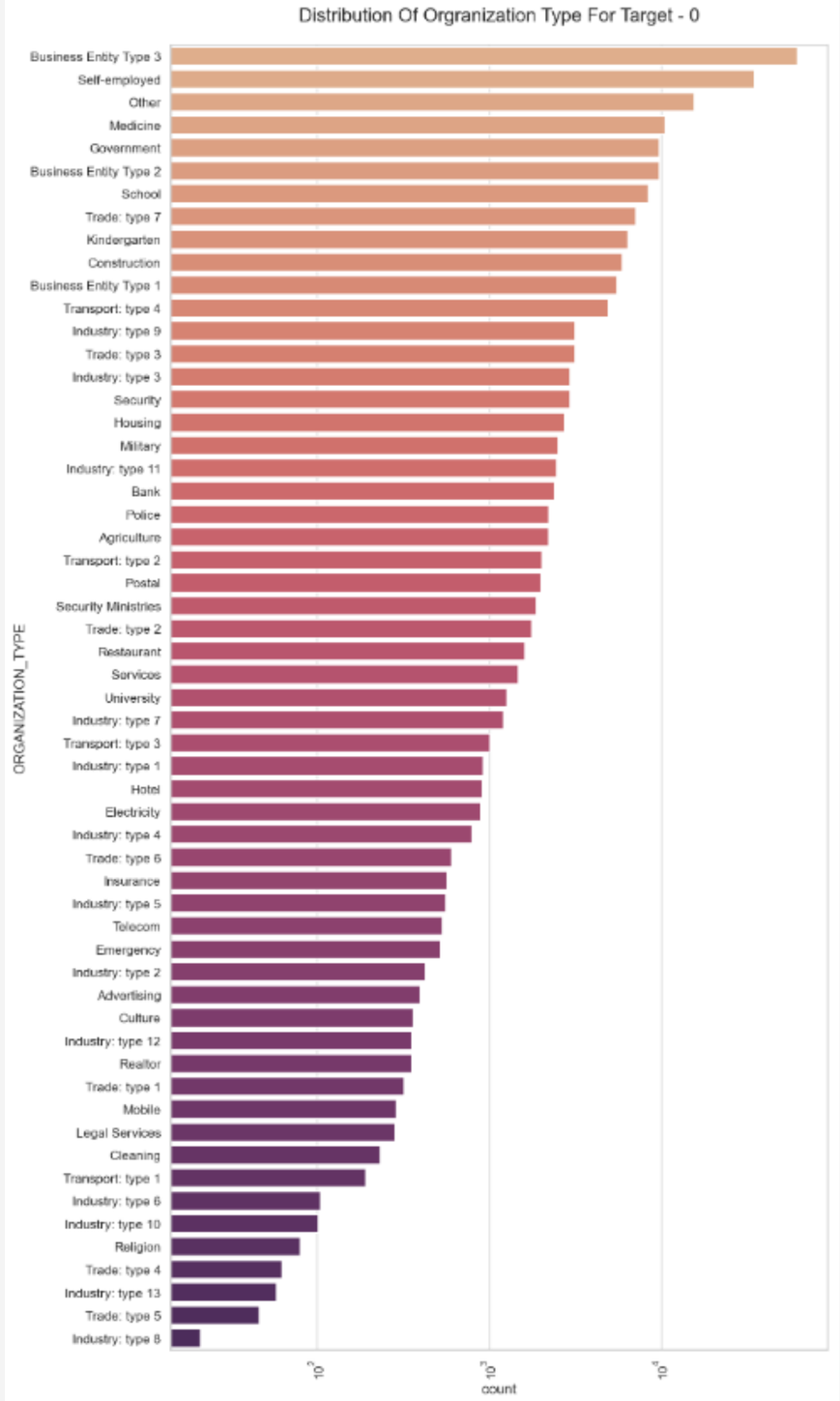
- For contract type ‘cash loans’ is having higher number of credits than ‘Revolving loans’ contract type.
- For this also Female is leading for applying credits

Distribution of organization type

Points to be concluded from the graph:

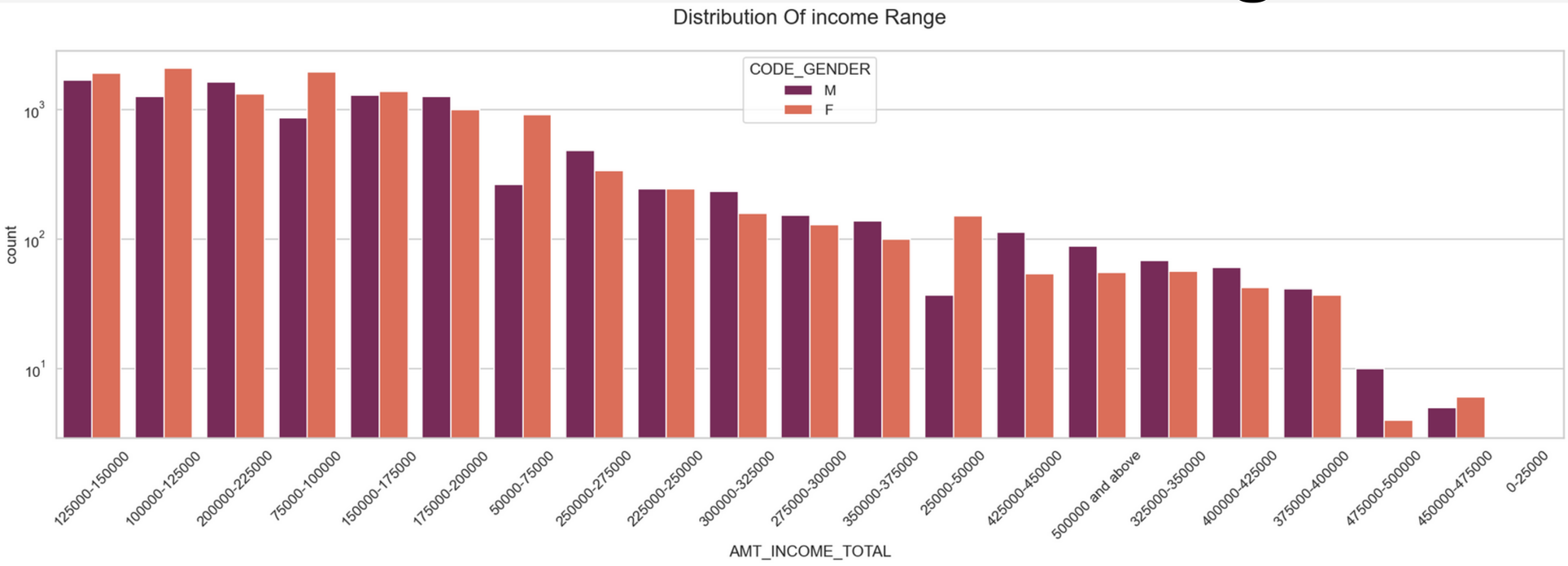
- **Clients which have applied for credits are from most of the organization type ‘Business entity Type 3’ , ‘Self employed’ , ‘Other’ , ‘Medicine’ and ‘Government’.**

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



UNIVARIATE ANALYSIS FOR TARGET 1

(Distribution of Income Range)



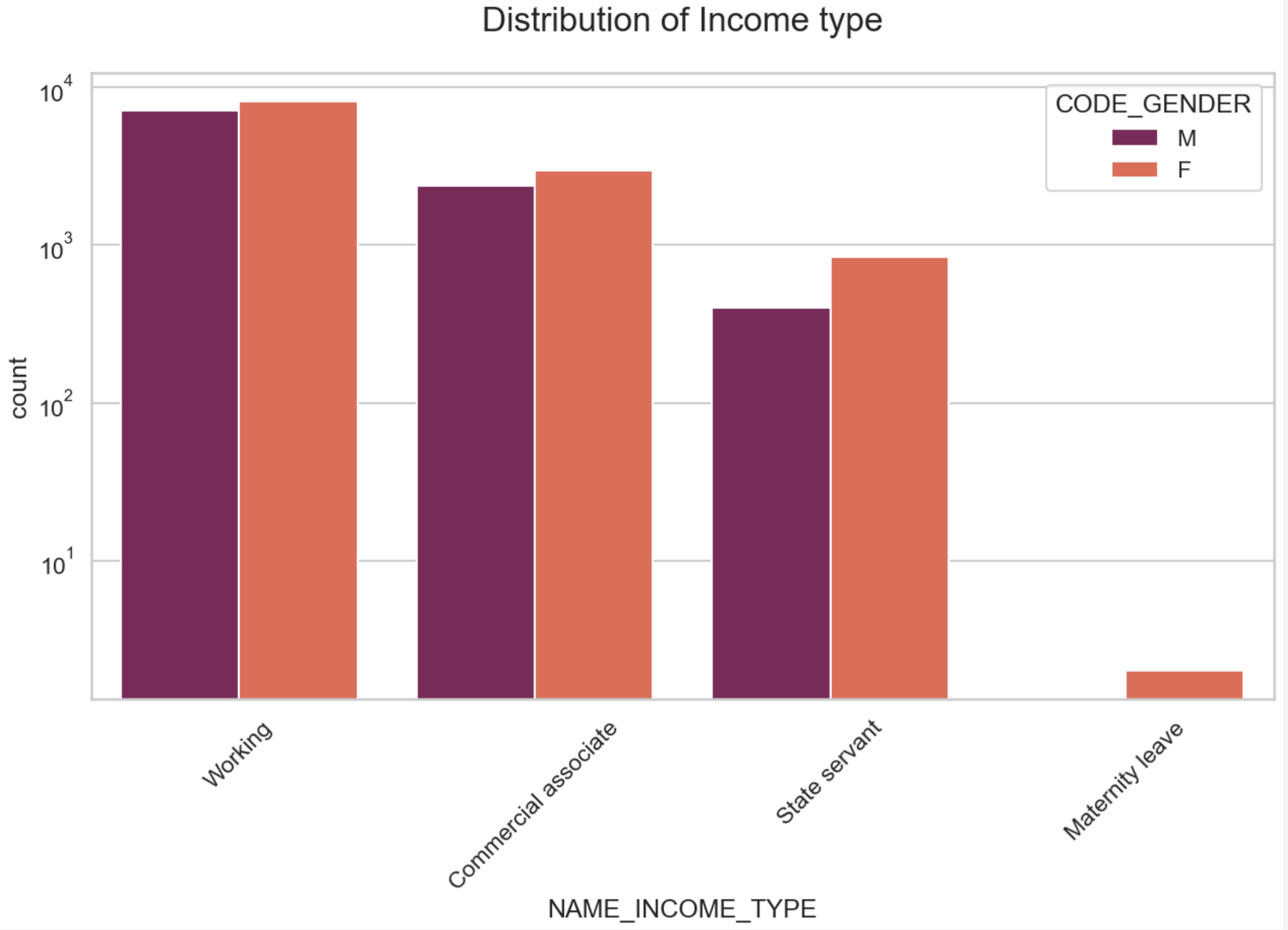
POINTS TO BE CONCLUDED FROM THE GRAPH:

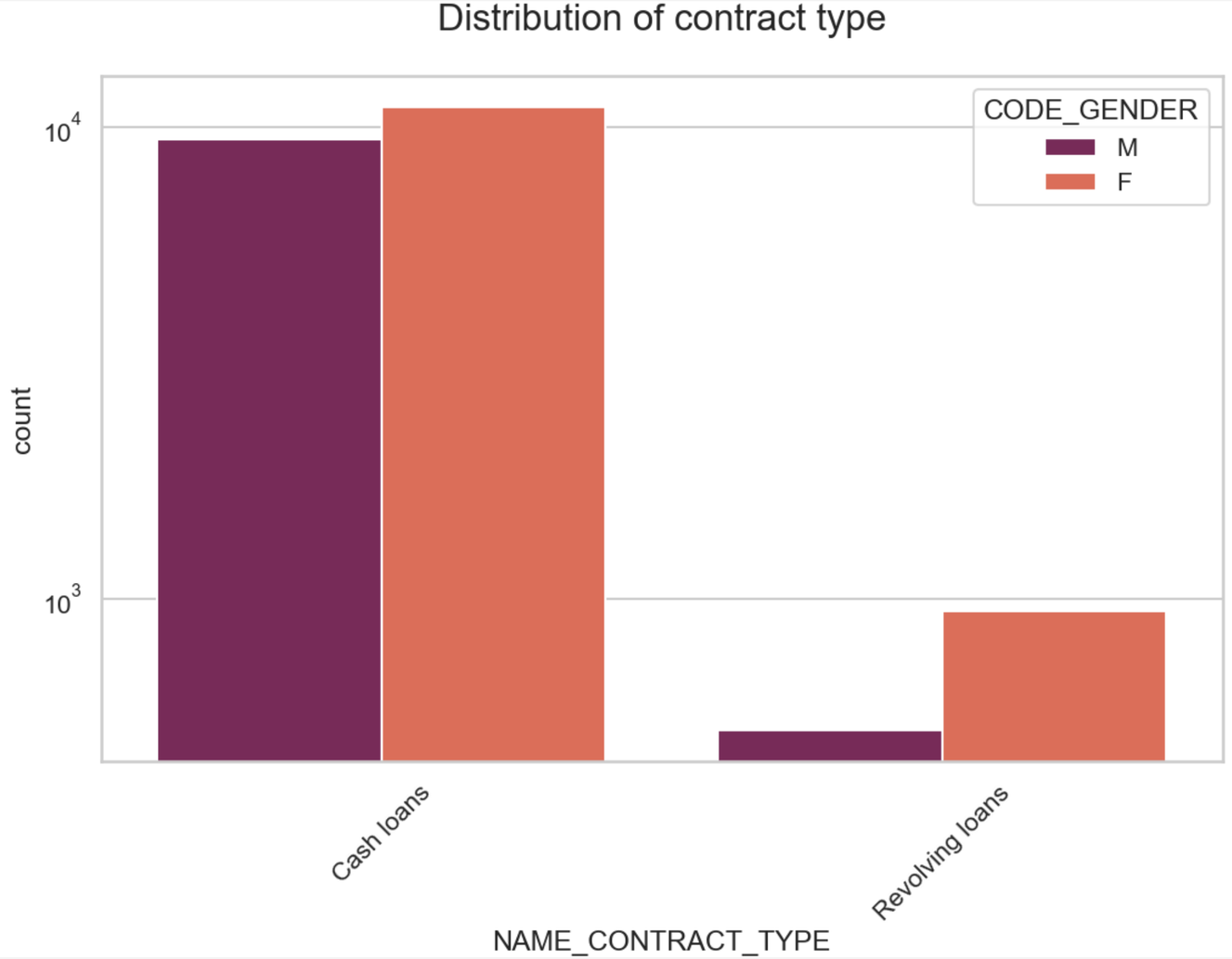
- **Male Counts are higher than Females.**
- **Income range from 100000 to 200000 is having more number of credits.**
- **Very less count for income range 400000 and above.**

Distribution of Income Type

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.**
- **2. For this Females are having more number of credits than male.**
- **3. For type 1: There is no income type for ‘student’ , ’pensioner’ and ‘Businessman’ which means they don’t do any late payments.**





Distribution for contract type

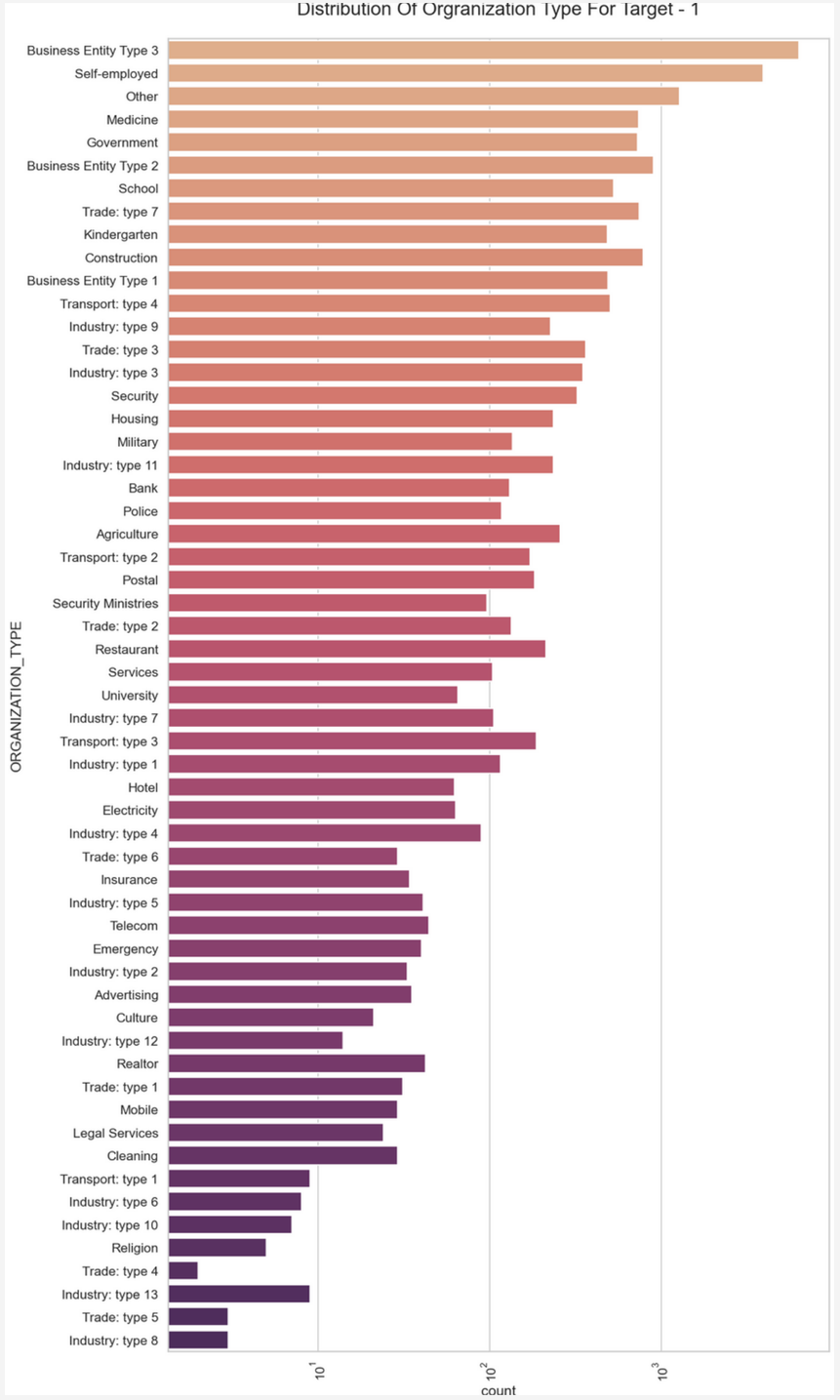
Points to be concluded from the graph on the left:

- For Contract type CASH LOAN is having higher number of credits than REVOLVING LOANS.
- Female is leading the plot.

Distribution of organization type

Points to be concluded from the graph:

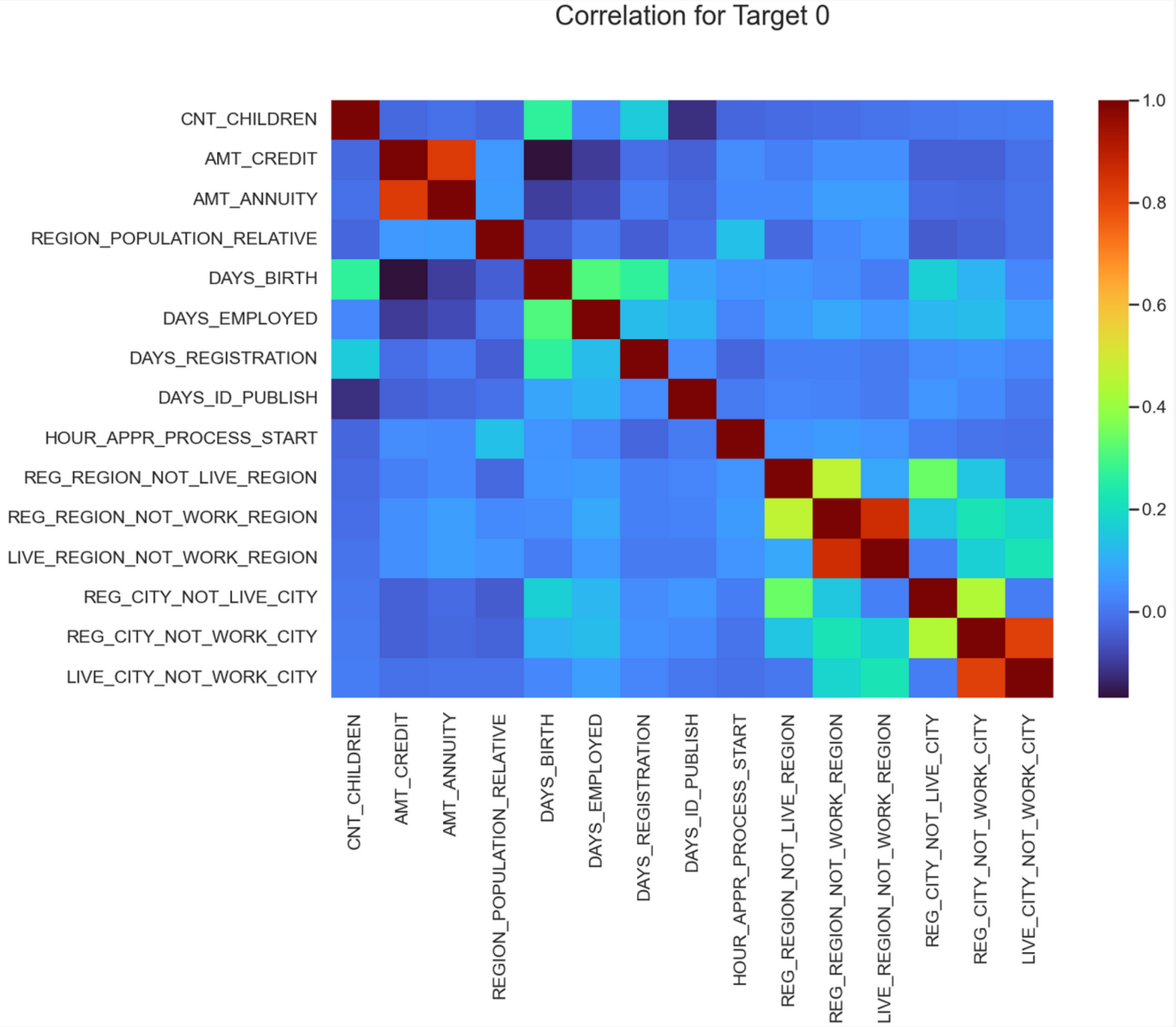
- **Clients which have applied for credits are from most of the organization type ‘Business entity Type 3’ , ‘Self employed’ , ‘Other’ , ‘Medicine’ and ‘Government’.**
- **Less clients are from Industry type 8,type 6, type 10, religion and trade type 5, type 4.**
- **Same as type 0 in distribution of organization type.**



CORRELATION FOR TARGET 0

**As we can see from correlation heatmap,
There are number of observation we can point out:**

- Credit amount is inversely proportional to the date of birth.
- Credit amount is inversely proportional to the number of children client have.
- Income amount is inversely proportional to the number of children client have.
- 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.

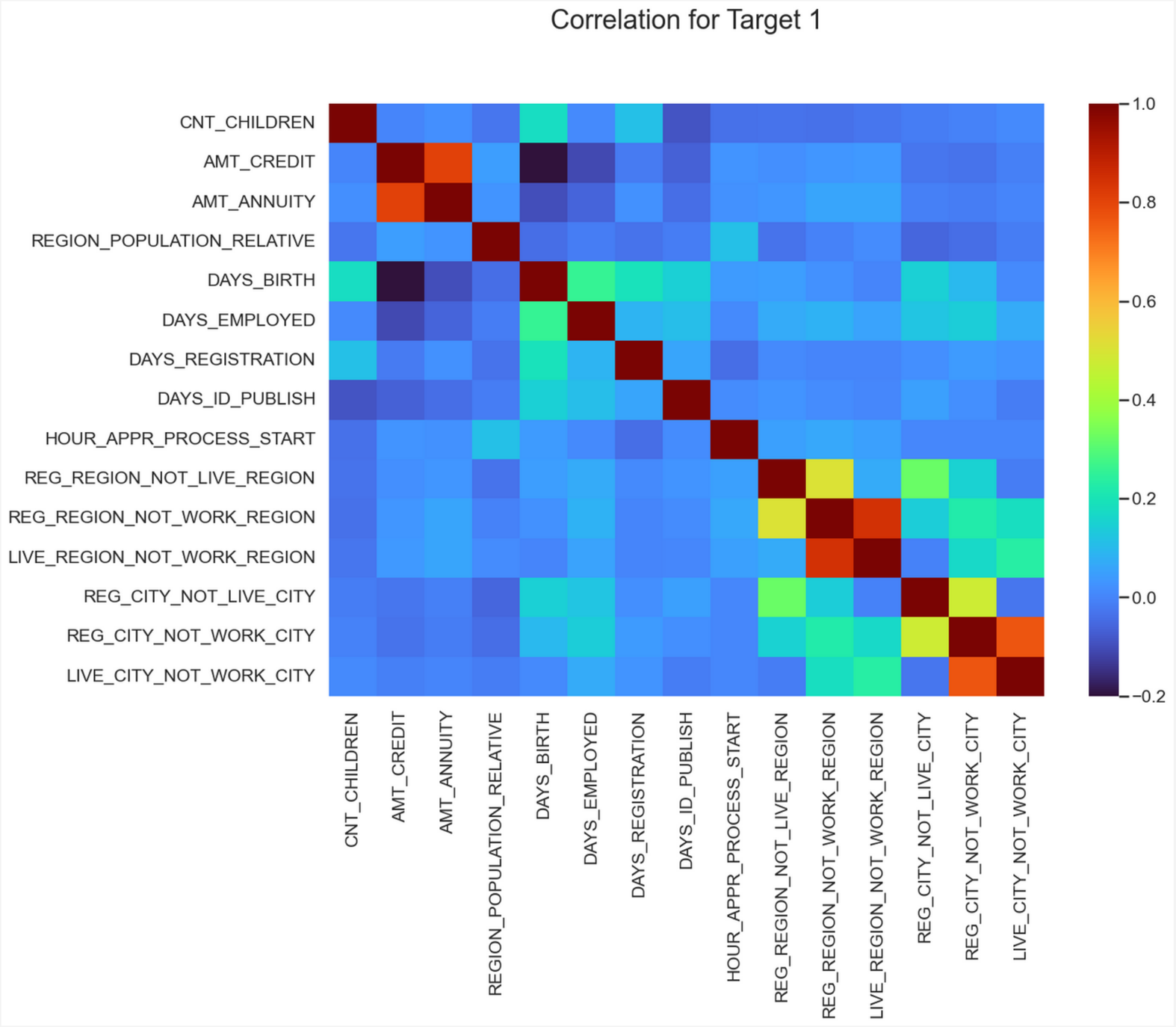


CORRELATION FOR TARGET 0

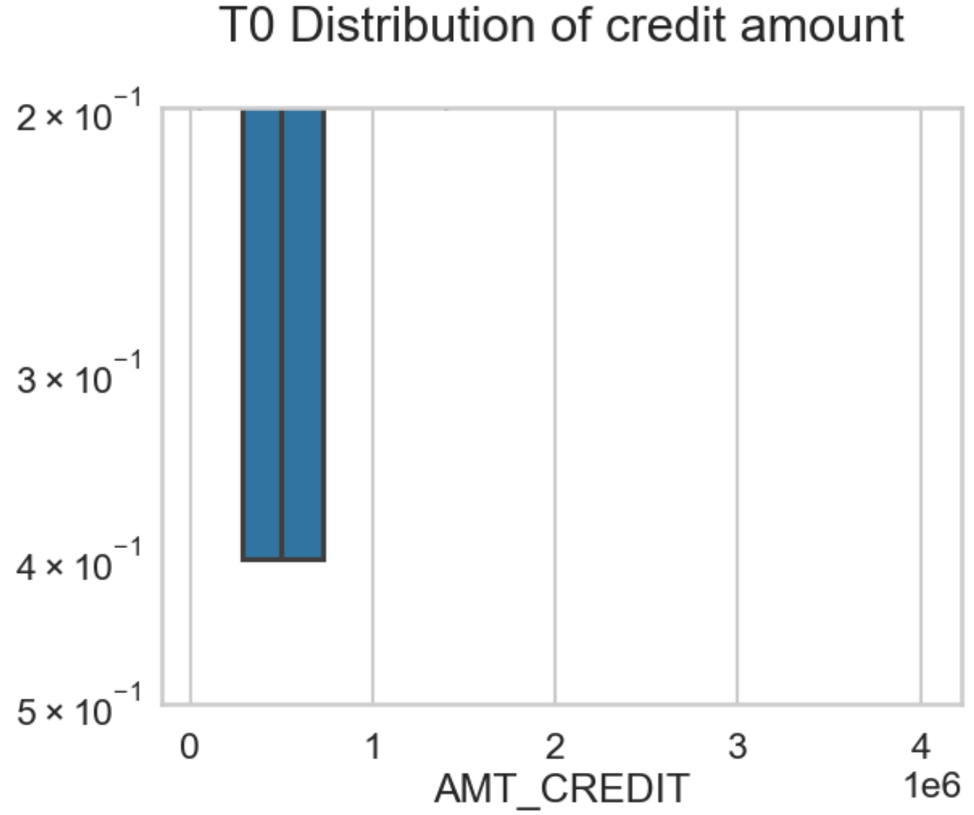
**This heat map for Target 1 is also having quite a same observation just like Target 0.
But for few points are different.**

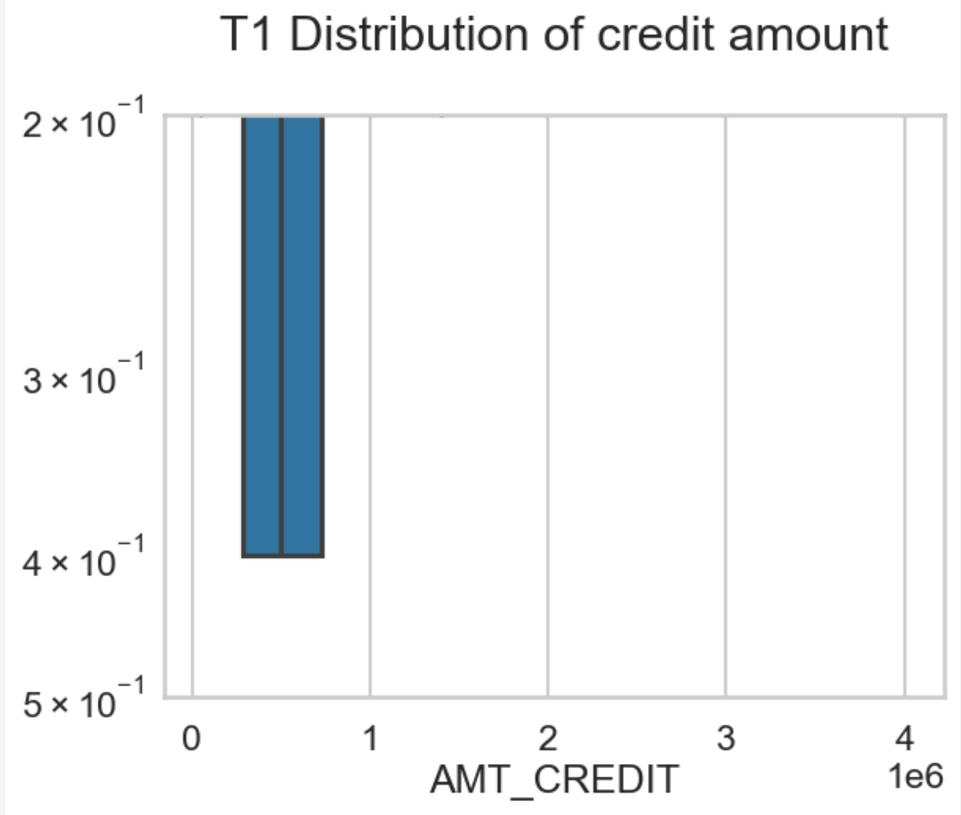
- The client's permanent address does not match contact address are having less children and vice-versa

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



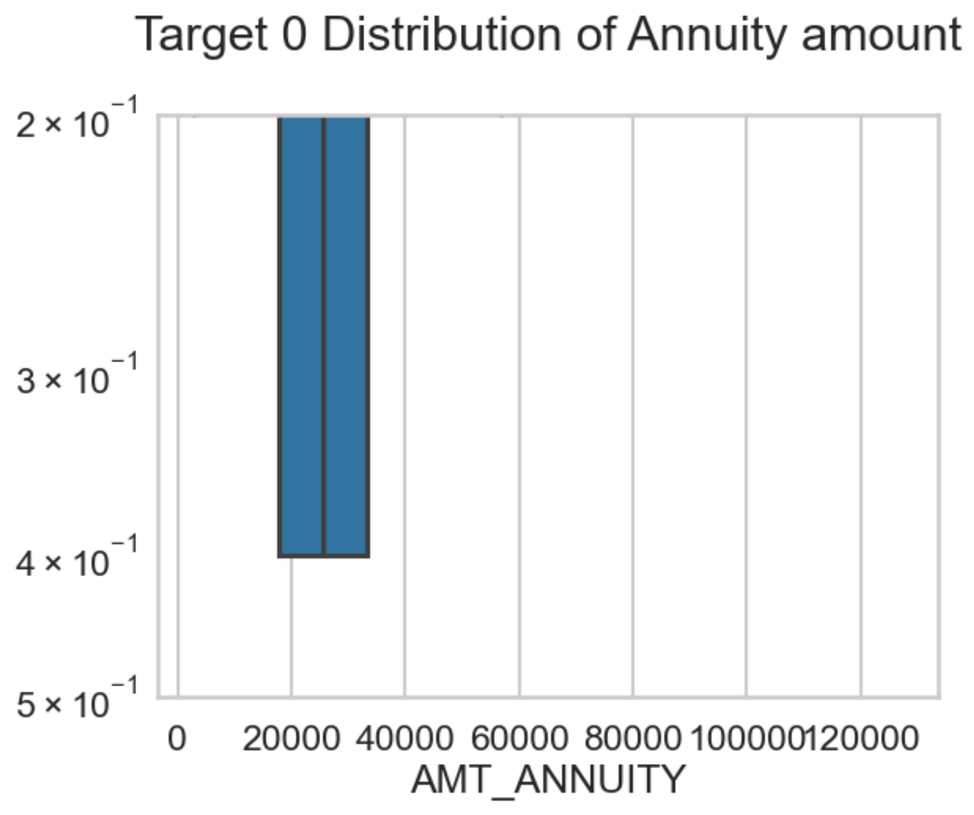
CONTINUOUS UNIVARIATE ANALYSIS
DISTRIBUTION OF ANNUITY AMOUNT FOR TARGET0 AND TARGET1

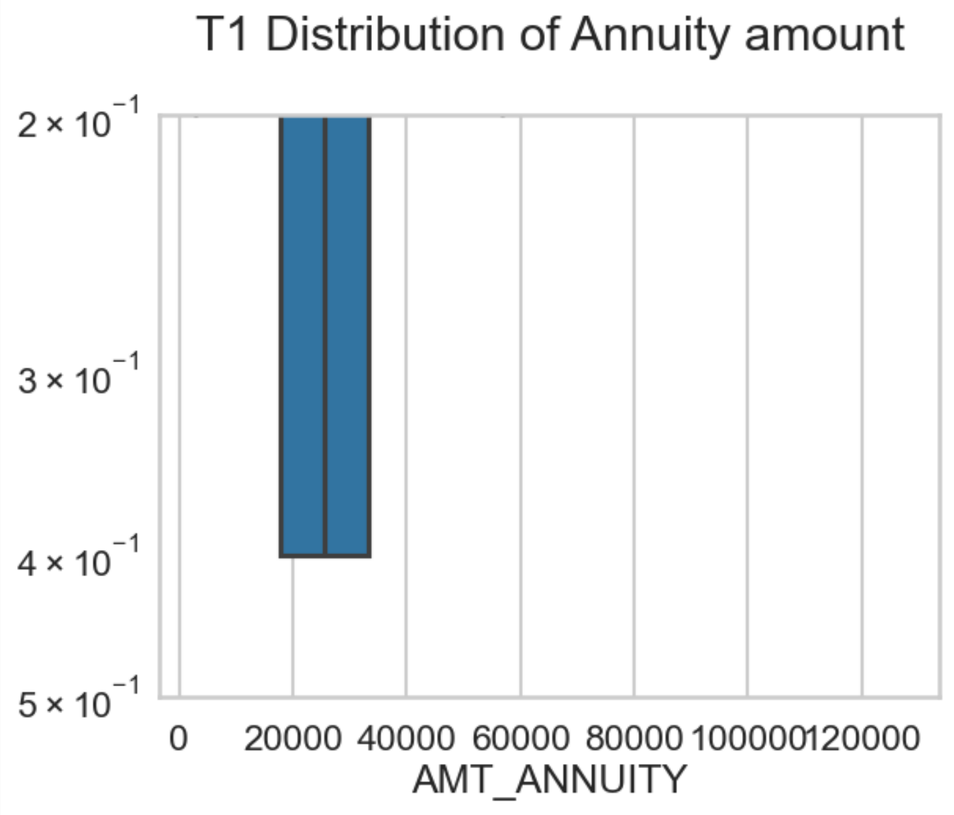




- Outliers are present in both
- 3rd quartile is narrow for both target 1 and target 0
- Most of the clients have credit amount in the 1st quartile.

CONTINUOUS UNIVARIATE ANALYSIS
DISTRIBUTION OF ANNUITY AMOUNT FOR TARGET0 AND TARGET1





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

BIVARIATE ANALYSIS

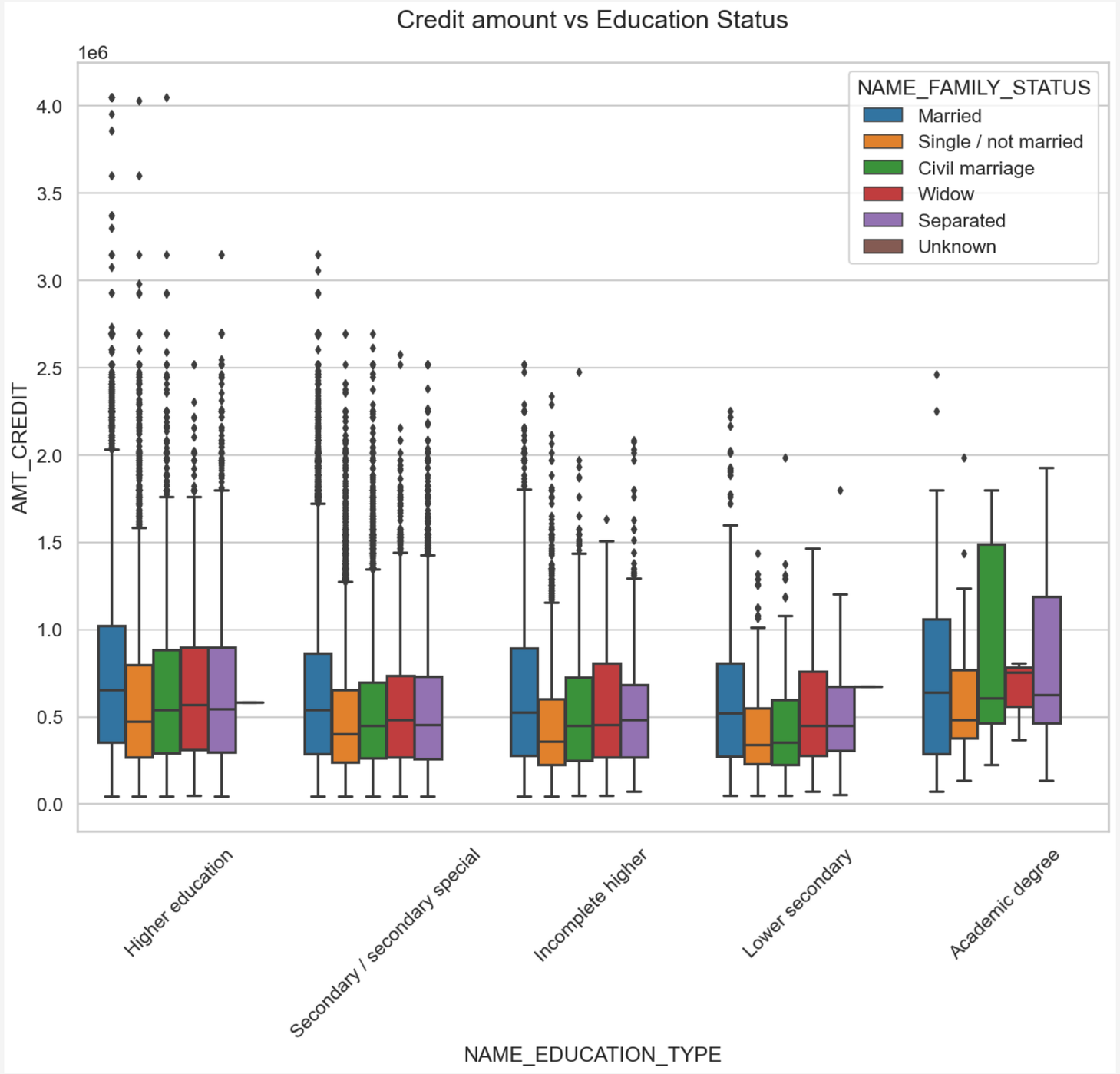
Credit amount vs Education Status

Based on the provided box plot, several conclusions can be drawn:

- **Family statuses categorized as 'civil marriage,' 'marriage,' and 'separated' are associated with a greater number of credits compared to other family statuses when considering individuals with an Academic degree education.**

- **Family statuses of 'marriage,' 'single,' and 'civil marriage' that possess a higher level of education tend to exhibit a greater number of outliers, indicating potential variations in credit behavior.**

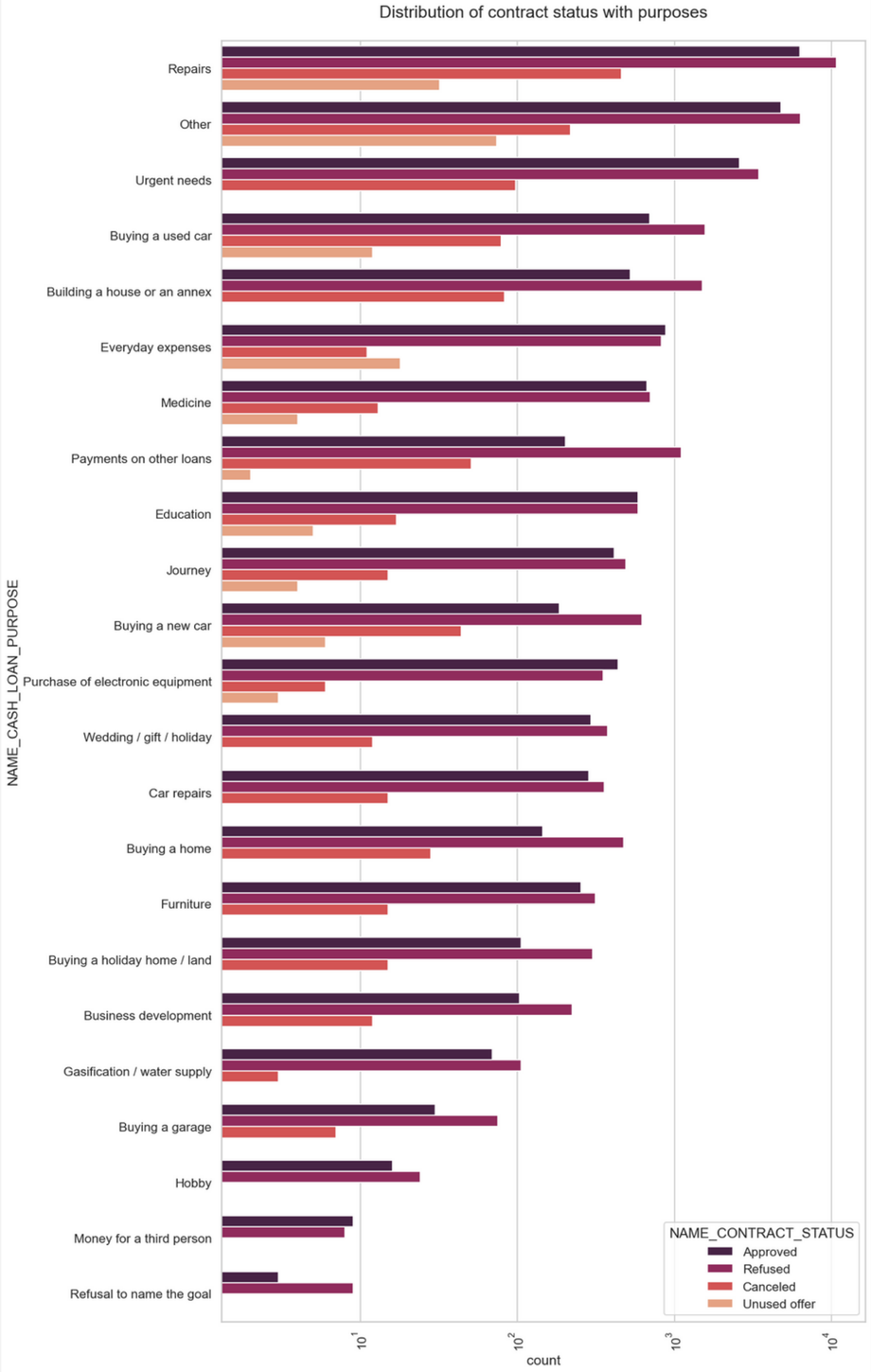
- **Notably, within the context of individuals with an Academic degree and a family status of 'civil marriage,' the majority of credits fall within the third quartile, highlighting the concentration of credits within this group.**



Univariate analysis after merging previous data

Distribution of contract status with purposes

- Most rejection of loans came from purpose 'repairs'.
- Paying other loans and buying a new car is having significant higher rejection than approves.

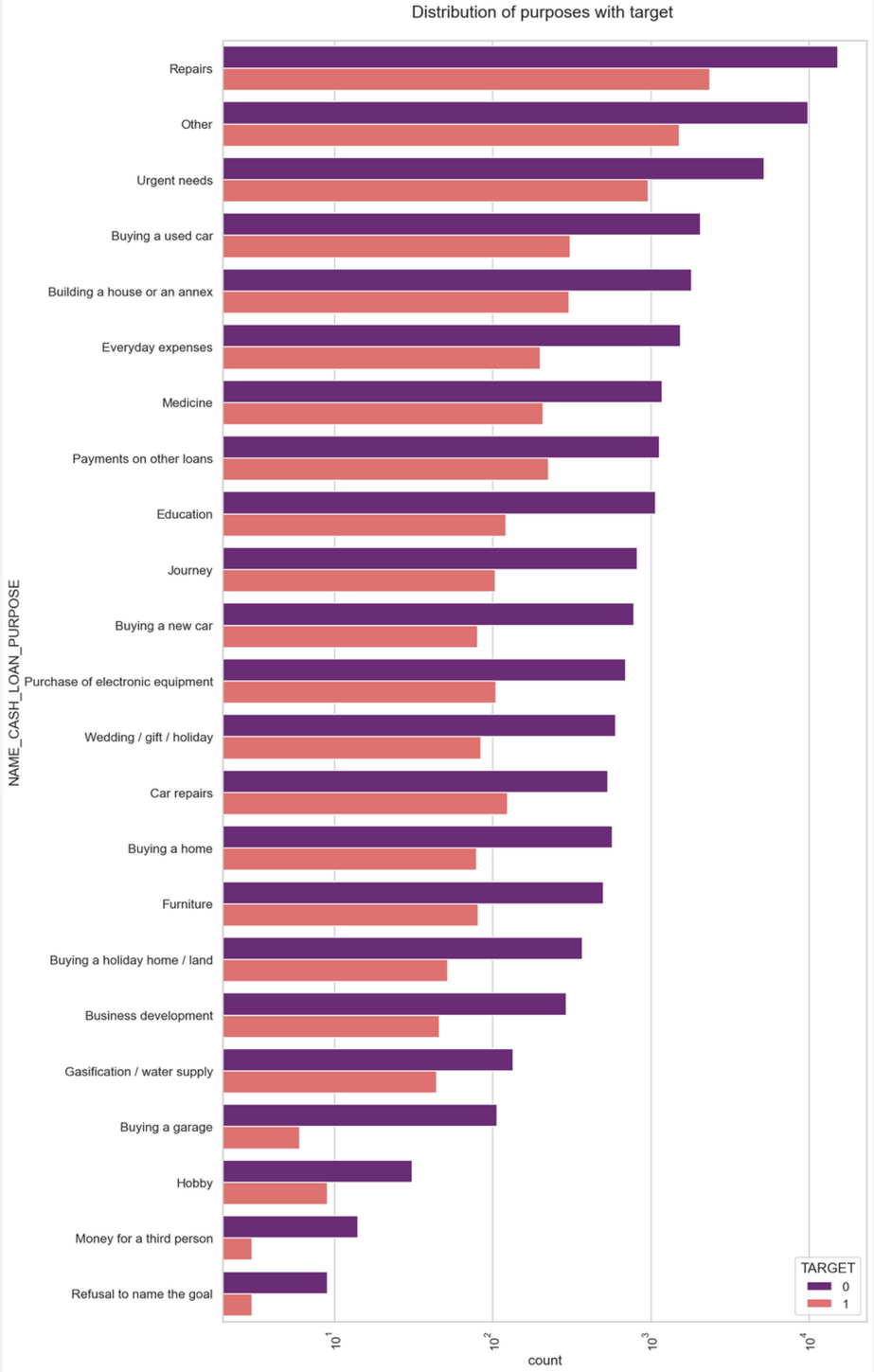


Distribution of purposes with target

Few points we can conclude from the plot:

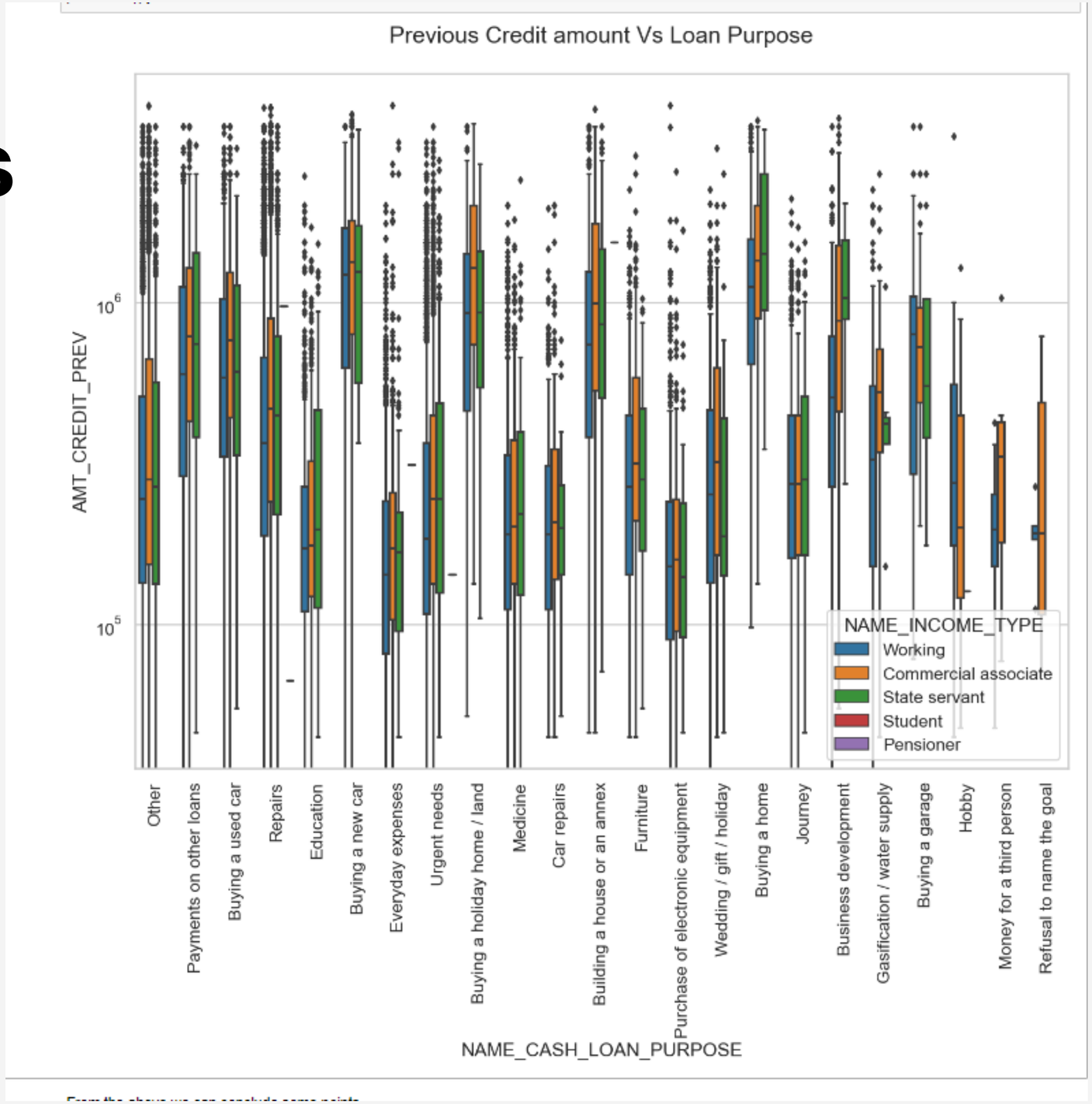
- Loan purposes with 'Repairs' are facing more difficulties in payment on time.
- There are few places where loan payment is significant higher than facing difficulties.
- They are 'Buying a garage', 'Business developemt, 'Buying land , Buying a new car' and 'Education'

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



Performing Bivariate analysis

Previous Credit amount vs Loan Purpose

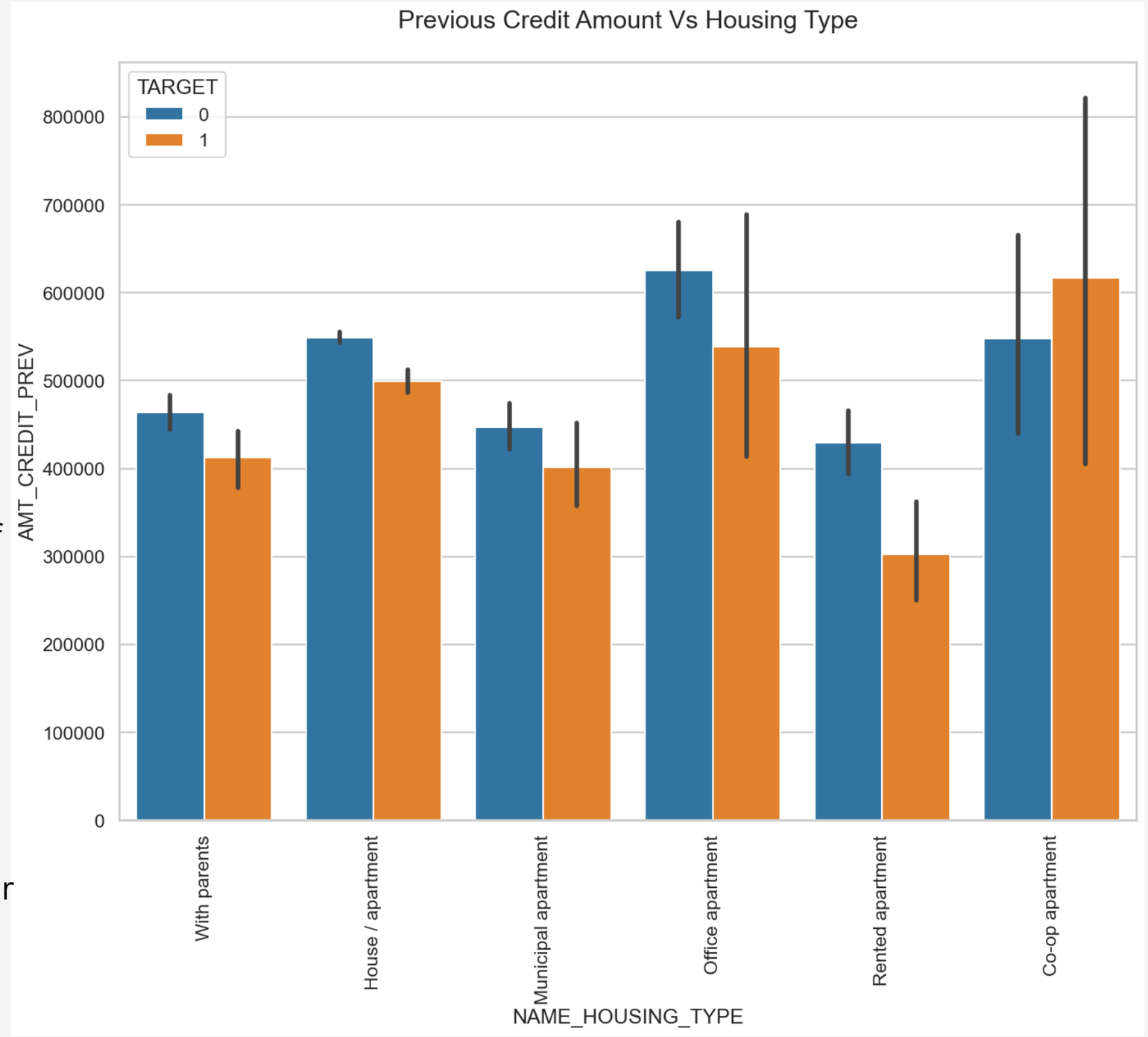


From the above we can conclude some points-

- 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.

Previous Credit amount vs Housing type

Here for Housing type, office appartment 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\appartment or miuncipal appartment for successful payments.



CONCLUSION

- For successful payments, banks should prioritize contract types such as 'Student,' 'Pensioner,' and 'Businessman.' Additionally, applicants seeking housing types other than 'Co-op apartment' should receive greater attention.

- It's advisable for banks to reduce their focus on clients with an income type of 'Working,' as this category demonstrates the highest frequency of unsuccessful payments.

- Notably, loans designated for the purpose of 'Repair' exhibit a higher occurrence of untimely repayments, warranting closer evaluation.

- To minimize the occurrence of unsuccessful payments, banks could target clients from the 'With parents' housing type, which exhibits the lowest frequency of payment issues.

Thank you!