# CREDIT EDA CASE STUDY

By :

Surya Kumar . K

# THE PROBLEM

## Company

Loan Providing Company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

## Context

Loan Providing Company wants to understand the **driving factors** behind loan default, i.e. the **driver variables** which are strong indicators of default.

The company can utilize this knowledge for its portfolio and risk assessment.
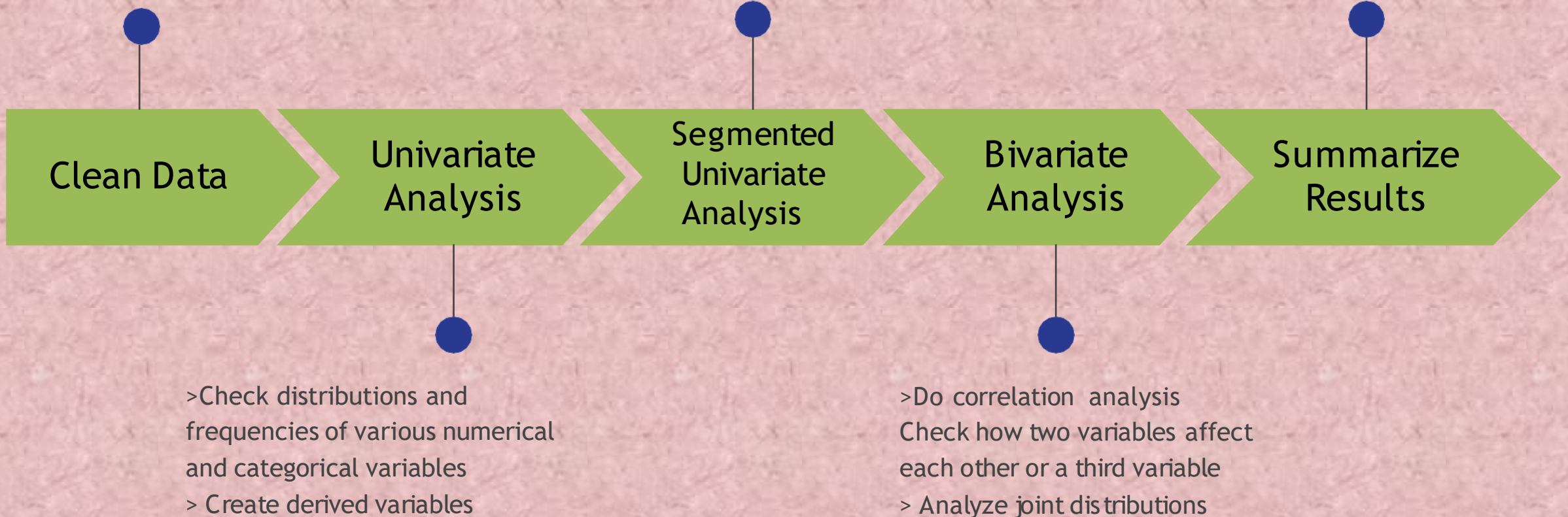
## Problem statement

As a data scientist working for Loan Providing Company analyze the dataset containing information about past loan applicants using EDA to understand how *consumer attributes* and *loan attributes* influence the tendency of default

# ANALYSIS APPROACH

>Drop columns with null values, all random values or single category value
>Convert values to proper int, float, date representations

>Analyze variables against segments of other variables
> Create derived variables

Publish insights and observations

| Clean Data | Univariate Analysis | Segmented Univariate Analysis | Bivariate Analysis | Summarize Results |

>Check distributions and frequencies of various numerical and categorical variables
> Create derived variables

>Do correlation analysis
Check how two variables affect each other or a third variable
> Analyze joint distributions

# Univariate Analysis

# CONTRACT TYPE FOR NEW APPLICATIONS

•Count of Cash loans is highest which is more than 1 Million (above 10^6).

•Count of Revolving loans is slightly above 100k (above 10^5).

•From this we can infer that people are getting more cash loans than revolving loans.



Contract Type for New Application
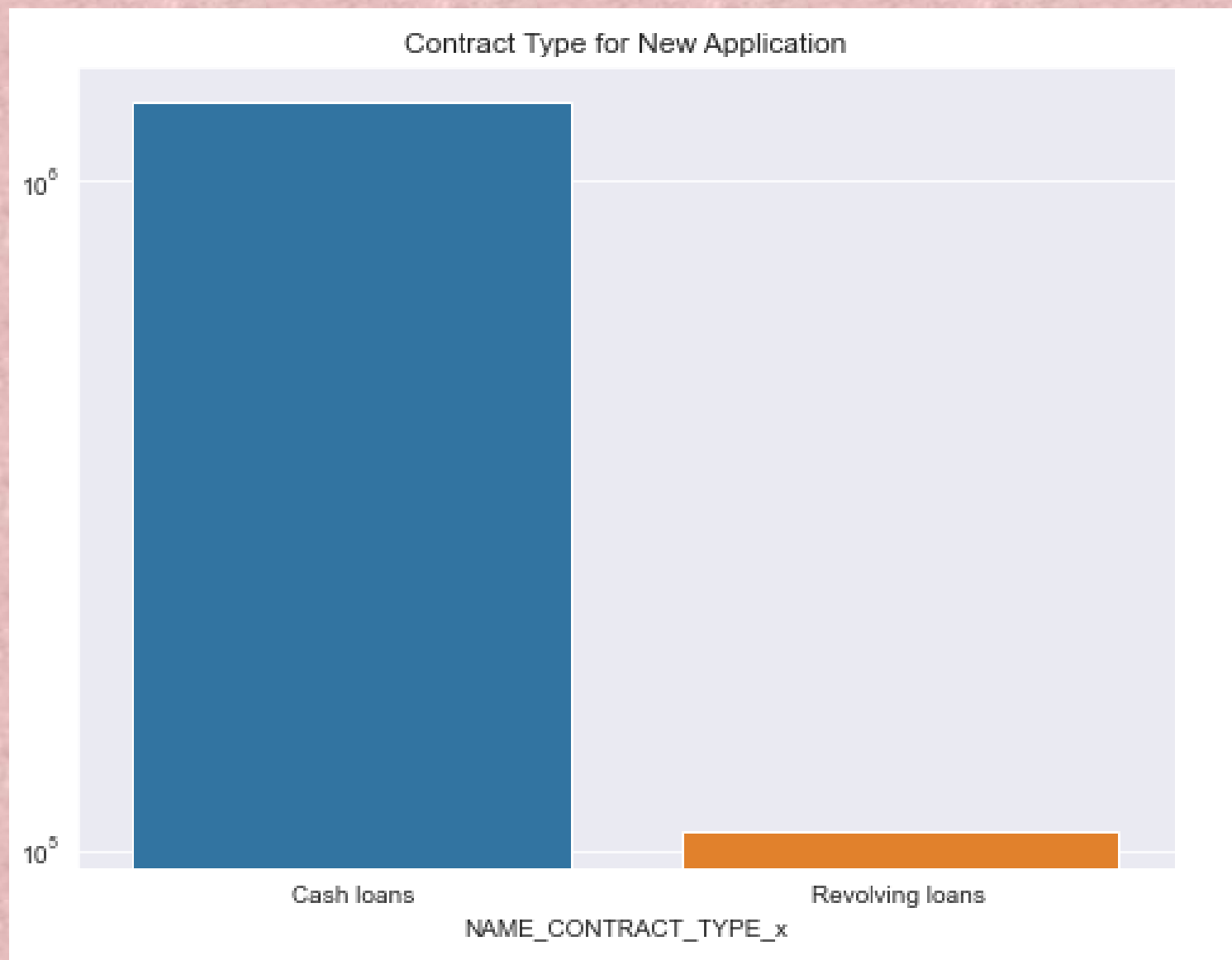
# INCOME TYPE FOR NEW APPLICATIONS

•From the plot we can infer that "Working" type has the highest count.

•And also we can infer that "Maternity leave" and "Student" type are the lowest.

•People with "Working" income type are getting more loans and people with "Maternity leave" and "Student" type are getting less number of loans.
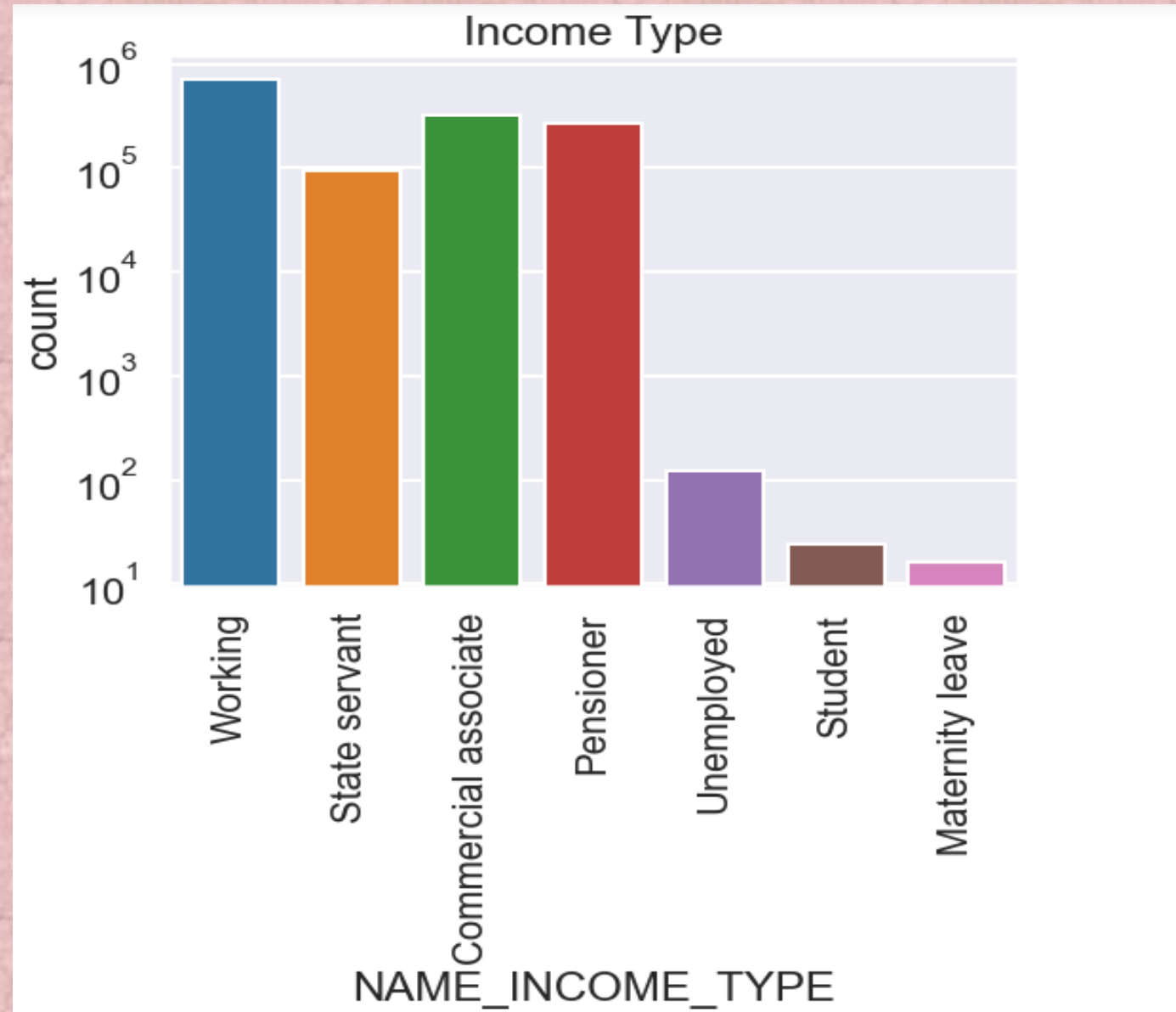
# ORGANIZATION TYPE FOR NEW APPLICATIONS

•From the plot we can infer that People working in "Business Entity Type 3" organization has the highest count.

•And also "XNA" type organization which means Not Applicable has second highest count, which means many people's organization type is unknown.
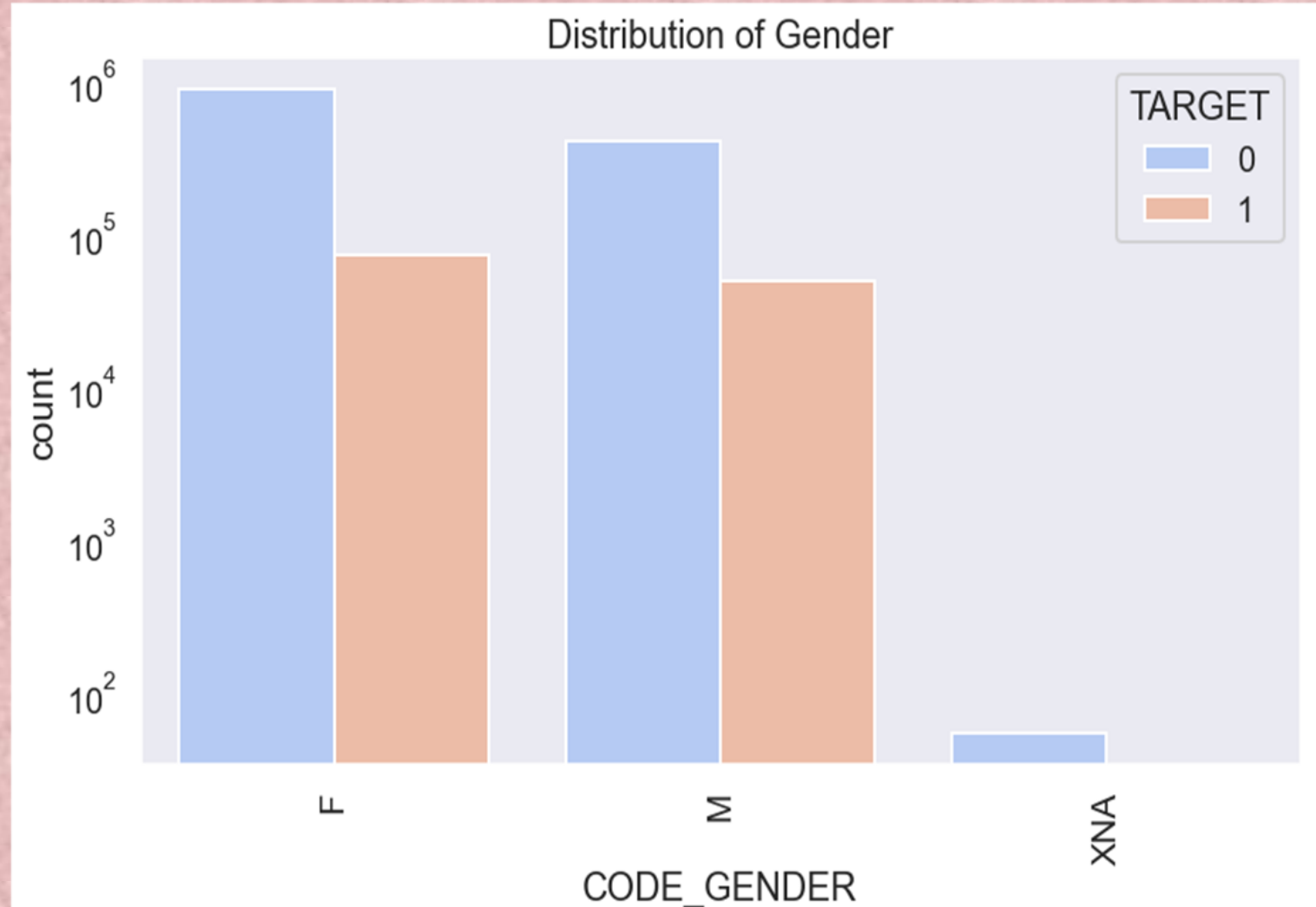
•And also we can infer that People working in "Industry Type 8" type organization has the lowest count.



Organization Type
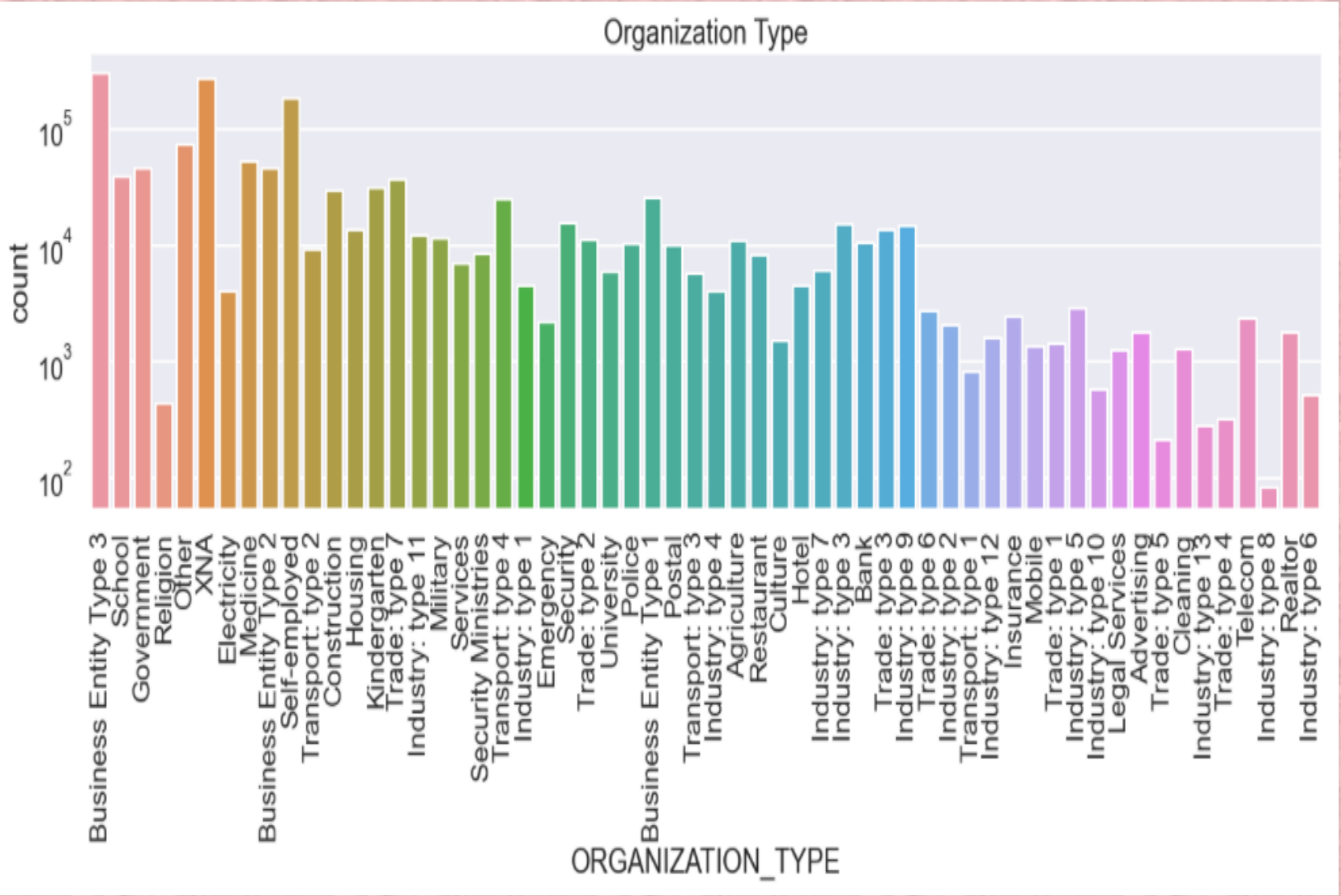
# Segmented Univariate Analysis

# DISTRIBUTION OF GENDER

- From the plot we can infer that Female applicants are more likely to be defaulter than male applicants.


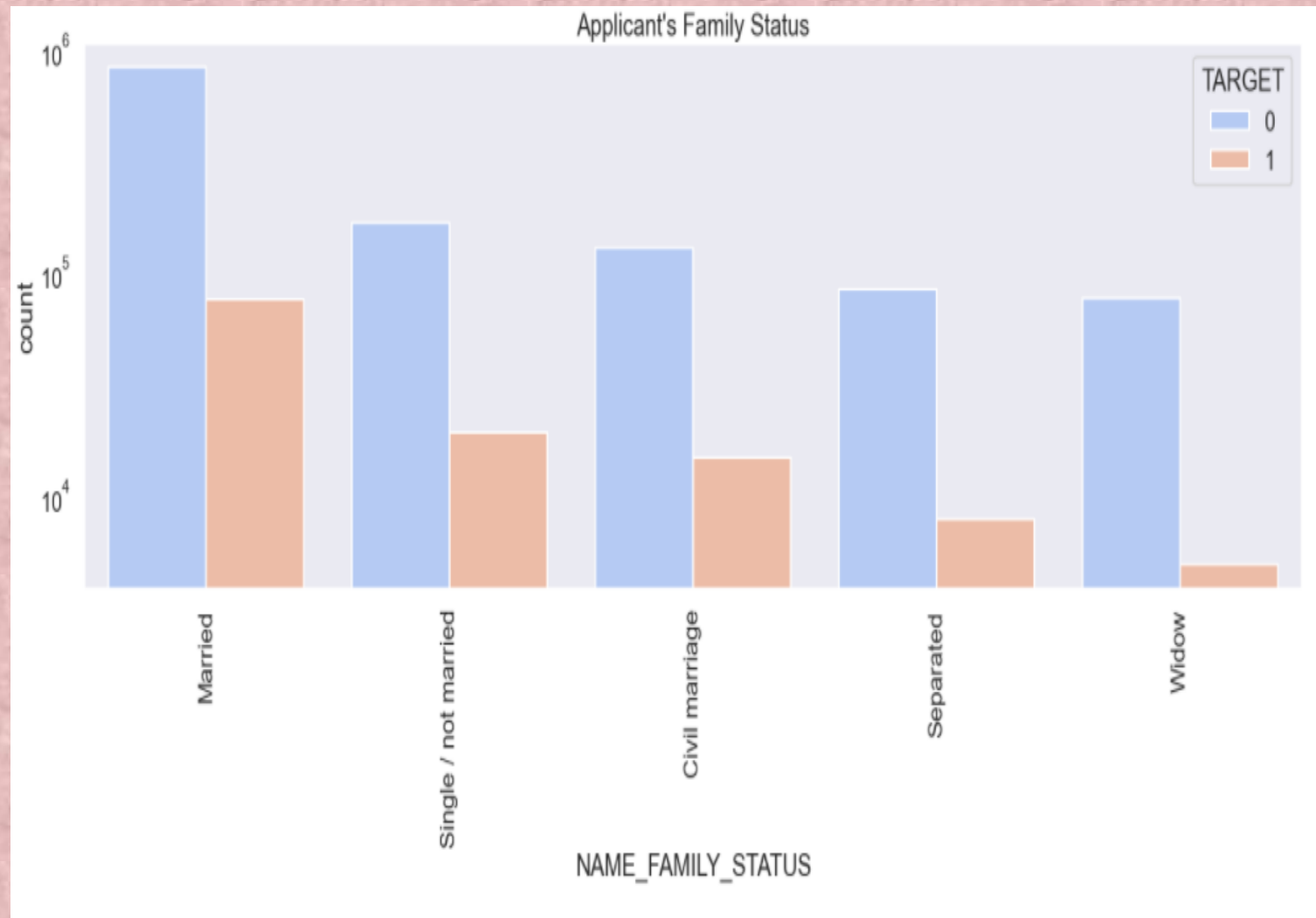- And also we can infer that Female applicants are good repayer of loans than male applicants.

# DISTRIBUTION OF APPLICANT'S INCOME TYPE

•From the plot we can infer that "Working" income type applicants are more likely to be defaulter

• And also we can infer that "Maternity leave" applicants are mostly a defaulter.

•We can infer that "Unemployed" type applicants have highest repayment ratio to that of default.
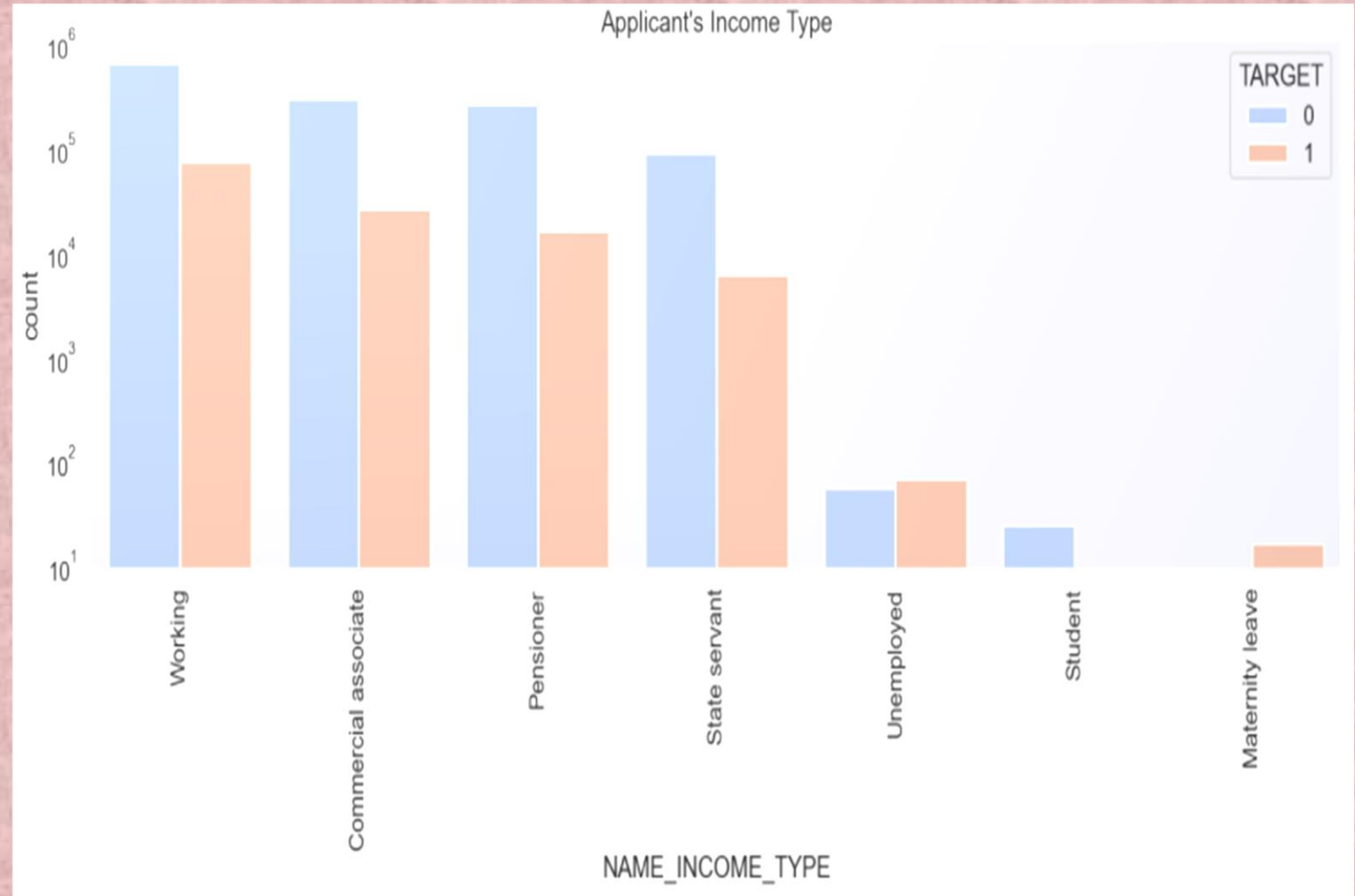
# DISTRIBUTION OF APPLICANT'S FAMILY STATUS

- From the plot we can infer that "Married" applicants are more likely to be defaulter.

- And also we can infer that "Widow" and "Separated" applicants are least defaulter.

# DISTRIBUTION OF APPLICANT'S HOUSING TYPE

- From the plot we can infer that applicants having "House/apartment" are more likely to be defaulter.

- And also we can infer that applicants having "Co-op apartment" are least defaulter.
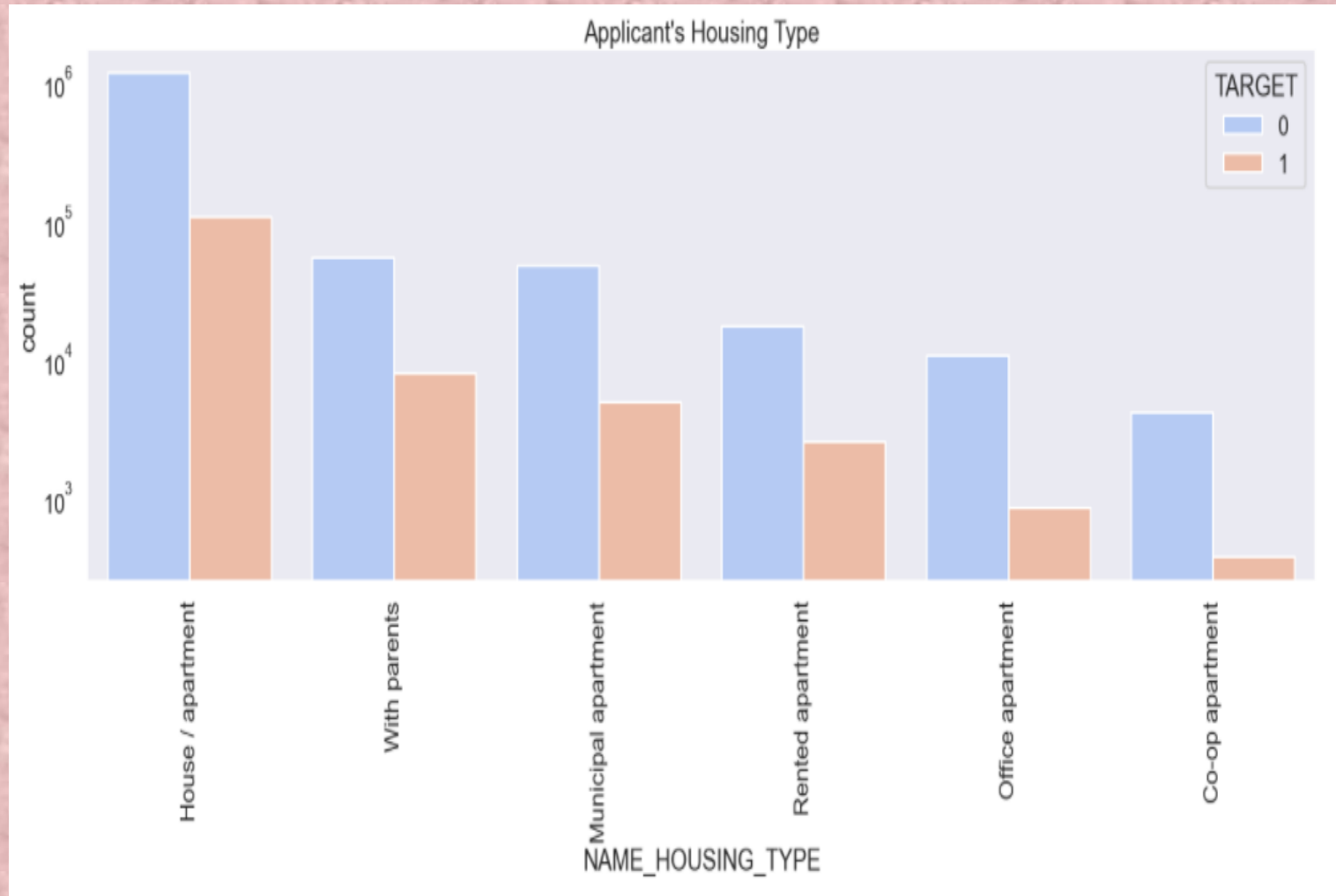
# Bivariate Analysis
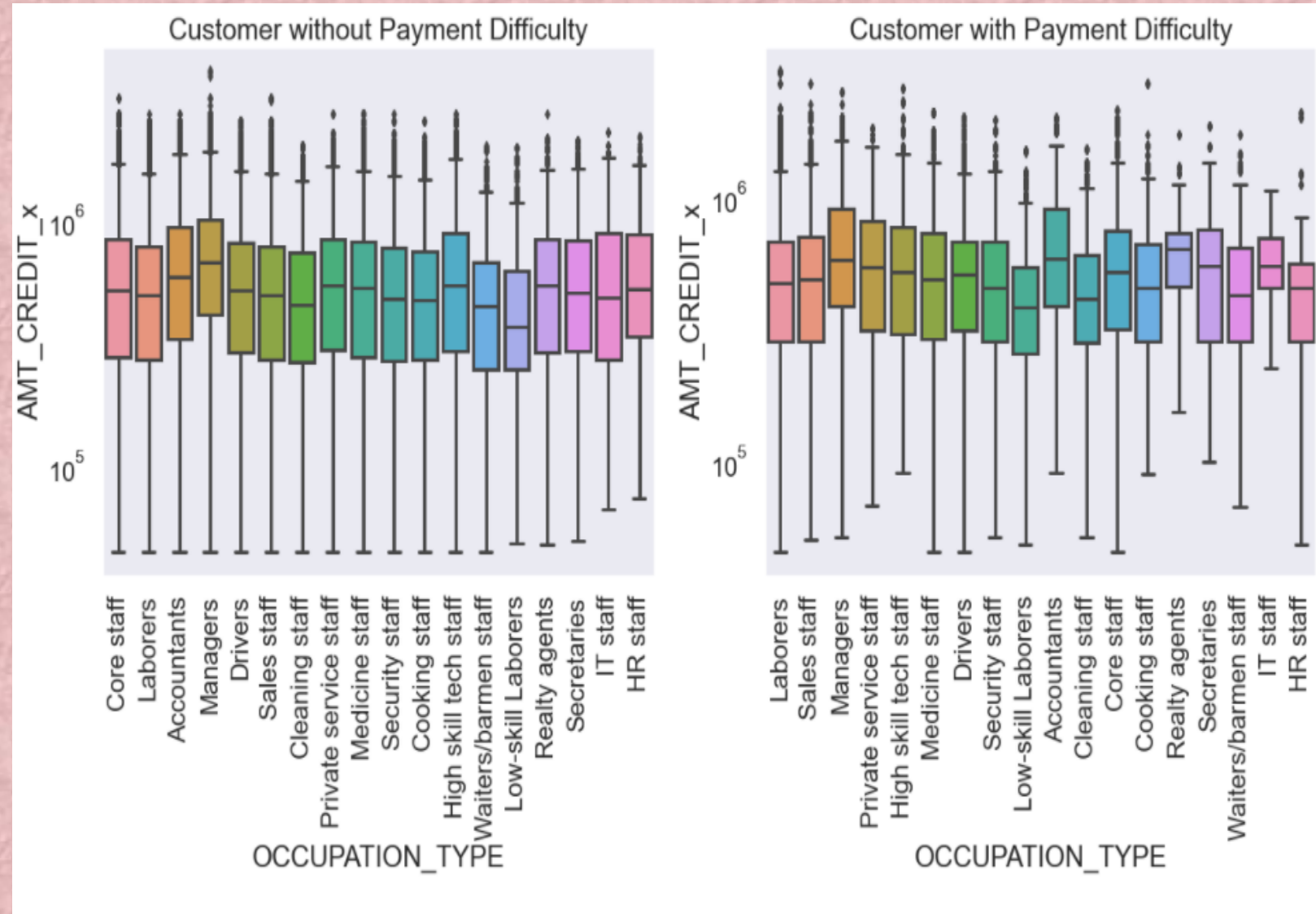
# APPLICANT'S EDUCATION TYPE VS AMOUNT CREDIT

- From the plot we can infer that applicant having academic degree takes more credit amount loans and then faces difficulty repaying them back.

- And also we can infer that applicant having lower secondary education takes less amount of loan and faces less difficulty in repaying them back.

# APPLICANT'S OCCUPATION TYPE VS AMOUNT CREDIT

•From the plot we can infer that Managers, IT staff, Reality agent, Accountants occupation type applicants are top clients that took high credit amount and facing difficulty paying them back.

•And also we can infer that "Security staff" and "Cooking staff" applicants have almost same median.

# APPLICANT'S HOUSING TYPE VS AMOUNT CREDIT

•From the plot we can infer that clients 'with parents' housing type has the lowest credit amount and hence they least likely to default the loan.

•And also we can infer that clients of "office apartment' housing type have highest credit amount in both target 0 and 1 and hence they are most likely to default loan.
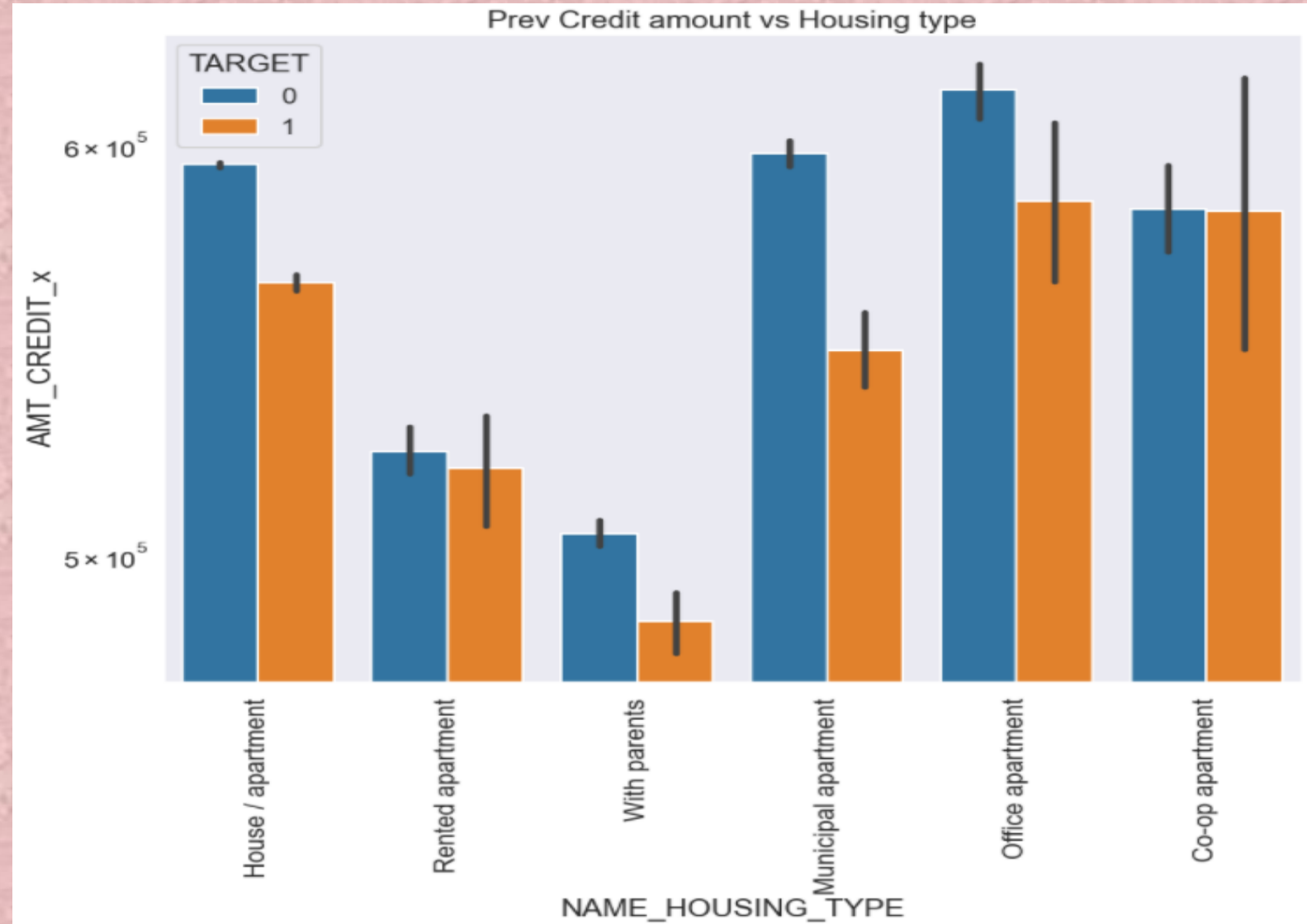
# CONCLUSION

- Bank should focus on applicants from housing type "**With parents**" as they are having least number of unsuccessful payments.

- Banks should focus less on applicants with income type "**Working**" as they are having most number of unsuccessful payments.

- Bank should avoid loans for "**Married**" applicants as they are more likely to be defaulter.

- Bank can avoid loans for applicants with "**academic degree**" as they takes more credit amount loans and then faces difficulty repaying them back or Bank can provide lesser credit amount loans with increased in interest rate.

- Bank can focus on applicants with "**lower secondary**" education as they takes less amount of loan and faces less difficulty in repaying them back.

- Bank can avoid loans for applicants from housing type "**office apartment**" as they are having highest credit amount and hence they are most likely to default loan.