



# CREDIT EDA CASE STUDY

PRESENTED BY - BHUVANA CHANDRA

# TASKS PREFORMED ON DATA SET

## **TASK 1:**

- IMPORTING CSV FILES.
- a) application\_data.csv
- b) Previous\_application.csv

## **TASK 2:**

- Checking the data its shape, info and describe

## **TASK 3:**

- Data cleaning
- Inspect Null values
- Impute Null values with mean or median
- Binning of continuous variables

# TASKS PERFORMED ON DATA SET

## **TASK 4:**

- Imbalance percentage for target0 and target1
- Splitting the data into two sets using target
- Univariate and Bivariate analysis on 0 and 1
- Correlation using heatmap

## **TASK 5:**

- Merging both data frames
- Univariate analysis and Bivariate analysis on merged data

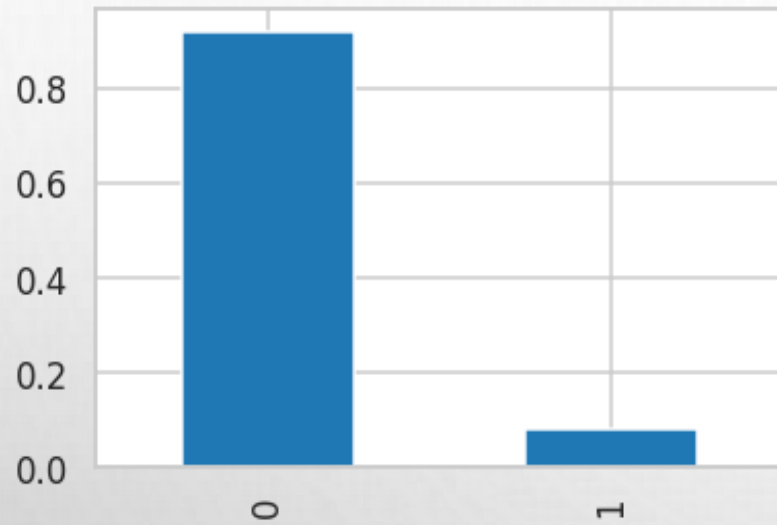
## **CONCLUSION**

## DETAILS ABOUT TASK 3

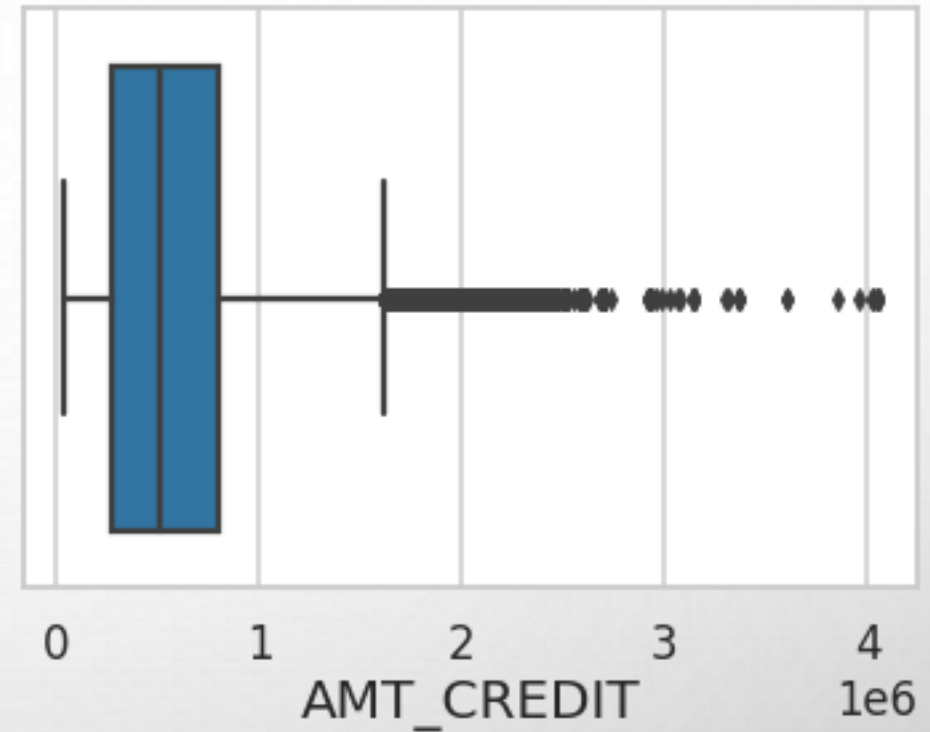
- Data quality check and find out null values
- Sum and percentage of null values
- If the percentage of null values are greater than 50% then drop the values in the dataset as they are no way related target variable
- If the null values percentage is less than 13% then replace the values with median(numerical values) or mean(categorical values)
- Finding outliers in the numerical values
- Binning of continuous variable

# TARGET CONSIDERATION AND AMT\_CREDIT OUTLIERS

Target Consideration



Target consideration bar plot  
Here we can see that target 0 is higher than  
Target 1



Here is the outliers of AMT\_CREDIT

The background of the slide is a light gray gradient, decorated with numerous realistic water droplets of various sizes. Some droplets are large and prominent, while others are small and subtle. They are scattered across the slide, with a higher concentration in the top-left and bottom-right corners, creating a clean, modern, and fresh aesthetic.

# **CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET 0**

# DISTRIBUTION OF INCOME RANGE

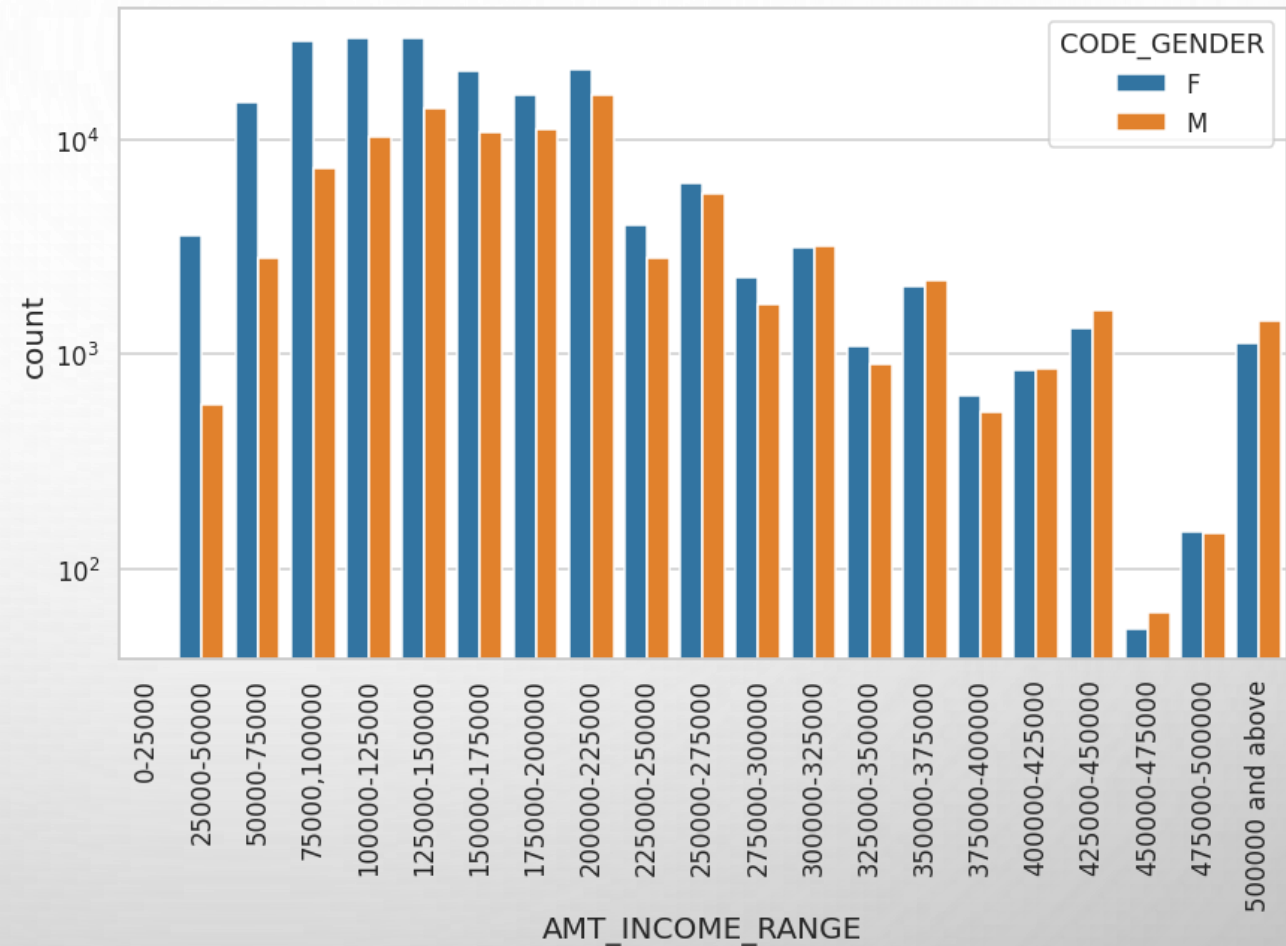
WHAT WE OBSERVED FROM THE PLOT?

1. INCOME RANGE FROM 25000 - 225000  
FEMALES ARE MORE THAN MALES

2. INCOME RANGE FROM 225000 AND  
ABOVE FEMALES AND MALES ARE ALMOST  
EQUAL IN NUMBER

3. VERY LESS COUNT FOR INCOME RANGE  
400000 AND ABOVE.

Income Range Distribution w.r.t Gender



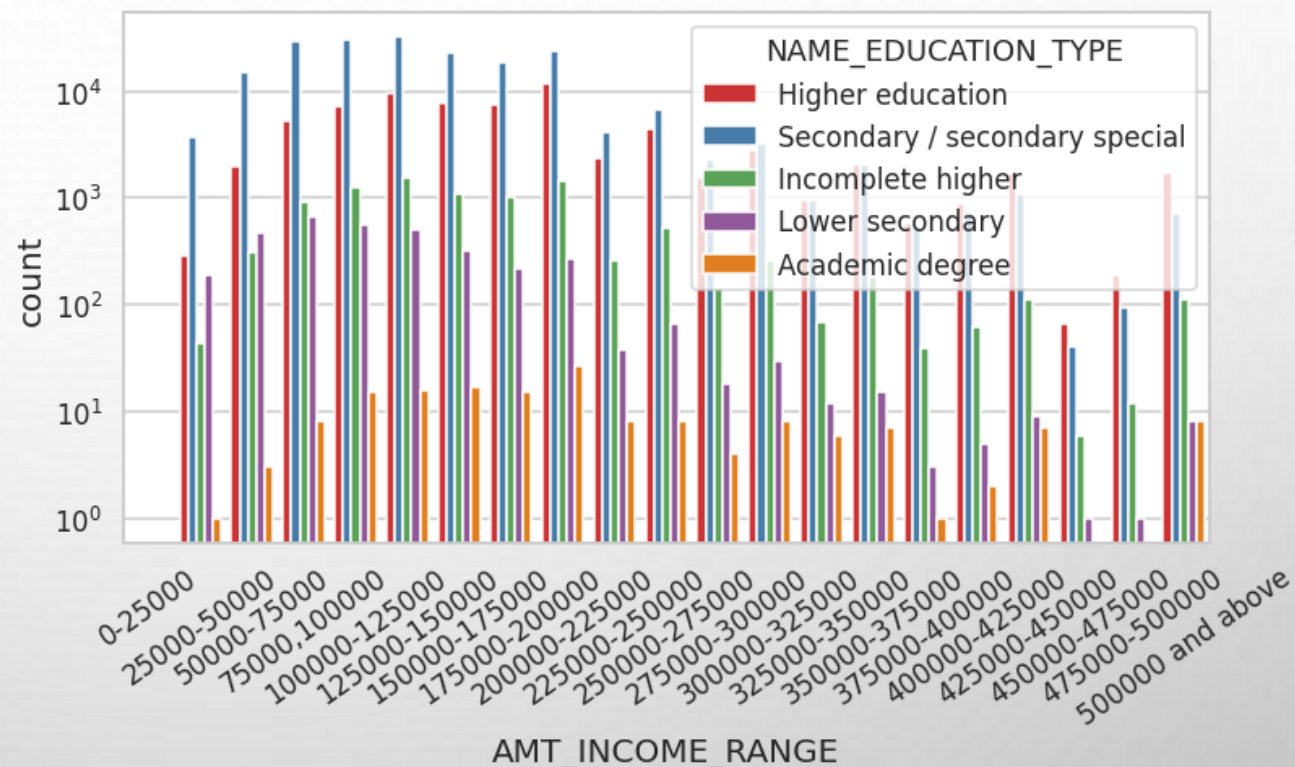


# DISTRIBUTION OF INCOME RANGE W.R.T EDUCATION

WHAT WE OBSERVED FROM THE PLOT?

- THE ABOVE PLOT INDICATES INCOME OF CLIENTS w.r.t THEIR EDUCATION AND IS OBSERVED THAT CLIENTS WHO ARE HAVING SECONDARY EDUCATION ARE MORE IN NUMBER FROM INCOME 0-225000
- CLIENTS WHO HAVE STUDIED HIGHER EDUCATION ARE IN THE INCOME RANGE OF 300000 ABOVE
- CLIENTS WHO HAVE ACADEMIC DEGREE ARE MORE IN INCOME RANGE 175000 -200000
- CLIENTS WHO HAVE STUDIED LOWER SECONDARY ARE MORE IN INCOME RANGE 50000 -75000

Income Range Distribution w.r.t Education



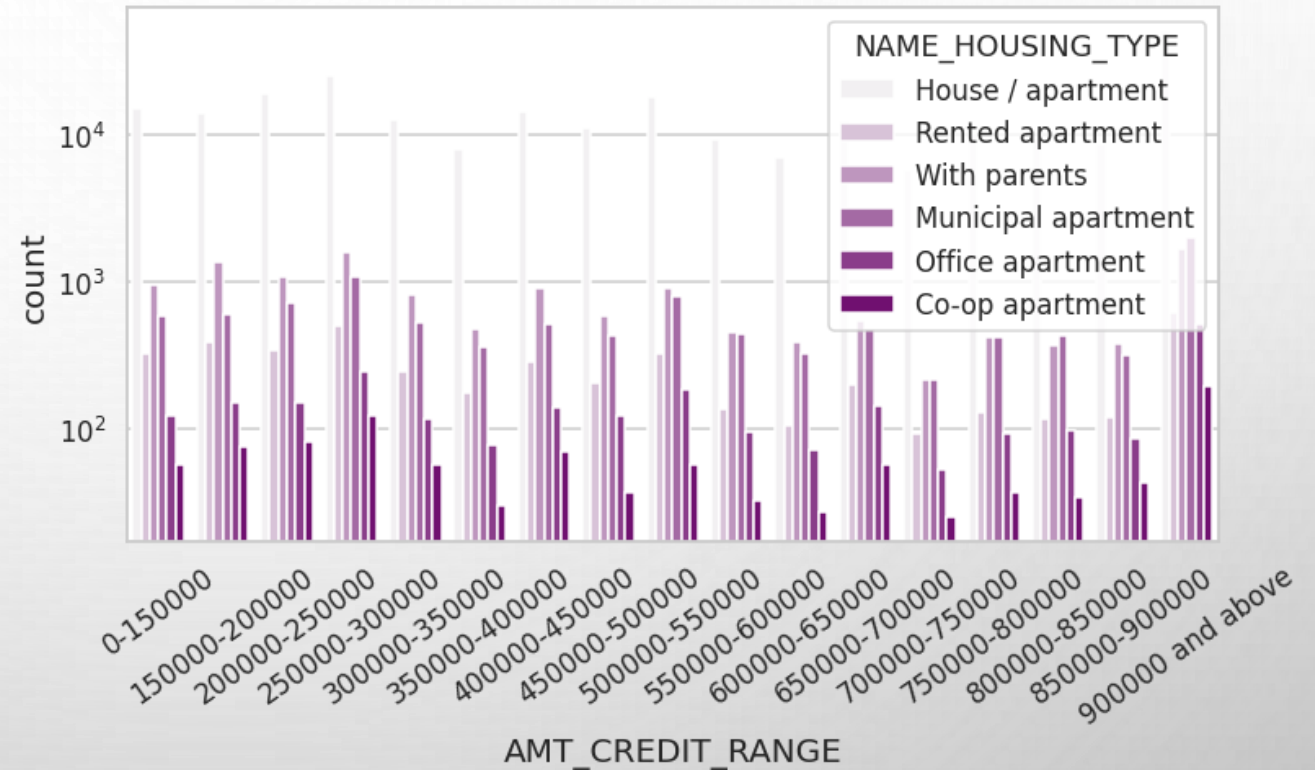


# DISTRIBUTION OF CREDIT RANGE

WHAT WE OBSERVED FROM THE PLOT?

- IT IS OBSERVED THAT CLIENTS HAVING HOUSE/APARTMENT ARE HAVING MORE CREDITS THAN OTHER HOUSING TYPES
- WE CONCLUDE THAT THEY HAVE NO ISSUES IN LOAN REPAYMENT

Credit Range Distribution w.r.t Housing Type

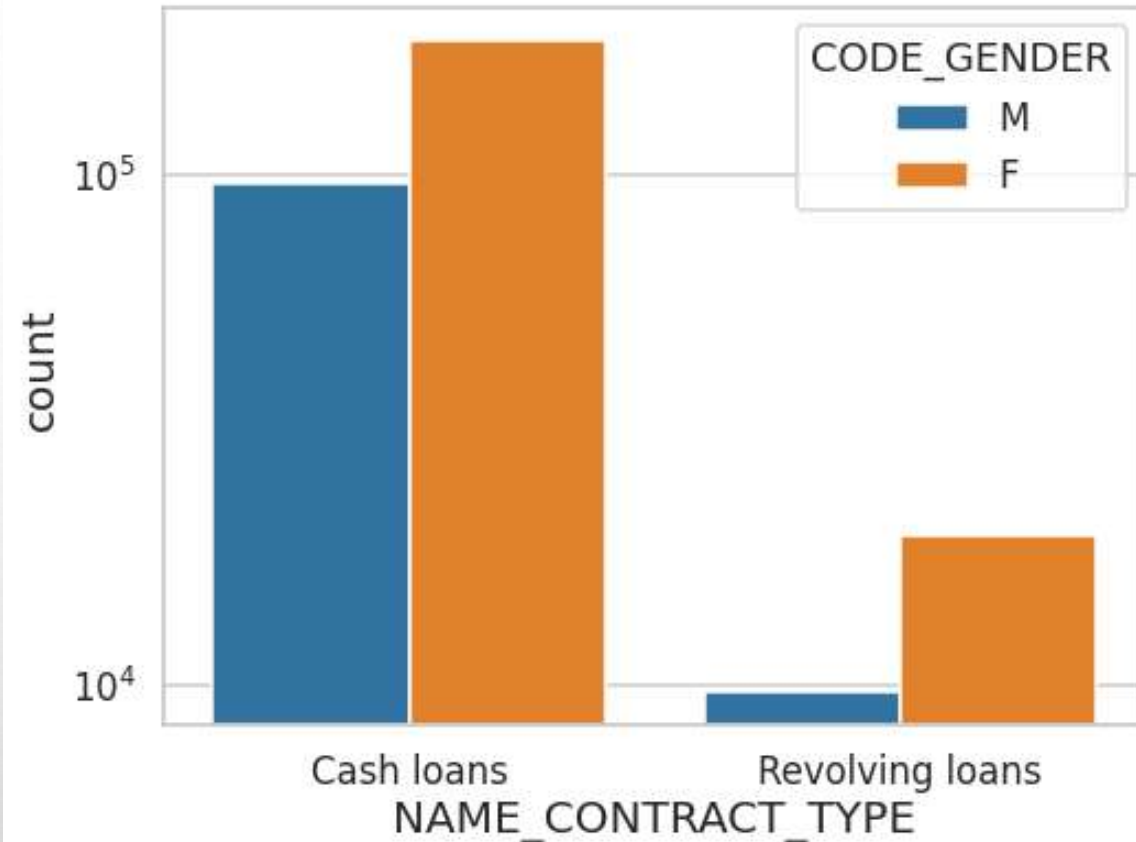


# DISTRIBUTION OF CONTRACT TYPE

WHAT WE OBSERVED FROM THE PLOT?

- IT IS OBSERVED THAT NAME\_CONTRACT\_TYPE RANGE DISTRIBUTION W.R.T GENDER THAT COUNT OF FEMALES ARE MORE THAN MALES AND THEY HAVE NO ISSUES FOR LOAN REPAYMENT

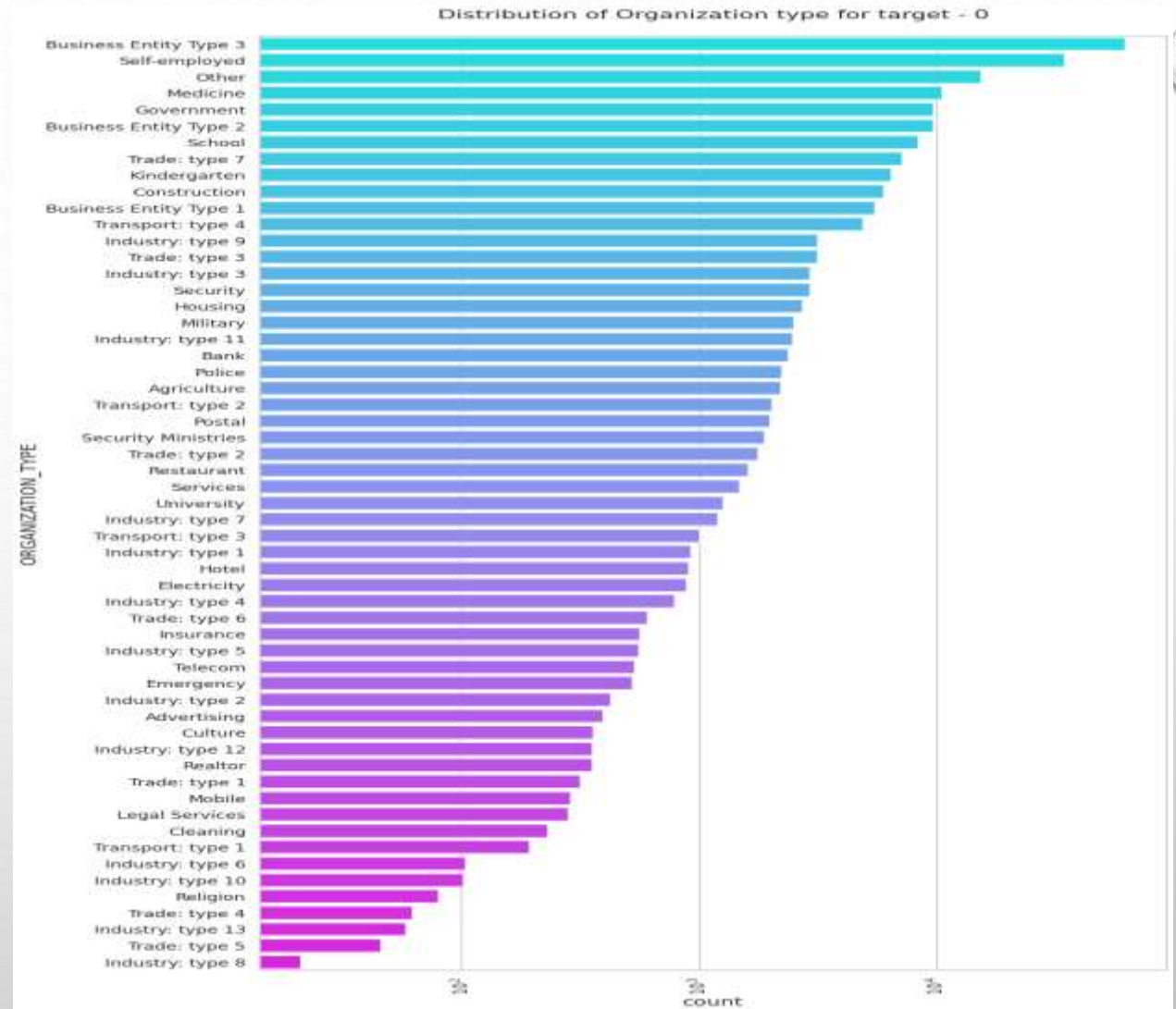
Name Contract Type Range Distribution w.r.t Gender



# DISTRIBUTION OF ORGANIZATION TYPE

WHAT WE OBSERVED FROM THE PLOT?

- IT IS OBSERVED THAT CLIENTS WHO HAVE APPLIED FOR CREDITS ARE FROM MOST OF THE ORGANIZATION TYPE: BUSINESS ENTITY TYPE3, SELF EMPLOYED, OTHER, MEDICINE AND GOVERNMENT.
- LESS CLIENTS ARE FROM INDUSTRY TYPE6, TYPE 10, RELIGION AND TRADE TYPE 5, TYPE4.



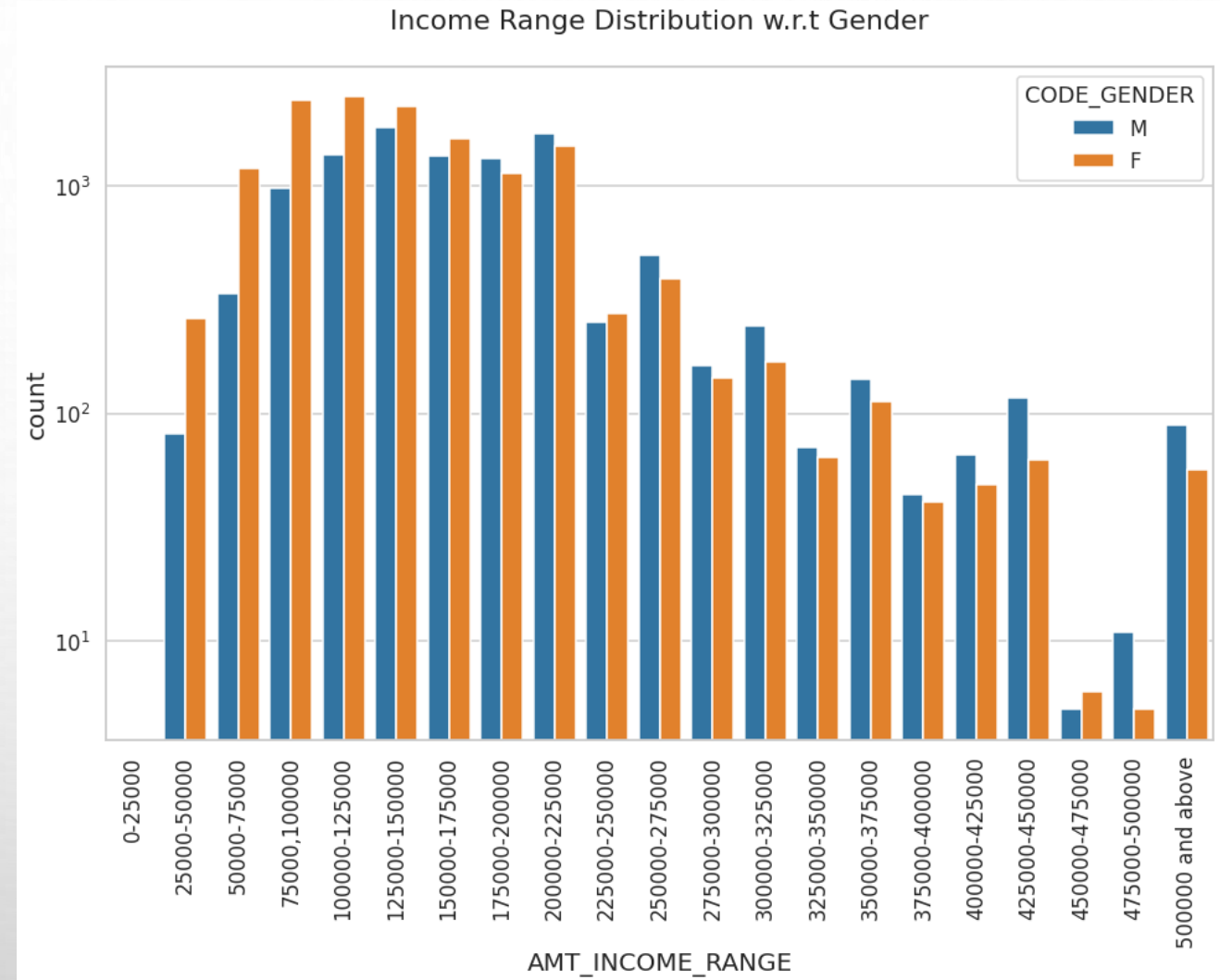


# CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET 1

# DISTRIBUTION OF INCOME TYPE

## WHAT WE OBSERVED FROM THE PLOT

- INCOME RANGE FROM 25000-175000 FEMALES ARE MORE IN NUMBER THAN MALE CLIENTS AND INCOME RANGE FROM 175000 AND ABOVE MALES ARE MORE IN NUMBER THAN FEMALES
- BOTH FEMALES AND MALE CLIENTS ARE HAVING PAYMENT ISSUES FROM INCOME RANGE 175000 AND ABOVE

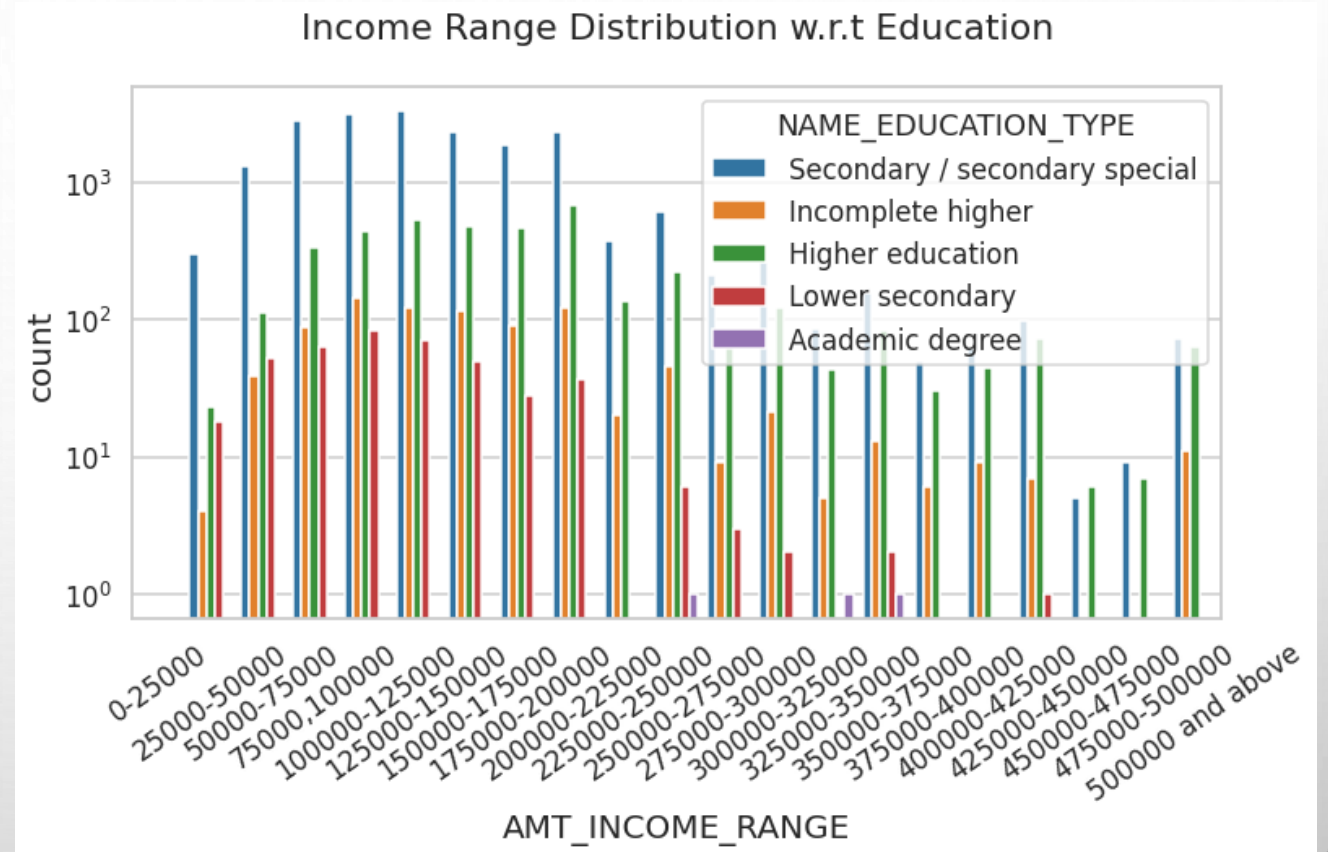




# DISTRIBUTION OF INCOME TYPE W.R.T EDUCATION

WHAT WE OBSERVED FROM THE PLOT

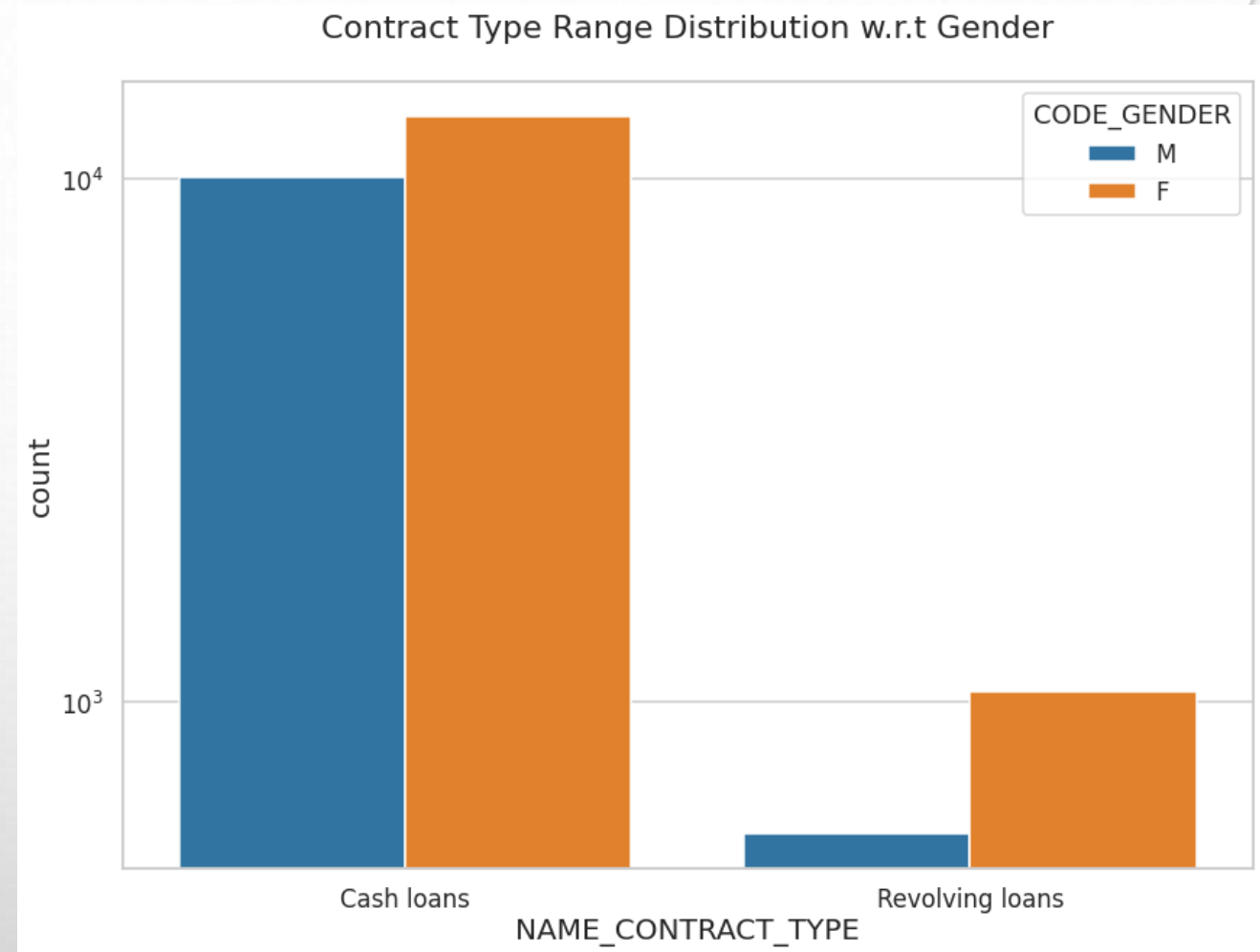
- IT IS OBSERVED THAT CLIENTS HAVING SECONDARY EDUCATION ARE MORE IN NUMBER IN TARGET1 DATA FRAME.



# DISTRIBUTION OF CONTRACT TYPE

WHAT WE OBSERVED FROM THE PLOT

- IT IS OBSERVED THAT FEMALE COUNTS ARE MORE IN LOAN TYPES

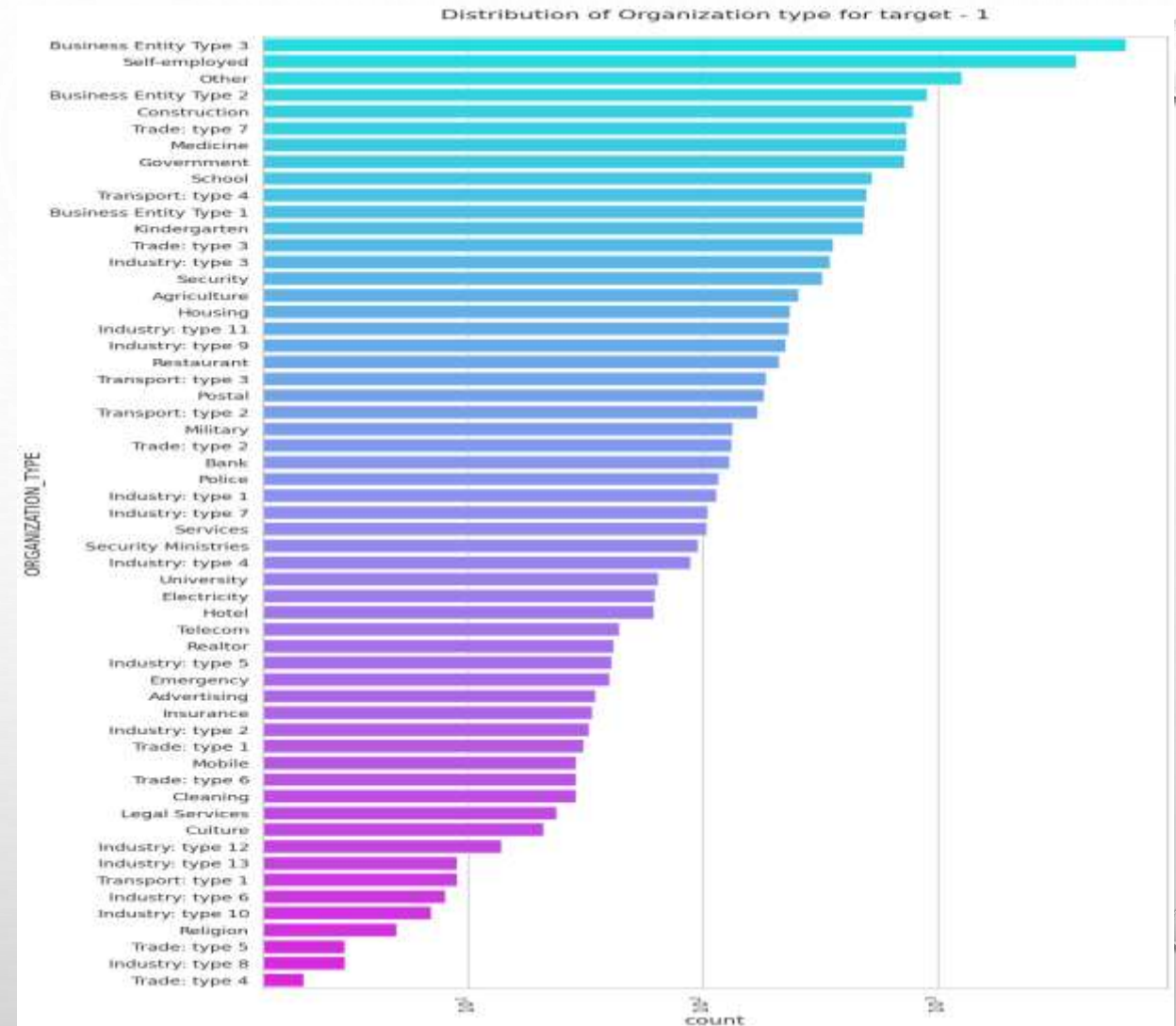




# DISTRIBUTION OF ORGANIZATION TYPE

## WHAT WE CONCLUDED FROM 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 OF TARGET 0



- AS WE CAN SEE FROM ABOVE CORRELATION HEATMAP, THERE ARE NUMBER OF OBSERVATION WE CAN POINT OUT:

1.CREDIT AMOUNT IS INVERSELY PROPORTIONAL TO THE DATE OF BIRTH, WHICH MEANS CREDIT AMOUNT IS HIGHER FOR LOW AGE AND VICE-VERSA.

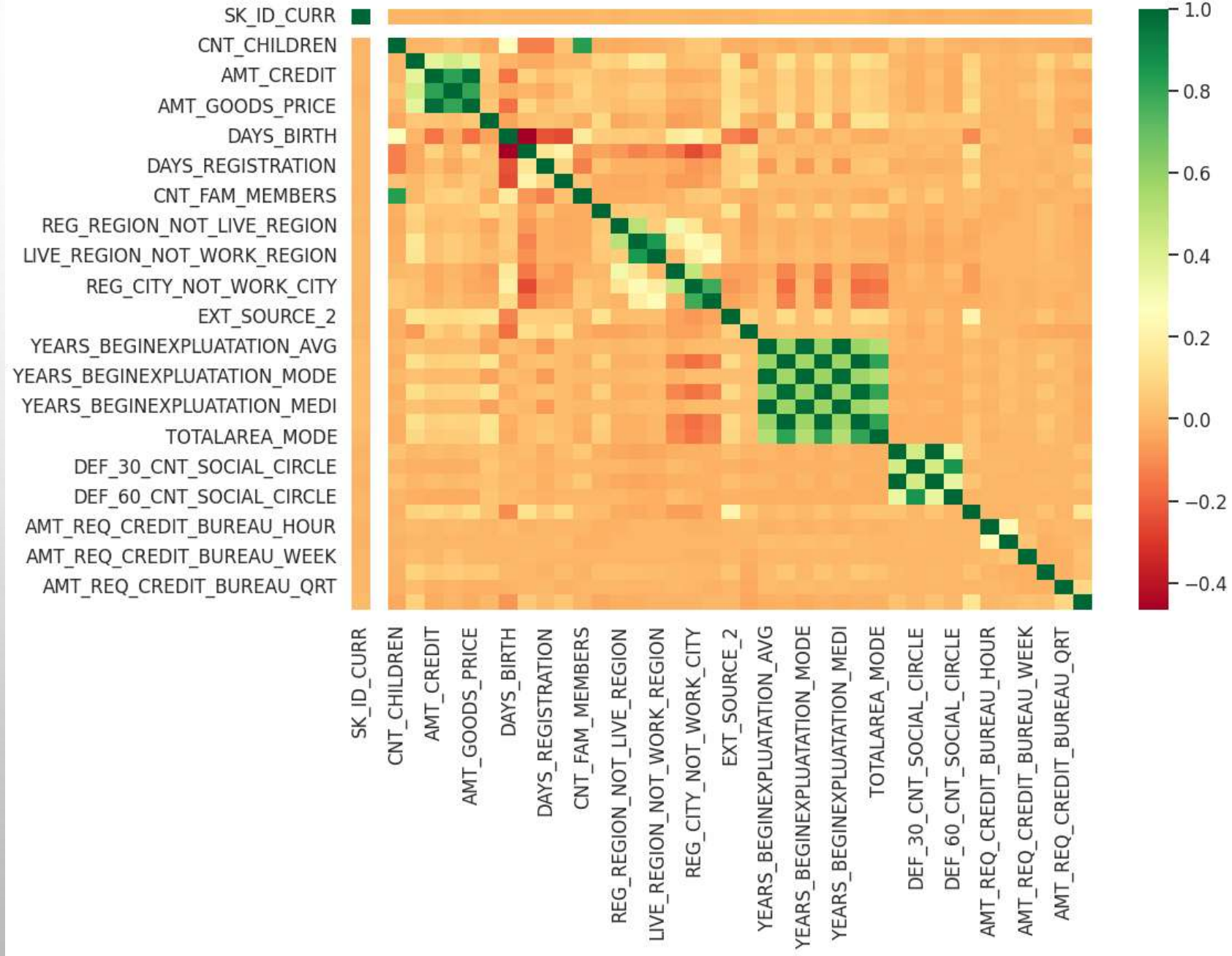
2.CREDIT AMOUNT IS INVERSELY PROPORTIONAL TO THE NUMBER OF CHILDREN CLIENT HAVE, MEANS CREDIT AMOUNT IS HIGHER FOR LESS CHILDREN COUNT CLIENT HAVE AND VICE-VERSA.

3.CREDIT AMOUNT IS HIGHER TO DENSELY POPULATED AREA.

4.THE INCOME IS ALSO HIGHER IN DENSELY POPULATED ARE



# Correlation for target 1



# CORRELATION FOR TARGET 1

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

The slide features a light gray background with a subtle radial gradient. In the top-left and bottom-right corners, there are clusters of realistic water droplets of various sizes, rendered with soft shadows and highlights to give them a three-dimensional appearance. Centered in the upper half of the slide is a faint, circular watermark containing a stylized geometric pattern.

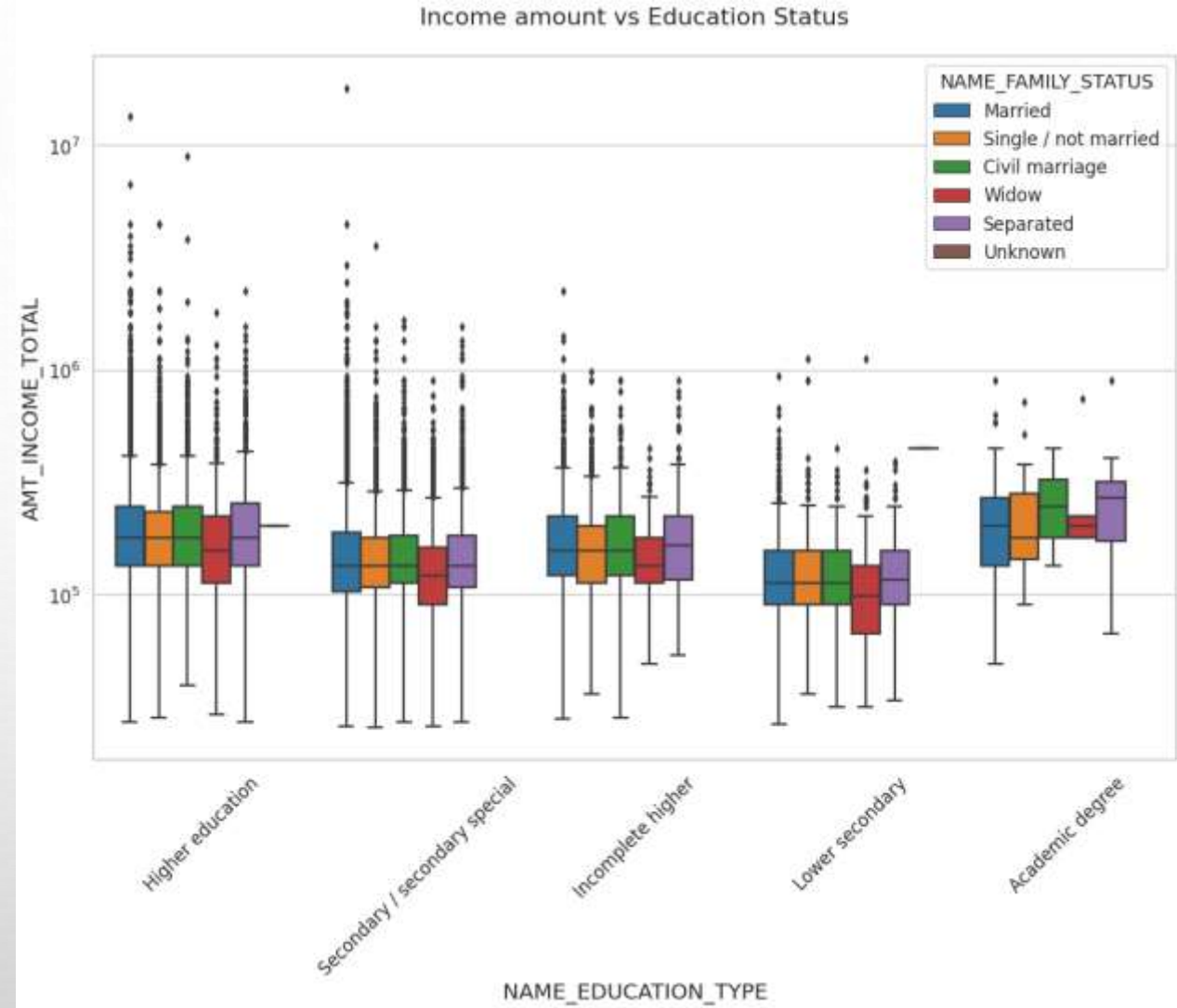
# BIVARIATE ANALYSIS FOR TARGET 0



# INCOME AMOUNT VS EDUCATION STATUS

WHAT WE OBSERVED FROM THE PLOT

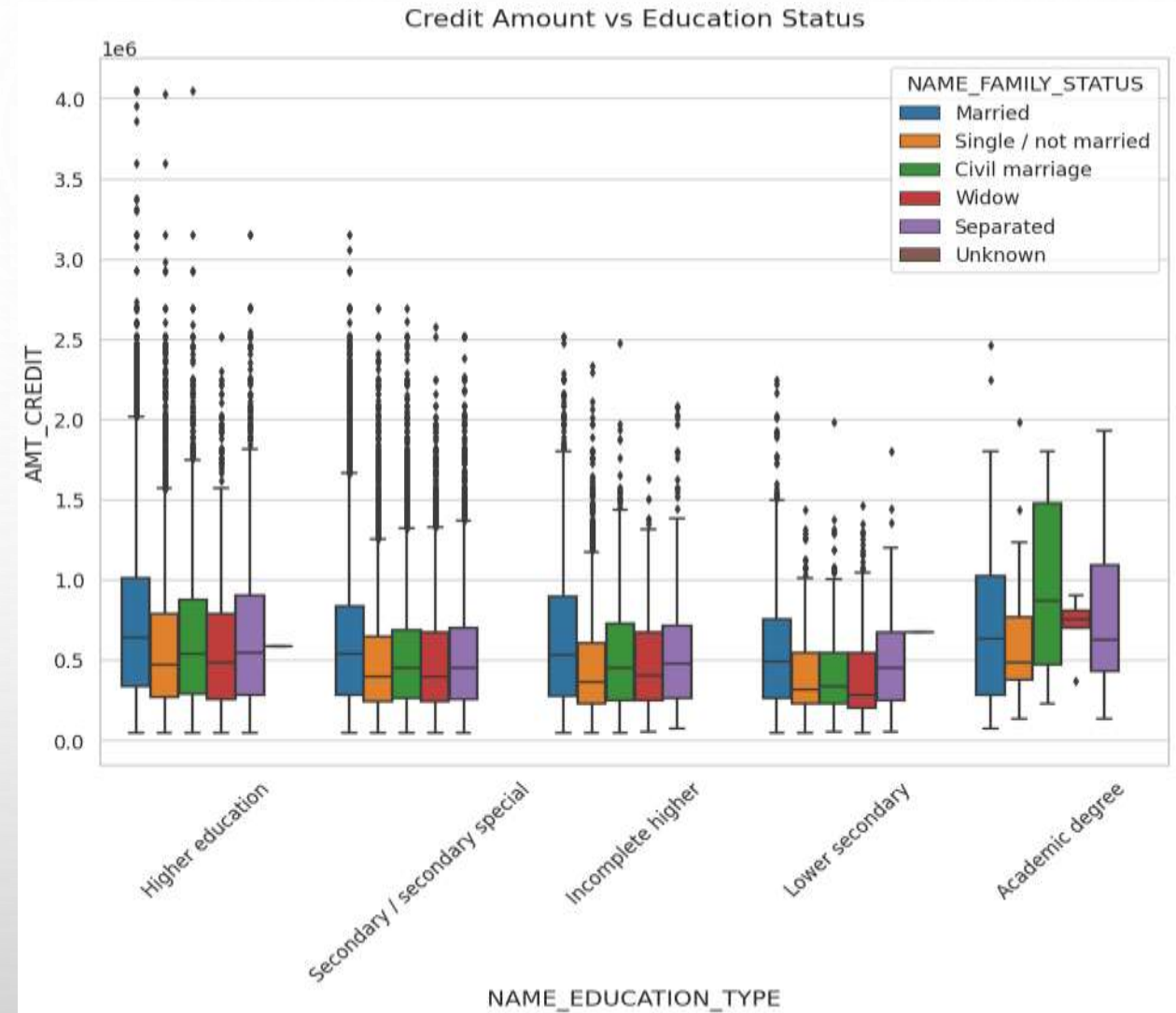
- IT IS OBSERVED THAT CLIENTS WITH HIGHER EDUCATION ,THE INCOME AMOUNT IS MOSTLY EQUAL WITH FAMILY STATUS
- LOWER SECONDARY OF CIVIL MARRIAGE FAMILY STATUS ARE HAVE LESS INCOME AMOUNT THAN OTHERS.



# CREDIT AMOUNT VS EDUCATION STATUS

## WHAT WE OBSERVED FROM THE PLOT

- IT IS OBSERVED THAT CLIENTS HAVING FAMILY STATUS OF 'CIVIL MARRIAGE', 'MARRIAGE' AND 'SEPARATED' OF ACADEMIC DEGREE EDUCATION ARE HAVING HIGHER NUMBER OF CREDITS THAN OTHERS.
- AND ALSO, CLIENTS HAVING HIGHER EDUCATION OF FAMILY STATUS OF 'MARRIAGE', 'SINGLE' AND 'CIVIL MARRIAGE' ARE HAVING MORE OUTLIERS.



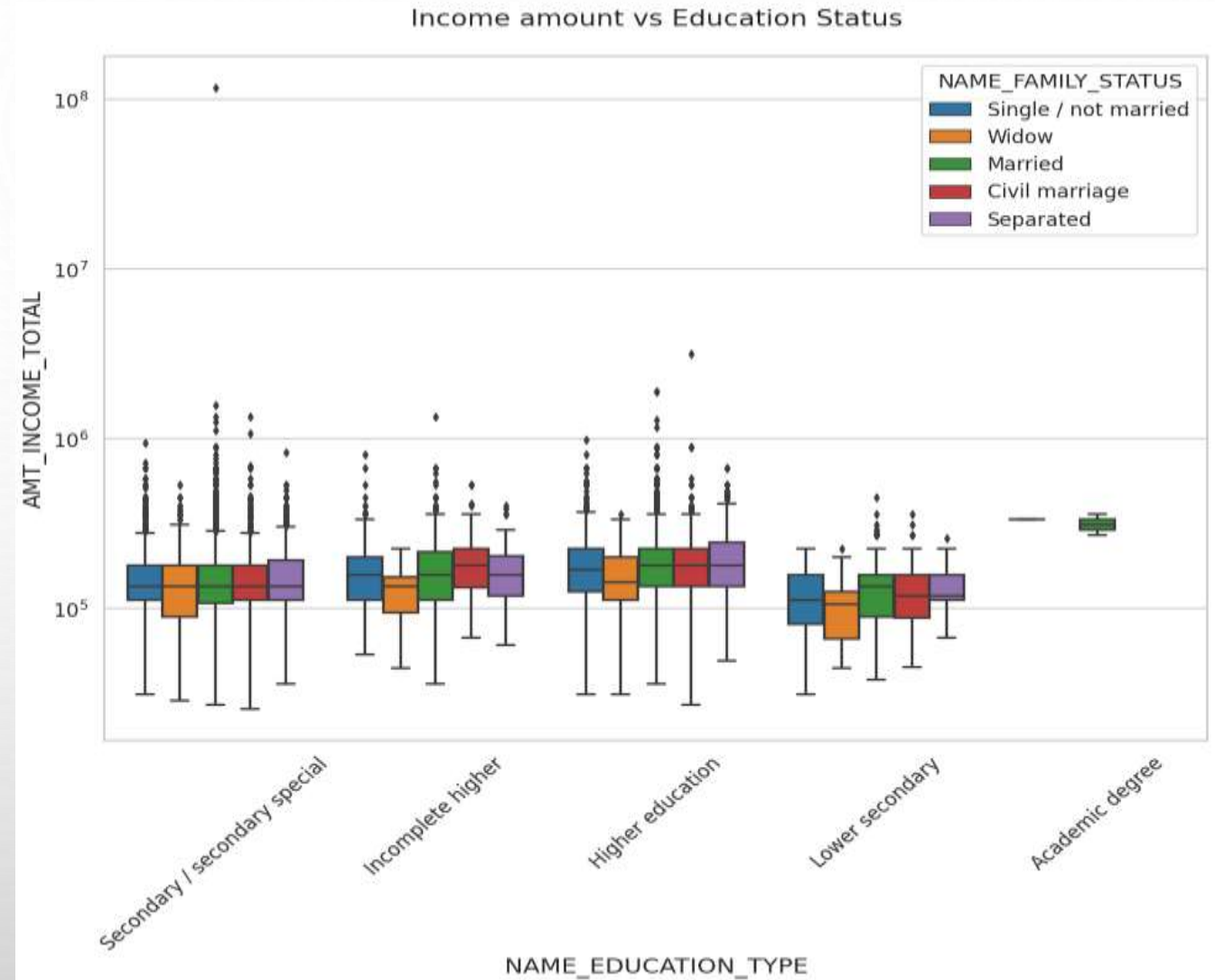


# BIVARIATE ANALYSIS FOR TARGET 1

# INCOME AMOUNT VS EDUCATION STATUS

## WHAT WE OBSERVED FROM THE PLOT

- IT IS OBSERVED THAT THE PLOT INDICATES THAT HIGHER EDUCATED (FAMILY STATUS: MARRIED, WIDOW, SINGLE, CIVIL MARRIAGE AND SEPARATED) CLIENTS HAVING ALMOST SAME INCOME STATUS IN TARGET '1' DATA FRAME.
- IT IS OBSERVED THAT THE PLOT INDICATES THAT SECONDARY EDUCATION(FAMILY STATUS: MARRIED,WIDOW,CIVIL,MARRIAGE,SEPARATED ,SINGLE) CLIENTS ARE ALMOST HAVING SAME MEDIAN VALUE

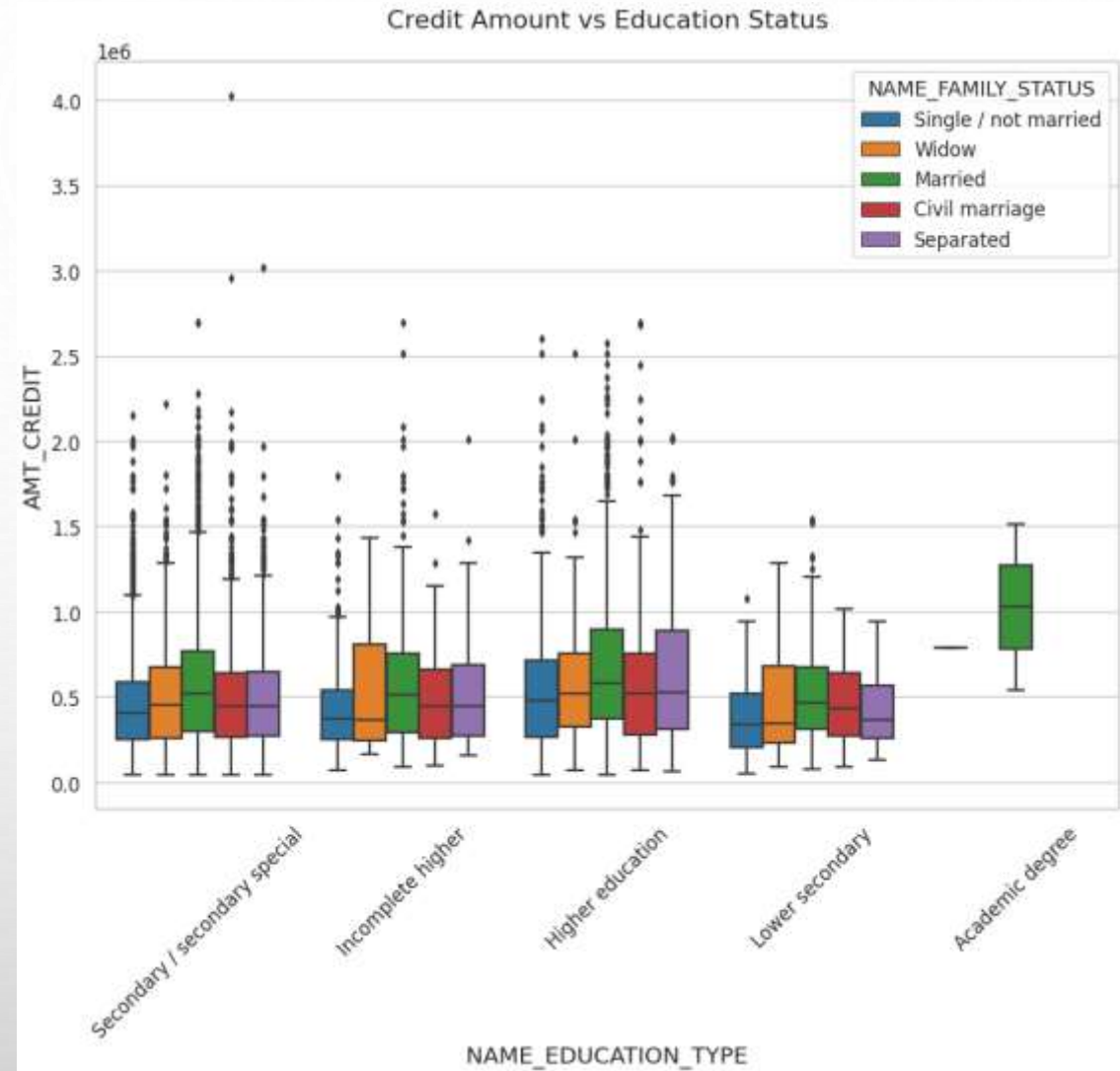




# CREDIT AMOUNT VS EDUCATION STATUS

WHAT WE OBSERVED FROM THE PLOT?

- IT IS OBSERVED THAT CLIENTS HAVING ACADEMIC DEGREE HAVING MORE NO OF CREDITS.
- IN THE SECONDARY EDUCATION AND INCOMPLETE EDUCATION FAMILY STATUS (MARRIED AND SEPARATED) ARE LIKELY IN THE SAME CREDIT COUNT.

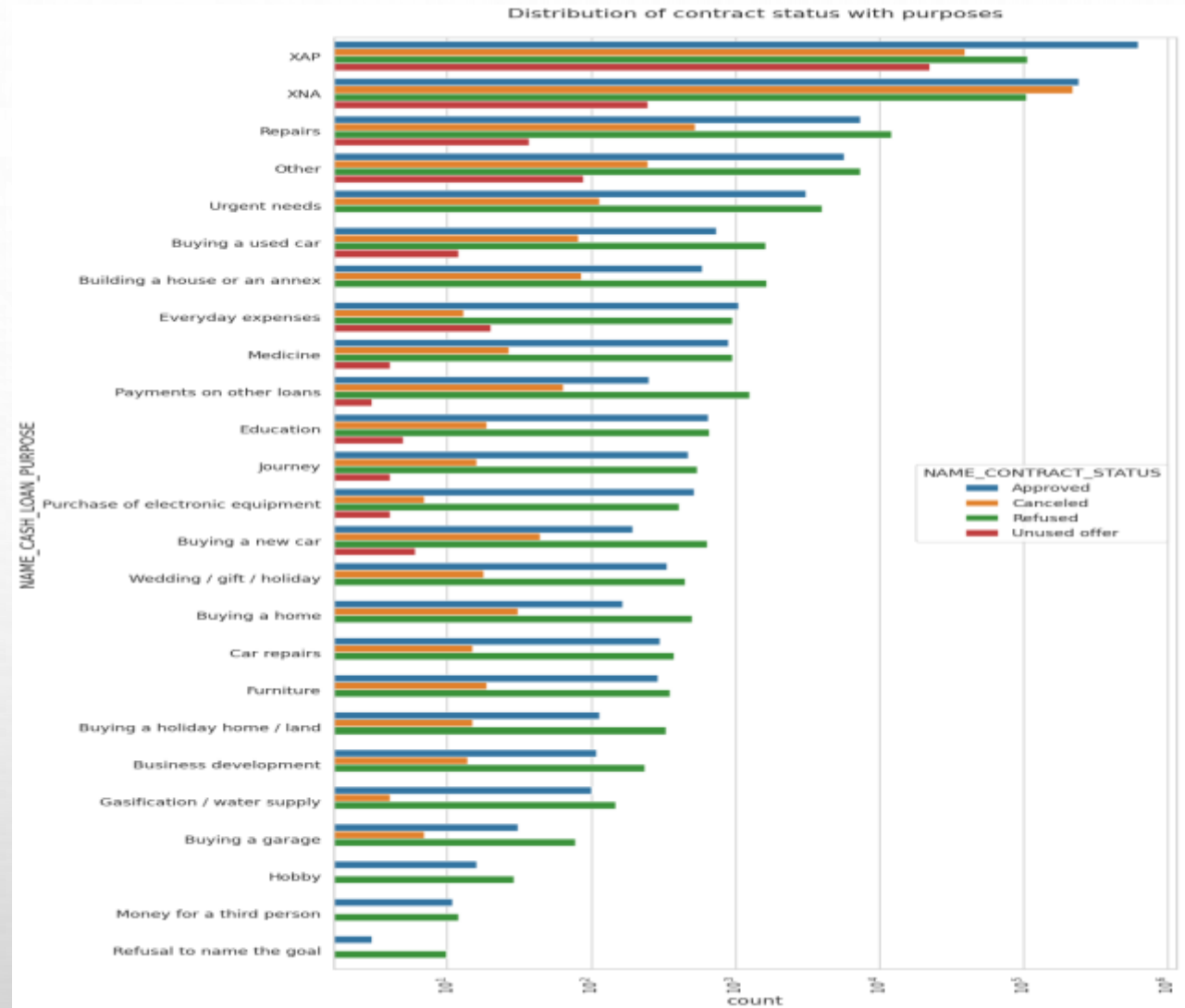


The background of the slide is a light gray gradient. It is decorated with several realistic water droplets of various sizes, some with highlights and shadows, located in the top-left, top-right, and bottom-right corners.

# UNIVARIATE ANALYSIS AFTER MERGING PREVIOUS DATA

# DISTRIBUTION OF CONTRACT STATUS WITH PURPOSES

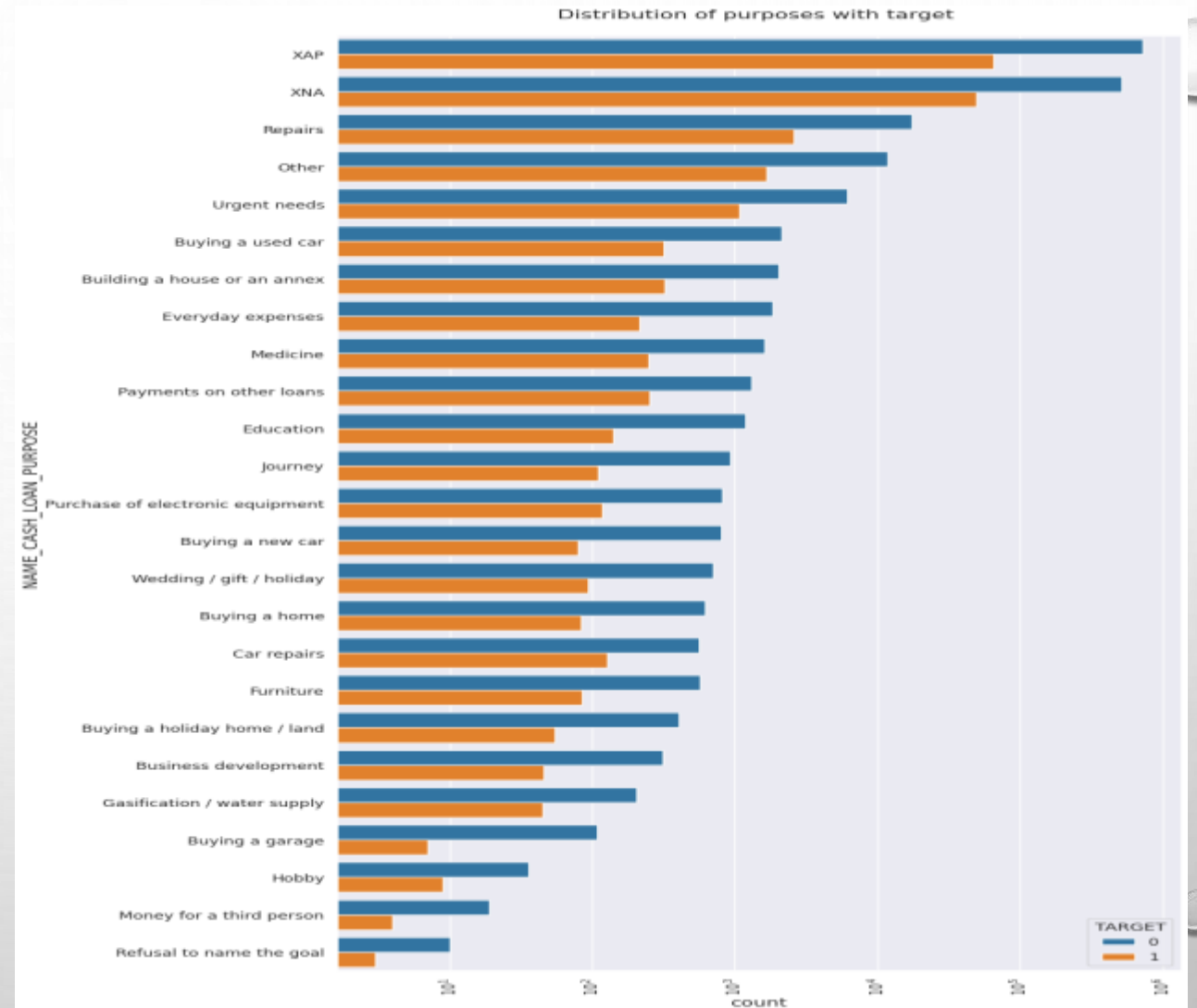
- RESULTS OBSERVED FROM UNIVARIATE ANALYSIS AFTER MERGING.
- XAP & XNA ARE MORE IN NUMBER OF APPROVED LOANS.
- MOST REJECTION OF LOANS CAME FROM REPAIRS PURPOSE.
- FOR EDUCATION PURPOSE BOTH APPROVED AND REJECTED ARE ALMOST EQUAL IN NUMBER AND SIMILARLY IN JOURNEY PURPOSE ALSO.
- PAYMENT ON OTHER LOANS AND BUYING A NEW CAR HAS MORE REJECTIONS THAN APPROVAL.





# DISTRIBUTION OF PURPOSES WITH TARGET

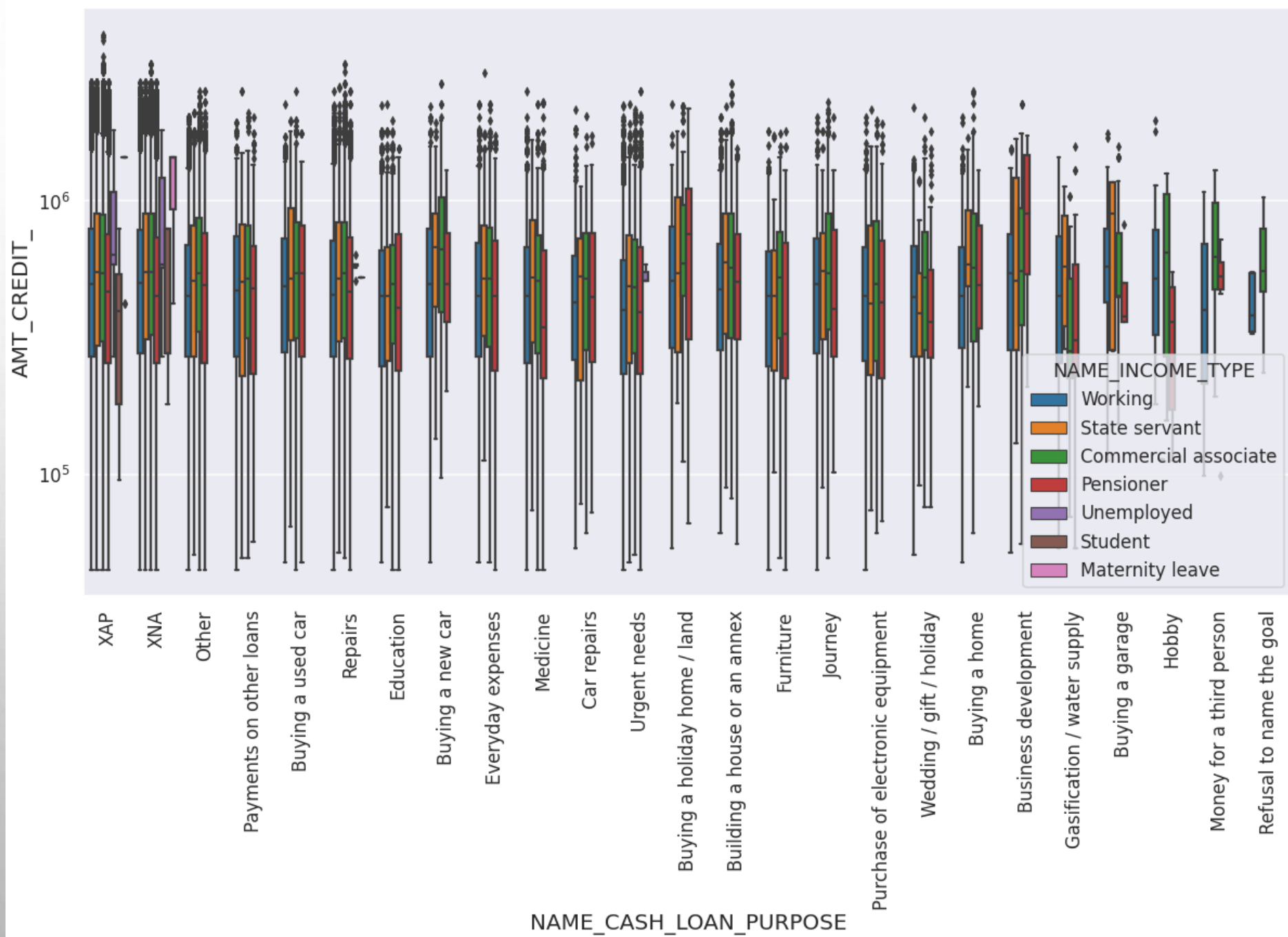
- RESULTS OBSERVED FROM UNIVARIATE ANALYSIS AFTER MERGING
- IN TARGET 0 & 1 XAP IS MORE IN NUMBER
- LOAN PURPOSES WITH REPAIRS ARE FACING MORE DIFFICULTIES IN LOAN REPAYMENT
- BUYING A GARAGE PURPOSE IS GOOD AND NOT FACING DIFFICULTIES
- SIMILARLY IN 'BUYING LAND', 'BUSINESS DEVELOPMENT', 'BUYING NEW CAR', 'EDUCATION' ARE FACING LESS DIFFICULTIES.
- HENCE BANK SHOULD CONCENTRATE ON THESE PURPOSE AS THERE ARE MINIMAL PAYMENT DIFFICULTIES



The slide features a light gray background with a subtle radial gradient. In the top-left and bottom-right corners, there are clusters of realistic water droplets of various sizes, rendered with soft shadows and highlights to give them a three-dimensional appearance. Centered on the slide is the title text.

# PERFORMING BIVARIATE ANALYSIS

Previous Credit amount vs Loan Purpose



# PREVIOUS CREDIT AMOUNT VS LOAN PURPOSE

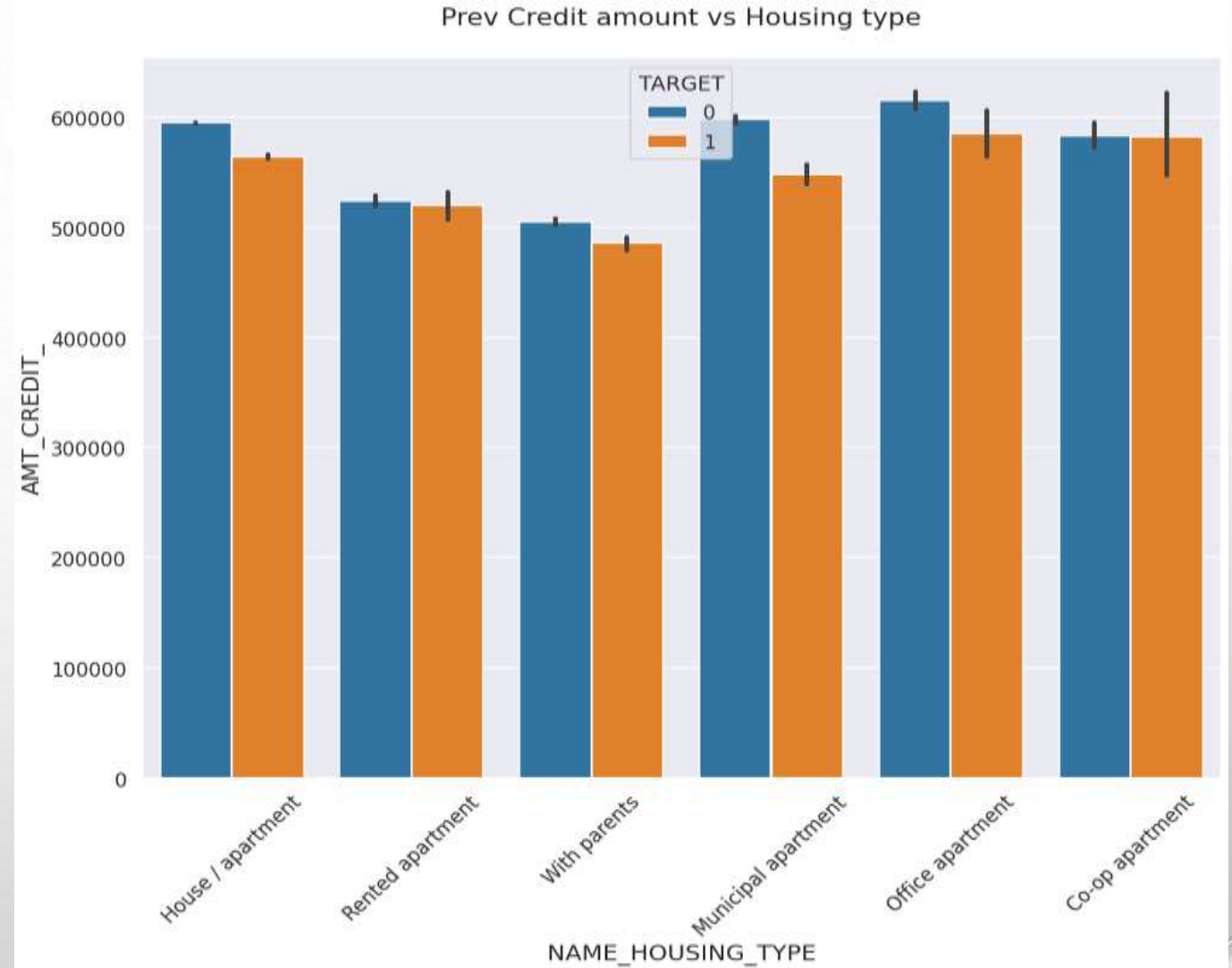
WHAT CAN BE CONCLUDED FROM THE GRAPH?

- ITS IS OBSERVED THAT THE LOAN PURPOSE OF 'BUYING A HOUSE', 'BUYING A LAND', 'BUYING A NEW CAR' AND 'BUILDING A HOUSE' IS HIGHER COMPARED TO OTHERS
- INCOME TYPE OF STATE SERVANTS HAVE A SIGNIFICANT AMOUNT OF CREDIT APPLIED.
- 'MONEY FOR THIRD PERSON' AND 'HOBBY' ARE HAVING LESS CREDITS APPLIED.

# PREVIOUS CREDIT AMOUNT VS HOUSING TYPE

WHAT CAN BE CONCLUDED FROM THE GRAPH?

- ITS IS OBSERVED THAT HOUSING TYPE , OFFICE APARTMENT HAS HIGHER CREDIT OF TARGET0
- CO-OP APARTMENT IS HAVING HIGHER CREDIT TARGET1
- HERE WE CAN CONCLUDE THAT BANK SHOULD AVOID GIVING LOANS TO HOUSING TYPE CO-OP APARTMENT AS THEY ARE HAVING DIFFICULTIES PAYING LOAN.
- BANK CAN FOCUS MAINLY ON HOUSING TYPE WITH PARENTS OR HOUSE/APARTMENT OR MUNICIPAL APARTMENT AS THEY ARE SUCCESSFUL PAYEES OF LOAN



# CONCLUSION

- BANK SHOULD AVOID CLIENTS WITH OFFICE APARTMENT AND CO-OP APARTMENT HOUSING TYPE HAVING MORE NUMBER OF PAYMENT DIFFICULTIES
- BANK SHOULD FOCUS ON BUSINESS ENTITY TYPE 1 (ORGANIZATION TYPE) AS THEY HAVE LESS NO. OF PAYMENT DIFFICULTIES
- BANKS SHOULD FOCUS LESS ON INCOME TYPE 'WORKING' AS THEY ARE HAVING MOST NUMBER OF UNSUCCESSFUL PAYMENTS.
- AND ALSO WITH LOAN PURPOSE 'REPAIRS' IS HAVING HIGHER NUMBER OF UNSUCCESSFUL PAYMENTS ON TIME.
- HOUSING TYPE 'WITH PARENTS', 'HOUSE/APARTMENT', 'MUNICIPAL APARTMENT' ARE HAVING LEAST NUMBER OF UNSUCCESSFUL PAYMENTS.

The background is a light gray gradient. In the top-left and bottom-right corners, there are several realistic water droplets of various sizes. A faint, circular logo is centered in the upper half of the image. The logo features a central emblem surrounded by text in a circular arrangement, which appears to be 'WORLD HEALTH ORGANIZATION' at the top and 'OFFICE OF THE DIRECTOR' at the bottom.

**THANK YOU**