CREDIT EDA CASE STUDY

BY

CHENAMMA PRIYA RANGARAJ

INTRODUCTION

- This case study aims to give you an idea of applying EDA in a real business scenario. In this case study, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

BUSINESS UNDERSTANDING - 1

- 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 as their advantage by becoming a defaulters. Suppose you work for a consumer finance company which specialises 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.

BUSINESS UNDERSTANDING - 2

- 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):

1.Approved: The Company has approved loan Application

2.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.

3.Refused: The company had rejected the loan (because the client does not meet their requirements etc.).

4.Unused offer: Loan has been cancelled by the client but on different stages of the process.

- In this case study, we will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

BUSINESS OBJECTIVES

- This case study aims to identify patterns which indicate if a client has difficulty paying their installments 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.
- To develop your understanding of the domain, you are advised to independently research a little about risk analytics -understanding the types of variables and their significance should be enough).

DATA UNDERSTANDING

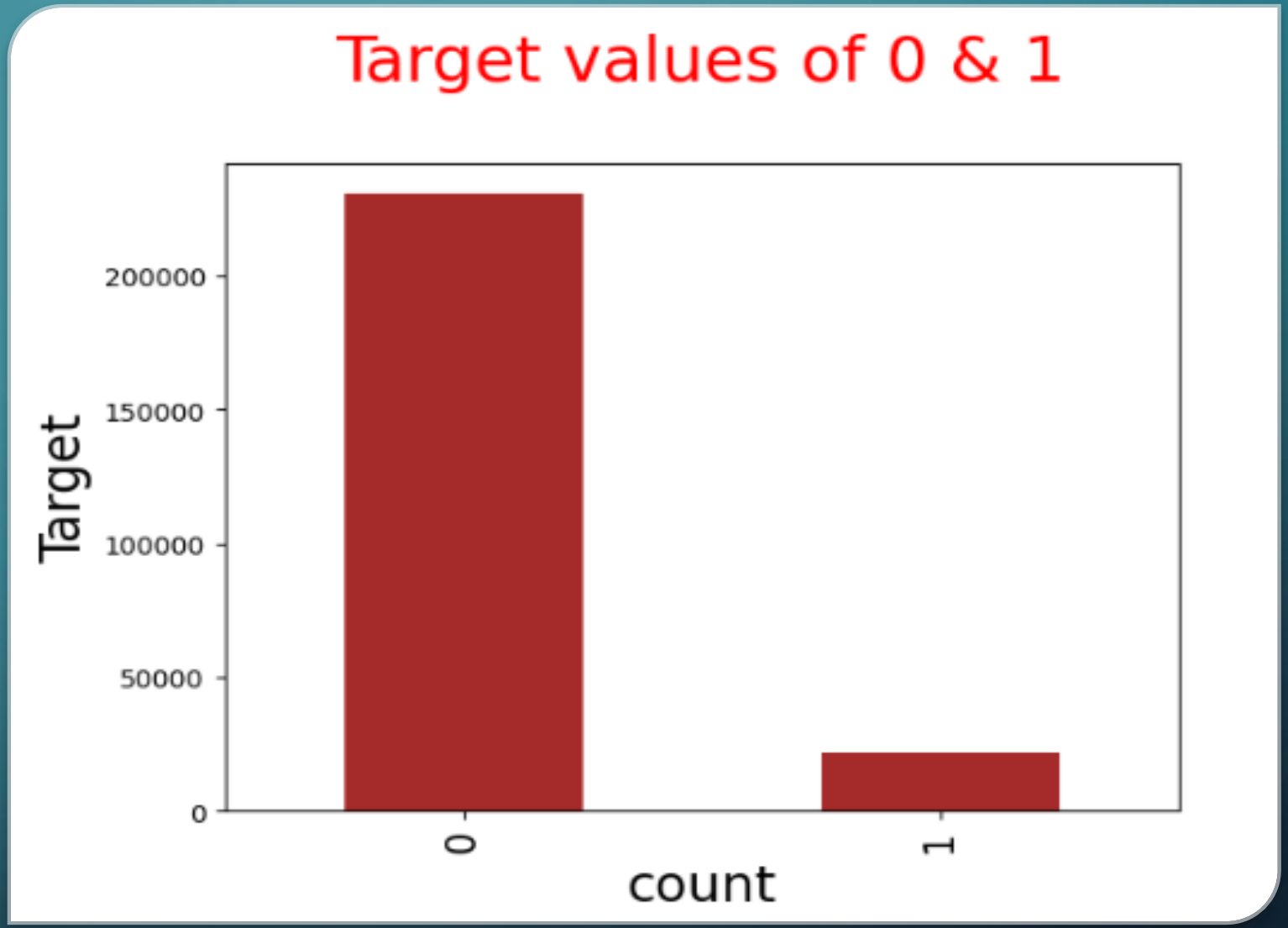
- This dataset has 3 files as explained below:
- *1. 'application_data.csv'* contains all the information of the client at the time of application. The data is about whether a **client has payment difficulties.**
- *2. 'previous_application.csv'* contains information about the client's previous loan data. It contains the data 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.

ANALYSIS OF INFORMATION OF THE CLIENT AT THE TIME OF APPLICATION

CHECKING IMBALANCE FOR TARGET

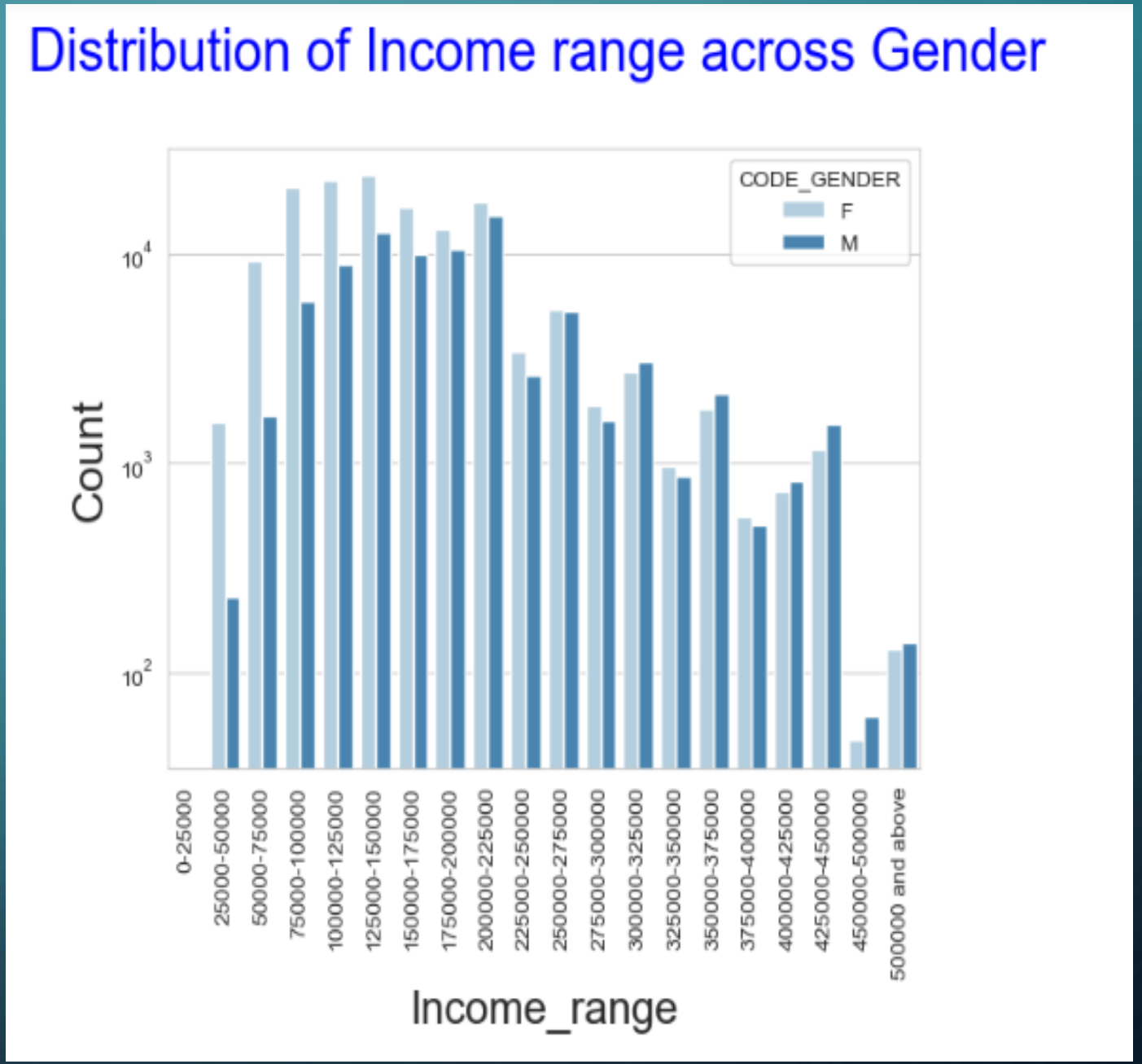
ANALYSIS OF IMBALANCE FOR `TARGET`

- We have imbalance in `TARGET` variable based on the % of observations
- TARGET` value 1 represents client with payment difficulties (he/she had late payment more than X days on at least one of the first Y installments of the loan). This is only 8.07% of the data
- TARGET` value 0 represents all other cases than 1. This is 91.93% of the data

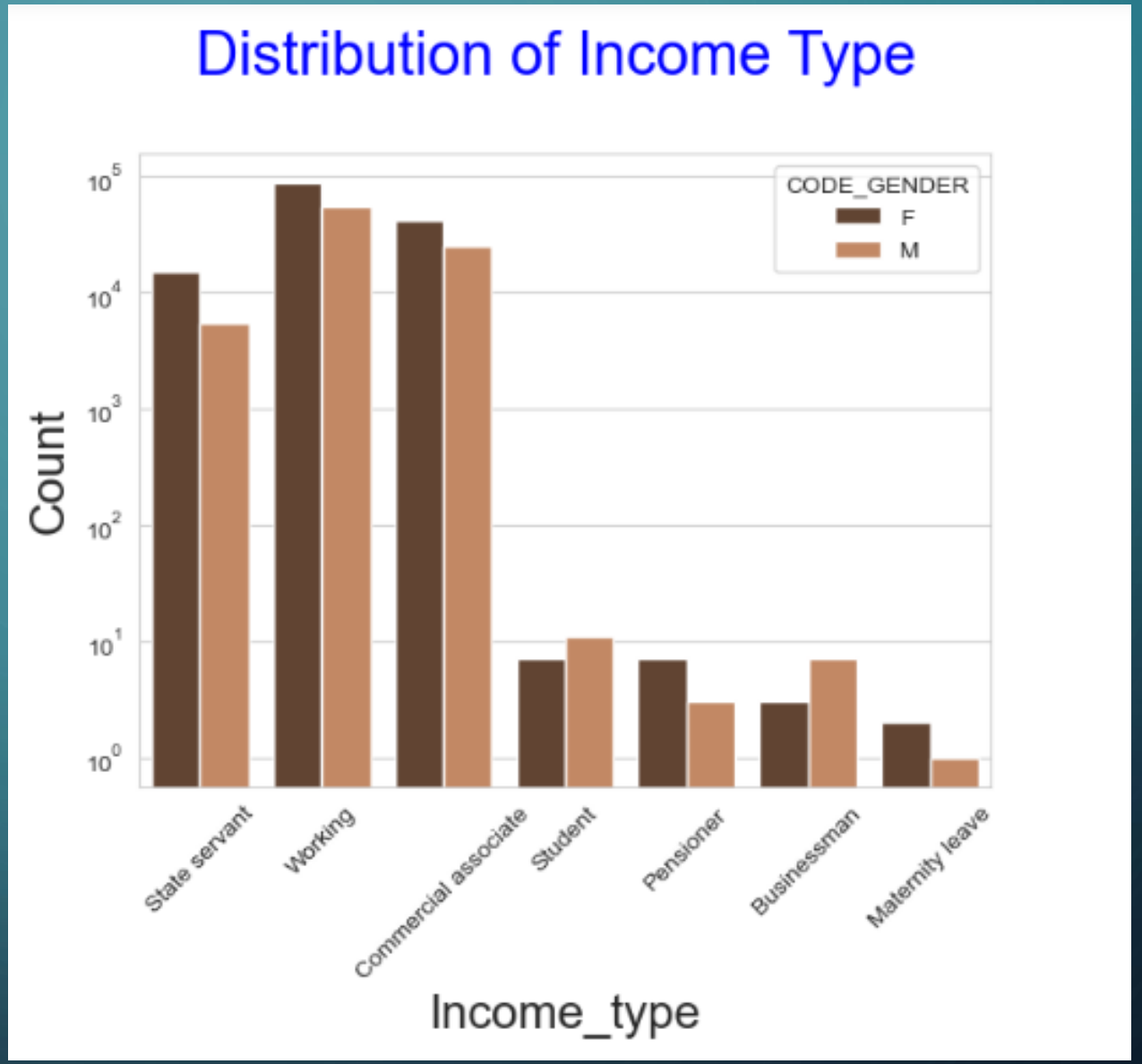


UNIVARIATE ANALYSIS FOR TARGET 0

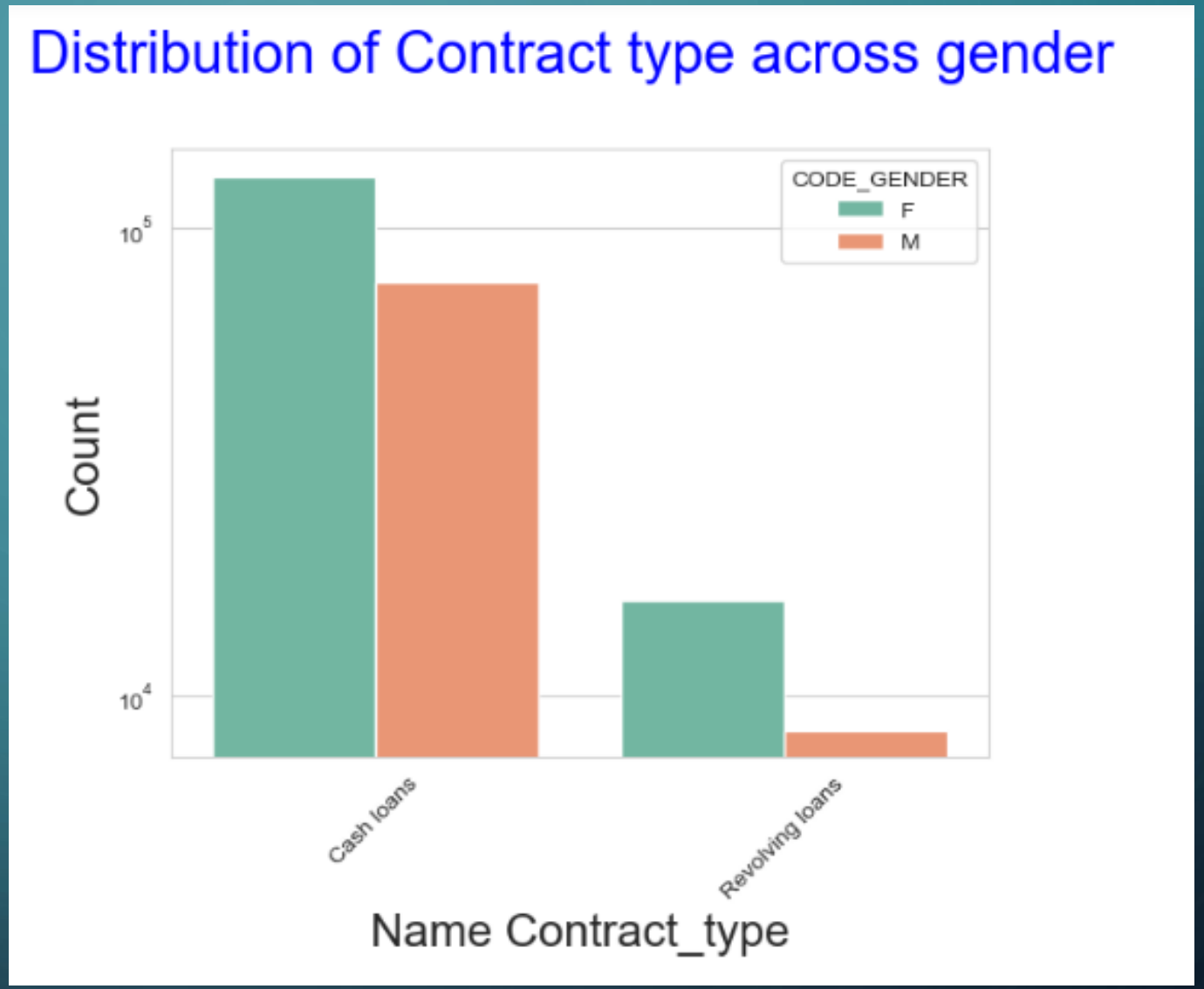
ANALYSIS OF INCOME RANGE OVER GENDER



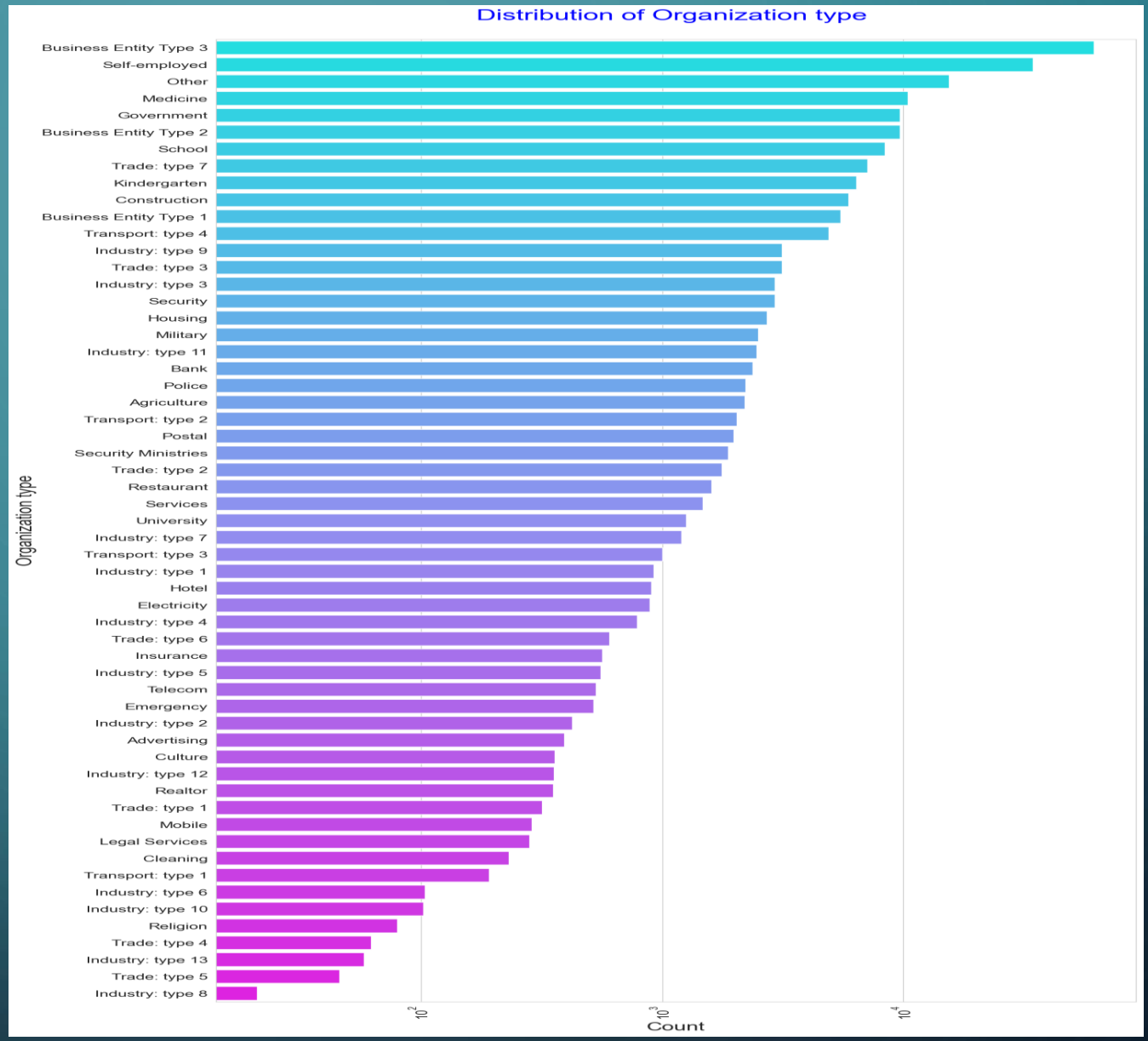
ANALYSIS OF INCOME TYPE ACROSS GENDER



ANALYSIS OF CONTRACT TYPE ACROSS GENDER

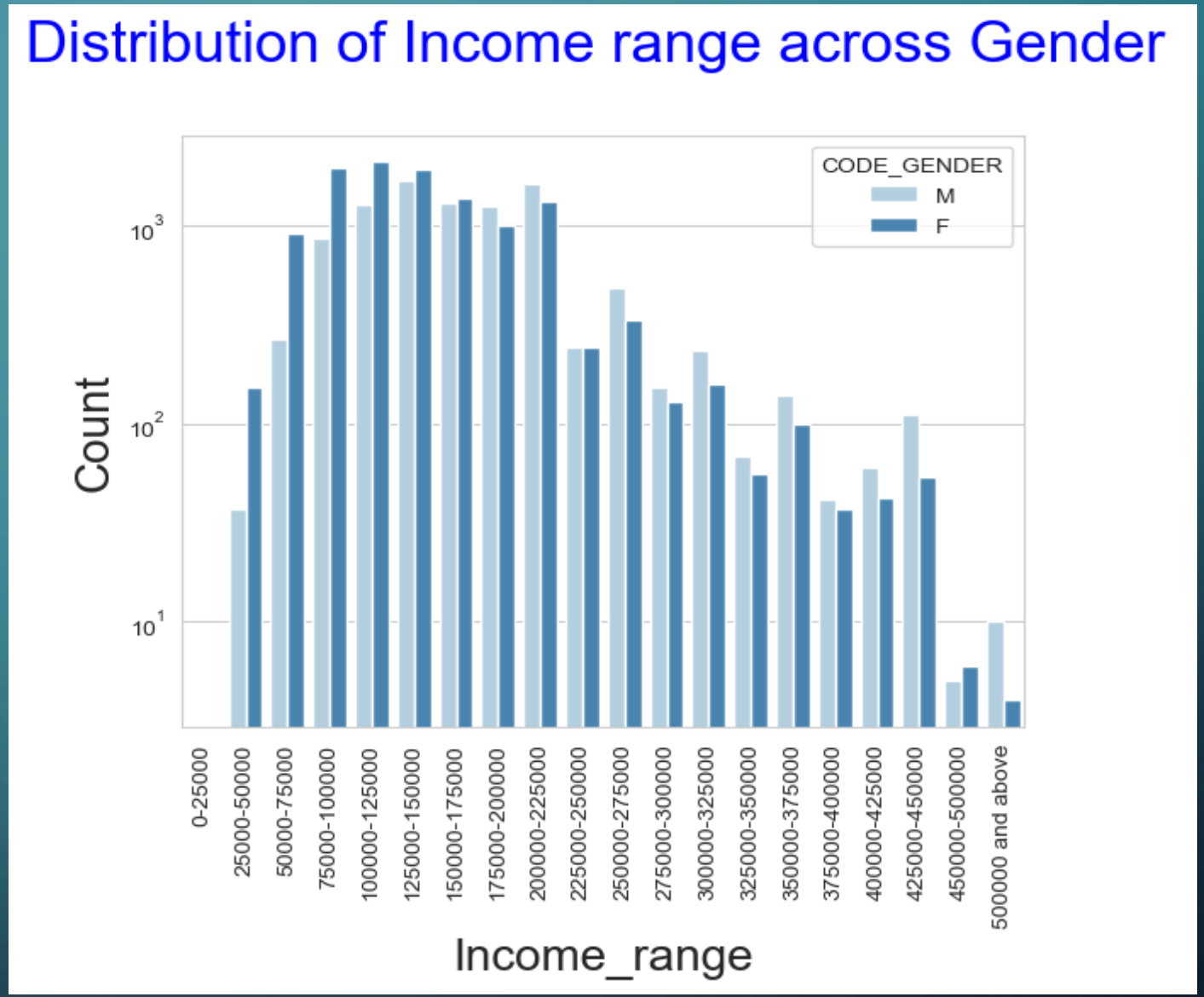


ANALYSIS OF DIFFERENT ORGANIZATION TYPE

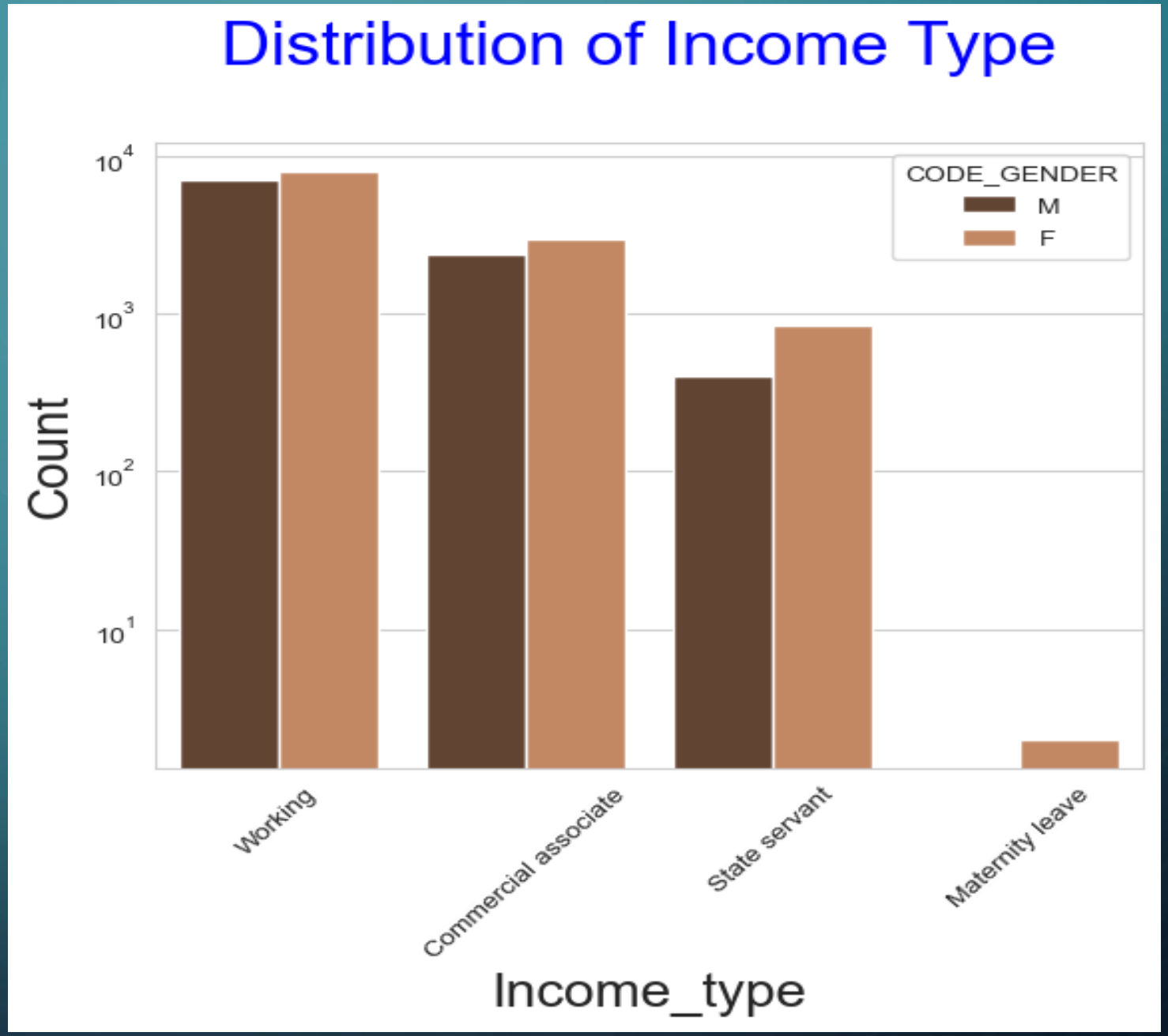


UNIVARIATE ANALYSIS FOR TARGET 1

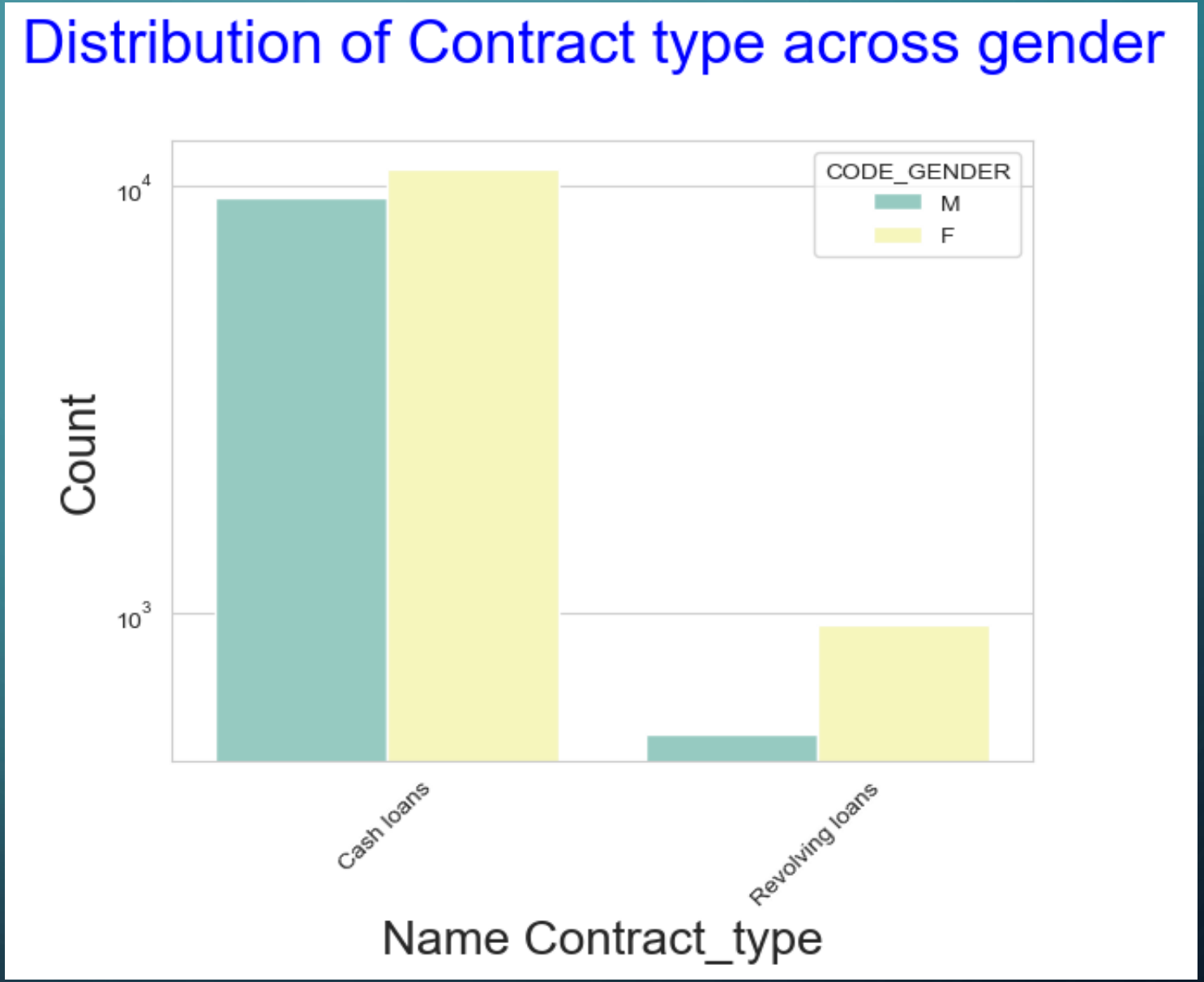
ANALYSIS OF INCOME RANGE ACROSS GENDER



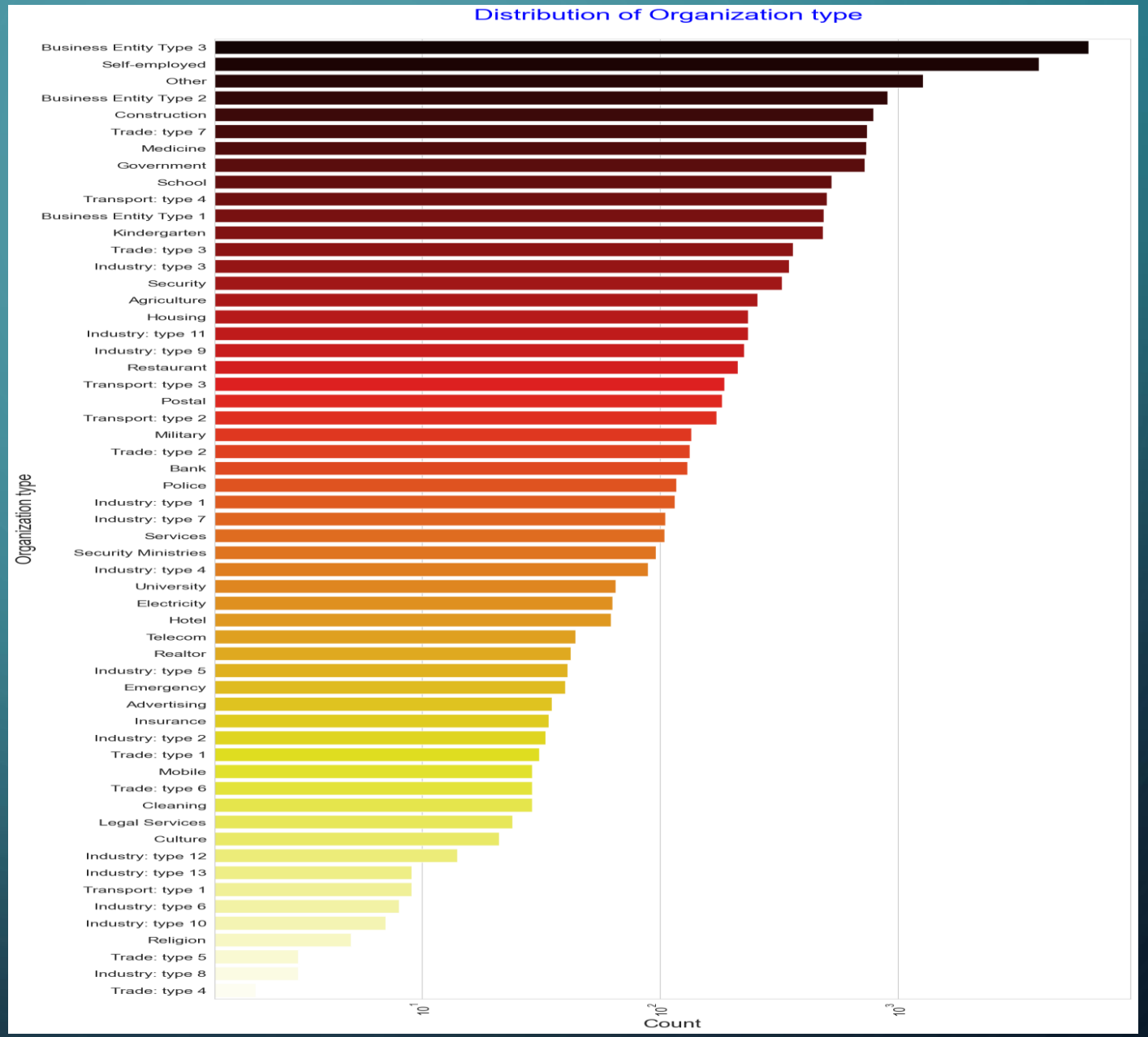
ANALYSIS OF INCOME TYPE ACROSS GENDER



ANALYSIS OF CONTRACT TYPE ACROSS GENDER

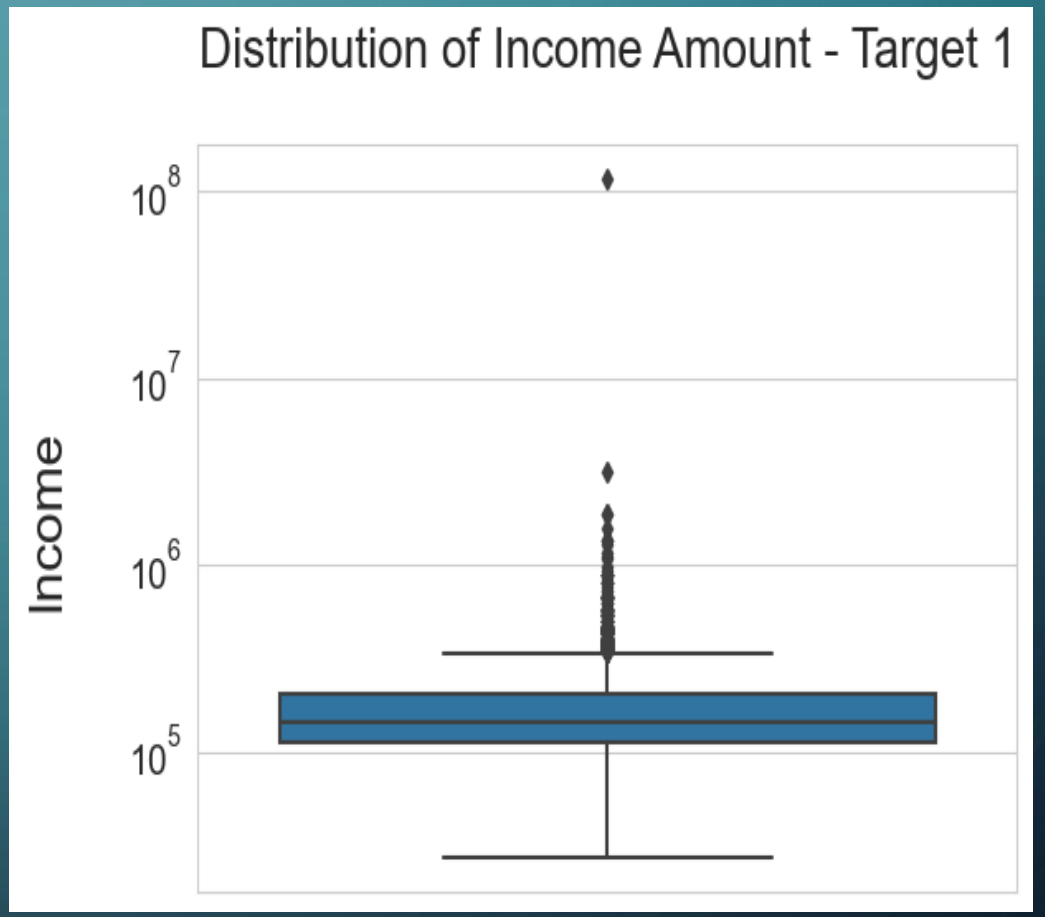


ANALYSIS OF ORGANIZATION TYPE ACROSS GENDER

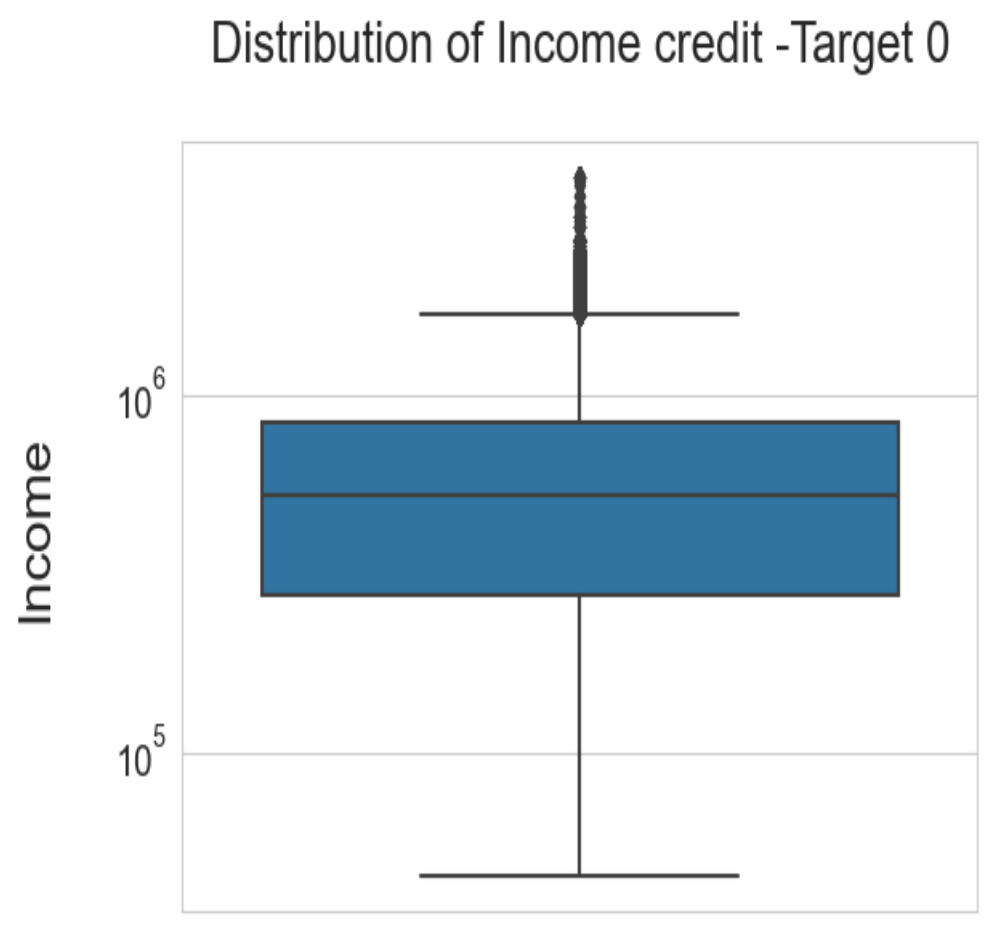


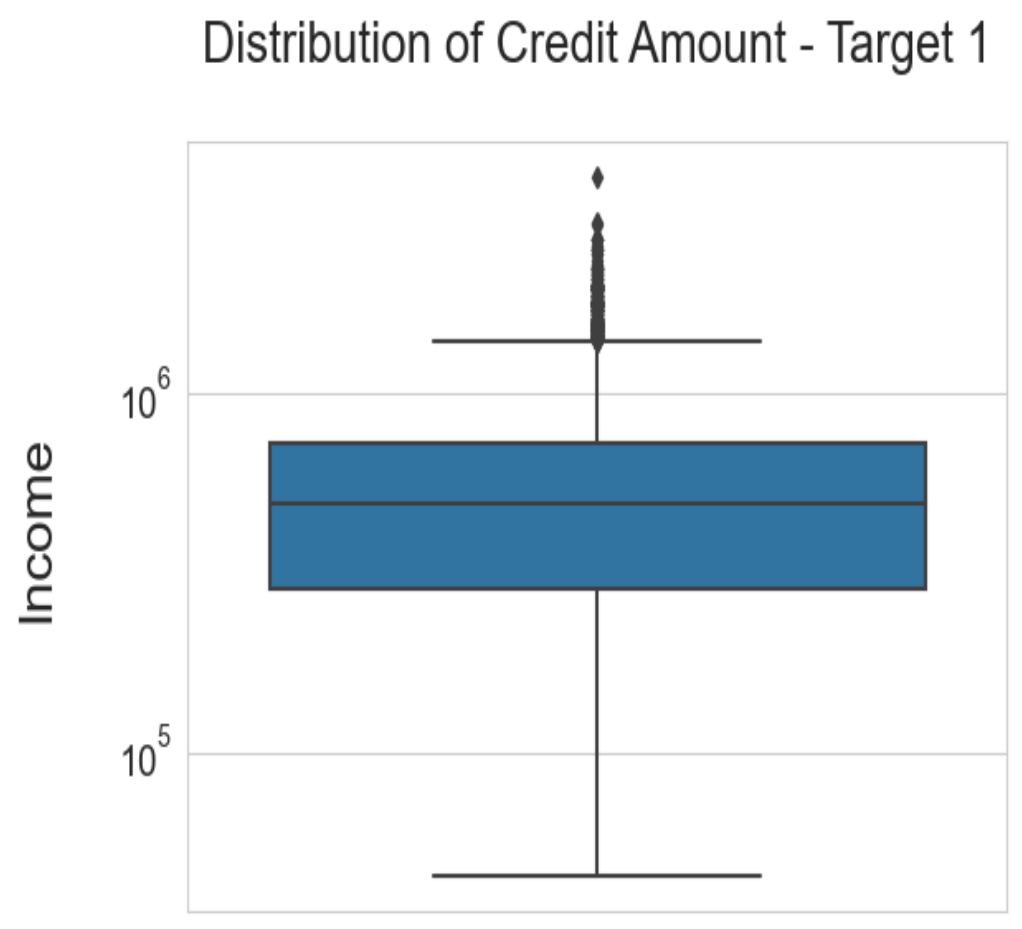
OUTLIER ANALYSIS

ANALYSIS OF INCOME AMOUNT

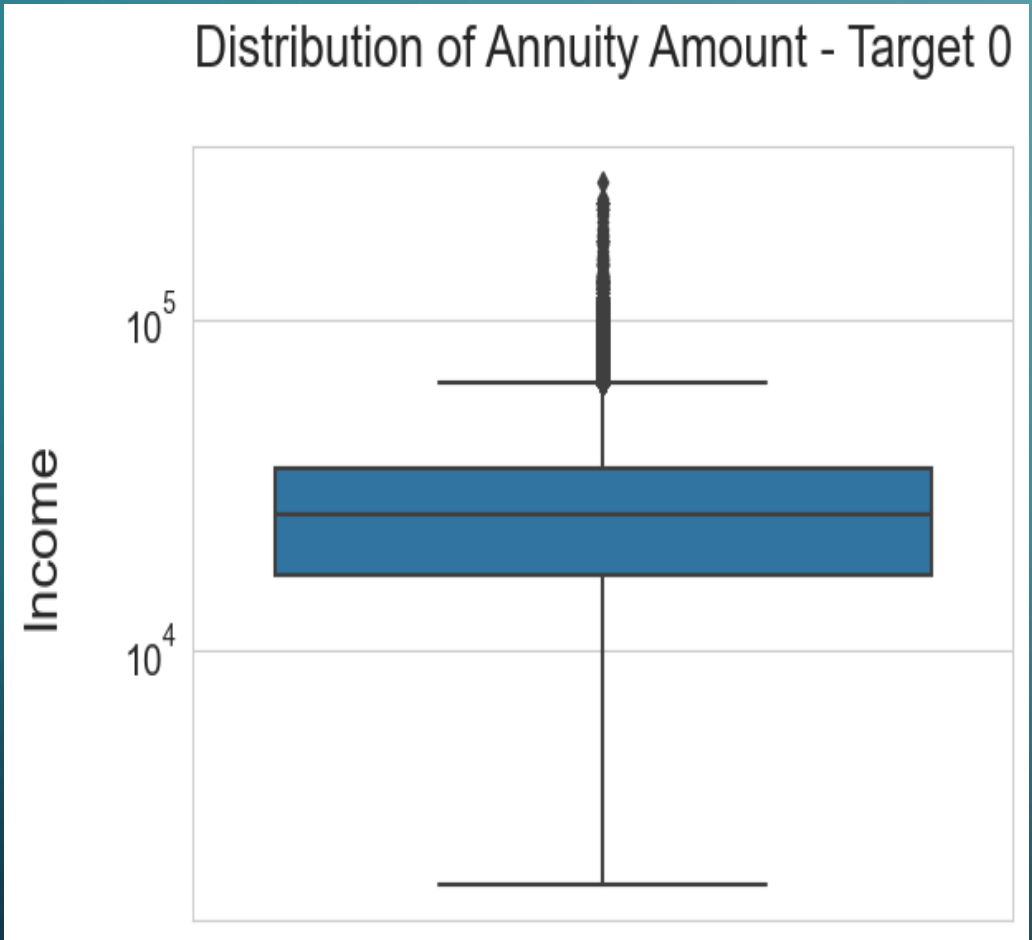


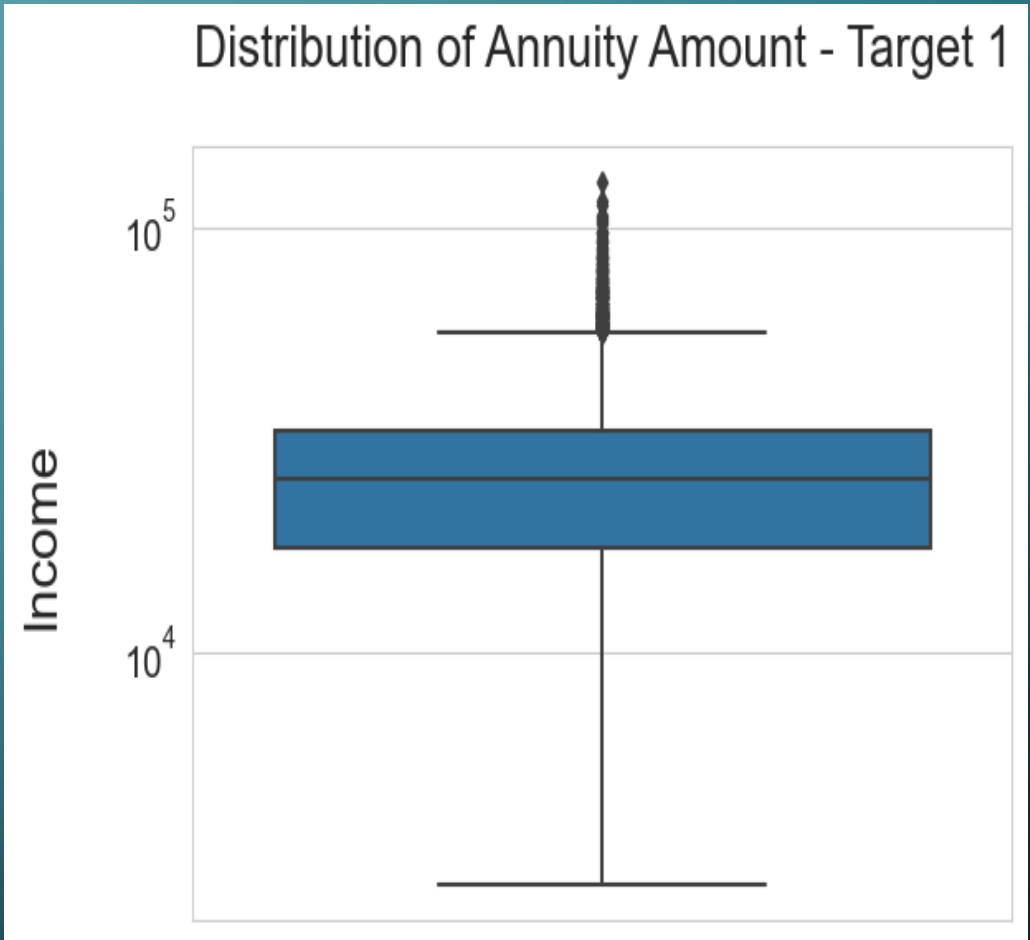
ANALYSIS OF CREDIT INCOME





ANALYSIS OF ANNUITY AMOUNT



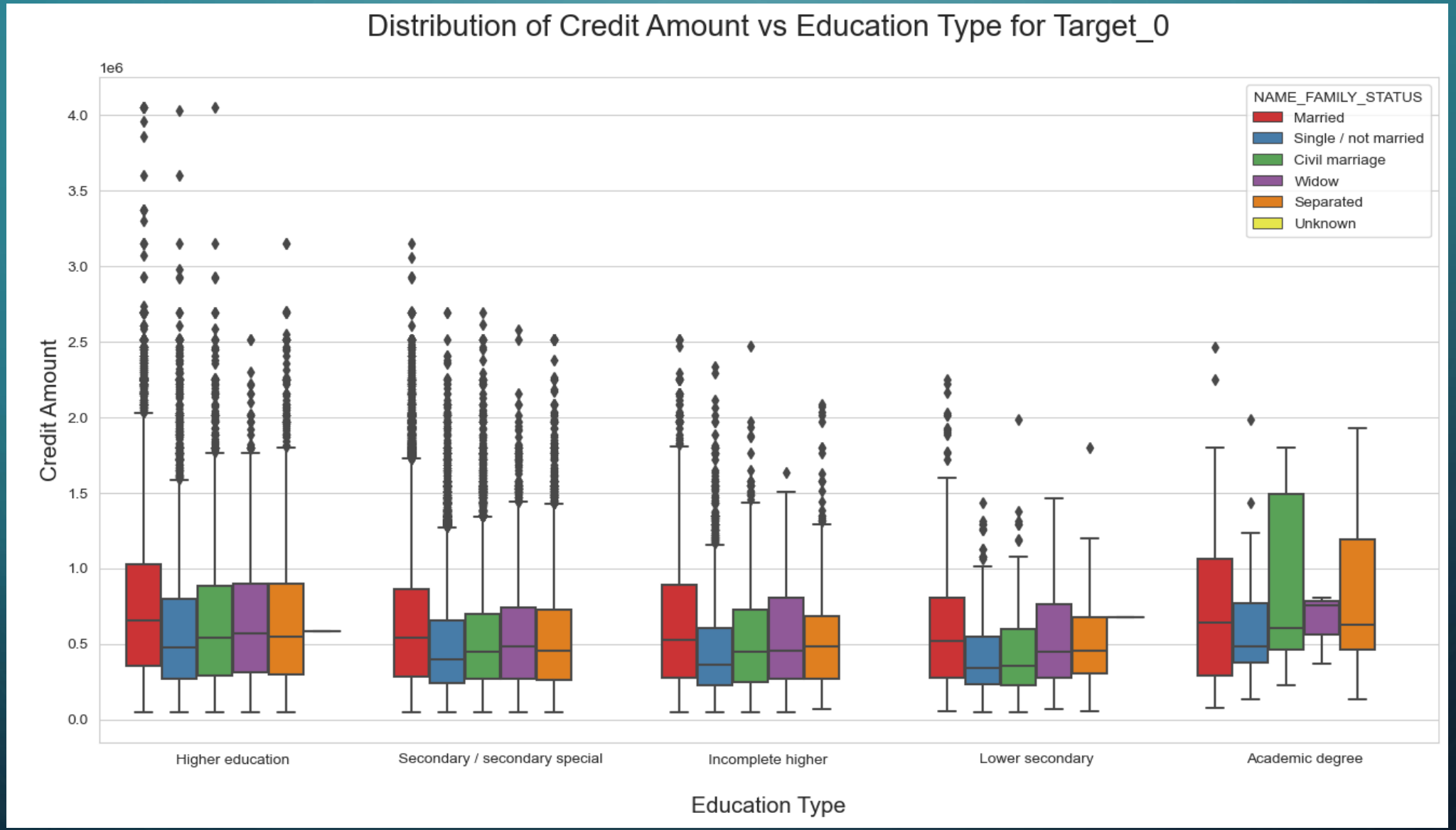


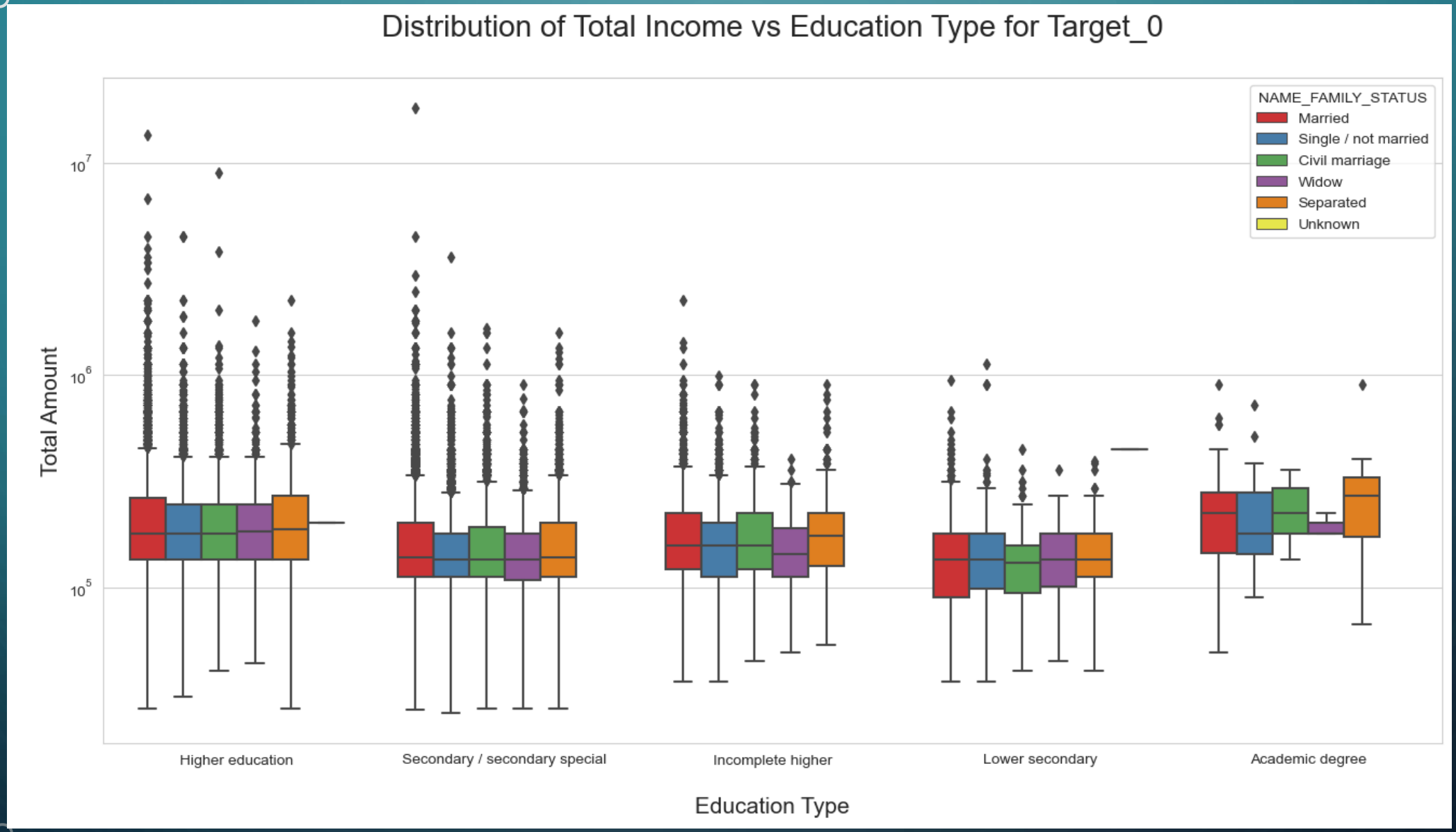
BIVARIATE ANALYSIS

ANALYSIS OF INCOME VS CREDIT

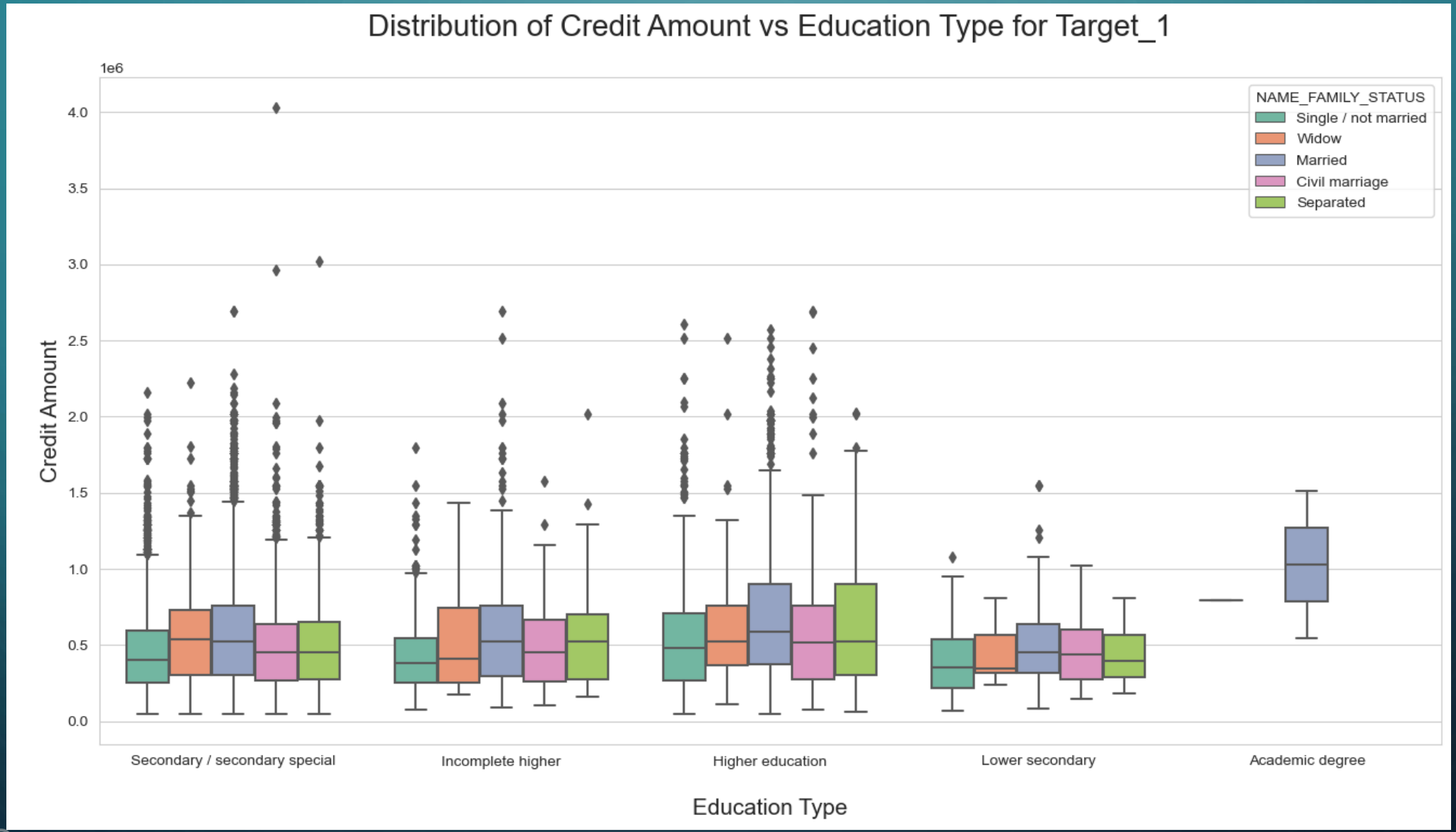
ANALYSIS OF INCOME VS GOODS PRICE

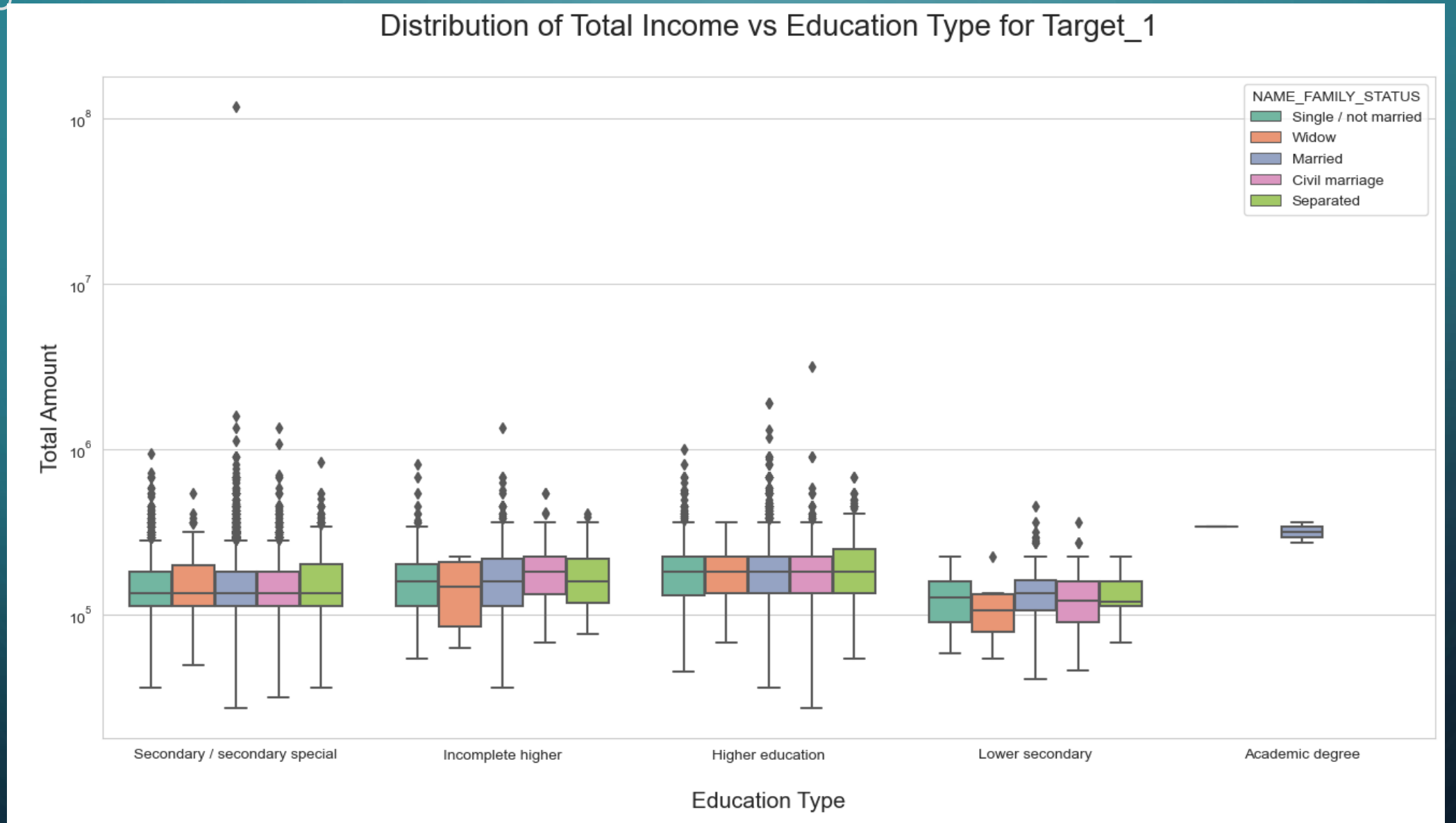
MULTIVARIATE ANALYSIS ON TARGET 0


Distribution of Credit Amount vs Education Type for Target_0
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Credit Amount
NAME_FAMILY_STATUS
Married
Single / not married
Civil marriage
Widow
Separated
Unknown
Higher education
Secondary / secondary special
Incomplete higher
Lower secondary
Academic degree
Education Type

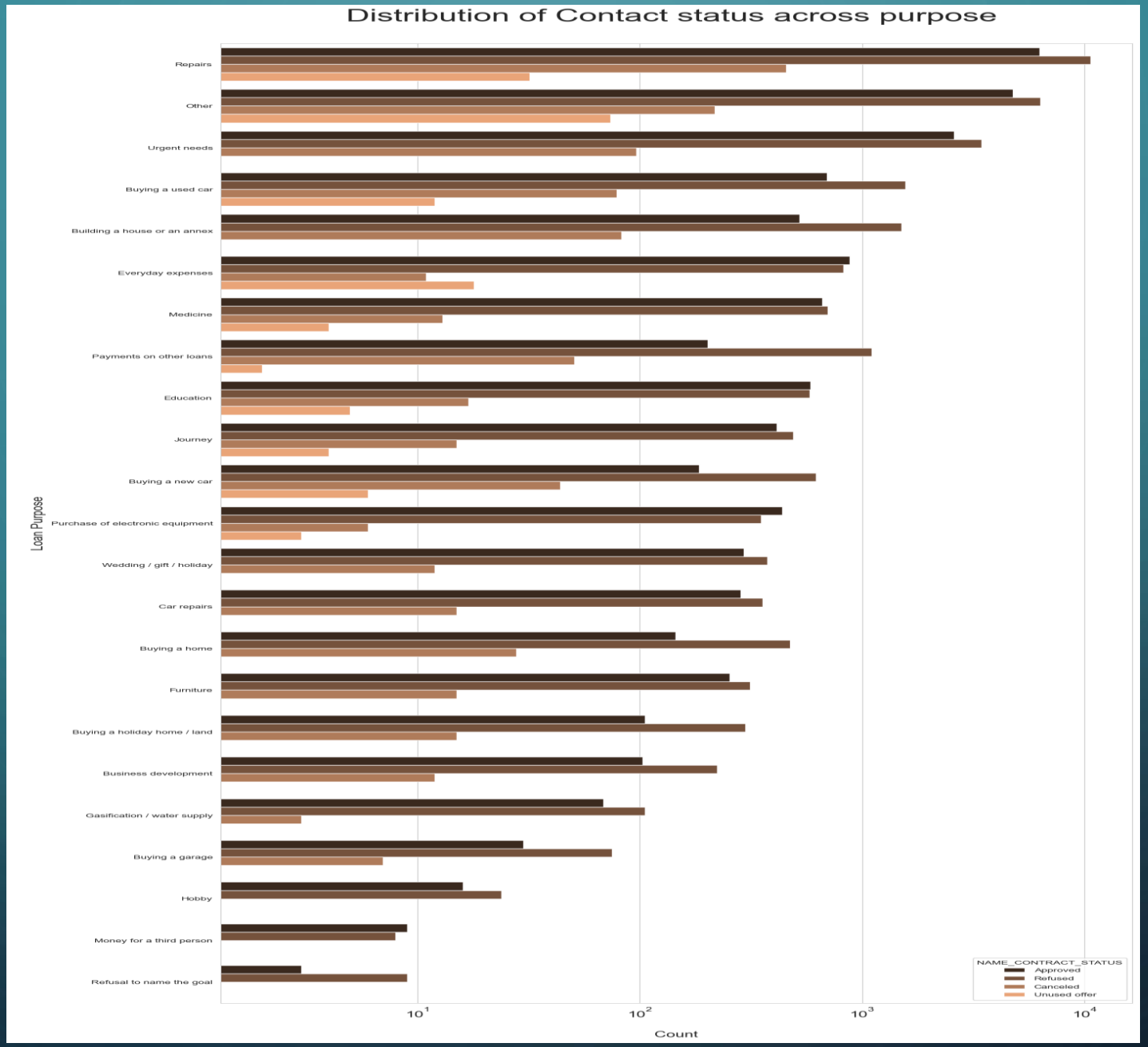

Distribution of Total Income vs Education Type for Target_0
NAME_FAMILY_STATUS
Married
Single / not married
Civil marriage
Widow
Separated
Unknown
Total Amount
10^7
10^6
10^5
Higher education
Secondary / secondary special
Incomplete higher
Lower secondary
Academic degree
Education Type

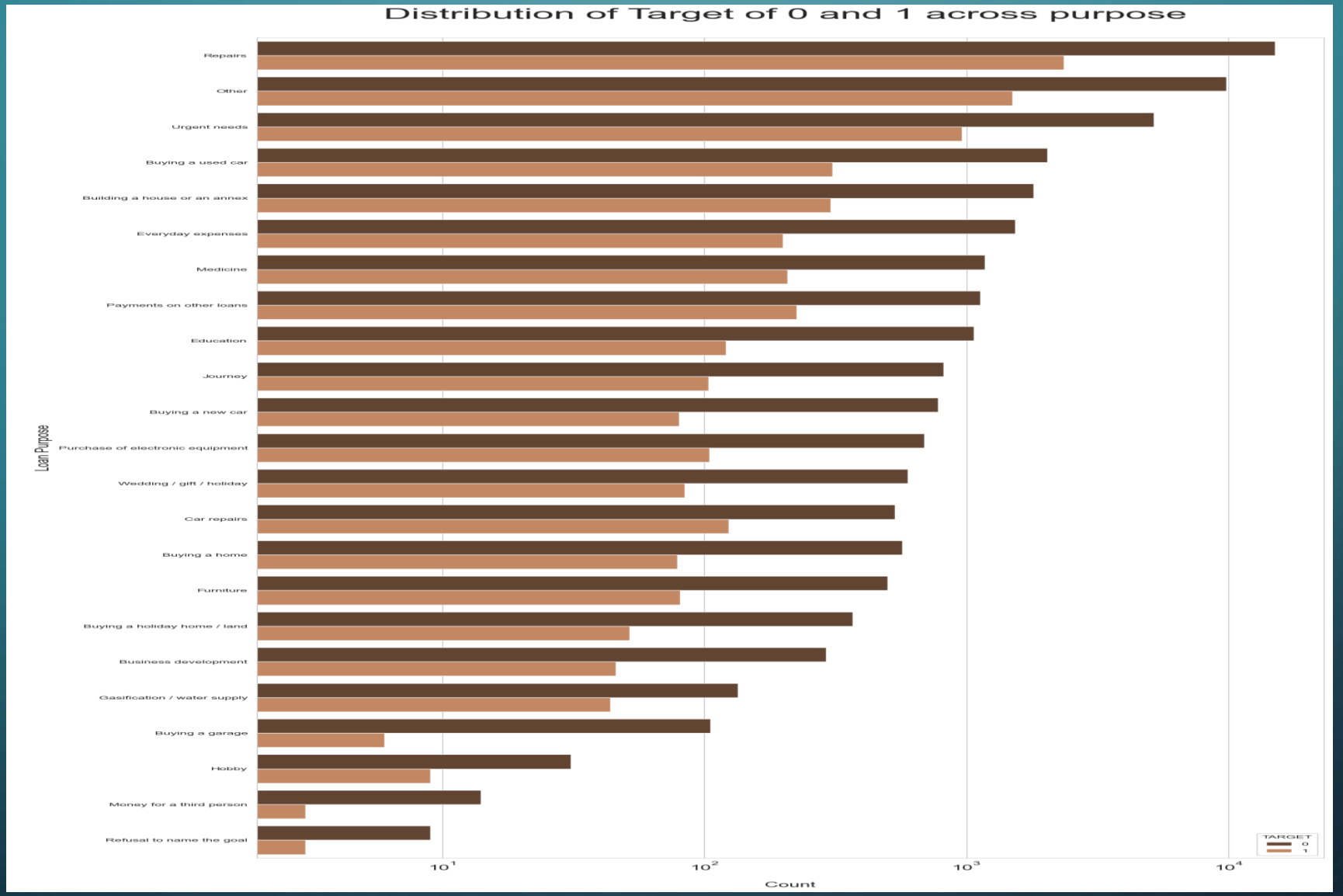
MULTIVARIATE ANALYSIS ON TARGET 1

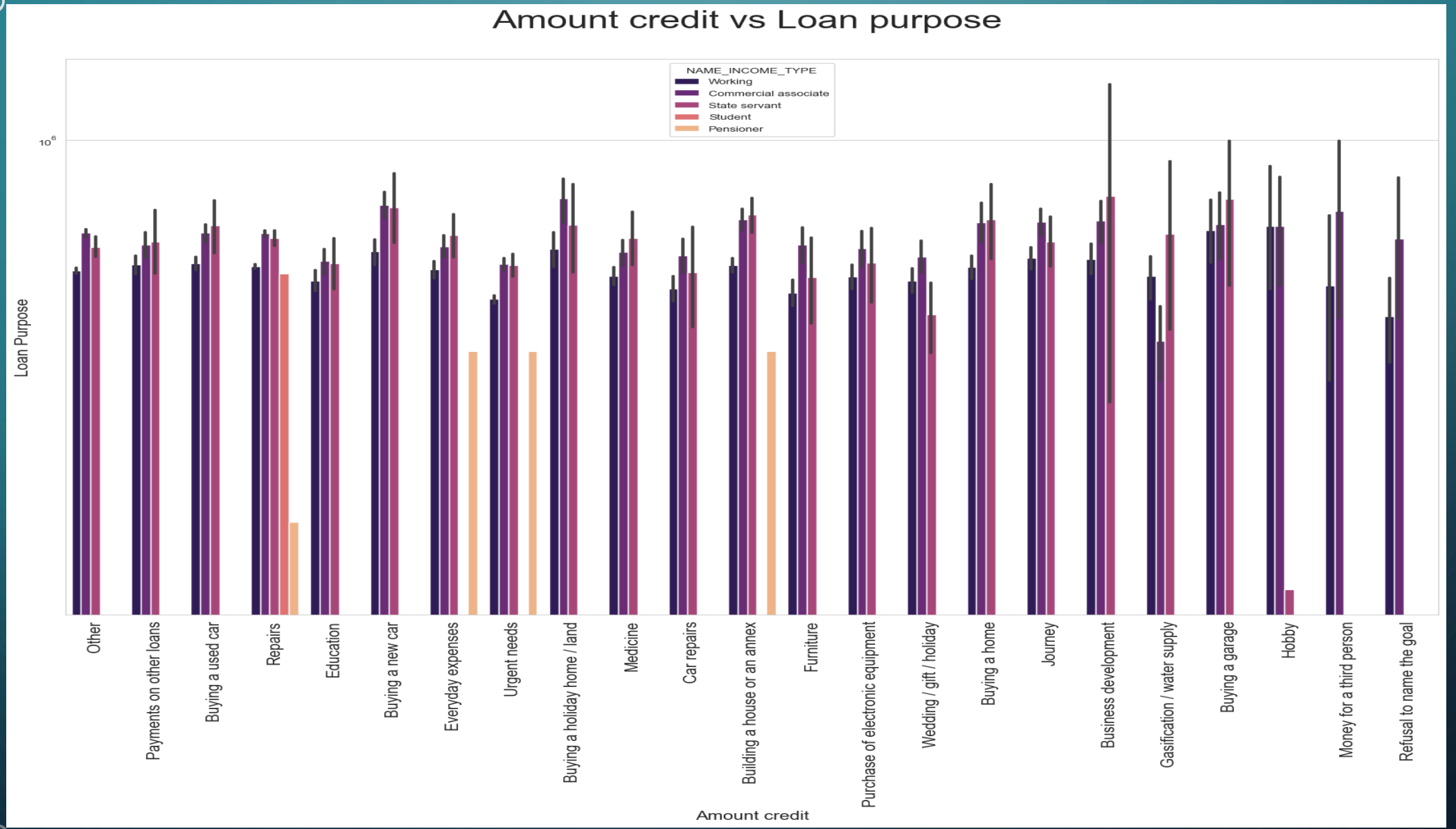

Distribution of Credit Amount vs Education Type for Target_1
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Credit Amount
NAME_FAMILY_STATUS
Single / not married
Widow
Married
Civil marriage
Separated
Secondary / secondary special
Incomplete higher
Higher education
Lower secondary
Academic degree
Education Type

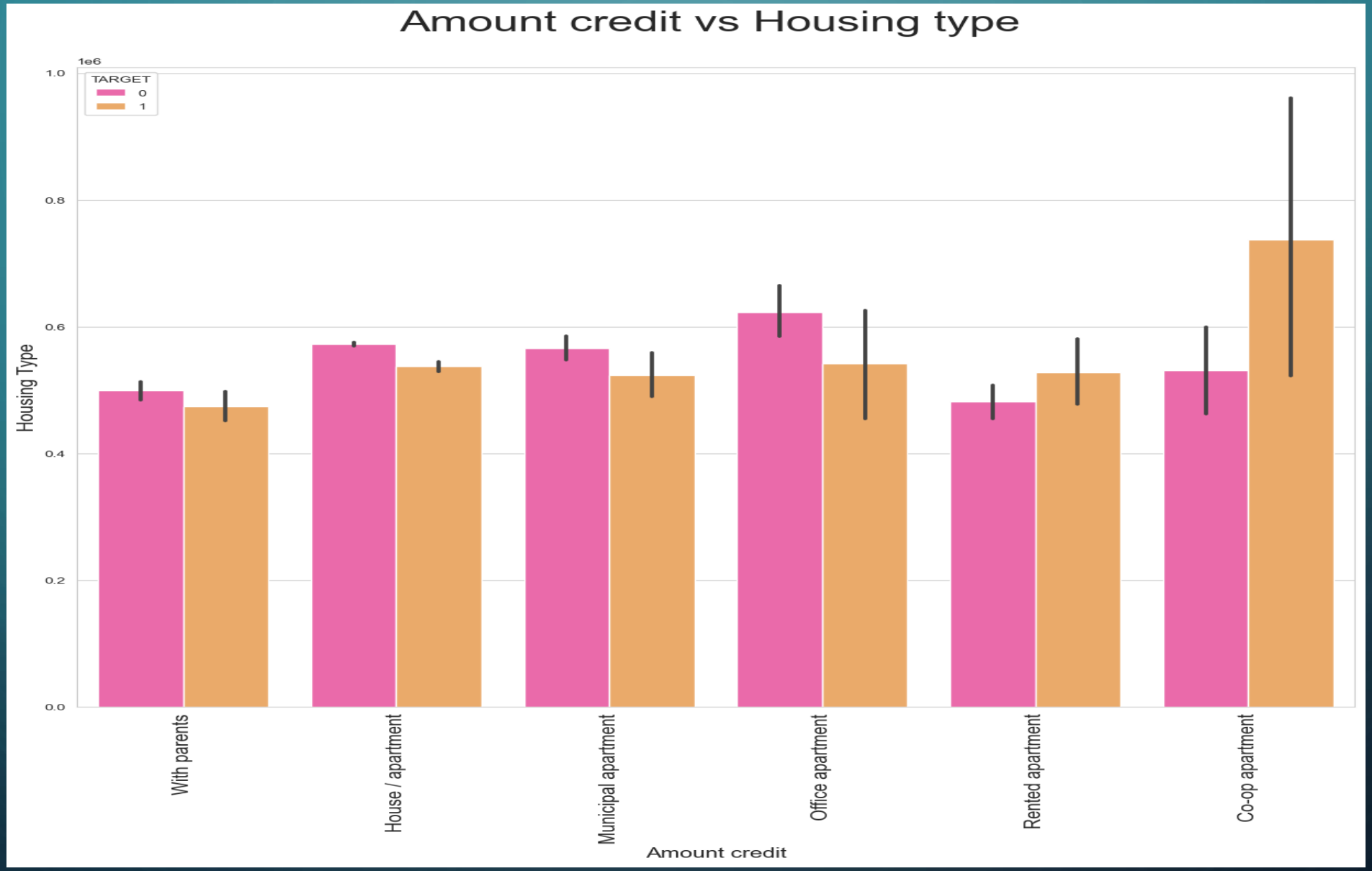

Distribution of Total Income vs Education Type for Target_1
NAME_FAMILY_STATUS
Single / not married
Widow
Married
Civil marriage
Separated
Total Amount
10^8
10^7
10^6
10^5
Secondary / secondary special
Incomplete higher
Higher education
Lower secondary
Academic degree
Education Type

ANALYSIS OF MERGED INFORMATION ABOUT THE CLIENT'S PREVIOUS LOAN DATA AND CURRENT LOAN APPLICATION


Distribution of Contact status across purpose
Repairs
Other
Urgent needs
Buying a used car
Building a house or an annex
Everyday expenses
Medicine
Payments on other loans
Education
Journey
Buying a new car
Purchase of electronic equipment
Wedding / gift / holiday
Car repairs
Buying a home
Furniture
Buying a holiday home / land
Business development
Gasification / water supply
Buying a garage
Hobby
Money for a third person
Refusal to name the goal
Loan Purpose
10^1
10^2
10^3
10^4
Count
NAME_CONTRACT_STATUS
Approved
Refused
Canceled
Unused offer


Distribution of Target of 0 and 1 across purpose
Repairs
Other
Urgent needs
Buying a used car
Building a house or an annex
Everyday expenses
Medicine
Payments on other loans
Education
Journey
Buying a new car
Purchase of electronic equipment
Wedding / gift / holiday
Car repairs
Buying a home
Furniture
Buying a holiday home / land
Business development
Gasification / water supply
Buying a garage
Hobby
Money for a third person
Refusal to name the goal
Loan Purpose
10^1
10^2
10^3
10^4
Count
TARGET
0
1


Amount credit vs Loan purpose
NAME_INCOME_TYPE
Working
Commercial associate
State servant
Student
Pensioner
10^6
Loan Purpose
Other
Payments on other loans
Buying a used car
Repairs
Education
Buying a new car
Everyday expenses
Urgent needs
Buying a holiday home / land
Medicine
Car repairs
Building a house or an annex
Furniture
Purchase of electronic equipment
Wedding / gift / holiday
Buying a home
Journey
Business development
Gasification / water supply
Buying a garage
Hobby
Money for a third person
Refusal to name the goal
Amount credit


Amount credit vs Housing type
1e6
1.0
0.8
0.6
0.4
0.2
0.0
TARGET
0
1
Housing Type
With parents
House / apartment
Municipal apartment
Office apartment
Rented apartment
Co-op apartment
Amount credit

CONCLUSION

1. Banks should approve loans for Office, apartment, and apartment housing types as there are fewer payment difficulties

2. Banks should provide loans to Repairs and Other purposes

3. Banks should provide loans to Business Entity and Self-employed persons so that they will repay the loan without any issues

4. Working people, especially female employees are the best to target for loans