EDA_Assignment_Aditi Tomar

Introduction

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

Understanding Of Business

The people's incomplete or nonexistent credit history makes it difficult for the loan providers to grant loans to them. As a result, some customers take advantage of this and start defaulting. Assume you are employed for a consumer finance company that specializes in providing urban clients with a range of loans. To analyze the patterns found in the data, EDA is required. It will guarantee that those who apply and are able to repay the loan are approved.

Upon receiving a loan application, the company must choose whether to approve the loan based on the applicant's profile. The bank's choice carries two different kinds of risks:

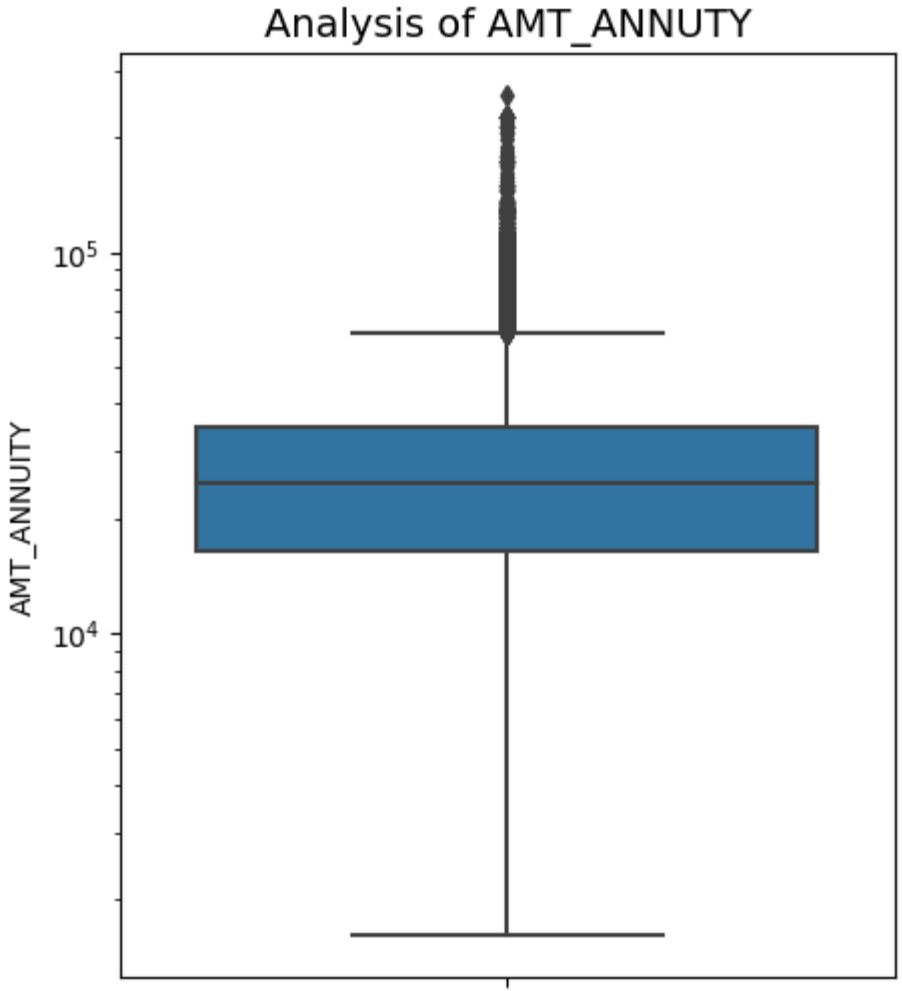
The information on the loan application at that time is included in the data below.

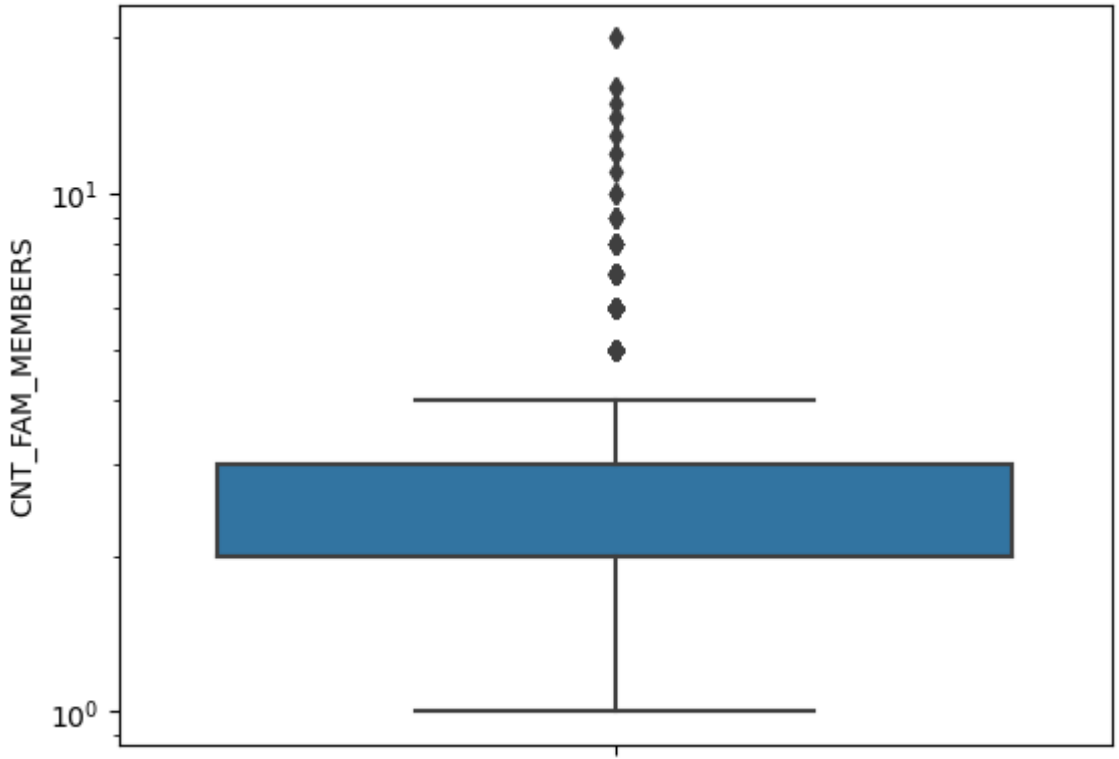
Business Objectives

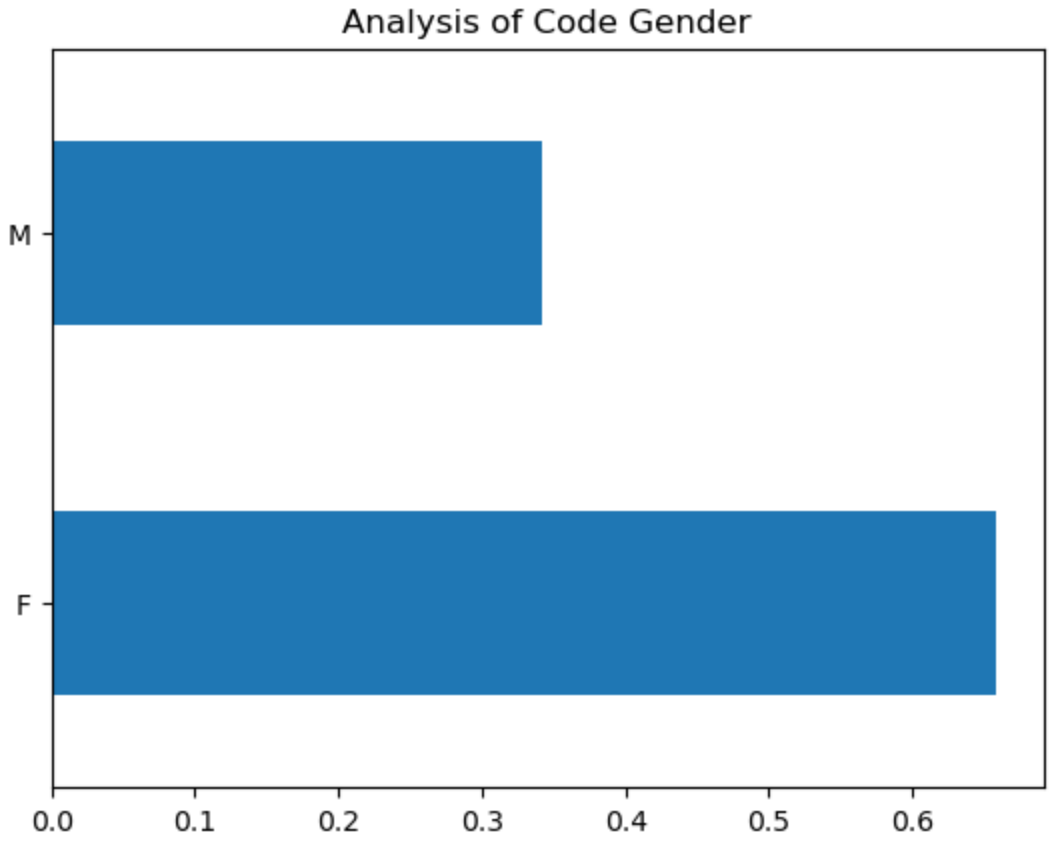
The case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

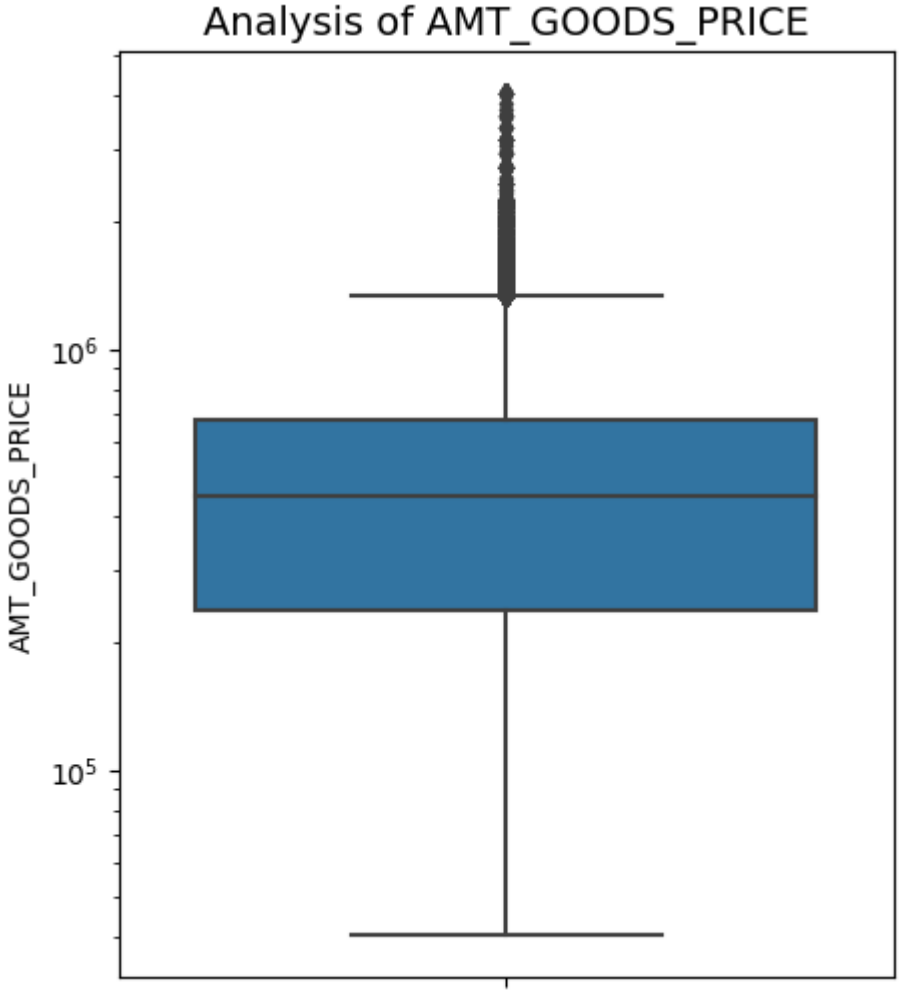
If we put it diffrently, Company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. Company can utilise this knowledge for its portfolio and risk assessment.

Steps in visualization

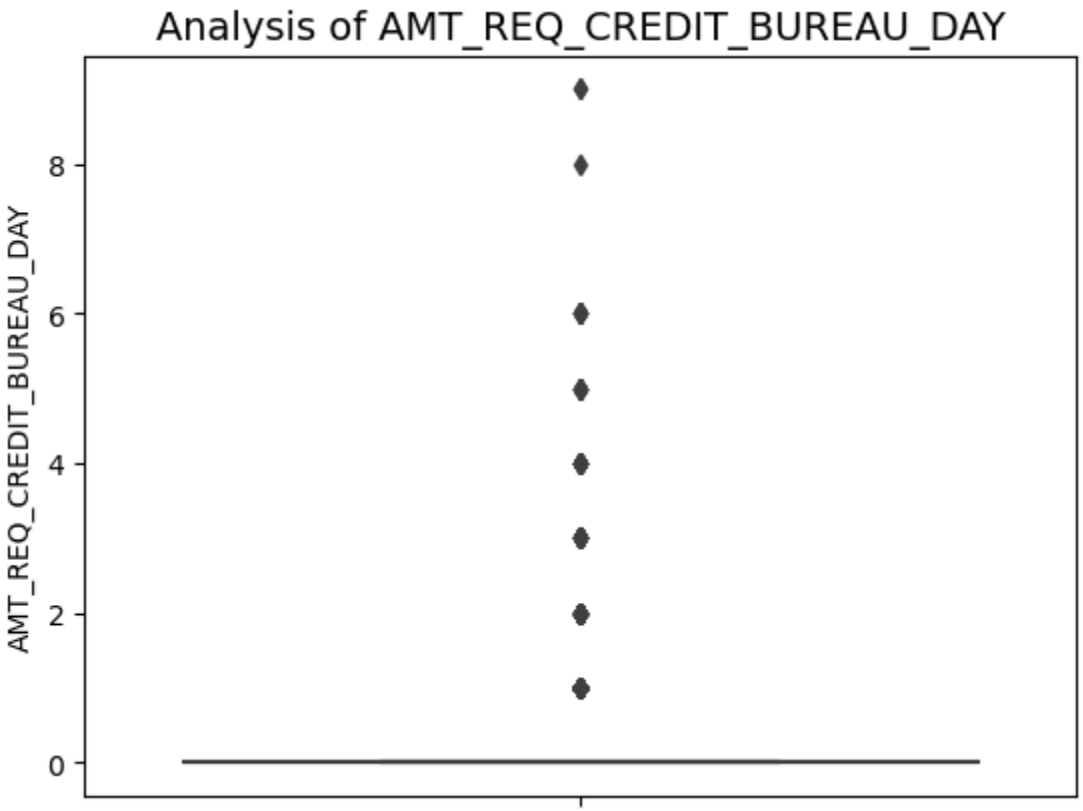

Analysis of AMT_ANNUTY
AMT_ANNUITY
10^5
10^4


CNT_FAM_MEMBERS
10^1
10^0

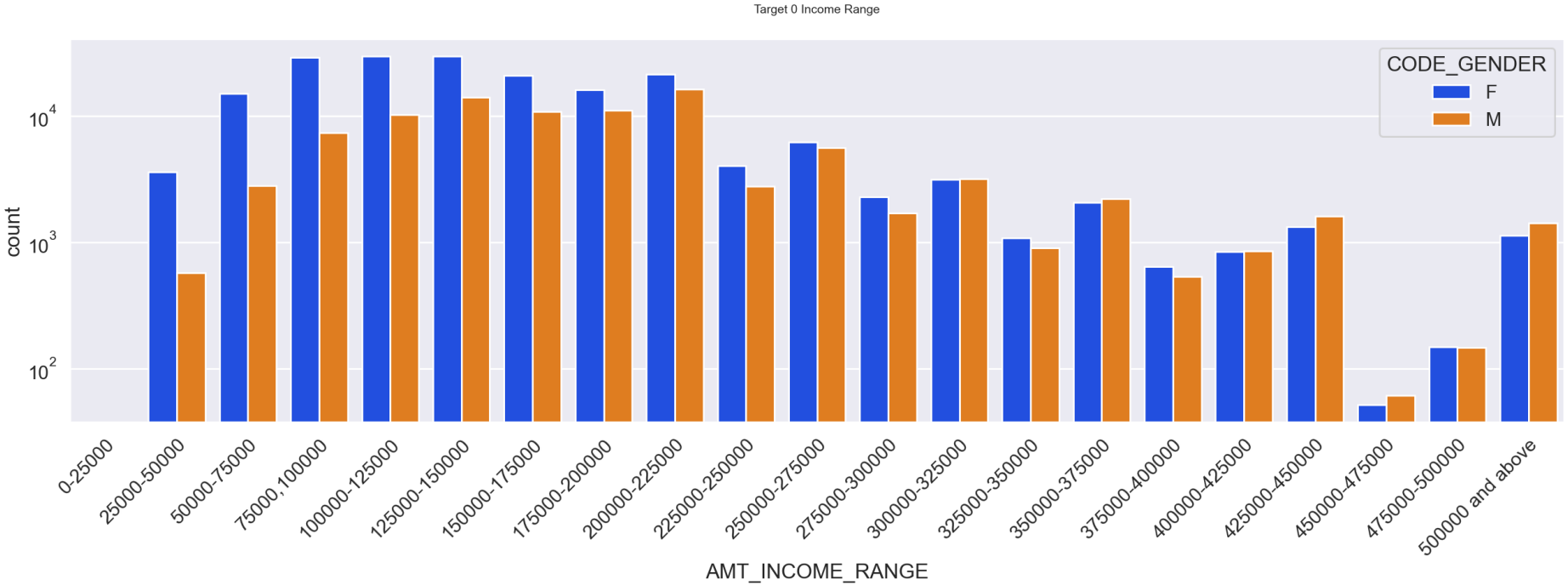

Analysis of Code Gender
M
F
0.0
0.1
0.2
0.3
0.4
0.5
0.6

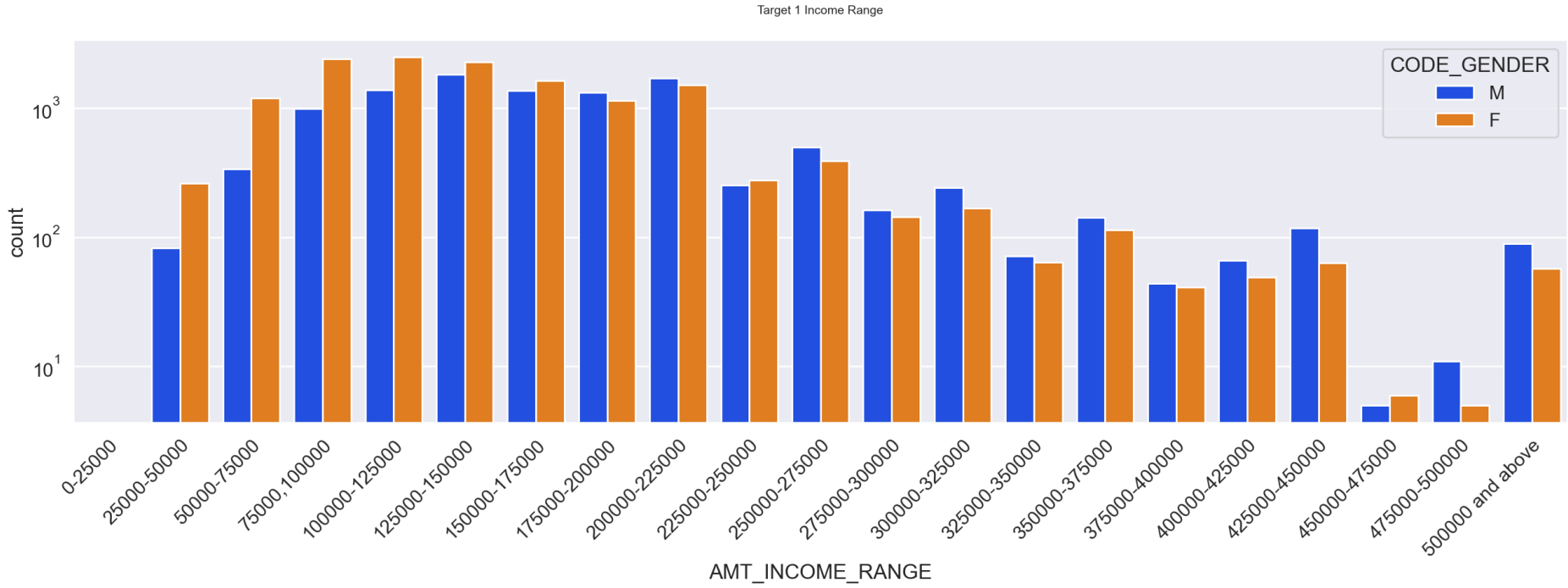

Analysis of AMT_GOODS_PRICE
AMT_GOODS_PRICE
10^6
10^5

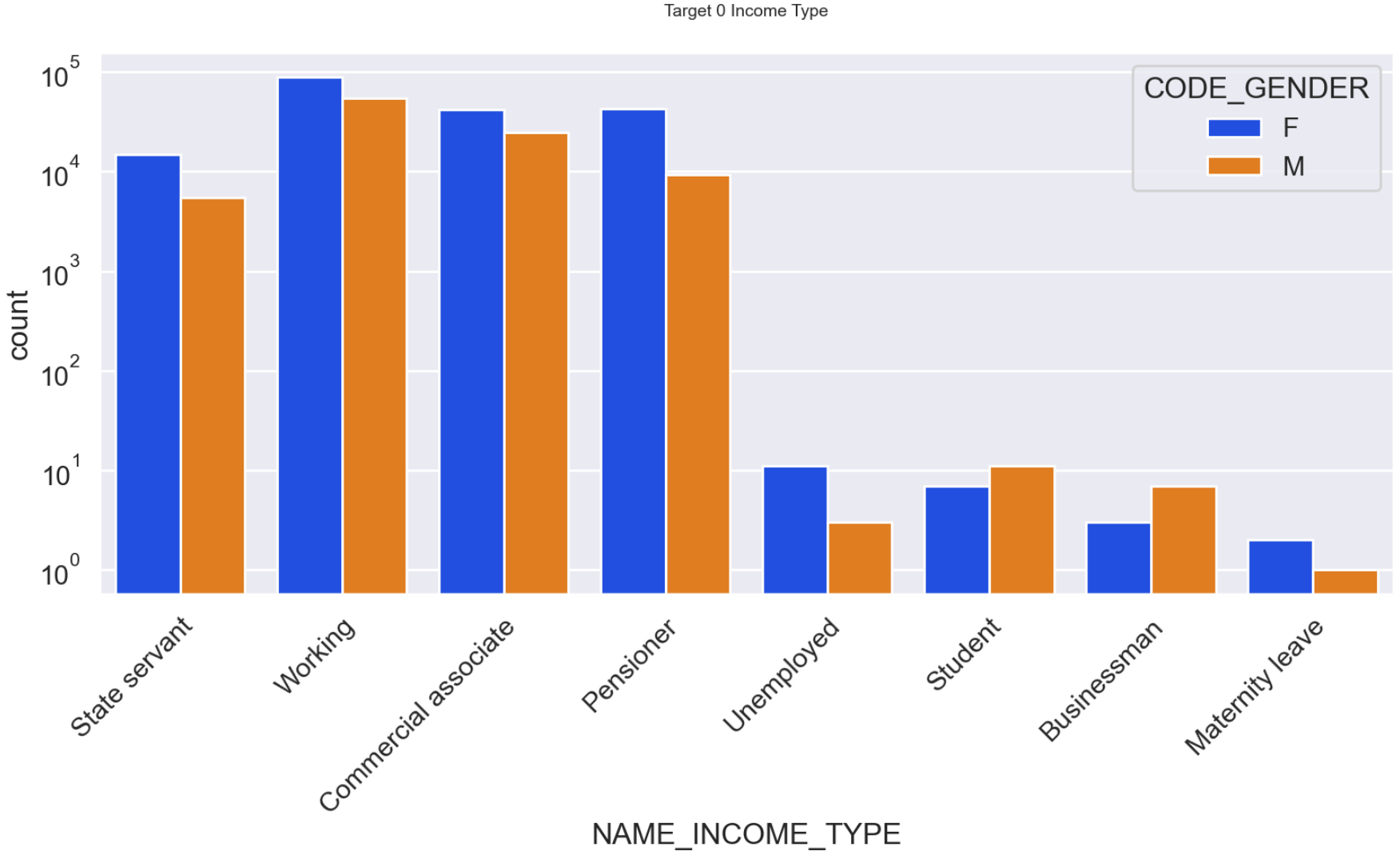
Conclusion:
From the BoxPlot we can see that there are severe outliners and their is quite a difference between 75% and max.So we are taking median value to replace those null values.


Analysis of AMT_REQ_CREDIT_BUREAU_DAY
AMT_REQ_CREDIT_BUREAU_DAY
0
2
4
6
8

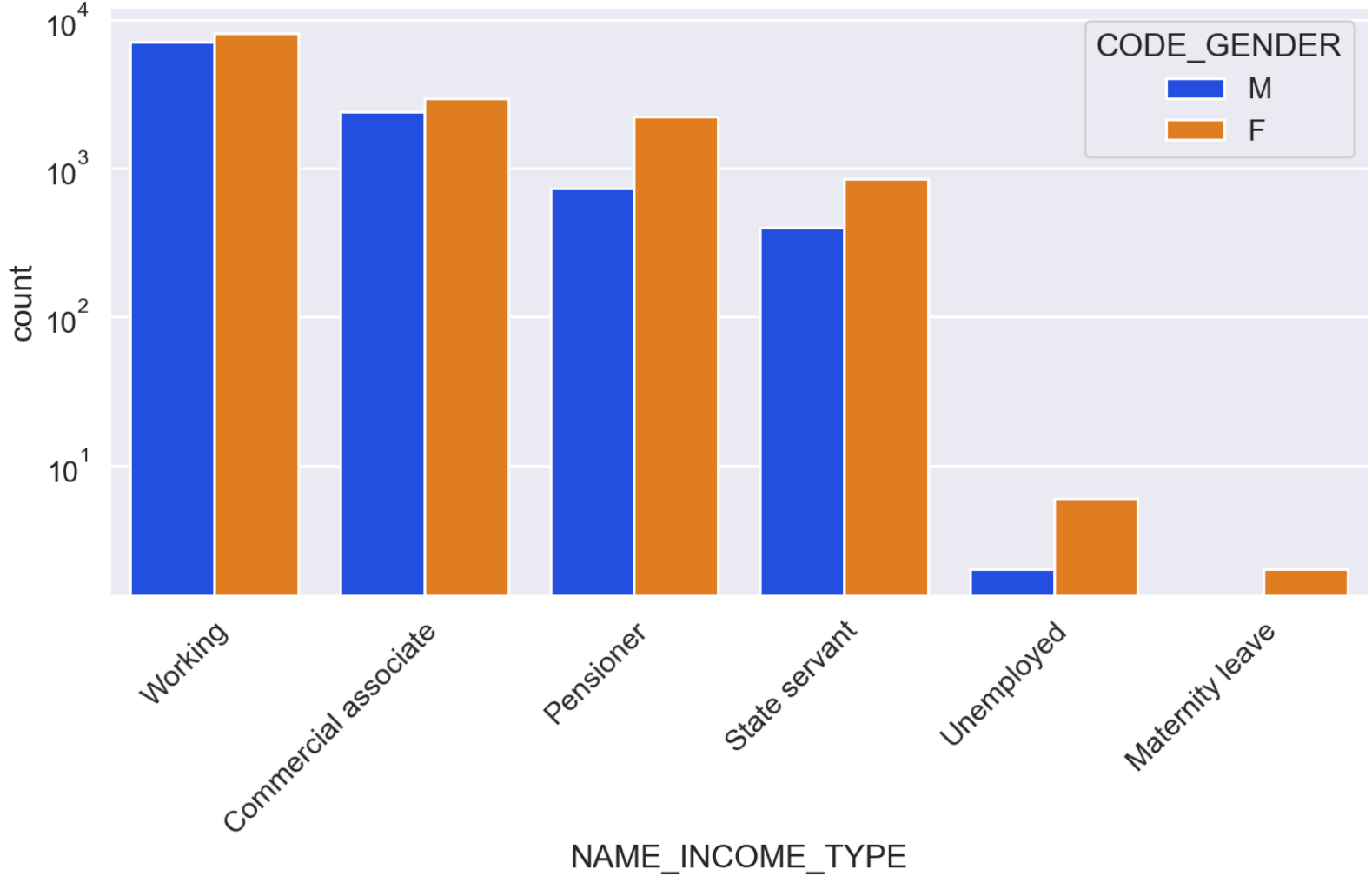
Handling outlier

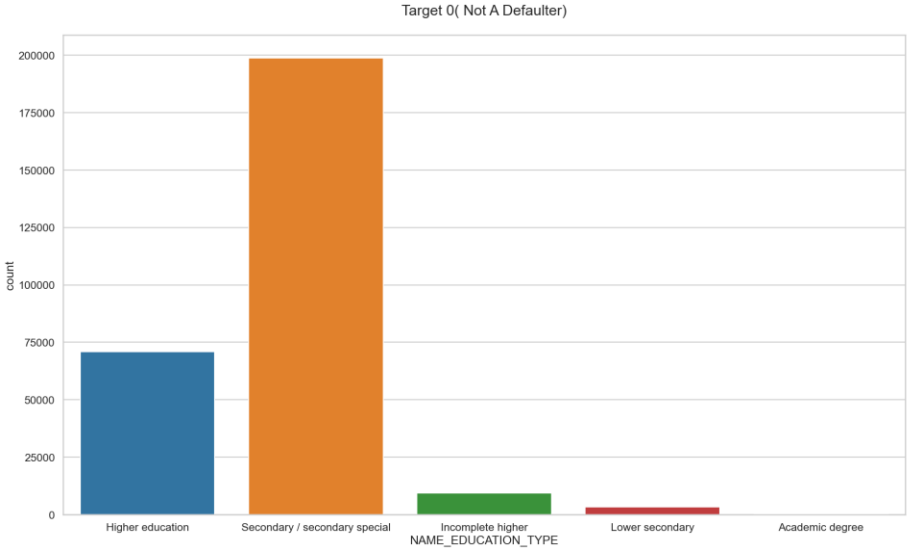

Target 0 Income Range
CODE_GENDER
F
M
count
10^4
10^3
10^2
0-25000
25000-50000
50000-75000
75000,100000
100000-125000
125000-150000
150000-175000
175000-200000
200000-225000
225000-250000
250000-275000
275000-300000
300000-325000
325000-350000
350000-375000
375000-400000
400000-425000
425000-450000
450000-475000
475000-500000
500000 and above
AMT_INCOME_RANGE

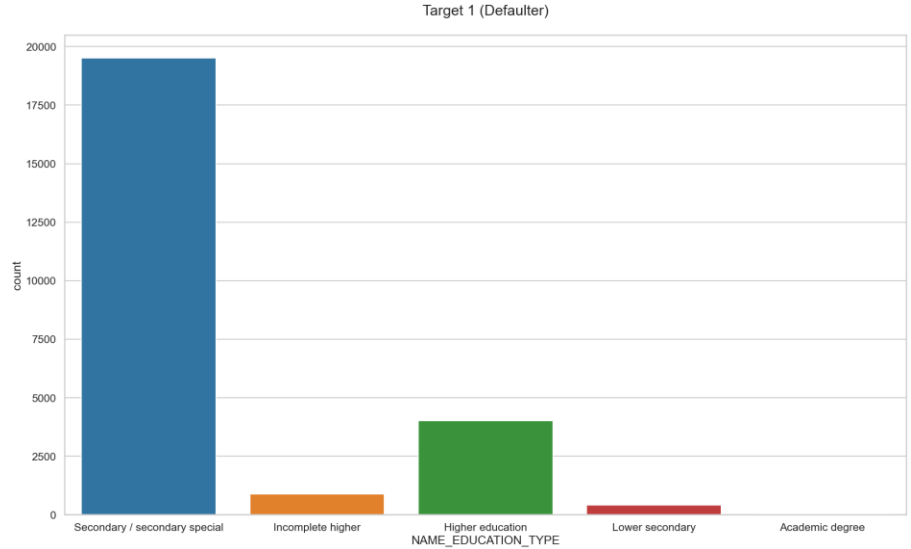

Target 1 Income Range
CODE_GENDER
M
F
count
10^3
10^2
10^1
0-25000
25000-50000
50000-75000
75000,100000
100000-125000
125000-150000
150000-175000
175000-200000
200000-225000
225000-250000
250000-275000
275000-300000
300000-325000
325000-350000
350000-375000
375000-400000
400000-425000
425000-450000
450000-475000
475000-500000
500000 and above
AMT_INCOME_RANGE

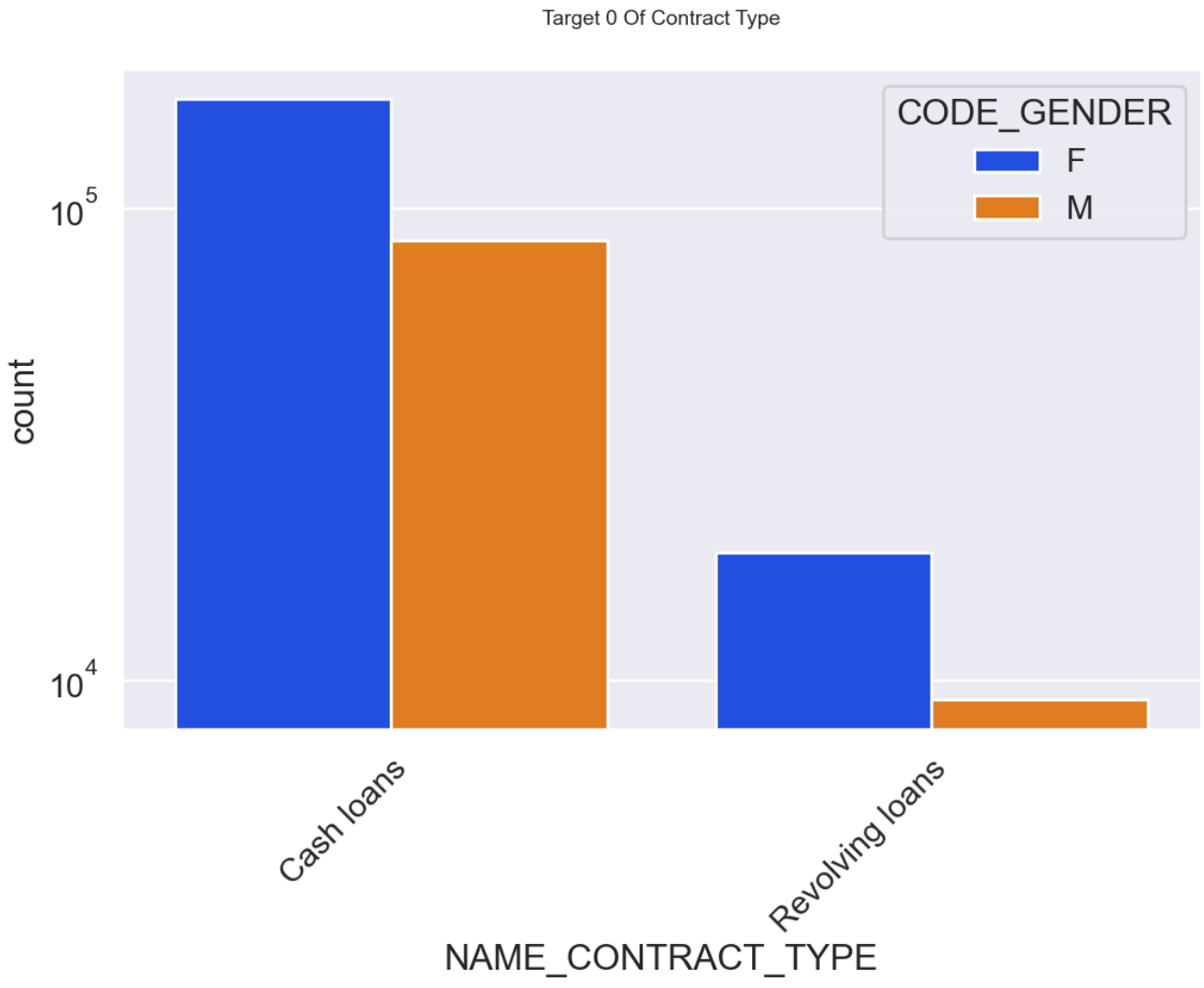

Target 0 Income Type
CODE_GENDER
F
M
count
10^5
10^4
10^3
10^2
10^1
10^0
State servant
Working
Commercial associate
Pensioner
Unemployed
Student
Businessman
Maternity leave
NAME_INCOME_TYPE

Target 1 Income Type

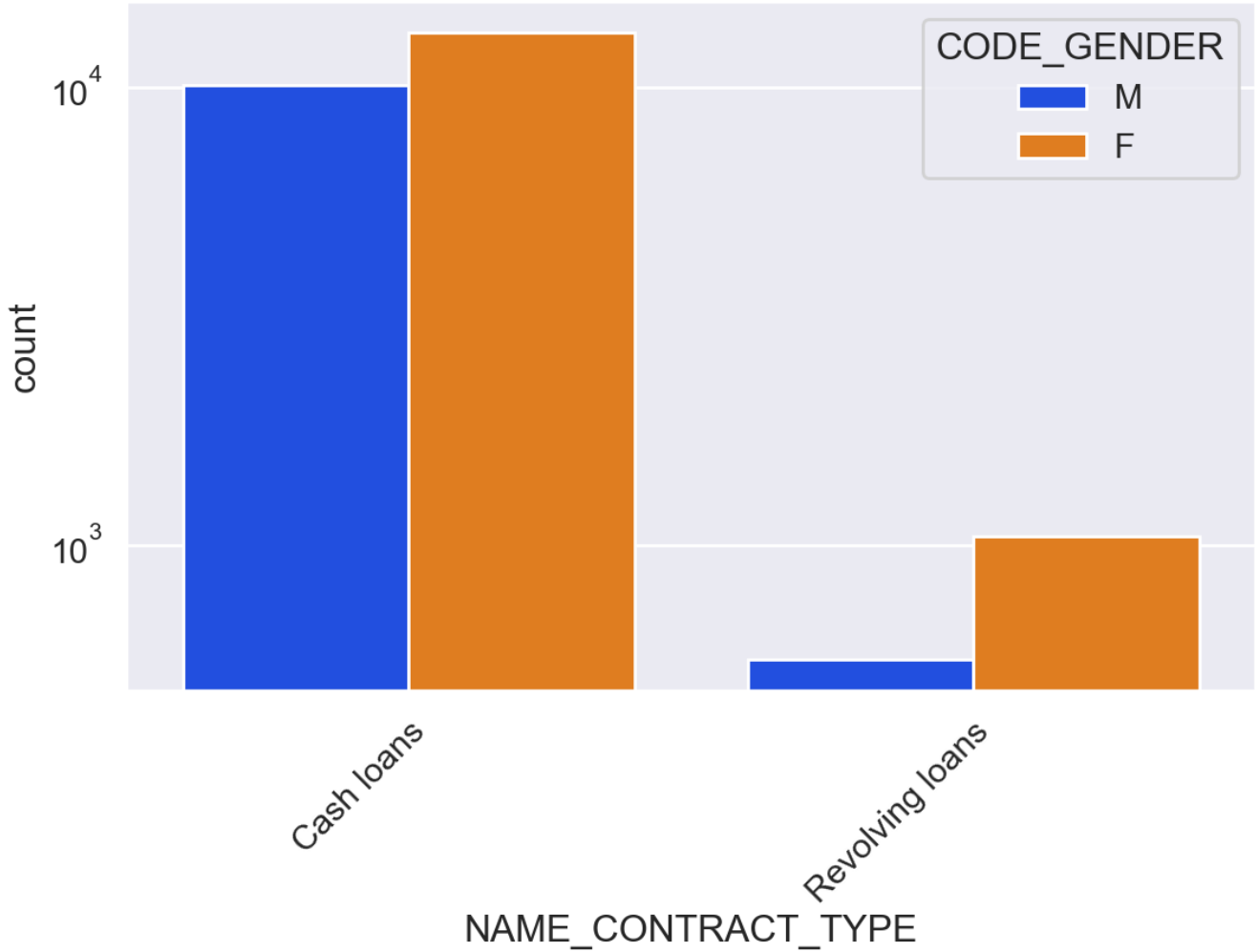



Target 0(Not A Defaulter)
count
200000
175000
150000
125000
100000
75000
50000
25000
0
Higher education
Secondary / secondary special
Incomplete higher
Lower secondary
Academic degree
NAME_EDUCATION_TYPE


Target 1 (Defaulter)
count
20000
17500
15000
12500
10000
7500
5000
2500
0
Secondary / secondary special
Incomplete higher
Higher education
Lower secondary
Academic degree
NAME_EDUCATION_TYPE


Target 0 Of Contract Type
CODE_GENDER
F
M
10^5
10^4
count
Cash loans
Revolving loans
NAME_CONTRACT_TYPE

Target 1 Contract Type



Conclusion

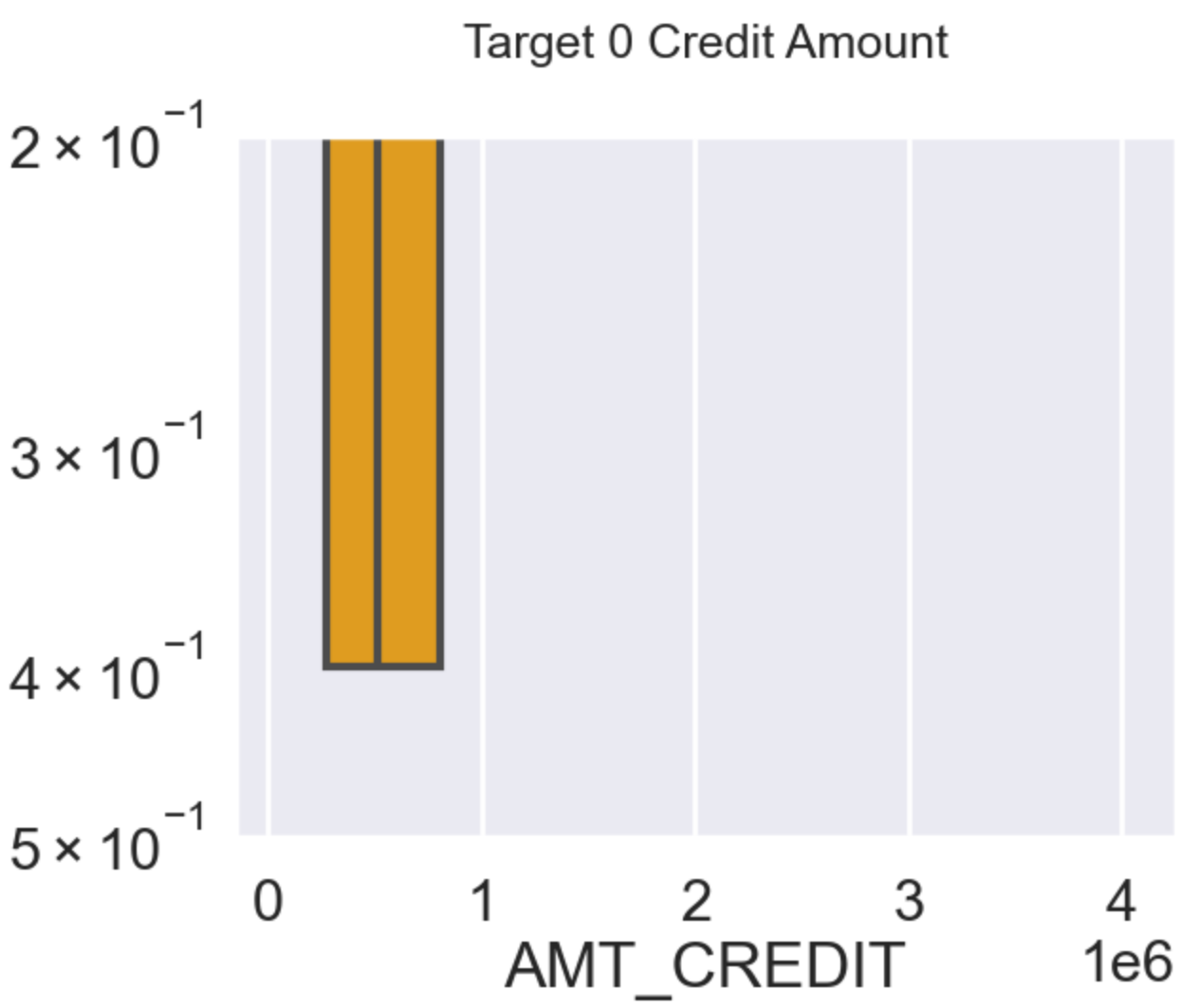
Points to be concluded from the above graph for Target=0 (Non-Defaulters).

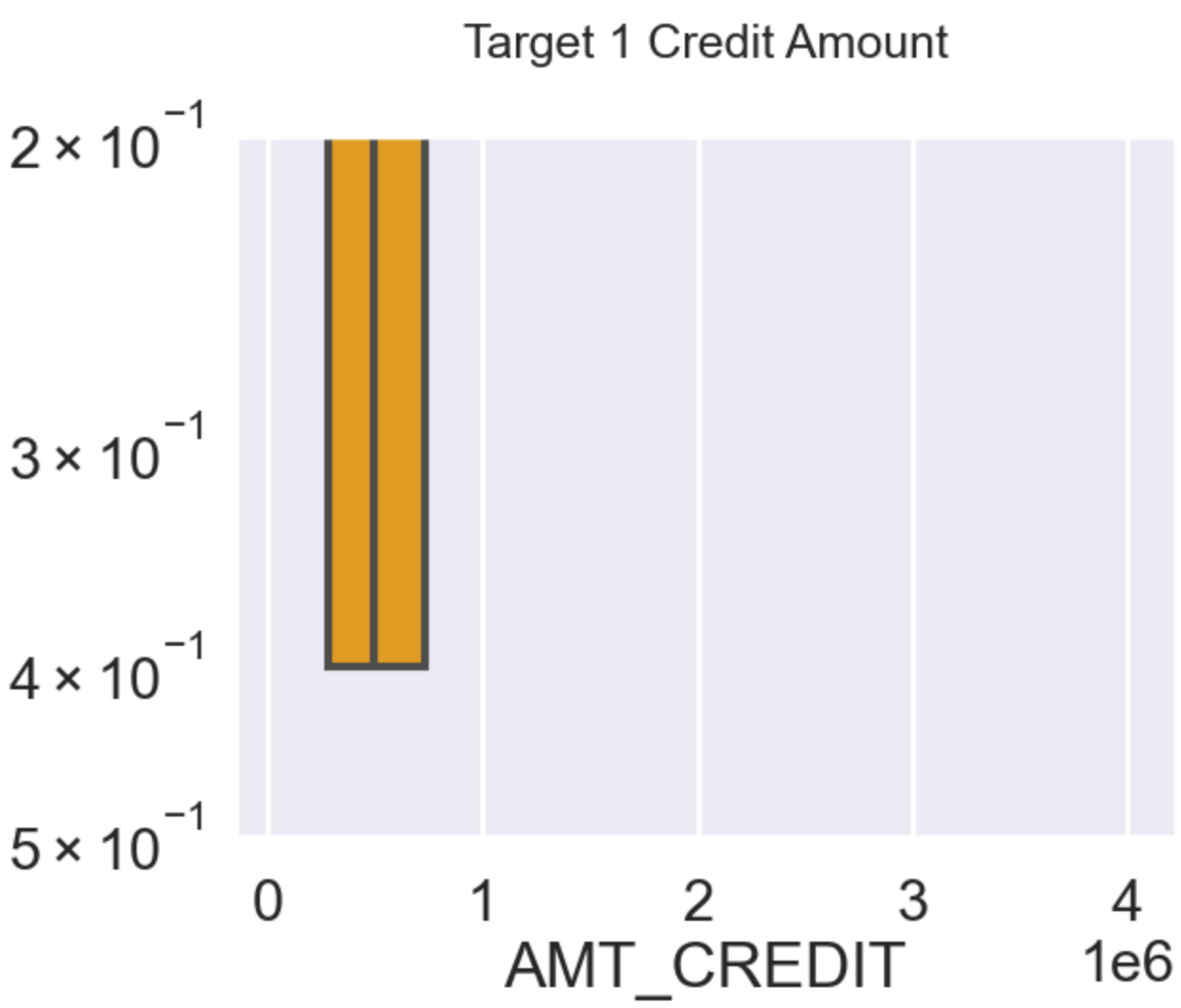
1. Cash Loan contracts have a higher number of credit than revolving loan contracts

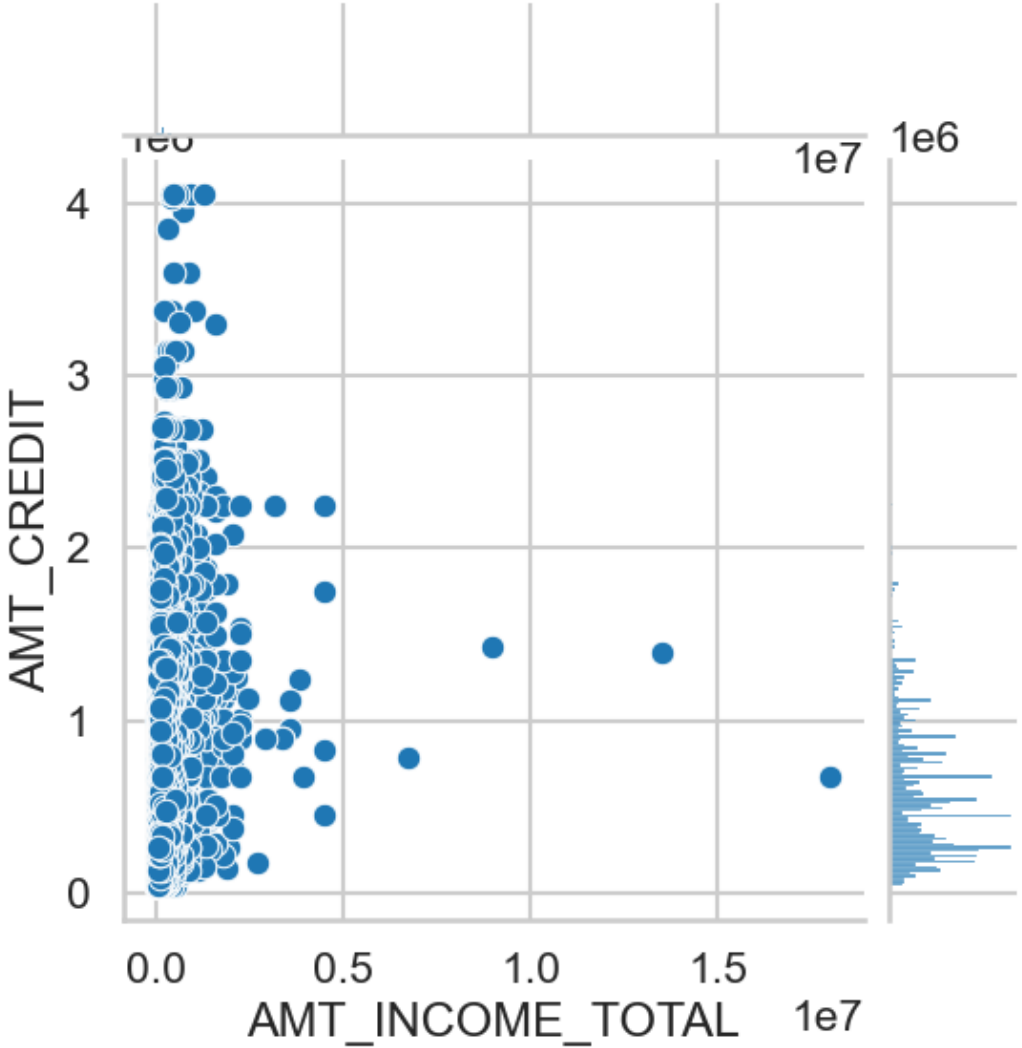
2. Count of female is more.

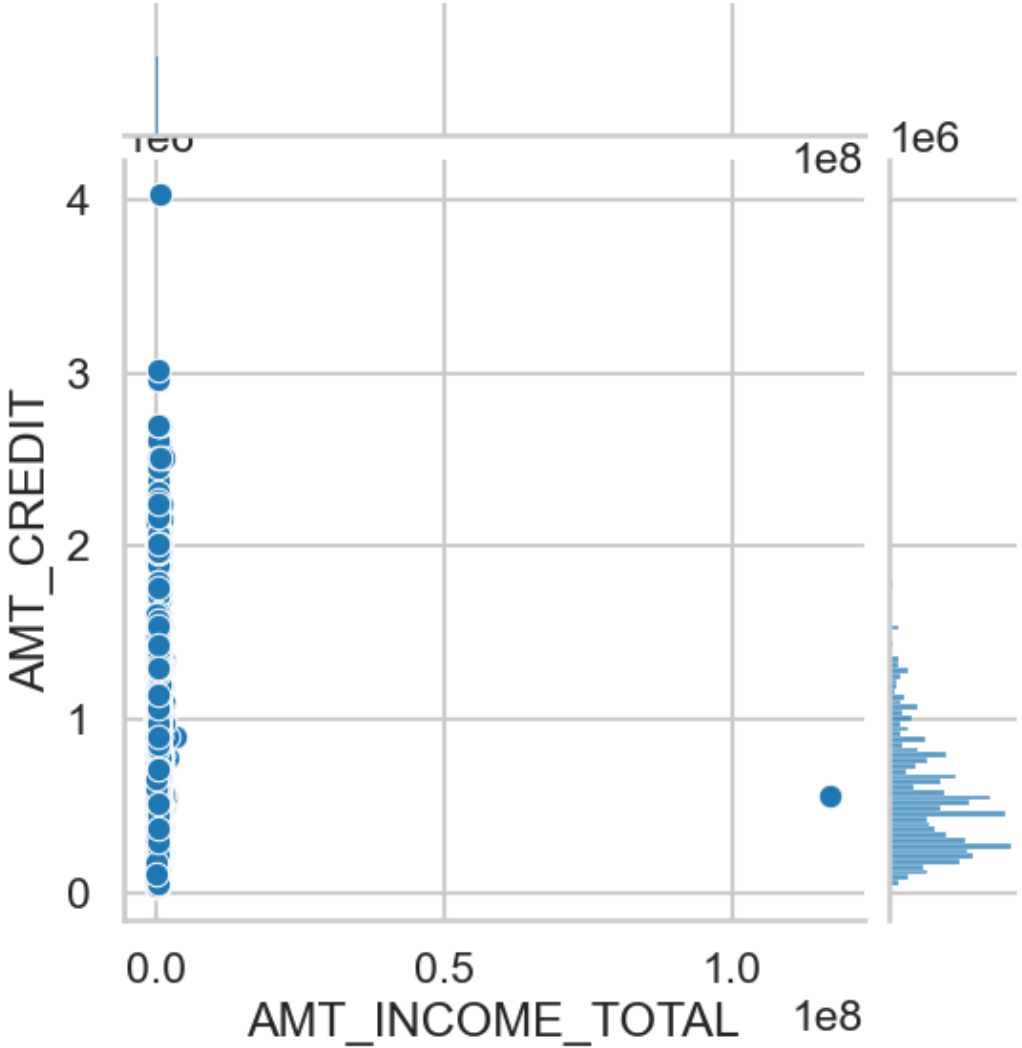
3.Cash Loan contracts have a higher number of credit than revolving loan contracts: This conclusion is based on the data shown in the graph, which suggests that among non-defaulters (Target=0), there is a larger proportion of cash loan contracts compared to revolving loan contracts.

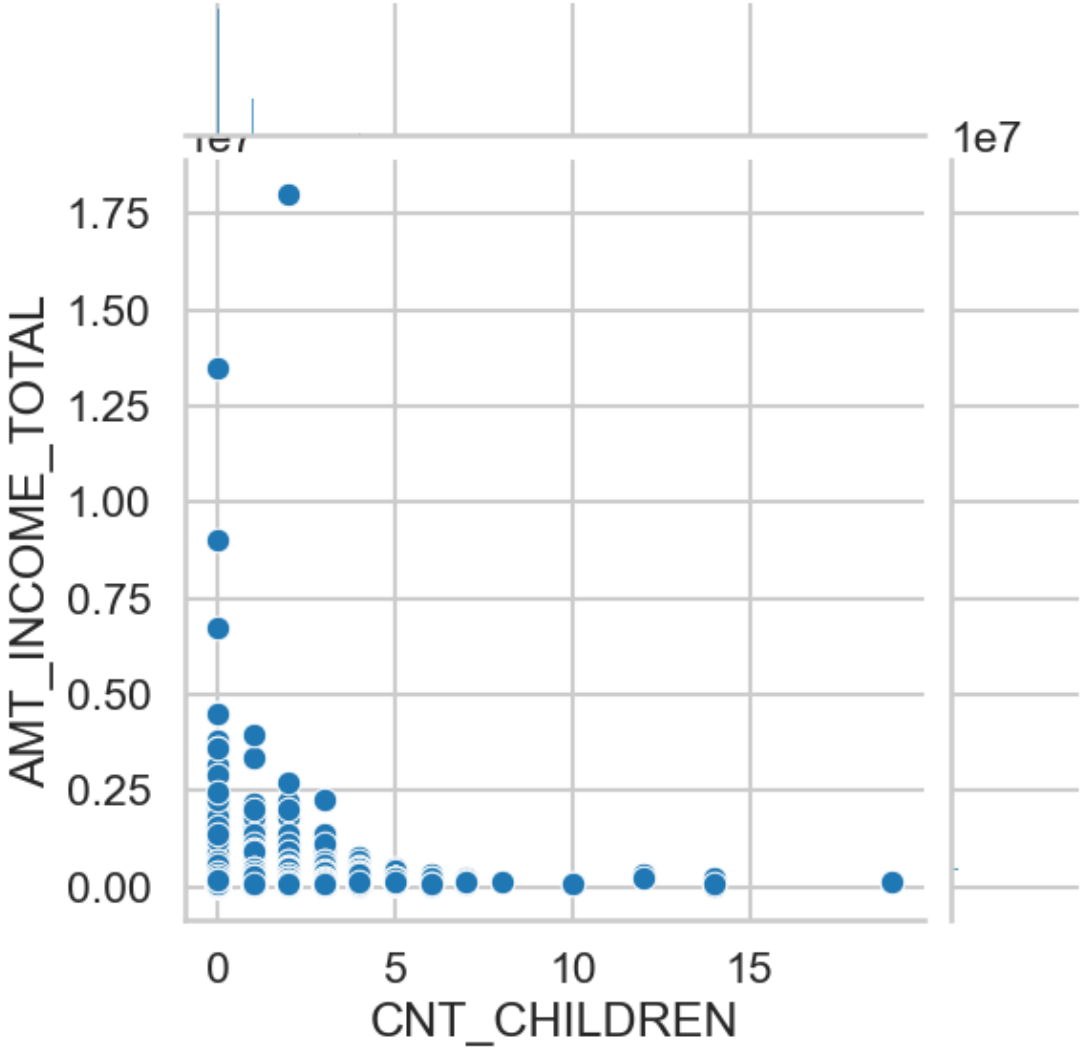
4.Count of females is more: The graph indicates that among non-defaulters, the count of females is higher than the count of males. This suggests that a greater number of females are categorized as non-defaulters in the dataset.

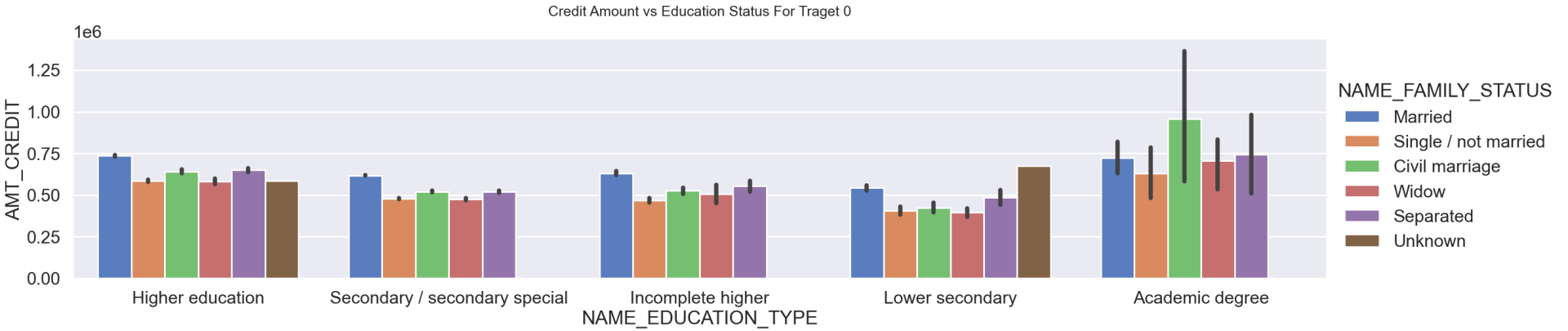

Target 0 Credit Amount
2×10^{-1}
3×10^{-1}
4×10^{-1}
5×10^{-1}
0
1
2
3
4
1e6
AMT_CREDIT

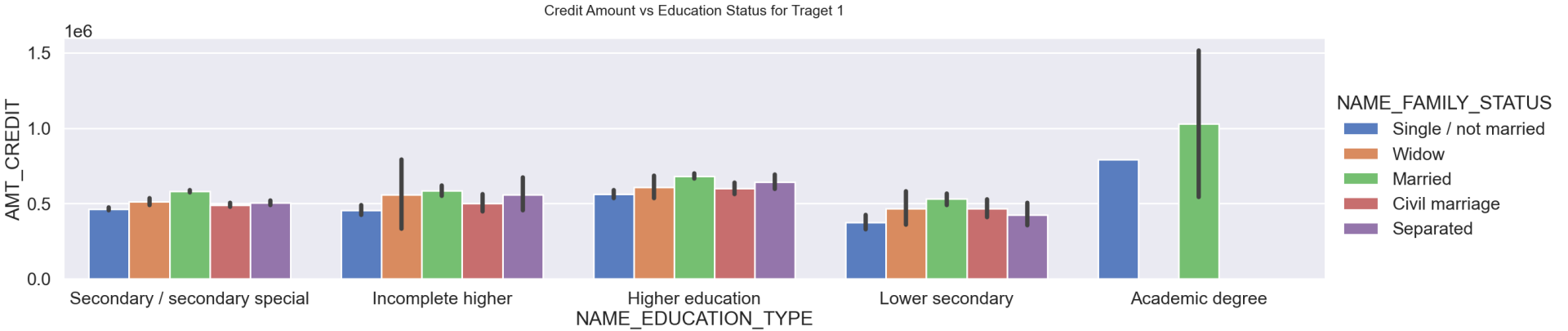

Target 1 Credit Amount
2×10^{-1}
3×10^{-1}
4×10^{-1}
5×10^{-1}
0
1
2
3
4
1e6
AMT_CREDIT


1e0
1e7
1e6
4
3
2
1
0
AMT_CREDIT
0.0
0.5
1.0
1.5
1e7
AMT_INCOME_TOTAL

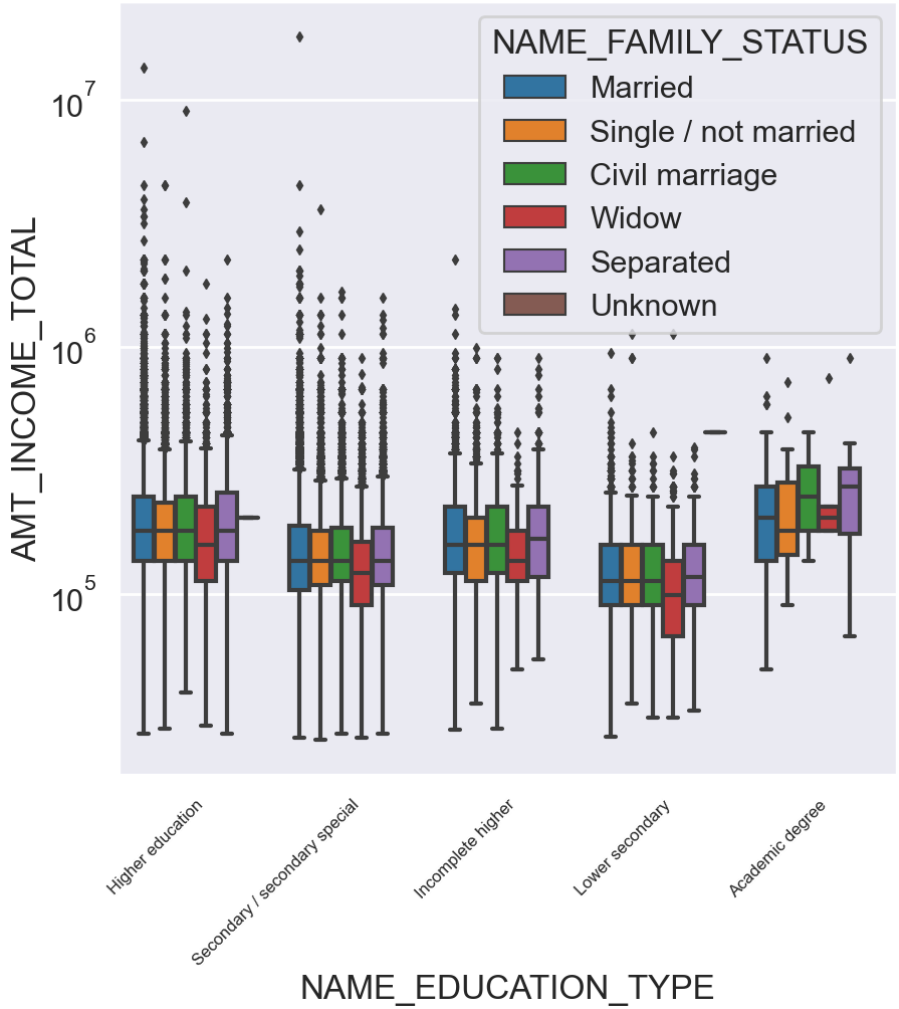

1e6
1e8
4
3
2
1
0
AMT_CREDIT
0.0
0.5
1.0
1e8
AMT_INCOME_TOTAL


1e7
1e7
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
AMT_INCOME_TOTAL
0
5
10
15
CNT_CHILDREN


Credit Amount vs Education Status For Traget 0
1e6
1.25
1.00
0.75
0.50
0.25
0.00
AMT_CREDIT
Higher education
Secondary / secondary special
Incomplete higher
Lower secondary
Academic degree
NAME_EDUCATION_TYPE
NAME_FAMILY_STATUS
Married
Single / not married
Civil marriage
Widow
Separated
Unknown


Credit Amount vs Education Status for Traget 1
1e6
1.5
1.0
0.5
0.0
AMT_CREDIT
Secondary / secondary special
Incomplete higher
Higher education
Lower secondary
Academic degree
NAME_EDUCATION_TYPE
NAME_FAMILY_STATUS
Single / not married
Widow
Married
Civil marriage
Separated

Income Amount vs Education Status For Non-Defaulters



conclusion:

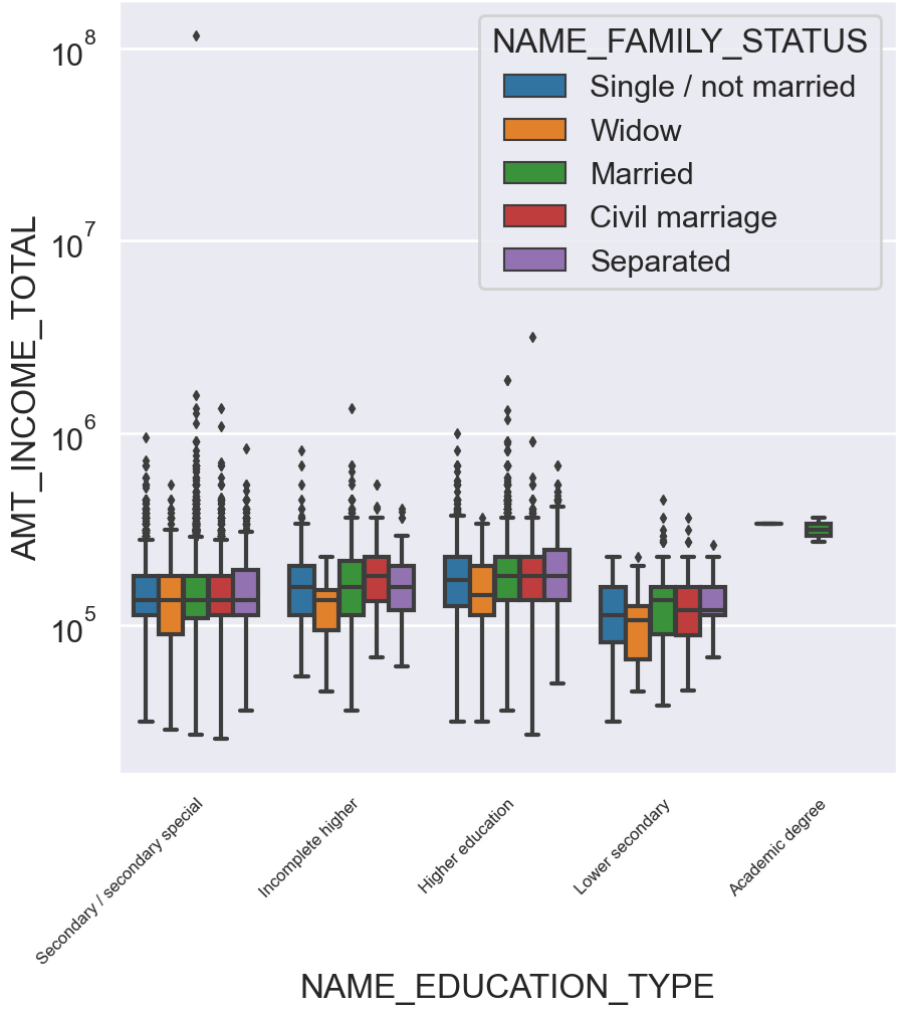
Points to be concluded from the above graph for target = 0 (Non-Defaulters).

1. For Education type 'Higher education' the income amount mean is mostly equal with family status. It does contain many outliers.

2. Less outlier are having for Academic degree but they are having the income amount is little higher that Higher education.

3. Lower secondary of civil marriage family status are have less income amount than others.

Income Amount vs Education Status For Defaulters



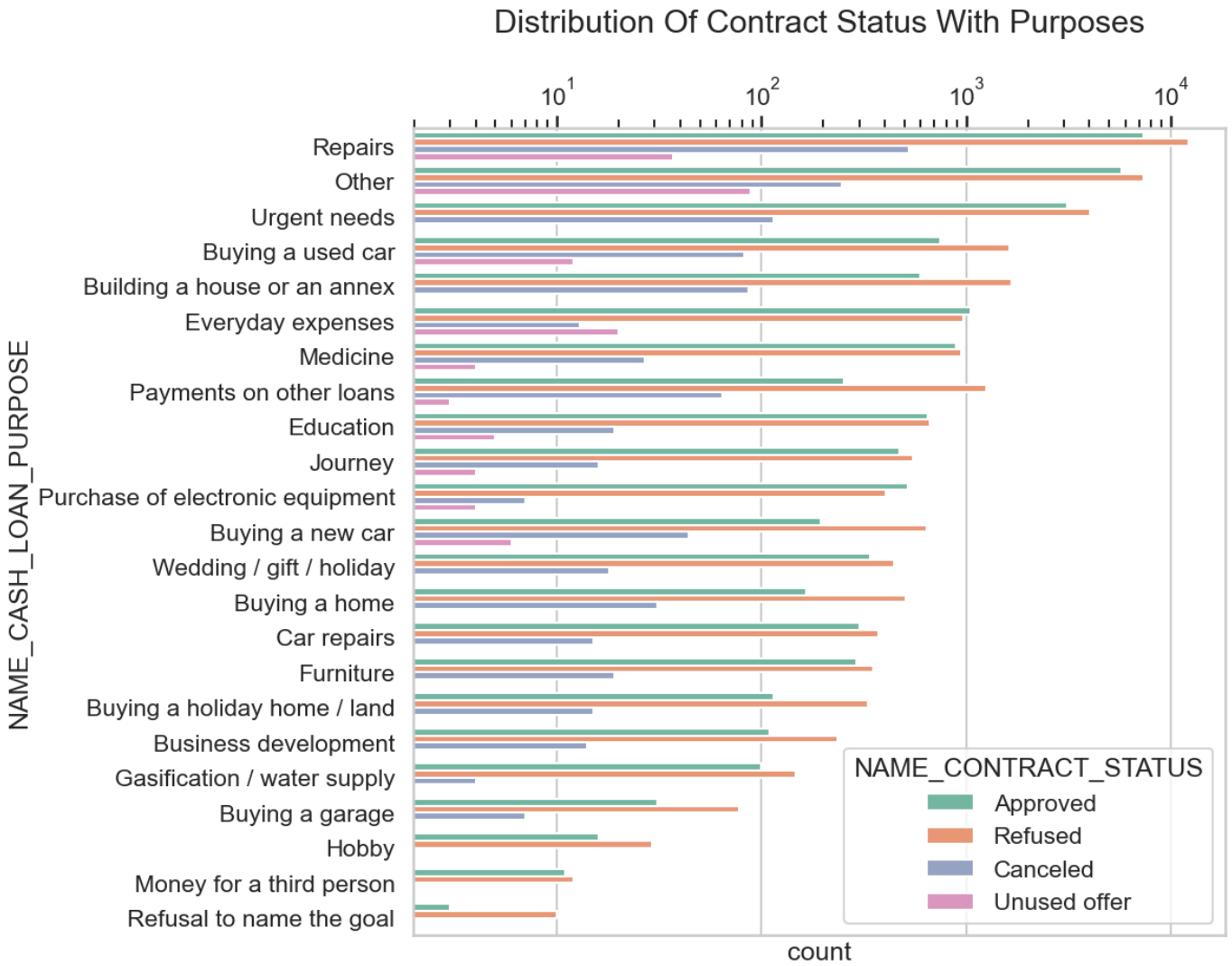
Conclusion:

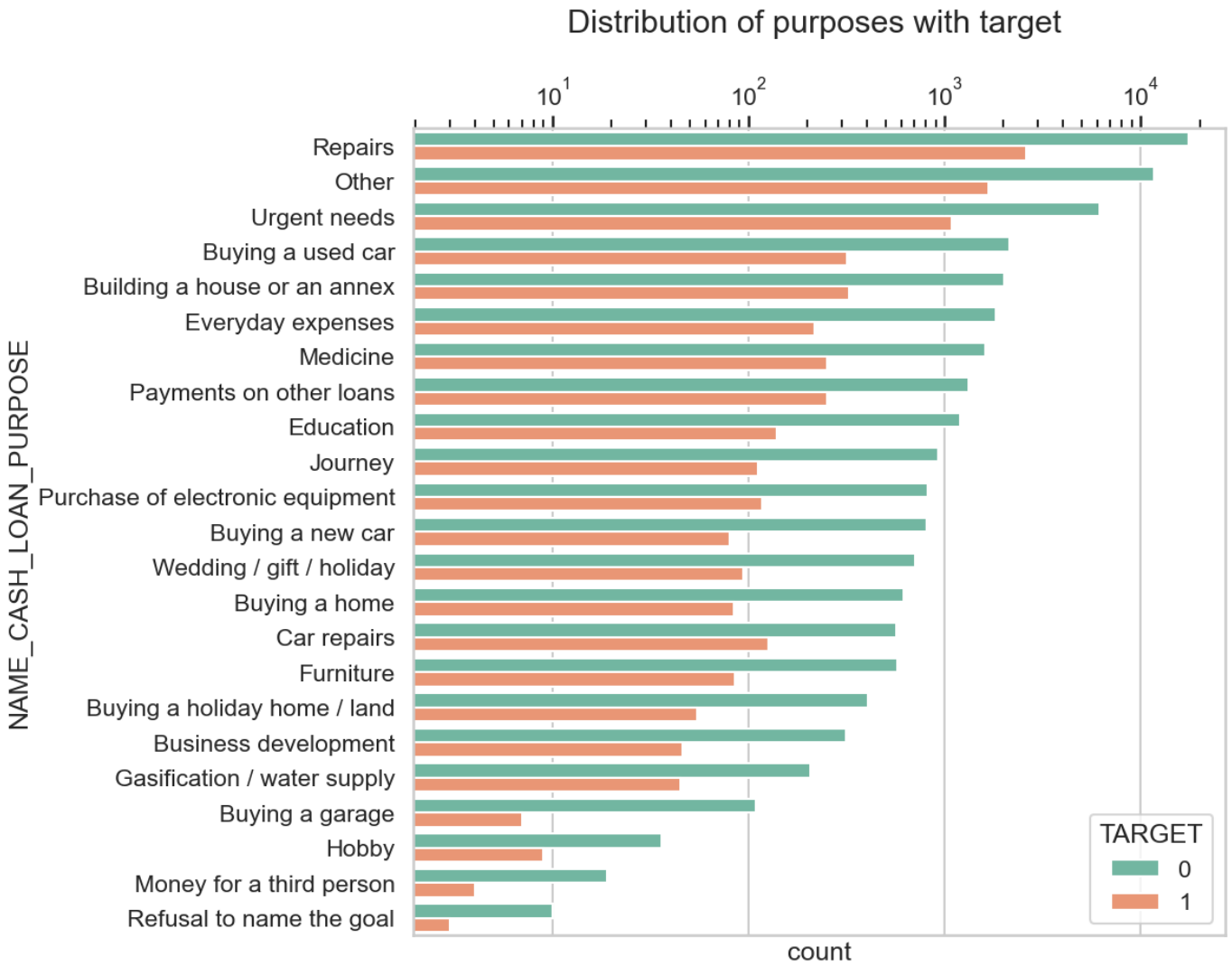
Points to be concluded from the above graph for target = 1 (Defaulters).

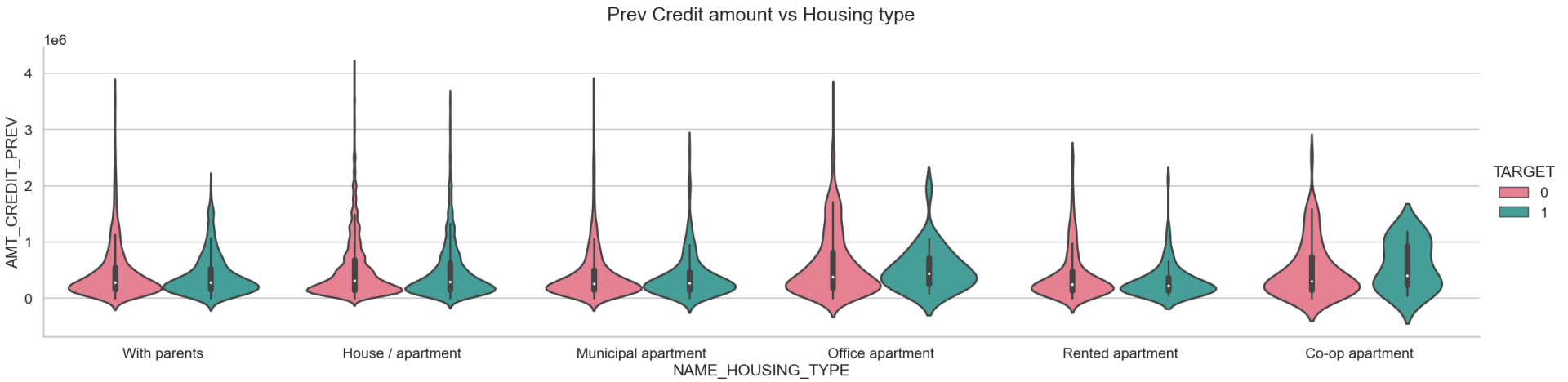
1. Have some similarity with Target0, From above boxplot for Education type 'Higher education' the income amount is mostly equal with family status.

2. No outlier for Academic degree but there income amount is little higher than that Higher education.

3. Lower secondary are having less income amount than others.


Distribution Of Contract Status With Purposes
10^1
10^2
10^3
10^4
NAME_CASH_LOAN_PURPOSE
Repairs
Other
Urgent needs
Buying a used car
Building a house or an annex
Everyday expenses
Medicine
Payments on other loans
Education
Journey
Purchase of electronic equipment
Buying a new car
Wedding / gift / holiday
Buying a home
Car repairs
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
NAME_CONTRACT_STATUS
Approved
Refused
Canceled
Unused offer
count


Distribution of purposes with target
10^1
10^2
10^3
10^4
Repairs
Other
Urgent needs
Buying a used car
Building a house or an annex
Everyday expenses
Medicine
Payments on other loans
Education
Journey
Purchase of electronic equipment
Buying a new car
Wedding / gift / holiday
Buying a home
Car repairs
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
NAME_CASH_LOAN_PURPOSE
TARGET
0
1
count


Prev Credit amount vs Housing type
1e6
4
3
2
1
0
AMT_CREDIT_PREV
With parents
House / apartment
Municipal apartment
Office apartment
Rented apartment
Co-op apartment
NAME_HOUSING_TYPE
TARGET
0
1

Conclusion:

We can conclude that bank should avoid giving loans to the housing type of co-op apartment as graph shows difficulties in payment.

Assignment Submission Guidelines

upGrad

3 Key Steps

- Ensure that the solution file/files are in the ***correct format***
- Ensure that the solution files are in a ***zipped folder***
- Upload the zipped folder on the upGrad platform

Summary

- Ensure that the solution file/files are in the ***correct format***
  - *Convert the PPT file to pdf using the online tool https://online2pdf.com/*
- Ensure that the solution files are in a ***zipped folder***
  - *Right click on the folder, which contains all your solution files, and then select Compress to ZIP file*
- Upload the zipped folder on the upGrad platform
  - *Download the zipped folder after uploading, and in case of any discrepancies, remove the previous submission and re-upload again.*