CREDIT EDA CASE STUDY

presentation

RAHUL GARIGANTI

Business Objectives

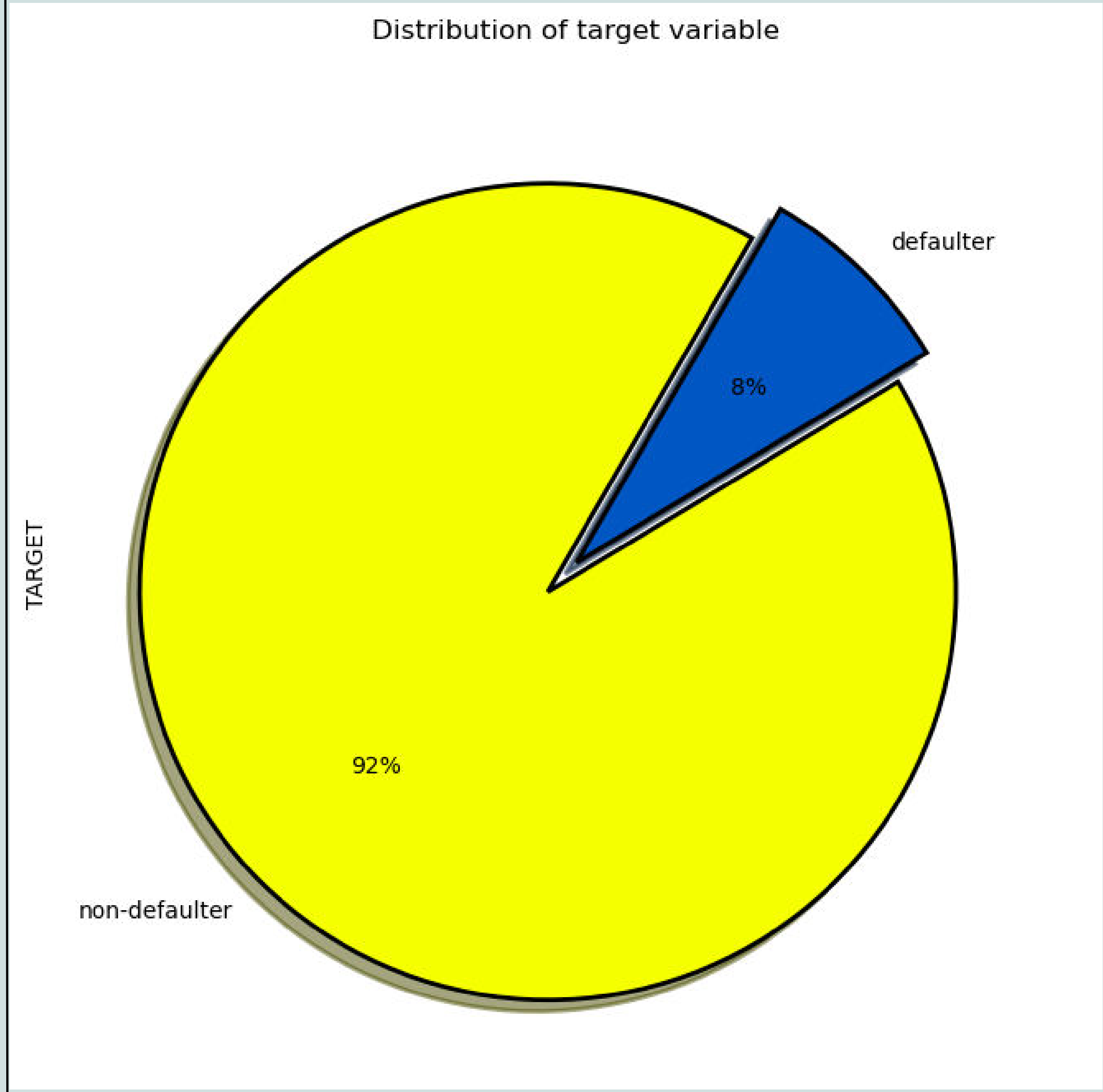
This case study aims to identify patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable 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 utilise 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.

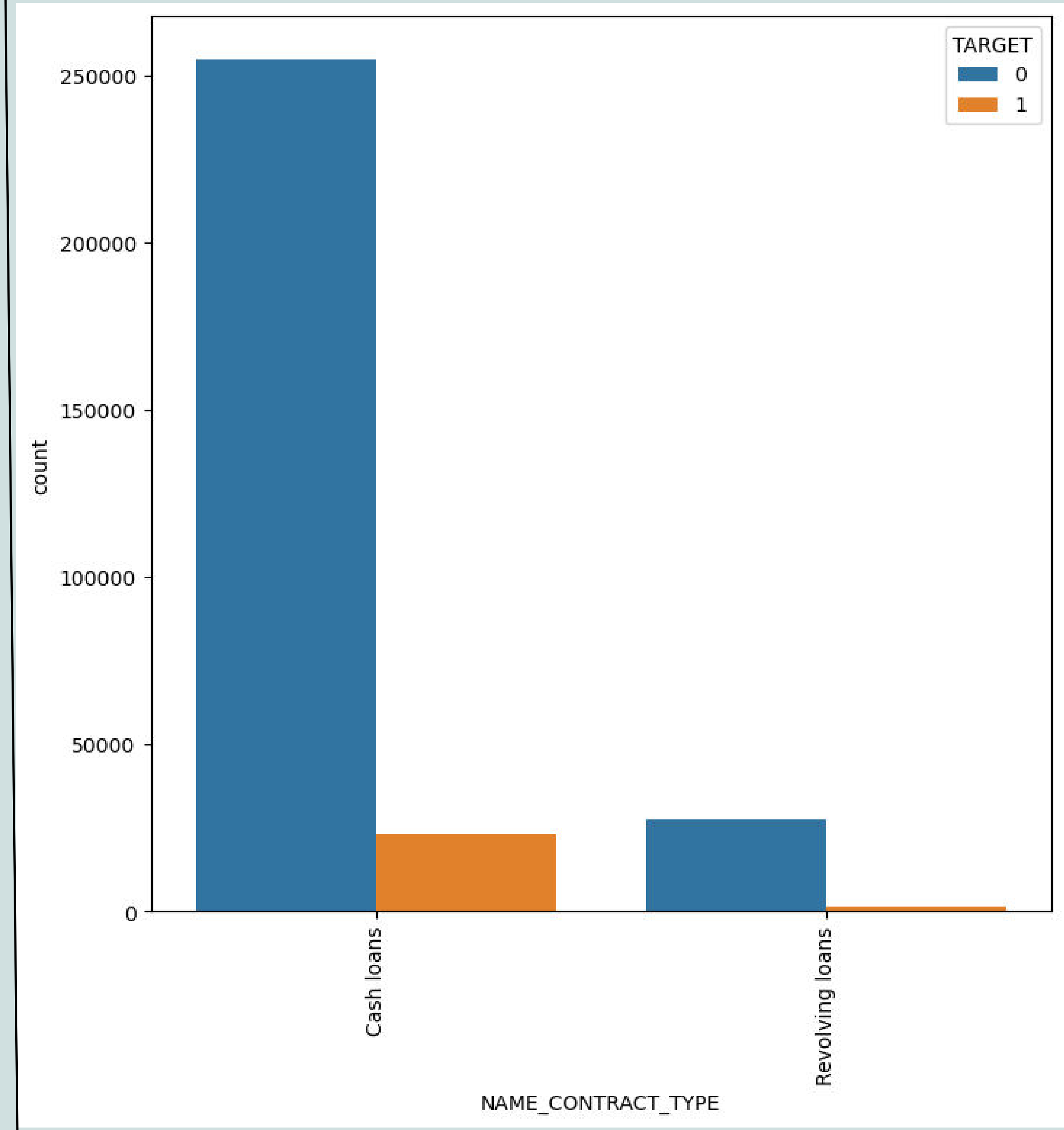
Data Imbalance

- This data is highly imbalanced
- Defaulter : Non-Defaulter = 8 : 92 = 2 : 23



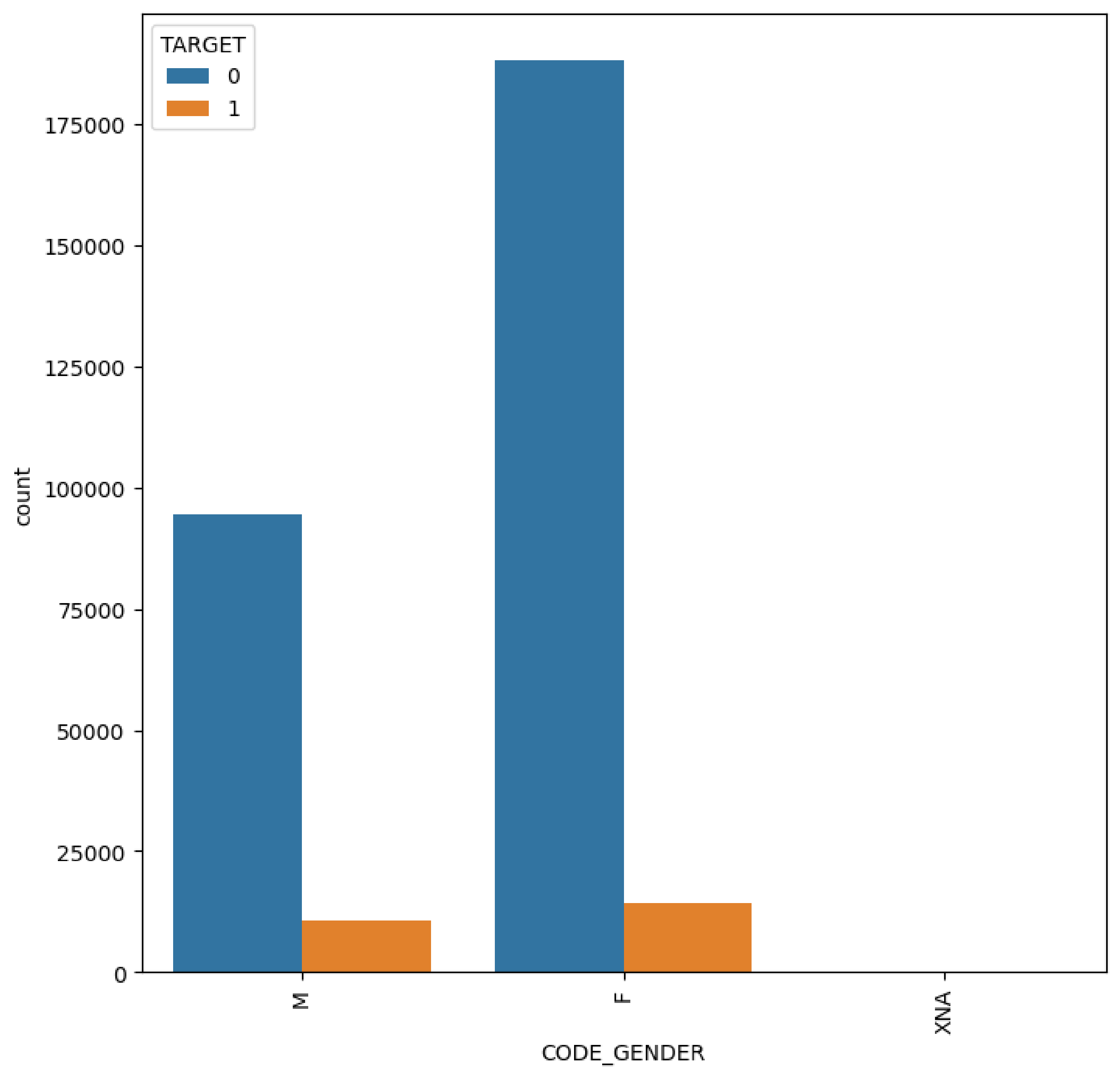
distribution of loans for differnt name contract types

- the plot consists of two types of name contract type
- among them both most of the loan that were given is of type cash loans
- most of the loans of cash type are are tured out to be a non defaulters
- so from this graph we can conclude that most of the loans that were given are of type cash



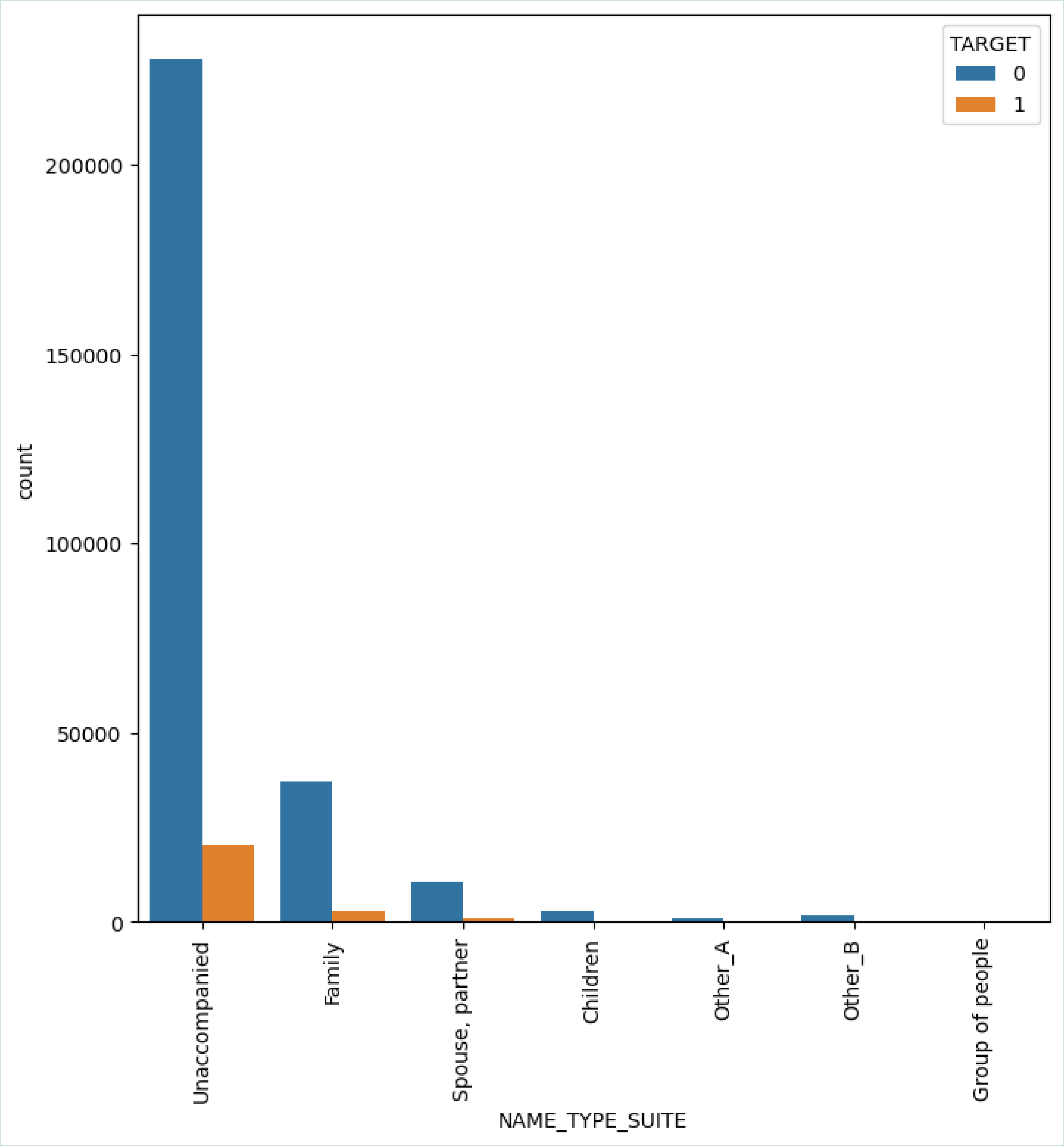
defaukt rate w.r.t code gender

- this plot shows the distribution of the loan for genders
- females are the one who got the most of the loans
- but we cannot say that males are more likely to default as this garph only shows the distribution of loans alon with target
- in order to conckude that we need to check the default ratio
- but here we can conclude that females have receved the most of the loans as compaired to males and others



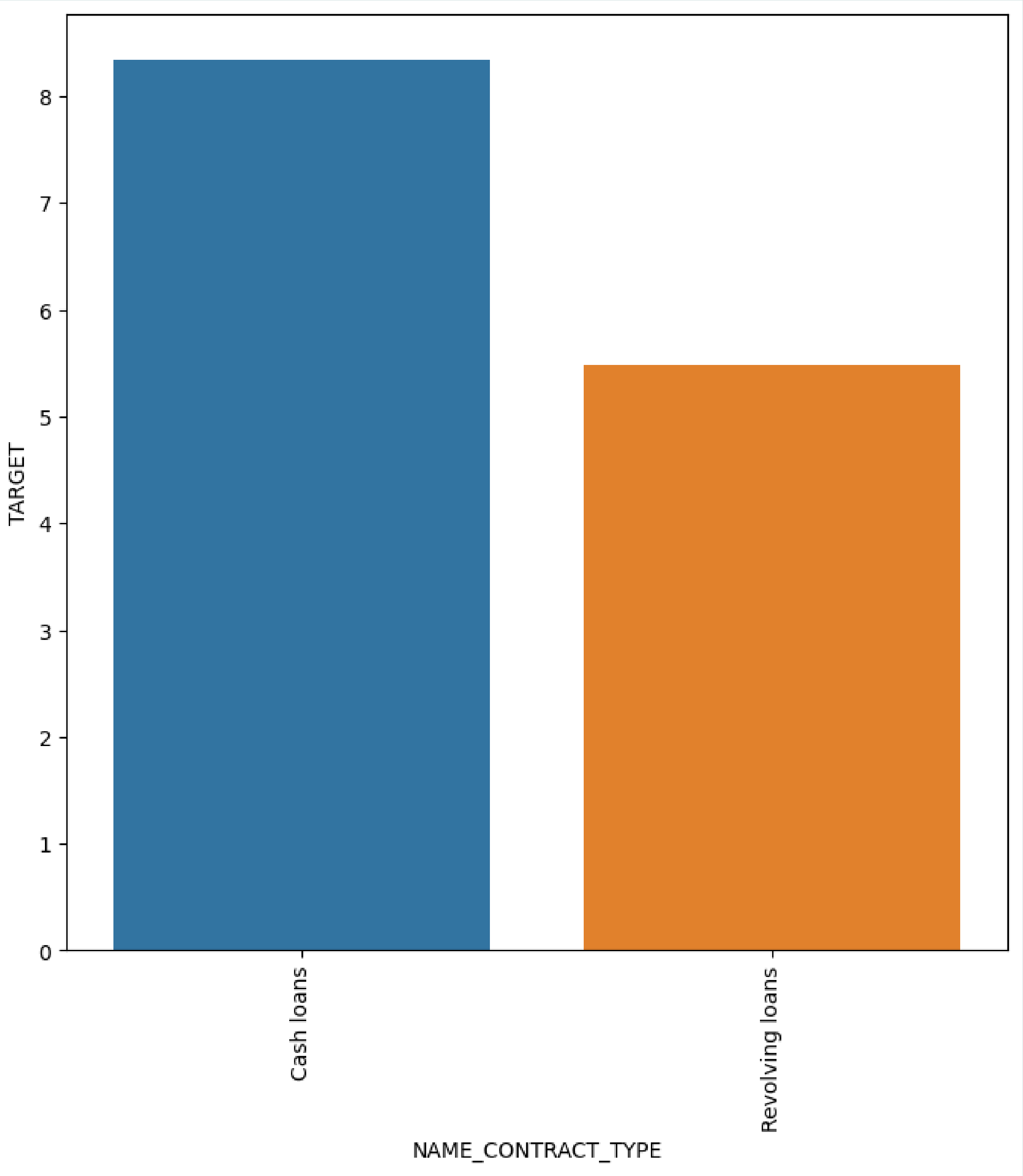
name type suite

- **this** plot represents name type suite
- here we can see that unaccompanied has taken the most of the loan
- among which most of the them are non defaulters
- here we can conclude that unaccompanied has the largest loan distribution as compaired to others
- but we cannot say it is the safest without looking at its default rate



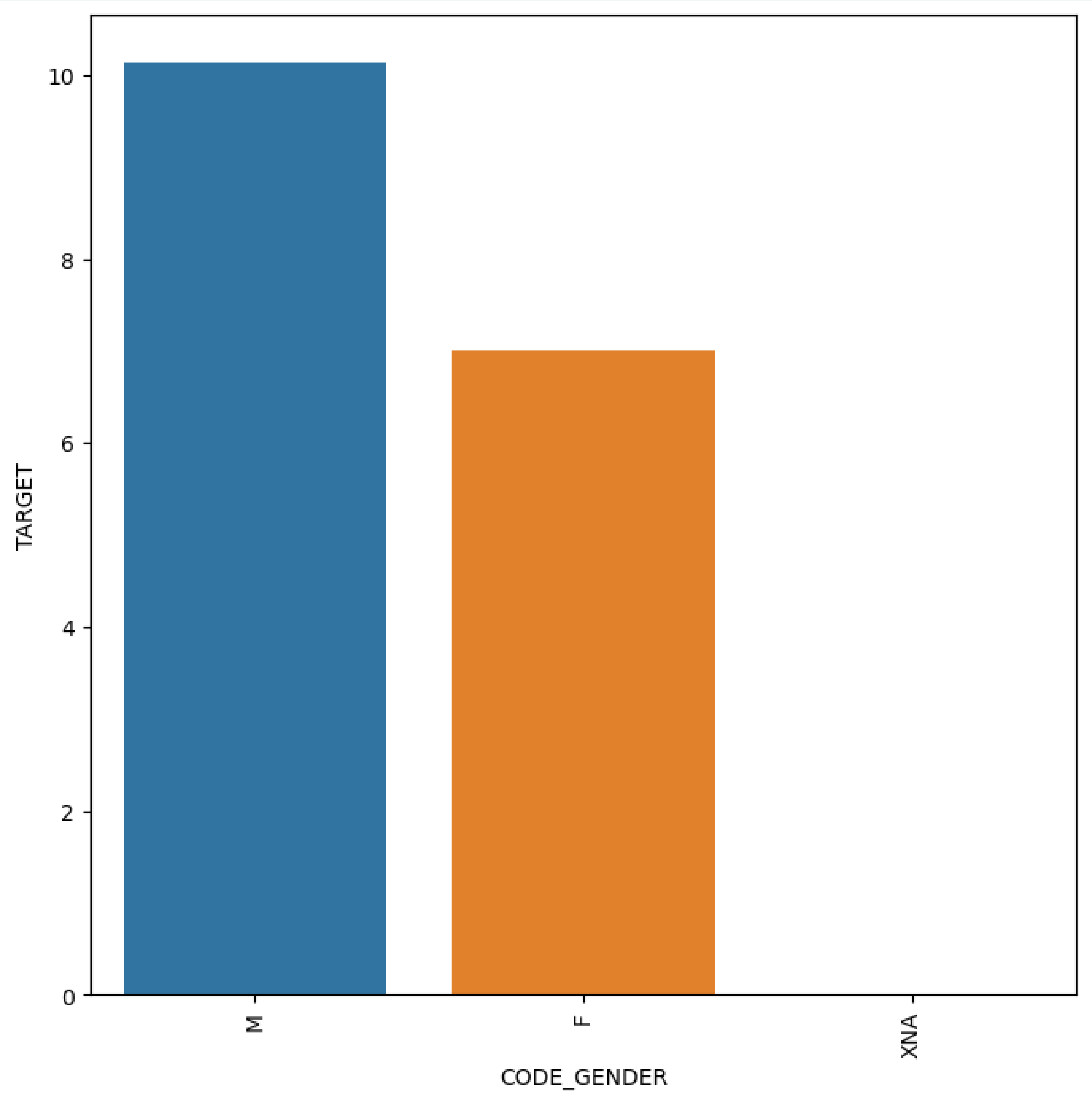
Univariate Analysis

- these graphs shows the default rate for each type of loans
- there are two types of loans that were given cash loans and revolving loans
- most of the loans that were given are of type cash loans
- we can see that default rate for cash loan is very high as compaired to that of revolving loans
- but as we can see the distribution of the cash loan is way more high than that of revolving loans
- so we can say that cash loan is the safest because it is manintaing the default rate arround 8 even after being the most given loan



default for code gender

- the graph consists of genders on the x axis and default rate on the y axis
- although females are the one who taken most of the loan i.e loan distribution is very high for the females as compaires to males
- but here as we are see that default rate is way higher for males
- so from these obsevation we can say that males are more like to default that females
- males are having the default rate of 10
- females are having the default rate of 7



default rate for NAME TYPE SUITE

- these is the graph of default rate vs name type suite

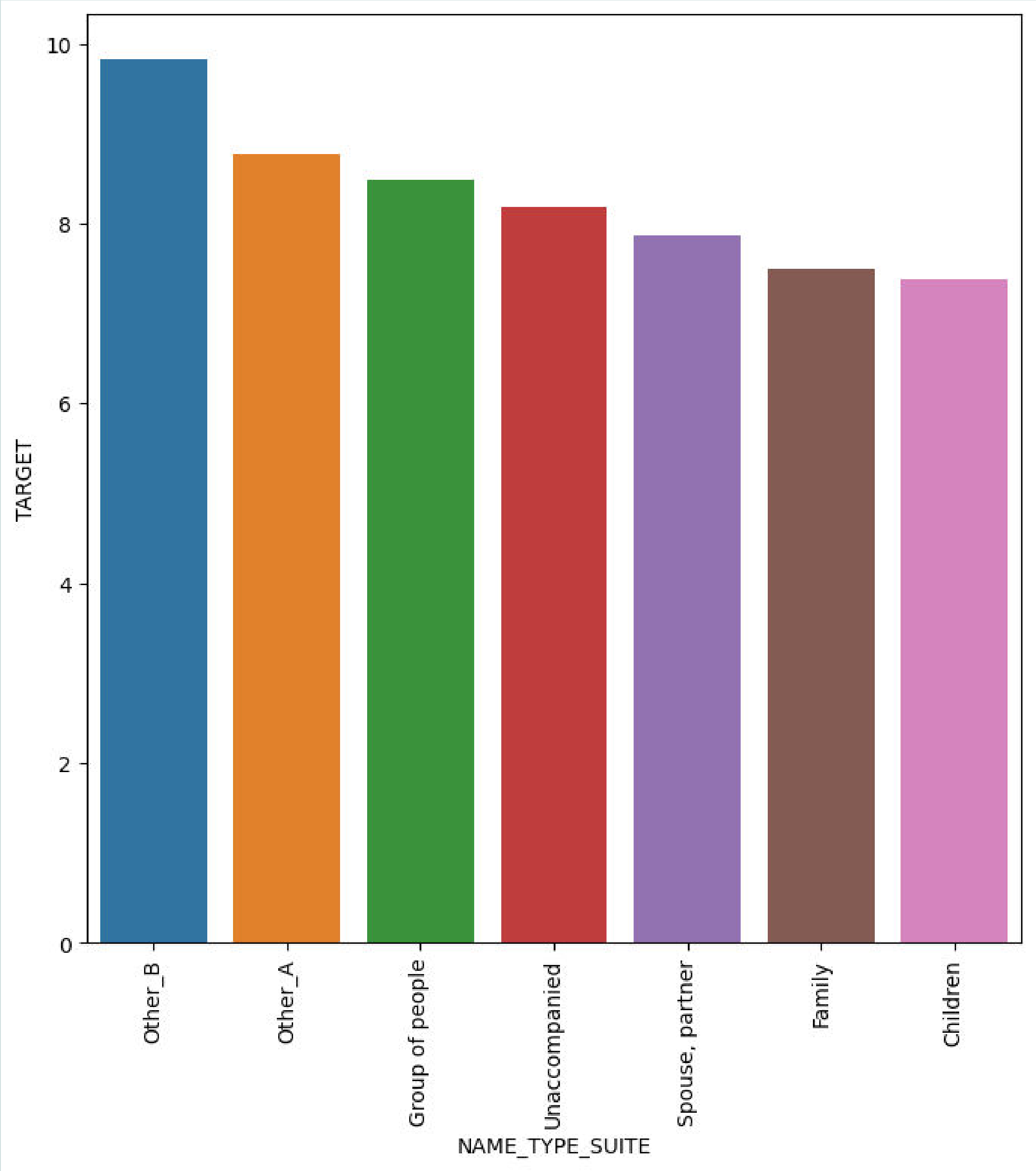
- there are seven types of name type suite

- if we were look at the distribution graph of name type suite then we will find that unaccompanied are the one who have taken most of the loans

- but here we can see that other_B have the highest default rate more than 9%

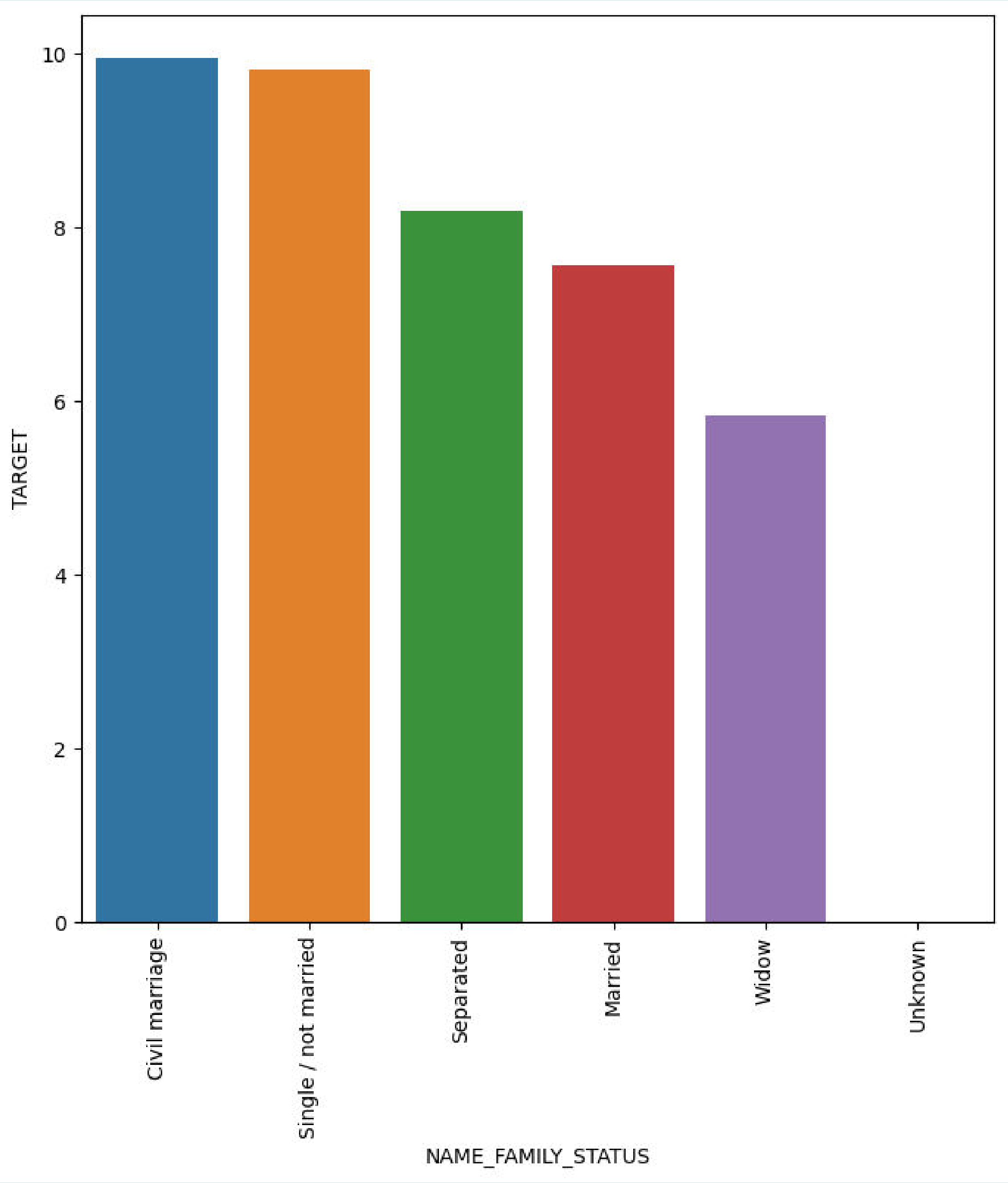
- if we look that unaccompanied it has one of the safest default rate despite being the most loans given

- so we can conclude that unaccompanid is the safest name type suite



default rate for differnt kinds of family sataus

- it is the graph that shows the default rate for each type of family satuts
- from the grapg we can see that civil marage has the highest default rate so we can say the it is the most risky among all
- married has the default rate of about 7.8% despite being having the most loans given
- so we can conclude that marraged people have the lower risk of default .
- although widow has the lowest default rate but when we look at the distribution the widow have taken the loan are vey few so .

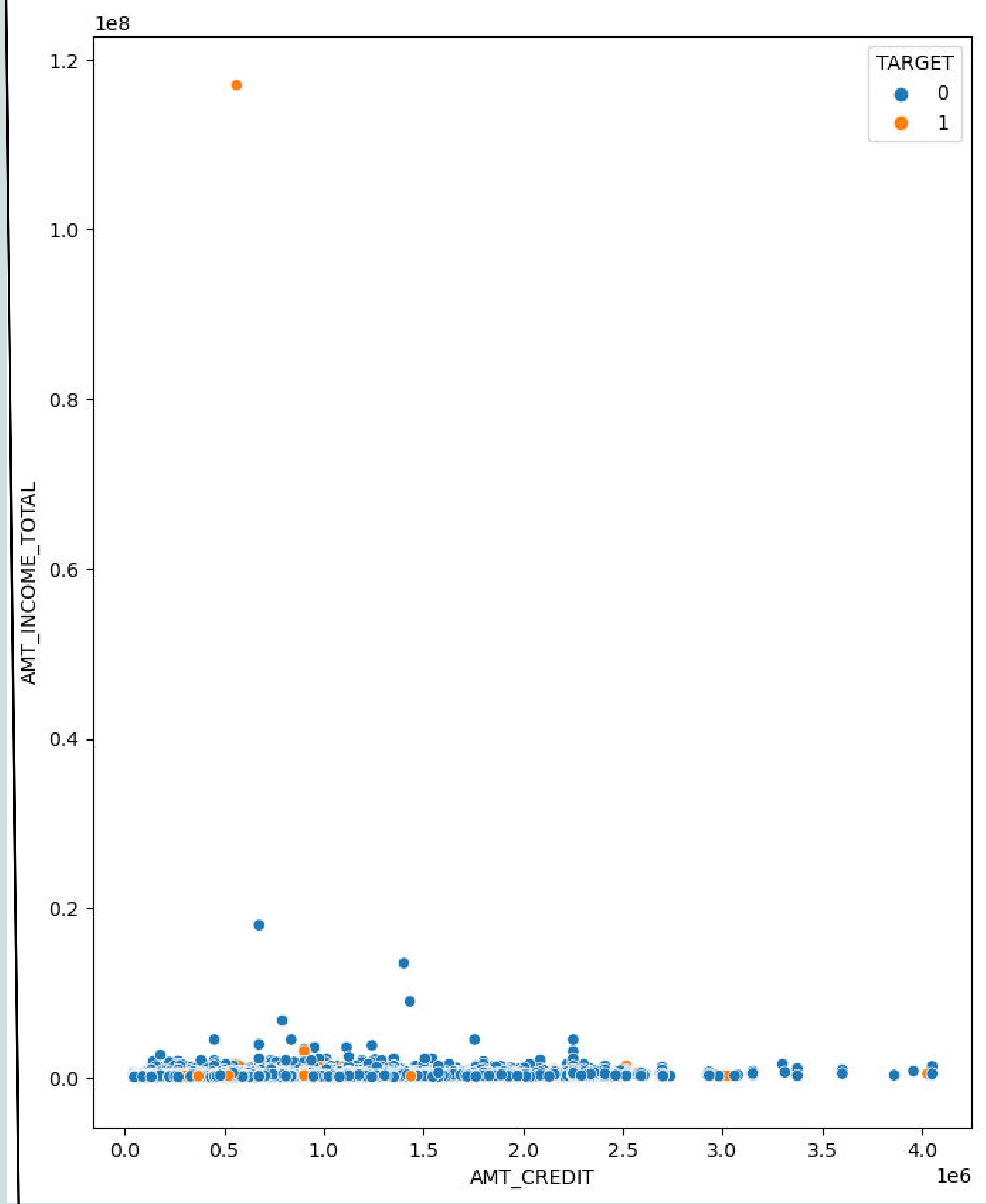


Bivariate Analysis and Multivariate analysis

- the plot is a multivarient plot because it consists of three features income,credited amount and target

- from the plot we can say that customer who have taken loan in between 1.5 million and 2.5 million with income less than 1 million are the safest segment

- where as customers with less than 1 million salary and credit amount in between 0.5 to 1.5 million are more likely to default

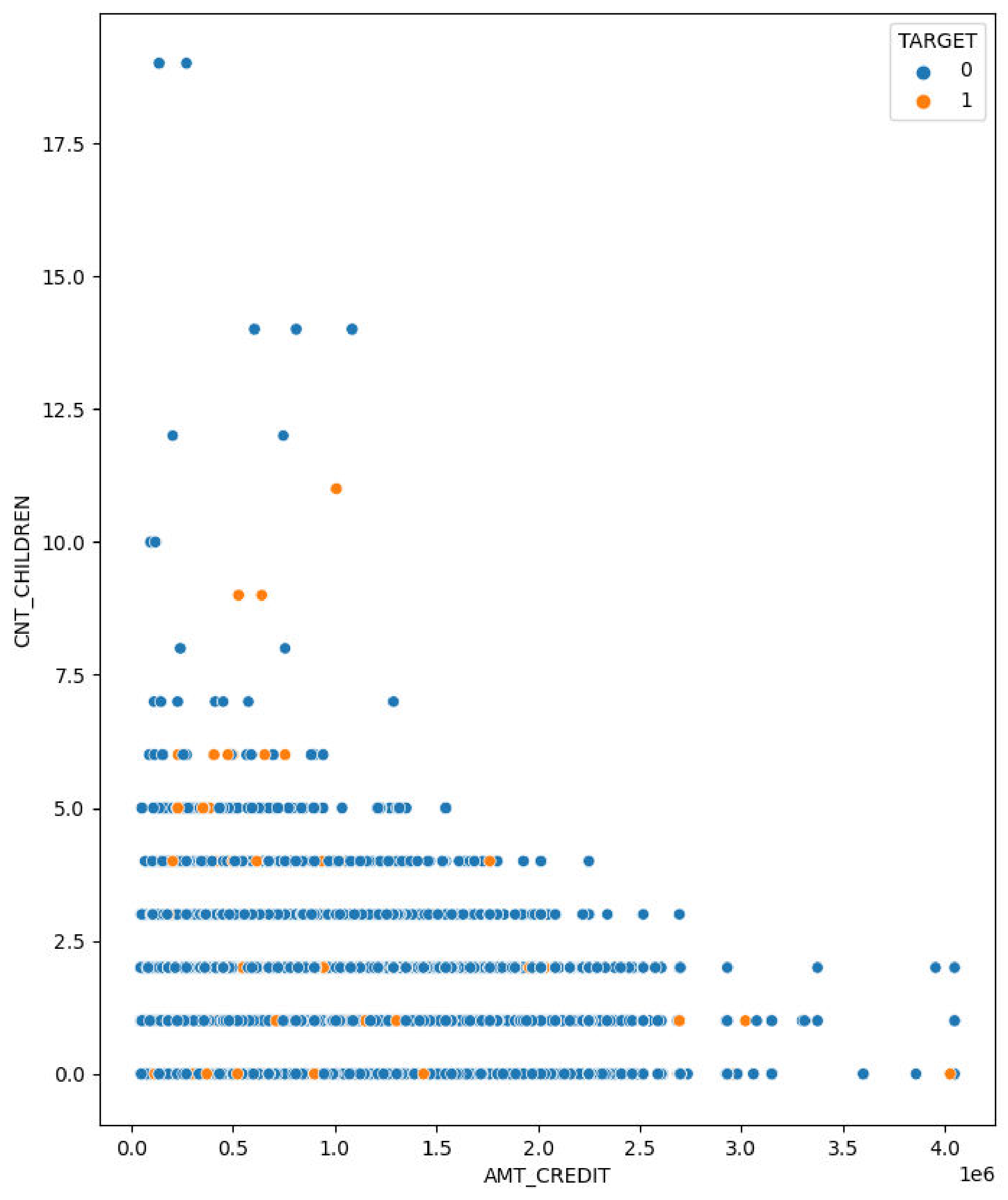


count children vs credit amount vs target

- the graph shows the relationship between amount credit , children count and target

- from the plot we can see that people who have children more than 5 and taken a loan less than 1.5 are more likely to turn default

- where the people who have less than 5 children and taken a credit in between 1.5 and 2.5 are the more like to repay the loan



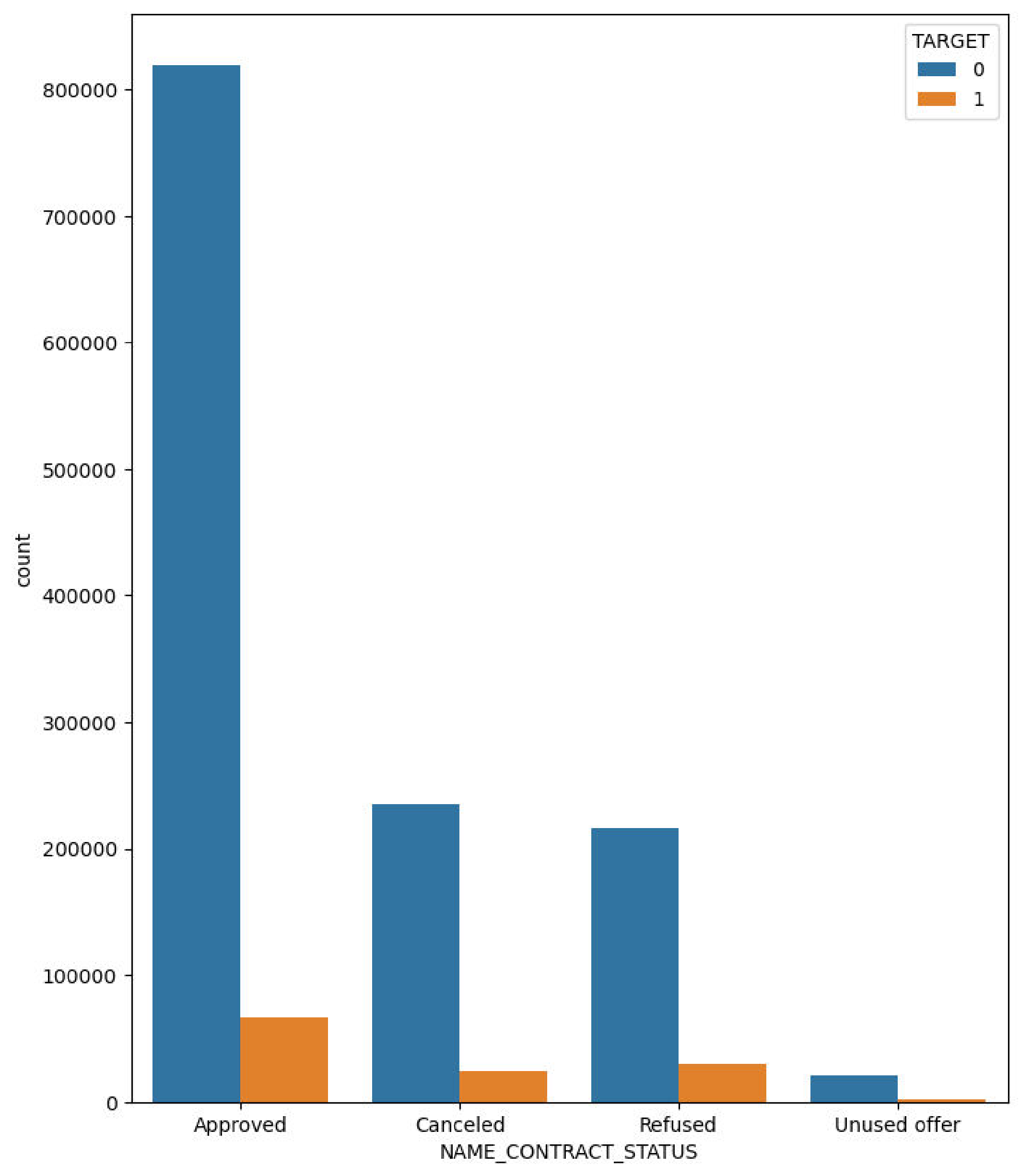
- the plot shows the relation between name contract statuts and target

- they are three types of name contract statuts approved , canceled, refused and un-used

- from the plot we can see that most of the loans are previously approved

- in approved the non defaulters are about 800k and defaulters are less than 100k

- to conclud which is the safest segment we need to check the next plot which is the rate of default



- the plot is showing the name contract status default rate

- in the above graph we saw that people who’s loan previously got approved receved most loans this time also

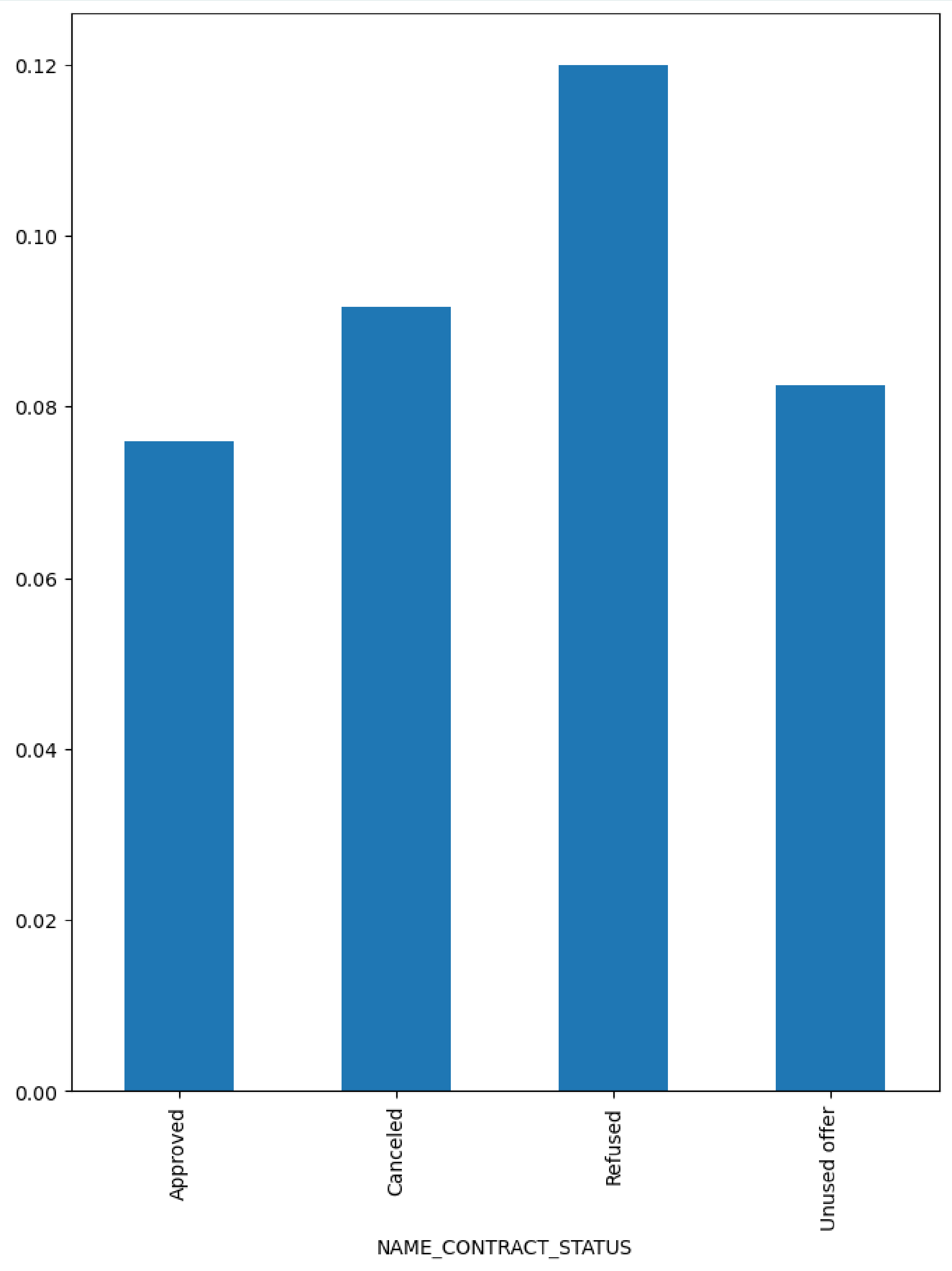
- turns out that was a good decision because desipiteb being the large distribution to the approved loans it still has the lowest default rate

- where as peoples who’s loan previously got reject and this time got loan have more number of default

- where we cannot make same comment to the unused as it’s data distribution it self is very low

- so now we can conclude that people who’s loan previously got rejected are more likely to default loans so they are the most risky segment

- where as people who’s loan previously got approved are more likely to repay the loan .



conclusion

- bank should focus more on cash loans because of its low default arrount 8% dispite being the given the most time
- males are more likely to default than females
- bank should focus on unaccompanied because they have low default ratio
- people with married family satatus are the safest
- people who are having income less than 1 million and taking loan in between 1.5 to 2.5 million are safest
- peole who are taking loan between 1.5 to 2.5 million and having chindren less than 5 are among the safest segment
- people with already approved loans are safest and people with previously refused loans are more risky

THANK YOU