CREDIT DEFAULTER ANALYSIS

By
Bibekananda sahoo

INTRODUCTION

OBJECTIVE

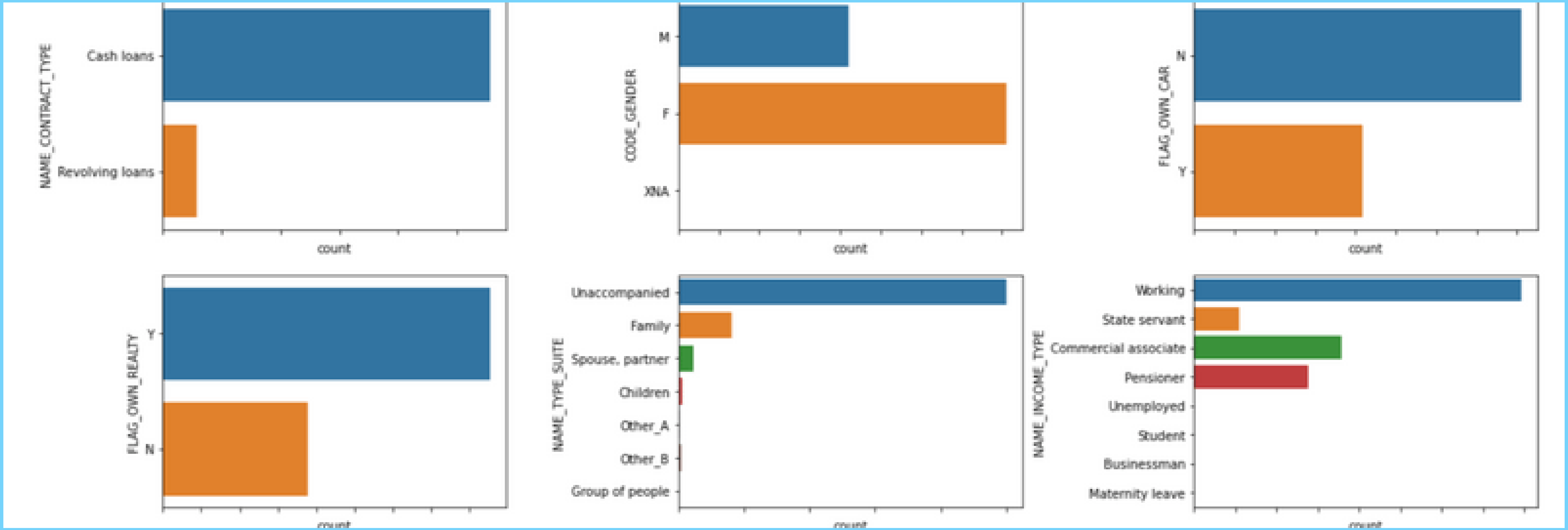
The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose we work for a consumer finance company which specialises in lending various types of loans to urban customers. We have to use EDA to analyse the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

Data Description

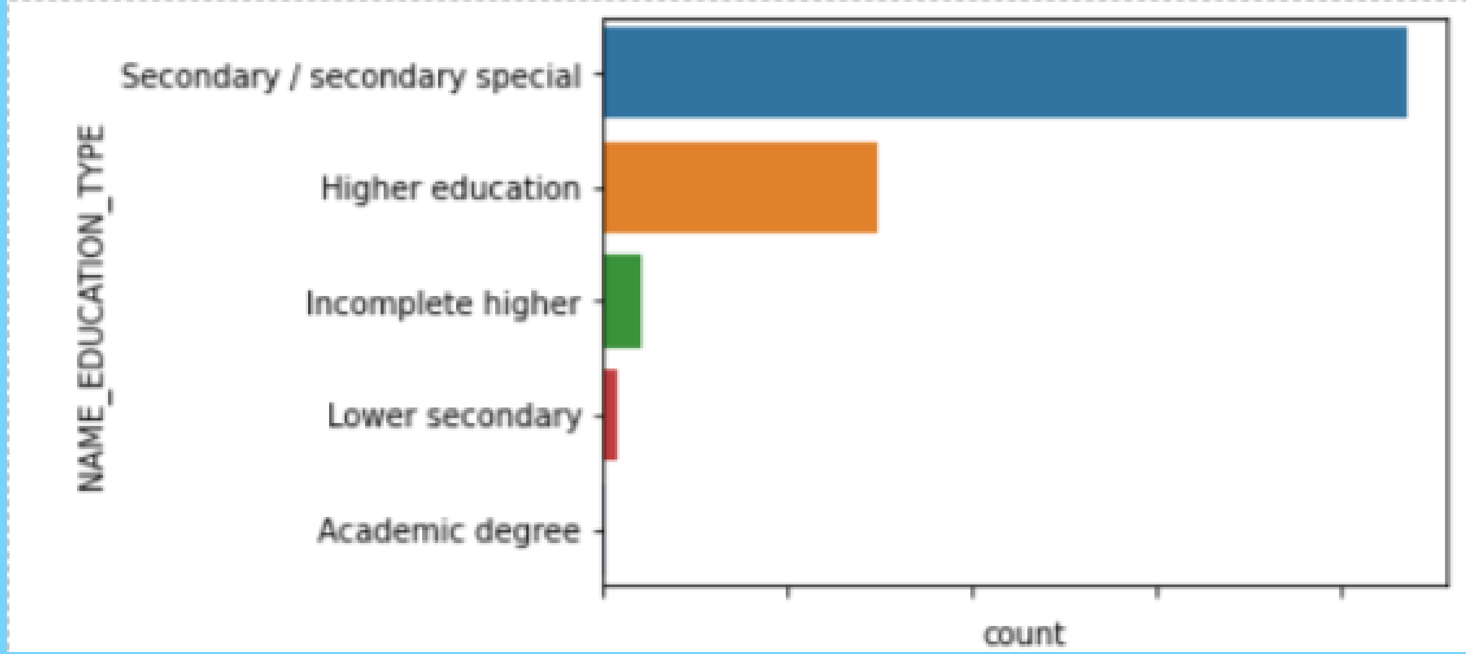
This dataset has 3 files as explained below:

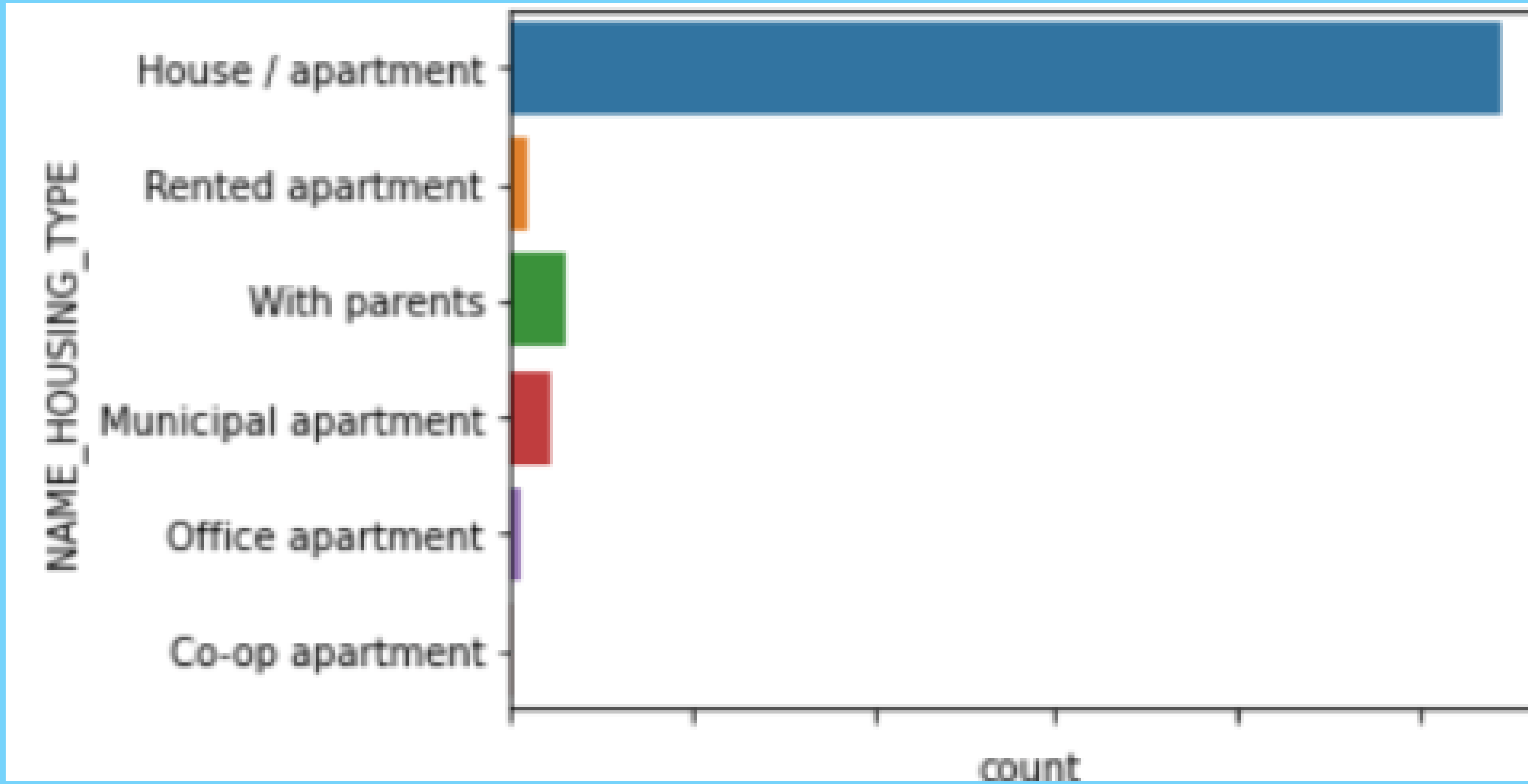
1. *'application_data.csv'* contains all the information of the client at the time of application.
The data is about whether a client has payment difficulties.
2. *'previous_application.csv'* contains information about the client's previous loan data. It contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.
3. *'columns_description.csv'* is data dictionary which describes the meaning of the variables.

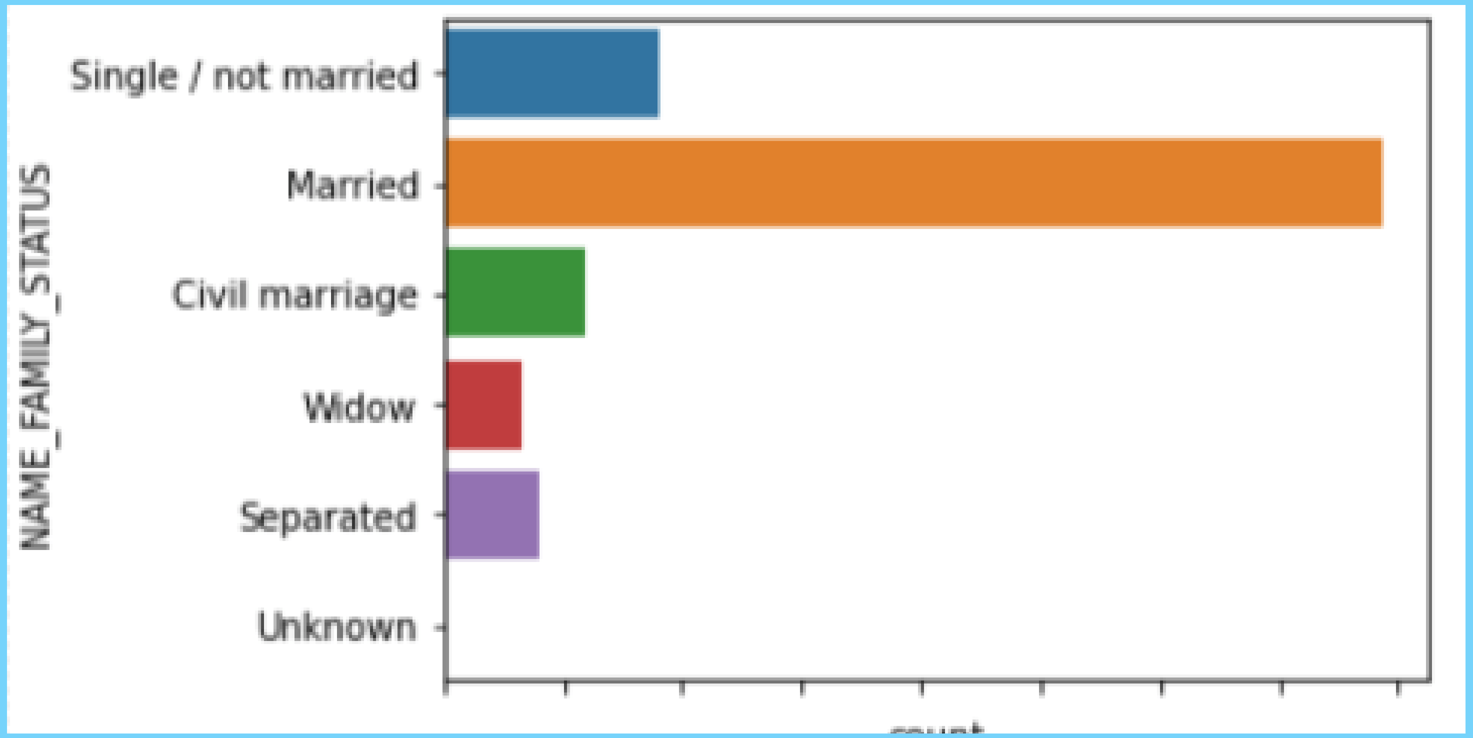
Categorical columns analysis (application_data)



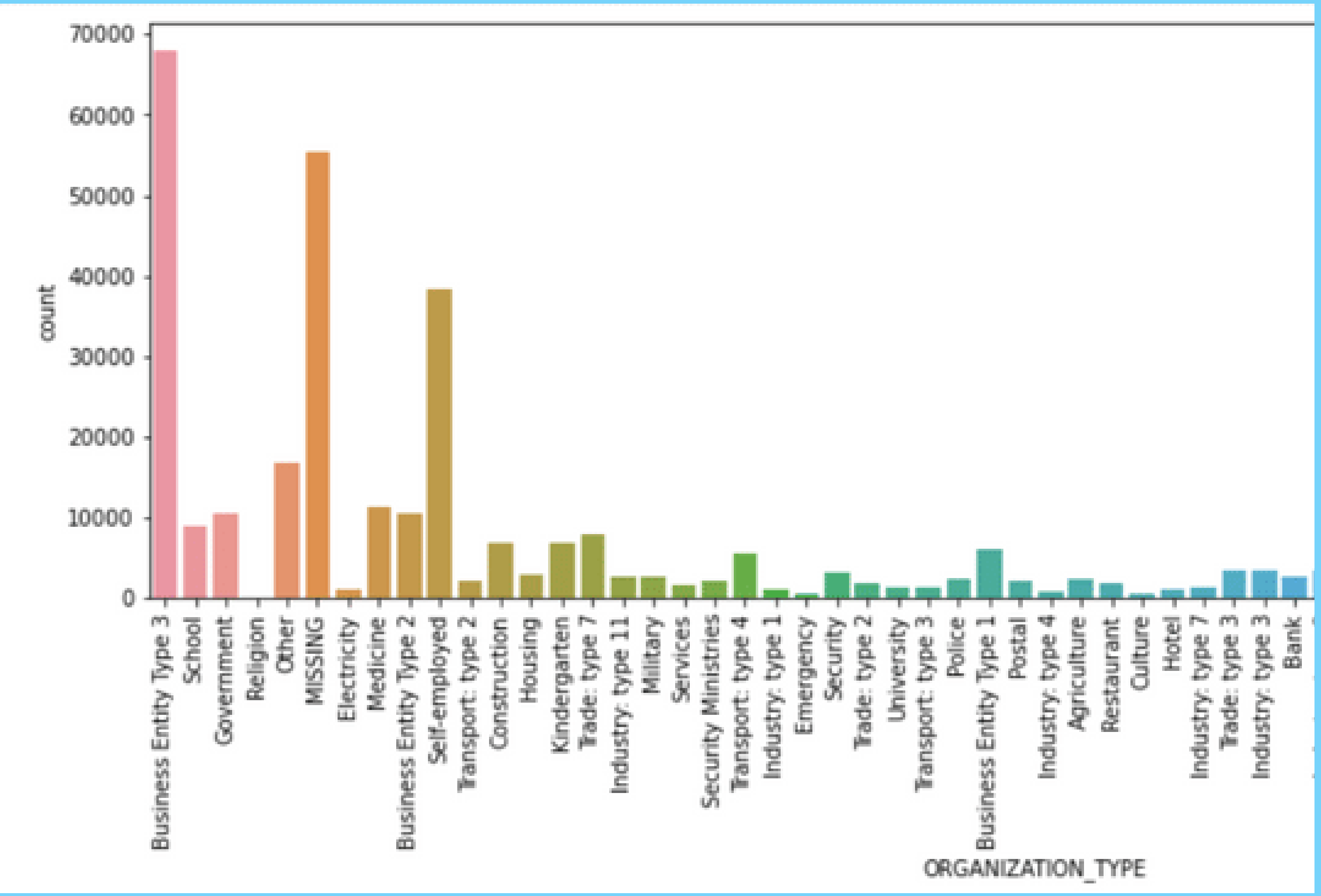
- 'cash loans' contract type is more then 'revolving loans'
- females are taking more loans then males
- There are more clients who don't have cars
- There are more clients having house or flat
- 'NAME_TYPE_SUITE' is mostly 'Unaccompanied' means clients applying the loan on there own
- The income type is mostly "working' and 'commercial associate'

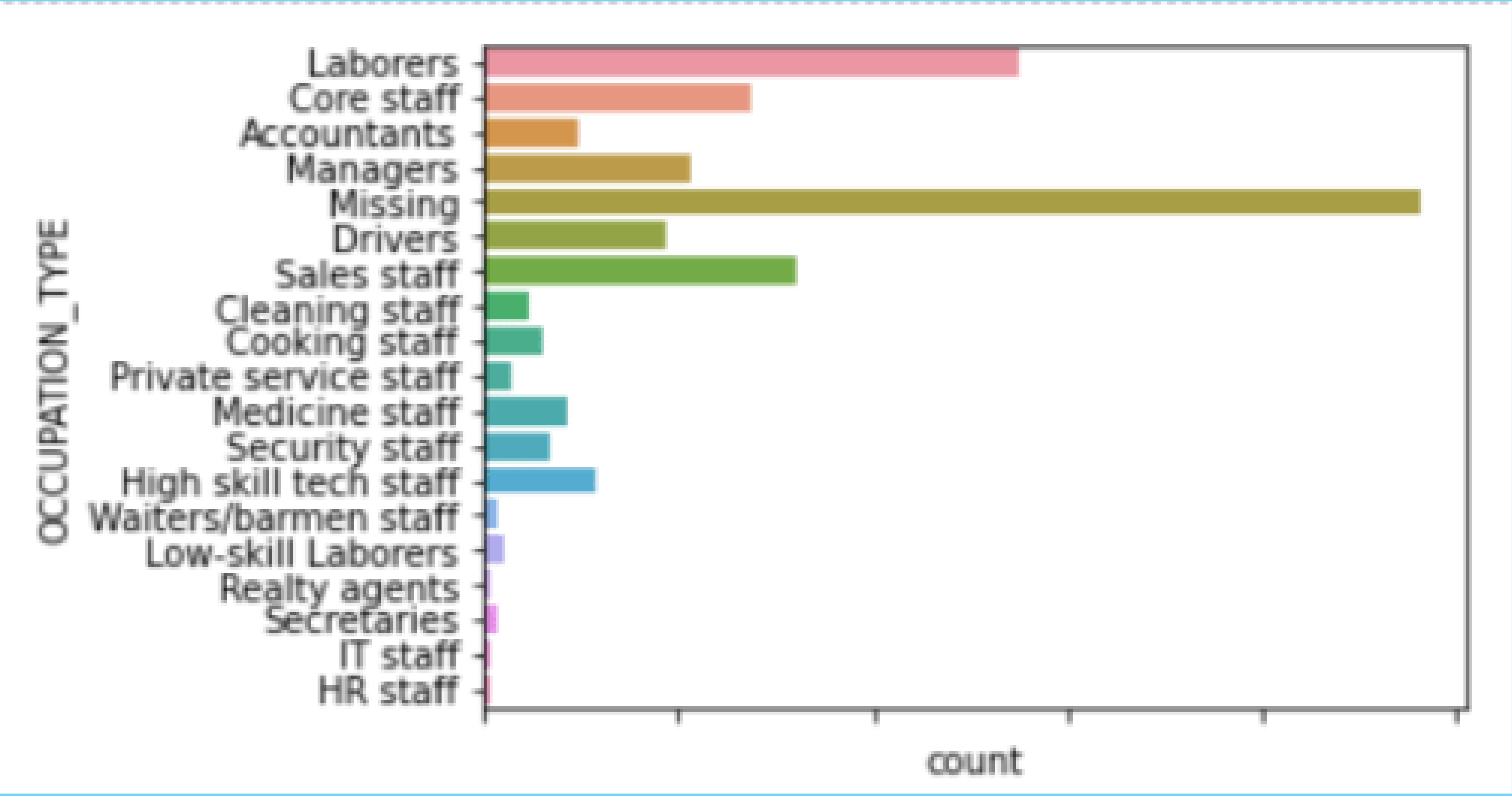






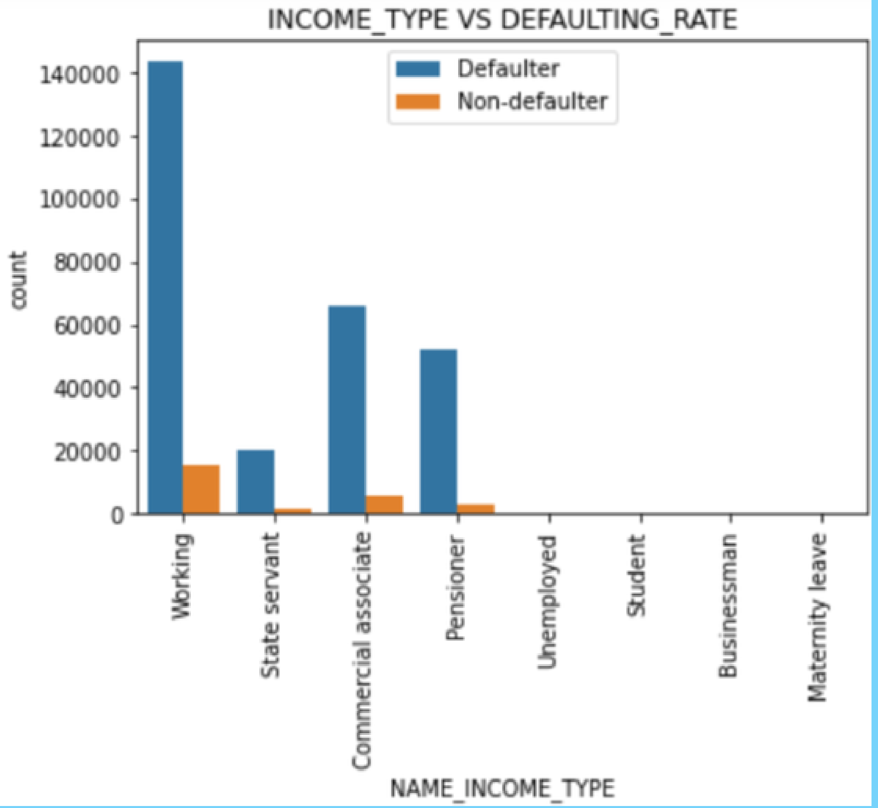
1. Education type is mostly 'secondary/secondary special' of the clients
2. most of the clients living on there apartment/house
3. There are more 'married' clints

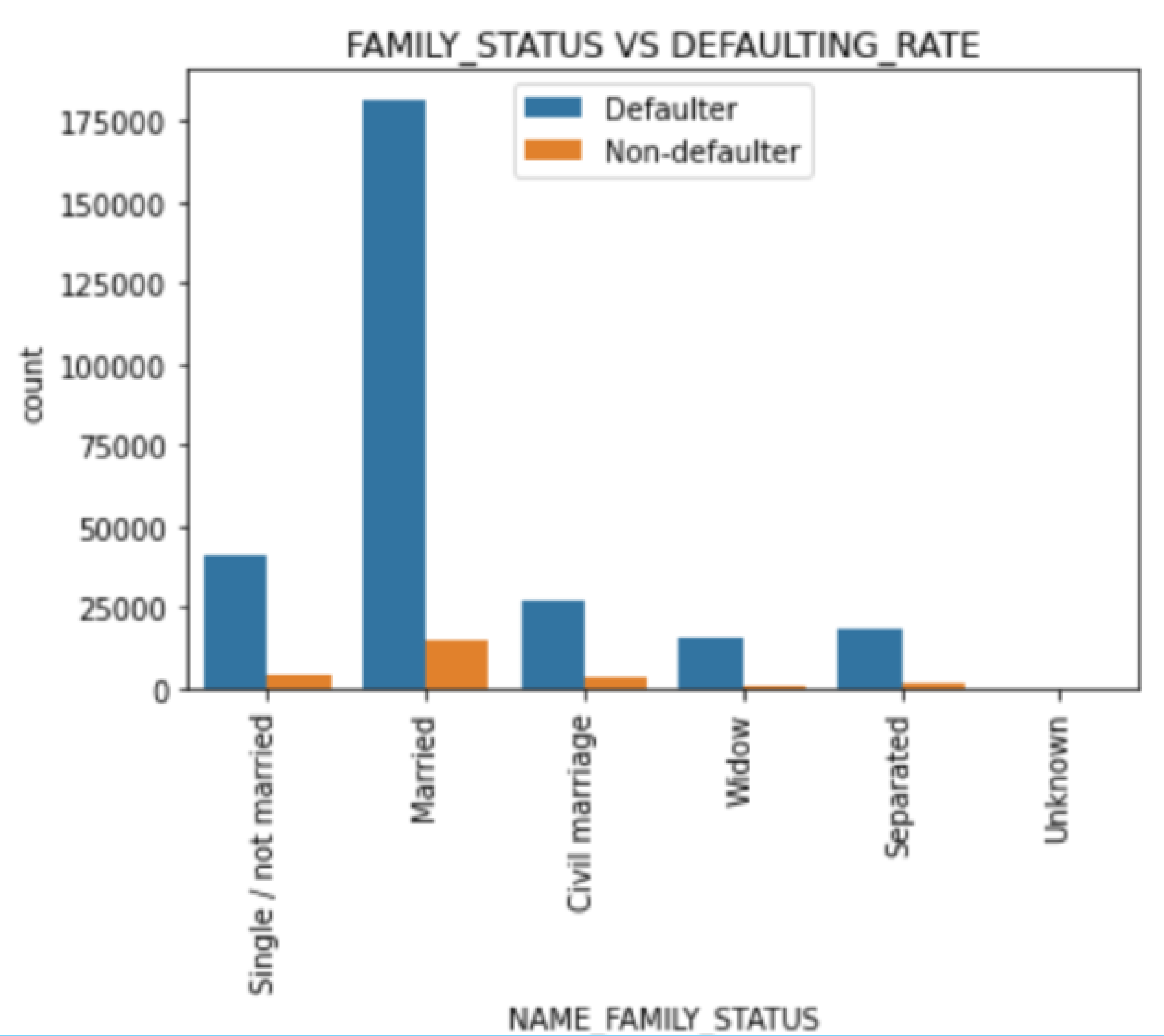




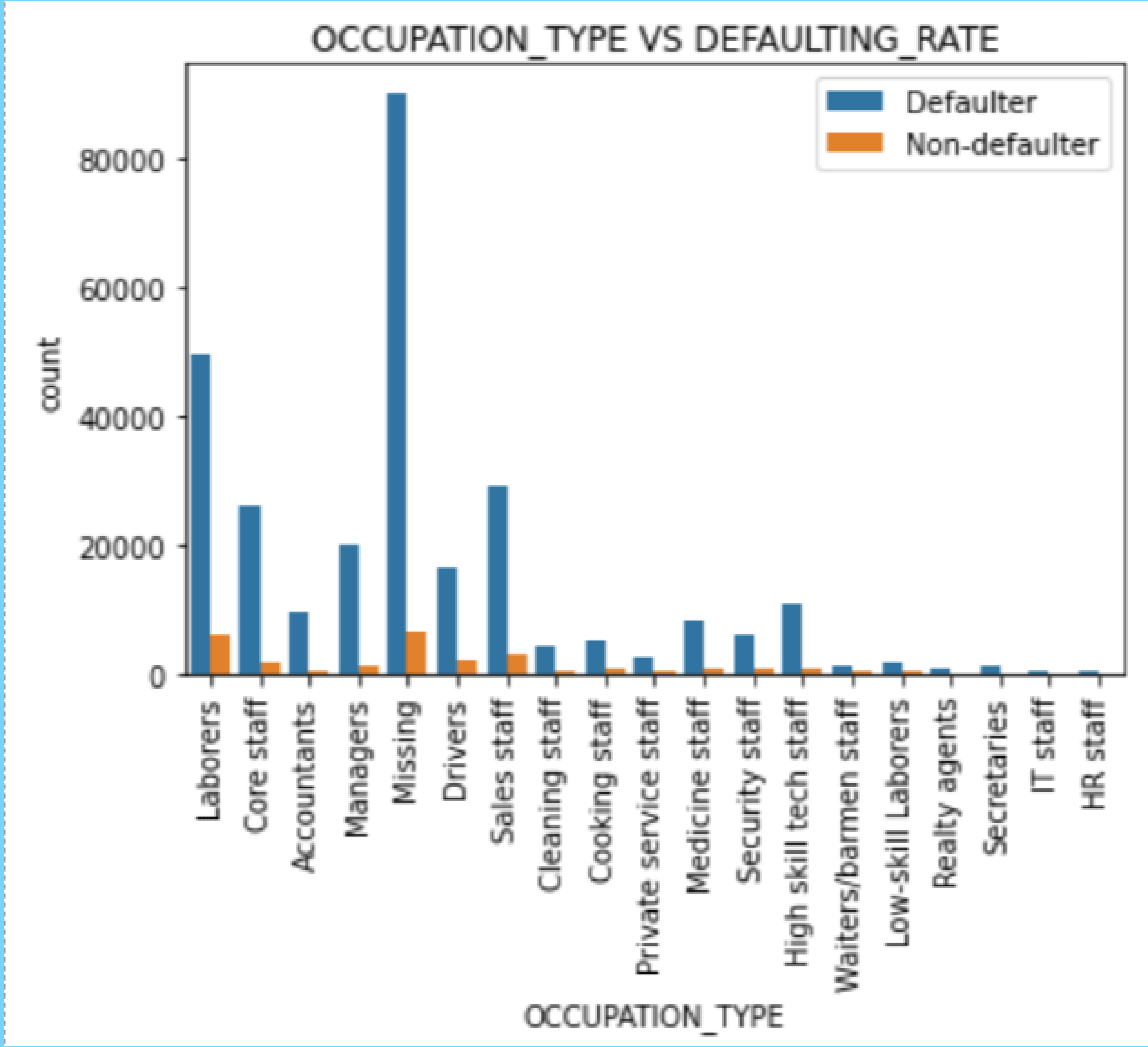
- ORGANIZATION_TYPE of most clients was 'Business Entity Type 3' followed by 'Self-employed' ,'Other','Medicine'
- most occupation type of clients are 'Laborers' followed by 'sales staff','core staff','managers'

Defaulters analysis in application_data

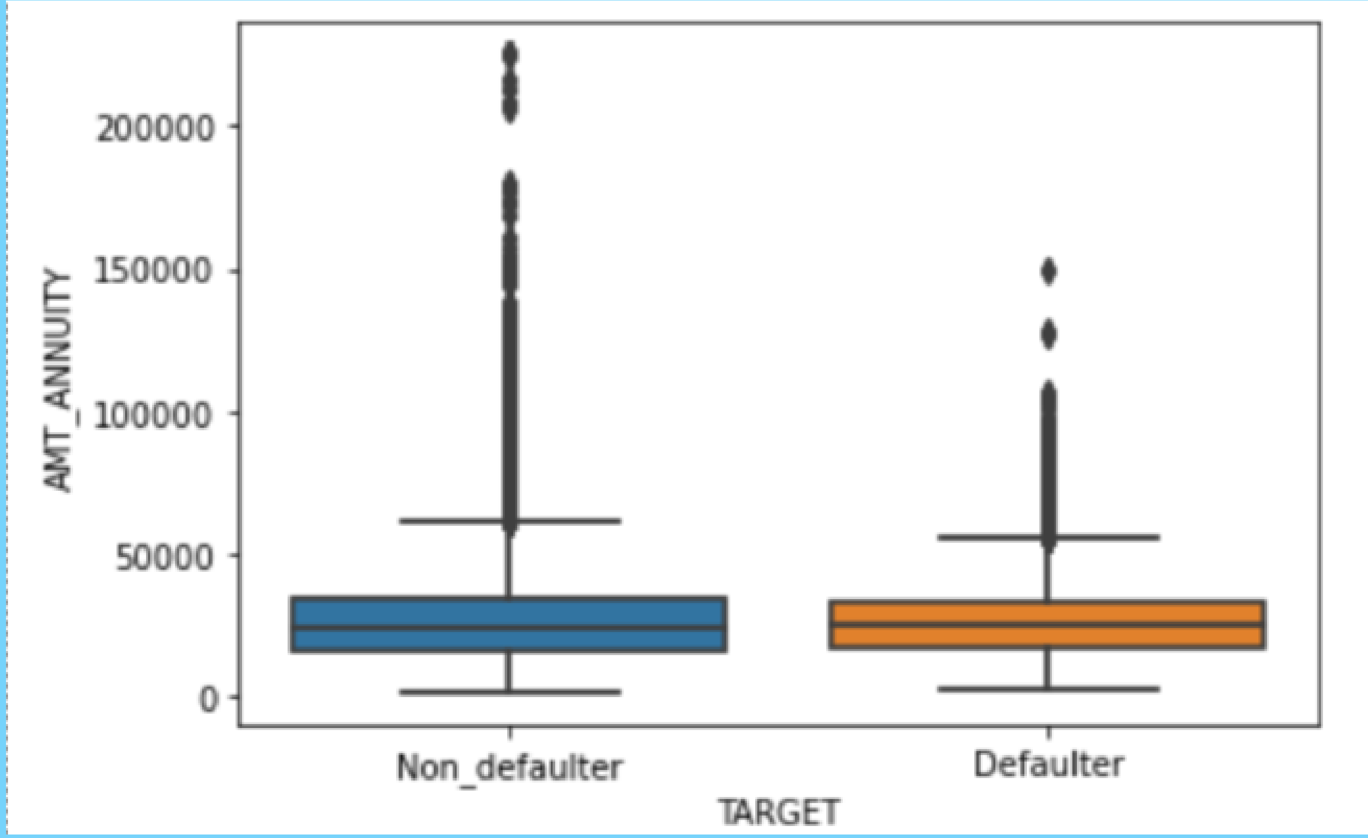


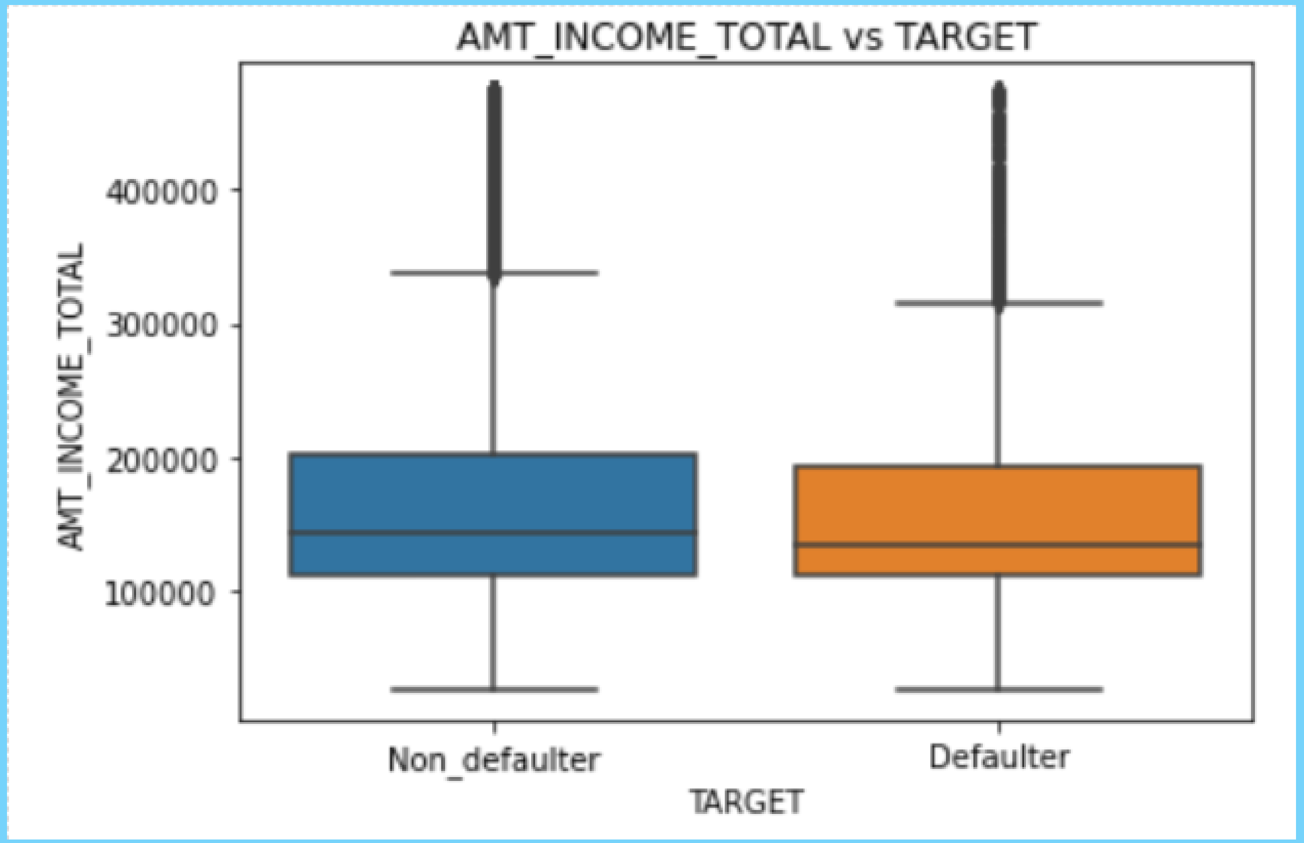


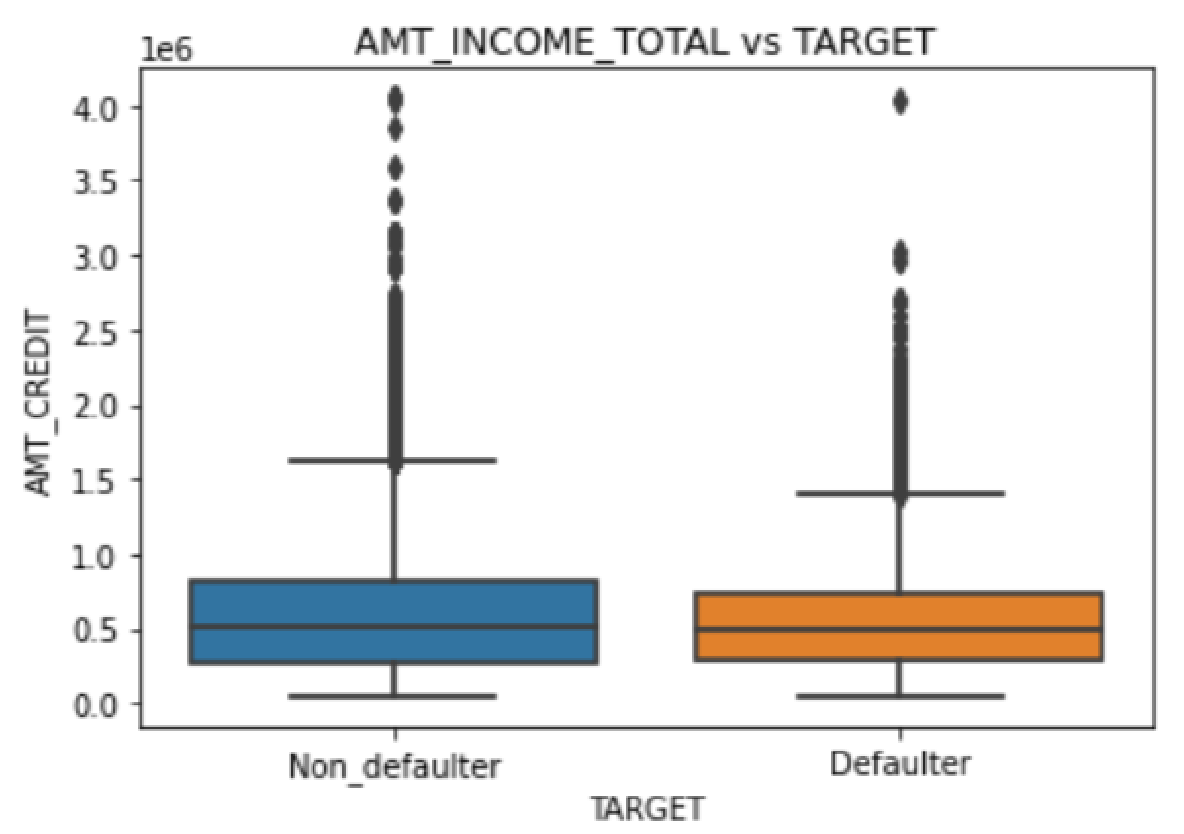
- 'single/not married' and 'civil marriage' clients are defaulting more
- 'married clients' are more intrested to take loan and also there defaulting percentage was low as compare to 'single/not married' and 'civil' marriage familt_status type
- clients having income type 'working','Commercial associate','Pensioner' are more tend to taking the loan and also there defaulting rate was less then 10%
- but 'unemployed' and 'Maternity leave' income type clients defaulting rate was so high



- defaulting rate of 'Laborers','Drivers','Cooking staff','Security staff','Waiters/barmen staff','Low-skill Laborers' are more then 10%
- on the other hand OCCUPATION_TYPE like 'Accountants','Managers','IT','HR' defaulting rates are less then 7%



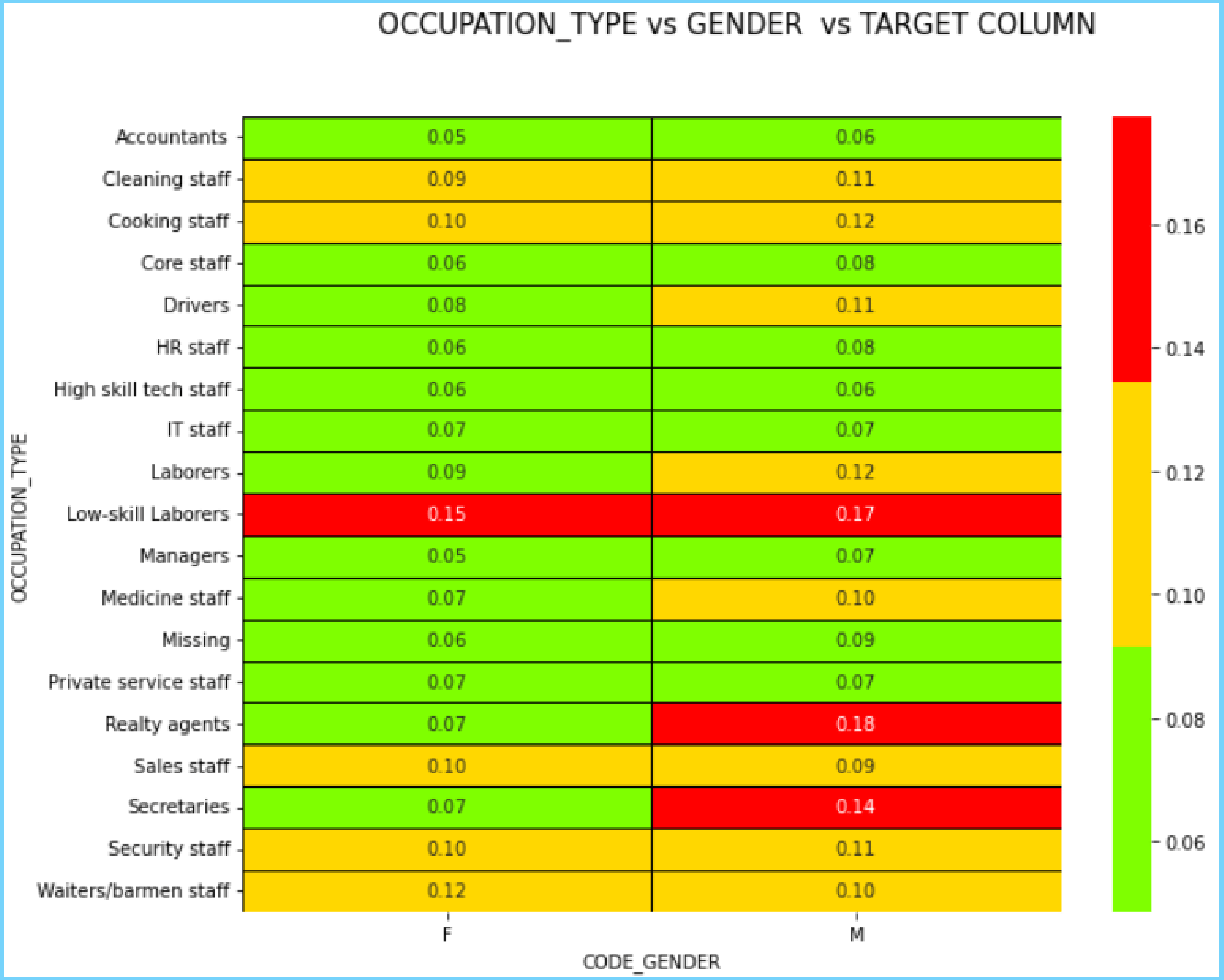




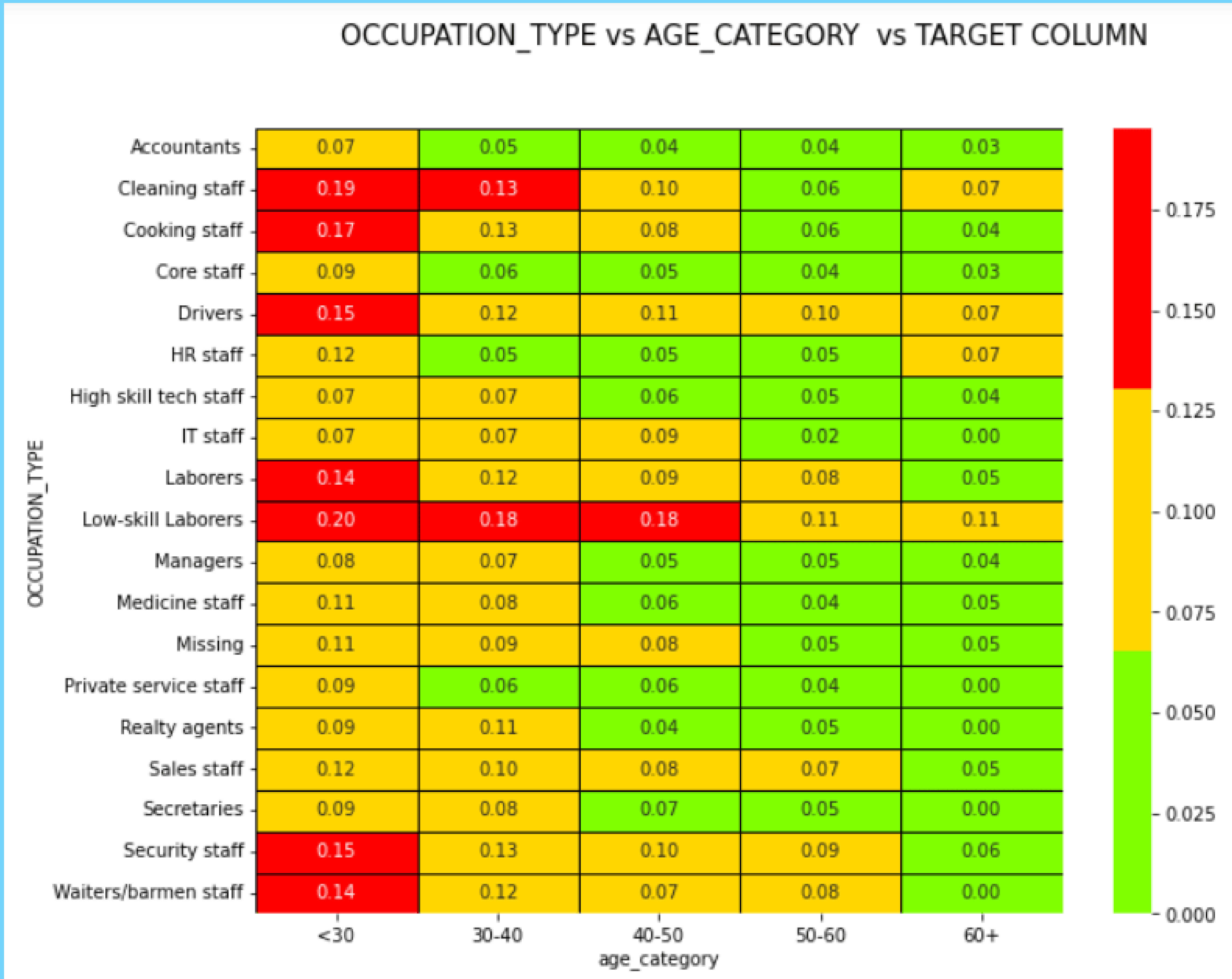
- mean of Amount annuity of defaulters and non defaulters are very much close to each other but in non defaulter case some clients having AMT_ANNUITY more then 150k
- mean of AMT_CREDIT of defaulters and non defaulters are very much close to each other
- mean of Total Income of defaulters and non defaulters are very much same

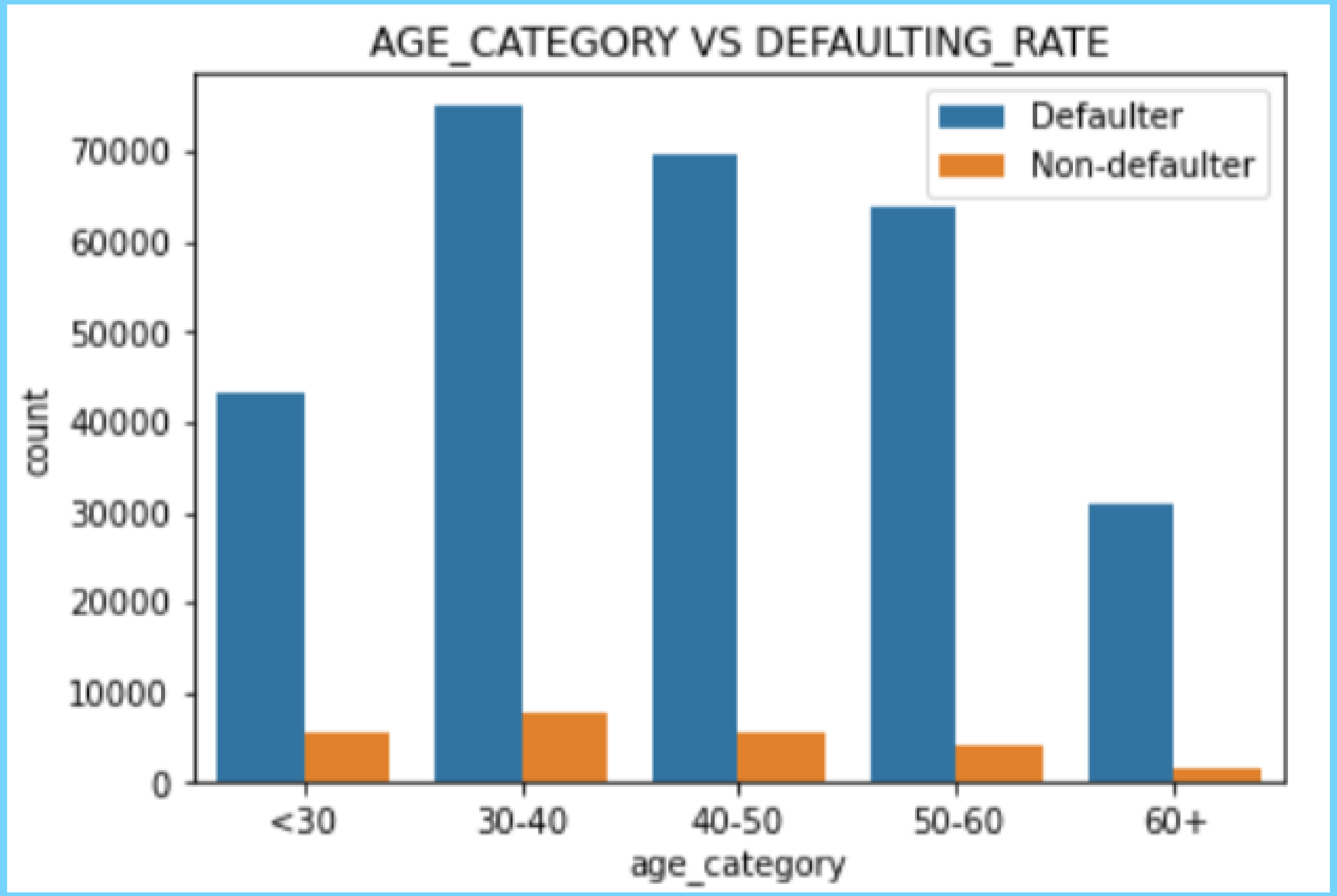
MULTIVARIATE ANALYSIS

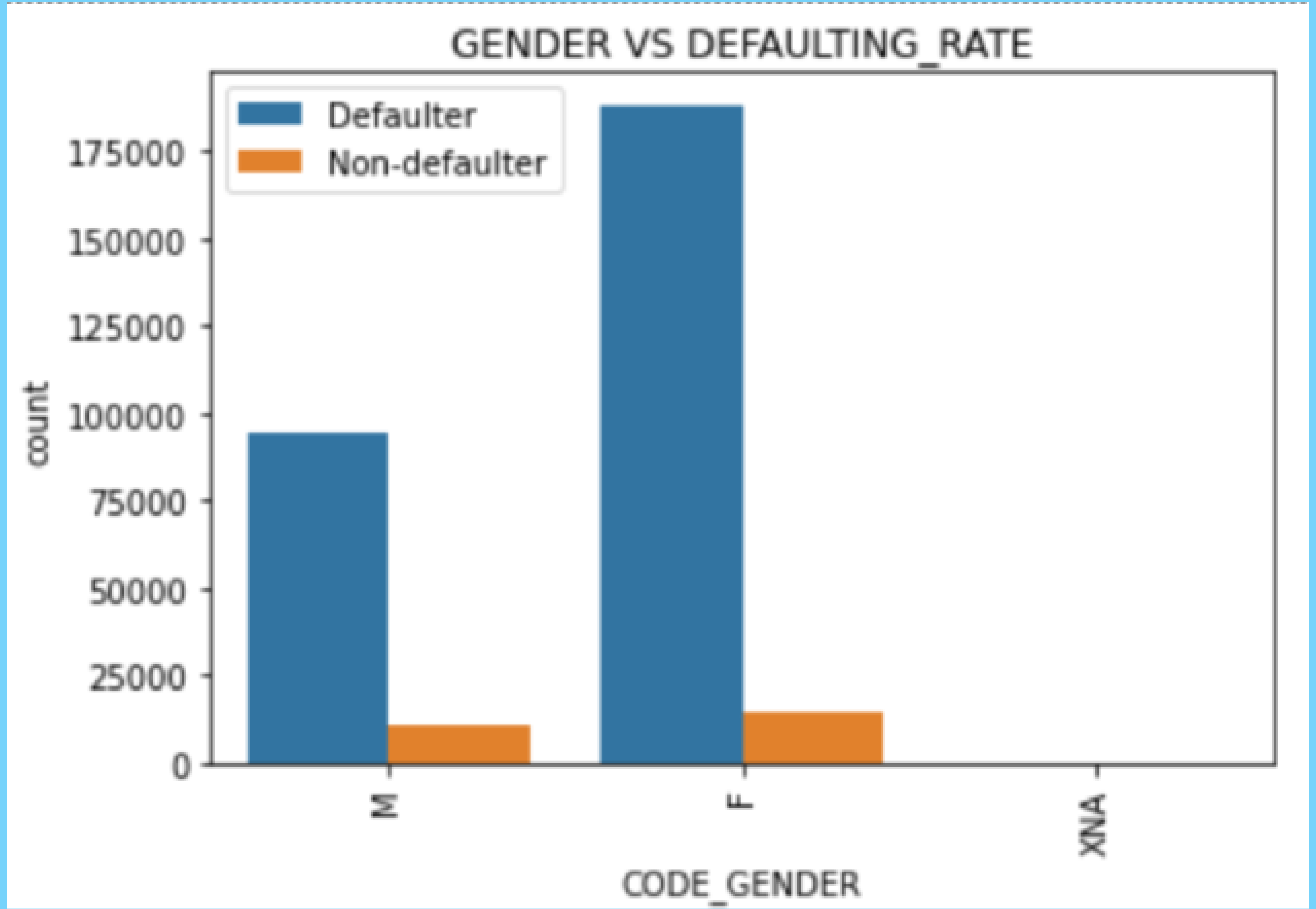
- Males have occupation-type reality agents, low-skill laborers, and secretaries have a high defaulting rate.
- Females having occupation type Low skill laborers having high defaulting rate.



- clients having age less than 30 and have occupations such as Cleaning staff, Cooking staff, Drivers, Low-skill Laborers, Laborers, Security staff, and Waiters/barmen staff having high defaulter rates.
- clients aged between 30 to 40 and have occupations such as Cleaning staff, Low-skill Laborers having high defaulter rates.

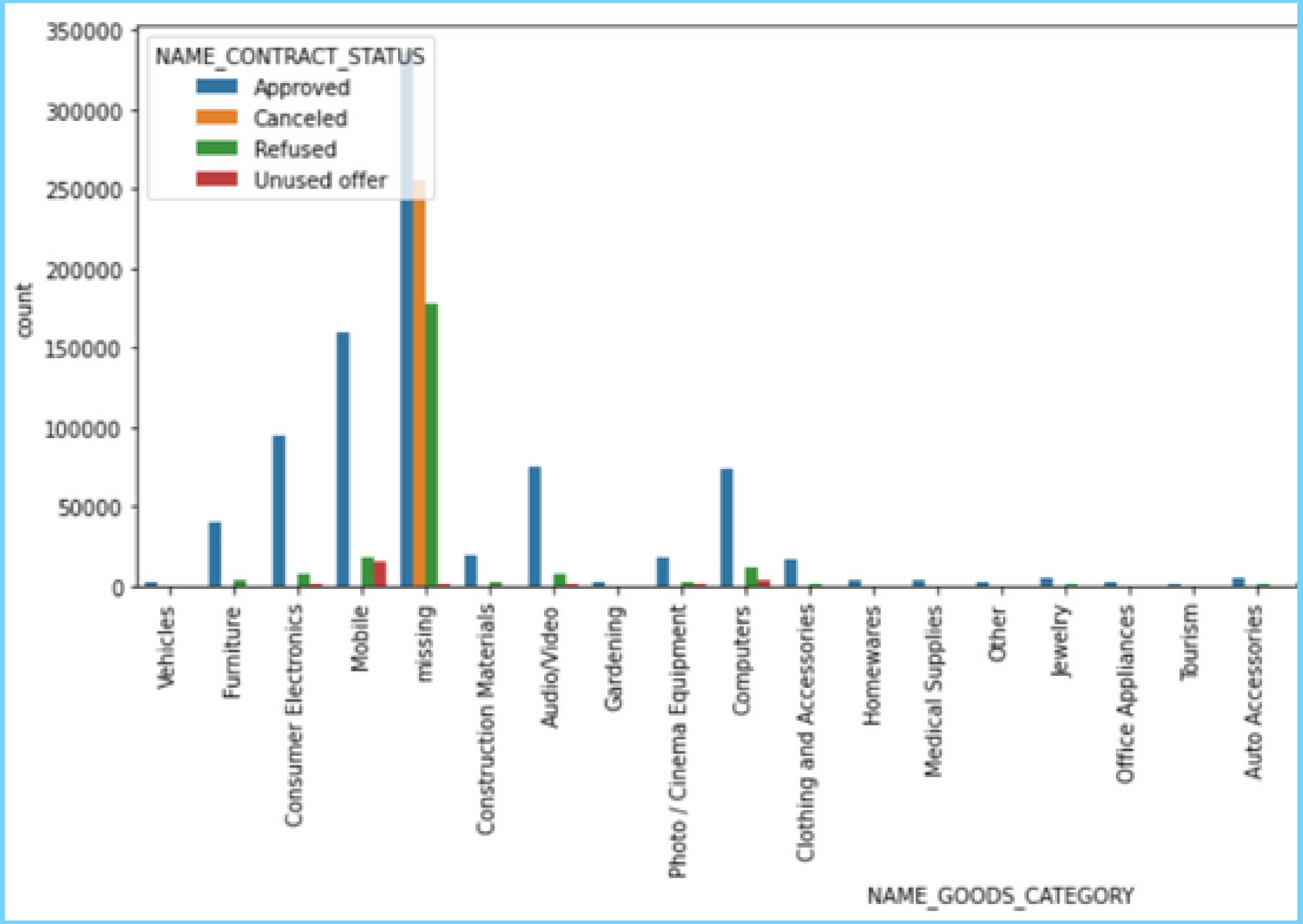


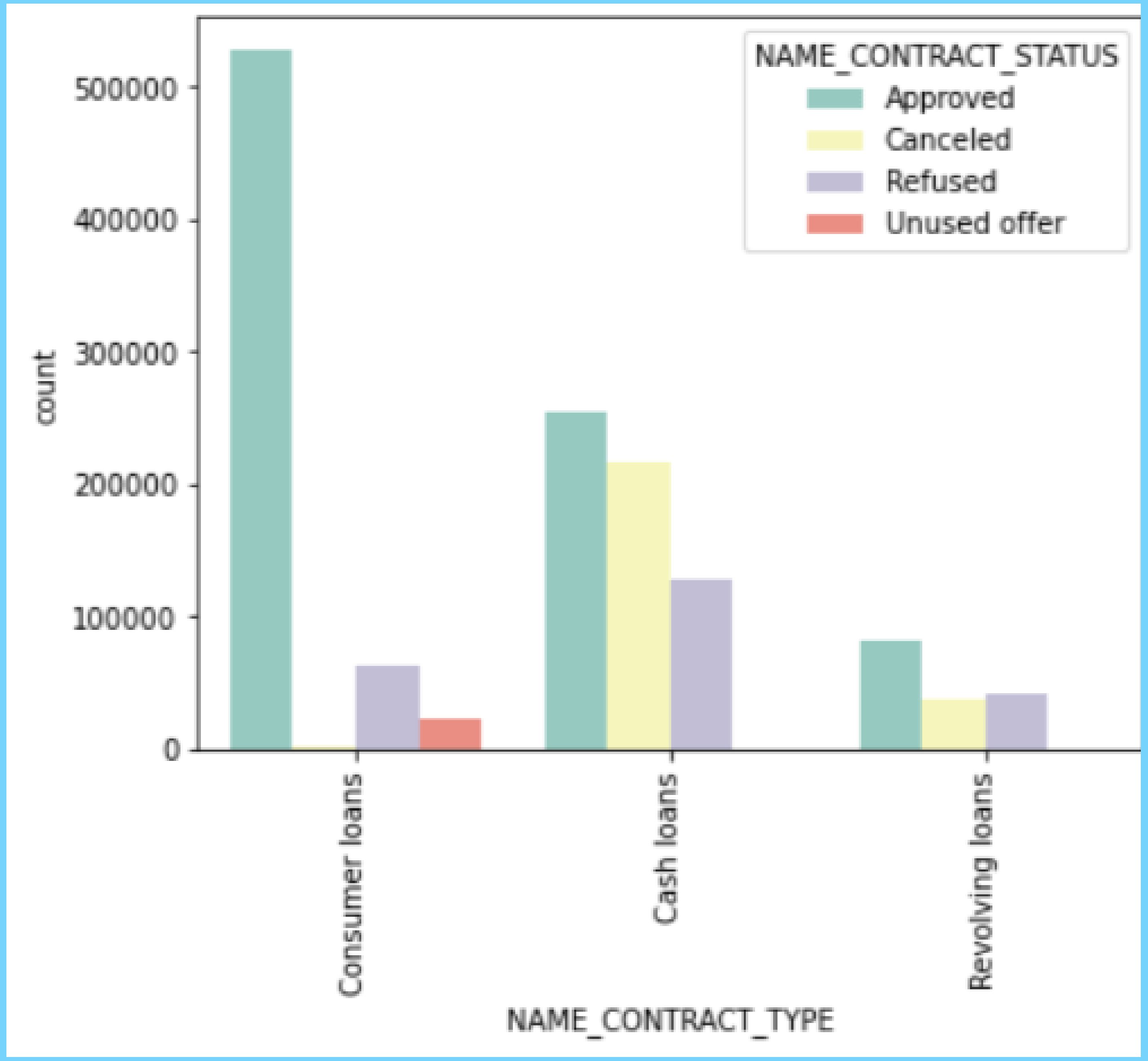




- age less then 30 defaulting more then other age category
- age greater then 60 are less defaulter
- Females are more intrested to taking loans and there overall defaulting rate was less as compare to males
- On the other hand males are defaulting more then females

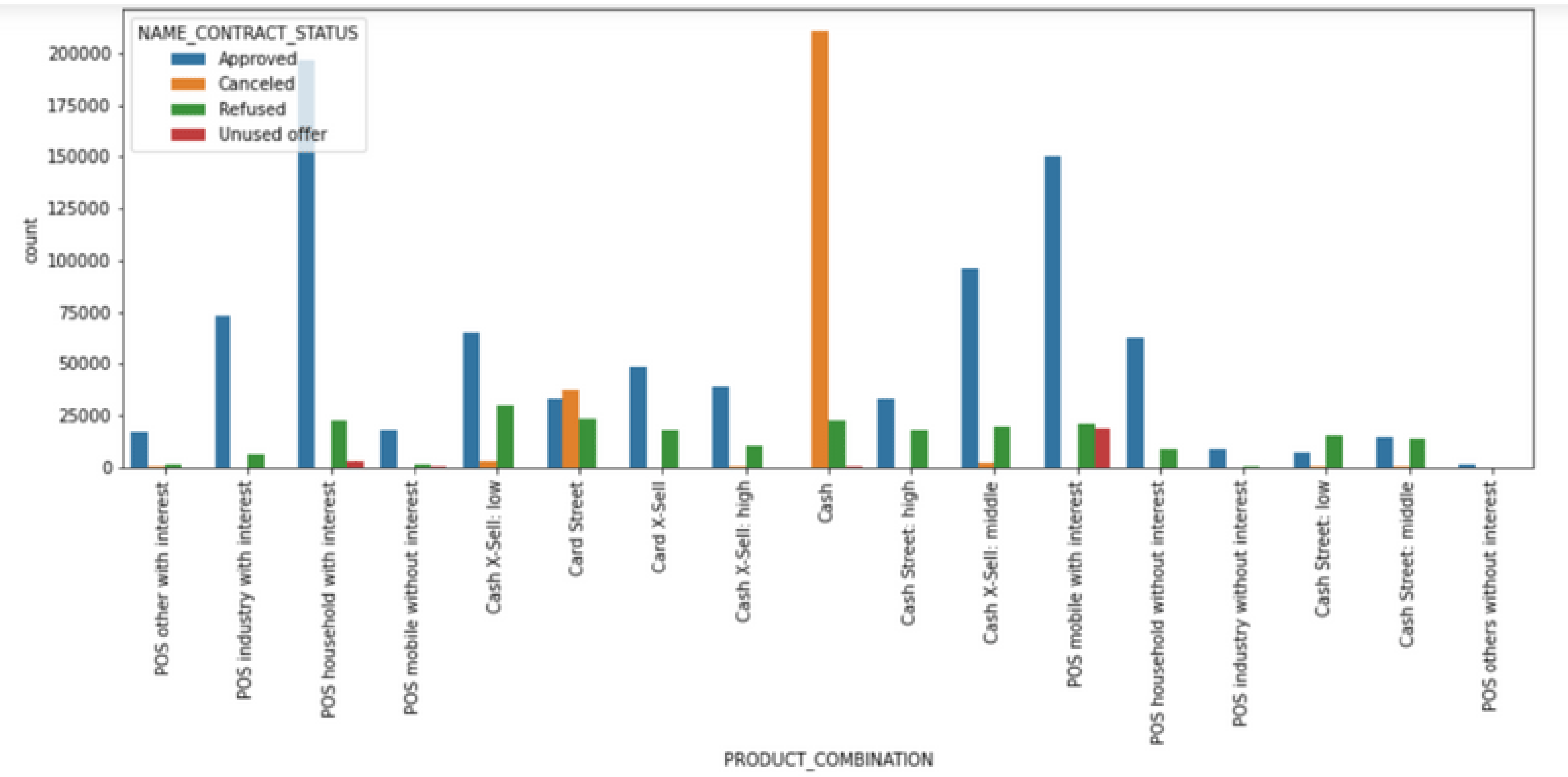
Categorical columns analysis (previous_application_data)



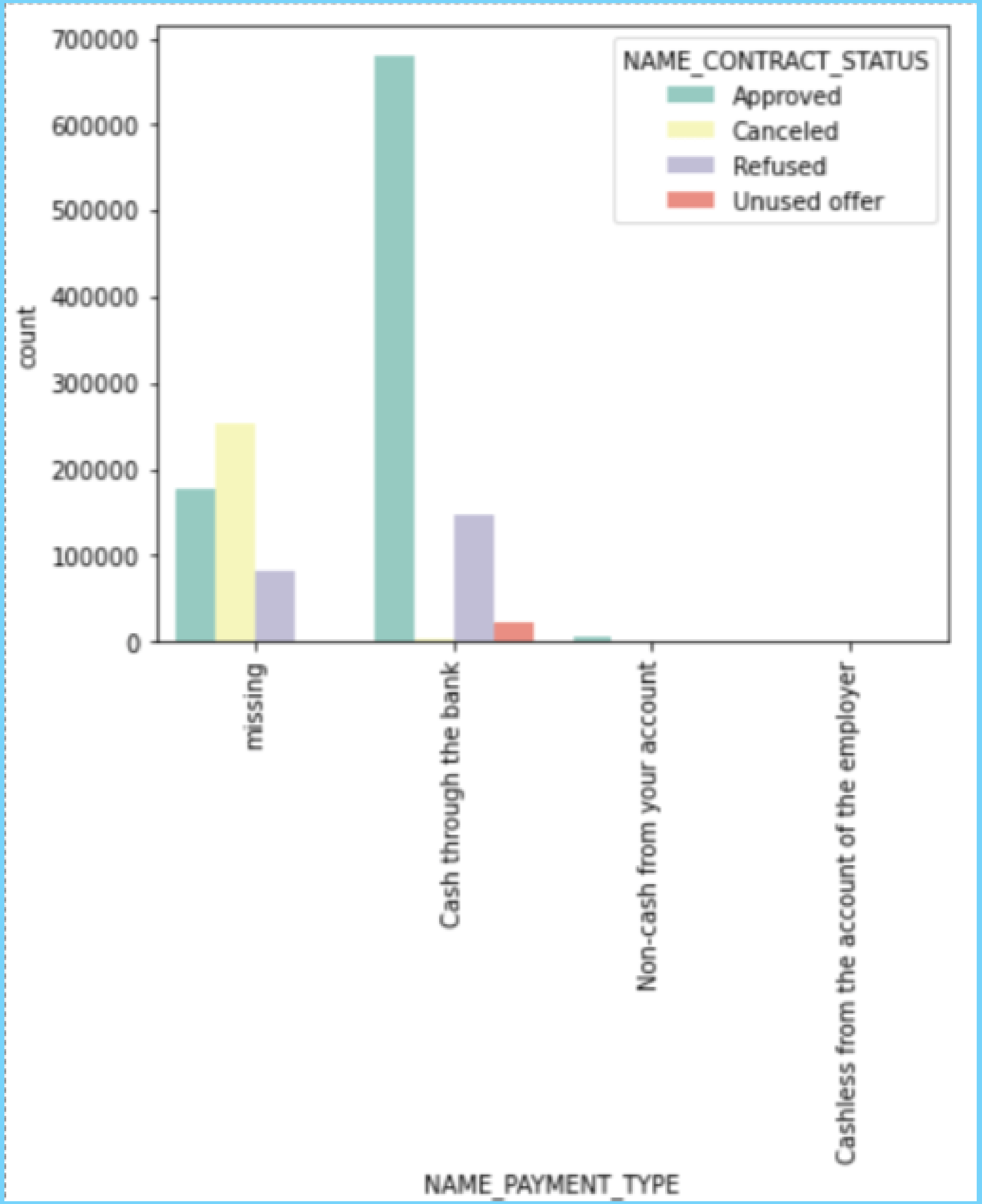


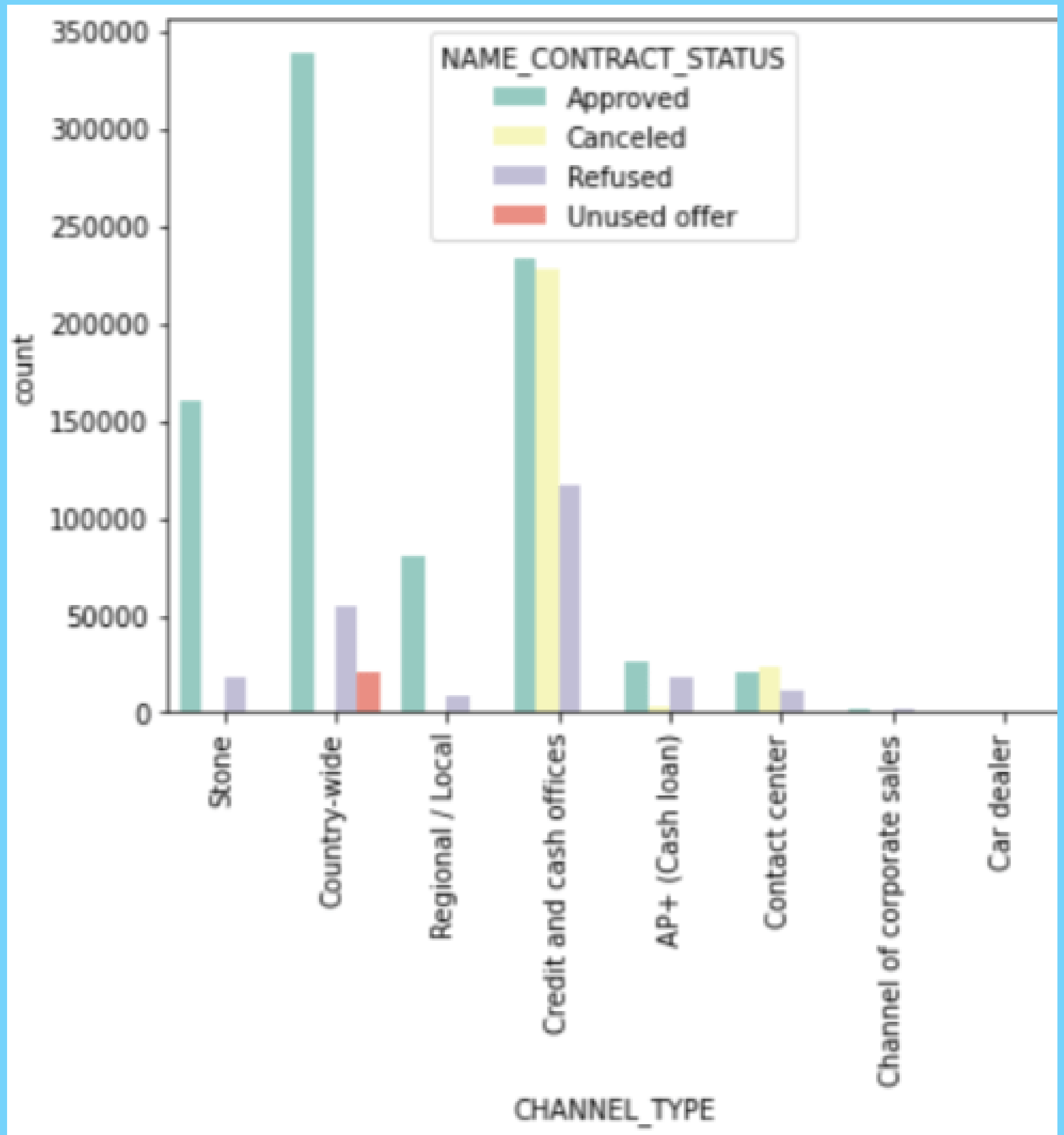
- privious_clients whose kind of goods is 'mobile','electronics' there approving rate is more
- most privious_clients are taking 'consumers loans' CONTRACT_TYPE also there approving rate was high

PRODUCT_COMBINATION vs NAME_CONTRACT_STATUS



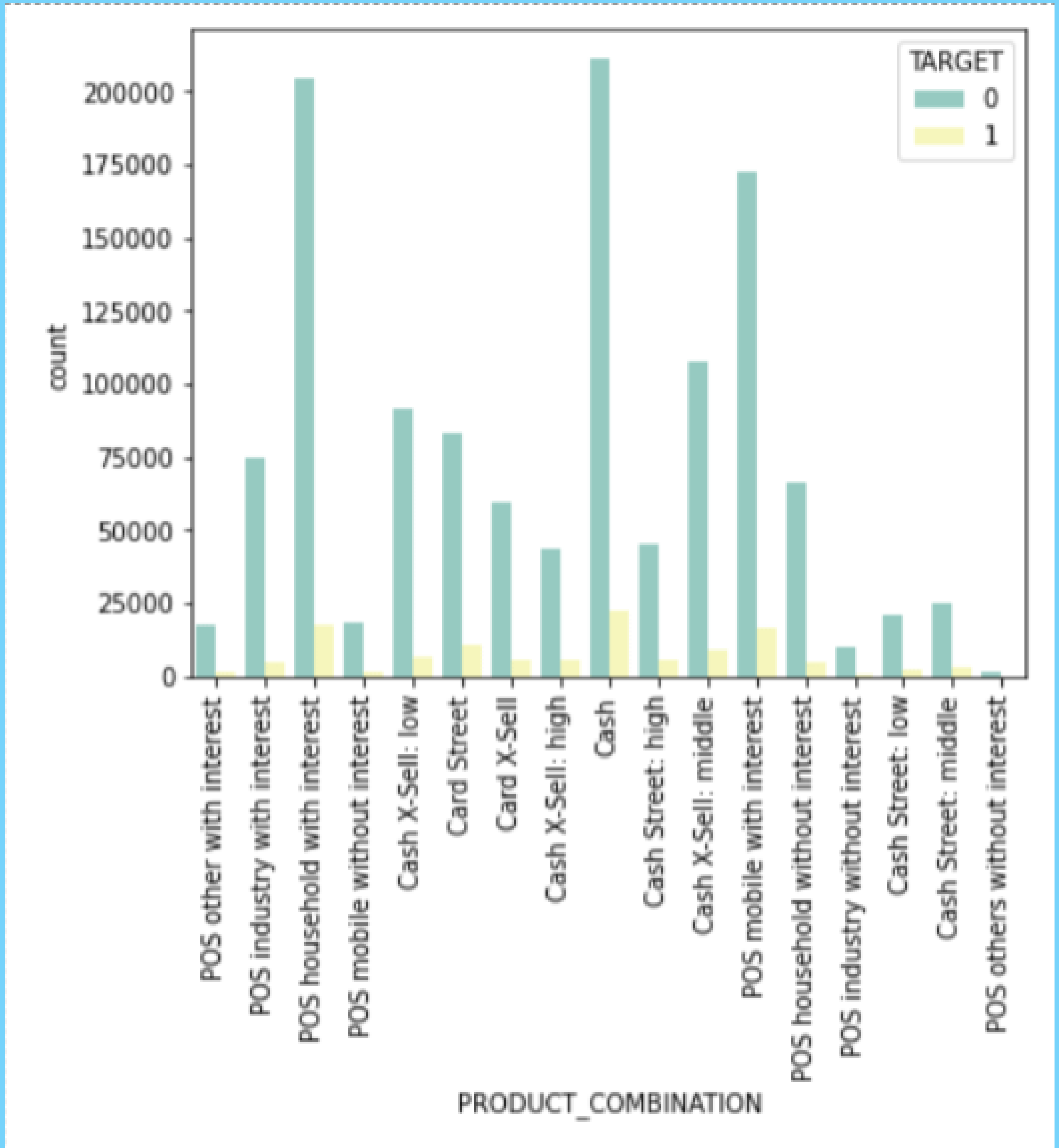
- Those privious_clients whose product_combination is 'Cash' there cancel rate was high as comapre to approving rate
- product_combination with 'POS household with interest' and 'POS mobile with interest' are high approving rate



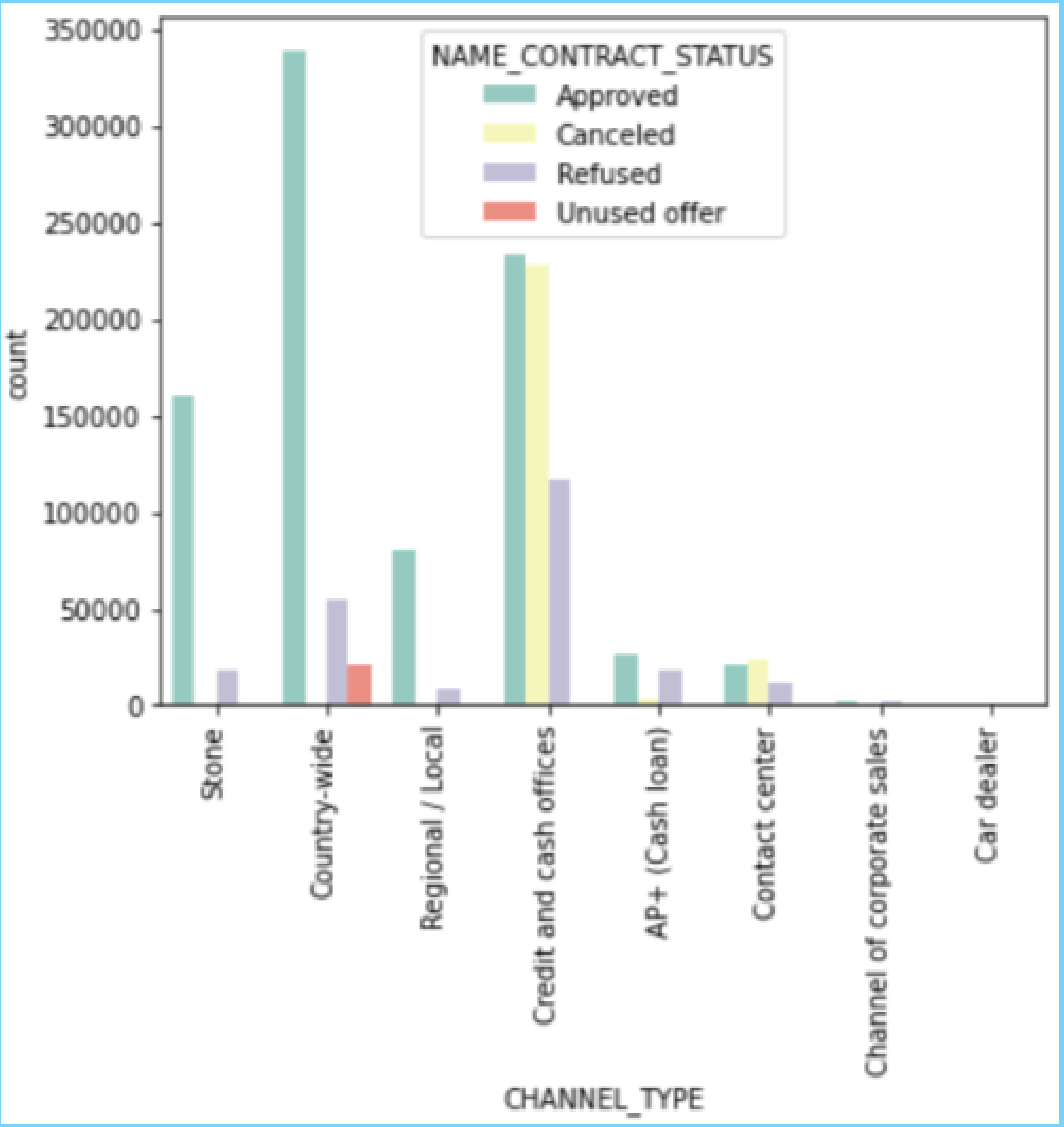


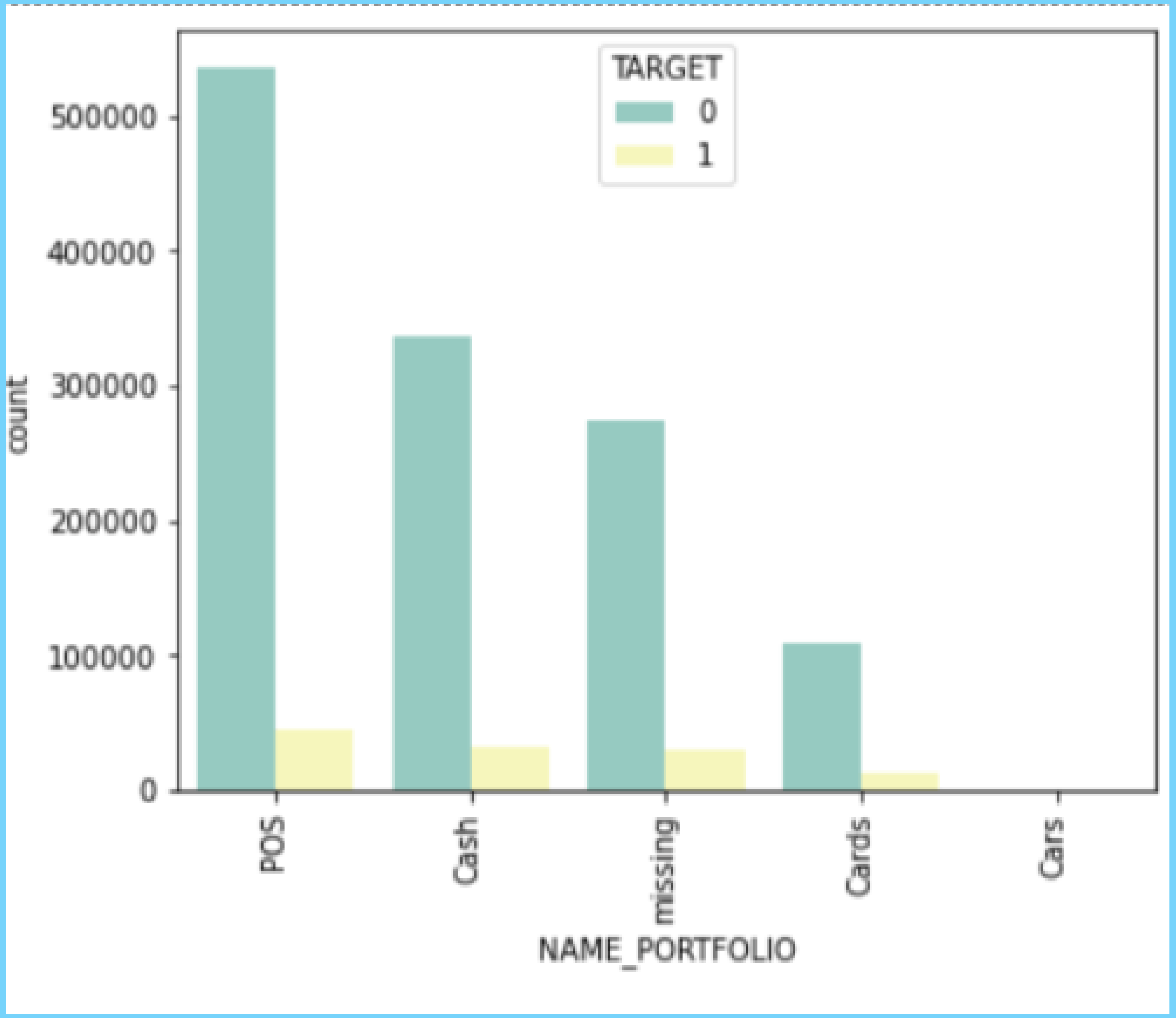
- Those privious_clients whose NAME_PAYMENT_TYPE is 'cash through the bank' were high approving rate then other types
- Those privious_clients whose CHENNEL_TYPE is credit and cash office were high cancel rate

Defaulters analysis in previous_application_data

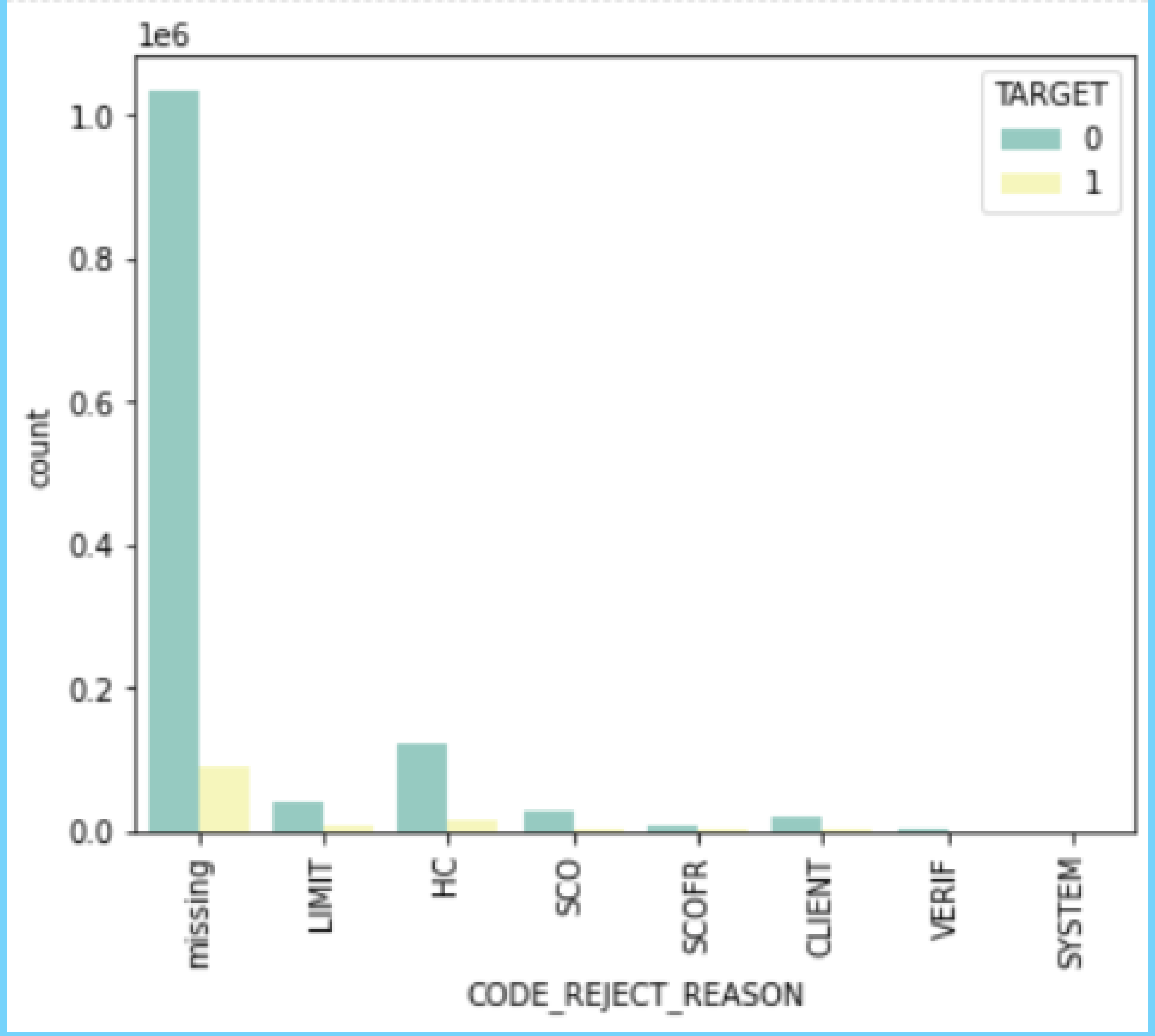


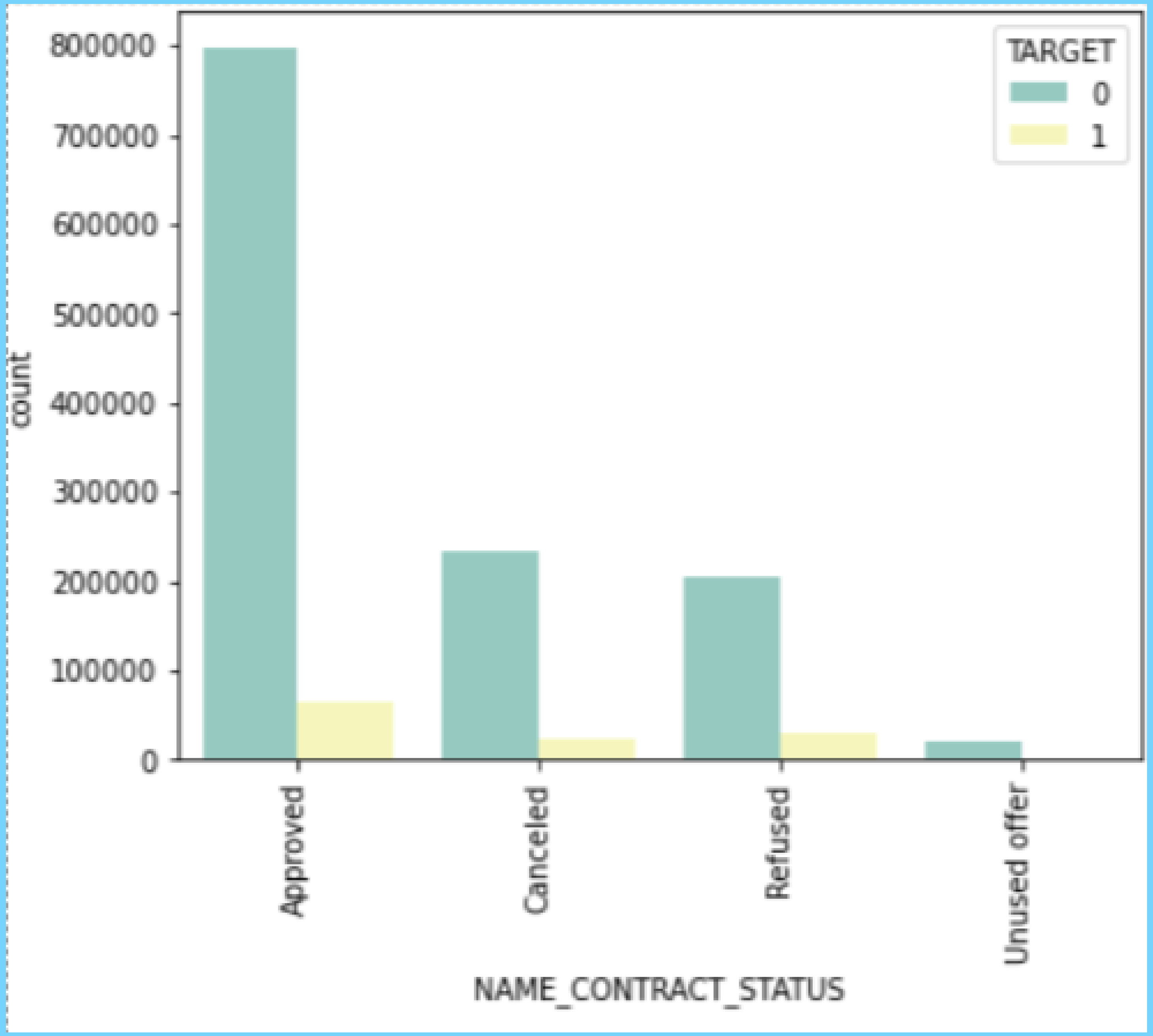
- ‘PRODUCT_COMBINATION’ is an important driving factor.
- These ‘PRODUCT_COMBINATION’ are more then 10% defaulting rate "Cash Street: middle" , "Cash Street: low" , "Cash Street: high" , "Cash X-Sell: high" , "Card Street".





- 12% loan applicatiant defaulted for AP+ (Cash Loan). ‘CHANNEL_TYPE’ is an important feature for analyzing ‘TARGET’ variable
- For “Cards” defaulter percentage is highest (10%). ‘NAME_PORTFOLIO’ is an important feature for analyzing ‘TARGET’ variable.





- ‘SCO’, ‘LIMIT’ and ‘HC’ are the most common reason of rejection.
- Those clients whom previous application was refused the defaulting rate was high around 13% and whom previous application was approved there defaulting rate was less around 8%

THANK YOU!