Credit EDA Assignment

CASE STUDY

Aashna Mehta
aashnam04@gmail.com

Objective

- 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.
- To understand the driving factors (or driver variables) behind loan default, (the variables which are strong indicators of default) which can be utilised by the company for the knowledge of its portfolio and risk assessment.

application_data

This dataset contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.

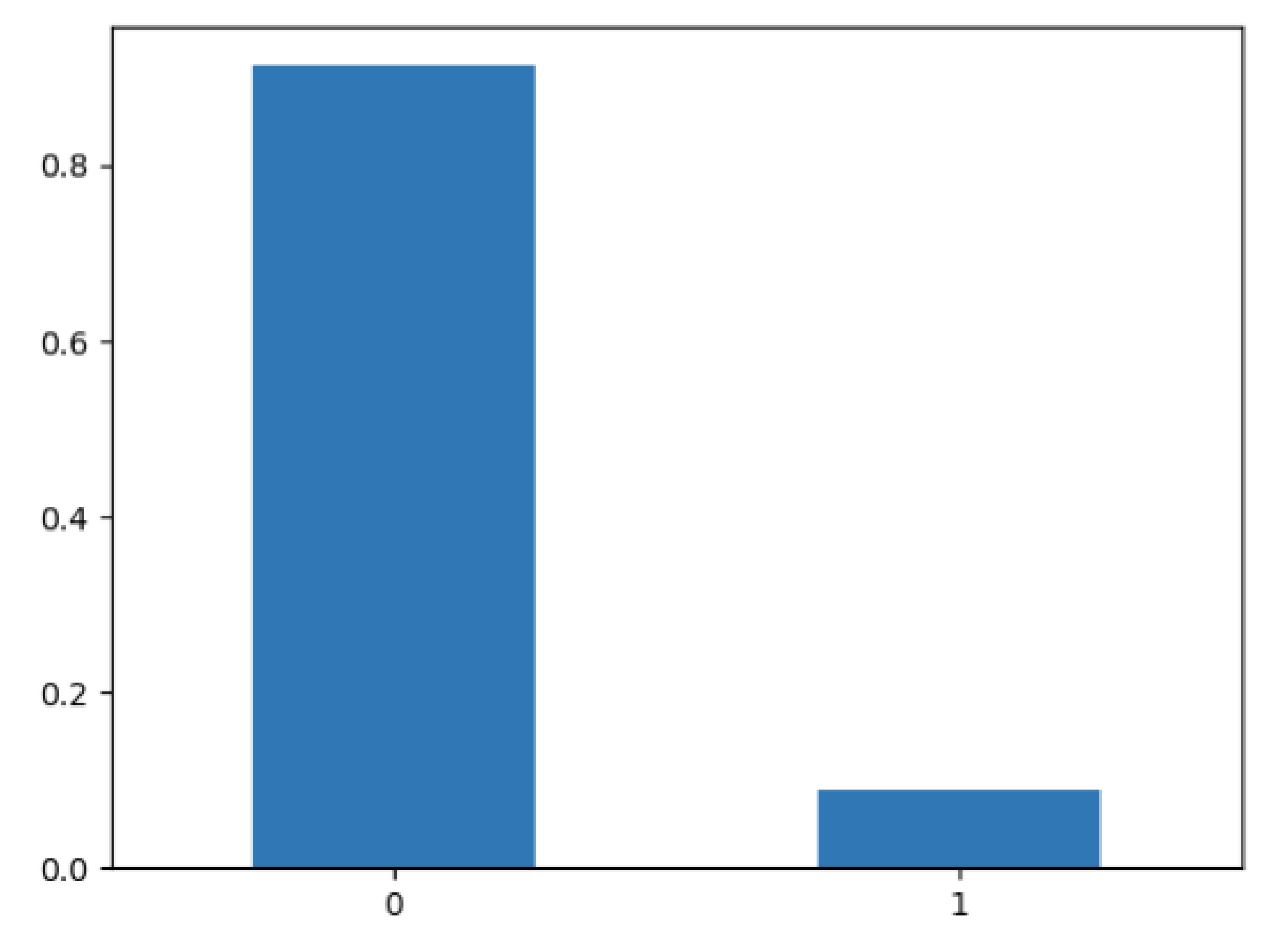
TARGET

This variable has two value counts – 1 and 0

1 - 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 in our sample
0 - all other cases

```
0    0.912148
1    0.087852
Name: TARGET, dtype: float64
```

The graph shows that the target = 0 i.e., clients having payment difficulties are approximately 9 times than the clients having payment difficulties.

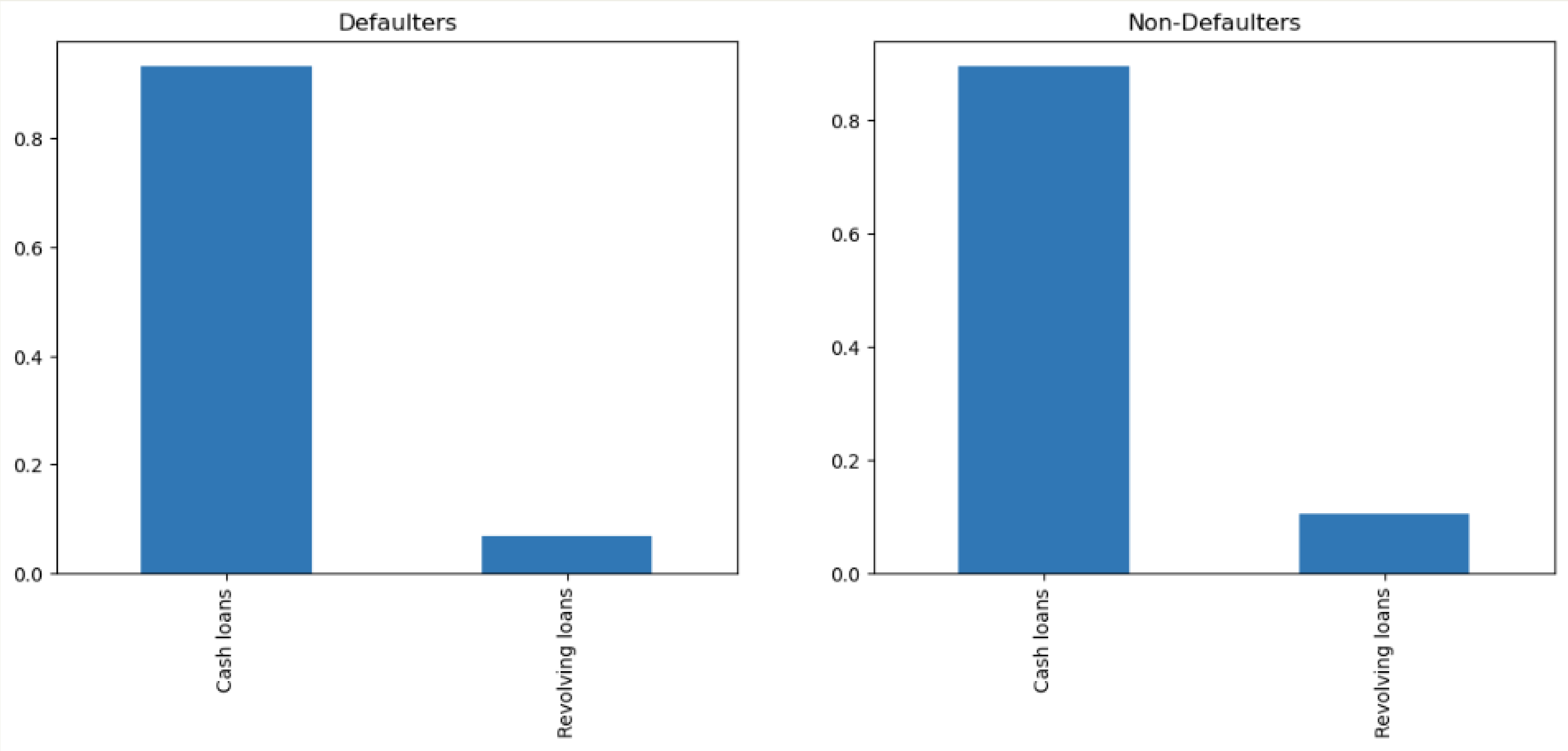


We will divide 'TARGET' into two different sections based on its value counts:

1 – Defaulters

0 – Non Defaulters

LOAN TYPE

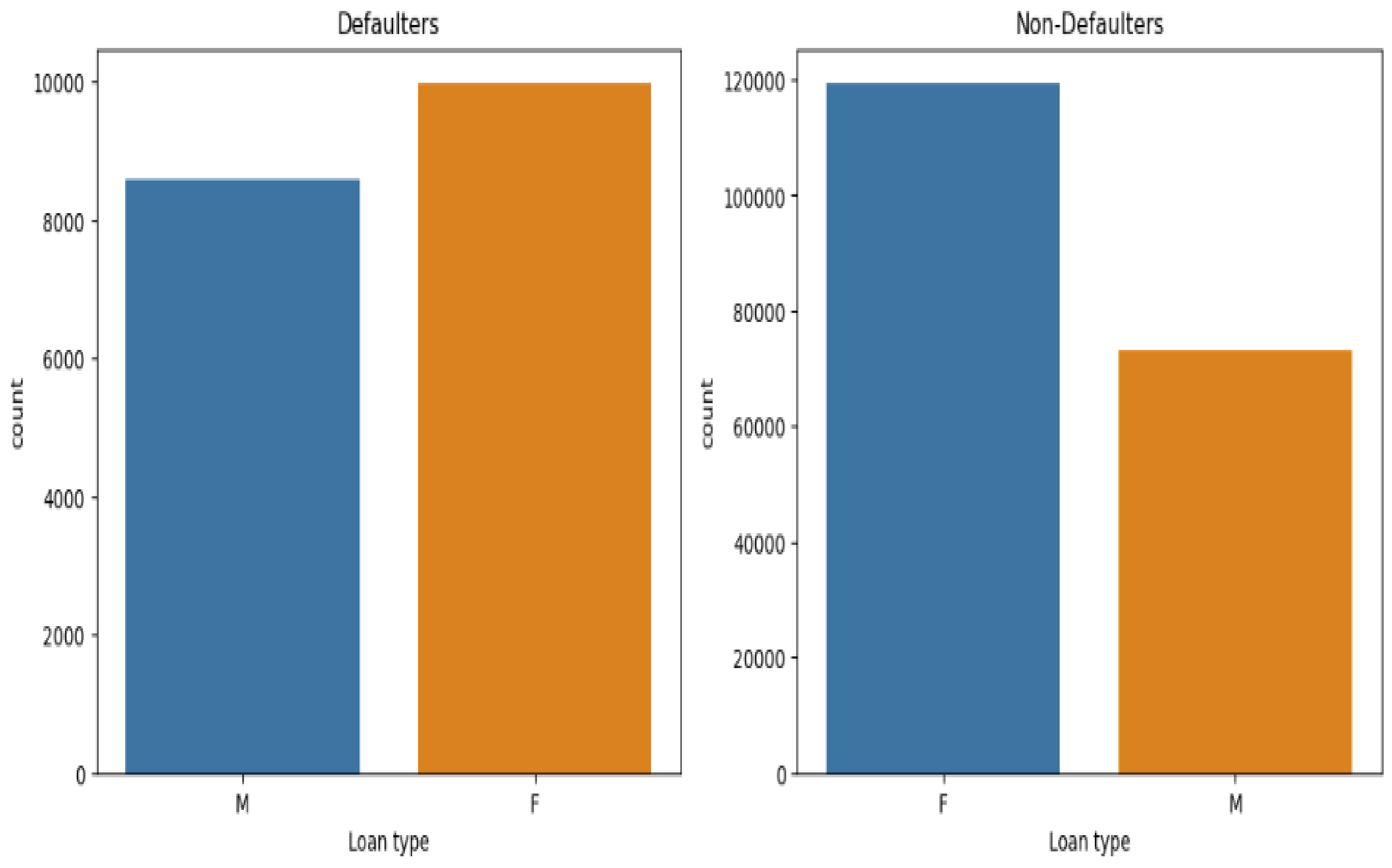


There are two loan types which are Cash Loans and Revolving Loans

The graph clearly shows that the Cash Loans has higher number of Defaulters and Non Defaulters

When we look at the Revolving Loans we see that the percentage for Non Defaulters in this category is higher

Hence, we can conclude that defaulters are higher in Cash as well as Revolving Loan type.



GENDER

From the graph we can see that the higher number of defaults are made by Female, and they are the highest number for Non-Defaulters as well.

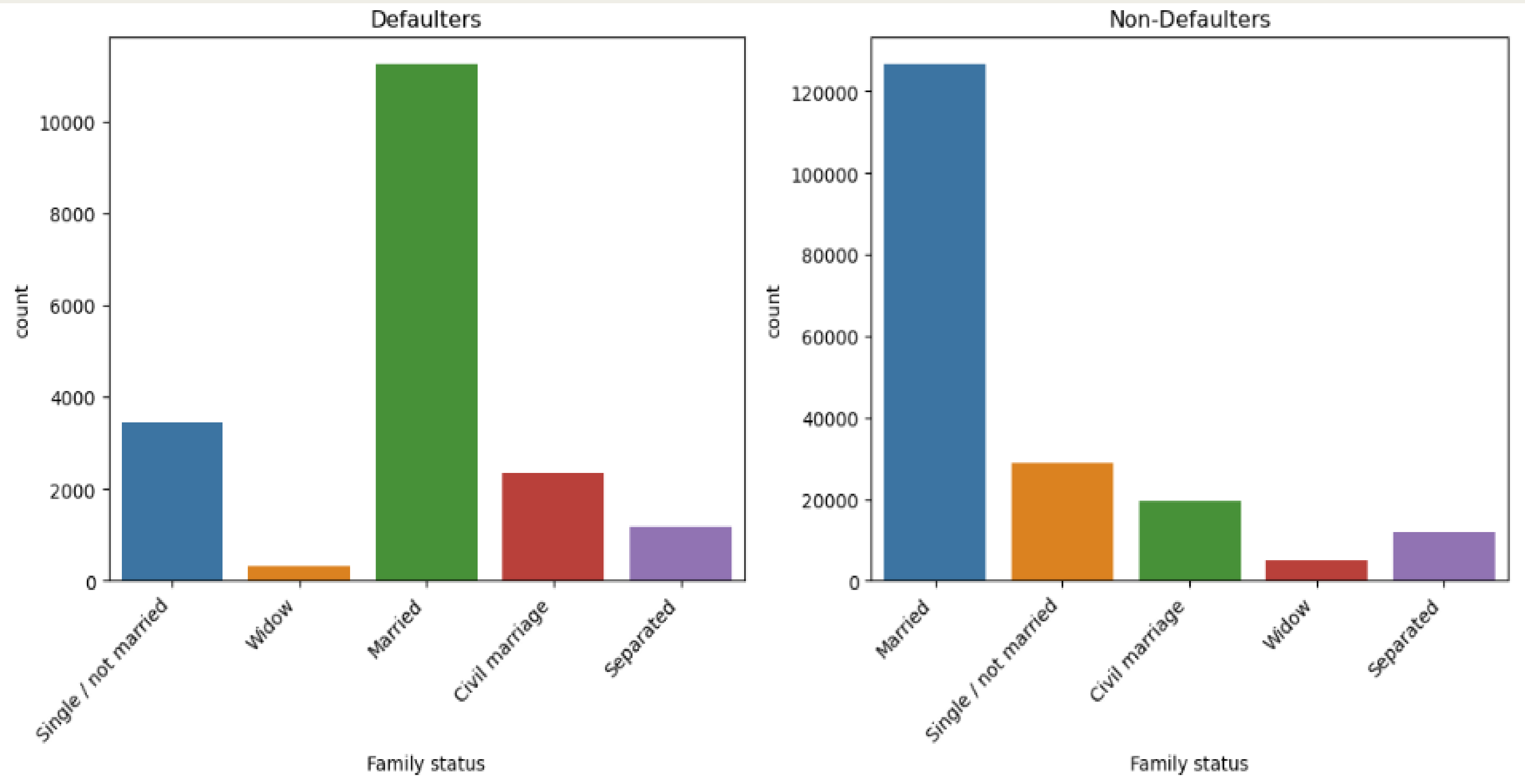
When Males are comparatively high as defaulters when compared individually to Non-Defaulters

FAMILY STATUS

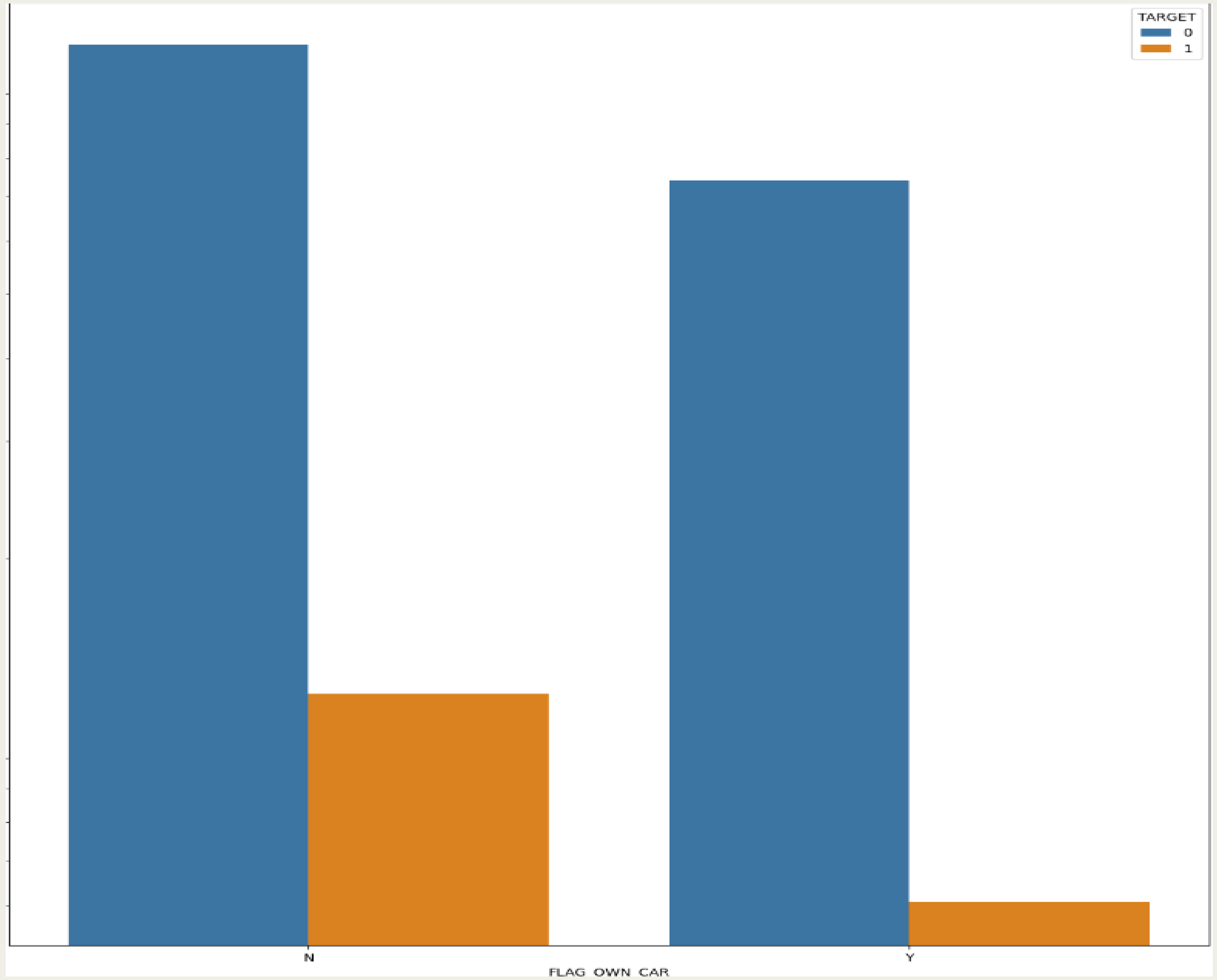
When compared by the Family Status it is observable that Married people are the highest number of Defaulters and Non – Defaulters.

This implies that people who are married are taking loans more comparatively.

And Widows are the least for the same.

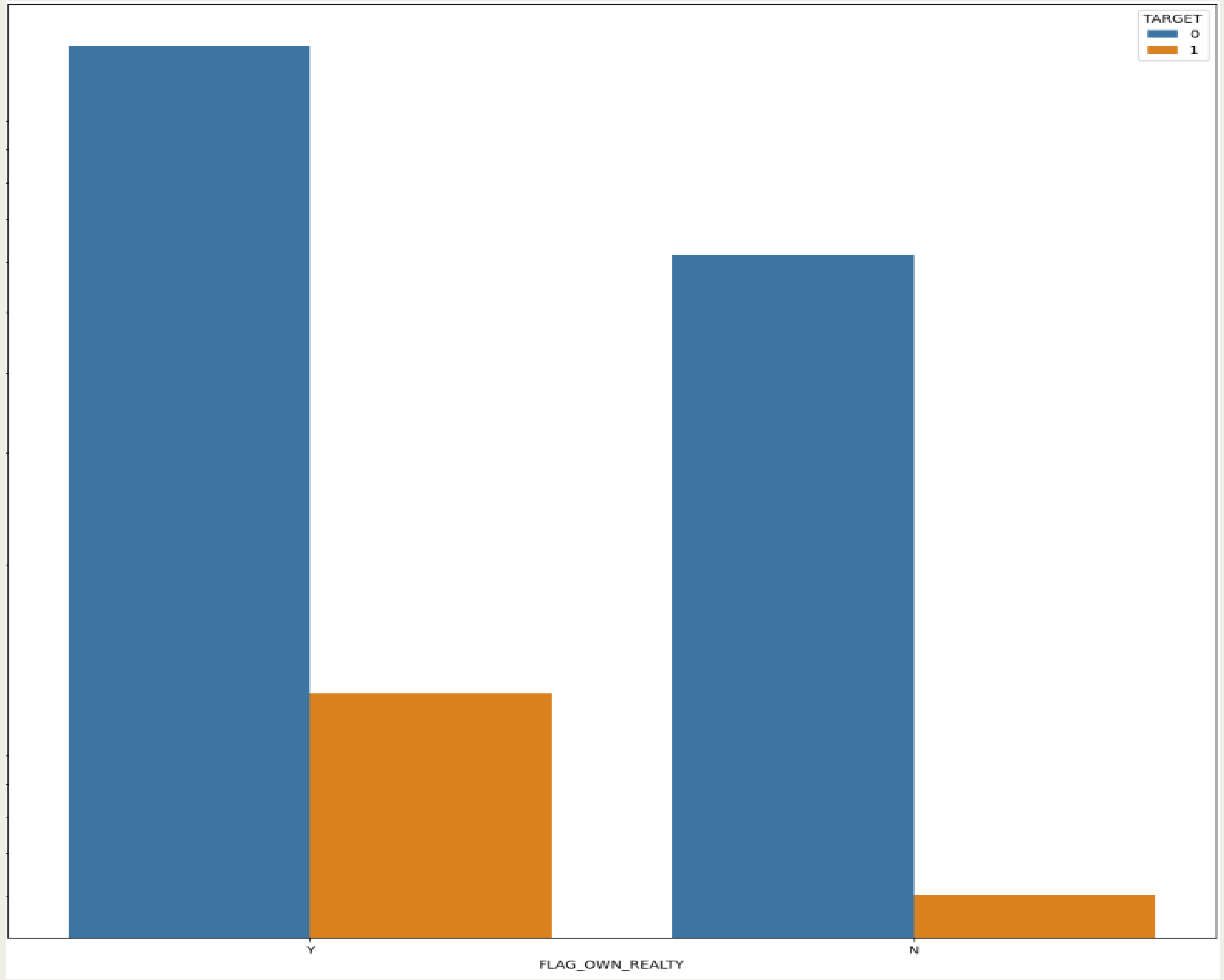


CAR OWNER

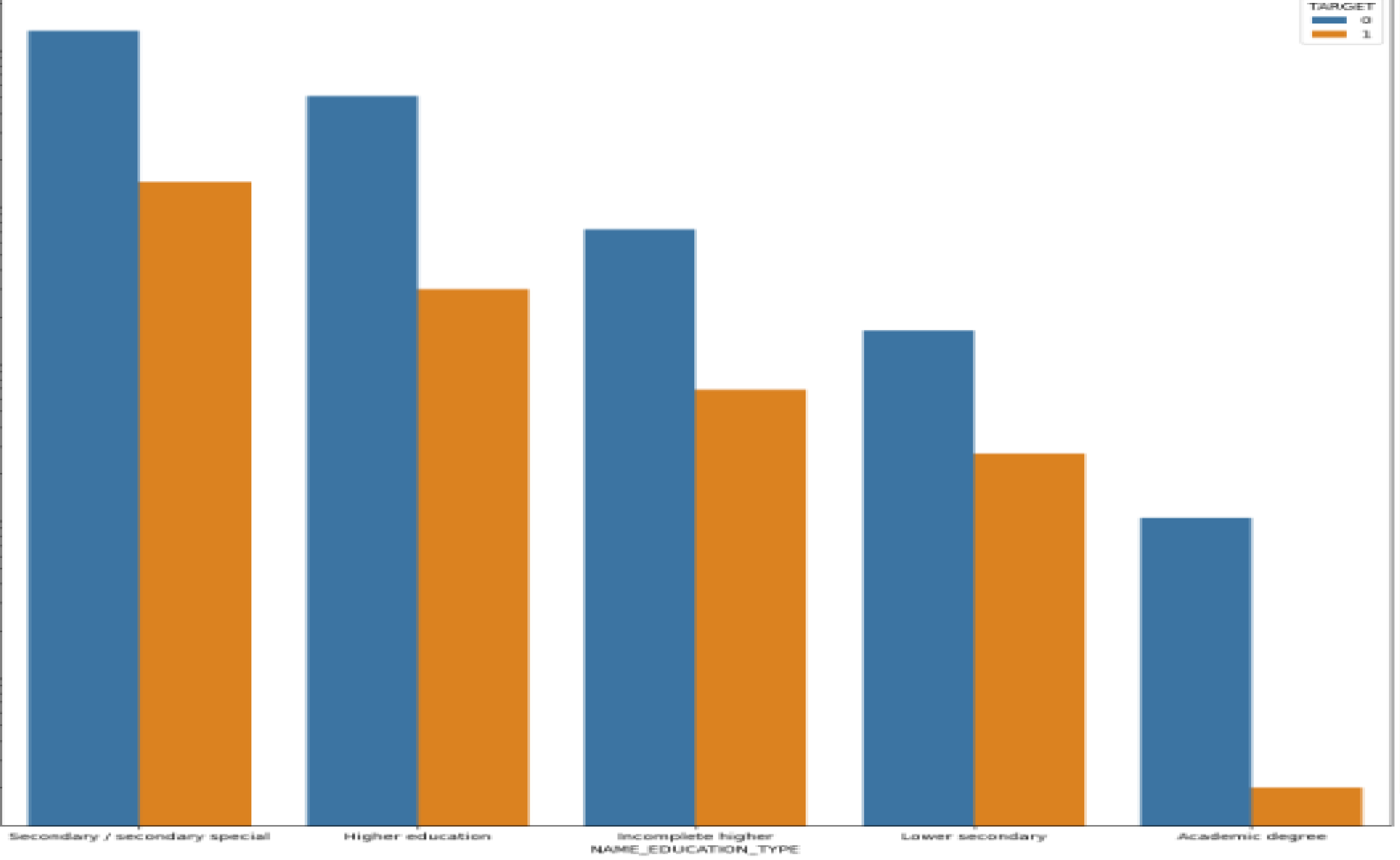


From the graph, people not owning car are more likely to make defaults then non defaulters.

FLAT OWNER



From the graph we can see that the higher number of defaults are made by people owning real estates showing that they are more likely to take loans as well.

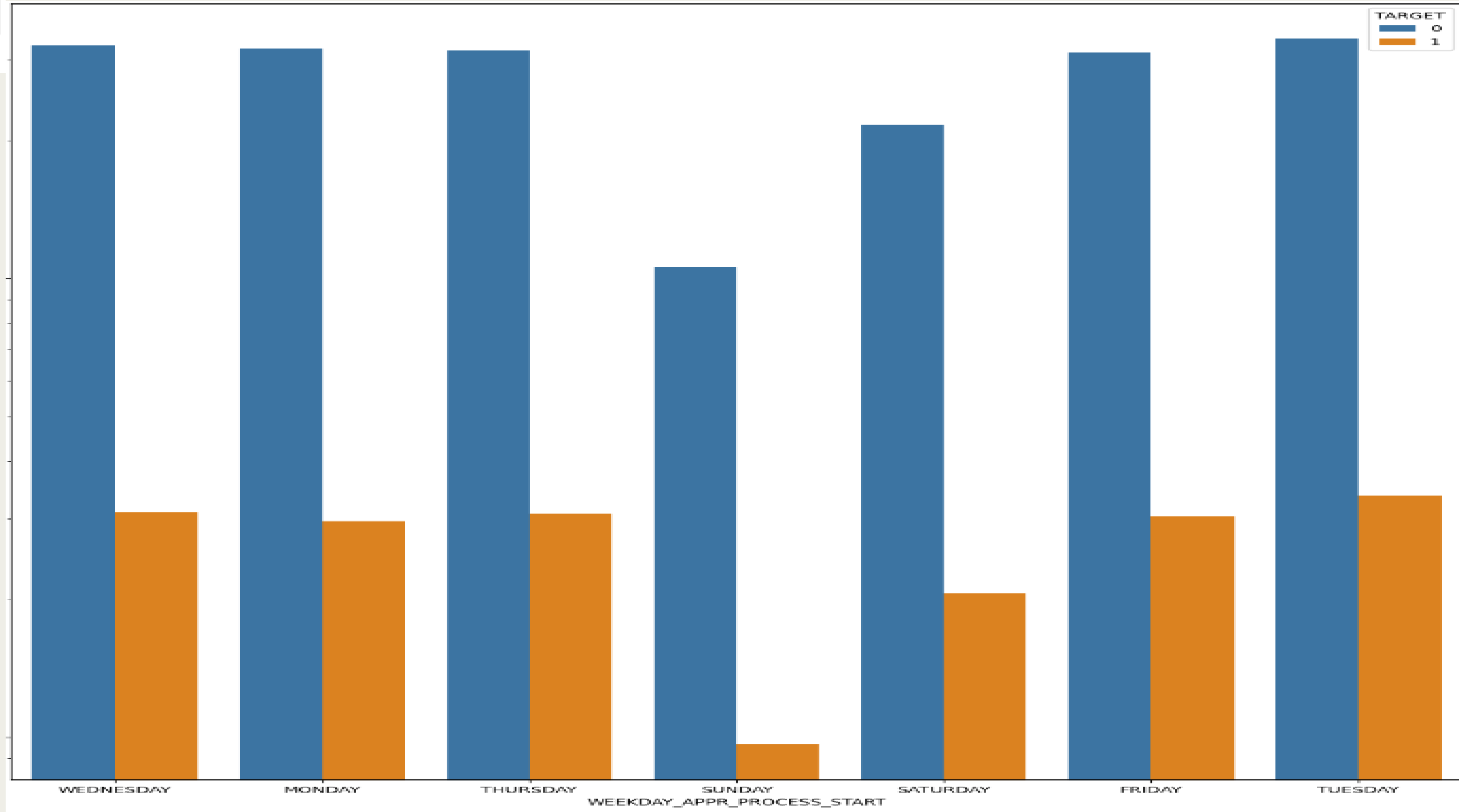


EDUCATION TYPE

We can also conclude that people educated till Secondary/Secondary Special in 'NAME_EDUCATION_TYPE' applies loans which are higher in number.

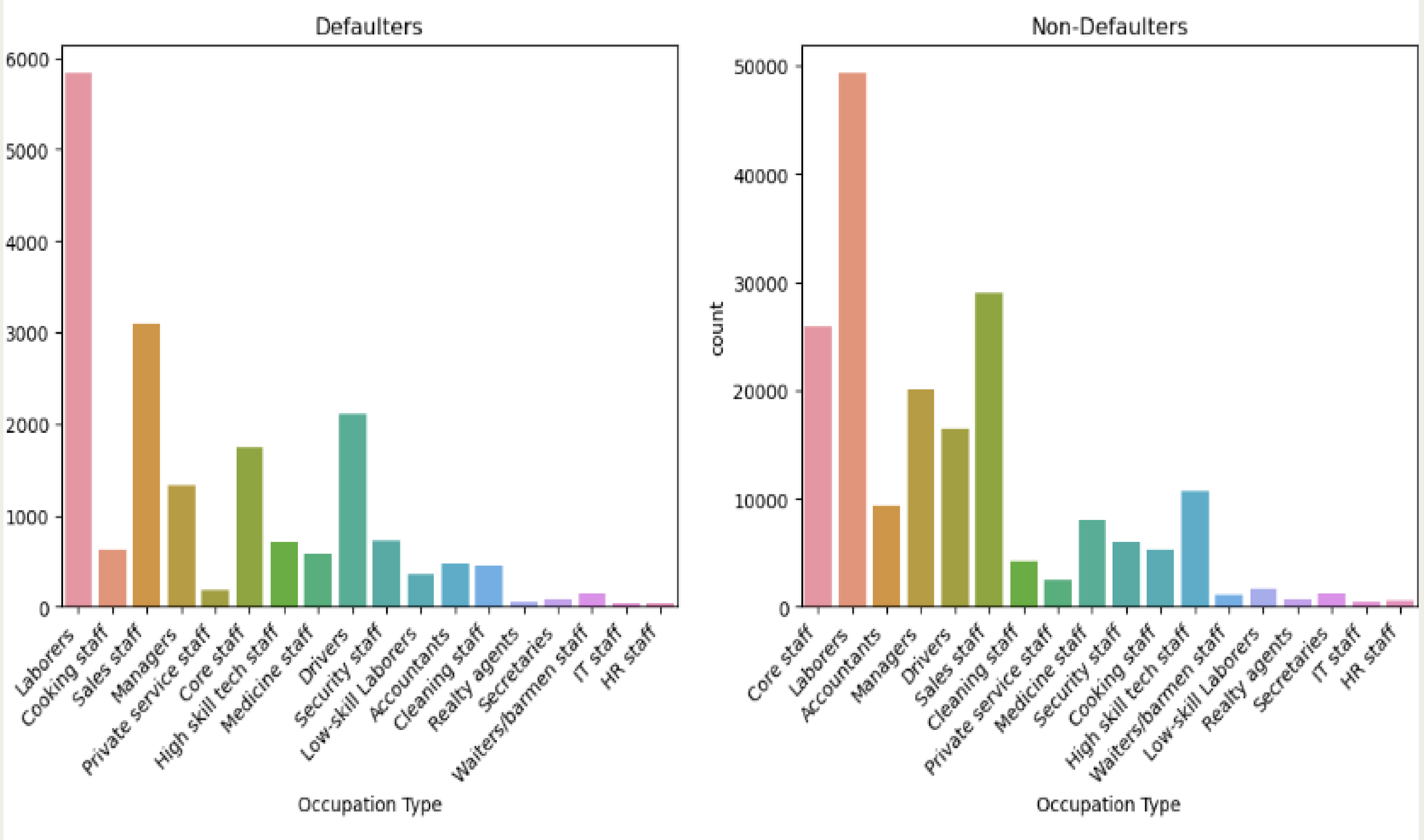
PROCESS START - WEEKDAY

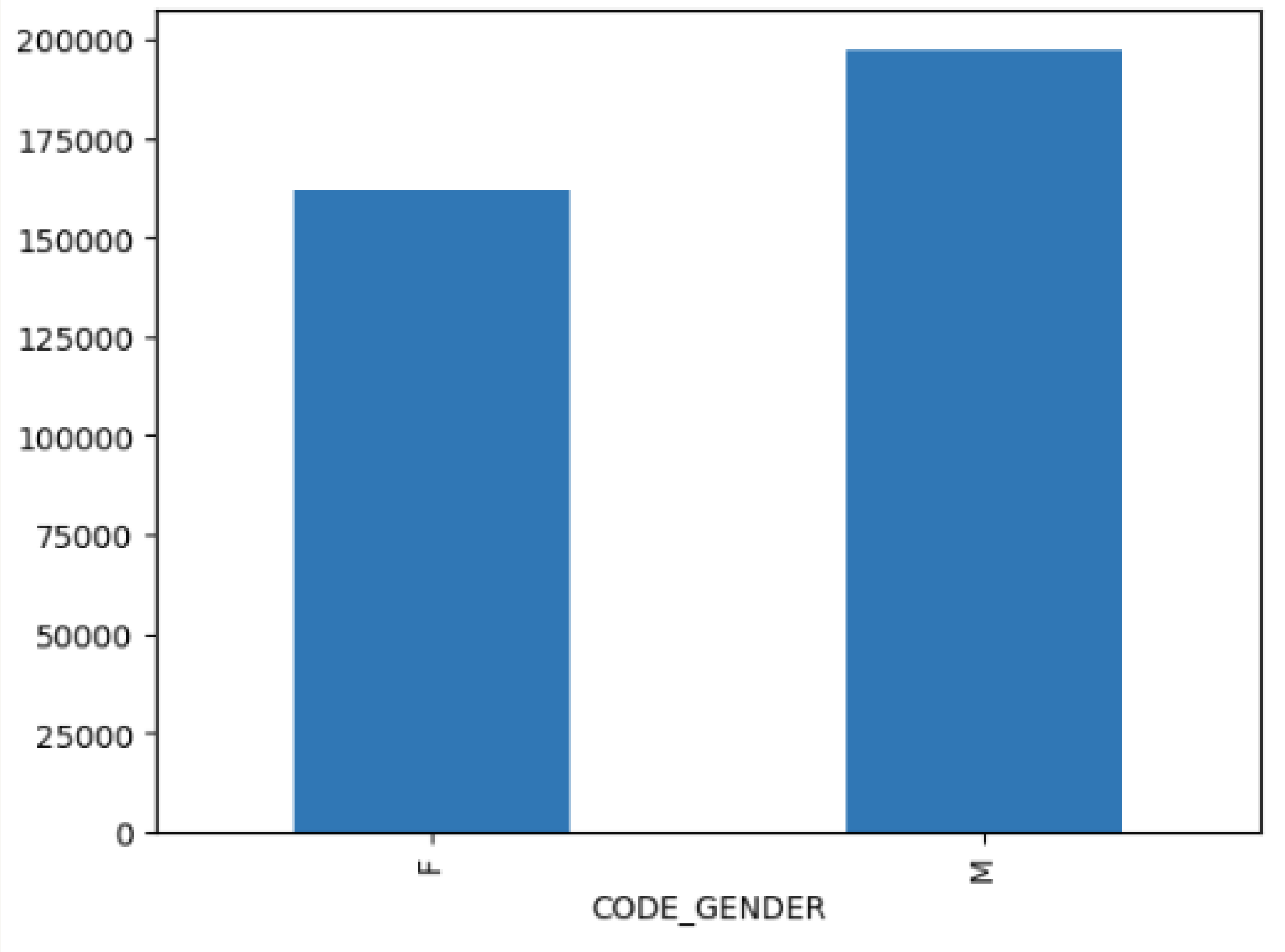
We can see that weekend i.e., Saturdays and Sundays are less busy for banks as the amount of loan applications are less when compared to weekdays.



OCCUPATION TYPE

Laborers are the category with the highest defaulters and non defaulters showing that they are the ones taking more loans which is quite self explanatory.



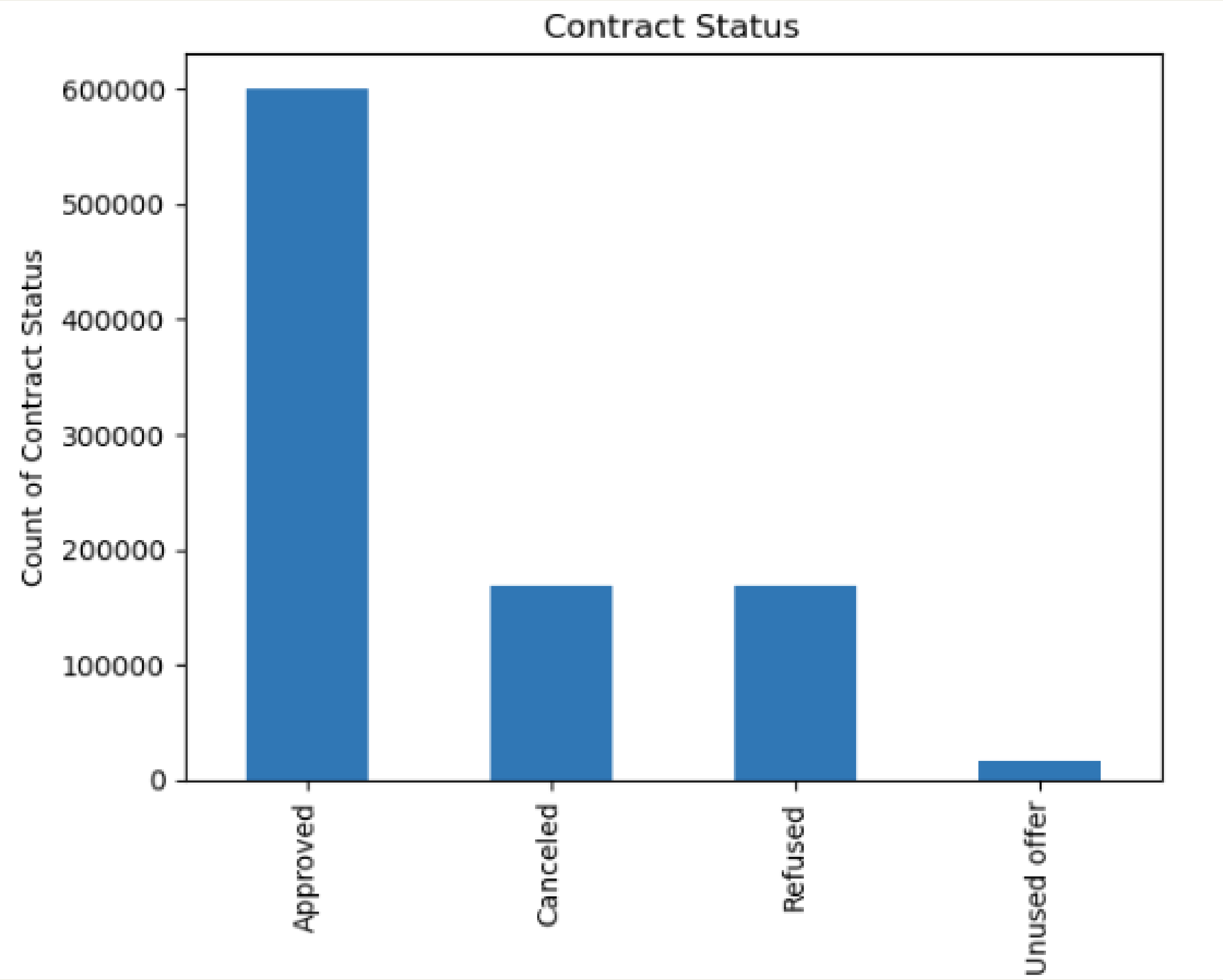


CODE GENDER AND TOTAL INCOME

Here, the CODE_GENDER is grouped by AMT_INCOME_TOTAL and we can see here that despite female count being more than male the mean for male is greater when we group by AMT_INCOME_TOTAL showing that male earns more than female

previous_application

This dataset contains information about the client's previous loan data. It contains the data on whether the previous application had been Approved, Cancelled, Refused or Unused offer.



CONTRACT STATUS

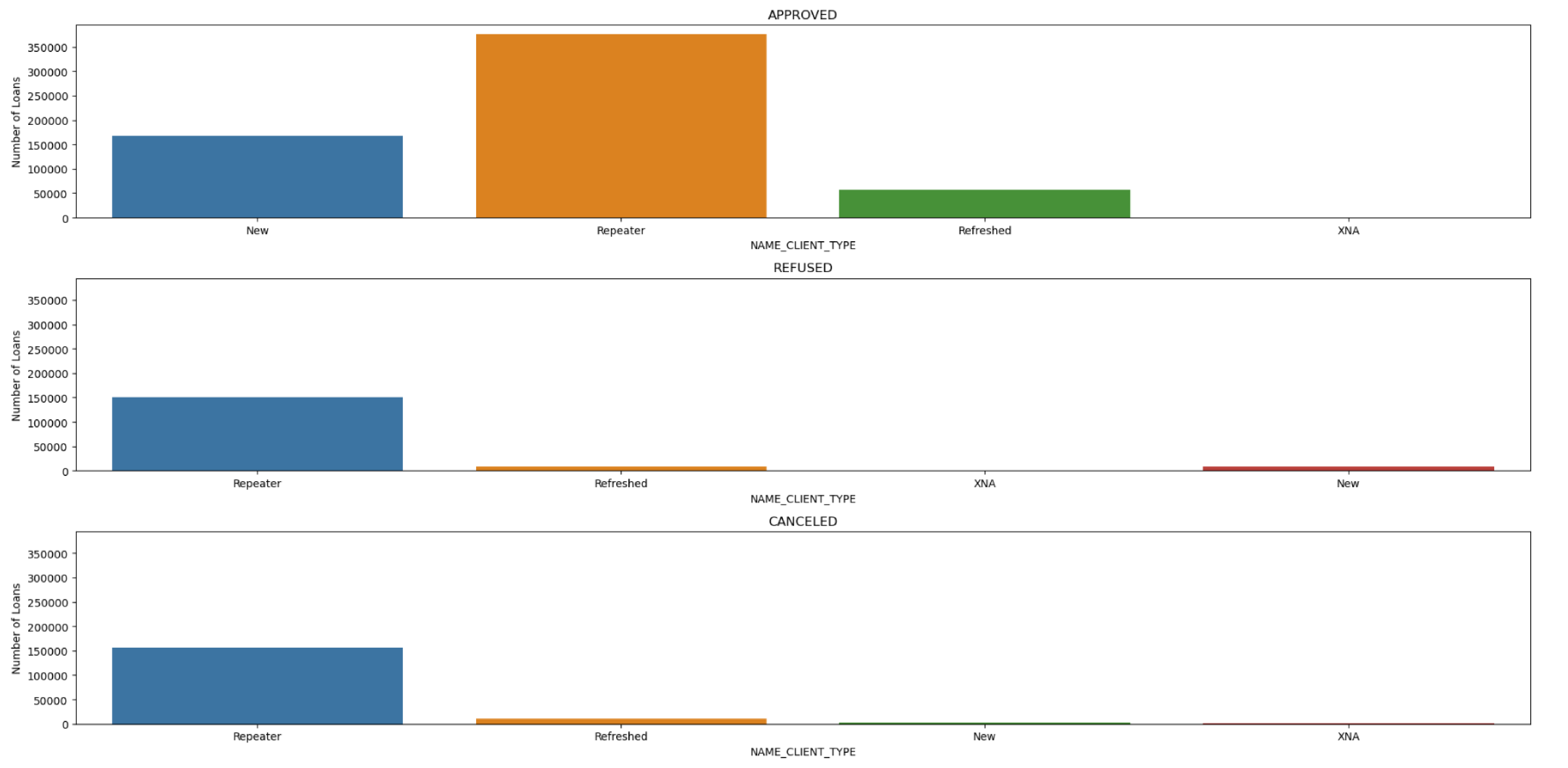
We see that maximum loan applications had been approved previously and the number of refused and canceled loan applications were almost same.

```
Approved        62.831278
Canceled        17.667004
Refused         17.663343
Unused offer     1.838375
Name: NAME_CONTRACT_STATUS, dtype: float64
```

Maximum loan applications were approved whereas almost equivalent loan applications were canceled or refused.

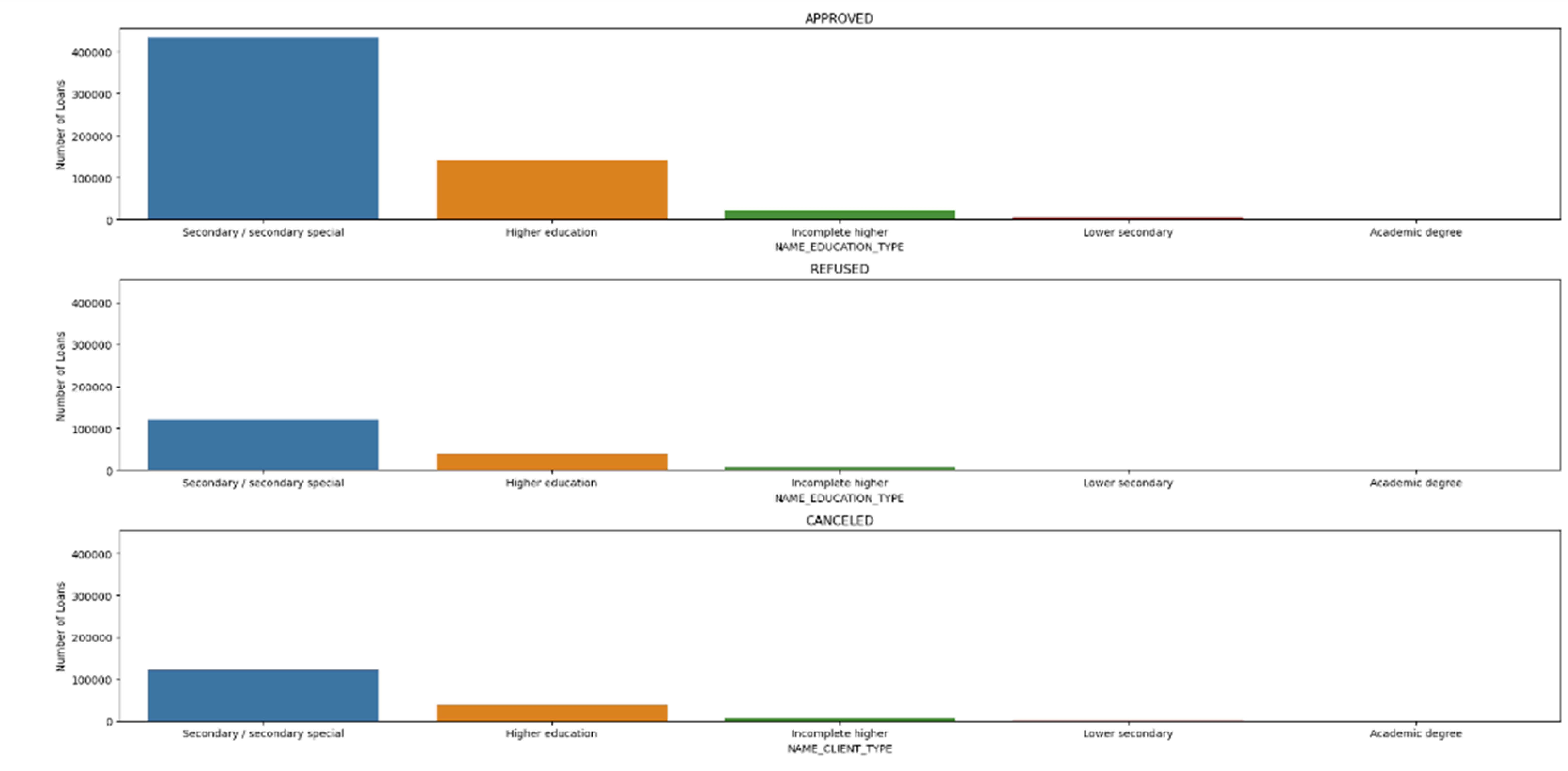
Just 2% of the loan applications are unused offer.

CLIENT TYPE



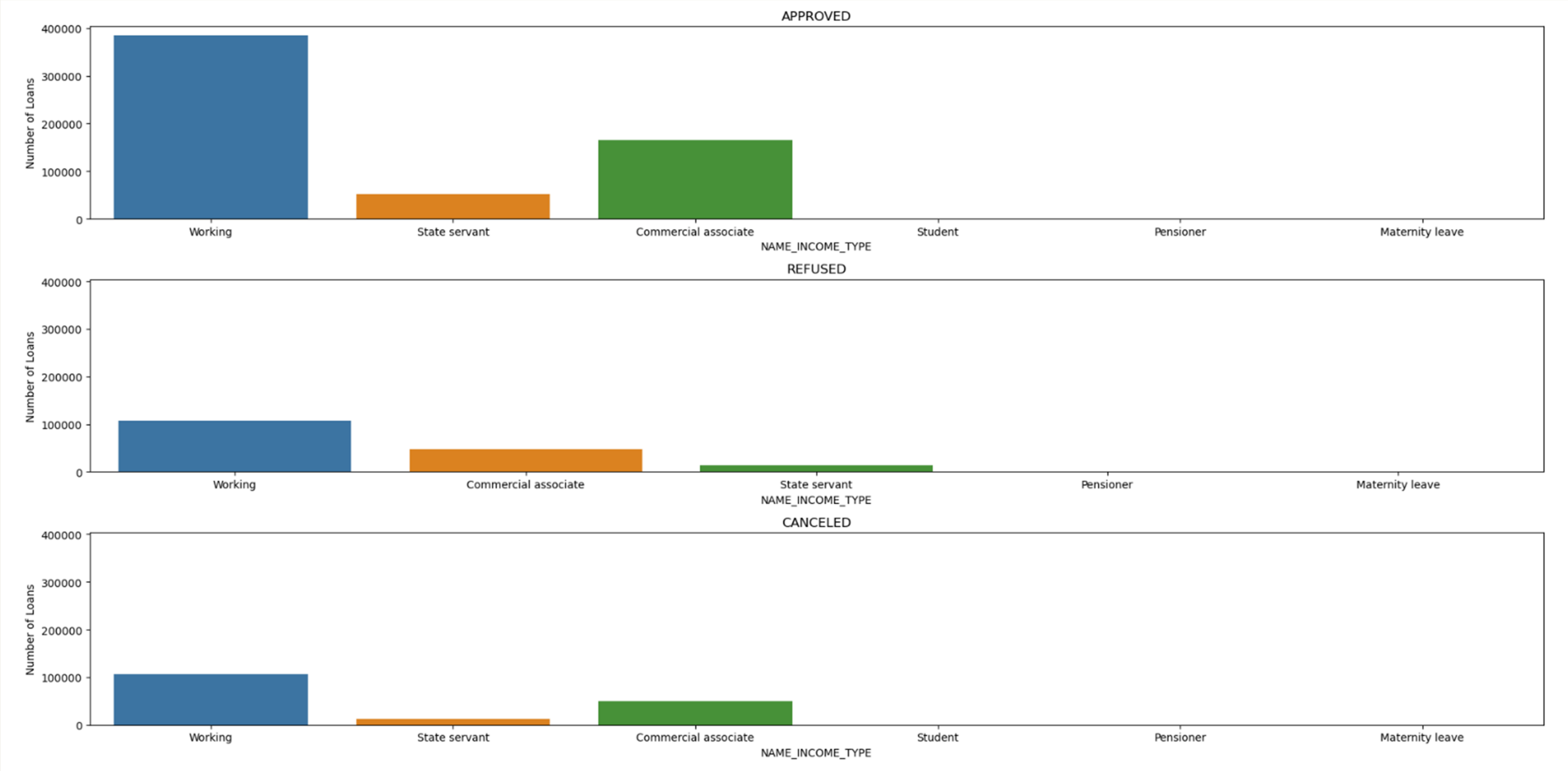
It is clear that the ‘repeater’ type of clients had maximum loan applications approved, refused and canceled. And ‘new’ type of clients had their loan applications approved when compared to canceled and refused.

EDUCATION TYPE



As observed earlier that Secondary/Secondary Special education type people used to apply for maximum loans, here we can see that this type used to get approved, canceled and refused the maximum with the second highest being the higher education type.

INCOME TYPE

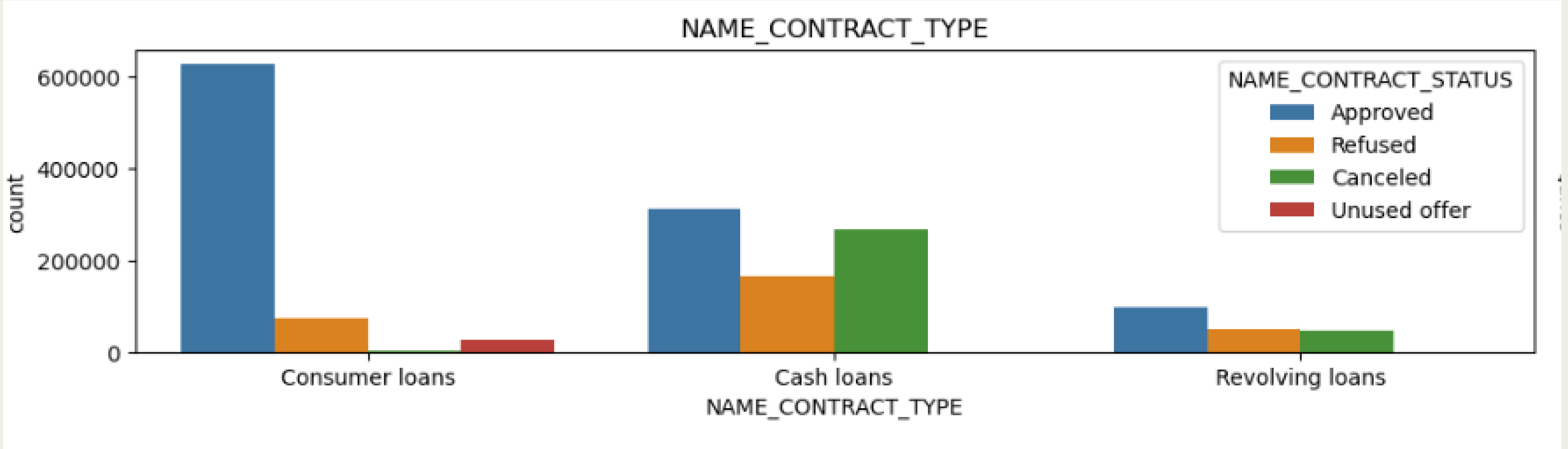


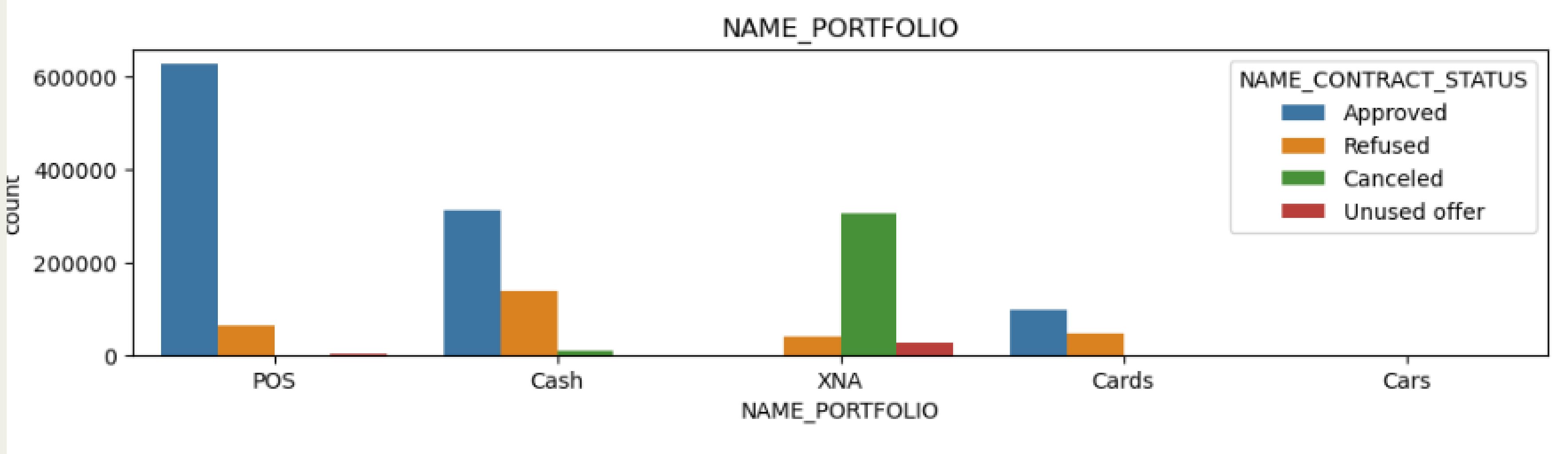
It is obvious that working income type used to get their loan applications approved, refused and canceled the most. As they are the ones taking more loans when compared to other categories.

CONTRACT TYPE

Consumer Loans are getting the highest approvals with no canceled loans being in this category.

Highest number of refusals is for cash loans than any other.





PORTFOLIO TYPE

POS is the portfolio with highest number of approvals.

And again, we can see that more cash loans have been refused than any other portfolio type.

CONCLUSION

- One reason for lower unused offers can be that this type had lower loan amounts.
- With more than half of the percent defaulters are working applicants, more exploration should be done for the same.
- With working income type, occupation as laborers and almost 75% not owning home are the ones making defaults in the payments and are also the ones getting the loans approved, hence more importance should be given for their analysis.
- It is more advisable to approve loan applications for the once having higher education for all professions as the chances of recovering the loans are comparatively faster in this case because of their higher income earning capacity.
- It is riskier to grant loans to the ones whose previous application were either canceled or refused.