

Credit EDA Assignment

Vedhavathi Nanjappa
DS 52

Business 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. 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.

Python Libraries Used

1. NumPy
2. Pandas
3. Matplotlib
4. Seaborn

Steps followed

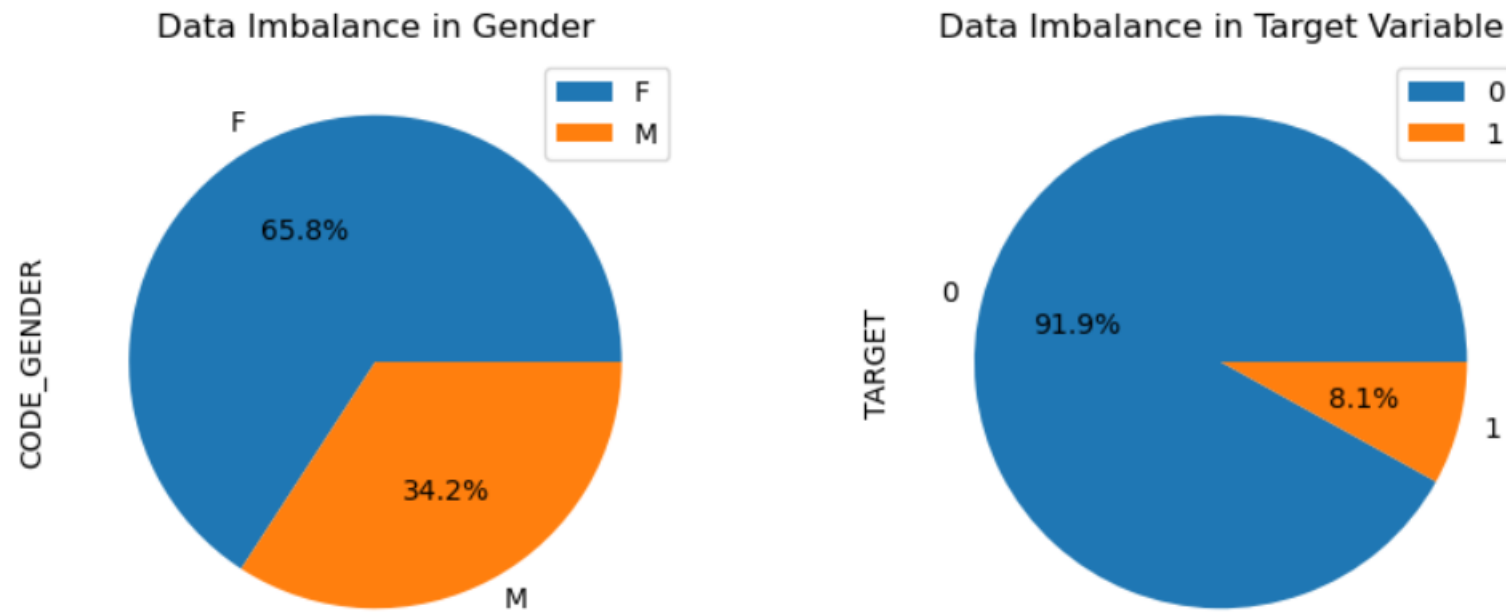
1. At first imported application_data.csv which mainly consists of applicants data having payment difficulties and also cases where payment is made on time.
2. Analysed the data set and cleaning of the data wherever needed
3. Imported previous_application.csv dataset , cleaning of this dataset and analysis of the data set
4. Finally merging of the above 2 datasets and analyse the whole dataset with some visualisation charts.

EDA steps followed (more in detail)

1. Import dataset
2. Understand the data and its various columns
3. Check for null values
4. Remove columns having more than 40% null values (Approach followed in this case study)
5. Impute rows (if needed)
6. Changing the datatype of columns if needed
7. Univariate, Bivariate and multivariate data visualization

Data Imbalance in Target column and Gender column

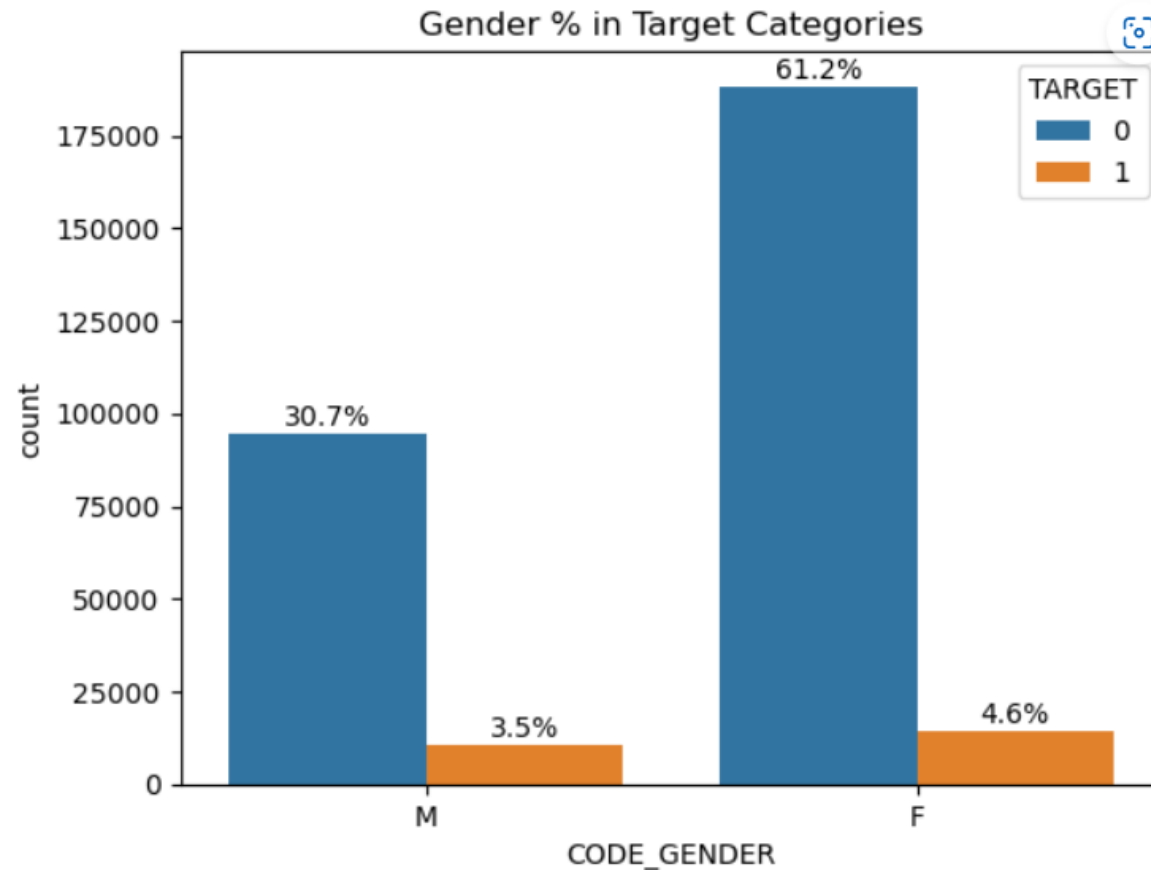
We can see that number of female applicants are more and also the number of non defaulters are more compared to defaulters



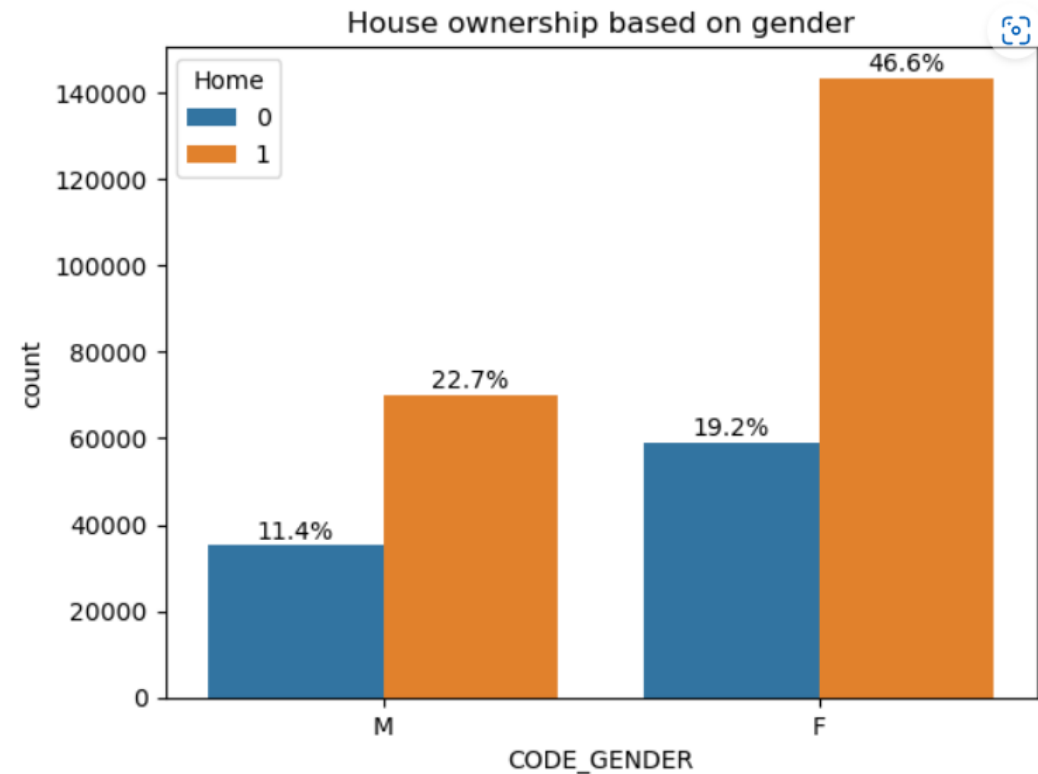
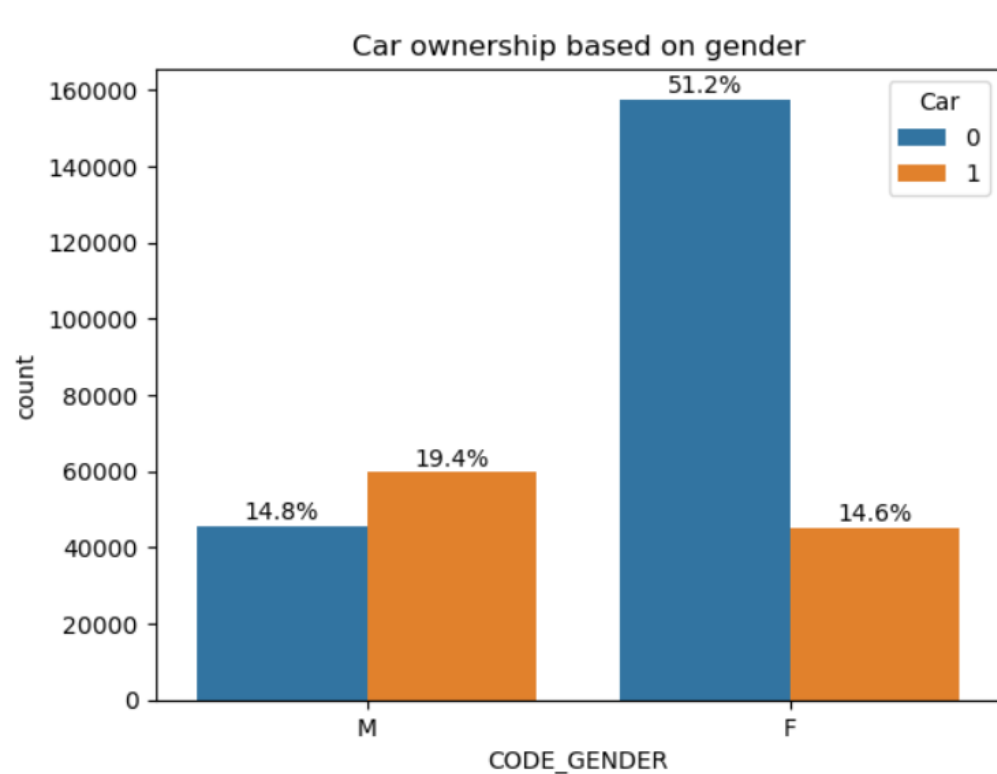
Defaulters based on Gender(Male or Female?)

The graph clearly shows that Females are the applicants who are not having much difficulty in making payments compared to men.

Inference: Females can be considered as better candidates.

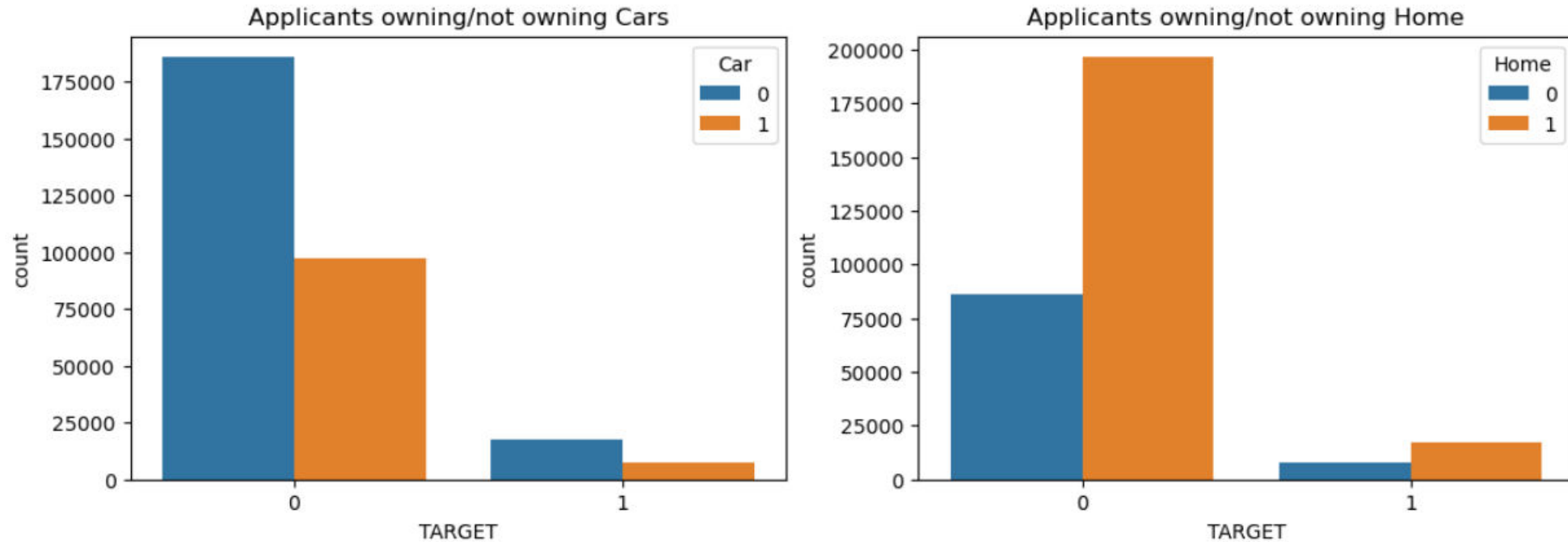


Which gender is interested in owning car/house?



We can see that Male population are interested in owning cars and female population are interested in owning house. Females are more interested in investing in real estate rather than depreciating asset like car.

Car ownership or House ownership?



Inference:

1. Applicants who do not own car are not facing challenges in making payments.
2. Applicants who own home are diligent in making payment.

Average age of applicants gender wise

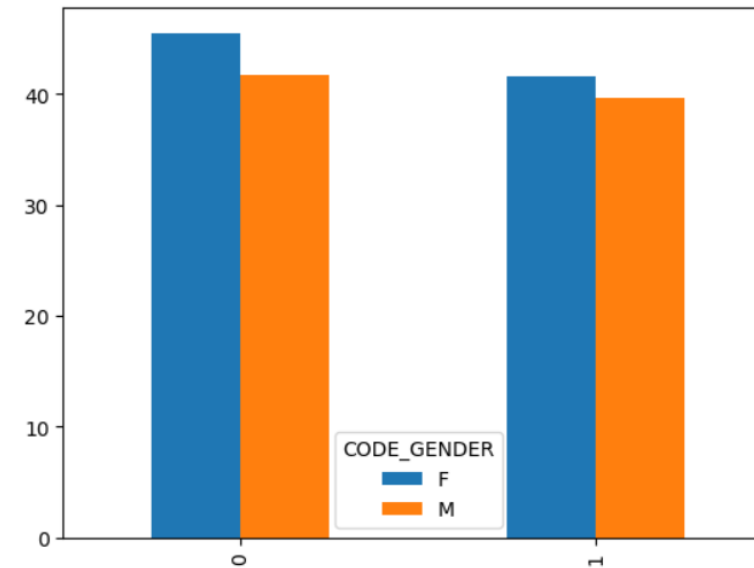
Female population with an average age of 45 are better applicants who are diligently paying the loans.

Out[356]:

	CODE_GENDER	F	M
	TARGET		
0	0	45.483001	41.689978
1	1	41.646366	39.625434

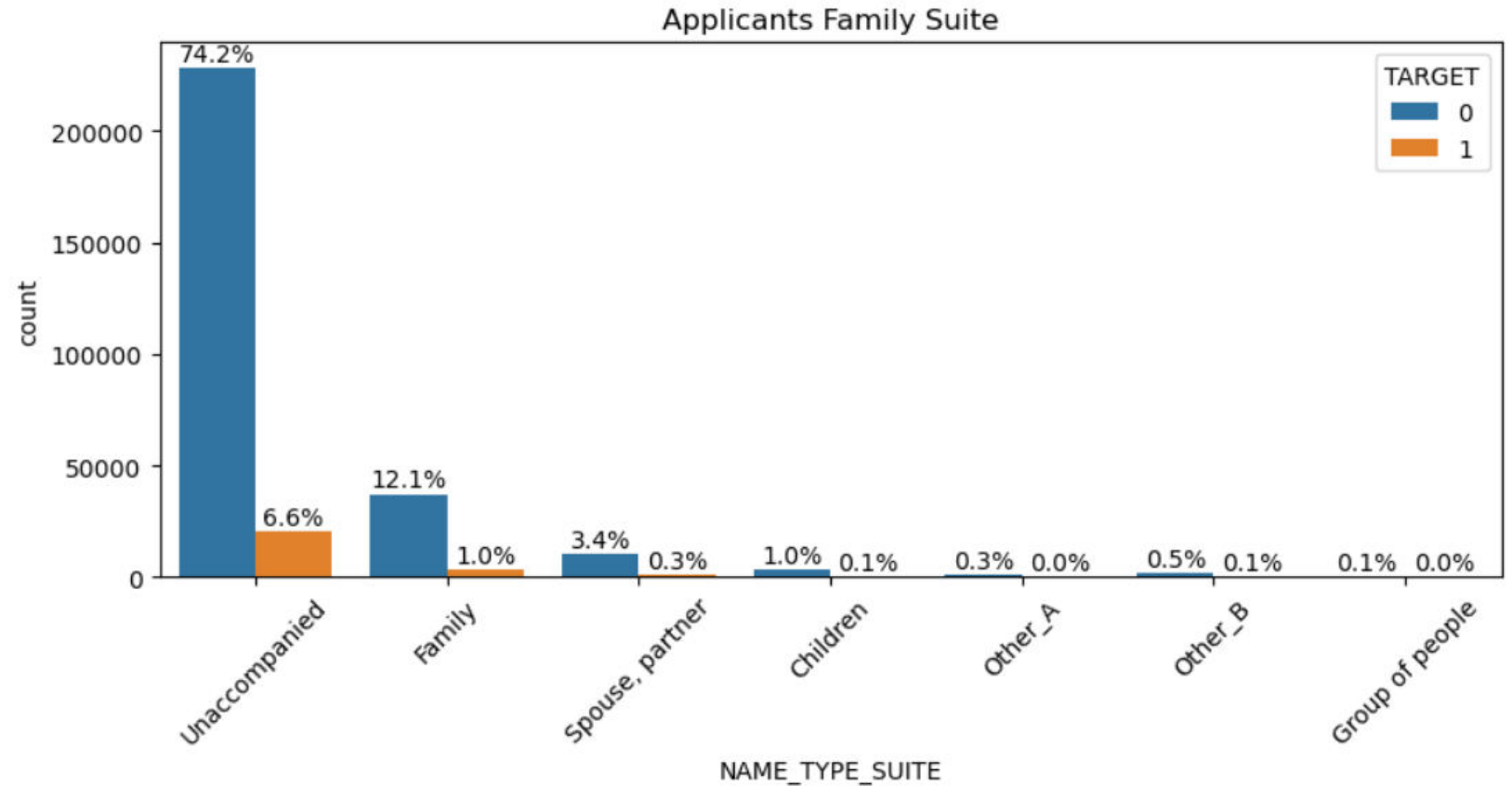
In [528]: `h.plot(kind = 'bar')`

Out[528]: `<AxesSubplot:xlabel='TARGET'>`



Who was accompanying client when he was applying for the loan

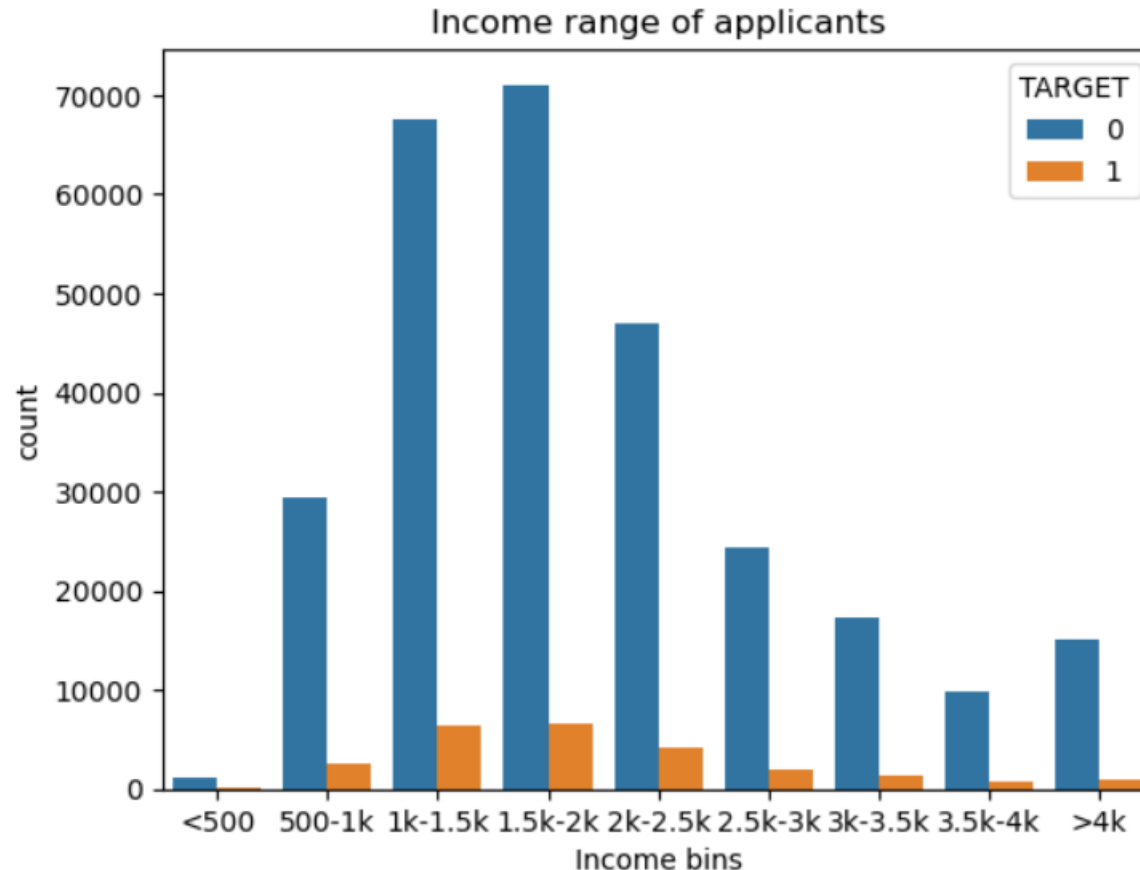
Most of the applicants in both the sections were unaccompanied when applying for loans. However more Target 0 applicants were accompanied by spouse, children and others. We can infer that these applicants have more family support which is why they are paying off loans may be even when they are in challenging situations.



Income range of applicants

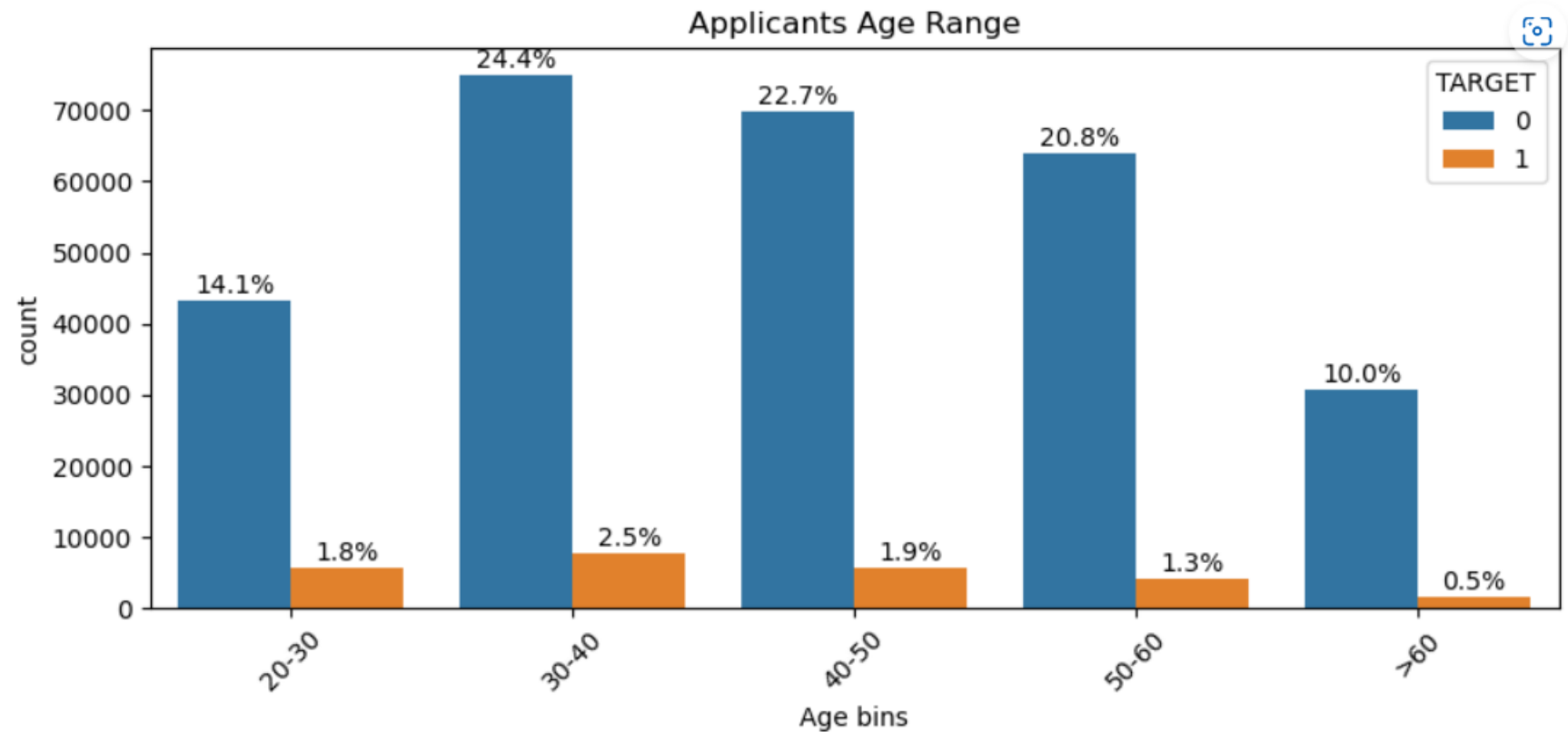
Note: Amount converted to dollars in the dataset (1 dollar = 81rupee)

We can infer that applicants in the income range of 1.5 – 2k dollars are good candidates in paying off loans



Applicants age range

Applicants in the age range of 30-50 year olds are making the payments diligently. This may also be attributed to the fact that these candidates would want to secure future before retirement.

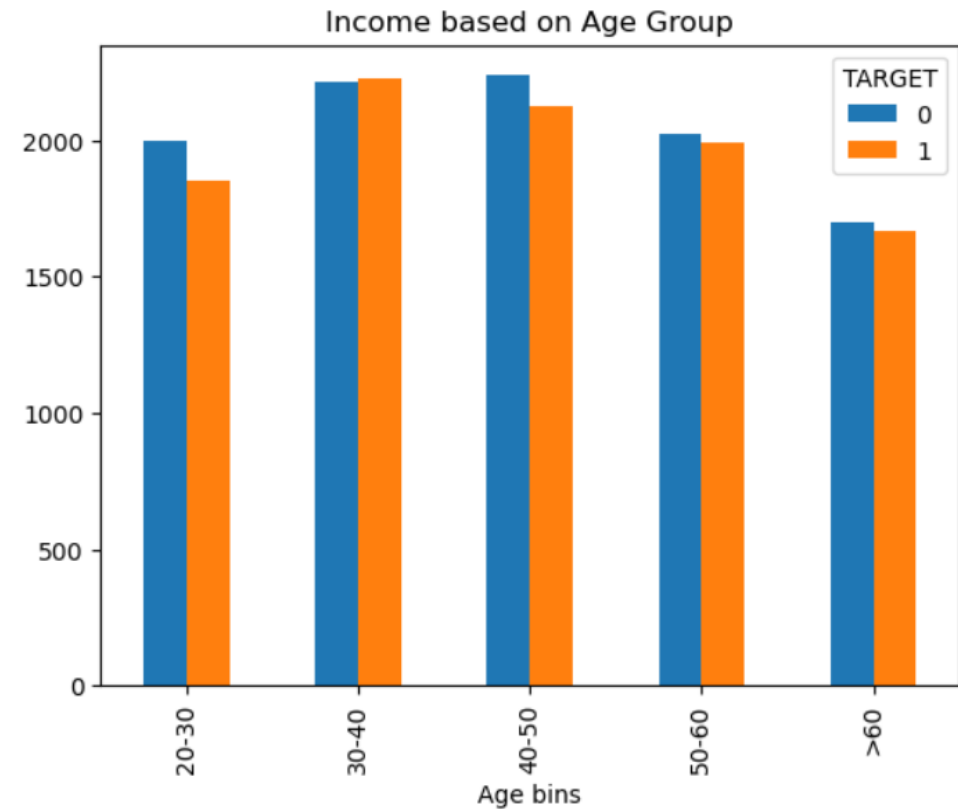


Income based on age group

Applicants in the age group 30 - 50 years have higher income range

Out[432]:

TARGET	0	1
Age bins		
20-30	1997.263183	1853.477365
30-40	2211.031067	2227.608849
40-50	2237.149351	2125.361799
50-60	2026.260211	1991.533520
>60	1700.421099	1665.445551

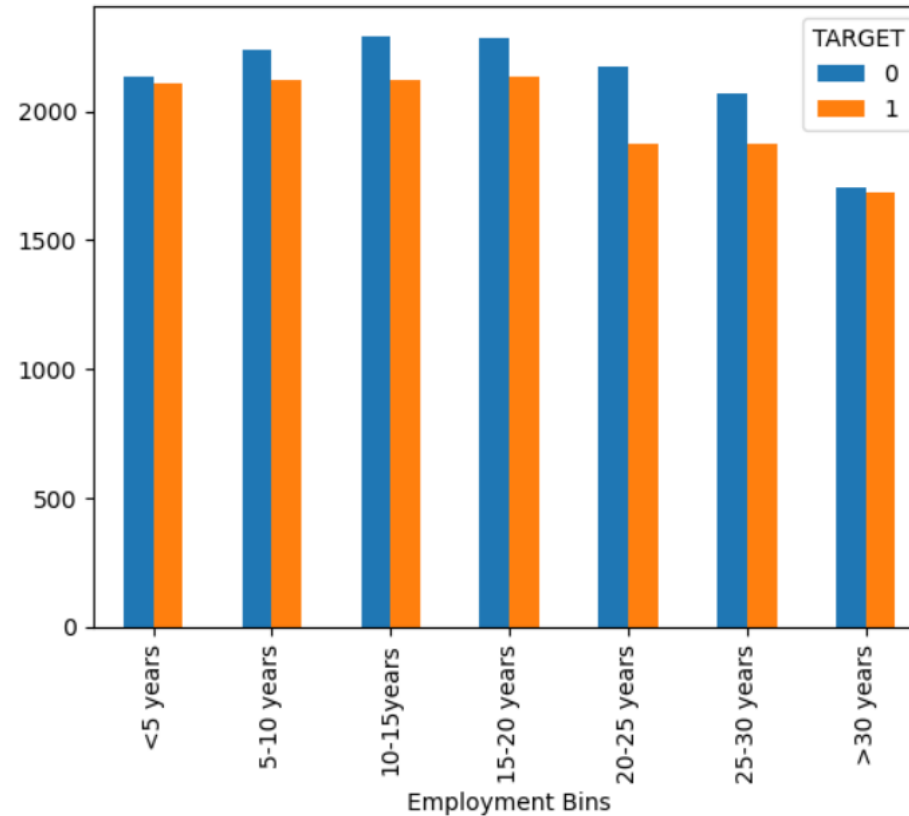


Income based on Employment experience

Applicants with 10-20 years employment exp have higher average income

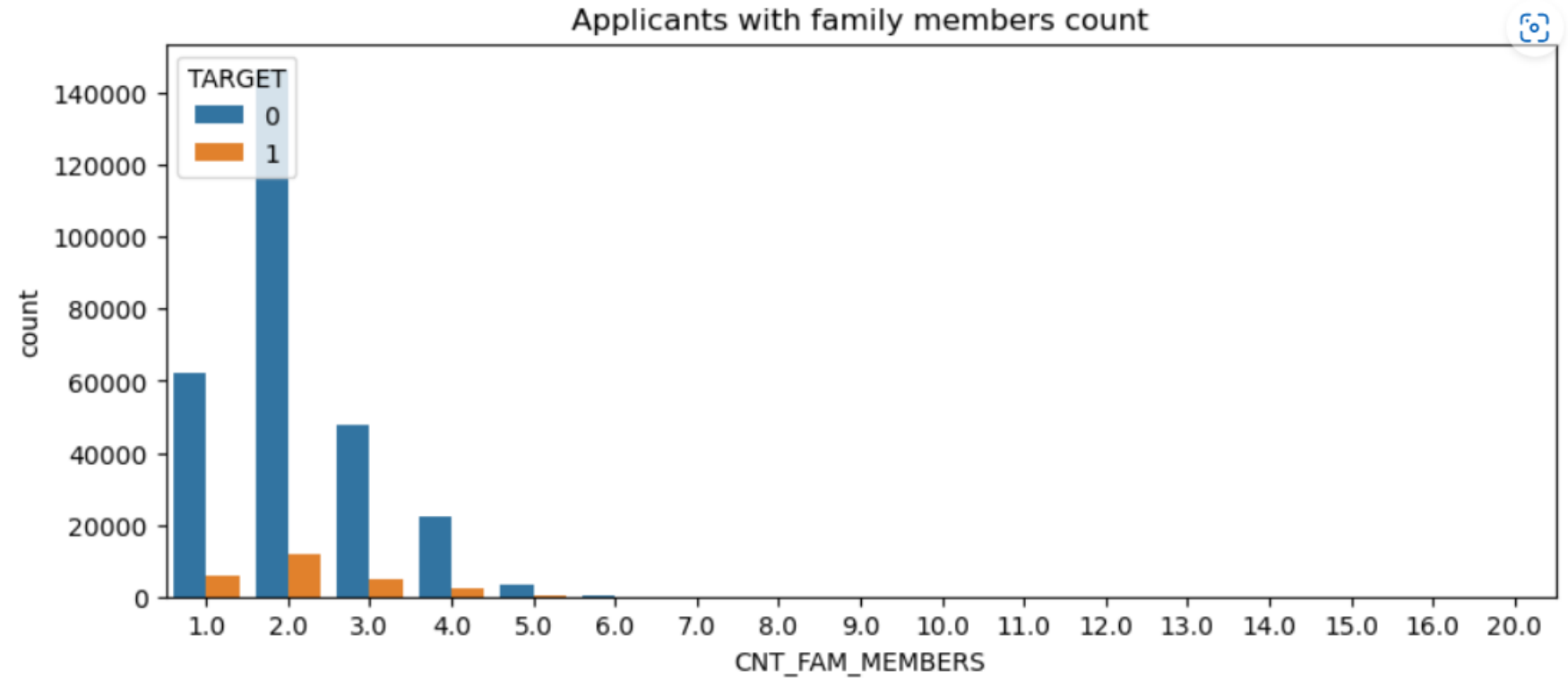
TARGET	0	1
Employment Bins		
<5 years	2138.143900	2108.326261
5-10 years	2241.962183	2122.698045
10-15 years	2293.141832	2124.039192
15-20 years	2286.622828	2133.535656
20-25 years	2174.992637	1875.590426
25-30 years	2067.654157	1871.482812
>30 years	1707.477346	1683.393722

Applicants with Higher Average Income based on Employment experience



Applicants with family member count

Applicants with 2 family members are not facing challenges in making payments. Could be just applicant and spouse in the family

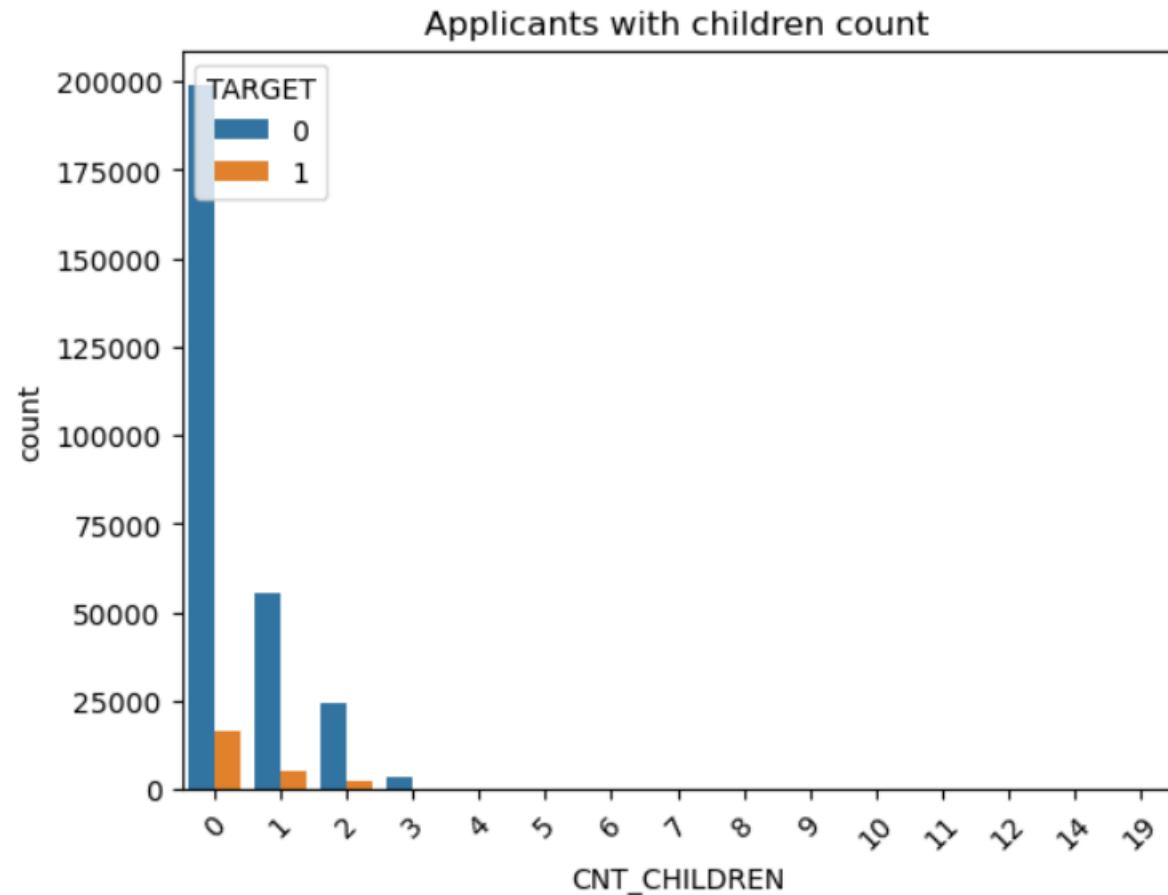


Applicants with children count

Applicants with zero kids are not facing challenges in payment.

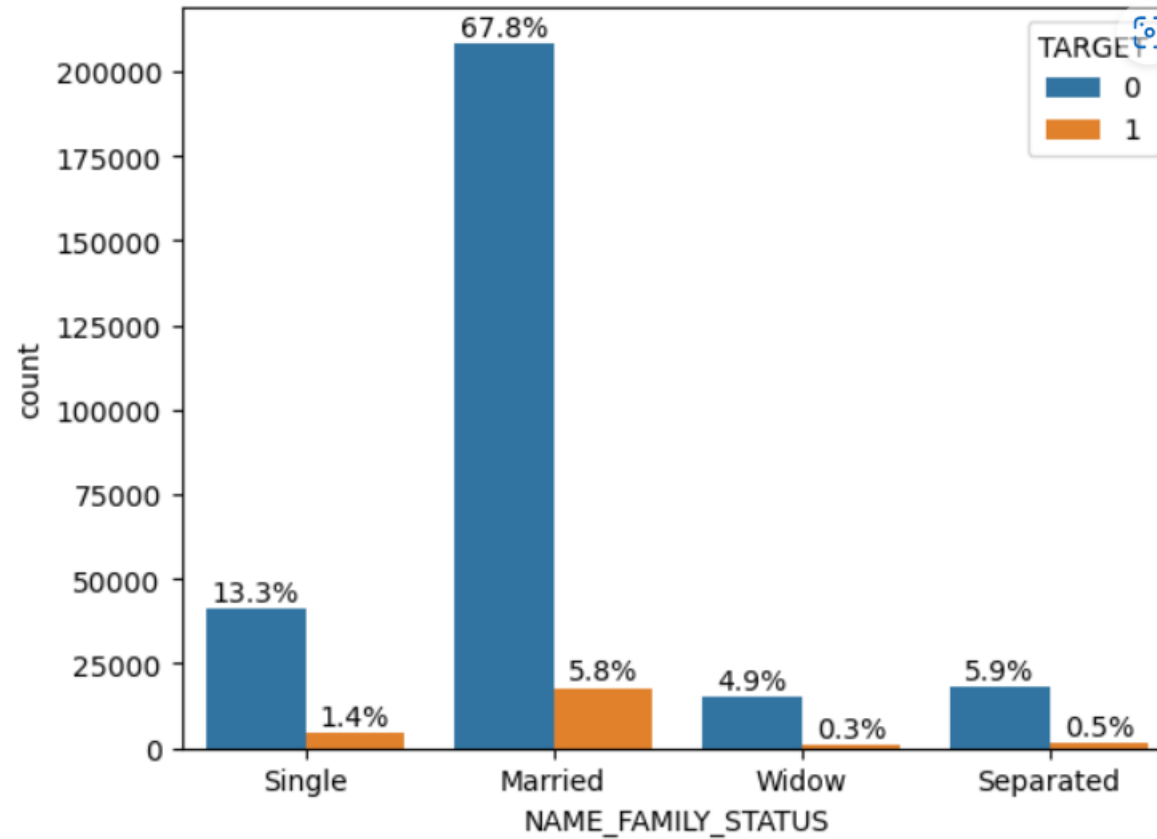
We may infer that most of these applicants are married and not having kids.

Or we may also infer that at the time of application, the applicants were married with no kids. Once kids were born the children status may not have been declared (Not sure)



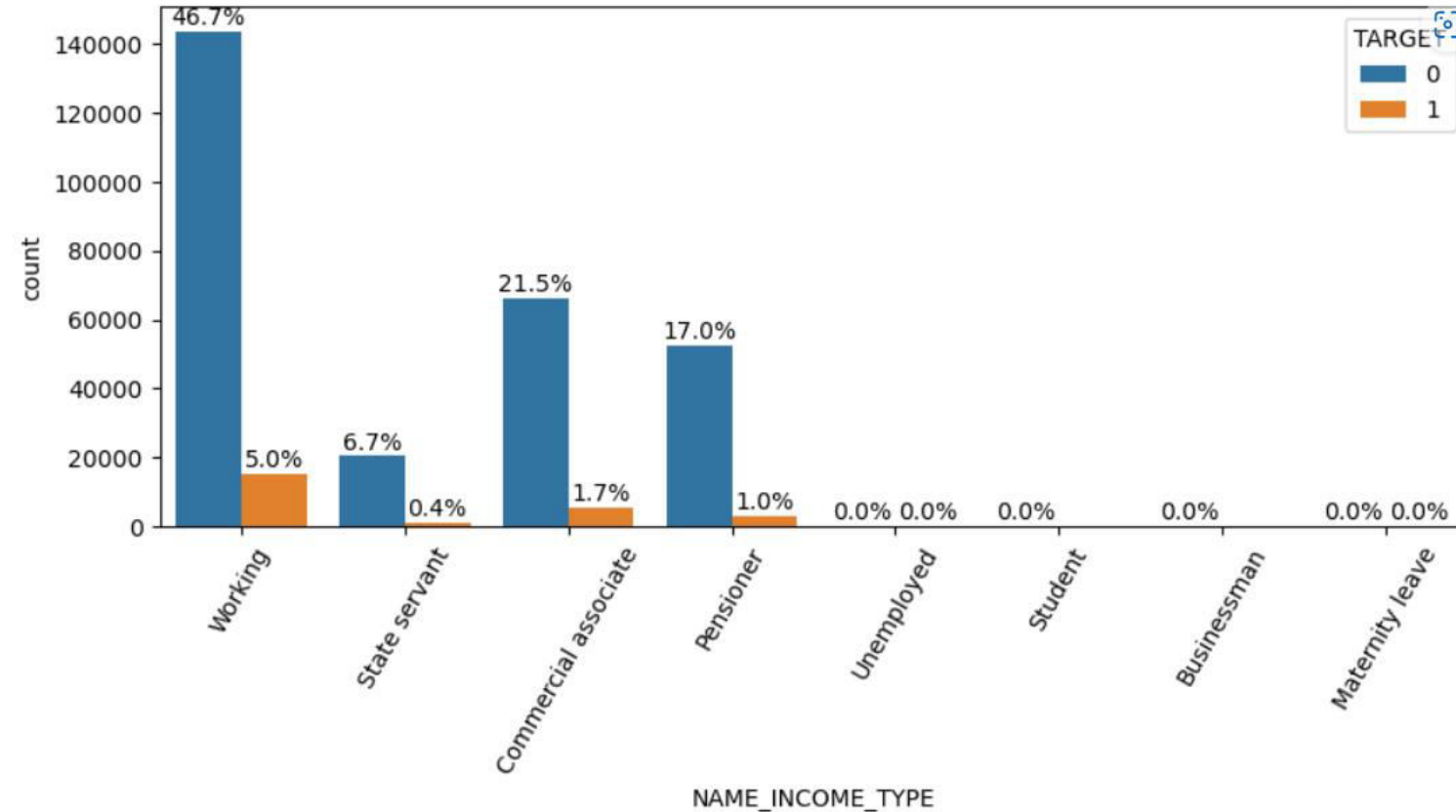
Marital status

Married people are better in paying off the loans. Married people would want to secure future



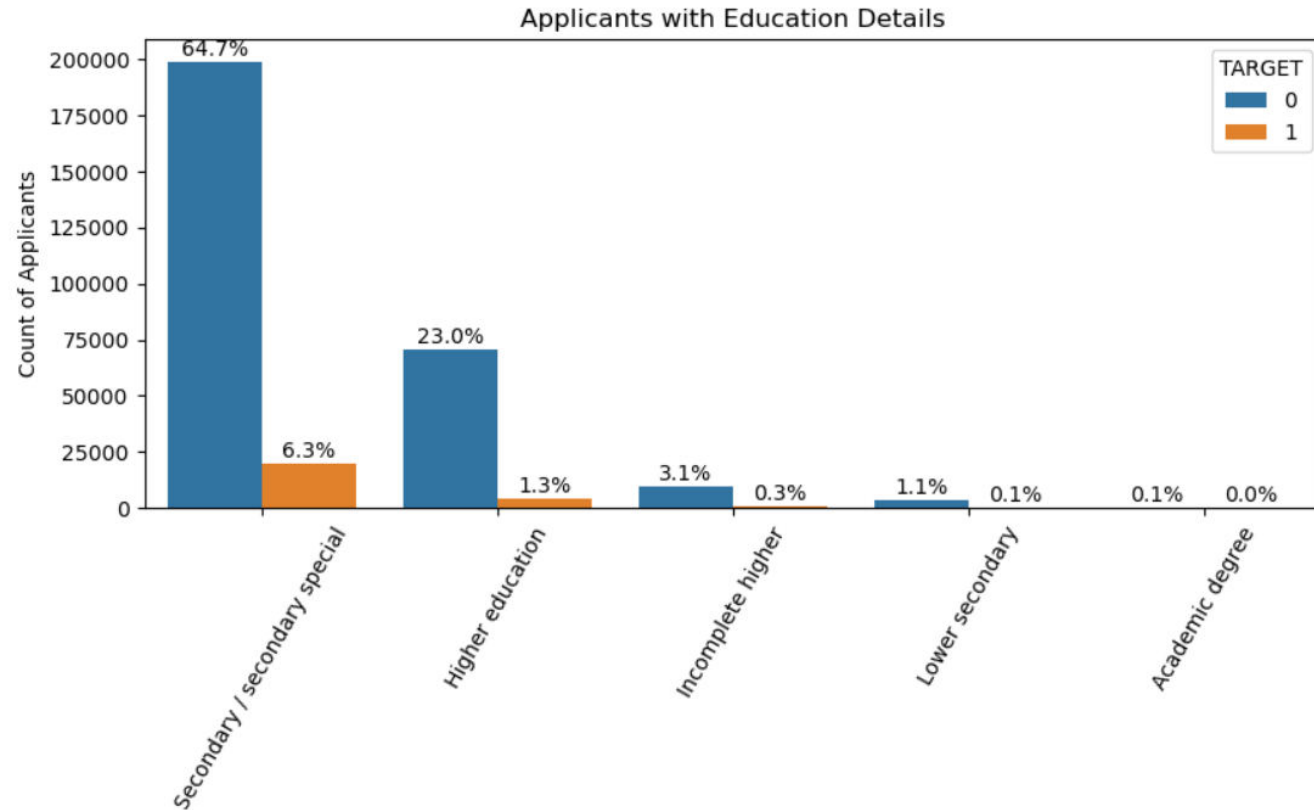
Income Type

Working Class Employees can be considered as the better applicants. They are better in paying off loans. Can also be attributed to the fact that they get monthly salary which is kind of fixed income as long as they are working



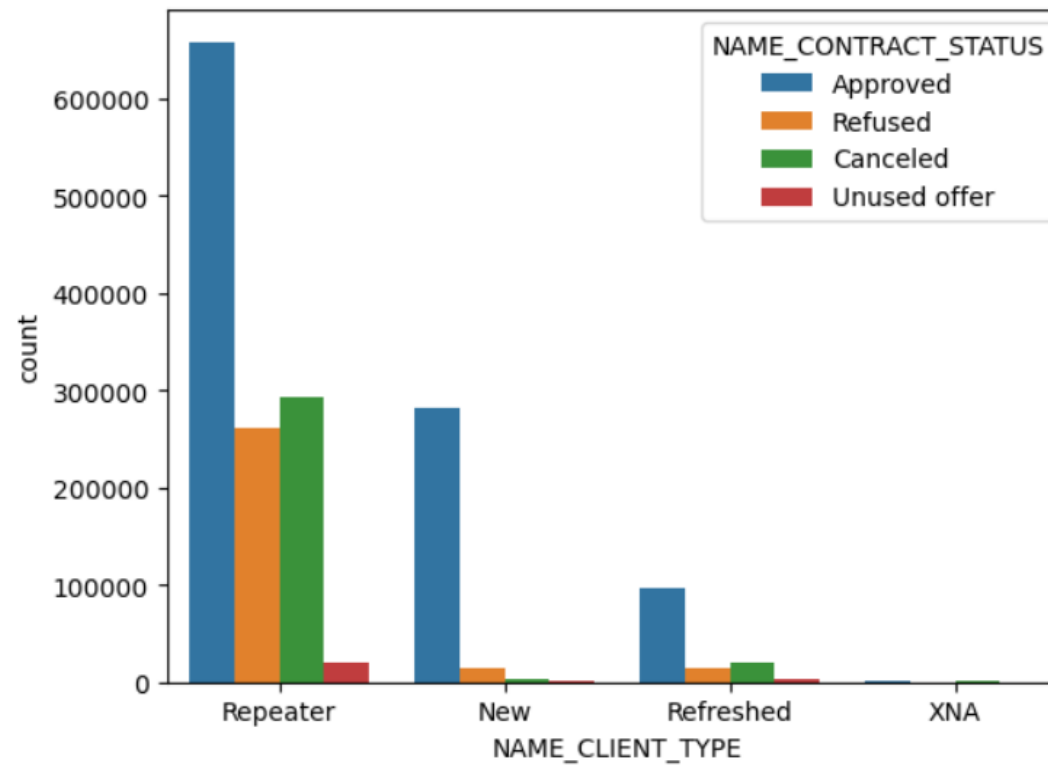
Education status

Secondary/ Secondary special applicants are not facing much challenges in payment. We may also infer that these people may be in stable jobs and would want to pay off loans and secure future.



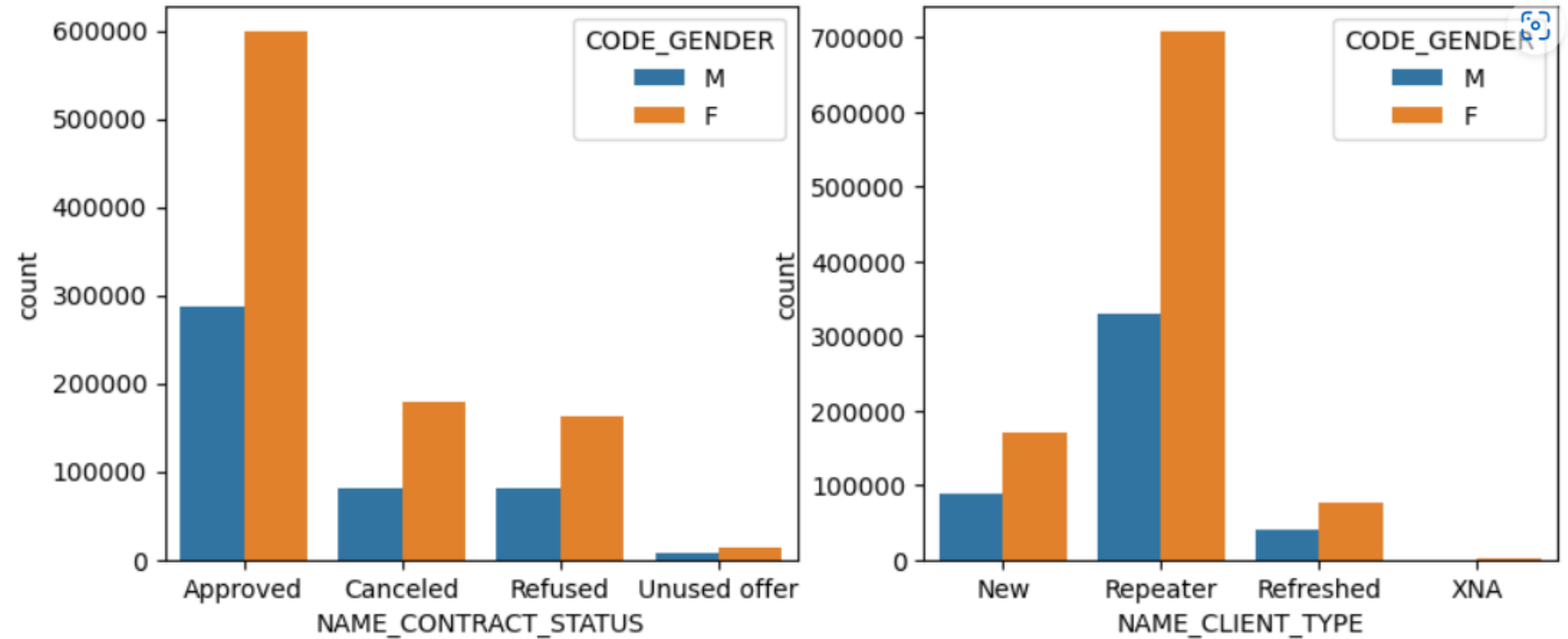
Client type and Contract Status

Most of the Applicant Applications who are REPEATERS have been approved



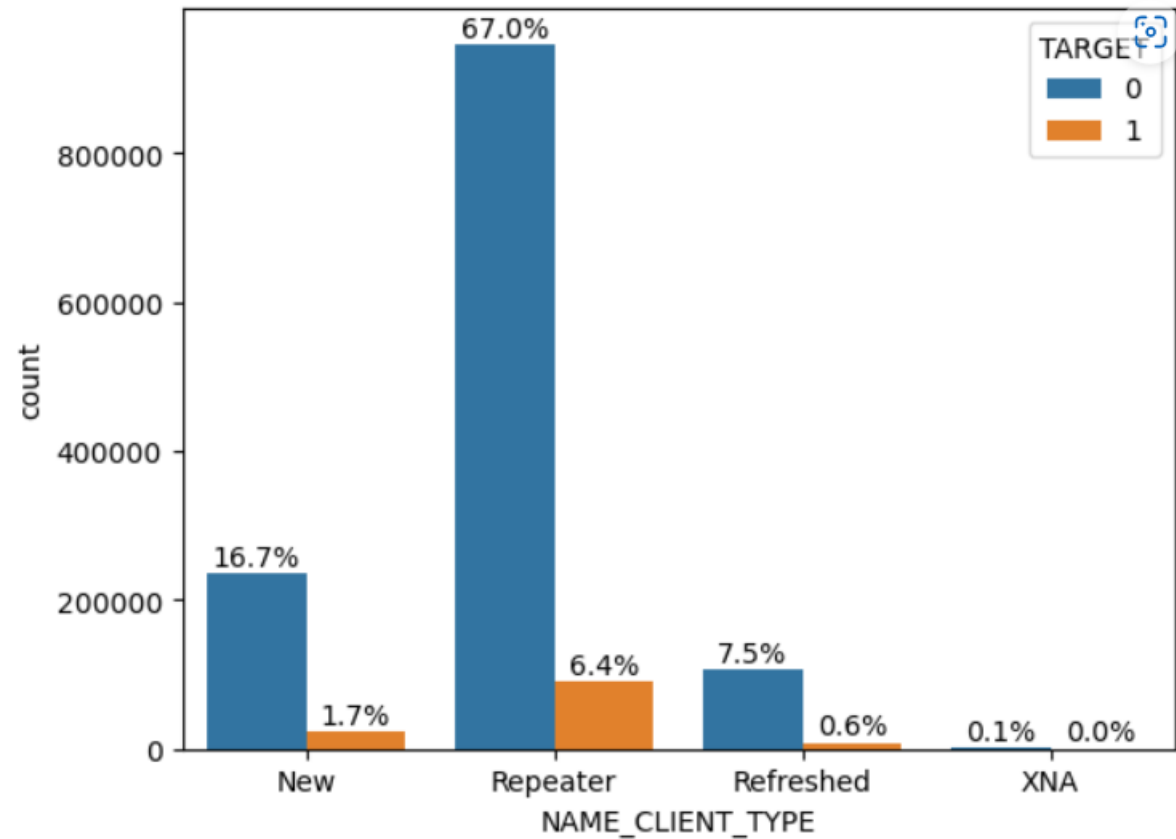
Gender based Contract Status

Female Applicants applications are more approved compared to males. Most of the applicants are Repeaters.

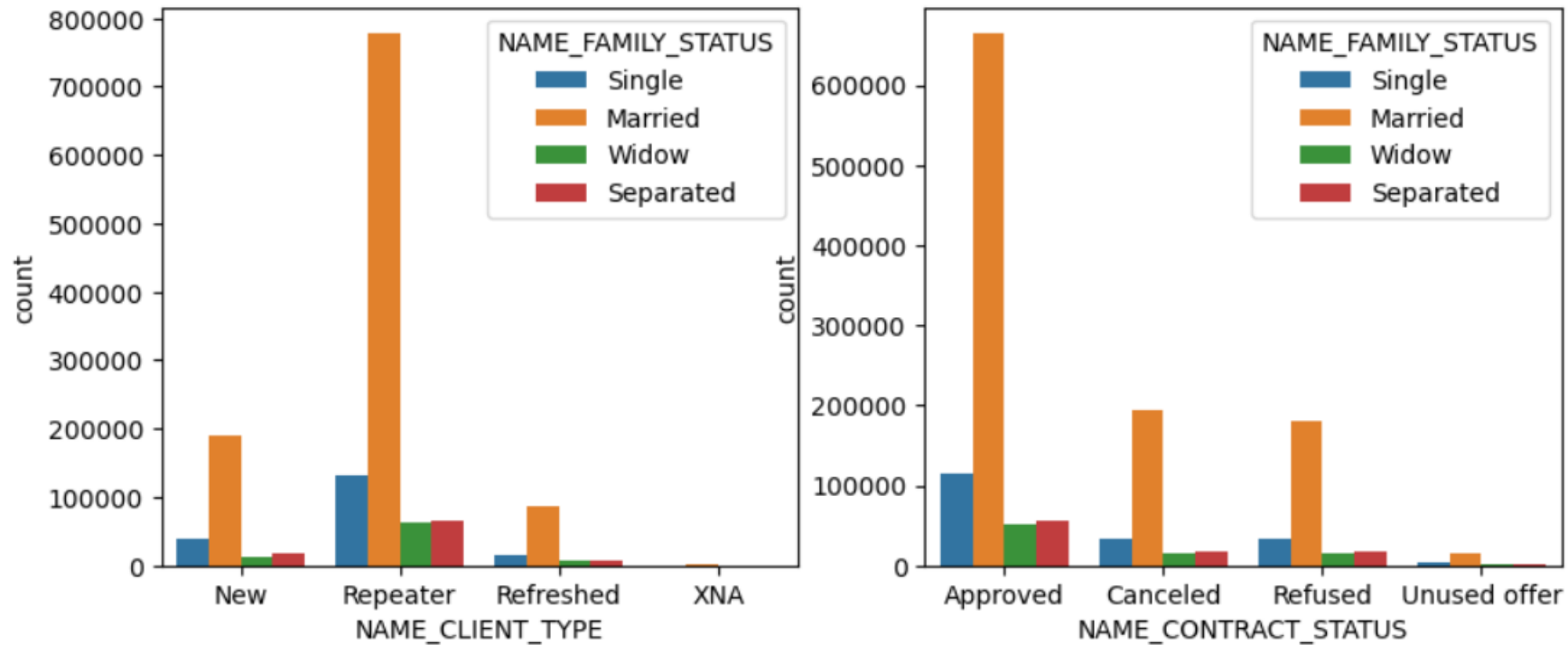


Which type of client applicants are provided loans and are good in making payments

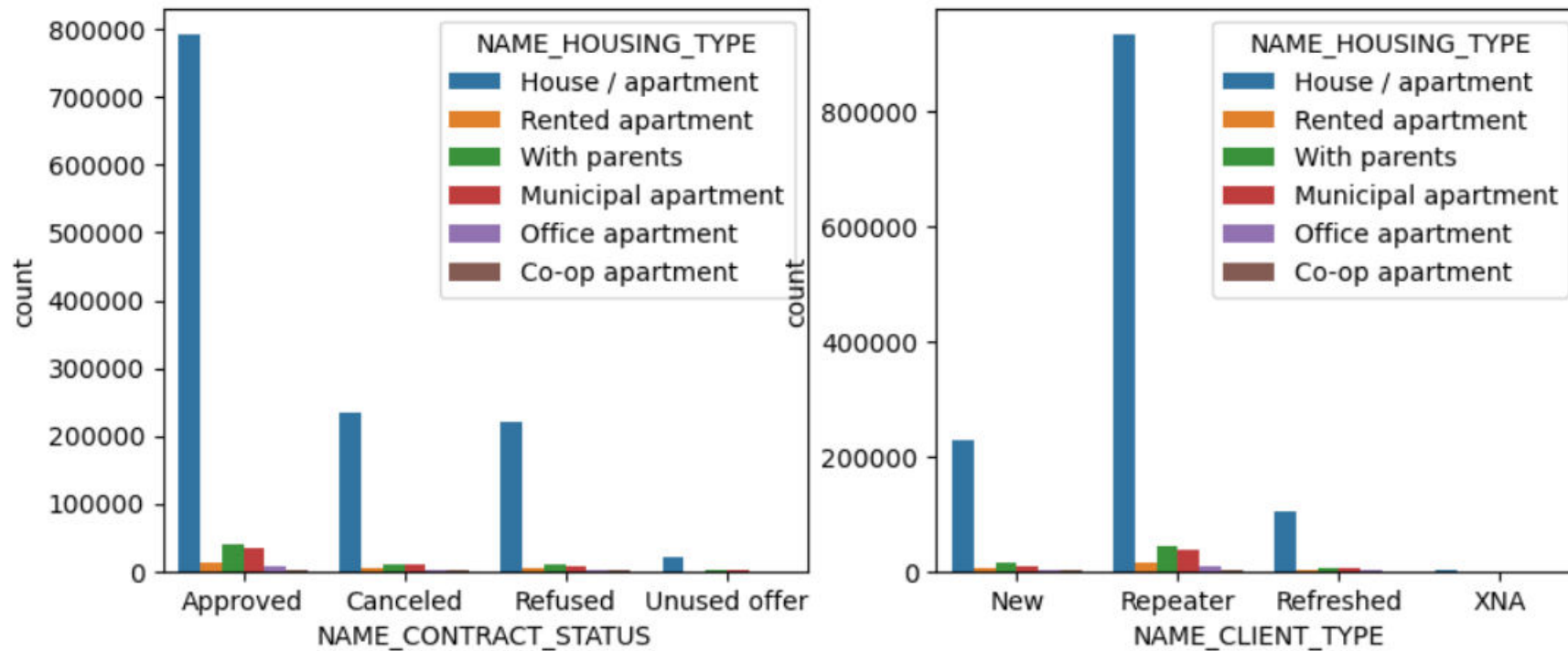
More number of Repeater Applications are approved and once approved they are diligent in making the payments



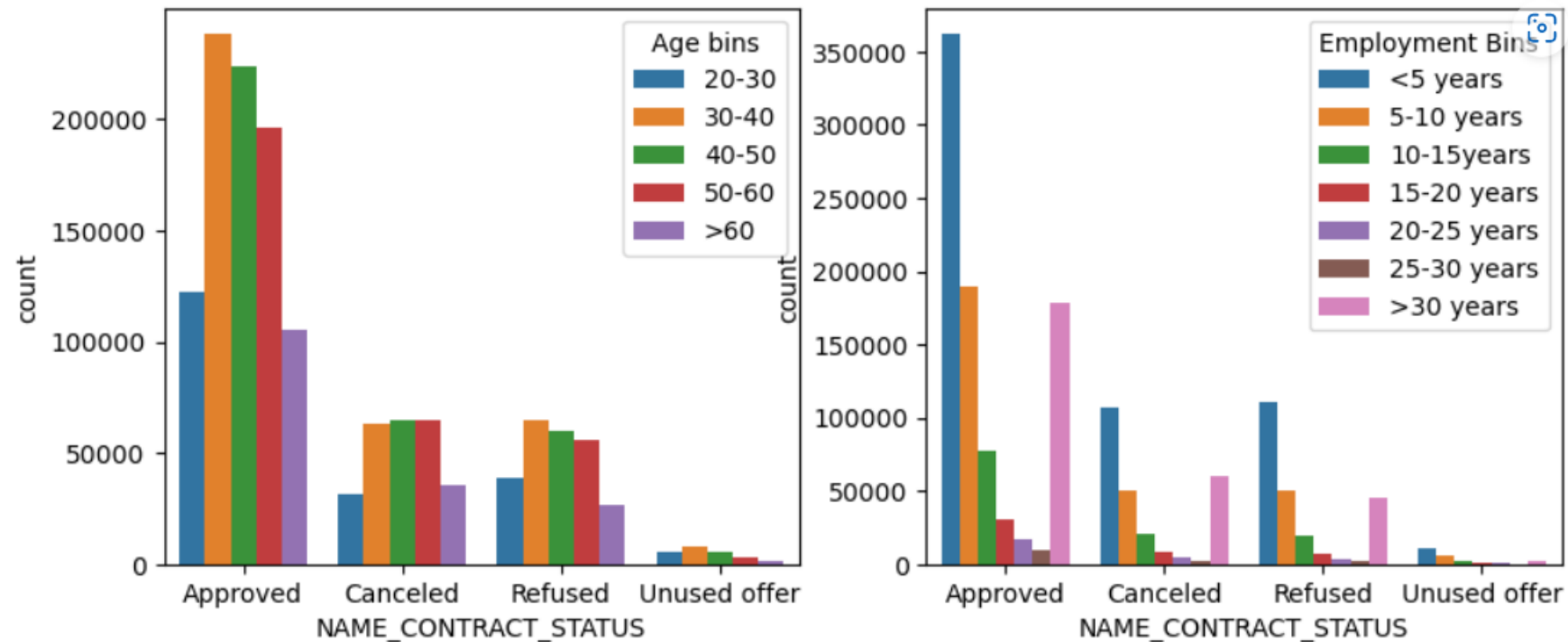
Married Repeater Applicants applications are mostly approved



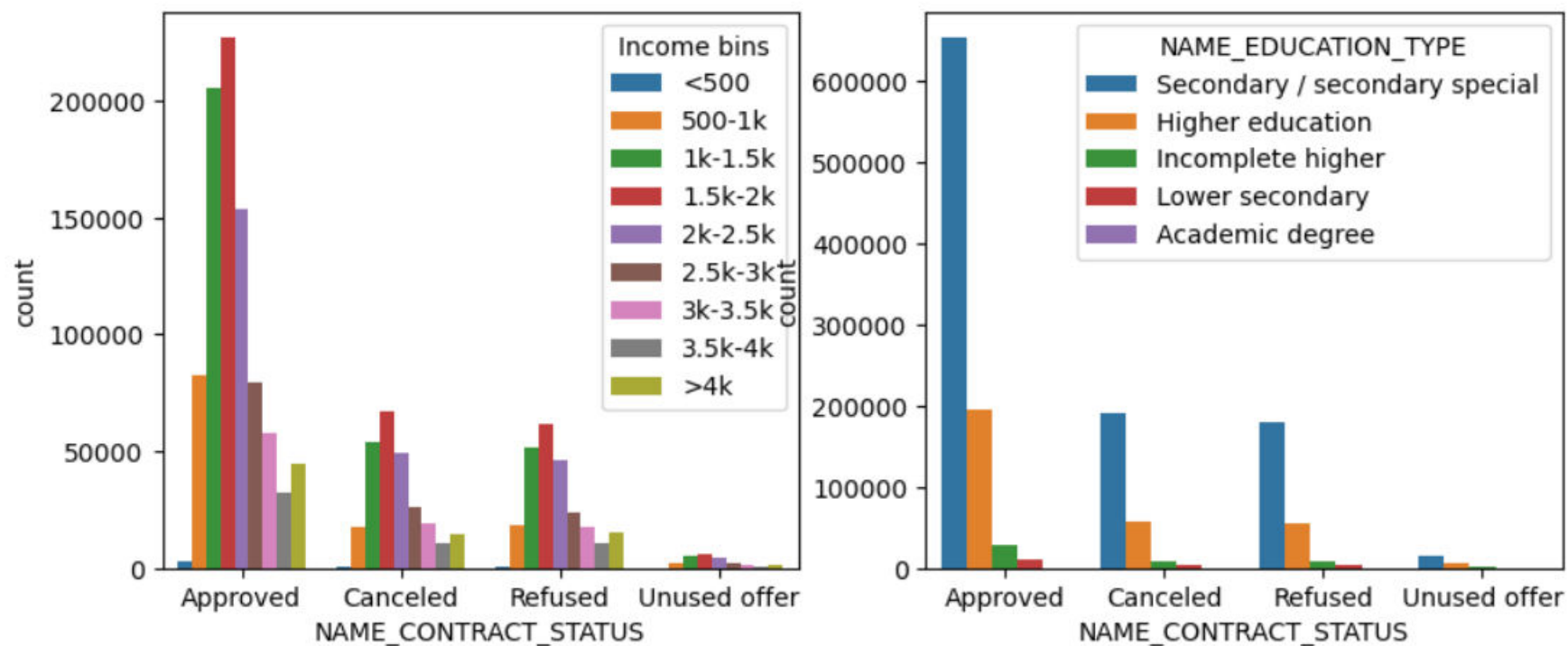
Applications of Applicants owning house/apartments are mostly approved



Applicants in age group 30-50 and employees with <5 year experience are approved.



Applicants with income range 1.5 – 2k dollars and having Secondary Education are mostly approved



Case Study Summary

1. Female applicants are better candidates. They invest more in real estate rather than depreciating assets like car. They seem to be more inclined towards securing the future. Candidates owning house are more likely to get the loan approved.
2. Working class applicants are the people who are diligent in making payments.
3. Married applicants with family member count of 2 are better candidates. Less financial burden probably.
4. Applicants with 1 or less children are better candidates.
5. Candidates in the age range of 30 – 50 years and having work experience of 10 – 20 years are better candidates. These people also have higher average income. And female population having an average age of 45 years are eligible for a good applicant status.
6. Applicants with an average income of 1.5 -2k dollars are more diligent in paying loans.
7. Applicants having <5 year experience are also more in number and these applications are also mostly approved.
8. Applicants having Secondary/Secondary Special Education are able to secure stable Working Type monthly jobs and better candidates in paying off loans
9. Applicants who are Repeaters are better candidates to whom loan can be provided.