

CREDIT EDA CASE STUDY

BY: JYOTSNA SHUKLA & SYEDA SHEEBA NYER

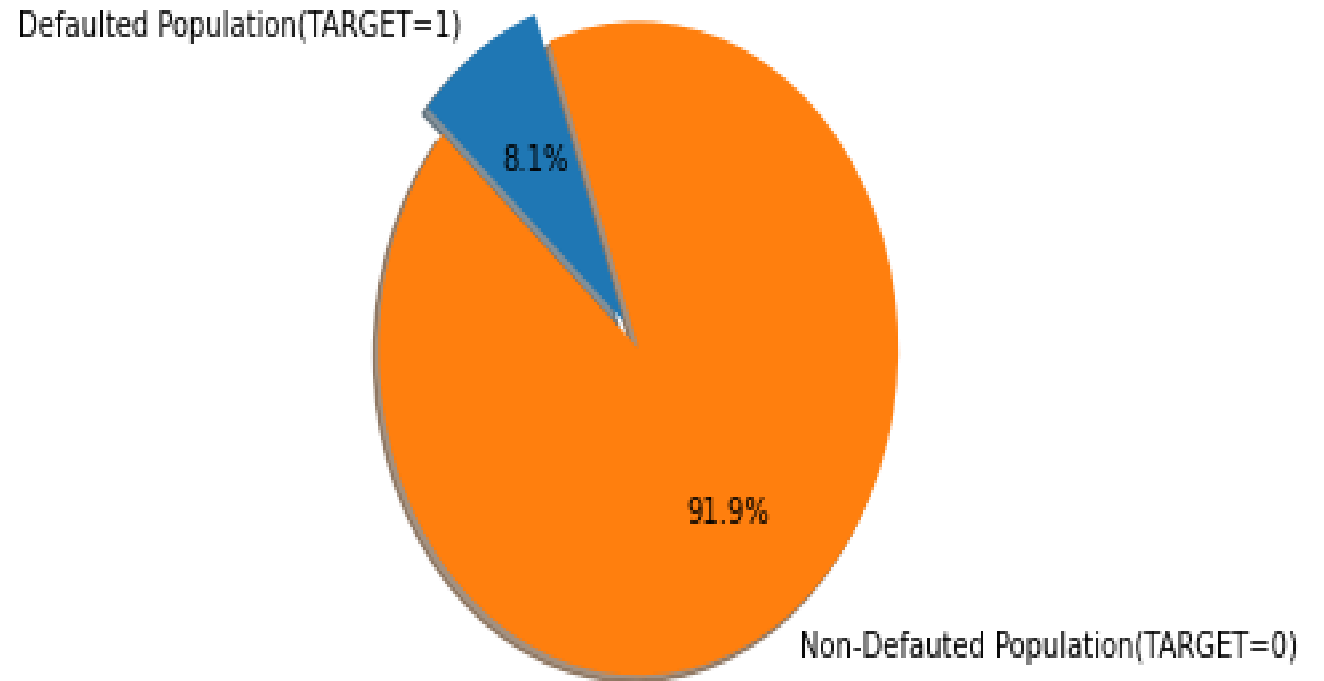
A solid orange horizontal bar spanning the width of the slide at the bottom.

Application Data

Imbalance Ratio:

We can see that the Application data has high imbalance with Defaulted population at 8.1% as compared to Non-defaulted population at 91.9% . The Imbalance ratio calculated is 11.39.

Data imbalance

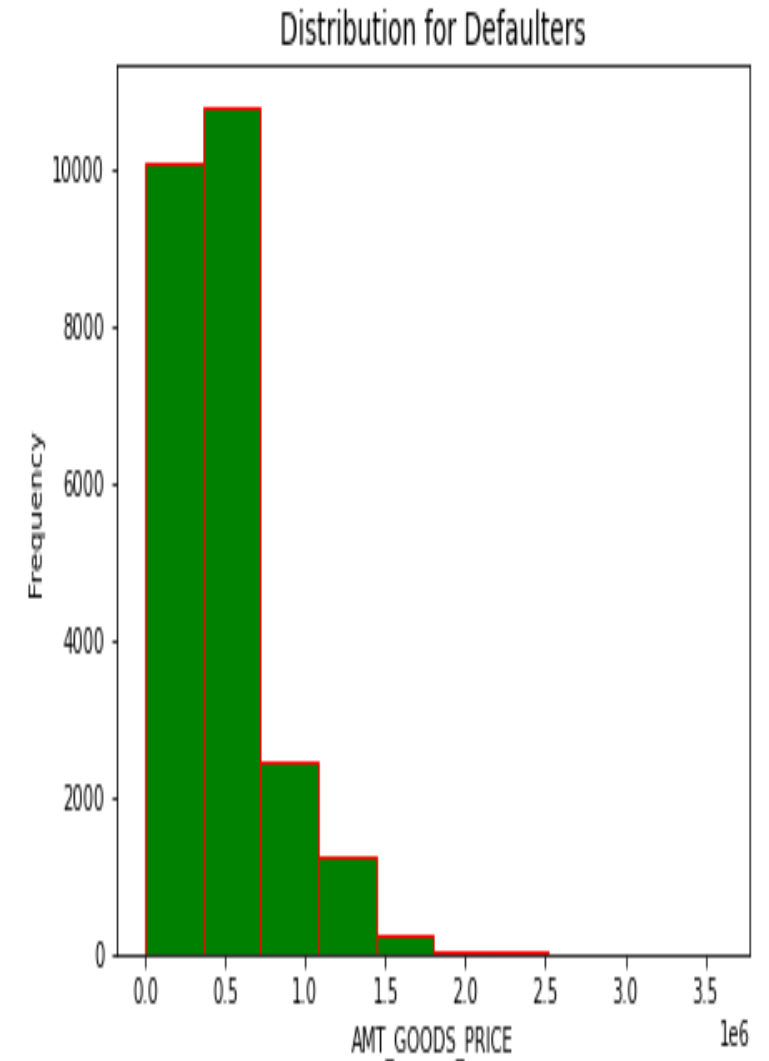
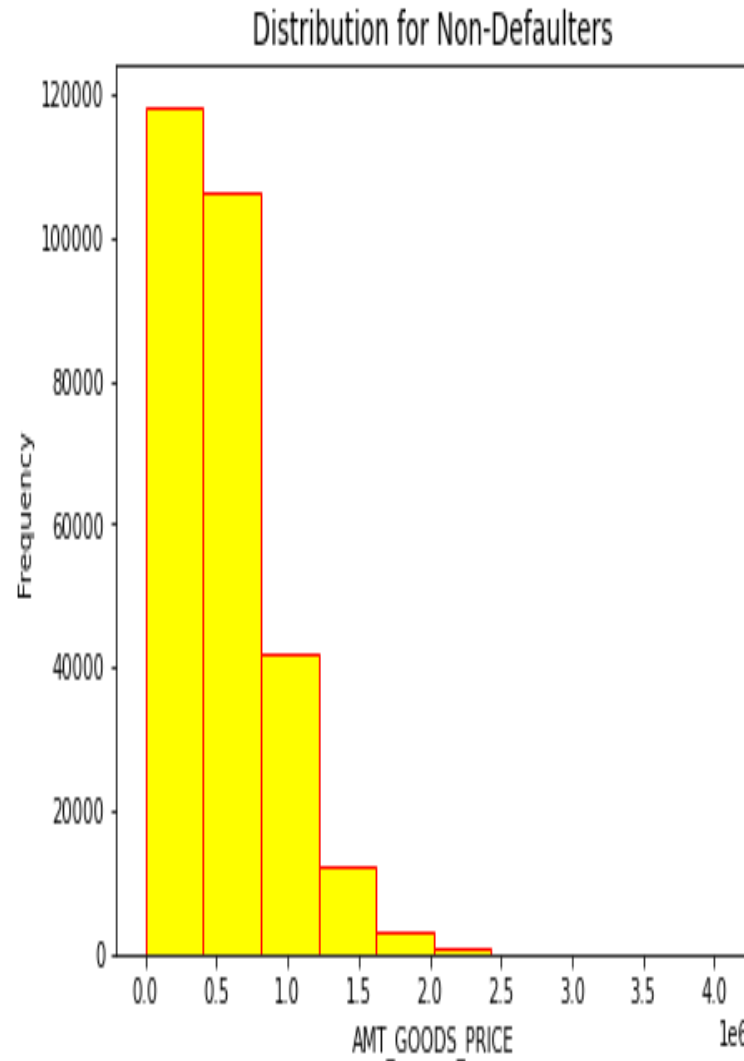


Univariate Continuous Analysis (wrt Target 0 & 1)

AMT_GOODS_PRICE:

Points to be concluded from the plot:

Defaults are higher for price of the goods for which the loan given is lesser (between 0 to 500,000)

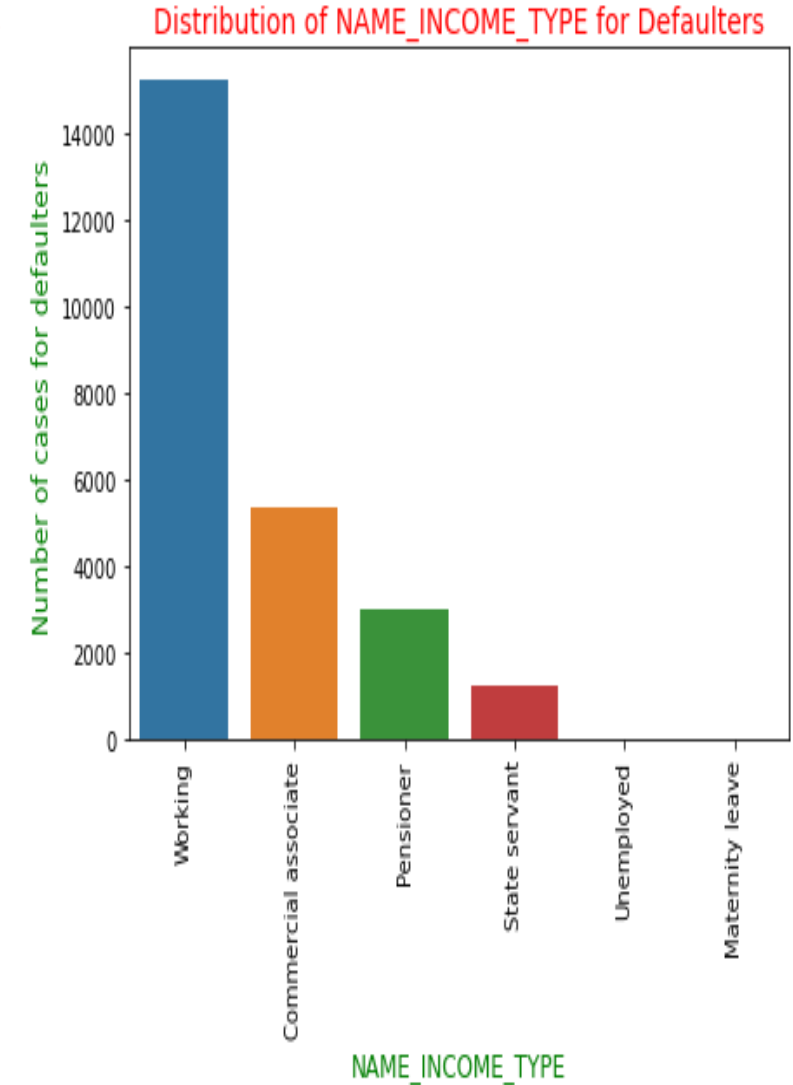
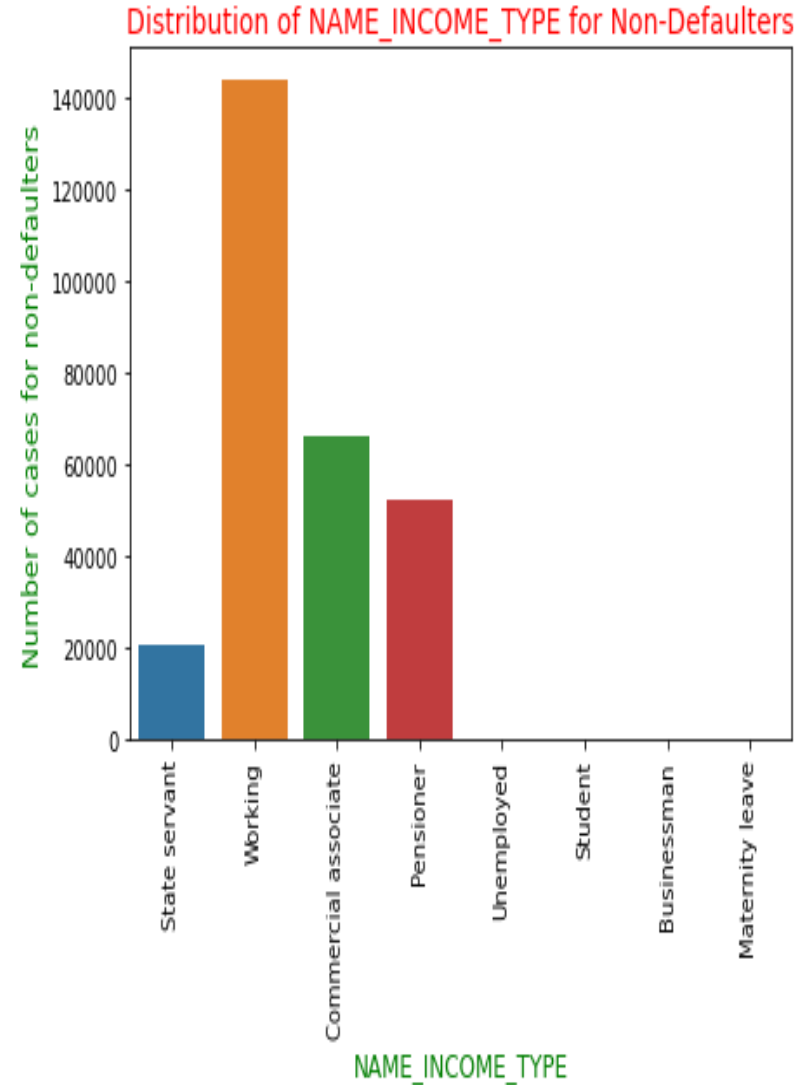


Univariate Categorical Analysis

NAME_INCOME_TYPE:

Points to be concluded from the plot:

We can notice that most of the defaults are working professionals and the least are students, unemployed, businessman & maternity leave.



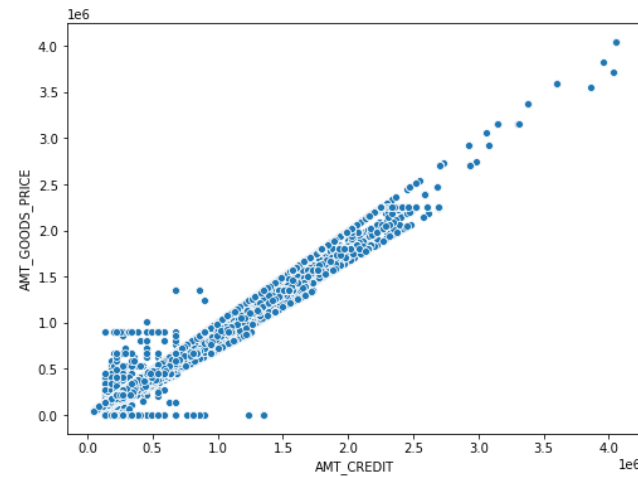
Bivariate Cont-Cont Analysis

AMT_CREDIT VS AMT_GOODS_PRICE:

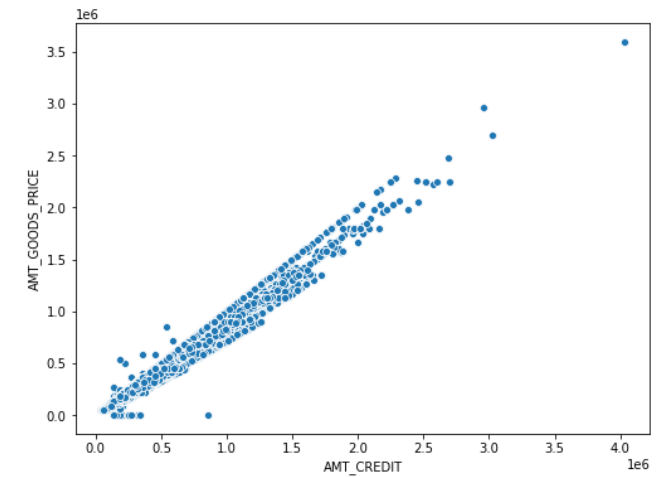
Points to be noted from the plot:

we can notice that defaulters are less for both Goods price & Amount Credit less than 500k.

CREDIT vs GOODS PRICE for Non-Defaulters



CREDIT vs GOODS PRICE for Defaulters

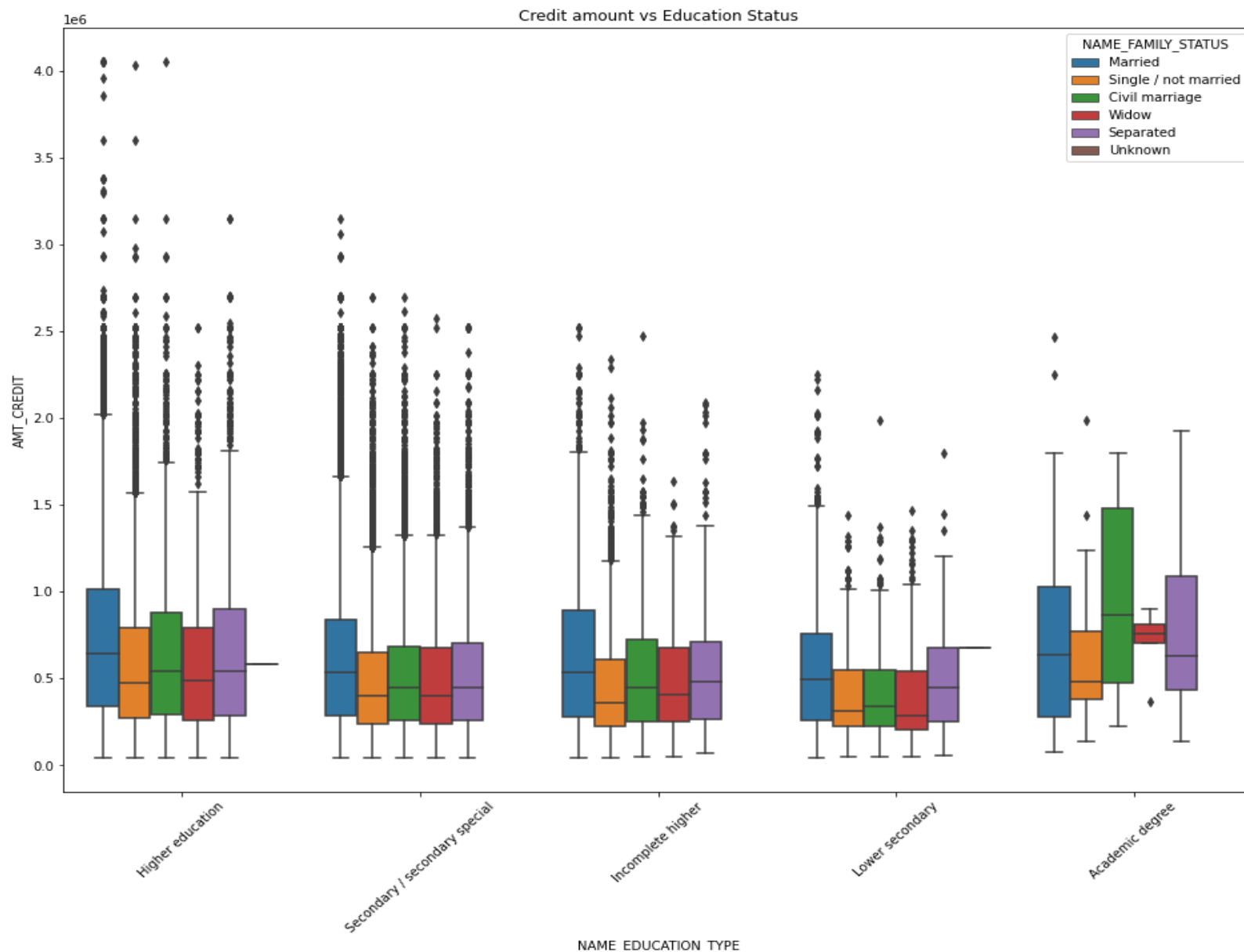


Bivariate Cont-Cat Analysis

CREDIT AMOUNT VS EDUCATION STATUS:

Points to be noted from the boxplot:

Family status of 'civil marriage', 'marriage' and 'separated' of 'Academic degree education' are having higher number of credits than others. Also, 'higher education' of family status of 'marriage', 'single' and 'civil marriage' are having more outliers. 'Civil marriage' for 'Academic degree' is having most of the credits in the third quartile.



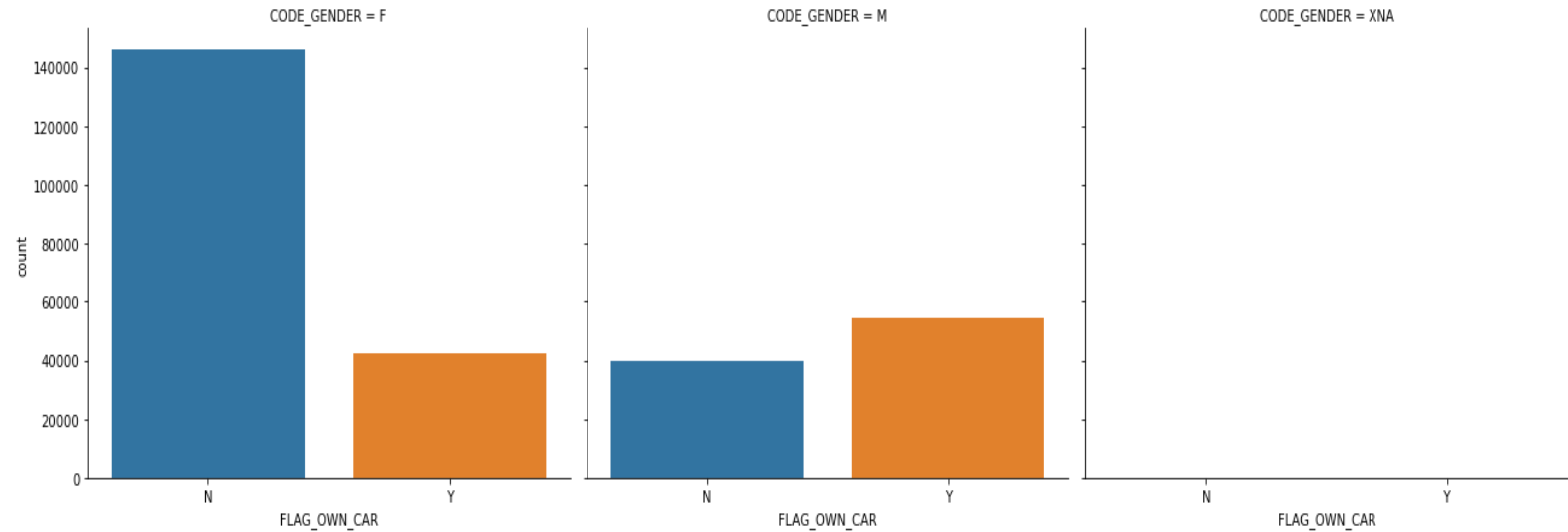
Bivariate Cat-Cat Analysis

FLAG OWN CAR VS CODE GENDER:

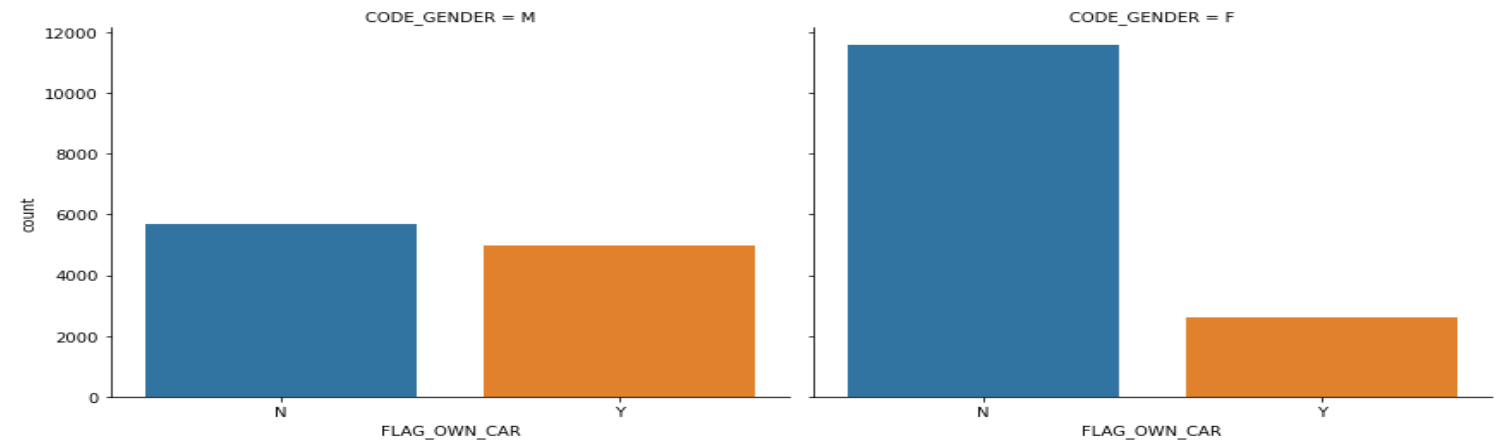
Points to be noted from the plot:

Males have higher default rate if they don't own car.

For non-defaulters:



For defaulters:



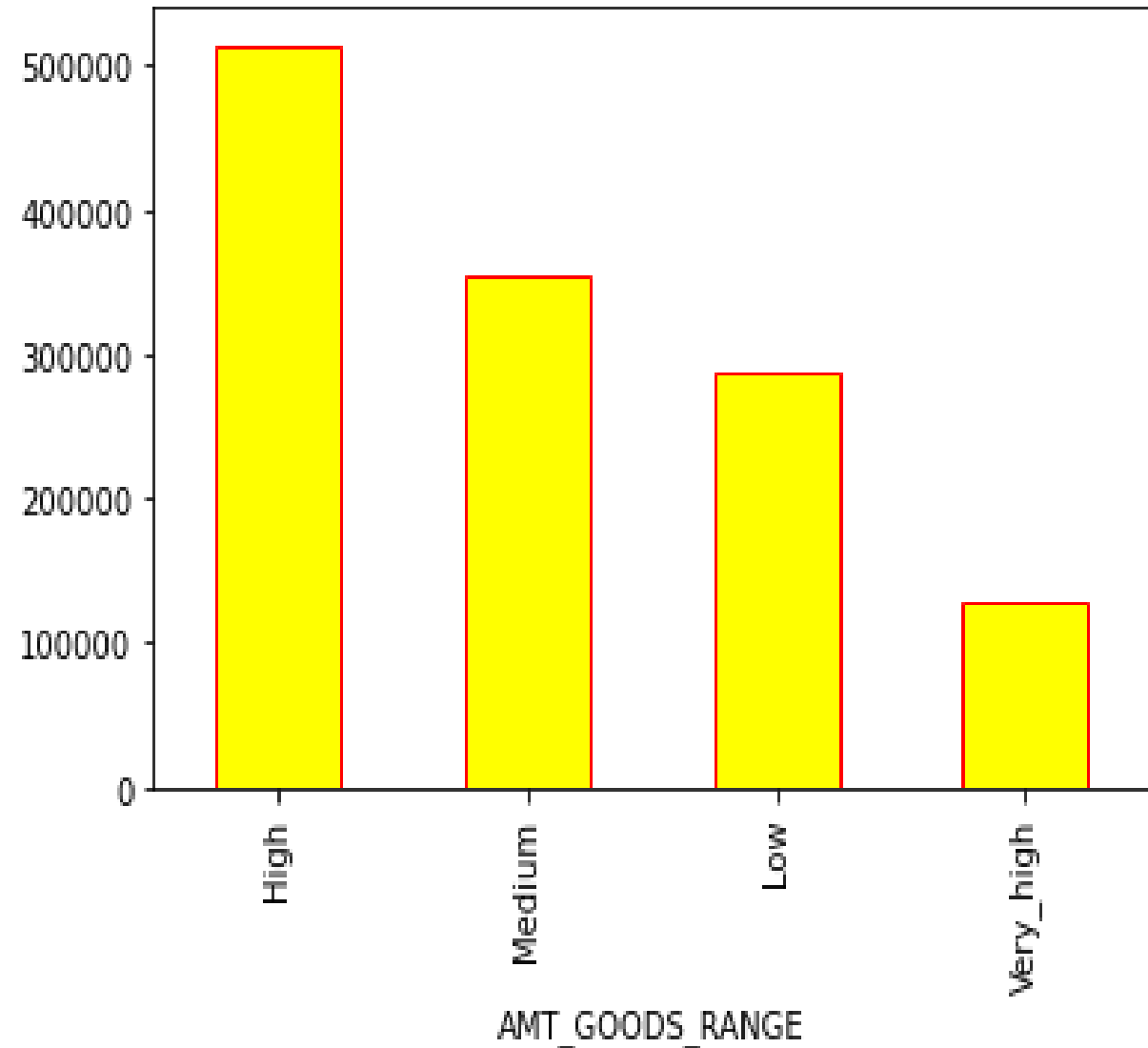
Previous Application

Univariate Continuous Analysis

AMT_GOODS_PRICE:

Points to be noted from the plot:

The mean of goods price is higher than its median. It means, in aggregate for all loans, frequency of higher goods price is more than lower ones.

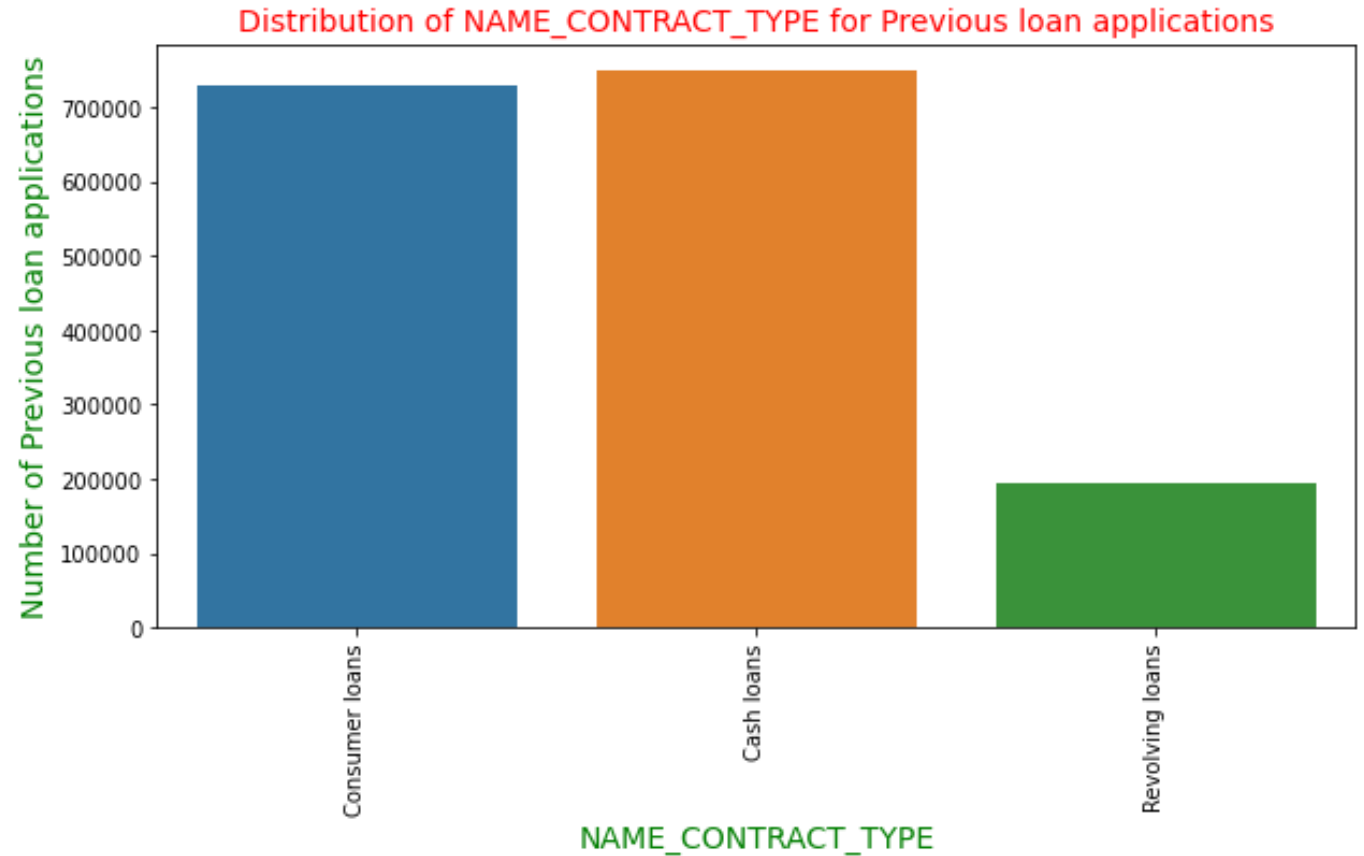


Univariate Categorical Analysis:

NAME CONTRACT TYPE:

Points to be noted from the plot:

Of all the previous loan applications, the revolving loans are minimal, while consumer and cash loans are comparable.

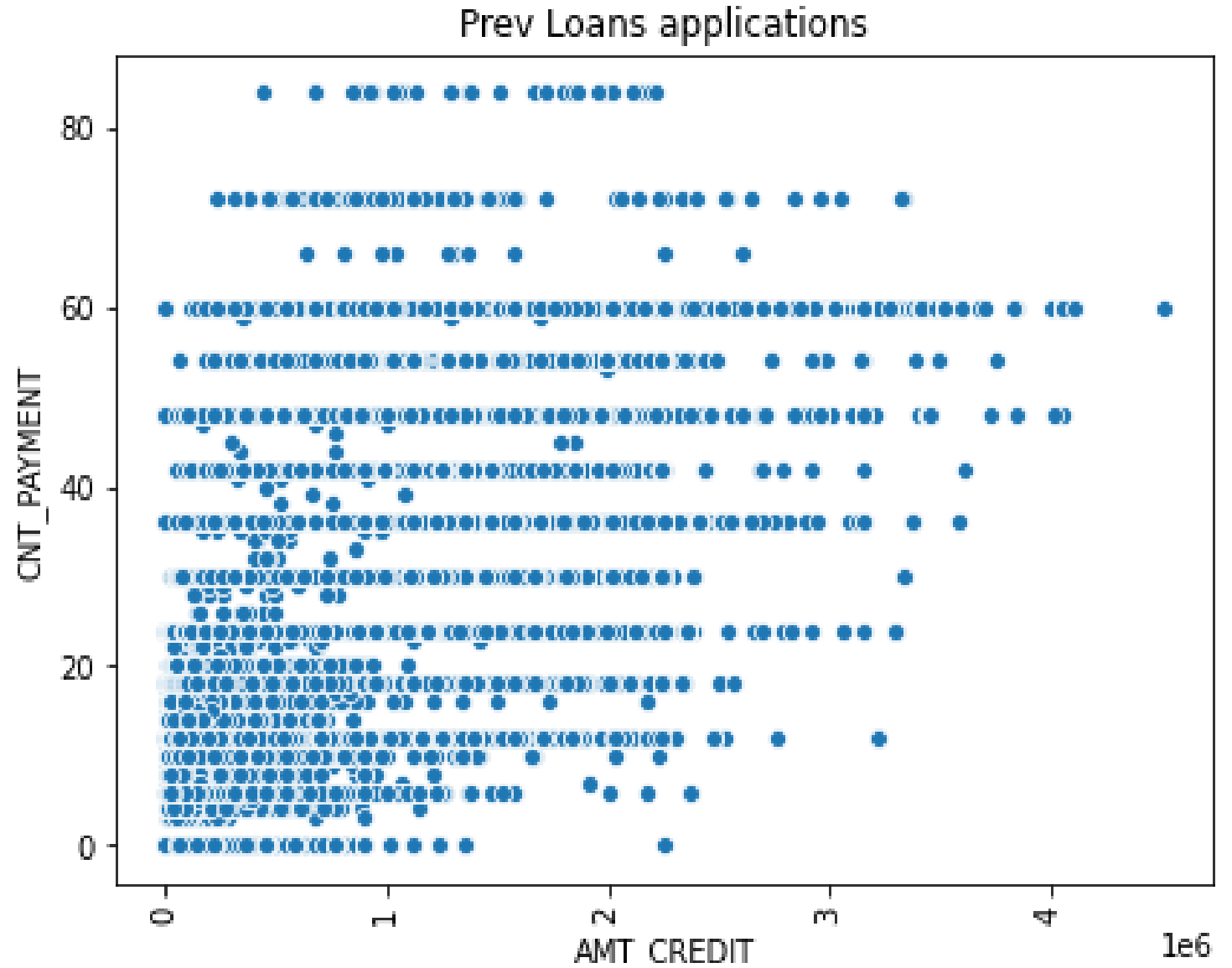


Bivariate Cont-Cont Analysis:

AMT CREDIT VS CNT PAYMENT:

Points to be noted from the plot:

There is no or very little relation between payments counts and credit amount for Previous loan applications.

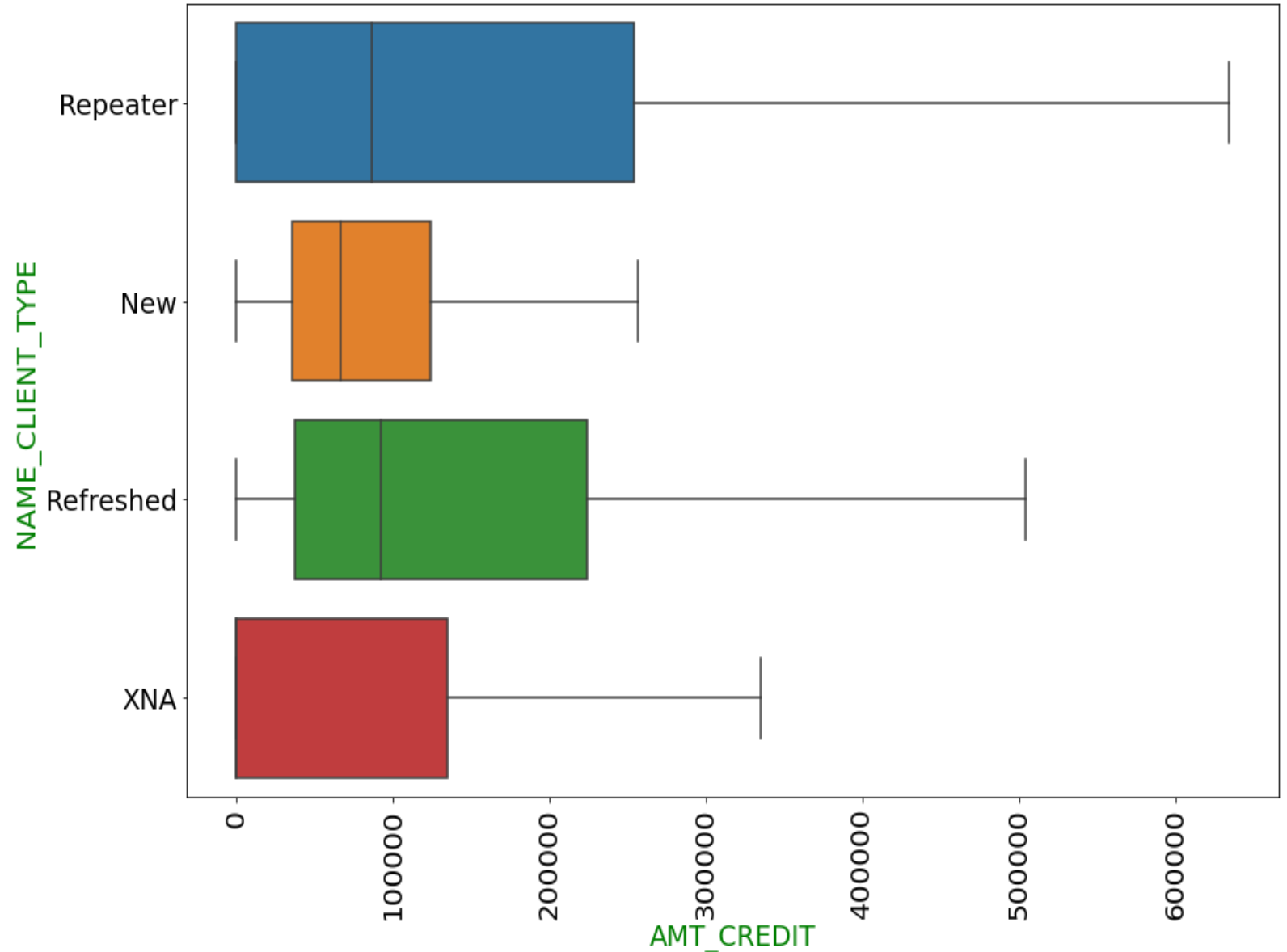


Bivariate Cont-Cat Analysis:

AMT_CREDIT VS NAME_CLIENT_TYPE:

Points to be noted from the plot:

Repeaters and refreshed client types have higher Credited amount and spread than new clients.

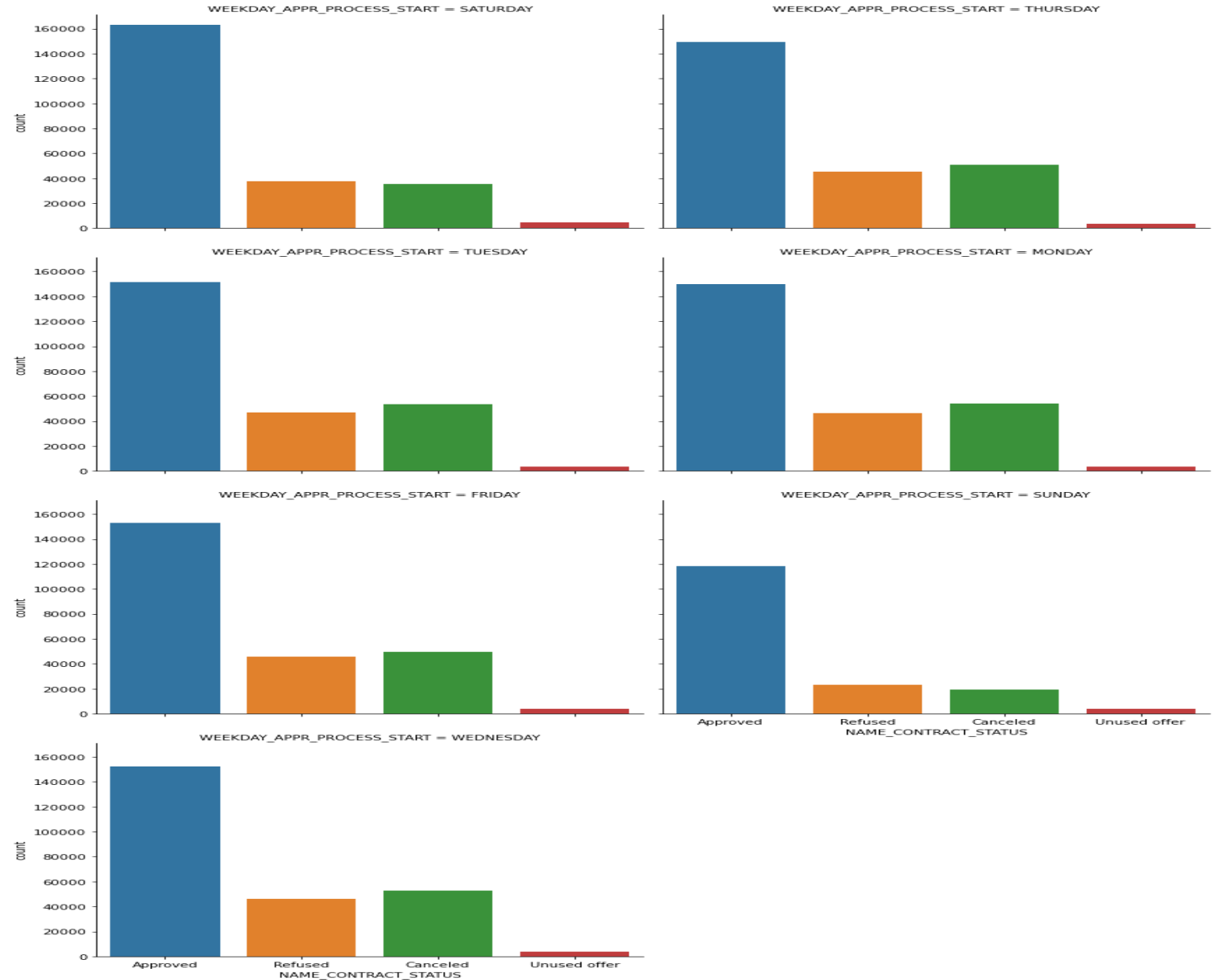


Bivariate Cat-Cat Analysis:

NAME CONTRACT STATUS VS WEEKDAY APPR PROCESS START:

Points to be noted from the plot:

The applications filed on
Sundays have lower cancellation
and Refusal rate in comparison to
other days!!!



CONCLUSION

- In previous application data, maximum population have chosen mid-range Payment counts, while lower numbers opted for very high payment counts. Also, the clients who cancelled the loans have higher average annuity by approximately 20K amount than approved ones. Thus, the bank should design loans so that loan amount having lower annuity and payment counts for better success.
- Of all the previous loan applications, the number of revolving loans are minimal, while numbers of consumer and cash loans are comparable. Bank should benchmark this with other banks, and take actions to increase revolving loans.
- Most of the clients are repeaters i.e. they are loyal clients and are using bank loan services again. Hence, bank should target current clients more aggressively for better return on investments on marketing.

- The maximum number of successful applications is through Country-wide channel followed by "Credit and Cash Offices" and "Stone" channels. Also, the Credit and Cash channel has a higher ratio of Cancellation and Refusal. Bank should look at how the conversion ratio of Credit and Cash offices could be improved through training its employees and better targeting. Bank also should invest more in its country wide channel as that leads to maximum conversion.
- In application data, Revolving loans are lesser in the defaulted population. Hence we can infer that revolving loans are comparatively safer. Bank should target clients requiring loans of this type.
- Banks should consider workings as they are applying for loans more in comparison to others.