



CREDIT EDA CASE STUDY

By Alfiya C j



Problem Statement

- **There are some bank data sets provided.**
- **Based on conditions, study and analyze the given data set.**
- **Understand and extract information from the bank data set provided and arrive at feasible conditions for risk management and profit enhancement by learning and modeling the patterns.**[1](#)

Exploratory Data Analysis

- **EDA is an analytics process to understand the data in depth and learn the different data characteristics, often with visual means.**

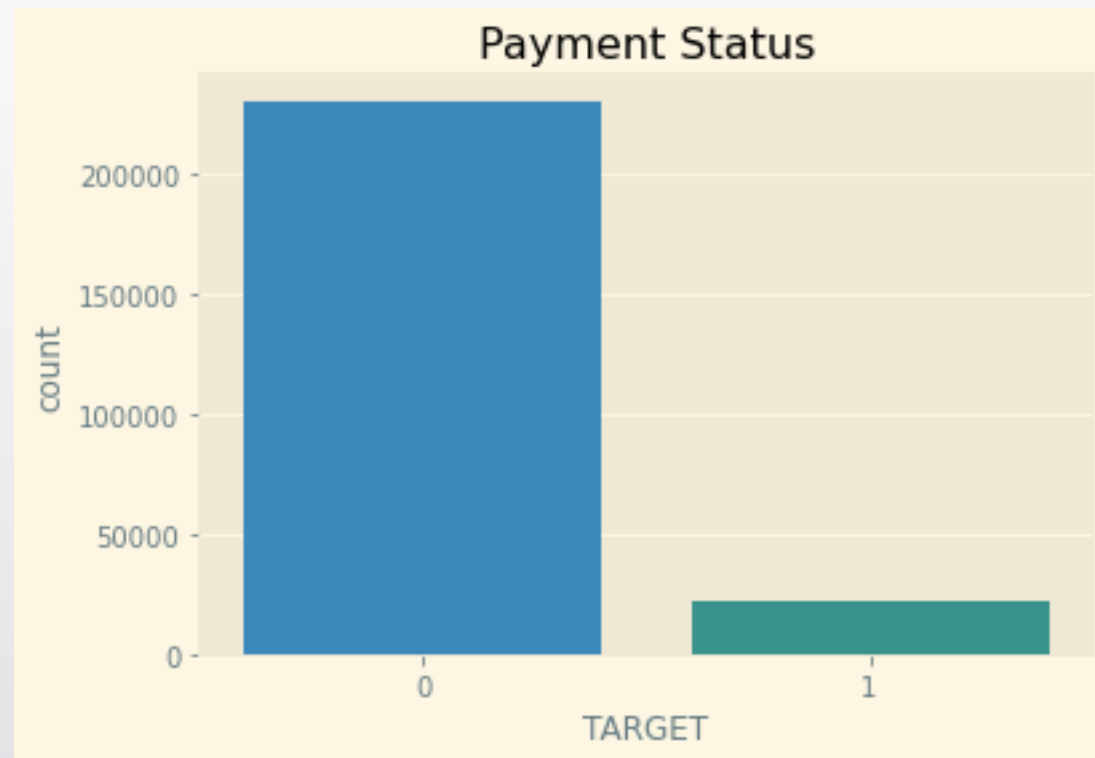
Steps Involved:

- Data collection
- Data cleaning
- Univariate Analysis
- Bivariate Analysis

Target imbalance

Observations:

- There is an imbalance between people who defaulted and who didn't default. More than 91% of people didn't default and 8% defaulted.



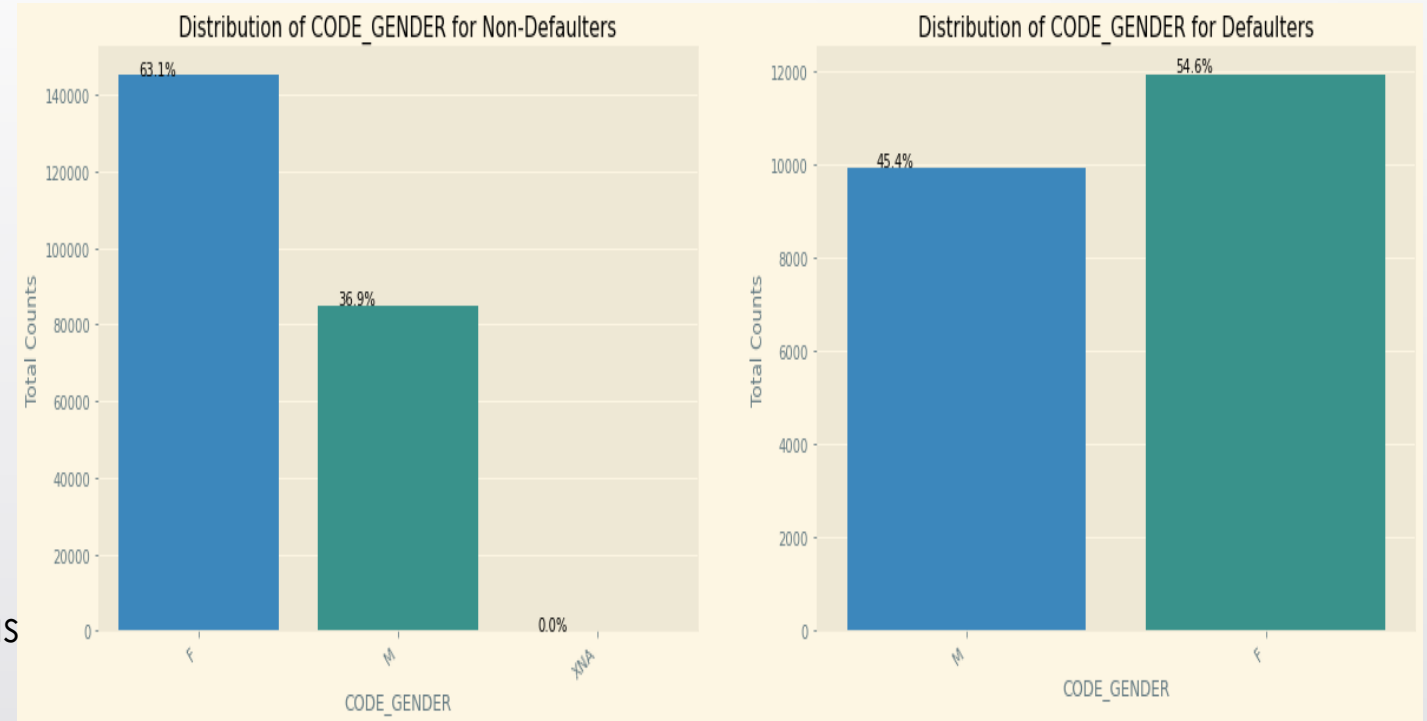


ANALYSIS ON CURRENT APPLICATION

Univariate Analysis on Code_Gender

Observations:

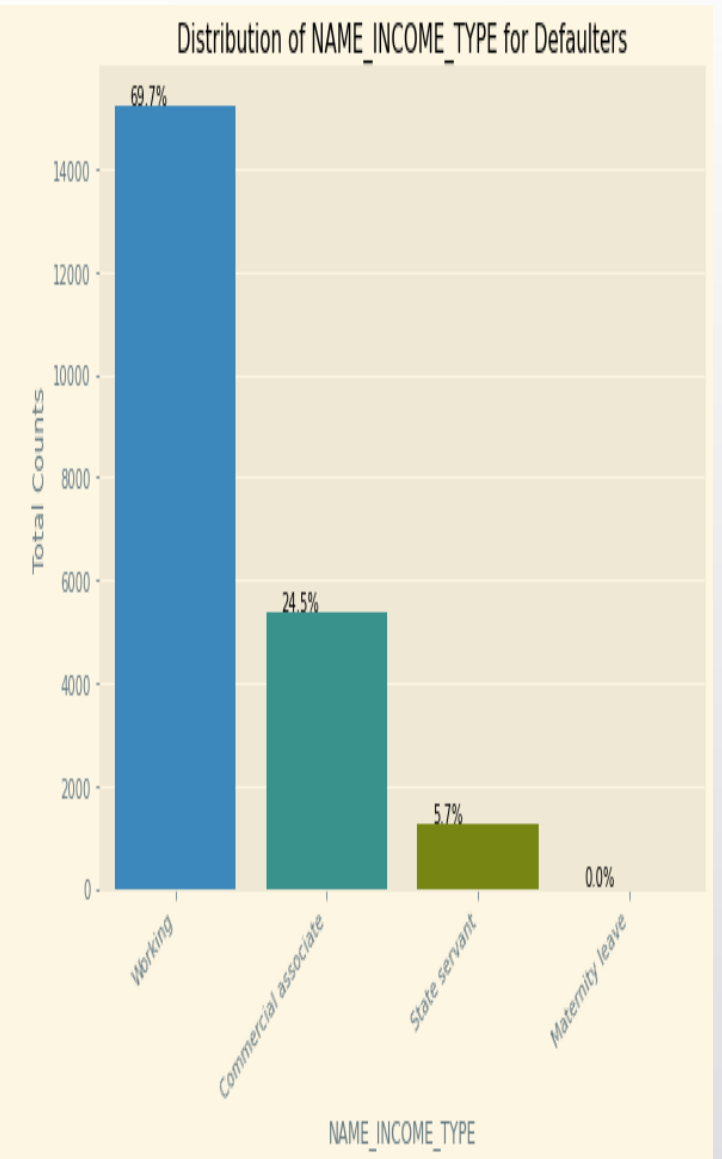
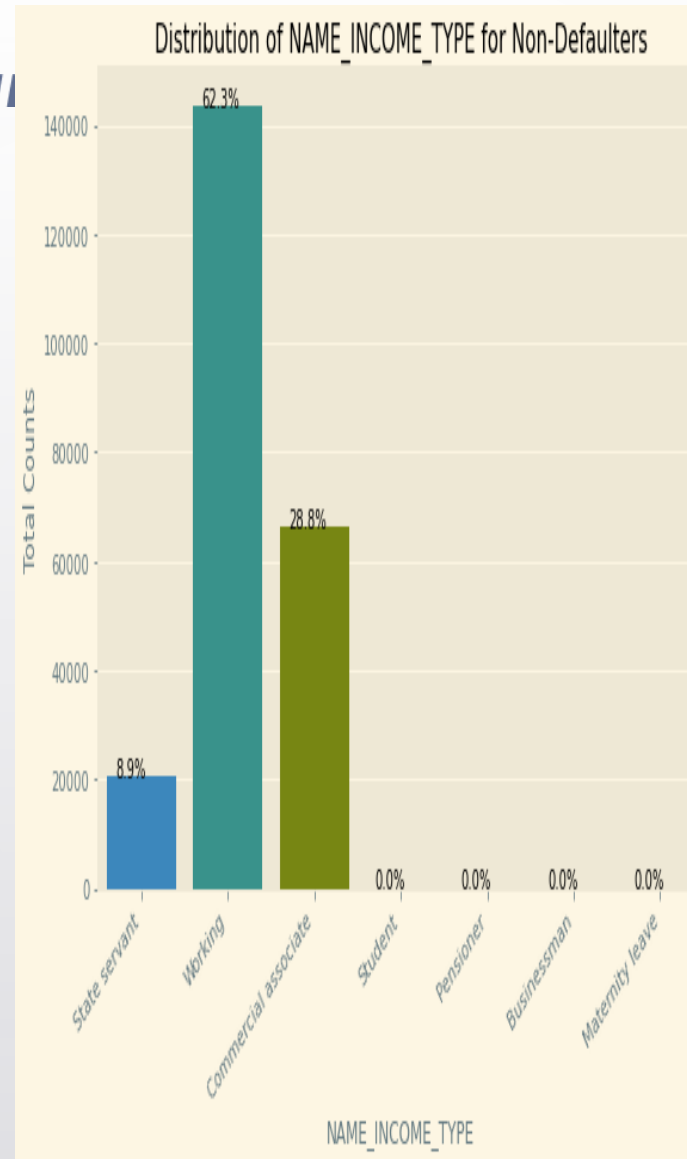
- Here, females contribute 63.1% to the non-defaulters and 54.6% to the defaulters.
- Females apply for loans more than males and hence the more number of female defaulters as well.
- Rate of defaulting of FEMALE is much lower compared to MALE.



Univariate ANALYSIS ON NAME_INCOME_TYP E

Observations:

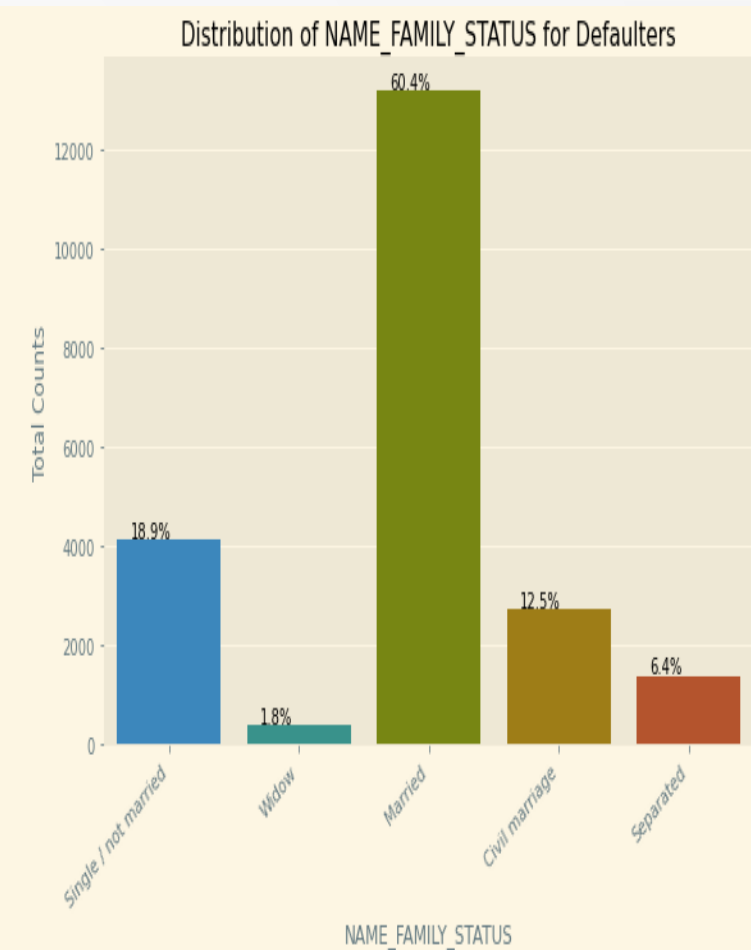
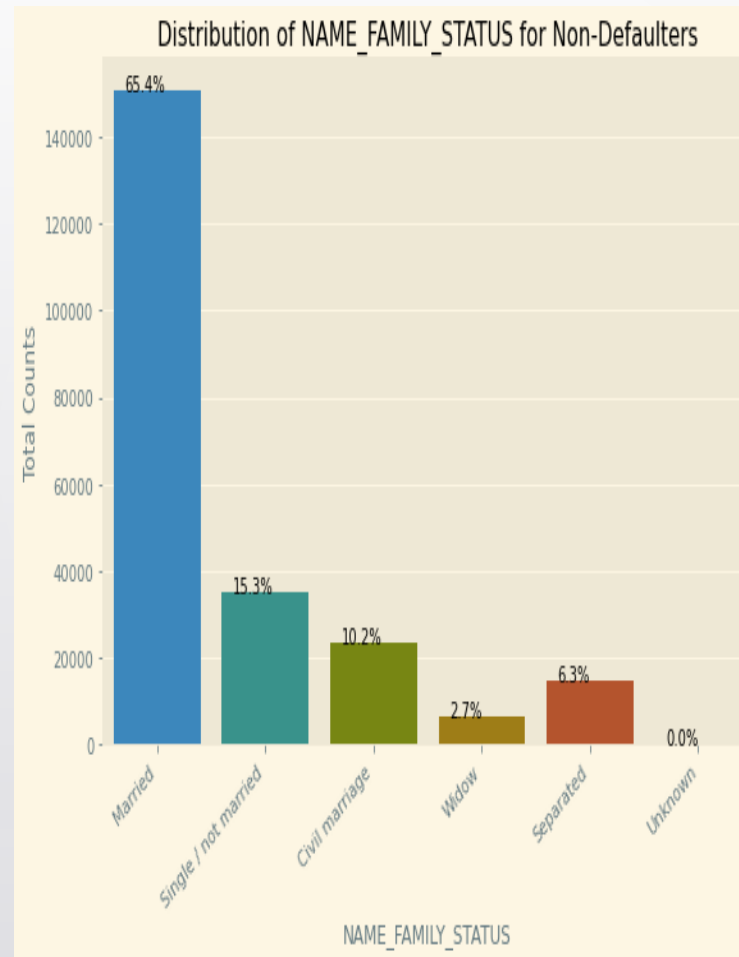
- It's obvious that the students and Business Men don't default.
- Most of the loans are distributed to working class people.
- Working class people contribute 62.3% to non defaulters and 69.7% of the defaulters.
- Chances of defaulting are more in Working class.



UNIVARIATE ANALYSIS ON NAME_FAMILY_STATU S

Observations:

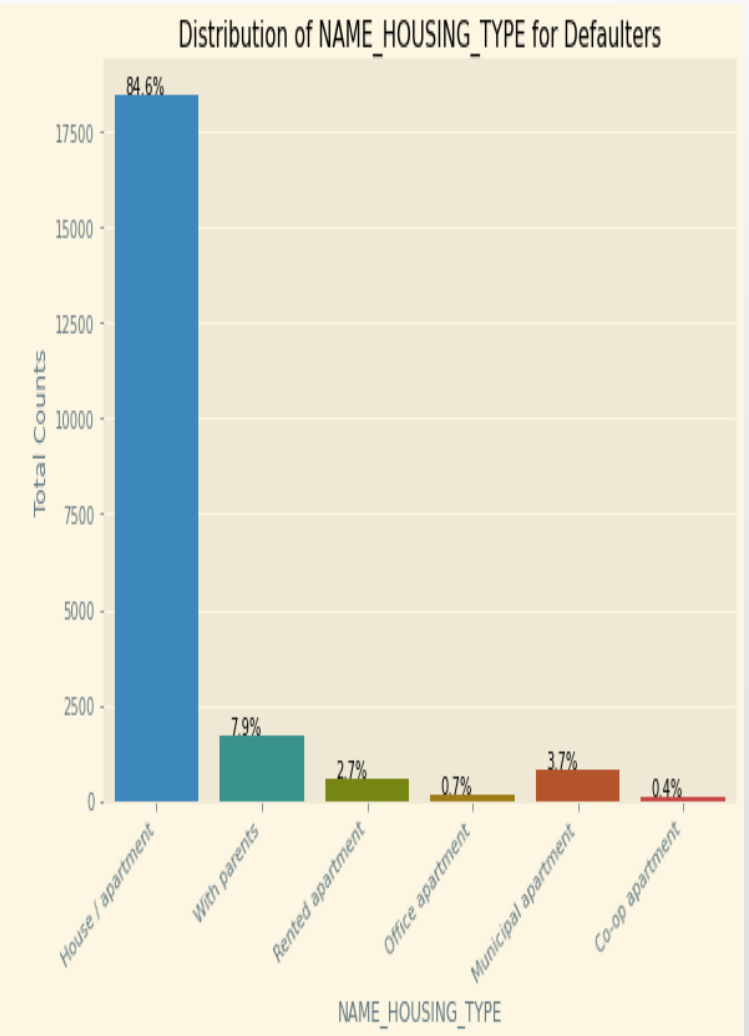
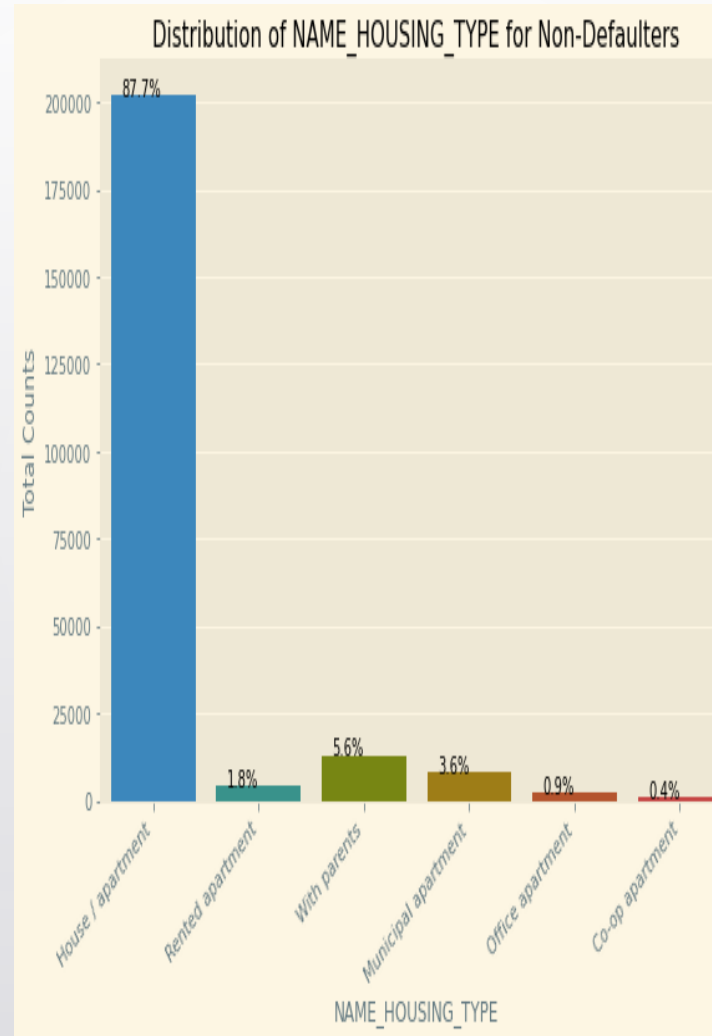
- Singles contribute 15.3% to Non Defaulters and 18.9% to the defaulters.
- More risk is for Singles



UNIVARIATE ANALYSIS ON NAME_HOUSING_TYP E

Observations:

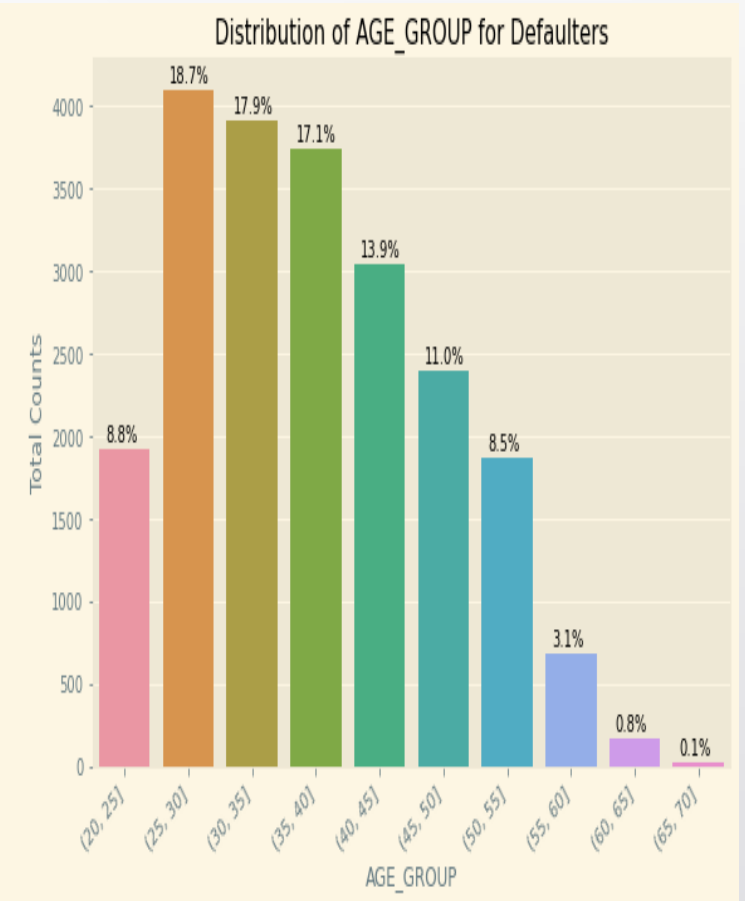
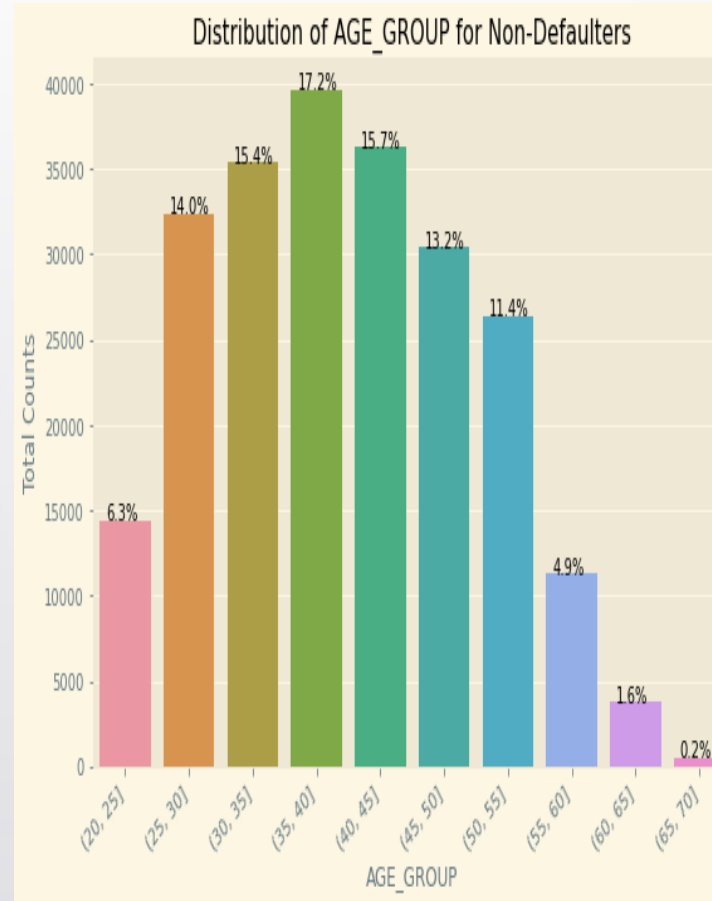
- People who have house or department tend to take more loans.
- People living with parents tend to default more often when compared with others.
- The reason will be increase expense due to looking after their parents with them.



UNIVARIATE CATEGORICAL ORDERED ANALYSIS ON AGE_GROUP

Observations:

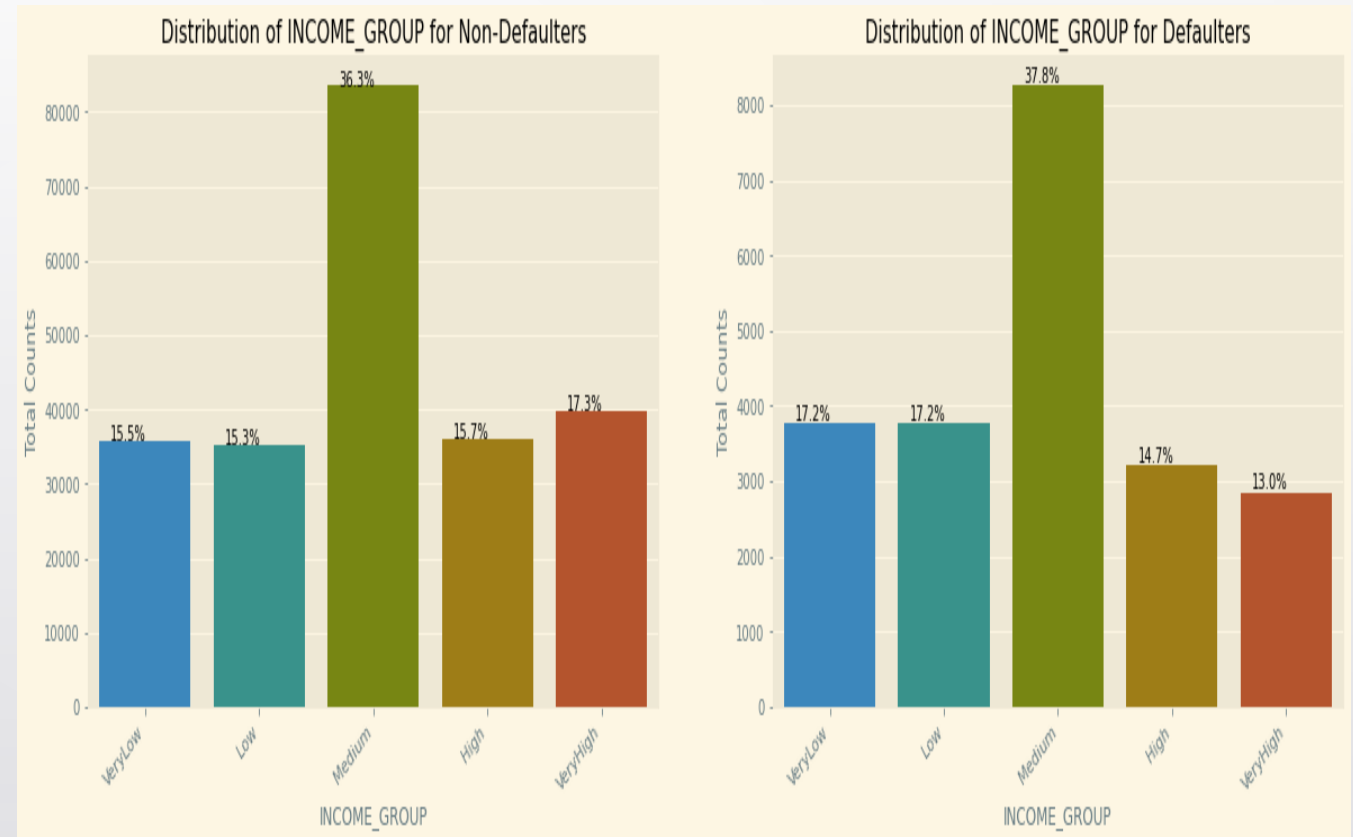
- 25-30 age group tend to default more often.
- With increasing age group, people tend to default less starting from the age 25.
- Reason might be changes in standard of living when they become adults and paid enough.



UNIVARIATE CATEGORICAL ORDERED ANALYSIS ON INCOME_GROUP

Observations:

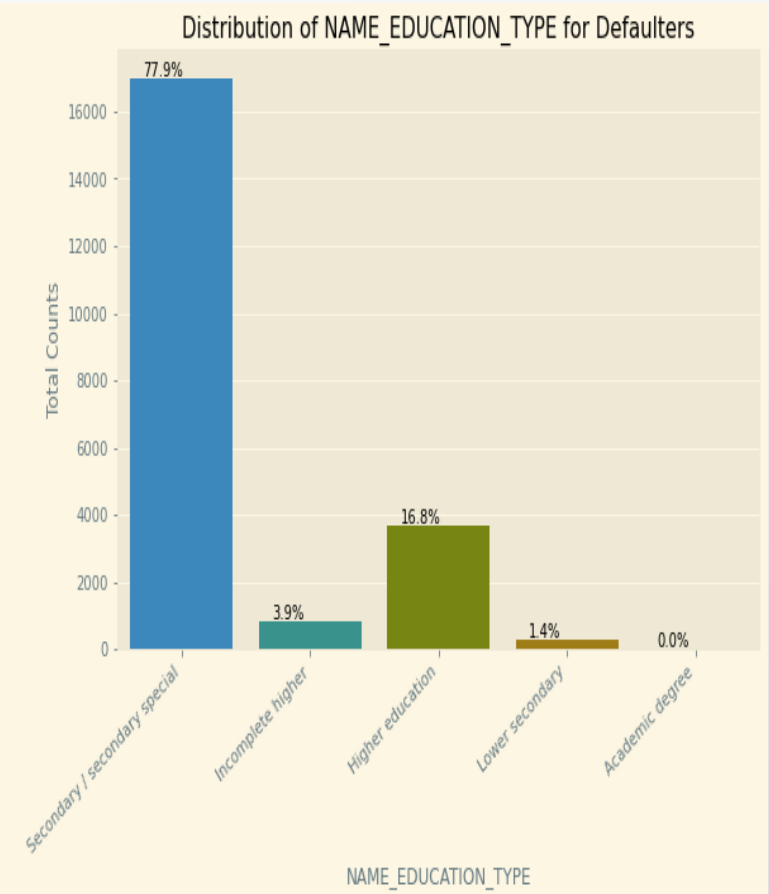
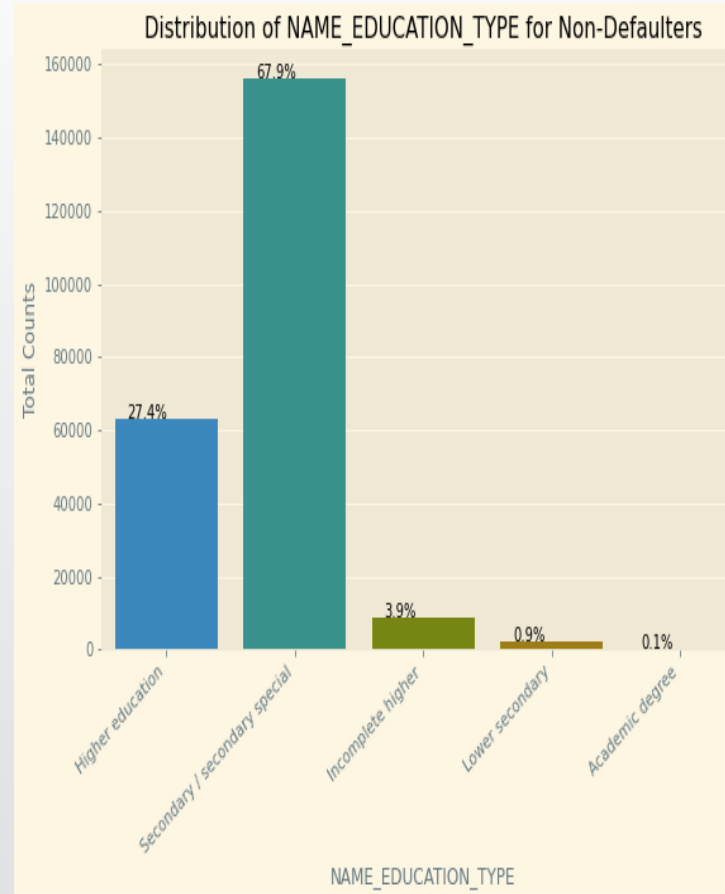
- Very High income group tend to default less often.
- They contribute 13% to the total number of defaulters, while they contribute 17.3% to the Non-Defaulters.
- This might be because they earn more and able to cope up with expenditures well.



UNIVARIATE CATEGORICAL ORDERED ANALYSIS ON NEW_EDUCATION_TYP E

Observations:

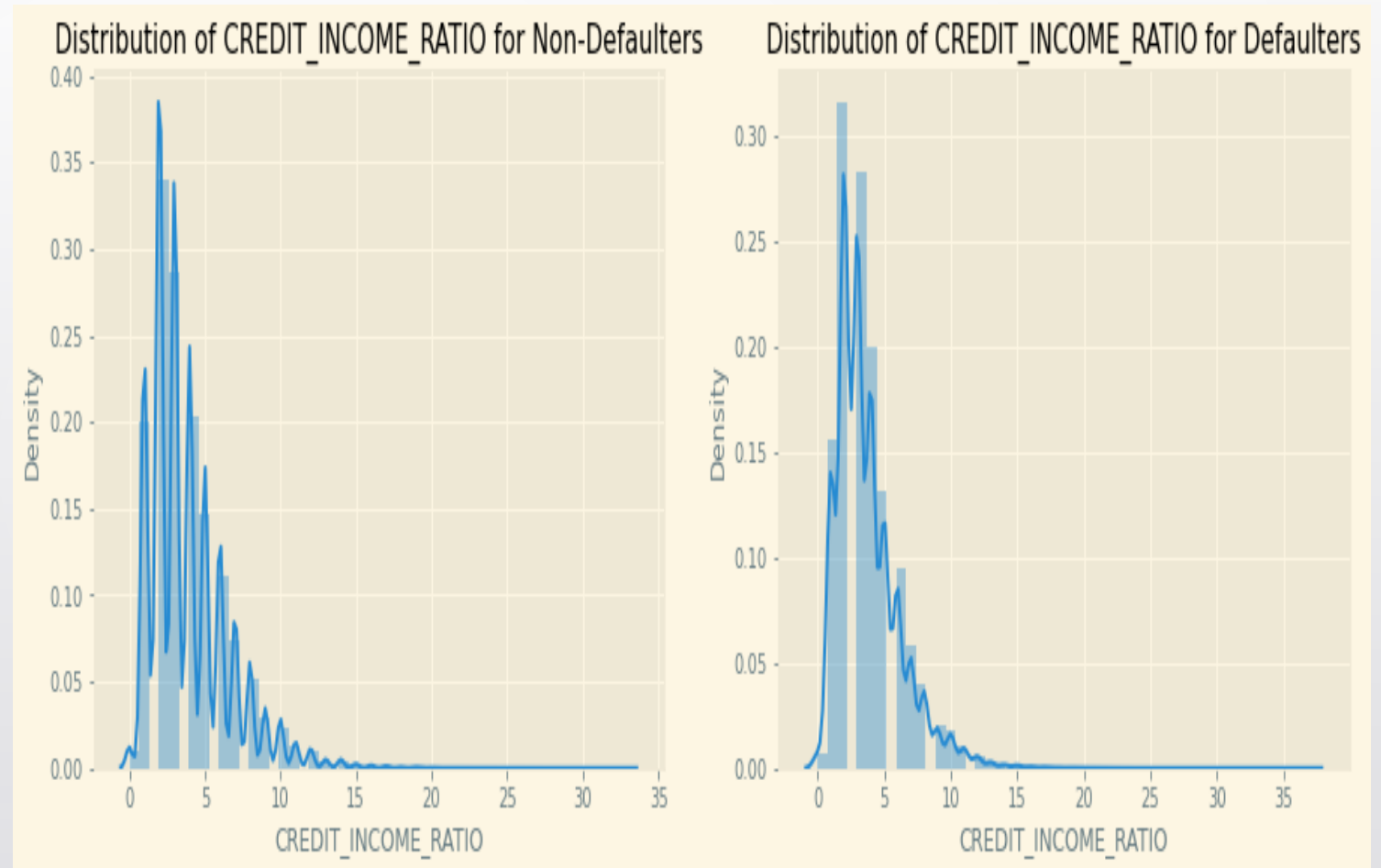
- Higher education streams tends to less likely default and all others equally default with secondary having more defaulters



UNIVARIATE CONTINUOUS ANALYSIS OF CREDIT_INCOME_RATIO

Observations:

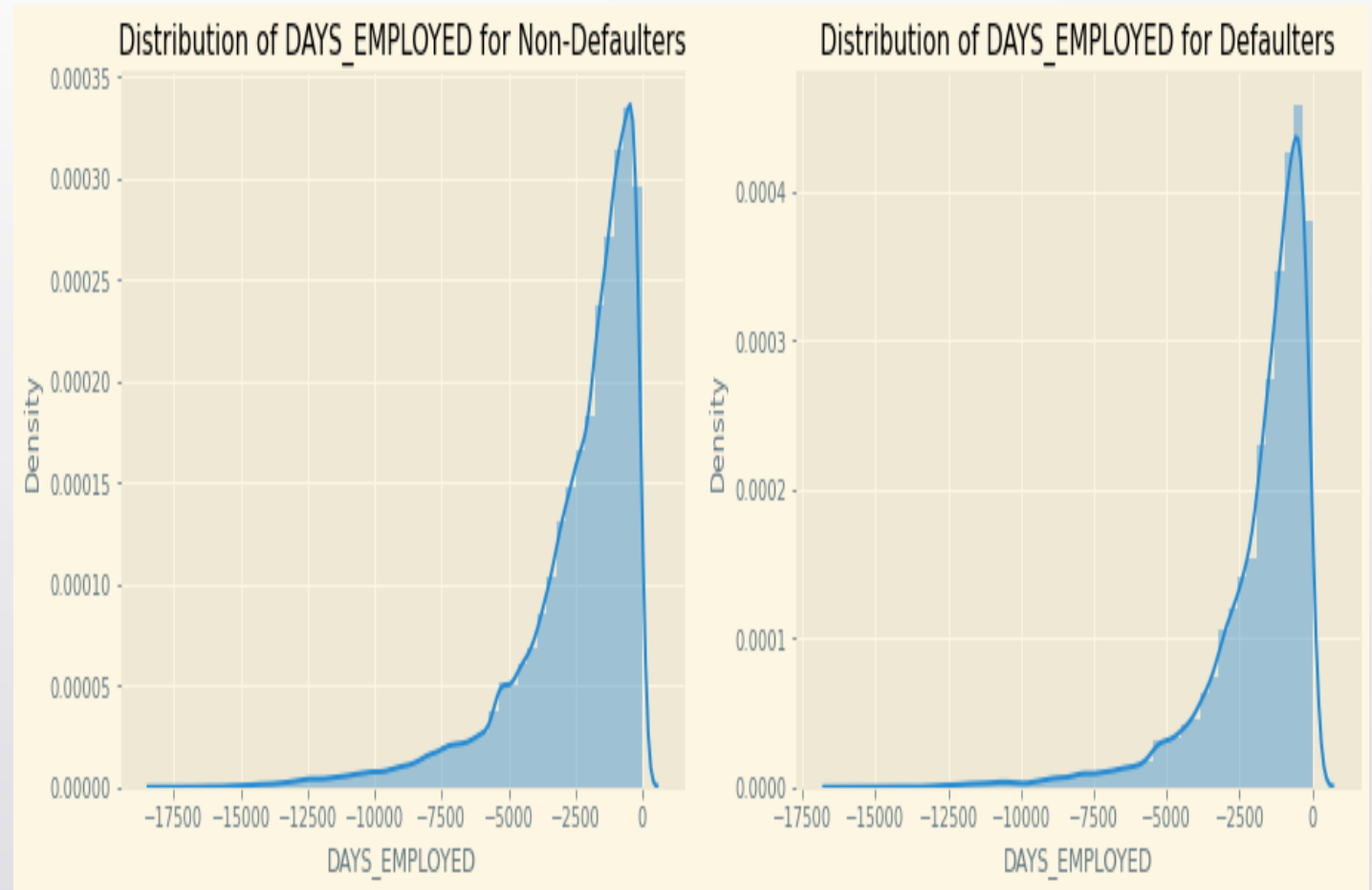
- The CREDIT_INCOME_RATIO is more for people who default.



UNIVARIATE CONTINUOUS ANALYSIS OF DAYS_EMPLOYED

Observations:

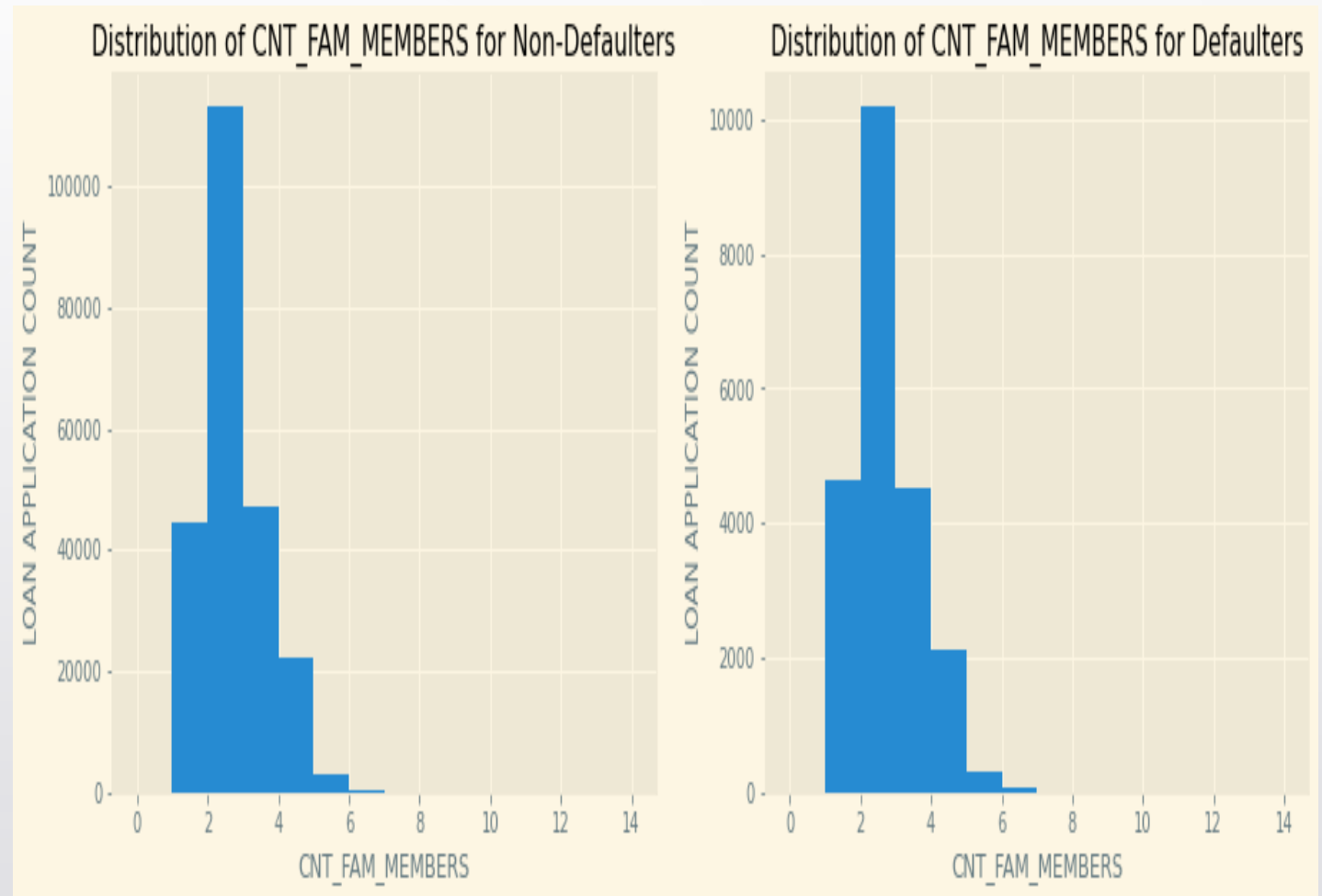
- Days employed is almost similar in both cases



UNIVARIATE CONTINUOUS ANALYSIS OF CNT_FAM_MEMBERS

Observations:

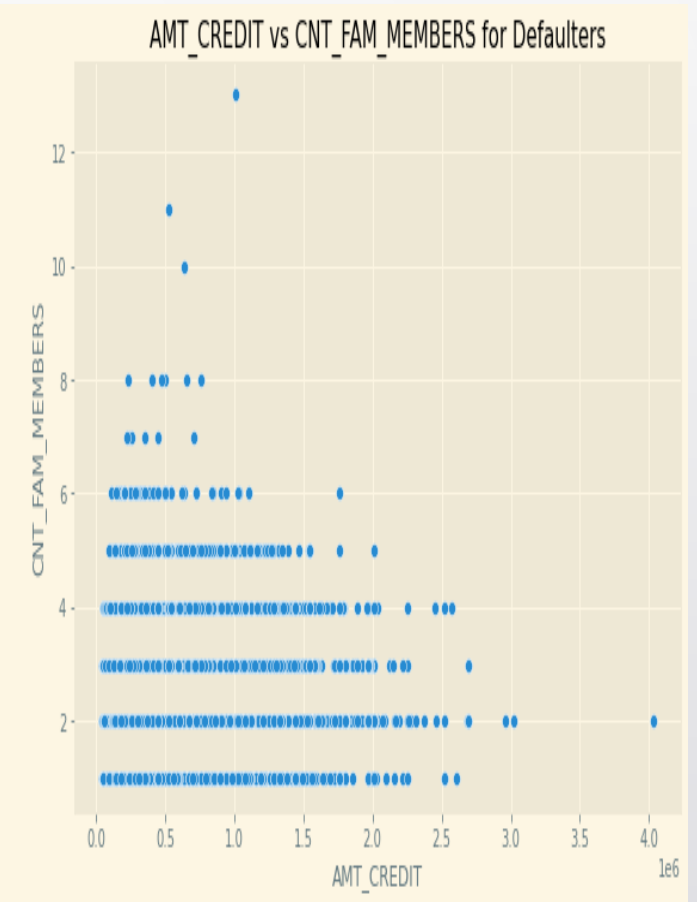
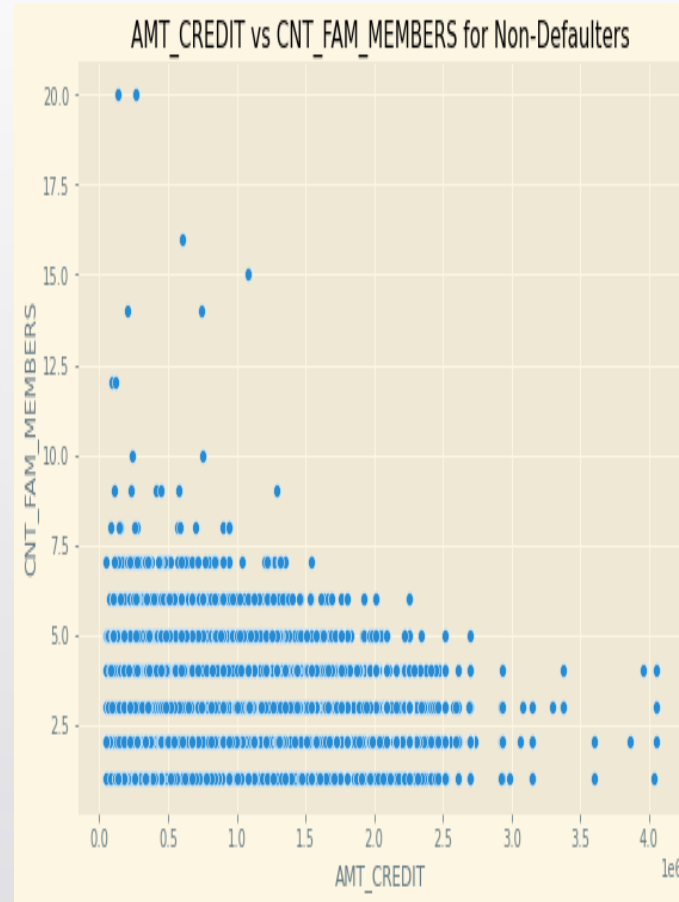
- It's obvious that, family of 3 applies for loan more often than others.
- Reason might be increased expense and less earning members.



BIVARIATE ANALYSIS BETWEEN AMT_CREDIT AND CNT_FAM_MEMBERS

Observations:

- Density in the lower left corner is similar in both the case.
- People defaults almost similar if the family is small and the AMT_CREDIT is low.
- larger families and people with larger AMT_CREDIT default less



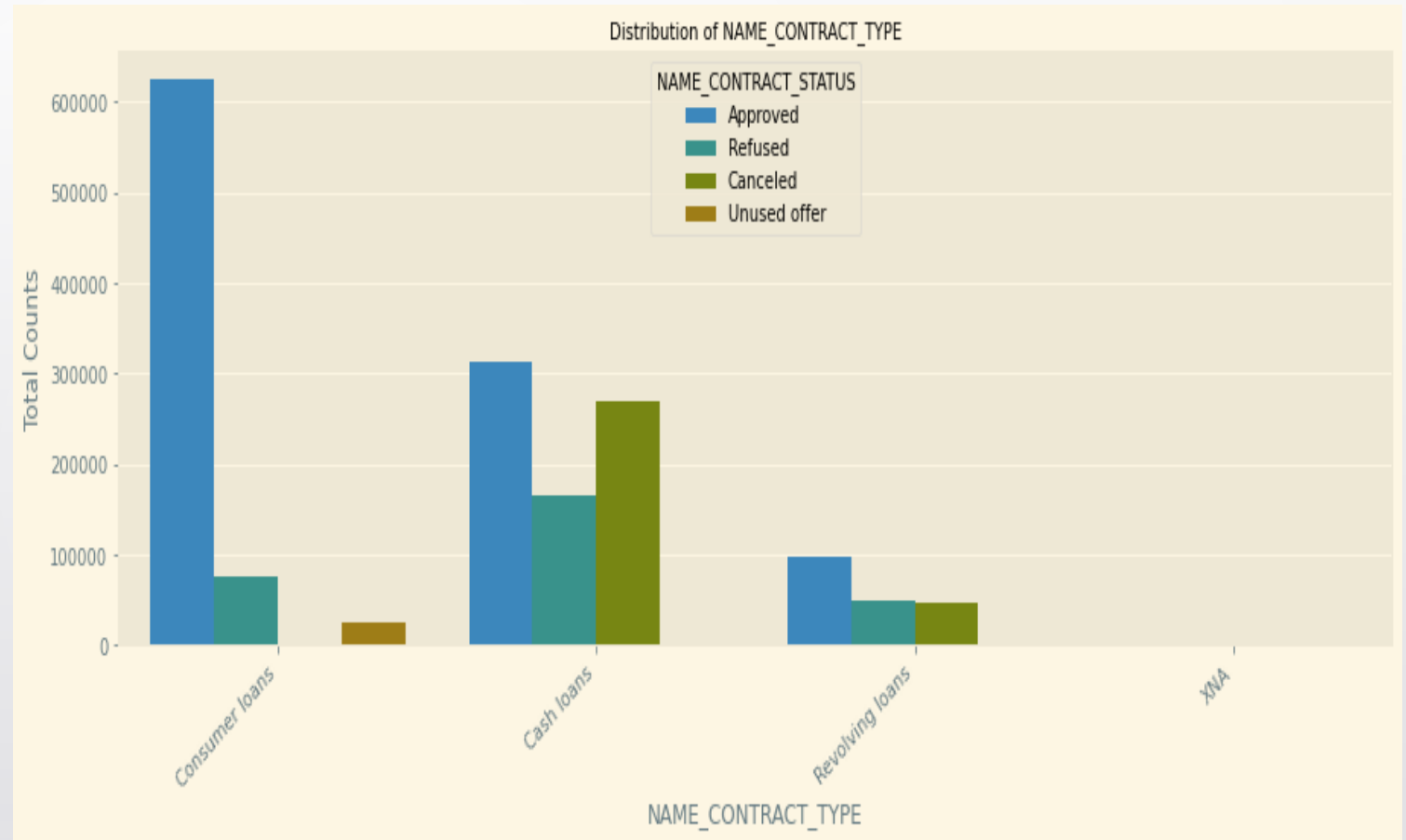


Analysis on Previous Application

UNIVARIATE ANALYSIS ON NAME_CONTRACT_TYPE

Observation:

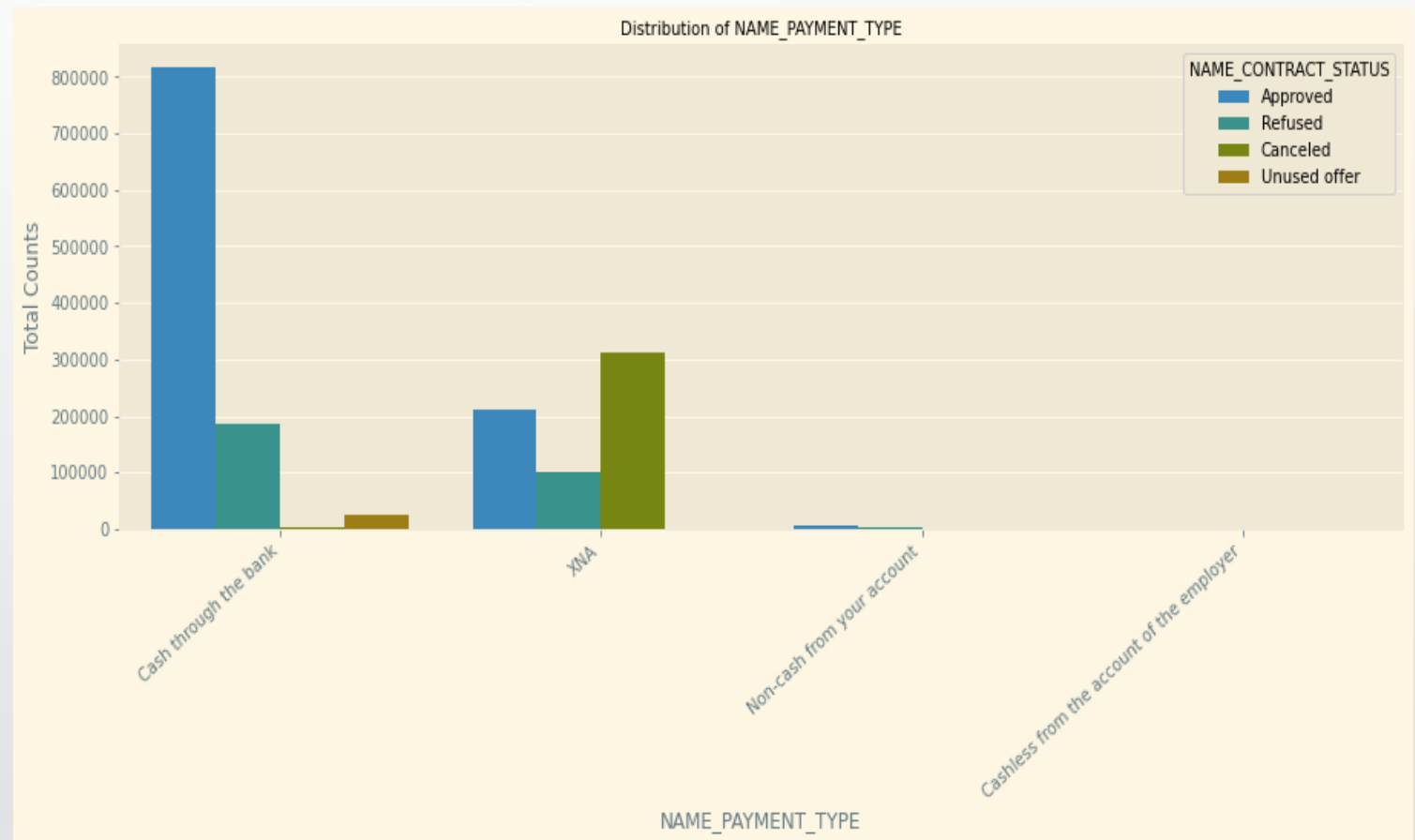
- Eventhough cash loans are not much accepted, applications 'Cash loan' and 'Consumer loan' are on a majority.



UNIVARIATE ANALYSIS ON NAME_PAYMENT_TYPE

Observation:

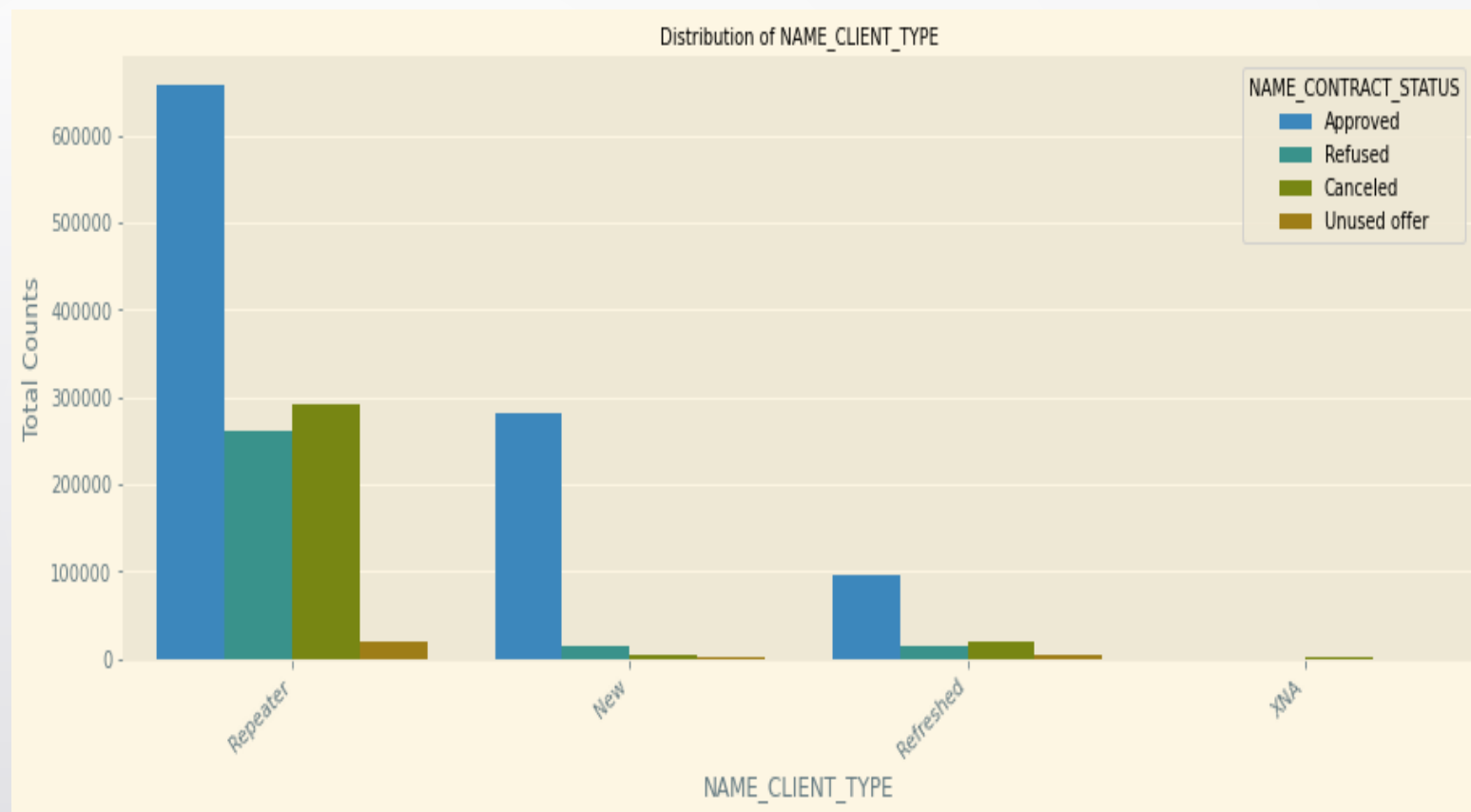
- Many chose to repay the loan using the 'Cash through the bank' option
- 'Non-Cash from your account' & 'Cashless from the account of the employer' options are not much used by customers.



UNIVARIATE ANALYSIS ON NAME_CLIENT_TYPE

Observations:

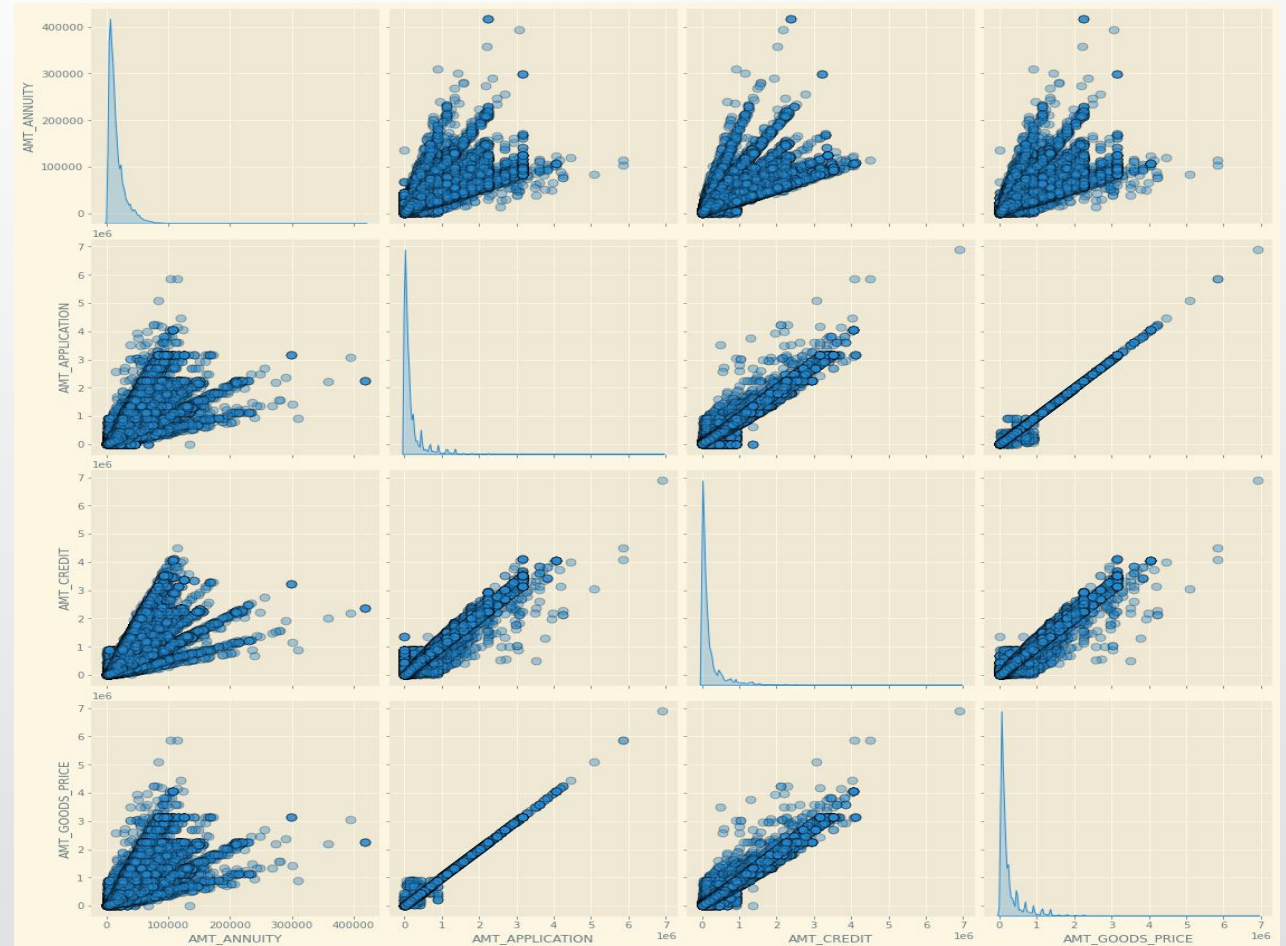
- Majority applications are from repeat customers.
- In the total applications 70% of customers are repeaters.
- Repeaters are mostly refused



CHECKING CORRELATION USING PAIRPLOT FOR BIVARIATE ANALYSIS ON NUMERICAL COLUMN

Observations:

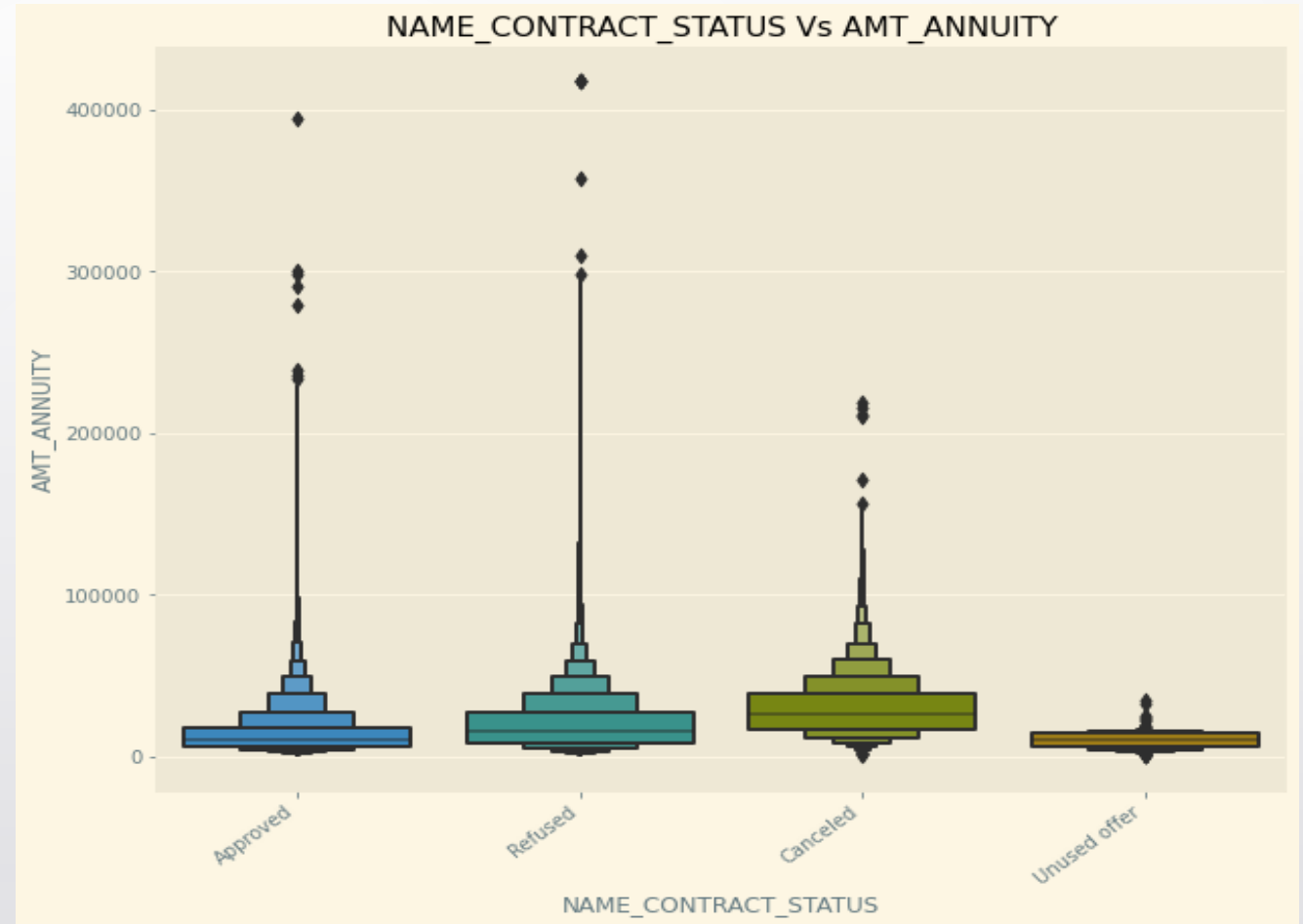
- Previous application's Annuity increase can increase: Goods price asked by client, Final credit amount on the previous application with bank's approval and Credit asked by client on the previous application
- Credit client asked is highly influenced by the Goods price of good that client has asked for on the previous application
- Final credit amount to the customer previously, after approval improves the application amount and also the goods price asked by client.



CHECKING CORRELATION USING BOXPLOT

Observations:

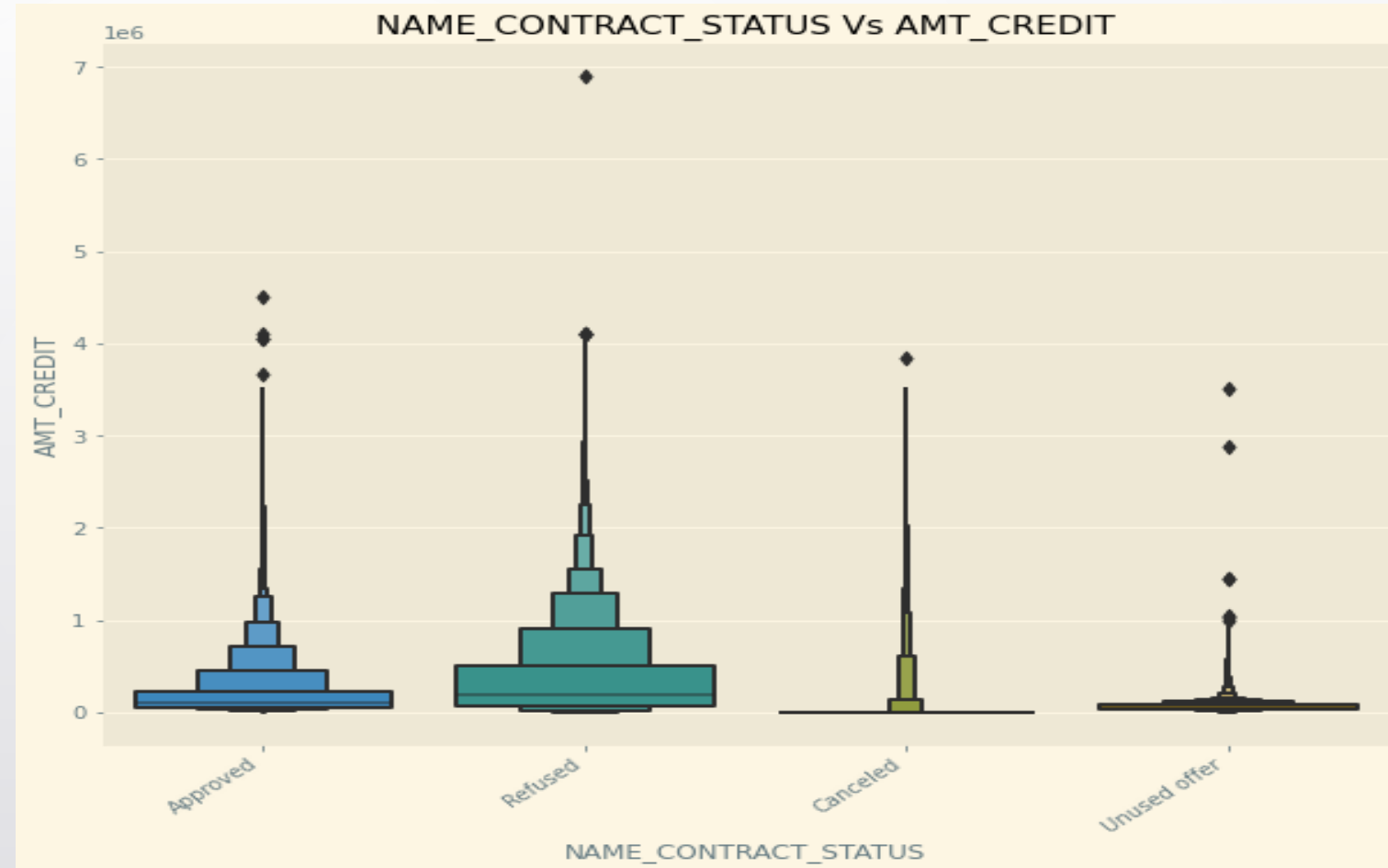
- Loan application for people with lower AMT_ANNUIITY gets canceled or unused always.
- Applications with too high AMT_ANNUIITY also got refused than others.



CHECKING CORRELATION USING BOXPLOT

Observations:

- it gets cancelled/unused when the AMT_CREDIT is too low.



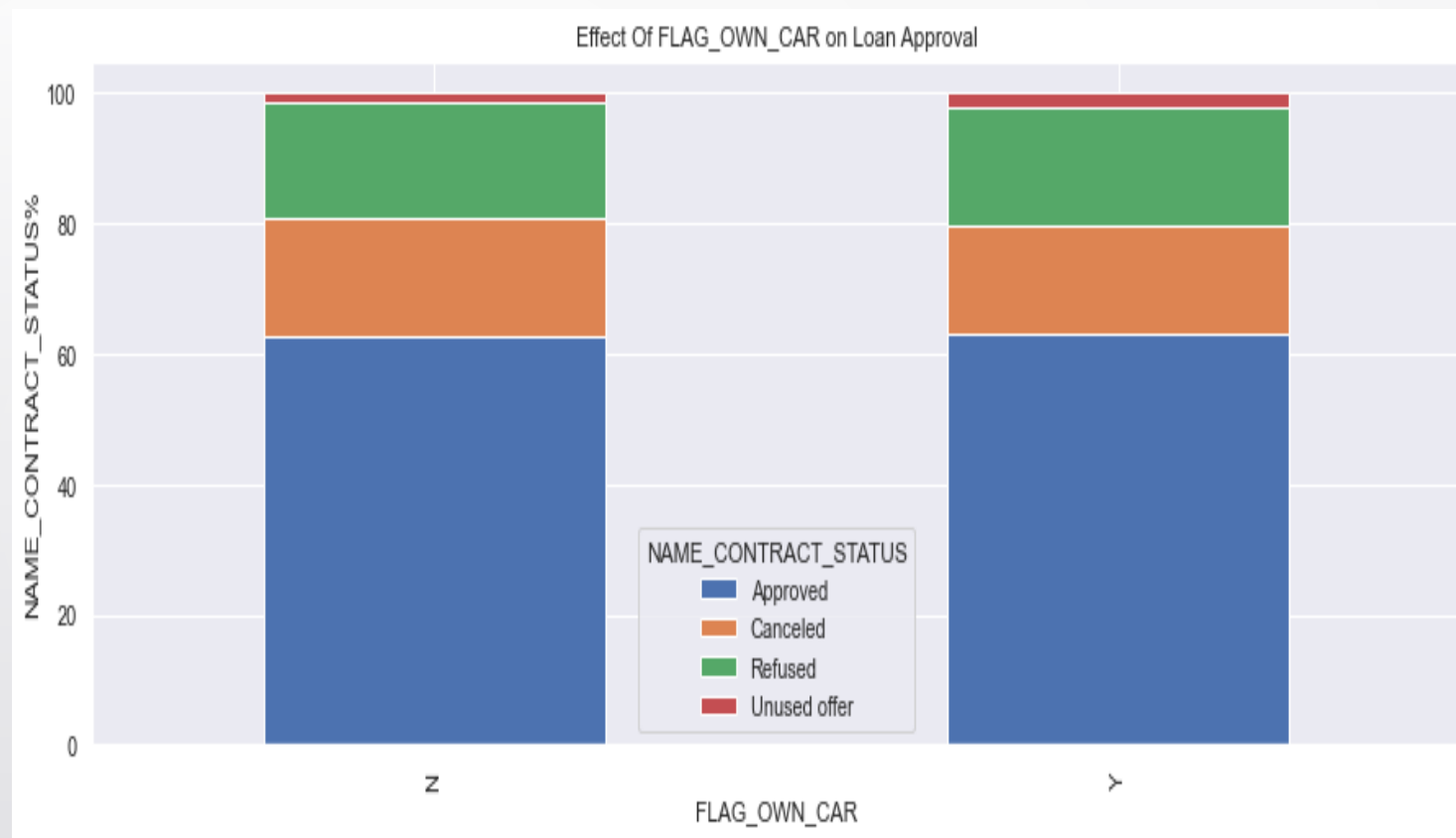


ANALYSIS ON BOTH PREVIOUS AND CURRENT APPLICATIONS MERGED

DISTRIBUTION BETWEEN FLAG_OWN_CAR AND NAME_CONTRACT_STA TUS

Observations:

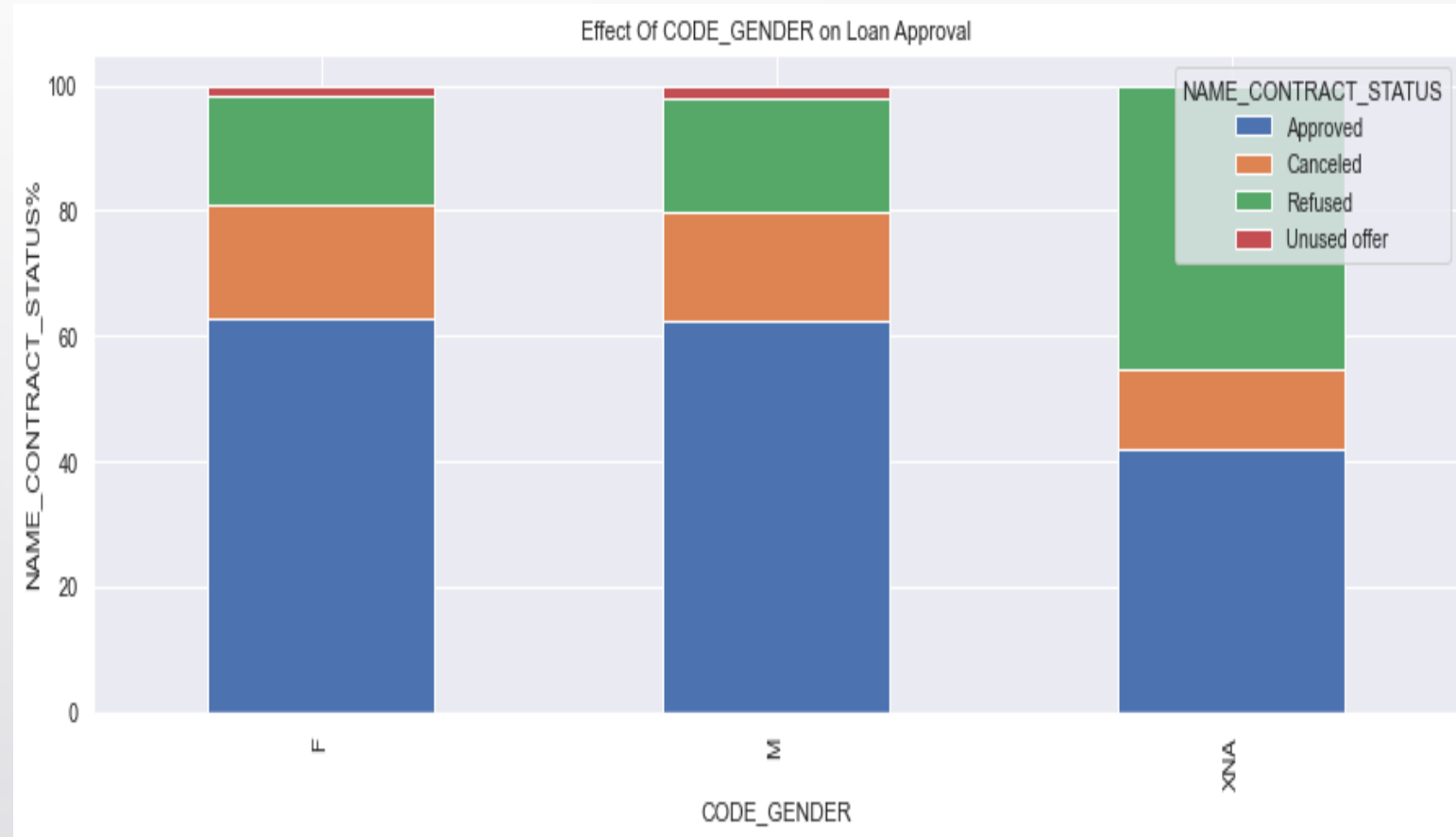
- Car ownership doesn't have any effect on application approval or rejection.
- But bank can add more consideration to car ownership while approving a loan amount since according to previous analysis, defaulting is less for people with car.



DISTRIBUTION BETWEEN CODE_GENDER AND NAME_CONTRACT_STATUS

Observations:

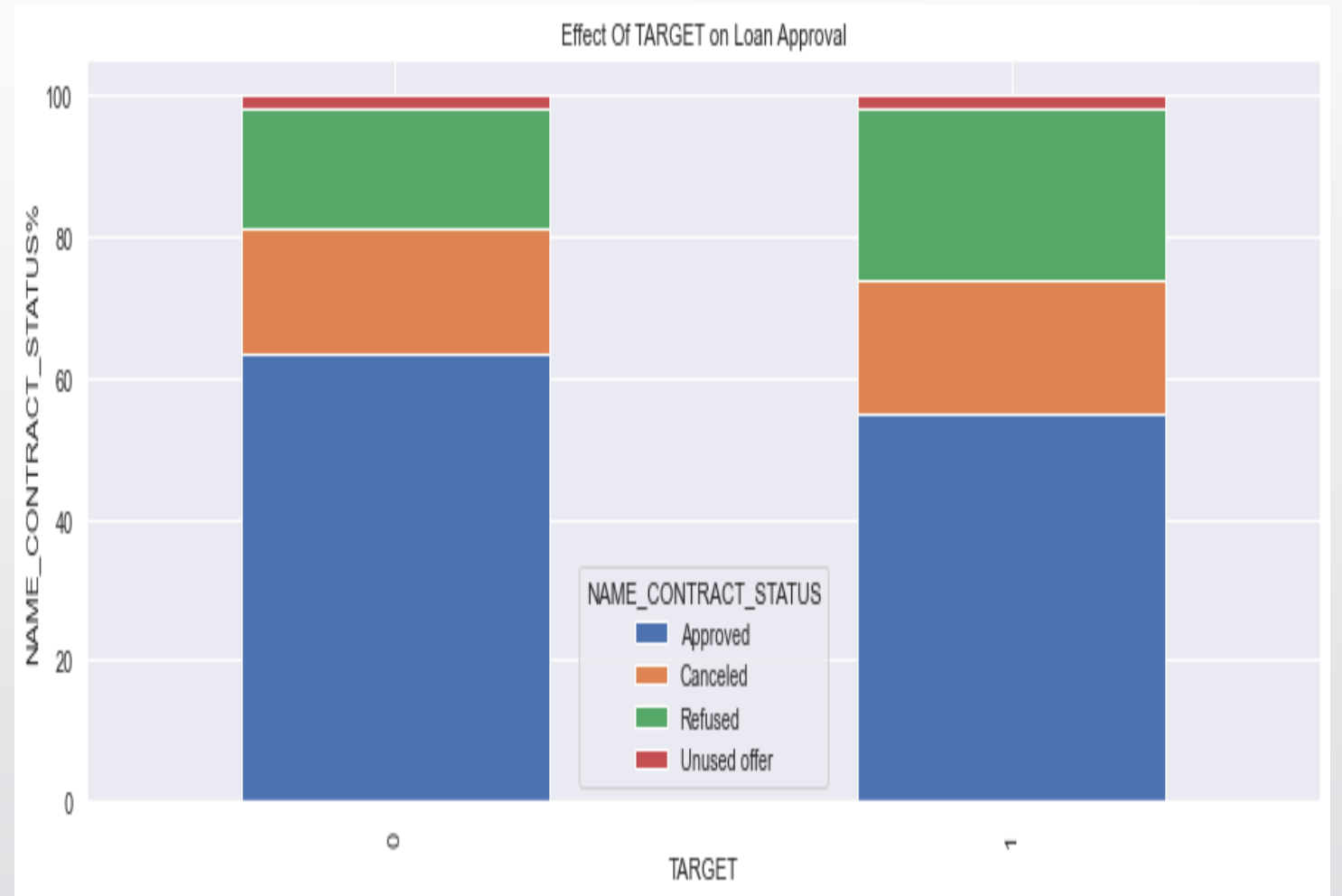
- Code gender doesn't have any effect on application approval or rejection.
- But according to our previous analysis, female have lesser chances of default.
- The bank can give consideration to females while approving a loan amount.



DISTRIBUTION BETWEEN TARGET AND NAME_CONTRACT_STATUS

Observtions:

- People who were approved for a loan earlier, defaulted less often
- People who were refused a loan earlier have higher chances of defaulting.





Conclusion

- Banks should focus more on 'Student' , 'pensioner' and 'Businessman' with housing 'type other than 'Co-op apartment' for successful payments.
- Banks should provide consideration for people from large families, since there are chances of increased number of earning members.
- Banks should focus less on income type 'Working' as they are having most number of unsuccessful payments.
- Get as much as clients from housing type 'With parents' as they are having least number of unsuccessful payments.
- Banks should give more considerations to people with high income since they default less.



Thank you