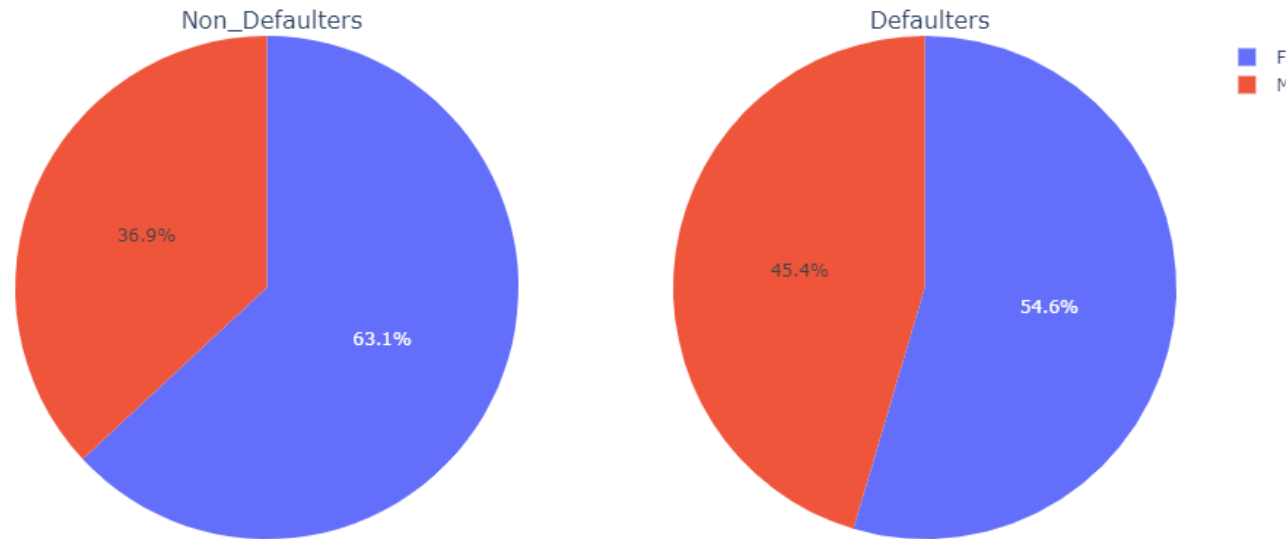# CREDIT EDA CASE STUDY

- Aakash N Kotkar

Purpose

- Provide credit risk analysis to bank based on applicant data.
- Suggest bank to whom they should give loan with lowest possible risk
- So bank can avoid NPA & control loss of business

Univariate Analysis on Application Data

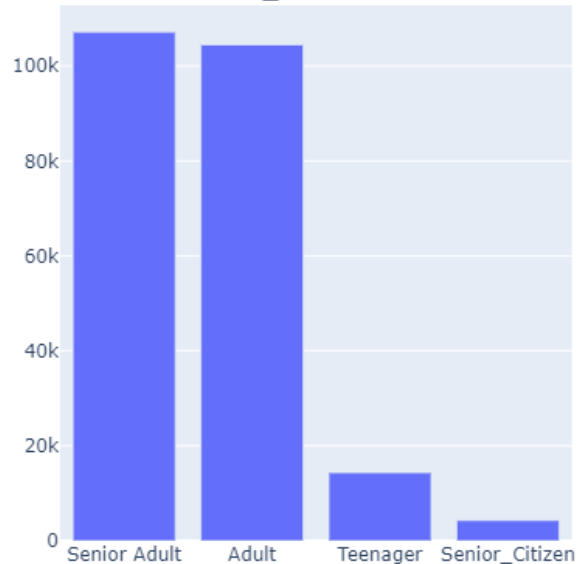# Gender Distribution – Defaulters vs Non-Defaulters

Non Defaulter
F – 1,45,246
M – 85,024

Defaulter
F – 11,919
M – 9,914

- By analyzing pie chart we can consider female Defaulters & Non- Defaulters percentage is high as compare to male.

- But if we consider both pie chart at same time we can under stand there is big difference between male Defaulter to Non-Defaulter percentage.

- So we can say that male & Female chances of defaulter loan are same male applicant chances of defaulting loan is little higher than female if we take ratio of defaulter & non defaulter.
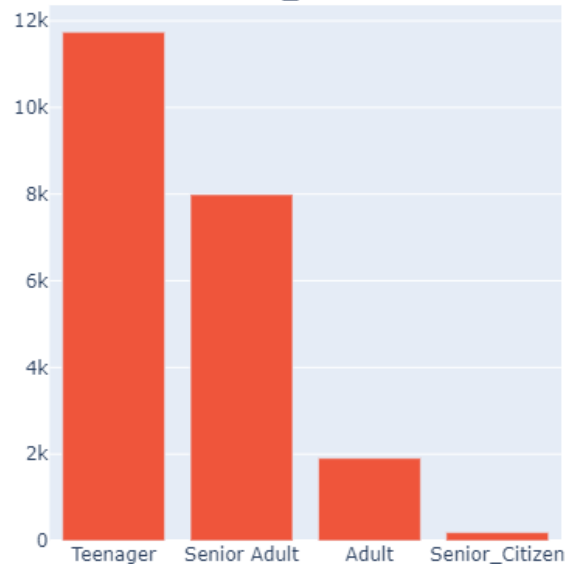
Age Distribution– Defaulters vs Non-Defaulters

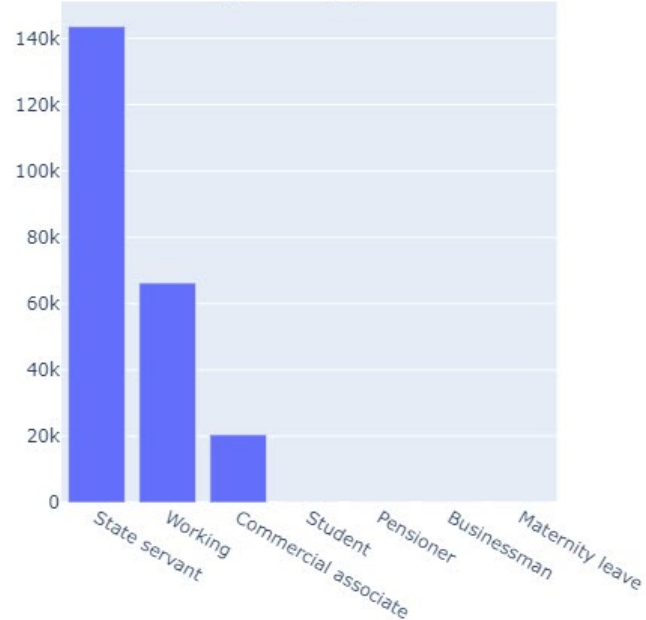
Non-Defaulters
Senior adult – 1,07,162
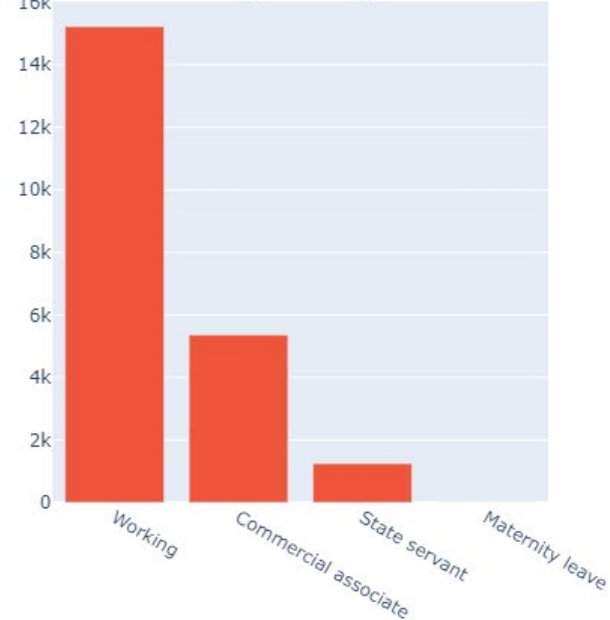
Defaulters
Senior adult – 11,742

- By analyzing Non-Defaulter bar chart we can consider senior adult & Adult less likely to default loan

- Where by analyzing Defaulter bar chart we can consider Teenager are more likely to default loan.

- In Non-Defaulter chart senior adult are good but in Defaulter chart senior adult are 2nd likely to default loan after teenager.

- But ratio between non-defaulter & defaulter for senior adult is nearly 9% so by considering other parameter we can give loan to senior adult.

INCOME TYPE – Defaulters vs Non-Defaulters


- By analyzing both chart we can consider giving loan to state servant is most likely to may be because state servant job is more stable.

- When we consider both chart we see 66K working professional are Non-defaulter but 15k defaulter also present so giving loan to working professional are higher.

- 20k commercial associate are non-defaulters & 5k are defaulters. So there is 20% to 25% that given loan is get defaulted.

FAMILY STATUS – Defaulters vs Non-Defaulters

Distribution of NAME_FAMILY_STATUS for Non-Defaulters
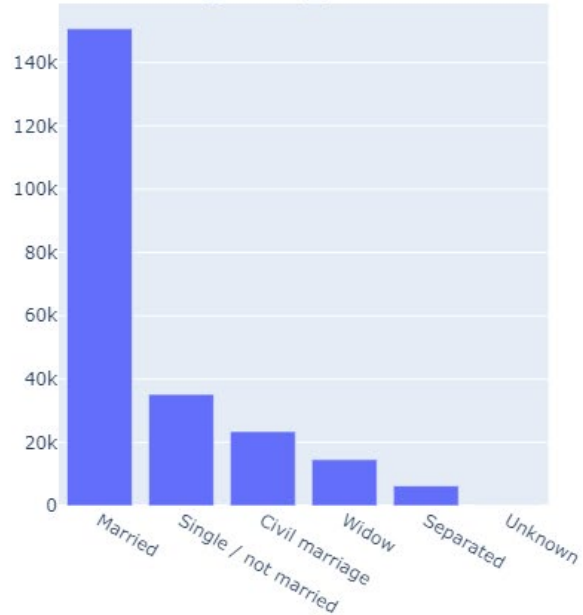
Distribution of NAME_FAMILY_STATUS for Defaulters
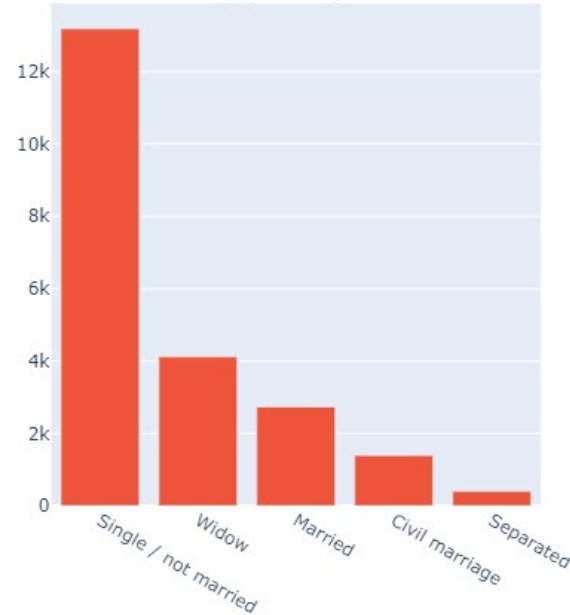
- We can clearly see that married people are less likely to default loan as compare to other family status applicant.

- Where 35k pay loan & 13k default the loan it means giving loan to singe person are risky.

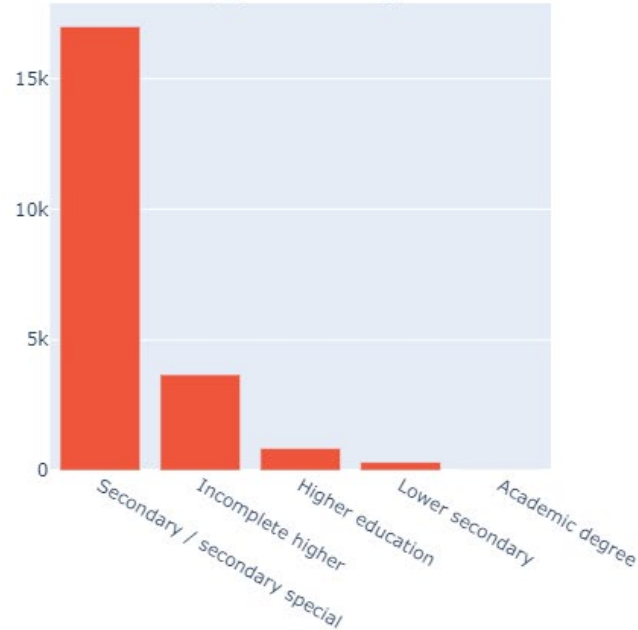- Giving loan to widow is also seams to be risky because 14k pay loan 4k default the loan.

EDUCATION_TYPE – Defaulters vs Non-Defaulters

Distribution of NAME_EDUCATION_TYPE for Non-Defaulters
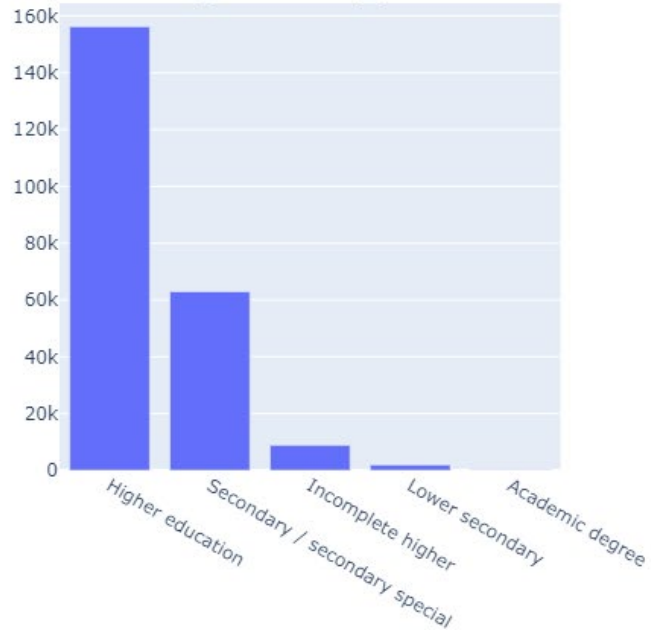
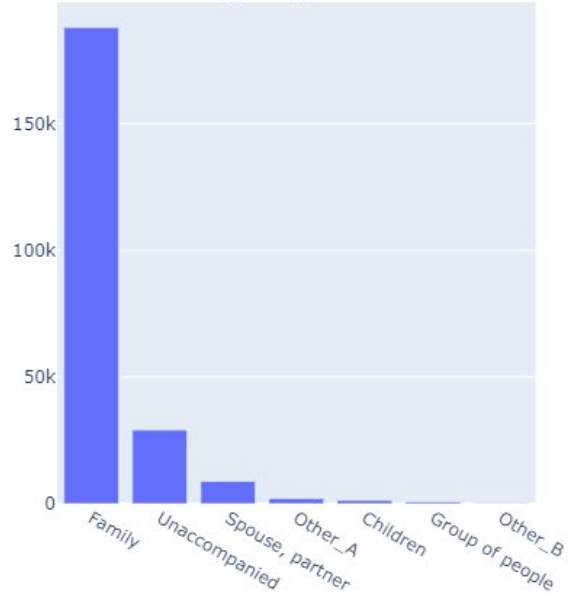Distribution of NAME_EDUCATION_TYPE for Defaulters

- We can see giving loan to Higher education is highly secure.

- Giving loan to secondary special is risky because even 62k applicant pay loan but 17k applicant default loan which means 25% to 30% from this category default the loan.
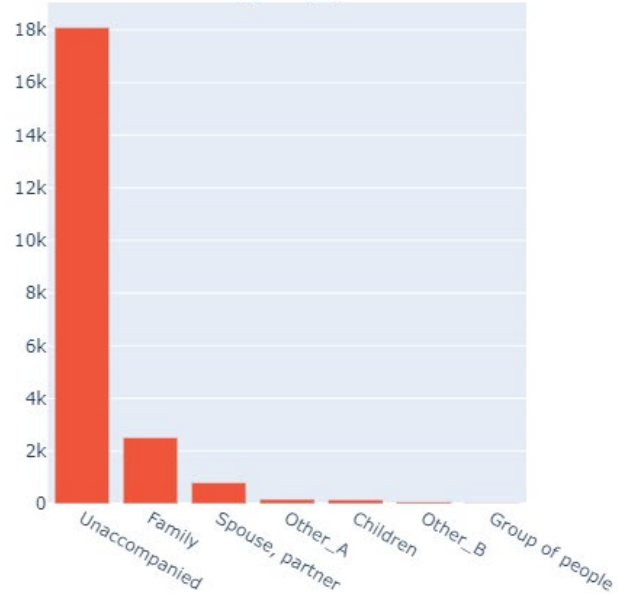
TYPE SUITE – Defaulters vs Non-Defaulters


Distribution of NAME_TYPE_SUITE for Non-Defaulters

Distribution of NAME_TYPE_SUITE for Defaulters

- So 29k unaccompanied applicant pay loan but 18k default the loan which is more than 40%.

- Bank should avoid giving loan to unaccompanied applicant.

- Where other hand we can see 187k family applicant pay loan & only 2.5k default the loan.

- Giving loan to family applicant is highly secure.

# CONTRACT TYPE– Defaulters vs Non-Defaulters
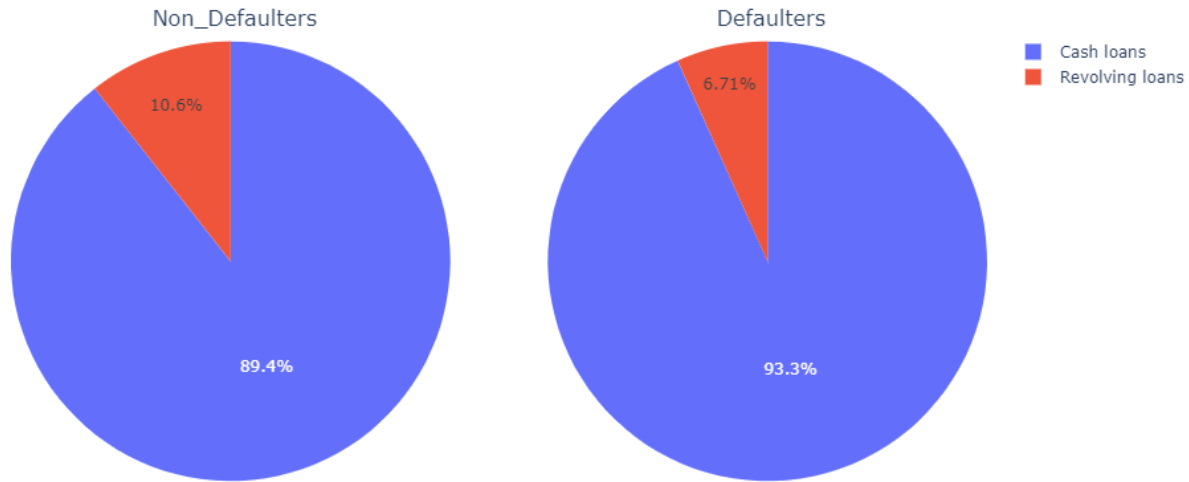
- We can see that in both pie chart Defaulter & Non defaulter percentage of cash loan type contract is high.

- It because there are many applicant for cash loan contract as compare to revolving loan type.

- But we can guess that giving loan to revolving loan applicant is little safe than cash loan type.

INCOME RANGE – Defaulters vs Non-Defaulters

Distribution of AMT_INCOME_RANGE for Non-Defaulters
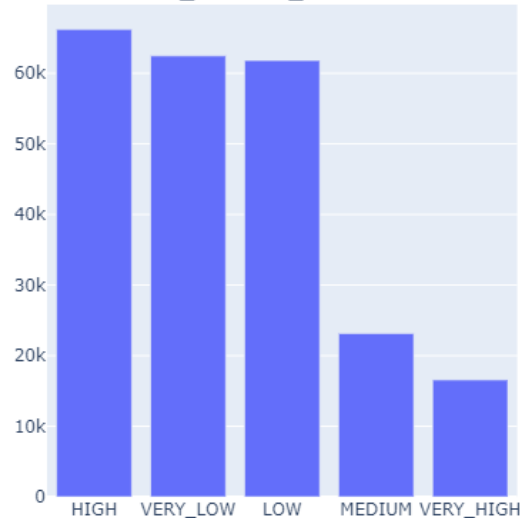
Distribution of AMT_INCOME_RANGE for Defaulters
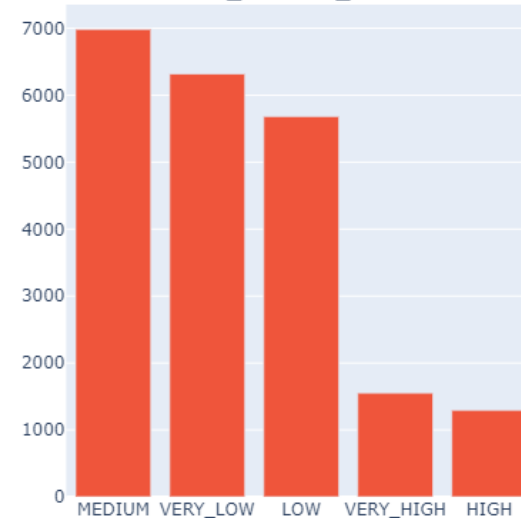
- We can see that high income people are less likely to default loan. May be because there earring is more than loan EMI.

- But giving loan to medium, low & very low income applicant is risky because nearly 10% applicant default loan from this category.

- If bank want to give loan to medium, low & very low category applicant this should consider saw other data related to them.

Bivariate Analysis on Application Data

Bivariate Analysis between Defaulter & Non- Defaulter

Target 0 (Loan- Non Defaulters)

Target 1 (Loan- Defaulters)
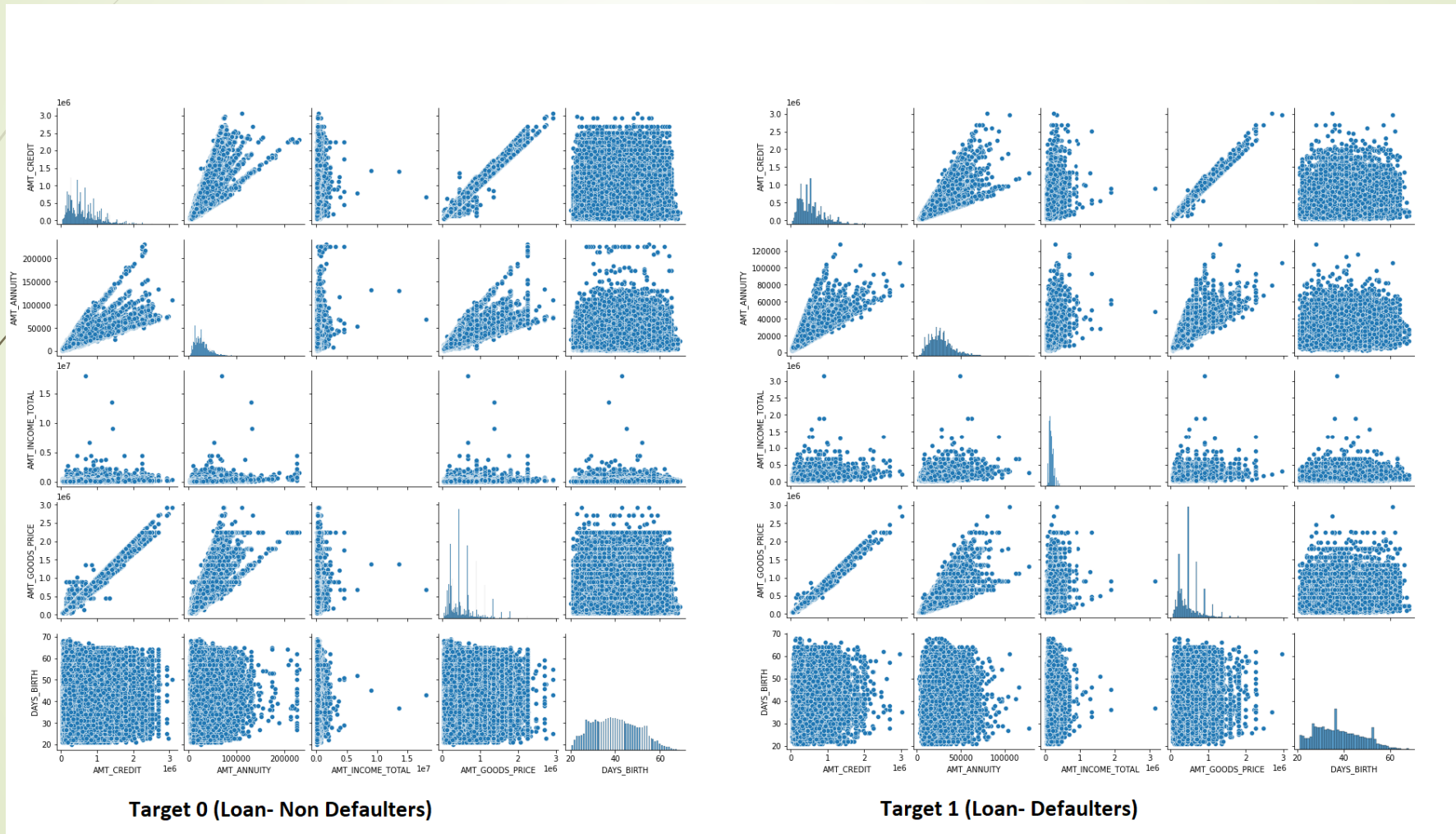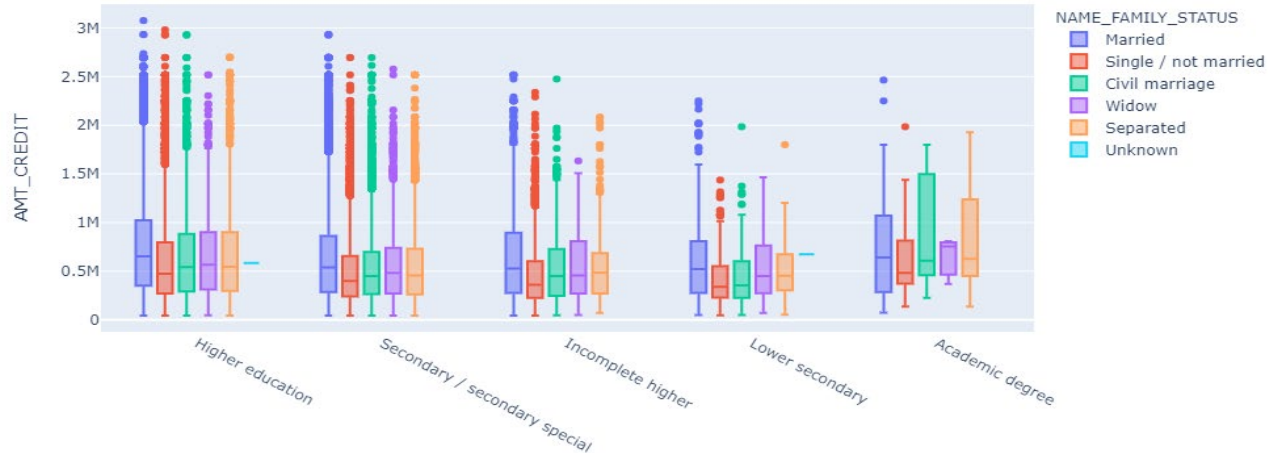
Bivariate Analysis between Defaulter & Non- Defaulter

- There are very few outliers between AMT CREDIT and AMT ANNUITY
- As the AGE increases, credit amount also increases
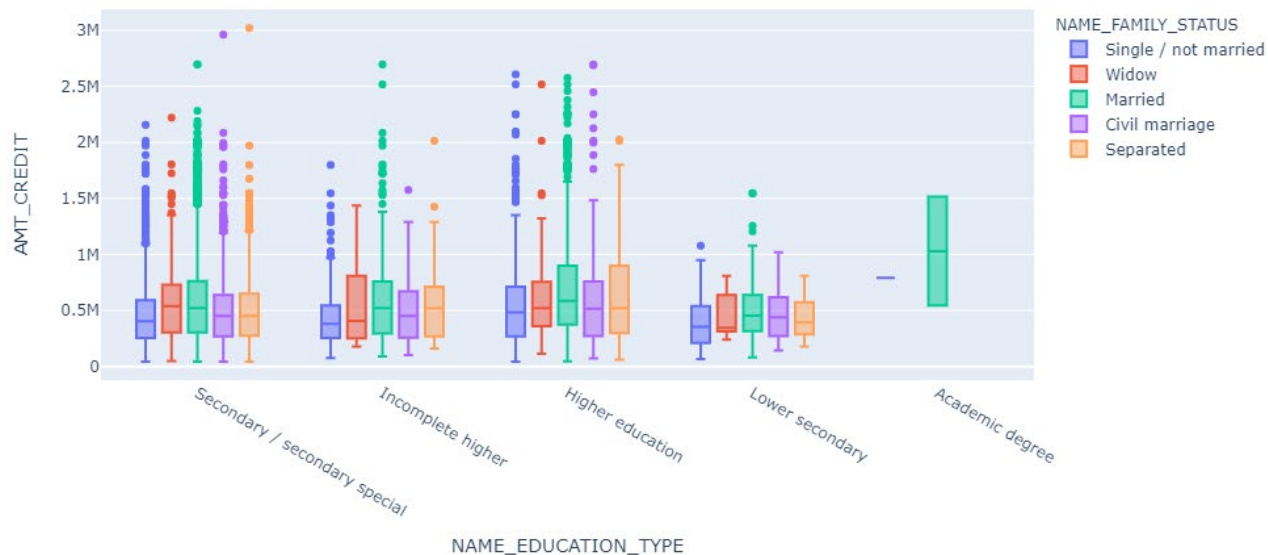- Income of the client increases with age

Bivariate Analysis - NAME_EDUCATION_TYPE vs AMT_CREDIT

Credit amount vs Education of Loan - Non Defaulters

Credit amount vs Education of Loan Defaulters

- The both graphs for Loan Payment Defaulters & Loan Non-Payment Defaulters appears to be similar.

- We say that Family status of civil marriage, marriage & separated of Academic degree education are having higher number of credits than others.

- Most of the outliers are from Education type Higher education and Secondary. Civil marriage for Academic degree is having most of the credits in the third quartile.

# Correlation between Application Data – Defaulter & Non Defaulters
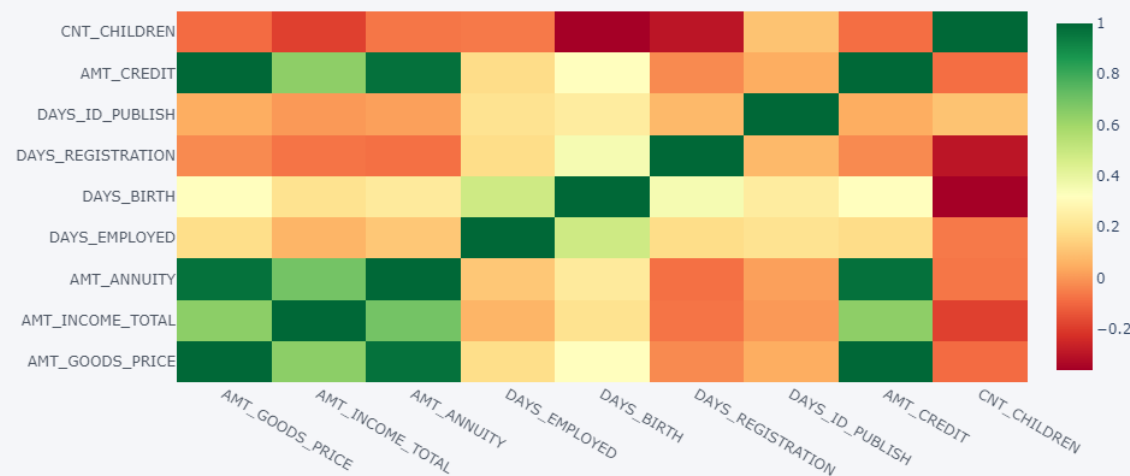
- We observe that there is a high correlation between credit amount and goods price. There appears to be some deviancies in the correlation of Loan- Payment Difficulties and Loan- Non Payment Difficulties such as credit amount v/s income.
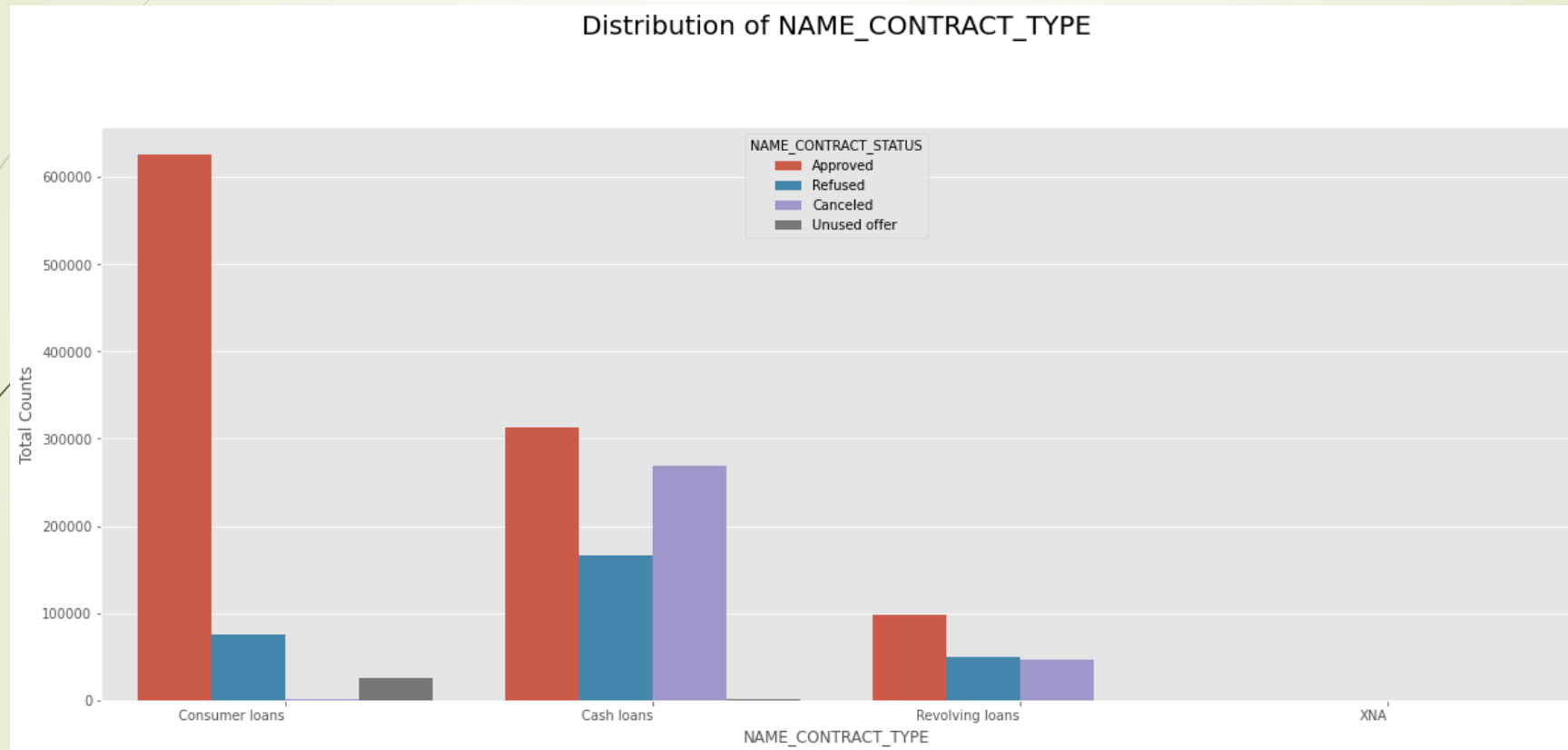
Univariate Analysis on Previous Application Data

Contract status of previous application
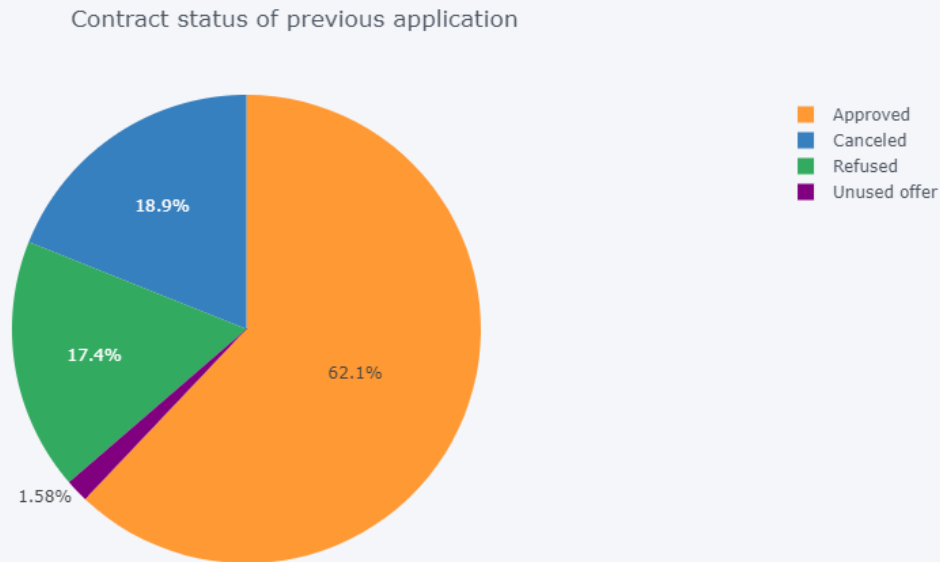
- We can observe most previous application Approved percentage is more higher than any other category.

- Unused offer are very less.

Contract Type of previous application

Distribution of NAME_CONTRACT_TYPE

- We can see that maximum consumer get approved but cash loan face some difficult get approved & refused & cancel also high.

Client Type - previous application
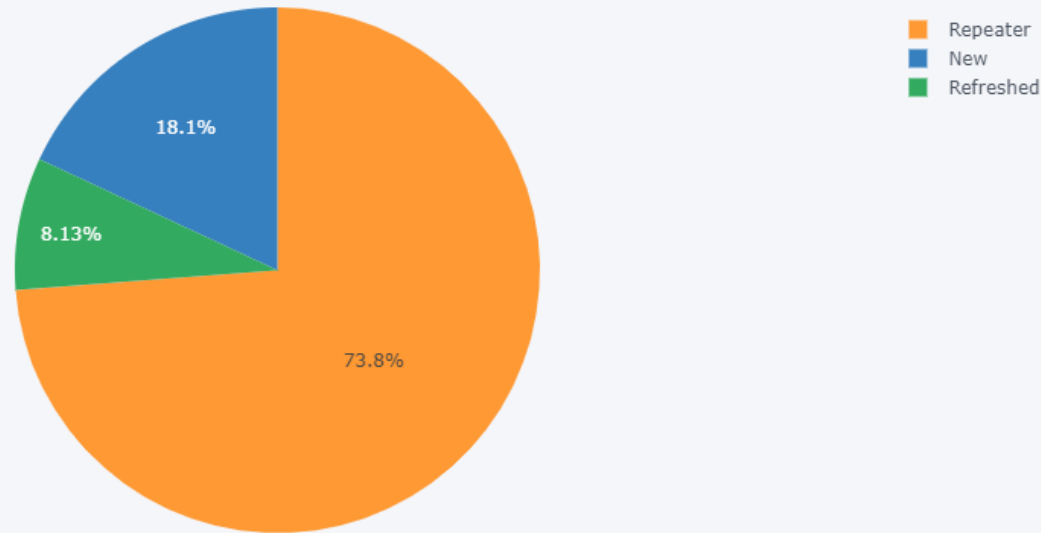

Distribution of NAME_CLIENT_TYPE
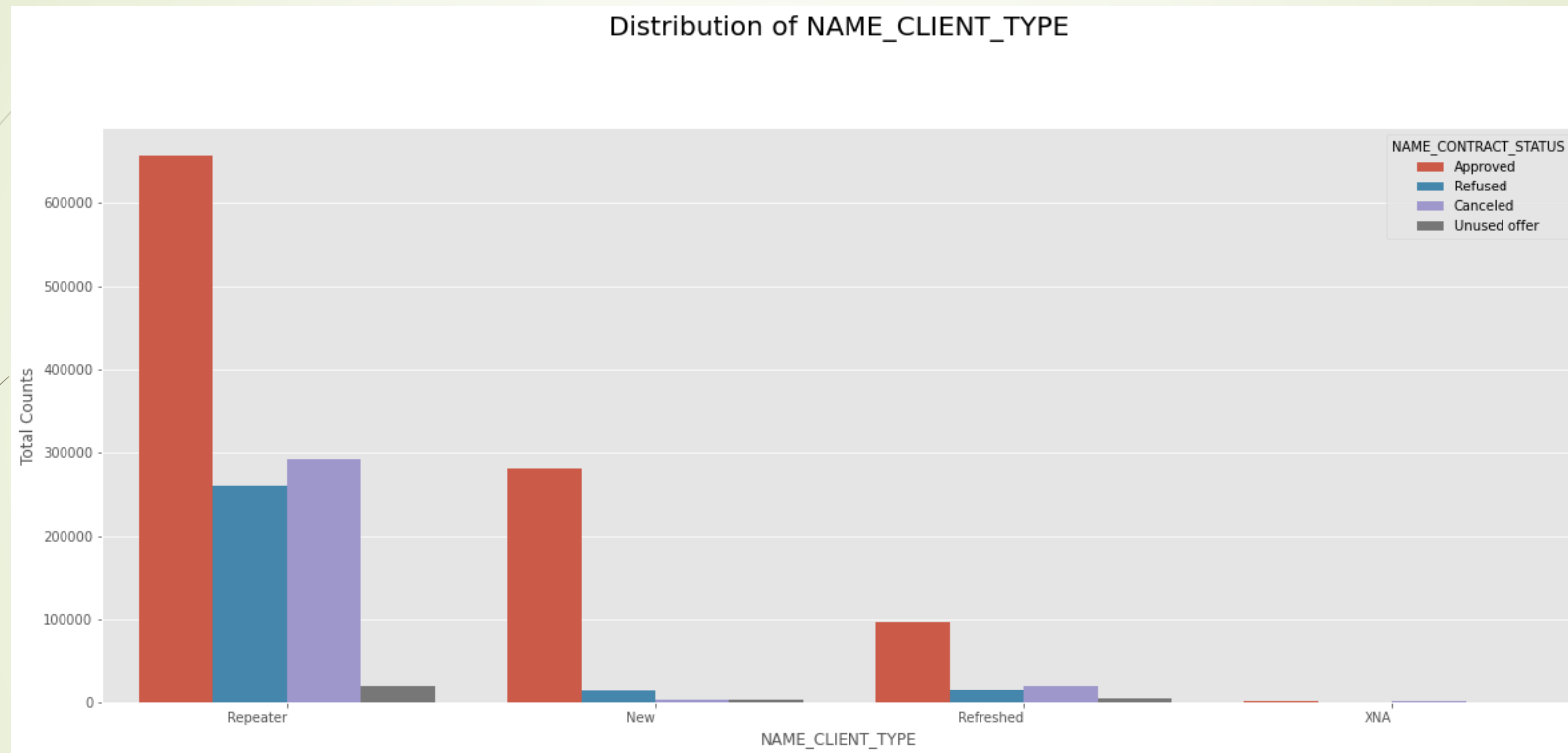
- We can observe from repeater maximum applicant get approved but there are still many refused & cancel applicant.

- Where New applicant & refreshed approved easily.

Previous application – old client or new client

Was the client old or new client when applying for the previous application

- Repeater
- New
- Refreshed

18.1%

8.13%

73.8%

- We can observe that maximum client are repeater.

- 18 % client are new.

- Only 8 % client are refreshed.

- So because bank have repeater client they have repeater client data also.

- They can give to repeater loan by observing client data.

GOODS CATEGORY – Previous application

What kind of goods did the client apply for in the previous application
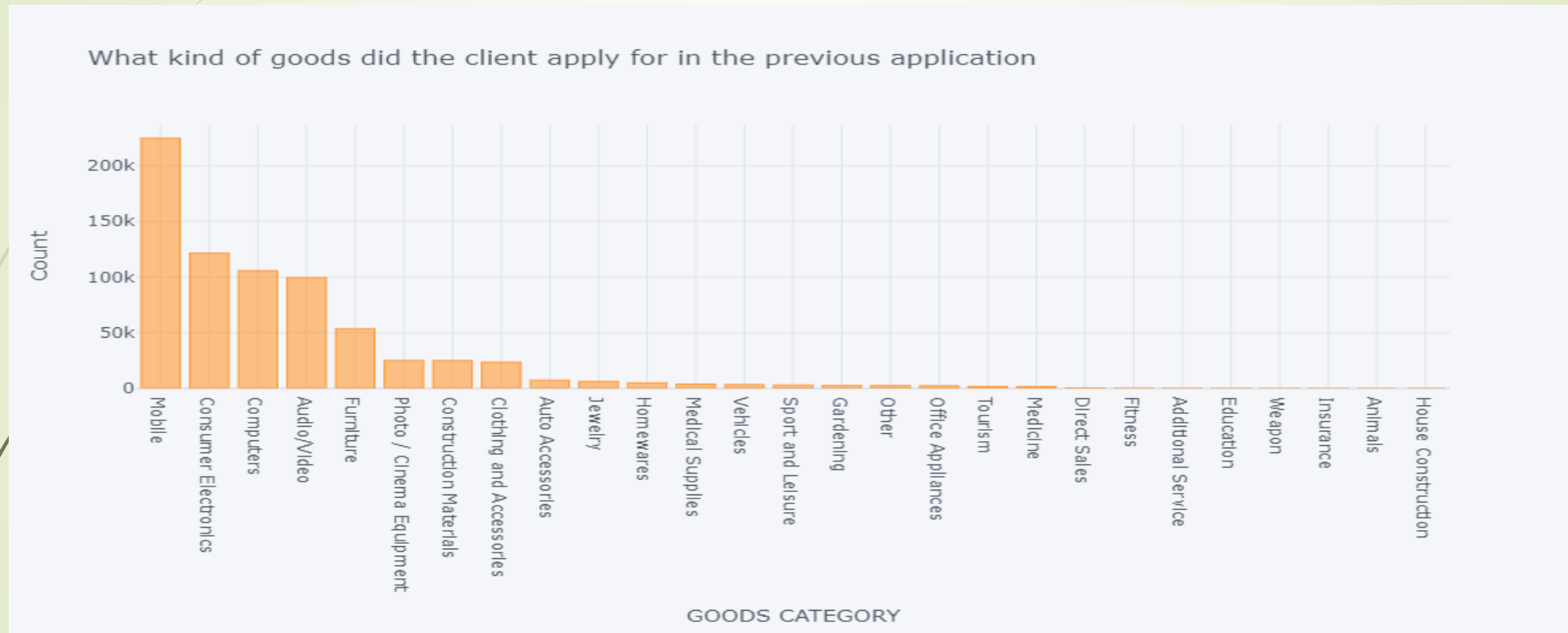
- We can see that maximum loans are for mobile, consumer, computers & audio/video category.
- We can say majority of loans taken for electronics product.

Reasons to Reject loans – previous application
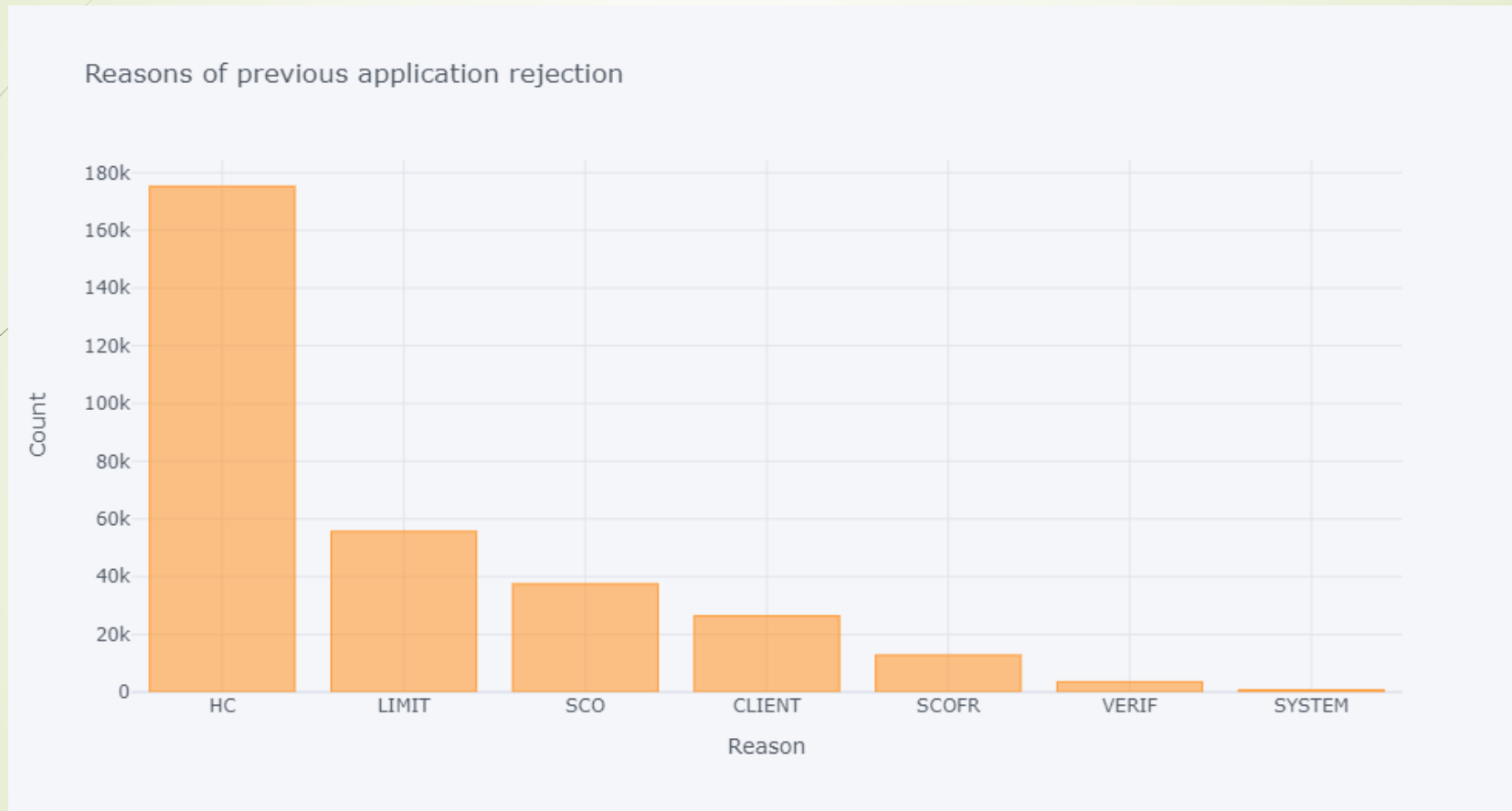
- We observe that HC is the reason majority of applications got rejected.
- Any details related to HC not given so we cannot say why HC rejection rate is high.

OWN CAR vs NAME CONTRACT STATUS
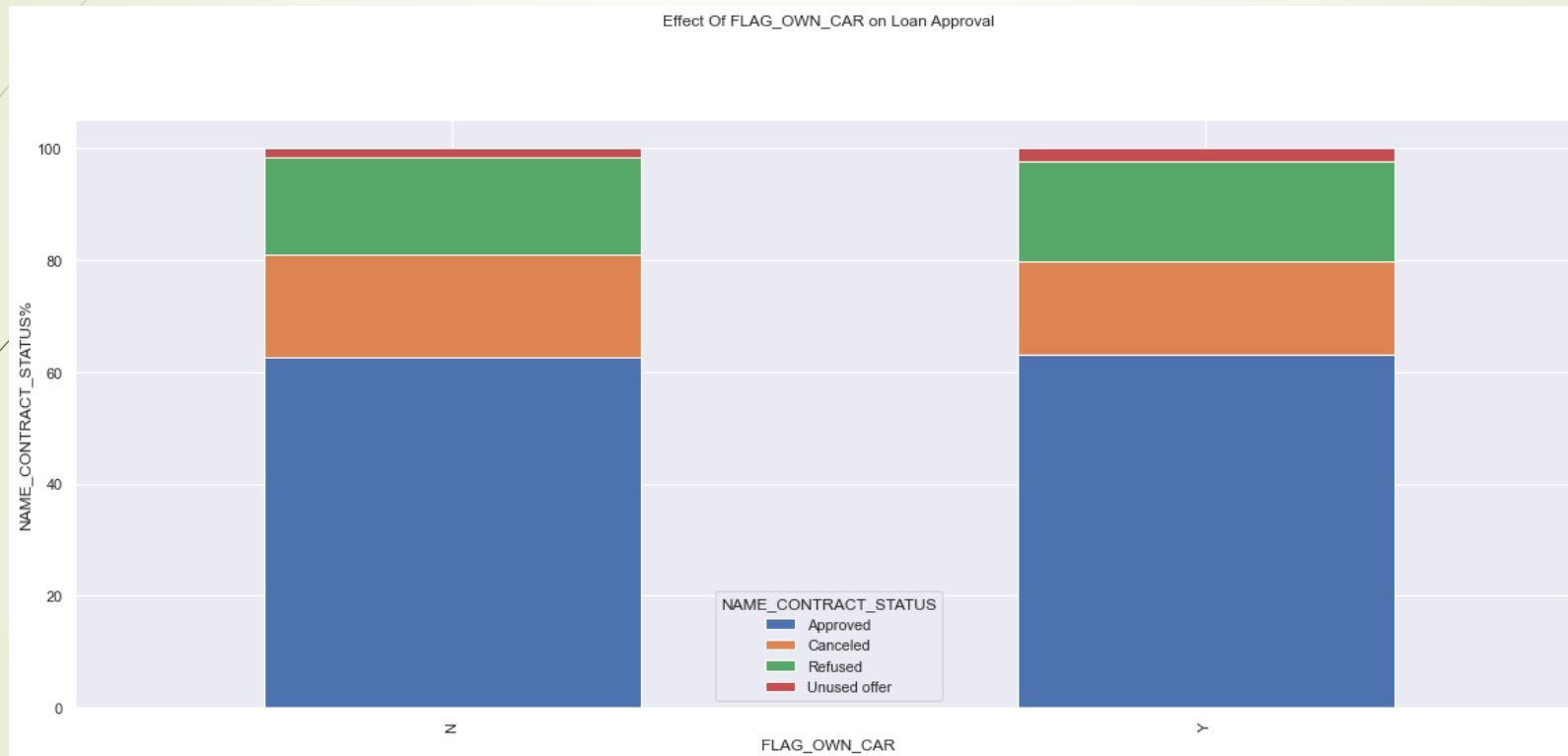
- We see that car ownership doesn't have any effect on application approval or rejection.
- But we saw earlier that the applicant who has a car has lesser chances of default. The bank can add more weightage to car ownership while approving a loan amount

CODE GENDER vs NAME CONTRACT STATUS
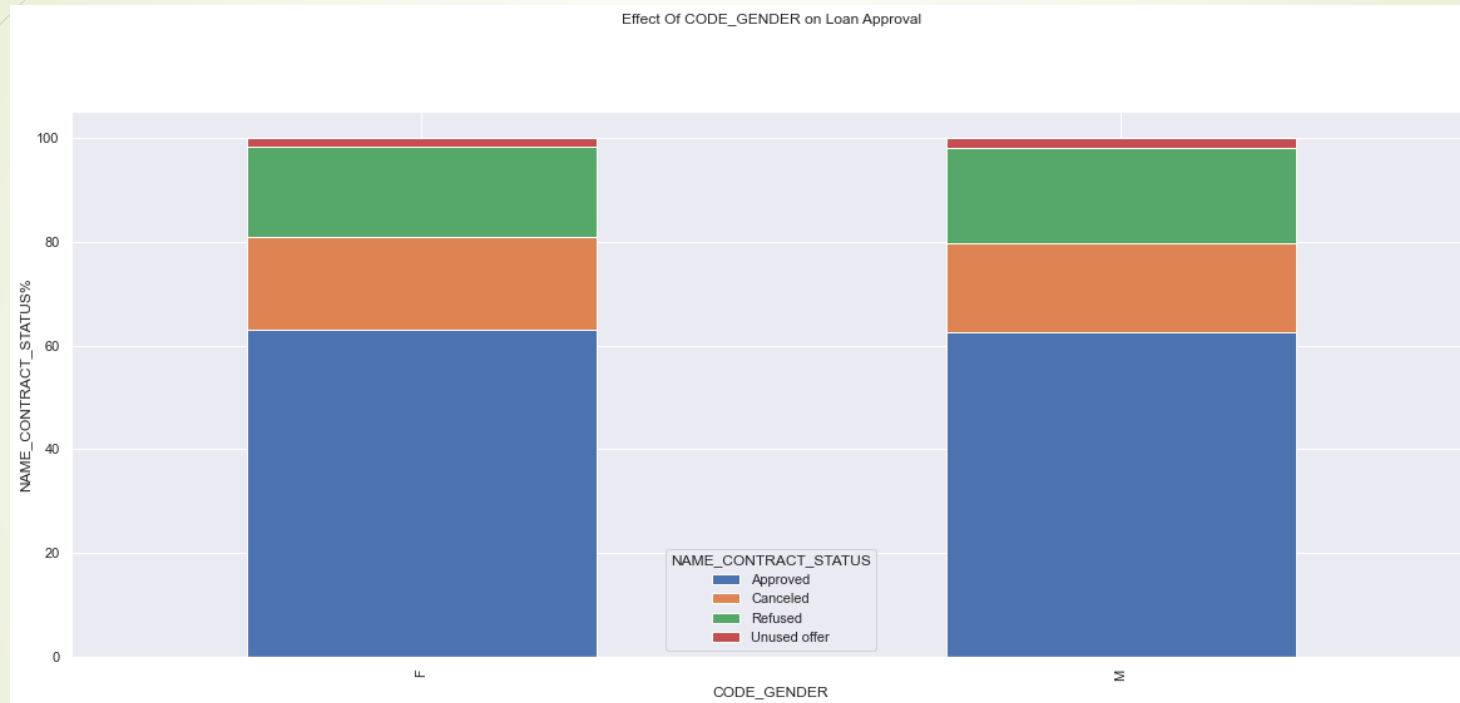
- We observe that code gender doesn't have any effect on application approval or rejection.
- But we observe earlier that female have lesser chances of default compared to males.
- The bank can add more weightage to female while approving a loan amount.

Defaulter & Non-Defaulter vs Contract status

Effect Of TARGET on Loan Approval
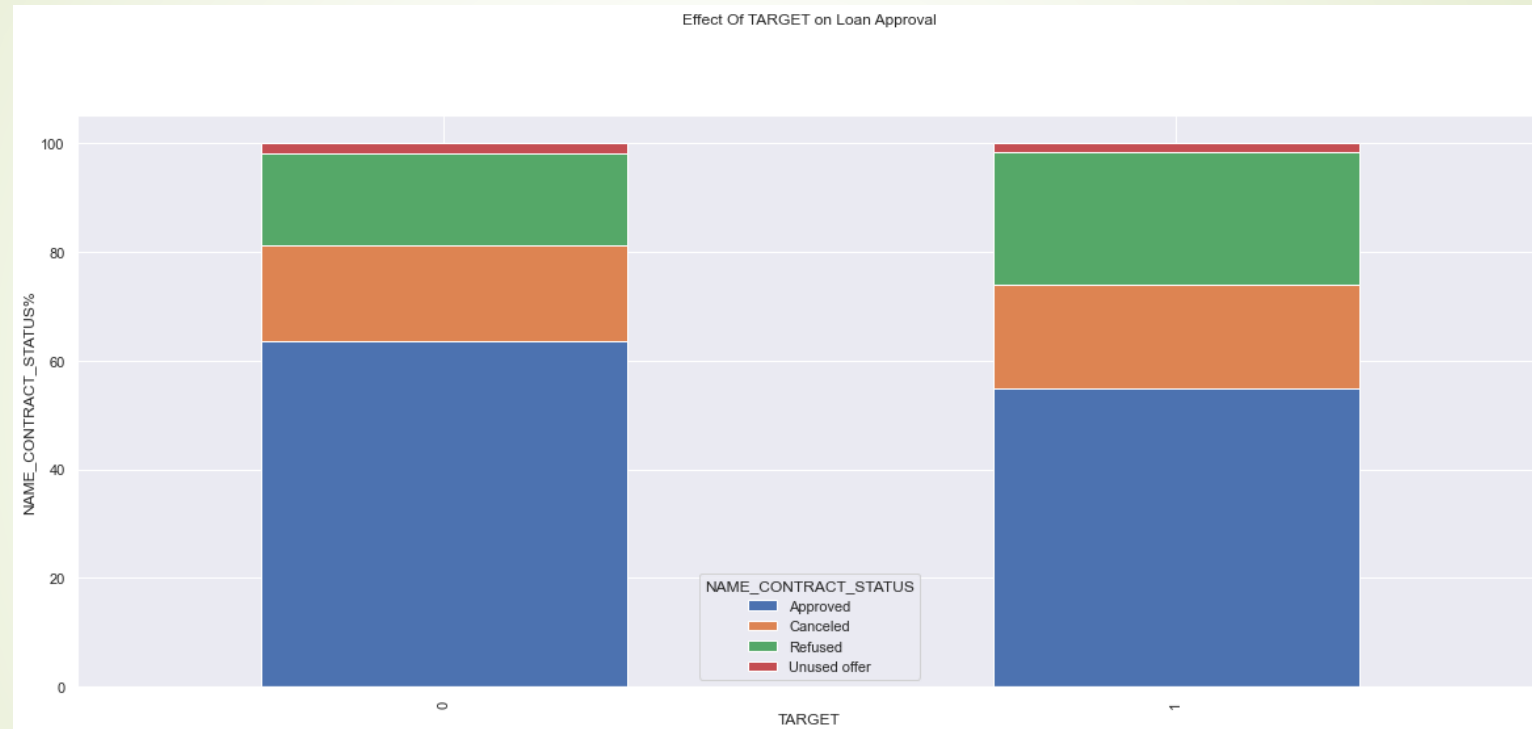
- We observe that the people who were approved for a loan earlier are defaulted less often where as people who were refused a loan earlier have higher chances of defaulting.

CONCLUSION

- Bank should focus on giving loan to higher education because they are less likely to default loan.

- Bank should give loan to higher income group because their chances to default the loan is less.

- *Give loan to clients from housing type With parents as they are having least number of unsuccessful payments.*

- Most of the previously cancelled client have actually repay the loan. Reason for cancellation which might help the bank.