

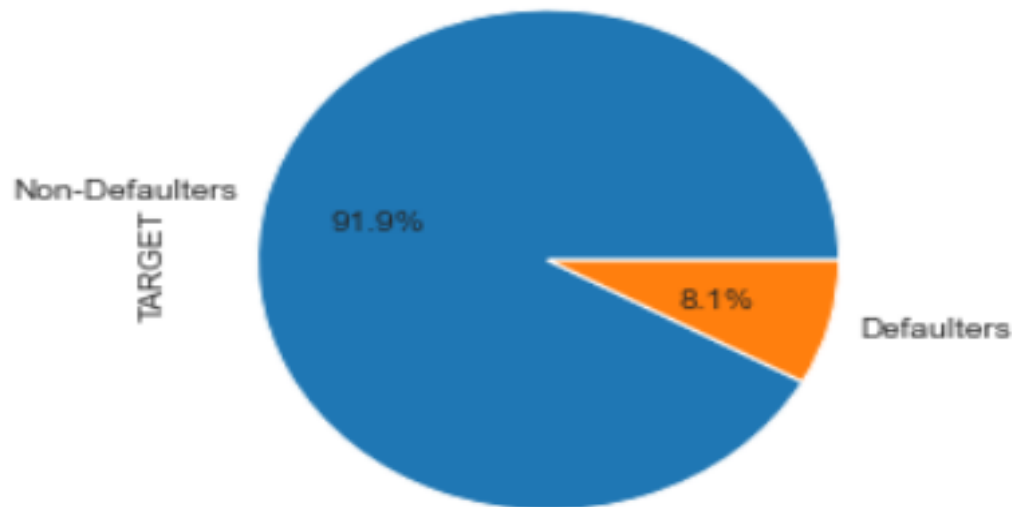
Credit EDA Case Study

BY
GAUTAM KODUKULA
NAVADEEP R JAGATHKARI

APPLICATION DATA

Uni Variate Analysis

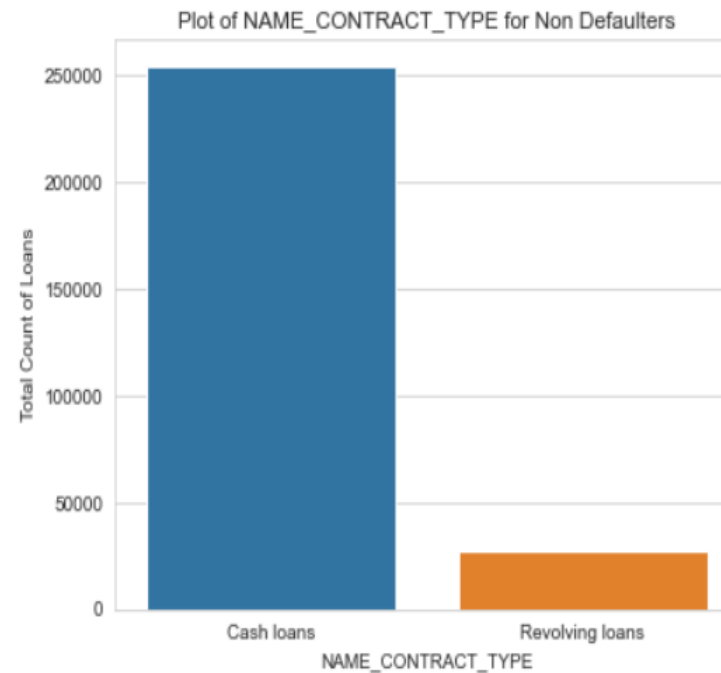
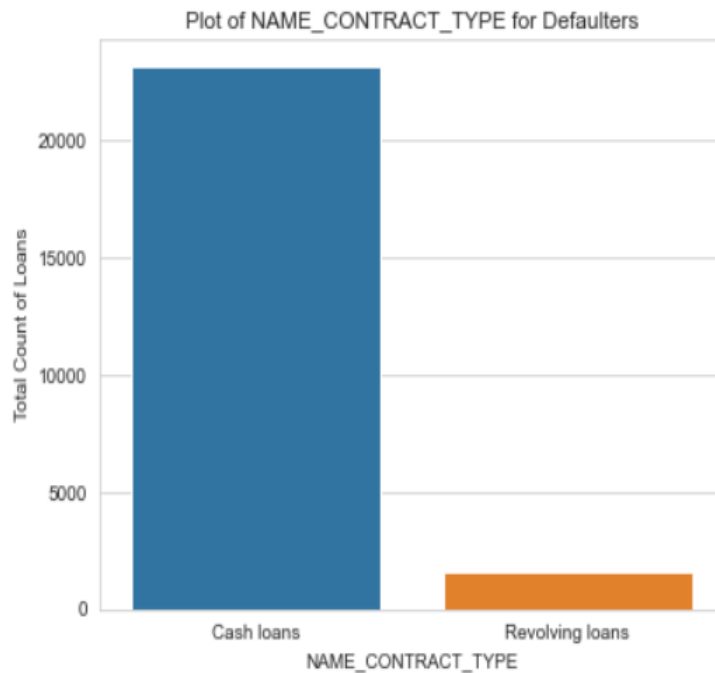
Distribution of Imbalance Ratio



Points can be inferred from the pie plot:

- The Non-Defaulter are more than Defaulters
- The Non-defaulters percentage is 91.9% and defaulters percentage is 8.1%

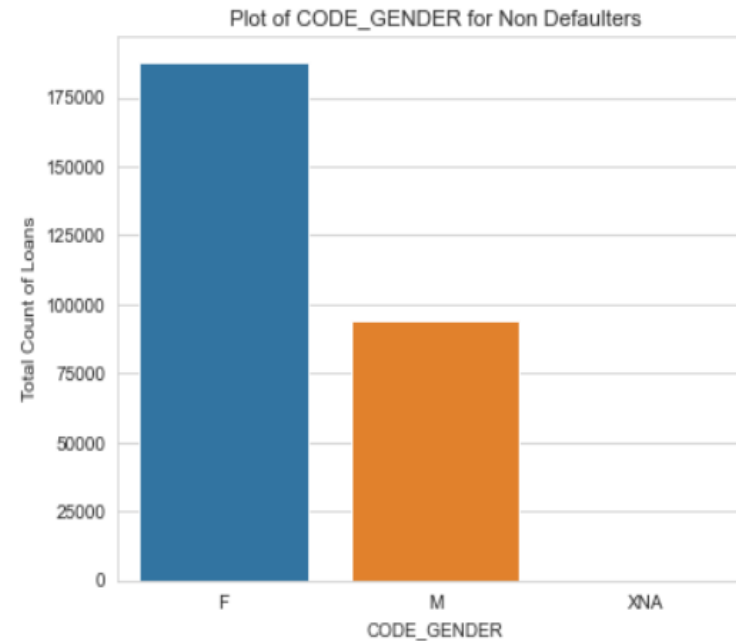
Count plot for Type of loan



The points can be inferred from the count plot:

- The Cash loans are popular than Revolving loans.
- The Non-Defaulters and Defaulter are more in Cash loans itself.

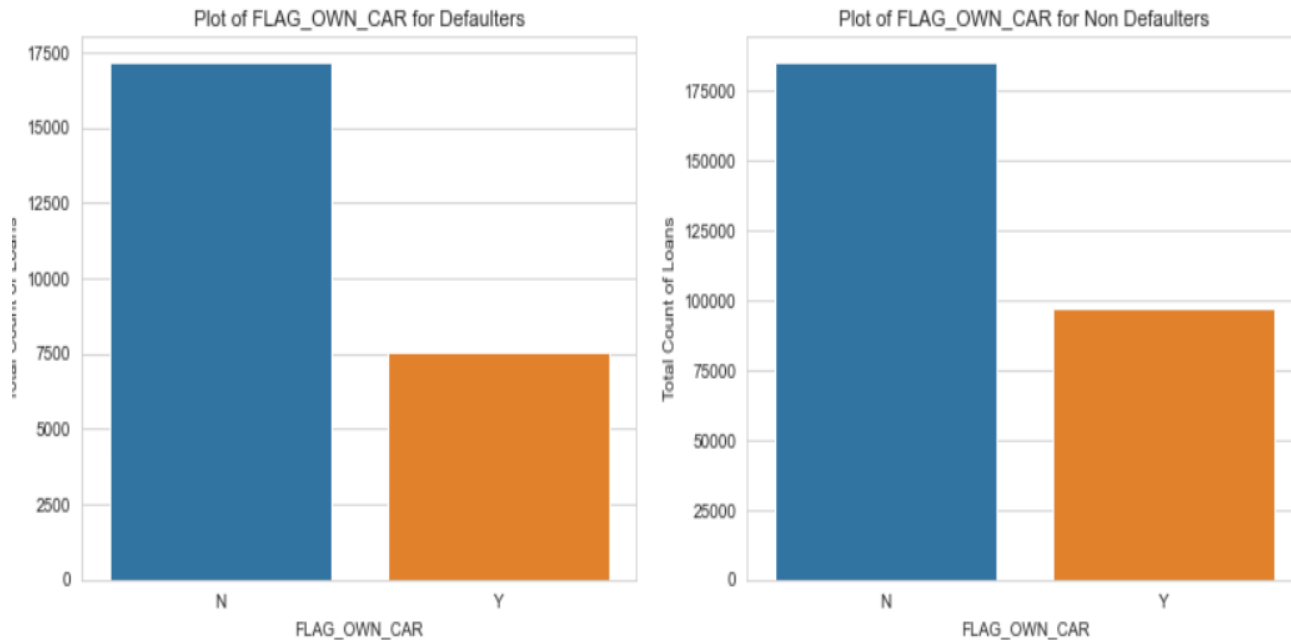
Count plot for Gender of the client



The points can be inferred from the count plot:

- Count of Loan transactions are more in females.
- The percentage of defaulters in females with respect to total female count is less.
- The percentage of defaulters is more in males with respect to total male count.

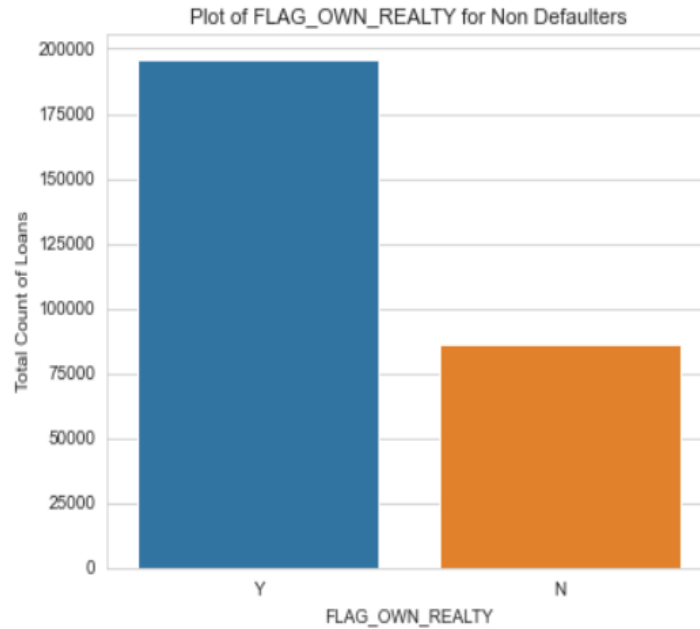
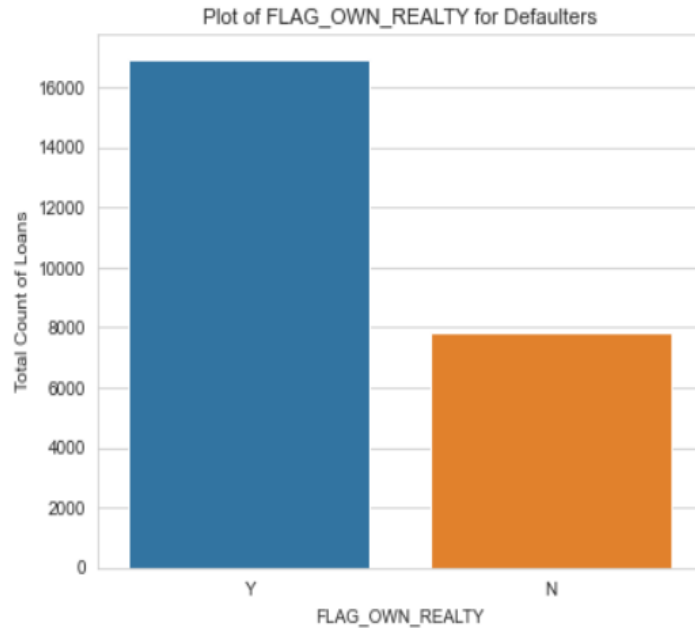
Count plot for if the client owned a car



The points can be inferred from the count plot:

- The clients who own a car are approximately half to those who don't own.
- The Defaulters are more in the group of clients who don't own a car.

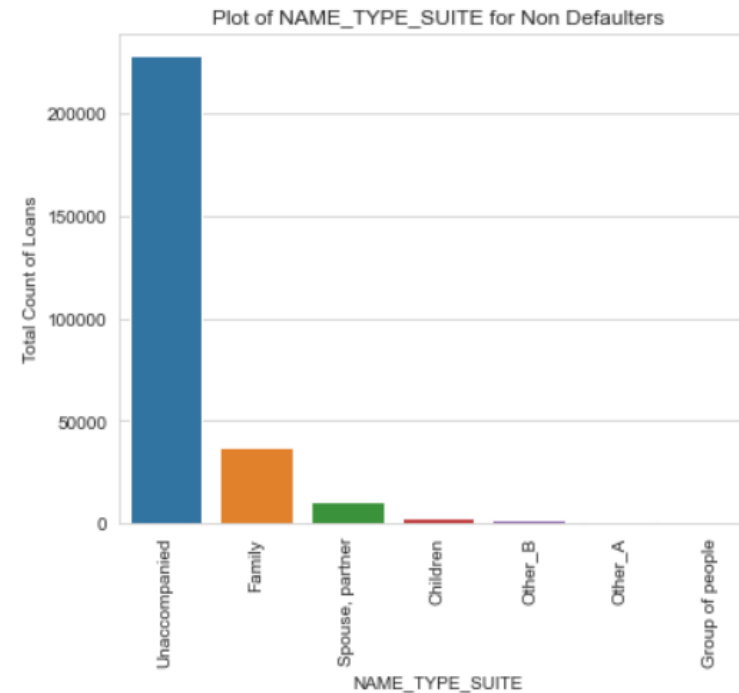
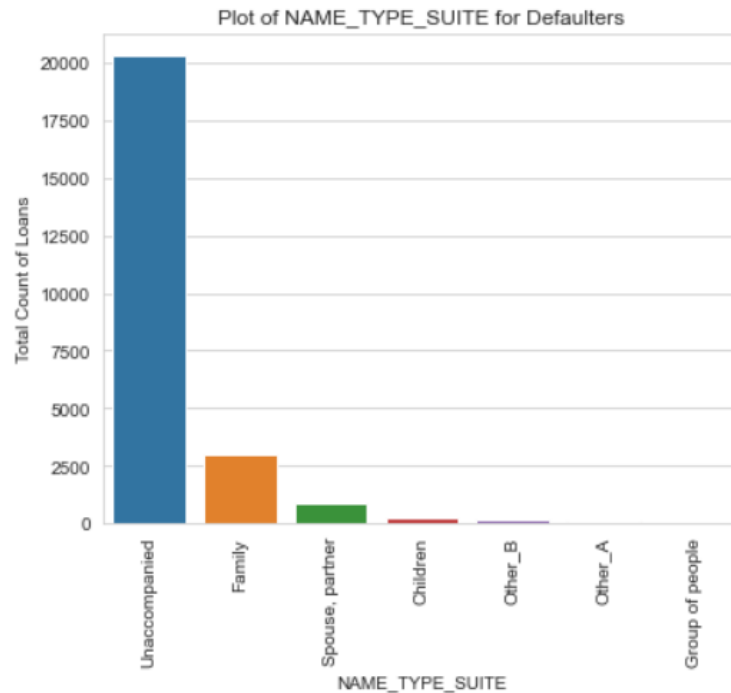
Count plot for if the client owned a house or flat



The points can be inferred from the plot:

- The clients who own a house/flat are more than who don't own.
- The defaulters are also more in clients who own a flat.

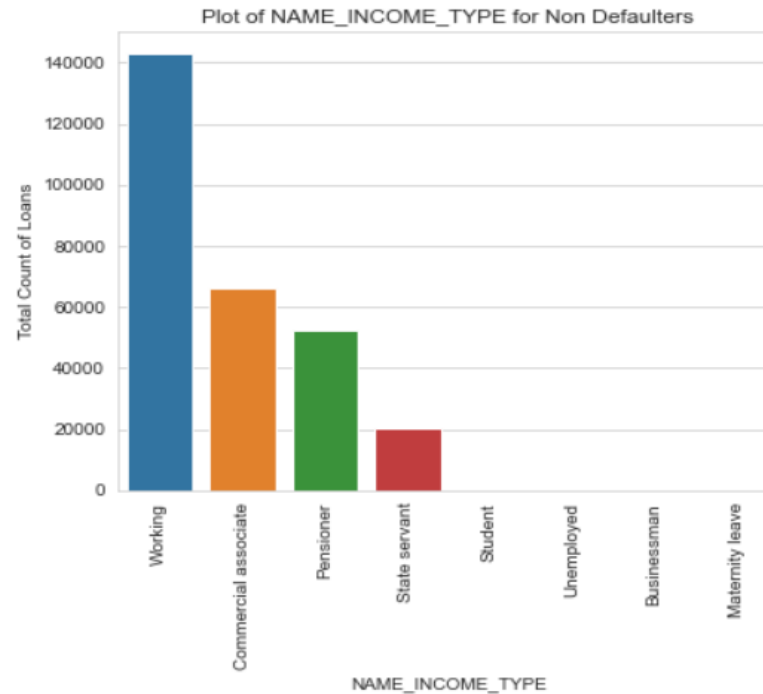
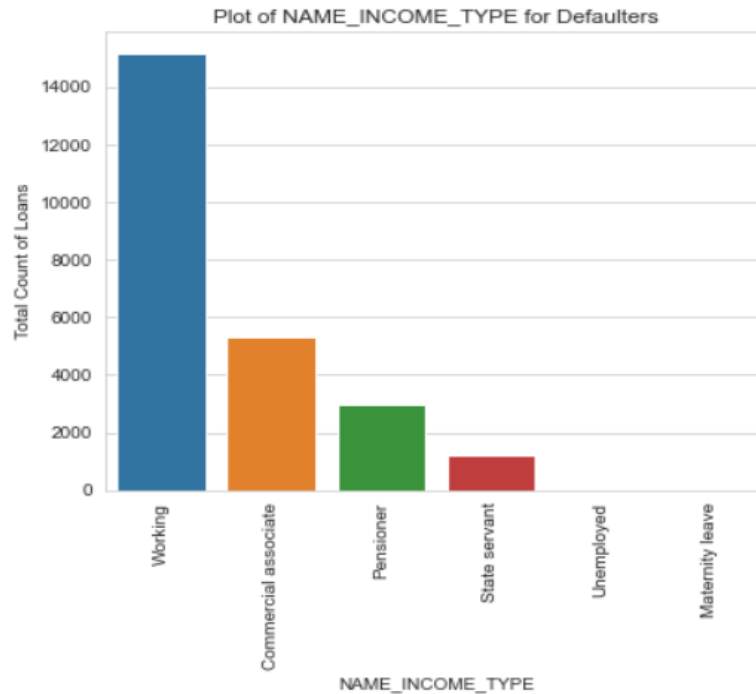
Distribution of client accompanied while applying loan



The points can be inferred from the plot:

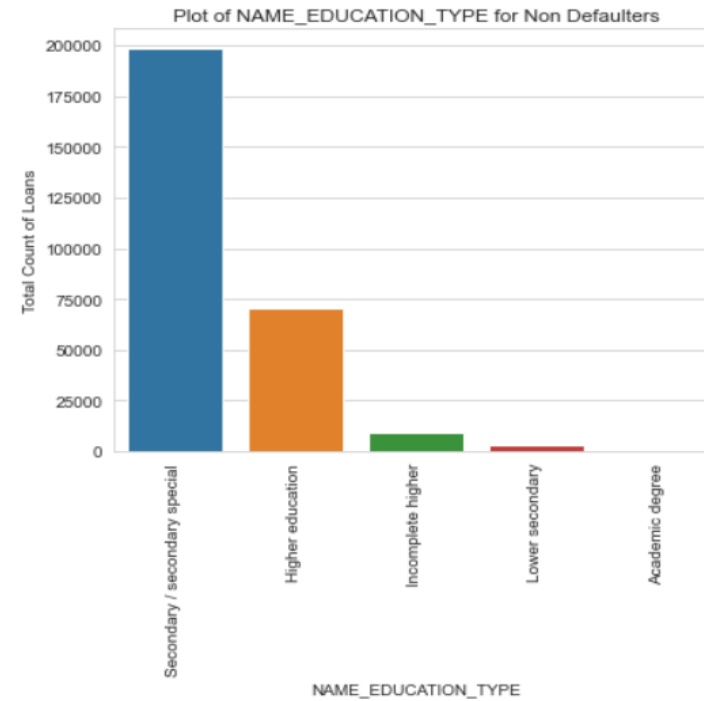
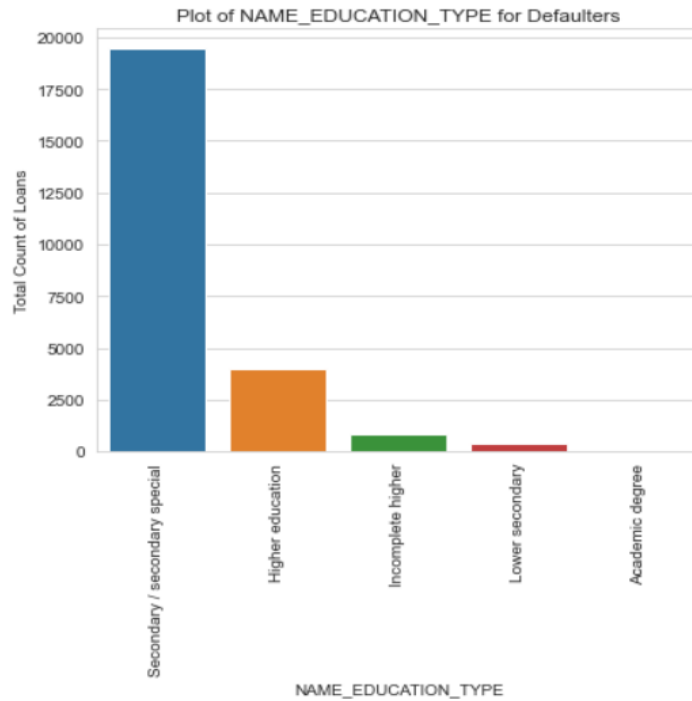
- The clients who don't accompany anybody while applying the loan are more compare to any group in both defaulters and non defaulters plot.
- The client who accompany with family while applying the loan are second highest in both defaulters and non defaulters plot.

Distribution of income type of the clients



- The points can be inferred from the plot:
- The clients in the working group are more in both defaulters and non defaulters plot.
 - The clients from maternity leave come out to be least.

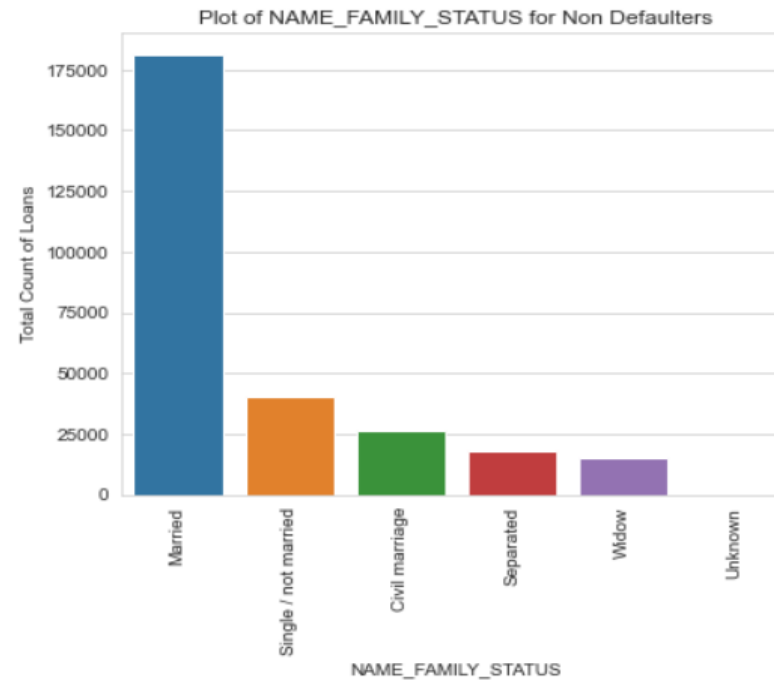
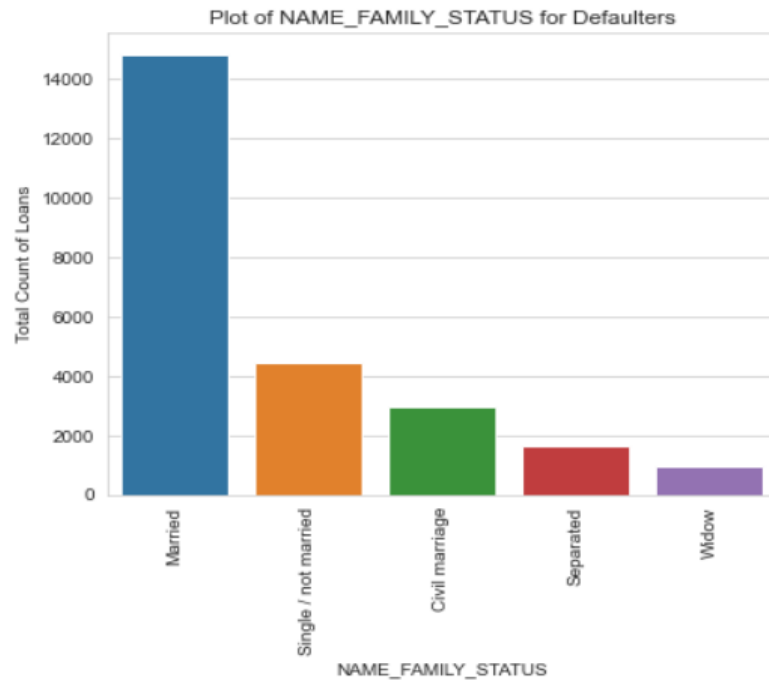
Distribution of educational status of the clients



The points can be inferred from the plot:

- The clients with educational background of secondary education are more than any educational status group.
- Secondary education is followed by higher education and so on.,

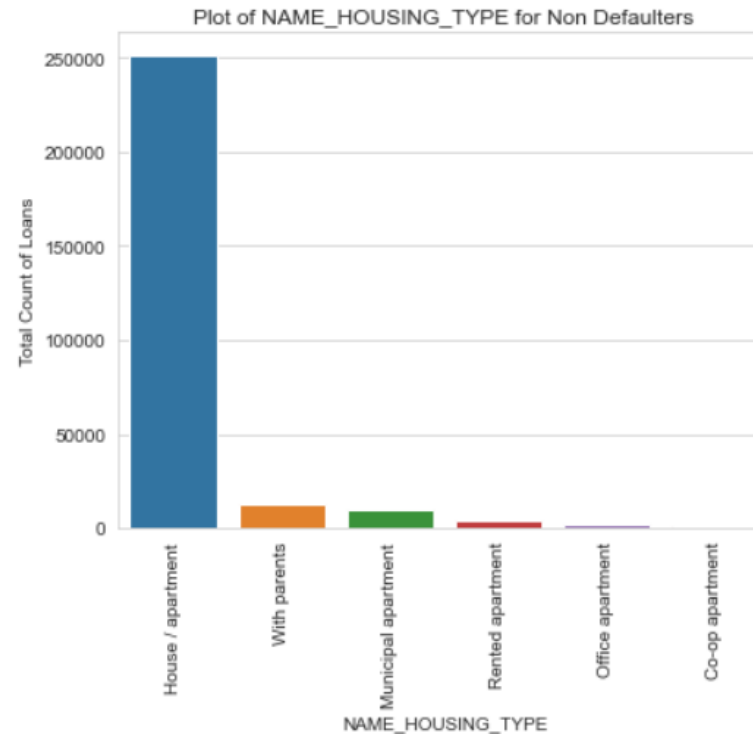
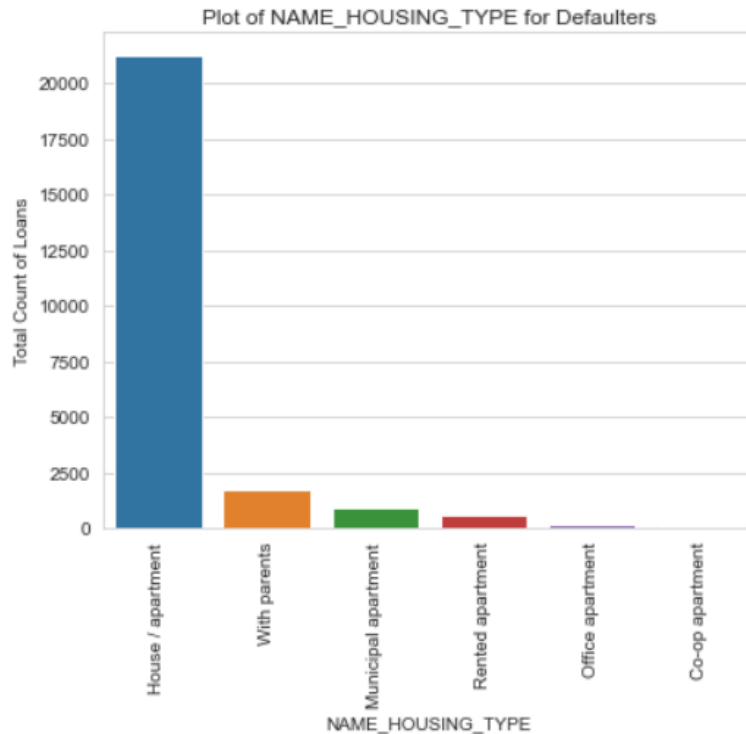
Distribution of family status of the clients



The points can be inferred from the graph:

- The clients who are married are the highest loan transaction group
- Married group is followed by single and so on.,

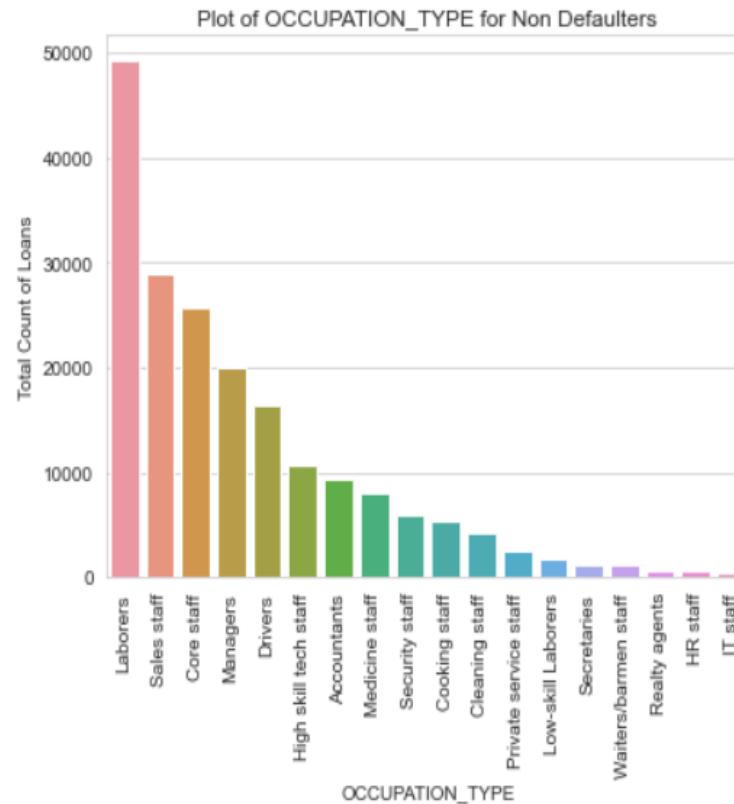
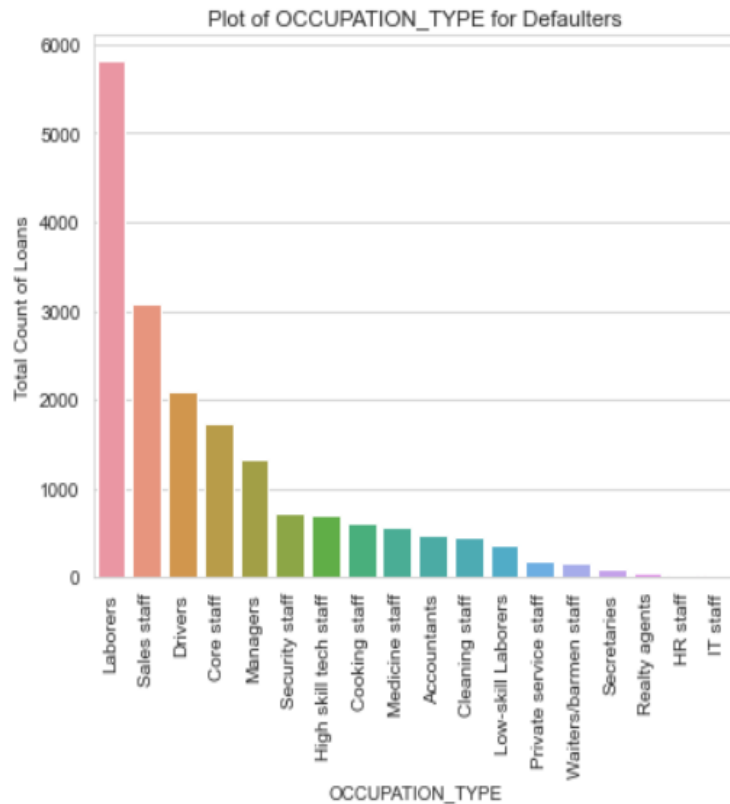
Distribution of housing situation of the clients



The points can be inferred from the plot:

- The clients who reside in own house/apartment with without parents tends to apply for more loans than any group.
- The same group came out to be more defaulters and non-defaulters.

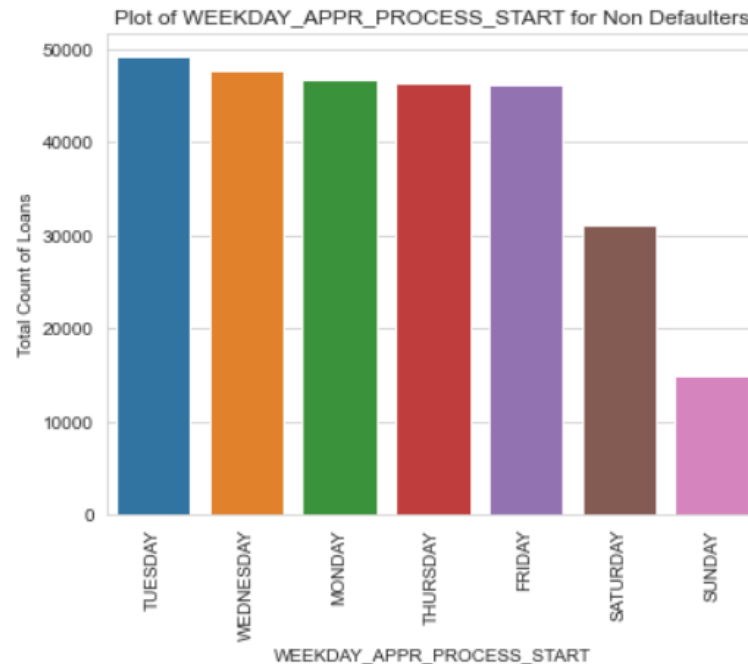
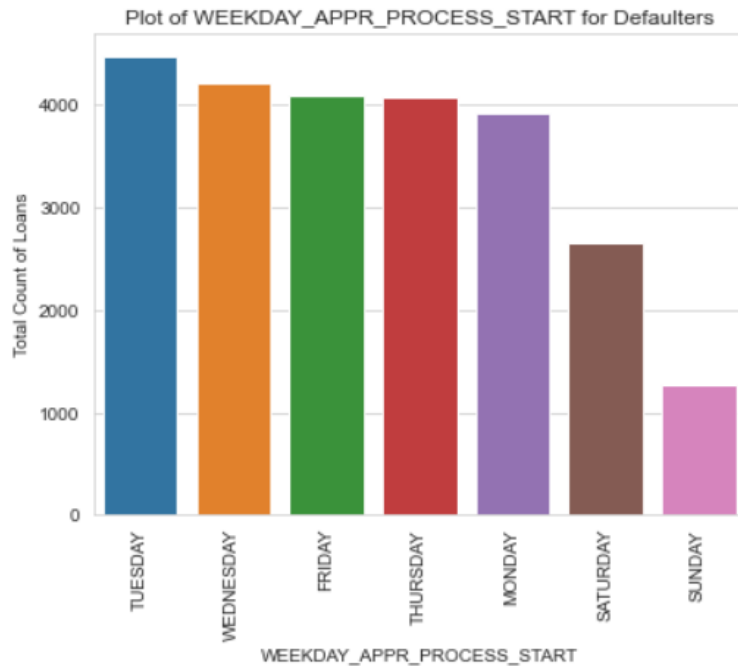
Distribution of Occupation Type



Points can be inferred from the graph:

- Labourers counts is more in both Defaulters and Non-Defaulters.

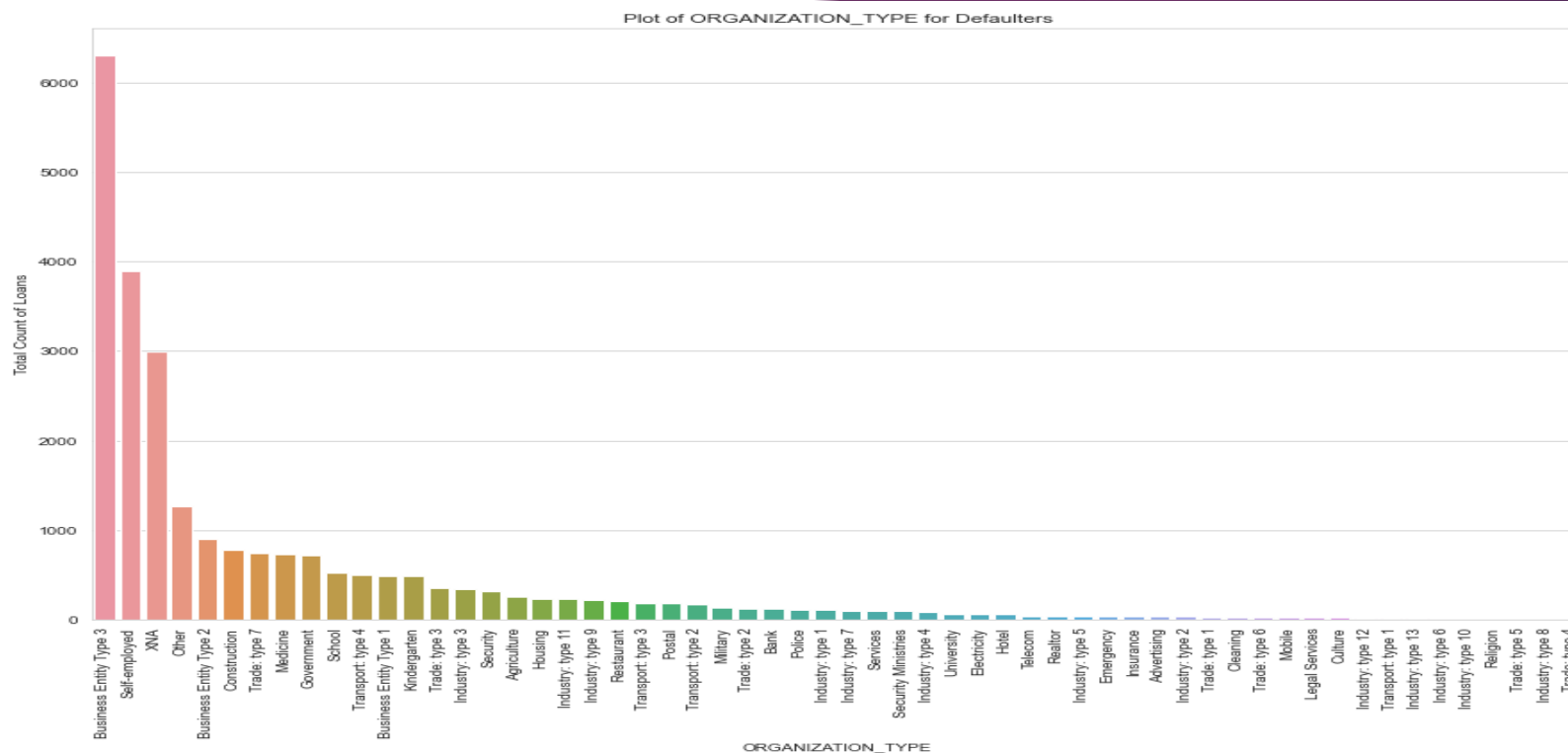
Distribution of Days of the week, clients applied for loans



The points can be inferred from the graph:

- Tuesday is most loan applied day in the week.
- As expected, the Sunday came out to be least.

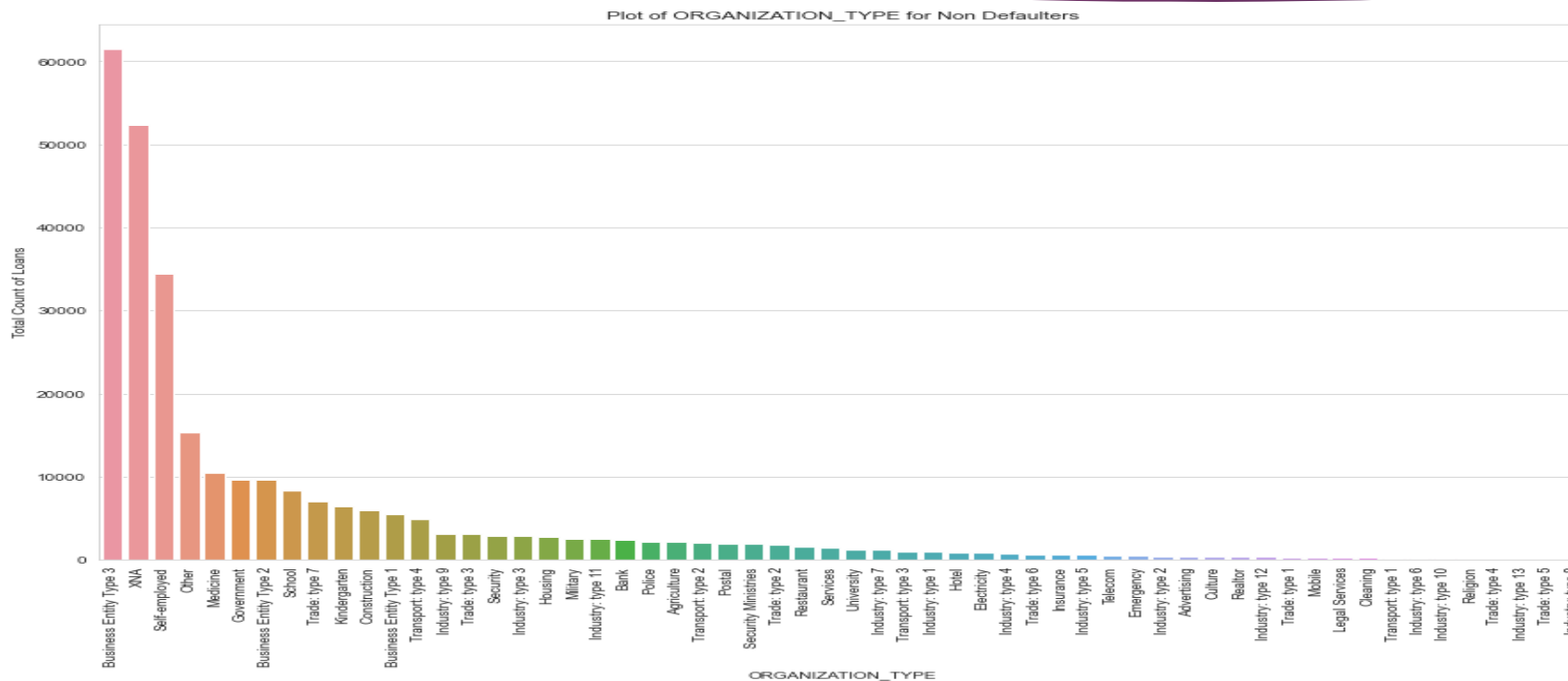
Distribution of type of organization(Defaulters)



The points can be inferred from the plot:

- The Business entity Type-3 organization is the highest defaulter.
- The self employed group is also one of the highest defaulters.

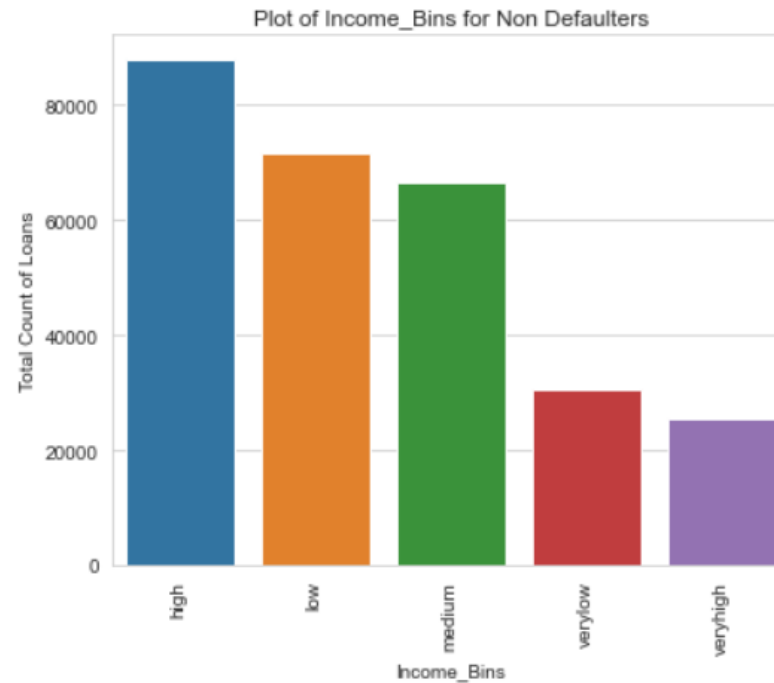
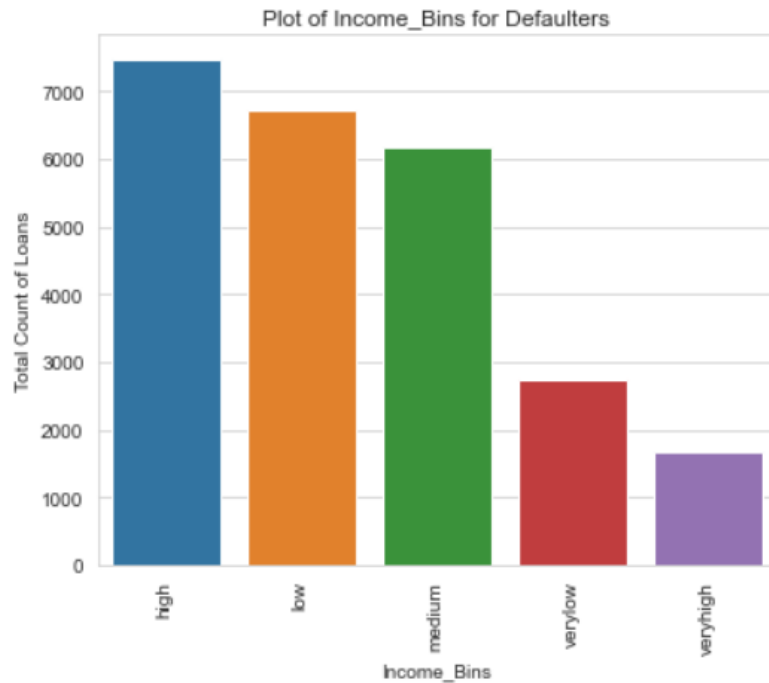
Distribution of type of organization(Non-Defaulters)



The points can be inferred from the group:

- The Business entity type-3 organization group has highest non-defaulters count followed by self-employed.

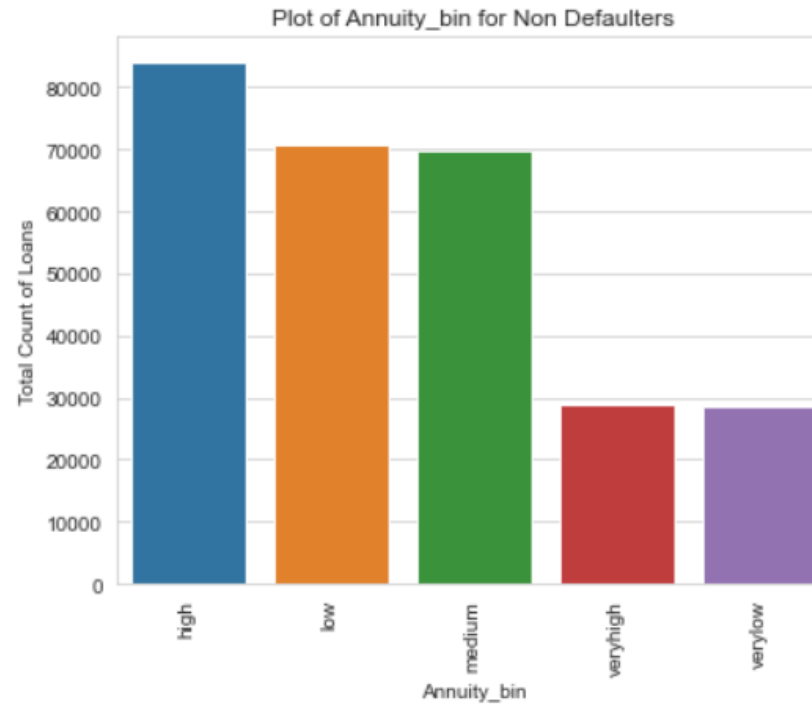
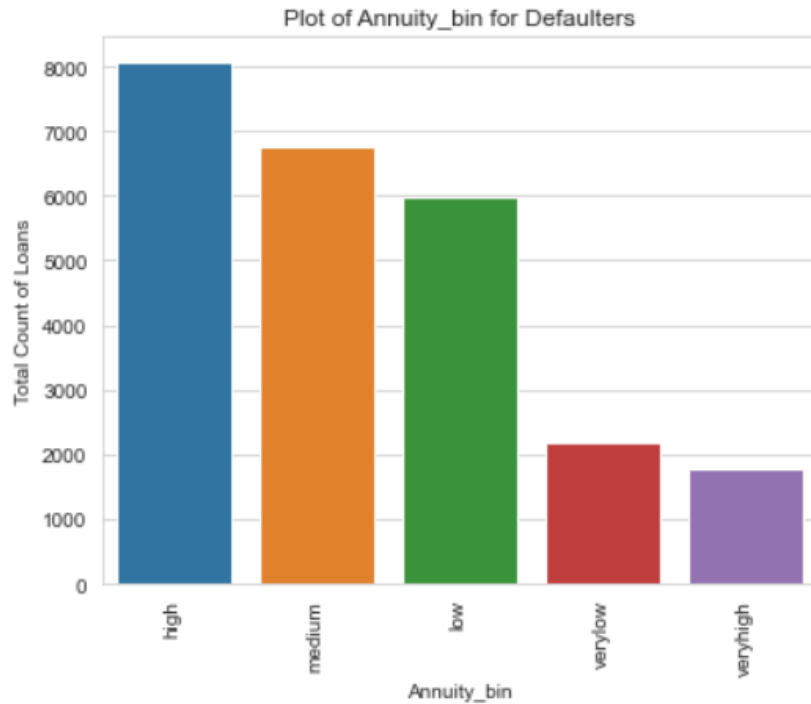
Distribution for income range of clients



The points can be inferred from the graph:

- The range of high income group has the highest loan transaction count.
- The same group came out to be highest defaulters.
- The very high income group is the lowest in both defaulters and non-defaulters.

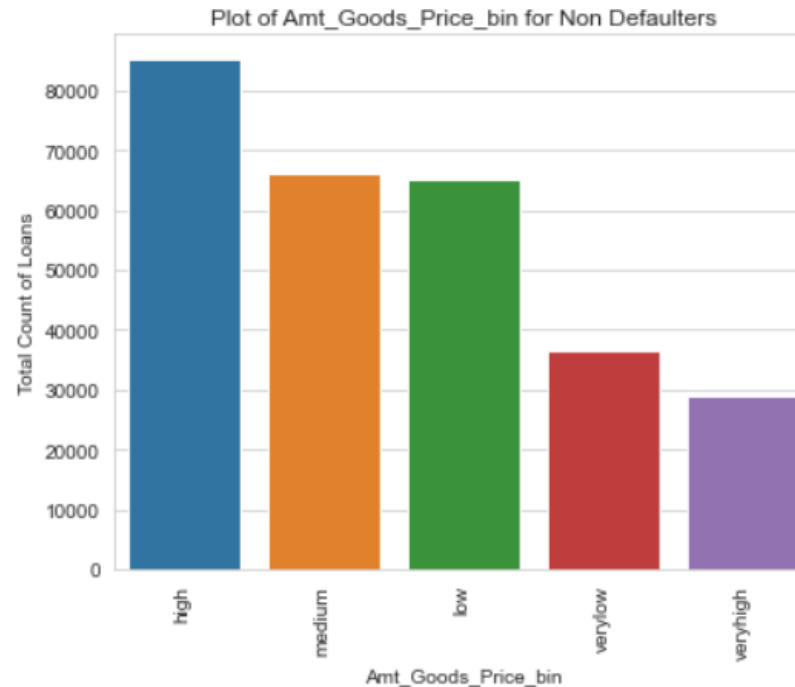
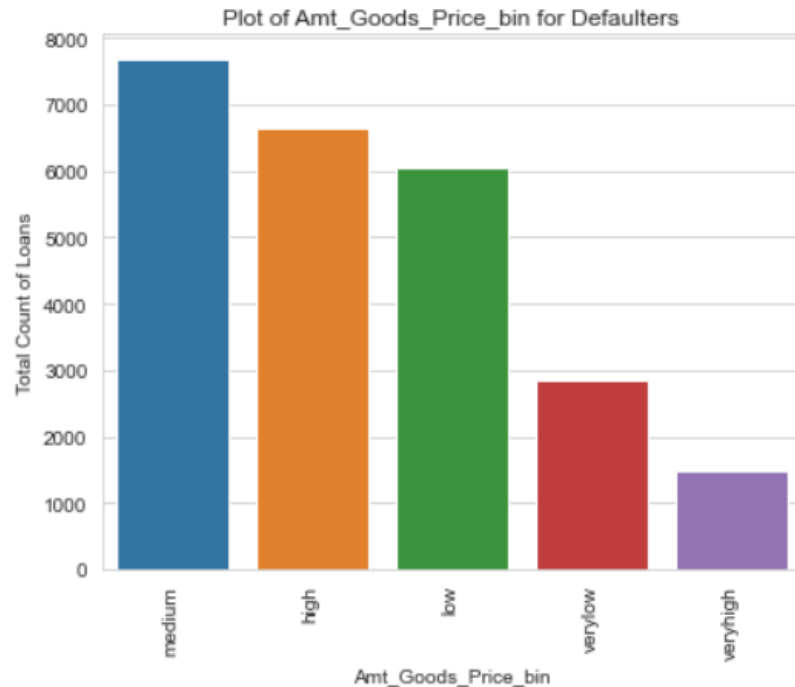
Distribution of loan annuity



The points can be inferred from the plot:

- The high Loan annuity group came out to be the highest loans receiver.
- The same group has highest loan defaulter count.

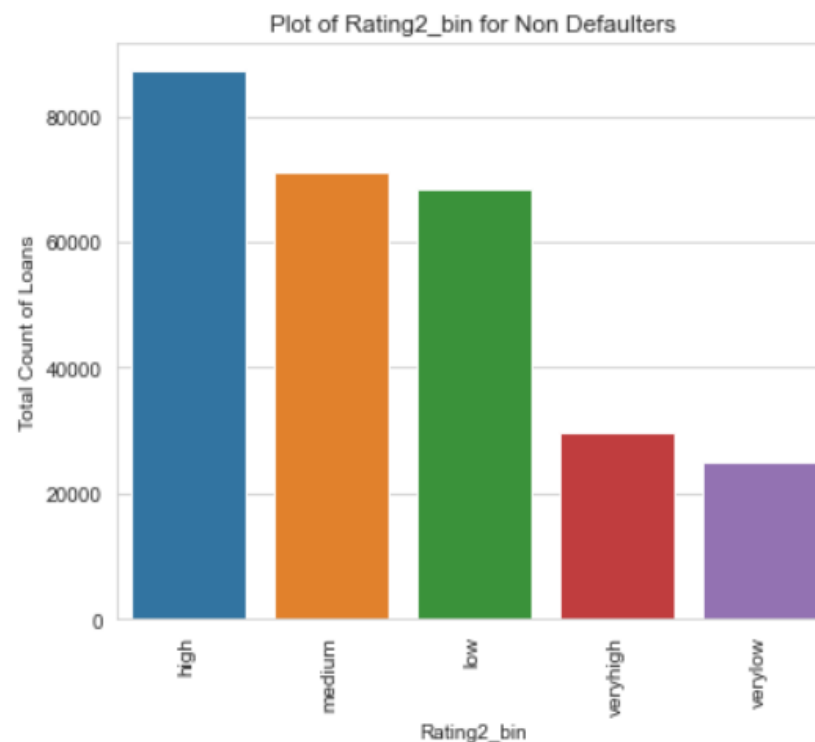
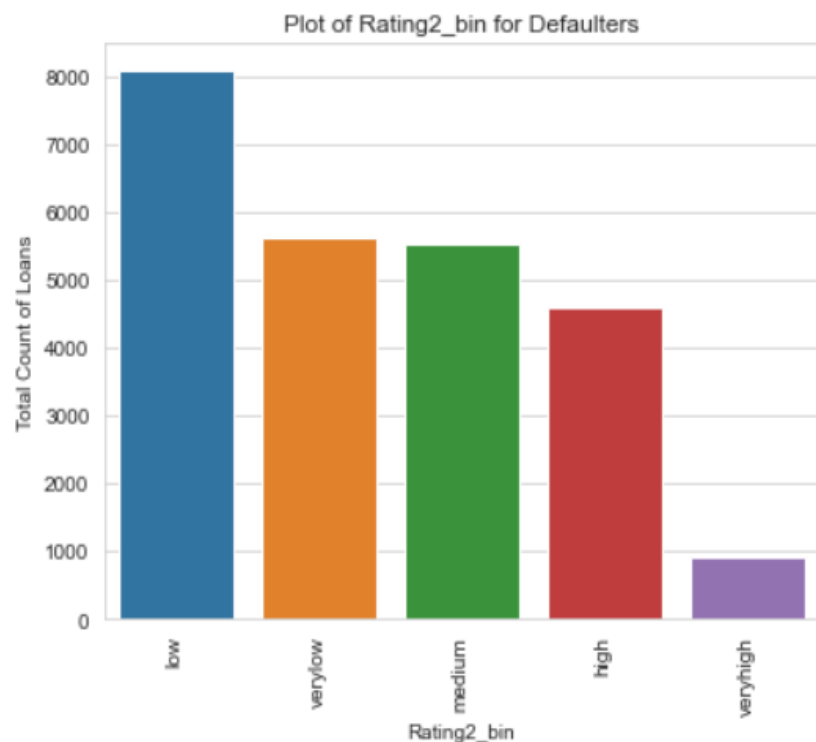
Distribution of goods price for which loans are given



The points can be inferred from the plot:

- Surprisingly, Medium goods price group are the highest loan Defaulters.
- The high good prices group are the highest non-defaulters.

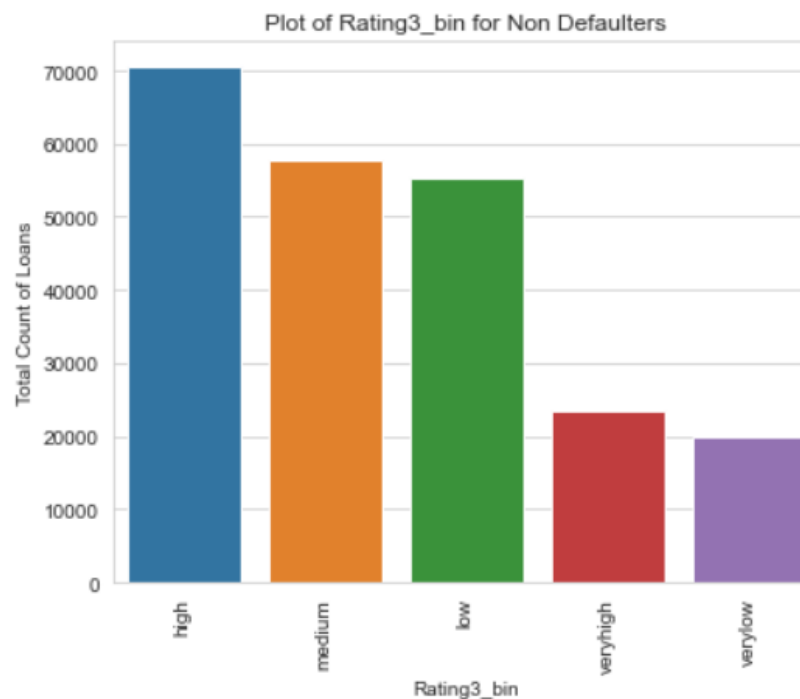
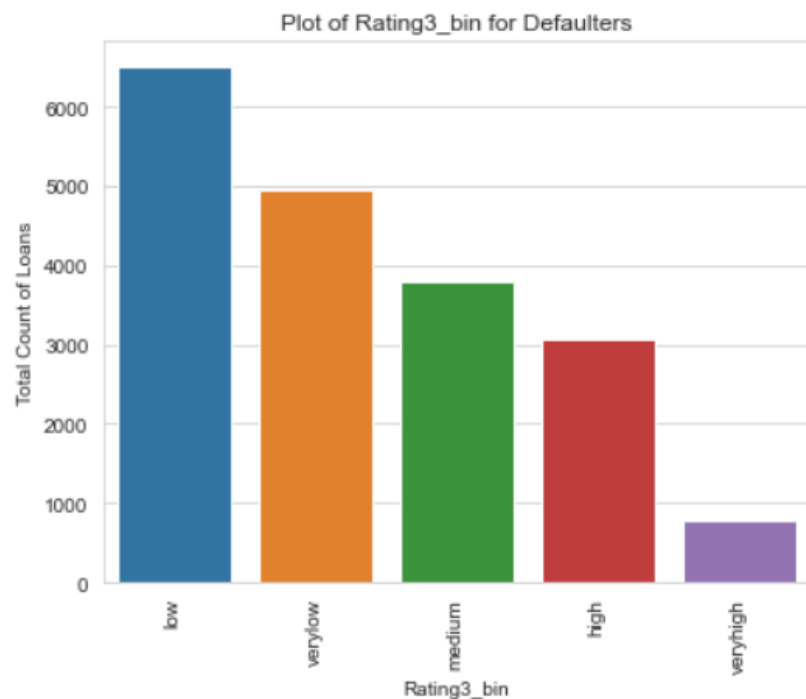
Distribution for count of normalized score from external source-II



The points can be inferred from the plot:

- As expected, The low rating group is the highest loan defaulters followed by very low rating group.
- The high rating group is the highest loan non-defaulters.

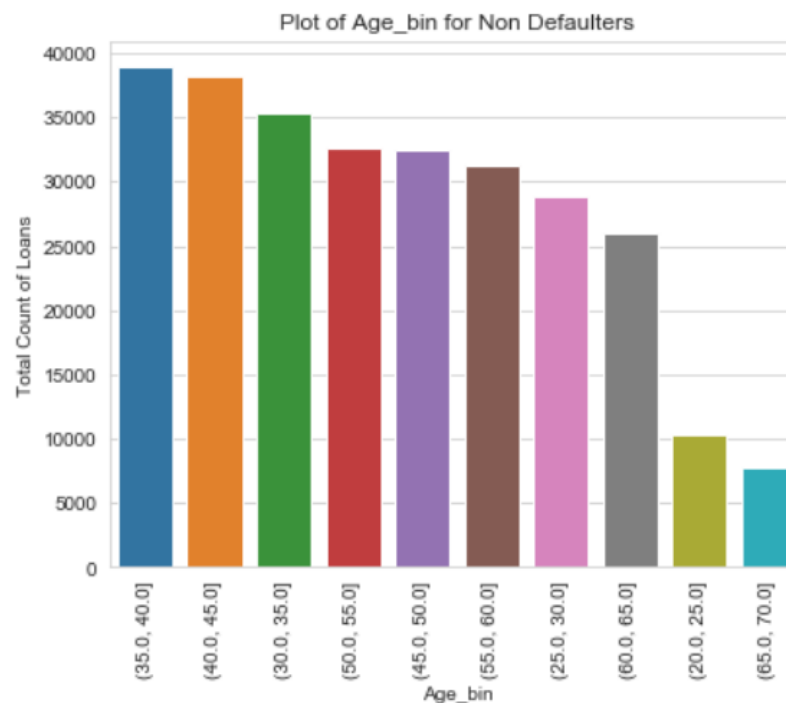
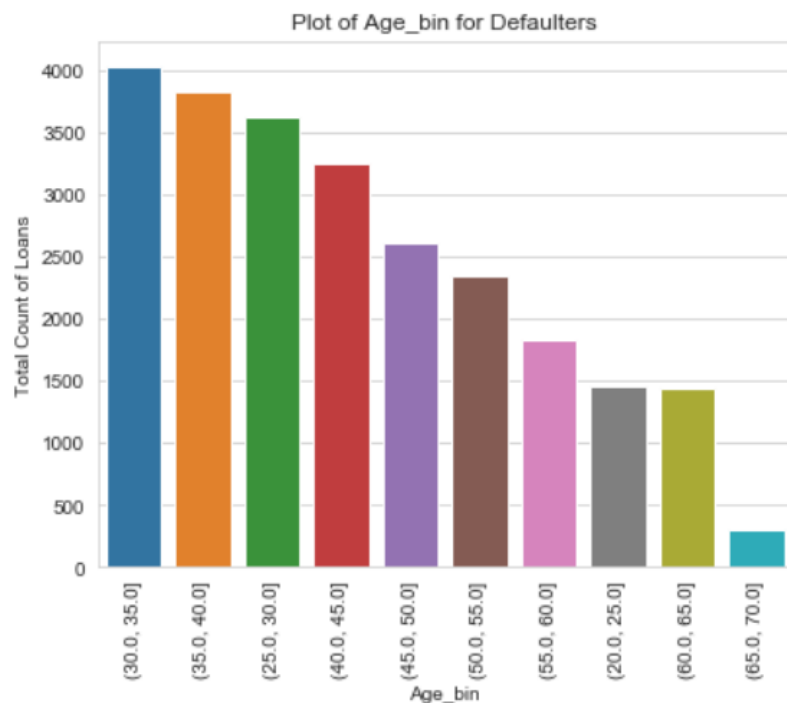
Distribution for normalized score from external source -III



The points can be inferred from the group:

- As expected, low and very rating groups came out to be highest non-defaulters
- The highest non-defaulters are high rating group

Distribution of age group

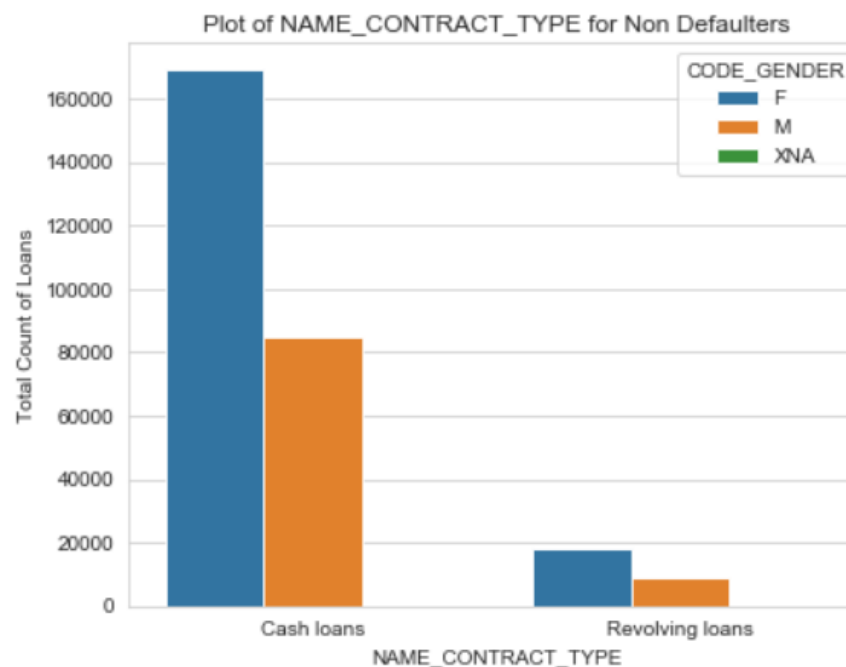
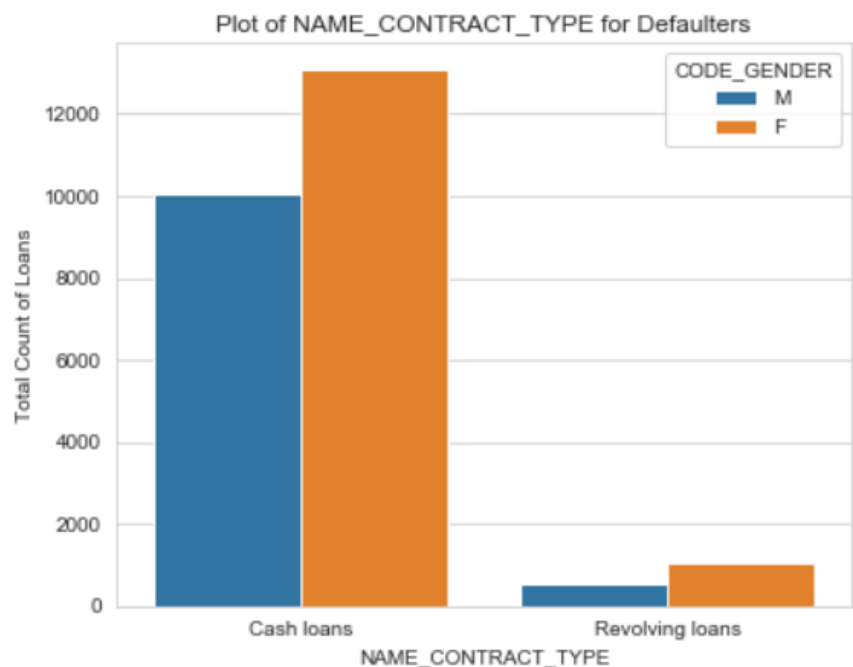


The points can be inferred from the group:

- The age between 30-35 are the highest defaulter group.
- The age between 35-40 are the highest non-defaulter group.

Bi- Variate Analysis

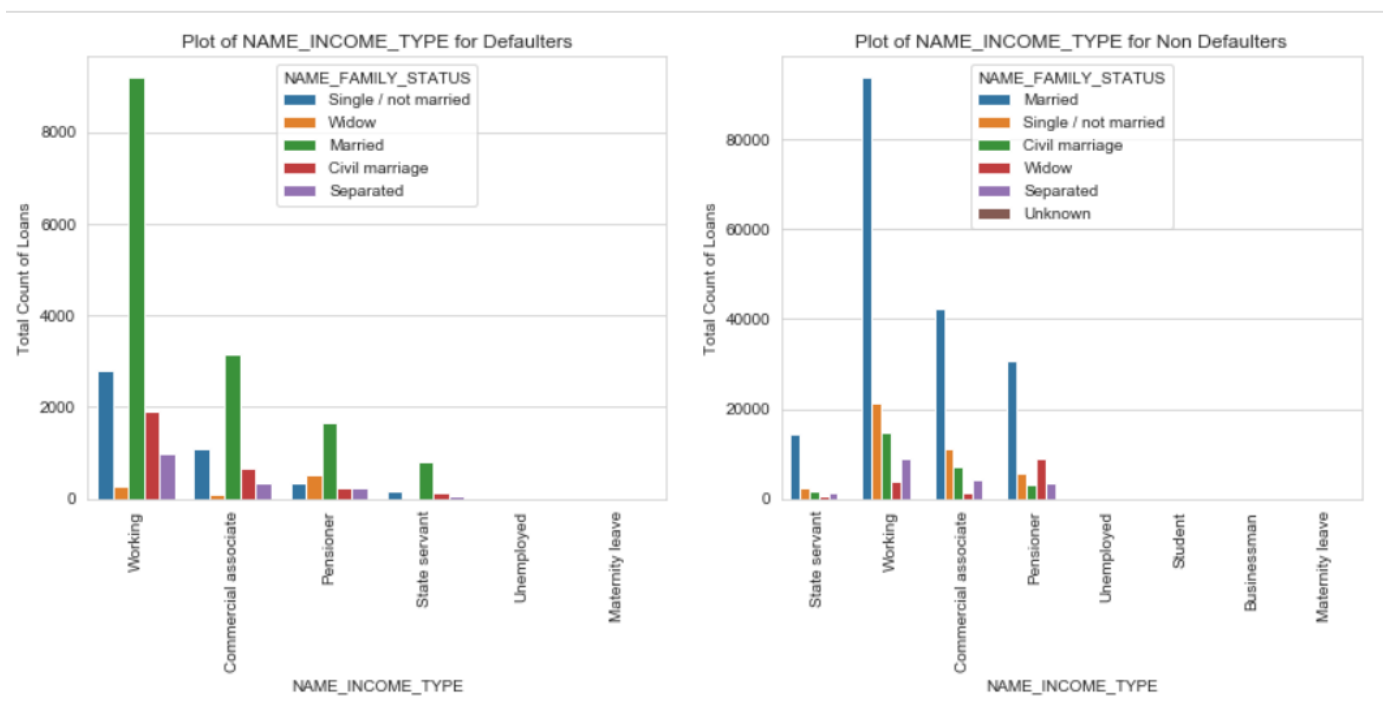
Type of loan Vs Gender of the client



The points can be inferred from the plot:

- The male gender has the highest loan non-defaulters count.
- Surprisingly, females are also the highest loan defaulters.
- As we inferred before cash loans are the popular irrespective to gender.

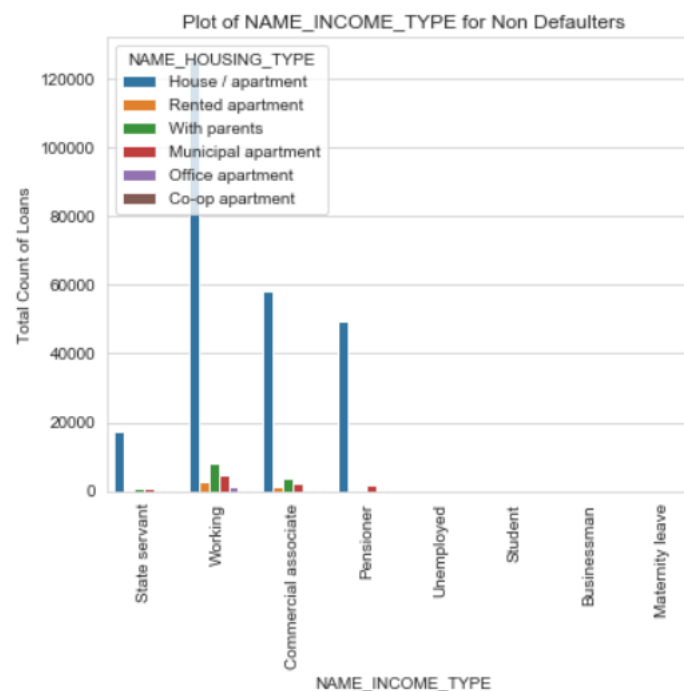
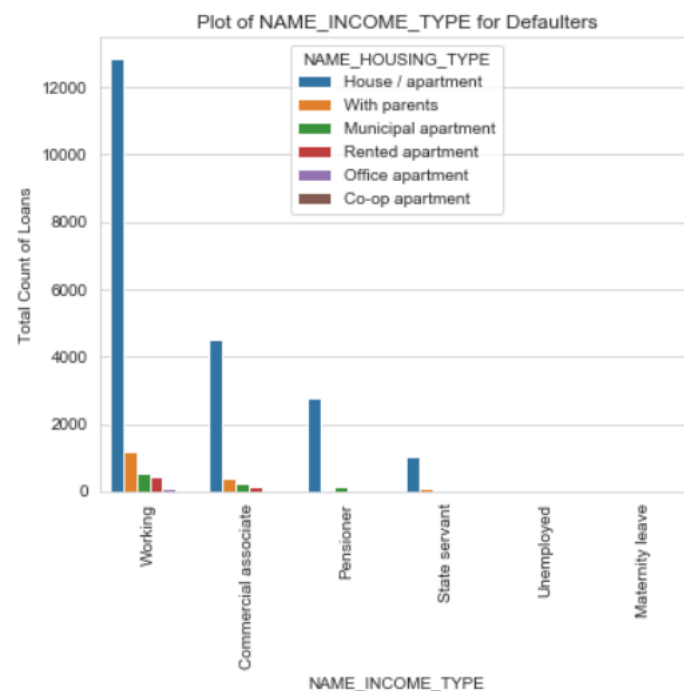
Income Type Vs Family Status



The points can be inferred from the plot :

- The working group who are married tends to be more defaulters than any group.

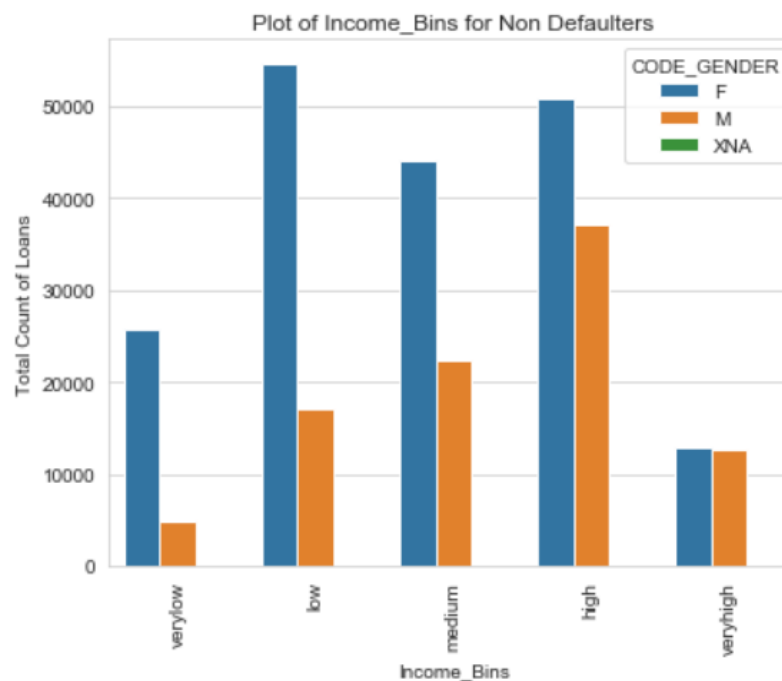
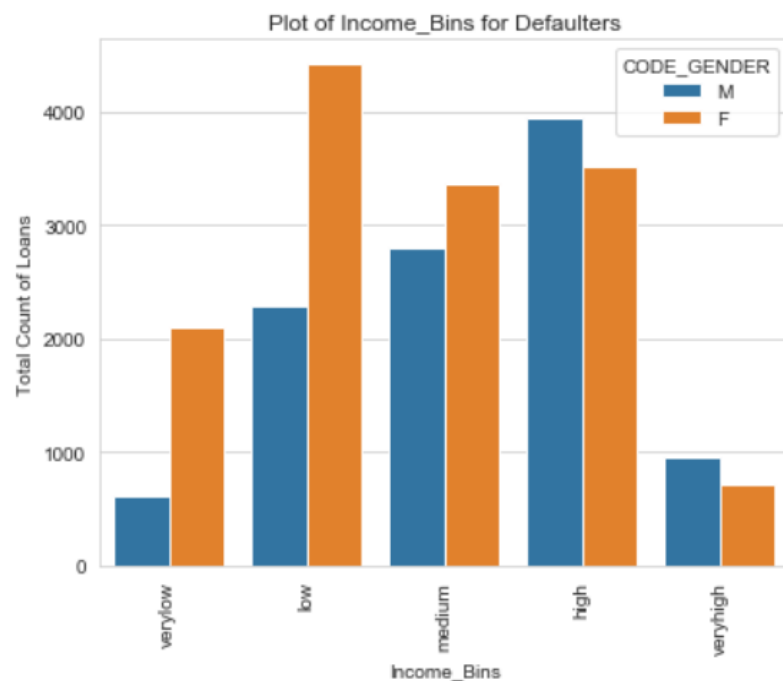
Income type Vs Housing type



The points can be inferred from the plot:

- The working group has highest counts, so the working group who has flat/apartment tends to have both non-defaulters and defaulters count.

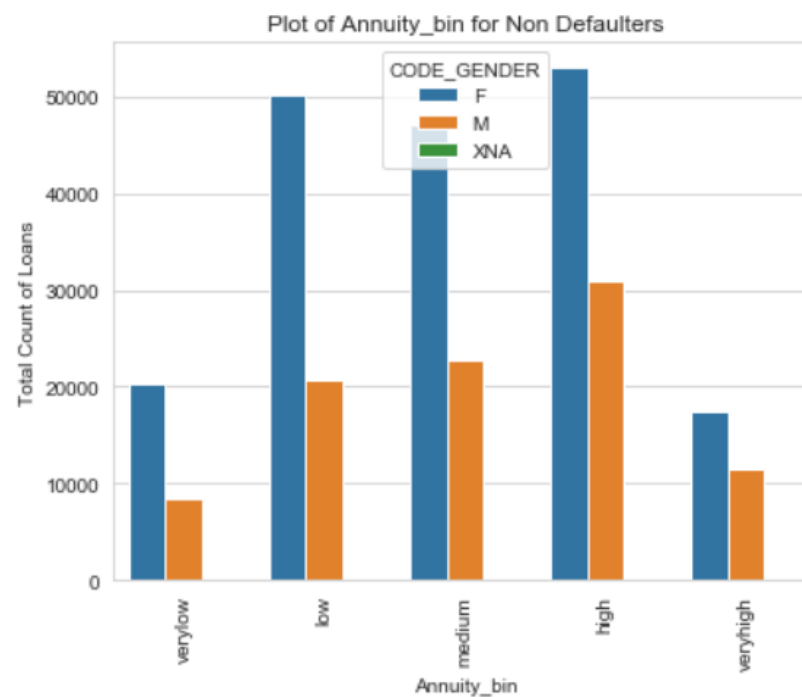
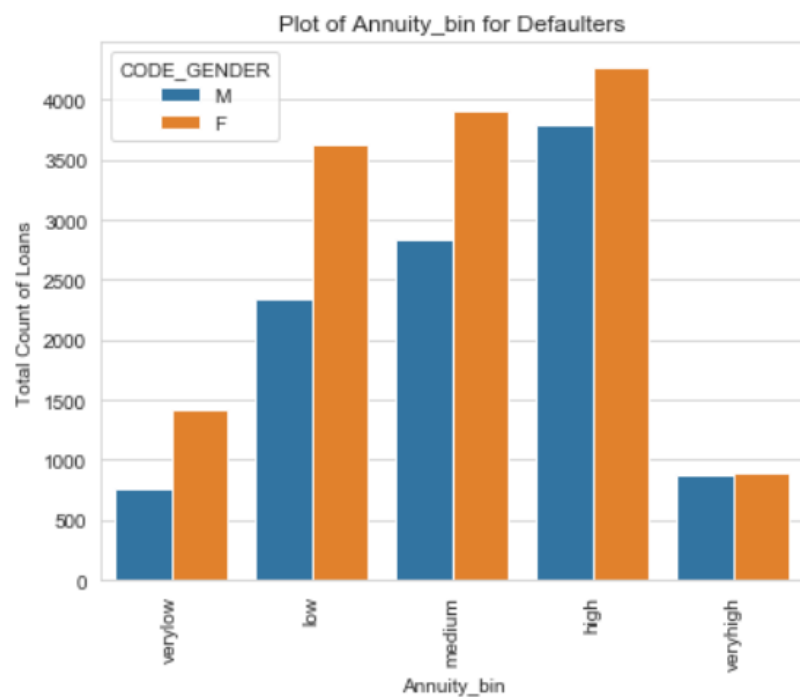
Income Group Vs Gender



The points can be inferred from the group:

- Males are the highest non-defaulters compared to female irrespective of income range.
- The females with below medium income groups has the highest defaulter count than male.
- But males with High-very high income group has more defaulter count.

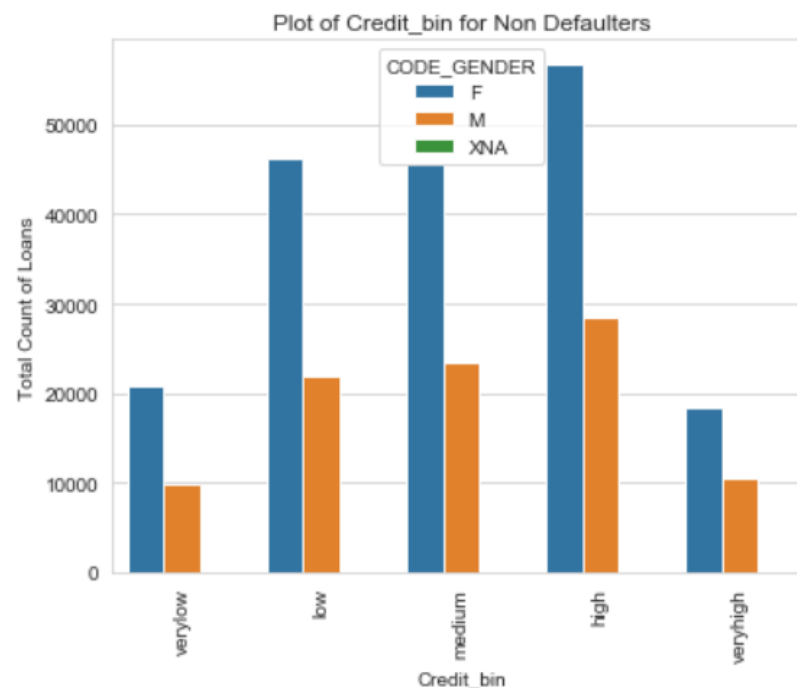
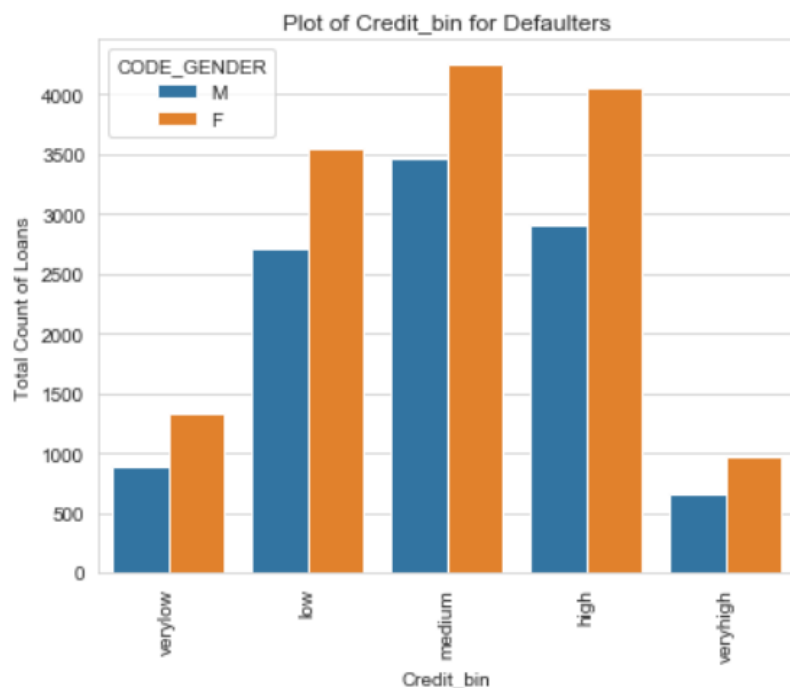
Annuity Group Vs Gender



Points can be inferred from the plot:

- Males are the highest non-defaulters irrespective to any Annuity group.
- Females are the highest defaulters irrespective to any annuity group

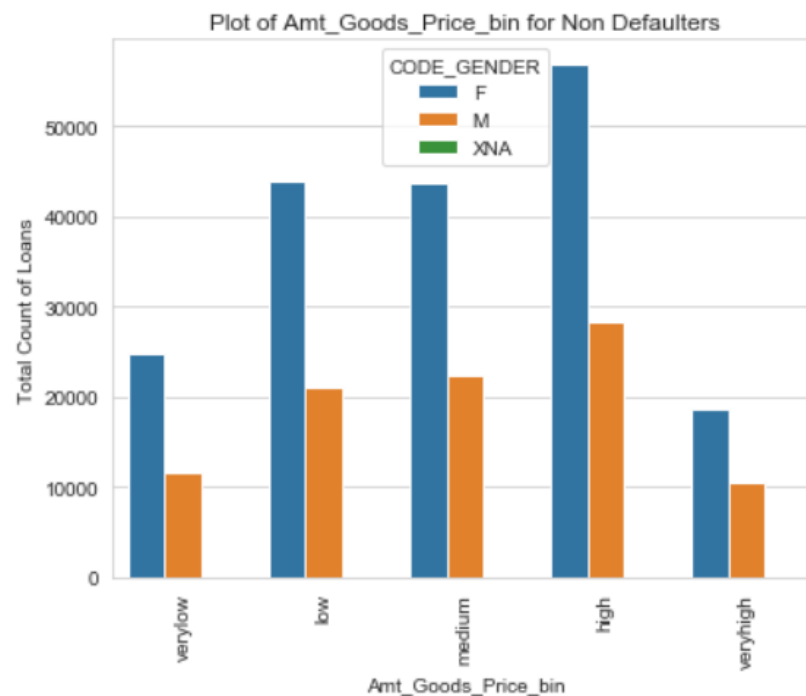
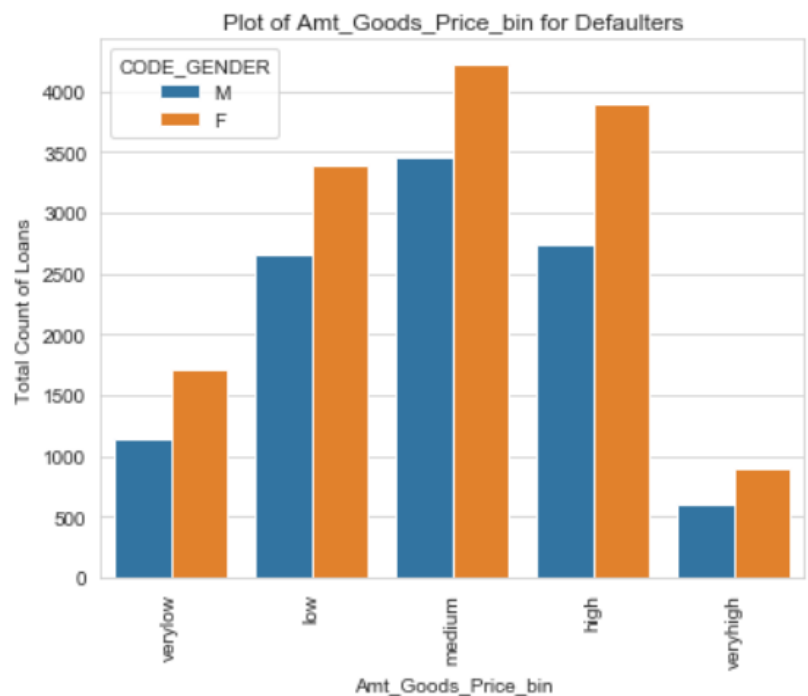
Credit Group Vs Gender



Points can be inferred from the group:

- Males has the most non-defaulters count than female irrespective to any credit range.
- Females with credit range between low to medium came to be more defaulters.
- Males with very-low credit group and very high credit group have more defaulters count.

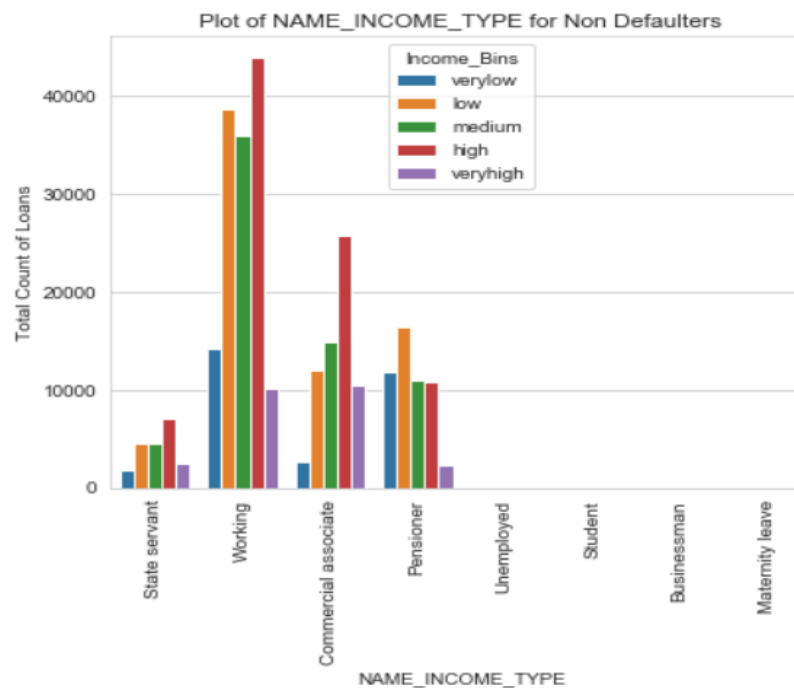
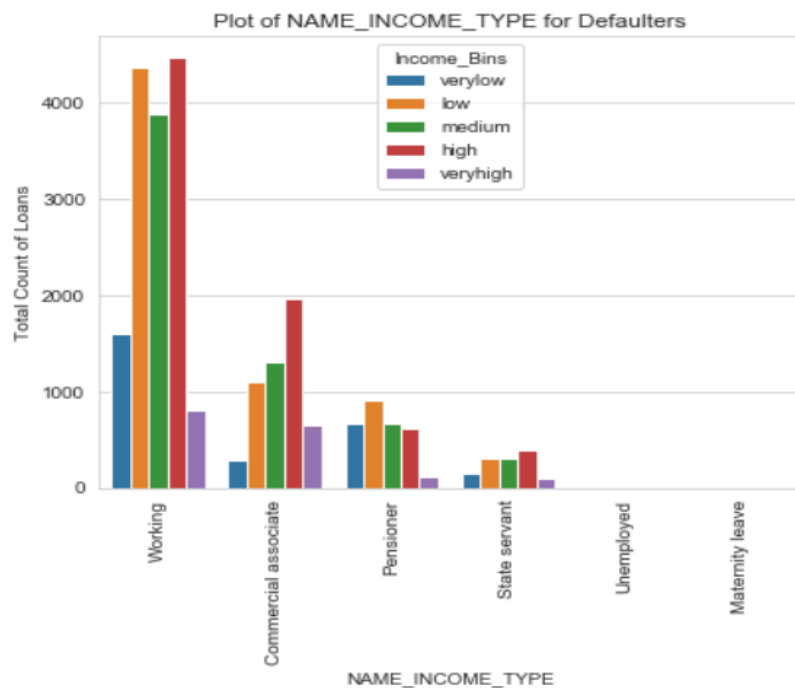
Goods-Price Groups Vs Gender



The points can be inferred from the plot:

- Males with any Goods-price range have the highest non-defaulter count
- Females with all goods-price range, except high group range having more defaulters count.

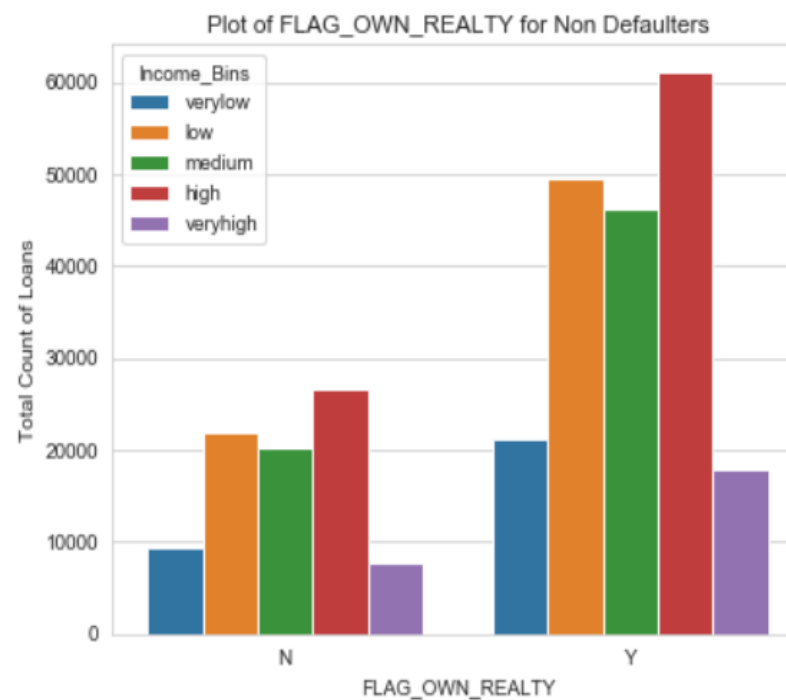
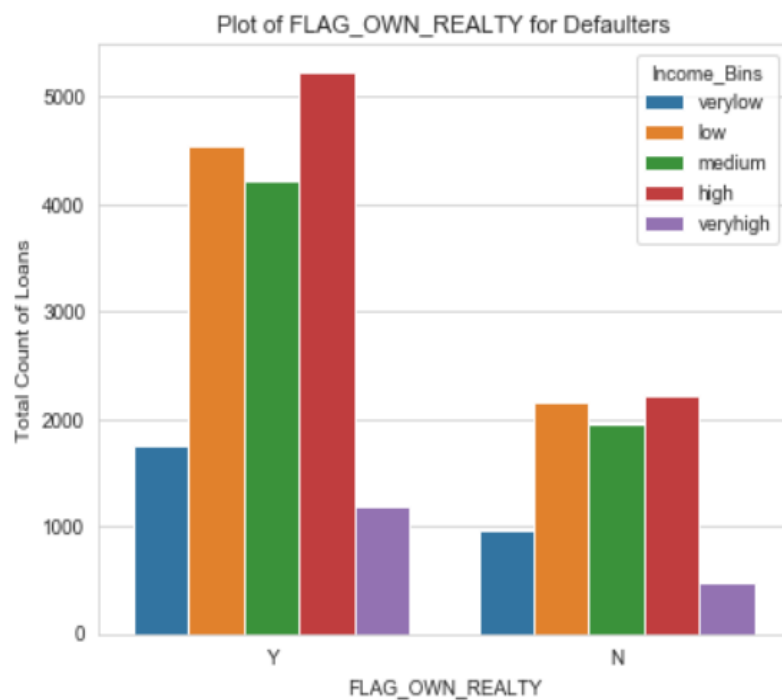
Income Type Vs Income Group



The points can be inferred from the plot:

- The working group has the highest defaulters and non-defaulters counts especially of high income group.
- As it can be concluded that the working group is the highest loan transaction group.

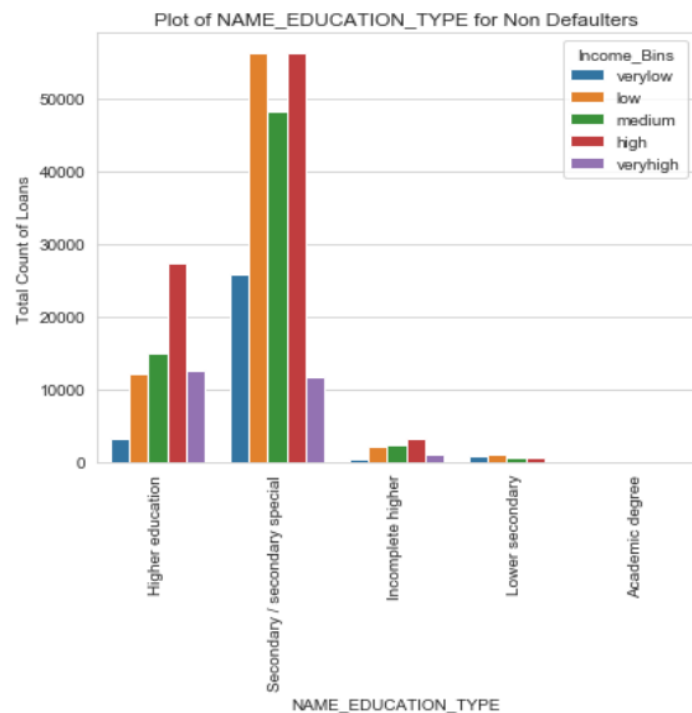
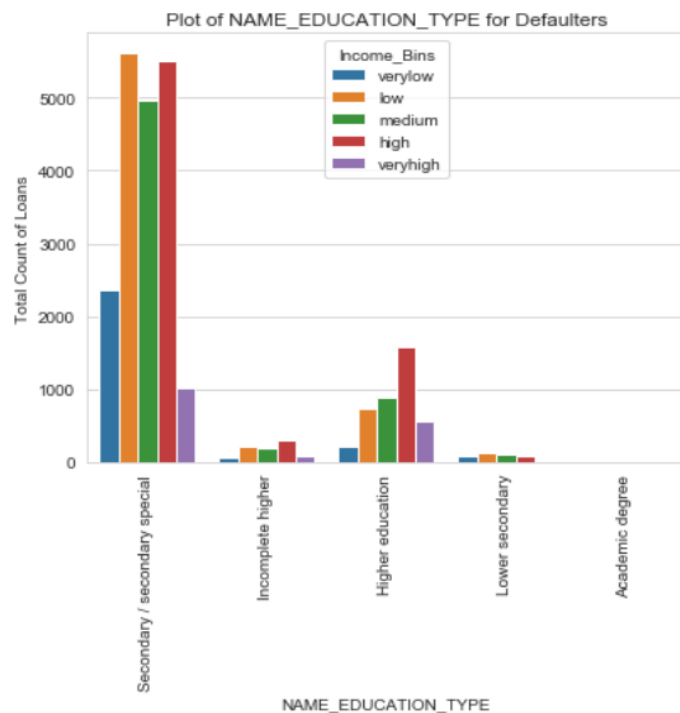
Client Owns flat/House Vs Income Group



The points can be inferred from the plot:

- The clients who own a flat are the most non-defaulters and defaulters especially with high income group.
- The clients who own a flat are the more loan receivers

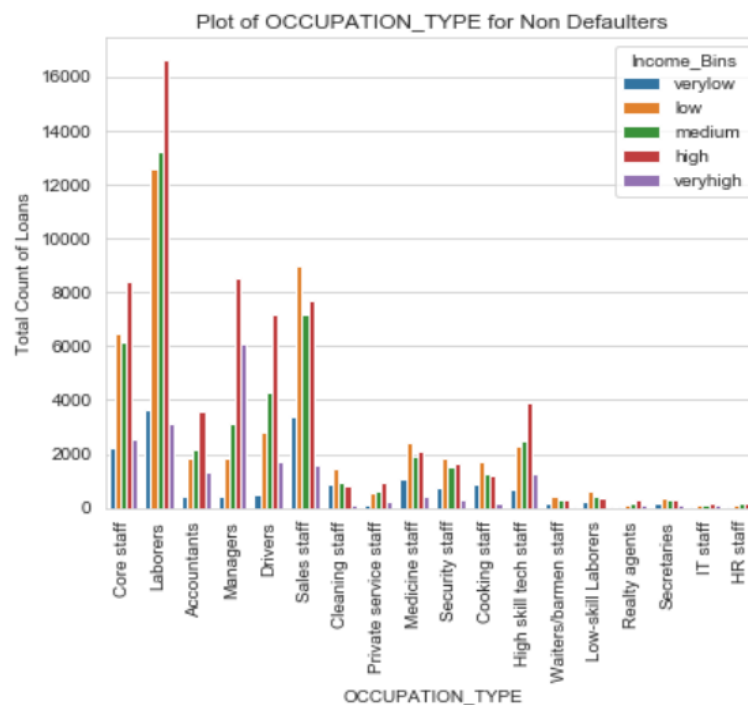
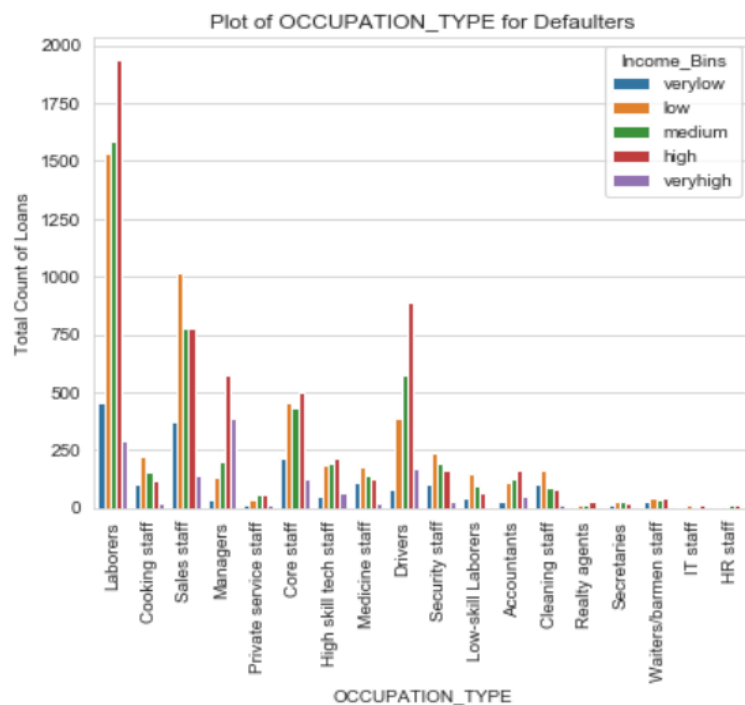
Educational Status Vs Income Group



The points can be inferred from the plot:

- The secondary education group with high and low income group has the highest non-defaulters and defaulters count.
- The higher education group with high income has the highest defaulters and non-defaulters count.

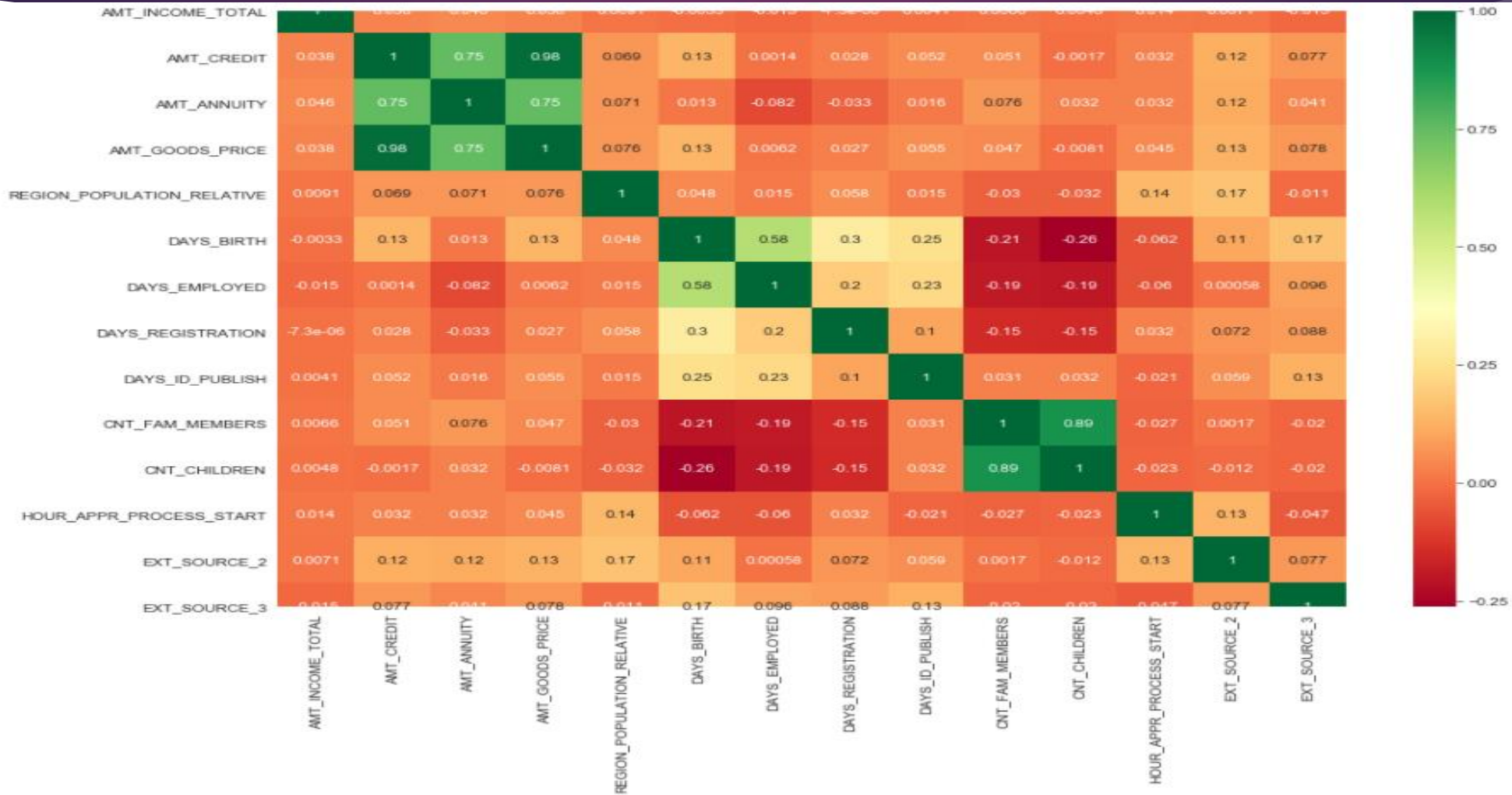
Occupation of client Vs Income Group



The points can be inferred from the plot:

- Laboure's with high income group has highest defaulters and non-defaulters count.
- The sales staff with low income group are also having high non-defaulters and defaulters count.

Correlation of Defaulters



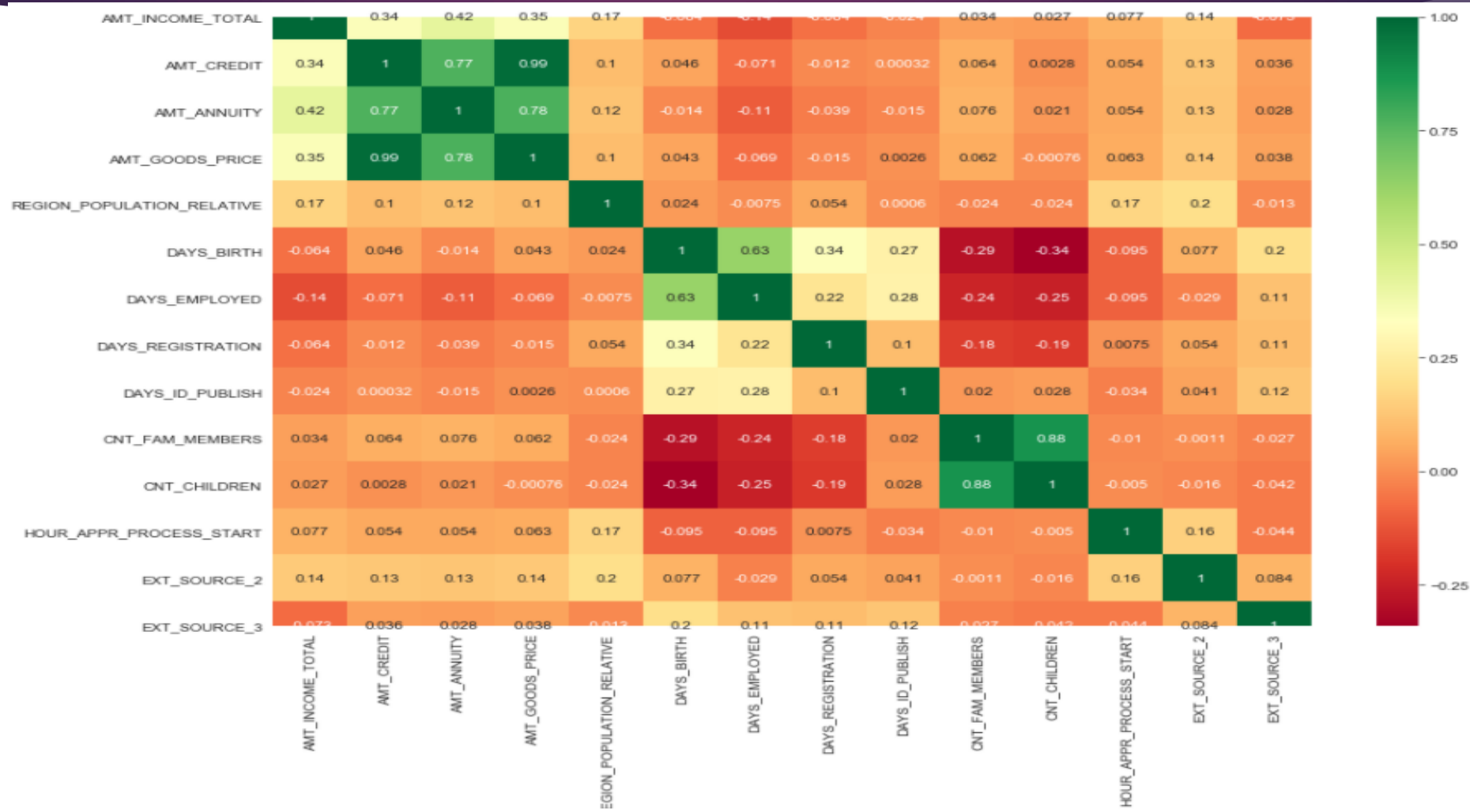
- ▶ The points to be inferred from the Defaulters heat map:
- ▶ The count of family-members is proportional to count of children with 0.89 correlation value. Family-members with more children could be more defaulters or Vice-versa.
- ▶ The amount-credited is proportional to goods-price with 0.98 correlation values, higher the goods price higher the amount credited are more defaulters or vice-versa.
- ▶ The annuity-amount is proportional with population relative with 0.71 correlation, The client lives in populated area with more annuity, came out to be defaulters.

Top 5 Correlation of Defaulters

EXT_SOURCE_3	EXT_SOURCE_3	1.000000
AMT_GOODS_PRICE	AMT_CREDIT	0.983101
CNT_FAM_MEMBERS	CNT_CHILDREN	0.885595
AMT_ANNUITY	AMT_GOODS_PRICE	0.752526
AMT_CREDIT	AMT_ANNUITY	0.752045
DAYS_BIRTH	DAYS_EMPLOYED	0.582620
	DAYS_REGISTRATION	0.298398
	DAYS_ID_PUBLISH	0.250291
DAYS_ID_PUBLISH	DAYS_EMPLOYED	0.228466
DAYS_REGISTRATION	DAYS_EMPLOYED	0.195712

dtype: float64

Correlation of Non-Defaulters



- ▶ The points can be inferred from the Non-Defaulters heat map:
- ▶ The Insights are mostly similar to the defaulters heat map.
- ▶ The credit-amount is proportional to days-birth with 0.46 correlation, which means credit amount is higher for low age clients.
- ▶ The credit amount received high for those who has less family members and less children.

Top 5 correlation of non- defaulters

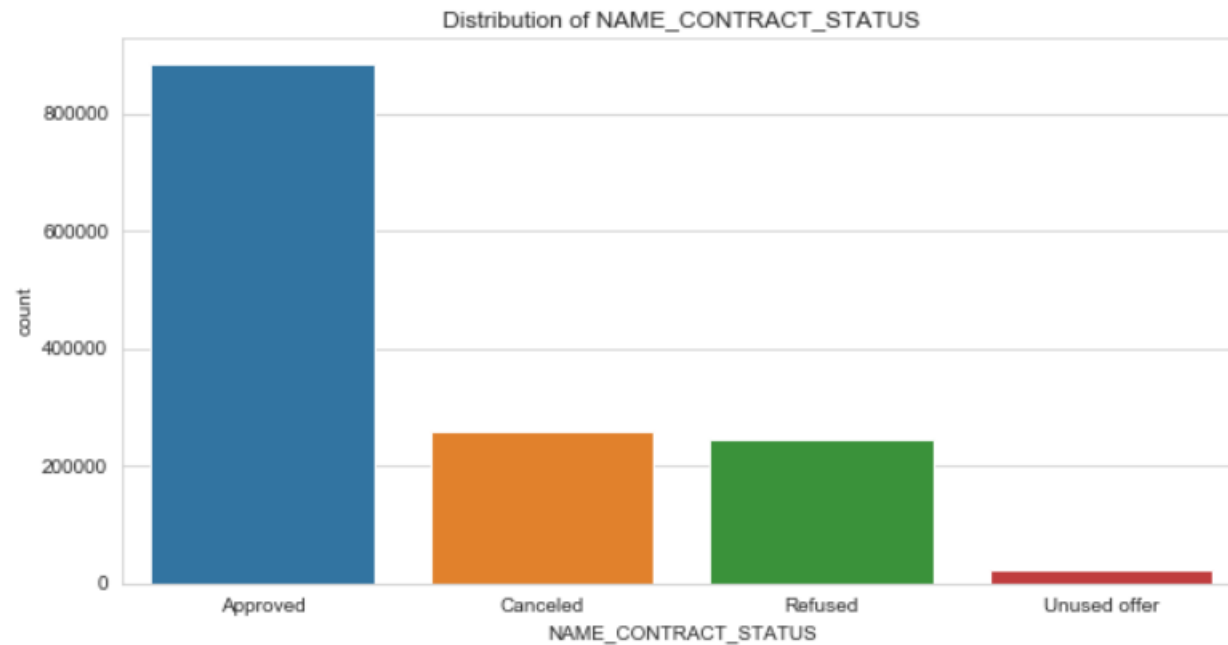
DAYS_BIRTH	CNT_CHILDREN	-0.261254
CNT_FAM_MEMBERS	DAYS_BIRTH	-0.206302
CNT_CHILDREN	DAYS_EMPLOYED	-0.193633
DAYS_EMPLOYED	CNT_FAM_MEMBERS	-0.187673
DAYS_REGISTRATION	CNT_CHILDREN	-0.149895
CNT_FAM_MEMBERS	DAYS_REGISTRATION	-0.145719
DAYS_EMPLOYED	AMT_ANNUITY	-0.081804
DAYS_BIRTH	HR_APPR_PROCESS_START	-0.061542
DAYS_EMPLOYED	HR_APPR_PROCESS_START	-0.059771
EXT_SOURCE_3	HR_APPR_PROCESS_START	-0.046657

dtype: float64

PREVIOUS APPLICATION

Uni-Variate Analysis

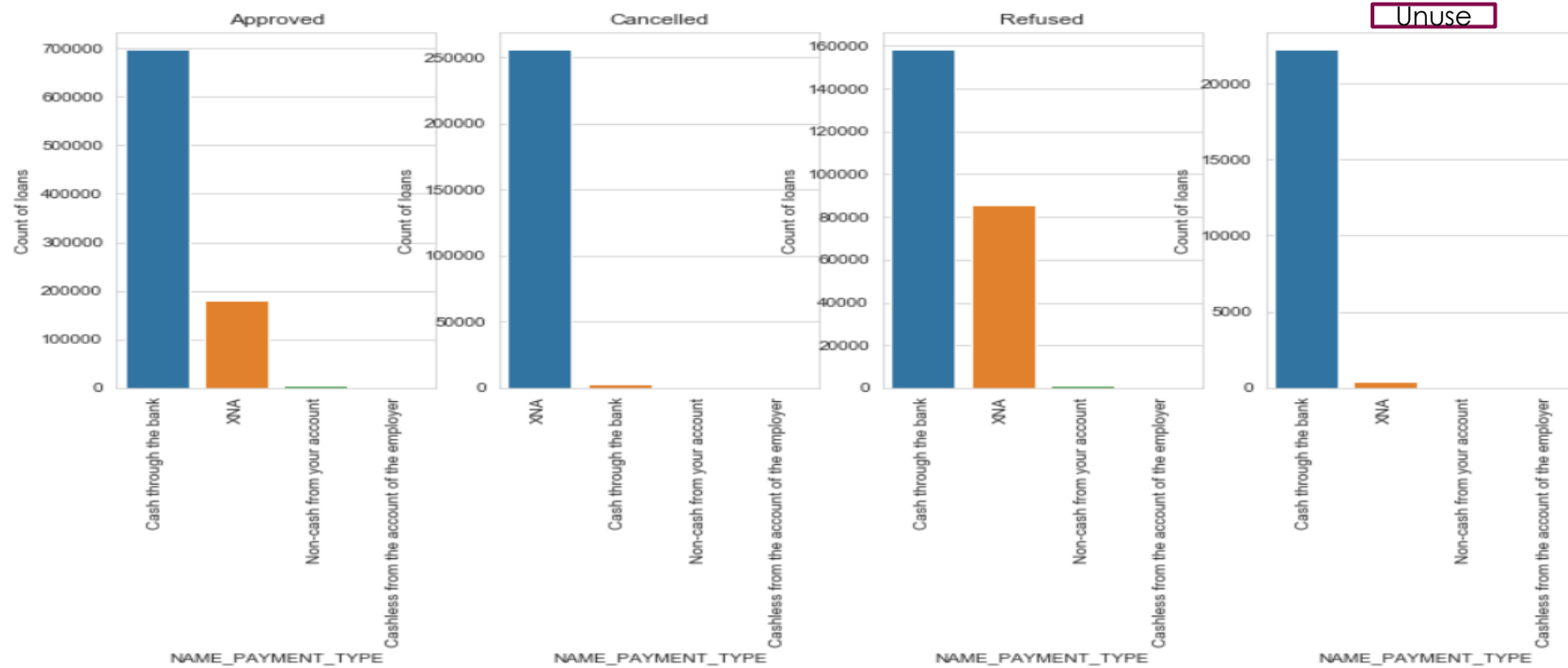
Distribution of Contract Status



The points can be inferred from the plot:

- The contract status for most of the customers is approved in the previous application data.
- The customers who unused the offer are less.

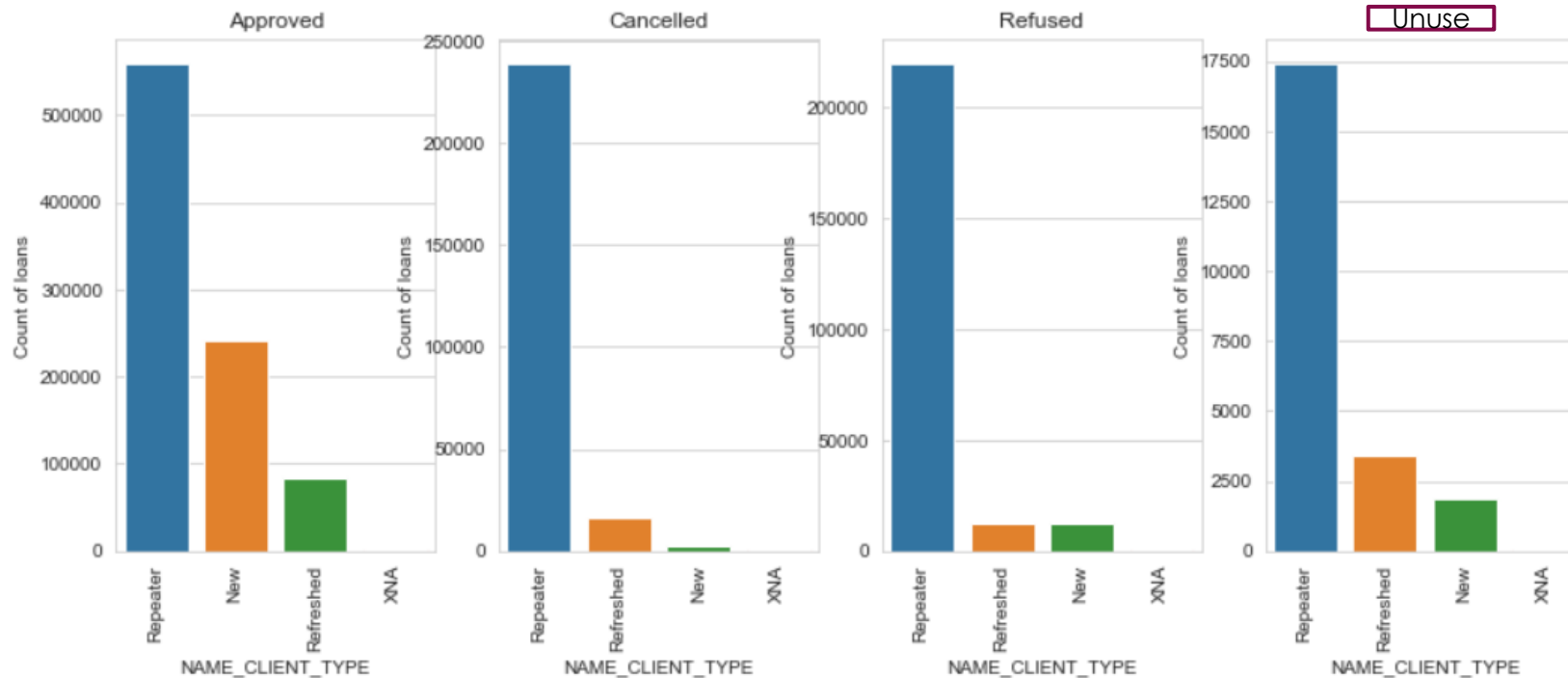
Distribution of Payment type



The points can be inferred from the plot:

- The clients who obtained the amount from the bank in cash mode is more irrespective to the status of their contract

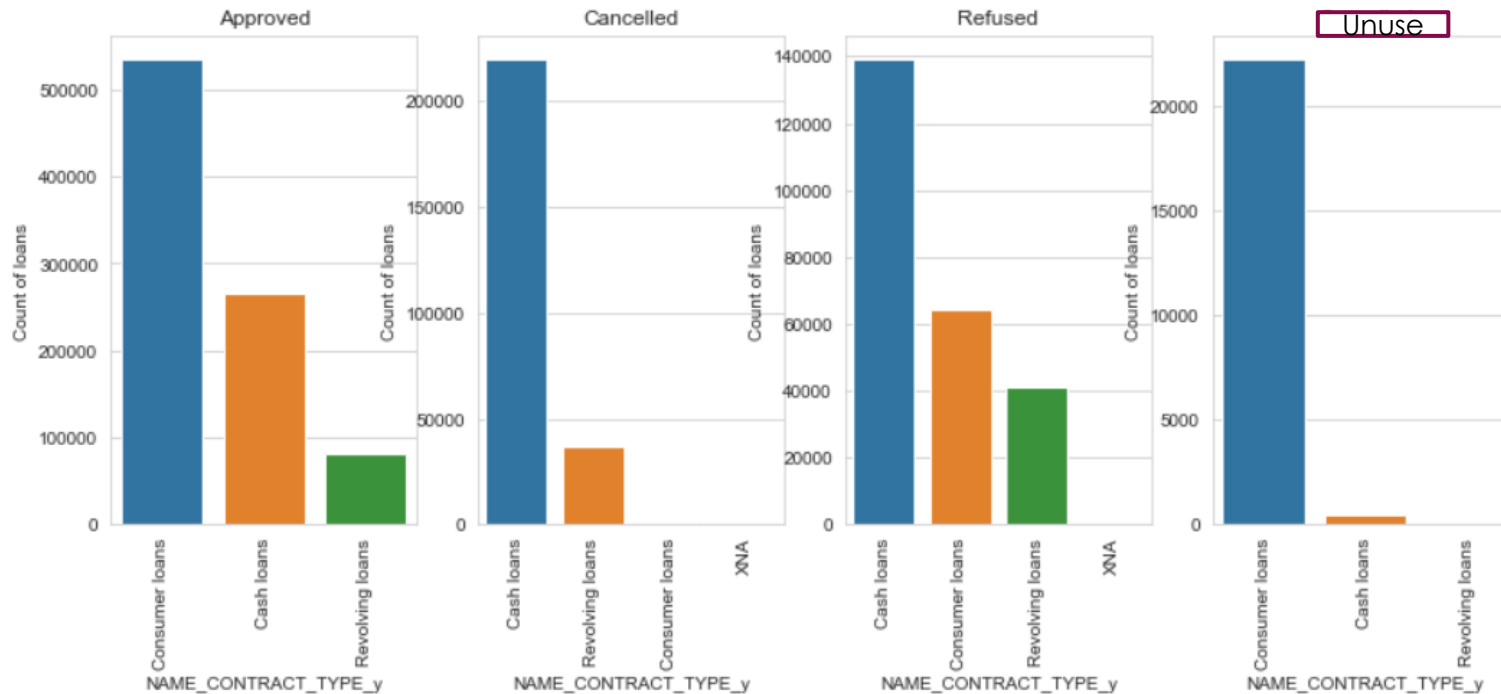
Distribution of Client type



The points can be inferred from the plot:

- The clients who repeat the loan irrespective to the contract status are more.
- The new customers who got loan approval are also more than rejected or refused.
- This means the clients who took loans previously are more willing to either to continue or repeat the loans.

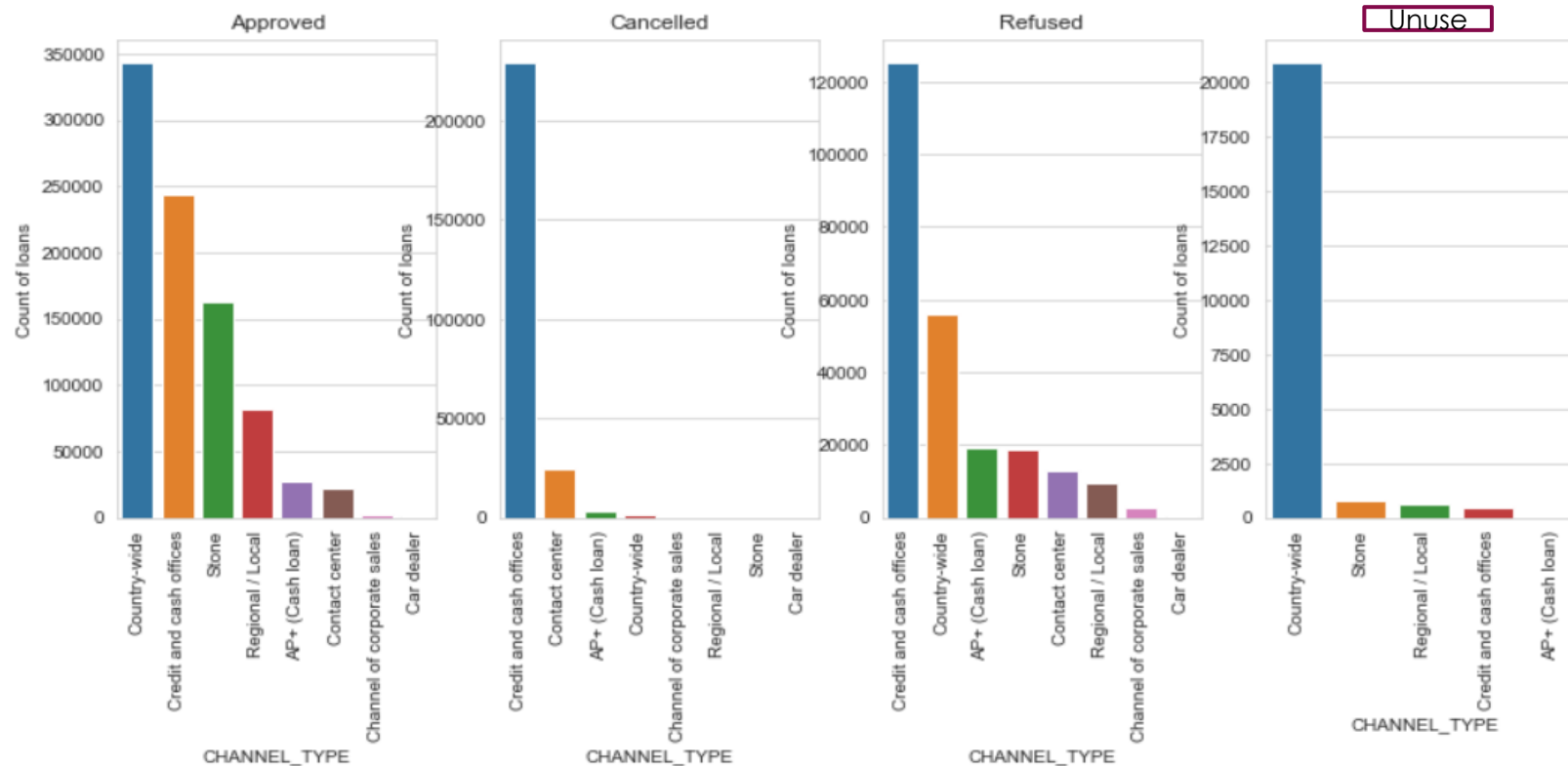
Distribution of Contract type



The points can be inferred from the plot:

- The clients who got approval are more from consumer loans type.
- The cash loans are more which got refused or rejected.
- Surprisingly, the cancelled contract types are also consumer loans.

Distribution of channel type

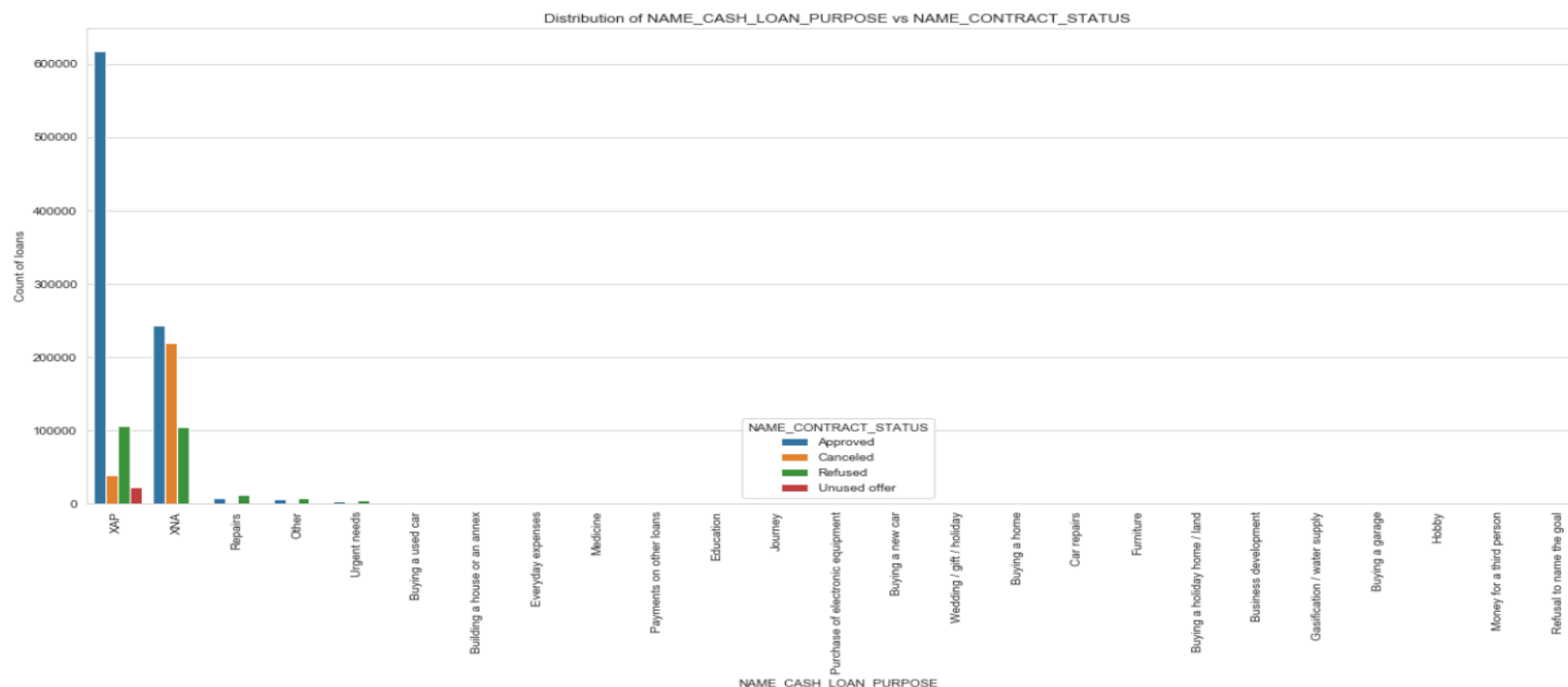


The points can be inferred from the plot:

- The clients through country-wide channel has most approved and cancelled contracts.
- This means the clients preferred to take loans through country-wide channel followed by credit and cash offices.
- The credit and cash offices possibly having ineffective process, for being cancelled or refused contracts

Bi-Variate Analysis

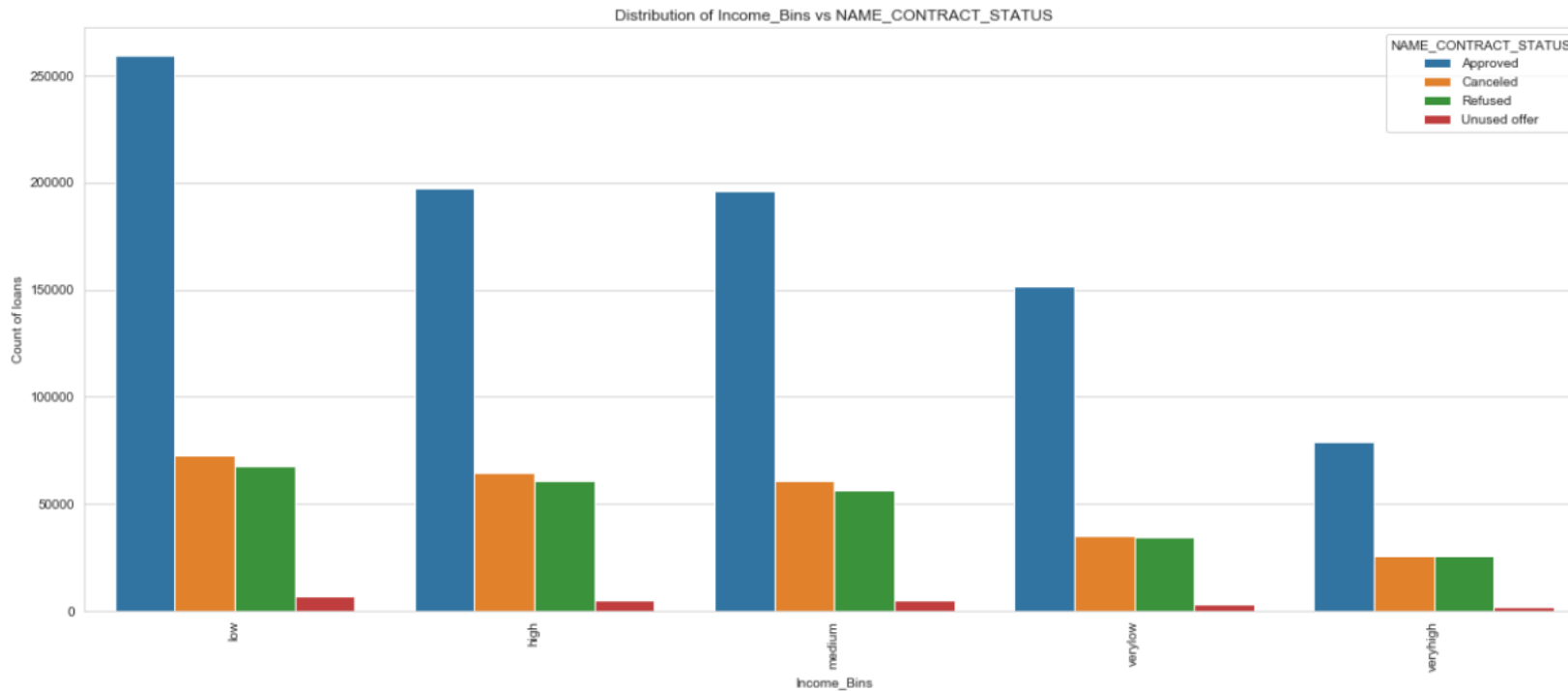
Distribution of Loan Purpose Vs contract status



The points can be inferred from the plot:

- The purpose of loans data is not collected and stored, so there is no proper insights to infer

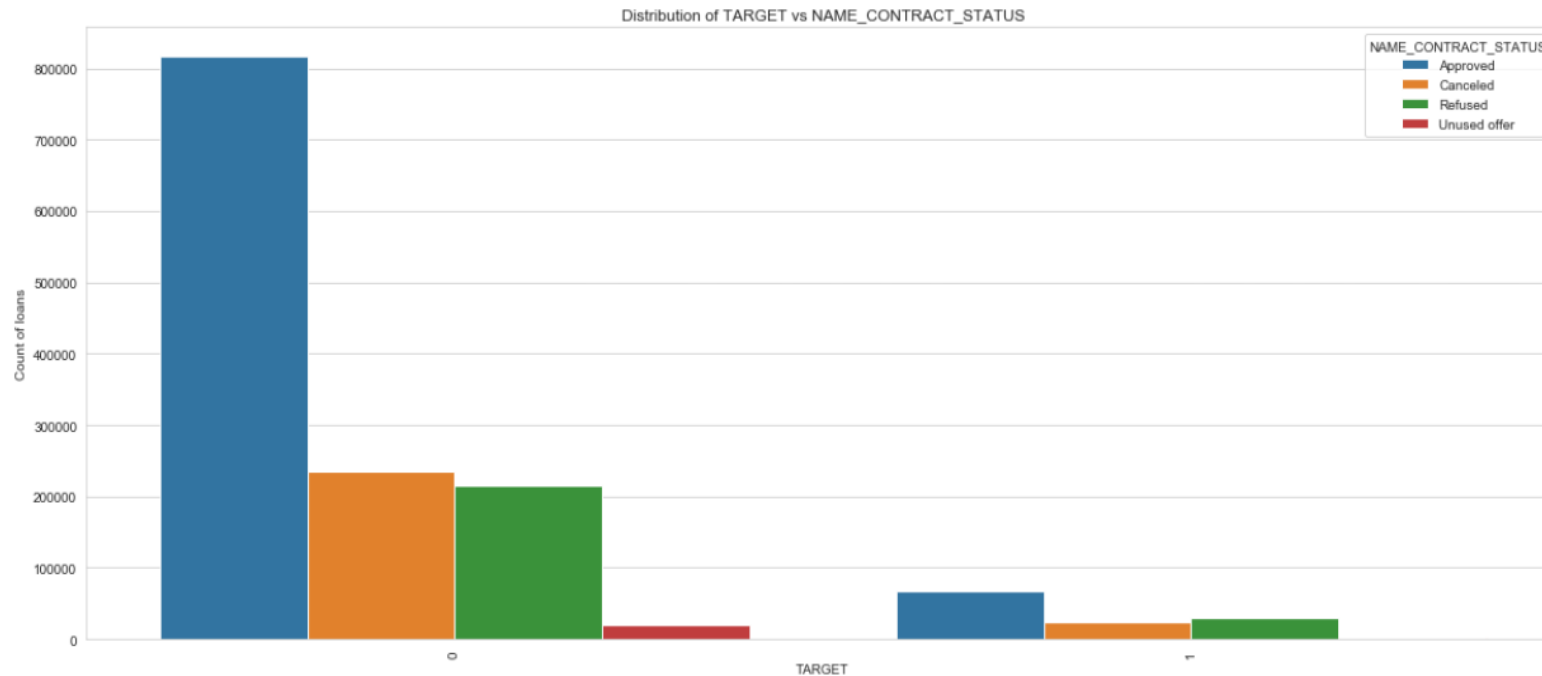
Income vs Contract status



The points can be inferred from the plot:

- The approval pf loan percentage is more irrespective to the income range.
- The low income range clients have higher approval count compare to any income group

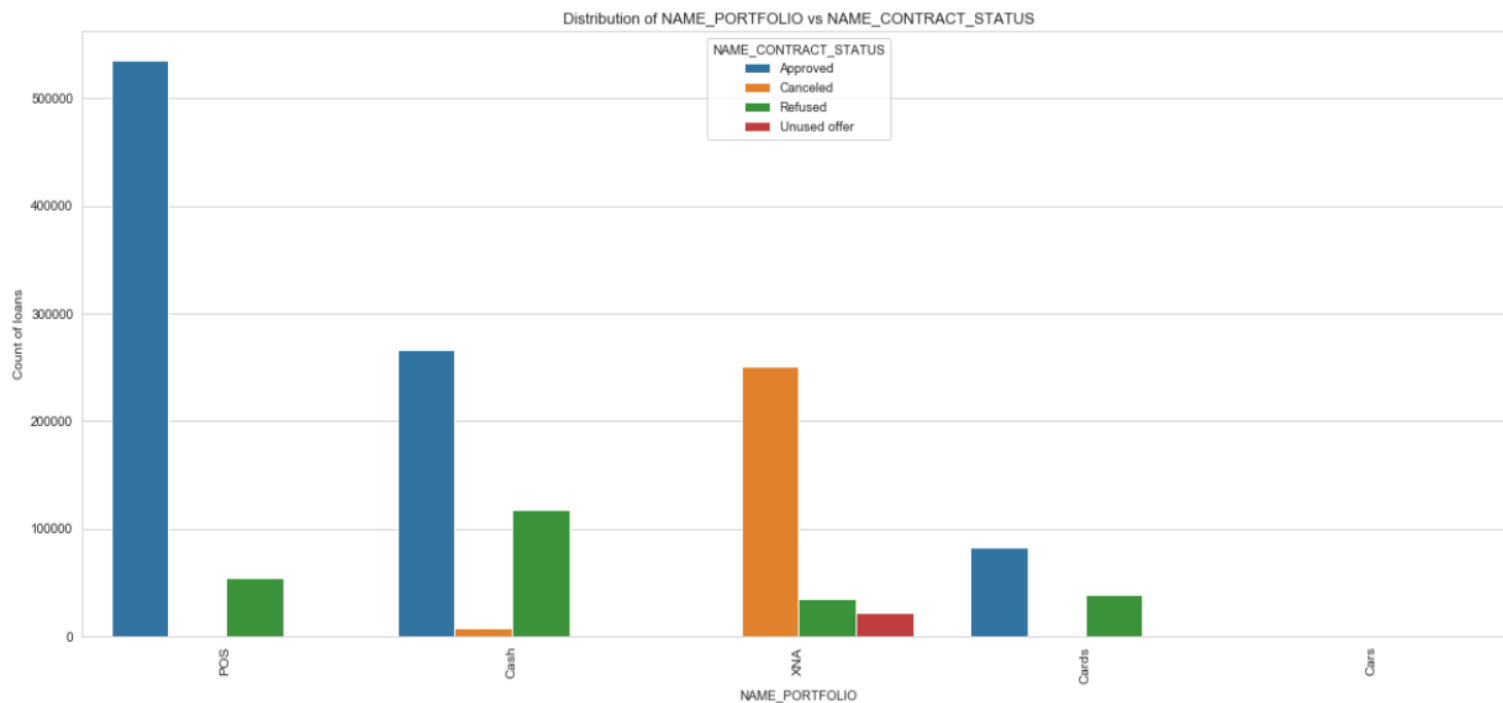
Distribution of Target variable Vs Contract status



The points can be inferred from the plot:

- As we already know the non-defaulters count is more than defaulters.
- The loan approval count for clients irrespective target is more.
- this means the percentage of loans approval is far more than any factors.

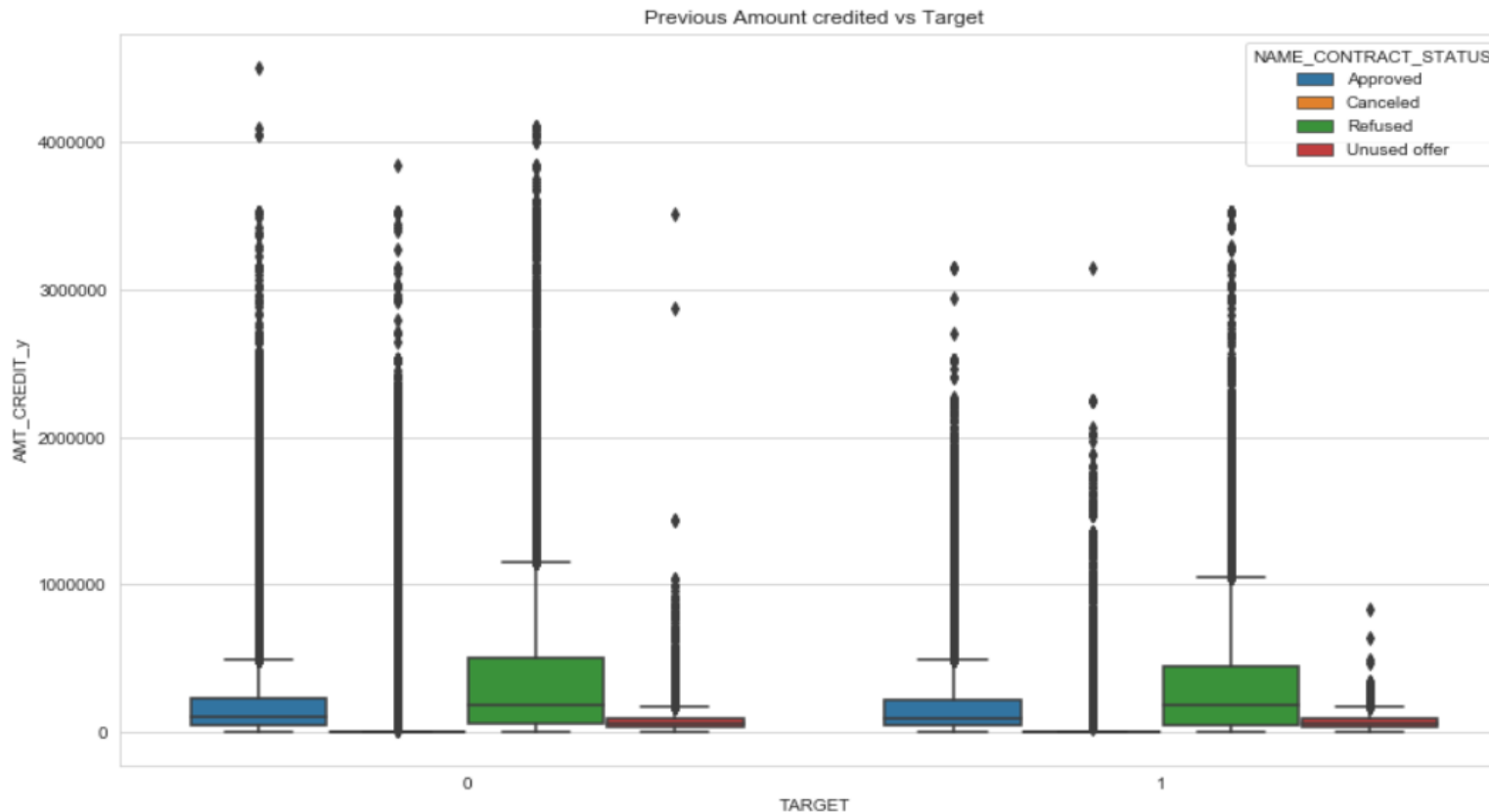
Portfolio Vs Contract Status



The points can be inferred from the plot:

- The loan process was done more for Point of Sale category followed by cash.
- Surprisingly, The contract status is approved for most of the loan appeals for POS's
- The car loans data is not available

Amount Credited Vs Contract Status



The points can be inferred from the plot:

- The Refused status of applications is more for those who applied for higher amount irrespective to the target variable.
- Which means higher the amount of loan the percentage of refusal increases, could be the safety concern for not being credible to the channels.

Conclusion

- ▶ It can be concluded that
- ▶ Banks must focus more on Pensioners and businessman with housing type as apartments tends to more successful.
- ▶ Working group with occupation type as Laboure's seems to be more defaulters, It could hurdle success rate of banks.
- ▶ The higher education with high income group is having high non-defaulters counts, so bank should focus the same for higher returns.
- ▶ Simultaneously, Despite being the Secondary education group has higher count of loan transactions, the percentage of defaulters is also more so banks should be cautious before approval of loans to this group.
- ▶ Percentage of females came out to be more defaulters than the percentage of males irrespective of Income, amount-credited, annuity, and the goods price.
- ▶ People who borrow loans under the channel credit/cash offices tends to more successful than through any channel.