

# CREDIT EDA ASSIGNMENT

SUBMITTED BY –  
GUNJIT KAPOOR

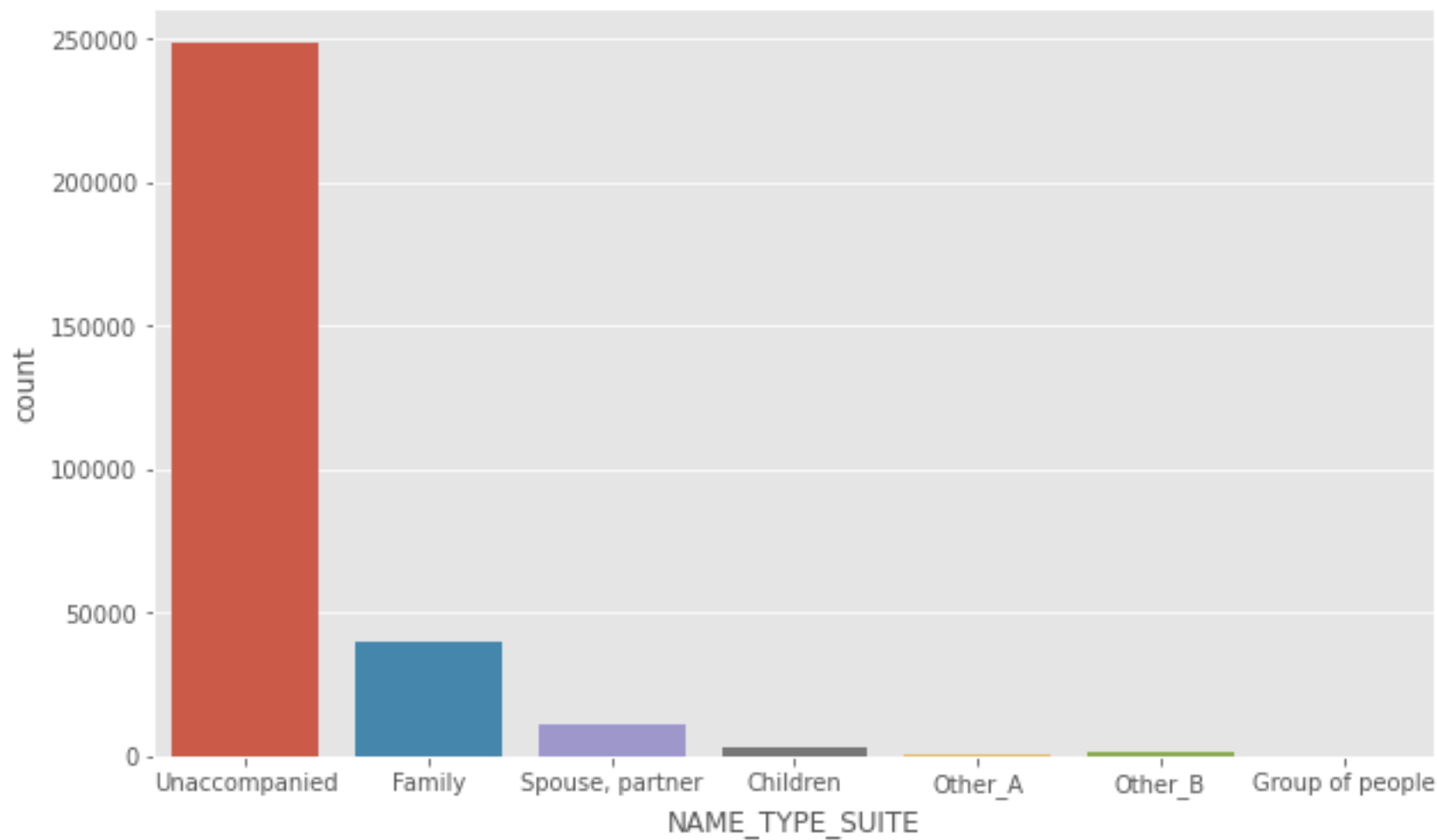
# ***Analysing the Application\_data set***

---

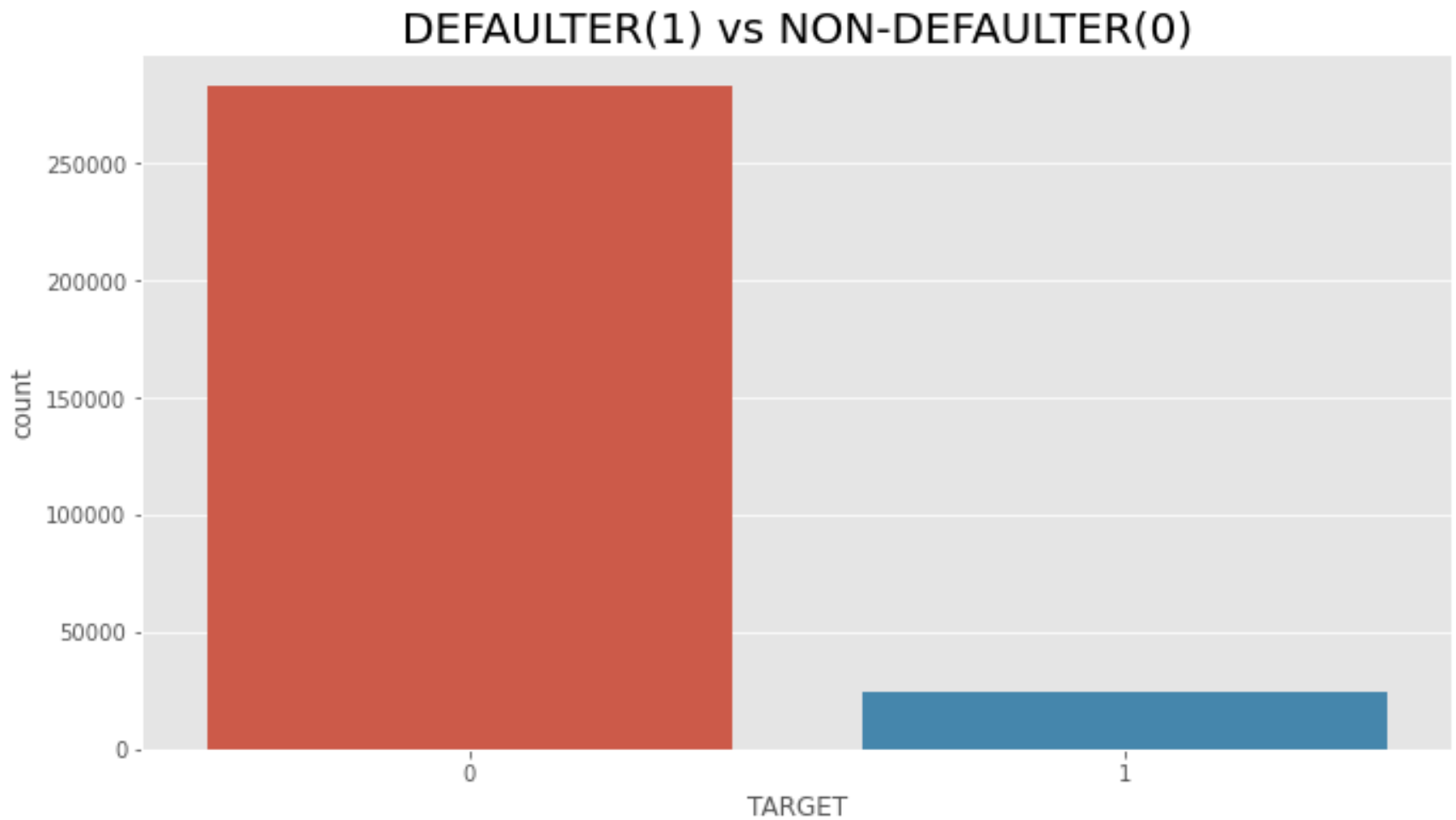
- ▶ The application data set contains the data about the new applicants , previous applicants who are again applying for loan.
- ▶ This data set contains around 122 columns .
- ▶ Let us see the insights that have been gained through the analysis of this data set.
- ▶ Through this data set we will analyse the credit worthiness which depends on various factors.
- ▶ The required plots with the insights gained from them will help to do correct decisions.



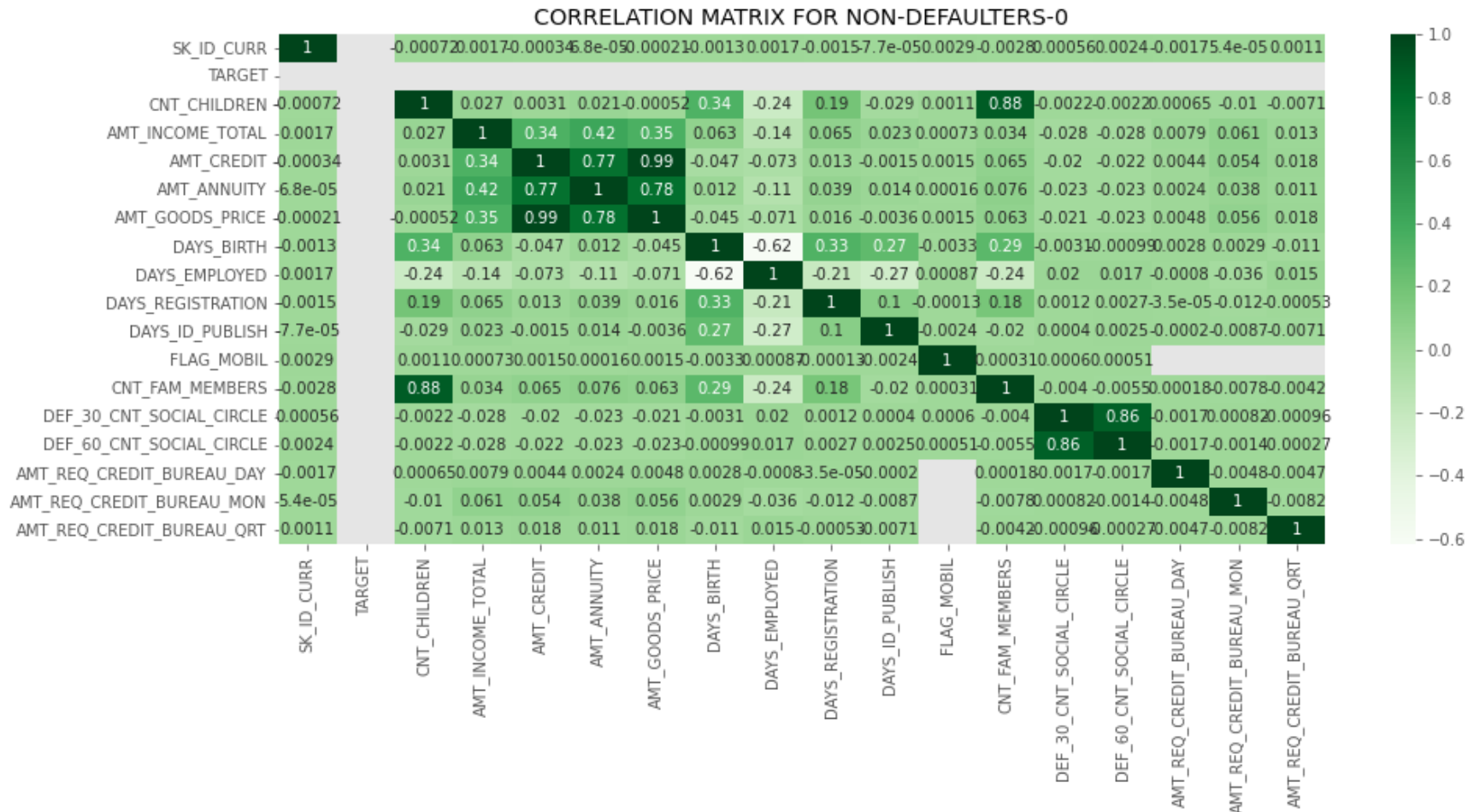
# As it is clear that NAME\_TYPE\_SUITE is a categorical variable, so mode can be used to impute the column and a count plot can be used for presentation. **Here mode is clearly 'unaccompanied' category.**



The above analysis shows that 92 % clients are non-defaulters and 8 % are defaulters. This is clearly a case of highly imbalanced dataset.



# # CORRELATION MATRIX FOR NON-DEFAULTERS-0



## TOP 10 CORRELATIONS FROM THE NON-DEFAULTER-0 MATRIX

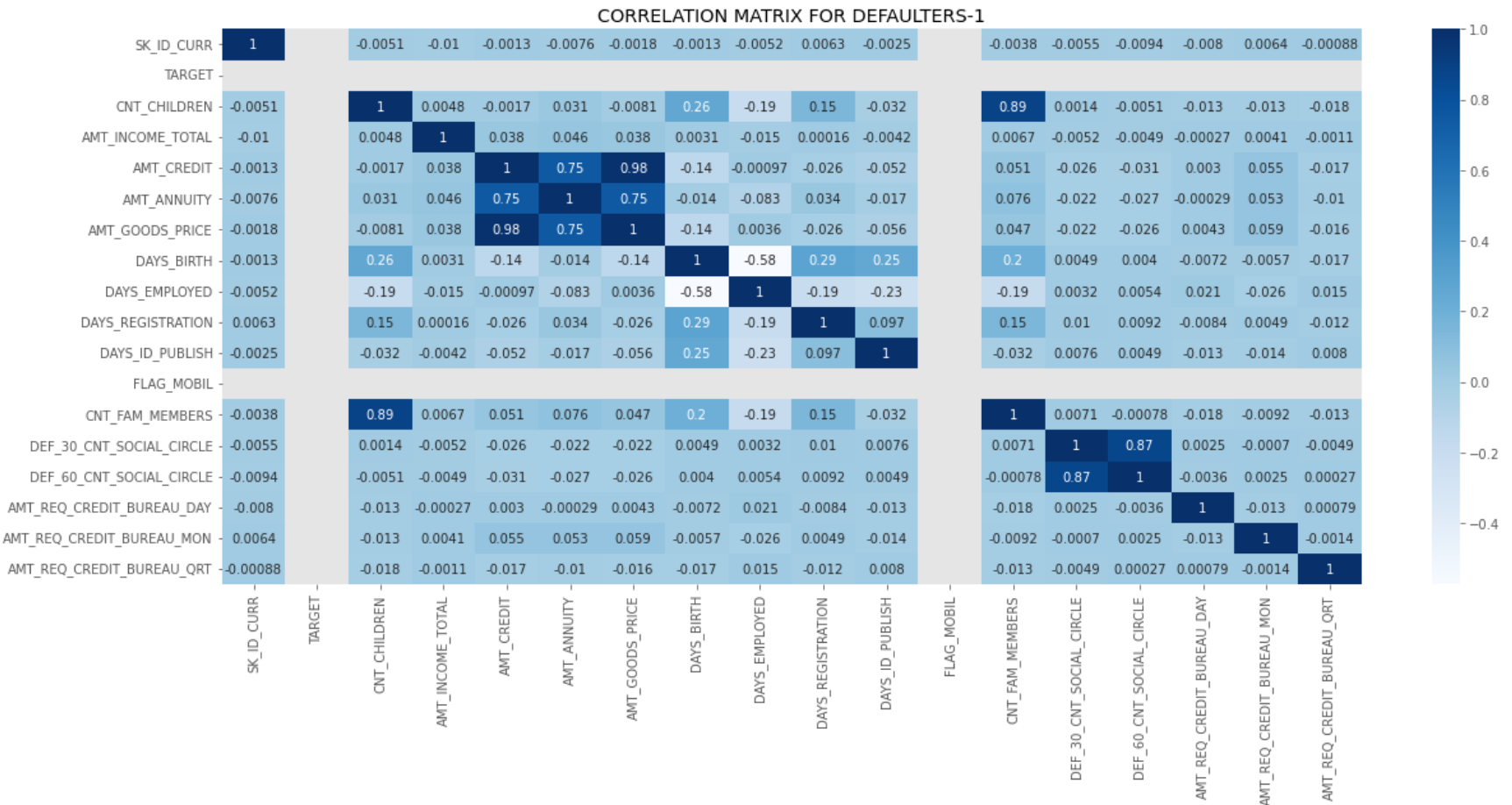
---

#SR.NO.	VAR1	VAR2	Correlation_Value
▶ #1	AMT_GOODS_PRICE	AMT_CREDIT	0.99
▶ #2	CNT_FAM_MEMBERS	CNT_CHILDREN	0.88
▶ #3	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.86
▶ #4	AMT_GOODS_PRICE	AMT_ANNUITY	0.78
▶ #5	AMT_ANNUITY	AMT_CREDIT	0.77
▶ #6	AMT_ANNUITY	AMT_INCOME_TOTAL	0.42
▶ #7	AMT_GOODS_PRICE	AMT_INCOME_TOTAL	0.35
▶ #8	AMT_CREDIT	AMT_INCOME_TOTAL	0.34
▶ #9	DAYS_BIRTH	CNT_CHILDREN	0.34
▶ #10	DAYS_REGISTRATION	DAYS_BIRTH	0.33

---



# # COORELATION MATRIX FOR DEFAULTERS-1



## TOP 10 CORRELATIONS FROM DEFAULTERS-1

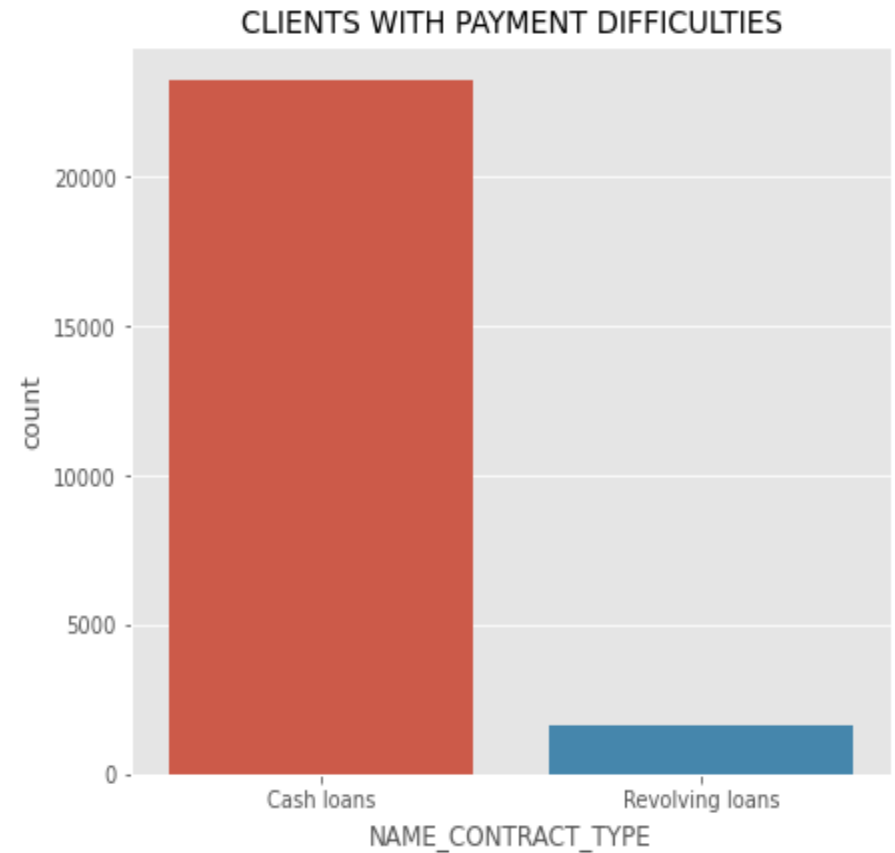
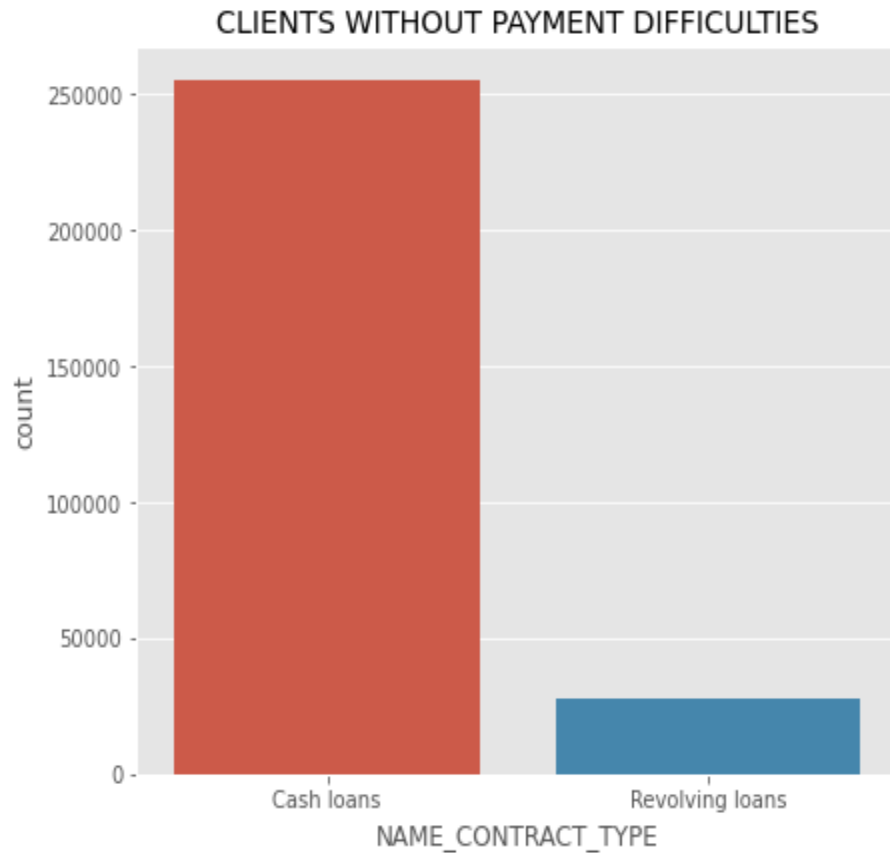
---

#SR.NO.	VAR1	VAR2	Correlation_Value
▶ #1	AMT_GOODS_PRICE	AMT_CREDIT	0.98
▶ #2	CNT_FAM_MEMBERS	CNT_CHILDREN	0.89
▶ #3	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CI	0.87
▶ #4	AMT_GOODS_PRICE	AMT_ANNUITY	0.75
▶ #5	AMT_ANNUITY	AMT_CREDIT	0.75
▶ #4	DAYS_REGISTRATION	DAYS_BIRTH	0.29
▶ #5	DAYS_BIRTH	CNT_CHILDREN	0.26
▶ #6	DAYS_ID_PUBLISH	DAYS_BIRTH	0.25
▶ #7	CNT_FAM_MEMBERS	CNT_CHILDREN	0.20
▶ #8	DAYS_REGISTRATION	YEARS_BIRTH	0.15
▶ #9	CNT_FAM_MEMBERS	AMT_ANNUITY	0.076
▶ #10	DAYS_REGISTRATION	DAYS_EMPLOYED	0.034

---

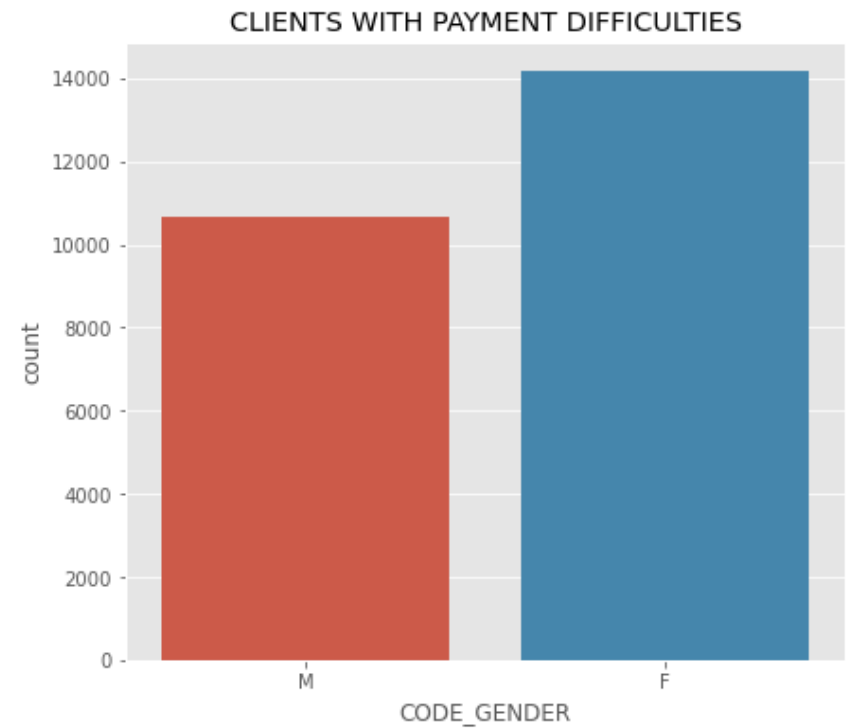
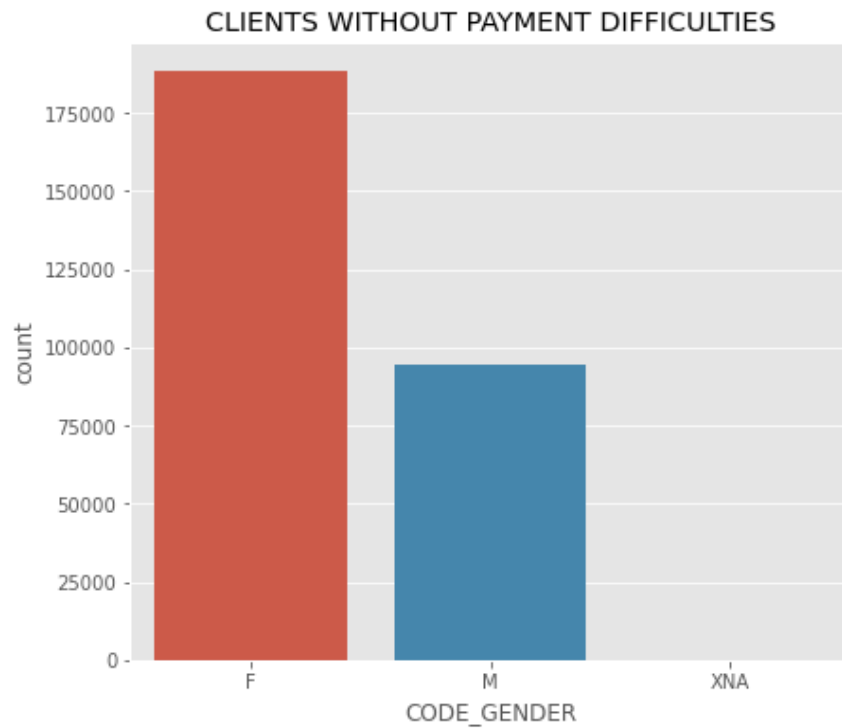


**As shown clients with or without payment difficulties are taking cash loans in large numbers**

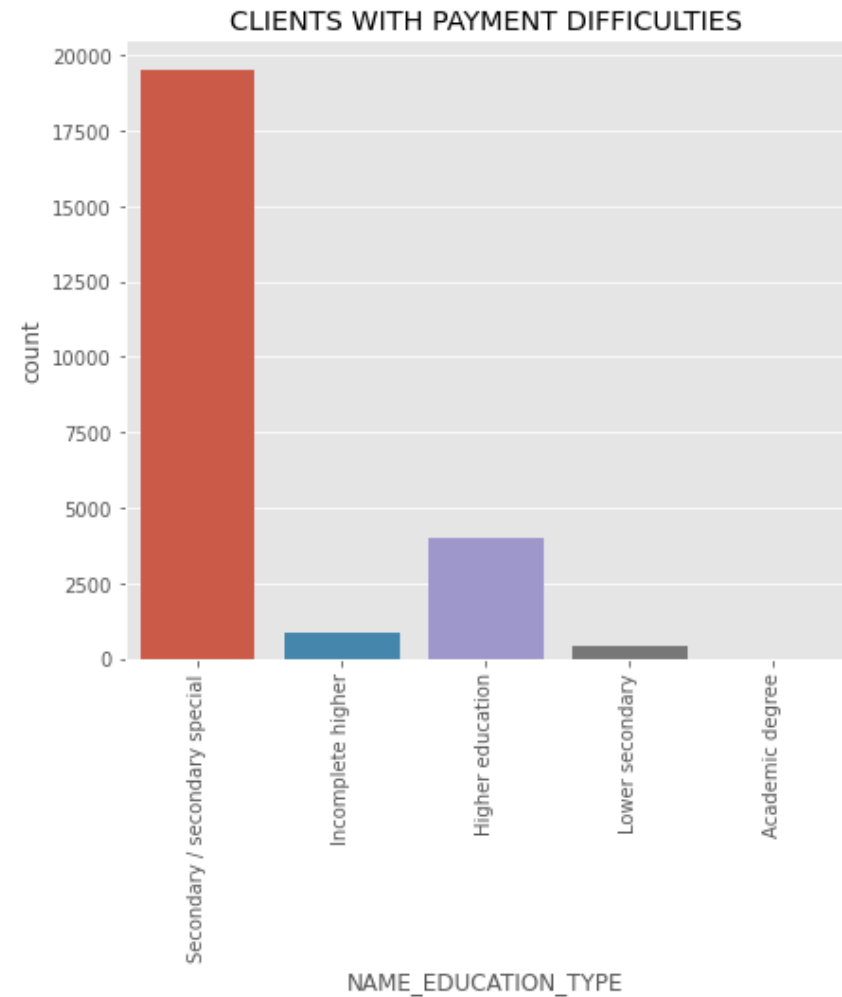
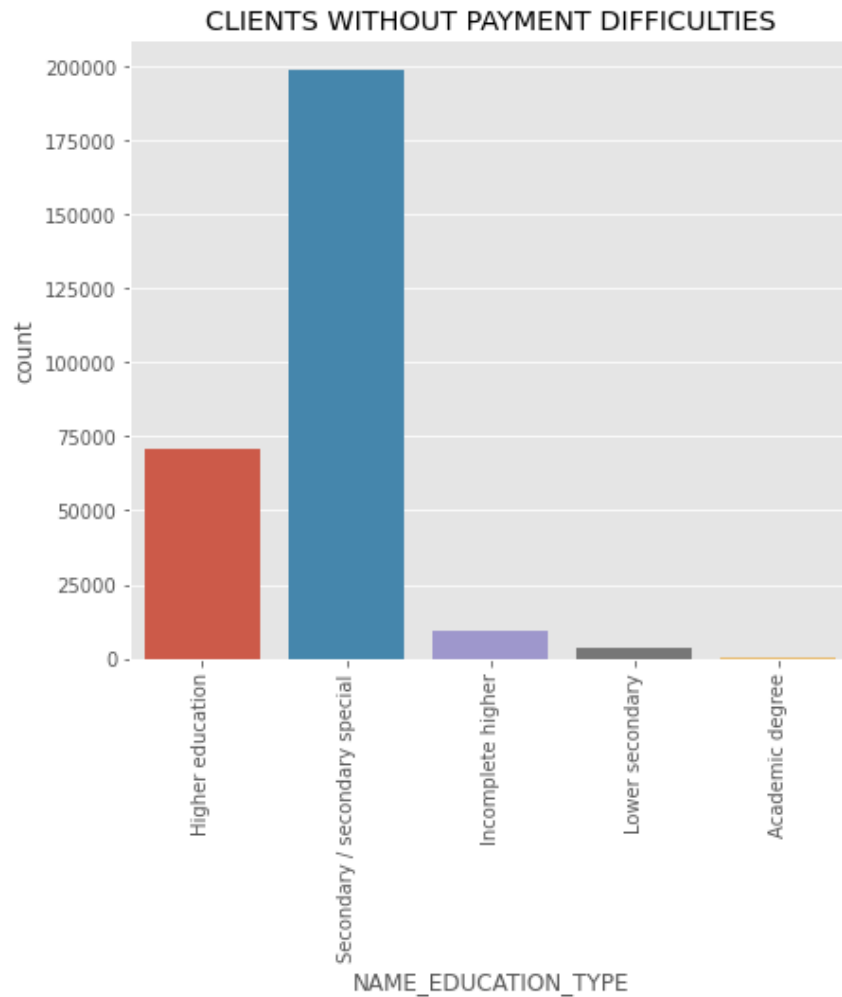


# Females are more in both cases with/without payment difficulties

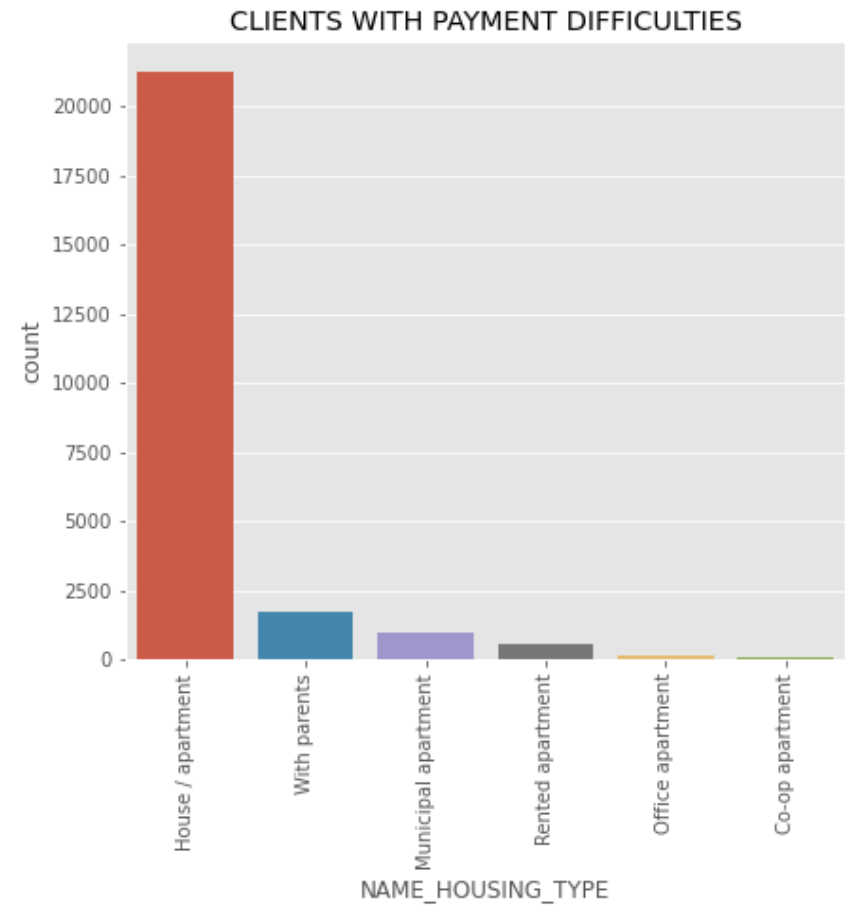
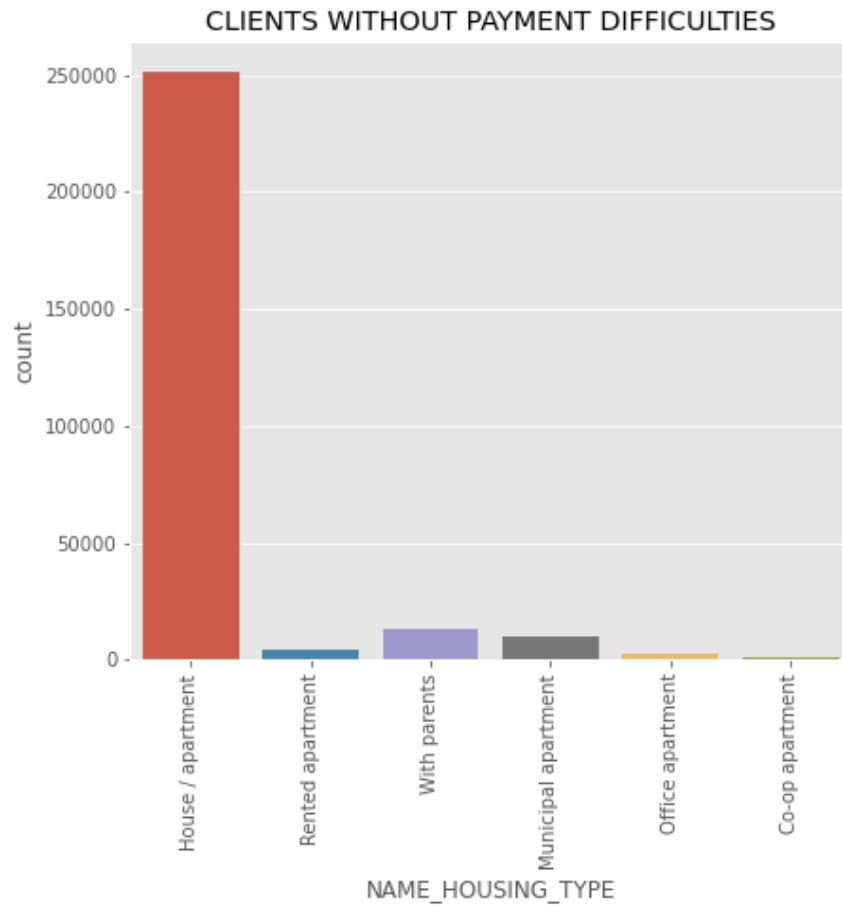
---



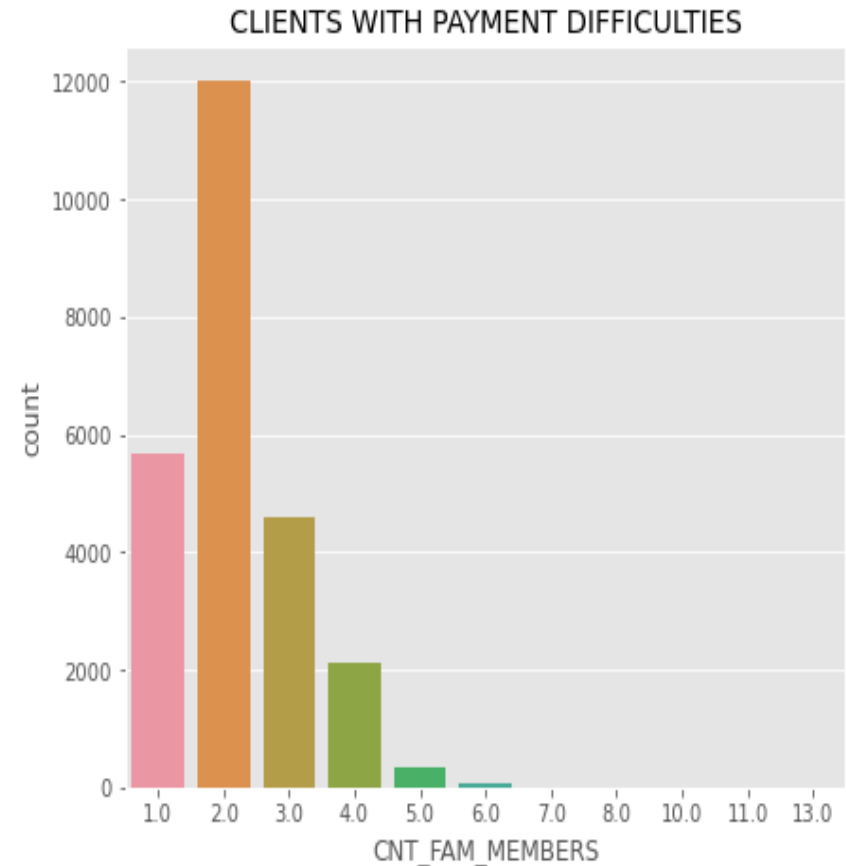
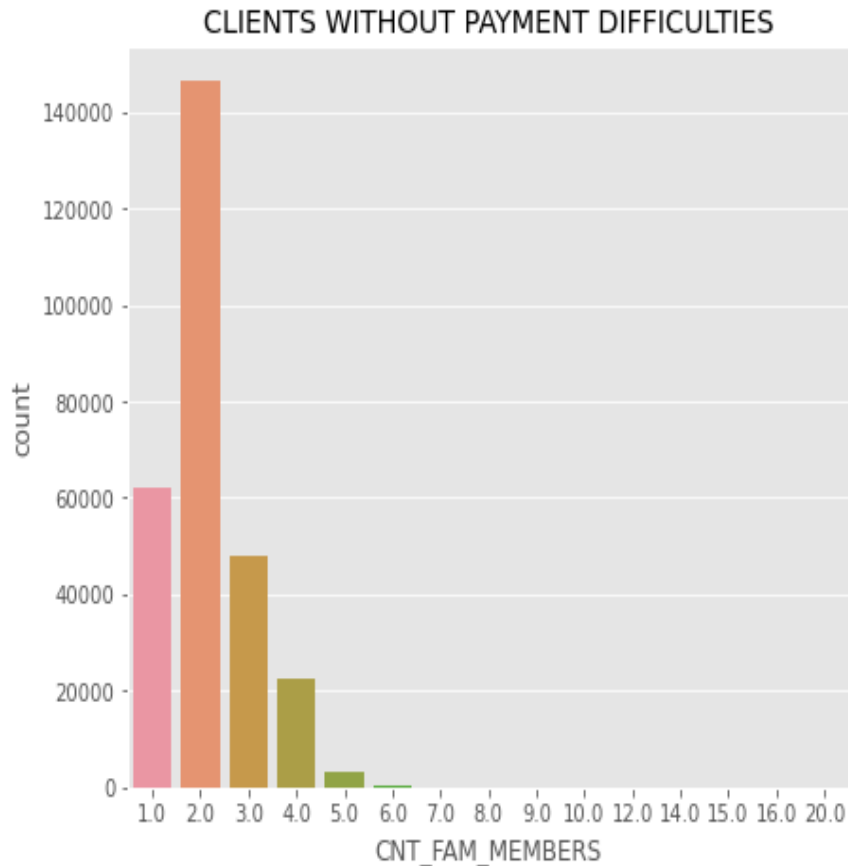
## Clients with and without payment difficulties belong to secondary education category.




Clients are more having a house/apartment in both the cases i.e they may be taking loans for the house/apartment renovation etc

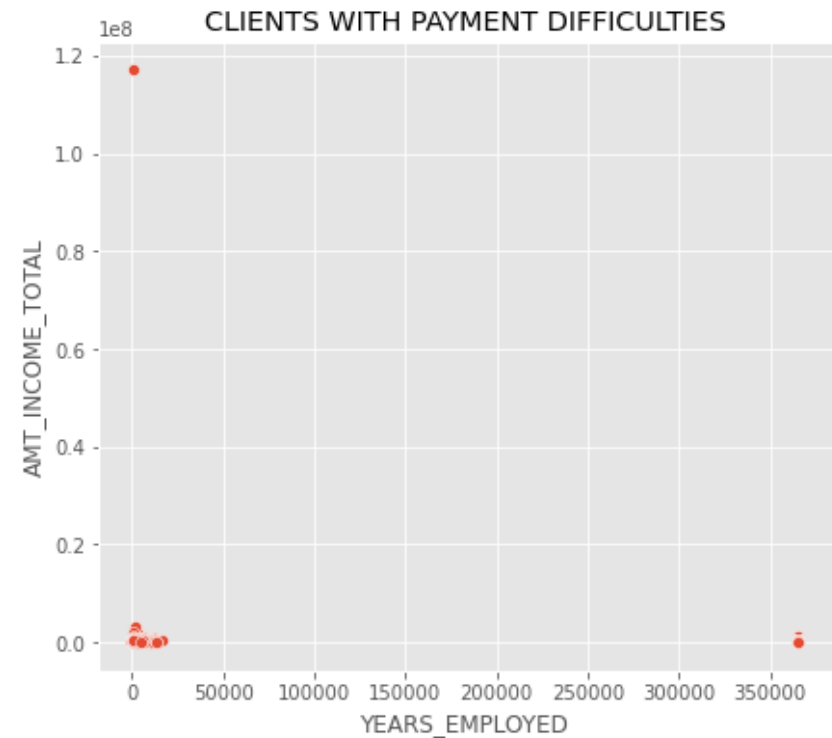
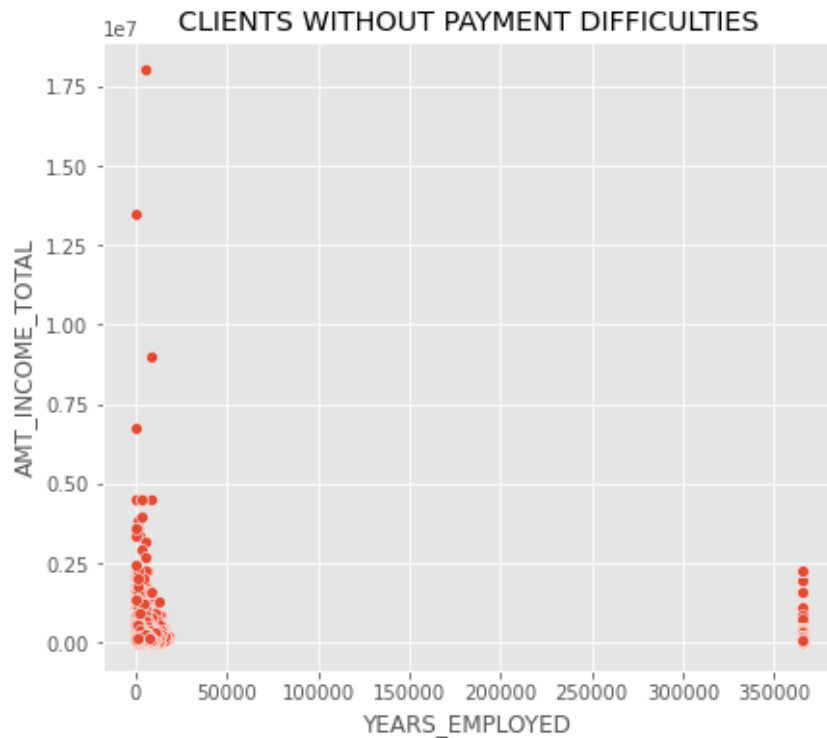


## Clients with 2 family members are applying for loan (maximum in number) with/without payment difficulties.

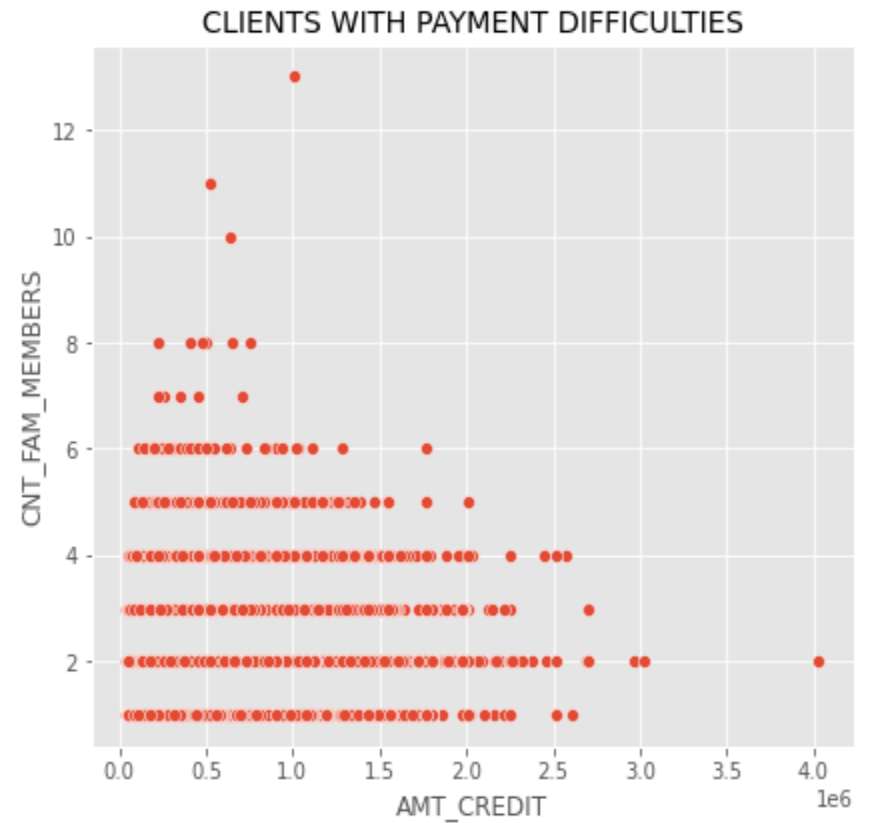
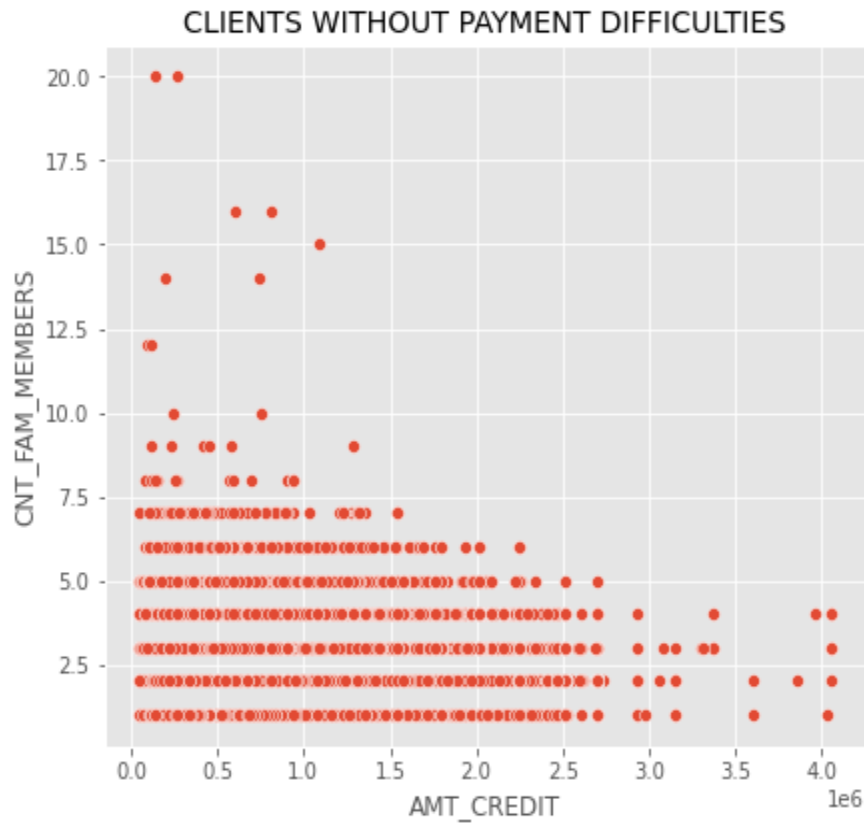


Total years for which the clients have been employed/working is approximately less than 50 years and having good income level in case of without payment difficulties.   
In case of payment with difficulties the clients have been employed for comparatively lesser number of years and have low income level.

---

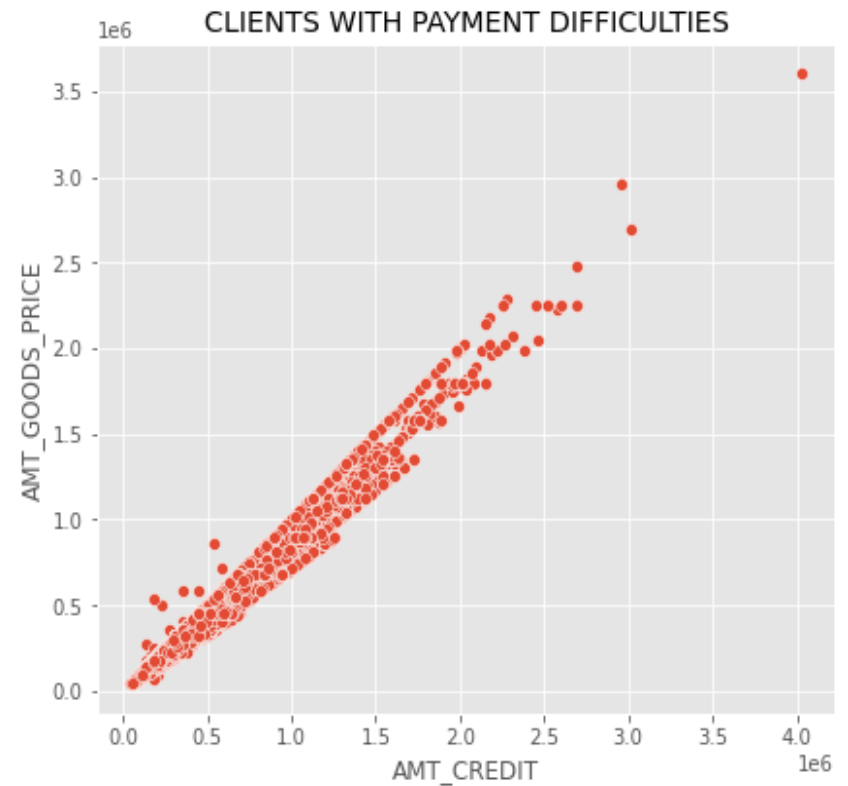
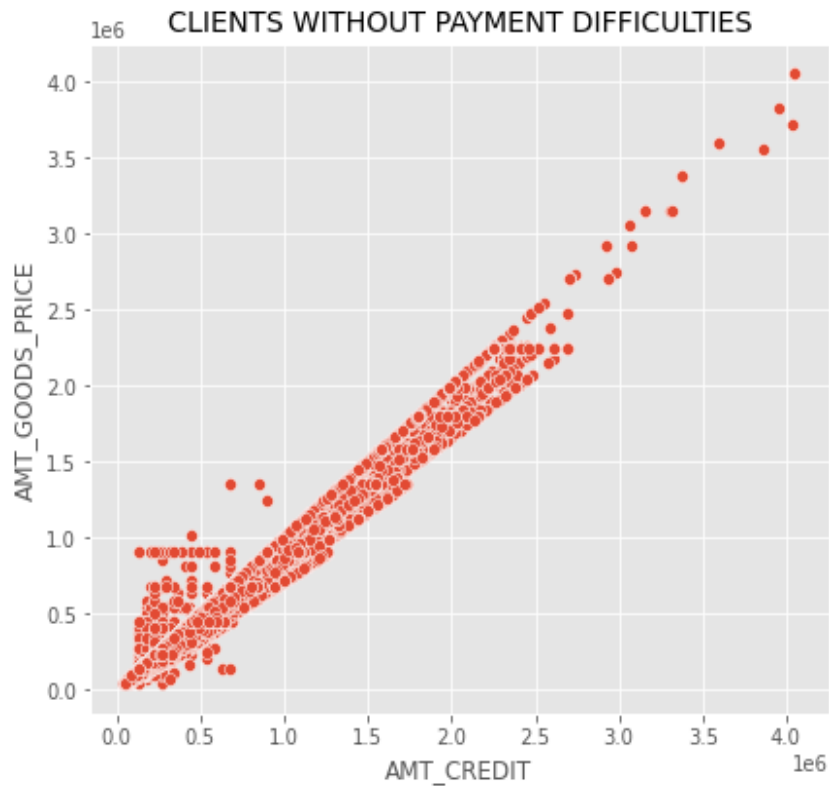


**Clients with lesser count of family members take small loans and are more in number. Clients taking large amount of loan are defaulting less with a comparative good number of family members .**



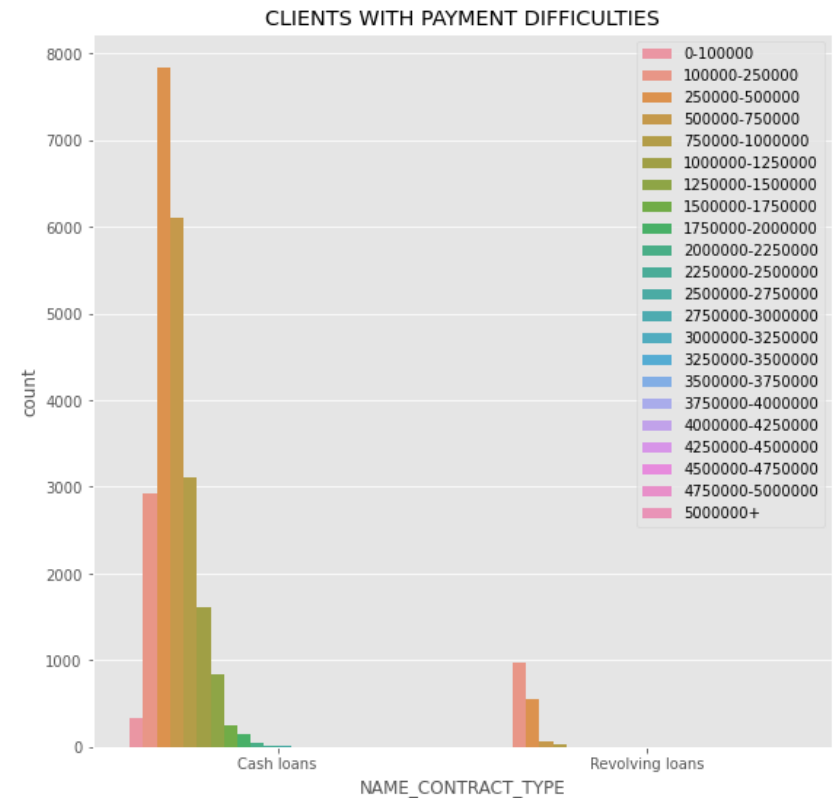
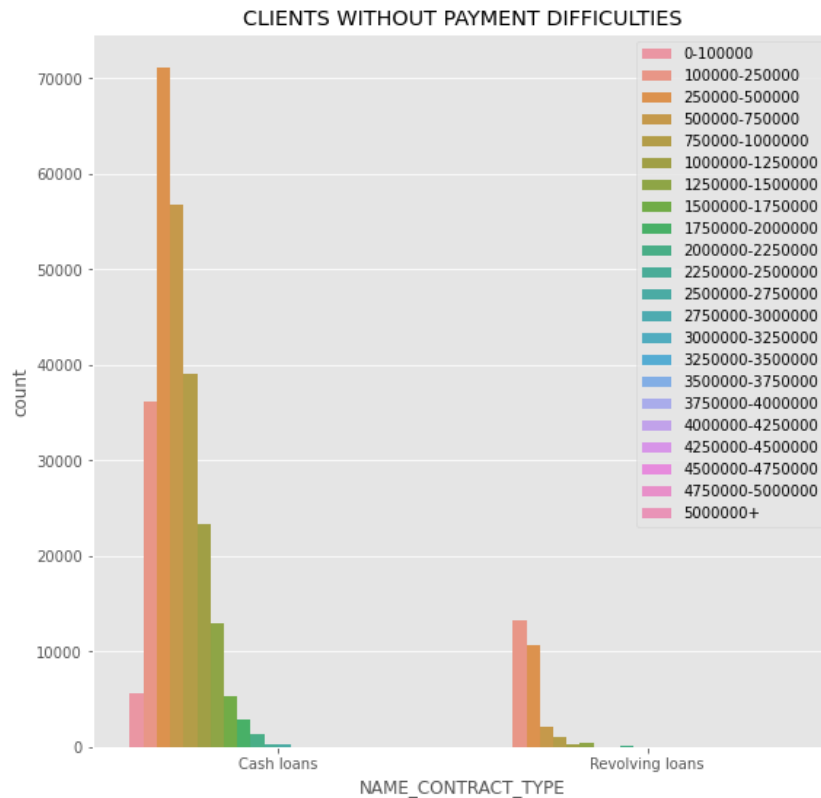
It is clear from below that goods price has positive correlation with credit amount which should be the case.

---

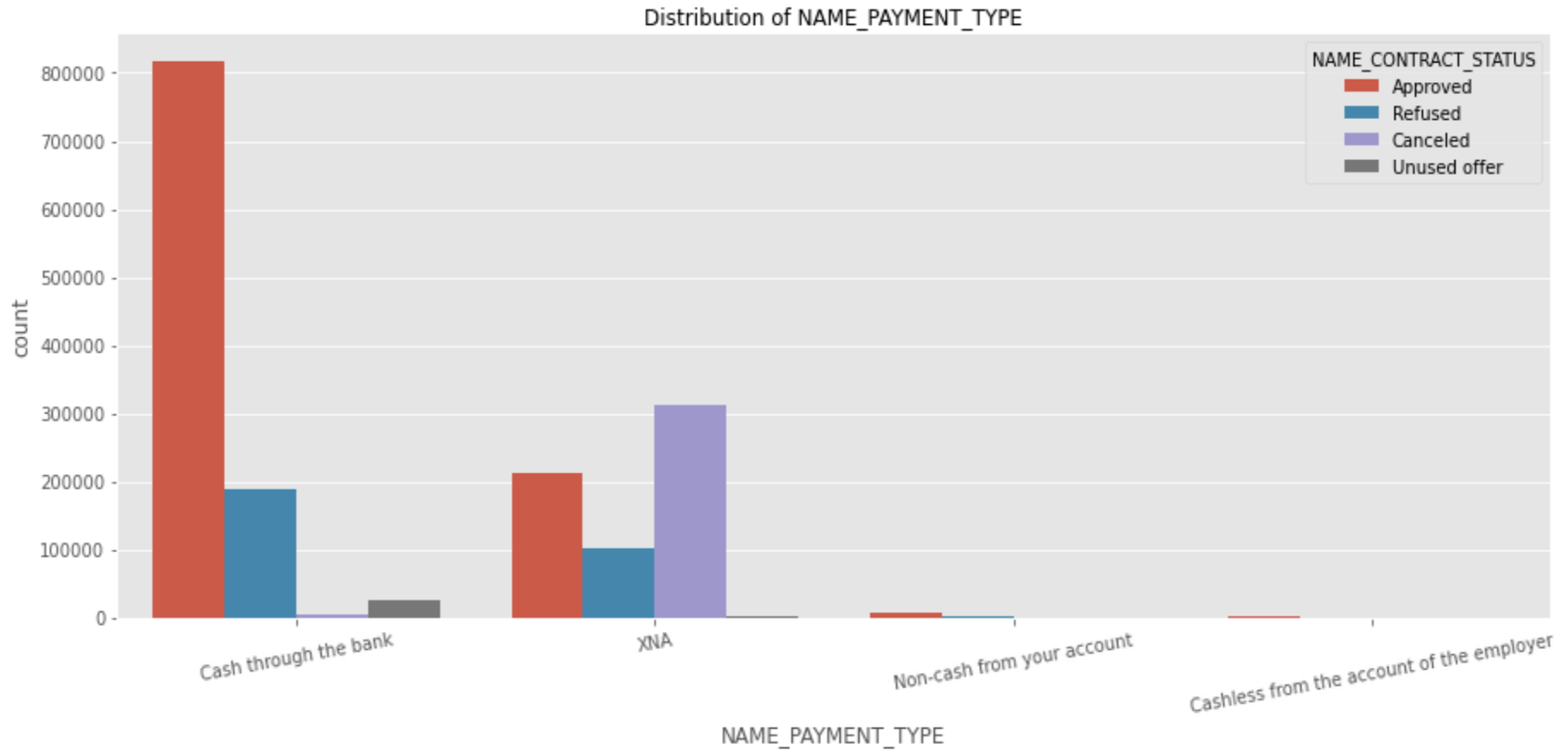




As shown below clients with/without payment difficulties are preferring cash loans instead of revolving loans for the different credit groups formed.

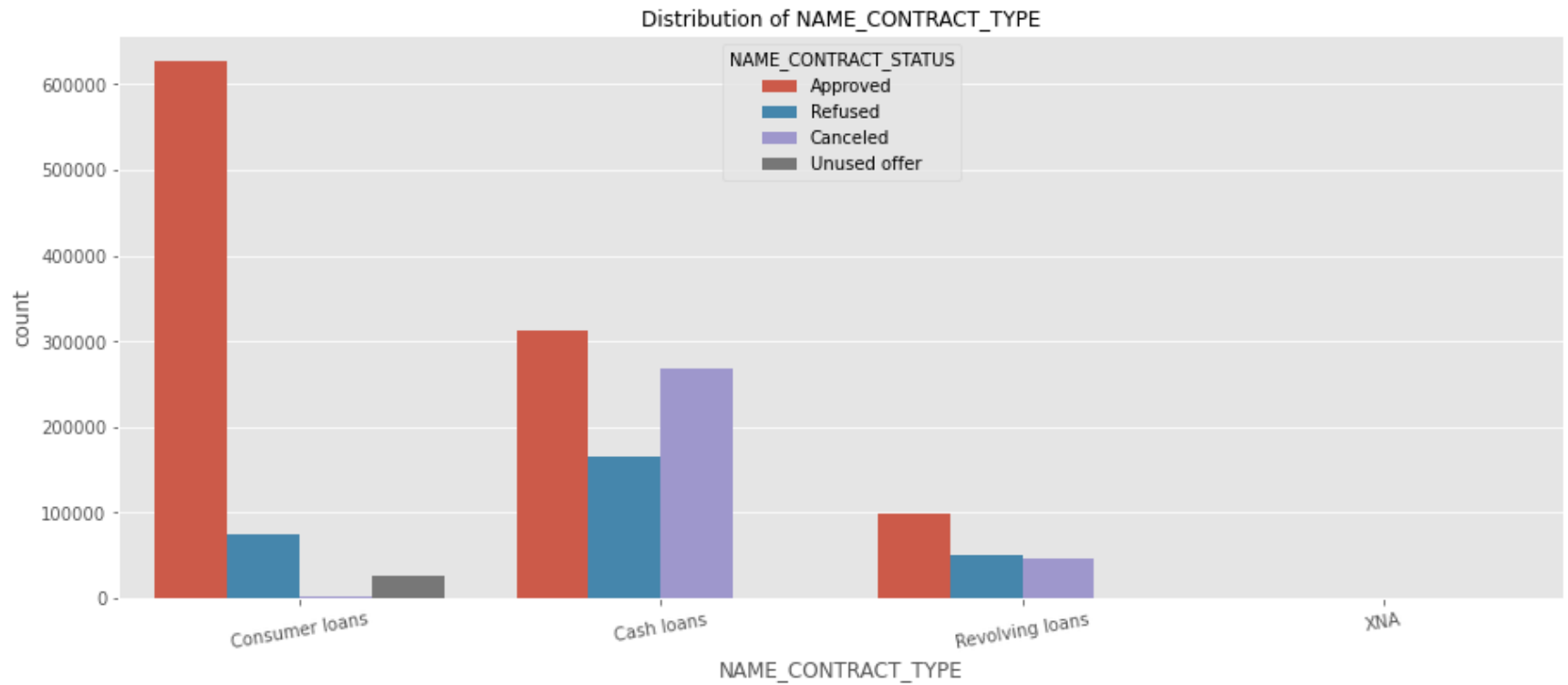


As it is clear from below that the clients have preferred the 'cash through the bank' payment option over other methods of payment

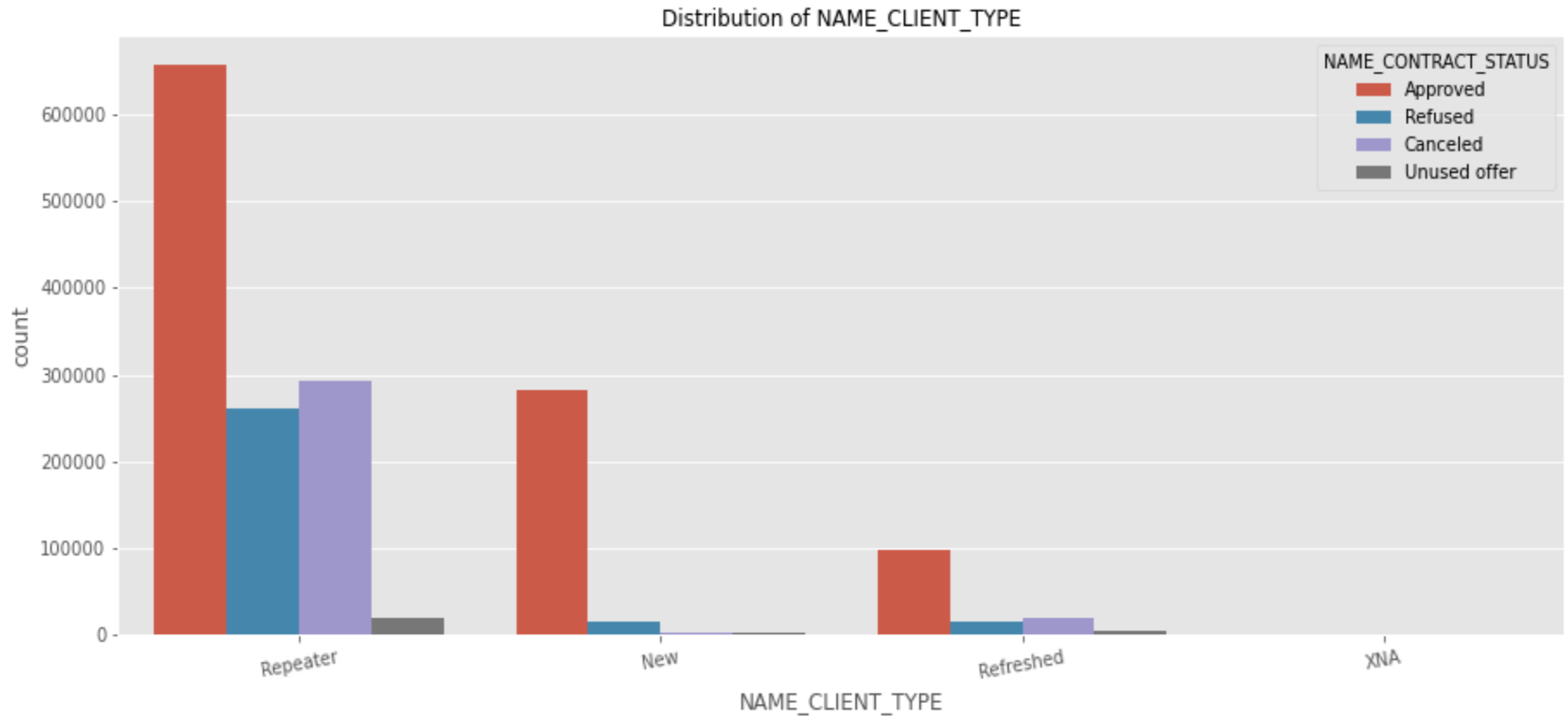


Below graph shows that clients ask for consumer and cash loans in large number and after that 3rd position

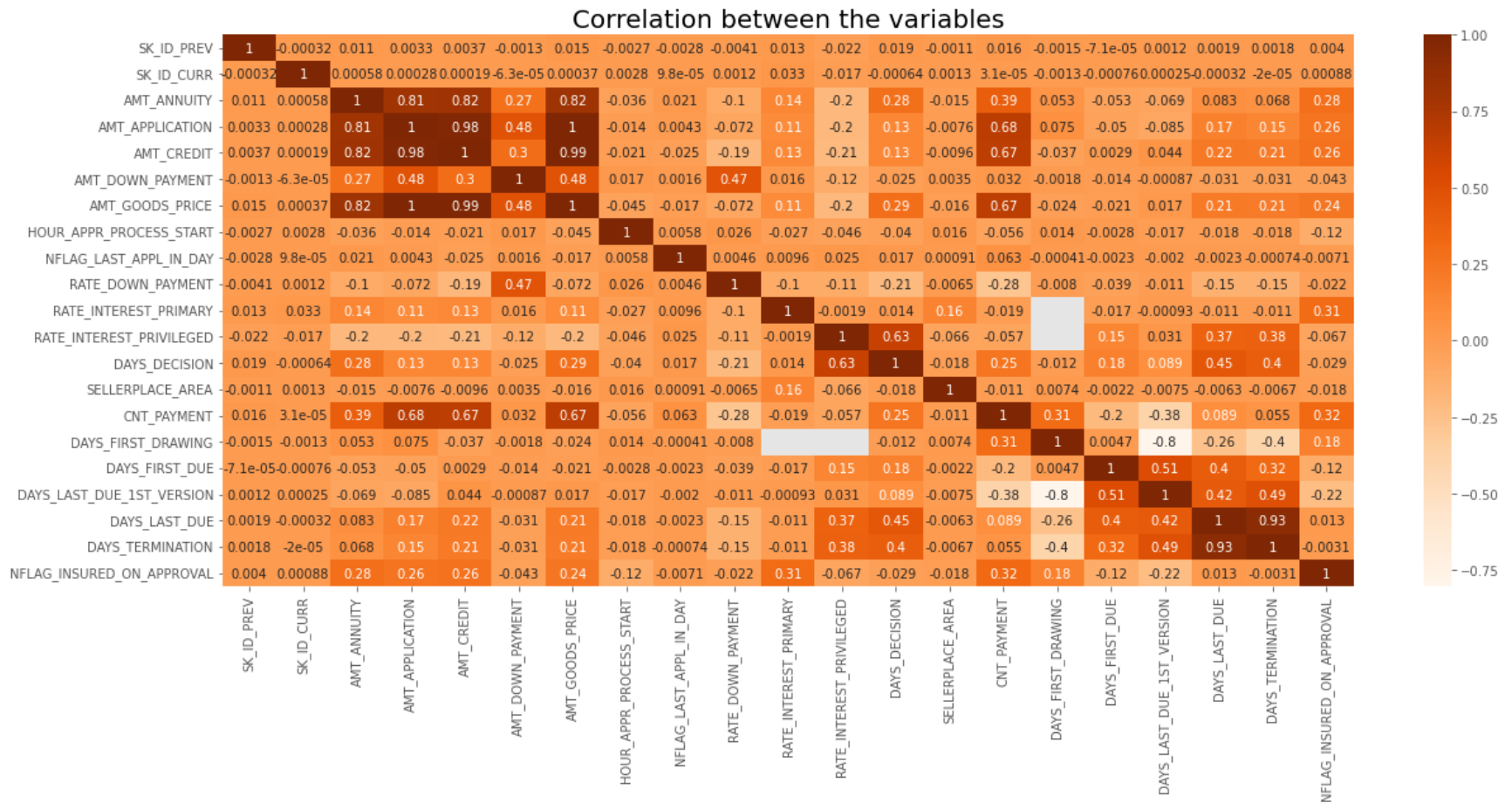
comes for revolving loans. The loans which are refused in large numbers are cash loans ●



The graph above shows that the loan applications which have been approved is highest in case of repeaters and the 2nd category is new clients. The highest number of applications have been rejected in case of repeaters.



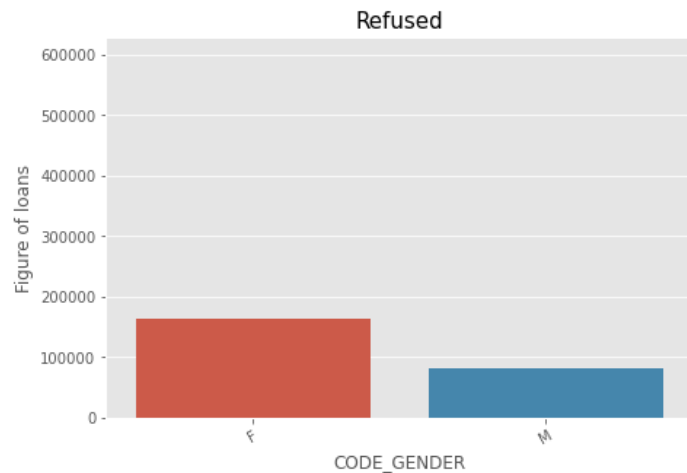
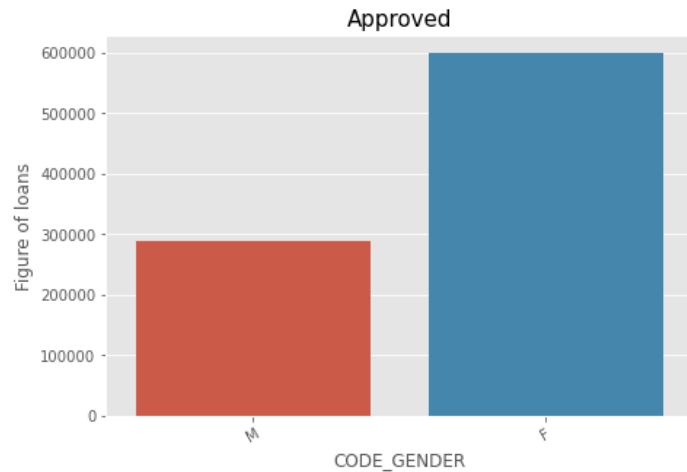
# # Correlation between the variables in previous\_application dataset.



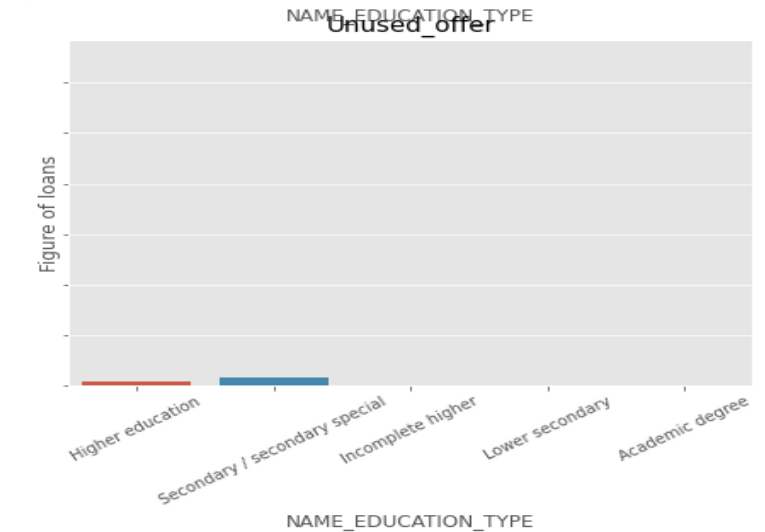
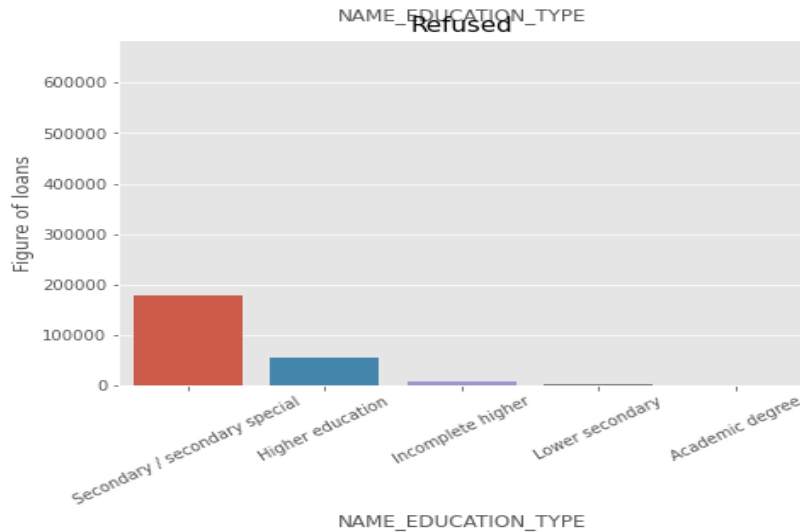
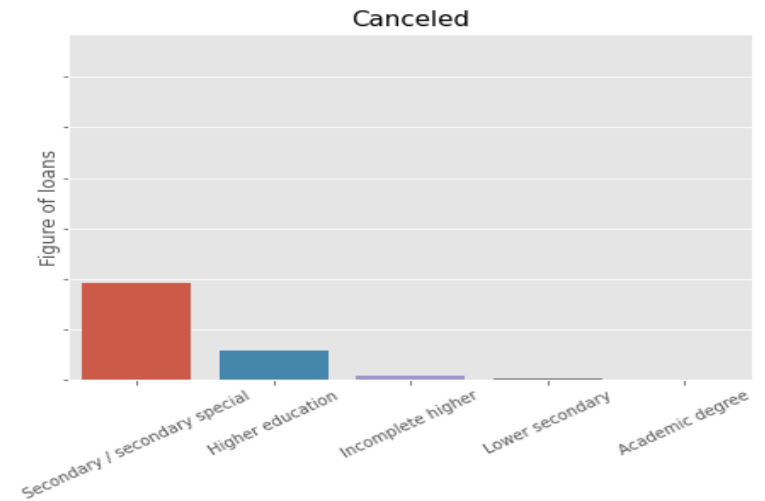
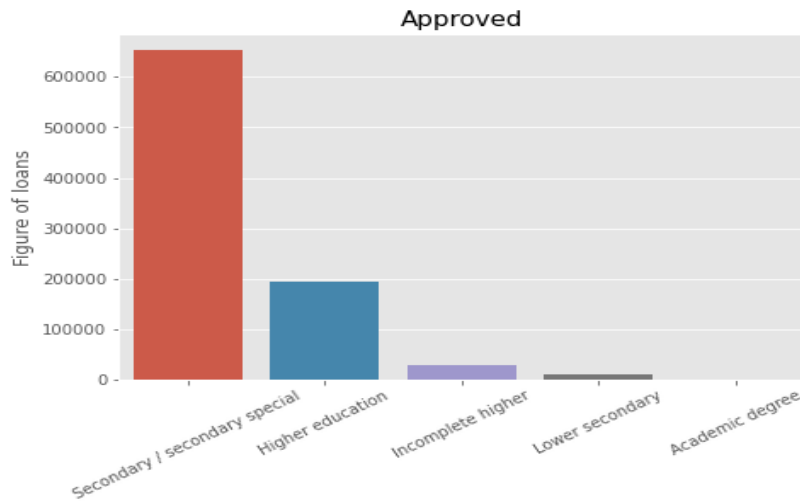
## As per the correlation matrix and heatmap plotted below, the top 10 correlations are given below;

#Sr. No.	Var1	Var2	Correlation value
▶ #1.	AMT_GOODS_PRICE	AMT_APPLICATION	0.99
▶ #2.	AMT_GOODS_PRICE	AMT_CREDIT	0.99
▶ #3.	AMT_CREDIT	AMT_APPLICATION	0.98
▶ #4.	DAYS_TERMINATION	DAYS_LAST_DUE	0.93
▶ #5.	AMT_GOODS_PRICE	AMT_ANNUITY	0.82
▶ #6.	AMT_CREDIT	AMT_ANNUITY	0.82
▶ #7.	AMT_APPLICATION	AMT_ANNUITY	0.81
▶ #8.	CNT_PAYMENT	AMT_APPLICATION	0.68
▶ #9.	CNT_PAYMENT	AMT_CREDIT	0.67
▶ #10.	CNT_PAYMENT	AMT_GOODS_PRICE	0.67

**Females in all the four cases are higher in number as compared to males  
i.e females are applying for loan in larger numbers as compared to males.**

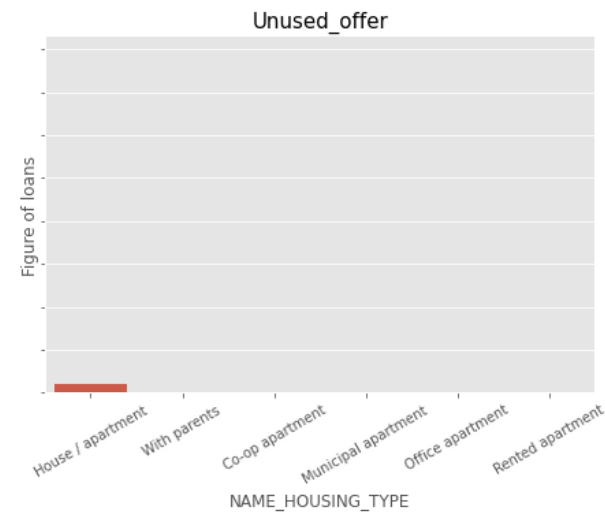
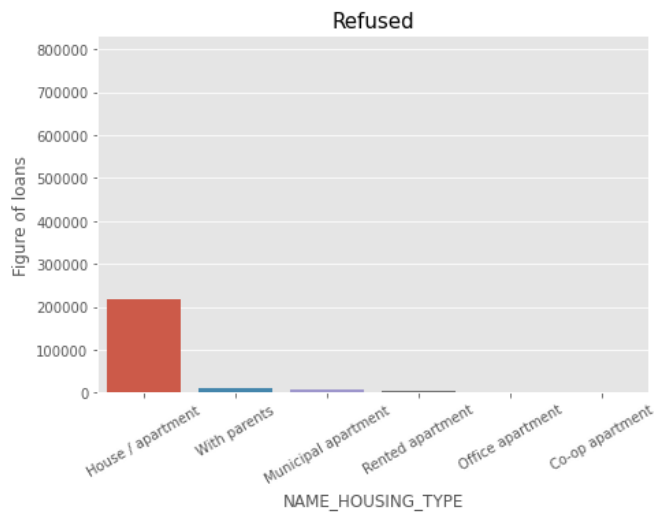
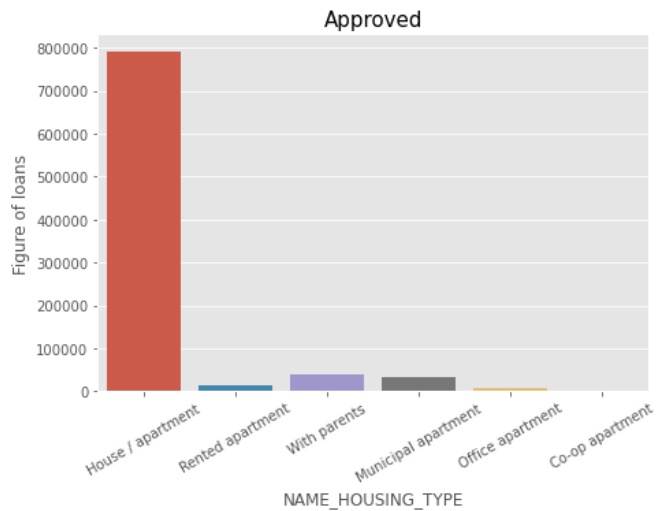


The loans have been approved,cancelled,refused and kept as unused offer in highest numbers for the secondary/secondary special category of eduacted clients.

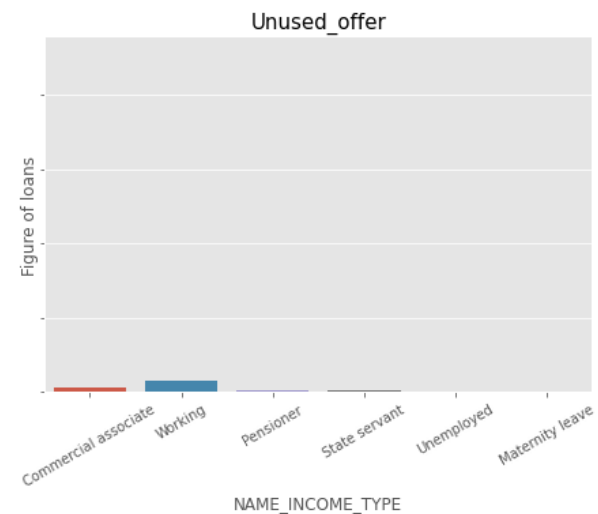
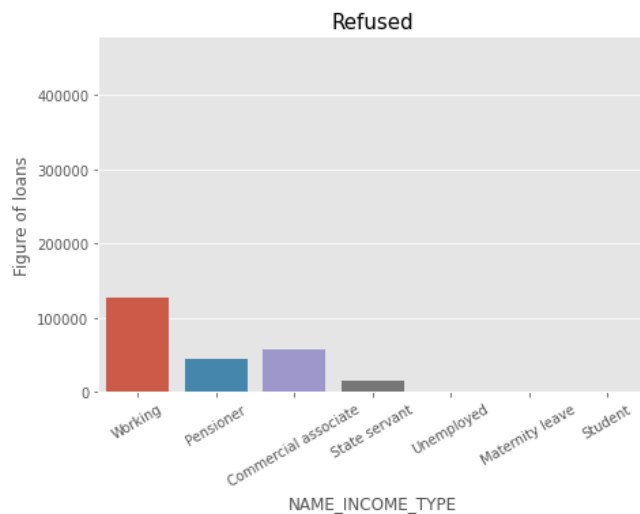
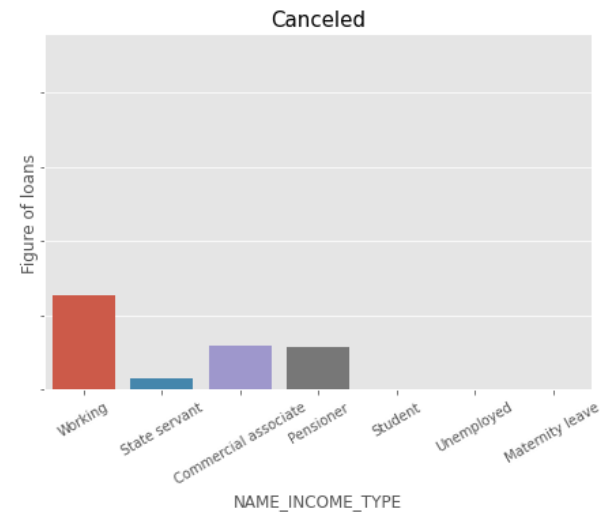
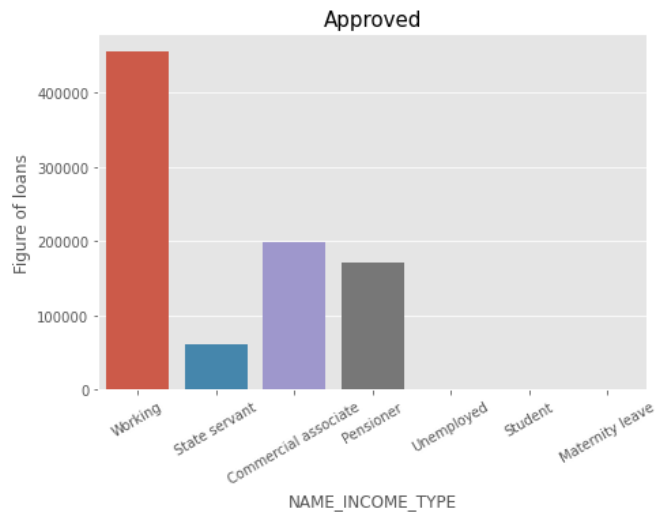




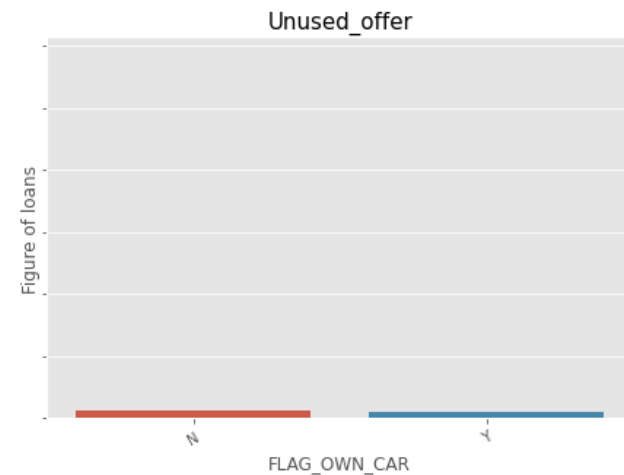
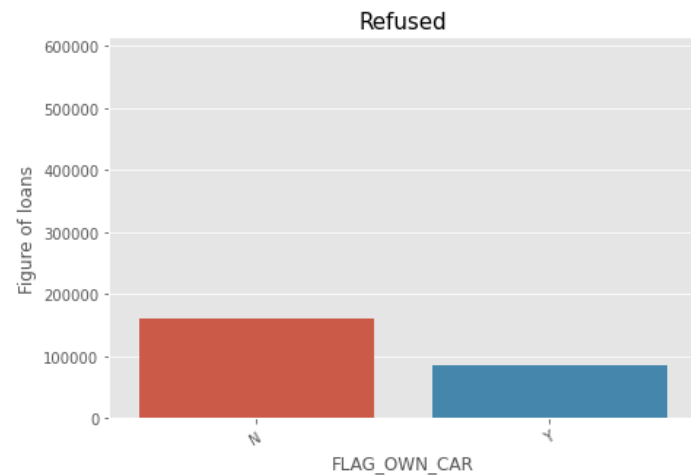
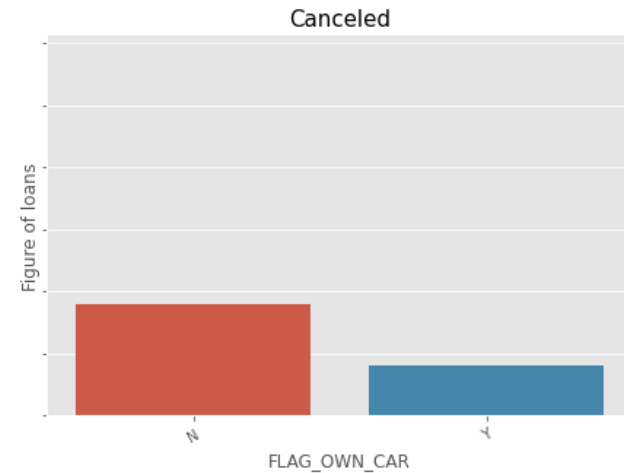
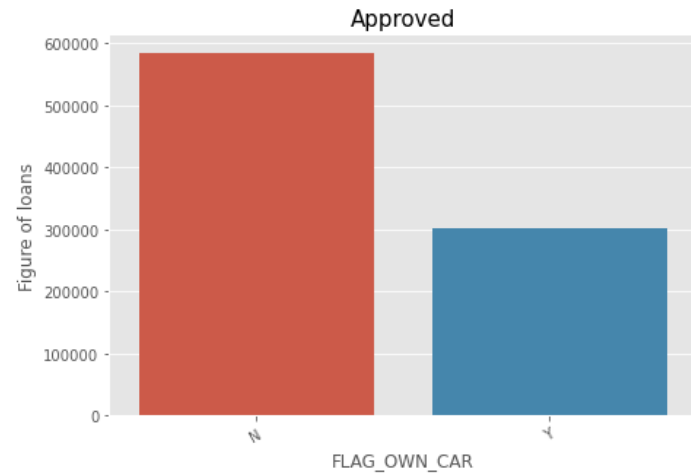
The loans have been approved,cancelled,refused and kept as unused offer in highest numbers for the clients having their own house/apartment.



The loans have been approved,cancelled,refused and kept as unused offer in highest numbers for the clients in the working category.

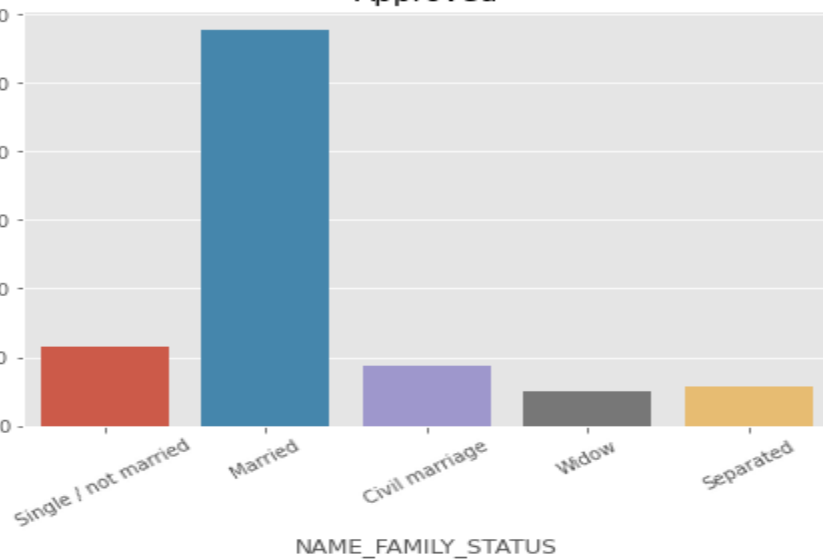


The loans have been approved,cancelled,refused and kept as unused offer in highest numbers for the clients not having an own car.

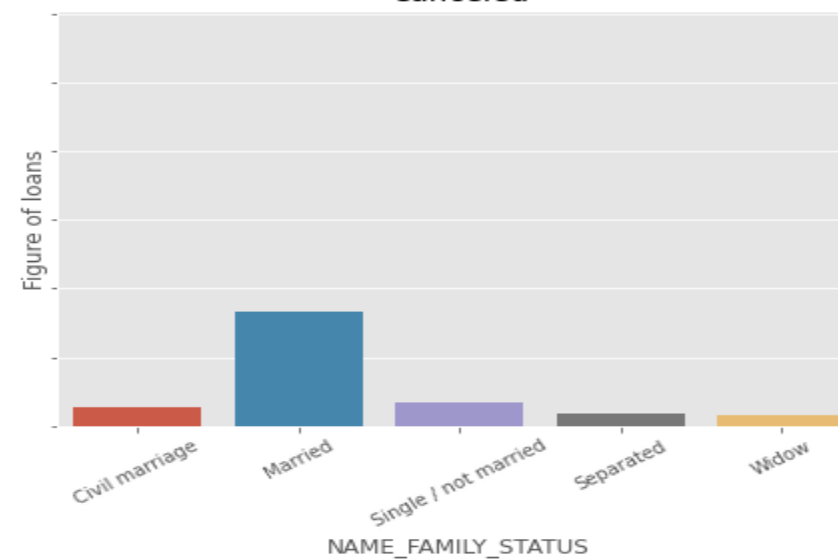


The loans have been approved,cancelled,refused and kept as unused offer in highest numbers for the clients in married category.

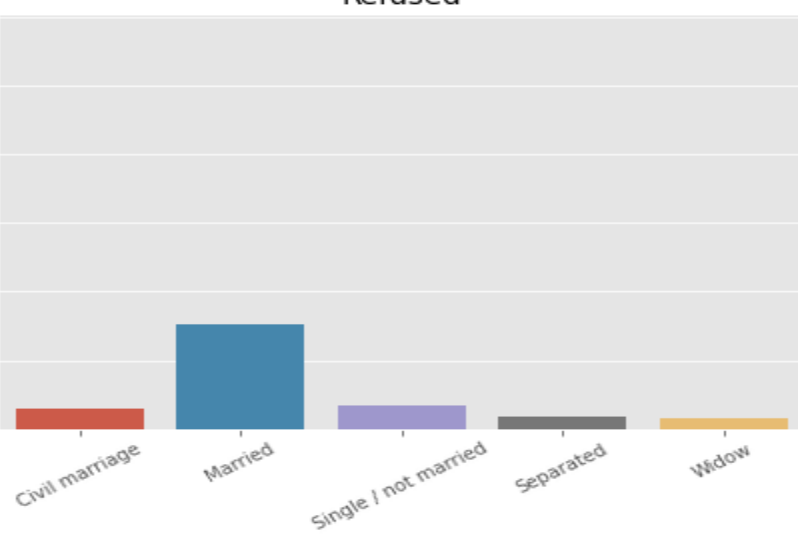
Approved



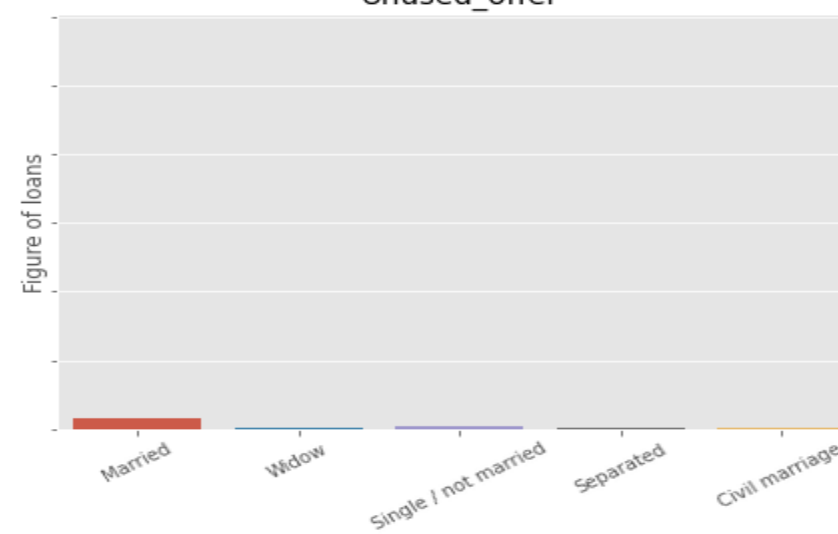
Canceled



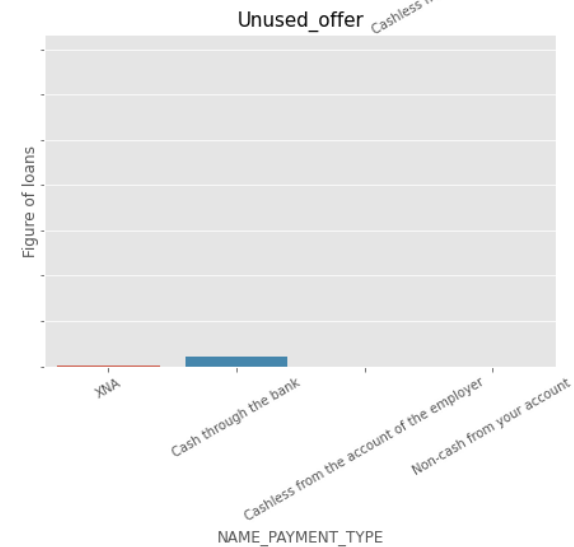
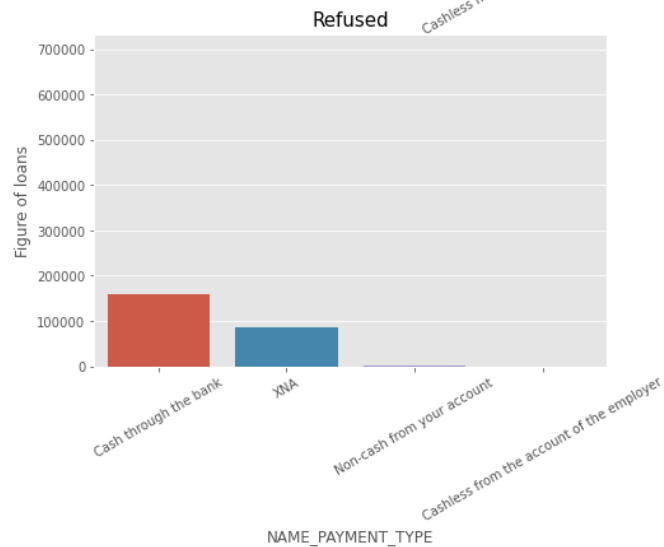
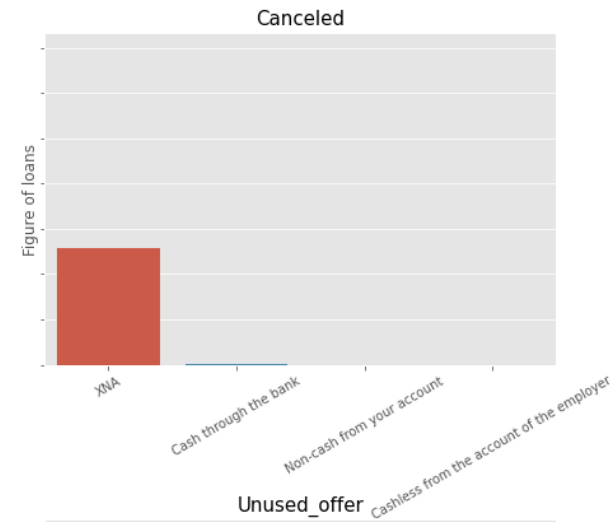
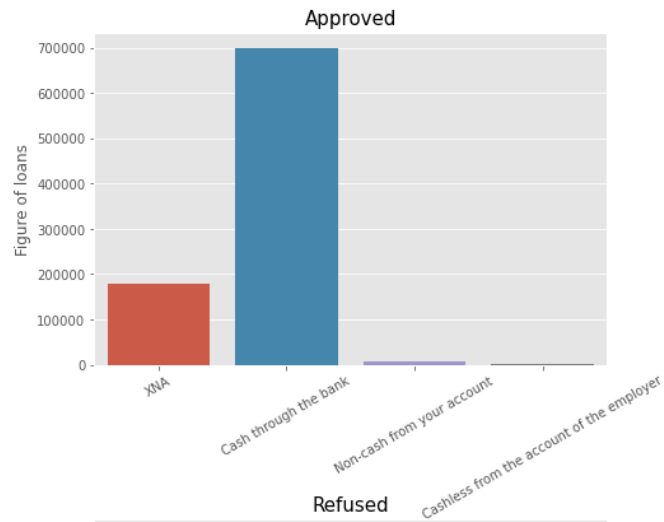
Refused



Unused\_offer



The loans have been approved, refused and kept as unused offer are highest for the clients who have opted for payment option as 'cash through the bank' and cancelled for the clients who opted XNA mode of payment.

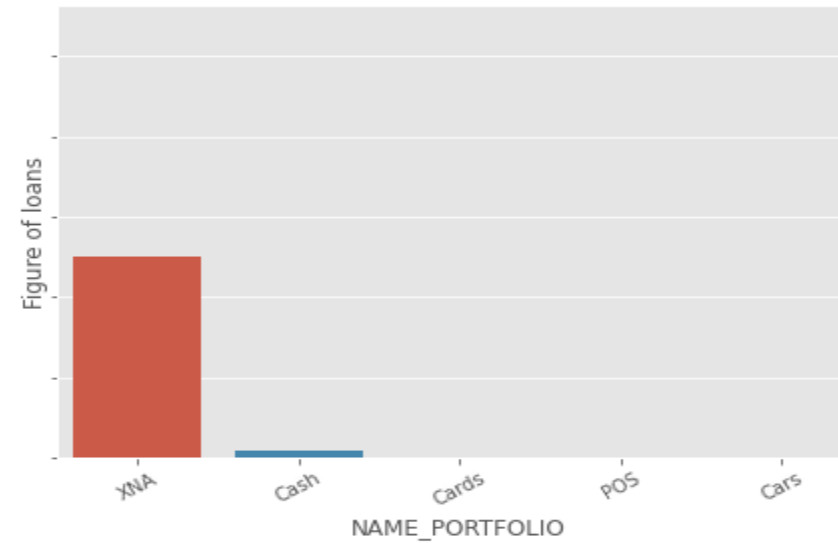


**The most loans which have been approved have been applied through POS and through all the other modes have been either cancelled,refused or kept as unused offer.**

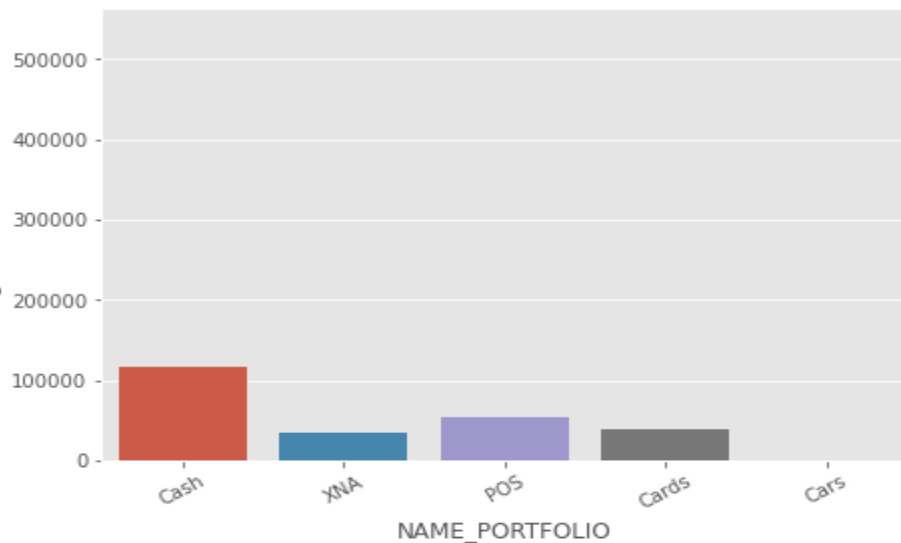
Approved



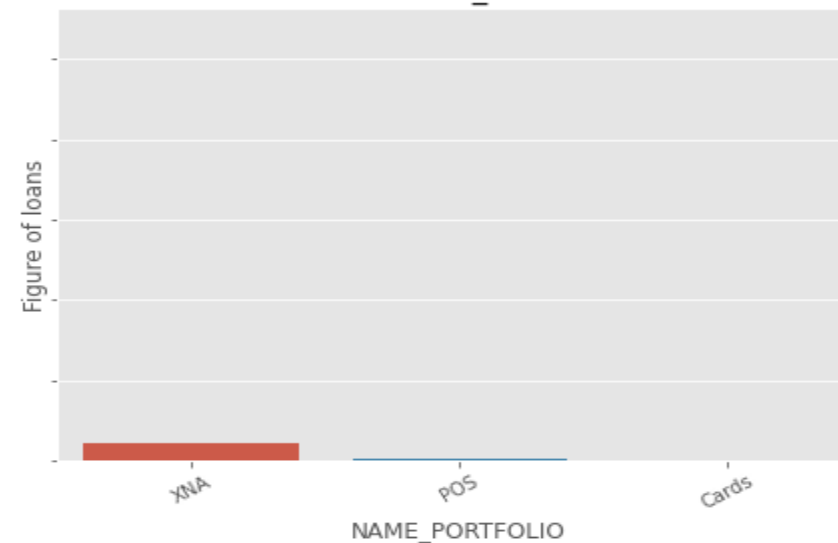
Canceled



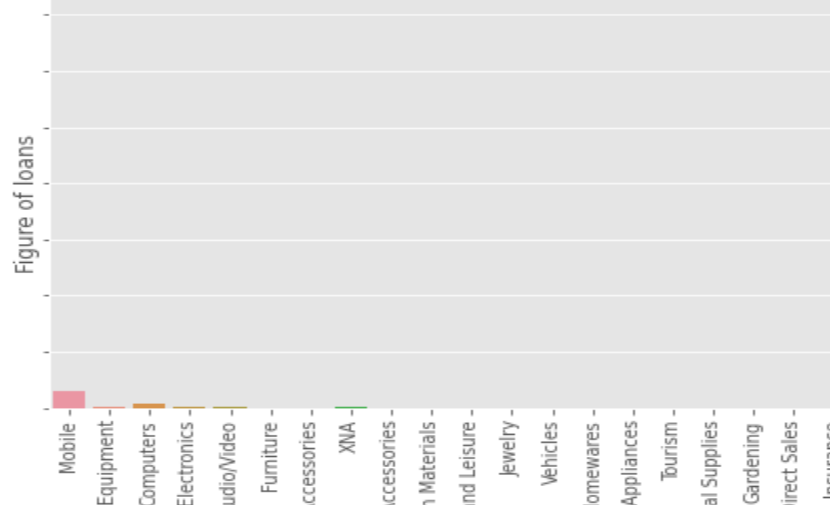
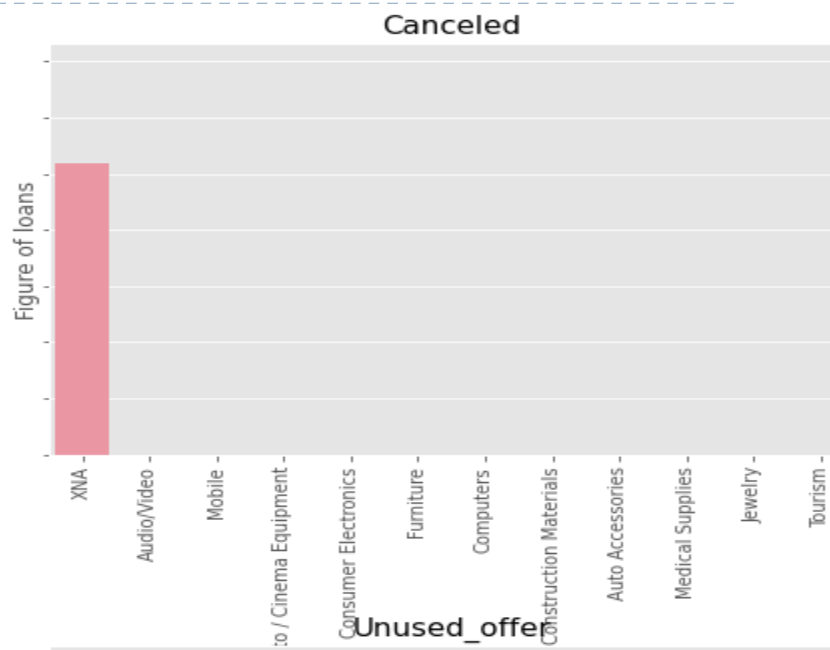
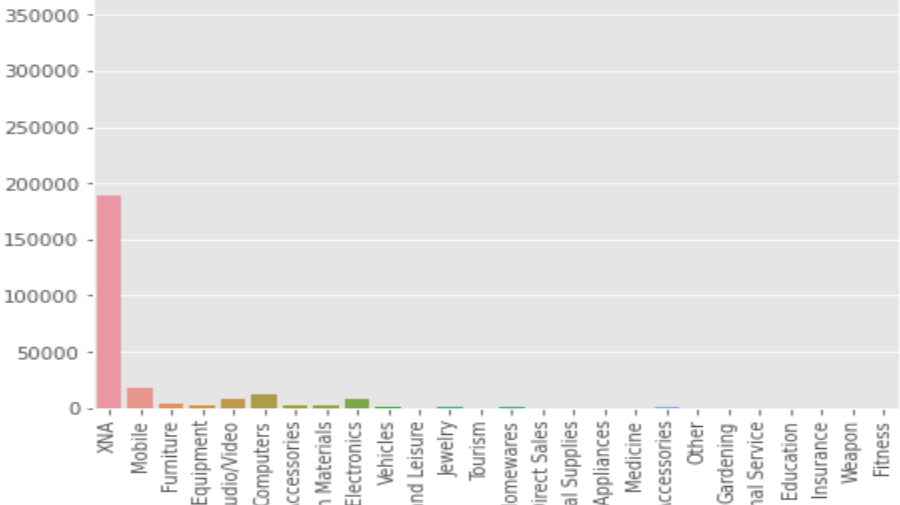
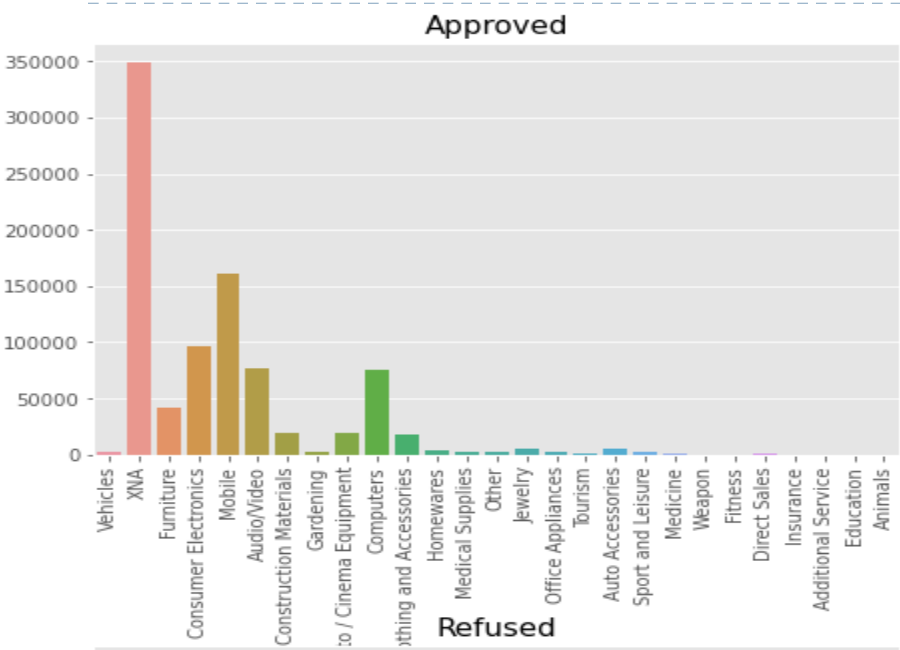
Refused



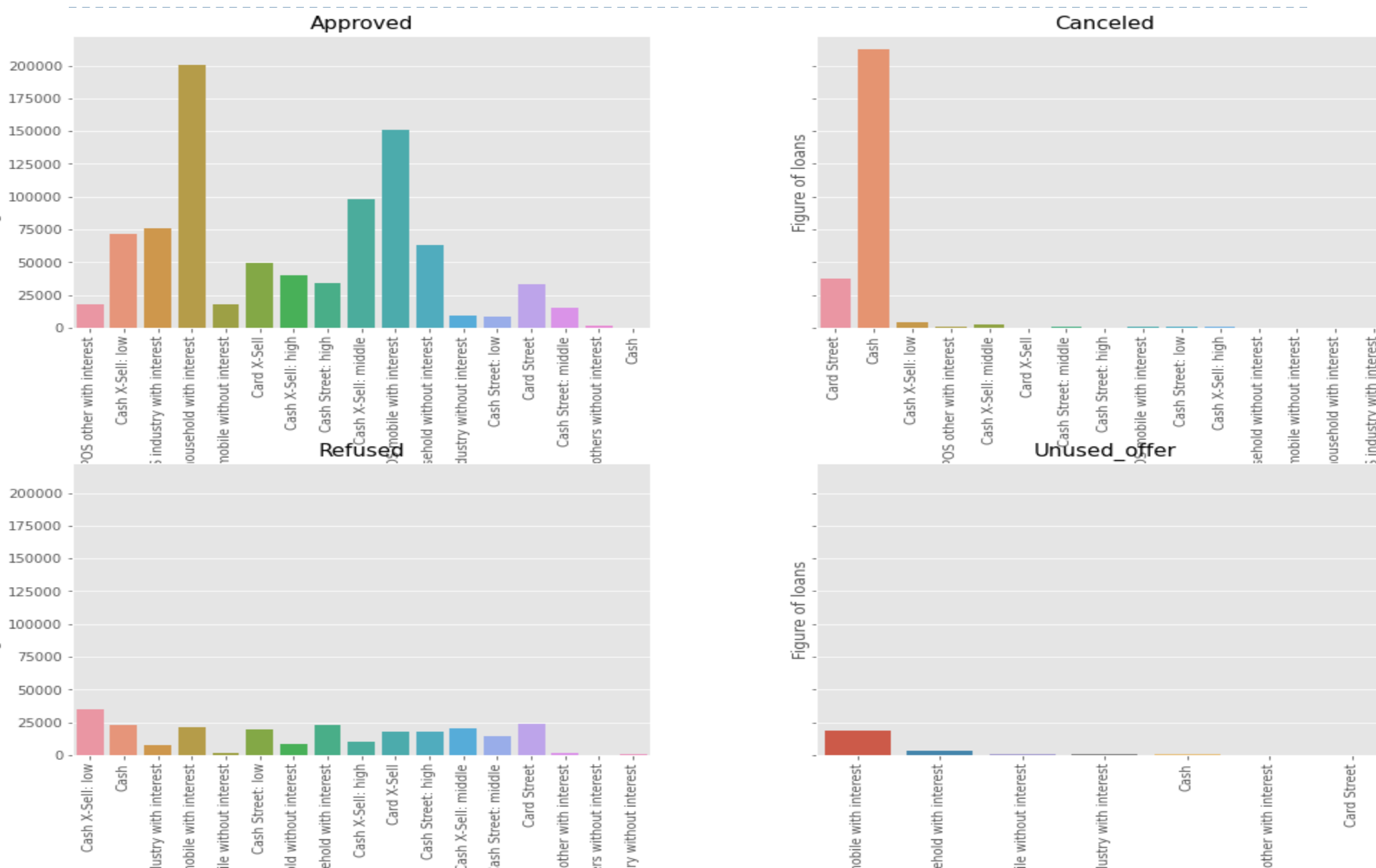
Unused\_offer



The XNA category is highest in loan approval, cancelled,refused and unused offers are maximum in case of mobile.



The most approved loan category is POS household with interest , most cancelled loan category is cash loans and most unused offers include POS mobile with interest.





Maximum approved loans have been applied via country-wide ans credit and cash offices channel type. However loans applied through credit and cash offices have largely been cancelled and refused. The unused offers are maximum in case of country-wide channel type.

