

Presentation on Credit EDA Case Analysis

By Priya Narayan



Problem Statement

A consumer finance company specializing in lending various types of loans to urban customers aims to understand the driving factors behind loan default, in order to utilize this knowledge for its portfolio and risk assessment.

Upon receipt of a loan application, the company has to decide for loan approval based on the applicant's profile. However, the following risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

The company wants a system which can ensure that its consumers who are capable of repaying the loan are not rejected and also to identify patterns which indicate if a client has difficulty paying their instalments. Identification of such applicants using EDA is the aim of this case study.

Analysis Approach

UNDERSTANDING PROBLEM STATEMENT

DATA CLEANING

DATA IMPUTATION

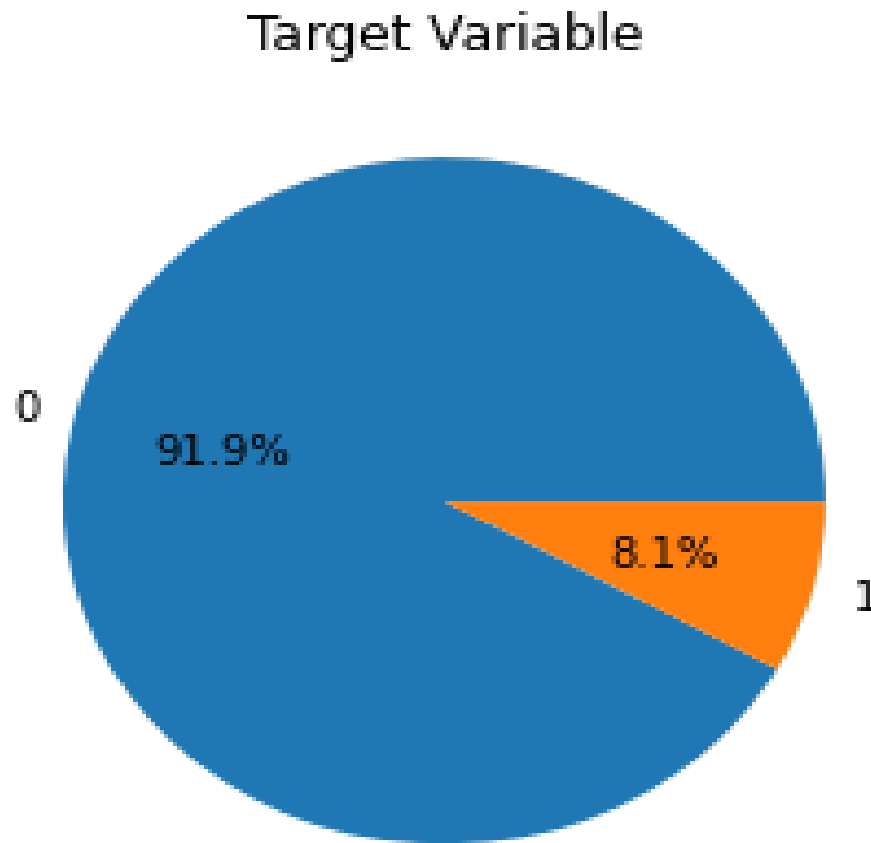
CHECK FOR DATA IMBALANCE AND UNIVARIATE & BIVARIATE ANALYSIS

MERGING OF APPLICATION DATA WITH PREVIOUS DATA

DATA ANALYSIS

CONCLUSION

DATA IMBALANCE

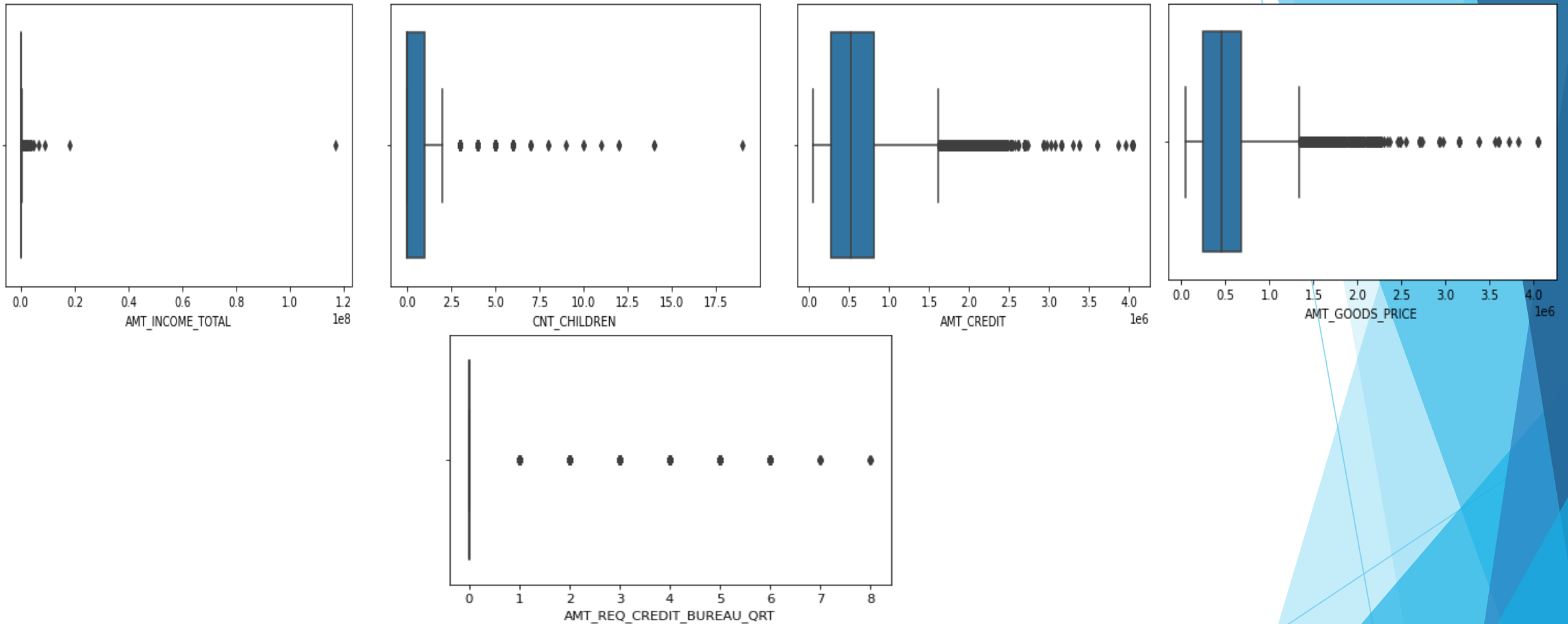


➤ Ratio of 0:1 =11.33:1

#This is an imbalanced dataset as it clearly indicates that for every 1 there is most number of 0's.

➤ Ratio of imbalance dataset in percentage is 91.93% and 8.1%

OUTLIERS ANALYSIS OF APPLICATION DATASET



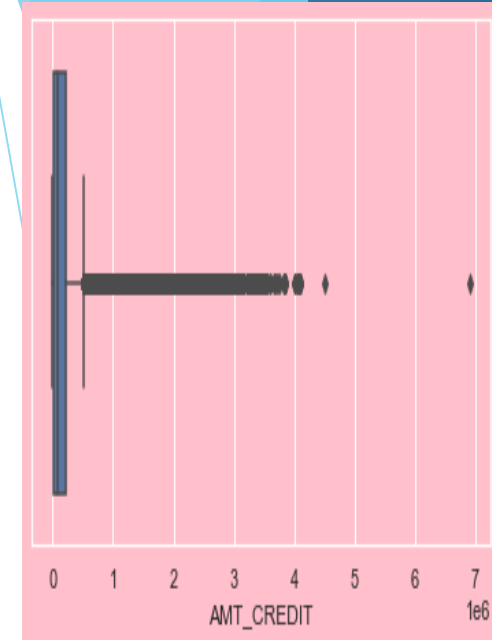
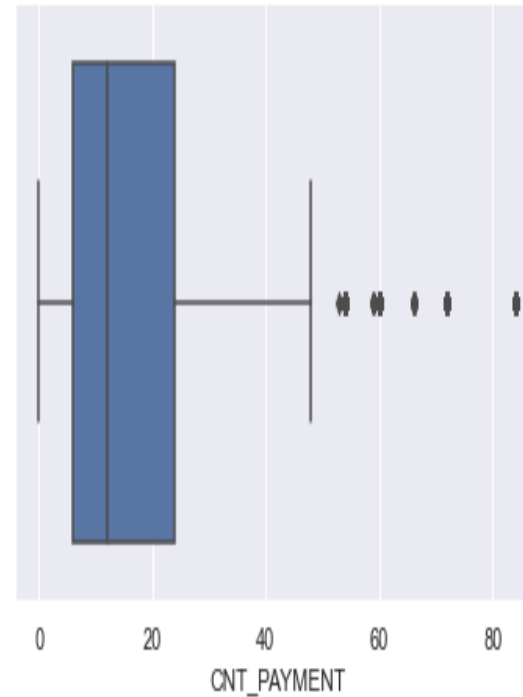
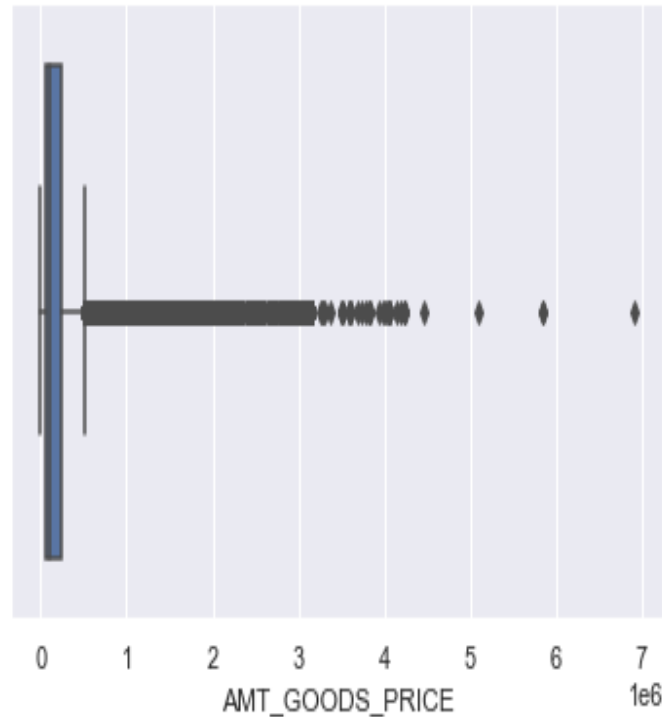
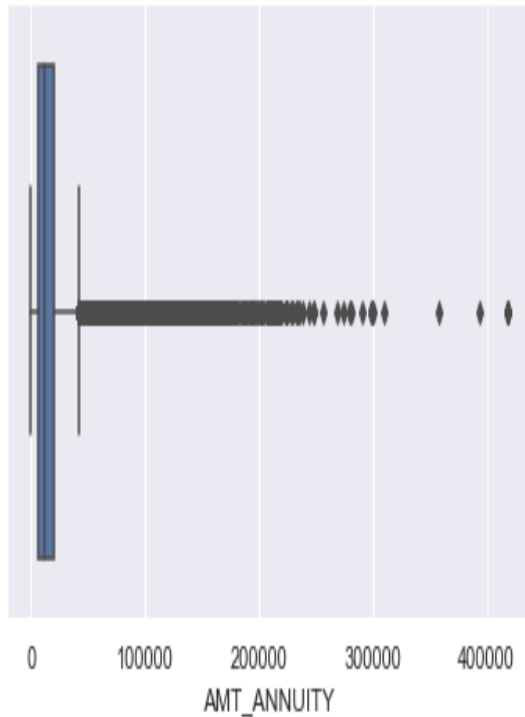
INFERENCES:-

AMT_INCOME_TOTAL, CNT_CHILDREN, AMT_CREDIT, AMT_GOODS_PRICE, AMT_REQ_BUREAU_QRT have some number of outliers.

AMT_CREDIT is greater than AMT_INCOME_TOTAL in all the cases.

AMT_INCOME_TOTAL has huge number of outliers which indicates that few of the loan application have high income compared to others.

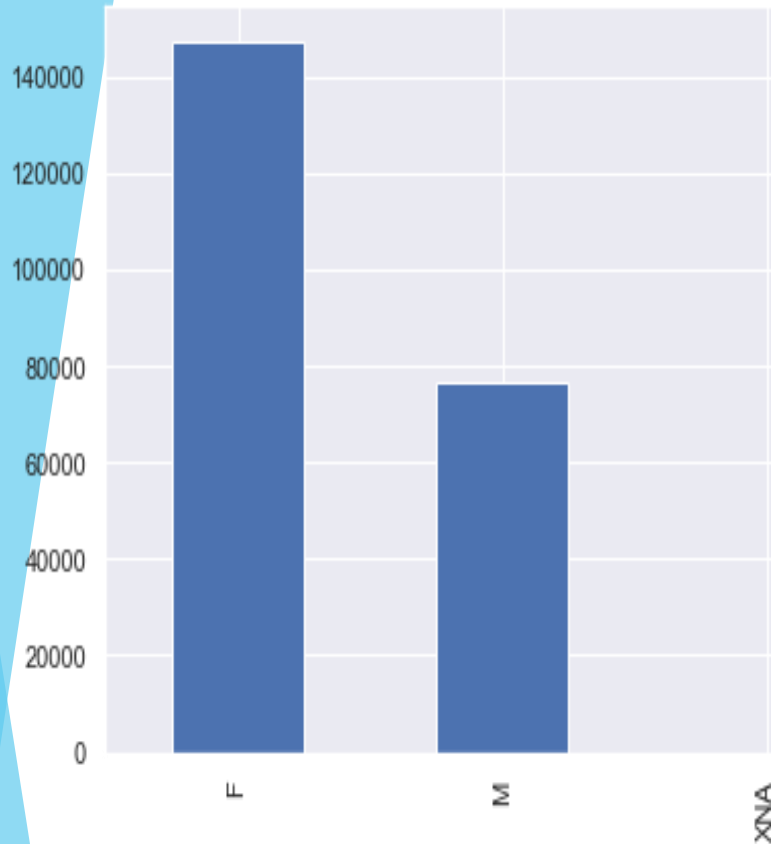
OUTLIERS ANALYSIS OF PREVIOUS APPLICATION DATASET



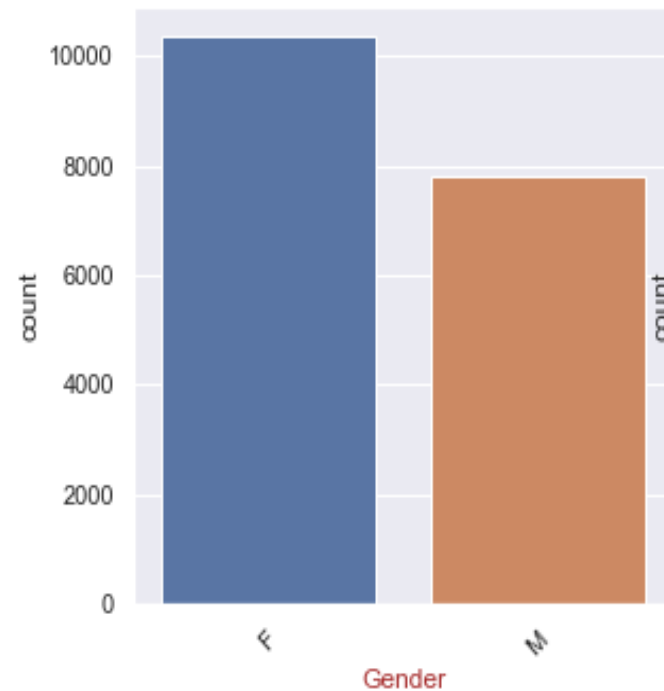
- AMT_ANNUIITY : We can see large no. of outliers .
- CNT_PAYMENT : Outliers are present in less no. in comparison to others.

GENDER CODE

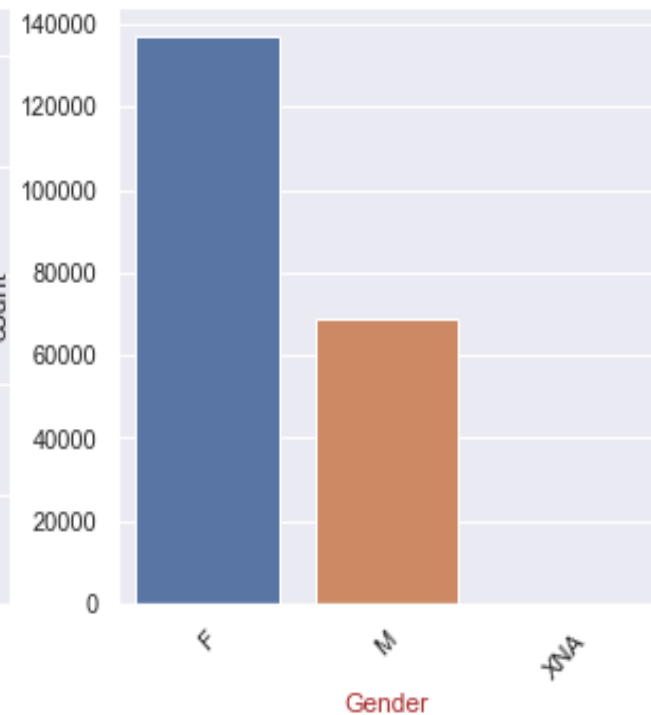
CODE_GENDER VARIABLE



Non-Defaulters Gender Wise



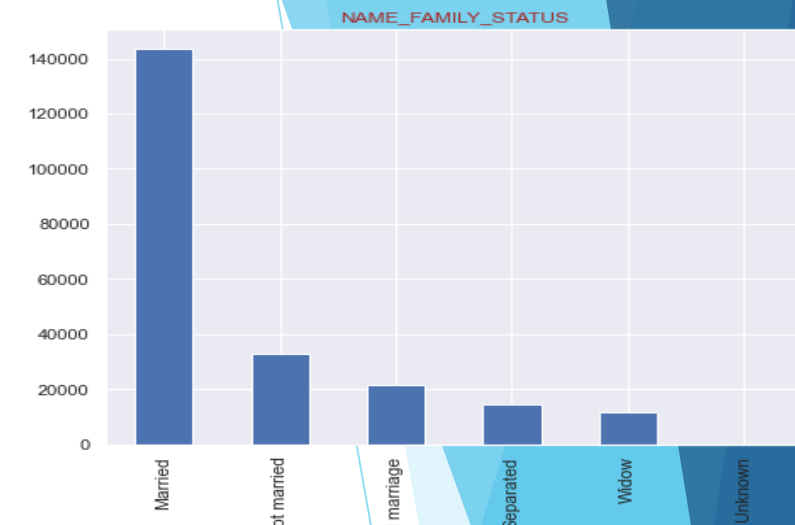
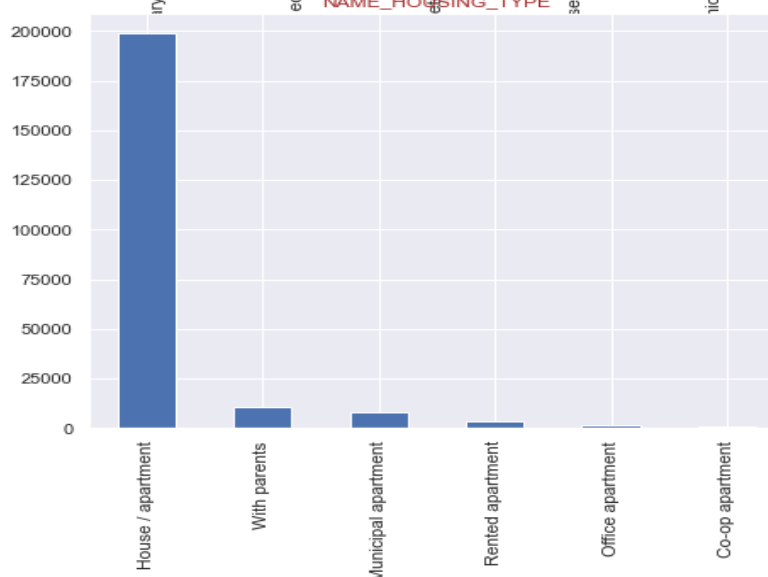
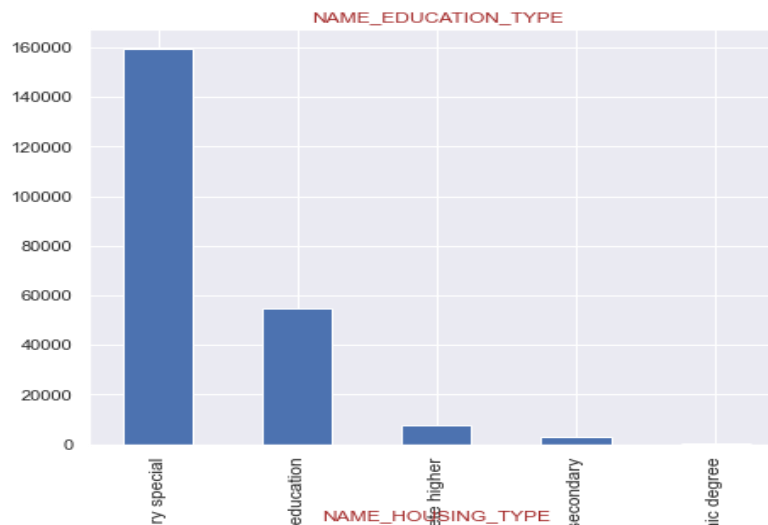
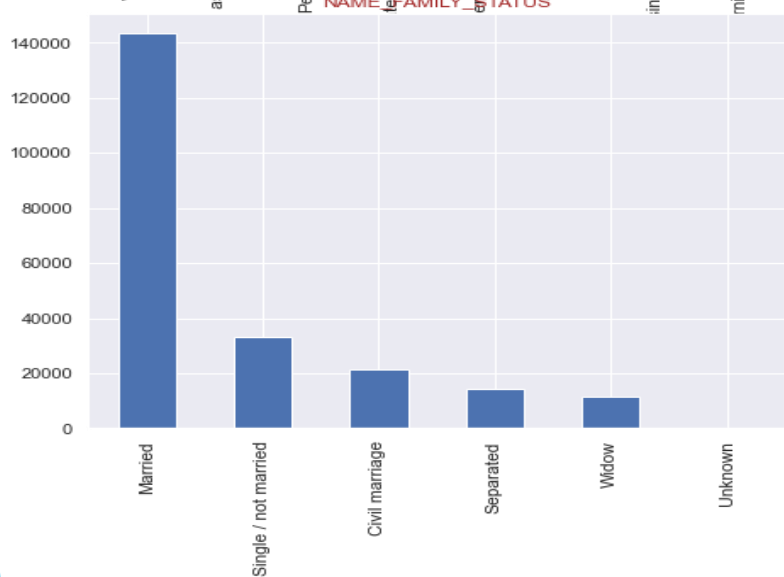
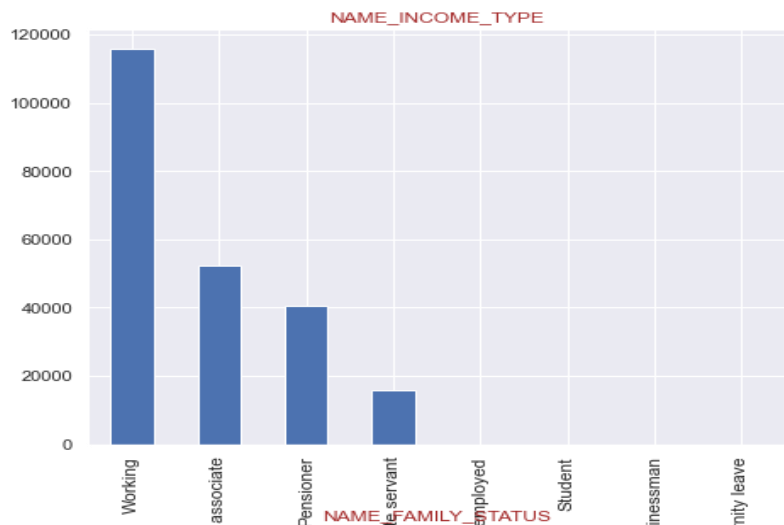
Defaulters Gender Wise



INFERENCES:

The number of females client is almost double than male client.

UNIVARIATE ANALYSIS



Inferences:

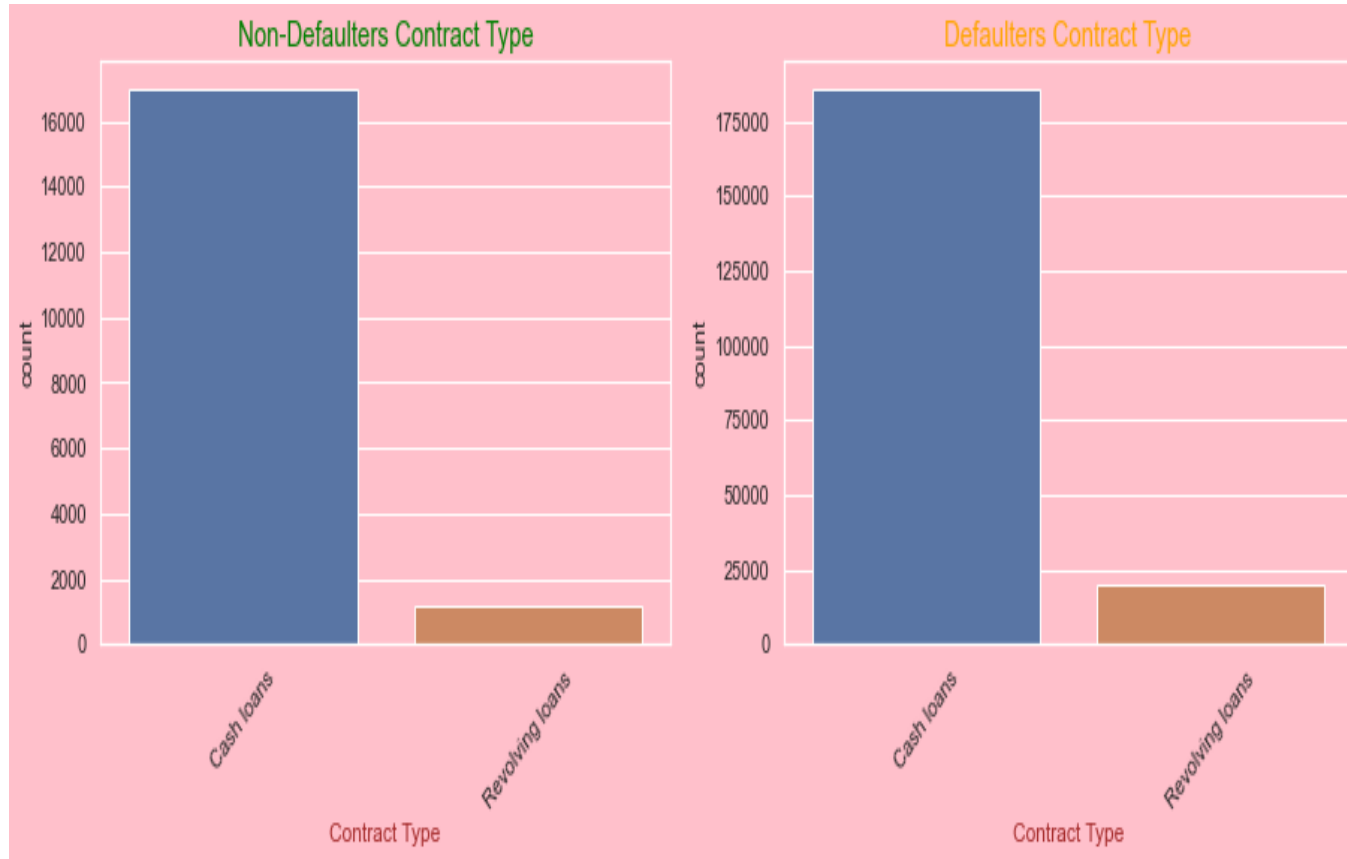
NAME_INCOME_TYPE: Working, Commercial associate and Pensioner having higher percentage of applicants.

#NAME_EDUCATION_TYPE: Most of the loans are provided either for Secondary or Higher education purposes.

#NAME_FAMILY_STATUS: Most of applicants are married

#NAME_HOUSING_TYPE: The applicants are staying mostly in their own house or else in apartments.

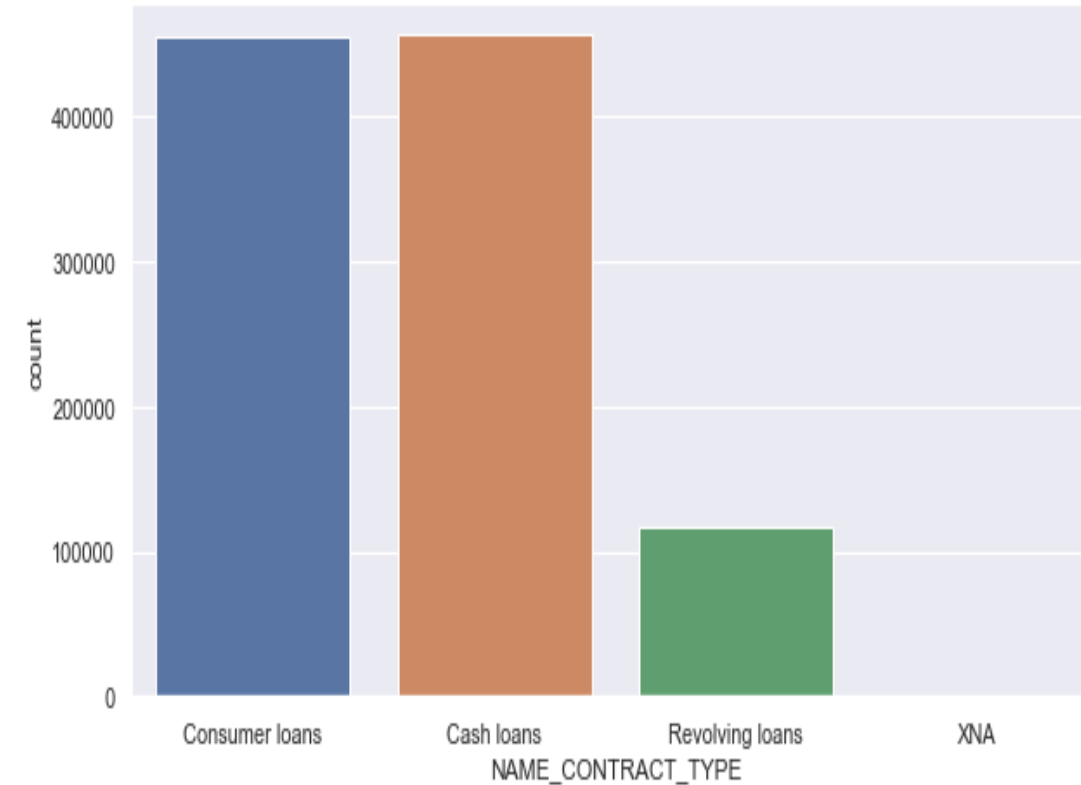
CONTRACT TYPE



Inferences:

Nearly both defaulter and non-defaulters opt for cash loans contract type.

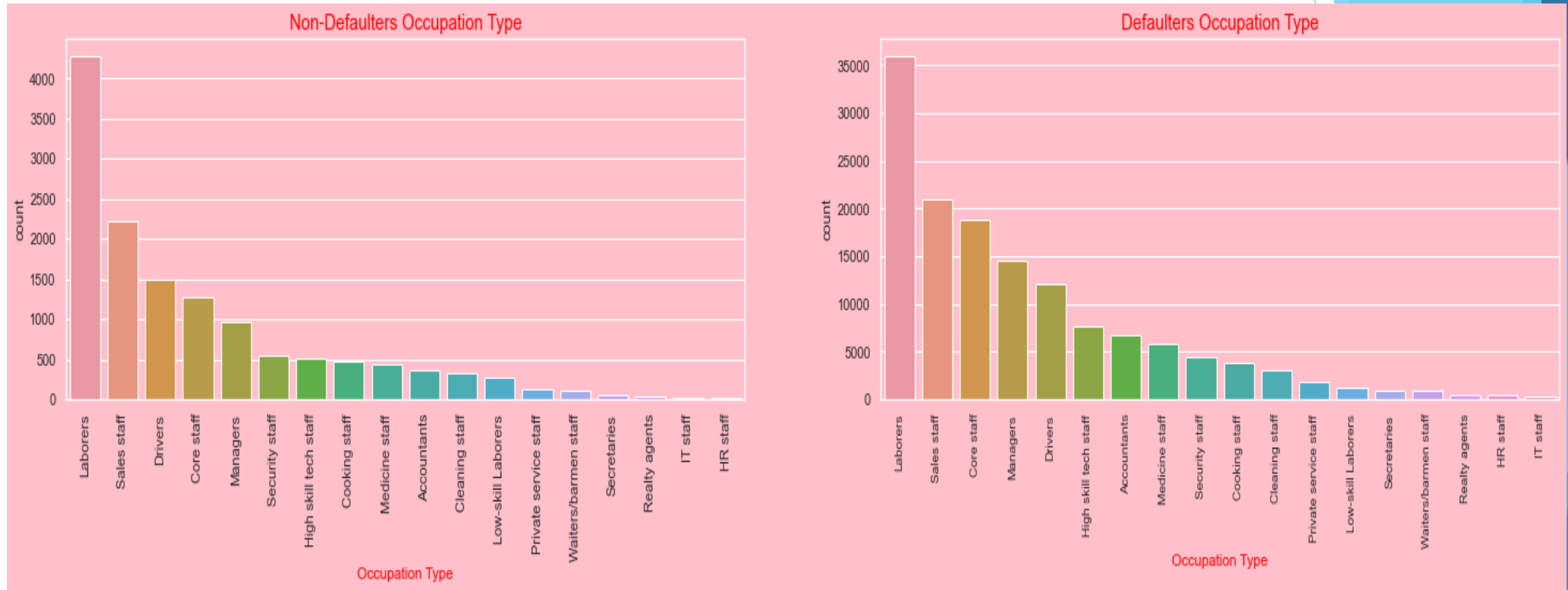
NAME_CONTRACT_TYPE



Inferences:

Cash loans and consumer loans are mostly preferred than other type of loans.

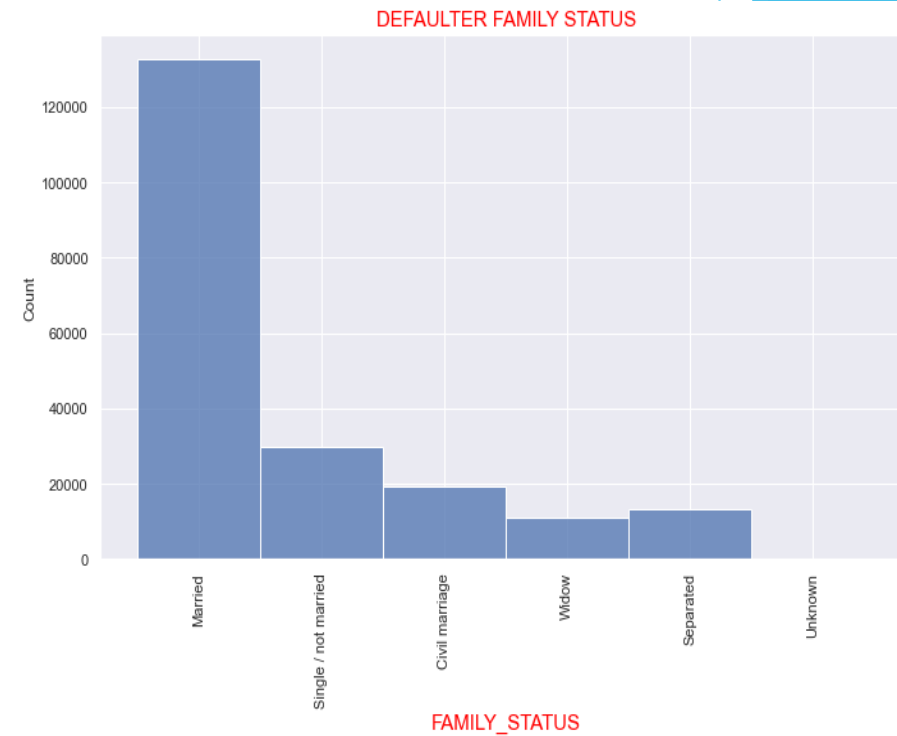
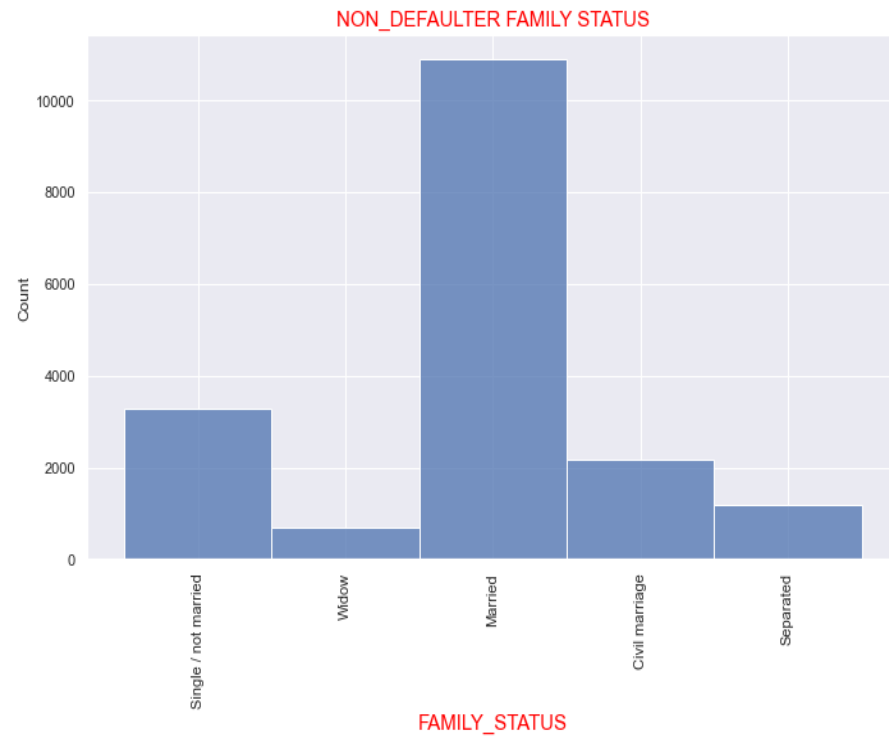
OCCUPATION TYPE



INFERENCES:

Most of the defaulters are laborers by occupation. The laborers are lower skilled people, who are struggling with the loan repayment and become defaulters.

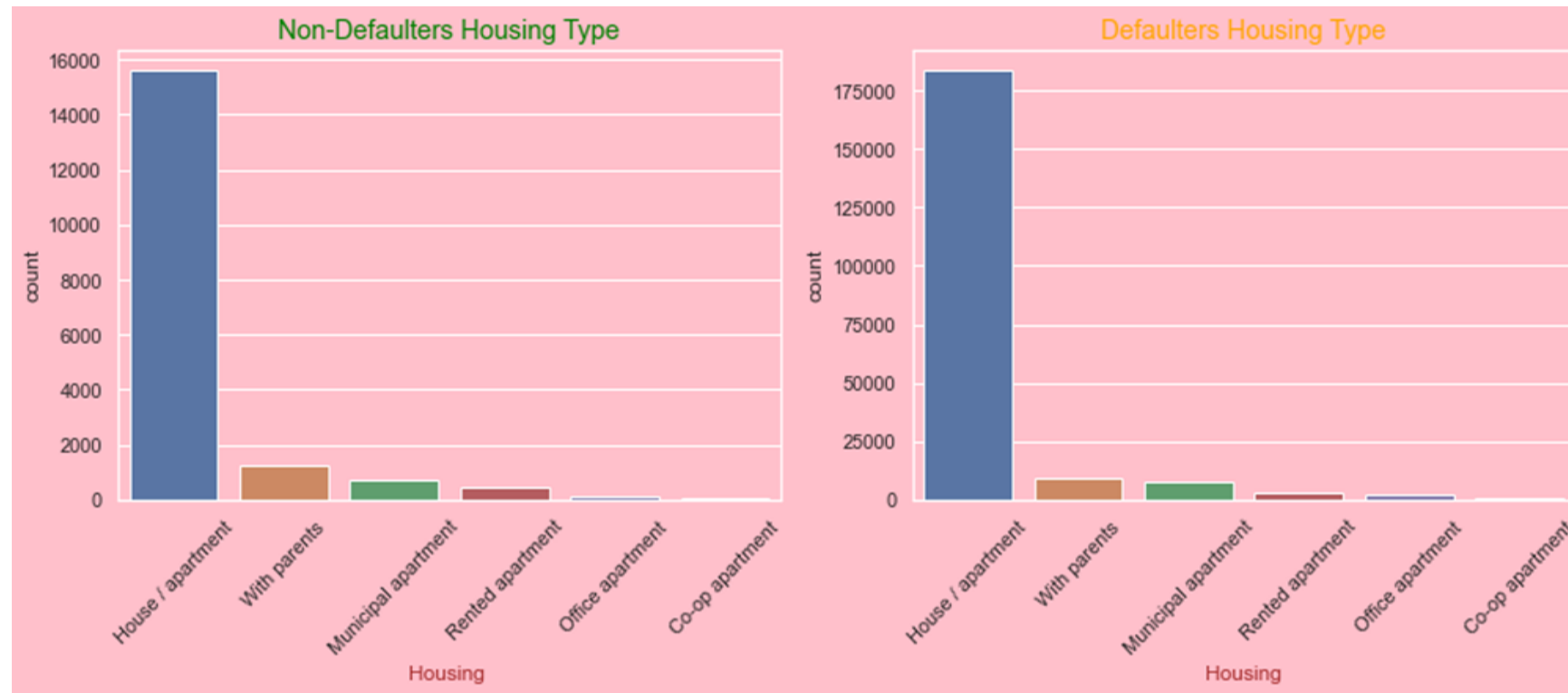
FAMILY STATUS



INFERENCES:

Married people are mostly defaulters compared to Single/Not Married people.

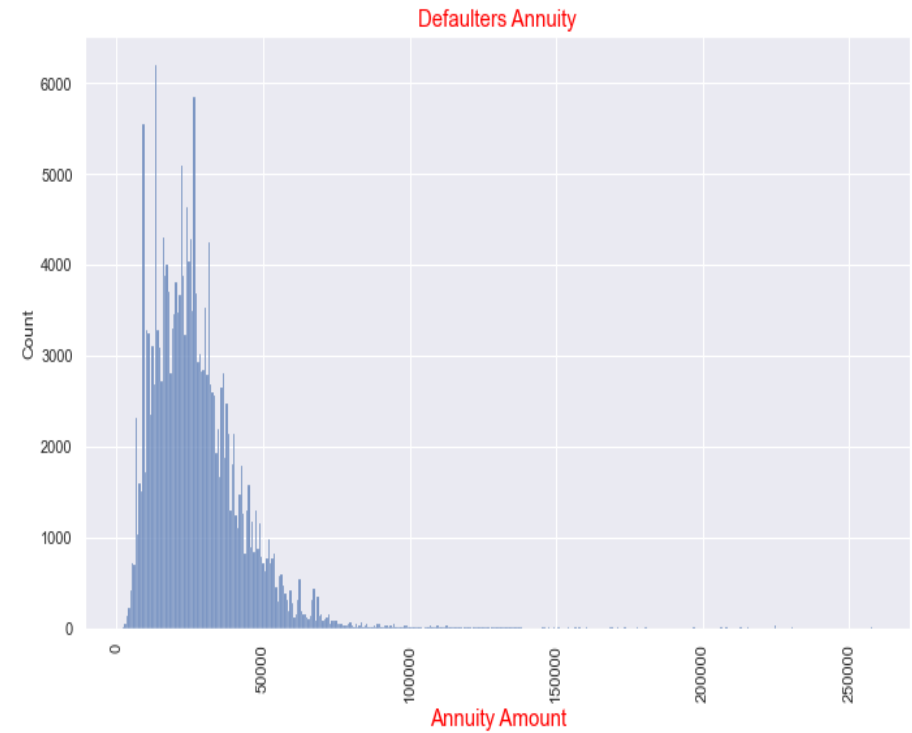
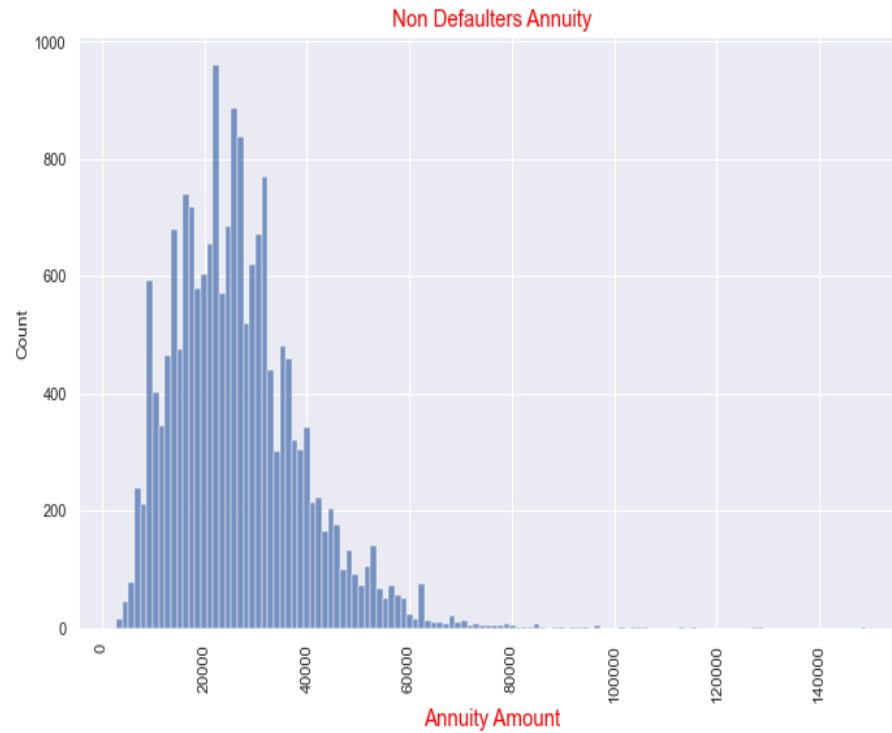
HOUSING TYPE



INFERENCES:

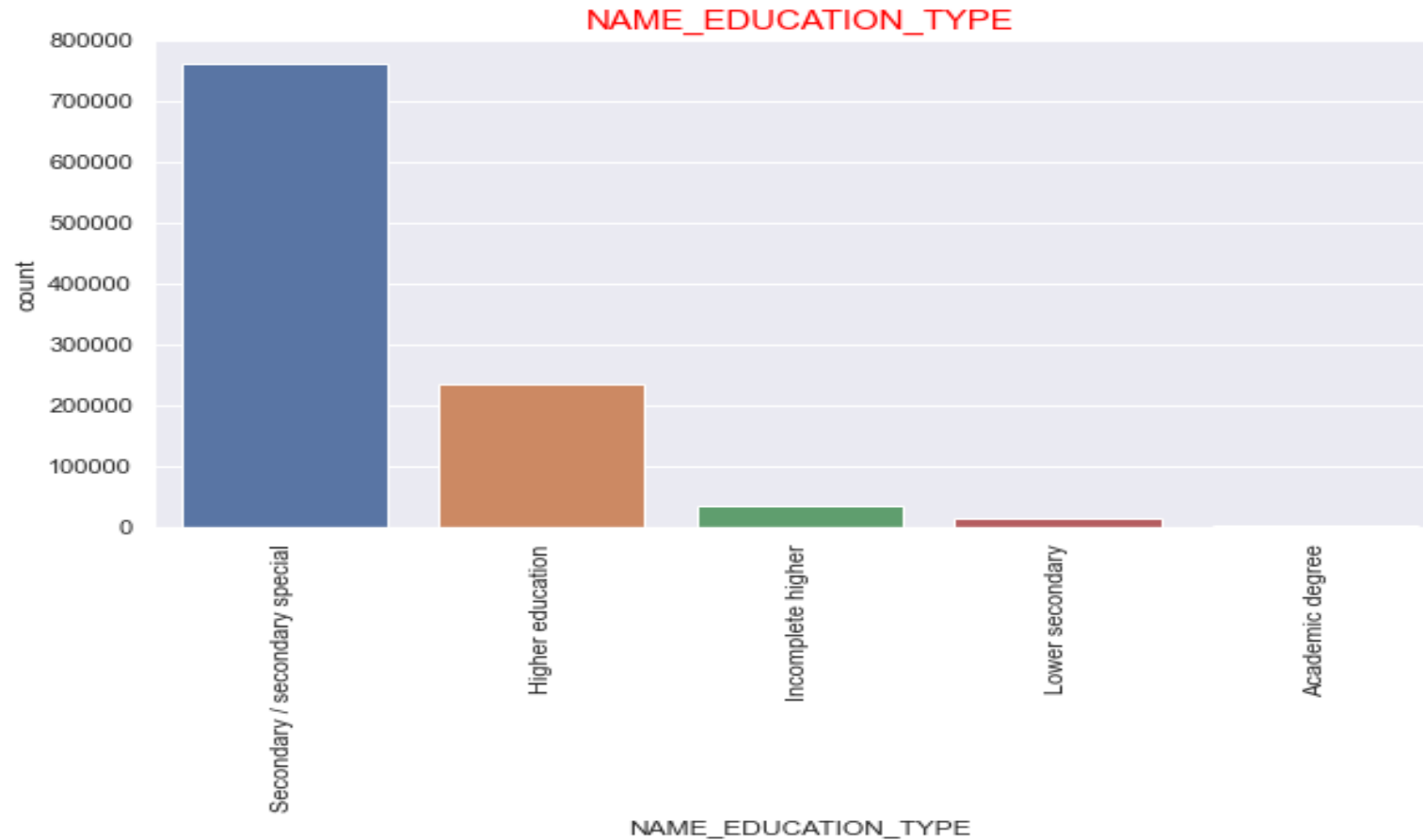
Most of the Defaulters are having their own house/apartment.

ANNUITY AMOUNT



INFERENCES:
Defaulters are having loan annuity between 15000 to 30000.

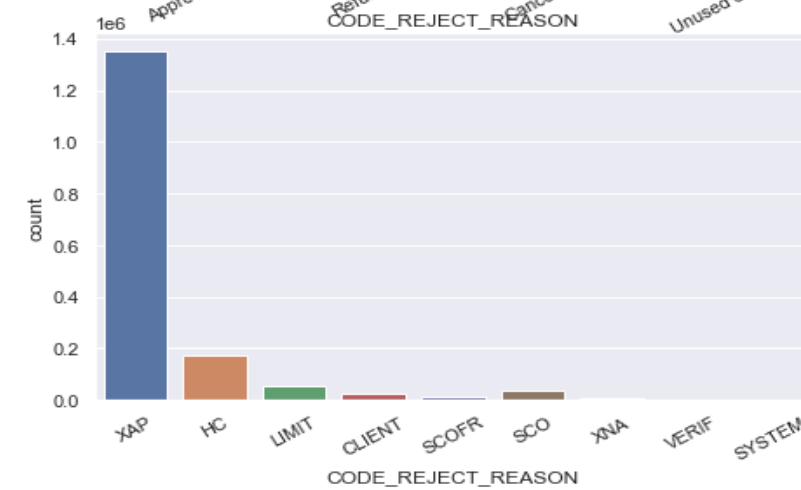
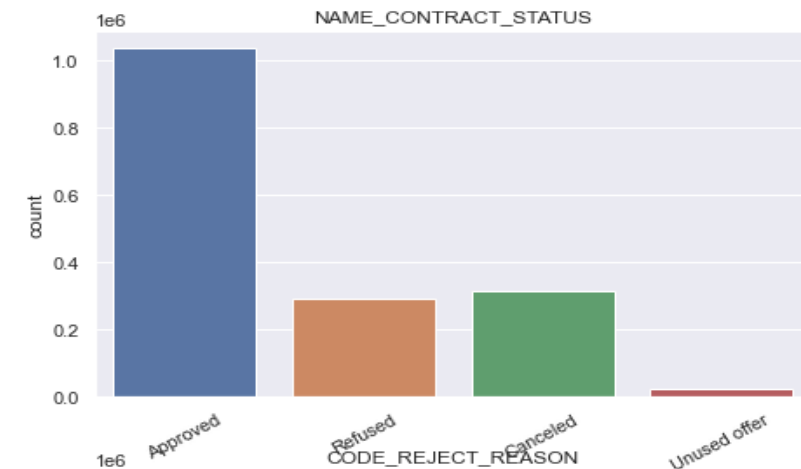
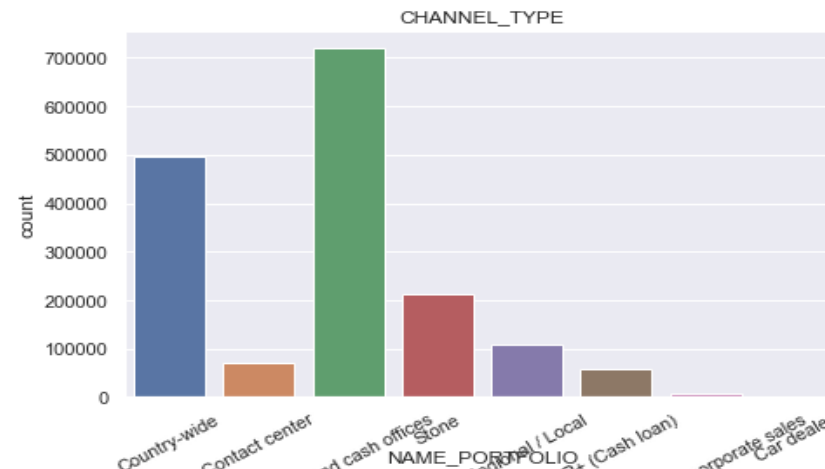
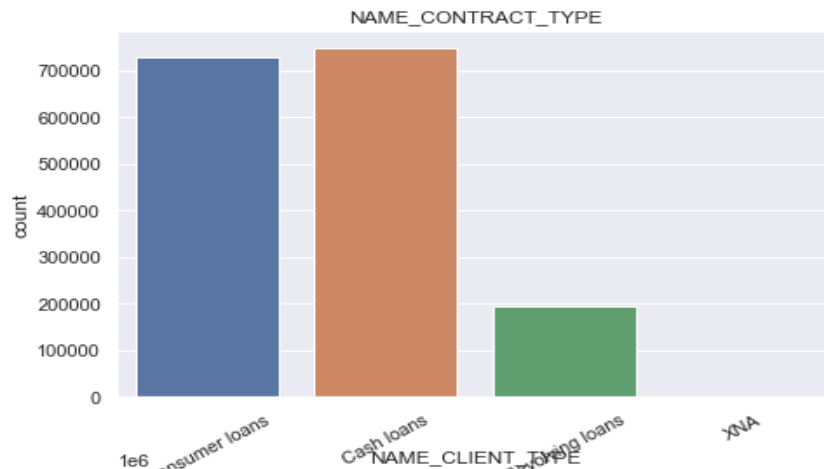
EDUCATION TYPE



INFERENCES:

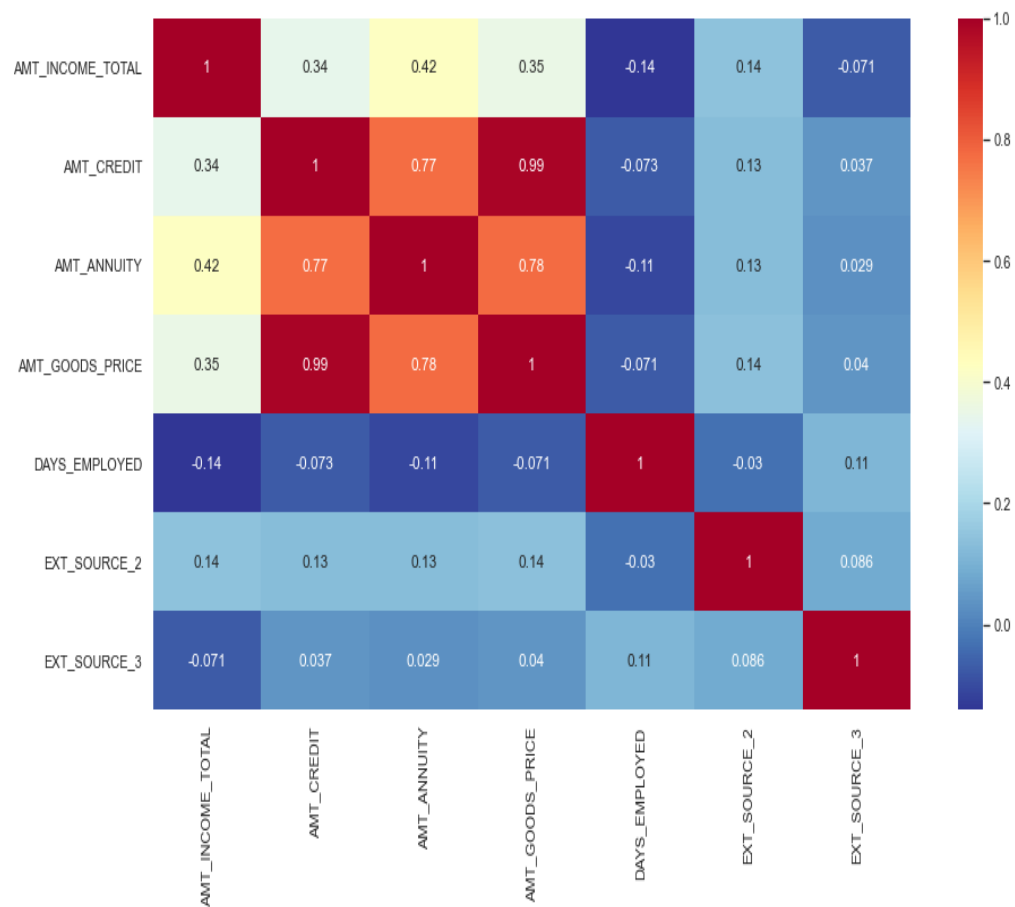
Majority of the clients have Secondary/Secondary special education , followed by clients with higher education .

BIVARIATE ANALYSIS



INFERENCES:

- NAME_CONTRACT_TYPE: More Consumer Loans and Less Revolving Loans
- CHANNEL_TYPE: Except Contact center, AP+ CashLoans and Car dealers all other are high in numbers.
- NAME_CONTRACT_STATUS: Mostly approved, very less Refused and very few loans are cancelled.
- NAME_CLIENT_TYPE: Most of the Clients are repeater client. NAME_PORTFOLIO: POS are in huge numbers compare to Cash and cards.



INFERENCES FROM DEFAULTER DATAFRAME:

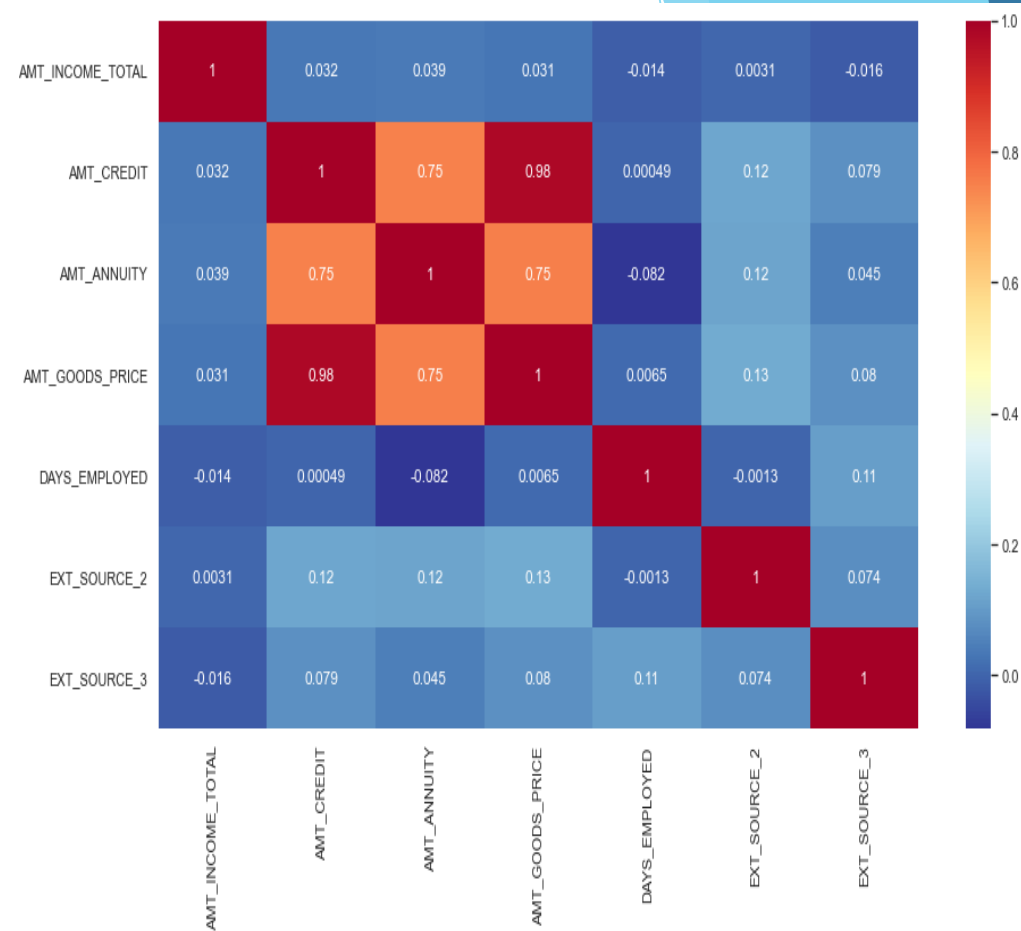
The more the correlated percentage means the high value for being a defaulter.

0.98% = AMT_CREDIT & AMT_GOODS_PRICE.

0.75% = AMT_CREDIT & AMT_ANNUITY.

0.75% = AMT_GOODS_PRICE & AMT_ANNUITY.

0.58% = Age & DAYS_EMPLOYED



INFERENCES FROM NON-DEFAULTER DATAFRAME:

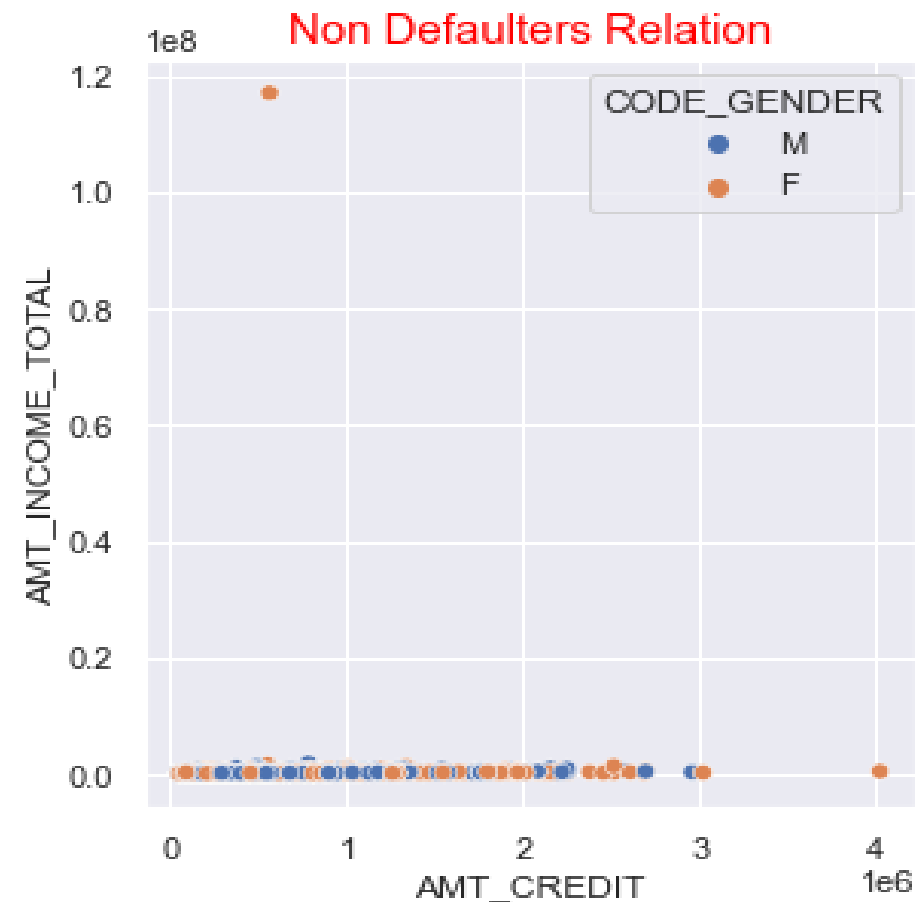
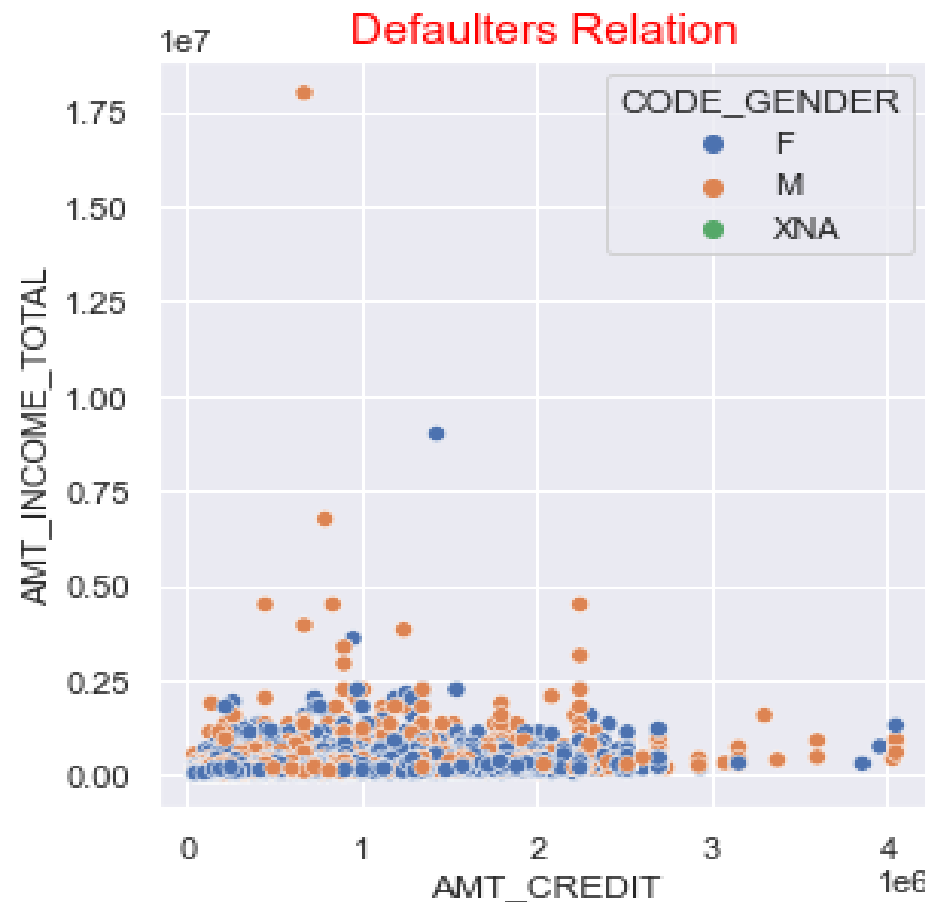
The more the correlated percentage means the high value for being a Non-defaulter.

0.99% = AMT_CREDIT & AMT_GOODS_PRICE.

0.77% = AMT_CREDIT & AMT_ANNUITY.

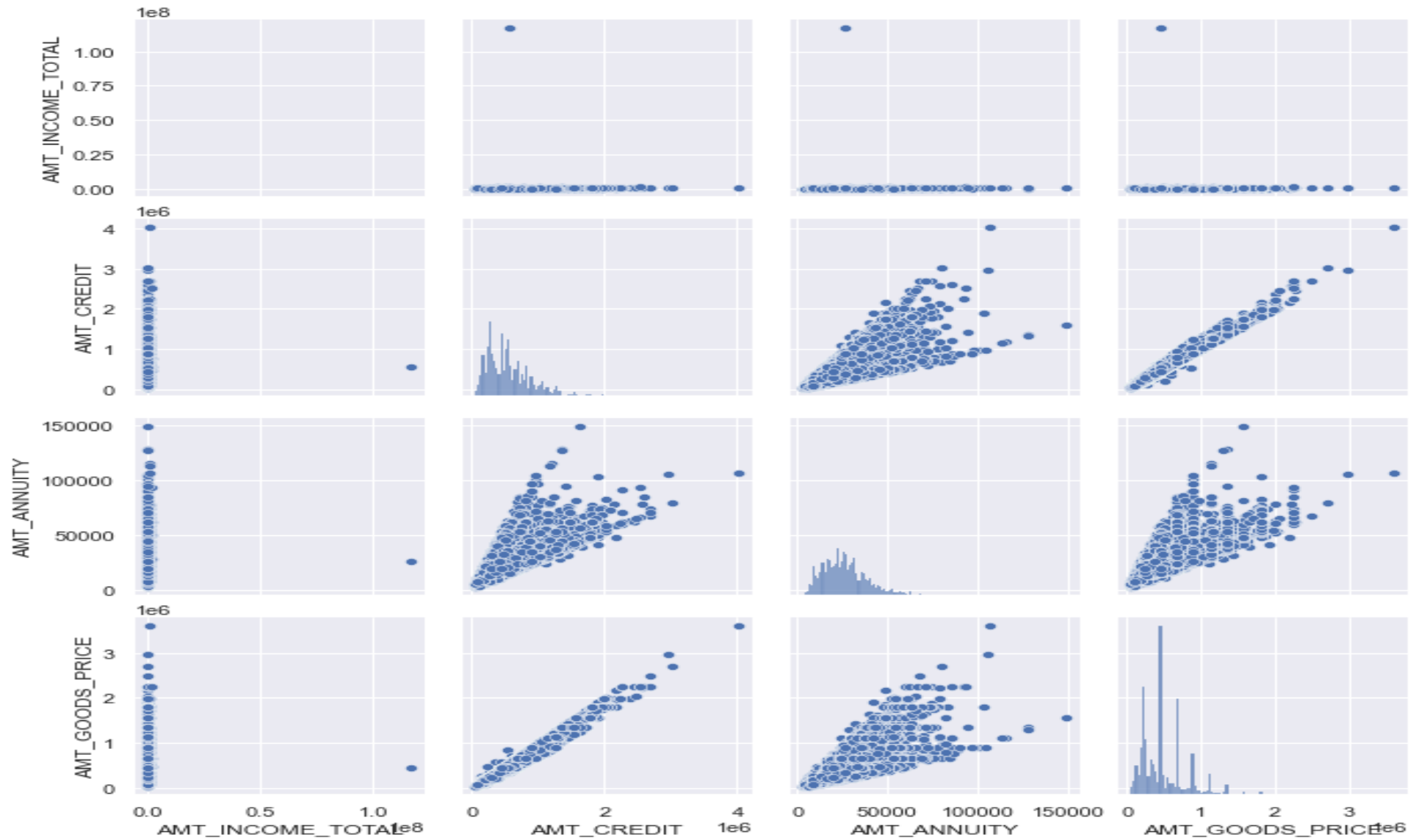
0.77% = AMT_GOODS_PRICE & AMT_ANNUITY.

0.63% = Age & DAYS_EMPLOYED



For Defaulters: There are more congested values in initial or lower areas of both AMT_CREDIT and AMT_INCOME_TOTAL.

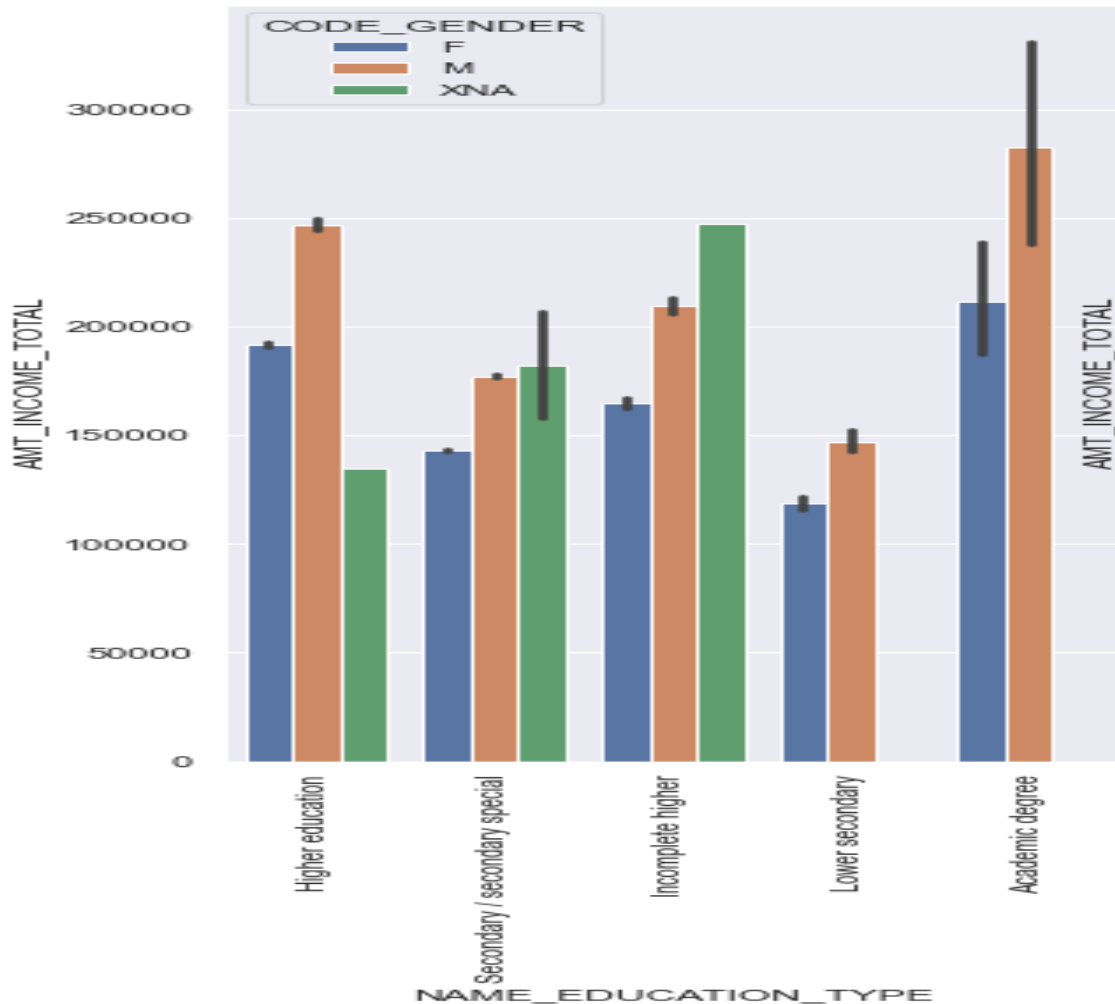
For Non-Defaulters: The values are evenly distributed between AMT_CREDIT and AMT_INCOME_TOTAL. There is no way to distinguish between Male and female. It looks like all are having same ratio in this plot, which makes it hard to figure it out here.



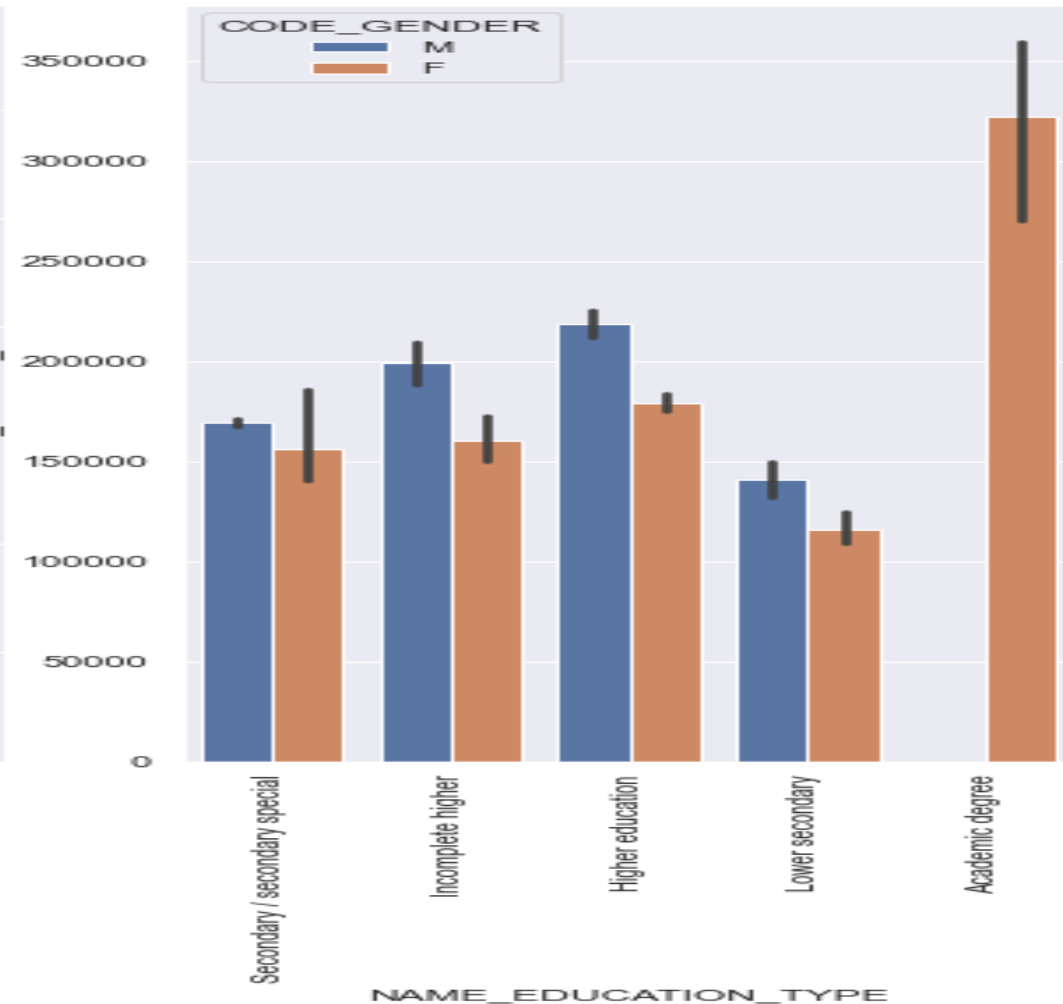
INFERENCES:

Very high correlation between `AMT_CREDIT` and `AMT_GOODS_PRICE`.

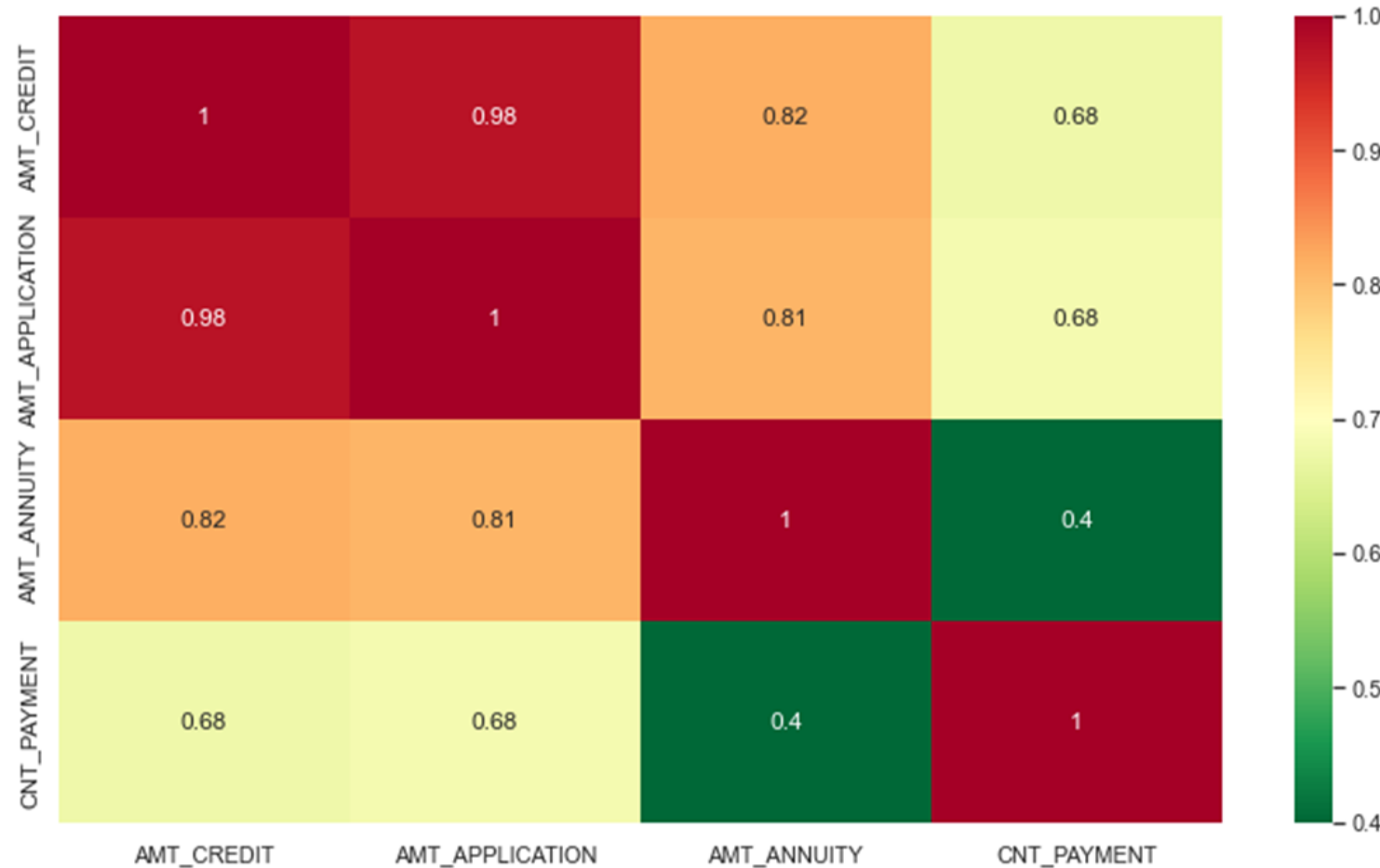
Defaulters Relation



Non Defaulters Relation

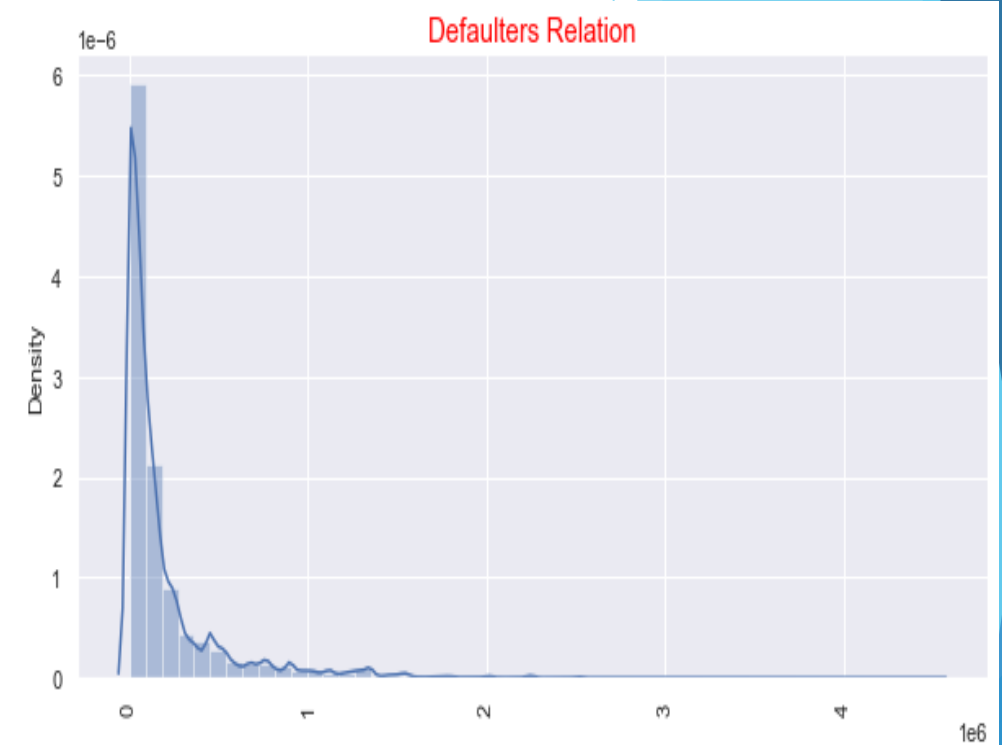
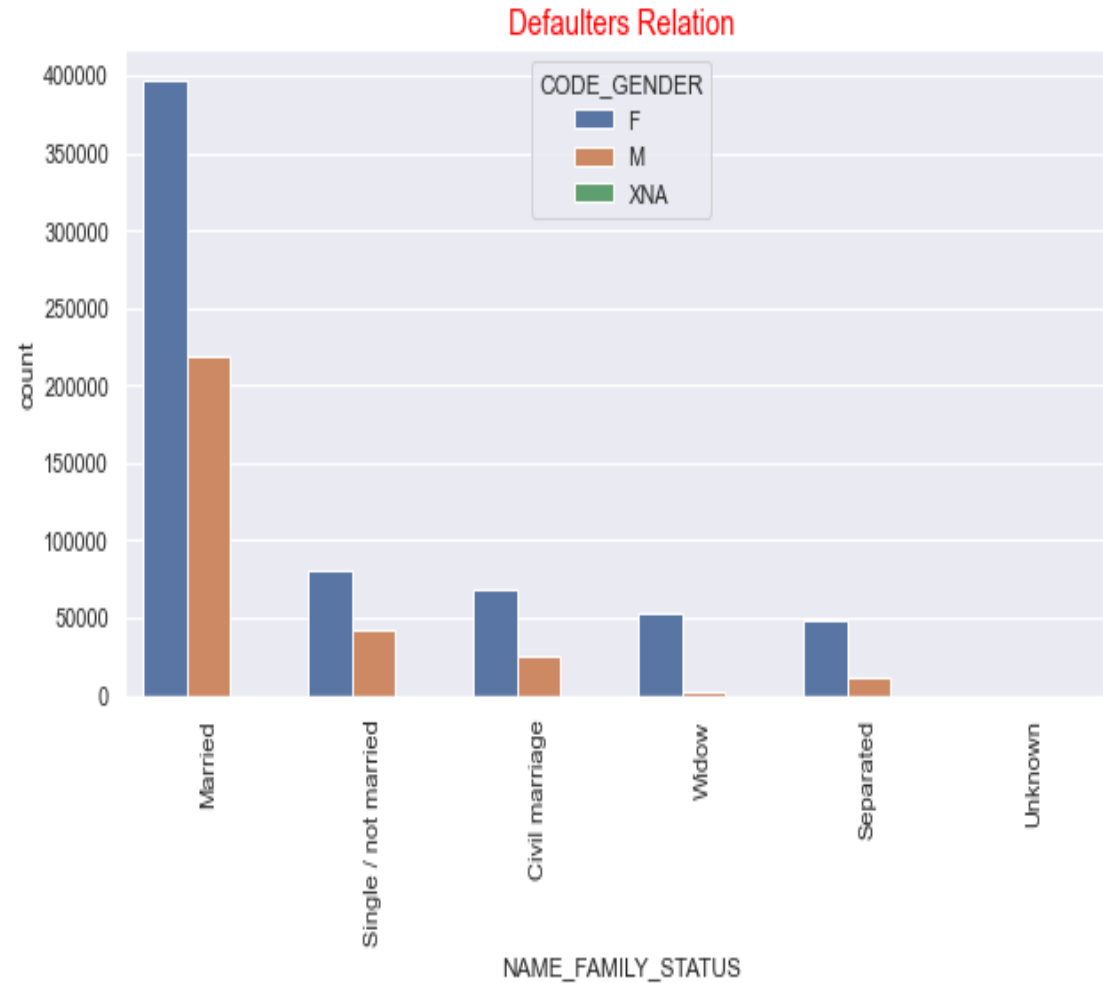


- **For Defaulters:** Female Academic Degree holders are earning more but still they are in default list where as not a single Academic male holders are in defaulter list. But for rest of the education type we can see that males are earning more as well as being the more defaulters.
- **For Non Defaulters:** Here also Female Academic degree holders are more and they are in high no. when it comes to non defaulters. Nearly same amount non defaulter are there for guys who studied Higher education. Applied to both male and female. Female in all educational aspects are more non-defaulters than male.



INFERENCES:

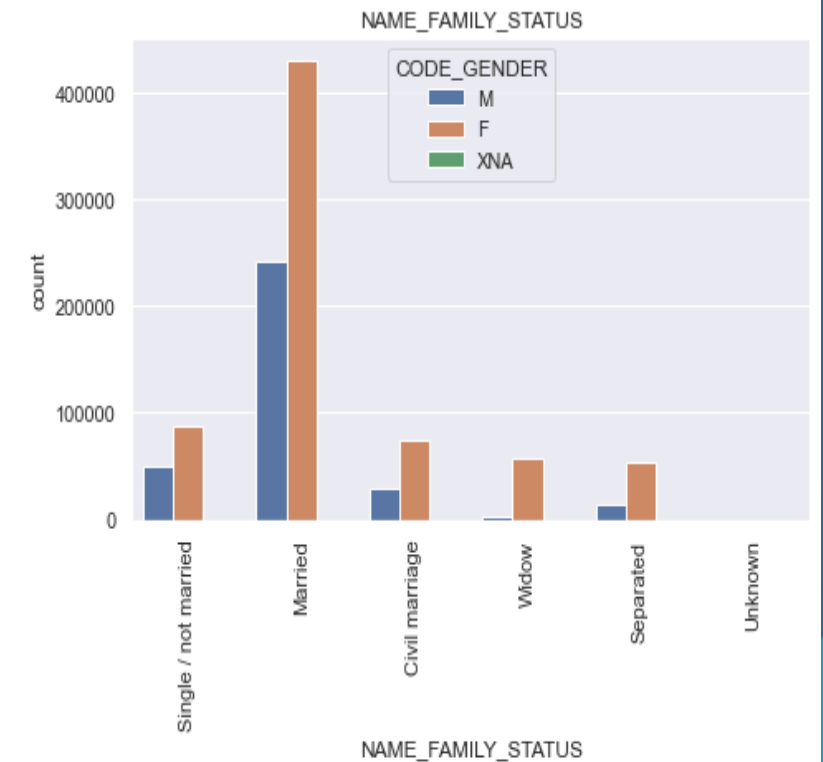
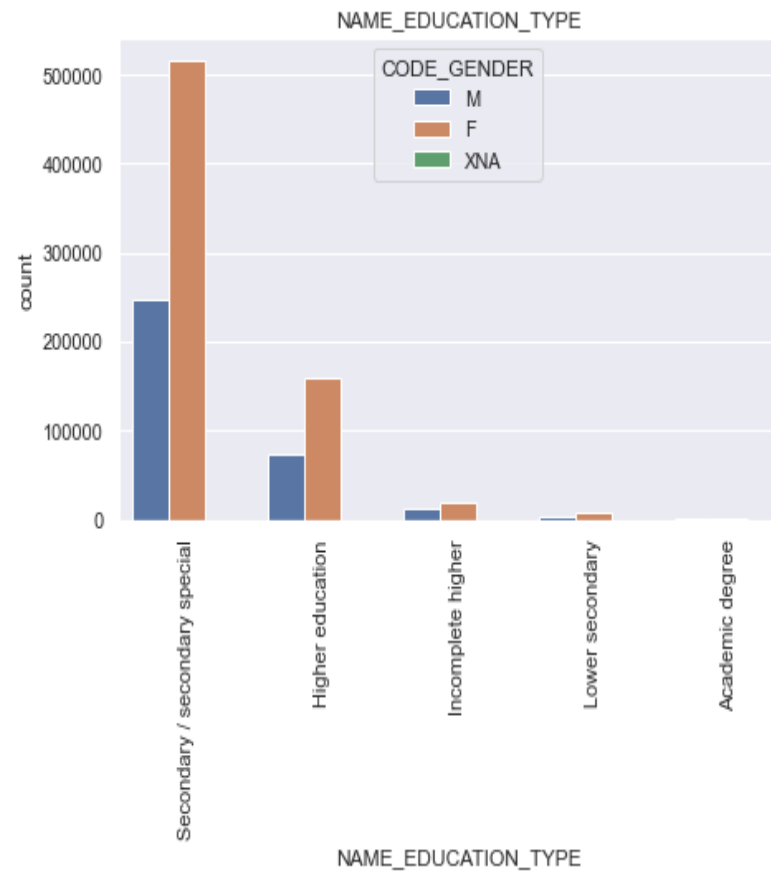
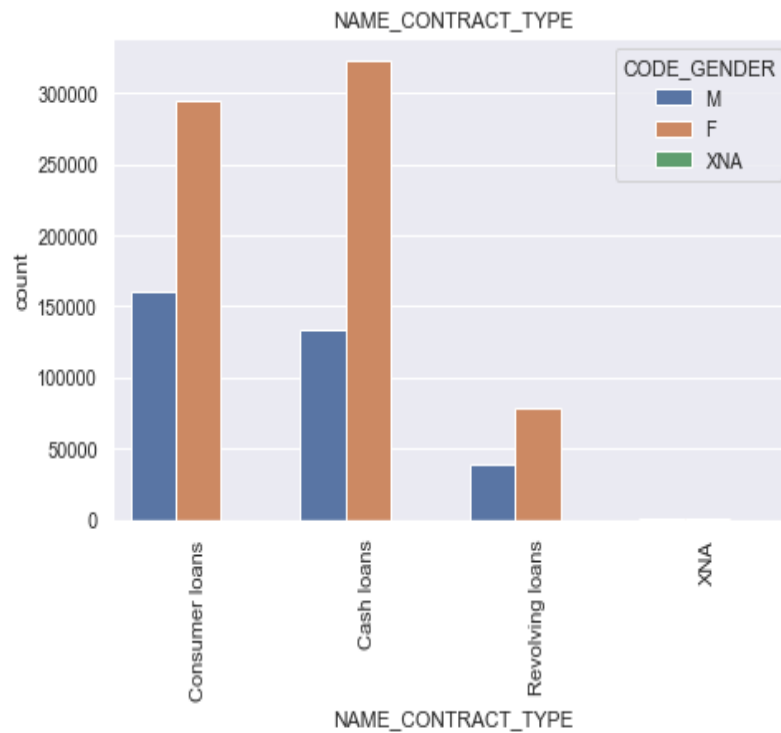
- AMT_CREDIT & AMT_APPLICATION : Highly correlated, nearly 99%.
- AMT_CREDIT & AMT_ANNUIITY : Highly Correlated, nearly 86%.
- AMT_ANNUIITY & AMT_APPLICATION : Highly correlated, nearly 86%



INFERENCES:

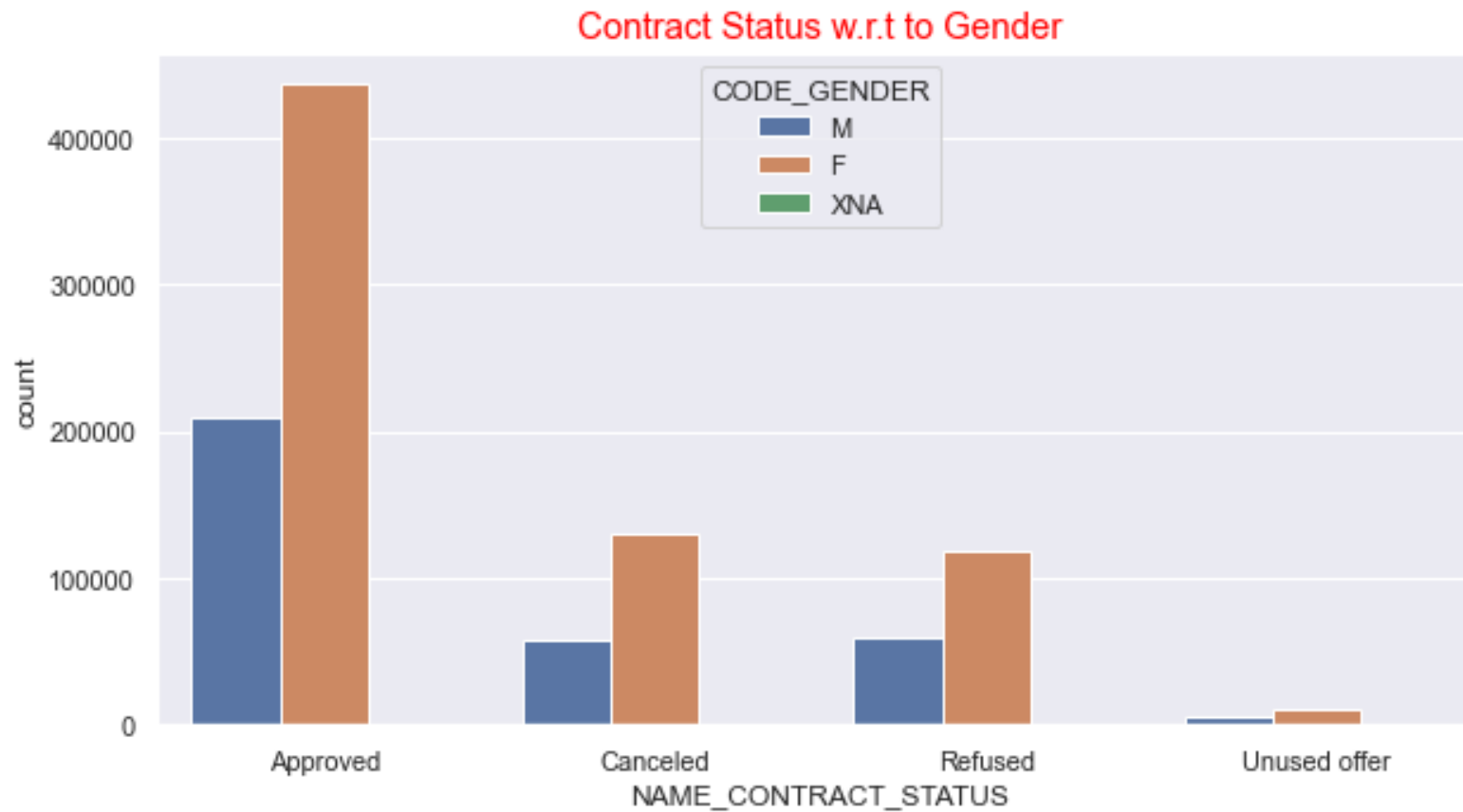
- Married people are struggling to pay the loan bill in comparison to single and separated people.
- More no of approved people are married.
- From next time onwards, we need to give loan to those who are either single or divorced.

MERGING OF APPLICATION DATA WITH PREVIOUS DATA



INFERENCES:

- Consumer loans are more taken by Female than Male.
- Secondary/Secondary Special Education type of female are more than male.
- More Married female member are taken loans compare to male.



INFERENCES:

- Female contracts got approved more compared to male.
- Female contracts also got refused more compare to male but that is nearly negligible and Less in numbers compare to approval.

CONCLUSION

- ❑ Female with consumer loans type are in more no. of defaulters.
- ❑ Married people are struggling to pay the loan bill compared to single and separated people.
- ❑ From next time onwards we need to give loan to those who are either single or divorced.
- ❑ Secondary/ secondary special people with education are struggling to pay their loan bill and becomes defaulters.
- ❑ Lower credit amount people are having high chance to becoming a defaulters than higher credit amount people.
- ❑ Female Gender are more likely to not face payment difficulties than the male and hence it is recommended to approve more loans of Female Gender than the male gender, at the same time Females are High in number than males who face difficulties in repayment.
- ❑ The Repeater applicant has High chance of non-Defaulting and also has high chance of defaulting when compared to new applicants.

A top-down view of a wooden desk with various stationery items. In the center is a white rectangular card with the words "THANK YOU" printed in a large, black, serif font. Surrounding the card are several items: a blue and black pen in the top left, a silver pen in the top center, a blue ruler in the top right, a black pen in the middle right, a blue eraser in the bottom left, and a silver pen in the bottom left.

**THANK
YOU**