CREDIT EDA ASSIGNMENT

Onkar Suryawanshi

INTRODUCTION

Credit risk is the possibility of a loss resulting from a borrower's failure to repay a loan or meet contractual obligations. Traditionally, it refers to the risk that a lender may not receive the owed principal and interest, which results in an interruption of cash flows and increased costs for collection. So to help any company to take decision on a loan for a customer by checking his previous records as well as background check with the help of the data.

To load the data using python libraries like numpy & pandas

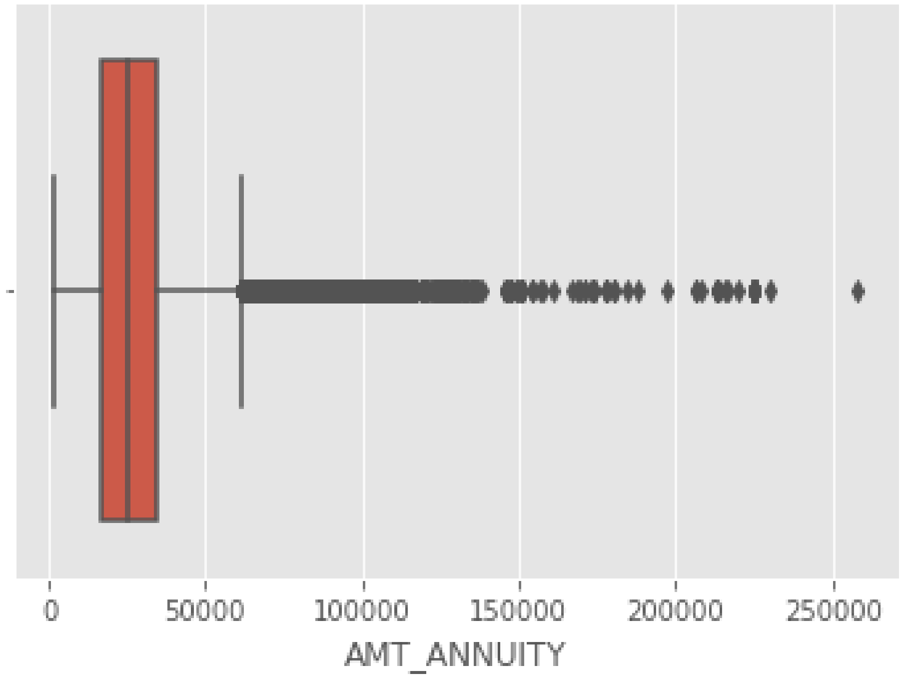
Understanding the data by & doing some analysis. Check the data if there are any missing values. Clean the data by imputing or deleting the unnecessary data

Check the data is imbalance or not. Use univariate & bivariate analysis.

Analysis the data after merging both the files

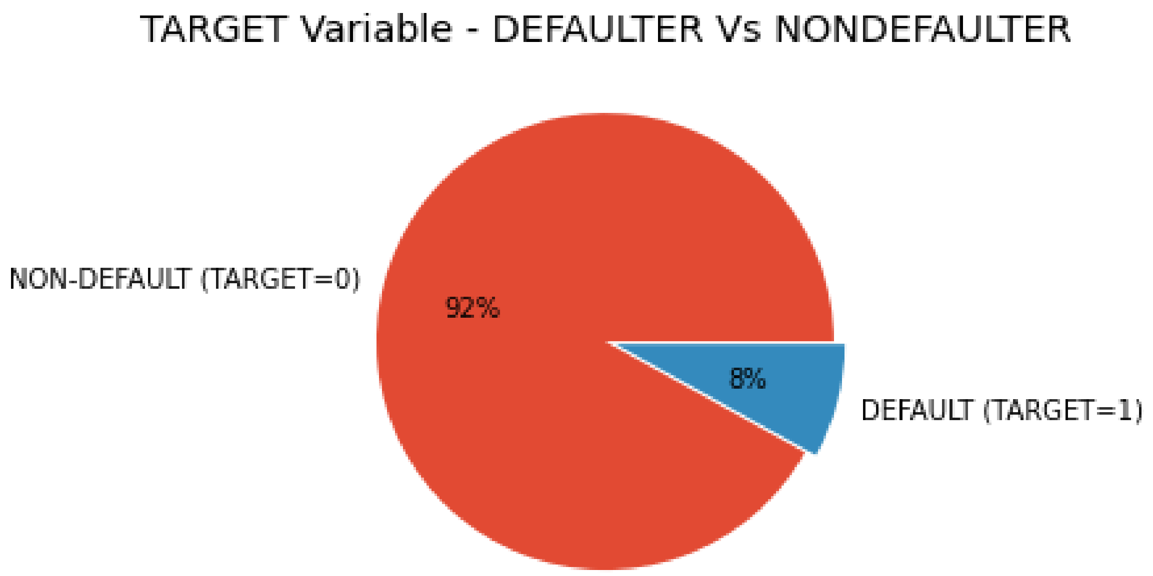
PROCESS FOR EDA

OCCUPATION TYPE IMPUTATION



‘AMT_ANNUITY’

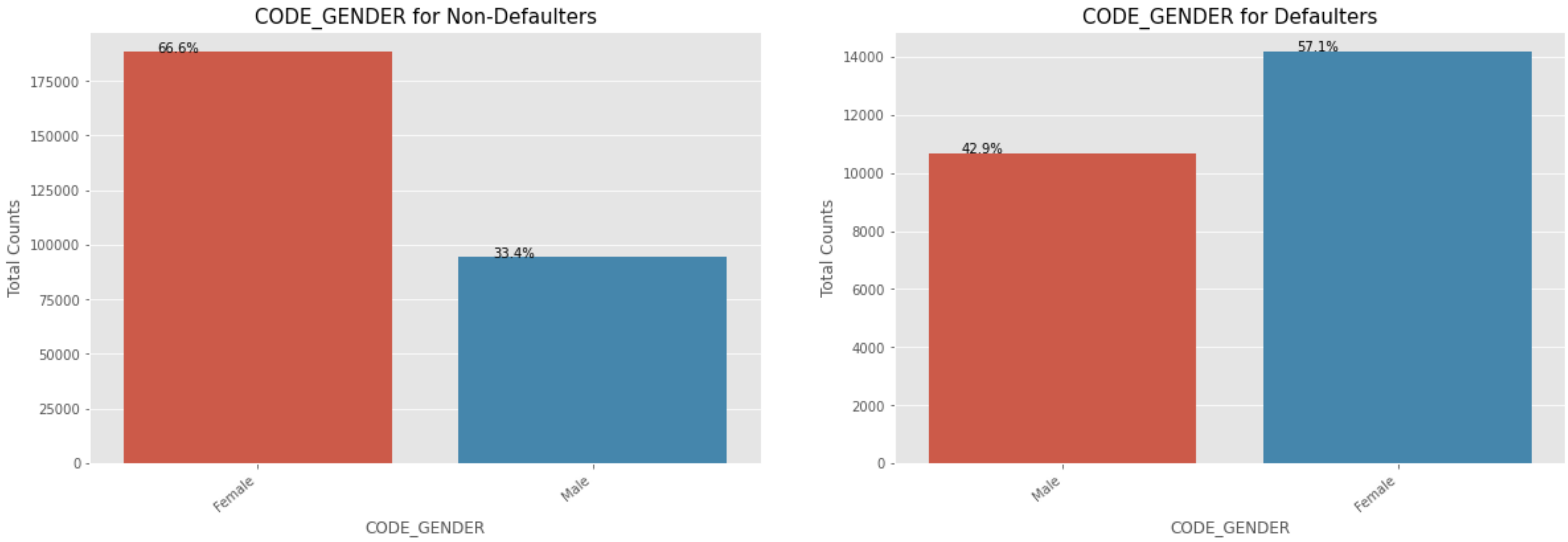
One of the column we can see there are outliers in this column. The solution for outliers is we can use median of this column to overcome the outliers



CHECKING IMBALANCE IN 'TARGET' COLUMN

'TARGET'

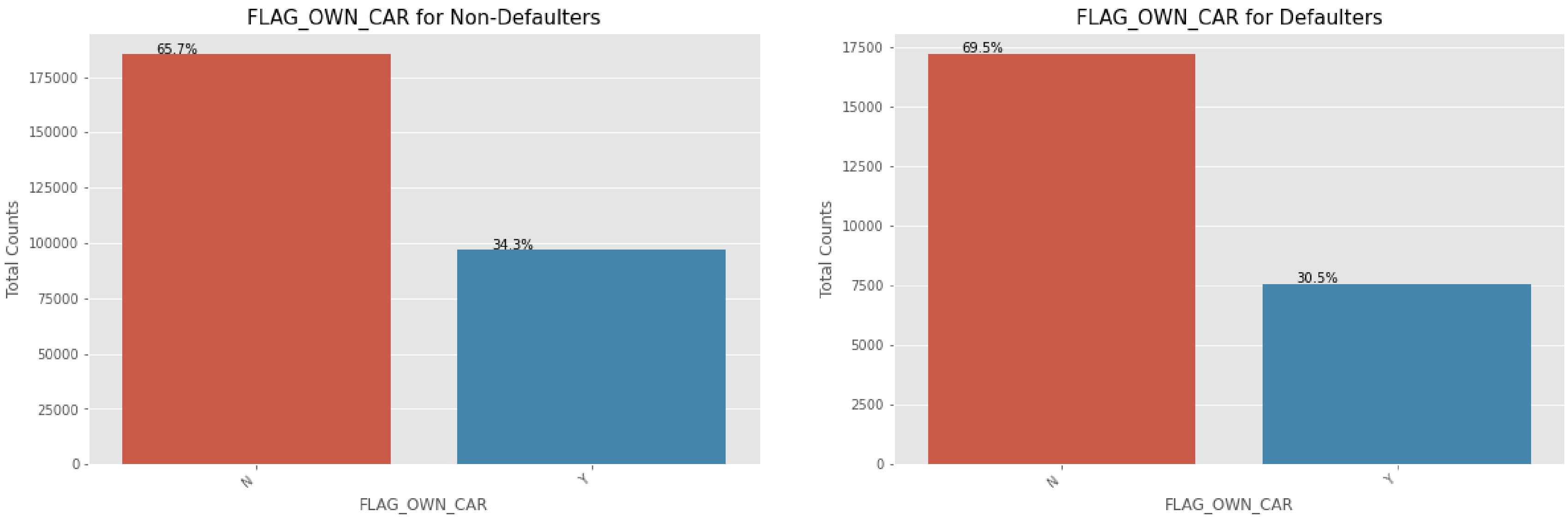
Above plot can tell that there is an imbalance between who has defaulted & who didn't. There are 92% of people didn't default & 8% who are defaulted



UNIVARIATE CATEGORICAL ORDERED ANALYSIS

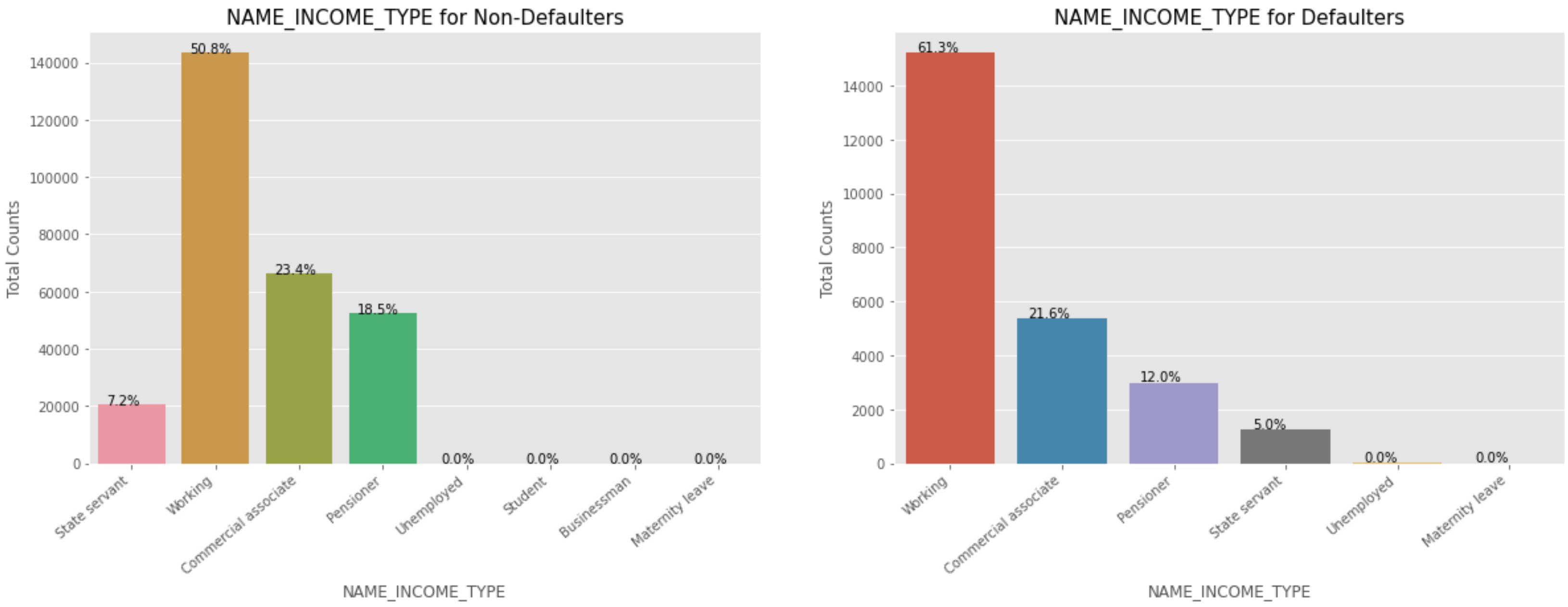
'CODE_GENDER'

Above we can see that Female contribute 66.6% to the non-defaulters & Male 33.4%. And also Female contributes 57.1% to the defaulters & Male 42.9%. Here we can conclude that more female applying for loans than males and hence the more number of female defaulters as well. But the rate of defaulting of FEMALE is much lower compared to their MALE.



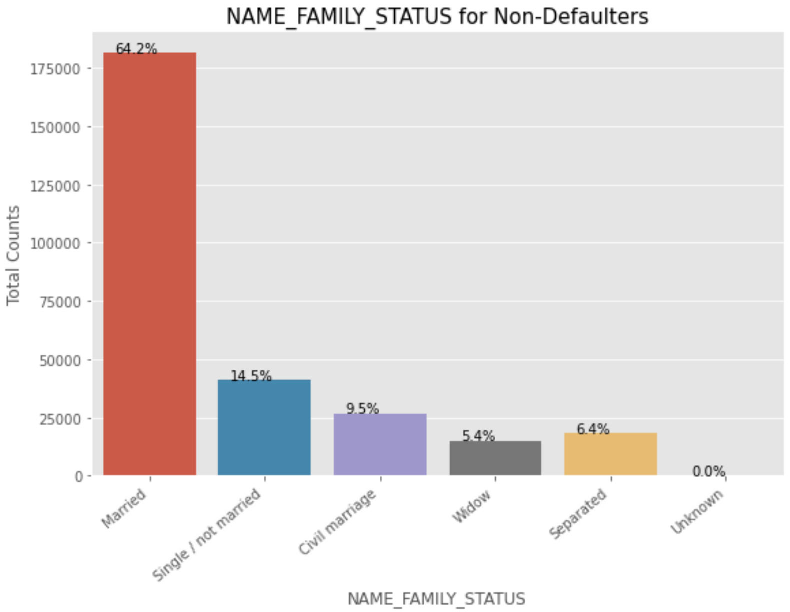
‘FLAG_OWN_CAR’

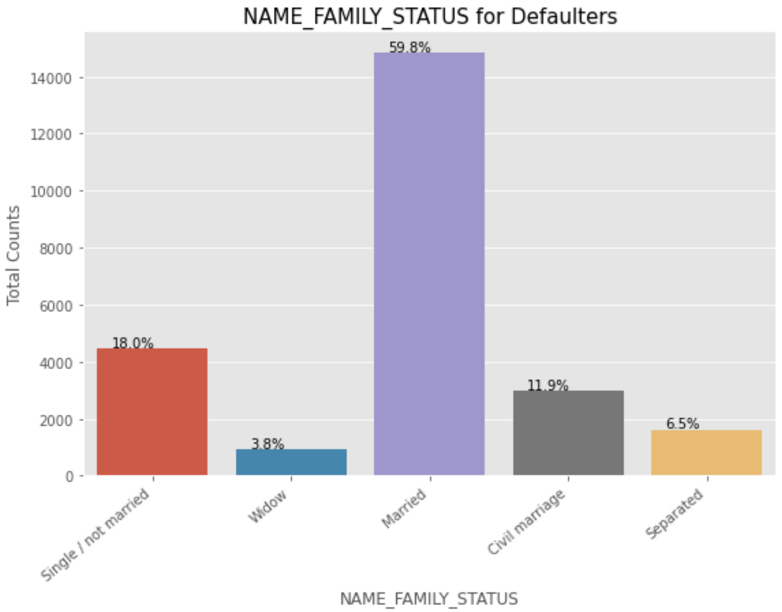
Above we can see that people with cars contribute 65.7% to the non-defaulters while 69.5% to the defaulters. Also we can see that people without cars contribute 34.3% to the non-defaulters while 30.5% to the defaulters. Looking at the percentages in both the charts we can conclude that the rate of default of people having car is low compared to people who are without cars



‘NAME_INCOME_TYPE’

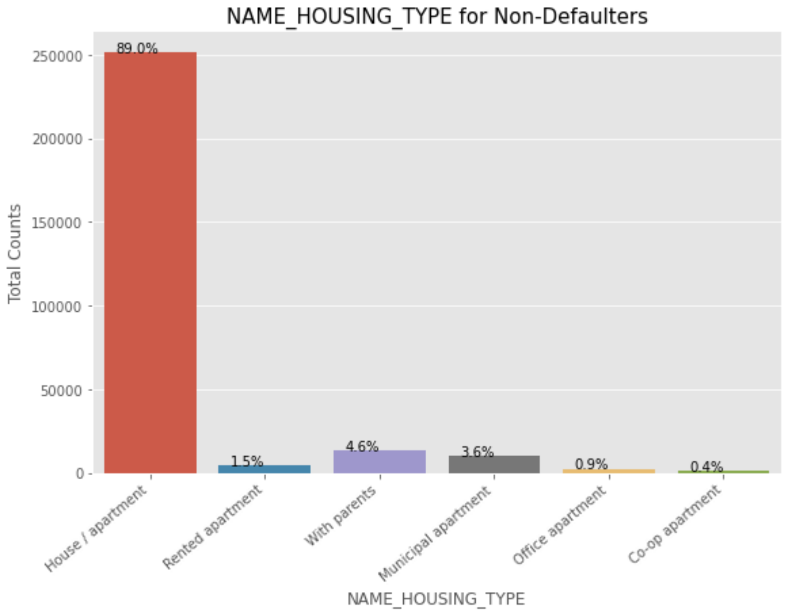
Above we can see that the students not default. The reason could be they are not required to pay during the time they are students. We can also see that the Businessman not default. Most of the loans are distributed to working class people. Also see that working class people contribute 51% to non defaulters while they contribute to 61% of the defaulters. Here we can say that working class people are more in defaulters list

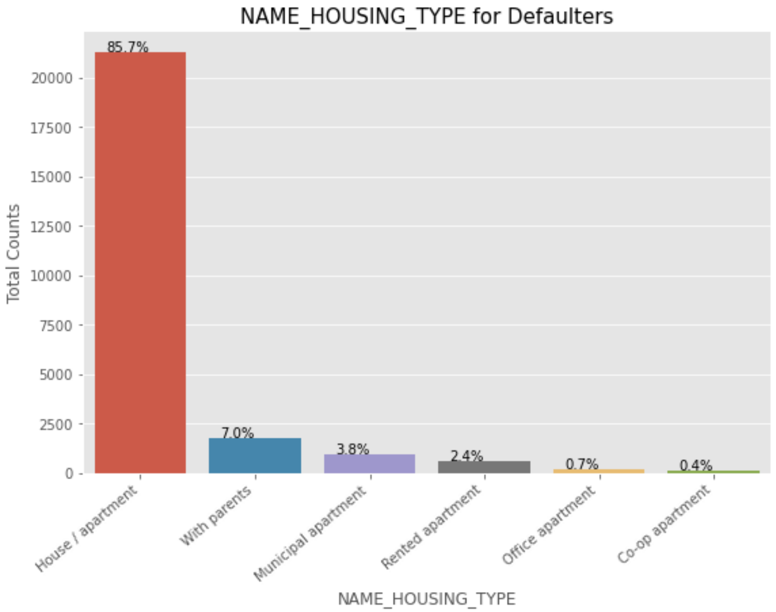




‘NAME_FAMILY_STATUS’

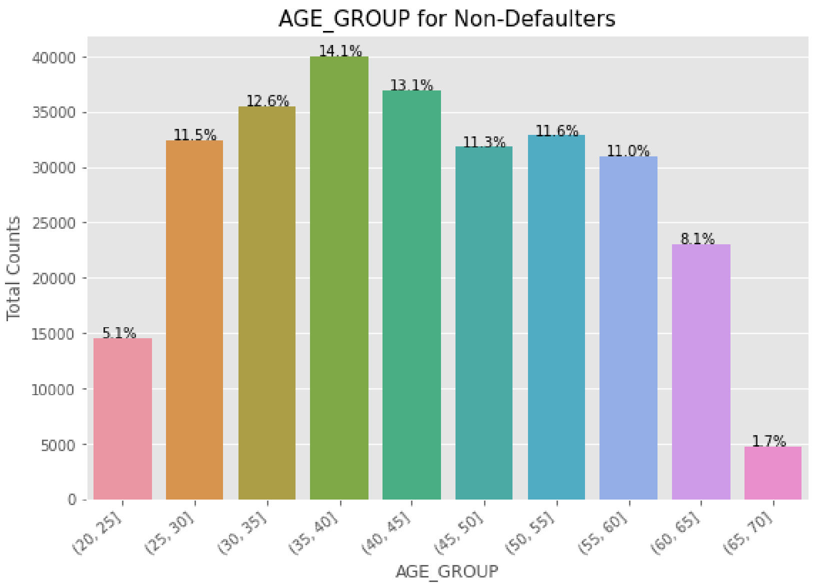
Usually we can see that married people are more for applying loans. As per the above graphs shows married people contribute 64.2% to non defaulters & 59.8% to defaulters there is more risk from married people

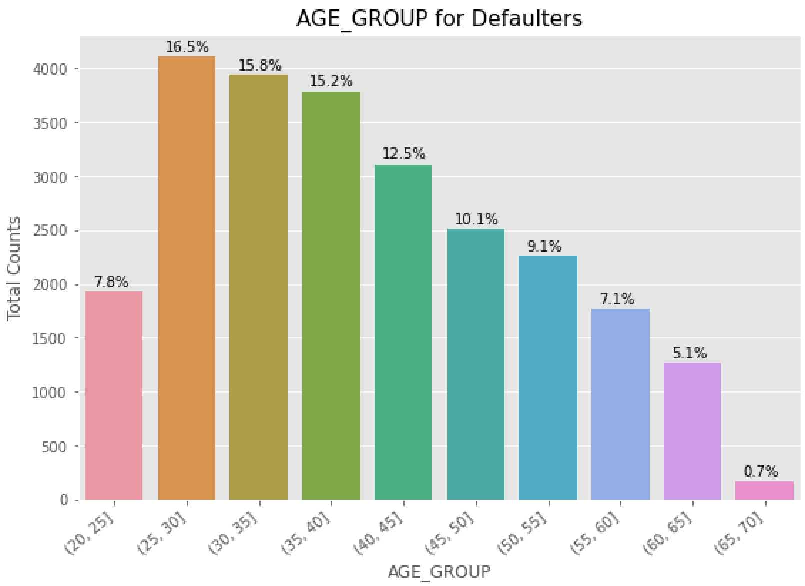




‘NAME_HOUSING_TYPE’

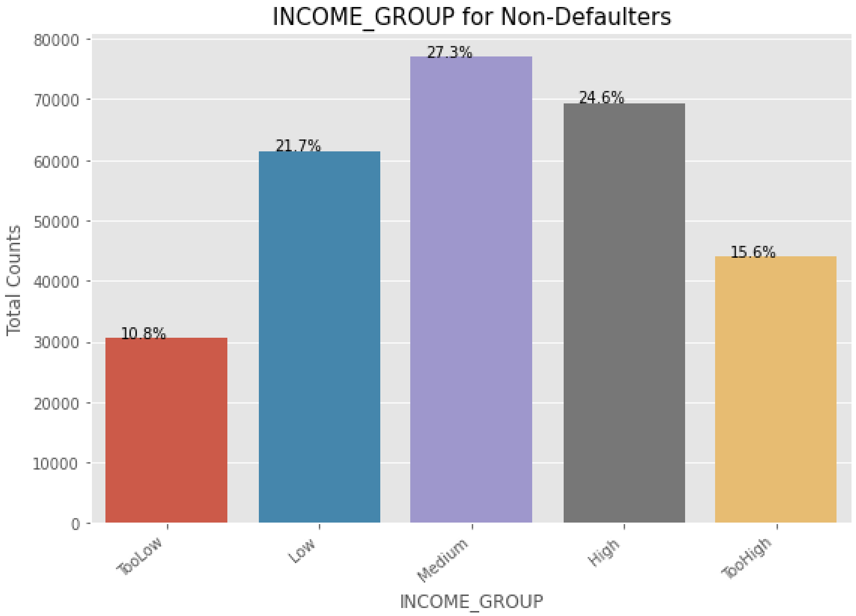
From the above graph we can easily say that people who has House/ Apartment those people apply more for loans

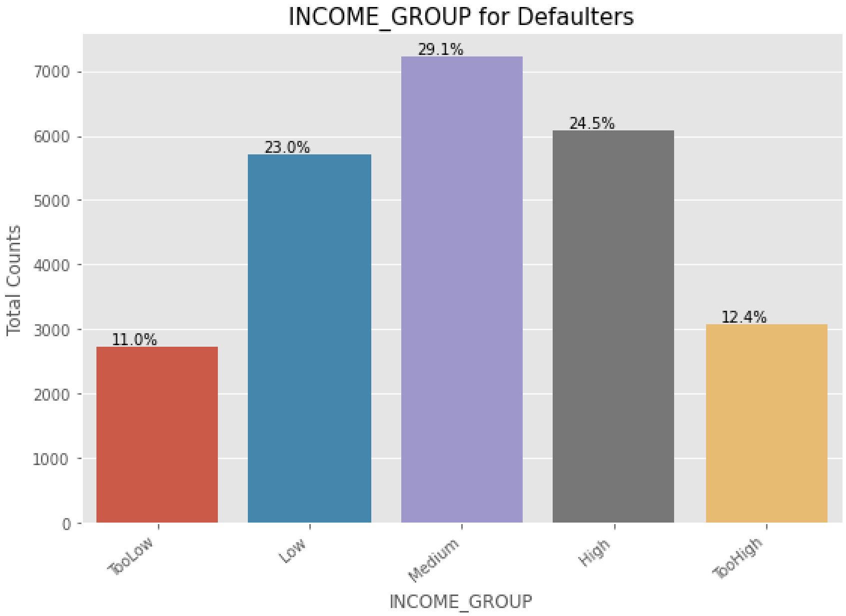




‘AGE_GROUP’

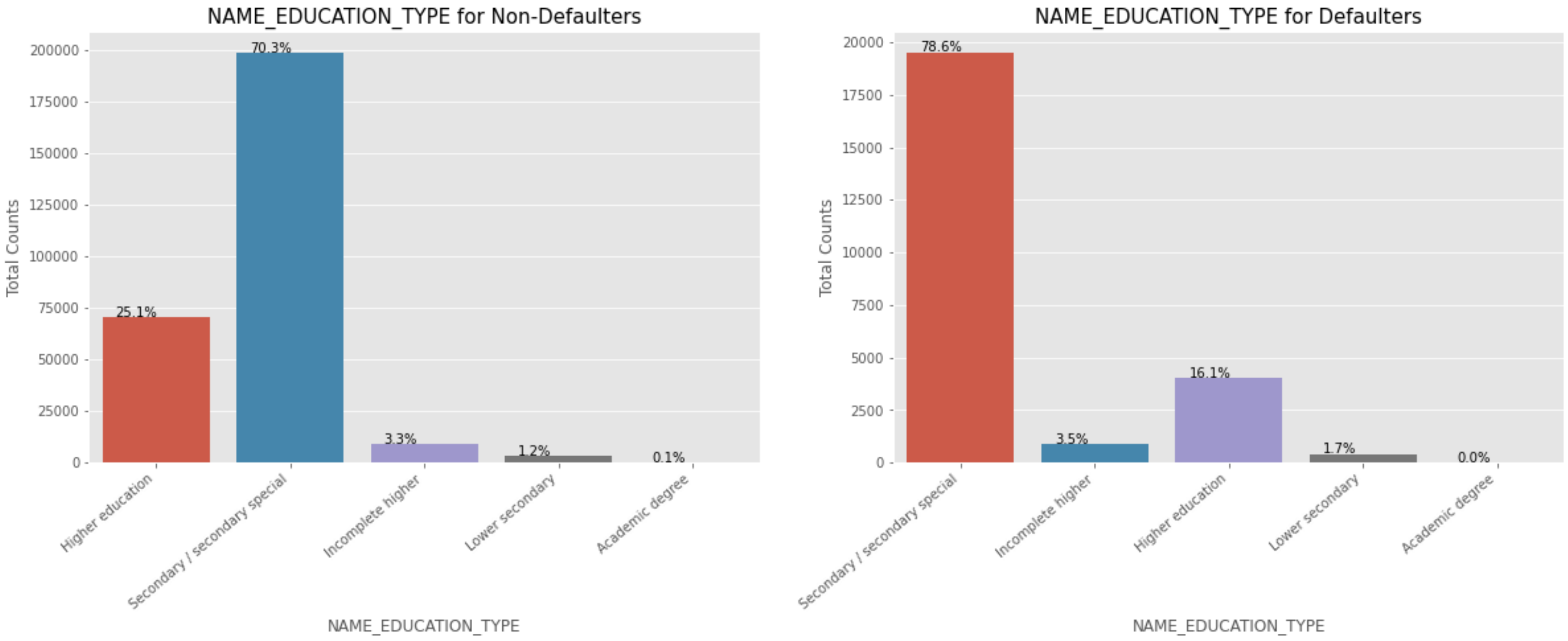
In the above graph we can see that 25-30 age group are more in default list





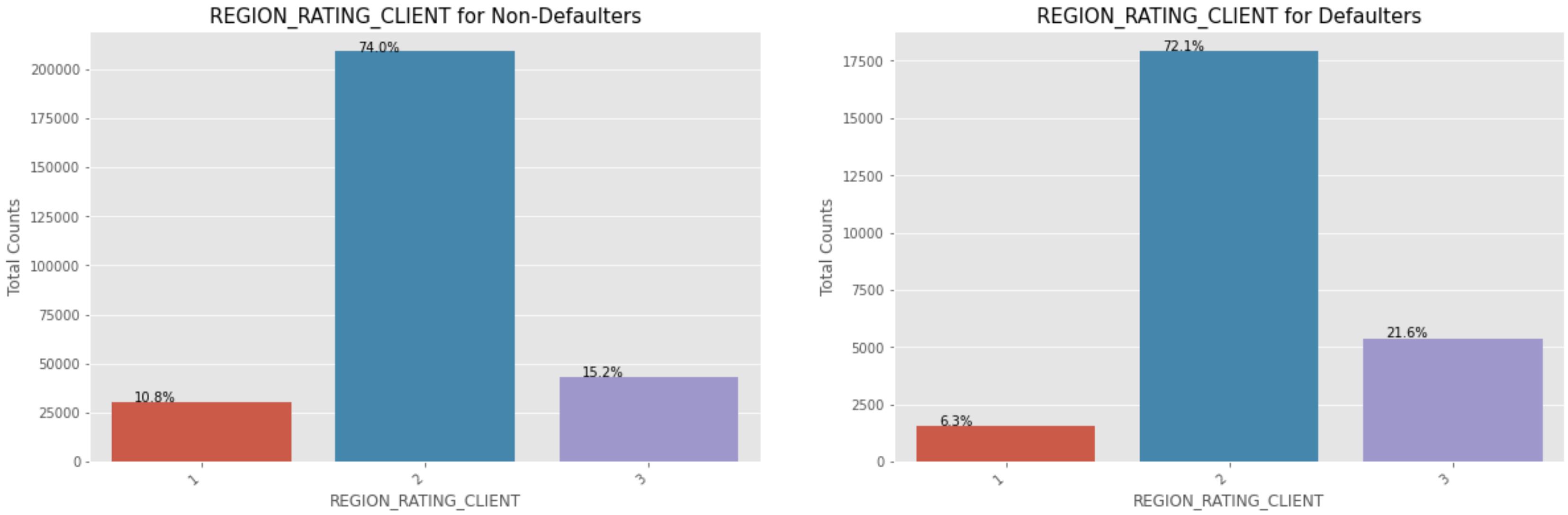
'INCOME_GROUP'

As per the above graph we can see that too high income group has less defaulters



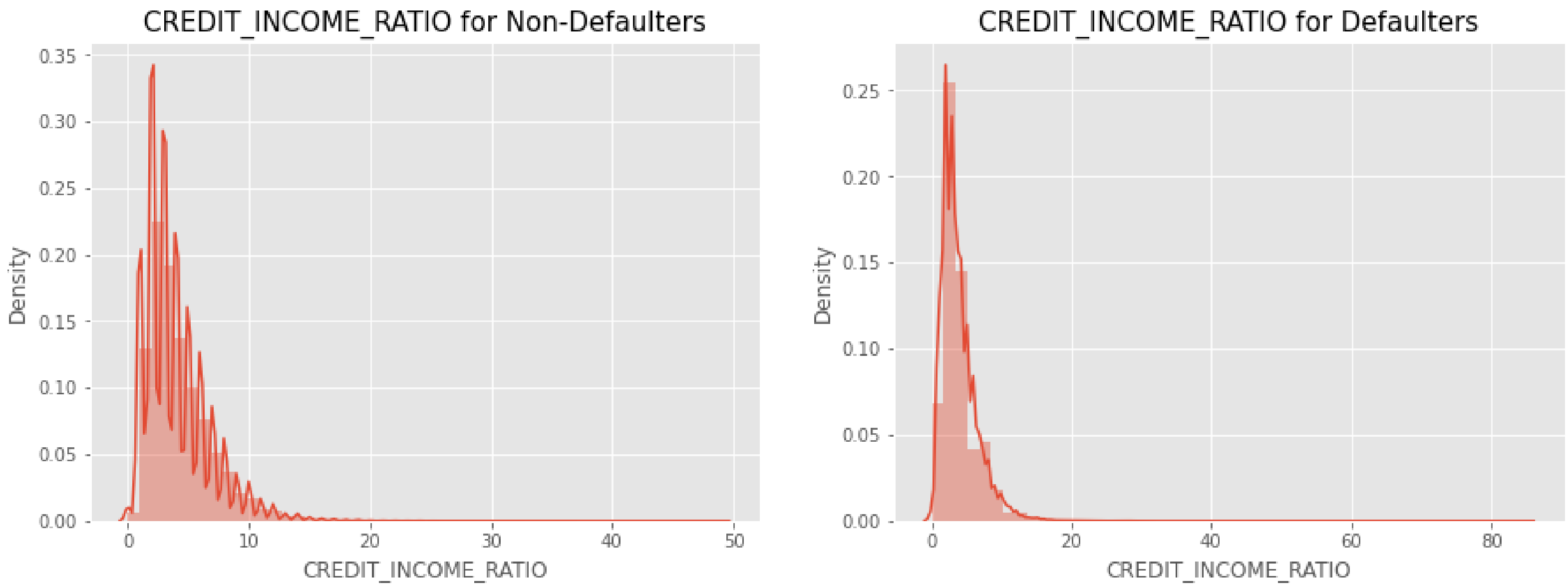
‘NAME_EDUCATION_TYPE’

In the above graph we can see that higher education people are less defaulters



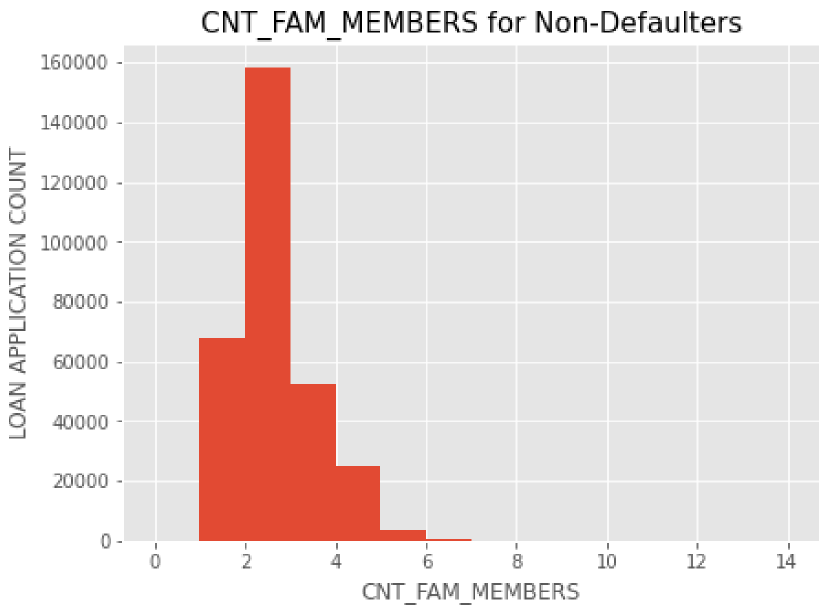
‘REGION_RATING_CLIENT’

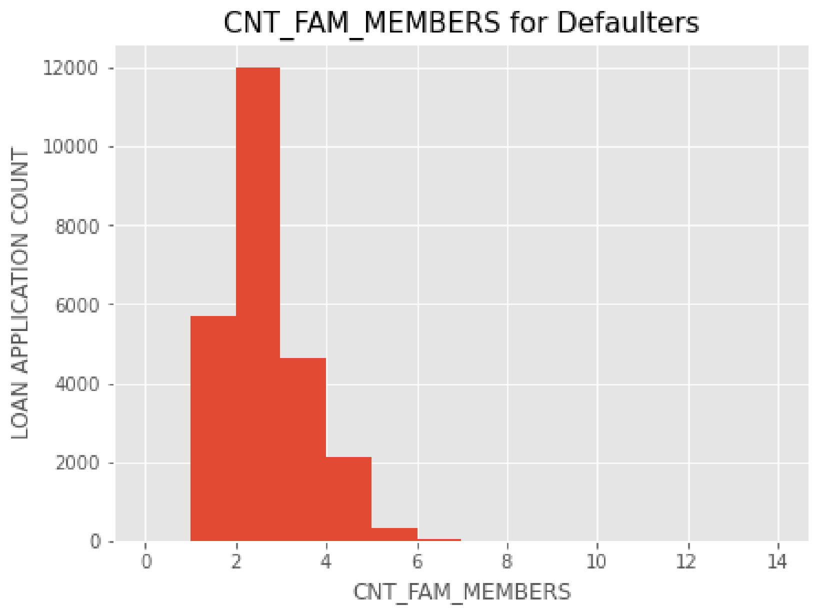
In the above graph we can see that 2 Rating people are more defaulters



‘CREDIT_INCOME_RATIO’

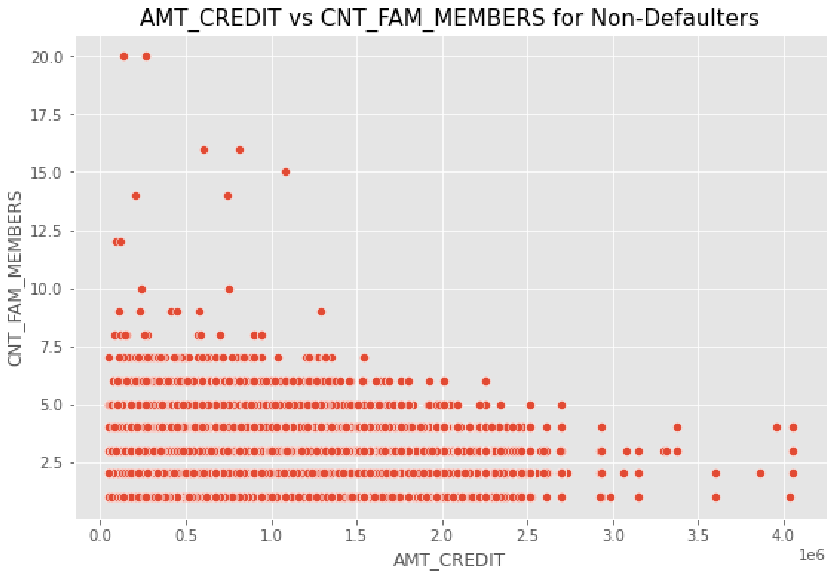
In the above graph we can see that CREDIT_INCOME_RATIO is more than 50 then it comes under defaulters

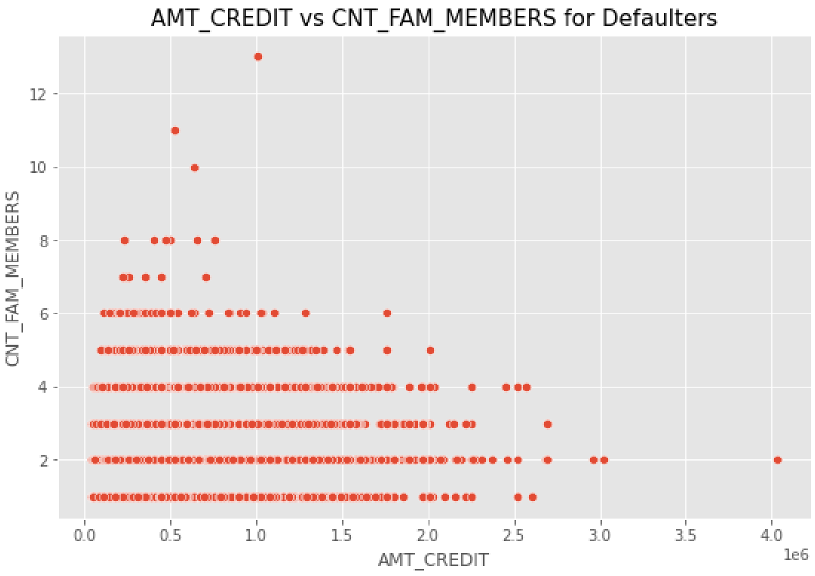




'CNT_FAM_MEMBERS'

In the above graph we can see that people of 2-3 members apply for more loans that other people

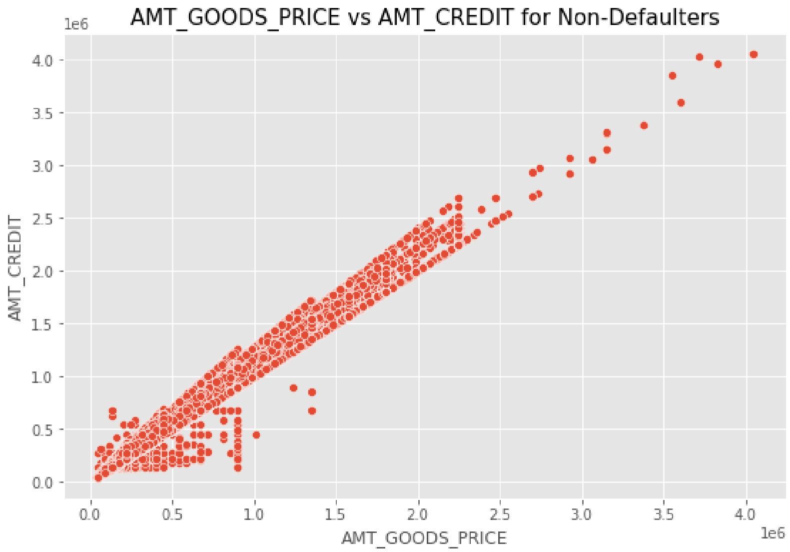


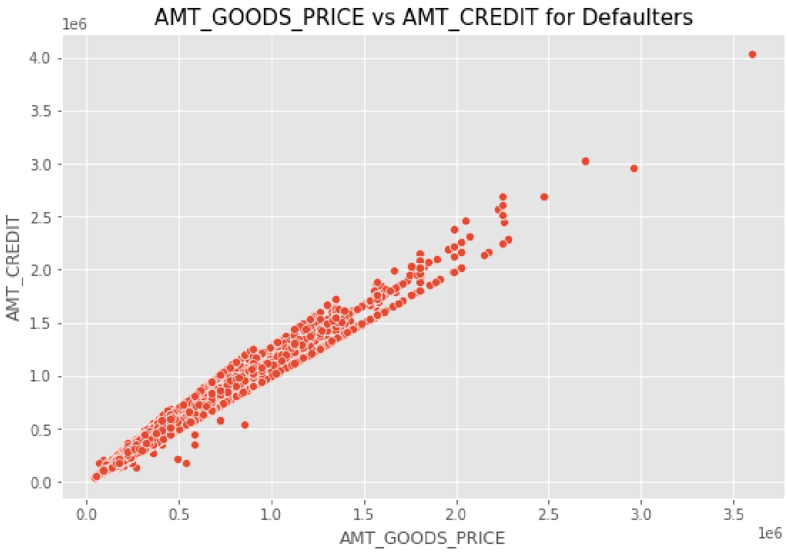


‘AMT_CREDIT’ VS ‘CNT_FAM_MEMBERS

In the above graph we can see that lower left corner is similar in both the case so larger families & people with high AMT_CREDIT are less defaulters. And people with low AMT_CREDIT are more defaulters.

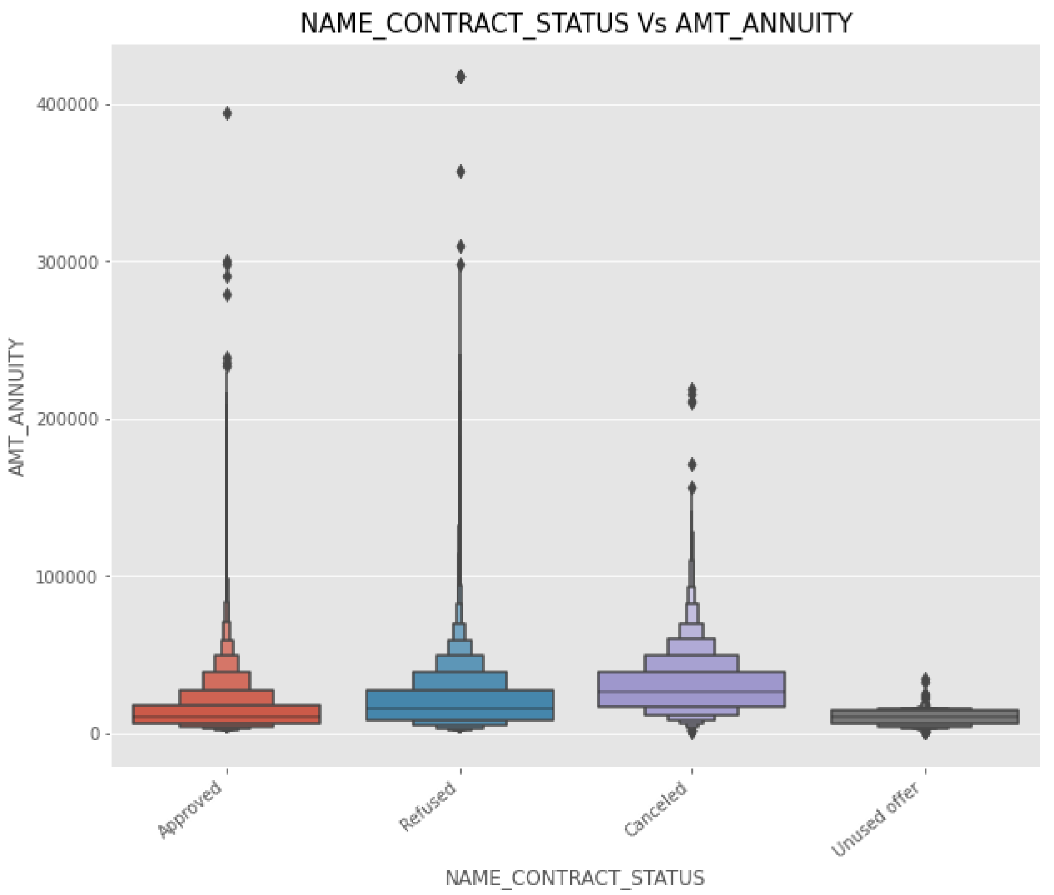
BIVARIATE ANALYSIS OF NUMERICAL VARIABLES





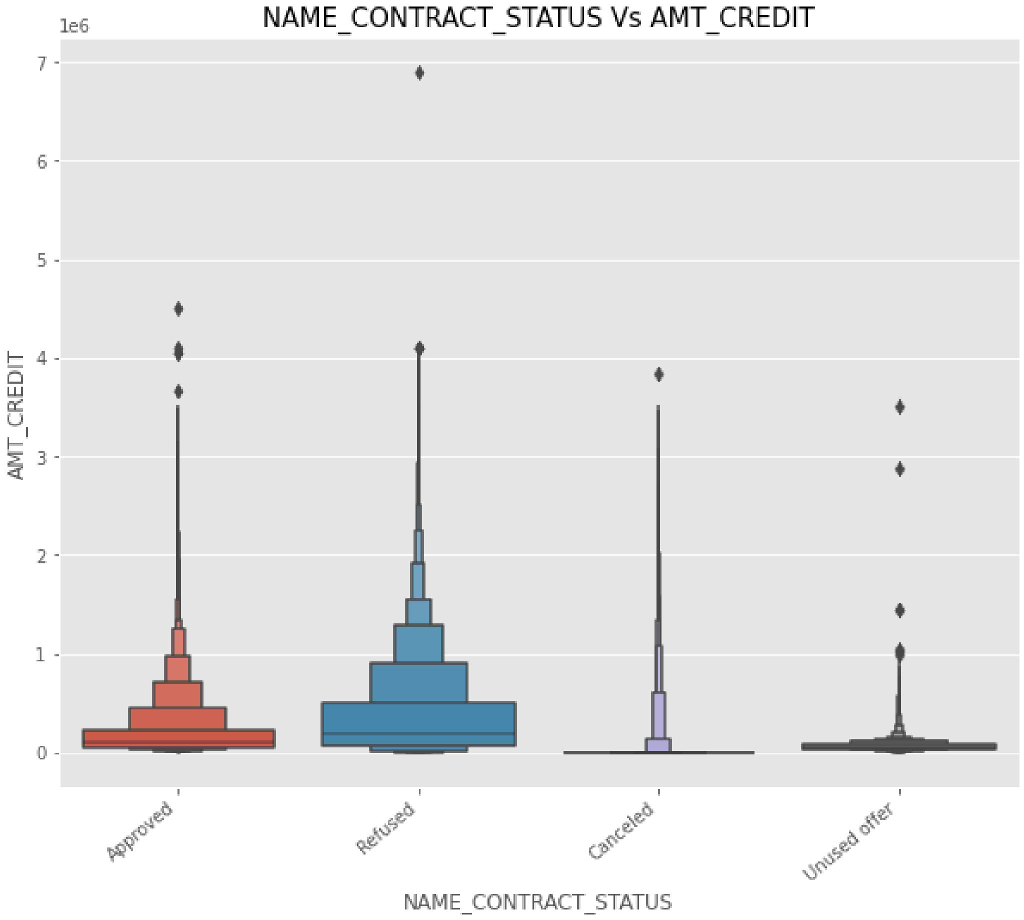
‘AMT_GOODS_PRICE’ VS ‘AMT_CREDIT’

In the above graph we can see that there are more non defaulters at low price and also less defaulters

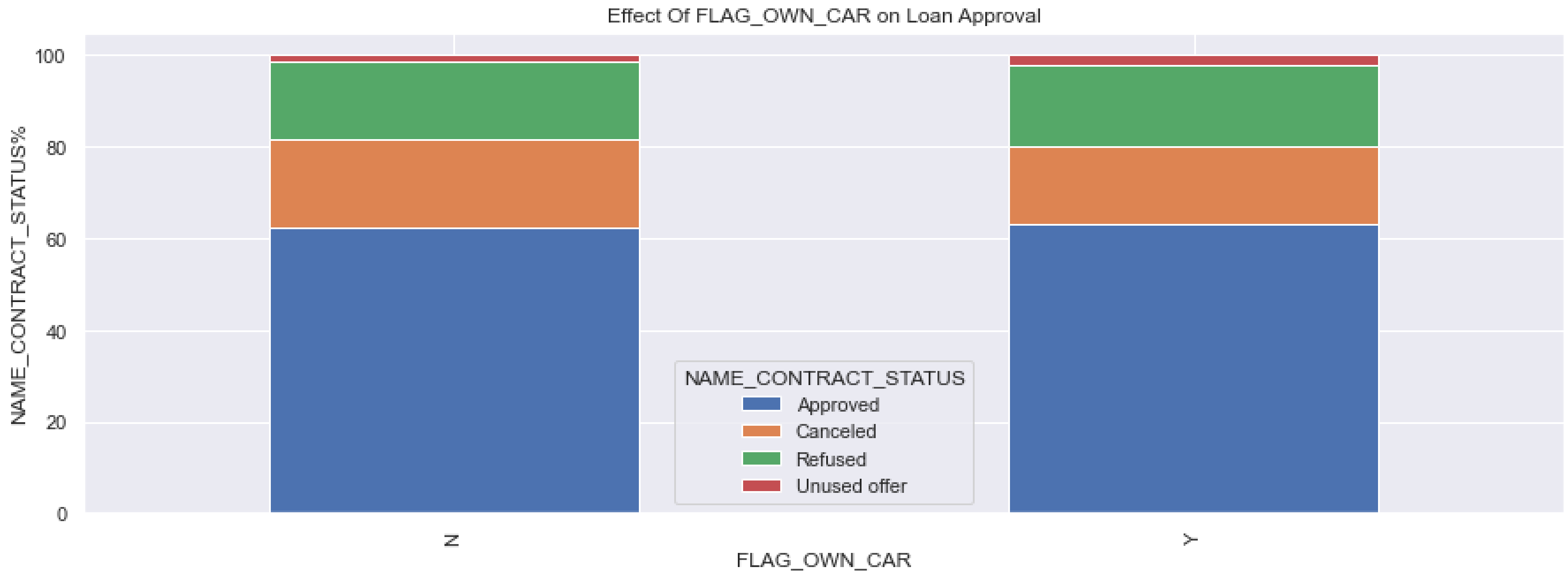


BIVARIANT ANALYSIS OF CONTRACT STATUS AND ANNUITY OF PREVIOUS APPLICATION

In above graph we can see that loan application for people with low AMT_ANNUITY gets cancelled and applications with high AMT_ANNUITY gets refused many times

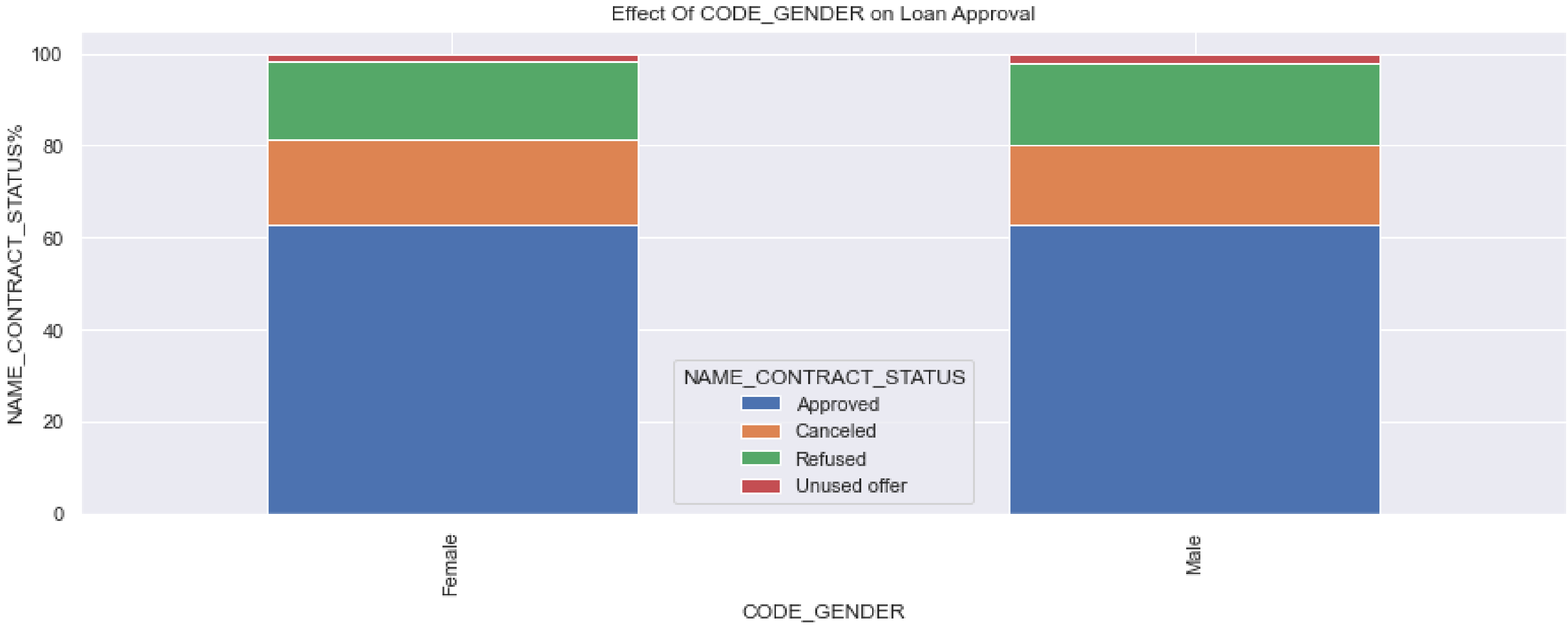


In the above graph we can see that AMT_CREDIT is too low then it gets cancelled.

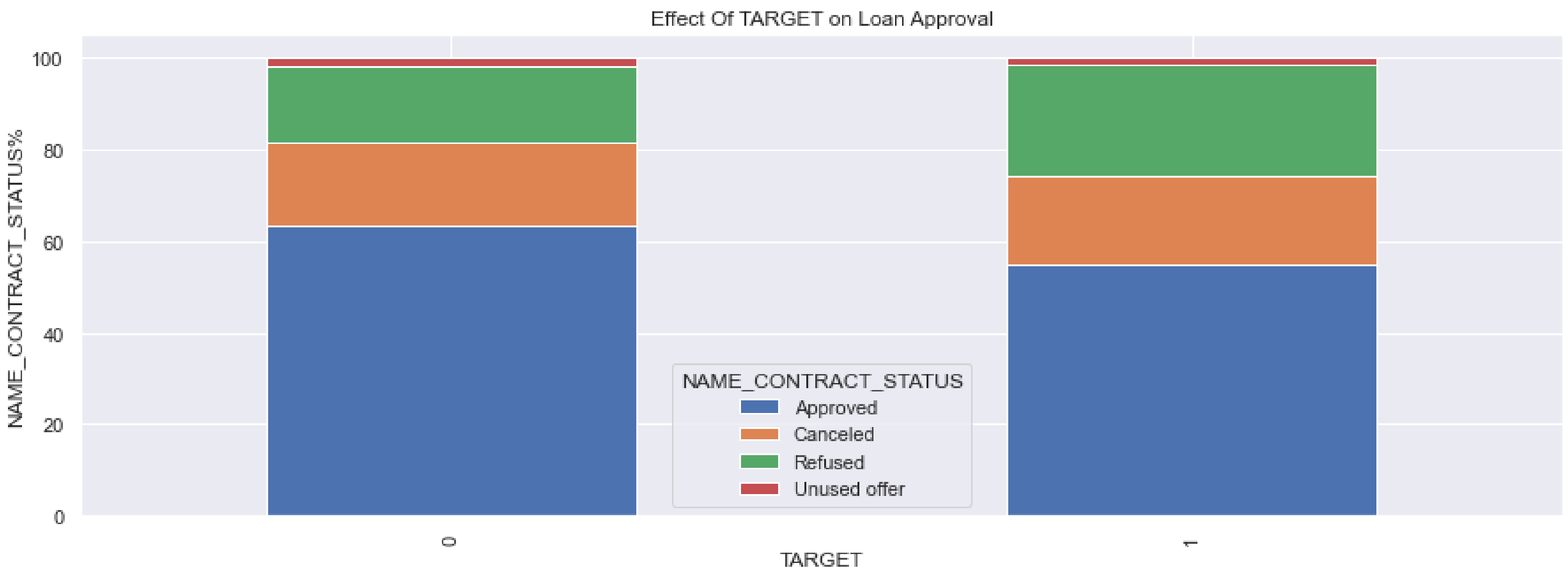


MERGING THE DATA FROM BOTH THE AND ANALYZING THE DATA

In the above graph we can see that people having car does not have any effect after merging the data.



In the above graph we can see that there is no effect for gender after merging the data.



In the above graph we can see that people who has approved loan previously are defaulted less & people who refused loan previously have high chances of defaulting.

Banks should give loans to people from income group businessman, student, pensioner

Banks should avoid giving loans to married people as the defaulters percentage is more in married people.

They should give loans to age group of 35-40 as these people have high income from other age group.

Bank should see if the people are from higher education background are less defaulters

LOAN RECOMMENDATIONS

THANK YOU

Onkar Suryawanshi