

# EDA\_CREDIT ASSIGNMENT

By SCHIFRA DARUWALLA

# PROBLEM STATEMENT

- To conduct an EDA on the Credit data to analyze who are the defaulters and who are the non defaulters.
- The goal also includes which of the defaulters are getting the loan approved and which of the non defaulters are getting their loans rejected.
- Both the cases cause the banks losses.

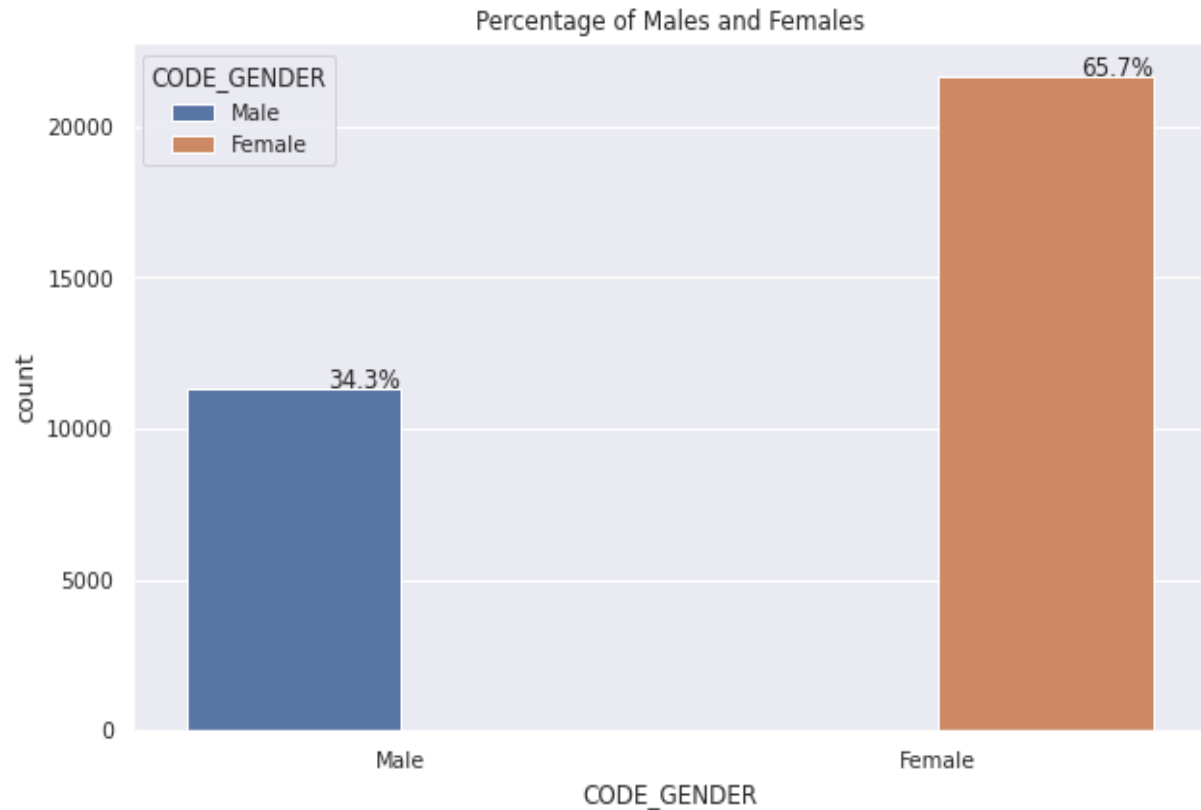
# ASSUMPTIONS

- The columns with more than 40% missing values were removed
- It is assumed that the XNA in the CODE\_GENDER column are those who do not identify as Male nor Female and for this reason XNA has been classified as Others.
- Any XAP category are assumed to be not applicable to the given categories and are considered as a miscellaneous category. XNA are assumed to stand for missing values or unknown.
- One of the values in the New contract type column was *cas*. It seems to be due to human error where *cas* was written in place of cash loan. Hence *cas* has been imputed as cash loan.
- The value of the DAYS\_EMPLOYEES which are greater than 300000 days (1000.66 years) was changed to the median value of 2226 days. This outlier could be due to human error as it fails the sanity test.

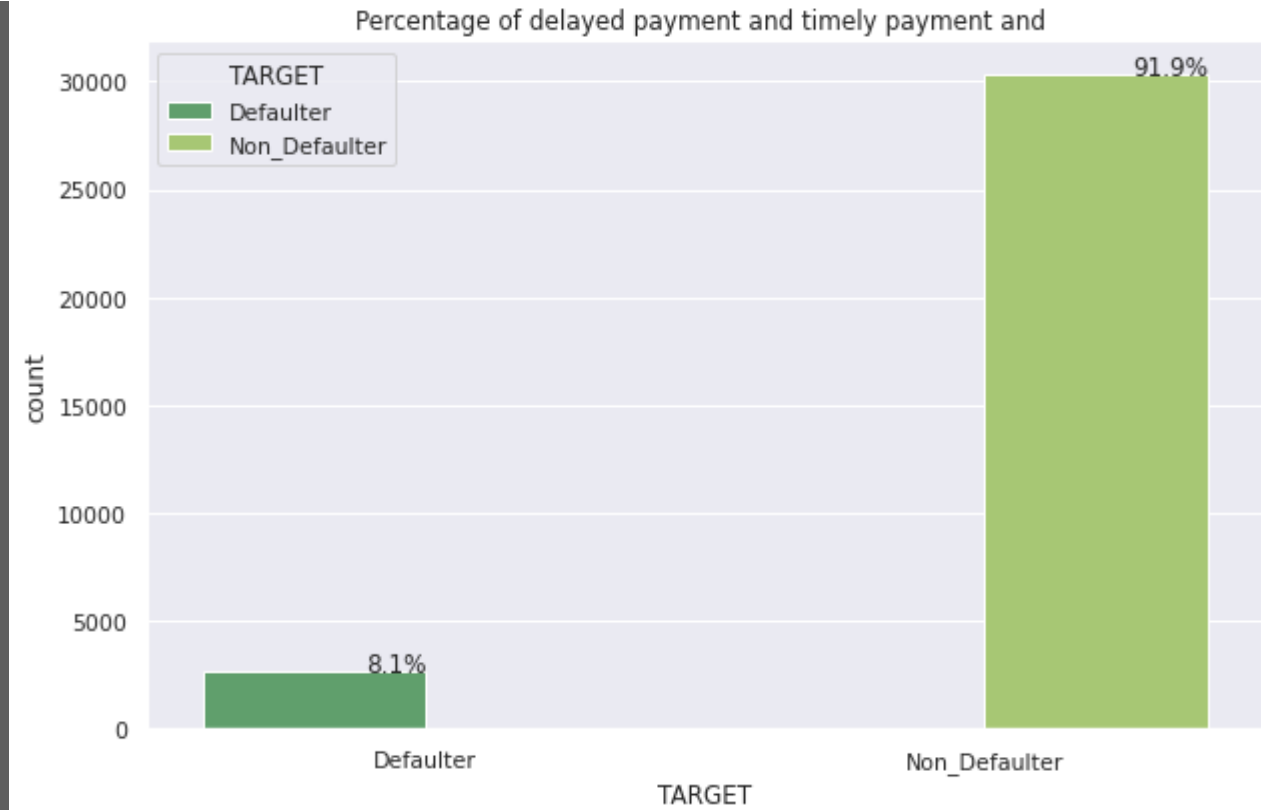
# APPROACH

- The approach is to keep the data as authentic as possible and only deal with the outliers and missing values when urgently required. This method is practiced as there are possibilities that these outliers are indicators of defaulters.
- First step is to remove the columns which have more than 40% missing values.
- Second step is to conduct a univariate analysis for the variables. If the variable is numerical, sanity test is performed and also outliers are checked. If the outliers are caused by the distribution of the data, then they are left alone, otherwise they are to be rectified according to the situation. Most missing values are left as they are but for not applicable category, they could be left as it is or categorized as Others.
- Third step is to conduct bivariate analysis and multivariate analysis.

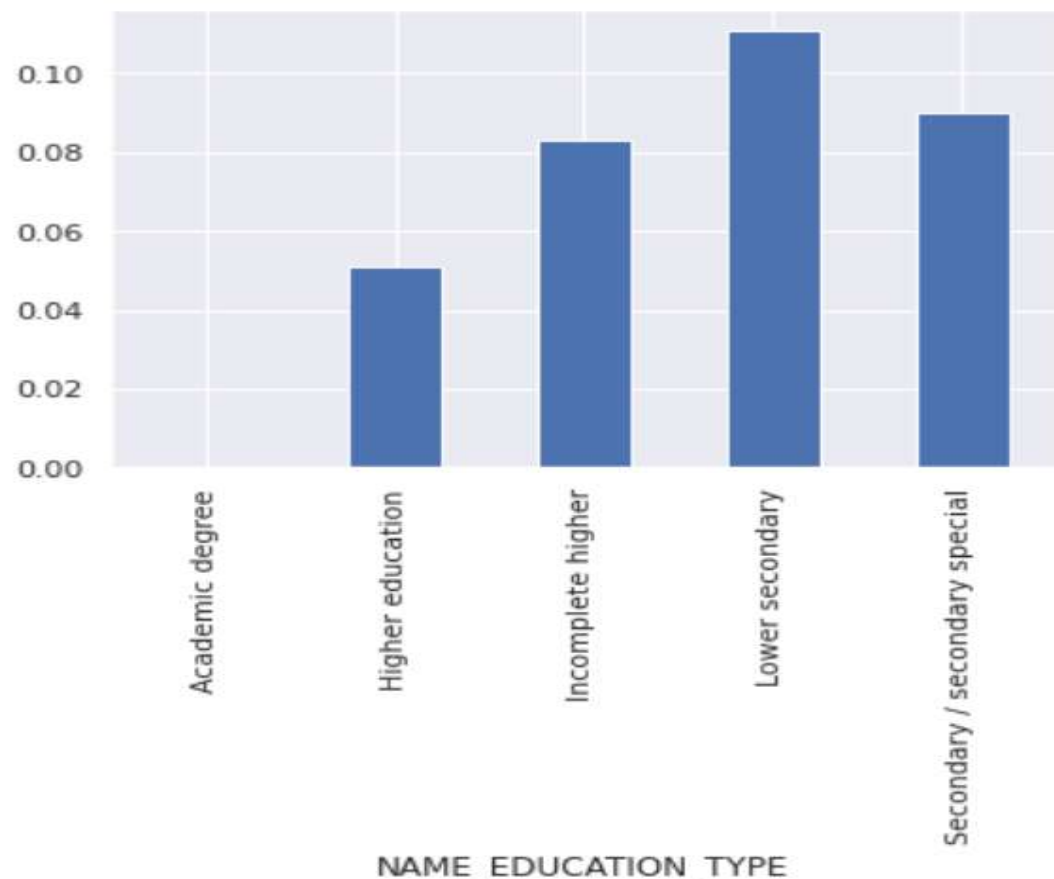
# RELEVANT ANALYSIS



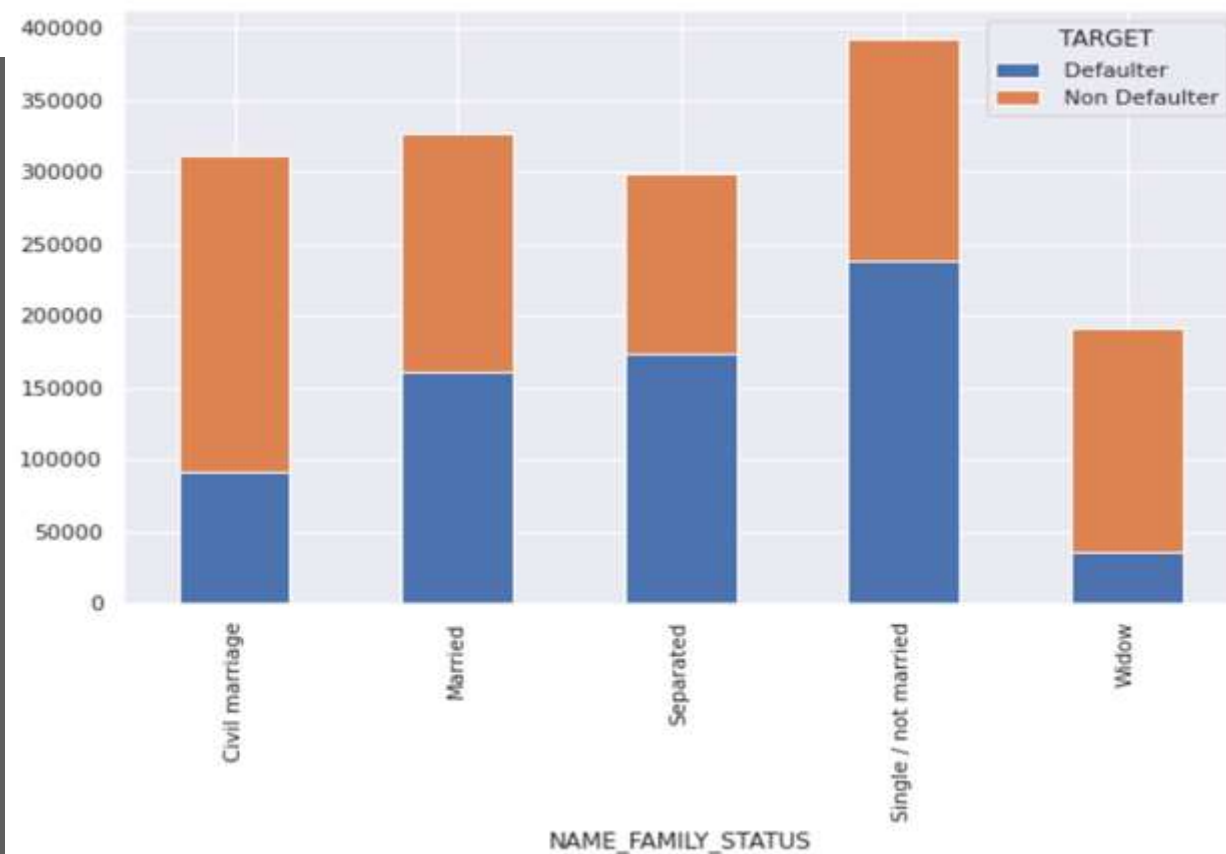
65.7% of the Females opt for loans while only 34.3% Males opt for loans



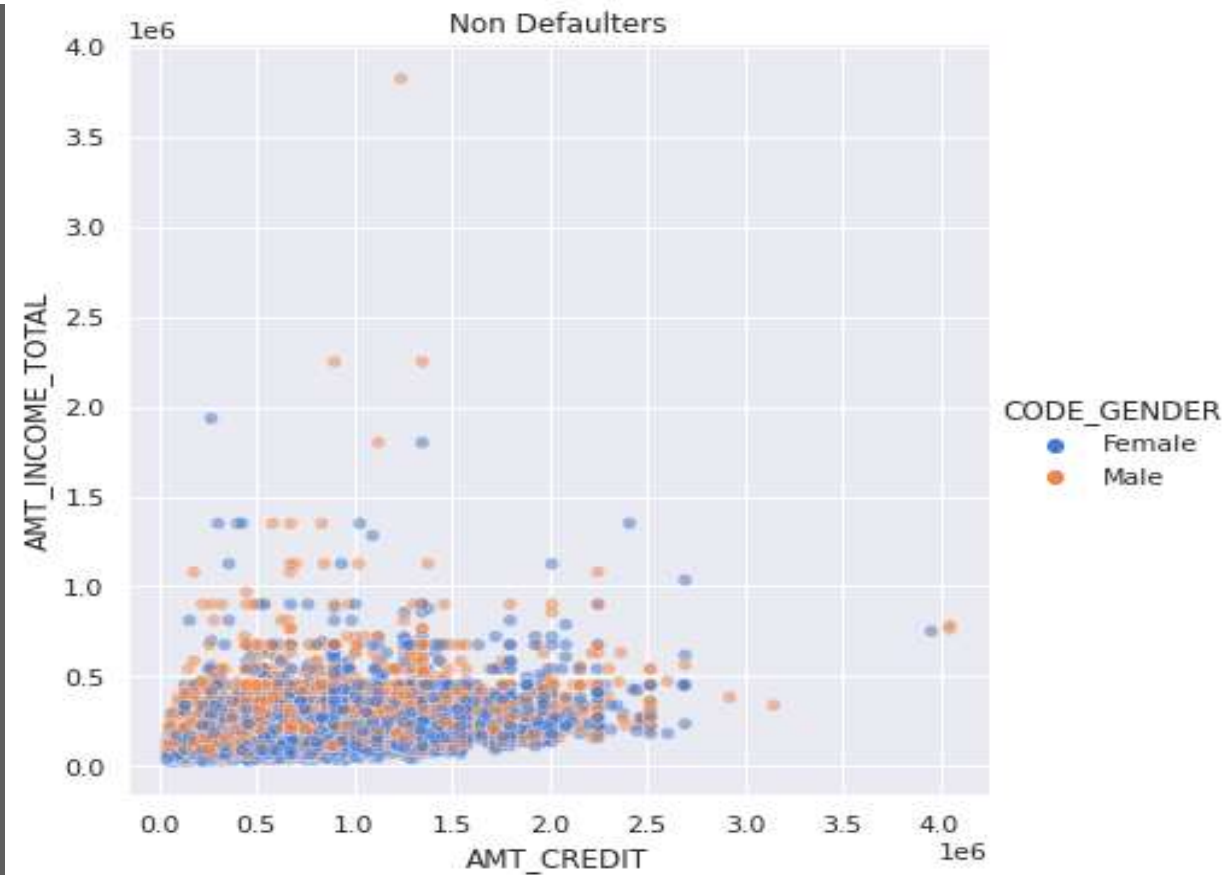
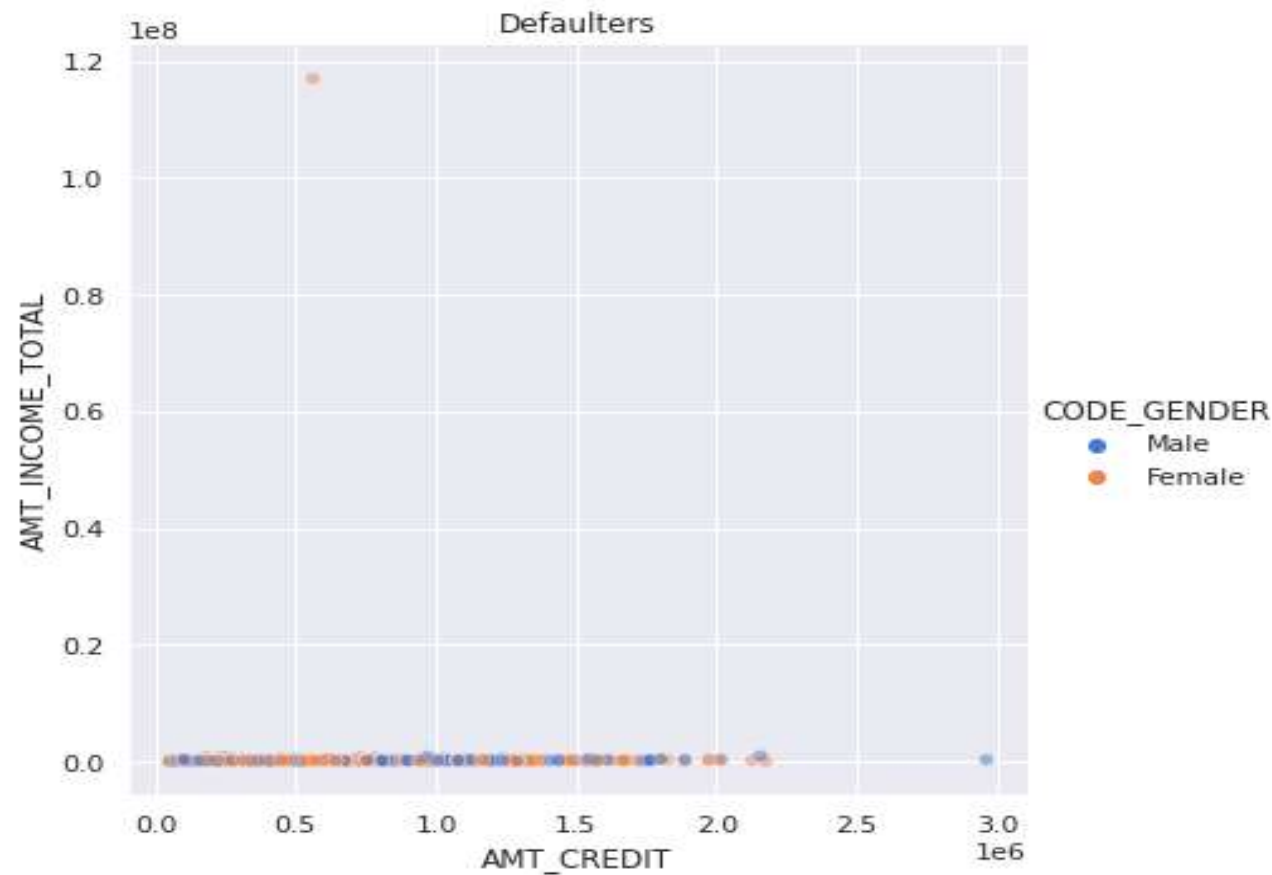
Majority of the customers are non defaulters (91.9%) while only 8.1% are defaulters.  
The data is unbalanced as the ratio of Non defaulters to defaulters is not 1:1 . It is 91.9% : 8.1% or 2759:242.



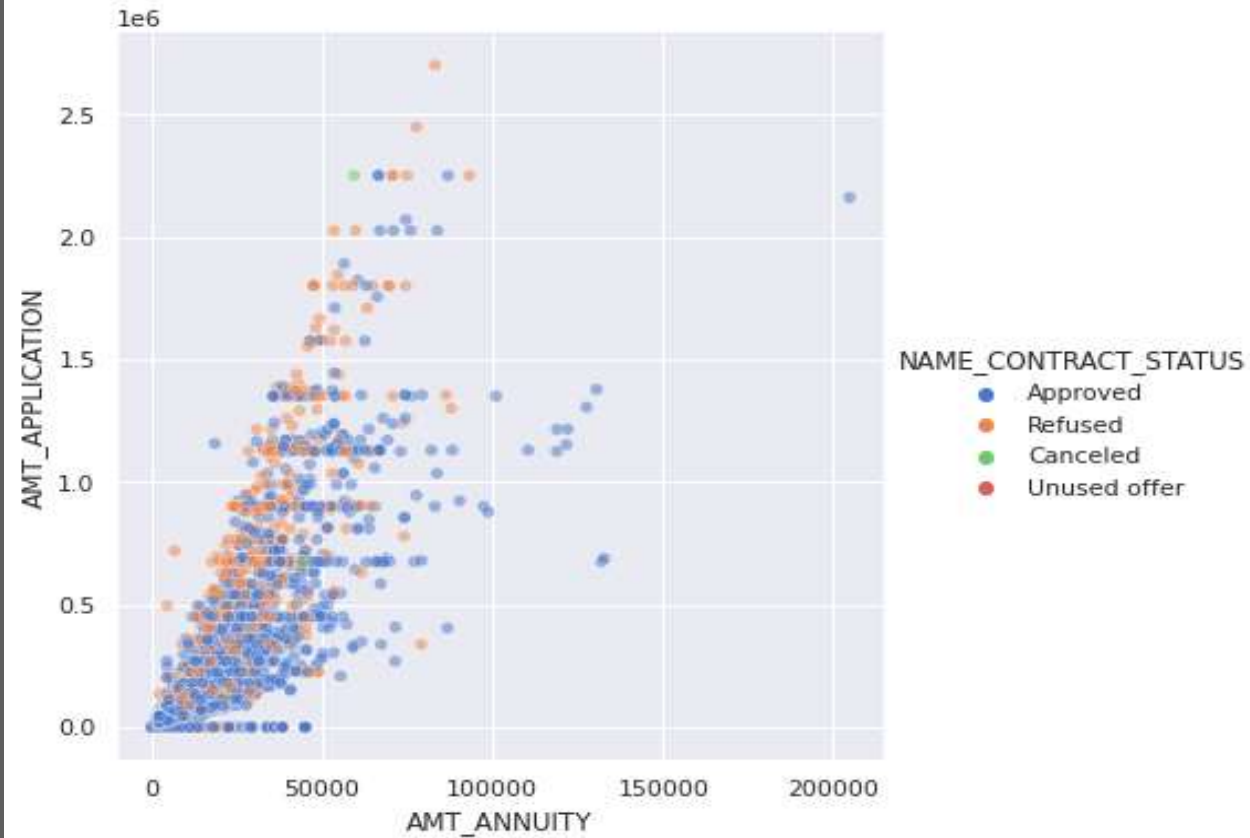
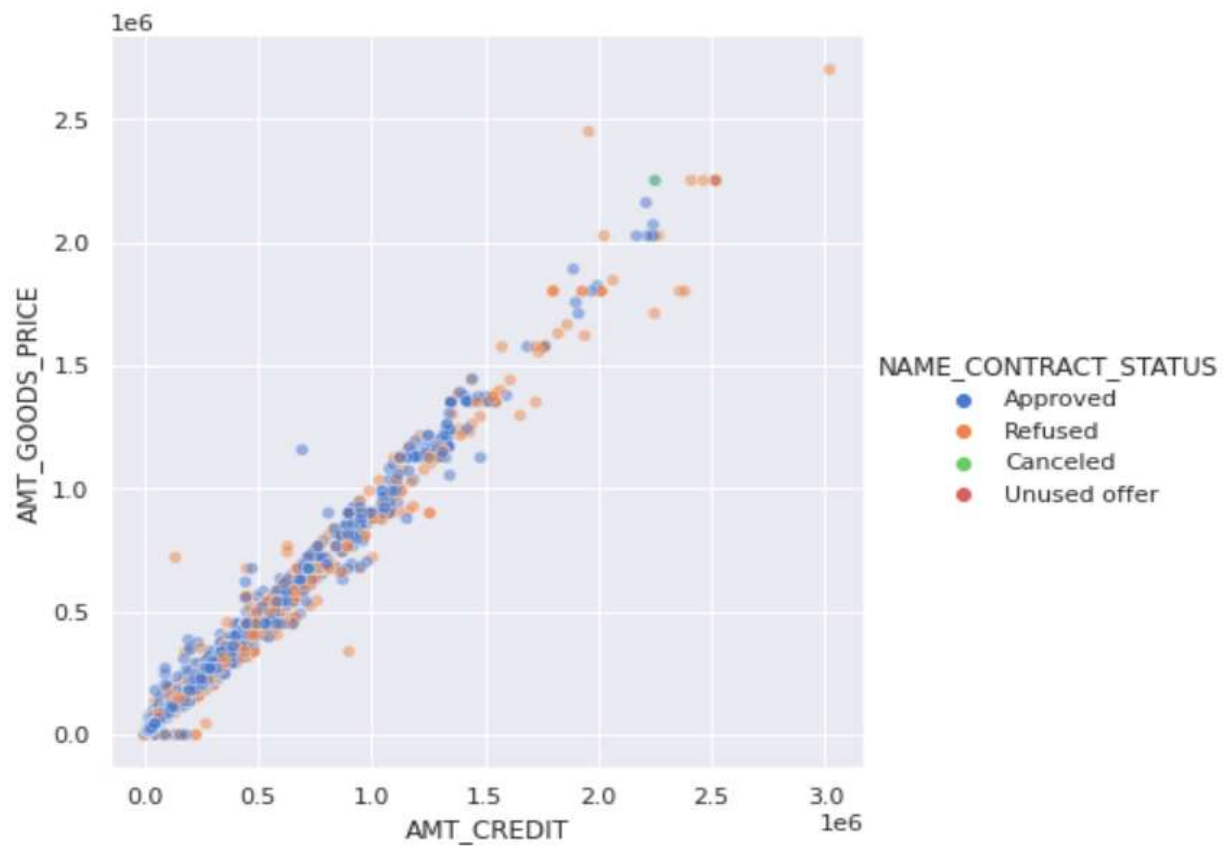
Those with Lower Secondary, secondary, incomplete education are more likely to default.



The first bar chart not married/ singles opted for loans the most, followed by married and civil married. The married, singles and separated defaulted much more.



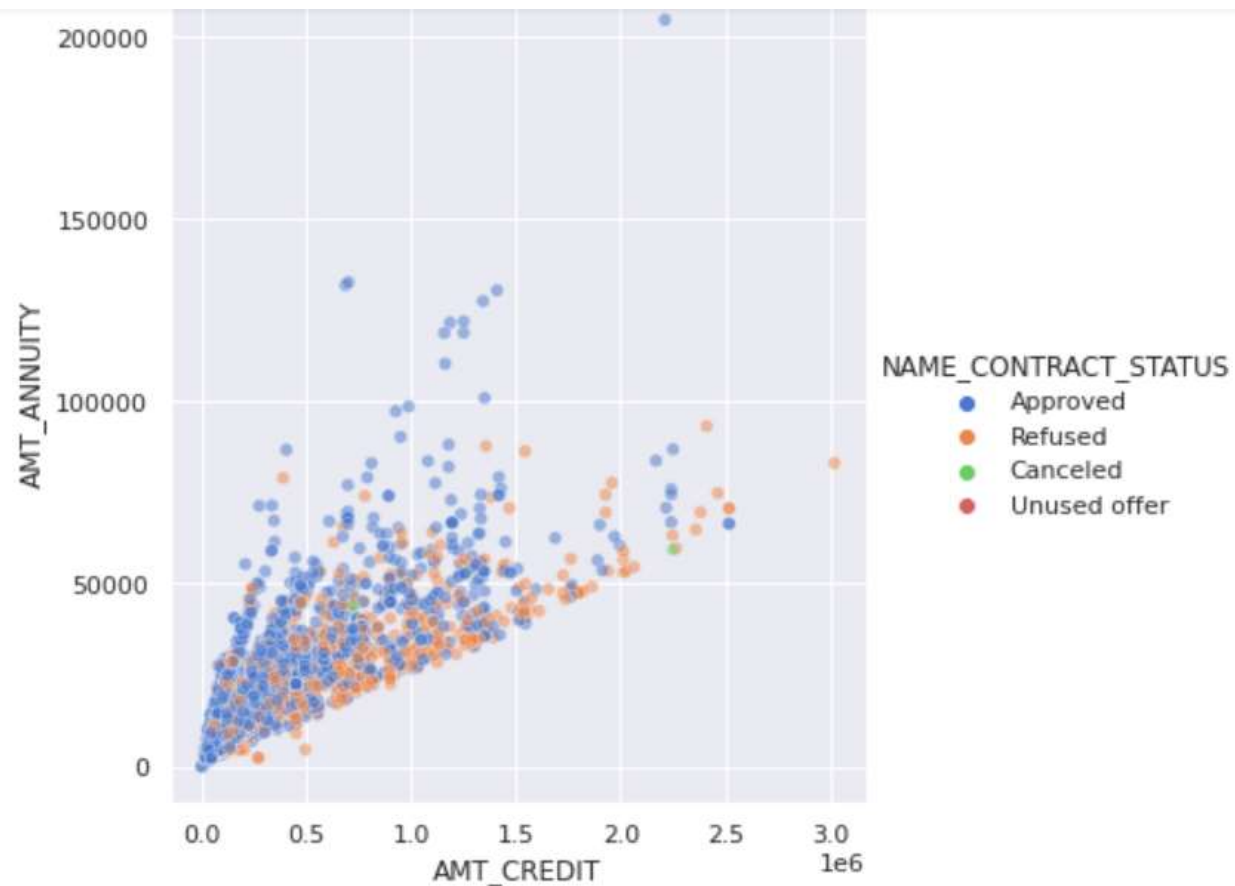
On the left scatterplot, the income rate of the client is 0.0 but takes credit of upto 2.0. Both male and female tend to default equally in this case. In the second case, most have an income rate of more than 0.0 and they take credit of upto 2.5. Most of them taking loans are females. Males tend to have higher income than females



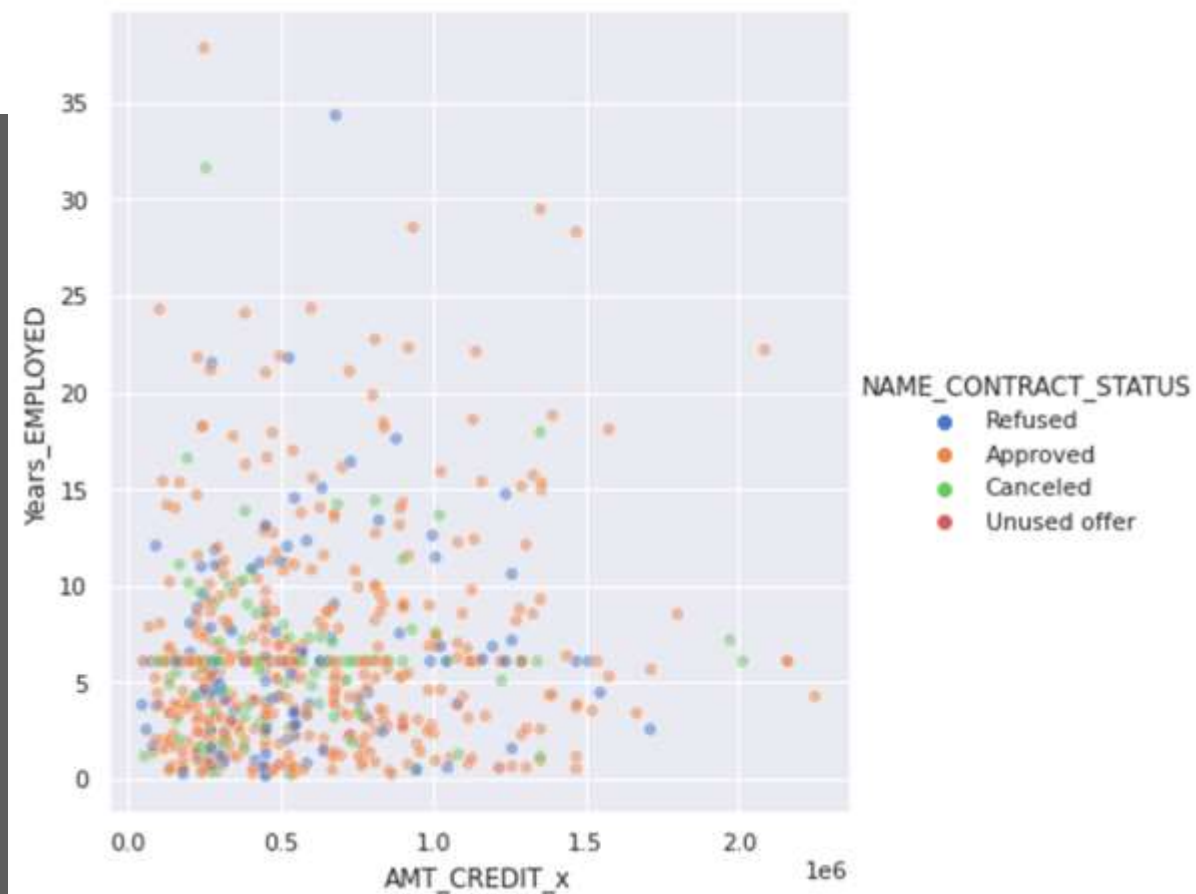
On the left scatterplot, most with lower credit amount of the loan and price of good for which the loan was taken out were approved. But for those beyond 1.5 tended to be rejected.

On the right scatterplot, those falling more towards loans annuity were accepted but those who asked for higher amounts in the loan application and took less loans annuity were refused loans.

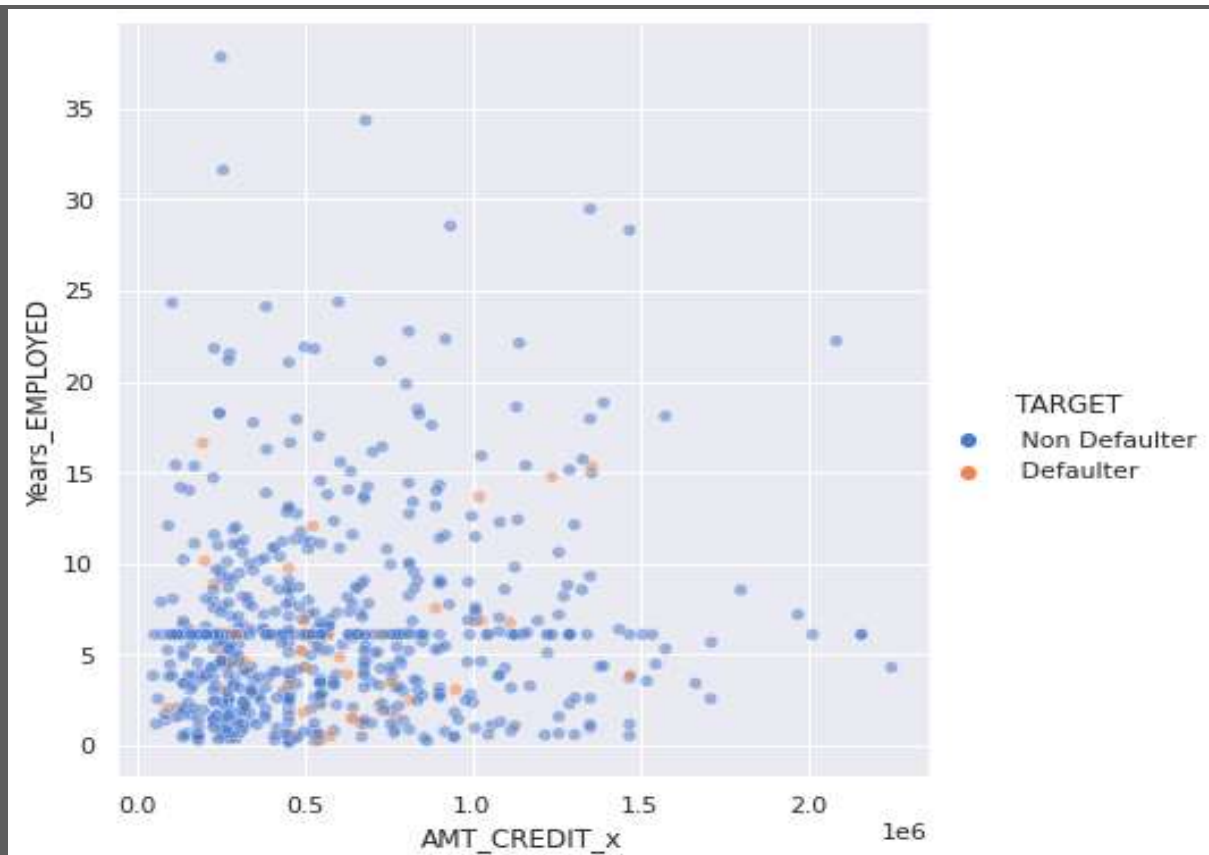
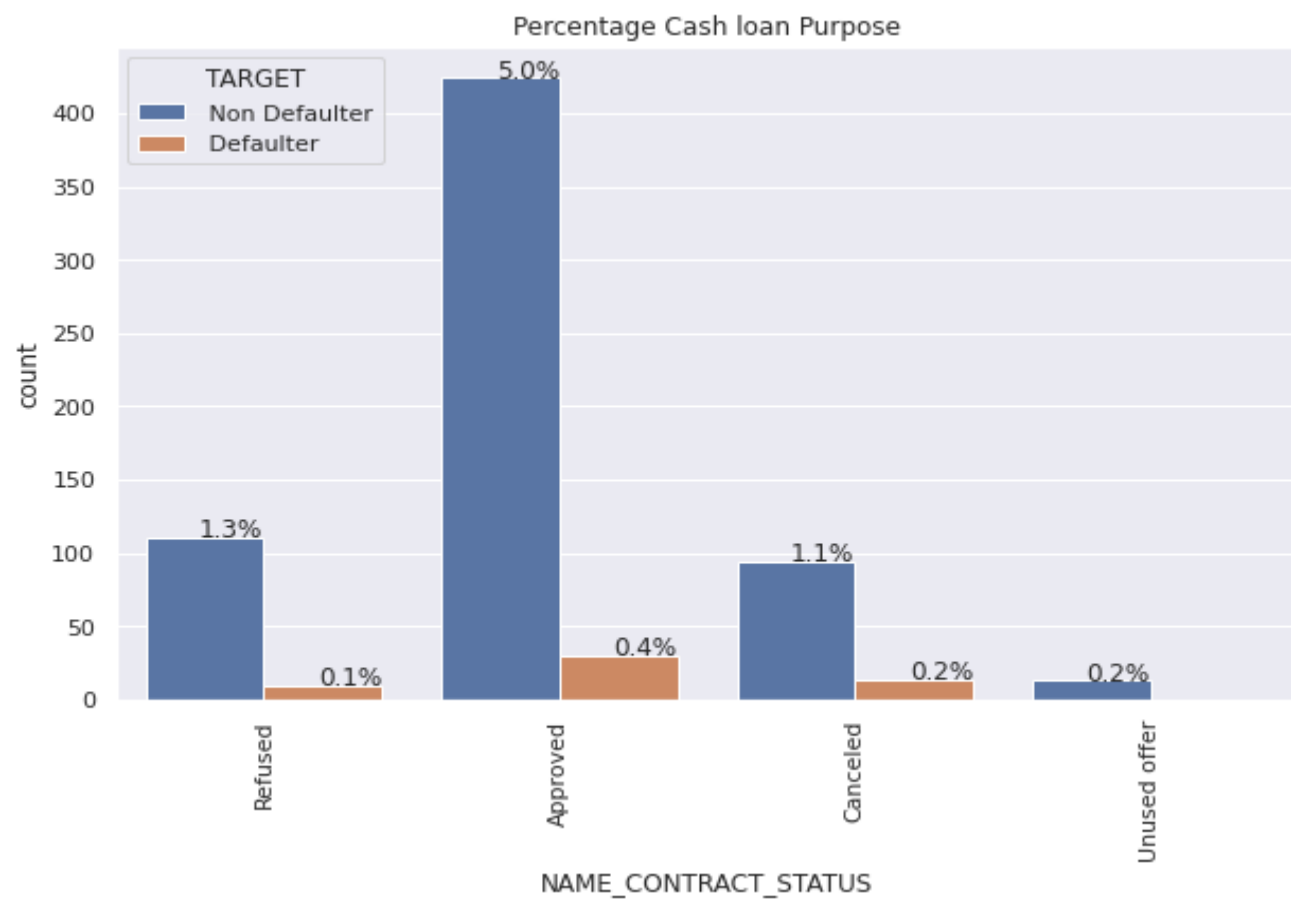




In the first scatter plot more loans were approved if they opted for higher loan annuity and lower credit amount of the loan. They were more likely to be rejected if they opted for higher credit amount and lower amount of loan annuity

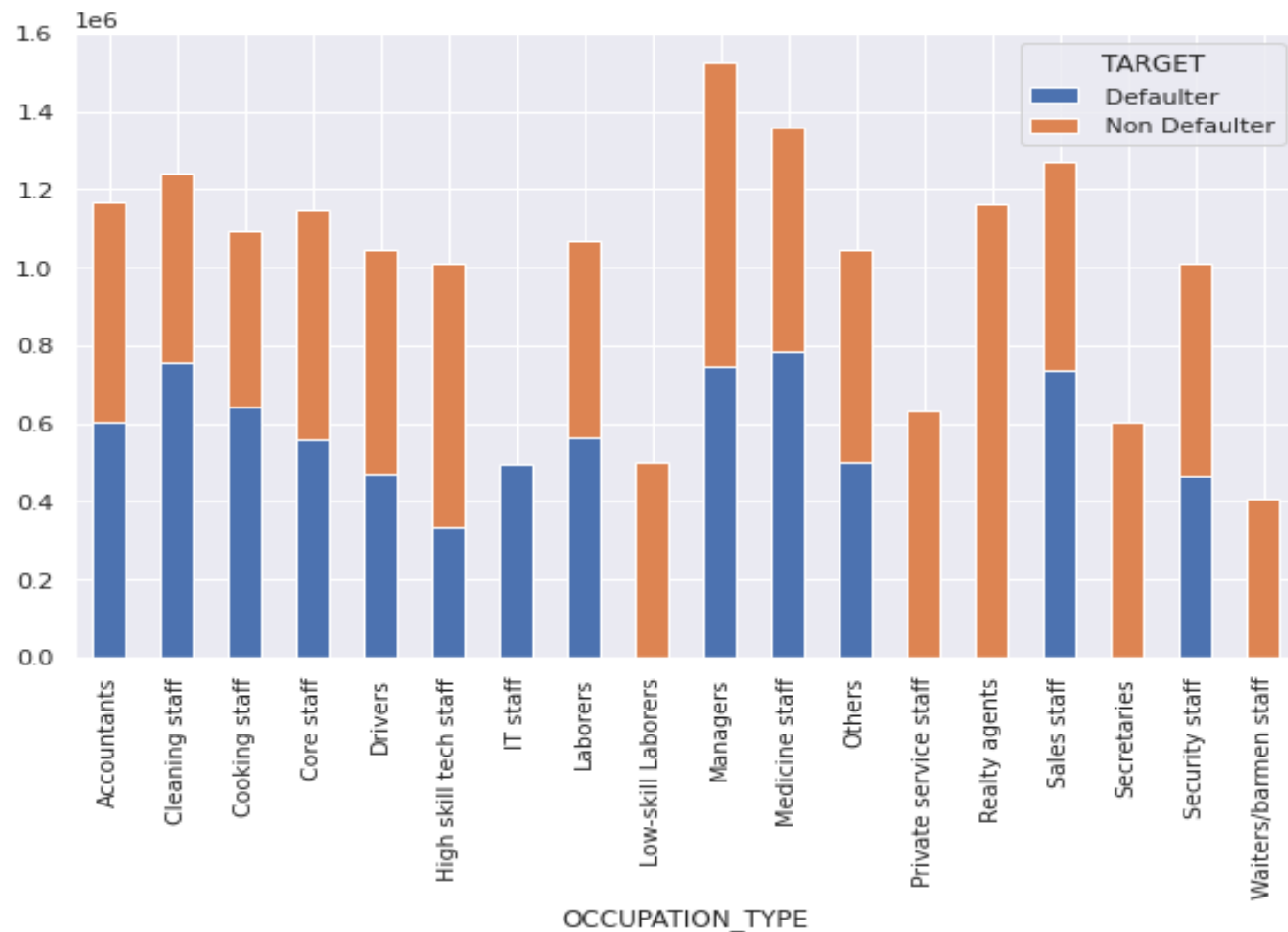


Those refused are found more densely to have worked for less than 5 years and opted for less than 0.5 % However, the refused ones are found outside these areas as well.

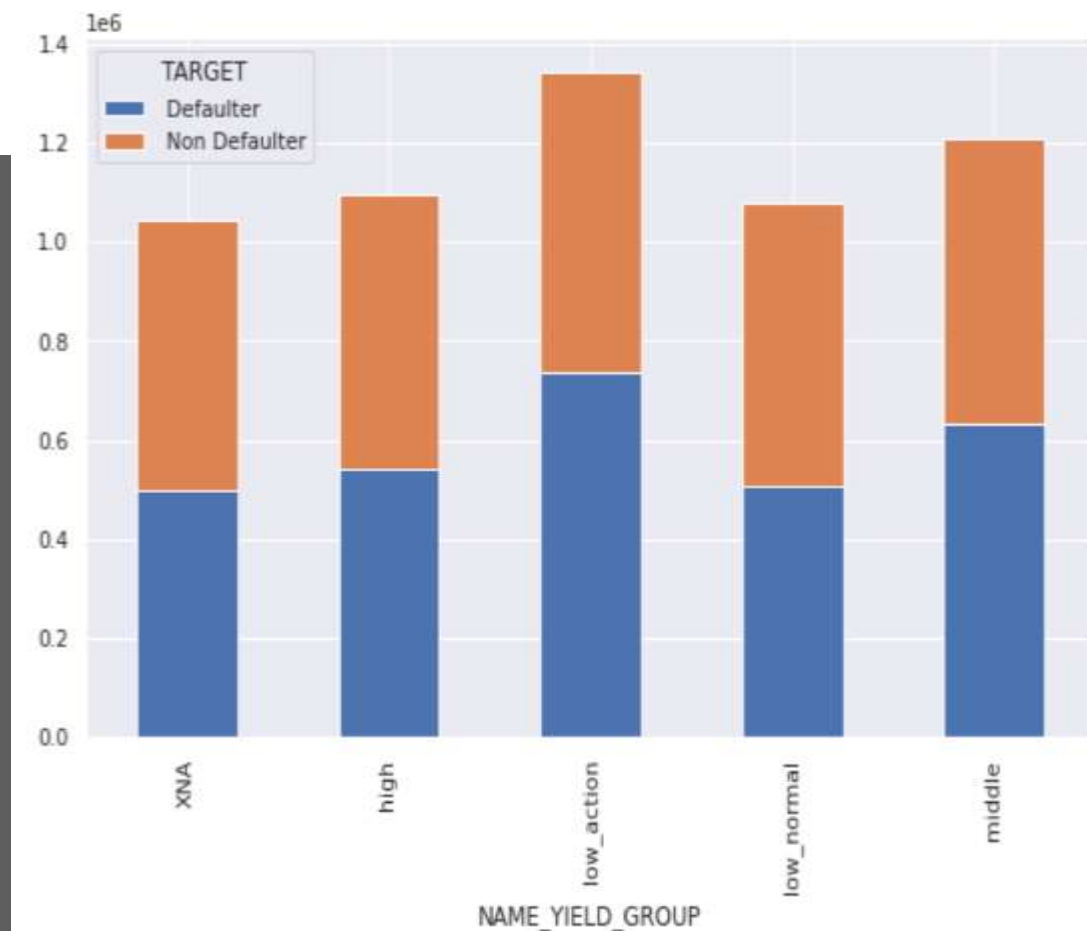


Out of those loans that are refused, only 0.1% were defaulters and 1.3% were non defaulters  
For those loans who were approved, only 0.4% were defaulters

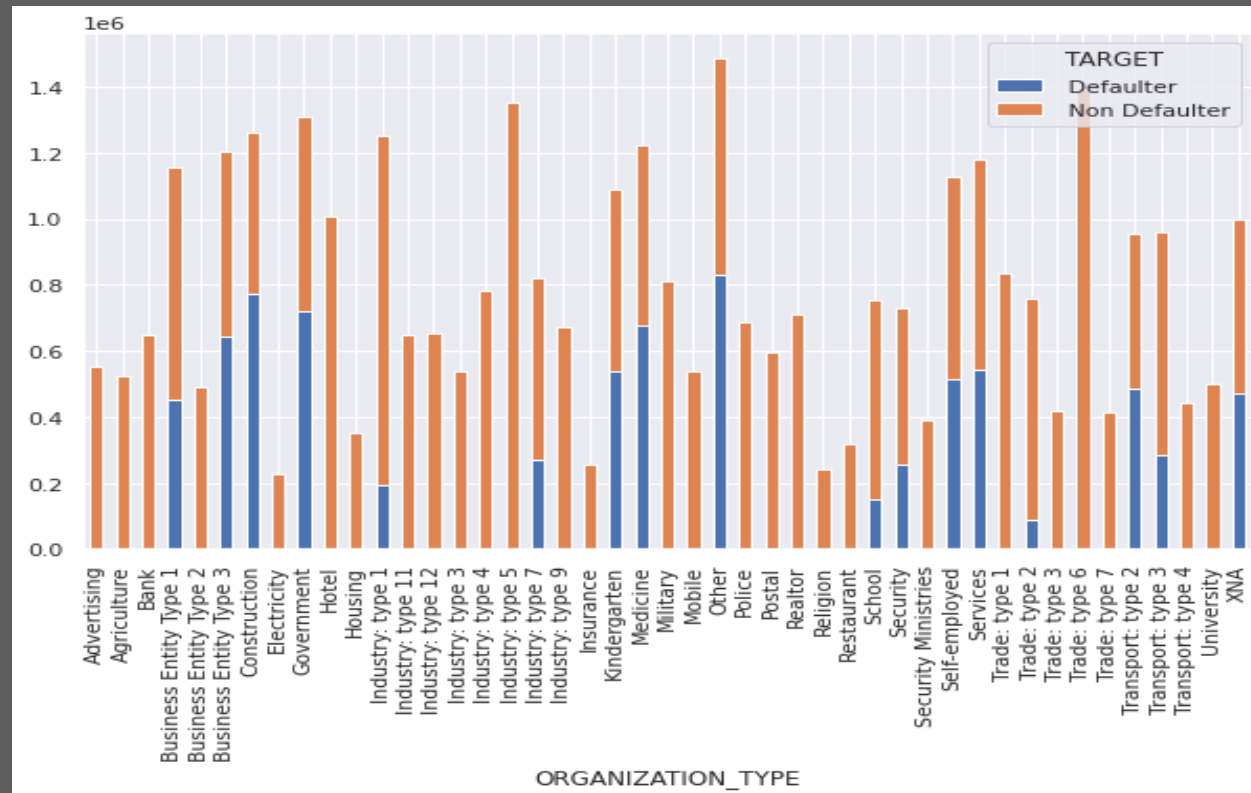
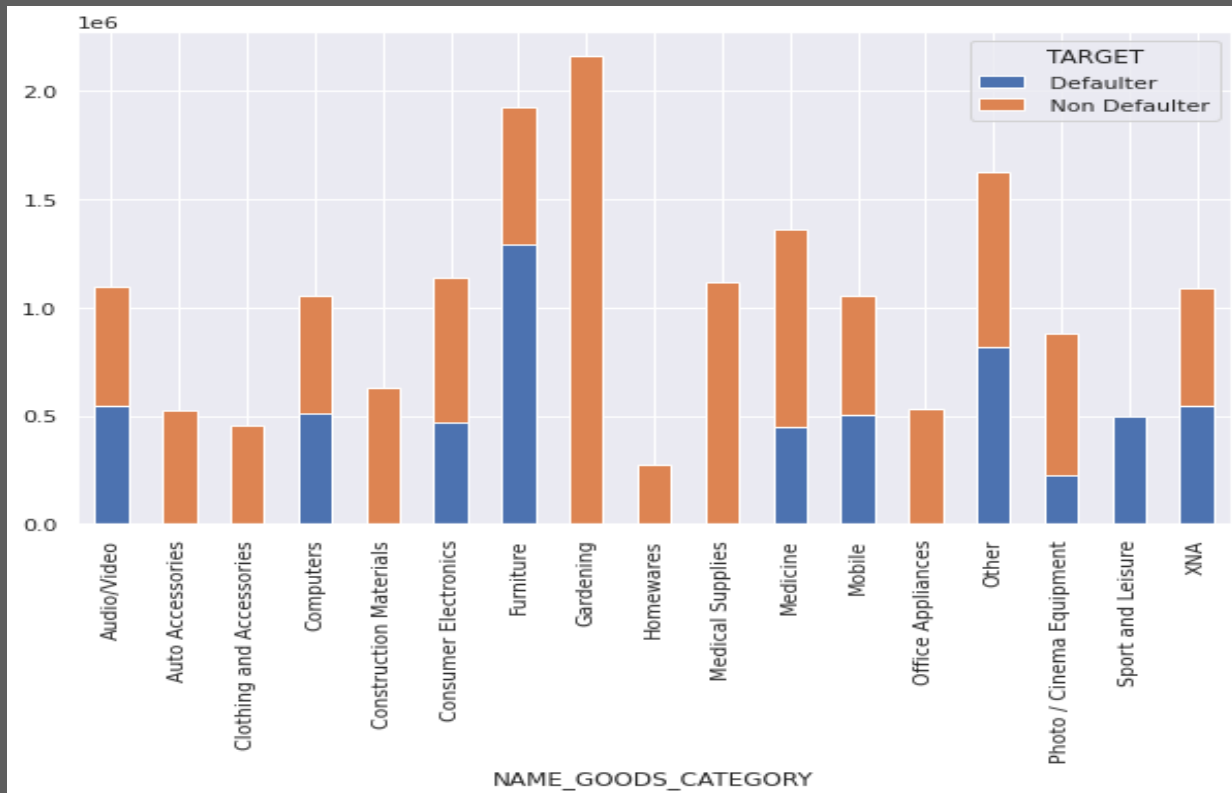
Currently defaulters mostly opted for 1.0 credit but tended to have worked for less 5 years



Currently Medicine staff, managers, sales staff, labours, accountants, cleaning staff, cooking staff, core staff, security staff, others and IT staff tended to default on the credit. Among IT staff, all of them have defaulted. \* No defaulters found in low\_skill labours, private service staff secretaries, waiters and barmen. These categories also opted for lower credit.

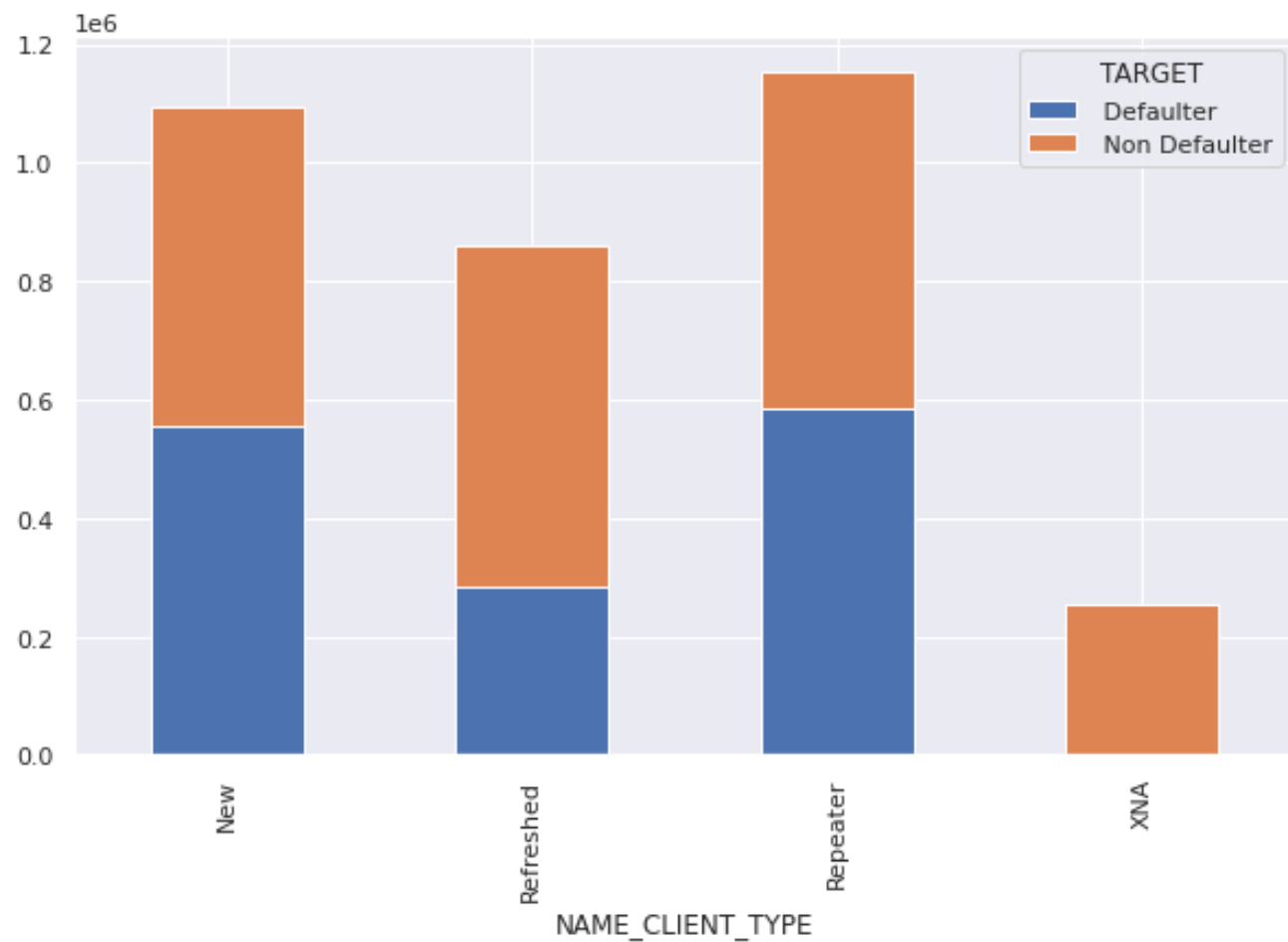


Those come under low action Interest rate Yield\_Group opted for more credit and also defaulted the most.

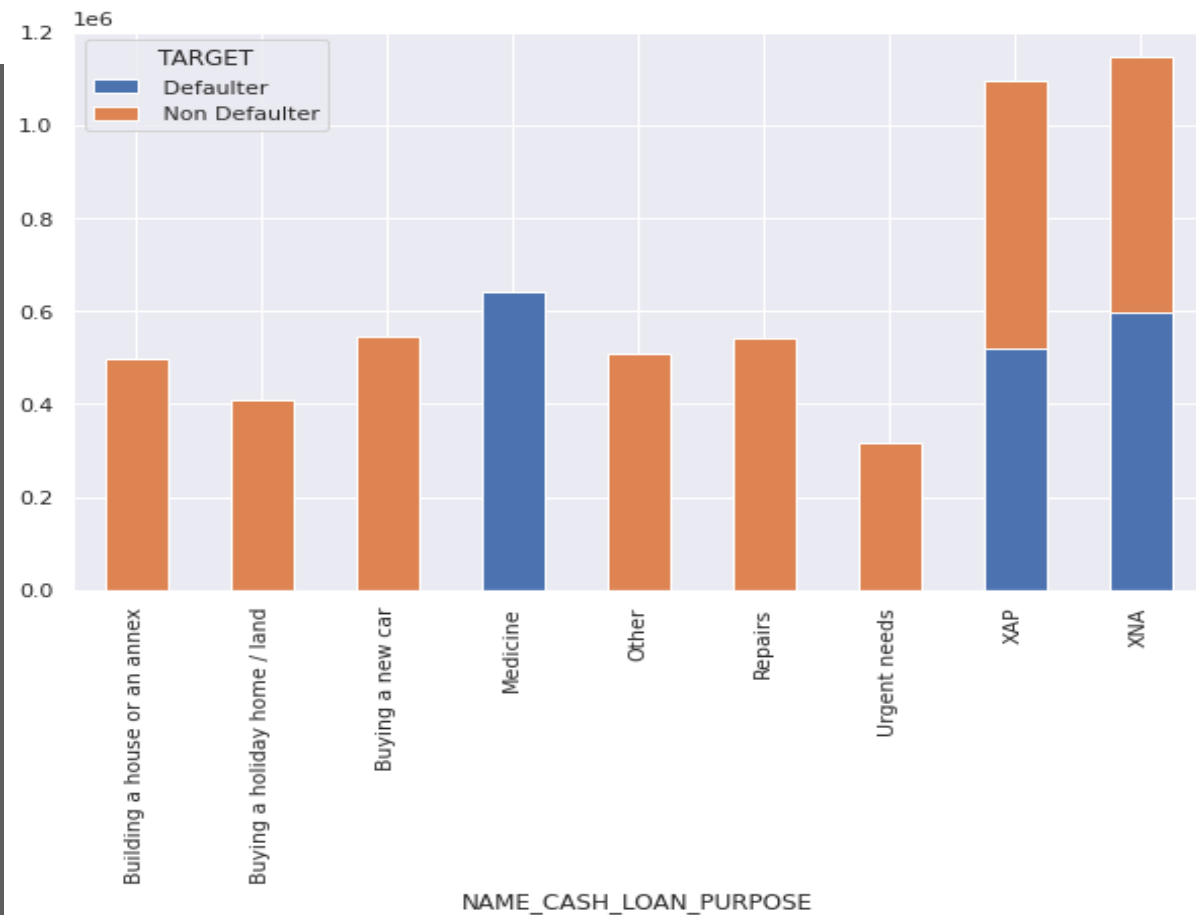


Currently those in XNA(unknown) group, medicine, mobile, sport and leisure, audio/video, computers, consumer electronics, furniture and photo/cinema equipment have defaulters. Furniture category have the highest amount of defaulters and they have taken the second highest amount of credit.

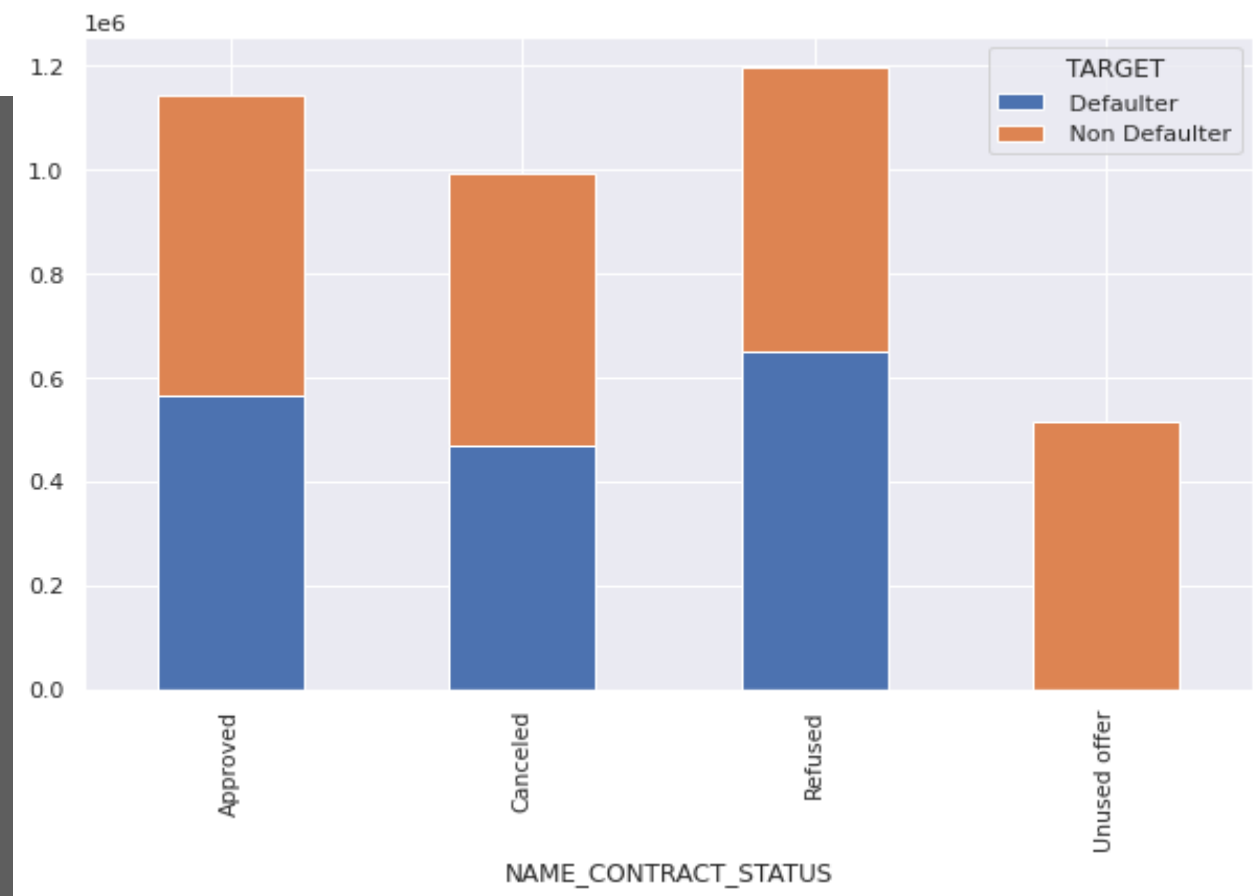
- Currently those in Medicine, Kindergarten, Other, Government, Construction, Business entity Type 3 and 1, transport type 2 and those falling in XNA defaulted heavily. These risky organizations were also granted higher amount of loans.



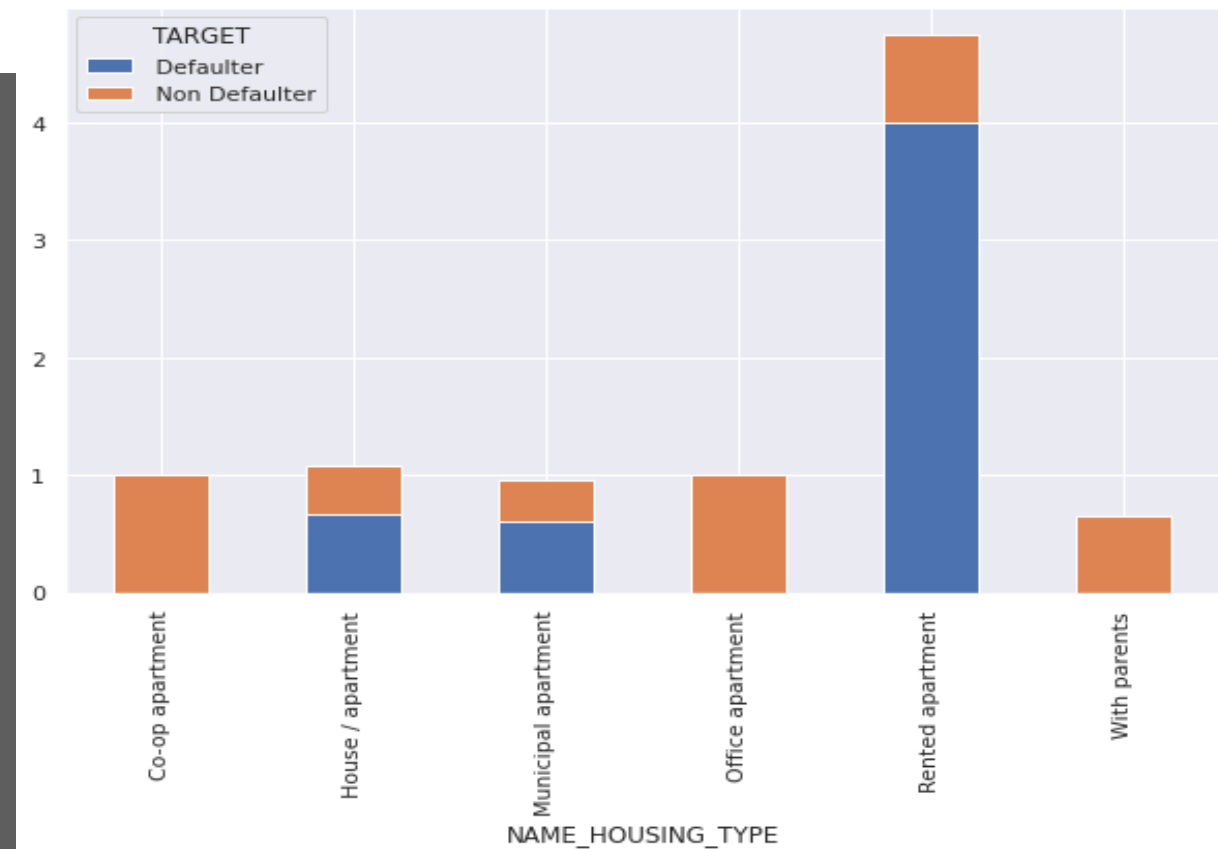
Currently Repeater clients and New clients take higher amount of credit and are more likely to default than refreshed clients.



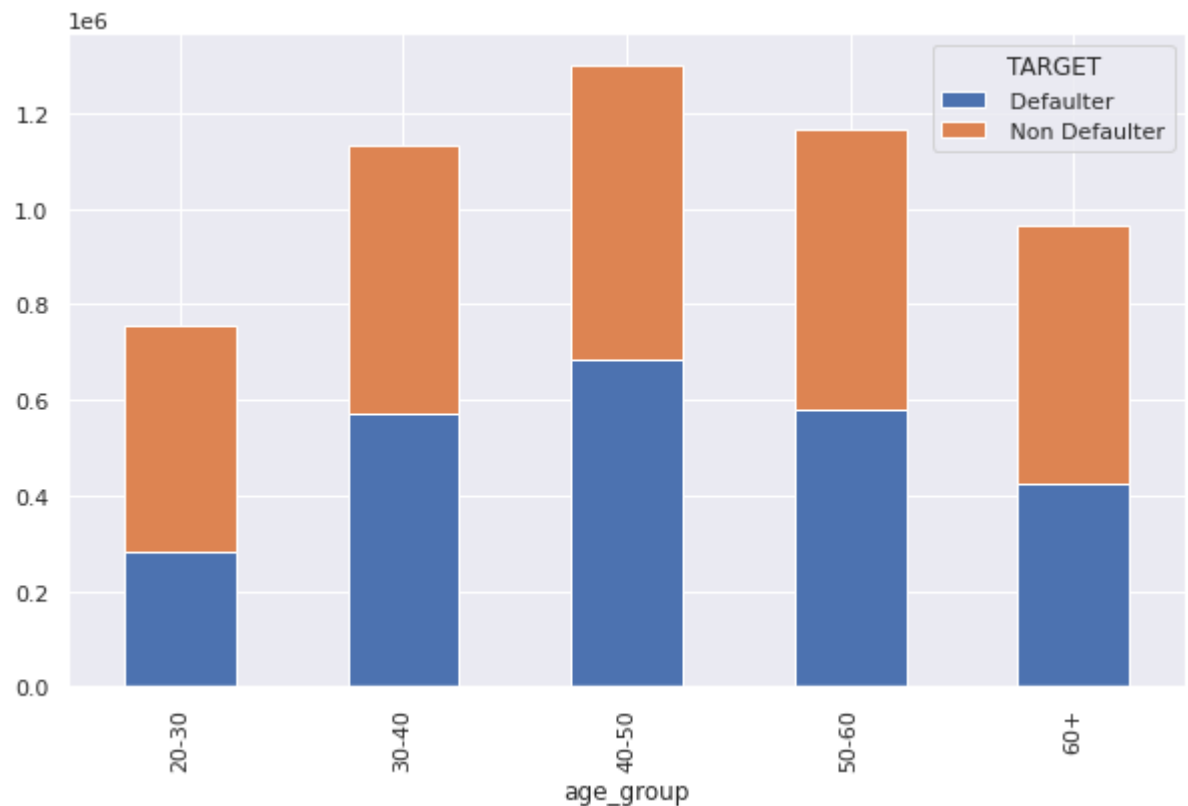
Currently those taking credit for Medicine, unknown purpose and miscellaneous purposes tend to default much more than other categories



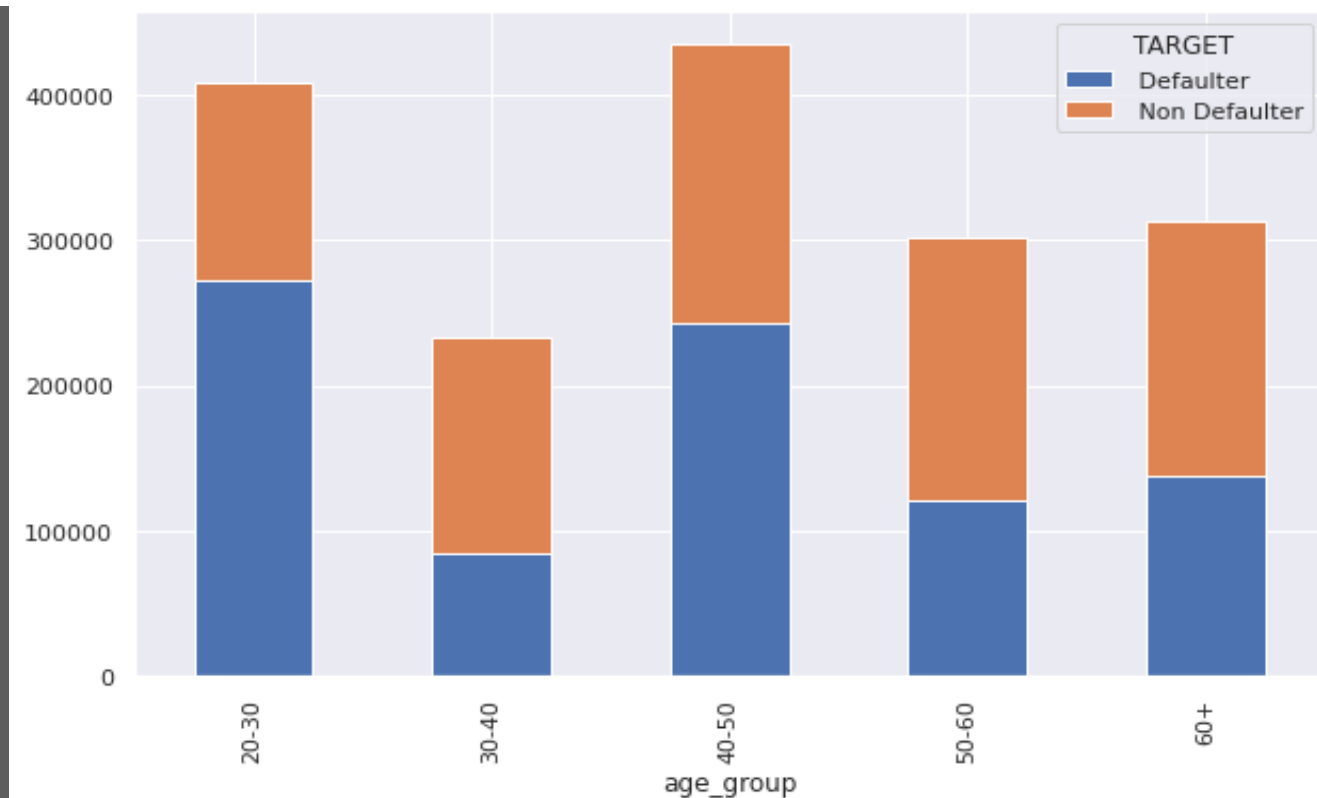
We can see that Currently many of the non defaulters have been refused and many defaulter have been approved. Both the categories, approved and refused opt for high credit amount



Currently those in rented apartments tend to have 4 children are most likely to default on the loans.



In the current case those in the age group 40-50 were more credit but has the highest number of defaulters. Those in the age group 20-30 and 60+ took less loans but have the least amount of defaulters



In the previous data, the 20-30 had the highest number of defaulters while 40-50 also defaulted heavily. The 60+ age group are still the less likely to default on loans as show in this graph

# RECOMENDATION

- The purpose of the loan should be checked in more details as certain organization and clients from certain occupation tend to default on the payment.
- Greater preference to be given to those have worked for more than 5 years.
- Loans should be offered more in the 60+ category. They are less likely to defaults on the loans.
- Clients with more than 2 children and are living in rented apartments are at risk of defaulting on the loans.
- New clients and Repeated clients are more in risk of defaulting. More loans should be offered to refreshed clients.
- Those come under low action Interest rate Yield\_Group opted for more credit and also defaulted the most.
- Avoid offering loans to those with secondary, lower secondary and incomplete education. More loans should be offered to those who have academic degree and higher education
- The not married/singles are likely to default on the loan. Less loans should be granted to them and more loans should be granted to widows.