Presentation on Credit EDA Case Study

Purpose

Credit risk analysis will help the company to make a decision for loan approval base on the applicant’s profile. Which controls loss of business to the company and avoid financial loss for the company.

STEPS

1. Dataunderstading and sourcing
2. Data quality check
3. Data check for outliers and imbalance in data.
4. Data analysis by univariate, segmented univariate, Bivariate analysis and correlation
5. Merging of application data with previous application data
6. Conclusion, Recommendations and Risks

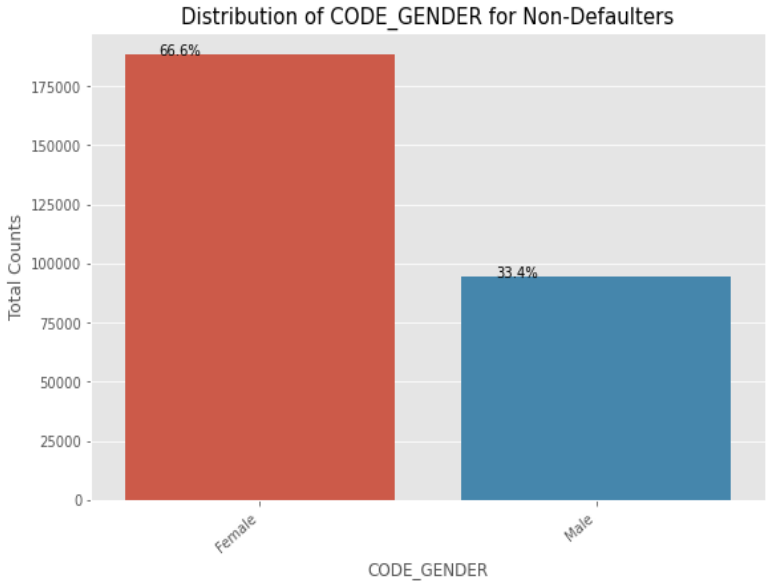
Data Analysis For Application Data

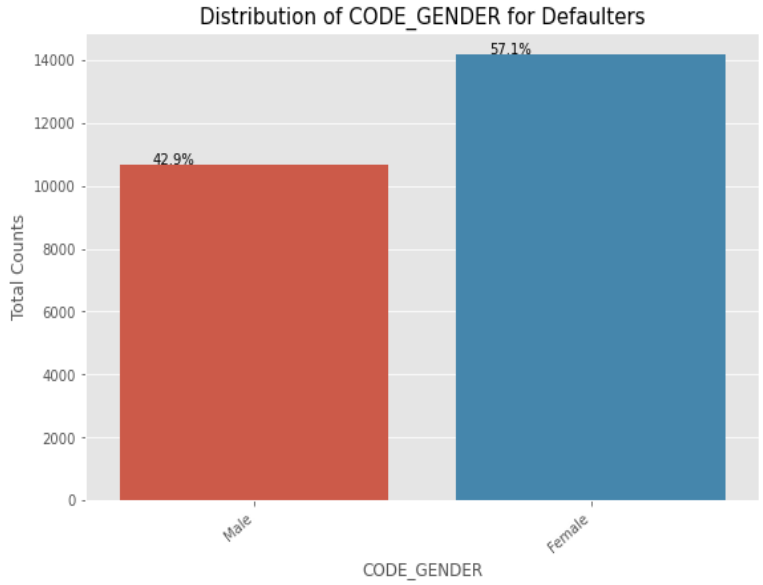
1.

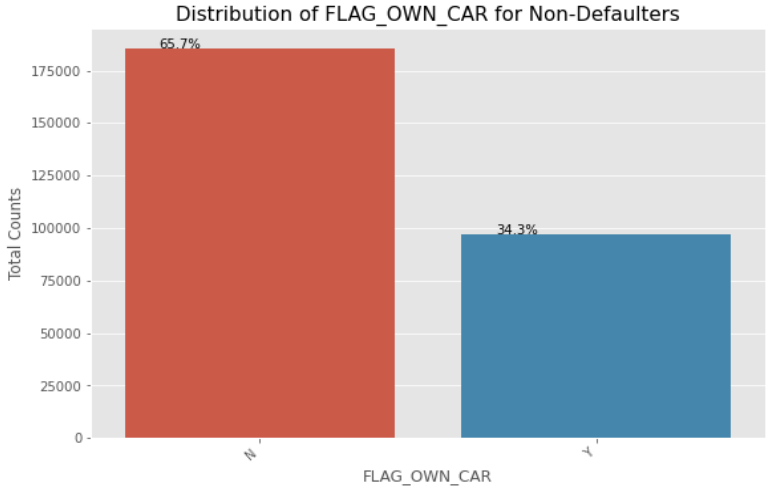
We can see that Female contribute 67% to the non-defaulters while 57% to the defaulters. We can conclude that We see more female applying for loans than males and hence the more number of female defaulters as well. But the rate of defaulting of FEMALE is much lower compared to their MALE counterparts.

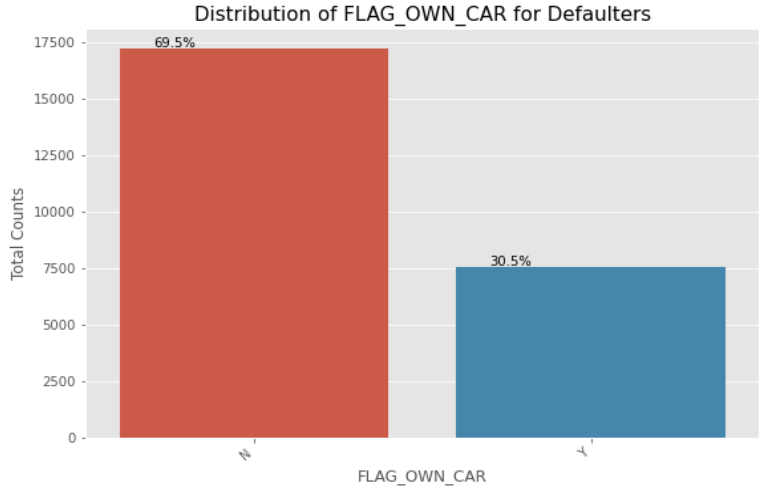
2.

We can see that people with cars contribute 65.7% to the non-defaulters while 69.5% to the defaulters. We can conclude that While people who have car default more often, the reason could be there are simply more people without cars 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 don't.

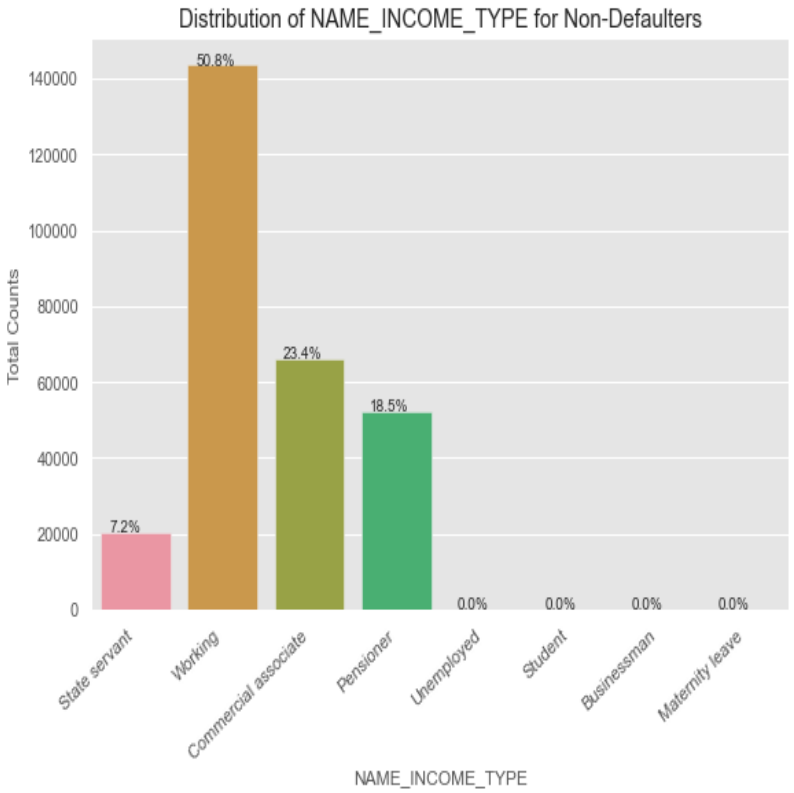


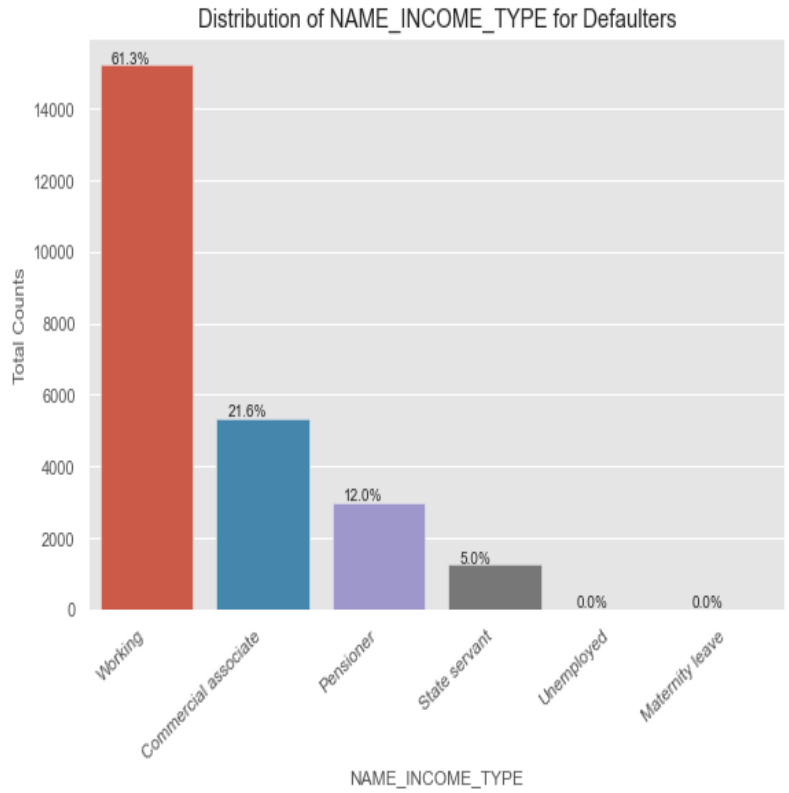






- We can notice that the students don't default. The reason could be they are not required to pay during the time they students.
- We can also see that the BusinessMen never default.
- Most of the loan are distributed to the working class people
- We also see that working class people contribute 51% to non defaulters while they contribute to 61% of the defaulters, Clearly, the chances of defaulting are more in their case.



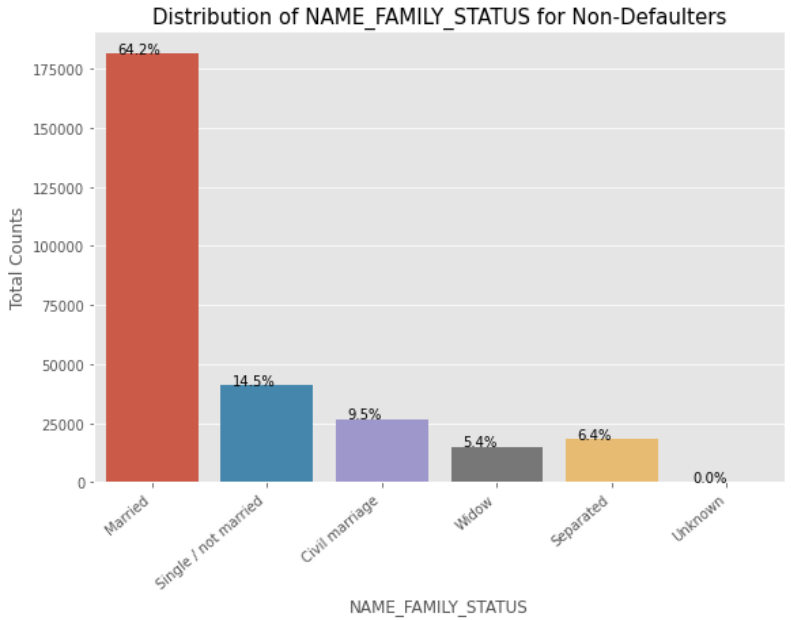


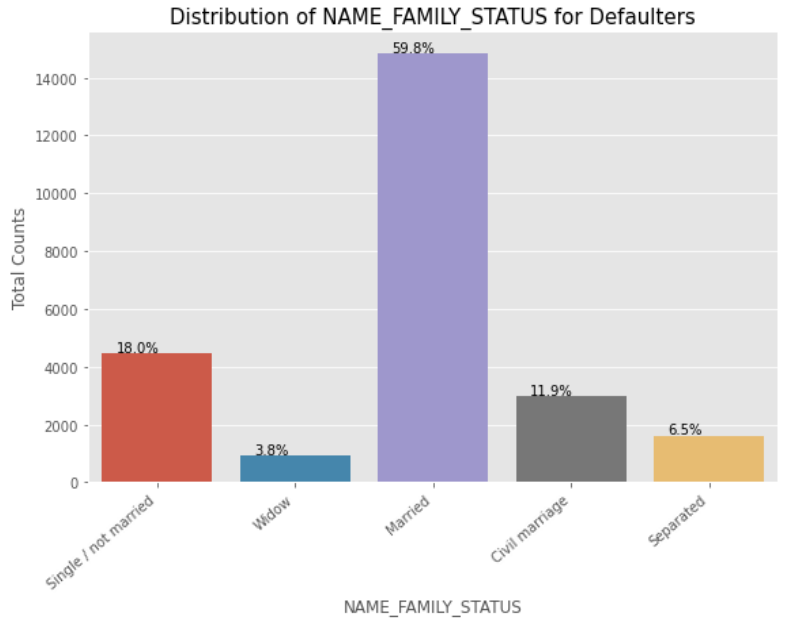
1.

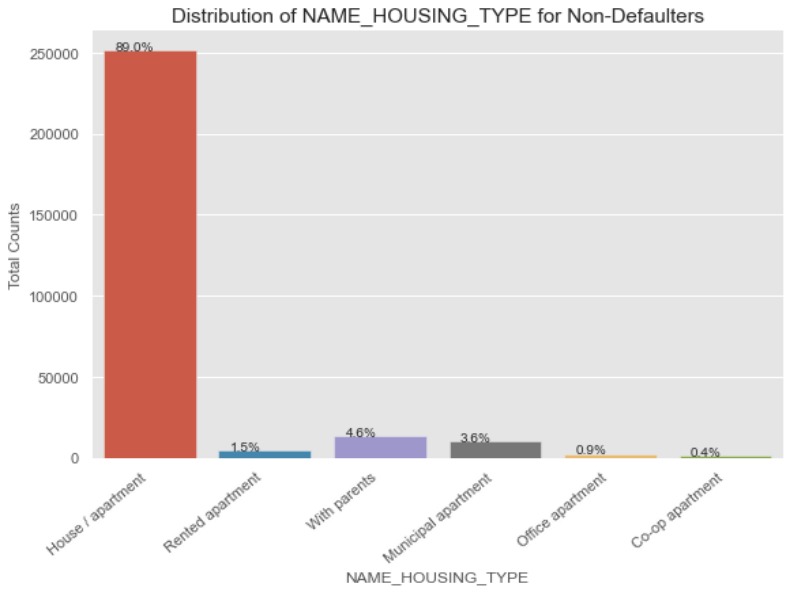
Married people tend to apply for more loans comparatively. But from the graph we see that Single/non Married people contribute 14.5% to Non Defaulters and 18% to the defaulters. So there is more risk associated with them.

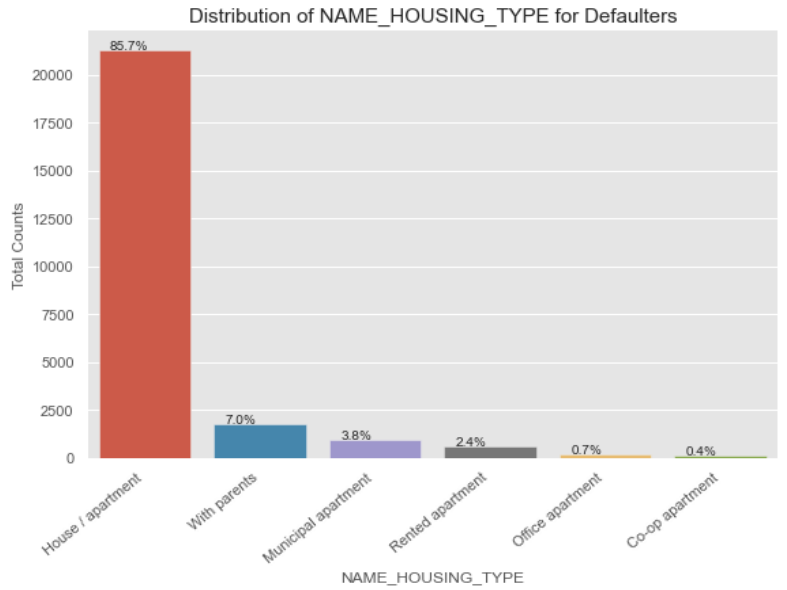
2.

It is clear from the graph that people who have House/Appartment, tend to apply for more loans. People living with parents tend to default more often when compared with others.The reason could be their living expenses are more due to their parents living with them.









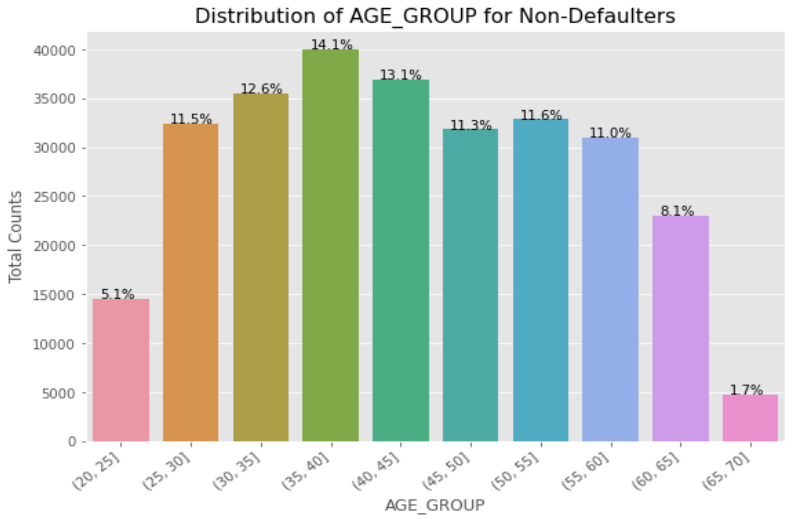
Univariate Categorical Ordered Analysis

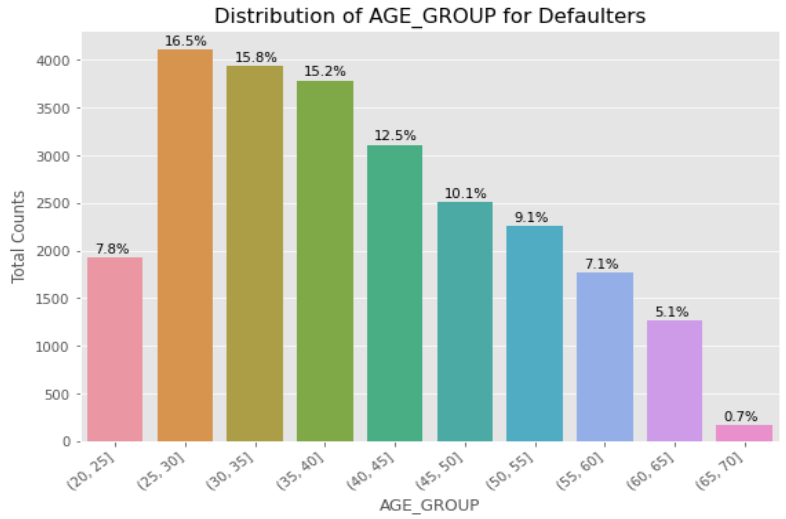
1.

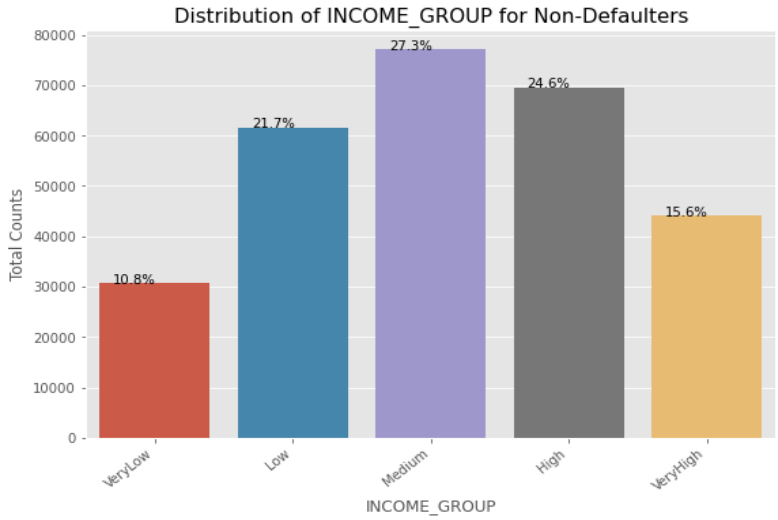
We see that (25,30] age group tend to default more often. So they are the riskiest people to loan to. With increasing age group, people tend to default less starting from the age 25. One of the reasons could be they get employed around that age and with increasing age, their salary also increases.

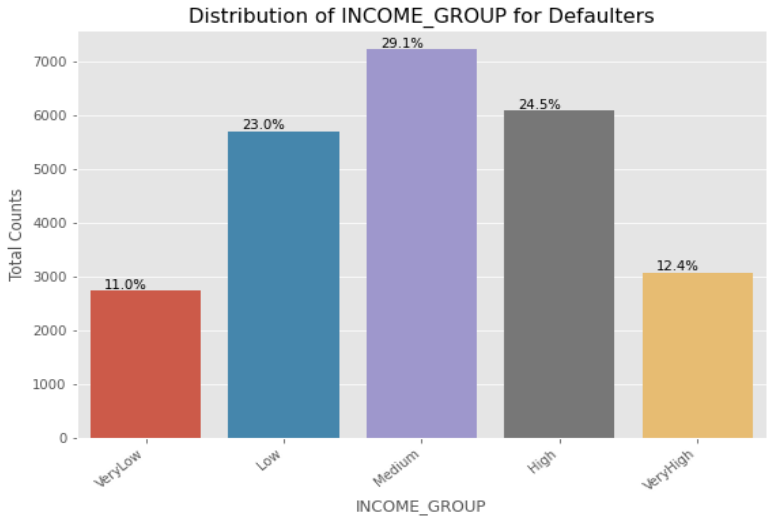
2.

The Very High income group tend to default less often. They contribute 12.4% to the total number of defaulters, while they contribute 15.6% to the Non-Defaulters.







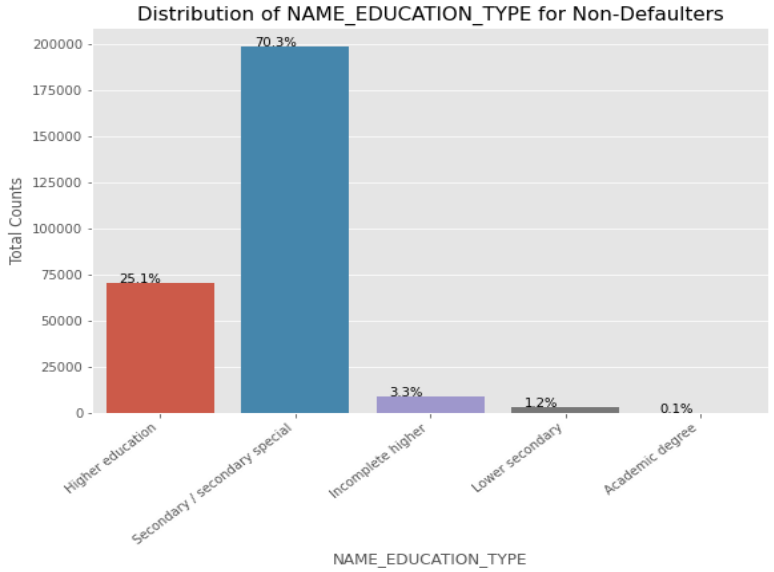


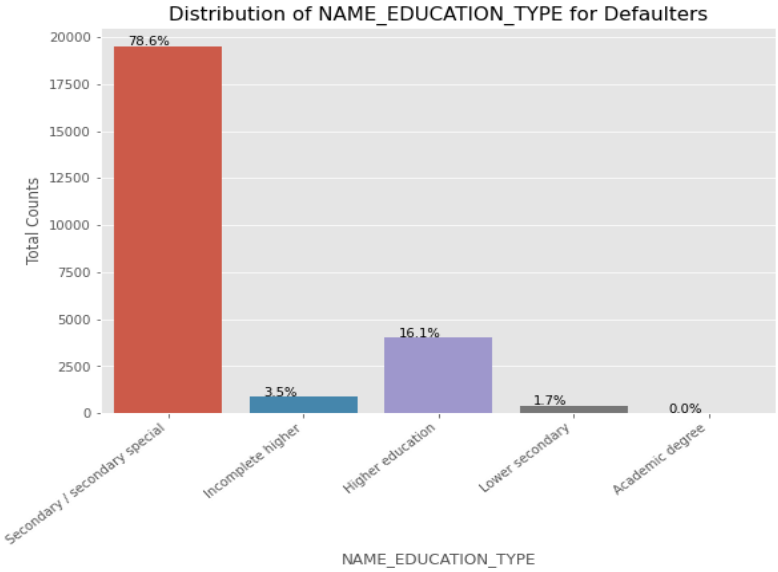
1.

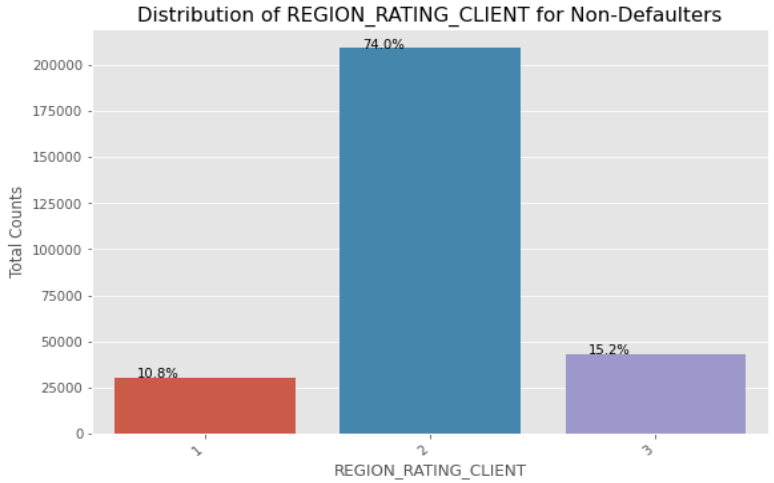
Almost all of the Education categories are equally likely to default except for the higher educated ones who are less likely to default and secondary educated people are more likely to default

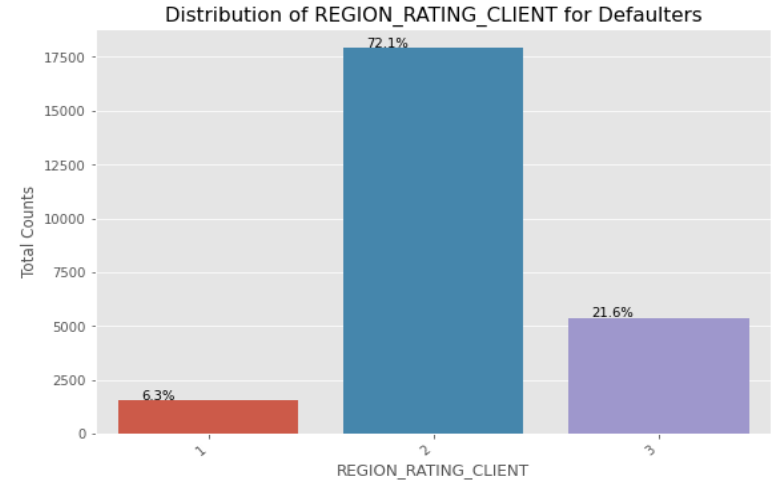
2.

More people from second tier regions tend to apply for loans. We can infer that people living in better areas(Rating 3) tend contribute more to the defaulters by their weightage. People living in 1 rated areas



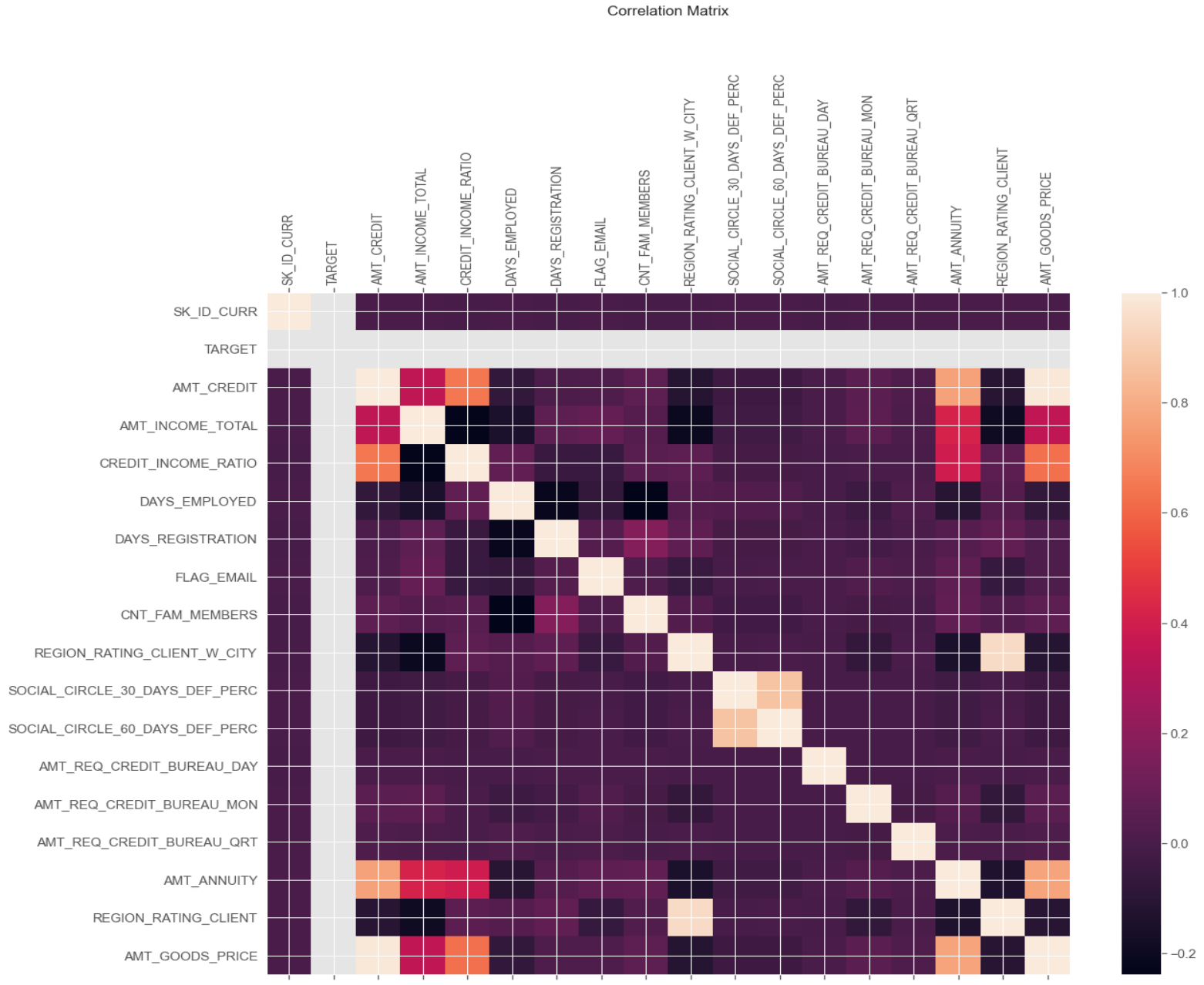




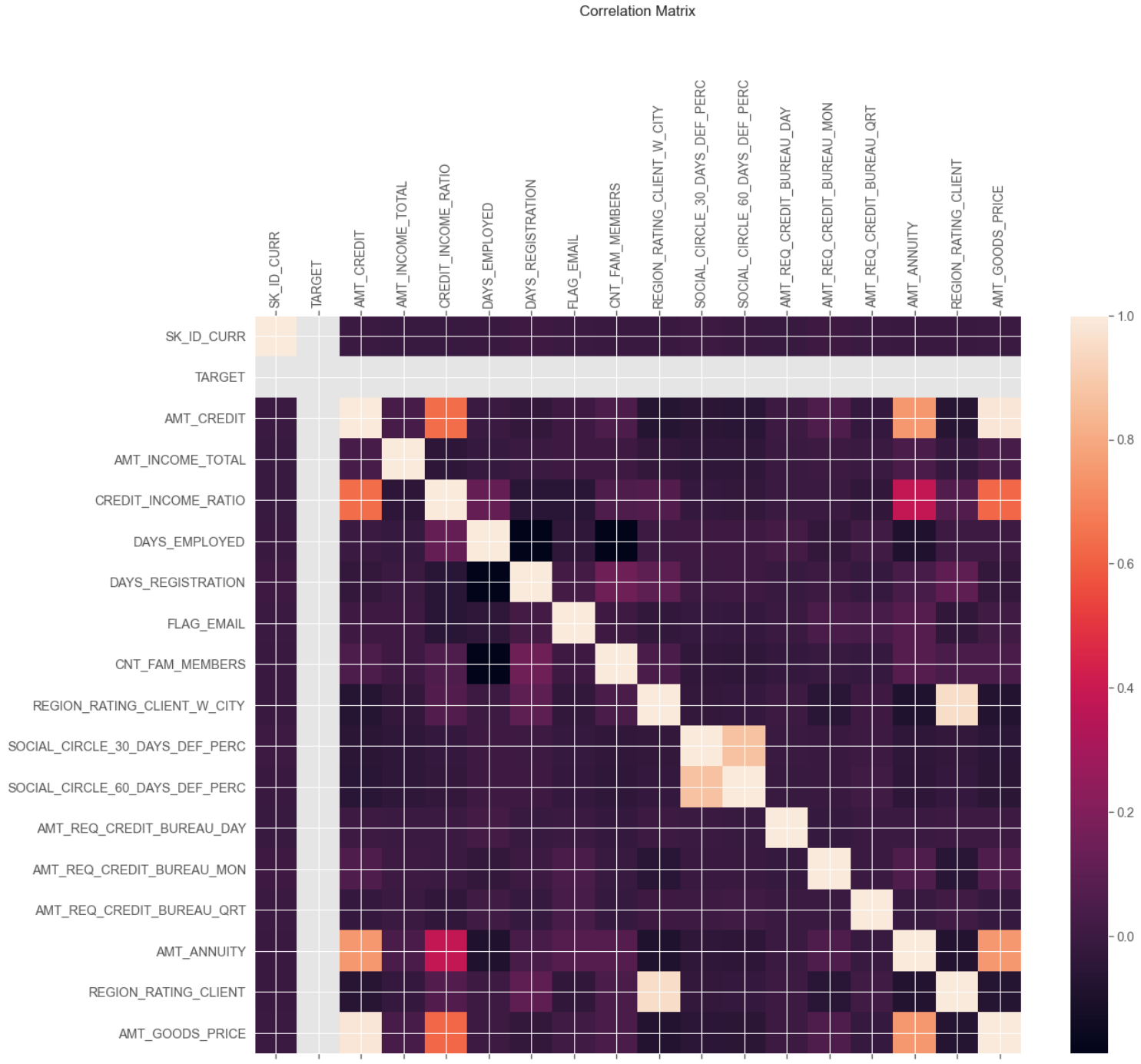


Correlation analysis for Target 0 and Target 1

This represent Correlation for Target 0



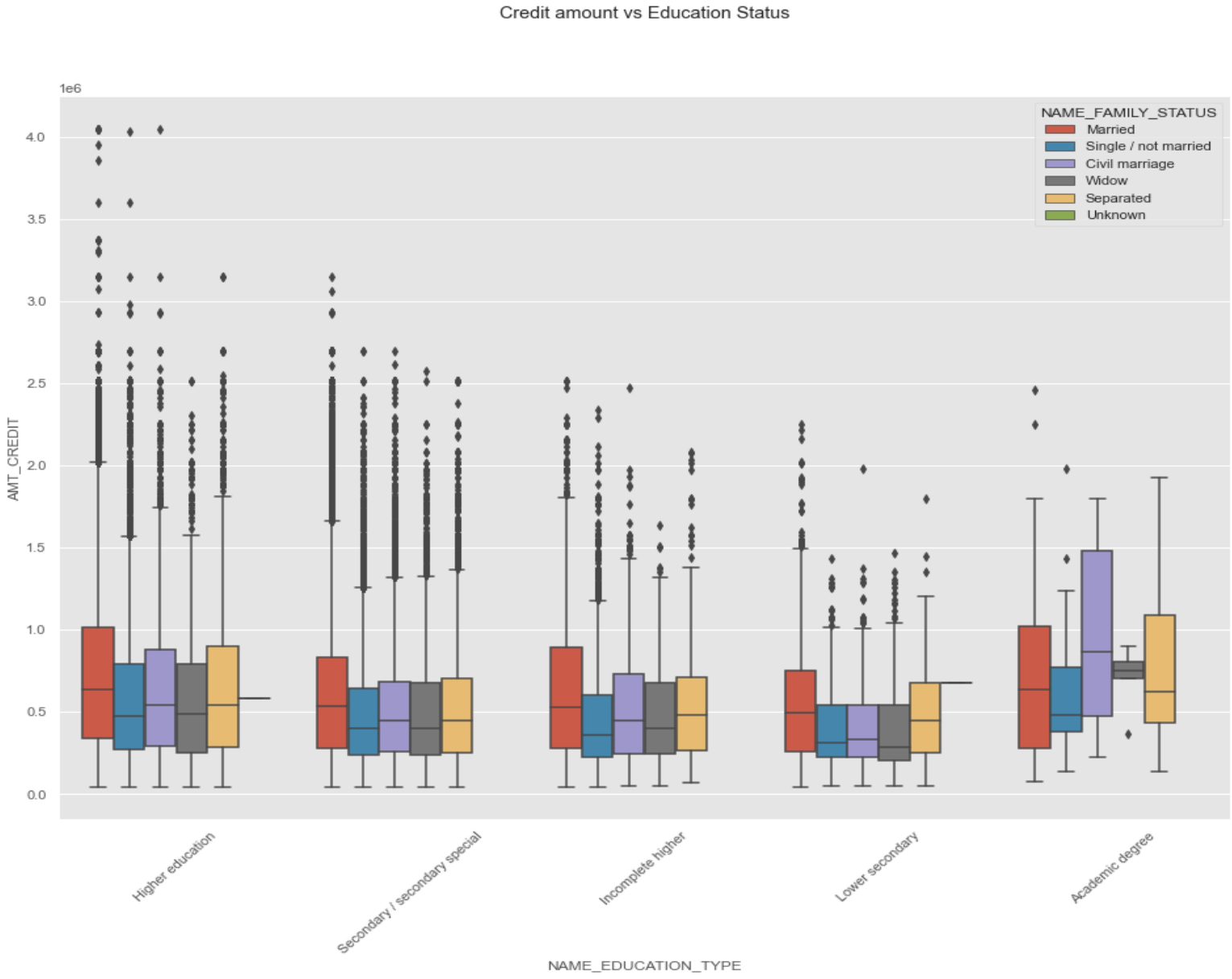
This represent correlation for Target 1



For Target 0

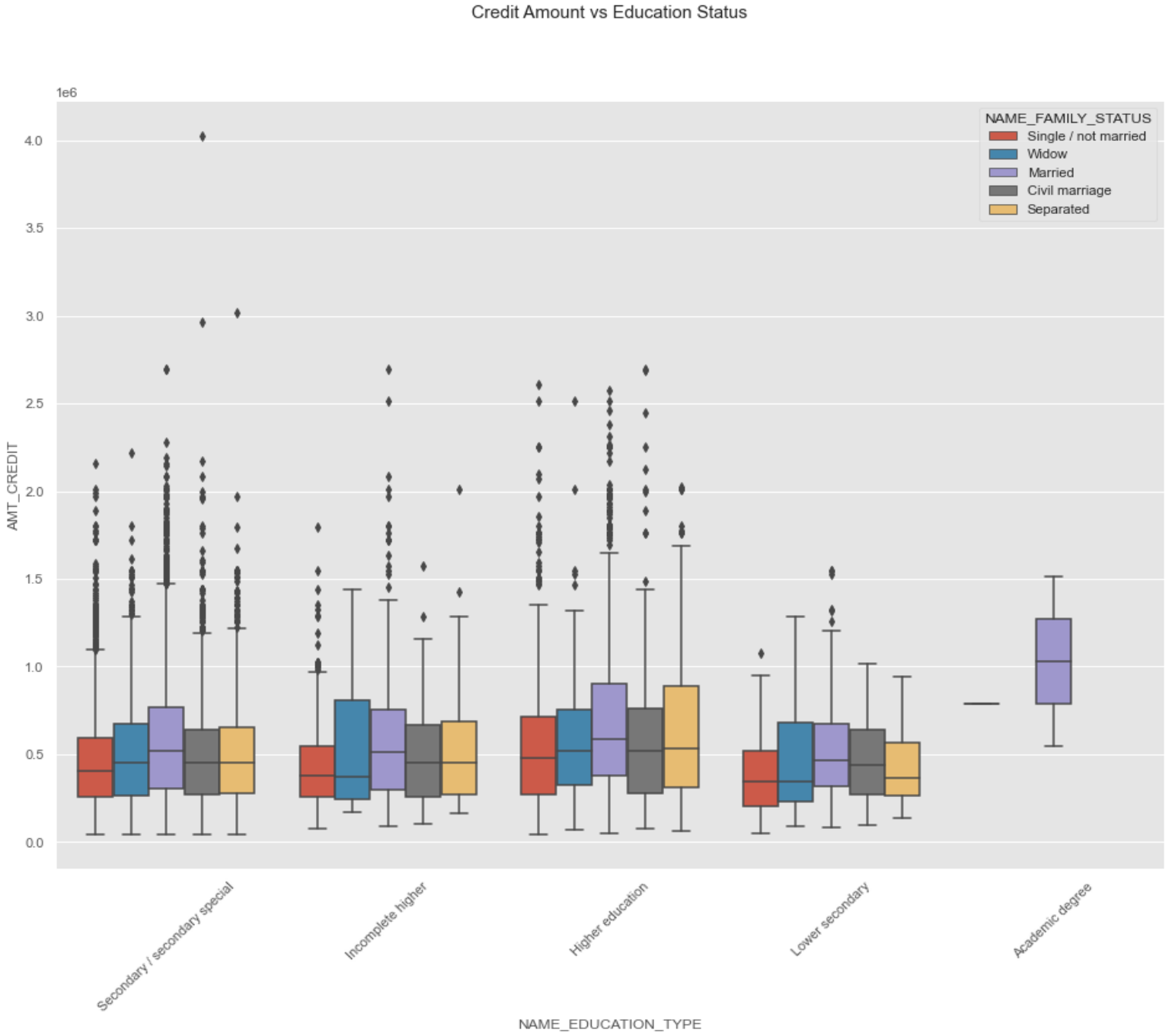
From the above box plot we can conclude that Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others. Also, higher education of family status of 'marriage', 'single' and 'civil marriage' are having more outliers. Civil marriage for Academic degree is having most of the credits in the third quartile

Bivariate Analysis of numerical variables



For Target 1

Quite similar with Target 0
From the above box plot we can say that Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others. Most of the outliers are from Education type 'Higher education' and 'Secondary'. Civil marriage for Academic degree is having most of the credits in the third quartile.

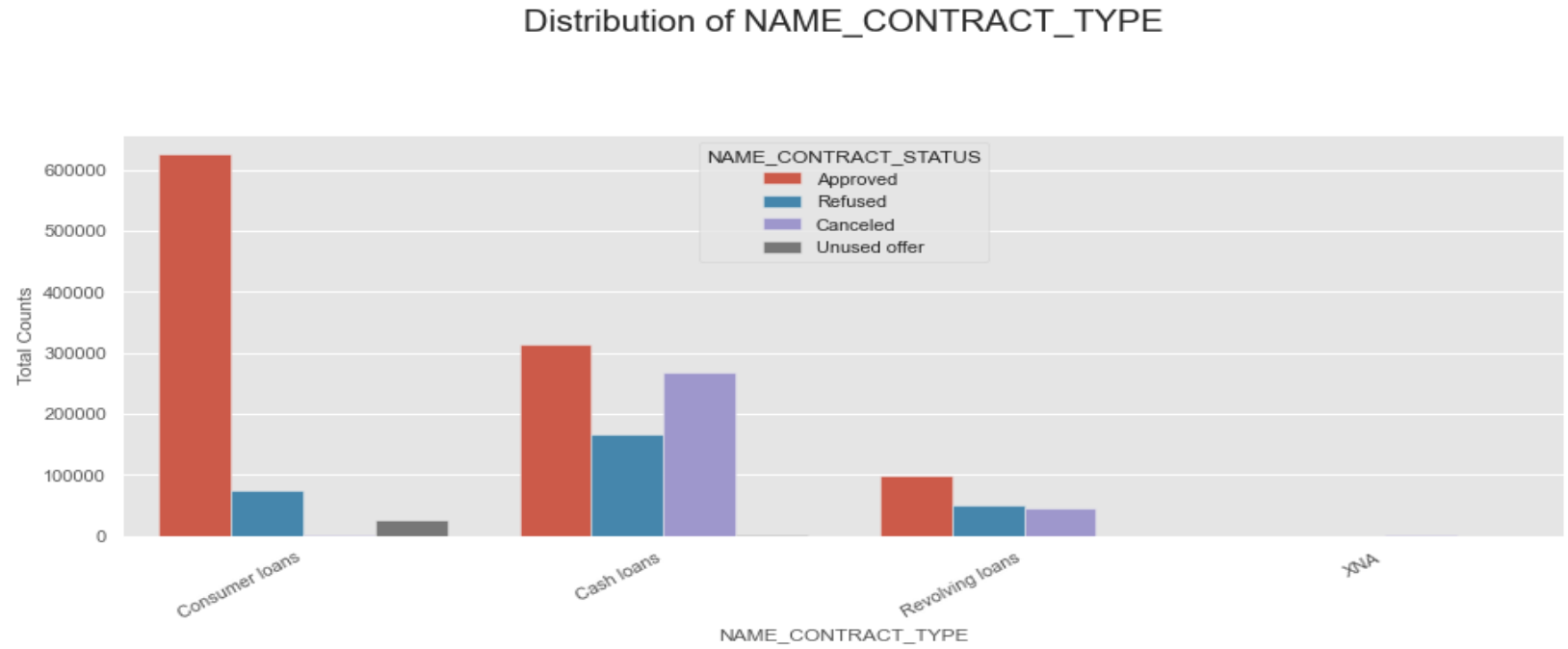


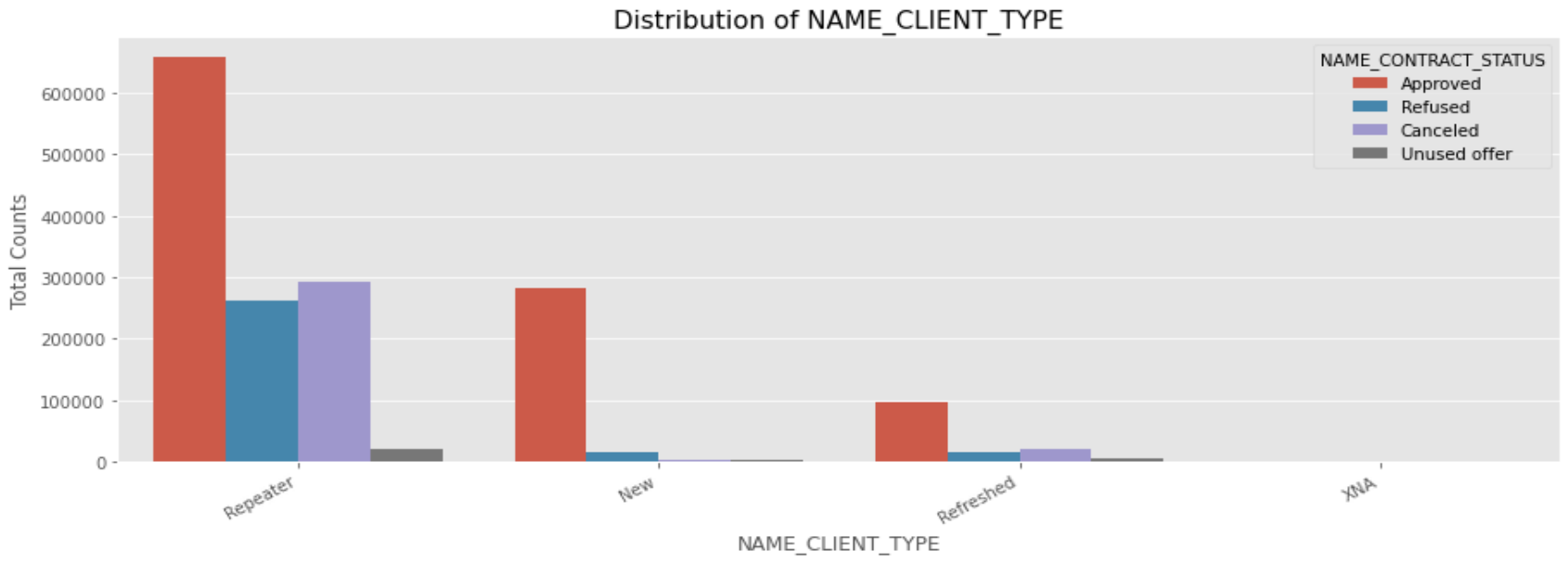
Data Analysis For Previous Application Data

1.From the above chart, we can infer that, most of the applications are for 'Cash loan' and 'Consumer loan'. Although the cash loans are refused more often than others.

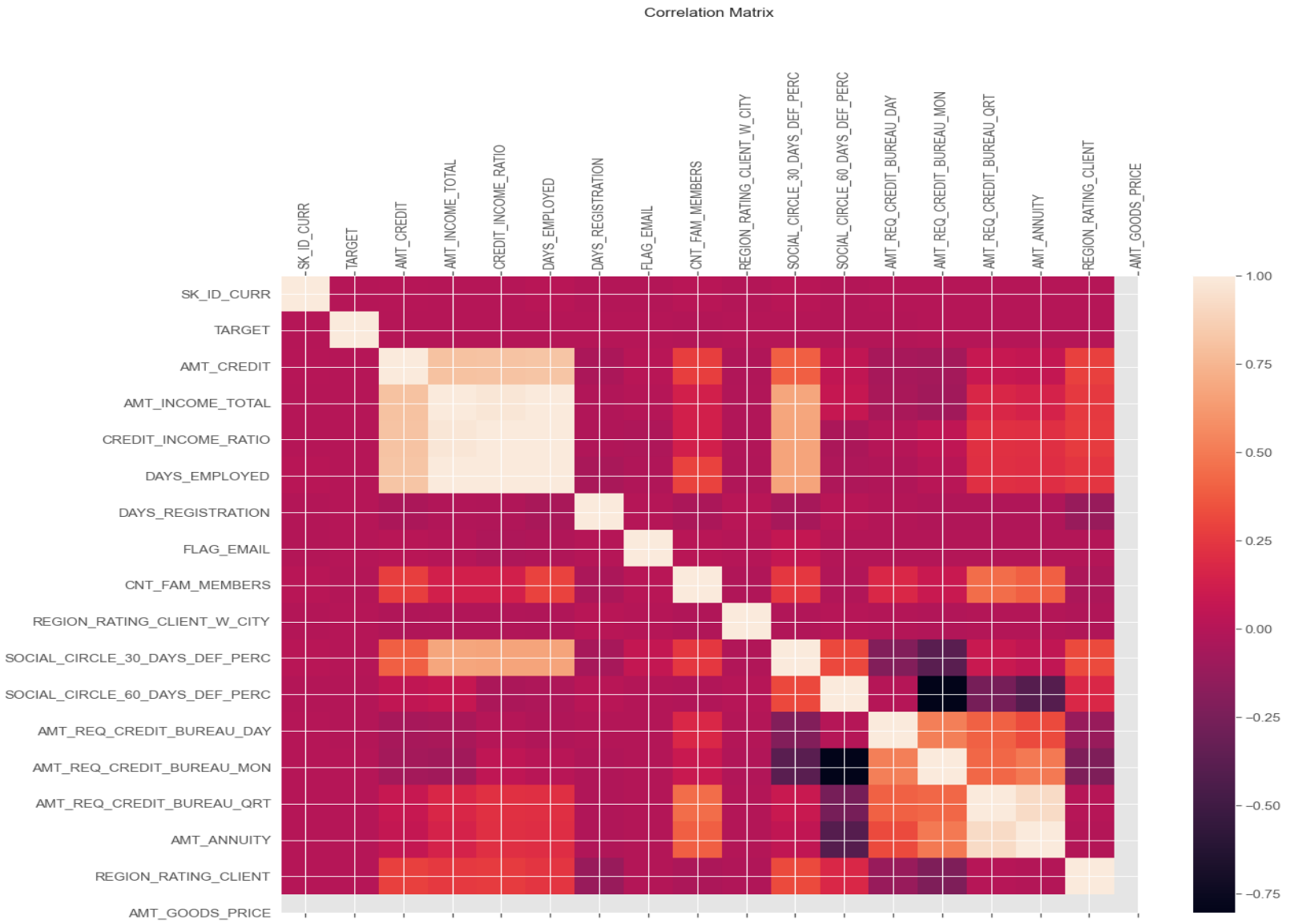
2. Most of the loan applications are from repeat customers, out of the total applications 70% of customers are repeaters. They also get refused most often.

Univariate analysis





Checking the correlation in the PreviousApplication dataset



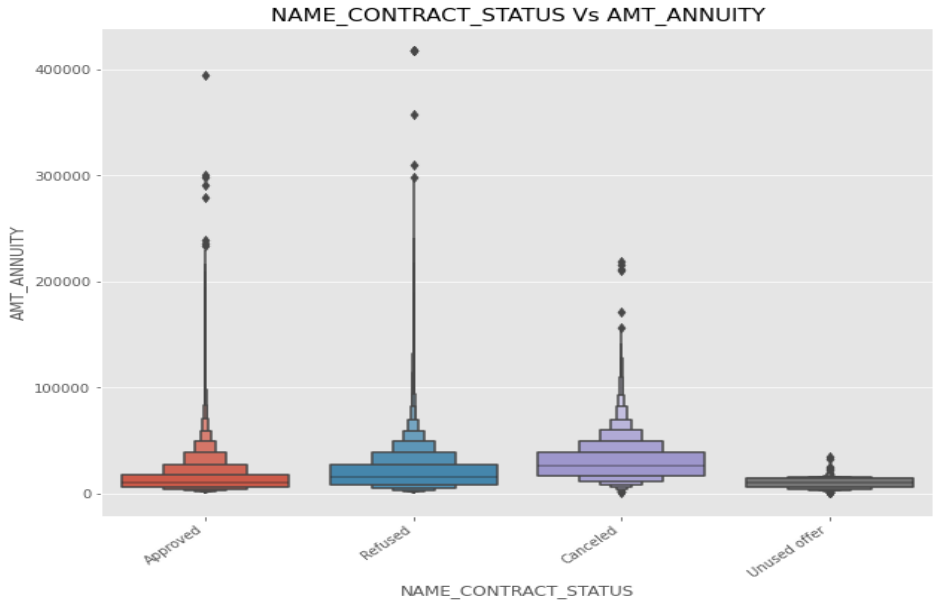
Using box plot to do bivariate analysis on categorical vs numeric columns

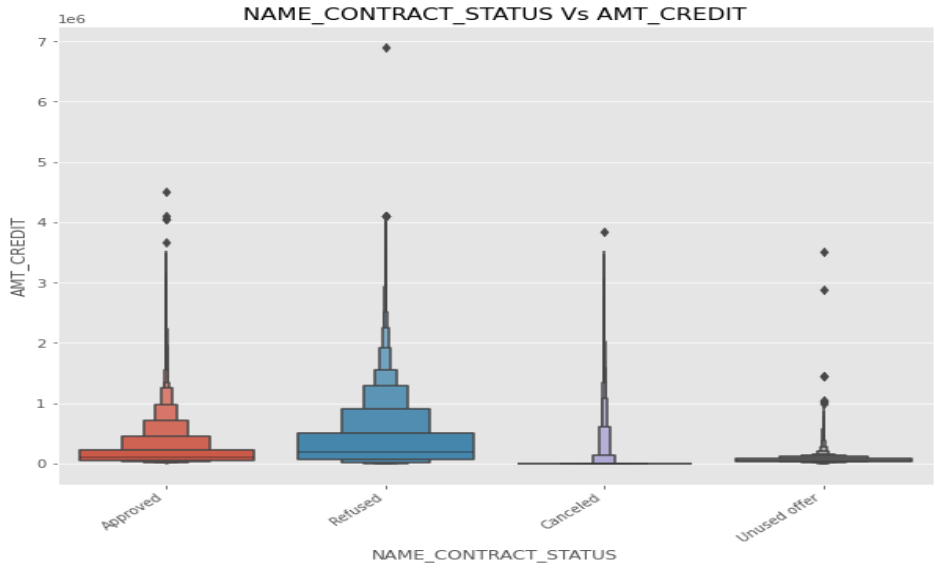
1.

From the above plot we can see that loan application for people with lower AMT_ANNUITY gets canceled or Unused most of the time. We also see that applications with too high AMT ANNUITY also got refused more often than others.

2.

We can infer that when the AMT_CREDIT is too low, it get's cancelled/unused most of the time.

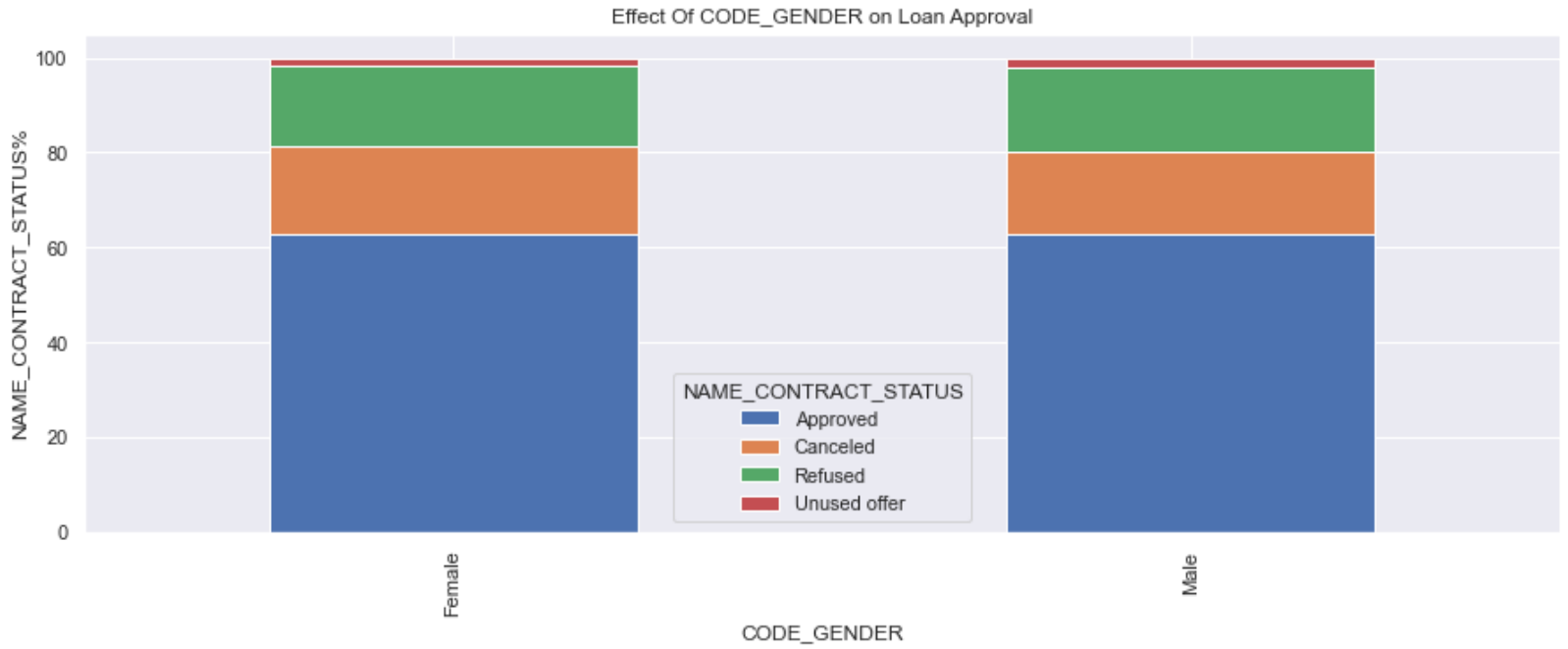




Merging the files and analyzing the data

We see that car ownership doesn't have any effect on application approval or rejection. But we saw earlier that the people who has a car has lesser chances of default. The bank can add more weightage to car ownership while approving a loan amount

1. We see that code gender doesn't have any effect on application approval or rejection. But we saw earlier that female have lesser chances of default compared to males. The bank can add more weightage to female while approving a loan amount
2. We can see that the people who were approved for a loan earlier, defaulted less often where as people who were refused a loan earlier have higher chances of defaulting.



CONCLUSION

1. Banks should focus more on contract type ‘Student’ ,’pensioner’ and ‘Businessman’ with housing ‘type other than ‘Co-op apartment’ for successful payments.
2. Banks should focus less on income type ‘Working’ as they are having most number of unsuccessful payments.
3. Also with loan purpose ‘Repair’ is having higher number of unsuccessful payments on time.
4. Get as much as clients from housing type ‘With parents’ as they are having least number of unsuccessful payments