
Credit EDA Assignment

— - Sai Venkatesh Malakala —

Agenda

1. Problem understanding
2. Solution Approach
3. Results of Univariate analysis
4. Results of Bivariate & Multivariate analysis
5. Conclusion

1. Problem Understanding

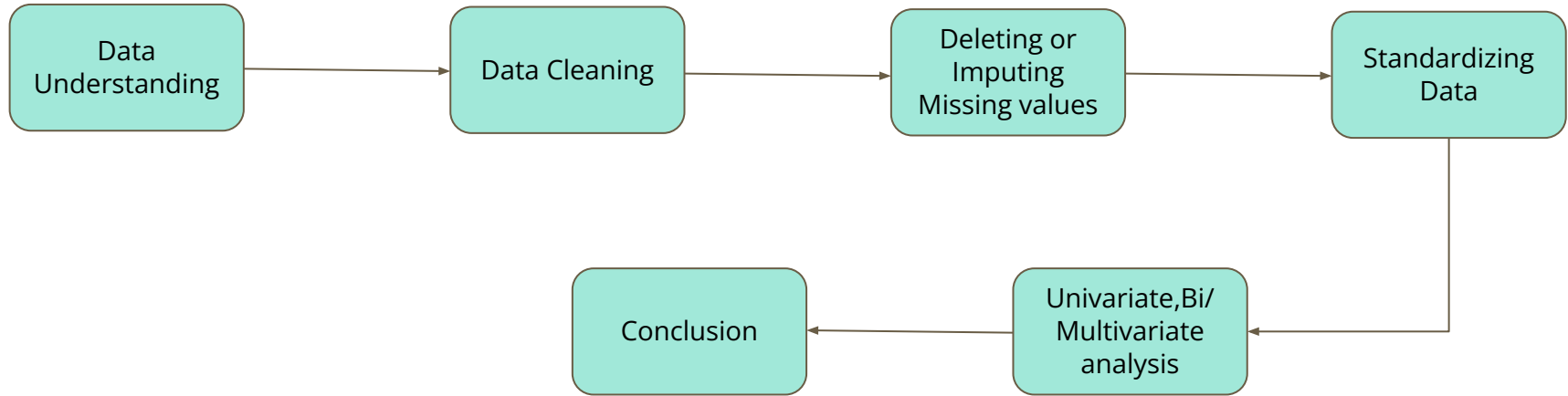
Current State

1. Client is a loan provider company & is facing difficulty in approving loans to the customers due to insufficient or non-existent credit history of the Customers
2. When the client receives a loan application, he has to decide for loan approval based on the applicant's profile. we can classify the loan applicants into 2 categories:
 - a. applicant is likely to repay the loan
 - b. applicant is not likely to repay the loan
3. Client is expecting a Risk assessment report based on the Present & Previous applicants data

Desired State

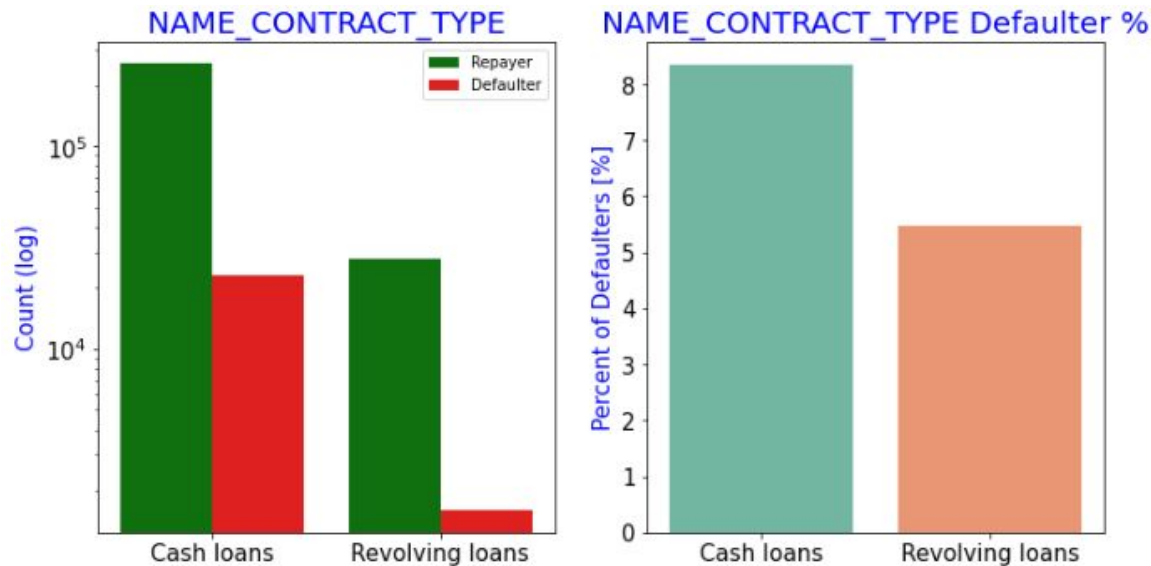
1. To identify patterns which indicate if a customer has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.
2. To understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default through Credit Risk Analytics Report

2. Solution Approach



3.1 Results of Univariate analysis

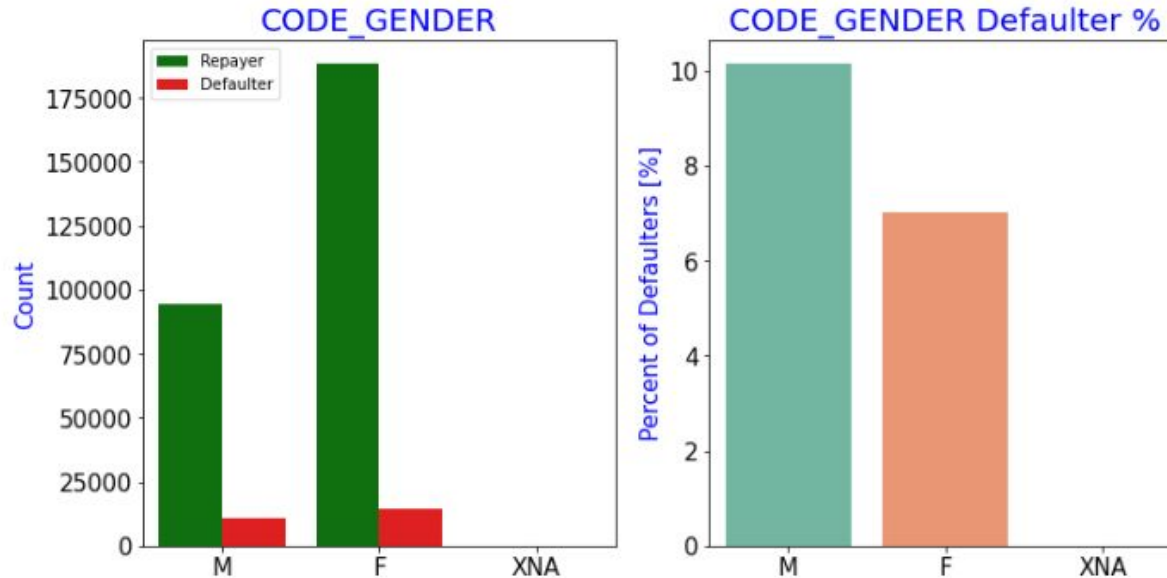
Variable:1 NAME_CONTRACT_TYPE



Conclusion: Revolving loans has more percentage of defaulters compared to the Cash loans

3.2 Results of Univariate analysis

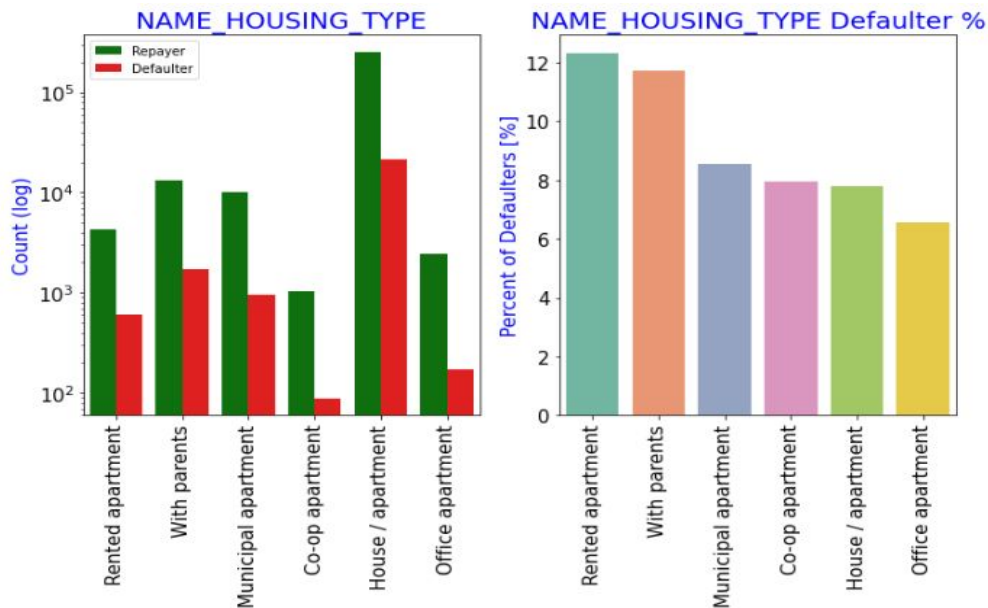
Variable:2 CODE_GENDER



Conclusion: Male has high tendency to be defaulters compared to Females, We can infer that from the above Bar chart.

3.3 Results of Univariate analysis

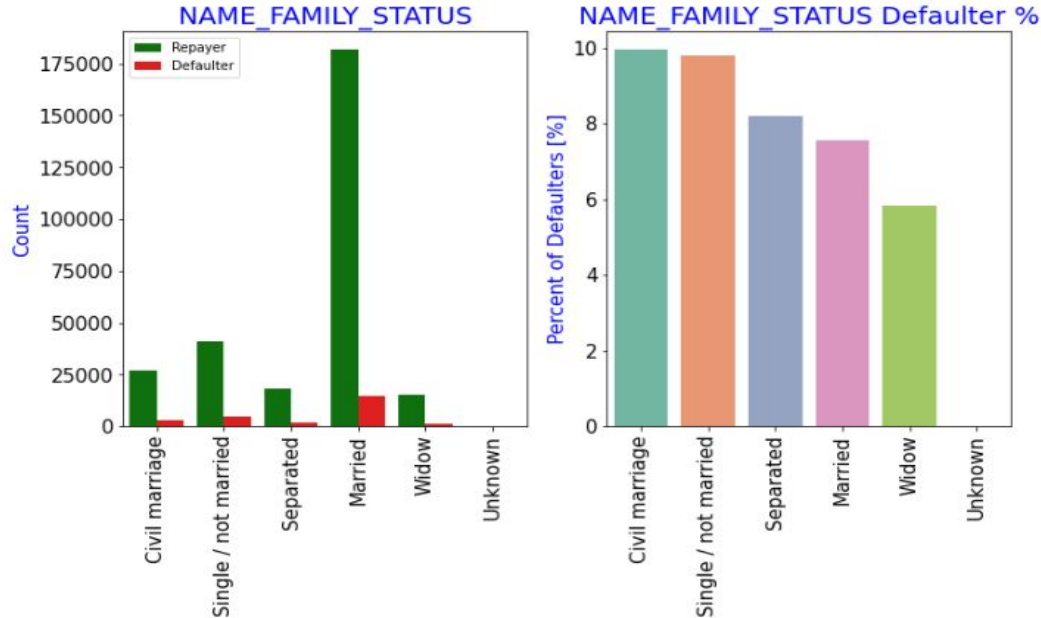
Variable:3 NAME_HOSING_TYPE



Conclusion: Majority of people live in House/apartment. People living in office apartments have lowest default rate & People living with parents (around 11.5%) and living in rented apartments(> 12%) have higher probability of defaulting

3.4 Results of Univariate analysis

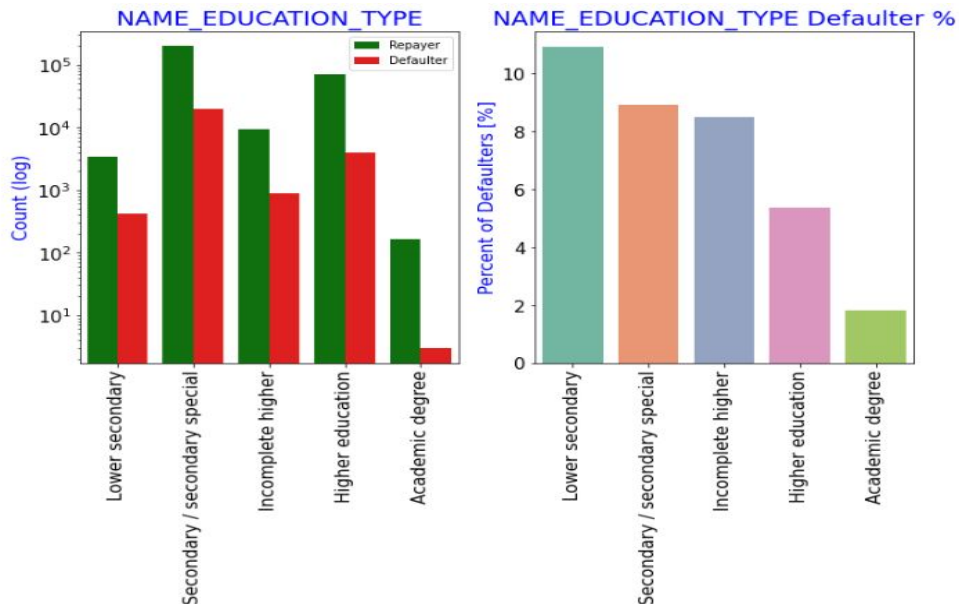
Variable:4 NAME_FAMILY_STATUS



Conclusion: Most of the people who have taken loan are married, followed by Single/not married and civil marriage. In terms of percentage of not repayment of loan, Civil marriage has the highest percent of not repayment (10%), with Widow the lowest (exception being Unknown).

3.5 Results of Univariate analysis

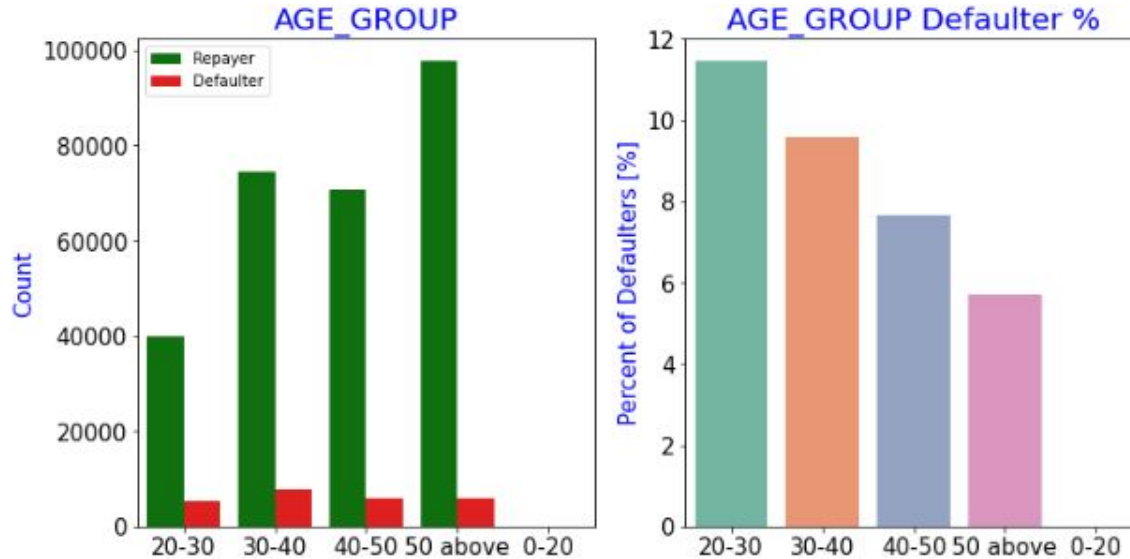
Variable:5 NAME_EDUCATION_TYPE



Conclusion: The majority of loan applicants did secondary education followed by Higher education. Applicants with Lower Secondary Education has most defaulters among others (> 10%) & the lowest rate of defaulters are the applicants who has Academic degree

3.6 Results of Univariate analysis

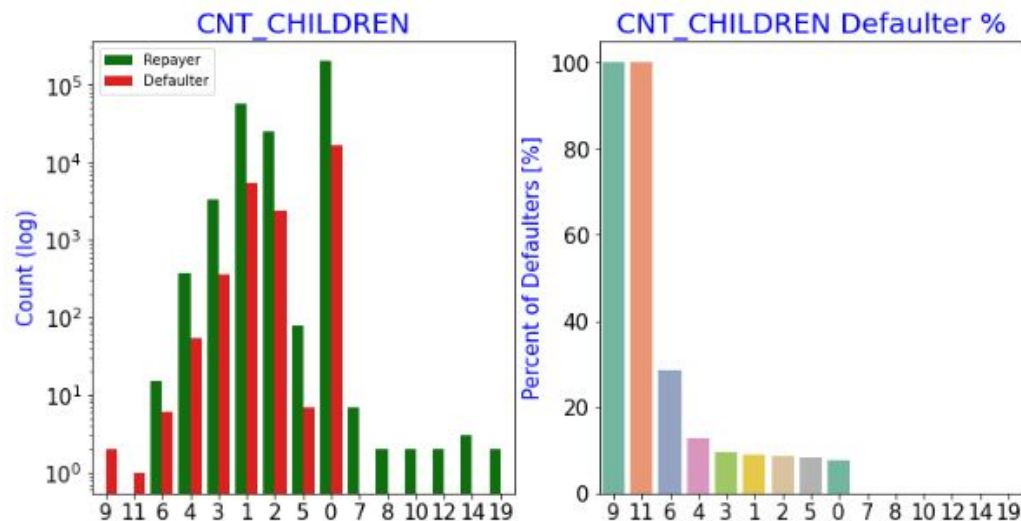
Variable:6 AGE_GROUP



Conclusion: Applicants whose age group is between 20-30 has more defaulters among other age groups & Applicants whose age group is Above 50 has lowest rate of Defaulters

3.7 Results of Univariate analysis

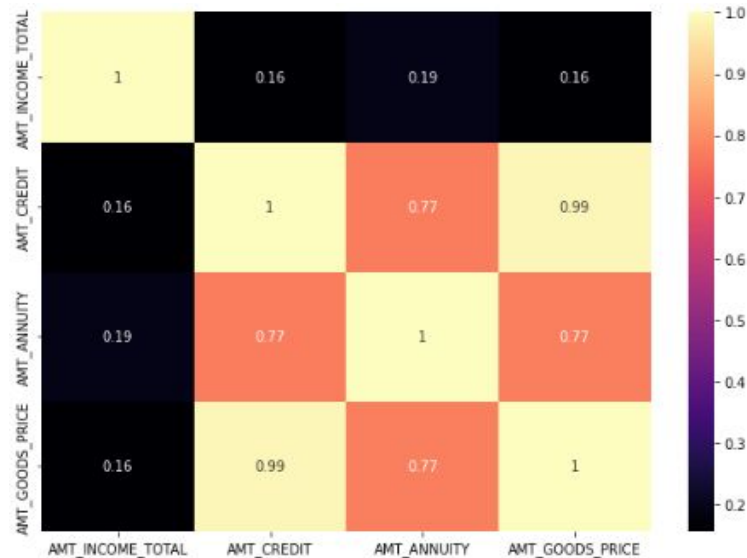
Variable:7 CNT_CHILDREN



Conclusion: Applicants with more number of children are most likely to be Defaulters Applicants with less than 3 children has low probability to be defaulters

4.1 Results of Bivariate/Multivariate analysis

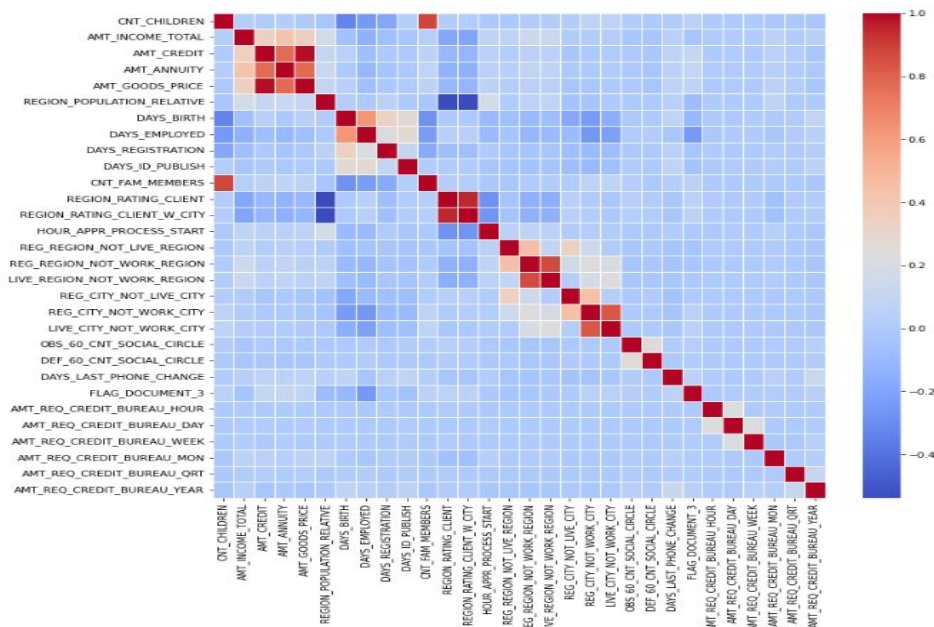
PLOT:01 Heat Map AMT_XX variables



Conclusion: High correlation between AMT_CREDIT and AMT_GOODS_PRICE - Applicants owning goods of high value can take loans of higher amounts.

4.2 Results of Bivariate/Multivariate analysis

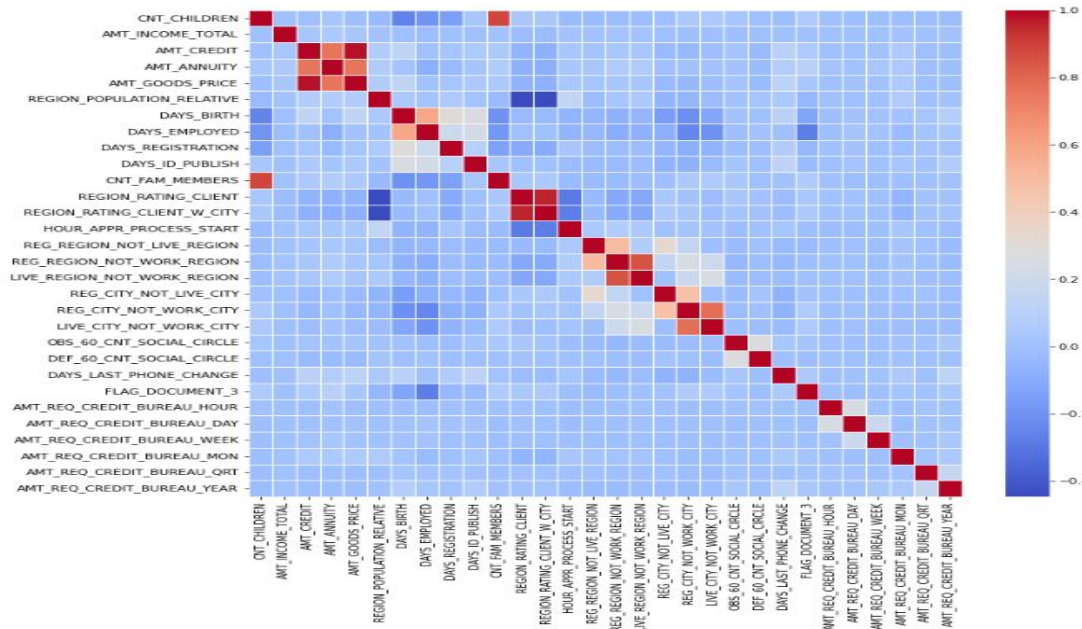
PLOT:02 Heat Map - Correlating factors amongst Repayers



Conclusion: Credit amount is highly correlated with amount of goods, price, loan annuity, & total income

4.3 Results of Bivariate/Multivariate analysis

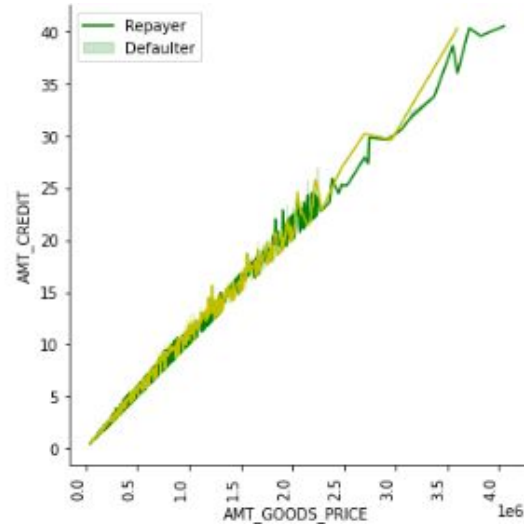
PLOT:03 Heat Map - Correlating factors amongst Defaulters



Conclusion: Credit amount is highly correlated with amount of goods price which is same as repayers. There is a severe drop in the correlation between total income of the client and the credit amount(0.038) amongst defaulters whereas it is 0.342 among repayers. There is a slight increase in defaulted count to observed count in social circle among defaulters(0.264) when compared to repayers(0.254)

4.4 Results of Bivariate/Multivariate analysis

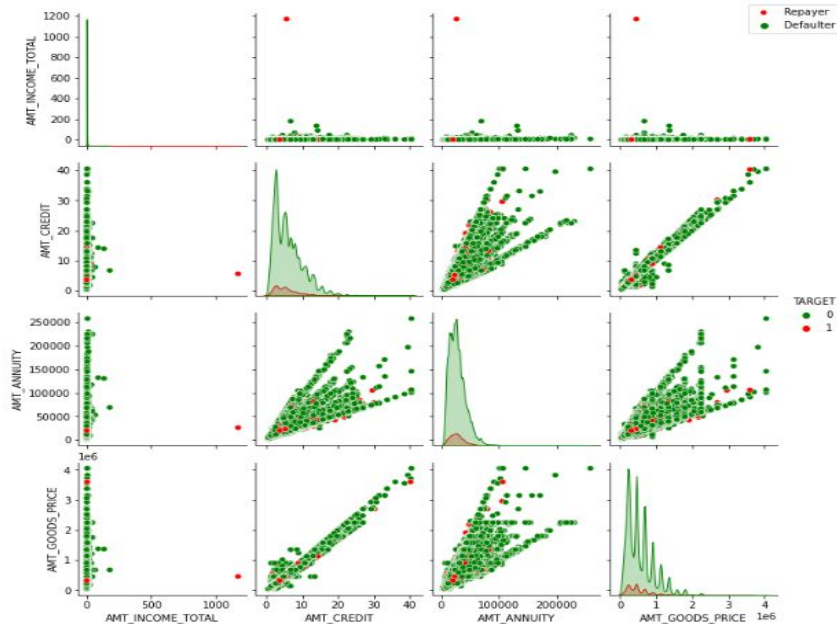
PLOT:04 AMT_GOODS_PRICE VS AMT_CREDIT



Conclusion: There is a rise in the number of defaulters when the loan amount is more than 3 Million

4.5 Results of Bivariate/Multivariate analysis

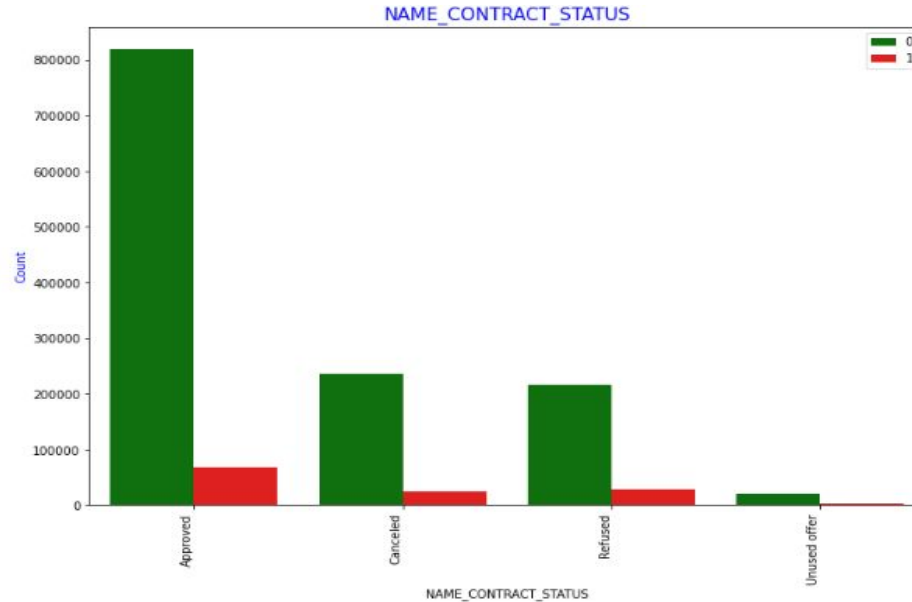
PLOT:05 Pair Plot



Conclusion: When $\text{amt_annuity} > 15000$ and $\text{amt_goods_price} > 3\text{M}$, there is a lesser chance of defaulters. AMT_CREDIT and AMT_GOODS_PRICE are highly correlated as based on the scatterplot where most of the data are consolidated in form of a line. There are very less defaulters for $\text{AMT_CREDIT} > 3\text{M}$. Inferences related to distribution plot has been already mentioned in previous distplot graphs inferences section.

4.6 Results of Bivariate/Multivariate analysis

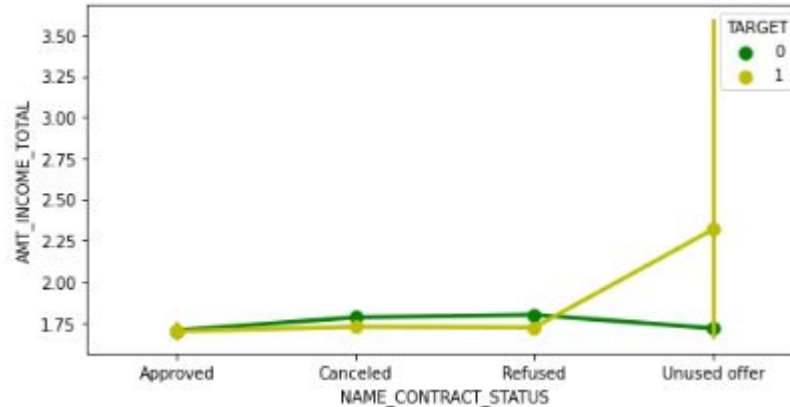
PLOT:06 Bar Chart - NAME_CONTRACT_STATUS



Conclusion: 90% of the previously cancelled client have actually repayed the loan. Revisiting the interest rates would increase business opportunity for these clients 88% of the clients who have been previously refused a loan has paid back the loan in current case. Refusal reason should be recorded for further analysis as these clients would turn into potential repaying customer.

4.7 Results of Bivariate/Multivariate analysis

PLOT:07 Point Plot - NAME_CONTRACT_STATUS



Conclusion: The point plot show that the people who have not used offer earlier have defaulted even when their average income is higher than others

5. Conclusions



Applicants with Income more than 700,000 are less likely to default



Applicants with 40+ year experience having less than 1% default rate



Applicants with zero to two children tend to repay the loans.



Applicants above age of 50 have low probability of defaulting.



Academic degree has less defaults.



Applicants with Trade Type 4 and 5 and Industry type 8 have defaulted less than 3%.