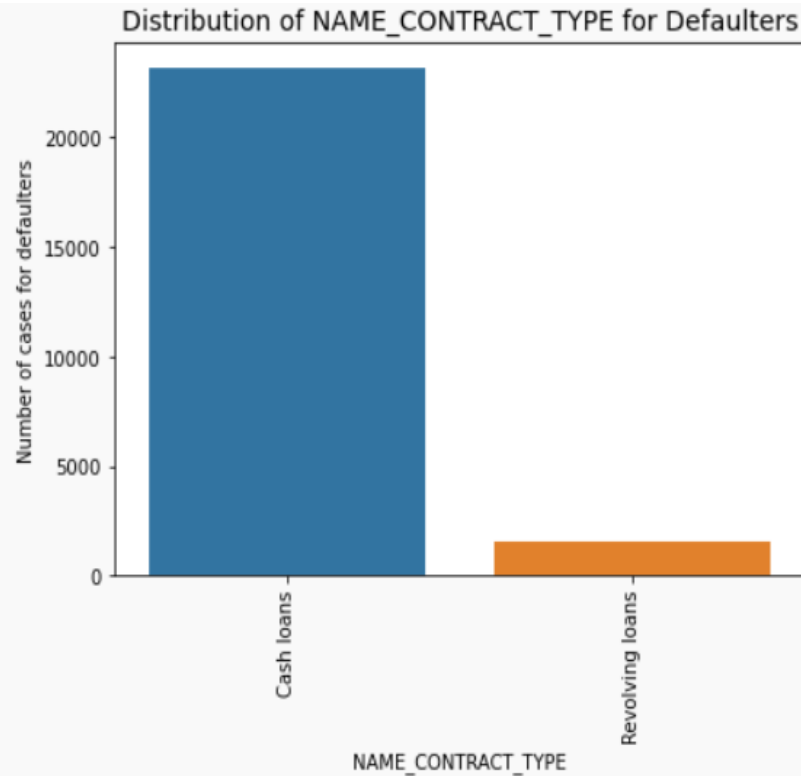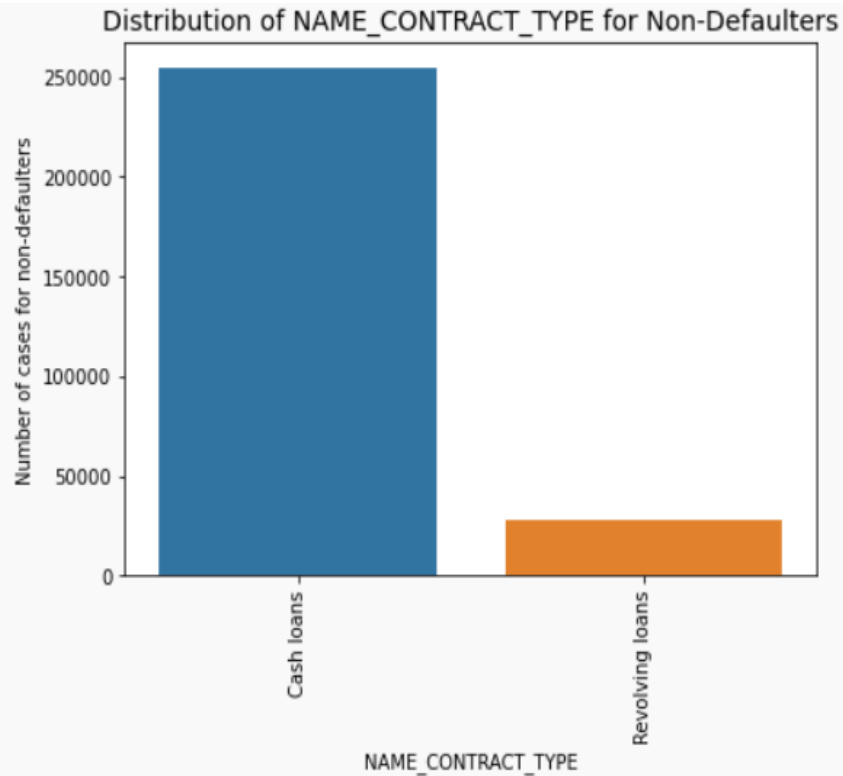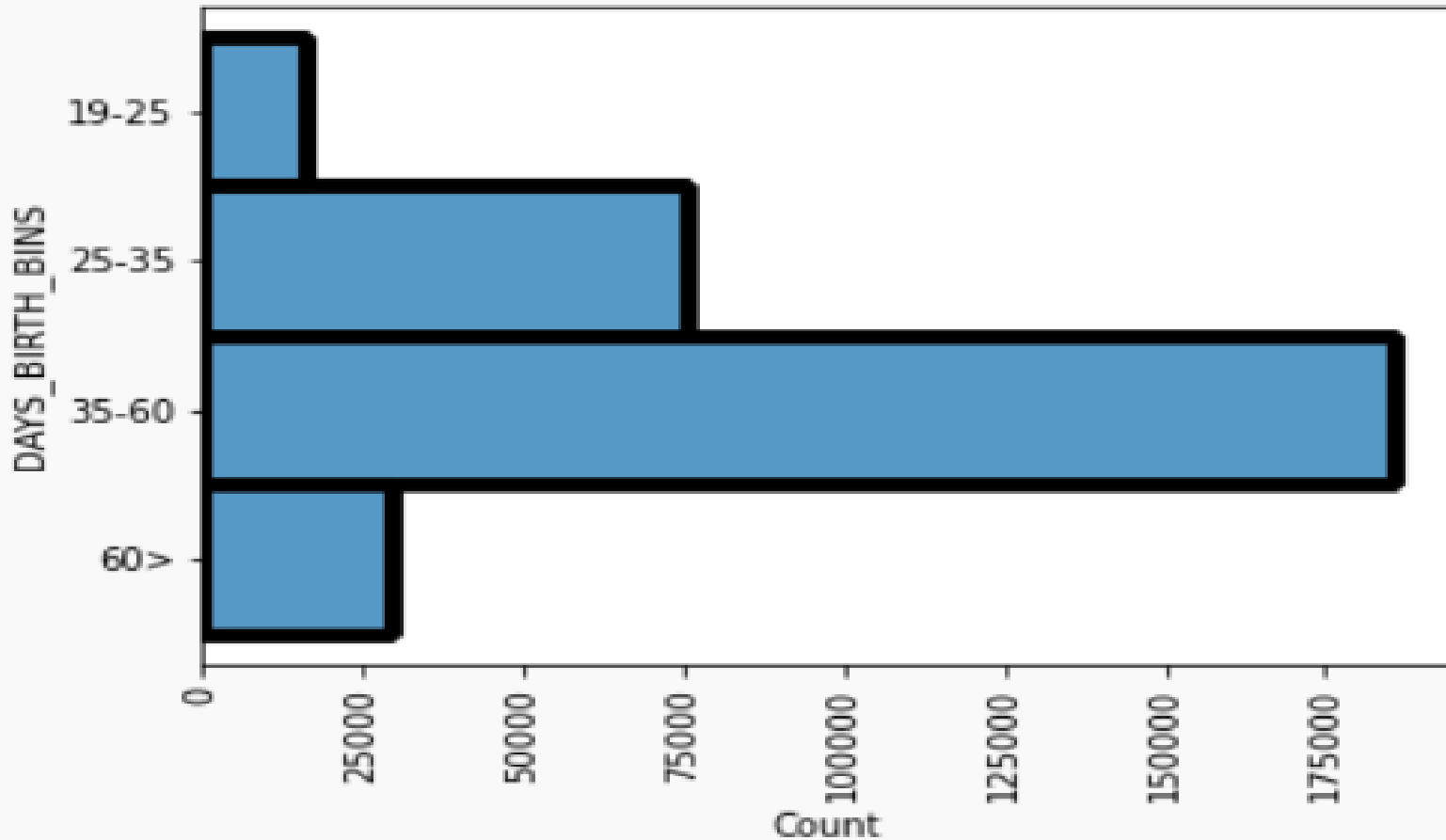# CREDIT EDA CASE STUDY

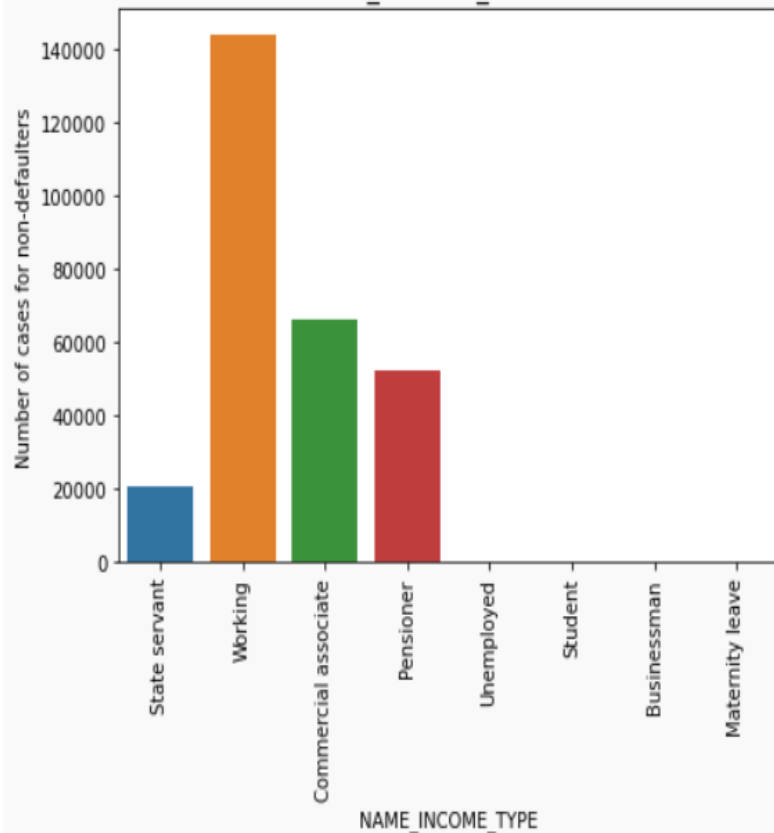By- Syed Sha Khalid

Cash Loans vs Revolving Loans

- We can notice that revolving loans are lesser in the defaulted population.

- Hence we can infer that revolving loans are comparatively safer.

- This may be attributed to the Nature of revolving loan as it is considered a flexible financing tool due to its repayment and re-borrowing flexibility
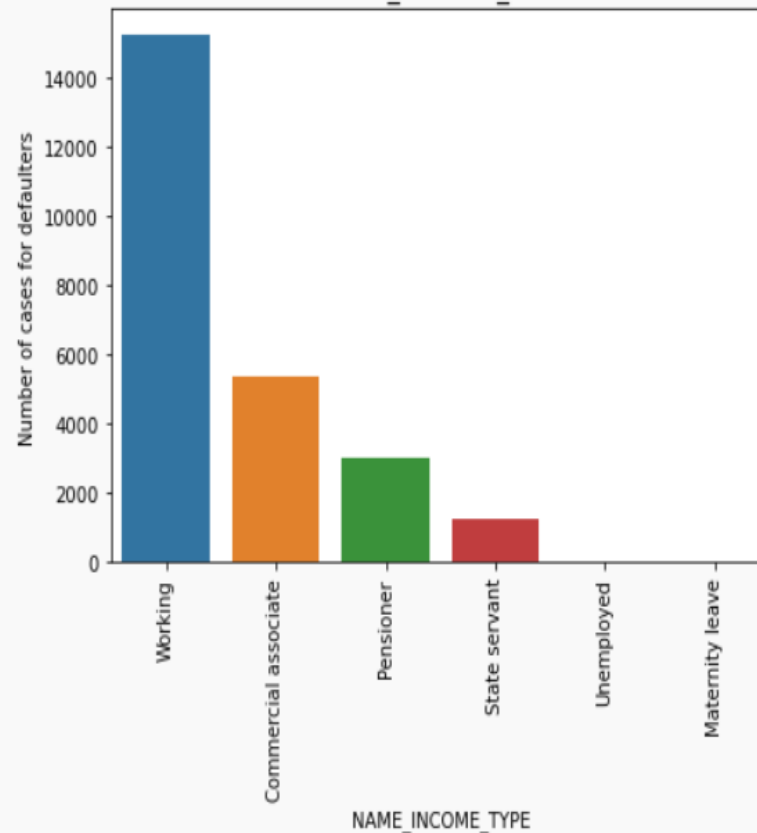
Age analysis after binning

- We can observe that most of our population lies between the age of 35-60.

- The least amount population lies in between the age group 19-25.

Distribution of NAME_INCOME_TYPE for Non-Defaulters / Distribution of NAME_INCOME_TYPE for Defaulters
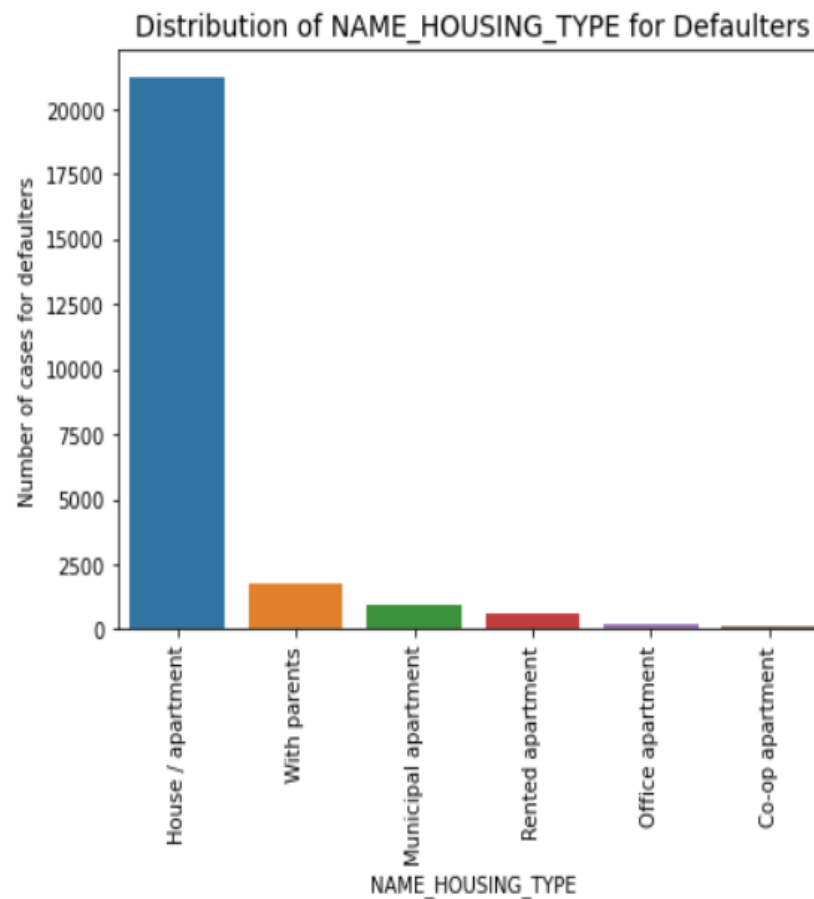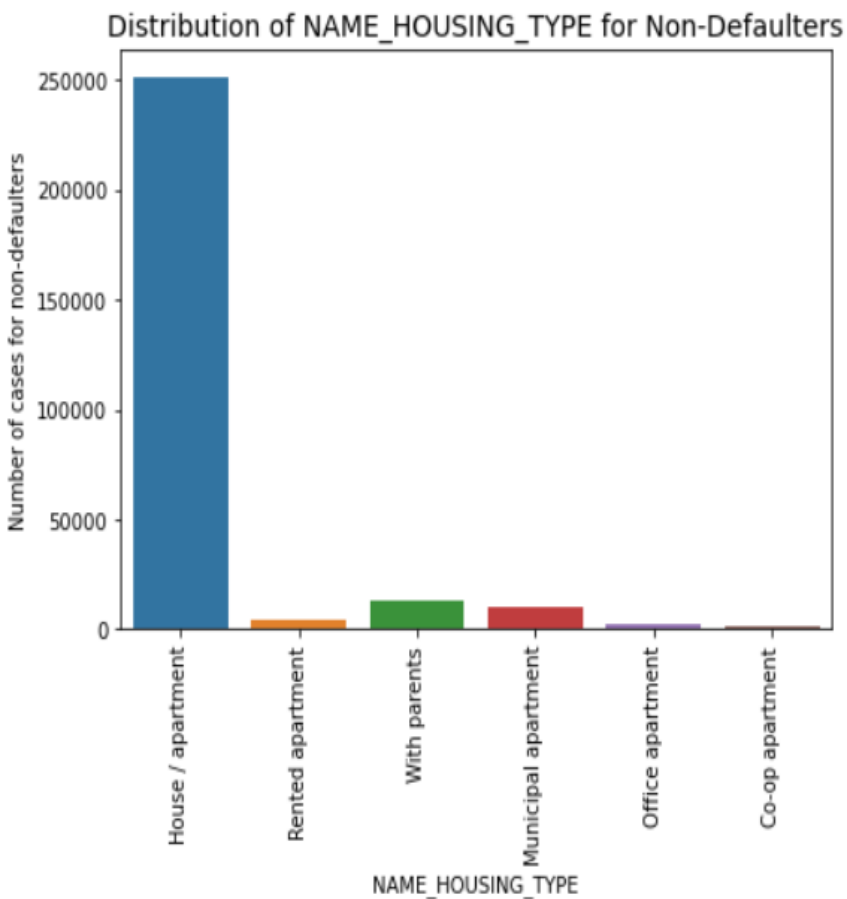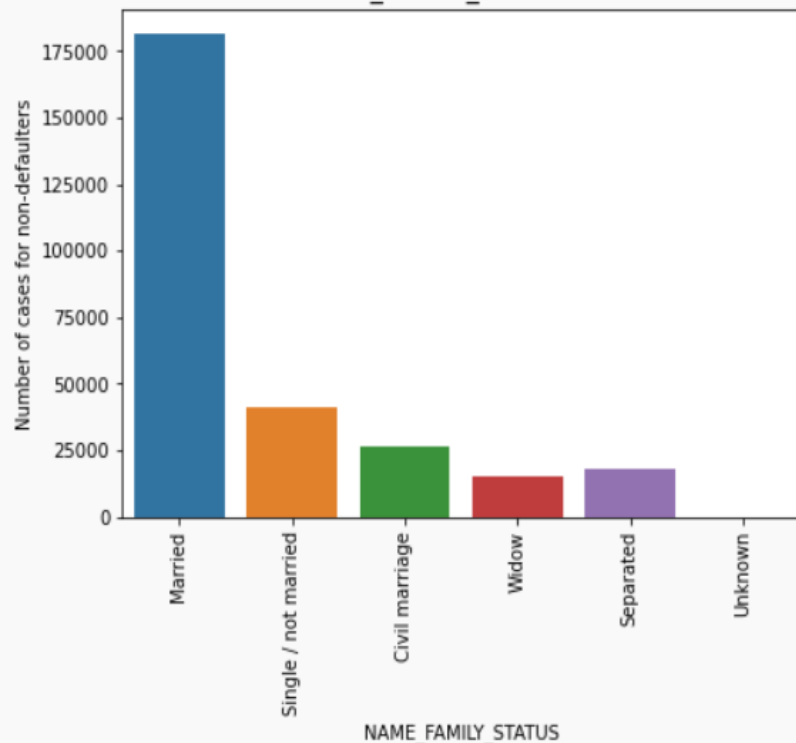
Income Types Distribution

- **Working population has the highest amount of defaulters**

- **State Servant have very low default population as they have secure jobs with steady income.**

Distribution of NAME_HOUSING_TYPE for Non-Defaulters / Distribution of NAME_HOUSING_TYPE for Defaulters

Housing type distribution

- Population living in Rented apartments and those living with parents have higher default rate.

- Living in rental apartment means a cash outflow towards rent and thus less cash left for repayment of loan.

- Living with parents may suggest that the income is not too high and thus difficulty in repayment of loan.

Distribution of NAME_FAMILY_STATUS for Non-Defaulters / Distribution of NAME_FAMILY_STATUS for Defaulters
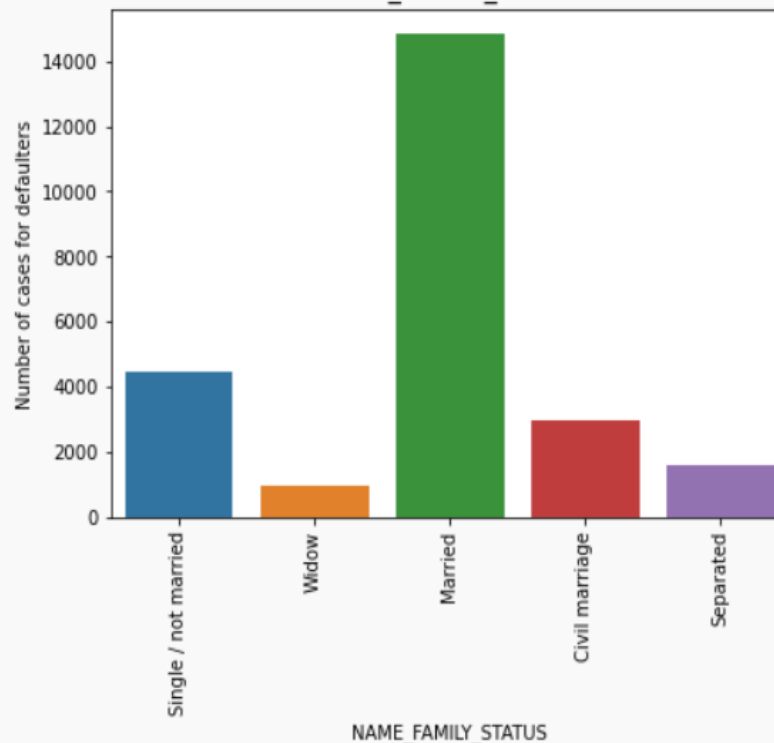
Family Status Distributions

- **The highest amount of non-defaulters are married people.**

- **On the other side we can also observe the highest amount from defaulters population are married.**

- **The least amount of defaulters are widow and separated.**

Correlation between Important columns (Target 0)

Correlation between important columns
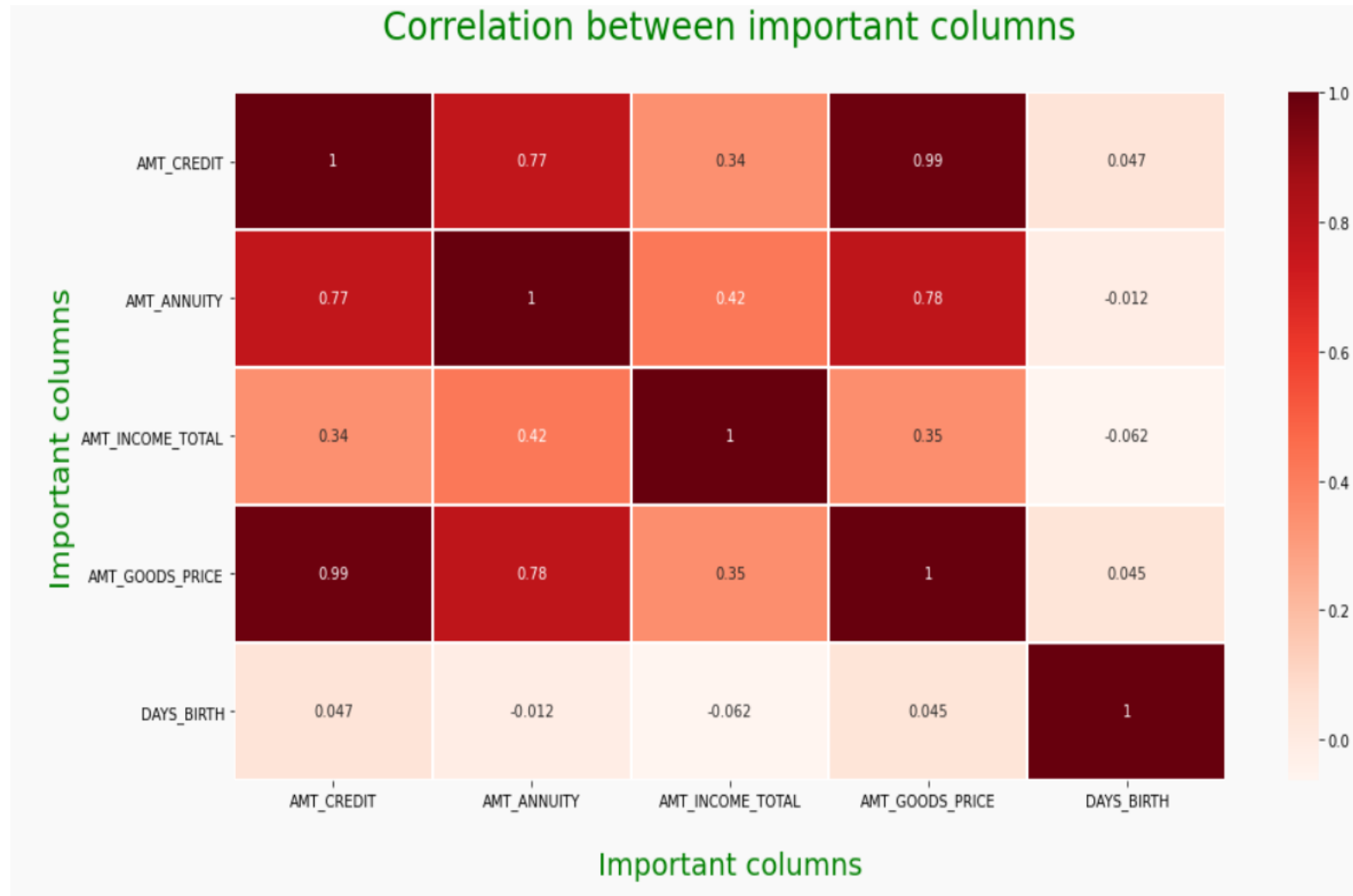
- Credit amount is related to Goods price as the higher the price the higher the amount of credit taken by customer and vice versa.

- Annuity amount and Credit amount is also related as higher credit amount has a higher annuity.

- Therefore annuity amount and Goods price amount is also correlated as higher price of goods results in higher credit amount which eventually results in higher annuity amount.

# Correlation between Important columns (Target 1)

Correlation between important columns
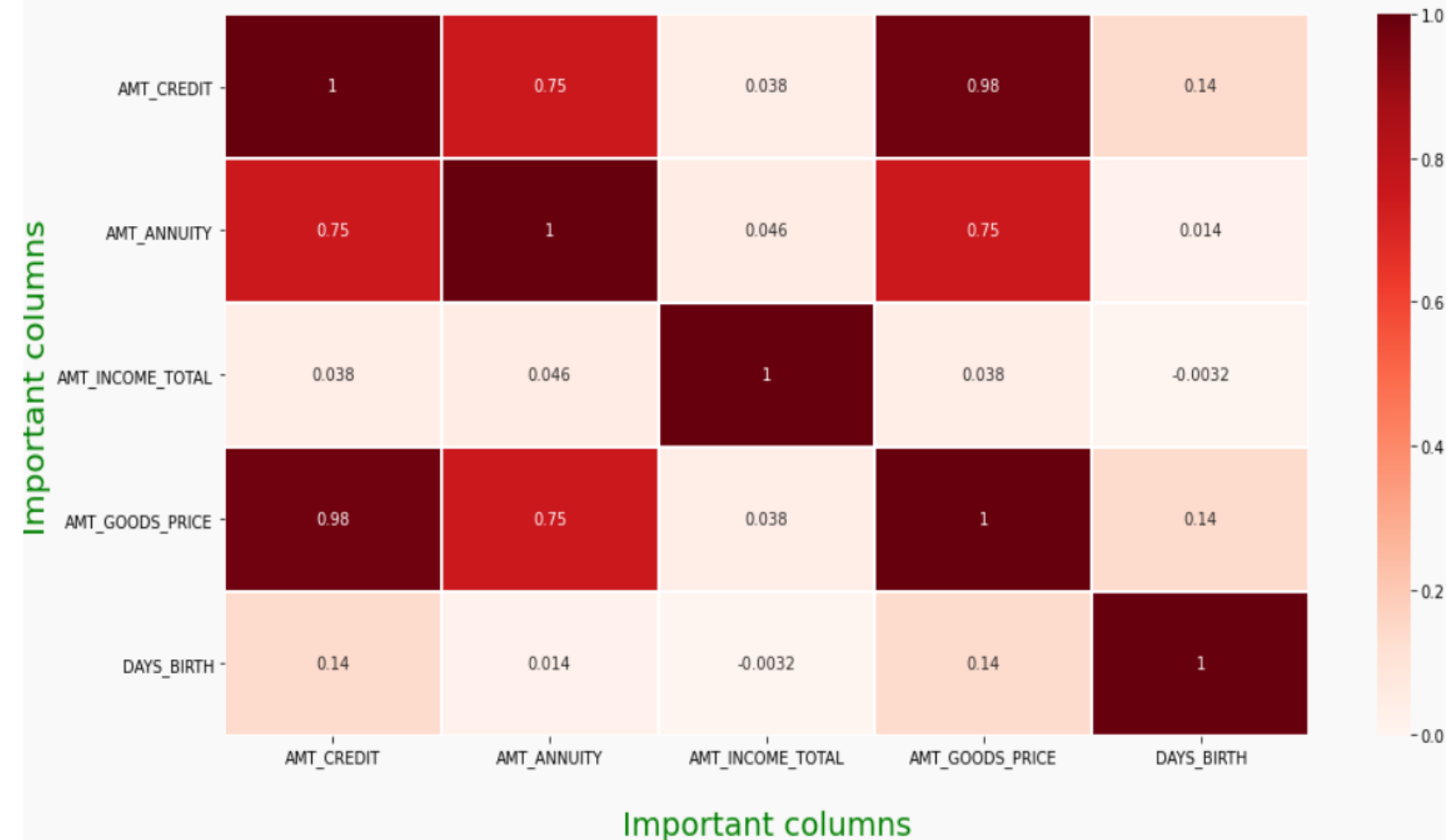
- The Target 1 population are people with payment difficulties.

- Target 1 heat map looks similar but there are a few changes like there is very little relation between Amount of total income and amount of annuity compared to Target 0 population
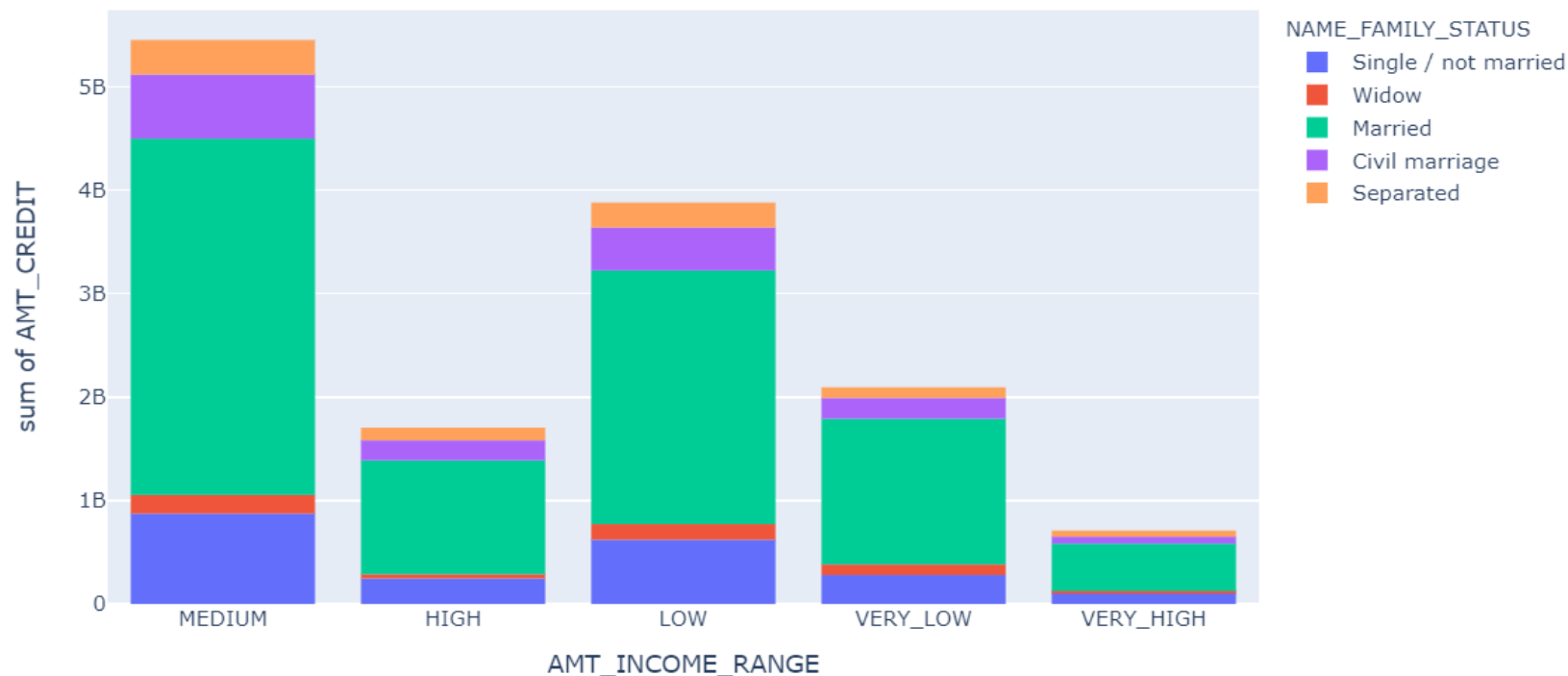
Credit amount vs Education of Loan-Payment Difficulties
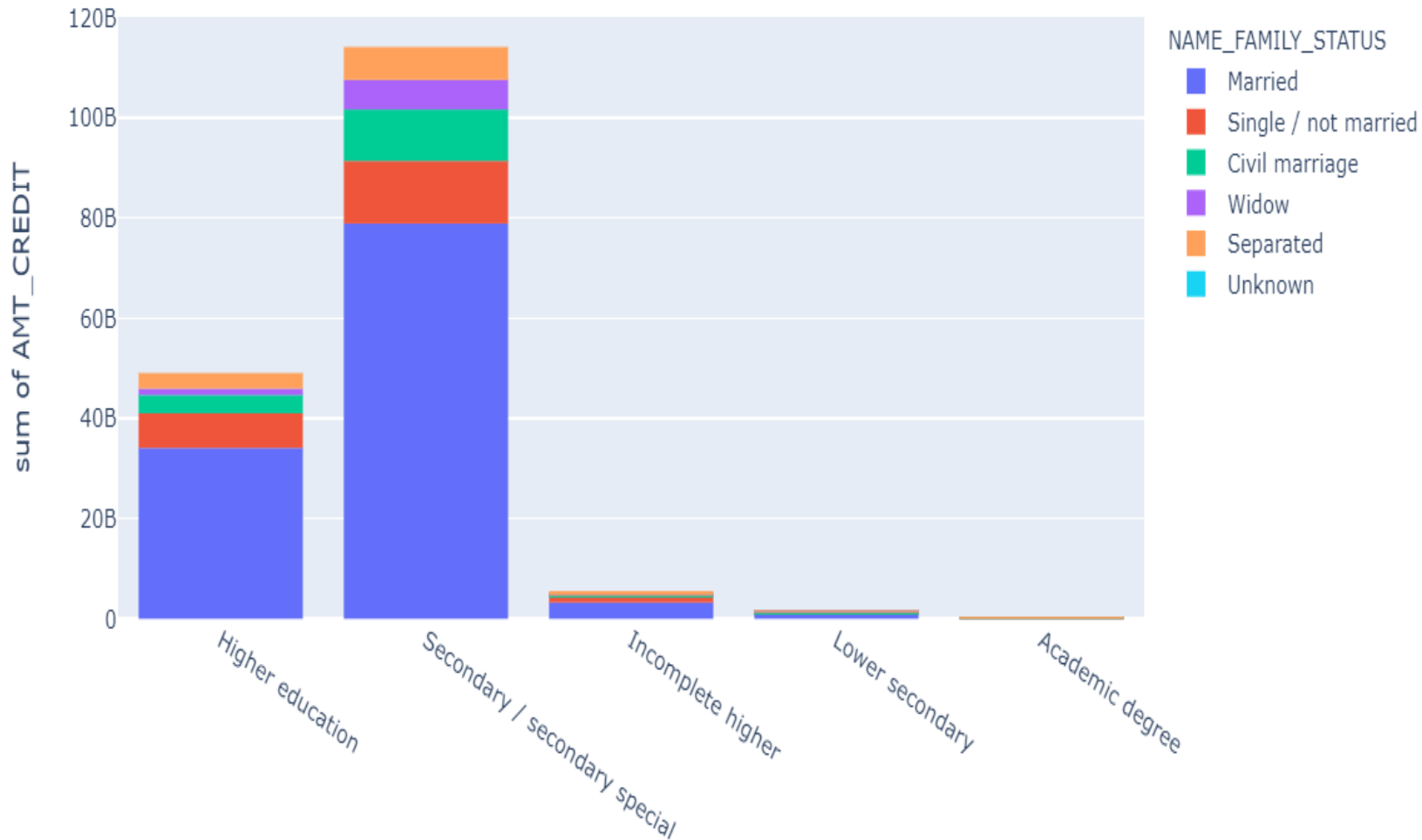
Credit amount vs Education of loan-Payment difficulties

- People with secondary and higher education have higher credit amount compared to other categories.

- The majority of higher and secondary credit amount is occupied by married people
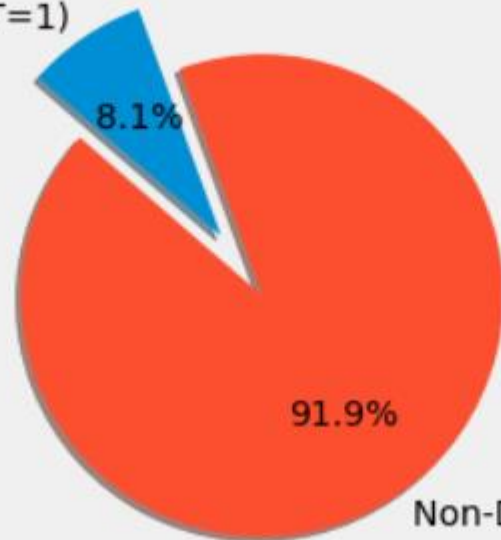
Income range vs Credit amount

- **The highest amount of credit is for people with medium income, followed by people with low income.**

- **The lowest amount of credit is for people with very high income, as people with very high income rarely need credit or loans.**
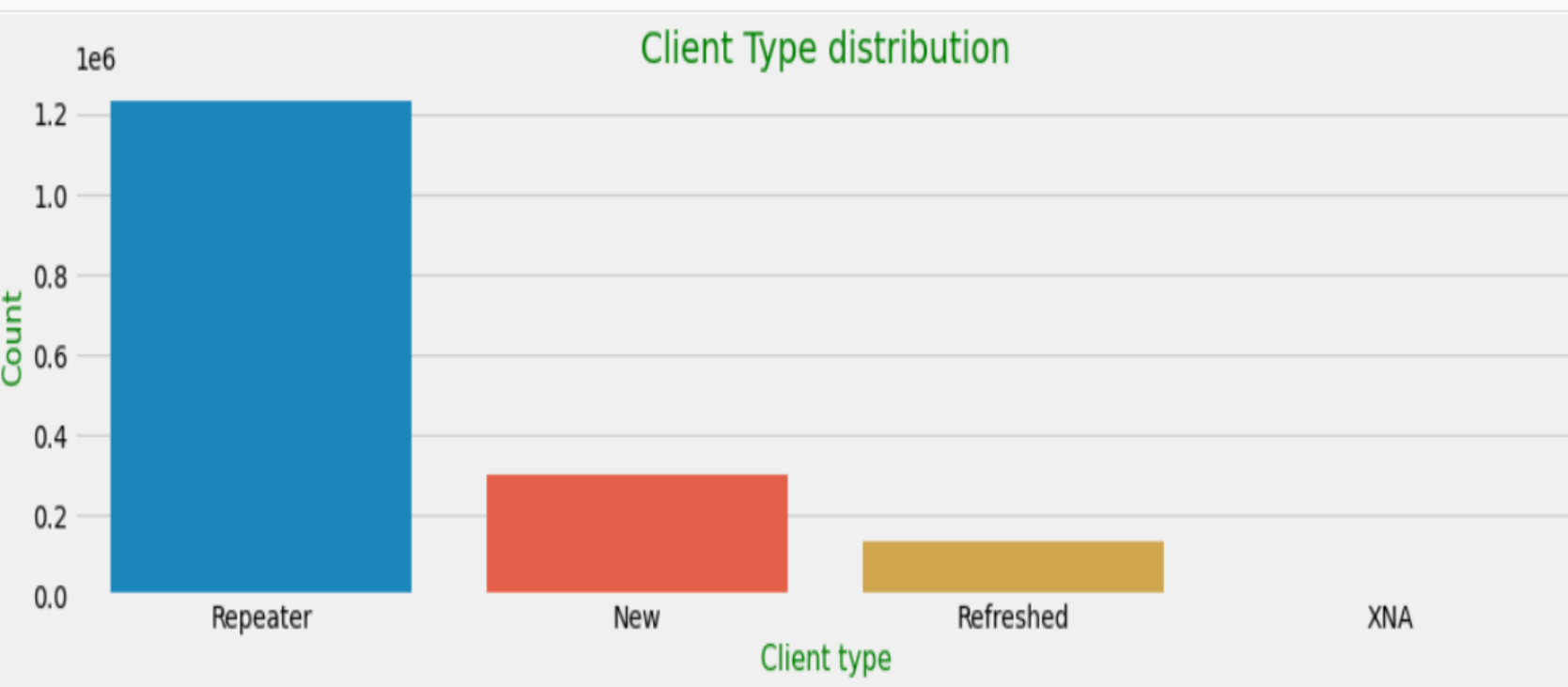
Data Imbalance

- **More than 90% of the population consists of people who are Target=0, i.e. people with no payment difficulties/defaulted on loans.**

Previous Data Application Analysis.

Client Type distribution

Client type while applying for the previous application

- We can observe that the client type is highest for repeater while applying for the previous application.

- The second highest population is for the new client type followed by refreshed.

Bivariate Analysis After Merging Previous Data

% of Loan Payment Difficulties for NAME_CONTRACT_STATUS and NAME_CLIENT_TYPE
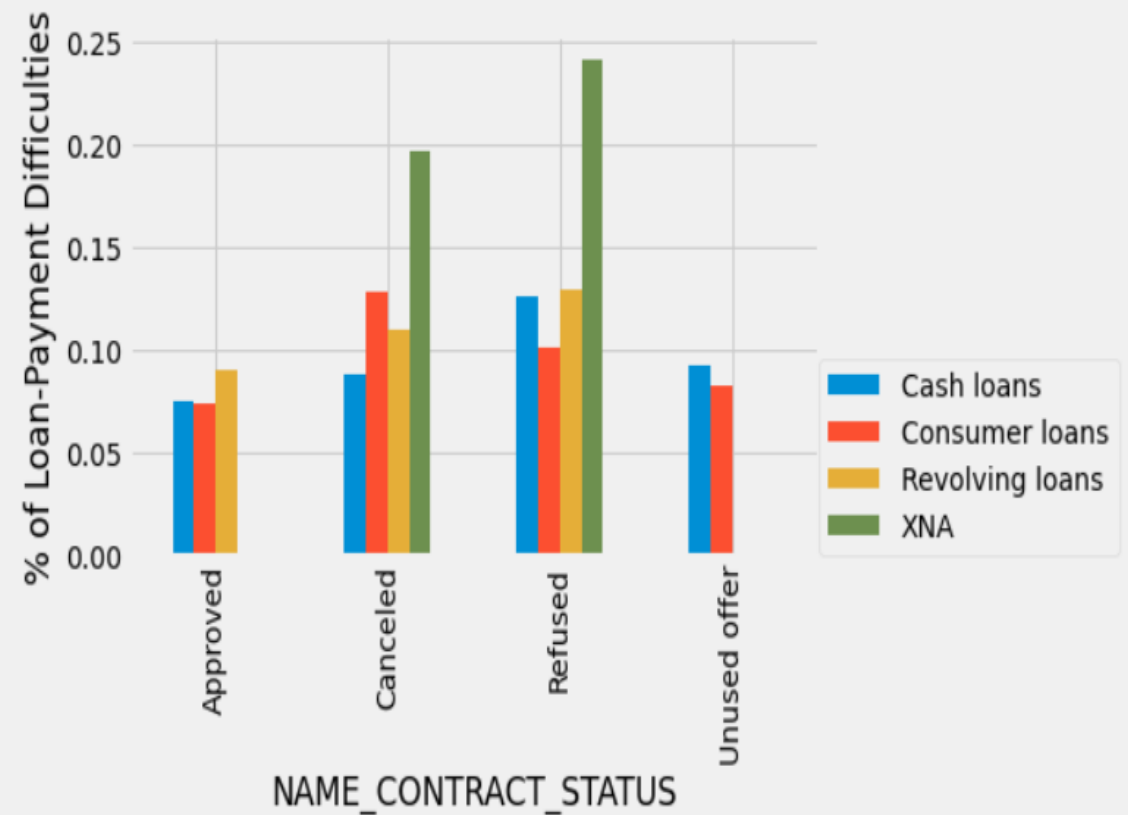
Contract status vs Client type

- The highest amount of approved loans are of new client type.

- The highest amount of cancelled loans are also of new client type.

- The highest amount of refused loans are also of repeater client type.

- The highest amount of unused offer are of new client type

Bivariate Analysis After Merging Previous Data

% of Loan Payment Difficulties for NAME_CONTRACT_STATUS and NAME_CONTRACT_TYPE

Contract status vs Contract type

- The highest amount approved loans are revolving loans.

- The highest amount of cancelled loans are consumer loans

- The highest amount of refused loans are revolving loans.

- The most unused offer are cash loans.

Conclusion

- Banks should focus less on income type 'Working' as they are having most number of unsuccessful payments.

- Banks should focus more on contract type 'Student' , ' pensioner' , 'state servant' and 'Businessman' for successful payments.

- Banks should accept more revolving loans as they are less defaulted compared to cash loans.

- Banks should also focus on people with Family status as 'widow' and 'separated' as they have the least amount of defaulters compared to other Family Status.