



EDA CASE STUDY

Credit Risk Analysis

: TUSHAR GEETE

Problem Statement :

In any loan application, two distinct risks are involved:

- 1.Risk Type 1: Denying a loan to a potentially creditworthy applicant results in missed business opportunities for the company.
- 2.Risk Type 2: Approving a loan for an applicant who is likely to default may lead to financial losses for the company.

An examination of the dataset has been conducted using Python within Visual Studio Code.

Analysis Performed :

Steps:

1. Check Missing Values: Determine which ones need handling and decide on the approach to handle them (all the outliers which needs to be taken care of have been explained in 'Case Study.ipynb').
2. Check Outliers and Data Imbalance: Assess the ratio of imbalance in the data. Identify
3. Top 10 Correlations: Find the most significant correlations between the client's payment difficulties and other variables within the Application and Previous App data frames.
4. Determine the Most Relevant Correlation.
5. Through out the Analysis 'Target0', 'Count_zeroes' are Non Defaulter and



File Name : Application Data

Ratio Of Defaulter and Non Defaulter in the given Data Set:

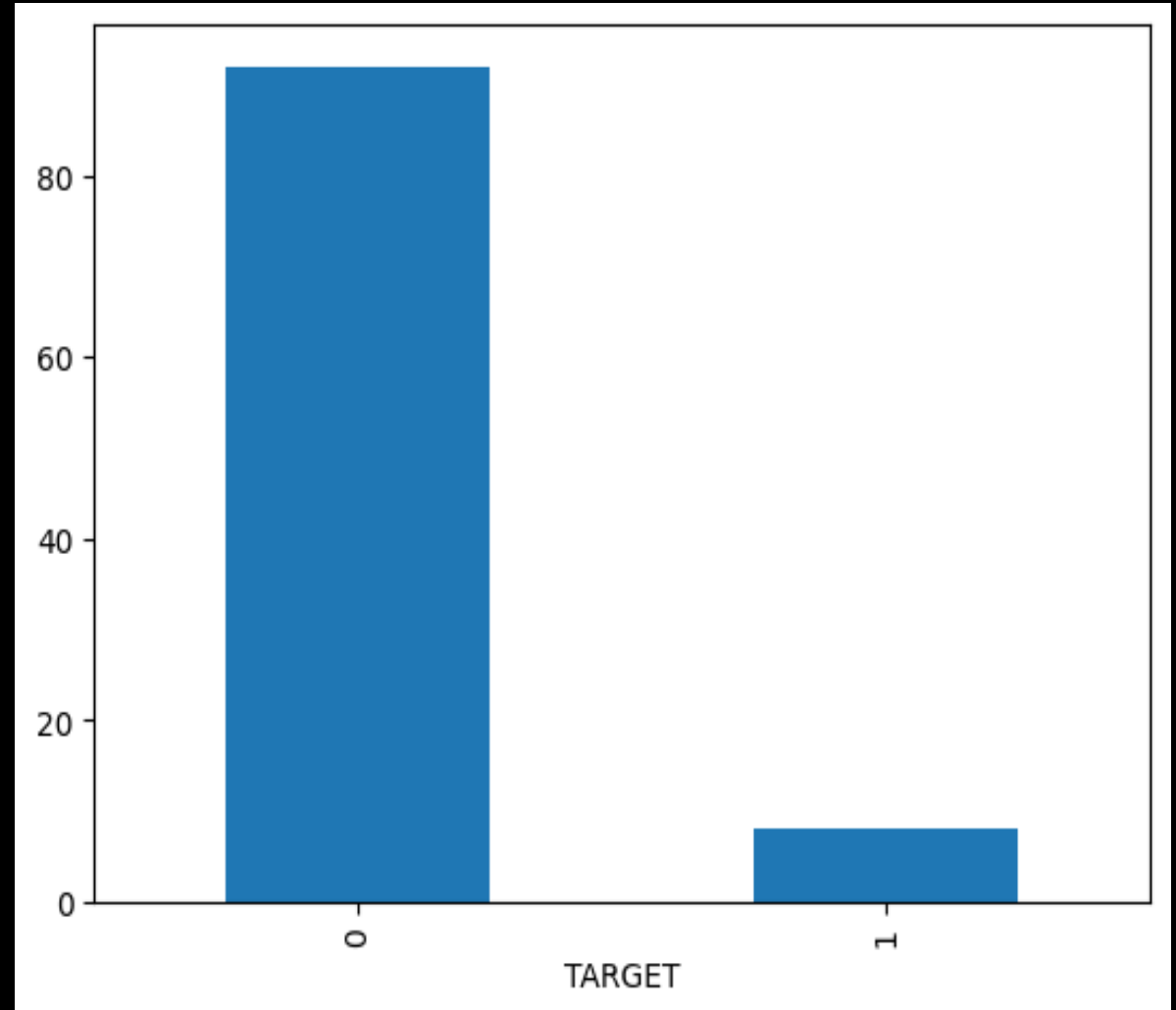
Here :

‘0’ Represents the percentage of Non – Defaulter in the Data Set and ‘1’ Indicates the Percentage of Defaulter.

Percentage Ratio :

Non – Defaulter : 91.93%

Defaulter : 8.07%



Ratio Of Males and Females in the Dataset:

Here :

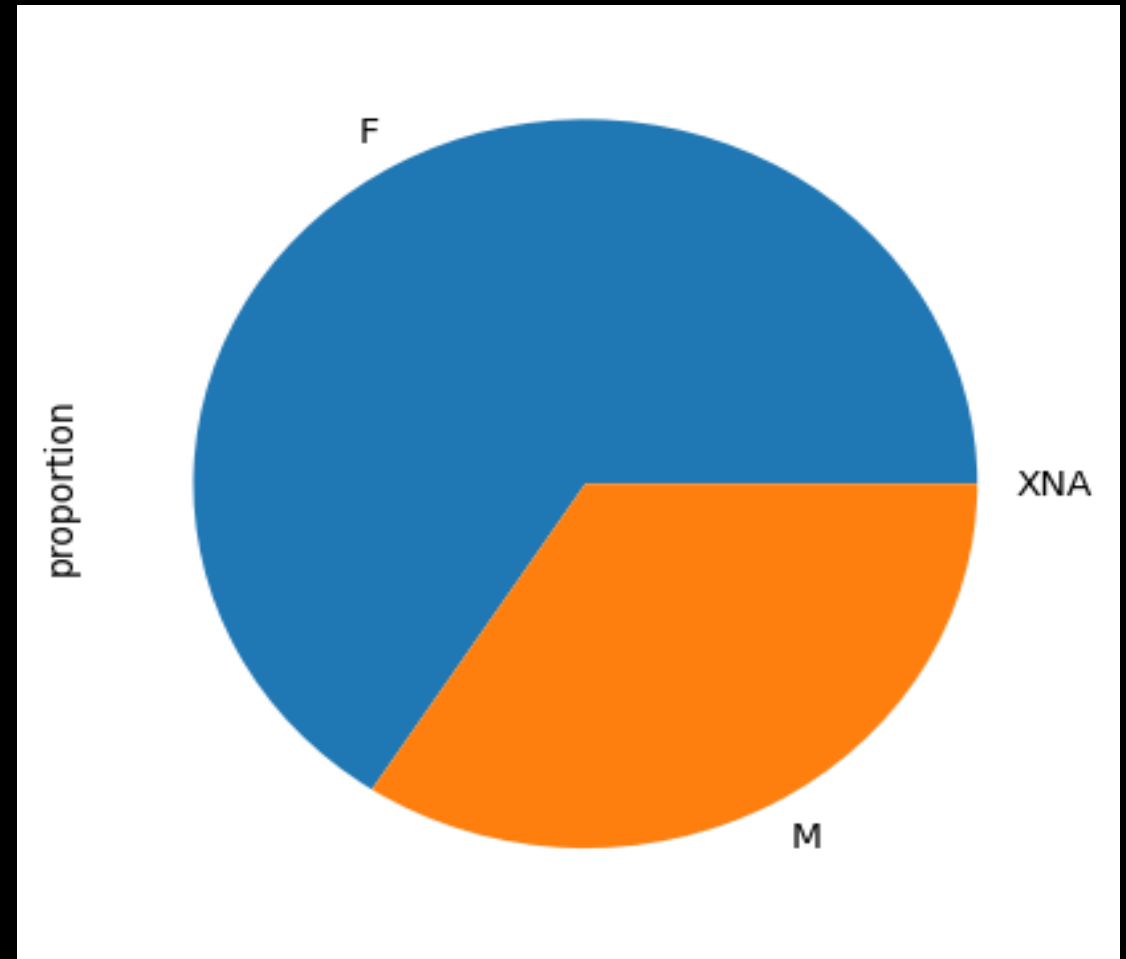
‘F’ Represents the percentage of Females in the Data Set and ‘M’ Indicates the Percentage of Males and ‘XNA’ represents the Missing Entries and we have shared the way to deal with it in file [Case Study.ipynb](#).

Percentage Ratio :

Females : 65.83%

Males : 34.16%

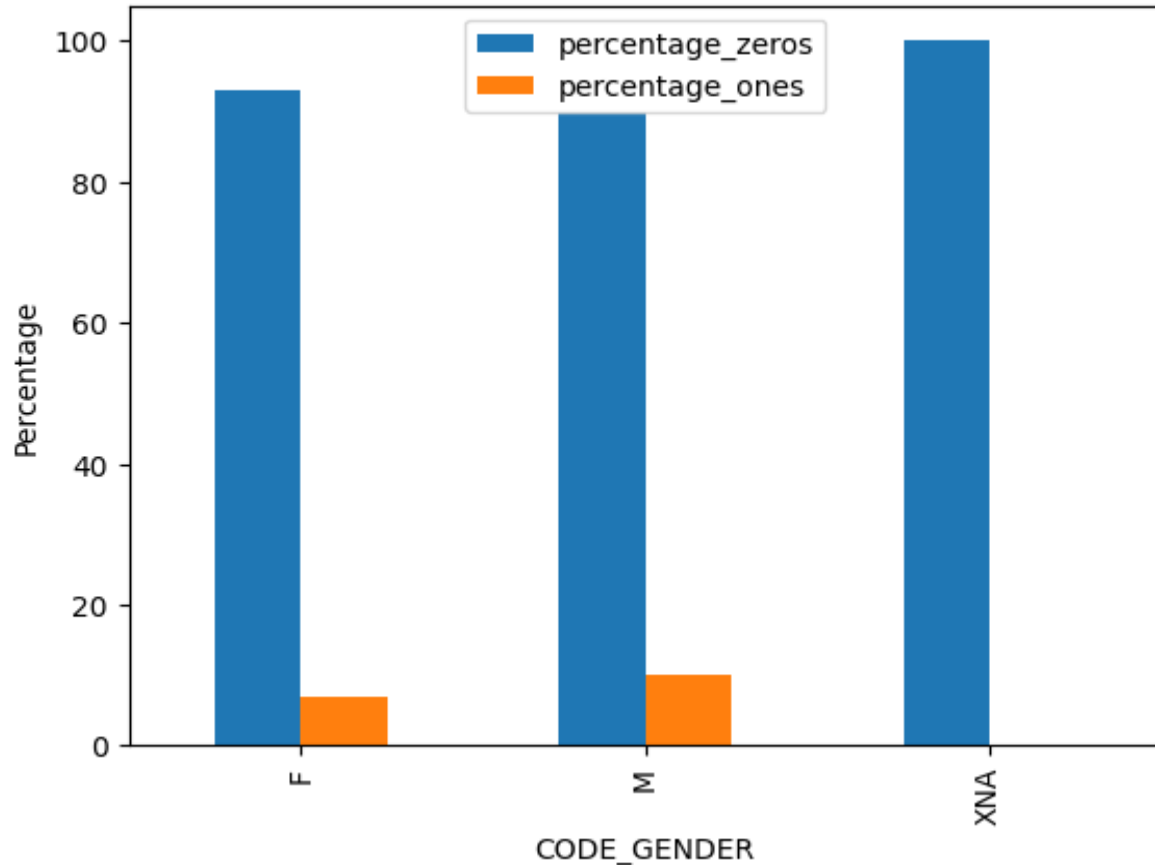
XNA : 0.001%



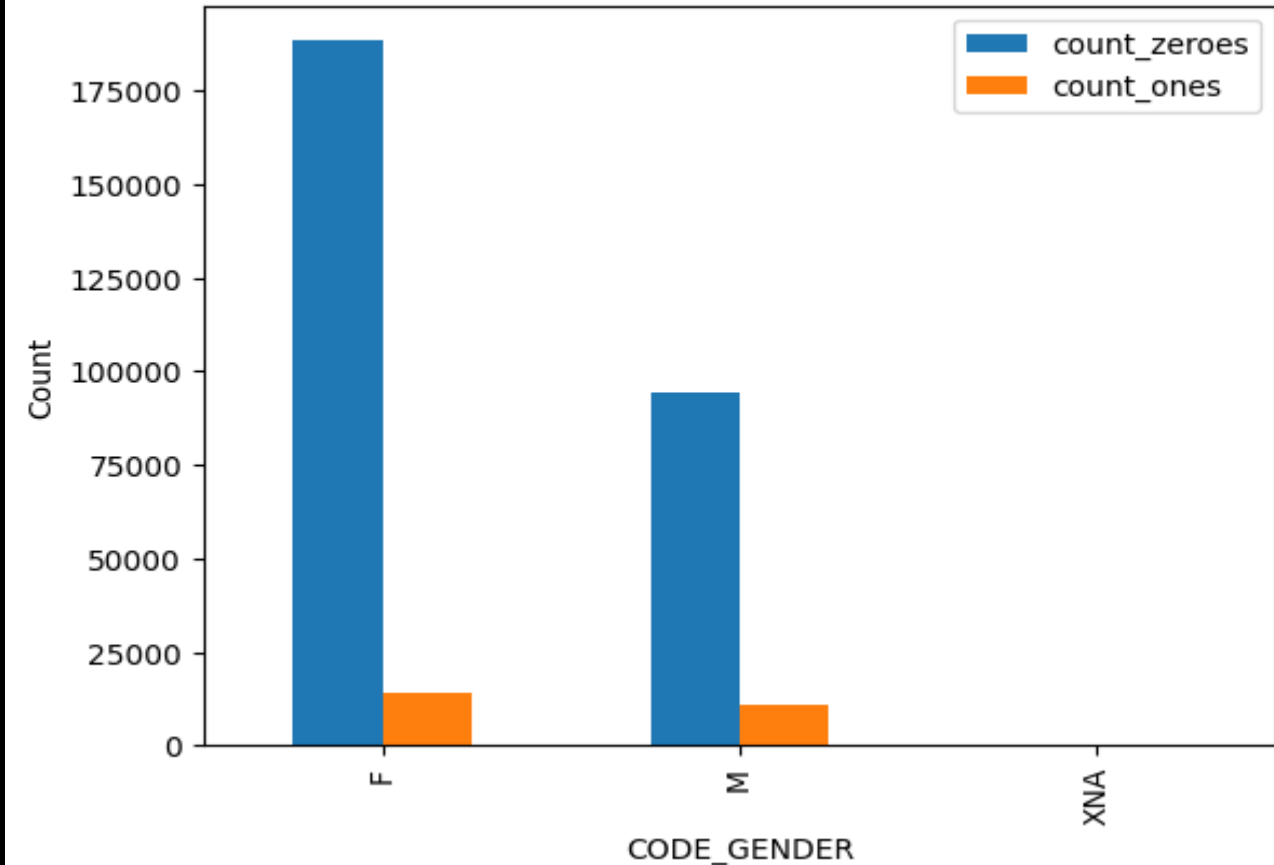
Analysis

Gender and Defaulter Relation:

Percentage of Zeros and Ones in TARGET by Gender



Count of Zeros and Ones in TARGET by Gender

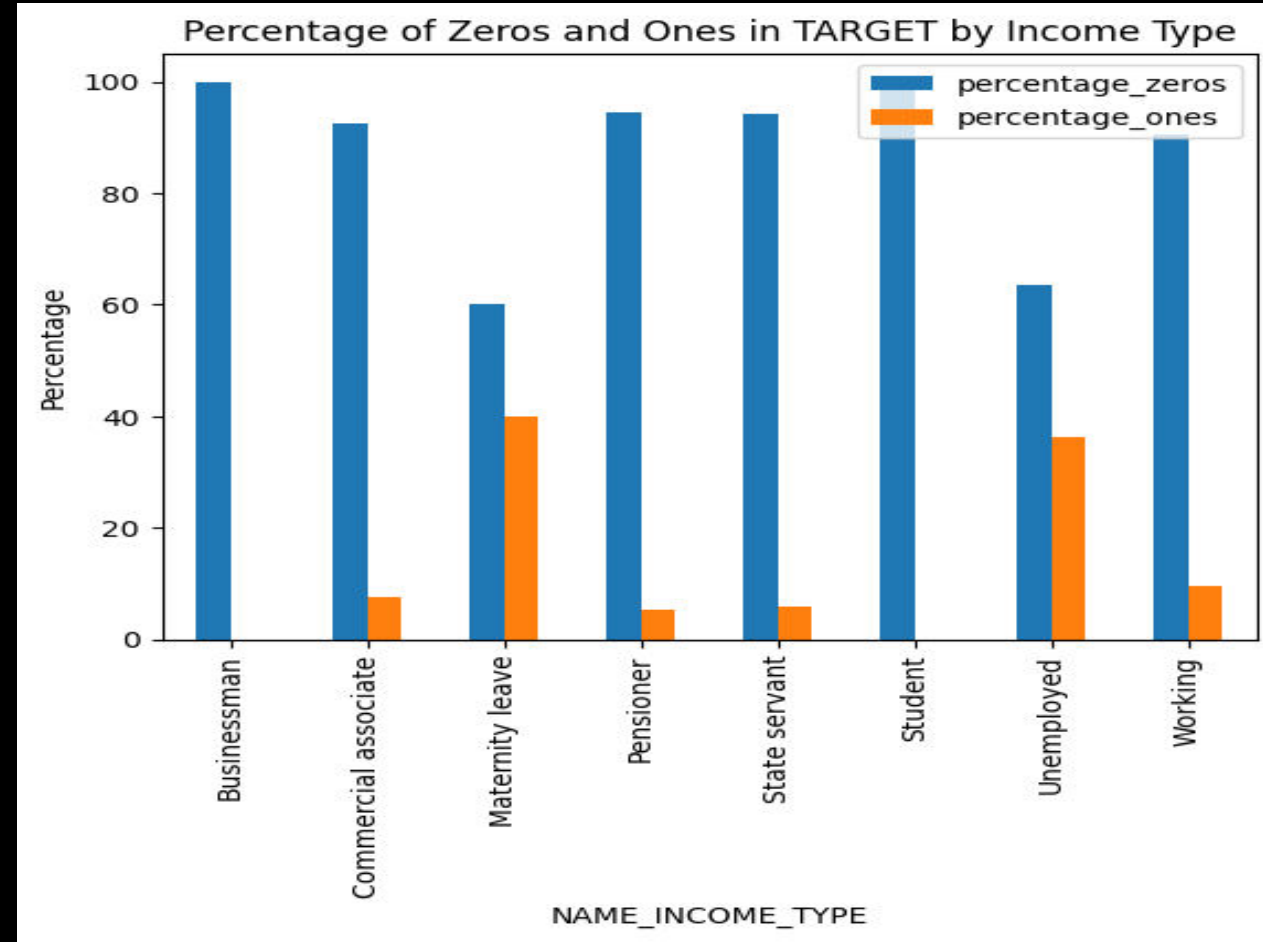
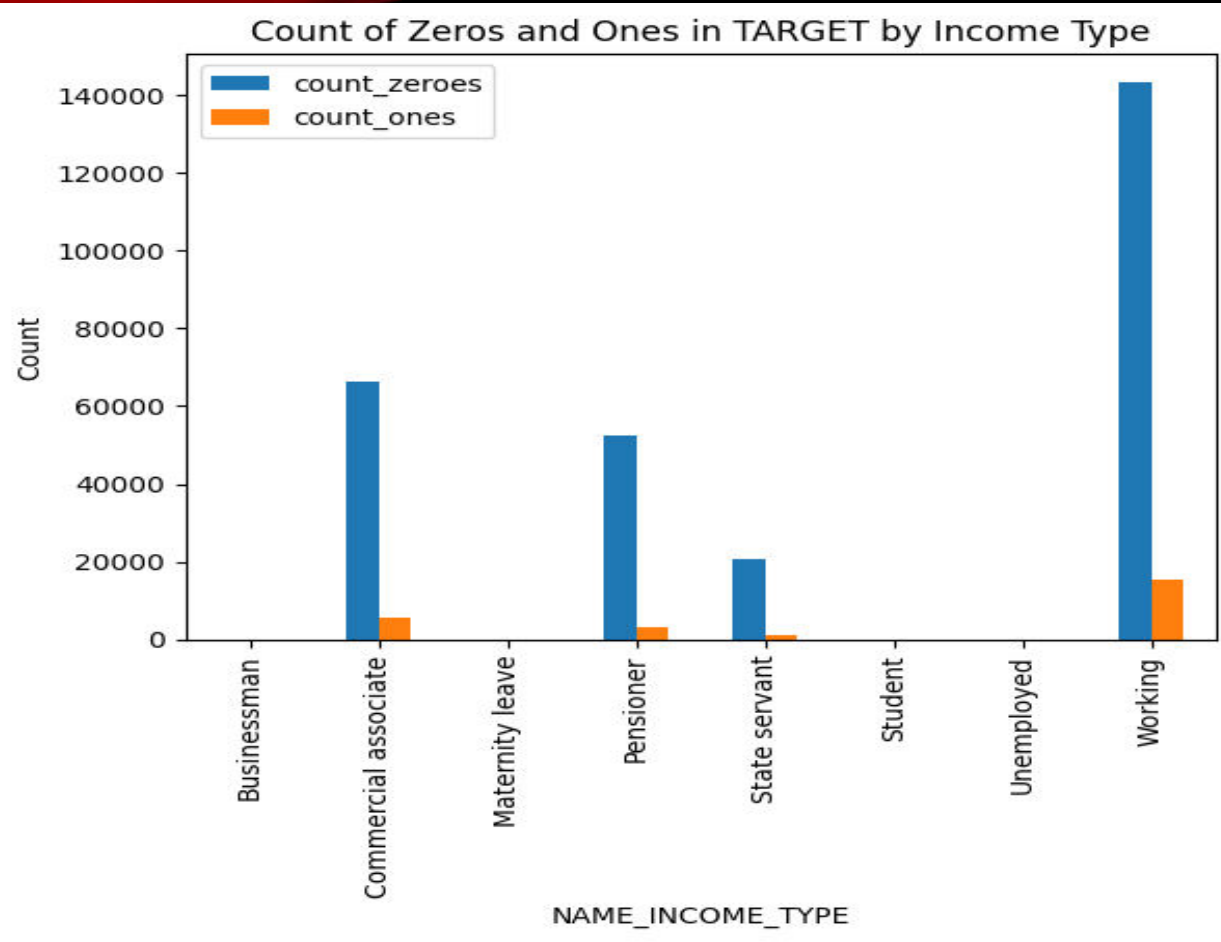


Observation :

1. The percentage of defaulters of Males is slightly higher than percentage of defaulters of Females.
2. The number of Females applied for the loan is higher.

Analysis

Income Type and Defaulter Relation:



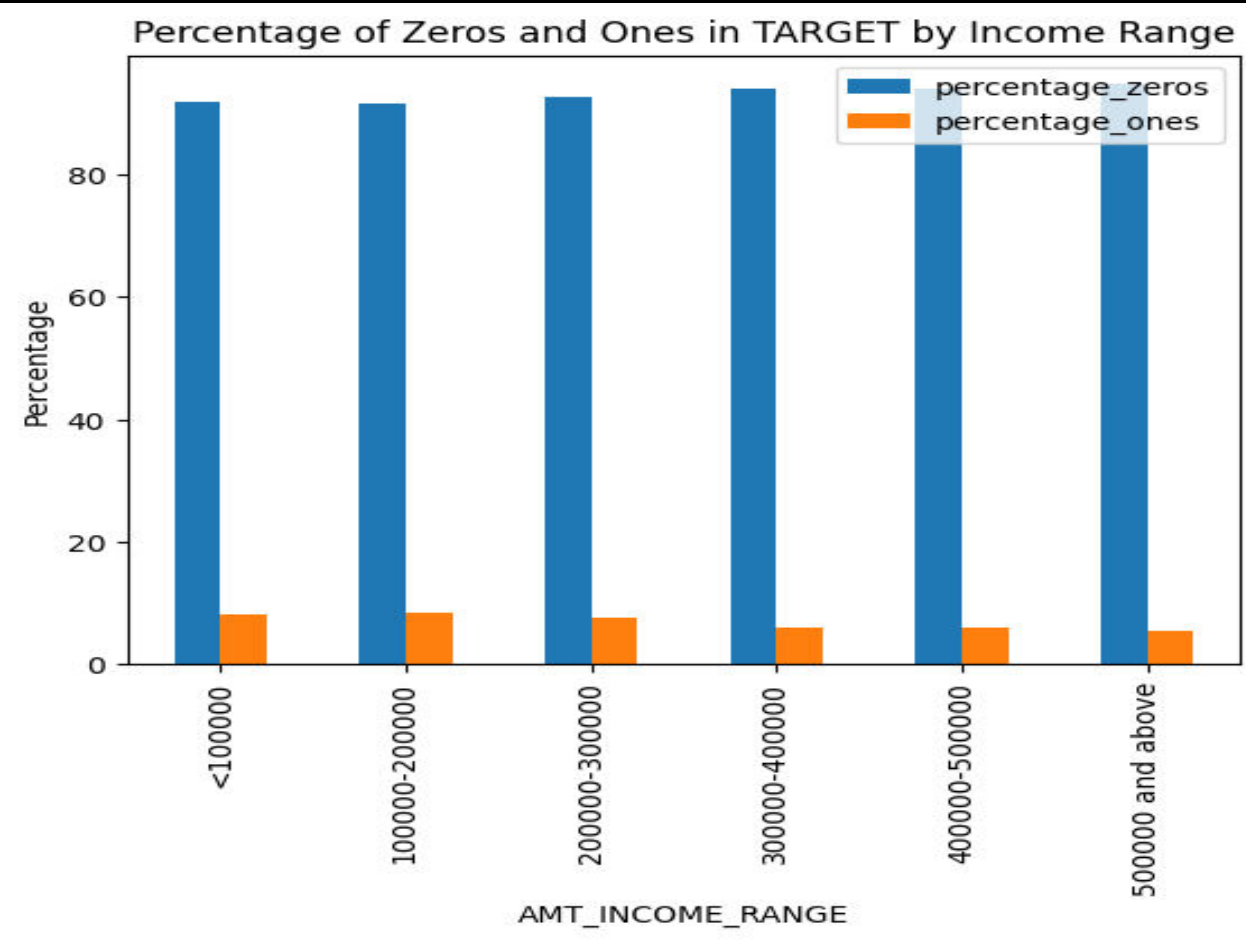
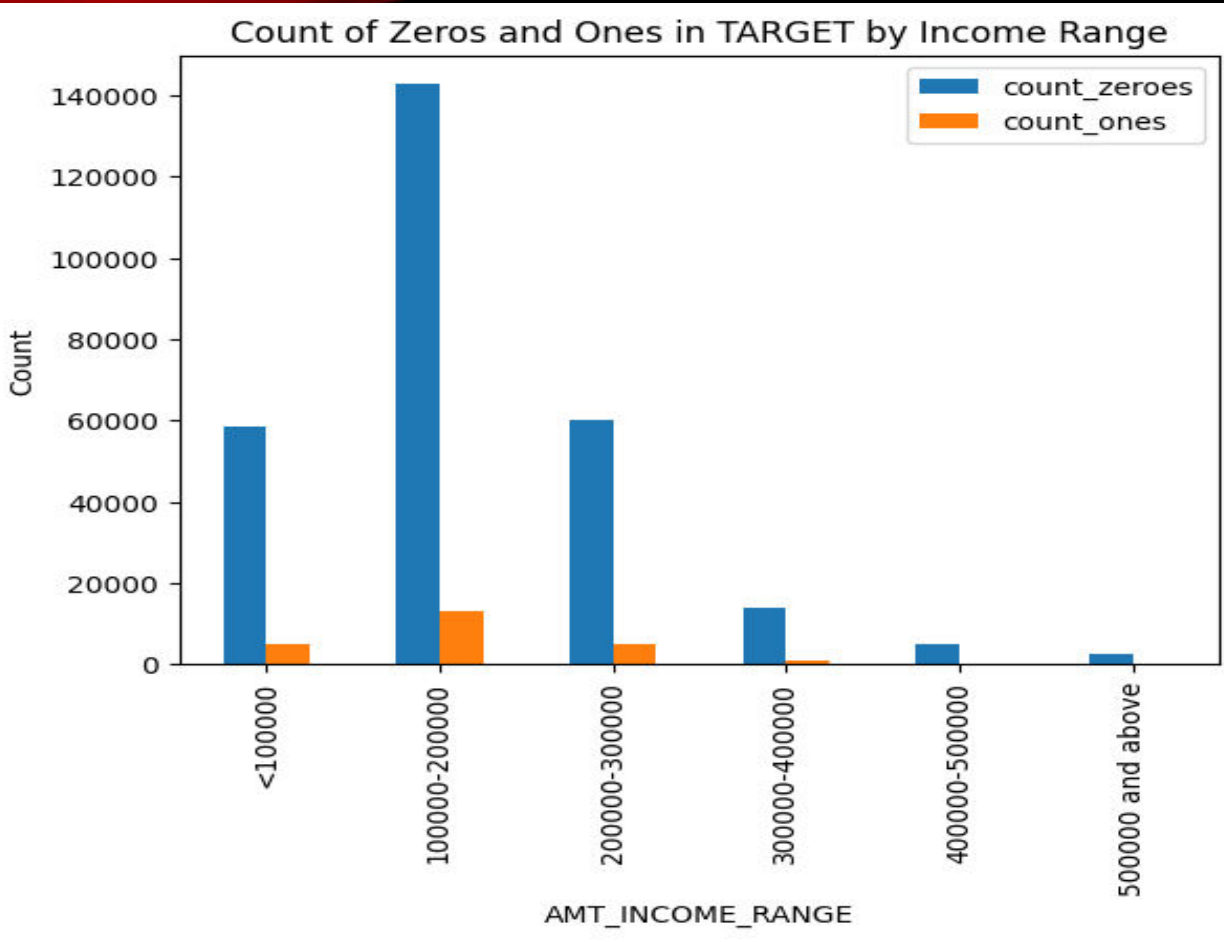
Observation :1. Working, Commercial associate, State servant and Pensioner are having lower percentage of defaulters in the same order as written.

2. Maternity is having highest percentage of defaulters followed by Unemployed Category.

3. Student and business have no defaulters.

Analysis

Income Range and Defaulter Relation:

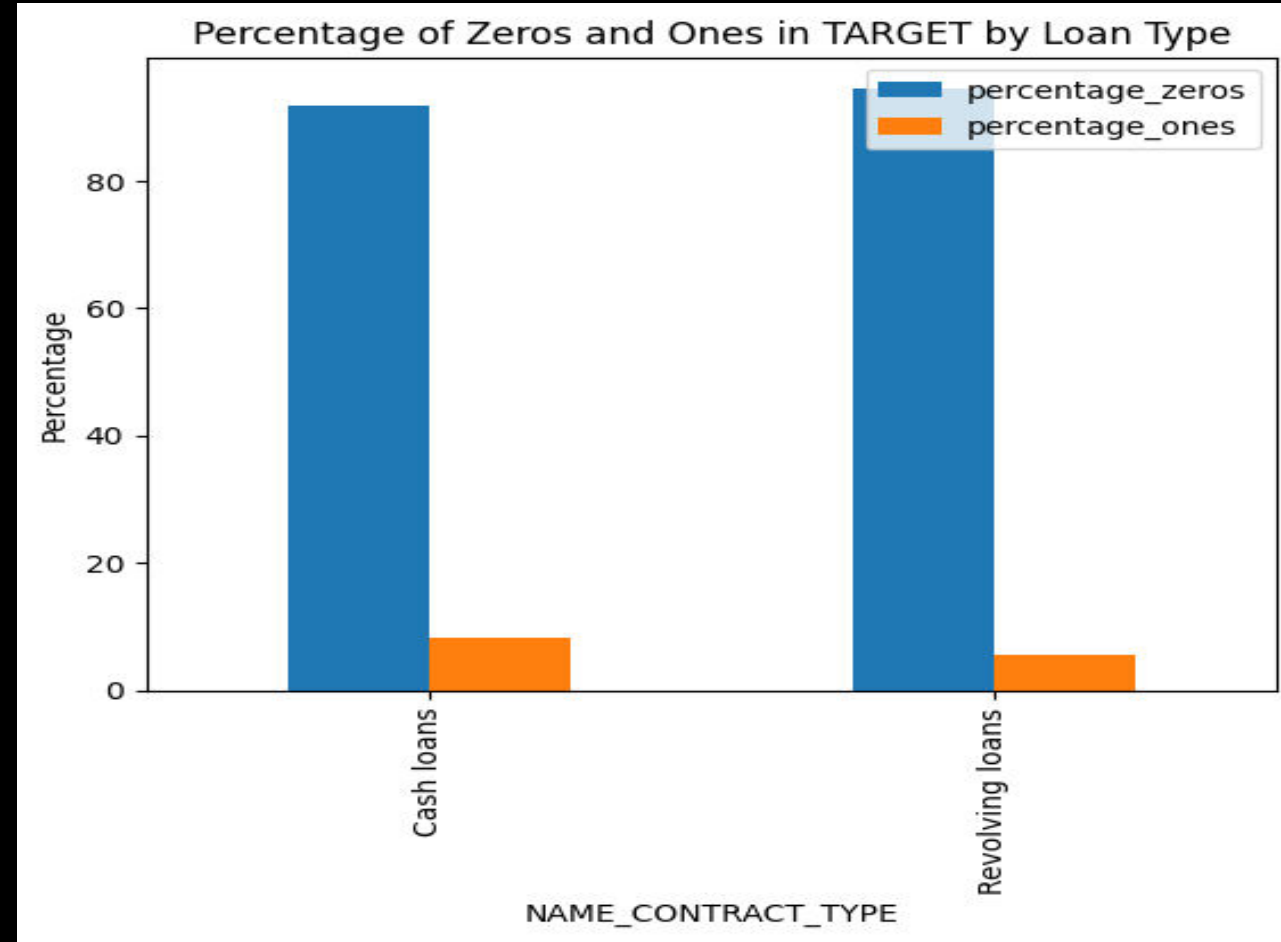
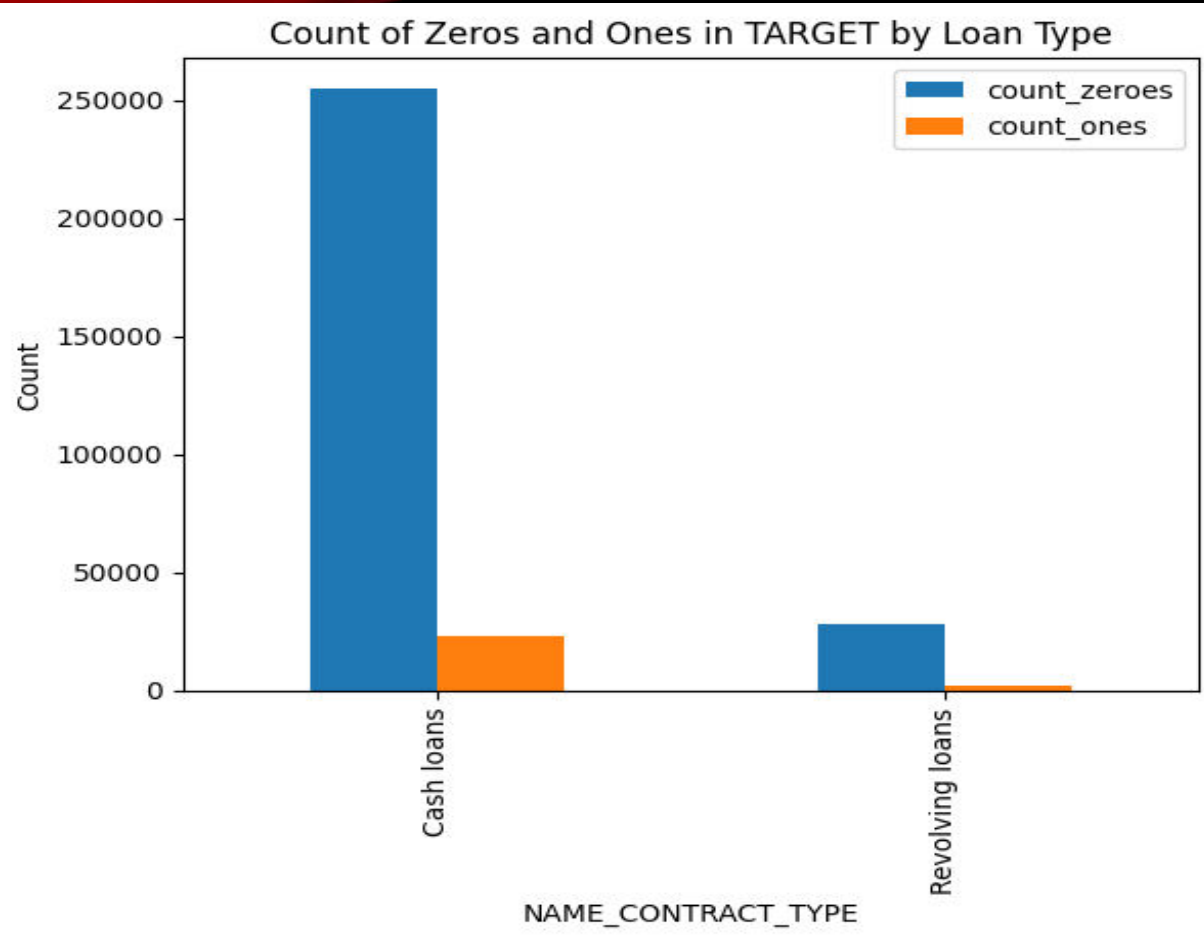


Observation :

1. People in range 100000-200000 have high number of loan and also have high number of defaulter.
2. People in range "500000 and above" has less defaulter percentage.

Analysis

Loan Type and Defaulter Relation:

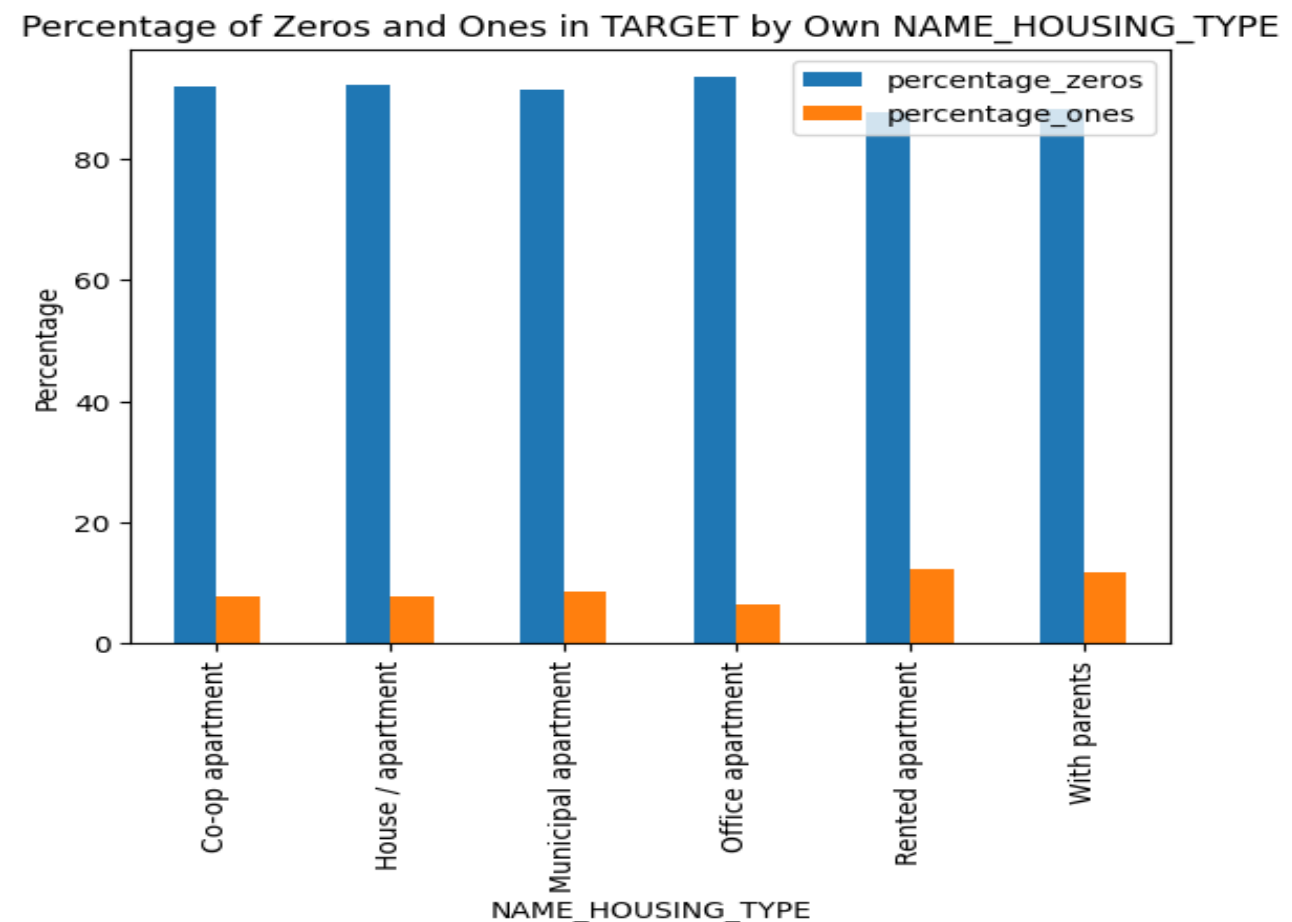
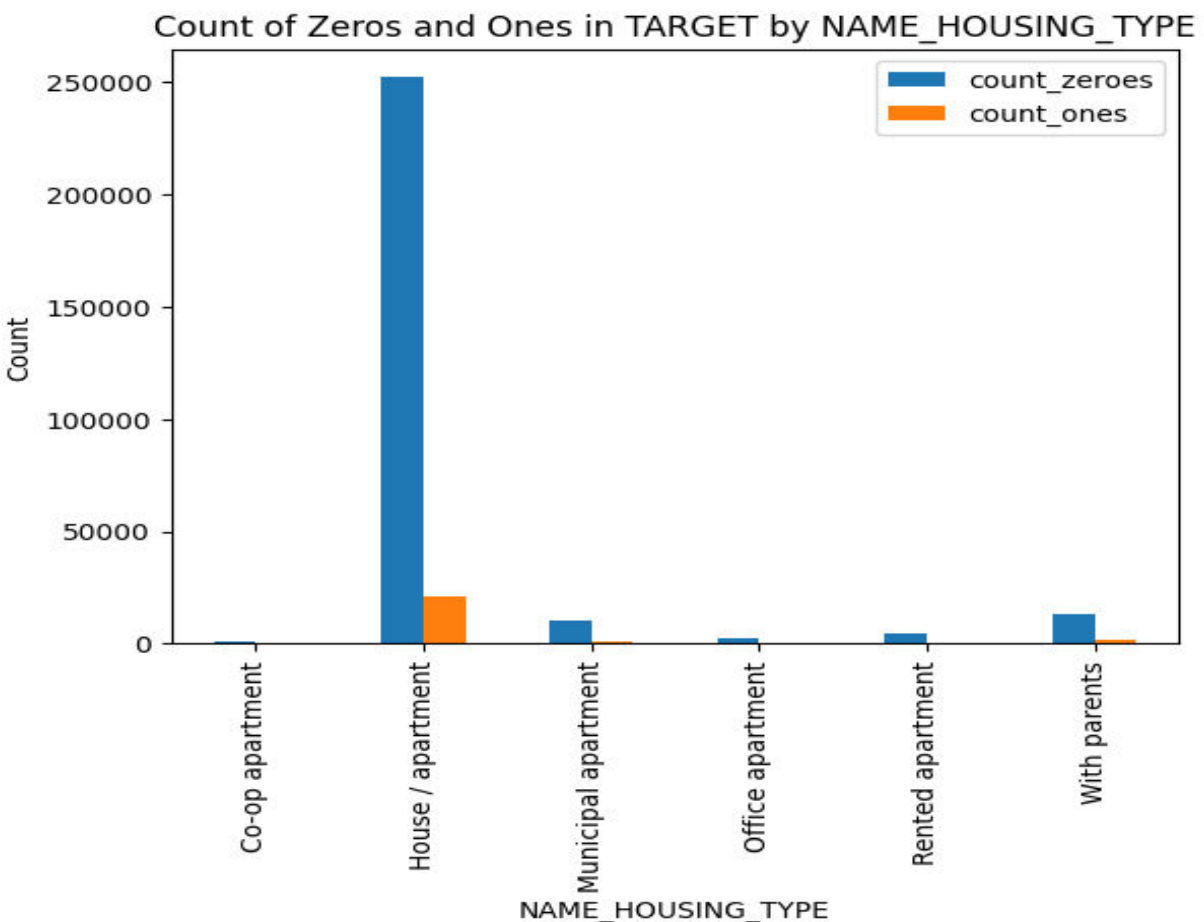


Observation :

1. The Defaulter percentage of "Cash loans" is higher than "Revolving loans".
2. People are more interested in "Cash Loans".

Analysis

Housing Type and Defaulter Relation:

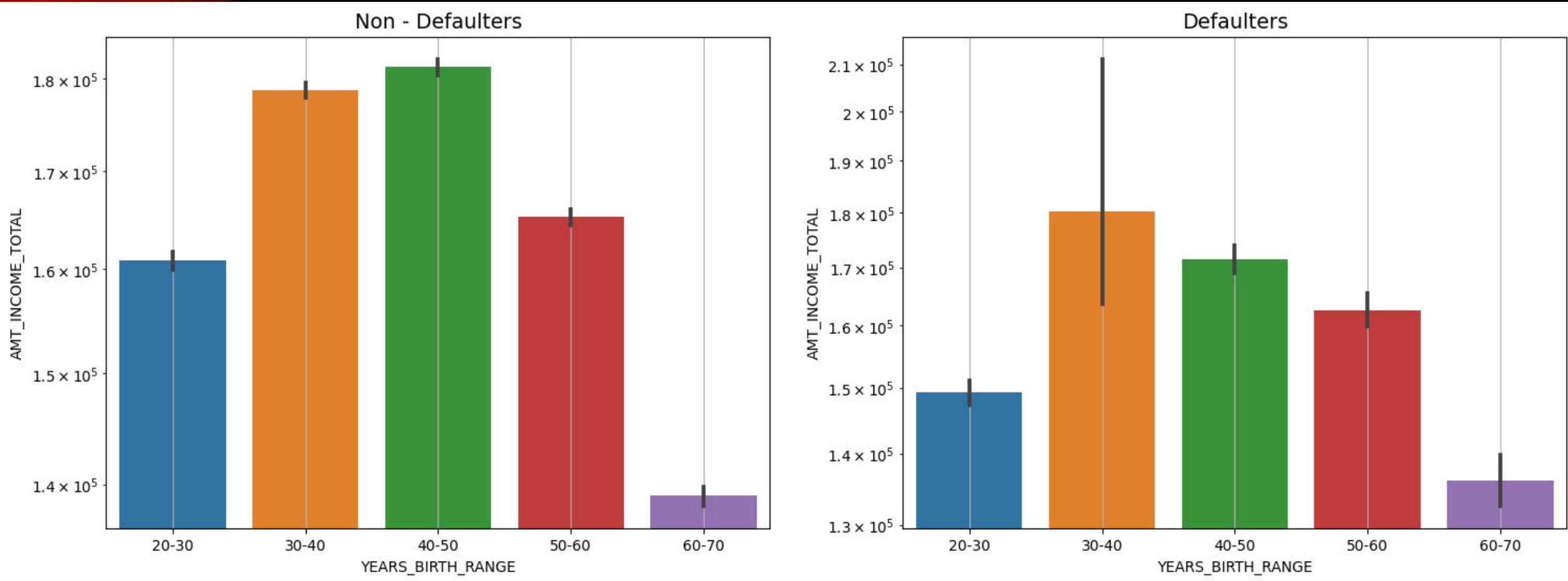


Observation :

1. People living in Office apartment has the lowest percentage of Defaulters.
2. People living in House/apartment have high number of loans and also high number of Defaulters.

Analysis

Age and Income :

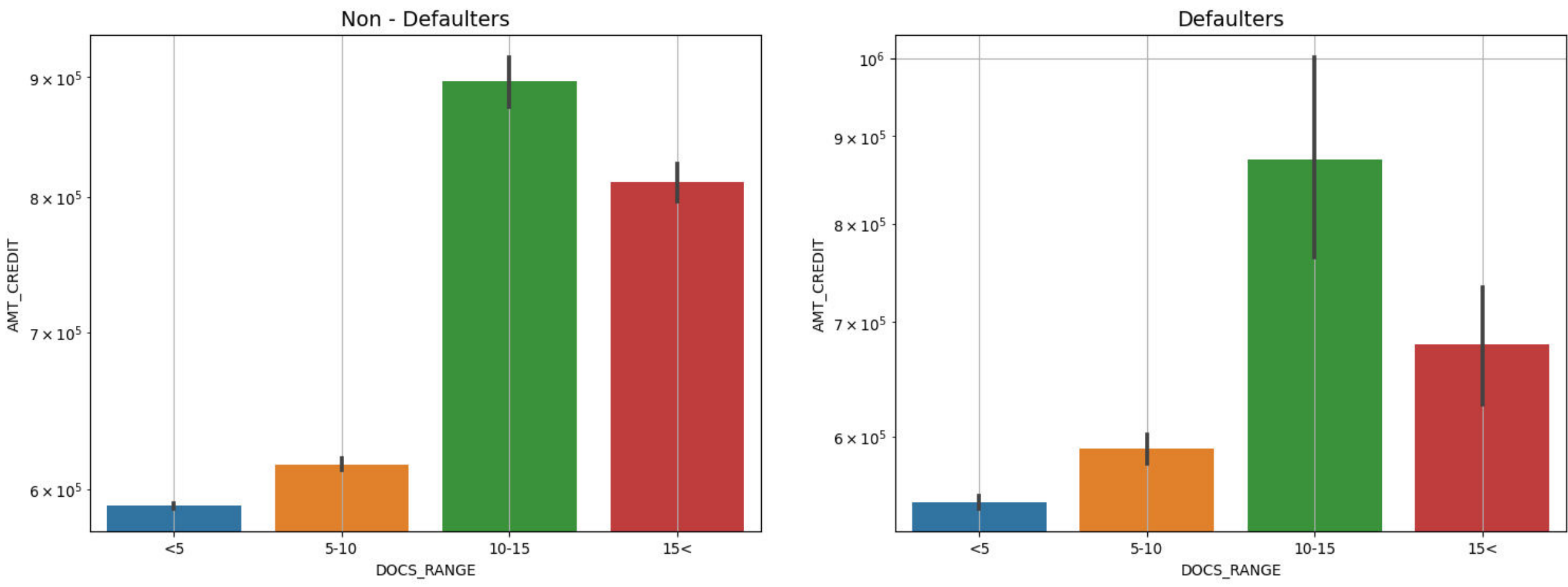


Observation :

1. People from age group 20 to 30 with over 150000 Income have higher chances for not being a Defaulter.
2. People from age group 40 to 50 with over 175000 Income have higher chances for not being a Defaulter.

Analysis

Document Submitted and Credit Amount :

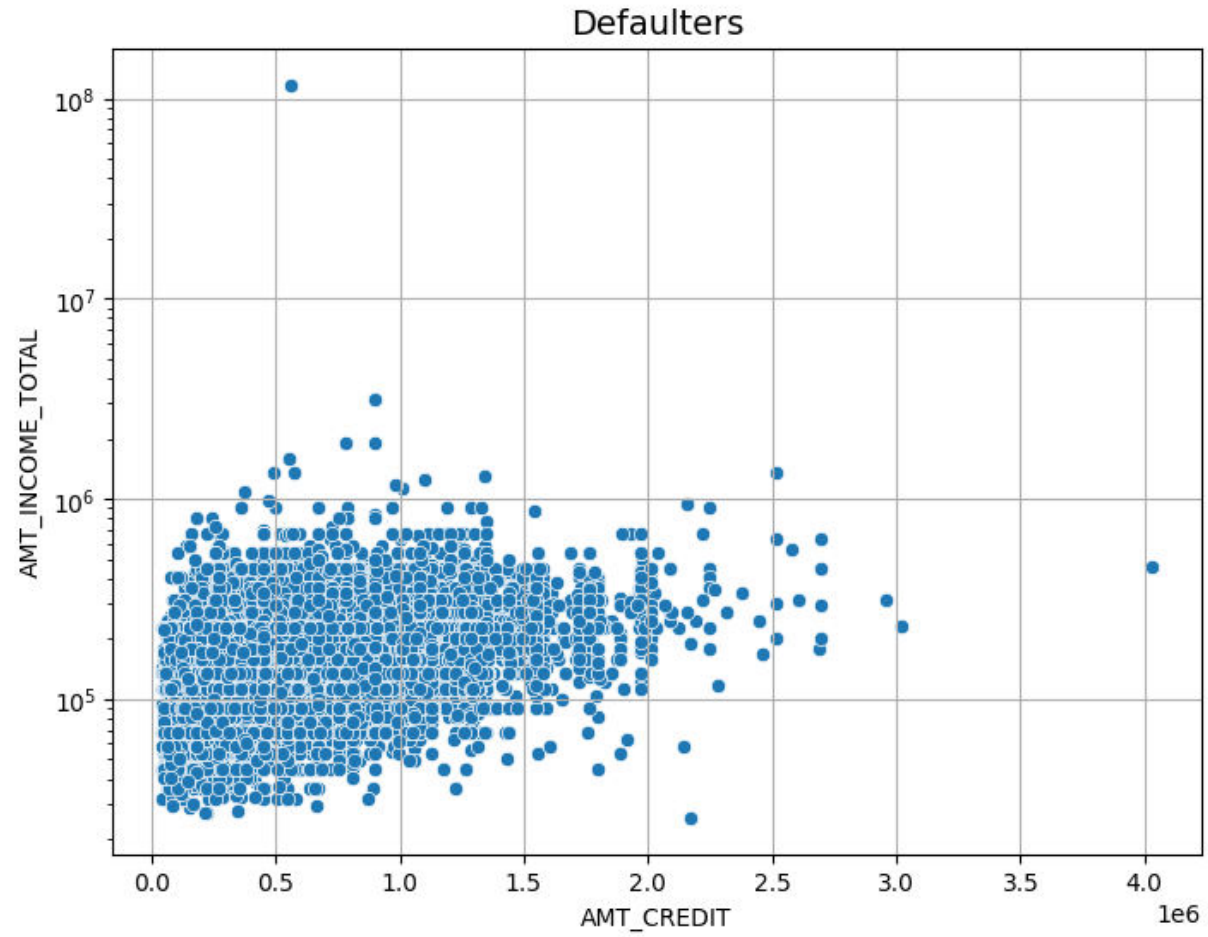
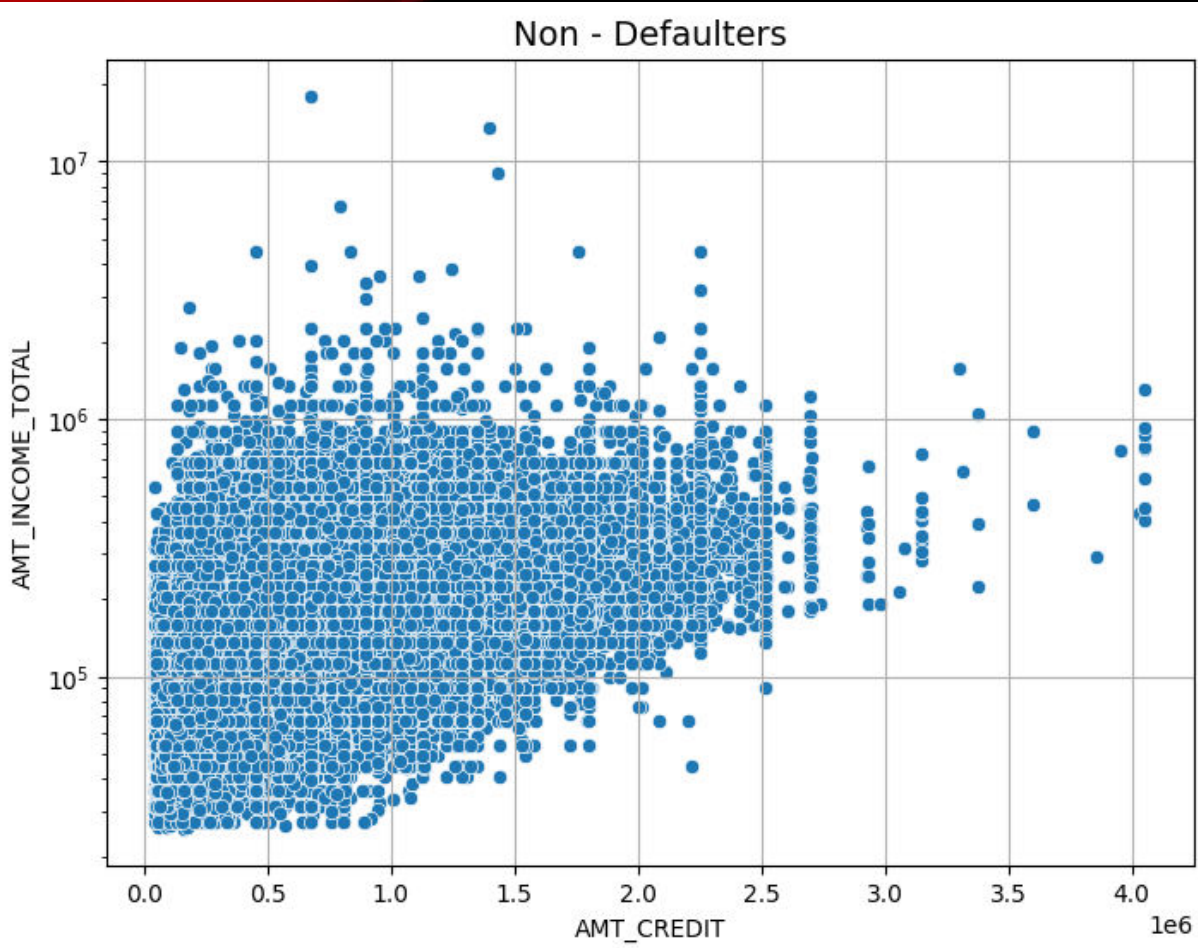


Observation :

People opting for over 700000 credit and providing 15+ documents have more chances of being a Non – Defaulter.

Analysis

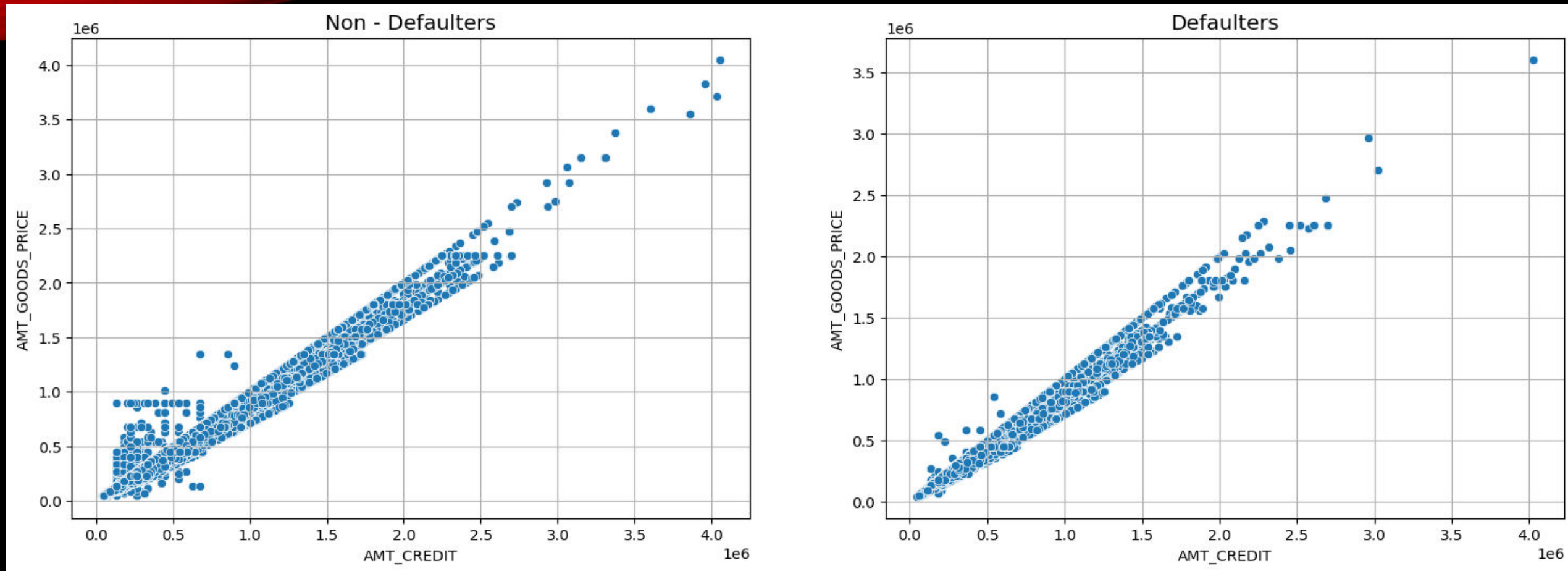
Credit Amount and Income :



Observation :
No decision can be made out of this information.

Analysis

Credit Amount and Goods Amount :

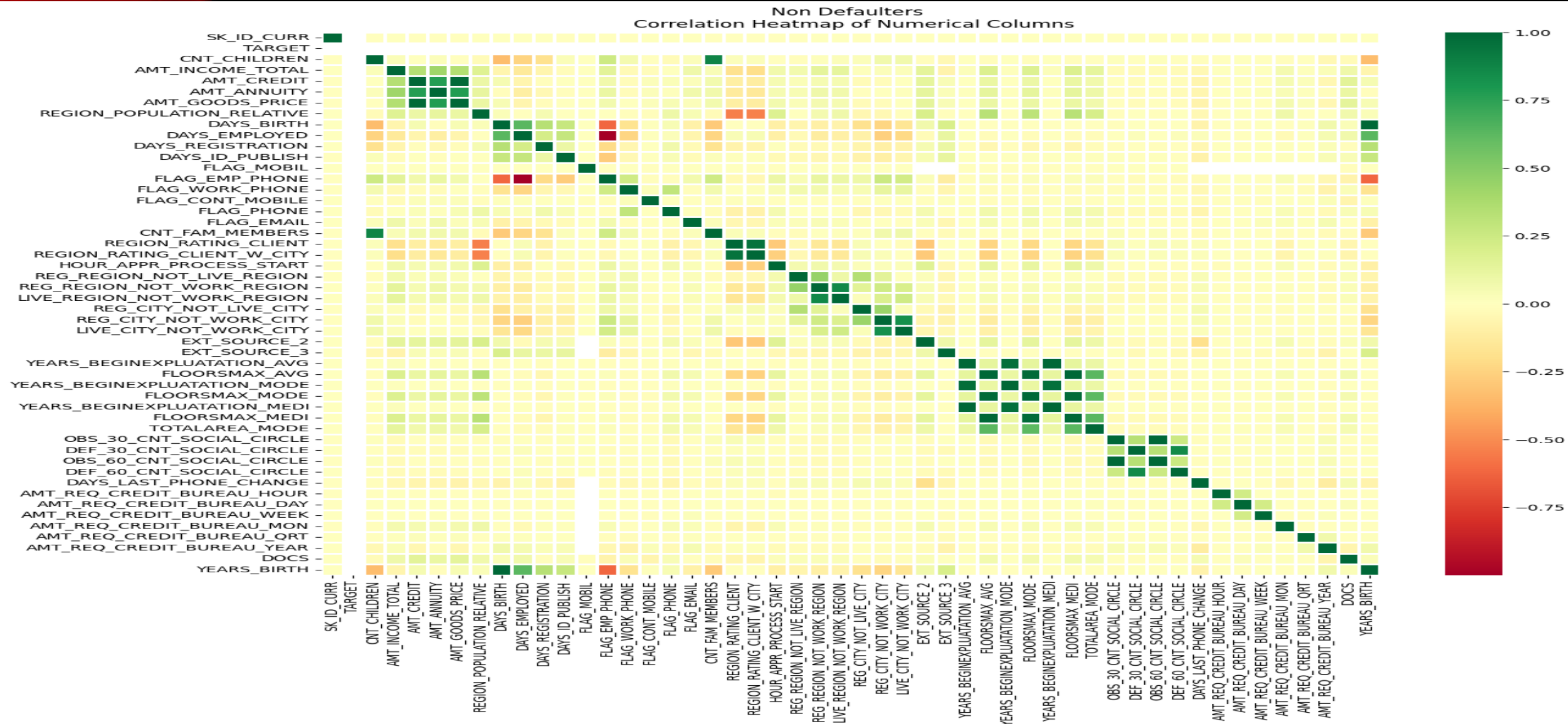


Observation :

1. As expected with increase in AMT Credited there is a raise in Goods Price as well.
2. Still it is not completely balanced, in many cases Credit Amount is higher than Goods Amount.
3. The number of defaulters reduces when the Goods Amount is higher than Credit Amount.

Analysis

Non – Defaulters Correlation :



Analysis

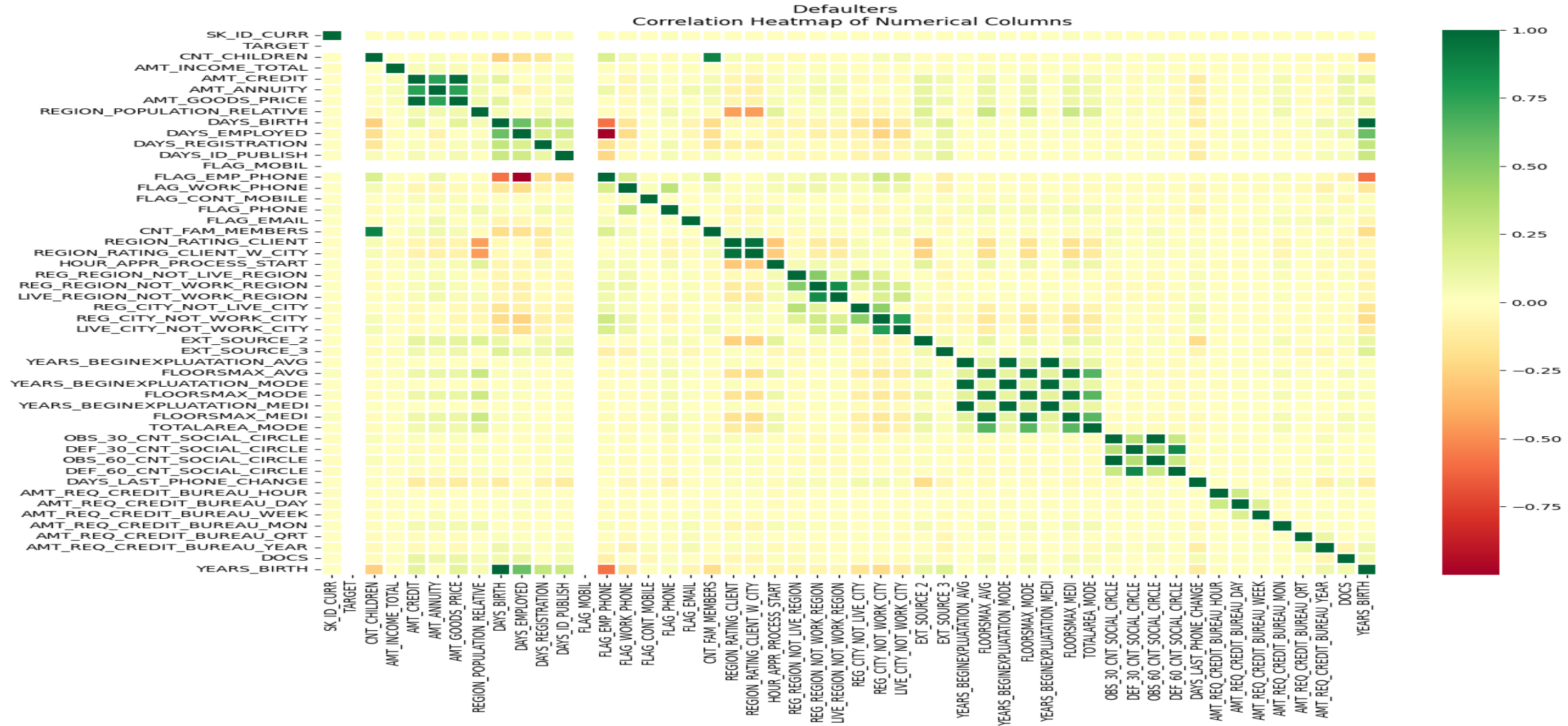
Non – Defaulter Correlation :

Top 5 Correlation Categories on the basis of above heatmap :

1. AMT_GOODS_PRICE - AMT_CREDIT
2. REGION_RATING_CLIENT_W_CITY - REGION_RATING_CLIENT
3. SOCIAL_CIRCLE_DEFAULT_RATE60 - SOCIAL_CIRCLE_DEFAULT_RATE30
4. CNT_FAM_MEMBERS - CNT_CHILDREN
5. AMT_GOOD_PRICE - AMT_ANNUITY

Analysis

Defaulters Correlation :



Analysis

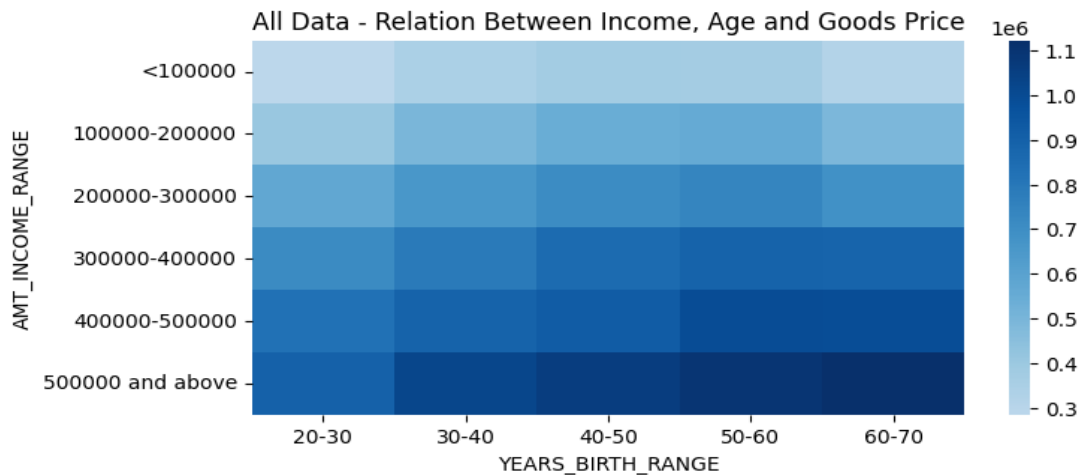
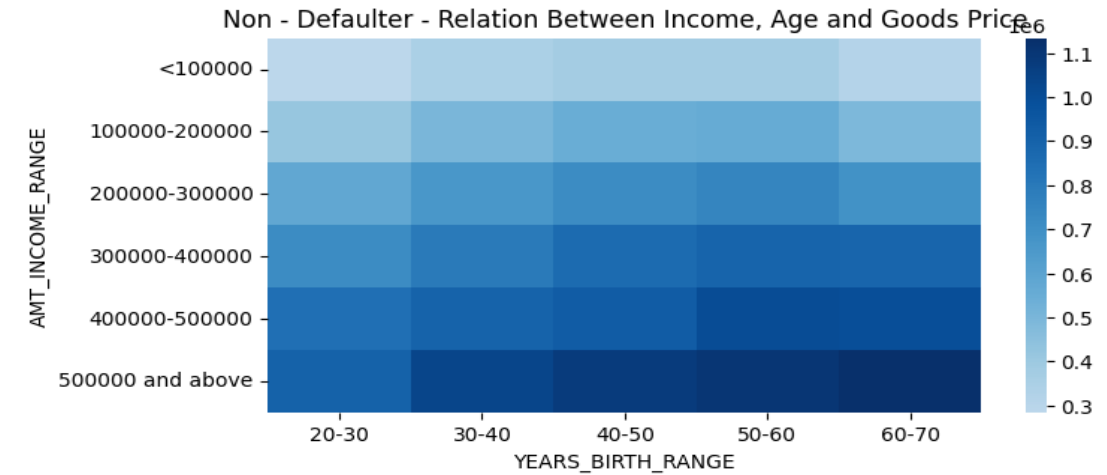
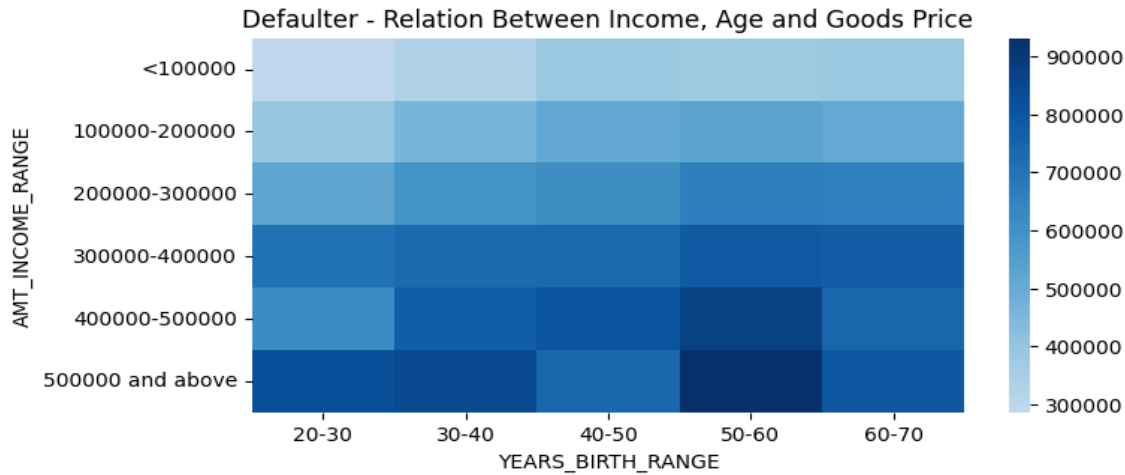
Defaulters Correlation :

Top 5 Correlation Categories on the basis of above heatmap :

1. *AMT_GOODS_PRICE - AMT_CREDIT*
2. *REGION_RATING_CLIENT_W_CITY - REGION_RATING_CLIENT*
3. *SOCIAL_CIRCLE_DEFAULT_RATE60 - SOCIAL_CIRCLE_DEFAULT_RATE30*
4. *CNT_FAM_MEMBERS – CNT_CHILDREN*
5. *AMT_GOOD_PRICE – AMT_ANNUITY*

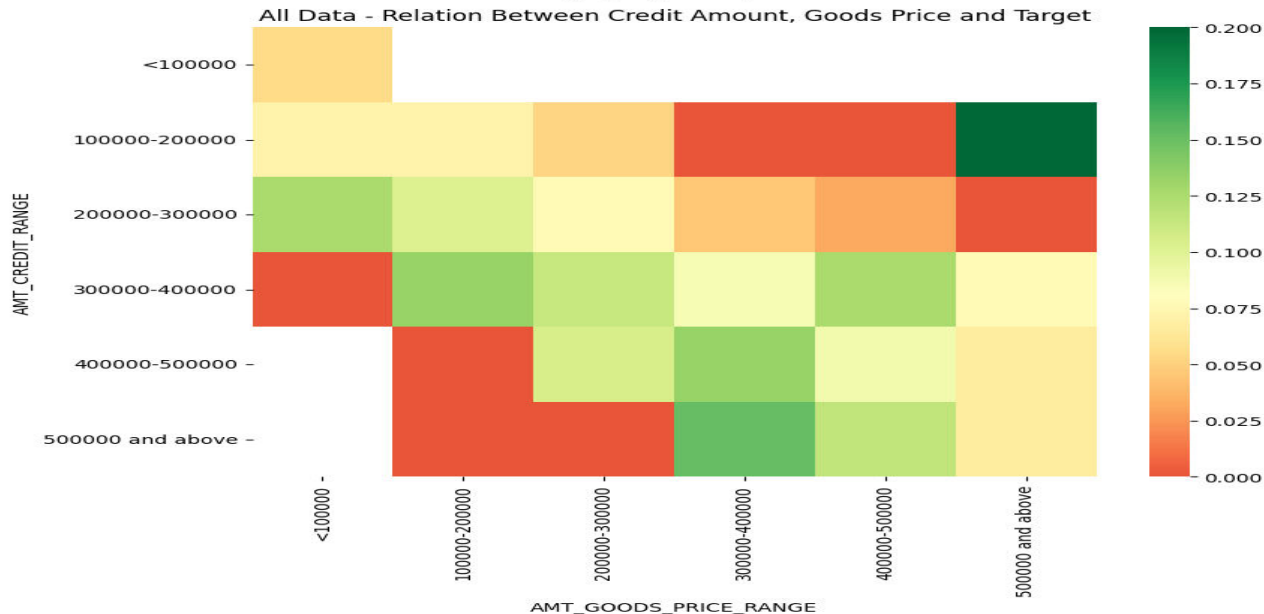
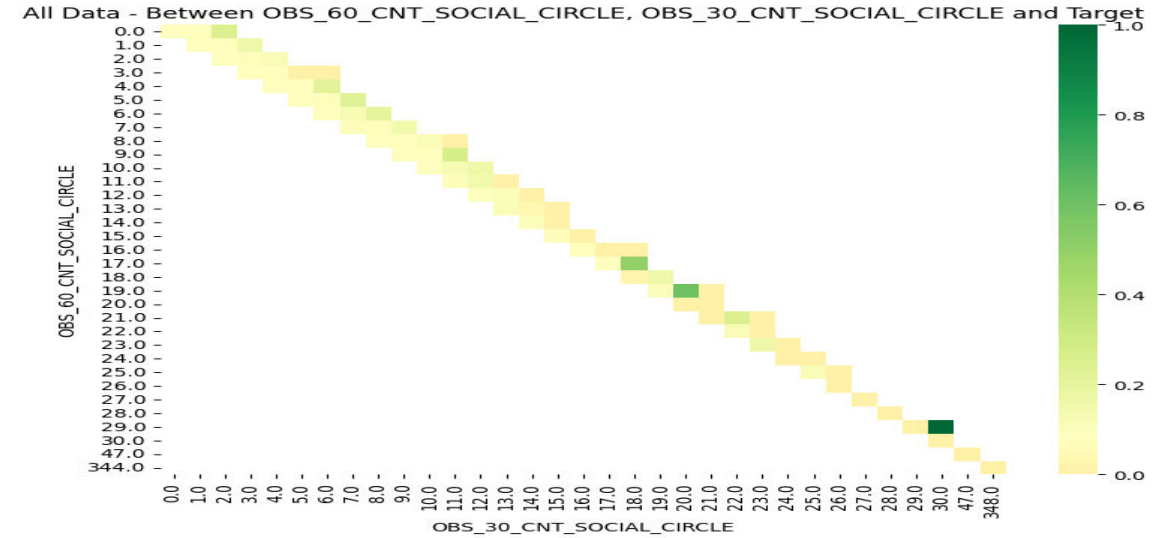
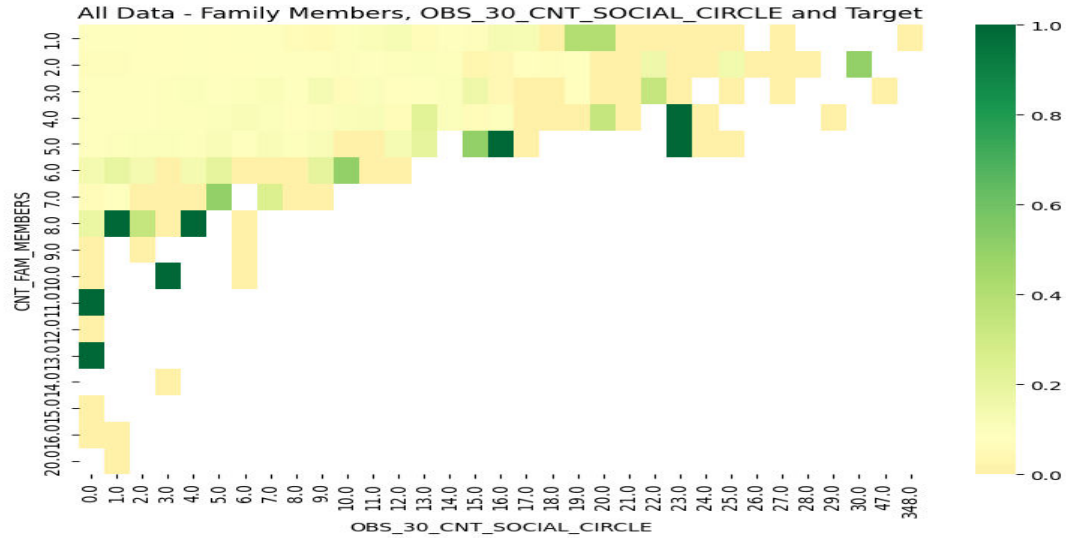
Analysis

Income, Age with values Goods Price :



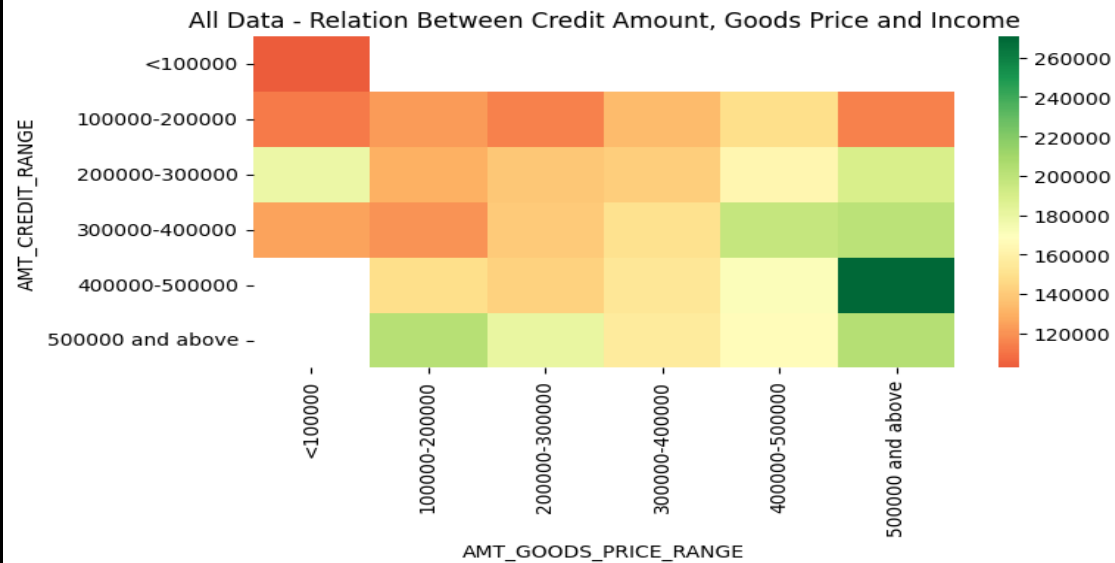
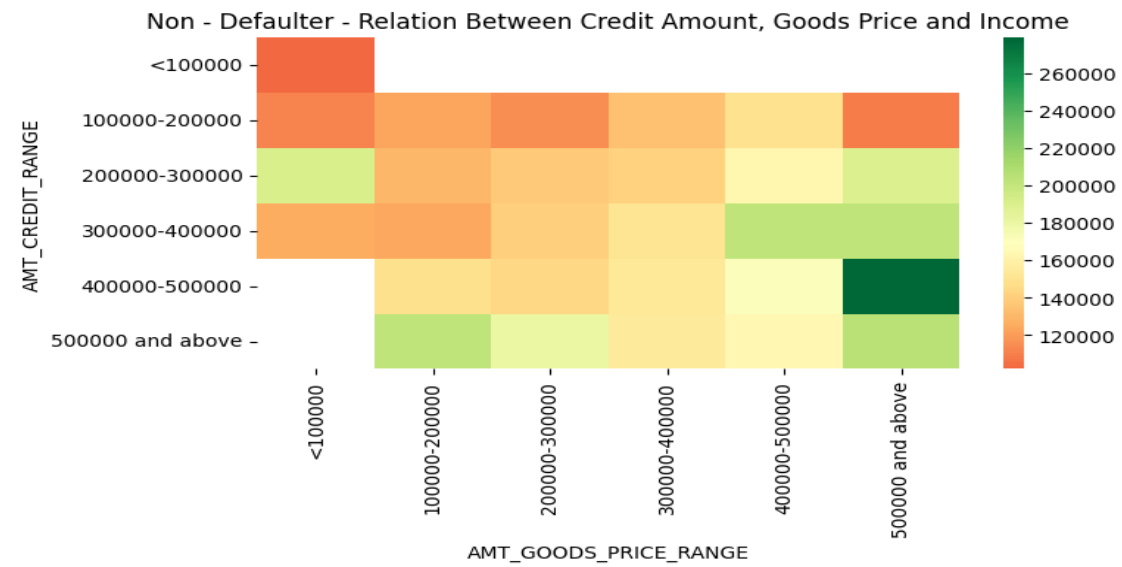
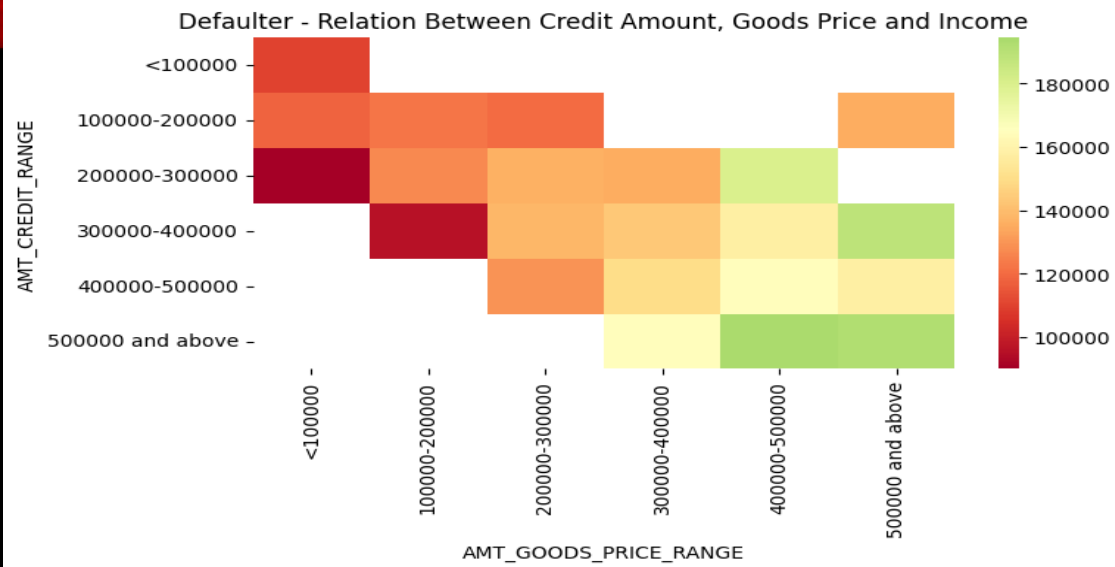
Analysis

Family Members, Social Circle 30 with values Target :



Analysis

Credit Amount, Goods Price with value Income:



Analysis

Combined Observations :

Observation :

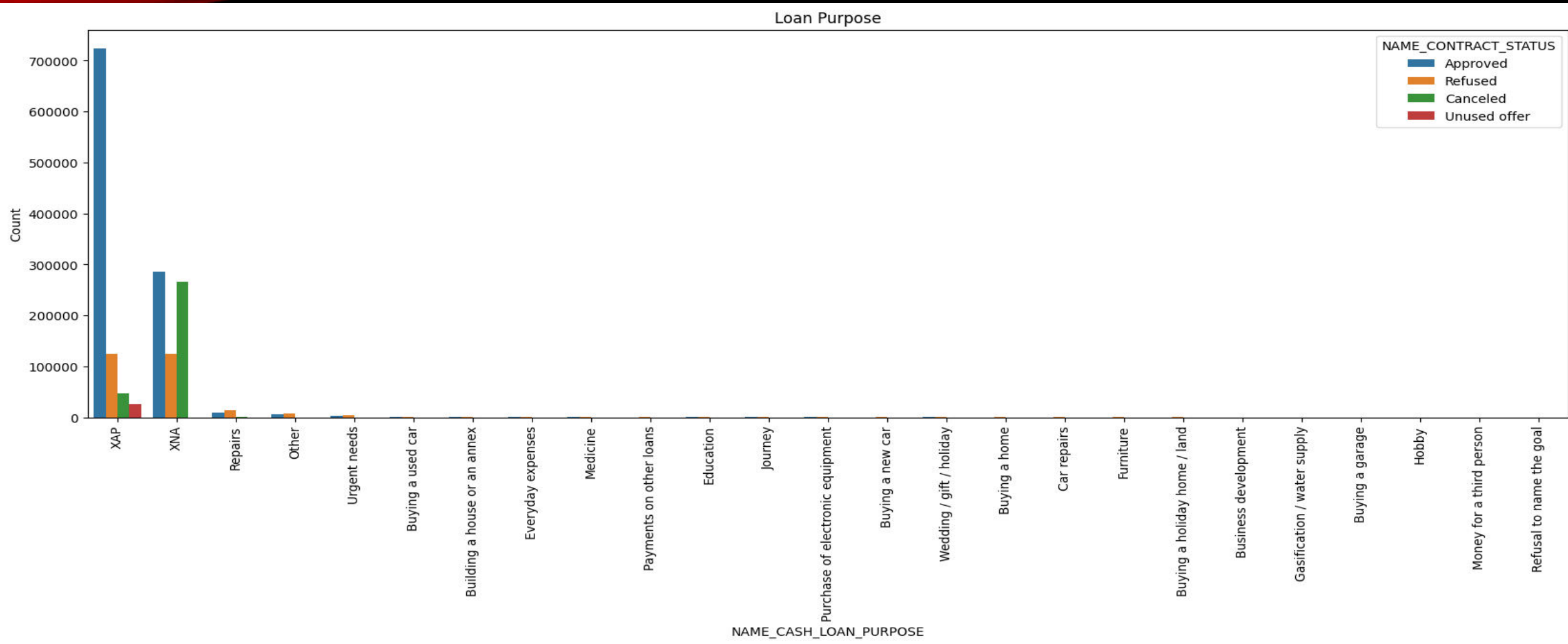
- 1. Within the age group of 50-60, individuals in the Very High income bracket exhibit a notable amount of credit. This trend may pose a risk of loss within the loan portfolio.*
- 2. Variables such as OBS_30_CNT_SOCIAL_CIRCLE and OBS_60_CNT_SOCIAL_CIRCLE reflect the client's social connections with observable 30/60 days past due (DPD). These variables demonstrate a clear correlation, particularly with Target 1, suggesting that during the approval process, this aspect warrants significant attention.*
- 3. The trend in DEF_30_CNT_SOCIAL_CIRCLE is upward, albeit with less data available for Target 1, resulting in a less densely populated graph.*
- 4. There appears to be a disproportionate increase in AMT_CREDIT and AMT_GOOD_PRICE relative to AMT_INCOME for Target 1, potentially contributing to defaults.*



File Name : Previous Application

Analysis

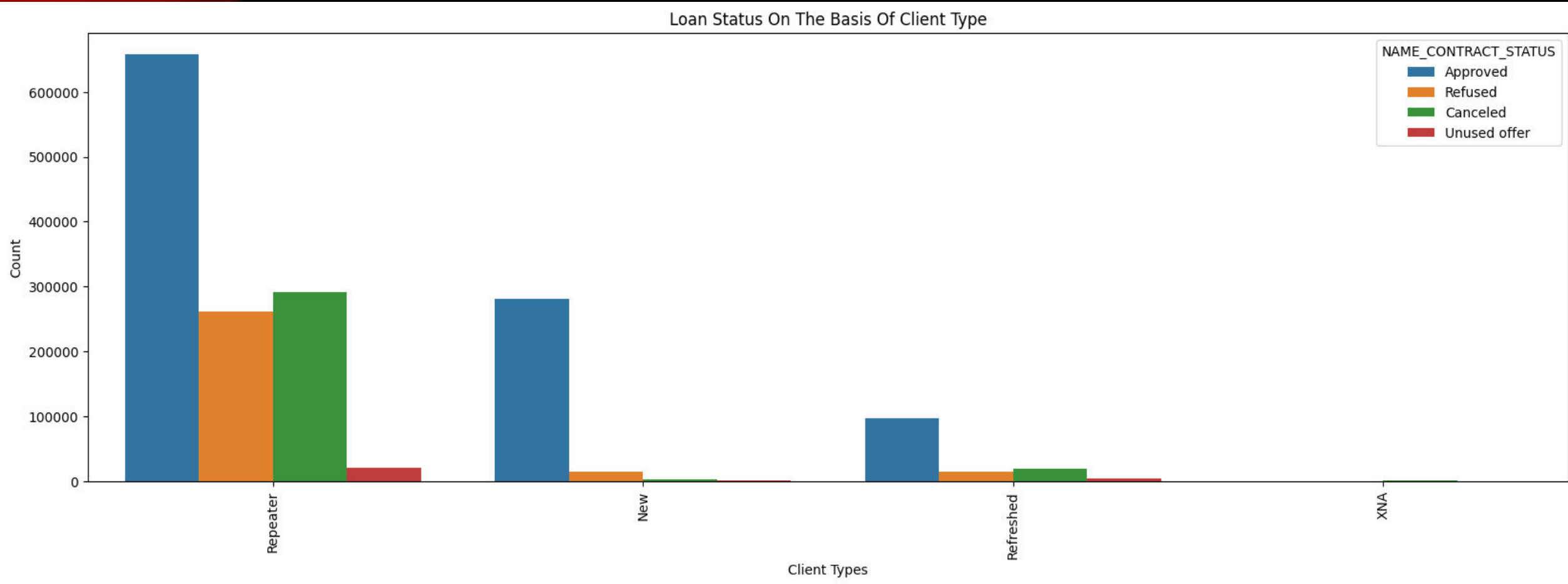
Loan Purpose and Contract Status:



Observation : "Repairs" been getting most rejected one

Analysis

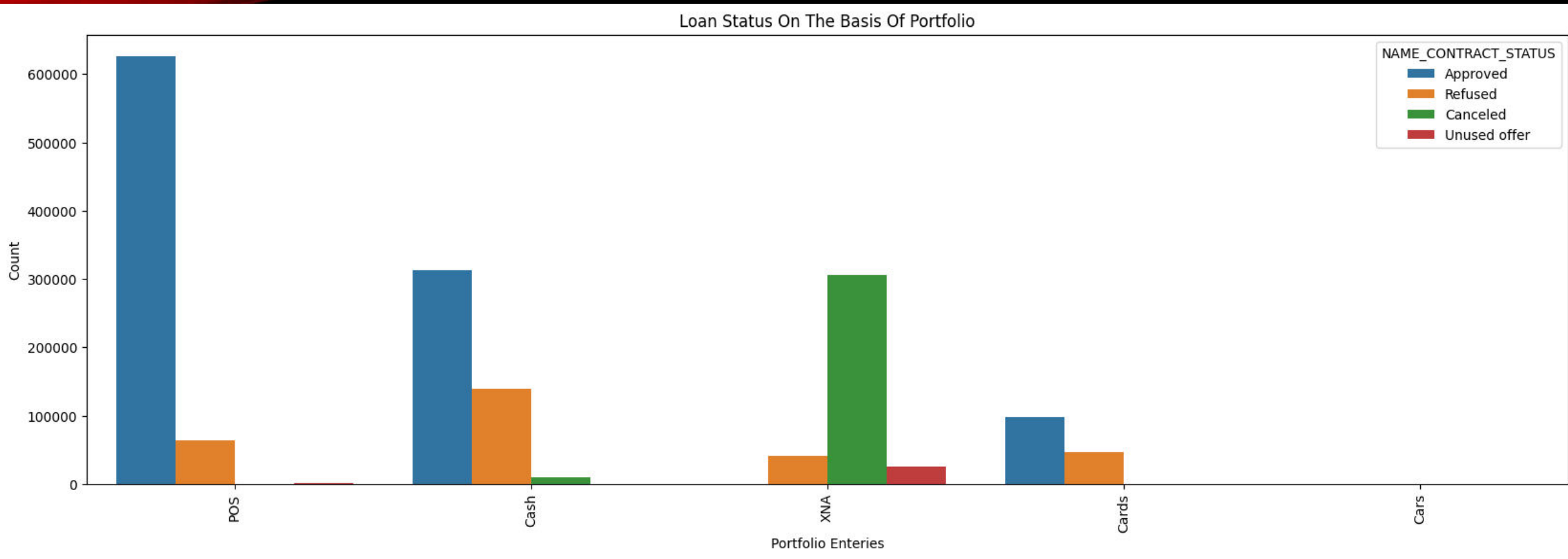
Client Type and Contract Status:



Observation : The bank has a larger number of repeat applicants across all categories: approved, refused, unused, and cancelled.

Analysis

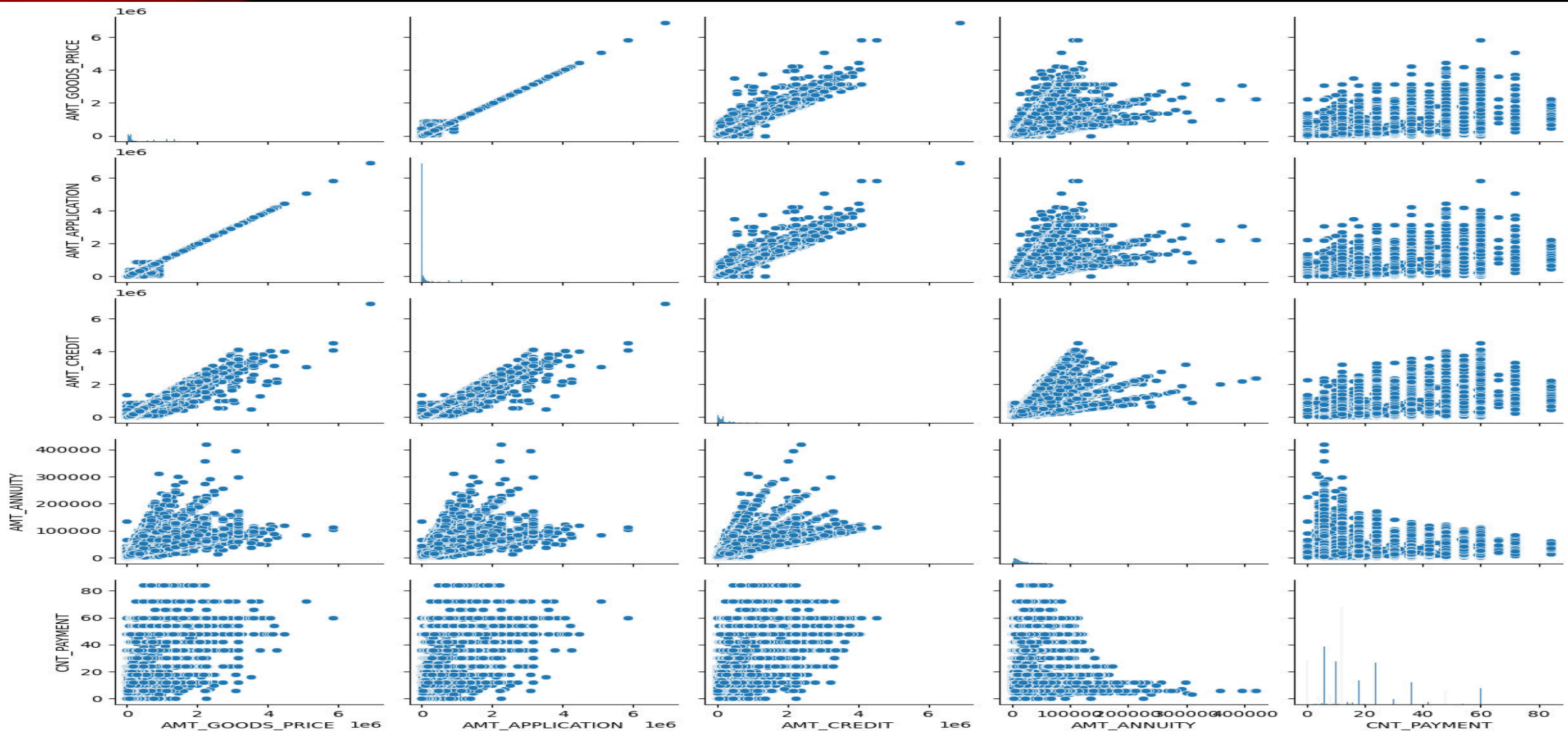
Portfolio and Contract Status:



- Observation : 1. Among approved applications, the consumer loan category exhibits the highest number of applicants.*
- 2. It appears that there are no cancelled loans in the cash loan category compared to the consumer loan category.*
- 3. A higher proportion of cash loans have been rejected compared to consumer loans.*
- 4. POS transactions appear to resemble consumer loans, and similarly, more cash loans have been rejected than*

Analysis

GOODS PRICE, APPLICATIONS, CREDIT, ANNUITY, PAYMENT :



Analysis

GOODS PRICE, APPLICATIONS, CREDIT, ANNUITY, PAYMENT :

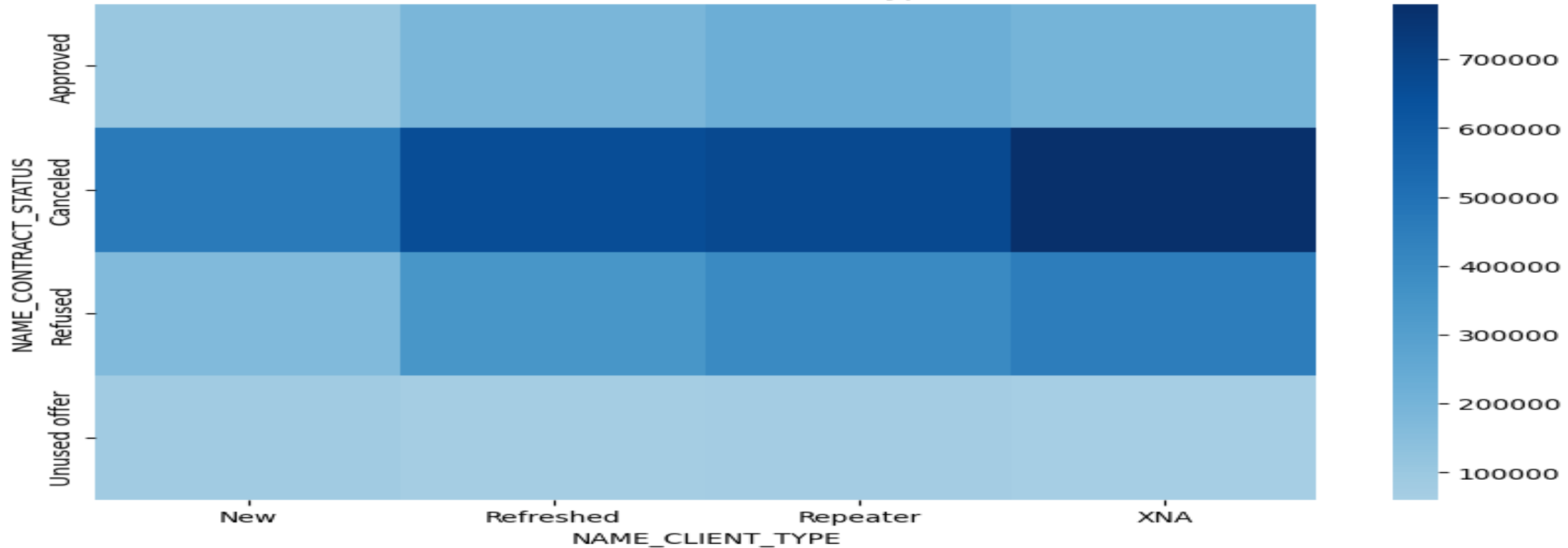
Observation :

1. *It is unsurprising to observe a strong correlation among variables such as AMT_GOODS_PRICE, AMT_ANNUITY, and AMT_APPLICATION. As the value of the goods being purchased increases, the demand for a larger loan amount naturally follows, leading to a correlation between these factors.*
2. *Likewise, there exists a notable correlation between AMT_CREDIT and AMT_GOODS_PRICE. This correlation is logical, as the credit amount typically aligns with the cost of the goods being financed.*
3. *One might expect to find a significant correlation between the column CNT_PAYMENT and AMT_CREDIT, indicating that higher credit amounts are associated with longer loan terms. However, upon examination of the data, such a correlation does not appear to be present.*

Analysis

Client Type, Contract Status with Value Goods Price:

Relation Between Contract Status, Client Type and Goods Price



Observation :

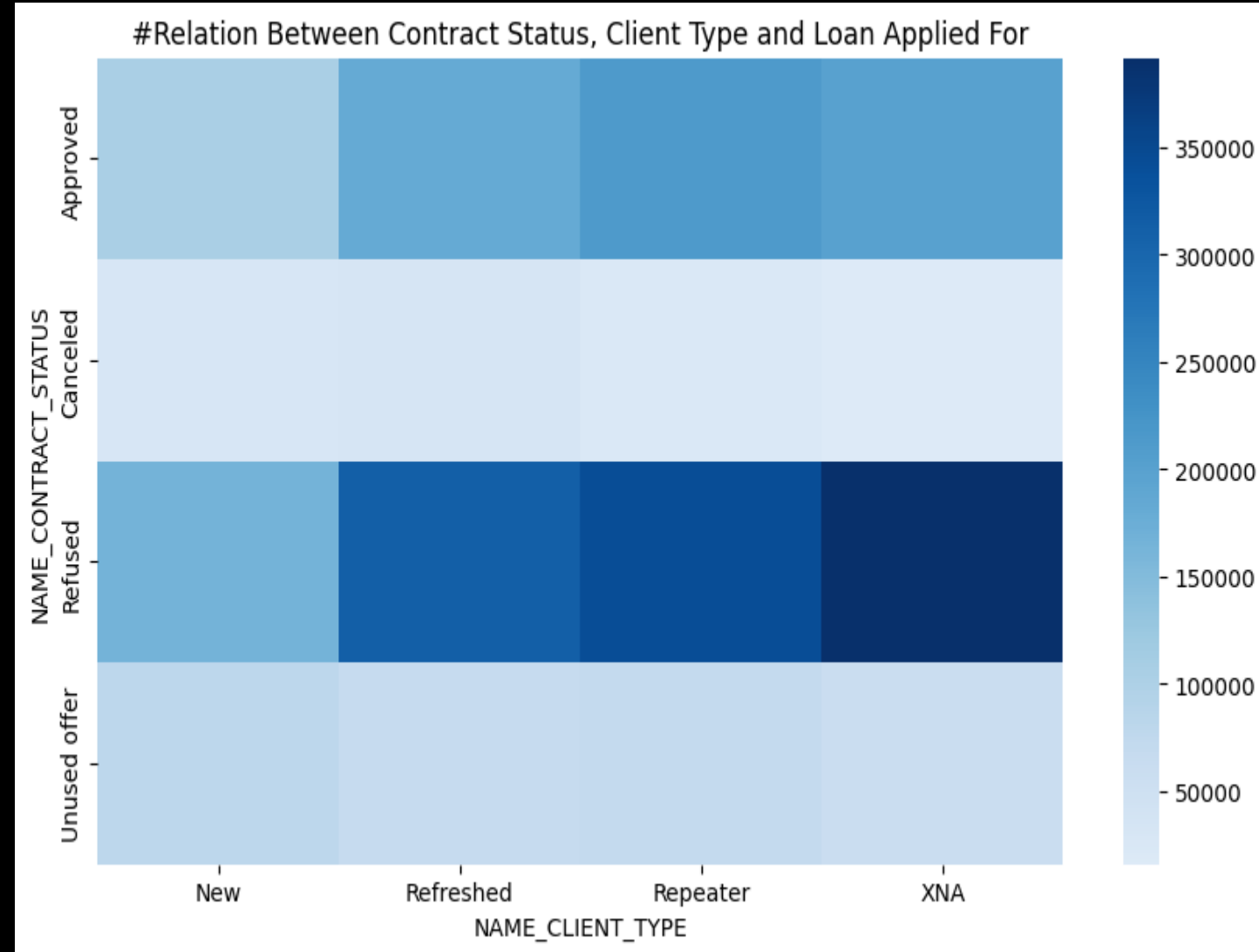
1. In comparison to other categories, all instances marked as cancelled or refused exhibit a higher value of goods.
2. The amount requested for unused offer applications is minimal.

Analysis

Client Type, Contract Status with value Loan Applied For:

Observation :

1. The amount requested for unused offer applications is minimal.
2. The amount of Refused applications is notably elevated fore Repeaters and Refreshed.
3. Repeat applicants tend to request higher amounts compared to new customers. This suggests that the bank may offer more favorable policies or interest rates to repeat applicants.





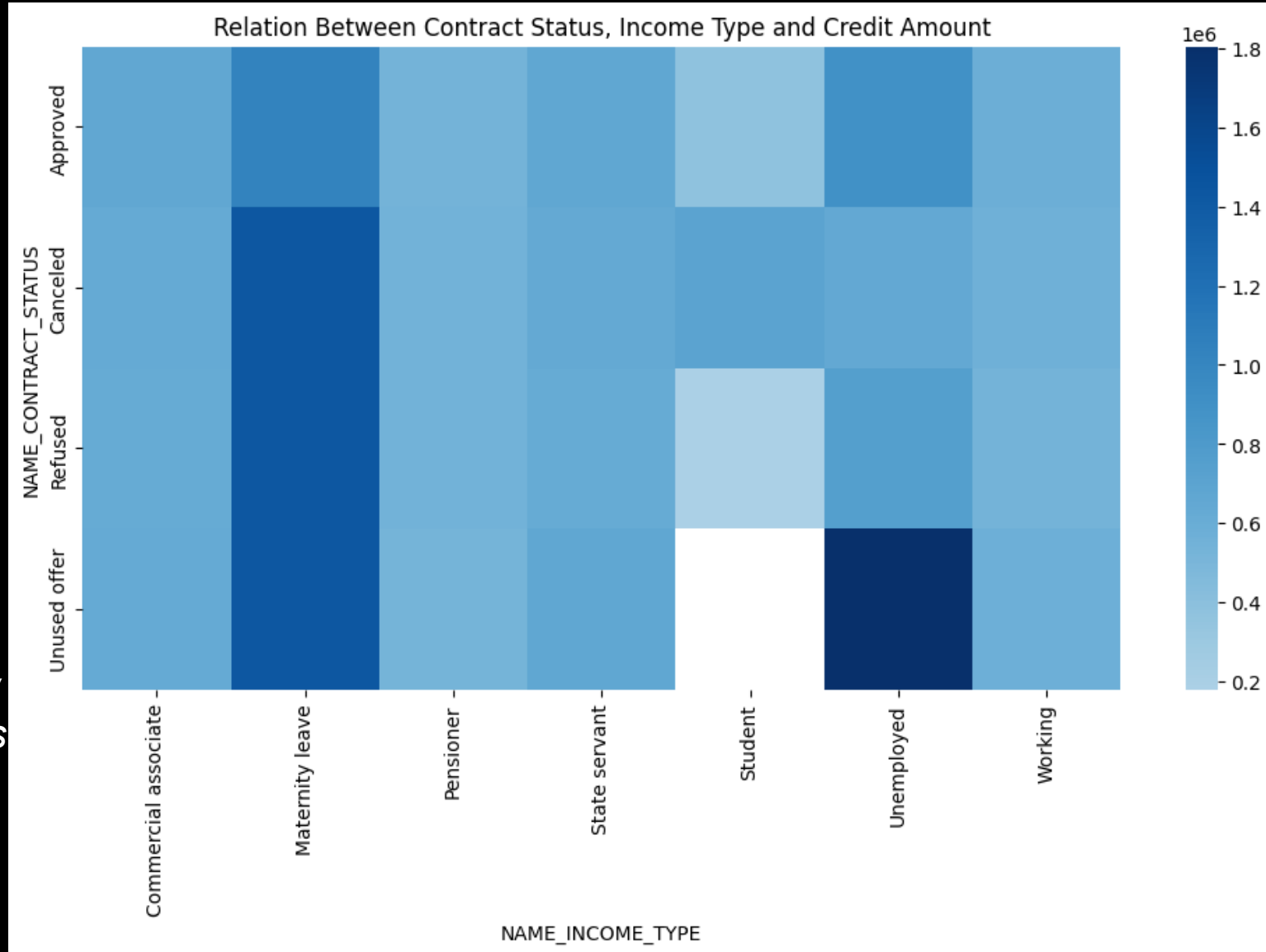
MERGIN APPLICATION DATA AND PREVIOUS APPLICATION

Analysis

Income Type, Contract Status with value Credit Amount :

Observation :

- 1. Unemployment and maternity leave are significant factors contributing to the provision of higher credit limits.*
- 2. Unused offers typically entail smaller credit values, potentially indicating why applicants are not utilizing them.*
- 3. Higher Credit Amount is being offered to Unemployed group which is Unused right now but could lead to higher numbers of Defaulters*

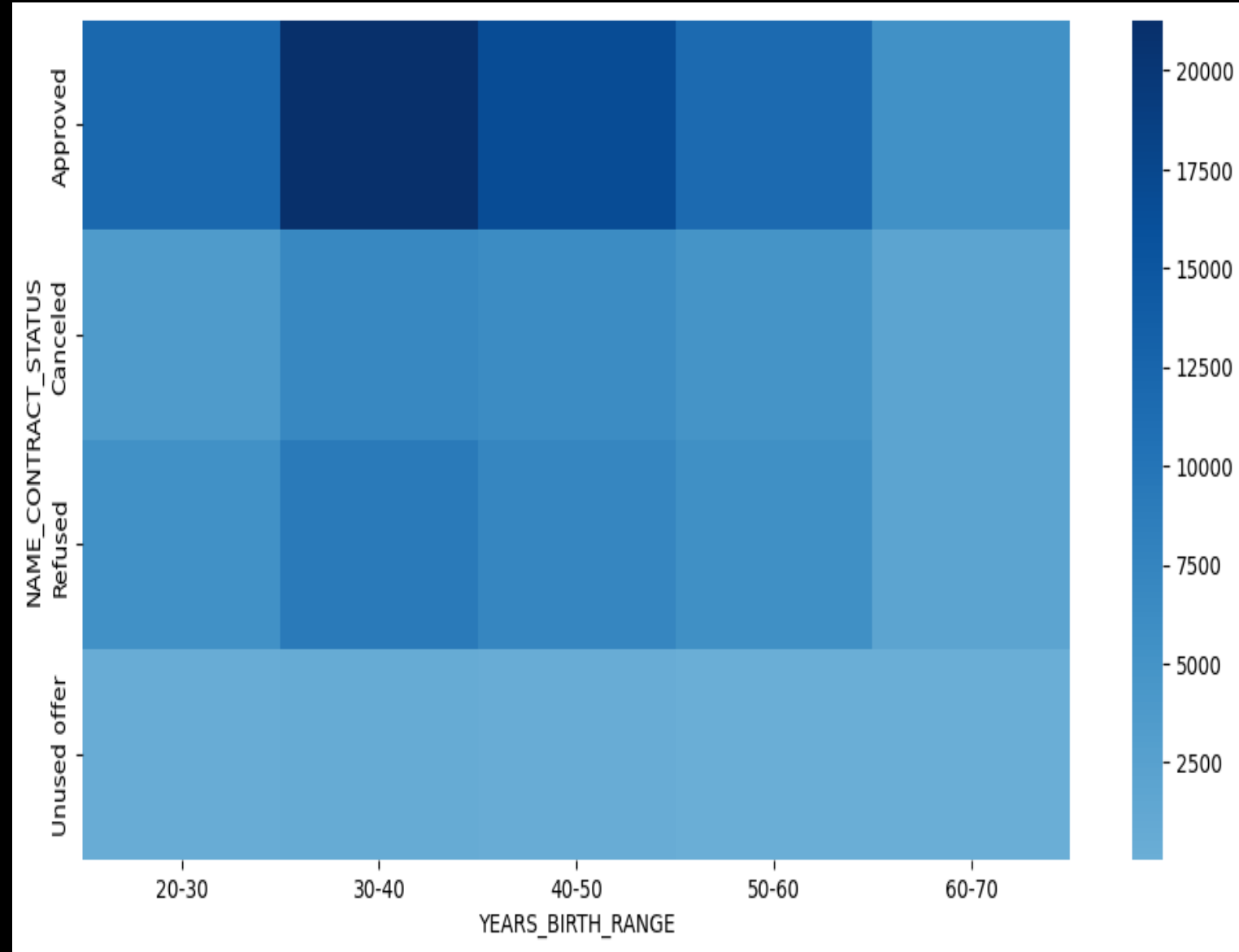


Analysis

Age, Contract Status with value Target :

Observation :

1. Given that Target 1 denotes defaults, higher values in the aforementioned matrix indicate a stronger correlation with default.
2. Age groups 30-40 and 40-50 with approved loan statuses exhibit elevated default rates.

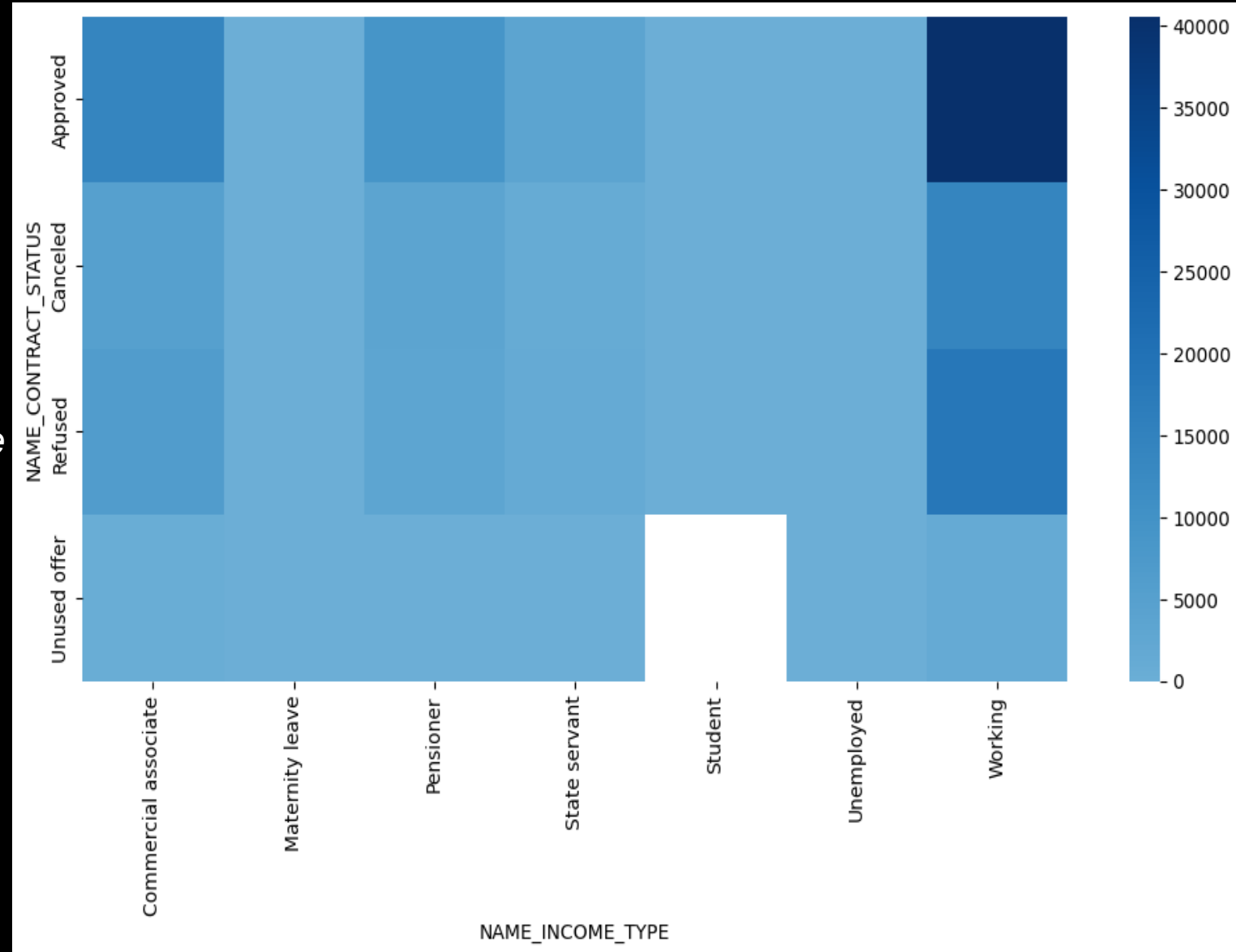


Analysis

Income Type, Contract Status with value Target:

Observation :

1. *Working Income Type with approved statuses have the highest default rates.*
2. *Previous applications with refusals, cancellations, or unused loans also experience defaults, posing a concern. This suggests that although the financial company rejected or cancelled previous applications, it approved the current ones, leading to defaults.*
3. *High number of applicants from the working class who were previously refused have now*





CASE SUMMARY // CONCLUSION

Summary / Conclusion

Possible Defaulters :

The following variables were identified in the analysis of the Application dataframe as leading indicators of default. When cross-referenced with approved loans that resulted in defaults, these findings were validated:

- 100000 – 200000 income earner*
- Age groups 30-40 and 40-50*
- Male gender*
- Unemployment status*
- Occupations such as laborers, salesmen, and drivers*
- Business type 3*
- Absence of property ownership*

Other Key Considerations :

- Higher Credit Amount is being offered to Unemployed group which is Unused right now but could lead to higher numbers of Defaulters.*
- Number of Bureau hits within specific periods: Zero hits are considered favorable.*
- Disparity between income and goods purchased: Low income coupled with high goods value raises concerns.*
- Previous applications with refusals, cancellations, or unused loans also exhibit defaults, indicating a noteworthy concern. This suggests that although the financial company refused or cancelled previous*

Summary / Conclusion

Possible Credit Worthy Group :

Key Consideration :

- Banks should prioritize customers with contract types such as 'Student,' 'Pensioner,' and 'Businessman' to enhance profitability.
- To mitigate financial losses, banks should reduce emphasis on customers with income type 'Working,' which exhibits the highest frequency of unsuccessful payments.
- Application Status 'Unused' typically involve lower loan amounts. Which could be a reason why it is not used.
- Female applicants demonstrate lower default rates and should be given additional consideration.
- Approximately 60% of defaulters are working applicants. However, this does not imply that working applicants should be automatically refused. Proper scrutiny of other parameters is necessary.
- Previous applications with refusals, cancellations, or unused loans also include cases where payments are made on time in the current application. This suggests that incorrect decisions may have been made in those instances.



THANK YOU