Case Study For Credit analysis using EDA

By Disha goyal

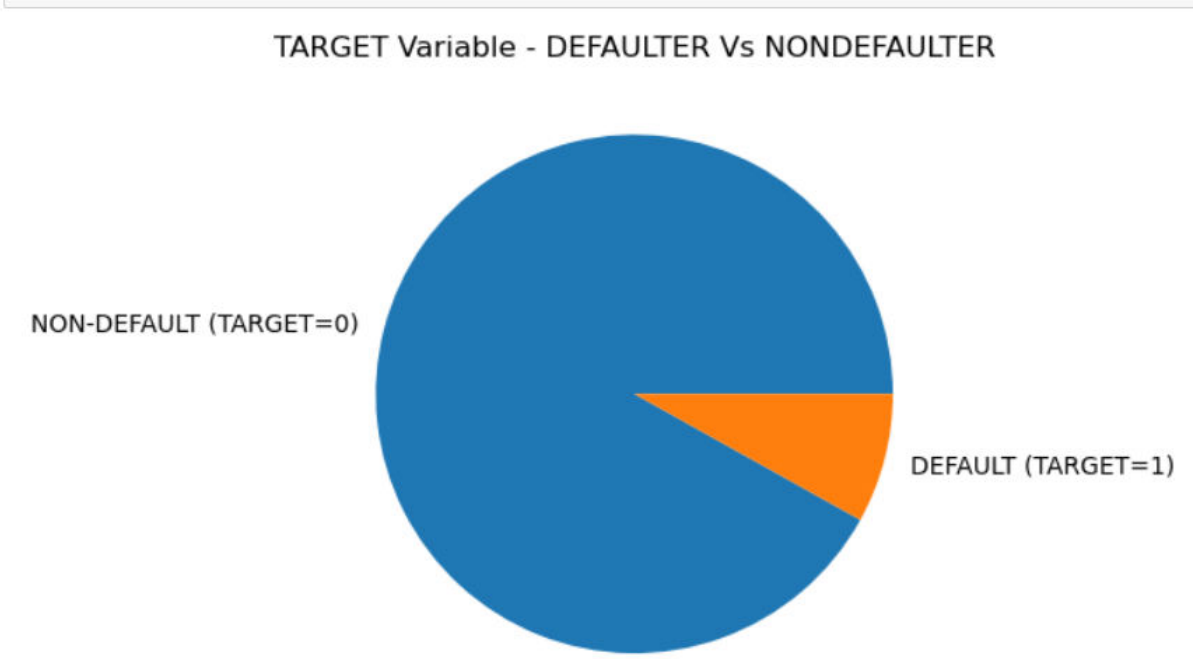
Introduction

- This analysis based on the previous history and present applications of clients.
- The purpose of this analysis to help the company to make decision for loan approval.
- By this analysis company controls the loss.
- This EDA will ensure that the consumers capable of repaying the loan are not rejected.

Steps involved in EDA

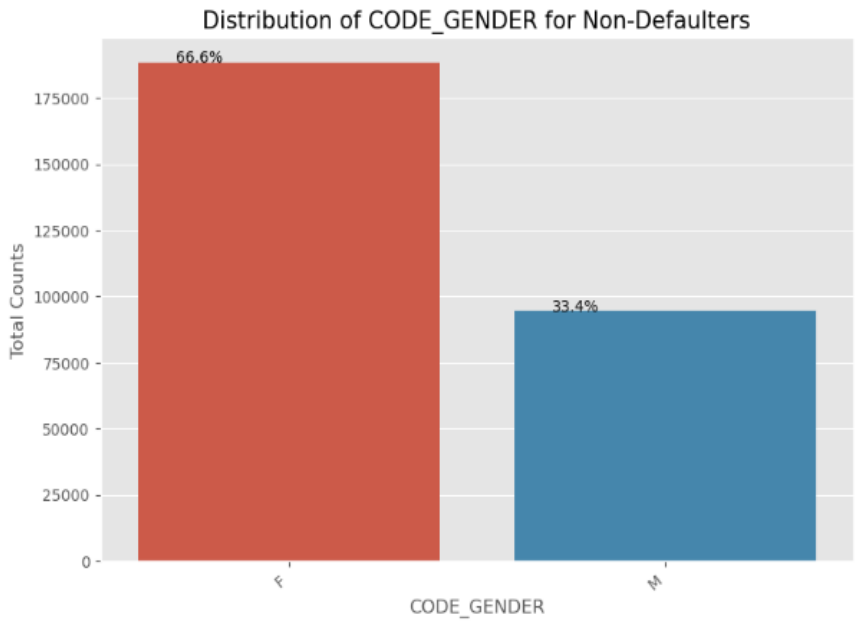
- Understanding the variables and load the application data
- Cleaning the data: Identifying missing values and treatment, Identify outliers, Imputation of missing values
- Find out the data types
- Data Imbalance
- Binning
- Univariate Analysis: Analyze one variable using histplot (for numerical variable),count plot(for categorical variables)
- Bivariant Analysis: Analyze two variable using bar chart, scatter plot
- Multivariant Analysis: Analyze more than two variables using heat map
- Finding out top 10 correlation
- Analysis of previous application data using above all steps
- Merge the both data files
- Plot the combined plot

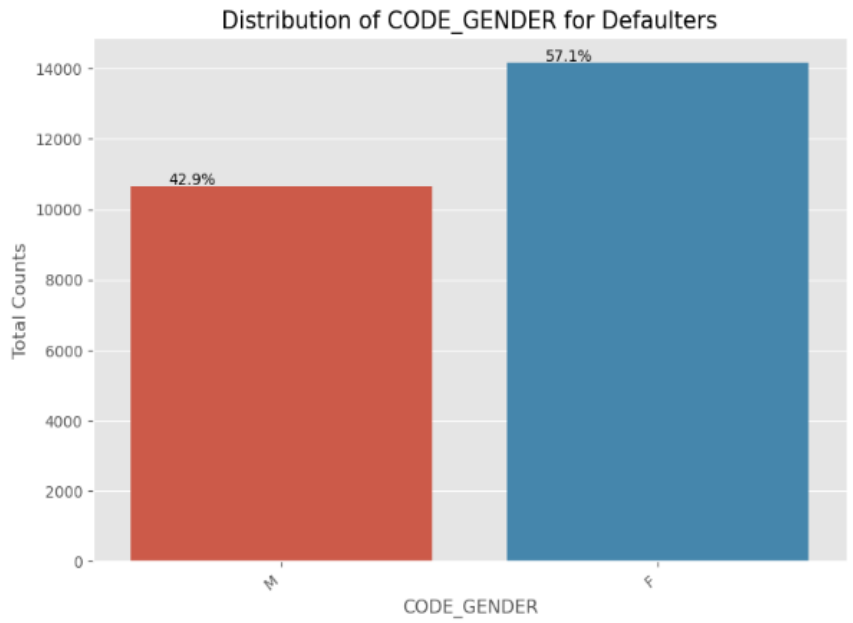
Imbalance in Target



- As we can see from pie chart In Target variable the no of non-defaulters are more than no of defaulters.

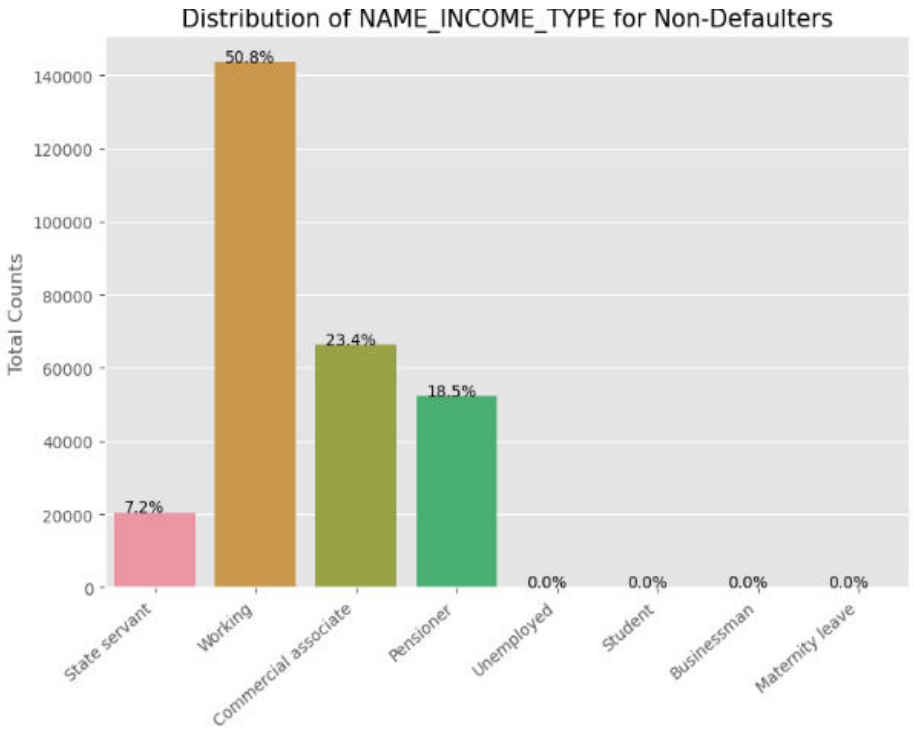
Univariant Analysis

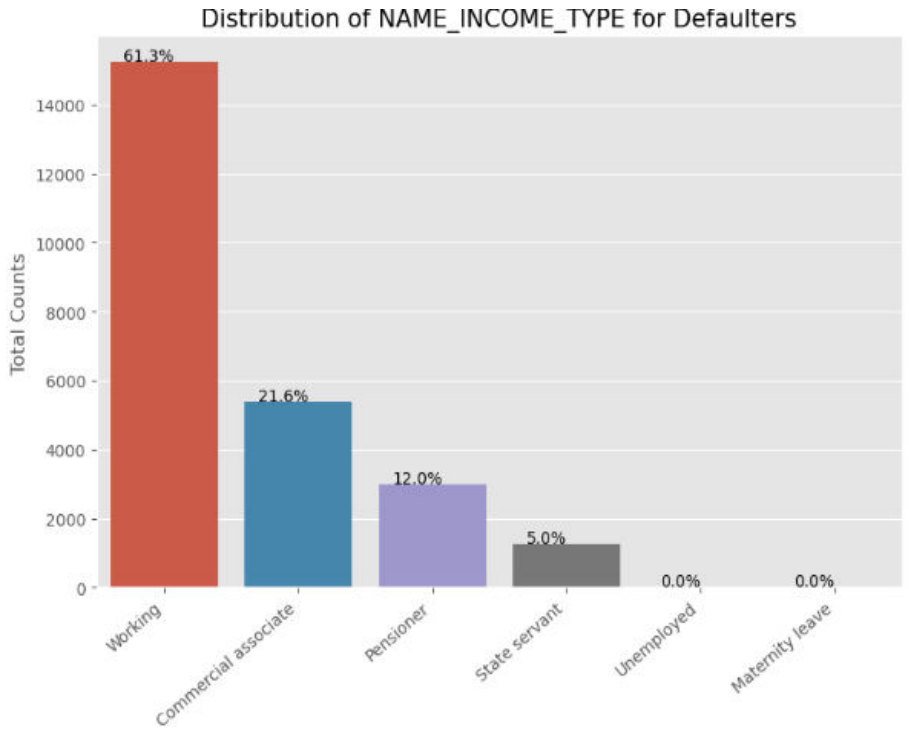




- As it is shown in above plot that 66.6% females are non-defaulter and 57% females are defaulter, so more no of females applied for loan than man , so females are more defaulter than man. but if we see the rate of being defaulter is less for female compare to man .

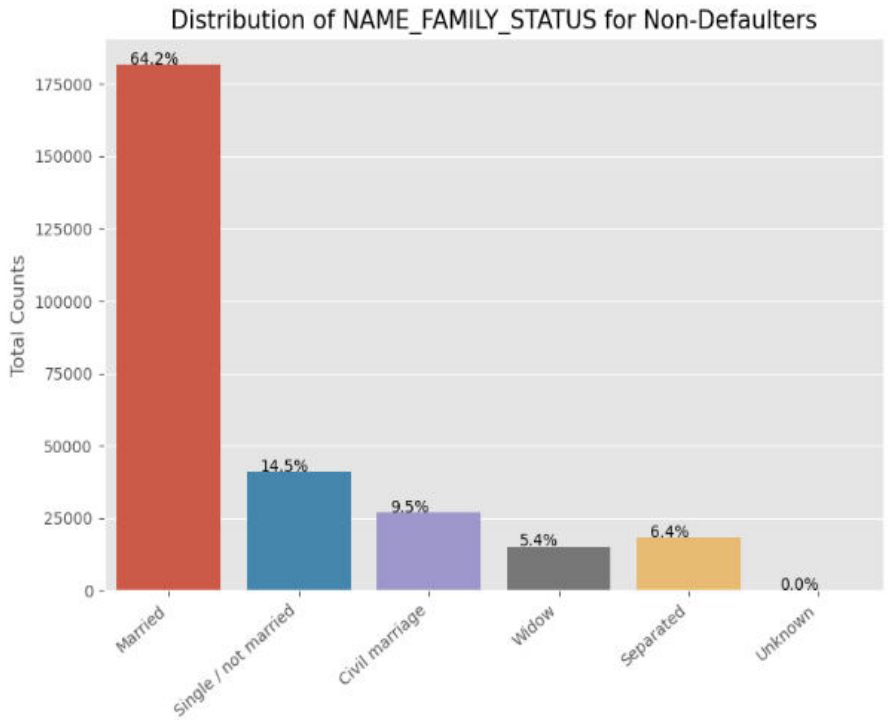
NAME_INCOME_TYPE Variable

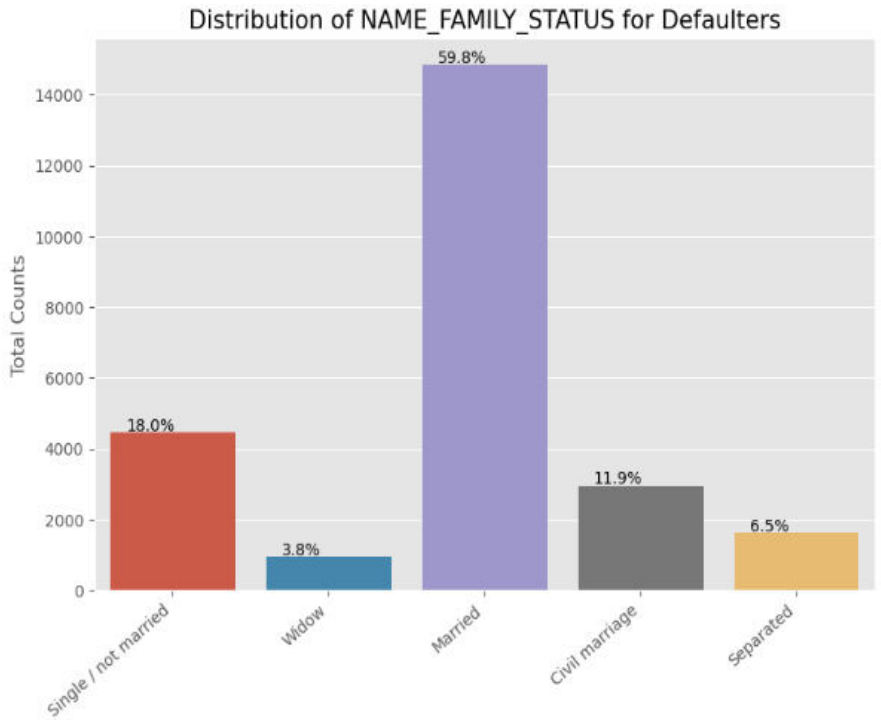




- As it is shown above working persons take more loans so chance of becoming defaulter is high in such category.
- Students are 0% defaulter because mostly they do not take loans, also businessman also 0% defaulters,
- so it is safe to give loans to businessman.

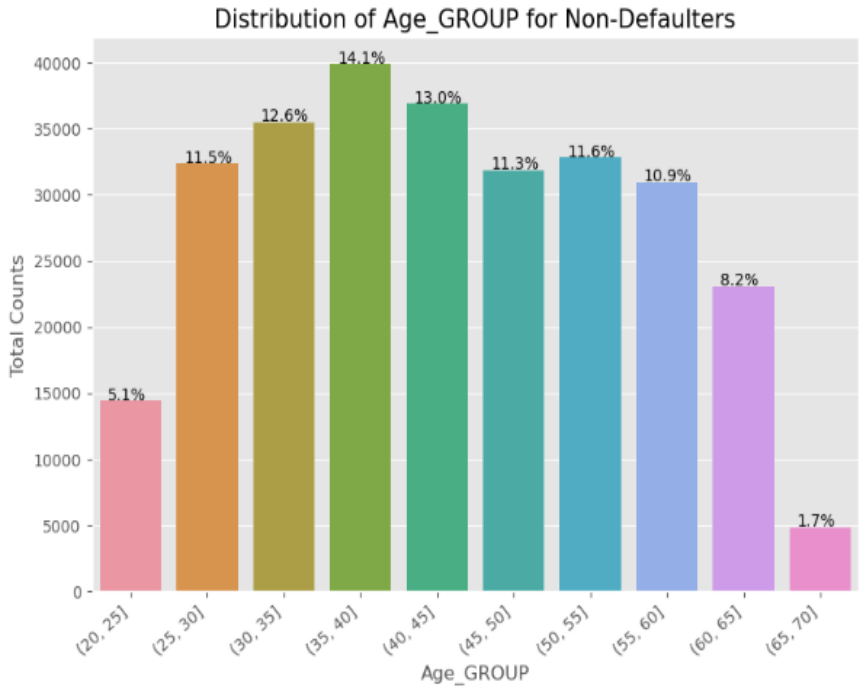
NAME_FAMILY_STATUS Variable

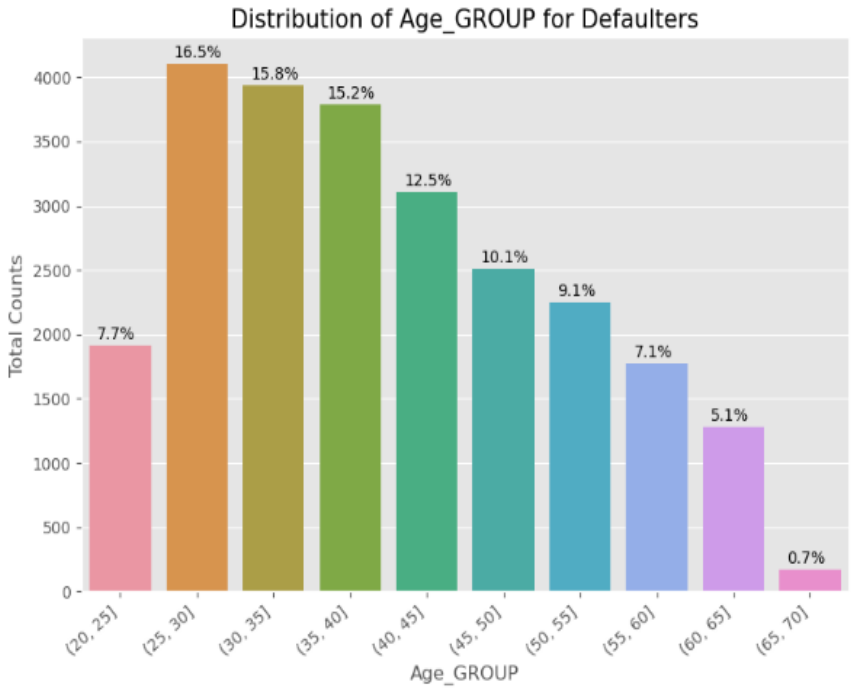




- Married people tend to apply for more loans comparatively.
- But from the graph we see that Single/non Married people contribute 14.5% to Non Defaulters and 18% to the defaulters.
- So there is more risk associated with them.

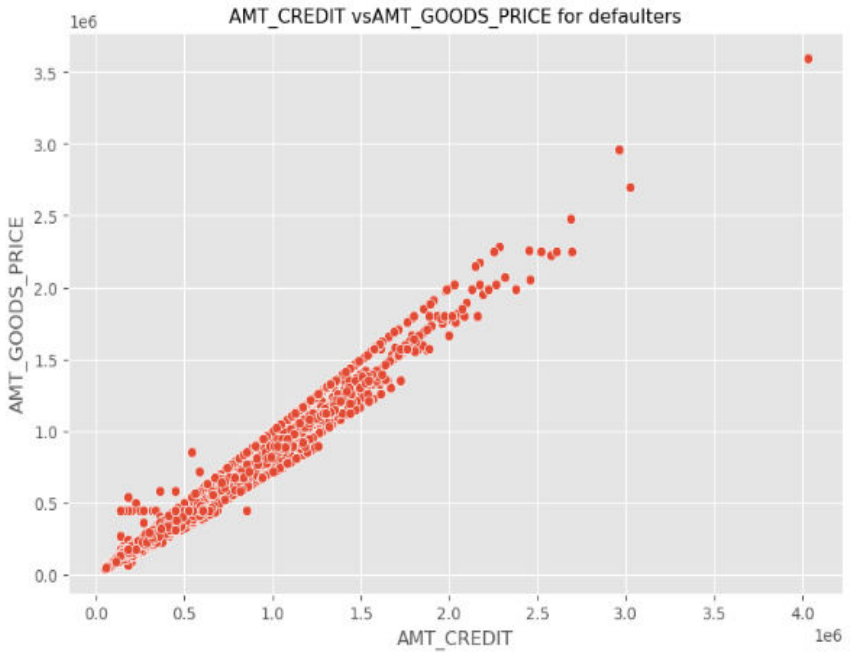
Age_GROUP Variable





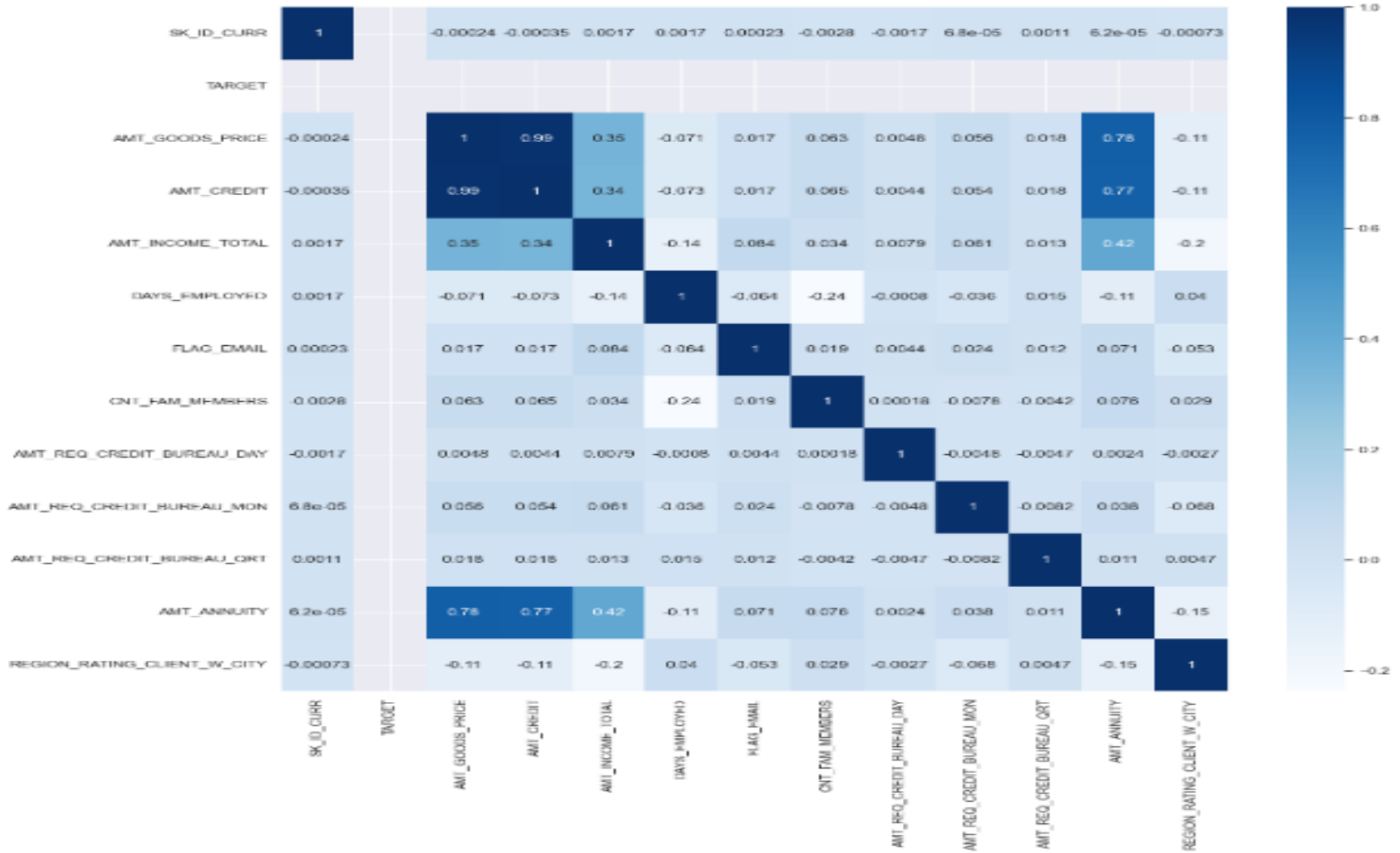
- From the above plot we can say that are group of 25-30 years are more defaulter than other categories, so we have high risk to give loan to such person.
- As the age increases the defaulter ration is less , means higher age person is safe to give loans.

Bivariant Analysis



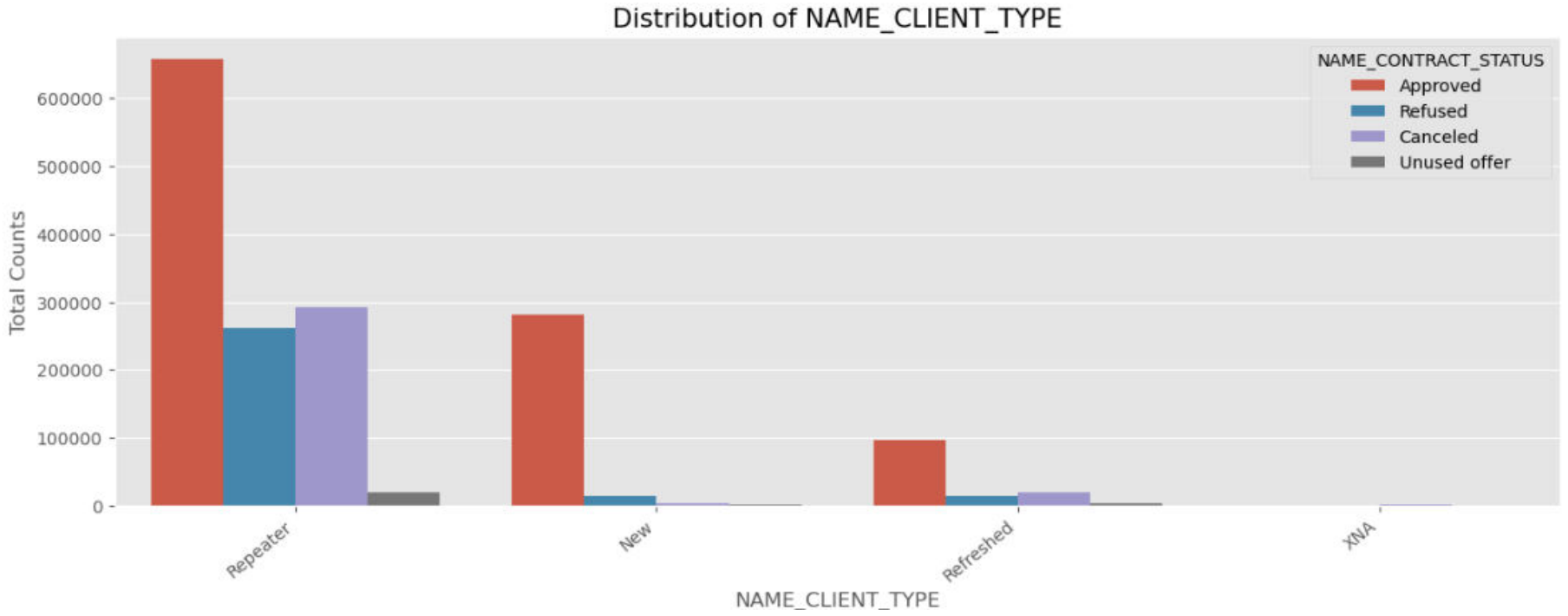
- As we can see that in lower left corner density is higher means for low loan amount no of non-defaulter are higher than defaulter.
- By increasing the goods price credit amount is also increases.

Multivariant Analysis



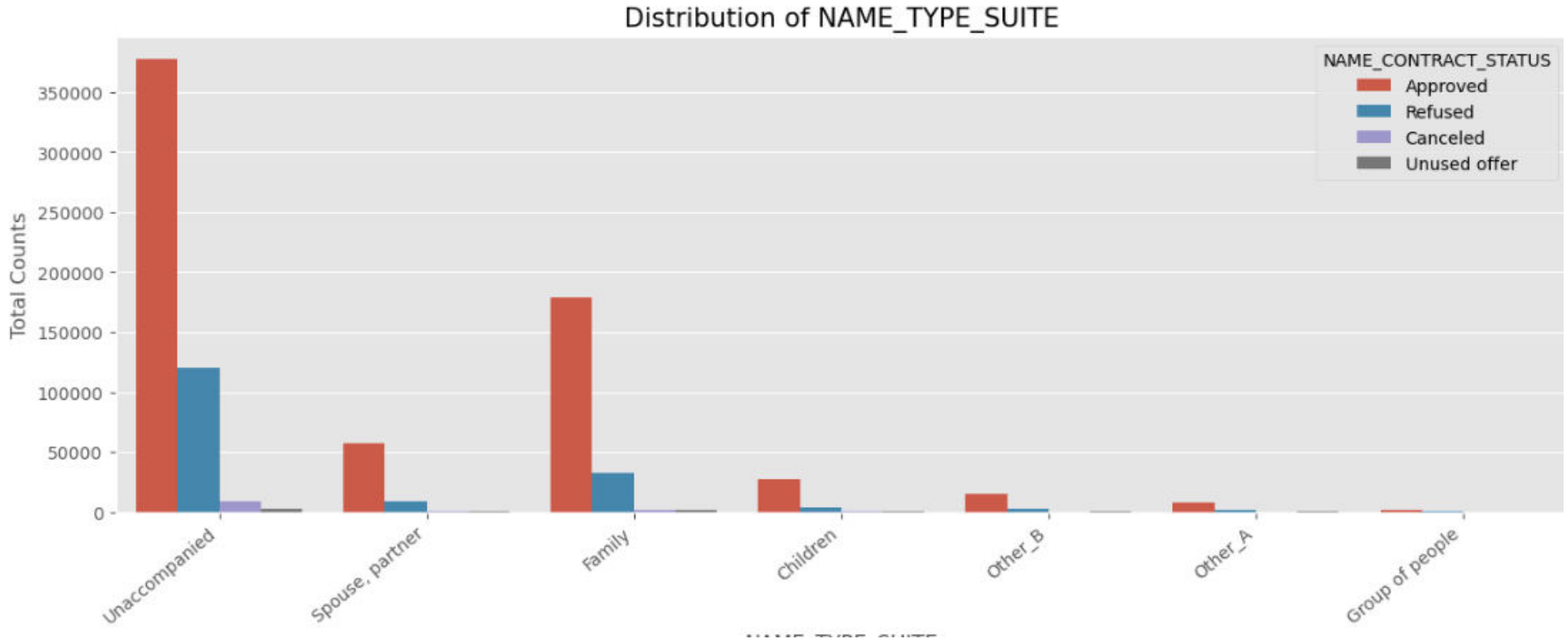
- This graph shows the top 10 correlation-9 of target 0.
- Here AMT_CREDIT and AMT_GOODS_PRICE are highly correlated.

Univariant Analysis for previous application data



- Repeaters are getting loans majorly. It is simple that bank is also interested such persons have previous loan history.

NAME_TYPE_SUITE Variable

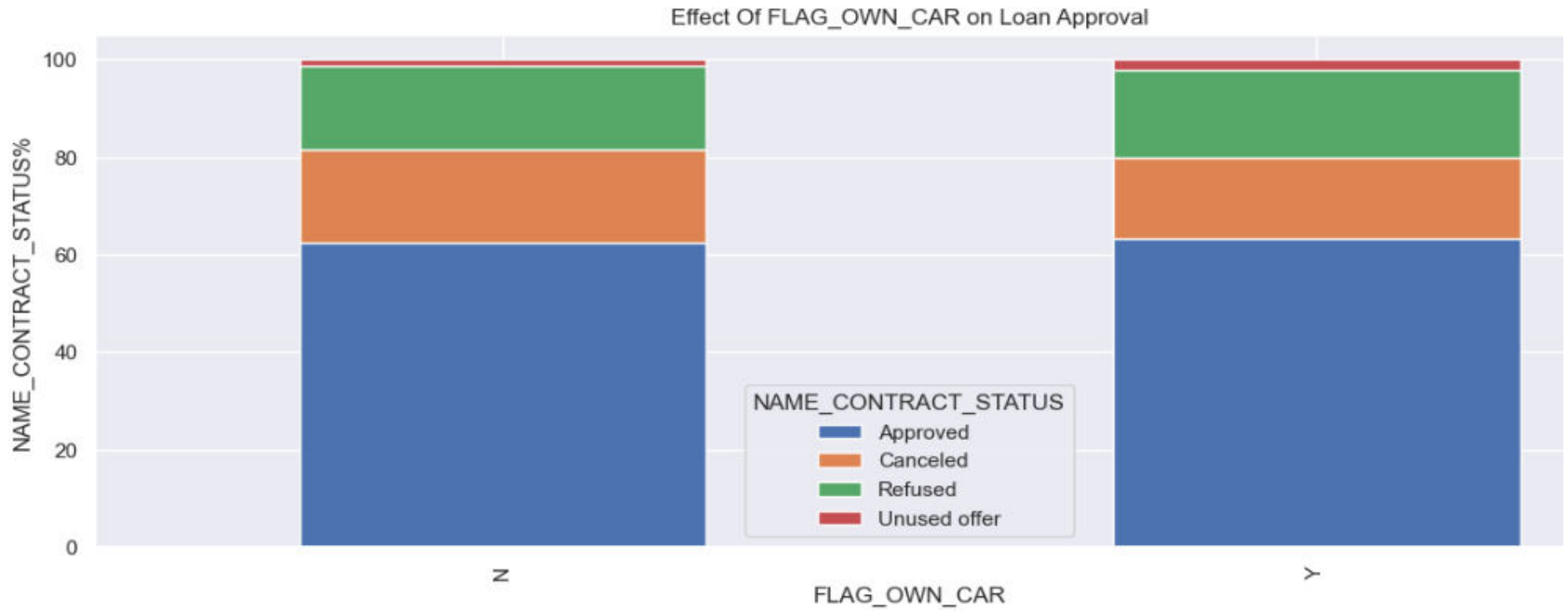


- Single person getting more loans.

Merged Data Analysis

- 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.

FLAG_OWN_CAR Variable



- As it is show from fig that response is not much effected by car ownership, but person who have cars is not as much defaulter.
- so bank prefer car own person more than not have.

Groups having less chance to be defaulter

- Old female clients.
- Person with high income category.
- Person with high education.
- old people of any income group.
- Person who are businessman.
- Any person who’s previous loan approved.
- widow

Risky Groups

- Male clients with marriage.
- Lower educated persons
- Previous refused loan status group