

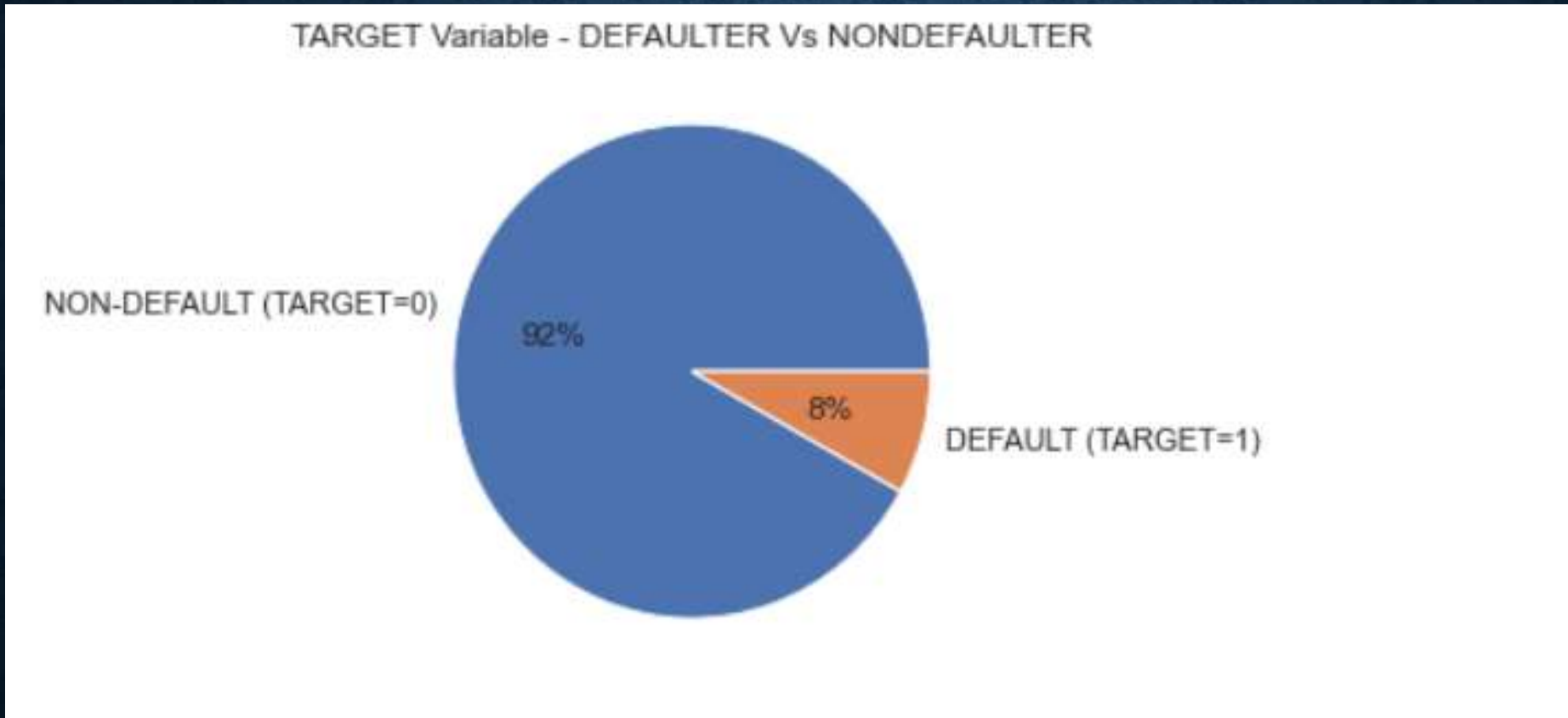
# **CREDIT EDA ASSIGNMENT**

Presentation by VIKRAM MOHANTY

# ANALYSIS ON NEW APPLICATION DATASET

- First we have to observe the dataset, search for the target variable.
- Now we have to clean the data and look for uneven rows and columns.
- Then we have to impute null values.
- After imputation and regular cleaning we can go further with the analysis.

## Checking for imbalance in the new application data

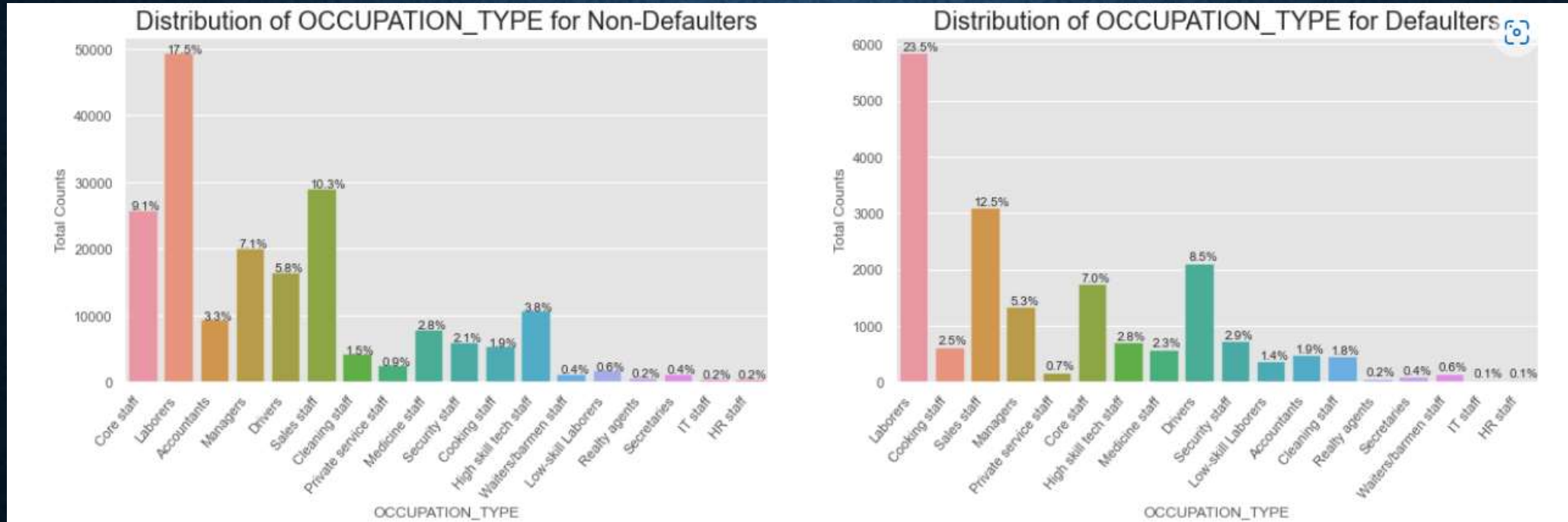


- By the graph we are seeing the imbalance in the dataset.
- By the graph we can see that there are 92%(Target=0) of non-defaulters and 8% of defaulters(Target=1).



# **UNIVARIATE ANALYSIS FOR NEW APPLICATION DATASET**

# PLOTTING GRAPH FOR DEFAULTERS AND NON-DEFAULTERS FOR OCCUPATION\_TYPE

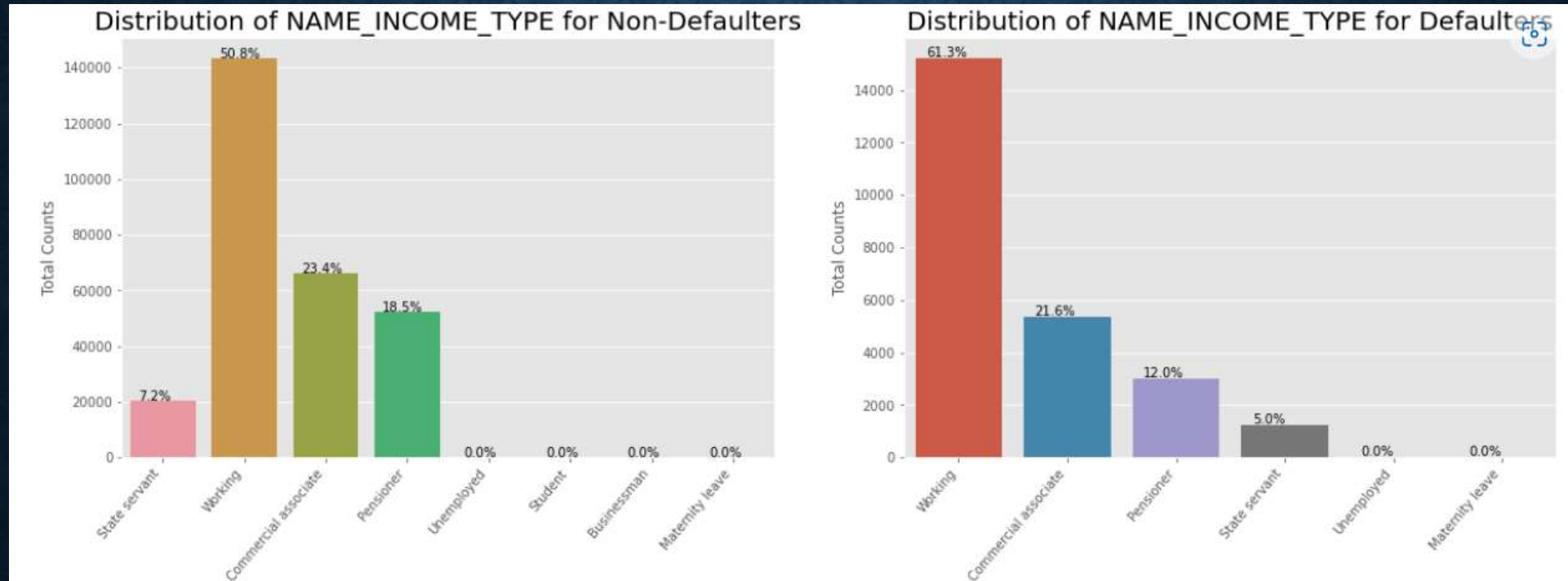


-Hence we can conclude that there are more number of laborers applying for loans and that's why we can see that there are more numbers of defaulters in laborers.

-The least number of defaulters and non-defaulters are from IT staff, HR staff, Realty agents.

-There is a risk of giving loans to drivers as their defaulters are more than non-defaulters.

# PLOTTING GRAPH FOR DEFAULTERS AND NON-DEFAULTERS FOR NAME\_INCOME\_TYPE

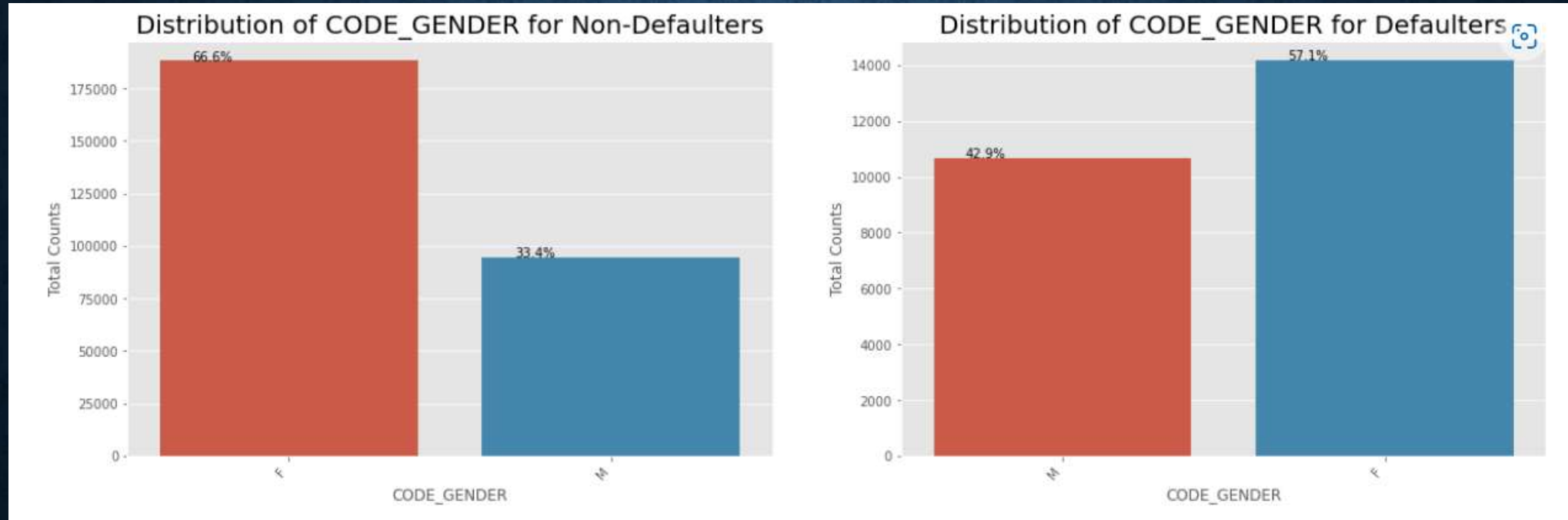


-we can see that student and businessman don't default as students are not allowed to pay while studying and businessman pay loans on time, while there are more number of defaulters than non defaulters in working sector.

-Businessman pay loans on time, while there are more number of defaulters than non defaulters in working sector.

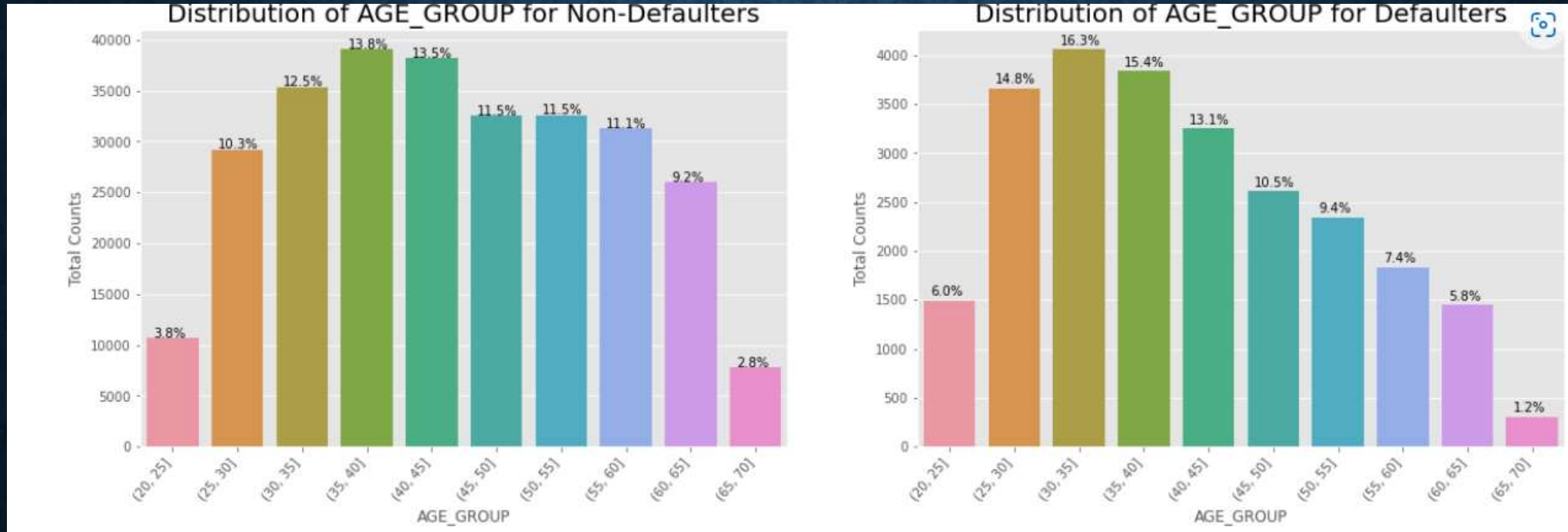


# PLOTTING GRAPH FOR DEFAULTERS AND NON-DEFAULTERS FOR CODE\_GENDER



-we can conclude by graphs that there are more number of non-defaulters in females than men. hence there are more number of defaulters in female as well.

# PLOTTING DEFAULTERS AND NON DEFAULTERS WITH AGE\_GROUP

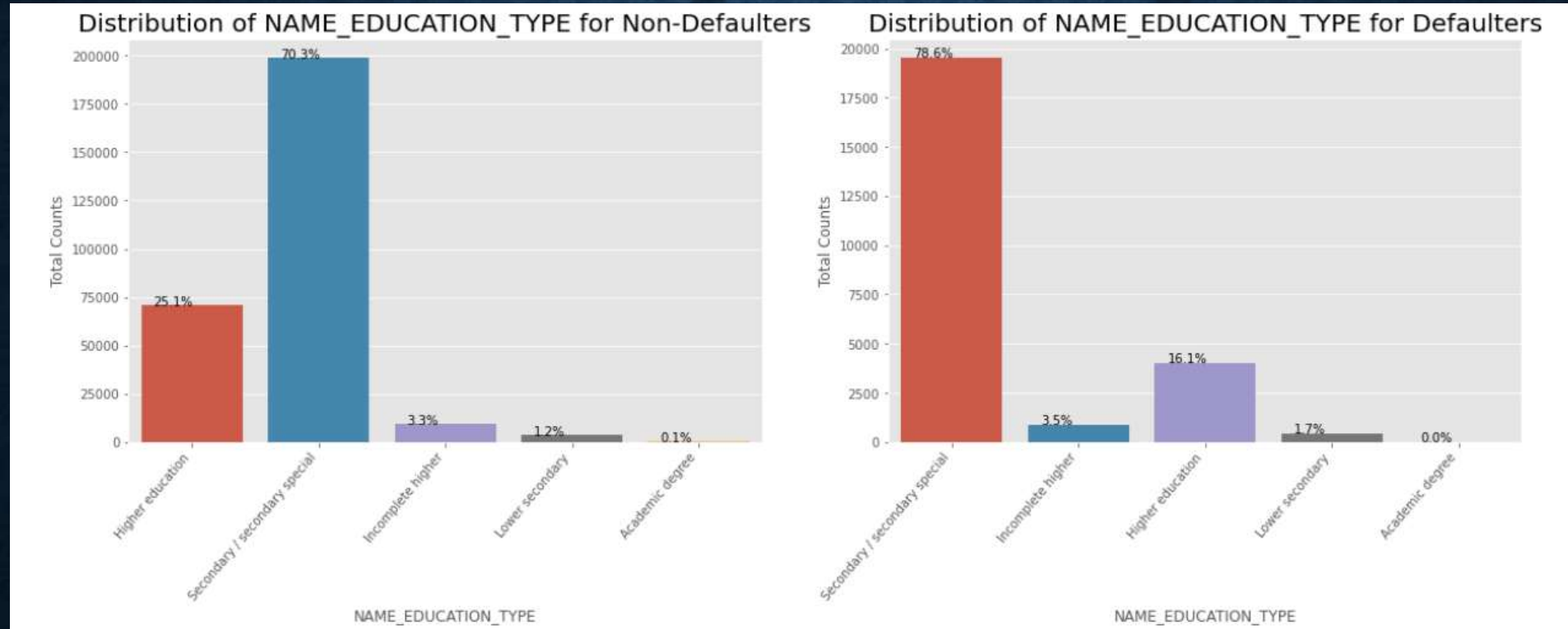


-As we can see from the graph as the age increases the default rate decreases we can say that after the age of(25-30) there is a increase in salary.

-Till the age of 30 the number of defaulters are more than non-defaulters we can say that till that time they are not employed or their salary is not high.



# PLOTTING DEFAULTERS AND NON DEFAULTERS WITH NAME\_EDUCATION\_TYPE

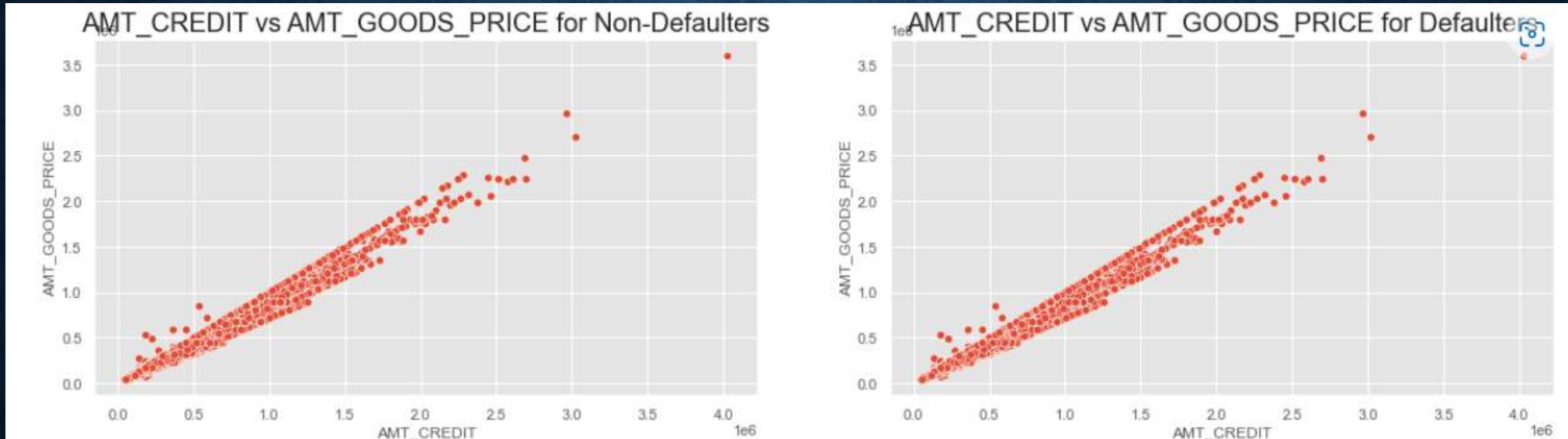


-As per graph higher education is 25.1% in non-defaulters and higher education has 16.1% in defaulters. which is not that big in difference in their percentage.

-secondary special also does have a percentage difference in defaulters and non-defaulters.

# **BIVARIATE ANALYSIS FOR NEW APPLICATION DATASET**

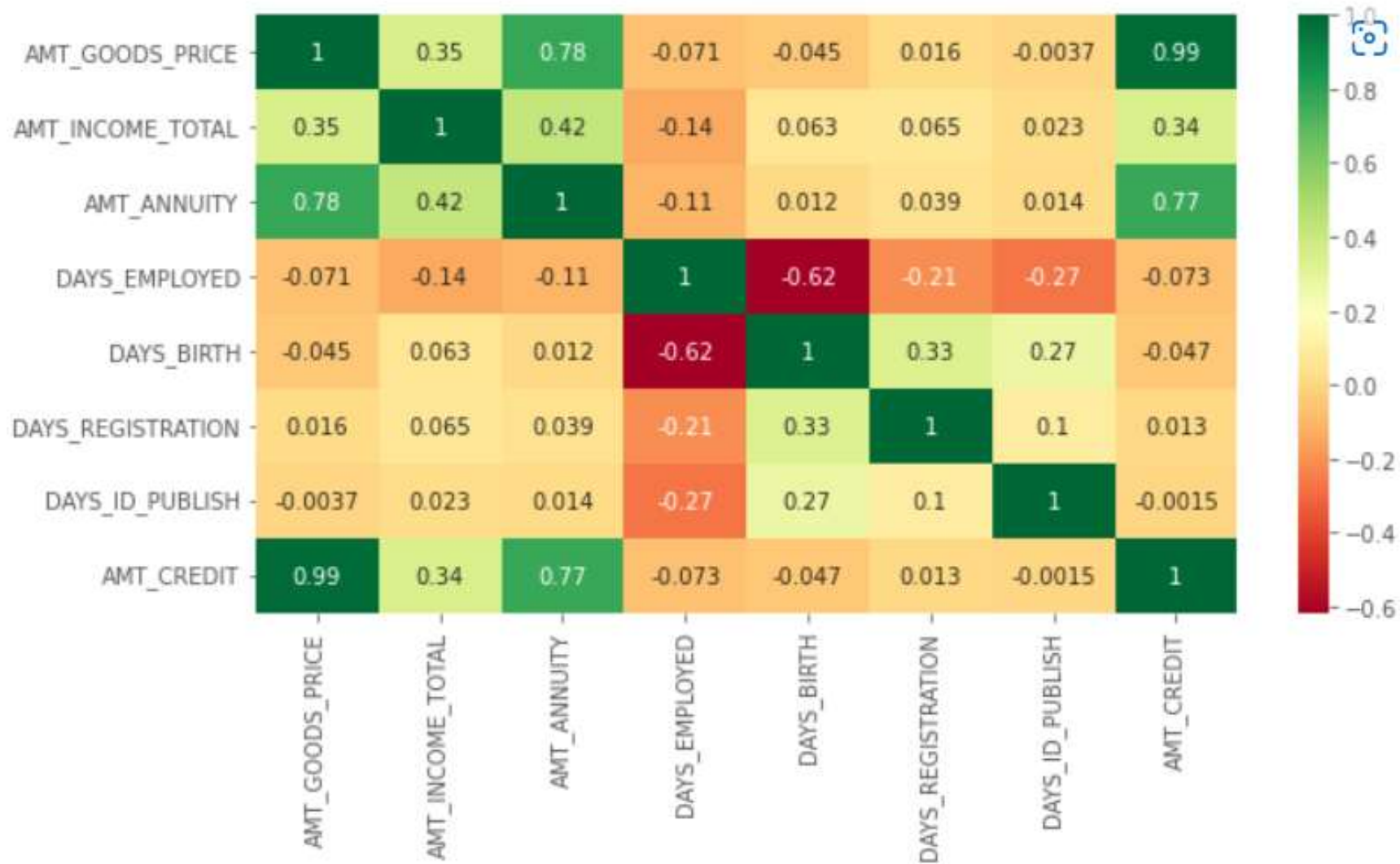
# PLOTTING SCATTER PLOT FOR 'AMT\_CREDIT','AMT\_GOODS\_PRICE'.



-As there is a linear relation between AMT\_GOODS\_PRICE vs AMT\_CREDIT as they are increasing linearly.



# HEATMAPS FOR NON-DEFAULTERS



# HEATMAP FOR DEFAULTERS



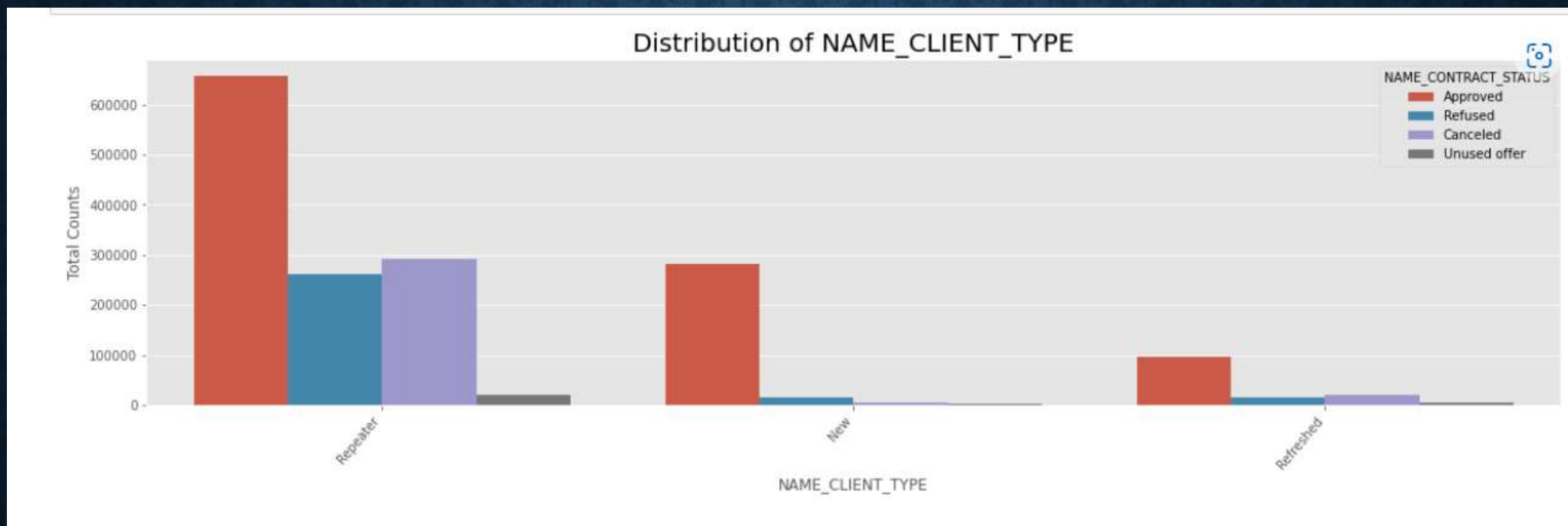
-By the heatmap we can see that there is a high correlation between good price and credit.

# ANALYSIS ON PREVIOUS APPLICATION DATASET

- First we have to observe the dataset, search for the target variable.
- Now we have to clean the data and remove the null values more than 50%.
- Then we have to replace XNA values with NaN.
- After imputation and regular cleaning we can go further with the analysis.

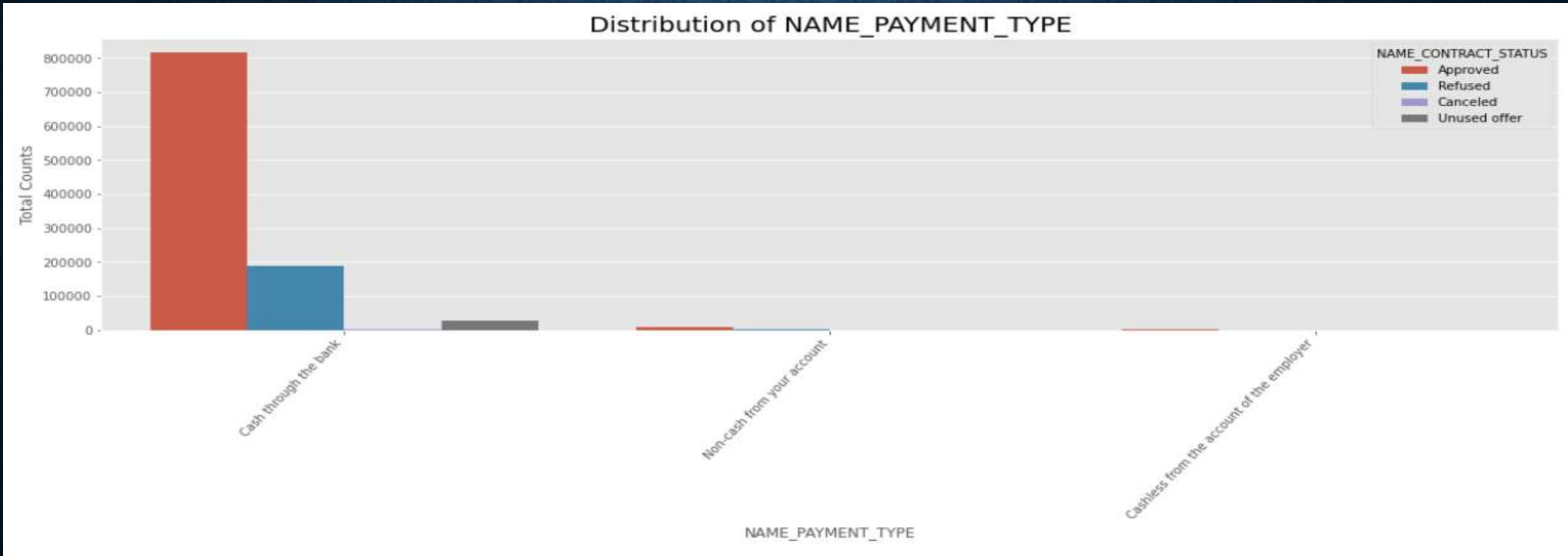


# UNIVARIATE ANALYSIS FOR PREVIOUS DATA



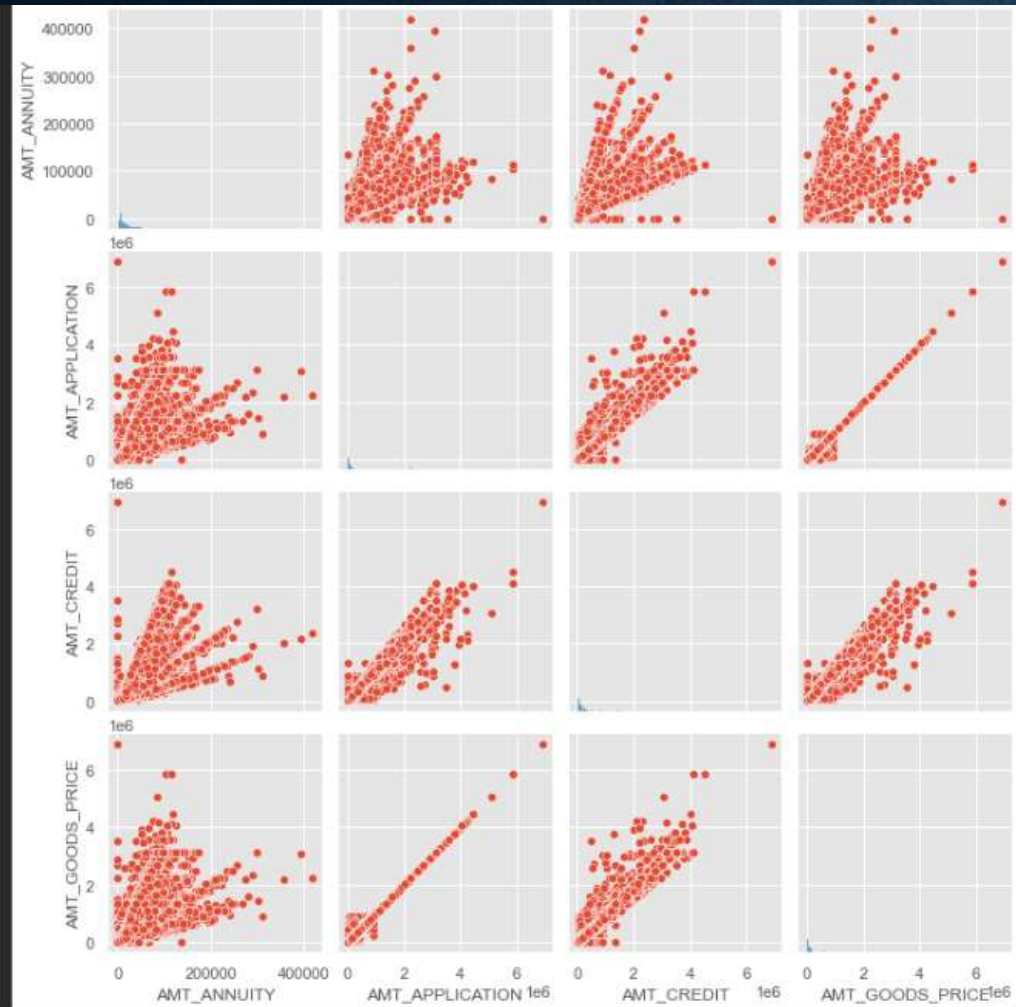
-Most of the loan applicants are from repeaters and hence the most of the cancelled and refused are by repeaters only.

# UNIVARIATE ANALYSIS FOR "NAME\_PAYMENT\_TYPE"



-Most of the loan are taken by 'cash through the bank' and the least was cashless from the account of the employer.

# BIVARIATE ANALYSIS FOR PREVIOUS APPLICATION DATA

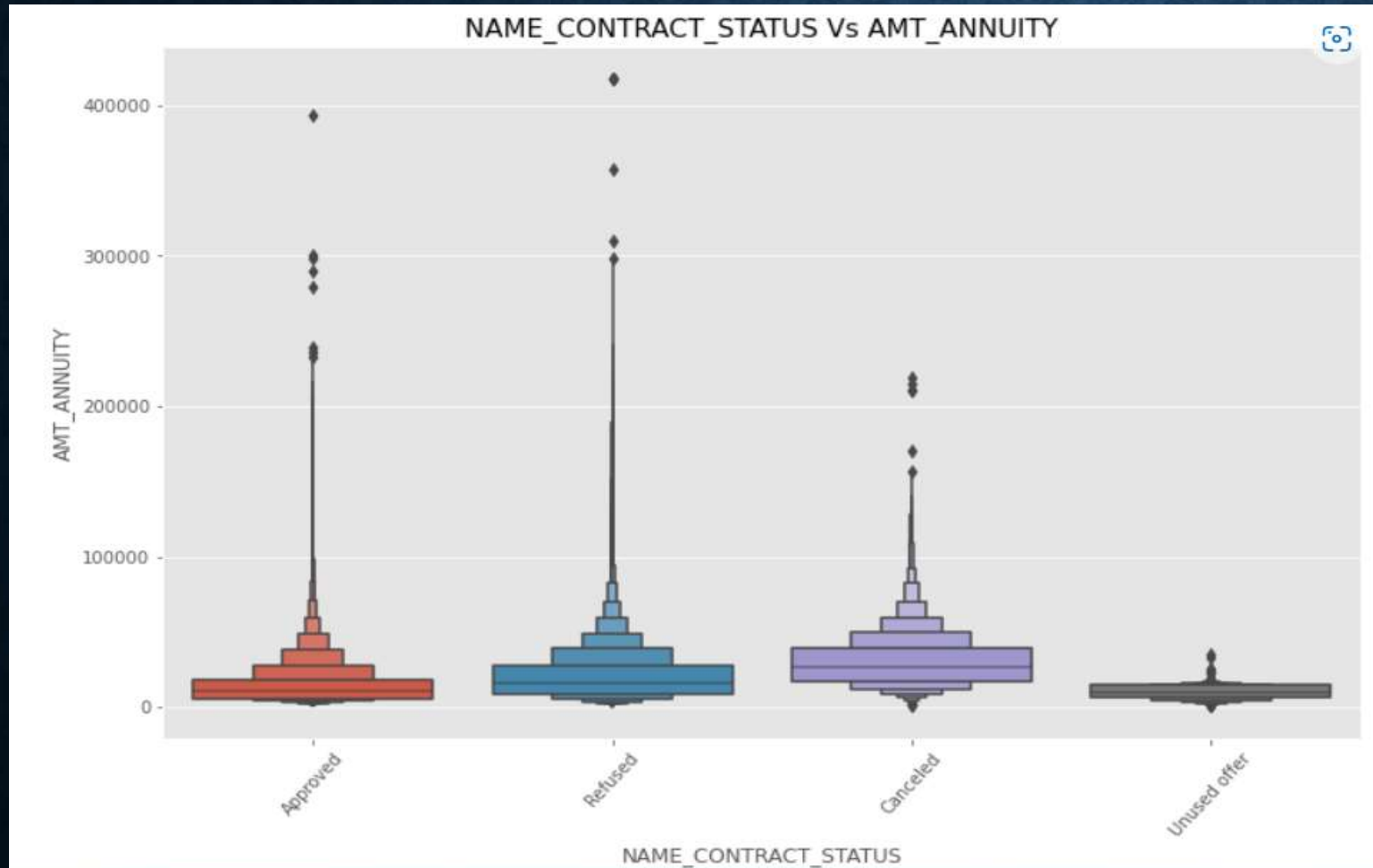


-In AMT\_ANNUIITY vs AMT\_APPLICATION, AMT\_CREDIT, AMT\_GOODS\_PRICE as the x-axis increases the y-axis increases too but not always.

- In AMT\_APPLICATION vs AMT\_GOODS\_PRICE, AMT\_GOODS\_PRICE vs AMT\_APPLICATIONS there is a linear increase.



# BIVARIATE ANALYSIS OF CONTRACT STATUS AND ANNUITY OF PREVIOUS APPLICATION



-We can see that loan application for people with lower AMT\_ANNUIITY gets canceled and people with high ANNUITY also get cancelled

# **MERGING BOTH THE DATASET (NEW APPLICATION DATA & PREVIOUS APPLICATION DATA)**

- Merging the data with inner join and on a similar column of SK\_ID\_CURR.
- Having a regular check on the combined data dataset.
- Finding the correlation between the merged data.
- Plotting graphs for correlation.

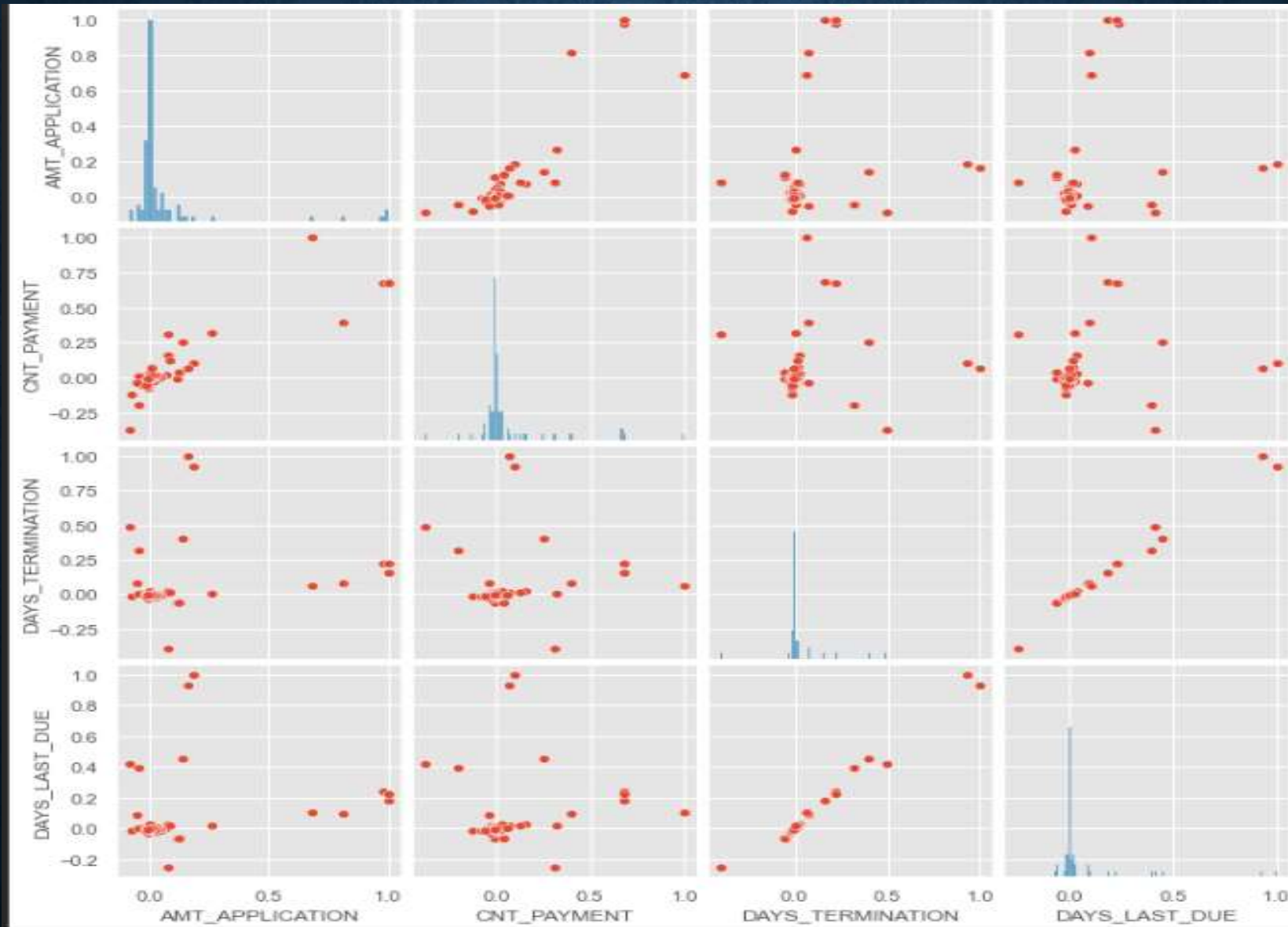
# UNIVARIATE ANALYSIS FOR MERGE DATA



-People who were approved for a loan earlier defaulted less where people who were refused a loan earlier have higher chances of defaulting.



# BIVARIATE ANALYSIS FOR CORRELATION OF MERGED DATA



**THE END**  
**THANK YOU**