



# CREDIT EDA CASE STUDY


By:

**Dinakar Nandanavanam**

- **Problem Statement:**

Identifying the process and pattern based on the client difficulties on payment of their loans and the action to be taken such as Rejecting the Loan, Reducing the Loan Amount or Lending the Loan at Higher Interest.

Based on Bank's Decision the following risks may arise:




If the applicant's history is good or he is likely to repay the loan, then not approving the loan leads to a Business loss to the company.

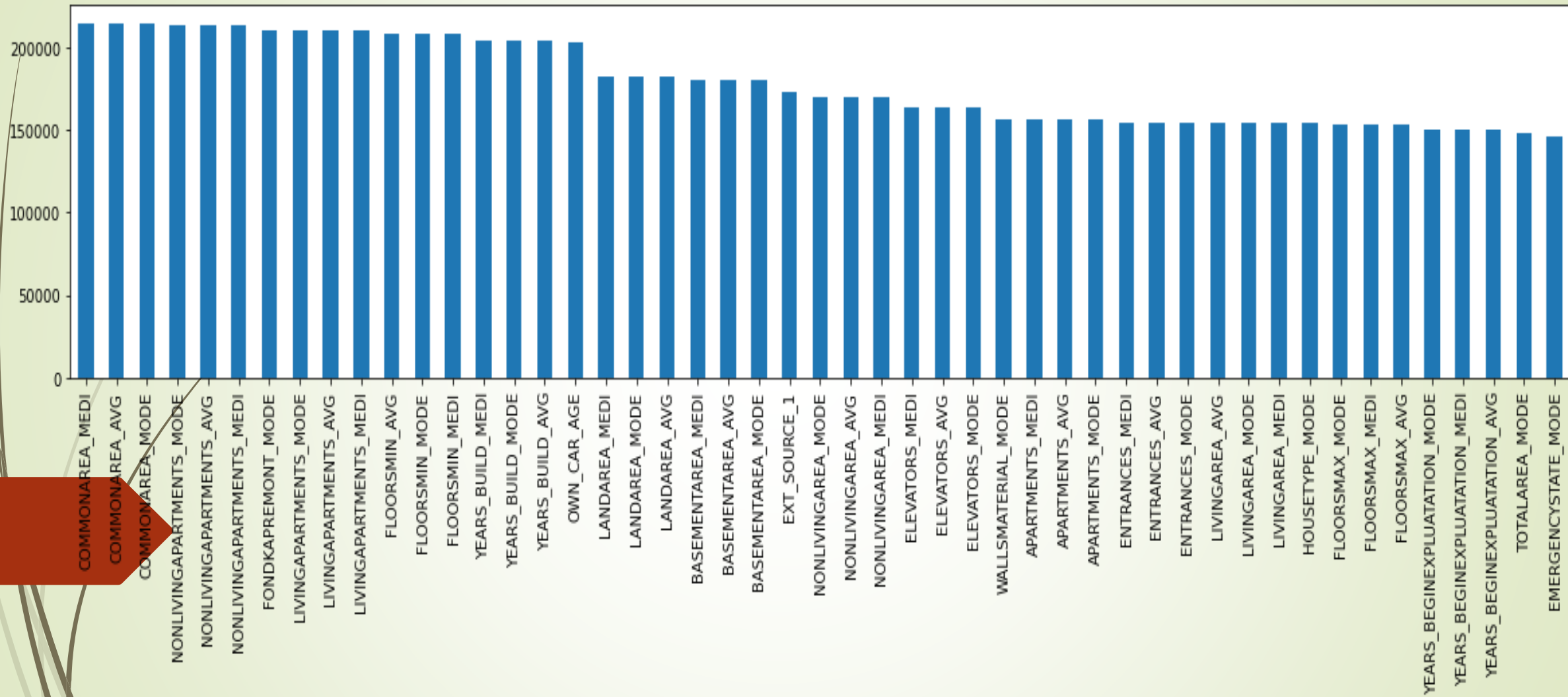
If the applicant is unable to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.



# Steps followed:

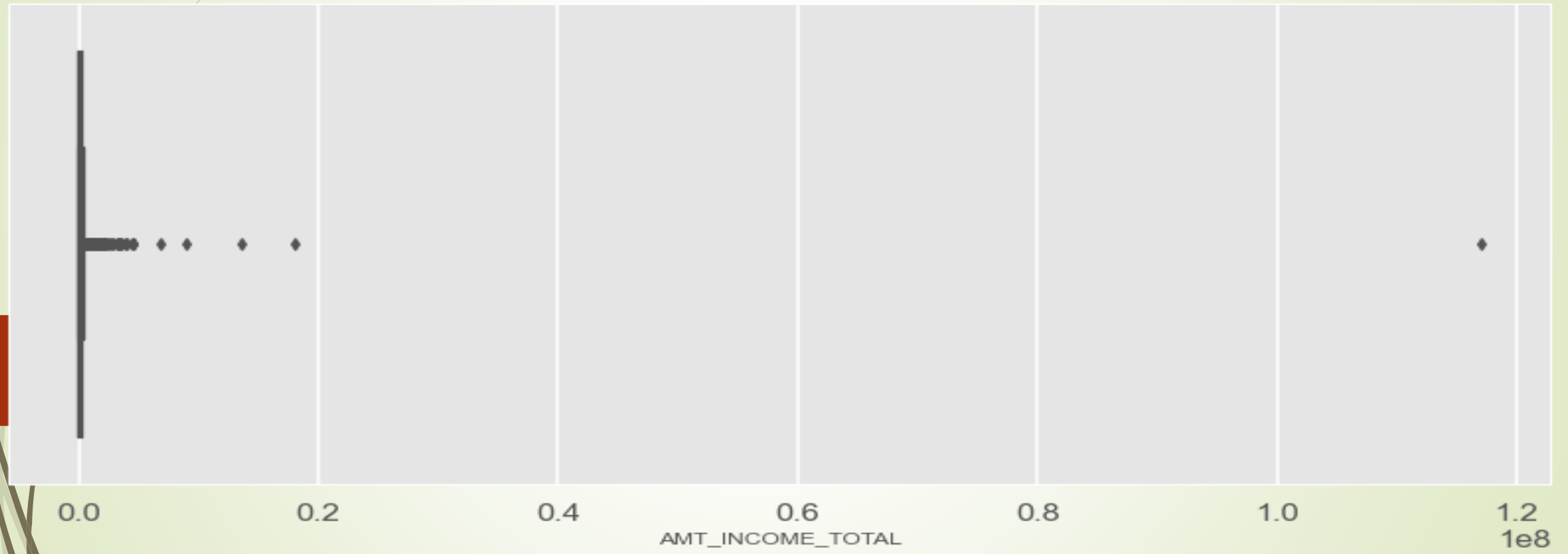
- Imputing Missing Values.
  - Checking Outliers.
  - Checking Data Imbalance ratio
  - Univariate and Bivariate Analysis on Data
  - Finding top correlation of Data
- 

List of Columns counts where NA values are more than 45%



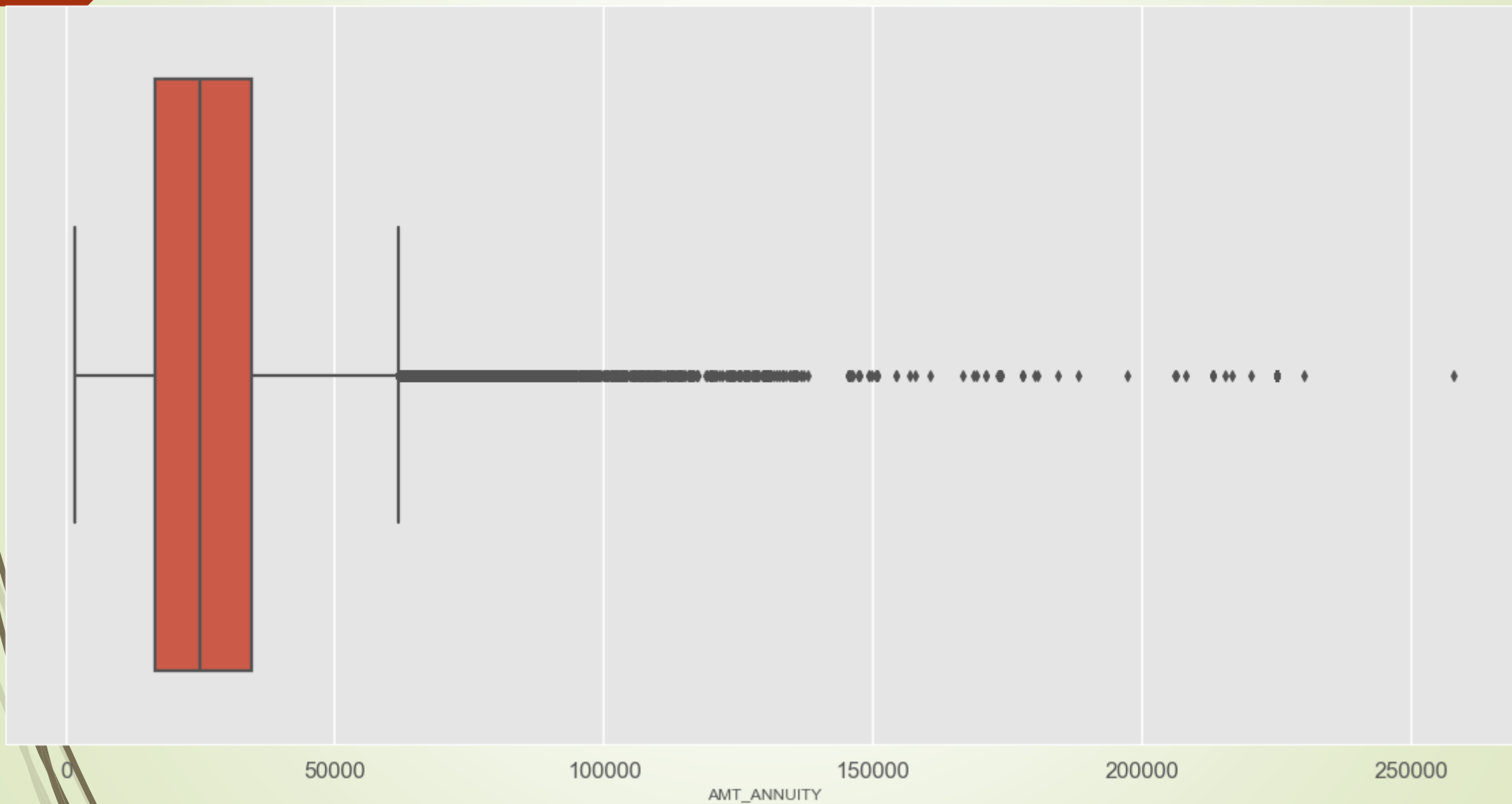
Outliers on  
AMT\_INCOME

AMT\_INCOME\_TOTAL Distribution



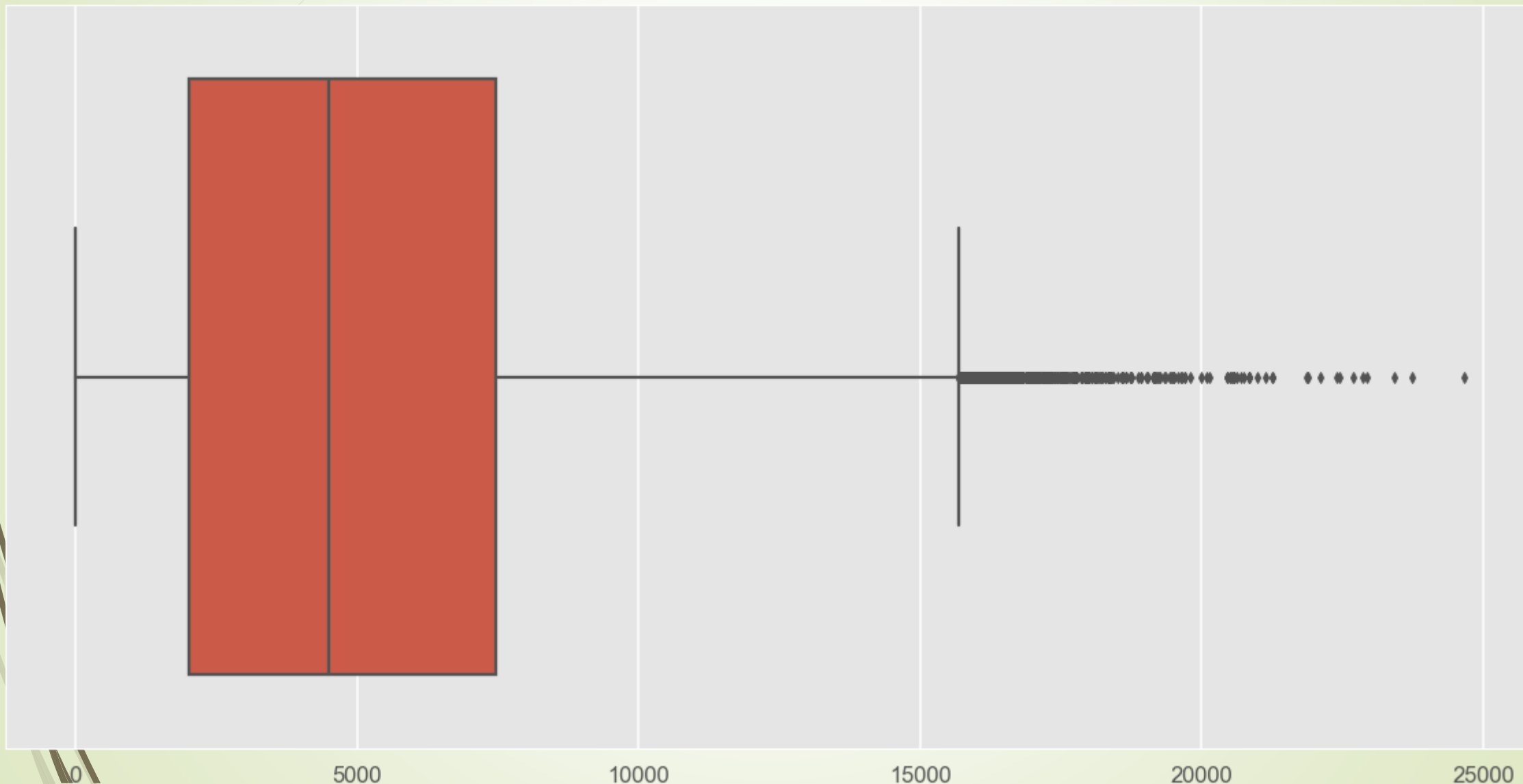
## Outliers on AMT\_ANNUIITY

Distribution of AMT\_ANNUIITY



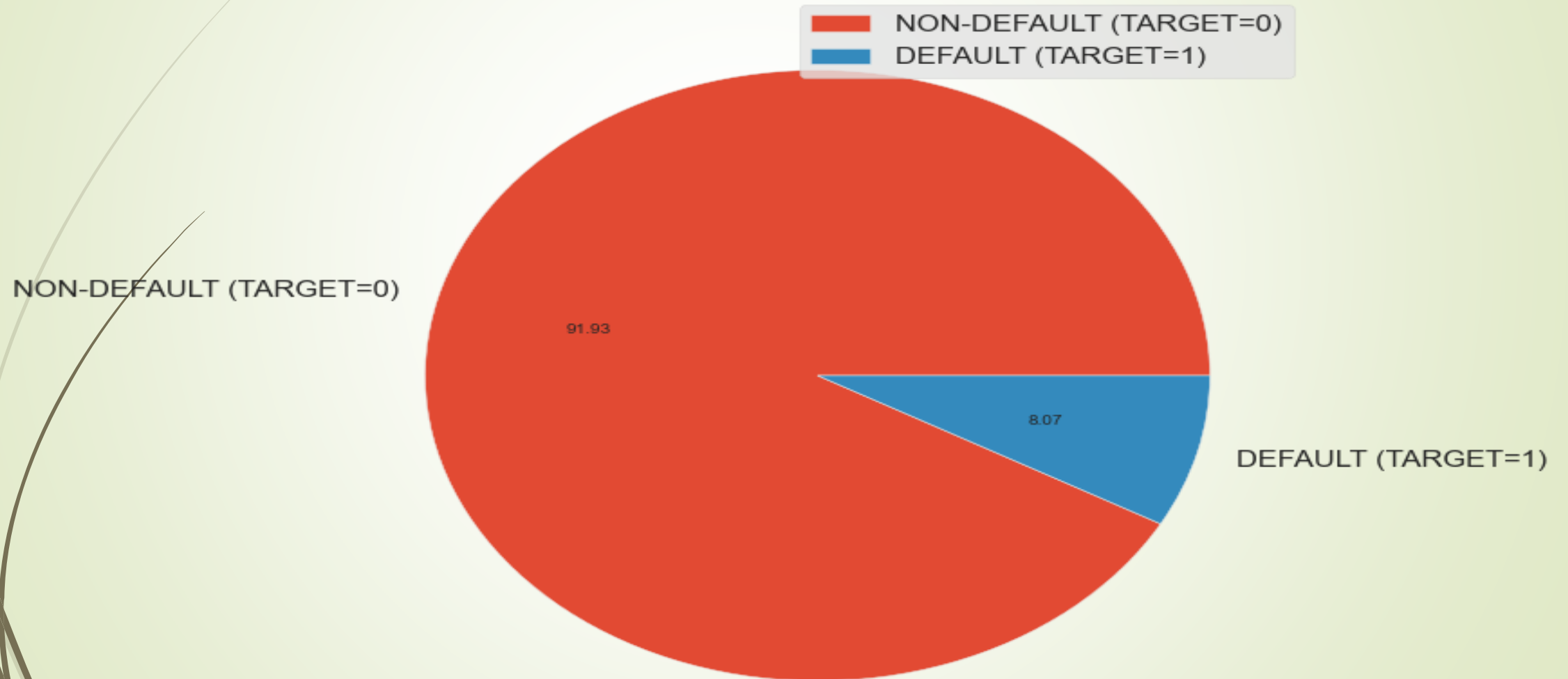
## Outliers on DAYS\_REGISTRATION

Distribution of DAYS\_REGISTRATION



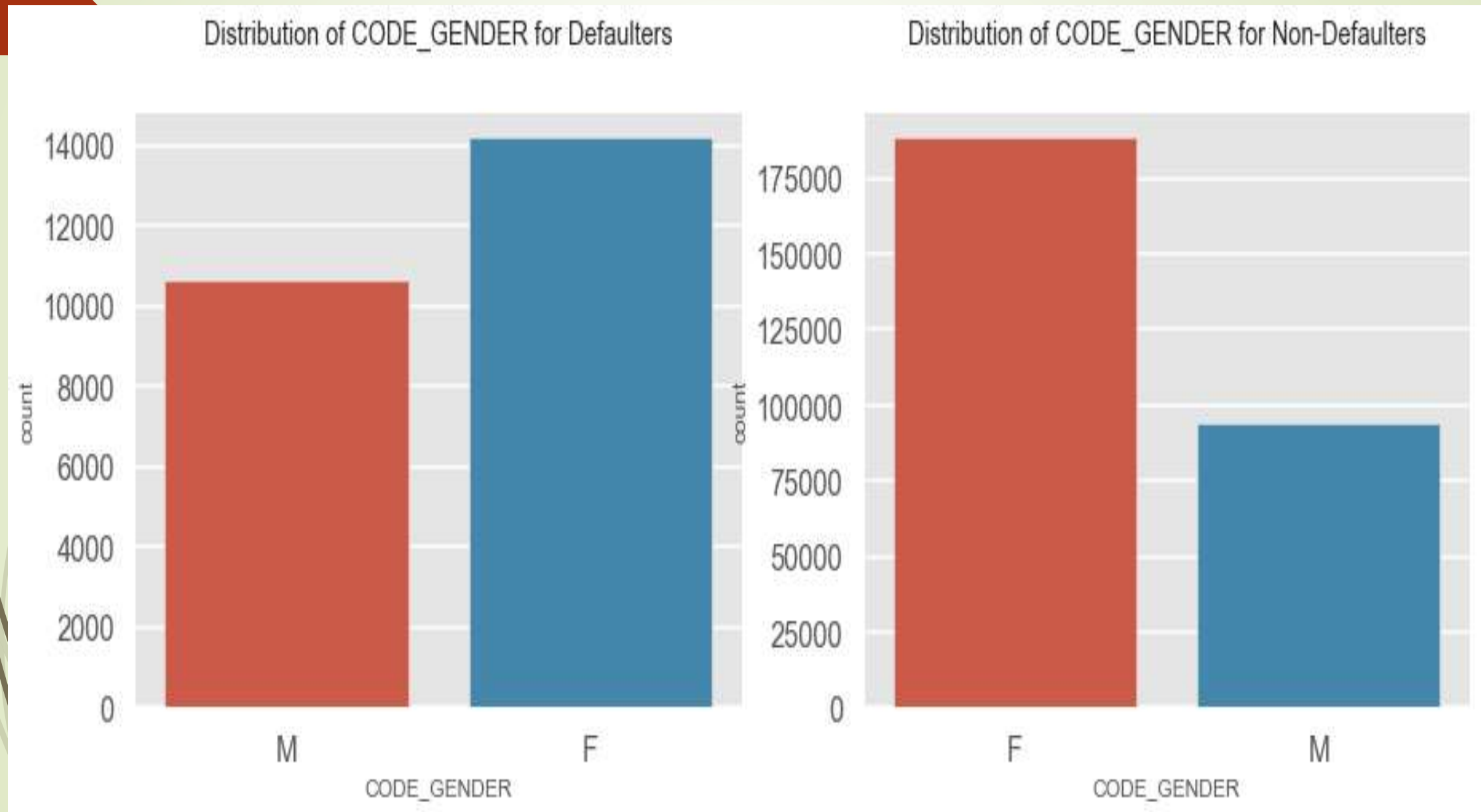
Proportion of Defaulter Imbalance  
is 8.07%

Target Imbalance



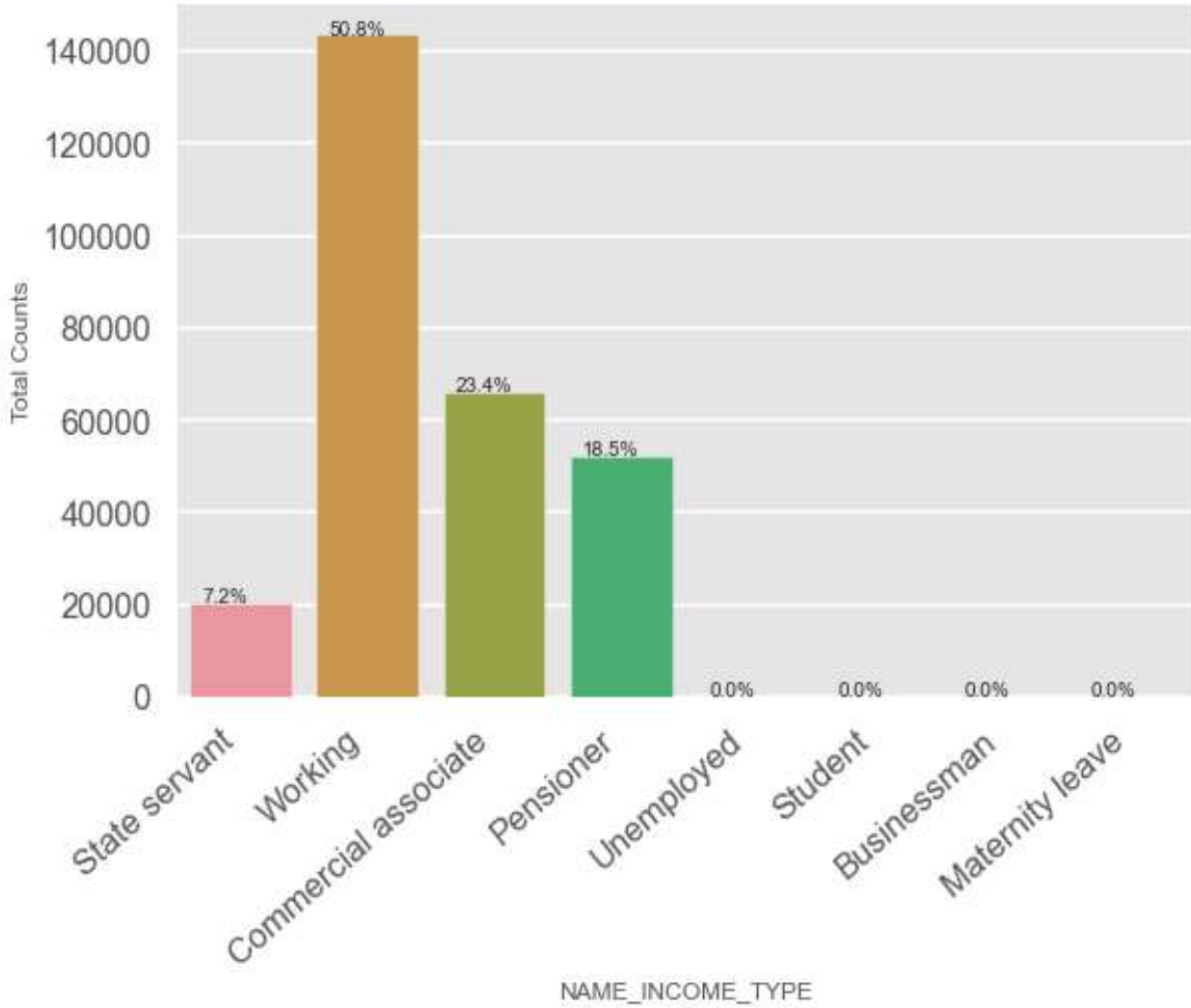


Analysis on CODE\_GENDER in both Defaulters and Non\_Defaulters

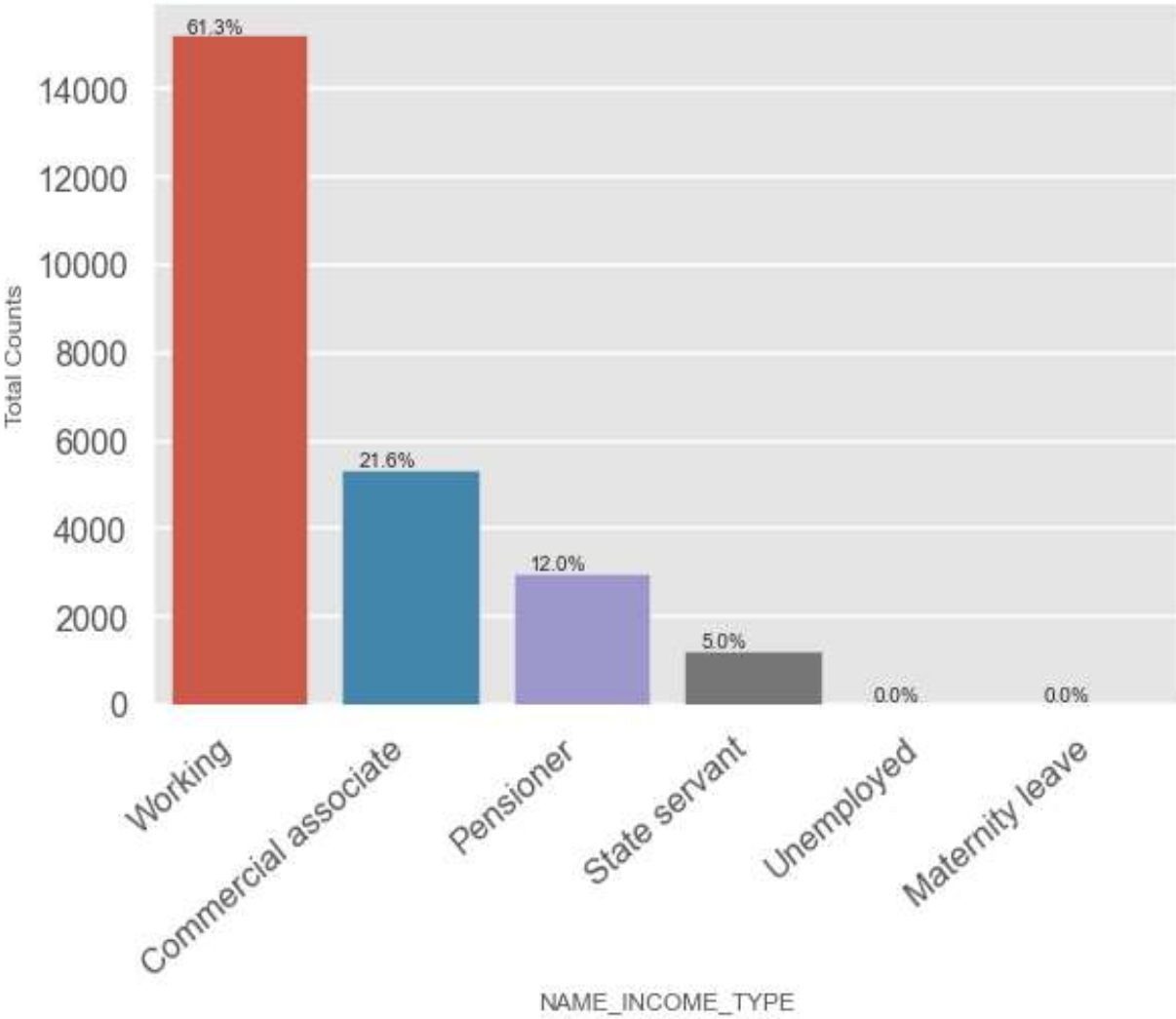


# Analysis on NAME\_INCOME\_TYPE in both Defaulters and Non\_Defaulters

Distribution of NAME\_INCOME\_TYPE for Non-Defaulters

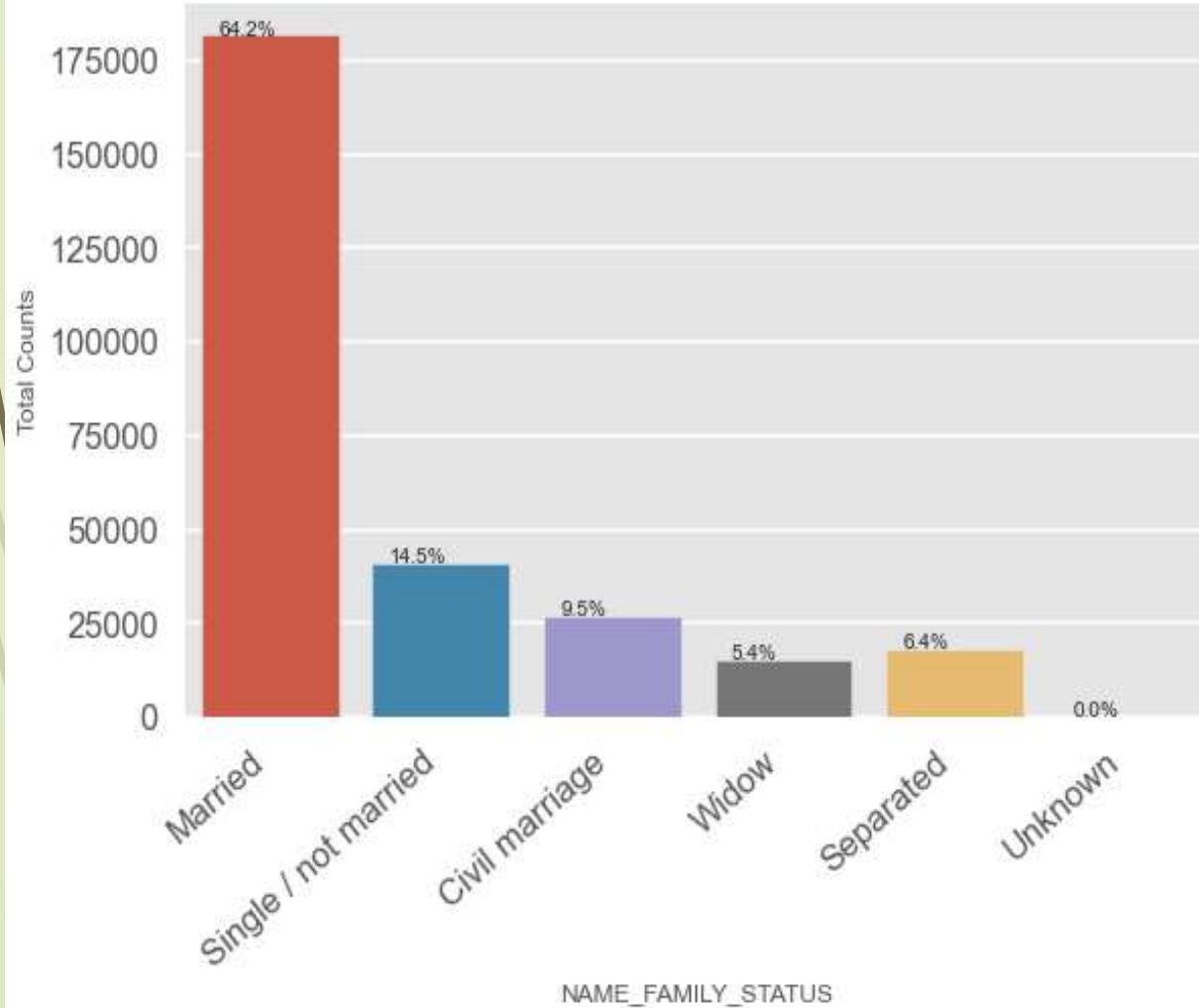


Distribution of NAME\_INCOME\_TYPE for Defaulters

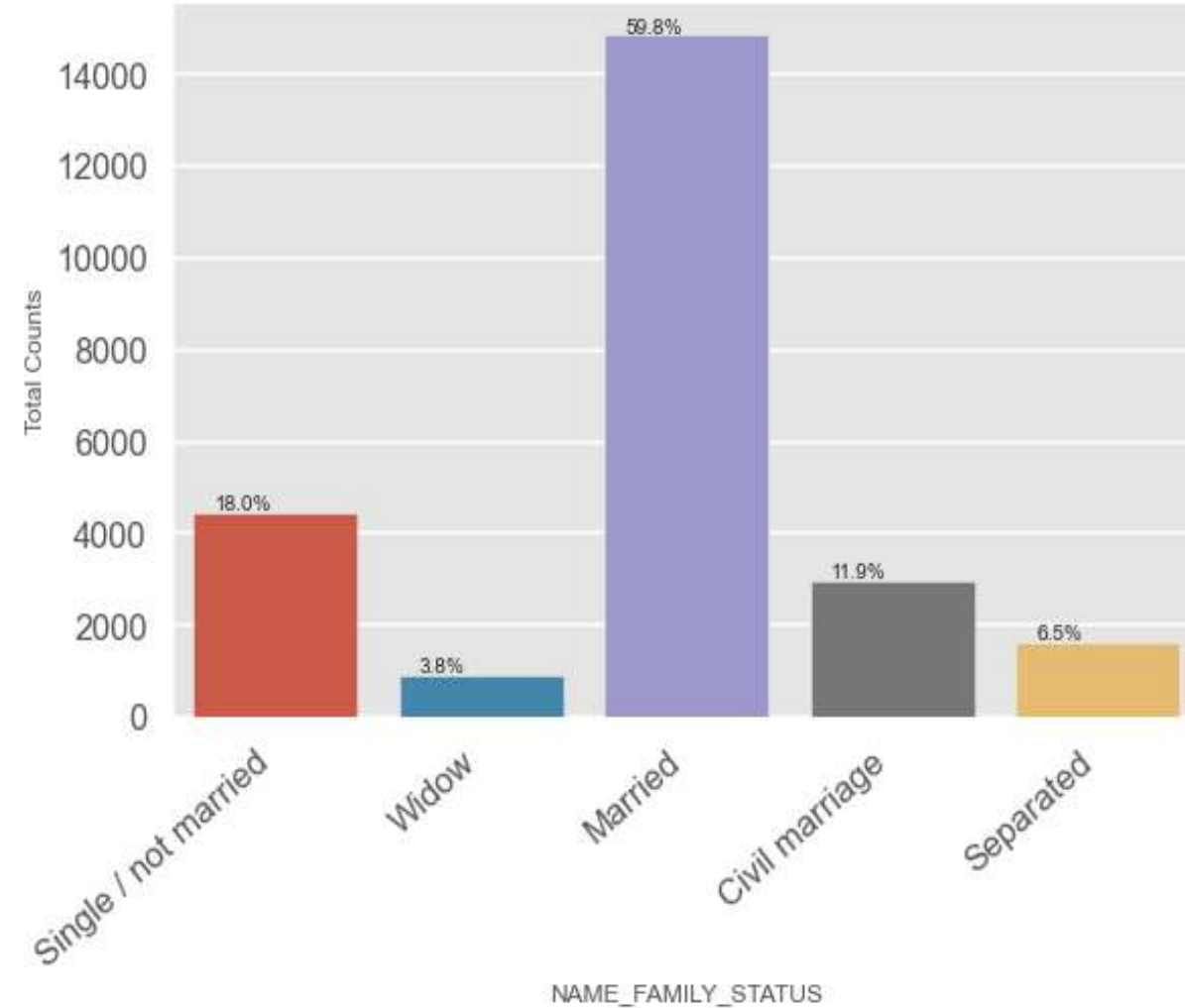


## Analysis on FAMILY\_STATUS in both Defaulters and Non\_Defaulters

Distribution of NAME\_FAMILY\_STATUS for Non-Defaulters

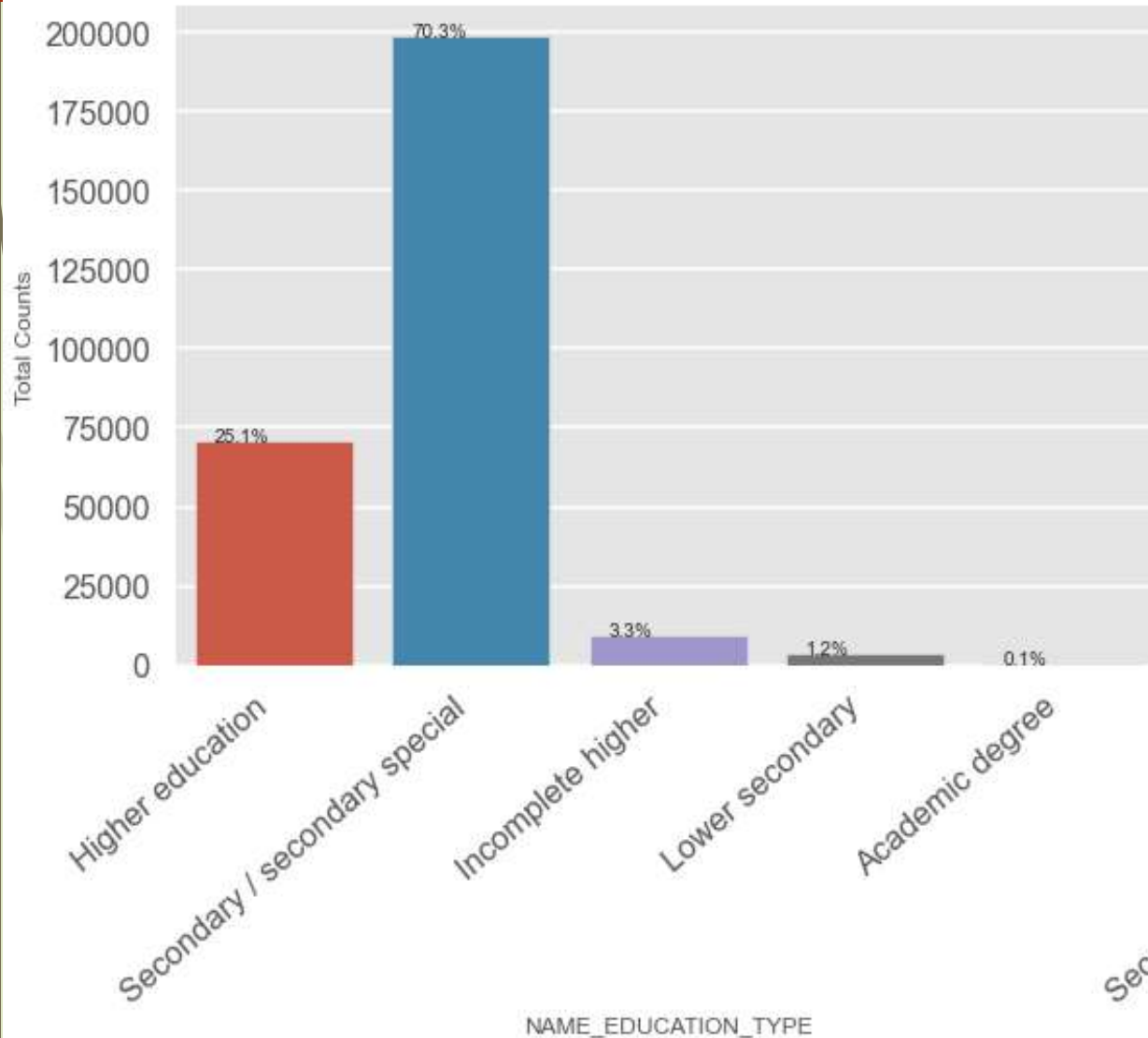


Distribution of NAME\_FAMILY\_STATUS for Defaulters

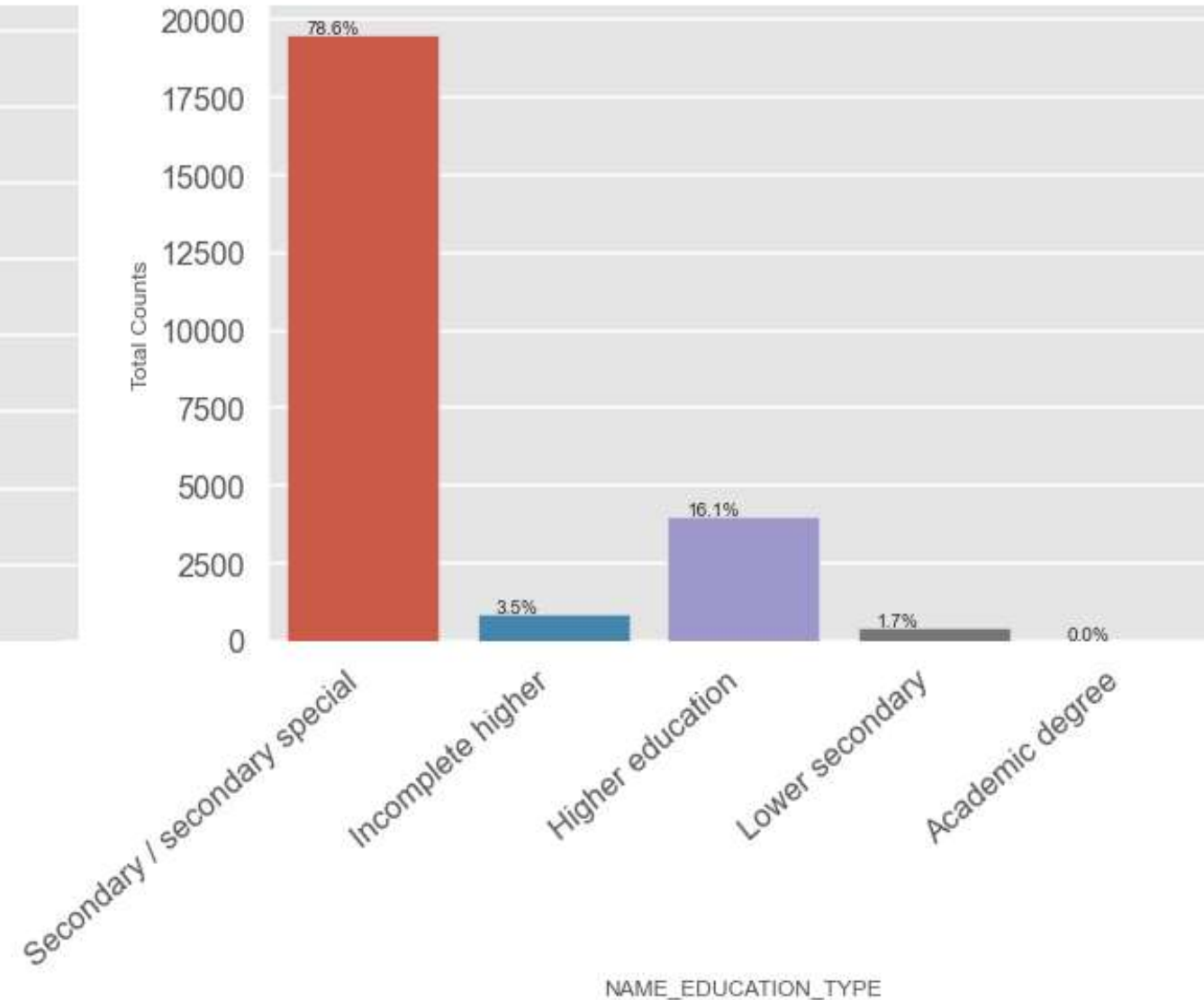


## Analysis on EDUCATION\_TYPE in both Defaulters and Non\_Defaulters

Distribution of NAME\_EDUCATION\_TYPE for Non-Defaulters

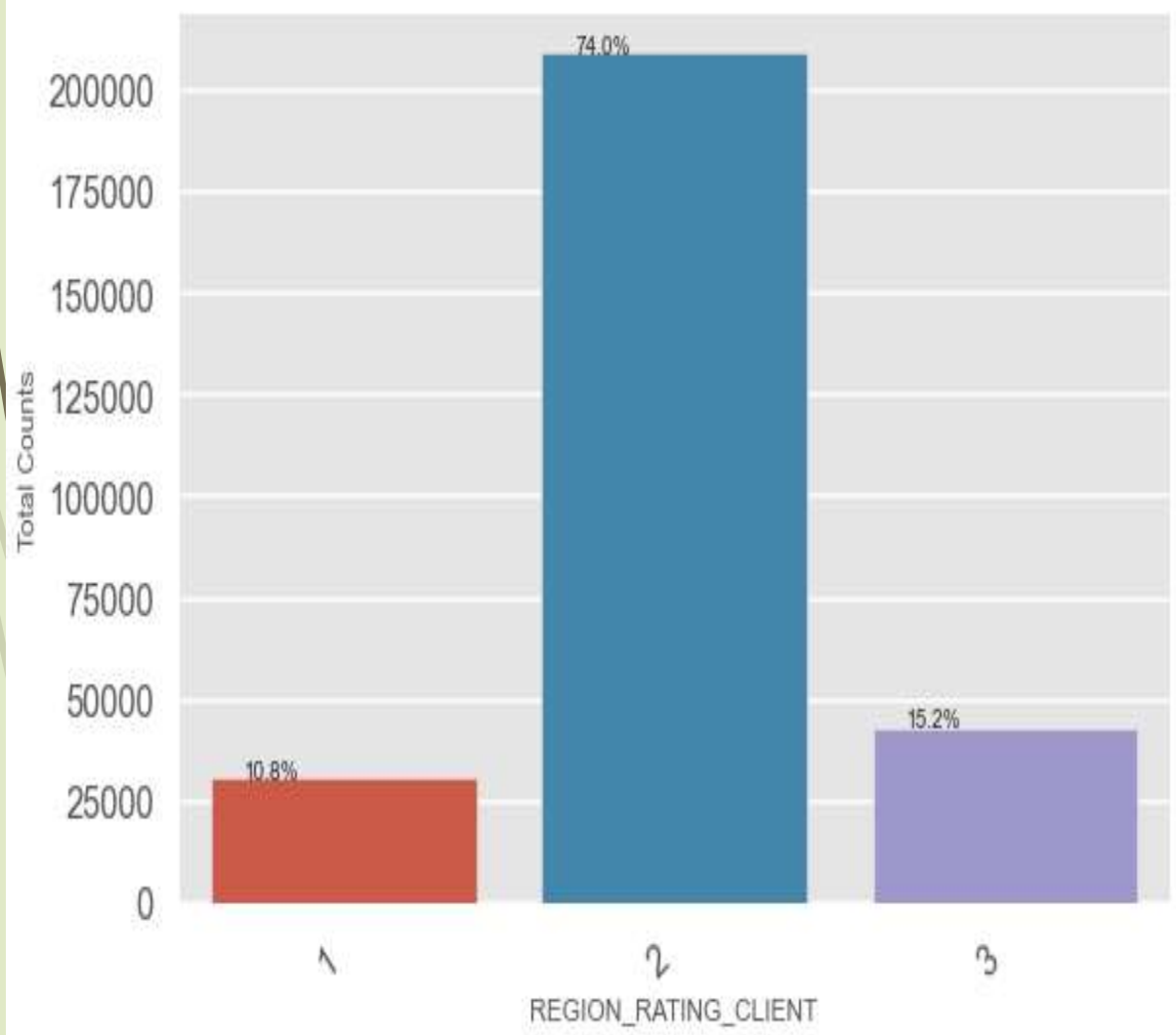


Distribution of NAME\_EDUCATION\_TYPE for Defaulters

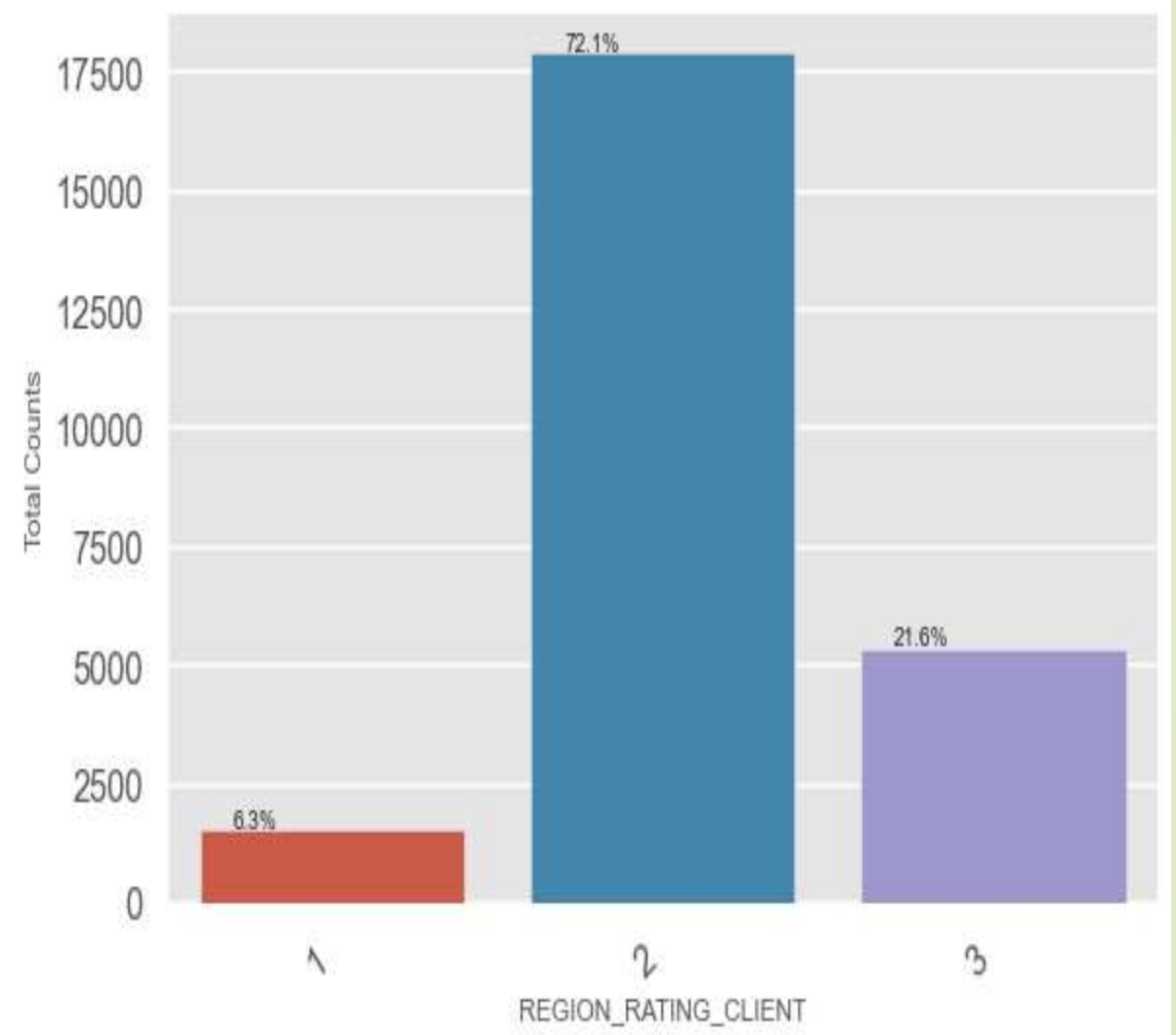


# Analysis on REGION in both Defaulters and Non\_Defaulters

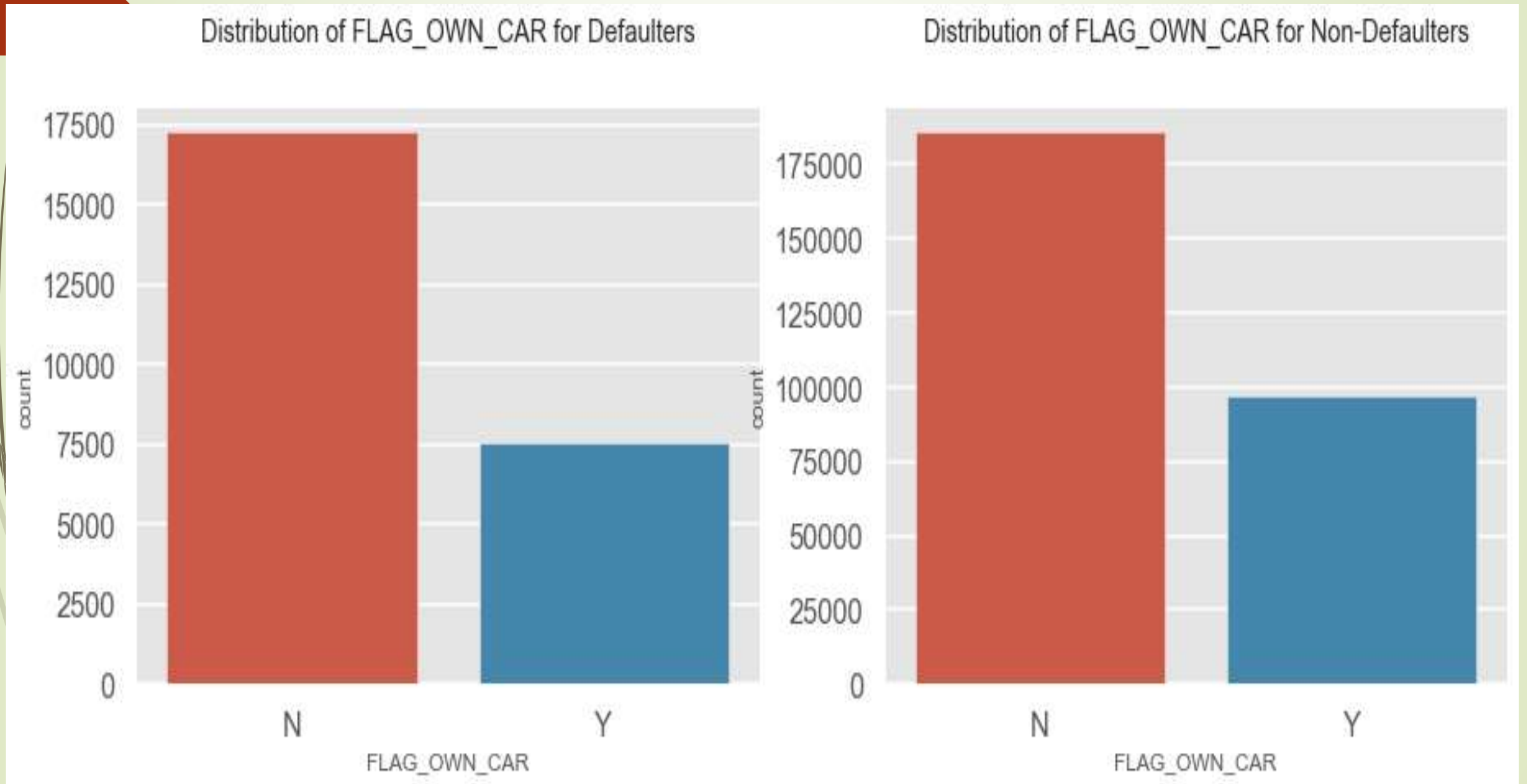
Distribution of REGION\_RATING\_CLIENT for Non-Defaulters



Distribution of REGION\_RATING\_CLIENT for Defaulters



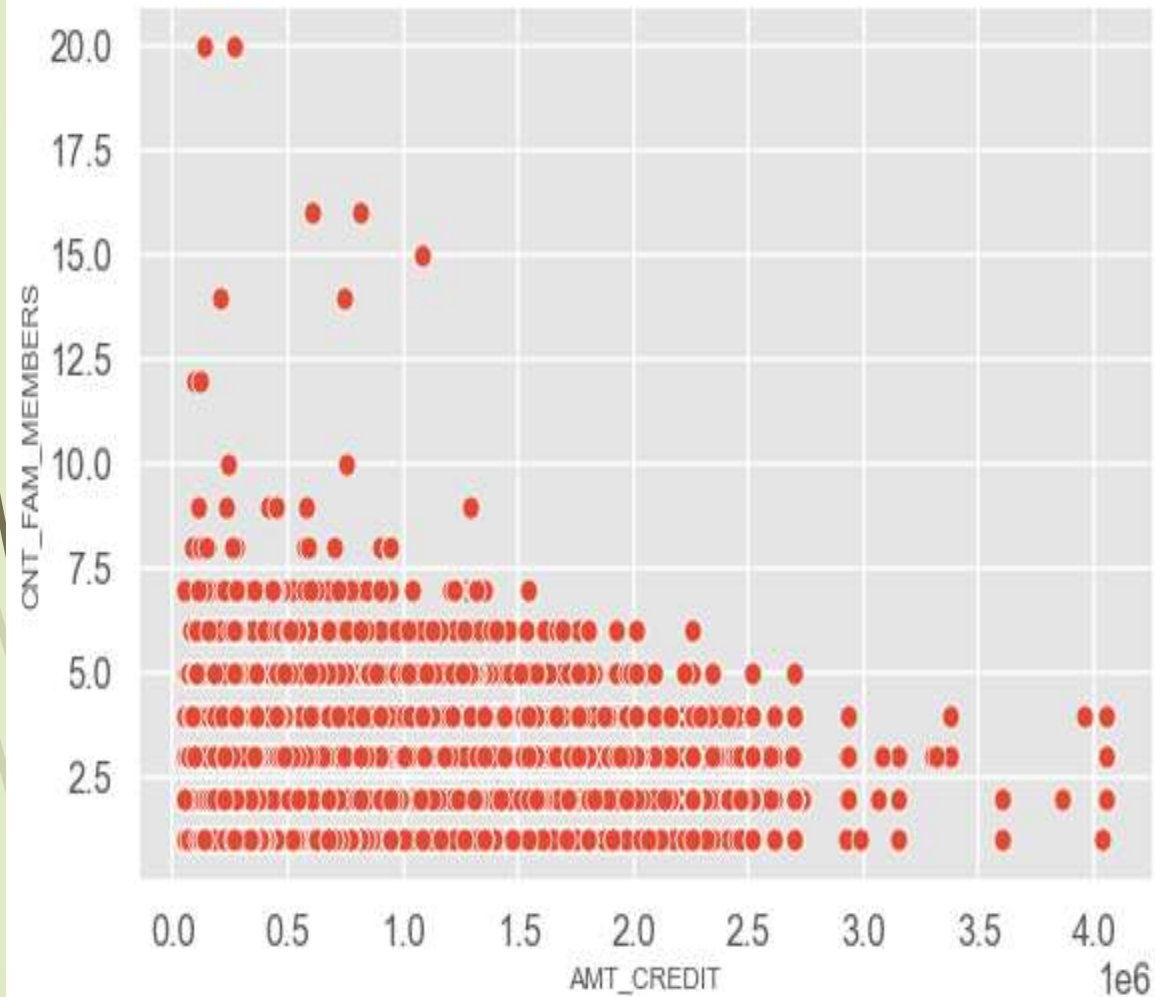
## Analysis on OWN\_CAR in both Defaulters and Non\_Defaulters



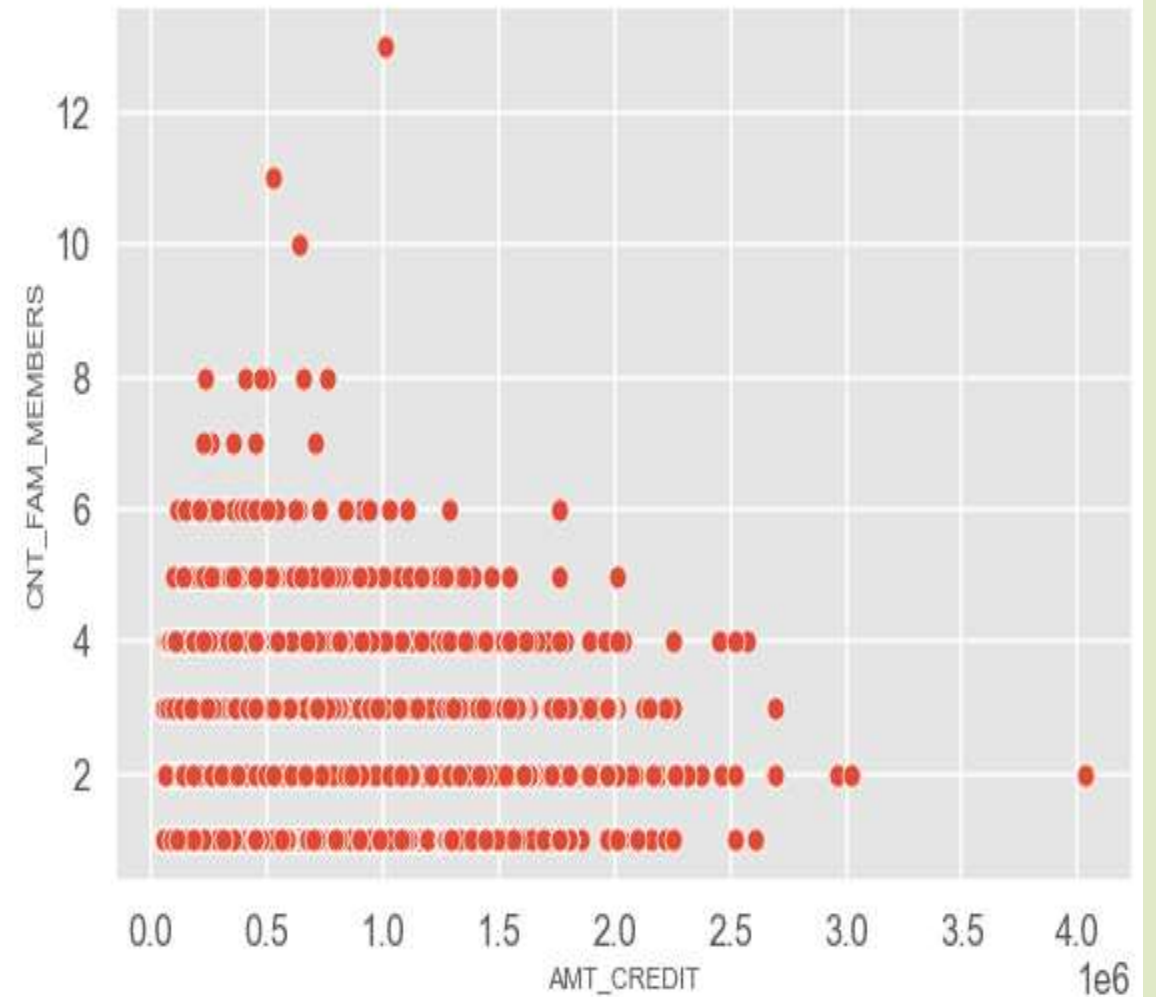


## Bivariate Analysis on numerical\_columns:

AMT\_CREDIT vs CNT\_FAM\_MEMBERS for Non-Defaulters

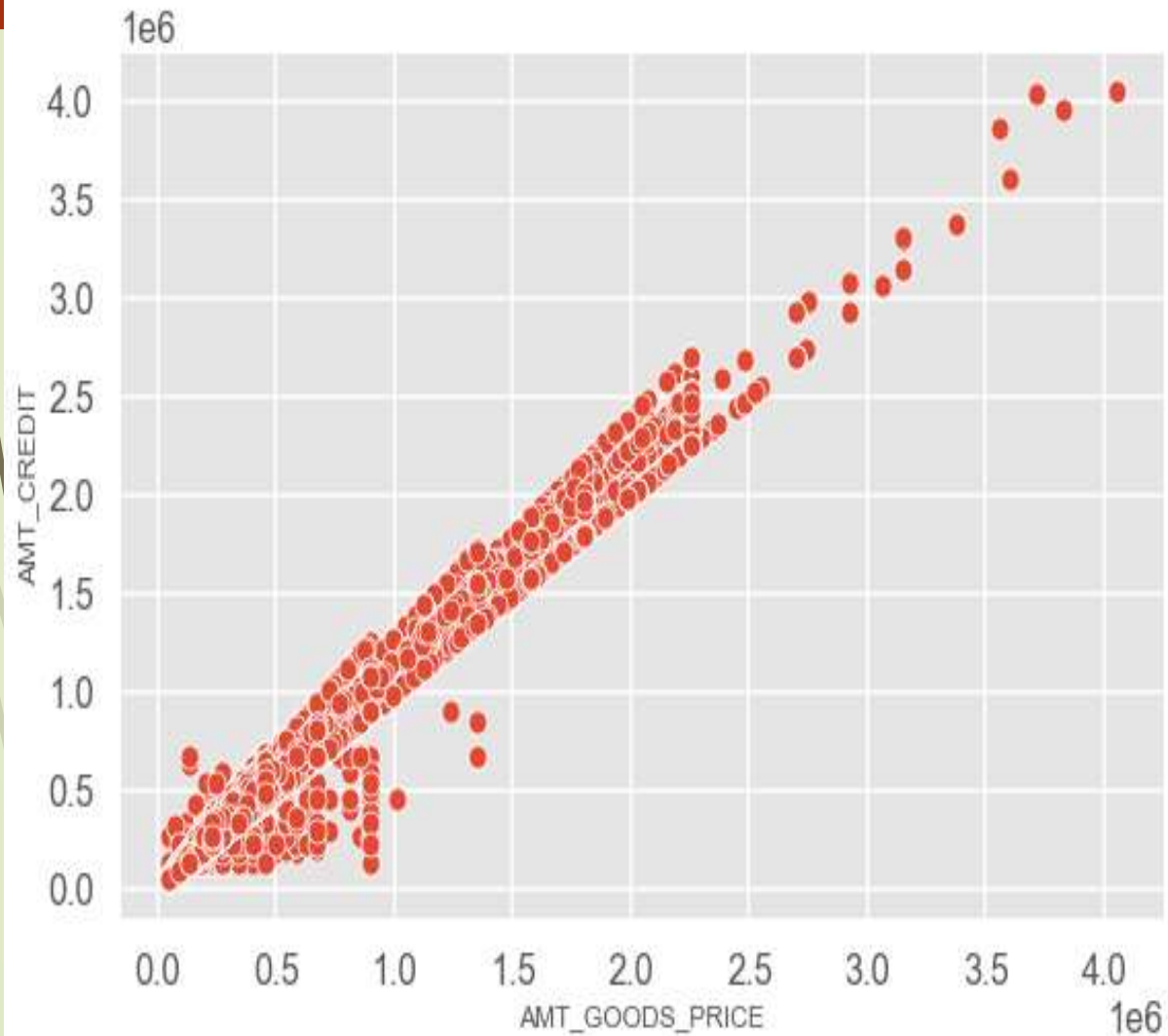


AMT\_CREDIT vs CNT\_FAM\_MEMBERS for Defaulters

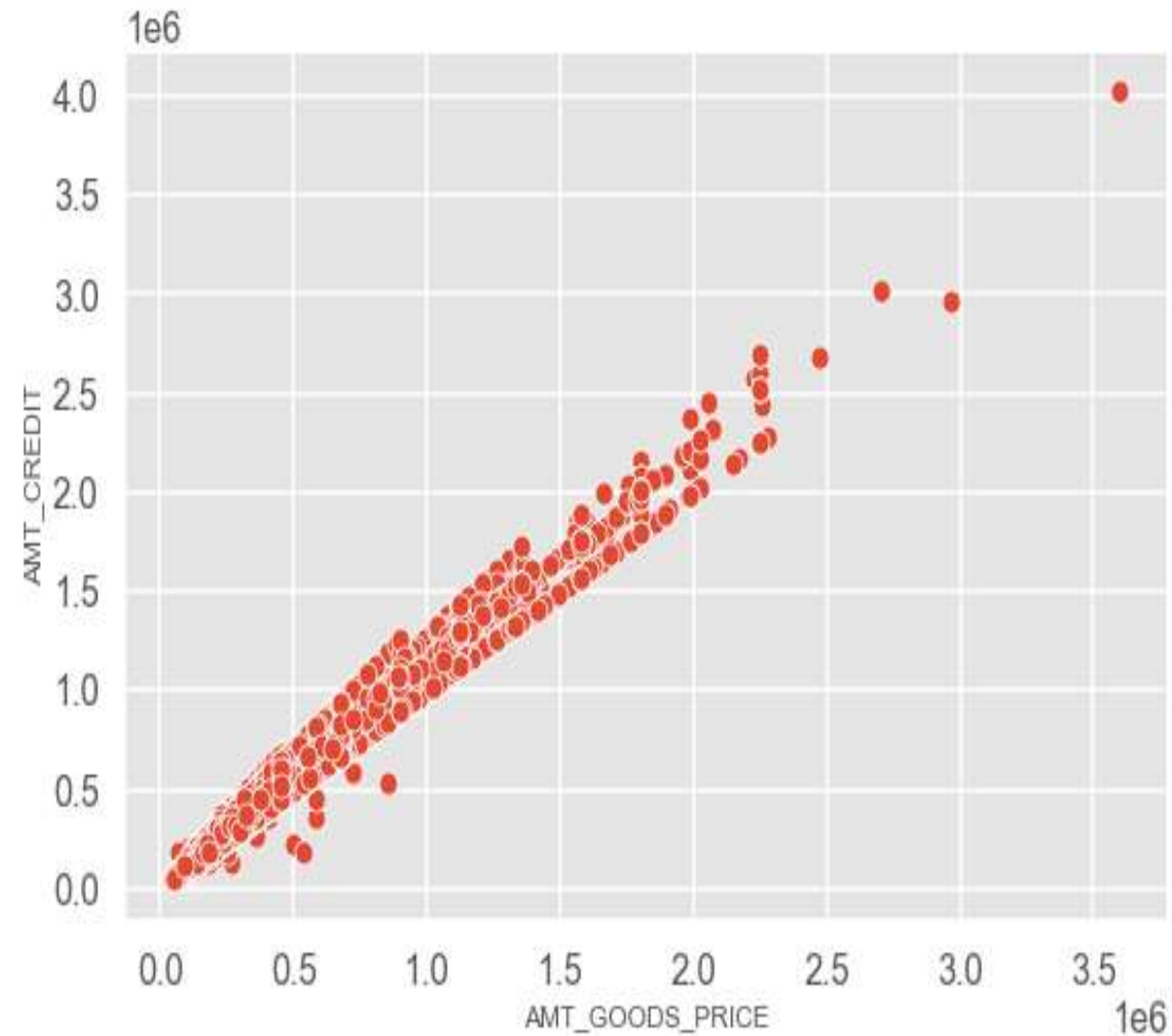


## Bivariate Analysis on numerical\_columns:

AMT\_GOODS\_PRICE vs AMT\_CREDIT for Non-Defaulters

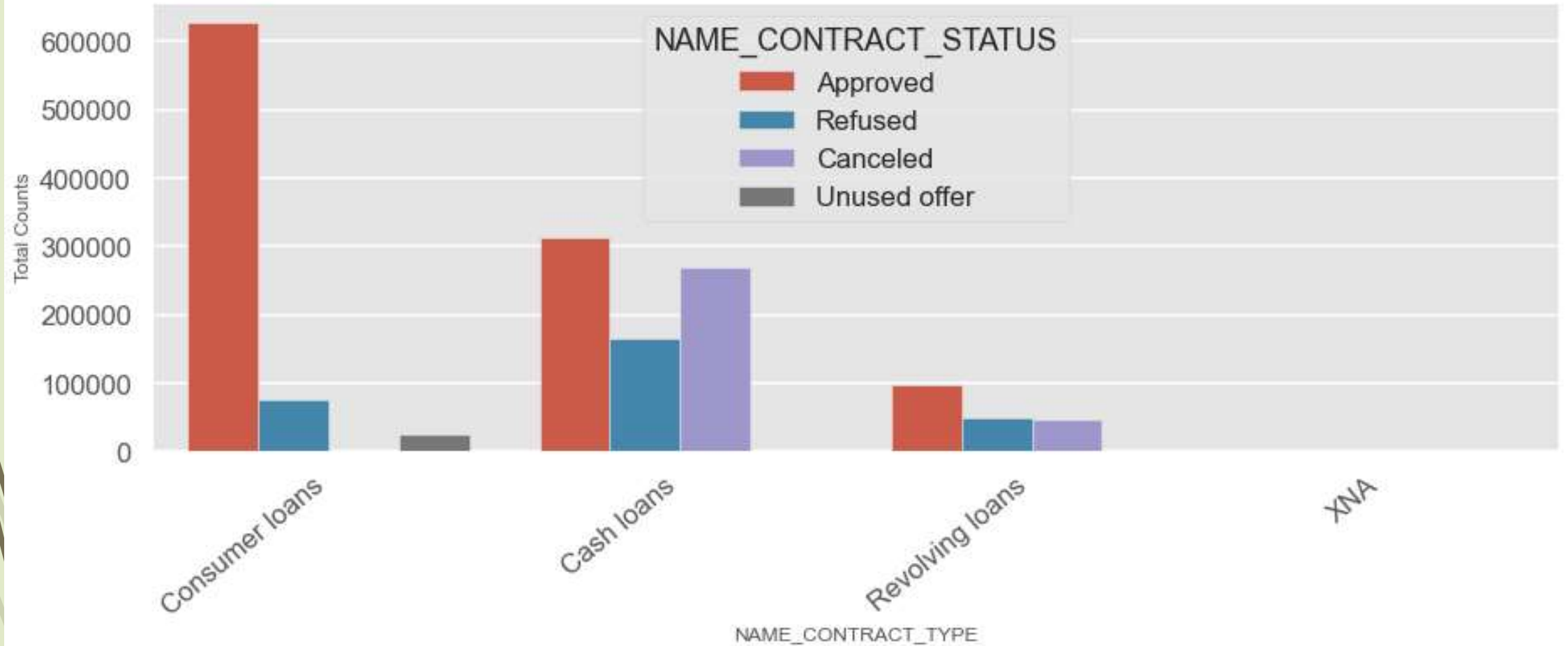


AMT\_GOODS\_PRICE vs AMT\_CREDIT for Defaulters

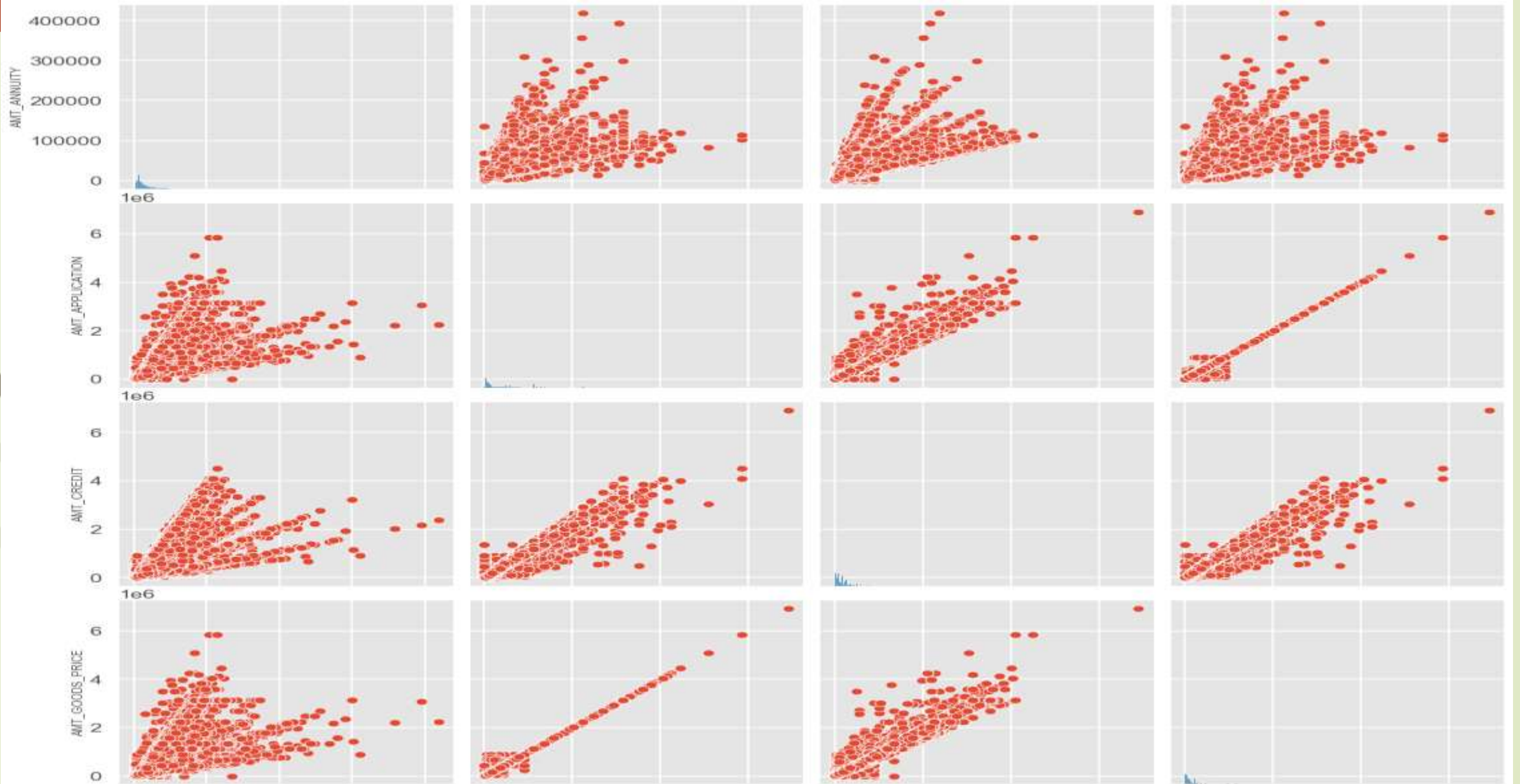




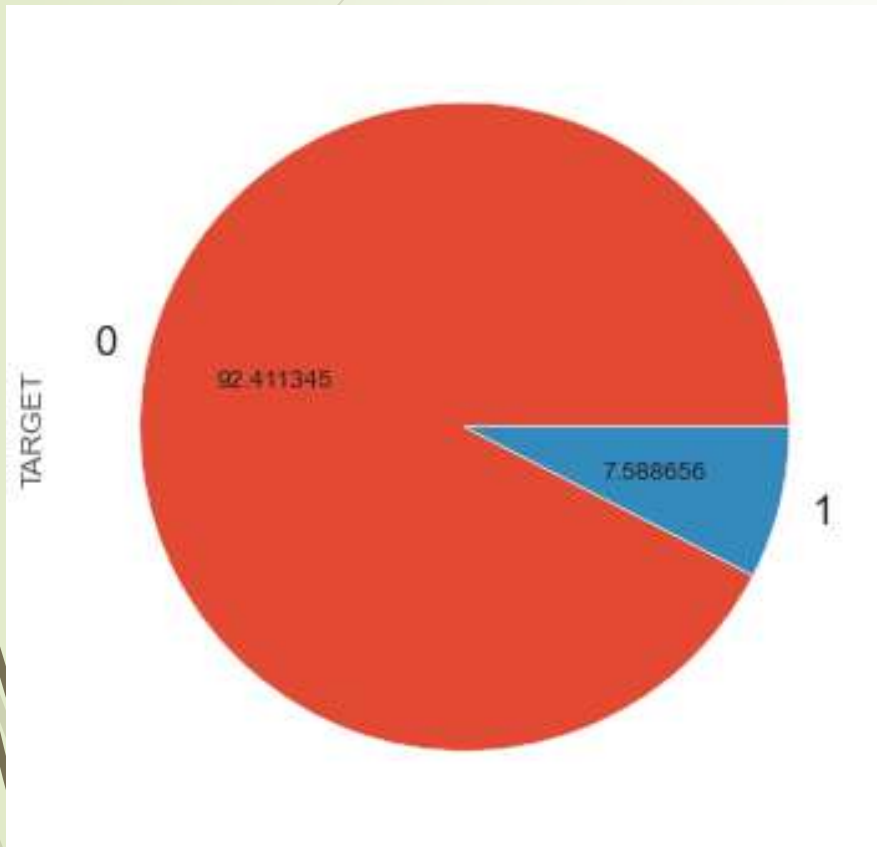
Distribution of NAME\_CONTRACT\_TYPE



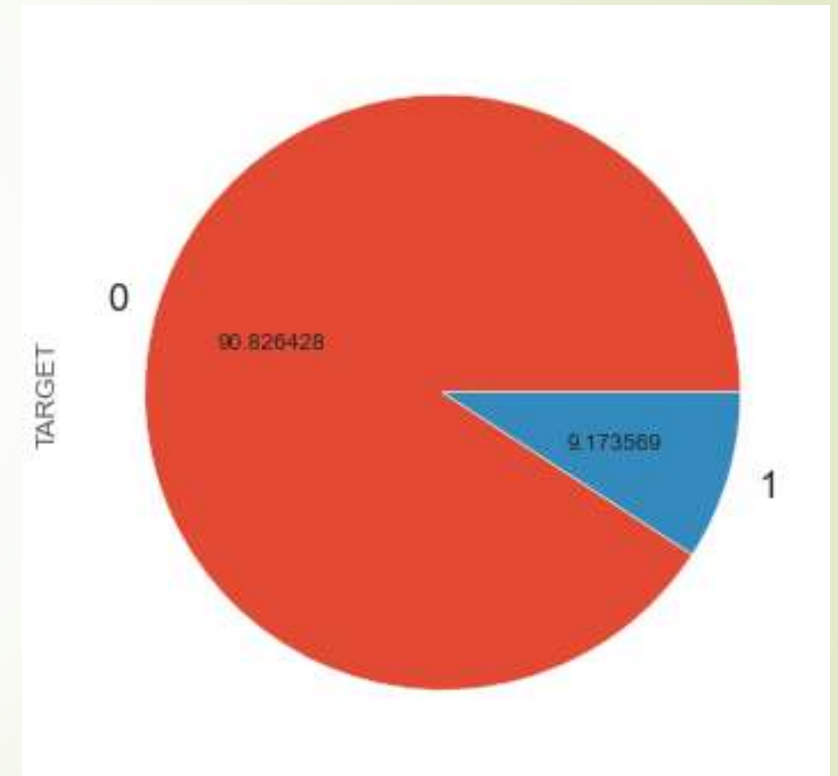
## Bivariate Analysis on numerical\_columns



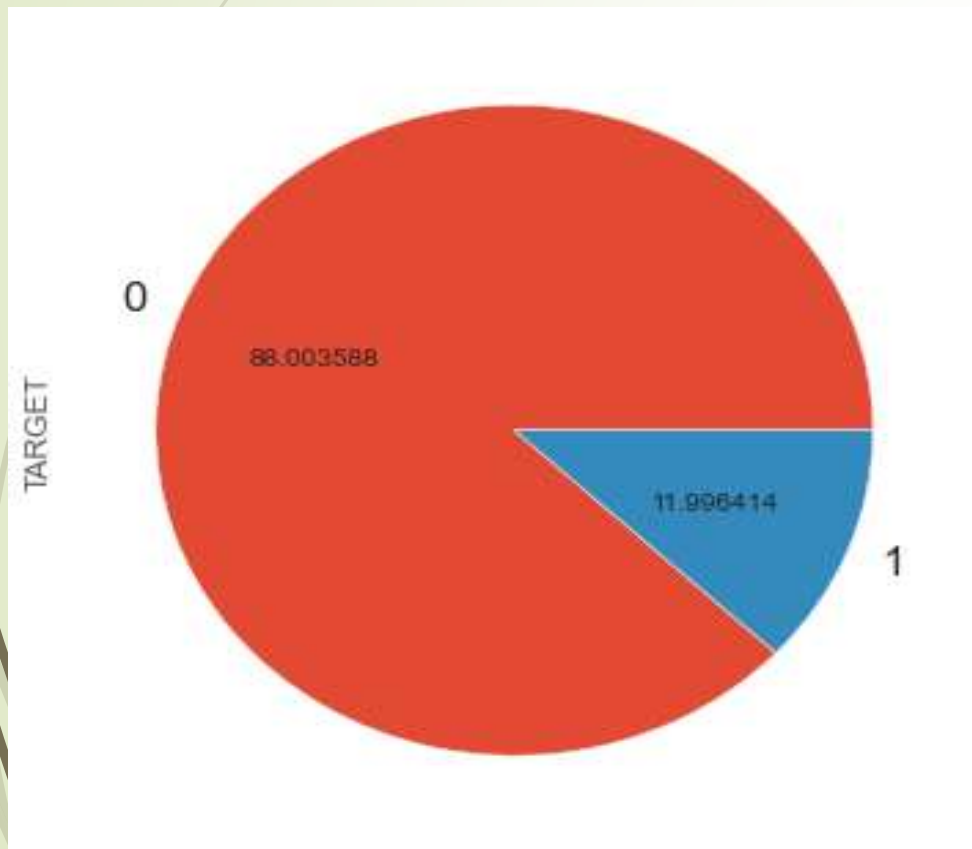
Contract Status  
Approved:



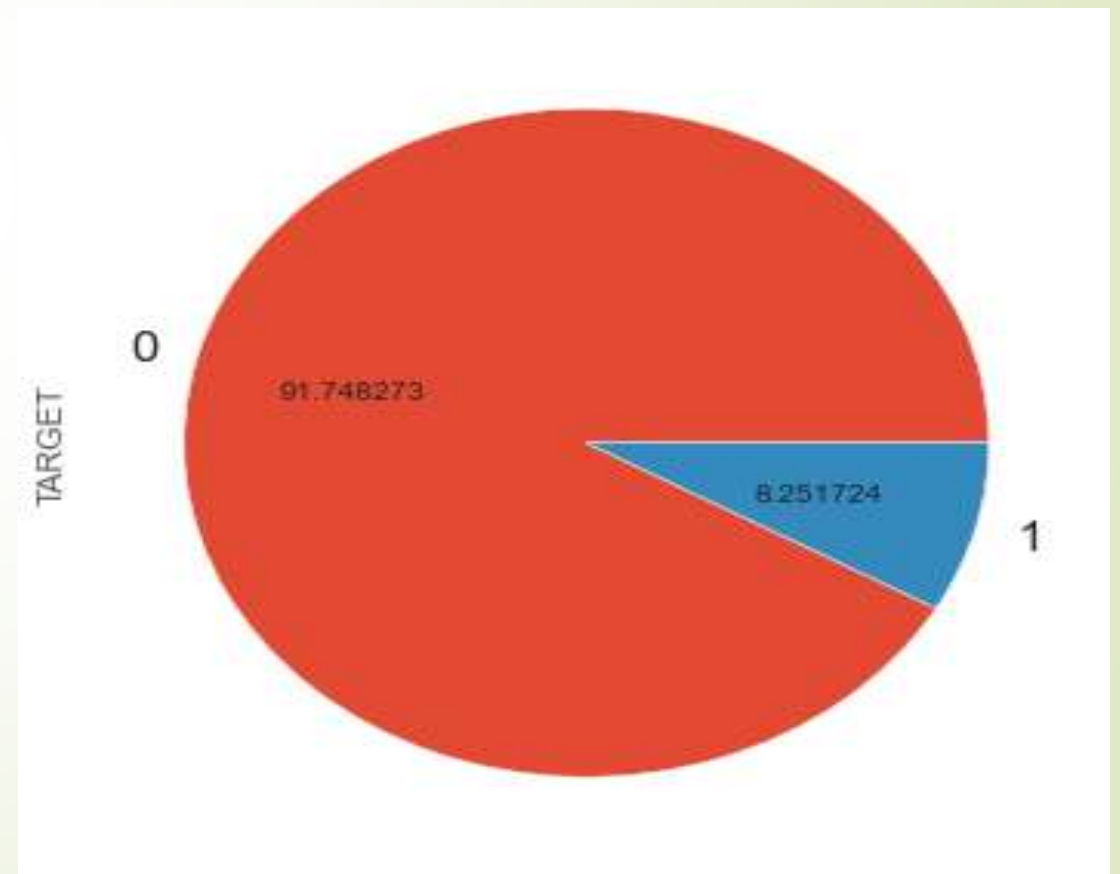
Contract Status  
Cancelled:



Contract Status  
Refused:



Unused  
Offer:





## Conclusions:

Previous applications with Refused, Cancelled, Unused loans have default, It shows that the company had Cancelled previous application but approved the current application.

60% of defaulters are Working applicants.

The Proportion of Defaulters is 8.7%.

Females are more to lend the Loan.

Revolving loans are less It Should be increased.