

# **Credited Exploratory** **Data Analysis**

**Abhijit Takalkar**

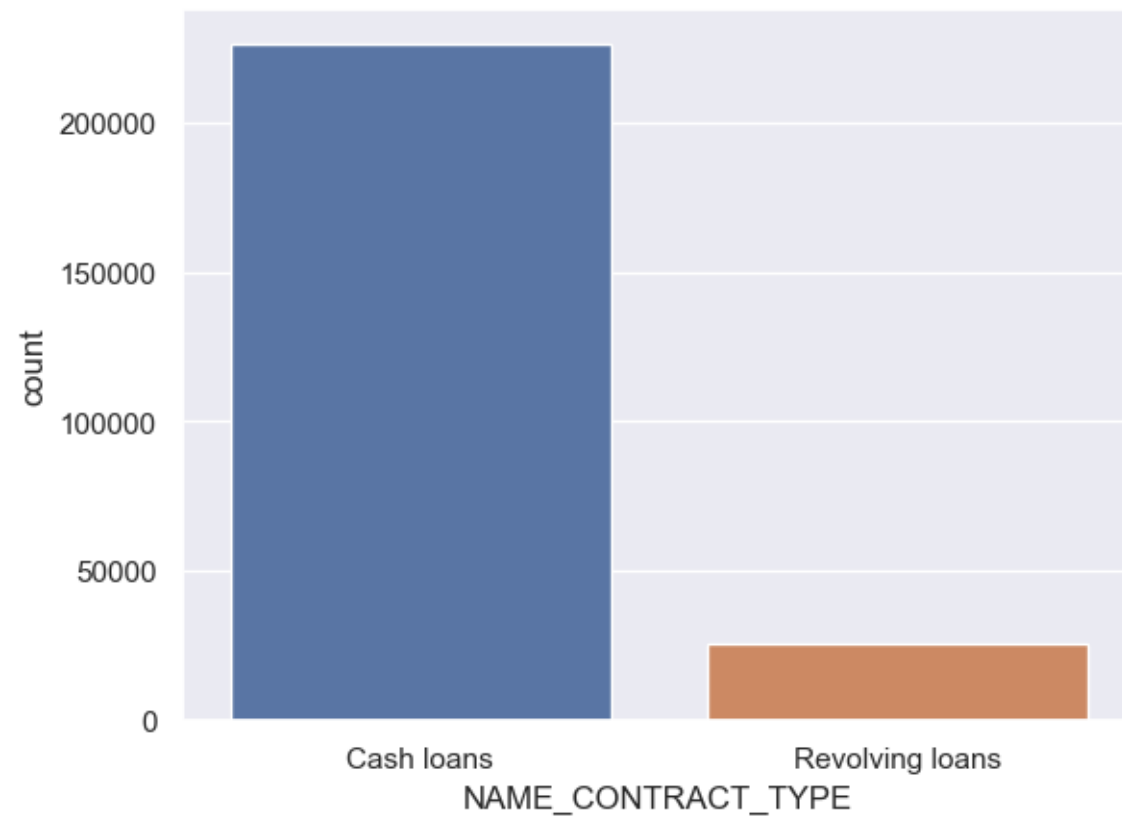
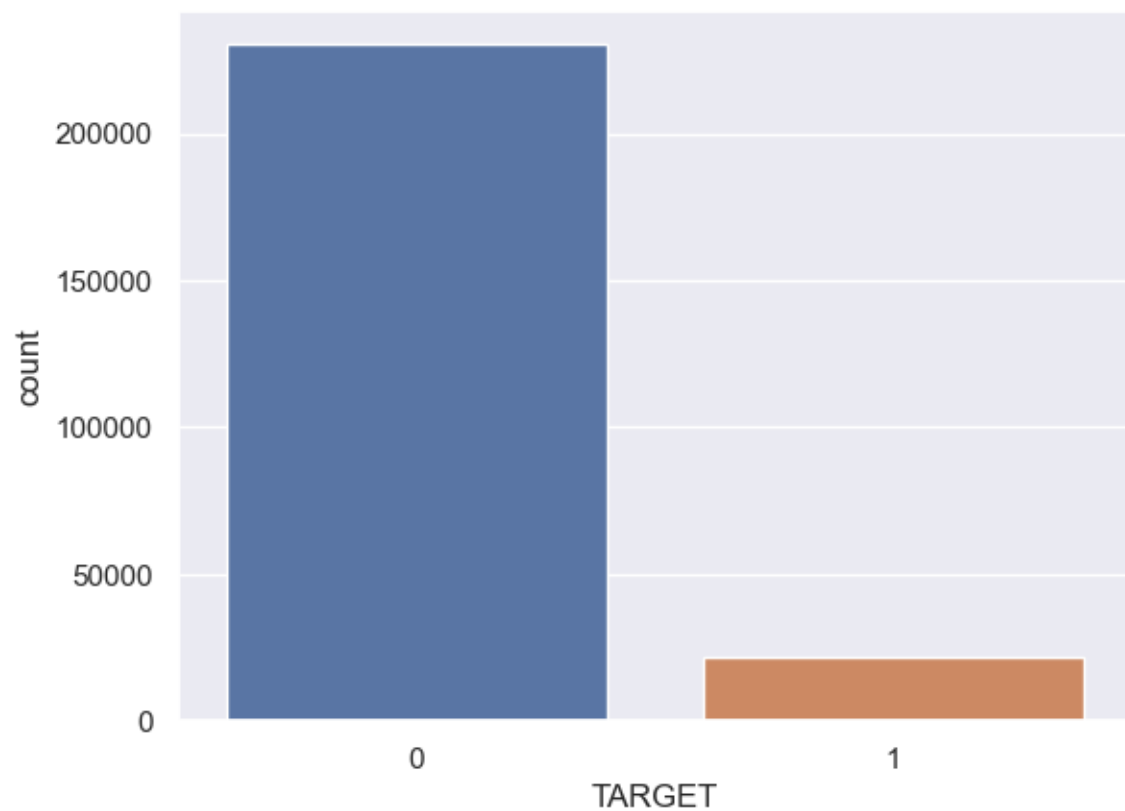
# Objective

- When the company receives a loan application, the company has to decide on loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:
- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely 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.

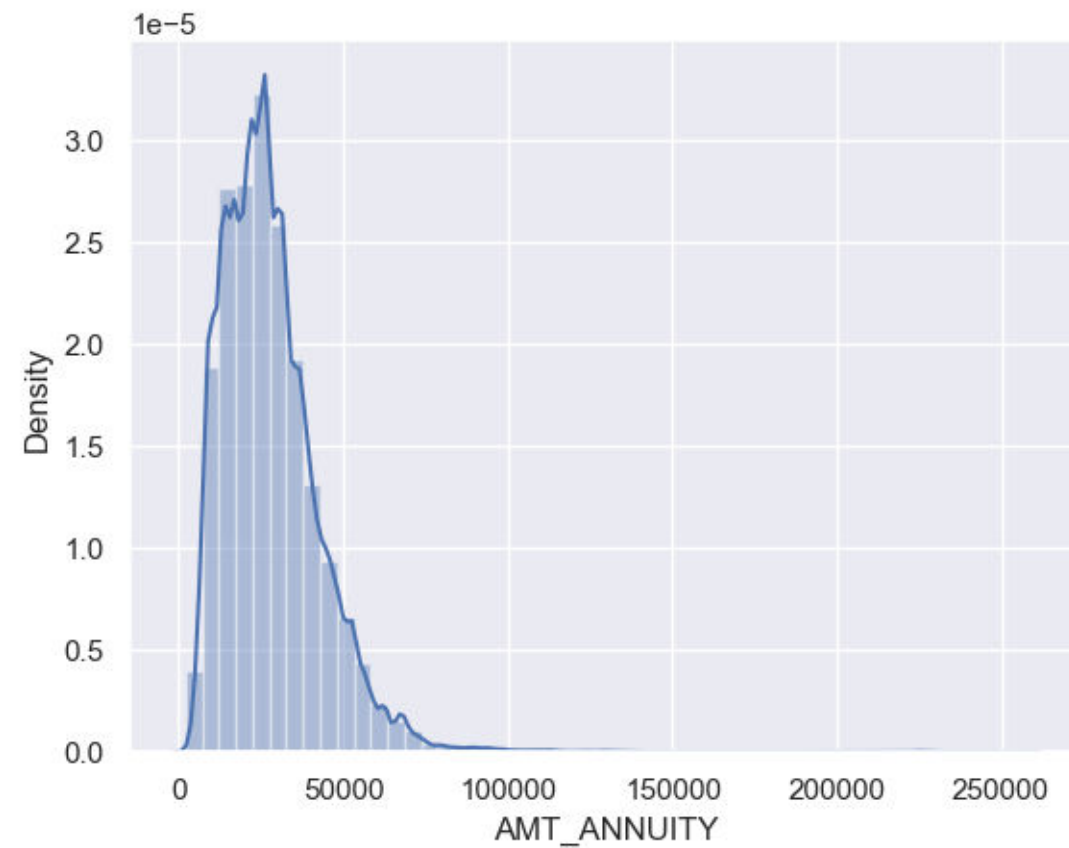
# Process

- Data Understanding & Sourcing
- Checking of data quality issues and binning
- Checking of data imbalance and univariate, bivariate, multivariate analysis and correlation
- Merging of both datasets i.e. application\_data & previous\_application
- Data analysis by univariate, bivariate, multivariate analysis
- Recommendation & Risks

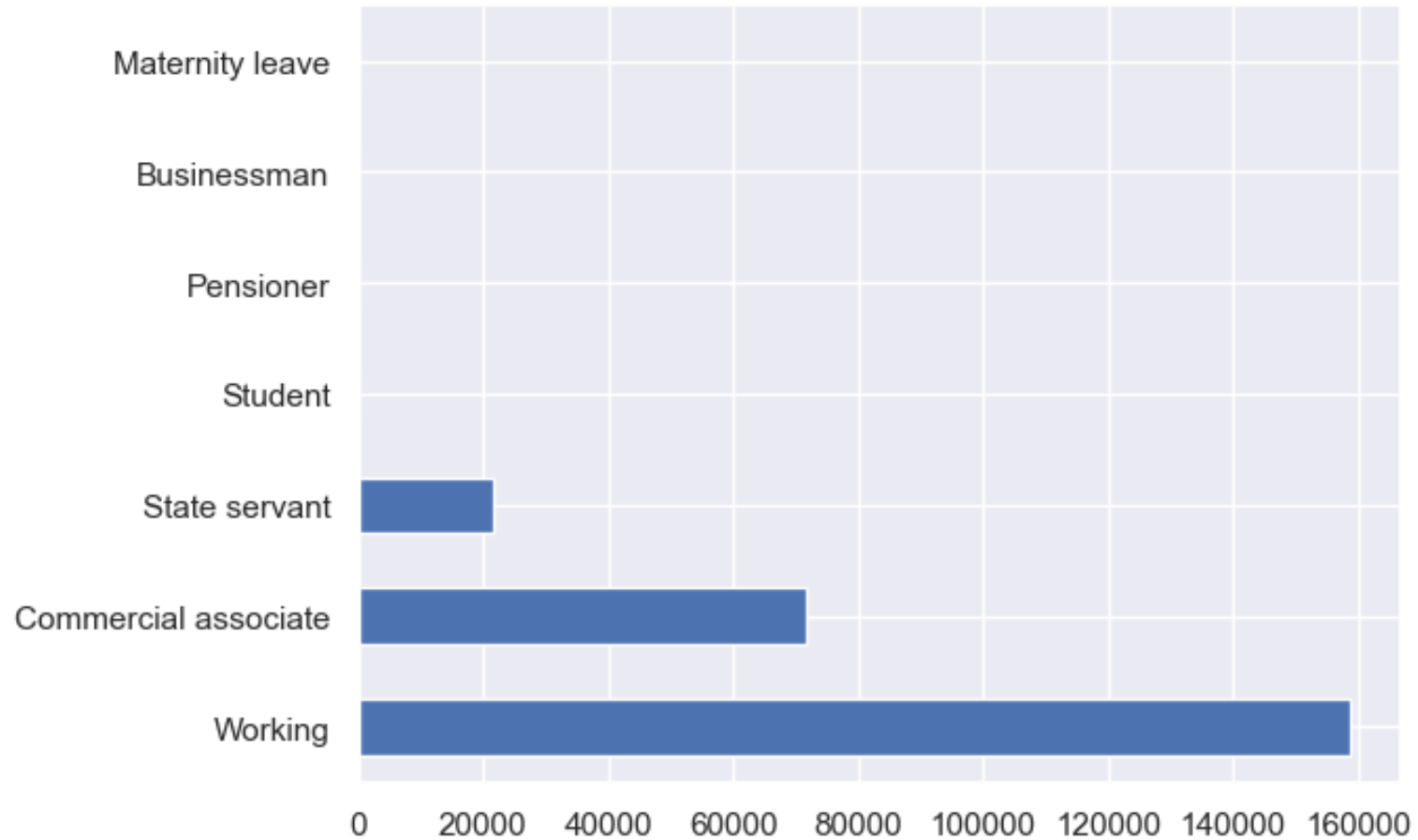
# Univariate Analysis of application\_data



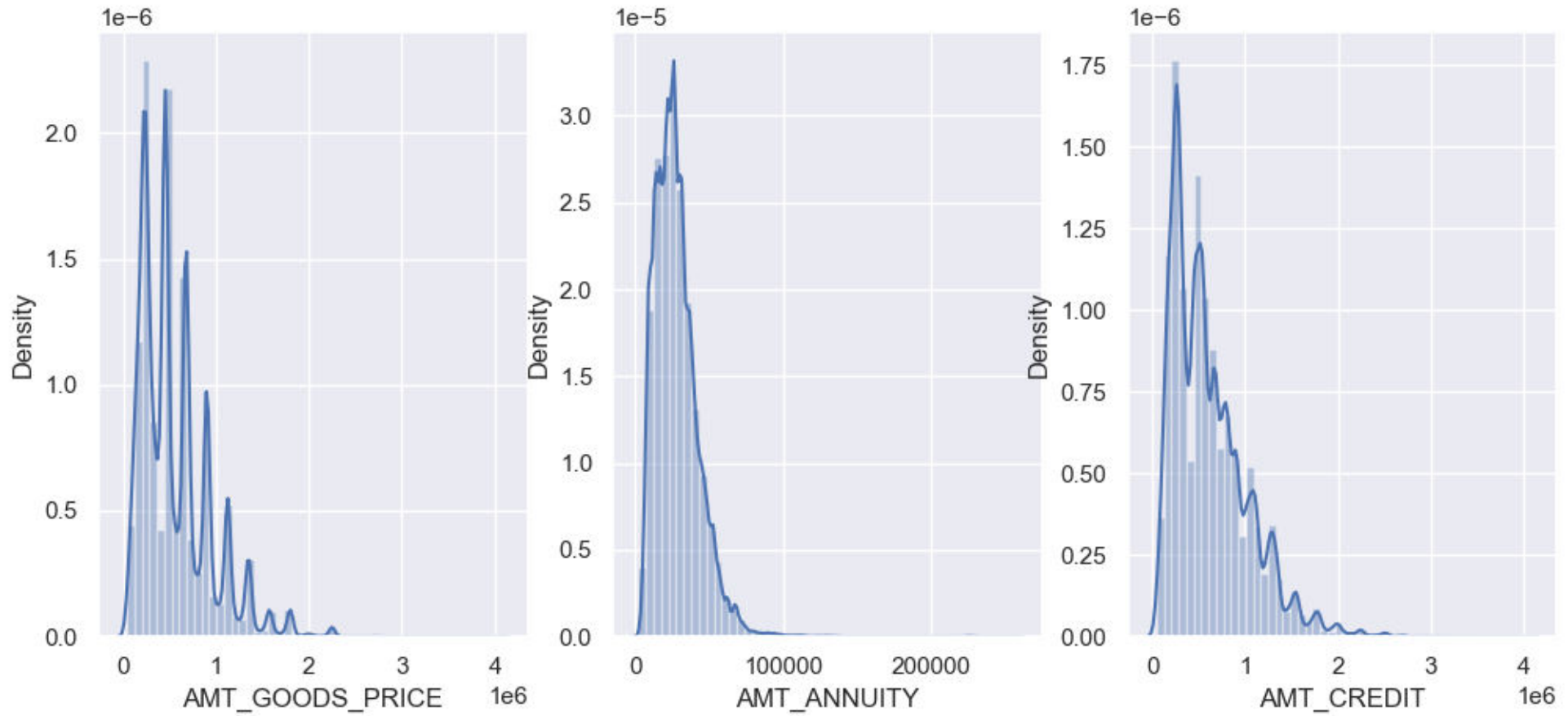
# Code Gender Amt Annuity



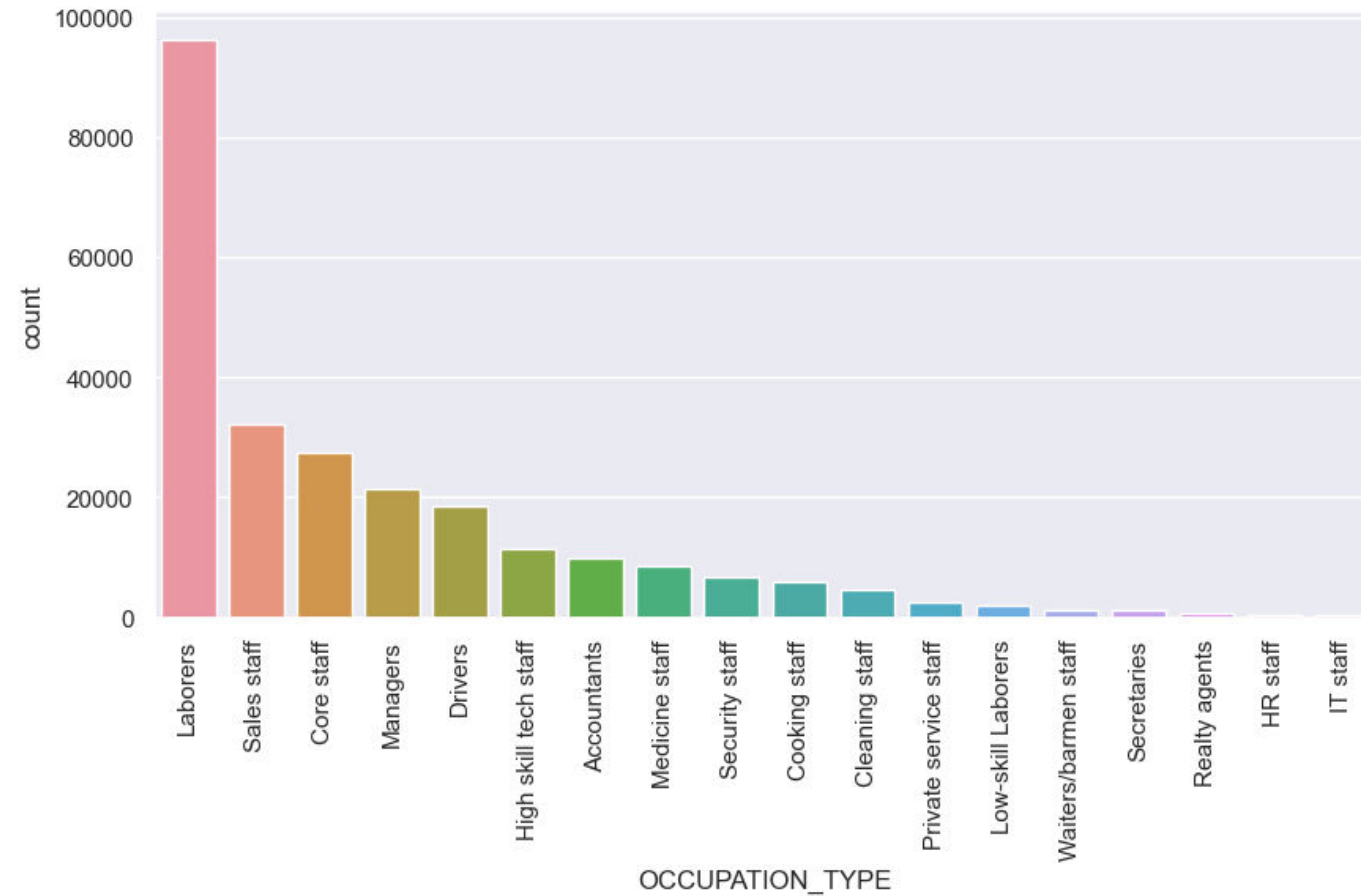
# Name Income Type



# Distplot

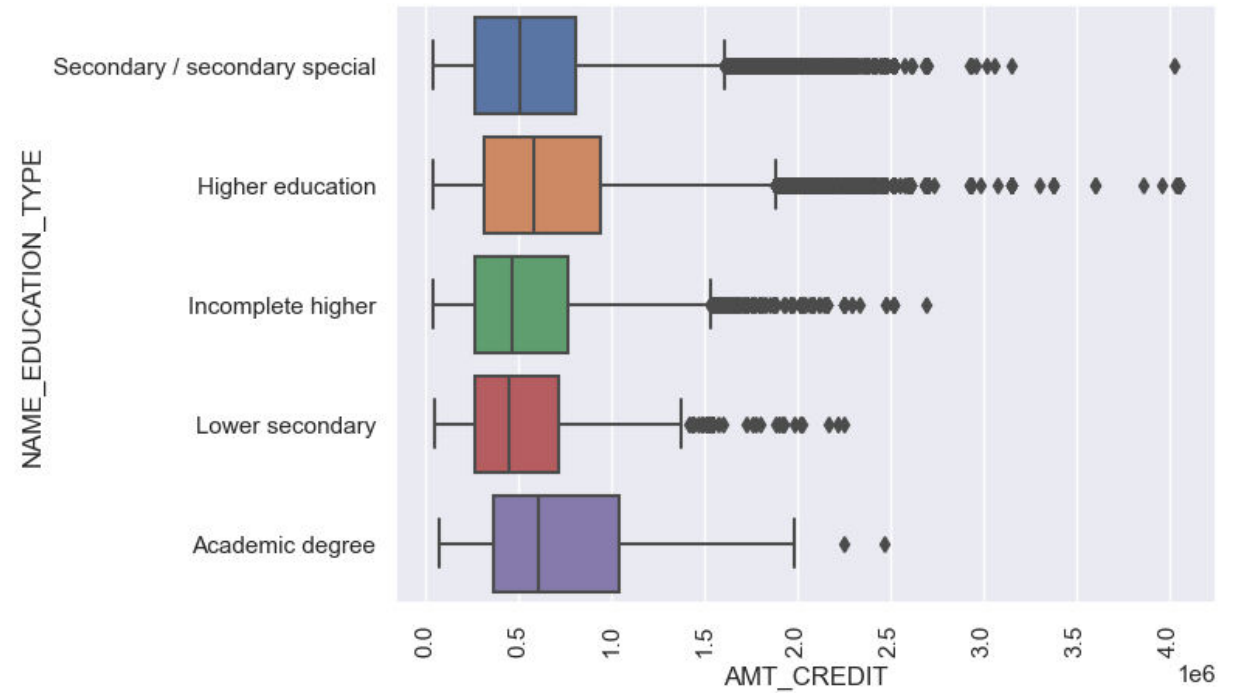
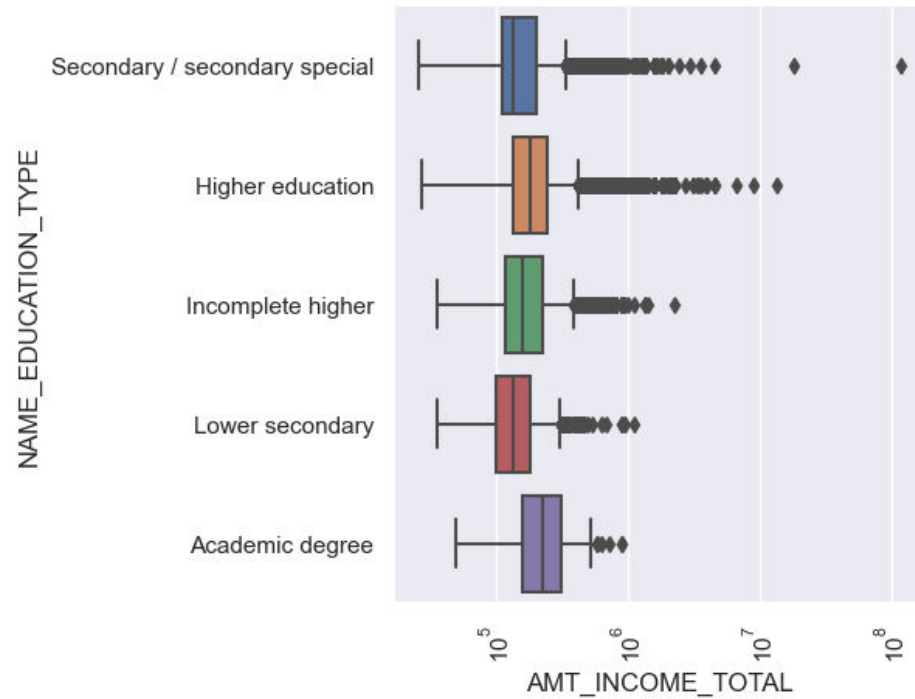


# Occupation Type

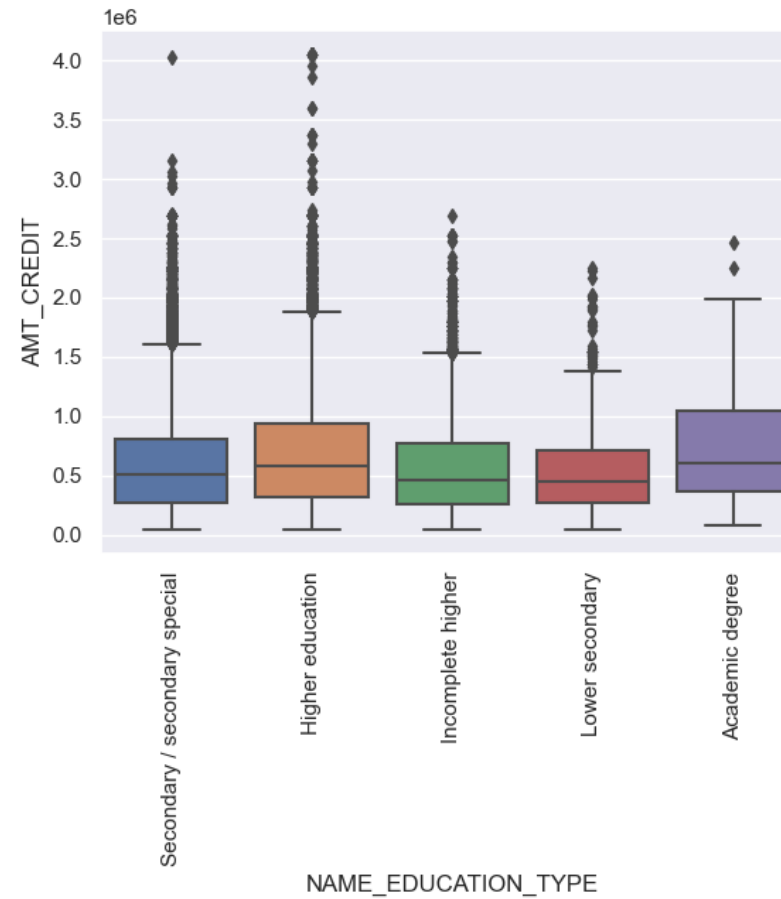




# Bivariate Analysis of application\_data



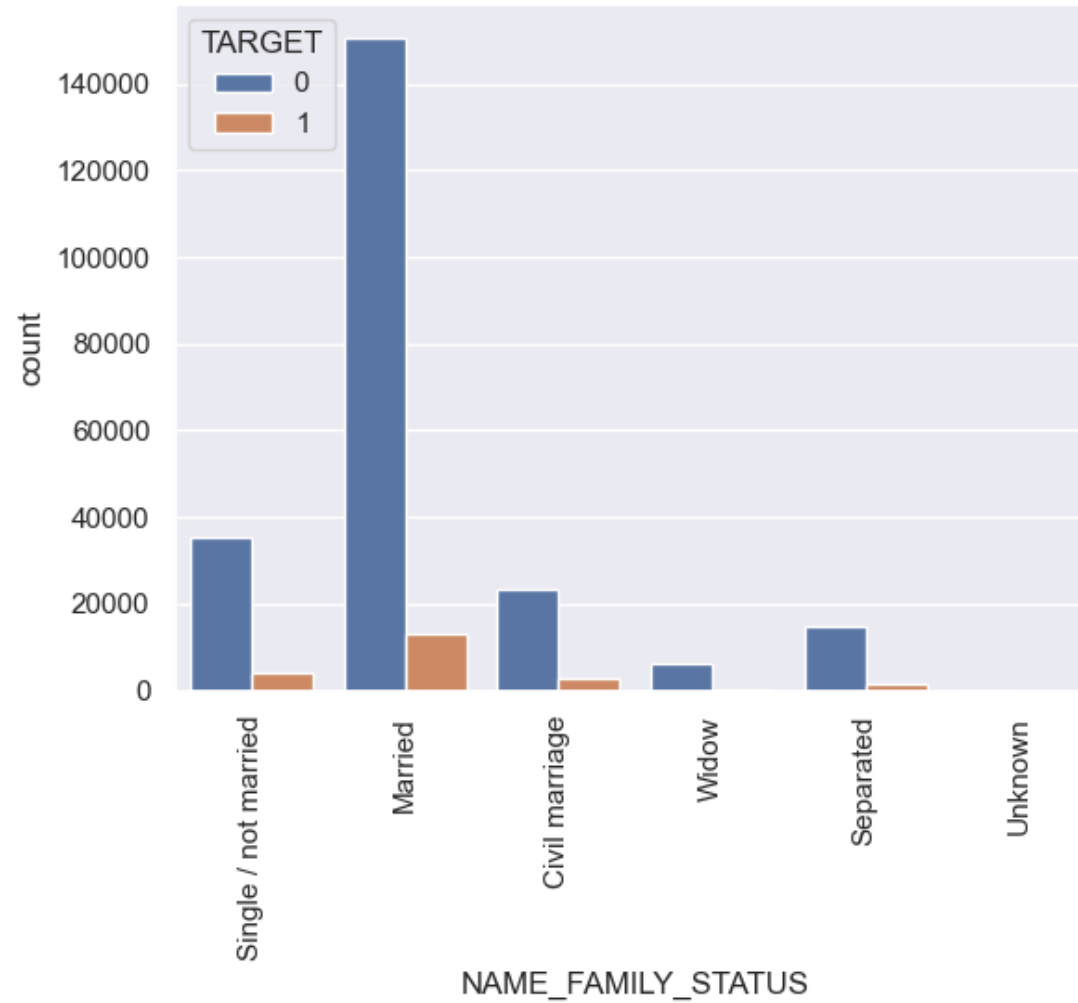
# Bivariate Analysis of Name\_Education\_Type Amt\_Annuity



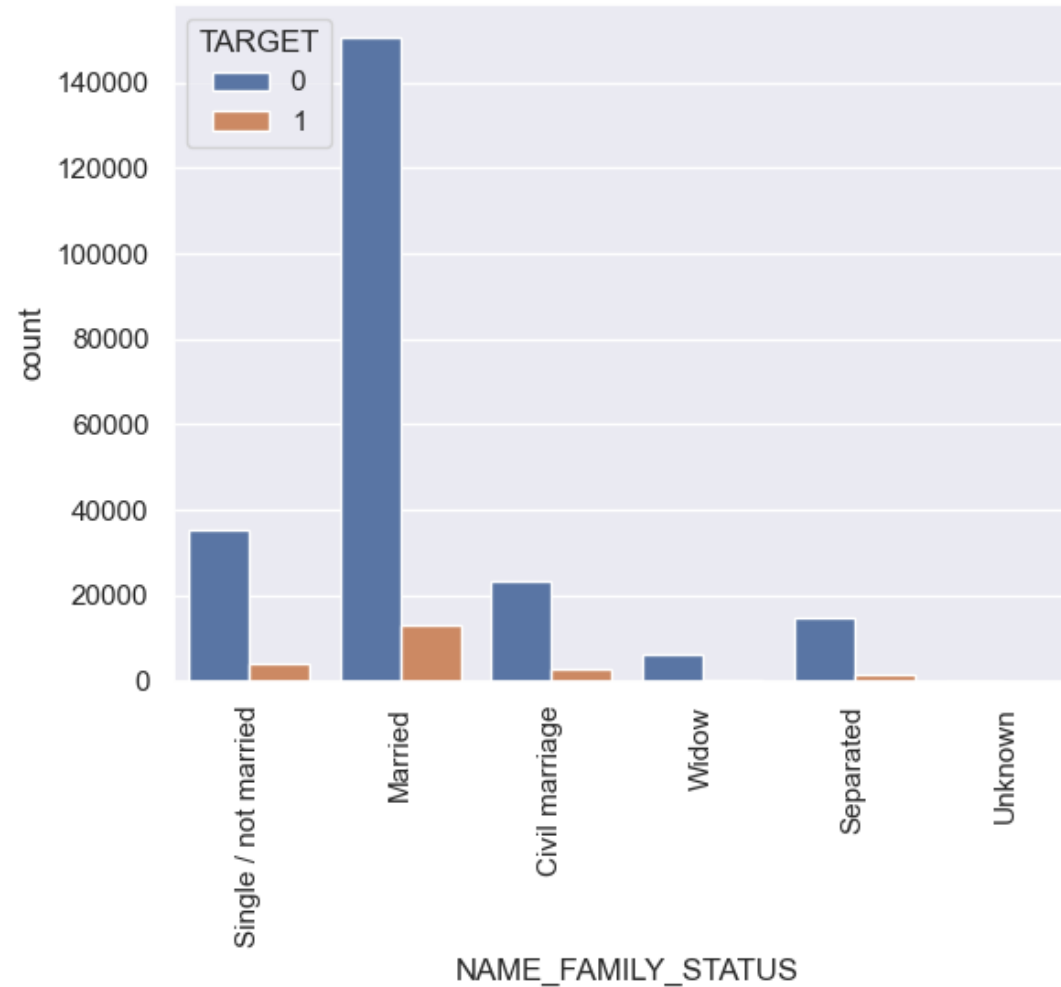
# Bivariate Analysis of Target and Code Gender



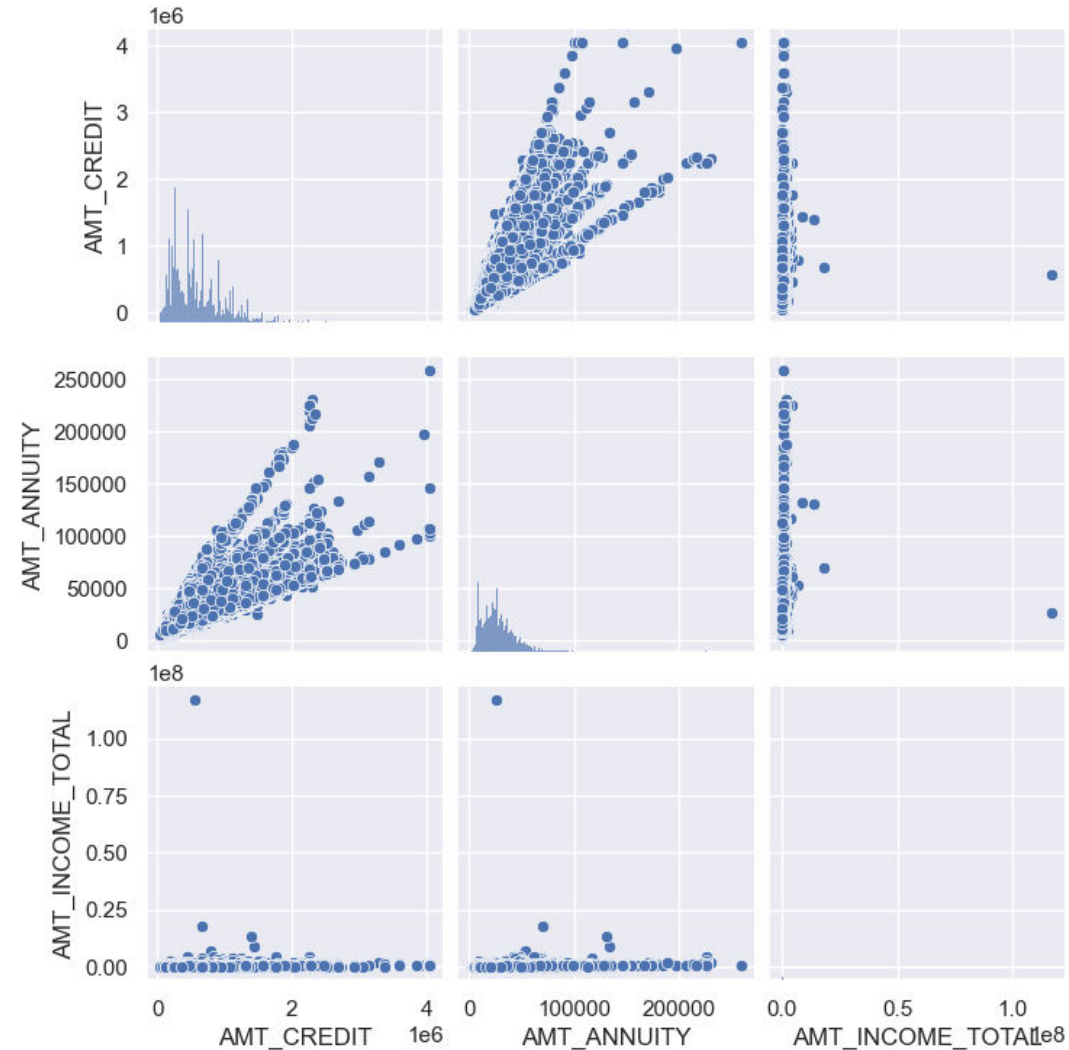
# Bivariate Analysis of Target Name Family



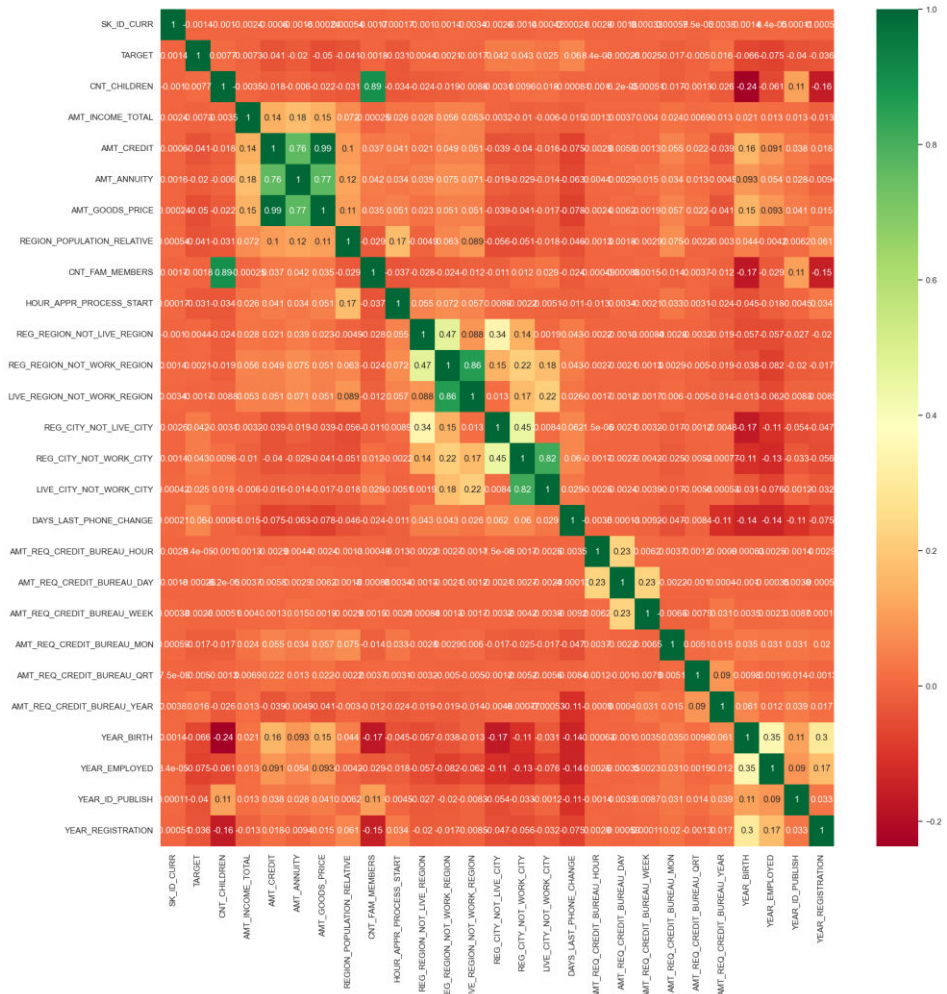
# Bivariate Analysis Target Name Family Status



# Correlation Amt Credit, Amt Annuity, AMT TOTAL INCOME

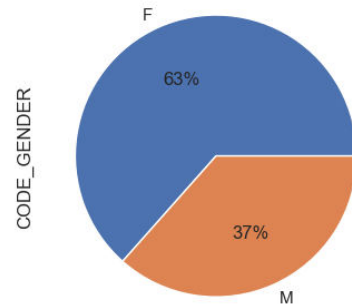


# Correlation Subplot Graph



# Bivariate analysis after merging both data

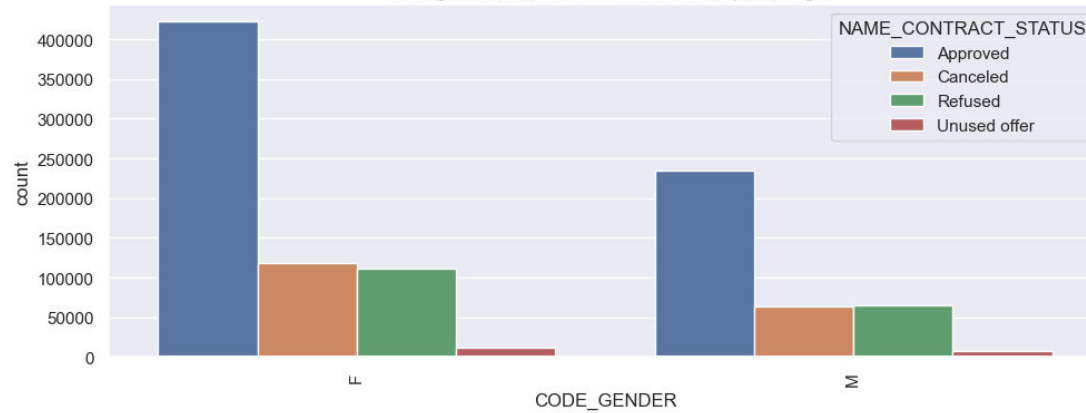
Plot for the column: CODE\_GENDER on overall data



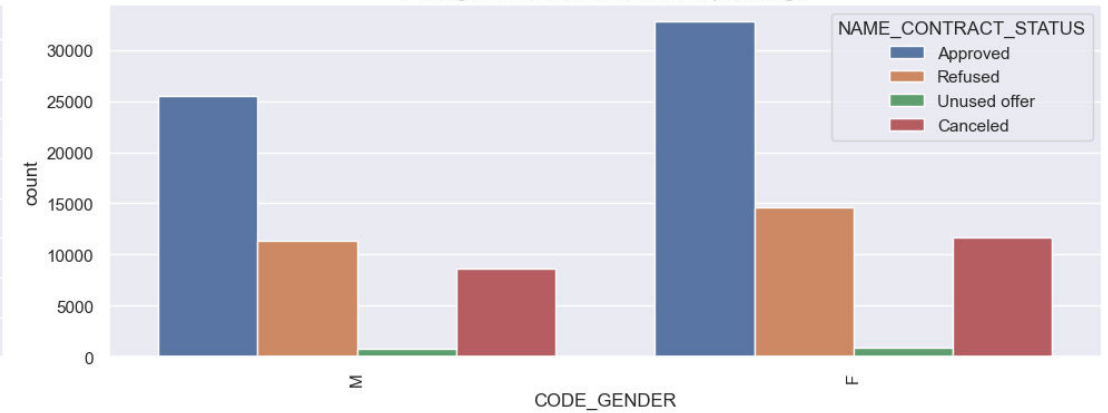
Plotting data for target in terms of total count



Plotting data for Non-Defaulter in terms of percentage



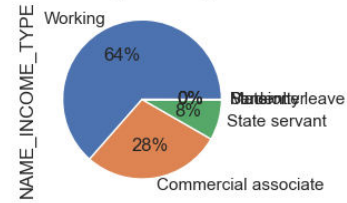
Plotting data for Defaulter in terms of percentage



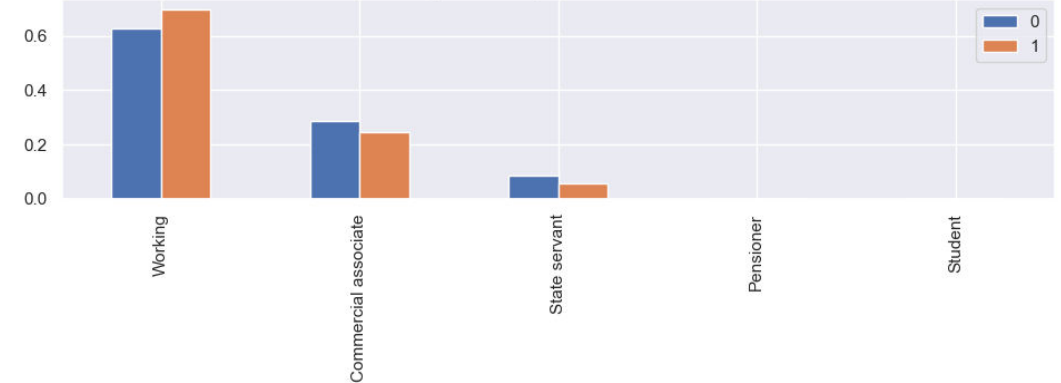


# Name Income Type Name Contract Status

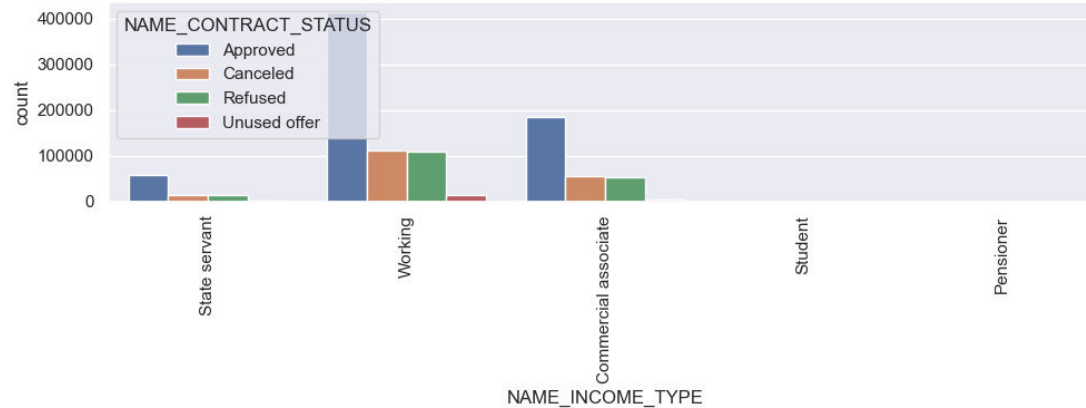
Plot for the column: NAME\_INCOME\_TYPE on overall data



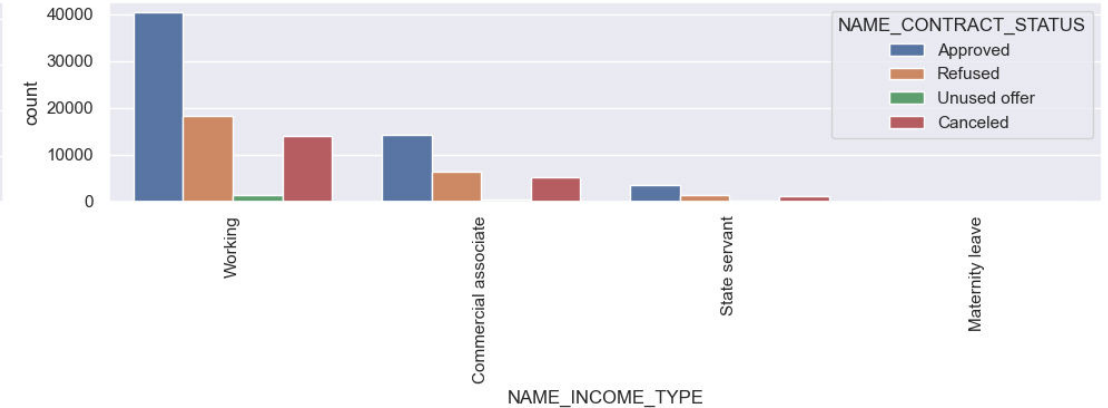
Plotting data for target in terms of total count



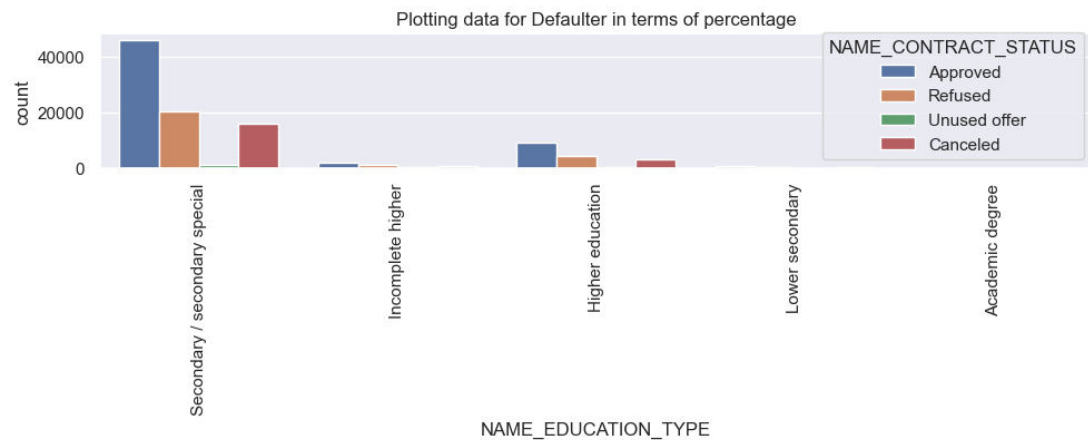
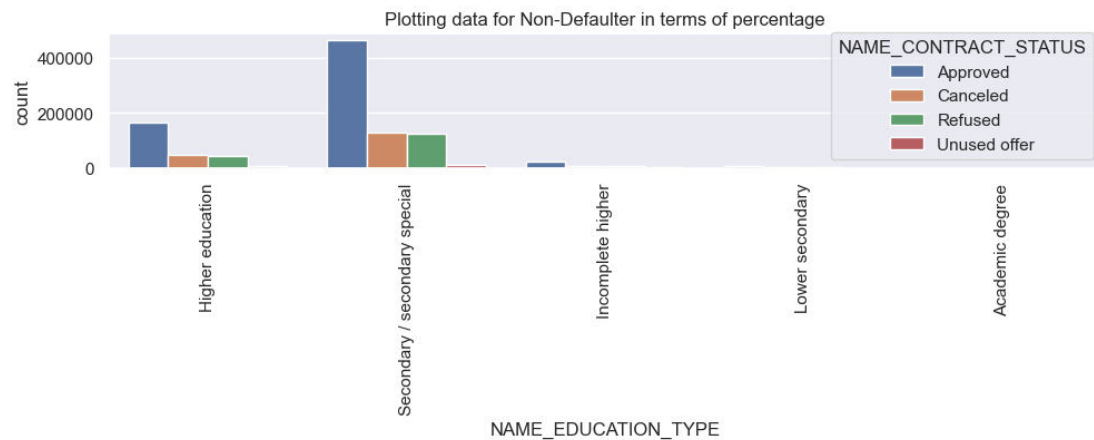
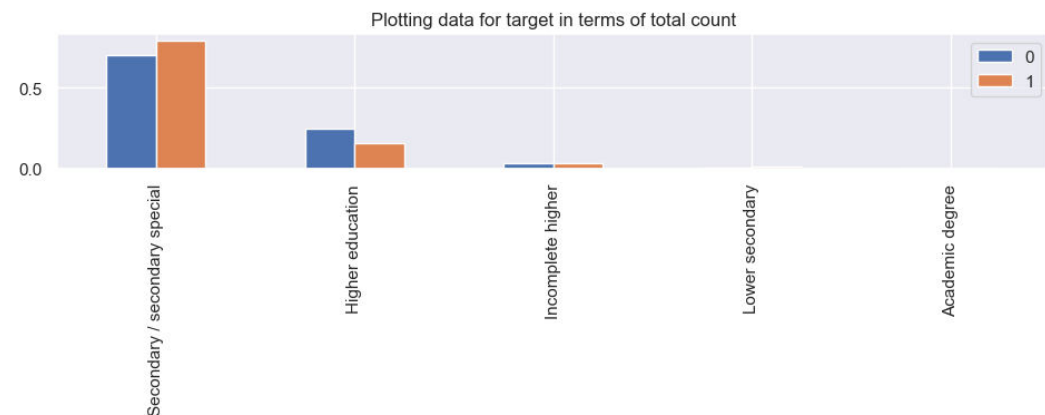
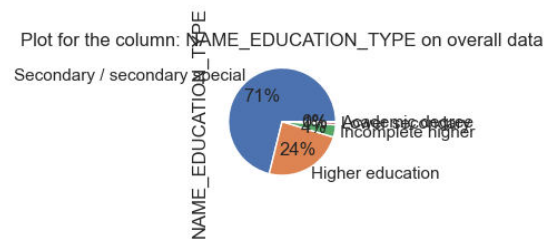
Plotting data for Non-Defaulter in terms of percentage



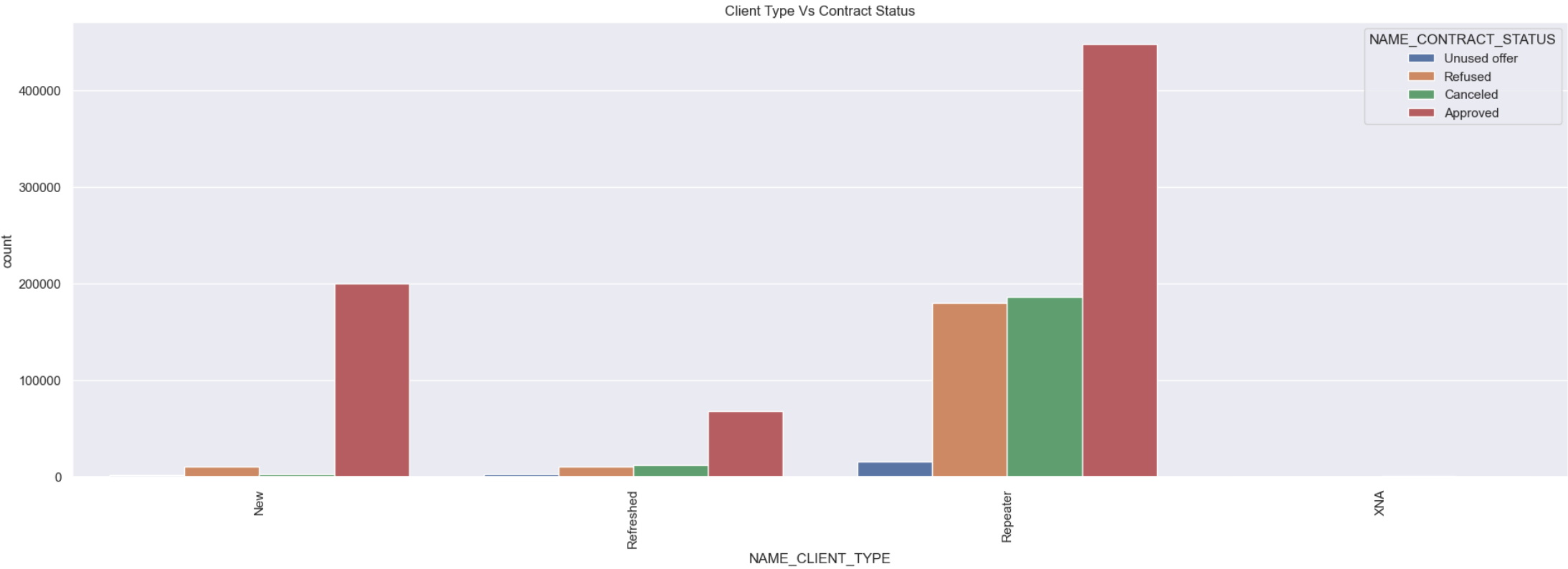
Plotting data for Defaulter in terms of percentage



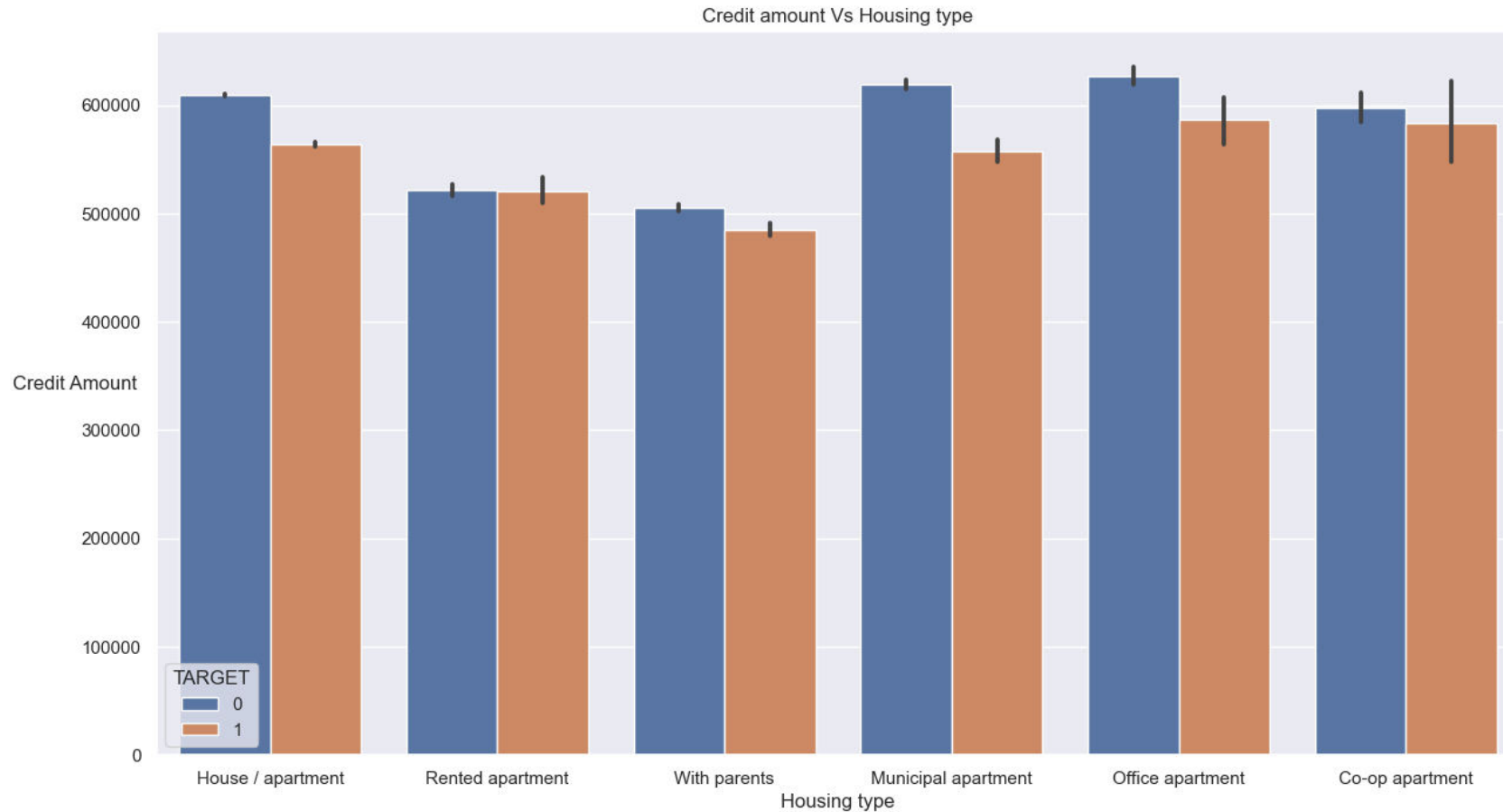
# Name Education Type Name Contract Status



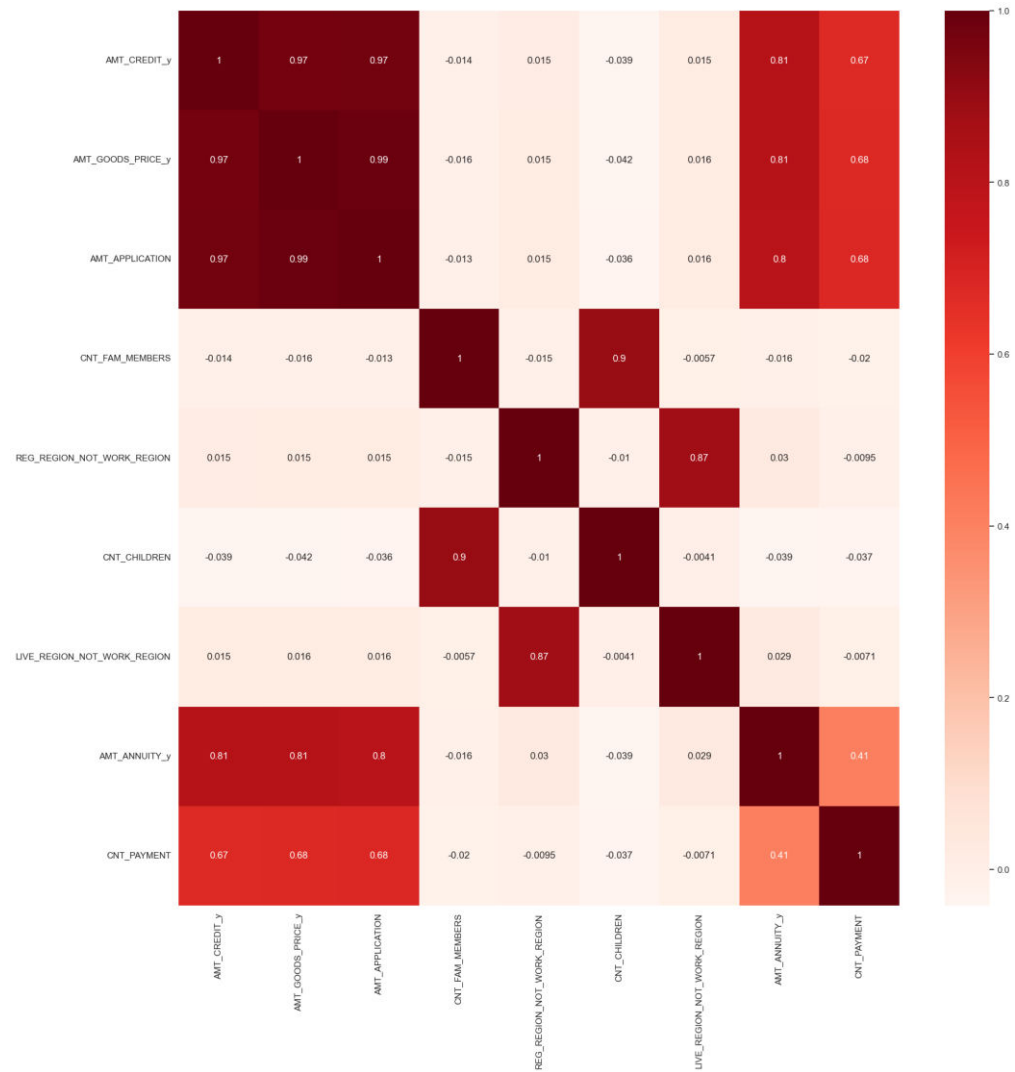
# Name Client Type Name Contract Status



# Bivariate Credit Amount Housing type



# Correlation after merging both datasets



# Recommendation & Risks

- Amount Annuity has a high positive correlation with Credit Amount, Goods Price, and Amount application
- CNT Family members and CNT children are highly correlated, more children means a big family
- Age and Annuity amount is highly correlated.

Thank  
You