Credit EDA Case Study

Preeti Kumari

Objective

- Credit risk analysis will help the company to make a decision for loan approval based on the applicant's profile which controls loss of business to the company and avoid financial loss for the company.

Steps

▶ Data understanding and reading data

▶ Check for data quality issues- missing values, incorrect data types, outliers and binning

▶ Check for data imbalance and univariate, segmented univariate and Bivariate analysis and correlation

▶ Merging of application data with previous application data

▶ Data analysis by univariate, segmented and Bivariate analysis

▶ Recommendations and Risks

Univariate Analysis of the Categorical data : CODE GENDER

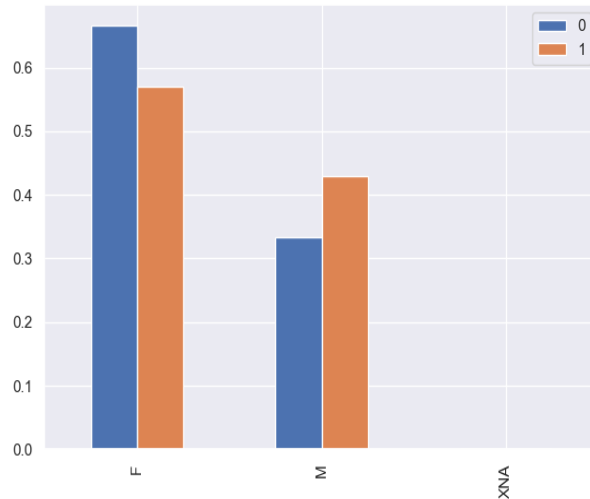
Plotting data for the column: CODE_GENDER
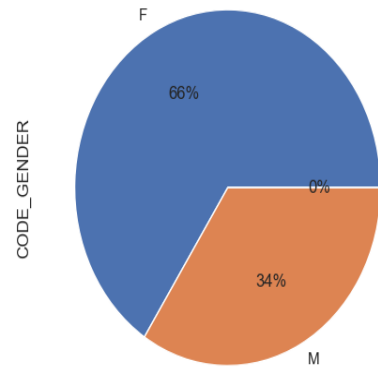
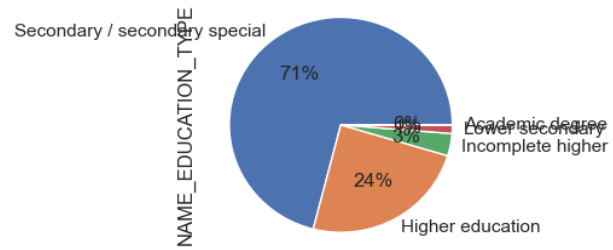Plotting data for target in terms of total count
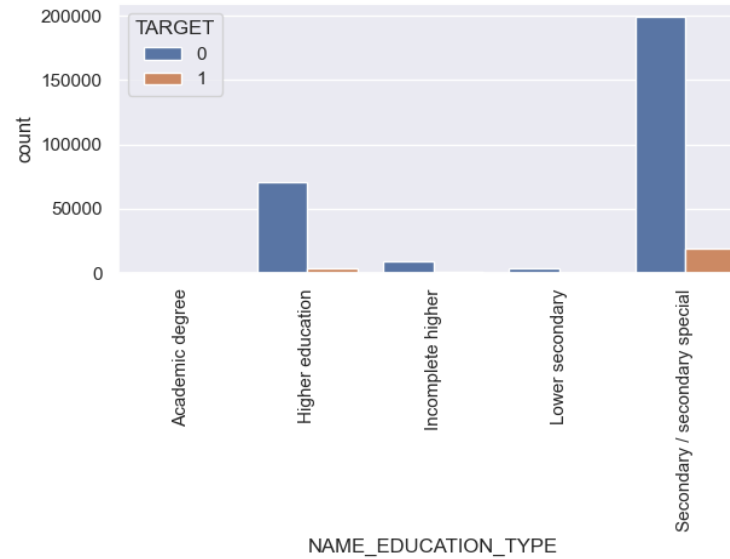

Plotting data for target in terms of percentage

CODE GENDER: Less number of males take loan but the defaulters are higher in case of males.

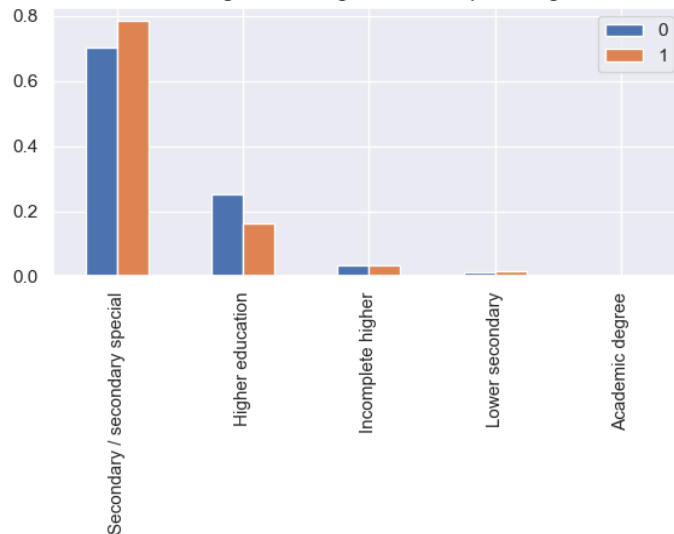# Univariate Analysis of the Categorical data :EDUCATION TYPE

Plotting data for the column: NAME_EDUCATION_TYPE

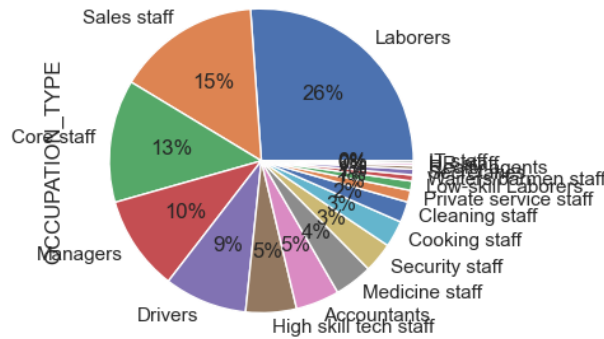Plotting data for target in terms of total count

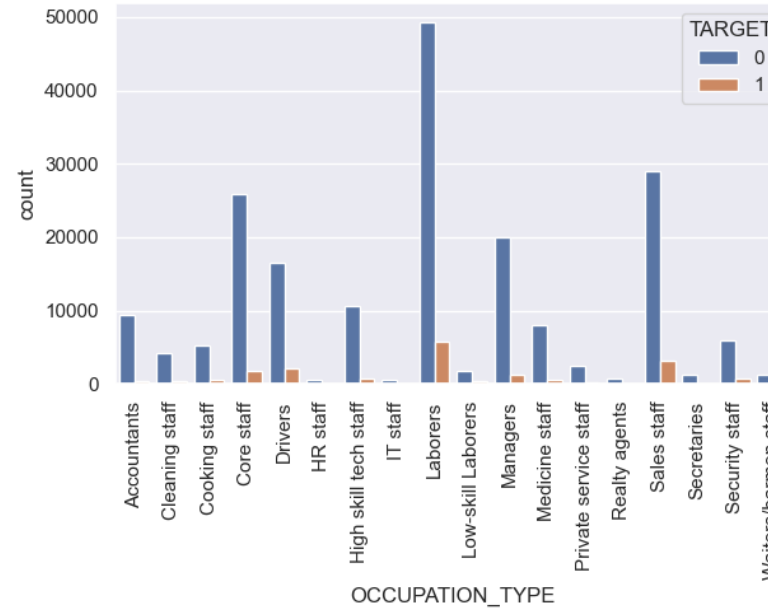
Plotting data for target in terms of percentage

EDUCATION TYPE: Most client take loan for secondary education followed by higher education. But the default rate in secondary education is much high and for higher education is much low.

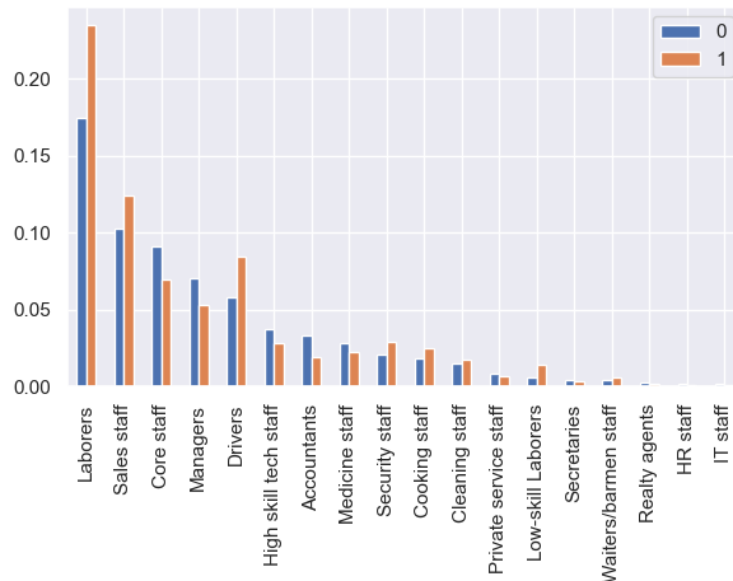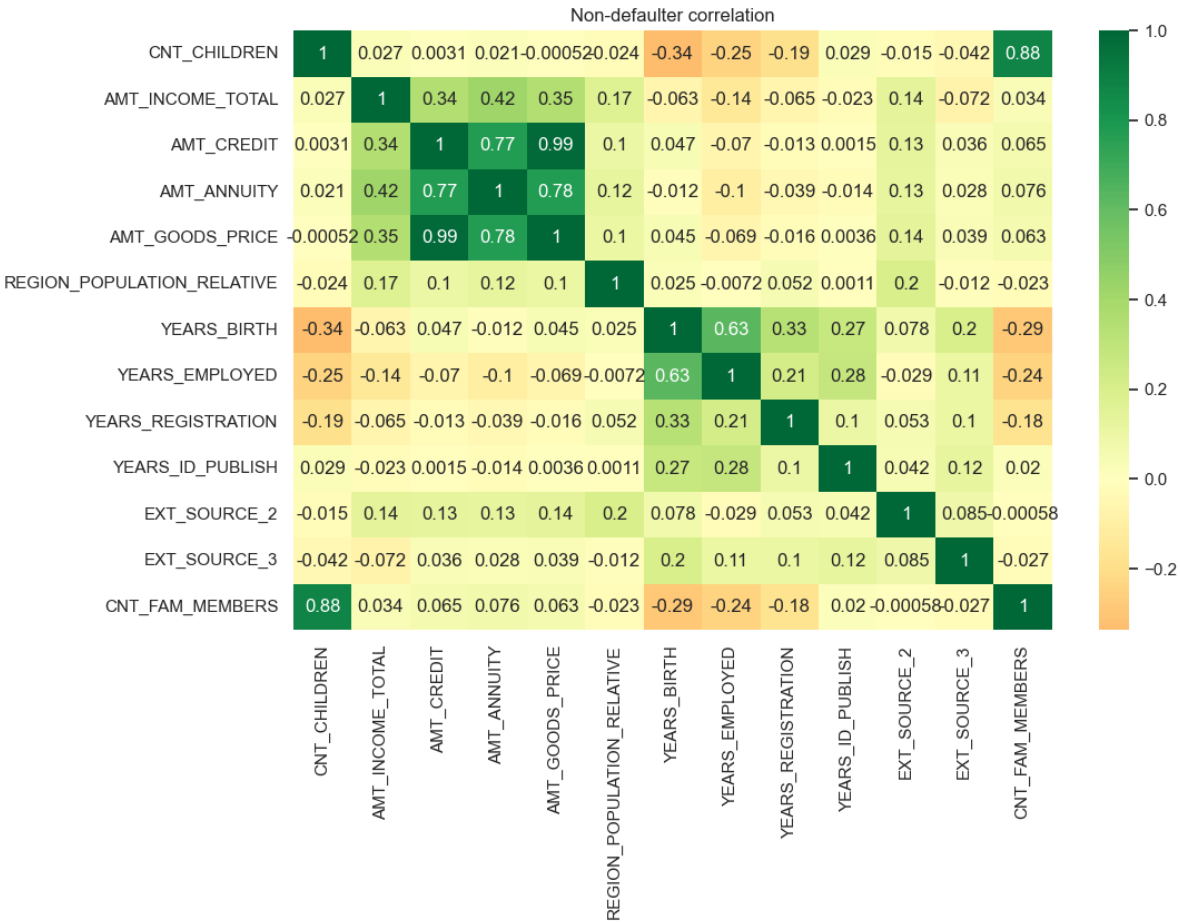# Univariate Analysis of the Categorical data : OCCUPATION TYPE

OCCUPATION TYPE: Laborers and different categories of staffs mostly take the loan, but the managers and the high skilled tech staffs are most reliable.
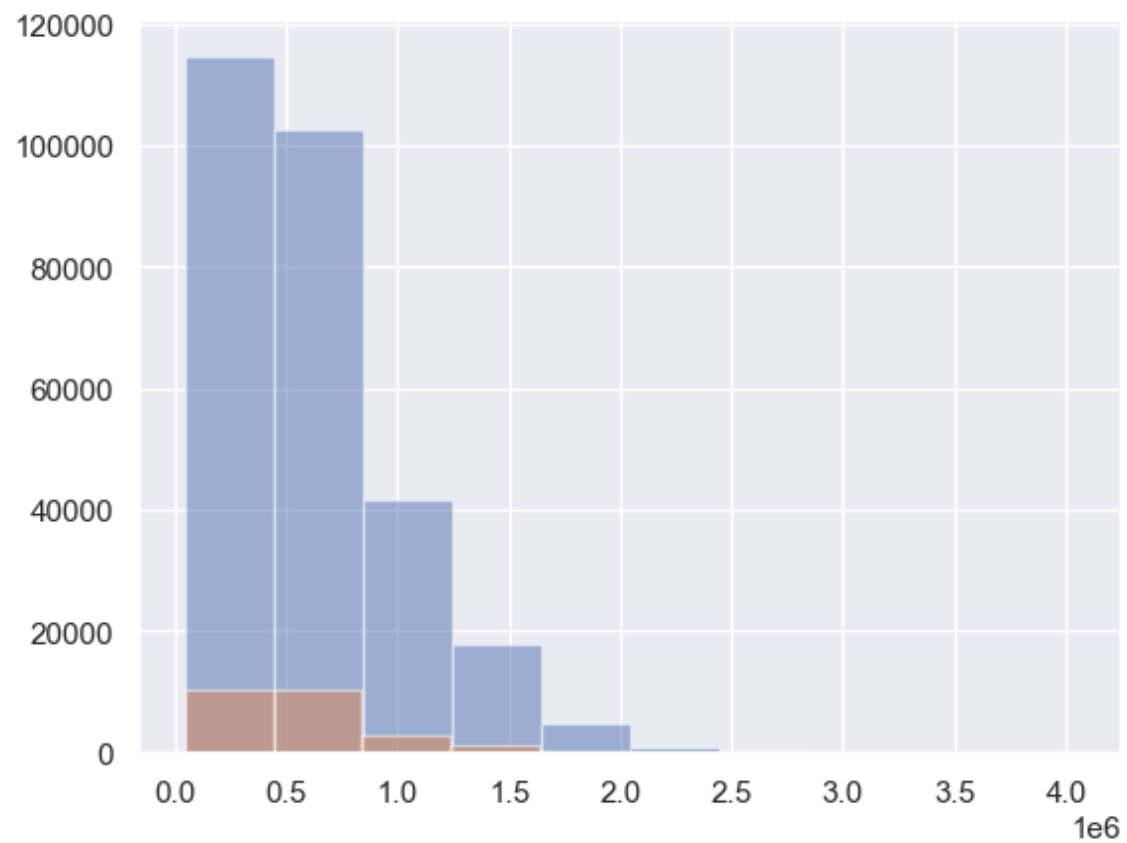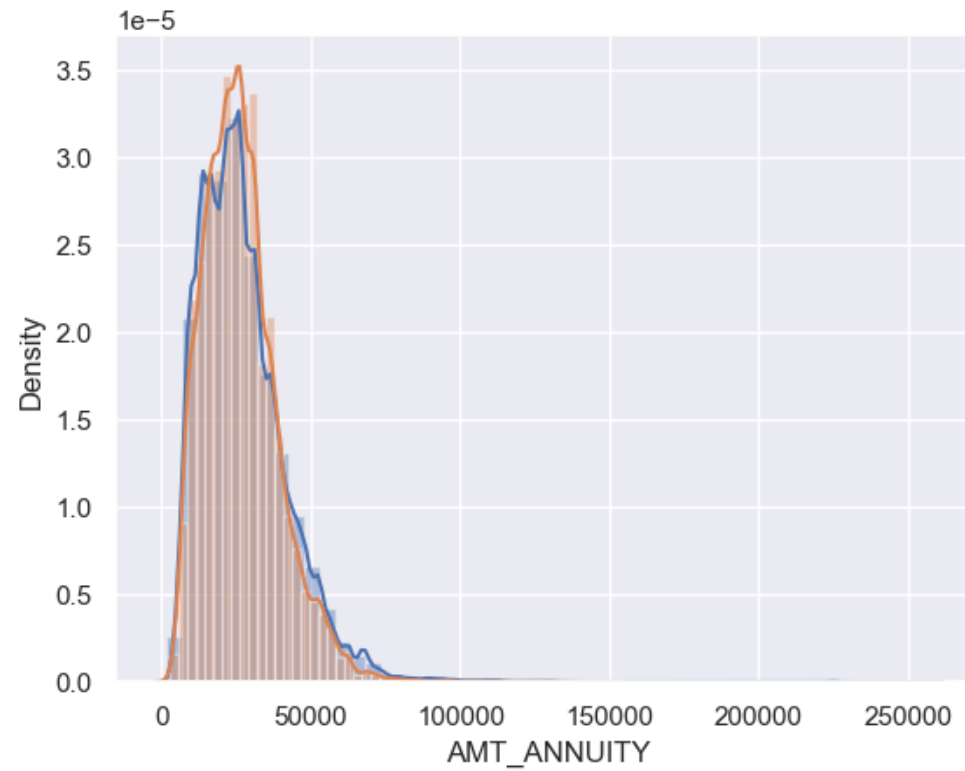
Univariate Analysis of Numerical columns
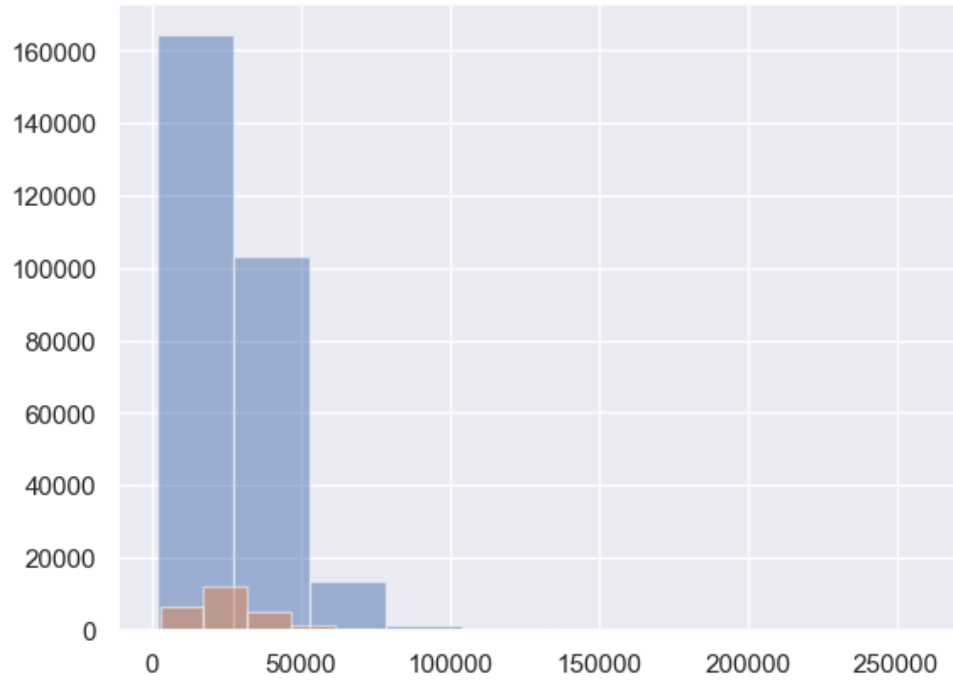
Non-defaulter correlation
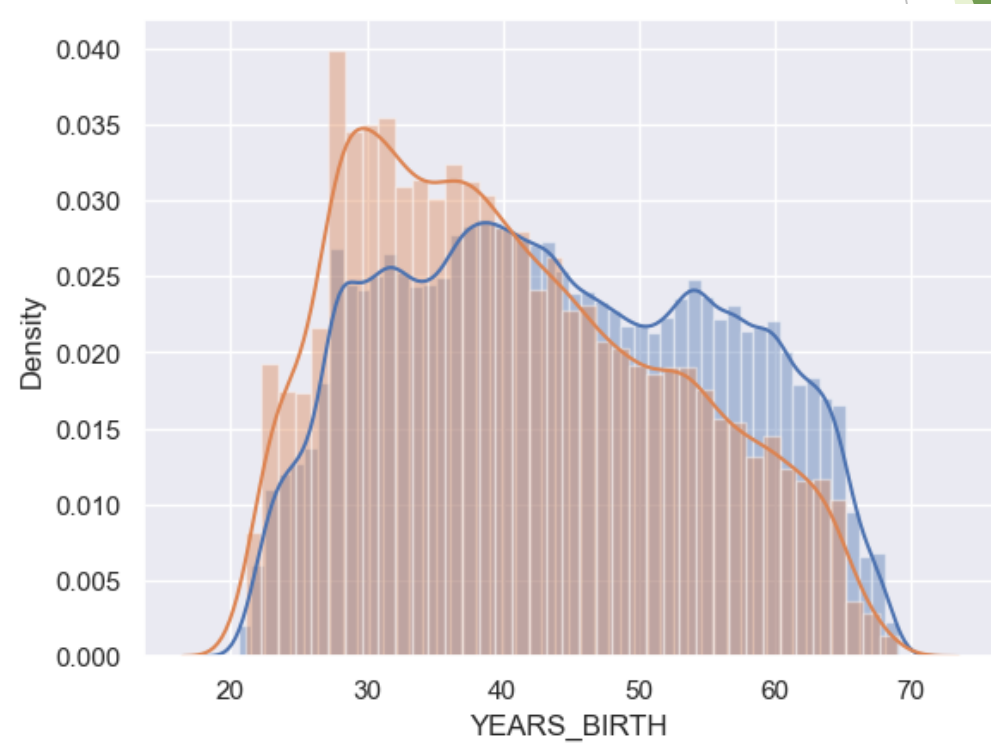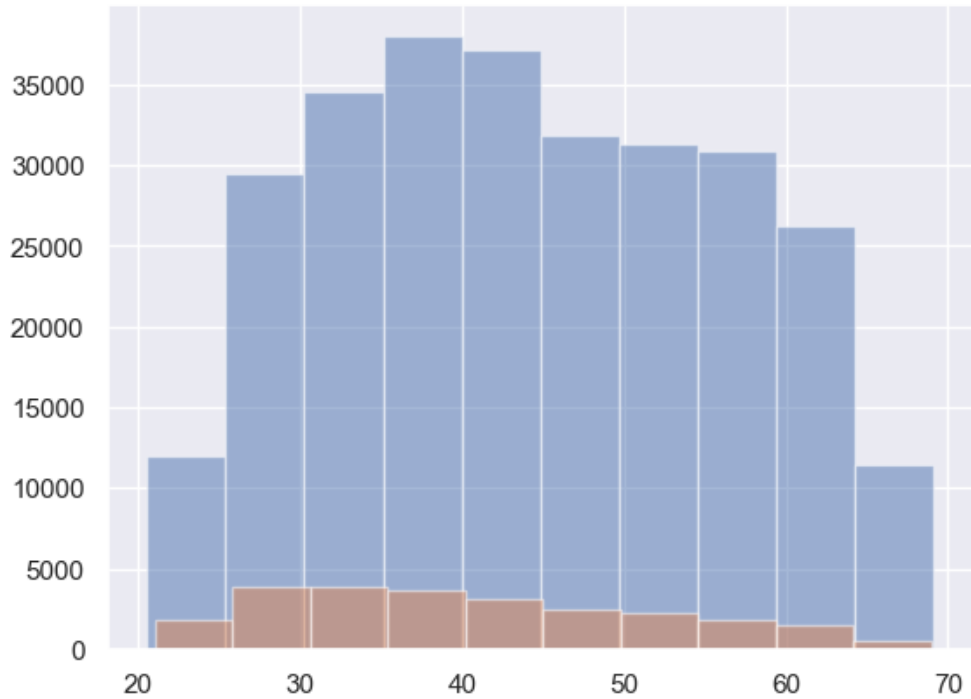
Defaulter correlation

AMT CREDIT

AMT ANNUITY
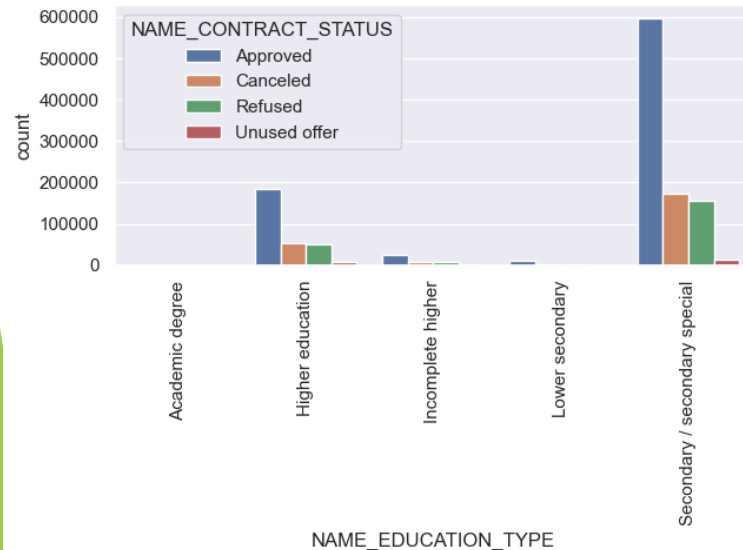
YEARS_BIRTH

Bivariate Analysis : NAME_EDUCATION_TYPE

Observation: People tend to make more loan for 'Secondary special' and their loan is also approved.
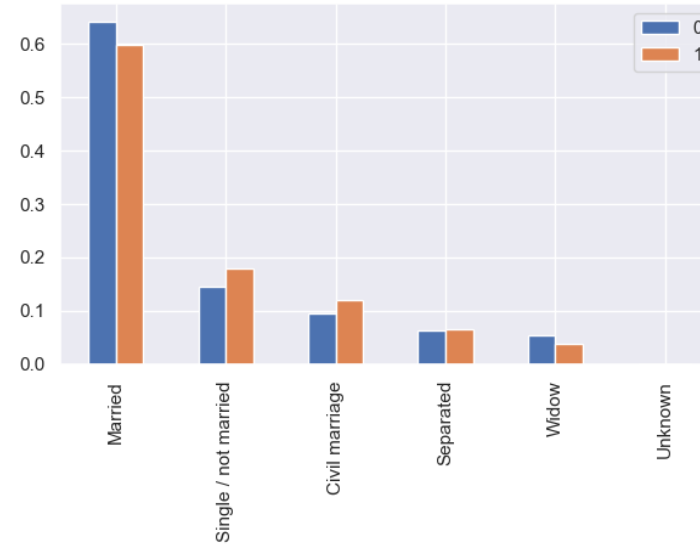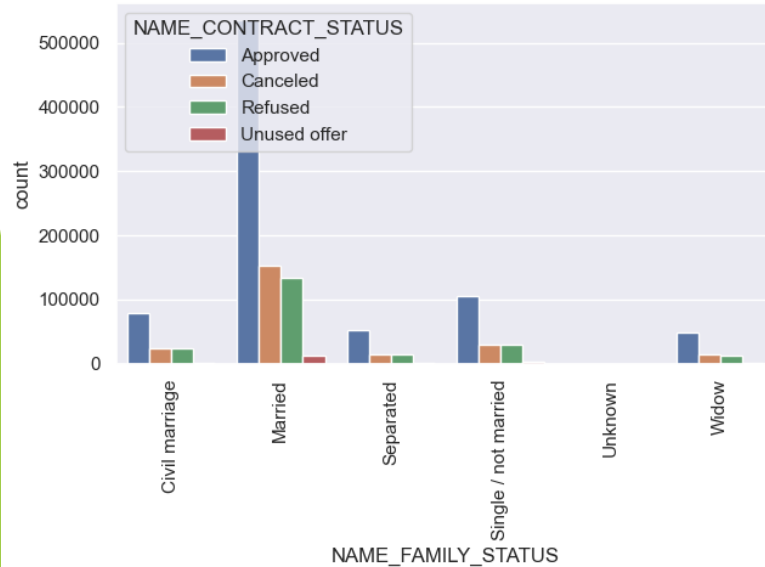
Bivariate Analysis : NAME_FAMILY_STATUS
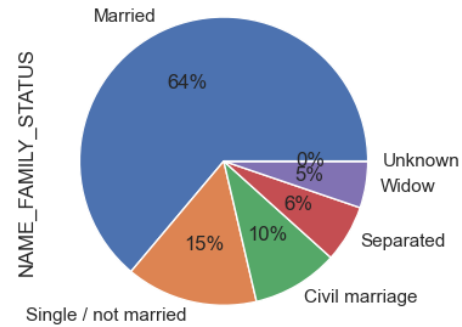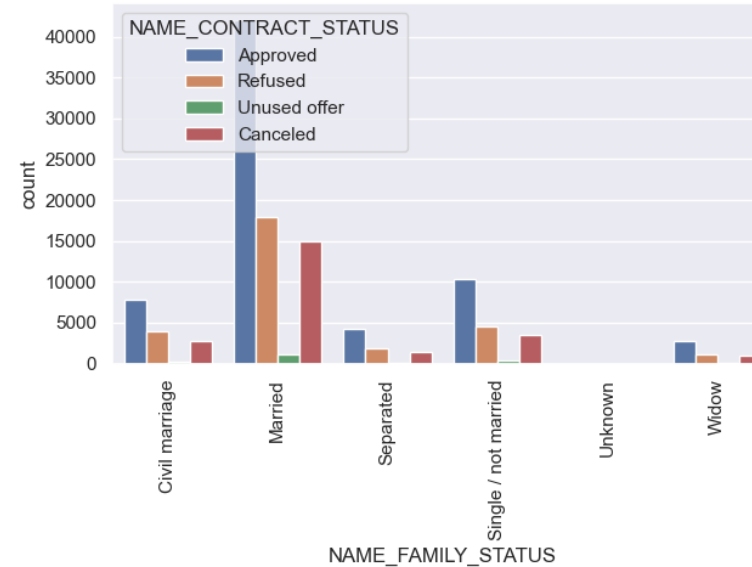
Observation: You can see, there is a clear difference for the categories for "Approved, Refused, Unused and Cancelled" for the category: Married. Married people tends to pay loan on time than Singles.

Bivariate Analysis : NAME_HOUSING_TYPE
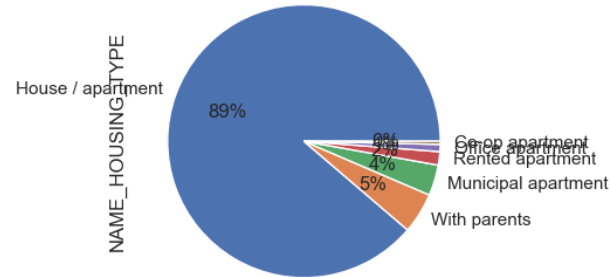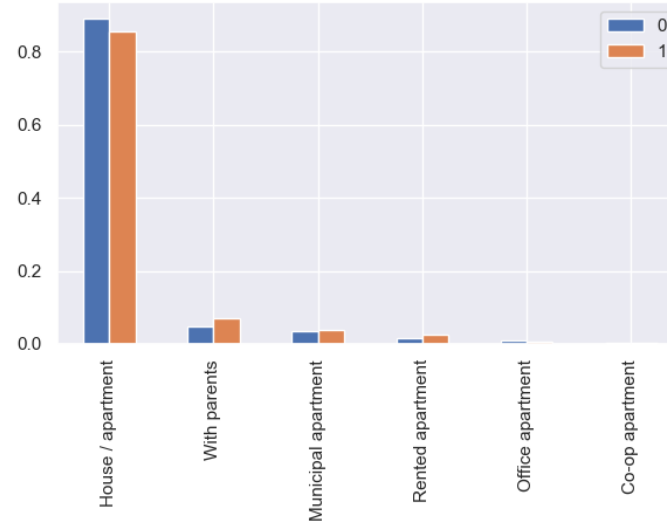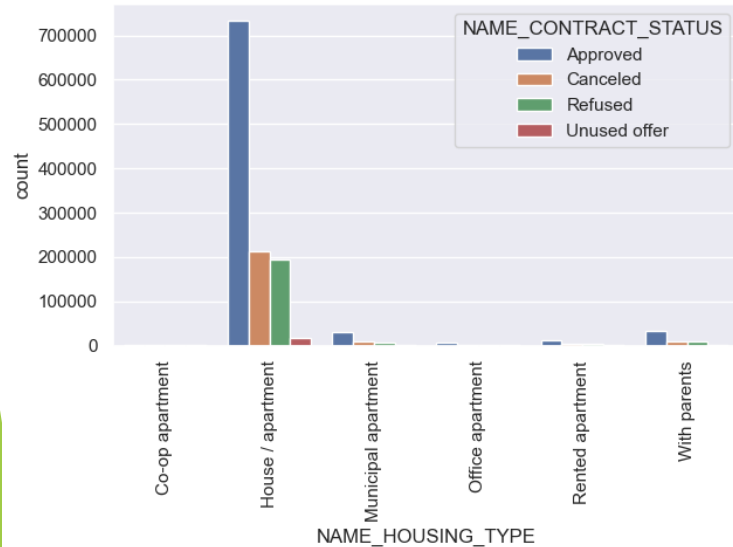

Observation: You can see, there is a clear difference for the categories for "Approved, Refused, Unused and Cancelled" for the category: House/apartment.

Bivariate Analysis : ORGANIZATION_TYPE
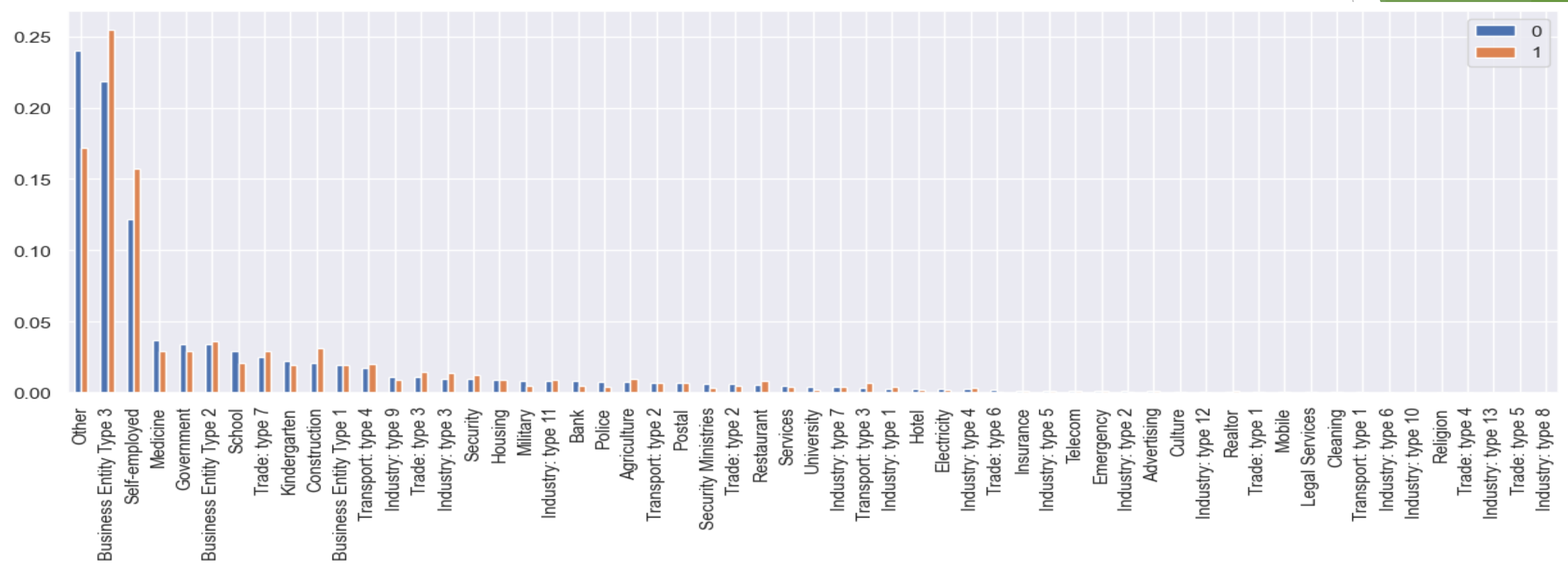
This columns seems to be most important. Business Entity Type 3 and Self-employed tends to be the maximum defaulter. The univariate analysis of this dataset was more fruitful than bivariate analysis.

Recommended group where loan can be credited and less chance to be defaulter

- Clients who are working as a state servant.

- Old people of any income group.

- Client with high income category.

- Female clients.

- Clients with higher education.

- Any client who's previous loan was approved.

- Refreshed client who has unused loan status previously.

Risky Groups

▶ Lower secondary educated clients are the most in number to be defaulted when their previous loans were cancelled or refused.

▶ Male clients with civil marriage.

▶ Previously refused loan status group.

▶ Business Entity Type 3 and Self-employed tends to be the maxiumum defaulter.

THANK YOU