CREDIT EDA ASSIGNMENT

Business understanding of loan approval based on applicant profile details

Problem Statement

- The aim of this analysis is to identify patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

Approach

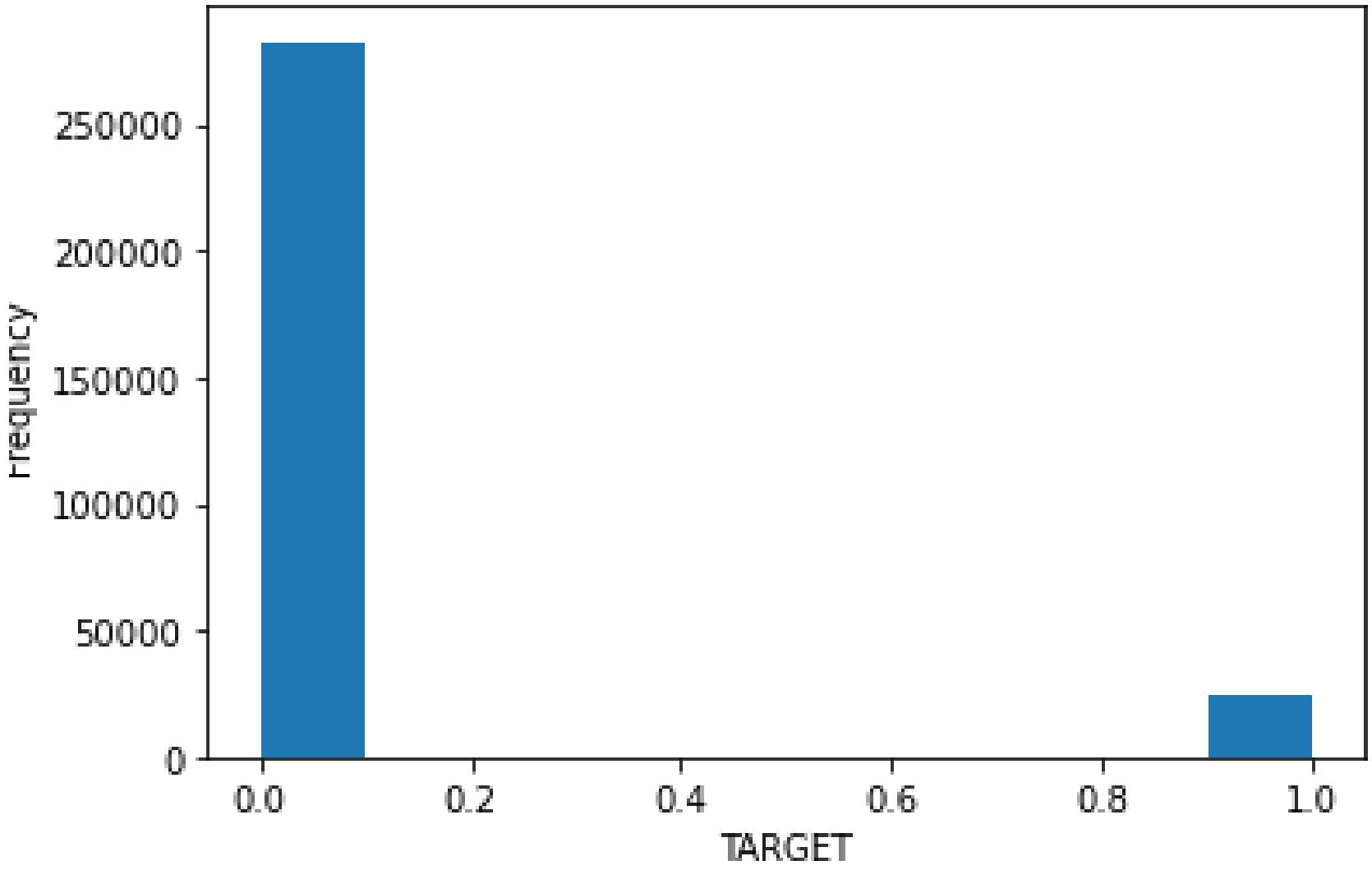
- Quick look at the data terminology and variables (Data identification)
- Loading data and understanding basic dimensions and structure
- Missing values and outliers checks
- Univariate and Bivariate analysis of both dataframes
- Final analysis after merging dataframes

Important Take aways from the EDA

- Univariate analysis of first dataframe "application_data.csv".
- Bivariate analysis
- Univariate analysis of data after diving in two sets for defaulters and non-defaulters
- Univariate & Bivariate analysis of second dataframe "previous_application.csv".
- Analysis of third dataframe merged from the first two dataframes.

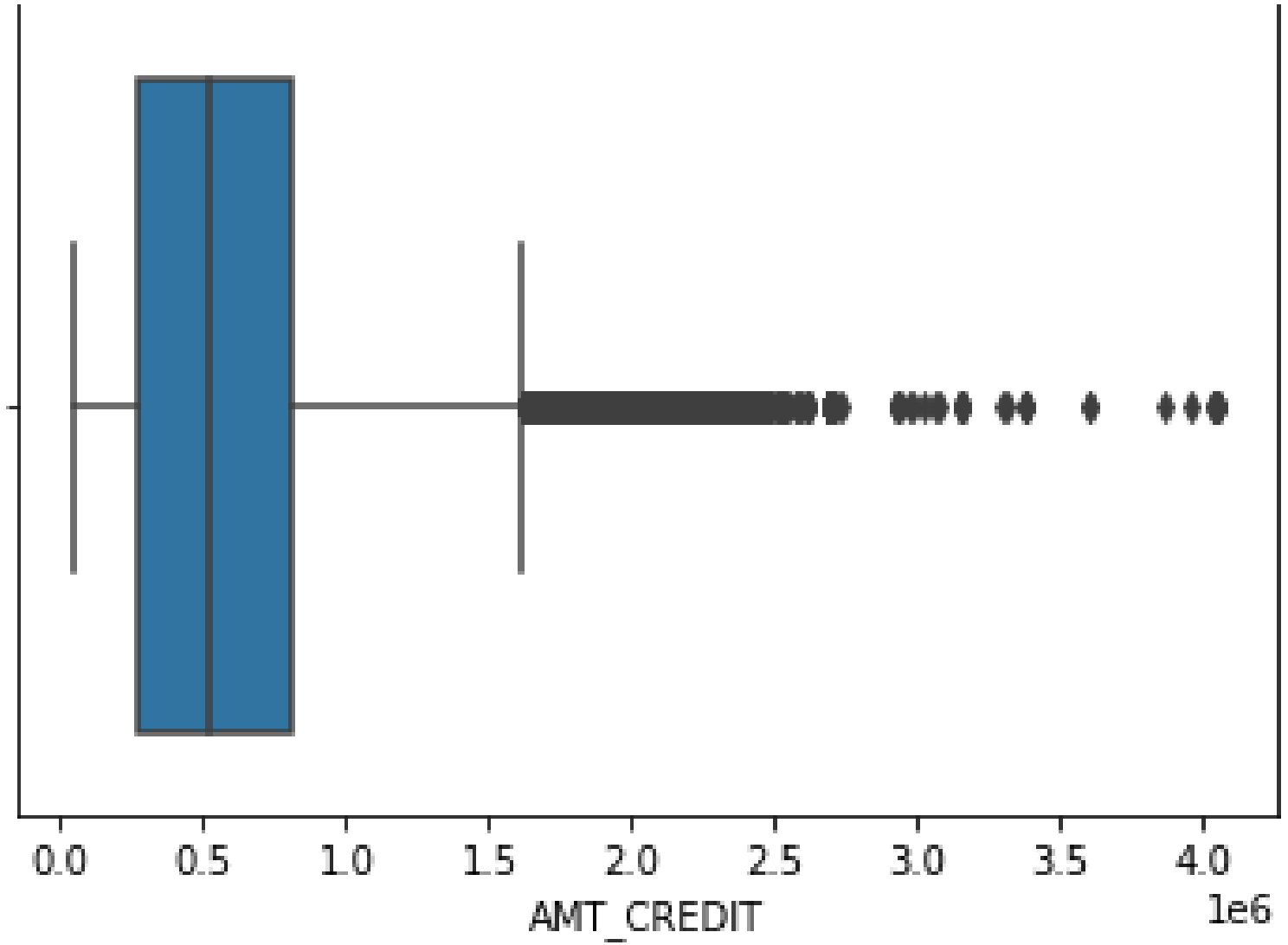
Histogram of the column TARGET

Here we see that majority of the applicants in the data are non-defaulters as the frequency of target being 0 is more



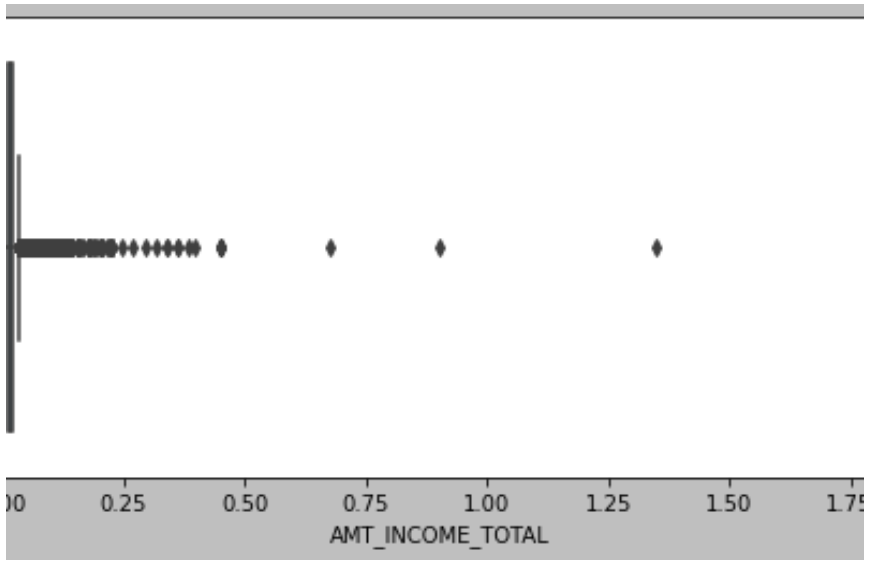
Boxplot graph for AMT_CREDIT column

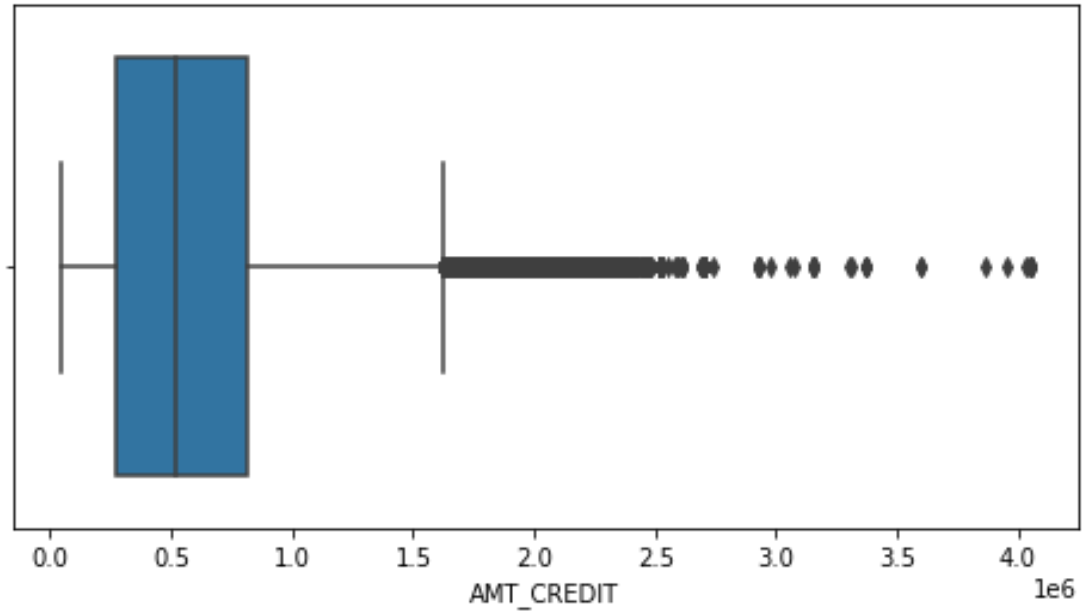
- Here we see a few outliers above the range of credit amount being 35L
- The major bulk of credit amount is given between the range of 15 - 25L



Univariate analysis after segmenting defaulters and non-defaulters

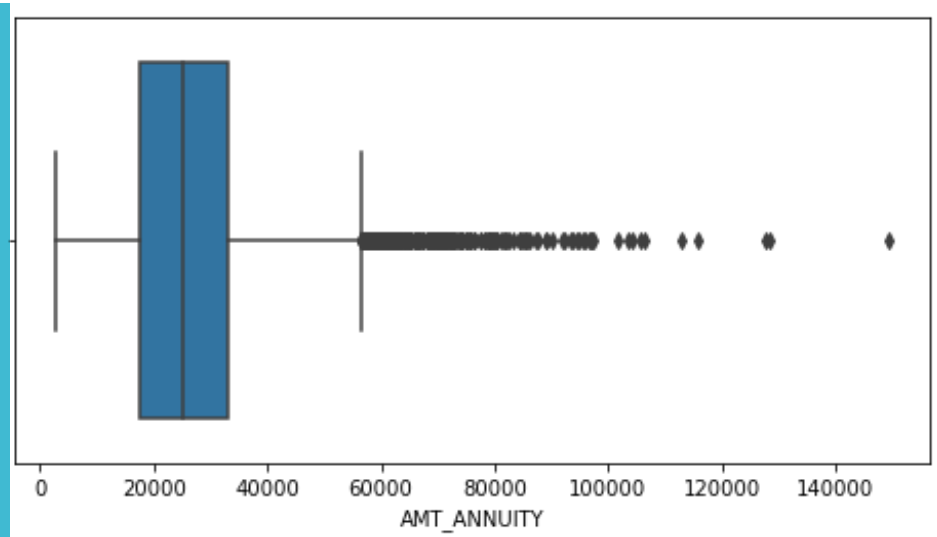
These graphs of segmented non-defaulters show a few outliers in the income total column and credit amount column.

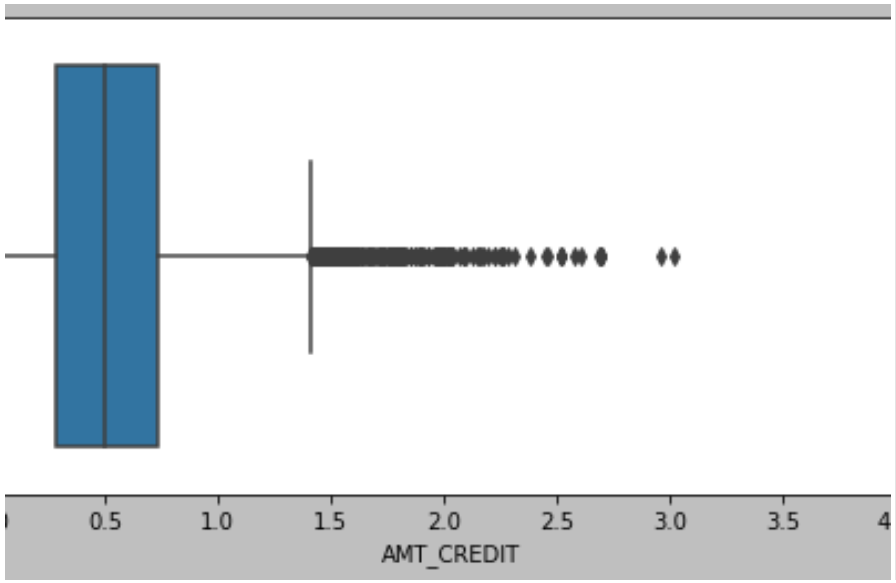




Segmented defaulters

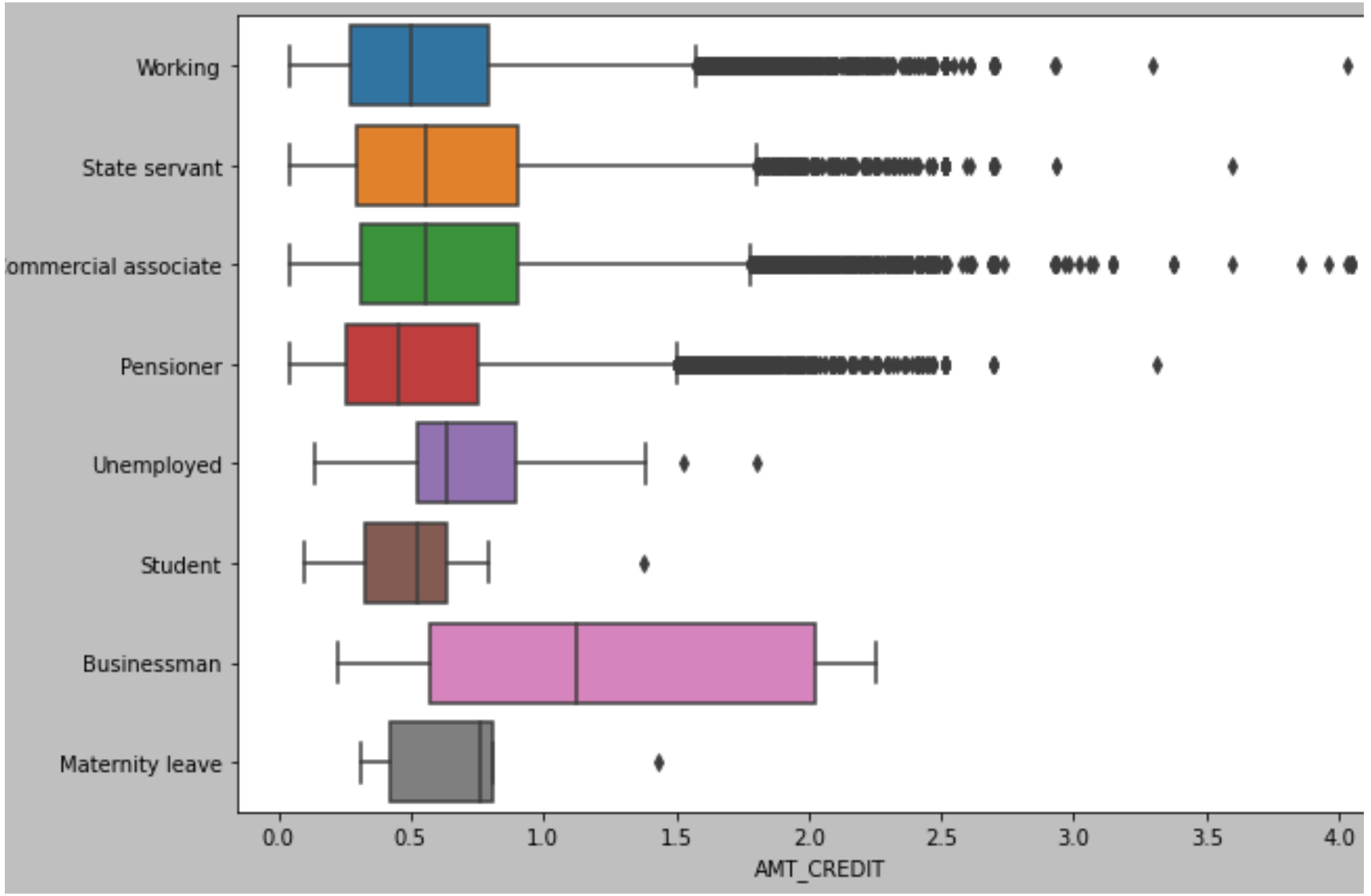
- Annuity for defaulters have a bulk of outliers with an amount above 60000 per year
- Overall credit amount being above 15L for outlier defaulters





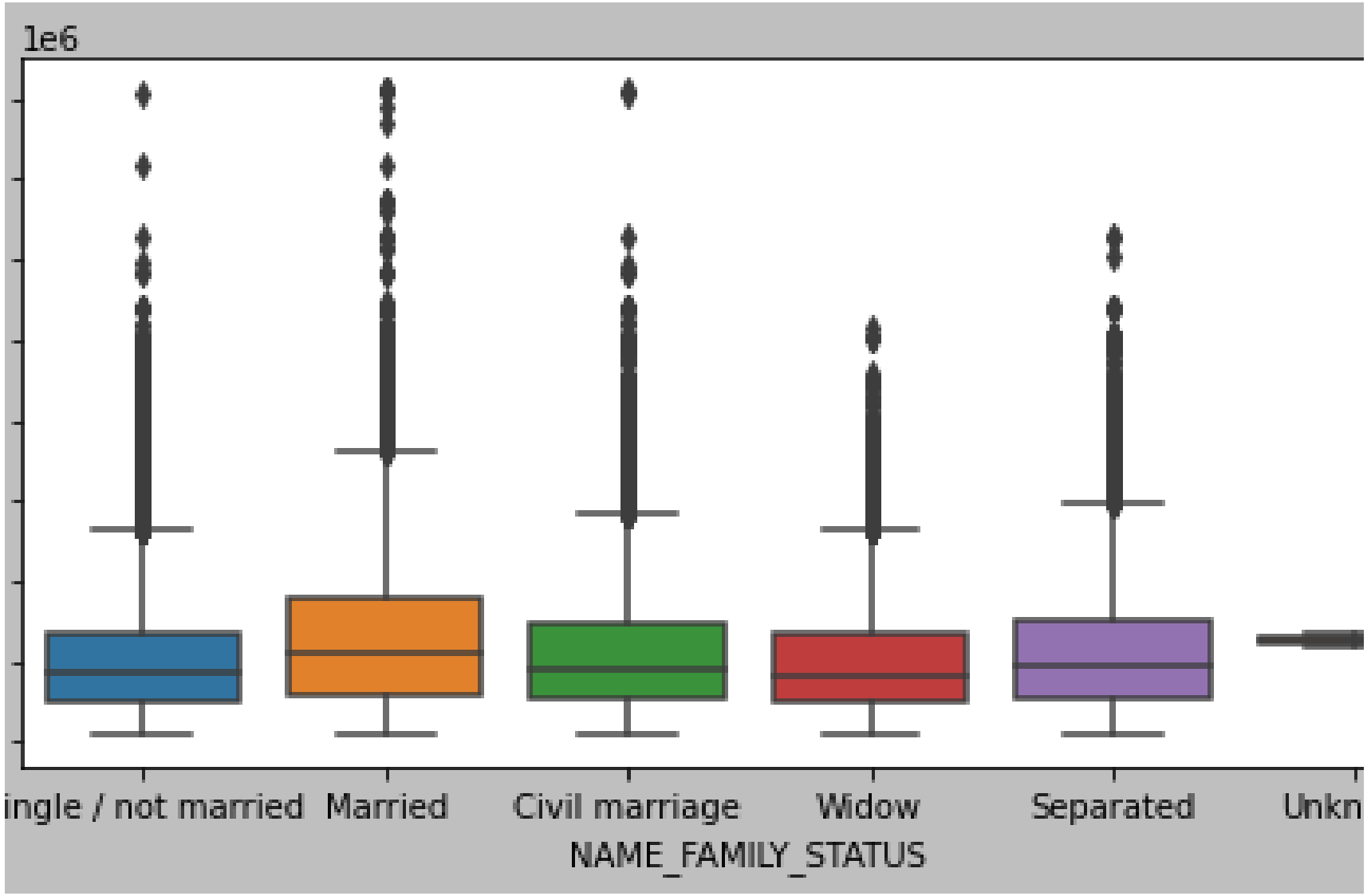
Credit data for defaulters based on their income type

Businessmen seem to have the highest credit amount bulk without any outliers and students have been offered the lowest range of credit



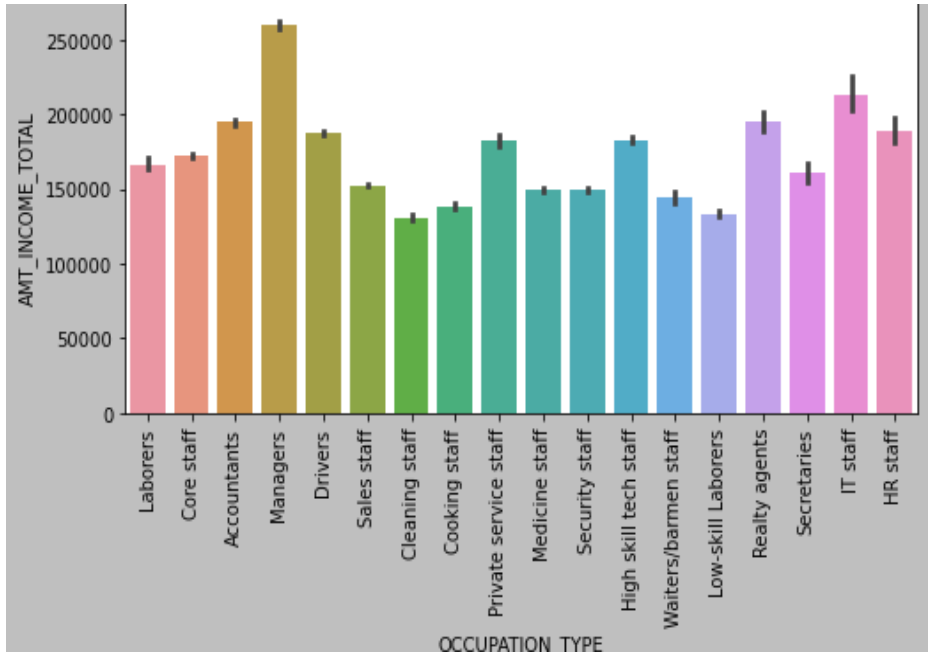
Defaulters family status vs credit

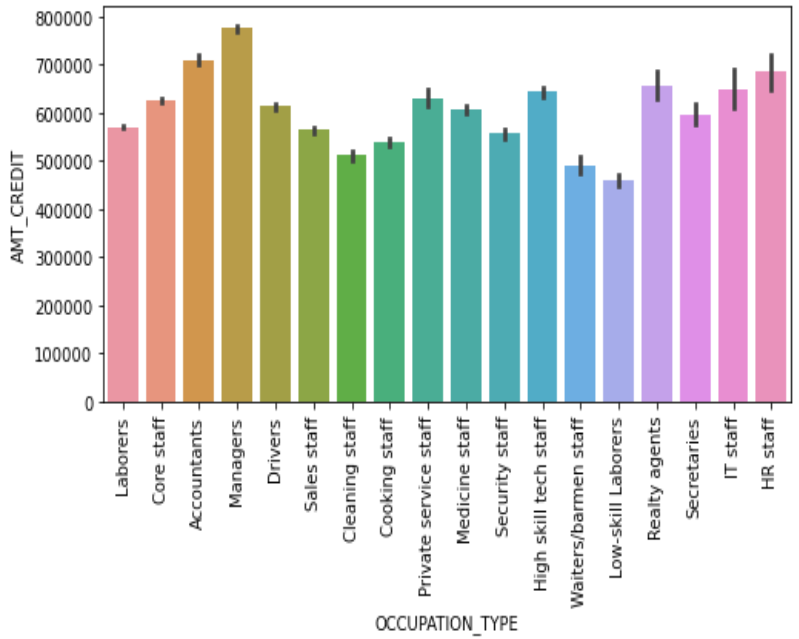
Married applicants have been issues highest credit amounts and single as well widowed the lowest in overall credit range

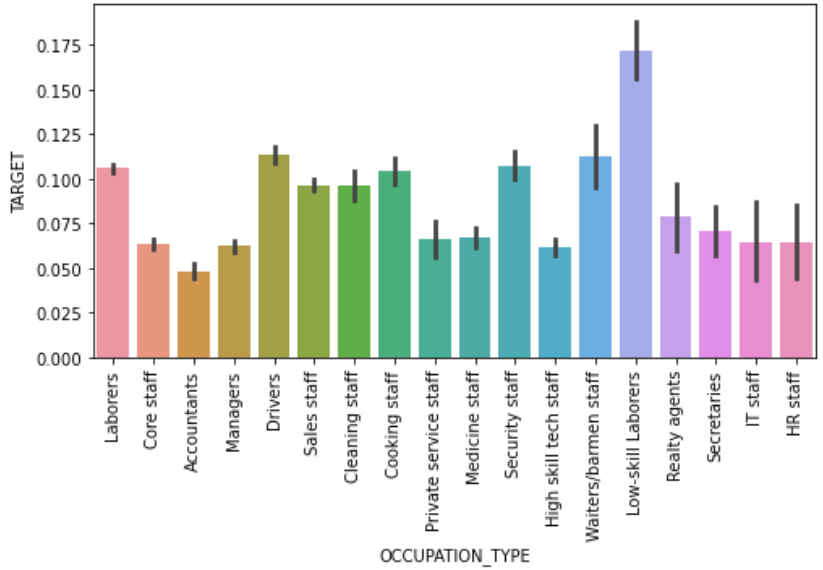


Occupation type vs various variables for defaulters

- Managers have been seen to have highest income and cleaning, cooking, security, waiters and lower skill labours have shown the lowest income
- The highest targeted income type is low skilled labour among the lower income types
- Although credit offered is not much varied from other low income types

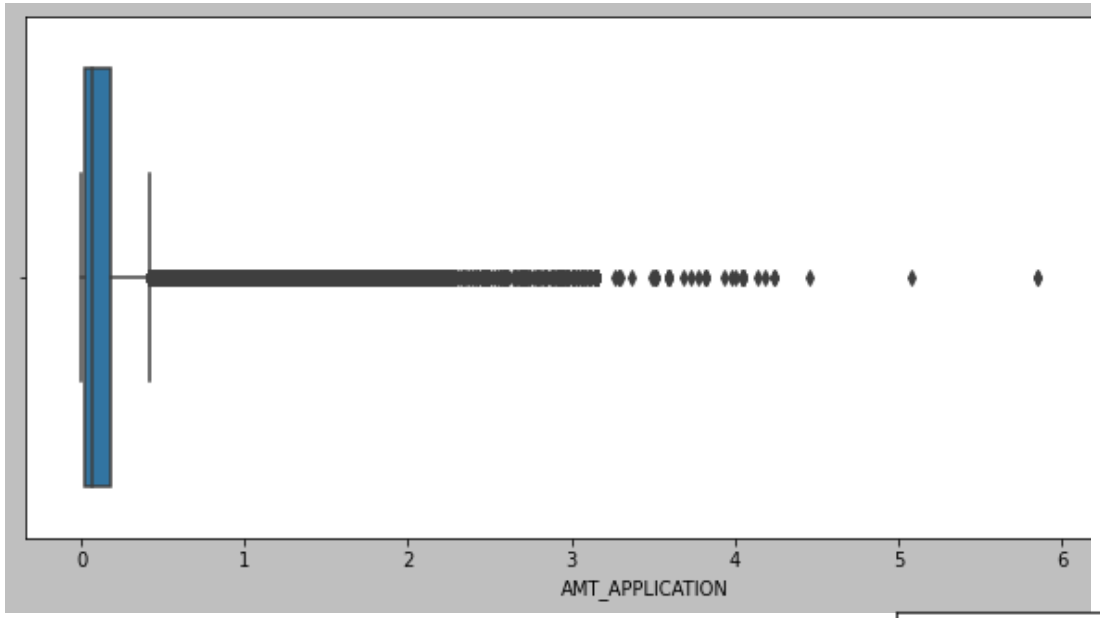


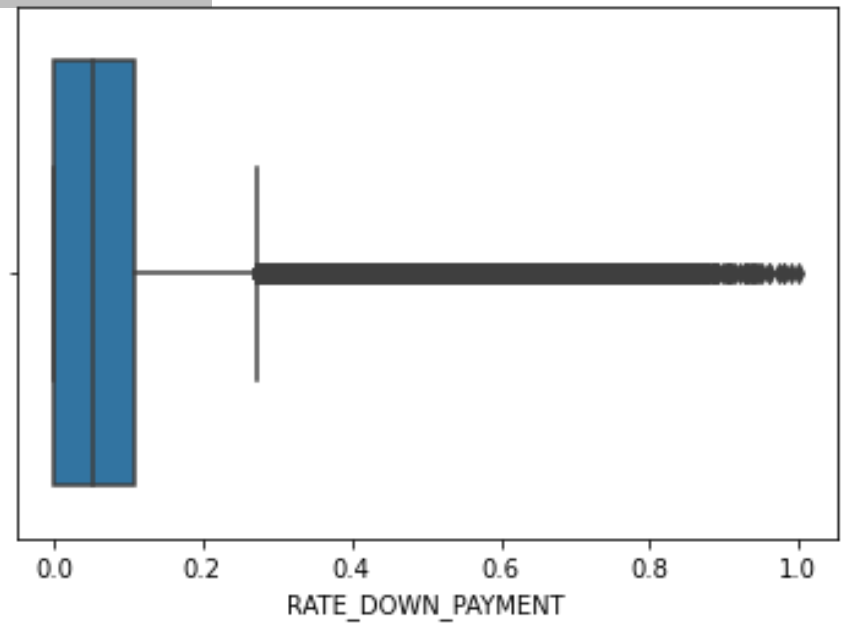




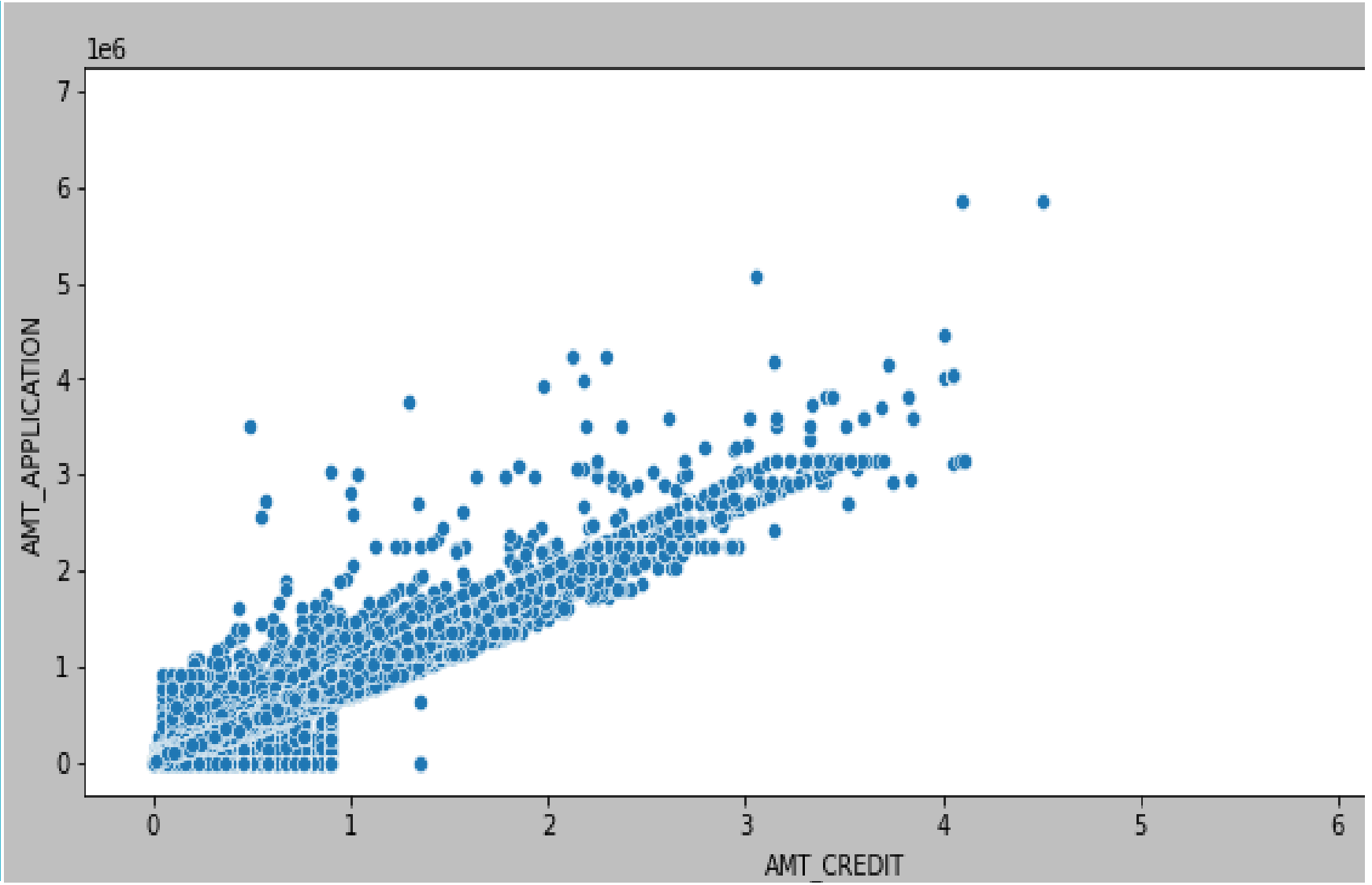
Univariate and Bivariate analysis from the data of previous applications

Rate of down payment for previous applications as per previous credit has major bulk in outliers





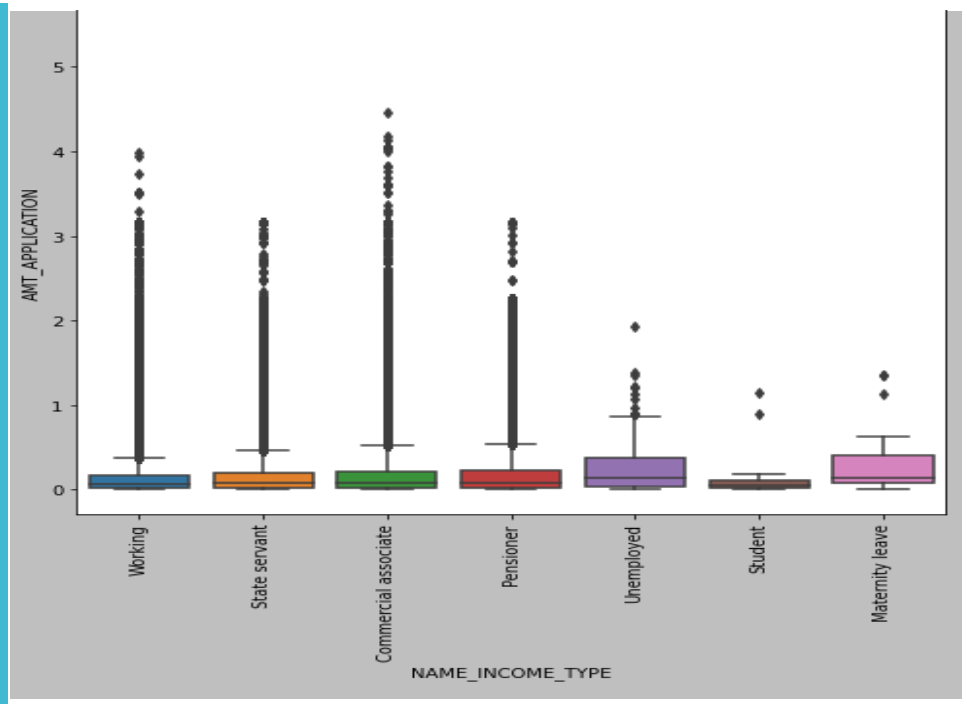
Credit amount vs application amount for previous applications has a steady increasing rate with some outliers

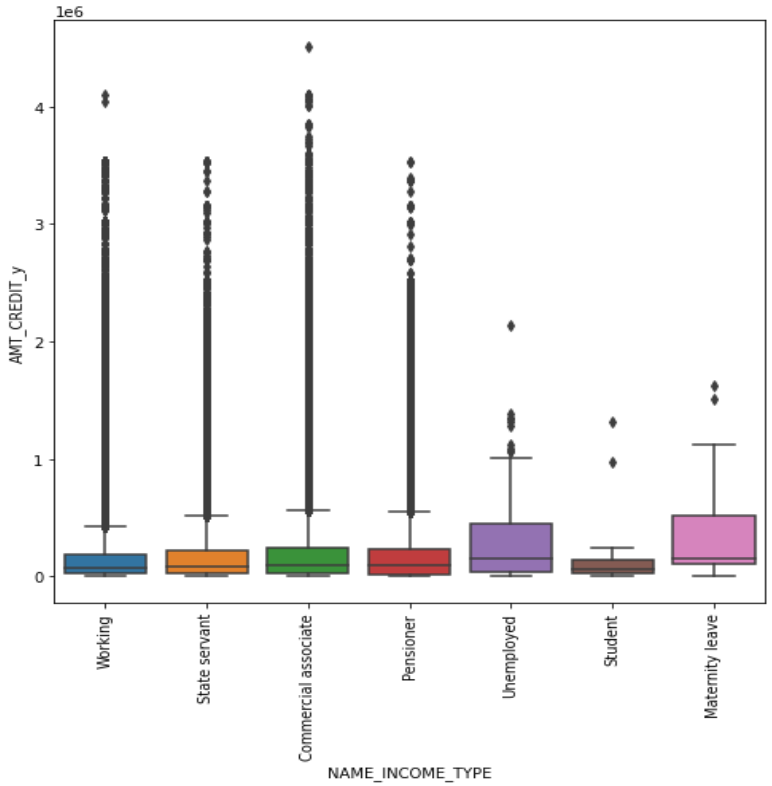


Analysis of merged data frames and their variables

Various income types vs their previous application and credit amounts

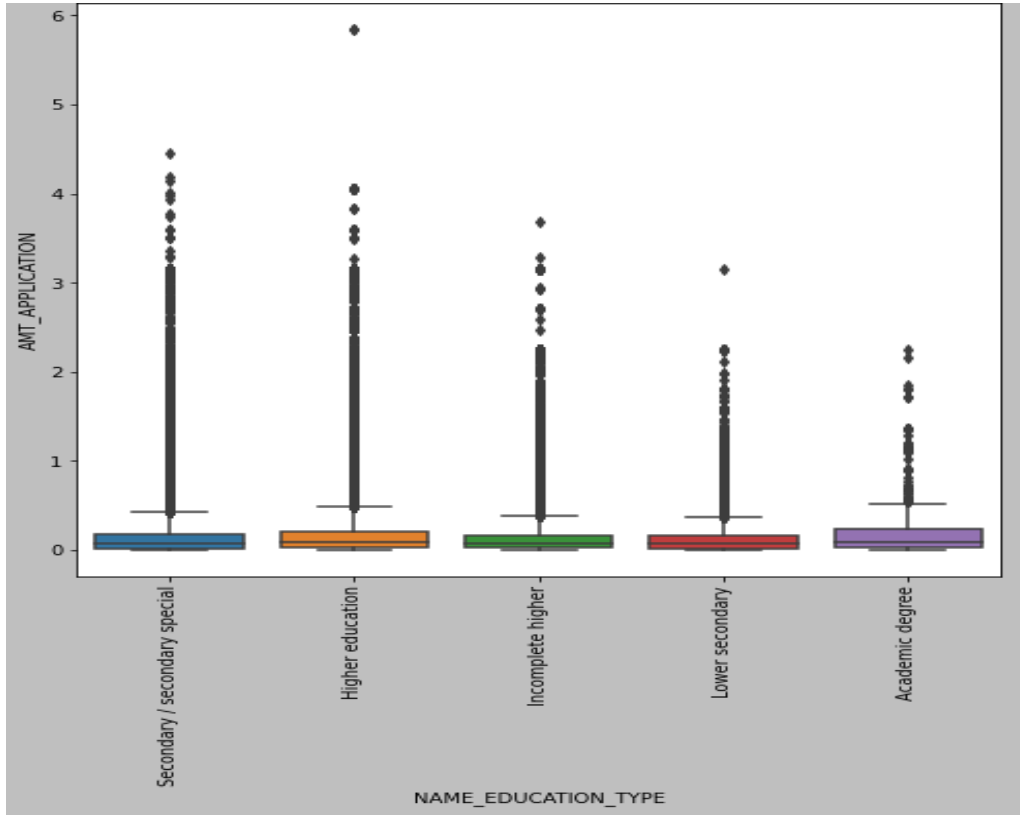
Students have lowest range of application and credit amounts from previous applications

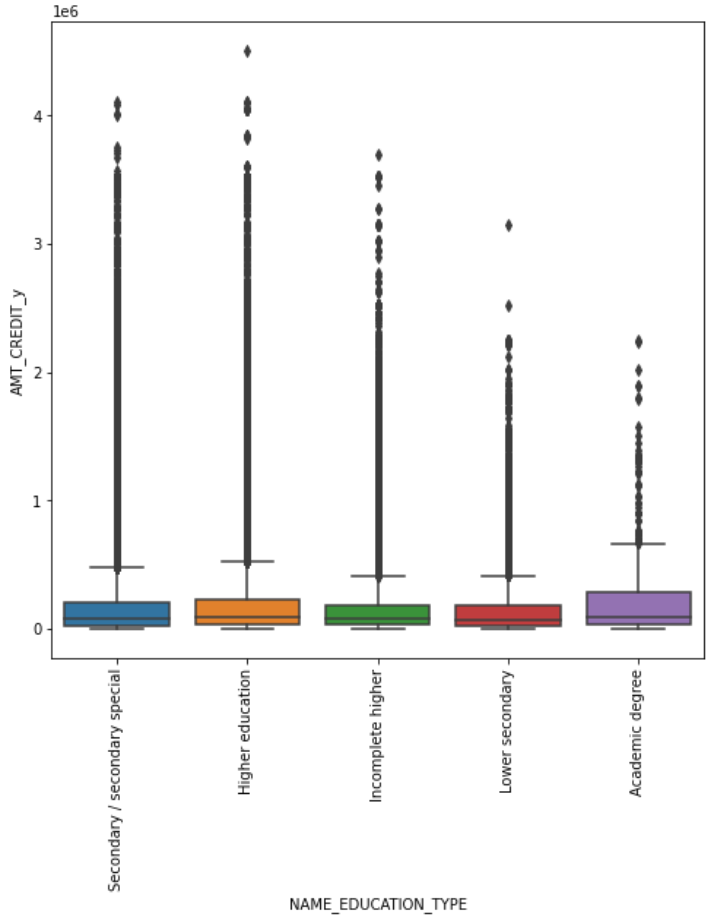




Various education types vs their previous application and credit amounts

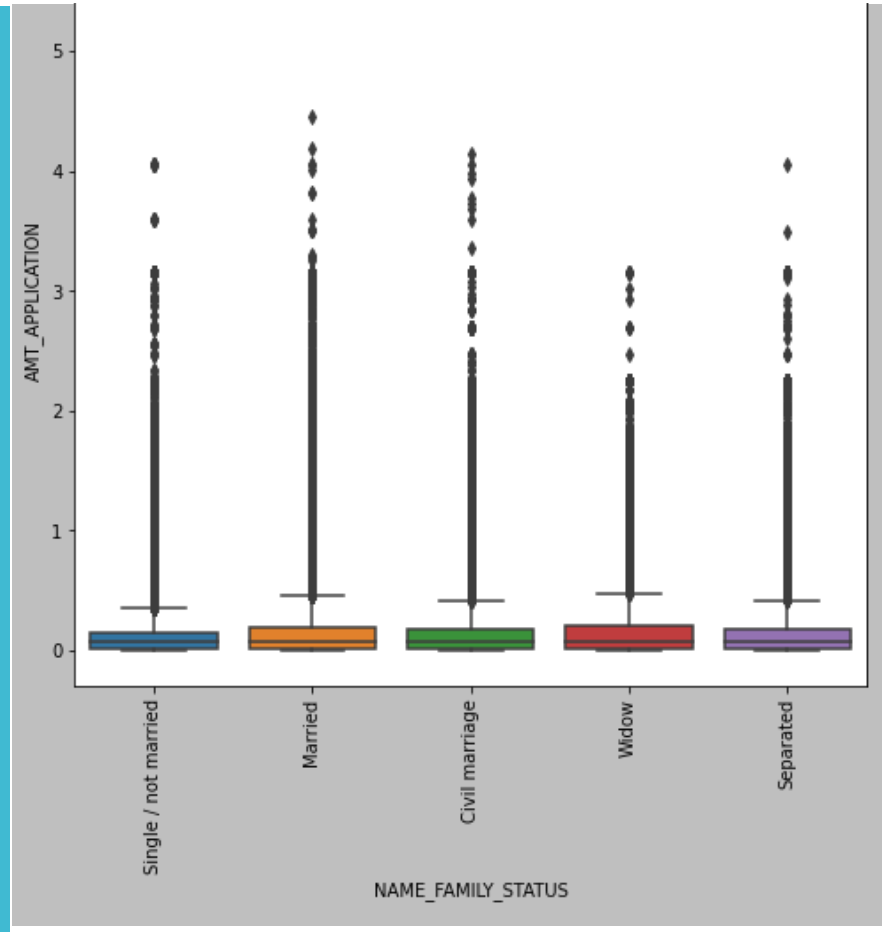
Incomplete higher education and lower secondary education have the lowest application and credit amount ranges from previous applications

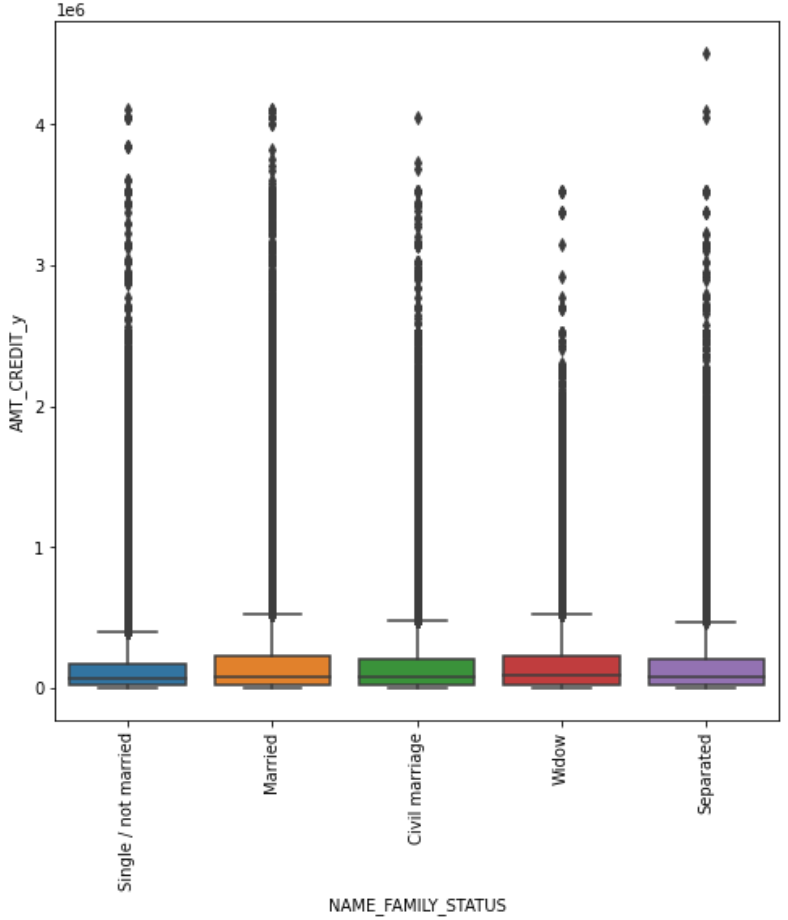




Various family types vs their previous application and credit amounts

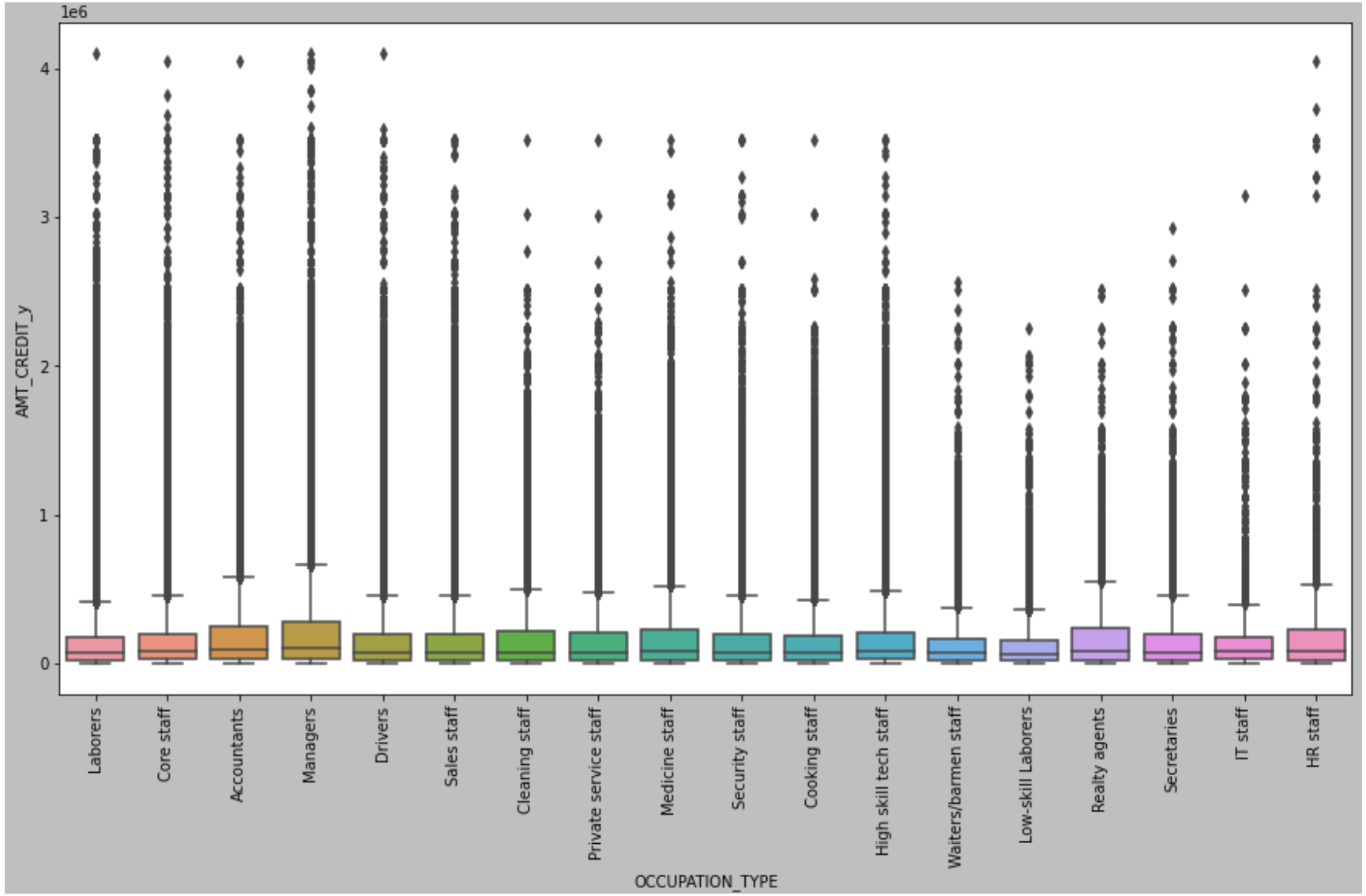
Single applicants having lowest range of application and credit amounts previously





Various occupation types vs their previous application and credit amounts

Low-skilled labourers and waiters have the lowest range of amounts



Conclusion

From the EDA of the given data frames we have reached on the conclusions:

- The highest targeted occupation type being low skilled labours, and having lower income from the current data and previous applications, should be considered the income type for risky defaulters
- The single and widowed defaulters have also shown risk from previous application data.
- Highest approval rate would be gained by the Managerial and Accountant income and occupation types. Other below average income types such as medicine staff and private service industry, cooking, cleaning and security staff although have on the lower income side have shown less risk of defaulting.