

CREDIT EDA ASSIGNMENT

- SWAPNIL SRIVASTAVA

Problem Statement

- ◆ Perform exploratory data analysis on a given set of Credit Data to help the company to decide for loan approval based on the applicant's profile.

Objective

- ◆ Credit Exploratory data analysis will help the company to make a decision for the loan approval by assessing the applicant's profile and by identifying the potential defaulters based on certain parameters.
- ◆ This will help the company to avoid financial losses and maintain a good loan record.

Data Set

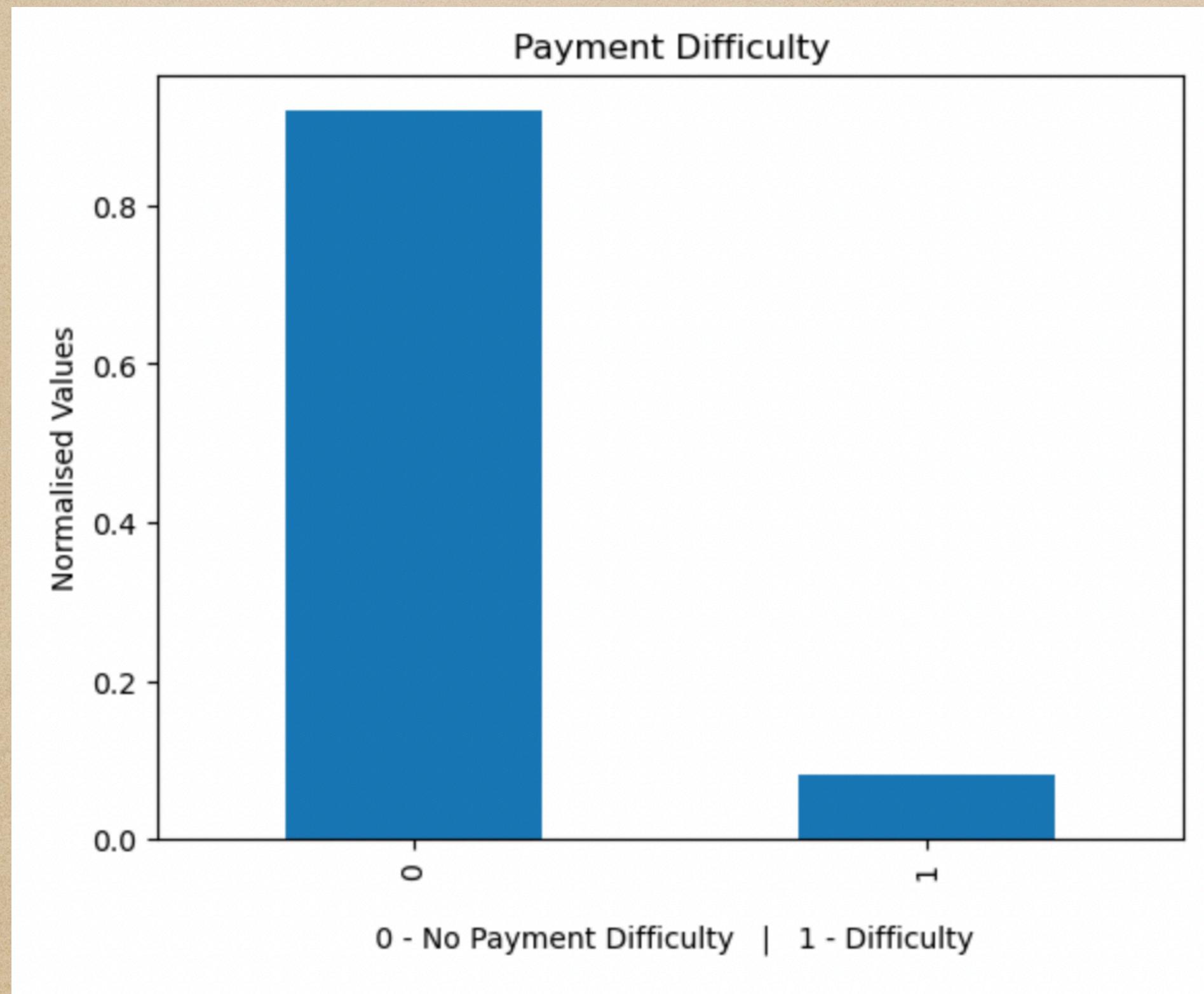
- ◆ application_data.csv -> contains the applicant's information at the time of loan application
- ◆ previous_application.csv -> contains the applicant's previous loan application information
- ◆ columns_description -> data dictionary containing the meaning of all the variables used in the other two datasets.

Steps required in EDA

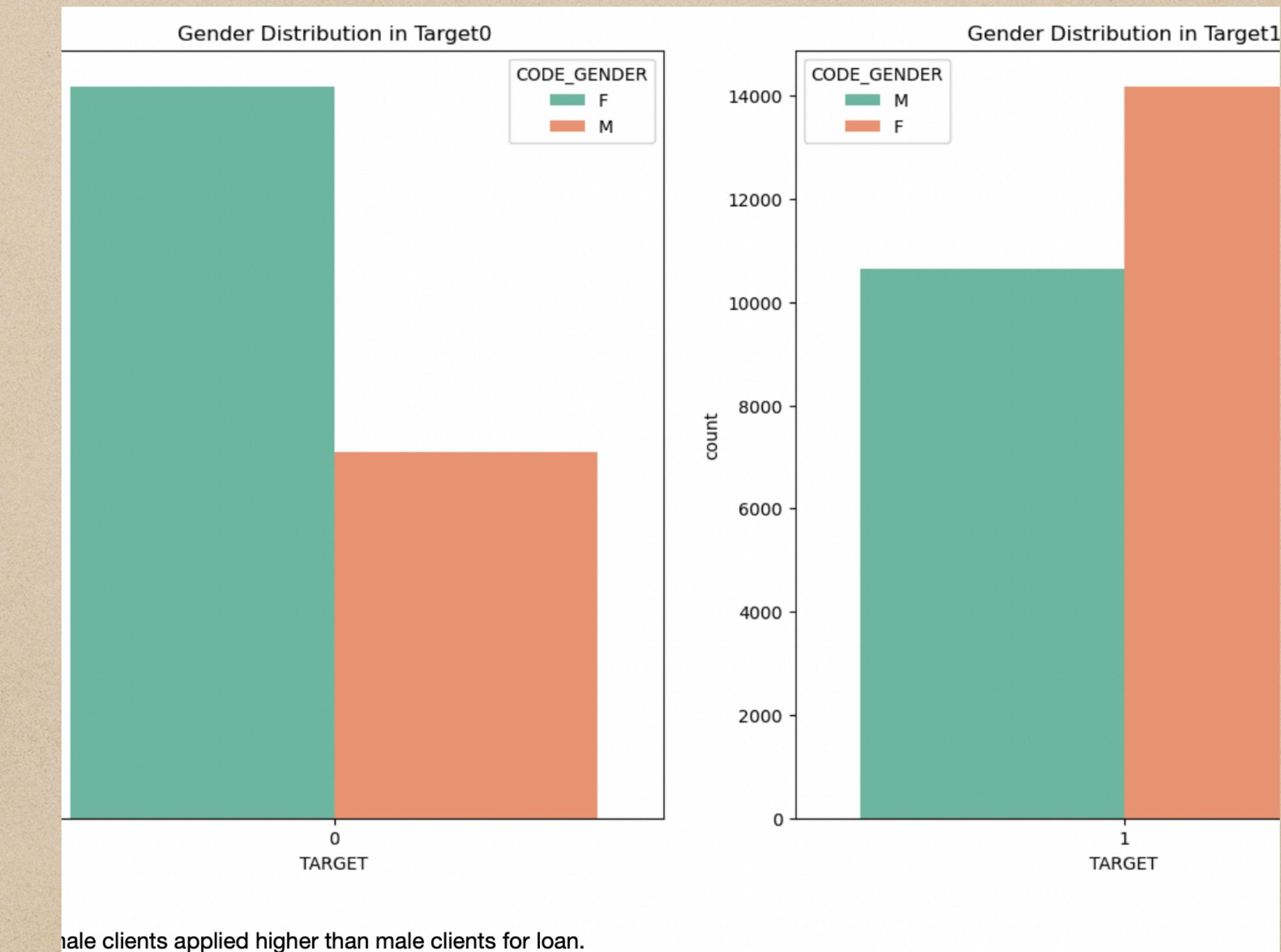
- ◆ Understand the variables -> Referring to the columns_description.csv file to understand the variables meaning and description. This is the first important step towards knowing the data in order to analyse it effectively.
- ◆ Import the necessary libraries -> Import the necessary libraries such as pandas, lumpy, matplotlib and seaborn into the Jupyter notebook.
- ◆ Import/Load the data -> Load the csv file into into the notebook to start the EDA.
- ◆ Check the overall structure of the data -> Get the overall structure of the data such as column names, some values, statistics, no. of rows and columns using basic commands.
- ◆ Identify columns with missing values -> Find out the columns containing missing values and observe the patterns, if any.

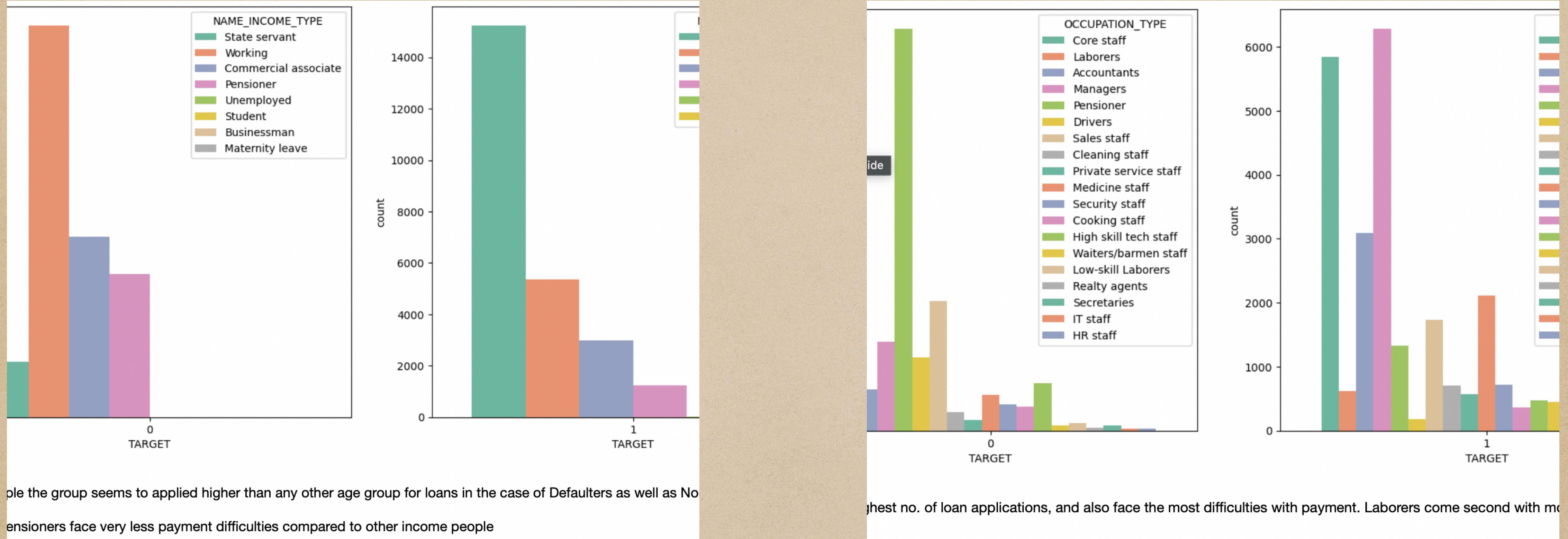
- ◆ Impute/drop missing value columns as per analysis -> Drop columns which have more than 40% missing values as they will not help in the analysis. Impute the other missing values with median or mode based on whether the column is numerical or categorical.
- ◆ Check for Outliers -> Identify the outliers in the dataset
- ◆ Univariate Analysis -> Analysing singles column variables to uncover patterns and derive insights. Eg: Histplot, box plot, dsistplot.
- ◆ Bivariate/ Multivariate Analysis -> Analysing multiple columns variables to establish any patterns in order to identify any dependency or relationship between them.
- ◆ List the observations -> Interpret and list down the observations derived from the EDA.

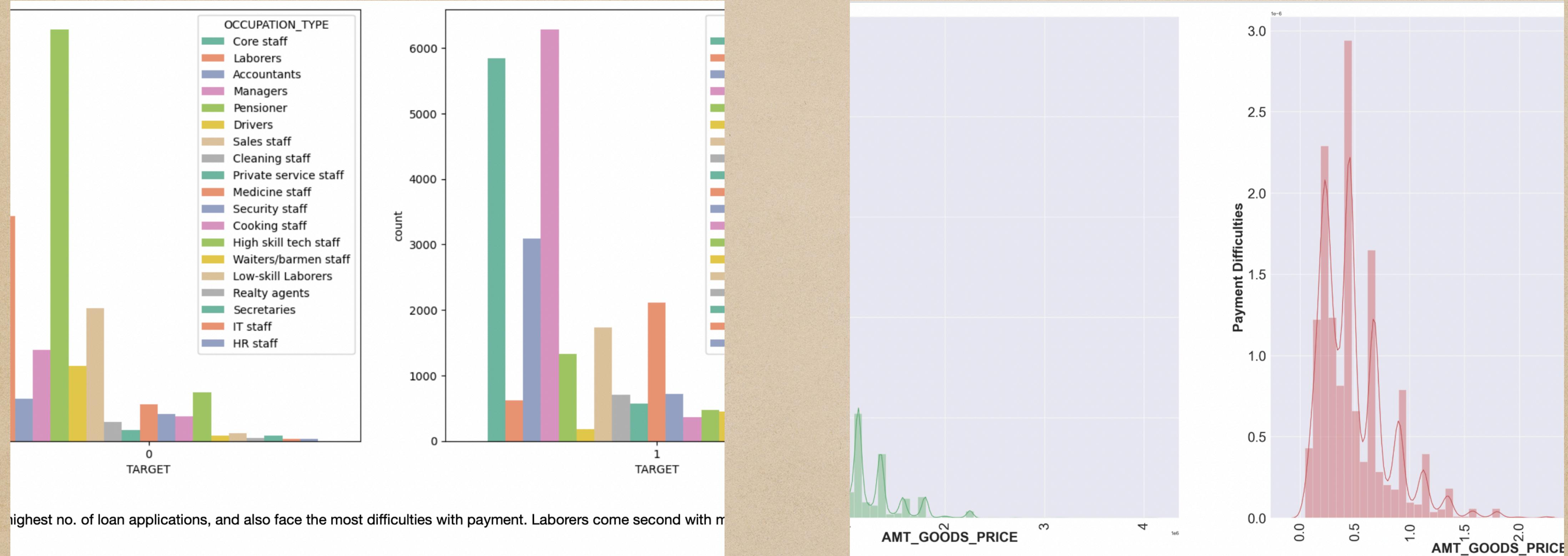
Univariate Analysis Graphs



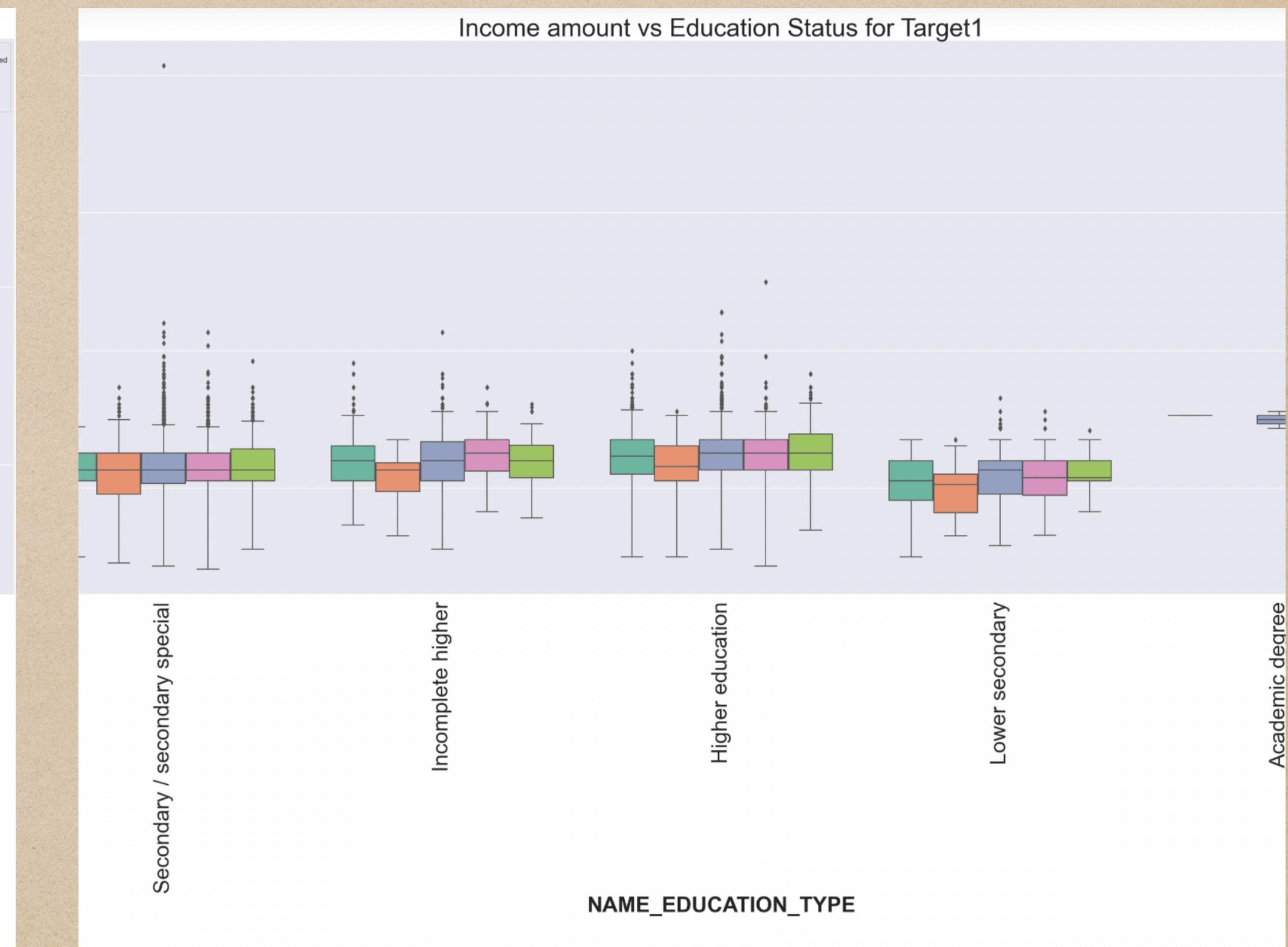
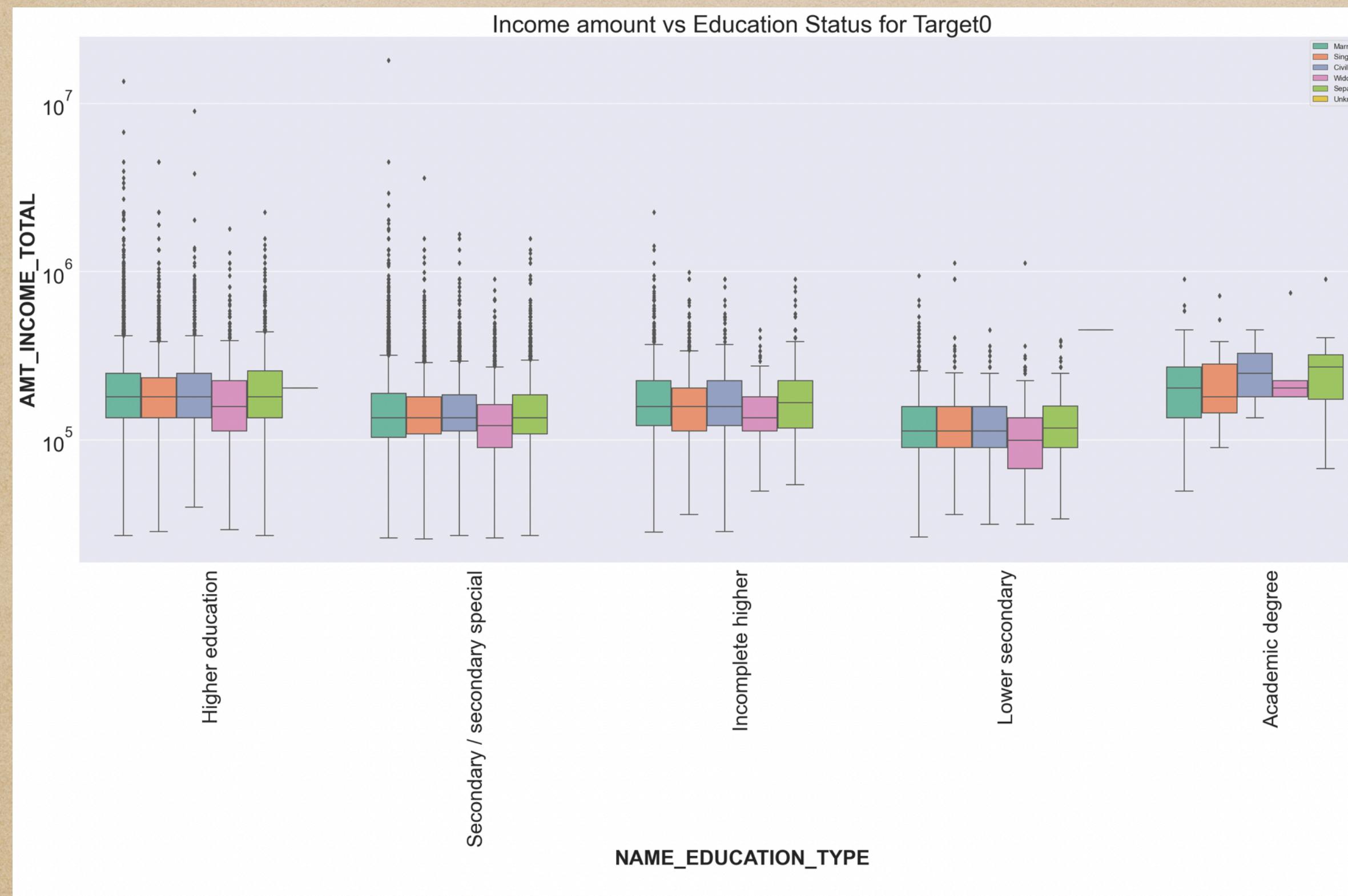
Target variable analysis

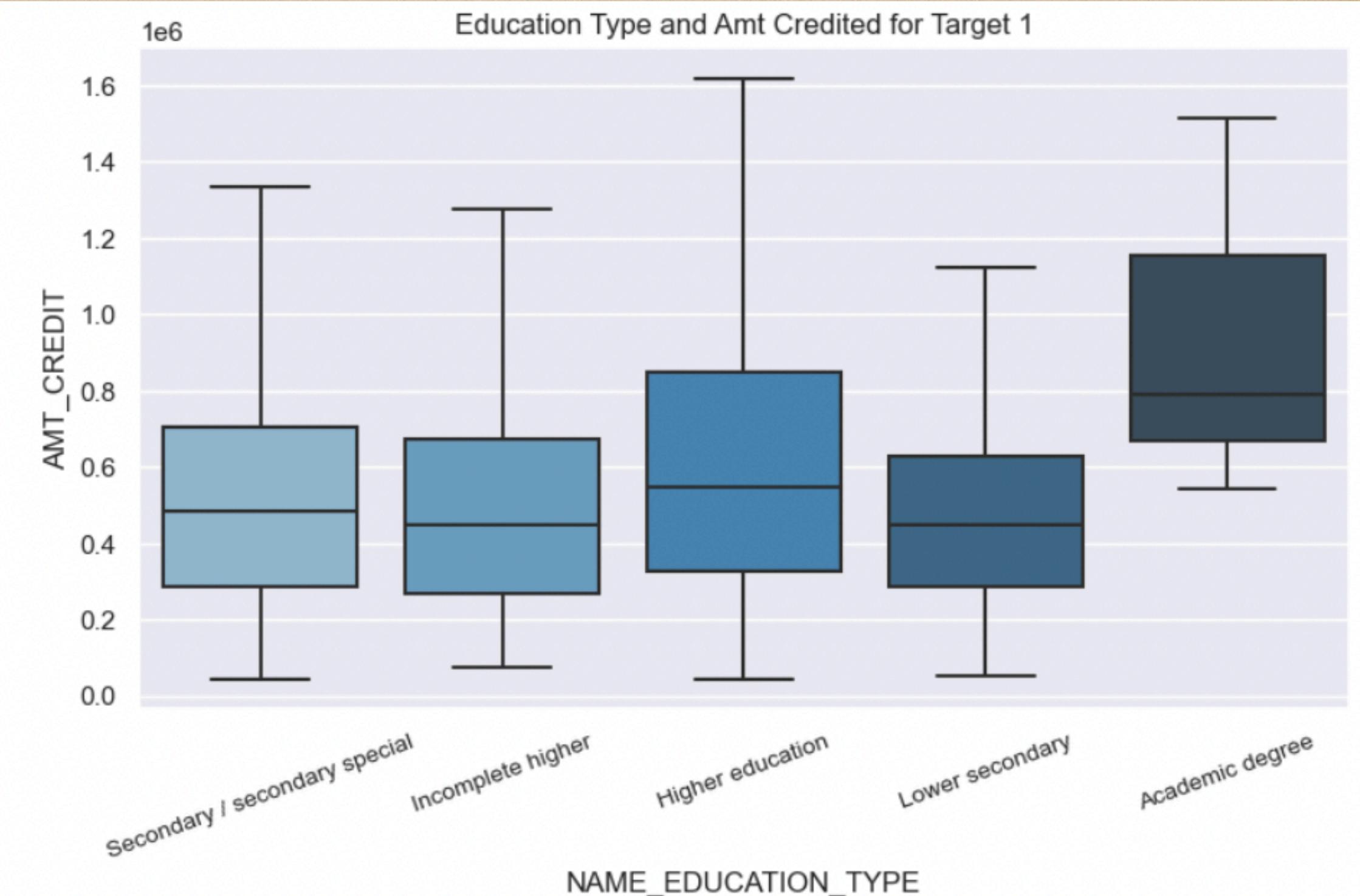
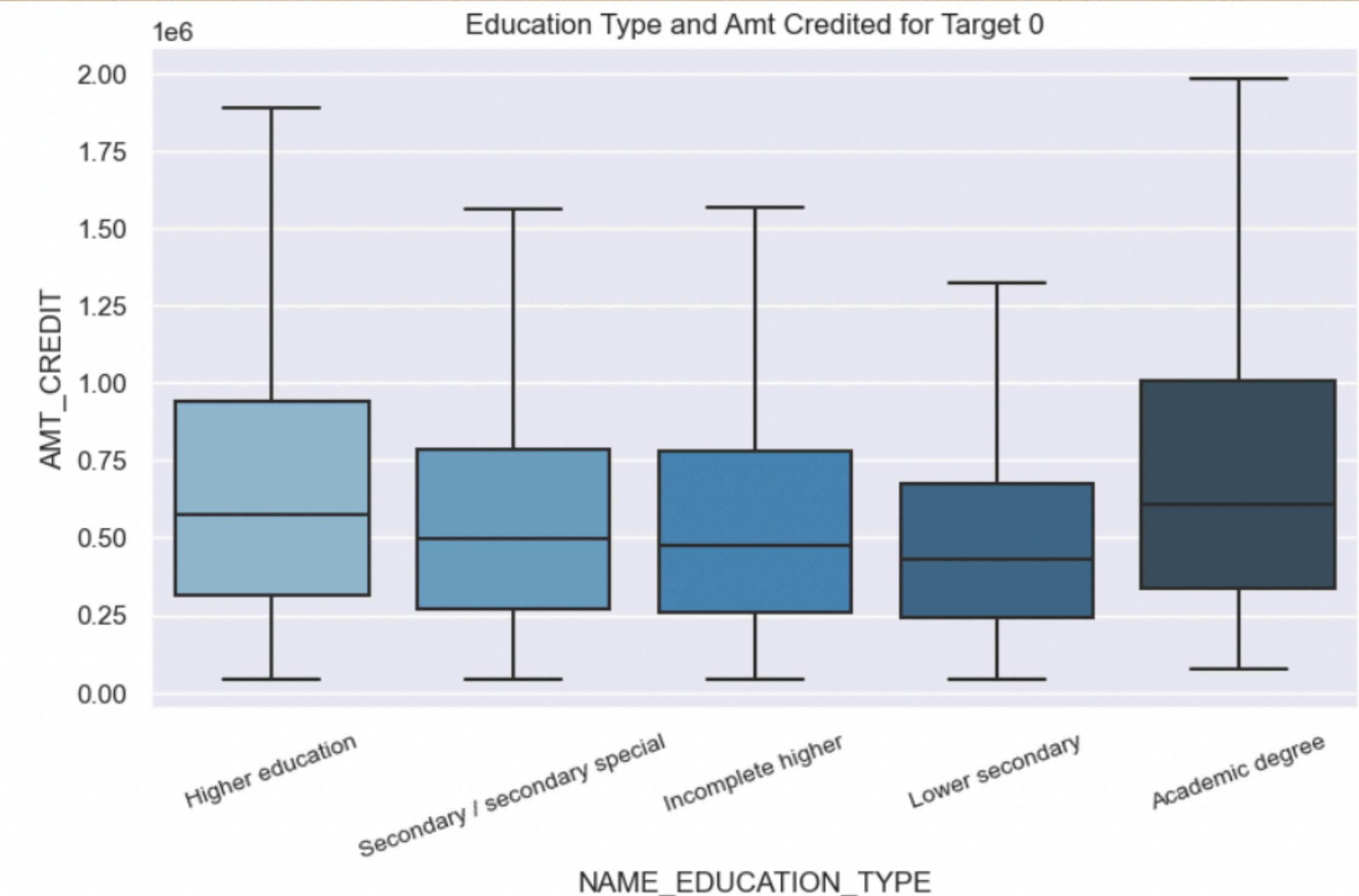


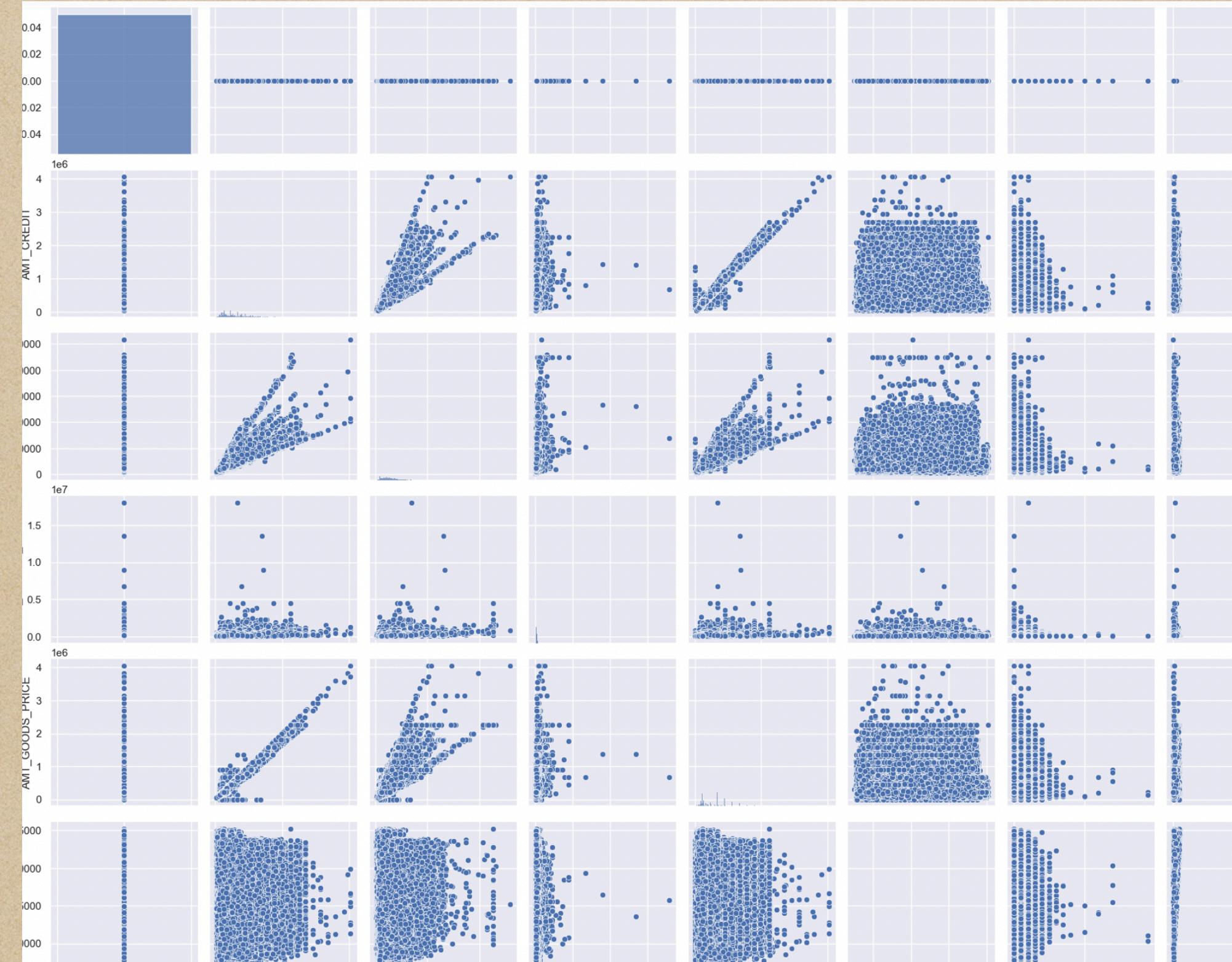




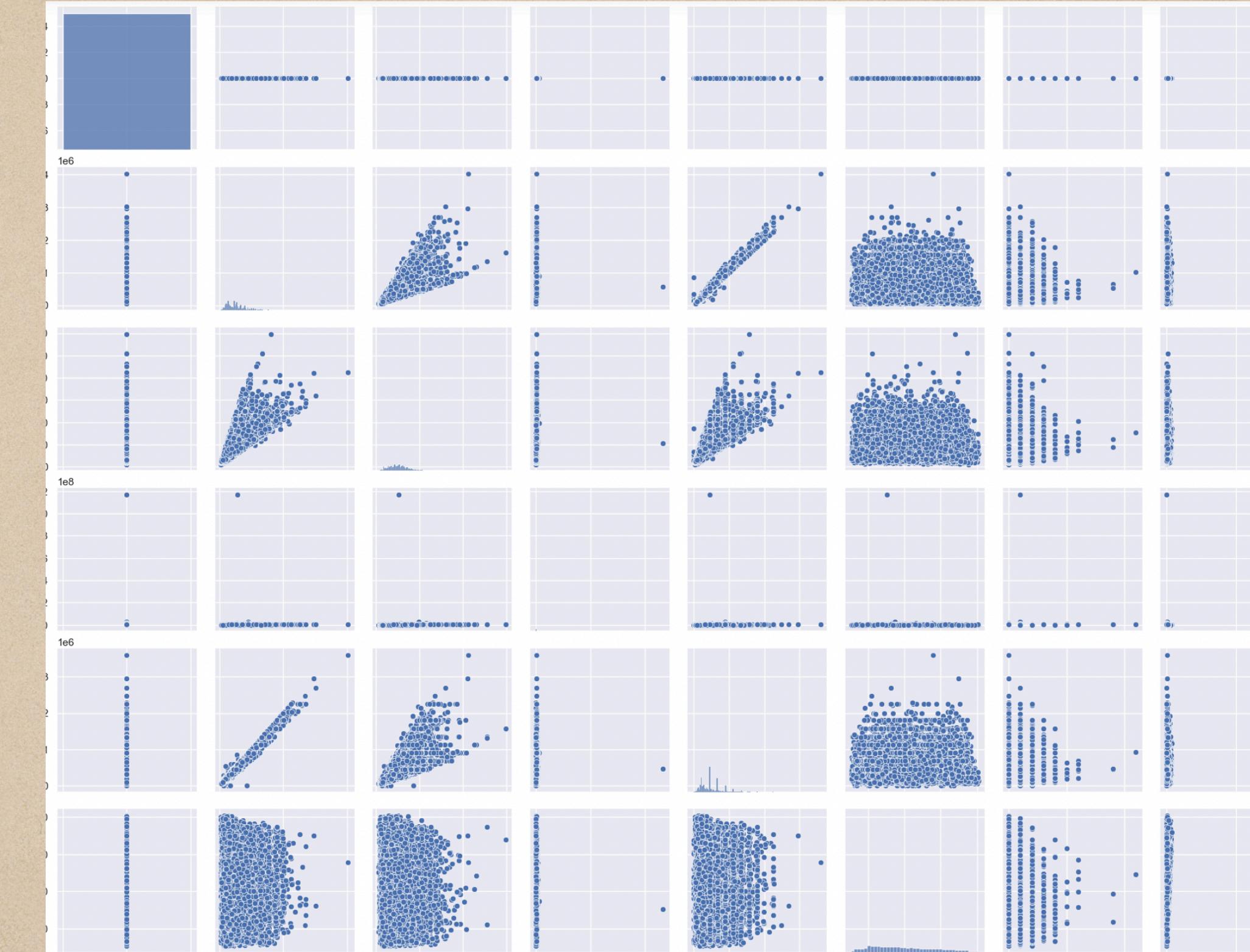
Bivariate Analysis Graphs







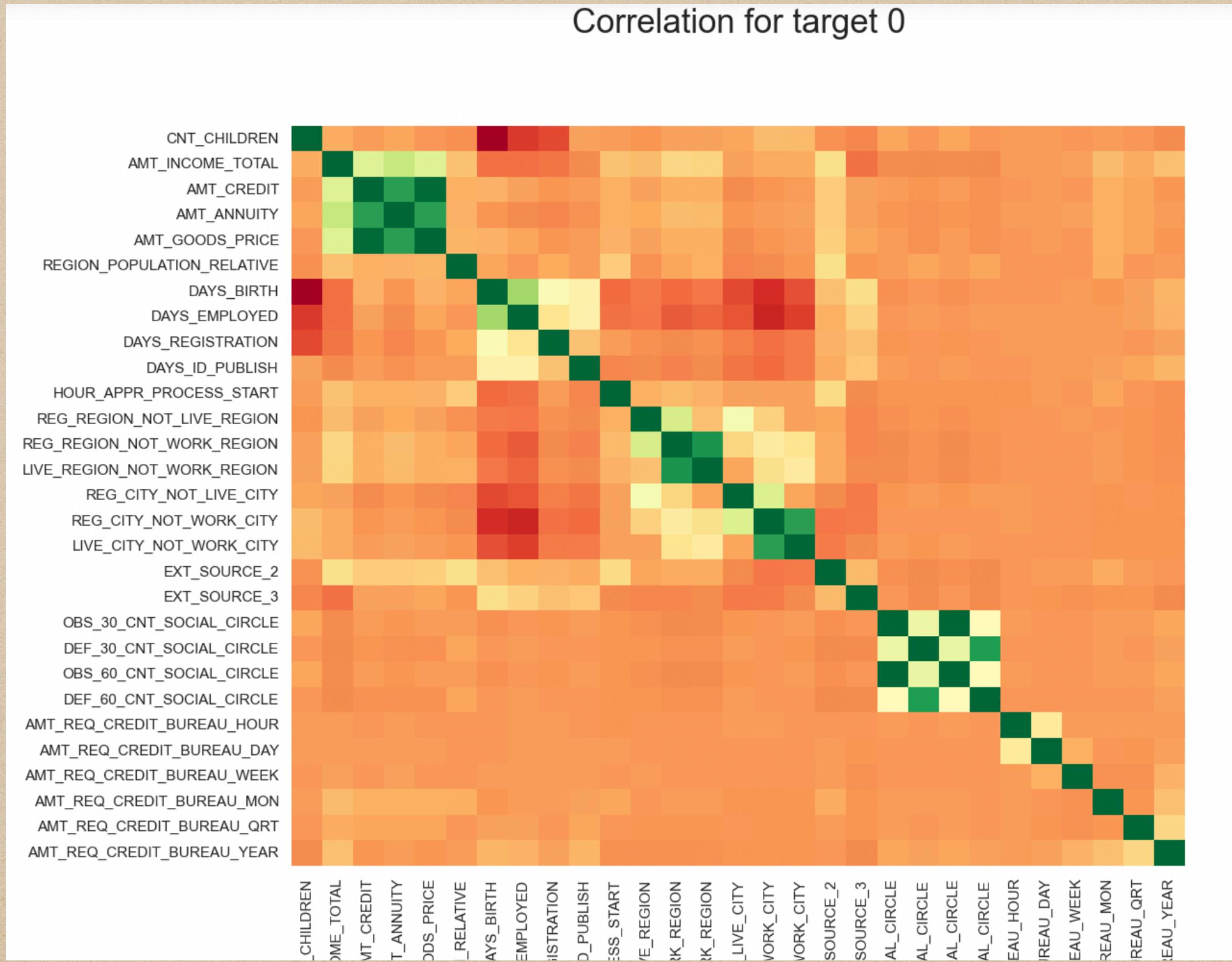
Target0



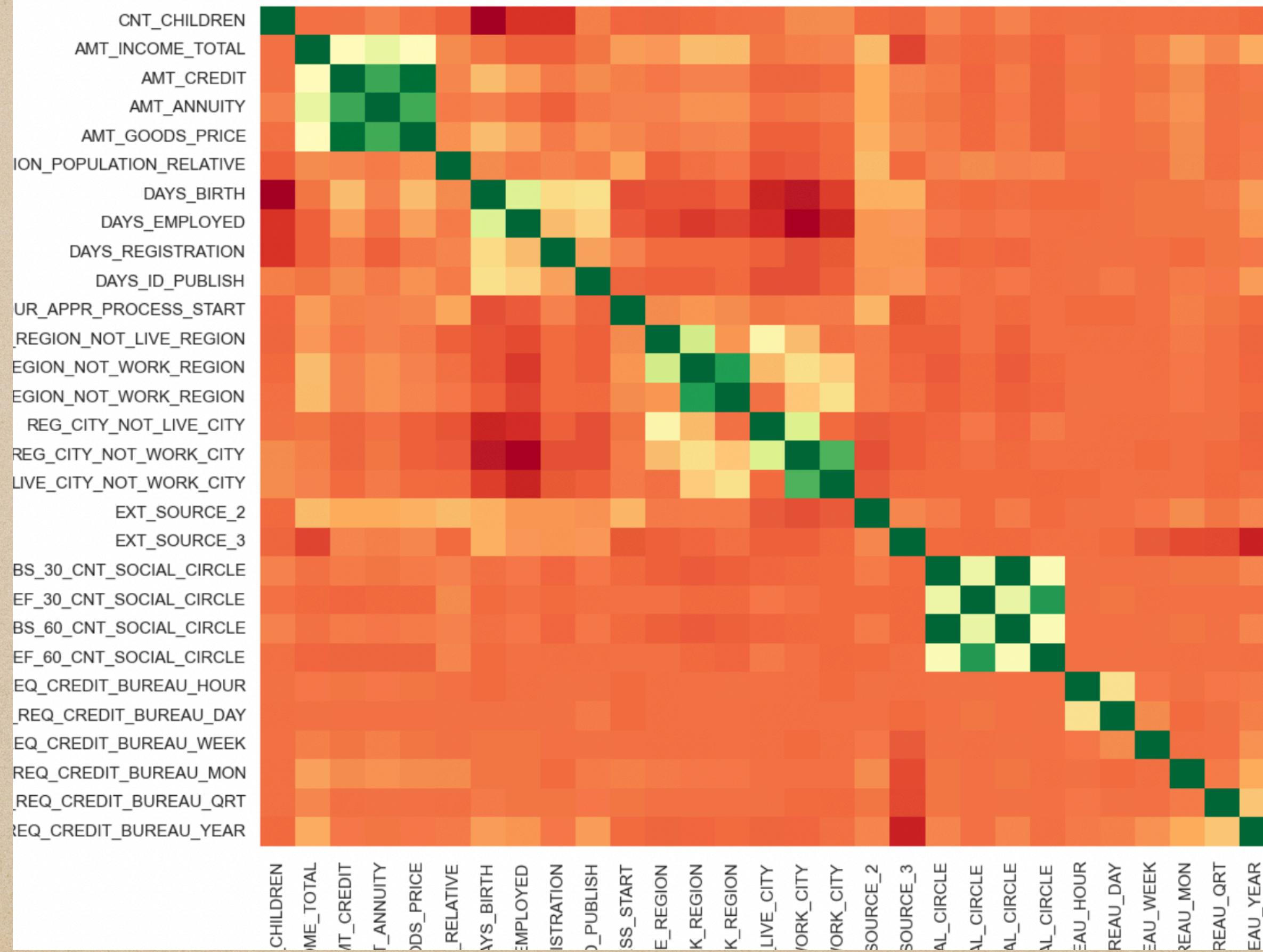
Target1

Multivariate Analysis

Correlation for target 0



Correlation for target 1



Observations

- ◆ 1 in every 11 applicant has a payment difficulty.
- ◆ It seems like Female clients applied higher than male clients for loan.
- ◆ 66.6% Female clients are non-defaulters while 33.4% male clients are non-defaulters.
- ◆ 57% Female clients are defaulters while 42% male clients are defaulters.Jj
- ◆ Working income people the group seems to applied higher than any other age group for loans in the case of Defaulters as well as Non-defaulters.
- ◆ State servants and pensioners face very less payment difficulties compared to other income people.

- ◆ Pensioners have the highest no. of loan applications, and also face the most difficulties with payment. Laborers come second with most payment difficulties
- ◆ The income amount for Married clients with an academic degree is much lesser as compared to others.
- ◆ (Defaulter) Clients have relatively less income as compared to Non-defaulters.
- ◆ Positive correlation between amount annuity and amount credit, meaning more credit so EMI increases

- ◆ Credit amount is higher to densely populated area.
- ◆ Most rejection of loans came from purpose 'repairs'.
- ◆ For education purposes we have equal number of approves and rejection
- ◆ Paying other loans and buying a new car is having significant higher rejection than approves.
- ◆ Those who refused to name goals had far more rejections than approvals.
- ◆ Loans with 'Repairs' are facing more difficulties in payment on time.
- ◆ There are few places where loan payment is significant higher than facing difficulties. They are 'Buying a garage', 'Business development', 'Buying land', 'Buying a new car' and 'Education'

- ◆ The credit amount of Loan purposes like 'Buying a home', 'Buying a land', 'Buying a new car' and 'Building a house' is higher.
- ◆ Students generally apply for a high loan credit, which makes sense as they need high money for education.
- ◆ Unemployed people too apply for a higher credit, maybe most of them are pensioners.
- ◆ We see that those with co-op apartments had significant amount of difficulty with payments, which is explanatory that the people in co-op apartments do not have very well income

Recommendations

- ◆ Banks should focus more on contract type 'Student' , 'Pensioner' and 'Businessman' with housing 'type other than 'Co-op apartment' for successful payments.
- ◆ Banks should focus more on income type 'Working' and 'Commercial Associates' as they are having most number of unsuccessful timely payments.
- ◆ Also with loan purpose 'Repair' is having higher number of unsuccessful payments on time.
- ◆ Get as much as clients from housing type 'With parents' as they are having least number of unsuccessful payments.

Thank You