Problem Statement

Purpose of this Analysis

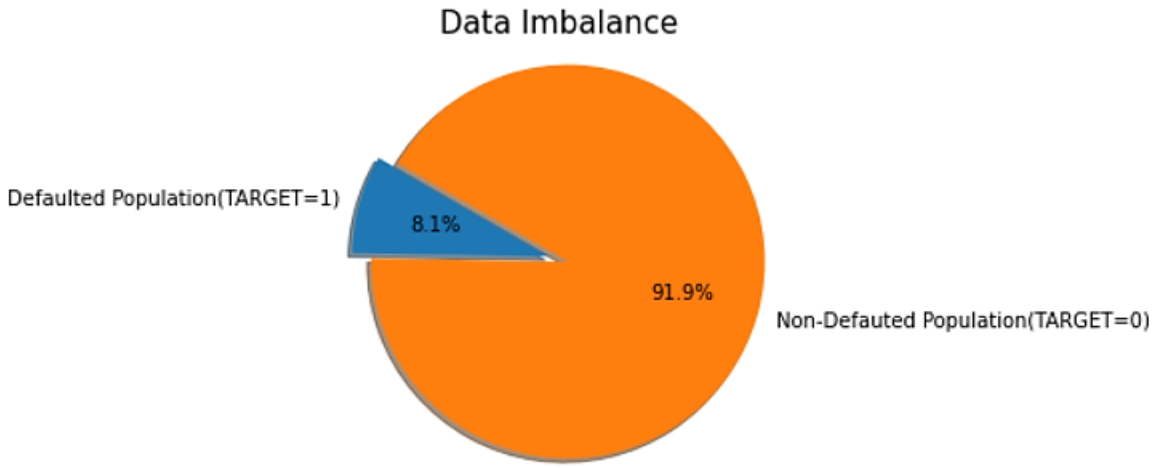
- To know the driving factors / driver variables of loan default.
- To identify patterns indicating if the Client has difficulty in paying their instalments on time.
- Based on above, to take actions such as denying the loan, reducing the amount of loan, lending to risky applicants at a higher interest rate, etc.

Why doing this?

- To understand how both Client and loan attributes influence the tendency of loan default.
- This insights can be used to target the right Clients / Consumers.
- To support the existing clients for their timely repayment, without affecting the company’s Profit.

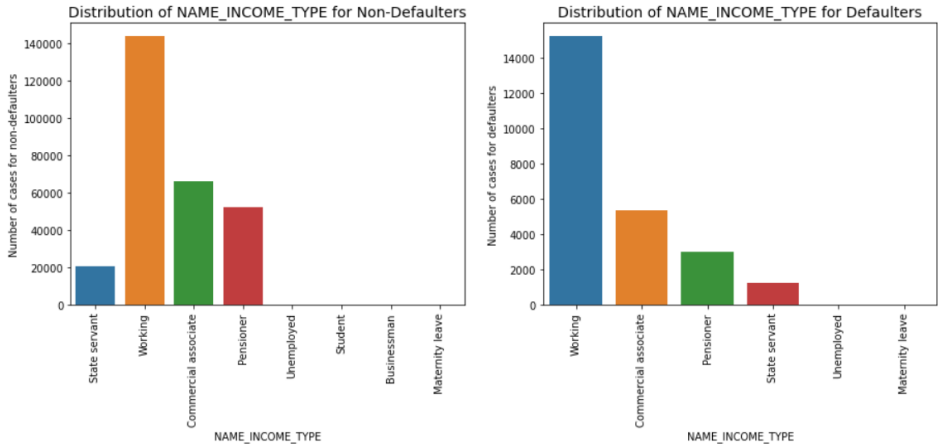
Target Variable

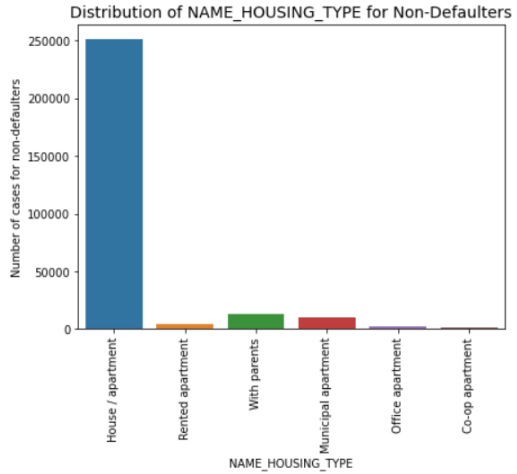
Our target variable is whether or not a client is able to pay back the loan on time
i.e., whether our client is a defaulter of loan or not.

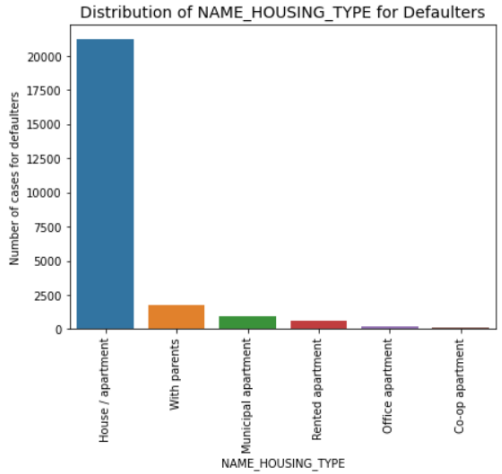


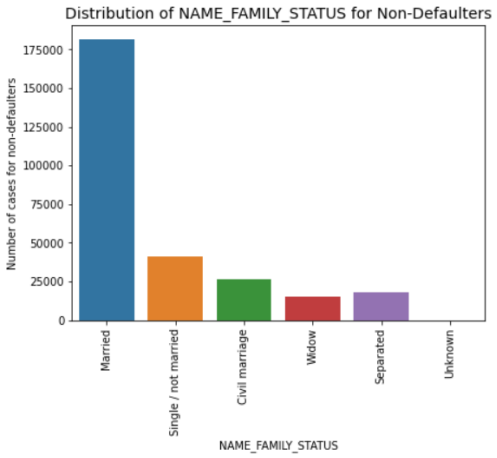
From the data imbalance, we could see that **91.9%** of our clients paid back the dues on time, whereas only 8.1% of our clients couldn't make it on time.

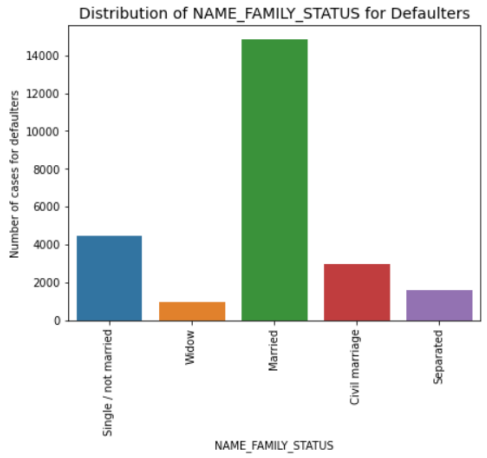
Bivariate Analysis for Unordered Categorical columns

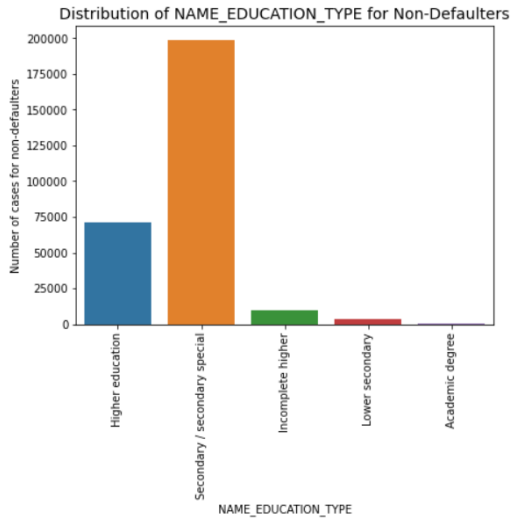


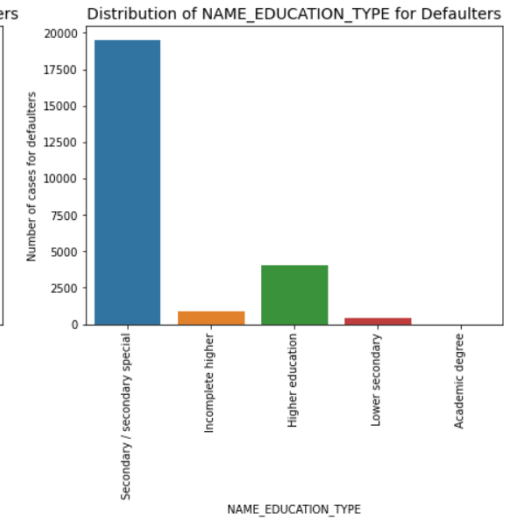










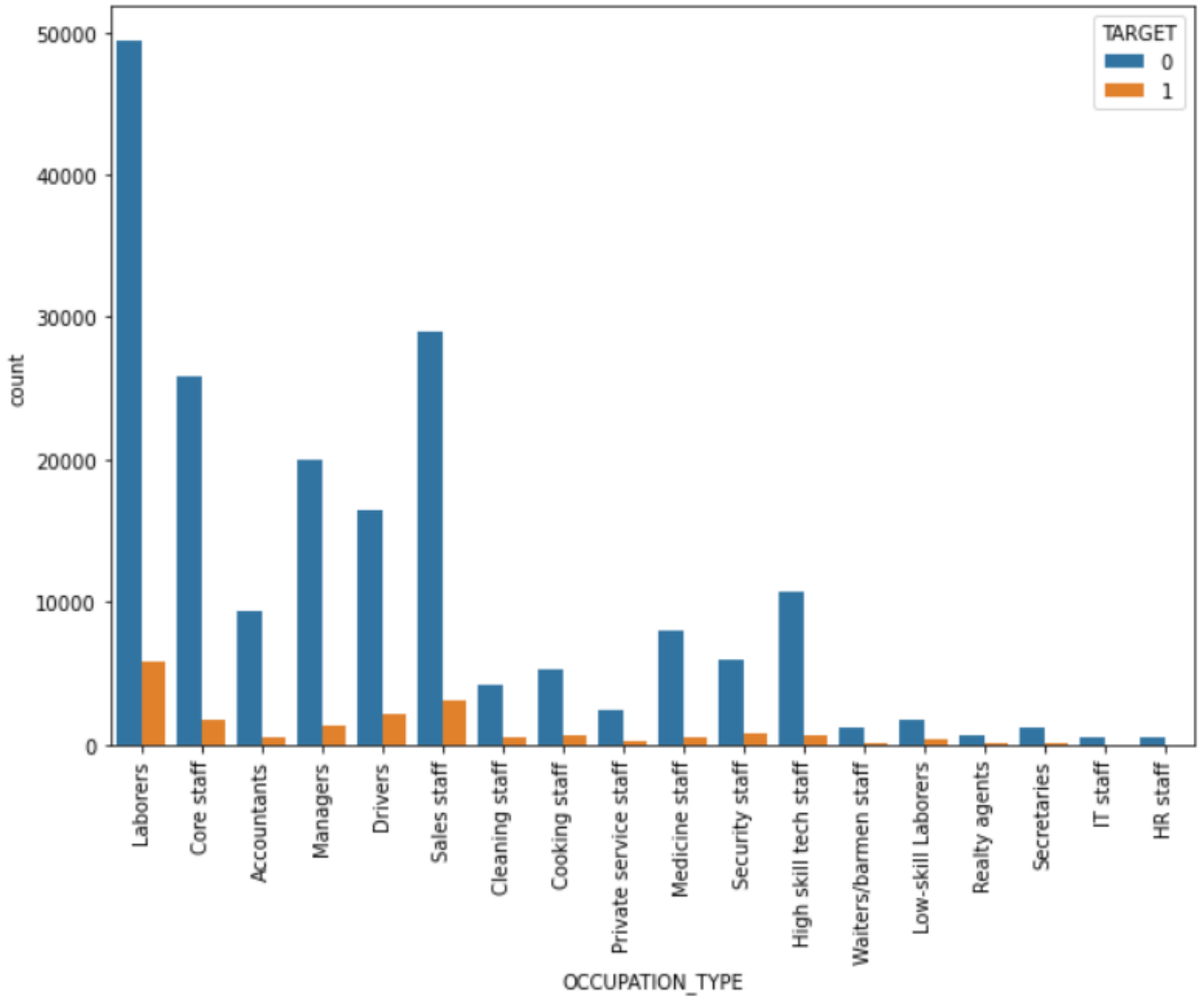


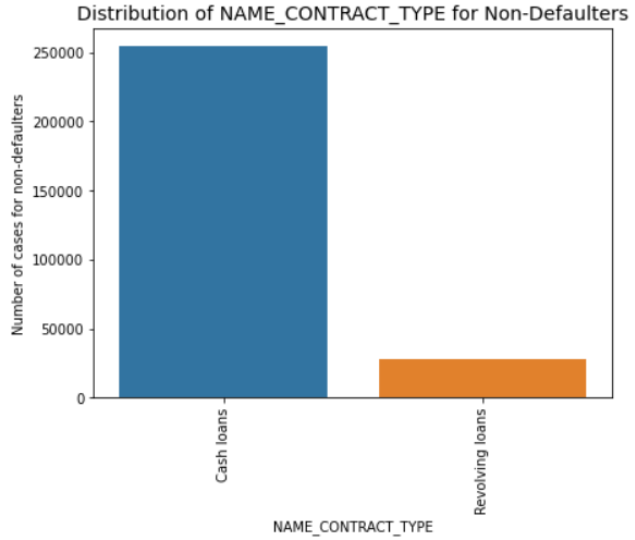
Bivariate Analysis for Unordered Categorical columns

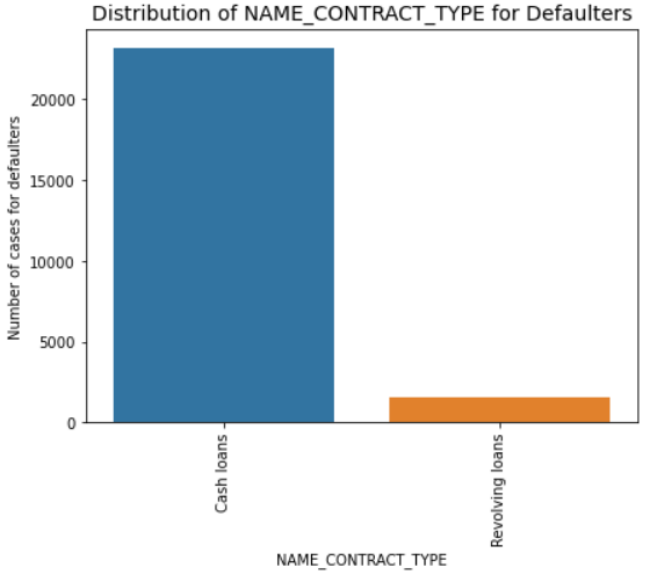
- Clients tend to follow the same pattern in different stages of their education, their family/relationship status, housing type, income/occupation type, be defaulting the loan or not.

- Client who has completed his/her secondary education and, a working-class professional, who is married and, living in their own house/apartment has the maximum tendency of loan default.

Bivariate Analysis for Ordered Categorical columns





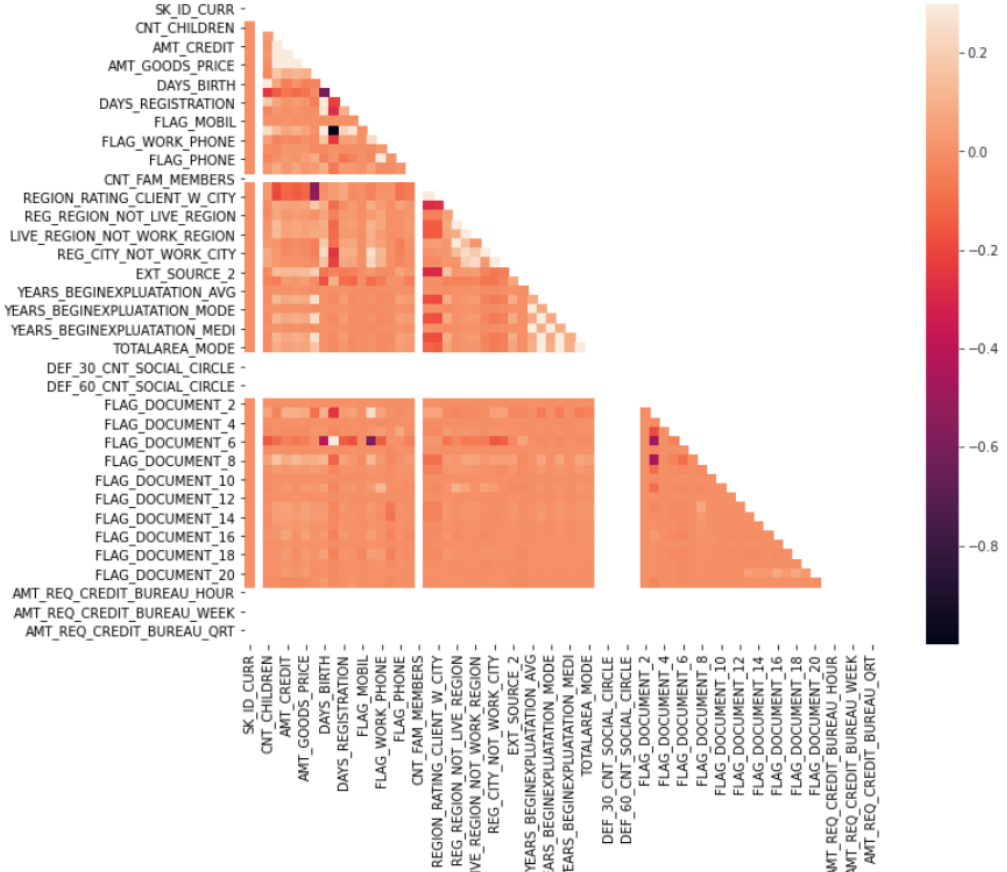


Bivariate Analysis for Ordered Categorical columns

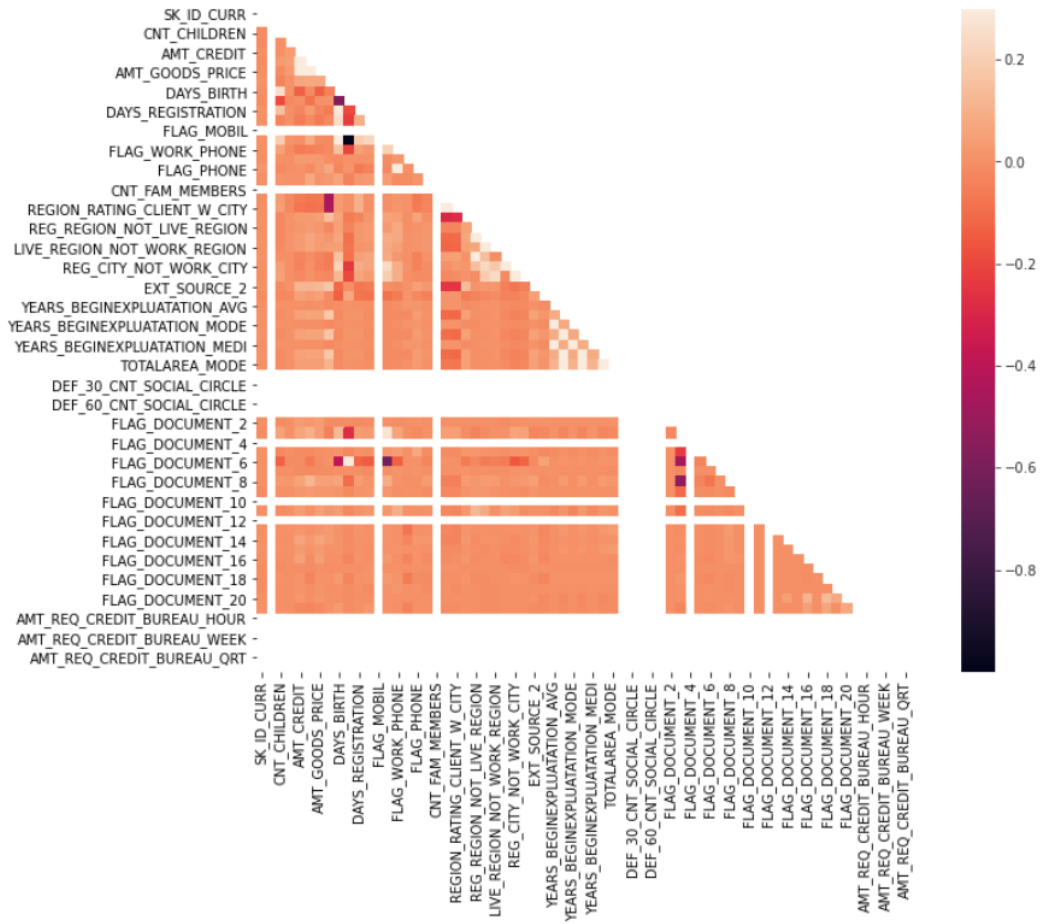
- We could see that the clients who are working as laborers have defaulted the least. However, we saw that this category also had the least number of missing values, which was left untouched.
- We could see that here, on contrary to our previous analysis, there are more defaulters from the clients who are self-employed/business than, the working-class professionals.
- We could see that there are more defaulters with cash loans compared to the revolving loans.
- This may be because the clients who are self-employed/business owners take revolving loans, as they may keep their cash flows in rotation. Whereas, a working-class professional take cash loans, as they are salaried.

Correlation for Numerical columns with targets = 0, 1

Target = 0



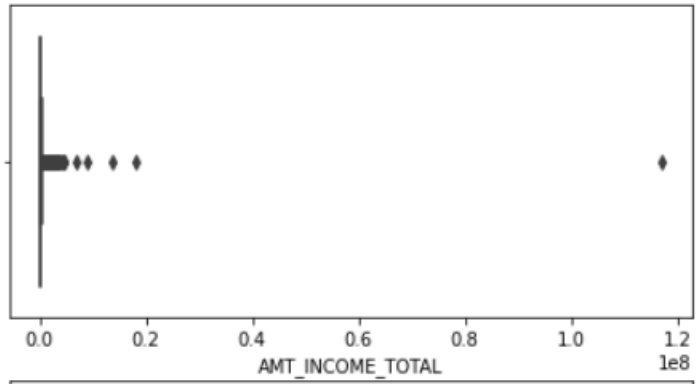
Target = 1

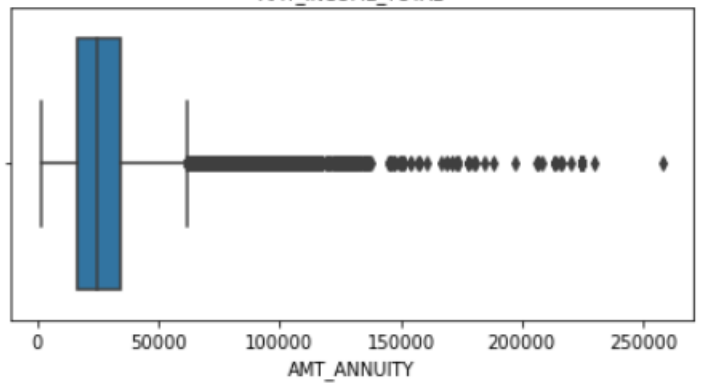


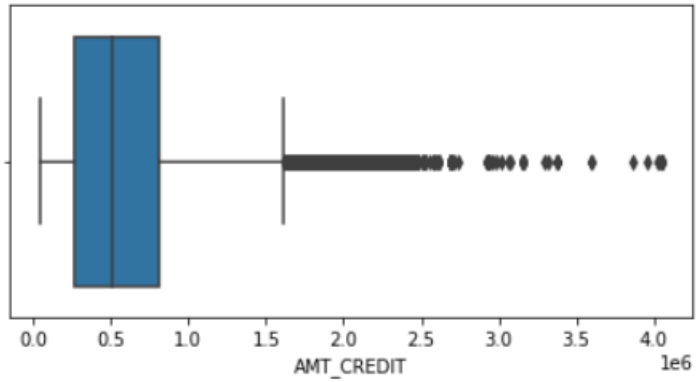
Correlation for Numerical columns with targets = 0, 1

- We could see that all the columns with both target = 0 and target = 1 have similar correlations.
- We could also see that the correlations are missing for some of the flag value columns.
- All the AMT columns have a good correlation with each other. Hence, we check if there are any outliers for the same.

Looking at the outliers for the Amount columns



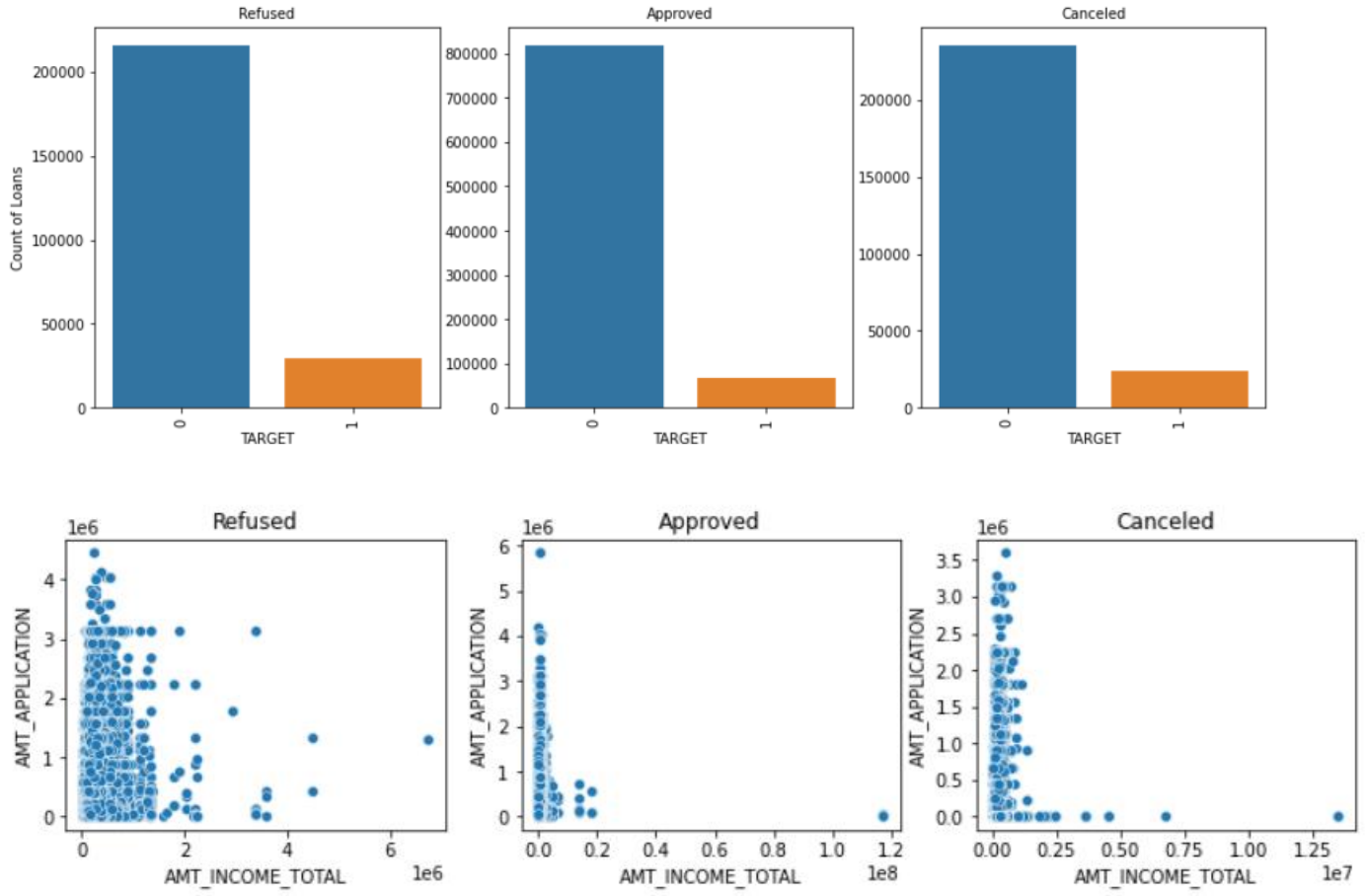




Looking at the outliers for the Amount columns

- We already saw that the vast majority of our clients are a working-class professionals, that is reflected in their income.
- Loan credit is relatively on the lower side for the vast majority of our clients. After which, the number of outliers are steadily decreasing from middle to high. This is because of the shift to self-employed/business owners clients.
- Annuity amount also follows the same pattern as that of Loan credit.
- We may say that from our last two plots, that a clients may borrow more money when their credit limit is high.

Multivariate Analysis of our Target variables

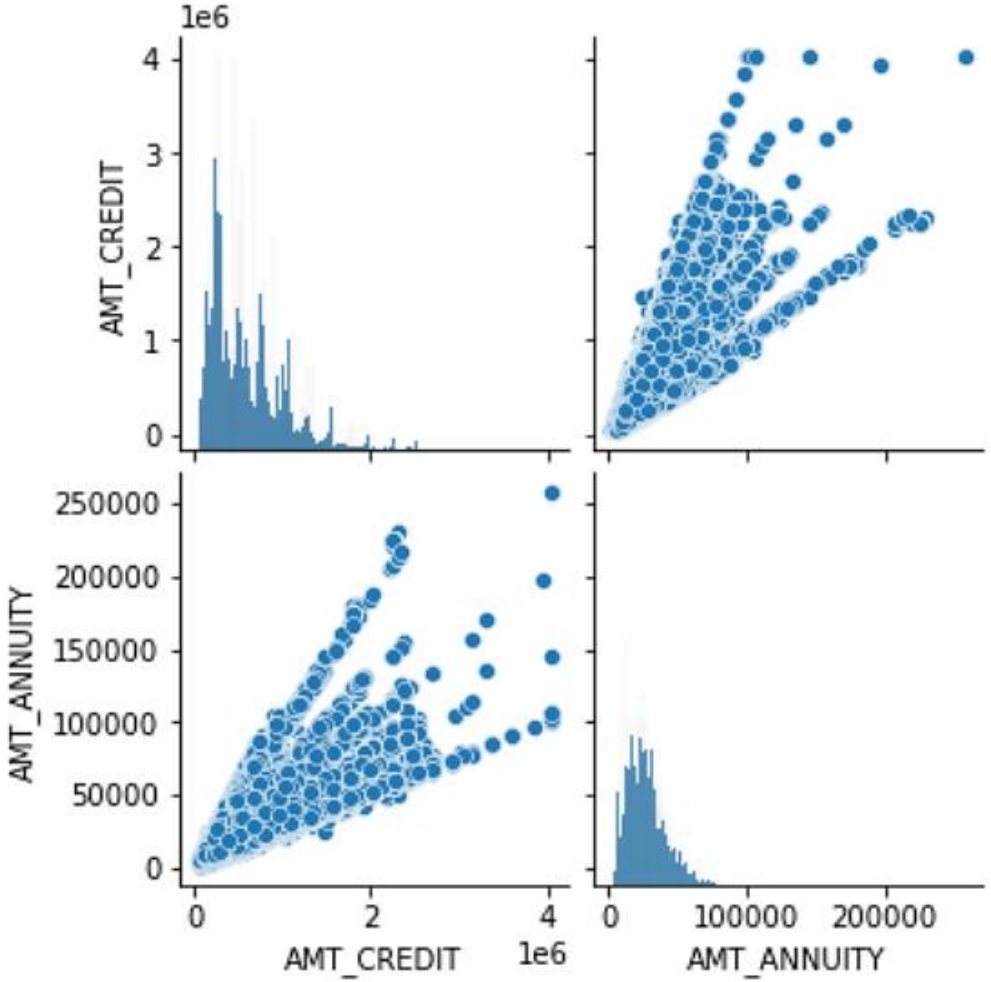


*There is some error in our data analysis hence, we didn't plot for "Unused Offer"

Multivariate Analysis of our Target variables

- Most of the loans were not defaulted, which was previously refused, approved, cancelled.
- From the scatterplots, we could say that higher the client's request for credit limit, more is the chances of that application being rejected.
- Cancelled applicants have a relatively moderate credit limit, which may have been sanctioned by competitor Banks.

Multivariate Analysis of our Target variables



Multivariate Analysis of our Target variables

- Plot (2,2,3) has a higher positive correlation than the Plot (2,2,2).
- Credit limit increasing, and the clients tendency to borrow more money, is proportional to each other.
- However, there is relatively more increase in credit limit while the client borrows more money, than vice-versa.
- Plot (2,2,1) follows near skewed distribution whereas, the Plot (2,2,4) follows a near normalized distribution.
- We may conclude from these pairplots that, there is an initial surge in the credit limit while a client starts borrowing more money, which then gradually decreases with respect to the amount borrowed.

Summary

- Most of our clients were able to payback their dues on time.
- We determined which clients have a hard time repaying the loan, based on their education, marital status, income level, residential status.
- We saw how two different types of loan, each influence different clients. We also contrasted the behaviour of working-class professional with self-employed/business owners, in regard to the loan repayment of each type.
- We compared how a client's income level affects him/her from getting their desired credit limit for their loan.
- We saw how a clients request for credit limit influences the decision of the loan to be sanctioned, by both the company and the consumer.
- We saw how their credit limit increases when a client begins to borrow more money from the Bank.

Hence, the major **driver variables** for loan default are:

- ✓ Social and Economic status of the **Client**.
- ✓ Type of Loan provisioned by the **Bank** – Cash & Revolving.