CREDIT EDA CASE STUDY

CRYSL LOBO

Problem Statement

The data given below contains the information about the loan application at the time of applying for the loan.

It contains two types of scenarios:

• The client with payment difficulties: he/she had late payment more than X days on at least one of the first Y installments of the loan in our sample dataset.

• All other cases: All other cases when the payment is paid on time.

The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

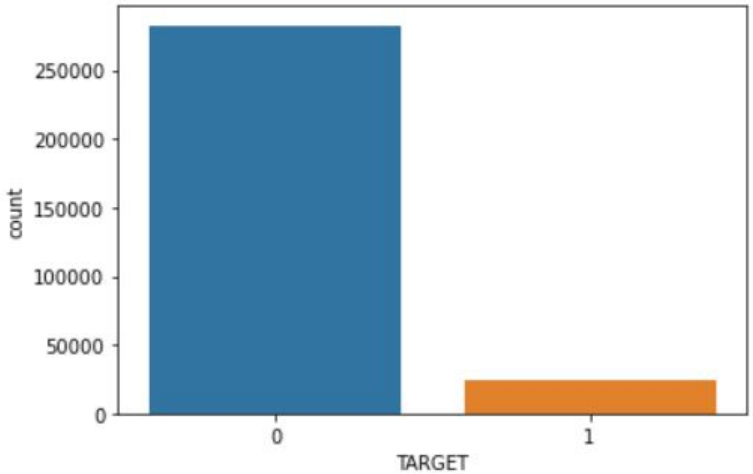
Problem Solving Methodology

- The required libraries needed for data cleansing and visualisation were imported.
- Data cleansing for columns was done wherever necessary and dropped the columns which had more than 30% of the missing values.
- Outliers are identified and handled wherever possible. Data imbalance was checked.
- Created new columns as per the requirements.
- Univariate/Bivariate Analysis of the relevant Categorical/numerical was done and insights are derived.
- Current and Previous application data was merged to derive insights based on bank Approval loan status.

Data Imbalance

The data was divided into two dataframe based on the Target column.

This helped us to perform Segmented Univariate Analysis on both the dataframes.

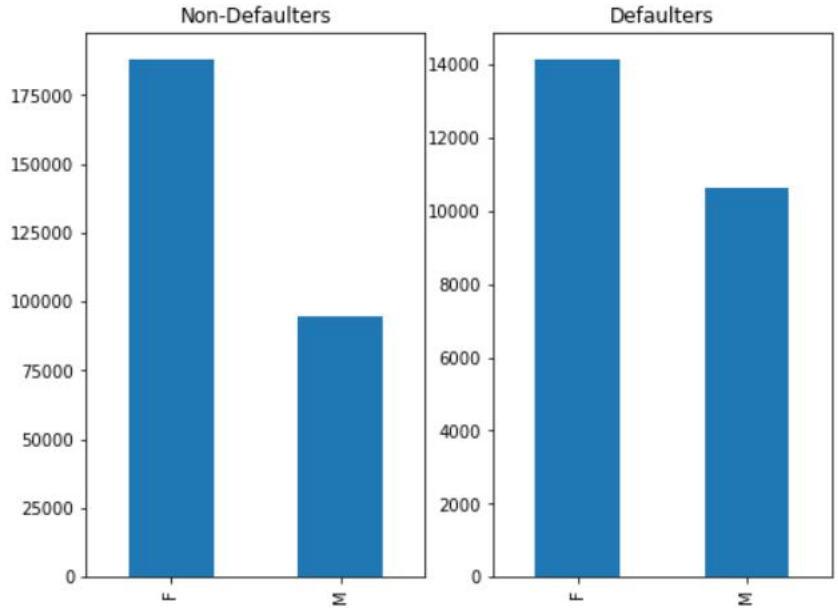


The ratio of imbalance is 11.39.

We can infer that there are 11 0's for every 1 present in the dataframe.

Univariate Analysis of Categorical Columns

Gender Column

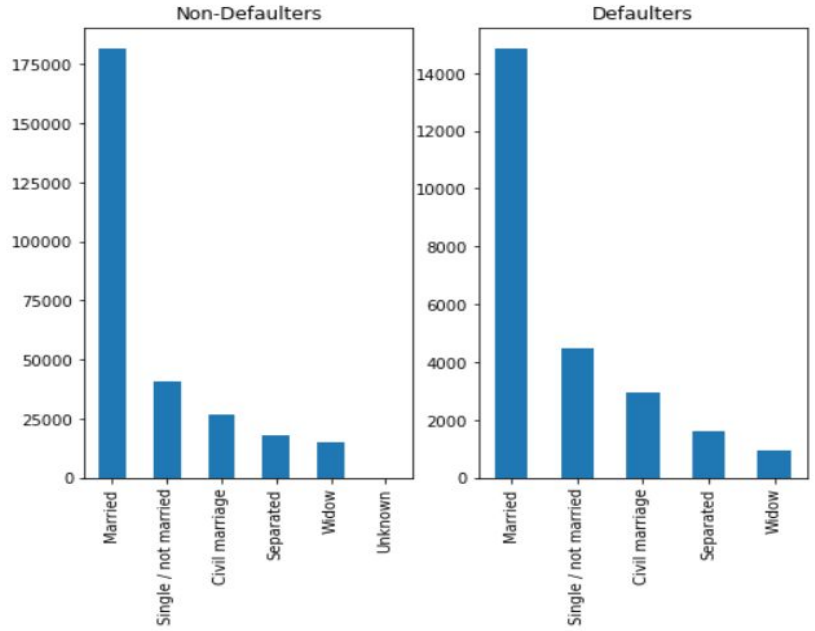


The count of females opting for loans are more than than of males.

Males are having more payment difficulties as compared to Females.

Univariate Analysis of Categorical Columns

Family Status Column

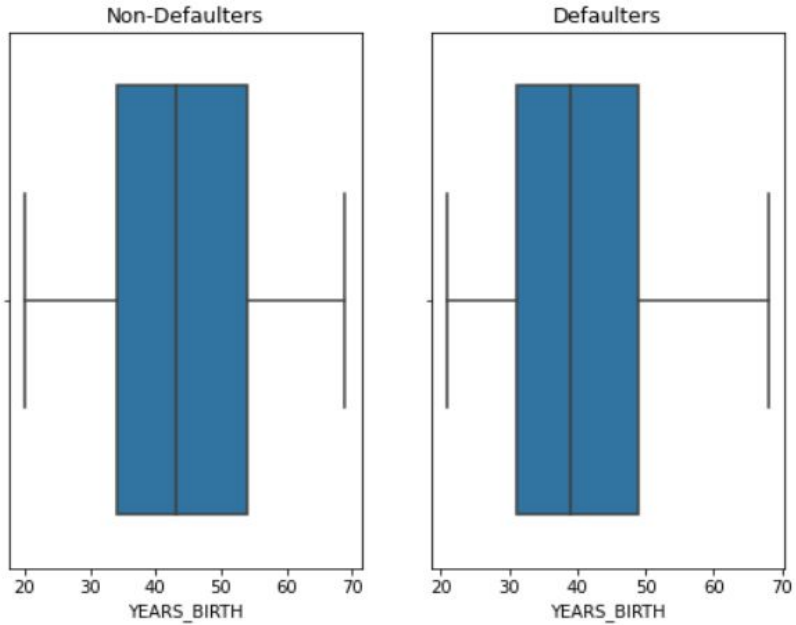


Majority of the loans are taken by the married clients.

This could be because married clients have dual source of income and are also capable to repay the loan.

Univariate Analysis of Numerical Columns

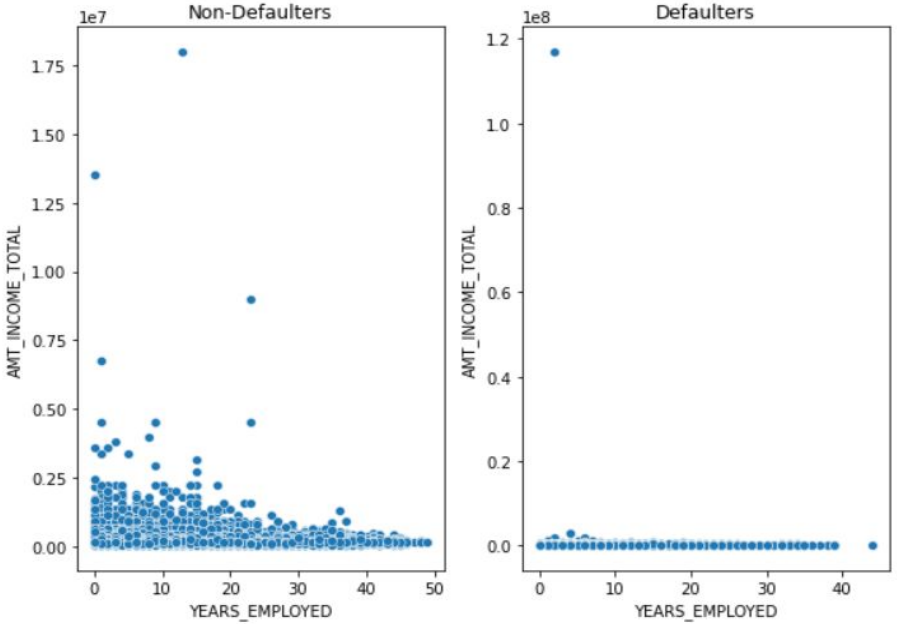
Years Birth Column



We can note that Non-Defaulters are between 34 to 54 years whereas Defaulters are between 31 to 48 years.

Bivariate Analysis

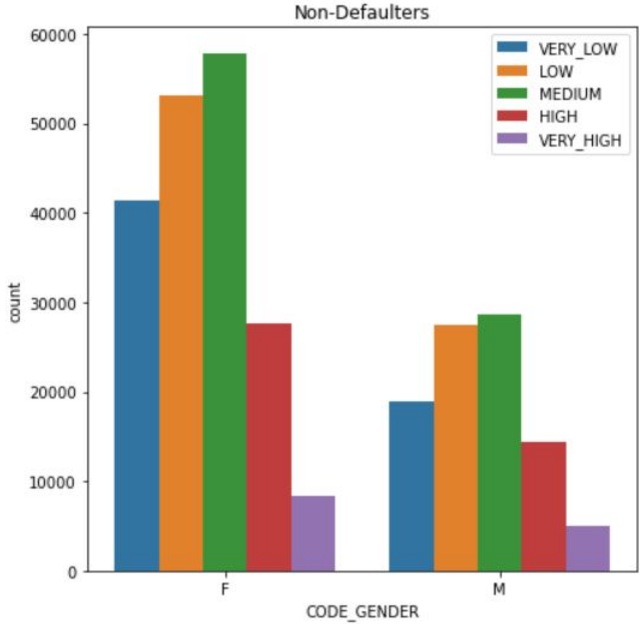
Years Employed vs Total Income

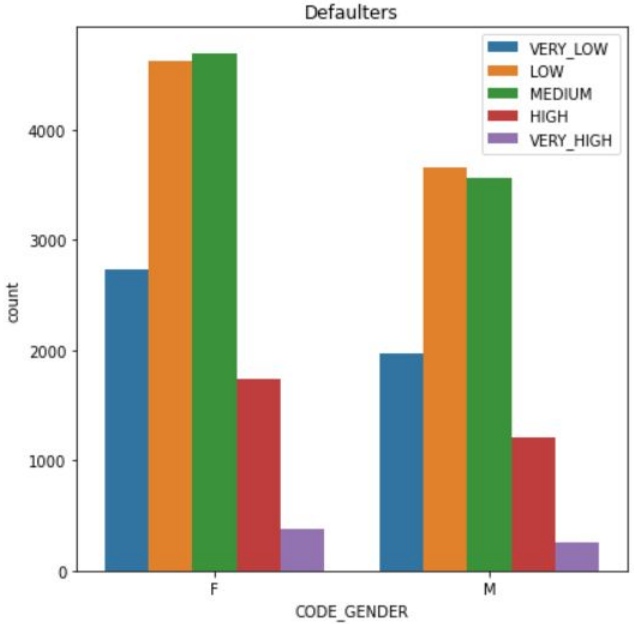


We can note that there is no correlation between Years Employed and Total Income.

Bivariate Analysis

Gender vs Credit

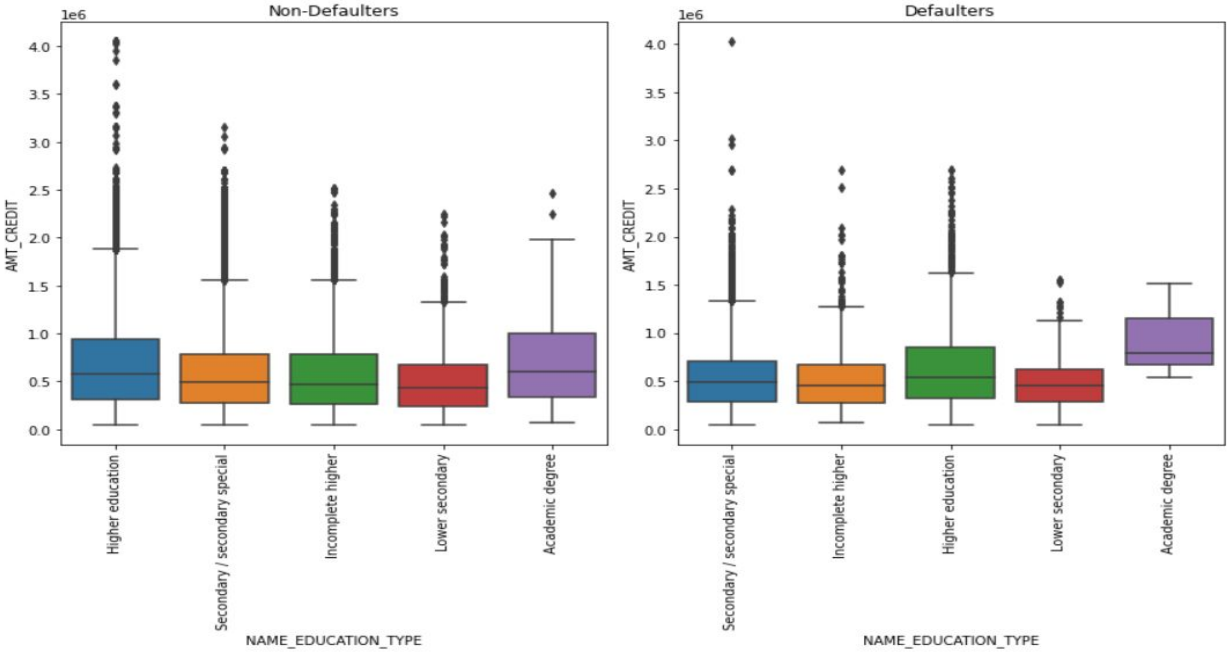




We can see that Females take more loans and also have high credit amount in both the cases when compared to Males.

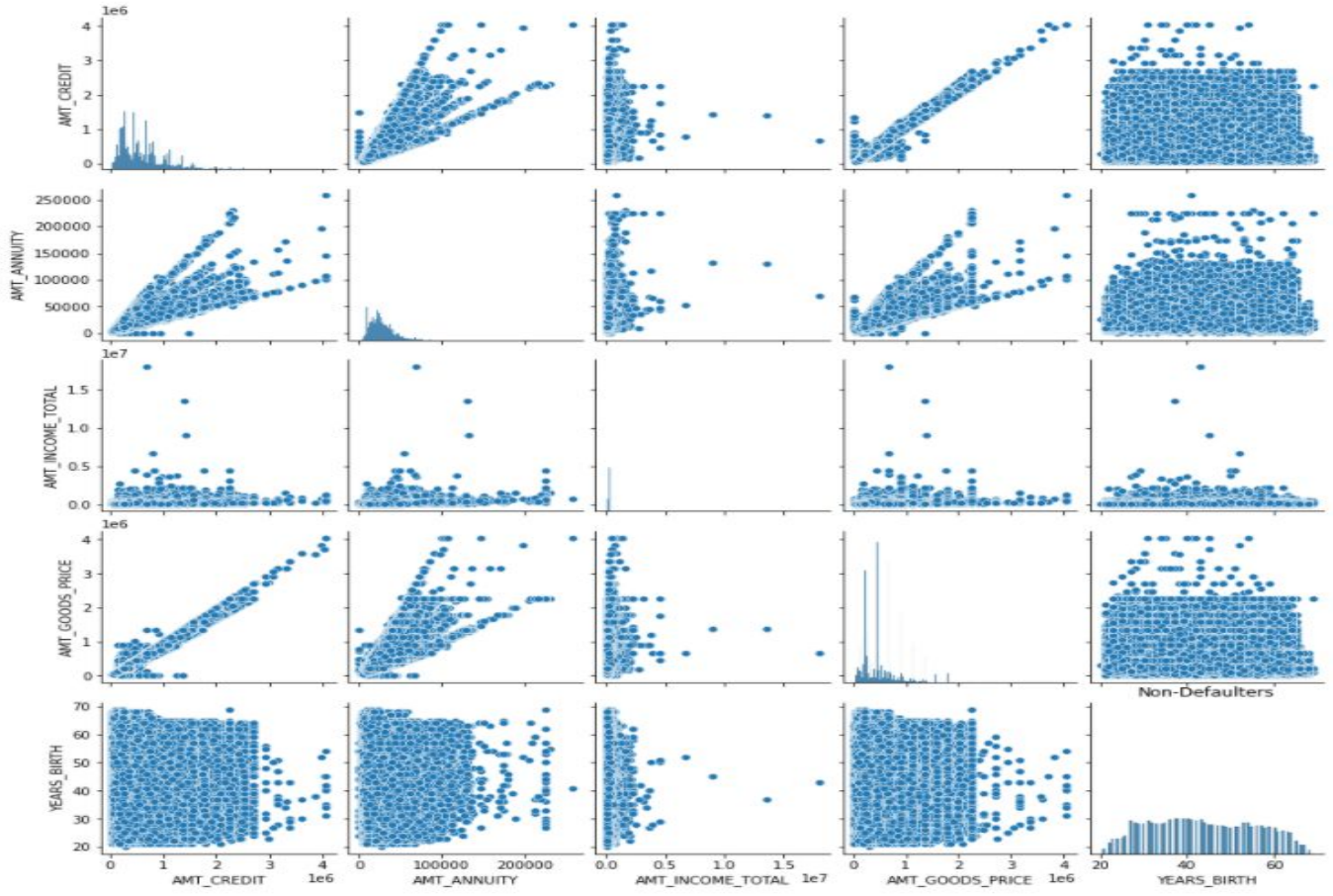
Bivariate Analysis

Credit vs Education Type

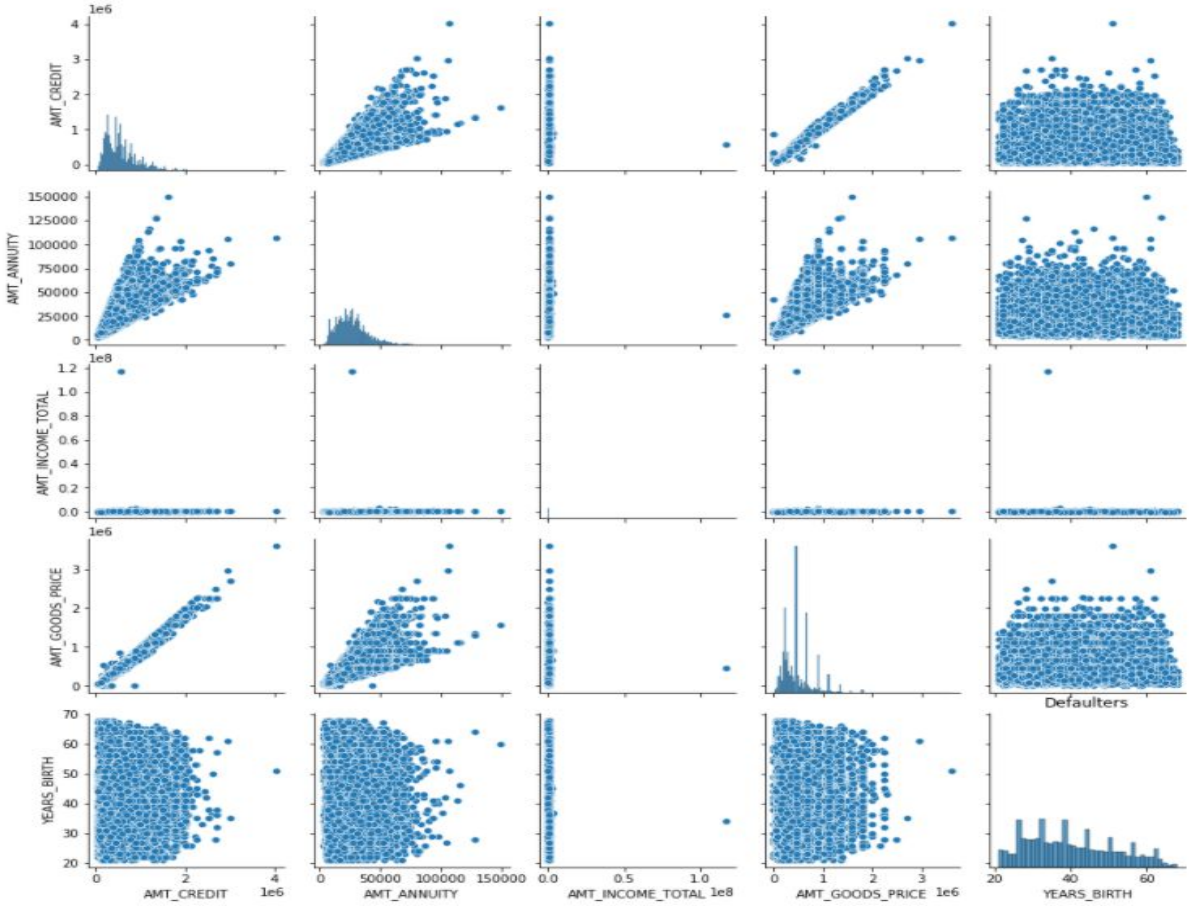


We can see that clients with Academic degree are more likely to default the loan payment.

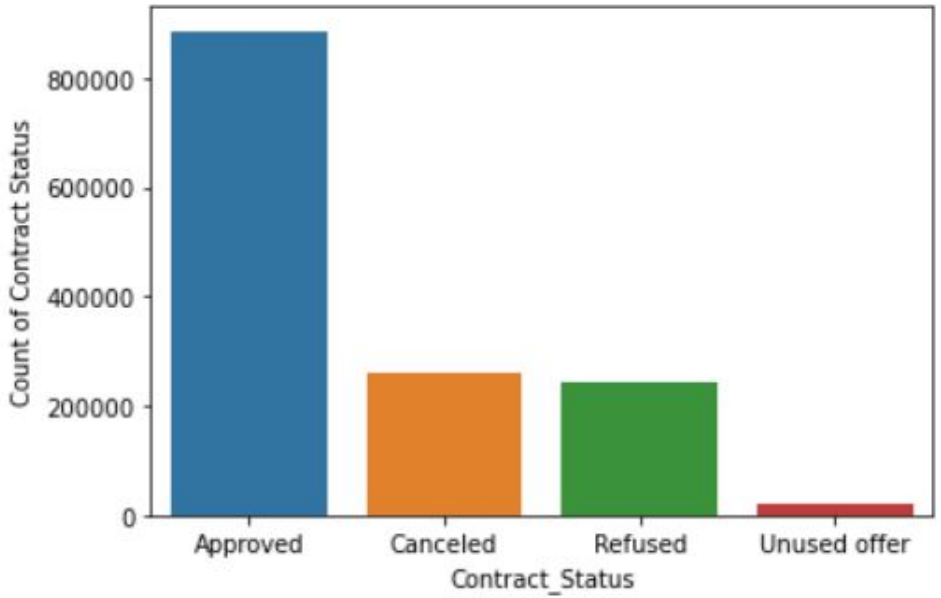
Pair Plot for Target - Non Defaulters



Pair Plot for Target - Defaulters

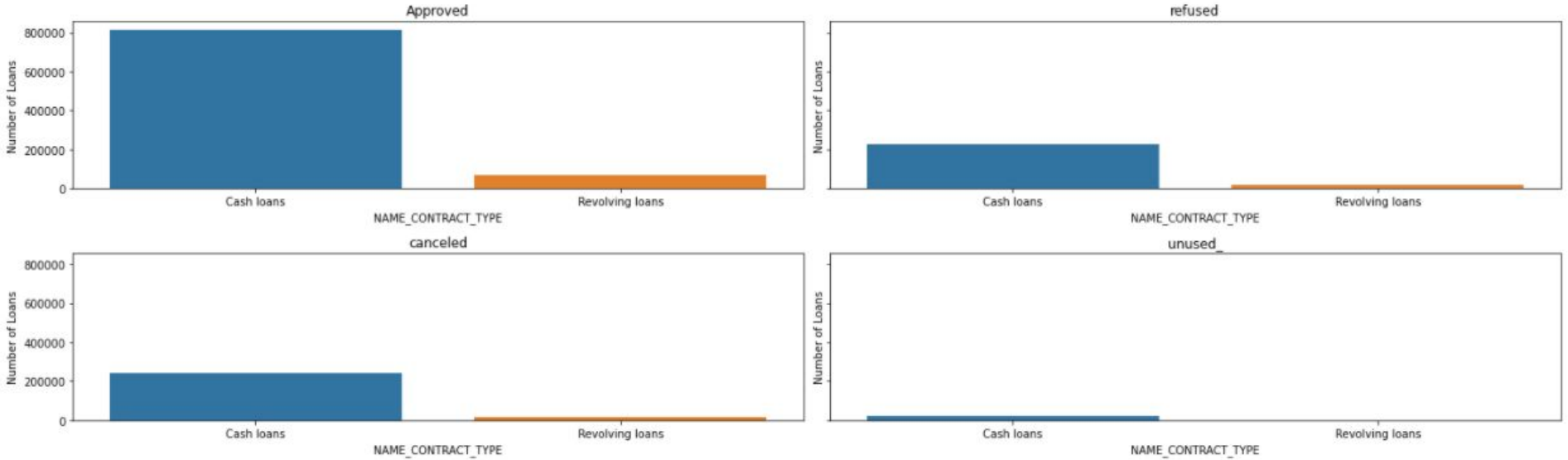


Merged Data Analysis



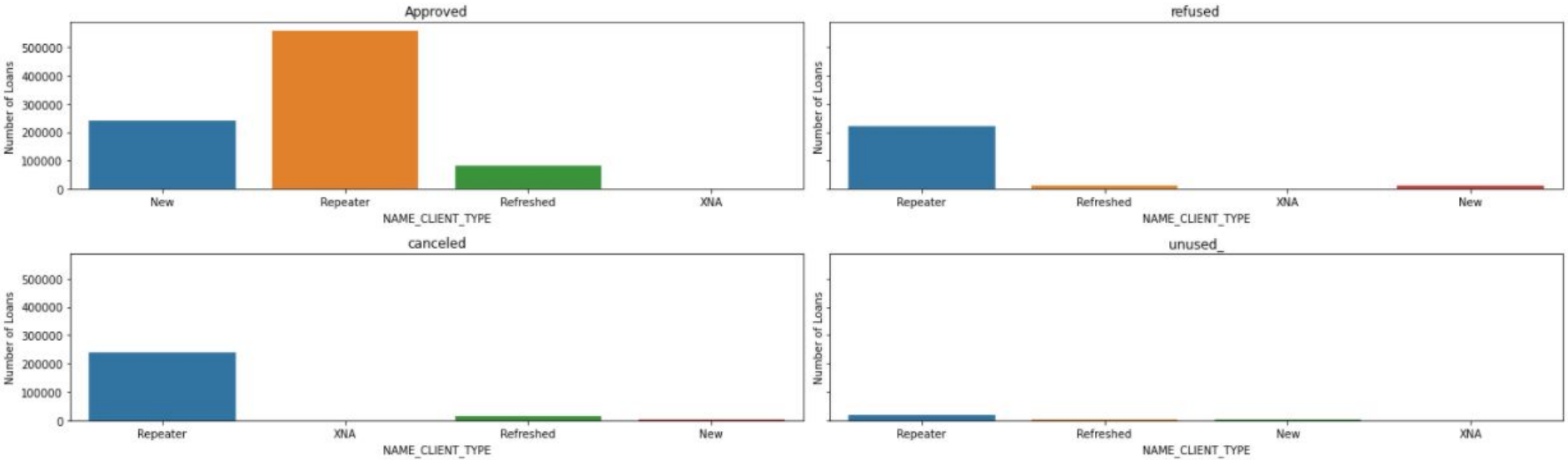
This is the bar graph after merging both the datasets i.e Application Dataset and Previous Dataset

Contract Status vs Contract Type



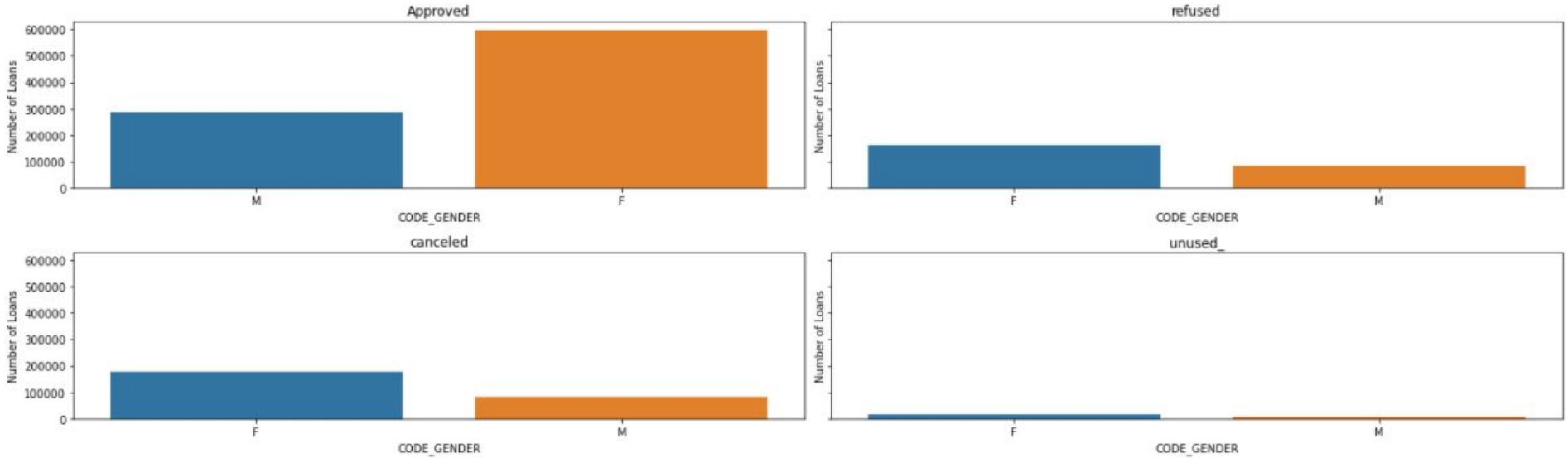
We can see that there are more approvals for Cash loans.

Contract Status vs Client Type



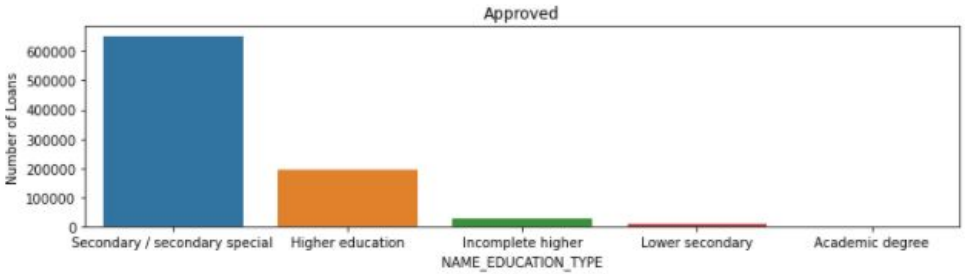
We can see that Repeater gets the most approvals.

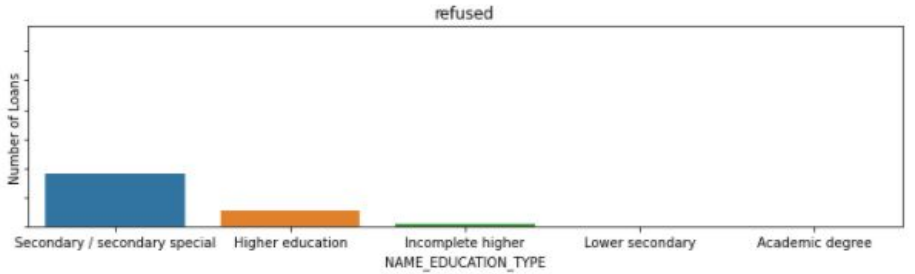
Contract Status vs Gender

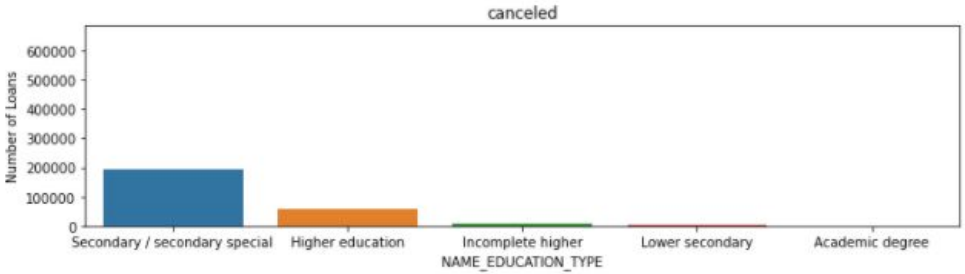


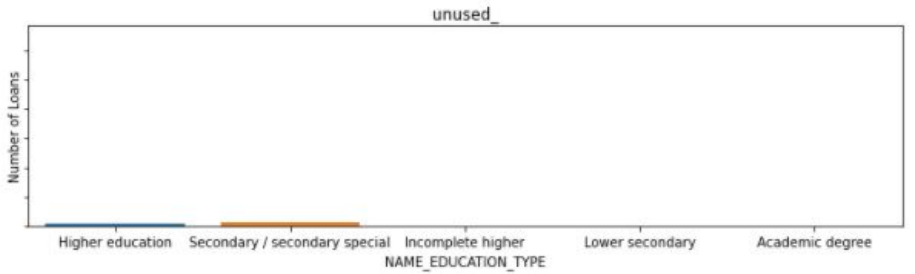
We can see that Females gets the most approvals.

Contract Status vs Education



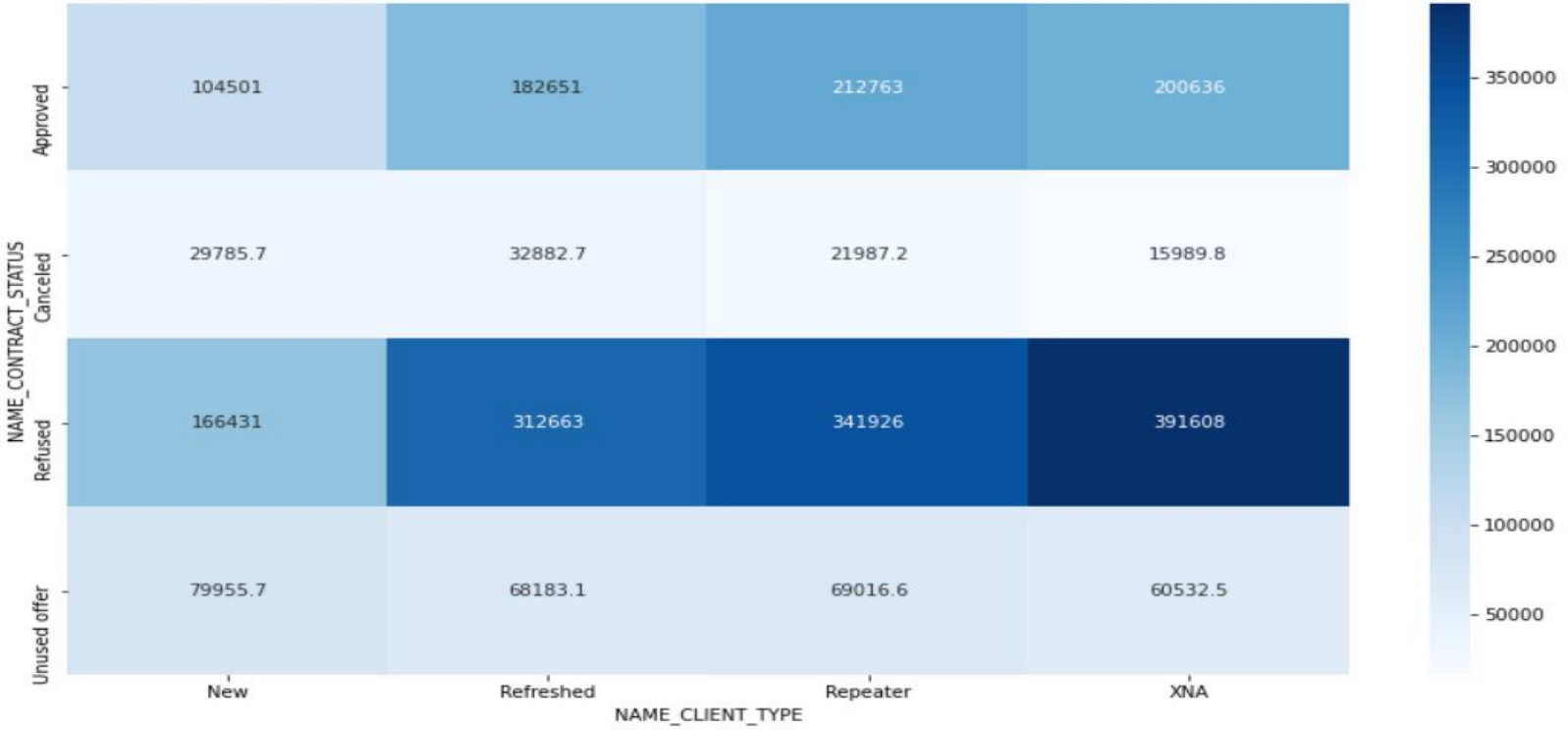




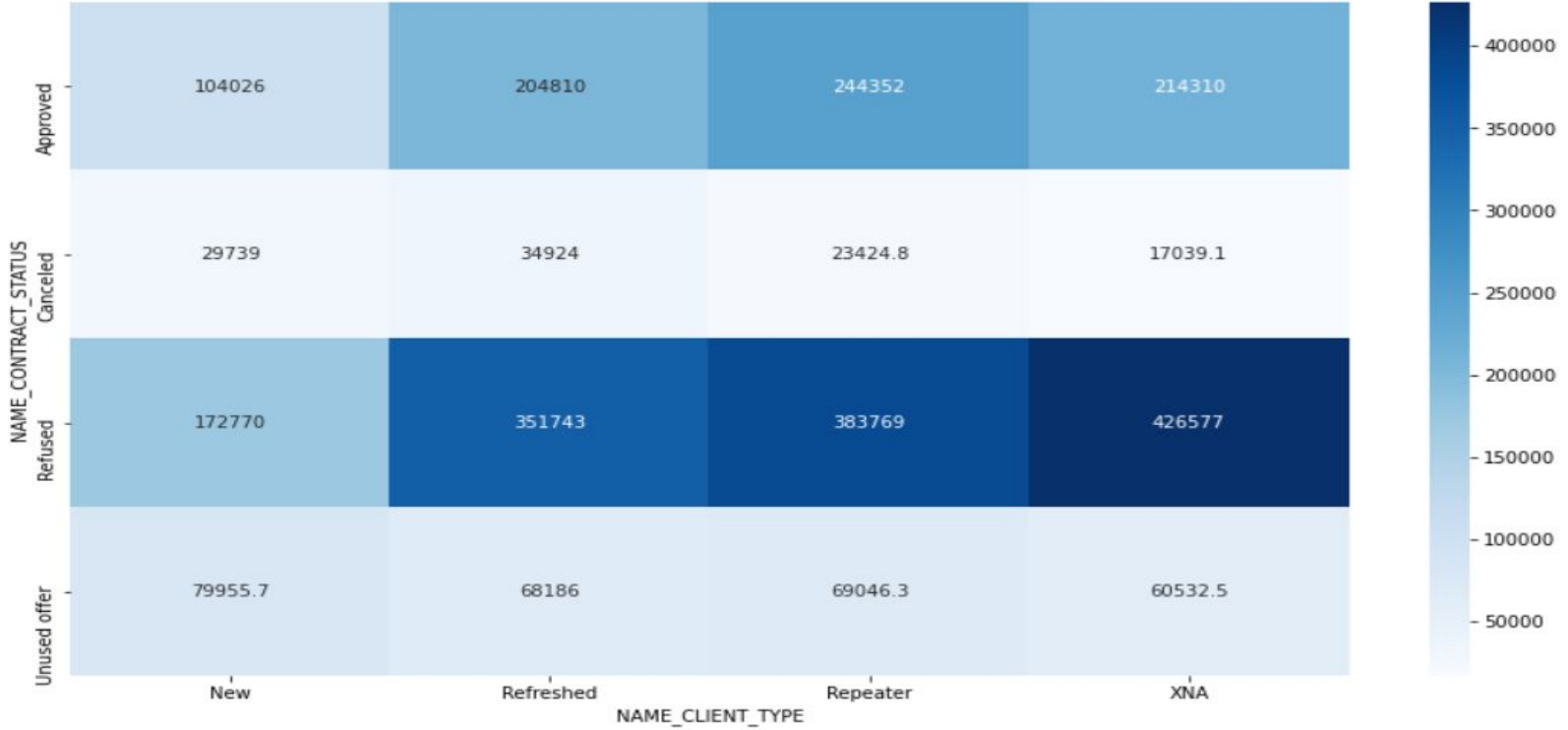


We can see that Secondary gets the most approvals.

Multivariate Analysis - Contract Status/Client Type/Application Amount



Multivariate Analysis - Contract Status/Client Type/Credit Amount



Conclusion

- Banks should focus more on education type 'Higher education' and avoid Secondary/secondary special, incomplete higher or lower secondary as they face paying difficulties.
- Avoid income type of 'Working' clients as they have high percentage of paying difficulties. Instead focus on Commercial associate, pensioner and State servant.
- Focus on clients from housing type ‘House/apartment’ as they are having less paying difficulties.
- Banks should focus on the client from age group of 41 to 70 as they will be financial stable and shows less paying difficulties.