PRESENTATION ON CREDIT EDA CASE STUDY

MEGHANA REDDY
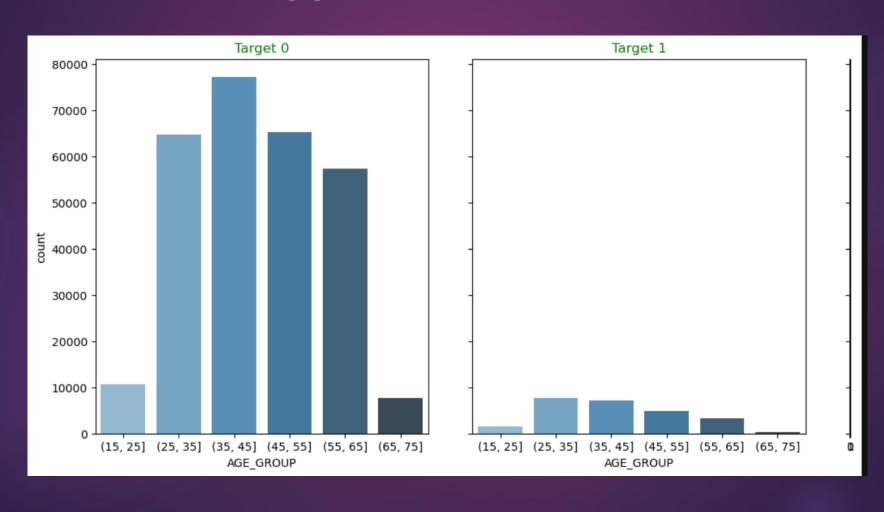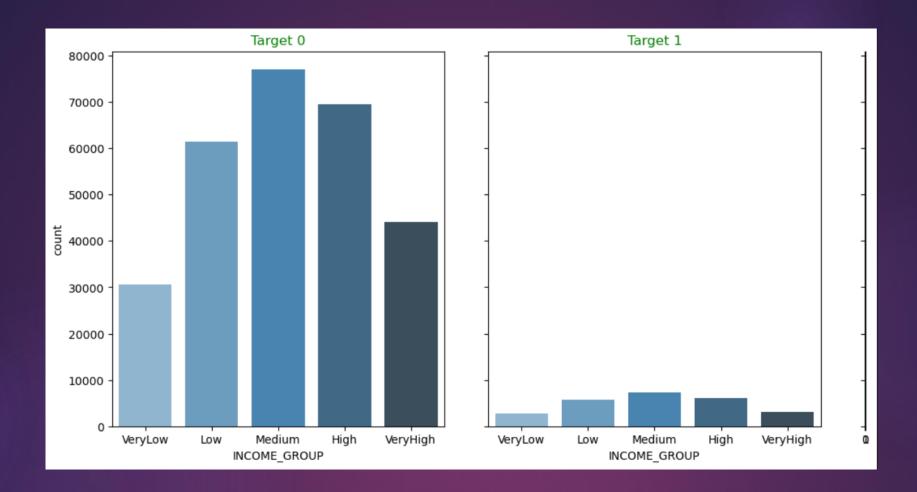
Purpose

▶ This case study aims to identify patterns which indicate if a client has difficulty paying their installment 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. In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.

Steps

- Data Loading and Inspection on Application data.

- Data Cleaning and Handling missing values.

- Data analysis by Univariate, Data correction & standardization, Bivariate, Correlation Analysis.

- Data Loading and Inspection on Previous Application.

- Data Cleaning, Data analysis by Univariate, Bivariate, correlation, Multivariate Analysis.

- Merged two Data frames.

- Case Summary

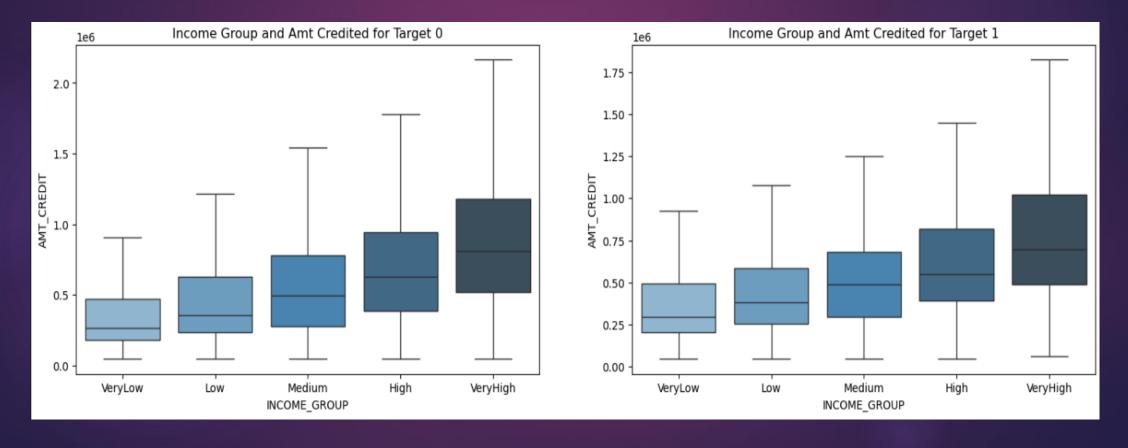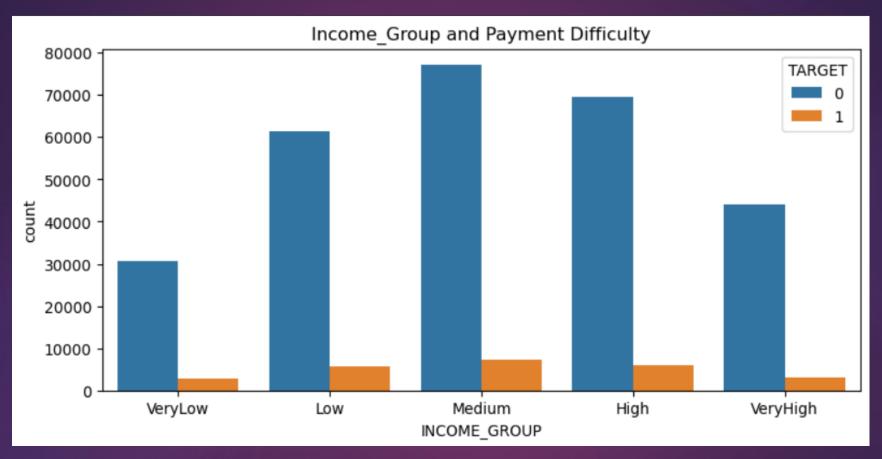# Univariate Analysis of continuous data in Application data

1. AGE_GROUPUP - 35-45 are more in TARGET 0. In Target 1 - 25-25 have higher share. Age does seem like influencing default.
2. INCOME_GROUP - Medium income group have more count in Target 0 and Target 1

Bivariate Analysis on Categorical and Continuous Variable
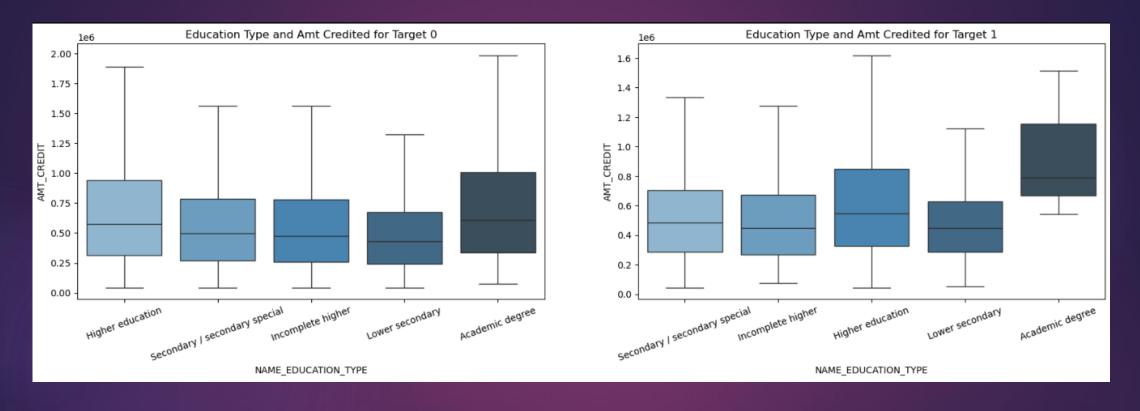
▶ INCOME_GROUP VS AMT_CREDIT for both data frames

Income Group and Payment Difficulty

Income_Group and Payment Difficulty

1.We can infer that though the maximum no of loans is given to Medium income group. Default value per loan is highest in High income group as the AMT_CREDIT is higher too. The loan book of the financial institution can get affected due to higher amount not being paid back.
2.The company must devise a different set of rules and policies while approving higher income group loans.

Bivariate Categoric to Continuous

1. Median of Loan values defaulting for Applicants with Academic degree is higher. But as we saw in a plot above, no of applicants with academic degree is miniscule
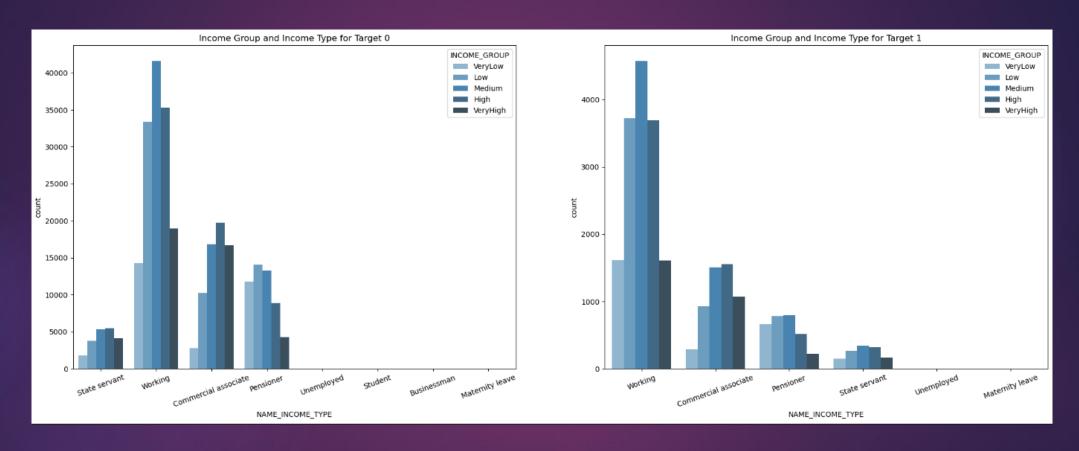2. No inference can be drawn from this analysis.

Bivariate categorical and categorical

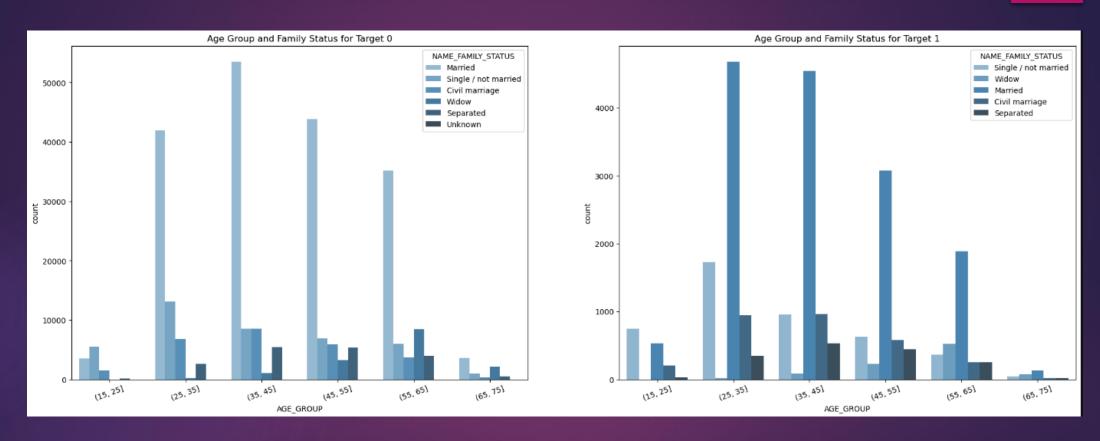1.As noted above data has more females as loan applicant.
2.As seen in plot above, though male applicants are lower, ratio of male applicants defaulting is higher.
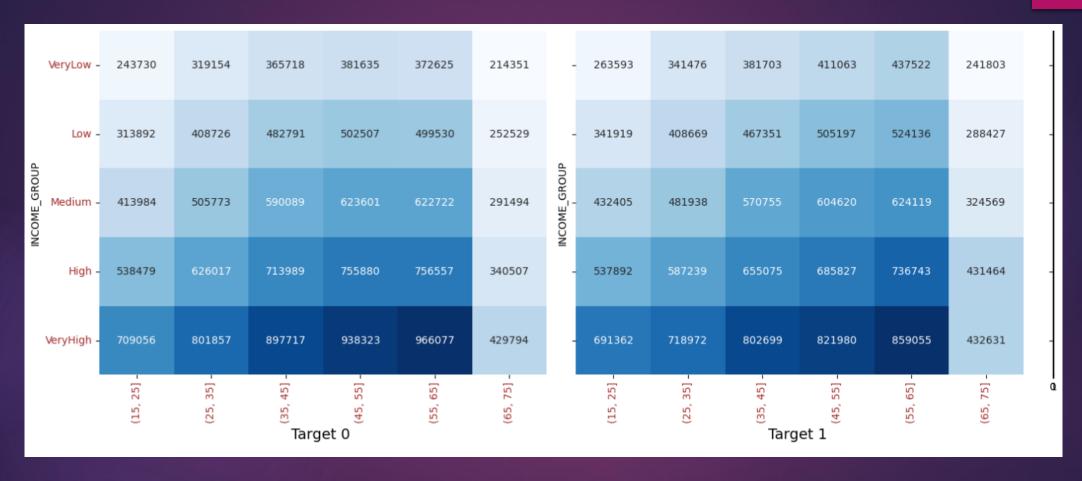
Income Group and Income Type

Medium income group with income type has almost 1 in 12 defaults. Higher than the avg 1 in 11 defaults.
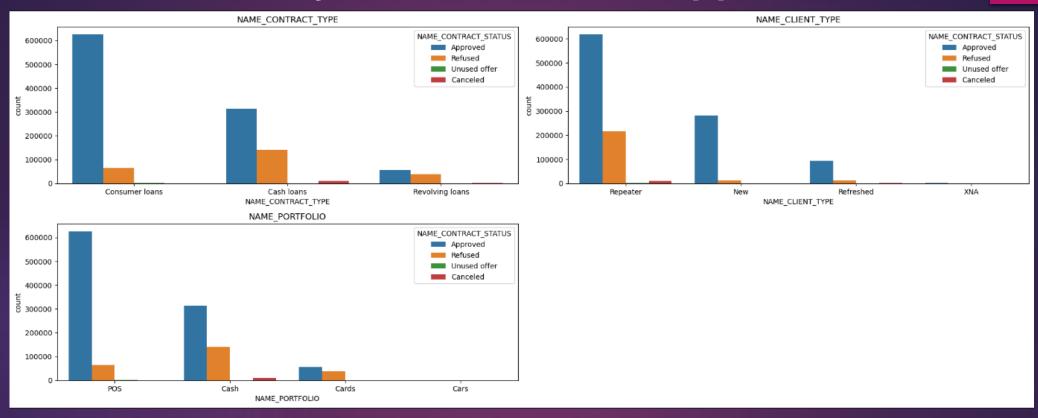
Age Group and Family Status

Married applicant in the age group 25-35 and 35-45 is the largest group of applicant with payment difficulties.
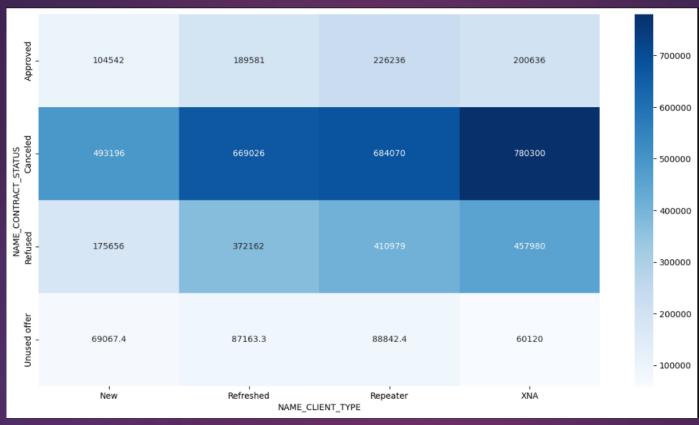
Correlation

Age Group 55-65 in Very High income group has high amount credit. As explained above, this could result as loss in loan book.

Bivariate Analysis on Previous Application

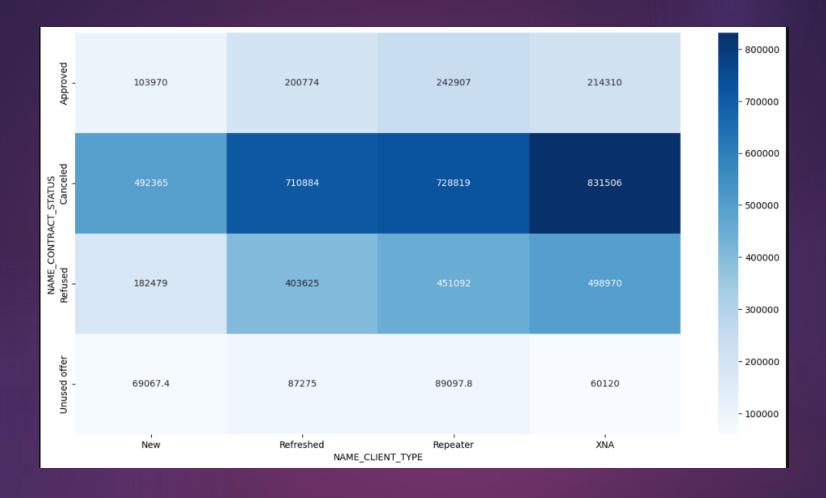1. More cash loans have been refused than consumer loans.
2. The bank has more repeaters in all approved, refused, unused, cancelled categories.
3. POS transactions seem to be consumer loans and similar to point 2 - more cash loans have been refused than POS.
4. There seem to be no cancelled loans in cash loan category than consumer loan.
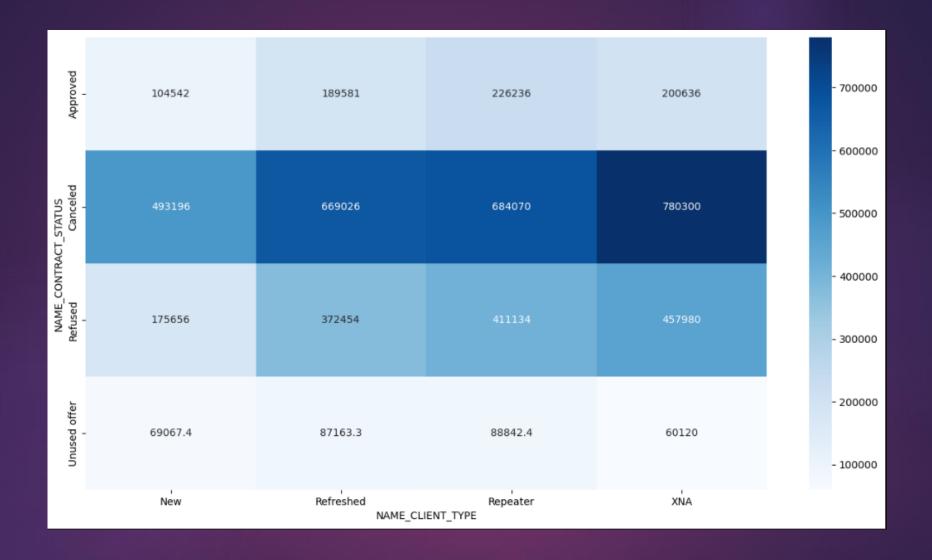5. In approved category, consumer loan has largest no of applicants.

Multivariate Analysis on Previous Application

1.Unused offer application amount is low.
2.Cancelled application amount is high. The bank may be refusing these possibly as the debt liability ratio of consumer must be going high due to the high amount and thus credit default risk.
3.Repeater's application amount is highest than the New customers. This may indicate that the back has more conducive policies/rate of interest etc for repeat applicants
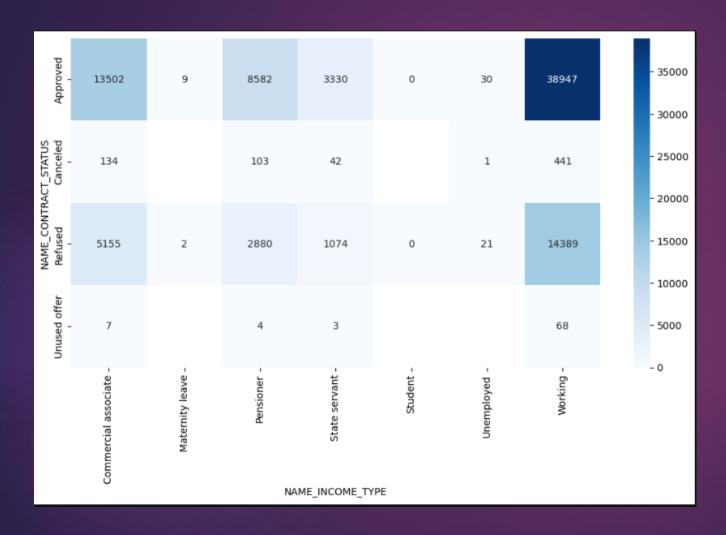
1.Unused offer CREDIT AMOUNT is low. This may be the reason for customer not using.
2.Unable to understand why cancelled and refused there should be any credit amount?
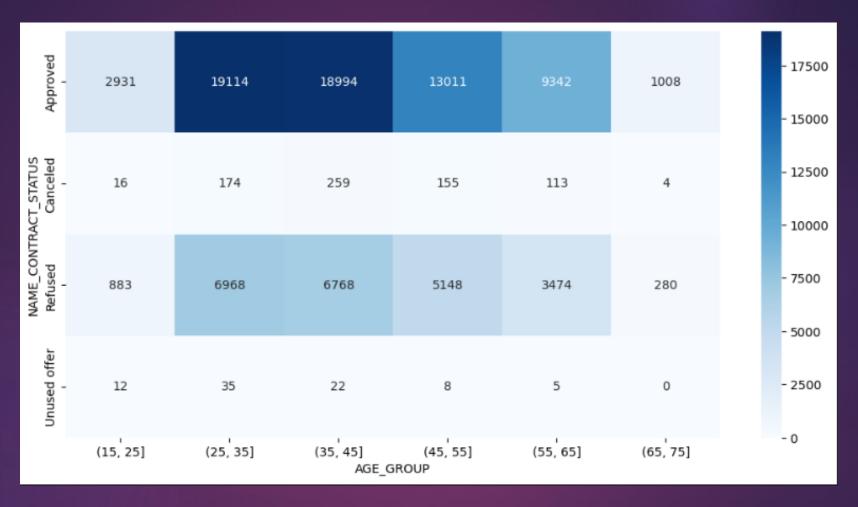
1.All cancelled and refused cases have higher value of goods than other categories

Merged Data frame

1. Since Target 1 is default, higher on the above matrix shows correlation to default.
2. Working applicant with Approved status have defaulted in highest numbers.
3. Previous applications with Refused, Cancelled, Unused loans also have default which is a matter of concern. This indicates that the financial company had Refused/cancelled previous application, but has approved the current and is facing default on these loans.
4. 14,389 applicants of working class were REFUSED earlier and now have defaulted

1.Since Target 1 is default, higher on the above matrix shows correlation to default.
2.Approved loans of age group 25-35 and 35-45 have higher defaults.
3.Refused, cancelled, loans in previous application have defaulted in current.

1. Higher credit offered to unemployed, maternity leave is a notable factor.
2. Unused offers have smaller credit values and possibly the reason why applicant is not using them.

Case Summary
Defaulters demography

▶ All the below variables were established in analysis of Application data frame as leading to default.   Checked these against the Approved loans which have defaults, and it proves to be correct   -Medium income   -25-35 years olds , followed by 35-45 years age group   -Male   -Unemployed   -Labourers, Salesman, Drivers   -Business type 3   -Own House - No   Other IMPORTANT Factors to be considered   -Days last phone number changed - Lower figure points at concern   -No of Bureau Hits in last week. Month etc – zero hits is good   -Amount income not correspondingly equivalent to Good Bought – Income low and good value high is a concern   -Previous applications with Refused, Cancelled, Unused loans also have default which is a matter of concern. This indicates that the financial company had Refused/Cancelled previous application but has approved the current and is facing default on these.

Credible Applications refused

▶ -Unused applications have lower loan amount. Is this the reason for no usage?-Female applicants should be given extra weightage as defaults are lesser.-60% of defaulters are Working applicants. This does not mean working applicants must be refused. Proper scrutiny of other parameters needed-Previous applications with Refused, Cancelled, Unused loans also have cases where payments are coming on time in current application. This indicates that possibly wrong decisions were done in those cases.