Lending Club Case Study

Batch: EPGML C55 July 2023

Group Members:

Mathangi M (mathangimr@gmail.com)

Kiran Prasad (kiranp37@gmail.com)

Problem Statement

Given the data provided about past loan applicants and whether they 'defaulted' or not, come up with an understanding on how various

- consumer attributes and
- loan attributes

influence the tendency of default.

Analysis Approach

Approach followed: Exploratory Data Analysis, including following steps:

- Data understanding
- Data cleansing & manipulation
- Derived metrics generation
- Univariate & segmented univariate analysis
- Bivariate analysis

Metrics

- Grade and sub-grade are type driven metrics which belong to ordinal variables.

- Employment length falls under internal variables as they follow a certain order but doesn't help us in any multiplication and division operations against defaulters.

- Based on the income type, we can determine the amount which comes under each income range which belongs to a derived metric calculation.

Observations

- Average annual income of defaulters are mainly less than 65000.

- Average dti is 14 or above for defaulters.

- Defaulted loans have average principal amount more than 12000.

- Majority of defaulters have rented apartments or their home under mortgage.

- More than 90% of defaulters belongs to low income group (<5 lakhs).

- Average pay ratio is higher than 1.275 for defaulters and this is very high compared to fully paid loans.

- There is a lot of defaulters who have taken loan amounts between 5000 to 10000.

- If the interest rate is between 10-15%, then the defaulters number increases.

- Large number of defaulters fall in the range around 10+ years while there are some of them similarly in the range of 0-2 years.

- Loan Amount and installment has a higher correlation of 0.93. The pay ratio and interest rate has a correlation value of 0.73.

EDA Results and Visualizations – Univariate Analysis
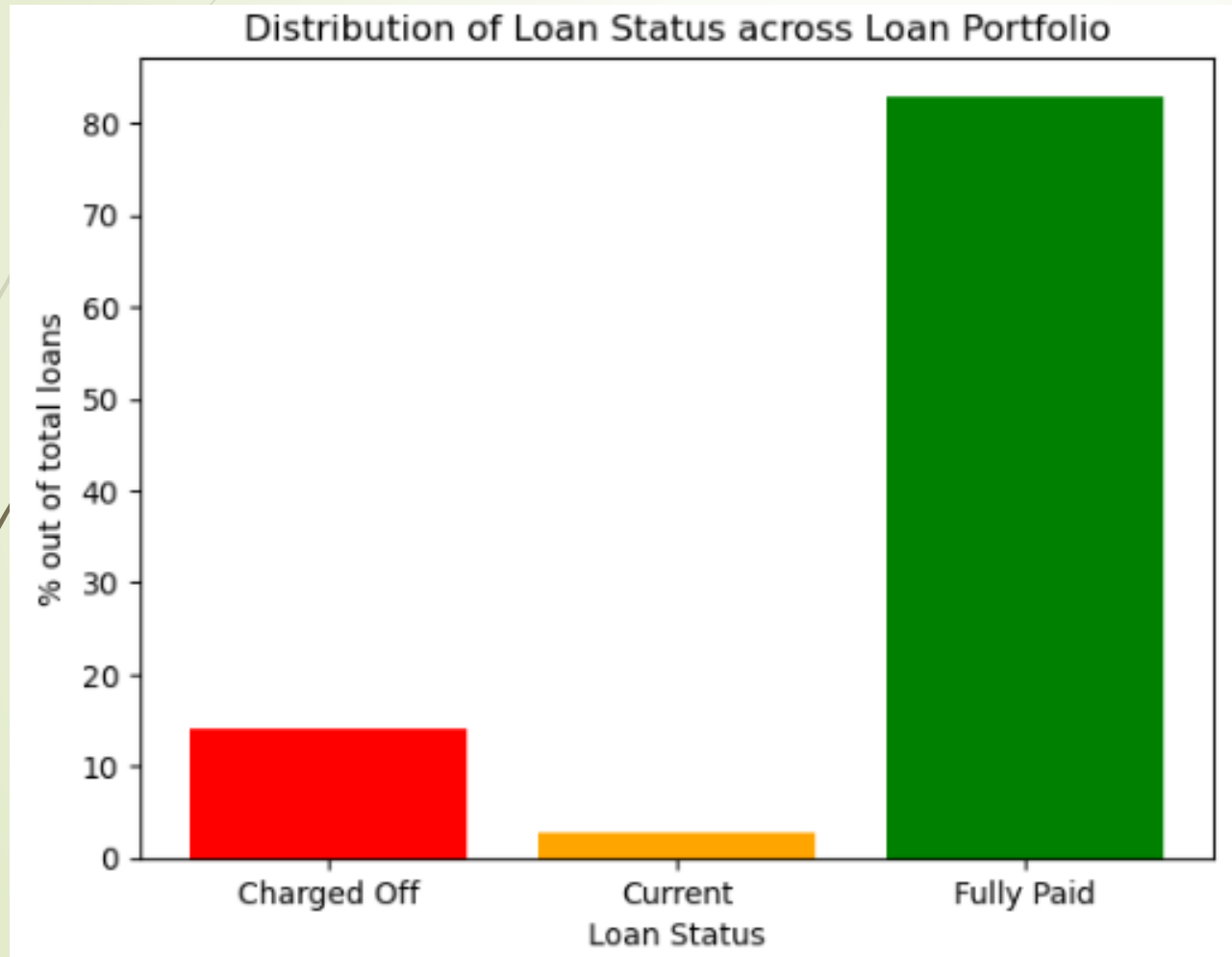
Loan distribution by Loan amount

Loan distribution by Loan amount
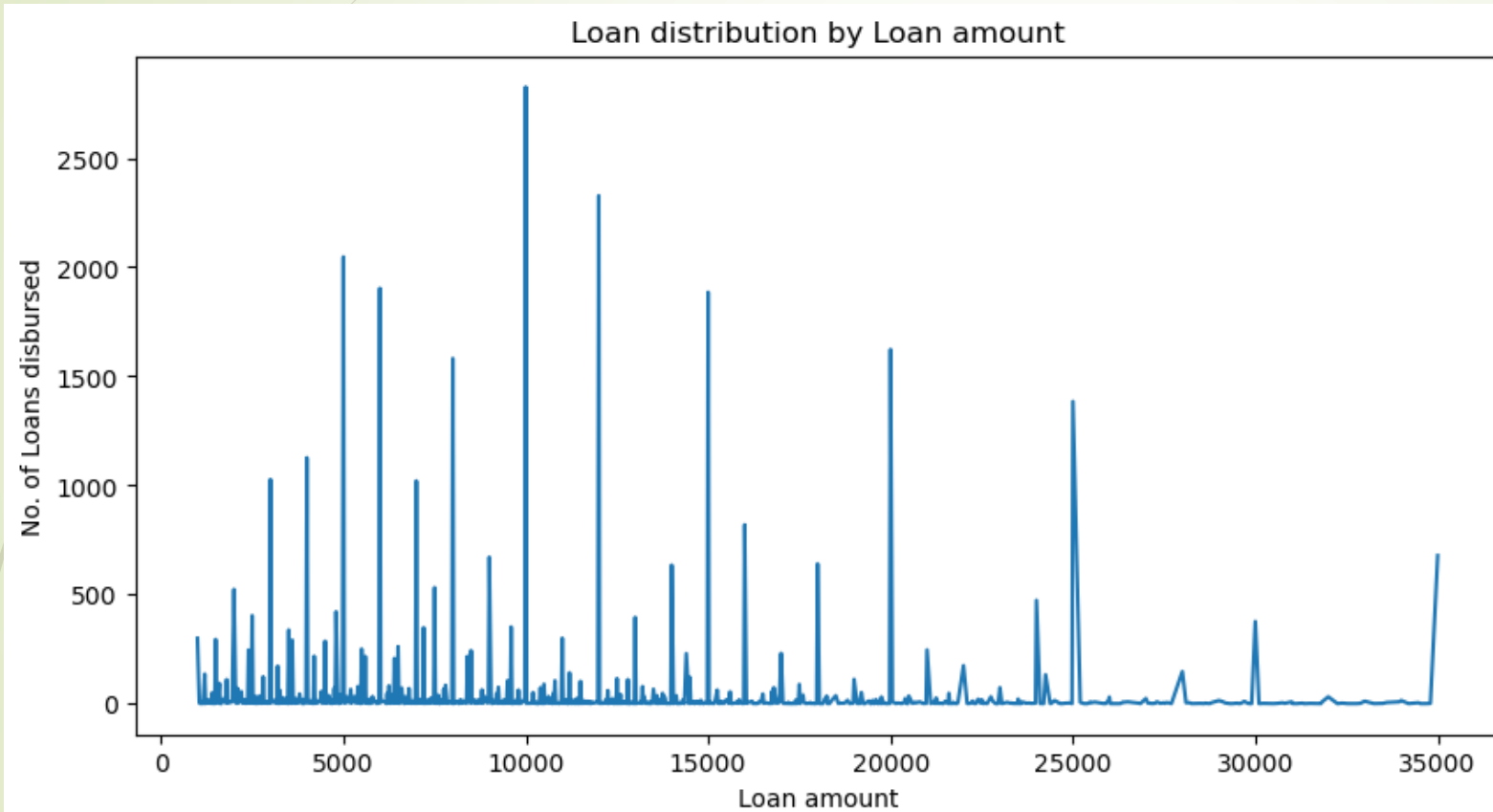
- Maximum loans disbursed are having amounts rounded to multiples of 1000s

- Highest number of loans for a single amount is of amount 10000.

- 50% of loans are ranging between 5000 - 15000
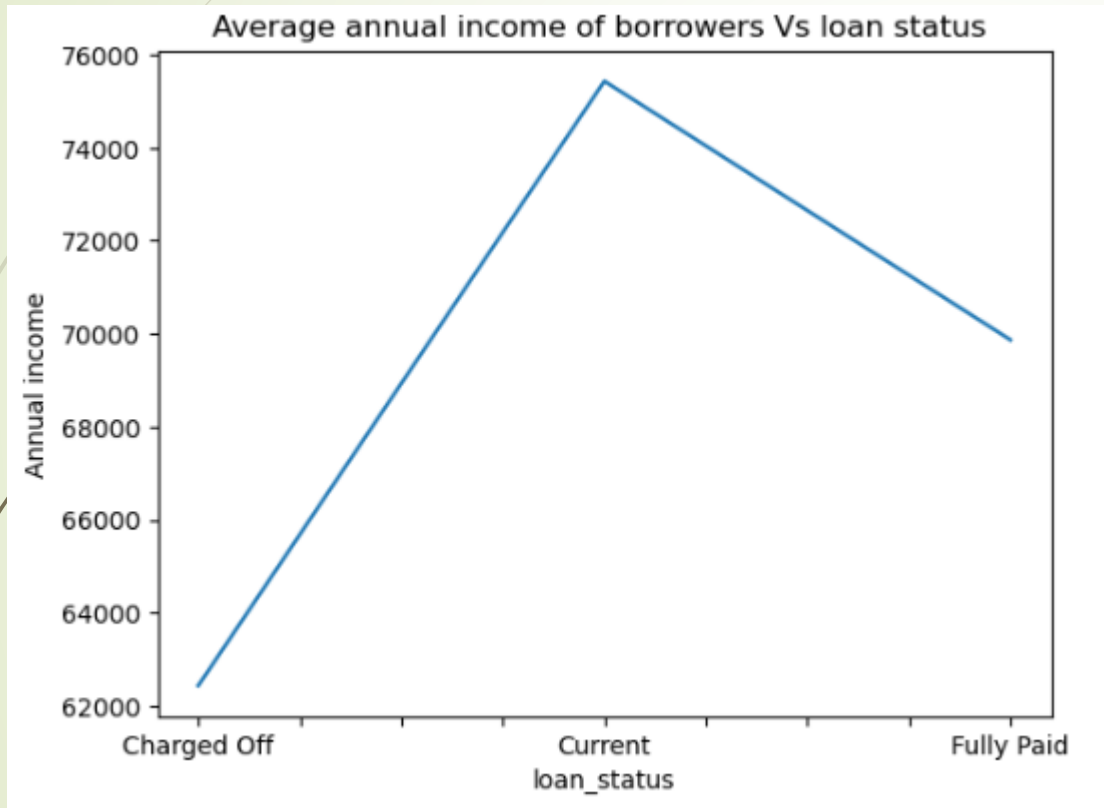
% distribution of Loan Status across the Loan portfolio

Distribution of Loan Status across Loan Portfolio

Around 15% of total loans disbursed are defaulted

Segmented Univariate Analysis

Average annual income of Borrowers Vs. Loan Status

Average annual income of borrowers Vs loan status

(chart: Annual income vs loan_status, with x-axis values Charged Off, Current, Fully Paid; y-axis Annual income from 62000 to 76000)

Data set is segmented based on loan status and average annual income is calculated for each segment.

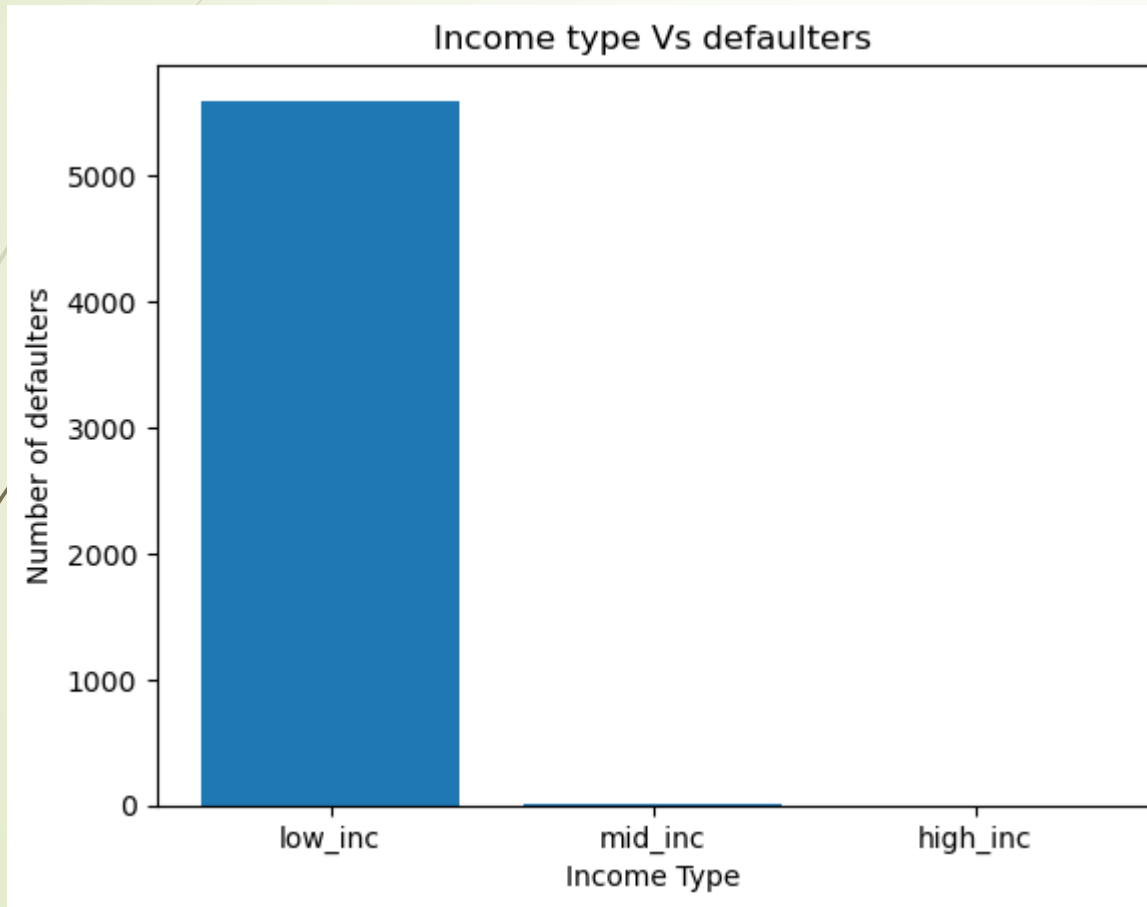All defaulted borrowers are having lower annual income (specifically, below 65000)

There are no defaulters with average annual income is below 70000

This is customer attribute influencing tendency of default

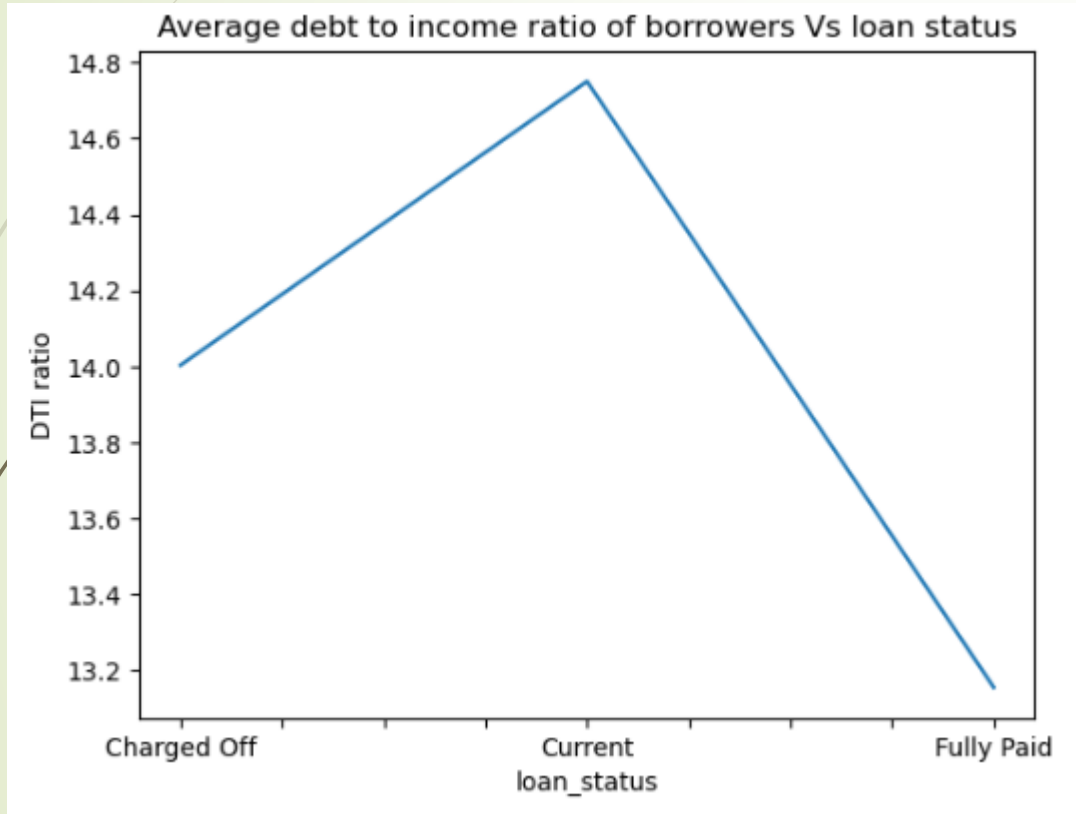# Annual Income of Defaulters
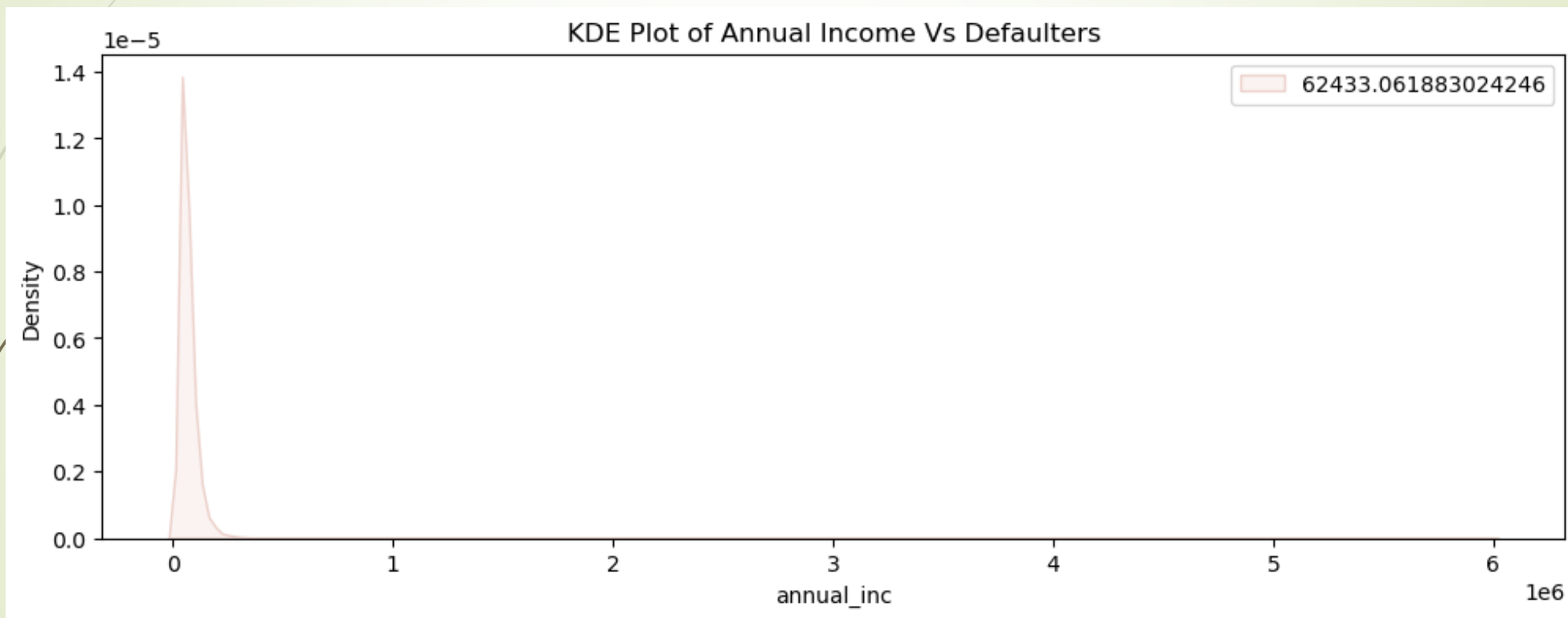
Income Type Vs. Defaulters

Annual income varies between 4000 to 60 lakhs. Hence customers are grouped into 3 categories:
1. low_inc = < 5 lakhs
2. mid_inc = Between 5 & 10 lakhs
3. high_inc = > 10 lakhs

More than 90% f defaulters belong to low income category
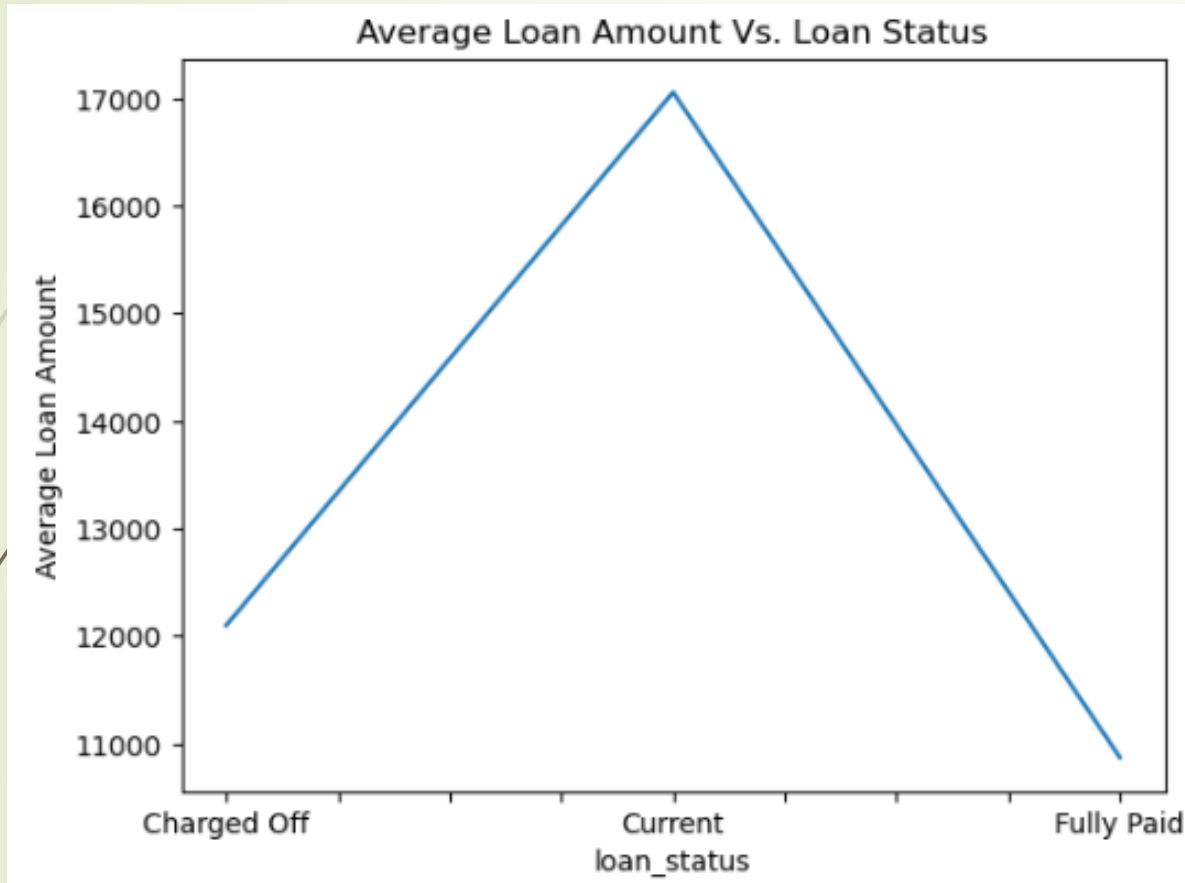
Average debt to income ratio of Borrowers Vs. Loan Status

Data set is segmented based on loan status and average debt to income ratio of borrowers is calculated for each segment.

All defaulted borrowers are having higher debt to income ratio (specifically, above 14)

This is customer attribute influencing tendency of default
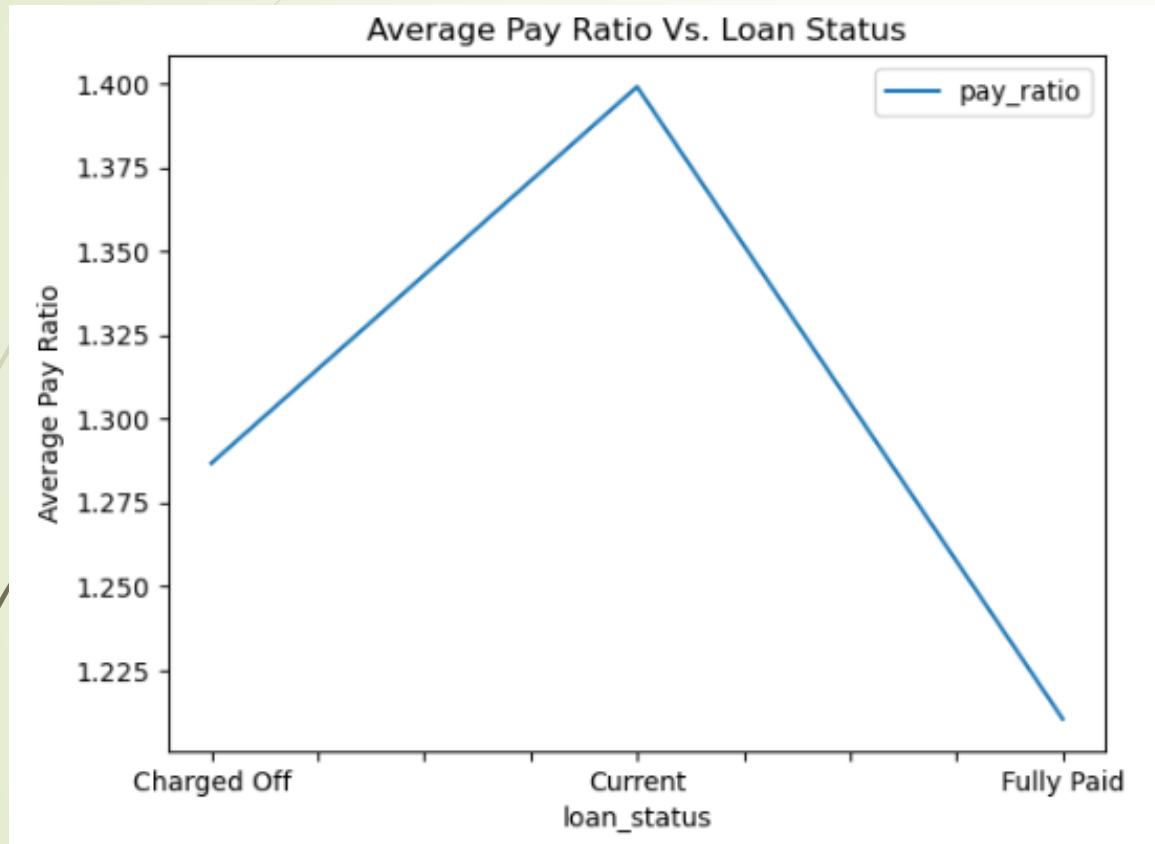
Average Loan Amount Vs. Loan Status

Data set is segmented based on loan status and average loan amount is calculated for each segment.

All defaulted borrowers are having average loan amount above 12000

This is loan attribute influencing tendency of default

Average Pay Ratio Vs. Loan Status

Pay ratio = ratio of total payable amount to loan amount

Data set is segmented based on loan status and average pay ratio is calculated for each segment.

Loans with pay ratio > 1.275 are likely to default
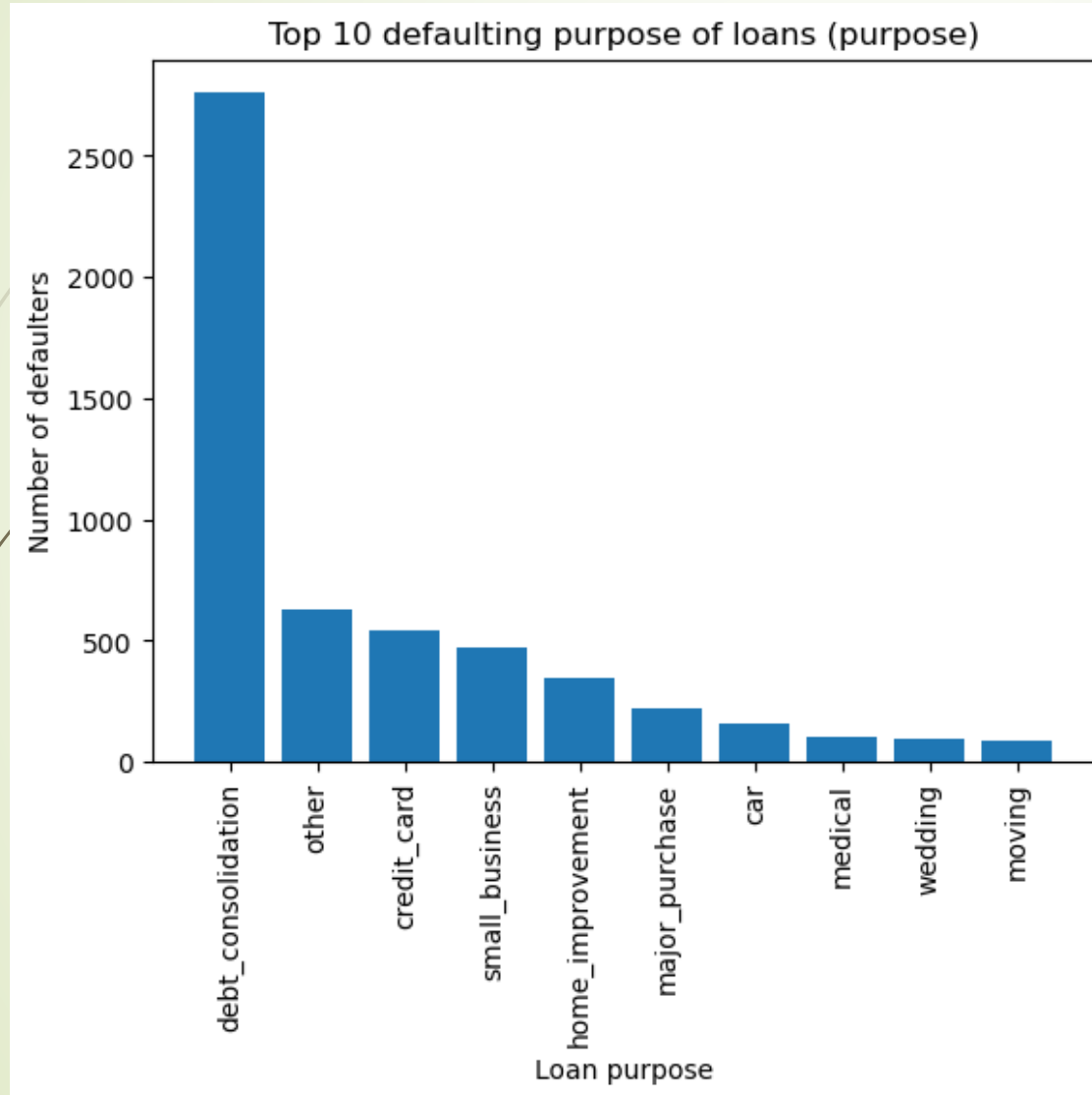
This is loan attribute influencing tendency of default

In above figures, the average values for currently active loans are higher than charged off and fully paid loans. This is because:

- Currently active loans are the least in number compared to others, thus average values goes up.

- All of them are higher term loans (60 months) with high loan amounts

- This also represents good borrowers

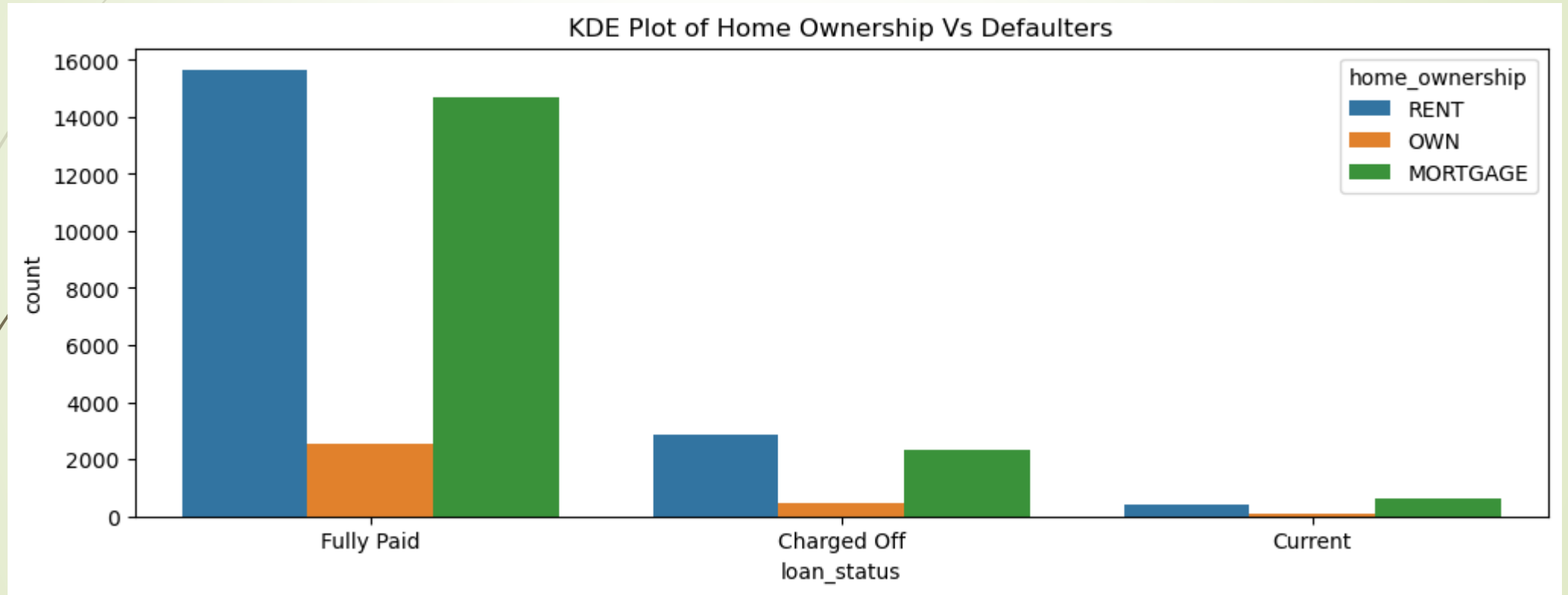# Bivariate Analysis

Top 10 defaulting 'purpose' of loans

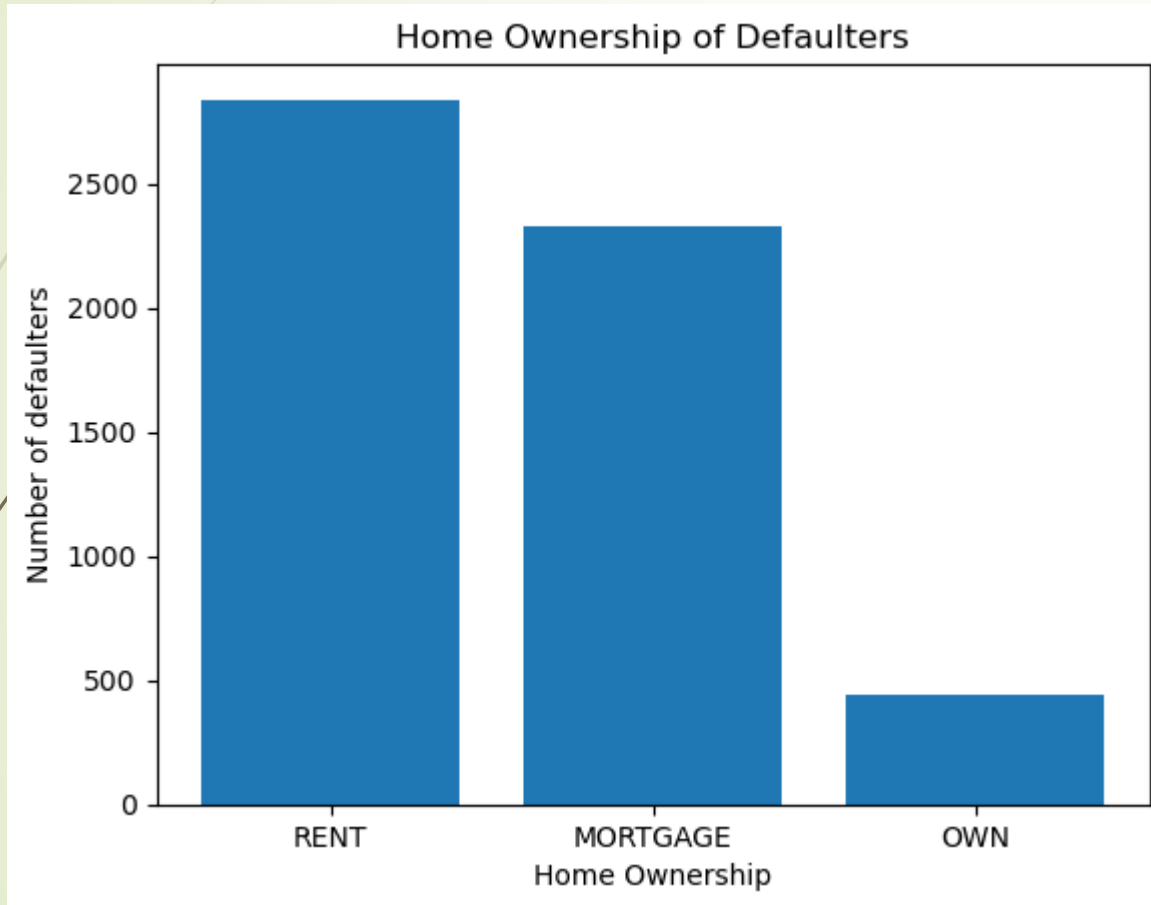Data set is segmented based on loan status and grouped based on the loan purpose.

Top 10 purpose covers more than 80% of total defaulters

This is loan attribute influencing tendency of default

Home ownership of Borrowers across Loan portfolio
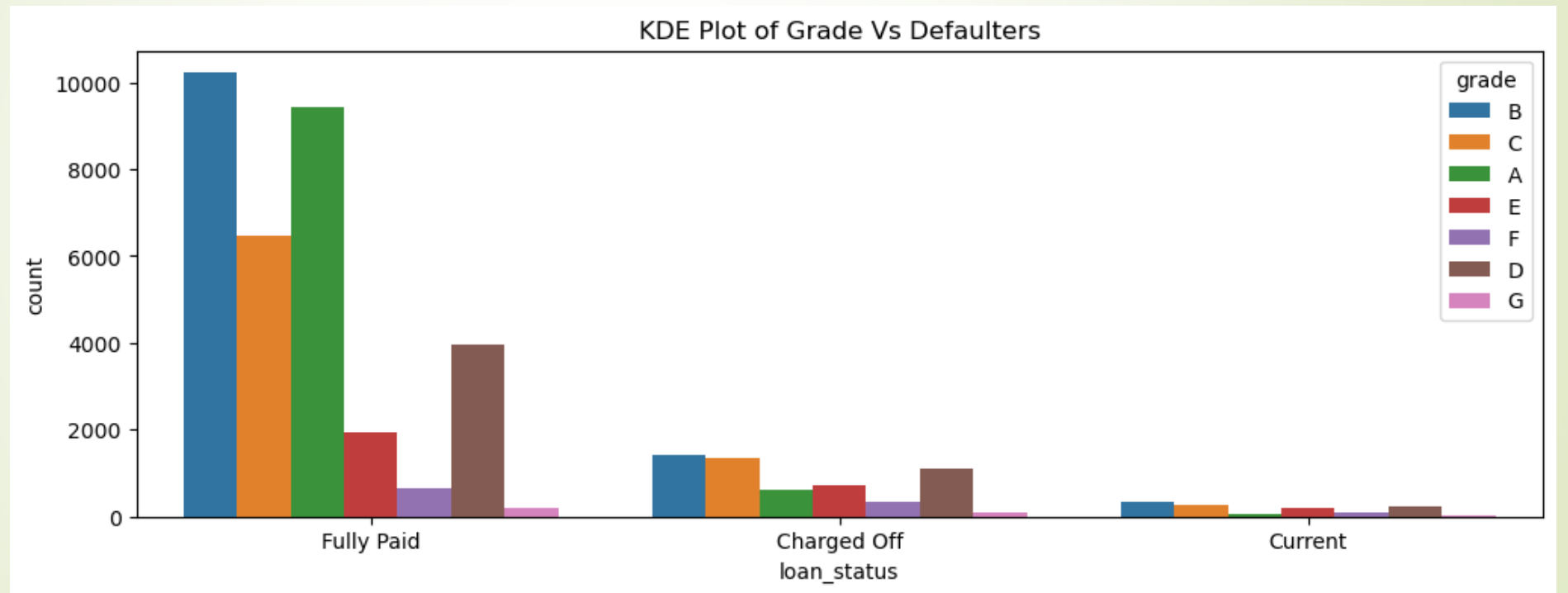
Home ownership of Defaulters

Data set is segmented based on loan status and grouped based on the type of home ownership
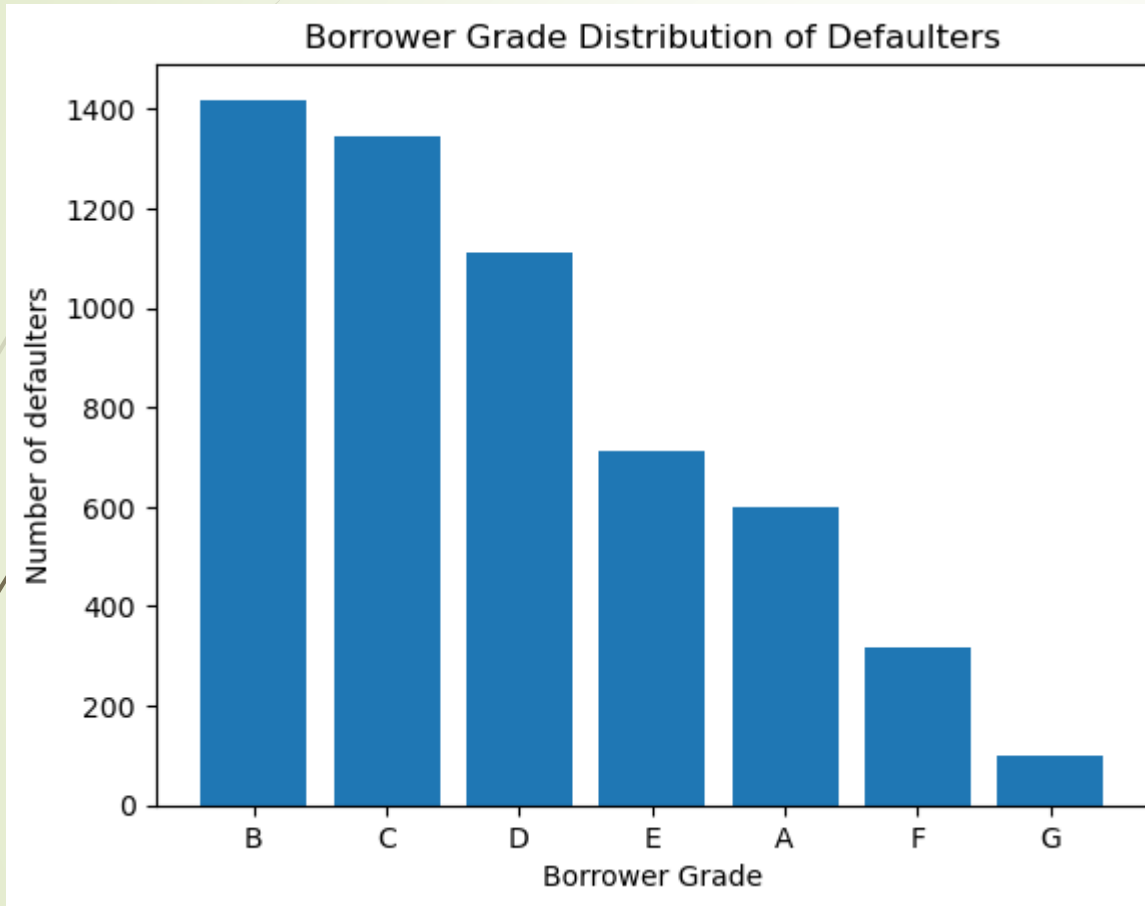
Largest group of defaulters are either on rent or have their homes mortgaged

This is costumer attribute influencing tendency of default

Borrower grade distribution across Loan portfolio

Borrower grade Vs. Defaulters

Borrower Grade Distribution of Defaulters

Data set is segmented based on loan status and grouped based on grade of borrowers

Largest group of defaulters are having grades B, C, D & E

Giving above average grades (B to D) to borrowers without proper evaluation by loan agents for increasing business can be a reason for this

This is customer attribute but decided by lender, influencing tendency of default
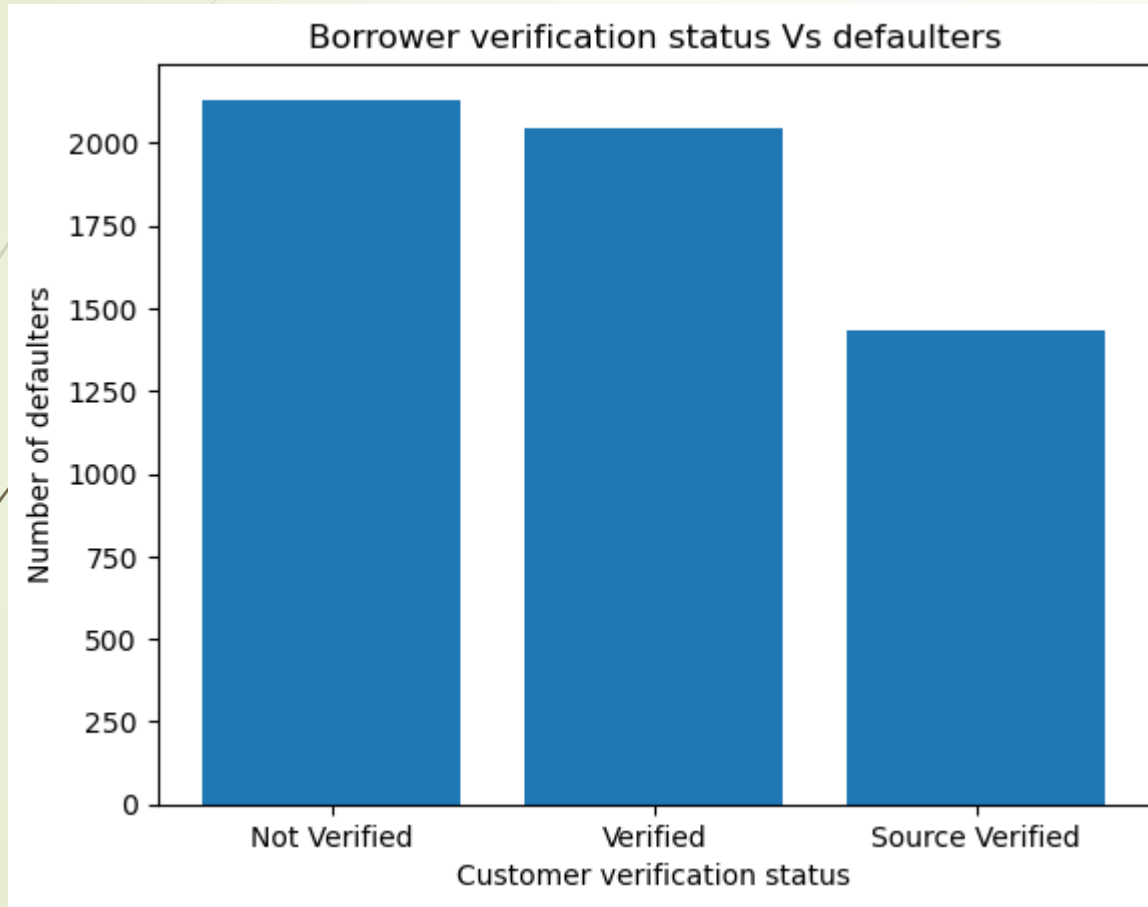
Borrower subgrade distribution for top 5 grades with maximum defaulters

Borrower subgrade distribution for top 5 grades with maximum defaulters

Borrower verification status Vs. Defaulters
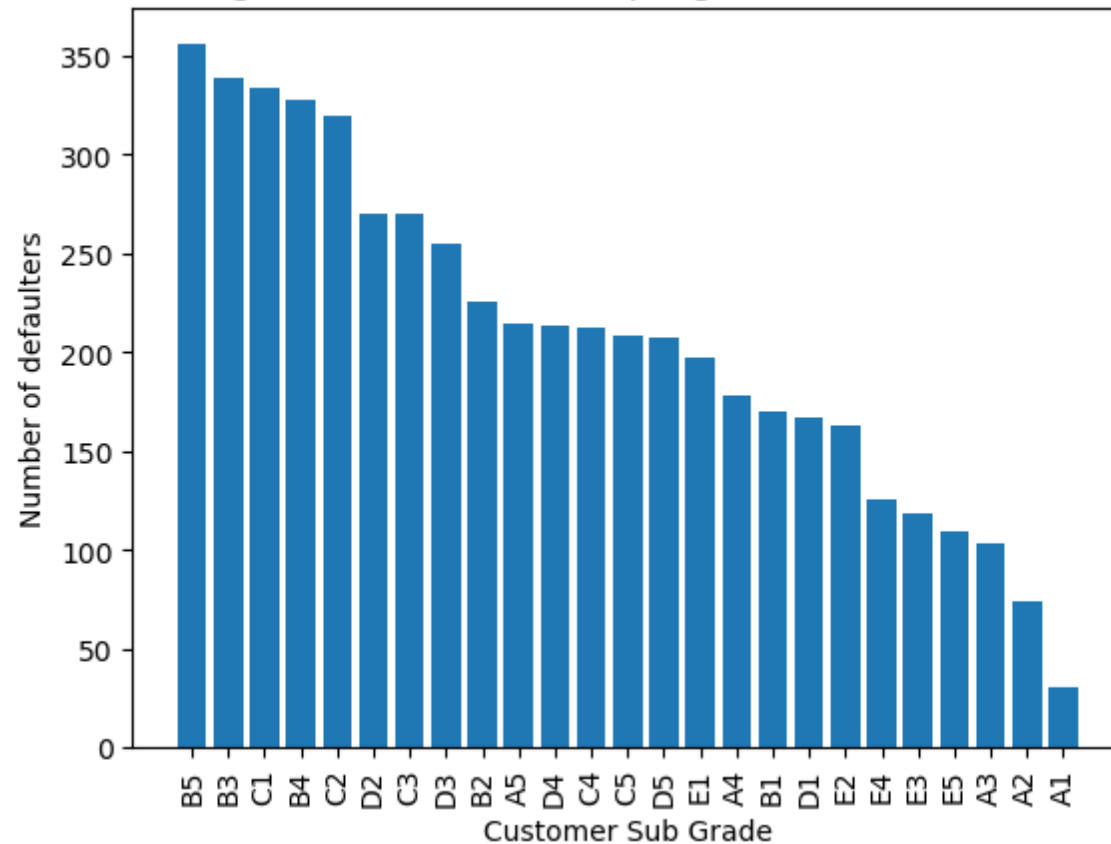
Borrower verification status Vs defaulters

Data set is segmented based on loan status and grouped based on verification status of borrowers.

Majority of defaulters are either not verified or source not verified.

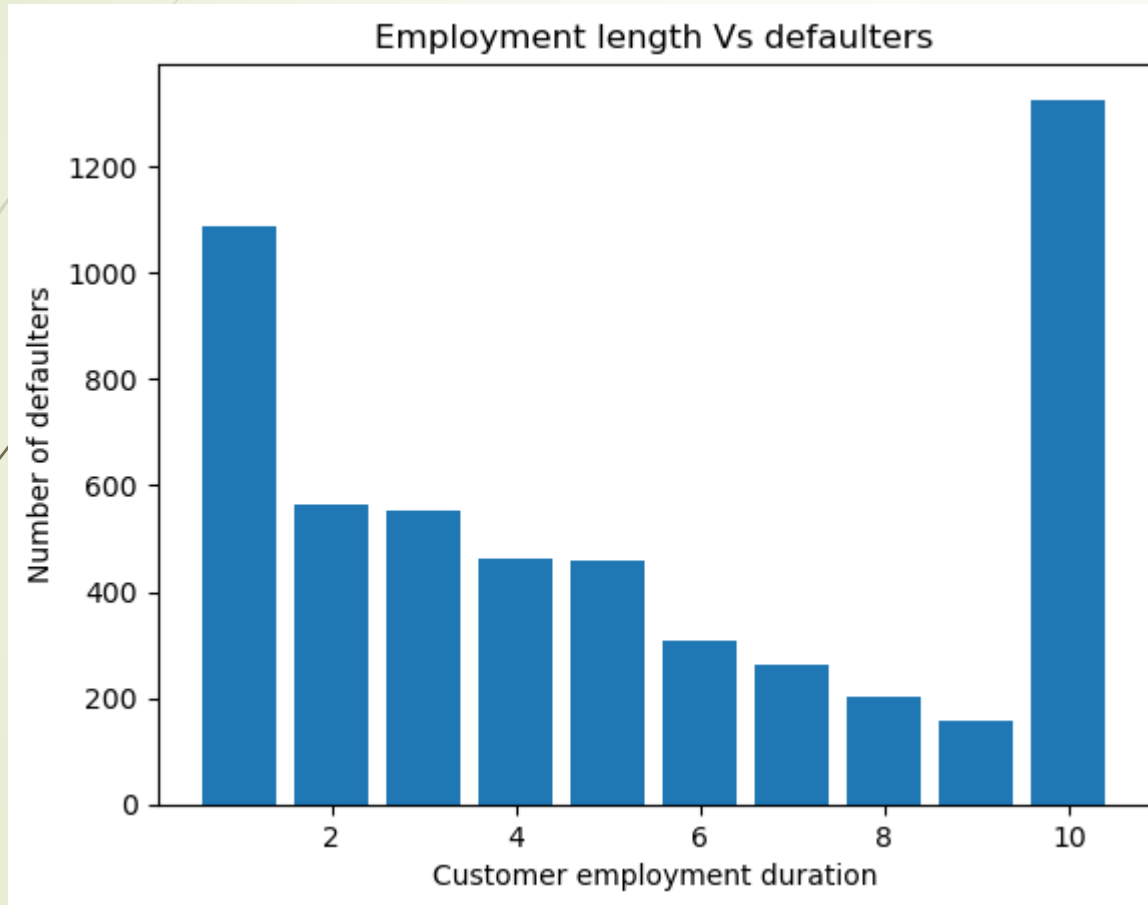This is customer attribute but decided by lender, influencing tendency of default

Annual income Vs. employment length of all Borrowers

Annual Income of borrowers Vs. employment Length
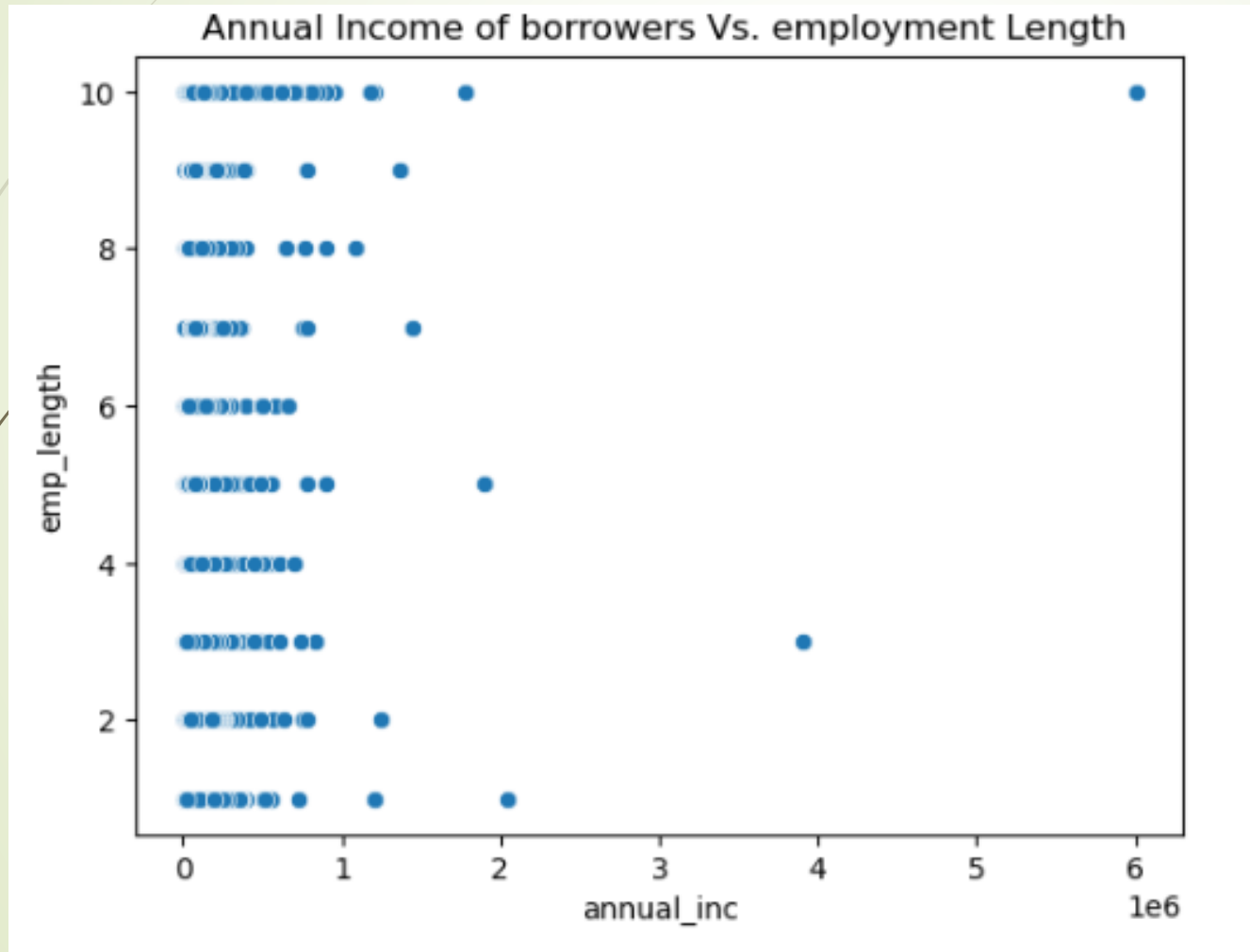
Majority of borrowers have annual income < 10 lakhs

Scatter plot is used to determine the vital relationship between two variables

Employment length of defaulters

This is customer attribute influencing tendency of default

Data set is segmented based on loan status and grouped based on employment length of borrowers

Majority borrowers are employed for less than 5 years

Largest single group of defaulters are employed for 10 or more years.
Filling high employment length without proper evaluation by loan agents for increasing business can be a reason for this.
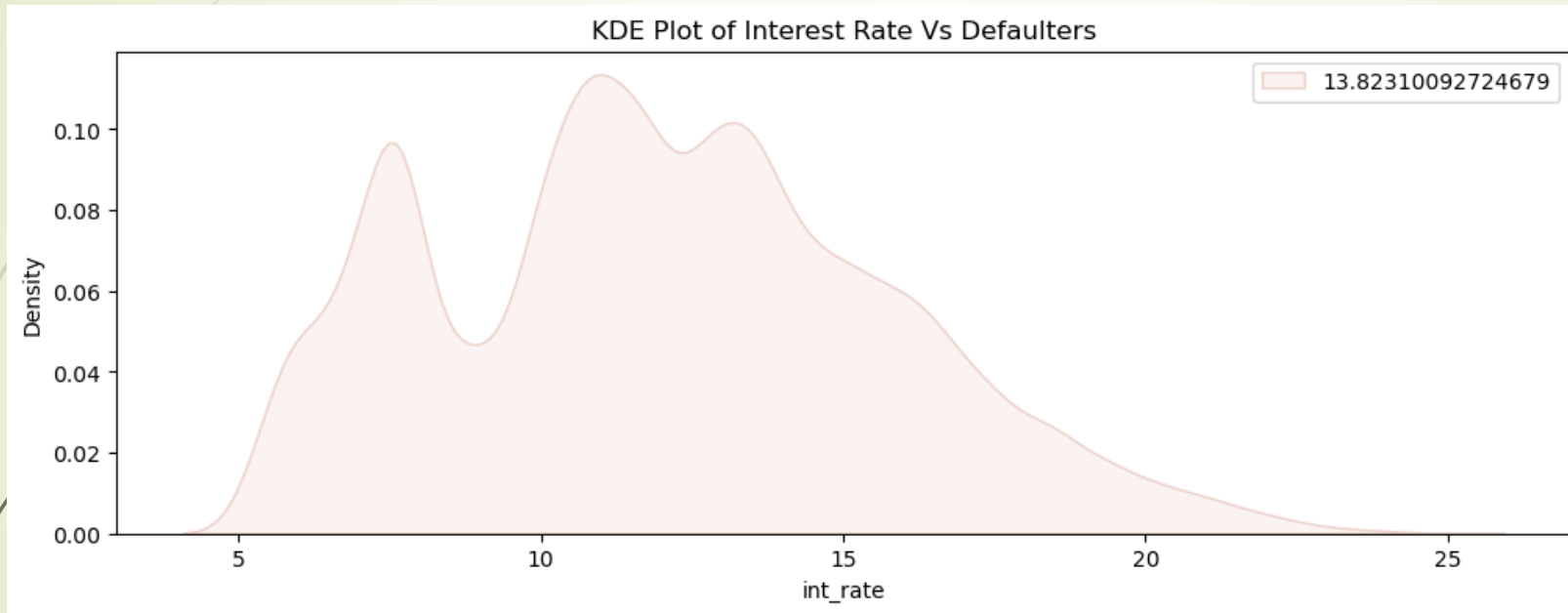
Term of loan Vs. Defaulters

Loan term Vs defaulters
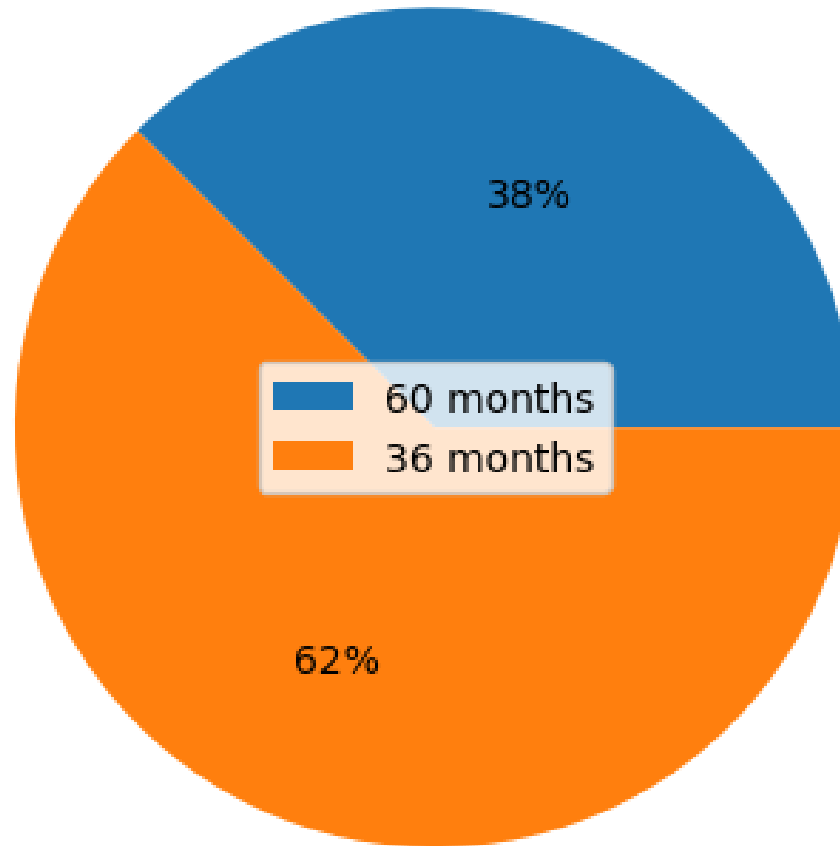
Short term loans (36 months) have more defaulting tendency

Interest Rate Vs. Defaulters

KDE Plot of Interest Rate Vs Defaulters

Majority loans defaulted have interest rates between 10% – 15%

Correlation of key driver variables

Correlation of key driver variables

Loan amount and instalment has a higher correlation of 0.93

Pay ratio and interest rate has a higher correlation of 0.73

Annual income and dti have a negative correlation, which shows that debt burden decreases as annual income increases.

Conclusion

- Customer attributes affecting tendency for defaulting:
  - Home ownership (rented or mortgaged)
  - Average annual income (less than 65000)
  - Employment length (less than 5 years)
  - Debt to income ratio of customer (Above 14)
  - Grade of borrower
- Loan attributes affecting tendency for defaulting
  - Loan Amount (> 12000)
  - Term of loan
  - Pay ratio
  - Interest rate