Lending Club Case Study Assignment

Submission

Group Facilitator: Jitendra Singh

Business Objective

Objective :

- The aim is to identify patterns which indicate if a person is likely to default, 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.

- **Pre conditions for analysis:**

  Use EDA to understand how consumer attributes and loan attributes influence the tendency of default

- **Use Case of this analysis:**

  The company can utilize this knowledge for its portfolio and risk assessment.

Problem solving methodology

Problem solving steps:

1>Read Data/understand data apply suitable encoding.

2>Identify missing values.

3>Replace missing values.

4> Drop un-necessary columns

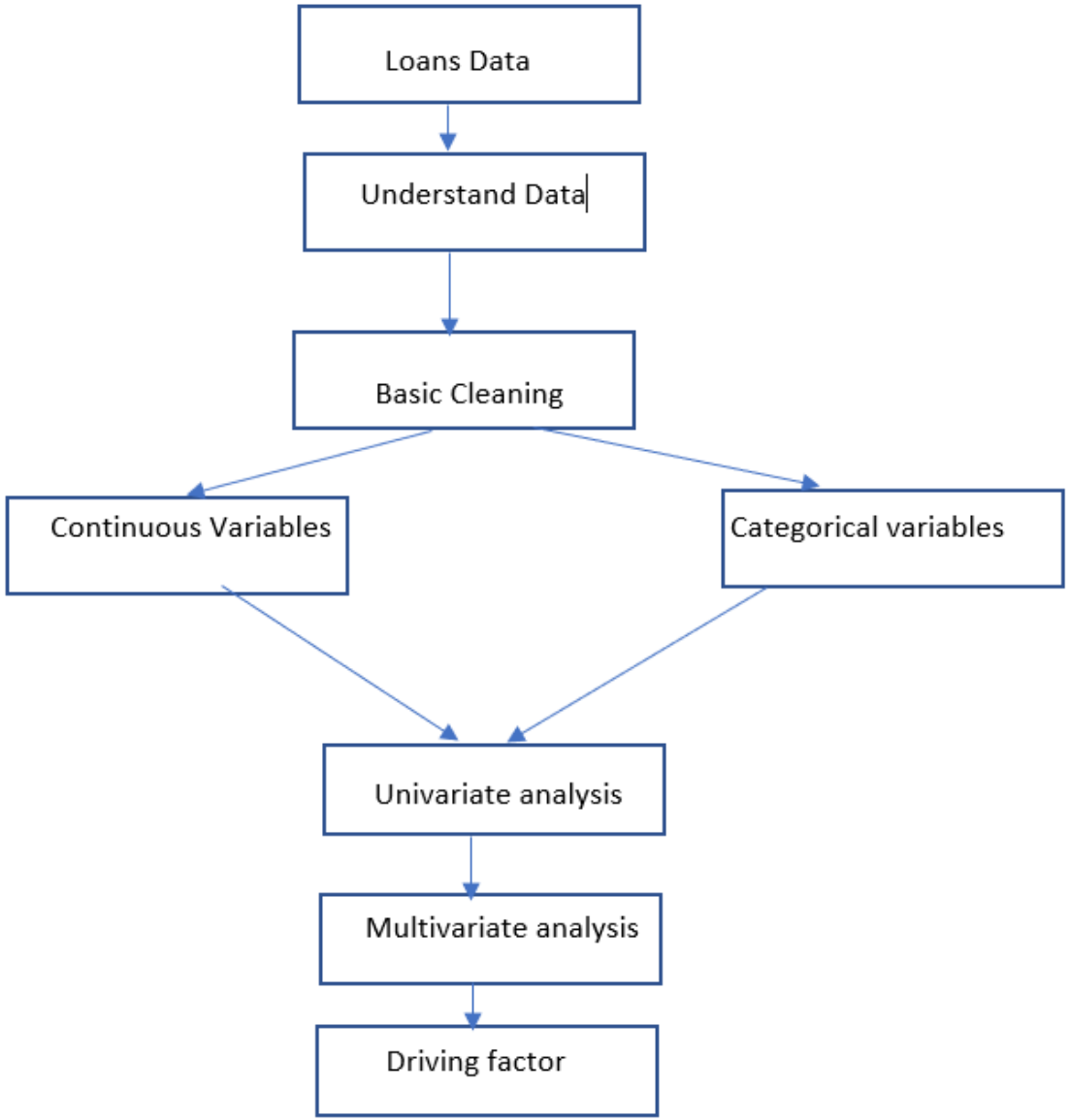
5> Segregate Continuous and Categorical data.

6>Perform Univariate analysis

7>Clean up after Univariate analysis

8> Bivariate or Multivariate Analysis.

9>Repeat above steps (3-7).



10>Draw conclusion on driving factor for loan default

Analysis

Loans data has 100 plus columns.

Post cleaning of the columns the , Removing columns with all null and columns with more than 70 percent null we get around 57 columns

Business understanding of the important columns is as below:

for Loan_status =Charged Off what are the factors driving it .

dti: ratio of individual’s monthly debt payment to his or her monthly gross income

earliest_cr_line:an amount of money a person or company is allowed to borrow during a particular period of time from one or more financial organizations

mths_since_last_delinq: month since last detauulting to pay loan amount

revol_bal: similar to credit card ince amount is paid the credit linit is increased

total_pymnt: total payment done

ID: unique id for each row in the LC

Loan_amnt: Loam applied amount by borrower

Funded_amnt:amount committed to that loan at that point in time

Funded_amnt_inv: amount committed by investors for loan.

Analysis

term:duration of loan in months

grade : Business grading of loan

sub_grade : business driven sub grading

emp_title: employee occupation designation or tile

emP_lenght: employee employment duration (total experience)

home_ownership : whether the borrower is staying in rented , own or mortaged house

annual_inc: earning of the employee per year

verification_status: whether income earned was

verified issue_d : month on which loan was issued

purpose of the loan as claimed by borrower

title of the loan

addr_state: name of state where loan is taken

Derived metrics :Type driven metrics.

revol_util_percent derived from revol_util

int_rate_percent derived from int_rate

Year and month columns derived from issue_d

Analysis

Columns are segregated into Numeric and Non Numeric , basically Continuous and categorical data.

Box plots are plotted on Continuous Data to identify the outliers.

Based on outliers data is cleaned further by checking how prevalent is missing data , outliers and is there any pattern in data missing .

Plot is reverified again to see if there are more outliers .

Outliers are removed by considering only records within 3rd standard deviation Plus or minus record.

Multi variate analysis are performed as below:

Continuous and Continuous are plotted using scatter plot and analysed

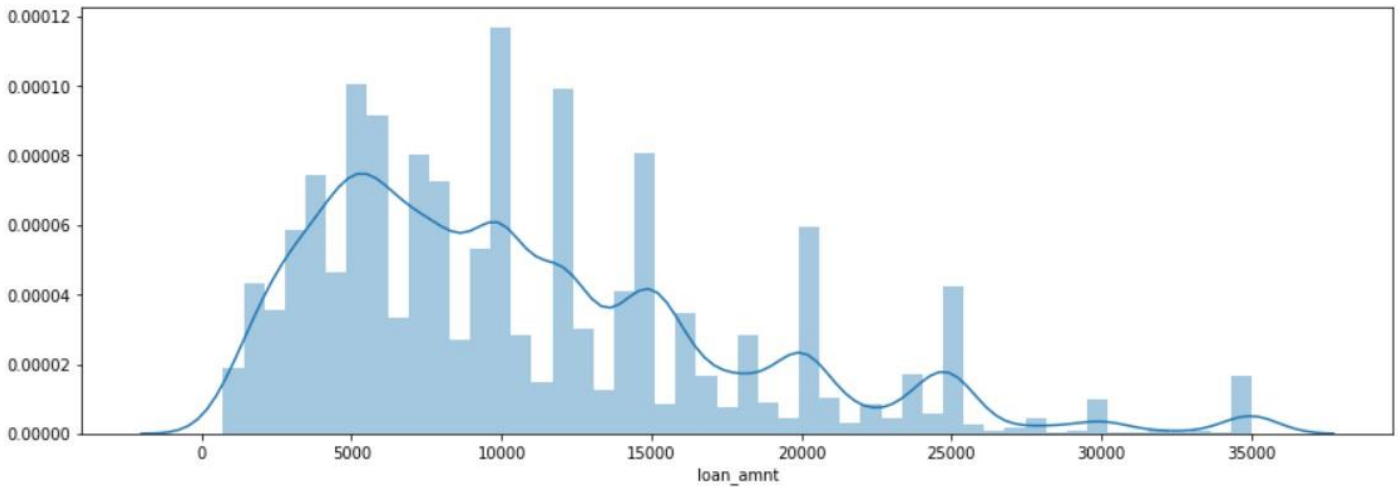
Categorical and Continuous columns are plotted using Box plot and analysed

Categorical and Categorical columns are plotted using Stacked and Box plots (with hue values)

Combining Univariate , Bivariate analysis and relationship between the column conclusion is drawn on default reason.

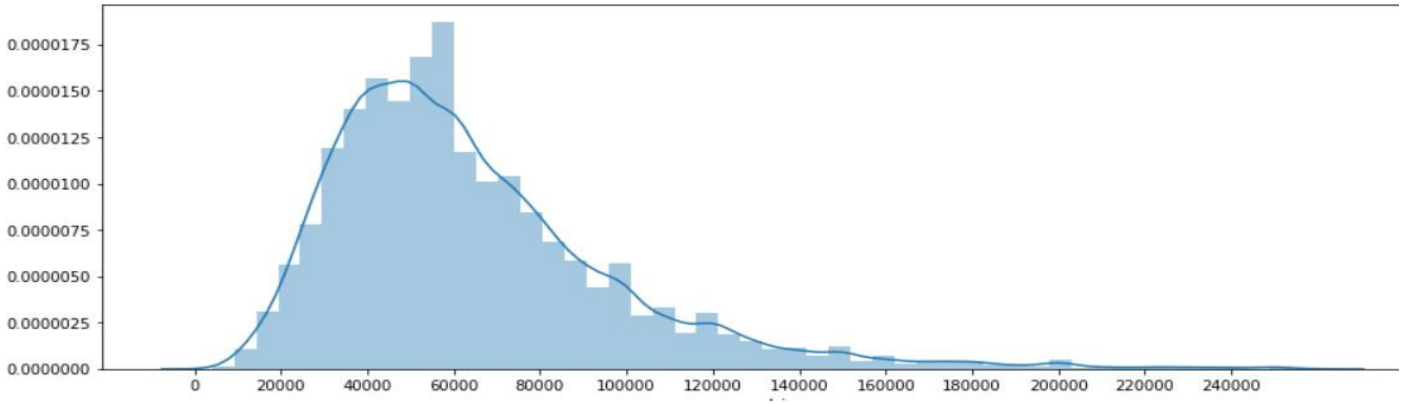
Results

Plot showing Loan amount Distribution (More amount is distributed around 1000 to 12000



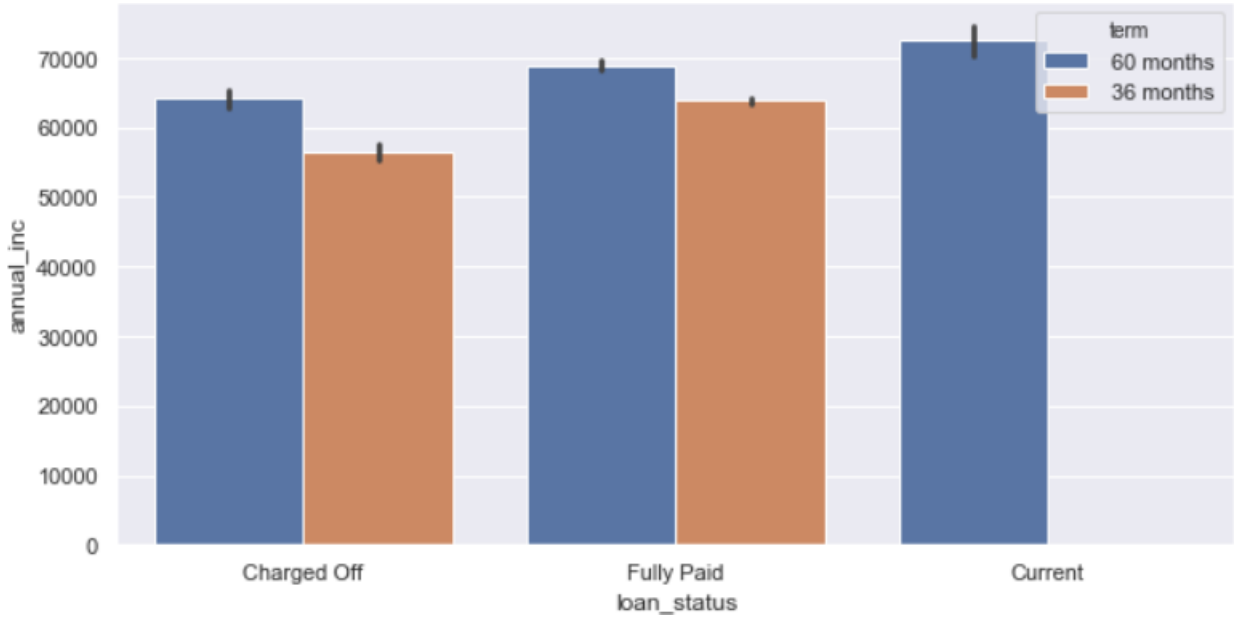
Plot showing annual_income :

Majority of annual inc is bet 4000 to 7000 range

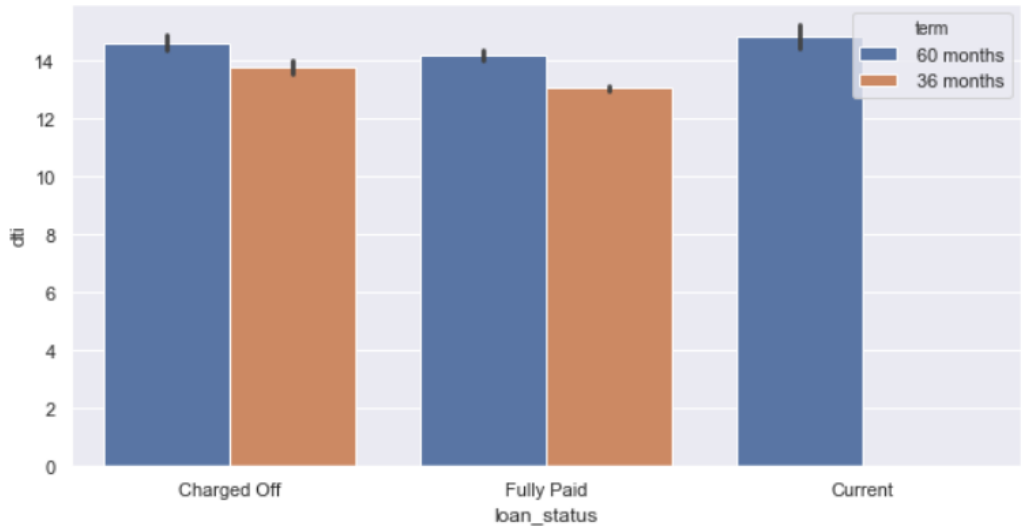


Results

Plot showing income of charged off loans are less compared to fully paid and in progress

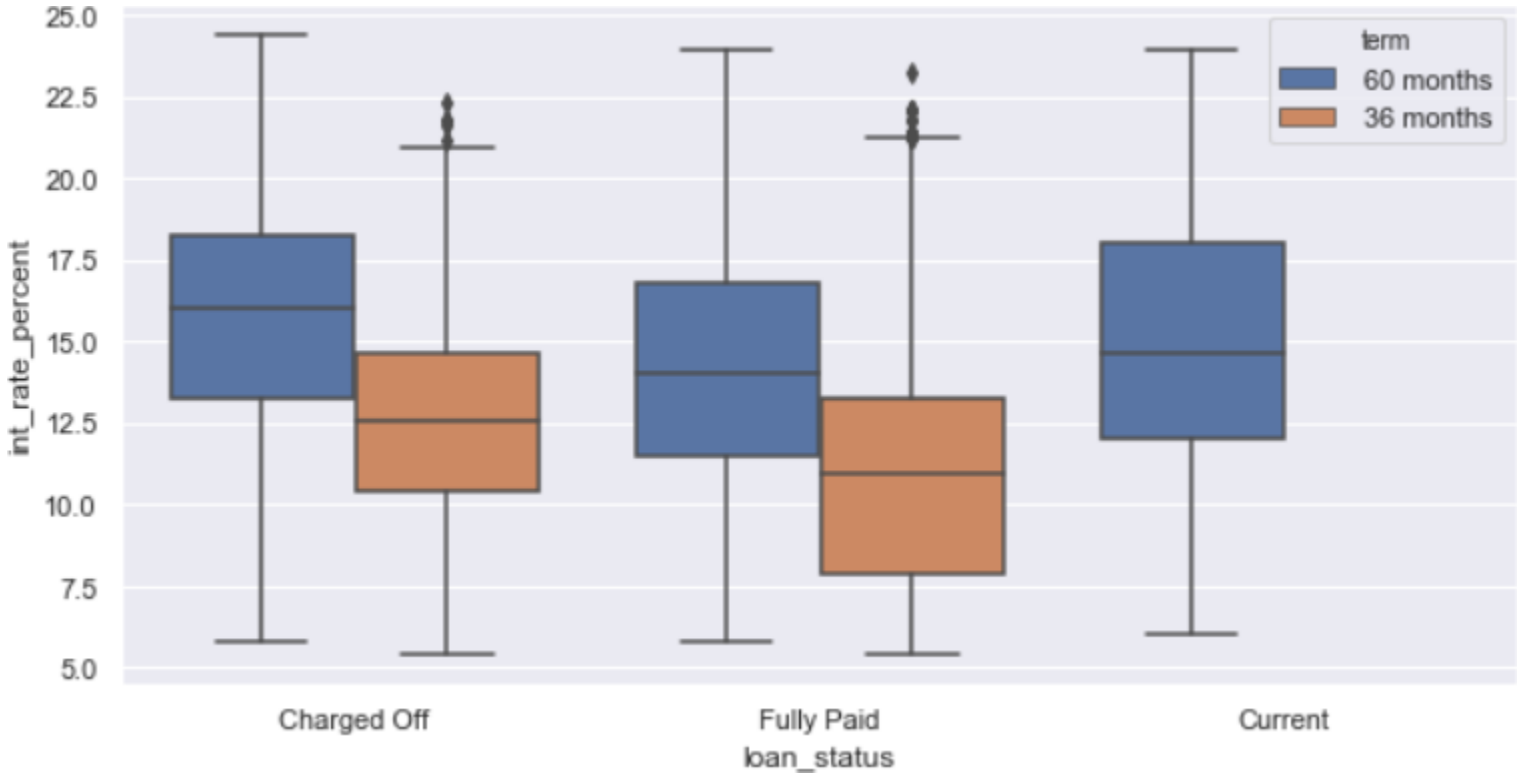


Plot showing DTI of charged off loan is higher



Results

Plot showing interest rate is higher for charged off loan



Conclusions

Conclusion:

Granting loan for parties in

1) lower income group
2) higher interest rate
3) higher instalment
4) higher dti results in defaulting of the loan .