Lending Club Case Study

How Lending Club can reduce loan defaults?

Sujay Kondanekar
Venkatesh H V

The problem

Company

Lending Club is a largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures

Borrowers can easily access lower interest rate loans through a fast online interface

Context

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

Problem statement

Identify the pattern of risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss

Challenges deep-dive

Data Cleaning → Data Analysis → Data Storytelling

Data Cleaning

- Remove nulls
- Fix missing values
- Remove outliers
- Standardise values
- Create derived values

Data Analysis

- Univariate
- Segmented Univariate
- Bivariate

Data storytelling

- Plot graphs and look for interesting trends
- All the columns were analysed with loan_status to check for correlation

Data quality issues addressed

- 50+ Invalid columns with not even single non-null value. All of them were dropped
- Rows with Null values were removed
- Outliers records were removed from multiple columns based on the Interquartile Rule
- Date fields were transformed to uniform format like %b - %d
- Special characters and text in the columns like %, <, “years” were eliminated

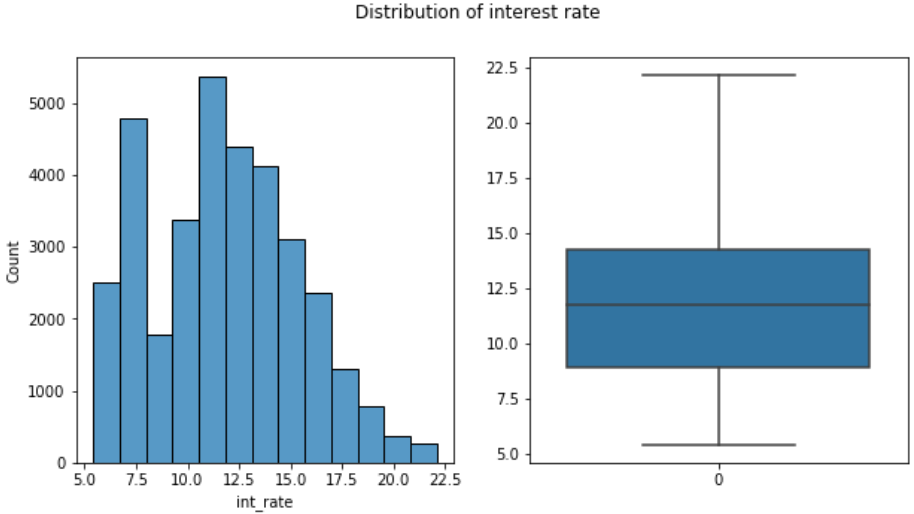
Solution

Driving factors for loan default

EDA of Lending Club data revealed many interesting factors that defaulting loan applicants exhibited in common

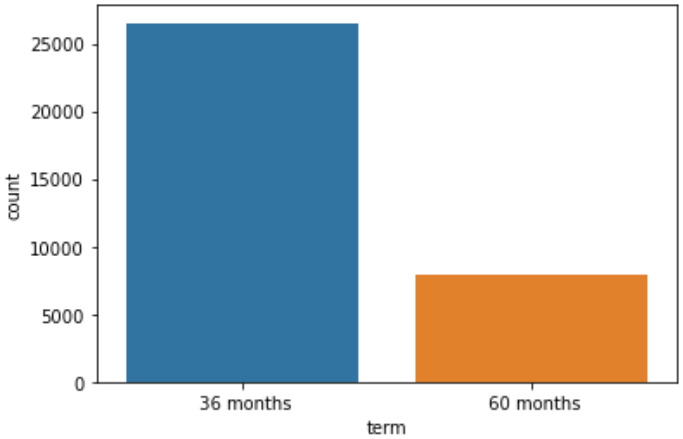
Data analysis in LC

Lending Club data Insights

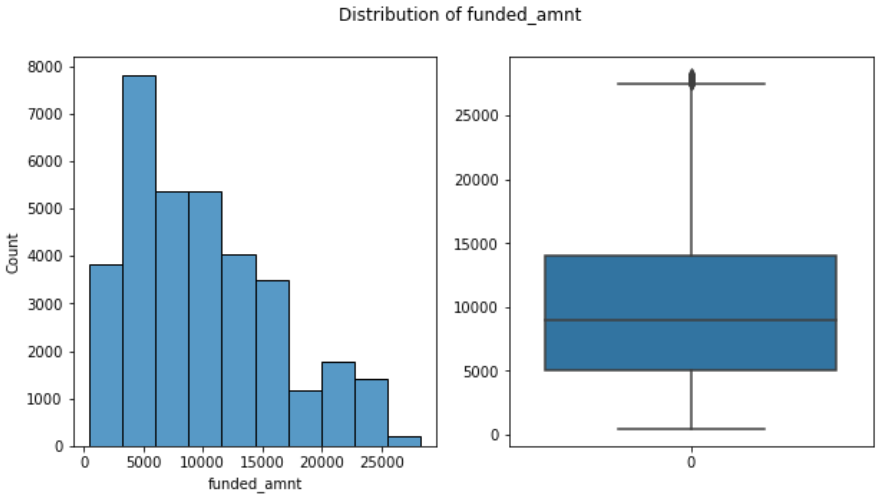


← LC interest rates varied from 5.4 to 22.1% with more loans offered around 11%

Only 2 terms (36 / 60 months) offered by LC as loan tenure and 36 term was issued highest →

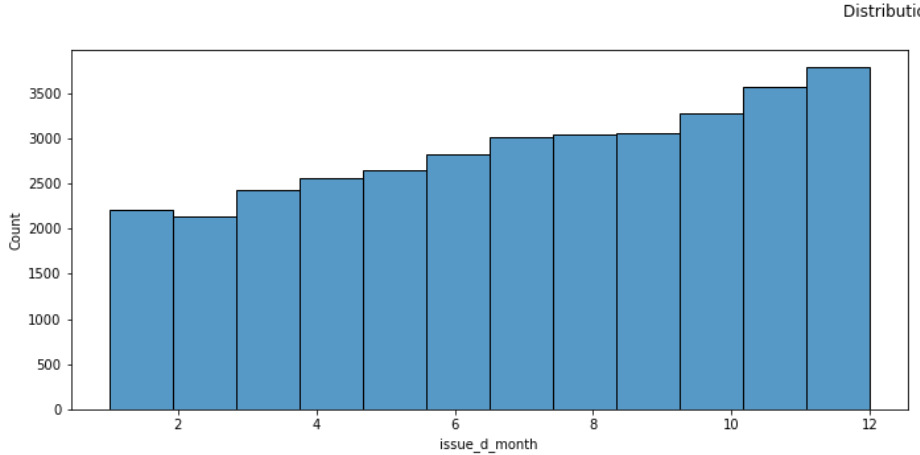


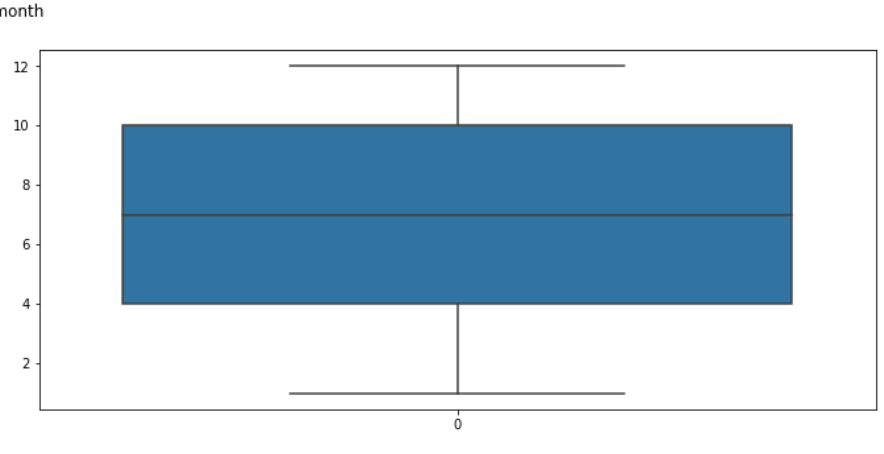
Lending Club data Insights



Maximum amount funded was 28250. 5000 was the most funded amount to borrowers. Median funded amount was around 9K

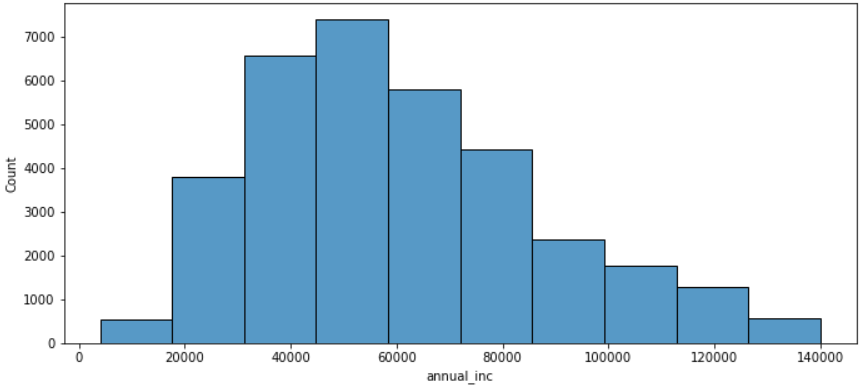
Lending Club data Insights





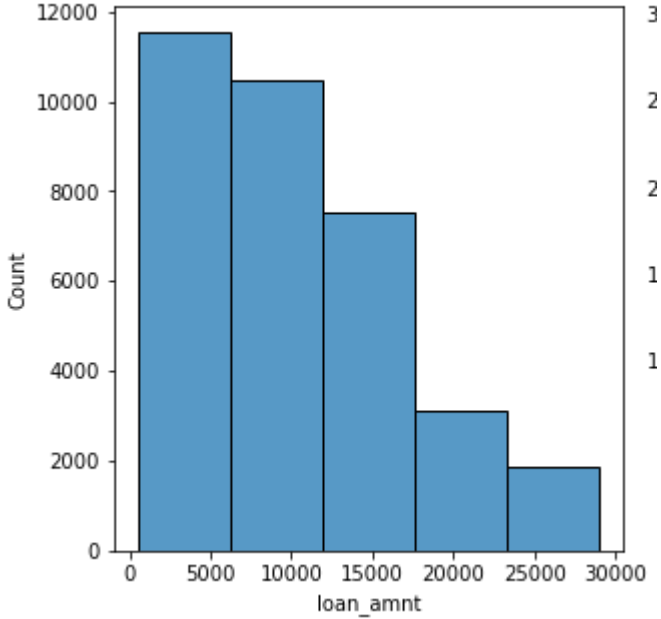
LC offered more loans towards the end of the year which coincides with Christmas and New year

Applicants’ Insights

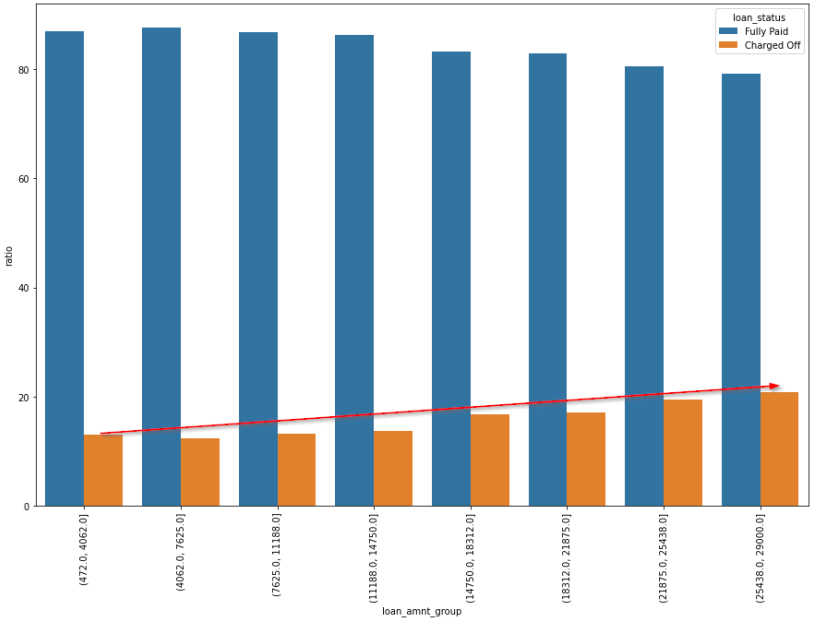


Annual income reported by most of borrowers was between 50000 to 60000

Maximum borrowers opted for upto 6K loan amount

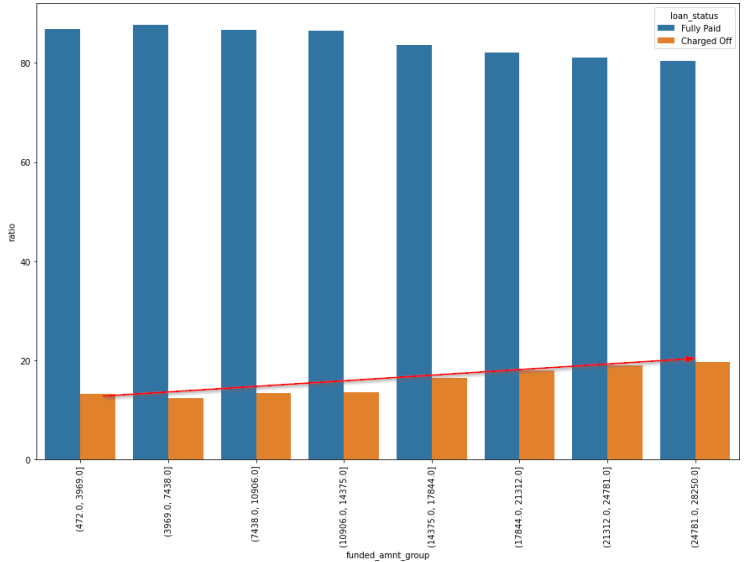


Applicants' Insights contd..

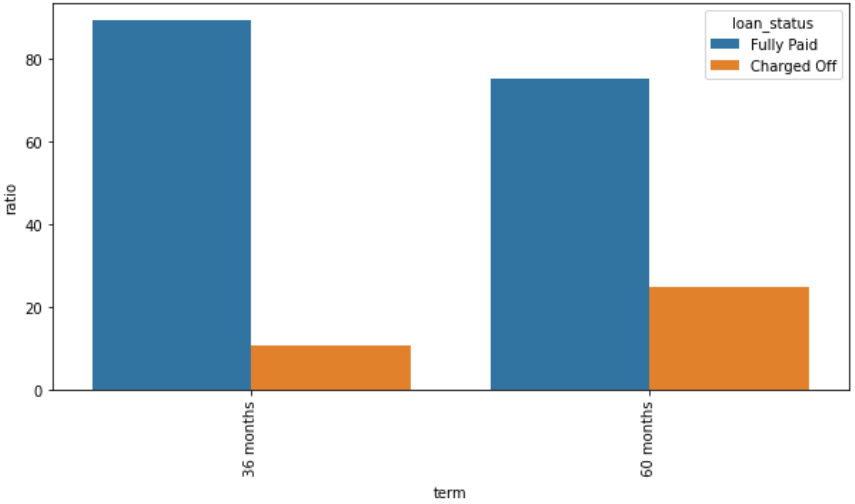


There is a positive correlation between loan amount and applicant defaulting

There is a positive correlation between funded amount and applicant defaulting

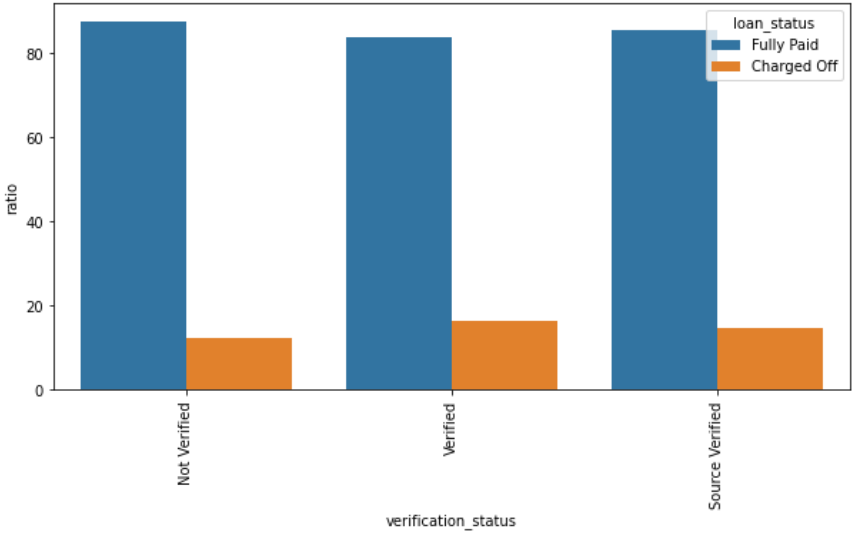


Applicants' Insights contd..

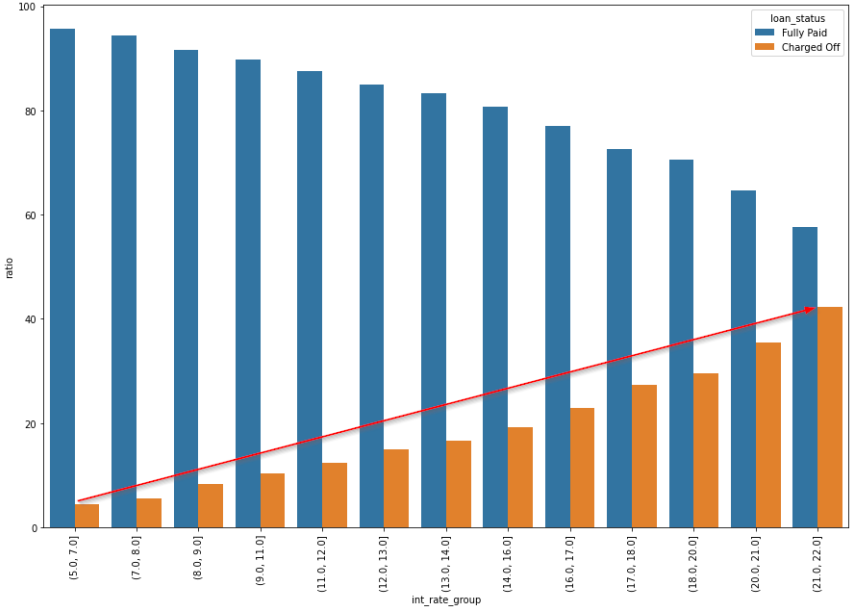


60 month term loans were most defaulted

Surprisingly, Verified borrowers defaulted most

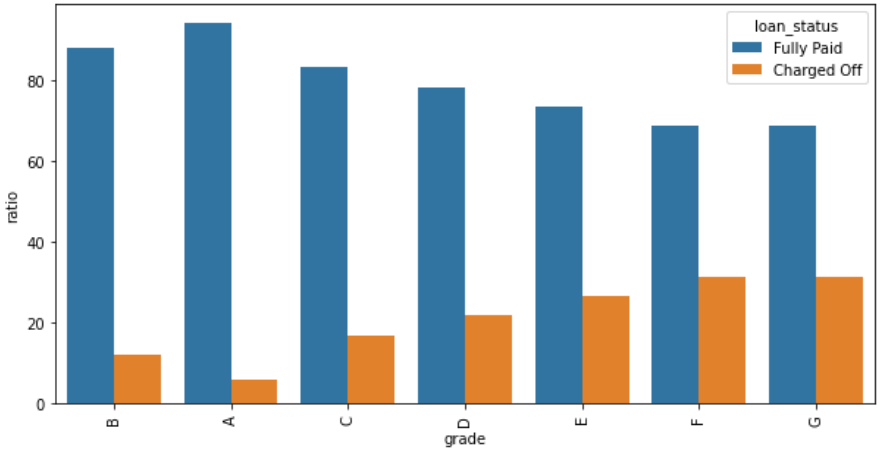


Applicants' Insights contd..

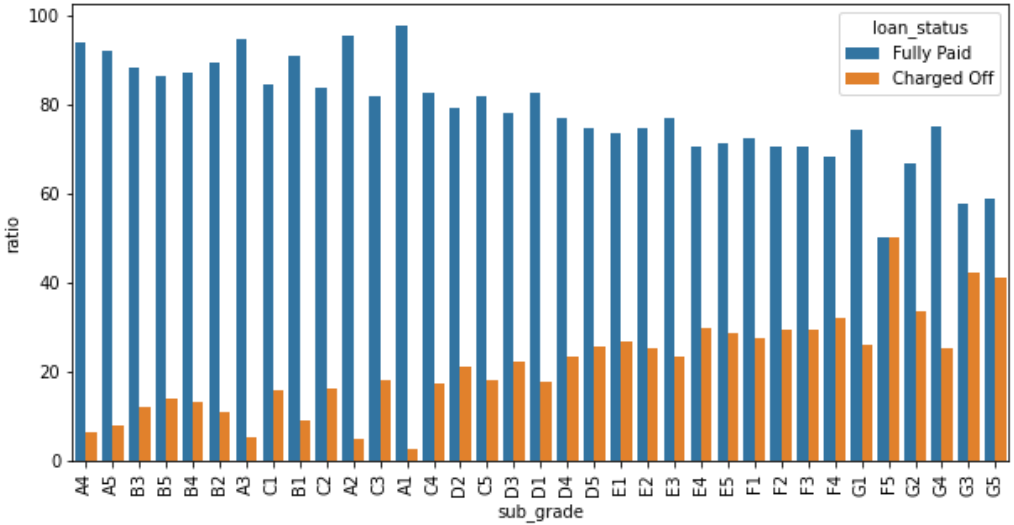


Positive correlation of interest rate with default rate is observed

Applicants with A grade loans defaulted less. As grade moved away from A towards G, default increased

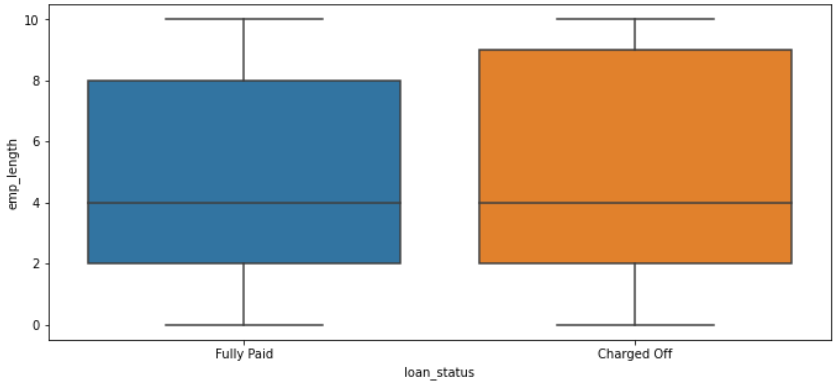


Applicants' Insights contd..

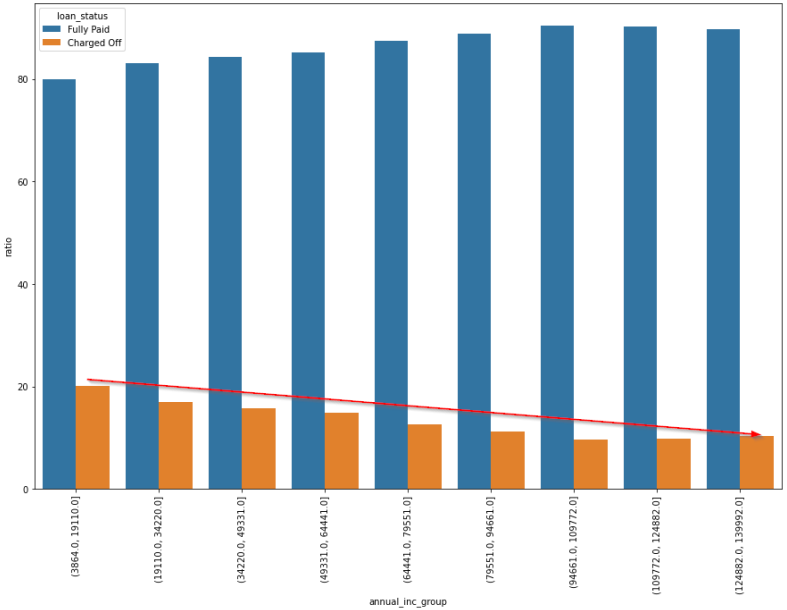


← F5 Subgrade loan borrowers defaulted at 50%

Borrowers with employment length between 4-9 years likely defaulted more based on the difference between Q3 and Q2 →

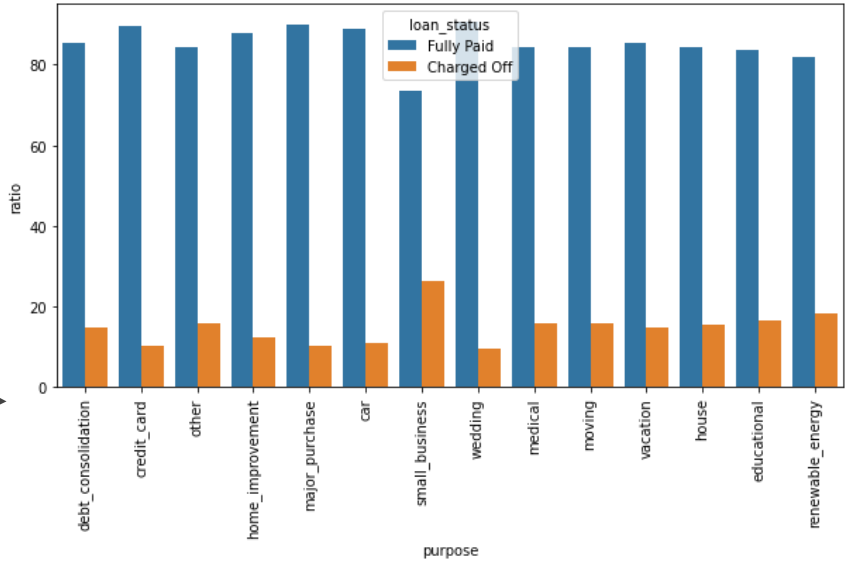


Applicants' Insights contd..

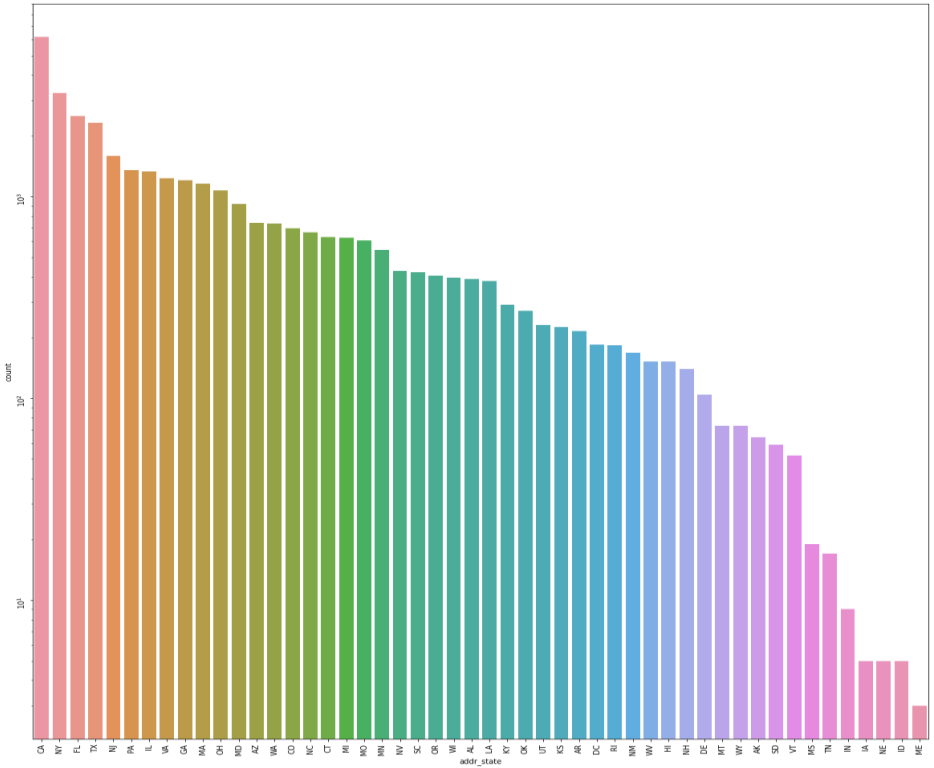


With higher annual income applicants defaulted less..
Annual income and loan status are -ve correlated

Loans borrowed for the purpose of small business defaulted more compared to others

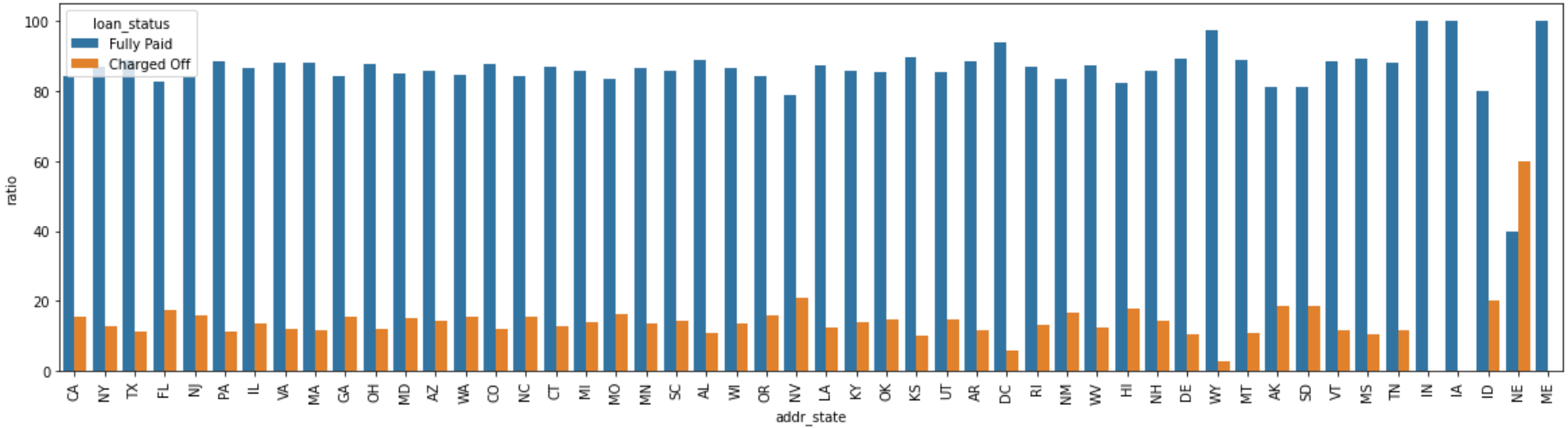


Applicants’ data Insights



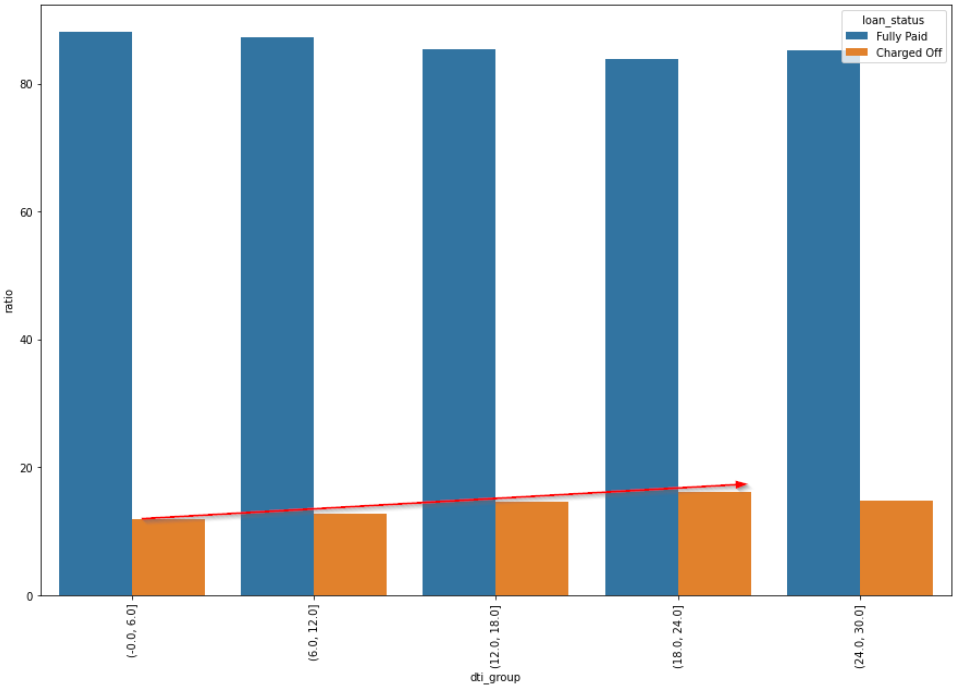
LC lends the most in the state of California

Applicants’ data Insights



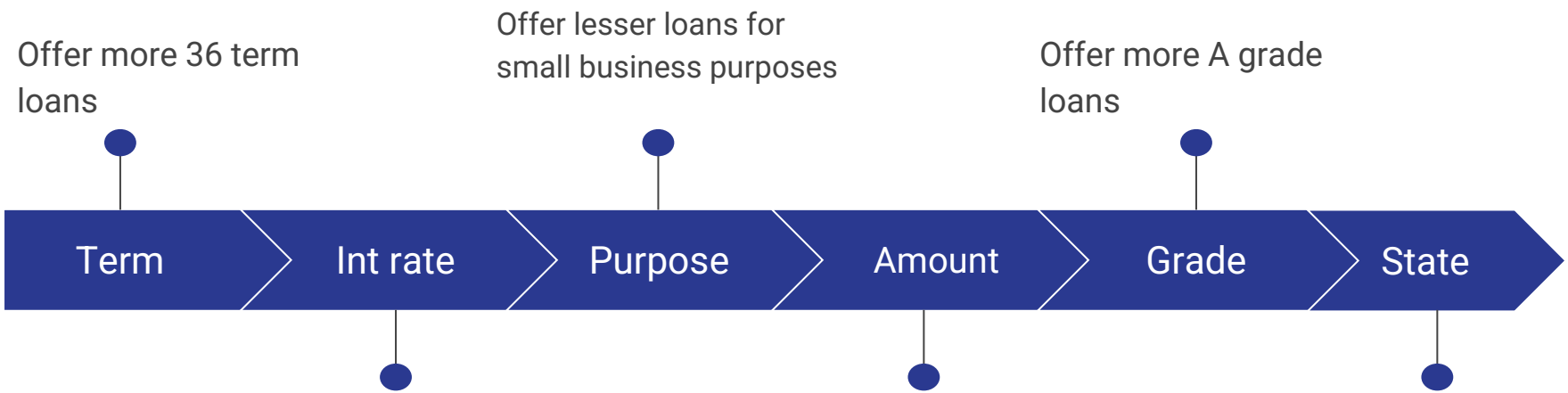
Defaults were high for borrowers from NE (Nebraska) and nil for borrowers from IN(Indiana), LA(Louisiana) and ME(Maine)

Applicants’ data Insights



Defaults increased when debt to income ratio increases

Inference : LC should follow below steps to reduce default loans



Impact

Lesser loan defaults