Lending Club Case Study

VARUN KHANNA

VARSHA S

Objective

The objective of the Lending Club case study is to analyse the factors influencing loan defaulters and predict the likelihood of borrowers repaying their loans. By leveraging historical loan data, the study aims to build a predictive model to assess credit risk, improve loan approval processes, and enhance decision-making strategies for managing loan portfolios.

Business Understanding

Lending Club is a peer-to-peer lending platform that connects borrowers with investors. To optimize its lending processes and minimize financial risk, it is essential to understand the factors that influence loan performance, such as borrower characteristics, loan attributes, and payment history. Using Exploratory Data Analysis (EDA), the goal is to uncover patterns and relationships within the historical loan data that can impact the likelihood of loan defaults or successful repayments.

Through EDA, we aim to identify key variables that affect loan outcomes, detect trends in borrower behaviour, and spot potential risk indicators. This understanding will guide Lending Club in refining its credit risk assessment models, setting more accurate interest rates, and implementing targeted strategies to reduce default rates and improve profitability.

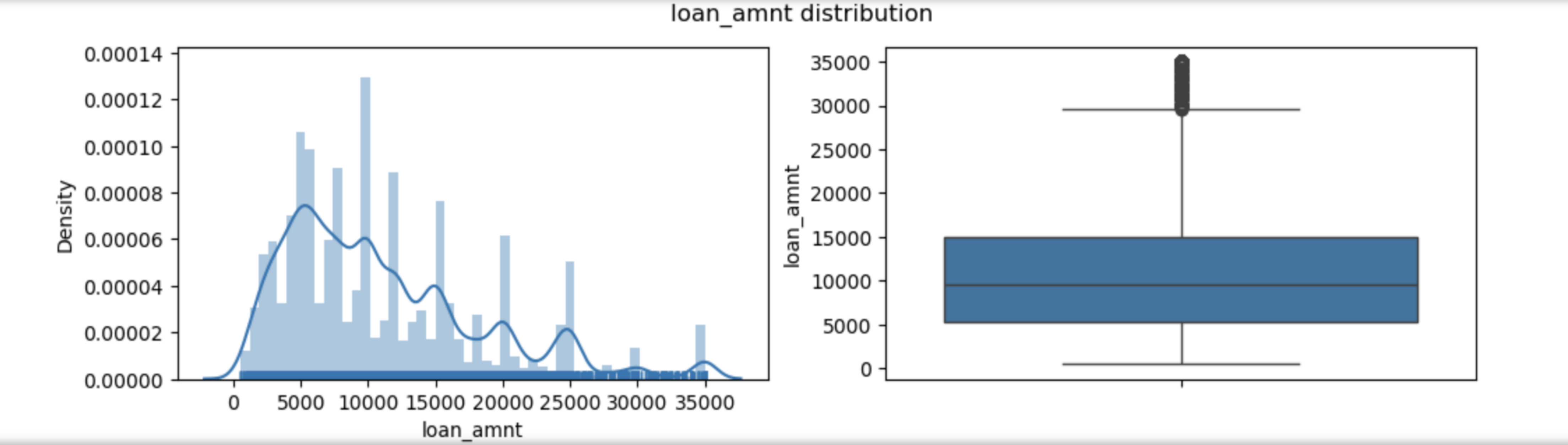
Dataset details

The loan dataset contains information about past loan applicants and whether they have defaulted or fully paid the loan.

The dataset contains information of approved loans only.

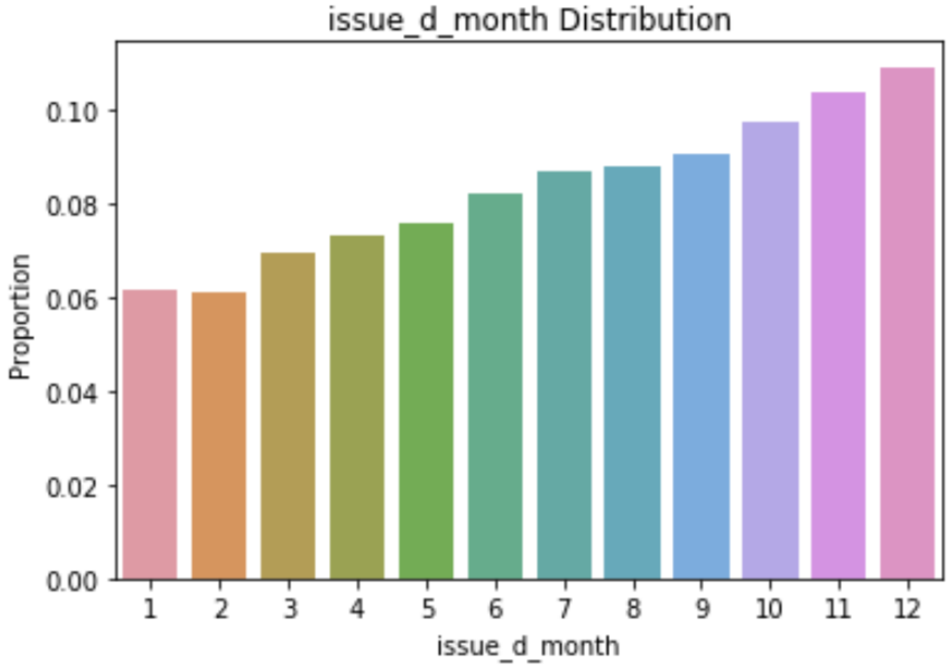
There are 3 loan statuses – Fully Paid, Charged-Off and Current.

Loan amount distribution



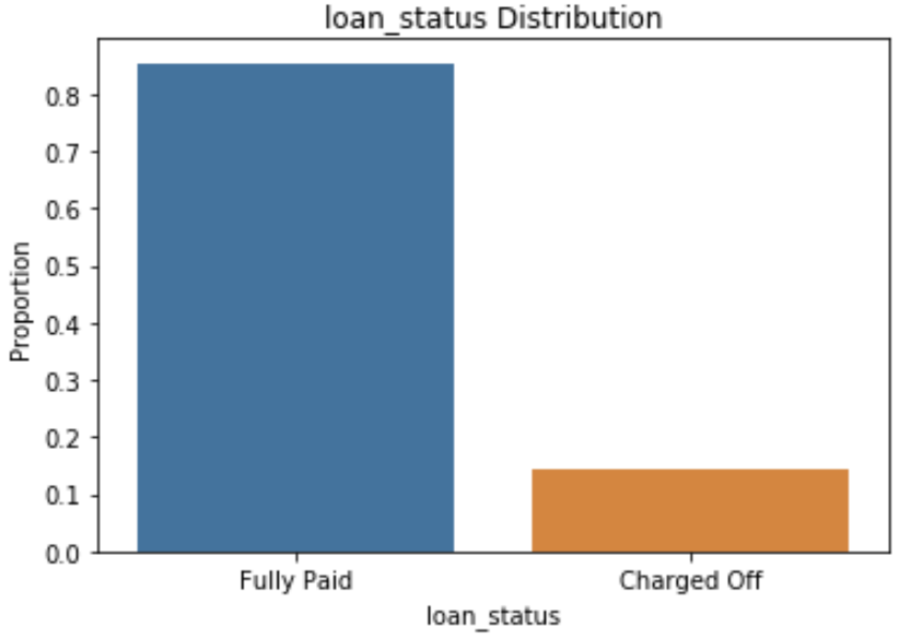
- Most of the people took loan of 10K, mean is also around the same value.
- 95% of the loans are below 25K.

Issued month distribution



More number of loans are issued in the last quarter of the year as bank employees would be trying to reach the year end target for the loans in financial year

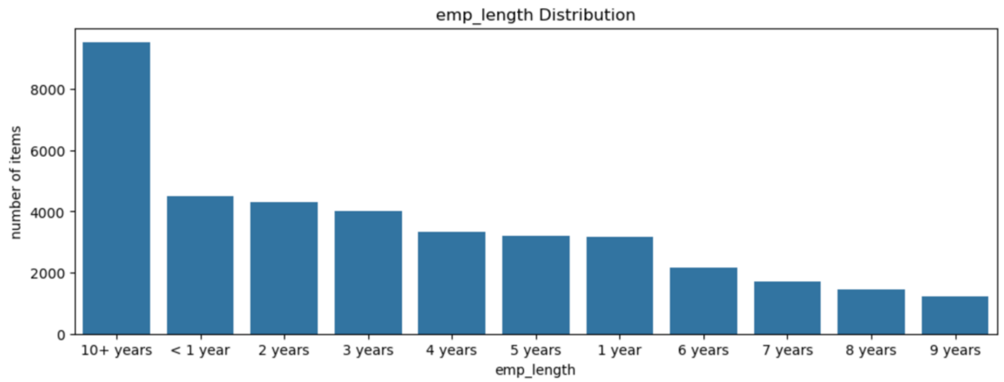
Loan status distribution

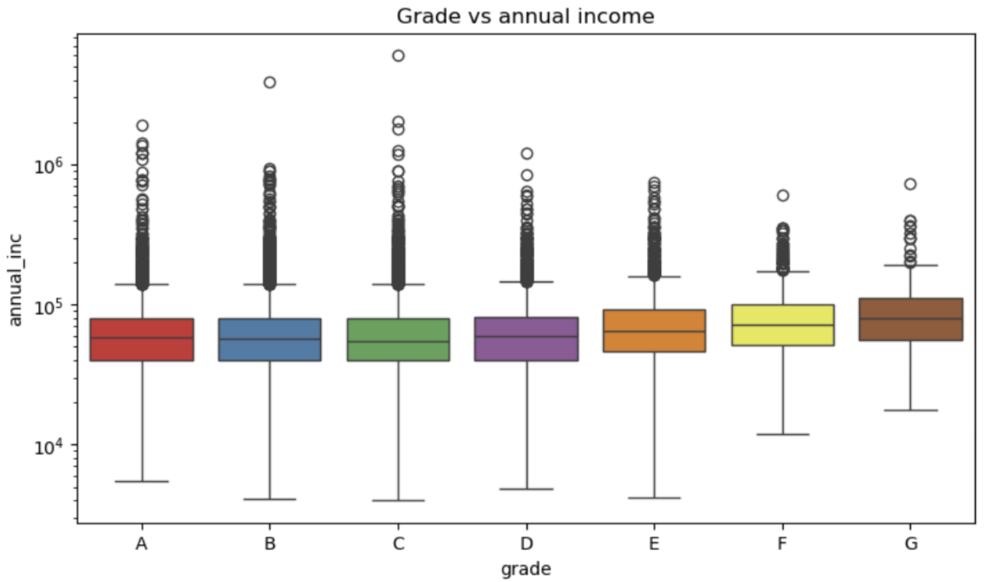


Nearly 85% of loans have been fully paid, and only around 15% of loans ##have been charged-off

Employment length distribution

Annual income vs Grade

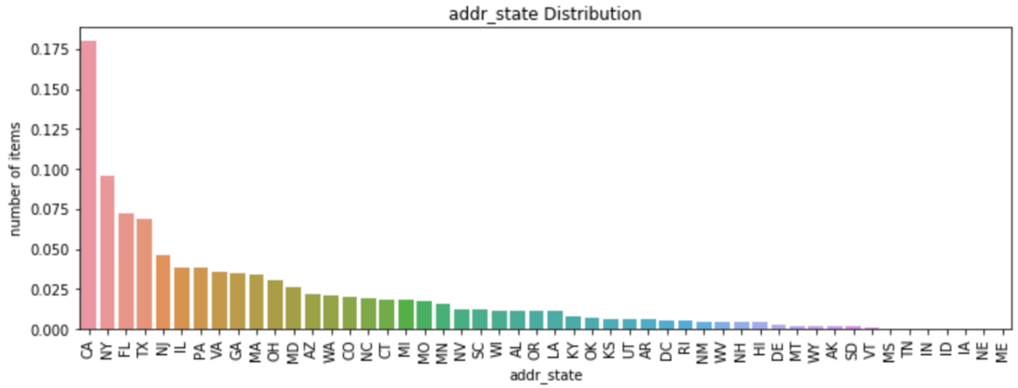




- People with most experience - 10+, are taking the maximum number
- From 2 to 9 years - as the experience increases the number of people taking loan decreases
- So as the experience increase salary increases & people have already made assets so loan taking tendency decreases

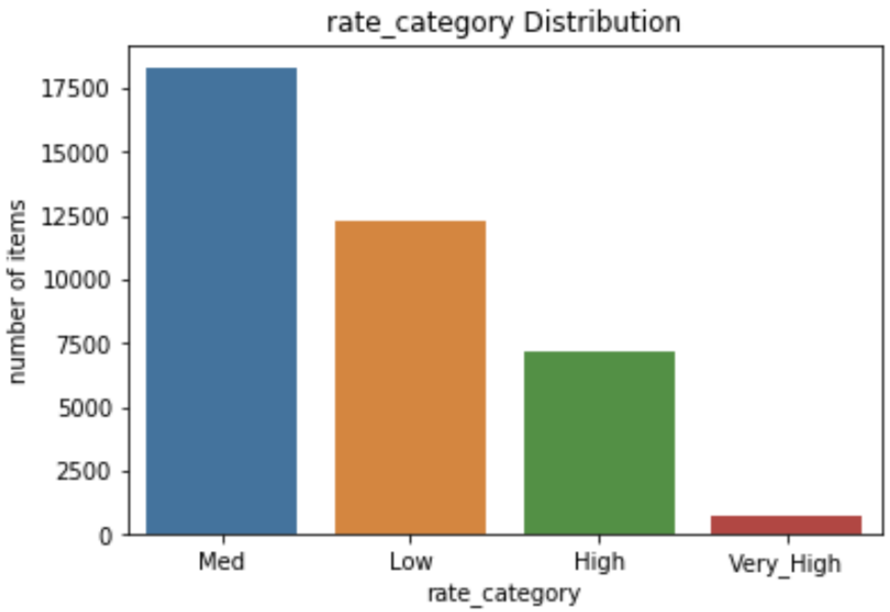
Annual income is comparatively higher for lower grades

Address state distribution



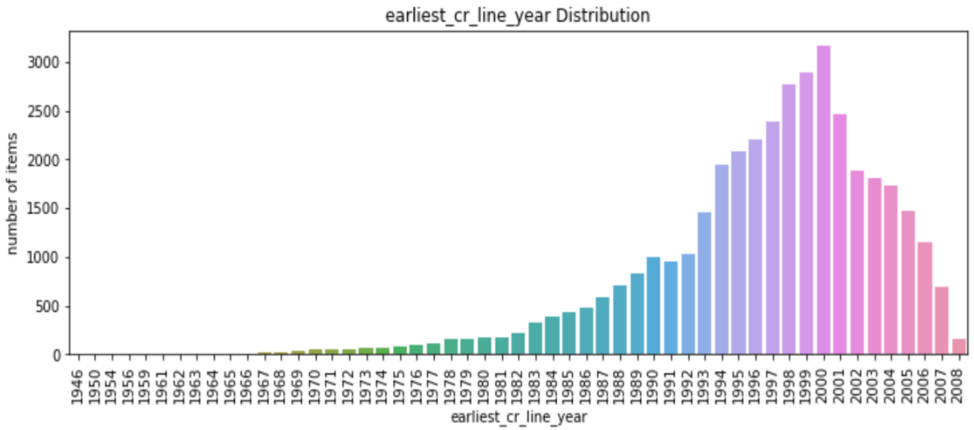
Maximum number of loans are taken by people living in California

Rate category distribution



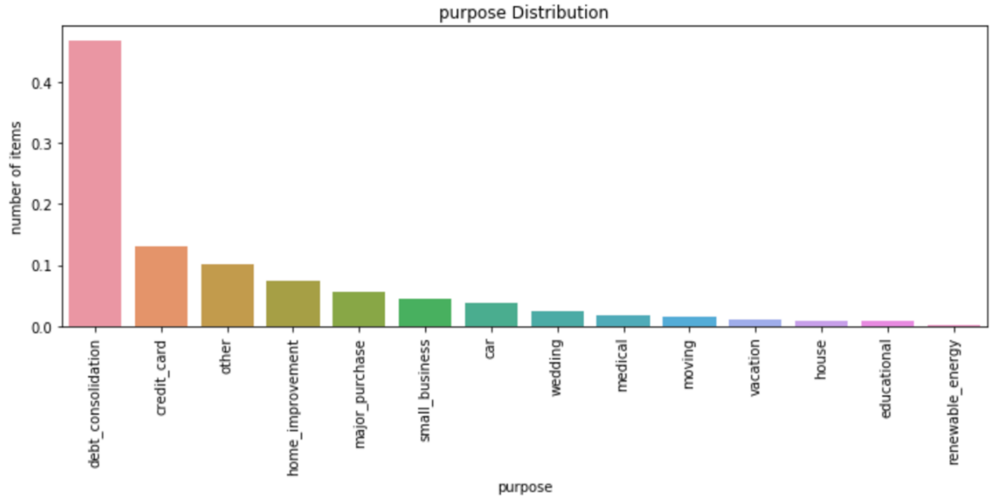
Out of the 4 categories for interest rate people take loans where interest rate is between 10% to 15%

Earliest credit line year distribution



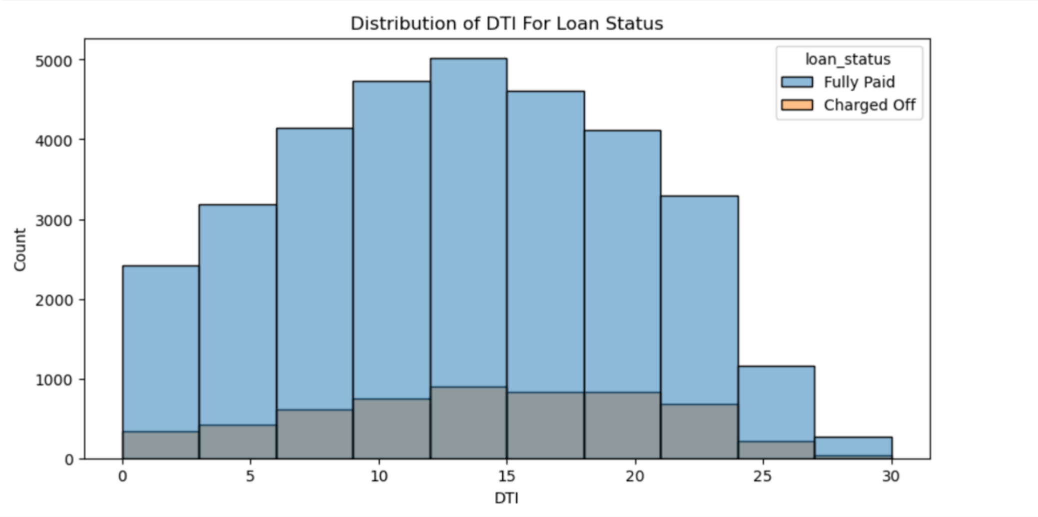
Many of Loan borrowers of LC have got earlier credit line in 2000 year

Purpose distribution



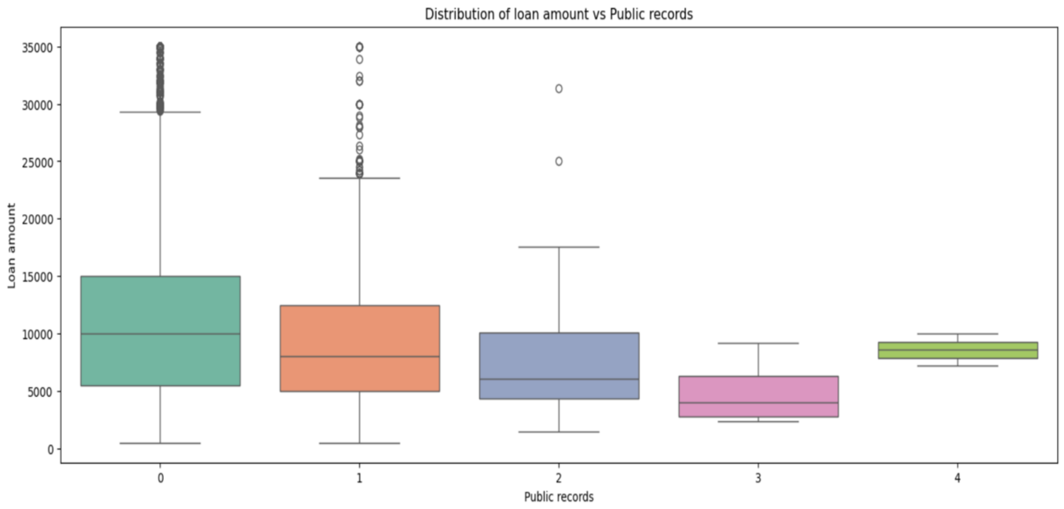
Maximum number of loans are taken for Debt consolidation purpose

Loan status vs DTI ratio



Higher the dti ratio, the higher is the chance of defaulting

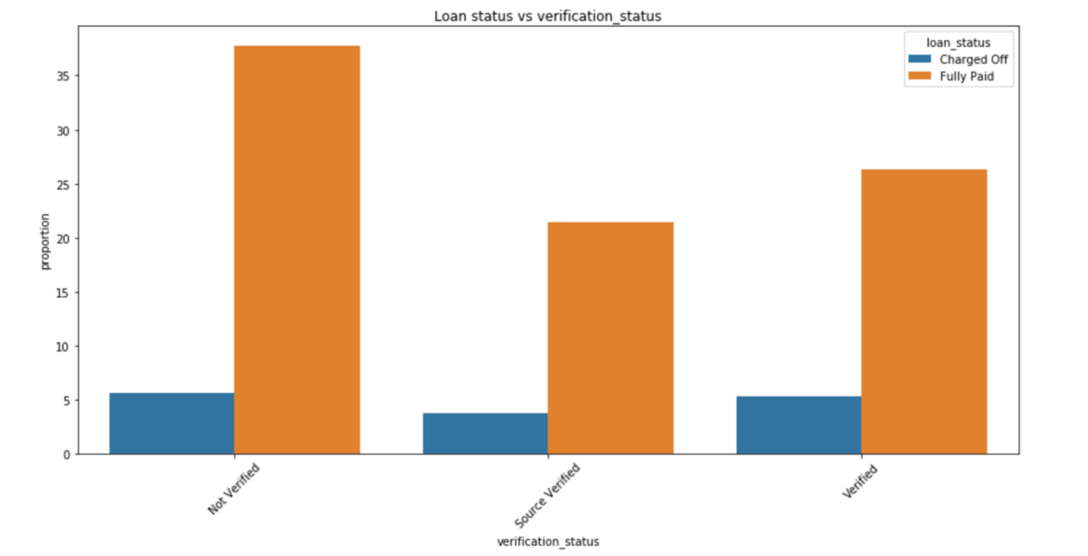
Loan amount vs Public records

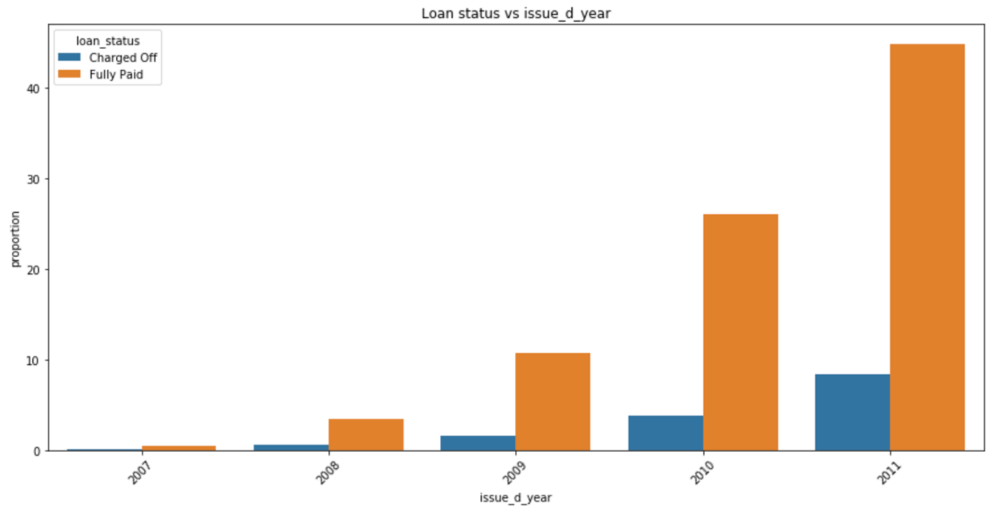


Borrowers with one or zero public records have taken big amount of loans

Loan status vs Verification status

Loan status vs Issued year

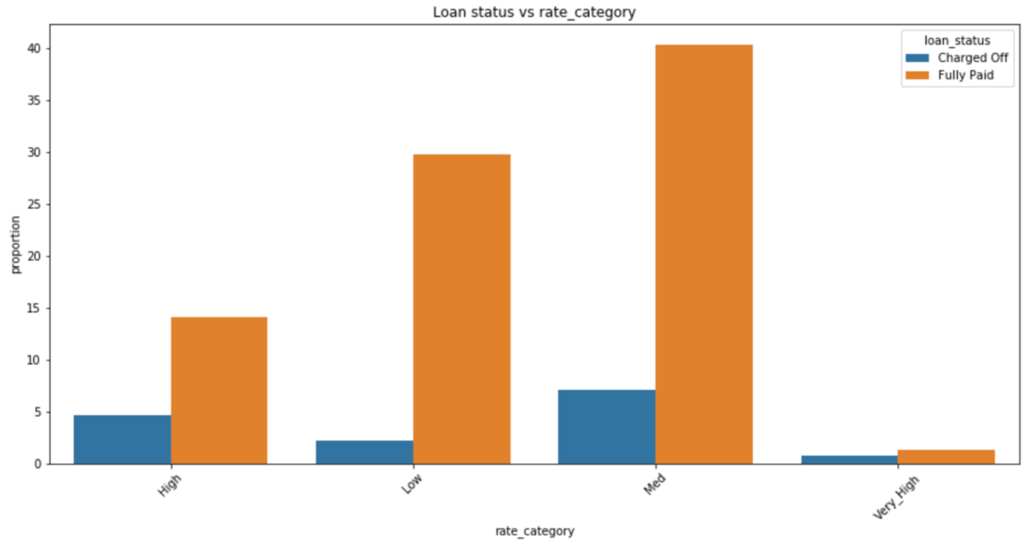




- Verified & not verified have same % of defaulters.
- Maximum no of borrowers do not have verified source of income but they have the least % of defaulters in comparison to others.

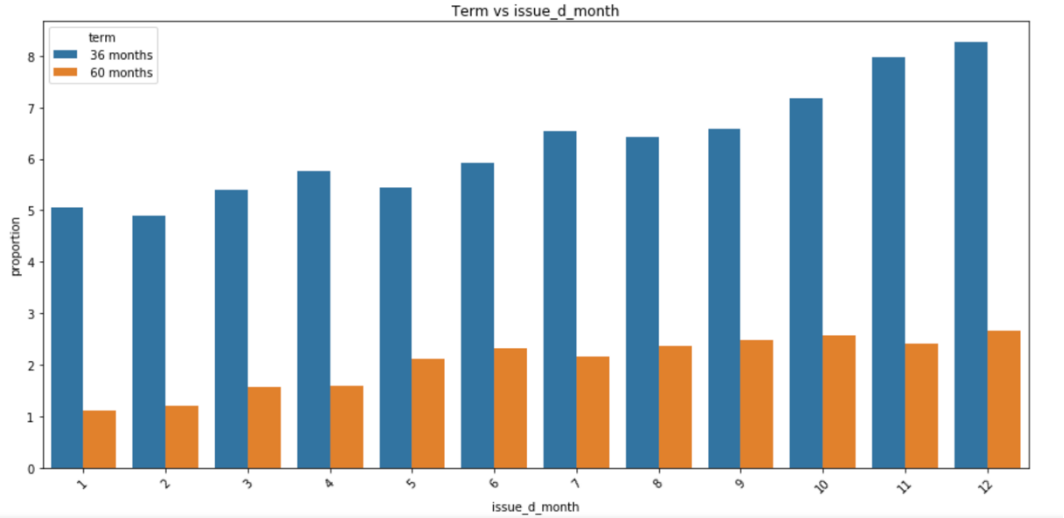
The defaulters have increased by almost 2.5 times as compared to fully paid borrowers who have increased by nearly 1.7 times from 2010 to 2011. So defaulters are increasing at more pace then fully paid

Loan status vs Rate category



- In Low category, the percentage of people defaulted is less than 10% & thus more such type of loans can be sanctioned & is beneficial for LC.
- In very high category & high, the percentage of people defaulted is more than 60% & 30% respectively, thus high interest rate loans have more chances of getting defaulted.

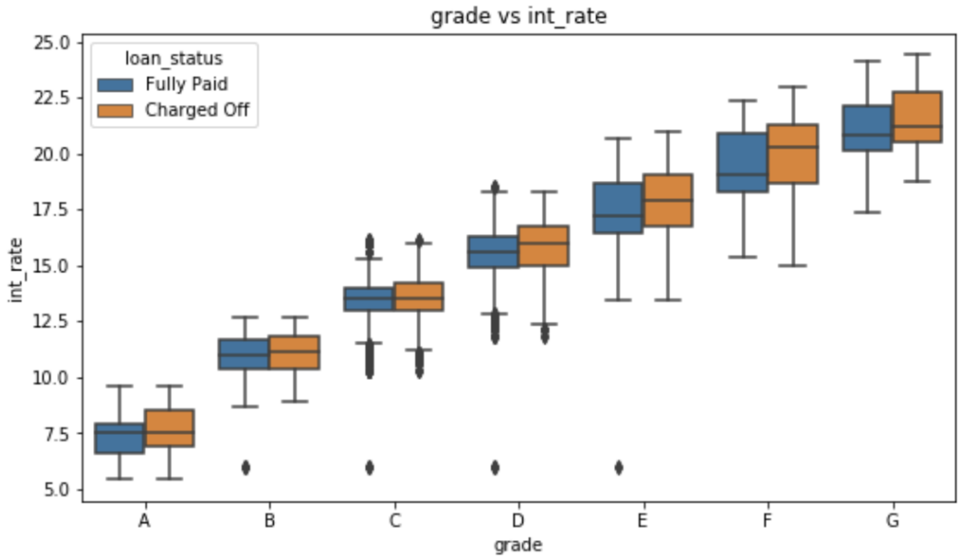
Term vs Issued month

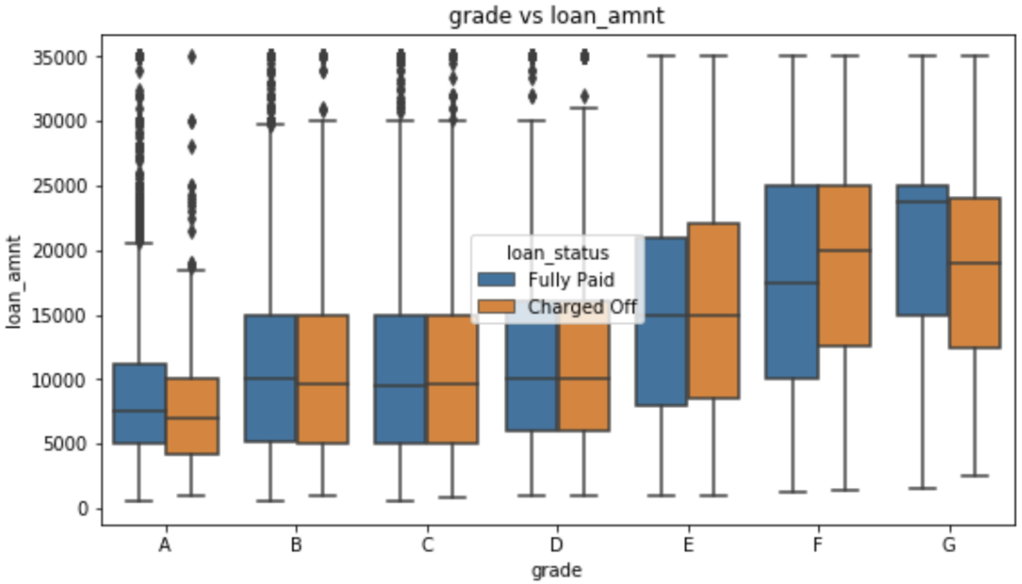


Throughout the year, there is a linear increase in the no of 36 months loans while for 60 month progression can be seen almost as a constant

Grade vs Interest Rate

Grade vs Loan Amount





As grade goes from A to G, the rate of interest interest increases linearly and so does the defaulters

- From A to G loan amount increases.
- B,C, D & E the defaulters & fully paid are almost same

Factors affecting loan status and preventing credit loss

1. Rate of interest
2. Employment length
3. Annual income
4. Loan term
5. Grade
6. Loan amount
7. Number of inquiries by borrowers
8. Number of public records of the borrowers

Recommendations

Inference	Resolution
Borrowers with higher incomes have lower default rates	Verify income levels & consider stable income in loan approval processes
Longer employment duration correlates with lower default rates	Sanction more loans to more experienced borrowers
Longer loan terms often increase the likelihood of default due to prolonged repayment periods	Emphasize on shorter period loans more
Higher interest rates correlate with higher charged offs	Offer competitive rates to lower-risk borrowers. More frequently access high interest loans
Larger loan amounts tend to have higher risk of defaulting	Implement stricter credit checks or take guarantee as assets for higher loan amounts.
Focus on grade A more than F,G & H	Try to give more grade A loans to the borrowers
As the inquiry increases the percentage of defaulters per inquiries also increases	Regularly Check for inquiries done by borrowers before giving out on loans
For 1 or more than 1 public records, the percentage of defaulters is more than 0 public rec	Less number of loans should be given to such borrowers