

LENDING CLUB CASE STUDY

-Analysis using Exploratory Data Analysis

By:

Umesh Kumar

Mydhili

Problem Statement:

- * Lending Club is a financing company that funds loans for a varied type of purposes for a wide range of audience in most of the US regions.
- * Analyze LC borrowers behavior i.e. if the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company OR if the applicant is **not likely to repay the loan**, i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

Objective:

- * We have a dataset with details of borrowers and their loan attributes which can be used to draw few observations on the who charged off (unable to repay)
- * What are the factors which might affect them being defaulted.
- * This enables LC to make certain decisions and being more cautious while evaluating the applicant's profile when they apply for a loan.

Approach:

*Data cleaning:

- Remove null value columns and rows.
- Remove outliers (e.g. annual income)
- Standardized row values (e.g. interest rate %)
- Replace null row values with mode (e.g. employee length)

*Derived Metrics:

- Added new columns for better analysis (e.g. Loan issue date)

*Univariate/Bivariate Analysis:

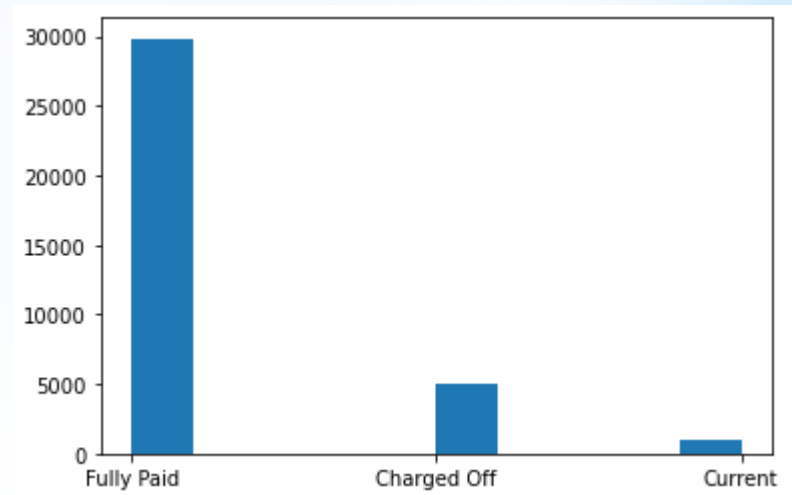
- Different plots to analyze data visually (e.g. hist, bar, pie)

*Conclusions:

- Observations made with plots behavior

Charged off vs fully paid borrowers

- This histogram shows ~14% of the people charged off (defaulted or unable to repay) among all the borrowers which is ~5k of them.
- A major chunk of the borrowers were able to repay amount on time and ~83% of them paid the loan amount in full
- Current status is not useful for us in this analysis since weren't sure whether pay the amount in full or they would default.



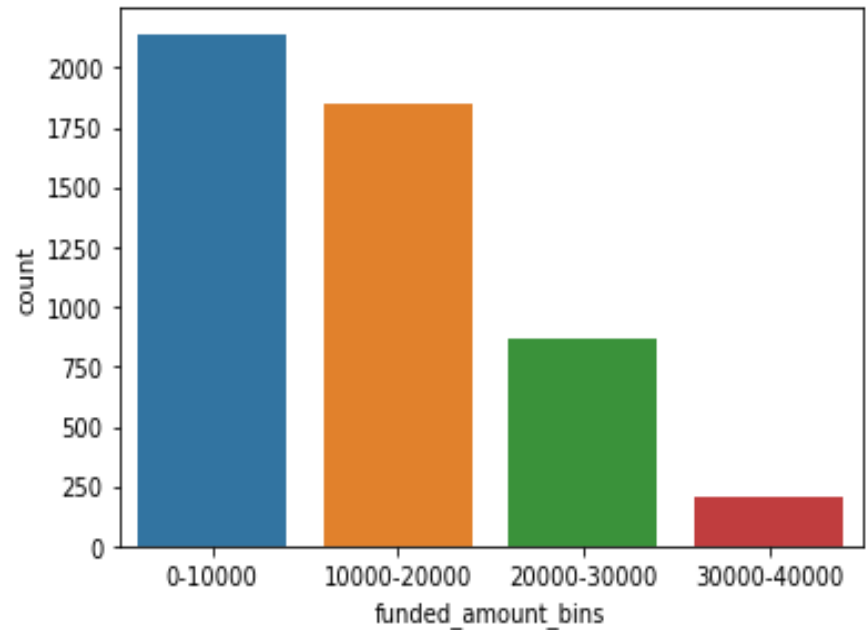
Fully Paid 0.830189

Charged Off 0.140985

Current 0.028826

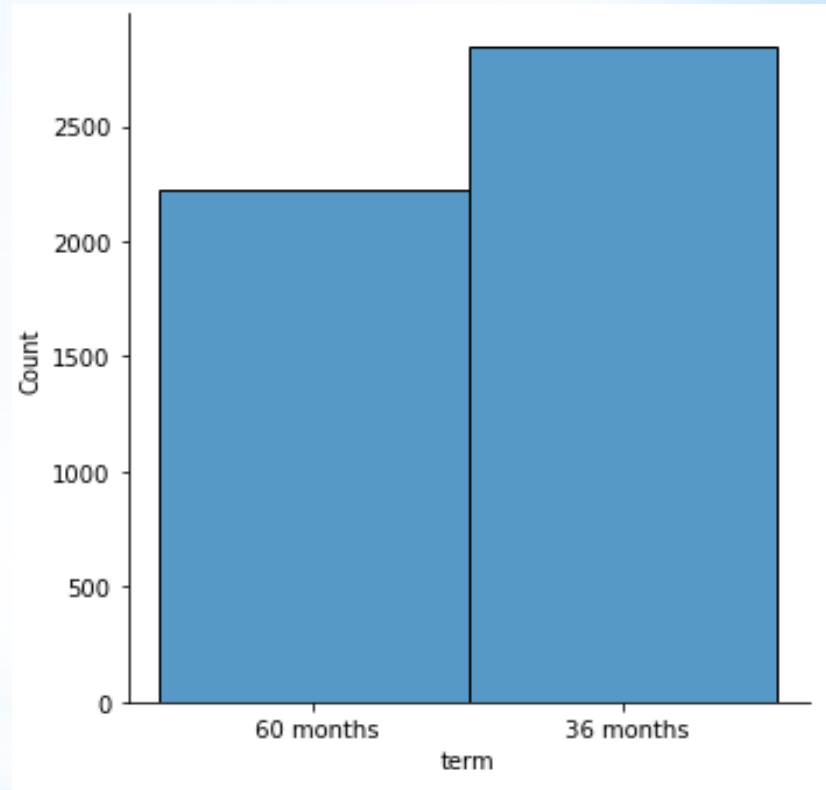
Charged off vs Funded amount

- Loan amount is the amount of loan applied by the borrower and funded amount is the amount we assume is the amount sanctioned.
- So, we have considered funded amount for column. (based on assumption)
- The plot shows, borrowers with lesser funded amount have charged off more.



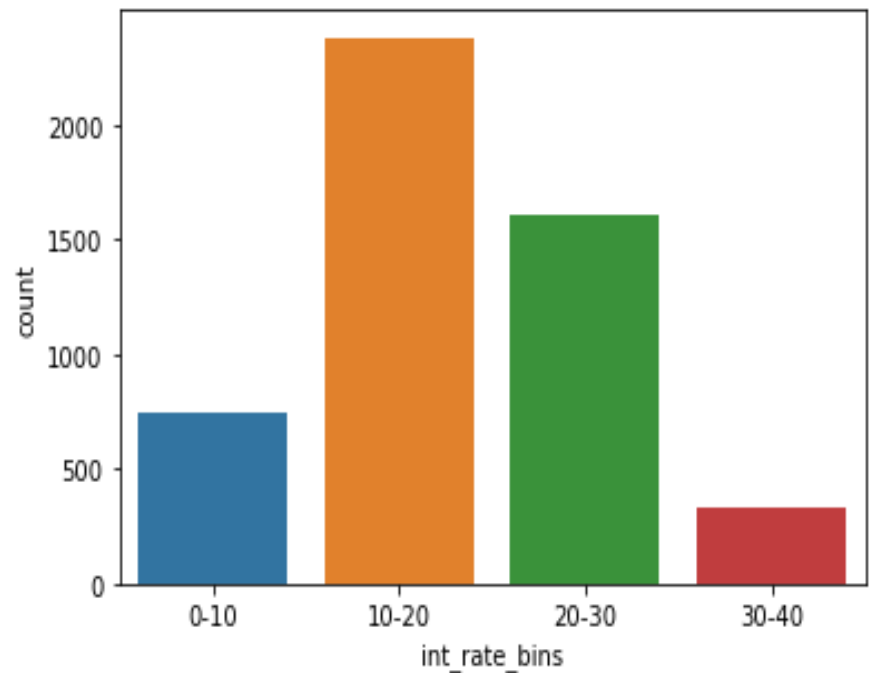
Term vs charged off

- Term is the duration of loan repayment period of borrowers, It is 2 terms here, 3 years and 5 years.
- This distribution plot shows, borrowers with 3 years loan repayment term charged off more than that of whom with 5 years term.



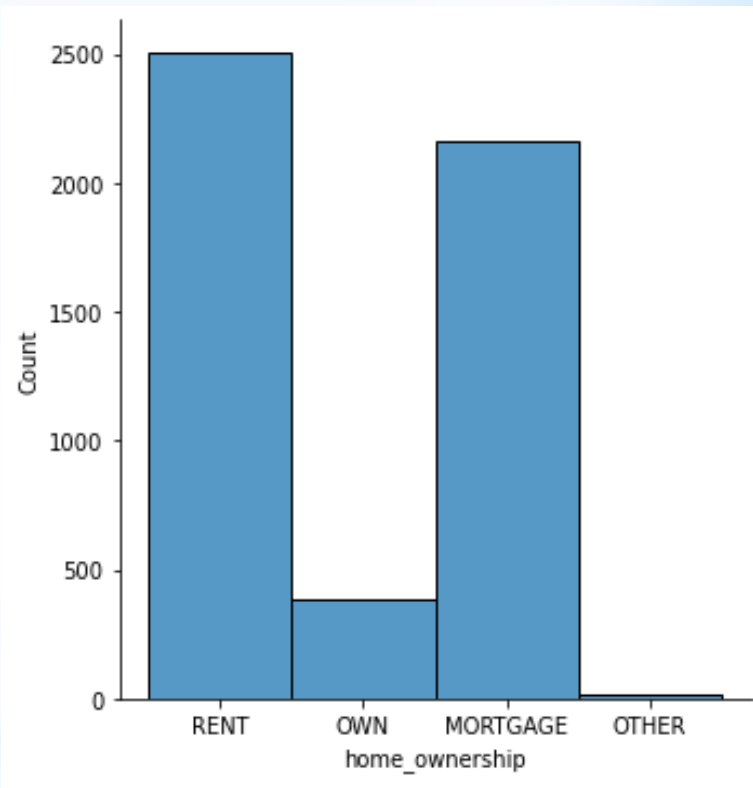
Interest rates vs charged off

- Interest rate is the percent of amount a borrower is imposed on the amount sanctioned.
- When interest rate is more than 10% charged off count is much higher, so Lending Club should be looking for amendments on the interest rate and to keep it as lower as possible.



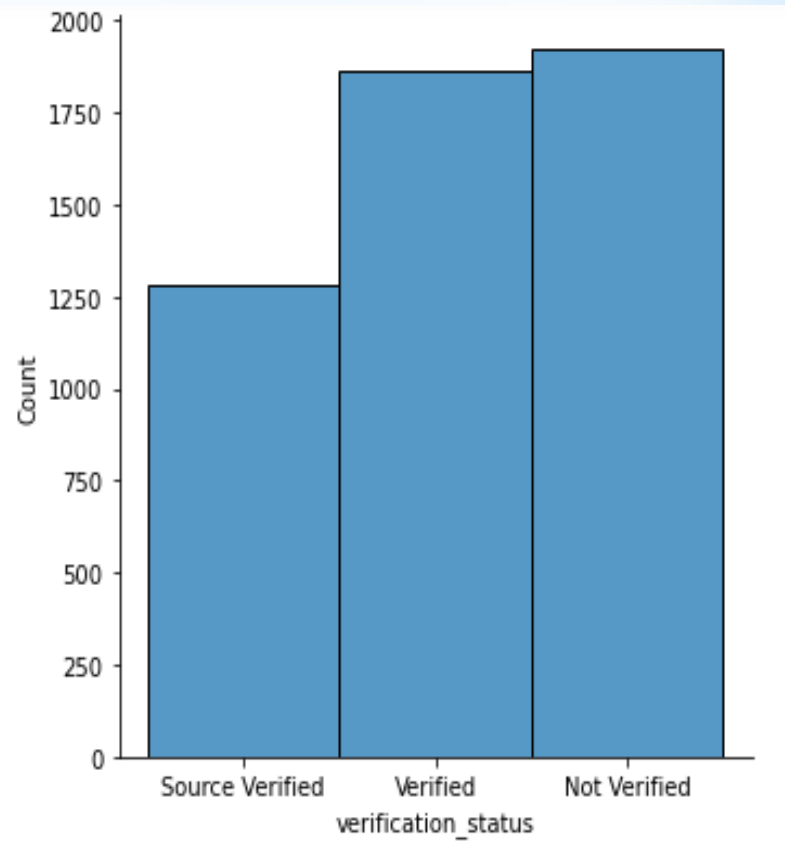
Home ownership vs charged off

- Home ownership is the status whether the borrower owns a housing property or are they on rent or on mortgage.
- This plot shows, borrowers who own a property charged off very much lesser than who rent or on mortgage.
- So, it's will be better if LC looks at the assets borrowers own when they issue loan, which might indicate financial status of them.



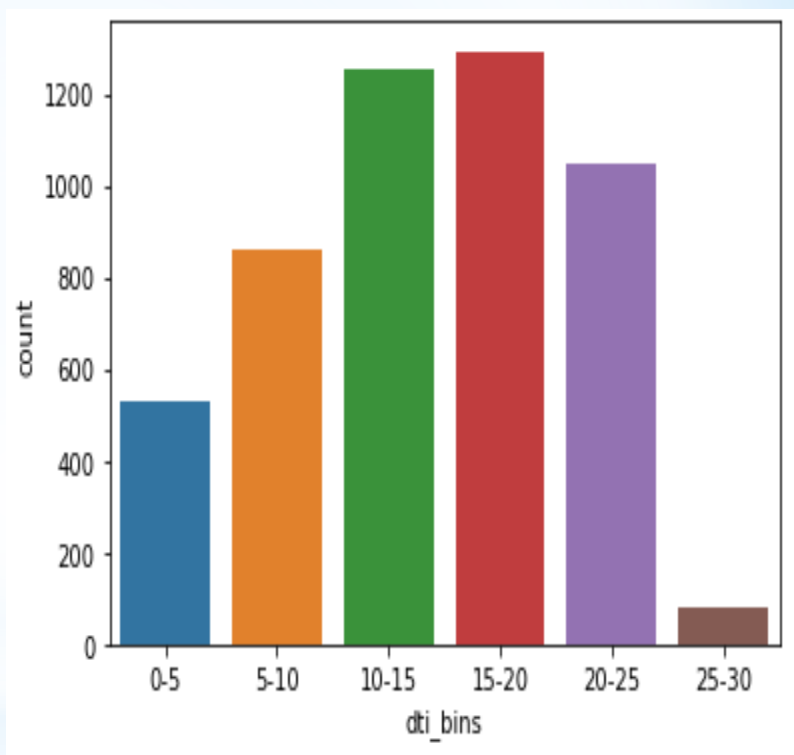
Verification status vs charged off

- Verifications status is assumed to be the status of the verification done by the Lending Club while issuing the funds for borrower.
- We assume Verified and Source verified as same and their combined count is greater than that of not verified status, so people whose background verification status is marked as verified, they charged off more which means the verification processes should be done properly by LC going forward.



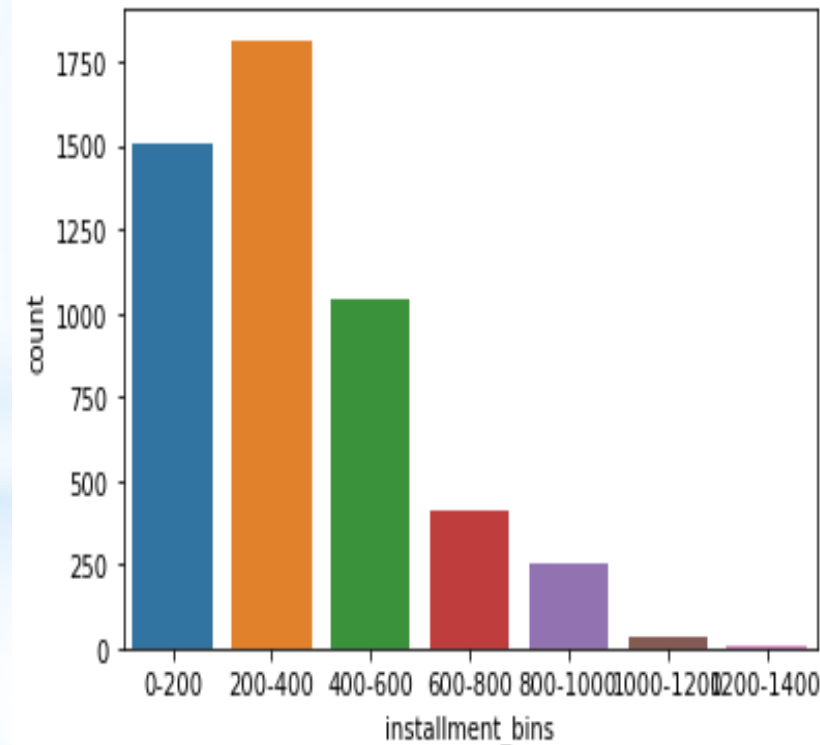
Debt to income ratio vs charged off

- DTI is the debt to income ratio which means the amount of outstanding debt a borrower has to pay to income that is annual income.
- When dti ratio is higher charged off rate is also higher, so when the dti ratio is lesser which means loan amount should be sanctioned in lower amounts based on their annual income.



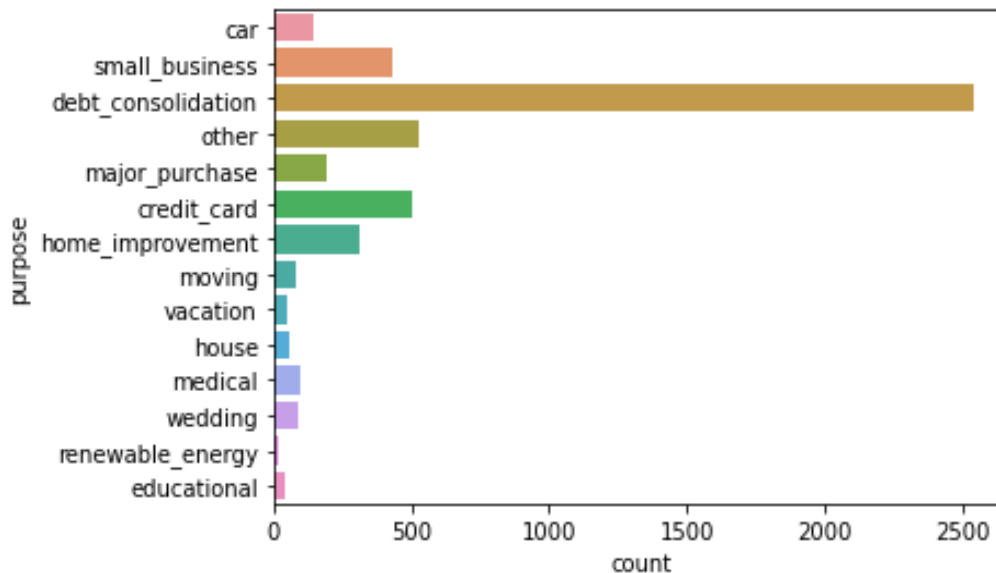
Installment vs charged off

- Installment is the monthly amount being re-paid by borrowers.
- When installment amount is lower borrowers charged off more, which might be due to their negligence to re-pay.



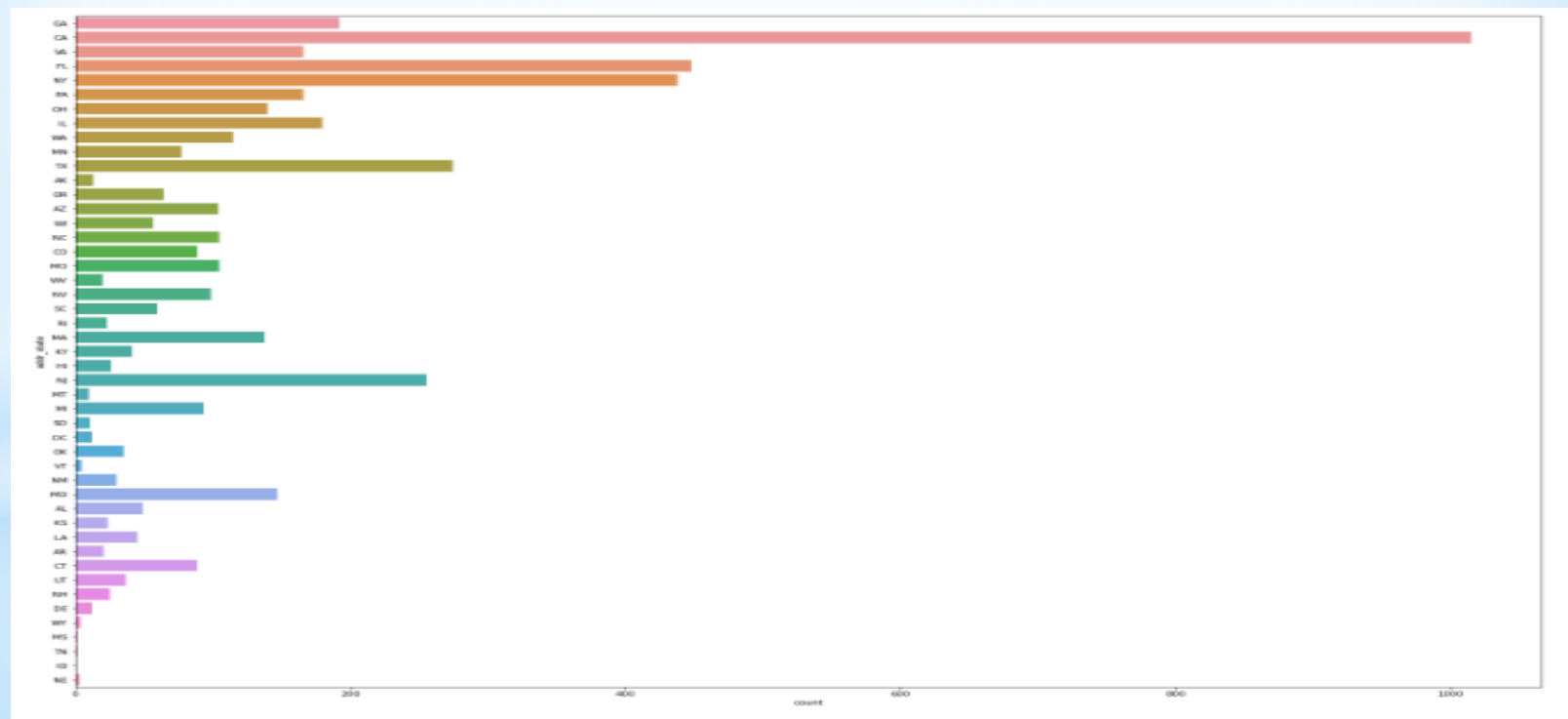
Purpose vs charged off

- Based on the below plot, the purpose of loan stated by borrower as debt consolidation, credit card, others, small business seem to be charged off the most.



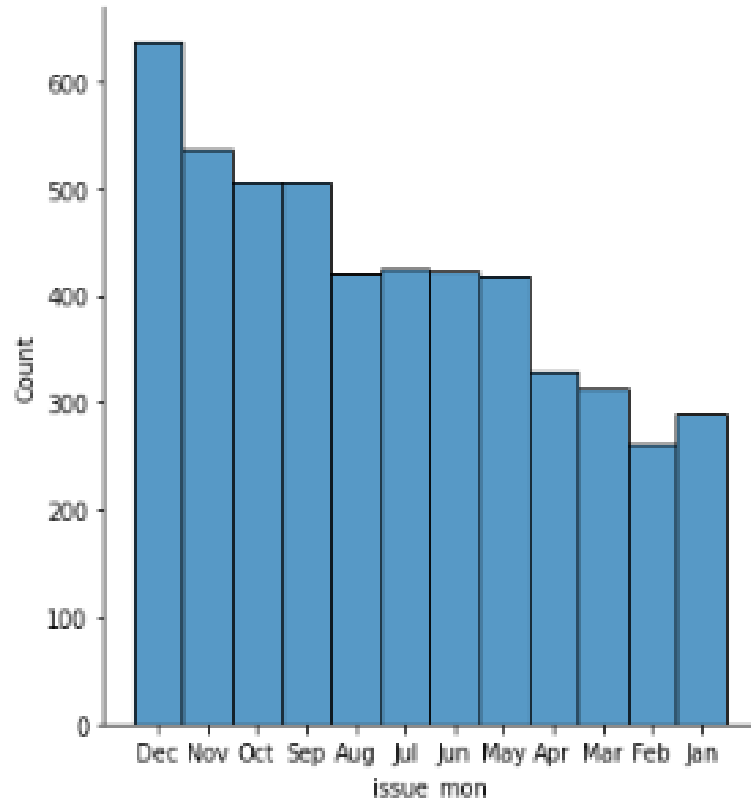
Address state vs charged off

- Based on the below plot, location of borrower like CA, NY,FL charged off to the highest, So Lending Club should be implementing more review while sanctioning loans to these location applicants.



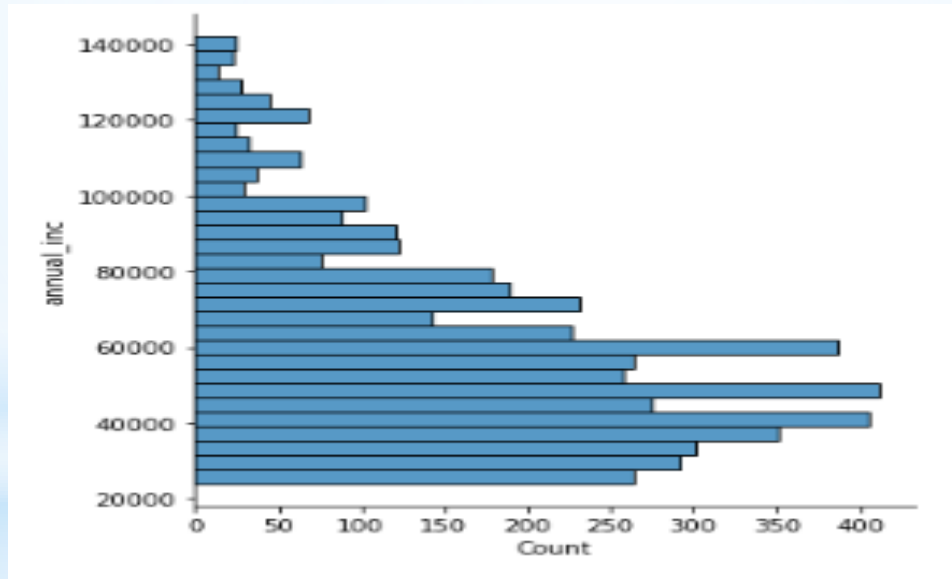
Issue month vs charged off

- Issue month is the derived column from our earlier steps, to observe the pattern of loan issuance across different months.
- We see in December month where charged off count is the highest than other months, which is the holidays month and people might opt loan for vacation. On a different note, From May to December we see the pattern as charged off higher.



Annual income vs charged off

- Annual income are the earnings of an individual yearly.
- Based on the above plot, borrowers whose annual income is lesser charged off more eg. $\leq 60k$.



Observations:

- DTI should be considered while issuing loan.
- Annual income plays a role to determine capability of borrower's loan repayment.
- Interest rates can be as low as possible to avoid borrowers being defaulted.
- Lending club might need to be cautious while issuing loans towards calendar year end.
- Location-wise verification can be sorted to make decision while sanctioning loans.
- Verification process must be improved.

*THANK YOU