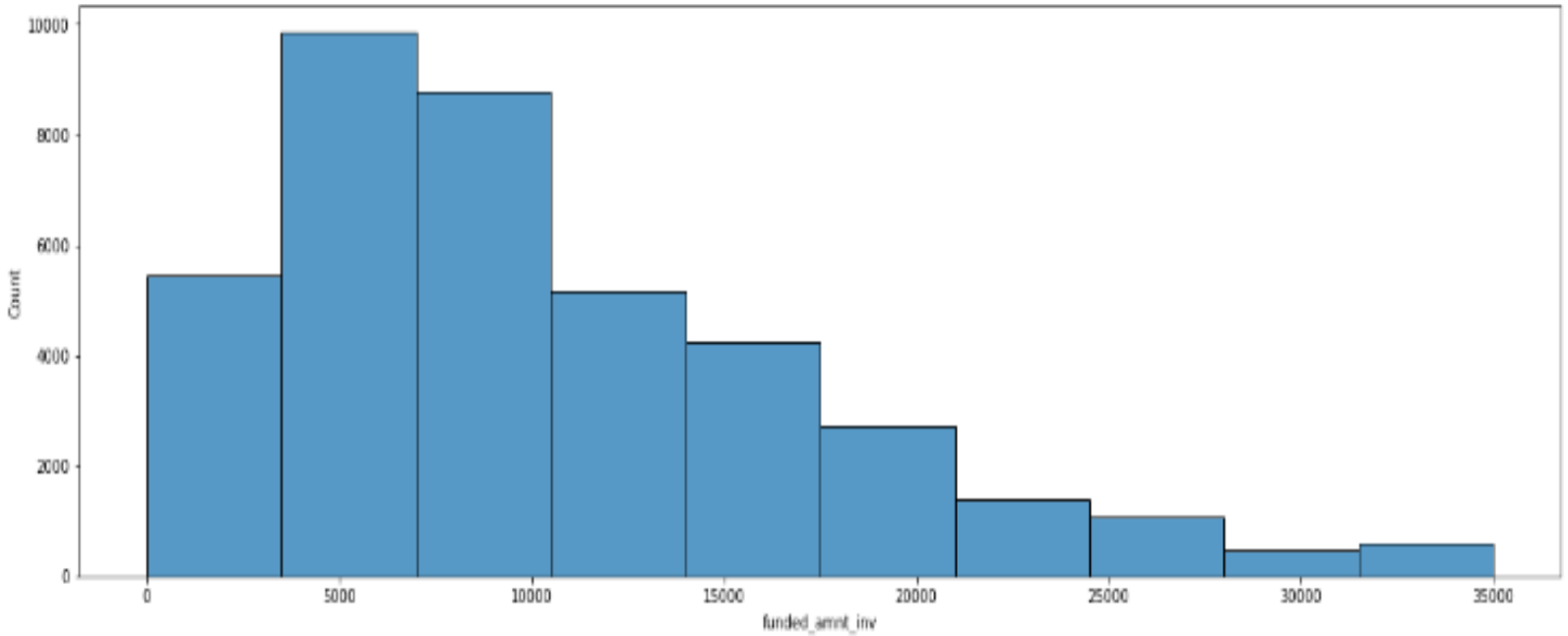
Lending Club Case Study

- Data Analysis in the case study should be done on the clean and filtered dataset so that inference/insights from the given dataset can be done more accurately.
- Loan data set has 111 columns and more number of rows where it has around 50 columns which has the value as null/NaN/NA so before we actually start with the analysis, we shall filter all those unwanted columns from the dataset.
- Out of the filtered columns , only 20+ columns are enough to perform the required analysis and infer the results.
- Driven factors/variables for the analysis are as follows :

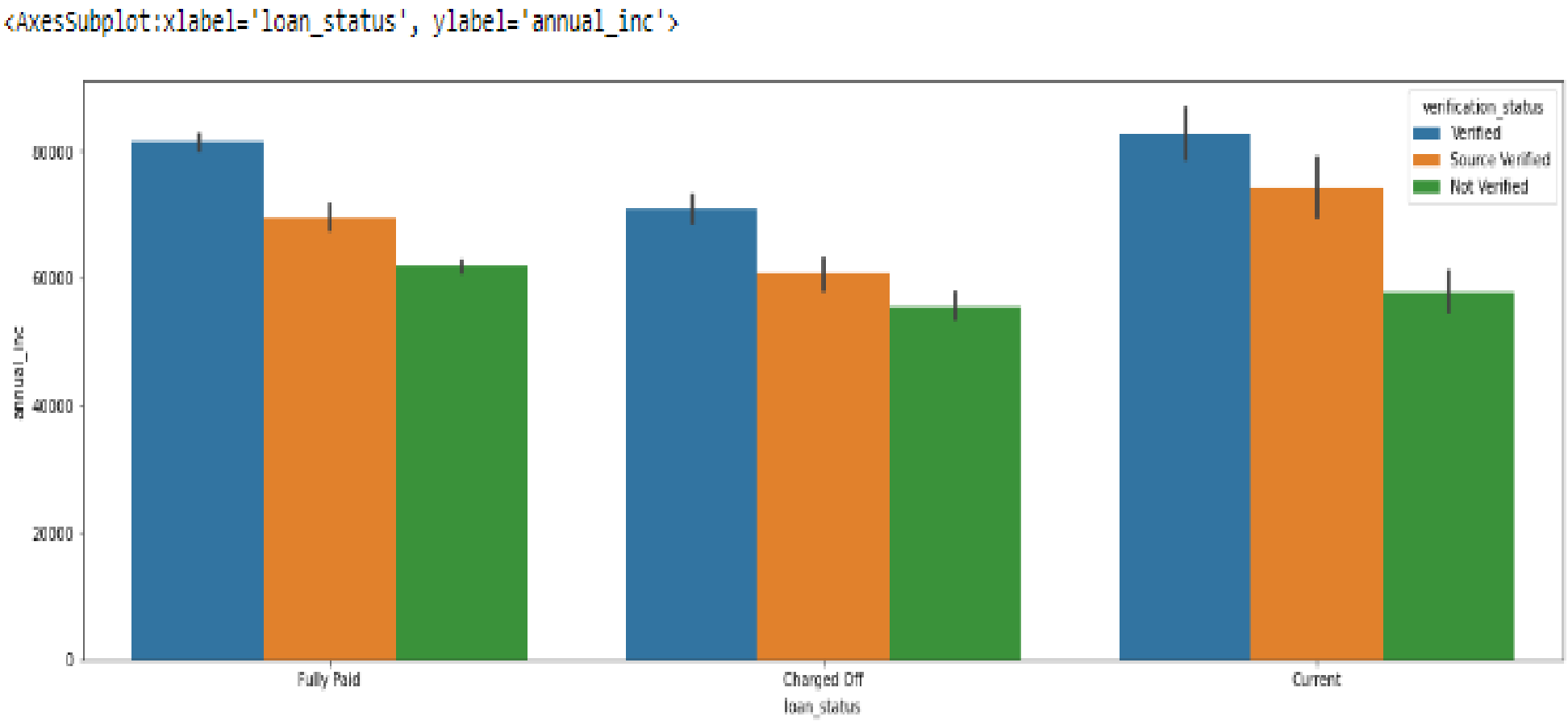
1. int_rate
2. emp_length
3. pub_rec
4. loan_amnt
5. funded_amnt_inv
6. term
7. annual_inc
8. loan_status
9. verification_status
10. issue_d
11. total_pymnt
12. recoveries
13. last_pymnt_d
14. delinq_2yrs
15. revol_util

To analyse the dataset, I have performed various univariate and bivariate analysis to get more insights from the given case. Analysis on amount invested by the investors gave the following insight:

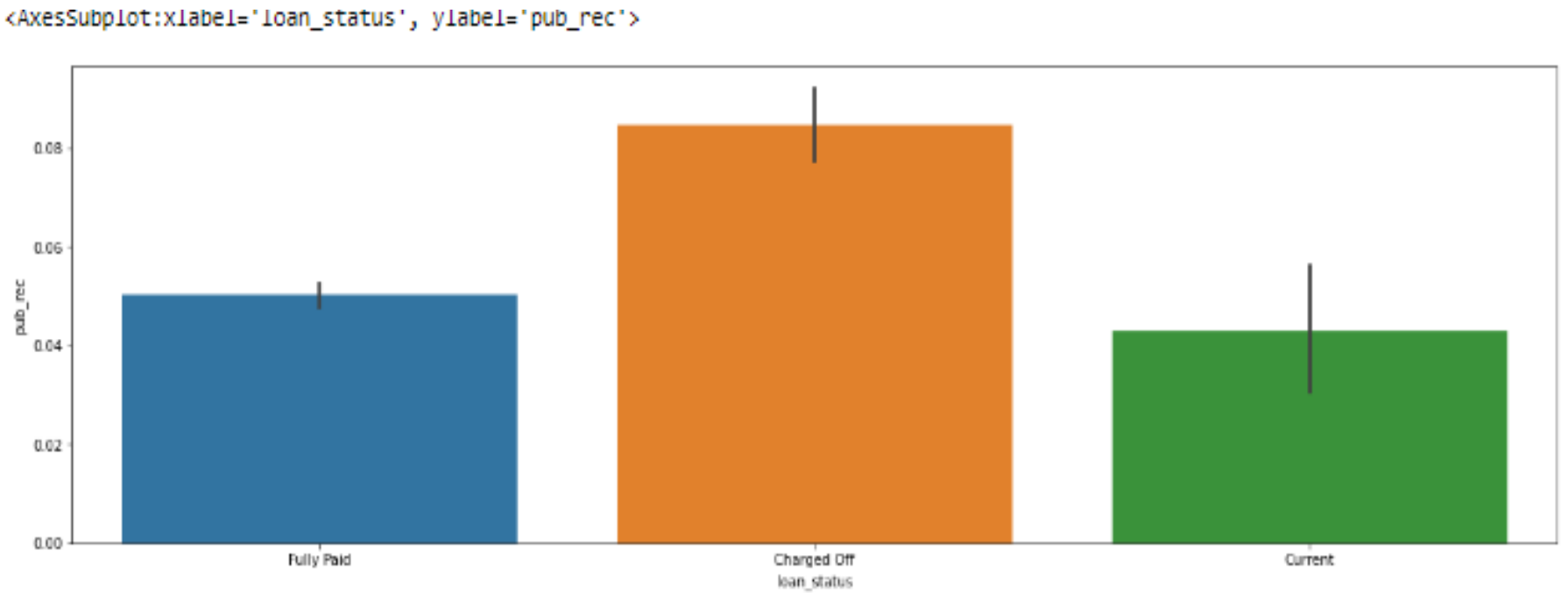


From the above plot, it can be observed that most of the investors have investerd for the loan amount between 5000 to 10000 .But still few burrowers did not lend the amount less than 5k even though the loan was requested ,investors can give the amount in this range with higher interest rate to invest and get the returns back.

Bivariate analysis on the annual income and the loan status infers that these two factors help in identifying the loan approval, rejection and also on identifying the defaulters.

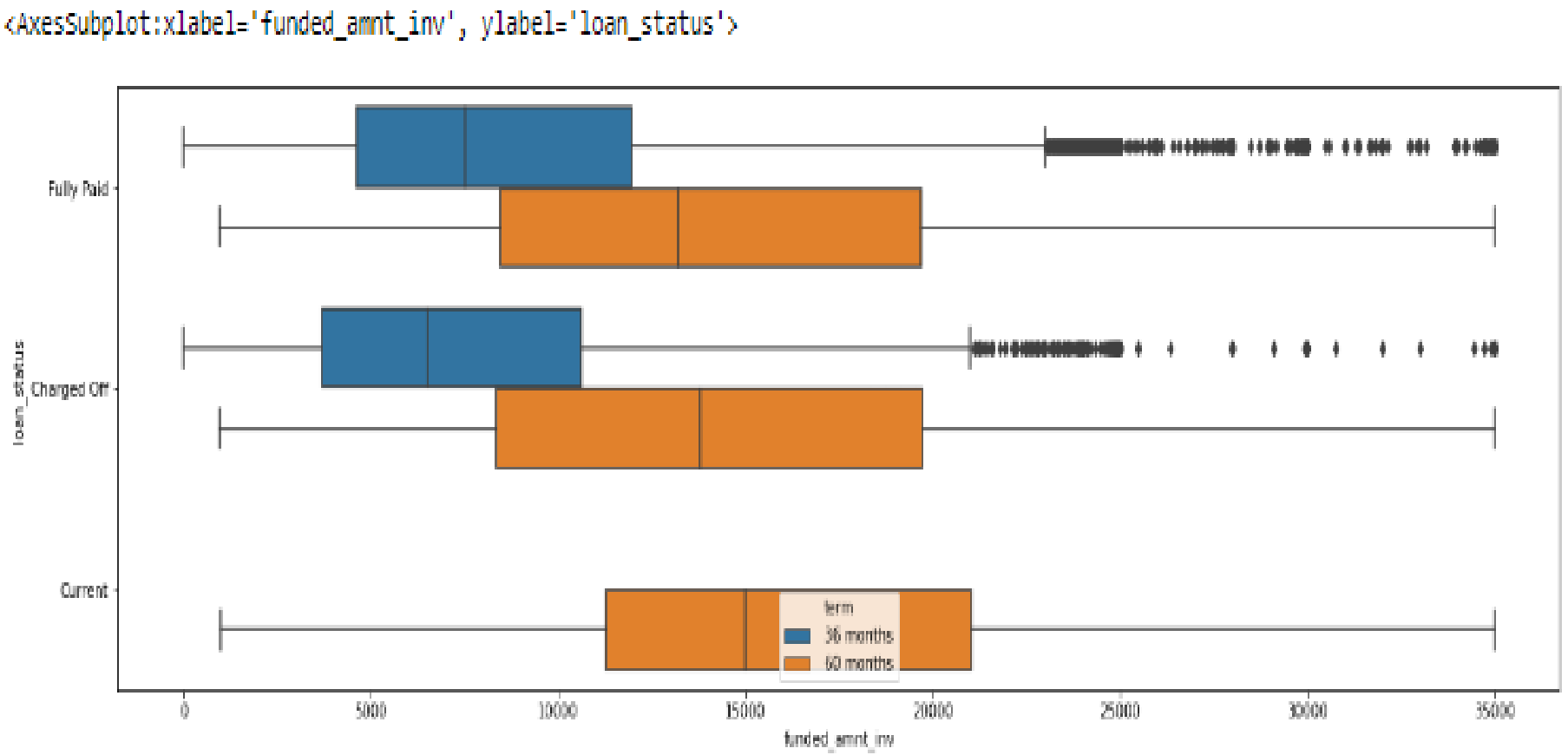


The main driving factor in identifying the defaulters is their background check and also the public records if they have any reports of bankruptcy or derogatory records. Following graph shows that people with more number of pub_rec are more likely to be defaulted so loan should not be approved for these category.



From the above plot it is clear that borrowers with derogatory public records are the most charged off and among the current payers, there are borrowers with derogatory public records so investors should check the background thoroughly before actually approving the loan so that financial loss can be avoided by greater extent

Analysis on the amount invested by the investors and the terms chosen and also the respective loan status, infers that higher the number of terms , more likely they are charged off.



Bivariate analysis on the amount invested by investors and the employment length of the borrowers, they are more likely to affect each other. Higher the employment length, more likely the loan gets fully paid and they are the most current payers.

Out[40]: <AxesSubplot:xlabel='funded_amnt_inv', ylabel='emp_length'>