

Lending Club Case Study

By - Karan S Yadav

Understand the company and the problem statement

Problem Statement

- Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who **default** cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.
- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.
- The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Company

Lending Club company is the 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.

Analysis steps

Clean

- Drop columns with null values
- Convert to proper types

Univariate analysis

- Take a single variable and perform the analysis

Segmented univariate analysis

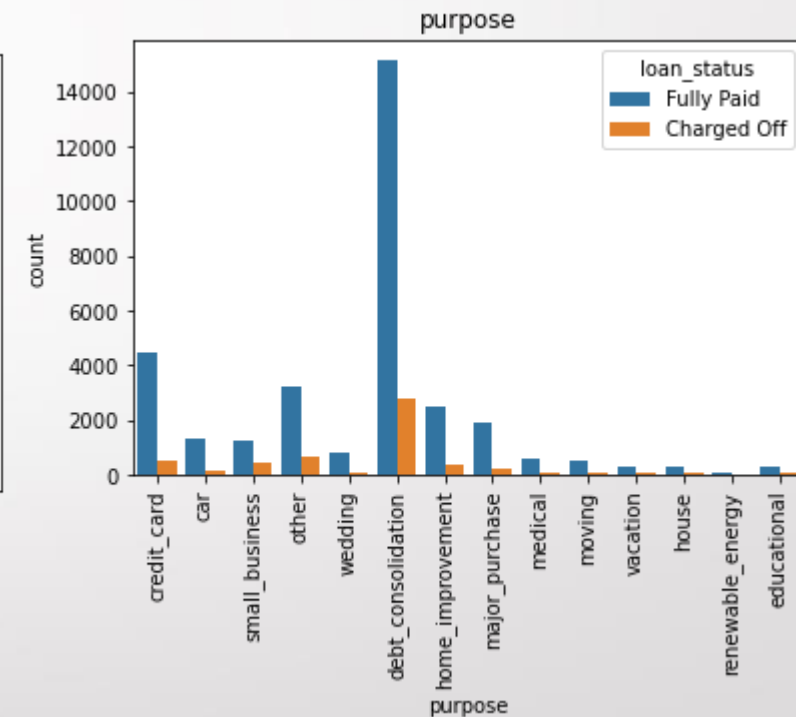
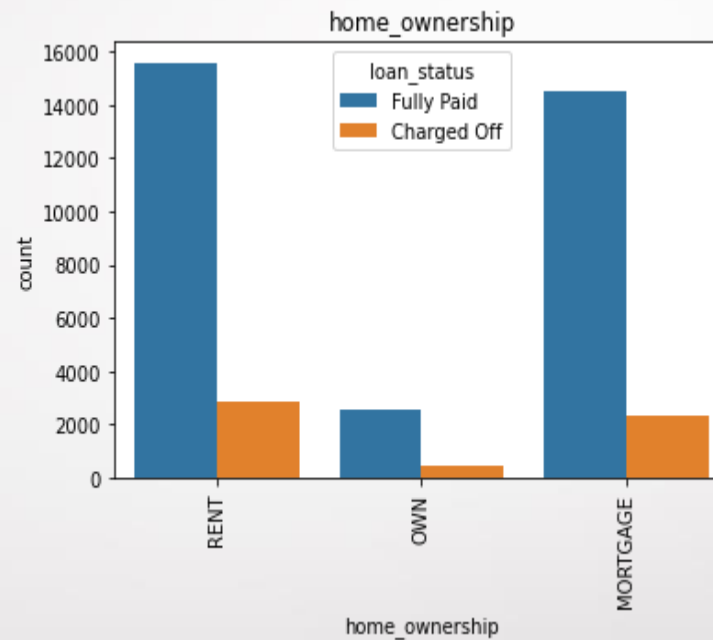
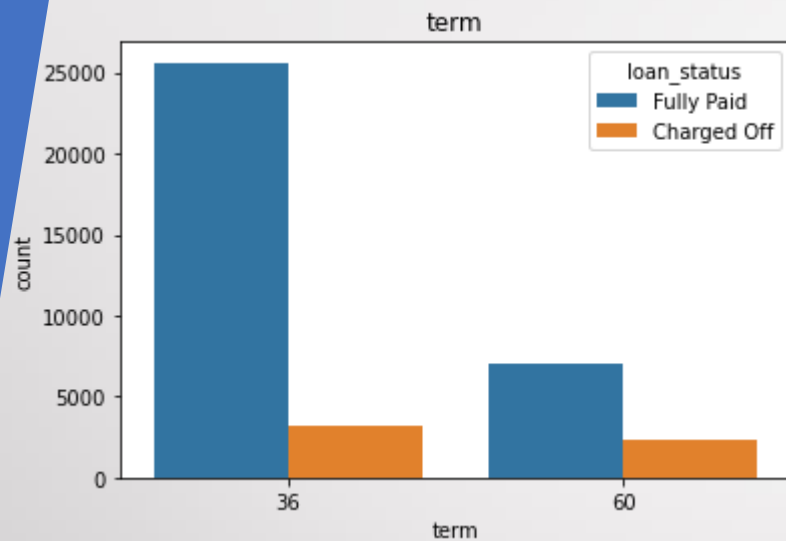
- Analyse a single variable based on other variable

Bivariate analysis

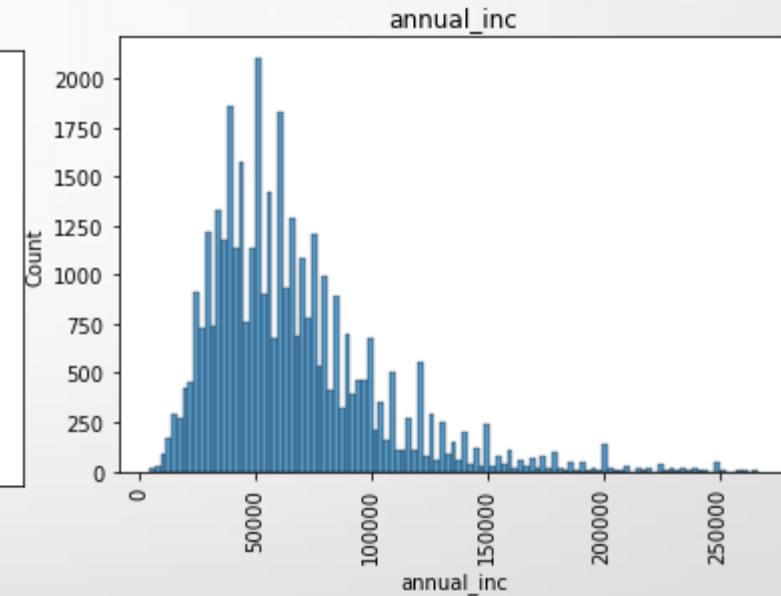
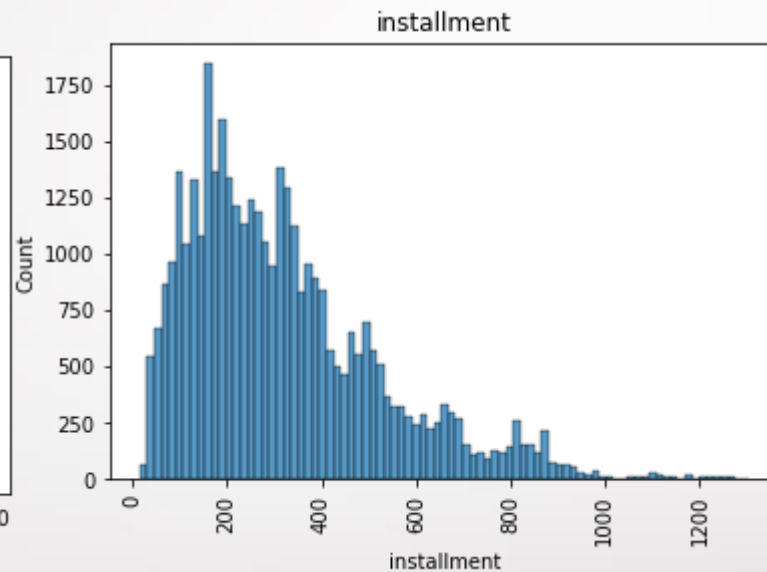
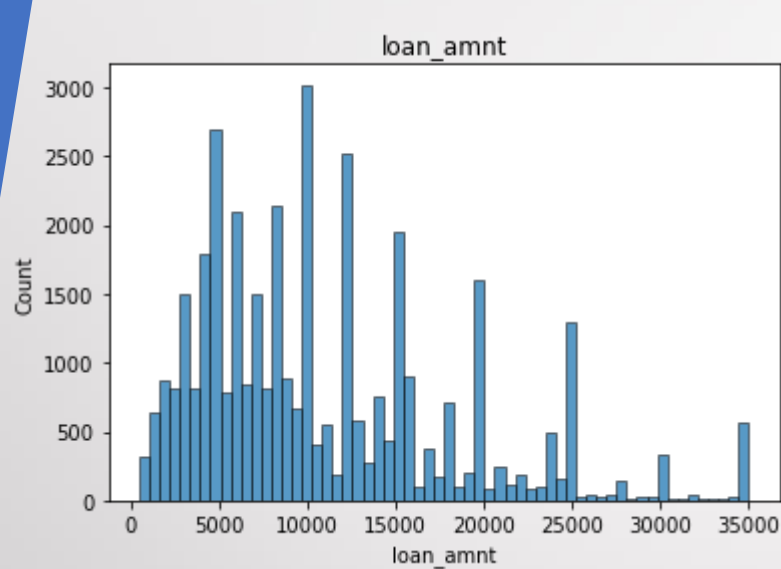
- Analyse the relation between two variables

Conclude

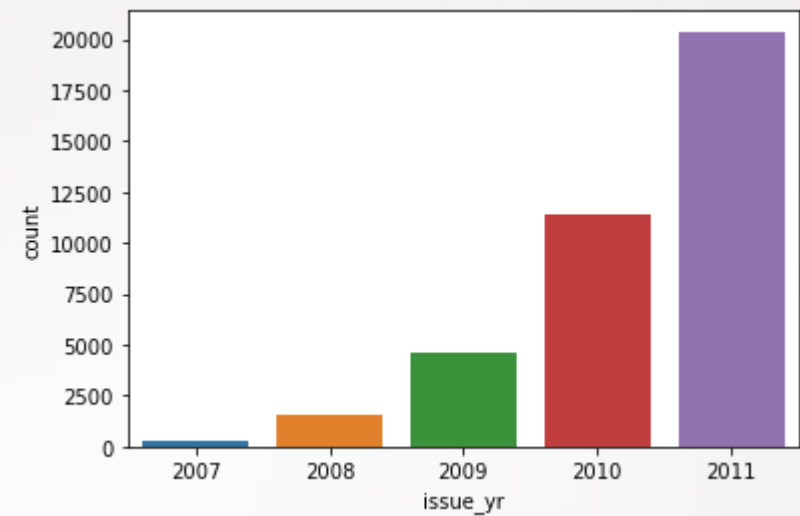
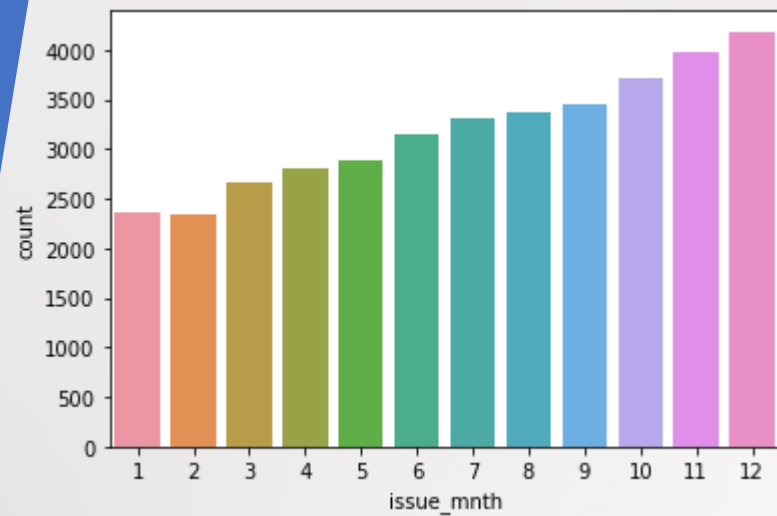
- List out the possible key observations made



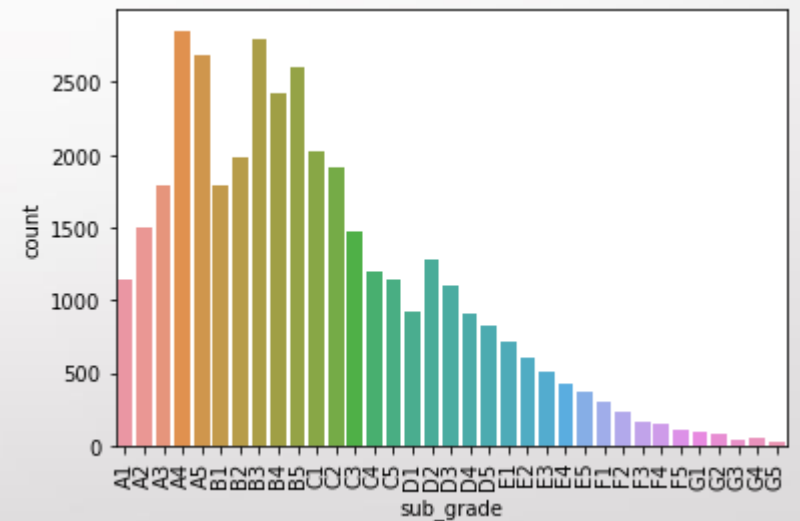
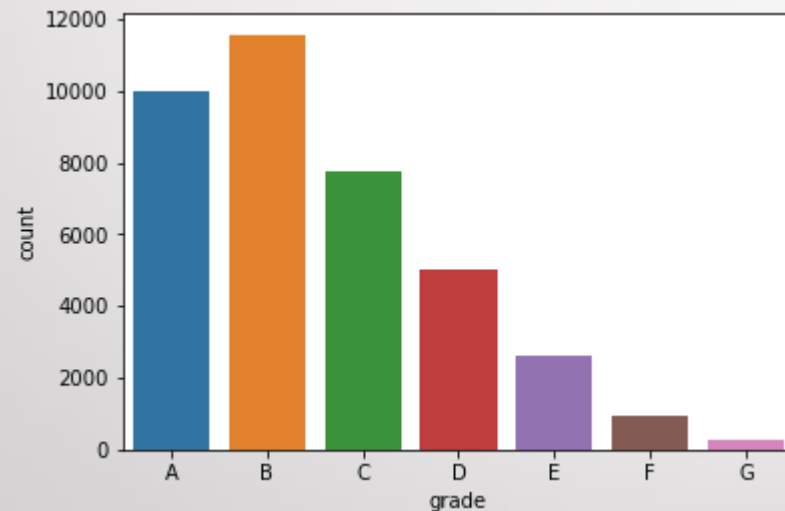
- It can be noted that most loan have a term of 36 months and even the fully paid/charged off ratio is more.
- People with OWN home tend to take less loans.
- The loan applicants are more in numbers for debt consolidation, credit card, home improvement and major purchase



- We can see there is an unusual peak in the loan amount in the vicinity of amounts such as 5k, 10k and so on
- People tend to choose lesser installments
- Most people who apply loan come in the below half of the mean annual income

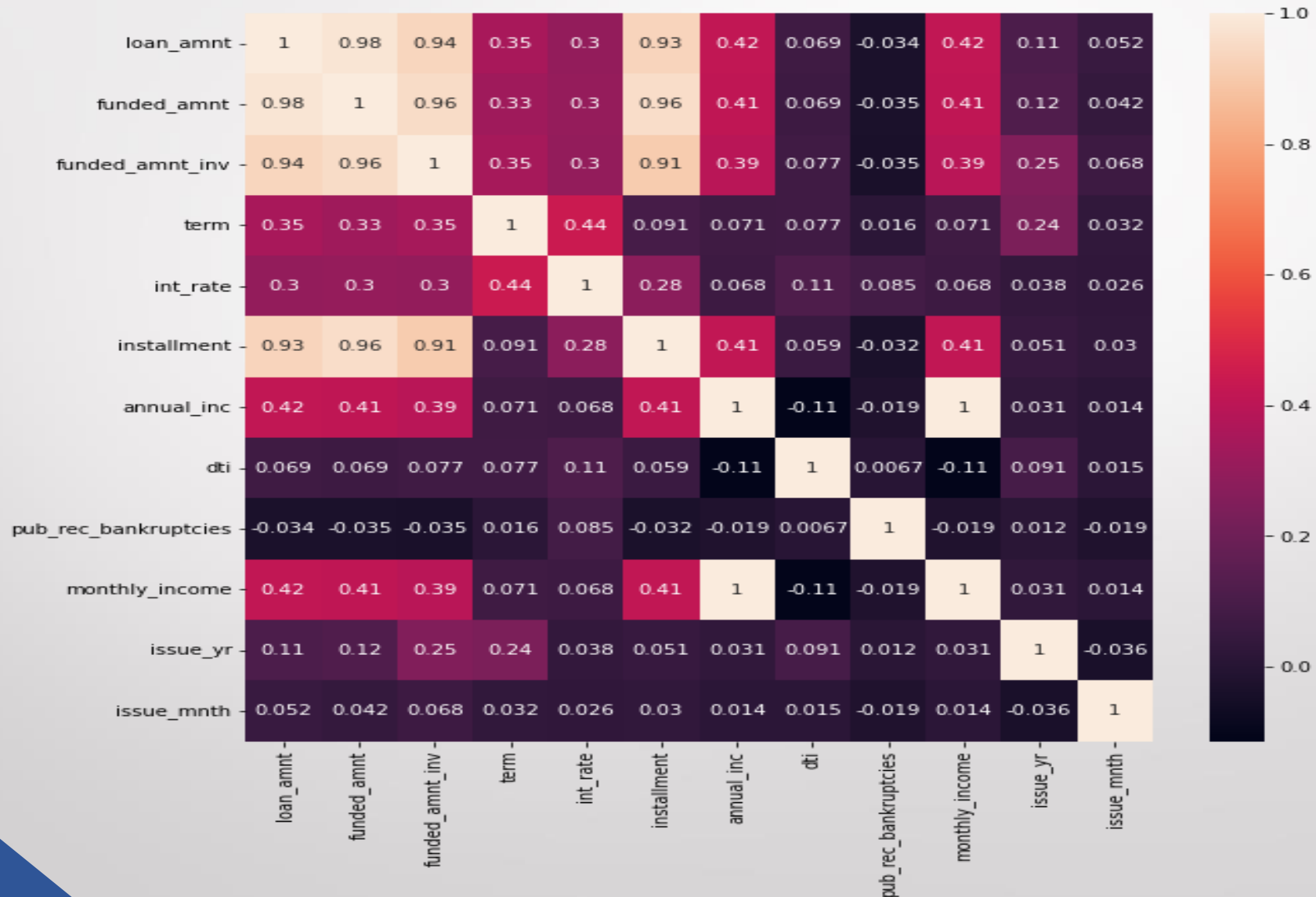


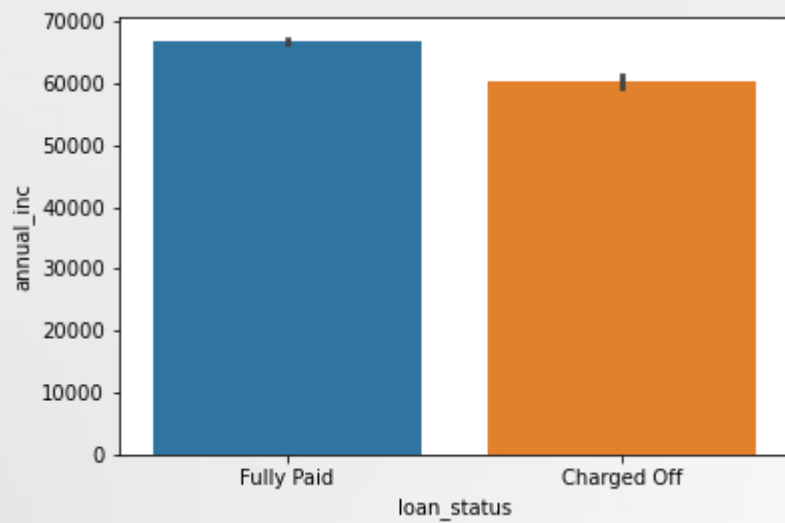
- We can see that the loan issued is increasing linearly by the month, meaning more people tend to apply for loan close to the year end
- We can see that the loans issued are exponentially increasing year after year



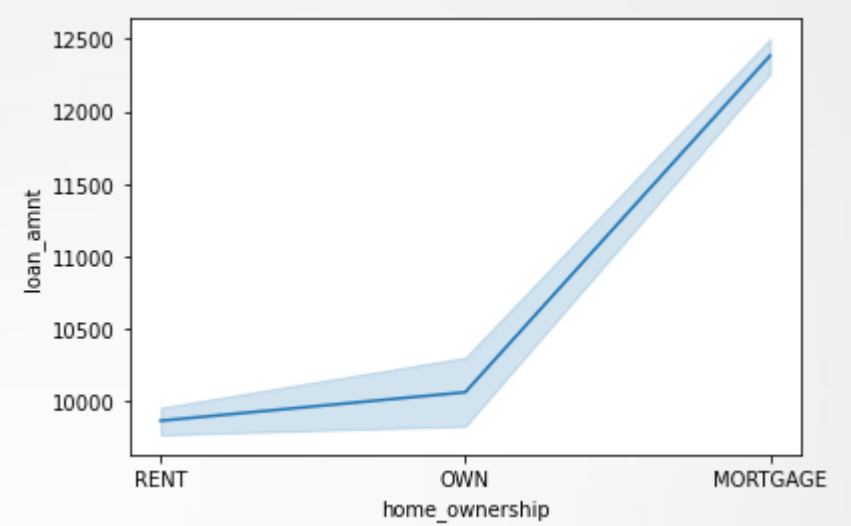
- As grade increases, the annual income also increase. Hence, the higher grades tend to apply for loan less compared to others
- As sub grade increases, the annual income also increase. Hence, the higher grades tend to apply for loan less others

- From the Heatmap, 'loan_amnt','funded_amnt' & 'funded_amnt_inv' are highly correlated. dti and annual income/ monthly income have negative correlation

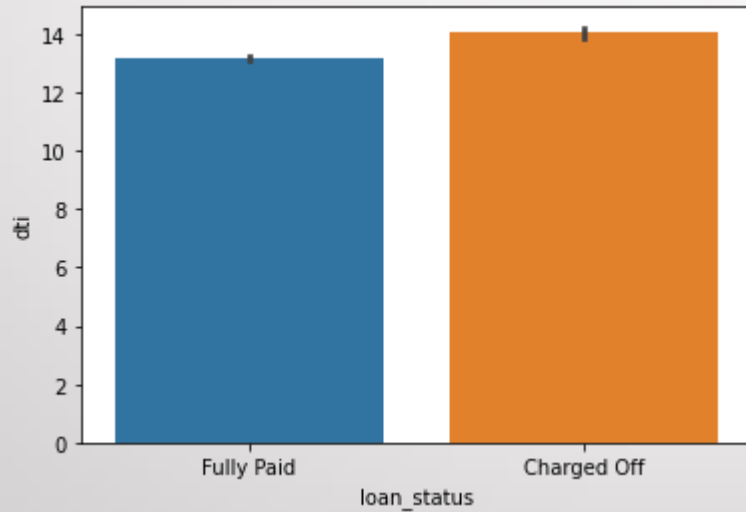




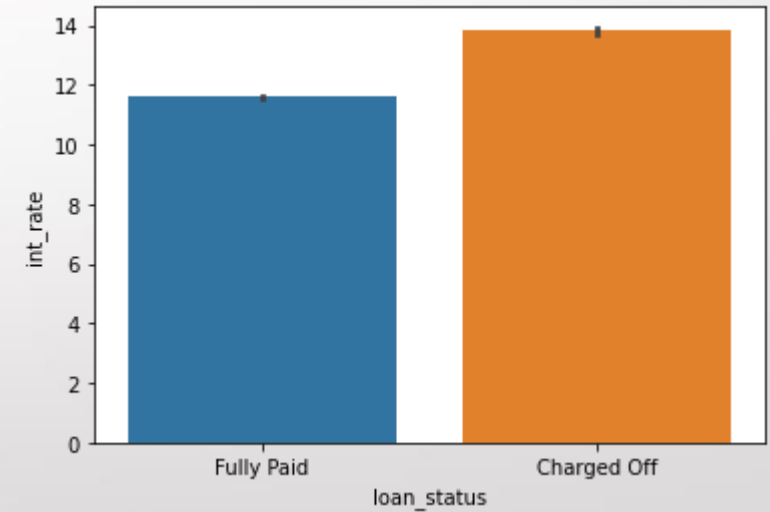
- We can see that people with higher annual income tend to fully pay the loan

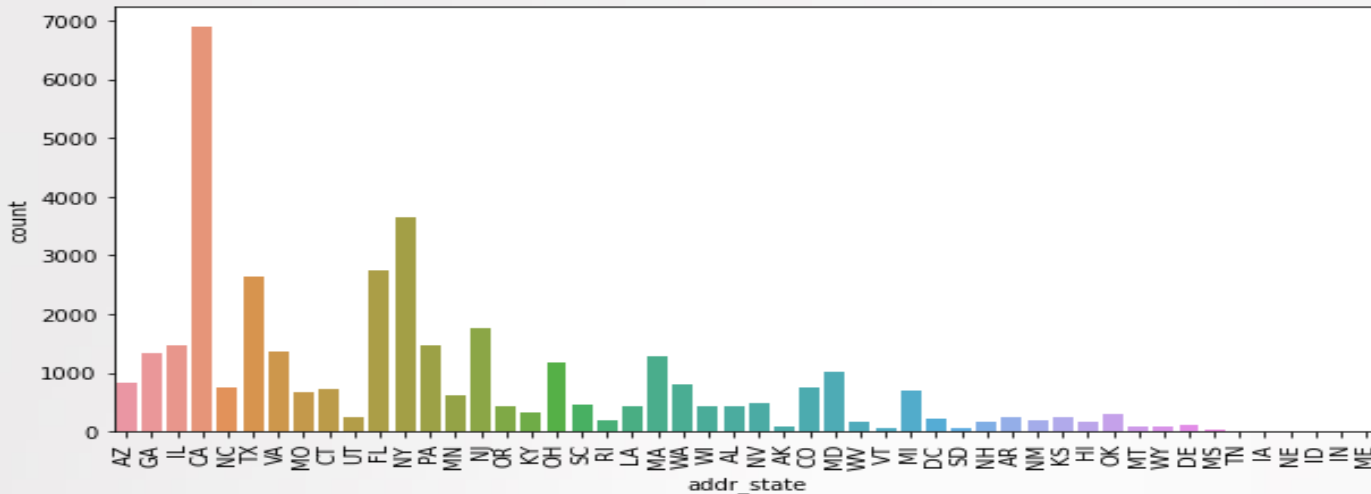


- We can see that the loan amount approved is more for Mortgage

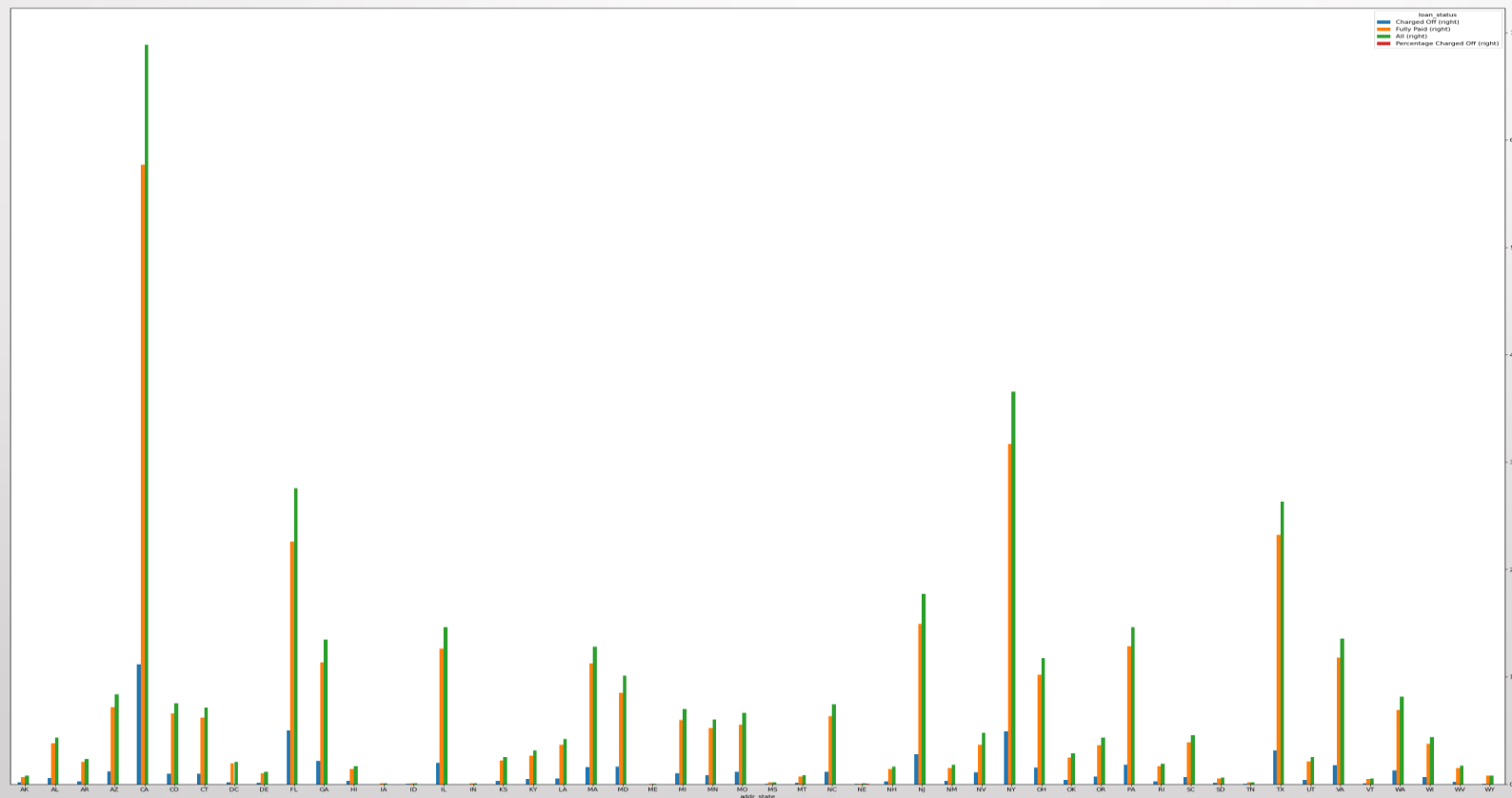


- We can see that when dti becomes more, people tend to default
- We can see that as interest rate increases, people tend to default.
- We can see that as annual income increases, people tend to fully pay the loan.

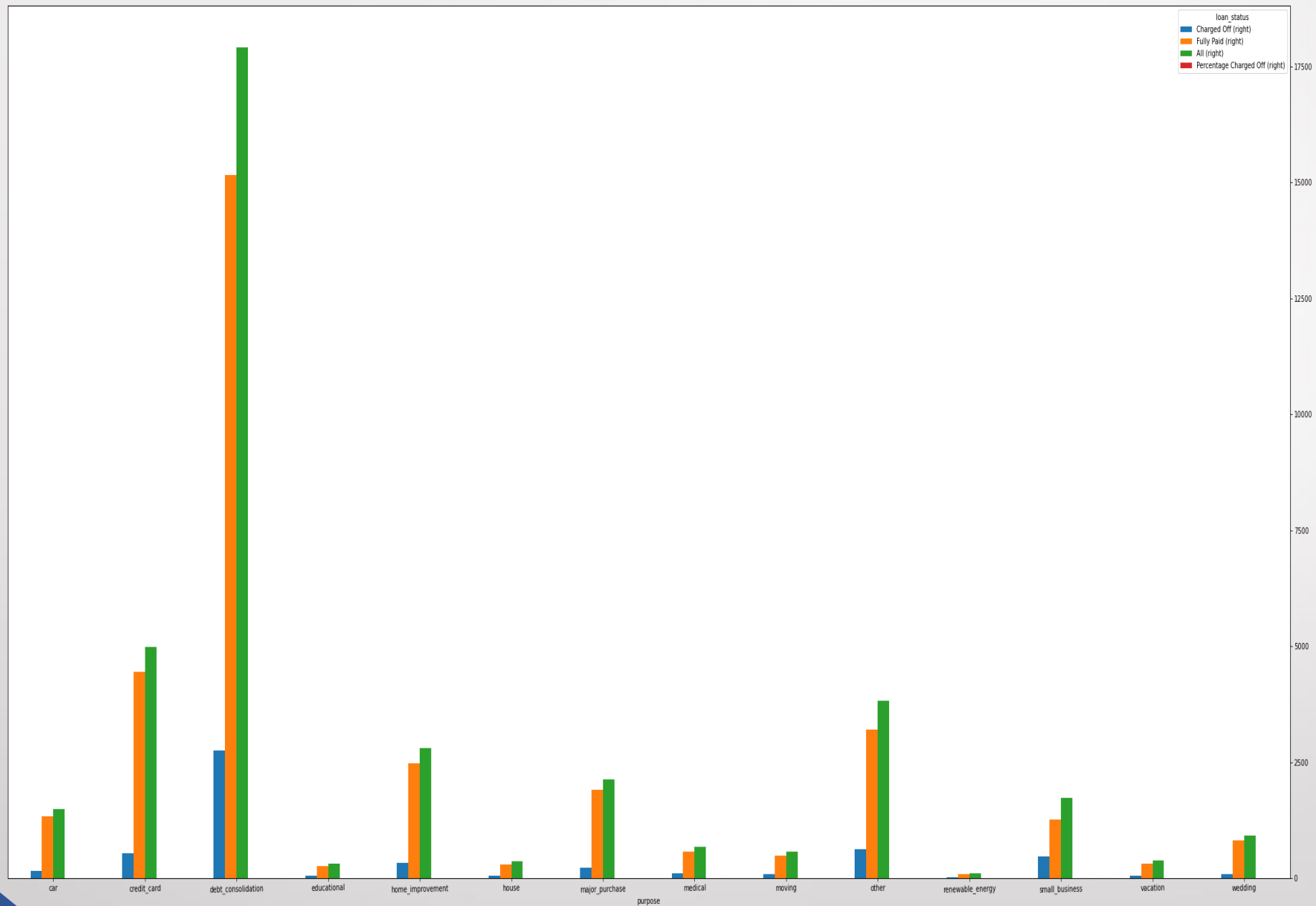




- We can see the states where people apply for loan. People in CA apply for the loan the most

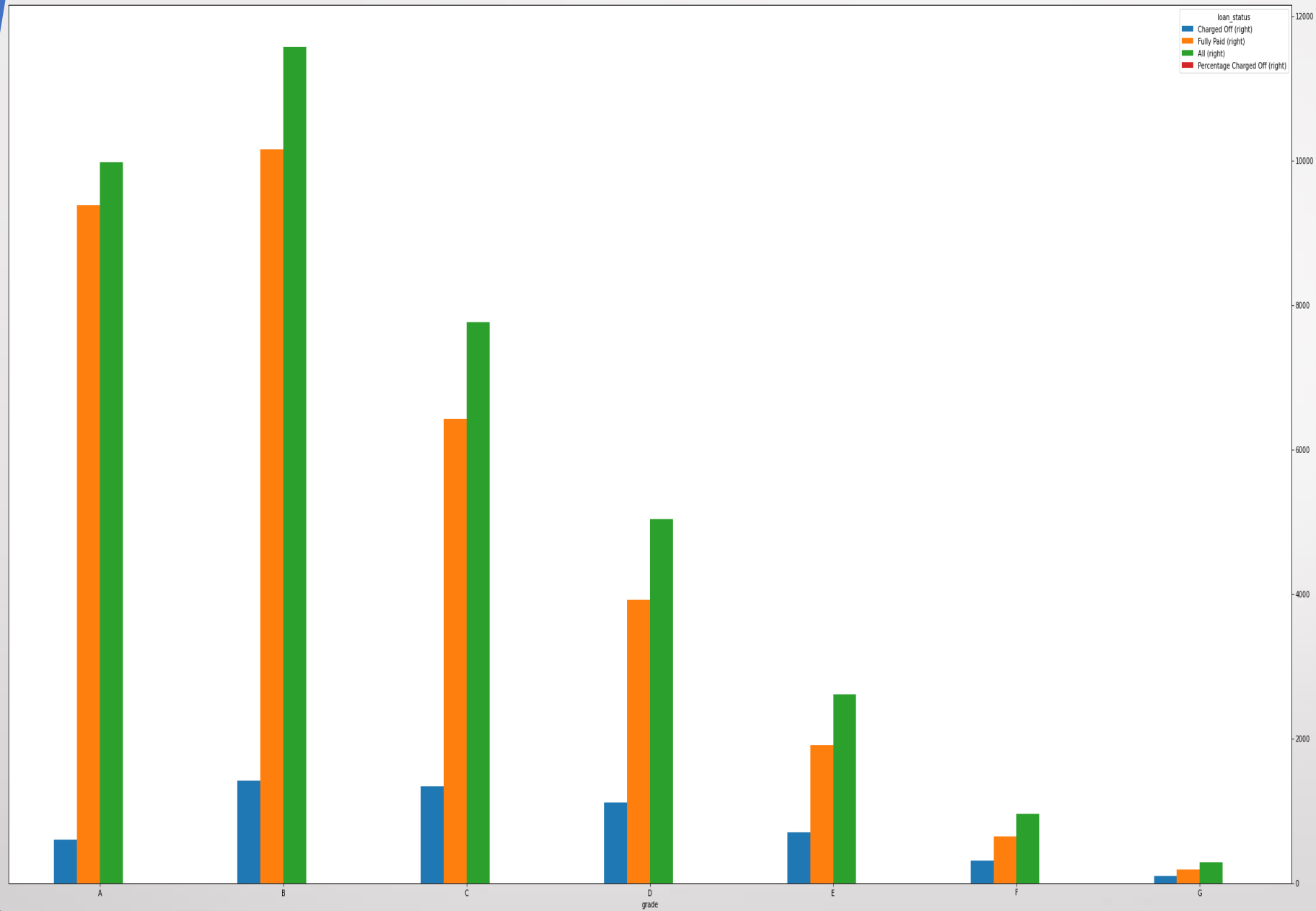


- We can see that CA has most number of people applying for loan and also the percentage charged off is relatively less.
- Places such as AK, NV, FL and SD have higher percentage of charged off

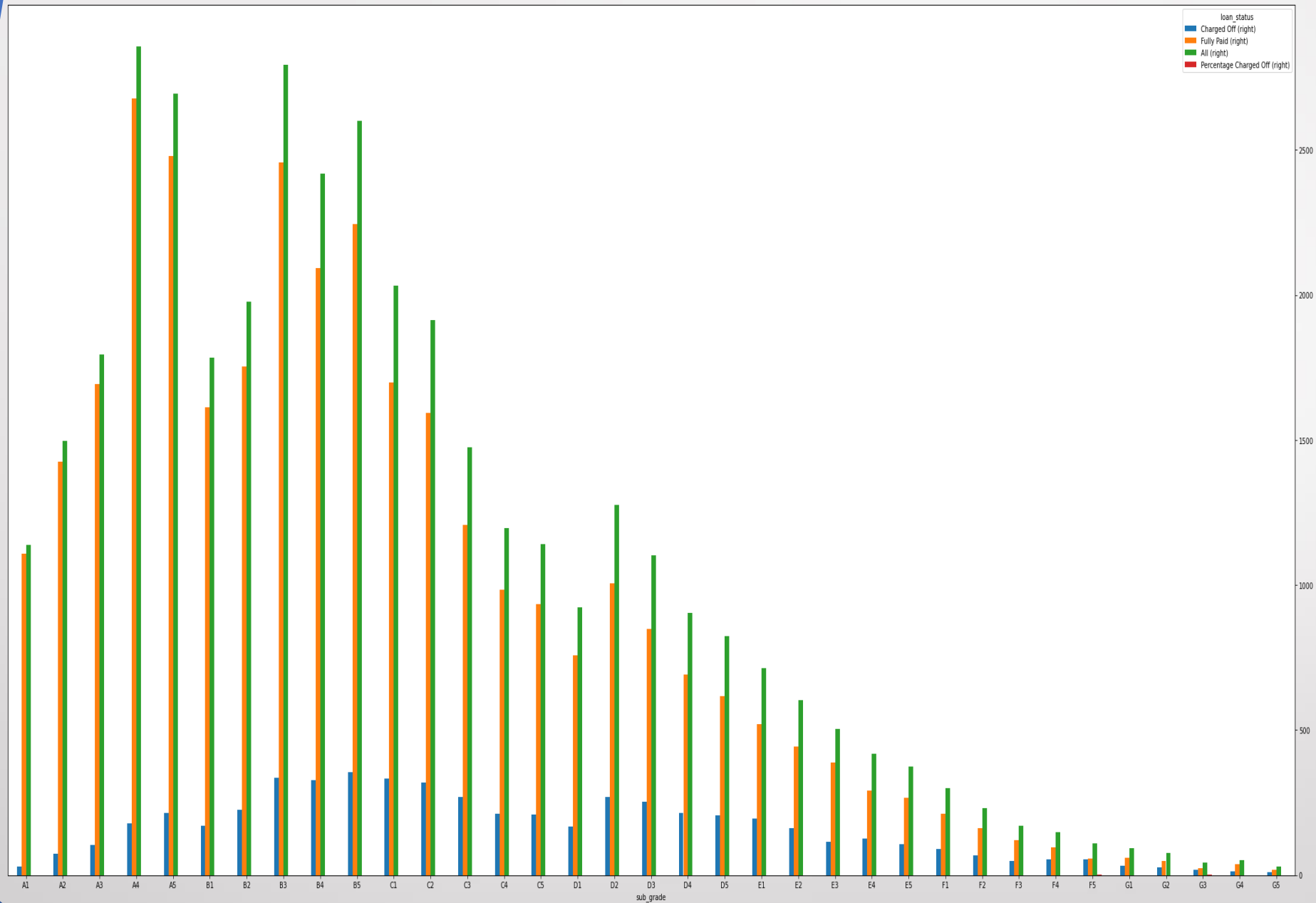


-We can see that debt consolidation is major reason for applying loan and the percentage charged off is comparatively less.

-For small business, renewable energy and education the percentage charged off is high

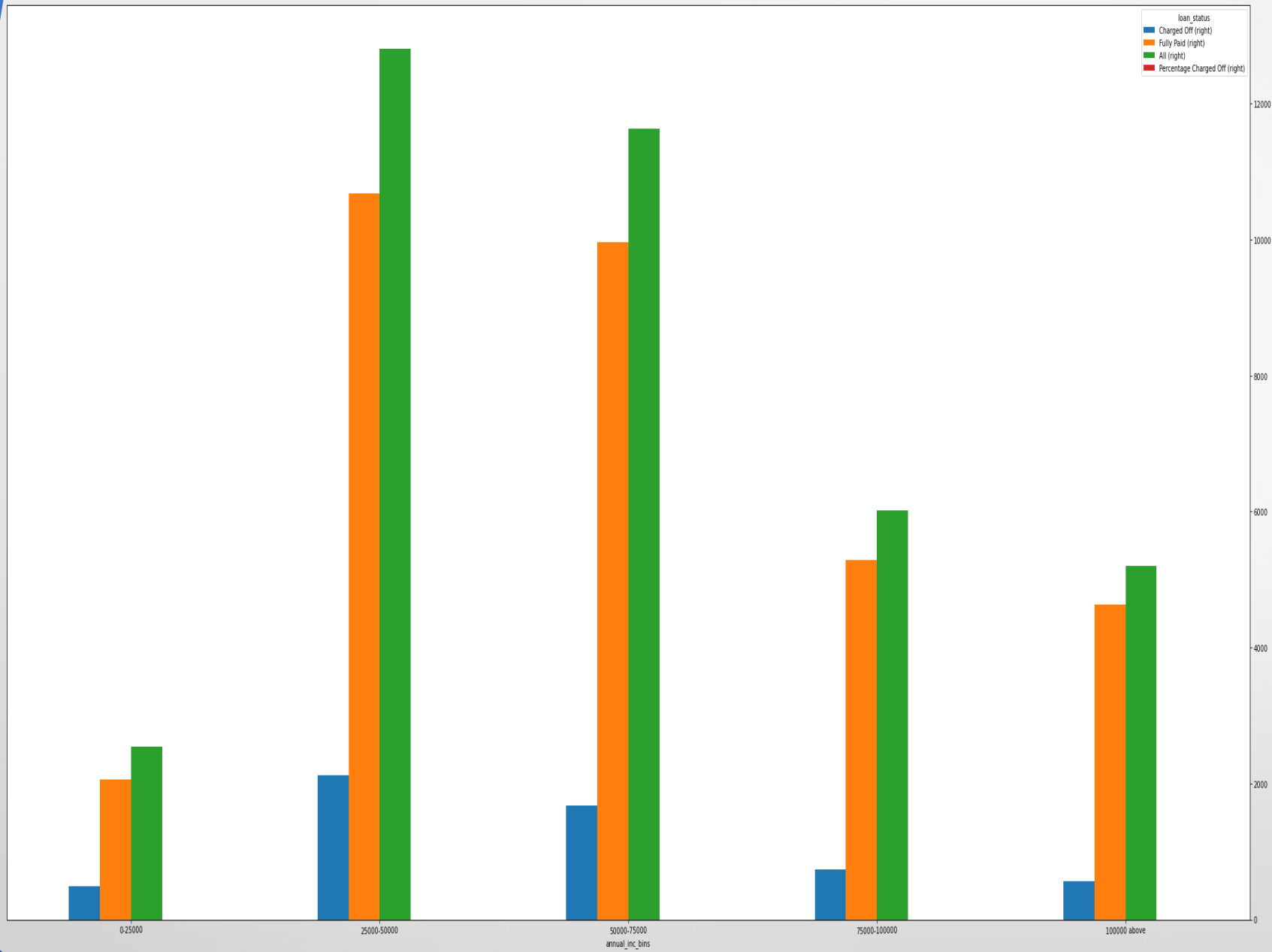


-We can see that as the grade increases the percentage charged off also increase linearly.

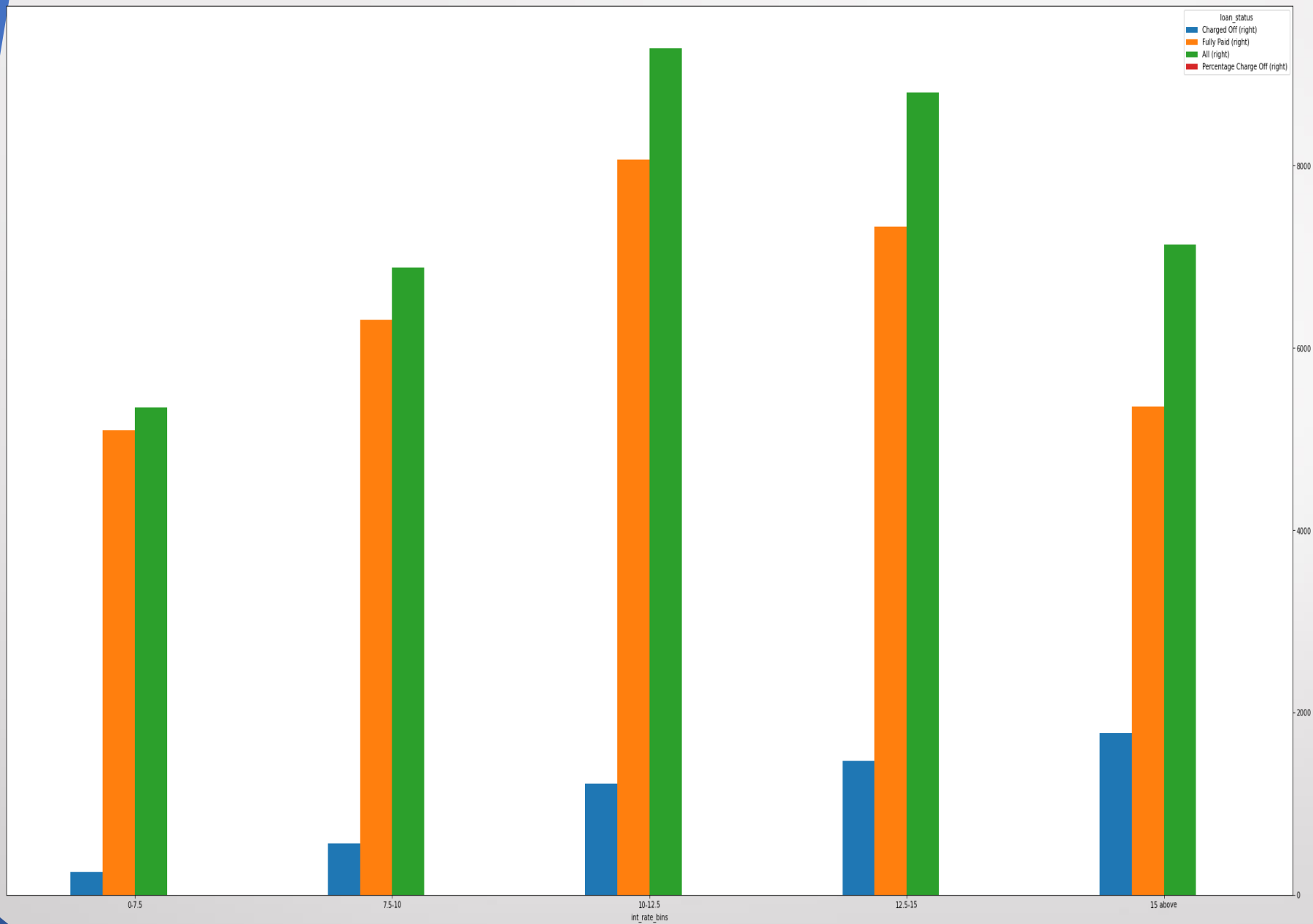


-We can see that the number of loan applications is more for lower grades,

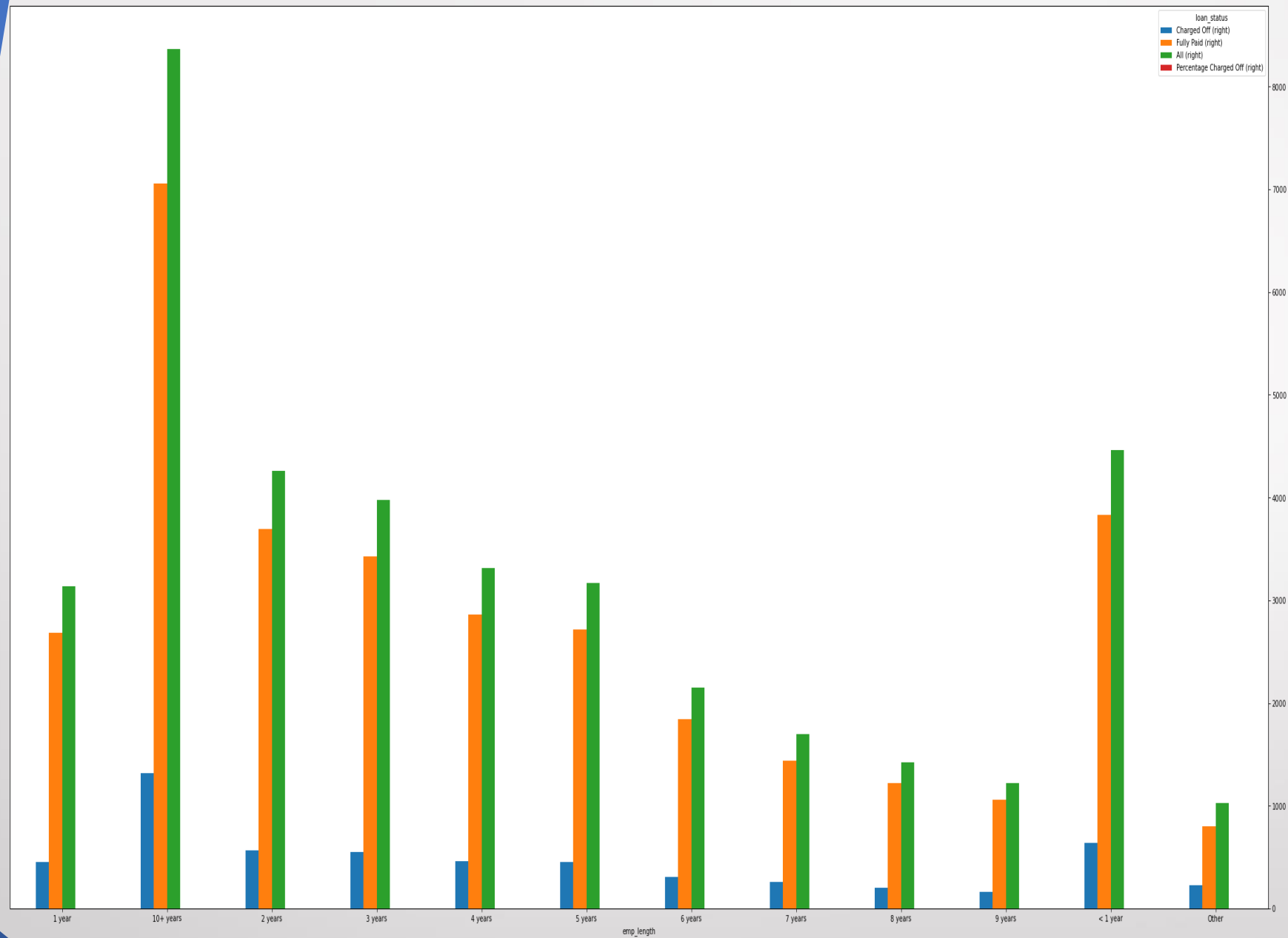
-We can also note that the percentage charged off is increasing as the sub grade increases.



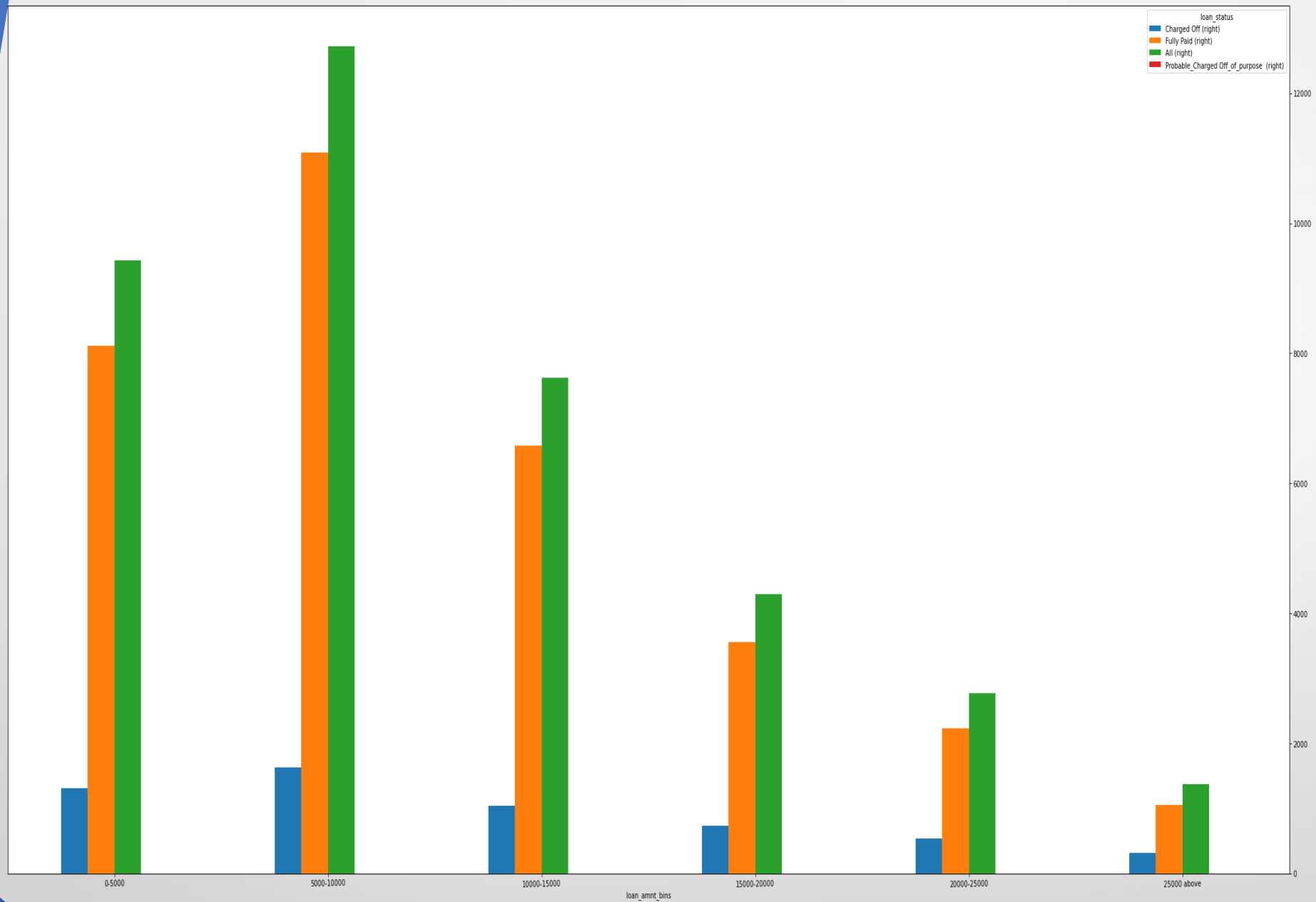
-We can see that people with lesser annual income tend to default



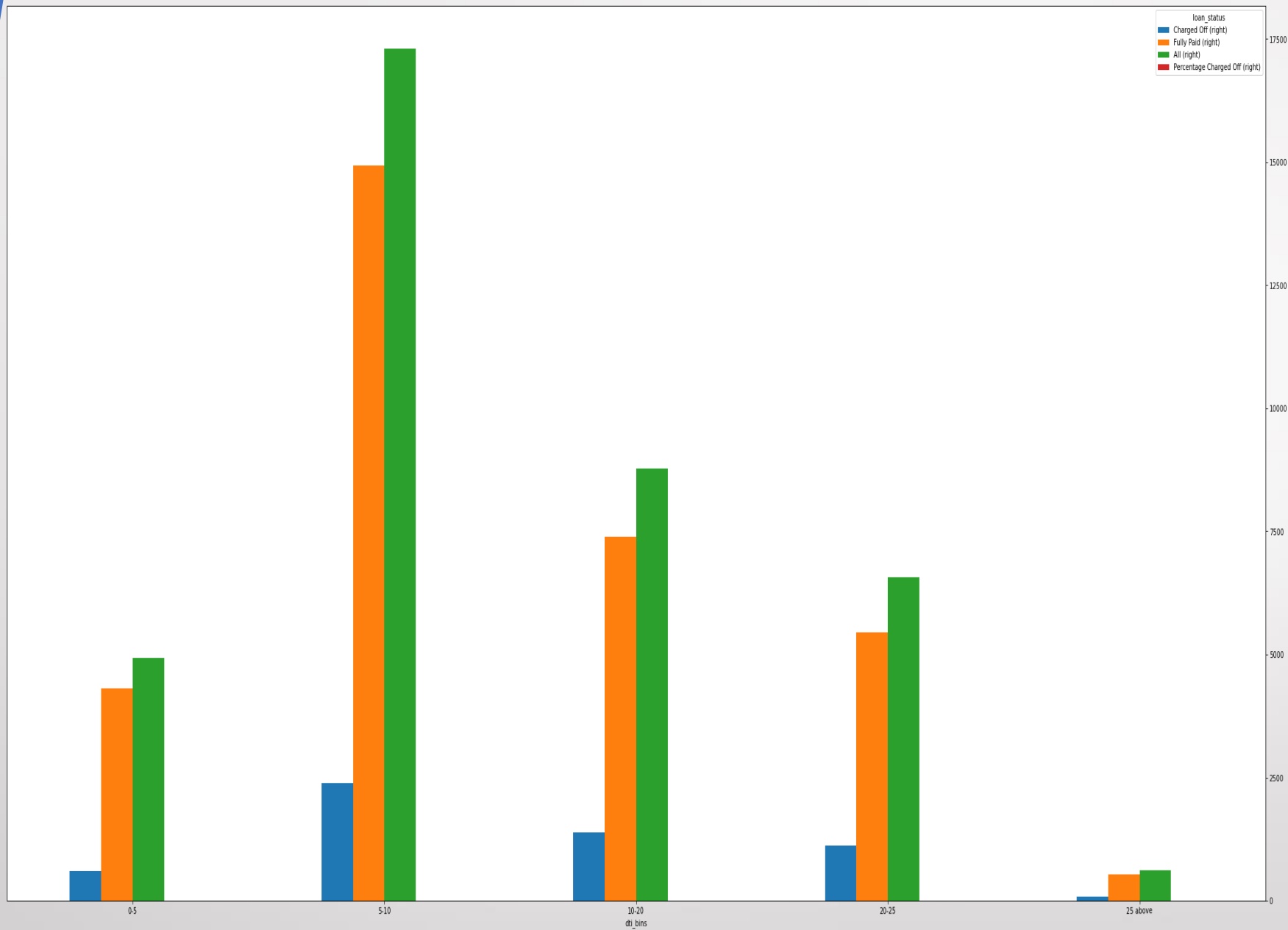
-We can see, as the interest rate increases, the chances of a person defaulting is more



-We can see that people with 10+ year and Other (not mentioned/ recorded) tend to default more.



- We can see that loan amount in range 20k-25k and 25k-or-above, have higher percentage of charge off



-Employees in the range of dti 10-20 and 20-25 have the most charge off



Thank you