

Lending Club Assignment

A Case Study on Exploratory Data Analysis

Presented By : Rajneesh Dubey

Problem Statement

- Exploratory Data Analysis on the historical data of Lending Club. The Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.
- The data given contains information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.
- We have the historical data of the applicants with their loan data. With this, 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.

Data Analysis – Attributes Used

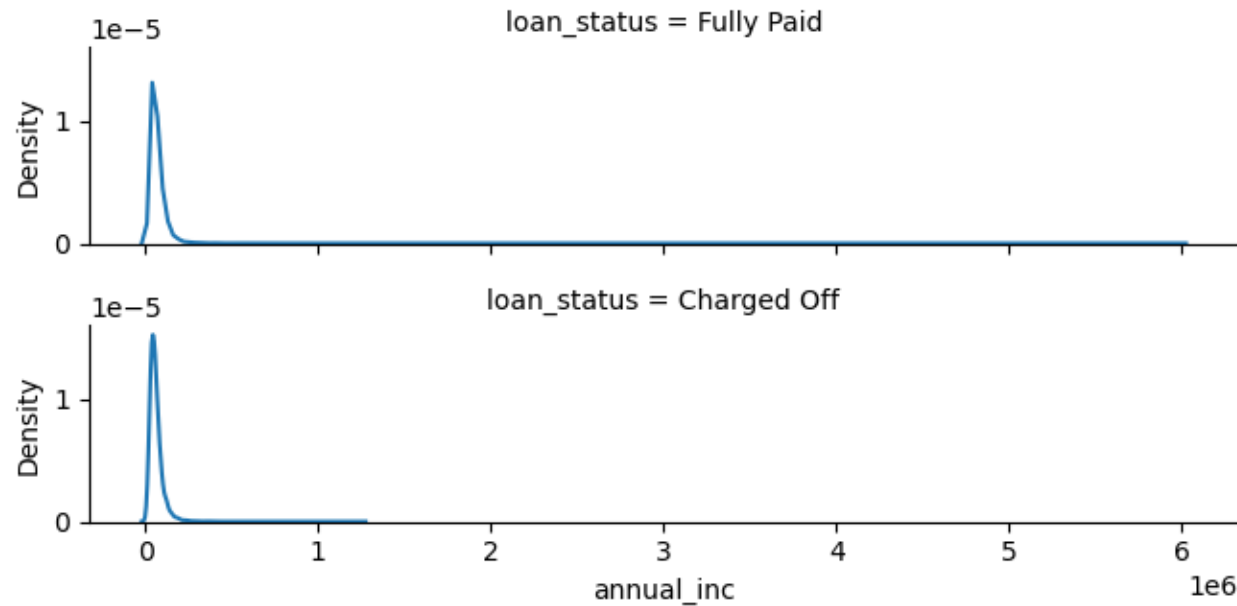
1. **loan_status** : status of the loan i.e. 'Fully Paid' or 'Charged Off'
2. **loan_status_ind** : status of the loan in boolean i.e. 1 for 'Fully Paid' or 0 for 'Charged Off'
3. **int_rate_num** : interest rate at which the loan is granted. Removed trailing % from int_rate column and created this new column.
4. **annual_inc** : Annual income of the applicant
5. **loan_amnt** : the loan amount the is applied for approval.
6. **purpose** : Purpose for which the loan is applied. e.g. small business, debt consolidation, credit card etc.
7. **home_ownership** : Defines the status of home ownership of applicant e.g. RENT, OWN, MORTGAGE etc.
8. **grade** : this is the grading assigned to the loan based on various parameters.
9. **sub-grade** : sub group of grade attribute which is used categorize the quality of loan.
10. **debt-to-income** : This is a ratio of the debt an applicant possesses to the income of the applicant.
11. **term_in_month** : The column derived by trimming 'months' from term column. Defines the term for which the loan is taken. e.g. 36 months, 60 months.
12. **funded_amnt** : funded_amnt is the amount recommended/approved by Lending Club
13. **funded_amnt_inv** : It is the amount funded by investors.
14. **fund_less_than_approved** : Newly introduced. This is the difference between the amount approved by Lending Club and the amount funded by investors. This can hold significant meaning to analyse the impact on the credit risk.

Impact of Interest Rate on loan_status



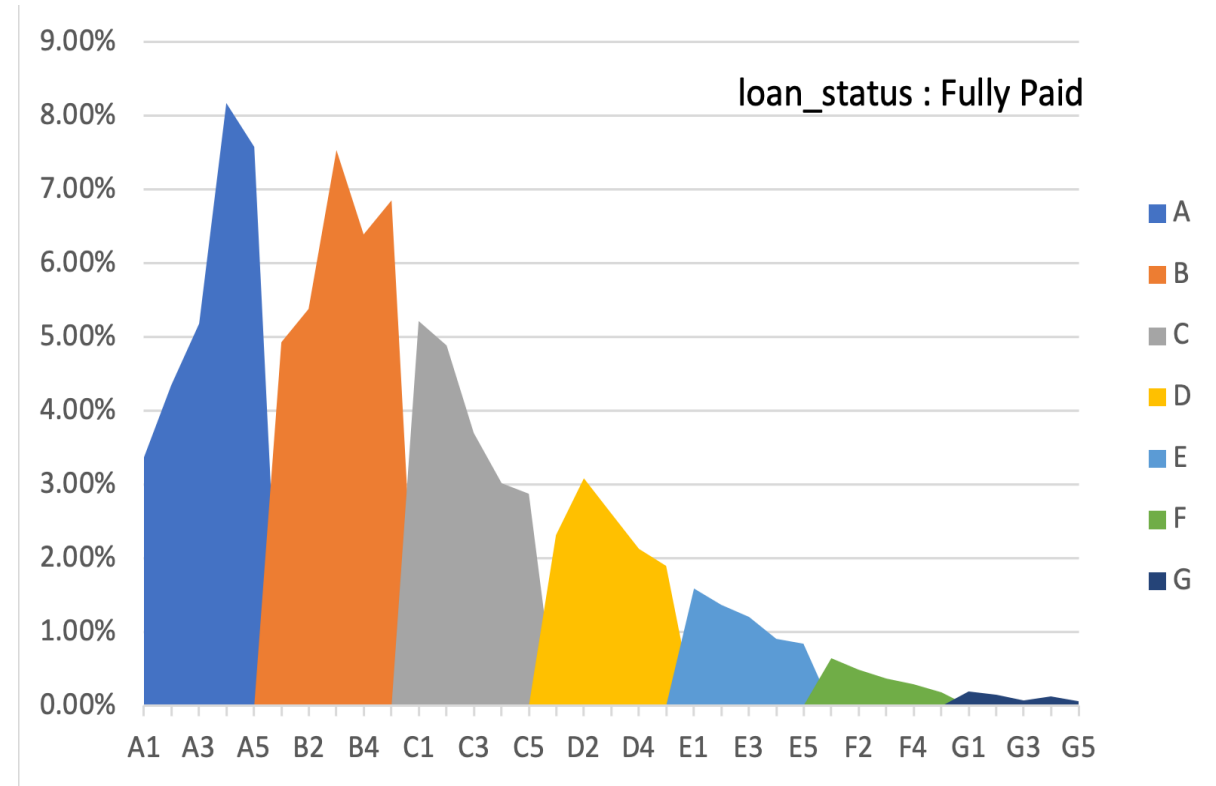
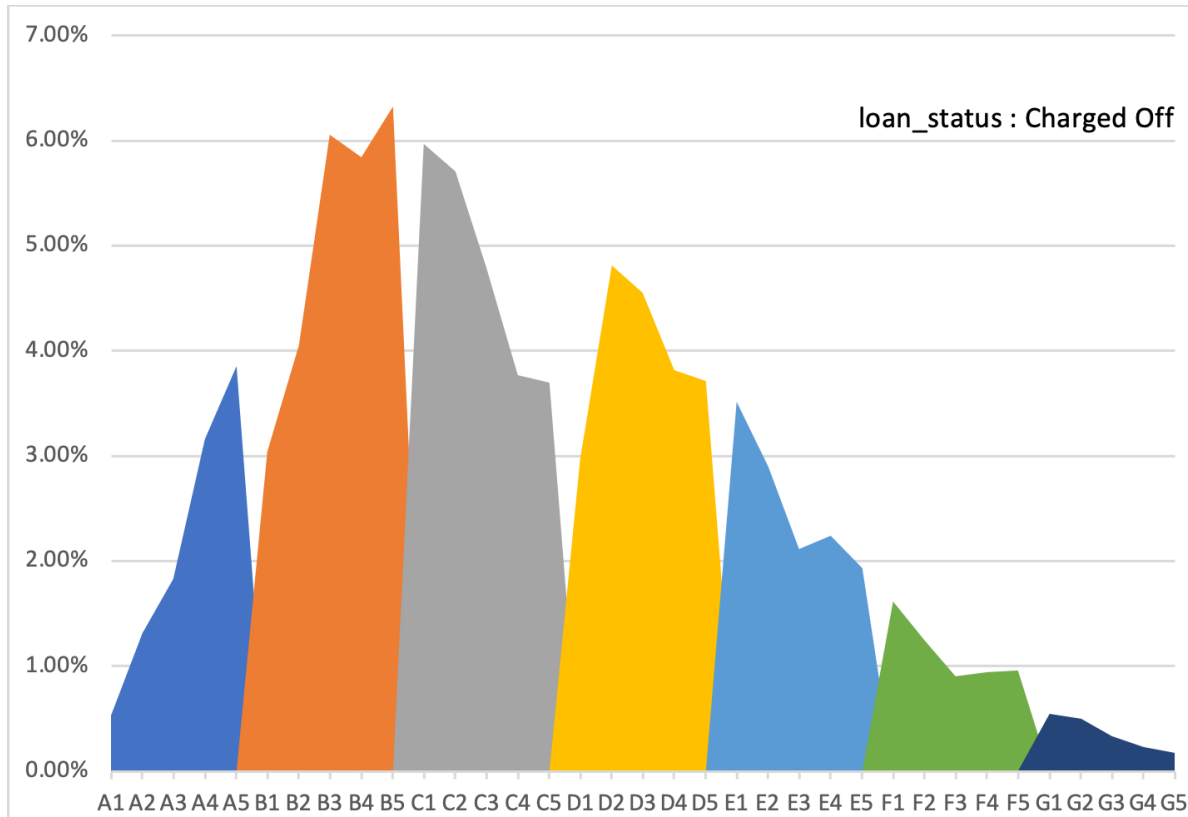
Conclusion : We observe that the applicants whose loan status is Charged Off i.e.Defaulters have slightly higher interest rate in the distribution. We can infer that the lower the interest rate,the easier it is to pay.

Impact of Annual Income on loan_status



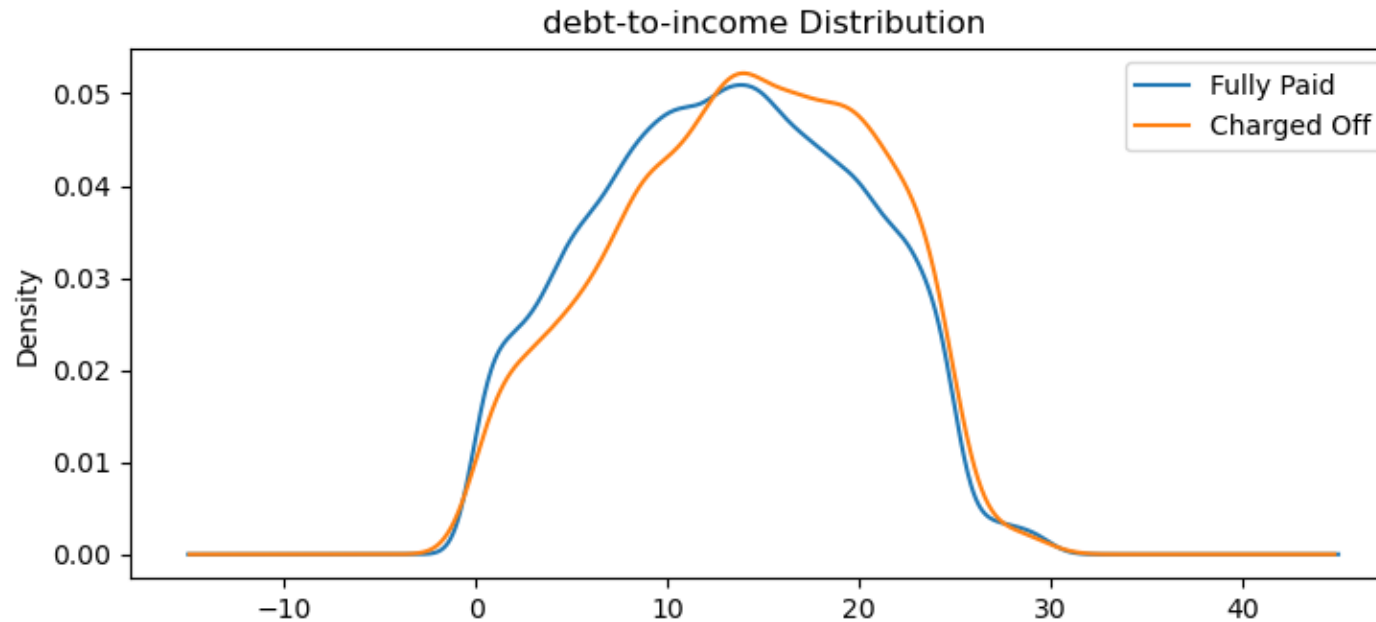
Conclusion : We observe that annual income of the applicants doesn't have any visible impact on the loan_status. Hence it hasn't been a major decisive factor, from the historical evidences.

Impact of loan gradings on loan_status



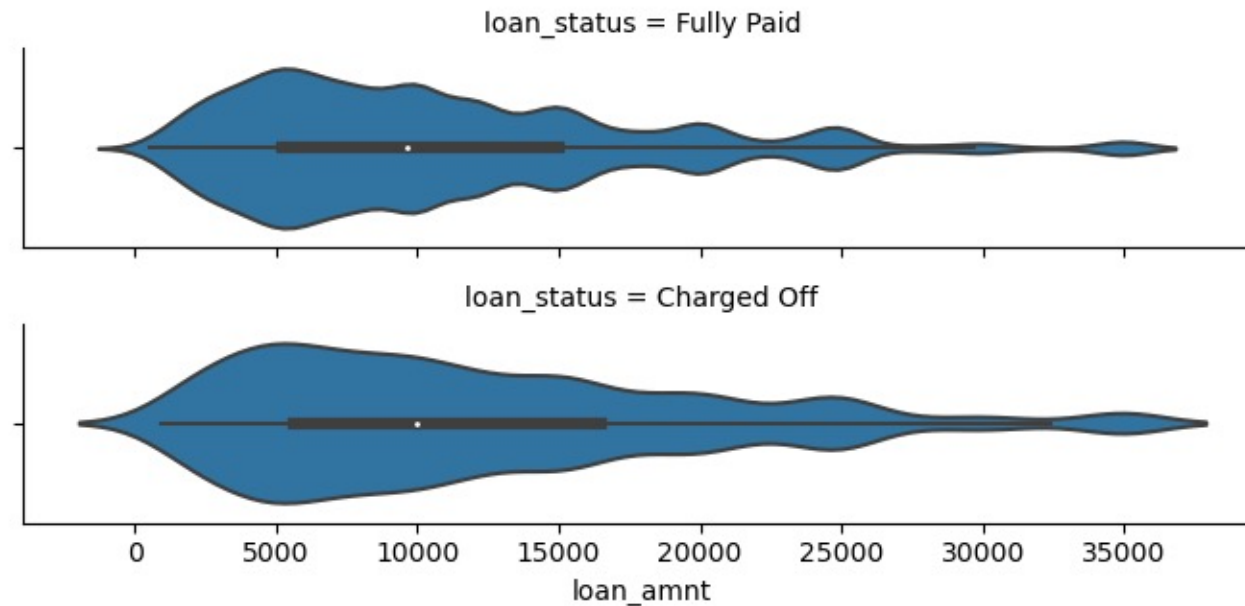
Conclusion : We see that in case of Fully Paid loan status, Grade A has more frequency. Whereas, In case of Charged-of, Grade A and Higher Grades are significantly lesser. Hence, it can be observed that Higher Grades have very less chance of Defaulting or Charging off.

Impact of DTI (debt-to-income) ratio on loan_status



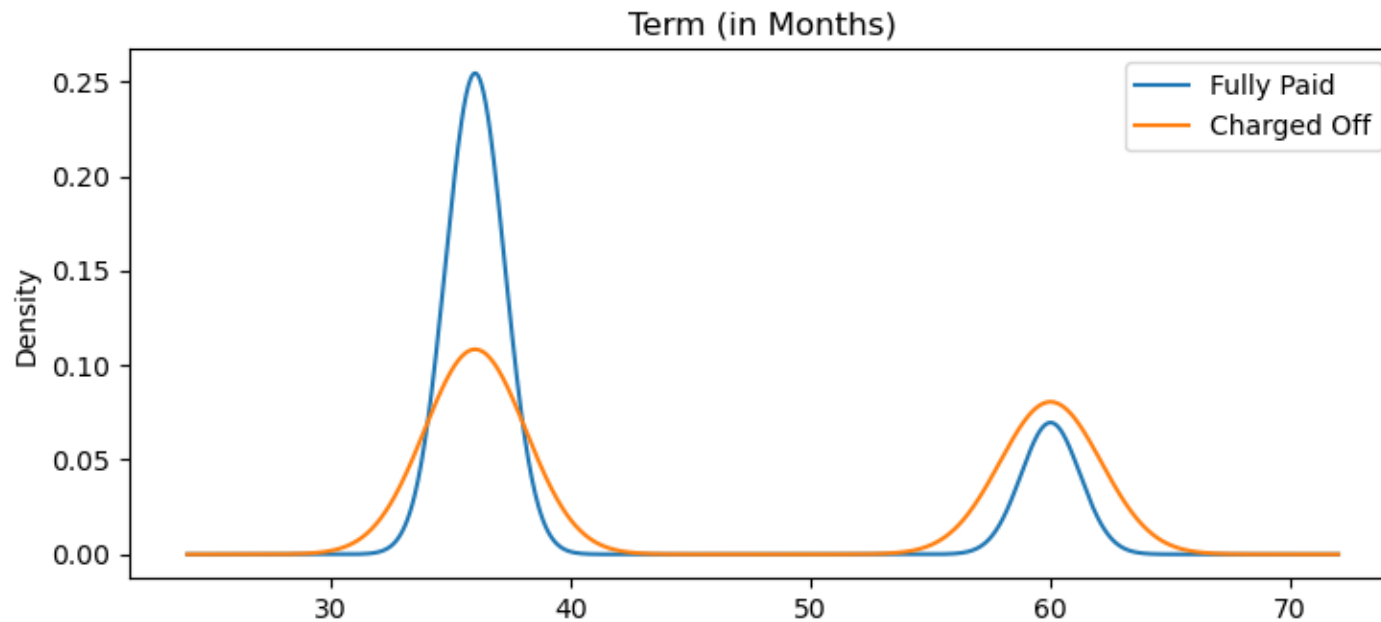
Conclusion : We observe that the applicants whose loan status is Charged Off i.e. Defaulters have slightly higher debt-to-income ratio. It makes sense as more the debt, the difficult it becomes to handle it.

Impact of loan_amnt on loan_status



Conclusion : We observe that loan_amnt of the applications doesn't have any visible difference on the loan_status. Hence it can't be considered as a major decisive factor, from the historical evidences.

Impact of Loan Term on loan_status



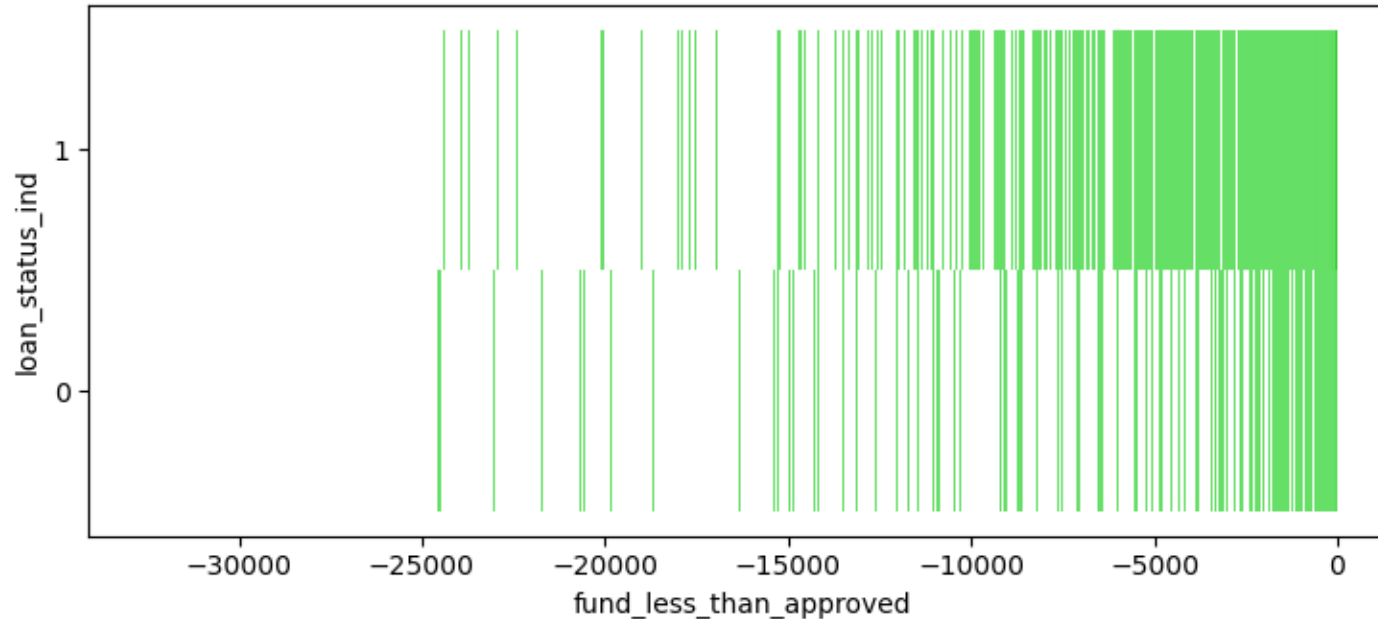
Conclusion : We observe that the applicants who have loan term as 36 months instead of 60 months. are more likely to Paid off the loan.

Impact of home_ownership on loan_status



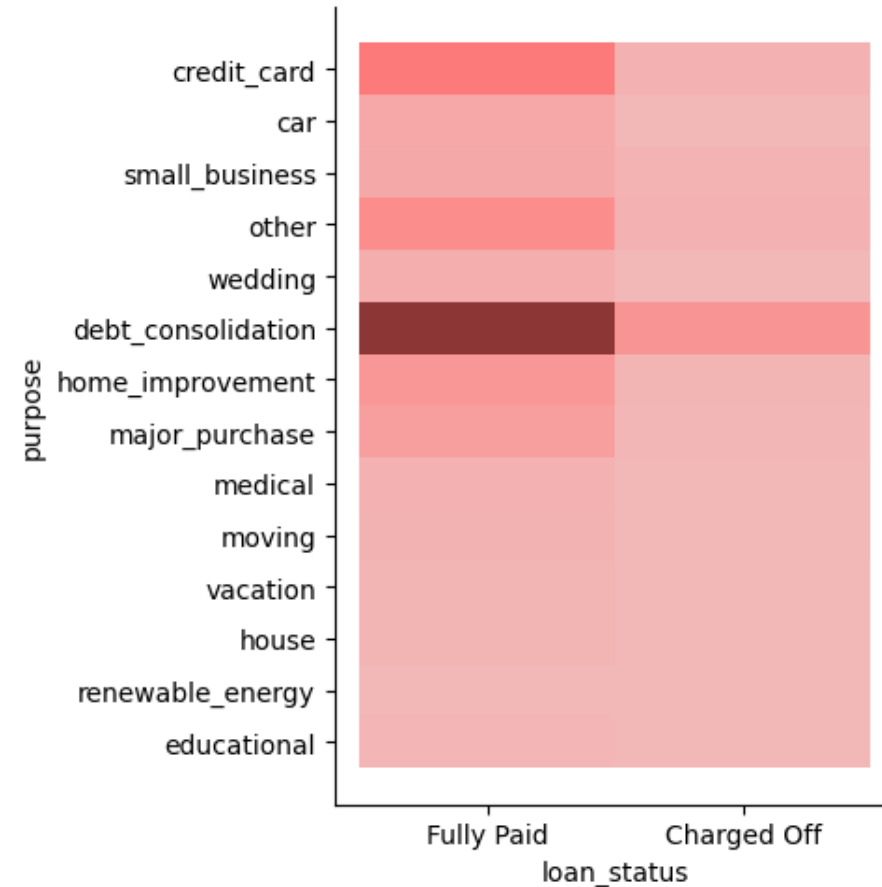
Conclusion : We see that home ownership is kind of proportionate in both the loan statuses. Hence, it doesn't make much impact on the ultimate loan status.

Impact of having fund_less_than_approved on loan_status



Conclusion : We observe that the applicants whose investors have been careful while allocating the funds and have allocated the appropriate funds (upon analysis) instead of approved amount. Such applications have seen significantly lesser default rates.

Impact of purpose on loan_status



Conclusion : We don't see any significant impact of the loan purpose in deciding the loan_status.