Lending Club Case Study

By

Anirban Ganguly

and

Anshul Goyal

Overview and Objective

- We were given a data set containing details of loans. The dataset had about 111 columns and 39717 rows. We were also provided with a data dictionary which explained each of these 111 columns
- Our task is to investigate the data, do exploratory data analysis and conclude on factors that effected a person to default on loans

Solution Approach

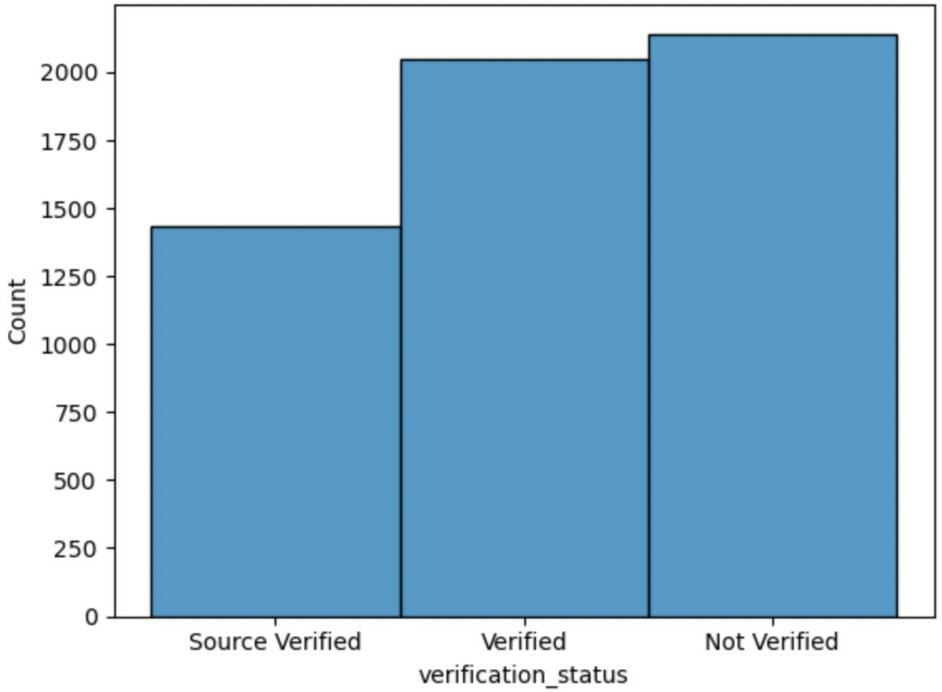
- We first did the data cleaning. After the data was cleaned, we ended up with 39 columns (we missed 2 columns with values 0, so the actual column count would be 37)
- Then we cleaned the rest of columns and converted the data as required
- Then we did univariate analysis of all those columns to understand the data distribution
- After that we tried to understand the correlation of all those columns with each other
- Finally, we started doing segmented/bivariate analysis on all relevant columns

Observations

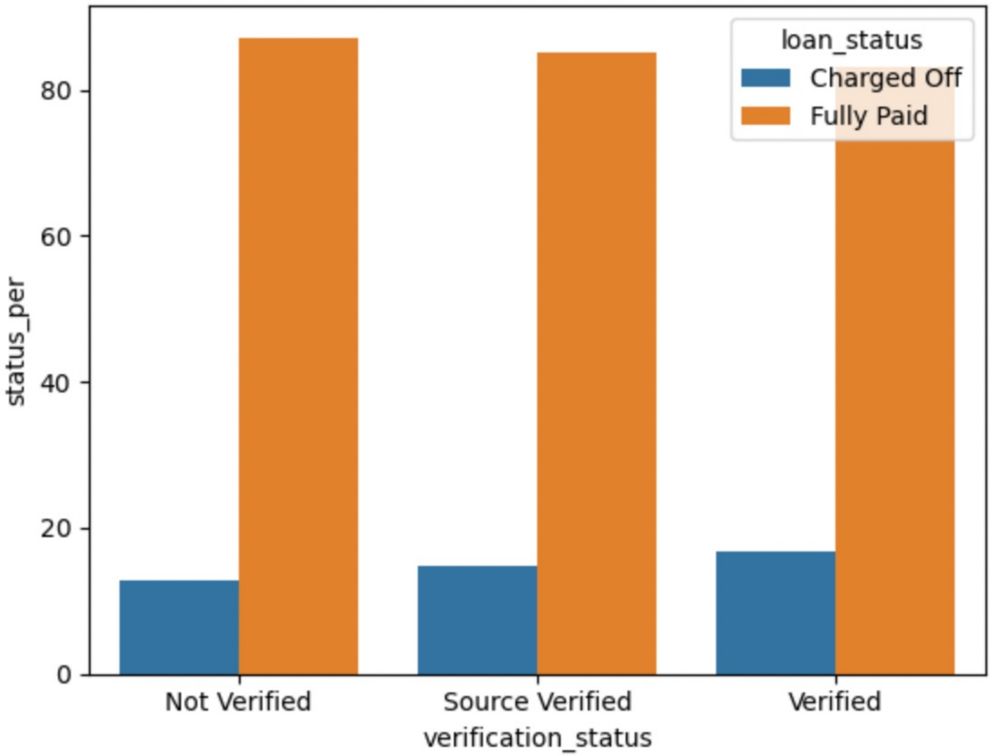
Based on the EDA the following slides have explains our observations and suggestions

Effect of Loan Default on Source Verification

- When we looked at the defaulted data, we found that a unverified persons has more chances of defaulting. Which is kind of makes sense.



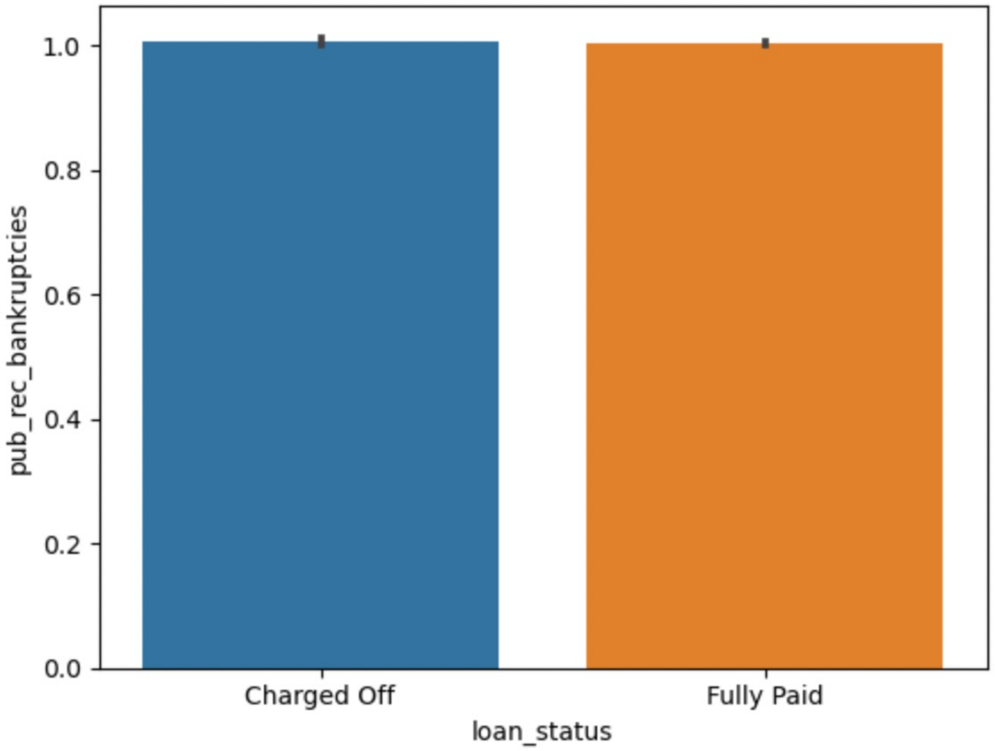
- But when we compared the data with non defaulted loans, this is what we got



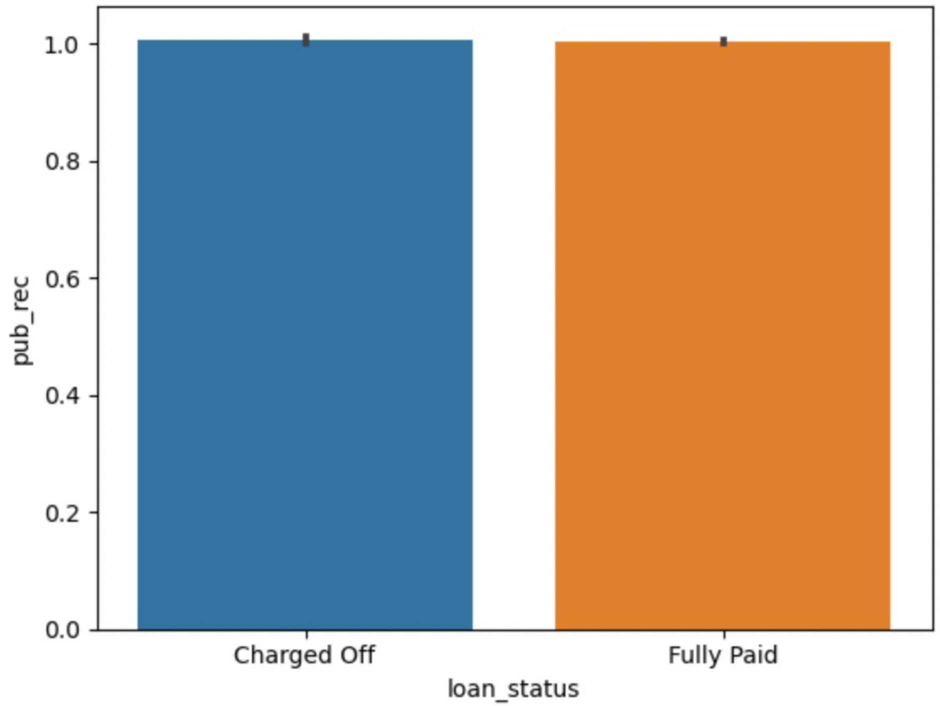
Conclusion – It looks like the verification is not done correctly that is why even the verified loans are getting defaulted. Also we should do more source verification to make sure that the loans are defaulting less

Effect of Public Record on Loan Defaults

- Public Record of bankruptcy

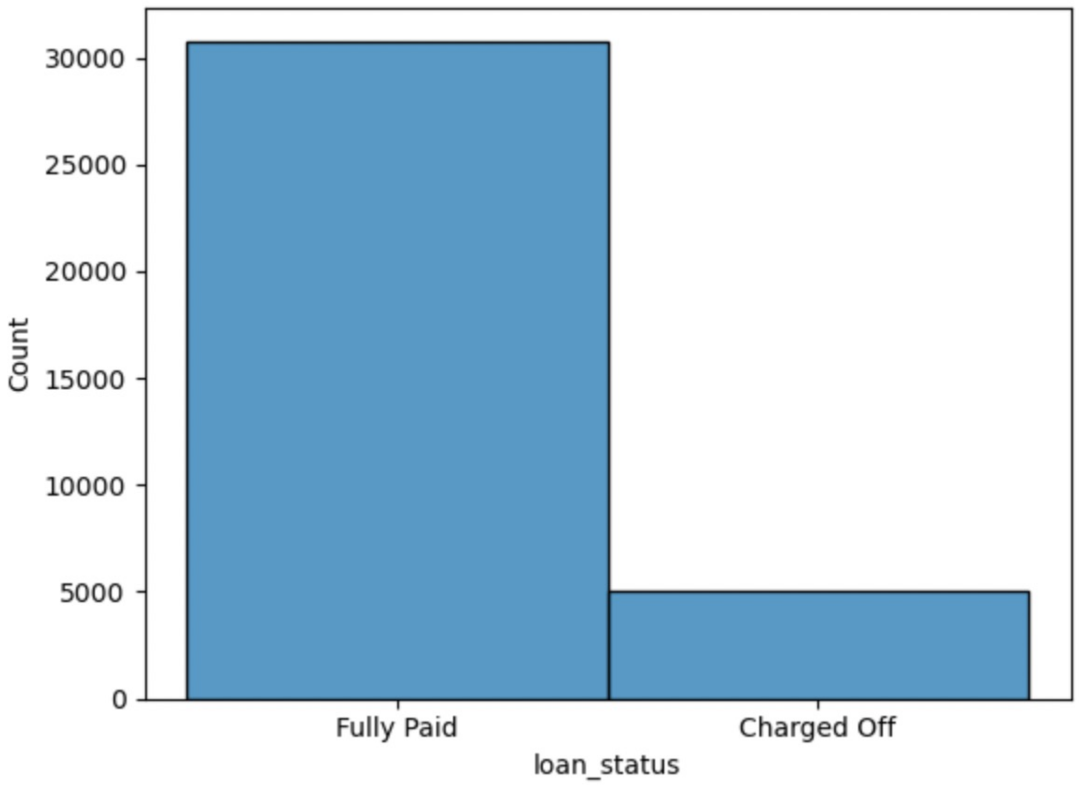


- Public Record of Derogatory remarks



Conclusion – It looks like even if the person was bankrupted in past or got a derogatory remarks they can be trusted with new loans

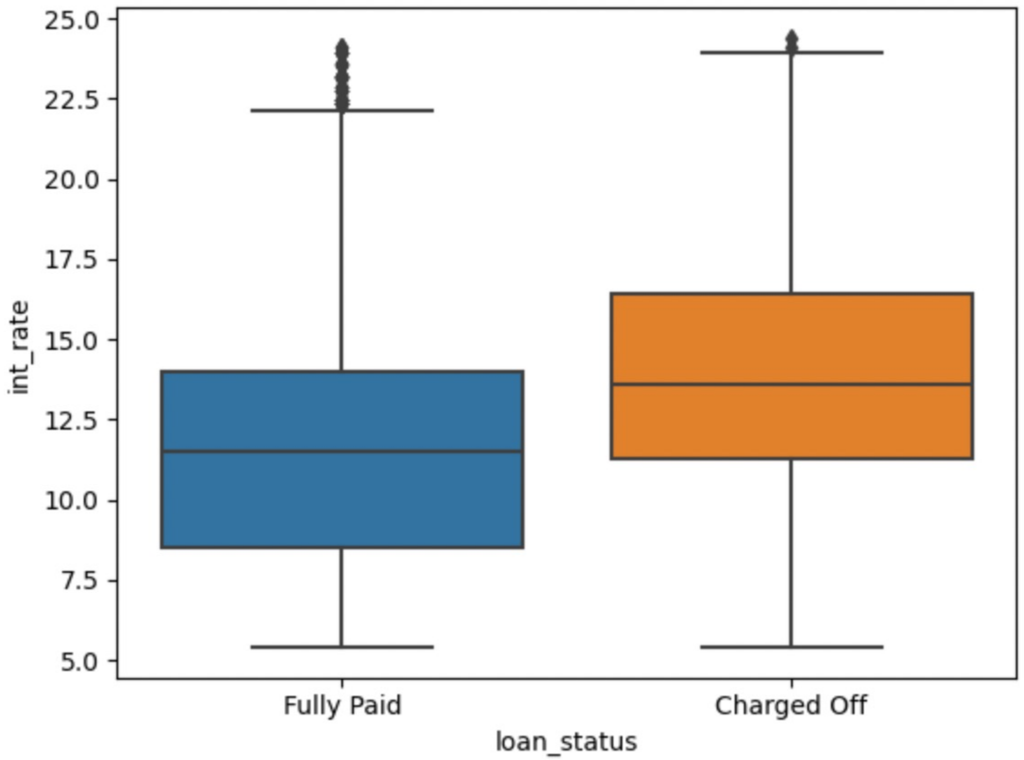
Effect of Public Record on Loan Defaults continued



- If the person does not have any derogatory remarks, there are higher chances that he would return the loan
- **Conclusion** – Though there is an almost 50:50 chance that a person with public records will default if the person does not have a public record, then the chance goes down pretty much. So, it is always a safer bet to give loan to a person with no public derogatory or bankruptcy records

Effect interest rate in loan defaults

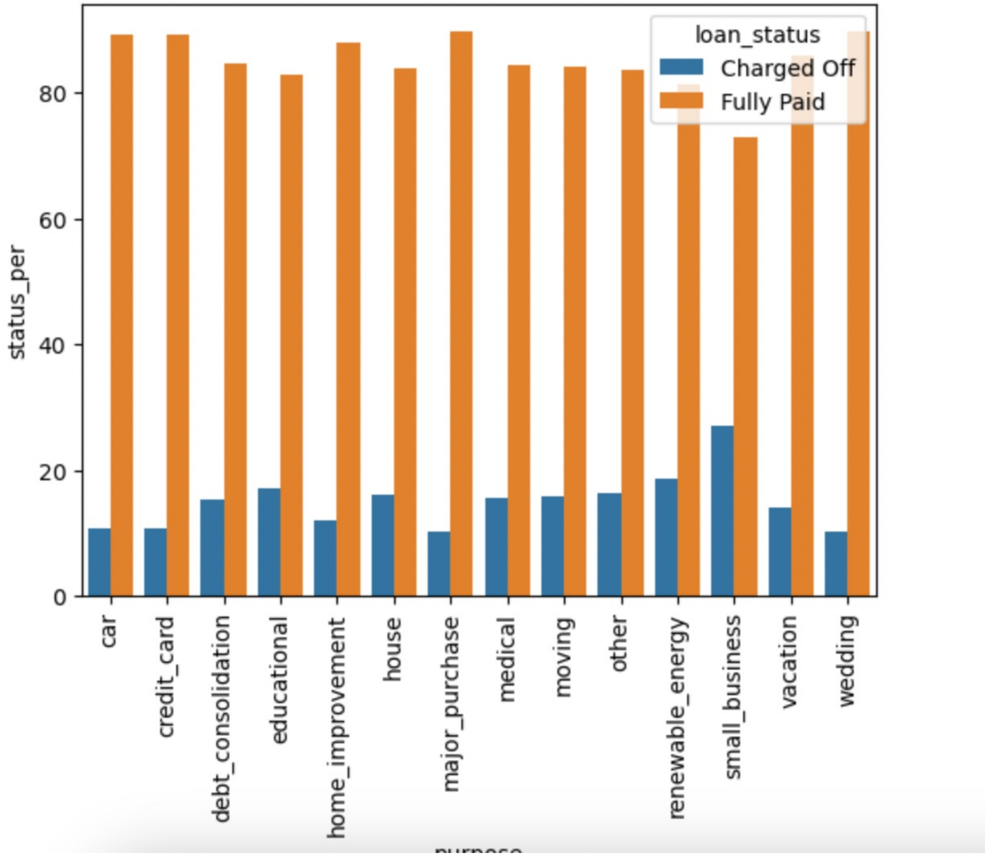
- If the loan interest goes over 13% there is a higher chance of defaults. In fact, while for non defaulted loans when the median interest rate value for non defaulted loans stay at 11% and the 75% percentile stay at 13%, for defaulted loans the median is at 13%



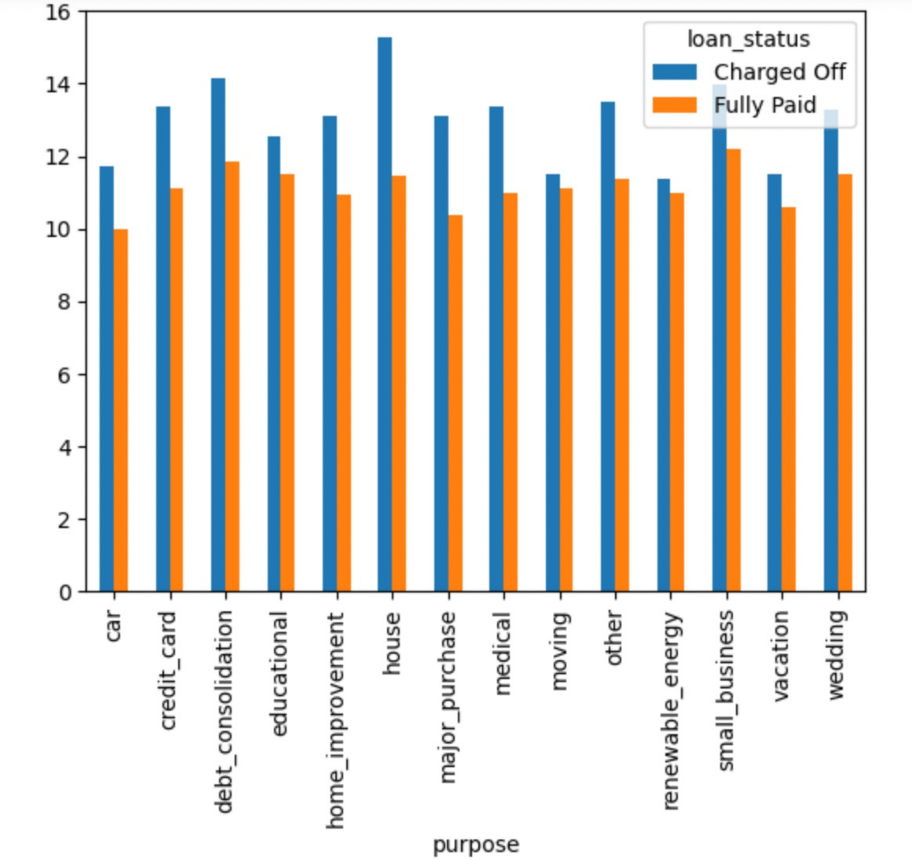
Conclusion – Try to keep the interest rate at 11%

Safer loan purposes

- We checked the loan types for payments

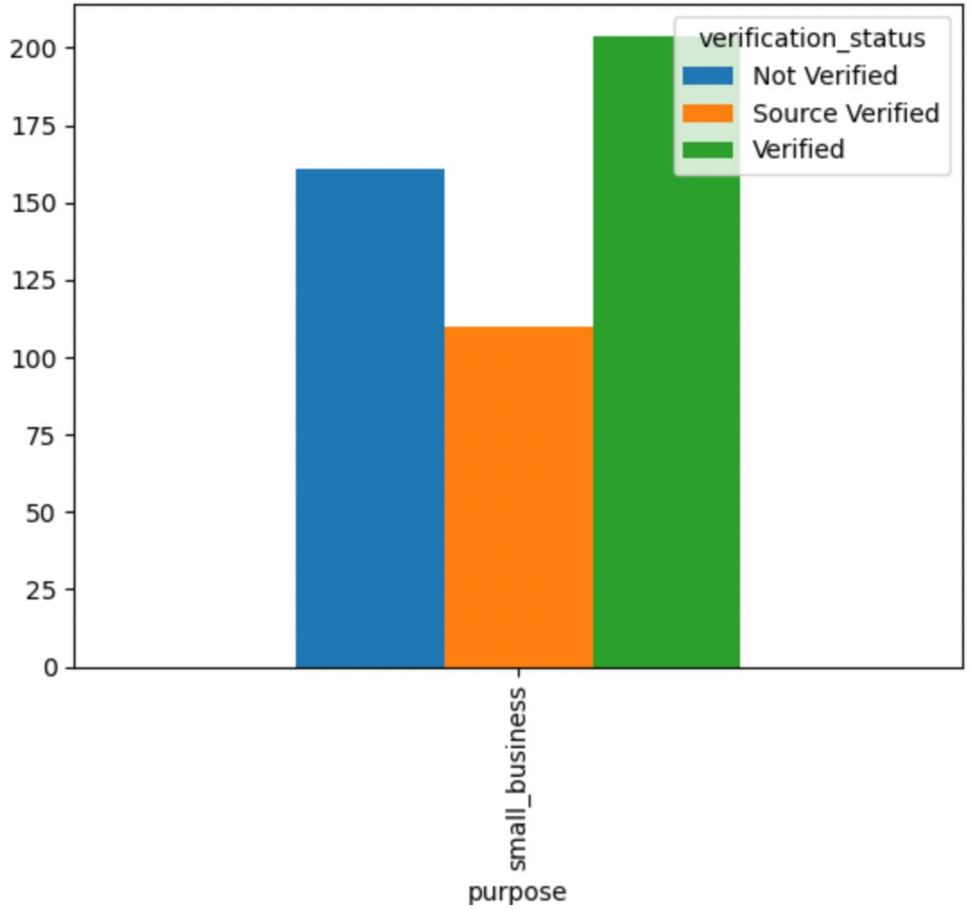


- We also checked if loan category have specific interests tied to it



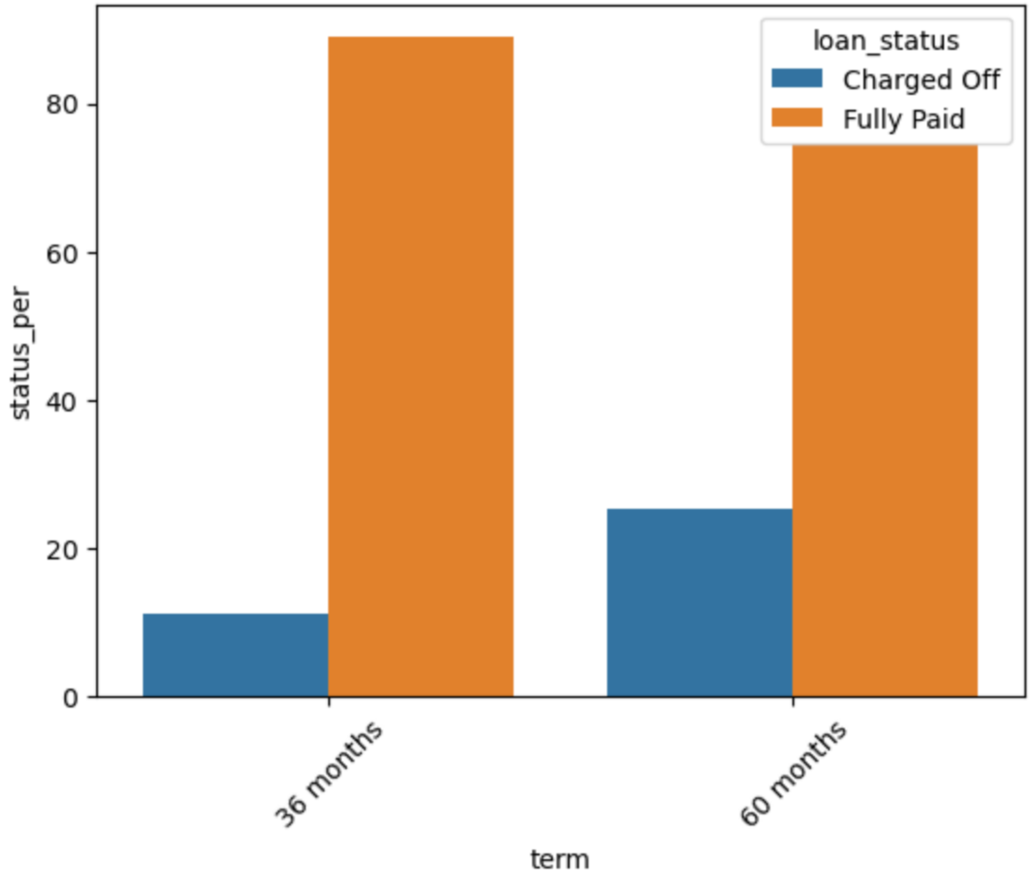
Safe Loan Purpose Continued

- We also looked if loan verification helps reduce loan defaults
- **Conclusion –** Though home loan and small business both tends to have higher loan interest rate. Small business tends to default more than any other loan types. We should try to be more vigilant while giving loan to small businesses. This means we should do *better source verifications*. Also, car, credit card, major purchase and wedding loans seems to have a higher chances of not defaulting.



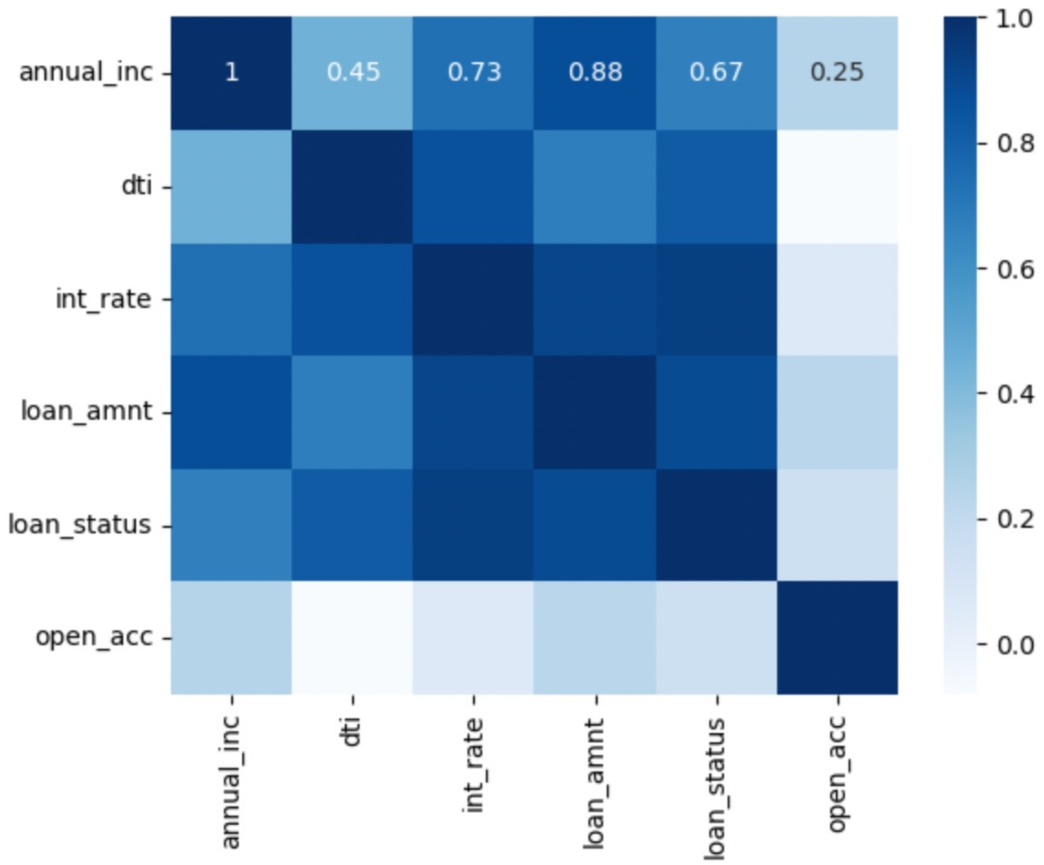
Effect of loan term on defaults

- We check the loans against defaulted and non defaulted.
- **Conclusion** – There is a higher chance that the longer terms loans may get defaulted.



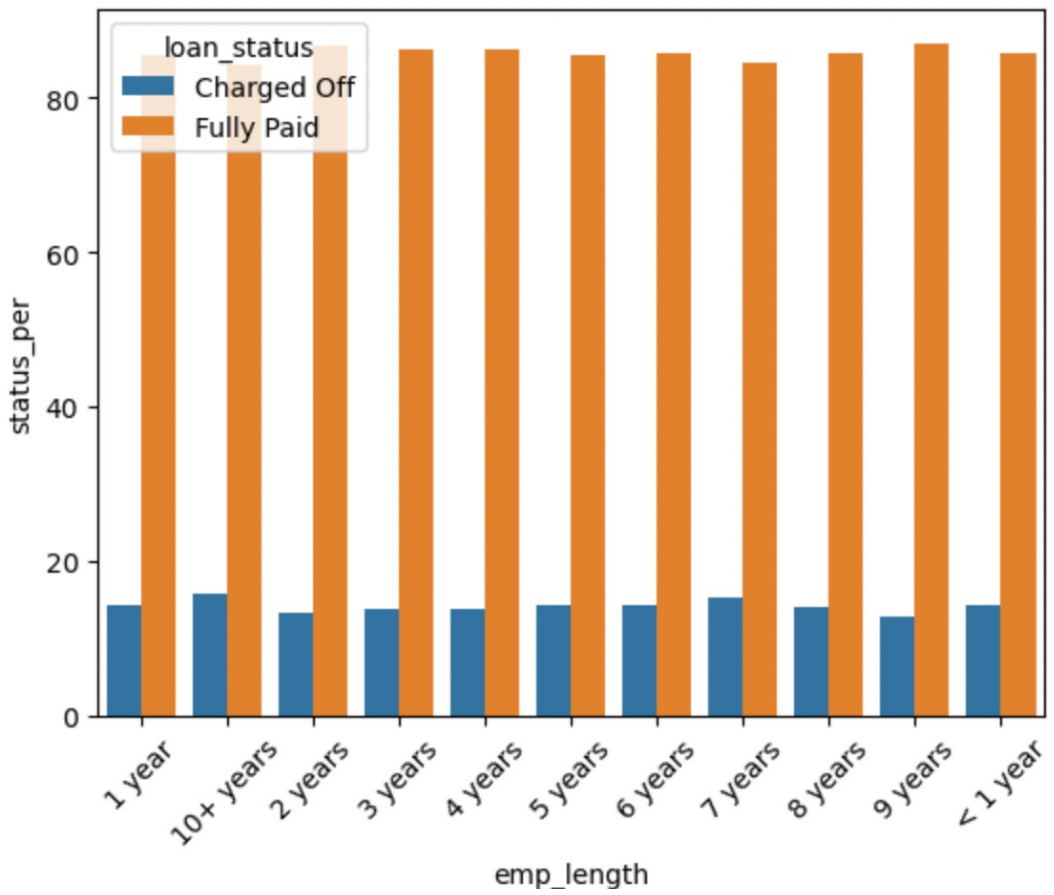
Effect of loan default on various parameters

- The loan status have higher correlation with loan amount, interest rate and annual income
- **Conclusion** – As the loan amount and interest rate increases so does the chances of loan default

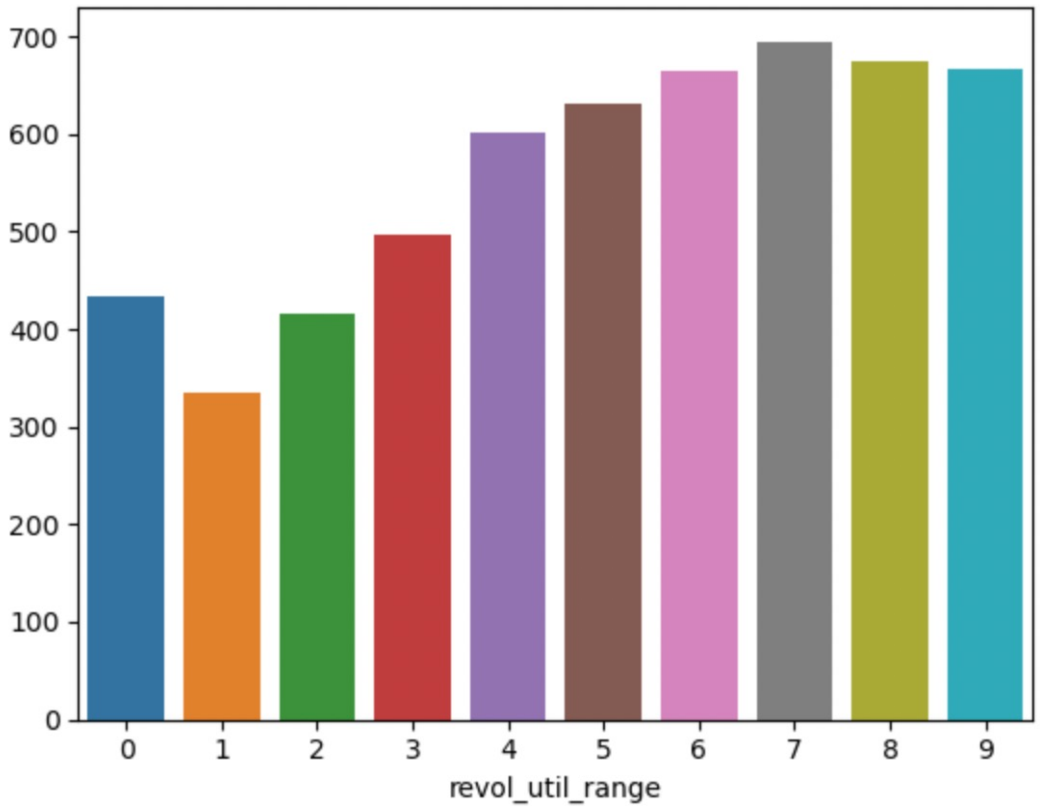


Does longer employment mean better loan repayment?

- All employment year gives similar loan repayment and loan defaulting values.
- **Conclusion –** The number of year of experience have not much effect on loan repayment. There seems to be a slightly higher chances of defaulting for people with less than 1 year or more than 10 years of experience.



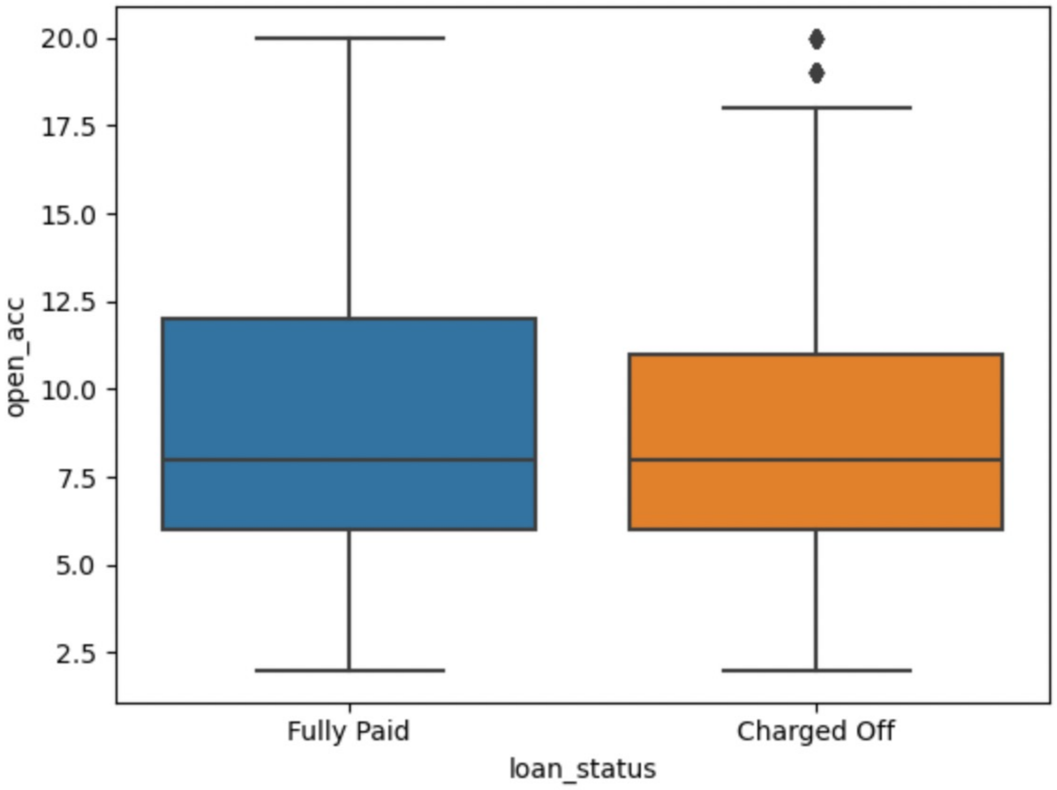
How number of credit line used effects loan defaults



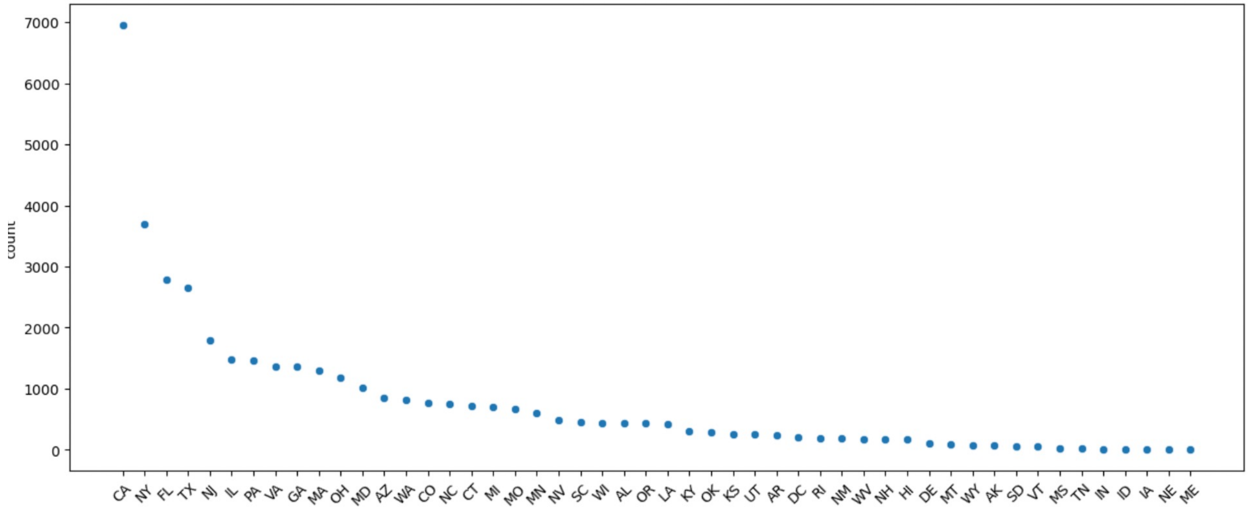
- From the data it looks like if there are more than 20% credits in use at the same time then there is a higher chance that the person would default.
- **Conclusion** – if the person is using more than 20% credit utilization think carefully about giving additional loans

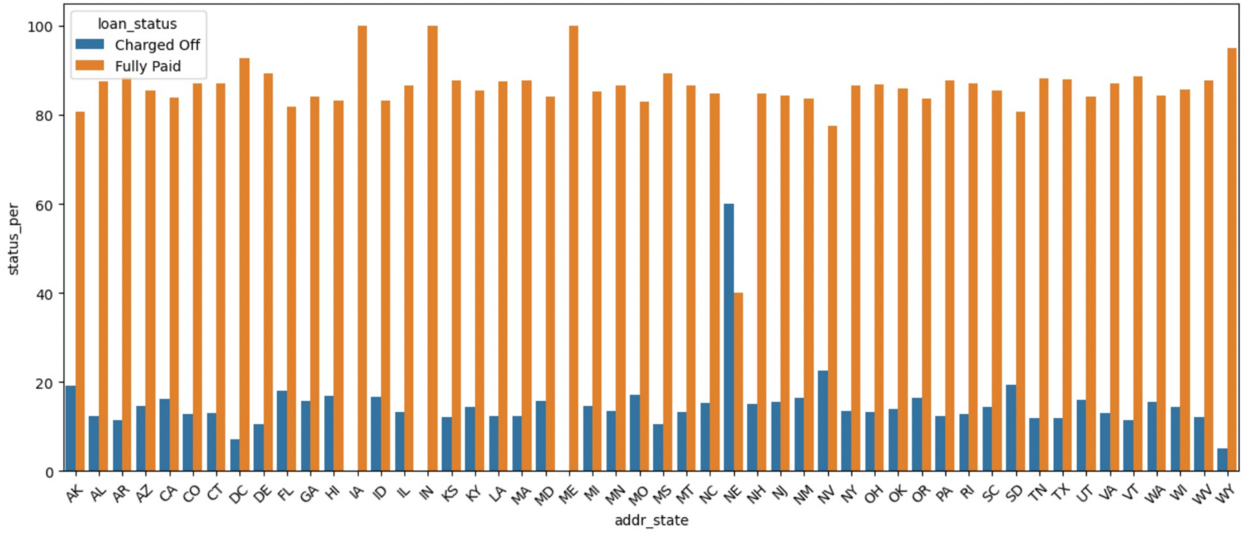
Effect of open accounts on defaults

- The non defaulted loans have slightly higher number of open accounts. That means, they have
- **Conclusion** – Since more than 20% active credit utilization can mean higher chances of defaults, our ideal customer should be someone who have additional unutilized credit lines



Effect of state on loan defaults





- Based on the number of loans in different states and the defaults that happened in difference states this is what we found
  - Among the top 10 states the following states have less chances of default than any other states – Texas, Illinois, Pasadena, Massachusetts
  - Among all states the following states are performing well – Wyoming, District of Columbia,
  - Among the states with less business, we have a very bad record at –Nebraska.

Final Conclusions

- Keep interest rates around 11% for all loans
- Be careful while giving loans to small business, make sure you are doing a source verification
- If the loan term is of 60 months there are higher chances of defaulting.
- An ideal customer should have around 20% credit utilization, but the more credit line available to the person the better.
- Try to do more business in Wyoming, District of Columbia, Texas, Illinois, Pasadena and Massachusetts
- As the loan amount and interest rate increases so does the chances of loan default
- Please come out with a better income verification process as the current process does not guarantee a proper payment of loans