Exploratory Data Analysis

Group project – Lending club

Le Nguyen Duc – norton0704@yahoo.co.uk

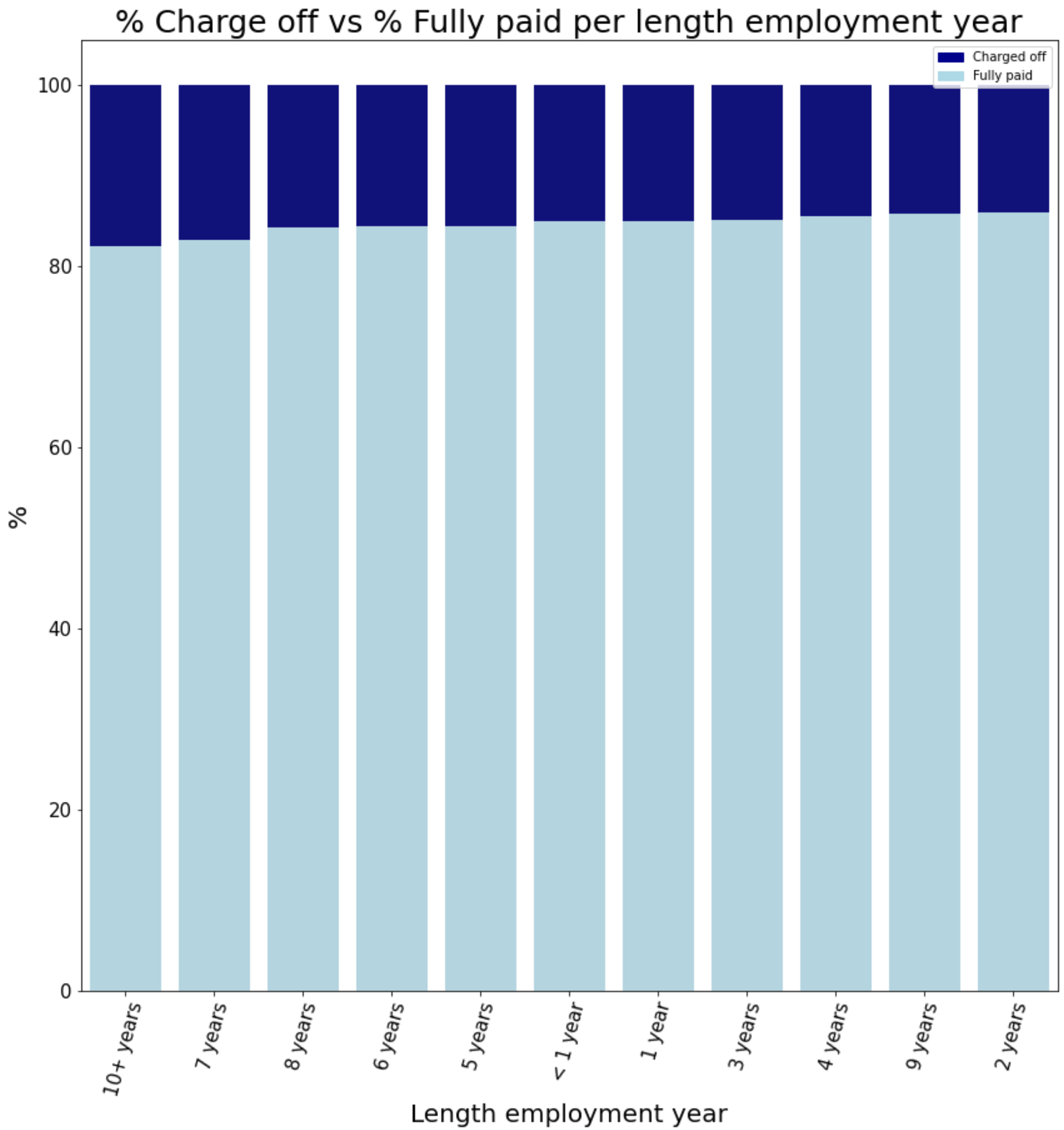
Tran Chi Thao - Thaotc.pte@gmail.com

Executive summary

- Data about lending customers from 2007 to 2011.
- Objective to detect patterns that may affect to bad debts.
- Only descriptive statistics analysis is applied.
- Univariate and bivariate analysis to determine some bad debt patterns.
- Those patterns may be a strong indicators for predictive analysis in future.

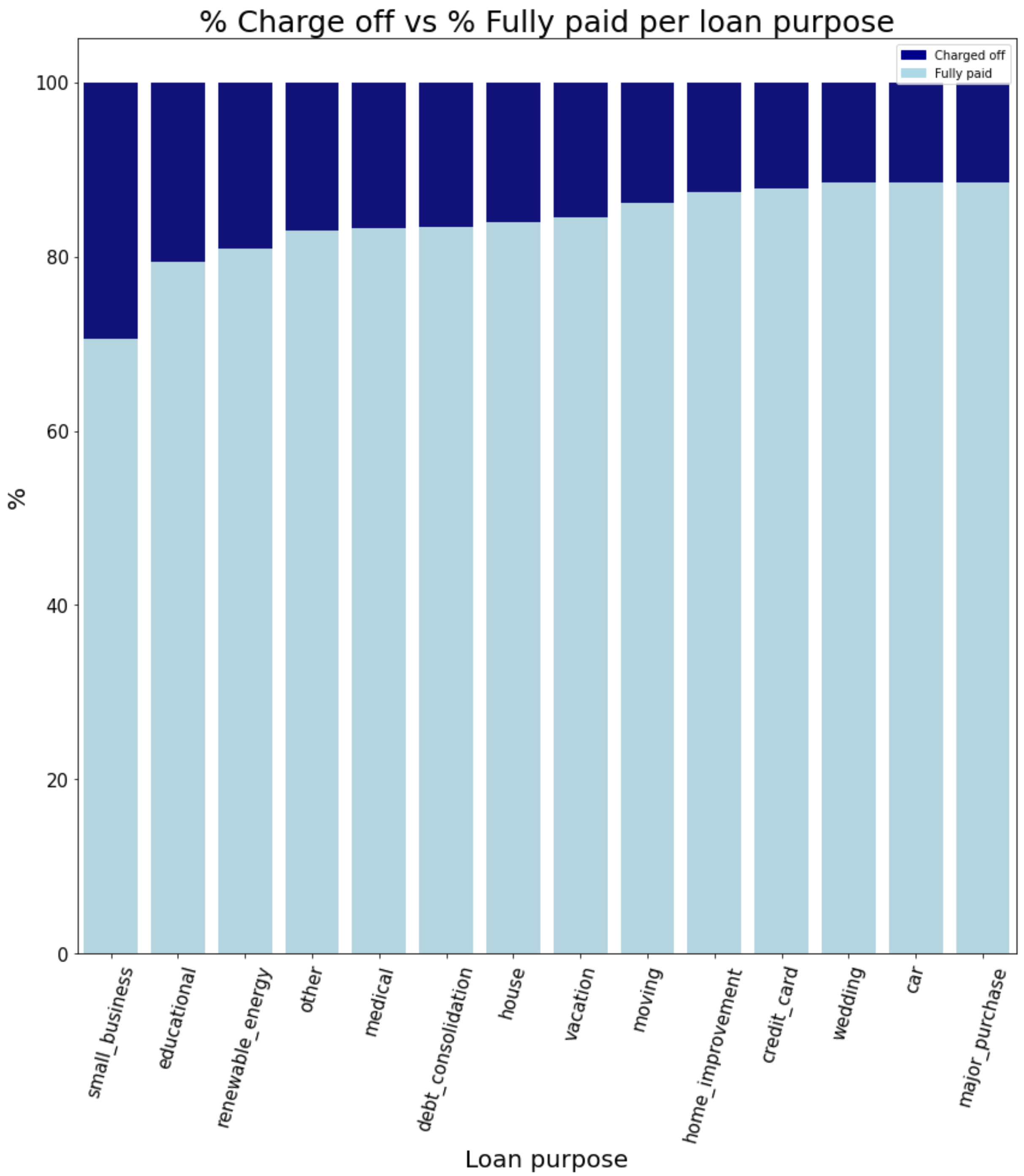
Employment year

- The chart on the right shows the contribution of charged off and fully paid (in term of loan amount).
- It's surprisingly that customers with more than 10 year of experience tend to less pay off the loan.



Loan purpose

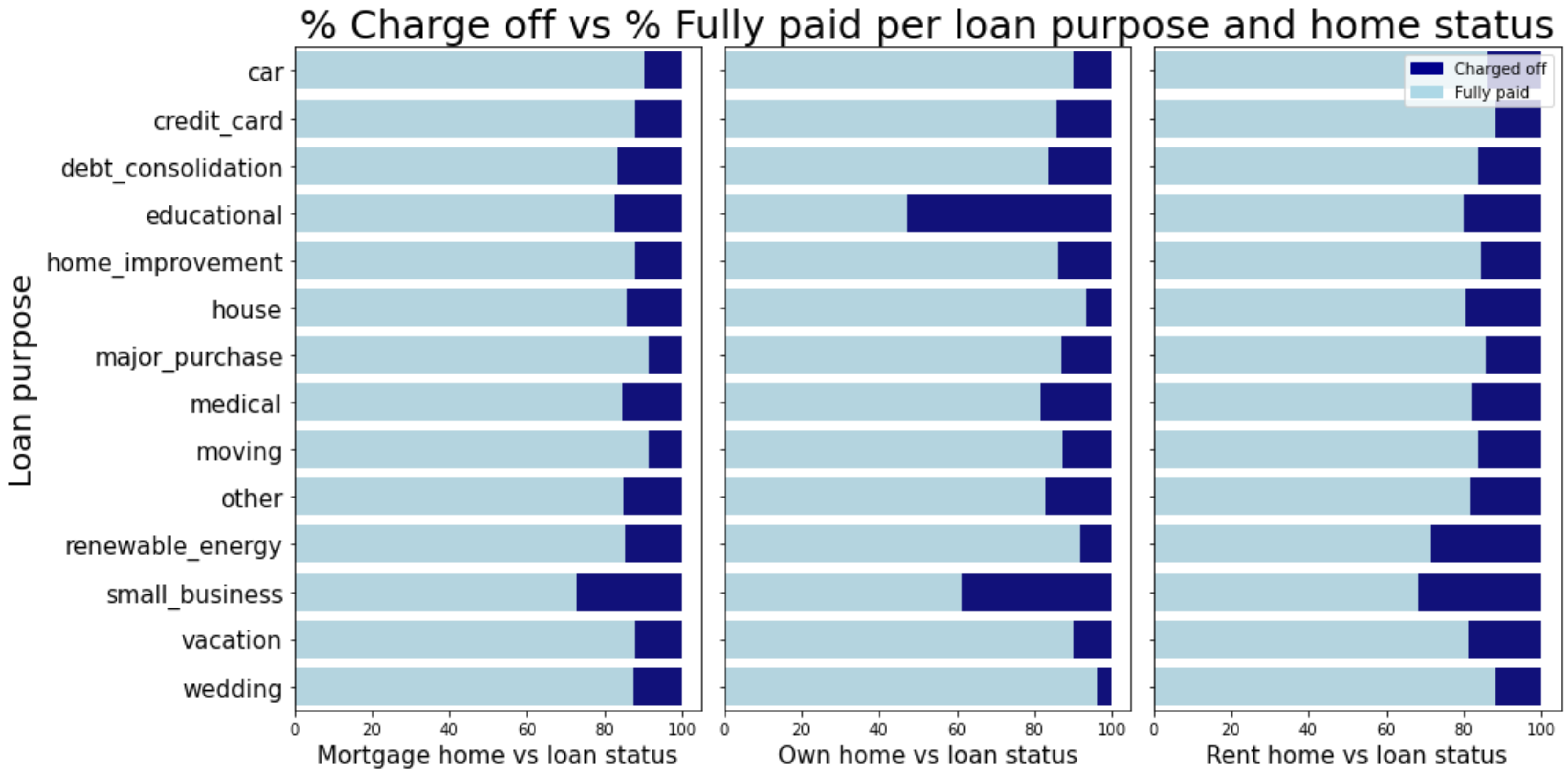
- The chart shows the percentage of charged off loan amount in term of loan purpose.
- If a customer applies a loan for educational or small business, there may be a high potential that that loan may lead to bad debt.



Loan purpose and home ownership status

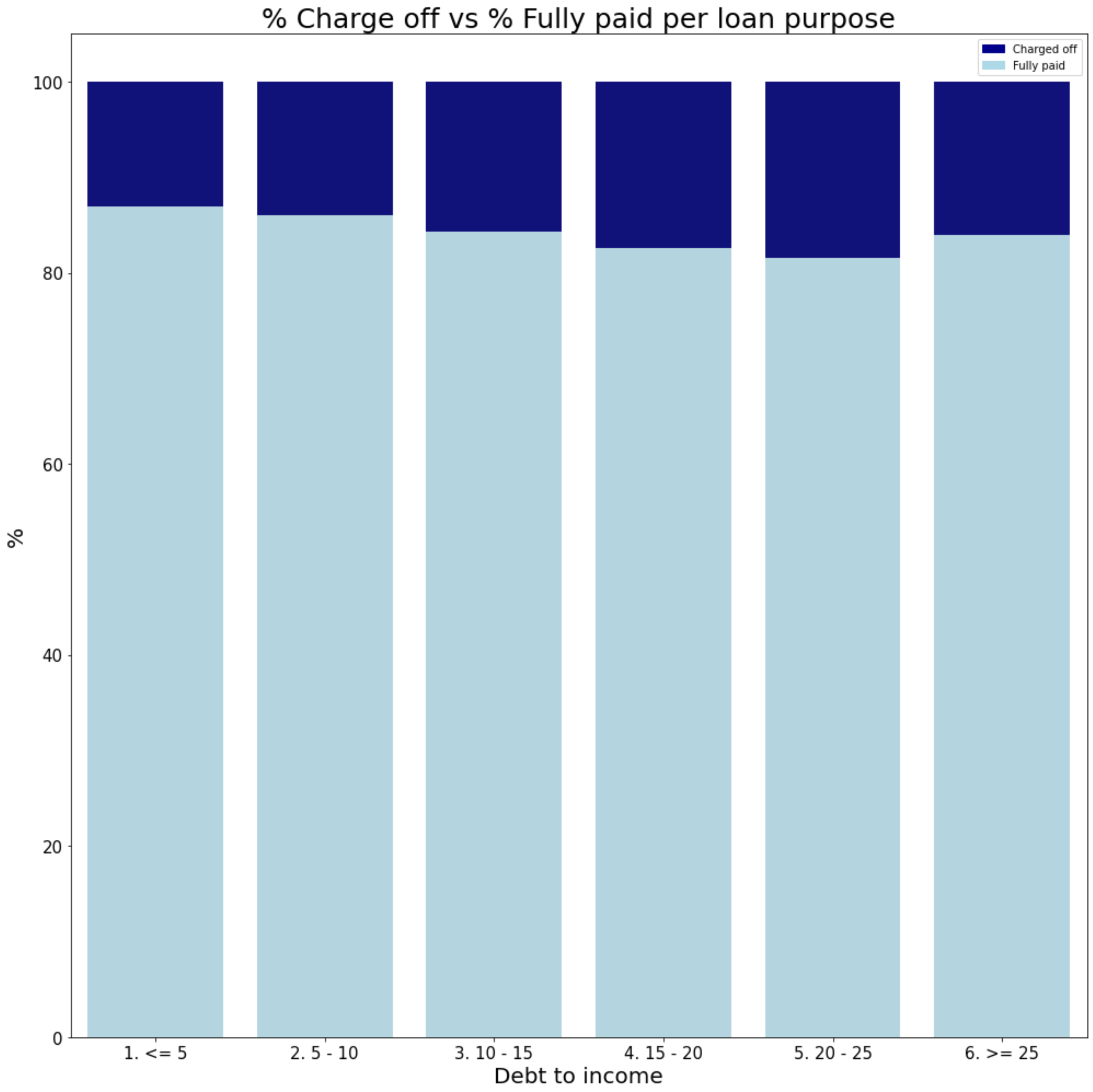
When adding home ownership status to loan purpose, we see there is an additional strong indicator, which are:

- own home - applying loan for educational
- own home - applying loan for small business



Debt to income ratio

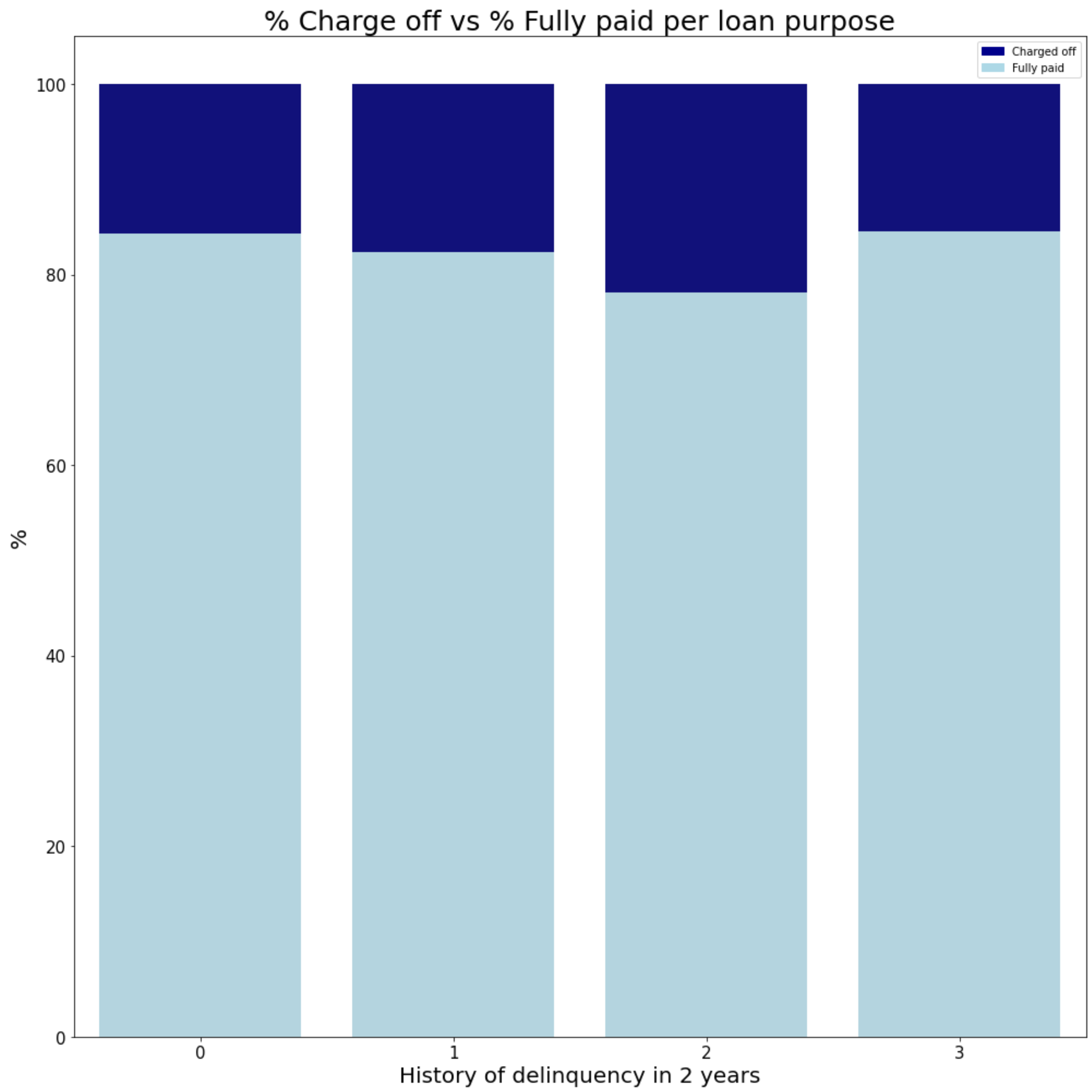
- Debt to income is an interesting metrics to determine customers' ability to pay back debts using their annual income. If customers have income way higher than debt, of course our loans are more likely to be paid back, vice versa.
- Surprisingly, the higher income customers have, the higher ratio bad debt is.



Delinquency in last 2 years

- An used-to-be-bad customers in history may be still in currently. Usually, during approval process, customers' credit history are also taken into consideration.

- Customers, who have been 2 times delinquency in last two years, have percent of charged off 40% higher than normal customers. It's look significant but the absolute number of loans are not very significant compared to normal customers. Therefore, it's not a very strong indicators and not very statistical significance.

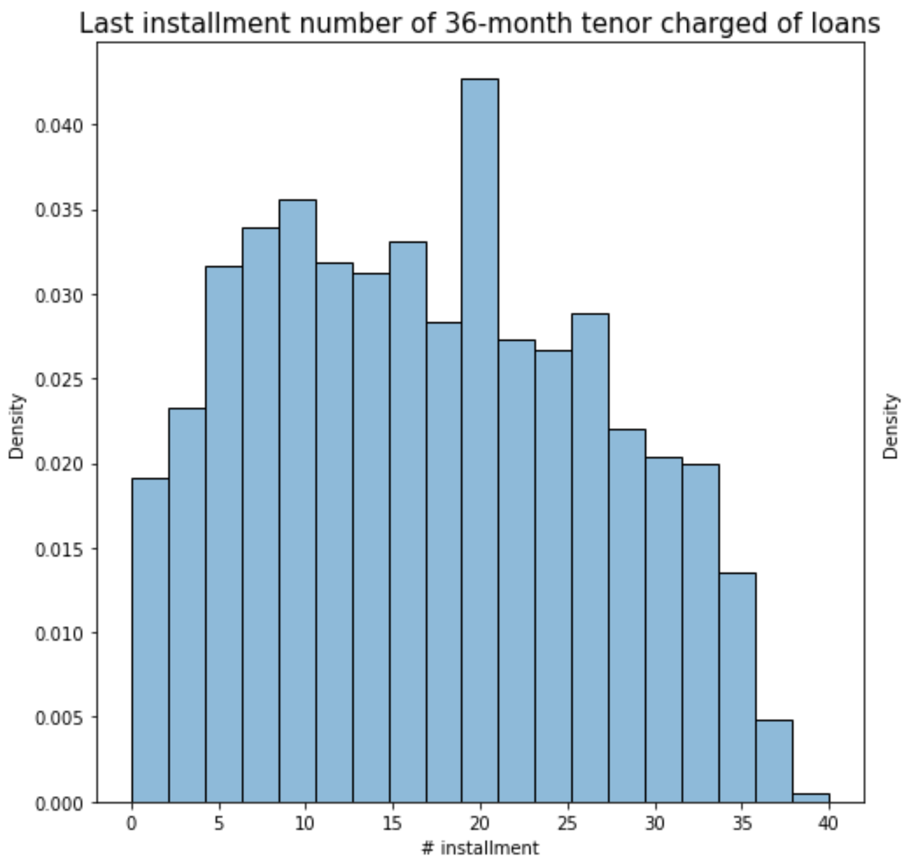


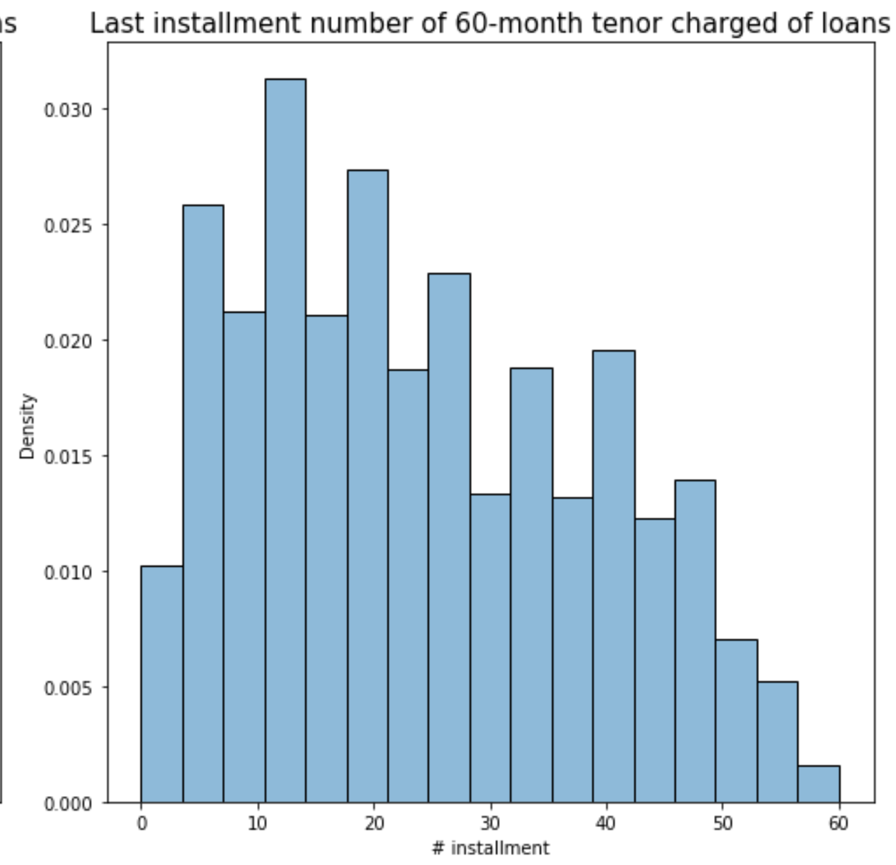
Post disbursement – last installment before bad debt

The two right charts show the last installment of all charged off loans. It helps us to determine the most last installment before customers stop paying.

- For 36-month tenor: the peak is at 10th installment.
- For 60-month tenor: the peak is also from 10th to 15th installment

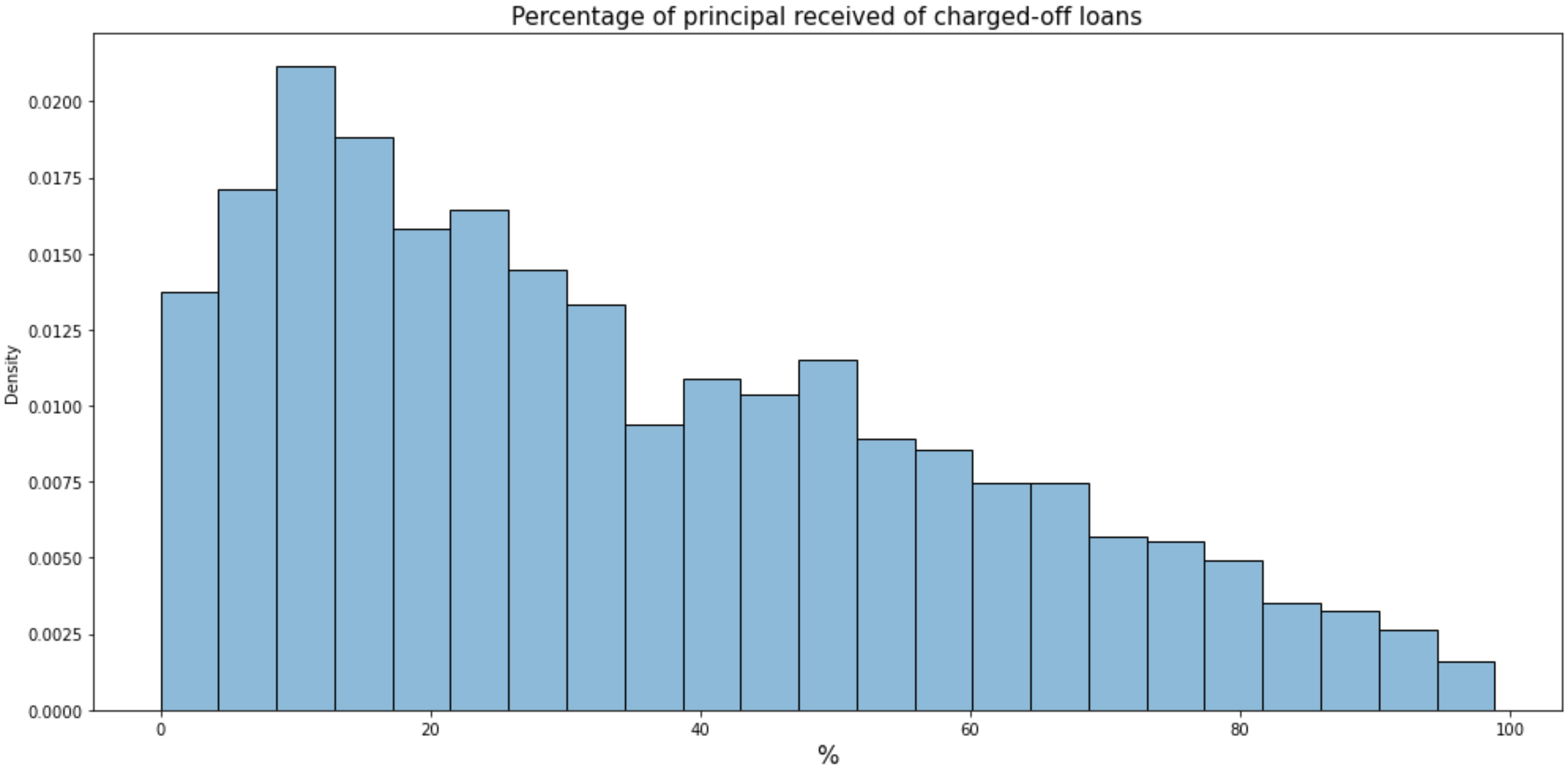
➔ 10th to 15th installments are period at which customers may decide to stop paying.





Post disbursement – percentage of principal received before bad debts

- Similarity, this chart focuses only in charged off loans.
- Since all loans are installment loans, which are principal are increasing and interest are decreasing over time.
- If customers tend to stop paying after 10th to 15th installments, the amount of principal received is about 15% of loan amount.
- This indicator may be also covered in previous indicator.



Conclusion

- Pre-disbursement indicators
  - Year of employment
  - Loan purpose (may or may not combine with home ownership status)
  - Debt to income ratio
- Post-disbursement indicators
  - Number of installments
  - Percent of principal received

Appendix notices:

- Either target customers of this lending club have very abnormal behaviors or
- Either this dataset is modified on purpose or
- Either we did any mistake in any step of analysis

  ➔ There are two very strange indicators such as the longer year of experience, the higher dti ratio, the higher charged off ratio.