Loan approval case study

Business goal

Reduce the number of defaulted loans without significantly reducing business potential by rejecting borrowers who will pay off.

Data issues and cleaning

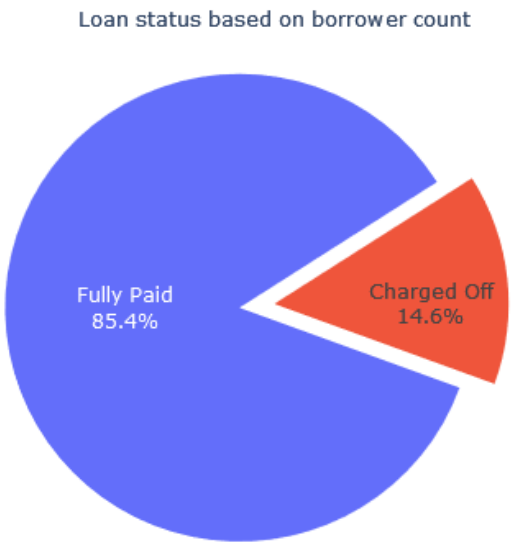
- Lot of variables contain high amount (>90%) of missing data. (solution: remove variables)
- Variables have unique data, not useful for analysis. (solution: remove variables)
- Rows still contain missing variables. Data Imputation (global/local mean, median, mode) (not applied in this assignment)
- Dates are read as string and converted to date type
- Term parsed and rewritten to remove space
- Interest rate: % removed, converted to float

Derived variables

- installment_monthly_income_ratio
- loan_amount_income_ratio
- funded_percentage
- issue_imonth

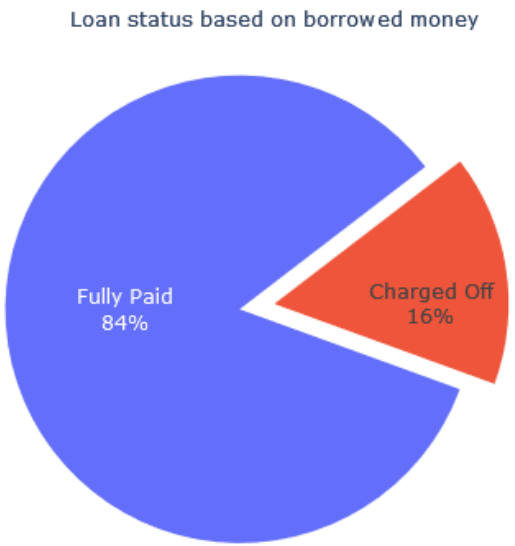
Charged Off distribution in data

- 14.6% of borrowers default on the loan.



Charged Off distribution in data

- If we check the borrowed amount (defaulted) it is even higher (16%), indicating that the average defaulted loan is slightly higher than the mean of loans. (see detailed distribution later)
- **The business goal is to reduce this number without significantly reducing business potential by rejecting borrowers who will pay off.**



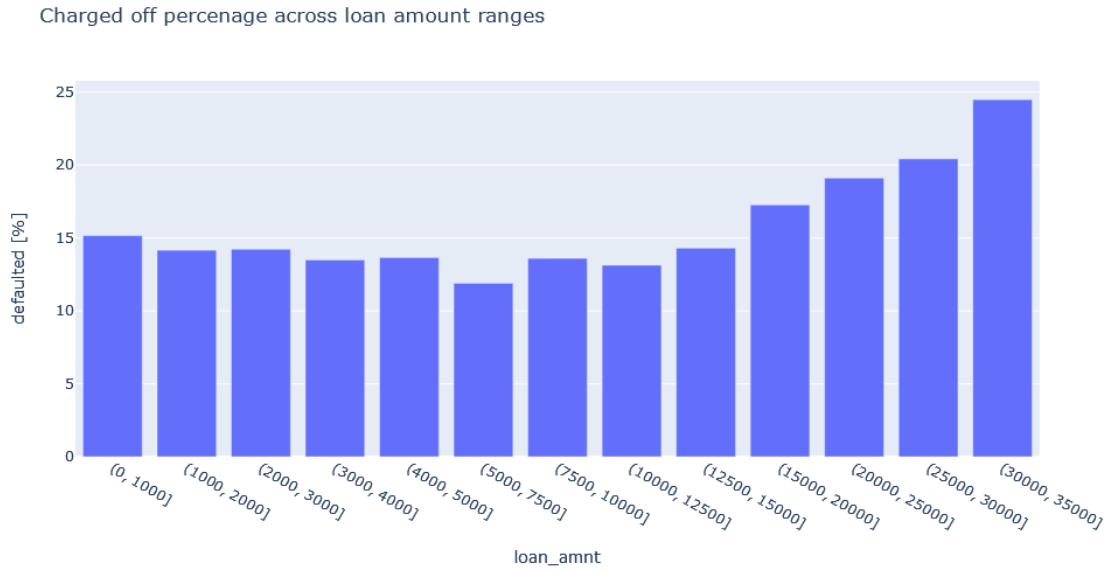
Drivers behind Charge Off

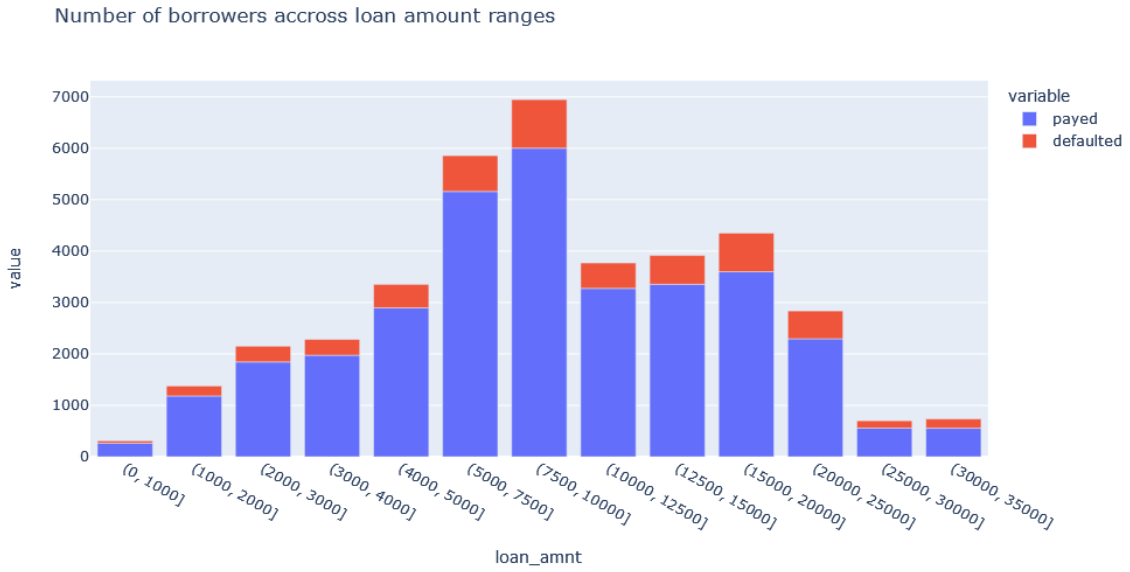
Loan amount

- The risk of charge off is the lowest around 5000-7500 and start to increase above 15000.

Recommendations

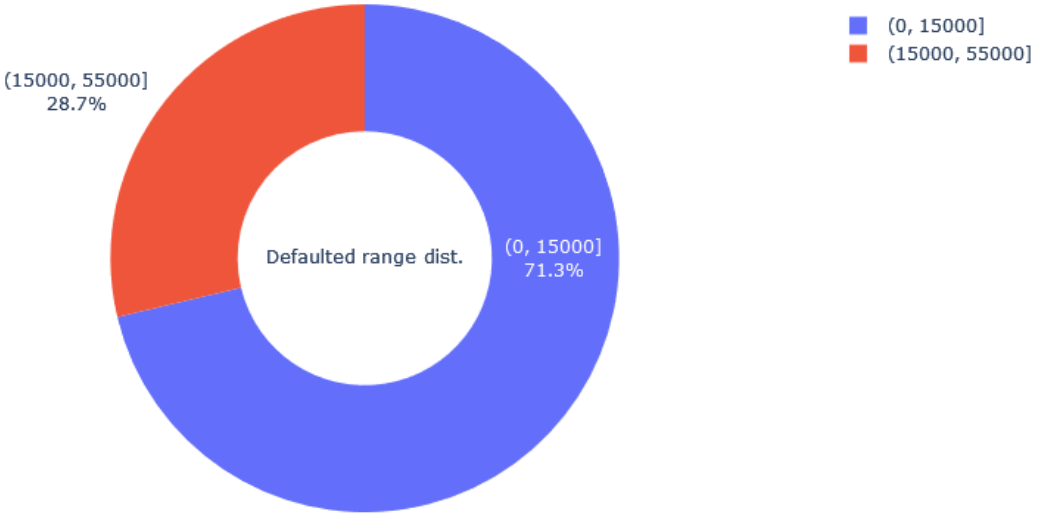
- On loans >15000 pay attention on other variables and potentially be more restrictive on the loan amount.





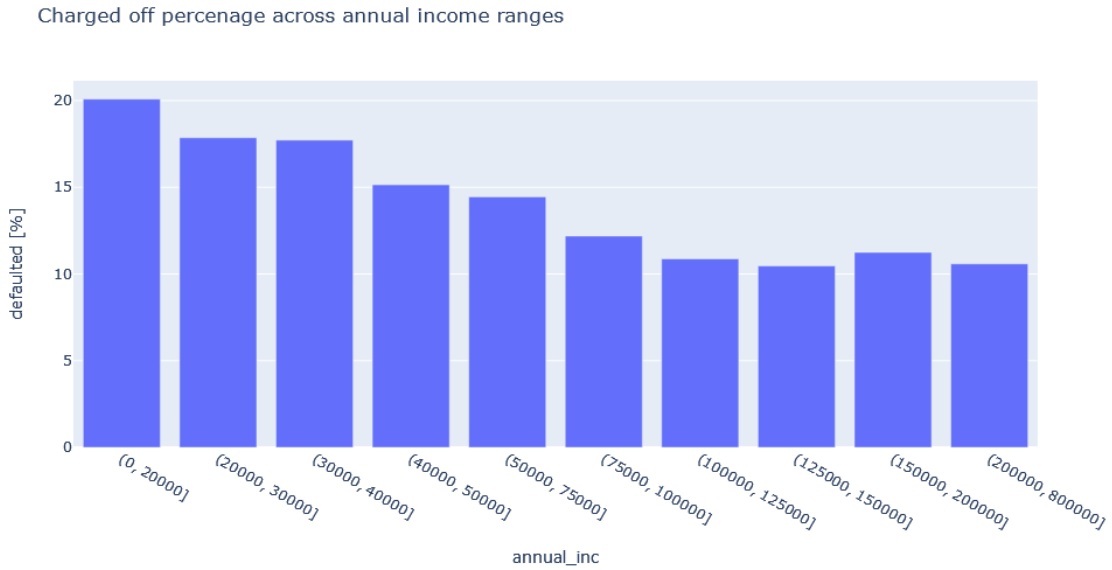
Loan amount

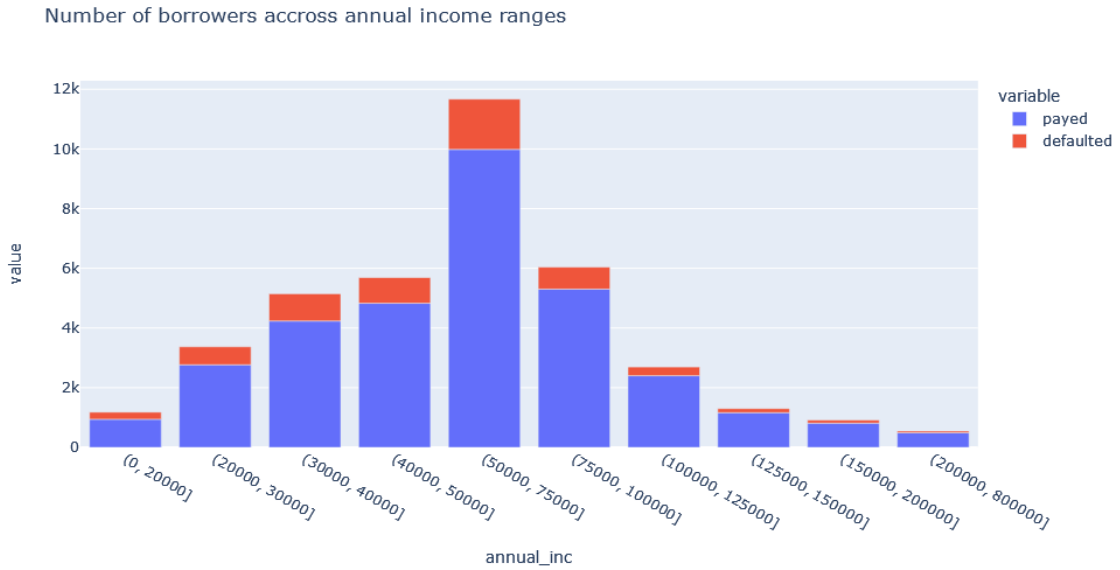
- Almost 3/4 of the loans are from the <15000 range. Currently these loans are in the lower risk range. If the company starts to target a market of higher loans, should follow the above mentioned recommendations.



Annual Income

- In general, lower annual income presents higher risk of default. Above 100k, income increase does not further decrease the risk of default.



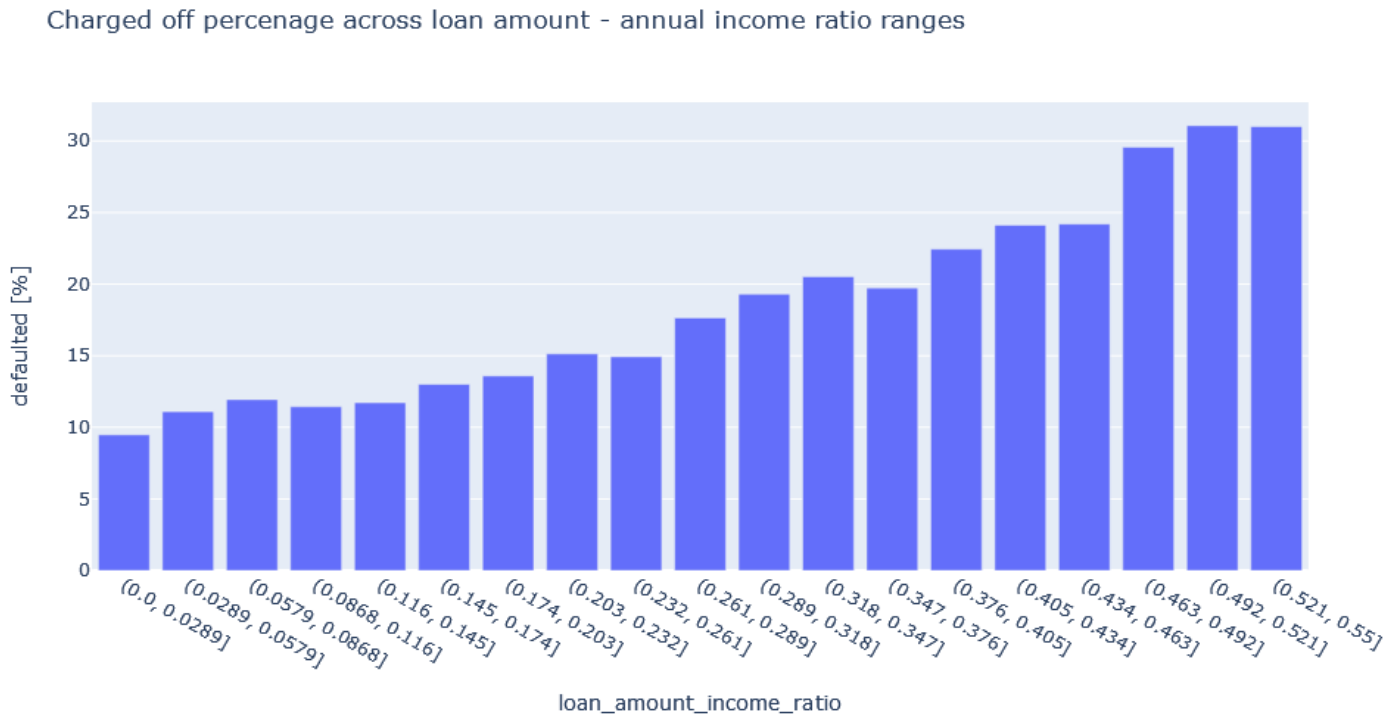


How borrowed amount relative to annual income affects default rate (percentage)?

- Higher borrowed amount - annual income ratio (borrowed / annual income) increases the risk of default

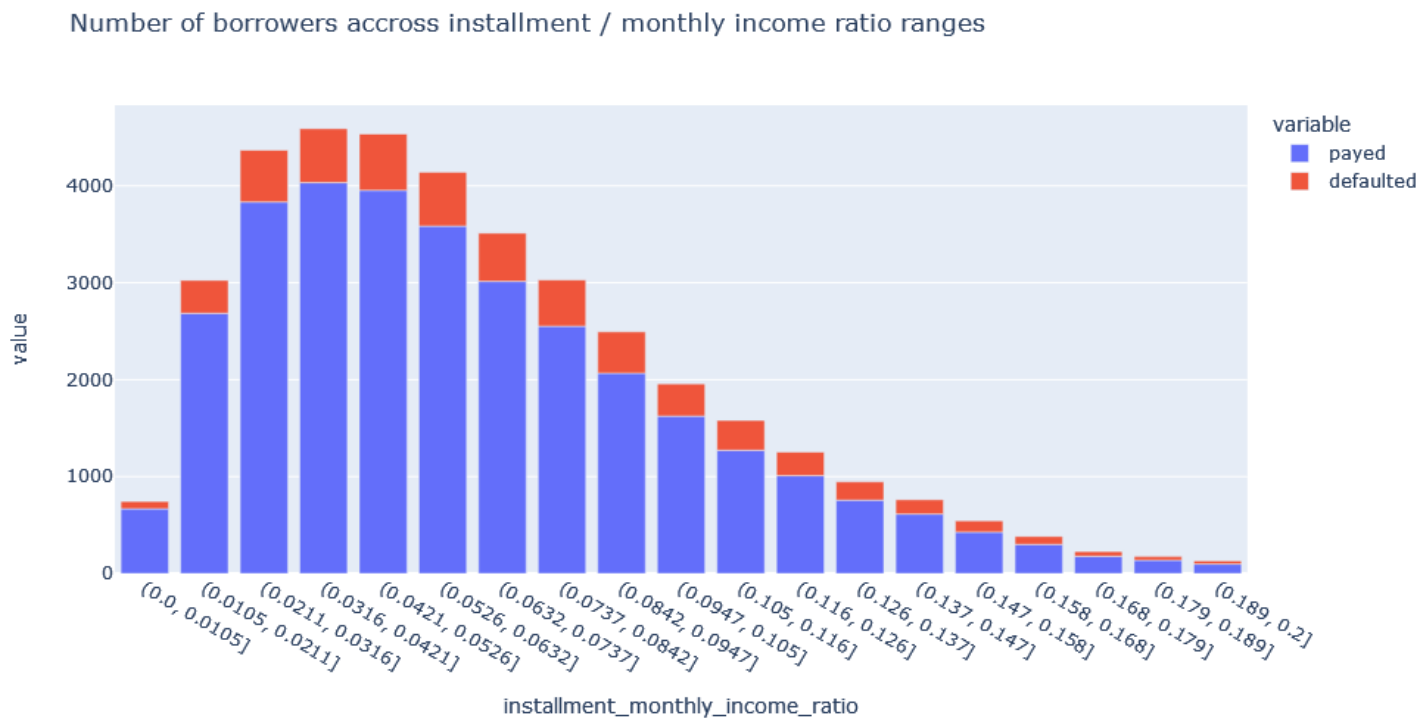
Recommendation

- for ratio >0.26 check other indicators and potentially decrease the accepted loan amount as well.



Instalment

- The number of borrowers across instalment / monthly income ratio ranges follows normal distribution



Instalment

- As the monthly instalment cost increase relative to the borrower's monthly income the chance of default increases.

Recommendation

- decrease the accepted loan amount therefore the ratio will decrease
- offering longer term might decrease ratio but * *see following analysis*

longer term effects default ratio respect to instalment / monthly income ratio

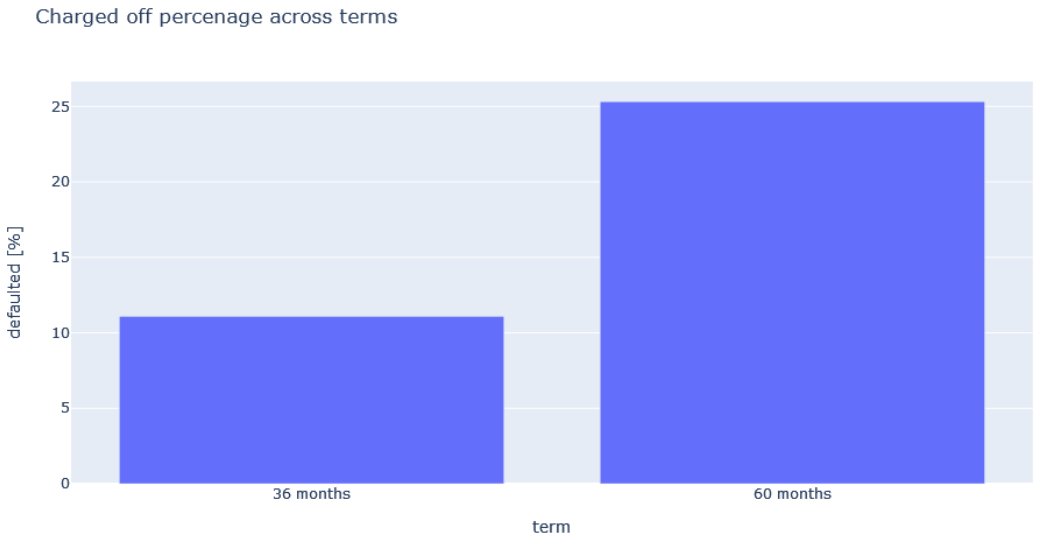
- Based on the data, longer term loans has a significantly higher risk of default.
- Even with lower instalment / monthly income ratio the chance of default is higher for longer term. It is **not recommended** to offer **significantly** longer term for those who has higher instalment / monthly income ratio.
- 40 or 48 months might be an option, but it needs data to prove.

Term

- In general longer term (60 months) has significantly higher (2x) risk of default.

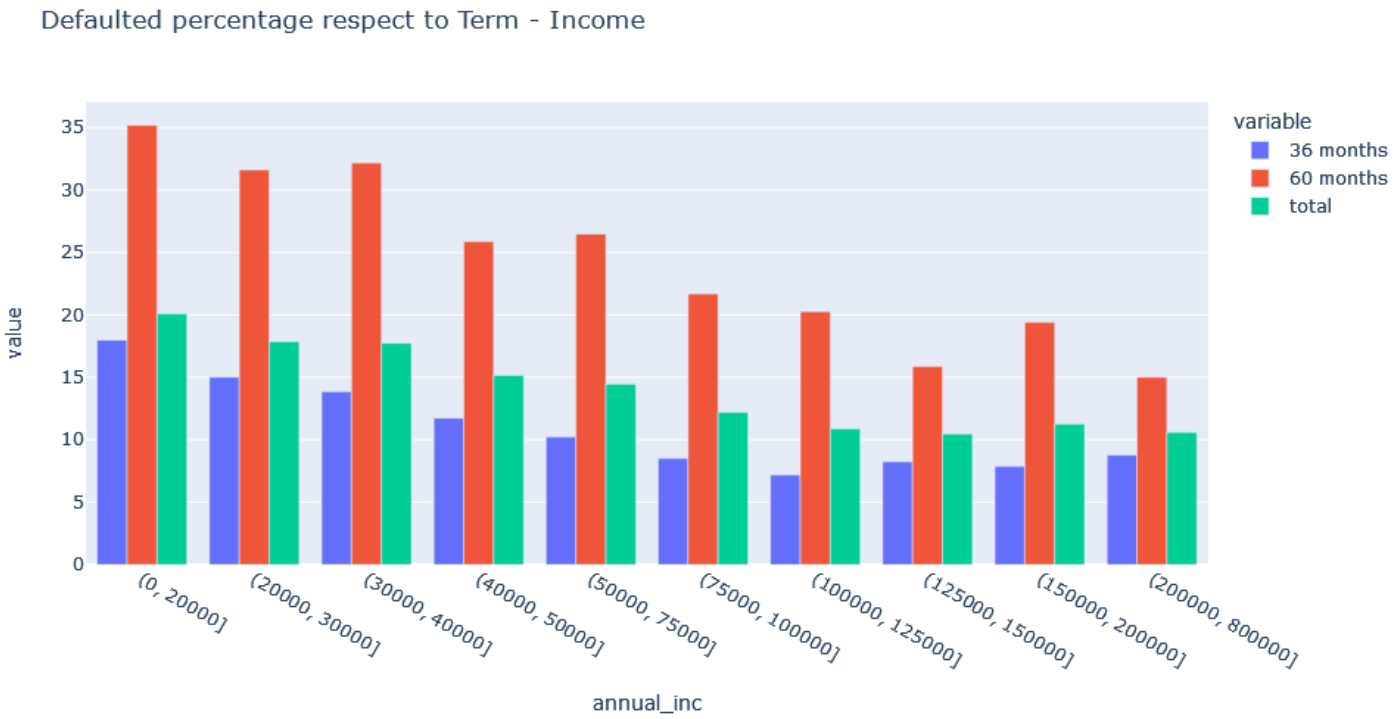
Recommendation

- Decrease the term to 36 if possible or decrease the loan amount to fit to the 36 months instalment. *As shown later decreasing the loan amount keeping the 60 months does not decrease the risk of default*
- Check other risk factors and reject the loan if 60 months term coexist with them



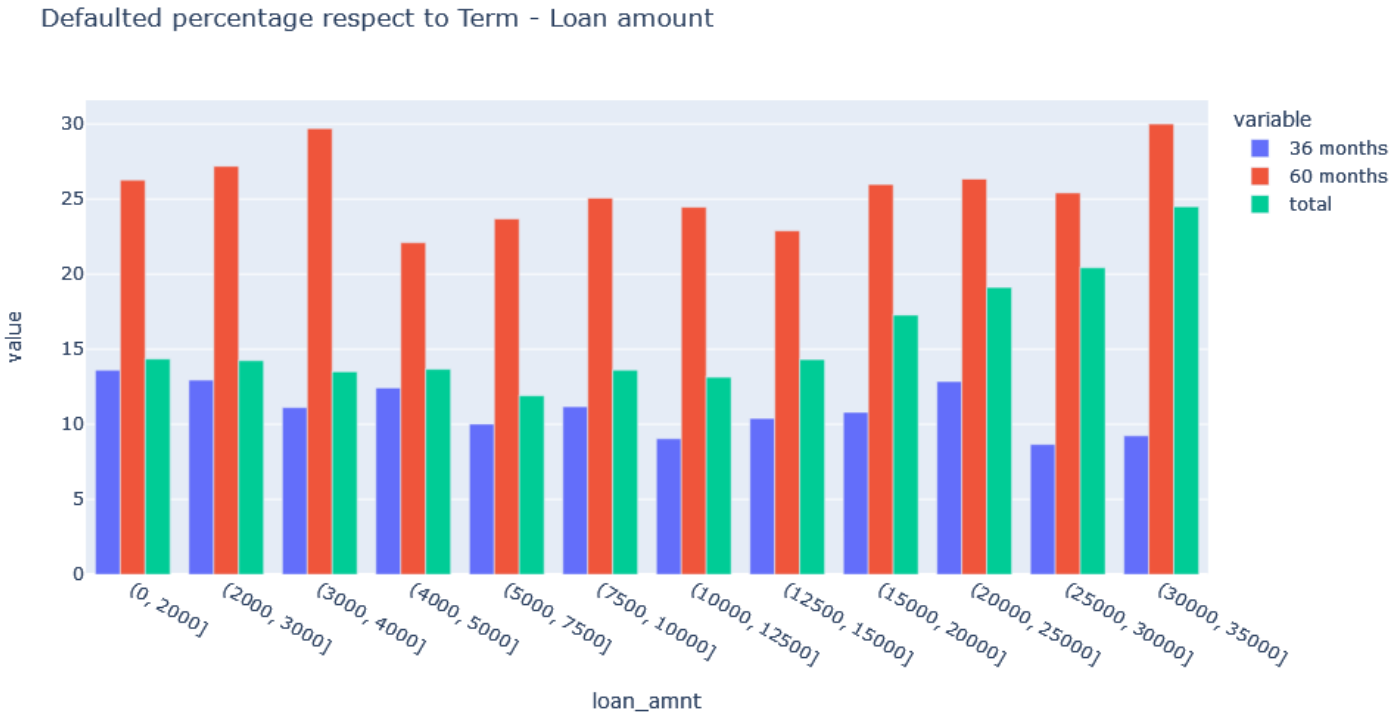
Defaulted percentage respect to Term- Income

- Annual income increase lowers the risk for both terms (36 months, 60 months), indicating that annual income feature (variable) is a consistent predictor.



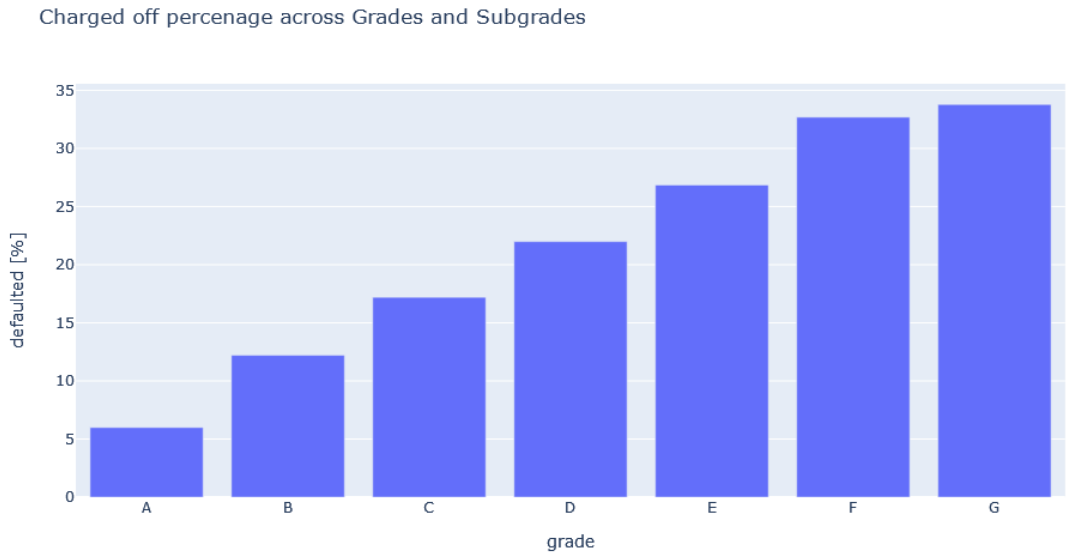
Defaulted percentage respect to Term- Loan amount

- Decreasing the loan amount keeping the 60 months term does not decrease the risk of default.



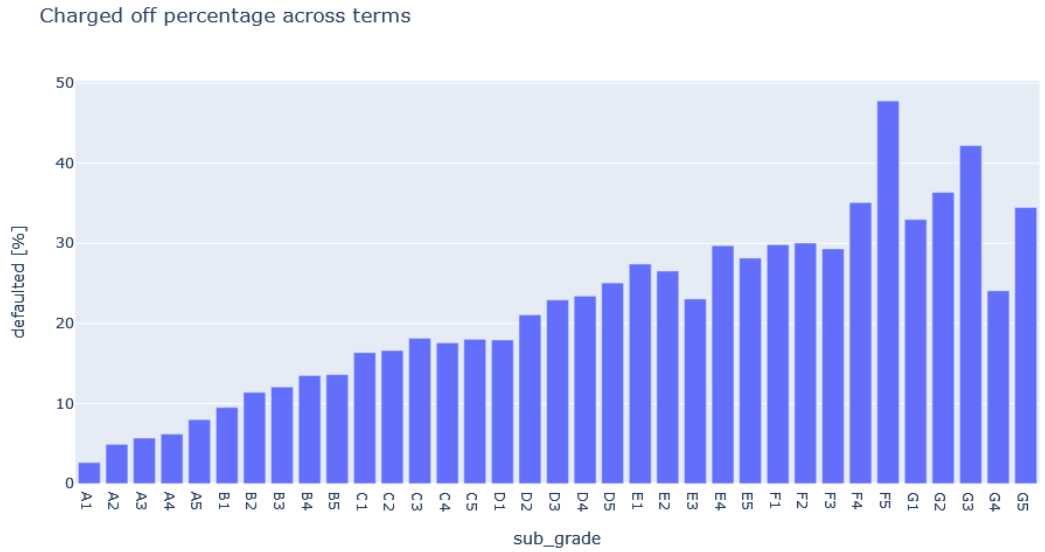
Grade

Grade is a strong indicator of charge off



Sub-grade

- Sub Grade is a strong indicator of charge off. It provides reliable and consistent prediction of default between A-D. E and above sub-grade does not give reliable indication.



Grade and Subgrade

Recommendation

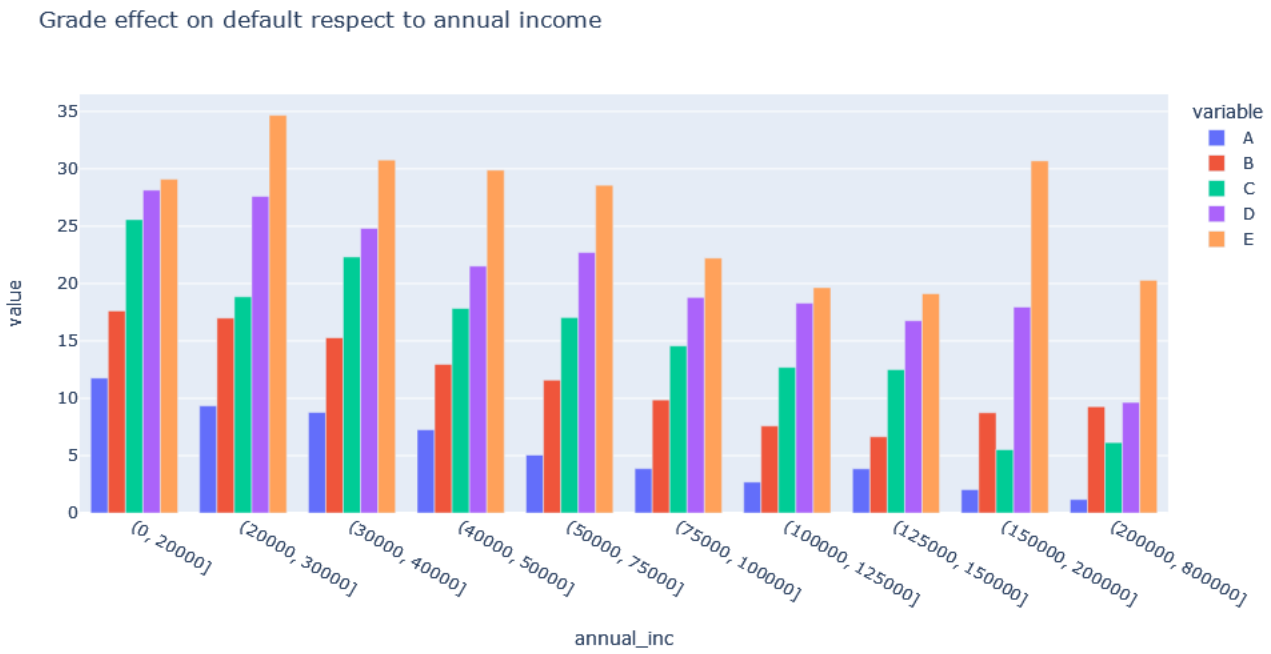
- Use grade and subgrade as a risk indicator between A-D.
- E and above use grade subgrade dose not consistent.
- As Grade decrease (A->G) consider other factors and decrease the loan amount or reject the loan.

Grade effect on default respect to annual income

- Both annual income and grade are strong and reliable indicators. F, G not included in the analysis because of low amount of data points per bin. But data shows that grade D, E are less consistent and incur higher risk.

Recommendation

- Annual income used in conjunction with grade can be a good predictor of default

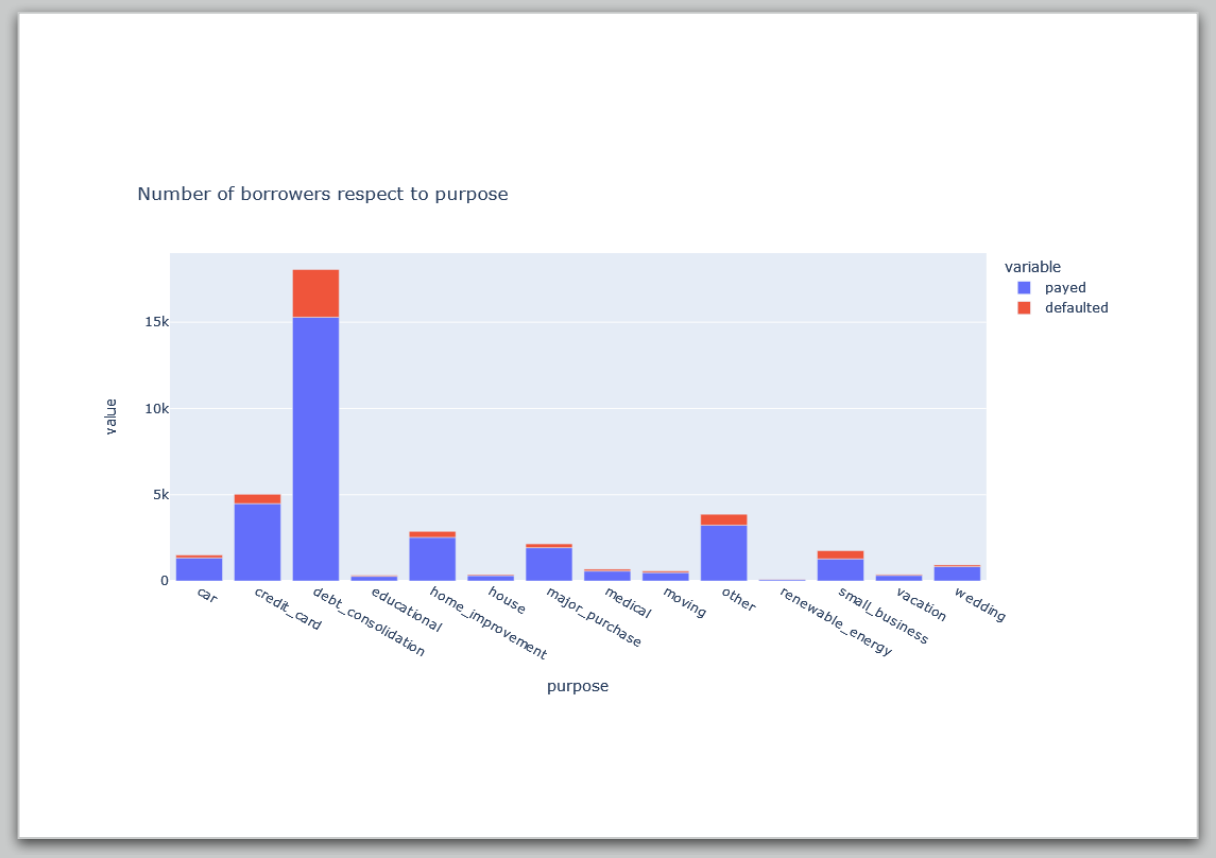


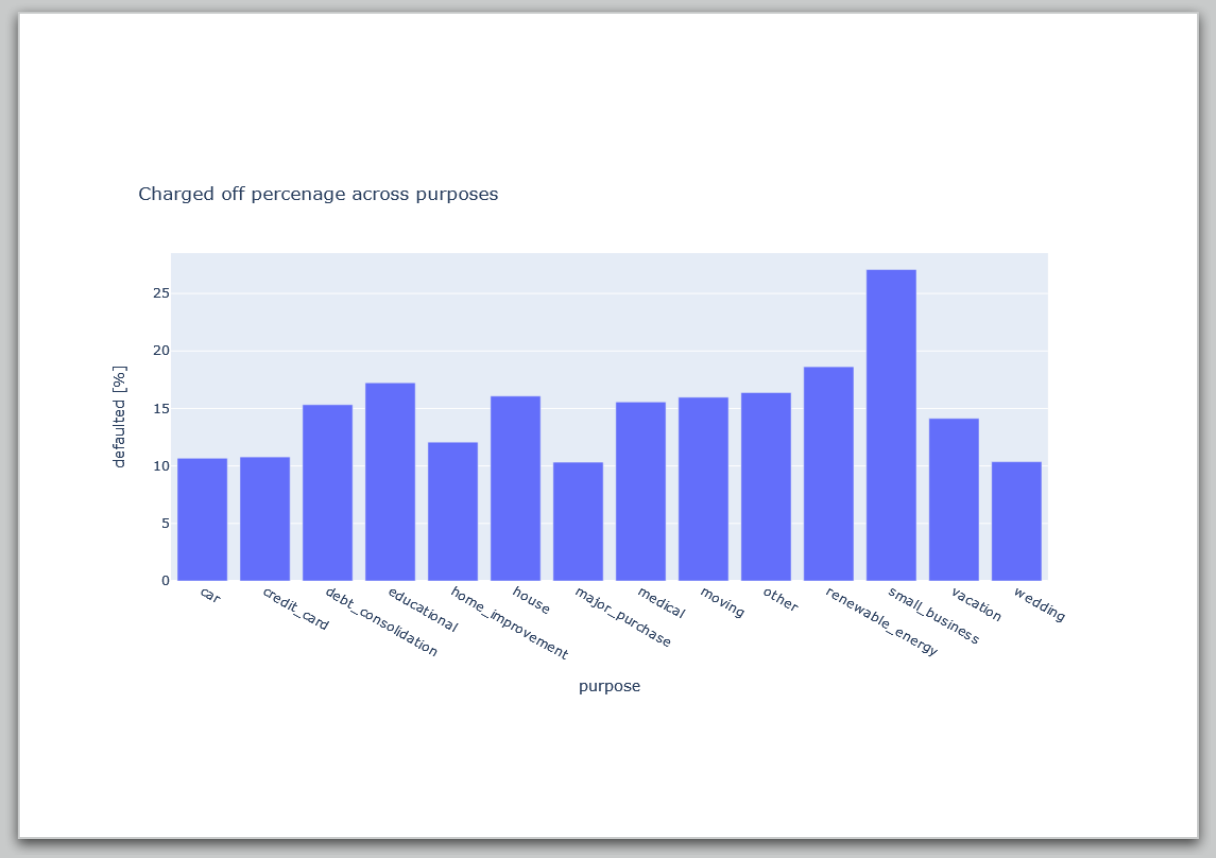
Purpose

- Small business loans carry a significantly higher risk of default.

Recommendation

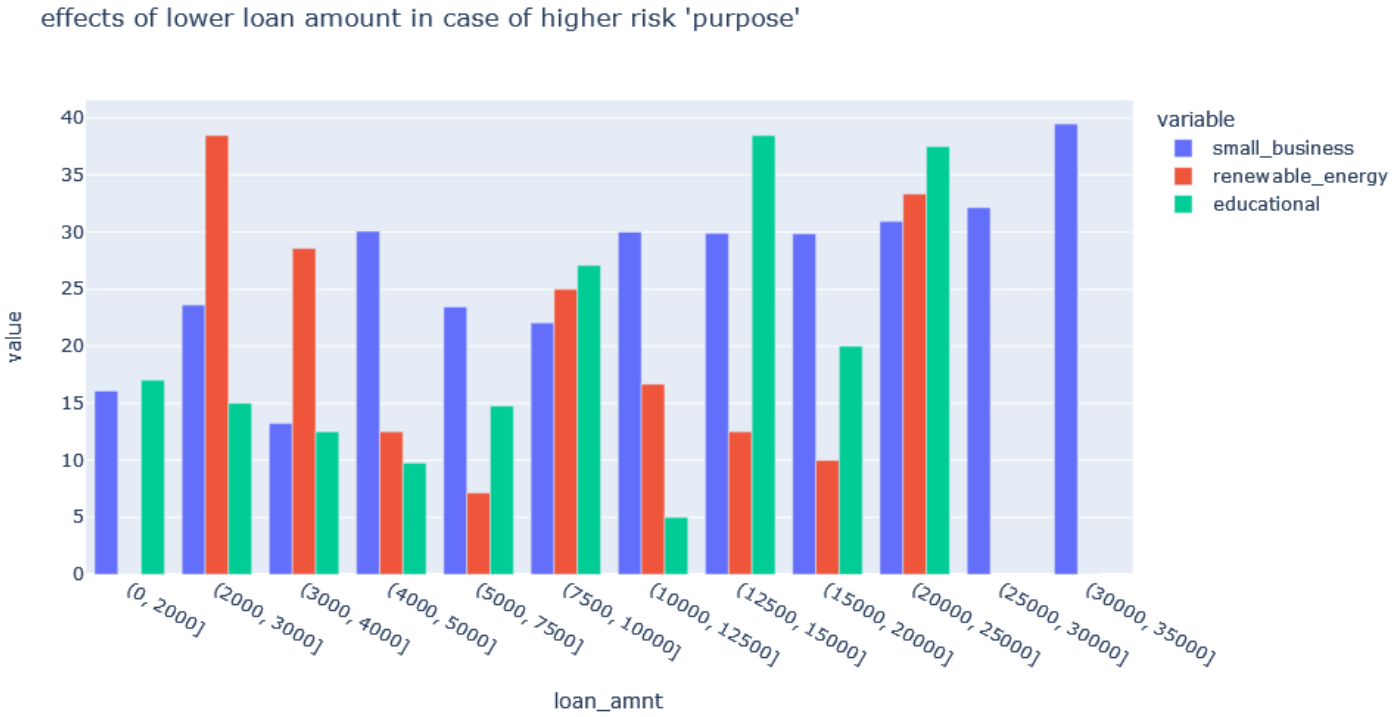
- Check other variables in case of small business loans





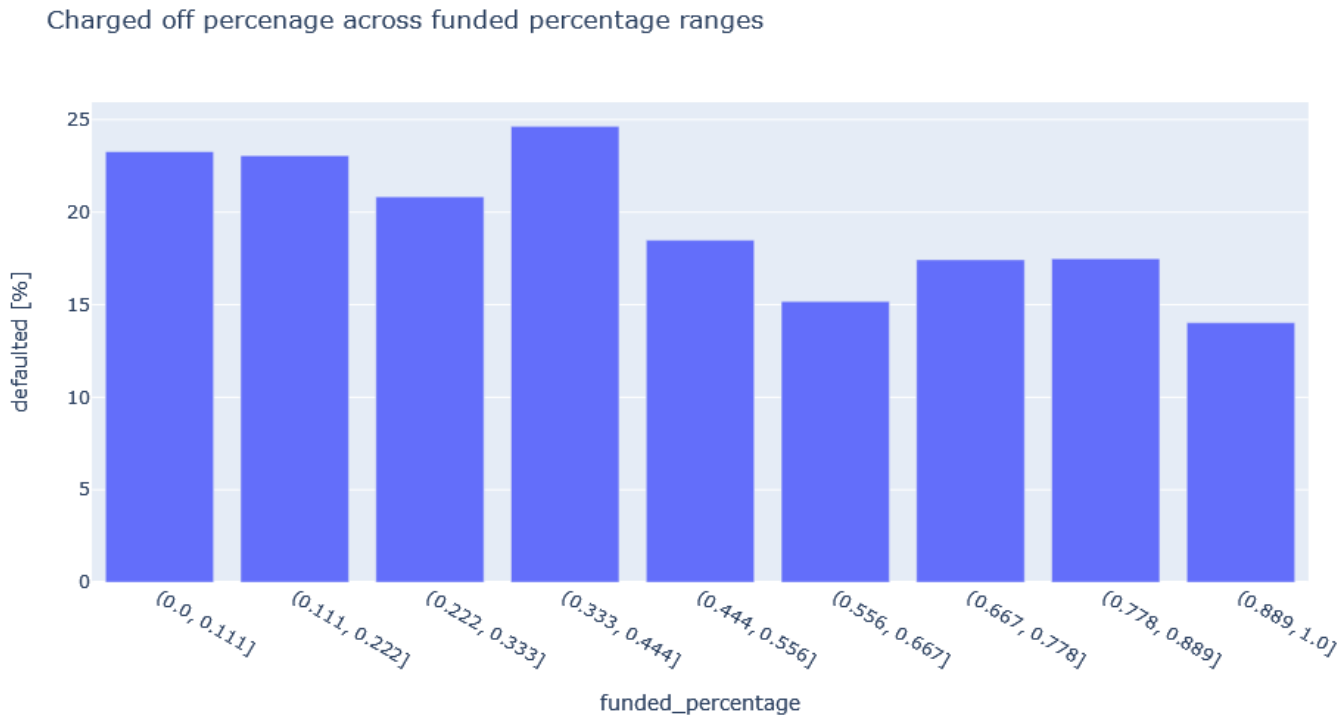
Possible effects of lower loan amount in case of higher risk 'purpose'

- small business: Lower loan amount can slightly decrease the risk.
- educational: Lower loan amount can decrease the risk.
- renewable energy: Lower loan amount can increase the risk.



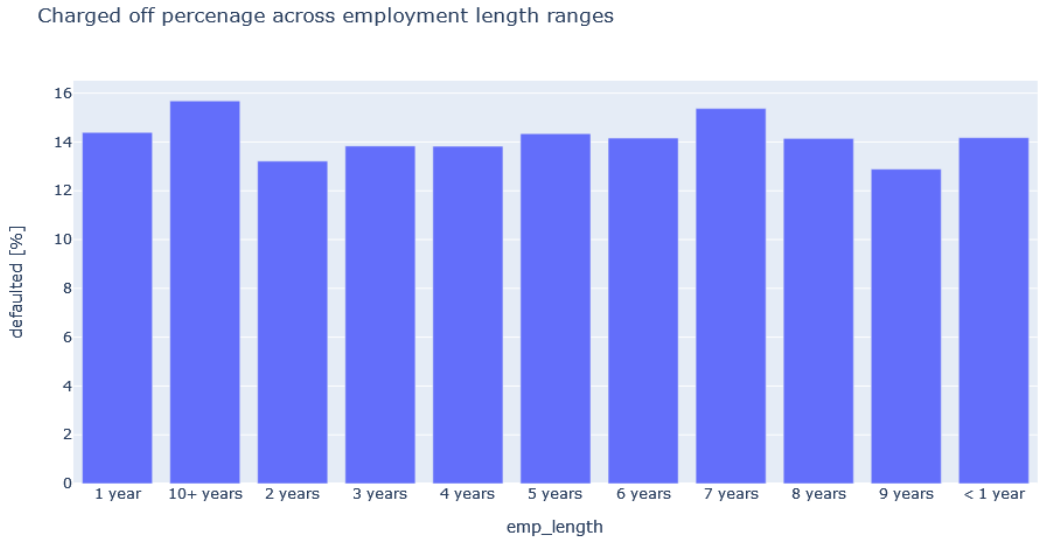
Funded percentage

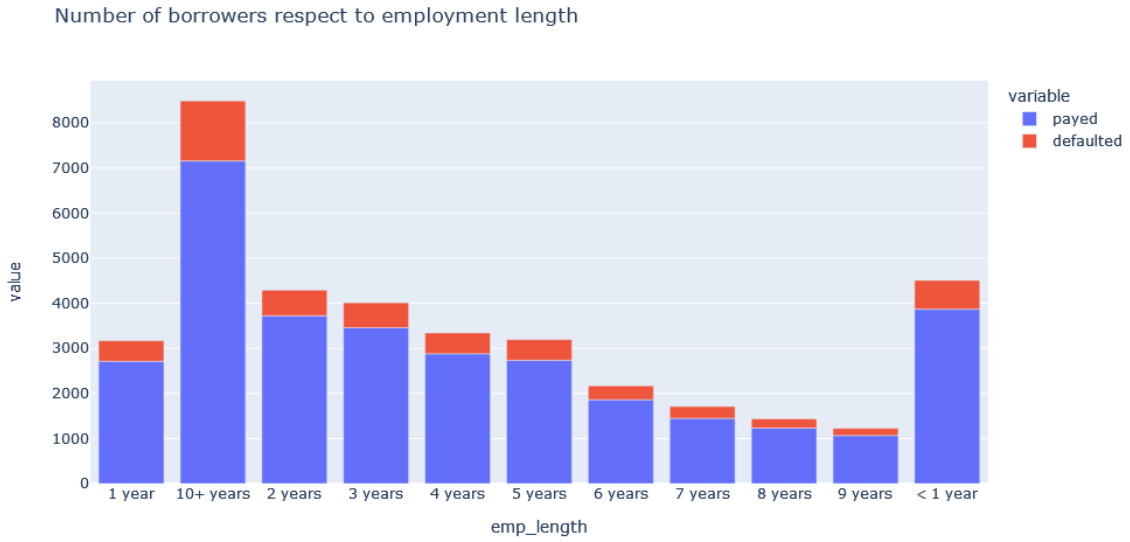
- Moderate underfunding (not providing the requested loan amount) does not increase the risk of default. Only substantial underfunding >50% can increase the risk.



Employment length

- Employment length is not a good predictor of default.
- Not recommended to use it.



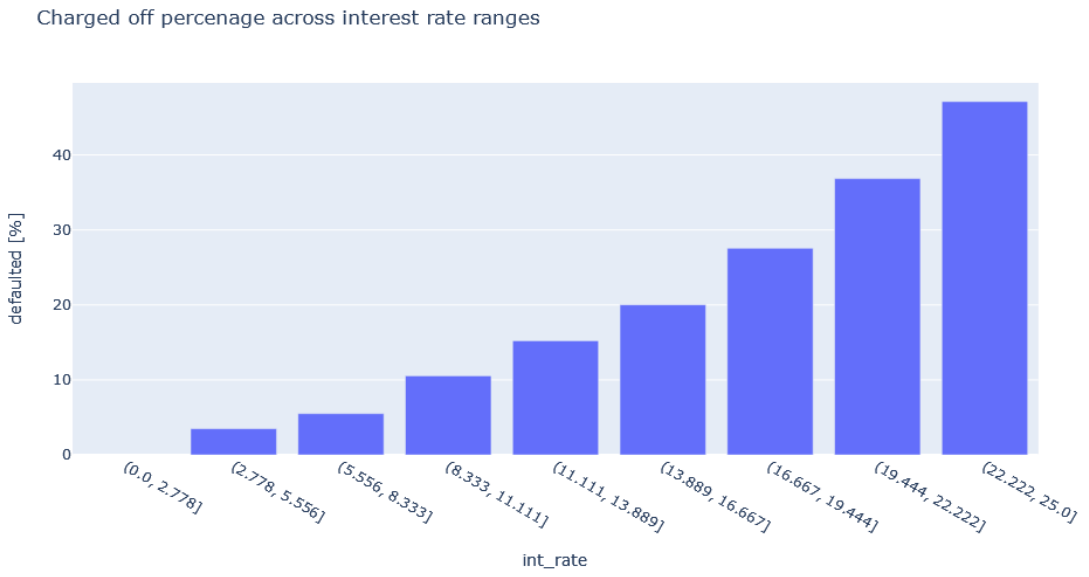


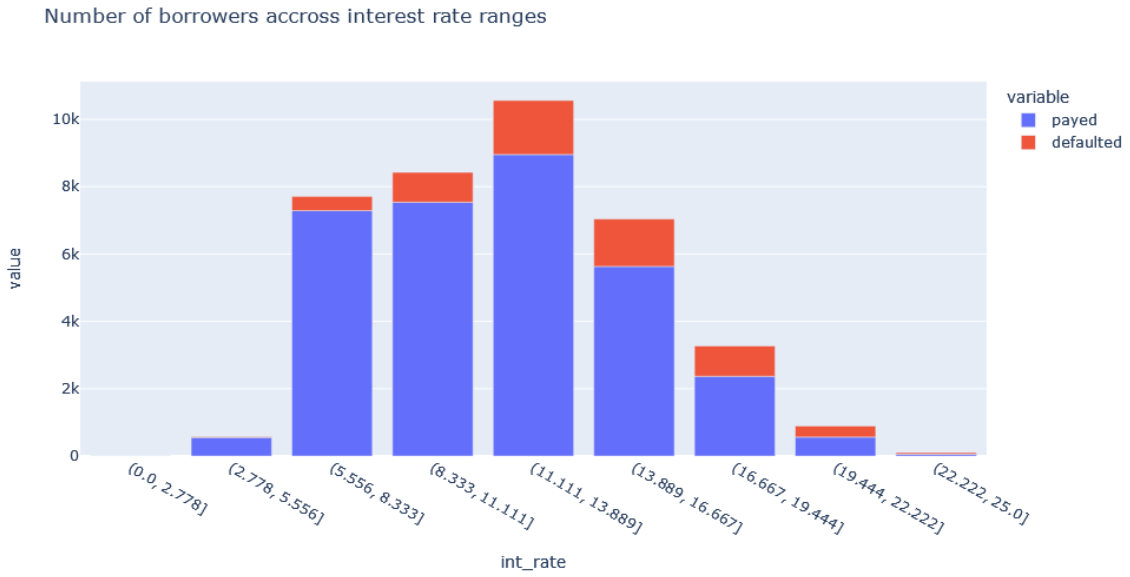
Interest rate

- Interest rate is a good predictor of risk. High interest rate has significantly higher default rate. It is a very strong indicator.

Recommendation

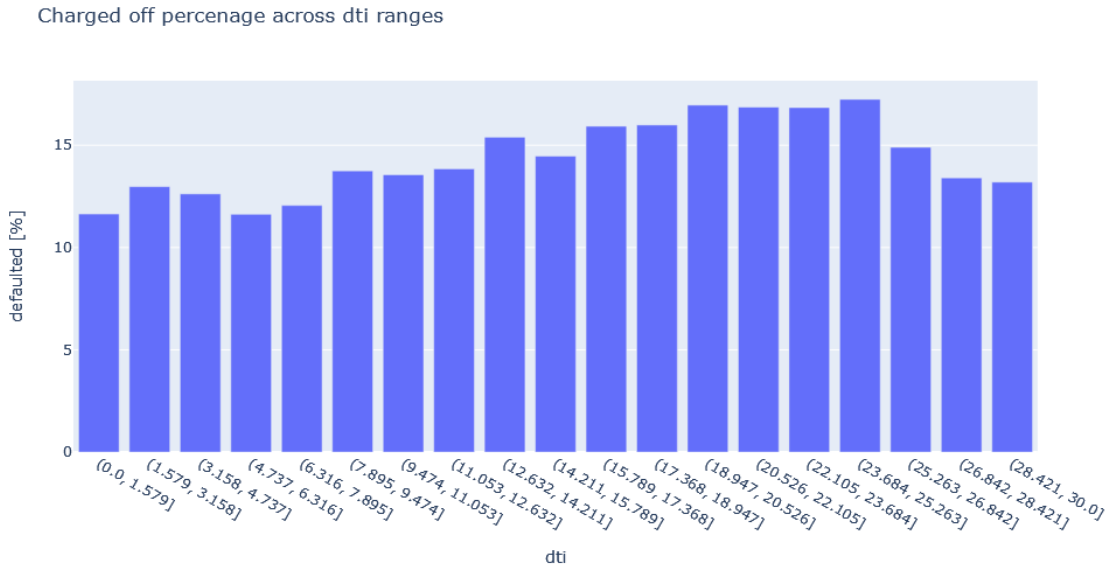
- If high interest rate combined with other strong negative predictors do not approve the loan.

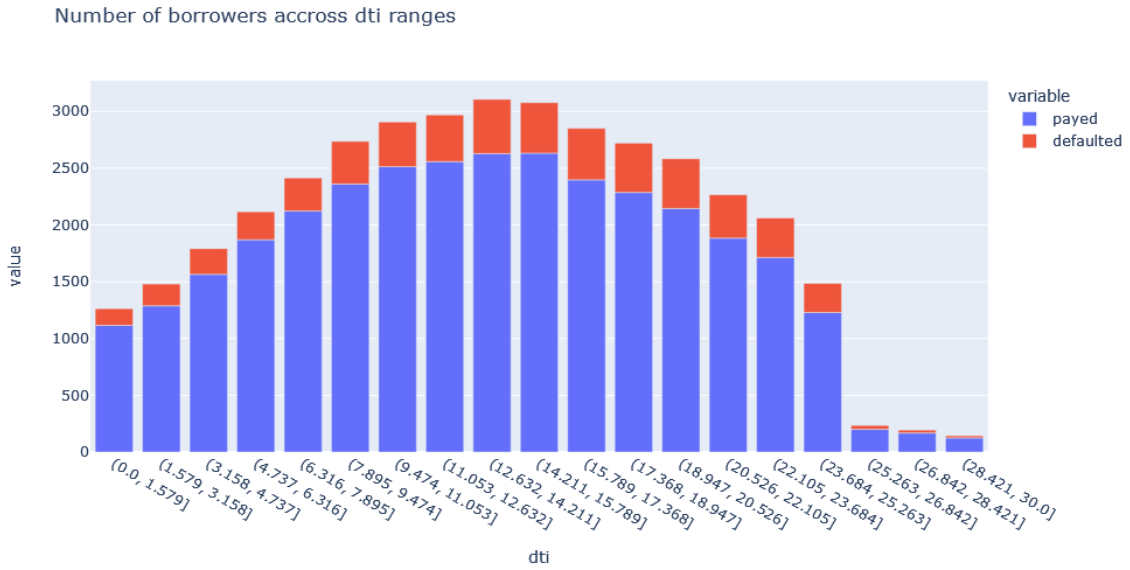




DTI

- DTI between 15 and 25 shows a higher risk of default. Above 25 default ratio drops*.
  **More data in this range could reinforce this observation.*
- **Recommendation**
- Not the best predictor, but can be used in conjunction with other stronger predictors.





Summary of Recommendations

- Annual Income (strong predictor)
- instalment / monthly income ratio
- Term (longer = higher risk) (strong predictor)
- Grade (strong predictor)
- Sub-grade (A-D reliable)
- Annual income used in conjunction with grade can be a good predictor
- Small business loans represent higher risk (check other predictors)
- Moderate underfunding does not increase default
- Employment length not relevant (do not use)
- High interest rate = high risk (strong predictor)
- Loan amount >15000 elevated risk, check other predictors