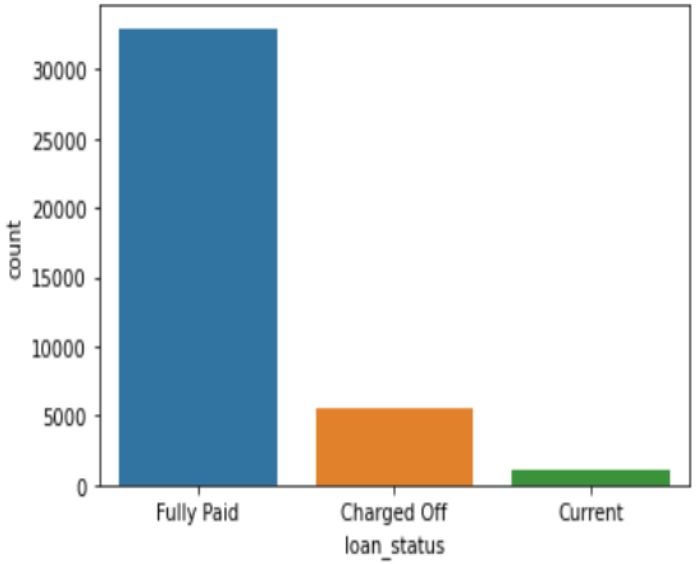
Case Study On Loan Default

Problem Statement

- When a company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:
  - If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
  - If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company
- 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.
- loan.csv data set is used for analyses.

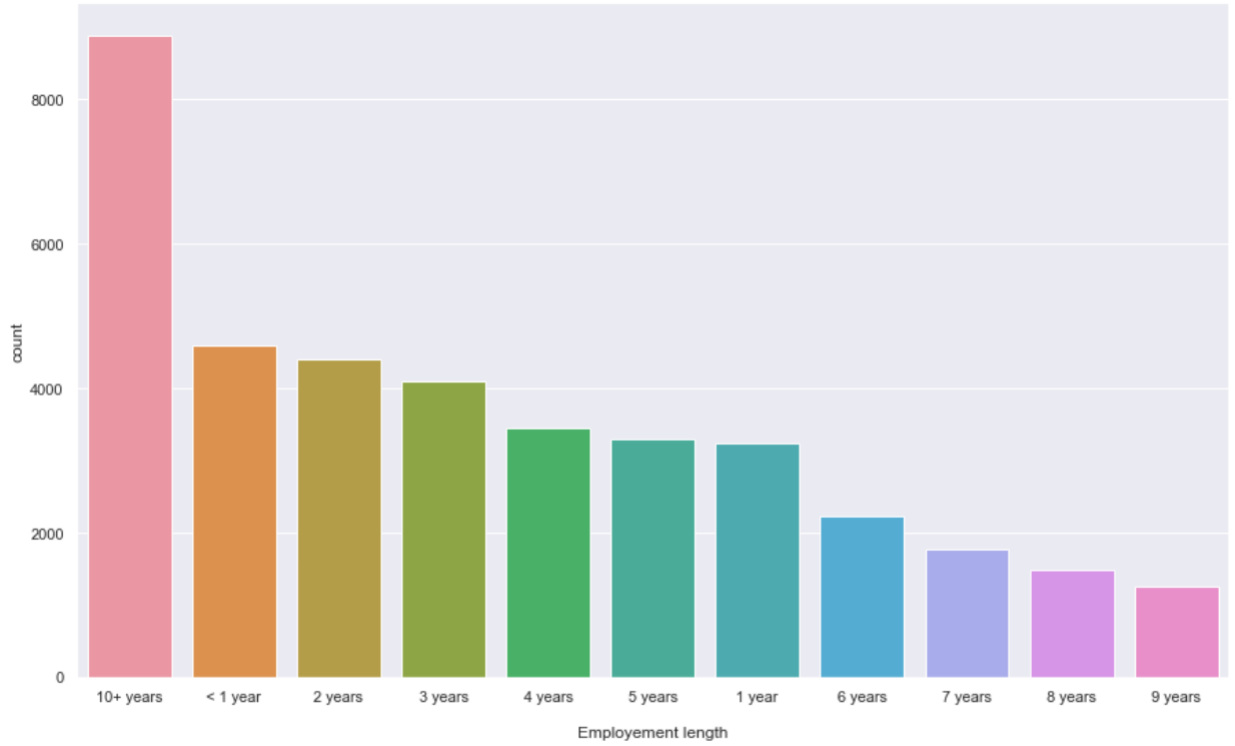
Univariate Analysis

Loan status analysis



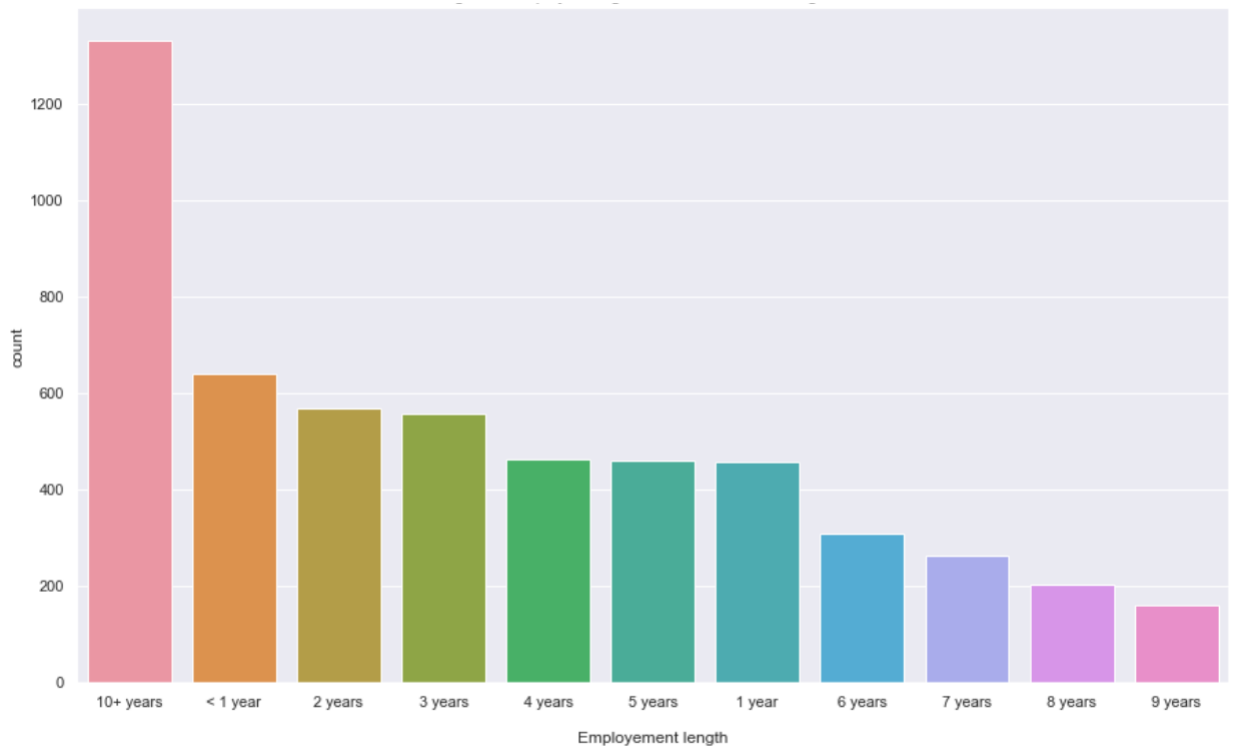
Above figure shows most of the loans are fully paid, and charged off loan few as compared to fully paid.

Employment length analysis



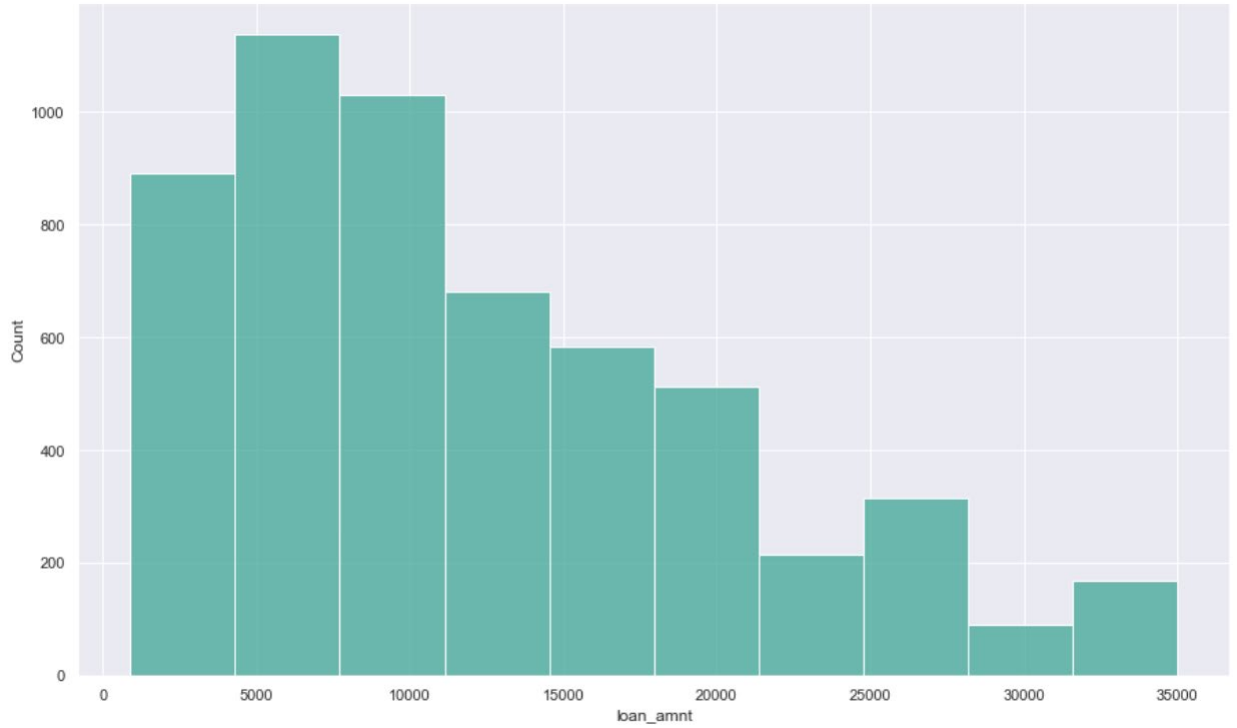
Above count plot is plotted for all type of loan status. From above count plot we can not derive any conclusive decisions. But here we get to know person/business which has 10+ years of employement length are more than any other lengths.

Employment length analysis of “Charged of” loan type



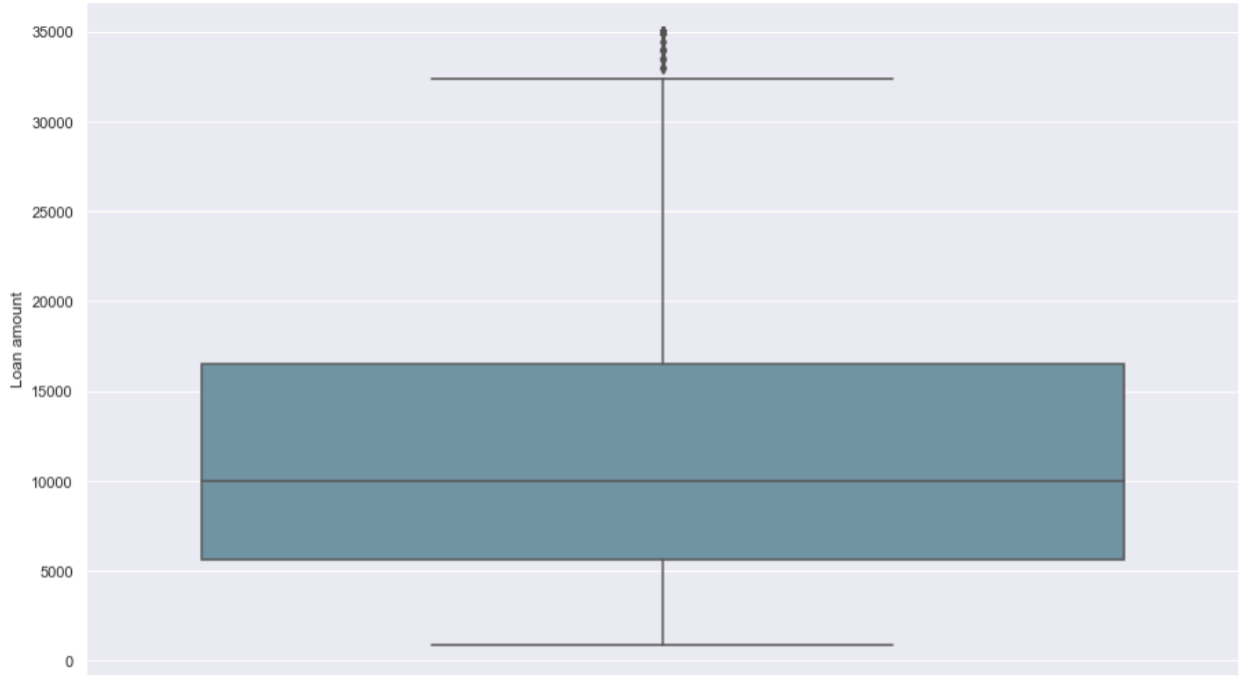
Above count plot is plotted for "Charged of" loan status. From above count plot we get to know most of the defaulters have 10+ years of employement length.

Loan amount analysis of “Charged of” loan type



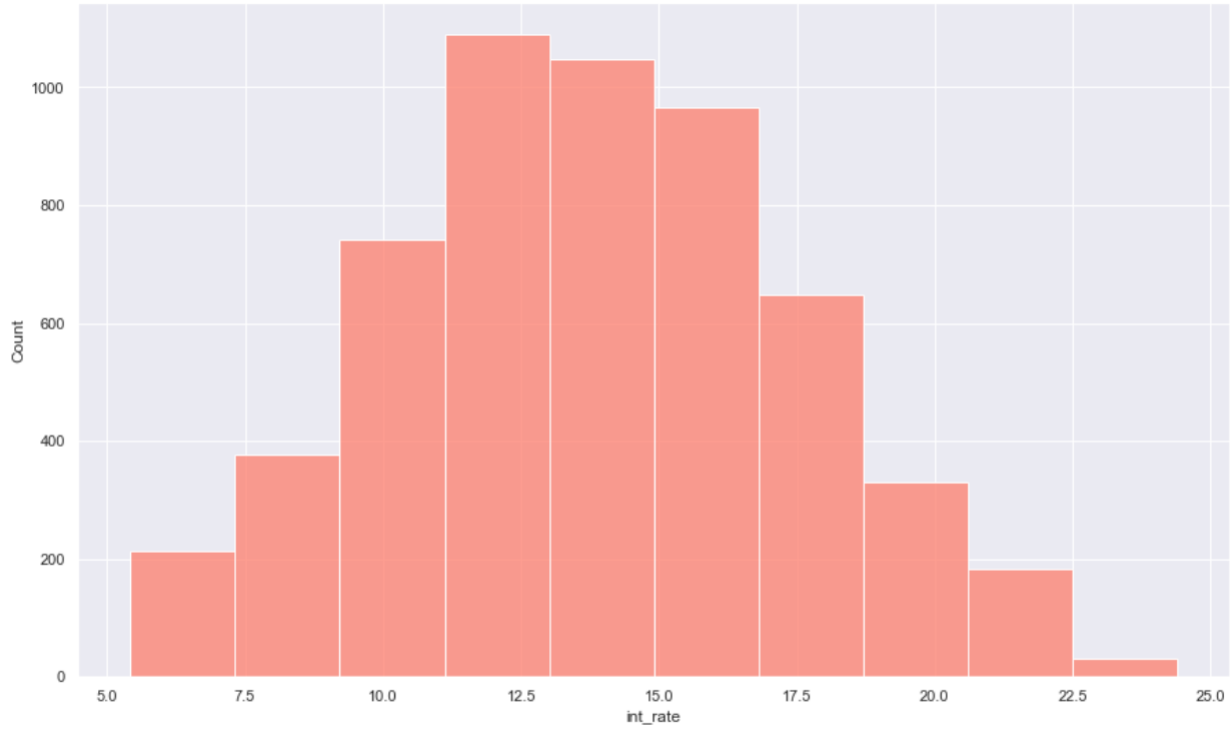
Above hist plot is plotted for "Charged of" loan status. From above hist plot we get to know most of the defaulters loan amount ranges from 900 to 15000

Loan amount analysis of “Charged of” loan type



Above box gives us little clarity on the loan amount spread out. Q1 as per above plot is 5600, median is 10000 and Q3 is 16500. Most defaulters lie between 5600 to 10000 loan amount

Interest rate analysis of “Charged of” loan type



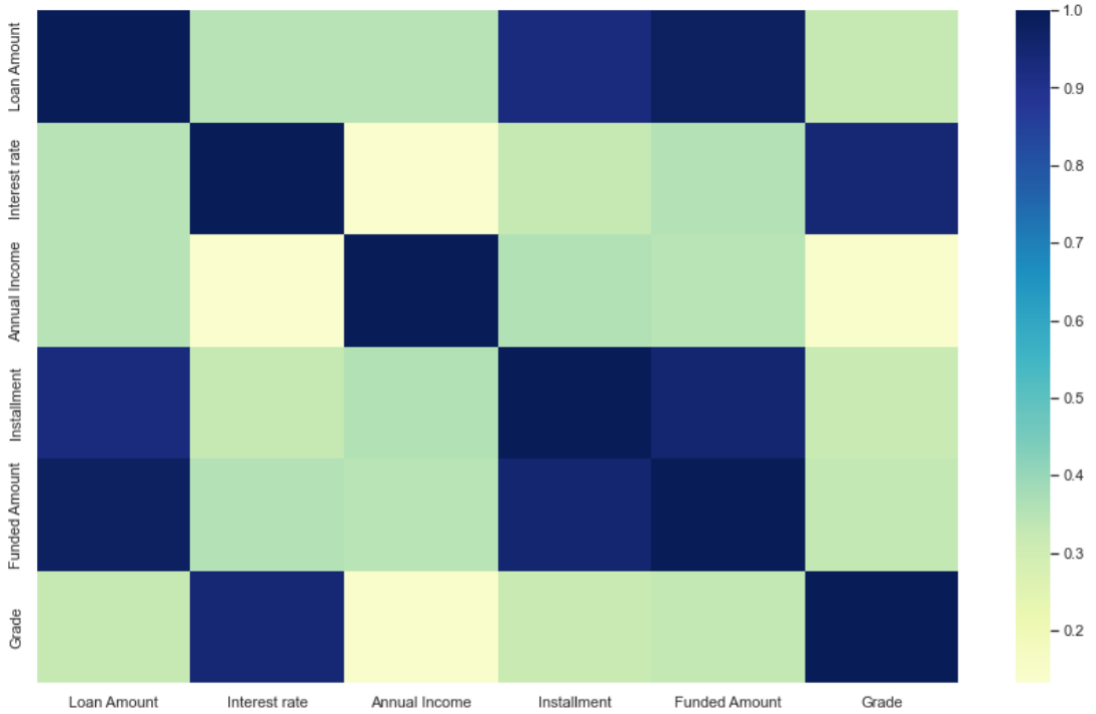
Above hist gives us little clarity that most defaulters lie between interest rate from 10% to 17.5%. So higher the interest rate more the chance of defaulting.

Bivariate Analysis

Loan amount and Interest rate analysis of “Charged of” loan type

Above scatter plot is for loan amount(y axis) and interest rate(x axis) for defaulters. From it we get to know that a loan defaulters cluster is getting created for loan amount between 2500 to 10000 and where interest range lies between 10.0% to 15.0%.

Loan Amount, Interest rate, Annual Income, Installment, Funded Amount, Grade analysis of “Charged of” loan type



Above heatmap denotes correlation between loan amount Loan amount, interest rate, annual income, installment and funded amount. Darker the block higher correlation. From above heat map we get to know below attribute have more correlation in loan defaulters:

- loan amount and installment
- loan amount and funded amount
- installment and funded amount
- interest rate and grade

Conclusion

- Chances of defaulters is more if interest rate of loan is between 10% to 17.5%
- Most of the defaulters are the person who has more than 10 years of employment length.
- Median of the defaulters is 10000
- Loan defaulters cluster is getting created for loan amount between 2500 to 10000 and where interest range lies between 10.0% to 15.0%.