

Lending Club Case Study

Nikita Sakore

Sanket Bhawe

PGDM in ML & AL : MLC33

About Project.

Its a case study of consumer finance company which specializes in lending various types of loans to customers. When any customers applies loan to them company must decide that where we should approve the loan or reject the loan and this approval and rejection is on the basis on some criteria and all these criteria says person likely repay the loan or not.

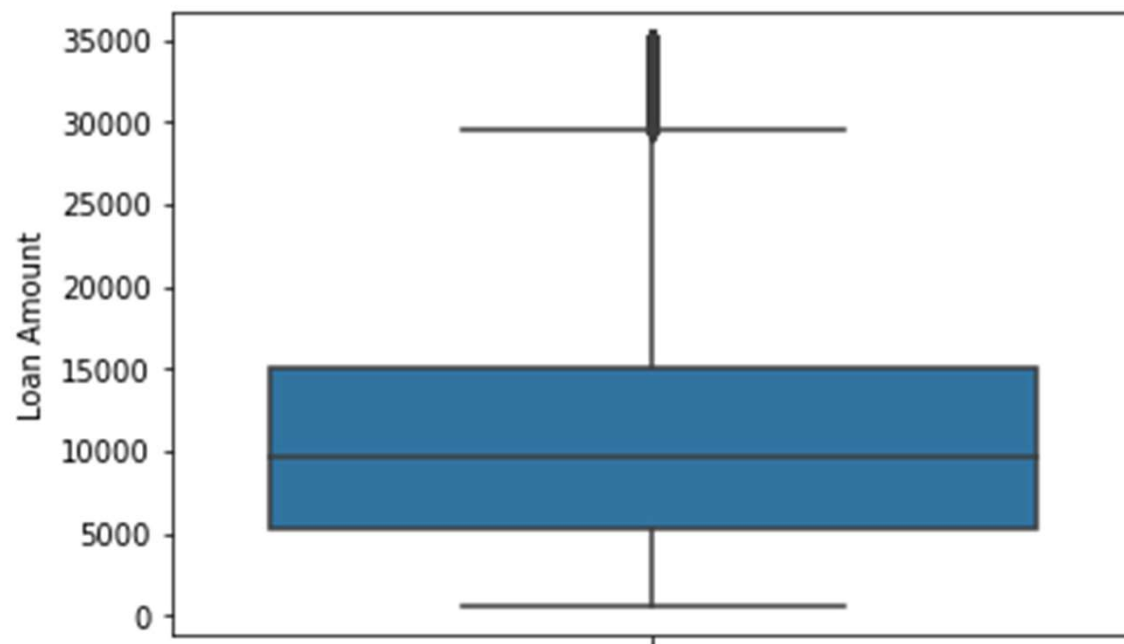
General Info

- The data given here contains the 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.
- Dataset that is being used contains the complete loan data for all loans issued through the time period 2007 to 2011.

Data Cleaning

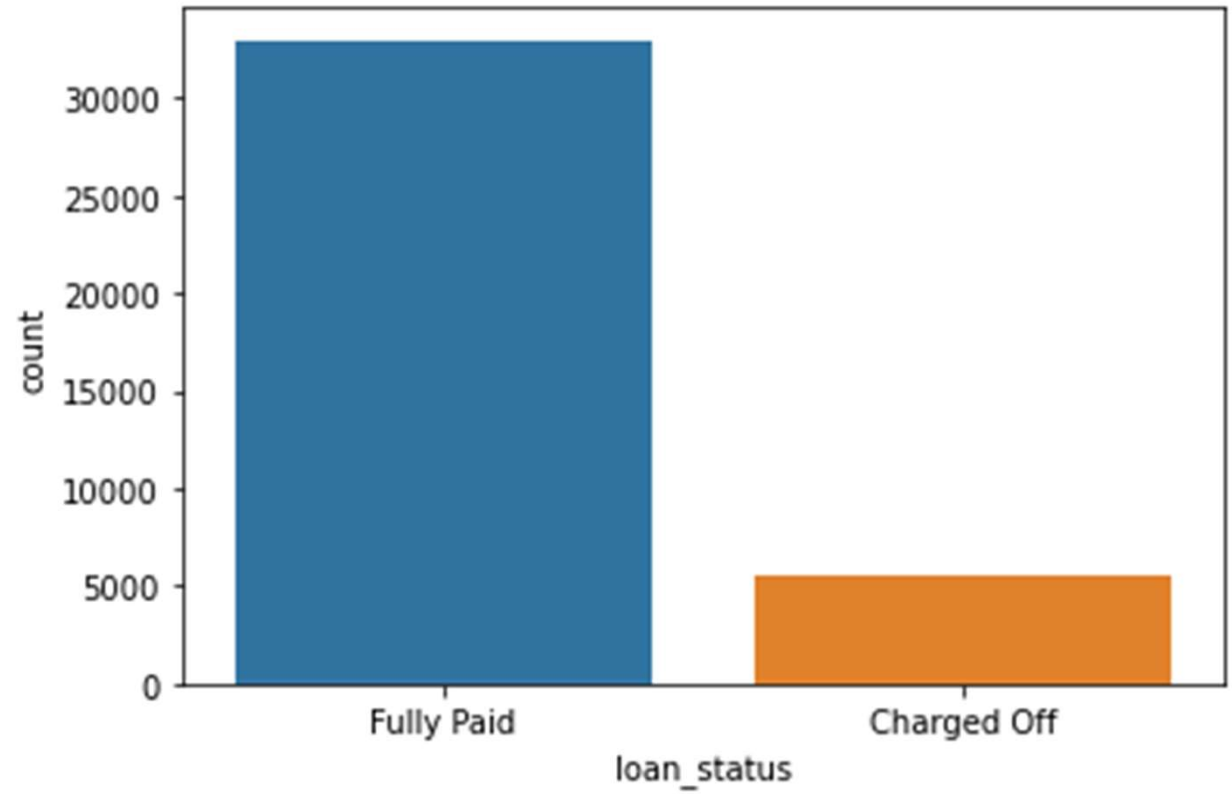
- **Missing Data Columns :** There are 111 columns(variables). After doing some analysis there are 54 columns have 100% missing or Null values which are useless for analysis. So, its better to delete them.
- **Single Valued Columns :** There are some columns that have only one unique value that means all rows have same value for those column.
- **No use columns :** There are some columns that are no use for our analysis like desc, title etc.
- **Add new columns :** We add new columns which are derived by already existed columns.

Univariate Analysis

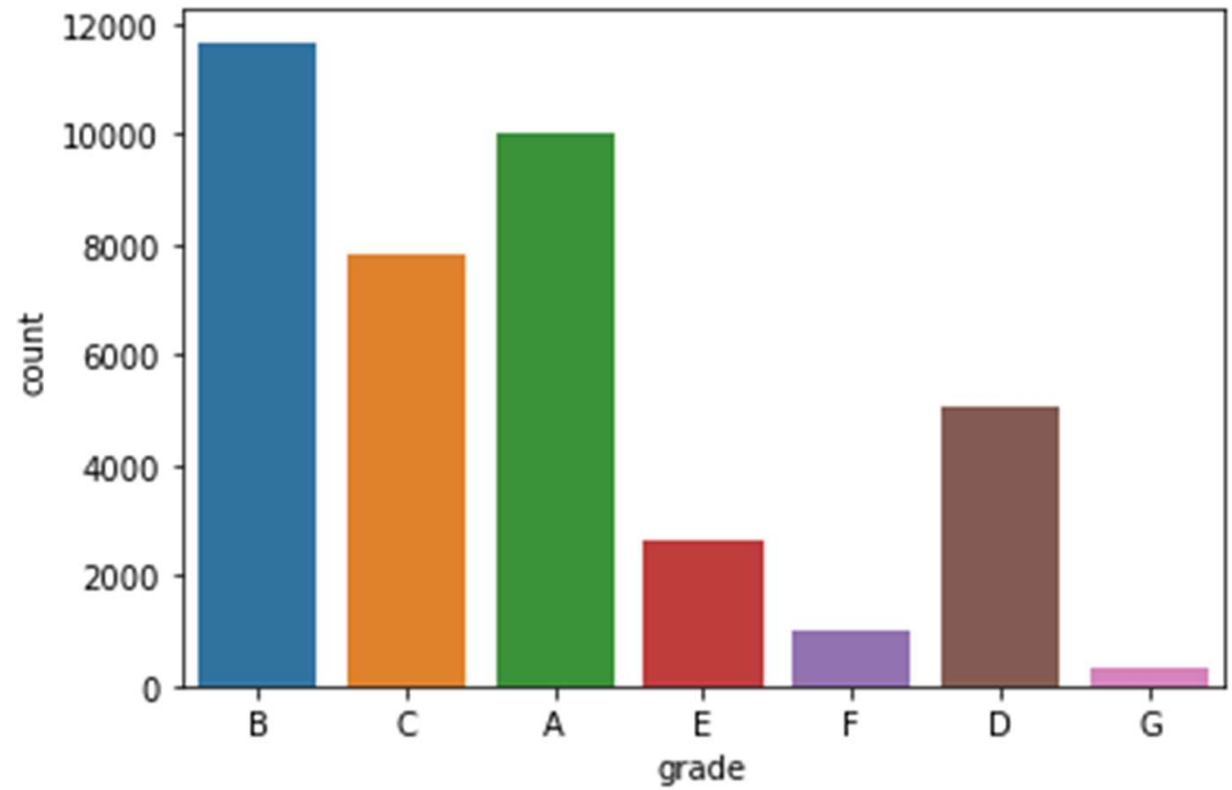


- Boxplot shows : Median Val = 10000 Range 0 - 35,000

Count Plot for Loan Status

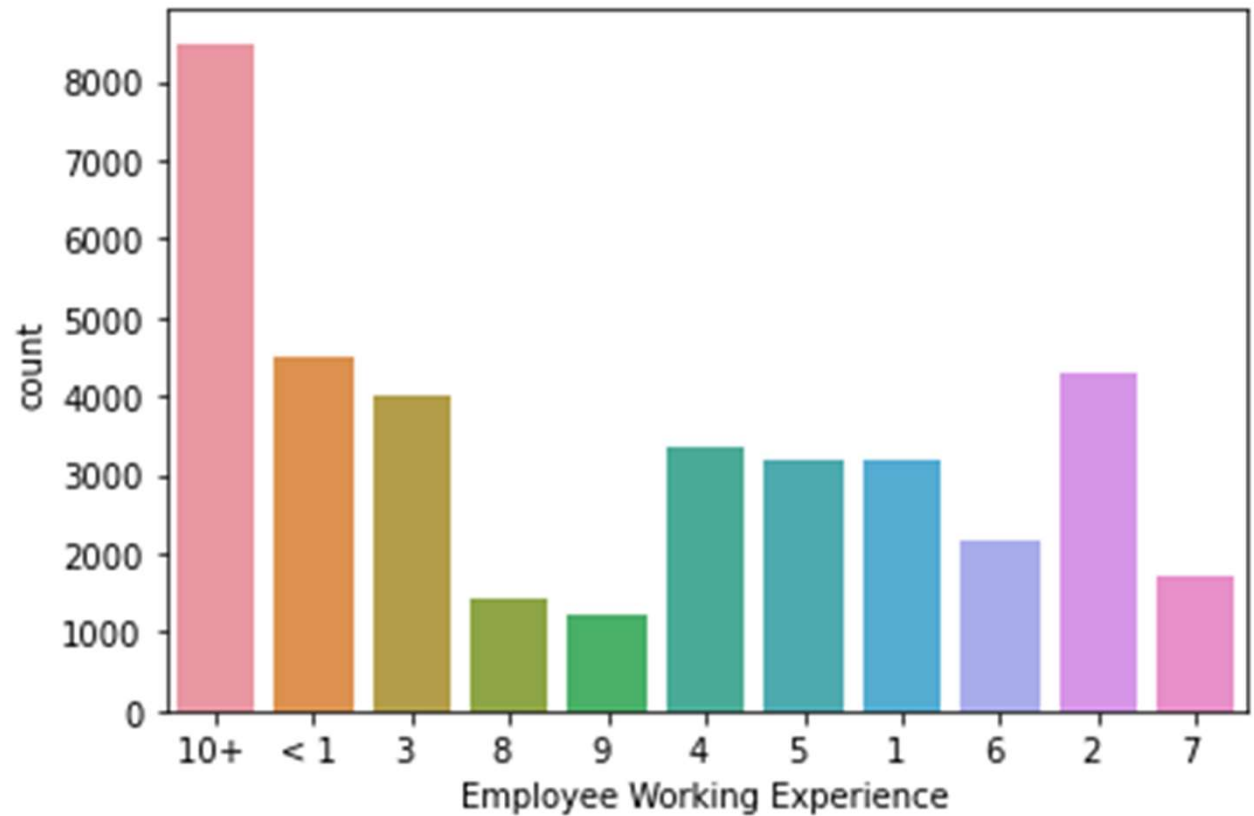


Count Plot for Grade



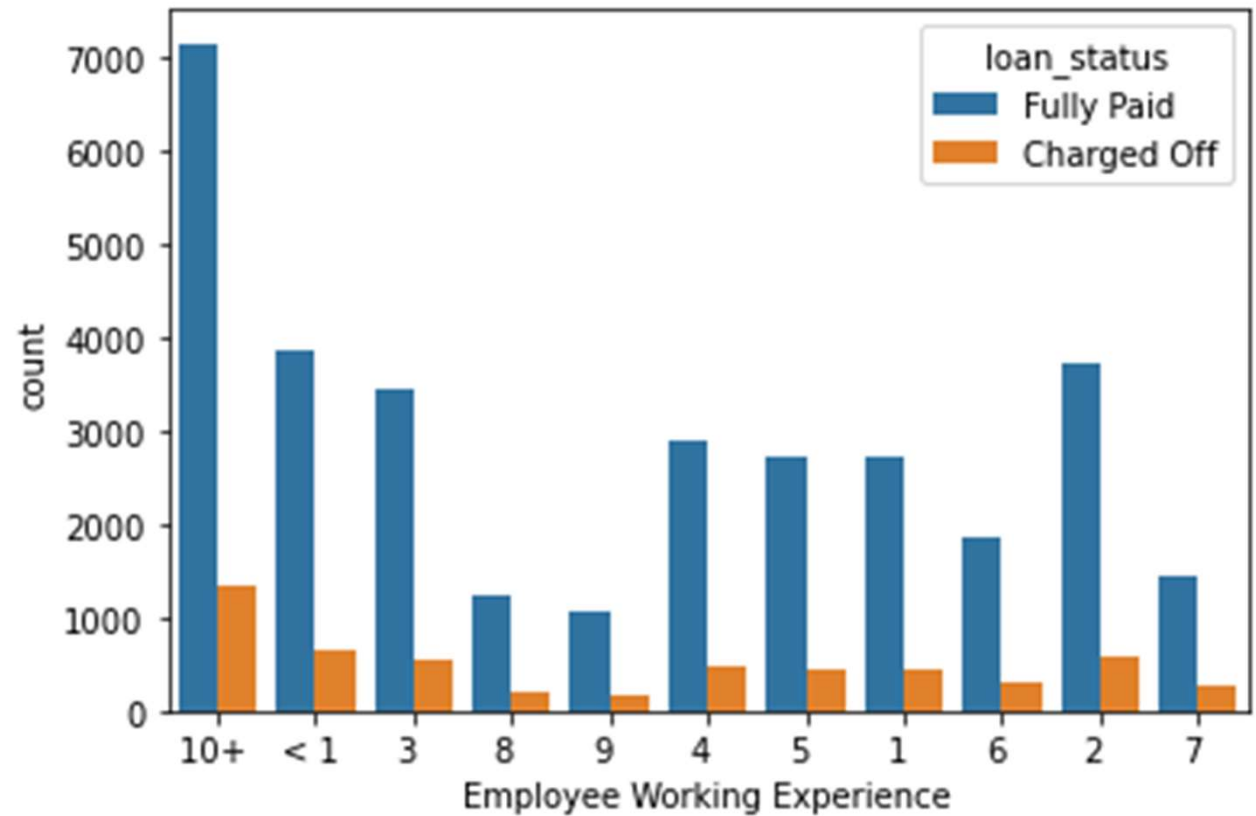
Count Plot for Employee Working Experience

- Employee work experience having 10+ years are most likely to apply for the loan

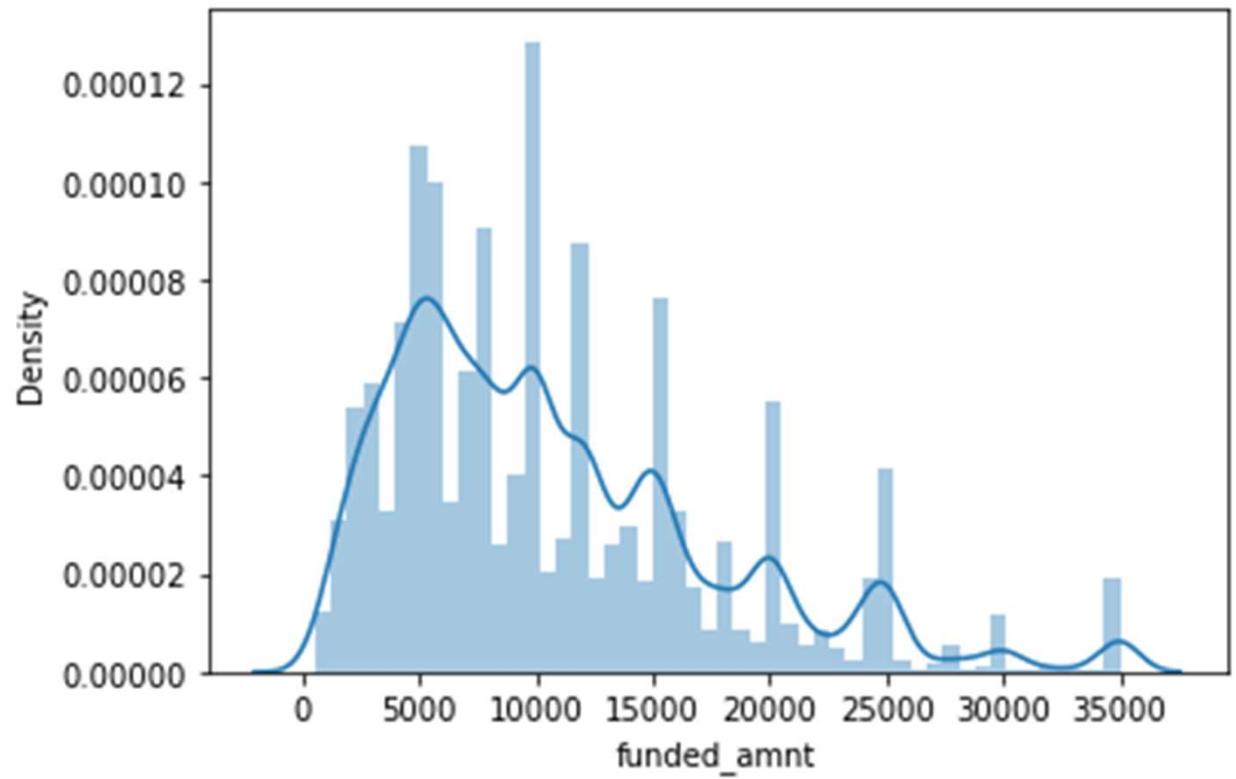


Count Plot for Employee Working Experience with loan status

- Employee work experience having 10+ years are most likely to apply for the loan for both fully paid and charged off.

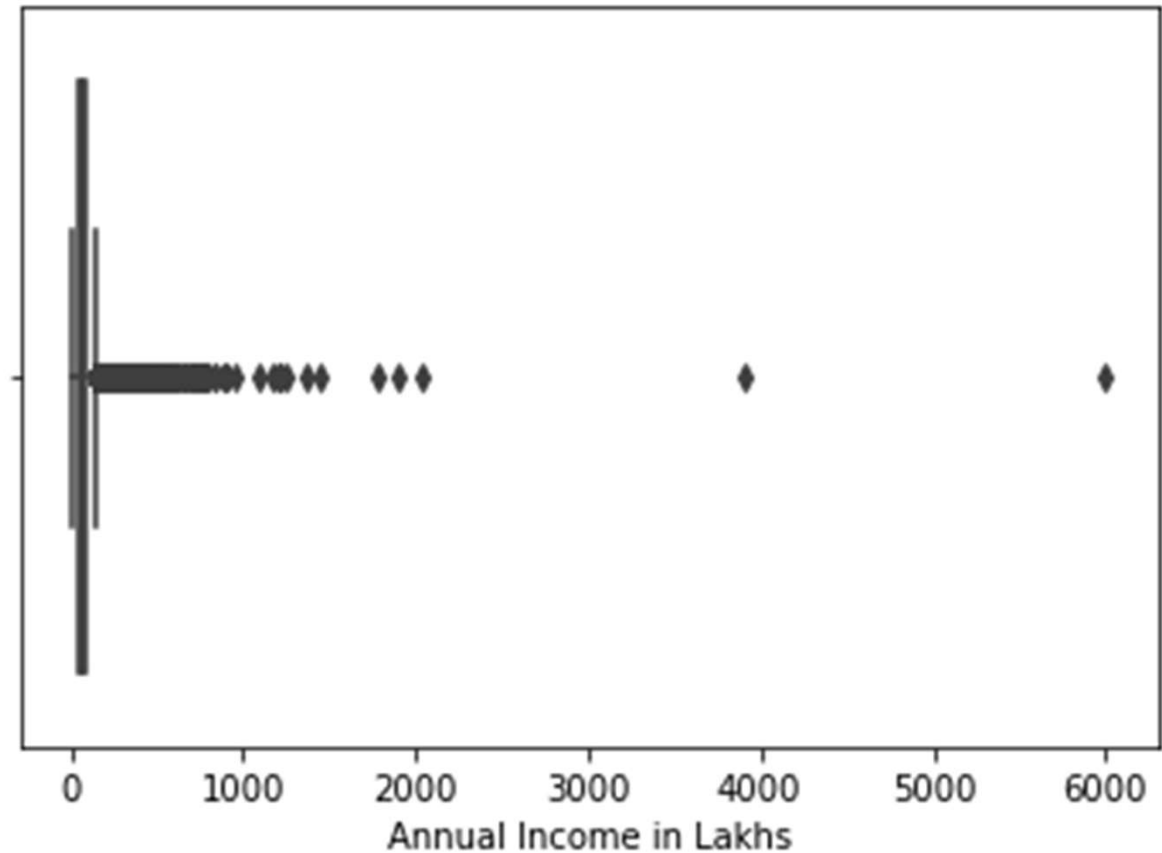


Distribution Plot for Funded Amount



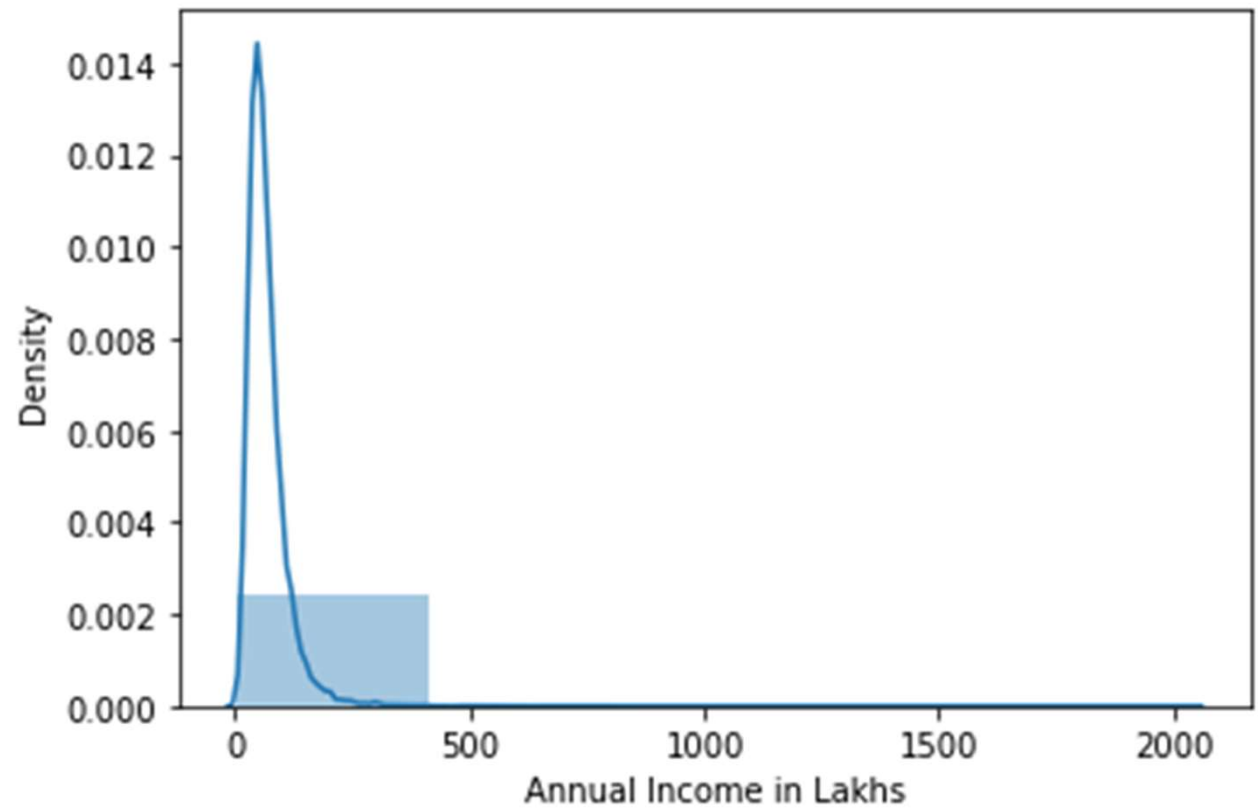
Box Plot for Annual Income in Lakhs

- Dataset has only 2 employee whose income is above 30 Lakh.

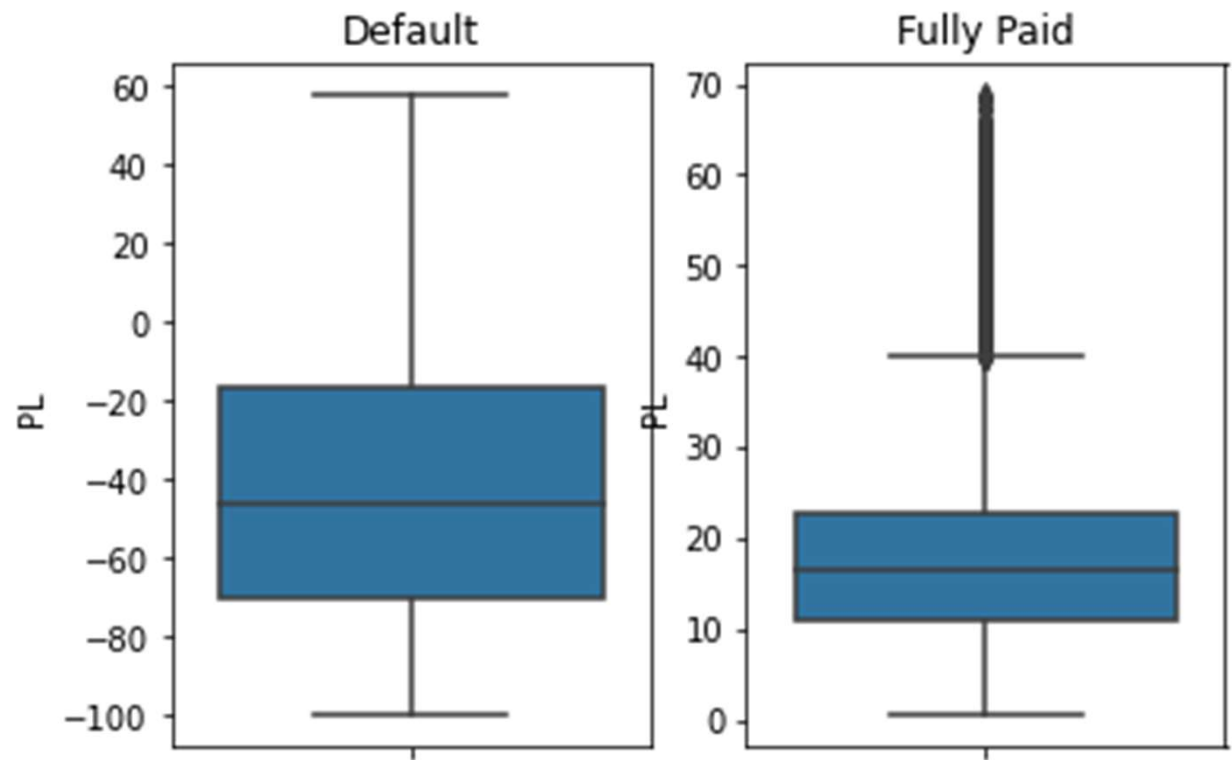


Distribution Plot for Annual Income in Lakhs

- Most employees apply for the loan if income is below 5L.

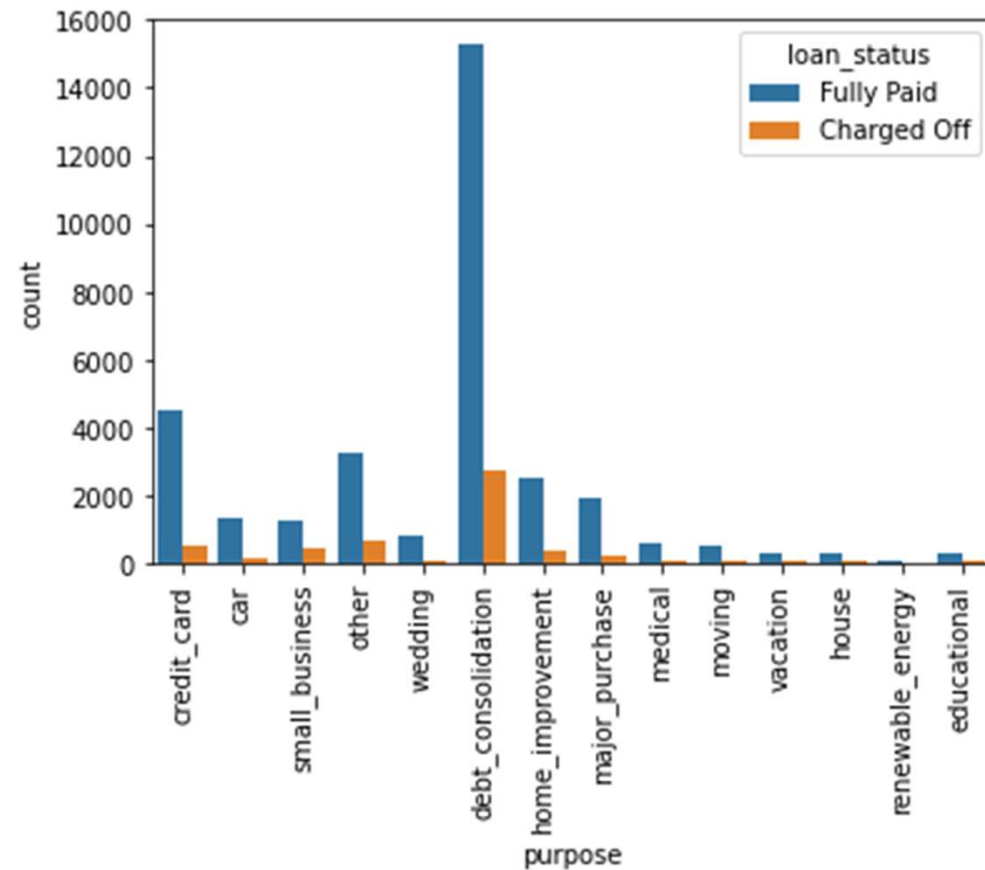


Box plot
Relationship
of Funded
Amount with
defaults



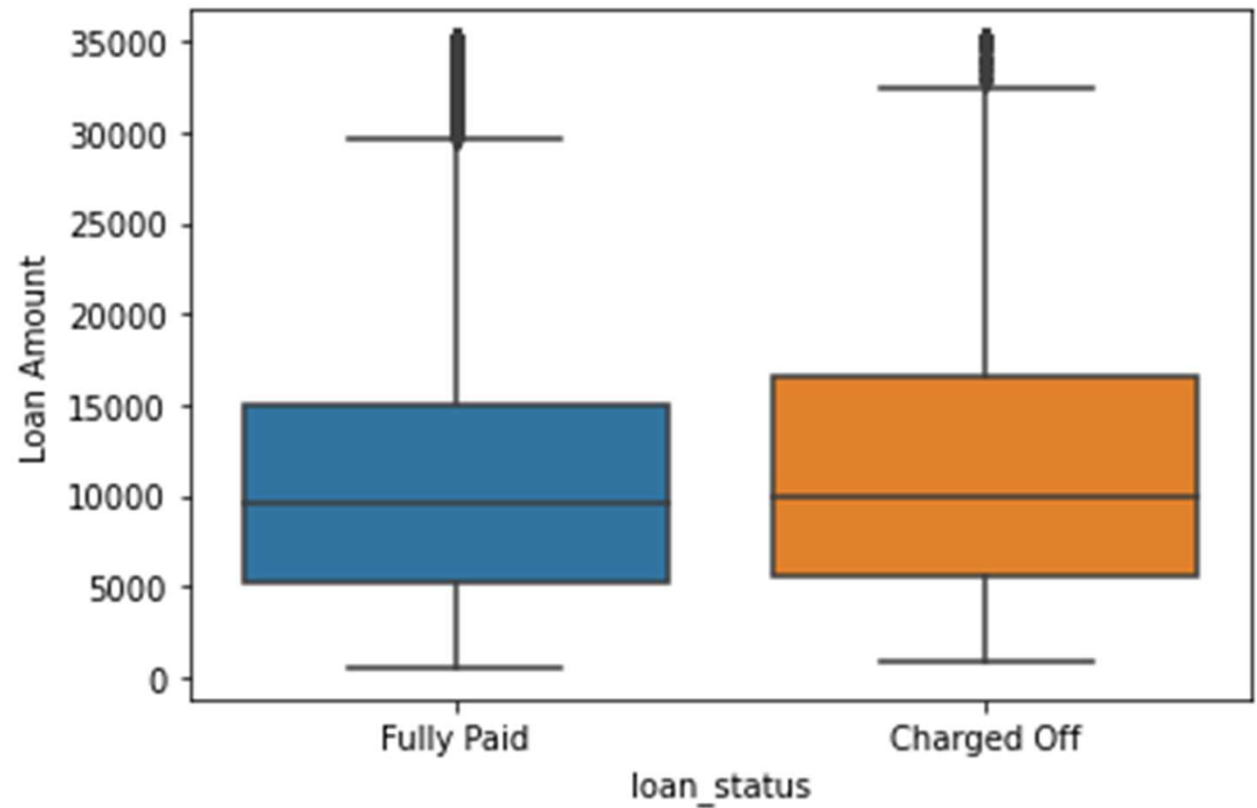
Count Plot for Purpose with loan status

- The debt consolidation loan are mostly fully paid.

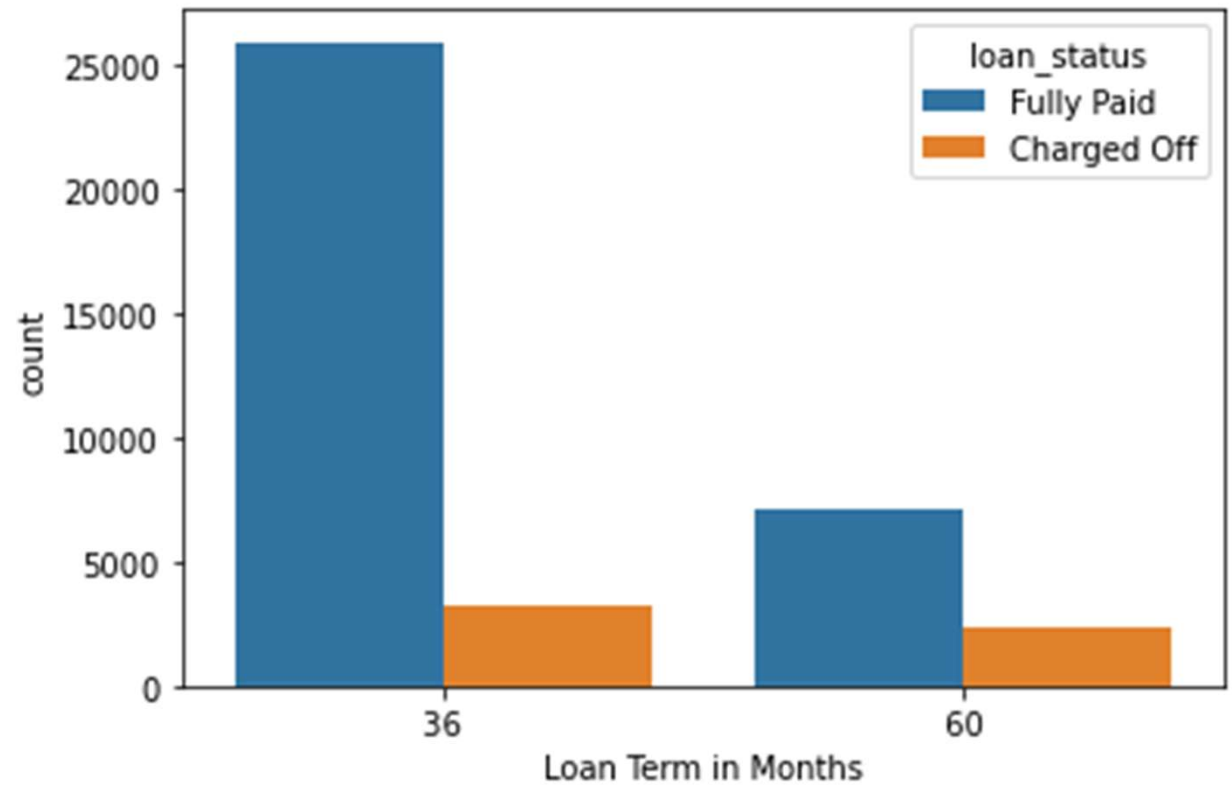


Box plot for
loan Status with
loan amount.

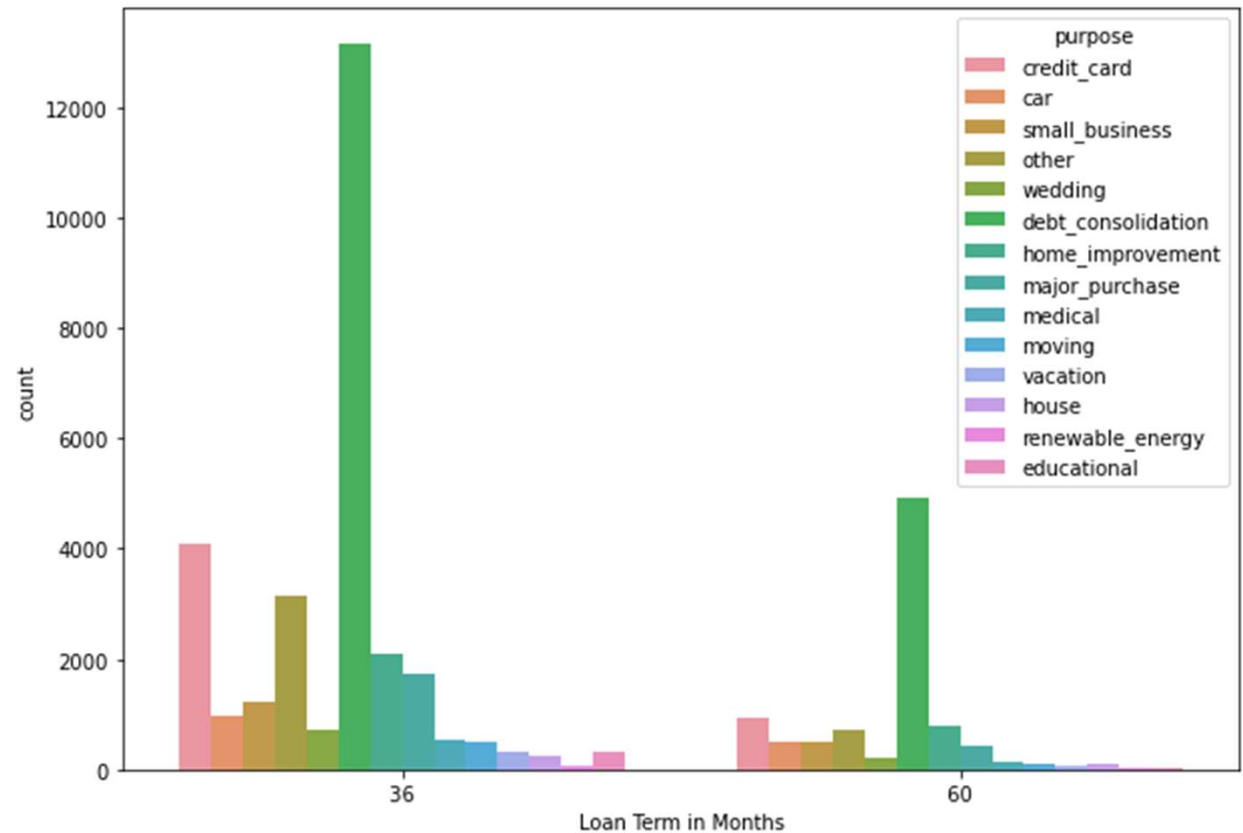
- Mean, 25% and 75% Loan amount of Fully paid and charged off is exactly same.



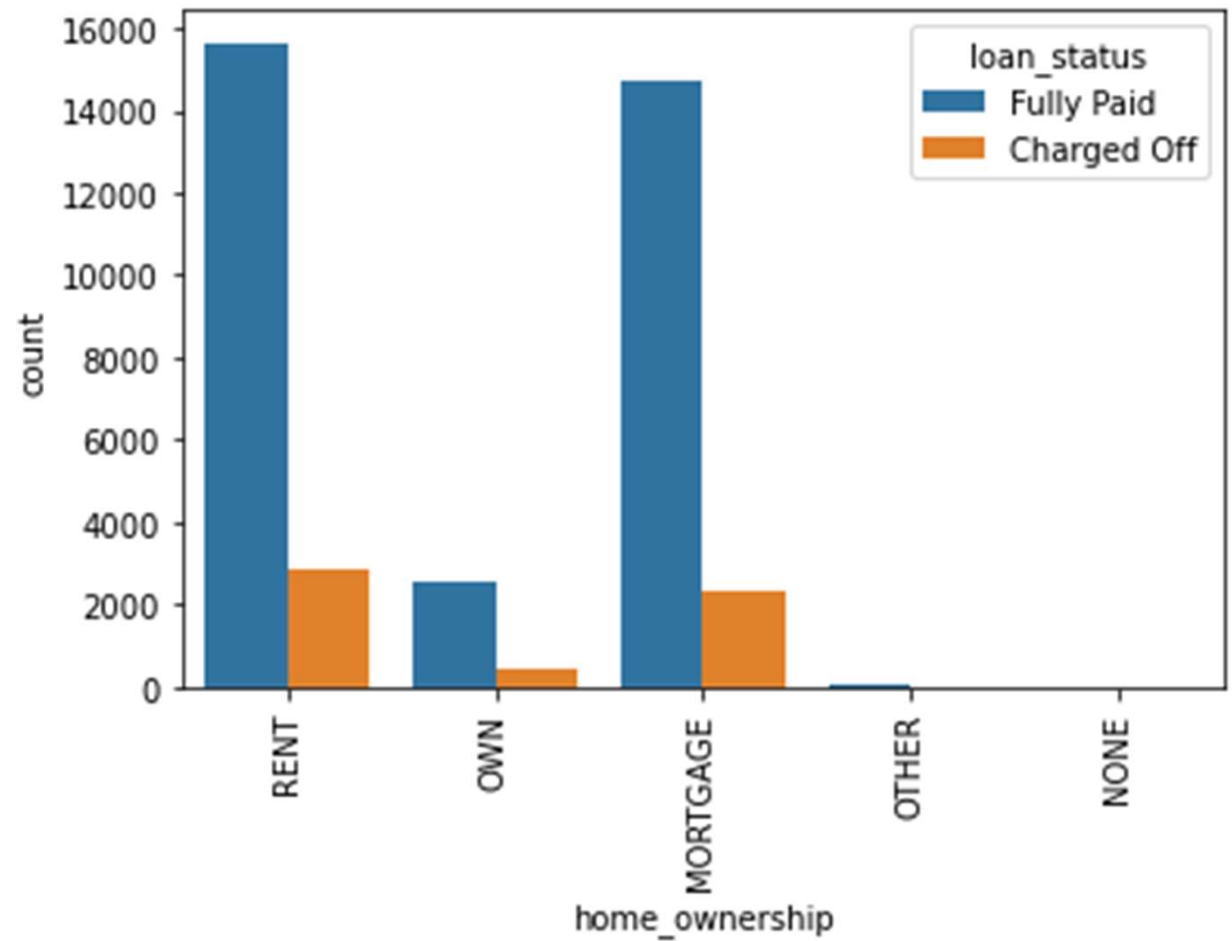
Count plot
term month
with loan
status



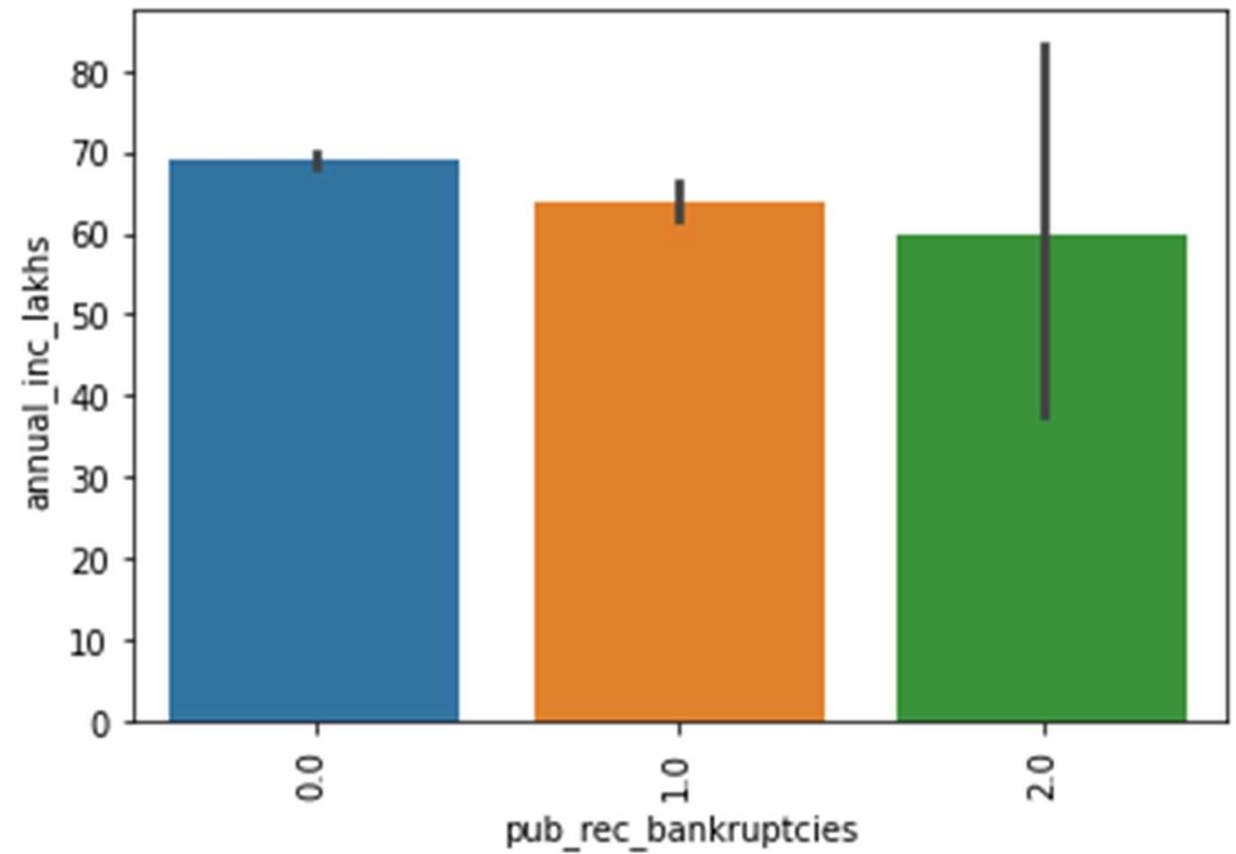
Count plot
term month
with purpose



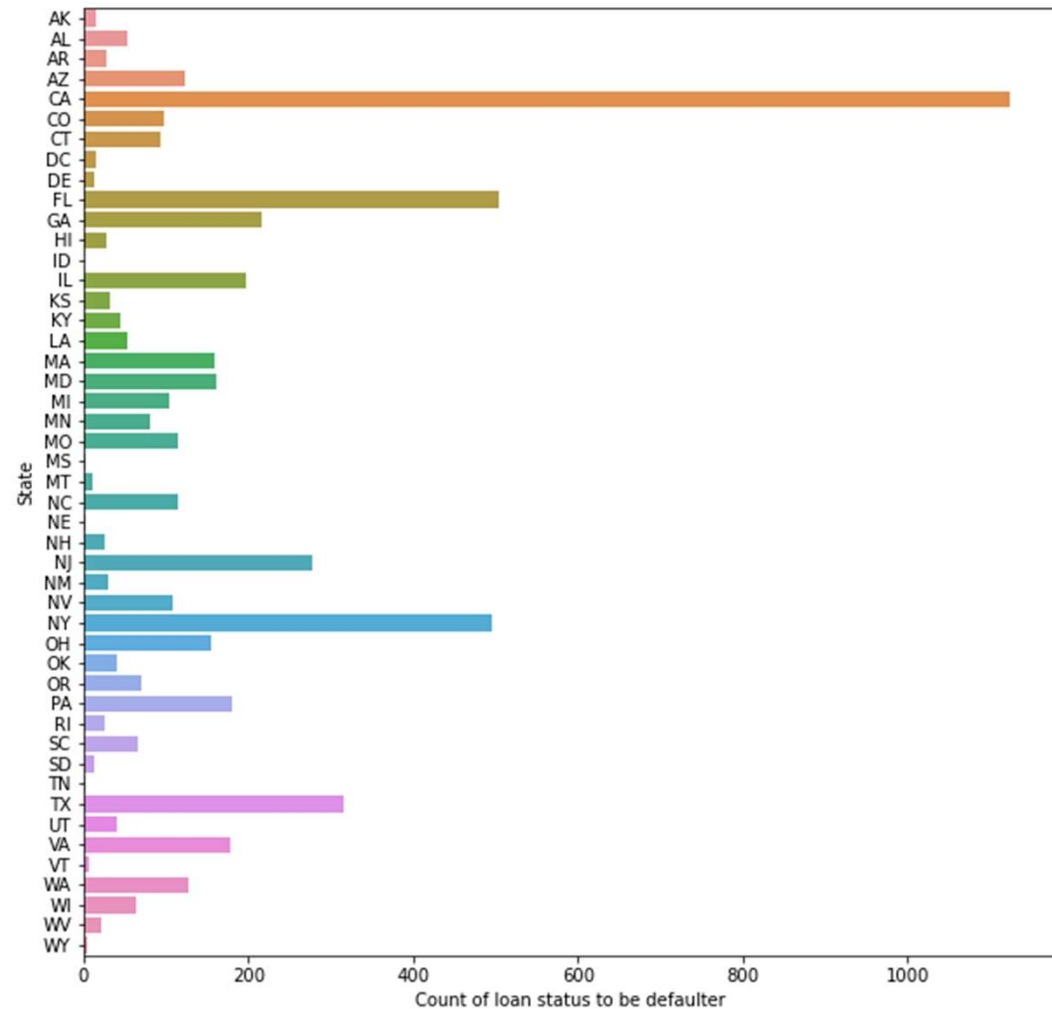
Count plot
home
ownership
with loan
status



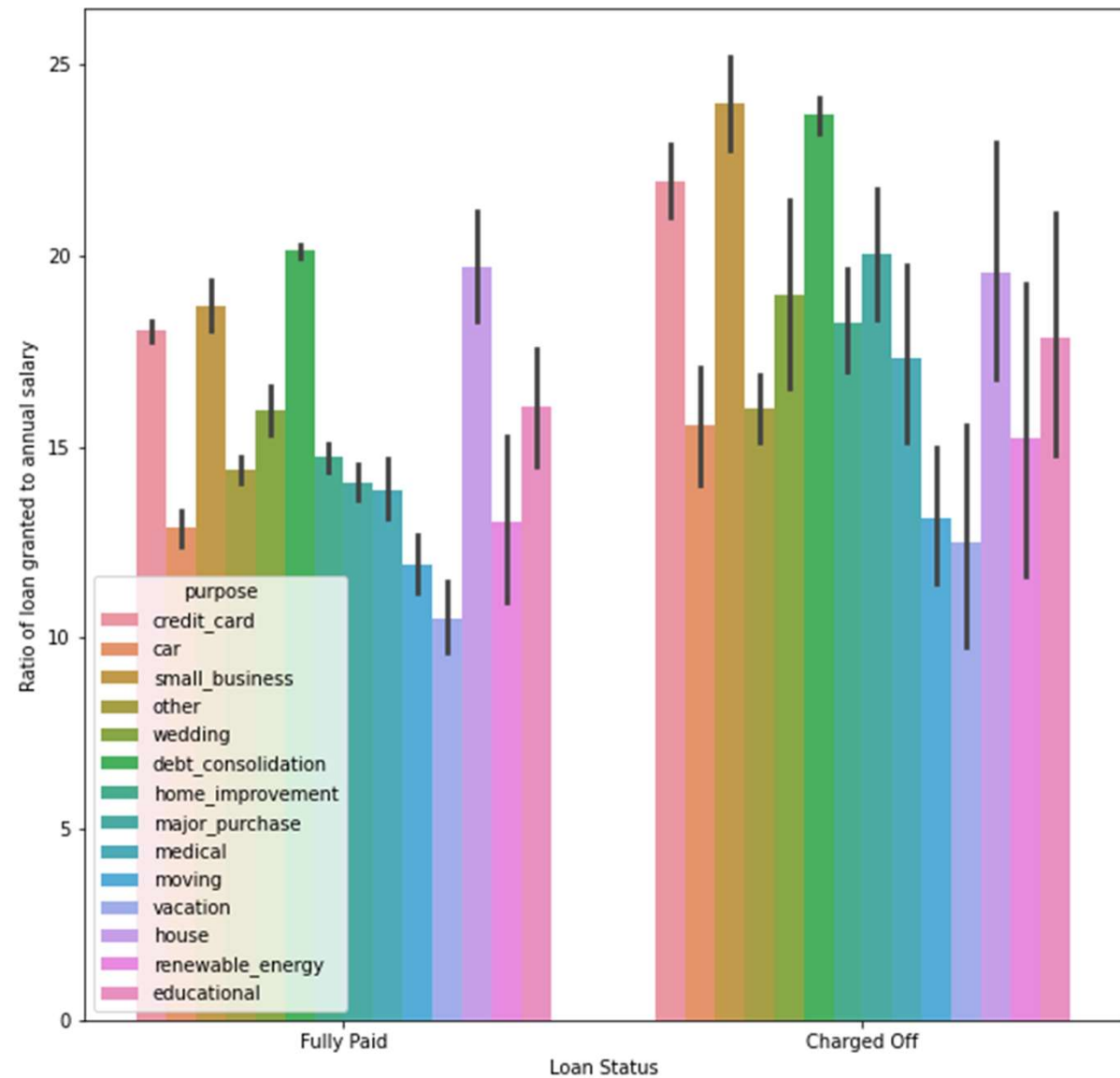
Bar plot
pub_rec_bankruptcies
vs
annual_inc_lakhs



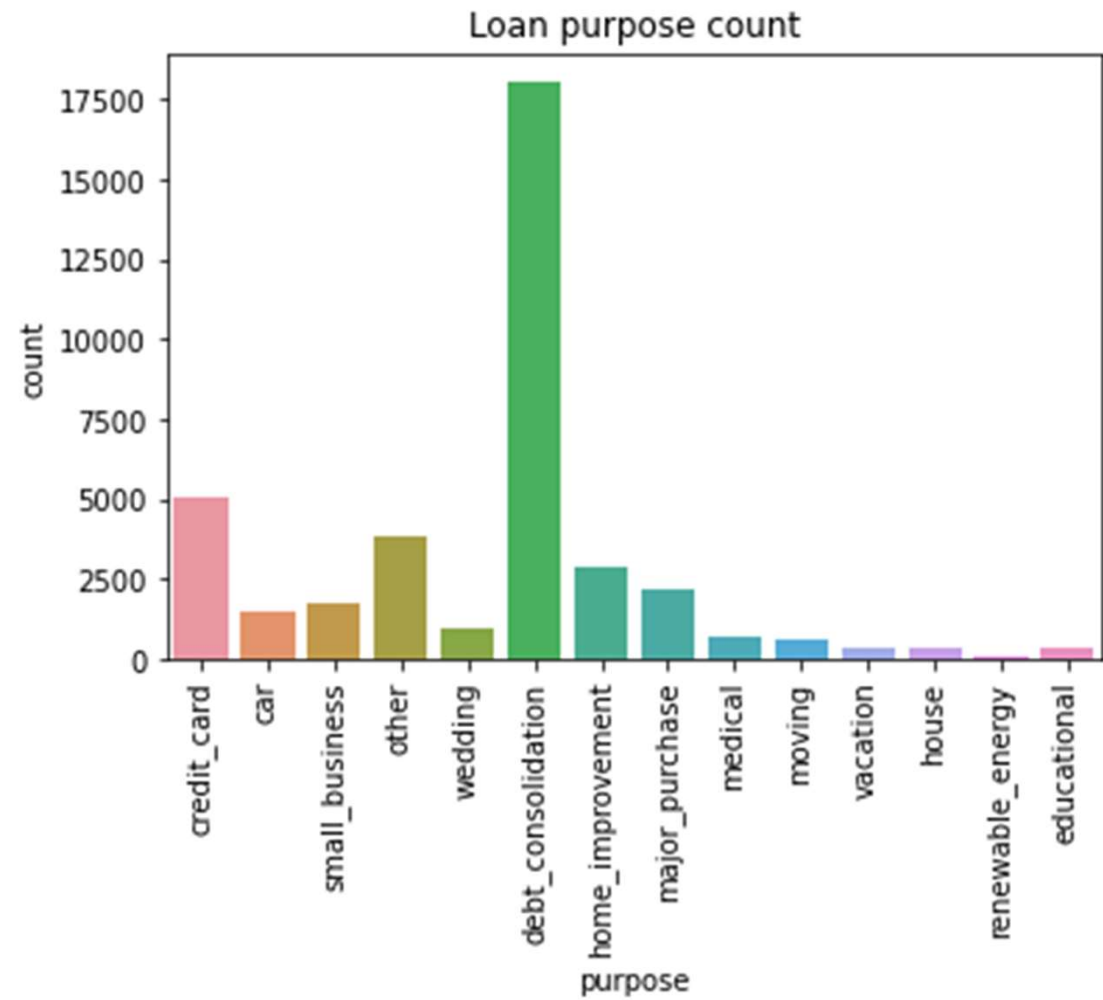
Bar plot Count
of loan status
to be default
vs
states



Bar plot
loan status
vs
Ratio of loan
granted to
annual salary



Loan Purpose Count



Conclusion

- Highest Rate Of Interest has more defaulters.
- Low grade have high tendency to default.