

LendingClub – Case Study

Problem Statement:

Lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed.

Objective is to use EDA and understand the driving factors (or driver variables) behind loan default, i.e., the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Data Cleaning

As part of Data Cleaning, following listed columns were removed from data:

- Columns which are part of Customer Behaviour, these variables will not be present at time of loan application and will be collected later.
- Columns for which all values are Null/NA
- Columns which have only one value or which has values/text in similar patten which are of no use like URL

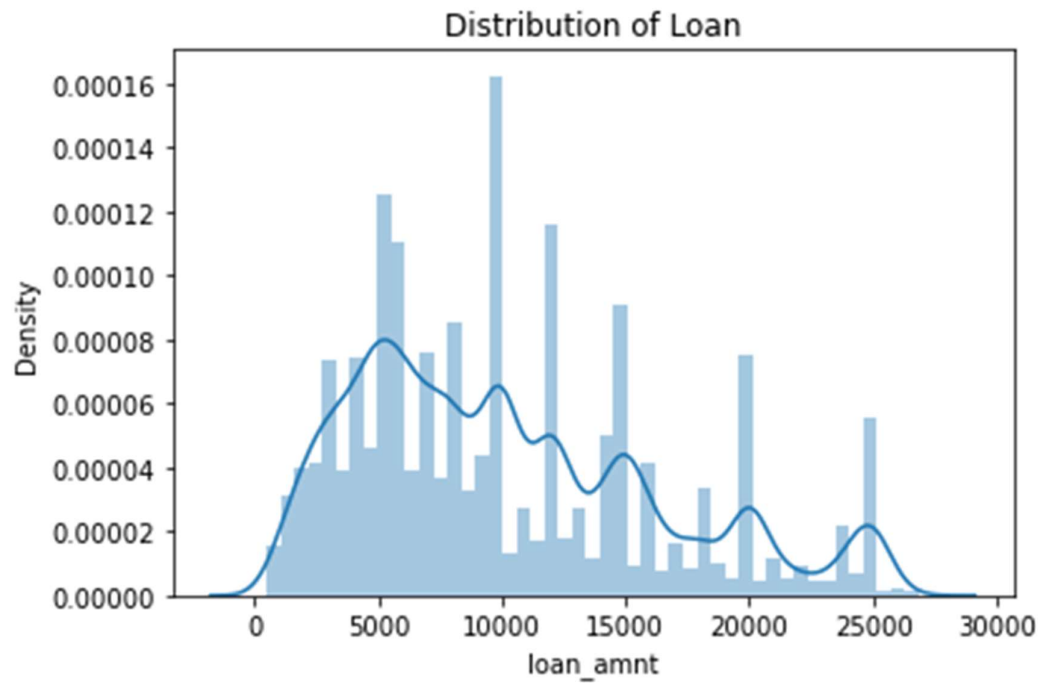
As next step in data cleaning, null/na values were handled, for which following approach was taken:

- For Categorical columns, "mode" value is used for fill empty values
- For Numerical columns, values are filled/removed depending on column type/values.

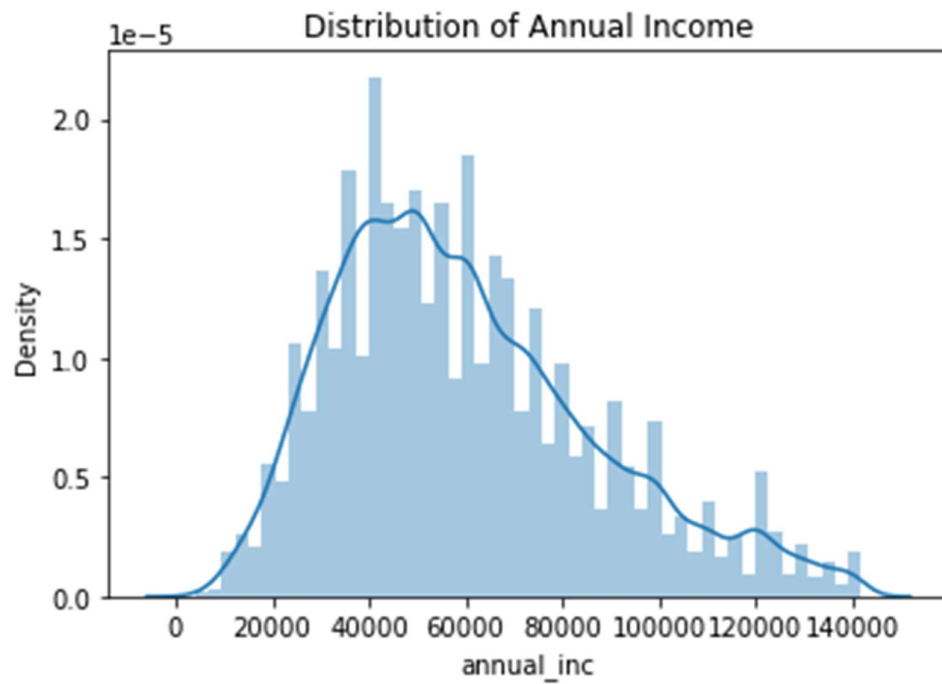
Uni-variant Analysis

Following steps were performed as part of univariant analysis:

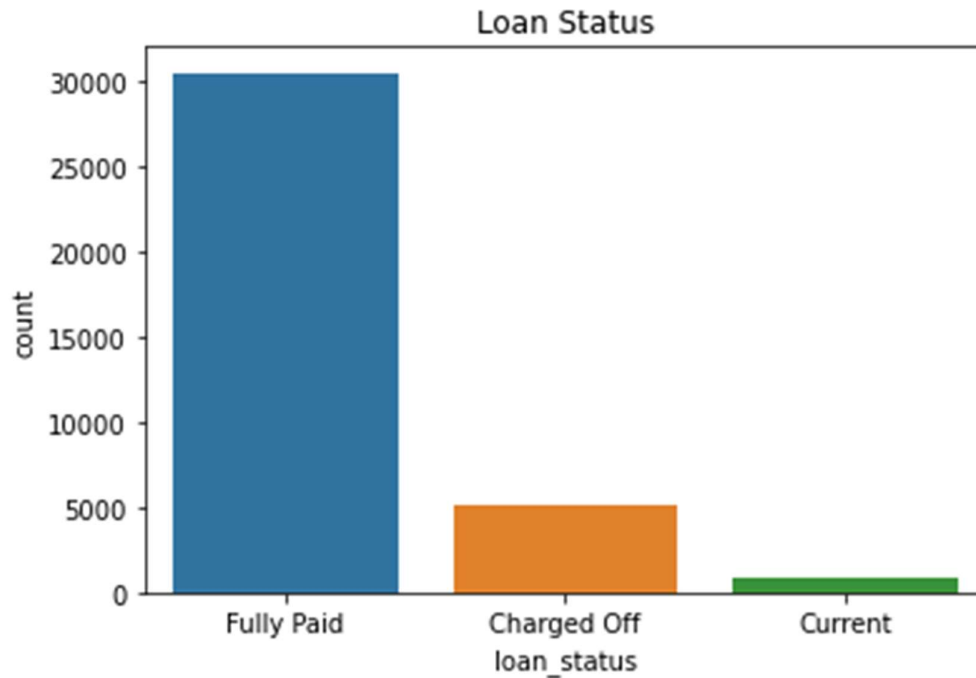
- Finding outliers in the columns and removing them:
 - From columns annual_inc etc outliers were removed.
- Create new columns like Year from existing data
 - New columns issue_month, issue_year were created from issue_d columns
- Applying binning/bucketing to create bins from existing columns:
 - Columns like annual_inc, int_rate were converted to bins for further analysis.
- Findings:



High density is observed in 20-60k range, most individuals have salary in 20-60k range.

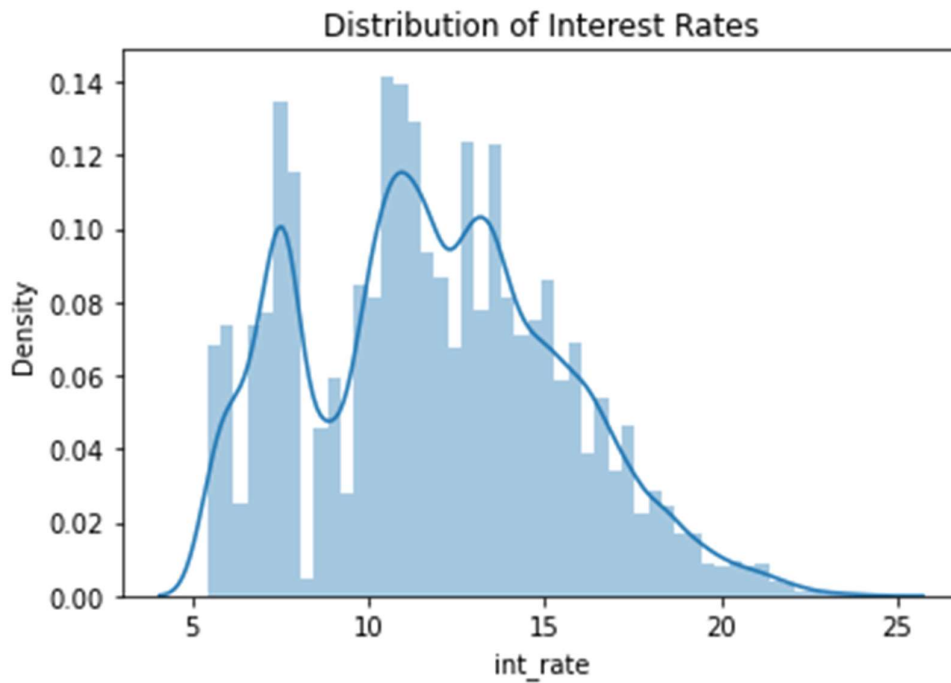


Higher density of loan is observed between 0-10k



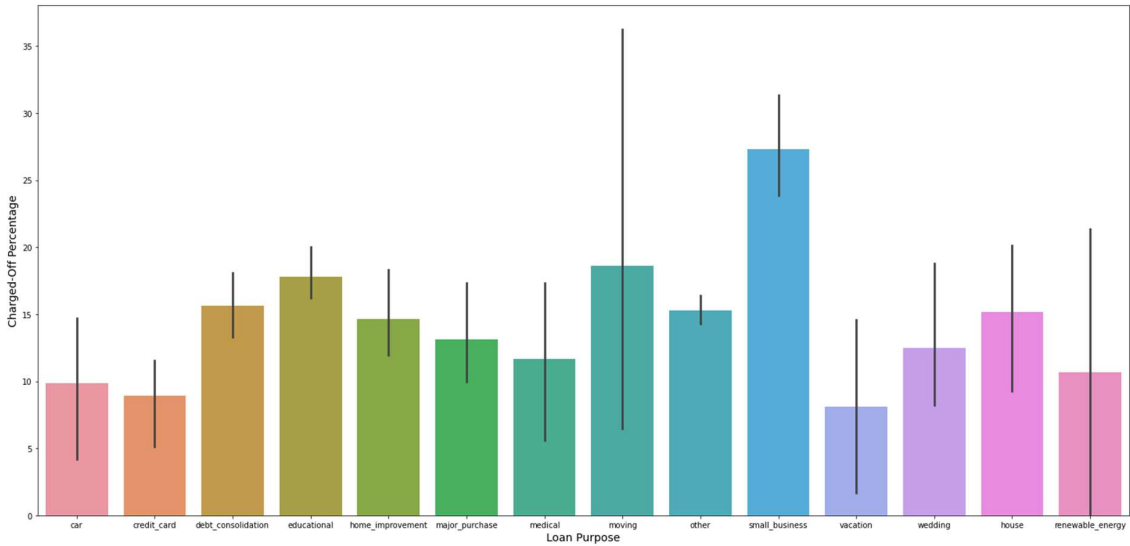
Fully Paid	30484
Charged Off	5177
Current	927

- Most of the loans are "Fully Paid"
- Around 14% of the loans are "charged-off"

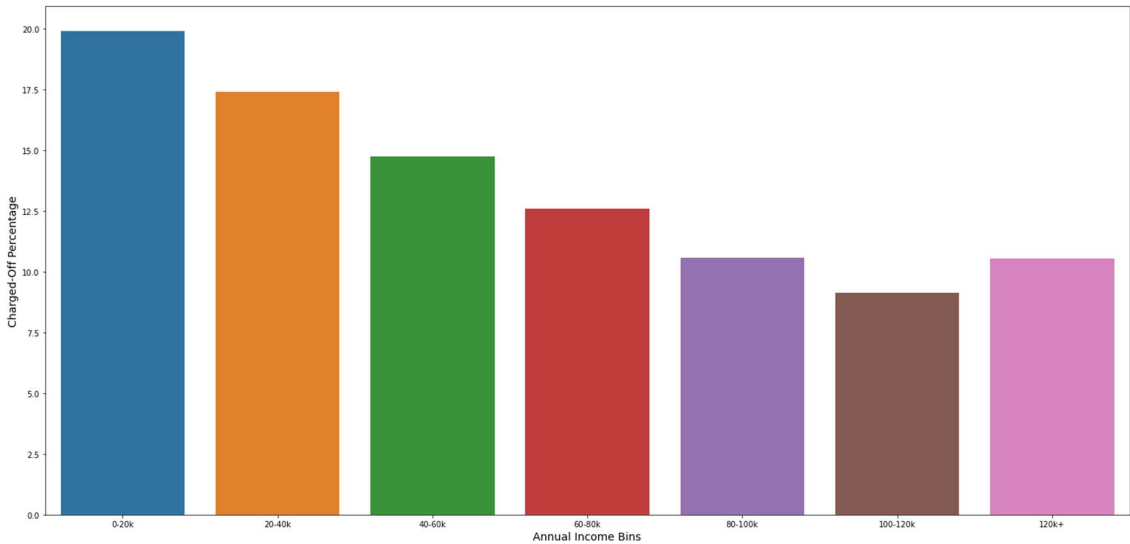


Most of the Interest rates are between 10-15%

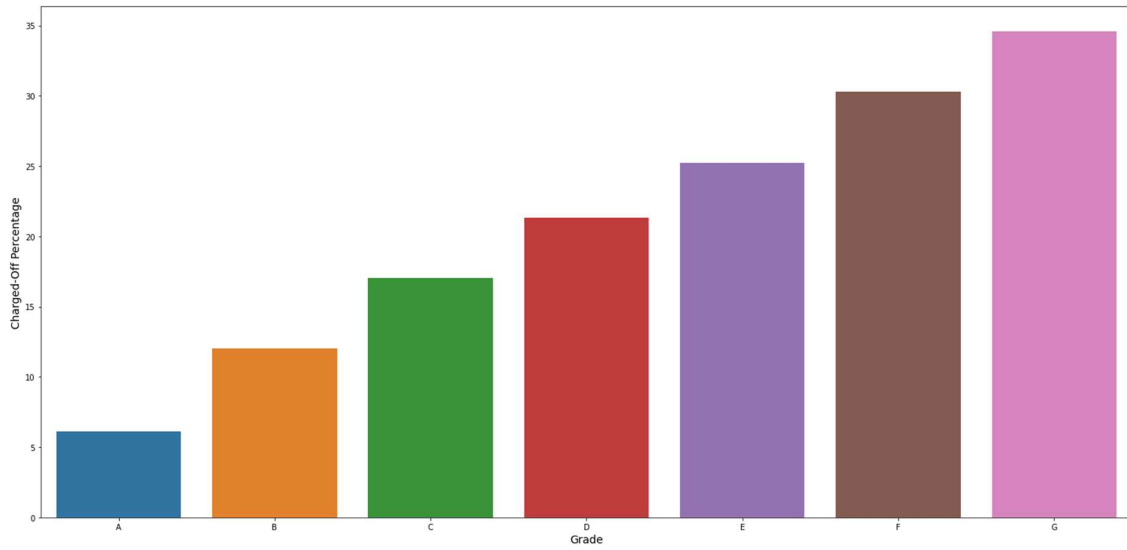
Bi-variant Analysis



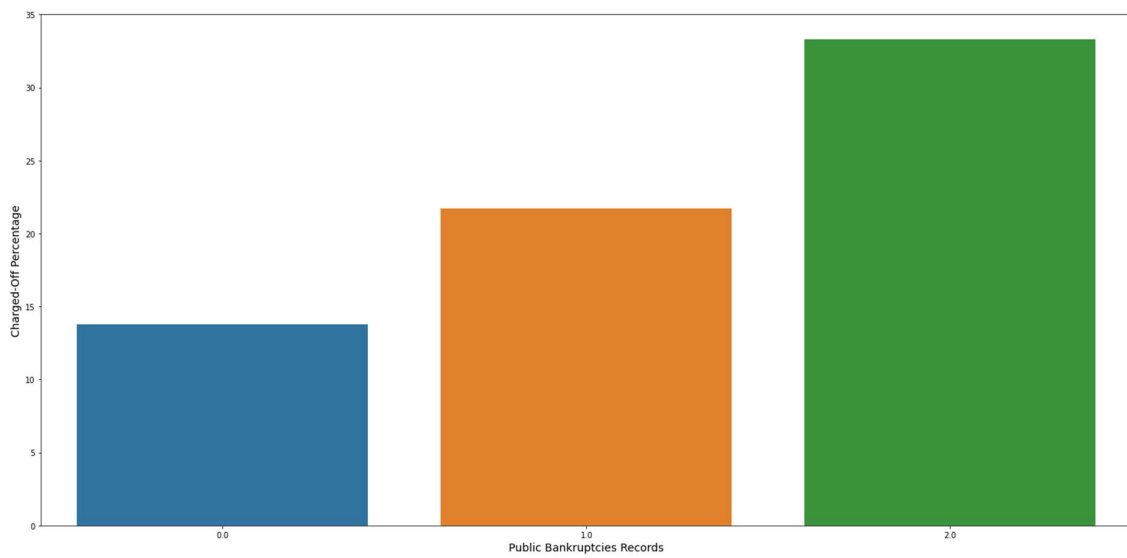
Loans taken for "small business" are the one for which "charge off" percentage is highest



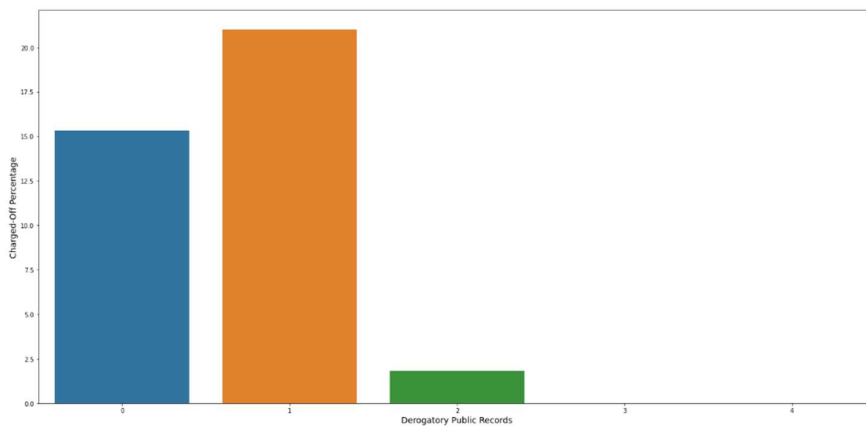
Less salary individuals (0-20k) are the one for which "Charged Off" percentage is highest.

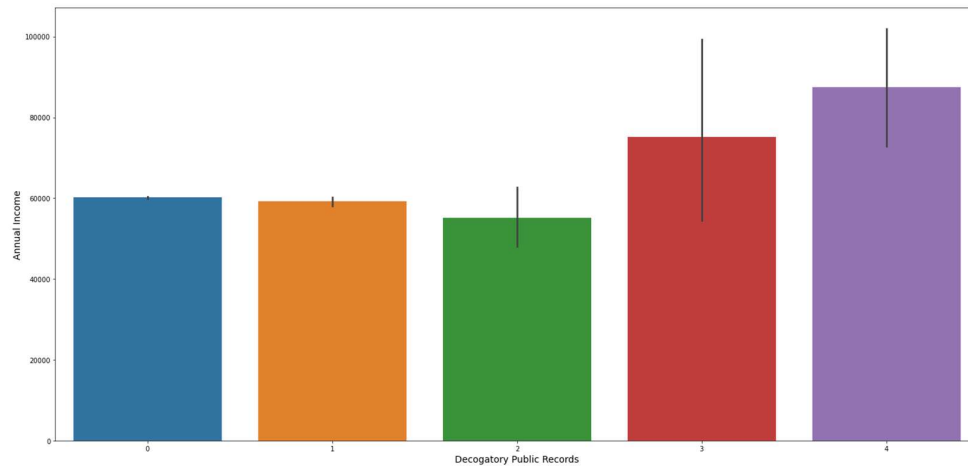


Individual with lesser grade tend to have higher "Charged Off" percentage.

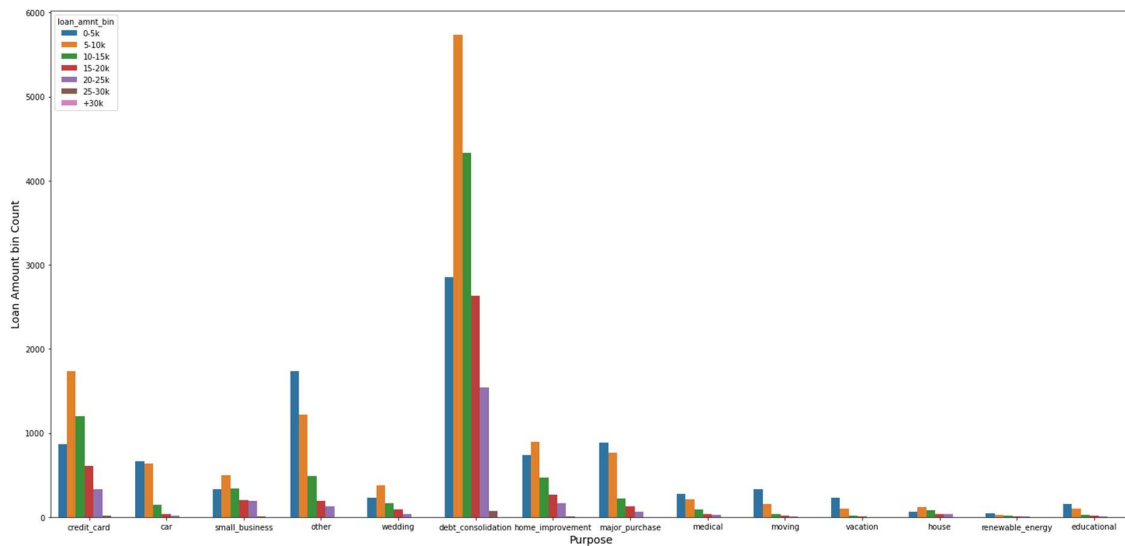


Individual with higher bankruptcies public records are having high charge-off percentage.

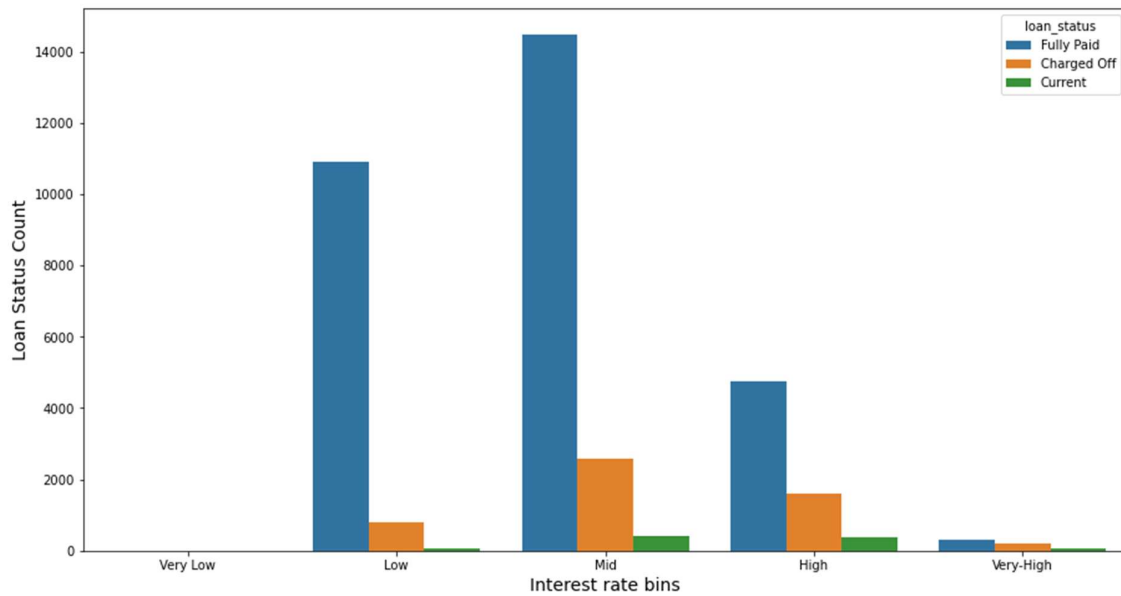




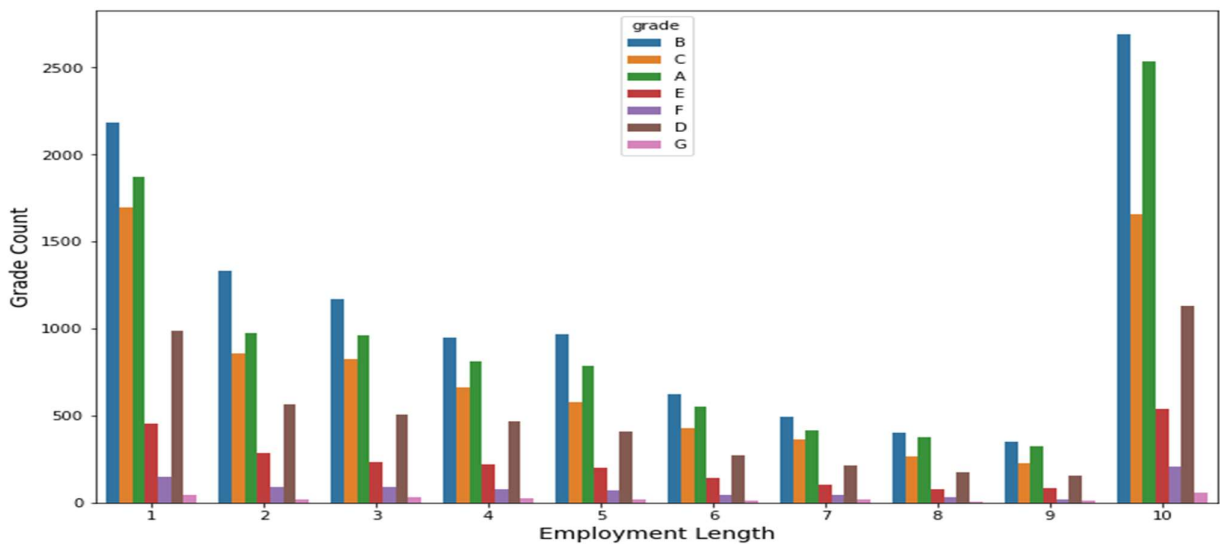
With increase in public derogatory records, annual income is also increasing, which could be the reason that with increase in pub_rec (public derogatory records) charged-off percentage decreased.



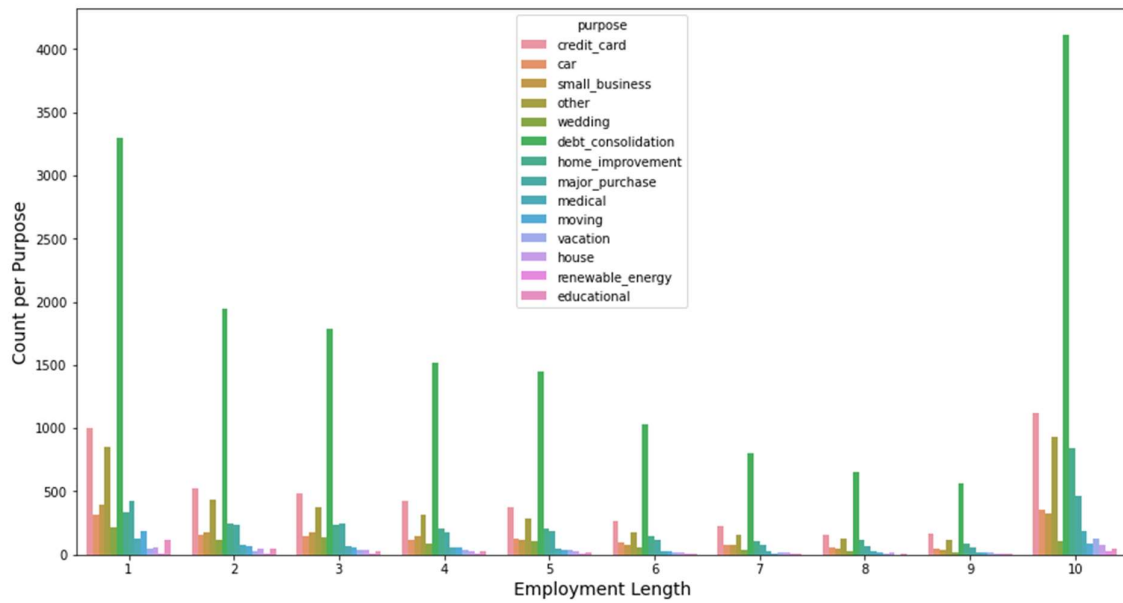
- Strongest reason to take loan is for "Debt Consolidation", which is common across all range of Annual income
- Normally, individual with salary range between 5-10k are the one taking most of the loan.



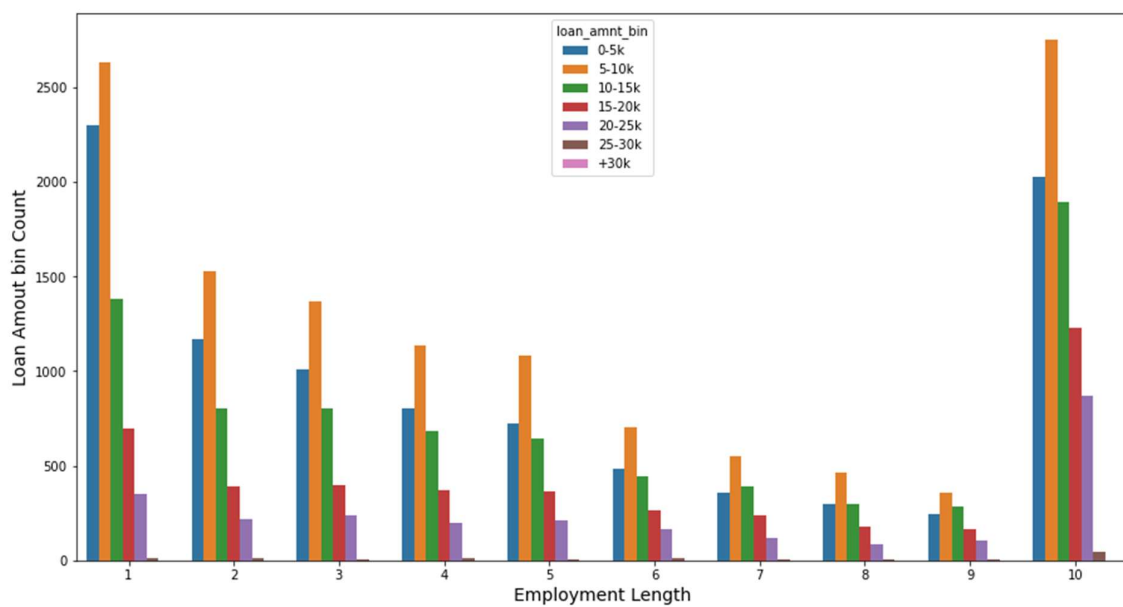
- As Loan interest increases, number of loans reduce significantly
- As loan interest increases, number of "Charged-off" increases.
- As loan interest increases, number of "Fully Paid" increases, maybe individual is trying to pay to loan fully then paying higher interest.



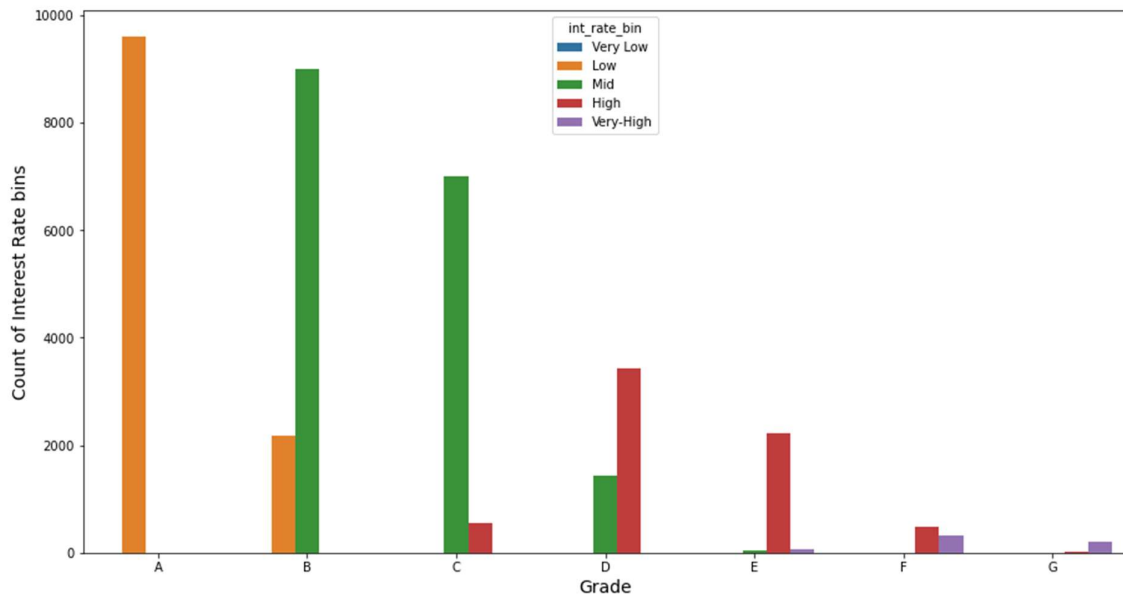
As Employee service length increases, Grade allocated to Individual increases. It makes sense, as number of years for individual increases, salary of individual increases and his paying capacity will also increase.



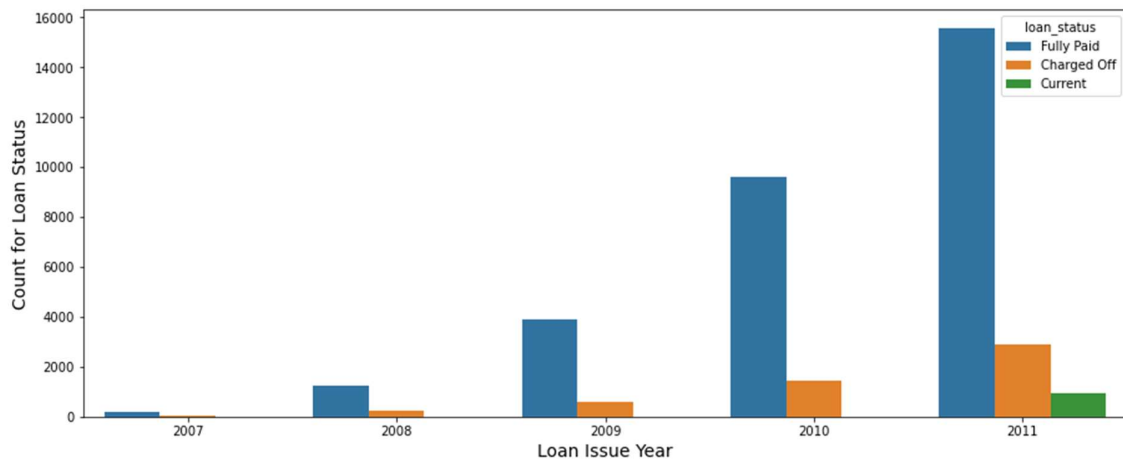
Across all individuals of employment service, loan is taken for Dept-consolidation



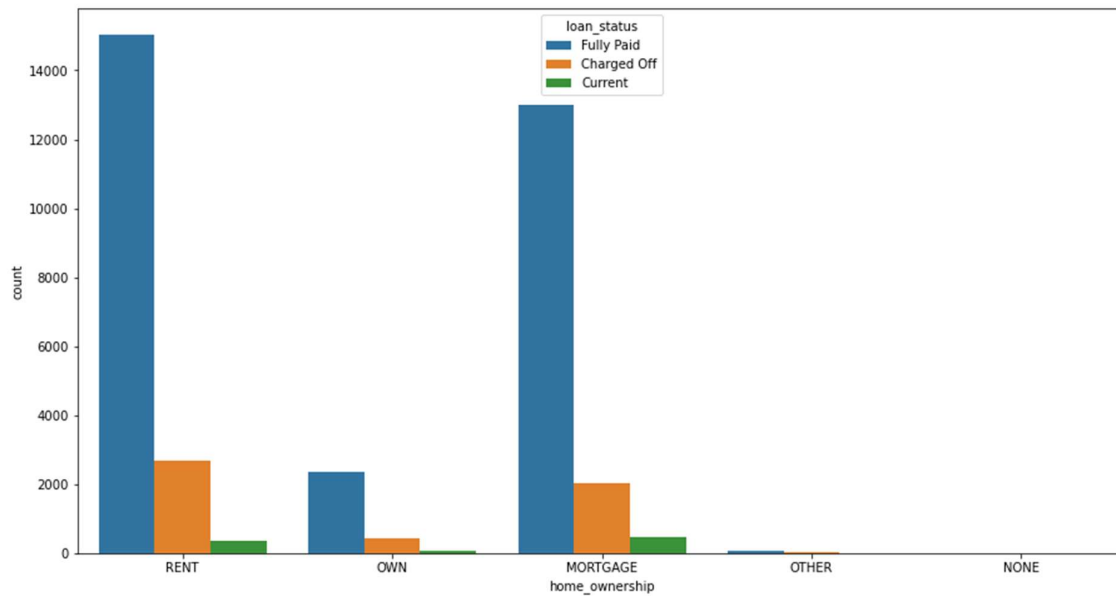
Across all obervation, mostly loan is taken for small quantity 5-10k, which makes sense as from earlier figures, mostly loan is taken for dept-consolidation



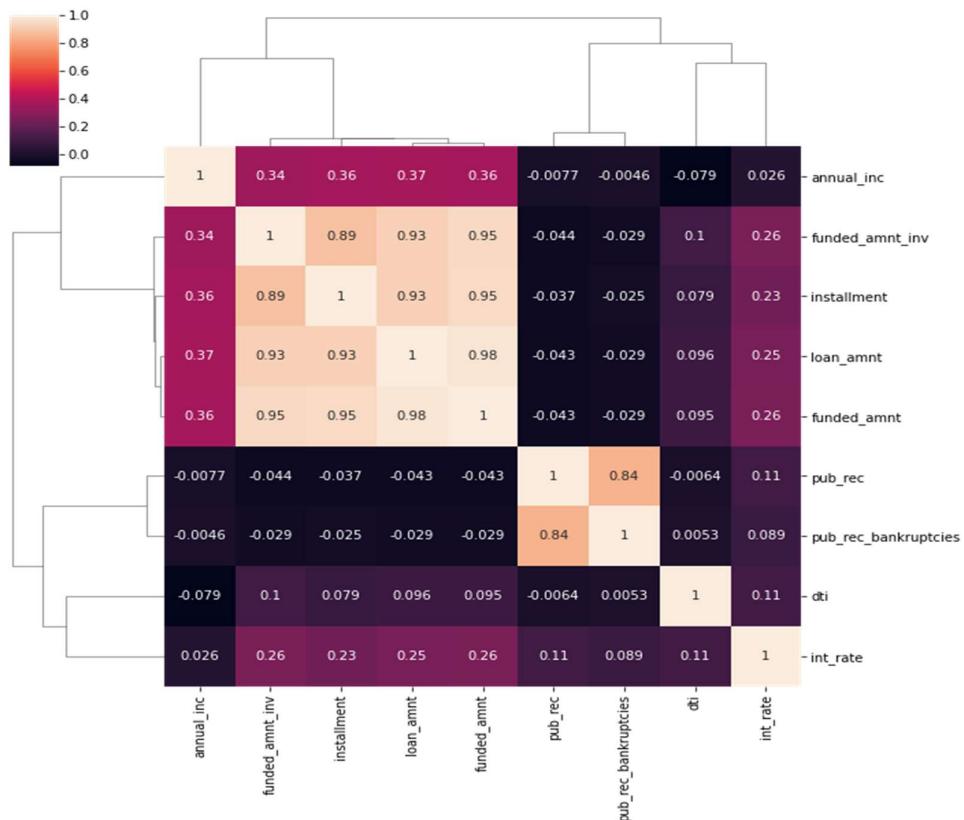
Interest rate increase with decrease in Grades



No of loans are increasing each year, which seems to be the reason for increase in Charged-off applications.



Individuals living in Rented or Mortgages house, take loan most and do have more defaulter compared to individual owning house.



Some of the highly co-related columns are following:

- funded_amnt, instalment, funded_amnt_inv, loan_amnt
- pub_rec and pub_rec_bankruptcies

Recommendation

- Loan taken with following characteristics are most likely to get Charged-off:
 - Loan is taken for "Dept-consolidation" or "small-business".
 - Loan amount is less: 5-10k.
 - When loan is taken at higher interest rates.
 - Loan is taken by Low salary individuals (0-20k)
- As salary of individual increase, his paying capacity increases and thus charge-off decrease.