LENDING CLUB CASE STUDY

- Case Study Submitted By:
  - M SOMASEKHARA REDDY(Group Facilitator)
  - SABYASACHI RATH

Problem Statement

Business Understanding overview:

You work for a **consumer finance company** which specializes in lending various types of loans to urban customers. When the 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 data given below 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.

When a person applies for a loan, there are two types of decisions that could be taken by the company:

1.Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:

1. Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
2. Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
3. Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

2.Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Case Study Implementation Steps

Data Loading and Understanding

Loading CSV file, Understanding Data Dictionary

Data Cleaning

Filter out 'current' row from loan status, Removing columns 30% of missing values and columns which are not useful for analysis, Removed % symbol from interest rate, extracted '+' from emplength, Some columns missing values replaced based on mode,meadian and with zero.

Removing outliers

Removed Outliers quantile .95 from Annual Income

Derive new columns

Derived new Columns

Explotary Data Analysis using Univariate, Segmented Univariate, Bivariate

EDA Data Visualization on important data using Univariate, segmented univariate and bivariate plots

Important Steps during Data Cleaning:

- Dropped the columns where the rows having more than 30% of missing values.
- Dropped the columns which are not useful in further Analysis.
- Fill up the nulls in the Employee length column with the zero value, if we try to fill with mode 0 its taking 10+ years due to that while plotting data 10+ years are more charged off instead of zero years' experience its leading to wrong analysis on data.
- Fill up the nulls in the Employee title column with the mode of zero value.
- Fill up the nulls in the Revolve util column with the mode of zero value.
- Remove % symbol from intrest rate and revol util column,it will be used for calculations.
- Extract digits from employee length column using regular expression.
- Converted amount columns into numeric data to find important correlation.
- Removed values which are present as Current from loan_status column due to this impact is not considered for analysis.

Remove Outliers:

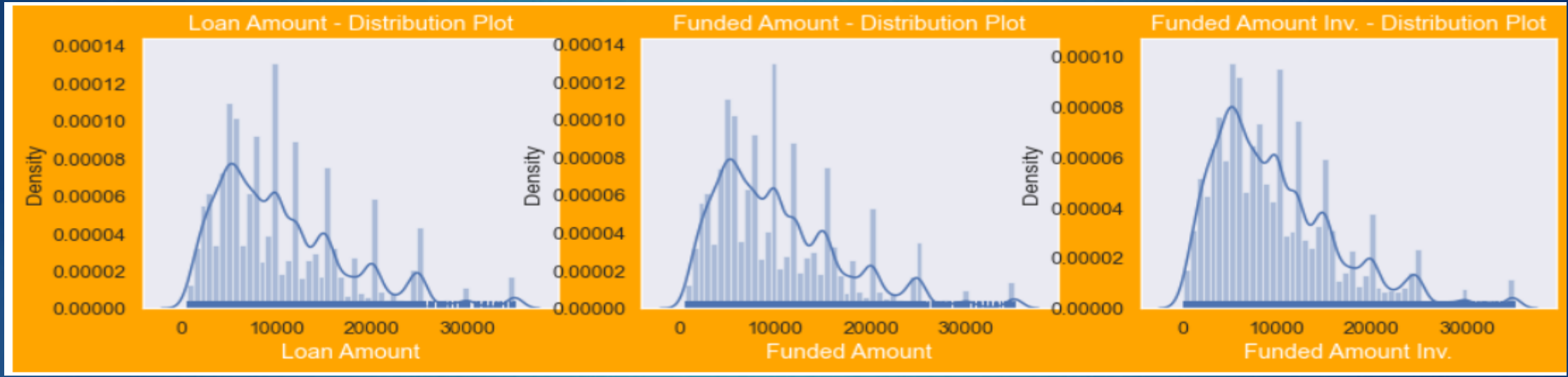
- Remove Outliers quantile .95 from Annual Income.

Derived Columns:

- Year, month, fund_amnt_cats, annual_inc_cats, int_rate_cats and dti_cats which can be used for further analysis.

Univariate Analysis

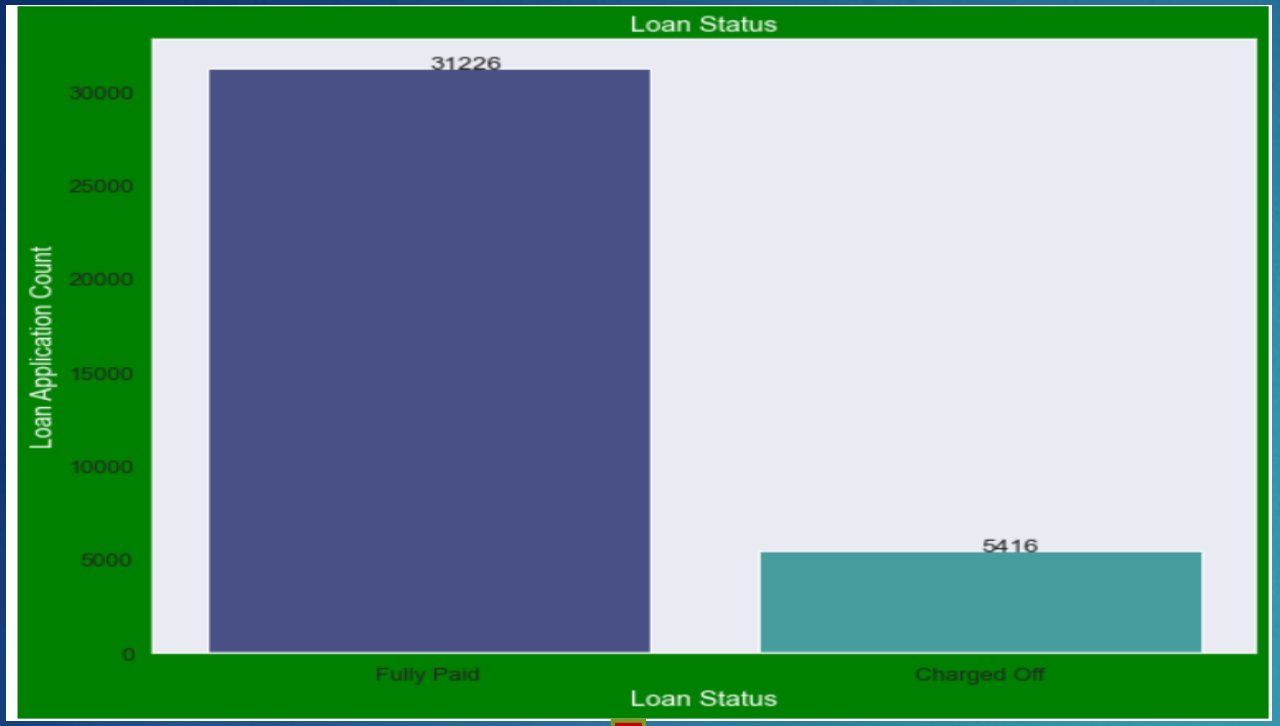
Univariate analysis using Distribution plot on Loan amount,Funded amount and Funded amount inv.



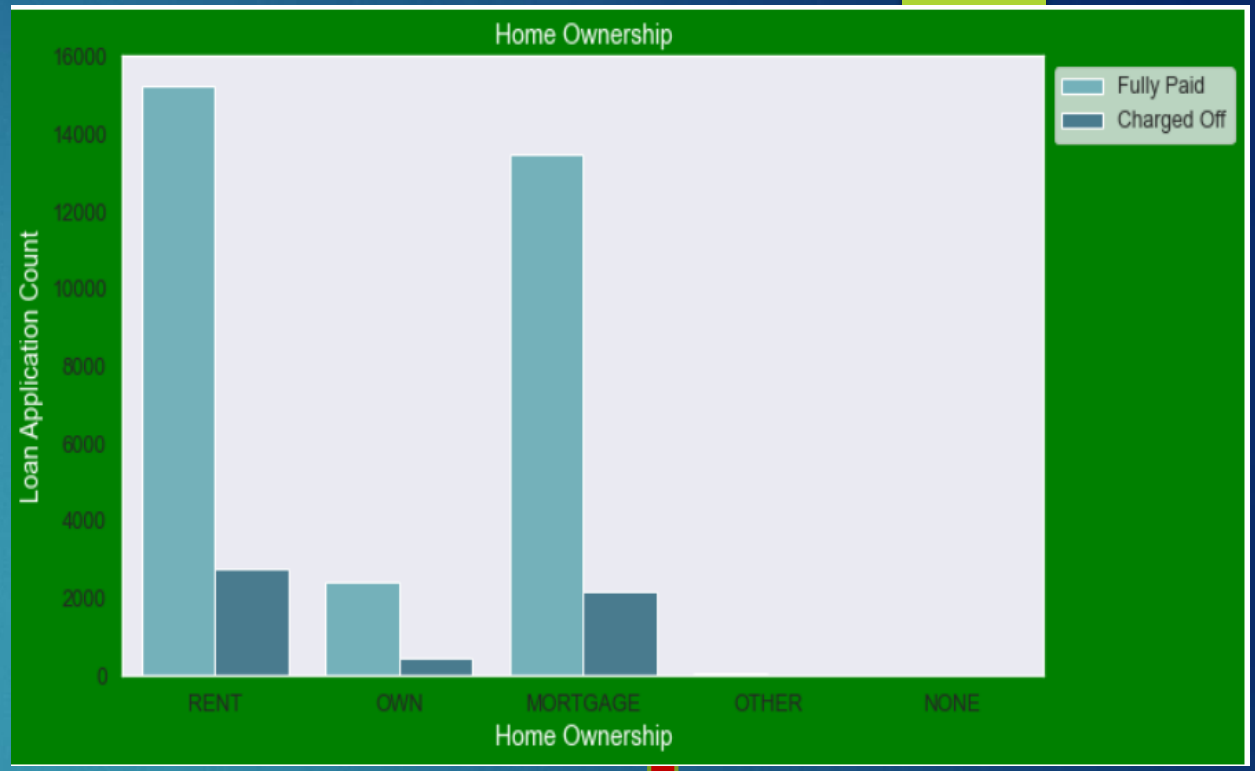
Observation:

- Distribution of amounts for all three looks close to similar.
- We will work with only funded amount column for further analysis.

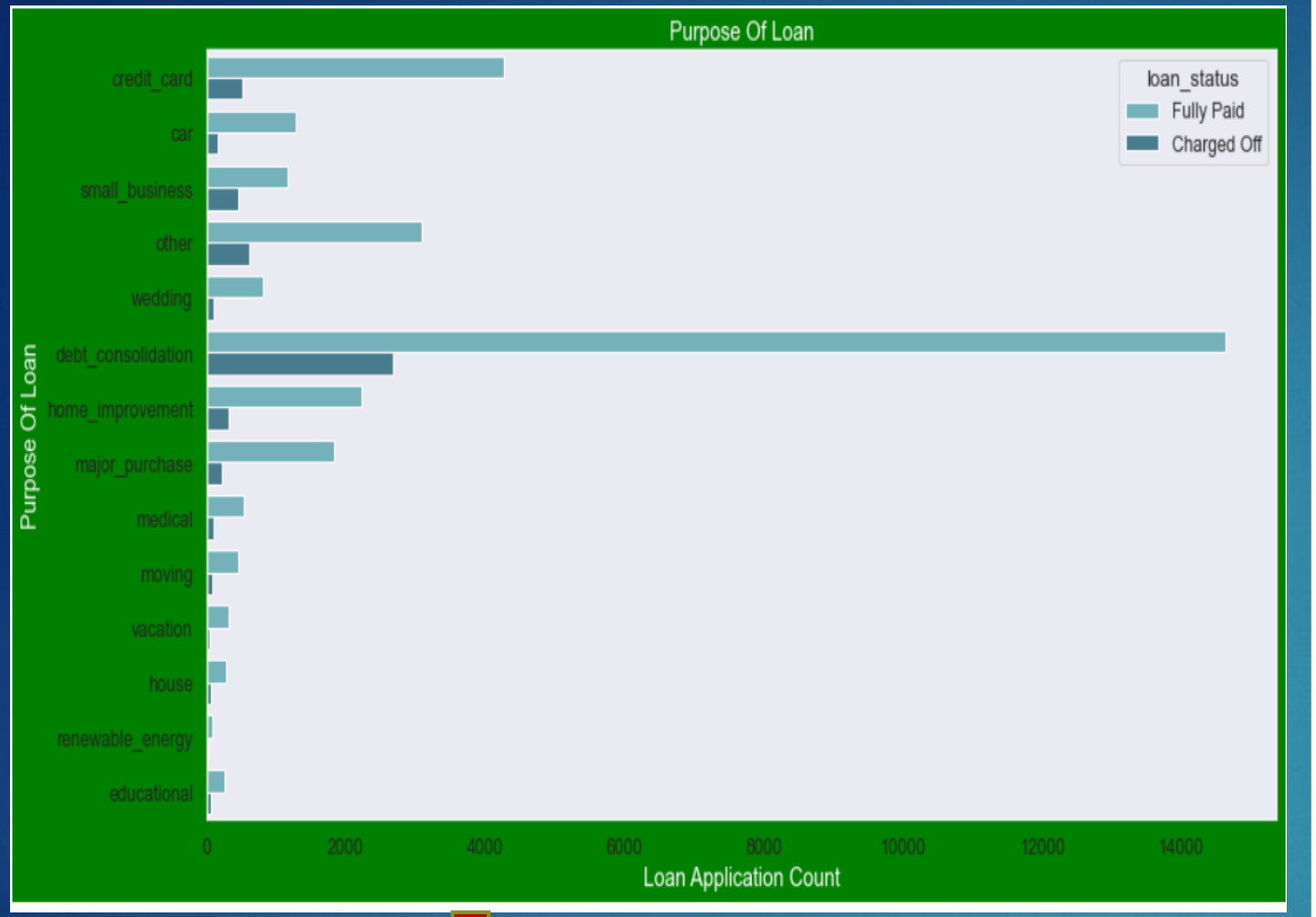
Segmented Univariate Analysis



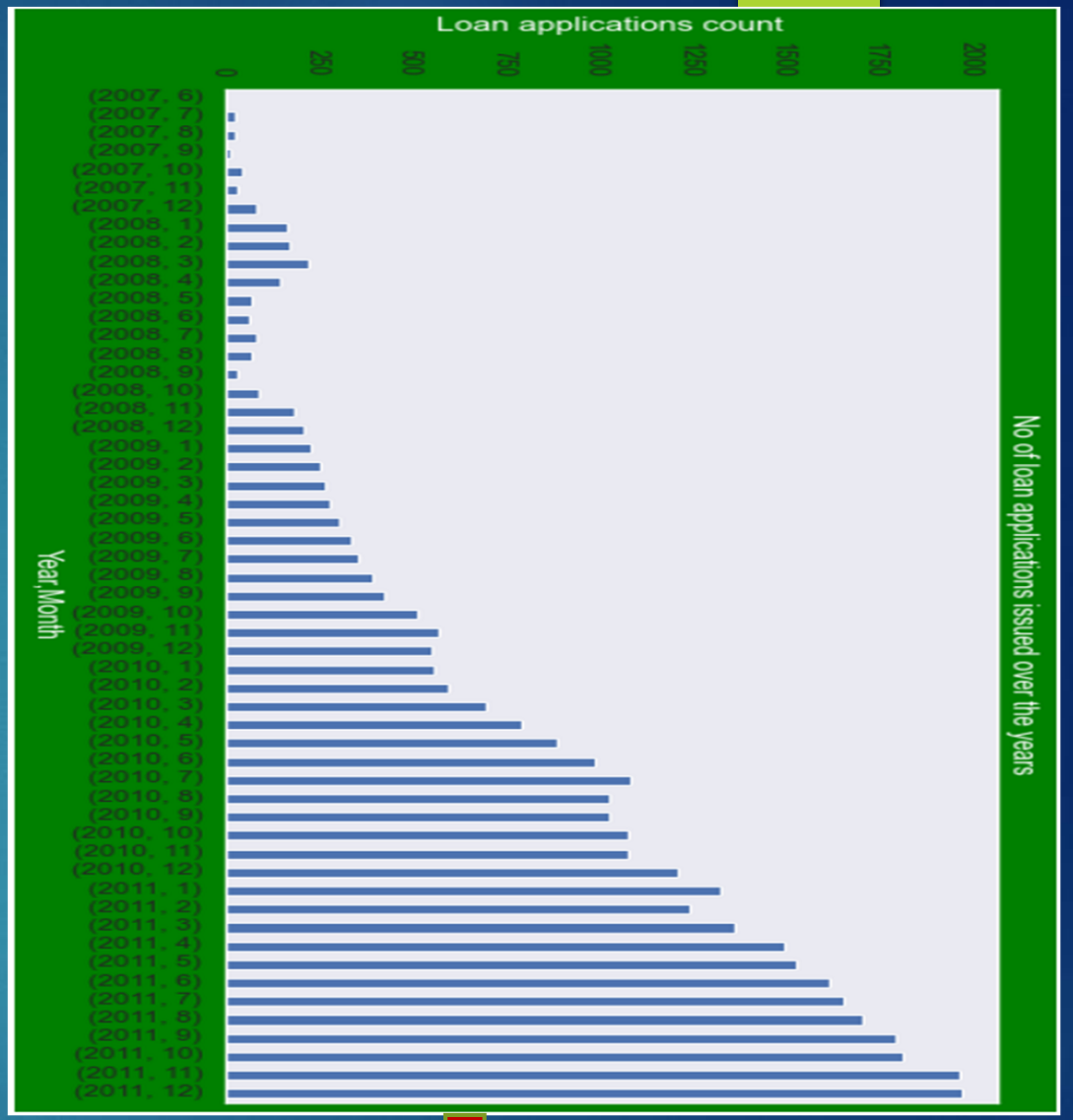
Plot shows close to 15% loans were charged off out of total loan issued



- Above plot shows that most of them living in rented home or mortgaged their home.
- Applicant numbers are high from these categories so charged off also high.

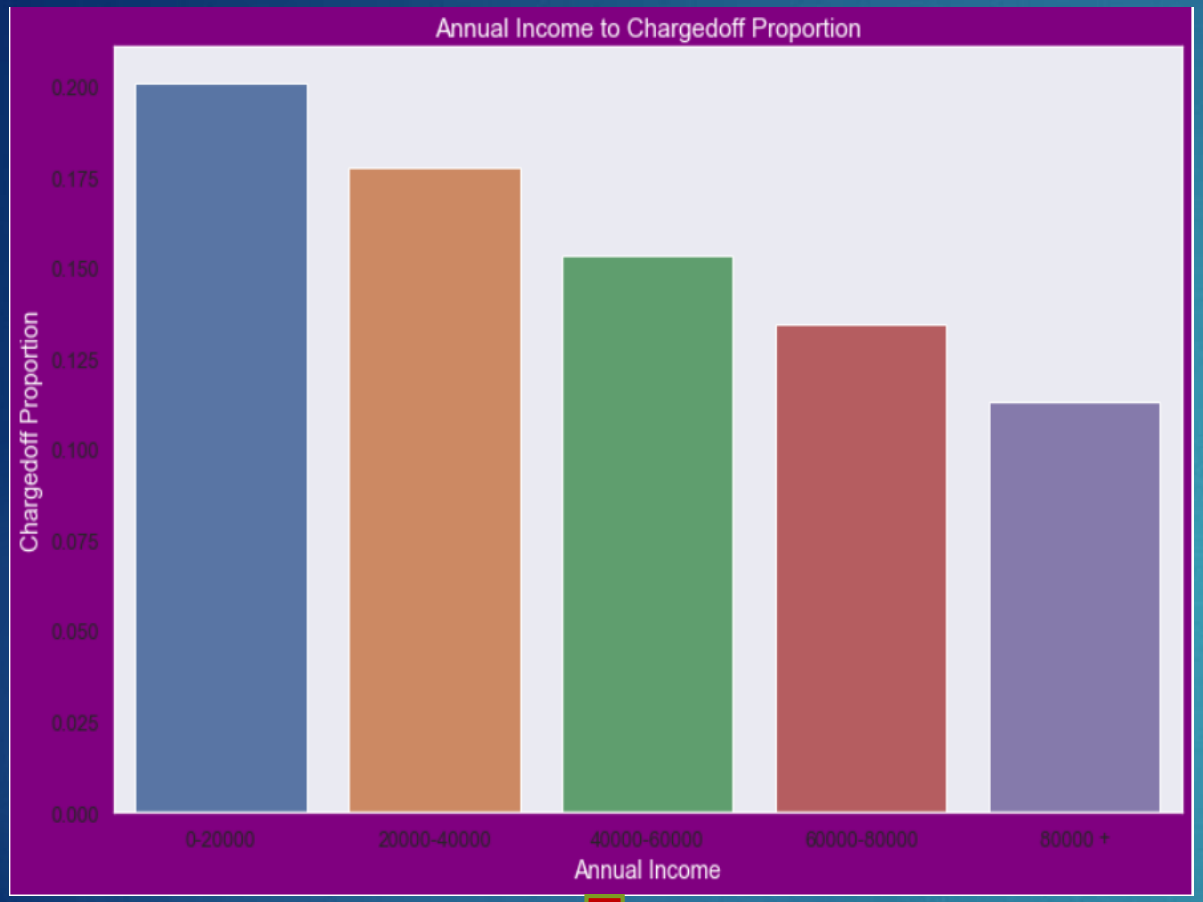


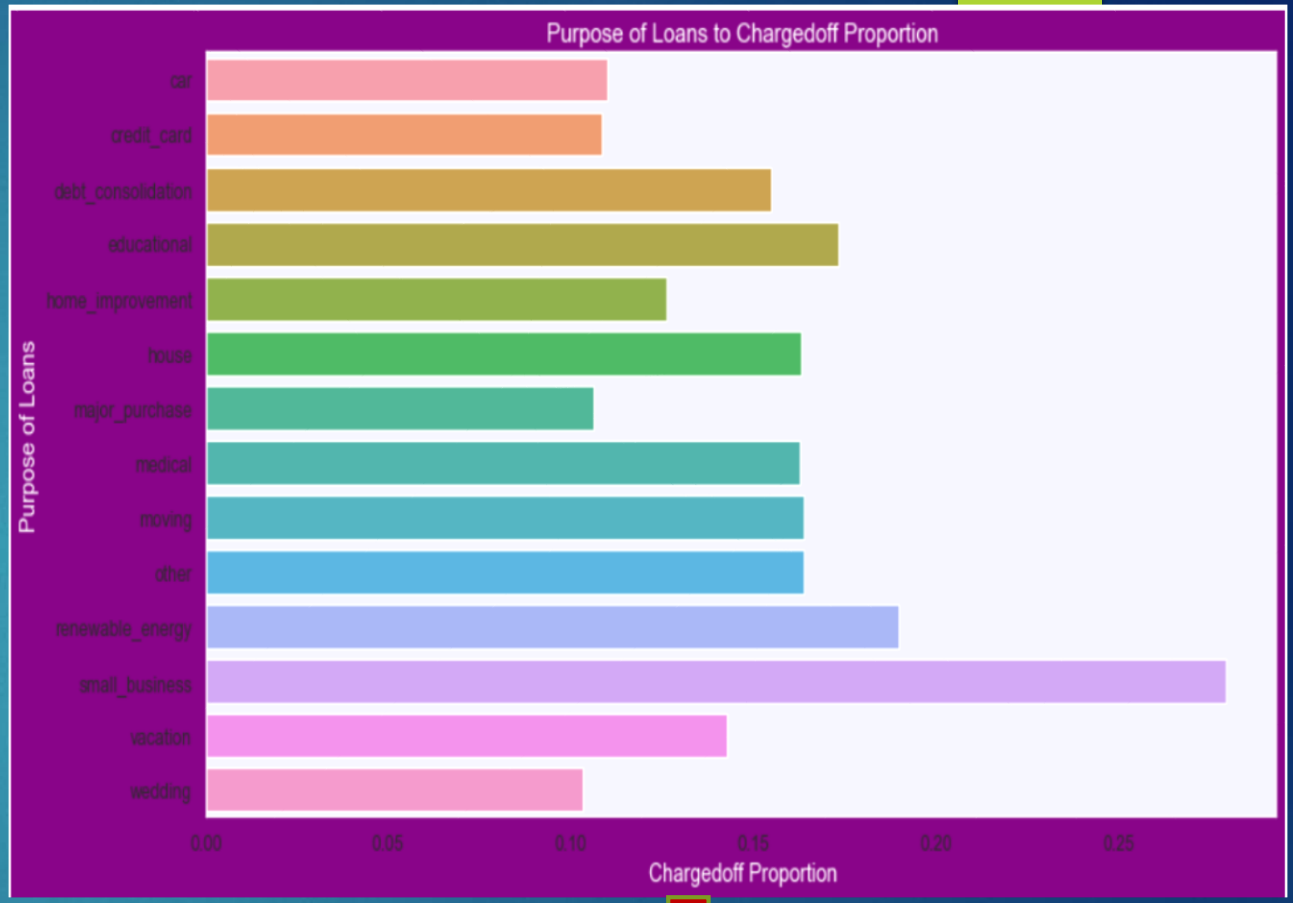
- Above plot shows loans were taken for the purpose of debt consolidation and paying credit card bill.
- Number of charged off count also high too for these loans.



- Count of loan application is increasing year to year.
- Increase in number of loan applications are causing to more Number of charged off applications.
- Number of loans issued in 2008(May-October).

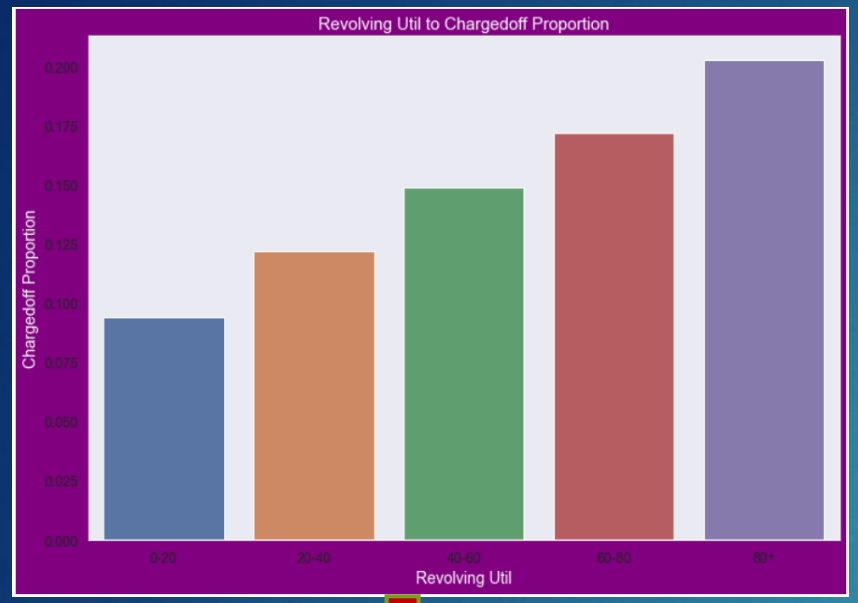
Bivariate Analysis

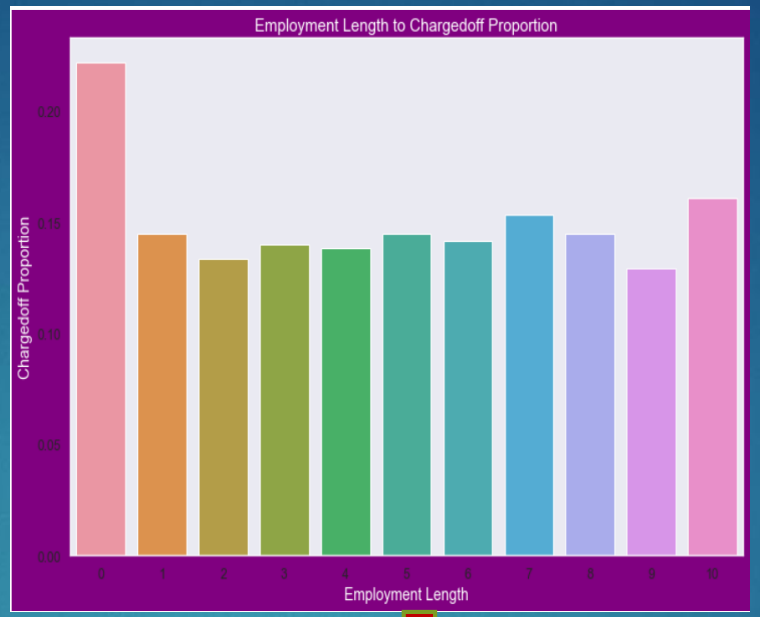


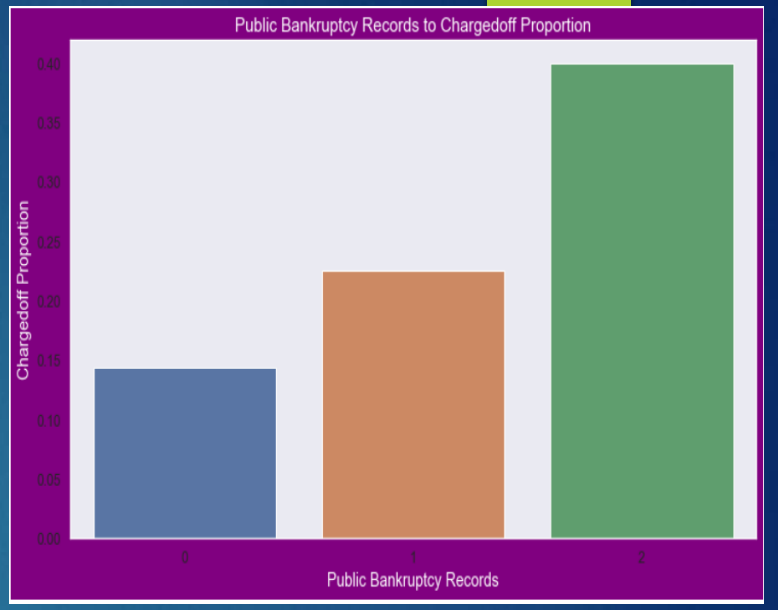


- Income range 80000+ has less chances of charged off.
- Income range 0-20000 has high chances of charged off.
- Notice that with increase in annual income charged off proportion got decreased.

- Small Business applicants have high chances of getting charged off.
- Renewable energy also next high changed off proportion .



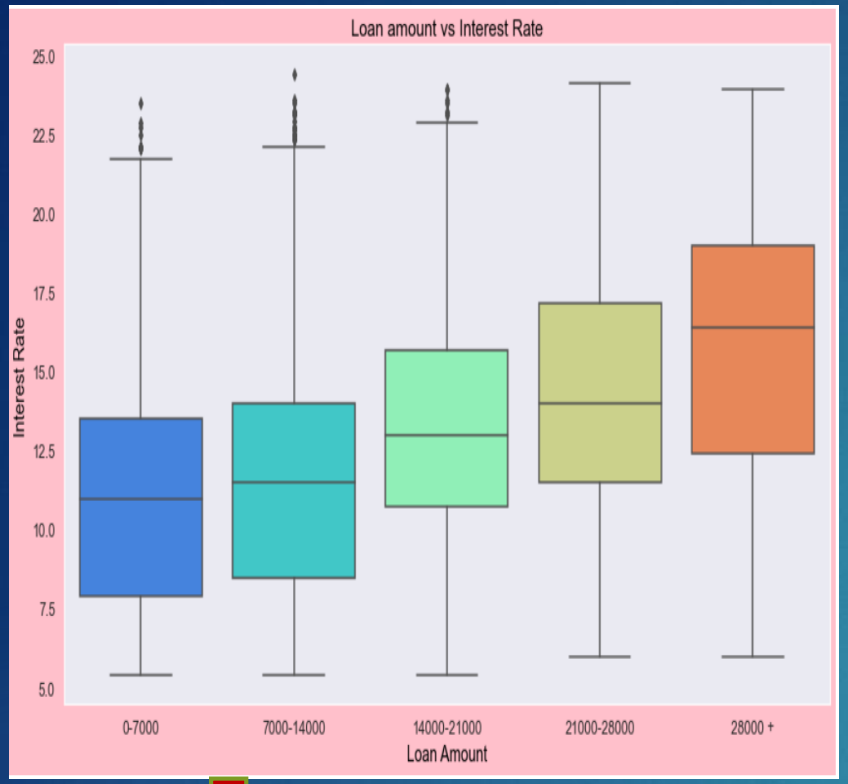




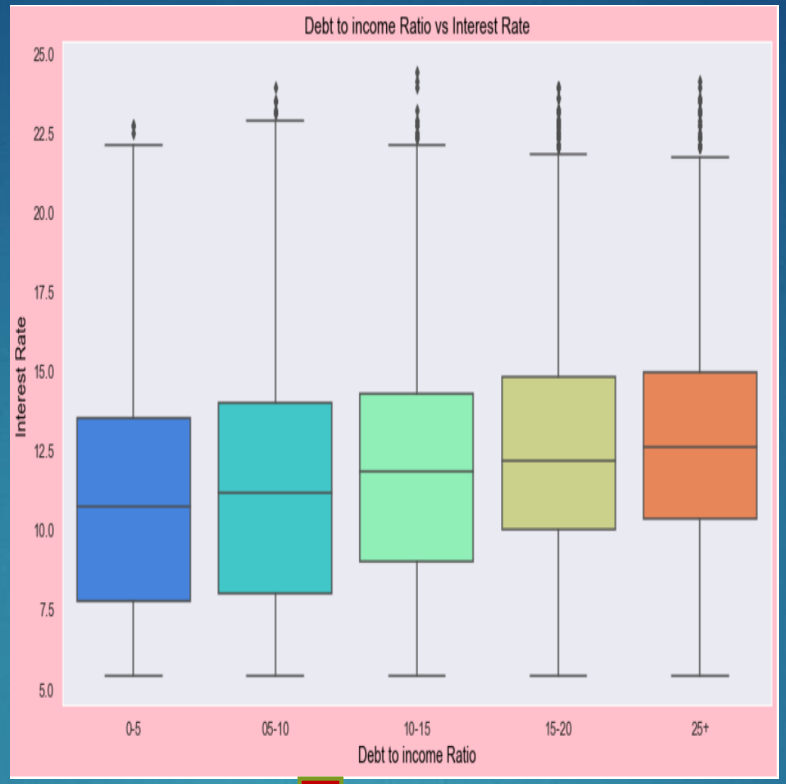
- Chances of charged off is increasing linearly with revolving line utilization rate.
- f the revolving utilization rate is more than 80% there is a maximum possibility of charged off.

- Those who are not working or have less than 1 year of work experience have high chances of getting charged off.
- However for the rest it is approximately similar. So employment length above 1 year has no impact on charged off.

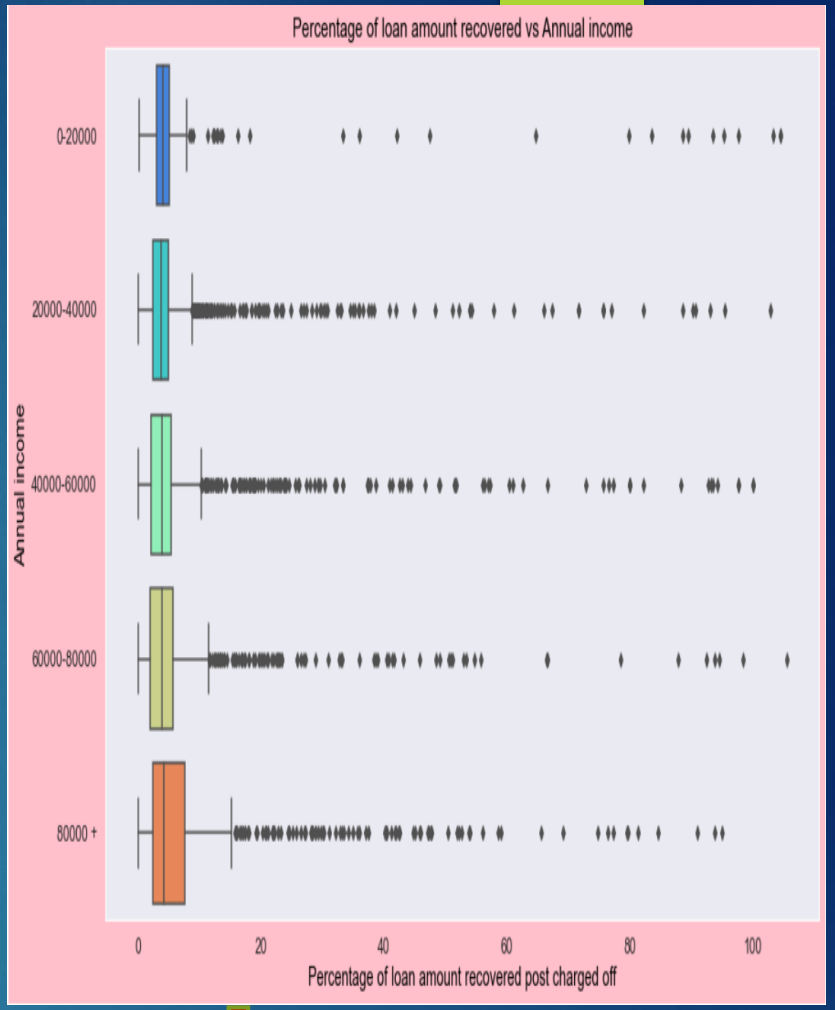
- pub_rec_bankruptcies count 2 has higher charged off proportion but those numbers are not significant to decide.
- People who have defaulted before are more likely to default in future as well.



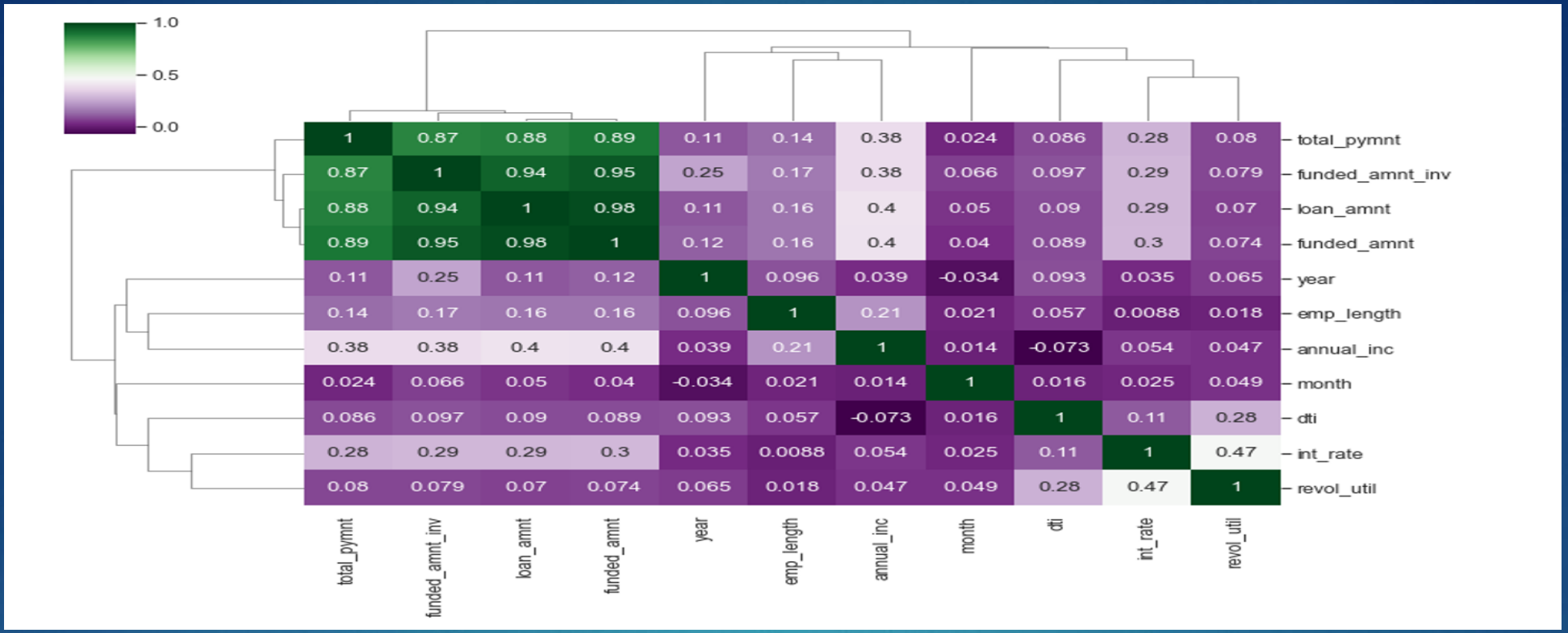
interest rate is increasing with increase in loan amount



- If your DTI is low enough you may get a lower interest rate.
- Plot shows no significant variation but there is slight increase in interest rate with increase in DTI.



- Higher percentage of loan amount is recovered when annual income is high.
- Plot shows no significant variation but there is slight increase in recovery percentage with increase in annual income.



- Loan amount-funded amount, loan amount-funded amount investor and funded amount-funded amount investor have positive and strong correlation.
- Annual income with DTI(Debt-to-income ratio) is negatively correlated, however the correlation between them is weak.
- There is a positive correlation between annual income and employment years.
- That means income increases with work experience
- A point to note is that DTI has limited impact on loan amount as per the dataset.
- Annual income and Loan amount are positively correlated.

Summary

- Loan amount-funded amount, loan amount-funded amount investor and funded amount-funded amount investor have positive and strong correlation.
- Annual income with DTI (Debt-to-income ratio) is negatively correlated, however the correlation between them is weak.
- Annual income and Loan amount are positively correlated.
- Number of charged off count is high for the purpose of debt consolidation and paying credit card bill.
- Count of loan application is increasing year to year, increase in number of loan applications are causing to more Number of charged off applications.
- Notice that with increase in annual income charged off proportion got decreased.
- Interest rate is increasing with increase in loan amount.
- Interest rate is increasing with increase in loan amount.