Lending Club Case Study

Developers Name :

1. Madhav Parashar
2. Nikhil Sharma

Problem Statement

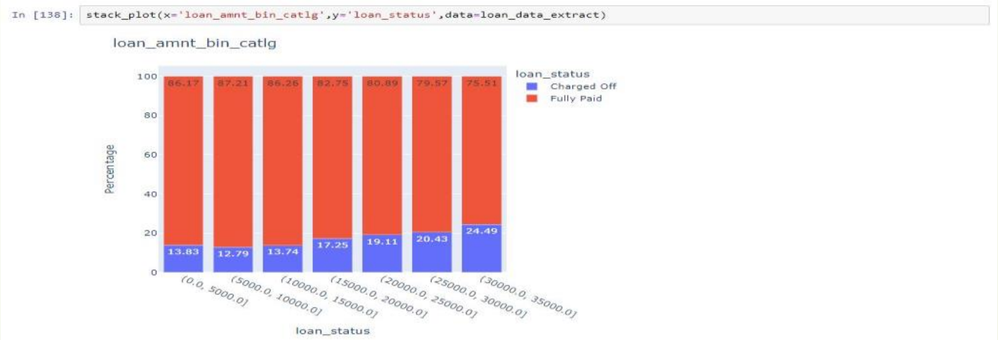
- This case study is to develop the basic understanding of risk analytics in banking and financial services and understand that how data is used to minimise the risk of losing money while lending to the customers.
- Here, We are having the loan data of customers with the below status.
  - **Fully Paid -:** Applicant has fully paid the loan (the principal and the interest rate).
  - **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'.
  - **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.
- So in our dataset we need to find out the types of risk that will help the Bank to make the decisions on the loan application.
  - 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.
- In this case study we are using the EDA to understand that how **consumer attributes** and **loan attributes** influence the tendency of default.

Analysis Approach

- Load the data using panda library.
- Data Cleaning
  - Remove columns those contains only null values.
  - Remove columns those having same values.
  - Remove columns having more than 30% null values.
  - Plot bar chart to see the distribution of ,missing values in each column.
  - Check object datatype and resolve them like remove ‘%’ etc.
- Select grade and sub-grade as categorical data.
- Check the employment duration.
- Check home ownership.
- Check last month payment was received or not.
- Check the last month LC.
- Calculate the credit history in terms of years.
- Create the new derived column to categorise if the emp_title is disclosed or not.
- Impute 'Not known' for the bankruptcies record where the data is not available.
- Delete the columns that do not add any weight to the analysis.

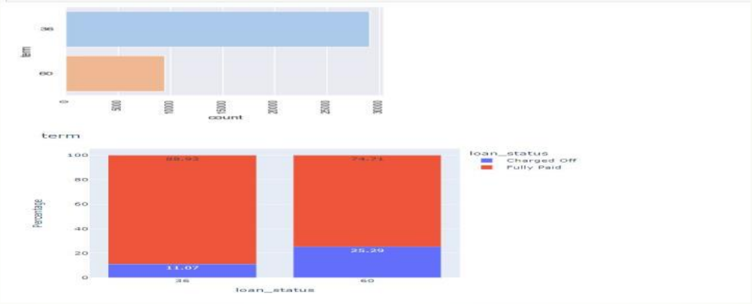
Results of Univariate

- Observation 1 -:
  - In general, as the loan amount is increasing, the percentage of the charged off loans is increasing.
  - The percentage of Charged off loans is highest for the loan amounts in range 30000 - 35000. But, the number of loans in this category are less.
  - From the count plot, the highest number of loans is for 5000 - 10000, followed by 0 - 5000 and 10000 - 15000. With in top 3 high volume categories, the laons that all with in 0 -5000 are slightly likely to be charged off.

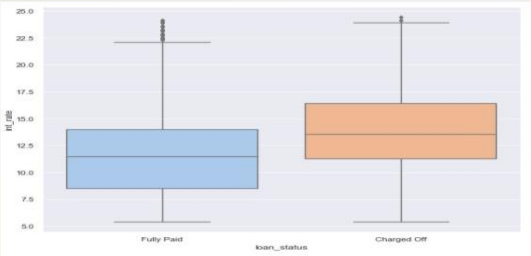


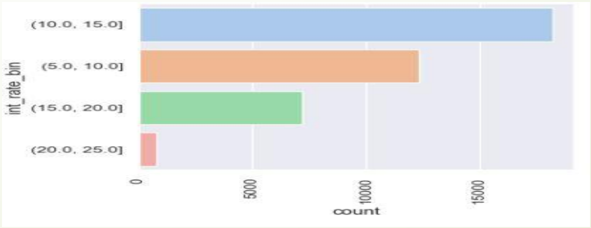
Results of Univariate

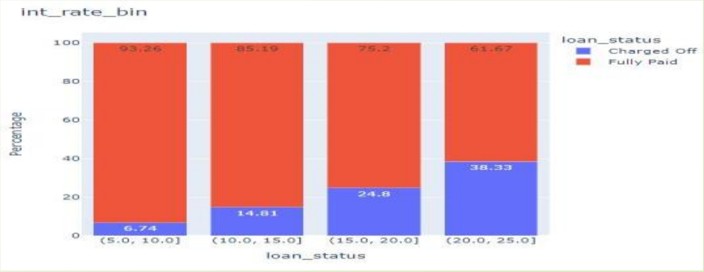
- Observation 2 :
  - Loans with longer tenure have high percentage of Charge Off applications. This indicates that customers who have longer commitments opted charge off due to varying financial circumstances over the time.



- Observation 3 :
  - In general, the higher the interest rate is, the higher is the percentage of Charged off Loans.
  - Loans with in interest rate 10 -15 are most frequent, followed by those with 5-10 and 15-20.
  - However, the percentage of chareged Offs is high in 10-15, followed by 5-10.
  - This analysis evidences, that there is some risk in issuing the High Interest Loans and additional due deligence is required while processing loan applications

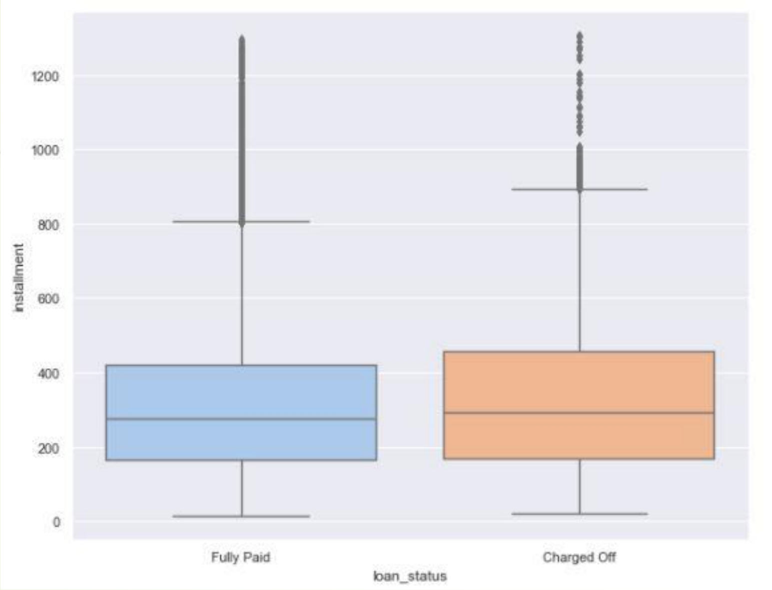


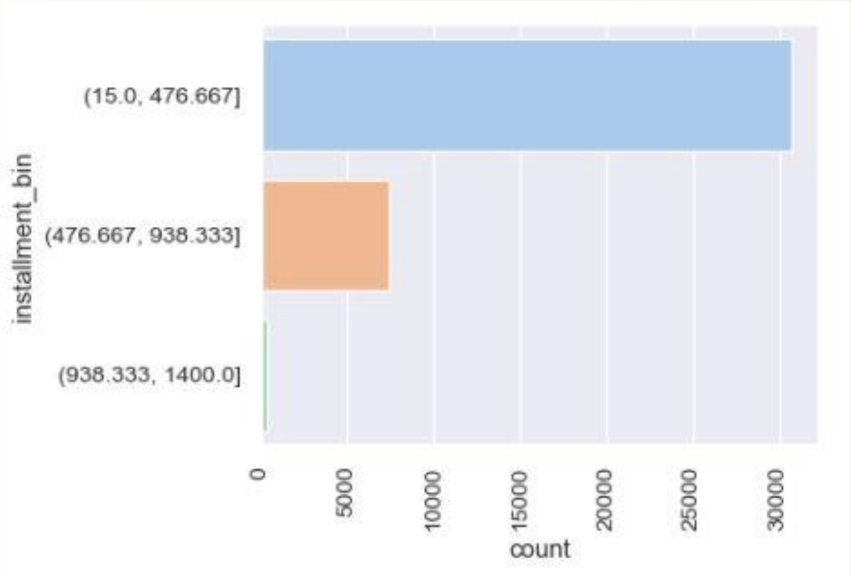


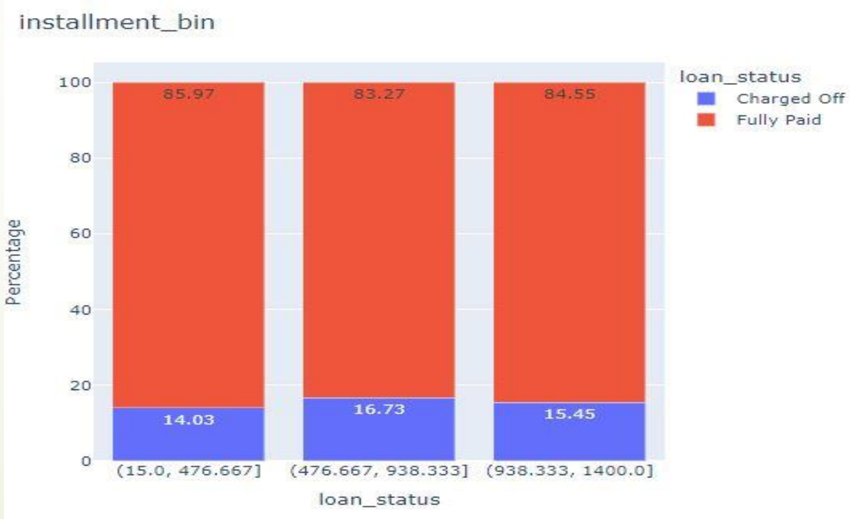


Results of Univariate

- Instalments

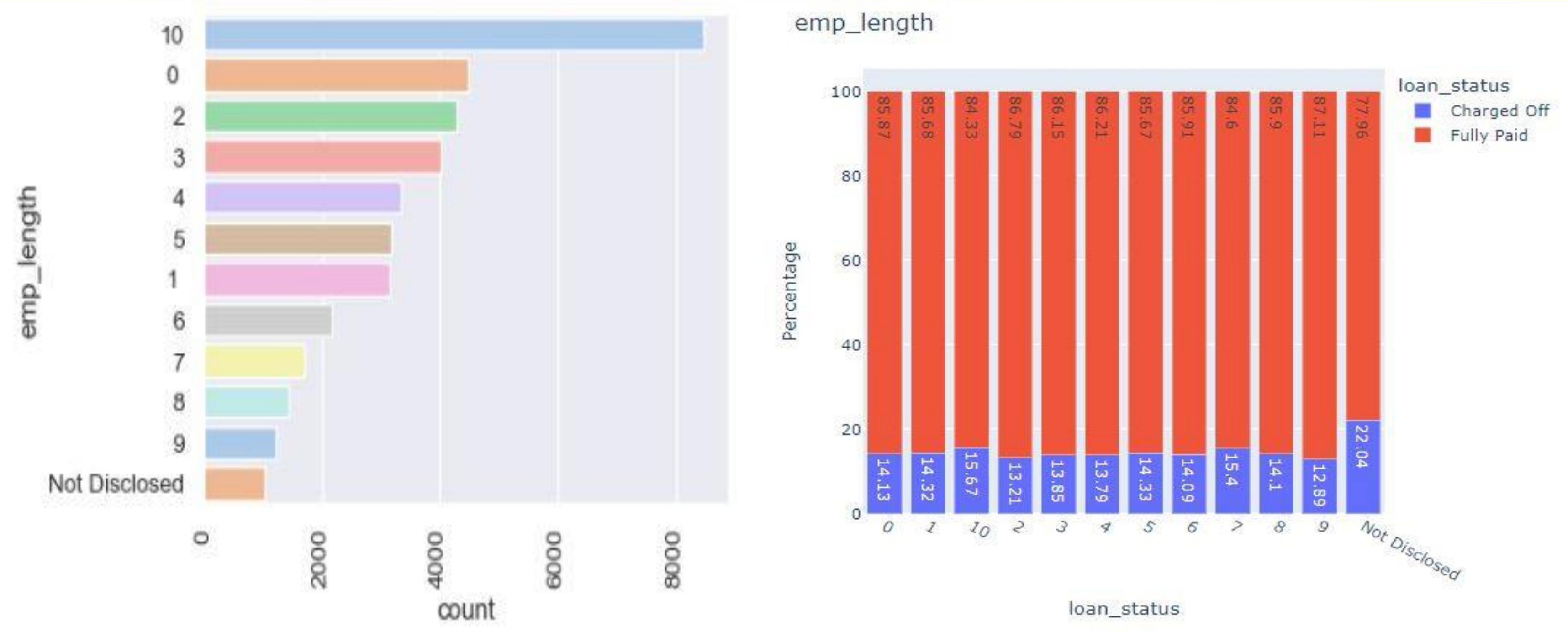






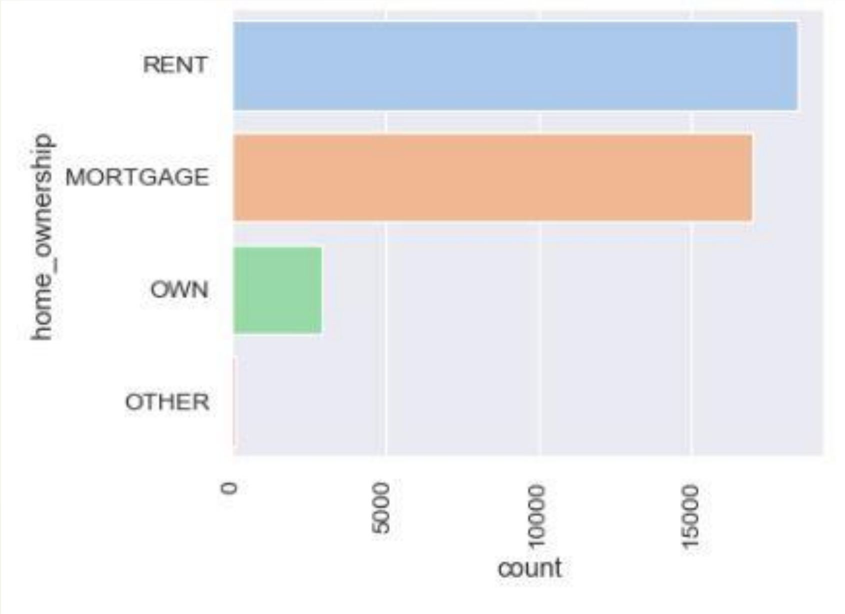
Results of Univariate

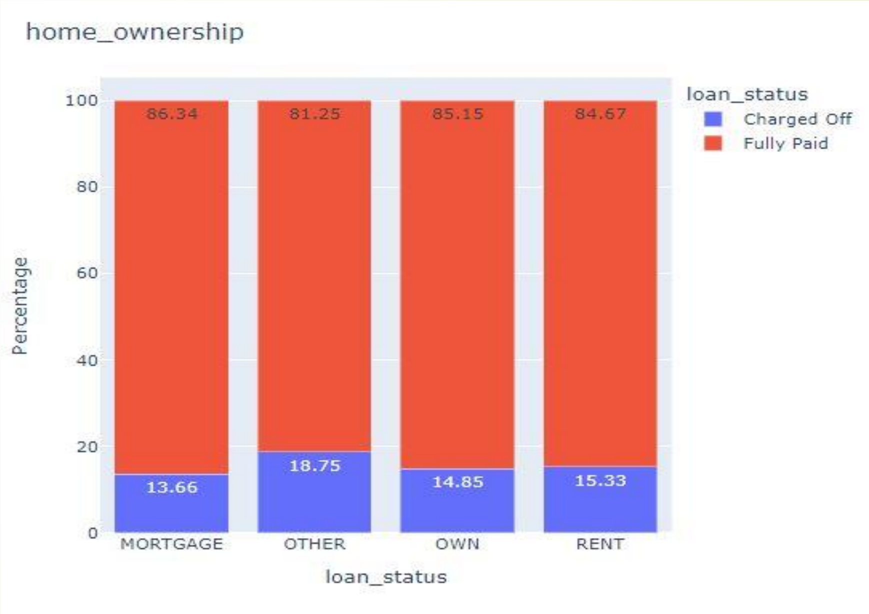
- Tenure of Employee
  - There is no clear trend of employment length impact on the charged off rate. Further investigation will be required to study this variable.



Results of Univariate

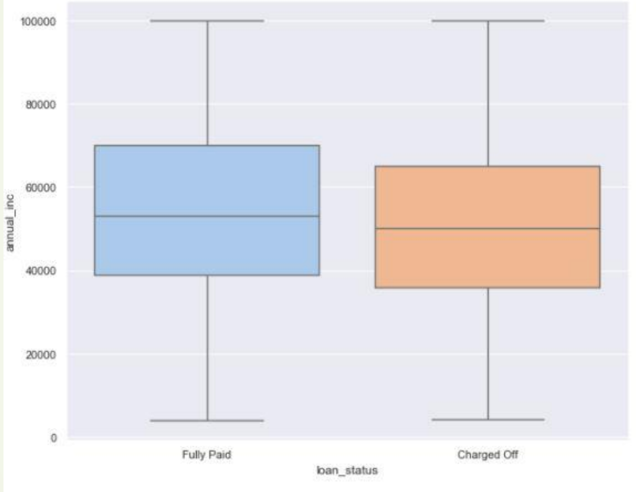
- Home_Ownership: type of home ownership rented, own or mortgage
  - The home ownership with 'other' are charging off the loans at higher rate.
  - This shows that 'other' category of the home ownership should be identified as risk. Further checks are required to understand the residency status.

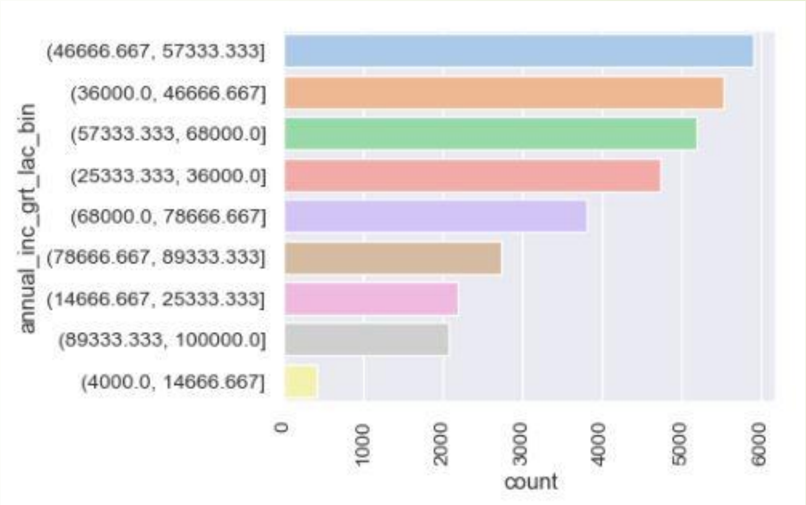




Results of Univariate

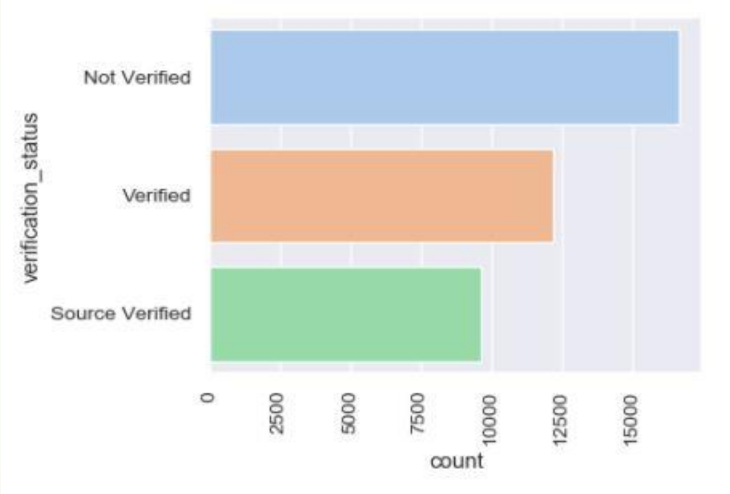
- Annual income
  - Employee greater than 1 Lac salary:
    - we can see that employees with greater than 1L rang, those are getting higher salary in range 4.5L t0 6.0L are more likely to charged off
  - Employee less than 1 Lac salary:
    - we can see that employees with greater than 1L rang, those are getting lower salary in range 4k t0 15k are more likely to charged off.

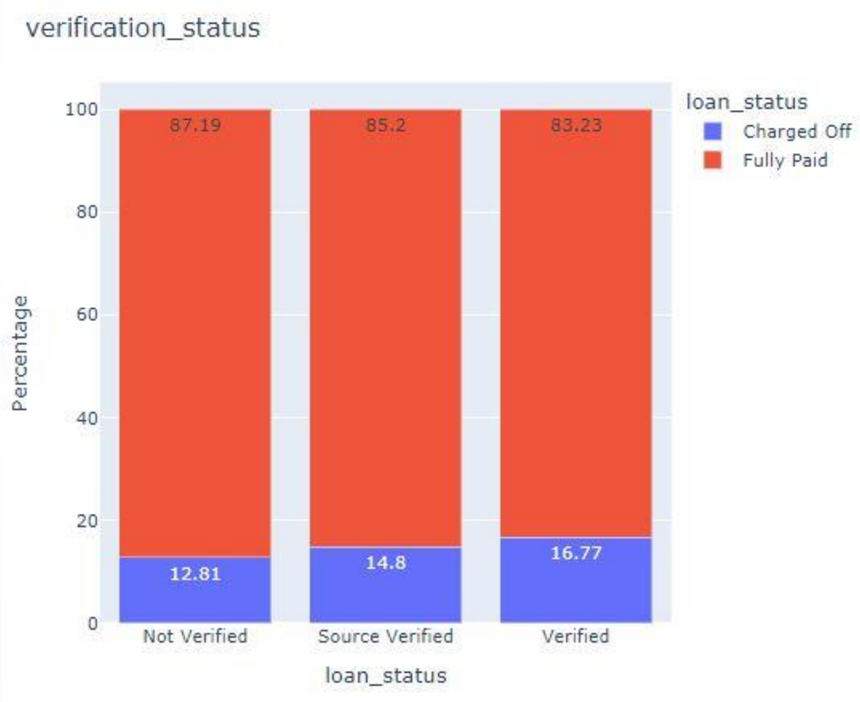




Results of Univariate

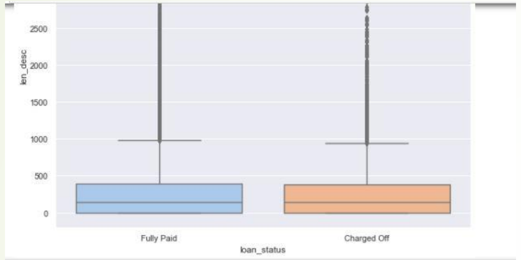
- verification status: Indicates if income was verified by LC, not verified, or if the income source was verified
  - In case of Income verification, the charge off rate is higher ,when the income has been verified.
  - This shows, that income verification checks have no effect in case of Risky applications. The income verification process will need to be reviewed.

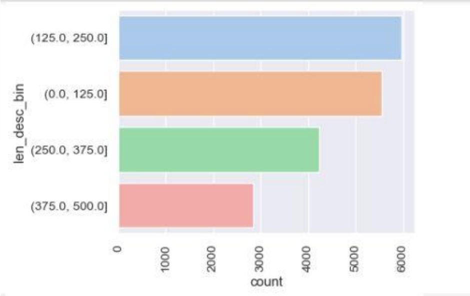


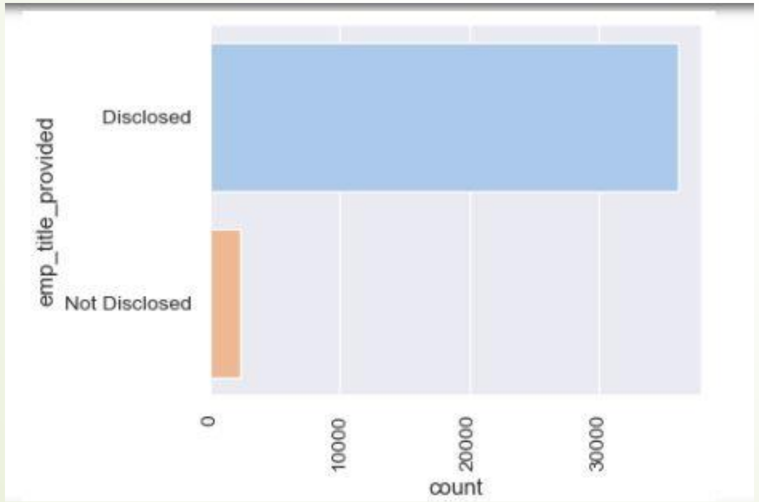


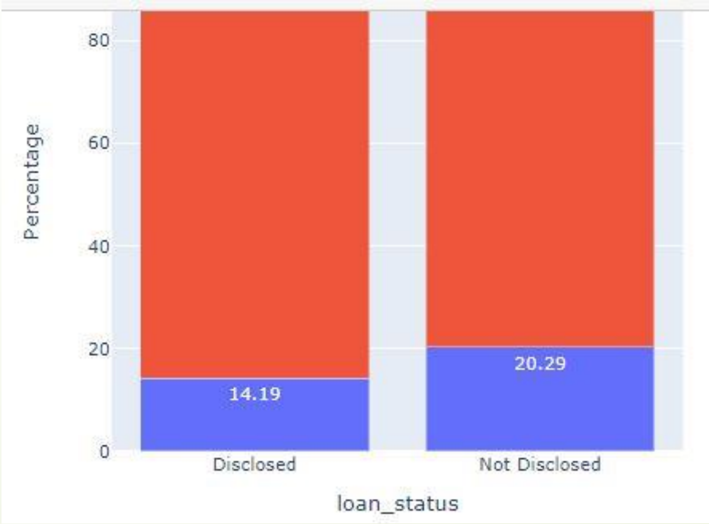
Results of Univariate

- emp_title_provided and Decription_length
  - Apparently, the rate of charged off applications are higher in cases, where the Employee title was not disclosed.
  - However, the 'Not Disclosed' is much less compared to 'Disclosed' category and that could have some bearing on this trend. Further Analysis will be required on Employee title to draw logical conclusions.



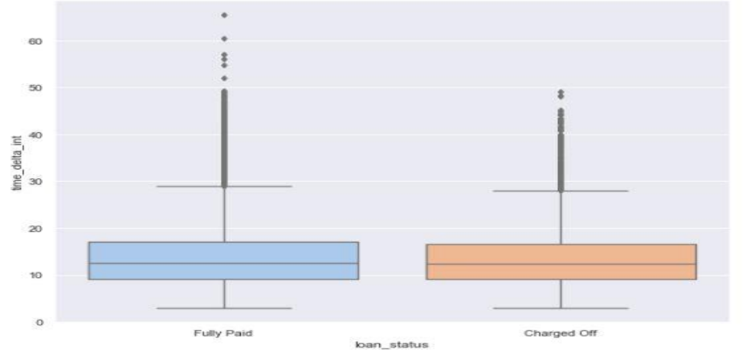


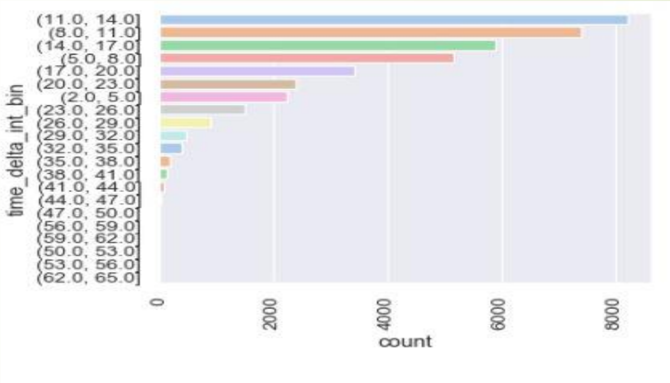


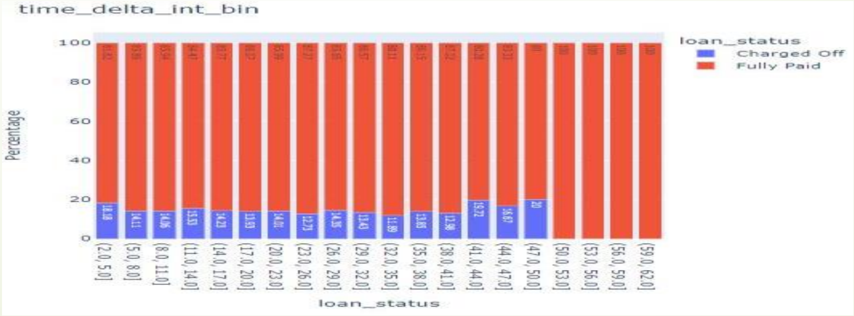


Results of Univariate

- Analyse Credit History - The difference in the issue date and earliest credit date
  - There is a mixed trend in the charged off loans when plotted against the credit age.
  - In general, a very low credit age meant, the financial profile of the applicant is not clear.
  - A high credit age meant, the financial profile of the applicant is superseded by the age of the application.
  - This trend is clearly visible in the above percentage stacked plot.

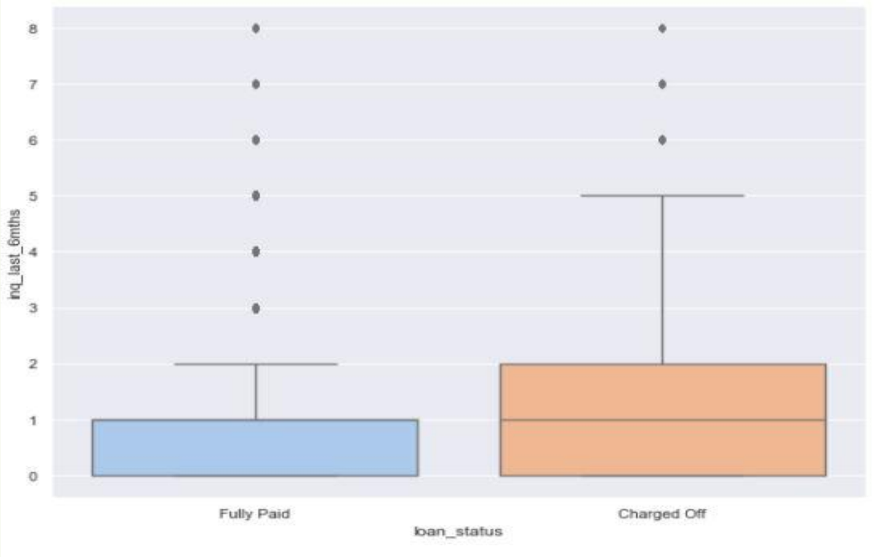


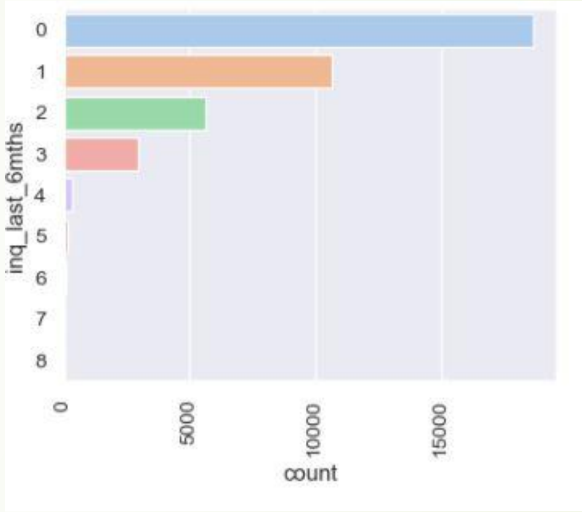


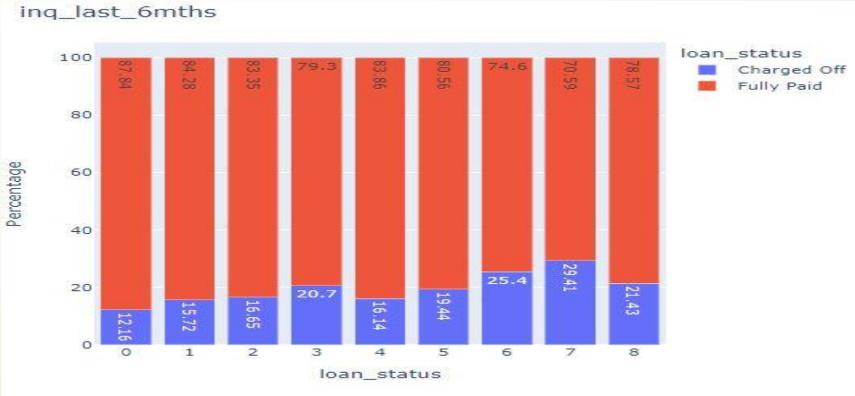


Results of Univariate

- inq_last_6mths
  - Considering both the count plot and stacked percentage plot, it is evident that, the higher the number of inquiries, the higher is the rate of charge off. This parameter shows the desperation of the customer for a credit line.

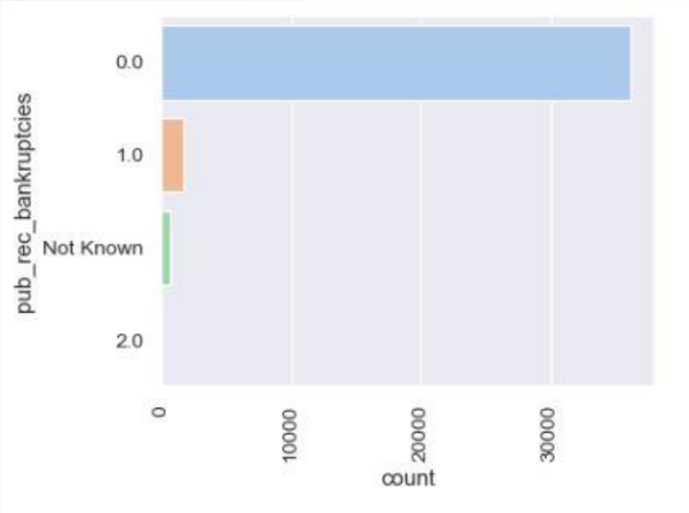


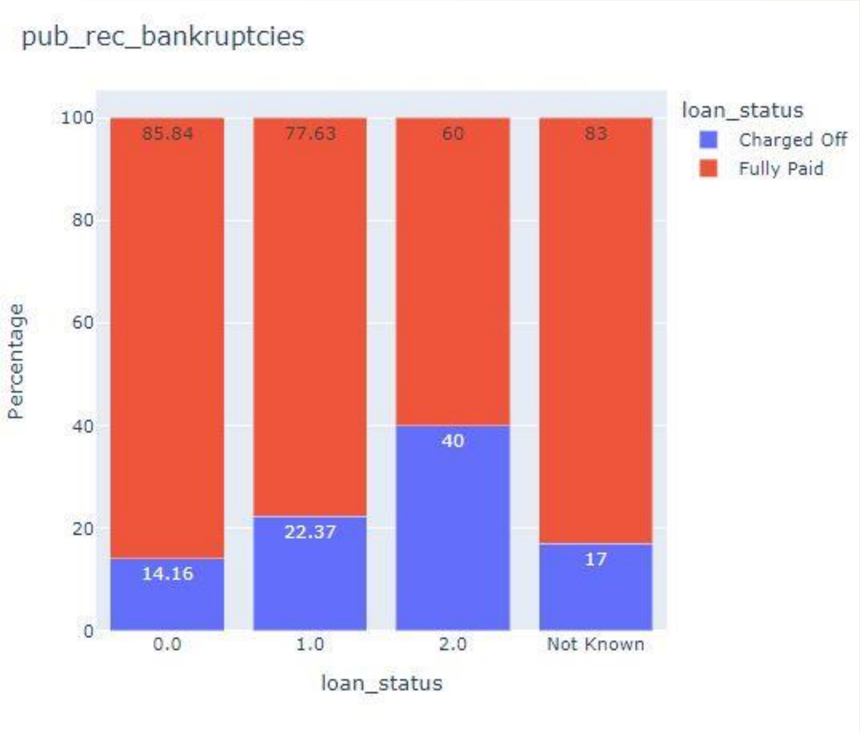




Results of Univariate

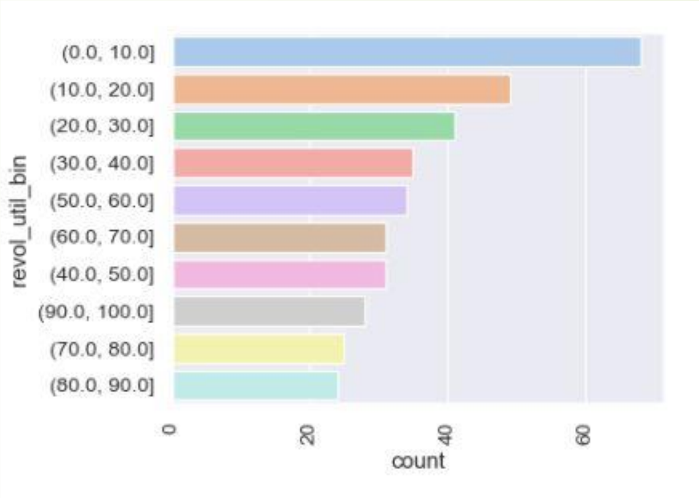
- pub_rec_bankruptcies: Number of public record bankruptcies
  - It is evident from the above chart that, the higher the bankruptcies are, the higher is the charge off rate.

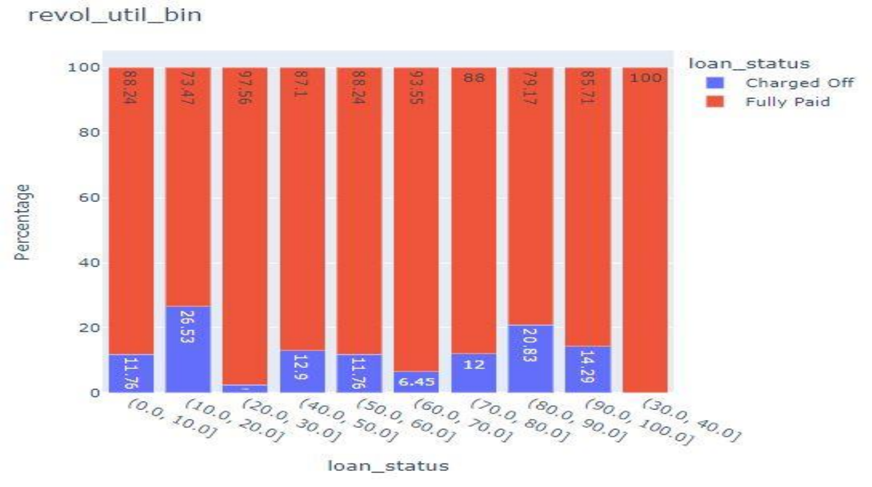




Results of Univariate

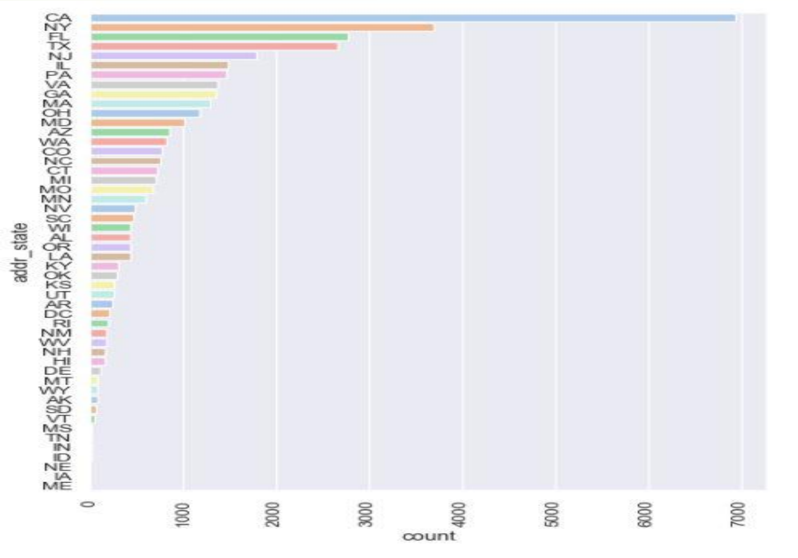
- revol_bal : Total credit revolving balance
  - There is no clear evidance of revolving balance on charged off rates

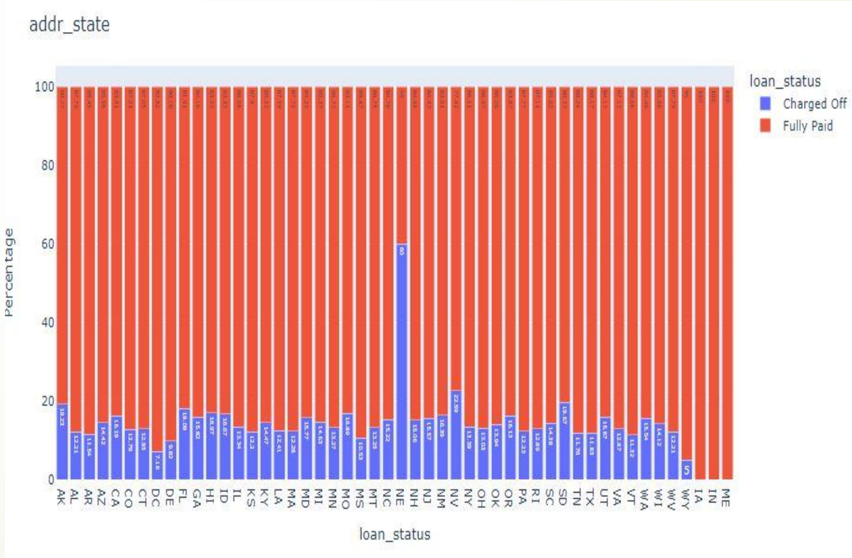




Results of Univariate

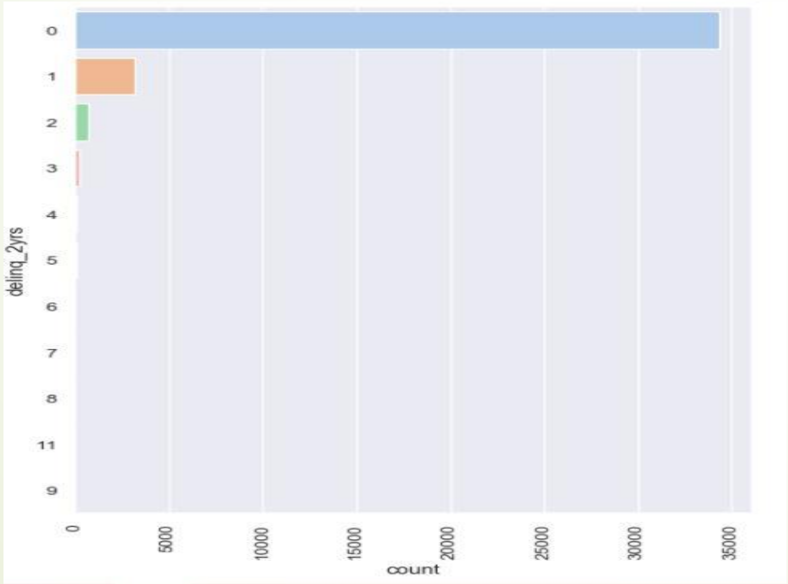
- addr_state
  - As per the above state 'CA' has highest number of charged off applications.
  - In terms of percentages, the loans are highly charged off in the state NE,NIU and AE.
  - Observing the number of loans applied from each of these states, it is clear, that the number of applications are less and hence
  - this behavior. Excluding the above outlier behavior, the state with highest charged off percentage is NV - Nevada.

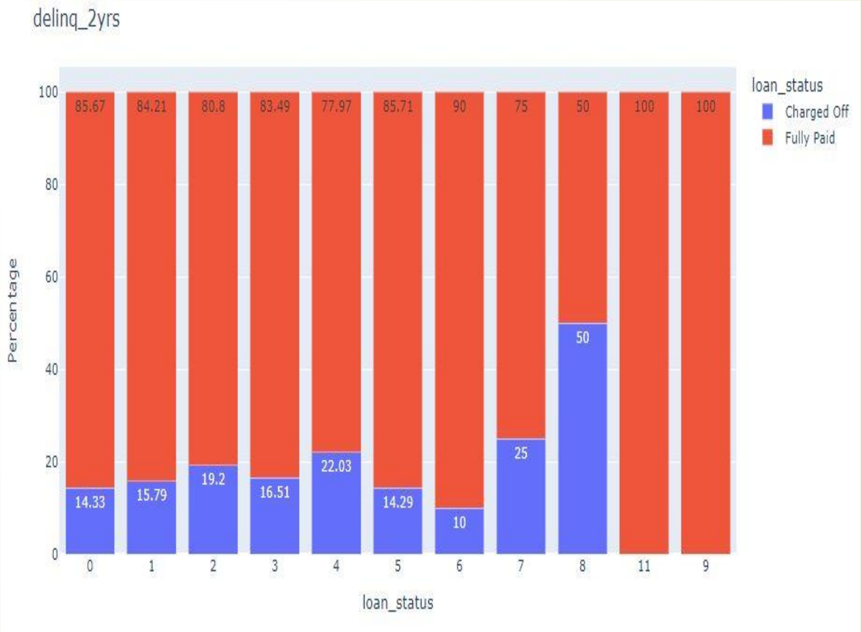




Results of Univariate

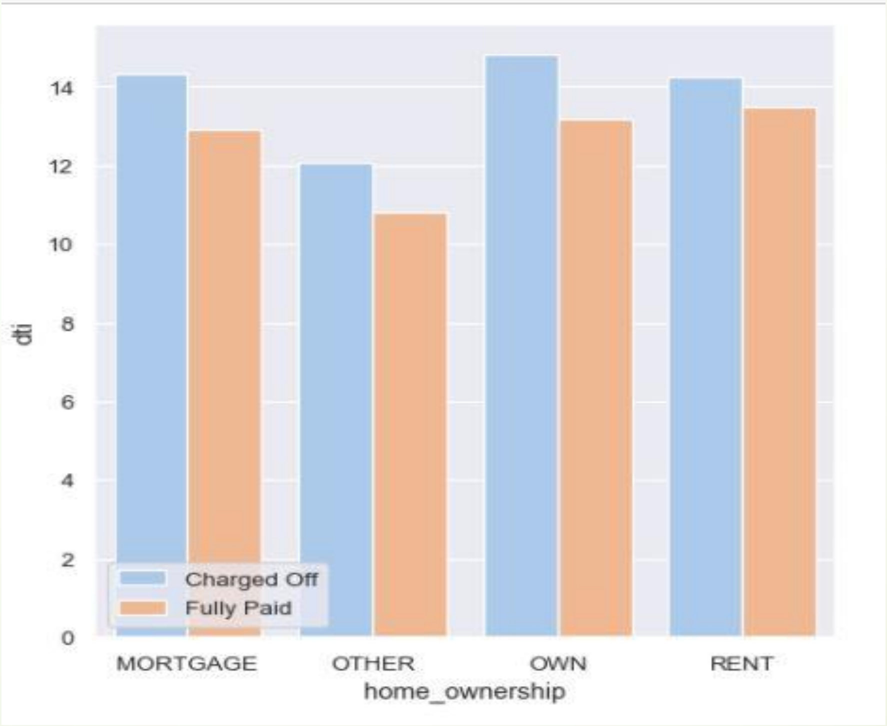
- delinq_2yrs : The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
  - In general, LC is not accepting the customers with higher count of 'delinq_2years'.
  - ('delinq_2years' - Denotes how many 30+ days past with the incidences of delinquency in last 2 years) The general trend is, as the count of variable increases, the percentage of charged off loans increases.
  - There is a definite chance of charging off loan when the delinq_2years is high. Additional checks are required to understand
  - the financial profile of the applicant in such cases.





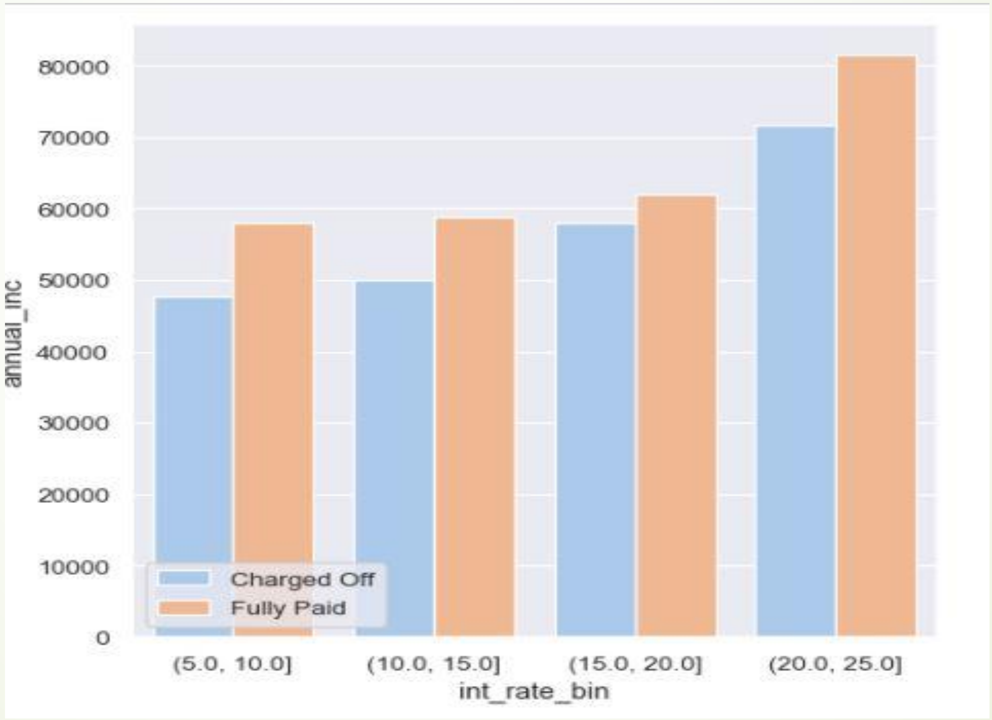
Results of Bivariate

- Anlysis of relation b/w homeownership, dti and loan status
  - People who has own house and DTI greater than 12 are more likely to get charged off.



Results of Bivariate

- Realtionship between DTI annual income and loan status
  - If annual income is greater than 70k and interest rate is 20% to 25%, then probability of getting charred off is higher.



Thank You