Lending Club Case Study

By Mrunmayee Dixit

Problem Statement

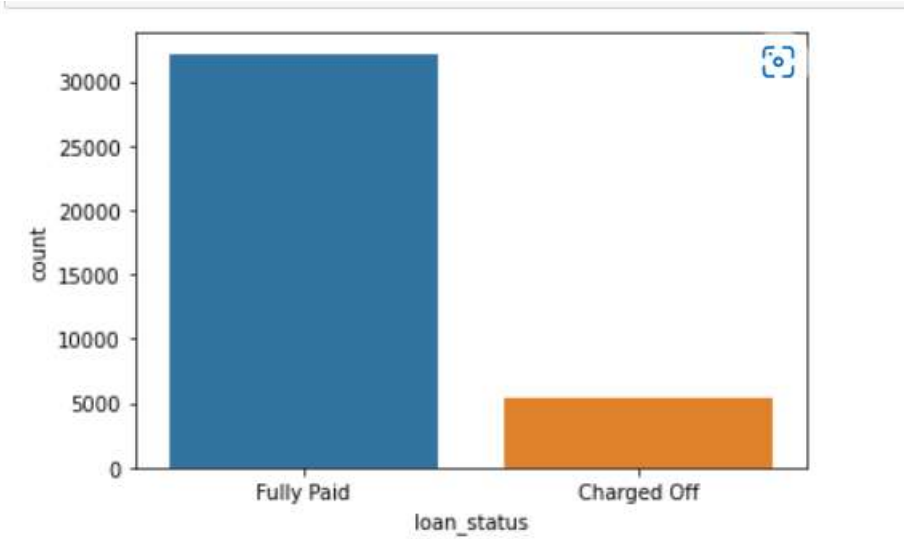
- We have the of the loan applicants for whom the loan was approved.
- We have to analyse the data and find the major reasons because of which the defaults occur.

Approach

- We first go though the complete data set
- Understand the metadata
- Do complete data cleaning process
- Then do Univariate analysis and bivariate analysis and draw conclusions from them

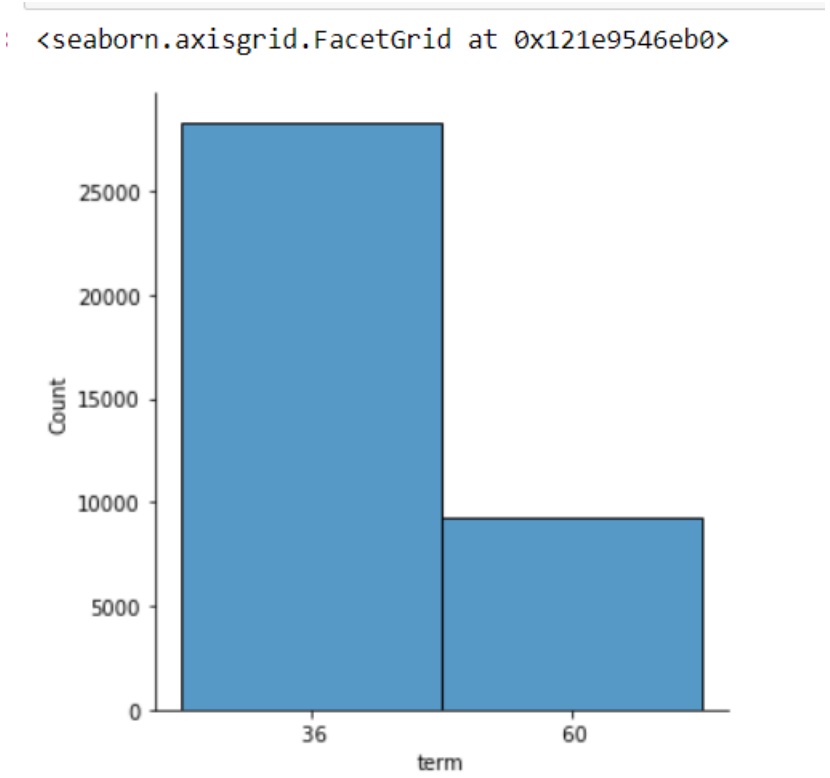
Univariate Analysis and Observations

- From above 'loan_status' plot we understand that there are around 14% of defaulters



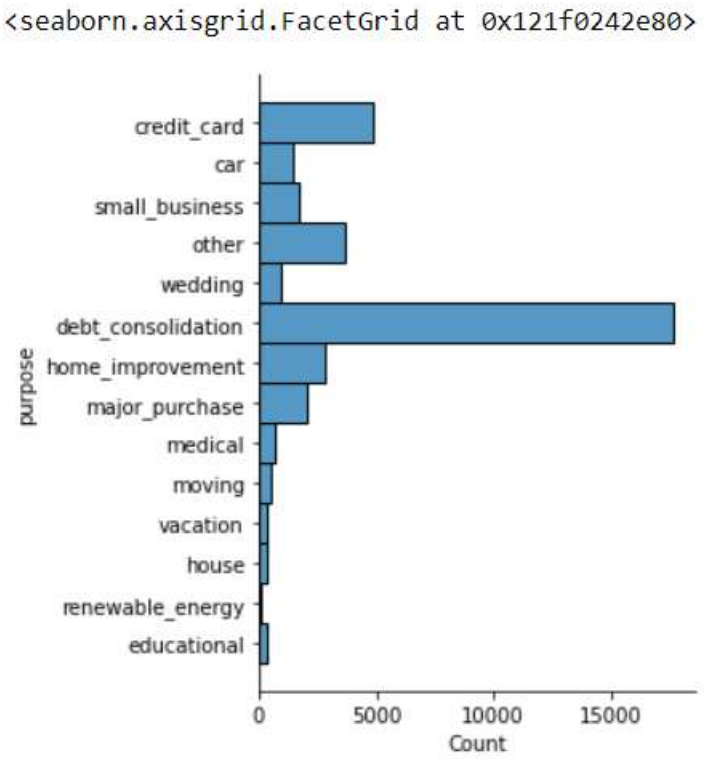
Univariate Analysis and Observations

- More number of loan applicants prefer to take short term loans than long term



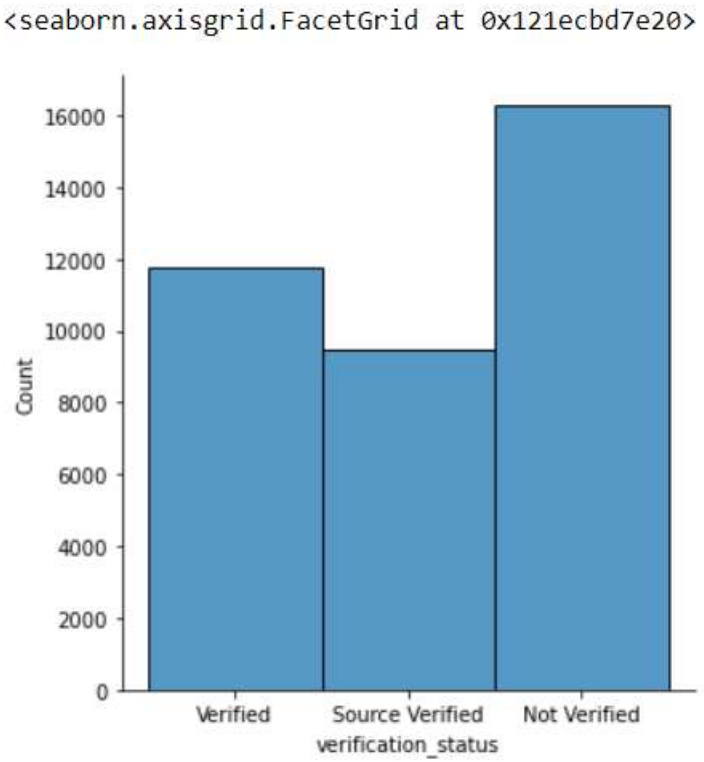
Univariate Analysis and Observations

- Maximum number of people take loan for dept_consolidation.



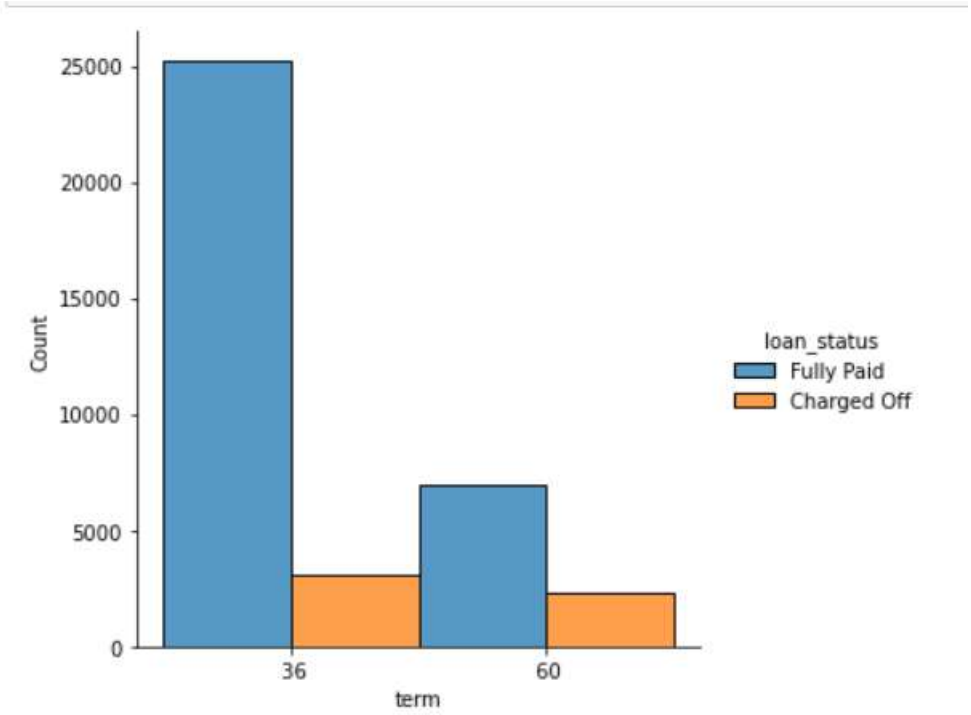
Univariate Analysis and Observations

- Only 25% loan applicants are source verified.



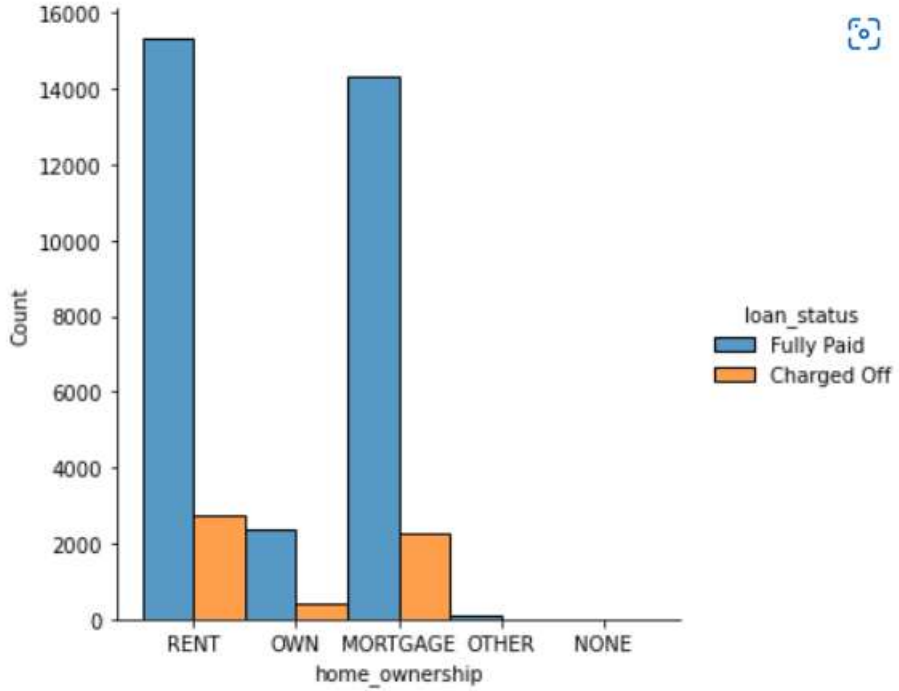
Bivariate Analysis and Observations

- Loan Applicants who have taken loan for longer duration are tent to become defaulters as compared to short term loan applicants



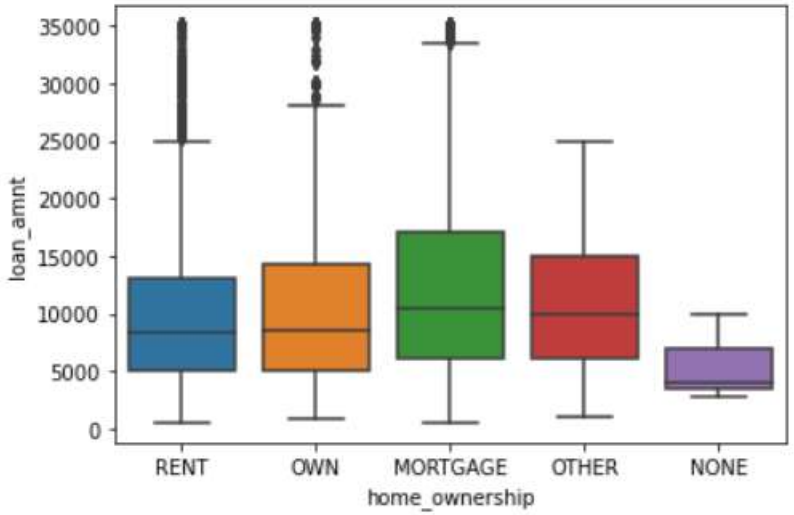
Bivariate Analysis and Observations

- Loan applicants are tend to default when they have rent the stay or they have taken loan for stay or property.



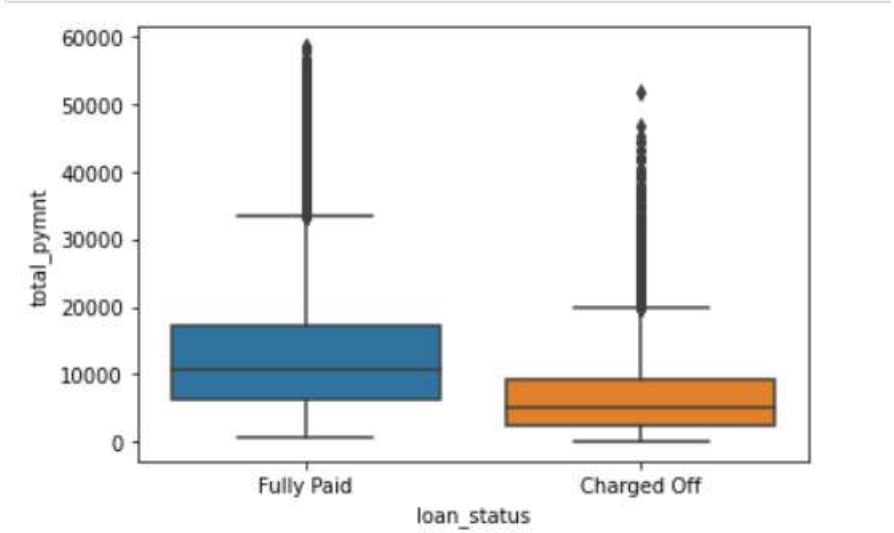
Bivariate Analysis and Observations

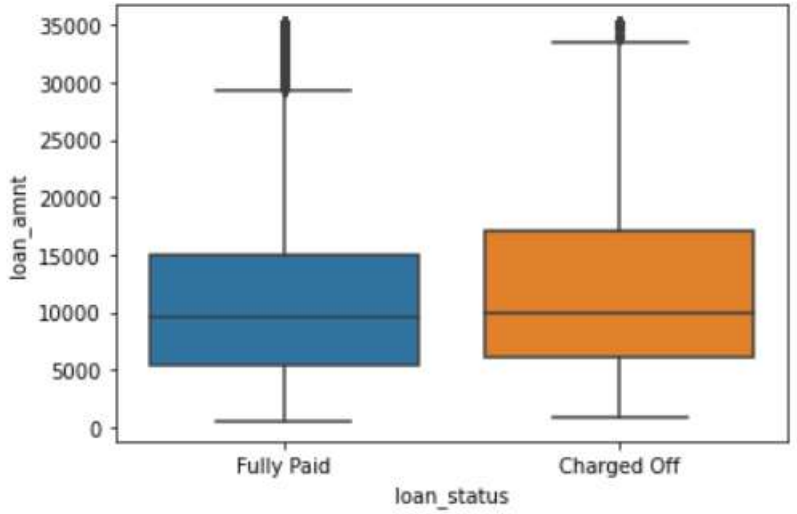
- Loan applicants who already have loan for home apply for higher range of loan amount



Bivariate Analysis and Observations

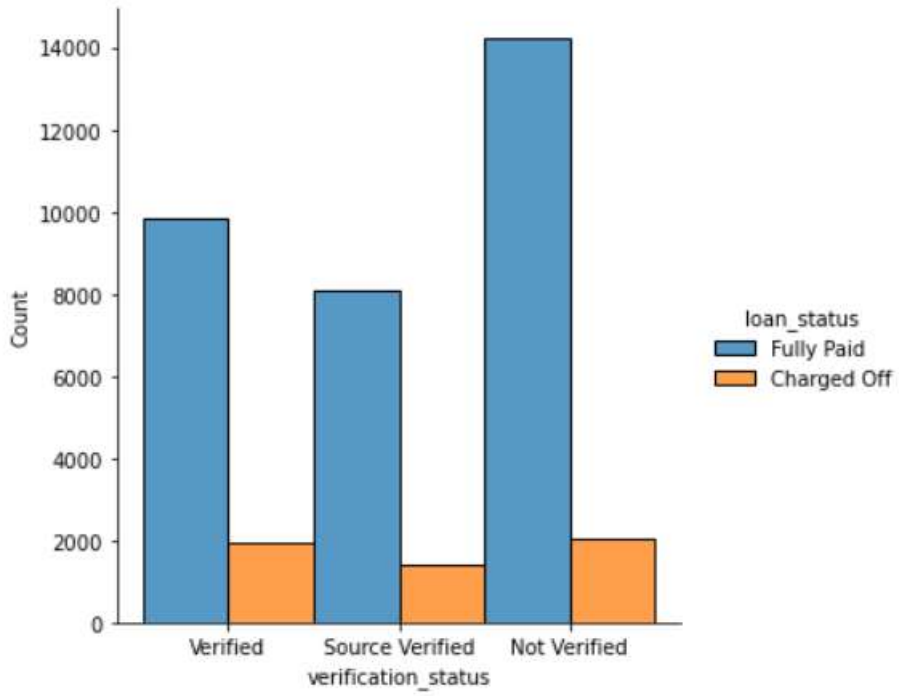
1. As we can see from above 2 plot, then we can say that 5000*5000(approx. number of defaulters * approx. mean of loan amt – approx. mean of total payment for charged off) i.e. mean loss for the company will be 25000000
2. maximum loss = approx. : 5000*7500 = 37500000
3. minimum loss = approx. 5000*2500 = 12500000
4. From above plot the applicants taken loan for higher loan amount tend to be defaulters





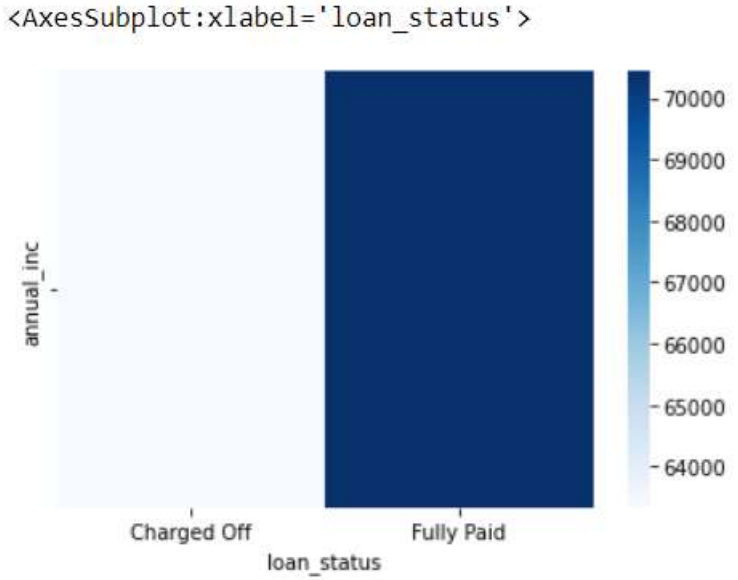
Bivariate Analysis and Observations

- If the loan applicant is source verified then chances of that person being defaulter is less.
- So, effort should be maid that all the applicants are source verified.



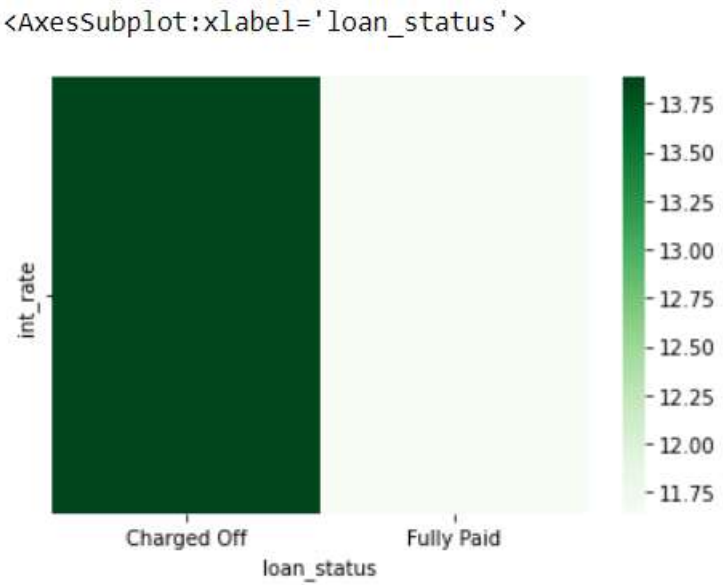
Bivariate Analysis and Observations

- Values = Annual Income
- Annual Income highly impact the defaulters
- Above plot shows that loan applicants with less annual income are the defaulters



Bivariate Analysis and Observations

- Values = Interest Rate
- High interest rate lead to more defaulters
- low interest rate result in less number of defaulters



Bivariate Analysis and Observations

- Loan applicants who have high 'dti' 1.e. above 14 percentage + Mortgage are defaulters.