

LENDING CLUB CASE STUDY

Prepared by:
Aman Vijaywargi
Ria Jennifer

Problem Statement

You work for a **consumer finance company** which specialises 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 company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Cont..

- When a person applies for a loan, there are **two types of decisions** that could be taken by the company:
- **Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below:
 - **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
- **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)

Objective of the analysis

- The data contains 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.
- In this case study, you will use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default.

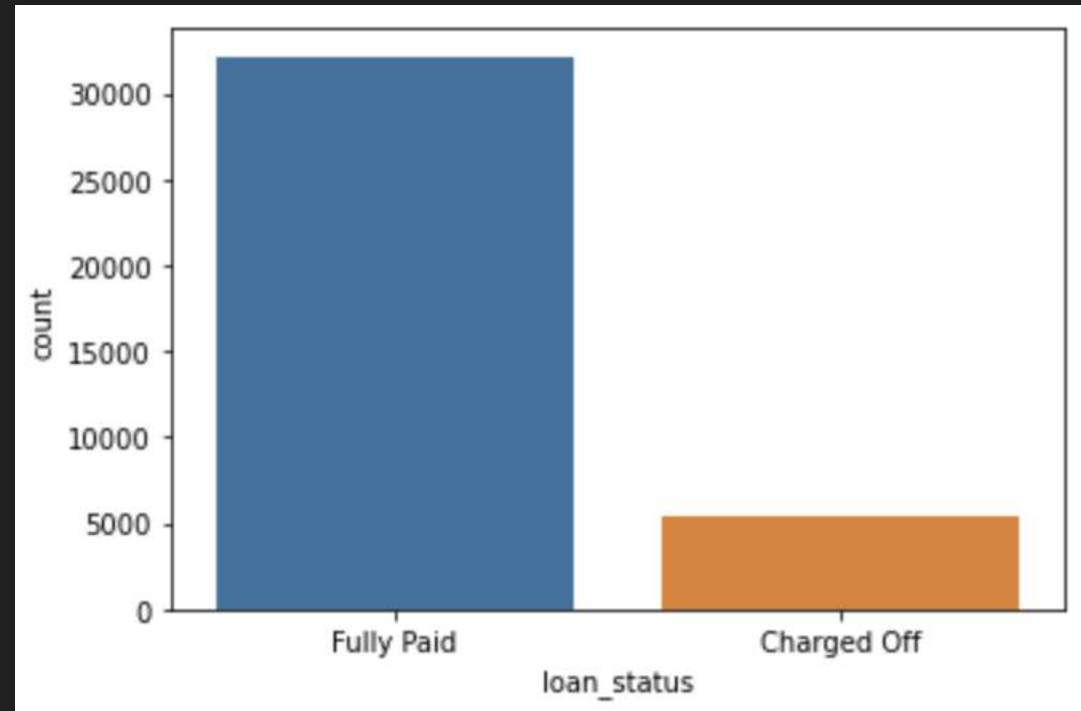
Tech info for analysing the test case

- Programming language: python
- Libraries: Numpy, Pandas, Matplotlib, Seaborn
- IDE used: Jupyter Notebook

Points followed to do analysis

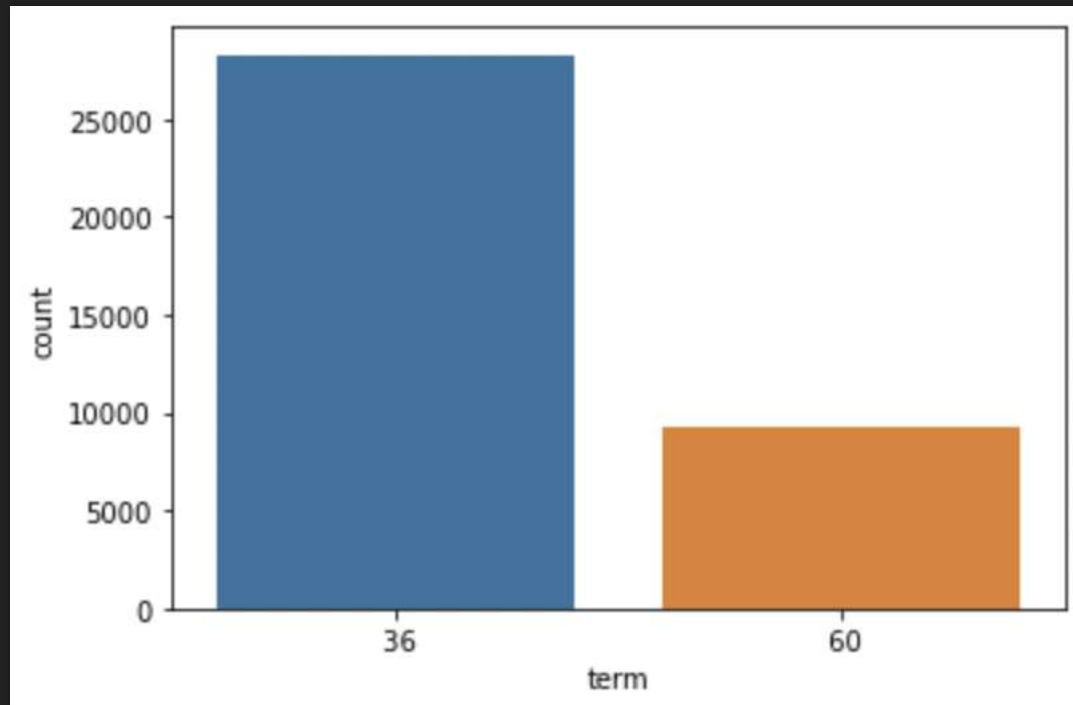
- Read the dataset and understand according to the column names.
- Find out any relation with the columns
- Begin with the data cleaning process i.e. removing rows and columns which are irrelevant for analysis, missing values, and if missing value percentage for a particular column is less add mean/mode accordingly.
- After cleaning, go for univariate and bivariate analysis

Univariate analysis



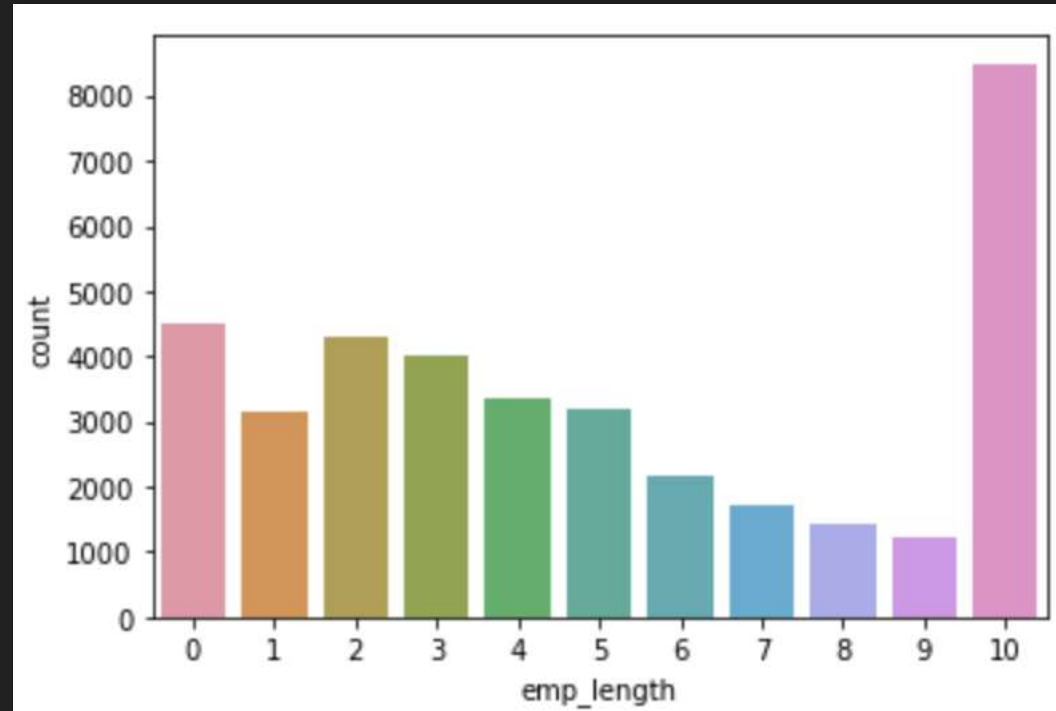
In the above diagram, the loan status data is displayed which shows there are around 14% defaulters among fully paid and defaulters

Univariate analysis



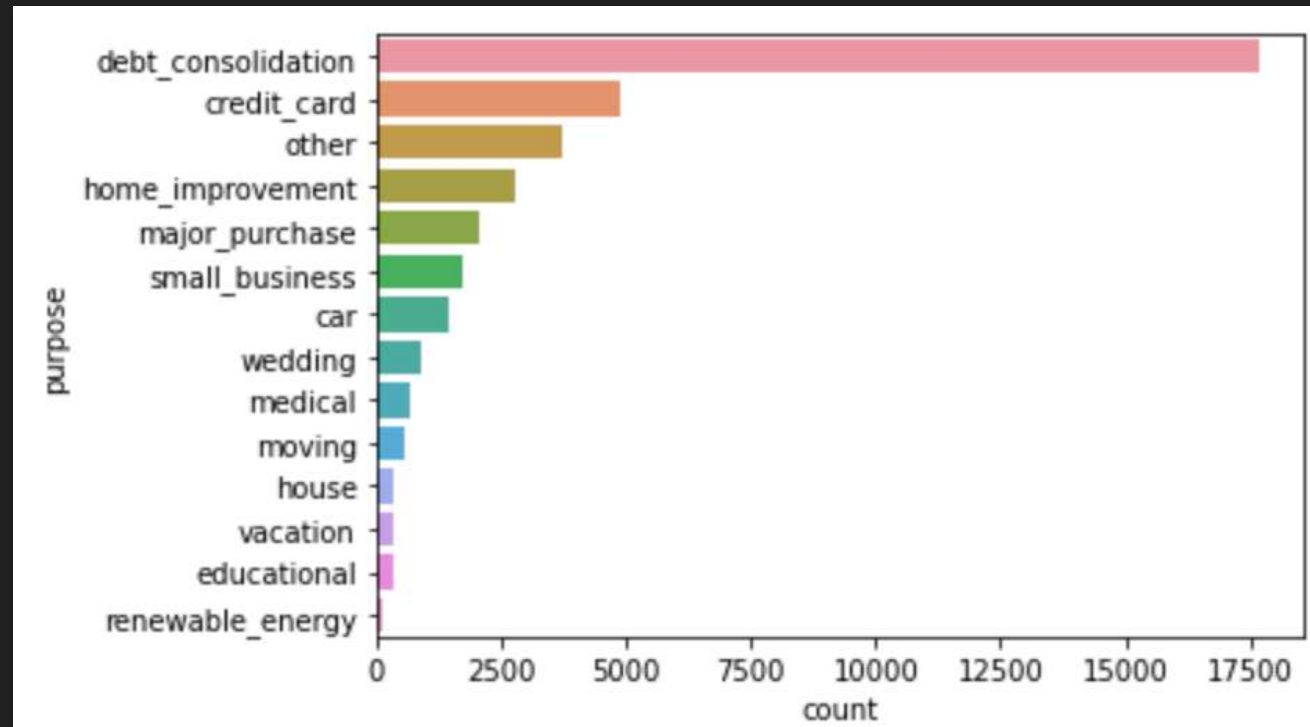
Short term period loans are taken more than long number of period

Univariate analysis



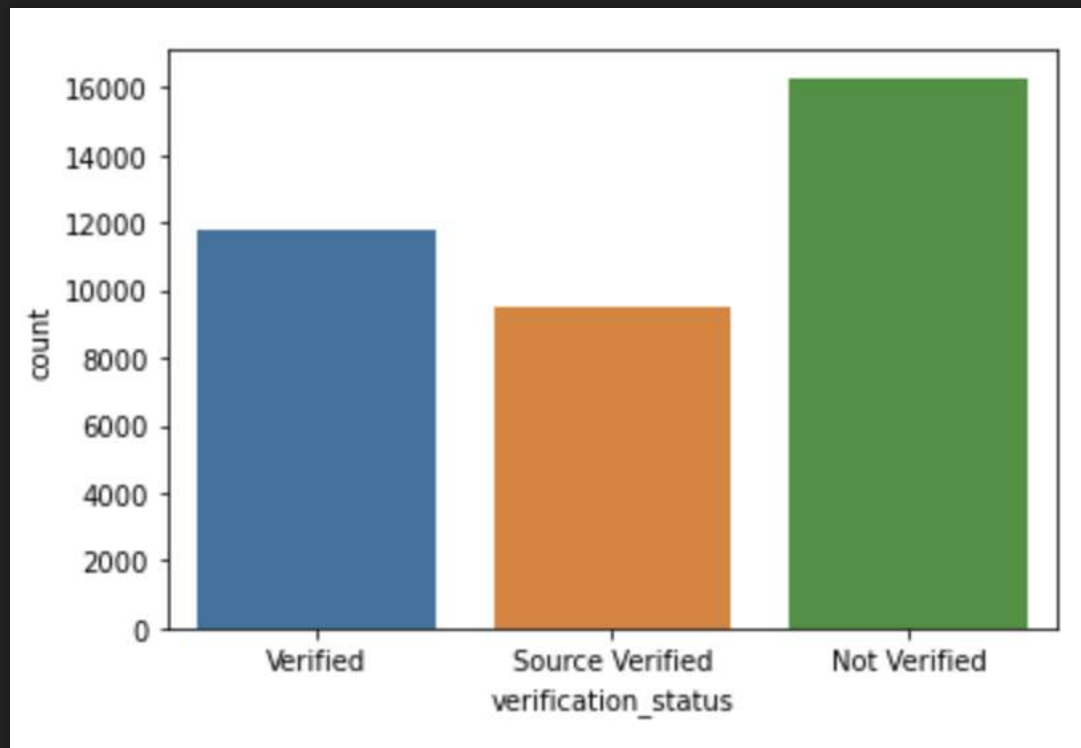
Maximum number of people take loan for dept_consolidation.

Univariate analysis



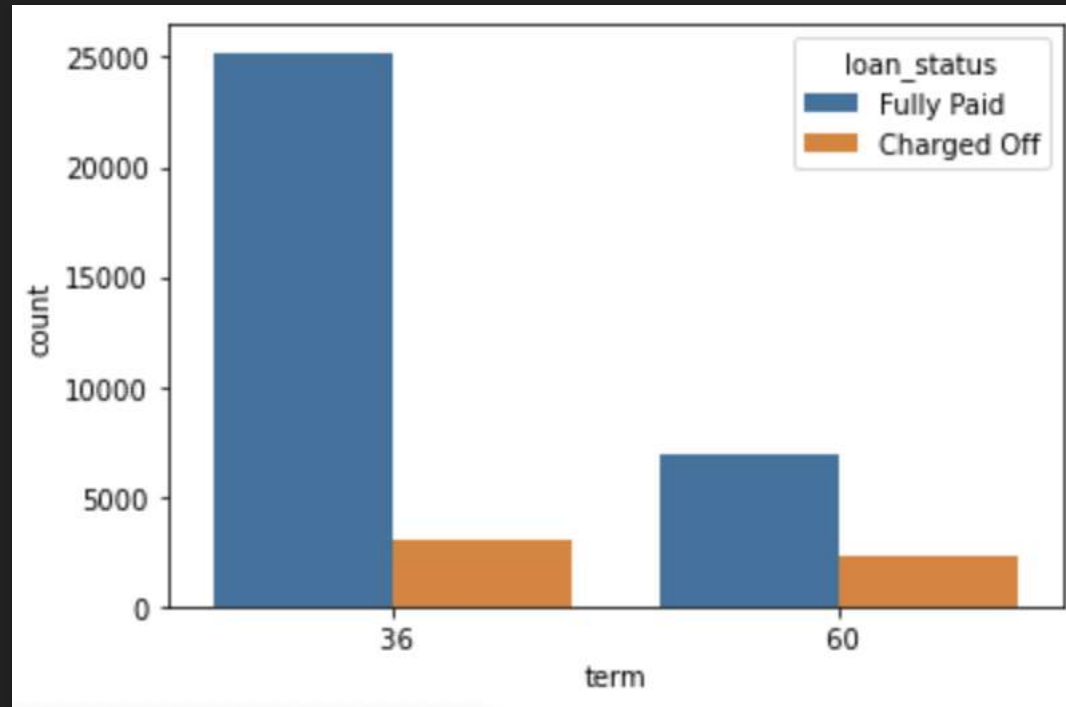
Maximum number of people take loan for dept_consolidation.

Univariate analysis



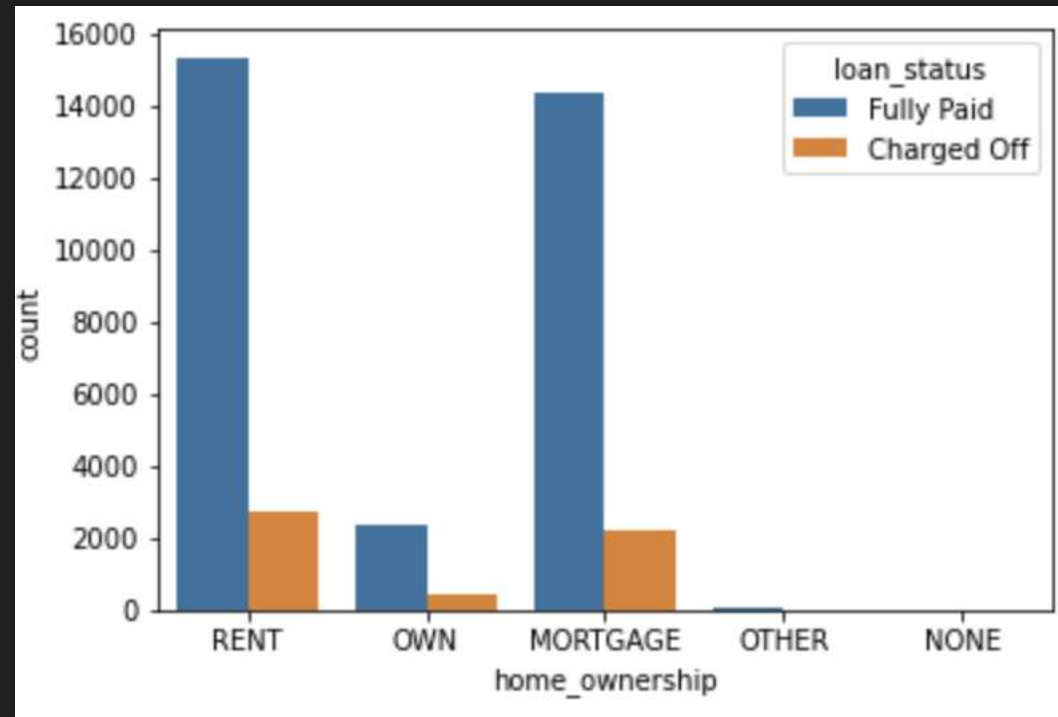
Only 25% loan applicants are source verified.

Bivariate analysis



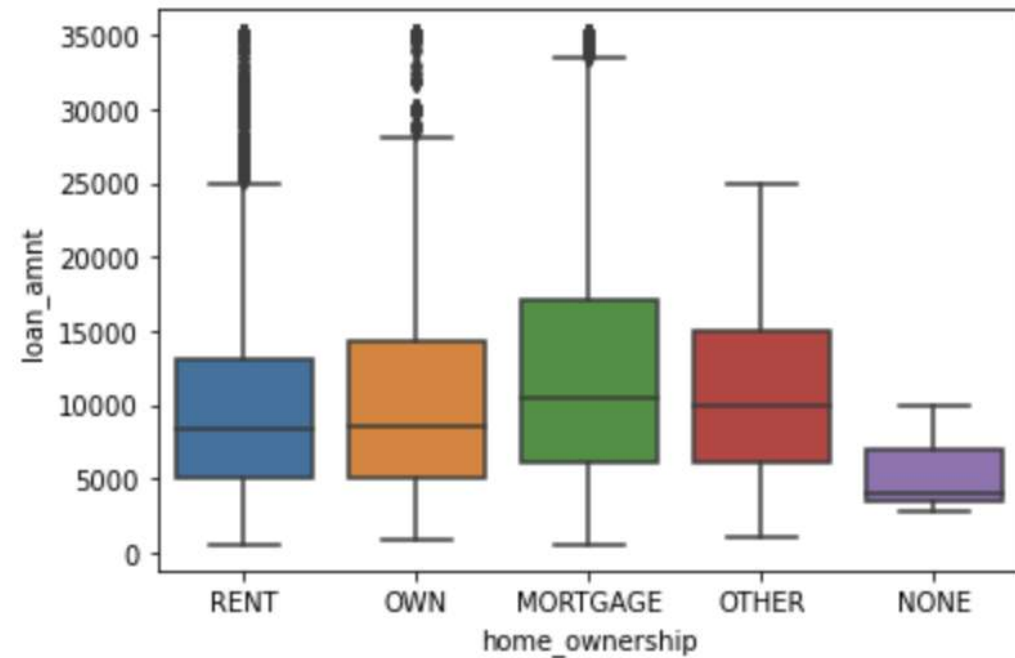
Loan Applicants who have taken loan for longer duration are tent to become defaulters as compared to short term loan applicants

Bivariate analysis



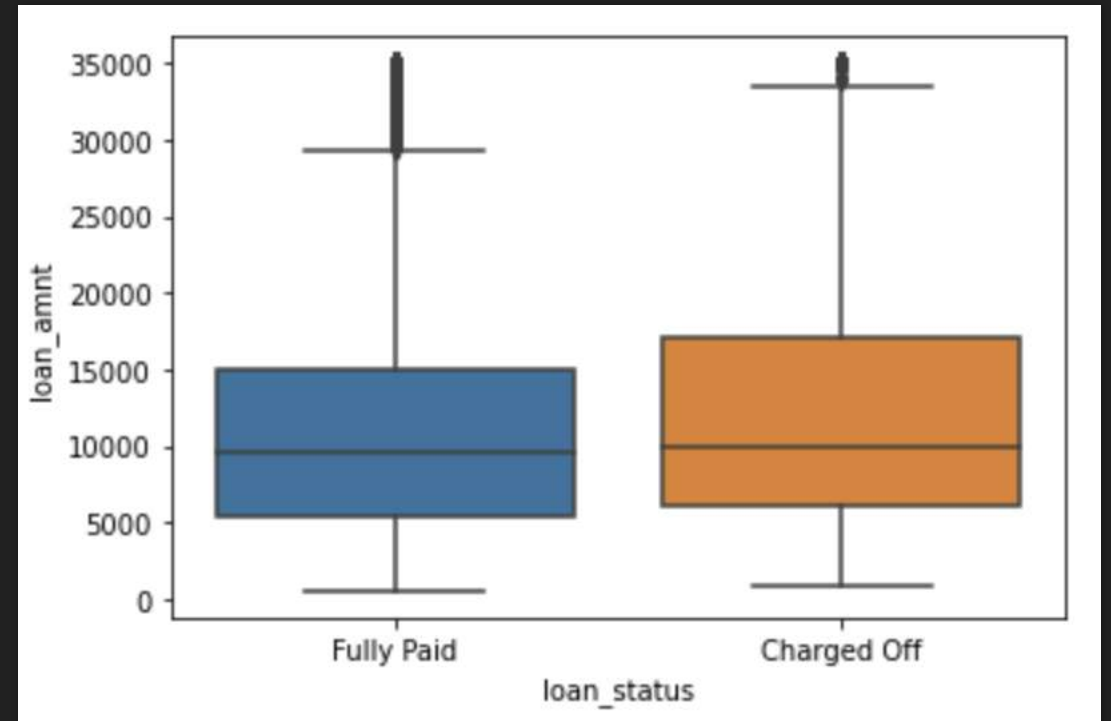
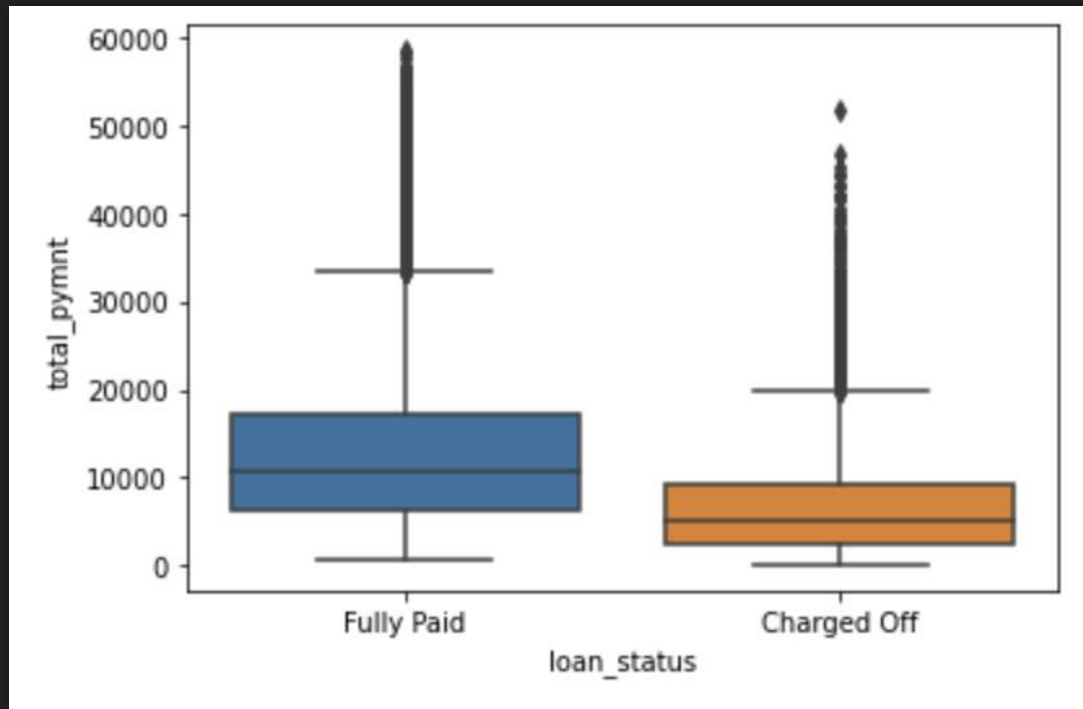
Loan applicants are tend to default when they have rent the stay or they have taken loan for stay or property.

Bivariate analysis



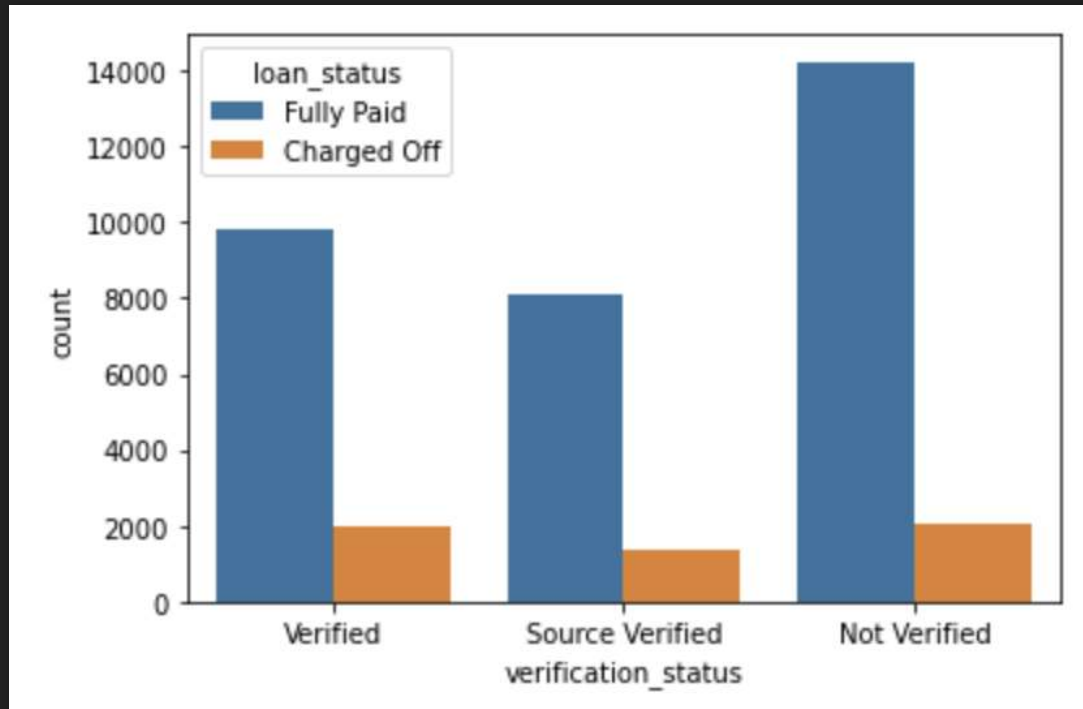
Loan applicants who already have loan for home apply for higher range of loan amount

Bivariate analysis



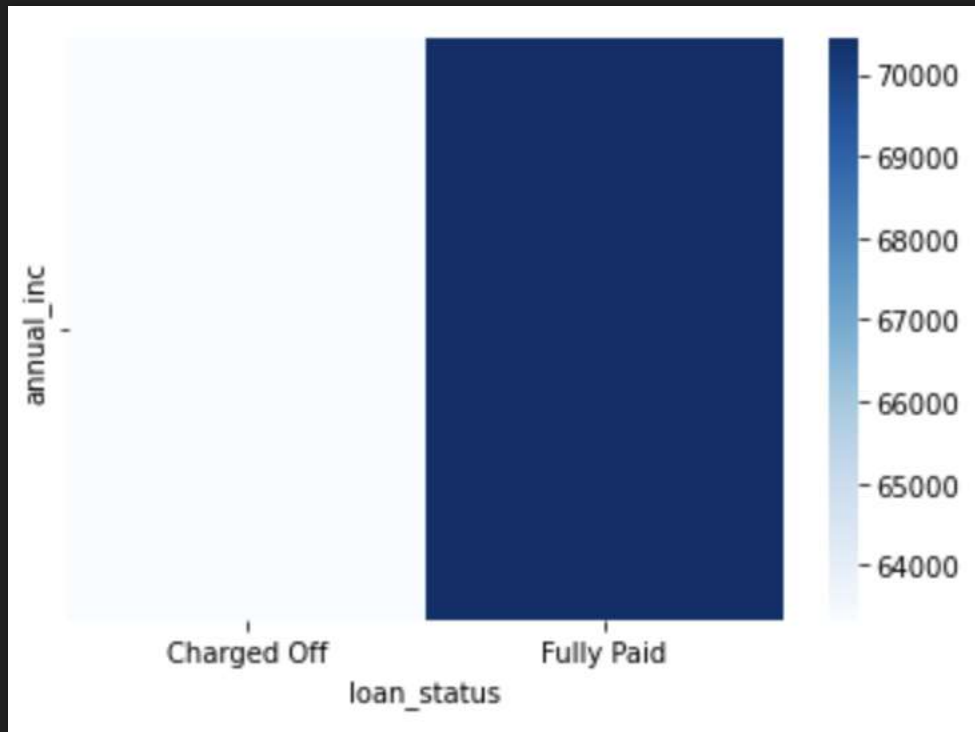
From above plot the applicants taken loan for higher loan amount tend to be defaulters

Bivariate analysis



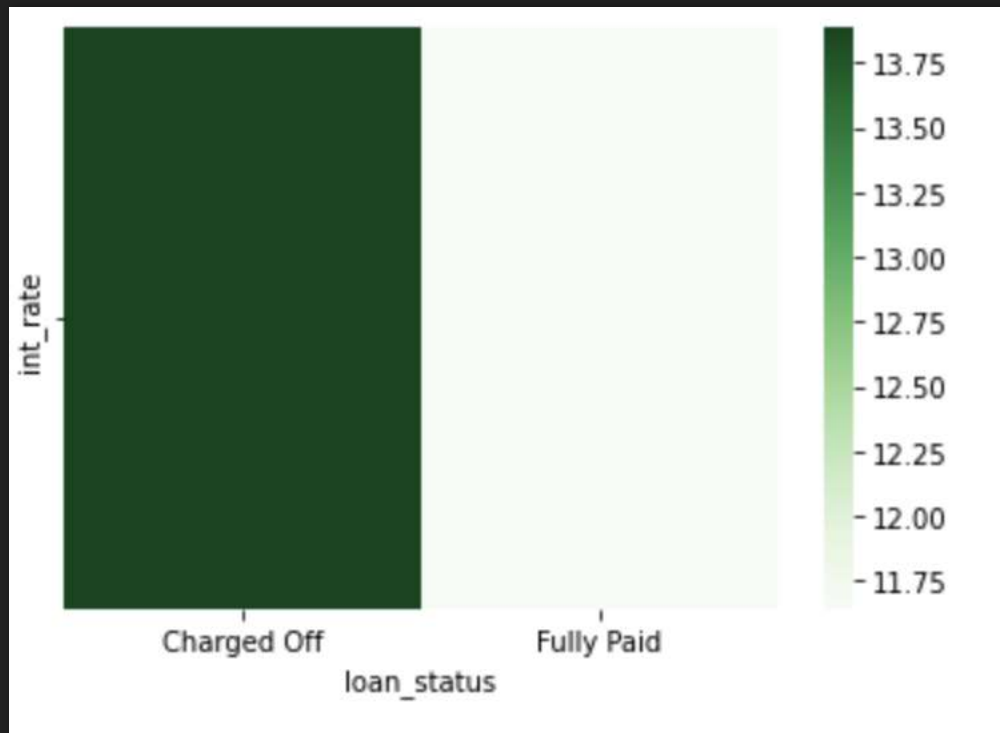
- If the loan applicant is source verified then chances of that person being defaulter is less.
- So, effort should be made that all the applicants are source verified.

Bivariate analysis



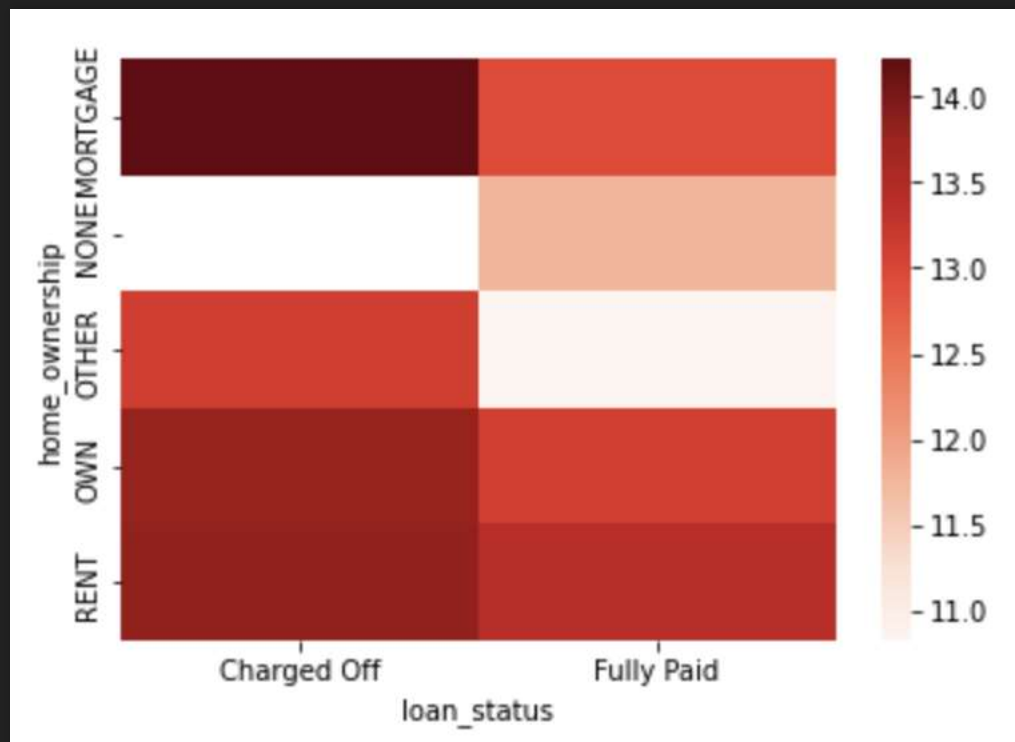
- Values = Annual Income
- Annual Income highly impact the defaulters
- Above plot shows that loan applicants with less annual income are the defaulters

Bivariate analysis



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

Bivariate analysis



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

CONCLUSION

- There were around 14% of loan defaulters
- People have taken short term loans more than long term loans
- The loans which were taken were majorly the people who already had loans and took to repay other loans
- Only 25% loan applicants are source verified
- If the interest rate is high, there will be high chance for user to become loan defaulter