Lending Club Case Study

Ashutosh Dutondkar

Srinivas Gopal S

Problem statement

- This is the largest online loan marketplace, providing multiple types of loans and a fast online interface which has easy access to lower interest rates.
- Lending loan to potential defaulters is the largest source of credit loss for the company.
- Identifying such applicants using EDA, thereby reducing the number of risky loan applications, is the goal of this case study.

Approach

<u>Following steps were undertaken during the analysis:</u>

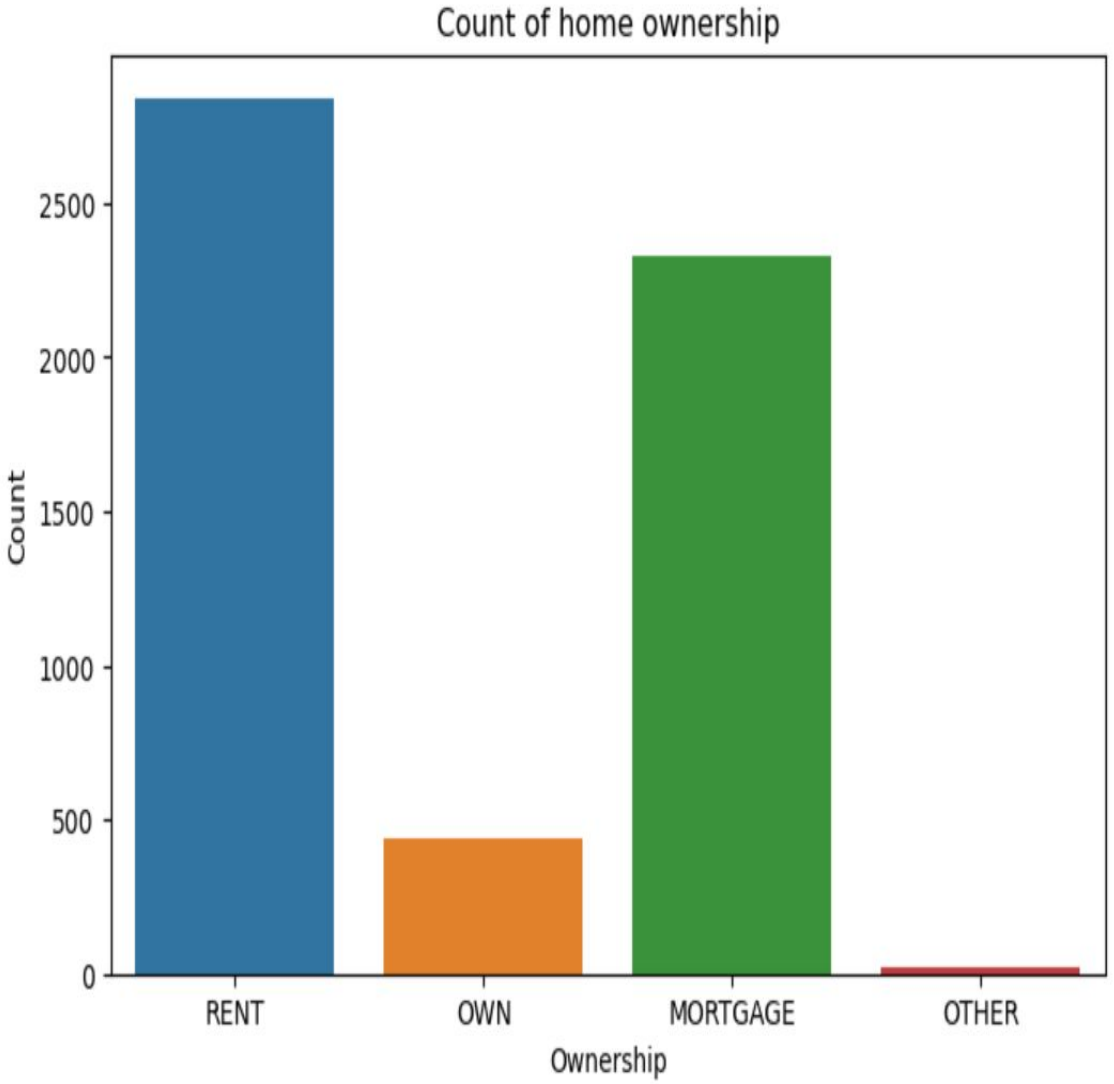
- Data cleaning and manipulation
- Univariate analysis & Segmented Univariate Analysis
- Bivariate analysis

Data cleaning and manipulation

Few of the important steps involved in data cleaning and data manipulation are listed below:

- Check for missing values
- Remove columns that have unusable values (e.g: NA)
- Convert columns to proper data type
  - Convert strings with “%” to float and dates with “months” to date formats
  - Standardize date format across all columns
- Assign meaningful and usable integers to string entries such as ‘NaN’
- At the end of the process, data had 39717 entries, each having 48 columns

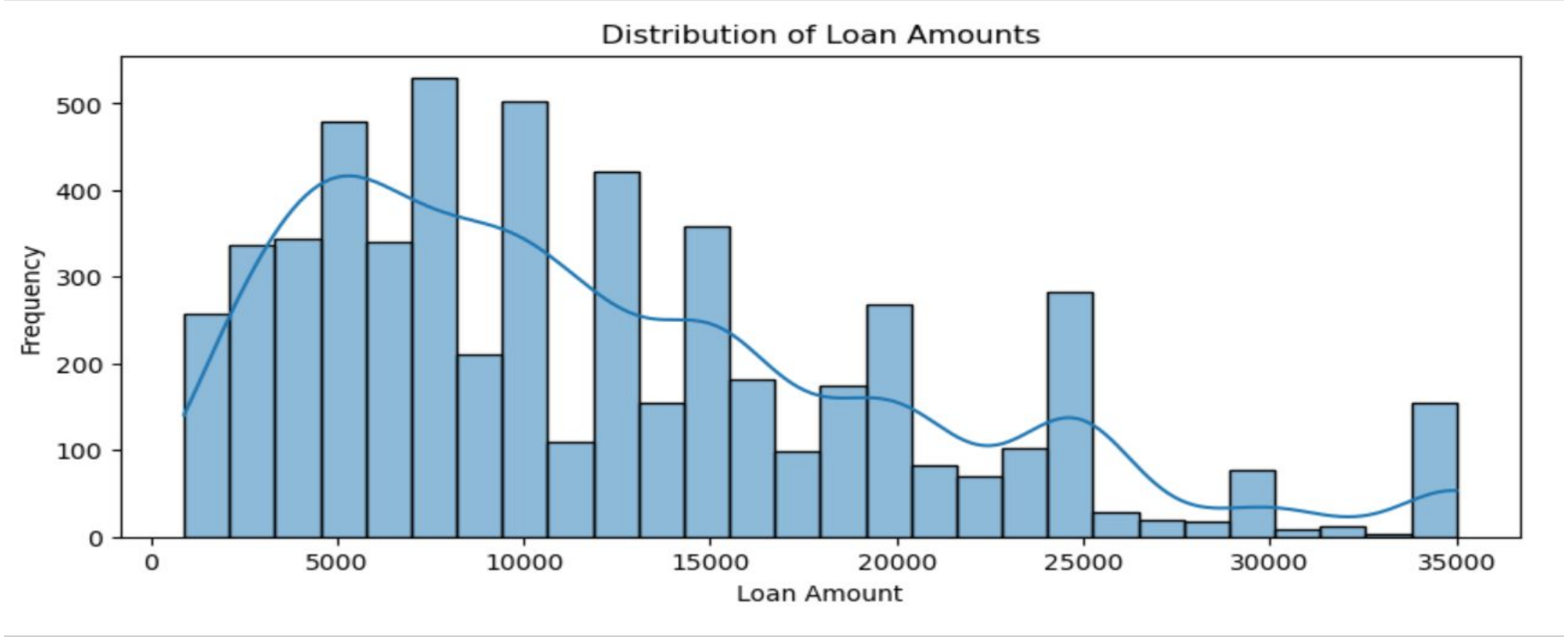
Univariate Analysis



Univariate analysis performed on **home_ownership** column of charged off customers revealed that most of them are either having a mortgage or are living on rent.

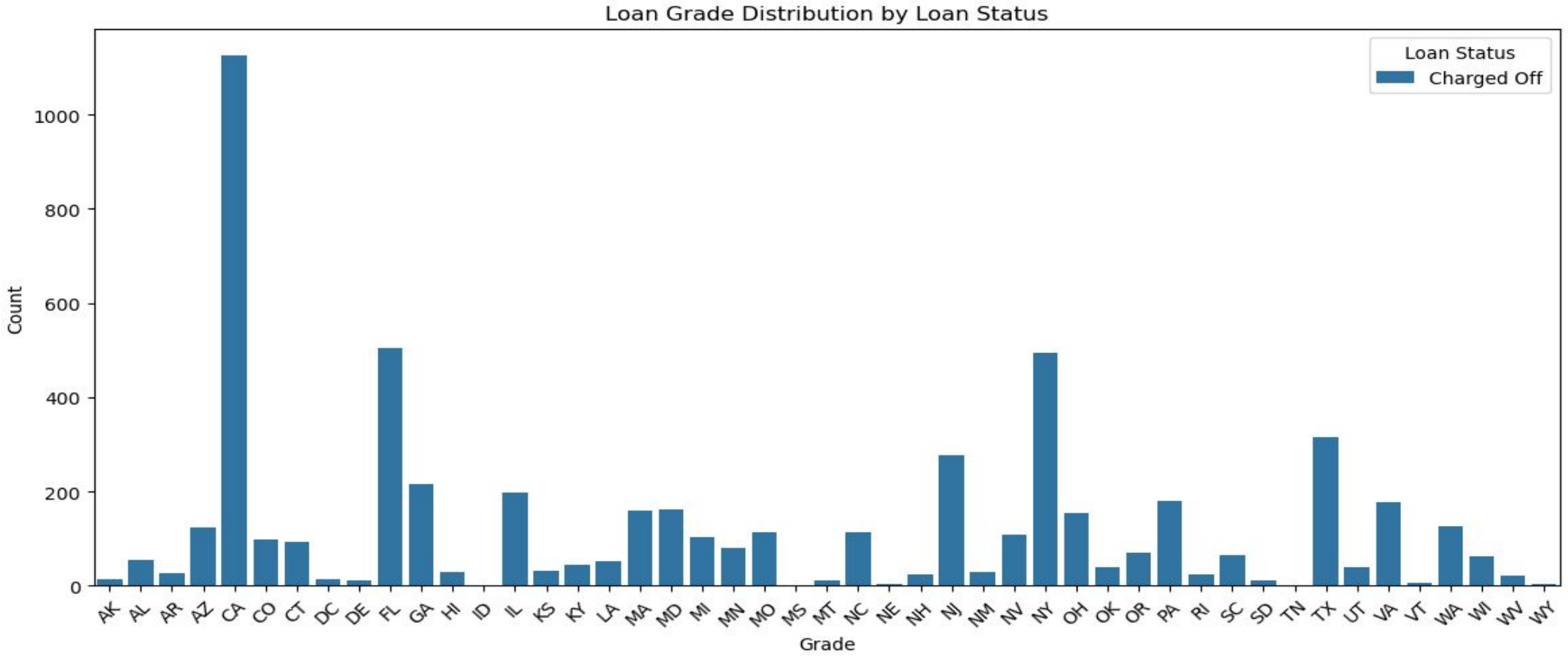
- More than 3500+ customers don't OWN a home
- With this we can say that the employee can do some extra verification for such consumers.

Univariate Analysis contd…



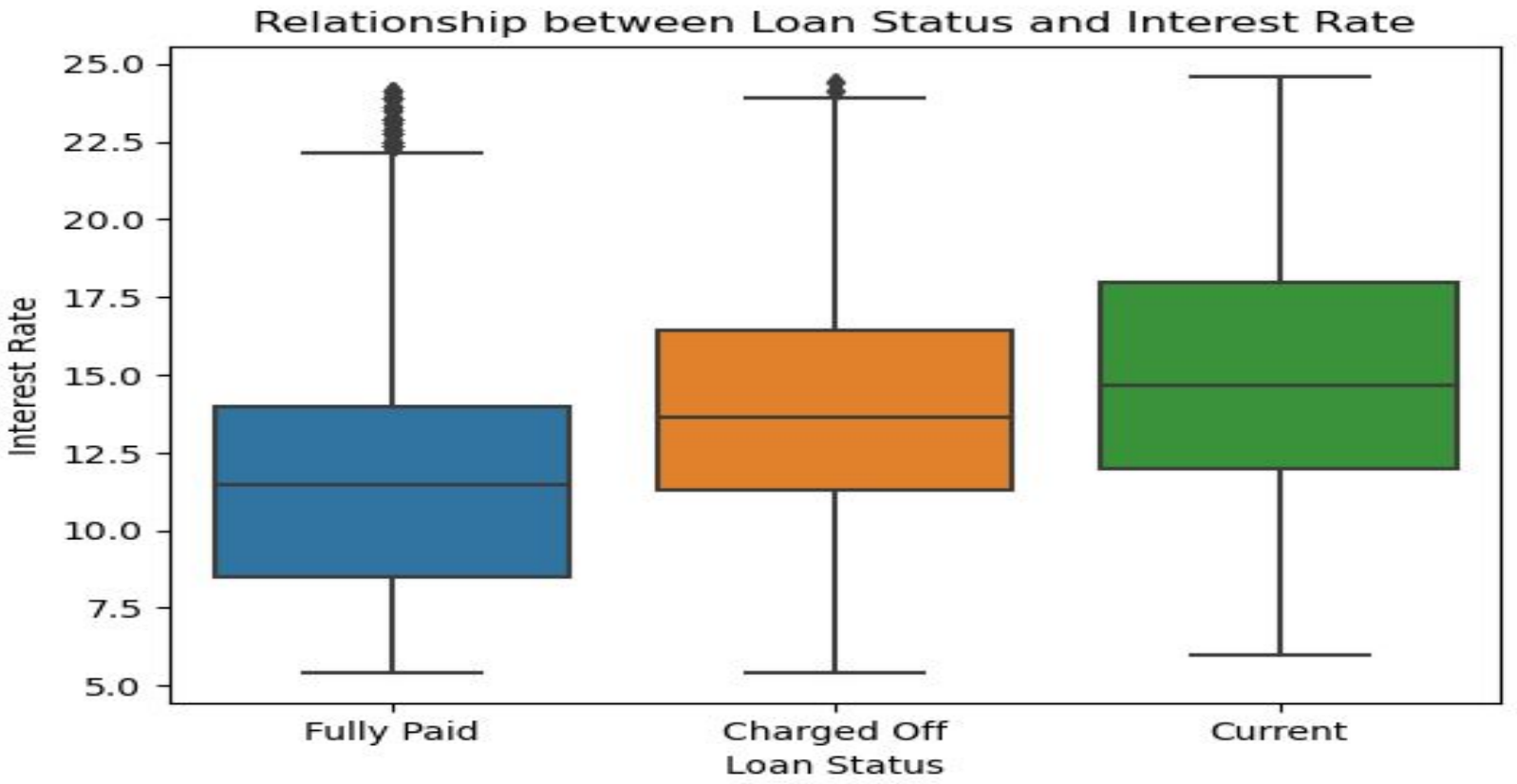
- From the analysis of disbursed loan amount, it was observed that the principal amount borrowed by defaulters were typically in the range of 5000 to 15000.
- We can ask the consumers more details for why they are taking the loan of such amount with some proofs required.

Univariate Analysis contd…



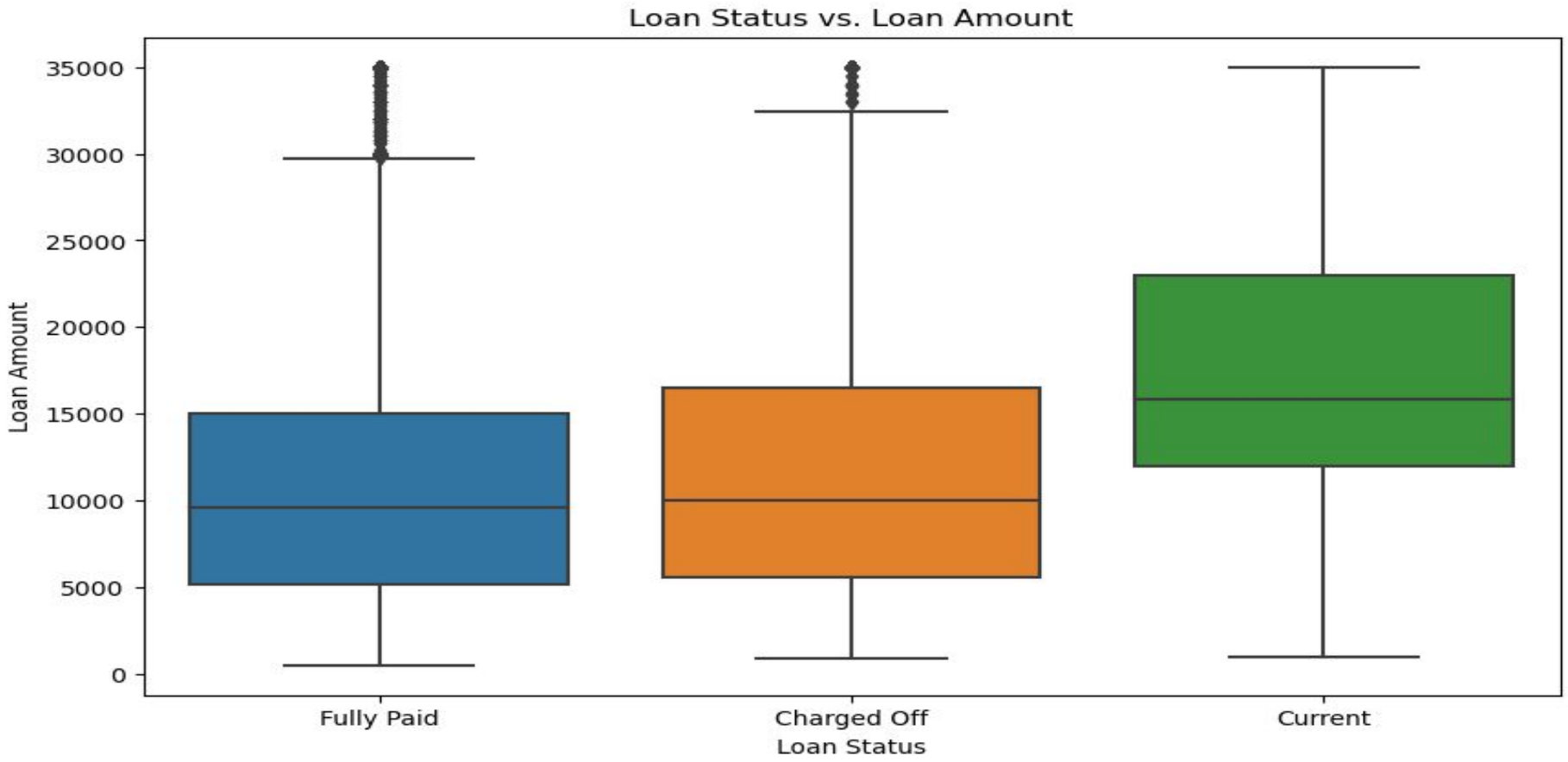
- From the above graph we can observe that California has the highest count of 1125 followed by Florida and New York.
- So when we have an application from the above states we need to do more detailed verification of the consumer & his purpose for taking the loan.

Bivariate Analysis



- We can observe that the median of fully paid loans, lower quartile of charged-off loans and lower quartile of ongoing loans have similar interest rate of >10.0% & <15.0%.
- So we can say the consumers prefer to have interest rates in the range of 10.0% to 15.0% you might be defaulter in the future.

Bivariate Analysis contd…



- In the above graph, we can observe that the lower quartile & median for closed loans and charged-off loans lie in the range of 5000 to 10000 while the median & upper quartile lie in the range of 10000 to 15000.
- Consumers tends to taking loan of such small amount might be defaulter in the future.

Conclusion

Based on data analysis performed previously, we can infer that additional verification needs to be performed under following conditions:

- Applicant does not own a home of his own
- Loan amount is between 5000 and 15000
- Applicant lives in the states of California, Florida or New York
- Applicants are ready for the interest rate of 10.0% to 15.0%

From the above data analysis we can say that the applicants with the above conditions should have more verification before processing the loan amount.
We can check for applicants history through his credit score.
We can have additional verification about his history.