



Lending Club Case Study

Nitish Sharma

Chetna Priyadarshini



Problem Statement

- Help a finance company identify risky loan applicants
- Understand consumer attributes and loan attributes that are likely to cause the account to be defaulted
- To serve above
 - Clean data removing null columns, replace NA with 0 where applicable, identify and remove rows that do not have any important information
 - Study distribution of data on various rows and perform outlier treatment
 - Understand correlation between different variables in people who defaulted on the loan and identify a common pattern
- Combine the rows together to arrive at a conclusion on risk factors for loan defaulters

Data Cleaning

Initial data: 39717 rows and 111 columns

Removed null columns: 39717 rows and 57 columns

Removed columns with only one value: 39717 rows and 51 columns

Removed columns which have more than 20% null values: 39717 rows and 47 columns

Removed columns with id/membership_id(all unique vals): 39717 rows and 45 columns

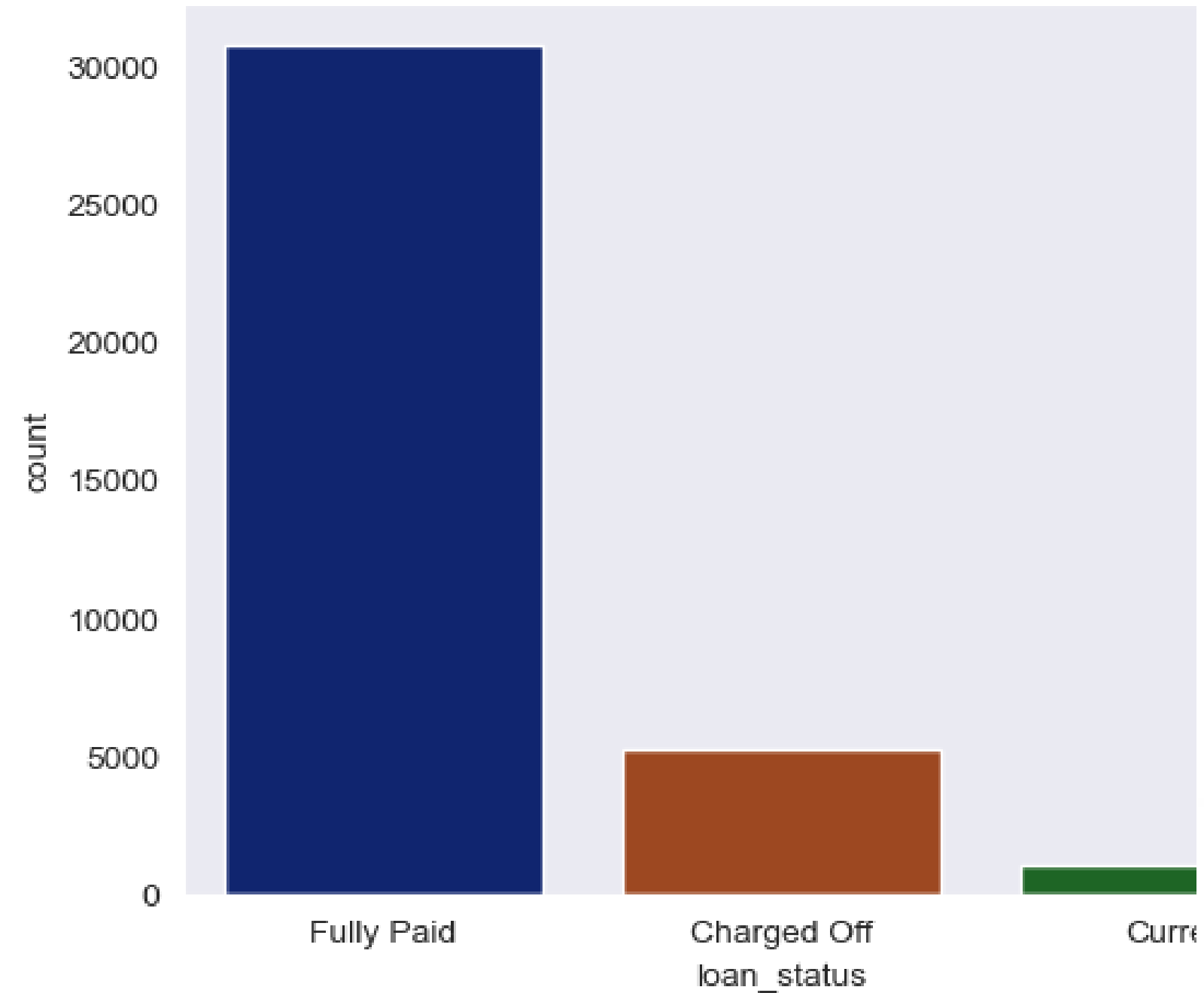
Removing columns not interesting for analysis: 39717 rows, 23 columns

Replace remaining null values with 0

Standardize column values removing trailing string of %, months, x, converting to numeric where possible

Remove outliers from Loan Amount, Annual Income, Debt to Income ratio columns

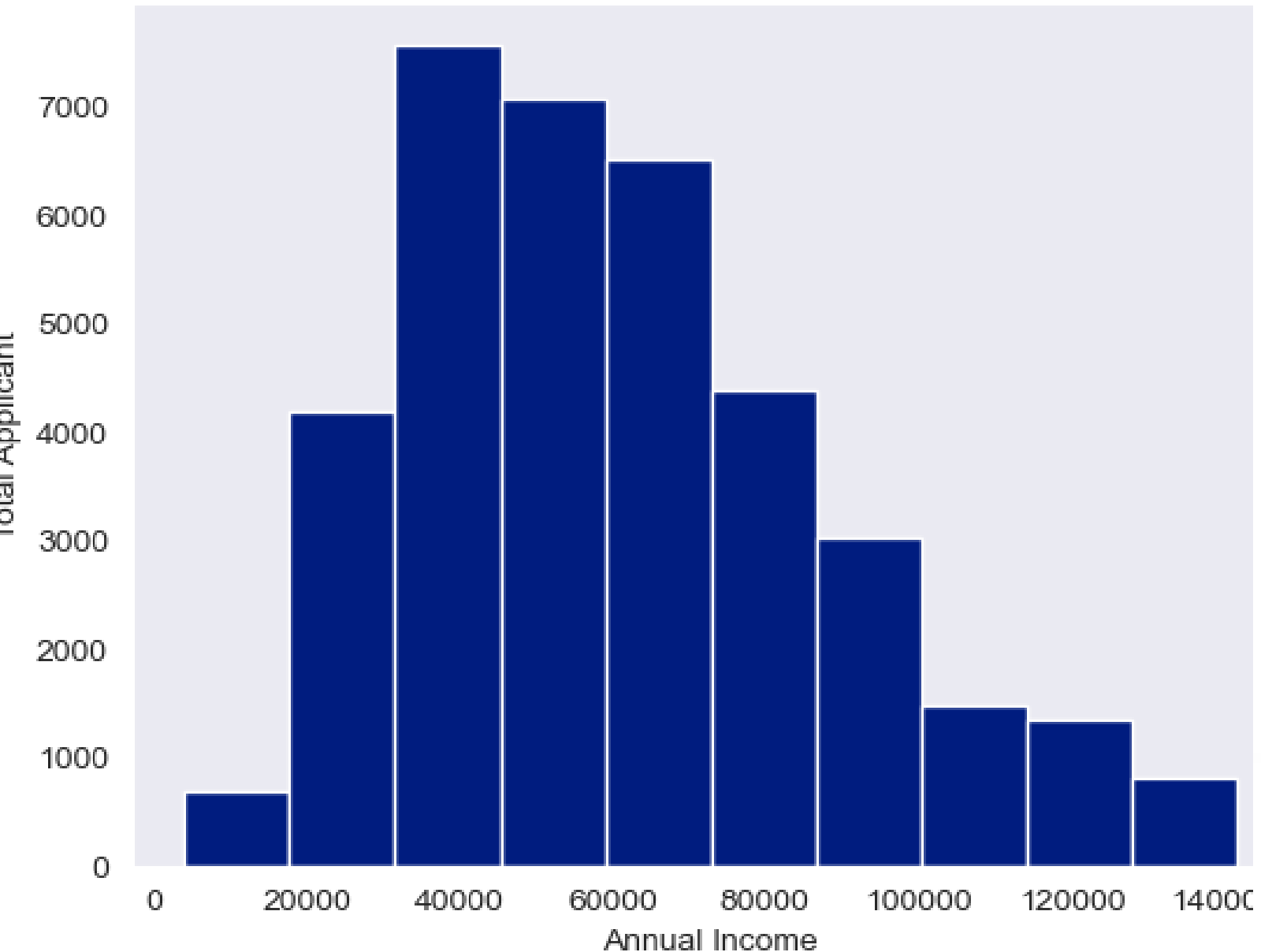
Loan Status



Univariate Analysis: Loan Status

We can see that most of the users in Fully paid category

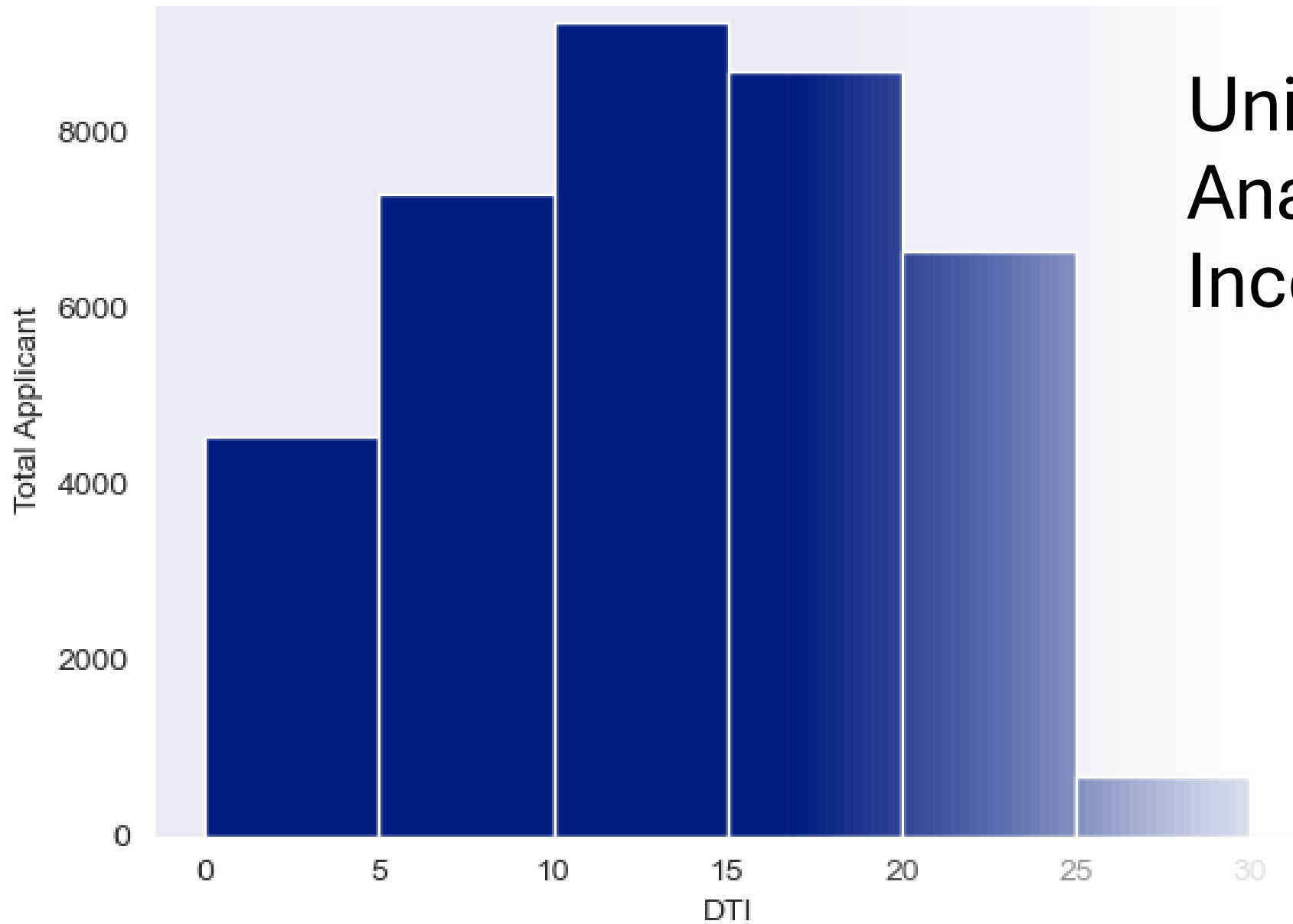
Annual Income of the Applicants



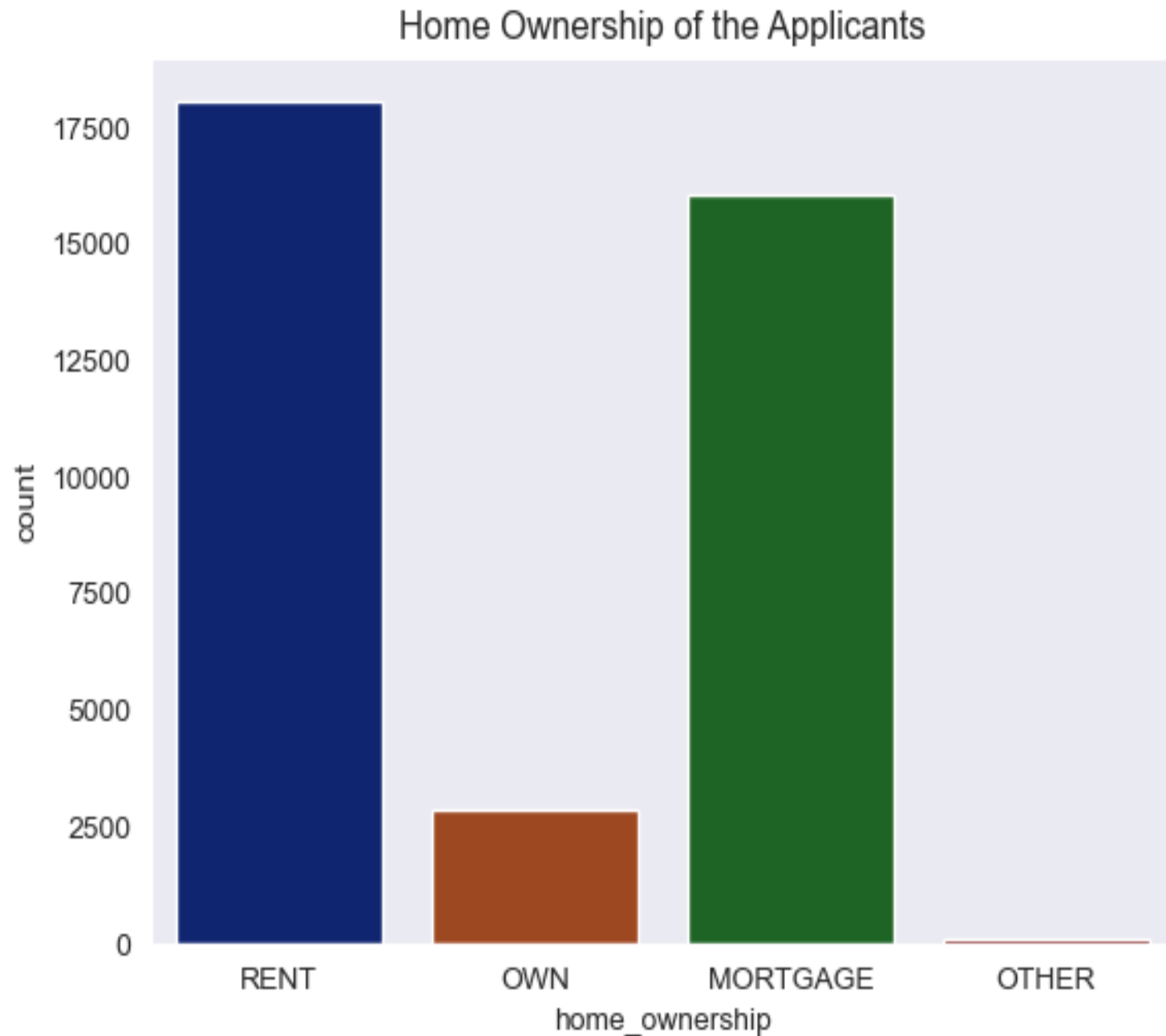
Univariate Analysis: Annual Income

- maximum number of loan applicants are having a annual income of 30000 - 40000.
- 4500 people have salary less than 30000.

Univariate Analysis: Debt to Income



- Composition of highest applicants with top 3 DTI falls under the categories as 10-15, 15-20 and 5-10 respectively.

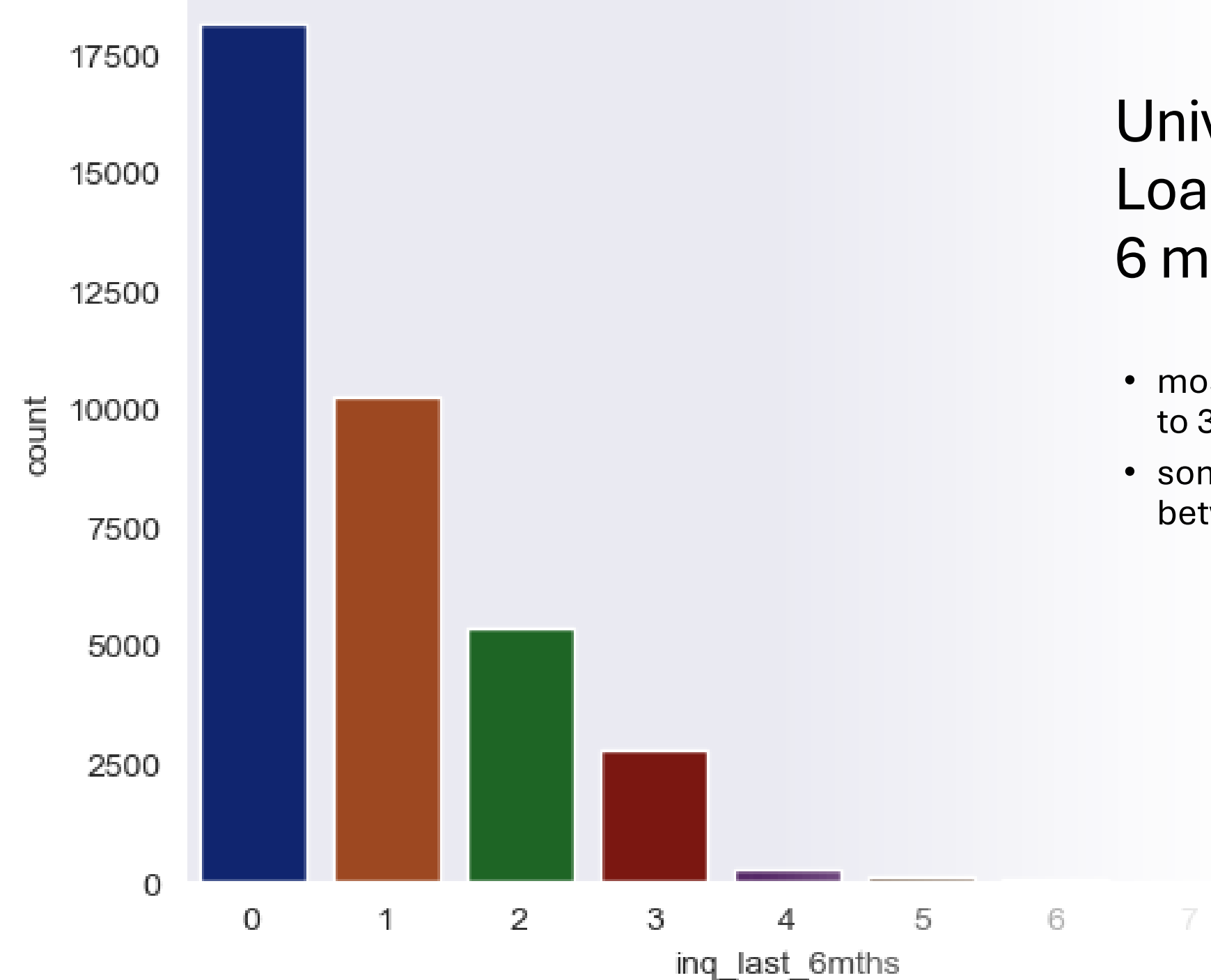


Univariate Analysis: Home Ownership

most of the applicants are either living on rent or property is mortgage.

Univariate Analysis: Loan enquiry in last 6 months

- most of the applicants have 0 to 3 enquiries.
- some applicants with enquiry between 4 to 8



Univariate Analysis : Loan Term

