Lending Club Case Study

Learned and applied the EDA technique, also develop a basic understanding of risk analytics in banking and financial services

Data Analysis

- Initially data set is having 39717 rows and 111 columns.
- Now we need to start with looking at the column values to do more deeper analysis.
- Removing or refactoring null null values.
- Removing outliers, etc.

Data Cleaning

- Changing object data type to required data type.
  - Eg: issue_d, last_pymnt_d, last_credit_pull_d changing these date formats from object to datetime.
  - Removing non usable values like '%', "months"
  - Removing extra spaces
  - Removing all the columns having null values

Observation 1: In the data set we have Fully paid
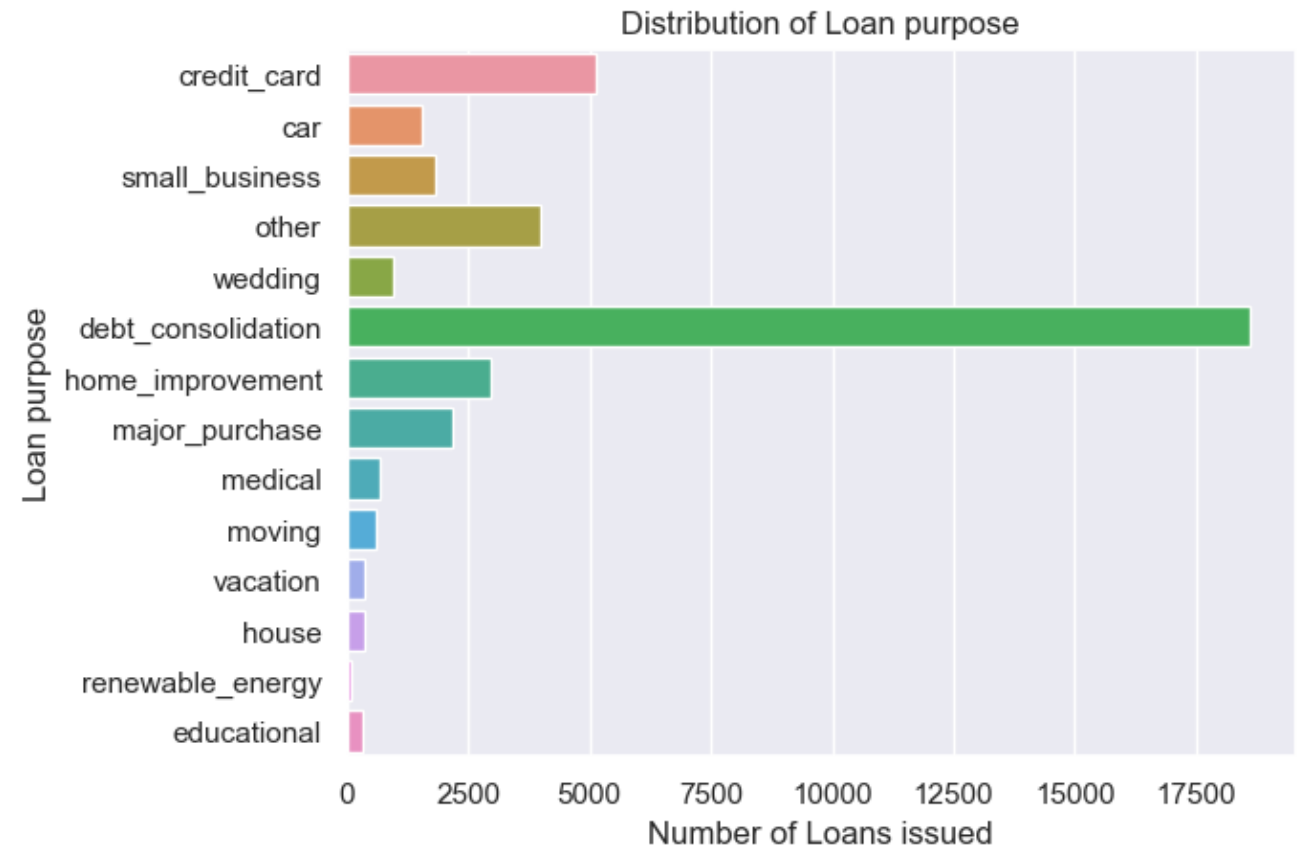loan status with the highest number.
1. Fully Paid   32950
2. Charged Off   5556
3. Current   1140

Distribution of Loan Status
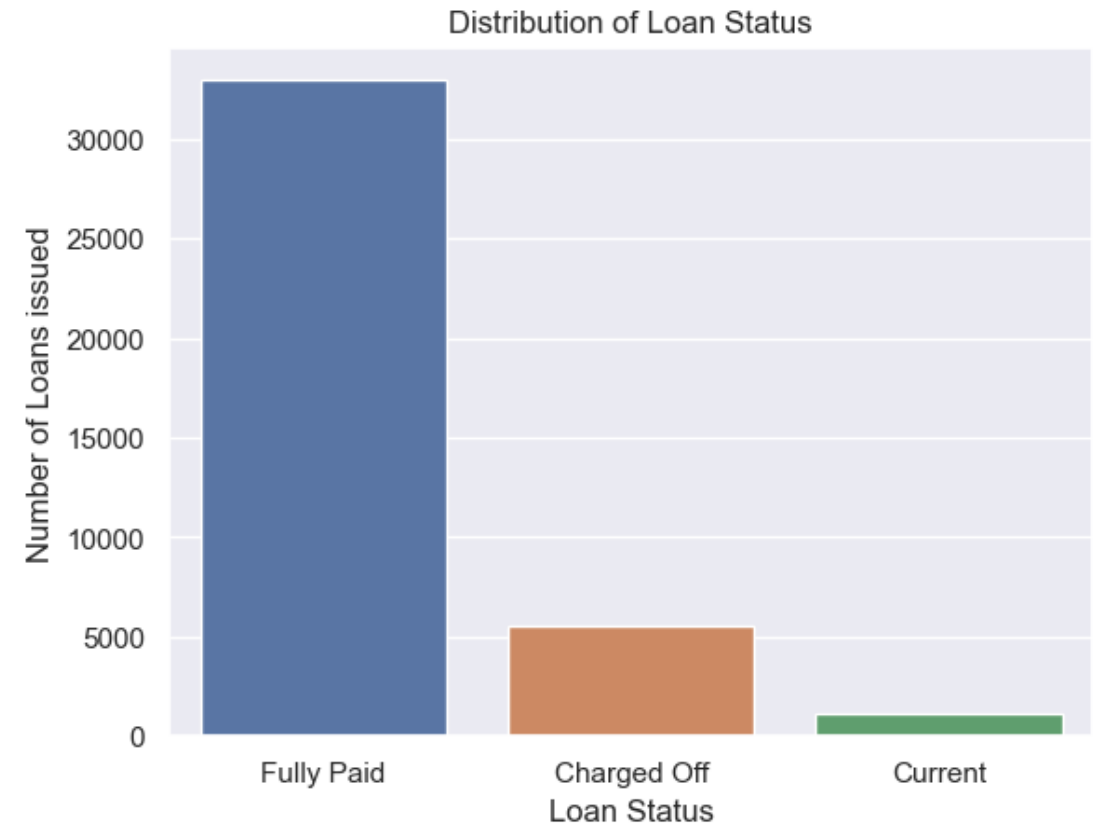
Observation 2: debt_consolidation is having the highest number of loan issued.

Top three loan status are:
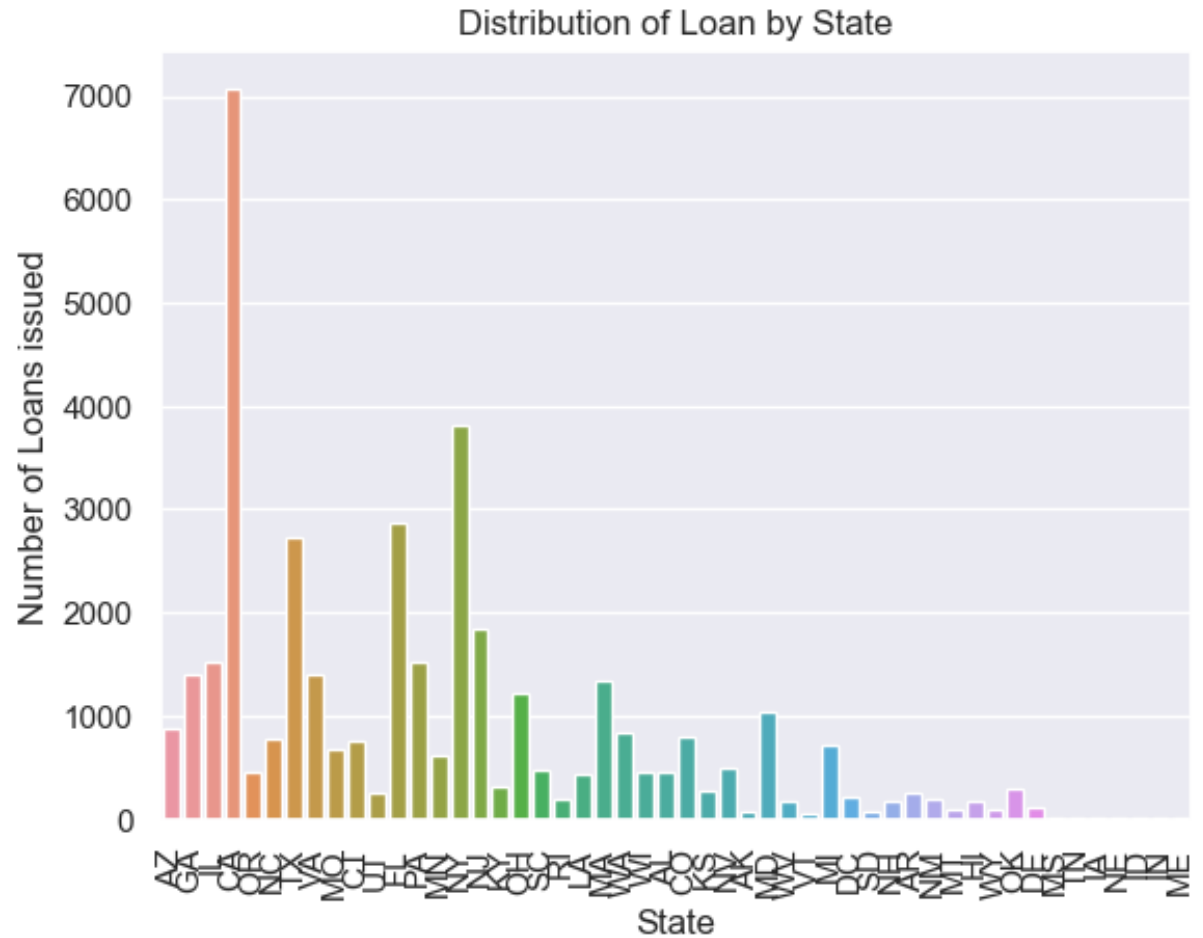1. debt_consolidation   18026
2. credit_card   5024
3. other   3855

Distribution of Loan purpose

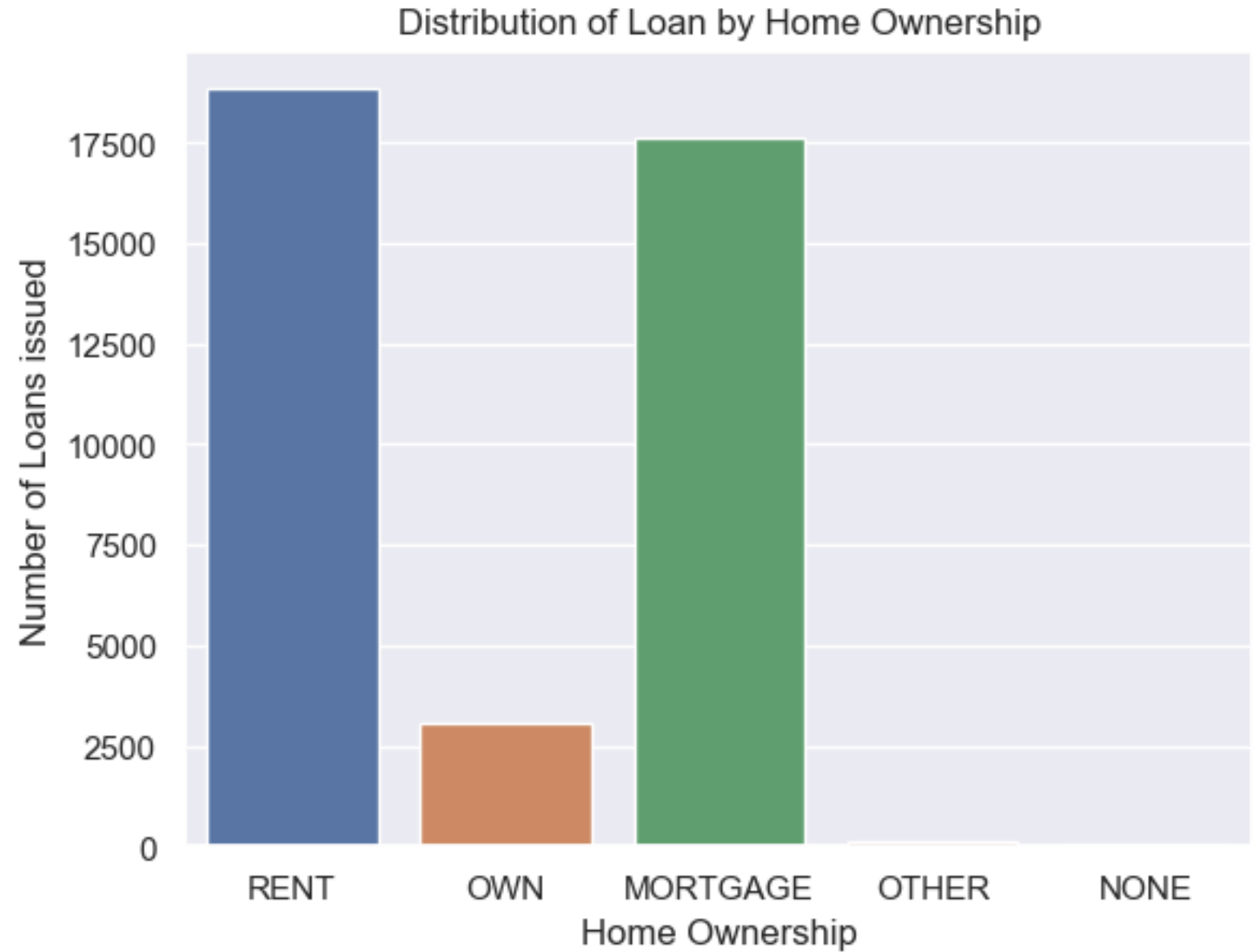Observation4: Number of loan issued in California state is the highest.

Top three states having heights number of loans are:
1. CA   6930
2. NY   3690
3. FL   2778

Distribution of Loan by State

Observation 4: The above countplot shows loan alocated to people having different house status.

We can see Rent with 18428 number of loan issued is the highest among all.
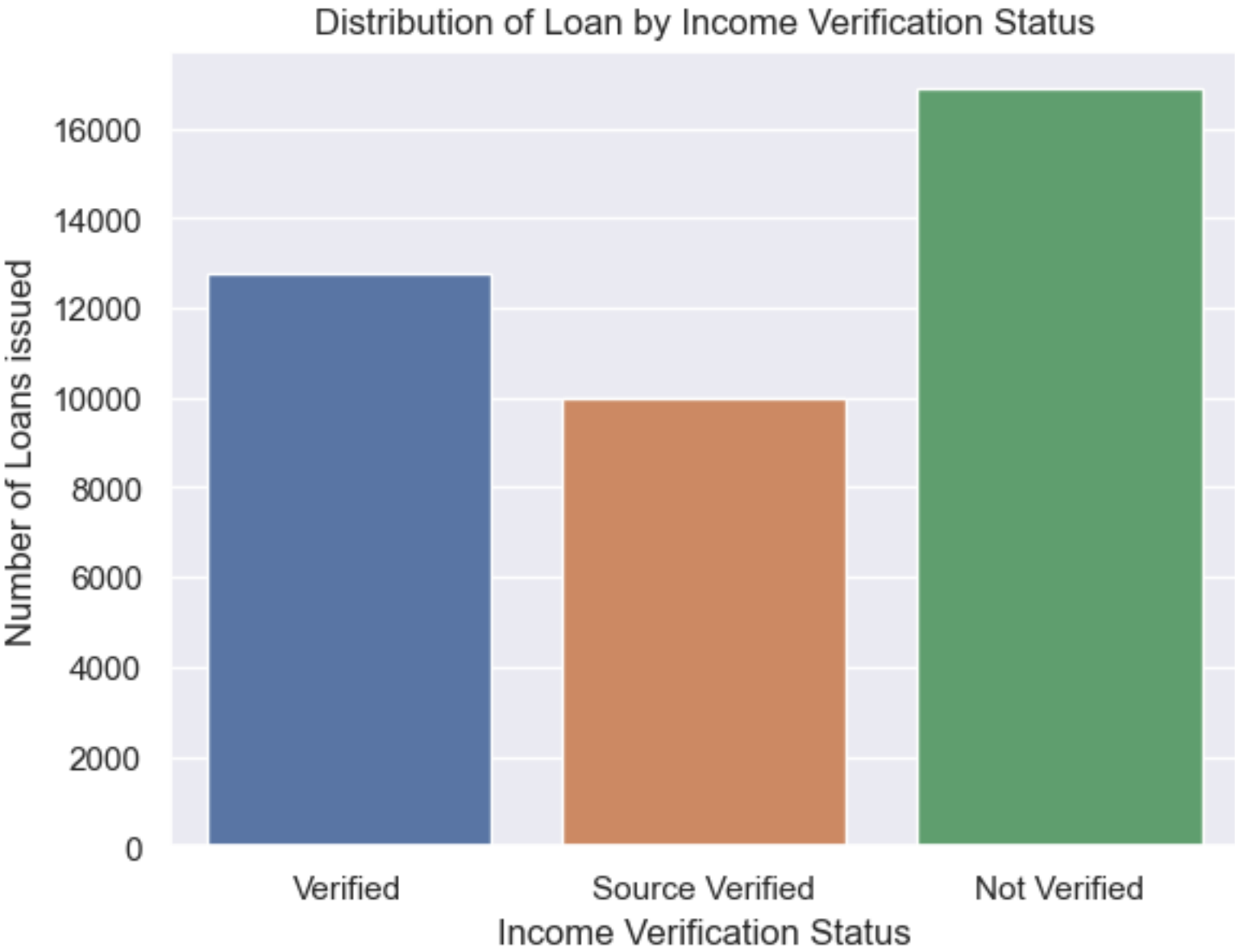

Distribution of Loan by Home Ownership

Observation 5: we can see the number of loan issued to 36 month is the highest. It could be possibly because of the consumer need, if the consumer wants to repay loan in only 36 months.
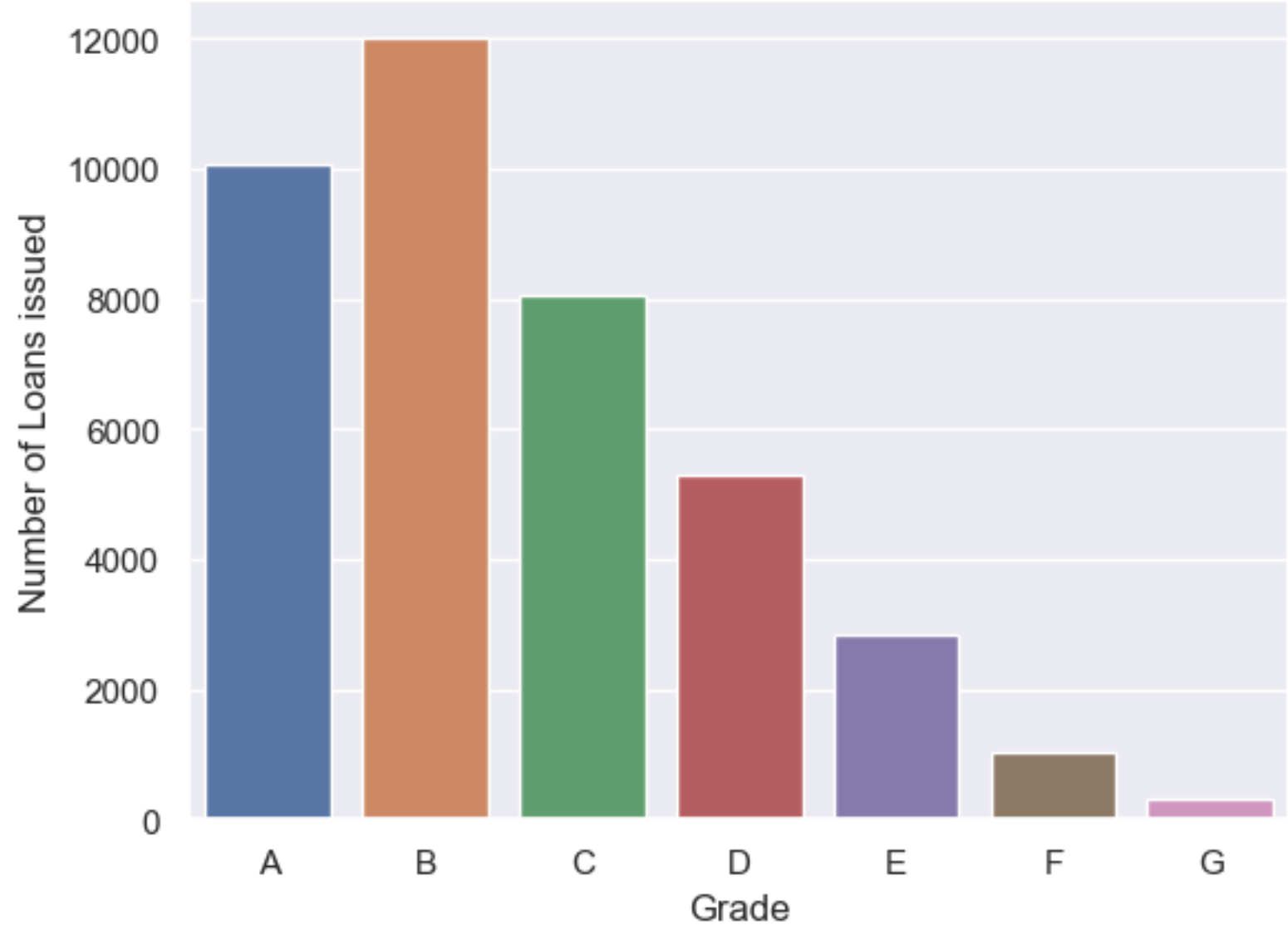

Distribution of Loan by Term Month
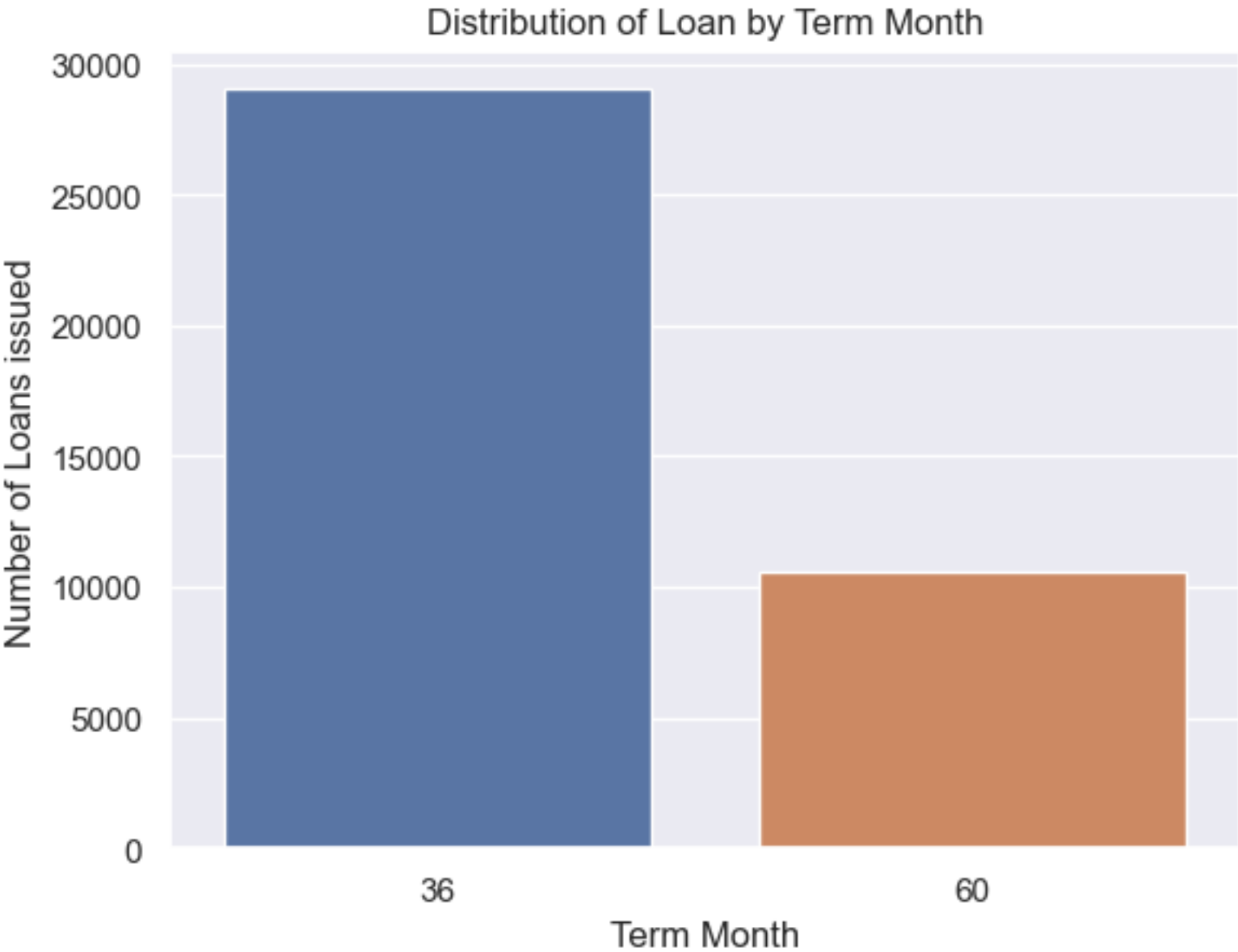
Observation 6: loan issued to non verified status is the highest.

Distribution of Loan by Income Verification Status

Observation 7: Number of loan issued to Greade B is highest.

Distribution of Loans by Grade

Observation 8: Subgrade A4 is having highest number of loan issued.

Top three subgrades:
1. A4   2873
2. B3   2824
3. A5   2711

Distribution of Loans by Sub Grade

Distribution of Loan Grades by Interest Rate
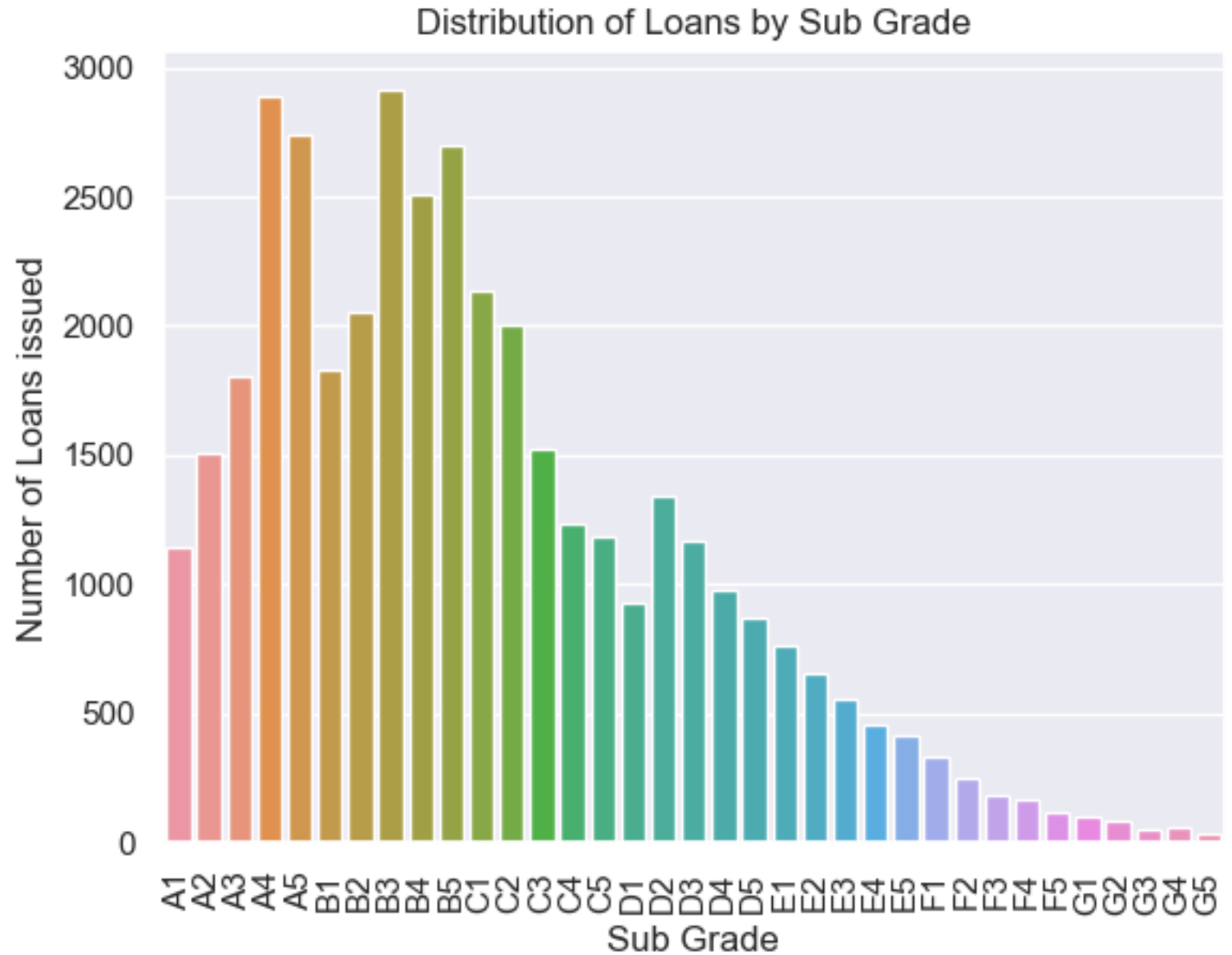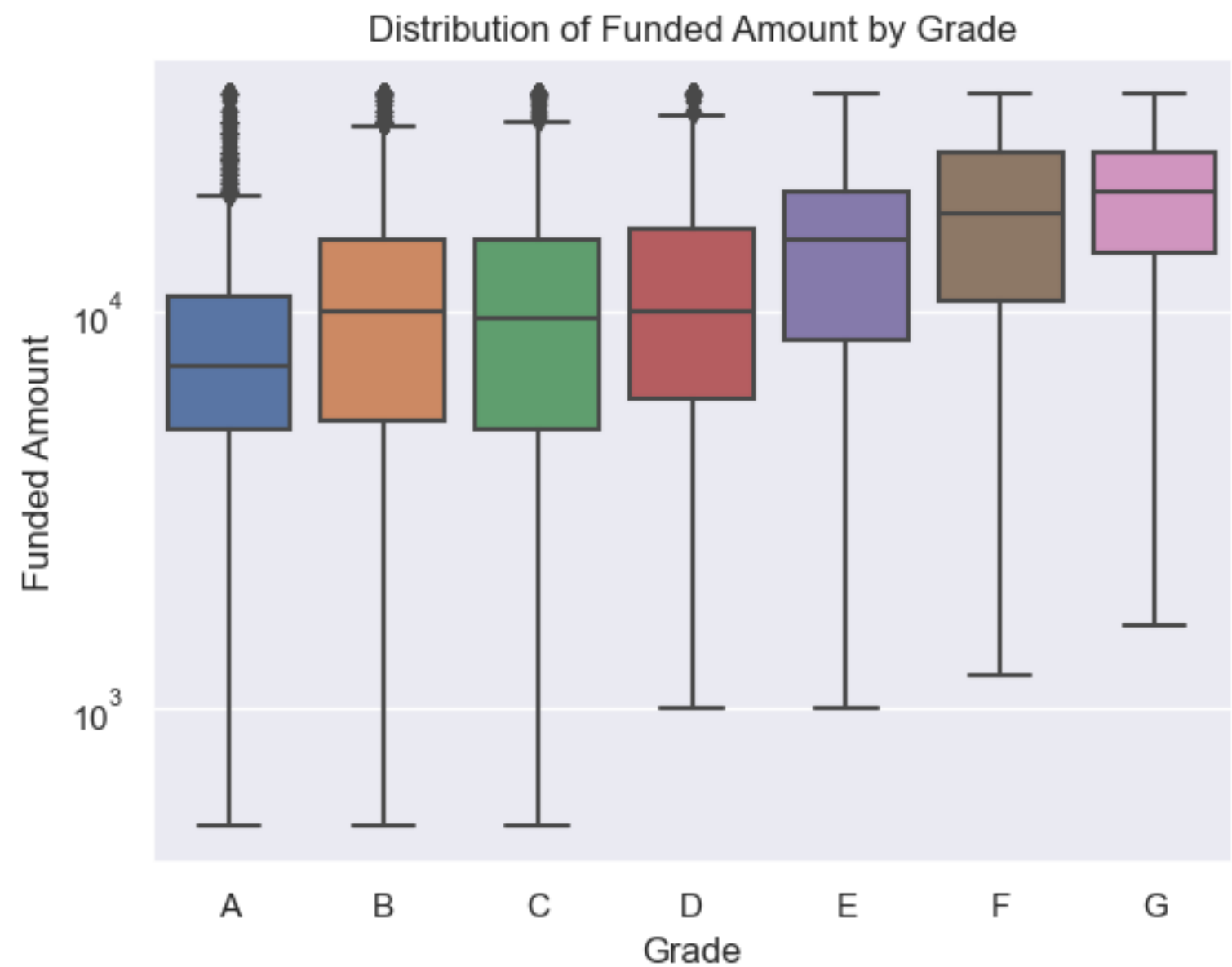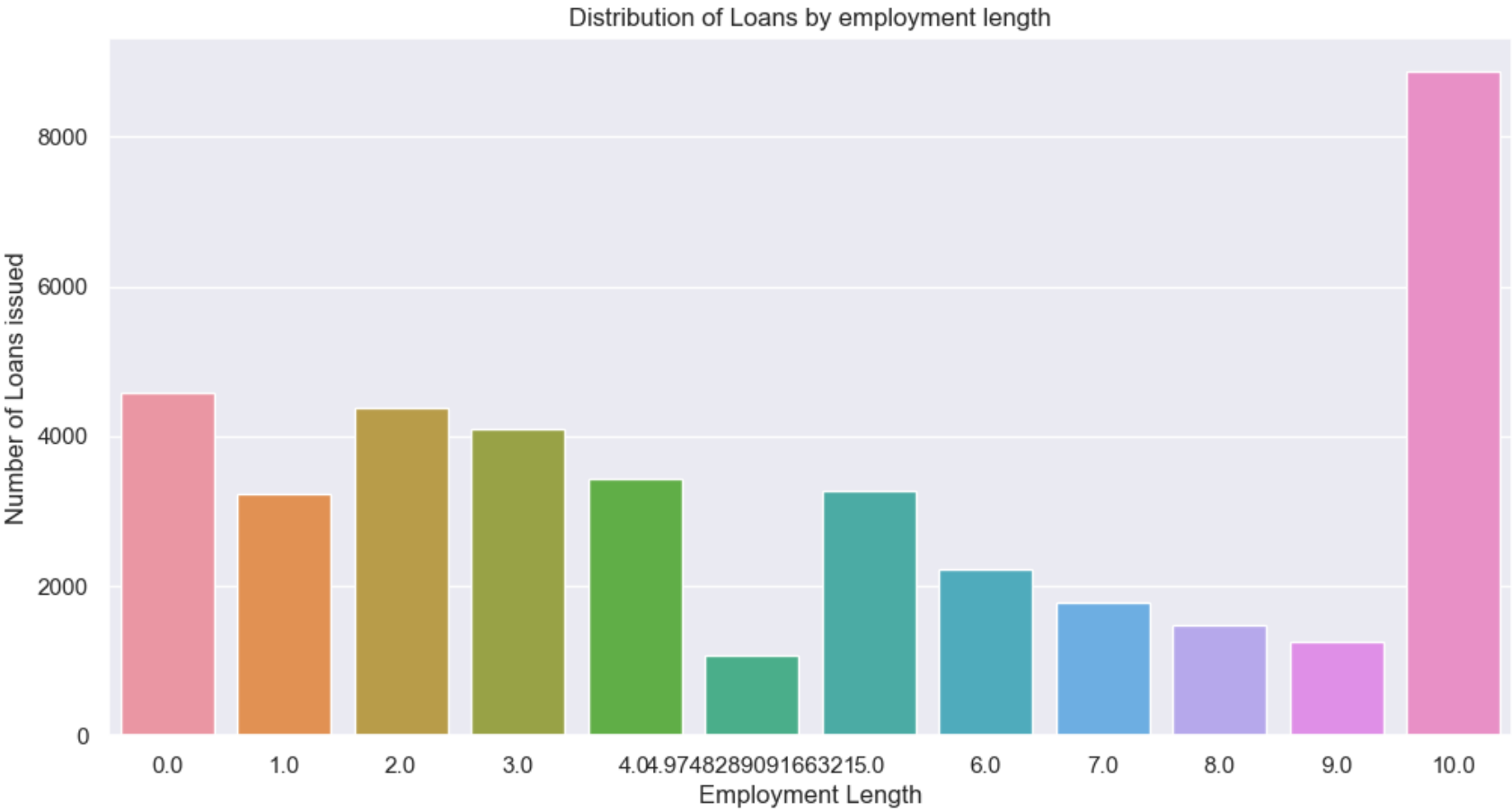
Observation 9: From grade A to G
Interest rate is increased

Observation 10: There are high funded amount in A,B,C and D grades

Distribution of Funded Amount by Grade

Observation 11: Majority of borrowers have been employed for at least 10 years.

Observation 12: There was a positive correlation between interest rates and ROI for borrowers who paid their loans, but for those who didn't pay the ROI was negative and it shows no correlation with interest rates.
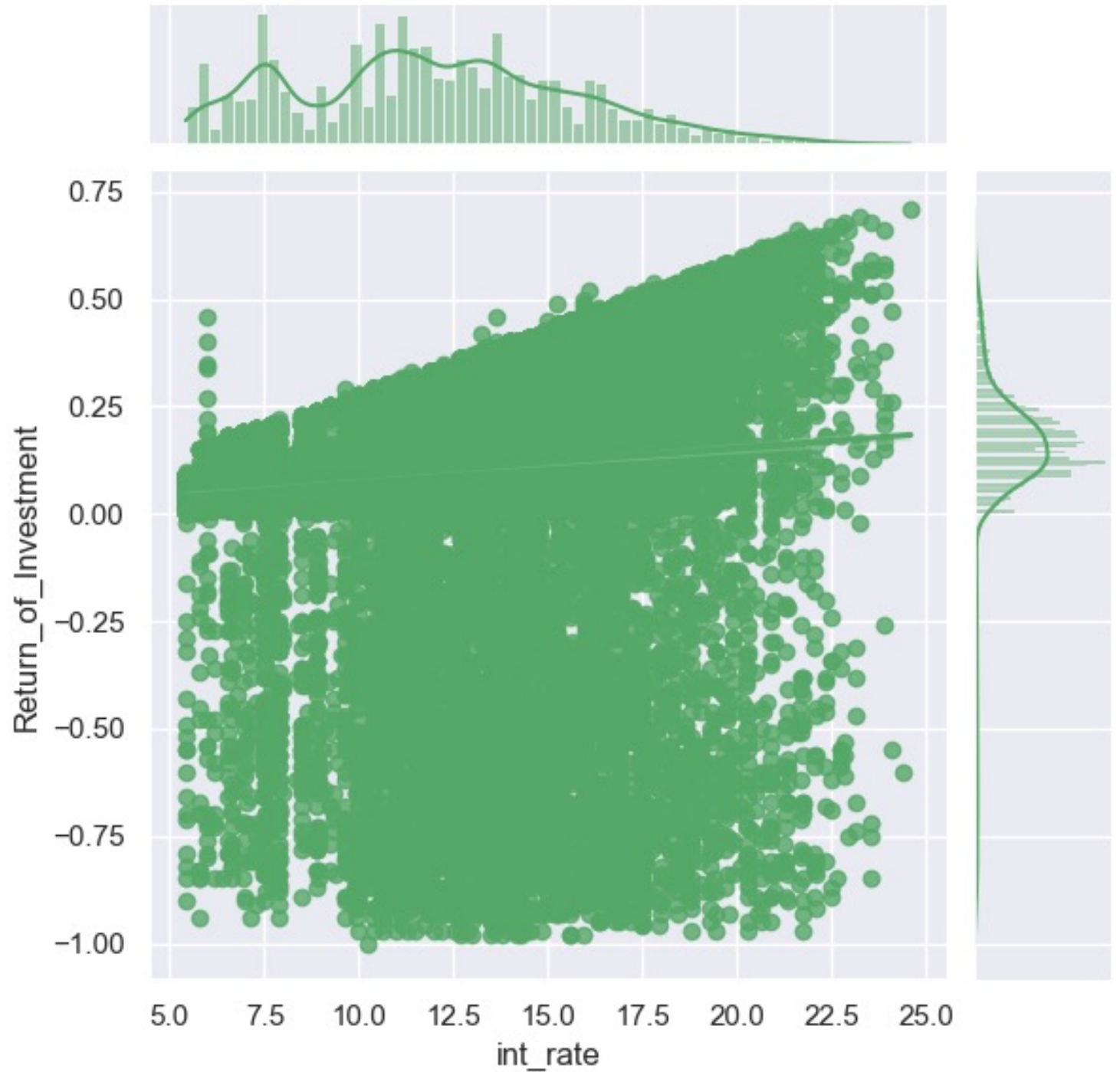
Observation 13: Borrowers with higher income has positive ROI while borrowers with lower income has negative ROI.
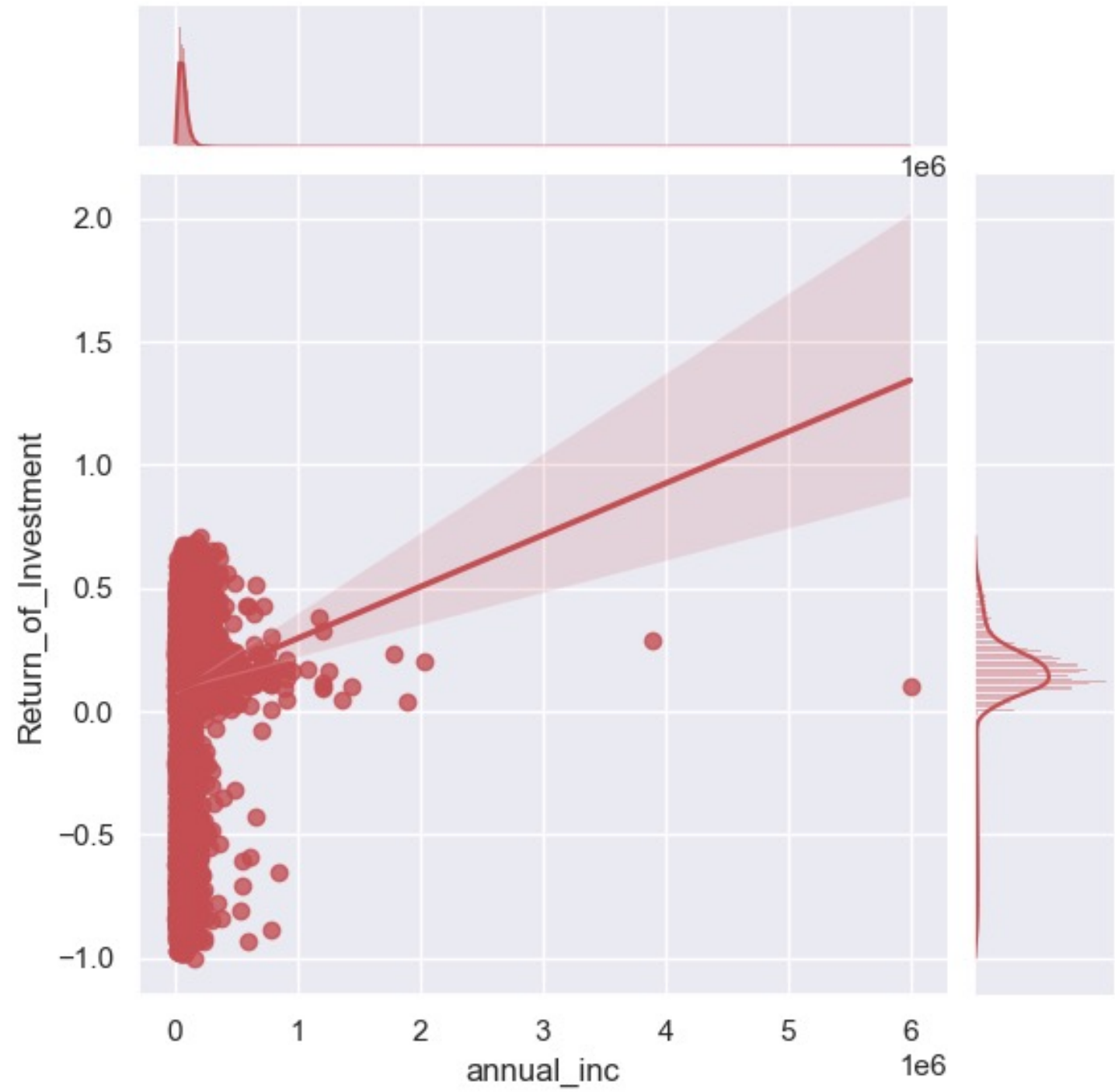
Observation 14: It shows there are more defaulters in RENT and MORTGAGE

Distribution of Loan Defaulters by Home Ownerships