Lending Club Case Study

Harish Shindhe

Harshit Sharma

Requirement / Objective…

- Customer loan data with Data dictionary is provided of a firm that lends money to customers for their various needs such as business, home, personal needs.
- Provide insights into data that the firm has captured so far with focus on customers / loans that have defaulted.
- Provide guidance on what variables both from a customer and loan perspective have high leverage on default of the loan and help the form to make informed decisions in future. Requirement…

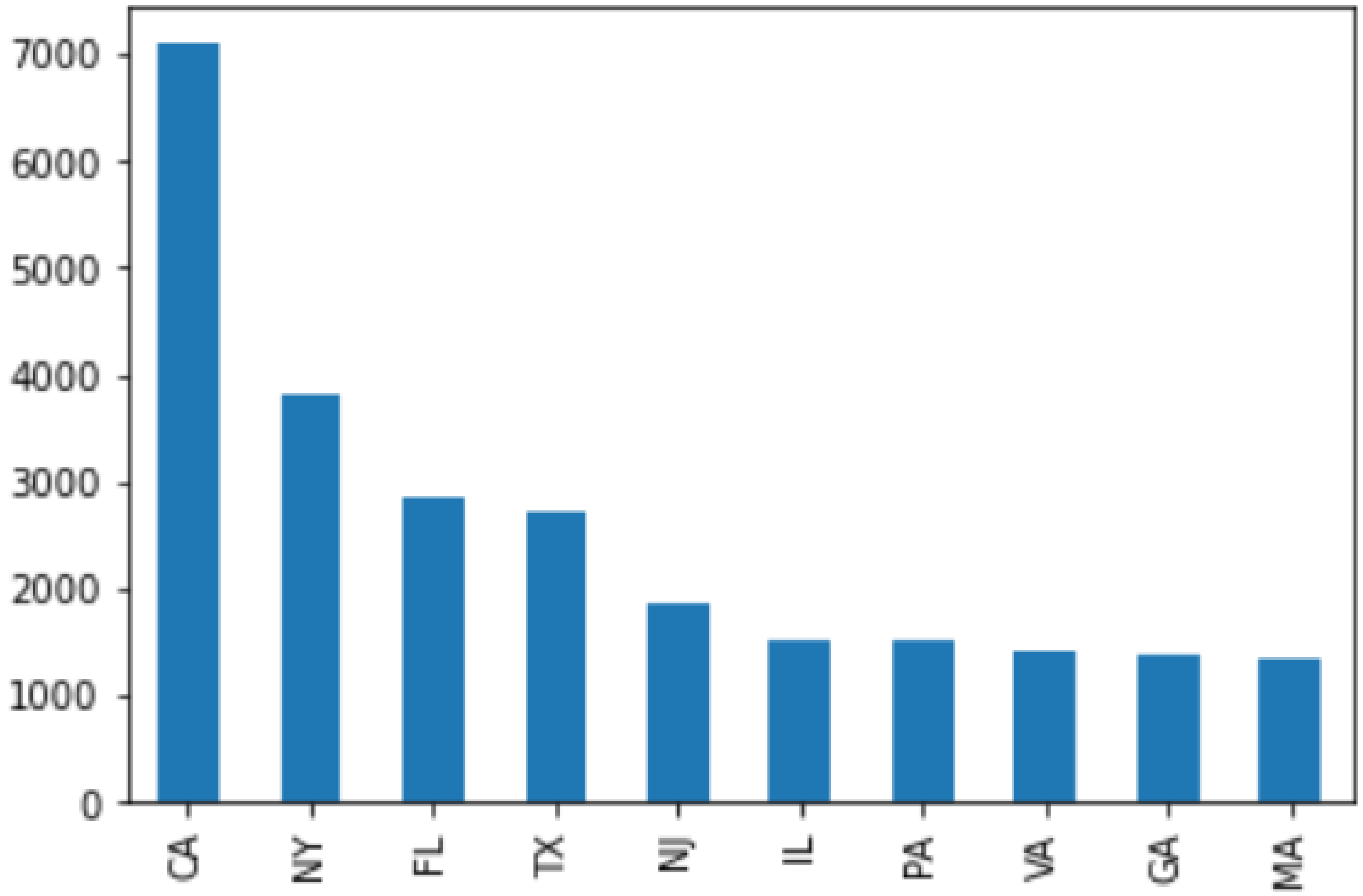
Data Cleaning…

- Remove all columns that have predominantly null values
- Remove all rows which have maximum null values in their columns if exists
- Remove all duplicate rows if exists
- Remove any duplicate columns if exists (columns that have similar values)
- Perform analysis of data dictionary and remove any columns that are not needed for analysis

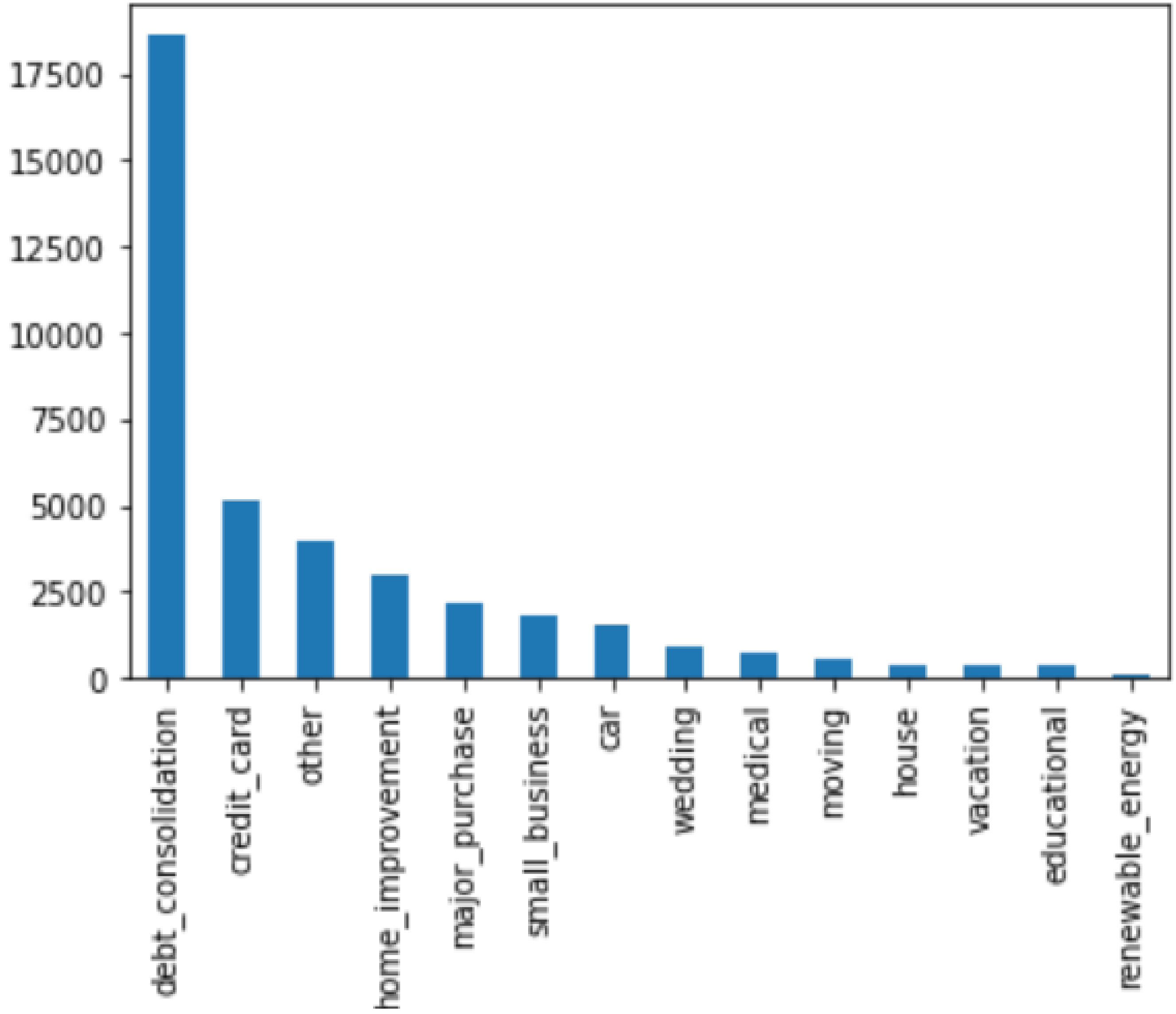
Data Modification…

- Convert the necessary variables from objects to numeric type
- If any of the values for the variables that need to be converted have characters, remove them.
- Split data time columns into separate columns like year, month
- Bin the data into necessary intervals so as to perform analysis

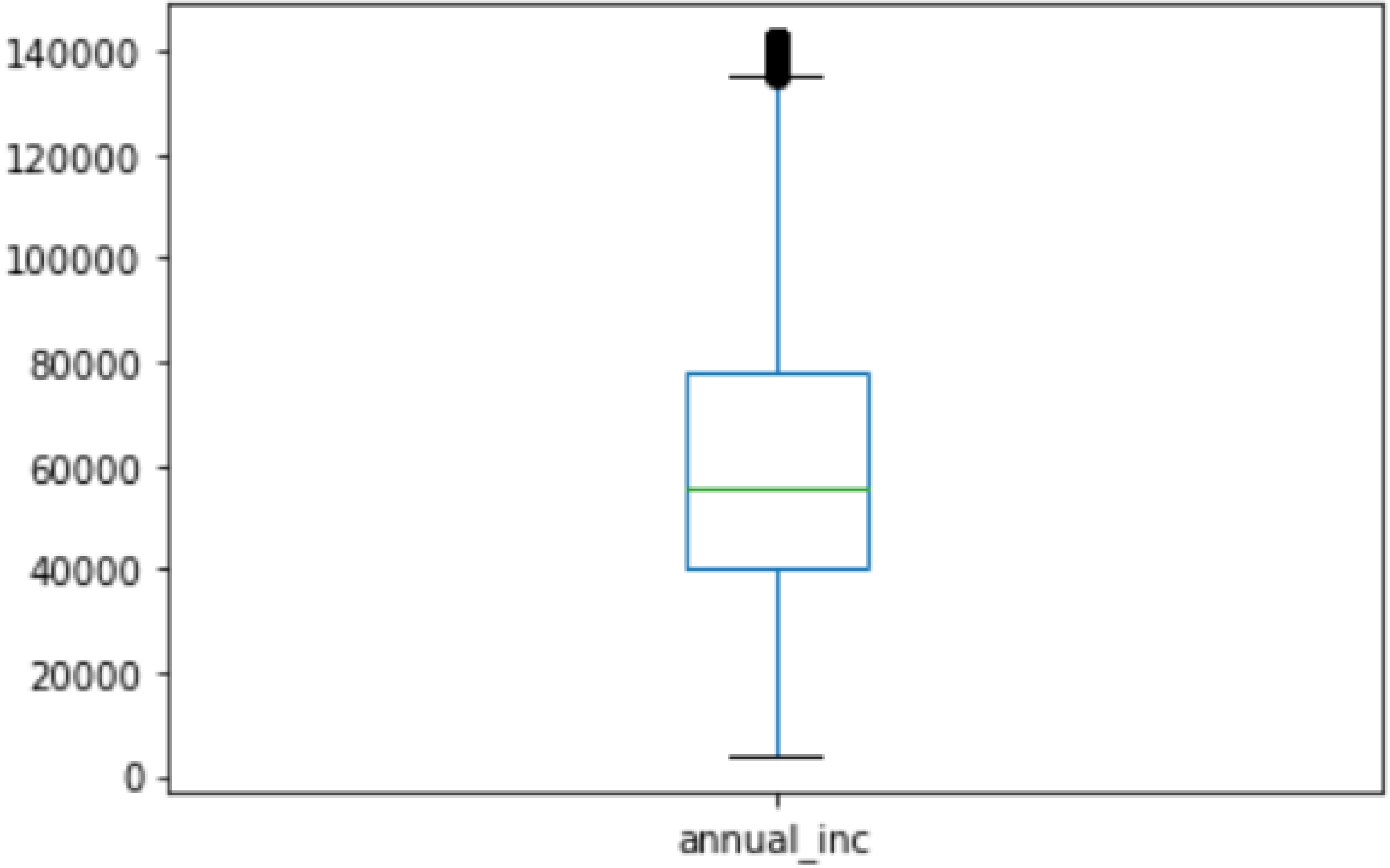
Top 10 states depicting the highest number of customers…



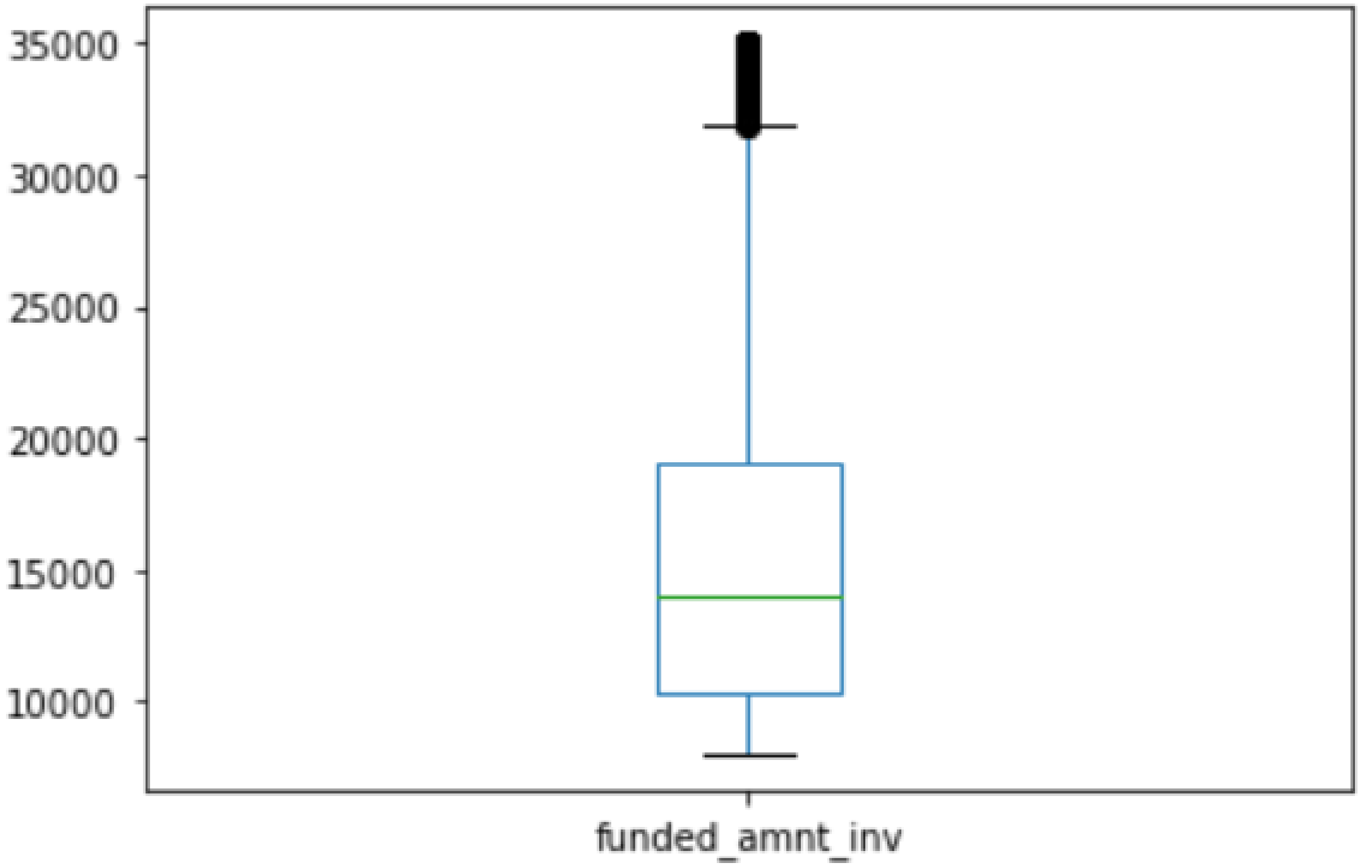
Main reasons why customers avail loans...



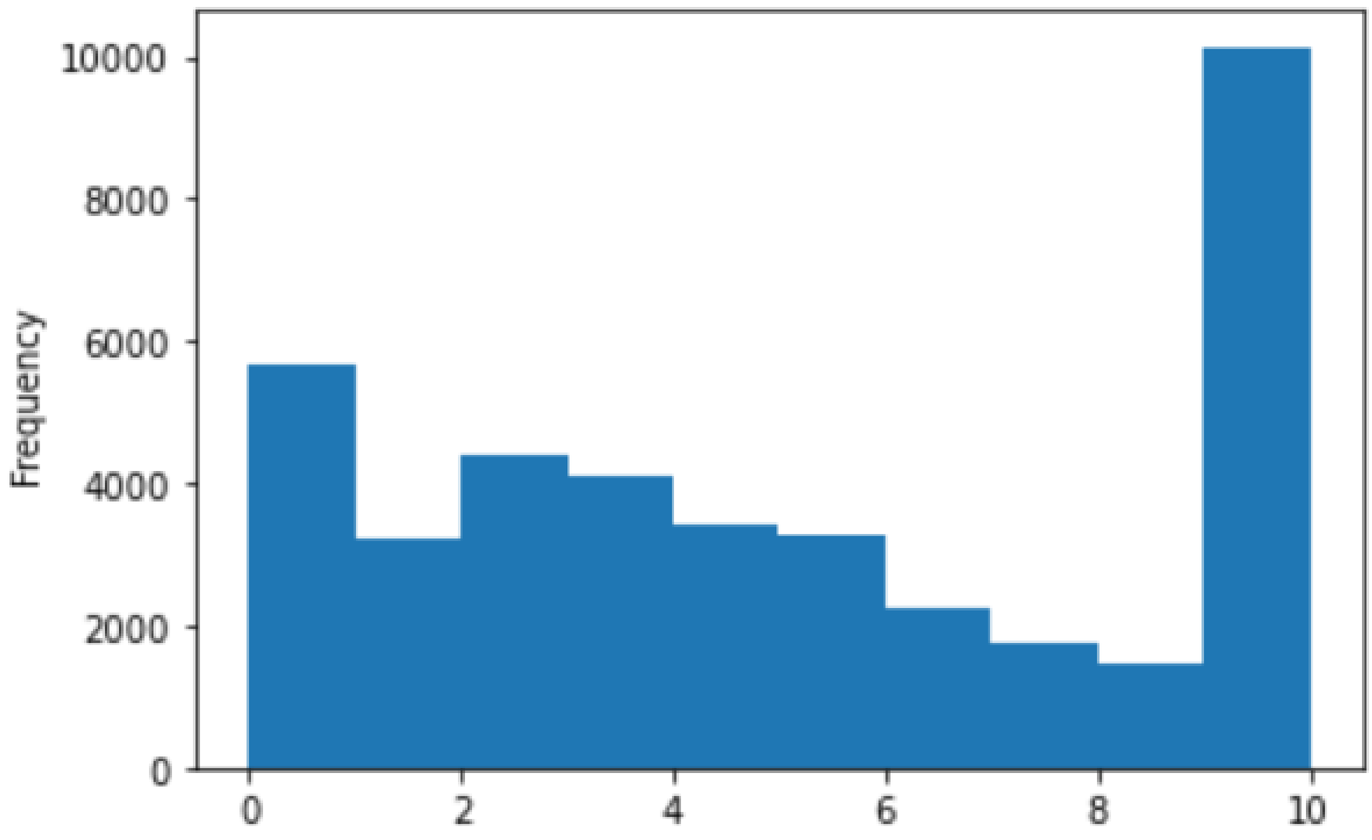
Customers who have availed loan are in the annual income range of 40 k to 80 k…



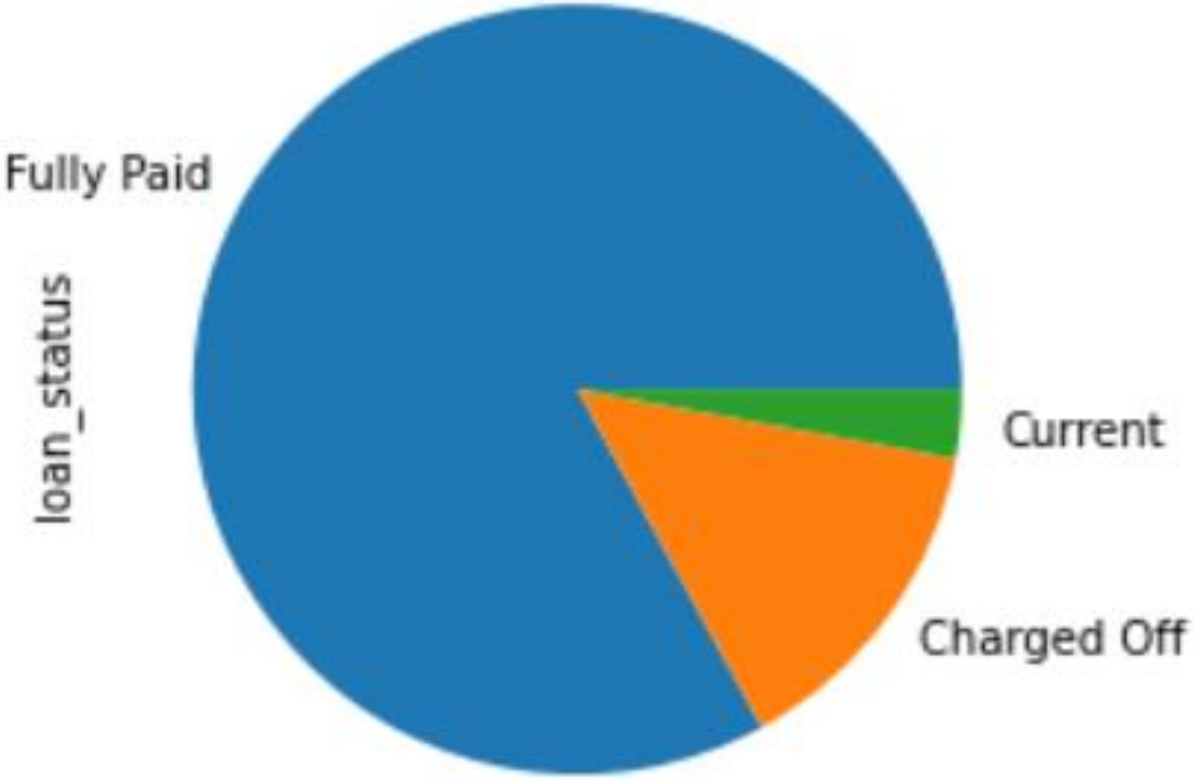
Maximum loans funded are in the range of 10 k to 20 k



Employees > 10 years - biggest benefitiaries of loan.
The second largest group is people with < 1 year

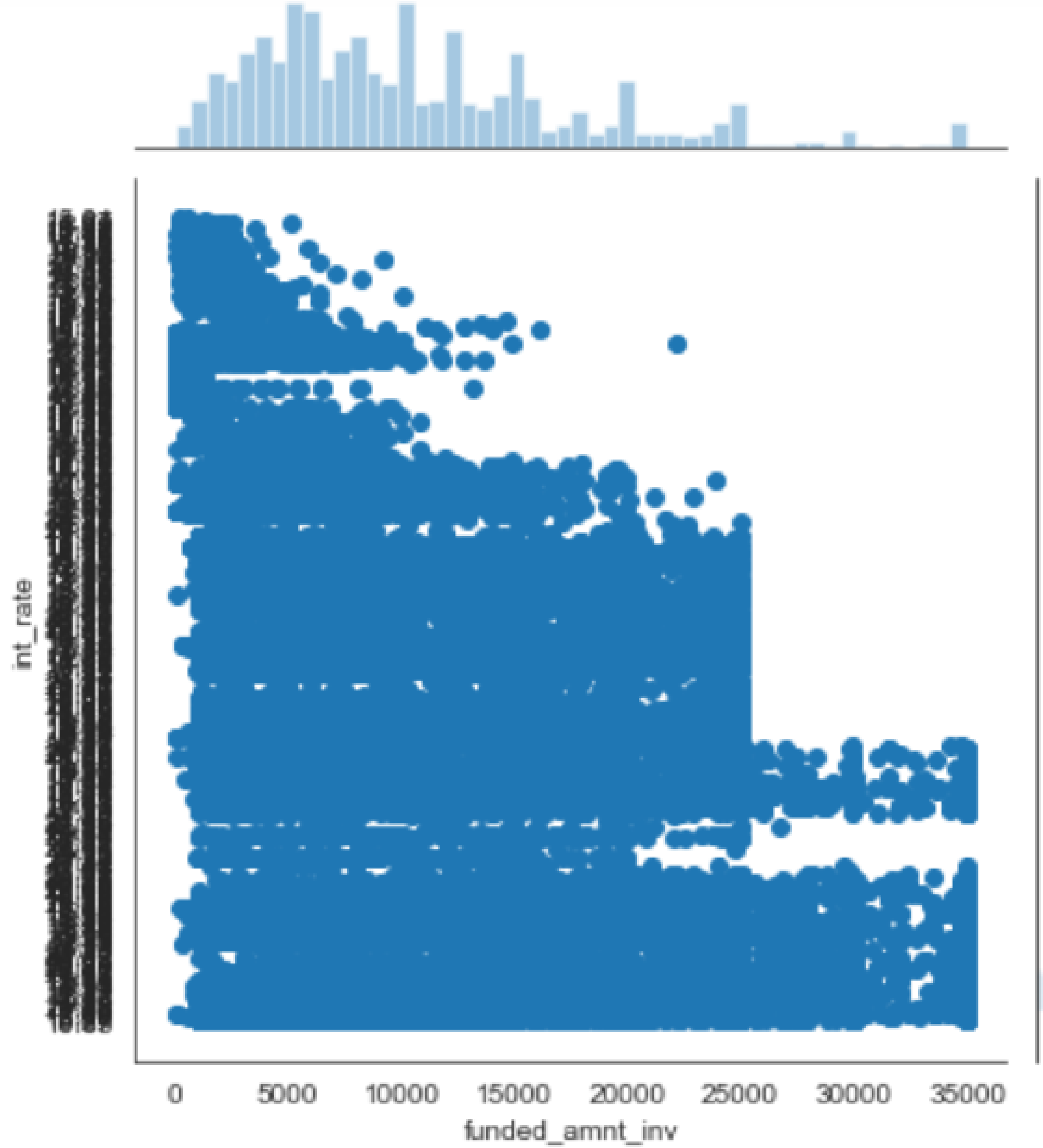


Most of the loans are already paid, unfortunately a good percentage are defaults

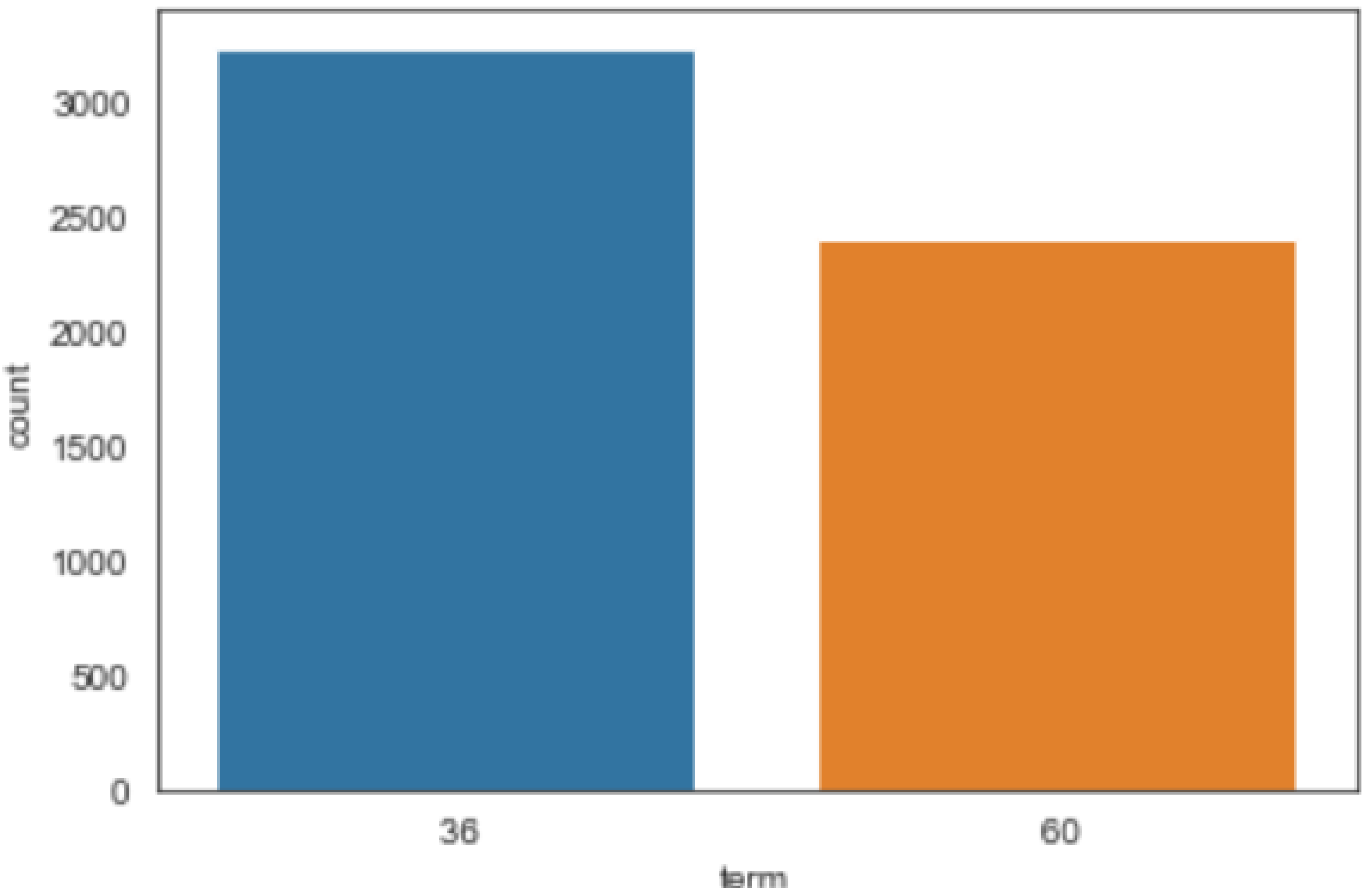


Lower amounts are funded at a higher rate of interest and seem to recede gradually as we approach the 25 k mark.

Any loan given after 25 k, there is a sharp decrease in interest rates

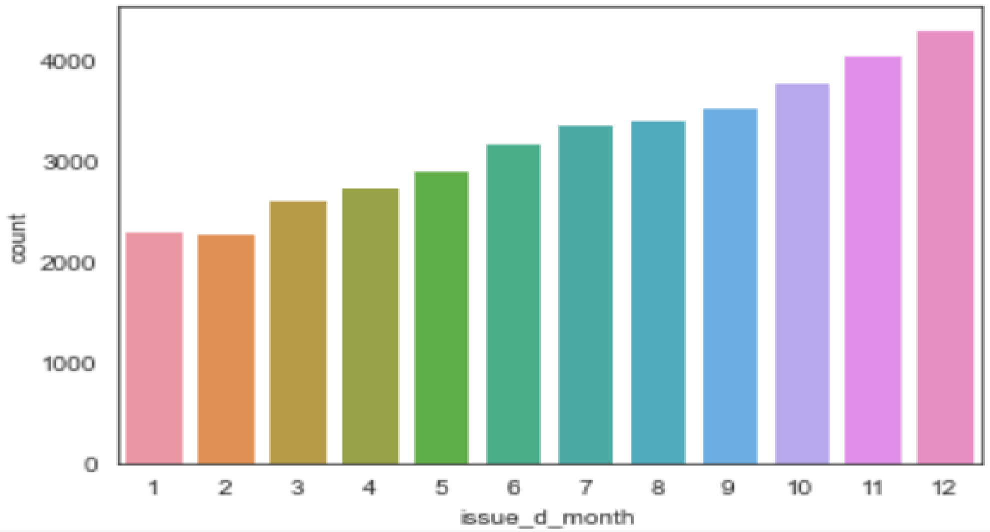


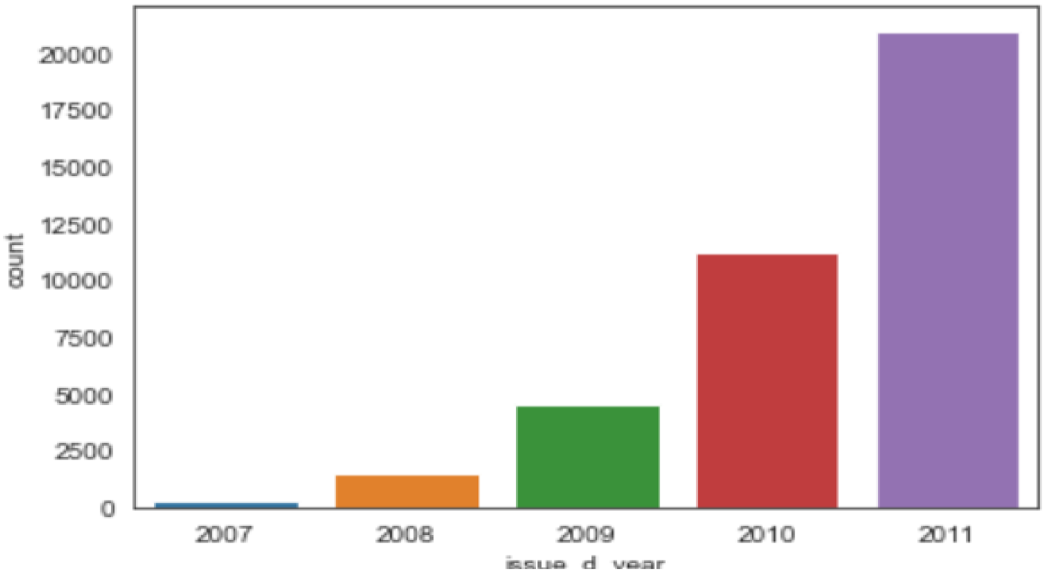
Looks like loans offered at 36 months are most defaulted



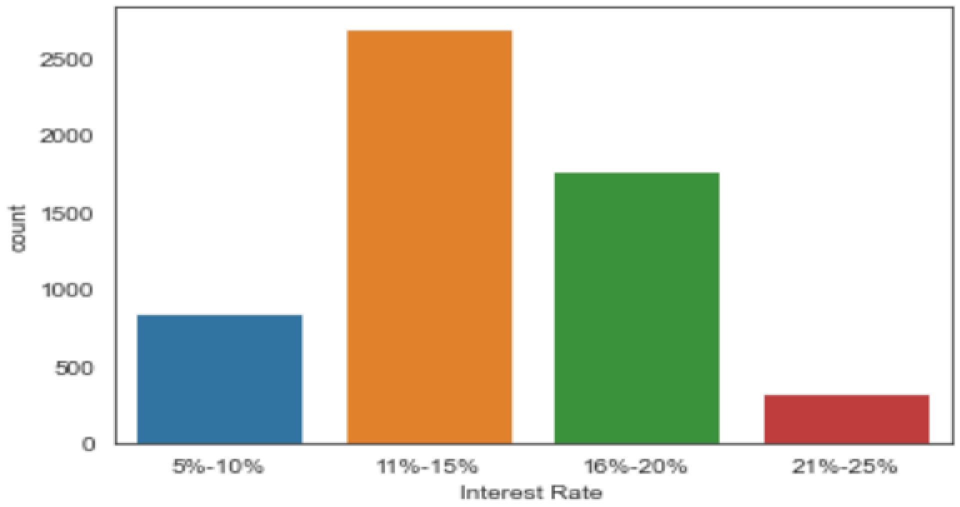
Looks like loans defaults are rising year on year
Loan provided in 2011 are the most defaulted

Loans taken during the month of December are the most defaulted.

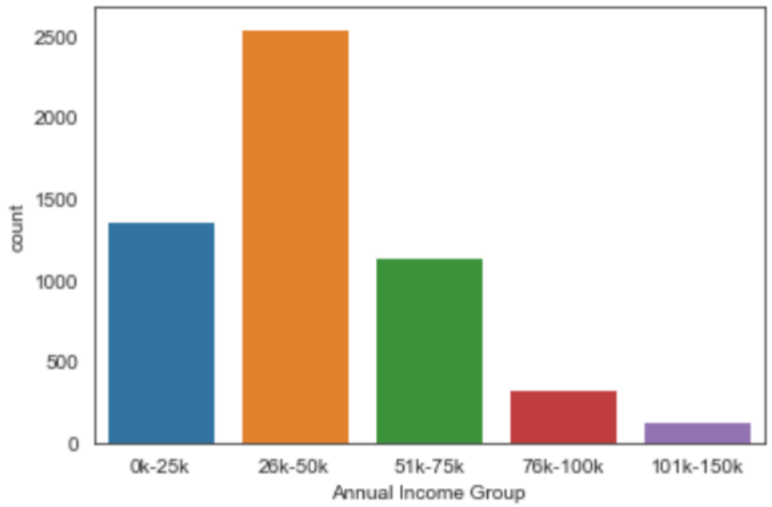


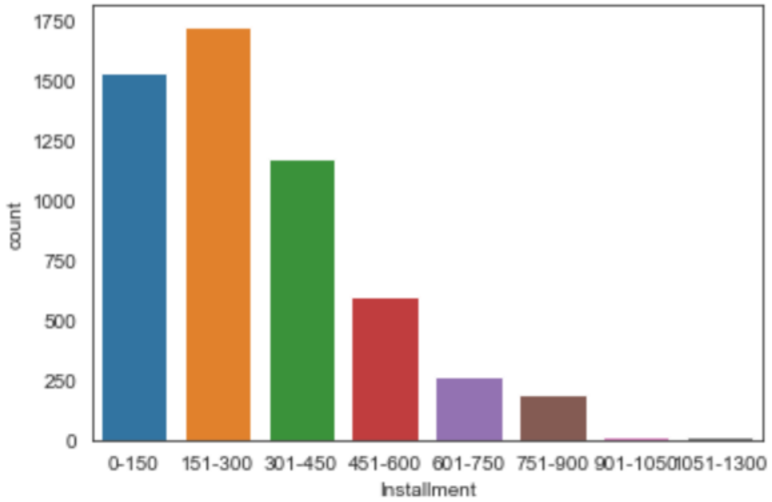


Looks like the loans that were offered in the interest rate of 11-15 % have defaulted the most



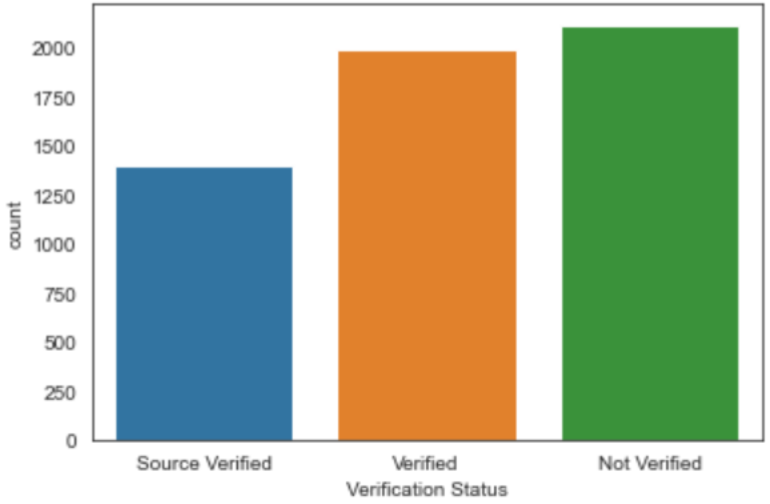
Looks like customers with annual income of 26-50 k have defaulted the most





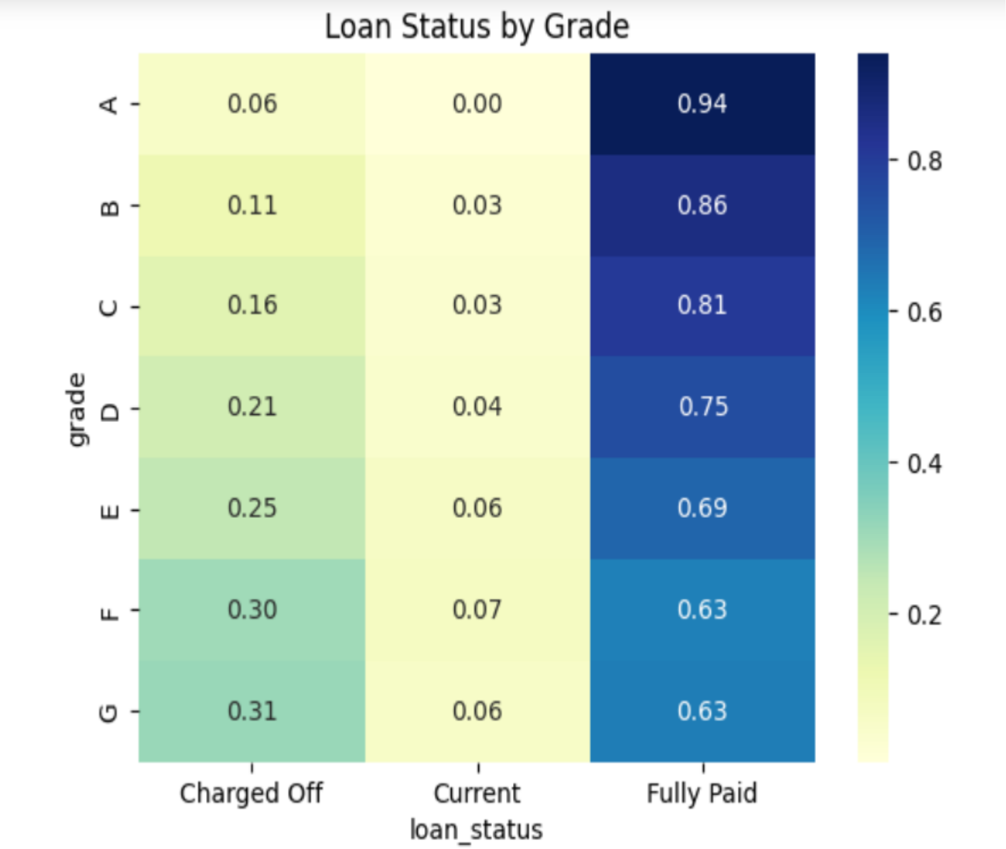
Looks like installments in the range of 151-300 are the most defaulted

Loans that were not verified are the largest that have defaulted



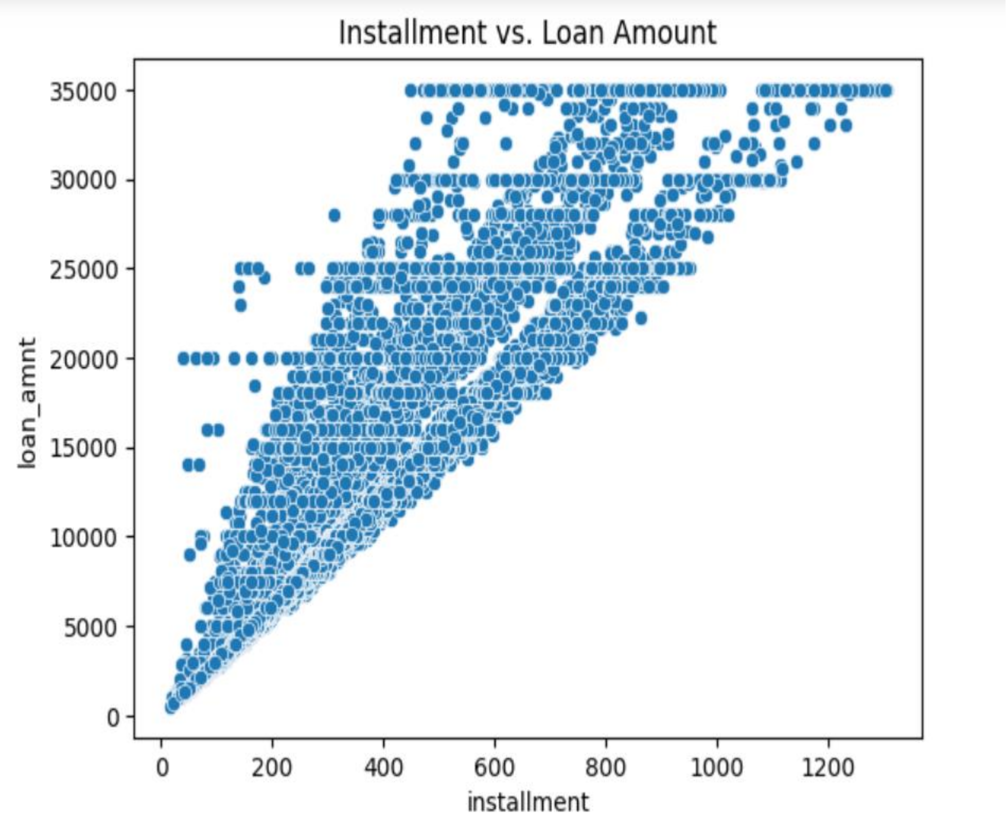
Loan status by Grade…

- There's a clear correlation between loan grade and loan status.
- As the loan grade decreases (from A to G), the percentage of charged-off loans increases, while the percentage of fully paid loans decreases.
- Grade A loans: Have an overwhelmingly high percentage of fully paid loans and a negligible charge-off rate, indicating a very low risk profile.
- Grades B and C: Also show a high percentage of fully paid loans but with a slightly increasing charge-off rate compared to A grade loans.
- Grades D, E, F, and G: Exhibit a significantly higher proportion of charged-off loans compared to fully paid loans, suggesting a higher risk associated with these loan grades.

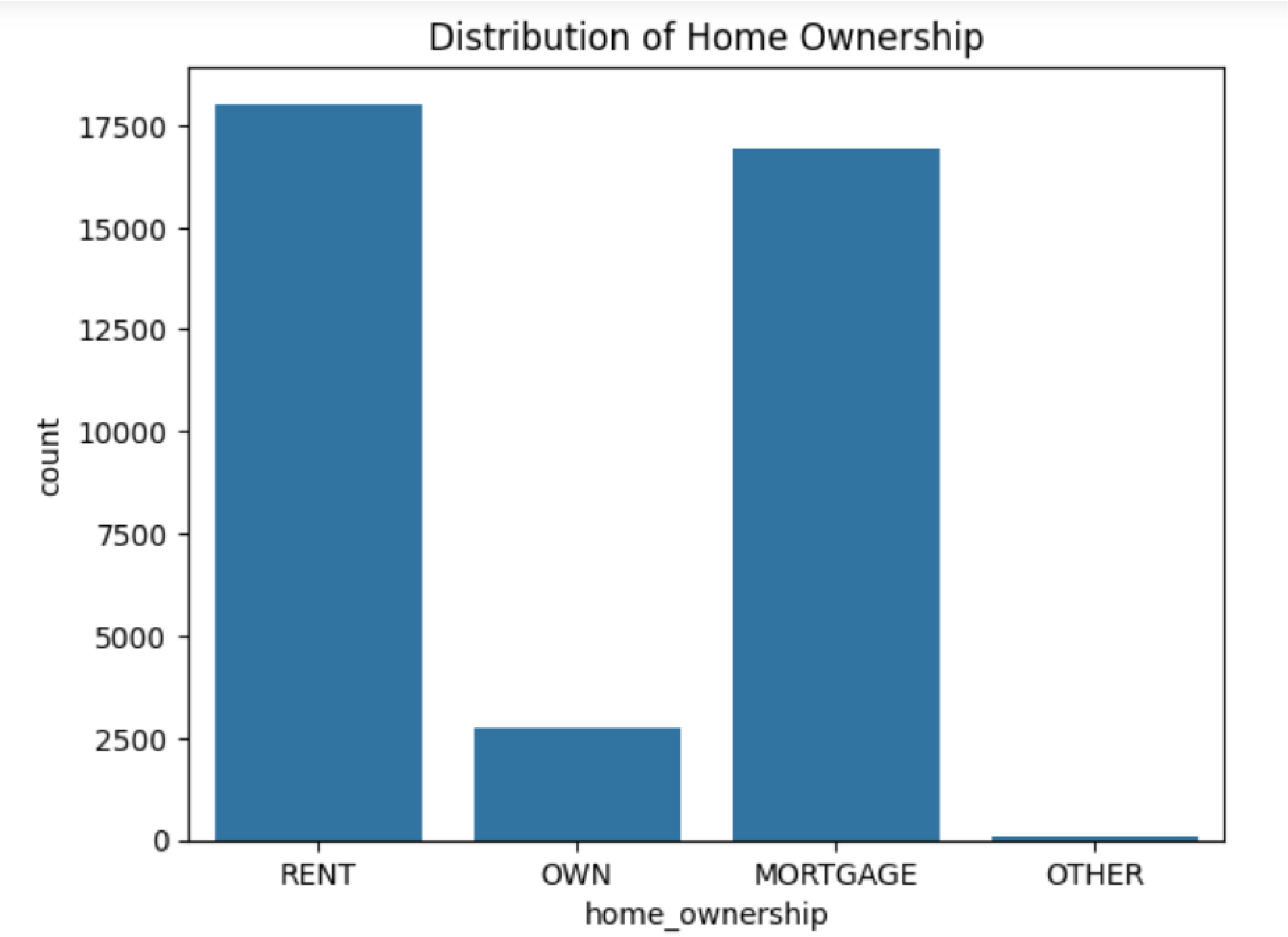


Instalment vs. Loan Amount…

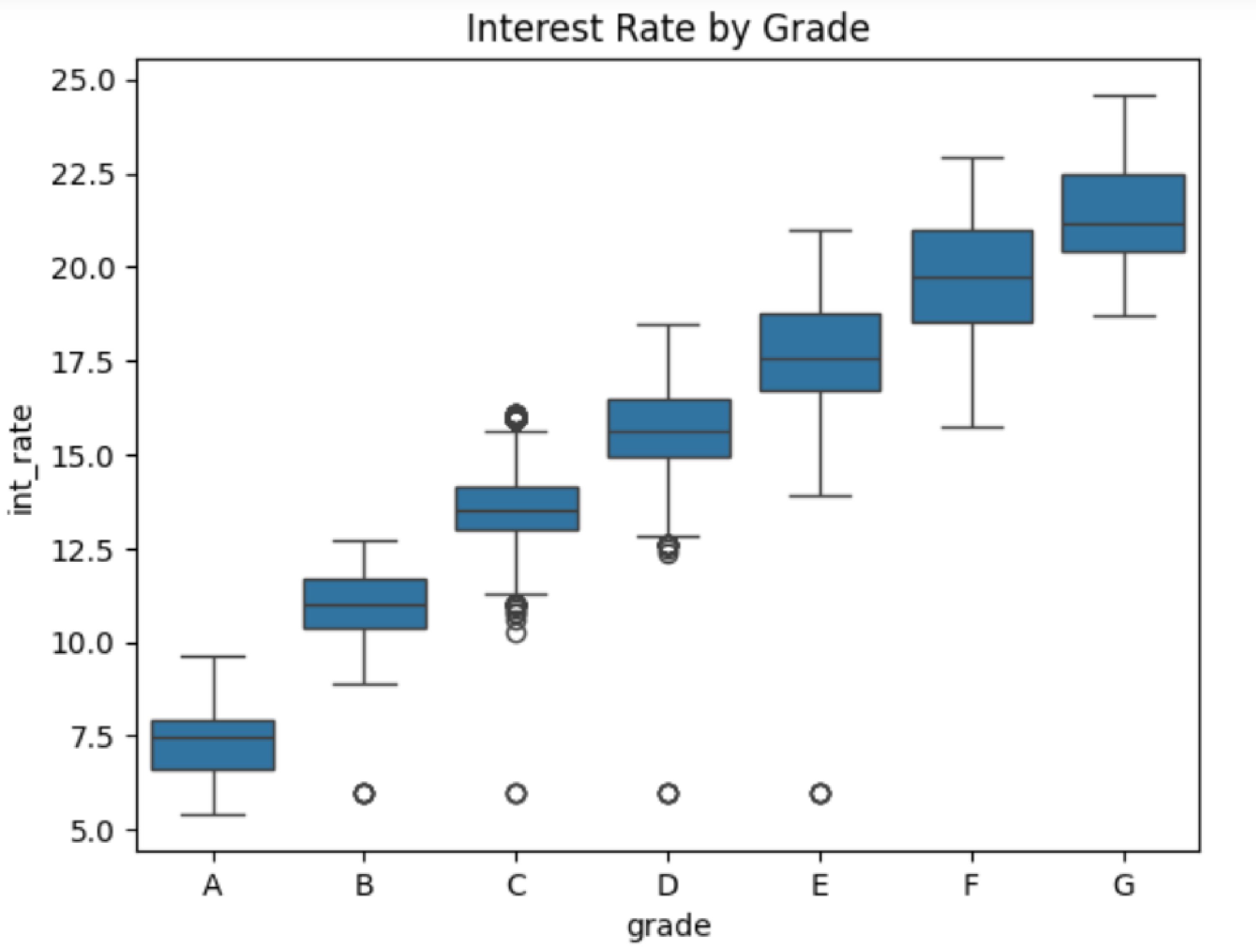
- The scatter plot illustrates a strong positive correlation between installment amount and loan amount.
- While there is some scatter in the data, the overall relationship appears to be linear, suggesting a consistent proportionality between the two variables.
- It also signifies that the higher the loan amount, the more will be the installment.



The majority of individuals in the dataset are renters, followed by homeowners with mortgages. Home ownership without a mortgage and other housing situations are less common.



There is a clear positive correlation between loan grade and interest rate. As the loan grade decreases (from A to G), the interest rate tends to increase.



Conclusion…

- **General Observaions**

    - Maximum loans were taken in California followed by Newyork, Florida, Texas and so on.
    - Debt consolidation is the number one purpose followed by Credit card payment.
    - Educational loan unfortunately figures as one of the last reasons, might indicate only self sufficient are able / willing to pursue education and there is lack of interest to procure a loan to pursue it.
    - Most of the people who have availed loan are in the annual income range of 40 k to 80 k
    - Most of the loans that have been funded are in the range of 10 k to 20 k
    - Employees with more than 10 years of experience are the biggest recipients of loan, The second largest group is people with < 1 years of experience.
    - Lower amounts are funded at a higher rate of interest and seem to recede gradually as we approach the 25 k mark
    - Any loan given after 25 k, there is a sharp decrease in interest rates
    - Most loans are concentrated in the lower grade categories (A, B, C), indicating a lower risk profile of the overall loan portfolio. There's a significant decrease in loan count as the grade deteriorates, suggesting fewer higher-risk loans.
    - The majority of individuals in the dataset are renters, followed by homeowners with mortgages. Home ownership without a mortgage and other housing situations are less common.

Conclusion…

- **General Observaions**

  - The majority of individuals in the dataset are renters, followed by homeowners with mortgages. Home ownership without a mortgage and other housing situations are less common.
  - The data signifies that the higher the loan amount, the more will be the installment.
  - There is a clear positive correlation between loan grade and interest rate. As the loan grade decreases (from A to G), the interest rate tends to increase.
  - Individuals with mortgages tend to have the highest median loan amount, followed by those who rent, own their homes outright, and those in the other category.
  - There are some outliers present, particularly for renters and Home Owners, suggesting a few individuals with considerably higher loan amounts compared to the majority.

Reasons for Default…

- Looks like loans offered at 36 months are most defaulted.
- Looks like loans defaults are rising year on year and loan offered in 2011 are the most defaulted
- Loans taken during the month of December are the most that are likely to be defaulted.
- Looks like the loans that were offered in the interest rate of 11-15 % have defaulted the most
- Looks like customers with annual income of 26-50 k have defaulted the most
- Looks like installments in the range of 151-300are the most defaulted
- As expected, the loans that were not verified are the largest that have defaulted but not much difference between the verified which might hint the current verification methods might be inadequate and needs a relook

- Most of the people who have availed loan are in the annual income range of 40 k to 80 k
- Most of the loans that have been funded are in the range of 10 k to 20 k
- Employees with more than 10 years of experience are the biggest recipients of loan, The second largest group is people with < 1 years of experience.
- Lower amounts are funded at a higher rate of interest and seem to recede gradually as we approach the 25 k mark
- Any loan given after 25 k, there is a sharp decrease in interest rates

Reasons for Default…

- There's a clear correlation between loan grade and loan status.
- As the loan grade decreases (from A to G), the percentage of charged-off loans increases, while the percentage of fully paid loans decreases.
- Grade A loans: Have an overwhelmingly high percentage of fully paid loans and a negligible charge-off rate, indicating a very low risk profile.
- Grades B and C: Also show a high percentage of fully paid loans but with a slightly increasing charge-off rate compared to A grade loans.
- Grades D, E, F, and G: Exhibit a significantly higher proportion of charged-off loans compared to fully paid loans, suggesting a higher risk associated with these loan grades.

Reasons for Default…

- **Strong Positive Correlations:**
- Loan Amount (loan_amnt) and Funded Amount (funded_amnt) have a near-perfect positive correlation (0.98), indicating that these two variables are almost identical.
- Loan Amount (loan_amnt) and Installment have a strong positive correlation (0.93), meaning larger loans tend to have higher installment payments.
- Funded Amount (funded_amnt) and Installment also have a strong positive correlation (0.96), reinforcing the relationship between loan size and installment.
- Total Payment (total_pymnt) and Funded Amount (funded_amnt) have a strong positive correlation (0.9), indicating that larger loans generally result in higher total payments.
- **Moderate Positive Correlations:**
- Interest Rate (int_rate) and Loan Amount (loan_amnt) have a moderate positive correlation (0.31), suggesting that larger loans tend to have slightly higher interest rates.
- Interest Rate (int_rate) and Term have a moderate positive correlation (0.45), indicating that longer loan terms are associated with higher interest rates.
- **Weak or No Correlations:**
- Annual Income (annual_inc) and most other variables show weak correlations, suggesting that income might not be a strong predictor of other loan characteristics in this dataset.
- Debt-to-Income Ratio (dti) also shows weak correlations with most variables, indicating limited influence on other loan factors.