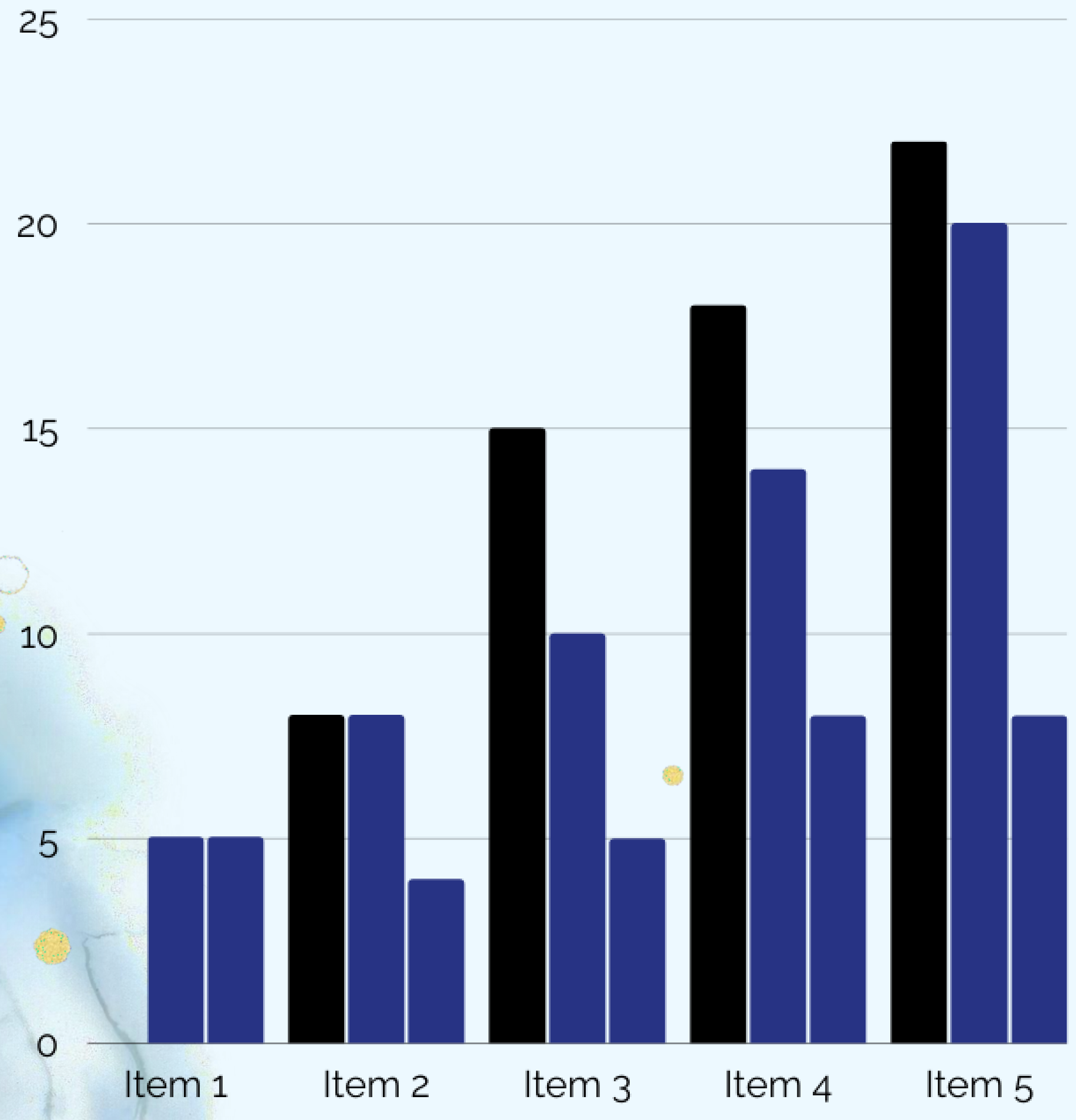
Lending Club Case Study



Presented By

Pallavi Gupta

Vasudha Srinivasaiah

Content

- Problem Statement
- Business Understanding
- Risk
- Analysis Approach
- Univariate Analysis
- Bivariate Analysis
- Conclusion

Problem Statement

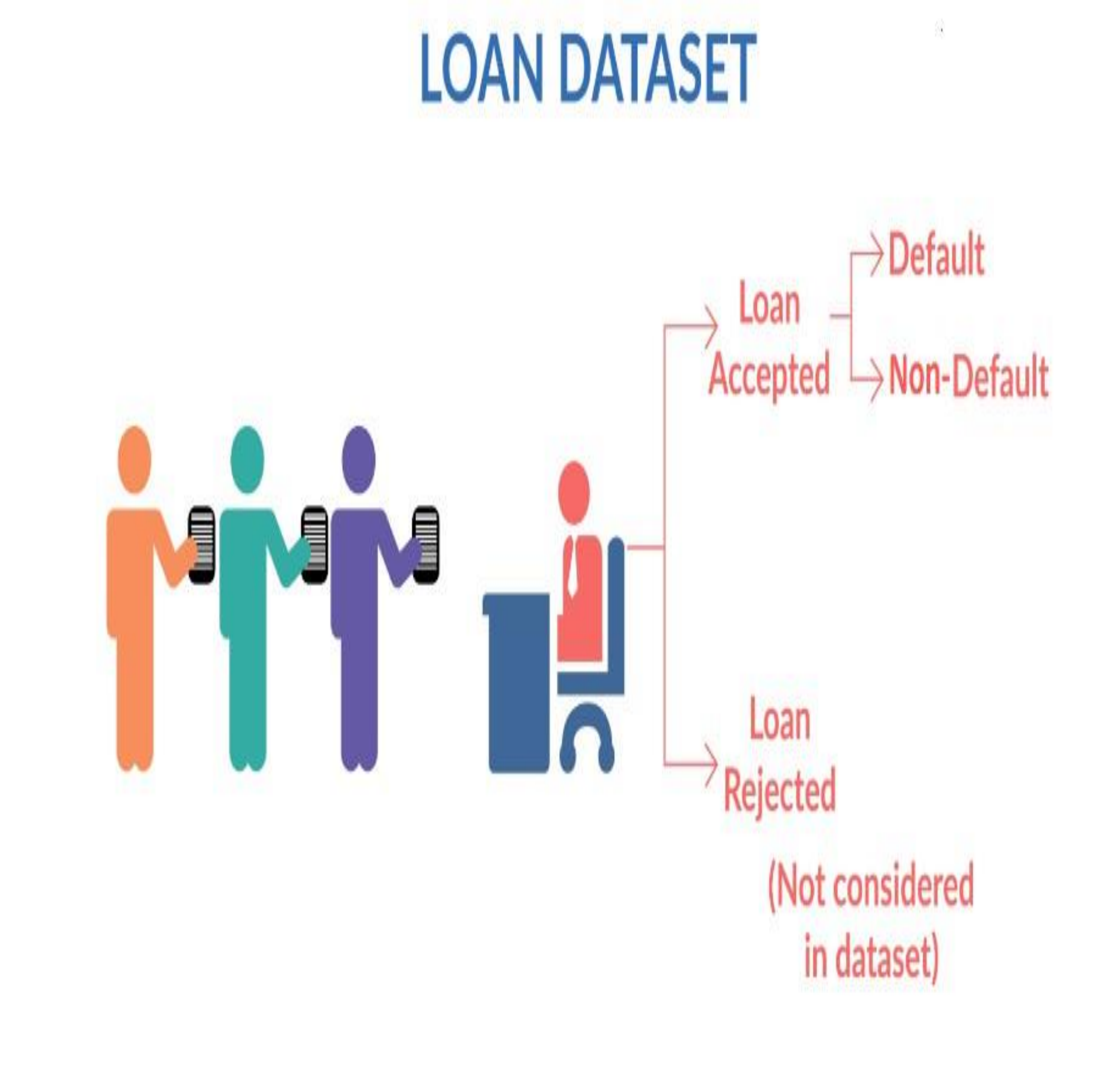
- Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment

Business Understanding

When a person applies for a loan, there are **two types of decisions** that could be
taken by the company:

1. **Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below:
    - **Fully paid**:
    - **Current**:
    - **Charged-off**:
2. **Loan rejected**: The company had rejected the loan (because the candidate does not meet their requirements etc.).



Risk

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.

Analysis Approach

Steps

Step1 – Data Cleaning
- Null value treatment
- Data type conversion
- Remove unwanted characters
- Outlier treatment

Step2 – Univariate Analysis
- Distribution
- Box plots
- Histograms

Step3 – Bivariate Analysis
- Distributions with respect to loan status

Step4 – Bivariate - Busines Case
- Pivote table
- Distribution of charge off proportion

Step5 – Conclusion

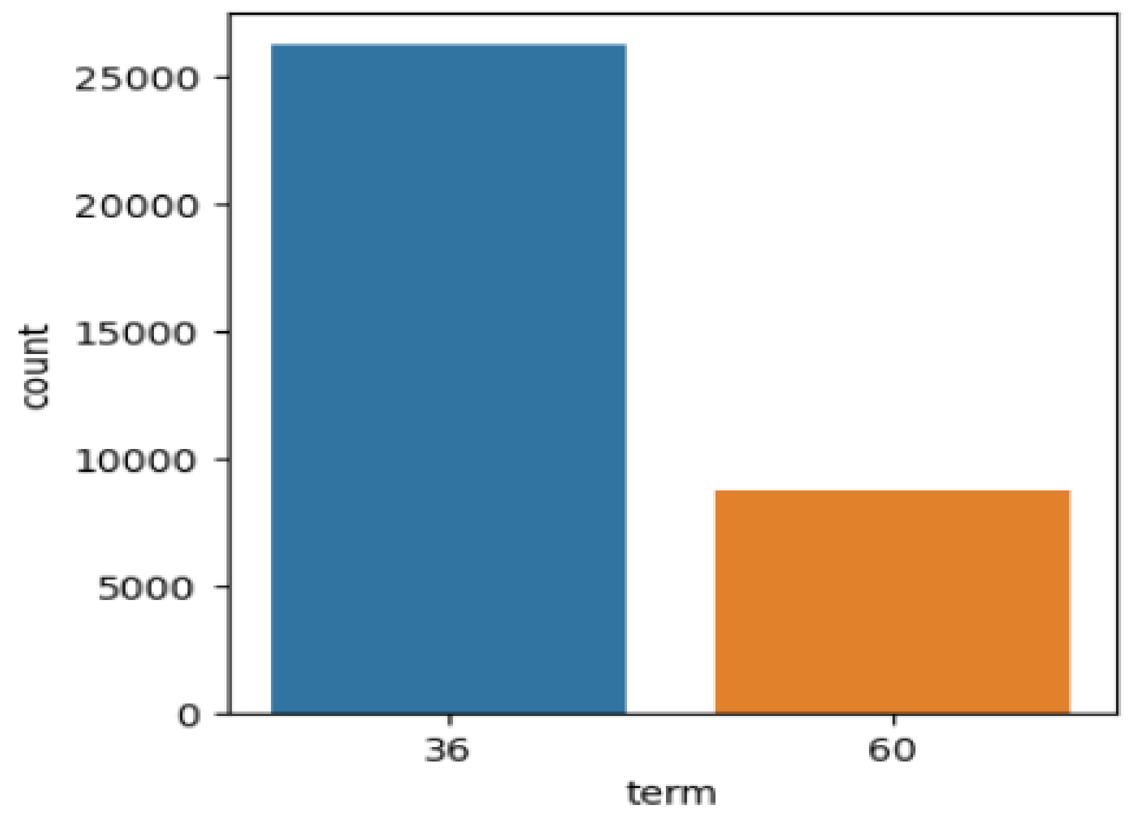
Data Cleaning and Manipulation

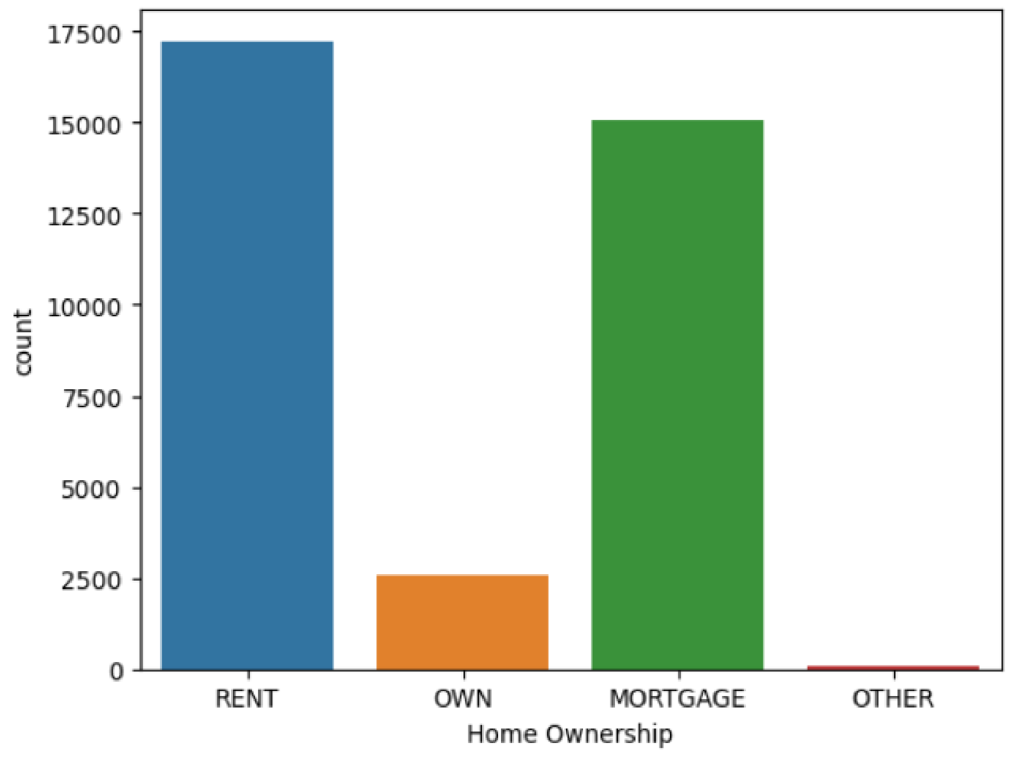
- Finding Duplicates
- Checking for Null values and dropping the columns with more than 50% NULL values
- Dropping columns with behavioral data
- Dropping columns with all unique values
- Dropping columns with 1 value
- Extracting new columns for year and month for column issu_date and earliest_cr_line.

Univariate Analysis

Term

- Loans were applied more for 36 months as compared to 60 months



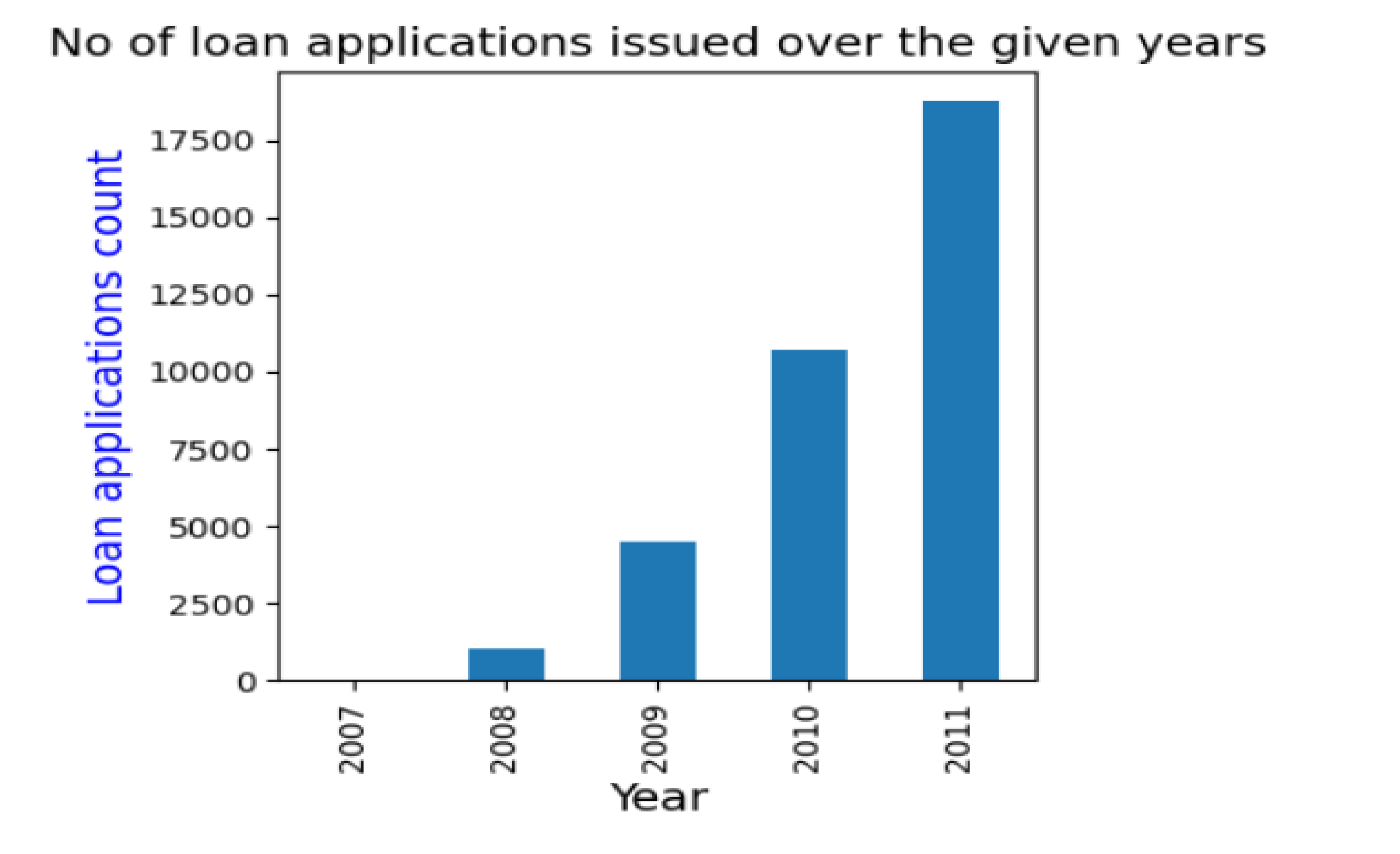


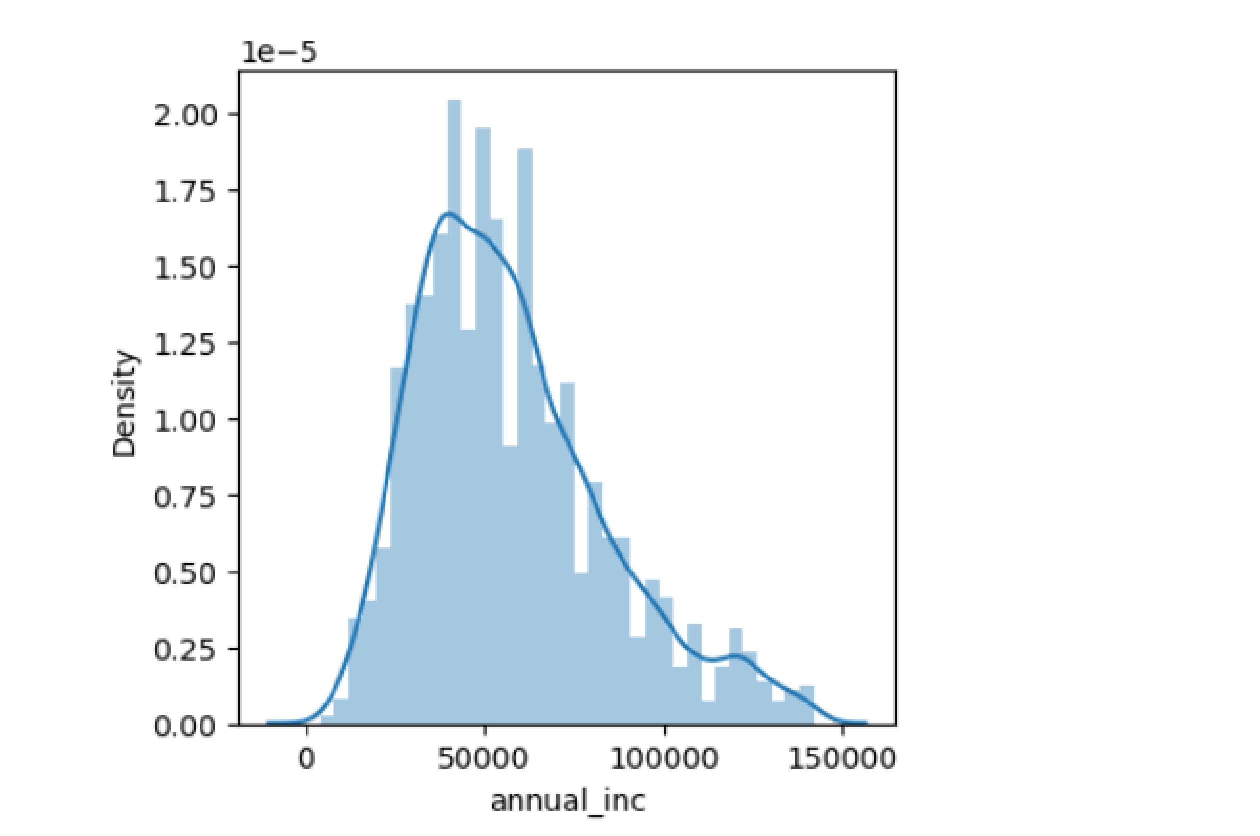
Home Ownership

- The people who applied for loan were having home ownership as rent

Year

- Loan applications increased every year and almost doubling every year

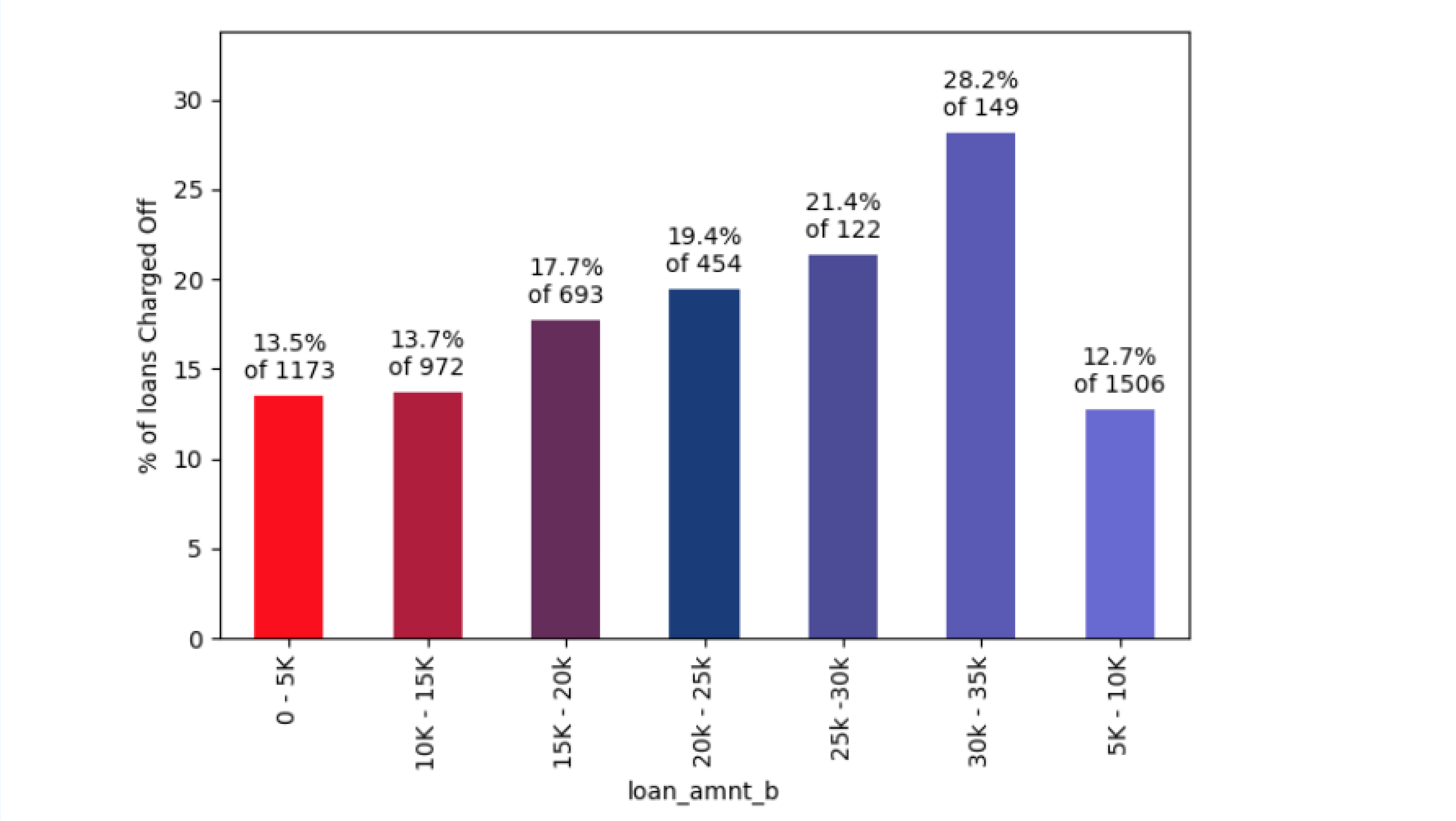




Annual Income

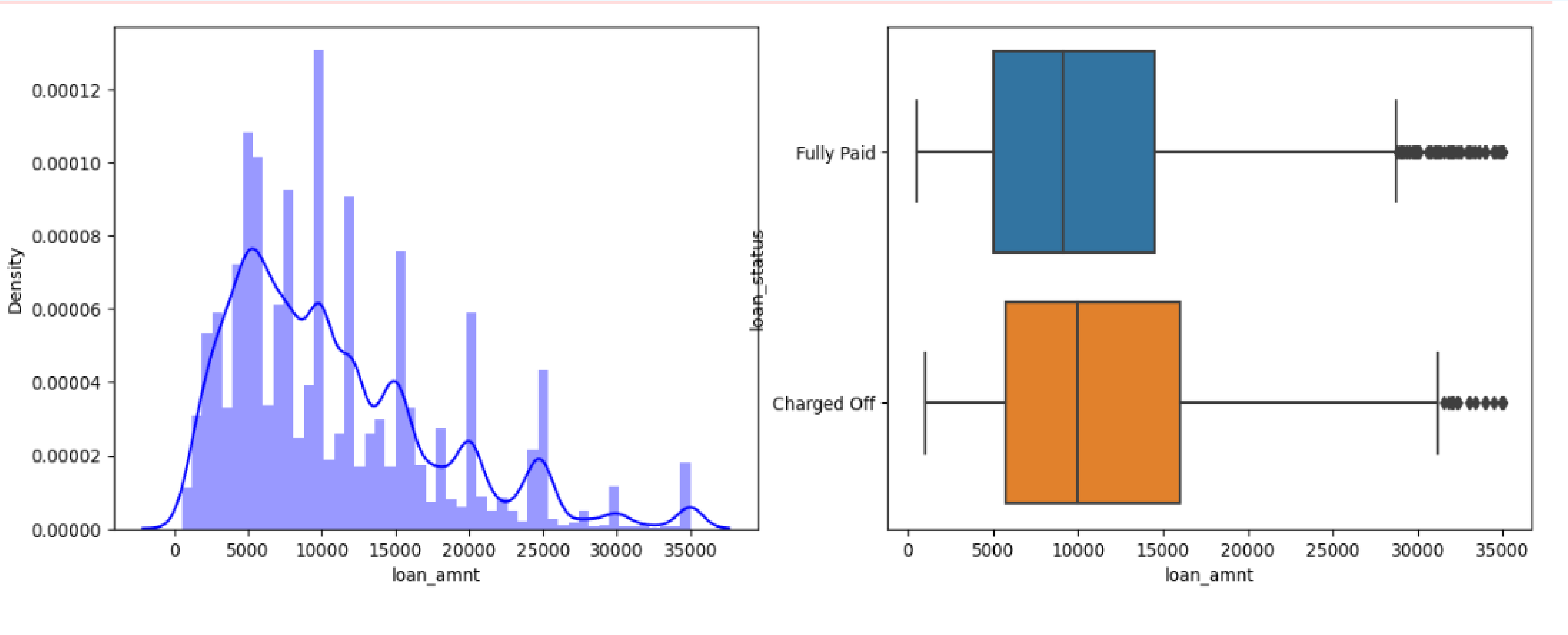
- Max loans were applied by applicants having annual income 30000-50000

Most of the loans were between 30k-35k and defaulter ratio is also high for this range



Loan Amount

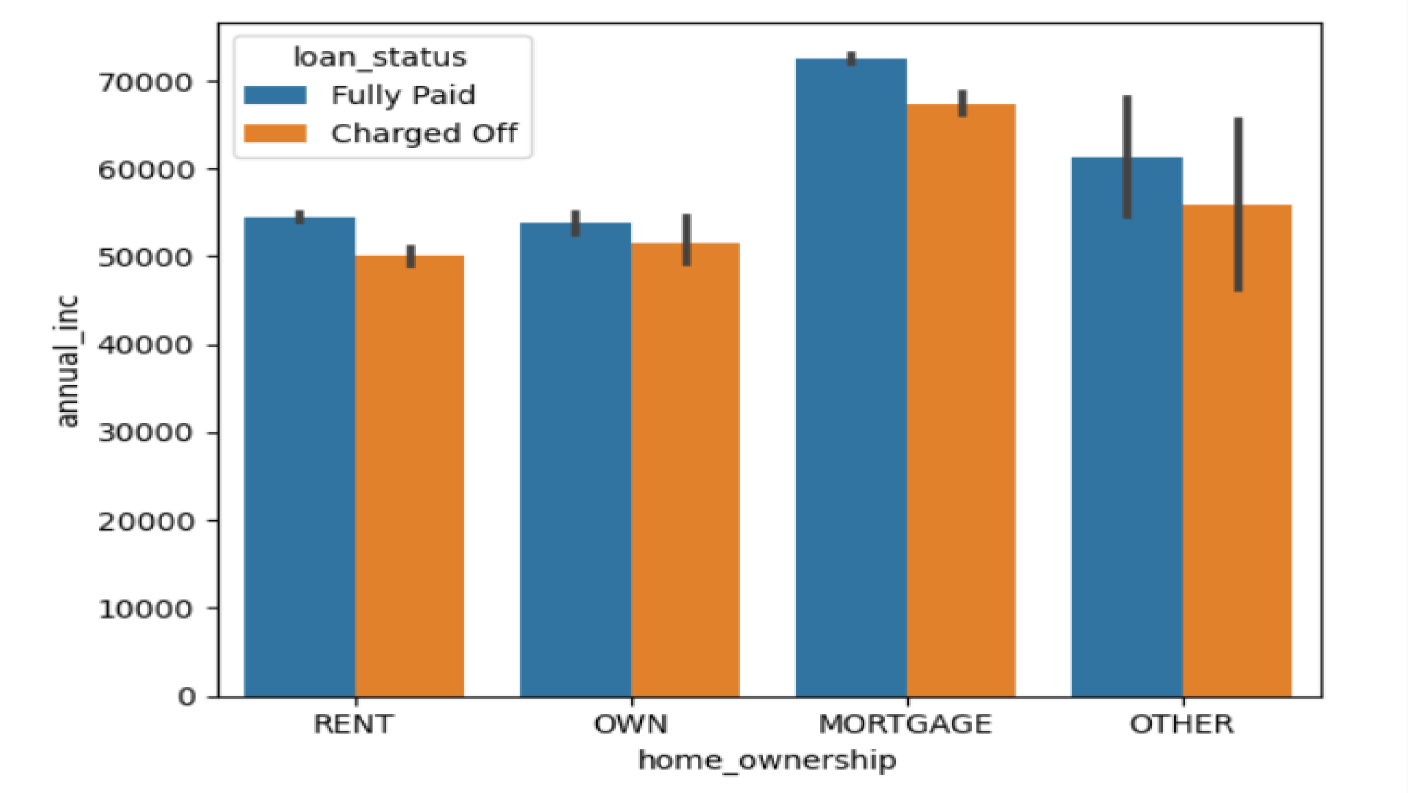
- The applicants applied for 5000-10000 loan amount most

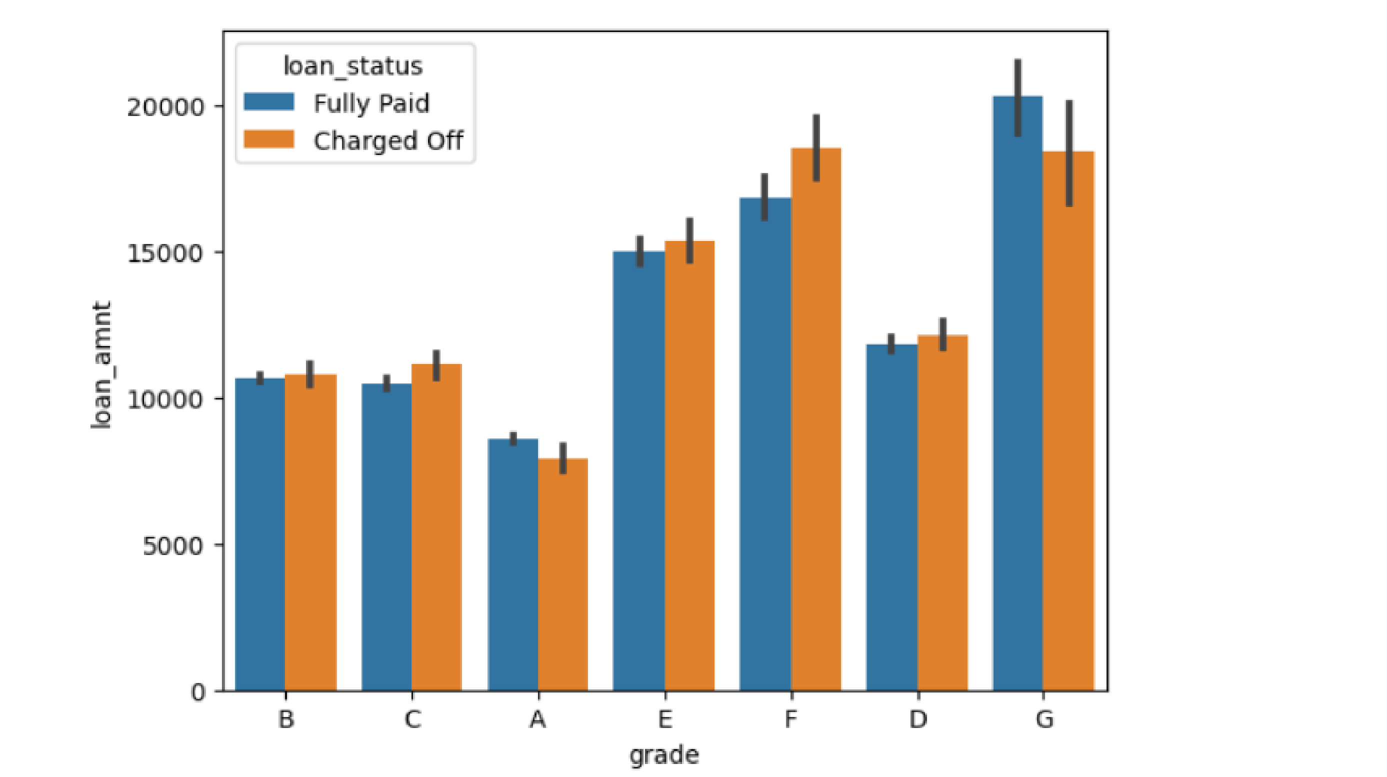


Bivariate Analysis

Home Ownership Vs Annual Income

Applications with Mortgage had high defaulters

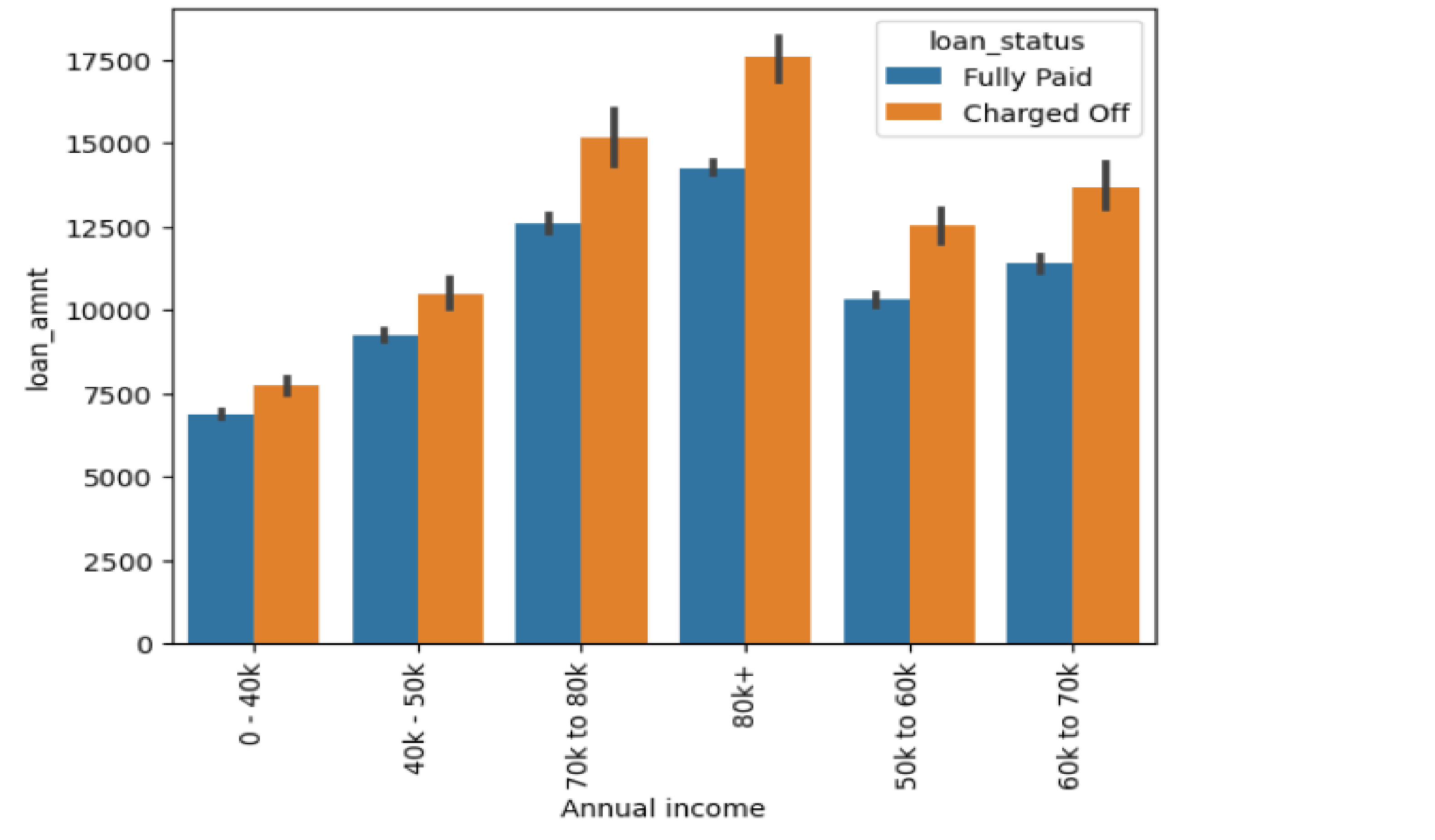


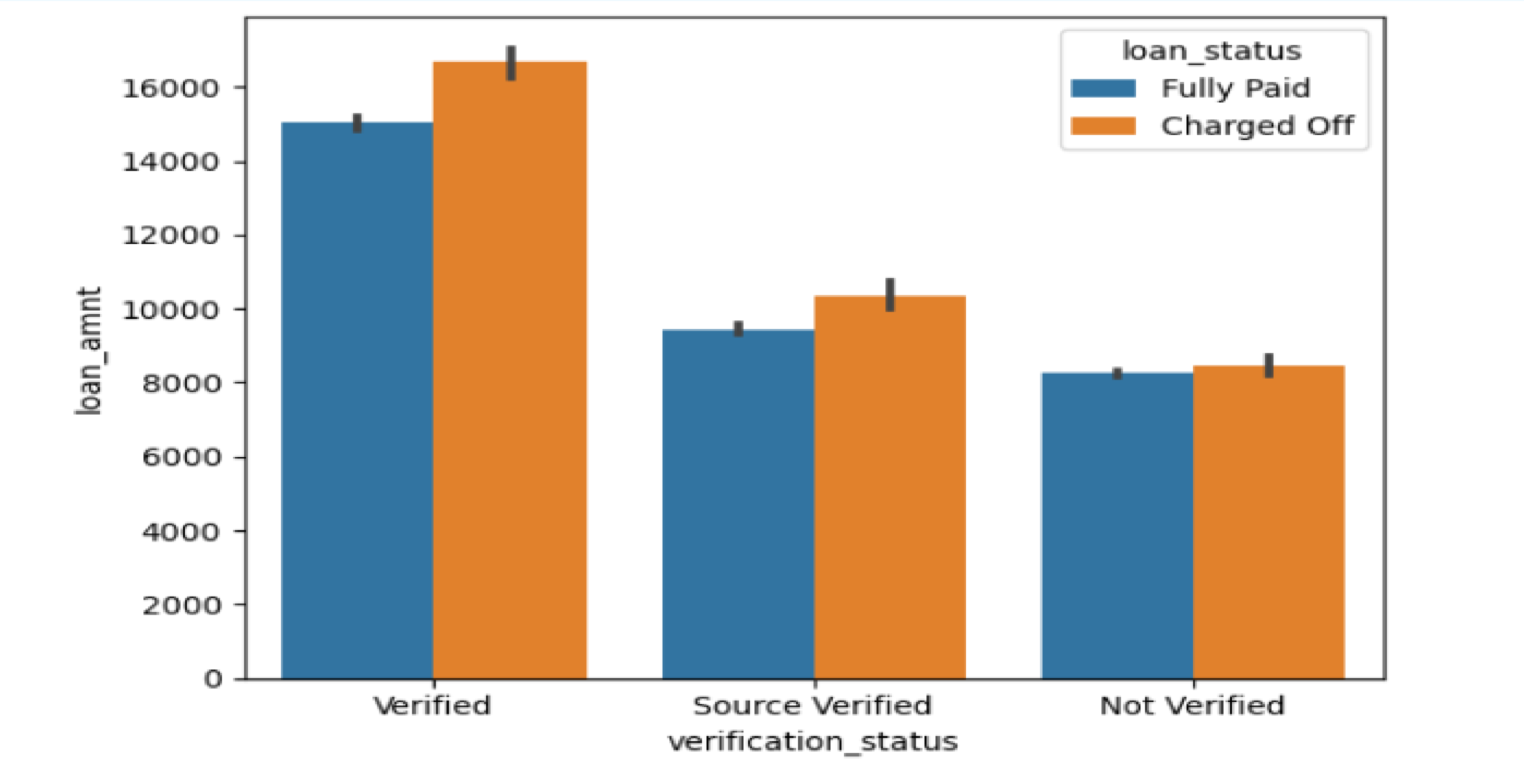


- As the Grade decreases (A B C D E F G) default rate increases

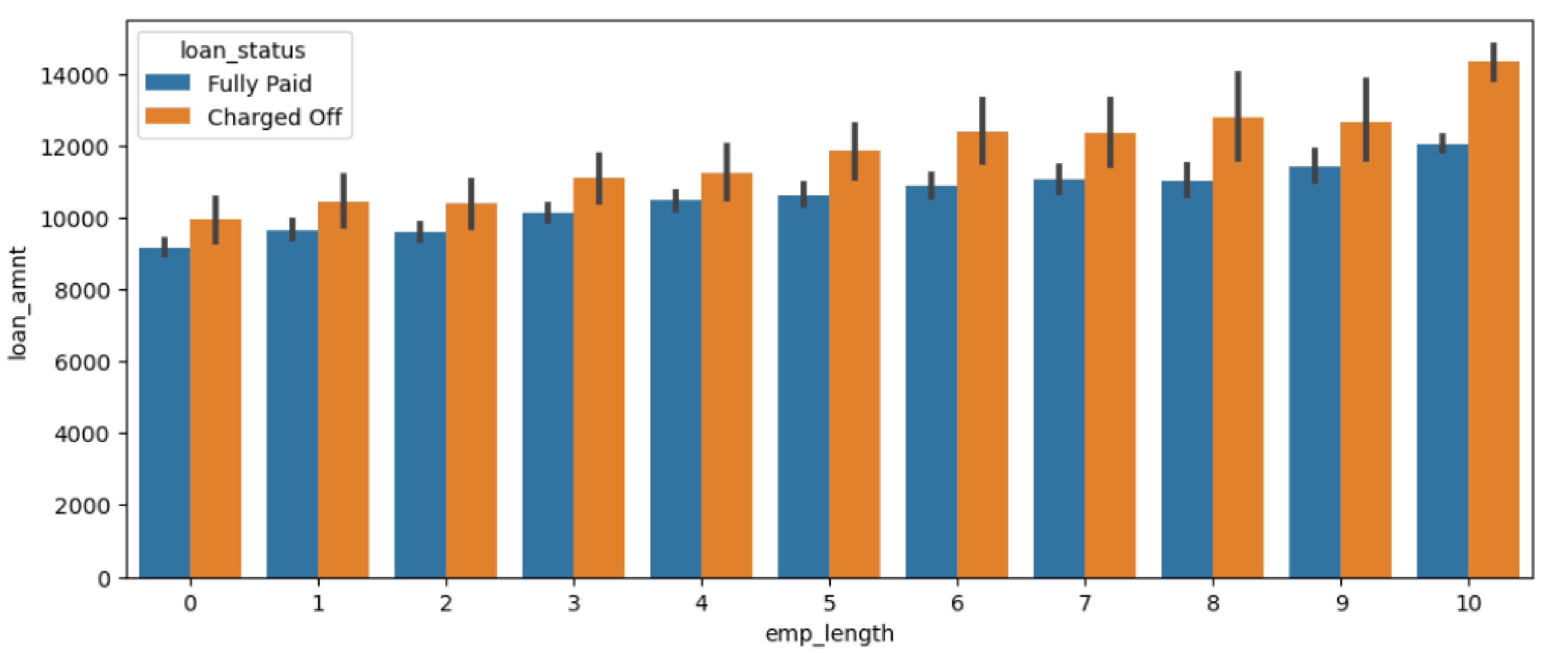
Annual Income vs Loan Amt on Loans data set (Paid or Charged Off)

Most of the loans were between 30k-35k and defaulter ratio is also high for this range

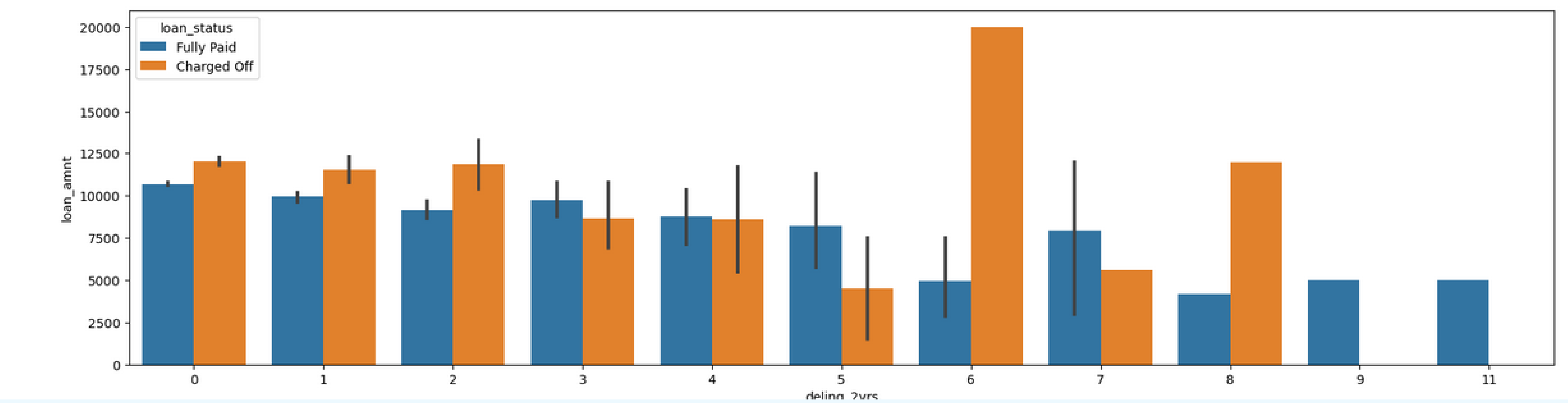




Verified Loans with higher loan amounts reflected higher Charged-offs.



The higher loan amounts with > 10 years of Emp work tenure had higher defaulters

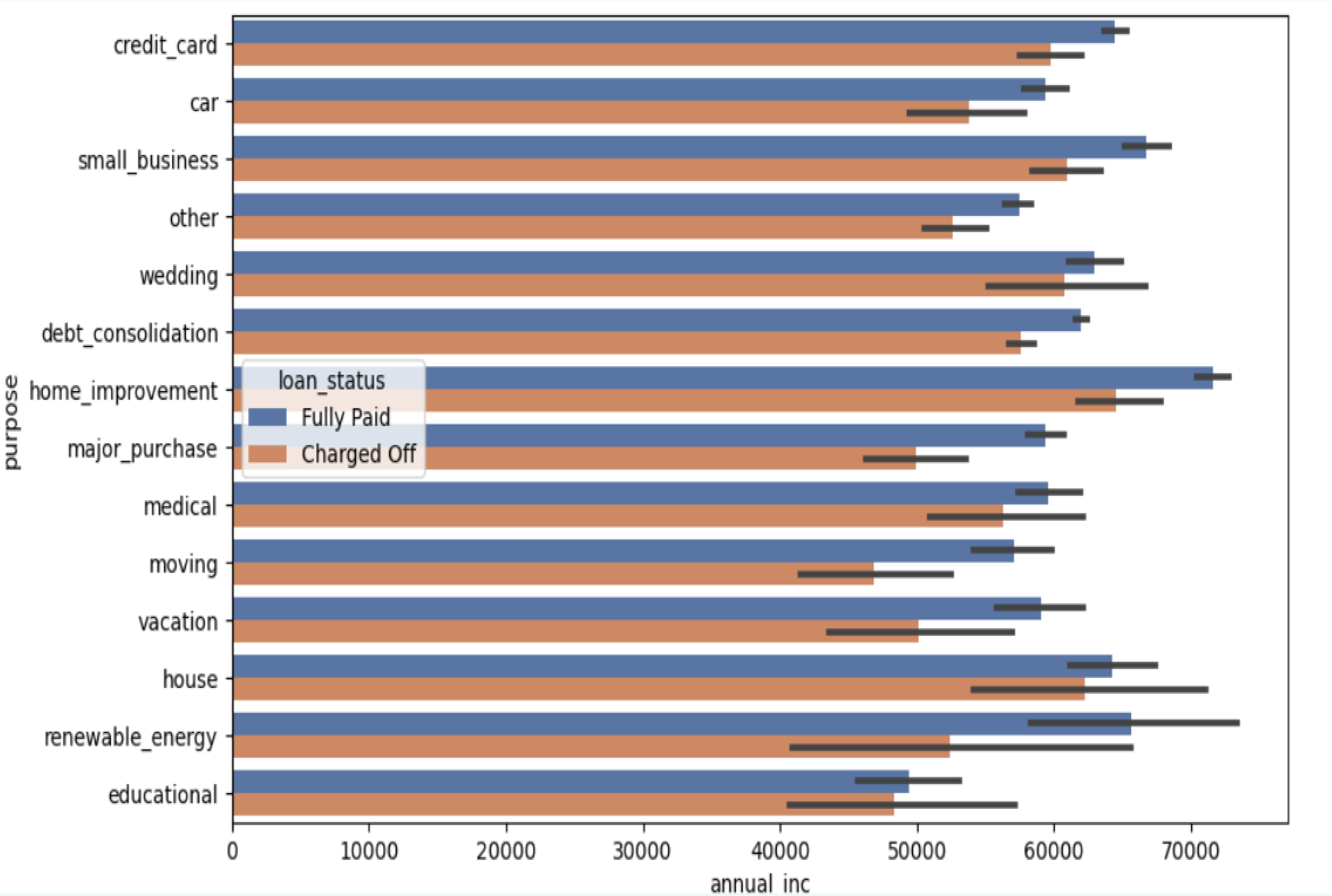


The number of loans issued decreases as the number of delinquencies increases. Loans should be restricted and not granted if there are two or more delinquencies. Also the charged-off is high while delinquencies is 6

Loan Purpose Analysis

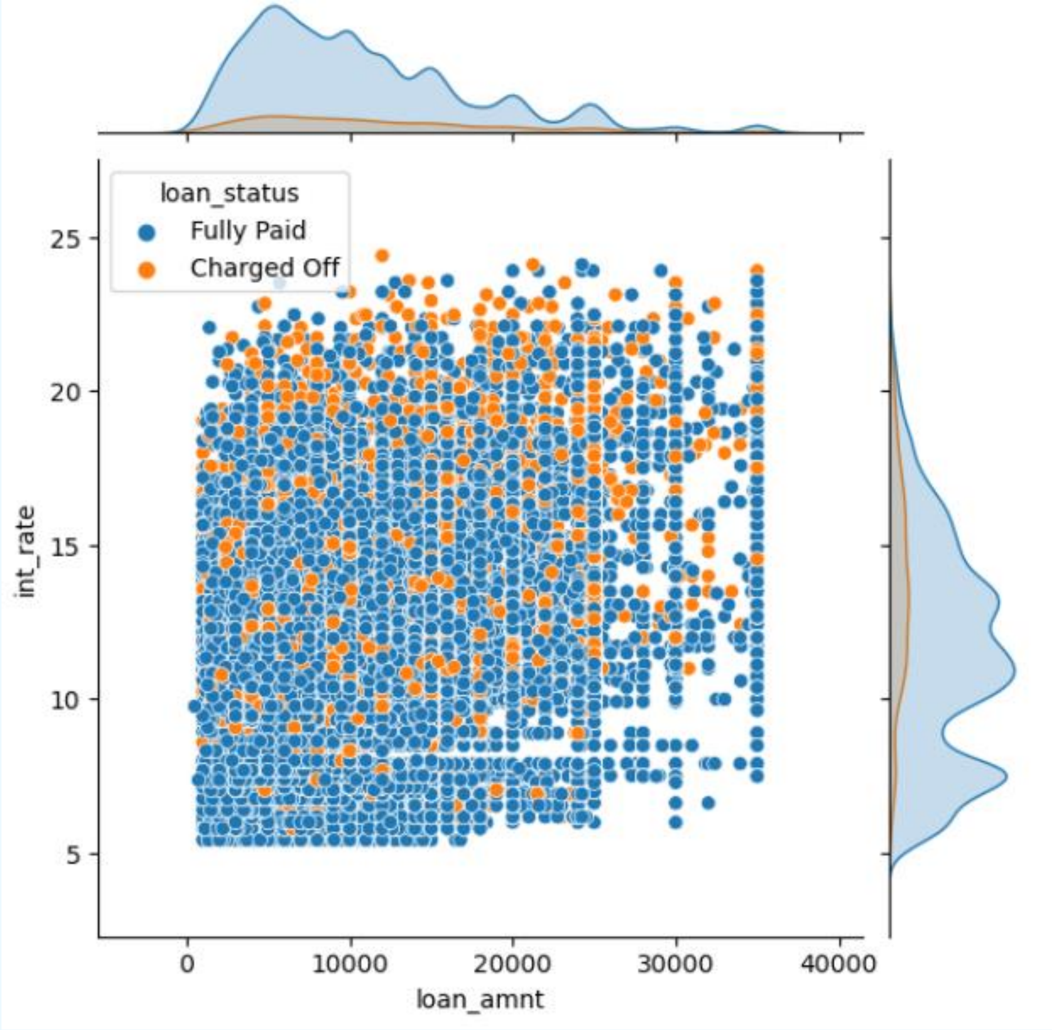
Though the number of loans applied and defaulted are the highest in number for paying debts, The annual income of those who applied isn't the highest. Applicants with higher salary mostly applied loans for their

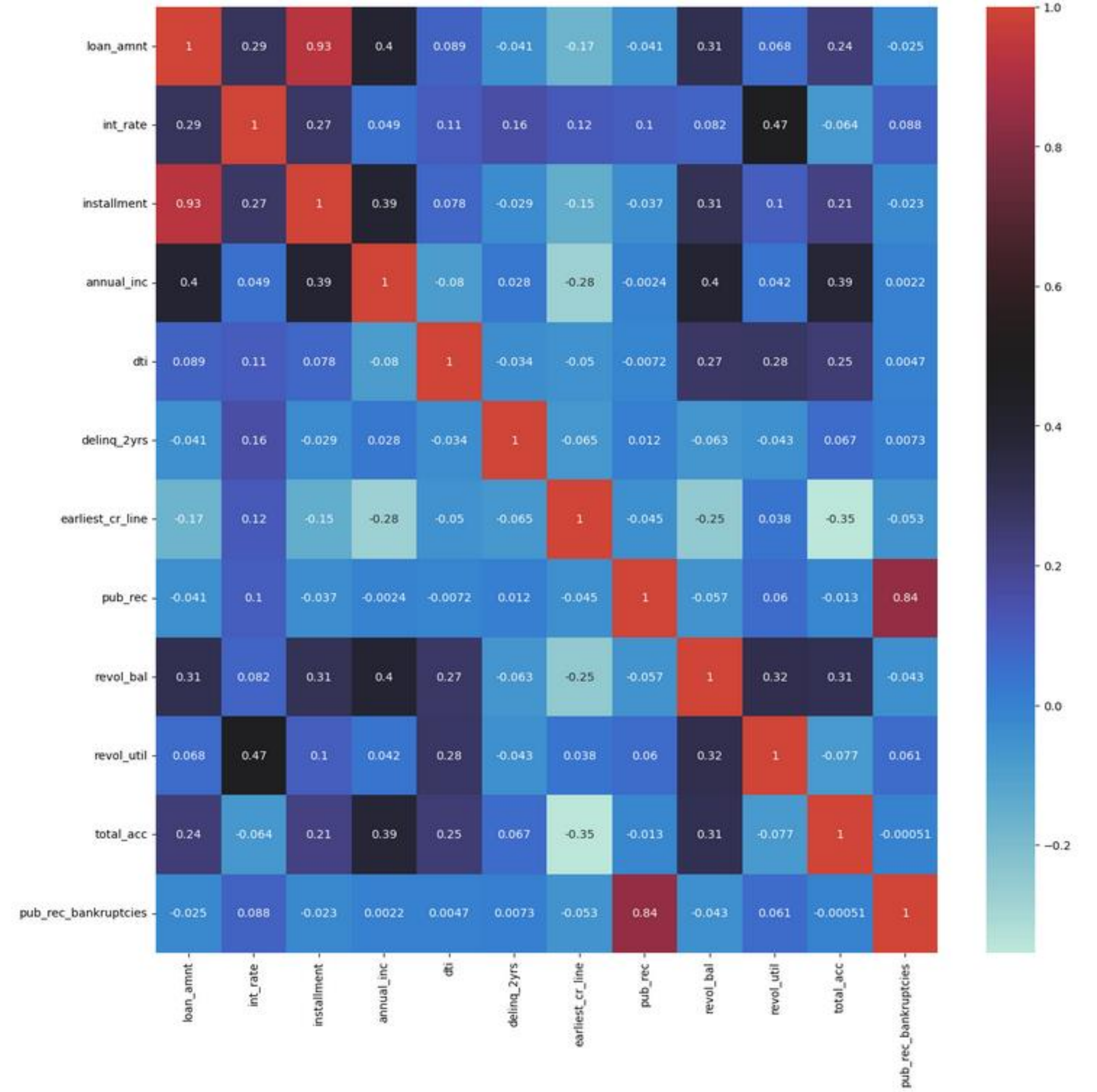
- Home renovation
- Buy new house
- Small business units



Correlation Matrix

- Correlation between loan amount, last paid amount interest rate, as the corporation charges higher interest rates for riskier loans.
- Also it indicates that Annual income and debt-to-income ratio (DTI) are adversely connected, but No link was discovered between delinquencies and the public record of bankruptcy.





Observations

- Across all the income groups, the loan amount is higher for people who defaulted. From the provided dataset, approximately 58% of loans were defaulted.
- The number of loans issued decreases as the number of delinquencies increases. Loans should be restricted and not granted if there are two or more delinquencies. The defaulters were high while delinquencies for the applied customers is 6.
- It’s seen that the interest rates are high for the defaulted loans.
- When the loan amount is high and interest rates are high, though there is a spread of Annual income from medium to high, the charged-off proportion is high.
- For approx 8 % of those seeking loans, 2000 had earliest credit line year. Fresher applicants had a greater default rate than the prev ones. The recession around 2000 may have had effect on the increased defaulters post year 2000.

Conclusion

- Verified loan applicants are defaulting more than those who are not verified. The company should review its verification process to ensure it effectively assesses applicant creditworthiness and consider improvements or adjustments.
- Applicants living in rented or mortgaged houses are more likely to default. This information can be considered in the underwriting process to assess housing stability and its impact on repayment ability.
- The overall default rate is approximately 14%. Any variable that results in default rate >16.5% can be termed as Business Risk
- Low income range has higher default proportion