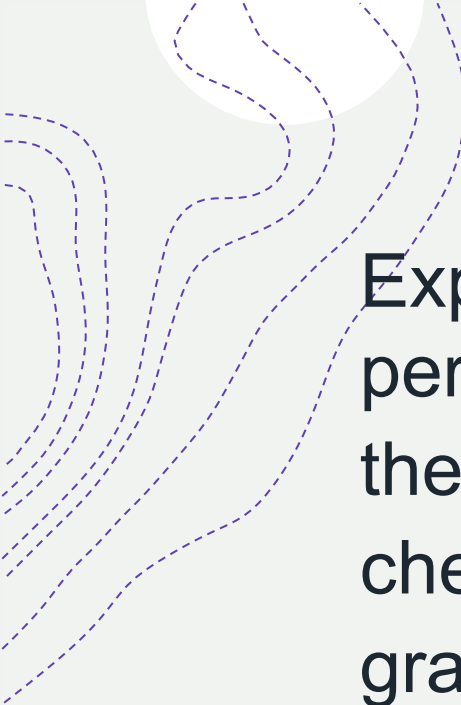


Lending Club Case Study

By Anupam Jain & Neelam Tripathi ML C59

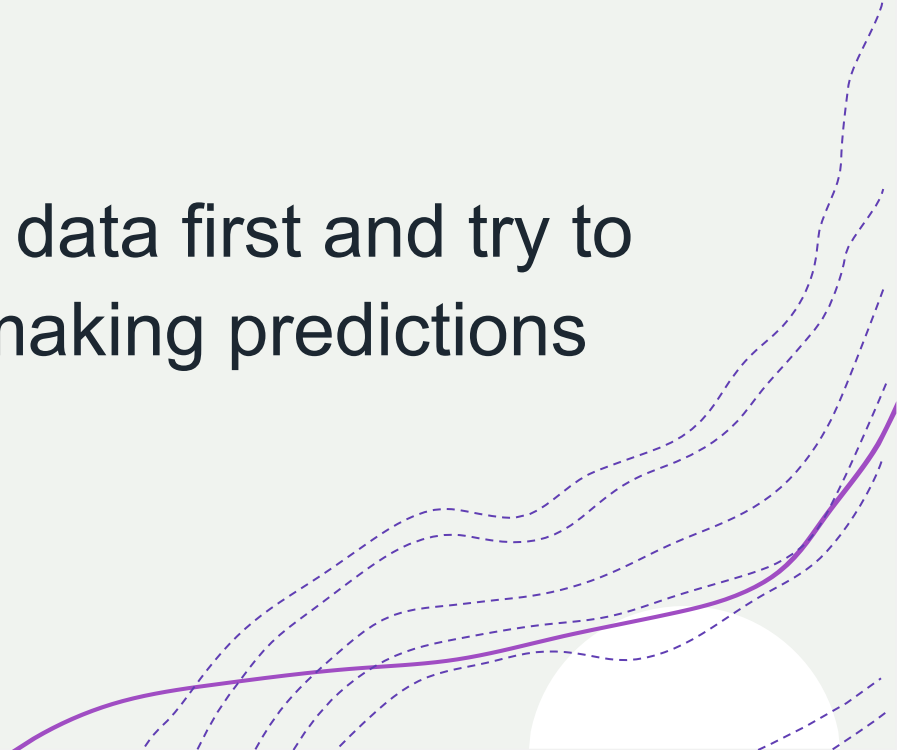


Exploratory Data Analysis (EDA)



Exploratory Data Analysis is the process of performing investigations or analysis on data to discover the patterns or to summarize their main characteristics or to check assumptions with the help of summary statistics and graphical representations.

It is a good approach to understand the data first and try to gather as many insights from it before making predictions



EDA explained on Lending Dataset

Data Dictionary : Our data has 86999 rows and 116 columns

1	LoanStatNew	Description
2	acc_now_delinq	The number of accounts on which the borrower is now delinquent.
3	acc_open_past_24mths	Number of trades opened in past 24 months.
4	addr_state	The state provided by the borrower in the loan application
5	all_util	Balance to credit limit on all trades
6	annual_inc	The self-reported annual income provided by the borrower during registration.
7	annual_inc_joint	The combined self-reported annual income provided by the co-borrowers during registration
8	application_type	Indicates whether the loan is an individual application or a joint application with two co-borrowers
9	avg_cur_bal	Average current balance of all accounts
10	bc_open_to_buy	Total open to buy on revolving bankcards.
11	bc_util	Ratio of total current balance to high credit/credit limit for all bankcard accounts.
12	chargeoff_within_12_mths	Number of charge-offs within 12 months
13	collection_recovery_fee	post charge off collection fee
14	collections_12_mths_ex_med	Number of collections in 12 months excluding medical collections
15	delinq_2yrs	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2
16	delinq_amnt	The past-due amount owed for the accounts on which the borrower is now delinquent.
17	desc	Loan description provided by the borrower
18	dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations,
19	dti_joint	A ratio calculated using the co-borrowers' total monthly payments on the total debt obligations, excluding r
20	earliest_cr_line	The month the borrower's earliest reported credit line was opened
21	emp_length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and
22	emp_title	The job title supplied by the Borrower when applying for the loan.*
23	fico_range_high	The upper boundary range the borrower's FICO at loan origination belongs to.
24	fico_range_low	The lower boundary range the borrower's FICO at loan origination belongs to.
25	funded_amnt	The total amount committed to that loan at that point in time.

Steps performed for EDA

Step 1 : Importing the data using pandas

Step 2 : Checking basic features of data such as shape , data types , null values

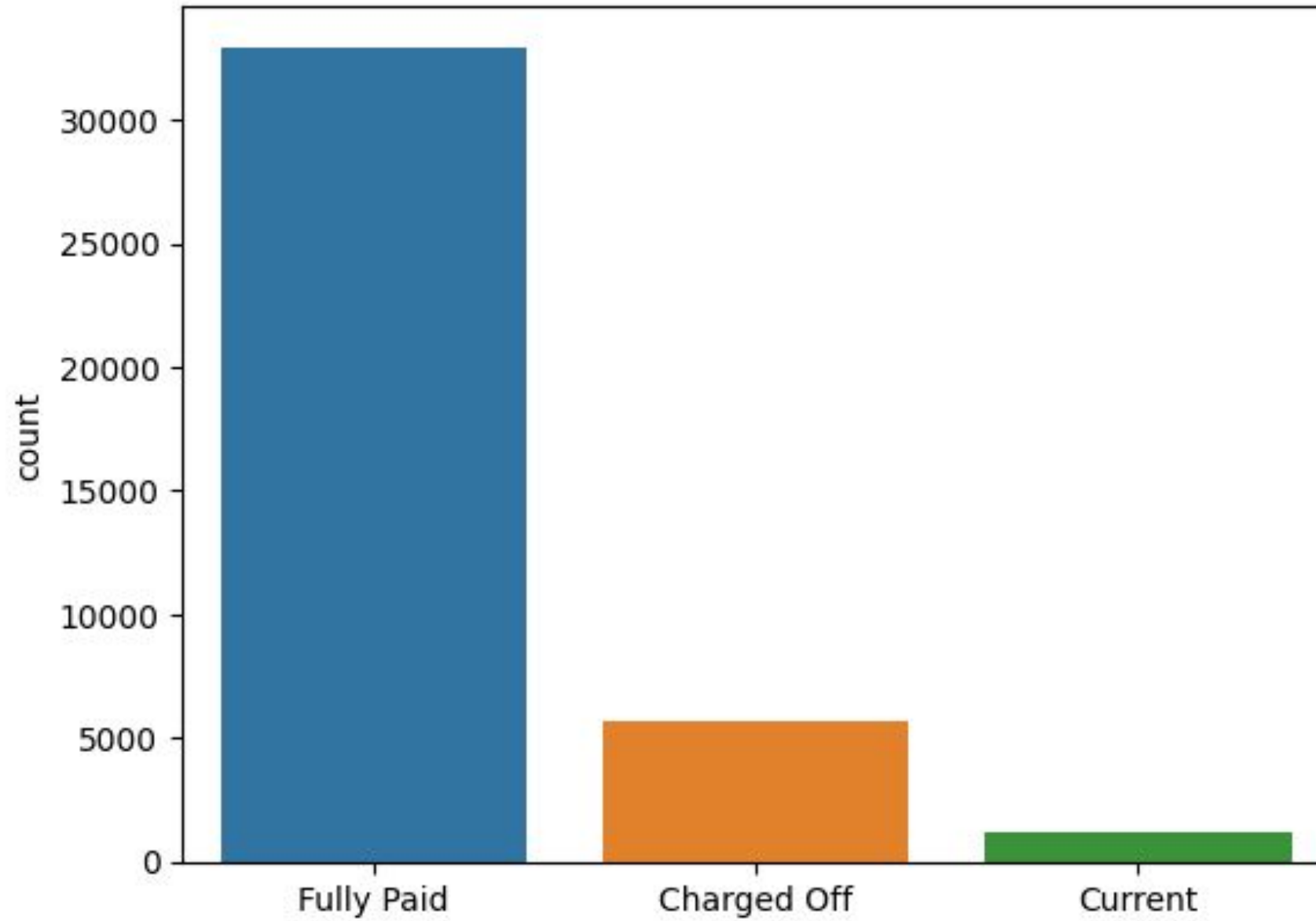
Step 3 : Treating null values

Step 4 : Dropping unnecessary columns

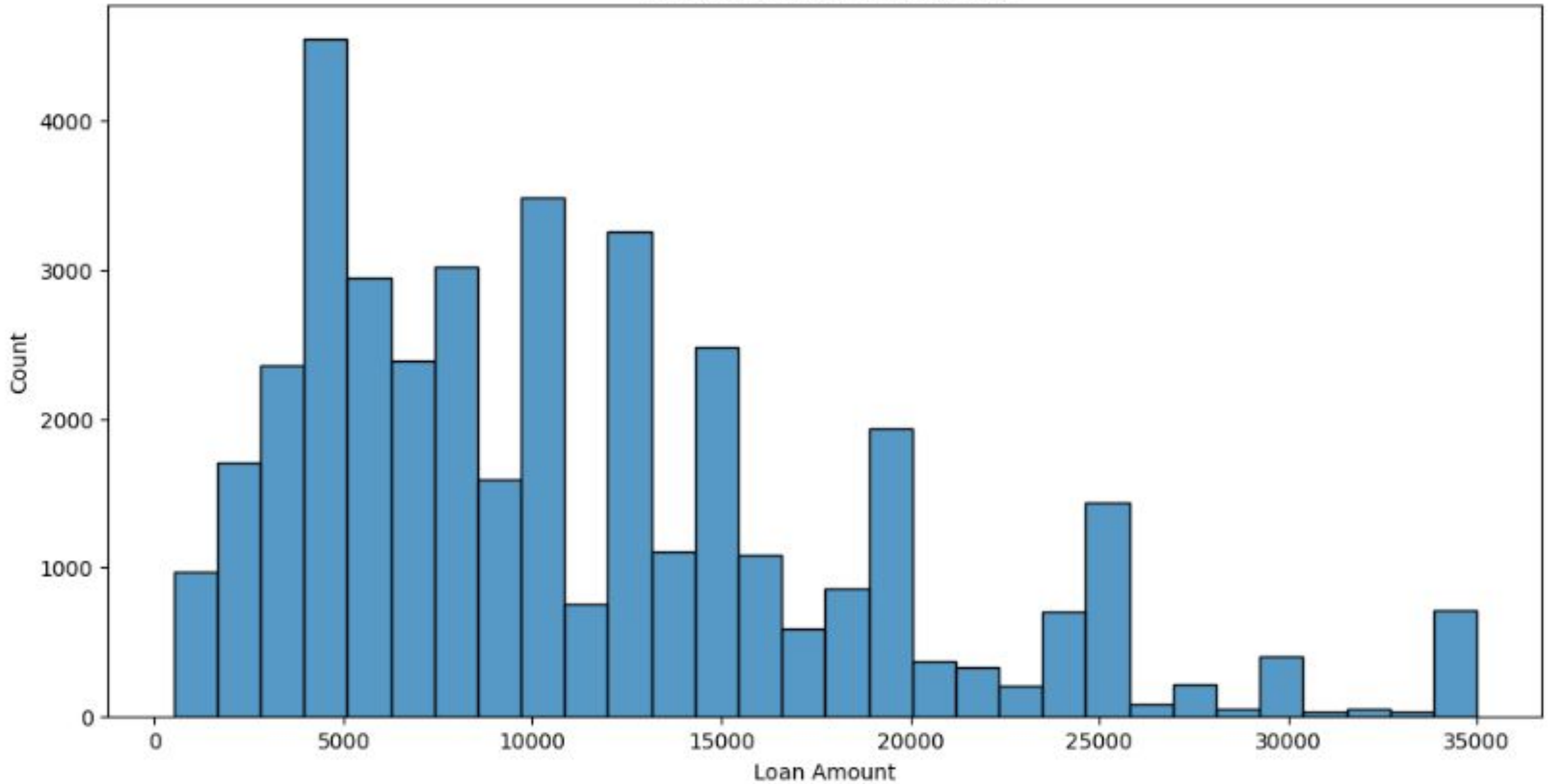
Step 5 : Checking and treating Outliers

Step 6 : Performing Analysis and Visualization on the Data

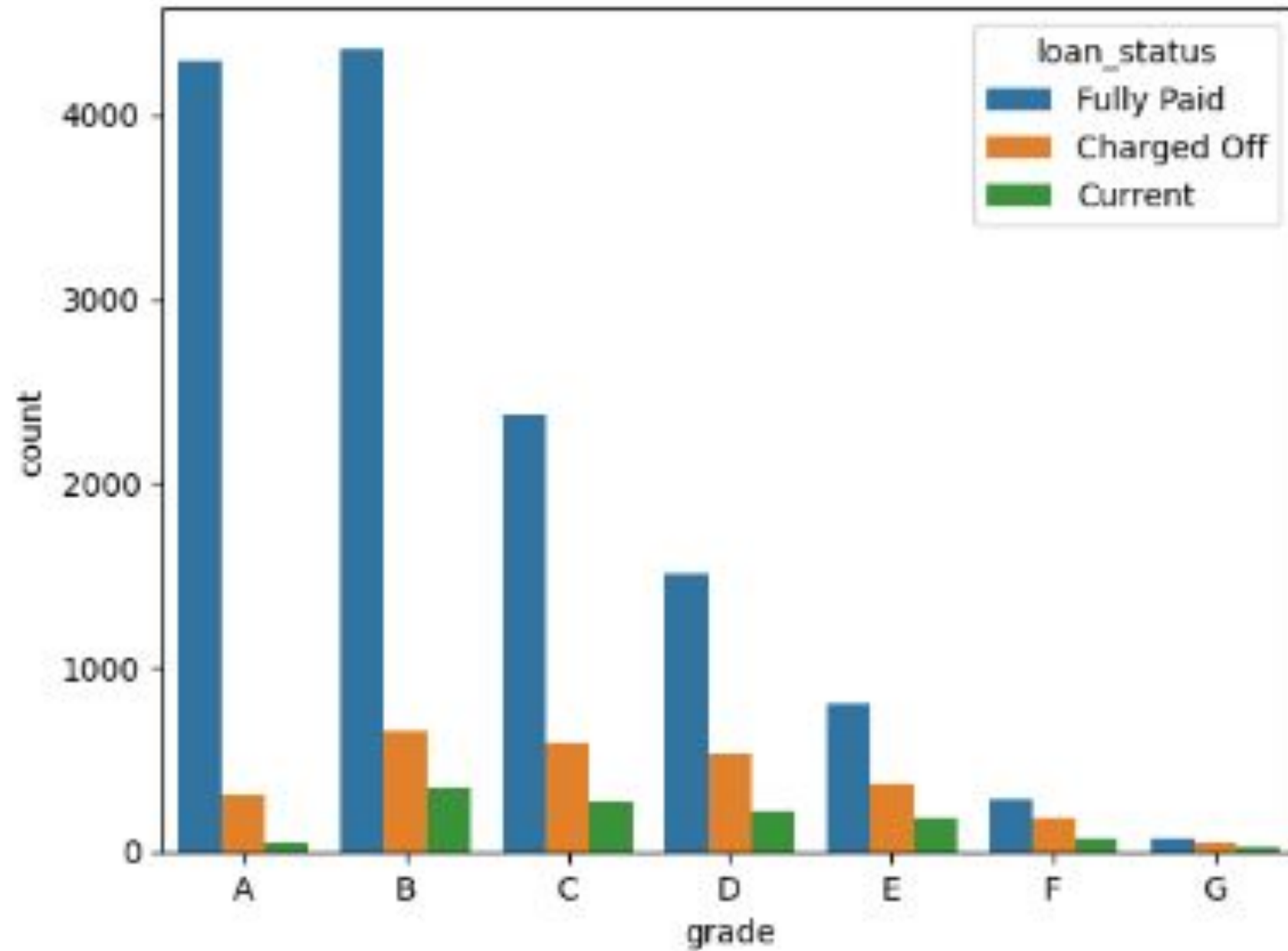
Distribution of Loan Status



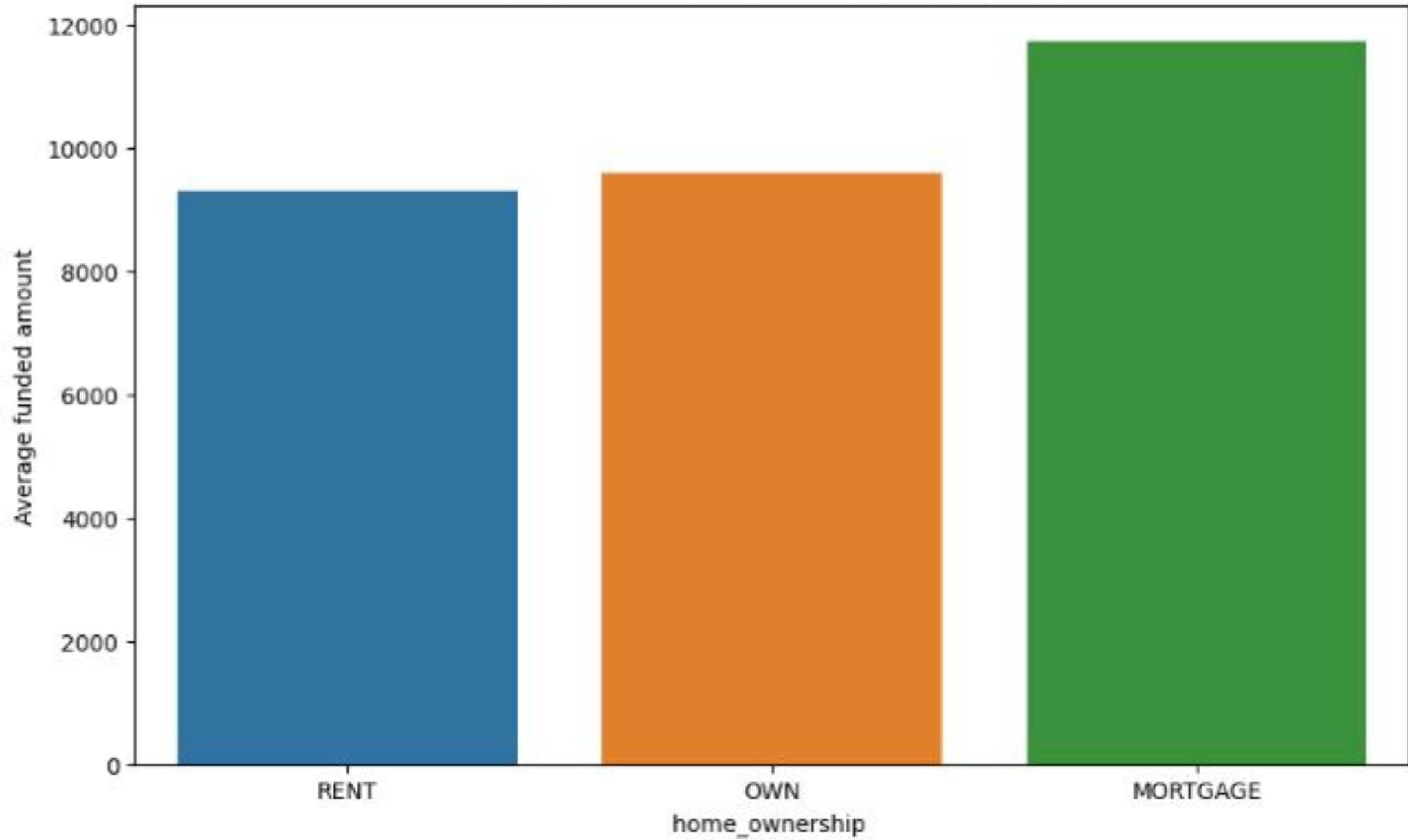
Distribution of Loan Amounts



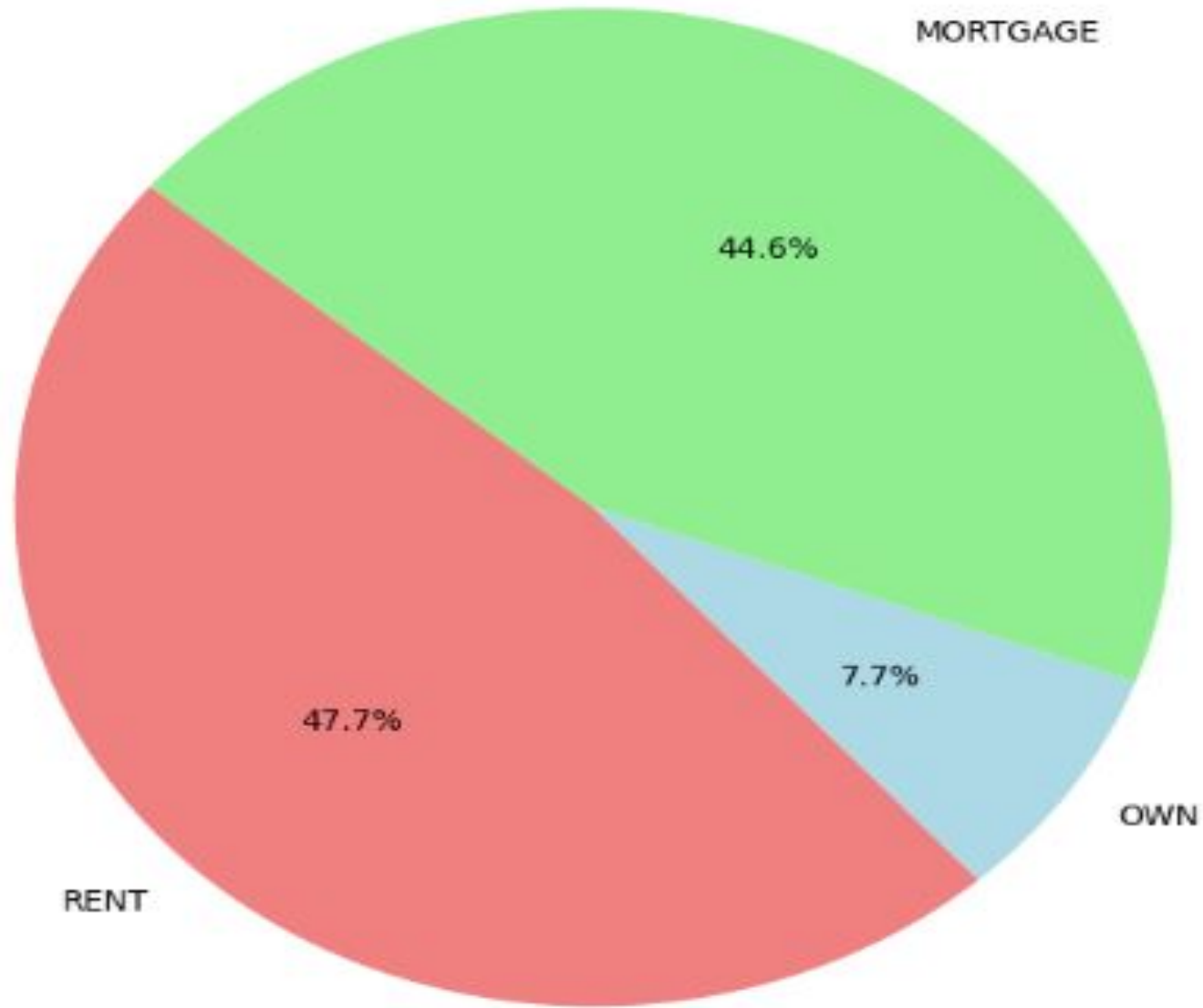
Loan Status Based on Grade



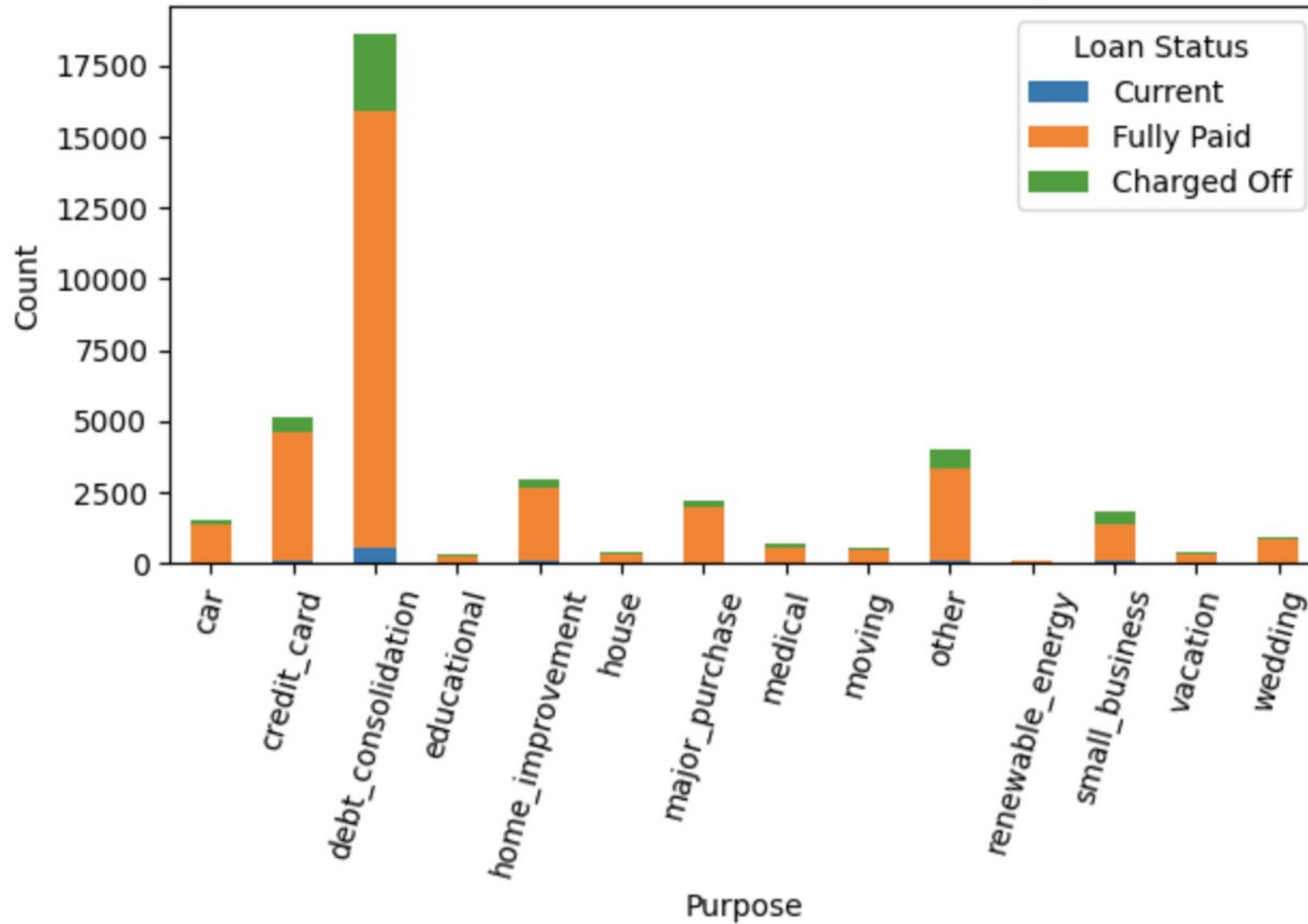
Impact of home ownership on funded amount



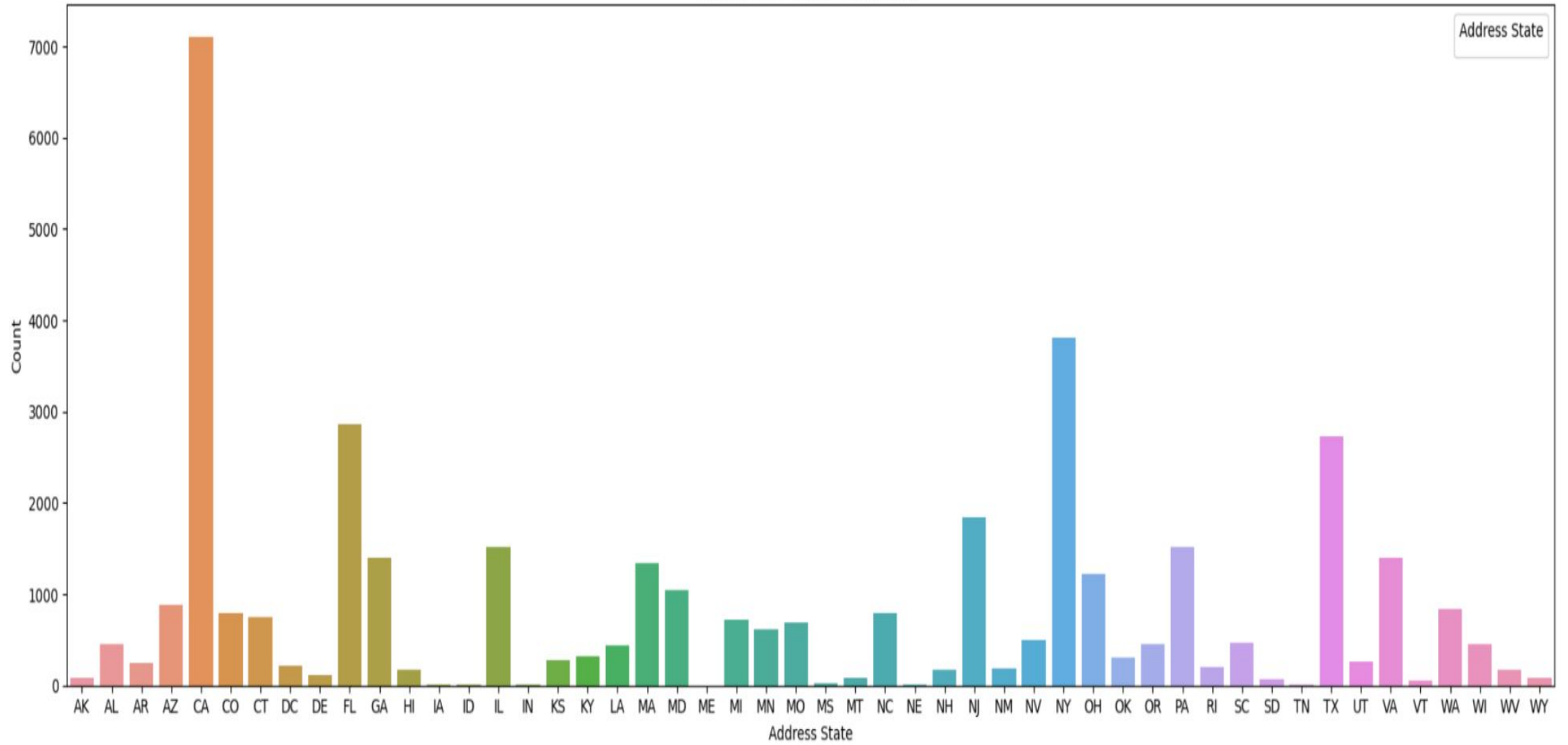
Distribution of Home Ownership



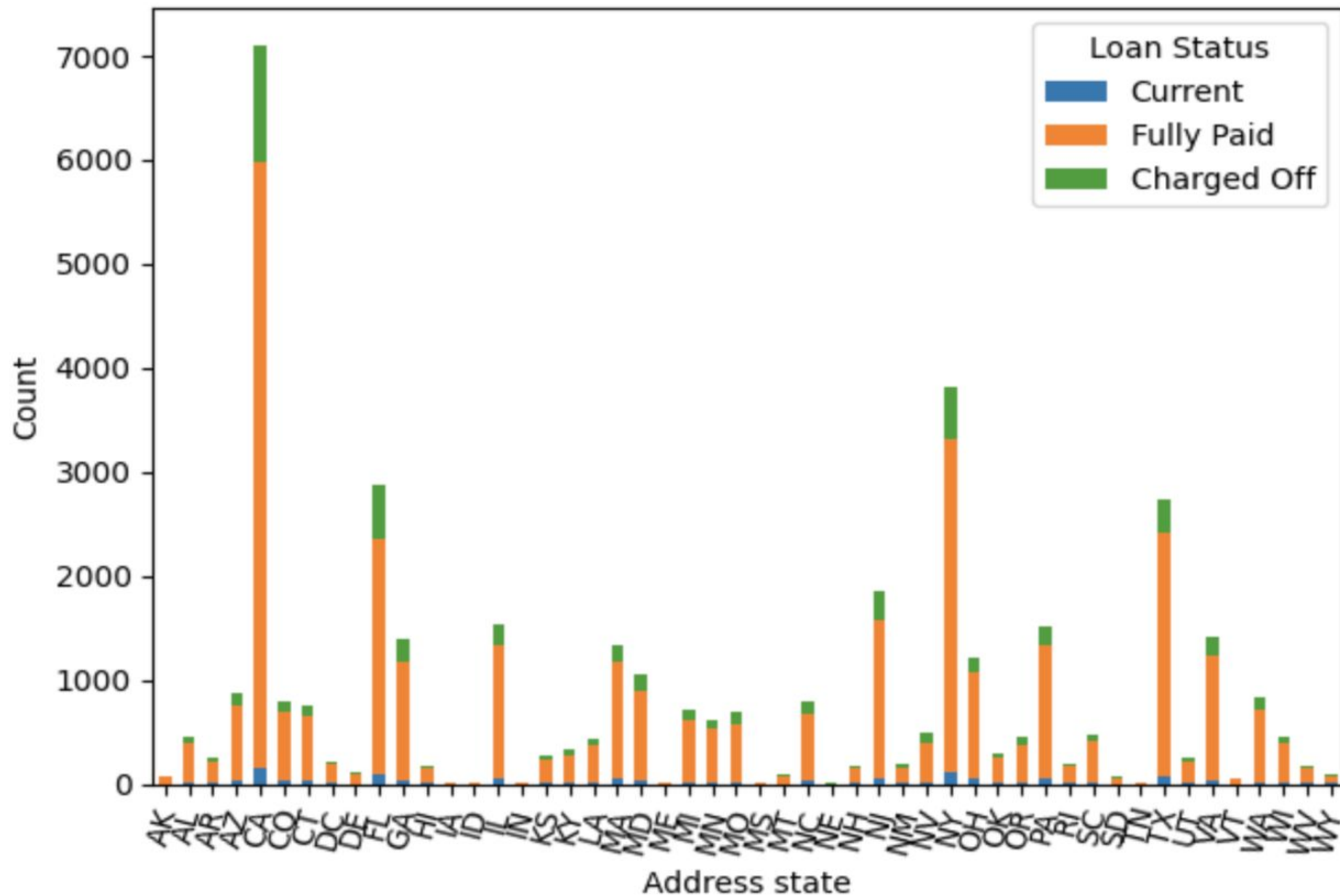
Loan Status Distribution for Each Purpose



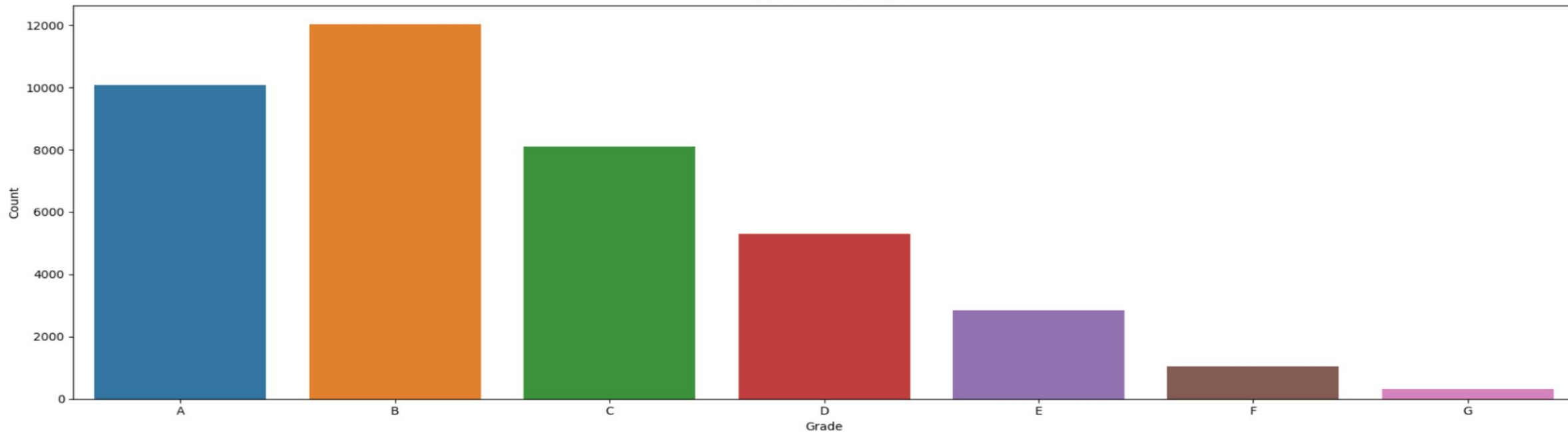
Loan adress state Distribution



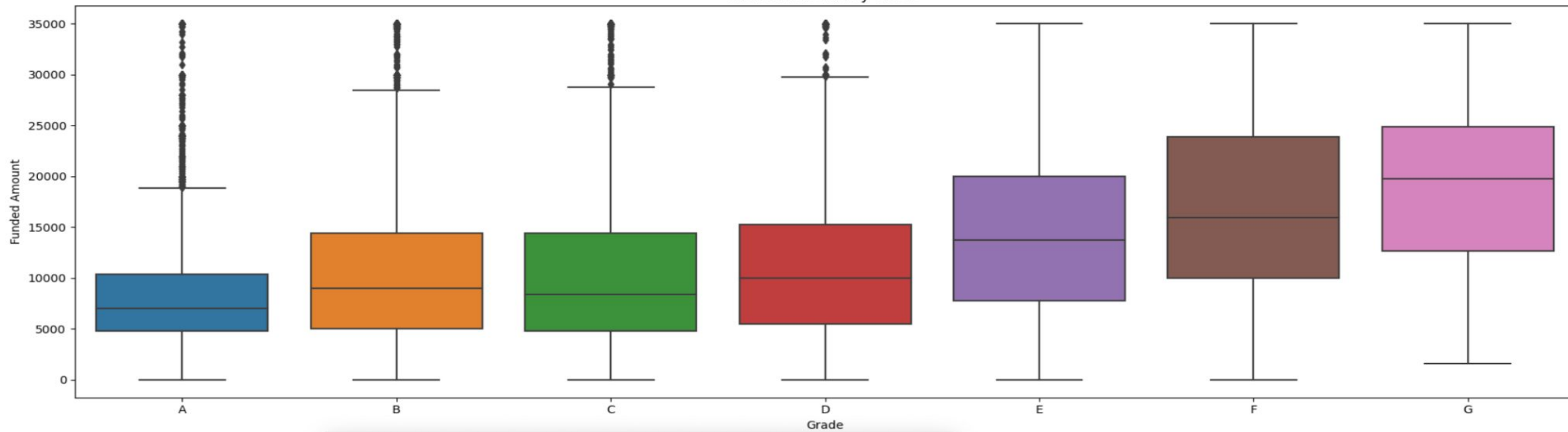
Loan Status Distribution for Address state



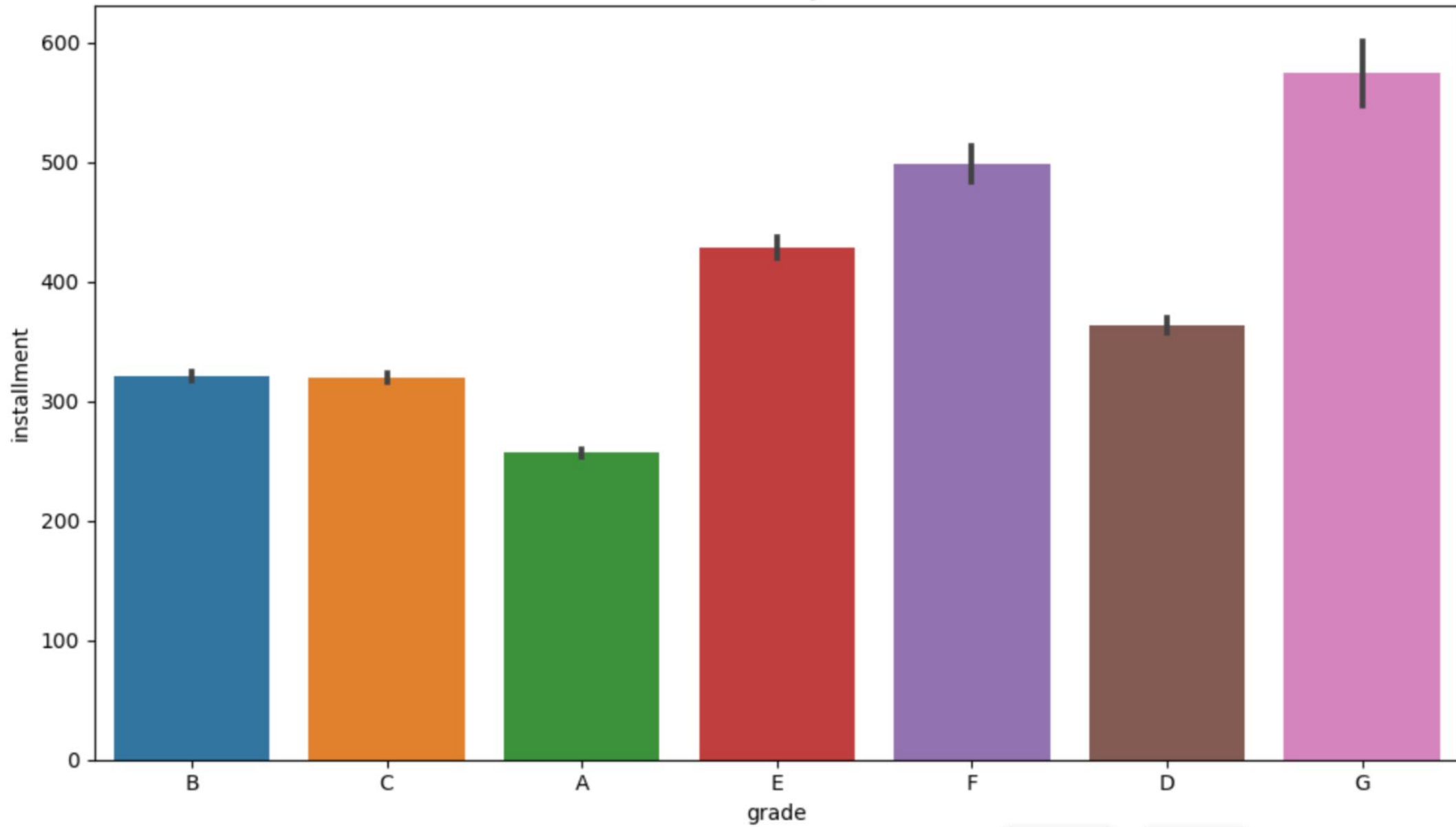
Loan Grade Distribution



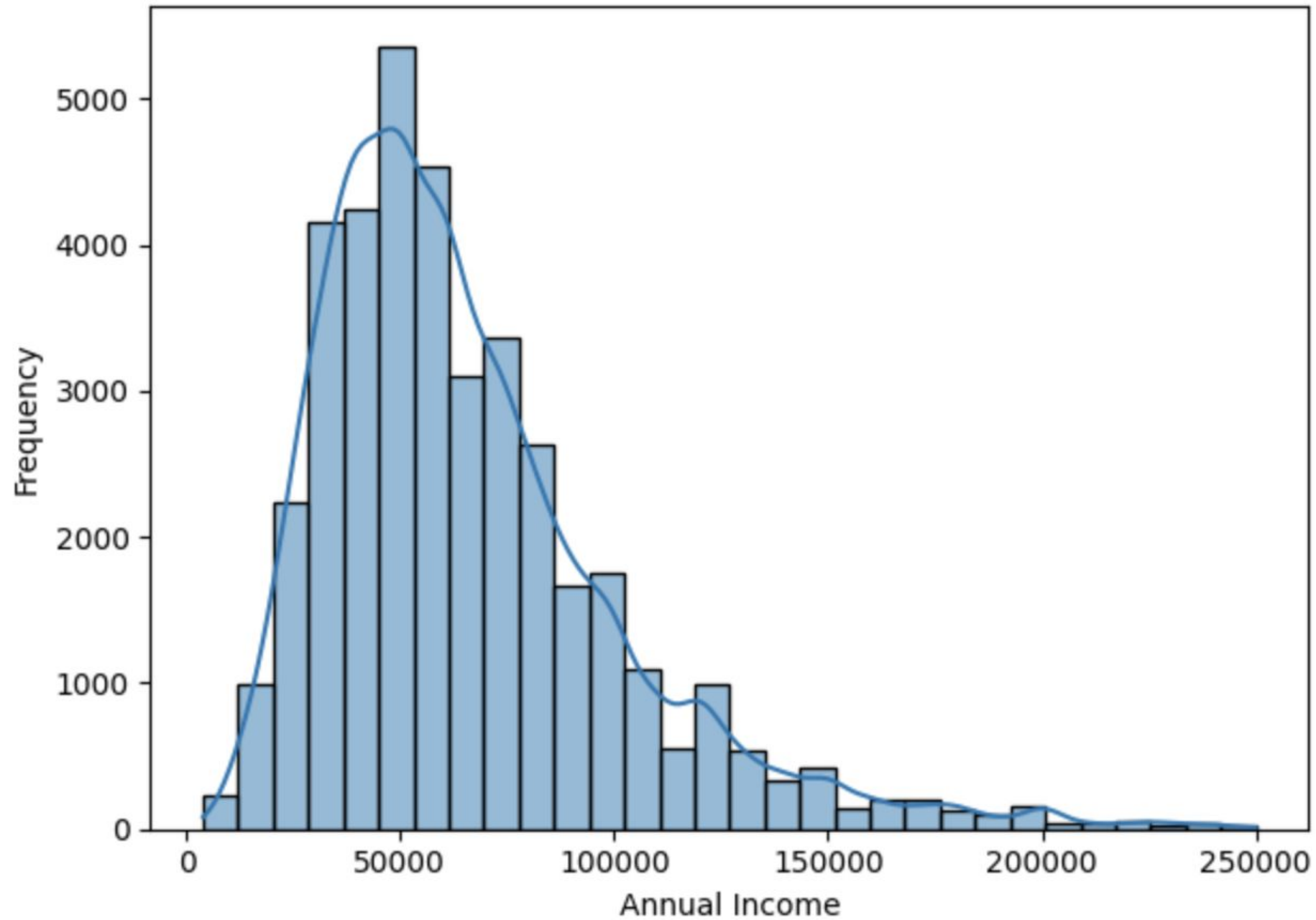
Funded Amount by Grade

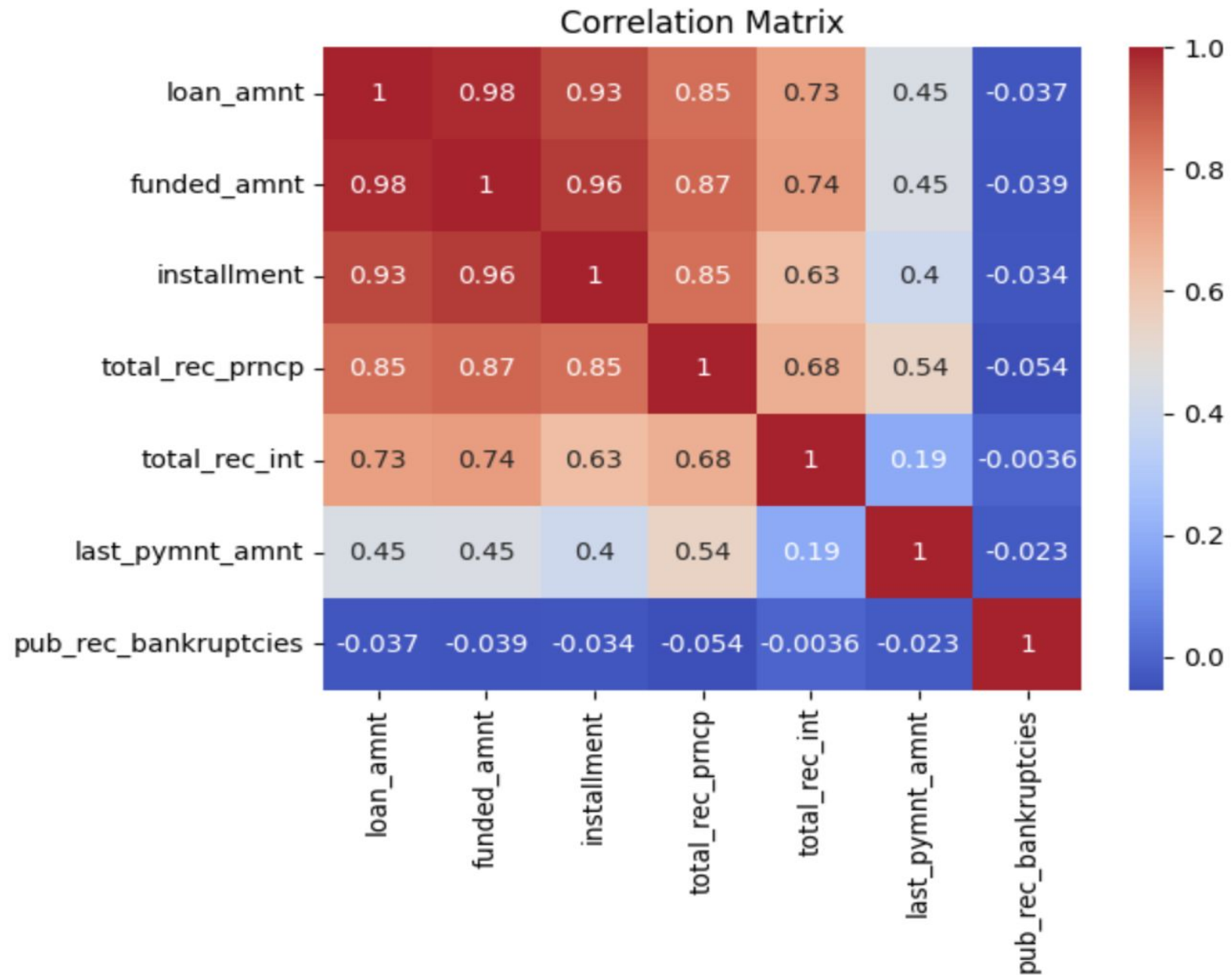


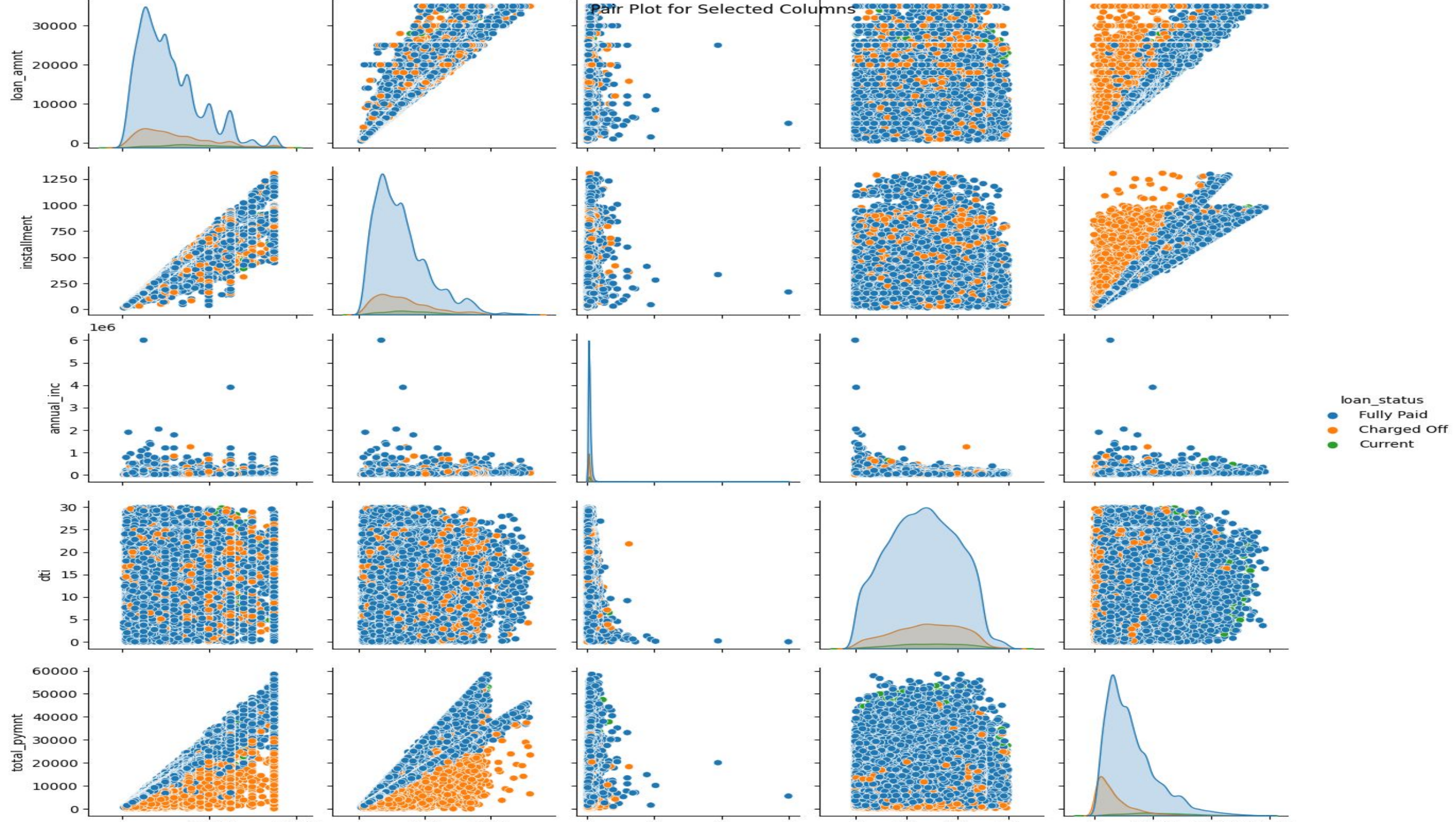
Installment by Grade

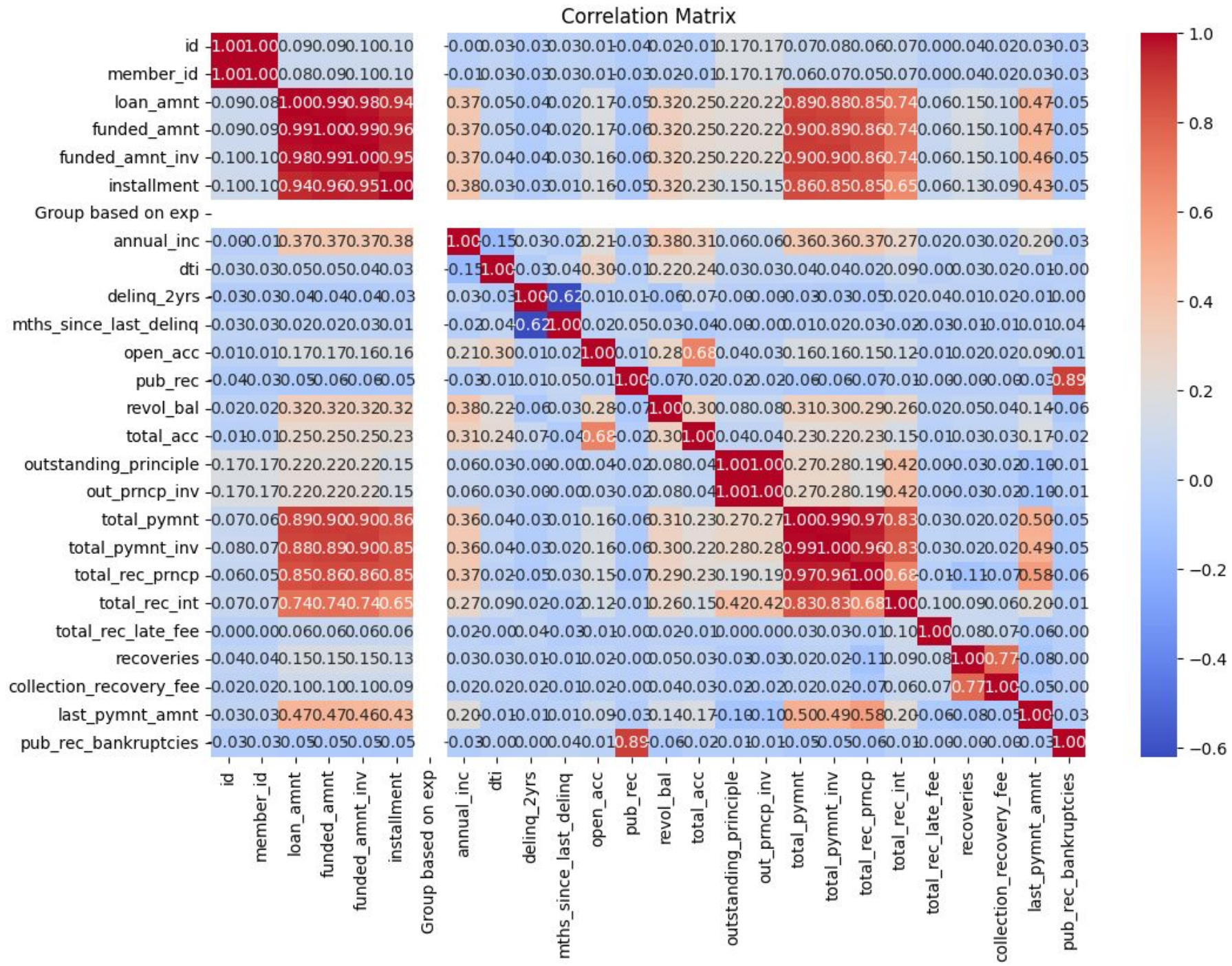


Annual Income Distribution









Conclusion

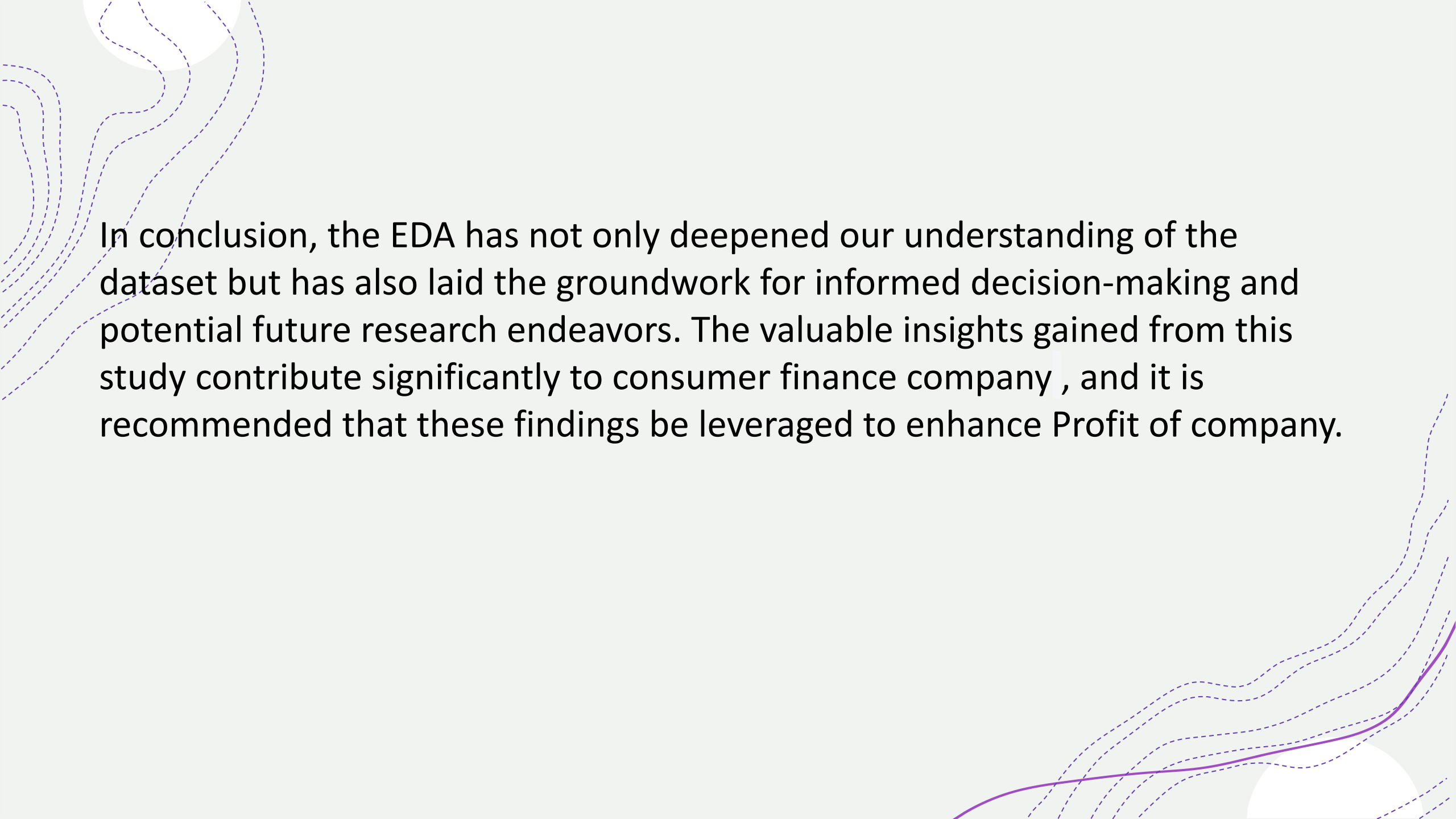
Implications

The insights derived from Lending Club Case Study, the EDA have several implications for consumer finance company, impacting loss of business to the company. Understanding these implications is crucial for making informed decisions and driving positive outcomes.

Recommendations

Based on the findings and insights obtained from the EDA, the following recommendations are put forth:

1. Impact of G grade people is more on funded amount.
2. Most of the people who take loan live in rented apartment.
3. Most of the people who take loans lie in Grade B.
4. Most of the people of CA state taken the Loan and paying full amount and non defaulter.
5. Most of the people taking loan for the purpose of debt consolidation and paying full amount and non defaulter.
6. Impact of G grade is more on loan amount.
7. People whose income is more the 50k take more loans.



In conclusion, the EDA has not only deepened our understanding of the dataset but has also laid the groundwork for informed decision-making and potential future research endeavors. The valuable insights gained from this study contribute significantly to consumer finance company , and it is recommended that these findings be leveraged to enhance Profit of company.