Landing Club Case Study

Contents

Problem Statement

- You work for a **consumer finance company** which specialises in lending various types of loans to urban customers. 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

Data Description

- Lending Club provided us with customer's historical data. This dataset contained information pertaining to the borrower's past credit history and Lending Club loan information. The total dataset consisted of over 39717records and 111 columns, which was sufficient for our team to conduct analysis. Variables present within the dataset provided an ample amount of information which we could use to identify relationships and gauge their effect upon the success or failure of a borrower fulfilling the terms of their loan agreement.

LoanStatNew	Description
acc_now_delinq	The number of accounts on which the borrower is now delinquent.
acc_open_past_24mths	Number of trades opened in past 24 months.
addr_state	The state provided by the borrower in the loan application
all_util	Balance to credit limit on all trades
annual_inc	The self-reported annual income provided by the borrower during registration.
annual_inc_joint	The combined self-reported annual income provided by the co-borrowers during registration
application_type	Indicates whether the loan is an individual application or a joint application with two co-borrowers
avg_cur_bal	Average current balance of all accounts
bc_open_to_buy	Total open to buy on revolving bankcards.
bc_util	Ratio of total current balance to high credit/credit limit for all bankcard accounts.
chargeoff_within_12_mths	Number of charge-offs within 12 months
collection_recovery_fee	post charge off collection fee
collections_12_mths_ex_med	Number of collections in 12 months excluding medical collections
delinq_2yrs	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
delinq_amnt	The past-due amount owed for the accounts on which the borrower is now delinquent.
desc	Loan description provided by the borrower
dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
dti_joint	A ratio calculated using the co-borrowers' total monthly payments on the total debt obligations, excluding mortgages and the requested LC loan, divided by the co-borrowers' combined self-reported monthly income
earliest_cr_line	The month the borrower's earliest reported credit line was opened
emp_length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
emp_title	The job title supplied by the Borrower when applying for the loan.*
fico_range_high	The upper boundary range the borrower's FICO at loan origination belongs to.
fico_range_low	The lower boundary range the borrower's FICO at loan origination belongs to.
›ded_amnt	The total amount committed to that loan at that point in time.
‹l_amnt_inv	The total amount committed by investors for that loan at that point in time.
	LC assigned loan grade
›ship	The home ownership status provided by the borrower during registration. Our values are: RENT, OWN, MORTGAGE, OTHER.

We required only the variables that had a direct or indirect response to a borrower's potential to default. To achieve this, we prepared the data by choosing select variables that would best fit this criteria.

Data Understanding

Dataset Attributes:

Primary Attribute

Loan Status:The Principal Attribute of Interest (loan_status). This column consists of three distinct values:

Fully-Paid: Signifies customers who have successfully repaid their loans.

Charged-Off: Indicates customers who have been labeled as "Charged-Off" or have defaulted on their loans.

Current: Represents customers whose loans are presently in progress and, thus, cannot provide conclusive evidence regarding future defaults.

For the purposes of this case study, rows with a "Current" status will be excluded from the analysis.

Decision Matrix:

Loan Acceptance Outcome-There are three potential scenarios

FullyPaid-Thiscategoryrepresentsapplicantswhohavesuccessfullyrepaidboththeprincipalandtheinterestrateoftheloan.

Current-Applicants in this group are actively in the process of making loan installments; hence, the loan tenure has not yet concluded. These individuals are not categorized as 'defaulted.

Charged-off-This classification pertains to applicants who have failed to make timely installments for an extended period, resulting in a 'default' on the loan.

Loan Rejection-In cases where the company has declined the loan application (usually due to the candidate not meeting their requirements), there is no transactional history available for these applicants. Consequently,this data is unavailable to the company and is not included in this dataset.

Data Cleaning Process

- To prepare the loan dataset for analysis, the following steps were taken:

1. Missing values were handled by imputation or removal.

2. Percentage strings (e.g., '83.70%') were converted to numeric values.

3. Irrelevant columns were removed, focusing on key fields like loan amount, interest rate, and loan status.

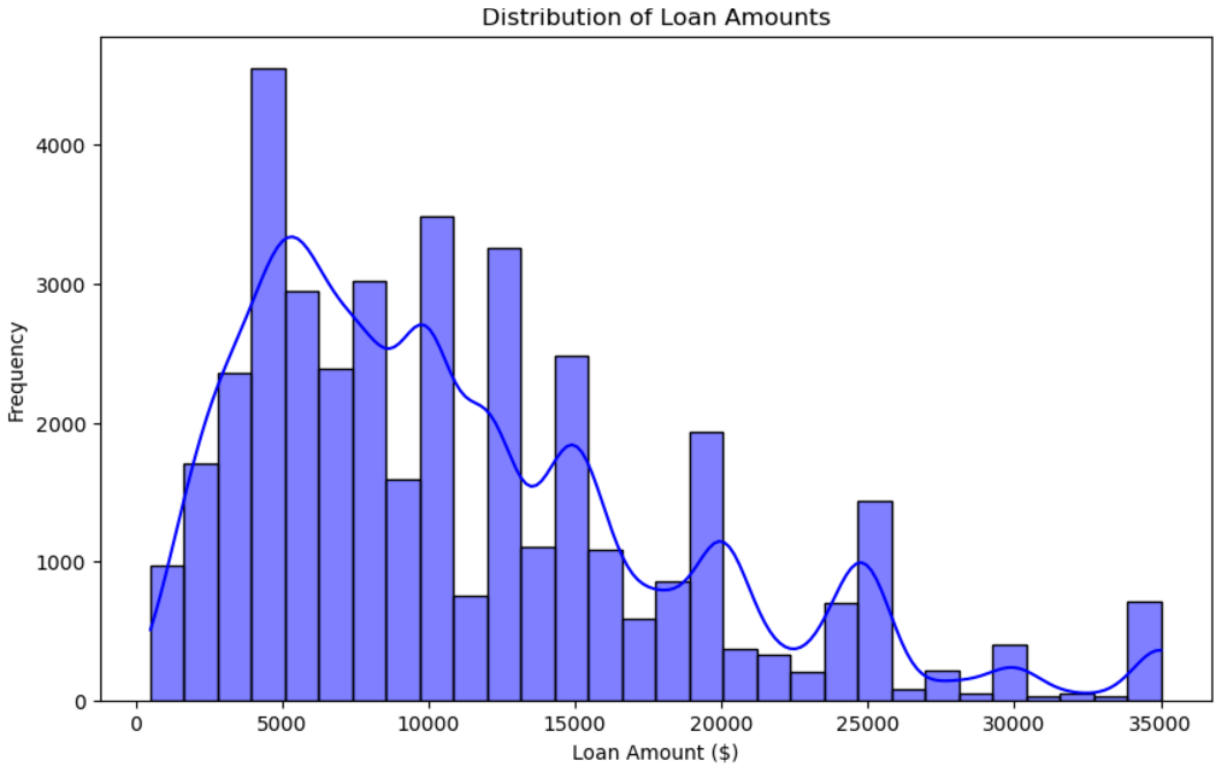
4. Data formatting issues in columns like 'revol_util' and 'int_rate' were resolved.

5. The dataset was verified for consistency and completeness before visualization.

Univariate Analysis

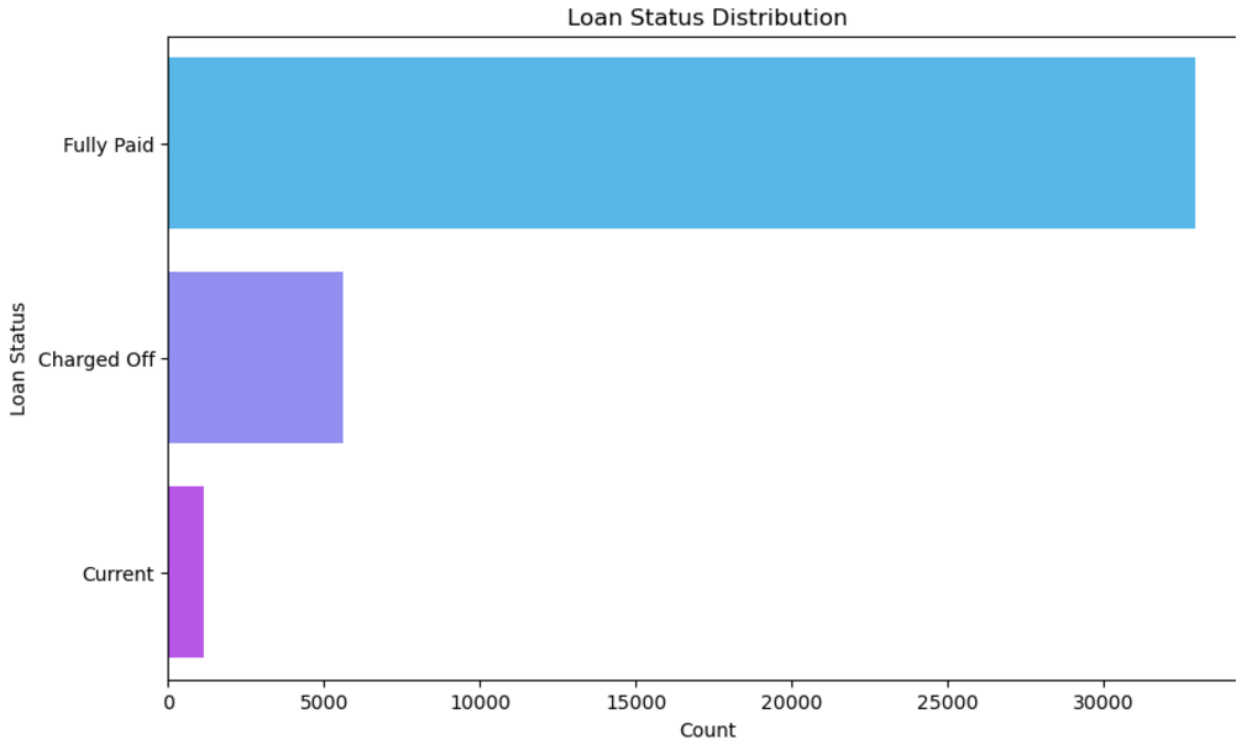
- Univariate analysis focuses on individual variables to understand their distributions and key characteristics.

- Graphs included:
- 1. Loan Amount Distribution
- 2. Loan Status Counts
- 3. Purpose of Loans



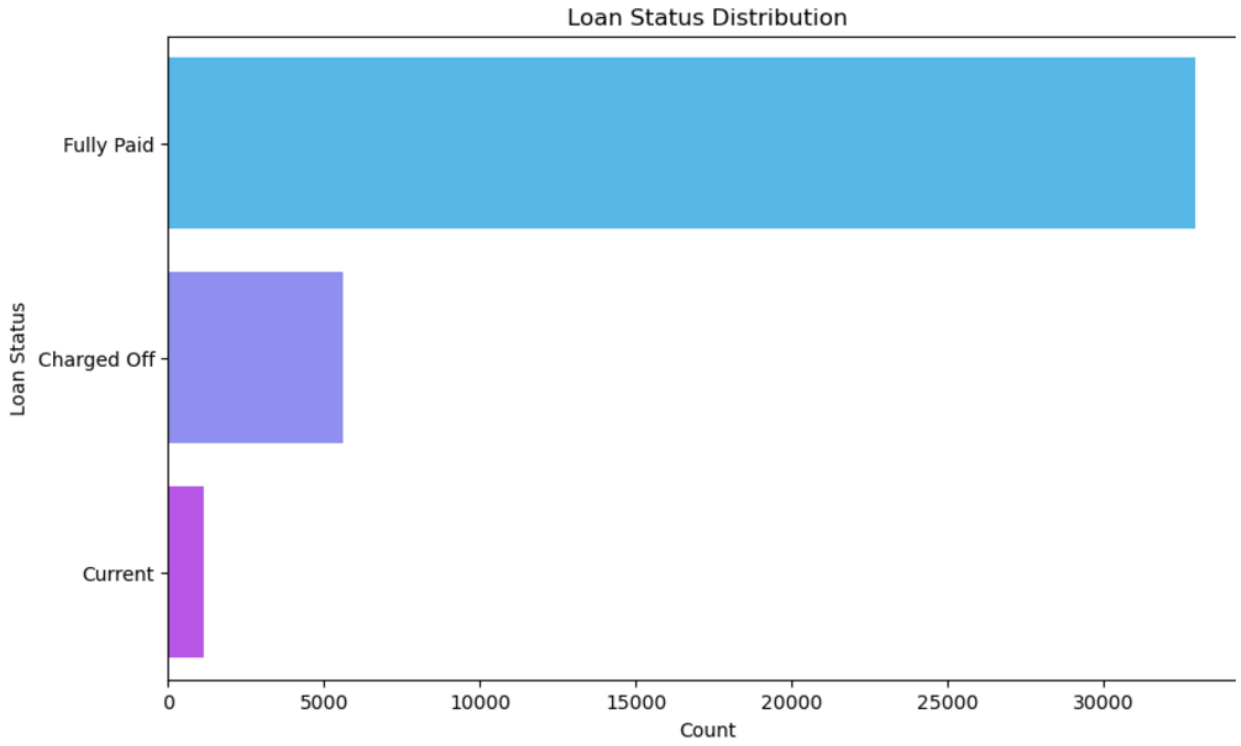
Loan Amount Distribution

Graph: Distribution of Loan Amounts.



Loan Status Counts

Graph: Loan Status Counts.



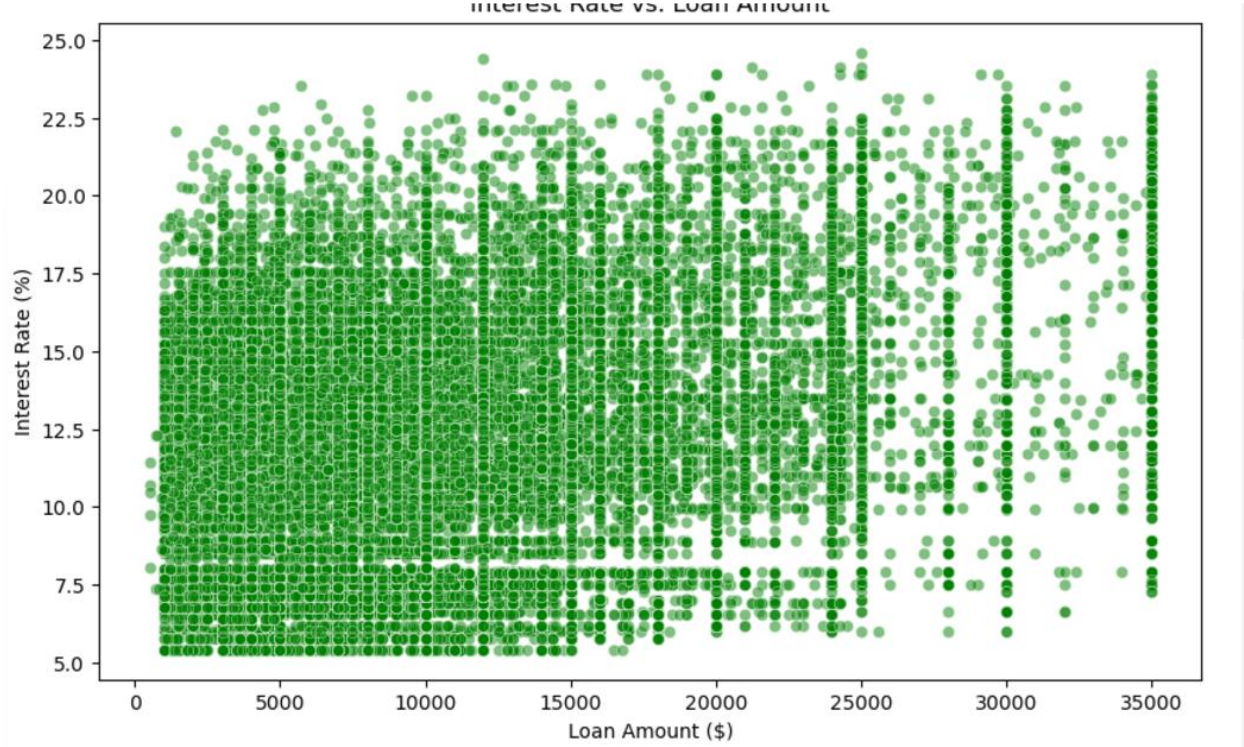
. Purpose of Loans

Graph: Loan Status Counts.

Bivariate Analysis

- Bivariate analysis explores relationships between two variables to identify trends or correlations.

- Graphs included:
- 1. Interest Rate vs. Loan Amount
- 2. Annual Income vs. Loan Amount



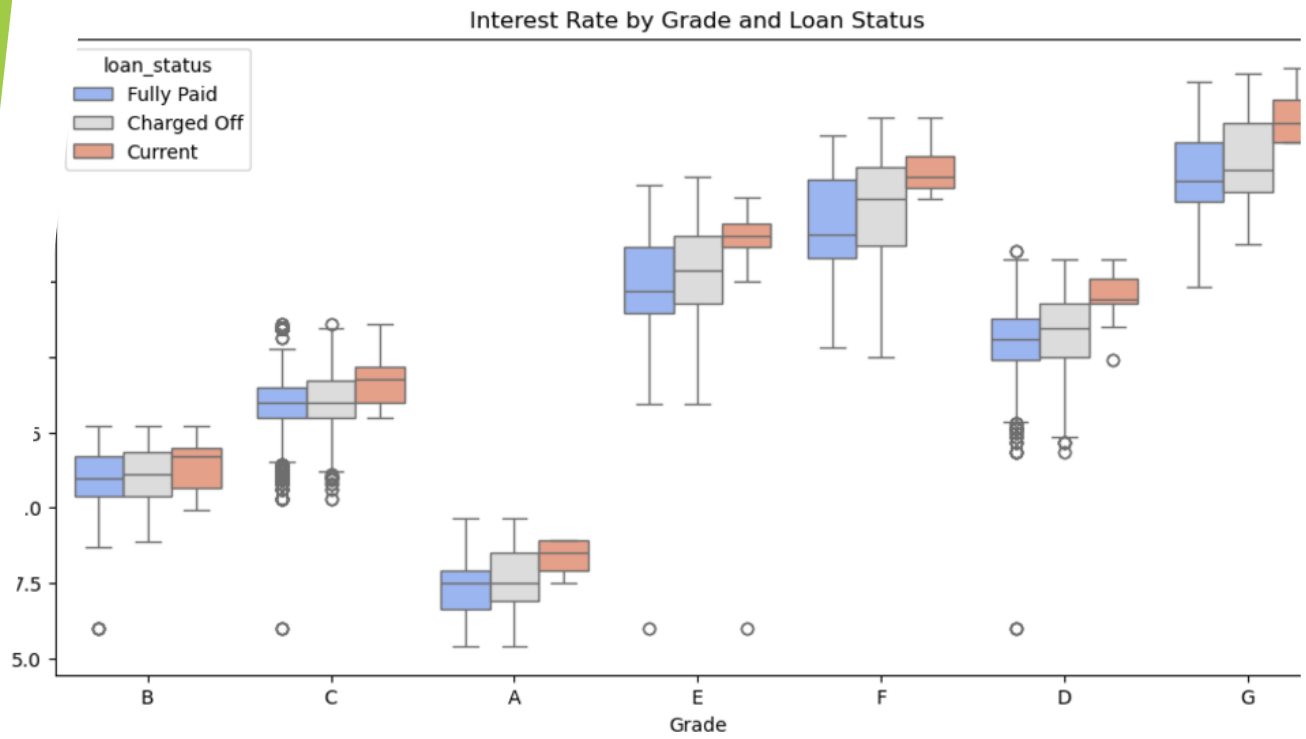
Interest Rate vs. Loan Amount

Graph: Correlation between Interest Rate and Loan Amount.

Multivariate Analysis

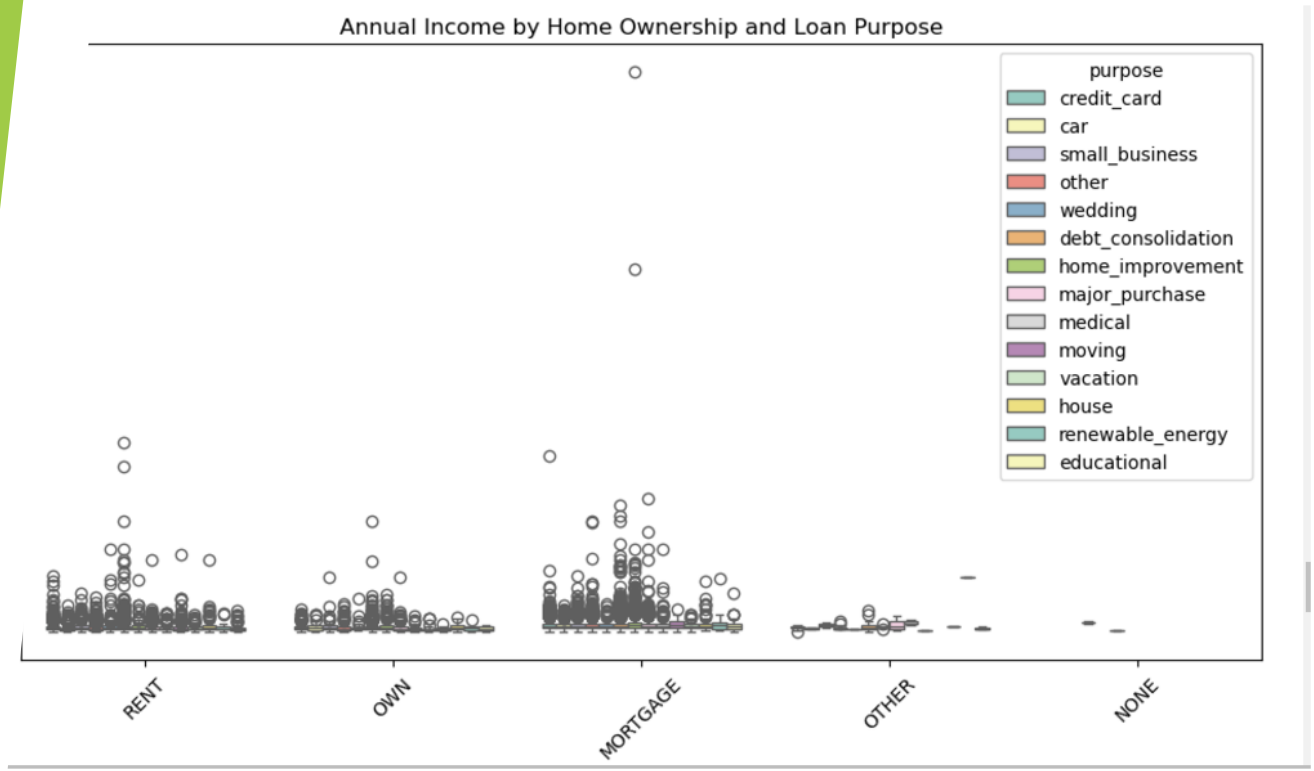
- Multivariate analysis investigates relationships among multiple variables simultaneously.

- Graphs included:
- 1. Grade vs. Interest Rate by Loan Status
- 2. Home Ownership vs. Annual Income by Loan Purpose
- 3. Loan Amount vs. Interest Rate Grouped by Purpose



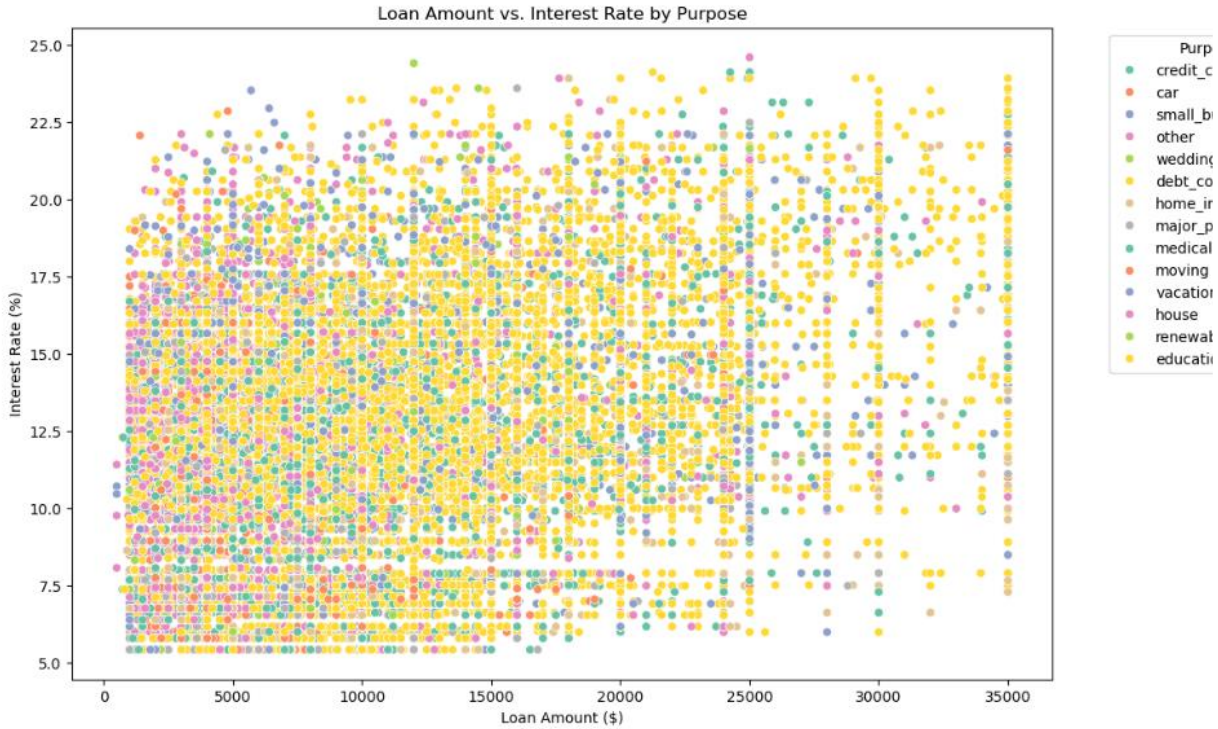
Grade vs. Interest Rate by Loan Status

Graph: Analysis of Grades and Interest Rates by Loan Status.



Home Ownership vs. Annual Income by Loan Purpose

Graph: Home Ownership vs. Annual Income by Loan Purpose



Loan Amount vs. Interest Rate Grouped by Purpose

Graph: Loan Amount vs. Interest Rate Grouped by Purpose

Conclusions

- Key insights from the analysis:

- 1. Loan amounts vary significantly across purposes and income levels.
- 2. Higher grades are associated with lower interest rates, reflecting better creditworthiness.
- 3. Loan status analysis highlights areas of potential default risk.
- 4. Multivariate analyses reveal complex interactions among purpose, income, and interest rates.

References & Useful Links

- **Technologies & Packages Used:**

  Python 3.11.4 https://www.python.org/

  Matplotlib 3.7.1 https://matplotlib.org/

  Numpy 1.24.3 https://numpy.org/

  Pandas 1.5.3 https://pandas.pydata.org/

  Seaborn 0.12.2 https://seaborn.pydata.org

- **GitHub Repository Link:**

https://github.com/Satish7049/lending-club-case-study

Thank You!