LENDING CLUB CASE STUDY

SUBMITTED BY:

BALAJI NARASIMHAN

BAGGAM BHASKARA SAIKUMAR PATNAIK

CONTENTS

PROBLEM STATEMENT

- **Objective:** To develop a model that assists a consumer finance company in making informed loan approval decisions by identifying patterns in applicant profiles that indicate the likelihood of loan default.
- **Context:** The company specializes in providing various types of loans. When a loan application is received, the company must decide whether to approve or reject the loan based on the applicant's profile. This decision involves two primary risks:
    1. **Business Loss:** If an applicant who is likely to repay the loan is not approved, the company loses potential business.
    2. **Financial Loss:** If an applicant who is likely to default is approved, the company incurs a financial loss.

PROBLEM STATEMENT

- **Scenarios:**
  1. **Loan Accepted:**
     I. **Fully Paid:** The applicant has fully repaid the loan.
     II. **Current:** The applicant is currently repaying the loan, and the tenure is not yet completed.
     III. **Charged-off:** The applicant has defaulted on the loan.
  2. **Loan Rejected:** The applicant did not meet the company's requirements, and thus, there is no transactional history available for these applicants.
- **Goal:** To use Exploratory Data Analysis (EDA) to understand how consumer and loan attributes influence the tendency of default. This analysis will help in making data-driven decisions on lending loans.

DATASET

- Dataset contains historical information about past loan applicants, their repayment status and Credit history. This data will be used to identify patterns and attributes that influence the likelihood of default.

DATA HANDLING AND CLEANING

- Load the dataset from the source CSV.
- Preview the data to ensure that the data has been loaded correctly.
- Drop all null values from dataset.
- There are multiple columns in the dataset that has single values which will not help in analysis therefore Dropping single valued columns.
- Remove the rows that have loan status as current and few column which are not useful for meaningful analysis and handle the missing values.

UNIVARIATE ANALYSIS

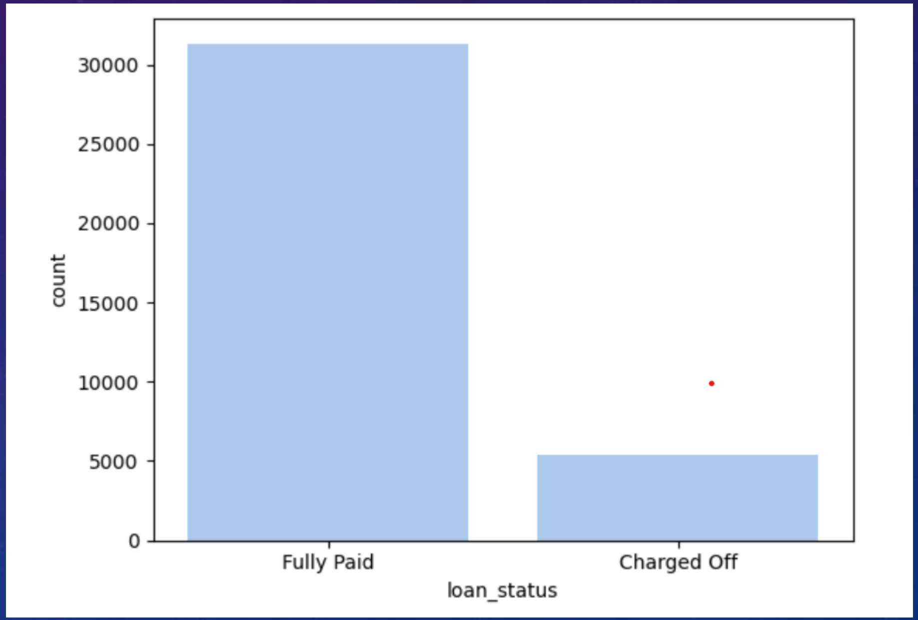
- Univariate analysis is used to describe, summarize, and find patterns in data from a single variable.
- Below mentioned few of the Univariate analysis from the use case.

UNIVARIATE ANALYSIS

Description

- This count plot visualizes the distribution of loan statuses in the dataset. It helps identify the proportion of loans that are fully paid, charged off, or in default.

Plot

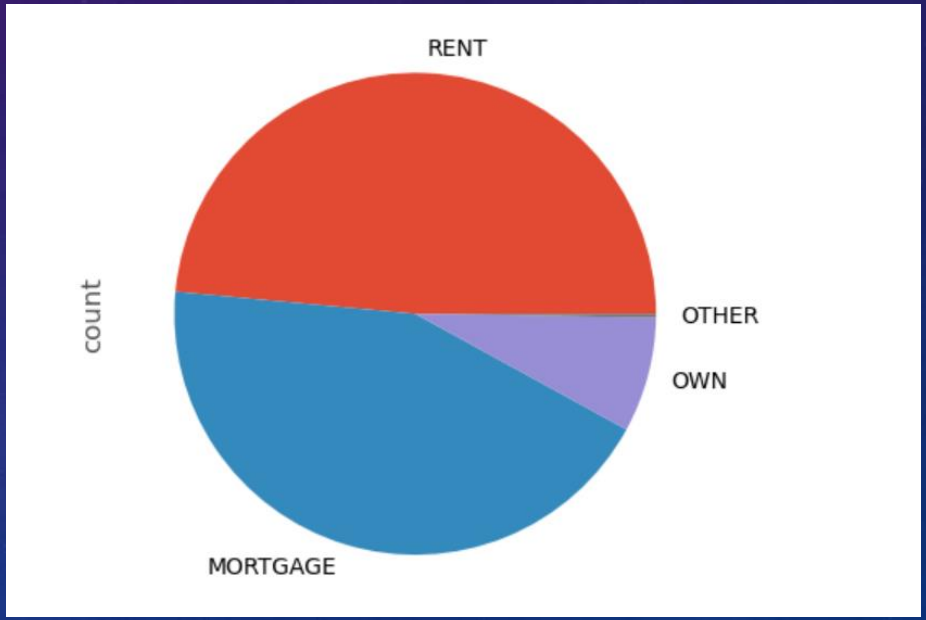


UNIVARIATE ANALYSIS

Description

- This count plot visualizes the distribution of loan statuses in the dataset. It helps The pie chart shows the distribution of home ownership statuses among borrowers. It provides insights into the most common types of home ownership in the dataset. identify the proportion of loans that are fully paid, charged off, or in default.

Plot

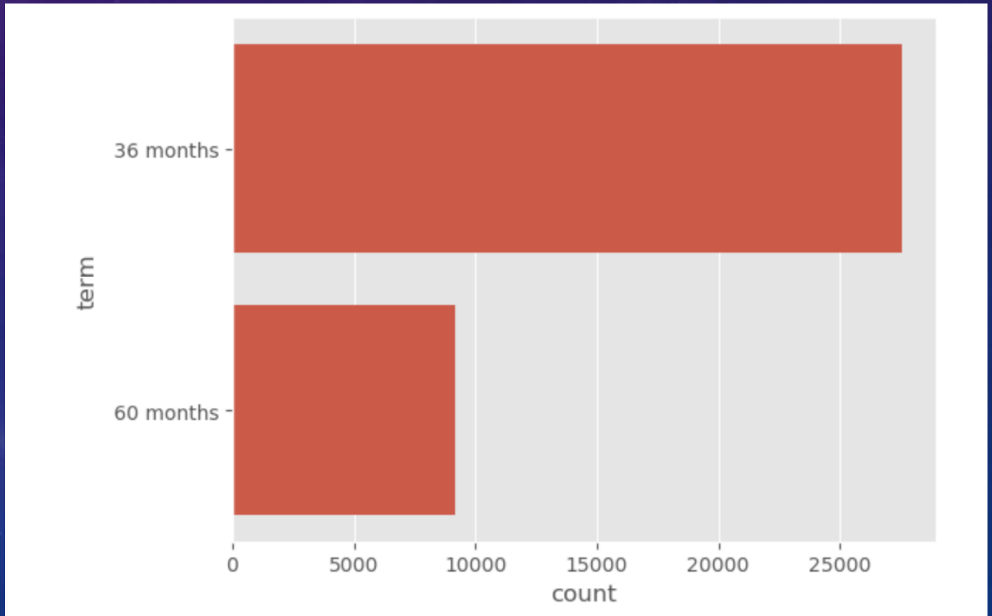


UNIVARIATE ANALYSIS

Description

- This count plot displays the distribution of loan terms. It helps understand the prevalence of different loan durations.

Plot

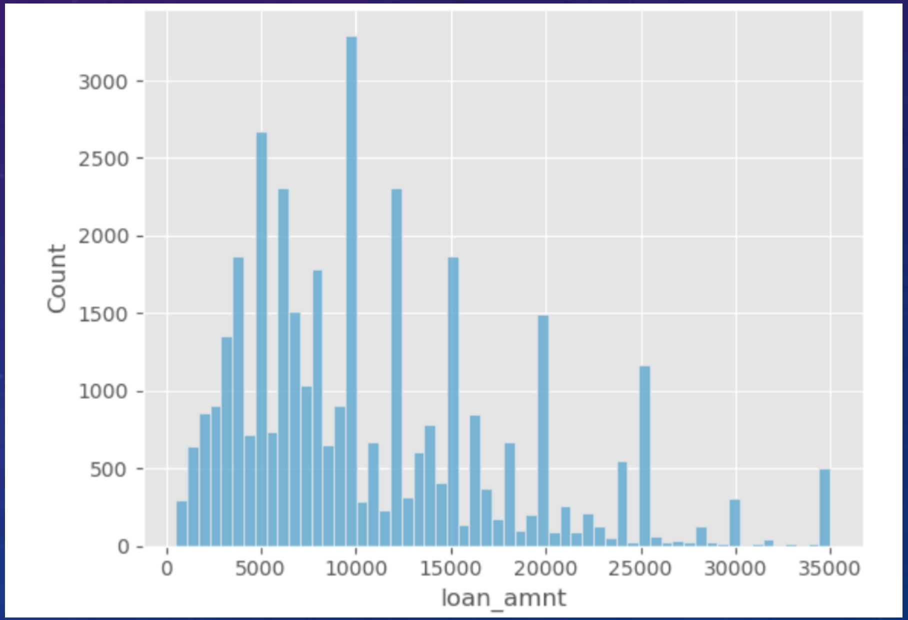


UNIVARIATE ANALYSIS

Description

- The histogram shows the distribution of loan amounts. It highlights the range and frequency of loan amounts requested by borrowers.

Plot

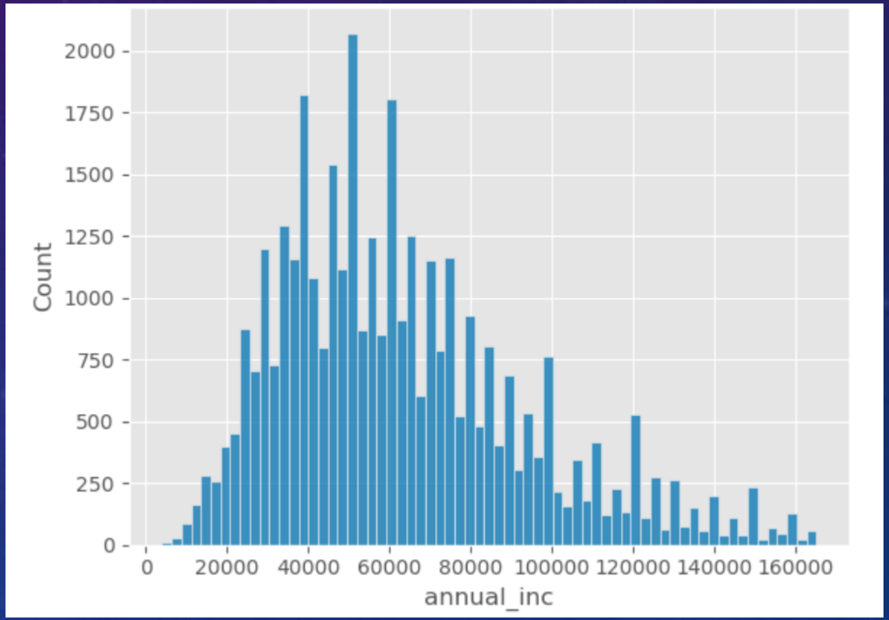


UNIVARIATE ANALYSIS

Description

- This histogram visualizes the distribution of annual incomes of borrowers. It helps identify the income levels of most borrowers.

Plot

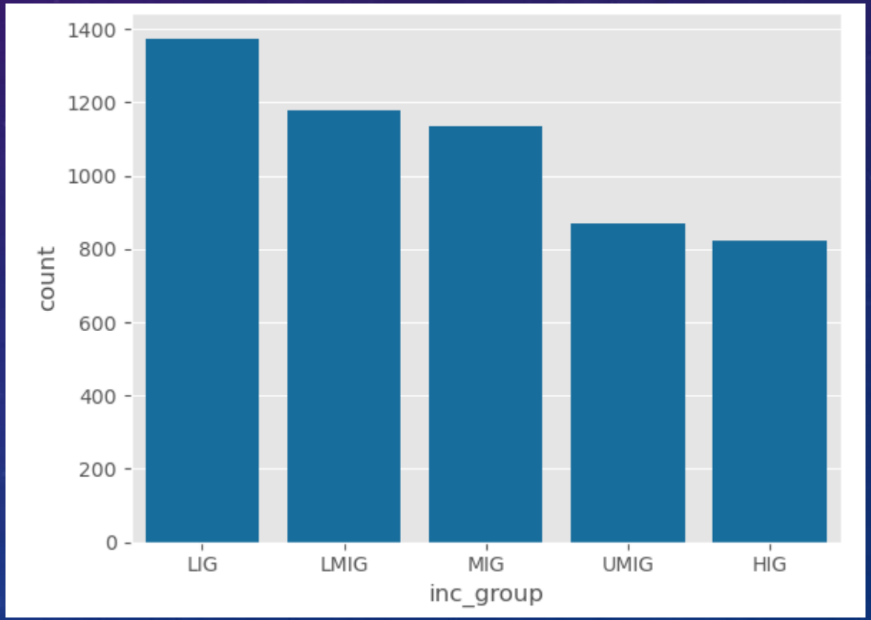


UNIVARIATE ANALYSIS

Description

- The count plot shows the distribution of income groups among defaulted loans. It provides insights into which income groups are more likely to default.

Plot

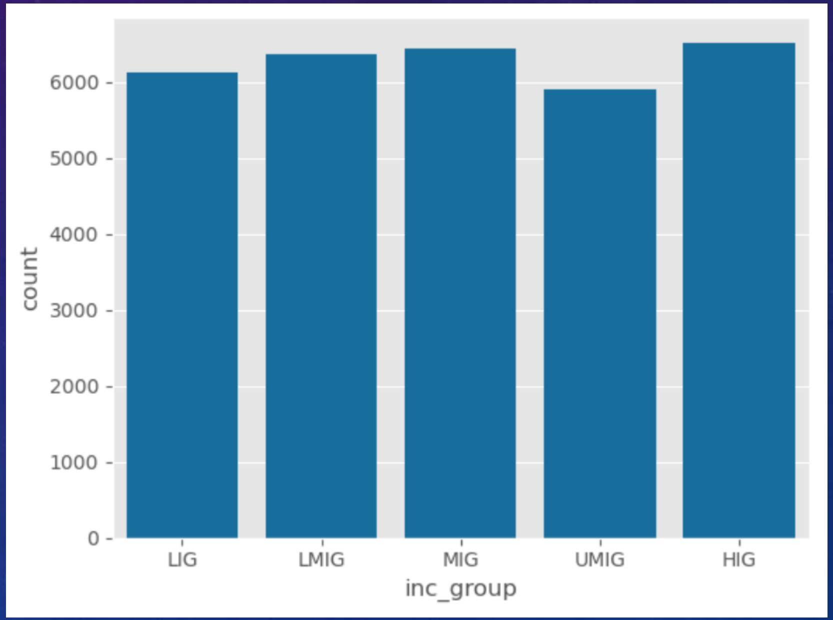


UNIVARIATE ANALYSIS

Description

- This count plot displays the distribution of income groups among paid-up loans. It helps understand which income groups are more likely to repay their loans.

Plot

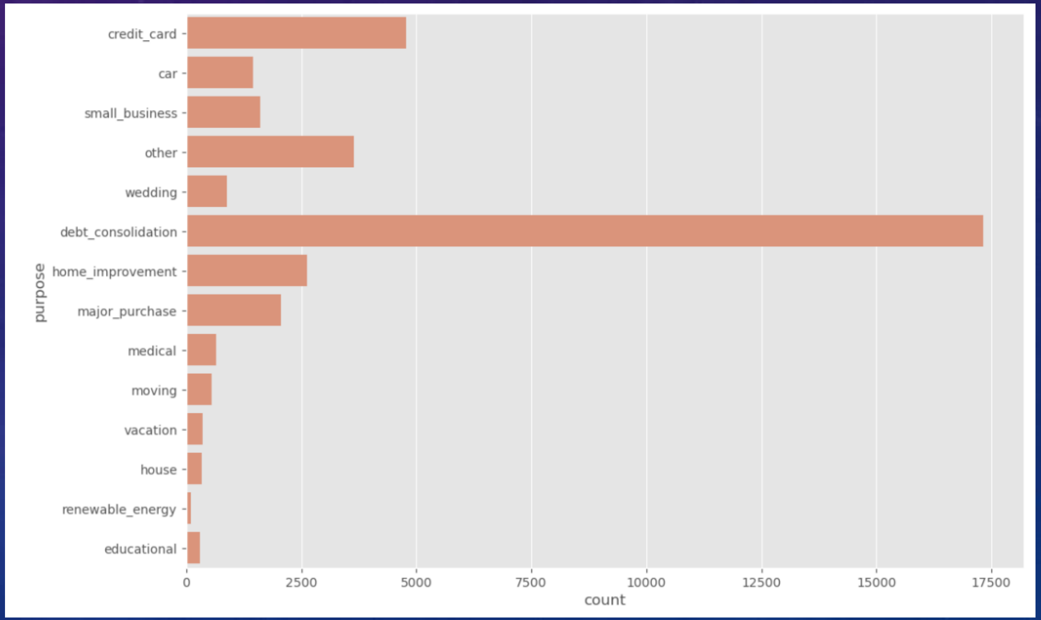


UNIVARIATE ANALYSIS

Description

- The count plot visualizes the distribution of loan purposes. It highlights the most common reasons borrowers take out loans.

Plot

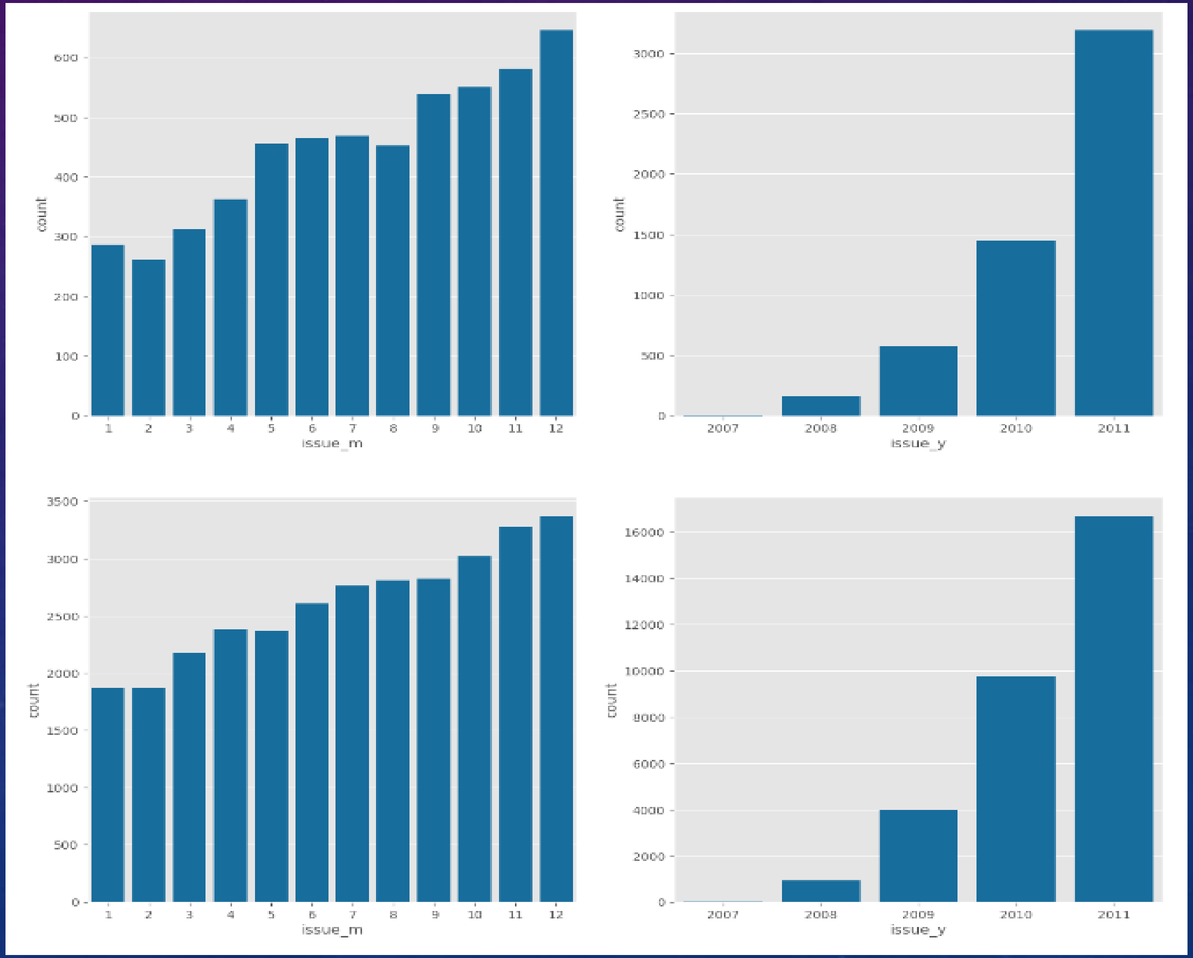


UNIVARIATE ANALYSIS

Description

- These subplots show the distribution of issue months and years for both defaulted and paid-up loans. They help identify any seasonal or yearly trends in loan issuance and repayment behavior.

Plot

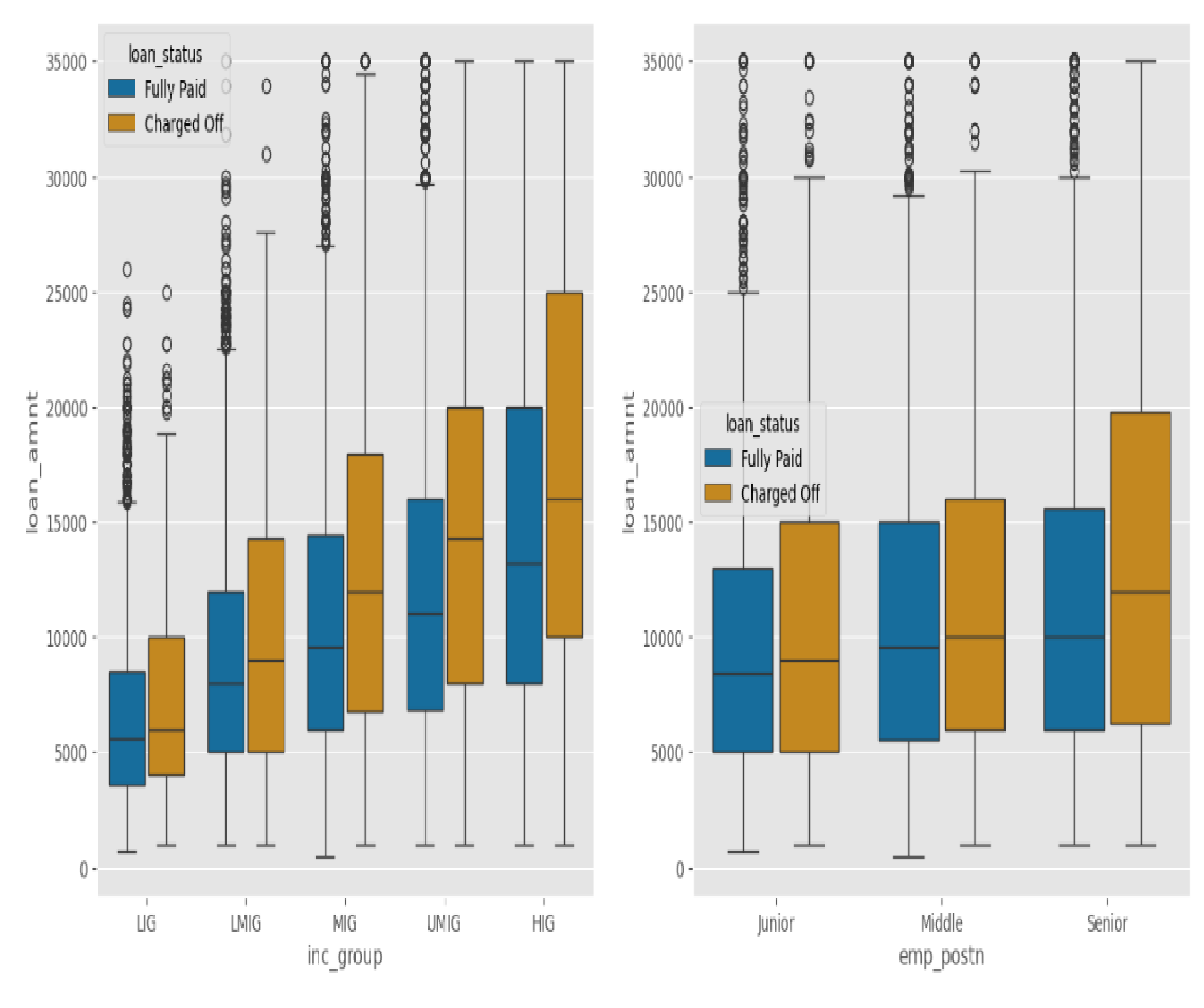


BIVARIATE ANALYSIS

- Bivariate analysis is a statistical method used to explore the relationship between two variables. It helps in understanding how one variable influences or is associated with another.

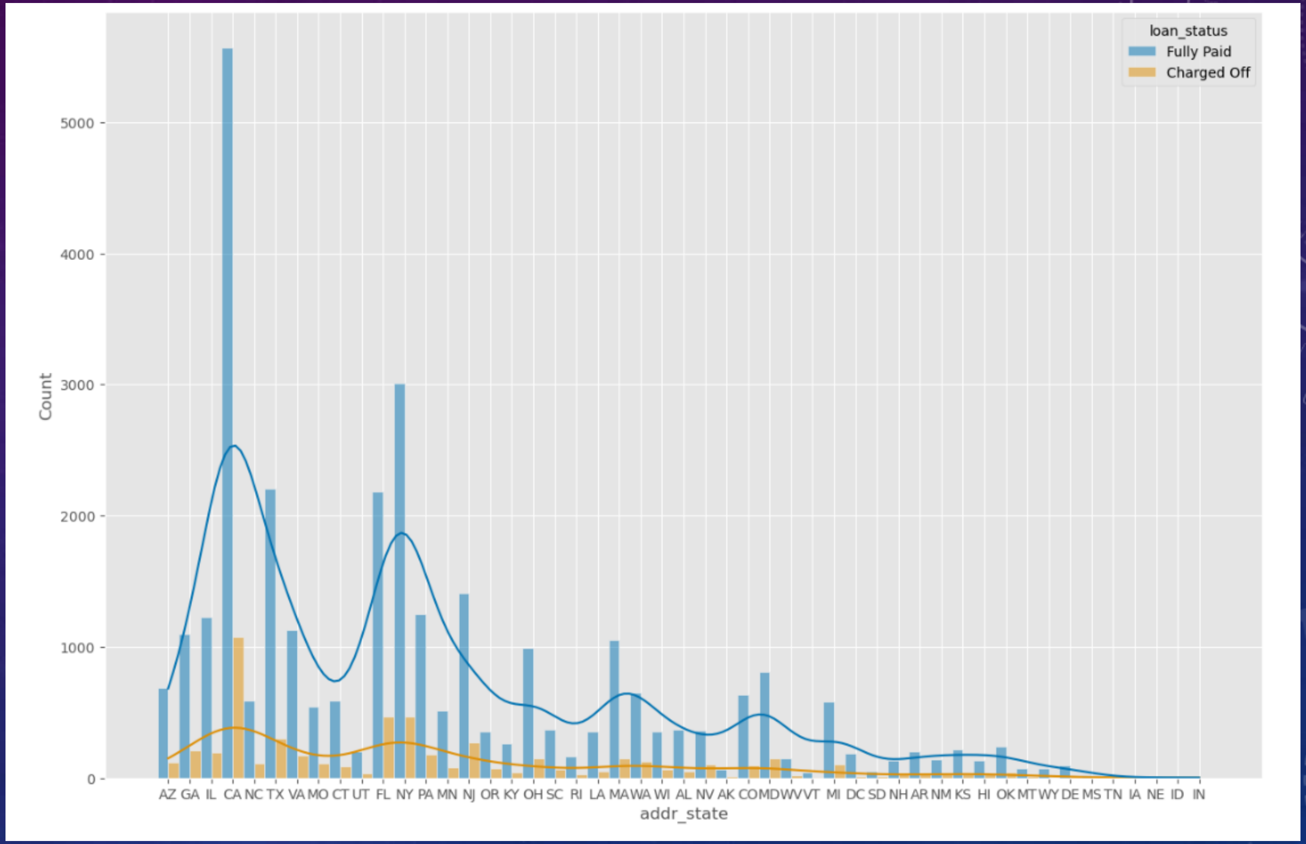
BIVARIATE ANALYSIS

- This plot shows the distribution of loan amounts across different income groups, segmented by loan status.
- This plot illustrates the variation in loan amounts based on different employment positions.



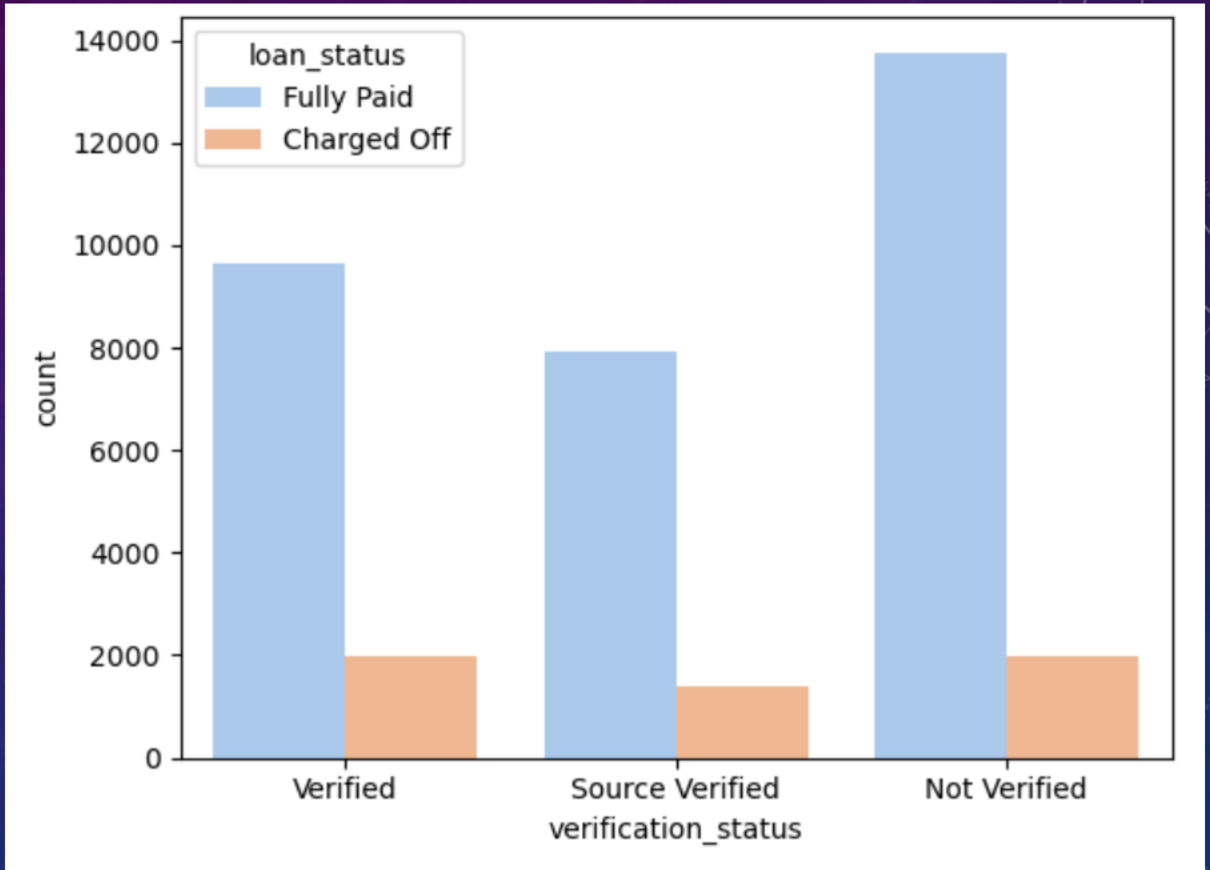
BIVARIATE ANALYSIS

This histogram displays the distribution of loans across different states, with a distinction between loan statuses.



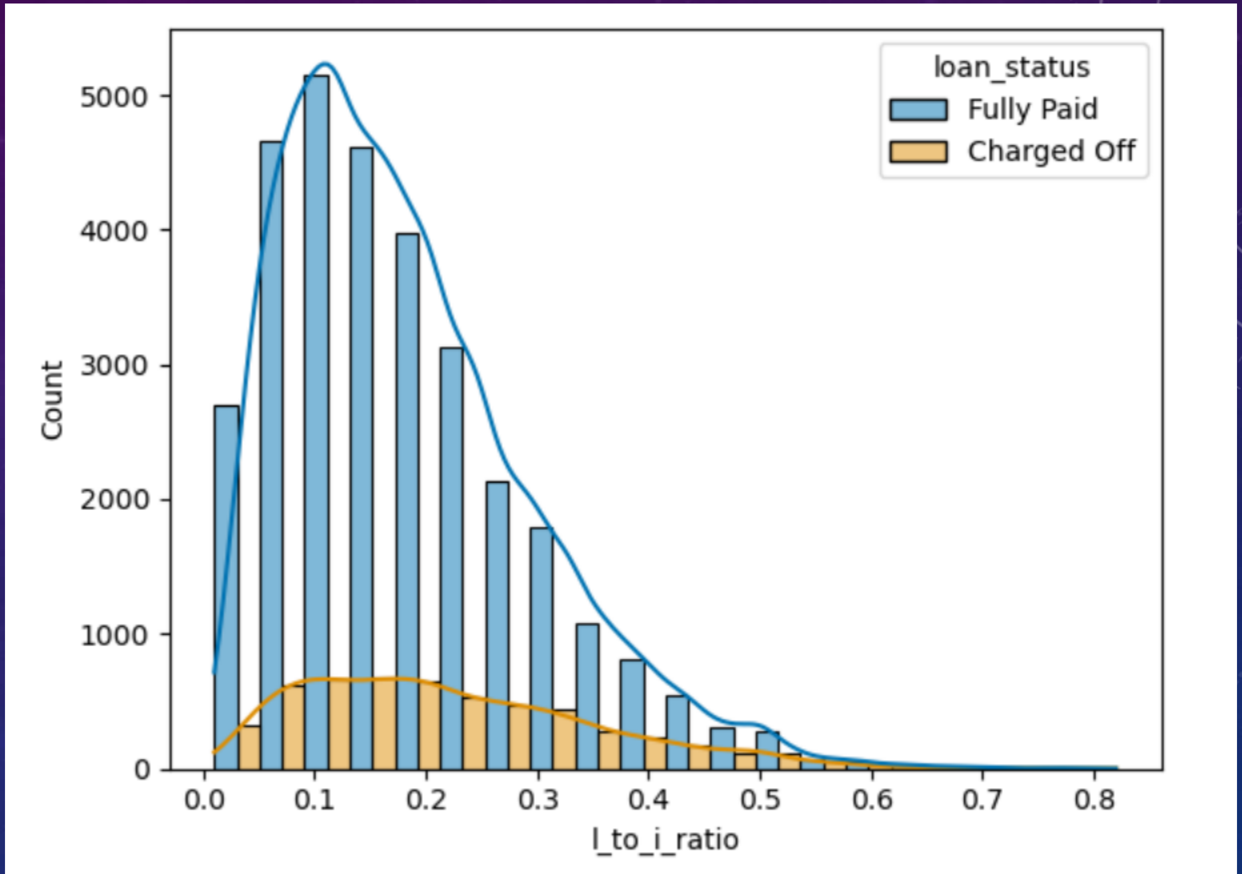
BIVARIATE ANALYSIS

This count plot shows the count of loans based on verification status, segmented by loan status.



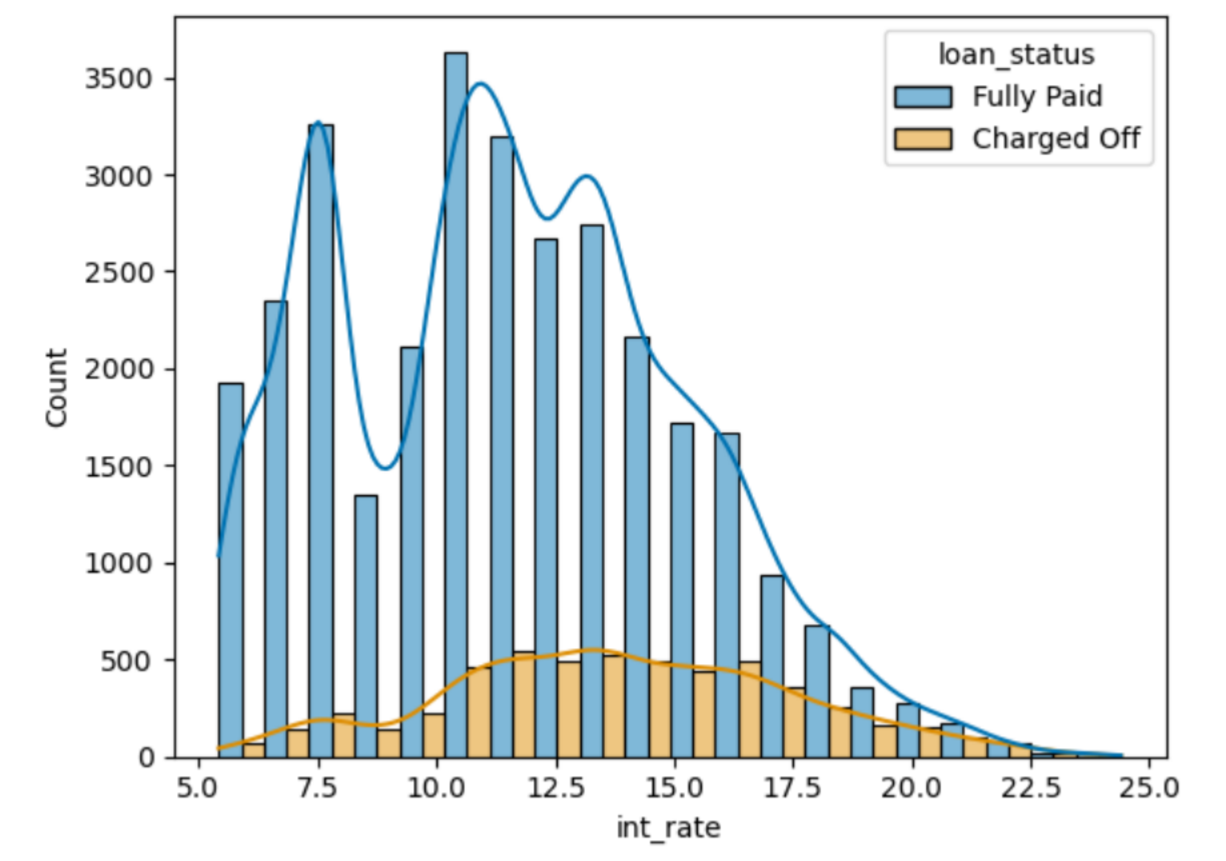
BIVARIATE ANALYSIS

This histogram represents the distribution of the loan-to-income ratio, categorized by loan status. It helps in understanding the relationship between borrowers' income and the loan amounts they receive.



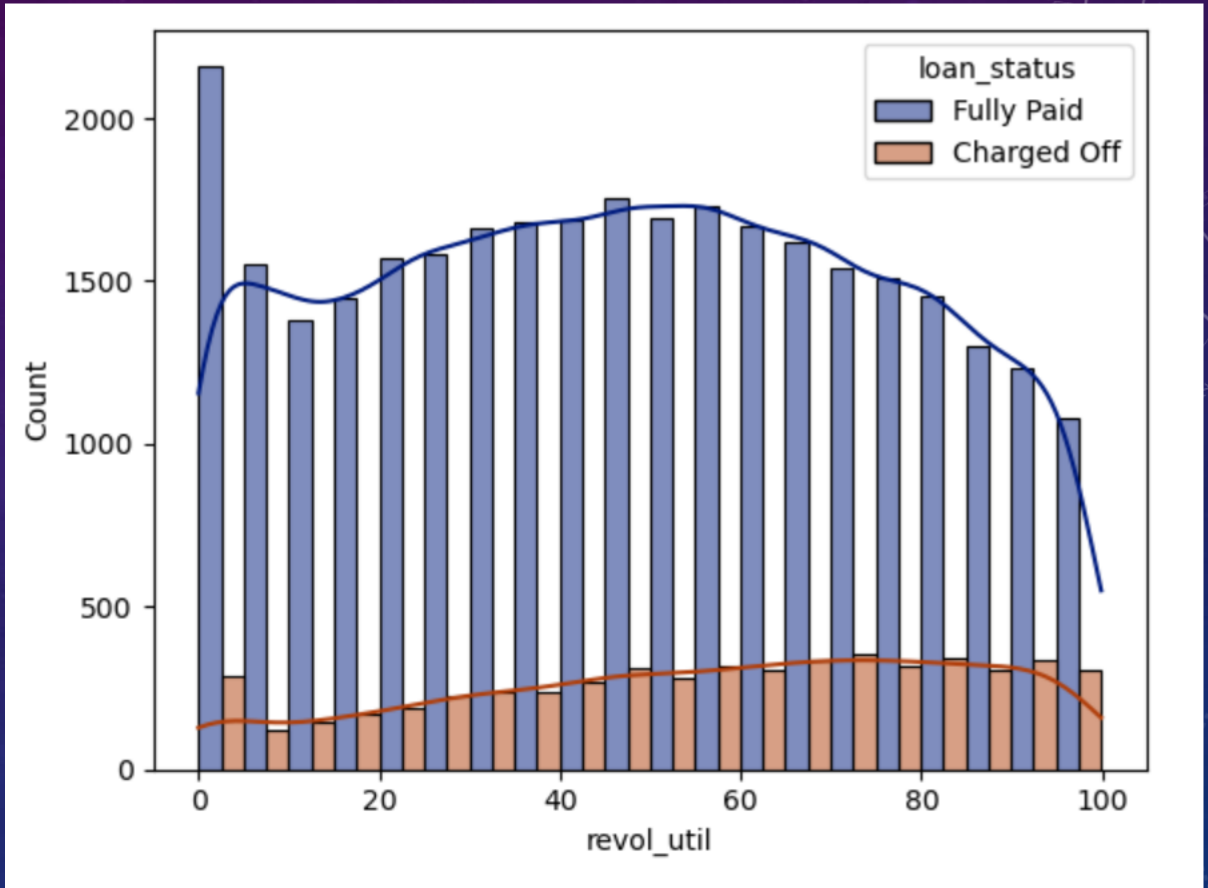
BIVARIATE ANALYSIS

This plot shows the distribution of interest rates for loans, segmented by loan status.



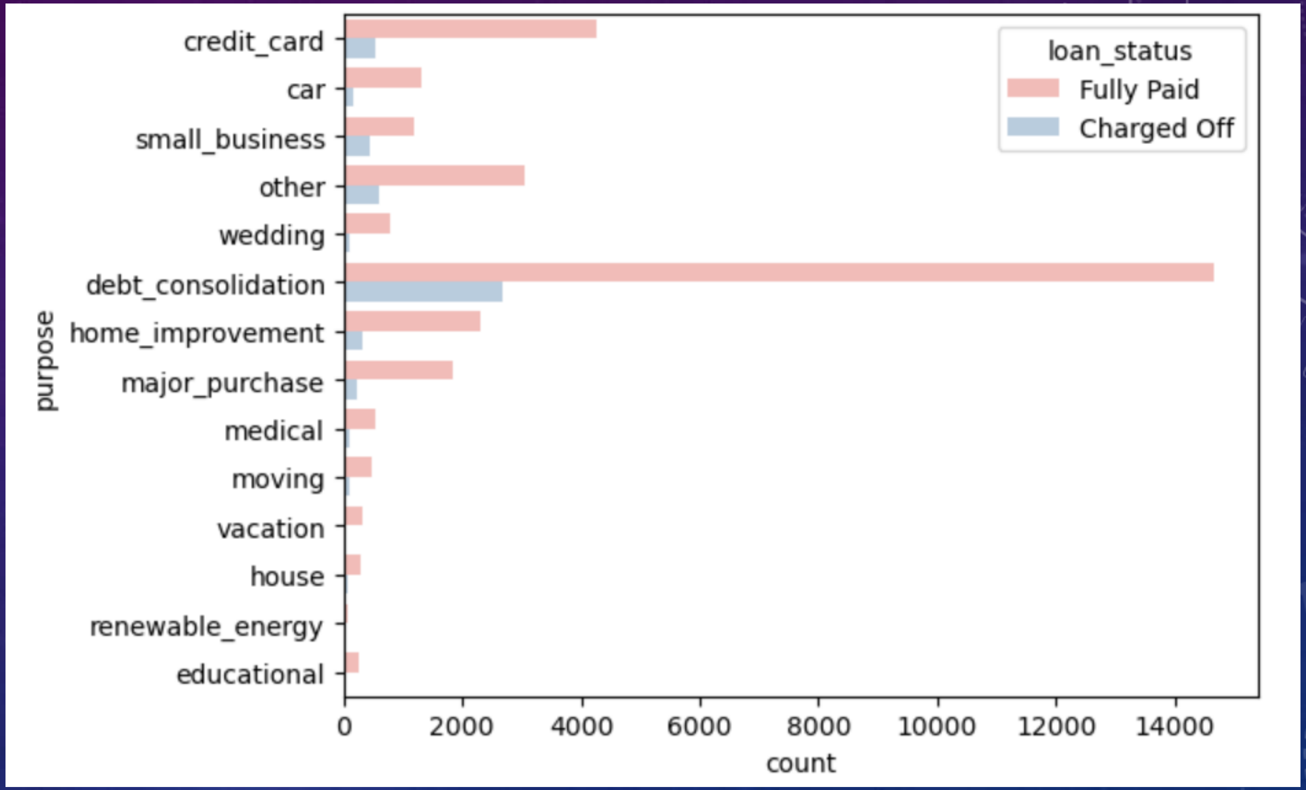
BIVARIATE ANALYSIS

This histogram displays the distribution of revolving utilization rates, categorized by loan status.



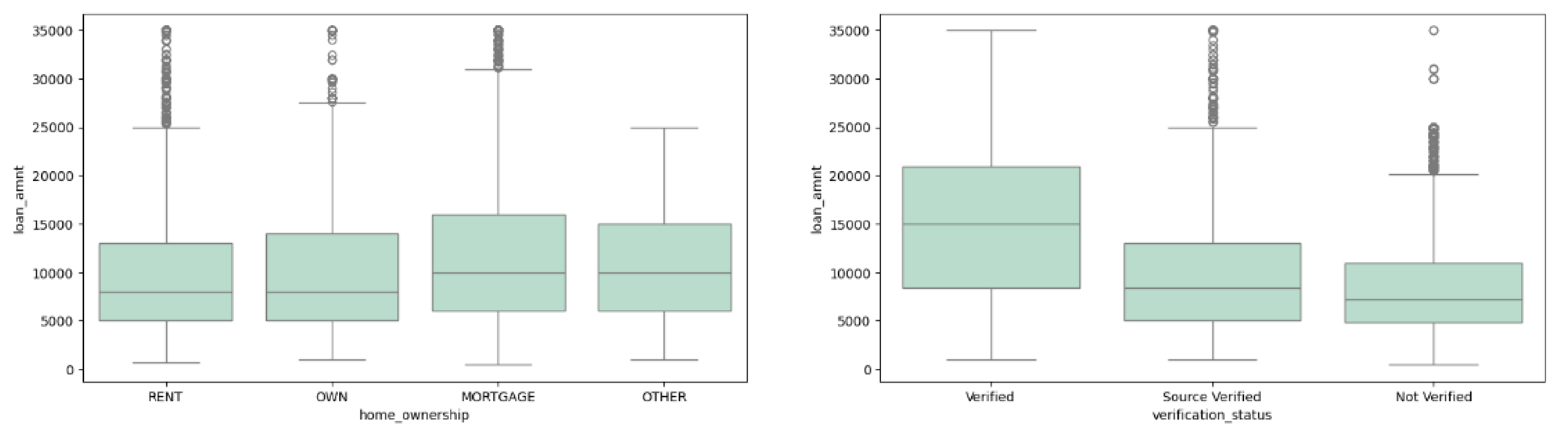
BIVARIATE ANALYSIS

This count plot shows the count of loans for different purposes, segmented by loan status. It helps in identifying the most common reasons for taking loans.



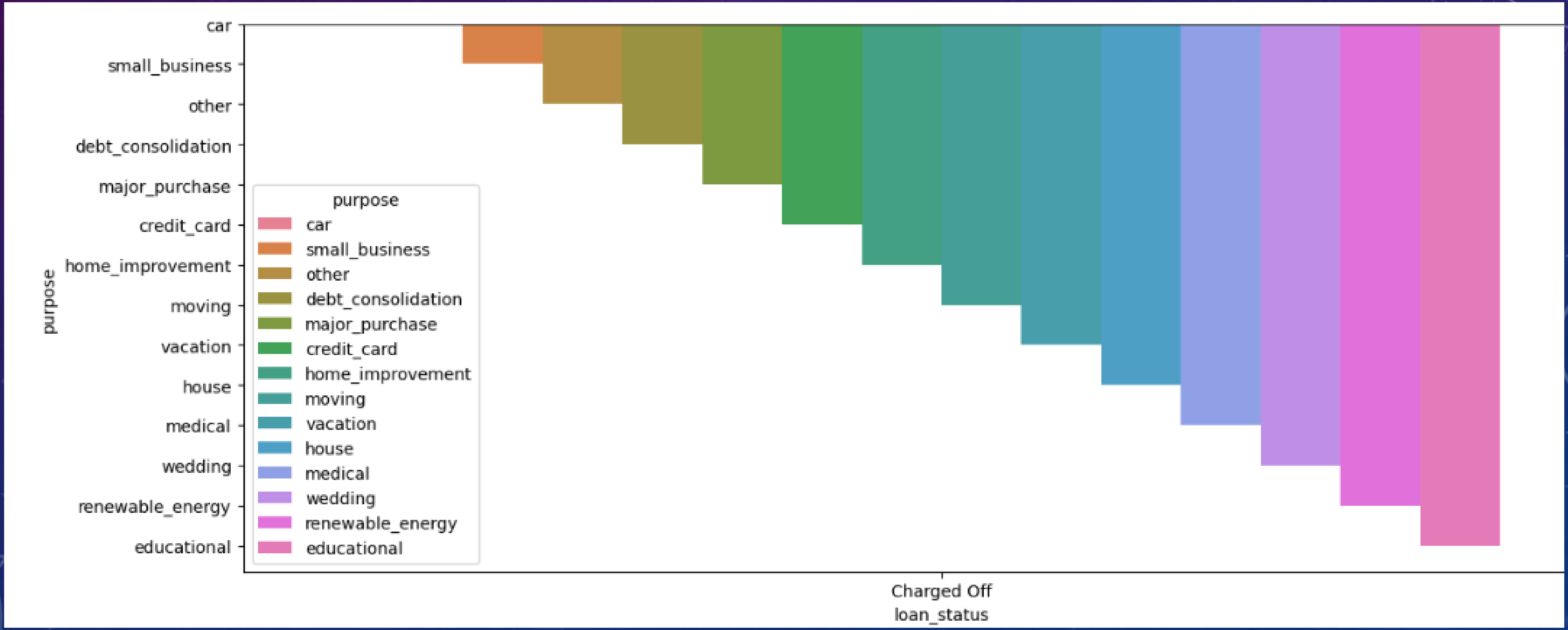
BIVARIATE ANALYSIS

- This plot illustrates the variation in loan amounts based on home ownership status. It provides insights into how owning a home affects the loan amounts borrowers receive.
- This plot shows the distribution of loan amounts based on verification status. It helps in understanding the impact of verification on the loan amounts.



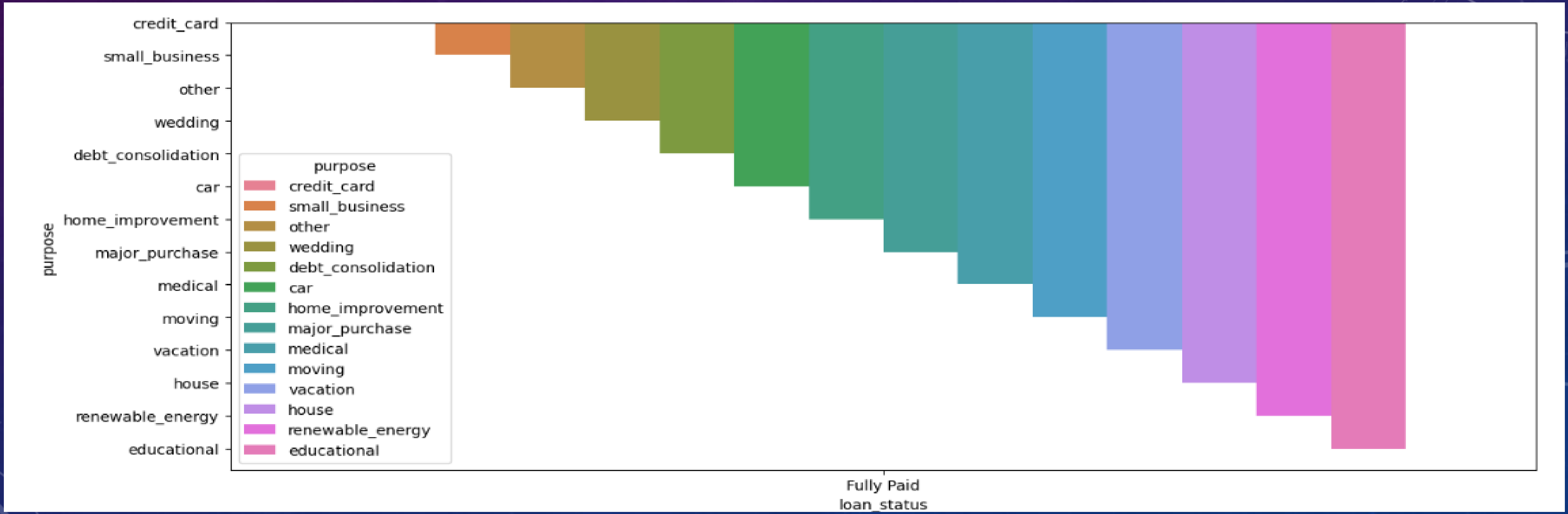
BIVARIATE ANALYSIS

This bar plot displays the distribution of defaulted loans based on their purpose. It helps in identifying which loan purposes have higher default rates.



BIVARIATE ANALYSIS

This bar plot shows the distribution of paid-up loans based on their purpose. It provides insights into which loan purposes are more likely to be fully paid.

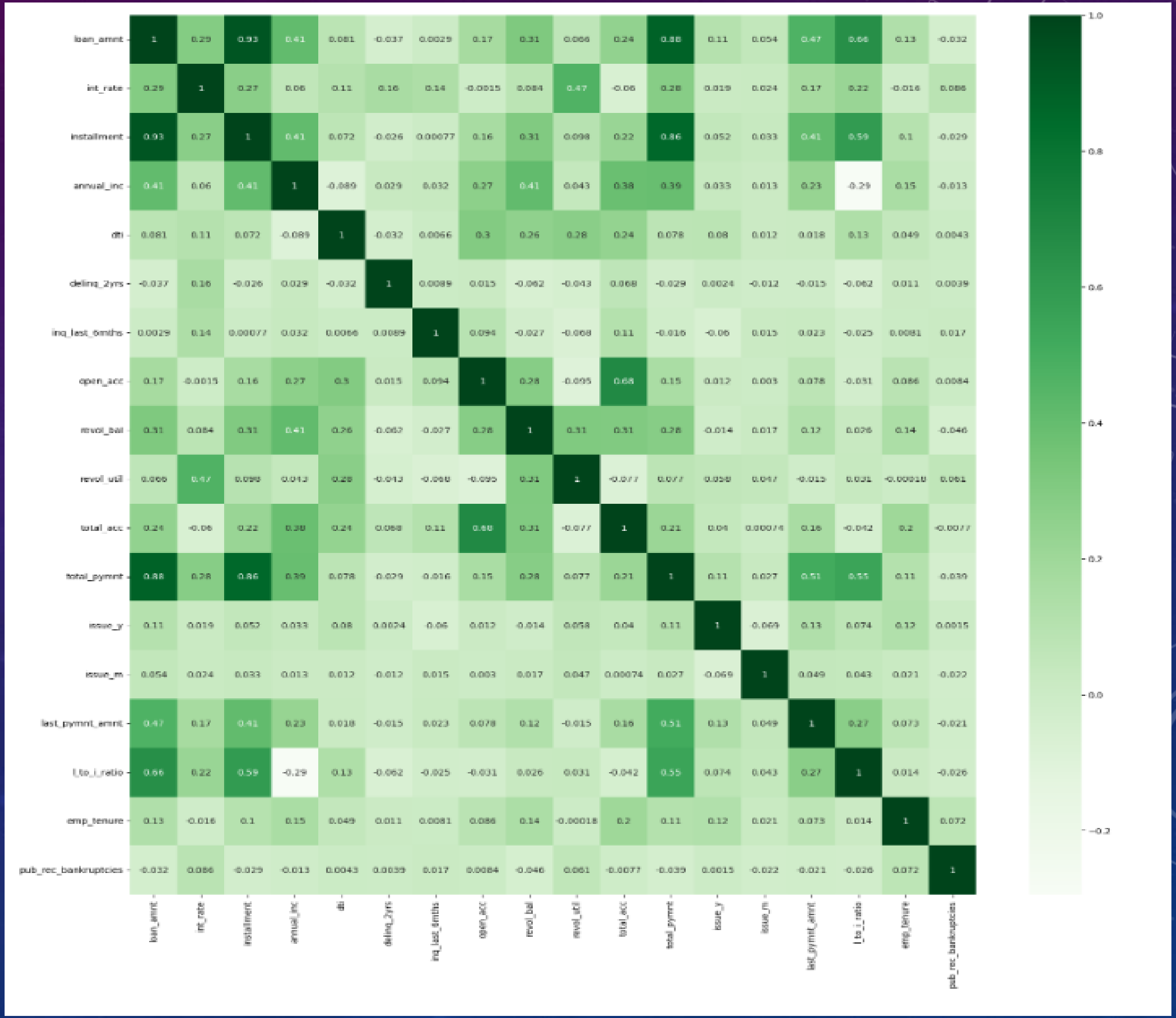


CORRELATION ANALYSIS

- Correlation analysis is a statistical method used to evaluate the strength and direction of the linear relationship between two quantitative variables.
- It helps in identifying which variables are related and how they influence each other, which is crucial for predictive modeling and decision-making.

CORRELATION ANALYSIS

The heatmap visualizes the correlation matrix of the selected numeric columns. Each cell in the heatmap represents the correlation coefficient between two variables.



CONCLUSION

- Borrowers who have a loan to income ratio of more than 30% are very likely to default
- Borrowers having high rate of revolving utilization are likely to default
- Loans taken for purposes Debt Consolidation, Credit Card, Small Business and Other are very likely to default
- Borrowers with employment tenure over have a risk of default
- Borrowers with over 100K annual income are likely to default compared to lower income group
- There is no relationship between the non demographic data and the risk of default

REFERENCES

Used Libraries	Documentation
NumPy - v1.26.4	https://numpy.org/
Pandas - v2.2.2	https://pandas.pydata.org/
Matplotlib - v3.8.4	https://matplotlib.org/
Seaborn - v0.13.2	https://seaborn.pydata.org/

GitHub Link - https://github.com/neodexter/Upgrad_Lending_Club_CaseStudy/tree/main

Thank you