Lending Club Case Study Assignment

Program: EPGP in AI-ML Aug 2024, IIIT- Bengaluru

Contributors:

- Santoshkumar Vagga
- Neeraj Kumar (Group Facilitator)

Problem statement

Objectives of Data Analysis:

Granting loans to high-risk applicants is a significant cause of financial losses, referred to as credit loss. This loss occurs when a borrower defaults or absconds with the borrowed money.

The primary goal of this analysis is to accurately identify these high-risk loan applicants, enabling the company to reduce such loans and, consequently, minimize credit loss. The case study focuses on using Exploratory Data Analysis (EDA) to identify such applicants.

The analysis aims to uncover the key factors that contribute to loan defaults—identifying the variables that serve as strong indicators of default. With these insights, the company can enhance its risk assessment and better manage its loan portfolio.

Analysis Approach

Given the dataset (loan.csv) along with its data dictionary, we identified several irrelevant columns that contained either a single value or a significant amount of missing data. These columns were removed from the dataset.

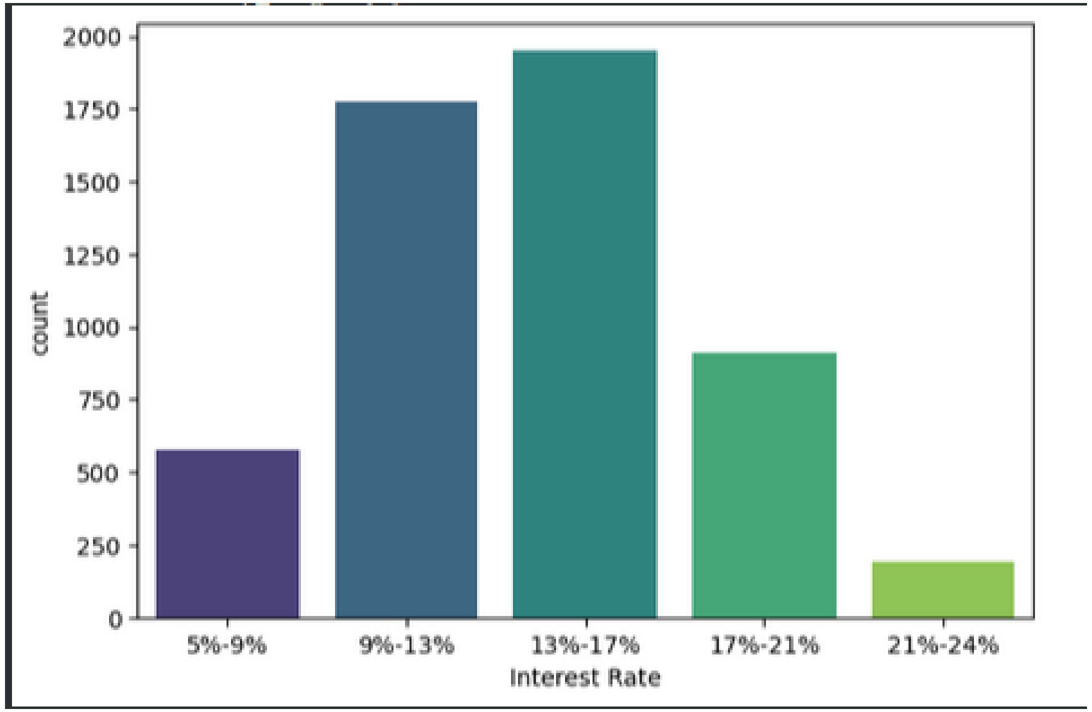
Our analysis focuses specifically on loans that are either fully paid or charged off, in order to identify patterns related to these outcomes. Loans that are still in progress were excluded from the analysis.

We addressed outliers, converted data types where necessary, standardized the values, and handled any missing data. Once the dataset was prepared, we conducted univariate and bivariate analyses, exploring various combinations of the target variable (loan status) and other relevant variables within the dataset.

Univariate Analysis – Interest rate

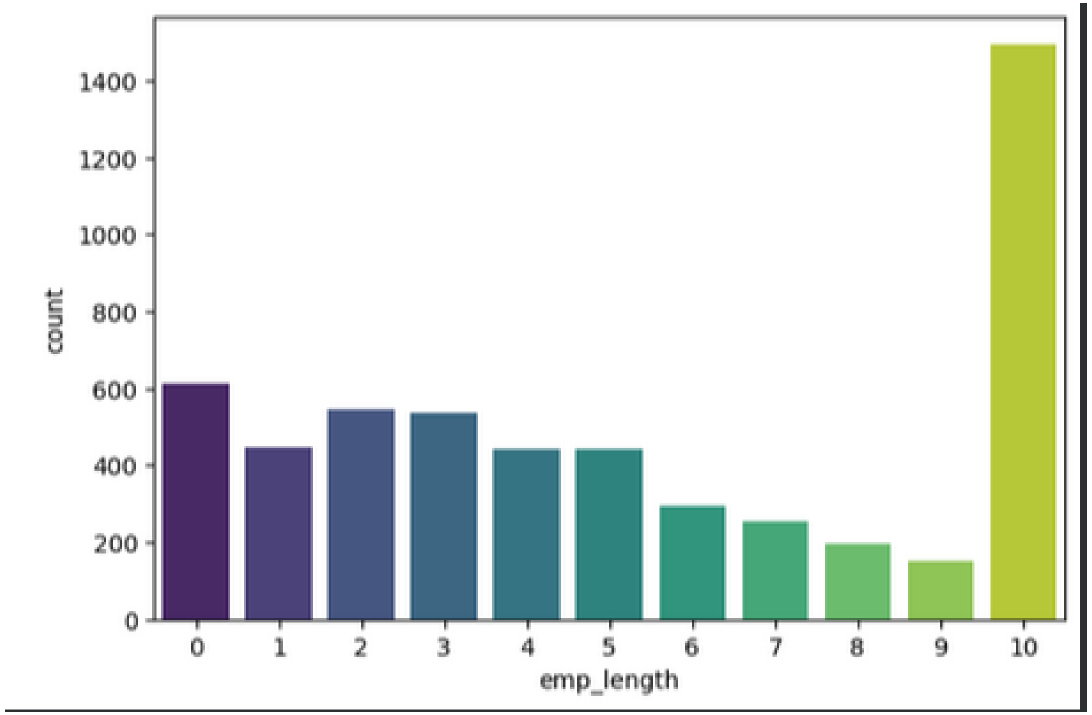
- **Applicants receiving an interest rate of 13%-17% have more chances of defaulting than other interest rates.**

- **Applicants receiving an interest rate of 21%-24% have more chances of defaulting than other interest rates.**



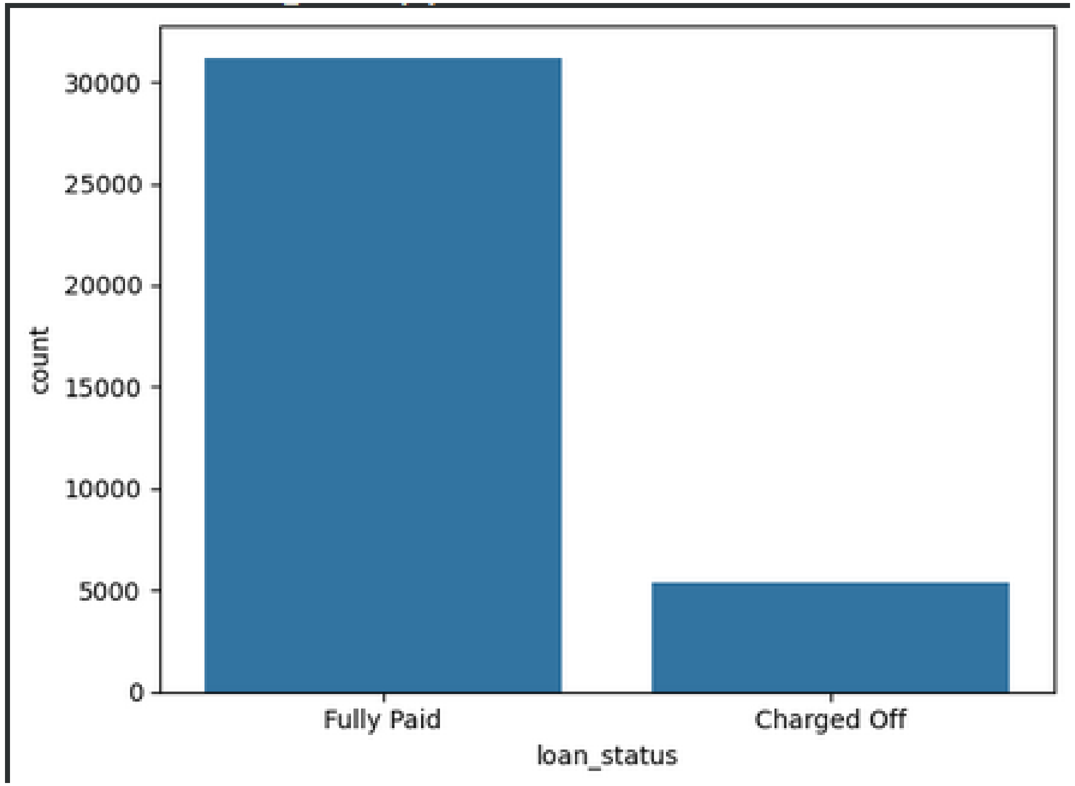
Univariate Analysis – Employment length

Applicants with an employment length of 10 years or more are more prone to defaulting.



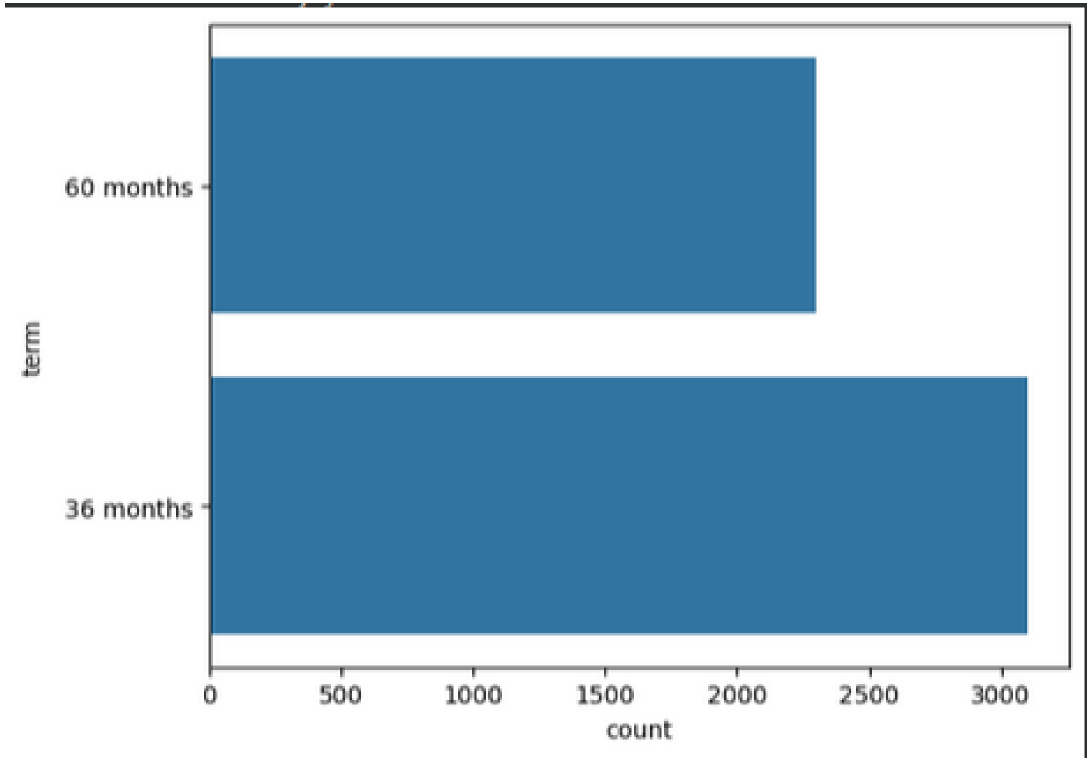
Univariate Analysis – Loan Status

- **The dataset includes 2 unique categories for loan status, which are likely "Fully Paid" and "Charged Off.**
- **The most frequent status is "Fully Paid".**
- **Out of the total entries, 31,204 loans (approximately 85%) are marked as "Fully Paid".**



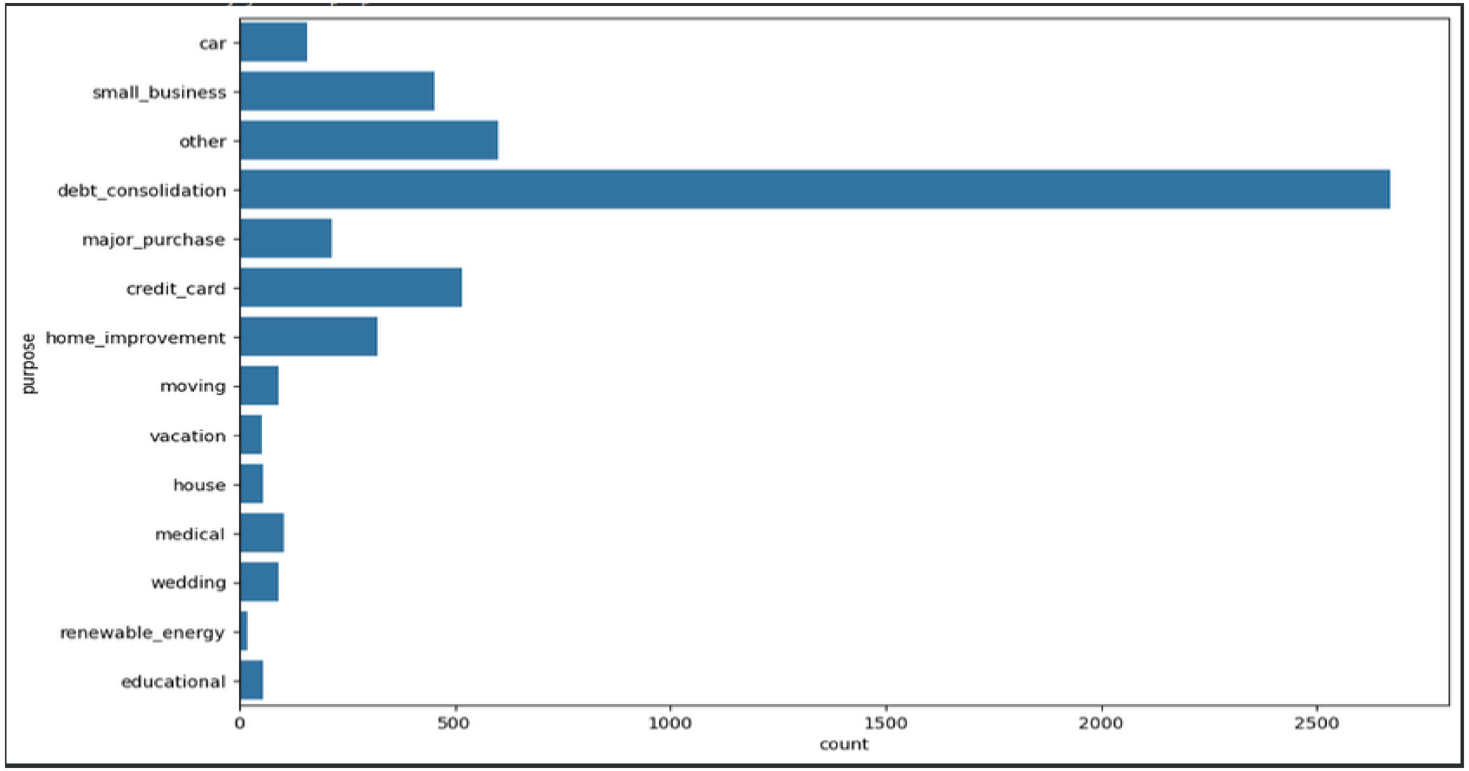
Univariate Analysis – Loan Tenure

- **Loans with a 36-month term**
- **have more probability of**
- **Defaulting compared to 60**
- **months Tenure**



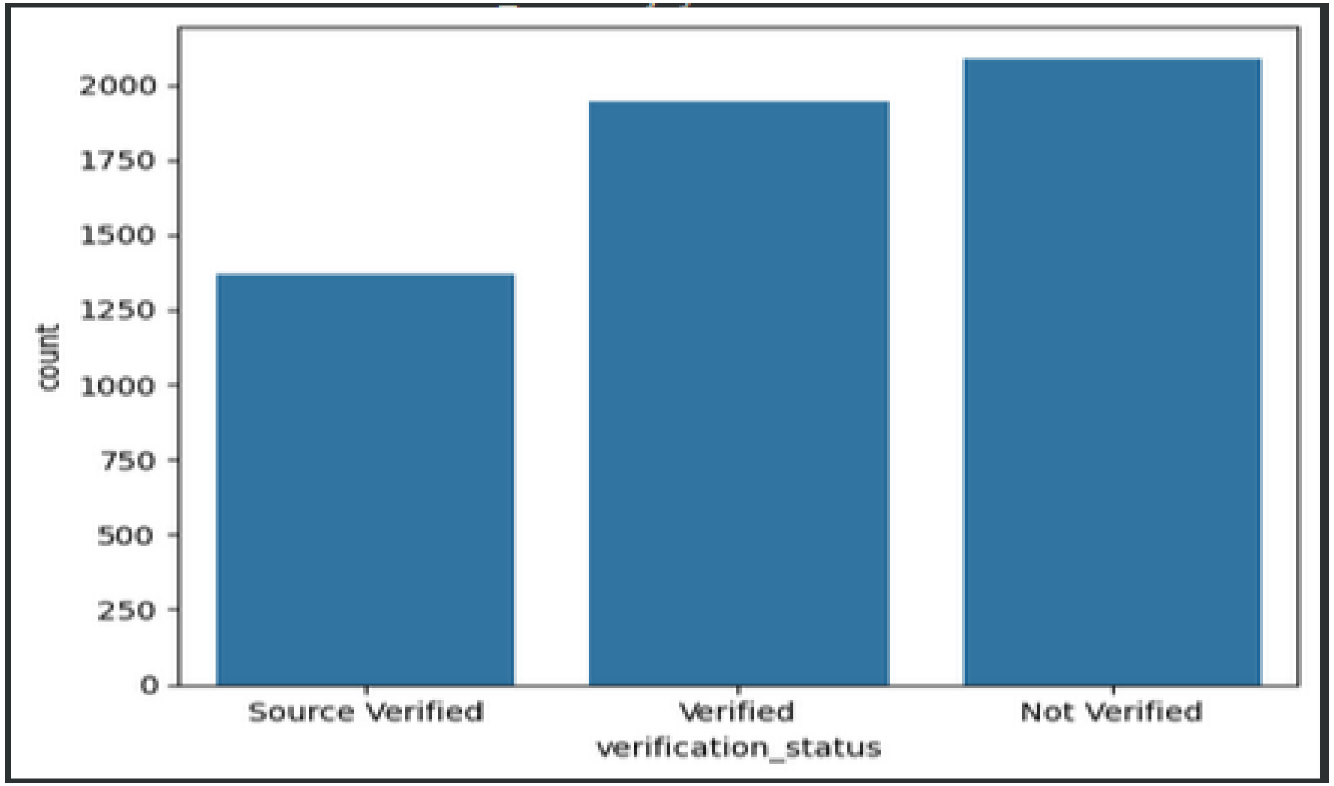
Univariate Analysis – Loan Purpose

Applicants with loan purpose as home improvement are more prone to defaulting than other loan purposes



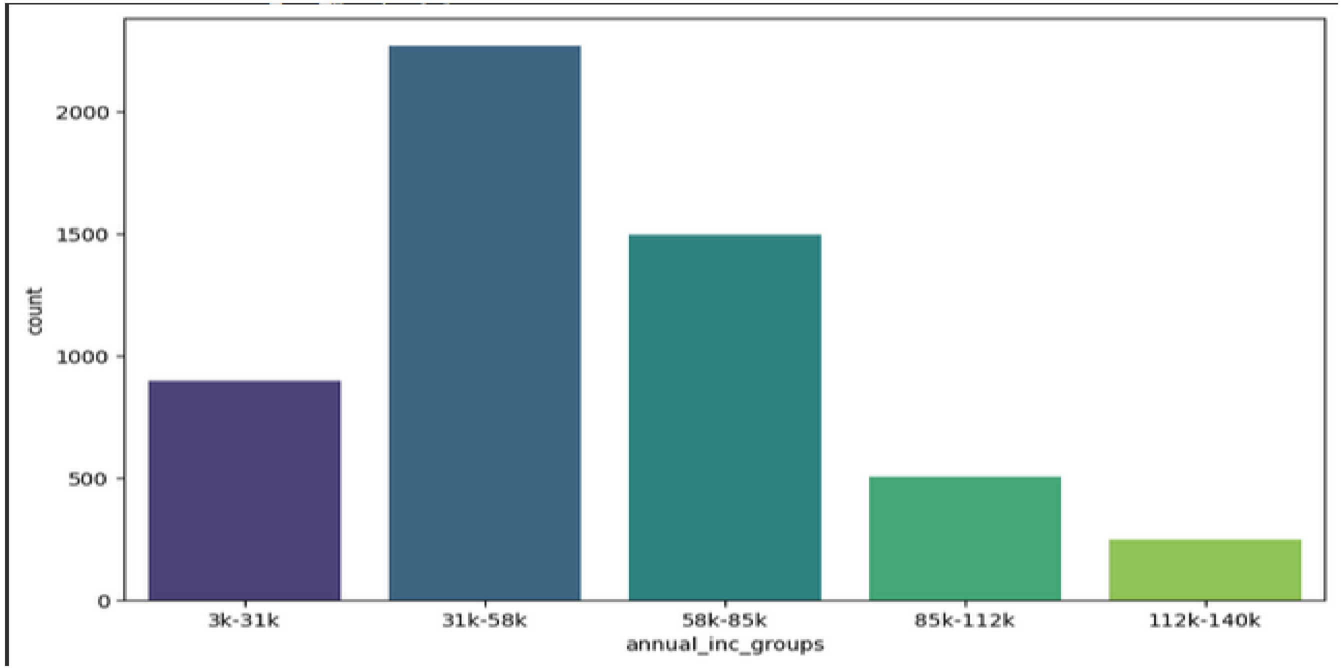
Univariate Analysis – Loan Verification Status

When loan status is "Not Verified“, such loans are more prone to defaulting.

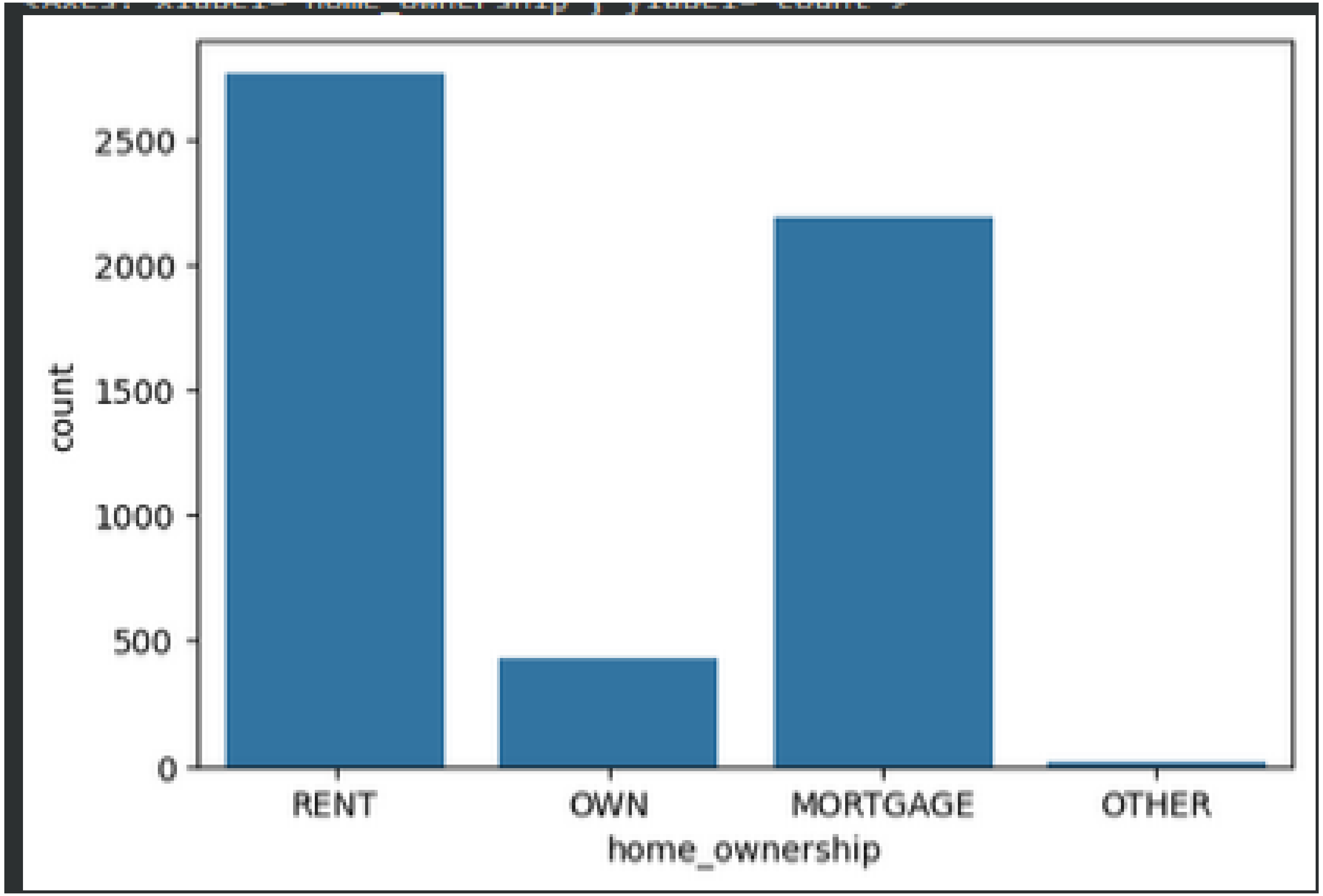


Univariate Analysis – Annual Income groups

Applicants with an income range of 31,201 to 58,402 are more prone to defaulting.



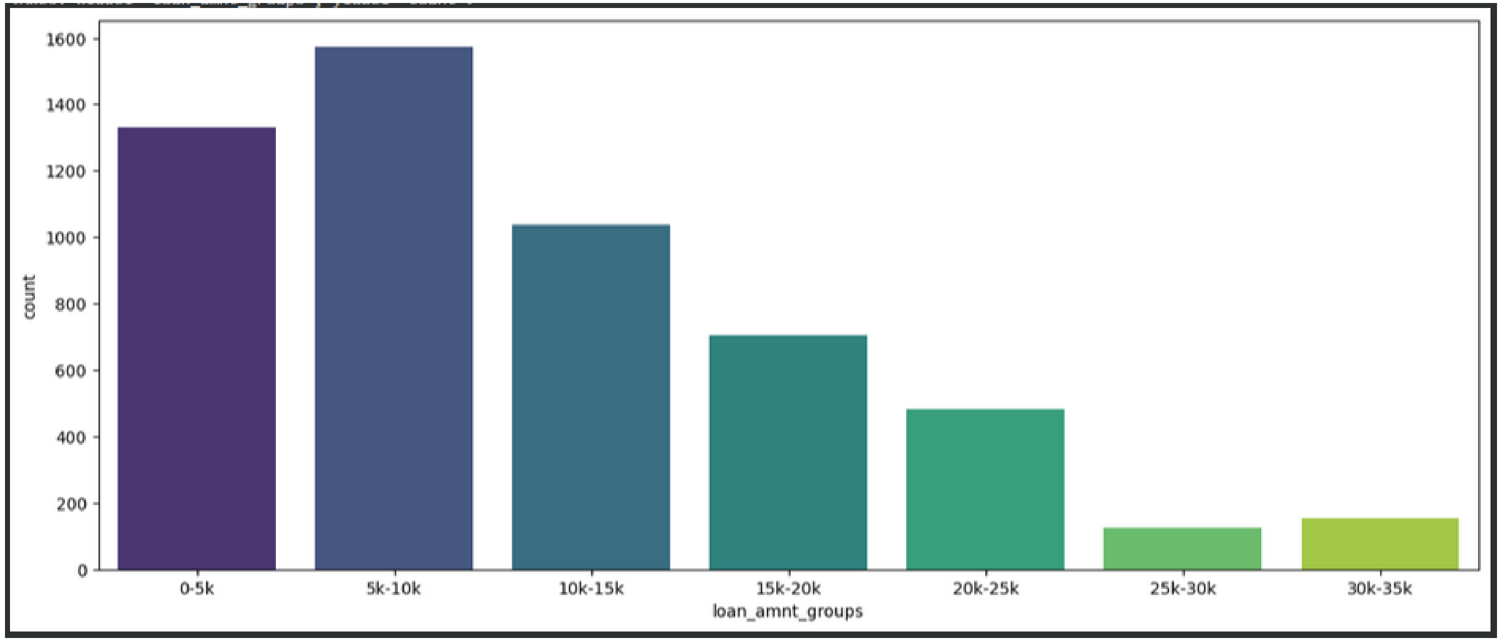
Univariate Analysis – Home ownership



Applicants with Rent homes are more prone to defaulting.

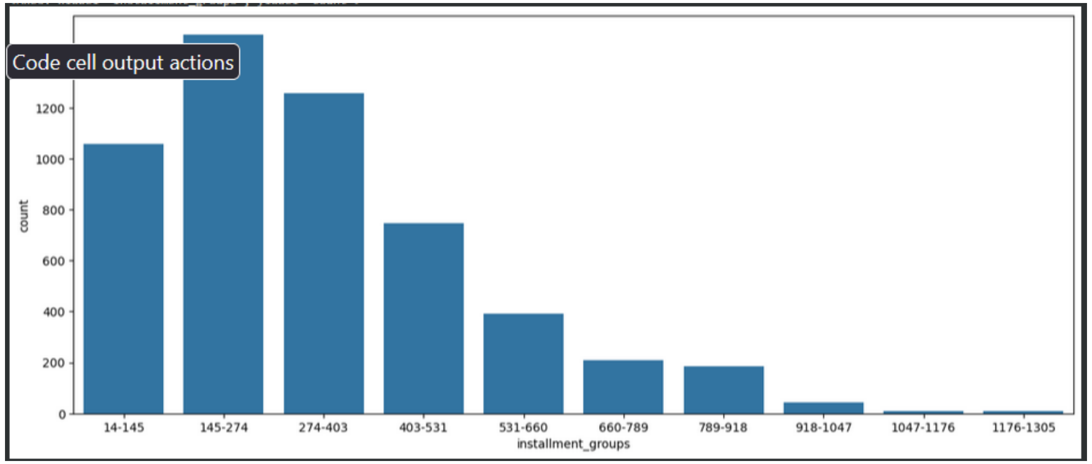
Univariate Analysis – Loan Amount group

Applicants with loan amounts ranging from 5k-10k are more prone to defaulting.



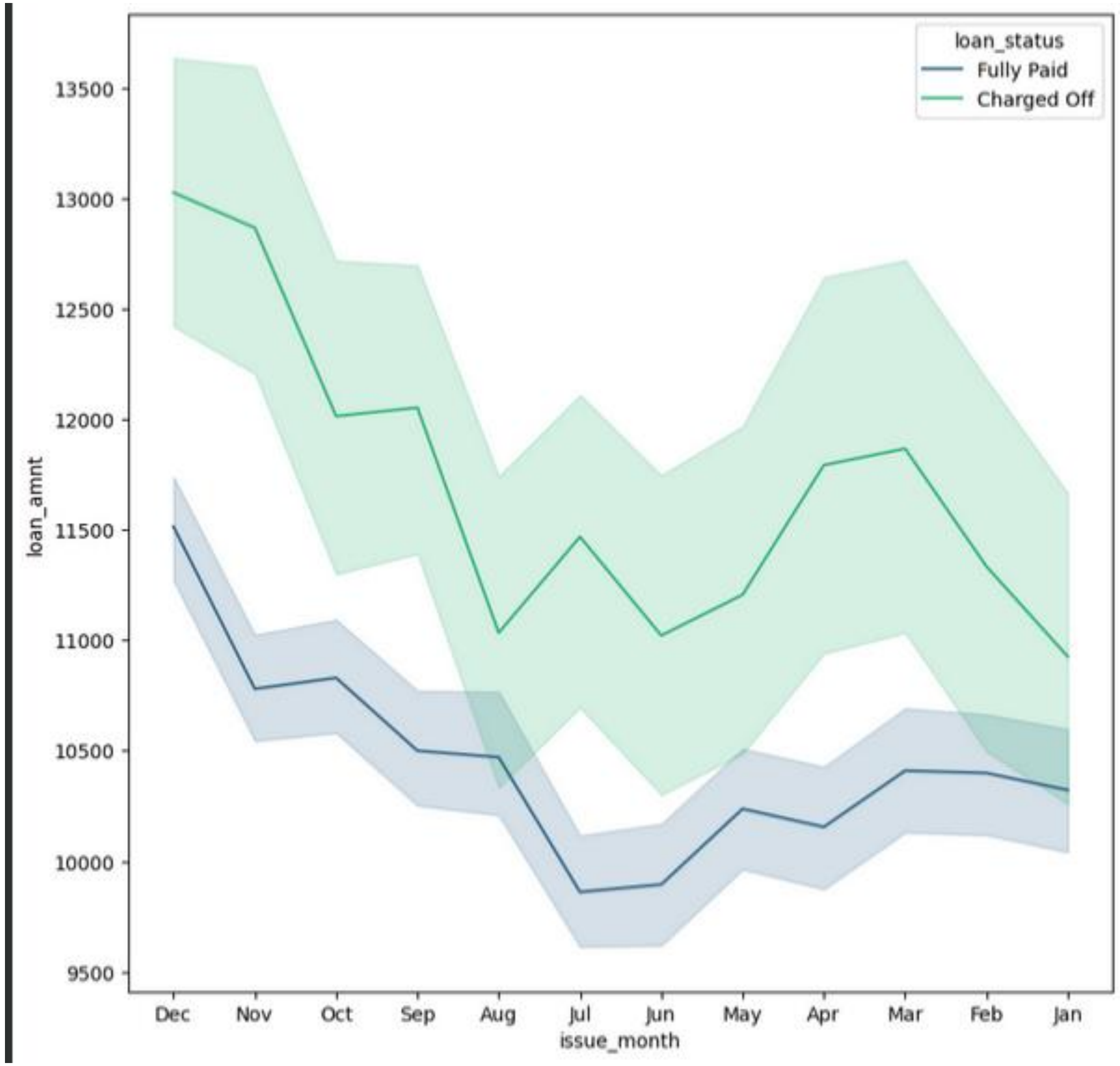
Univariate analysis – Loan Installment group

- Applicants with loan installment groups 145-274 are more prone to default, followed by range 247-403.



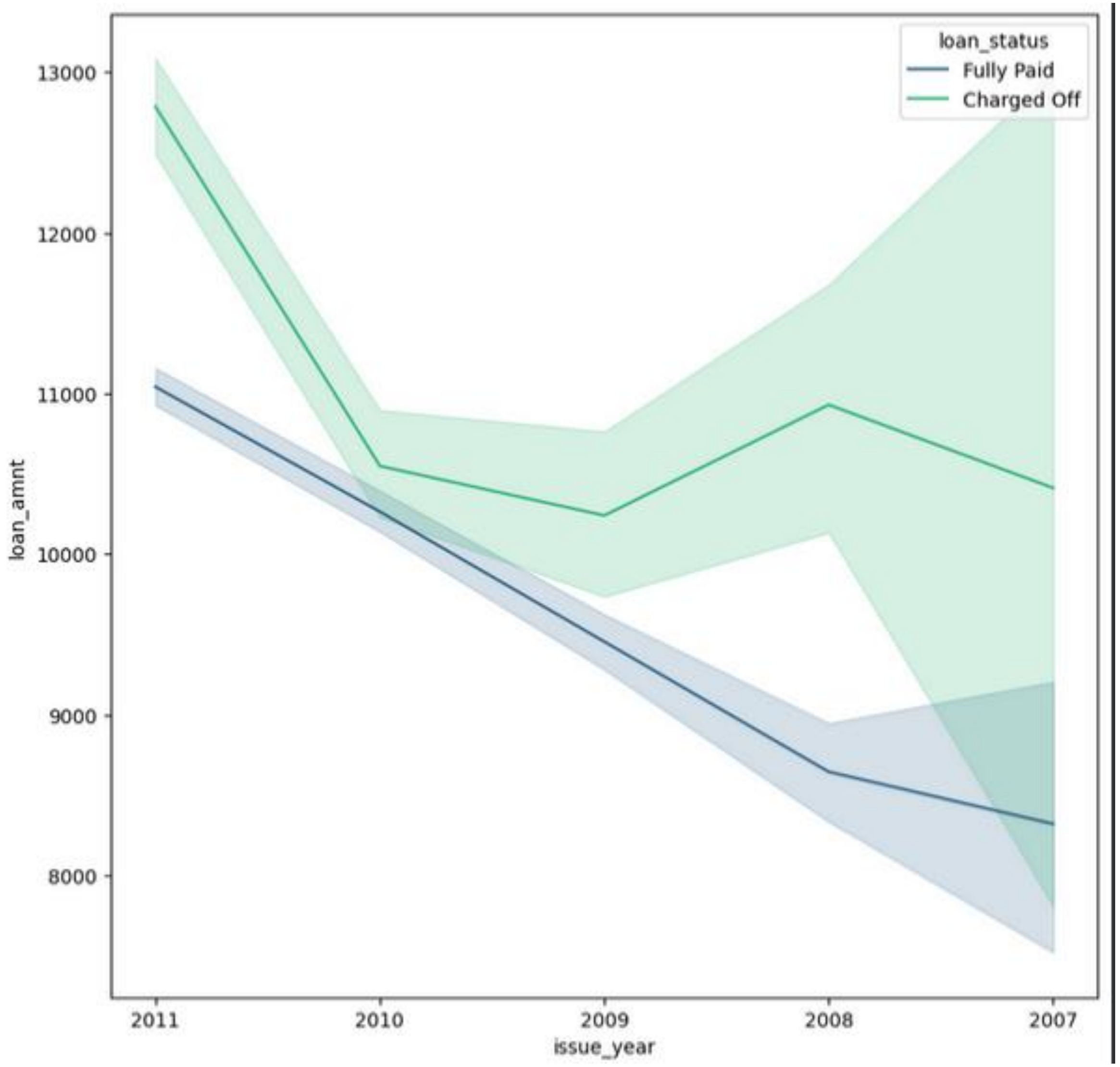
Derived analysis – Issue month

Loans issued in December month are more prone to default.



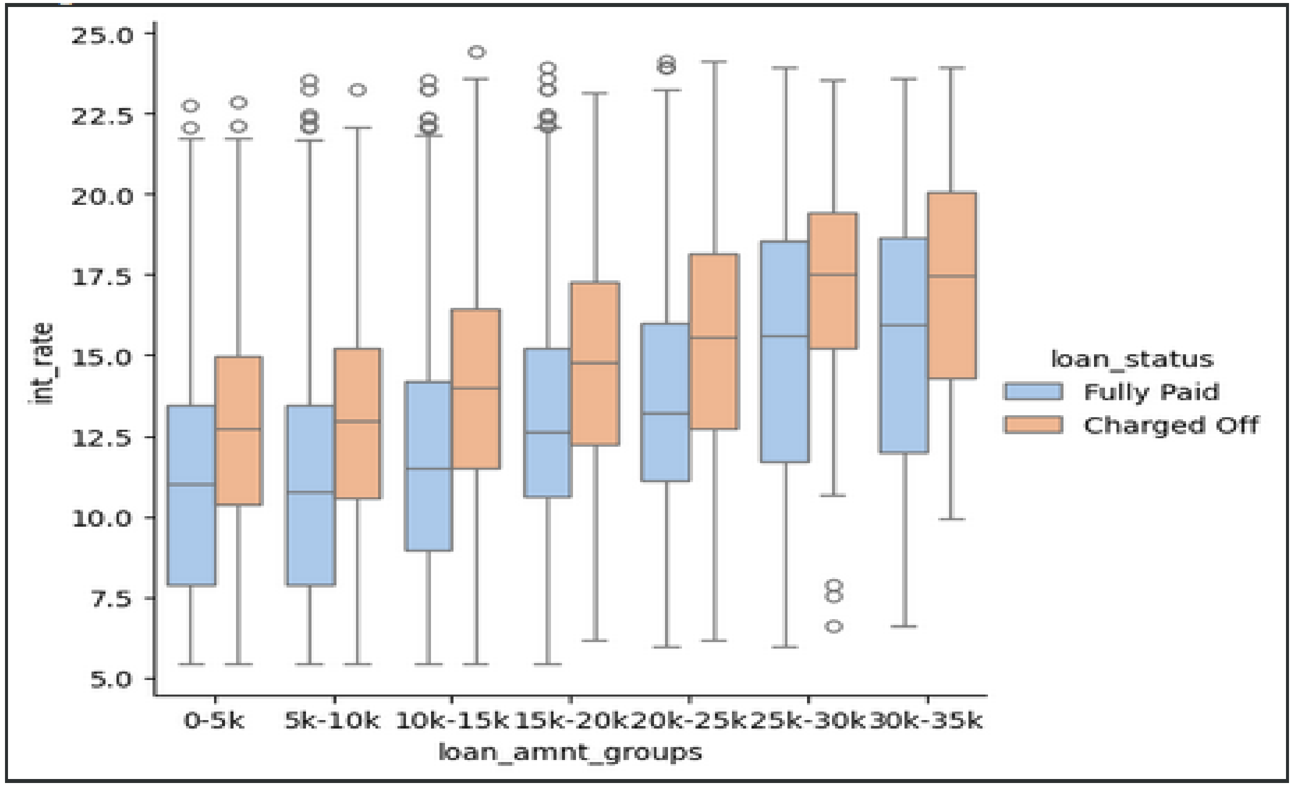
Derived analysis – Issue Year

Loans issued in year range 2011-2010 were more prone to default.



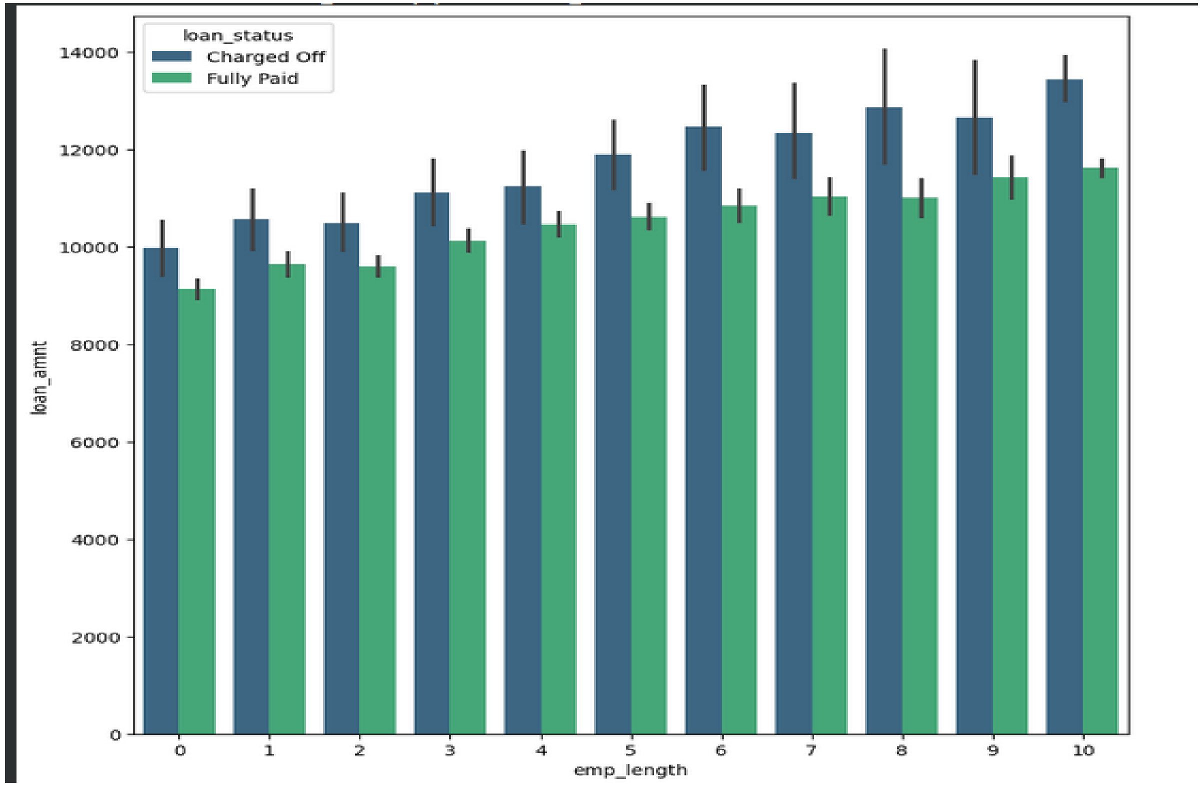
Bivariate Analysis – Interest Rate v/s loan amount groups

More the loan amounts, higher the interest rates. But defaulting is more with increasing interest rates.



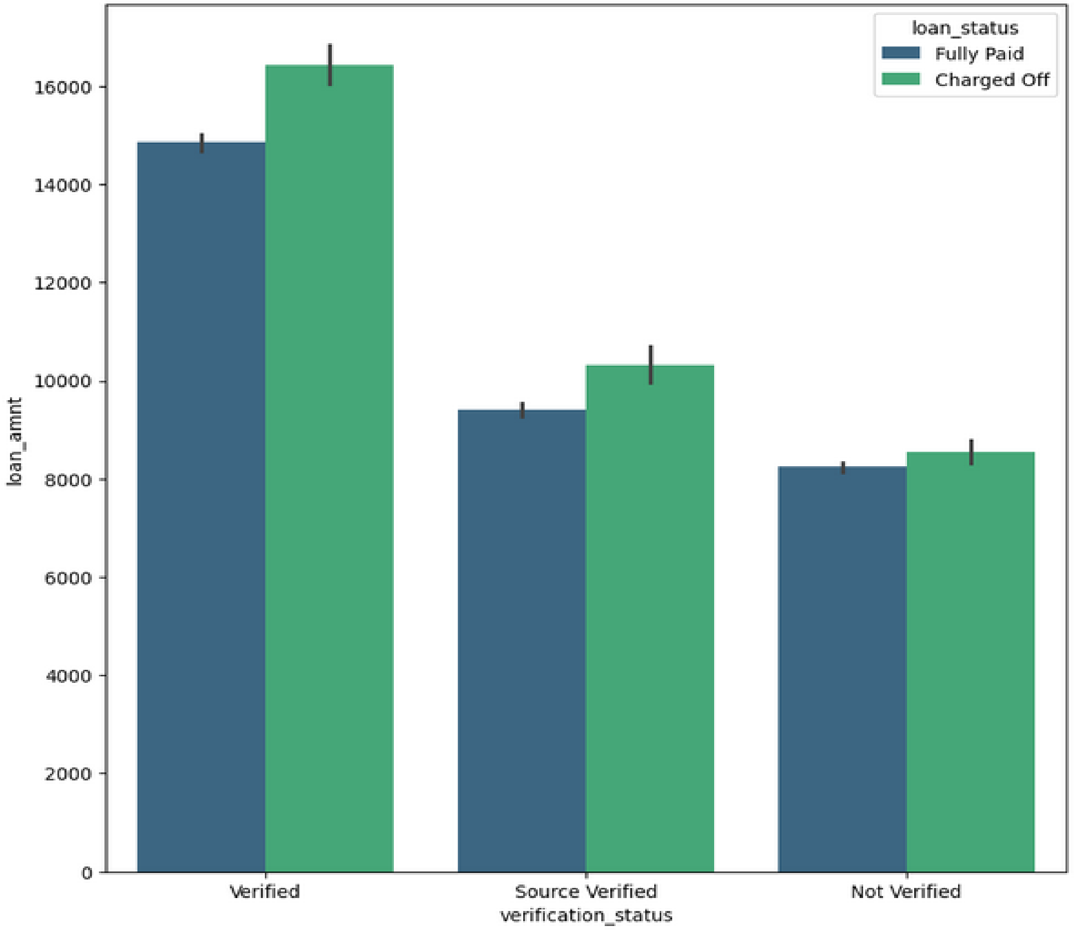
Bivariate Analysis – Loan Amount vs Employment length

Loan amount increased with higher employment length, also the trend of defaulting increased linearly with employment length.



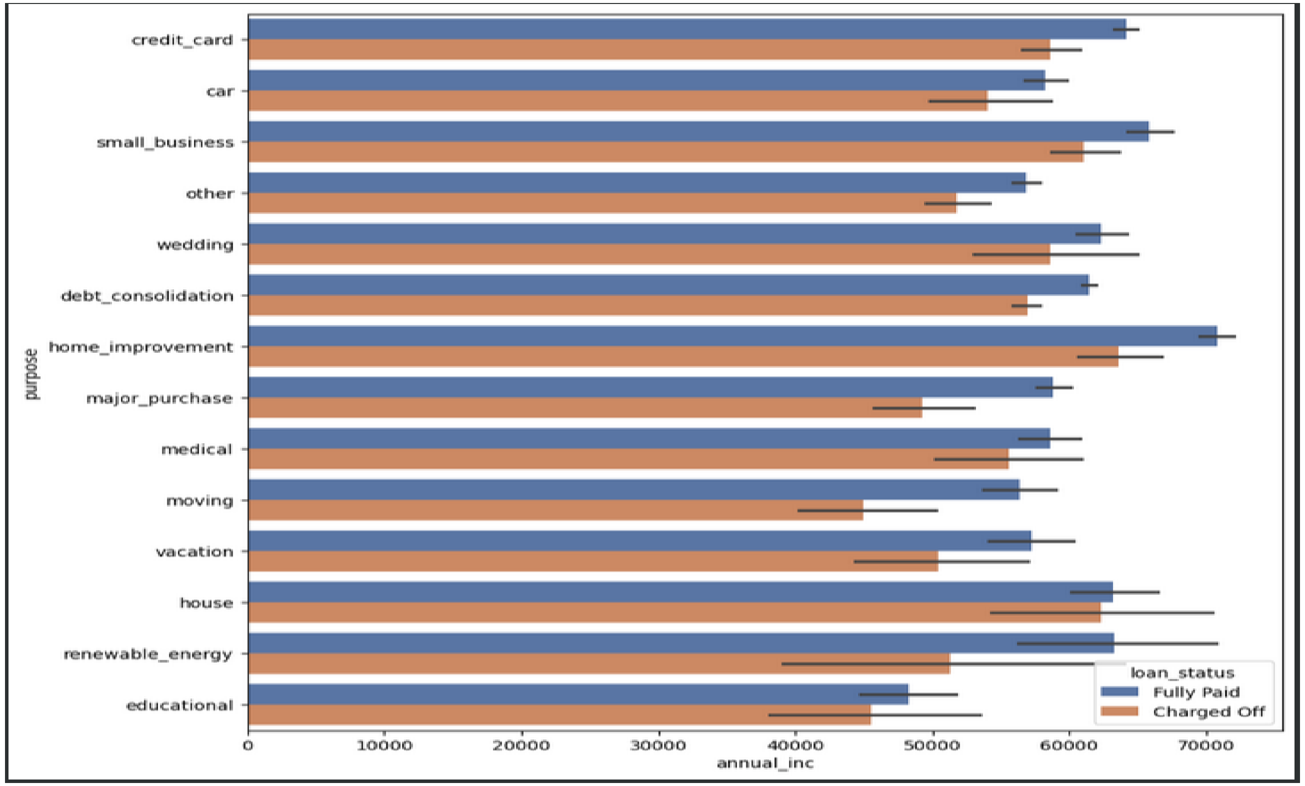
Bivariate analysis – loan amount vs verification status

All most all verification status loans have same fully paid loans vs defaulting loans.



Bivariate Analysis – Annual Inc & Purpose

Applicants with annual income 60k-70k and taking loan for purpose home improvement are more prone to defaulting



Insights

Higher Probability of Loan Defaulting:

- Applicants taking loans for *home improvement* with income of 60-70k.
- Applicants with home ownership as *MORTGAGE* and income of 60-70k.
- Applicants receiving interest rates of 21-24% with income of 70-80k.
- Loan amounts between 30k-35k with interest rates of 15-17.5%.
- Applicants taking loans for *small business* with loan amounts above 14k.
- Homeowners with *MORTGAGE* and loan amounts of 14-16k.
- Grade F loans with amounts between 15-20k.
- Employment length of 10 years and loan amounts between 12k-14k.
- Verified loans with amounts above 16k.
- Grade G loans with interest rates above 20%.

Recommendations for Lending Club agency for better profitability via minimum credit loss

Screening House Ownership:

Avoid high-risk applicants with house ownership status marked as 'RENT' or 'MORTGAGE' (especially when income is within the 60k-80k range), as these applicants have shown a higher likelihood of default.

Loan Purpose & Debt Consolidation:

Prioritize applicants with loan purposes other than debt consolidation, home improvement, or small business loans, as these purposes are associated with higher default rates.

Interest Rate Caps:

Set a maximum interest rate cap below 13%, as applicants charged rates in the 13-24% range are more prone to defaulting.

Recommendations for Lending Club agency for better profitability via minimum credit loss

Income Range Filters:

Avoid applicants with an income range of 31,201 - 58,402 and 60k-80k, especially when paired with loans for home improvement or other non-essential uses, as these income brackets appear at higher risk.

Loan Amount Control:

Be cautious with applicants seeking loans between $5,429 - $10,357 or $15k - $35k, particularly those with a DTI ratio between 12-18 and monthly installments in the range of 145 - 274.

Employment Length Verification:

Ensure that applicants' employment length is verified and consider lengthier employment as a factor. 10 years of employment is a critical cutoff point, with higher risk associated with longer durations.

Recommendations for Lending Club agency for better profitability via minimum credit loss

Credit

Credit Enquiries and Records:

Avoid applicants with 0 inquiries in the last 6 months and no derogatory public records—counterintuitively, this indicates a higher likelihood of defaulting. Verified records and inquiries are essential.

Loan

Loan Term Limit:

Prefer shorter loan terms. Defaults are more common in 36-month terms.

Grade

Lower Grade Applicants:

Be more cautious with applicants in Grade B, B5, F, and G.

Defaults increase when loans are issued to these grades, especially with high loan amounts or interest rates exceeding 20% for Grade G applicants.