

# Lending Club Case Study



Submitted By :  
Hema Lakshmi  
Dushyanth Bharadwaj  
Date: 21/08/2024

---

# CONTENTS

## Problem Statement

1. ☐ Data Description
2. ☐ Data Understanding
3. ☐ Data Cleaning & Pre-processing
4. ☐ Univariate Analysis
5. ☐ Bivariate Analysis
6. ☐ Multivariate Analysis
7. ☐ Correlation Analysis
8. ☐ Suggestions
9. ☐ References & Useful Links



# Problem Statement

- Lending Club, a Consumer Finance marketplace specializing in offering a variety of loans to urban customers, faces a critical challenge in managing its loan approval process. When evaluating loan applications, the company must make sound decisions to minimize financial losses, primarily stemming from loans extended to applicants who are considered “Risky”.
- These financial losses, referred to as Credit Losses, occur when borrowers fail to repay their loans or default. In simpler terms, borrowers labeled as “Charged-Off” are the ones responsible for the most significant losses to the company.
- The primary objective of this exercise is to assist Lending Club in mitigating credit losses. This challenge arises from two potential scenarios:
  1. Identifying applicants likely to repay their loans is crucial, as they can generate profits for the company through interest payments. Rejecting such applicants would result in a loss of potential business.
  2. On the other hand, approving loans for applicants not likely to repay and at risk of default can lead to substantial financial losses for the company.
- The objective is to pinpoint applicants at risk of defaulting on loans, enabling a reduction in credit losses. This case study aims to achieve this goal through Exploratory Data Analysis (EDA) using the provided dataset.
- In essence, 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.

# 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 39717 records 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.

# Data Understanding

Based on the dataset attributes and decision matrix you've provided, here's a concise summary of how to approach the analysis and the data interpretation:

Dataset Attributes:

Loan Status (Primary Attribute of Interest):

**Fully-Paid:** Loans have been completely repaid.

**Charged-Off:** Loans have defaulted or are not repaid.

**Current:** Loans are still active and in progress.

Note: For analysis, exclude rows with "Current" status.

# Decision Matrix

- Fully Paid: Applicants who have successfully repaid both the principal and interest.
- Charged-Off: Applicants who have defaulted on their loans due to non-payment over an extended period.
- Current: Applicants who are still making payments, hence their loan status is ongoing and cannot be determined as defaulted or fully repaid yet.
- Loan Rejection: Not included in this dataset; these are applicants who were declined for a loan and have no transactional history available.

# Analysis Approach

- **Data Filtering:** Remove rows with the "Current" loan status to focus only on loans that have been either fully repaid or charged off.
- **Outcome Evaluation:** Focus on evaluating the factors influencing loans that are Fully Paid versus those that are Charged-Off.
- **Modeling and Insights:** Use the filtered data to build predictive models or perform exploratory data analysis to understand the characteristics and patterns associated with fully paid versus charged-off loans.
- **Excluded Data:** Loan Rejection data is not available in this dataset and thus cannot be used for any analysis. Ensure to exclude these cases from any assumptions or conclusions drawn.

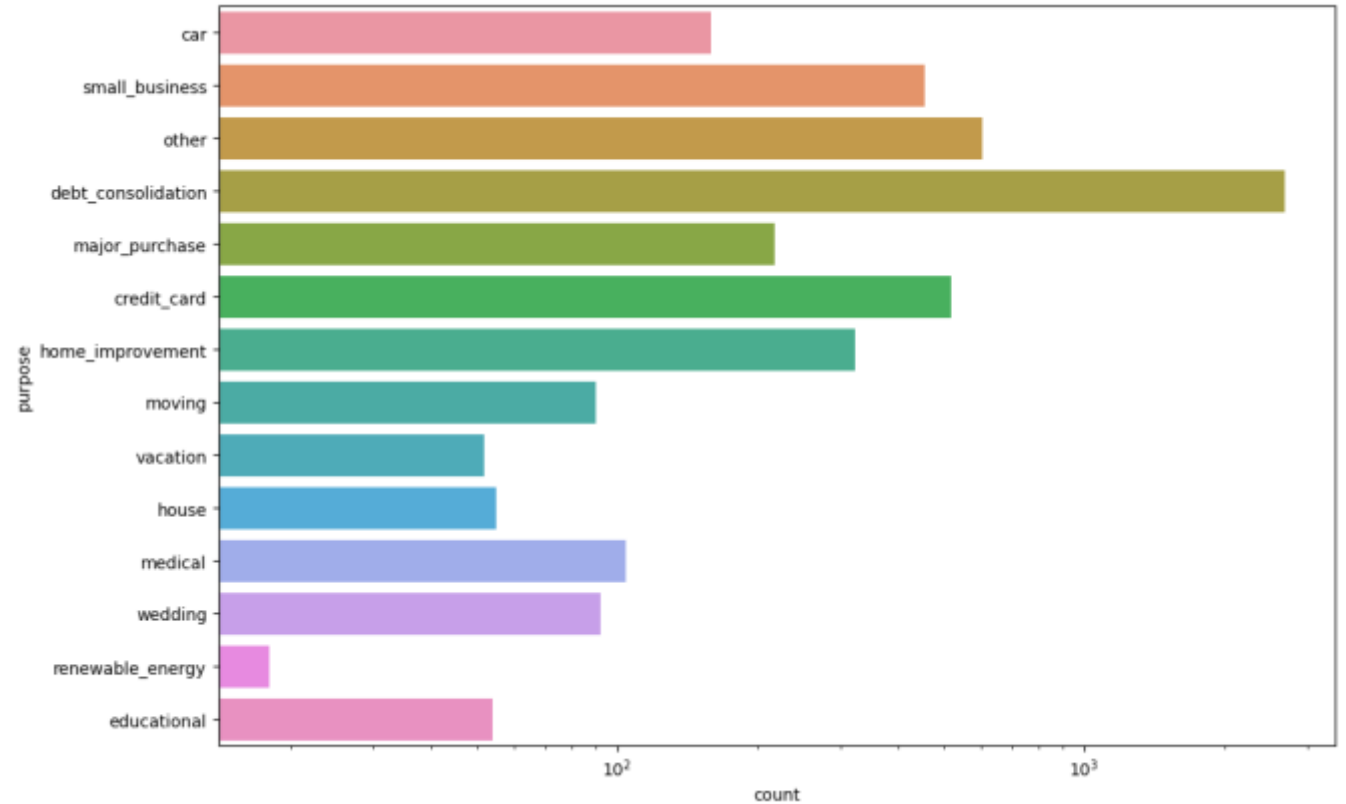
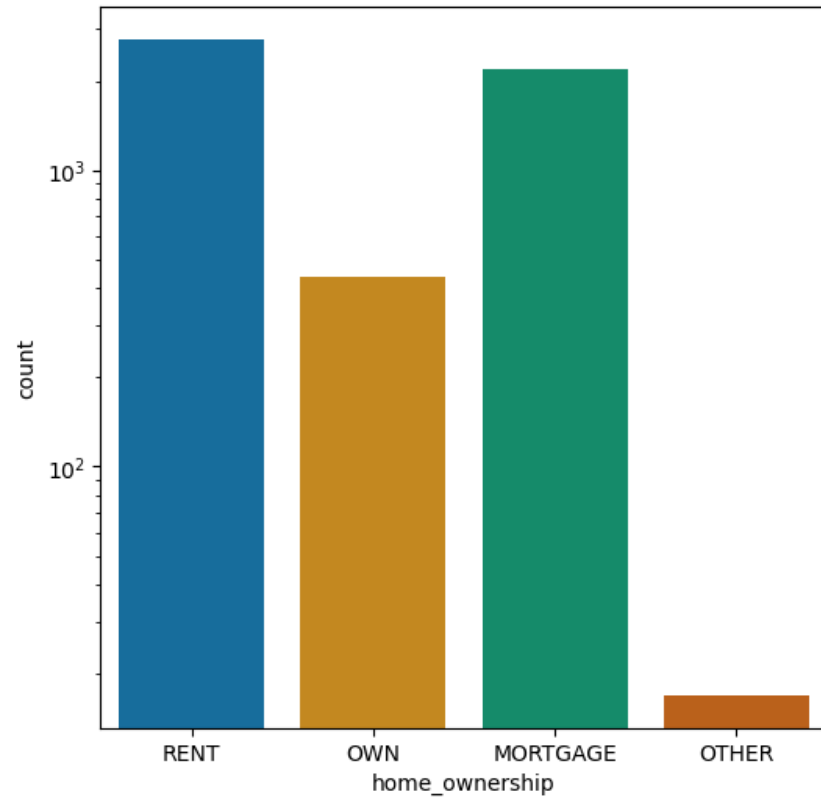
If you need further details on how to analyze the data or specific methods for modeling and interpretation, let me know!

# Data Cleaning & Pre-processing

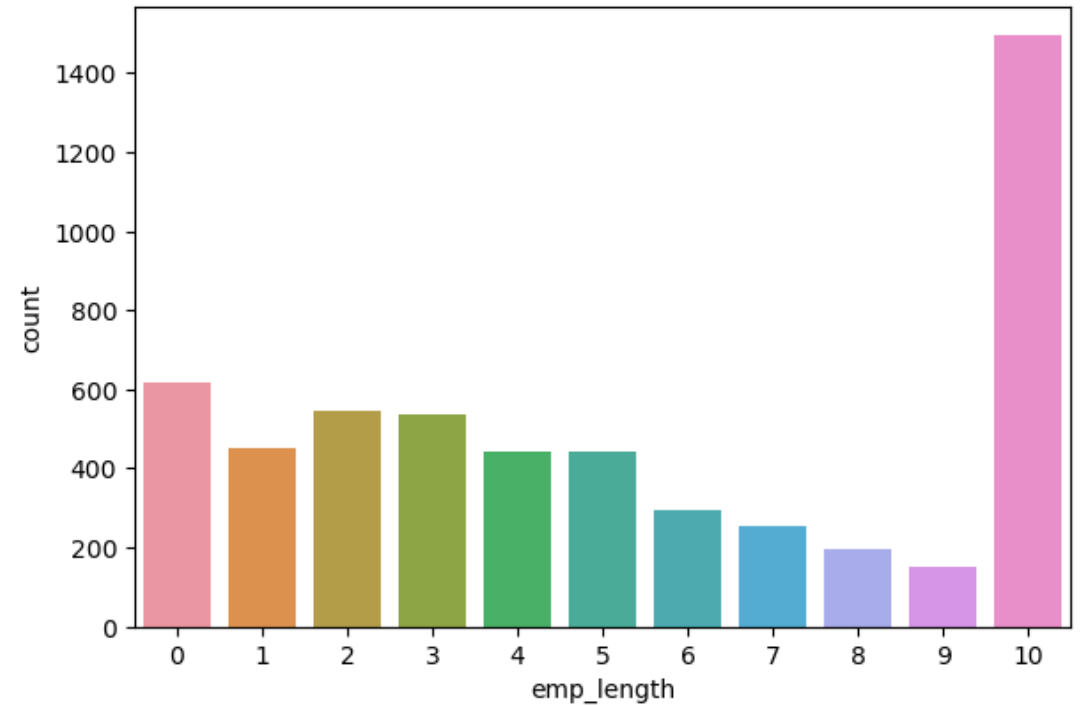
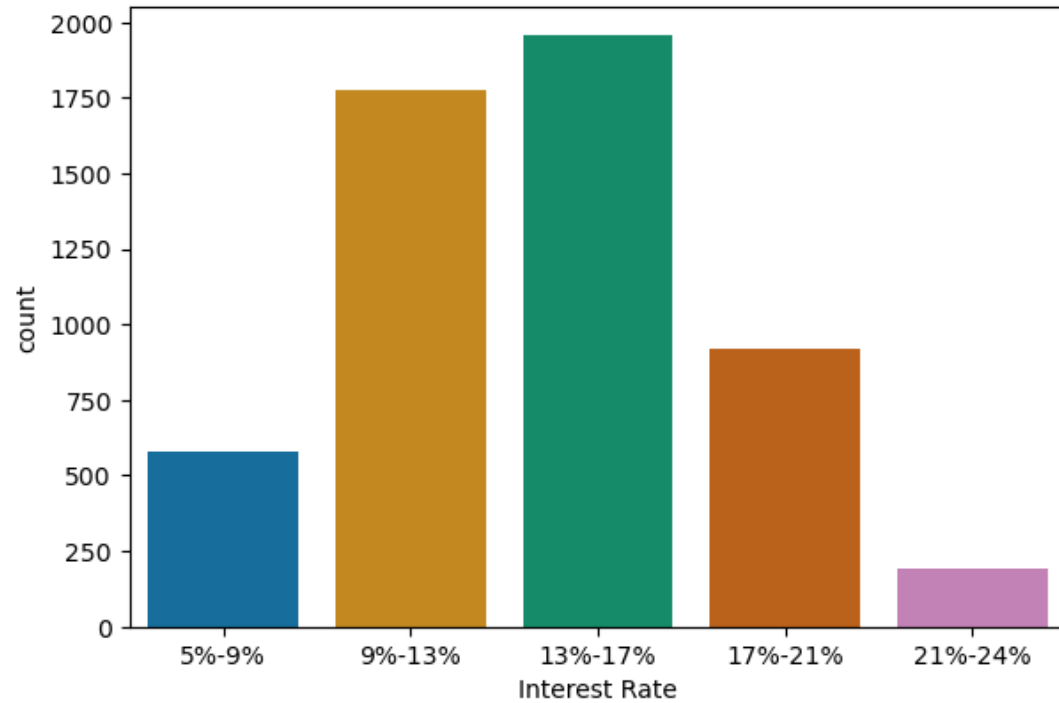
- 1.Loading data from loan CSV
- 2.Checking for null values in the dataset
- 3.Checking for unique values
- 4.Checking for duplicated rows in data
- 5.Dropping Records & Columns
- 6.Common Functions
- 7.Data Conversion
- 8.Outlier Treatment
- 9.Imputing values in Columns



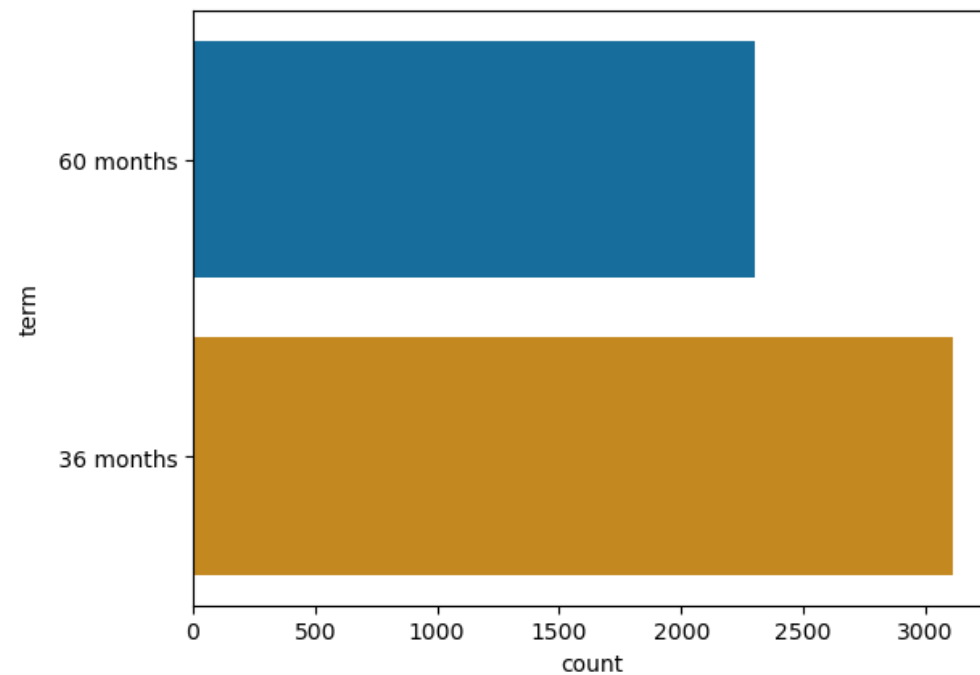
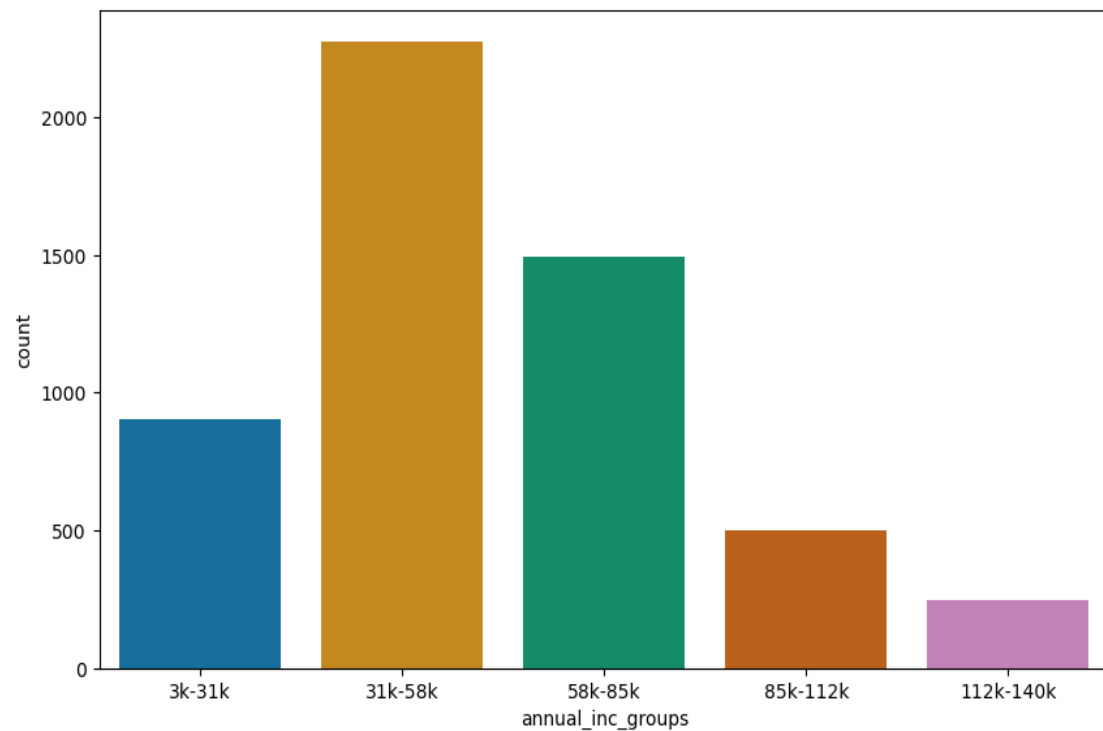
# Univariate Analysis (Quantitative Variables)



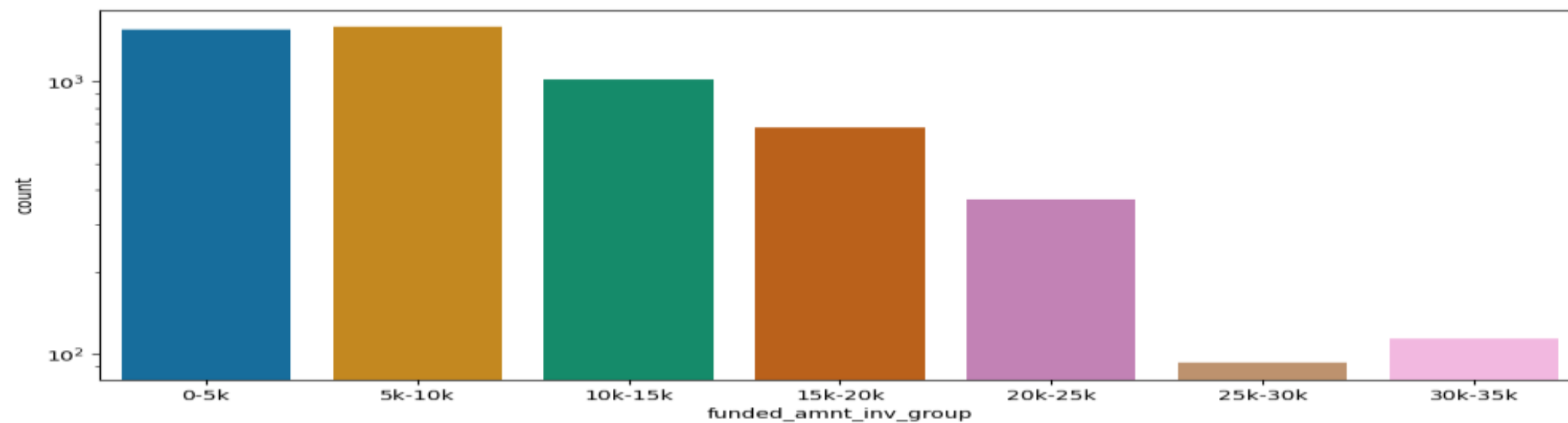
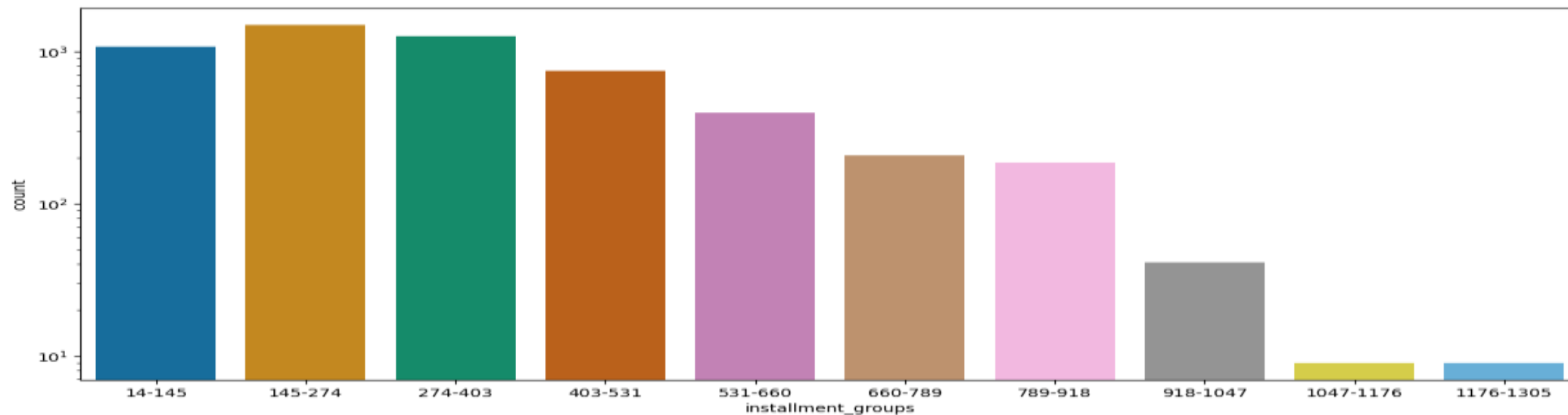
# Univariate Analysis



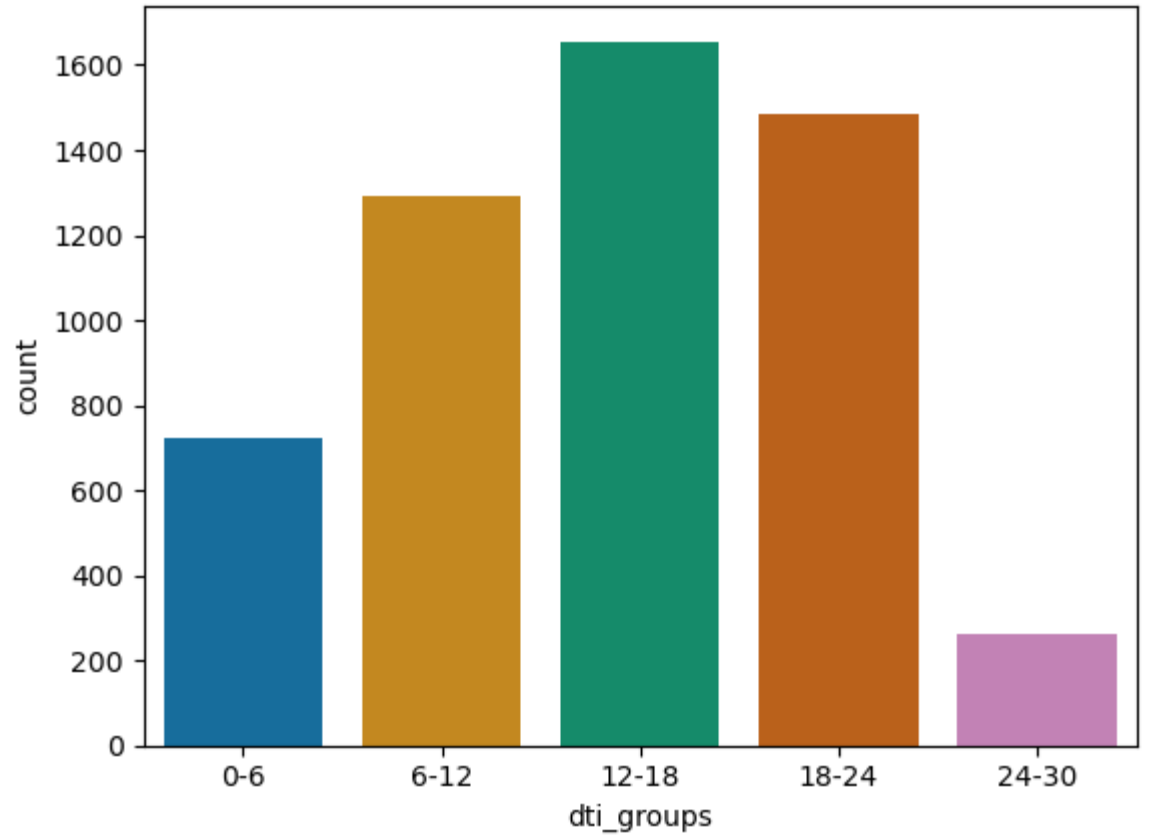
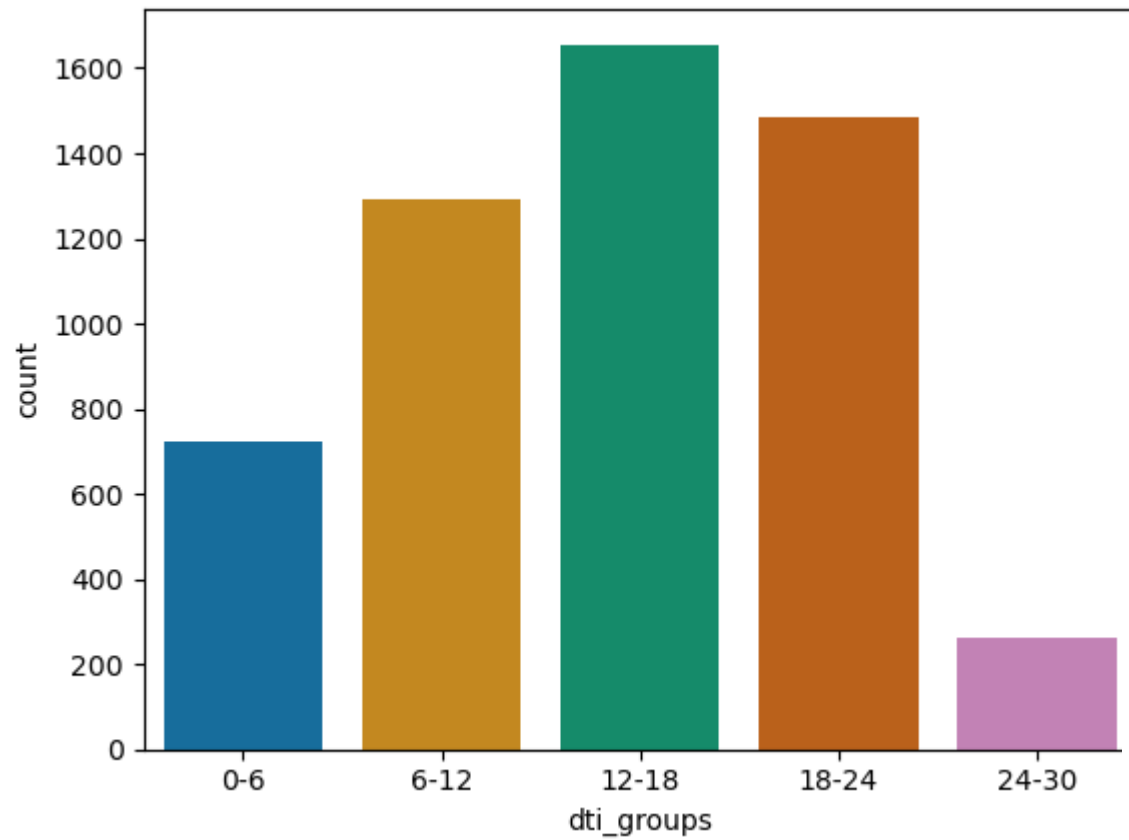
# Univariate Analysis

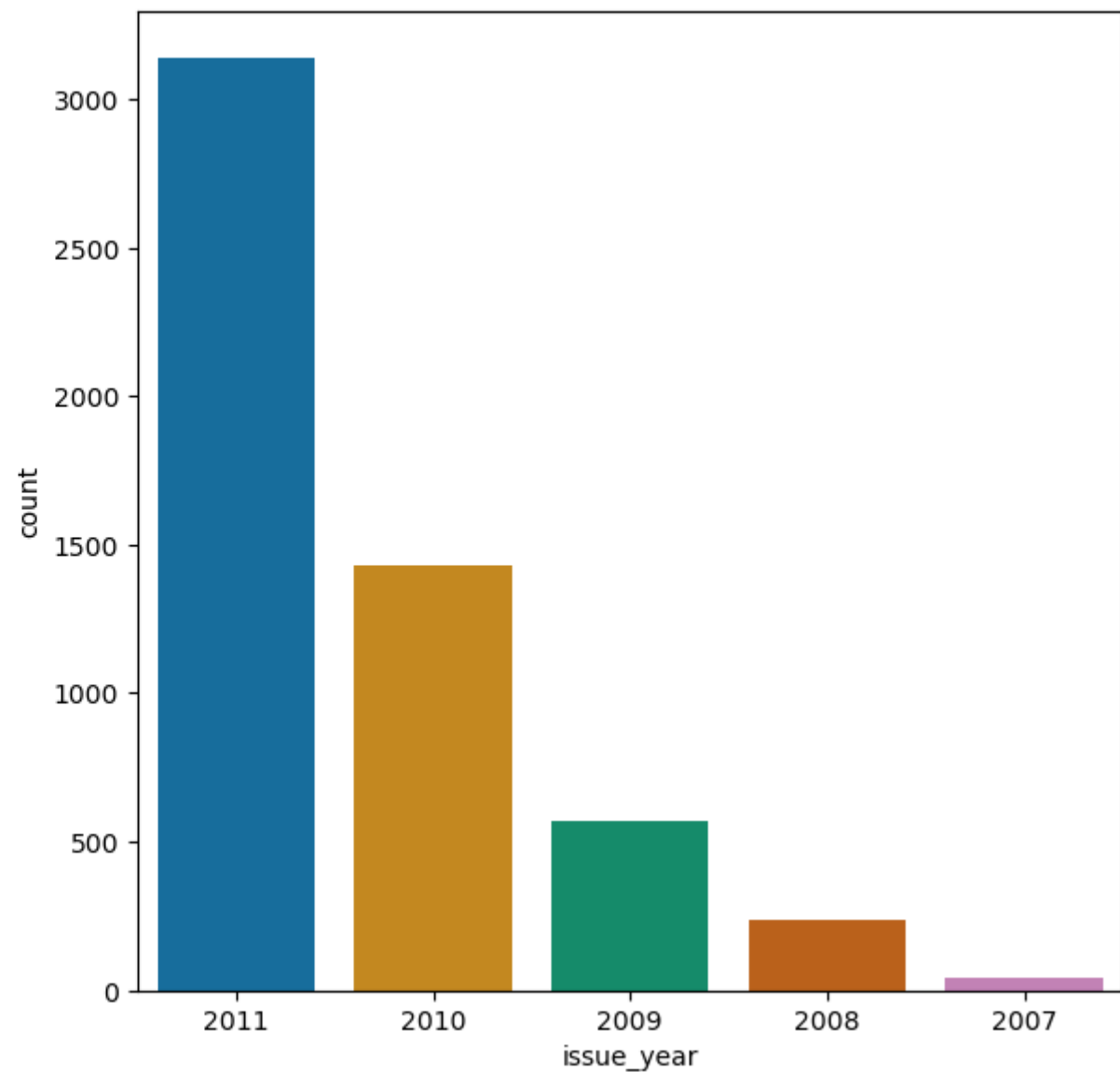
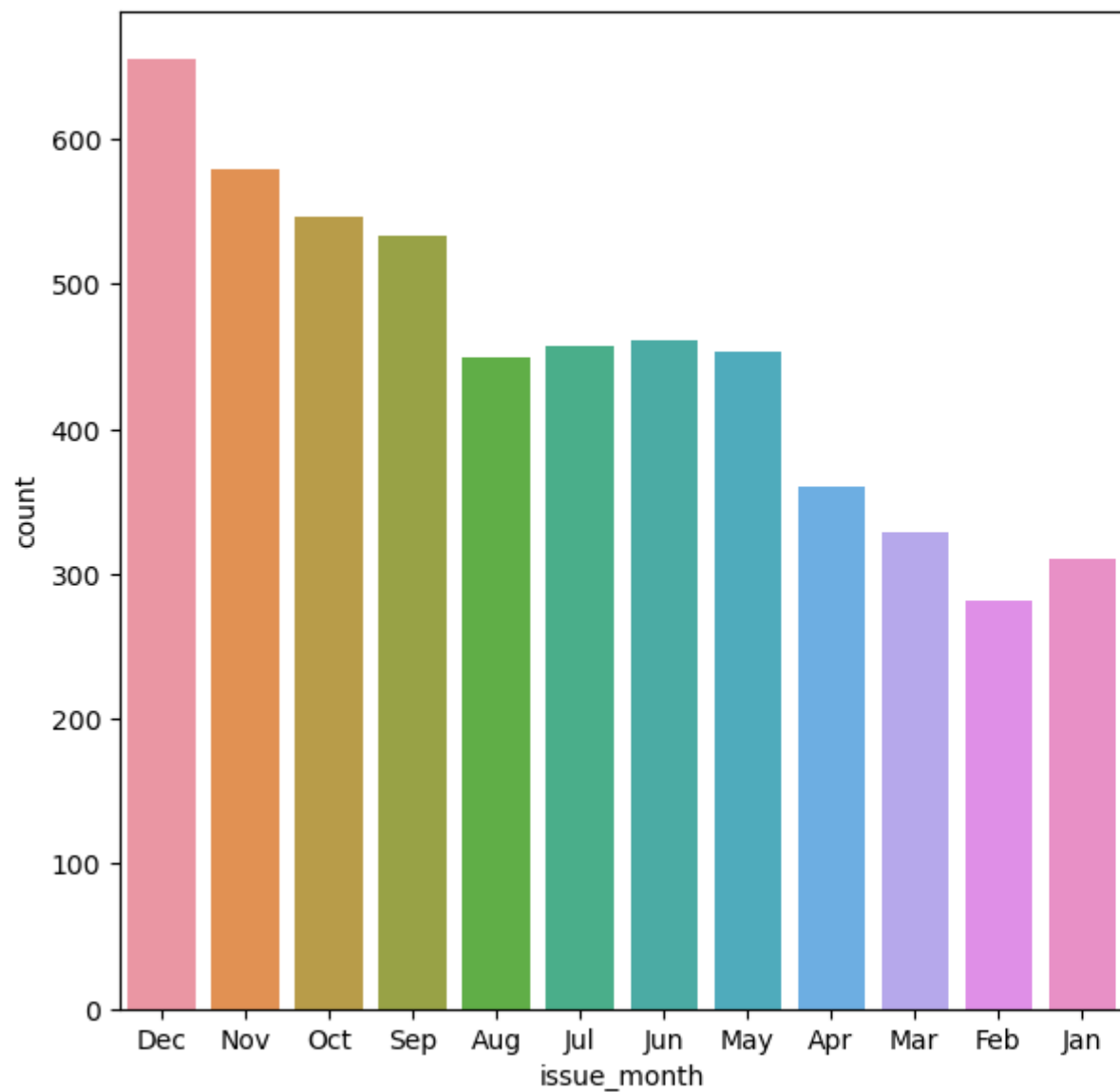


# Univariate Analysis



# Univariate Analysis





# Univariate Analysis - Observations:

- 1,561 loan applicants who defaulted had annual salaries below \$40,000. The lending company should be cautious when lending to individuals with lower annual incomes, ensuring thorough income verification and assessing repayment capacity more rigorously for this income group.
- Out of the 2,025 loan participants who defaulted, a significant number were in the 13%-17% interest rate bracket. To mitigate the risk of default, the lending company should consider offering loans at lower interest rates when feasible.
- 1,695 loan participants who defaulted had loan amounts of \$15,000 or more. The lending company should carefully evaluate applicants requesting higher loan amounts, ensuring they have a strong credit history and the ability to repay larger loans.
- 1,608 loan participants who defaulted received funded amounts of \$15,000 or more. The lending company should ensure that funded amounts match the borrower's financial capacity and conduct thorough credit assessments for larger loan requests.
- Among the loan participants who defaulted, 1,178 had very high debt-to-income ratios. The lending company should enforce strict debt-to-income ratio requirements to avoid lending to individuals with unsustainable levels of debt relative to their income.
- For those who defaulted, the majority had monthly installment amounts between \$160 and \$440. The lending company should closely monitor and assess applicants with similar installment amounts to reduce the risk of loan defaults.

# Bivariate Analysis

Bivariate analysis is a statistical method that involves the simultaneous analysis of two variables (factors). It aims to determine the empirical relationship between them. The analysis can be used to test hypotheses, identify patterns, or explore relationships between the variables.

✓ It was carried out for both Categorical and Quantitative Variables



# Bivariate Analysis (Unordered Categorical)

