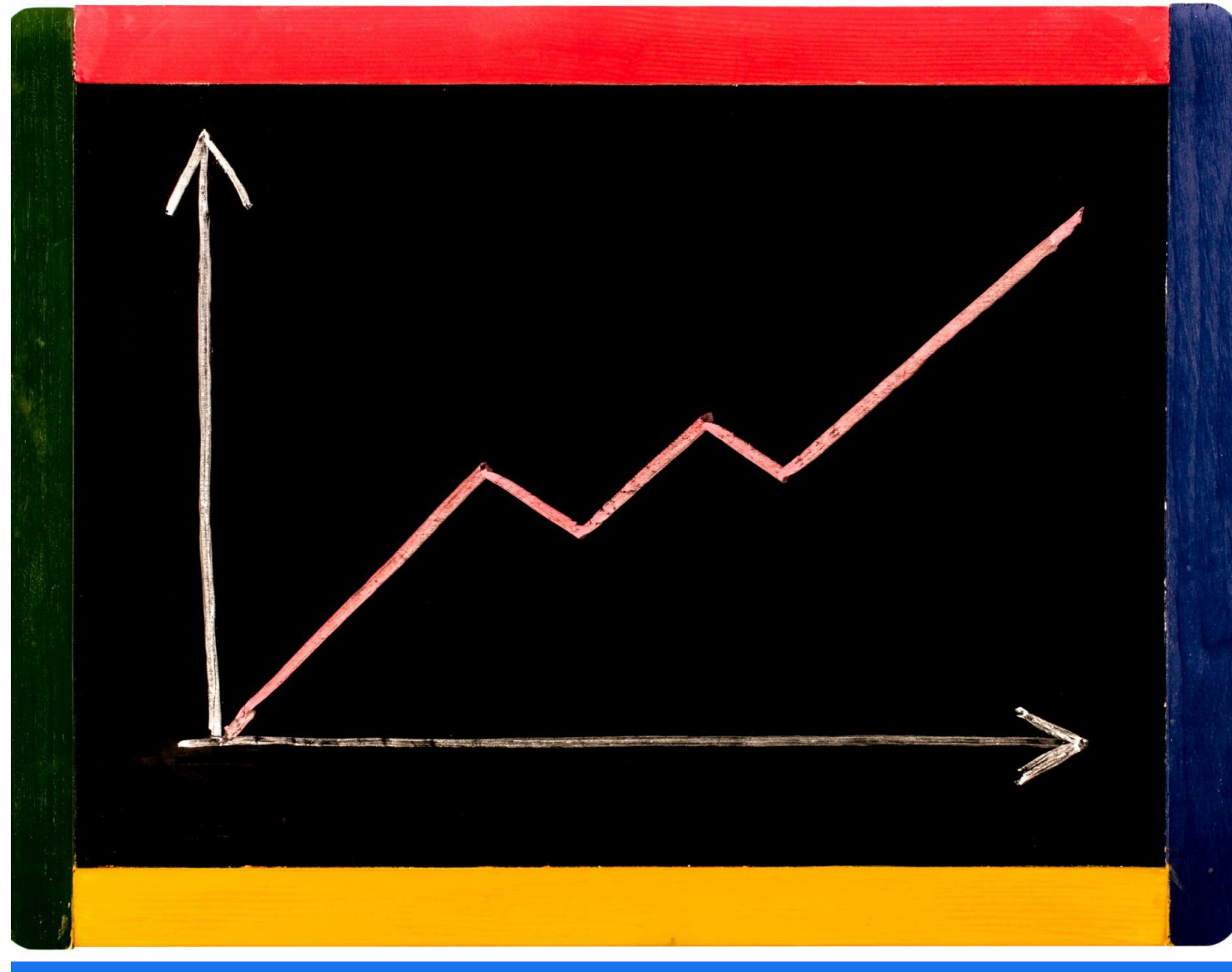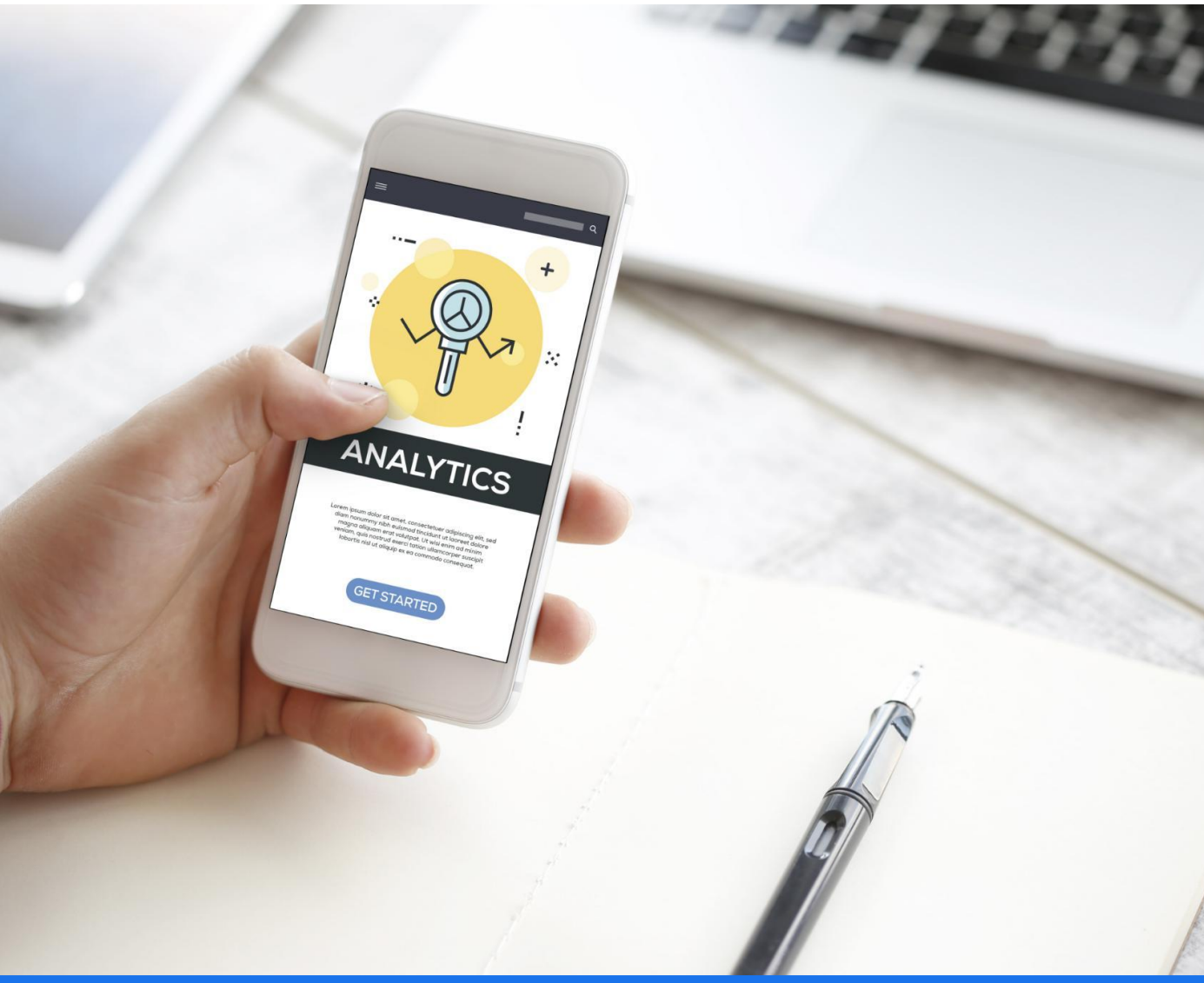# Case Study: Identifying Factors Behind Loan Defaults

EVALUATION BY POOJA KAMATH & HARIPRIYA PAMU

Agenda Items

- Problem Statement

- Analysis Approach

- Results of Univariate Analysis

- Results of Bivariate Analysis

- Key Insights and
  Recommendations

Introduction to the Problem

Background

- Lending Club operates as a peer-to-peer lending platform, facilitating loans between individual borrowers and investors. As the company grows, it faces challenges in accurately assessing the creditworthiness of applicants, which is critical for both borrower satisfaction and investor confidence.

Business Objective

- To identify variables which are strong indicators of default and potentially use the insights in approval rejection decision making.

Analysis Approach

1. **Data Cleaning**: Understand the data types and check for null values. Drop columns with too many null values or those that don't impact loan status. Remove columns with only one unique value as they are not useful.

2. **Data Imputation and Standardization**: Impute missing values with the mode for columns with low percentages of missing data. Standardize continuous values by removing symbols and converting text to numerical values. **Remove outliers** to ensure the analysis is not skewed by extreme values.

3. **Univariate Analysis**: Examining each variable individually to understand its distribution and identify patterns. Eg . Subset the data to analyze only defaulters. Identify patterns such as loan term, grade, employment length, and loan purpose that are associated with defaults. Visualizations like histograms and bar charts

4. **Bivariate Analysis**: Examining the relationship between two variables to identify correlations and interactions. Create buckets for continuous variables and analyze relationships between variables like interest rate, annual income, loan amount, and home ownership to identify strong indicators of default. Scatter plots, box plots, and other visualizations used

Approach of Univariate Analysis
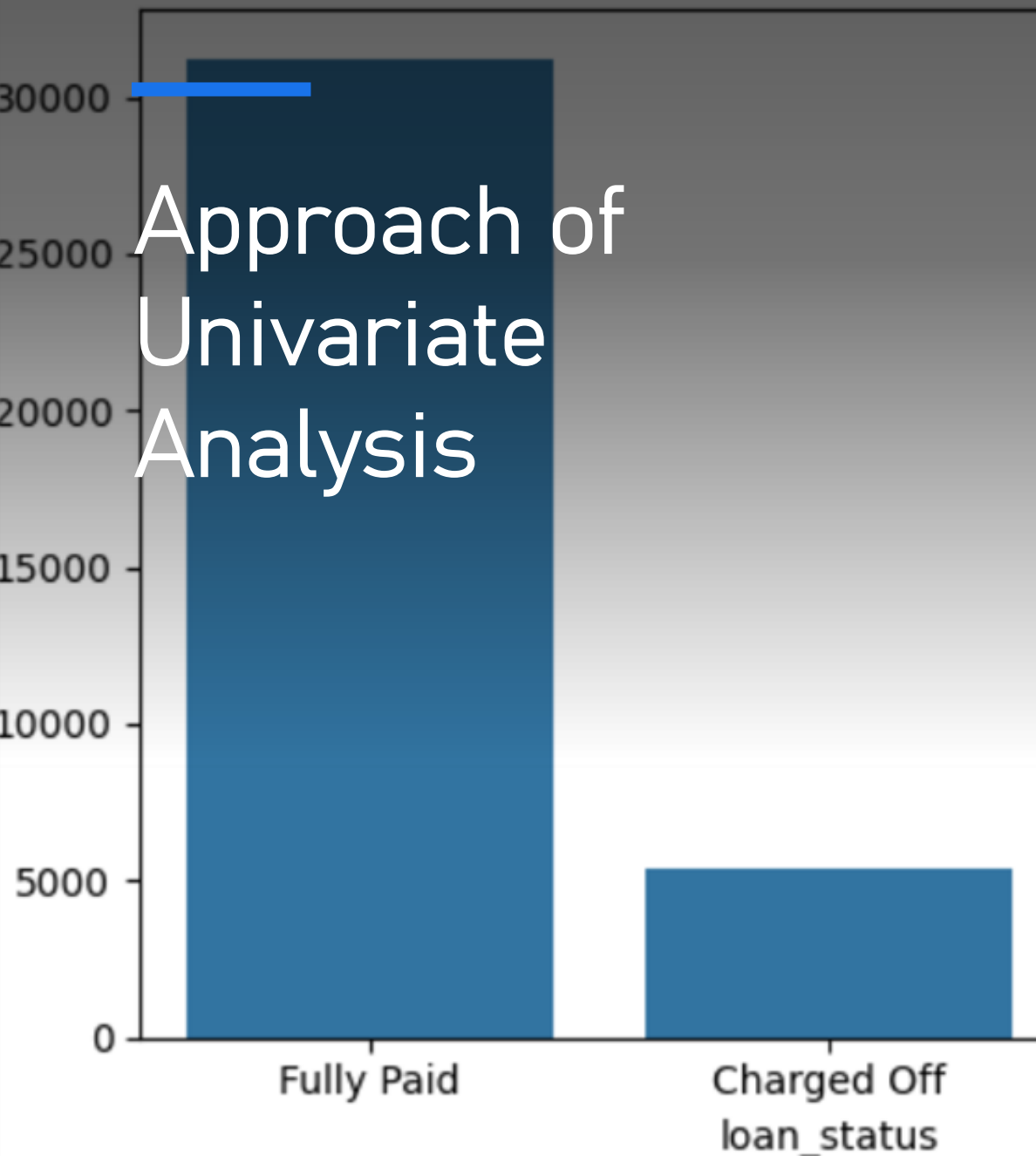
- **Subsetting the Data**: We focused on analyzing only the defaulters by removing 'Current' and 'Paid off' statuses from the loan status column. This allowed us to concentrate on the characteristics of defaulted loans.

- **Identifying Patterns**: We looked at various features such as loan term, grade, employment length, and loan purpose to identify patterns associated with defaults..

- **Handling Outliers**: We removed outliers to ensure that extreme values did not skew the analysis. This helped in maintaining the integrity of the statistical analysis.

- **Visualizing Data**: We used visualizations like histograms and bar charts to represent the distribution of variables and highlight key insights.

Visualizations for Univariate Analysis
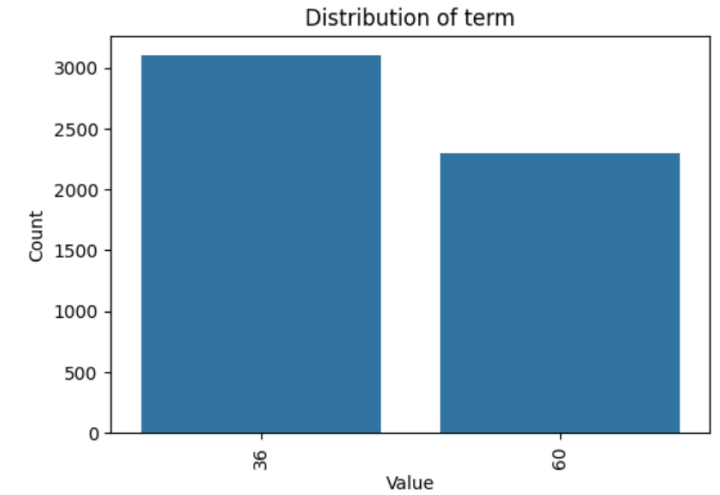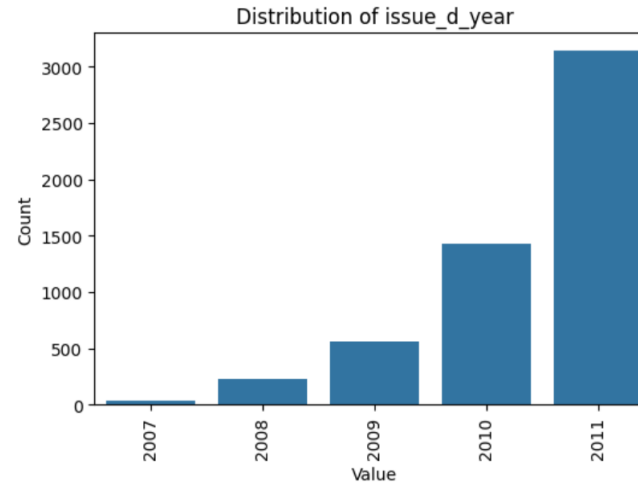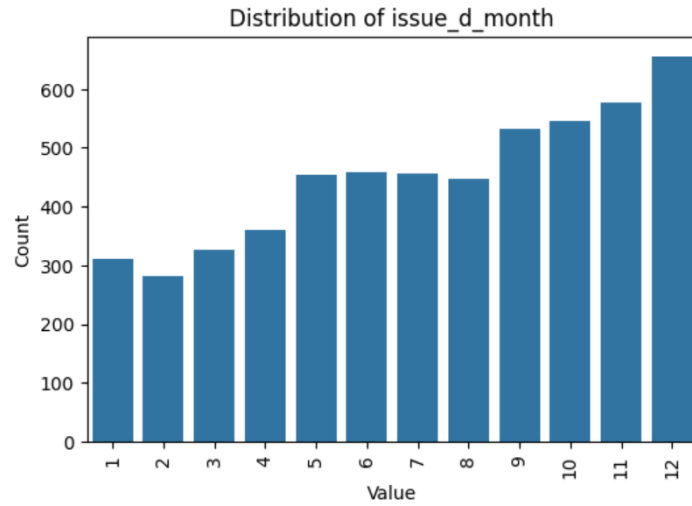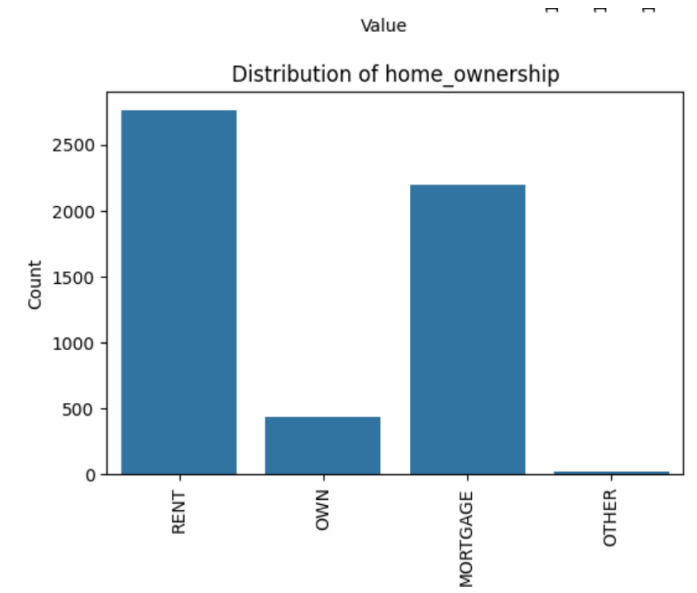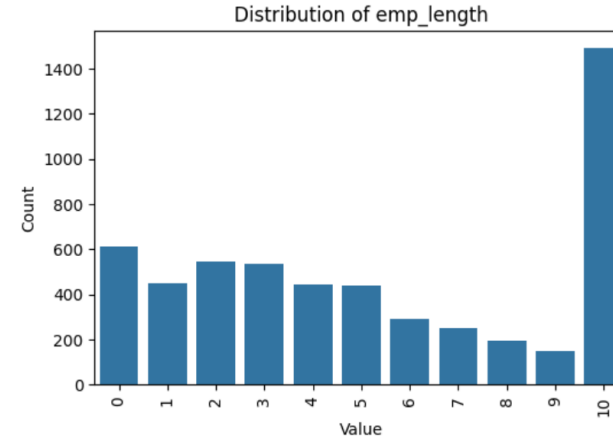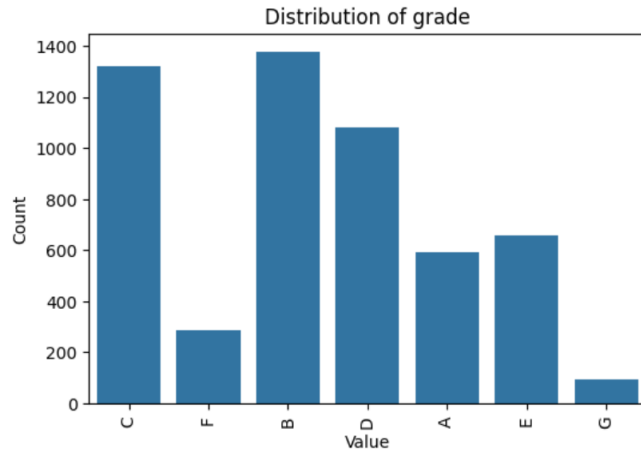
# Visualizations for Univariate Analysis
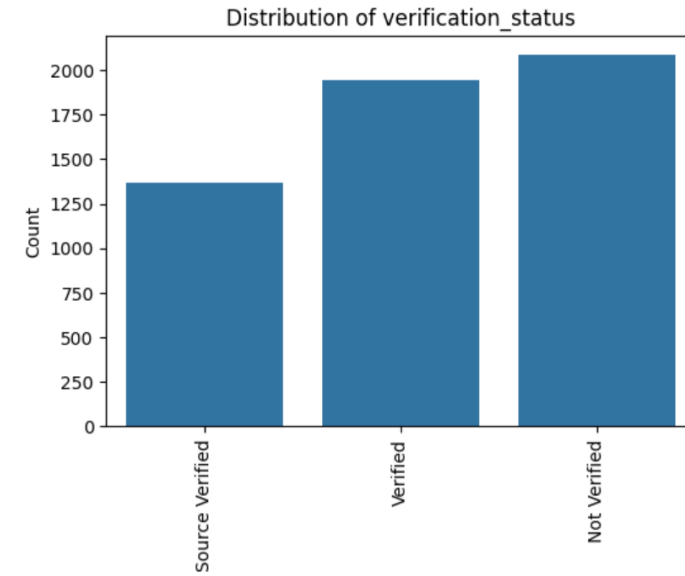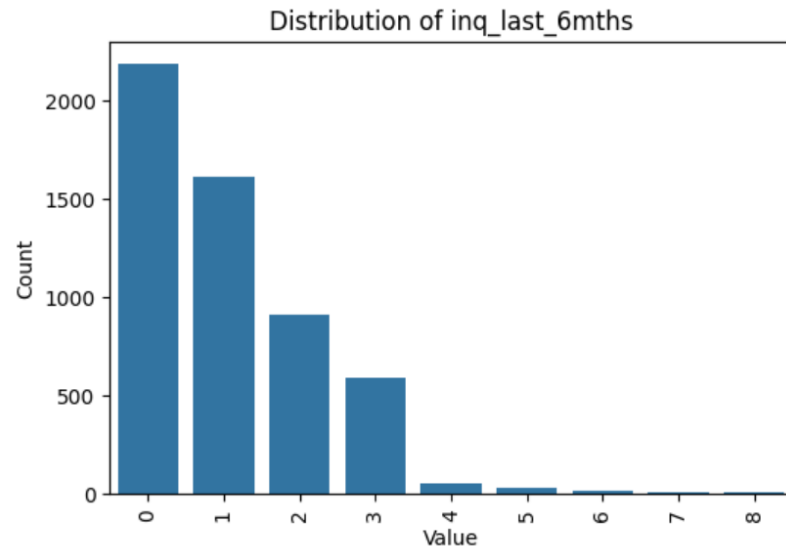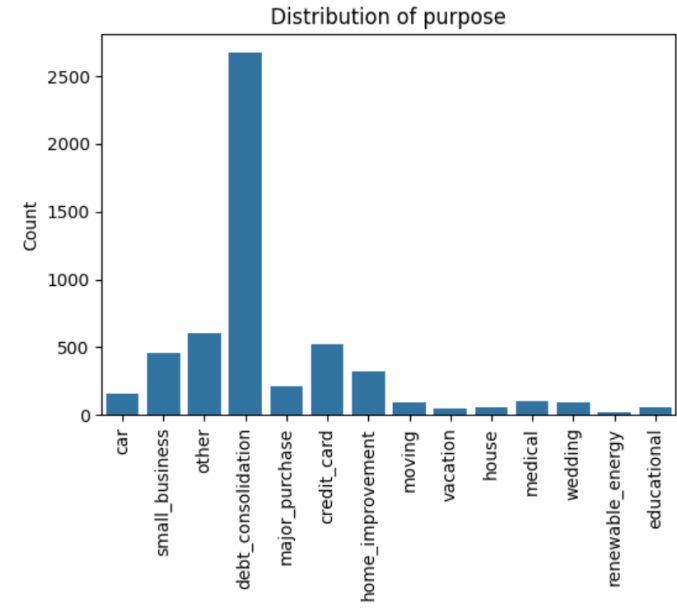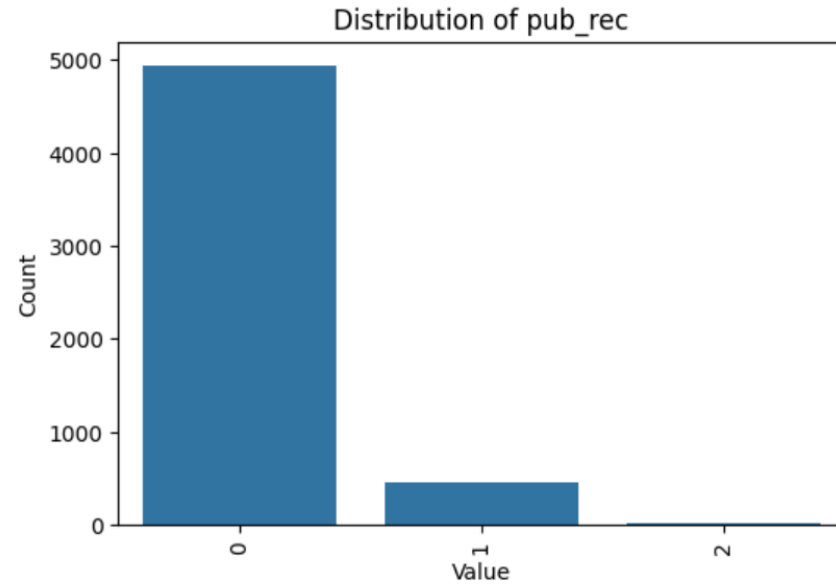
Countplot visualization shows

- Loans taken for 36 months term are most likely defaulted

- Most of the people who defaulted fall under Grade B or Grade C.

- Within Grade B, people with B5 sub grade are more prone to defaulting. Within Grade C, it is C1.

- Most of the defaulted loans are taken by people with employment length of 10.

- Most of the people who took loan in 2011 defaulted.

- People who took loan in last months of the year are more likely to default majority being in December.

- Most of the people who defaulted doesn't own the house and are living on Rent / Mortgage.

- Most People defaulted when the loan status is not verified.

- Most People who have done 0 enquiries in last 6 months defaulted.

- Debt Consolidation is the major purpose for taking the loan for the people who defaulted.
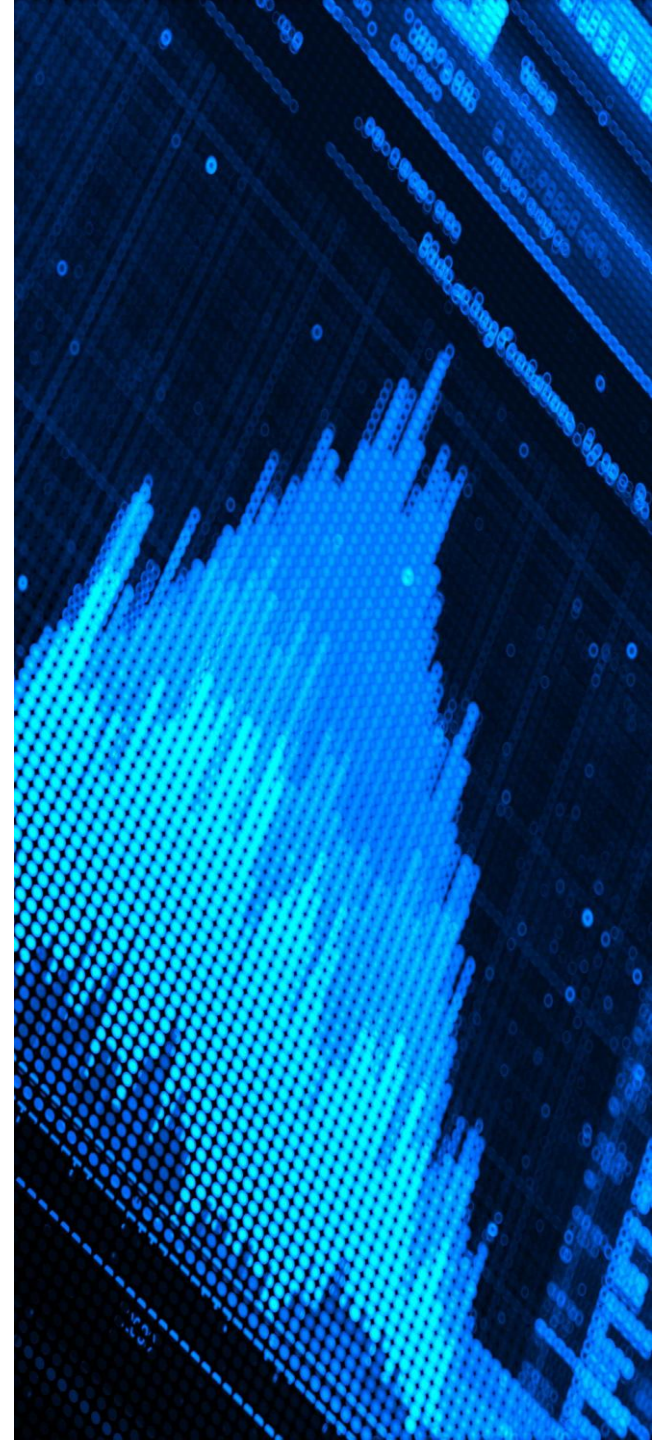
Visualizations for Univariate Analysis
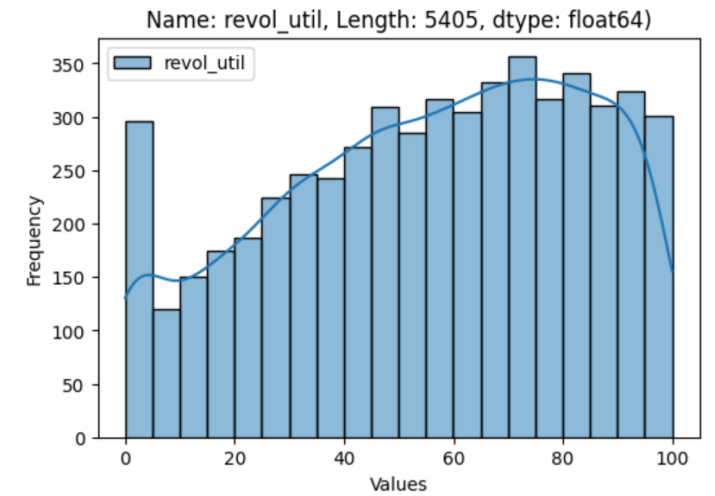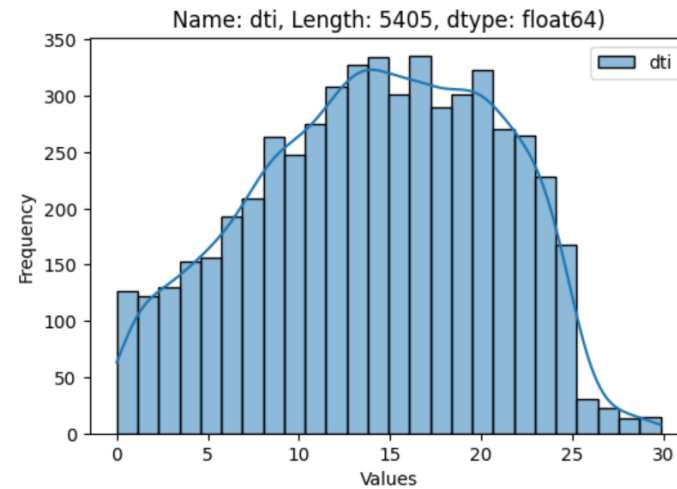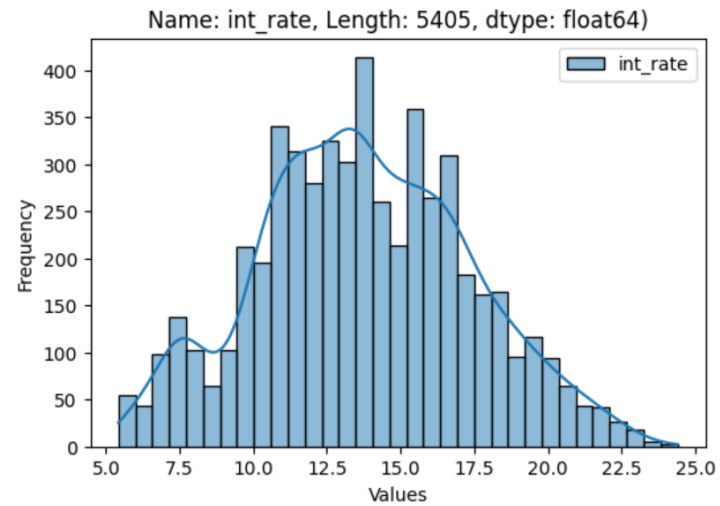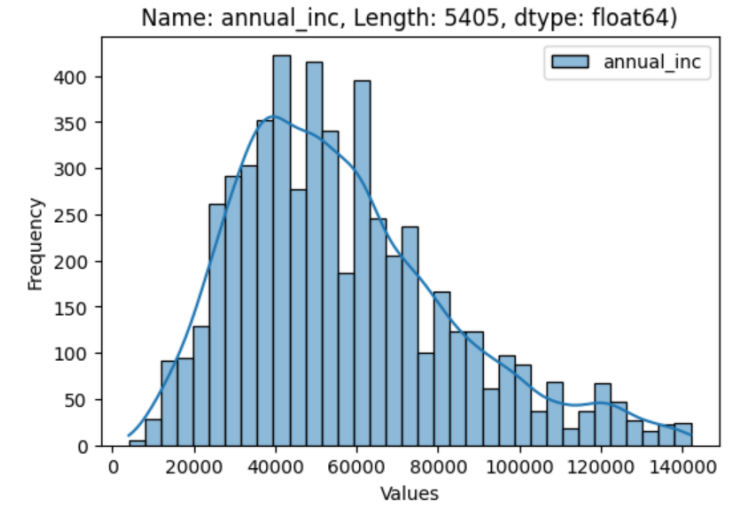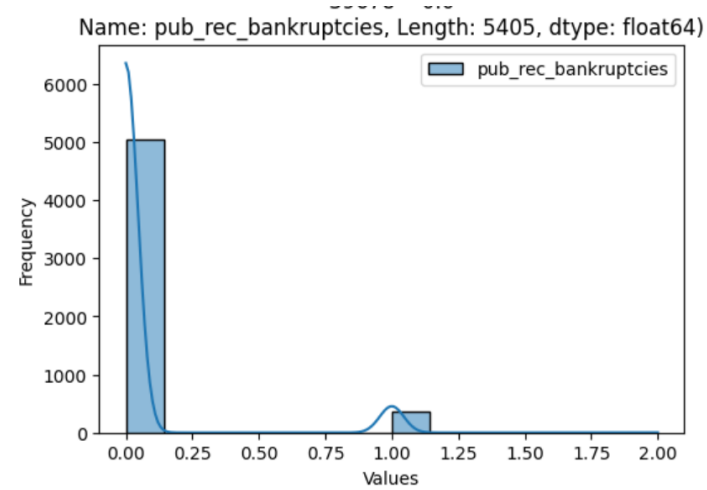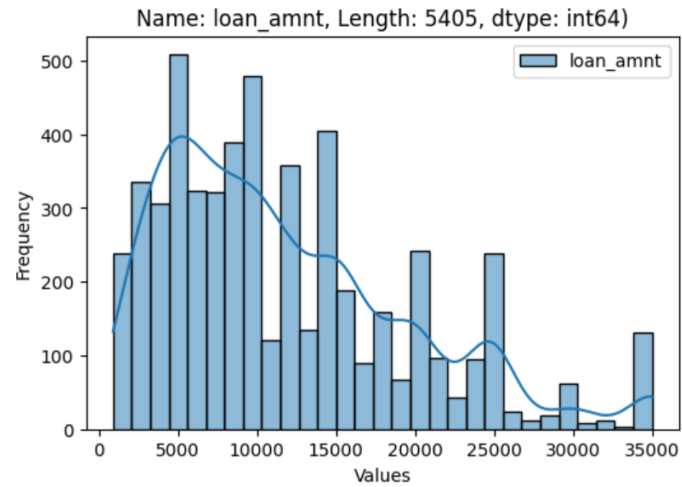
Histogram visualization shows

- Defaulted People are mostly applying for lower loan amounts (<15000).

- People who defaulted have increasing revol_util and the peak at 0.

- Loans with interest rate between 10% to 17% are majorly defaulted.

- Most of the loans defaulted are taken by people whose annual income is in middle (>20000 and <80000)

- Defaulted Loan applications increased till 20. And there is a major drop as DTI goes beyond 24.

- pub_rec_bankruptcies is not a valid parameter as it is completely zeros. So dropping it.

Approach of Bivariate Analysis

- **Creating Buckets for Continuous Variables**: We created buckets for continuous variables like interest rate and annual income to facilitate analysis. This helped in identifying trends and patterns more effectively.

- **Analyzing Relationships**: We analyzed the relationships between different variables to identify strong indicators of default. For example, we found that people with high-interest rates and higher annual incomes were more likely to default

- **Hypothesis Testing**: We formulated hypotheses based on the observed patterns and tested them using statistical methods.

- **Visualizing Interactions**: We used scatter plots, box plots, and other visualizations to represent the interactions between variables. This helped in identifying key insights and communicating the results effectively.
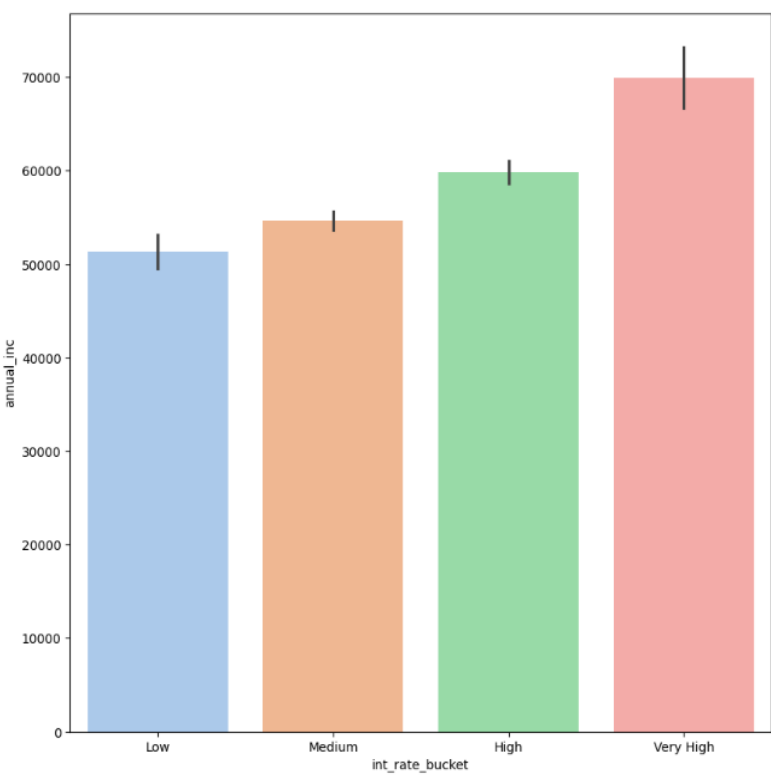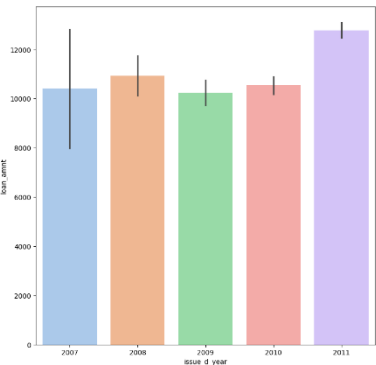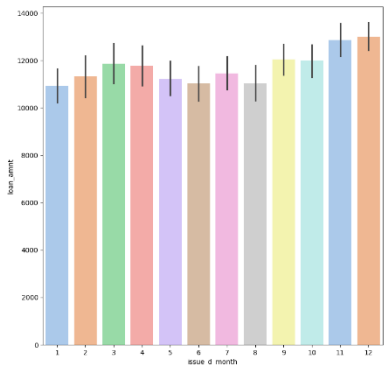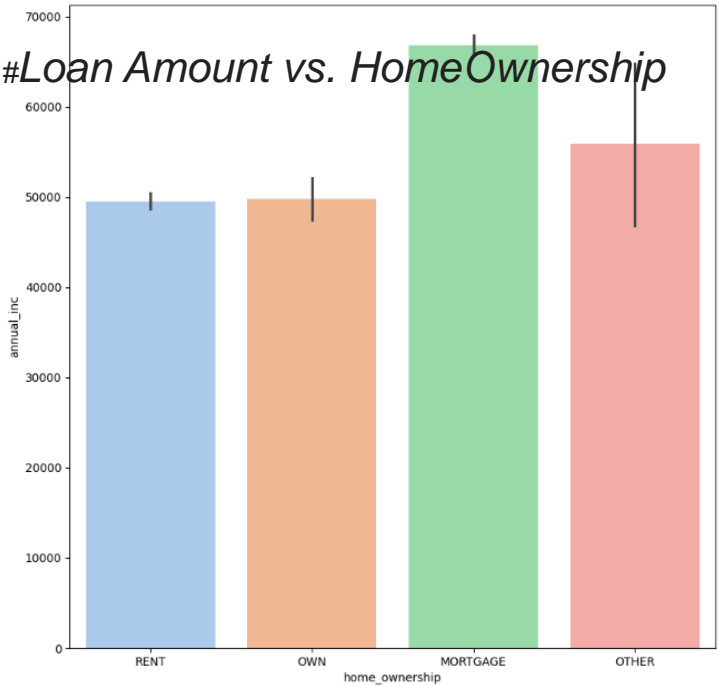
Analysing
Annual Income
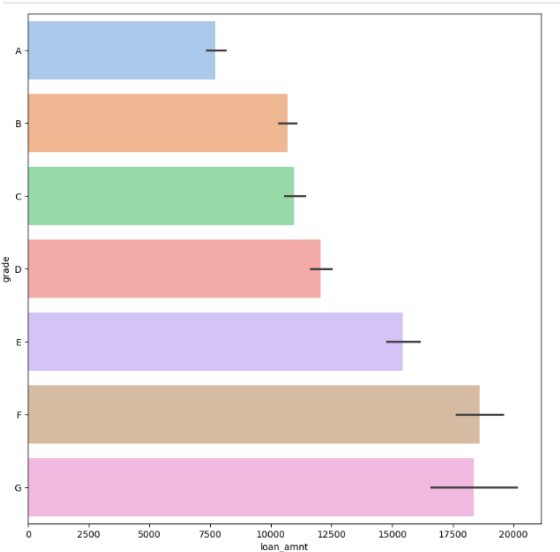with other
columns
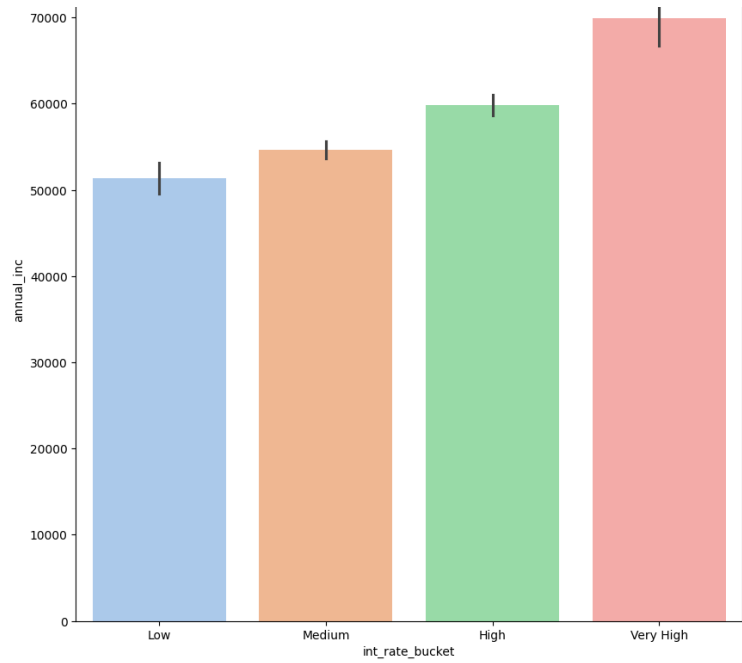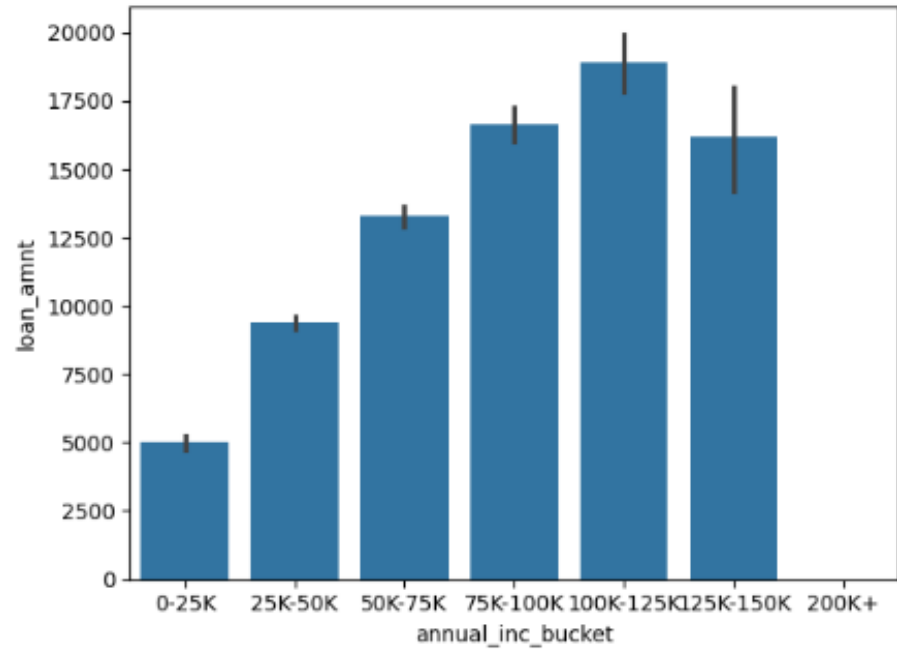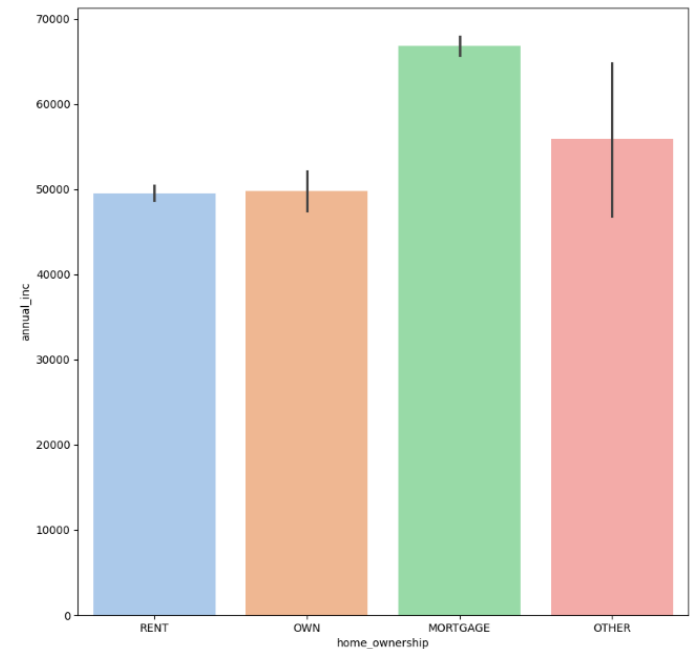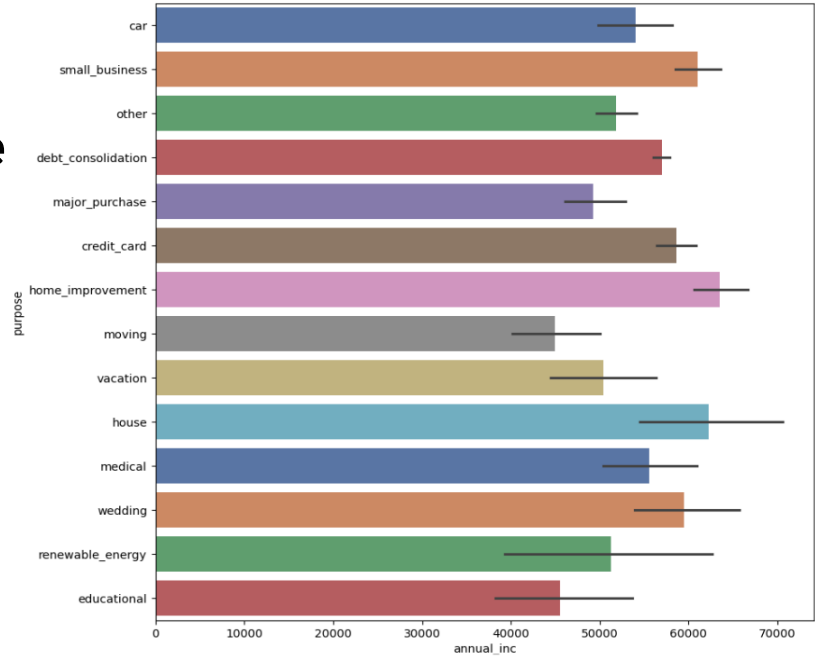
Loan Amount Vs. Purpose
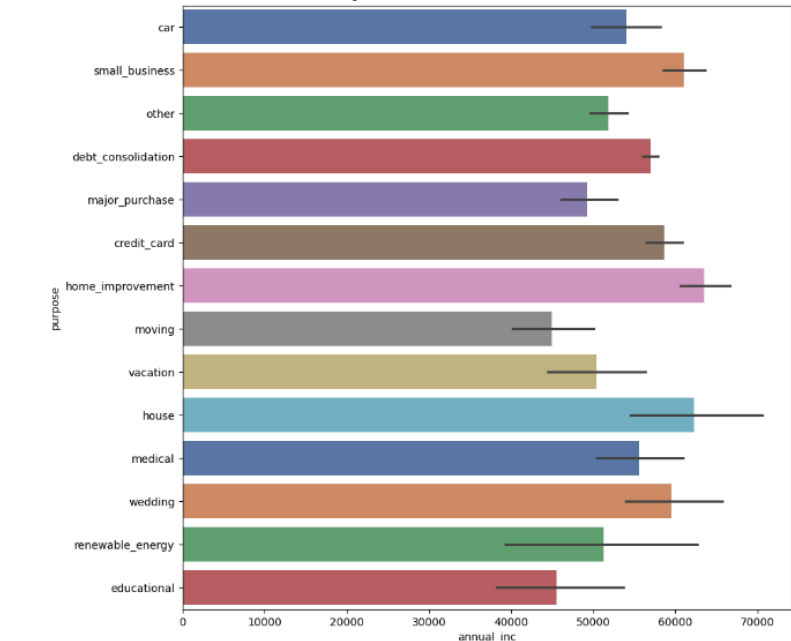
#*Loan Amount vs. Interest Rate*

#*Loan Amount vs. HomeOwnership*

#*Loan amount vs month issued and year issued*

#*Loan amount vs Grade*

Bivariate analysis shows

Below are the hypothesis for Defaulted people

- People took highest loan amount for small business

- People with highest loan amount took loans at highest interest rate

- People who took highest loan amount doesn't own house and live on mortgage

- People took highest loans in 2011 and mostly in last months of any year

- People with Grade F has taken highest loan amount

- People who are employed for more than 10 years took higher amount of loan

- People with Verified status took highest loan amount

#Loan Amt Vs Interest Rate

#Grade vs interest rate
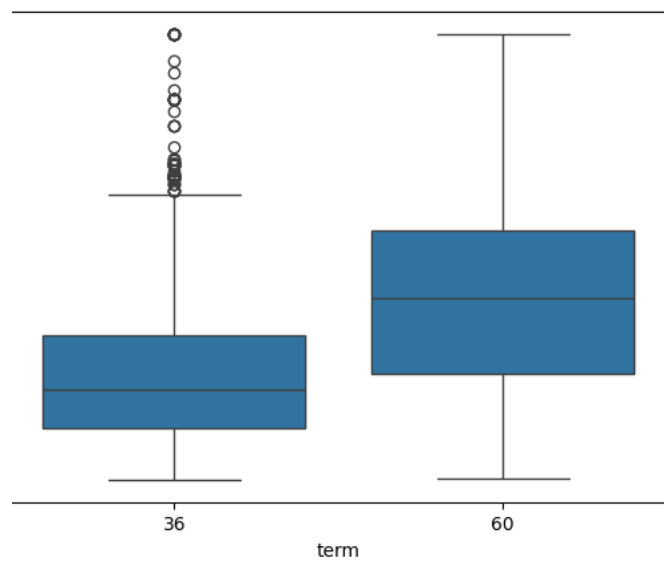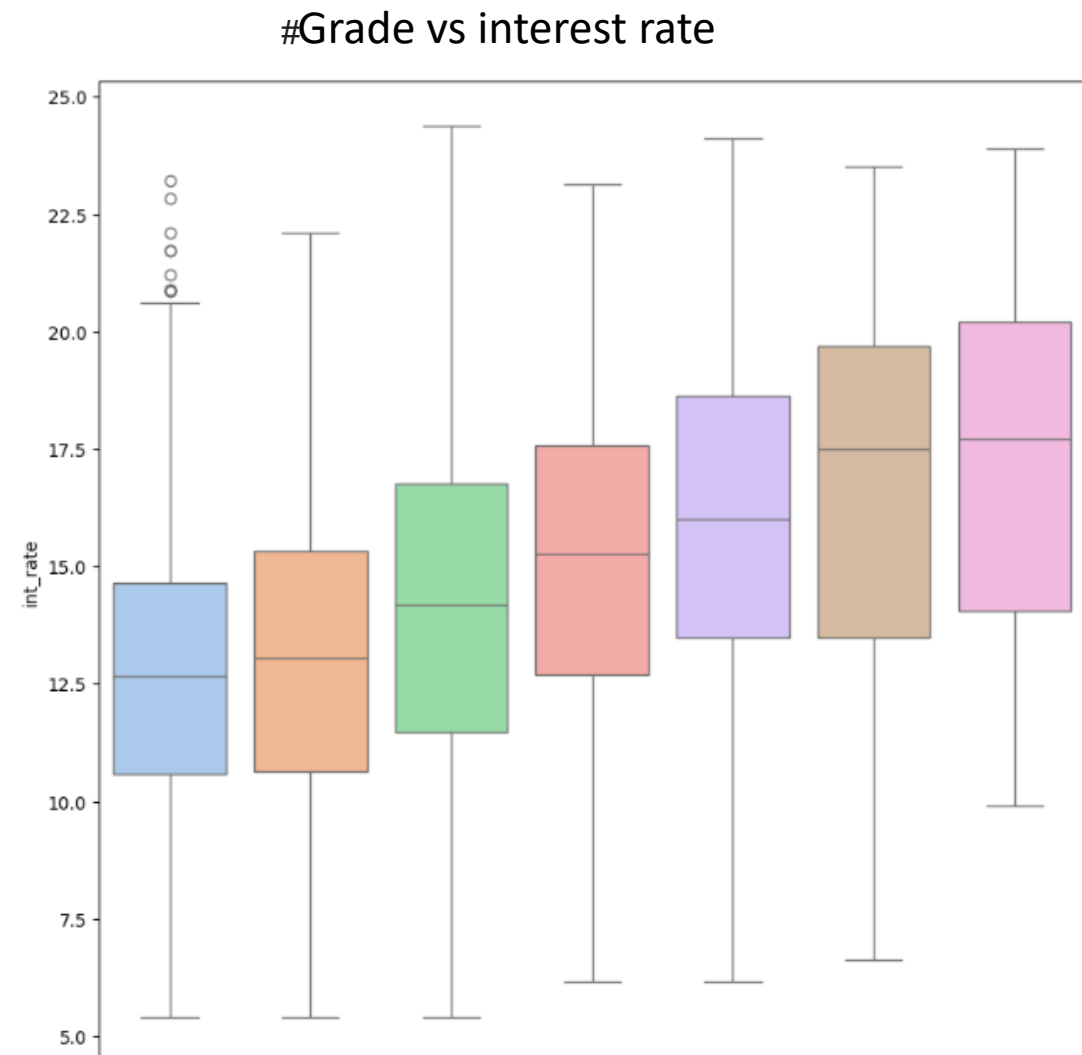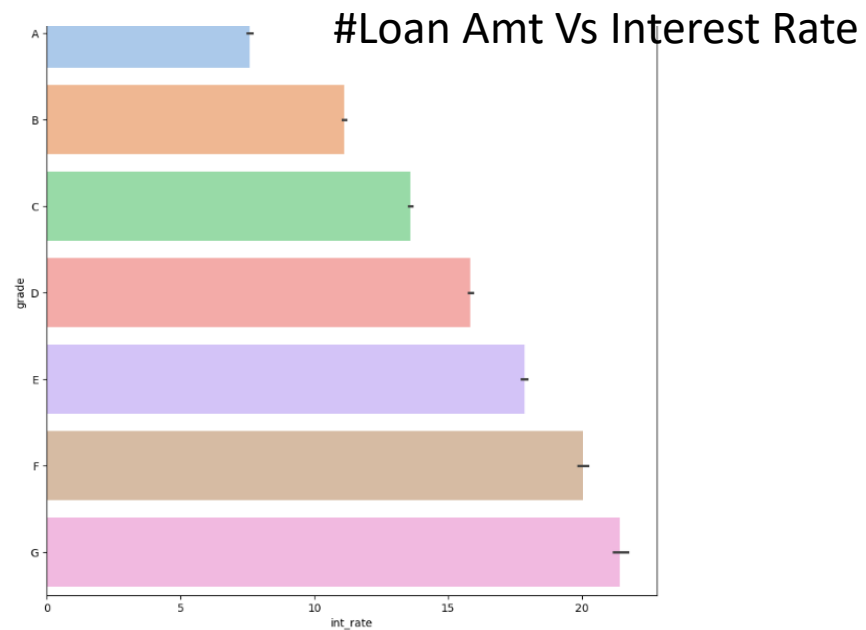
#Loan Amt for term

Probability of defaulting

Applicants taking loan for 'home improvement' or 'small business' and have income of 60k –70k

- Applicants who receive interest at the rate of 20%–25% and have an income of 60k–80k

- Applicants whose home ownership is 'MORTGAGE and have an income of 60K–70K

- Applicant with annual income of 100K–125K and have taken loan amount of 17.5K to 20K

- Applicants who have taken a loan in the range 30k – 35k and are charged interest rate of 15%–17.5%

- Applicants who have taken a loan for small business and the loan amount is between 12K to 14K

- Applicants whose home ownership is 'MORTGAGE and have loan of 14–16k

- When grade is F and loan amount is between 15k–20k

- When employment length is 10yrs and loan amount is 12k–14k

- When the loan is verified and loan amount is above 16k

- For grade G and interest rate above 20%

Key Insights and Recommendations

Based on above Observations, the factors that impact defaulting of loan are

- Annual Income

- Purpose

- Interest Rate

- Loan Amount

- Grade

- Home Ownership

Thank you
– Haripriya Pamu &
Pooja Kamath