# Learning Club Case Study

Batch: EPGP ML58

Anish V A

# Problem Statement

Consumer finance company 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.

The company aims to reduce credit loss by identifying risky loan applicants. Specifically, they want to understand the driving factors or variables that strongly indicate loan default.

The customers labeled as 'charged-off' are considered defaulters.

# Case study Objective

Conduct EDA to identify patterns which indicate if a person is likely to default, allowing the company to make informed decisions regarding risk assessment and portfolio management.

# Data Understanding

This section provides a summary of the dataset used for analysis:

| | |
|---|---|
| **Data Name** | loan.csv |
| **File Format** | .csv |
| **Number of Columns** | 111 |
| **Number of records (rows)** | 39717 |
| **Data Source** | UpGrad |

# Data Cleaning & Transformation

Data cleaning was conducted to enhance the quality and relevance of dataset for analysis

- Dropping columns with null values exceeding 40%
- Imputing null values with the median for specific columns
- Transforming data formats (e.g., removing symbols such as <, >, %) (columns - emp_length, init_rate etc.)
- Eliminating columns with uniform data (e.g., all 0, 1, n, p)
- Creating new year and month columns from existing date information (e.g. column - issue_d to issue_year, etc.)
- Generating categorical columns through bucketing for numerical features (e.g. column - dti to dti_bucket, etc. )
- Duplicated data were not identified
- Non-contributory columns were dropped

# Analysis

Analysis encompasses a exploration of the data employing univariate, bivariate, segmented, and multivariate analysis using visualization techniques to unveil patterns.

Objective of this part of EDA would be to understand the factors contributing to default of loan.

# Overview of Loan Status

Presenting an overview of the general loan status in our dataset. Subsequent slides will delve into specific patterns and factors contributing to the 'Charged Off' category.
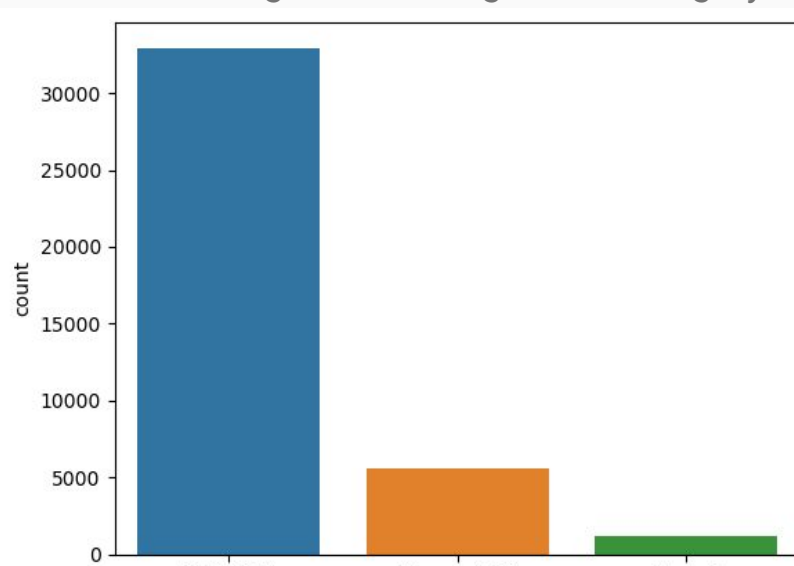
# Univariate analysis

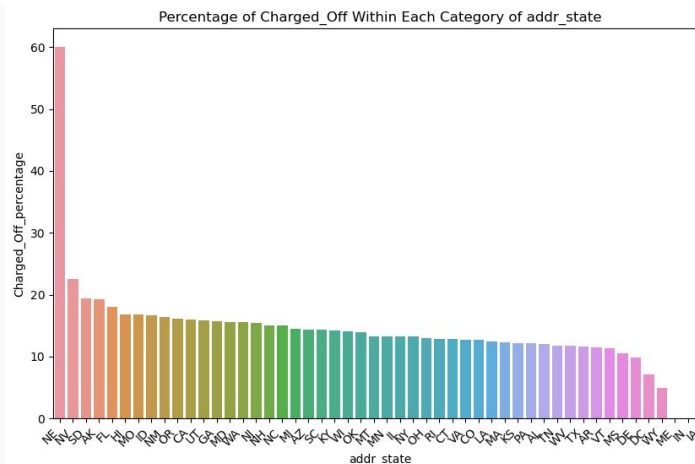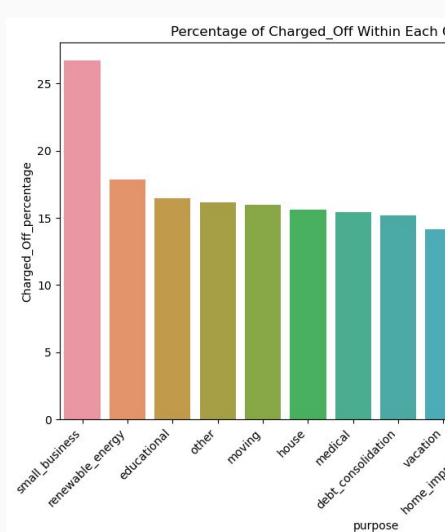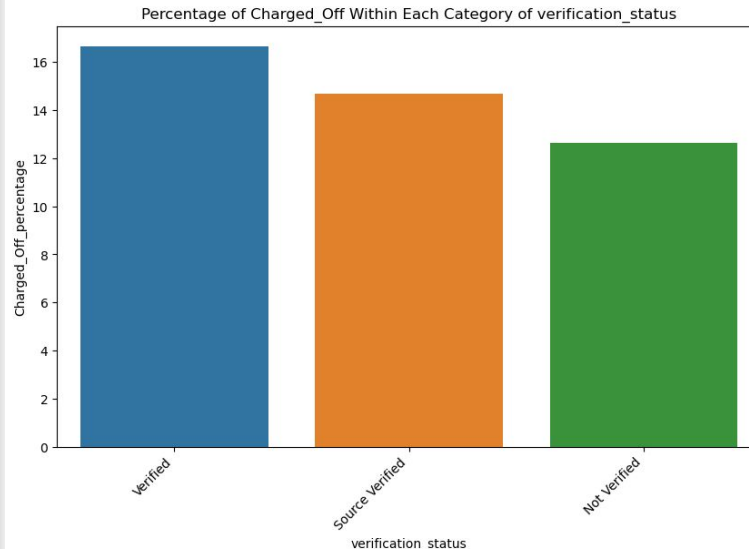Univariate analysis insights for categorical attributes.

- Loans with a term of 60 months exhibit a higher default rate.
- Grades G and F display higher default rates compared to others.
- Within G3 and F5 subcategories, default rates are notably high.
- Limited information is available for the "OTHER" category in home ownership.
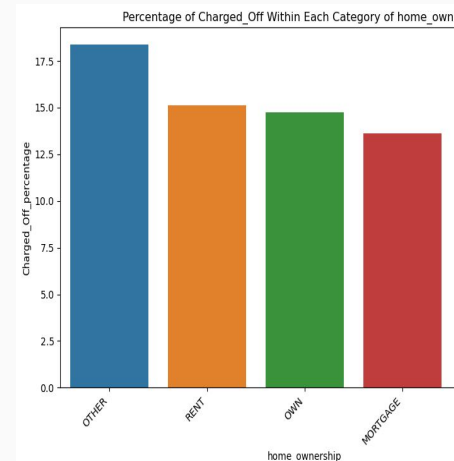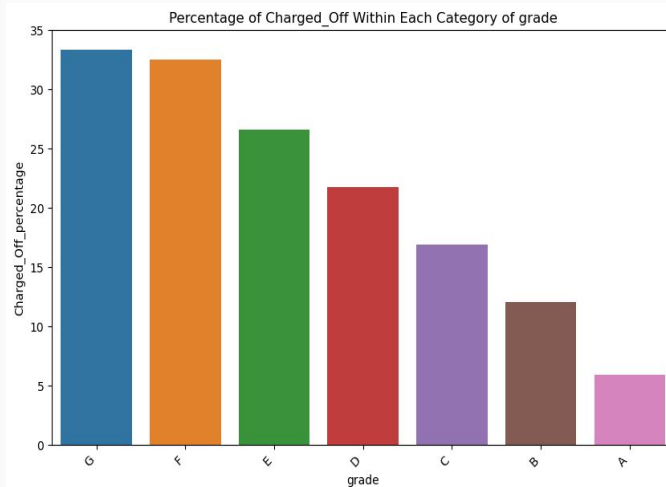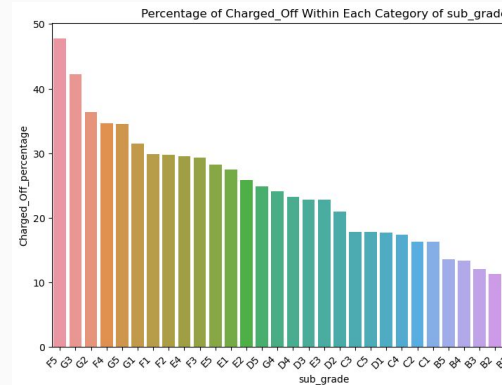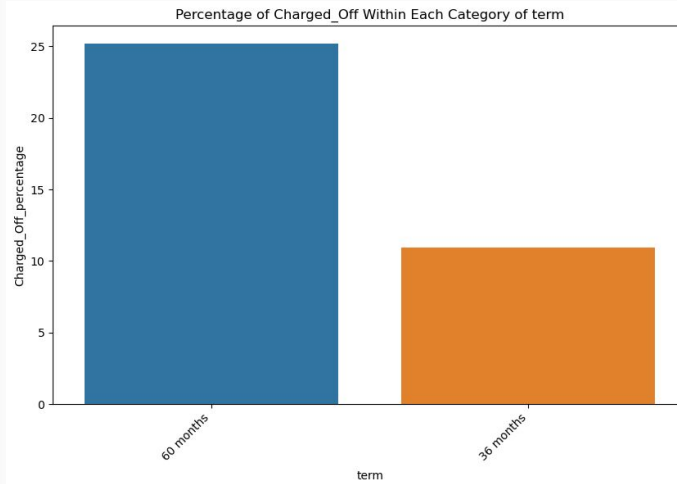
# Univariate analysis

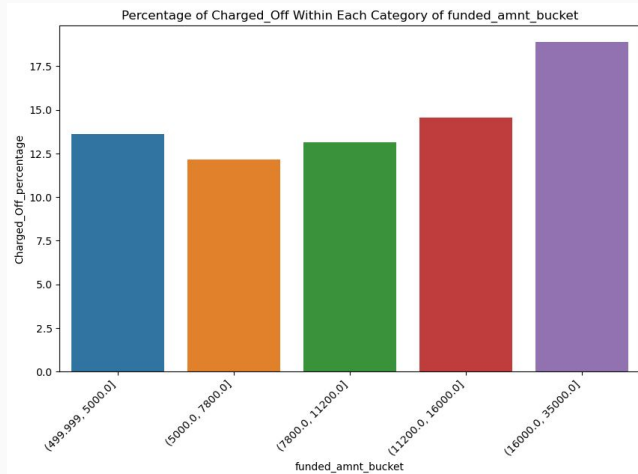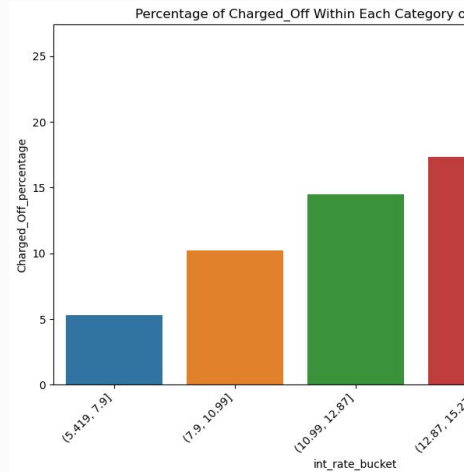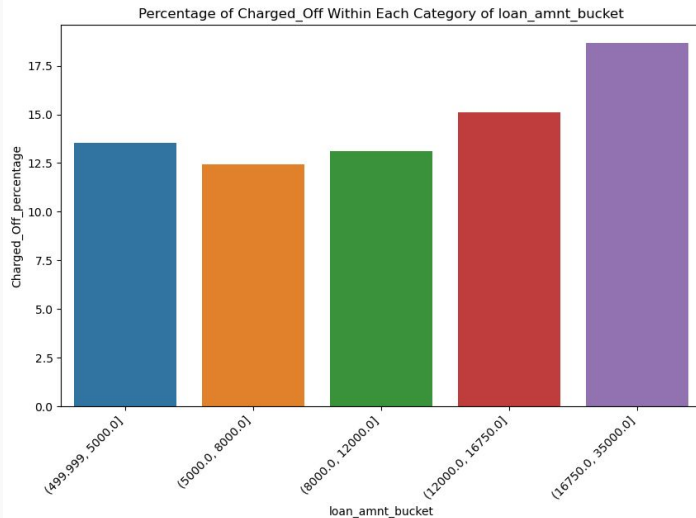Univariate analysis insights for categorical attributes.

- Verified applicants exhibit a higher default rate, suggesting potential issues in the verification process.
- Loans for small_business and renewable_energy categories pose higher risks of default rate.
- States such as NE, NV are identified as having higher default rates.
- Higher Debt to Income Ratio (DTI) depicts an increased likelihood of charged off loans.



Percentage of Charged_Off Within Each Category of verification_status



Percentage of Charged_Off Within Each Category of purpose



Percentage of Charged_Off Within Each Category of addr_state



Percentage of Charged_Off Within Each Category of dti_bucket

# Univariate analysis

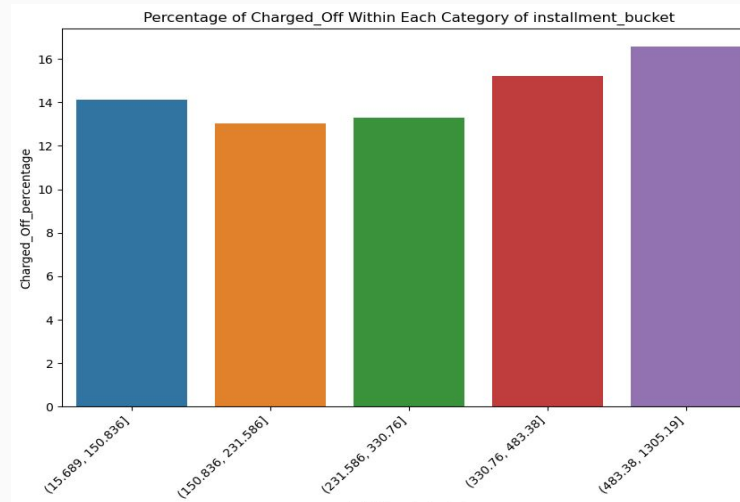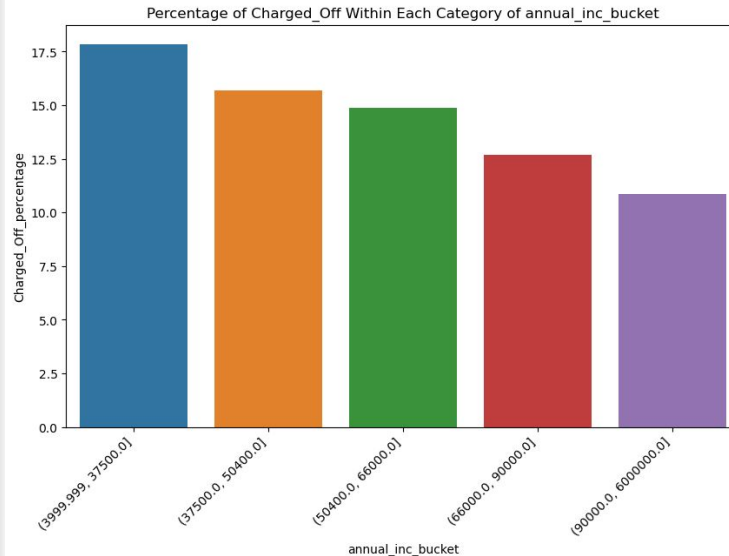Univariate analysis insights for categorical attributes.

- Higher loan amounts are associated with a greater chance of default.
- Higher interest rates contribute to a higher rate of charged off loans.
- Larger funded amounts correspond to an increased likelihood of loans being charged off.



Percentage of Charged_Off Within Each Category of loan_amnt_bucket



Percentage of Charged_Off Within Each Category of int_rate_bucket



Percentage of Charged_Off Within Each Category of funded_amnt_bucket

# Univariate analysis

Univariate analysis insights for categorical attributes.

- Higher annual income is linked to improved repayment likelihood
- Higher installment rates correspond to a higher rate of charged off loans.



Percentage of Charged_Off Within Each Category of annual_inc_bucket



Percentage of Charged_Off Within Each Category of installment_bucket
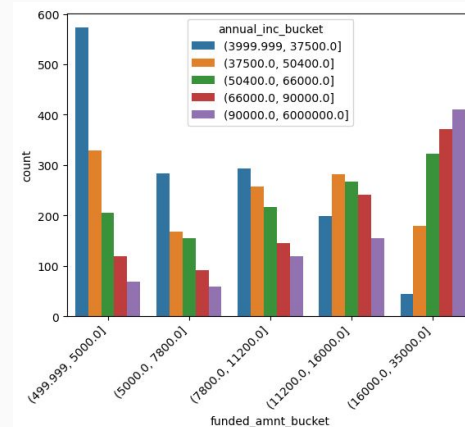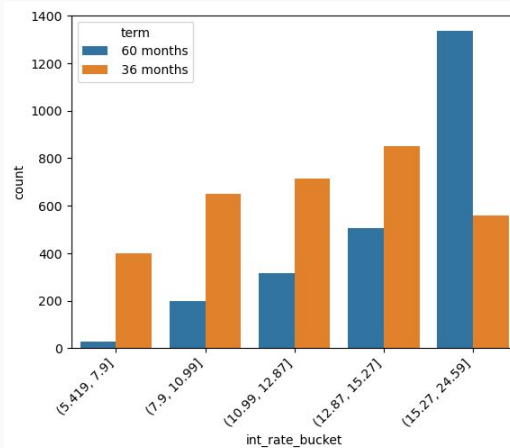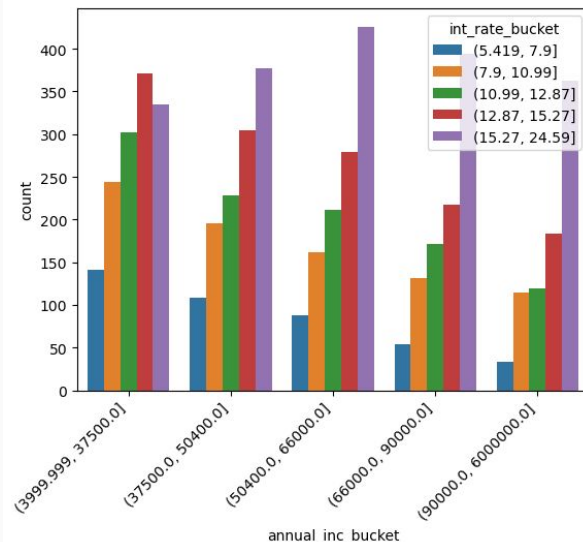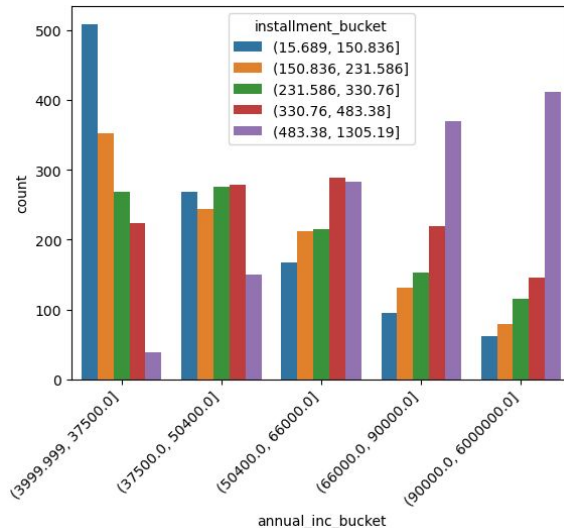
# Univariate analysis

Univariate analysis insights for numerical attributes.

- Predominantly, defaulters tend to have an employment length of 10+ years.
- Concentration of defaulted loans was observed in those sanctioned in the year 2011.
- Public records were not available for a considerable portion of applicants who defaulted on loans.
- The credit line year for a significant number of defaulted applicants centered around the year 2000
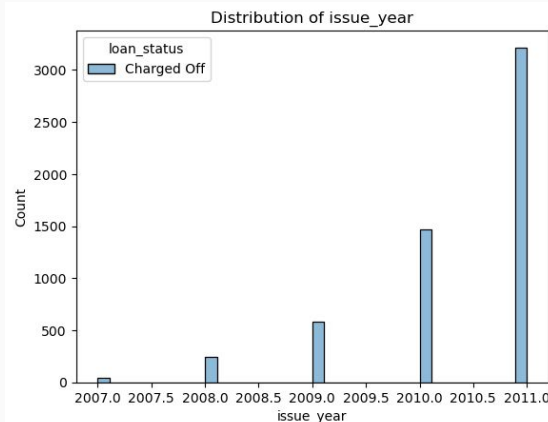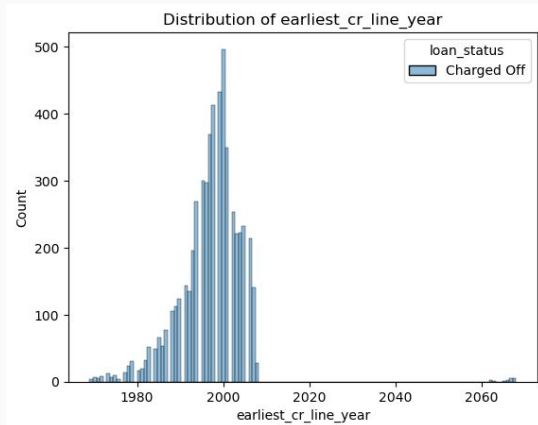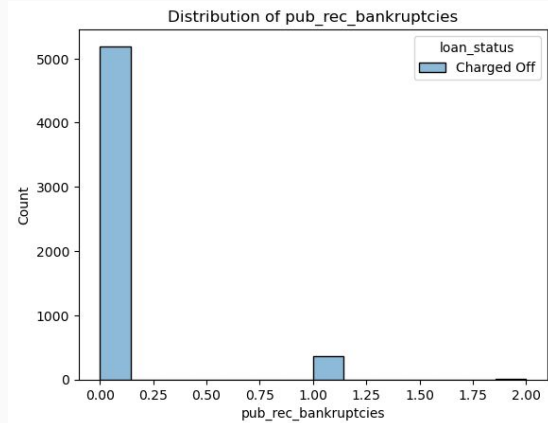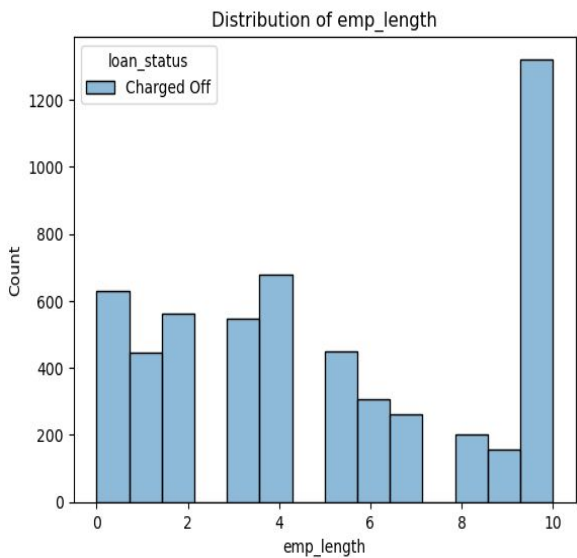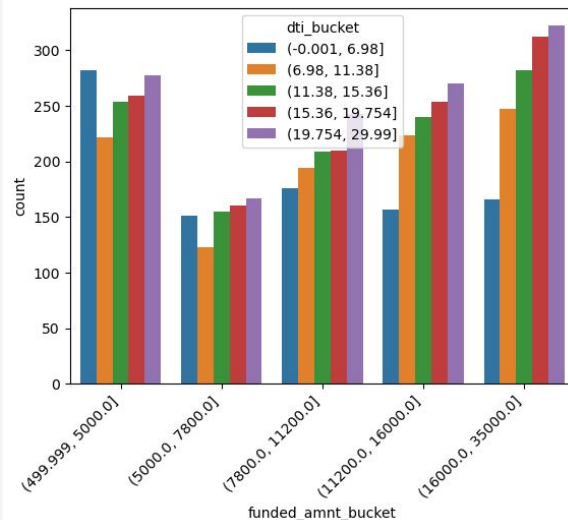
# Bivariate analysis

- Higher installment along with annual income contribute to a higher rate of charged off loans.
- Higher interest rate along with higher annual income contribute to a higher rate of charged off loans.
- Higher annual income along with higher funded amount contribute to a higher rate of charged off loans.
- Higher interest rate along with 60 months term contribute to a higher rate of charged off loans.

# Bivariate analysis

- Higher dti along with higher funded amount contribute to a higher rate of charged off loans.

# Multivariate analysis

- The loan amount, funded amount, and funded amount by investors have a strong correlation with each other.
- Installment shows a strong correlation with the loan amount, funded amount, and funded amount by investors.



Titanic dataset - correlation map
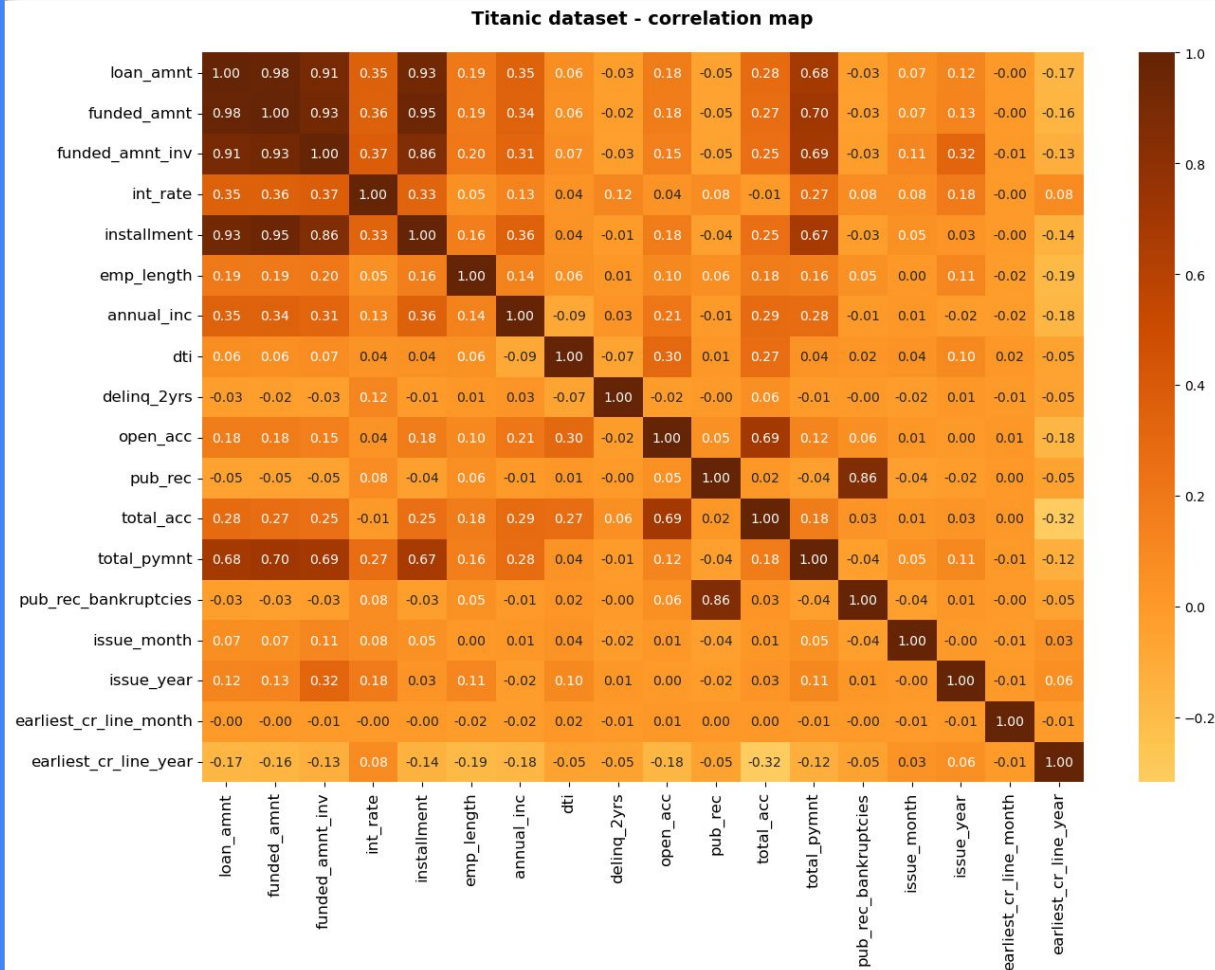
# Summary

Based on the analysis, following recommendations aim to enhance risk management to reduce default rates:

- Loans with a term of 60 months exhibit a higher default rate.
- Grades G and F display higher default rates compared to others.
- Within G3 and F5 subcategories, default rates are notably high.
- Loans for small_business and renewable_energy categories pose higher risks of default rate.
- States such as NE, NV are identified as having higher default rates.
- Higher Debt to Income Ratio (DTI) depicts an increased likelihood of charged off loans.
- Loans for small_business and renewable_energy categories pose higher risks of default rate.
- Higher loan amounts,interest rates,funded amounts correspond to an increased likelihood of loans being charged off.

# Thanks!

Email: anishnairva@gmail.com