Lending Club Case Study

Presented By :
Akshay Handoo
&
Karan Afuwale

Problem statement

- Providing loans to customers is a primary objective for banks. It is essential to lend to qualified customers to ensure profitability

- We need to examine loan histories to establish parameters that assess customer reliability in repaying loans.

- We aim to discern general loan patterns: who borrows, when they borrow, where they borrow from, and the reasons behind their borrowing decisions to identify the potentially risky customer

Objective

- Identify the key attributes that significantly predict loan default.

- Gain insights into the traits of high-risk loan applicants.

- Examine customer and loan attributes to reveal patterns and trends that signal a higher risk of default.

Analysis…

We have followed the following steps.

- **Data Understanding and Cleaning** :
  - Review Data Dictionary: Examine column names and their descriptions to comprehend the data structure.
  - Data Quality Assessment: Handling Missing Values, Outlier Detection.
  - Duplicate Records and Categorical Data Consistency
  - Column Selection: Identify and select the most relevant columns for a focused analysis.

- **Univariate Analysis**:
  - Numerical and Categorical variable analysis

- **Segmented Analysis**:
  - Explore the variables directly influence loan default rates.

- **Bivariate and Multivariate Analysis**:
  - Exploring Investigate interactions among multiple variables and their impact on loan status.

Data cleaning and manipulation

- Understanding data, data types, missing values and outliers
- Dropping the columns with more than 50% of missing values and with single values
- Remove all the rows with unwanted loan status (Current as it doesn't signify any roles here)
- Remove %, hyphens (-), converting variables to date time , int and float
- Remove outliers
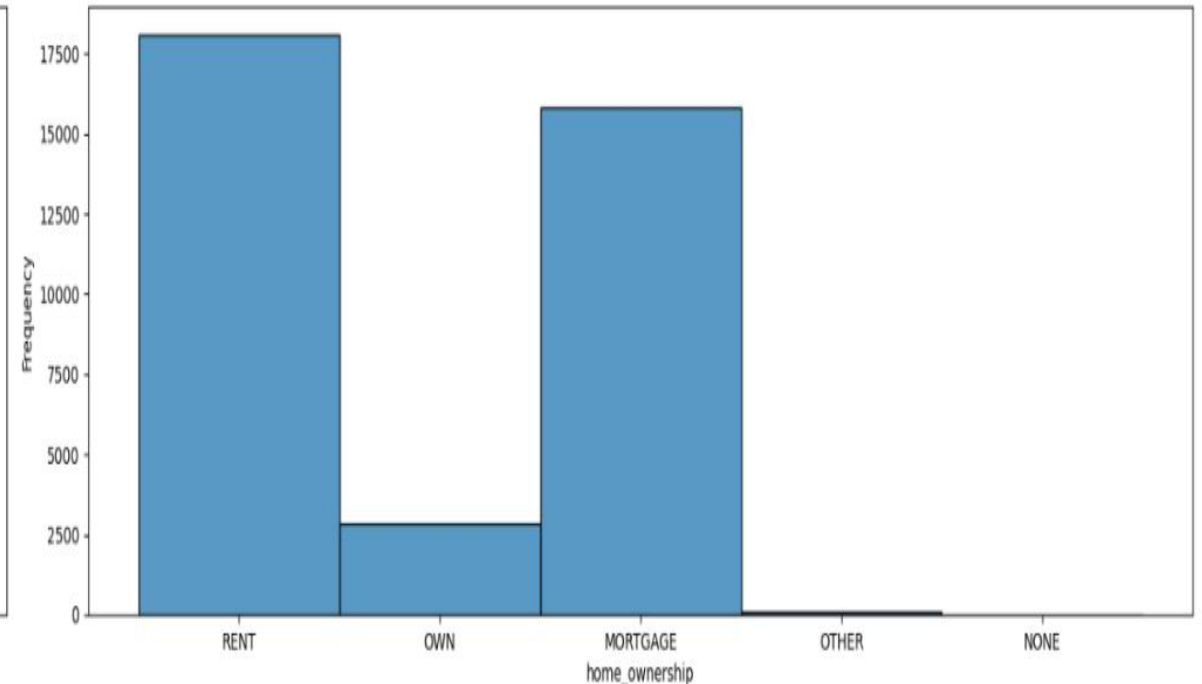- Converting some numerical variables to ordered categorical (Binning)

Final list of variables :

loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate
installment	grade	emp_length	home_ownership	annual_inc
verification_status	issue_d	loan_status	purpose	issue_year
addr_state	dti	open_acc	pub_rec	revol_bal
revol_util	total_acc	issue_month		

Univariate analysis…
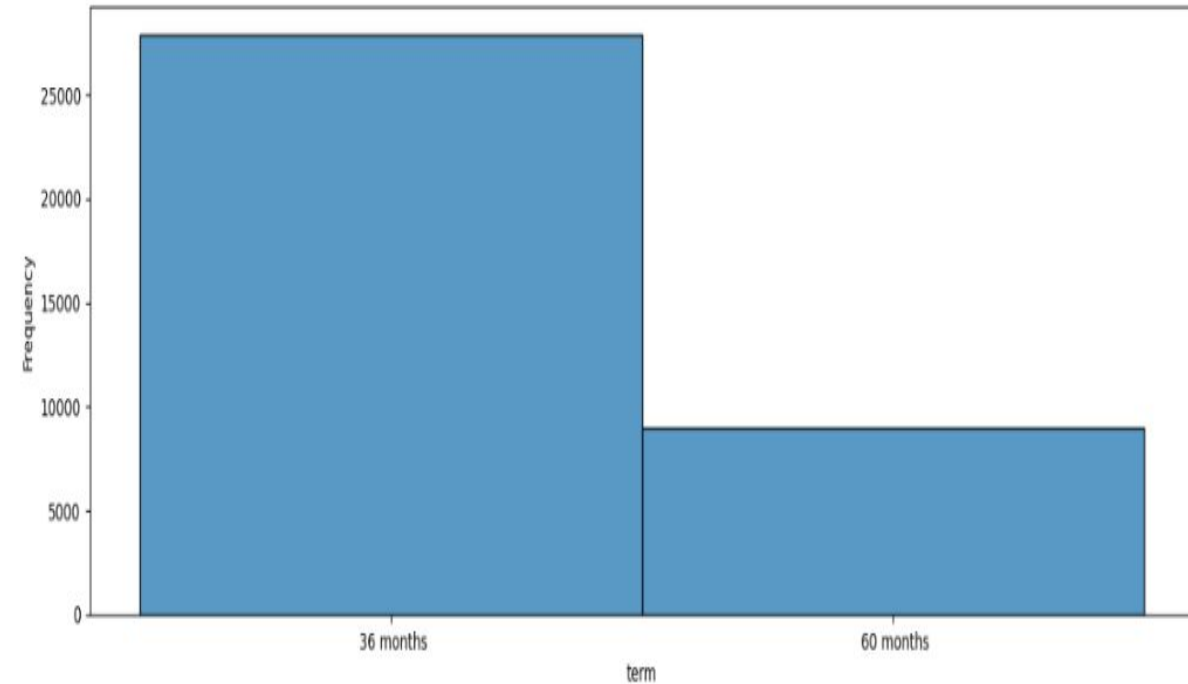
term

home_ownership
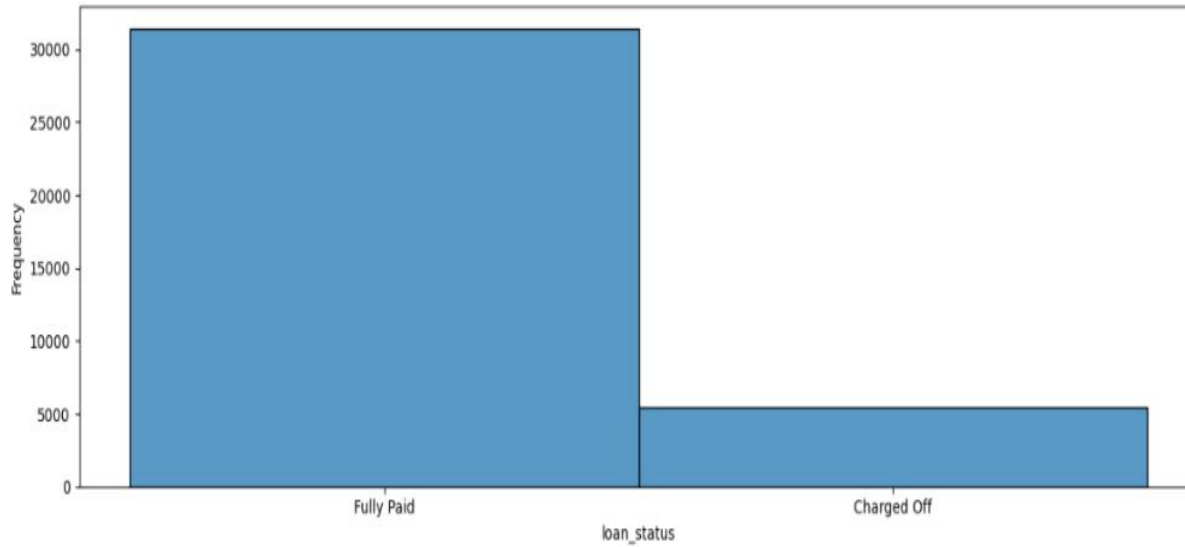
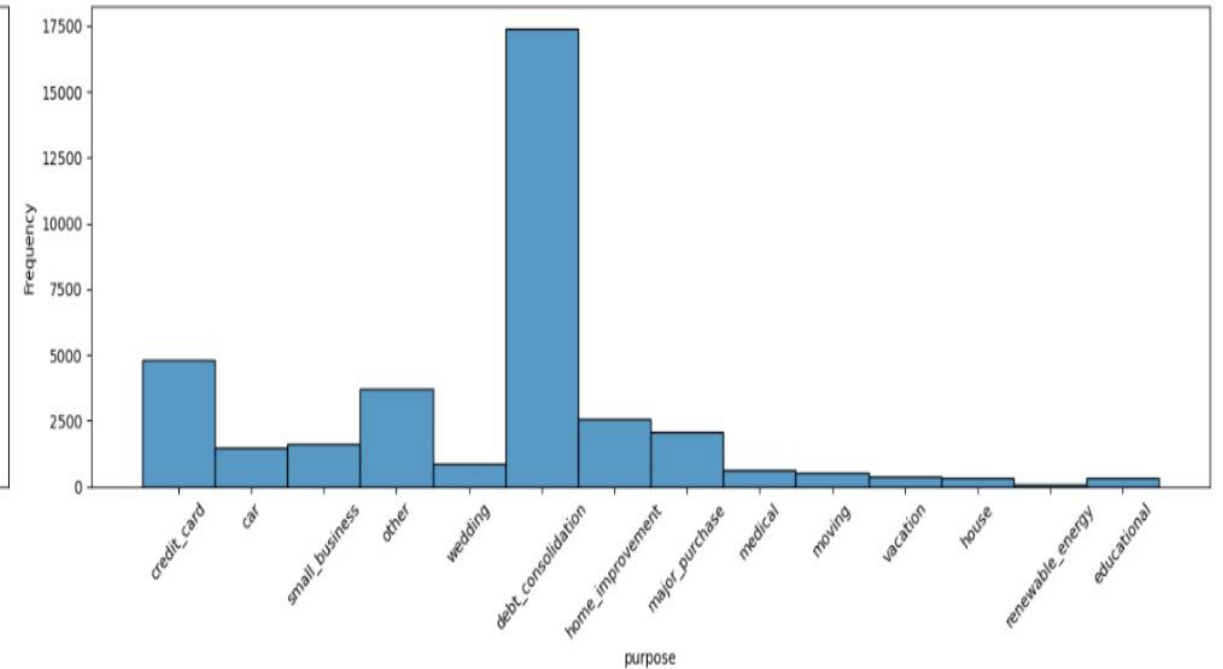Analyzing the "term" variable we can see more people are likely to take the loan for 36 months

Analyzing the "home_ownership" variable, we can see more people with rented house and mortgage prefer to take loan
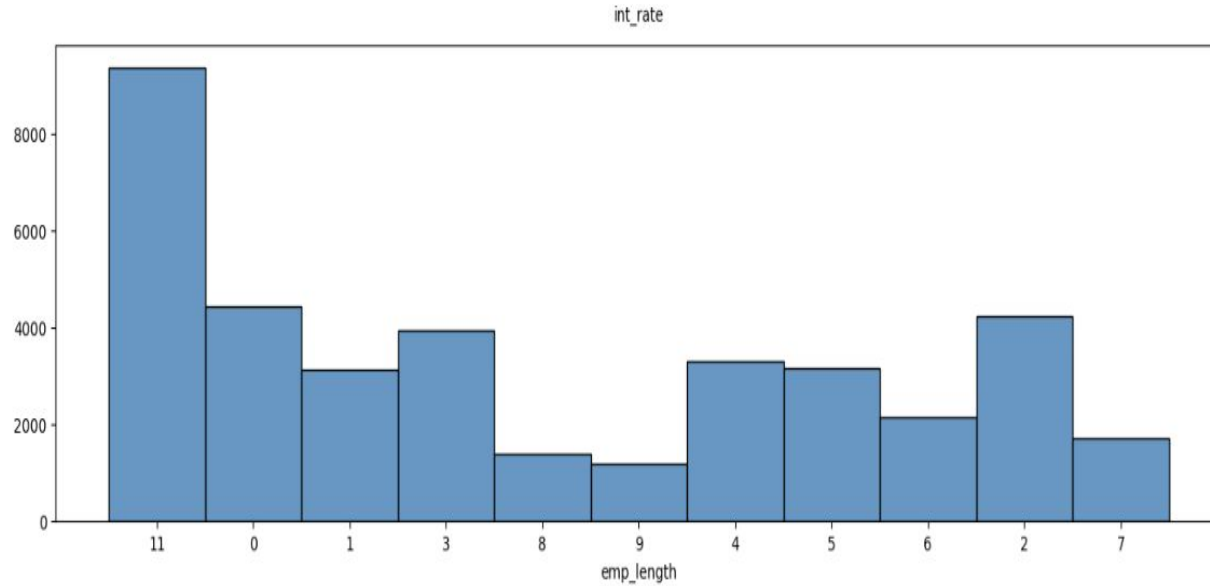
Univariate analysis…

loan_status

purpose

Analyzing the "loan_status" variable we can see higher number of people are paying the loans compare to default and it's a positive indicator for our analysis
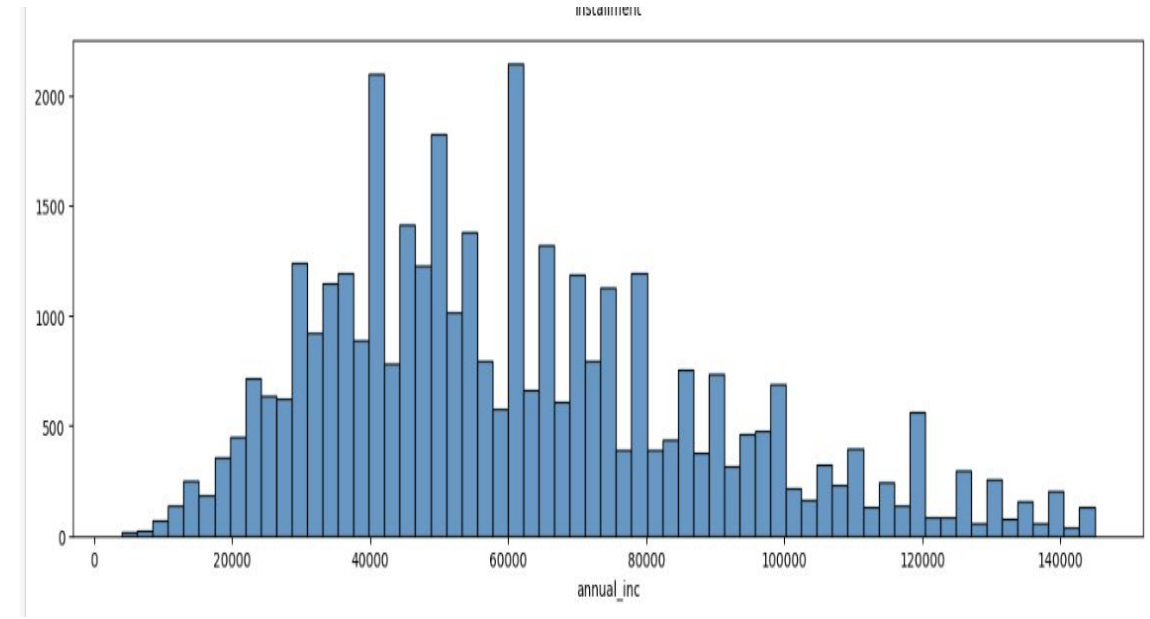
Analyzing the "purpose" variable, we can see more people are taking loans for the debt consolidation.

Univariate analysis…

emp_length

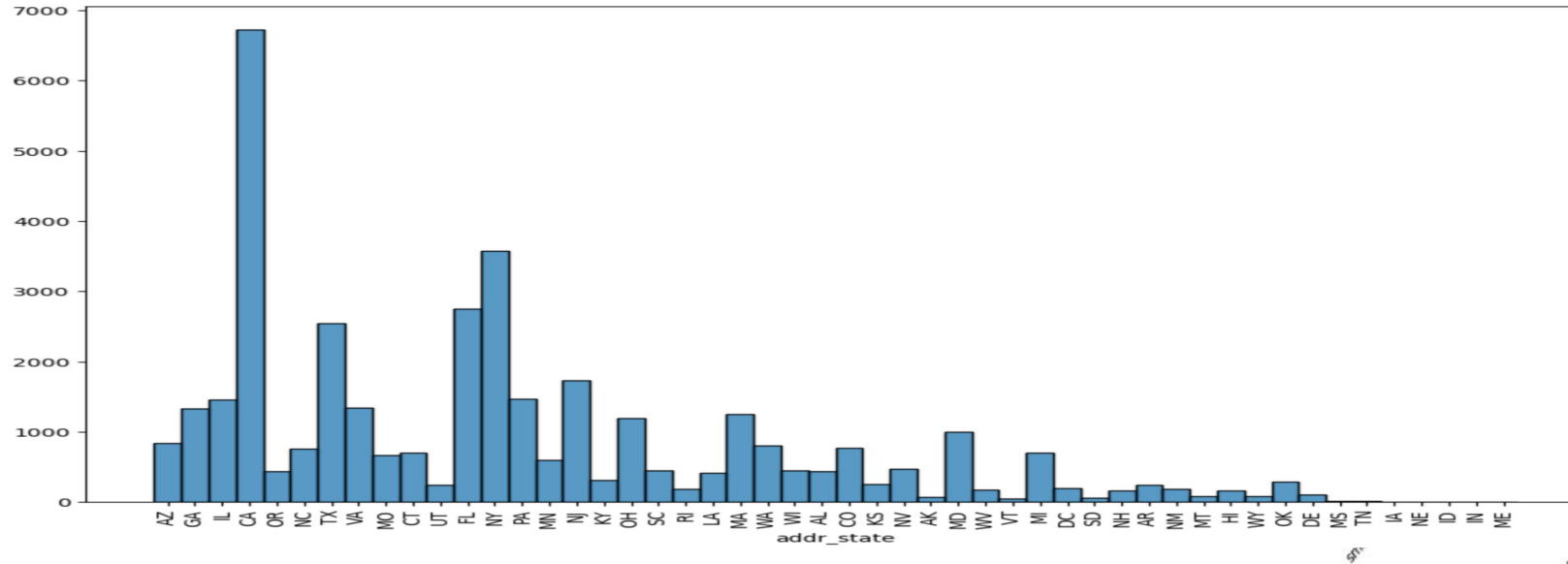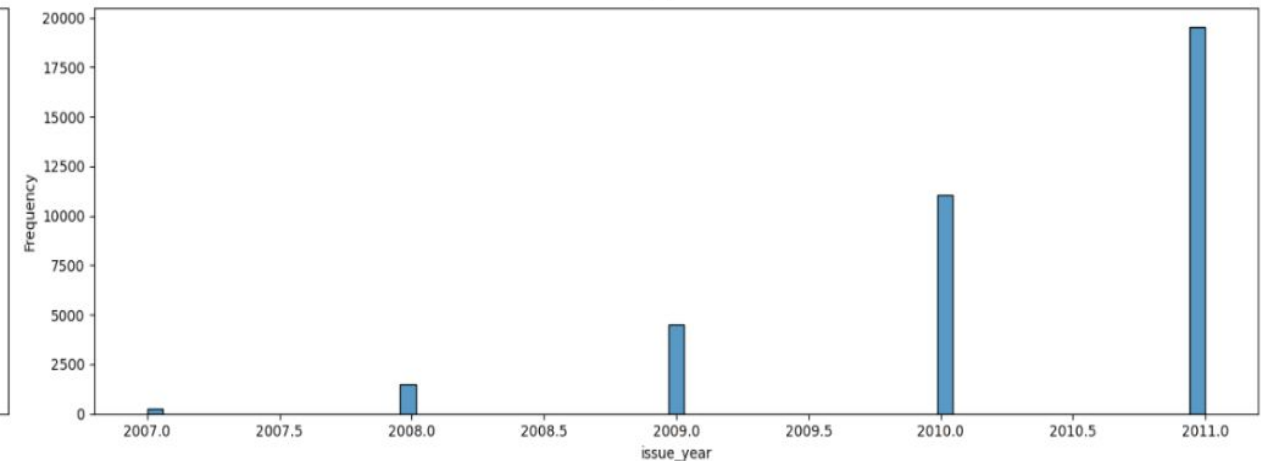Analyzing the "emp_length" variable we can see higher number loan taken by people above above 10+ years of employment
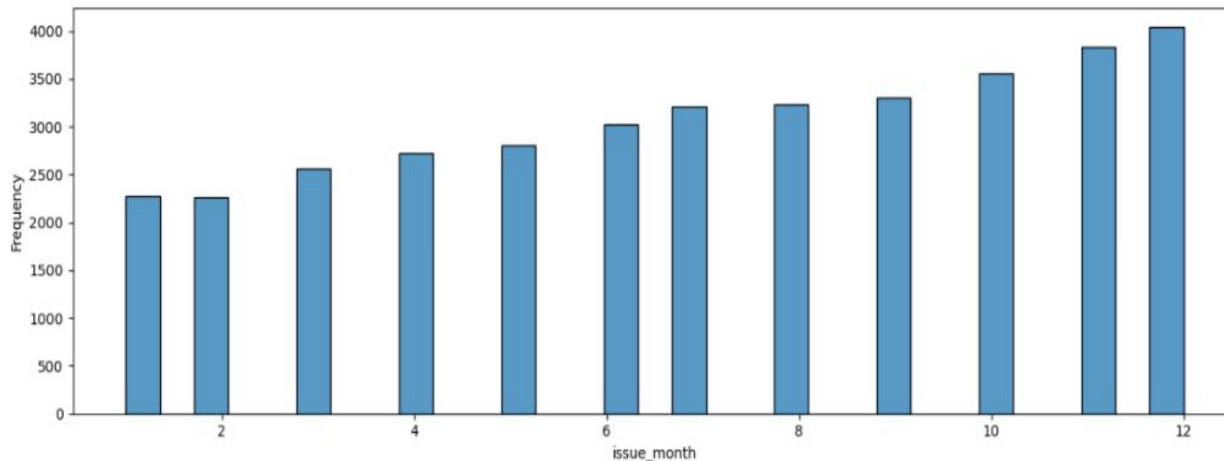

annual_inc

Analyzing the "annual_inc" variable, we can see, income avg income ranges from 40k to 100k
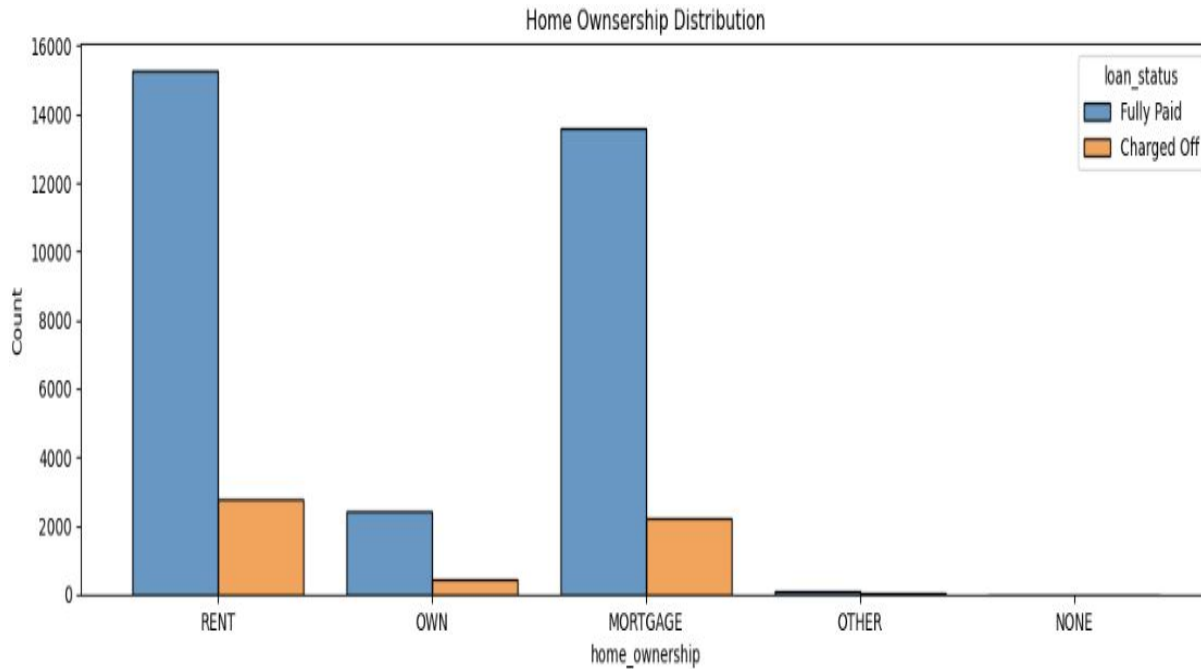
Univariate analysis…

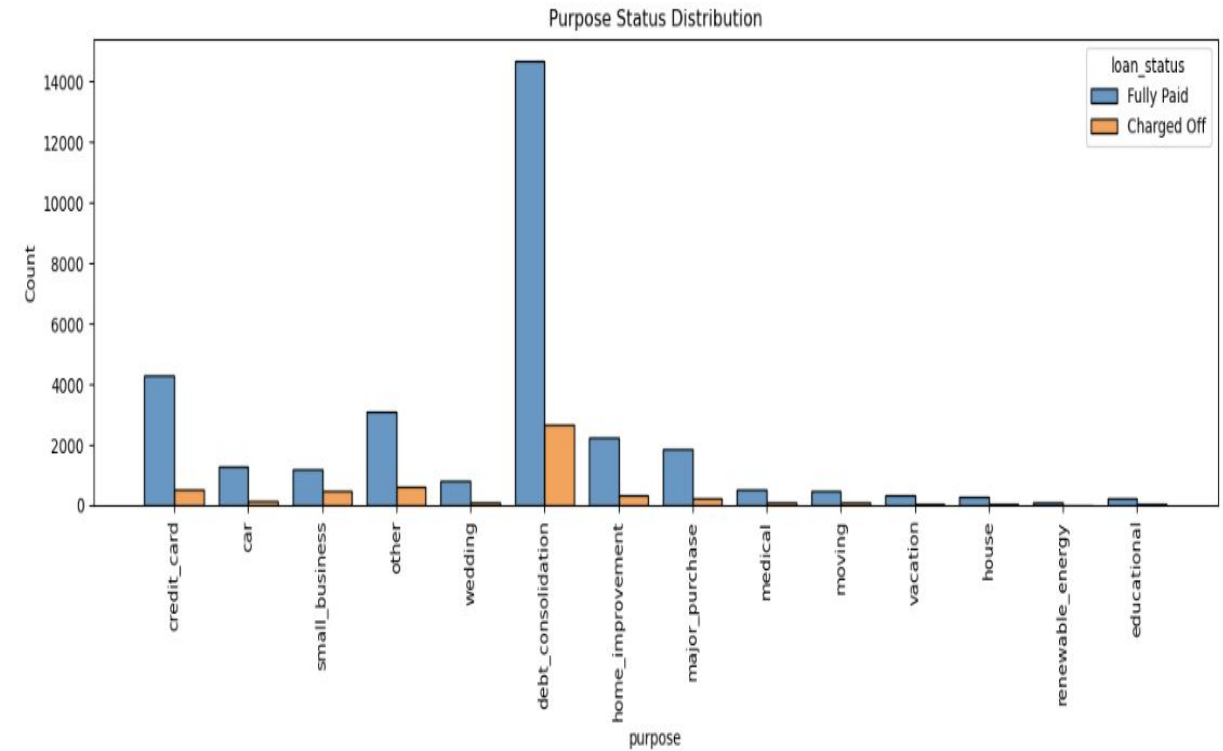Fom "addr_state" variable we can see CA state has highest loan uptake and so on

Fom issue_month and issue_year factor, we can see there is upward trend on loan applications
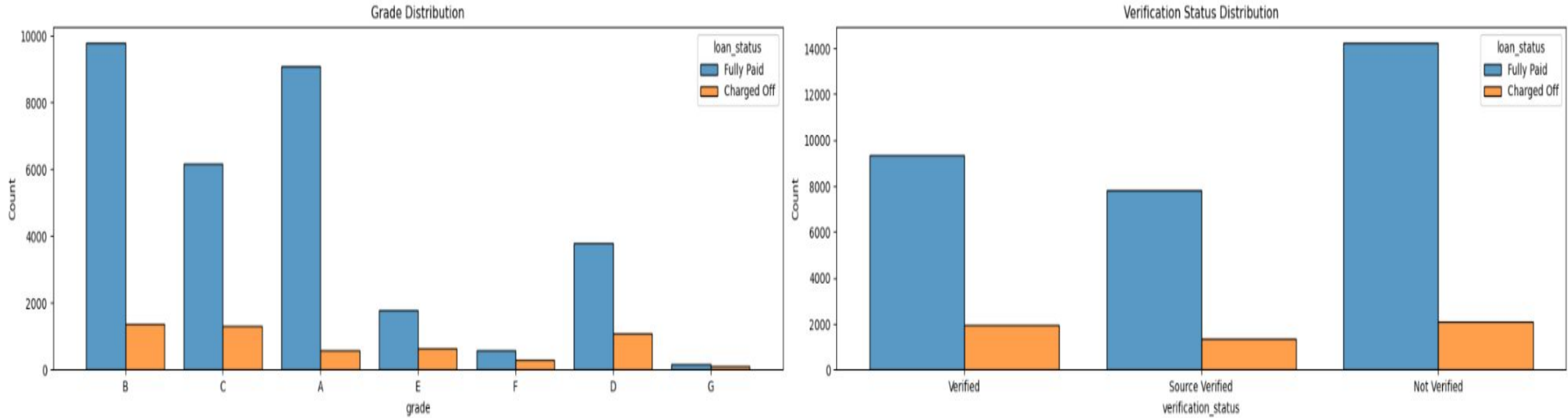
Segmented bivariate analysis…

Home Ownership Distribution

home_ownership vs charged off

higher charged off with people on RENT and
MORTGAGE


Purpose Status Distribution

purpose vs charged off

People with loan purpose of "debt consolidation"
are higher in chance of the charged off

Segmented bivariate analysis…

- Highest loan approval can be seen on B, A and C grade.
- Charged off can be seen on B, C and D grades.
- Similarly Charged off can be seen for the "Not verified" verification status

Segmented bivariate analysis…

Employee Experience Distribution

Interest Rate Distribution

Address State Distribution
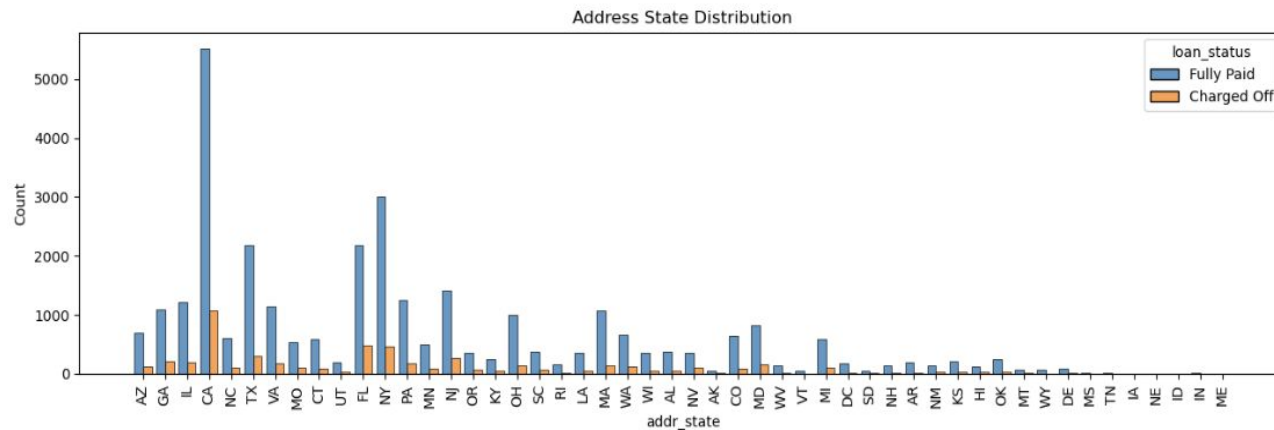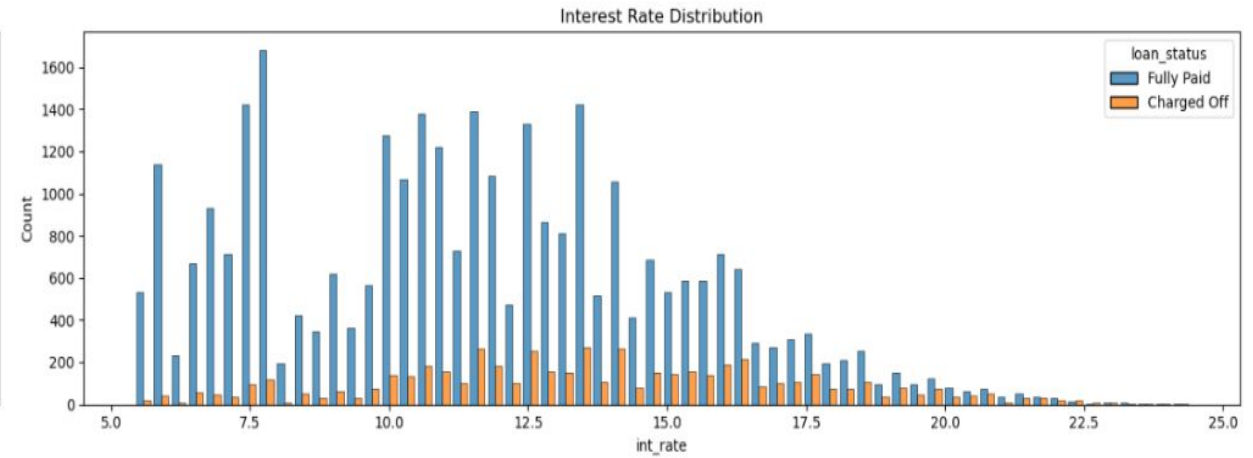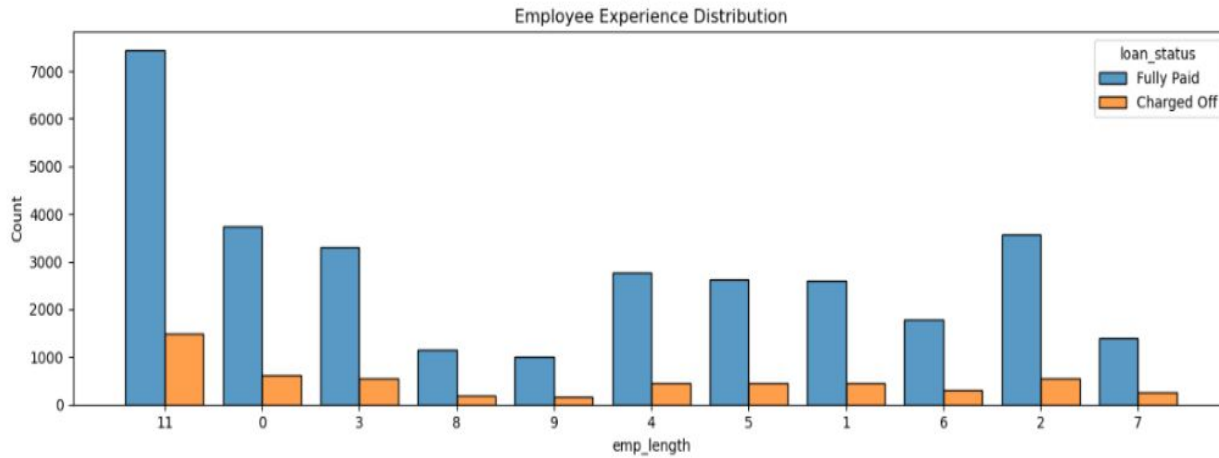
- Application borrowers with higher number of experience are likely to be on loan repayments, however more charged off can also been seen in same segment
- Loan applicant from CA states shows the trend of profitable repayment of the loan, however charged off amongst CA states are higher compare to other states.
- Applicant with flagged with charged off is almost seen in every segments of interest rate

SCATTER PLOT

Multivariate analysis…

Combined Heatmap of Correlation Matrix
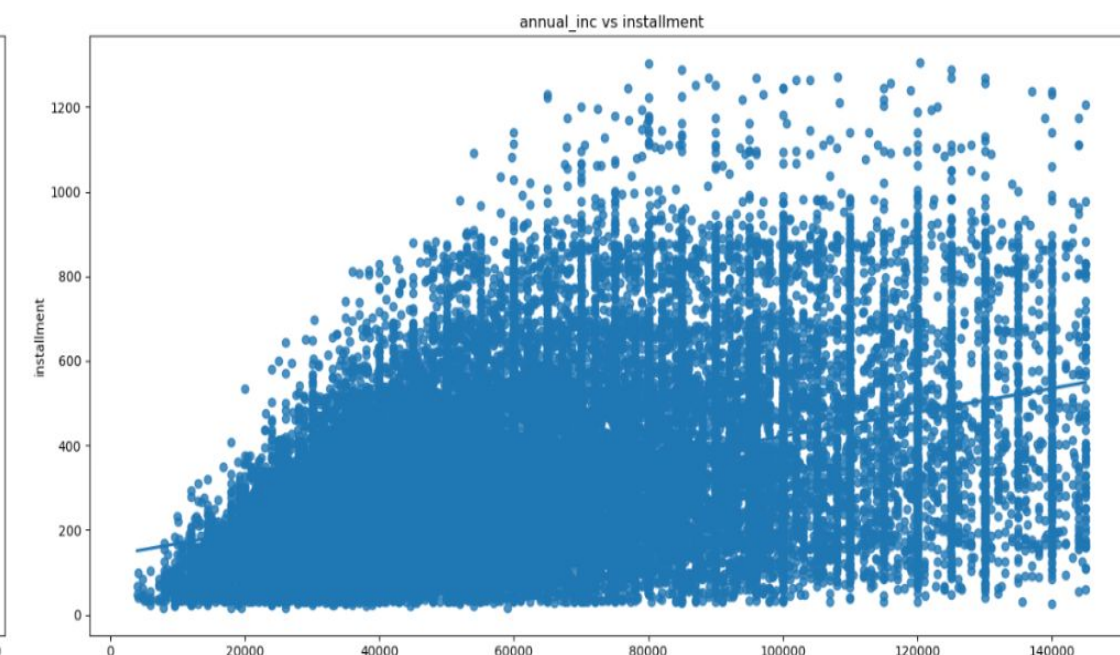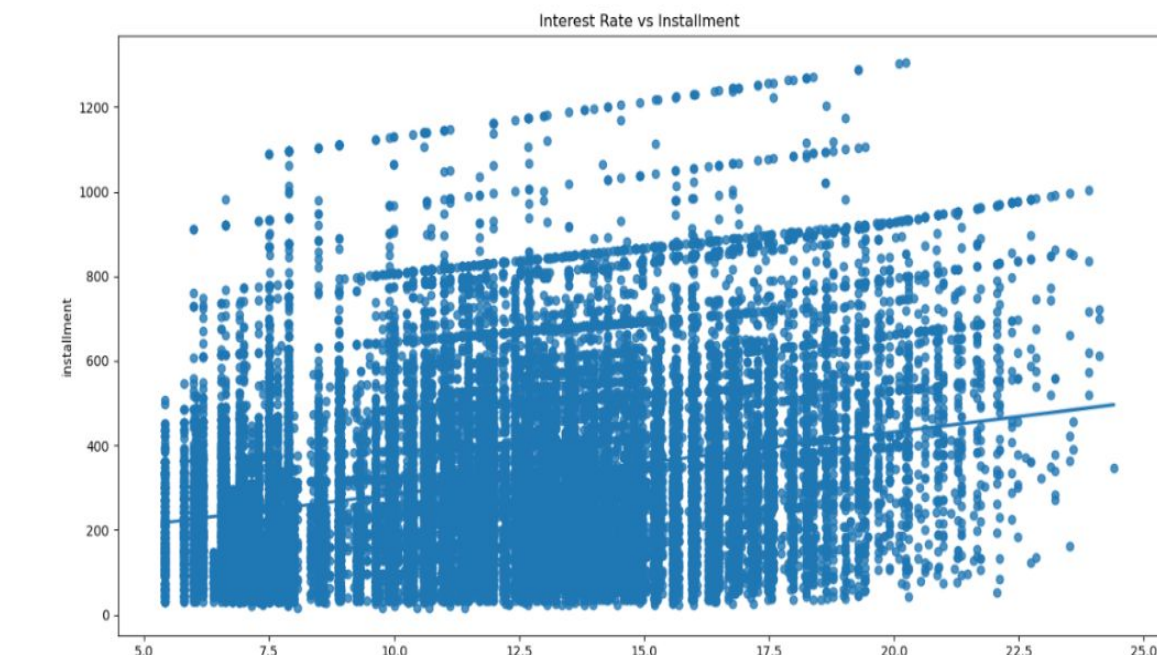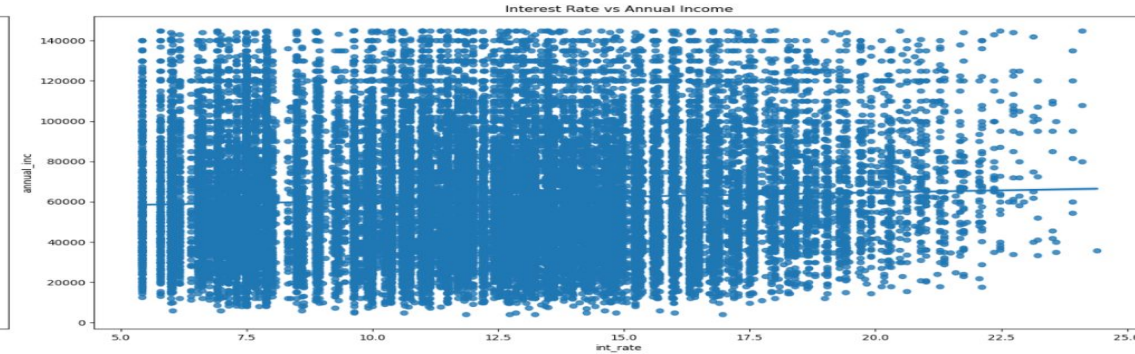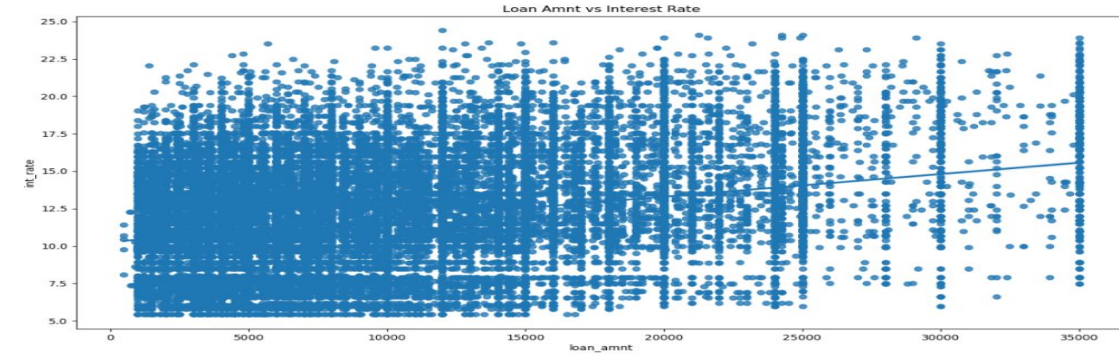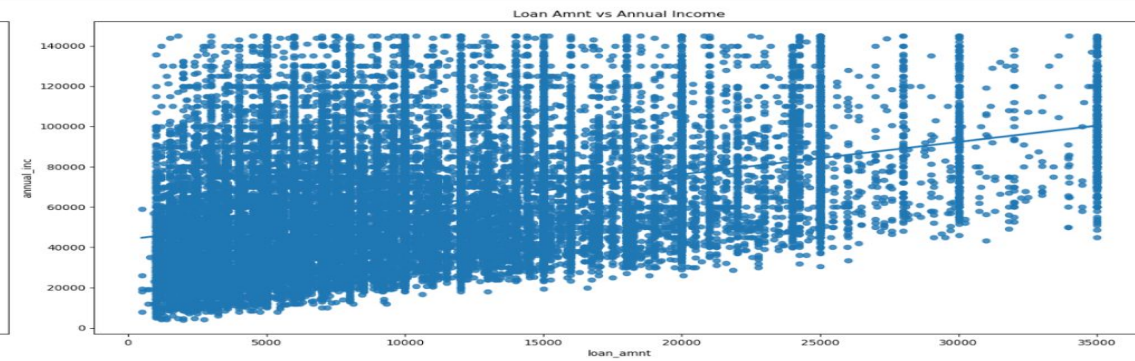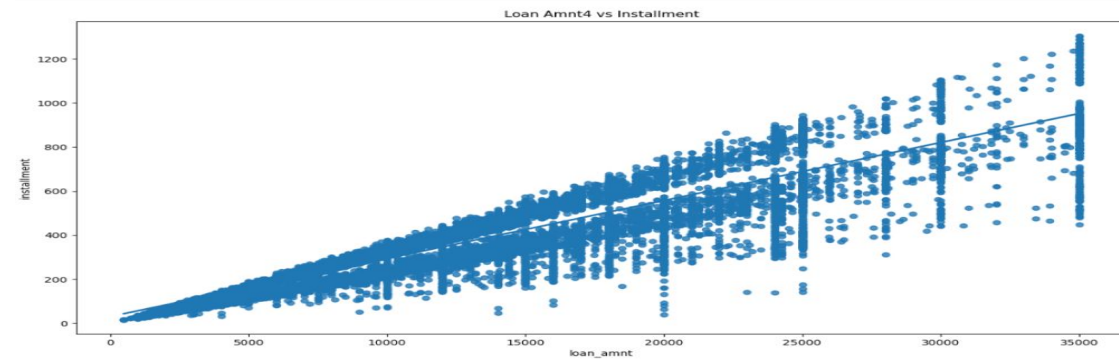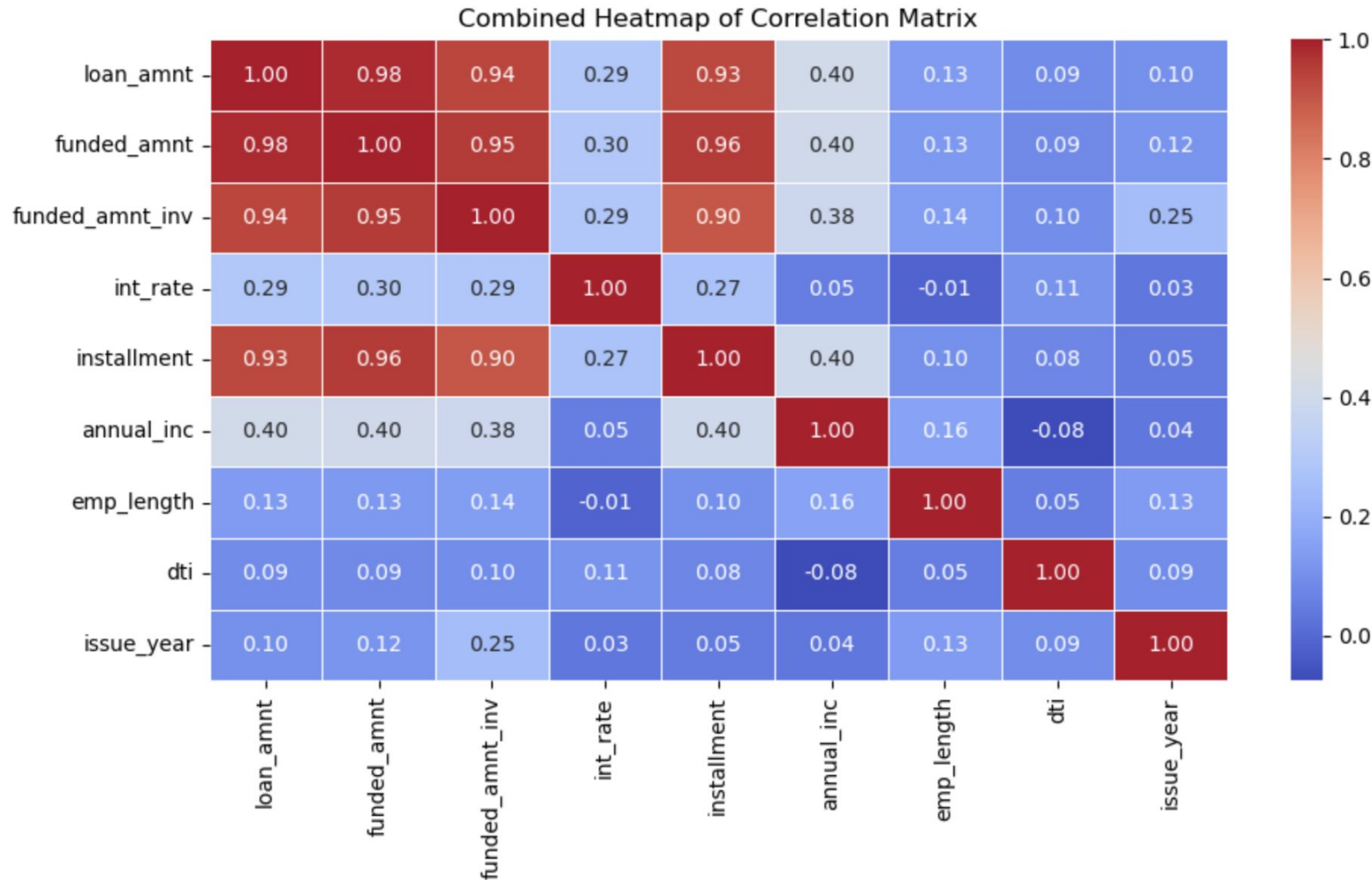
- Loan Amount, Funded Amount, Funded Amount Investment, and Installment show a strong positive correlation.

- Annual Income and Debt-to-Income Ratio (DTI), as well as Interest Rate and Employment Length, shows negative correlations.

Summary..

- Strengthen the verification process :
  - By reducing the "Non-verified" applicant, risks of the defaulter can significantly mitigated
- Consider below factor during the evaluation process
  - Integrate the various risk factors such as, employment length, term , grade during the evaluation process. It will provide more insights for the loan risk assessment.
- State and purpose based restrictions
  - Adding the policies for high risk geographic regions and specific loan purposes especially debt consolidations
- Variables with negative correlations
  - Negatively correlated variables such as annual income vs dit as well as interest rate vs employment length can be considered in evaluation process.