



Lending Club Case Study

Vishrut Mishra

Sireesha Tumpati

Content

- Problem Statement
- Dataset Description
- Dataset Cleaning & Manipulation
- Variables Identification
- Univariate Analysis
- Derived Matrices
- Segmented Univariate Analysis
- Bivariate Analysis
- Correlations
- Conclusions
- Advice to Business

Problem Description

Problem:

A consumer finance company which specializes in lending various types of loans to urban customers. The company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

Objective:

Use EDA to analyze the influence of consumer attributes and loan attributes on tendency to default the loan.

Constraint:

When a person applies for a loan, there are two types of decisions that could be taken by the company:

- ❖ Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:
 1. Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
 2. Current: Applicant is in the process of paying the installments, i.e. the tenure of the loan is not yet completed. These candidates are not labeled as 'defaulted'.
 3. Charged-off: Applicant has not paid the installments in due time for a long period of time, i.e. he/she has defaulted on the loan
- ❖ Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Dataset Description

- The dataset provided contains the information of the loan applicants of the consumer finance company.
- It has about 39717 applications with 111 attributes.
- The attributes can be classified into customer and loan attributes.

Dataset Cleaning & Manipulation

Below steps are performed to clean the whole dataset:

- Identified no headers, footers and duplicate rows to be deleted.
- Identified and dropped the columns which have more than 50% of empty records.
- Identified and dropped the columns with descriptive values, single value, redundant and behavioral in nature which will not influence the analysis to identify the driving factors behind the loan defaulters.
- Identified the rows which has missing value in a column and dropped those rows.
- Dropped the rows with the loan status '*current*', because they can be used to determine the defaulter's trend by the data of applicants who are currently paying their loan.
- Removed outliers from the dataset for column '*annual_inc*', which denotes the annual income of the loan applicants.
- Fixed the value (remove % from end) and changed datatype (from string to float) of the dataset column '*int_rate*' which denotes the loan interest rate.
- After these data cleaning activities , final data left with 36110 rows and 18 columns for data analysis.

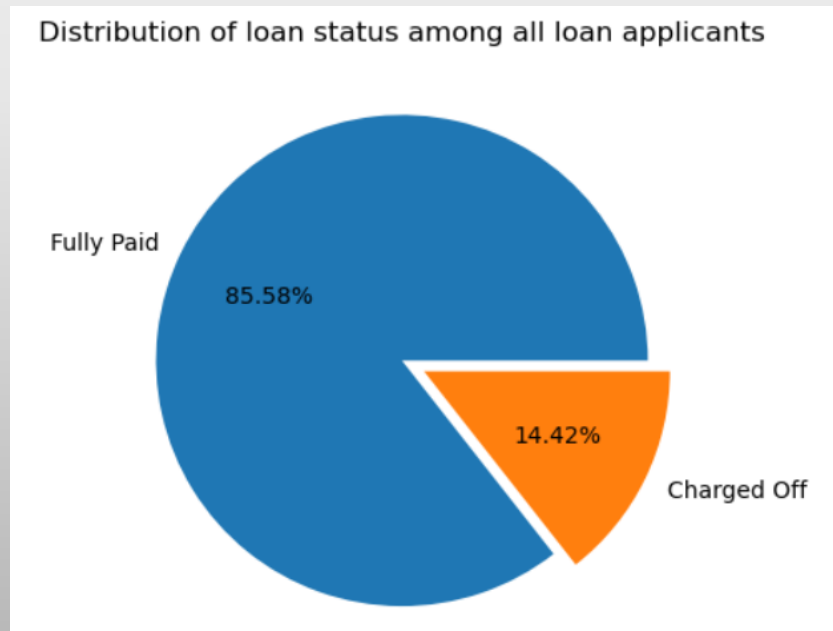
Variables Identification

Selected some of the variables from the dataset and divide them into three categories for better analysis:

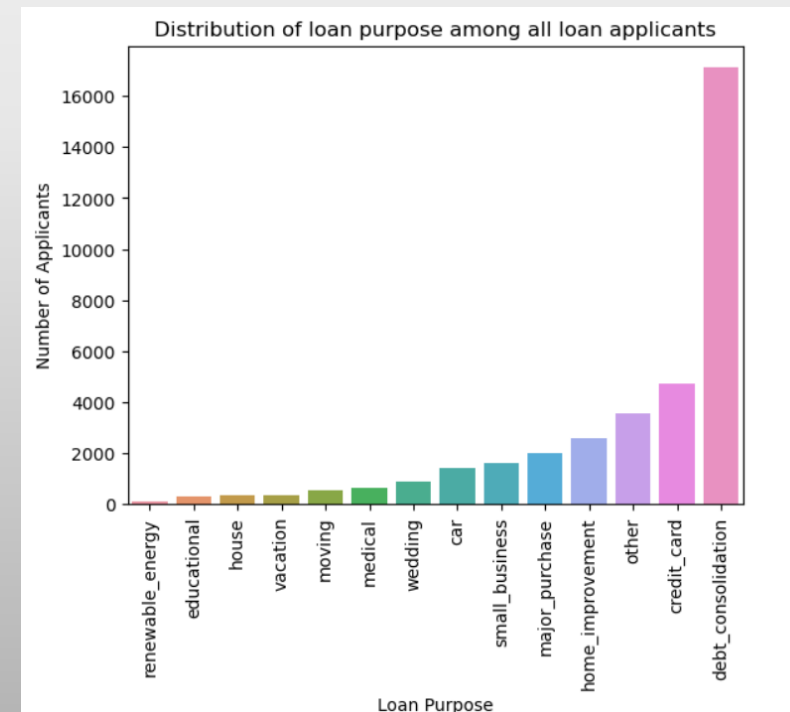
- i) Categorical Variables (loan status, loan purpose, loan term, employment length, home ownership, address state, grade, public record bankruptcies etc.)
- ii) Numeric or Quantitative Variables (annual income, funded amount, interest rate, debt to income ratio, installments etc.)
- iii) Target Variable (loan status)

Univariate Analysis

- Nearly 14.5% of loan applicants has 'charged off' or 'defaulter' status on their loan application.

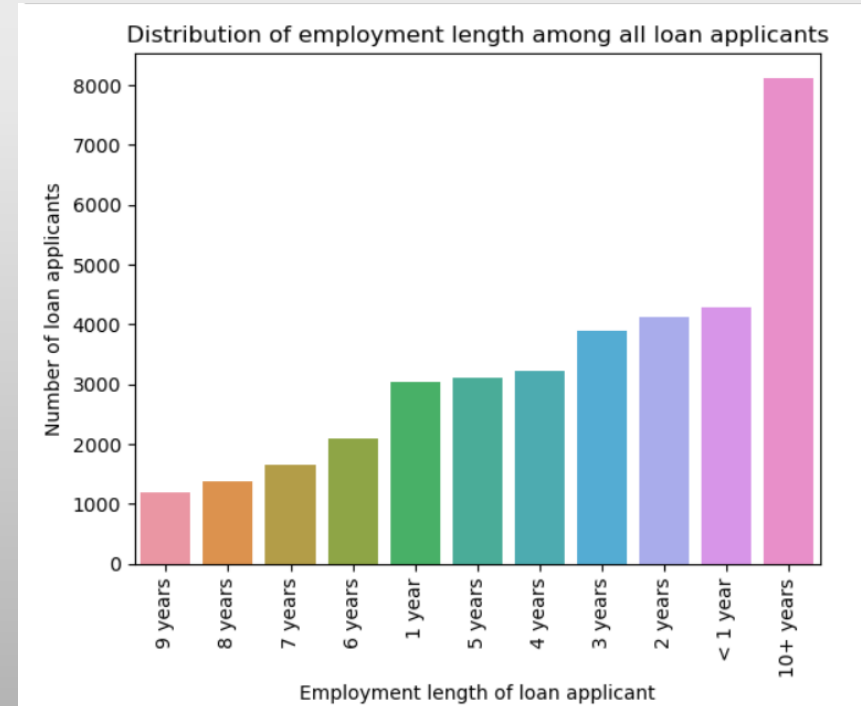
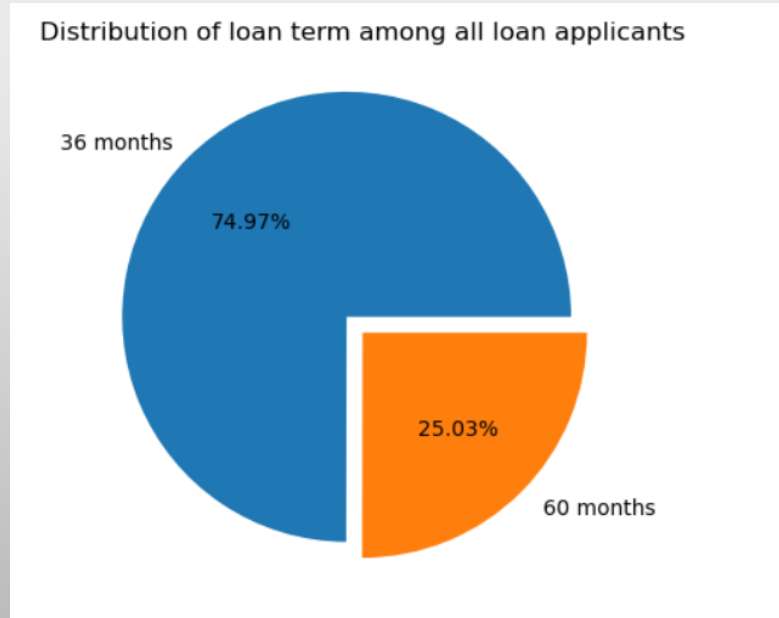


- Around 17K+ loan applications have a loan purpose as 'debt consolidation'.



Univariate Analysis

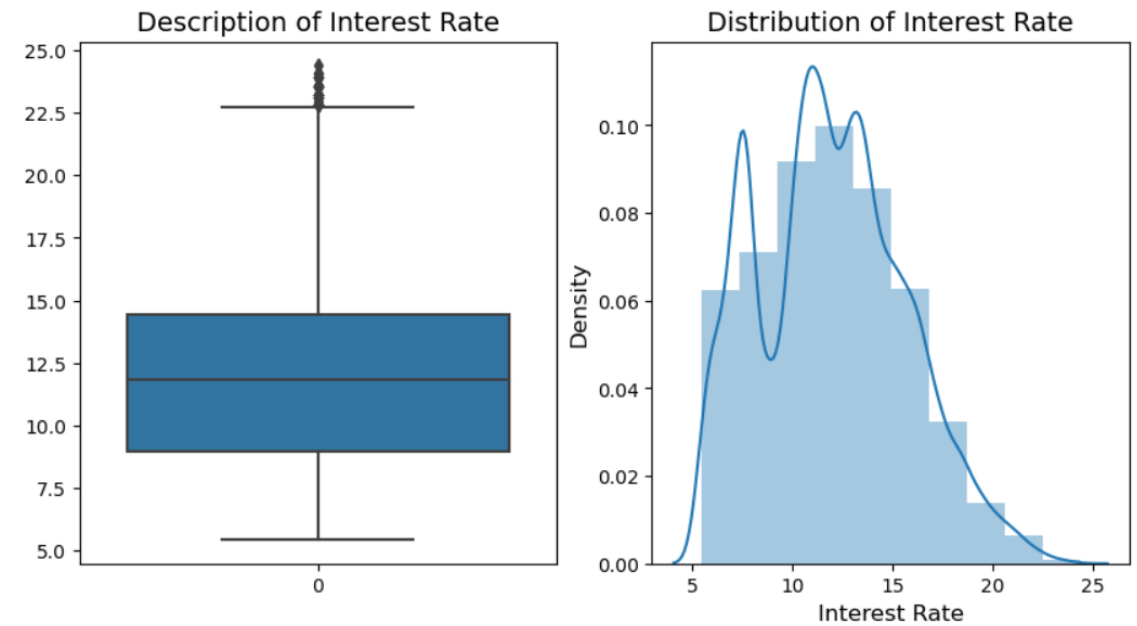
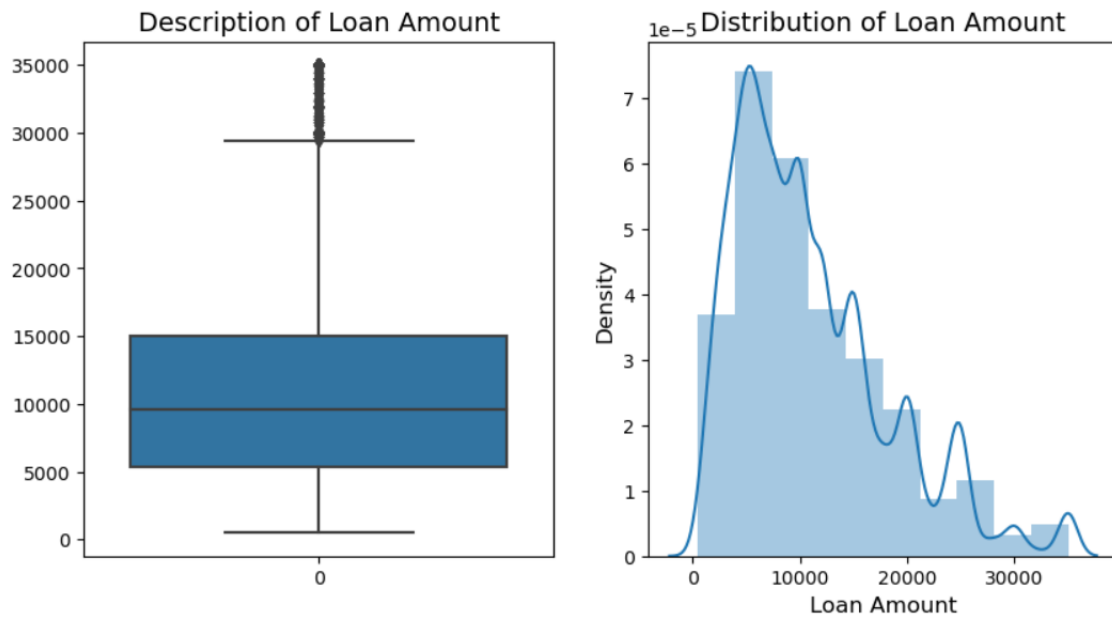
- Nearly 75% of loan applicants select '36 months' of loan term plan.
- Around 8K+ loan applications are of the consumers who has 10+ years of experience.



Univariate Analysis

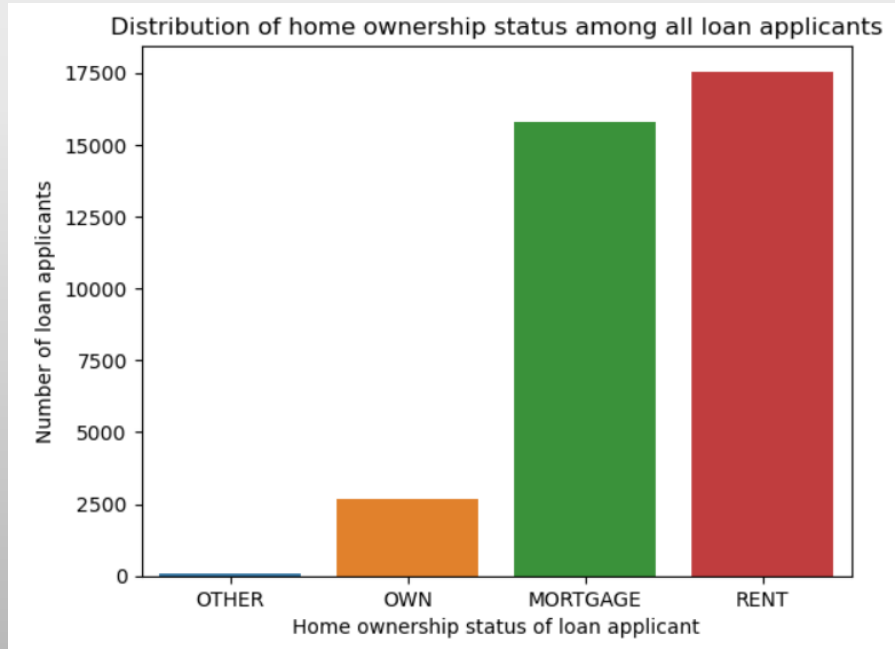
- Loan amount varies from 5K to 15K.

- Loan interest rate varies from 8% to 15%.

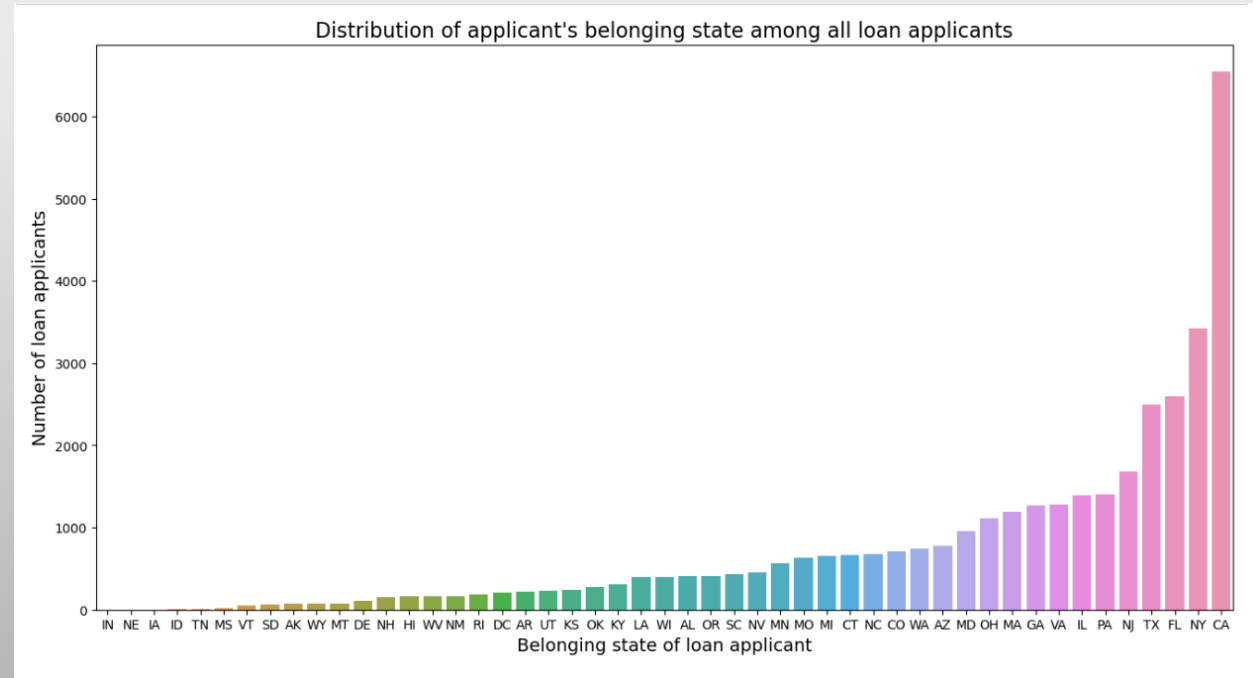


Univariate Analysis

- Around 17K+ loan applicants reside in rented house.

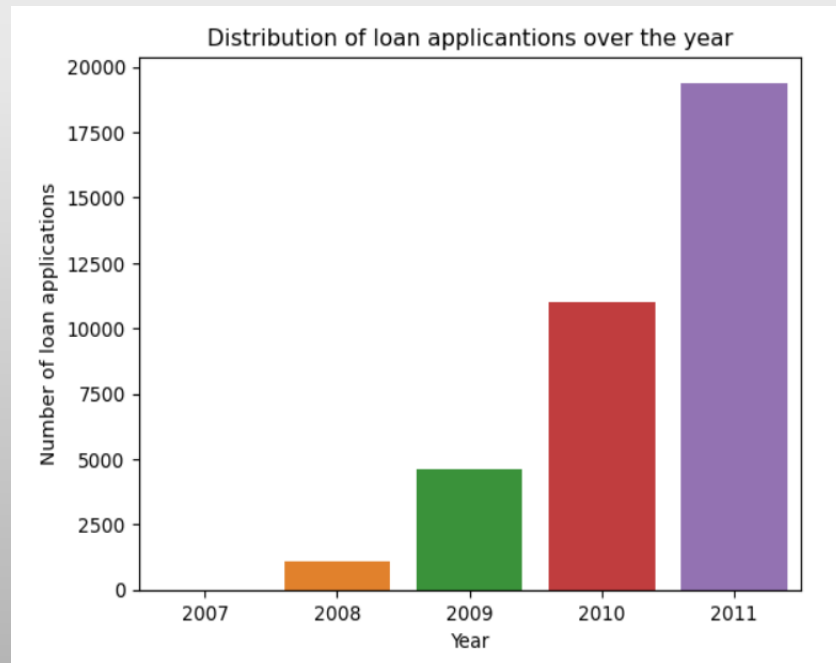


- From state CA, the highest number of loan applications are received.

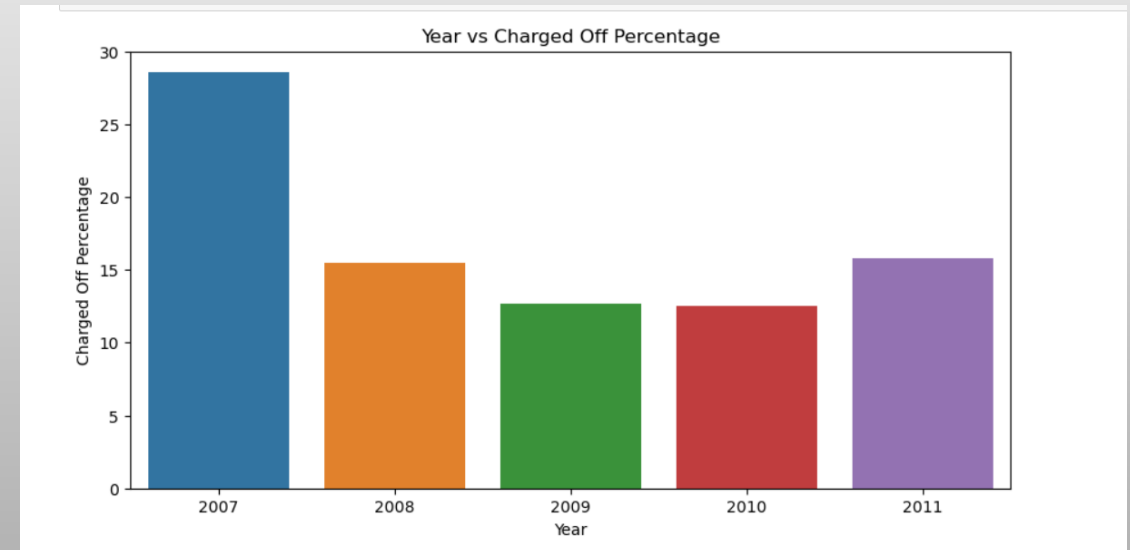


Derived Matrices

- After deriving year from loan issue date, we can say that the loan applications are increased over the year from 2007 to 2011.

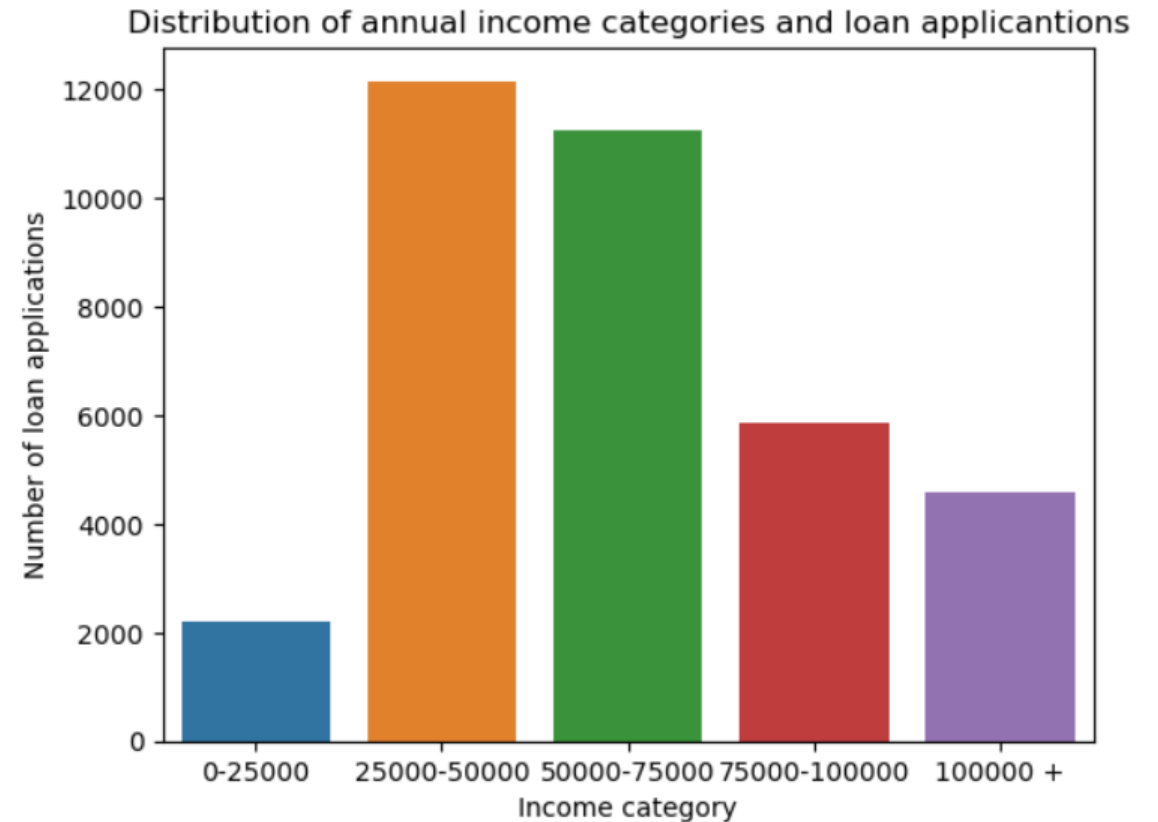


- From 2007 – 2009, the defaulter rate is decreasing.
- From 2009 – 2011, the defaulter rate is increasing.



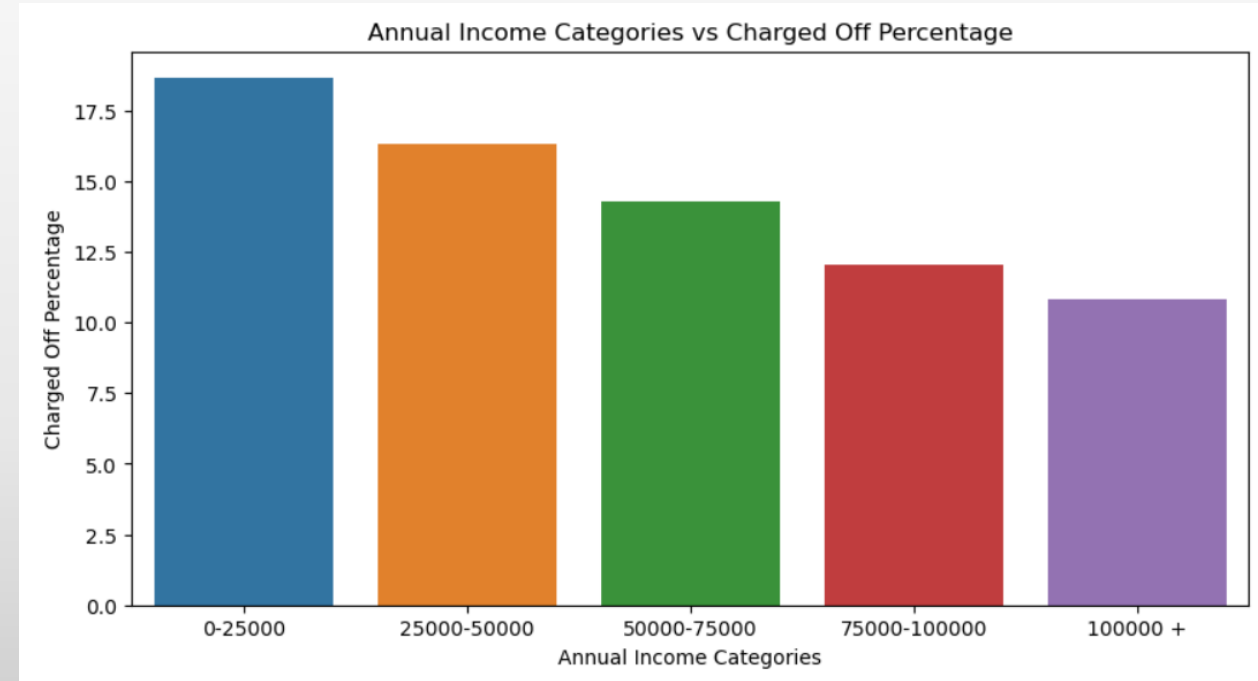
Derived Matrices

- We divided the annual income of loan applicants into the following income categories
 - Below 25K – Has nearly 2K loan applicants
 - Range between 25 to 50K – Has nearly 12K loan applicants
 - Range between 50 to 75K – Has nearly 11K loan applicants
 - Range between 75 to 100K – Has nearly 6K loan applicants
 - Above 100K – Has nearly 5K+ loan applicants



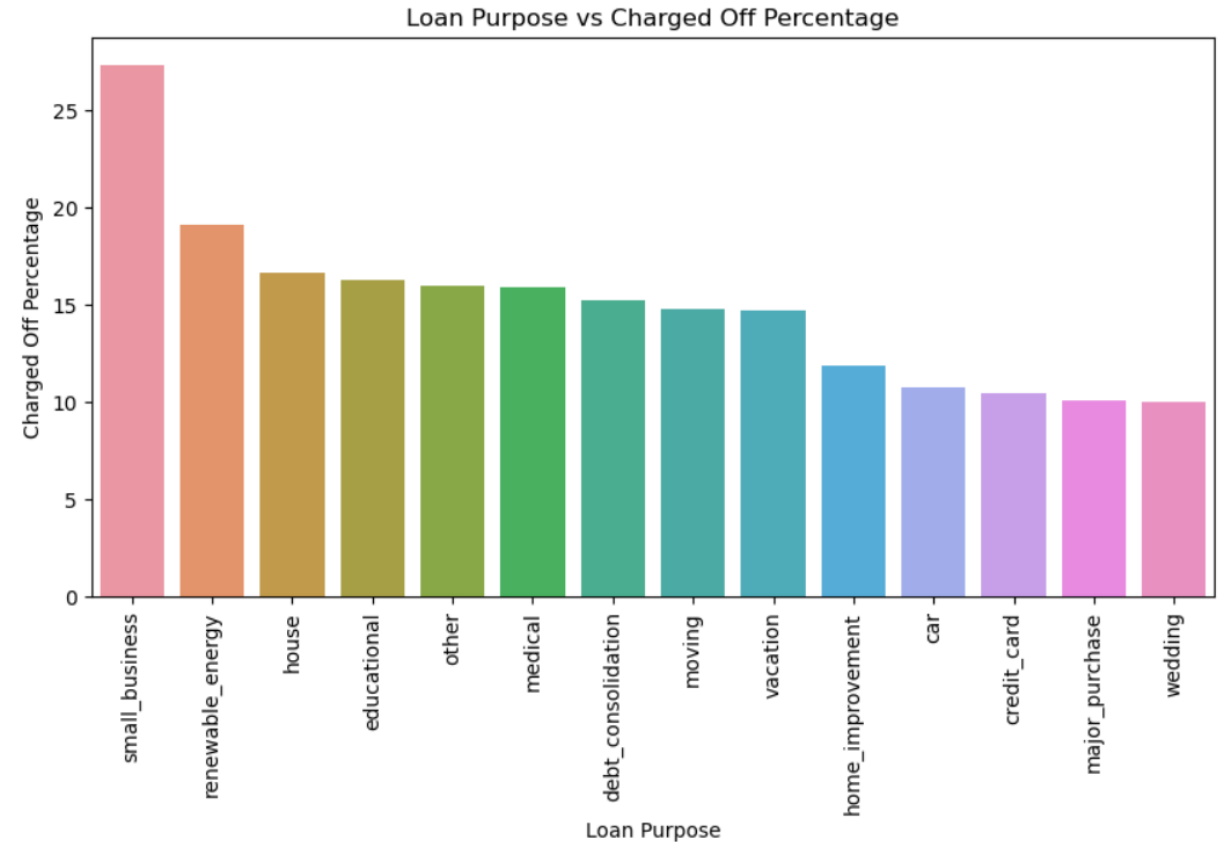
Segmented Univariate Analysis

- People who earn between 0-25K per year are most likely to be defaulter or charged off.
- People who earn 100K+ per year are less likely to be defaulter or charged off.



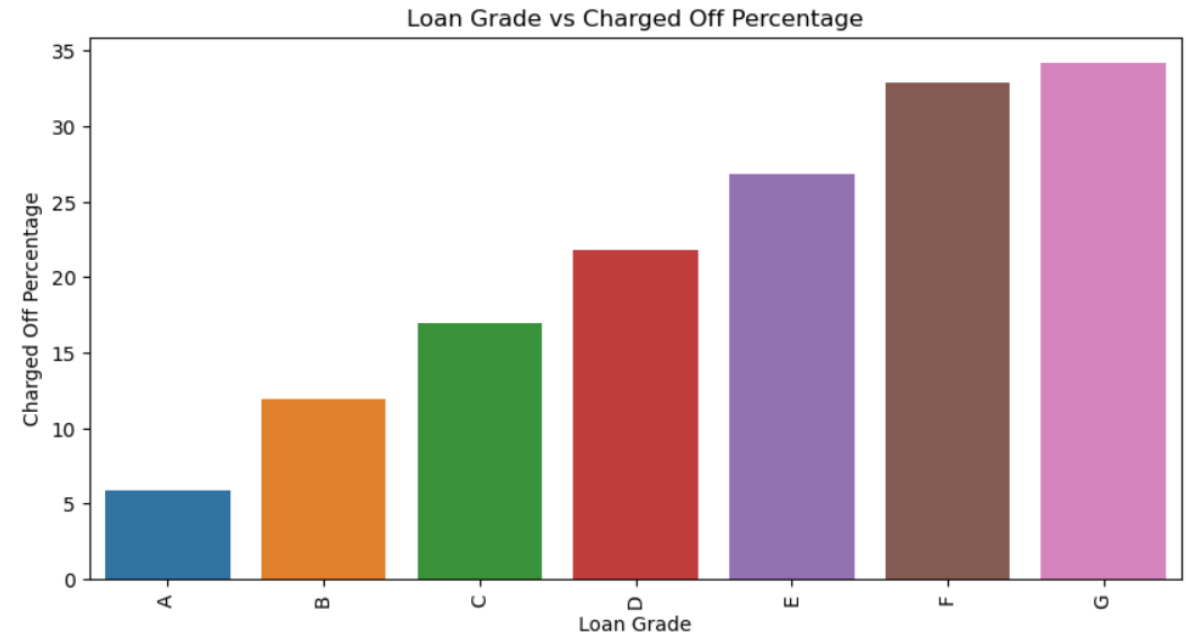
Segmented Univariate Analysis

- People who are taking loan for small business establishment purpose are most likely to be defaulter or charged off.
- People who are taking loan for their wedding purpose are less likely to be defaulter or charged off.



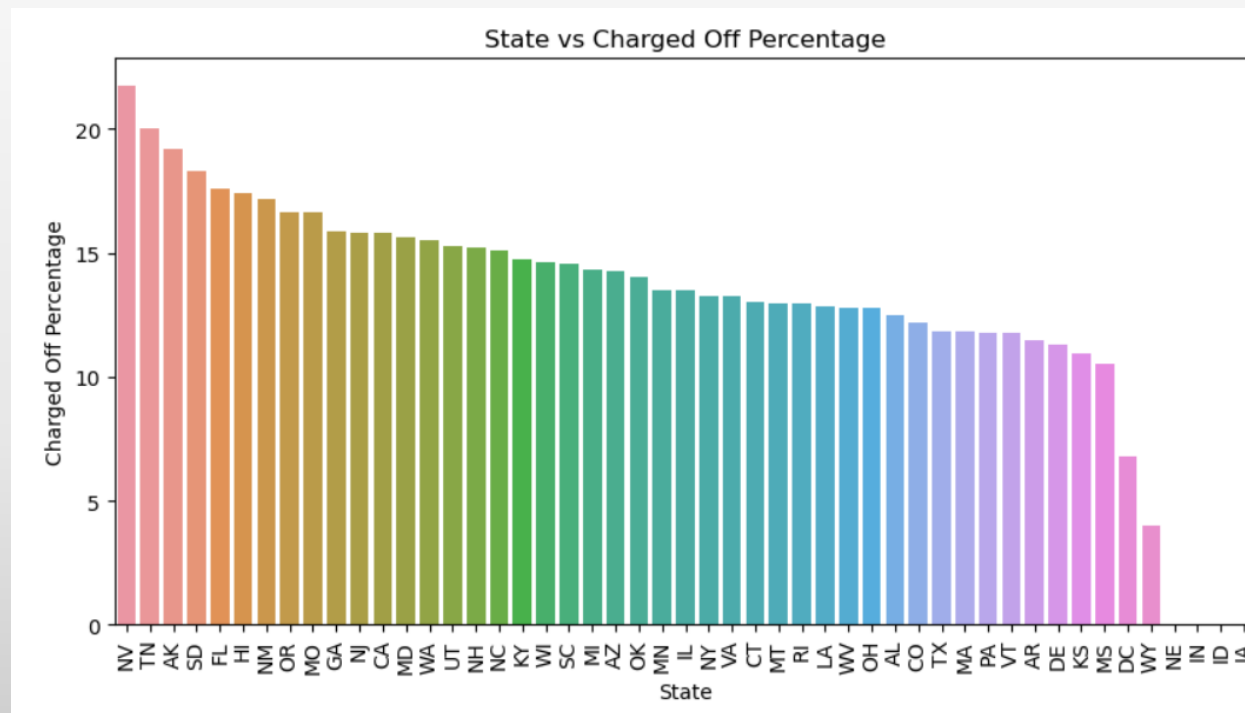
Segmented Univariate Analysis

- Loan grade F and G are most likely to be defaulter or charged off.
- Loan grade A is less likely to be a defaulter or charged off.



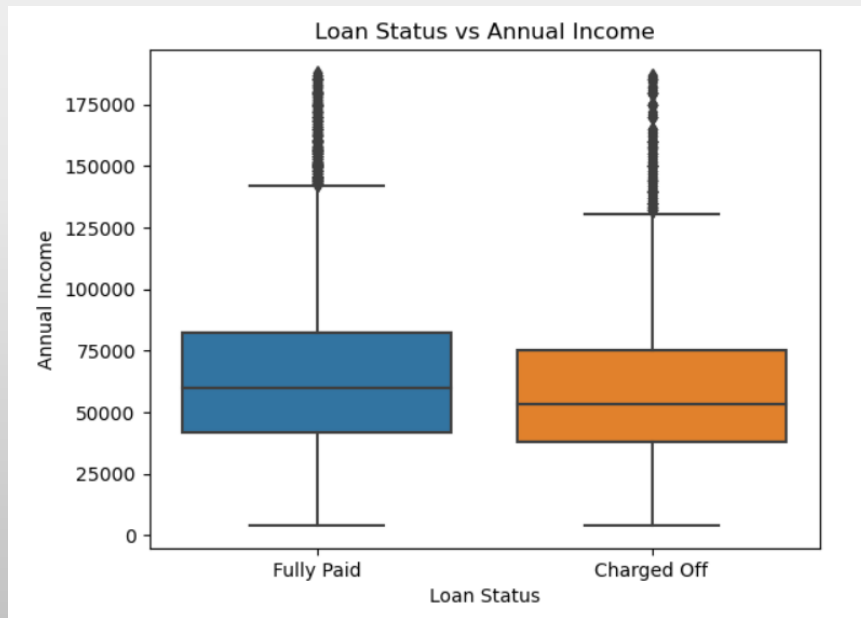
Segmented Univariate Analysis

- People who belong from state NV, TN and AK are most likely to be defaulter or charged off.
- People who belong from state IN, ID and IA are less likely to be defaulter or charged off.

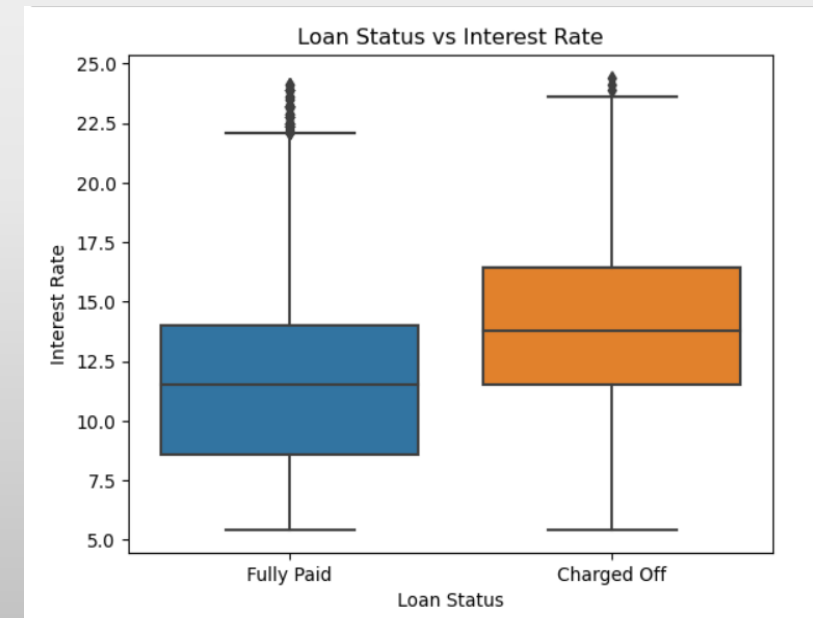


Bivariate Analysis

- Loan applicants who are charged off are drawing less salary than the loan applicants who fully paid their loans.

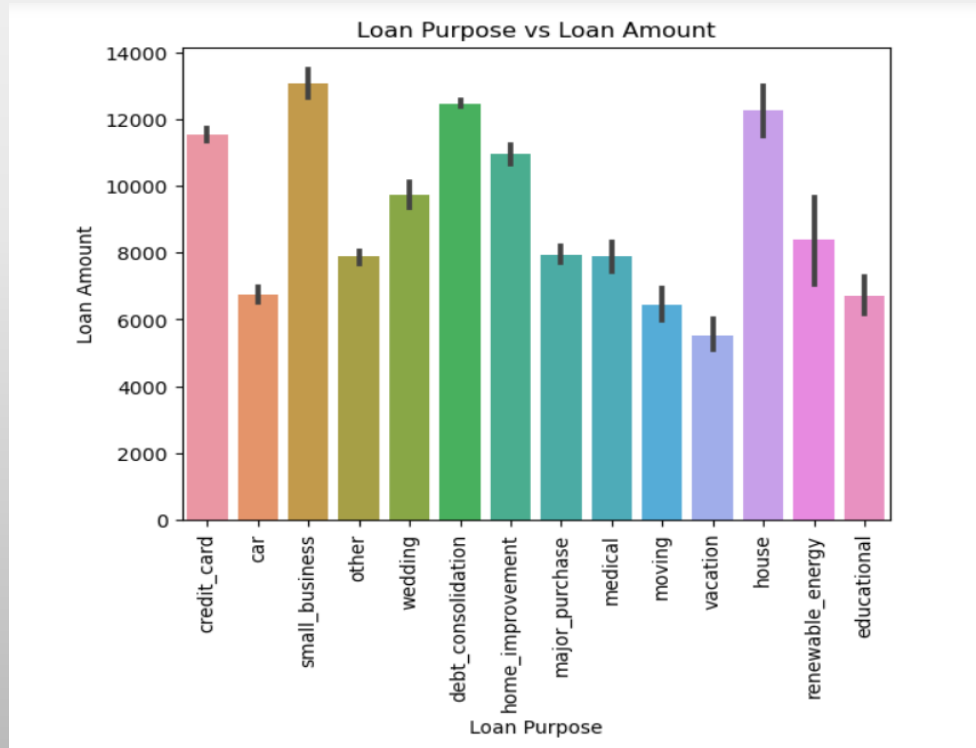


- Loan applicants who are charged off are paying loan with higher interest rate than the loan applicants who fully paid their loans.

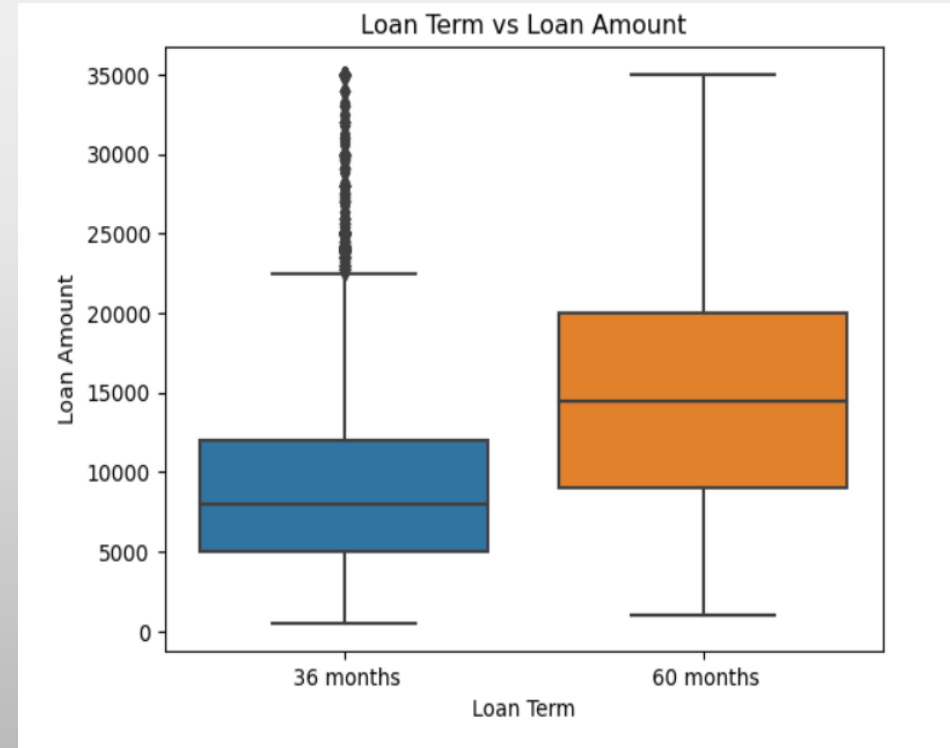


Bivariate Analysis

- Majority of loan amount is taken either for small businesses or for house.

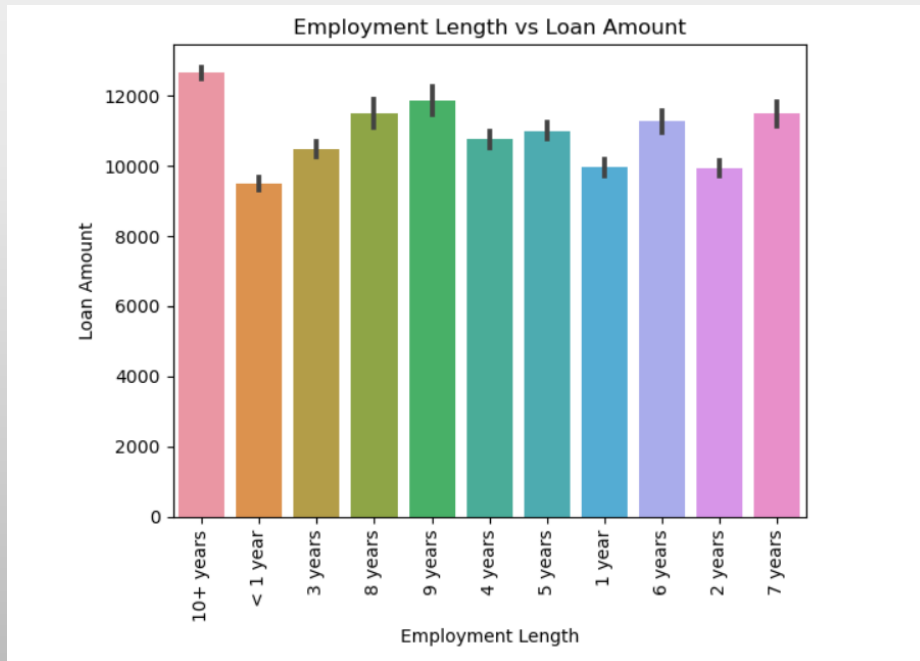


- 60 months loan term amount is greater than 36 months loan term amount.

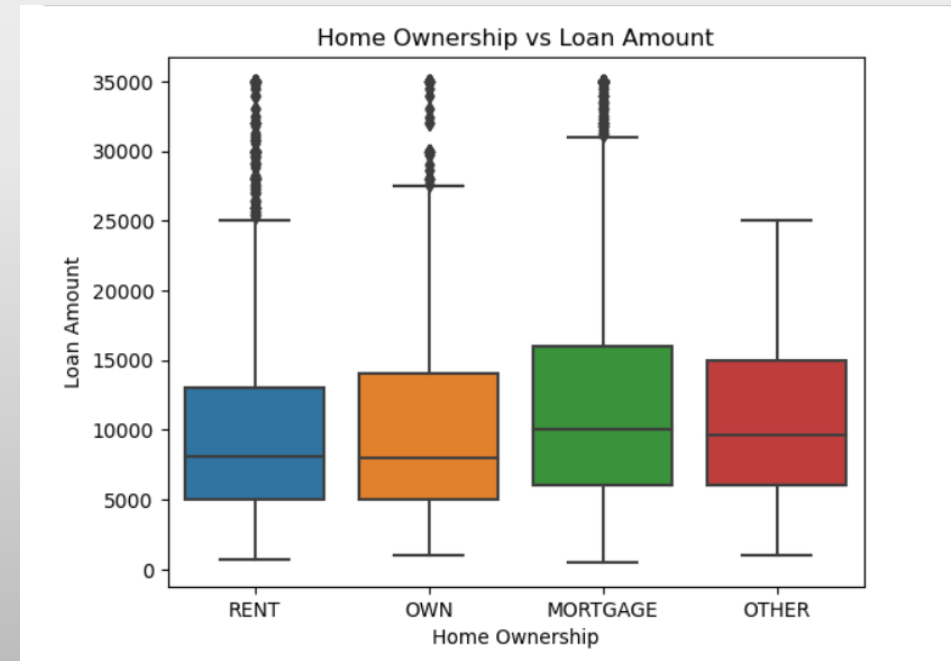


Bivariate Analysis

- Loan applicants who are taking highest amount of loan have more than 10 years of service or employment length.

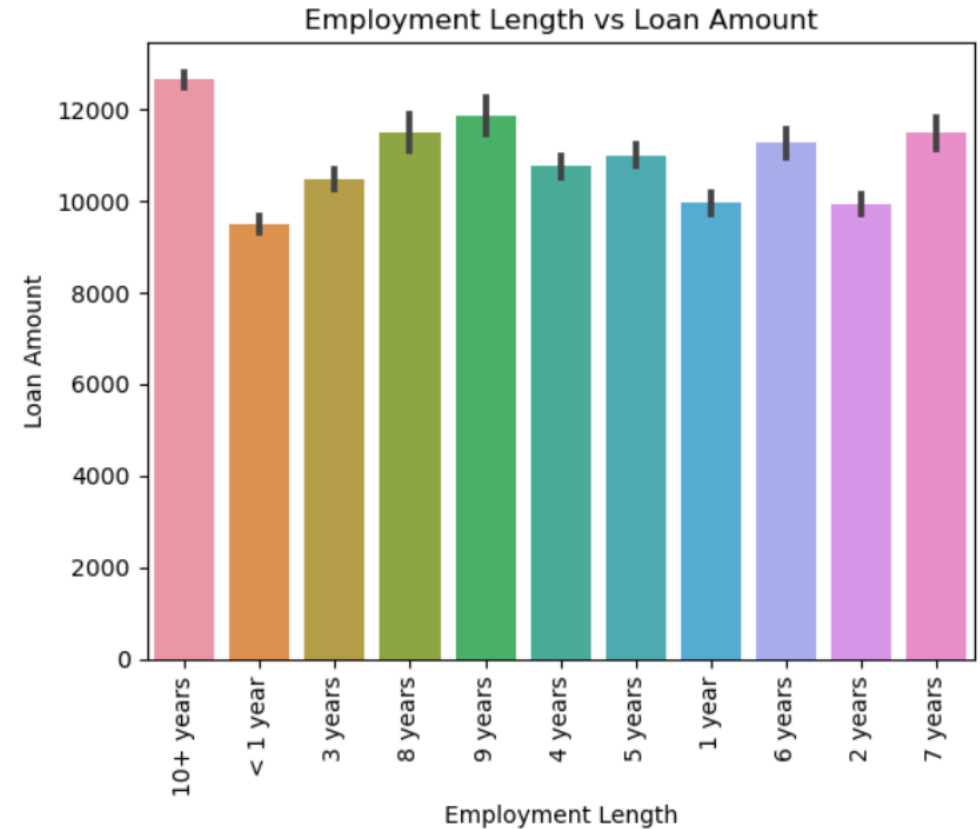


- Mortgage people are taking highest amount of loan.



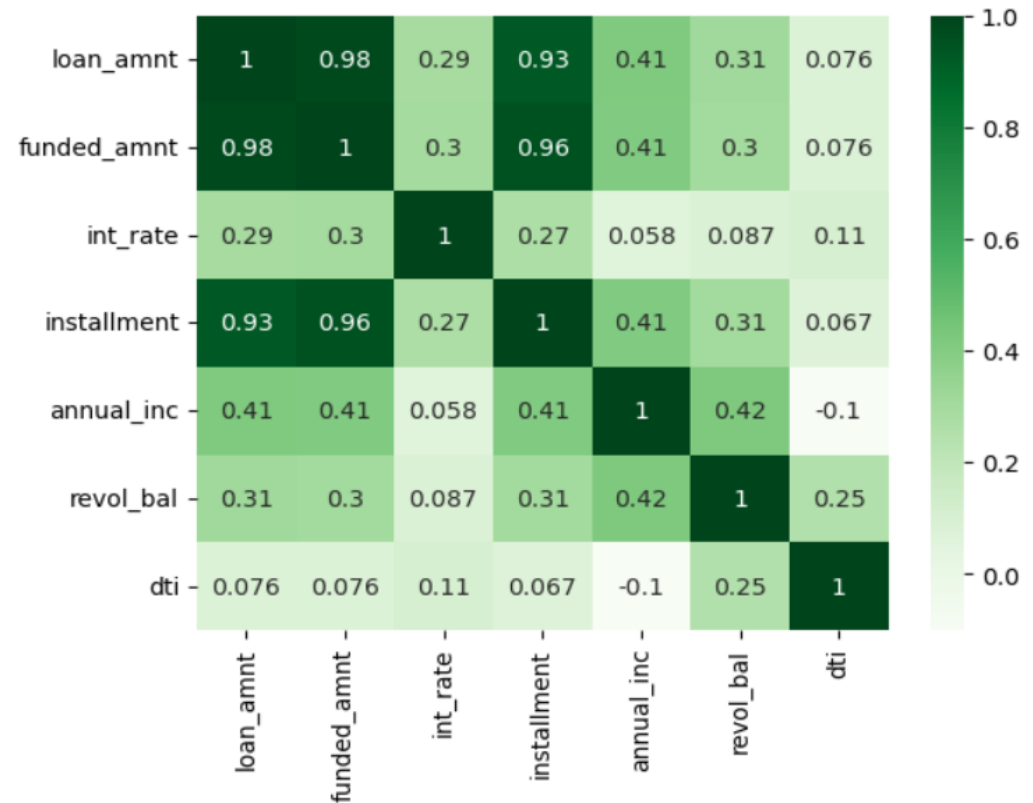
Bivariate Analysis

- People having 10+ years of service length, have mostly recorded their bankruptcy in any public forum.



Correlations

- Loan amount and Funded amount are highly co-related.
- Annual income and Interest rate are less co-related.
- Annual income and debt to income ratio are negatively co-related.



Conclusions

- Applicants with annual income more than 100K+ have less likely to be charged off and with income less than 25K+ most likely to be defaulters.
- High Interest Rate loans have high chances of charge off.
- Applicants having 10+ years of service length are taking highest amount of loan and hence have mostly recorded their bankruptcy in any public forum.
- Applicants who have taken loan for small business establishment purpose are most likely to be defaulter or charged off.
- Applicants who have taken loan for their wedding purpose are less likely to be defaulter or charged off.
- Applicants with loan grades A & B are likely with less defaulters while F & G have more.

Advice to Business

- Business should accept the loan application of a person having 100K+ annual salary as he is least likely to be a defaulter.
- Business should control to give loans to the people who are earning below 25K+ salary.
- Business should examine more on loan grades before approving the loan application, as grade A & B have less defaulters while F & G have more.
- Business should control their loan in state NV, TN and AK.
- Business should accept more loan application from IN, ID and IA.
- Business should be careful while approving the business loans and house loans.
- Business should accept more loans for the wedding or vacation purpose.
- Business should avoid giving loans to bankrupt.