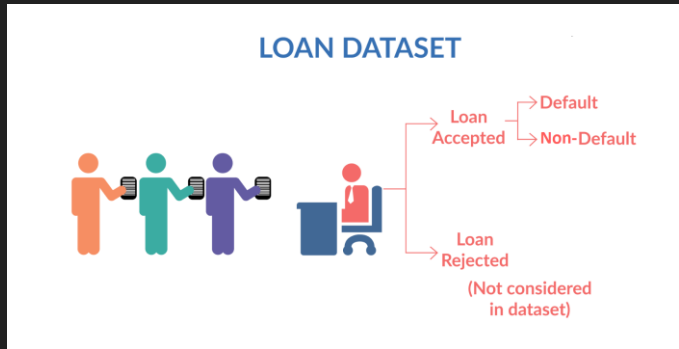
Lending Club Case Study

Nawal Thapliyal and Prabhakar Khandavalli

Overview

Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface

- Company wants to minimize the credit loss from loan defaults
- Company wants to utilize Exploratory Data Analysis for portfolio and risk assessment
- Objective of this case study is to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default
- EDA will be performed to understand how consumer attributes and loan attributes influence the tendency of default



Data Understanding

- Data set used for case study contains the complete loan data for all loans issued through the time period 2007 to 2011
- Dataset contain total of 111 columns and ~39.7K rows
- At high level data set contains below features
  - Loan amount, Loan Status (Fully Paid, Charged Off or Current), Interest rate, purpose of the loan
  - Term (36 or 60 months)
  - Customer address
  - Customer employment details
  - Annual income
  - Loan Grade and Sub Grade
  - Home ownership
  - Delinquency details
  - Revolving line utilization rate

Data Cleaning

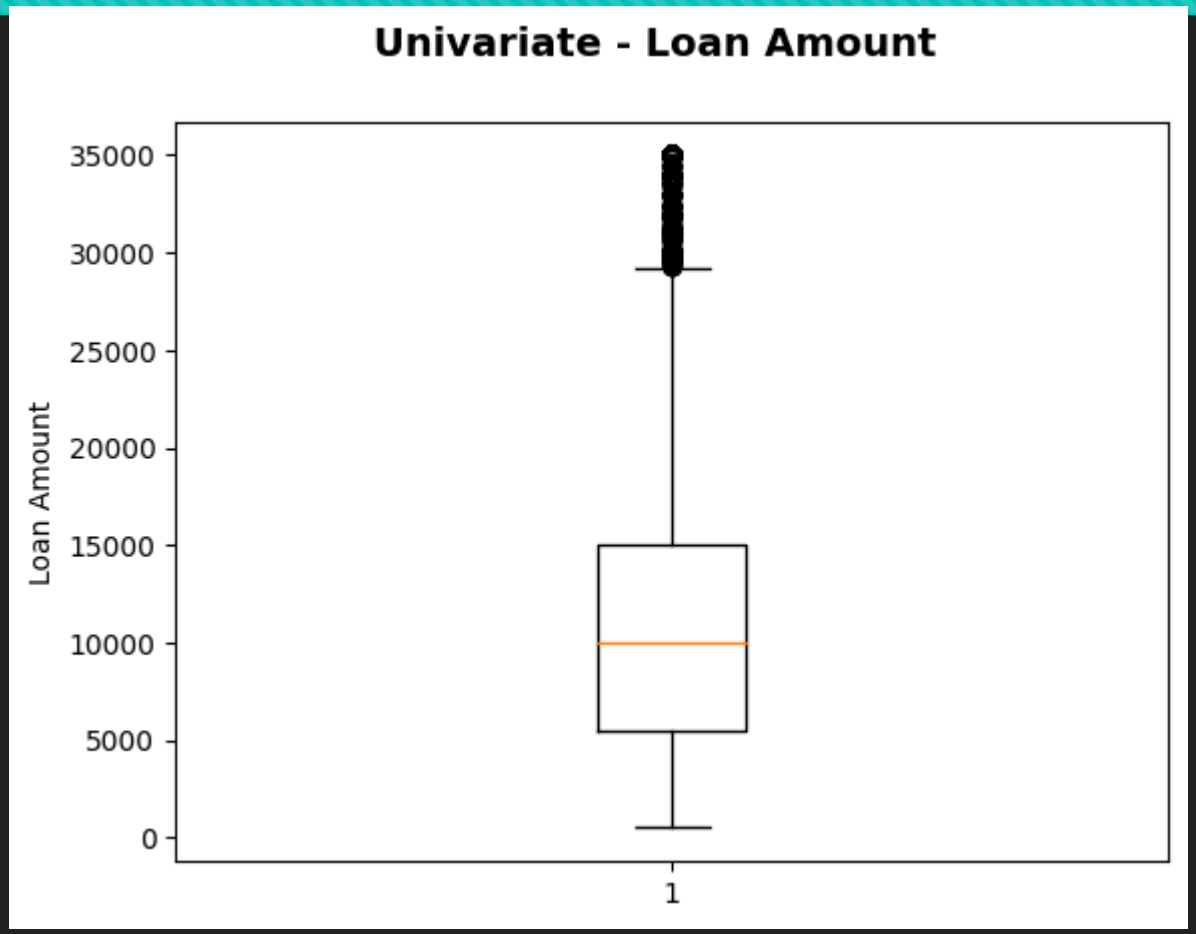
As part of data cleaning, following steps were performed on the dataset

- Fixing Columns
  - Removing features which has all '0', 'NA' or NaN values (ex. Tot_hi_cred_lim, total_bal_ex_mort, total_bc_limit etc.)
  - Removing features which has only one value (ex. tax_liens, collections_12_mths_ex_med, policy_code, delinq_amnt etc.)
  - Removing features which has no significance on the analysis (ex. id, member_id, funded_amnt, funded_amnt_inv etc.)
- Fixing missing values
  - Converting NaN values in the numeric columns to 0
- Standardizing values
  - Converting % columns into numeric columns to enable better visualization (ex. int_rate, revol_util)
  - Changing field datatypes from float to int wherever applicable (ex. mths_since_last_delinq)
  - Converting term to numeric value by removing 'months'
- Delete outliner rows
  - Delete rows which has annual income falling outside of 95^{th} percentile

Univariate Analysis

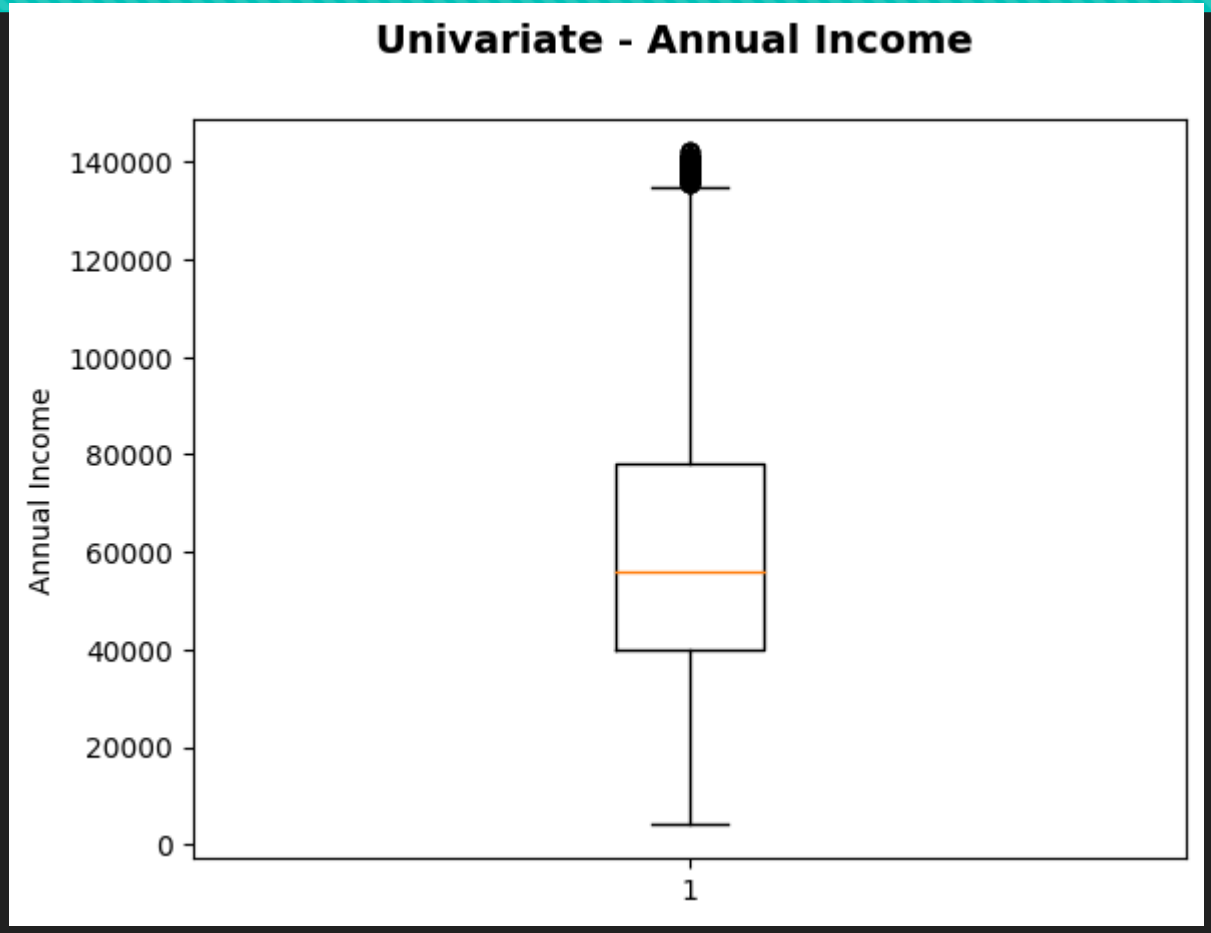
Loan Amount

- Percentile values for loan amount
  - 25% 5500
  - 50% 10000
  - 75% 15000



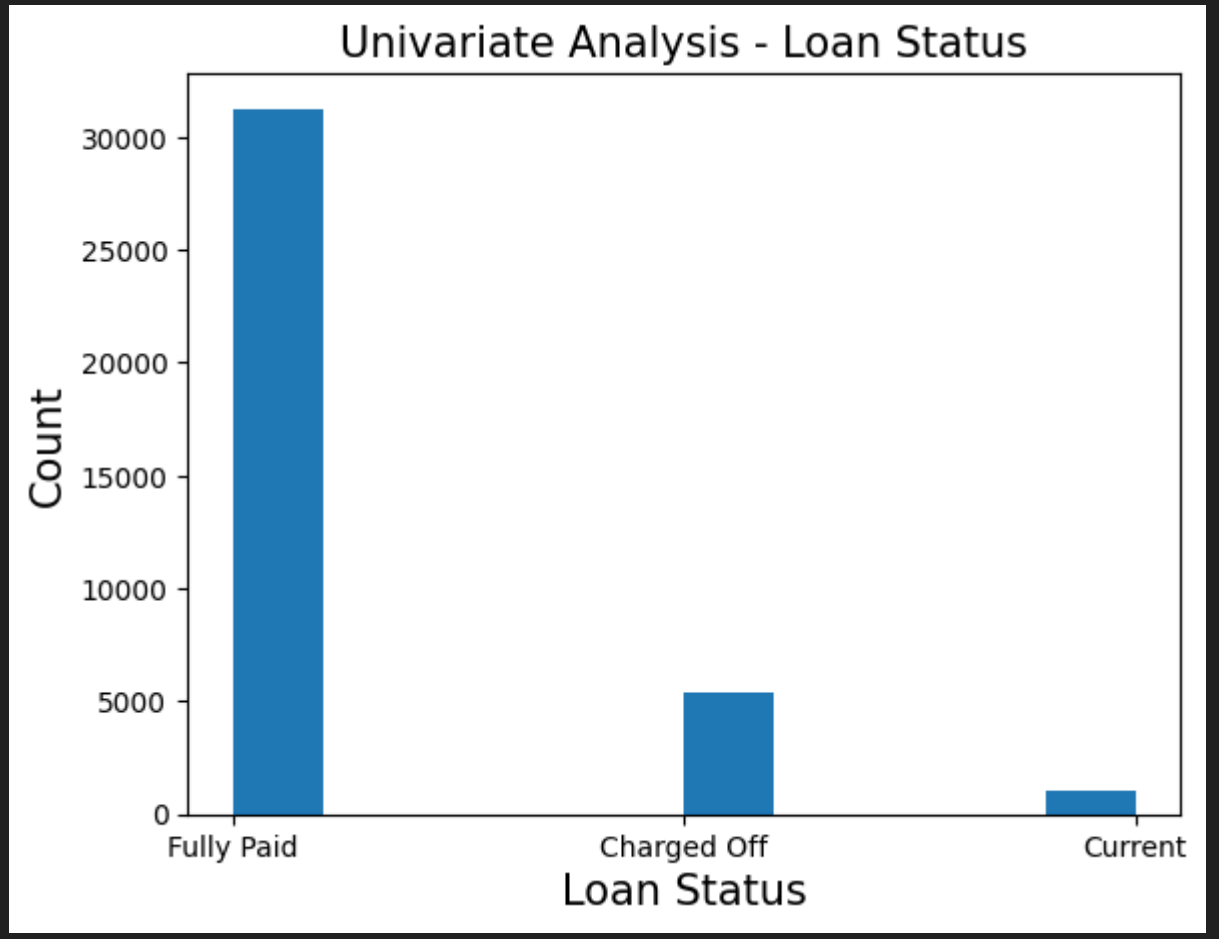
Annual income

- After removing outliners with 95th percentile, annual income percentile values are
  - 25% 40000
  - 50% 56000
  - 75% 78000



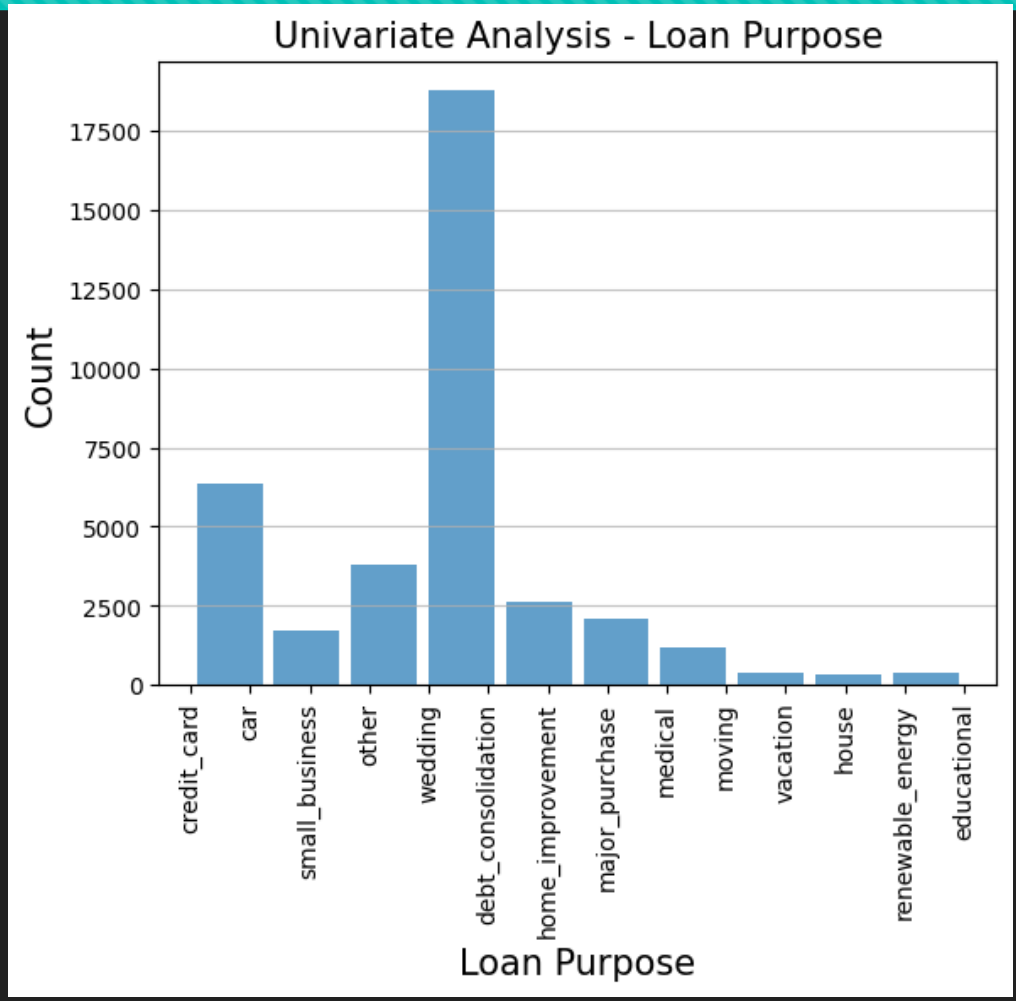
Loan Status

- Changed of loans are around ~13% of overall loans
  - Fully Paid 31256
  - Charged Off 5420
  - Current 1054



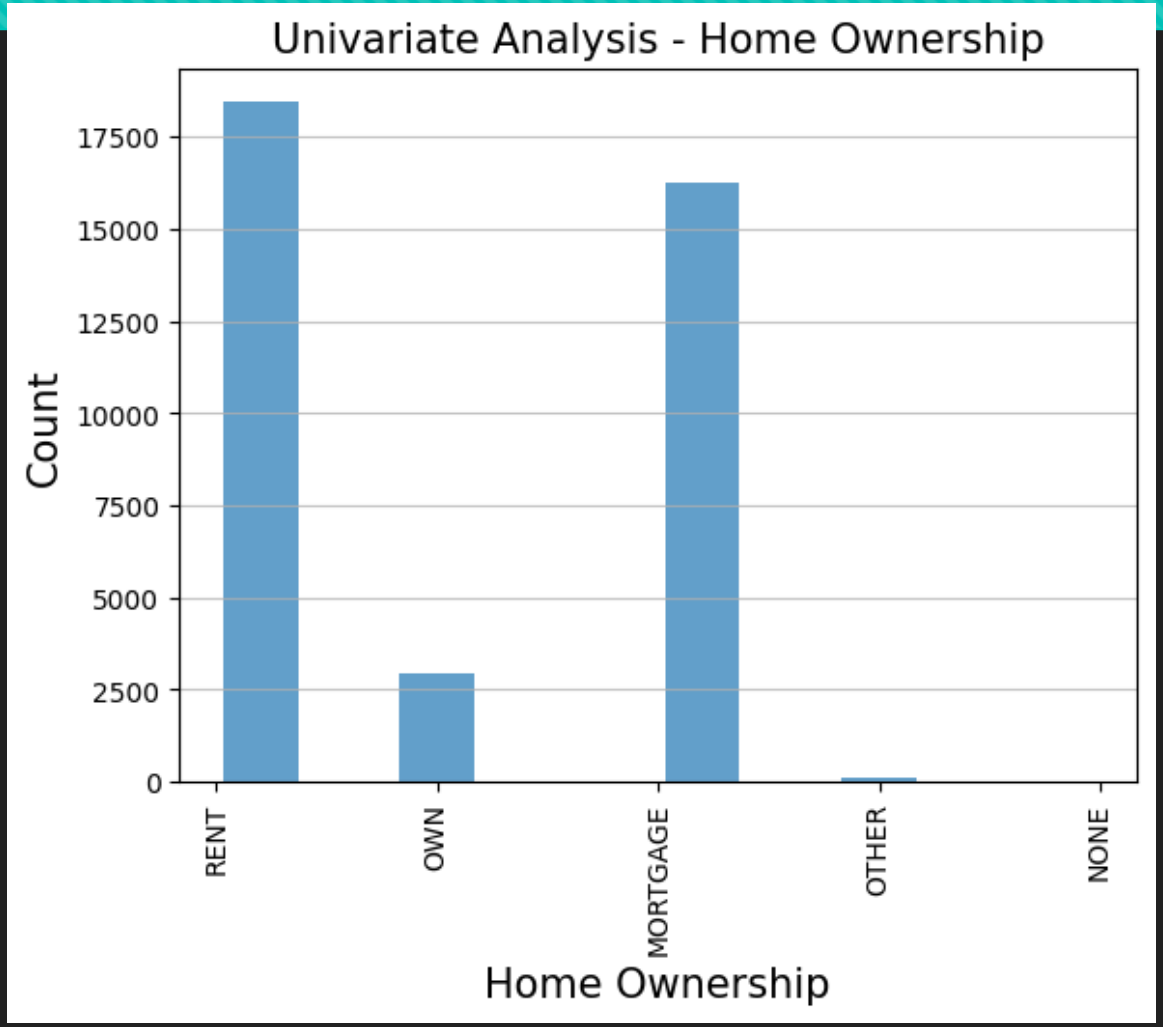
Loan Purpose

- Most number of loans are taken for either debt consolidation or credit card payment



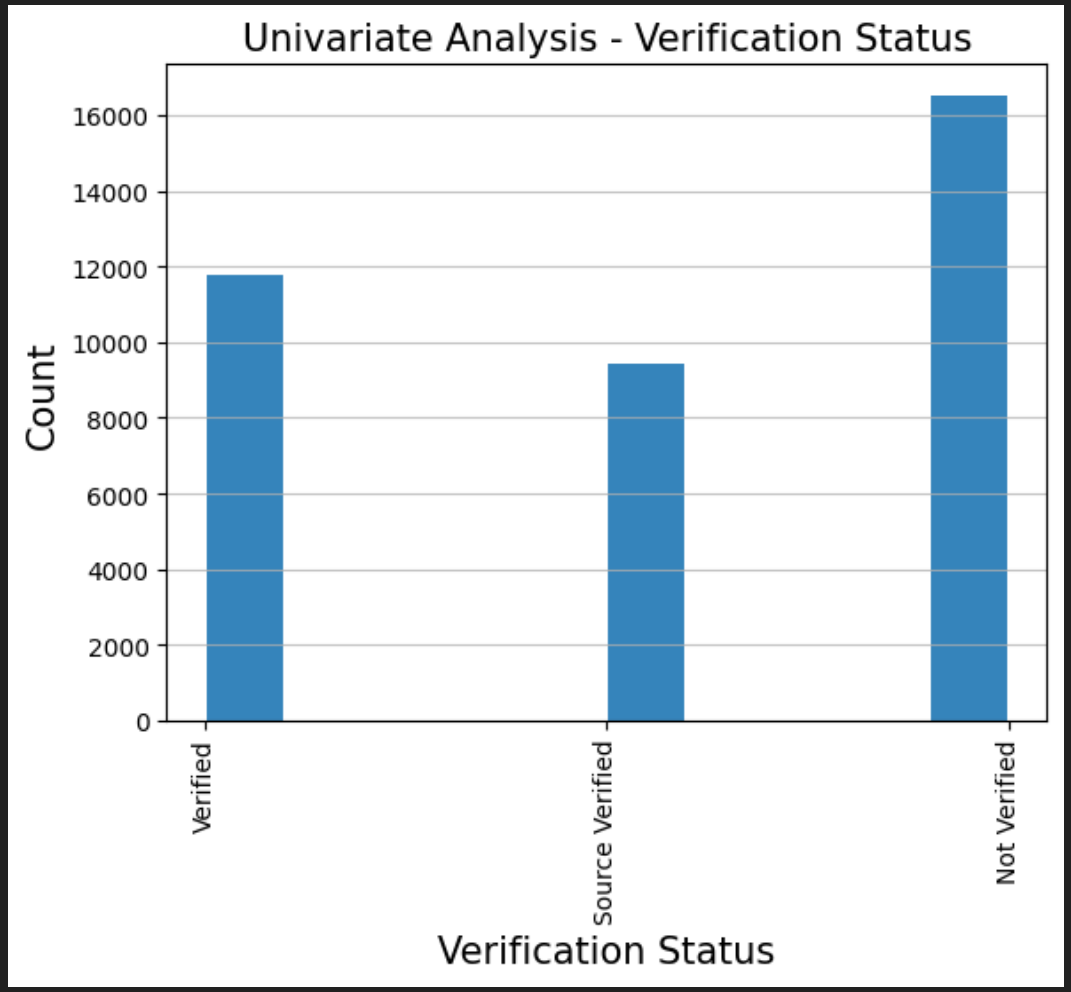
Home ownership

- People with home ownership status Rent or Mortgage are going for loans



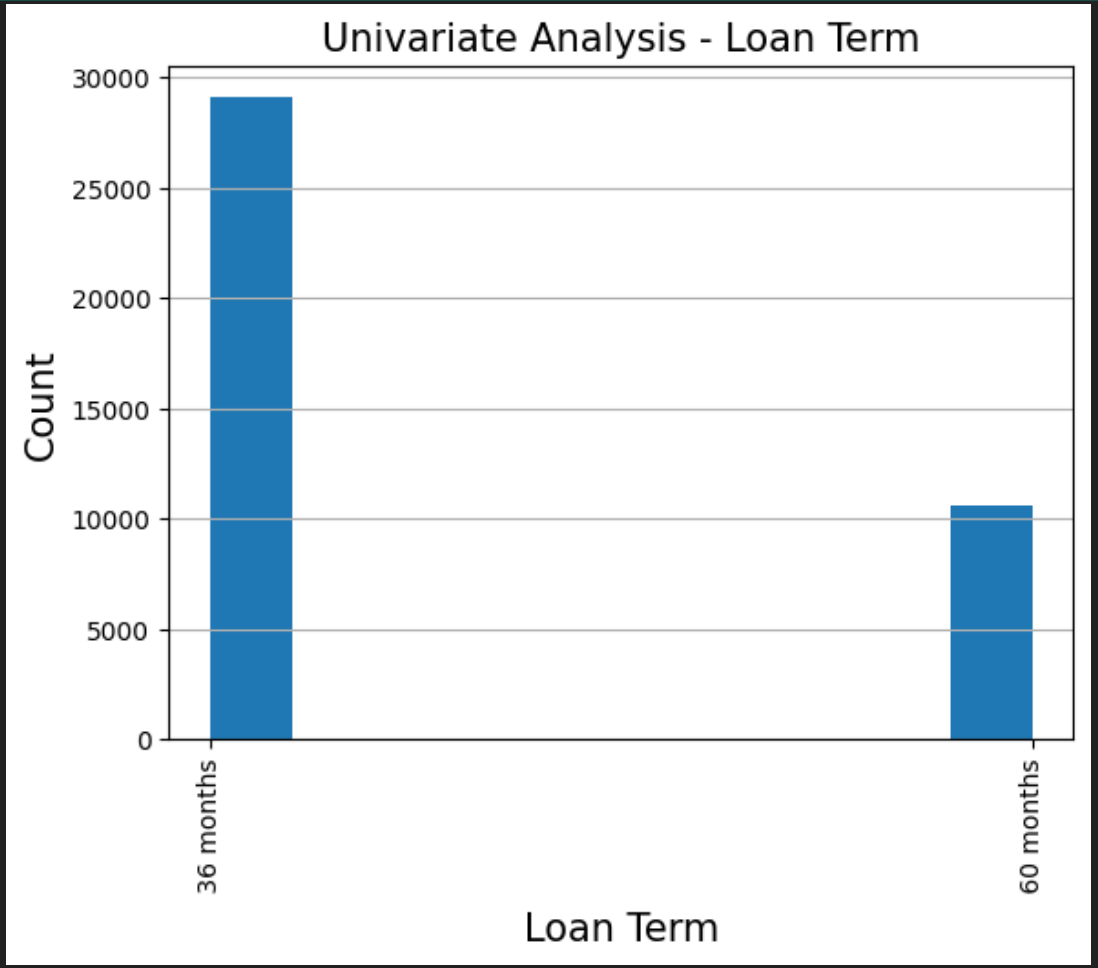
Verification status

- Dataset contains more people who are not verified



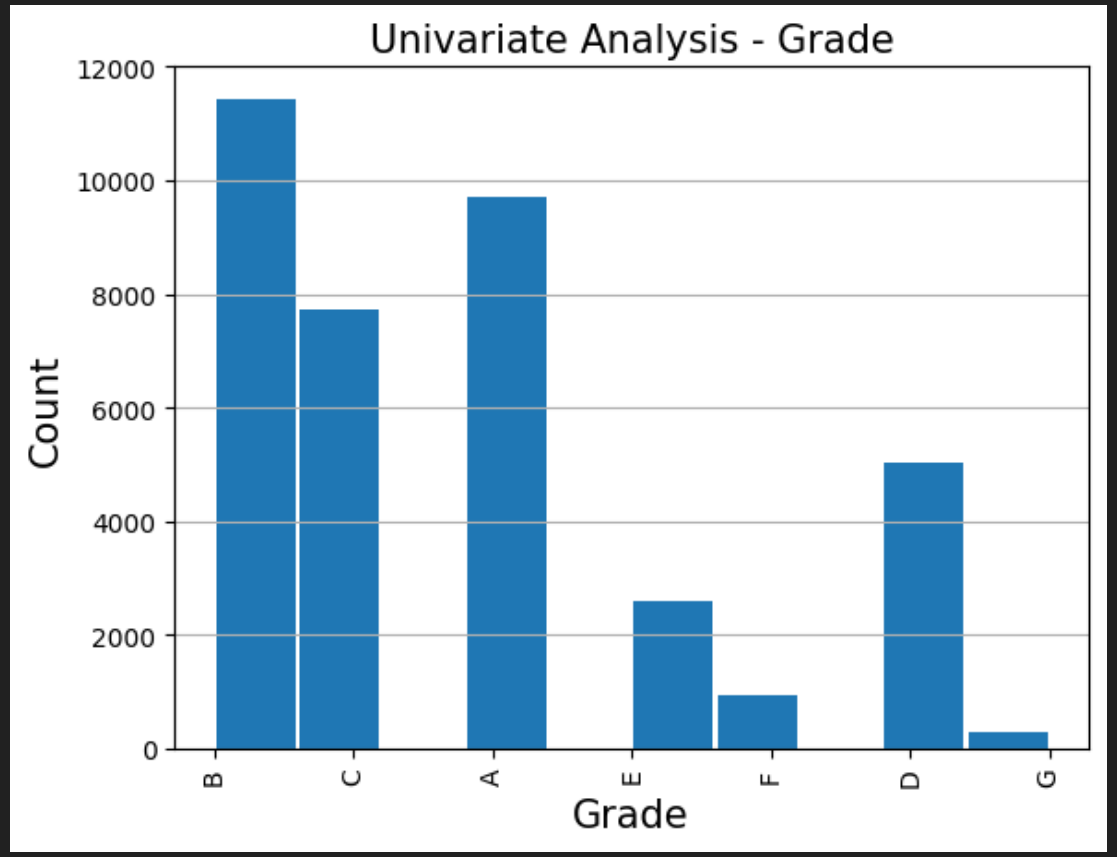
Loan term

- More people go with short term loans (36 months)



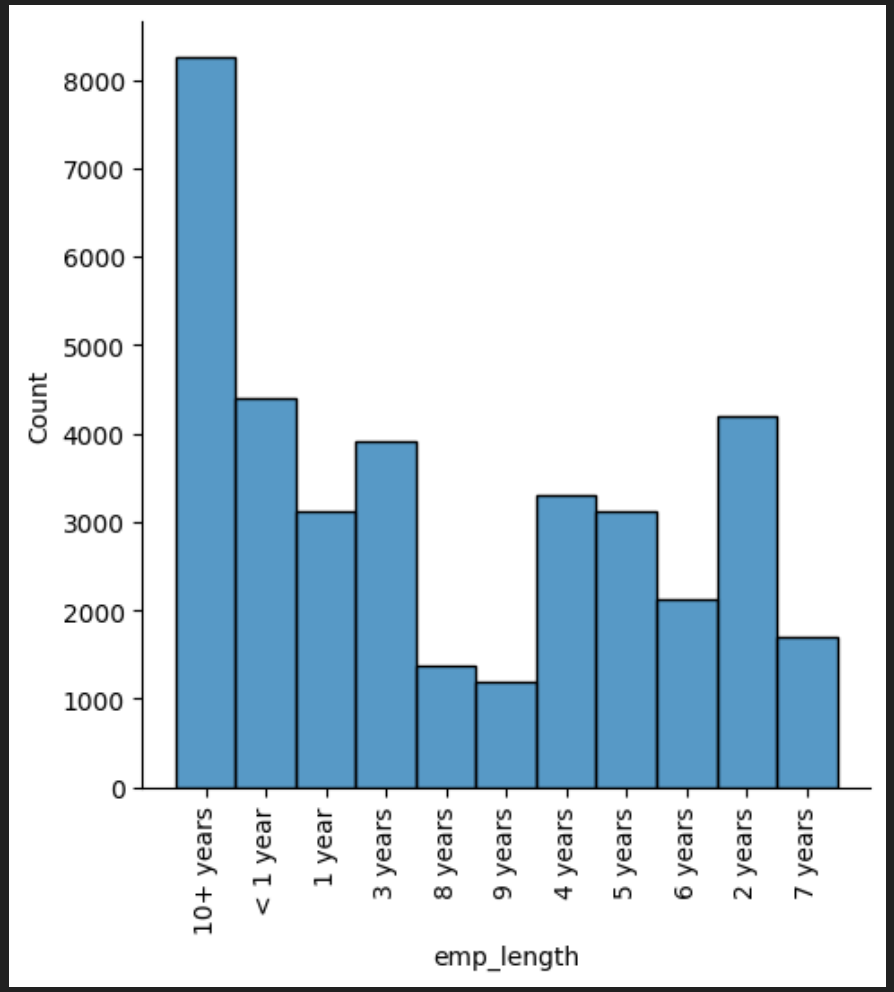
Grade

- Dataset has highest distribution of Grade A, B and C



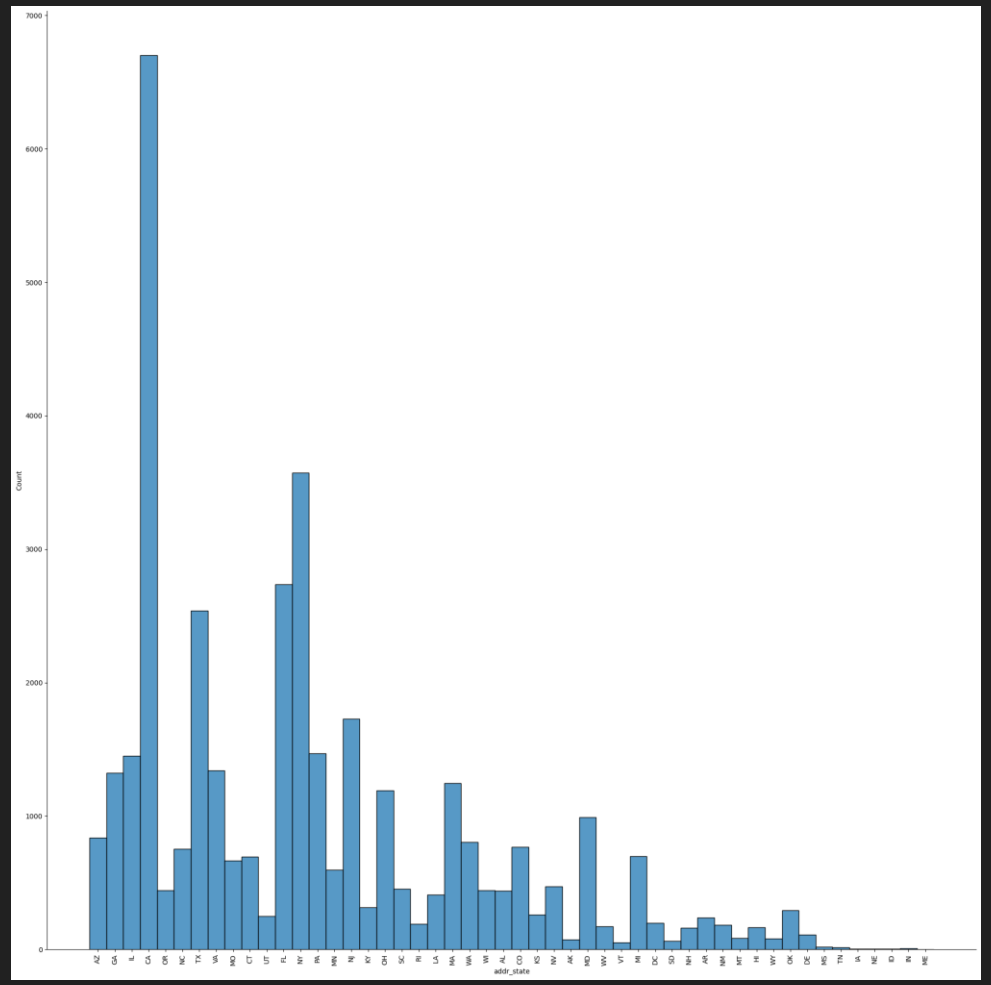
Employment Length

- Dataset contains more persons with 10+ years of employment length



Statewide distribution of loans

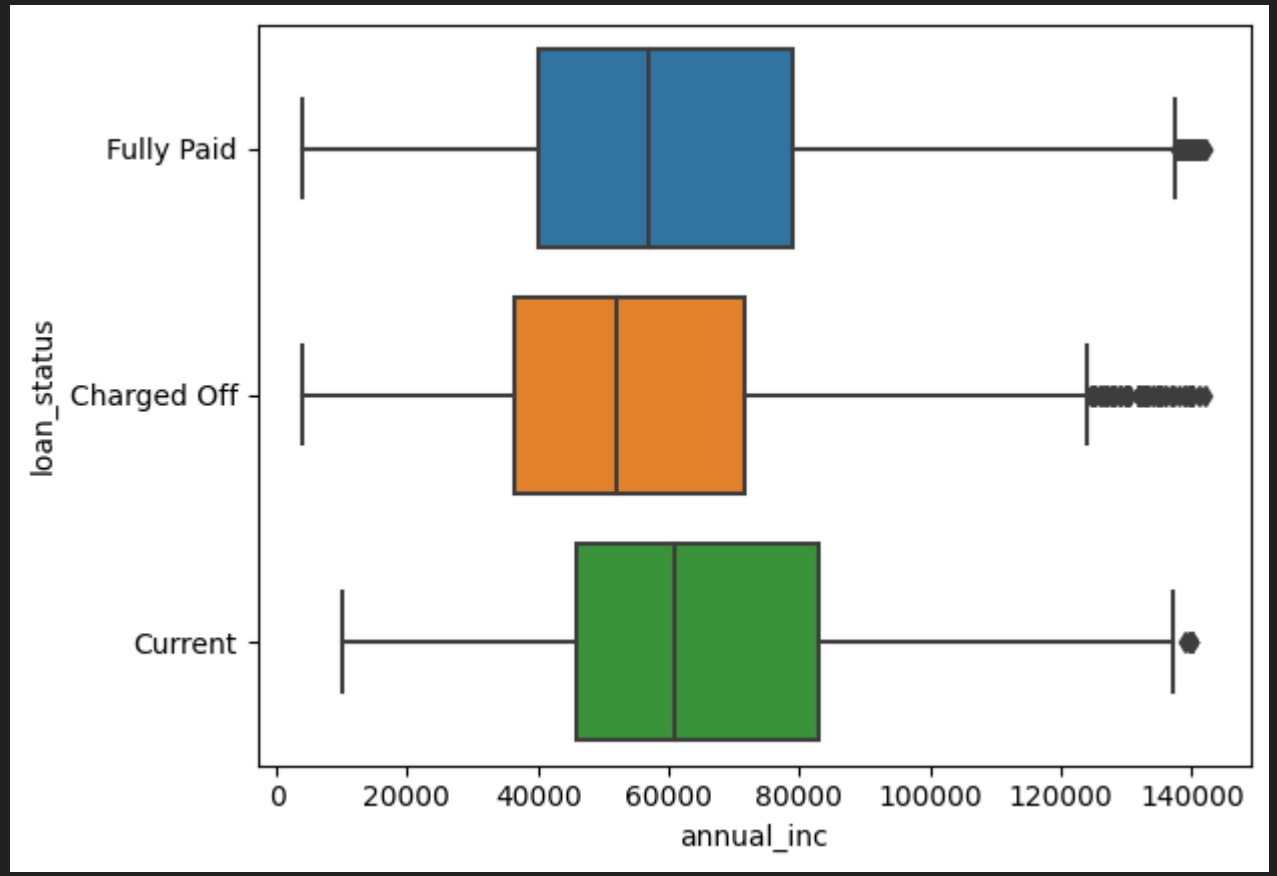
- California has more number of loans compared to other states



Segmented Univariate analysis

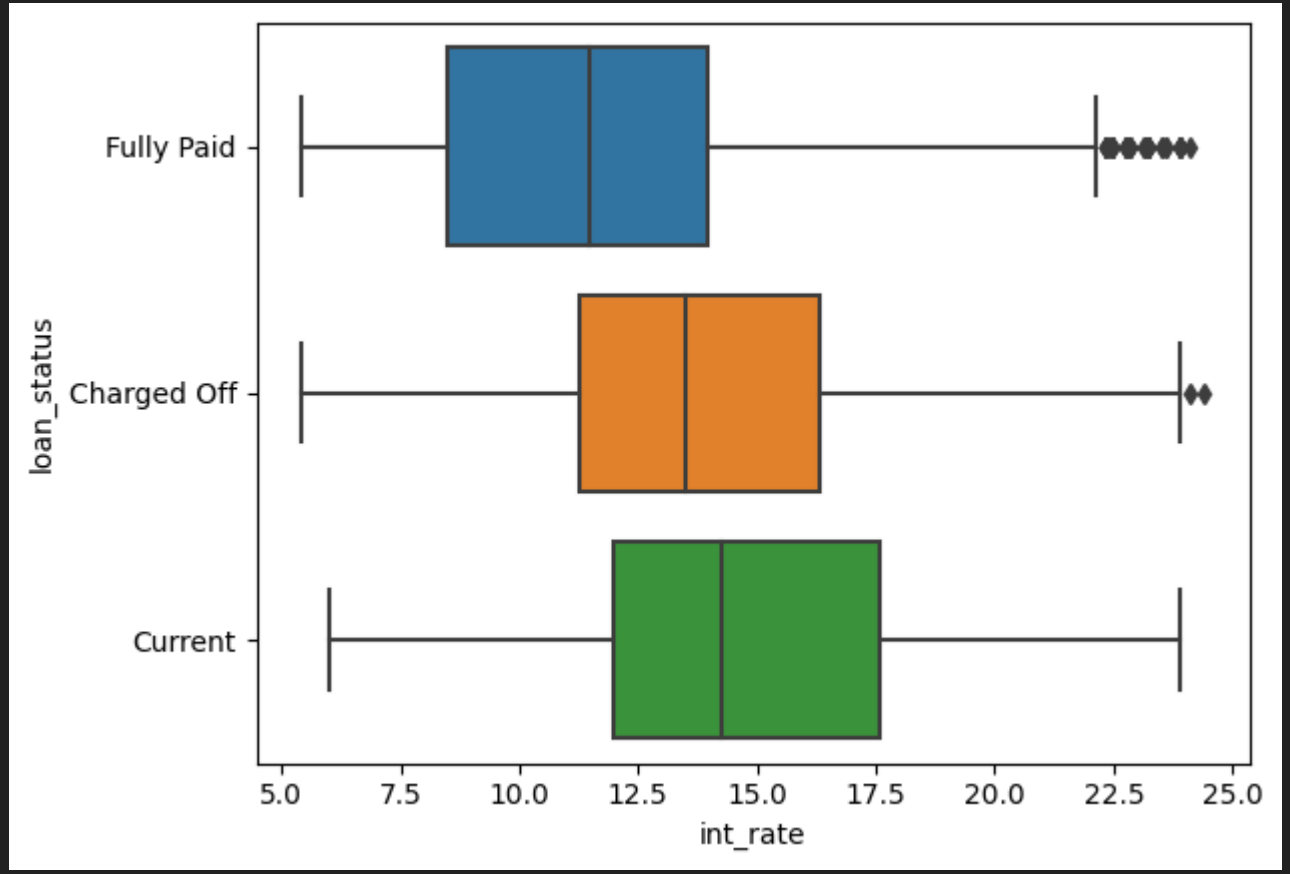
Annual Income group by Loan Status

- Median of annual income of persons who Charged off is relatively less compared to annual income of persons who Fully Paid or Current in payment



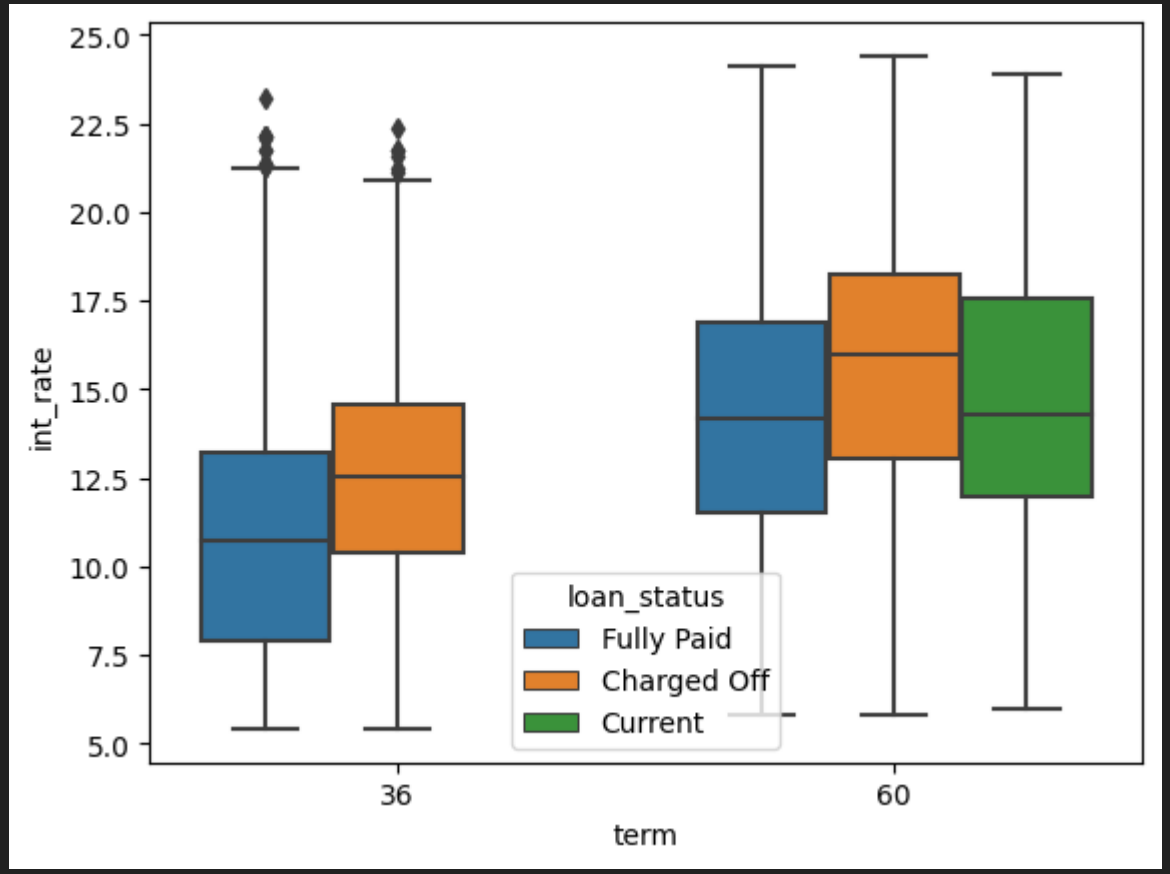
Interest Rate group by Loan Status

- Median of interest rates of charged off loans is couple of percentage points higher than fully paid loans



Interest rate variation with term

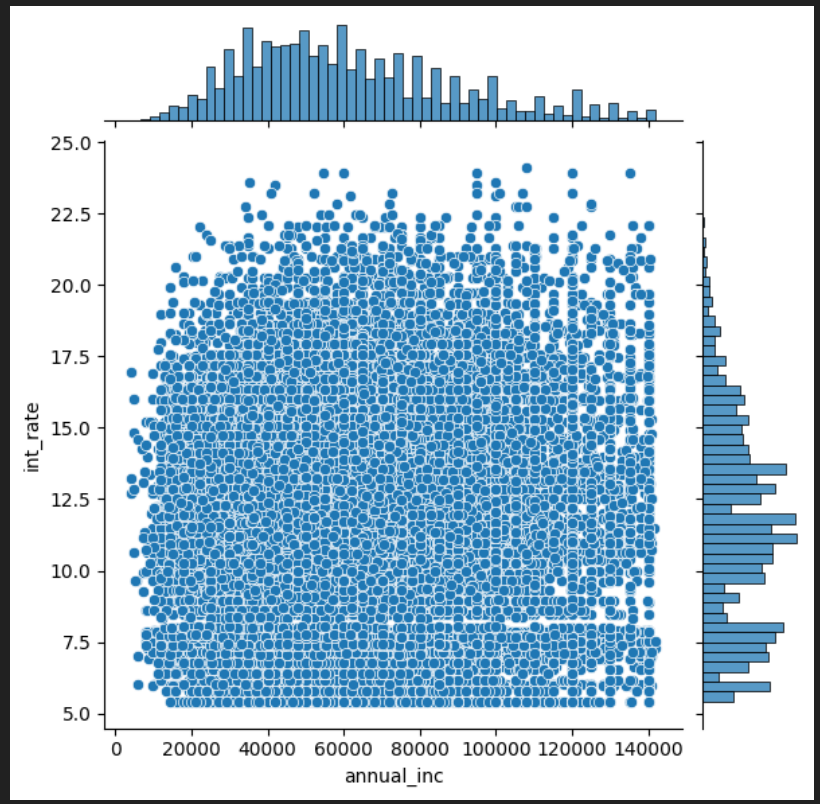
- For charged off loans with longer term has higher interest rate



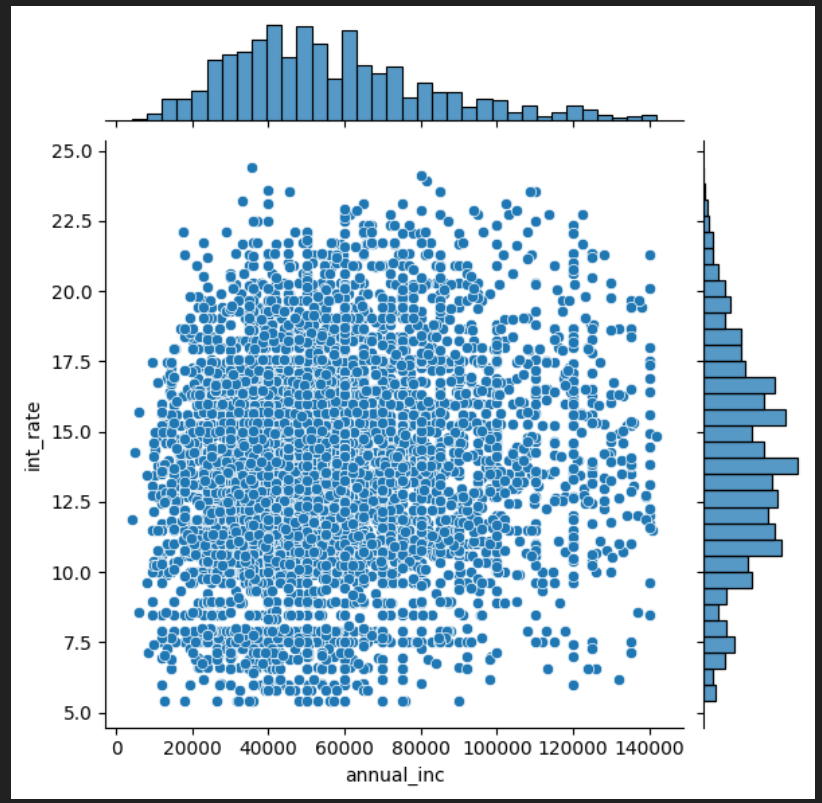
Bivariate Analysis

Bivariate analysis

- Charged off loans usually have higher interest rates and lowest annual income



Fully Paid

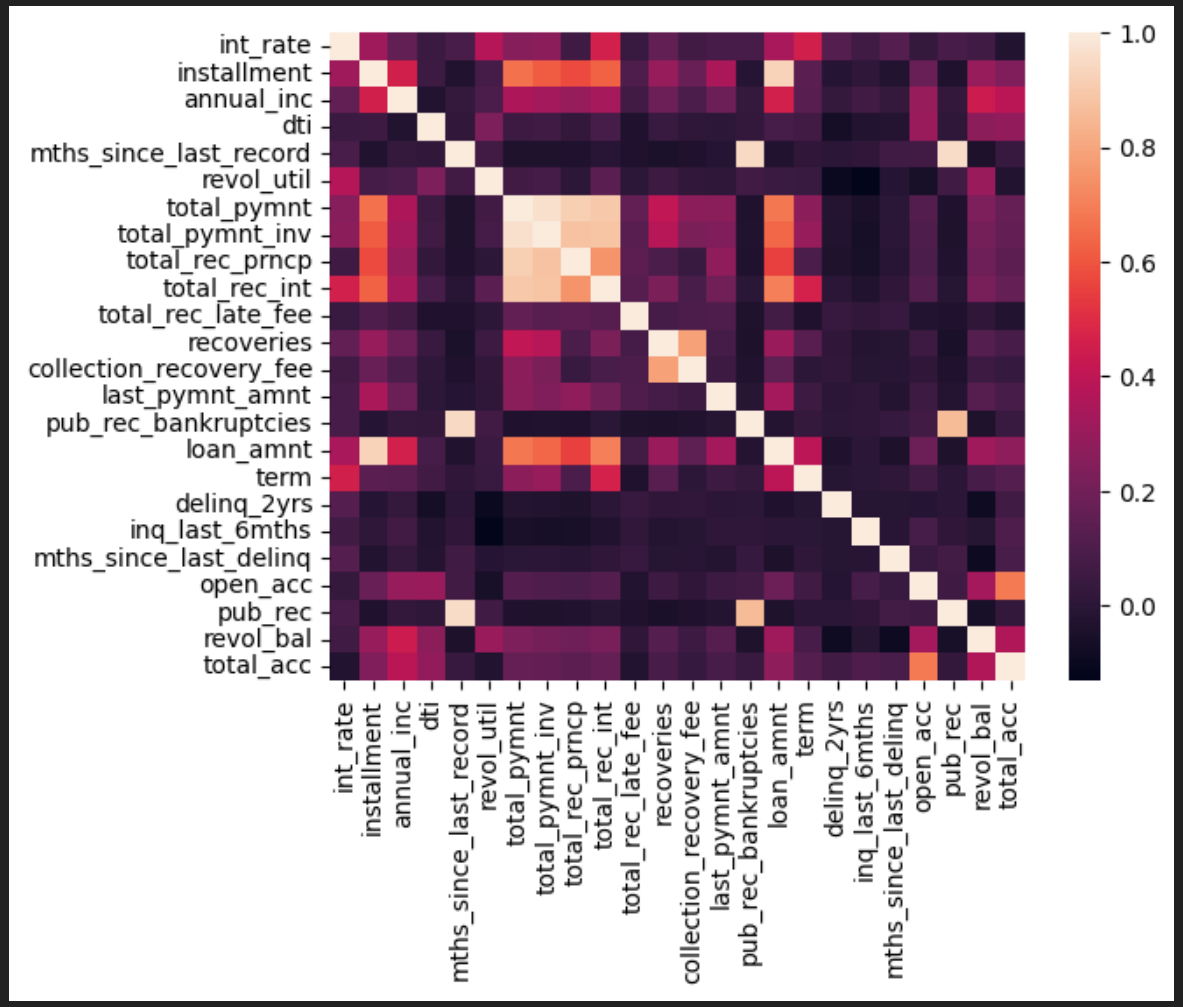


Charged Off

Mutivariate Analysis

Relationship between numerical columns

- Number of derogatory public records and Number of public record bankruptcies are correlated to borrowers delinquency and which is directly related to charge off



Insights

- Loan purpose
  - ***Small Business*** loans are more likely to be defaulted (28%) and should be one the driving variables during loan sanctions
  - Followed by Renewable Energy (19%) and Educational (17%)
- Number of inquiries in past 6 months
  - More the number of inquiries, more likely to be defaulted. Persons with 6 enquiries are ~30% likely to be defaulted and with 5 its ~25%
- Public derogatory records and public record bankruptcies
  - Both those features have relation to charge off
- Interest rate and term
  - Longer term with higher interest rate loans are more likely to be defaulted
- Revolving Utilization rate
  - Higher the revolving utilization rate more likely to default