LENDING CLUB CASE STUDY

Submitted By:

Ravi Shankar

Abhishek Ranjan

CONTENTS

- Problem Statement
- Data overview
- Understanding data
- Data cleaning
- Data Analysis
  - Univariate Analysis
  - Bivariate Analysis
  - Multi-variate
- Recommendations

PROBLEM STATEMENT

- Lending Club, a consumer finance marketplace that provides various loans to urban customers, faces a significant challenge in managing its loan approval process. When assessing loan applications, the company needs to make informed decisions to reduce financial losses, particularly from loans granted to applicants deemed "risky."
- The main objective of this exercise is to help Lending Club reduce credit losses. This challenge stems from two possible scenarios:
    - Identifying applicants who are likely to repay their loans is essential, as they can generate profits for the company through interest payments. Rejecting these applicants would mean missing out on potential business opportunities.
    - Conversely, approving loans for applicants who are unlikely to repay and pose a risk of default can lead to significant financial losses for the company.
- The company aims to identify the key factors (or driver variables) that contribute to loan defaults, specifically the variables that strongly indicate a likelihood of default. This understanding can be leveraged for better portfolio management and risk assessment.

DATA OVERVIEW

- Historical data provided by Lending Club.
- Contains borrower credit history and loan details.
- Size: Over 39,717 records and 111 columns.
- Data dictionary with descriptions

DATA UNDERSTANDING

- Target attribute: Loan Status
- Target attribute can have three distinct values
  - Fully Paid: This category includes applicants who have successfully repaid both the principal and interest of the loan.
  - Current: Applicants in this group are actively making loan installments, meaning the loan term is still ongoing. These individuals are not considered *defaulted*
  - Charged-off: This classification applies to applicants who have not made timely payments for an extended period, leading to a default on the loan.

Cont.

DATA UNDERSTANDING

- Customer attributes
  - Annual Income (annual_inc)
  - Home ownership (home_ownership)
  - Debt to Income (dti)
  - State (addr_state)
  - Employemnt year (emp_length)

- Loan attributes
  - Loan amount (loan_amnt)
  - Loan Term (term)
  - Interest rate (int_rate)
  - Loan date (issue_d)
  - Grade (grade)
  - Loan purpose (purpose)
  - Public record bankrupticies (pub_rec_bankruptcies)

DATA CLEANING

- Removal of columns having all null values
- Removal of columns which have single value
- Removal of columns which have post loan approval features
- Removal of columns which has string data like *emp_title*, *desc*, *title, url*
- Removal of entries which has loan status of *current* (as they don't help in analysis)
- Handling of missing values
  - Removal of entries which has missing values (~ 4.48% rows deleted)

Cont.

DATA CLEANING

- Modifying values
  - Removing string *months* from column *term*
  - Removal of char *%* from the interest rate colum *int_rate*
  - Modifying *emp_length* column by removing string *years and converting 10+ to 10 and converting < 1 to 0*
- Deriving columns *issue_month* & *issue_year* from loan issue column *issue_d*
- Removing outliers
  - Plotting boxplot for columns : *loan_amnt*, *int_rate* , *annual_inc*, *dti*
  - Use standard IQR technique to remove outlier
    - ❑ Only data falling between [Q1 – 1.5*IQR , Q3 + 1.5 * IQR] are kept for analysis

Cont.

DATA CLEANING

- Categorizing columns : Since some of the loan attribute columns like *loan_amnt*, *int_rate* are continuous numerical values , hence we can categorize this values into bucket (in new columns) for better analysis
  - Loan amount (*loan_amnt*): [*0-5k , 5k - 10k, 10k - 15k , 15k - above*] , column: loan_amnt_bucket
  - Loan interest rate (*int_rate*) : [*Very low, Low, Moderate , High, Very high*] , column : int_rate_bucket
  - Annual Income (*annual_inc*) : [*0-40k , 40k -50k, 50k - 60k, 60k - 70k, 70k - 80k, 80k - above*] , column: annual_inc_bucket
  - Debt to income (*dti)* : [*Very low, Low, Moderate , High, Very high*], column : dti_bucket

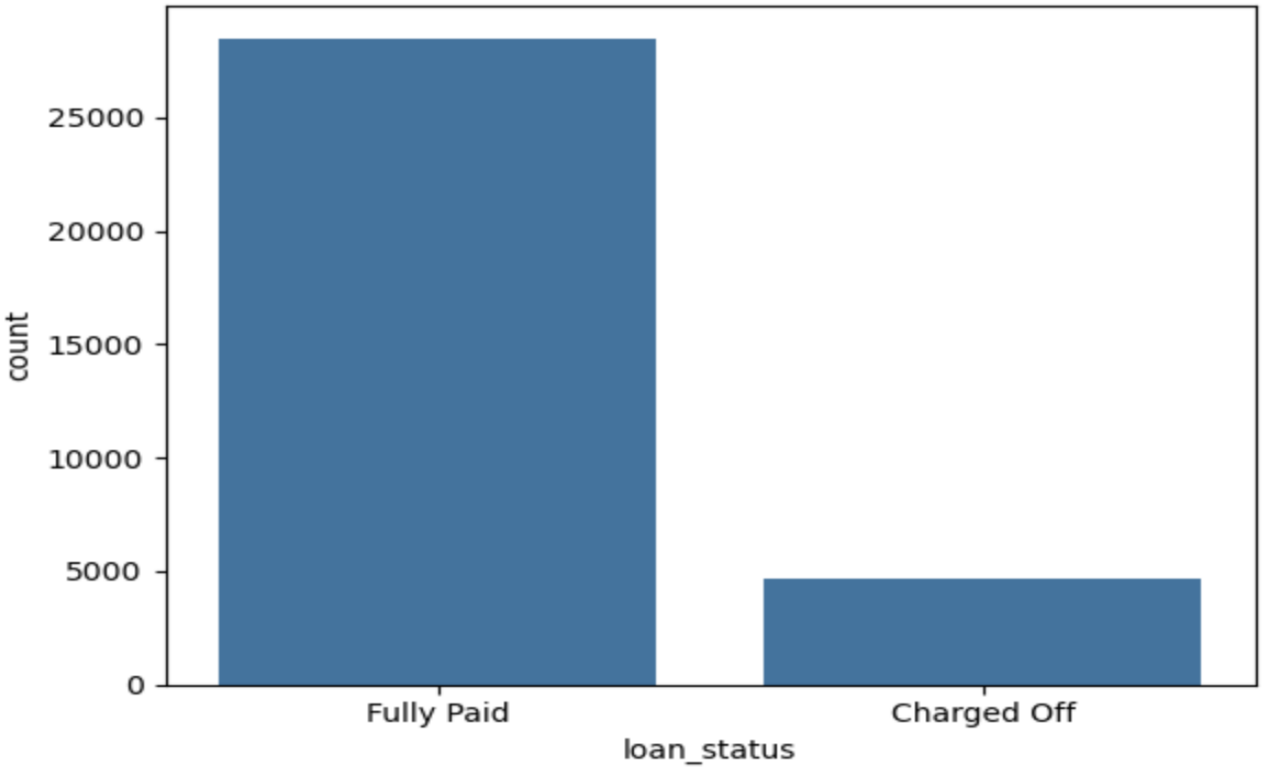
DATA ANALYSIS

- # of interested columns = 25
- Date-time : [*issue_d , issue_month, issue_year*]
- Numerical : [loan_amnt, funded_amnt, funded_amnt_inv, term, int_rate , installment, annual_inc, dti]
- Categorical : [*grade, sub_grade, emp_length, home_ownership, verification_status, loan_status, purpose, addr_state, pub_rec, pub_rec_bankruptcies, loan_amnt_bucket, int_rate_bucket, annual_inc_bucket, dti_bucket*]
- Target variable : [*loan_status*]

UNIVARIATE ANALYSIS

- Analysis of Loan status (loan_status)

Loan Status	
Fully Paid	28494
Charged Off	4699

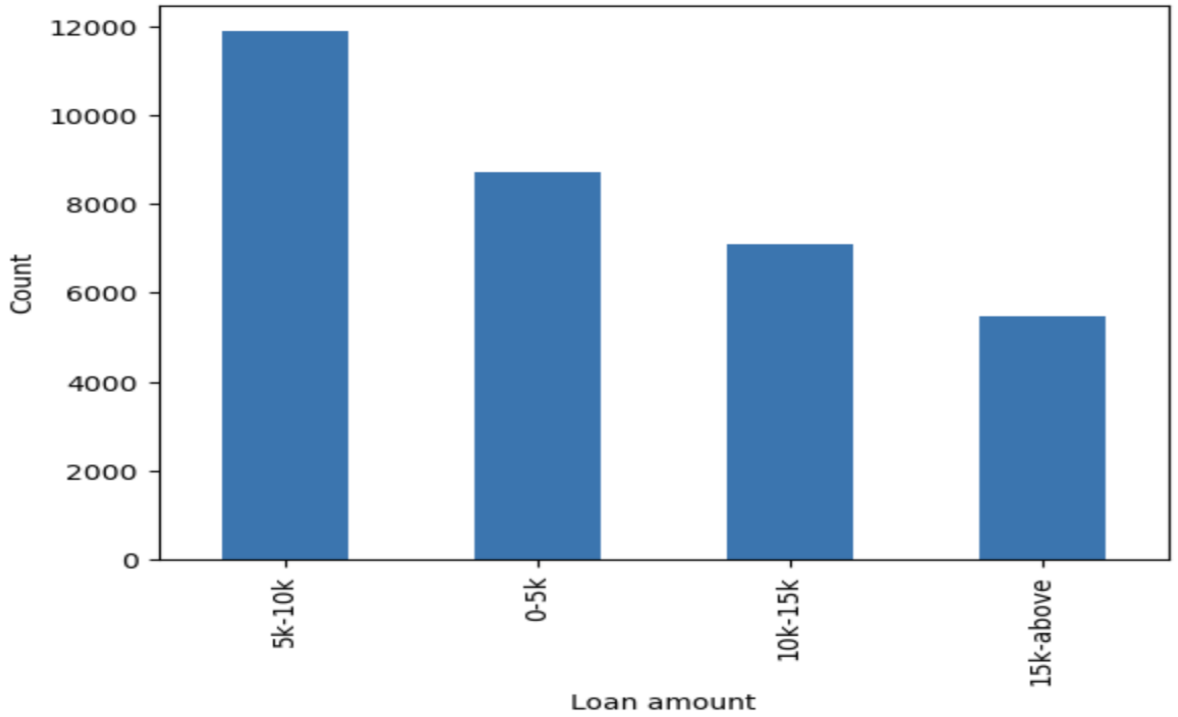


Observation : **14.16%** of loan in the data set has defaulted

Cont.

UNIVARIATE ANALYSIS

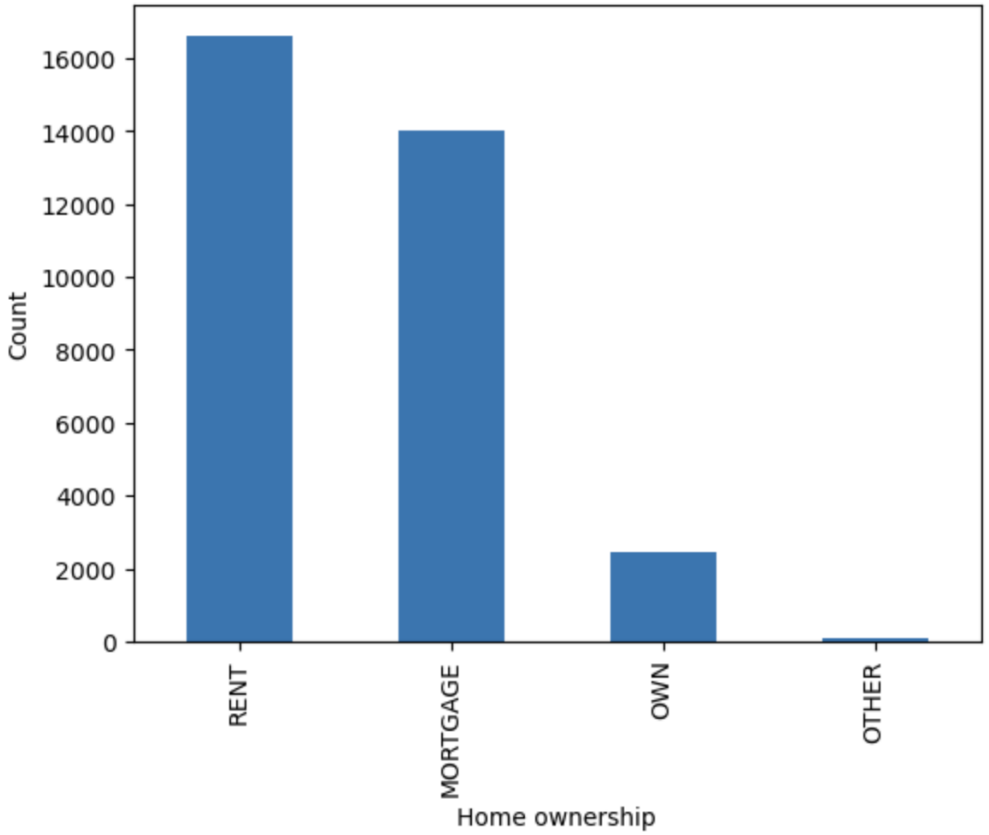
- Analysis of Loan amount (*loan_amnt_bucket*)



Observation : Most of the loan application is for amount between 5000 -10000

UNIVARIATE ANALYSIS

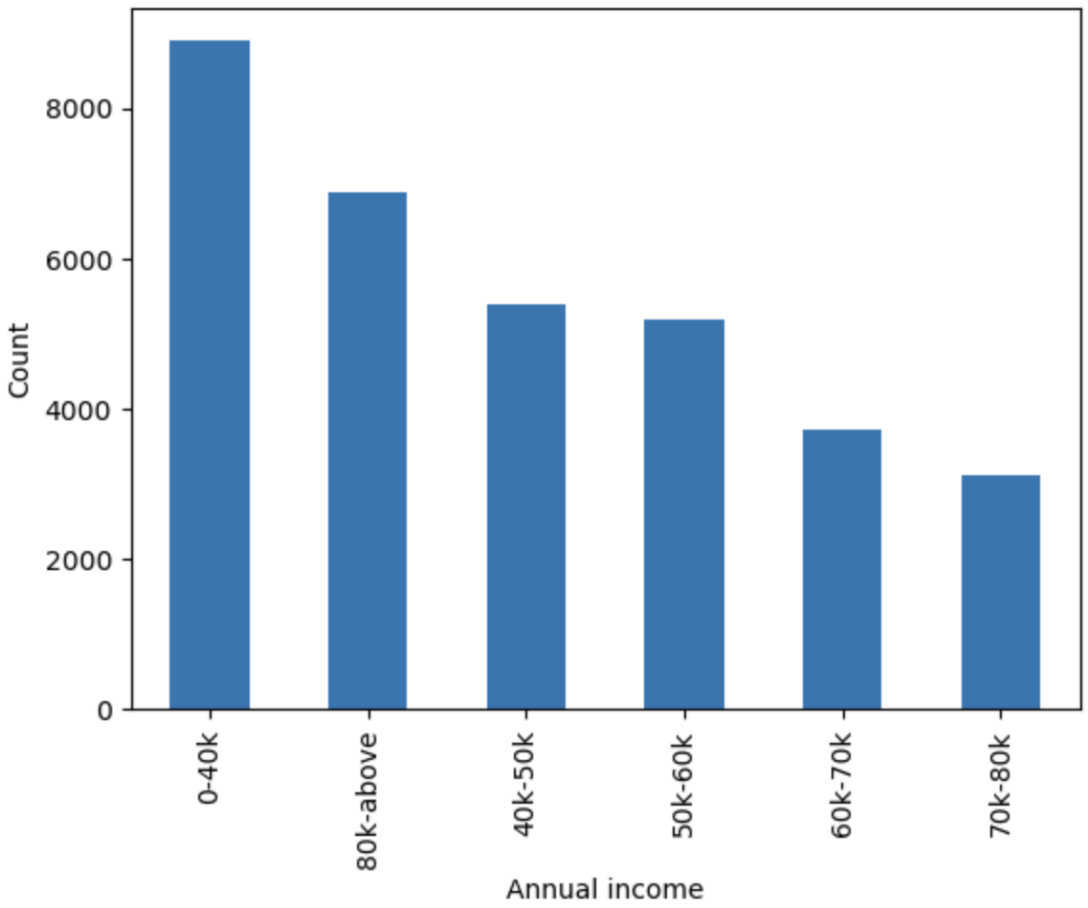
- Analysis of home ownership (*home_ownership*)



Observation : Majority of the loan applicants are ether living on *rent* or have house on *mortgage*

UNIVARIATE ANALYSIS

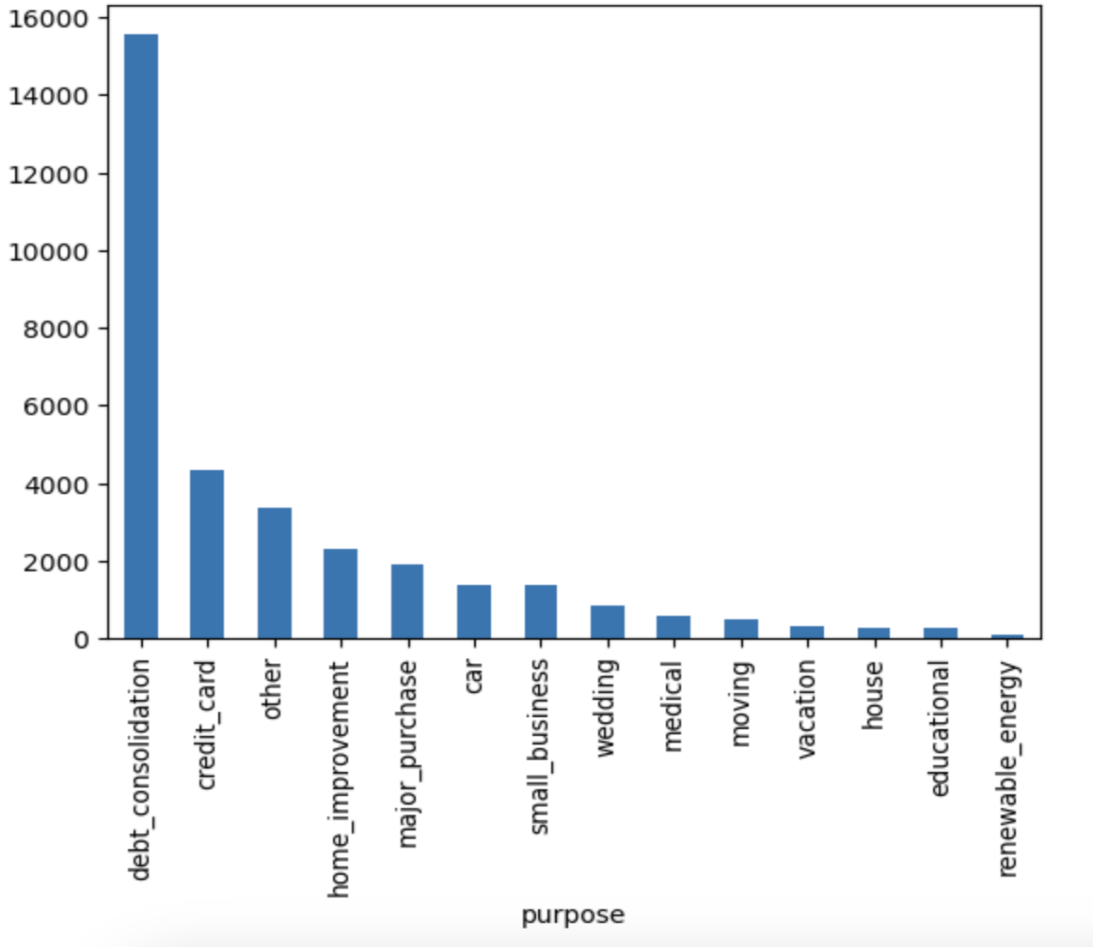
- Analysis of applicant annual income



Observation : Majority of the applicants has income between 0-40k

UNIVARIATE ANALYSIS

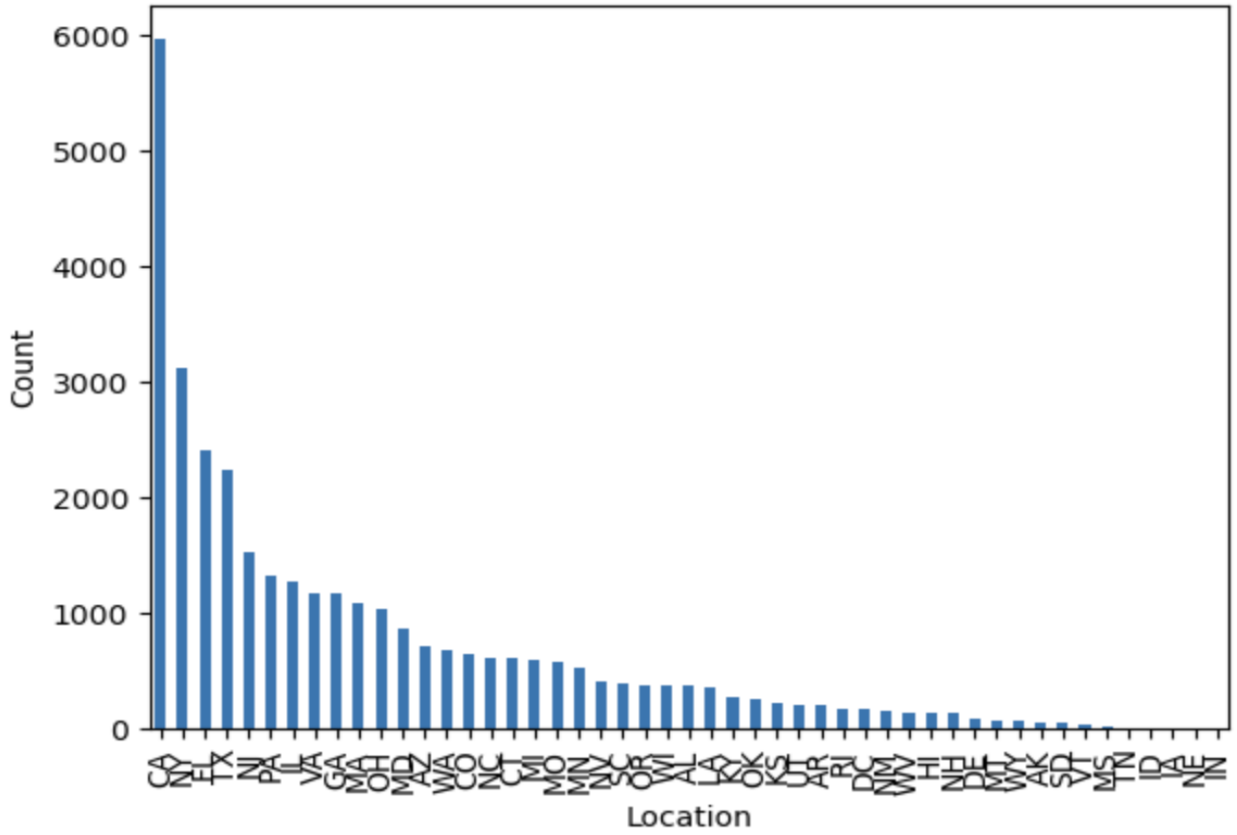
- Analysis of loan purpose (*purpose*)



Observation : Most of the applicants apply loan for the *debt consolidation* followed by *settlement of credit card*

UNIVARIATE ANALYSIS

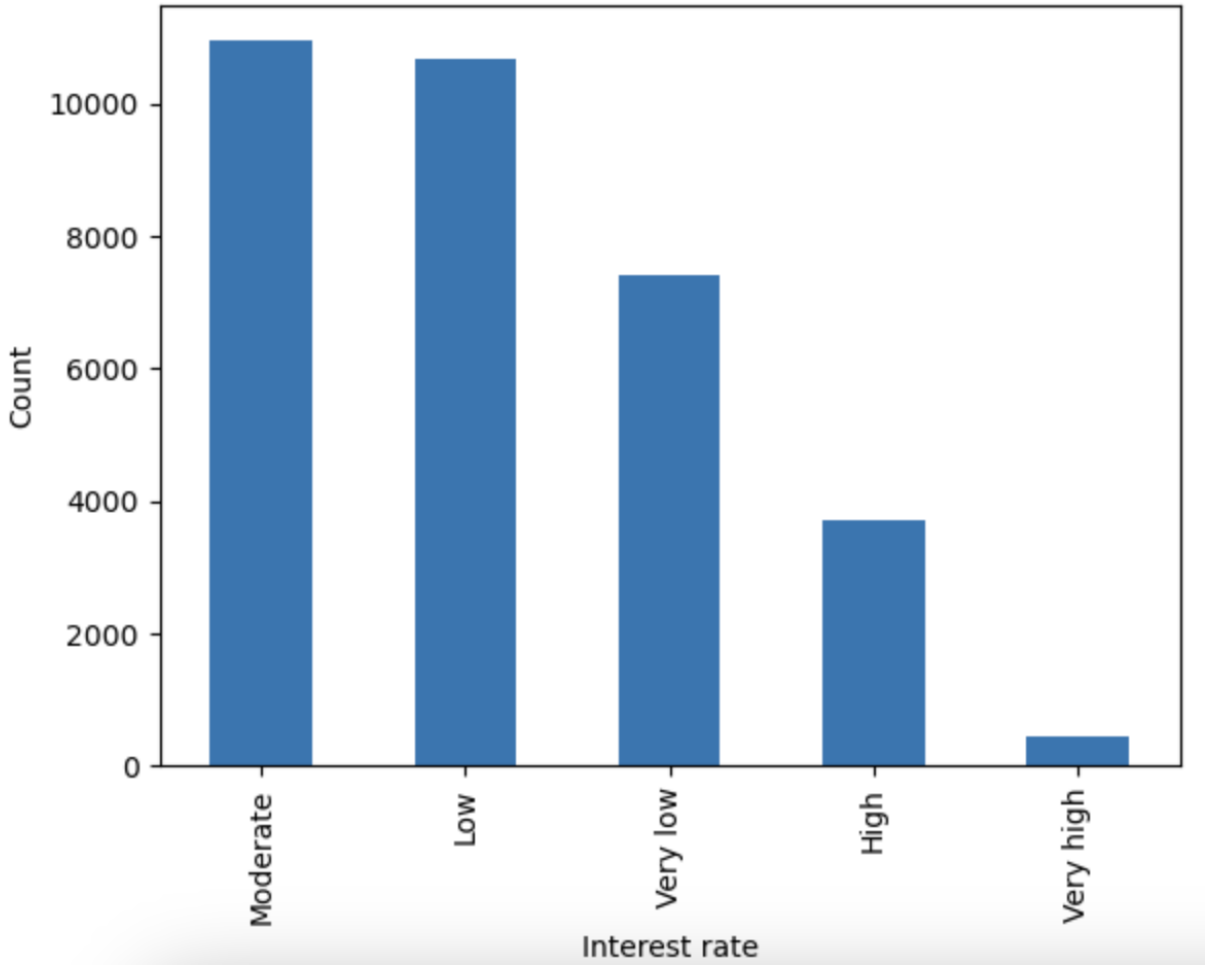
- Analysis of location (*addr_state*)



Observation : State CA has most of the applicants

UNIVARIATE ANALYSIS

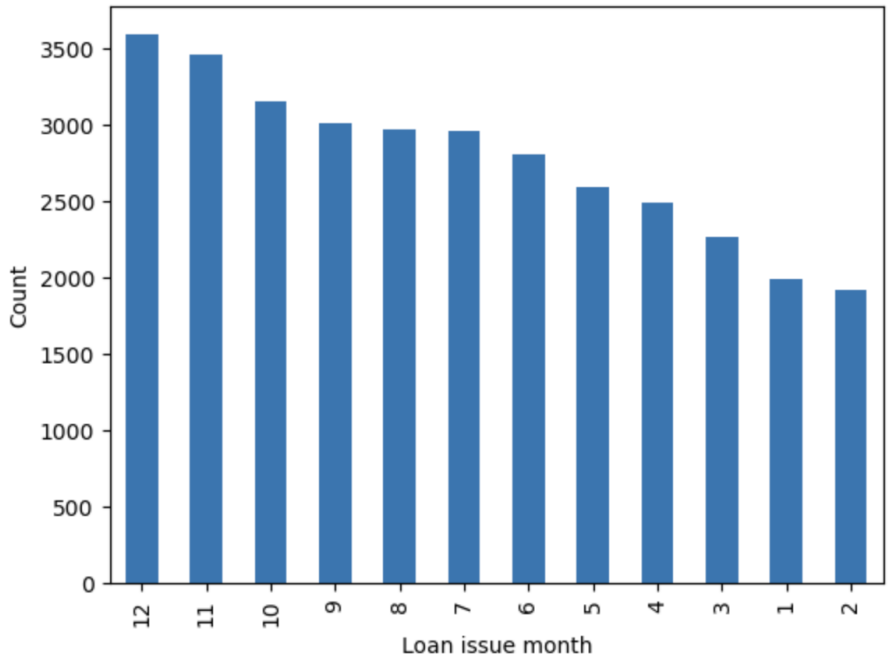
- Analysis of interest rate (*int_rate*)

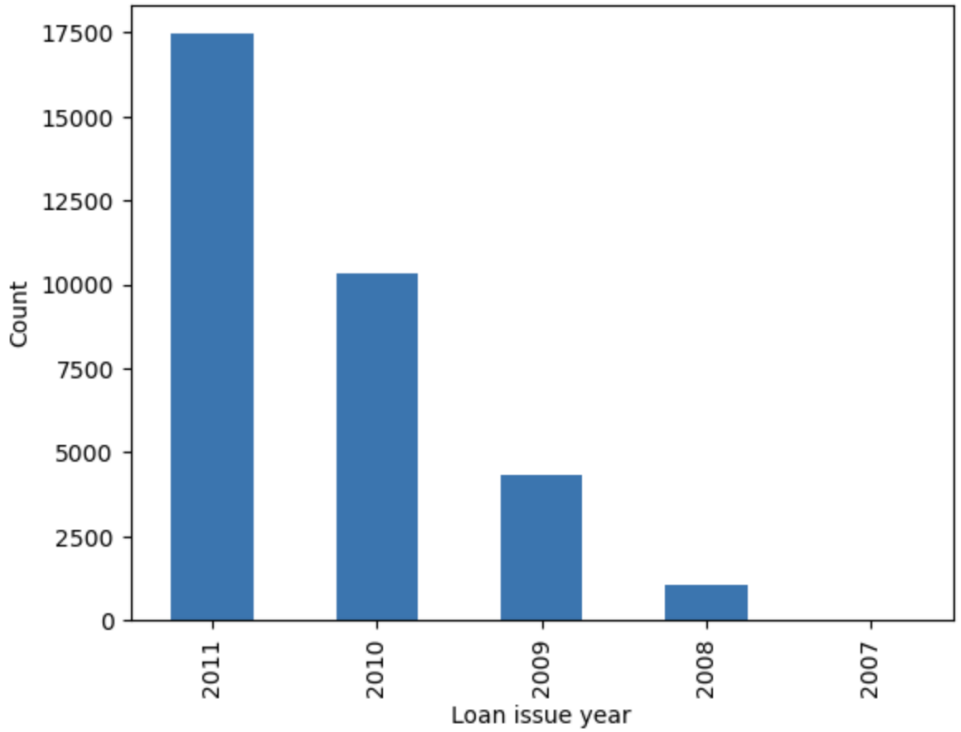


Observation : Most of the applicants have either low (<8%) or moderate(12% to 16%) interest rate.¶

UNIVARIATE ANALYSIS

- Analysis of month & year

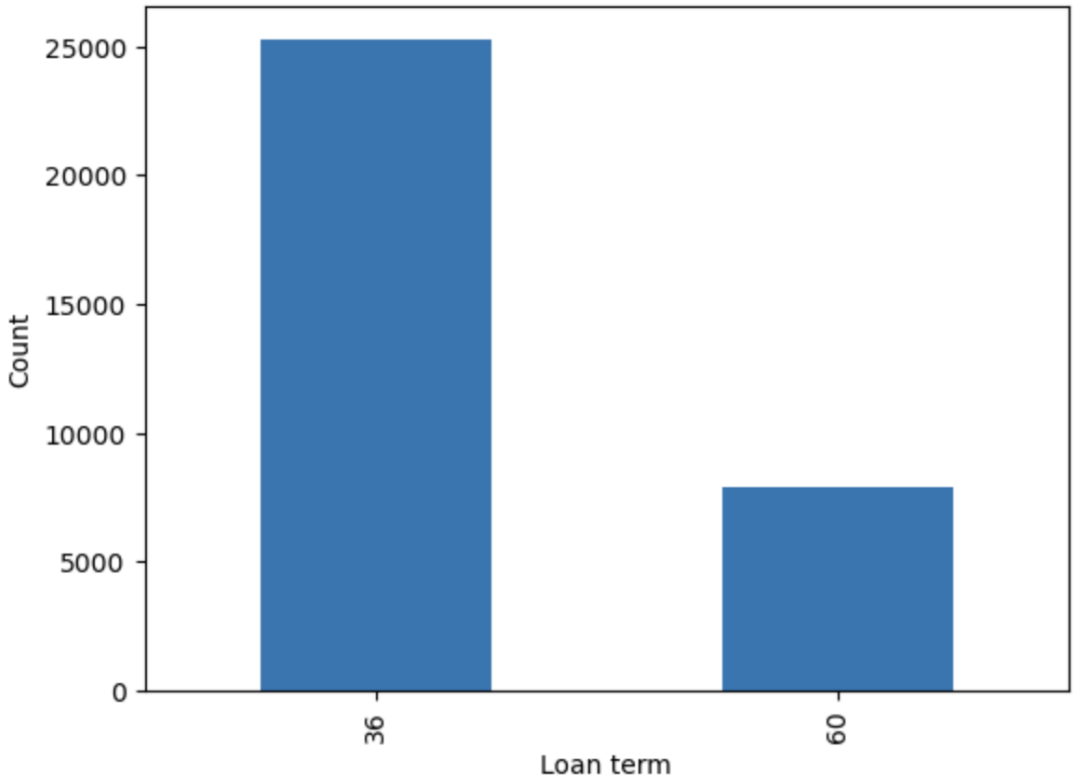




Observation : In year 2011 and month of dec the loan application is highest

UNIVARIATE ANALYSIS

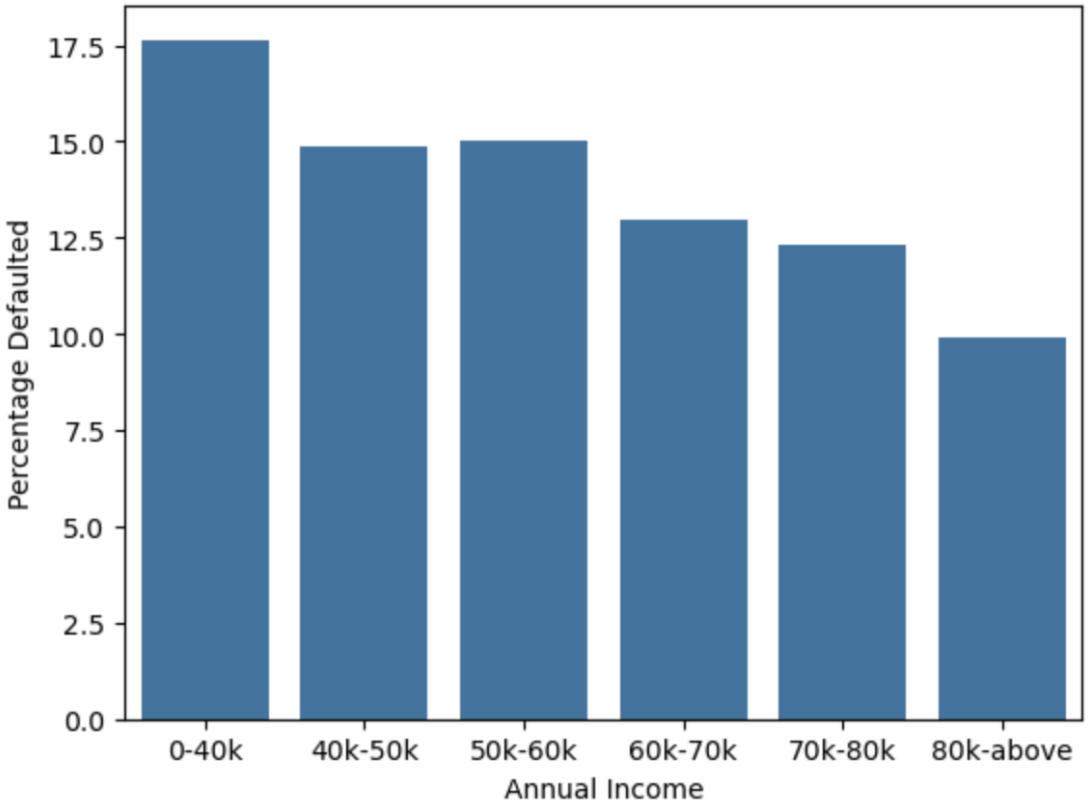
- Analysis of loan term



Observation : Most of the applicants took loan for 36 months term

BIVARIATE ANALYSIS

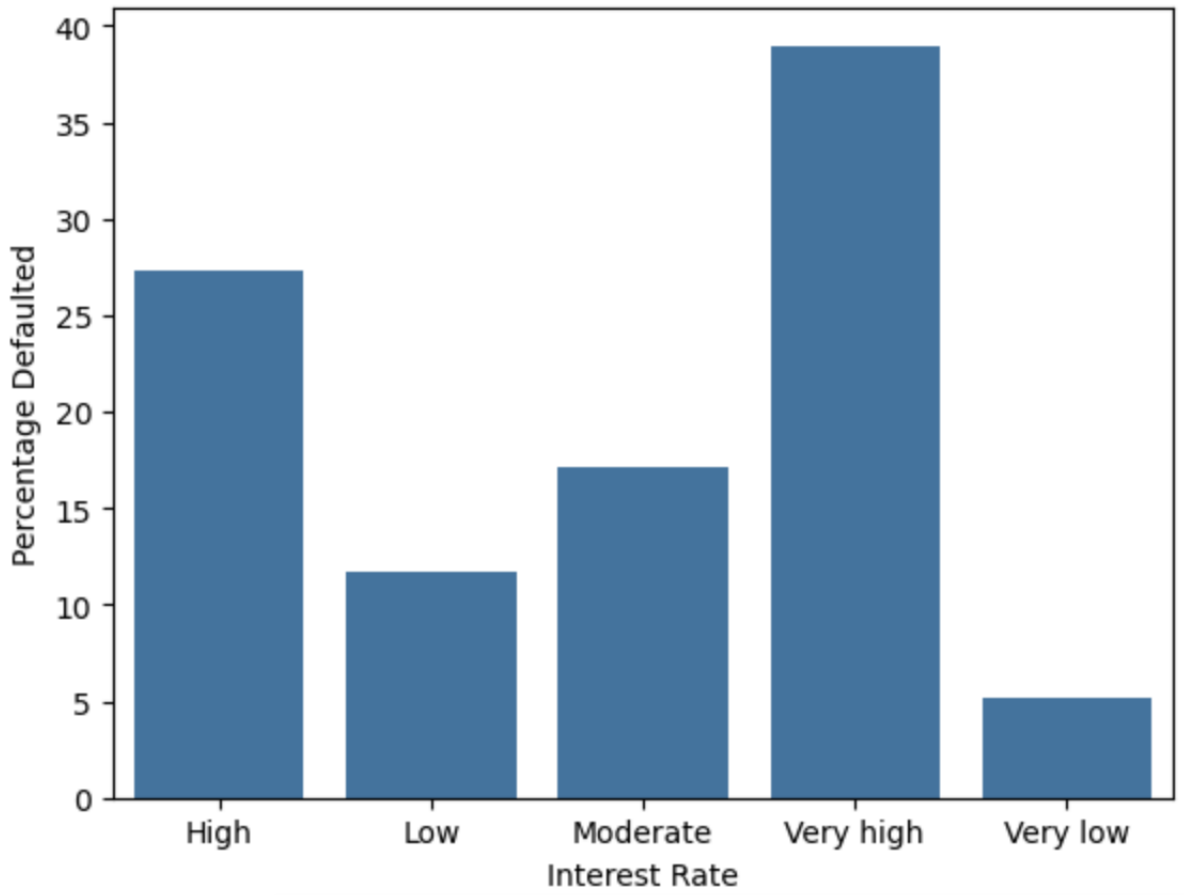
- Analysis of *annual_inc* vs *loan status*



Observation : Applicant with annual income in range 0-40k are most defaulted one¶

BIVARIATE ANALYSIS

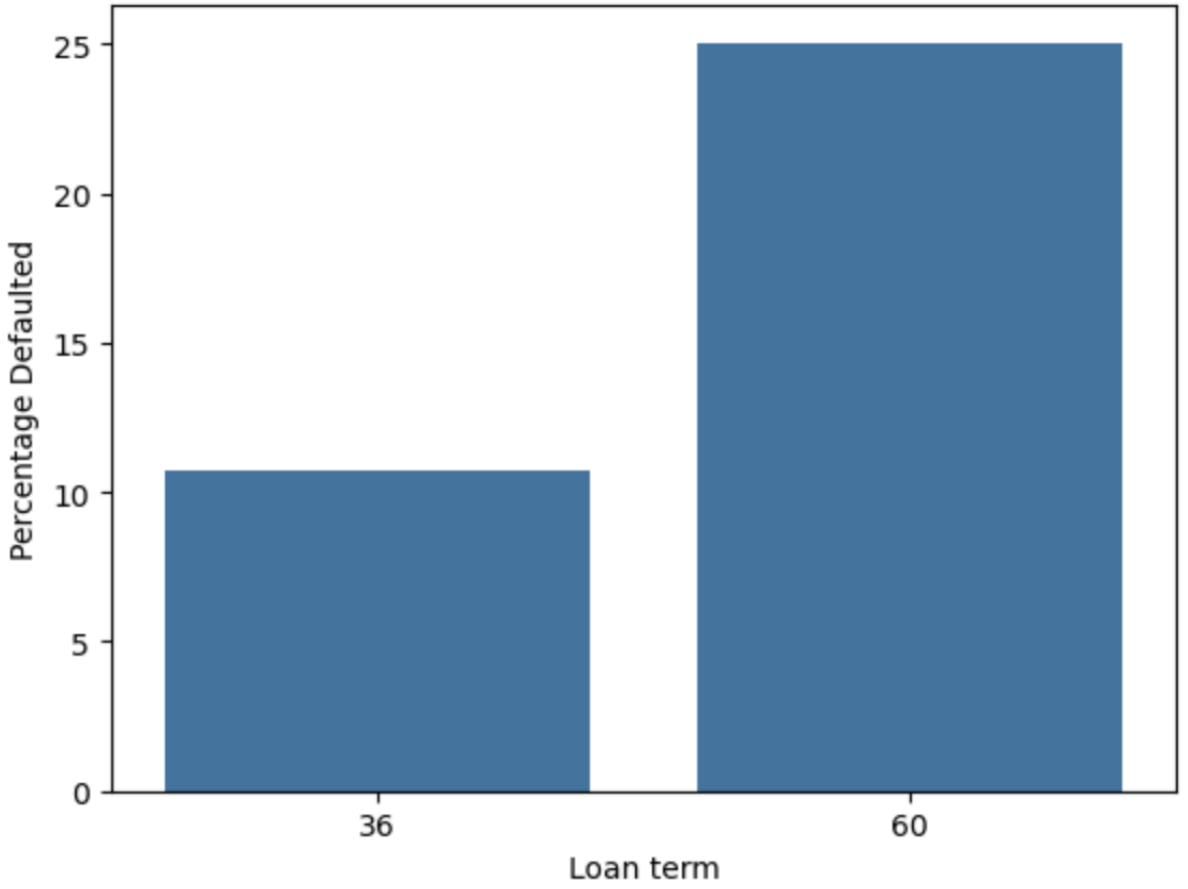
- Analysis of *int_rate* vs *loan status*



Observation : Applicants whose loan interest rate is more than 20%v are likely to get default

BIVARIATE ANALYSIS

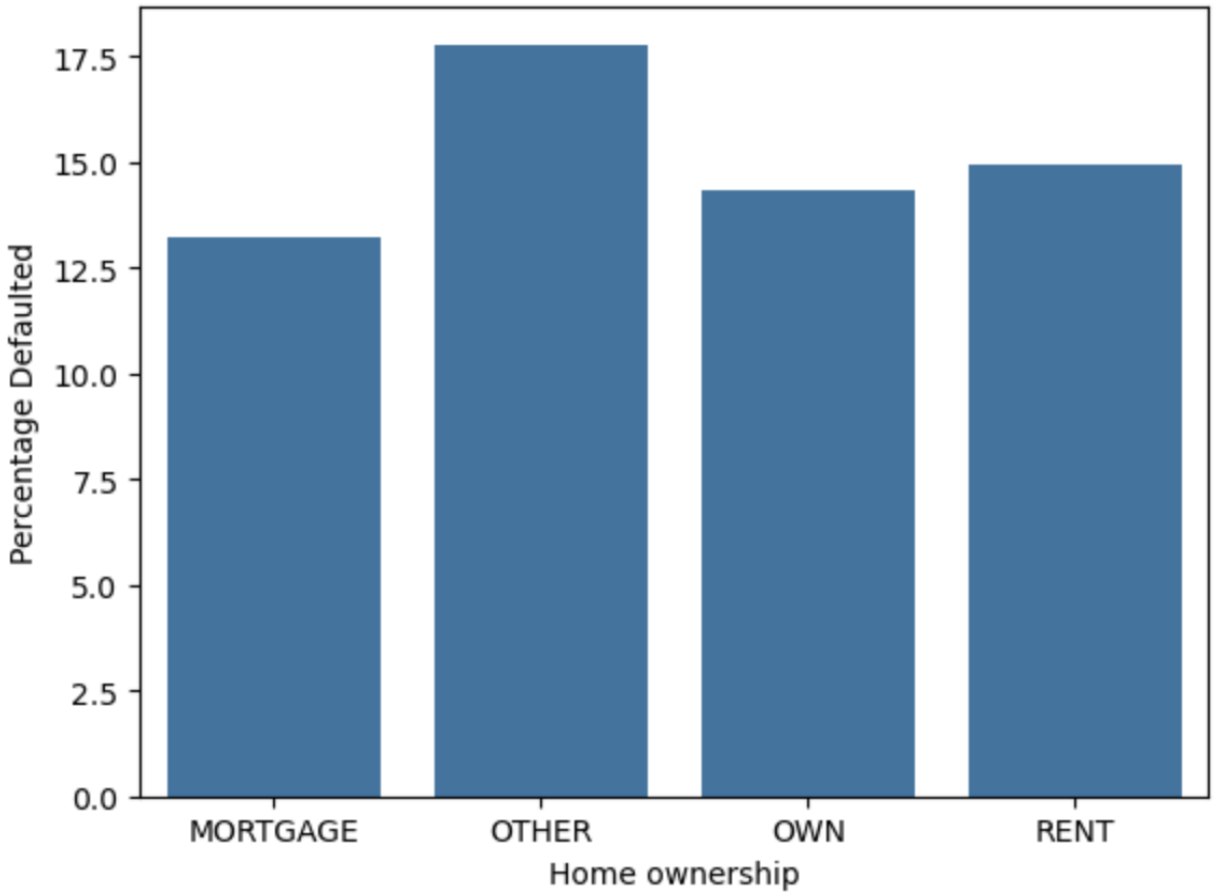
- Analysis of l*oan term* vs *loan status*



Observation : Applicants whose loan term is 60 months are likely to get default

BIVARIATE ANALYSIS

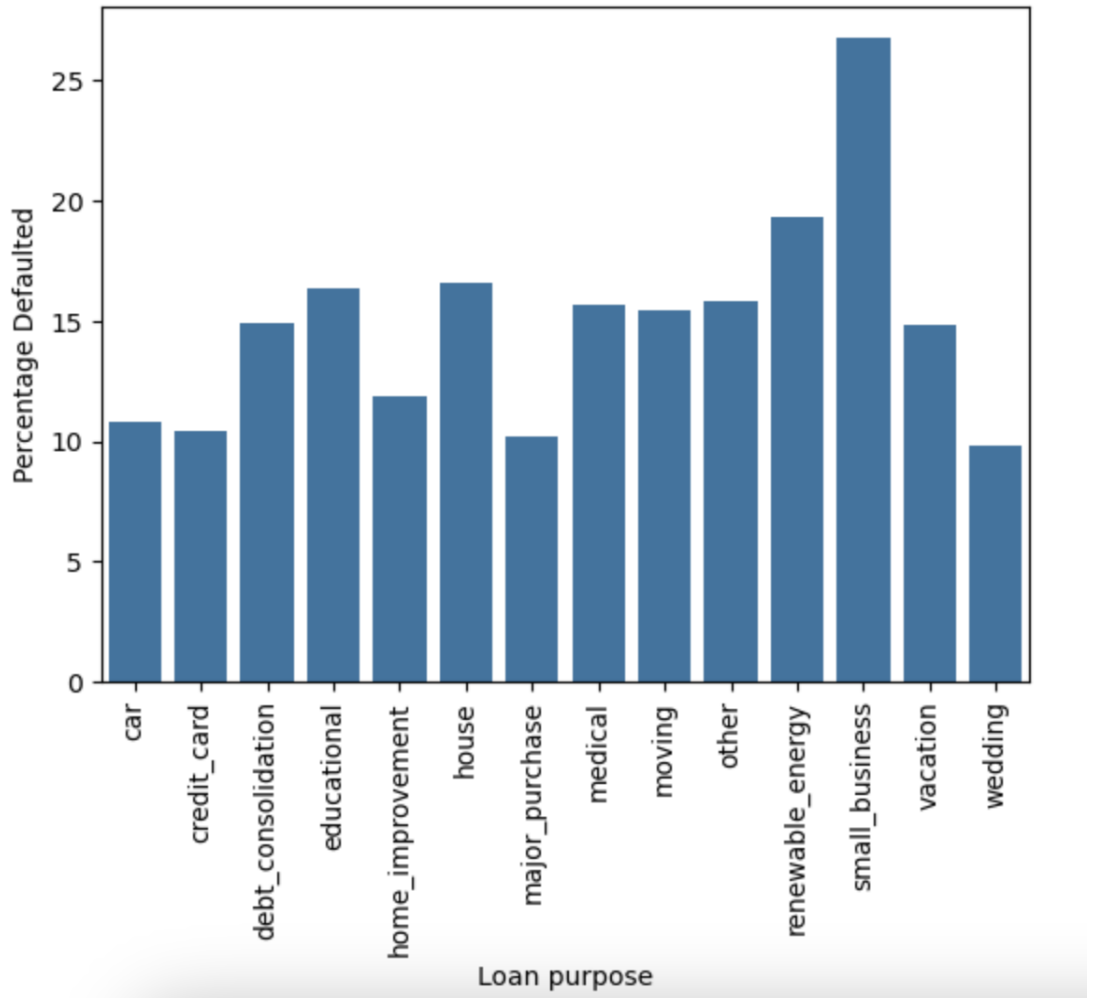
- Analysis of home ownership vs *loan status*



Observation : We cannot get clear indication if home ownership data can tell us about the loan status as almost all categorical value are equally defaulted

BIVARIATE ANALYSIS

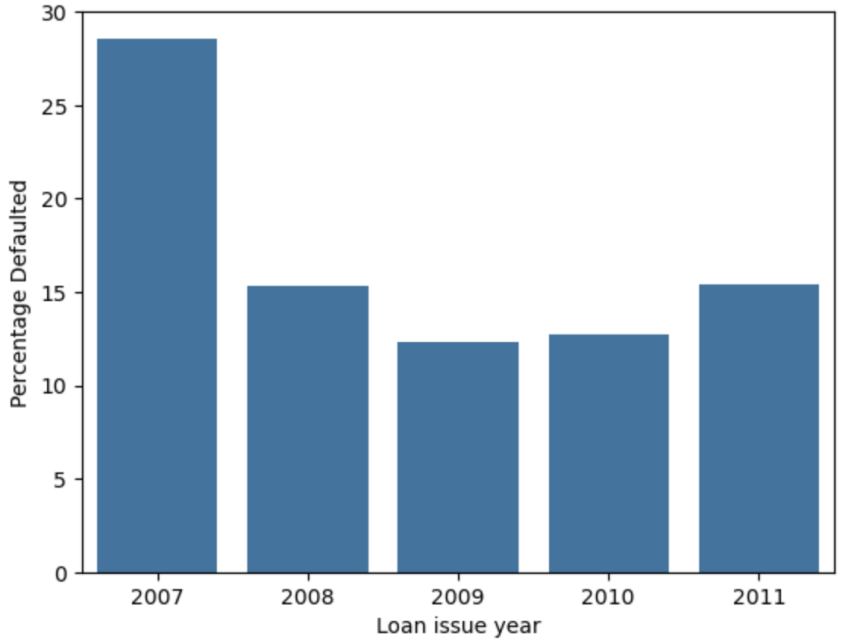
- Analysis of Loan purpose vs *loan status*

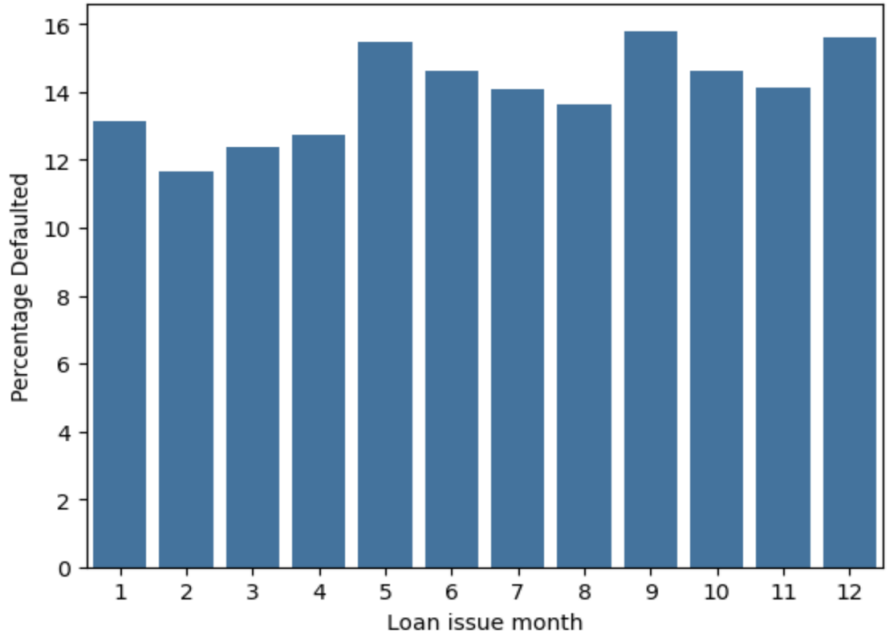


Observation : Applicants who availed loan for small businesses are most defaulted

BIVARIATE ANALYSIS

- Analysis of *Loan issue year/month* vs *Loan status*

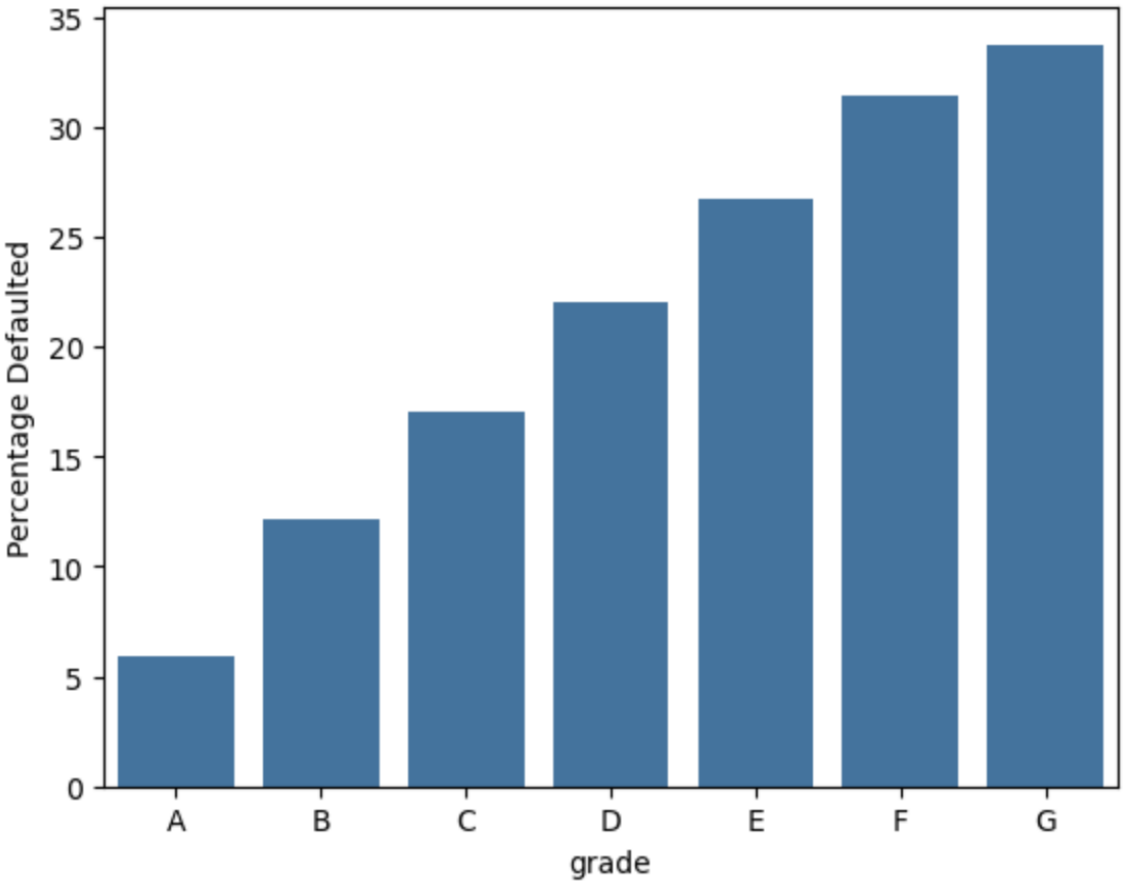




Observation : Loan issued in year 2007 and in months of sept is more likely to get default

BIVARIATE ANALYSIS

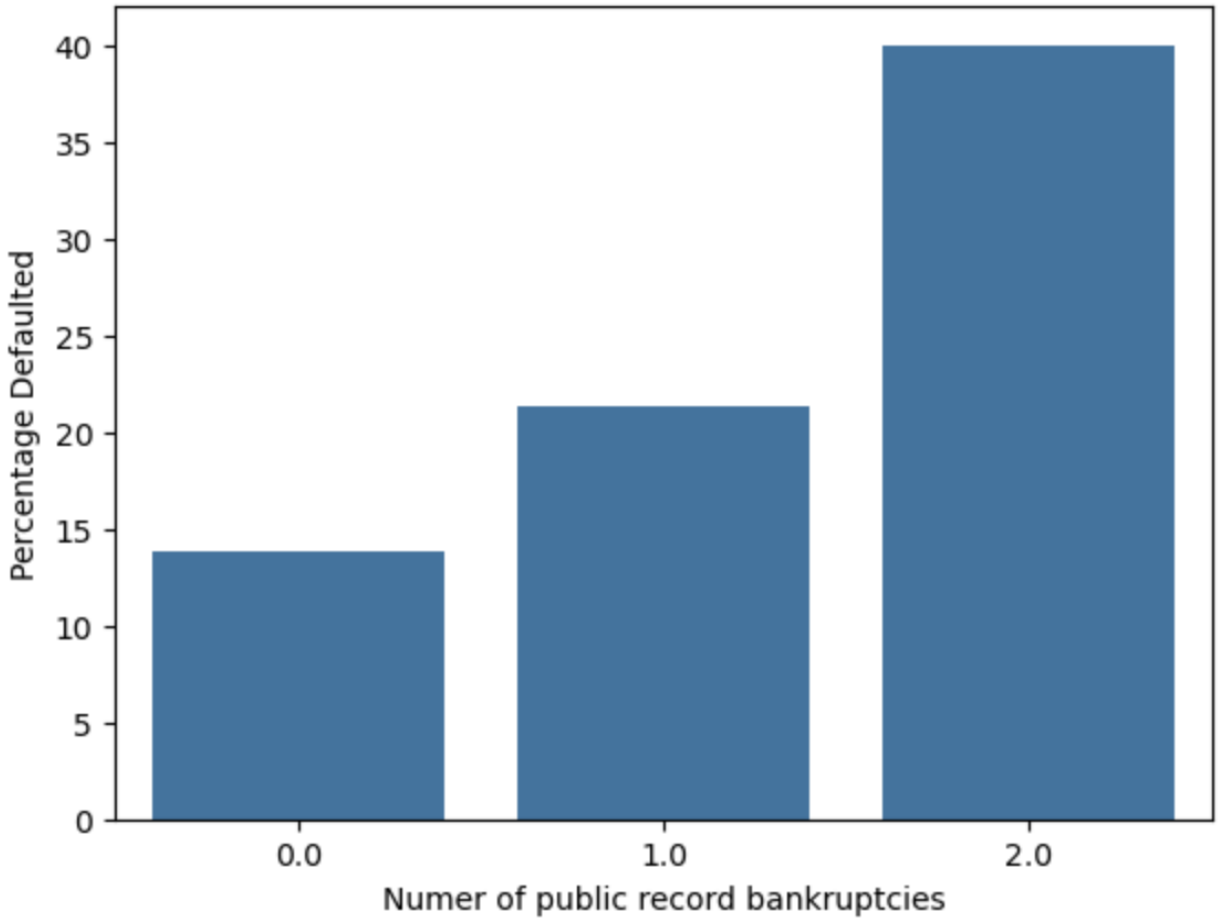
- Analysis of *Grade* vs *Loan status*



Observation : Grade G get default more followed by F

BIVARIATE ANALYSIS

- Analysis of Num of *Public record bankruptcies* vs *Loan status*



Observation : Public record bankruptcies of count 2 are more default

OBSERVATIONS

The analysis of charged-off loans for each variable indicates the following.
The likelihood of default increases when:

- Applicant has annual income in range 0-40k
- Applicant's interest rate is very high : > 20%
- Loan term is of 60 months
- Loan amount is more than 15k
- Applicants use the loan for small businesses
- Loan is issued in year 2007 or in month of sept to dec
- Applicant is from TX state
- Applicant has more that 1 bankruptcies in past
- DTI is more than 20%

RECOMMENDATIONS

- Consider strict check when applicants annual income is in range 0-40k
- Refrain from giving high interest rate of > 20%
- When lending loan for 60 months period, extra measures need to be taken.
- Implement rigorous audit and scrutiny measures when loan is being availed for small businesses.
- Take extra care when applicant is from TX state.
- Refrain from lending loan to applicants who have went bankruptcies more than 1 in past.
- Don't lend loan when DTI is more than 20%

THANK YOU