# Lending Club Case Study

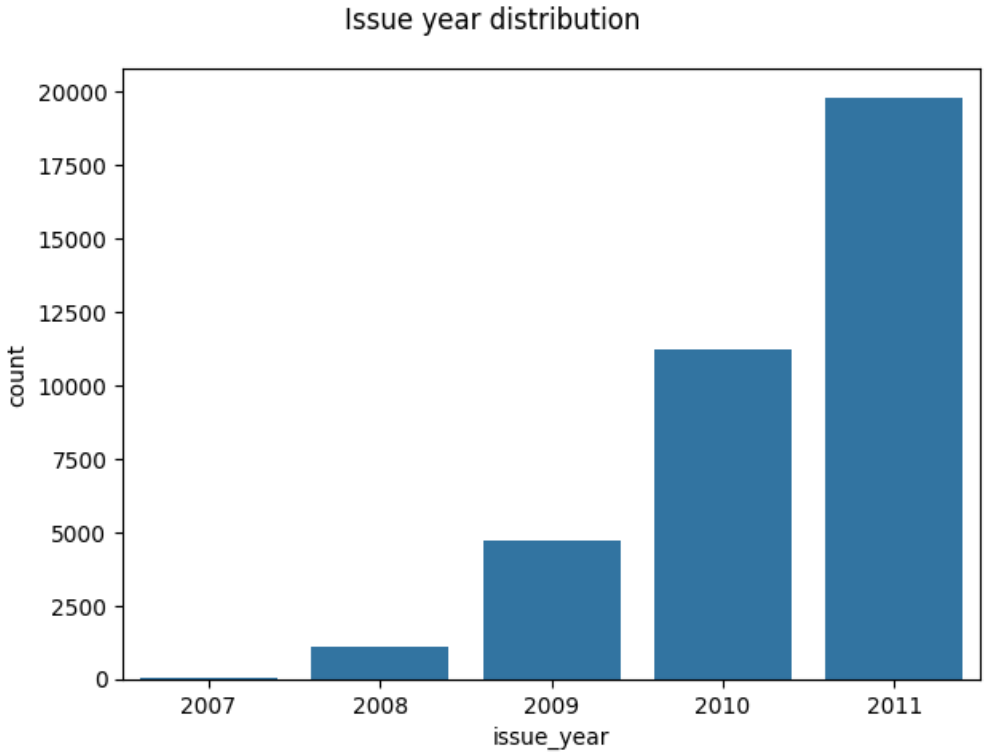
# Problem Statement

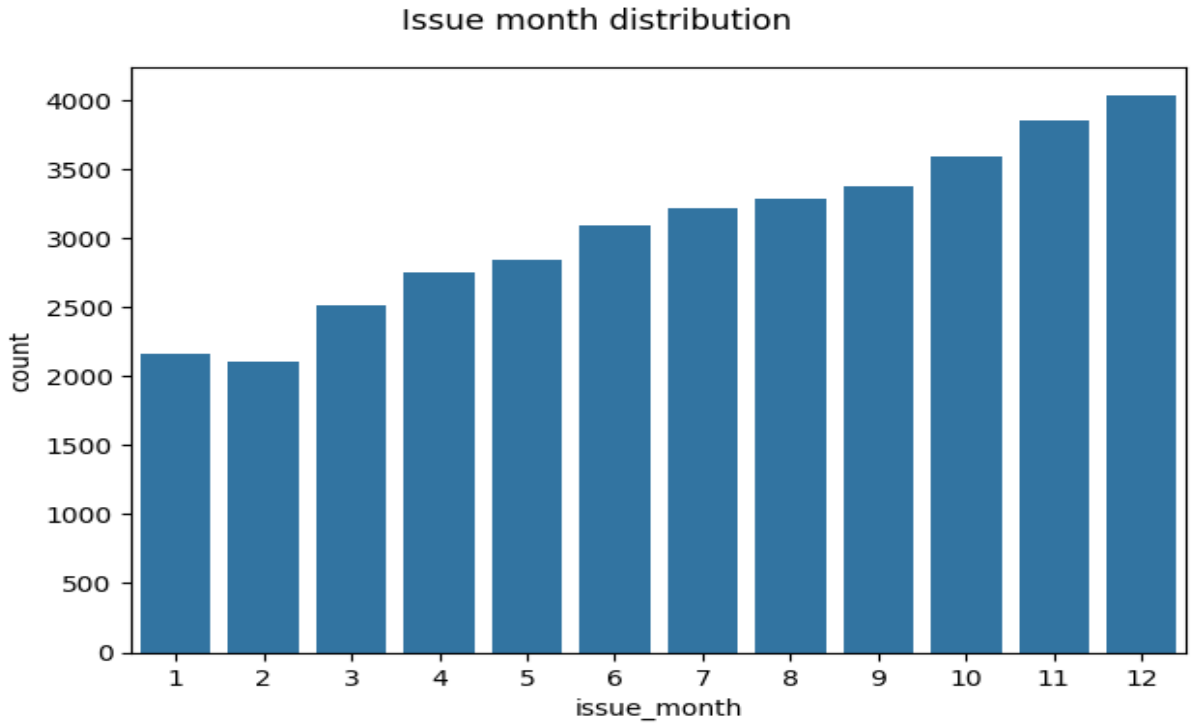
- Lending club is a large consumer finance company which specializes in lending various types of loans to urban customers.
- A dataset containing past data loan applications and whether they have defaulted has been provided for analysis.
- The purpose of this case study will be to use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency to default.
- The business-objective of the case study is to understand the driving factors behind loan default.

# Data Understanding

- The dataset consists of close to 40k records and 111 columns.
- Eliminated columns which have more than 25% of null values.
- Columns containing 1 unique values were removed. e.g. : tax_liens, delinq_amnt, chargeoff_within_12_mths & acc_now_delinq etc.
- Filtering the data :
    - ‘Current’ loans are not within our scope of analysis; hence we excluded those loans.
    - Remove rows which have null values in important fields
- Data conversion :
    - int_rate and revol_util has ‘%’ sign at the end. Removed them and converted to floating type data for performing numerical analysis.
    - Date fields issue_d and earliest_cr_line were converted to proper date format.
- Derived Metric :
    - Added the derived metrics year and month from issue_d and earliest_cr_line

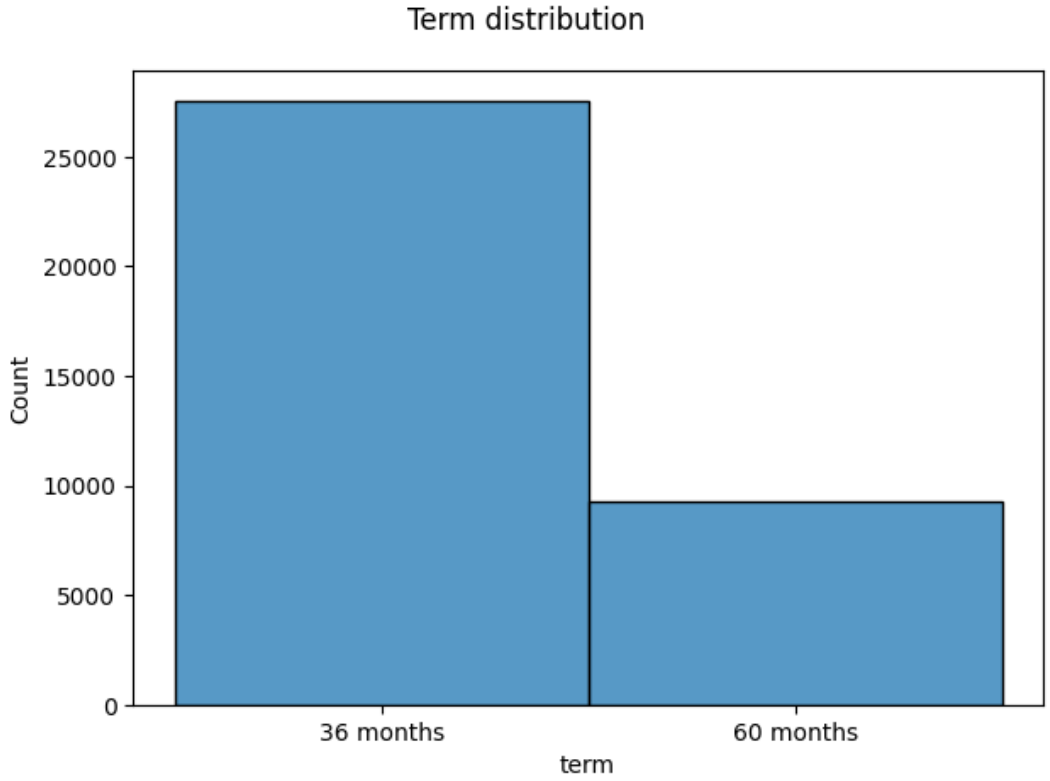
# Univariate Analysis

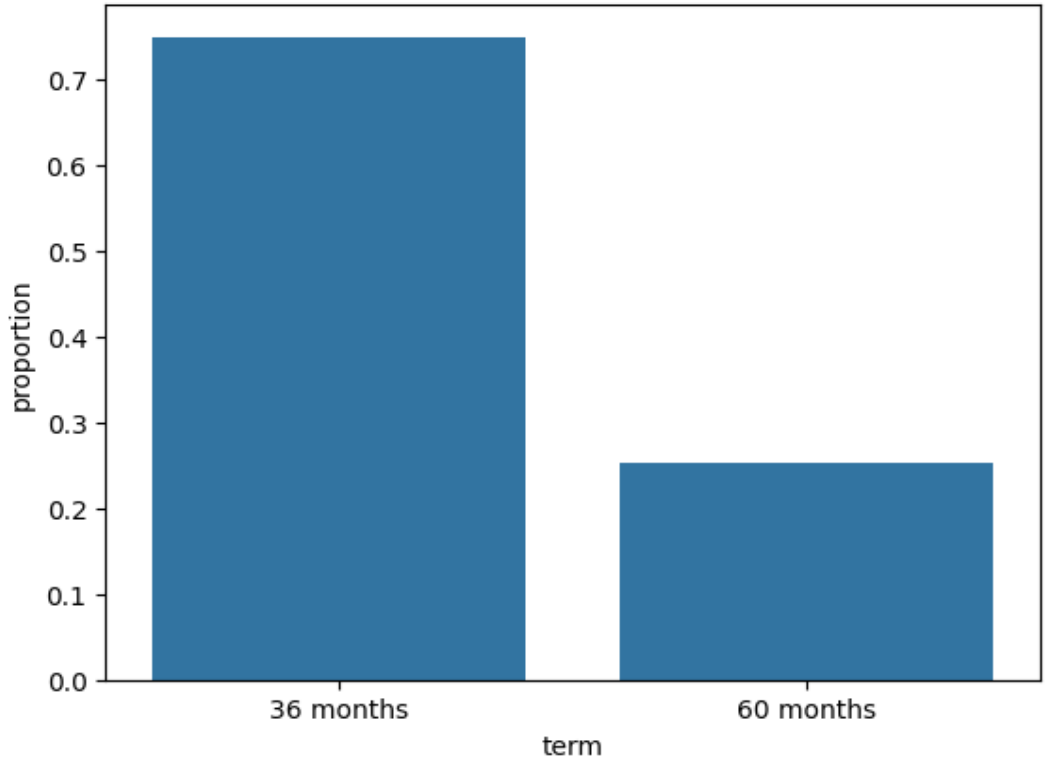



- From the plots it is evident that the number of loans increased significantly in the past 3 years.
- There is also a pattern emerging from the data, where higher number of loans are issued in the last few months of the year.

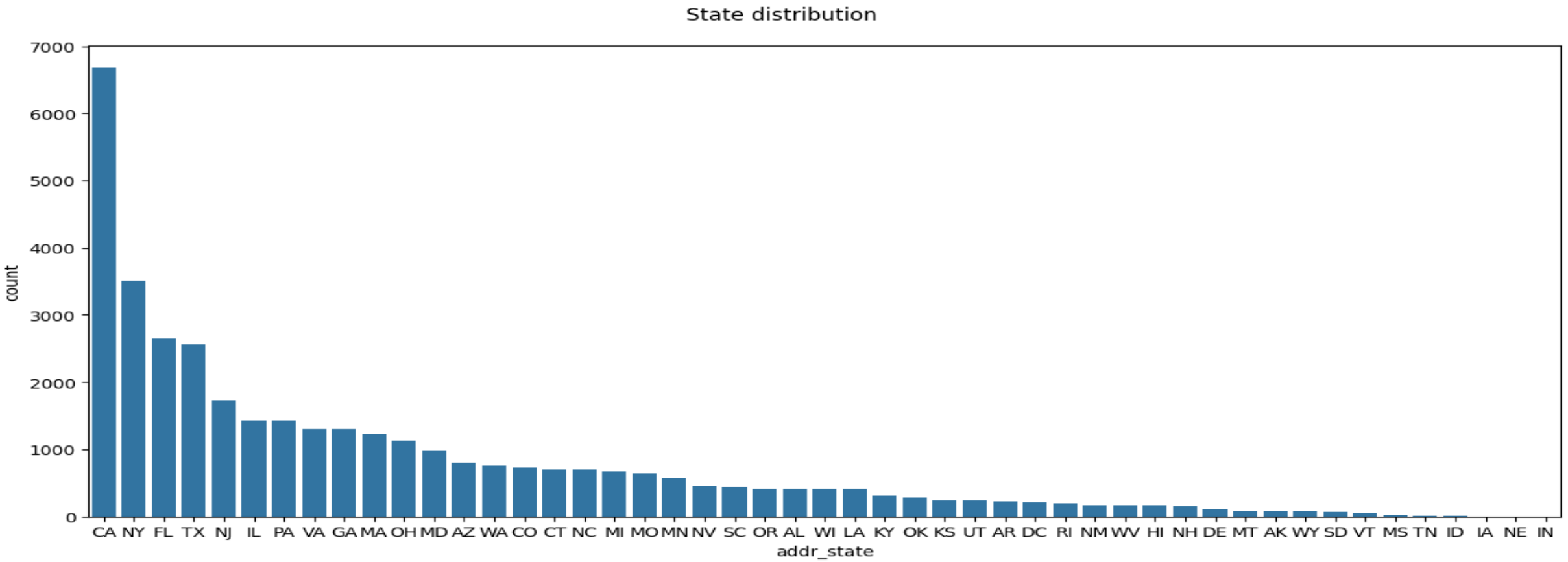
# Univariate Analysis



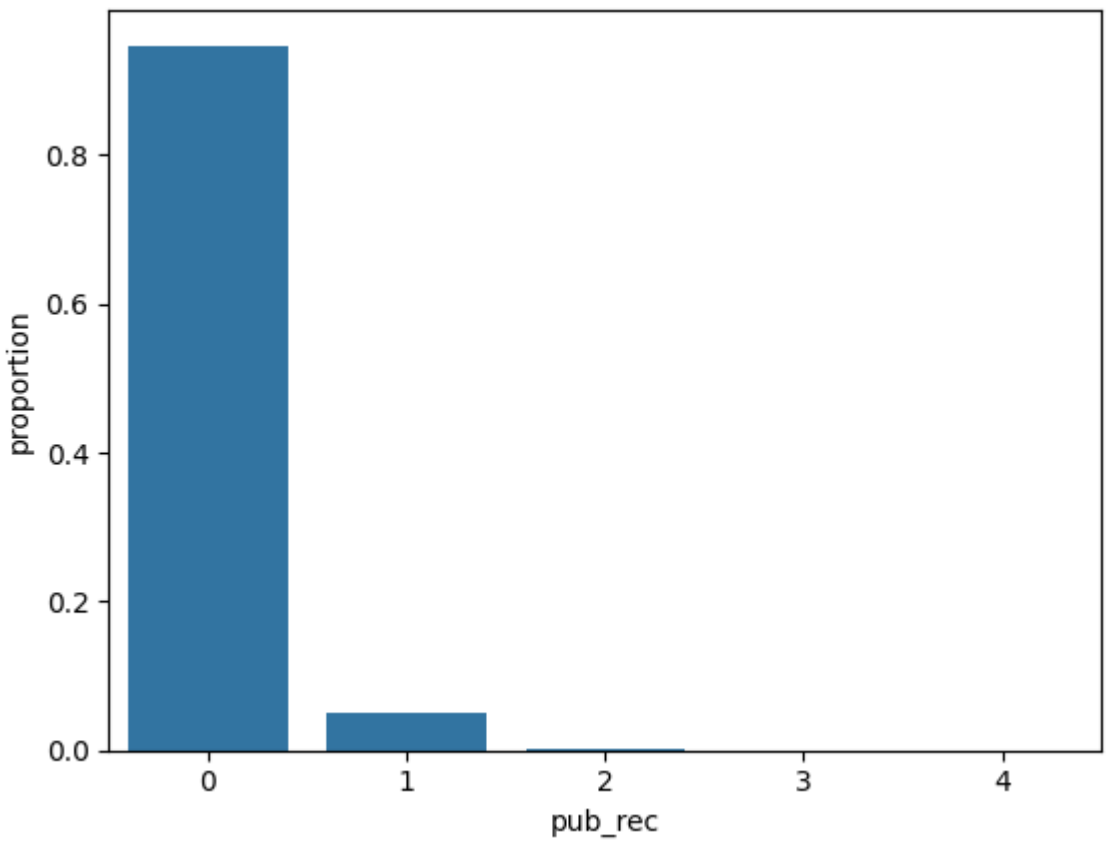

- As many as 75% of the applicants opted for a 36-month loan term

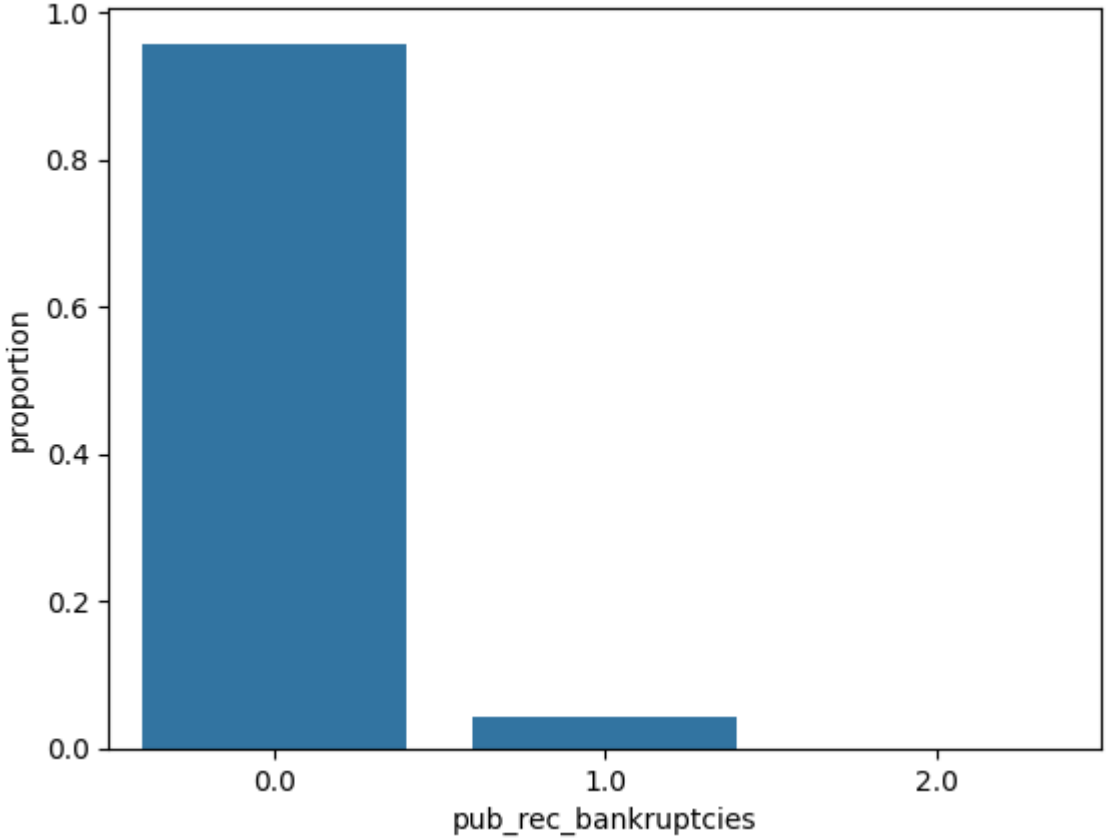
# Univariate Analysis




- Most loan applicants were from the states CA, NY, FL and TX.

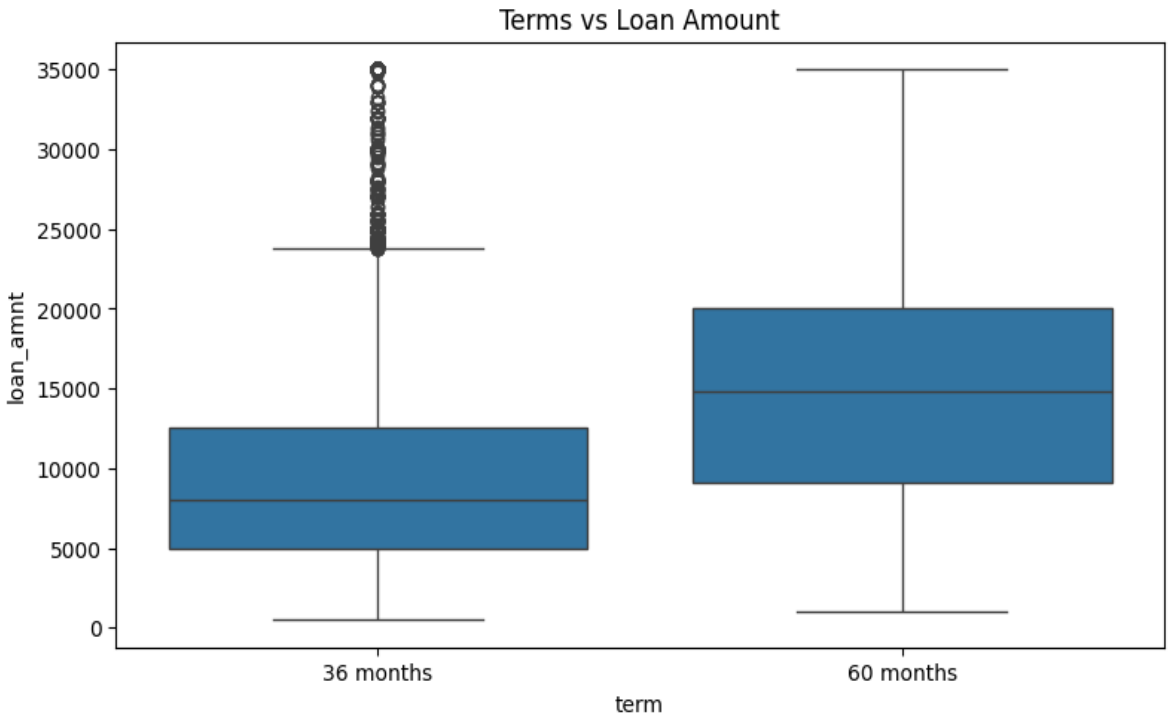
# Univariate Analysis

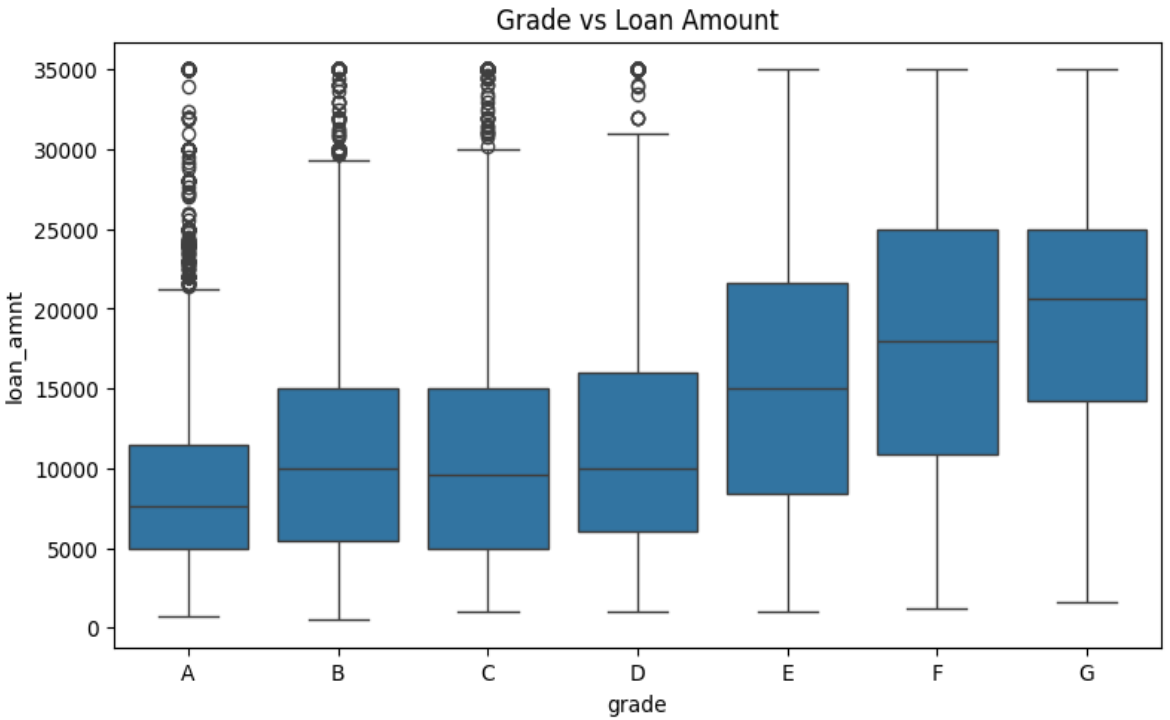



- Almost 95 % people don’t have a public derogatory record.
- 96% of the people don’t have a public bankruptcy record.

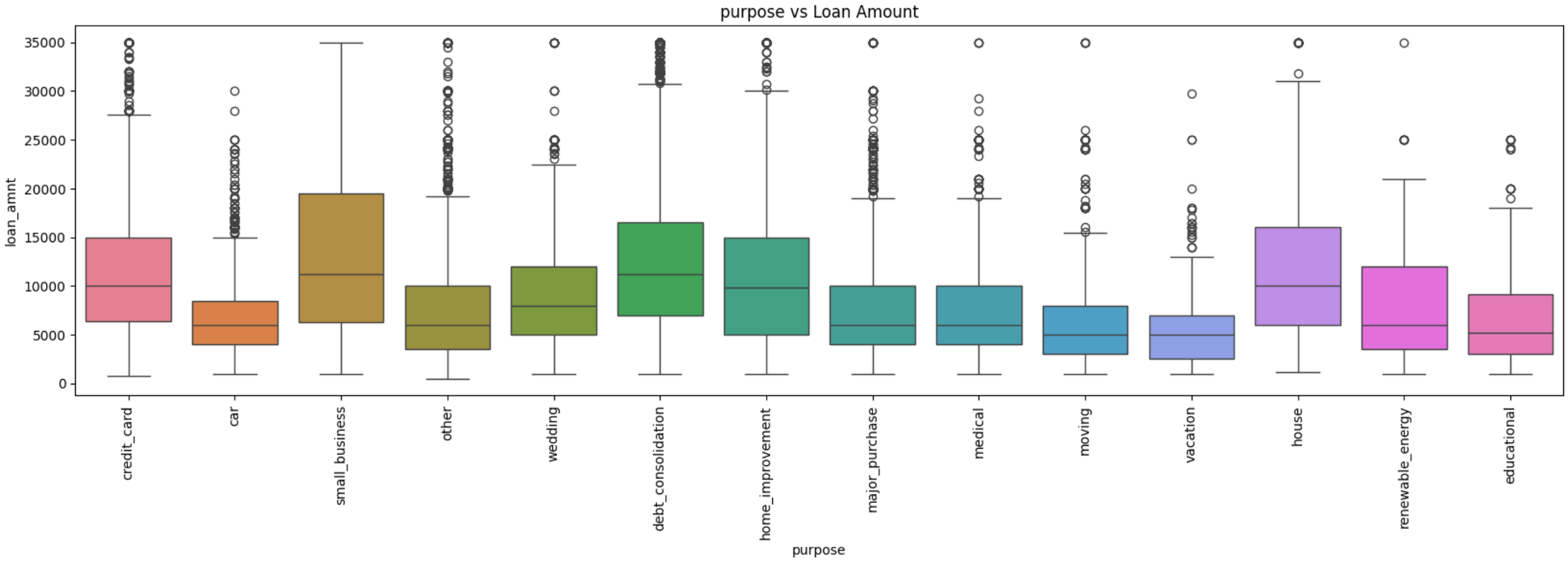
# Segmented Univariate Analysis



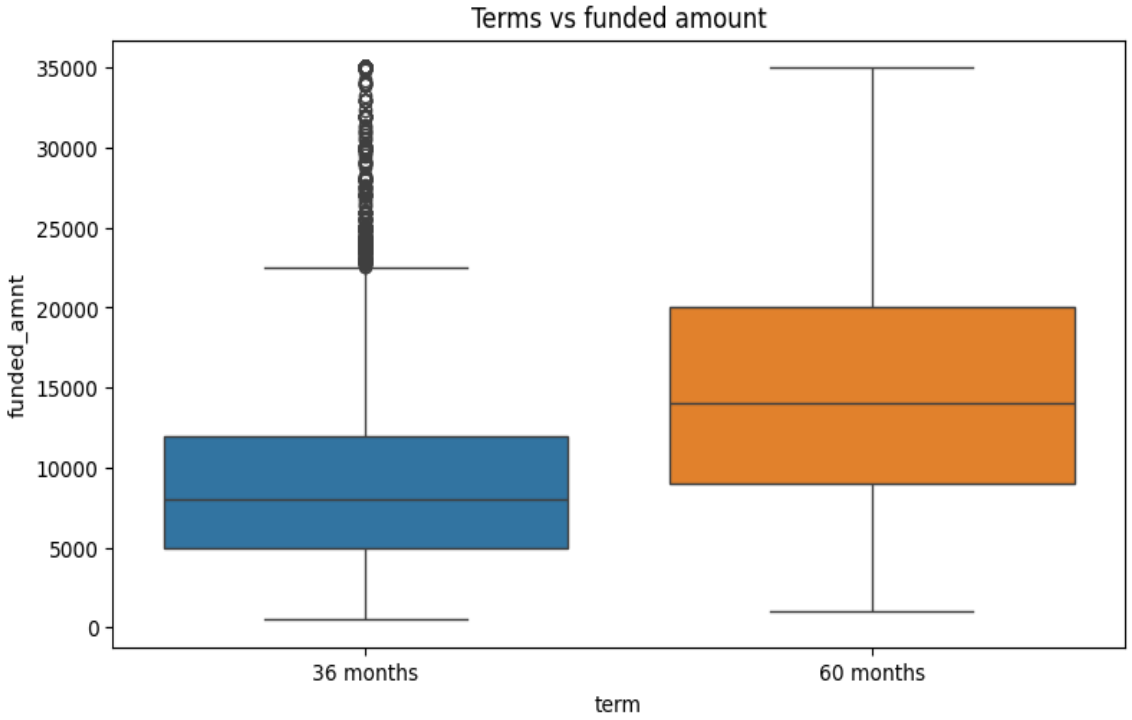

- Higher amount loans have high tenure i.e, 60 months.
- Grade 'G' and 'H' have taken max loan amount. As Grades are decreasing the loan amount is increasing.

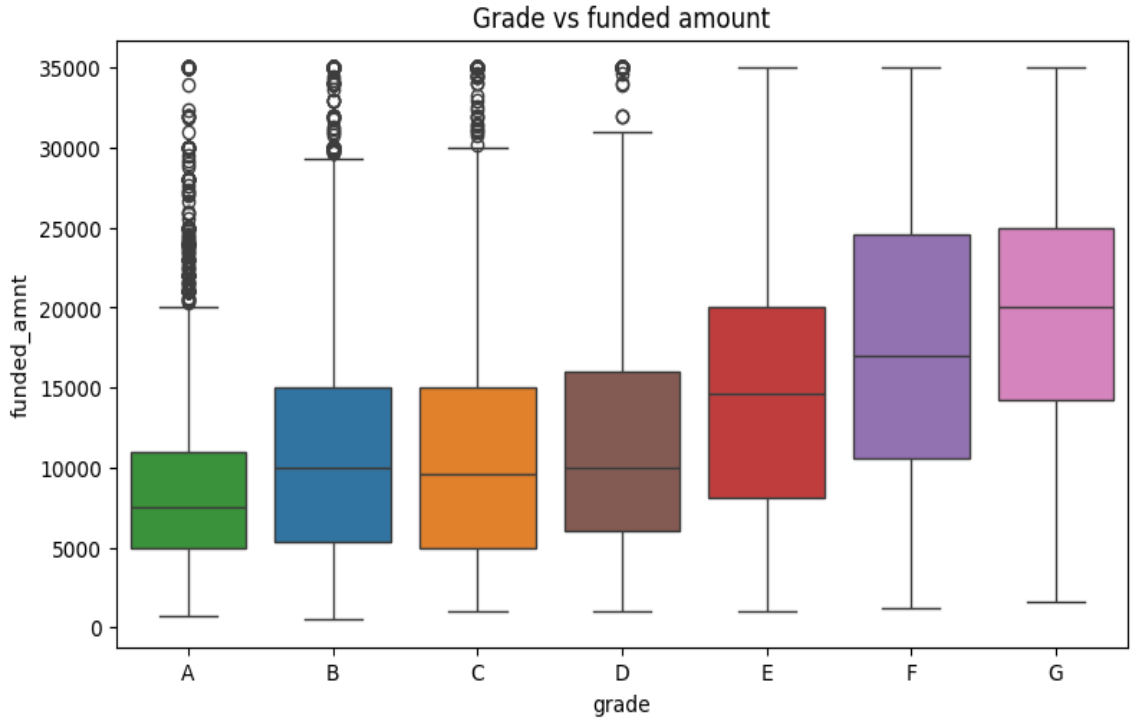
# Segmented Univariate Analysis




- Most loans are taken out for small businesses and houses.

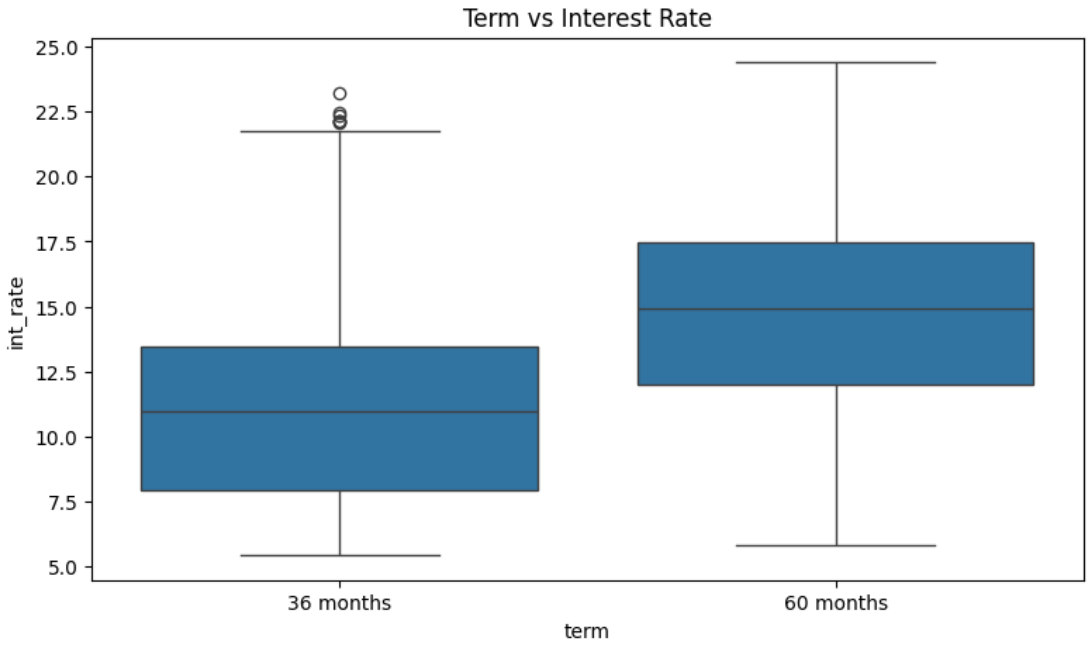
# Segmented Univariate Analysis

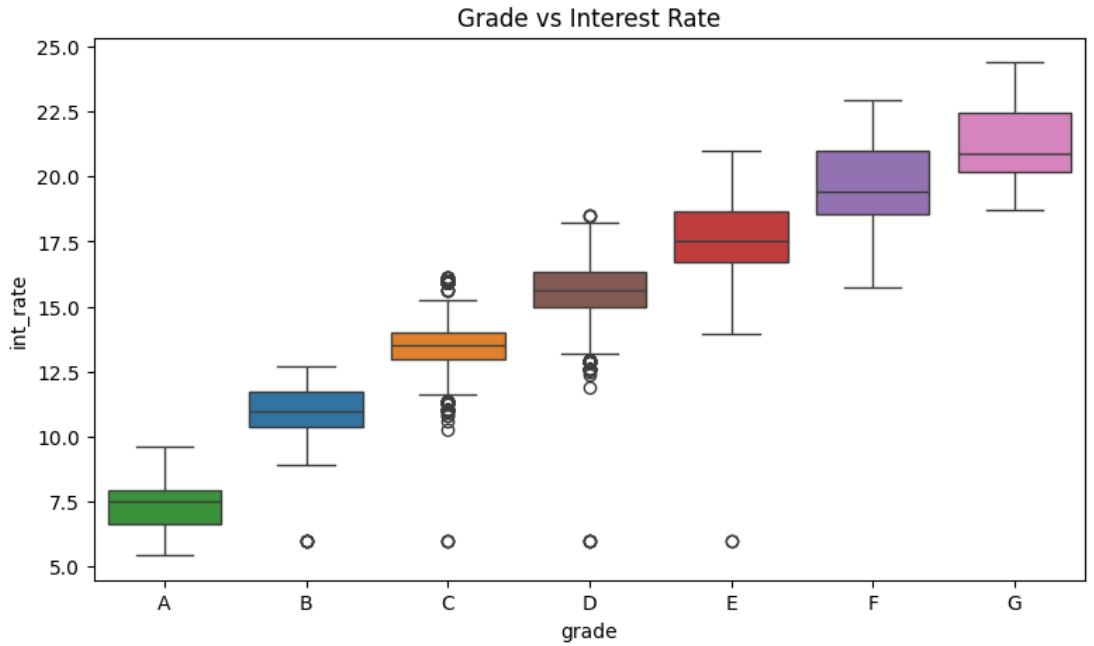



- Higher funded amounts have high tenure i.e, 60 months.
- Grade 'G' and 'H' have taken max funded amount. As Grades are decreasing the loan amount is increasing

# Segmented Univariate Analysis




- The interest rates are higher for Higher tenure loans. And Also Interest Rates are Higher as Grades are Lowering (A to G).

# Conclusions

- Higher DTI is associated with a higher risk of default.
- Lower credit grades (E, F, G) have a higher default rate.
- Longer loan terms (60 months) have a higher default rate.
- Certain purposes (e.g., small business) have a higher default rate.
- Borrowers with public derogatory records have a higher default rate.
- This analysis provides insights into loan default prediction and can be used for risk assessment and credit scoring