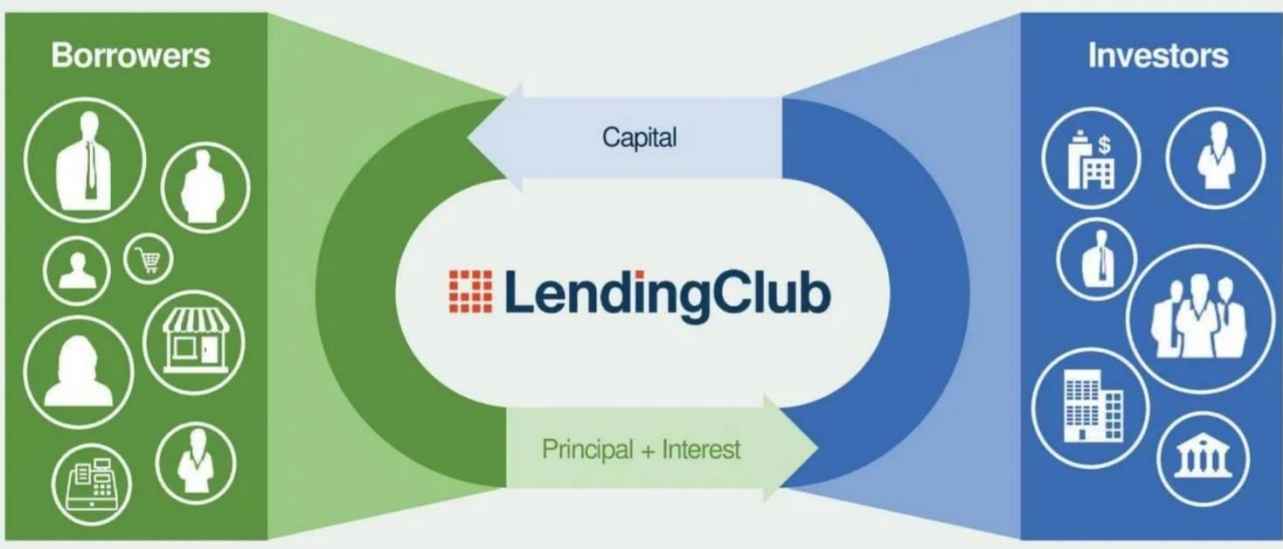# Lending Club

Case Study submission by:
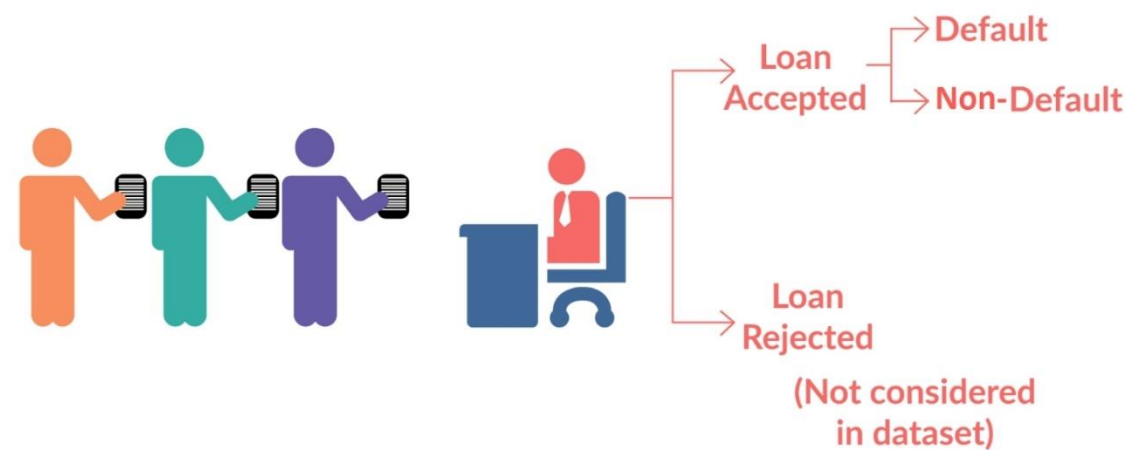
- **Abhinav Kulkarni**

- **Kumar Shekhar**
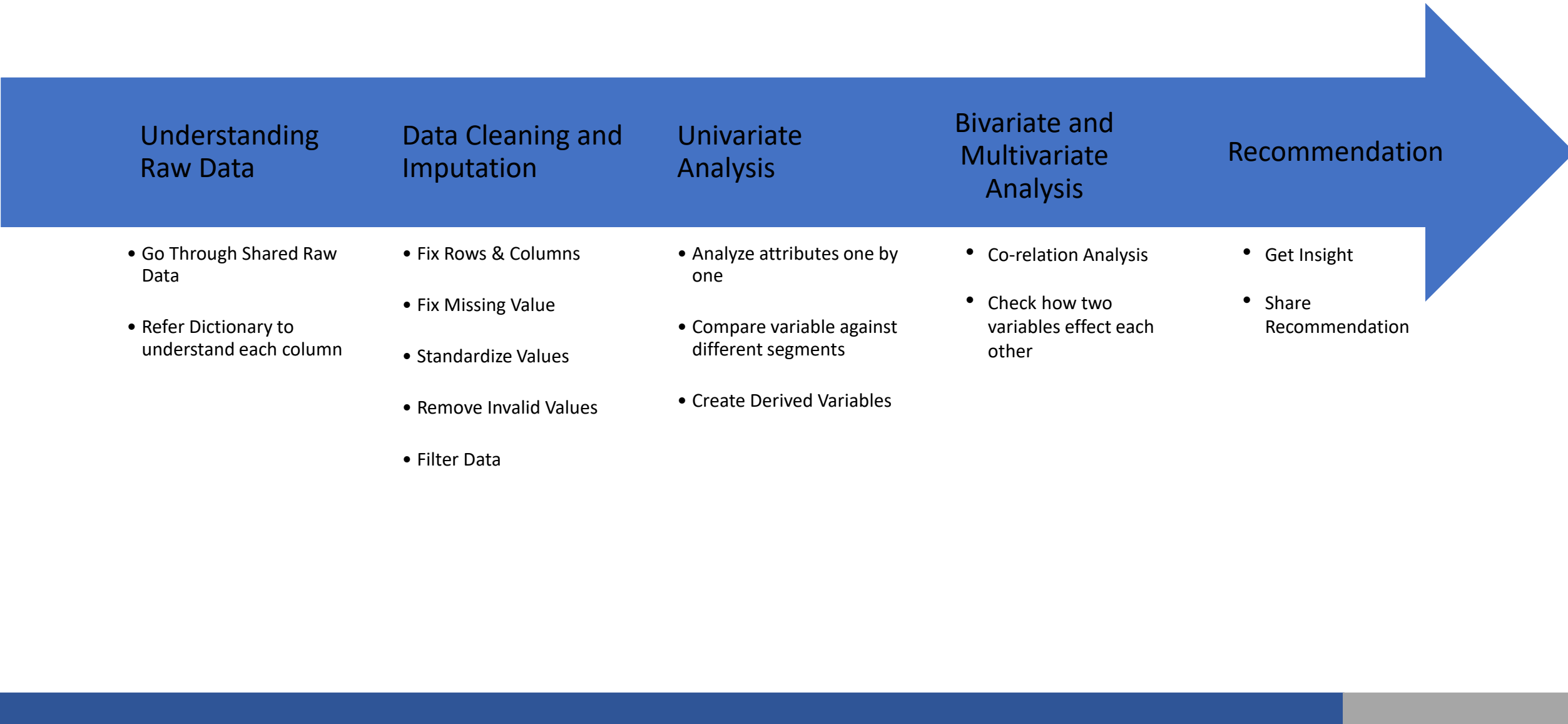
Basic Information

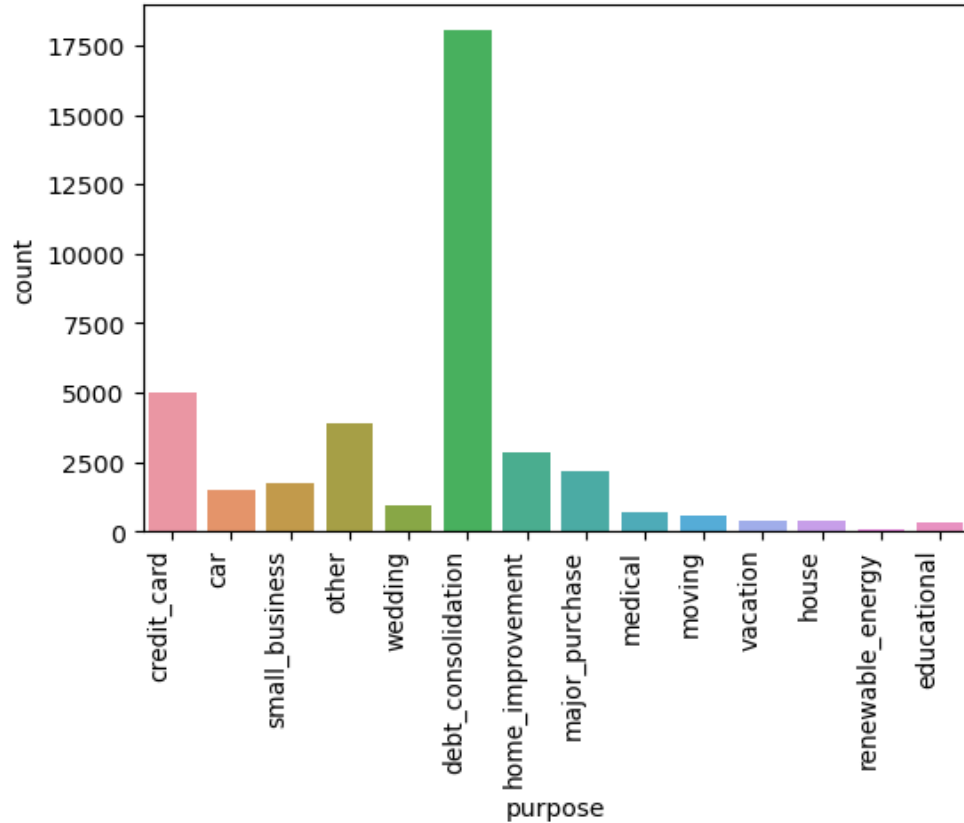❑ **What Is Lending Club?**

❑ **Loan Data Set**

Problem Statement

- This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

- Lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). In other words, borrowers who **default** cause the largest amount of loss to the lenders.

- Identification of risky loan applicants using EDA is the aim of this case study.

- In other words, the company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default.

Approach

Understanding Raw Data

- Go Through Shared Raw Data
- Refer Dictionary to understand each column

Data Cleaning and Imputation

- Fix Rows & Columns
- Fix Missing Value
- Standardize Values
- Remove Invalid Values
- Filter Data

Univariate Analysis

- Analyze attributes one by one
- Compare variable against different segments
- Create Derived Variables

Bivariate and Multivariate Analysis

- Co-relation Analysis
- Check how two variables effect each other

Recommendation

- Get Insight
- Share Recommendation

Loan Account Analysis

Majority of the loans have been distributed for the purpose of debt reconstruction, credit card, home improvement and other reasons.
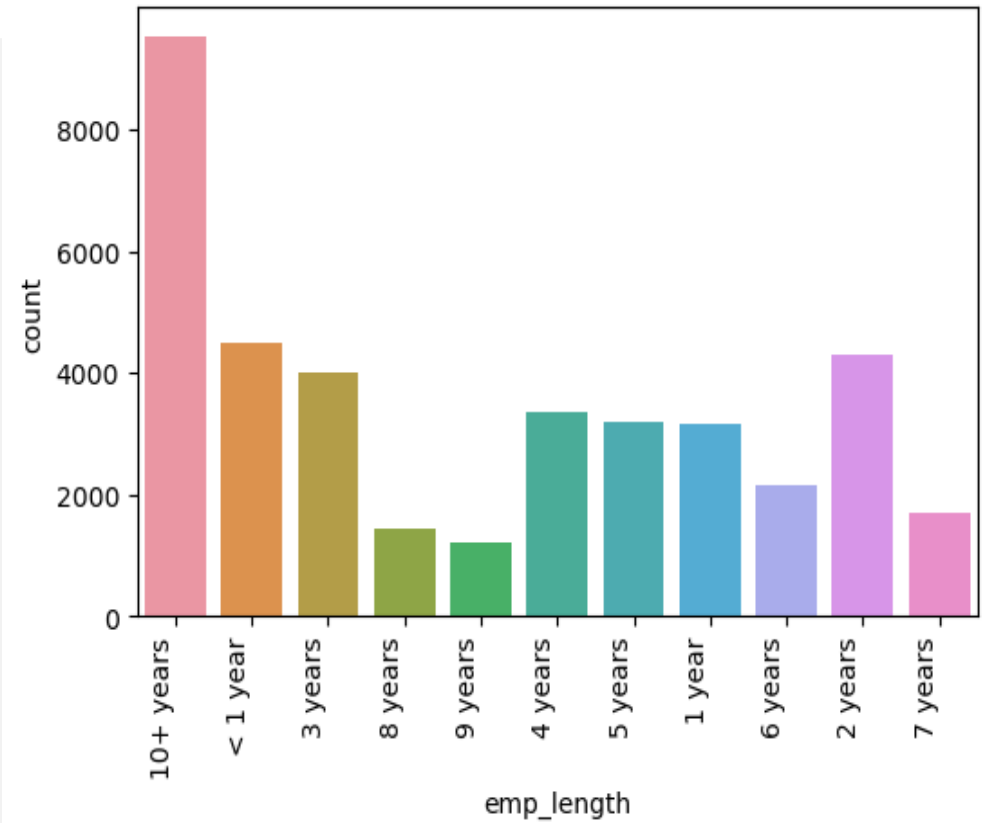
Majority of the loans have been distributed to the applicants with more that 10 years of employment, followed by less than 1year, 2 years and 3 years of employment.

Univariate Analysis : Part 1

- Approximately 16% loans have been defaulted on by the borrowers.
- Most of the loans distributed have 36 months tenor for repayment.
- Majority of the distributed loans fall under Grade B, followed by Grade A and Grade C.
- Most of the loan applicants have rented their home or have taken loan for home ownership purpose i.e. mortgage.
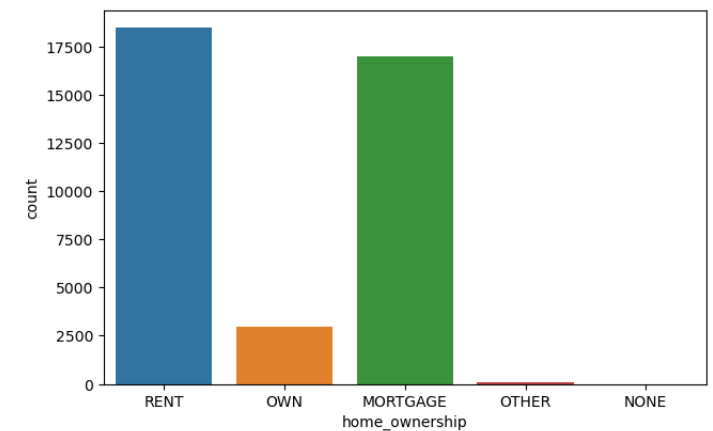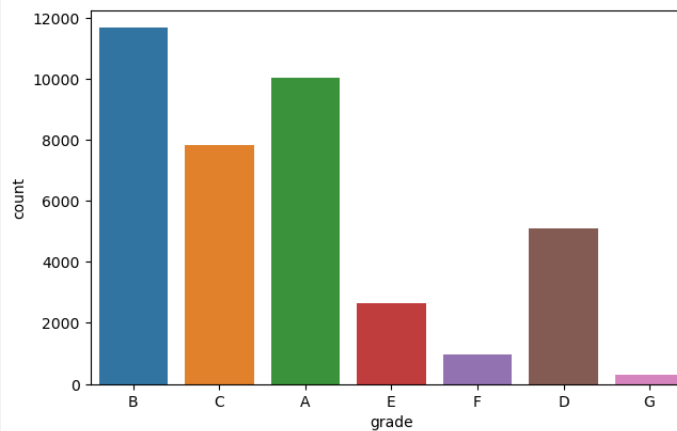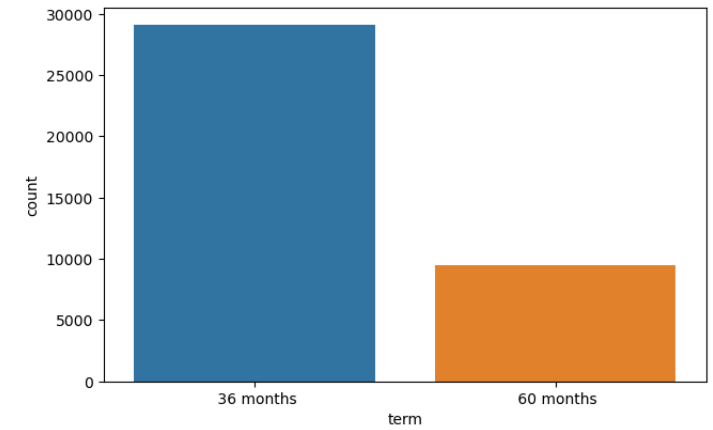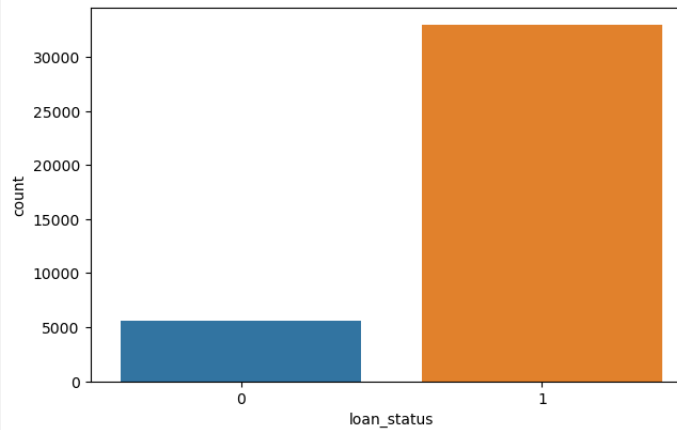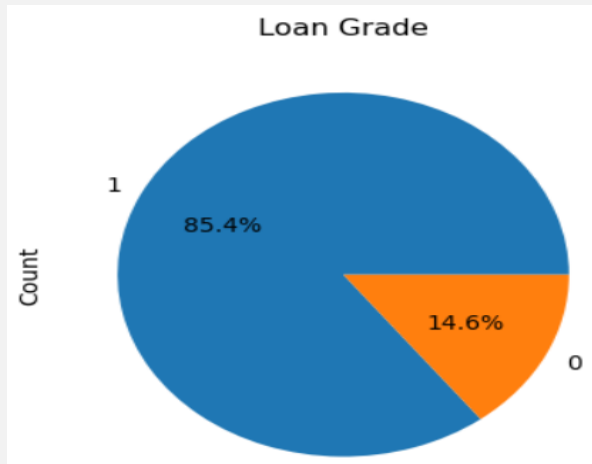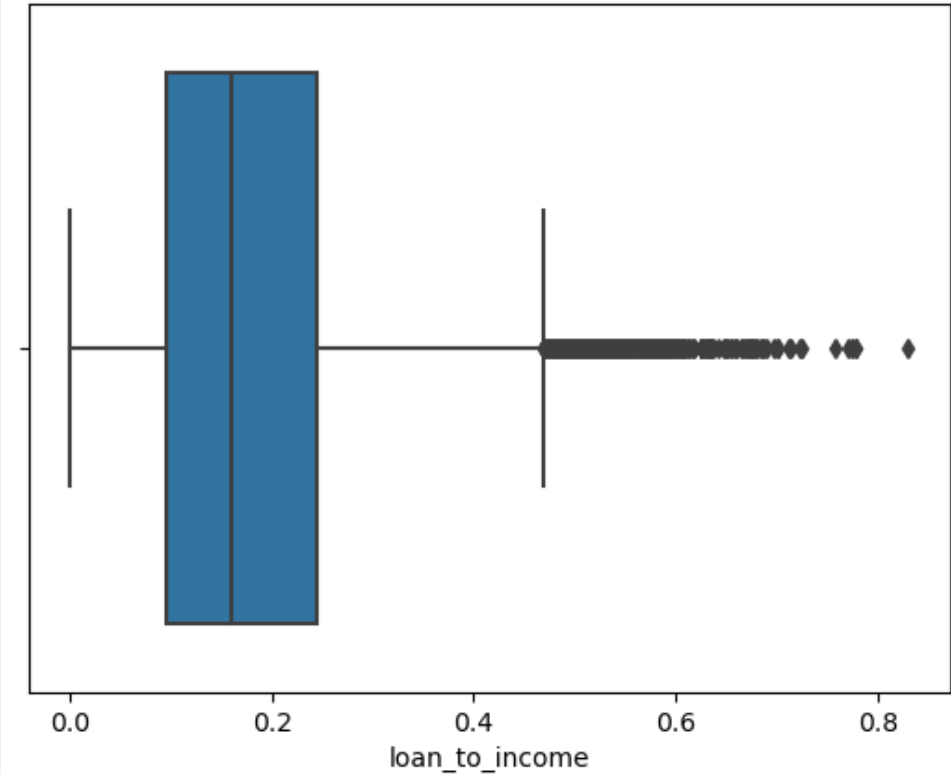
Univariate Analysis : Part 2

Distribution of loan amount applied by the borrower

Distribution of funded loan to annual income ratio
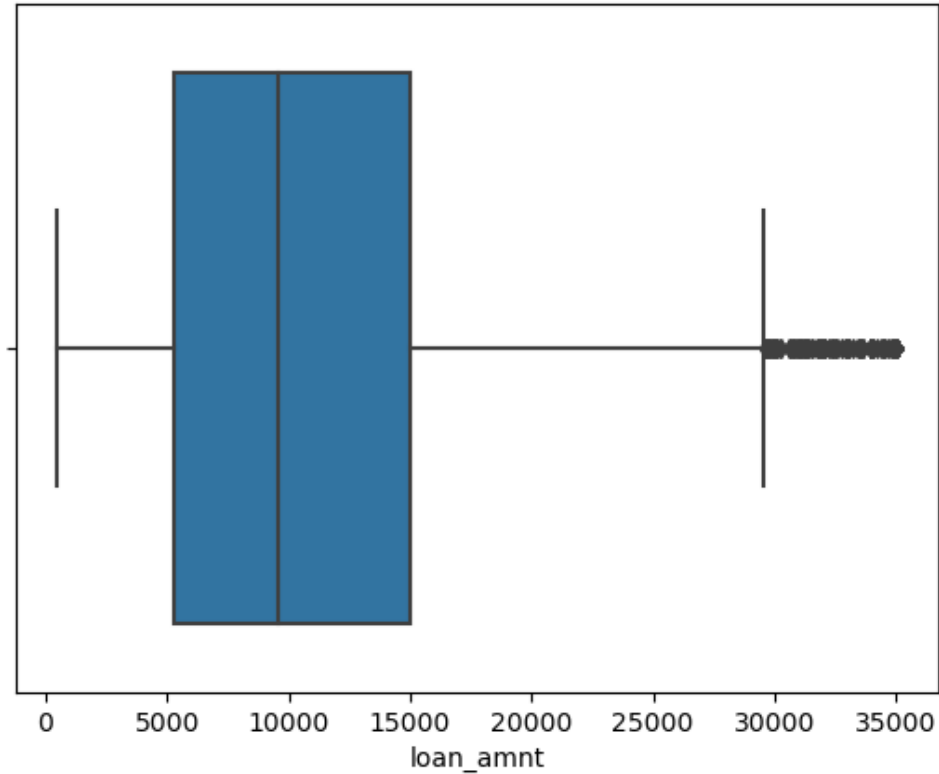
- Distribution of loan application amount is positively skewed and has significant presence of outliers.

- Distribution of funded loan to income ratio is positively skewed and has significant presence of outliers.

Bivariate Analysis: Loan Default vs. Loan Purpose

Plot for percentage wise break up of loan purpose vs loan status
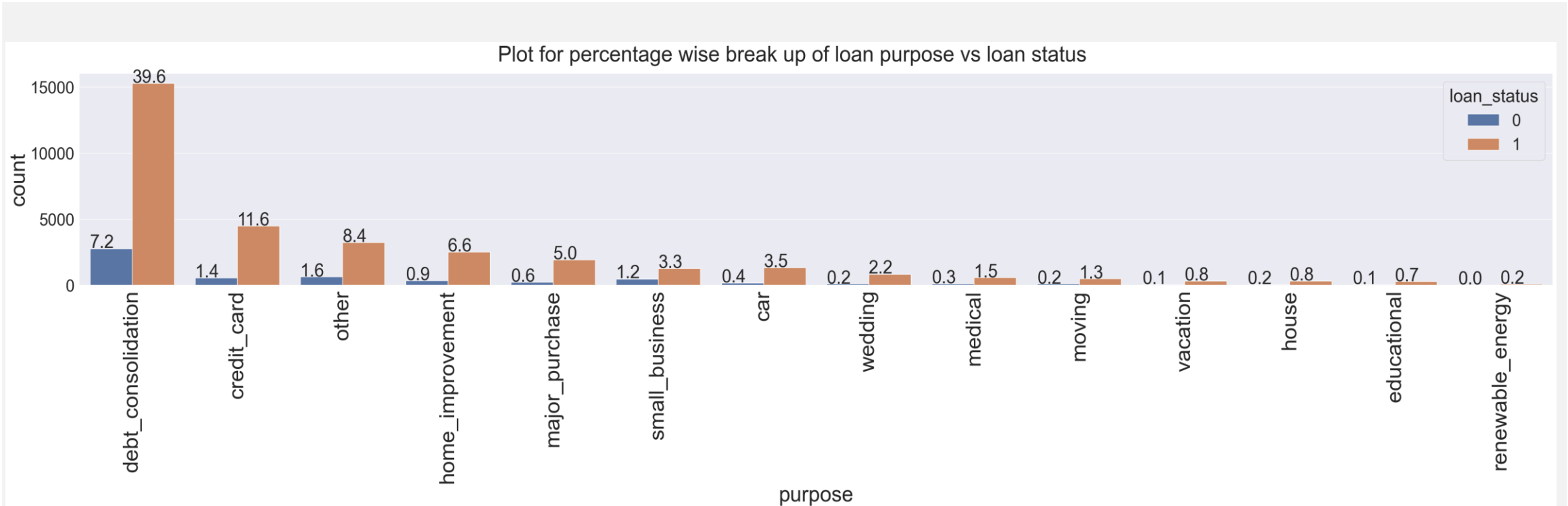
- Highest number of Default Cases are notices in Debt Consolidation, which is also the prime reason for which maximum loans are taken.
- Next highest default cases are observed for credit cards and other category.
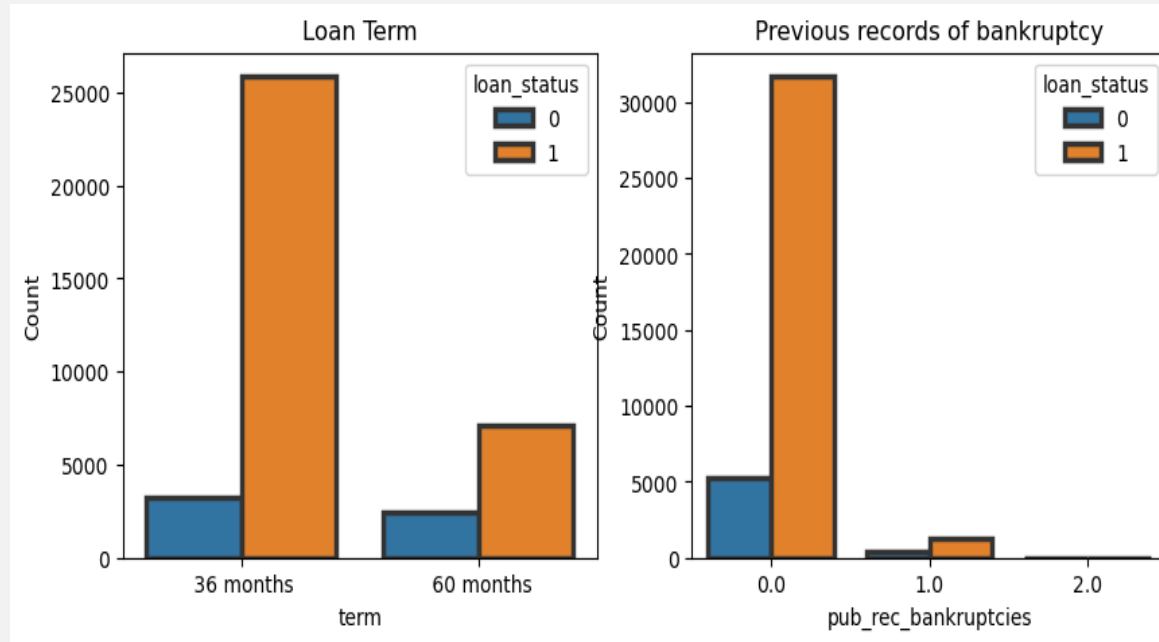
# Bivariate Analysis: Loan Default vs. Home Ownership and State
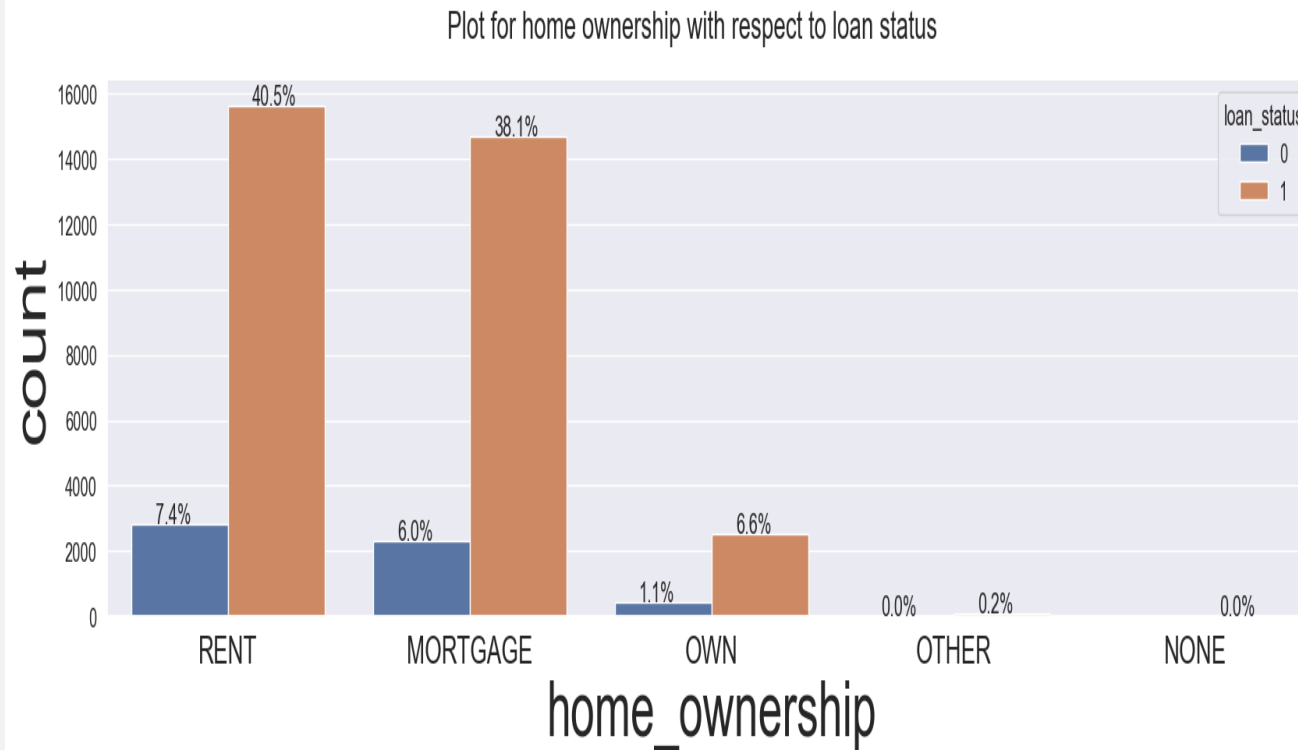
- Charged Off Cases are higher when people have Mortgage & Rent as accommodation compared to own house (as shown in above plot).
- Borrowers from CA has high chance of defaulting the Loan, followed by FL & NY (as shown in adjacent plot).
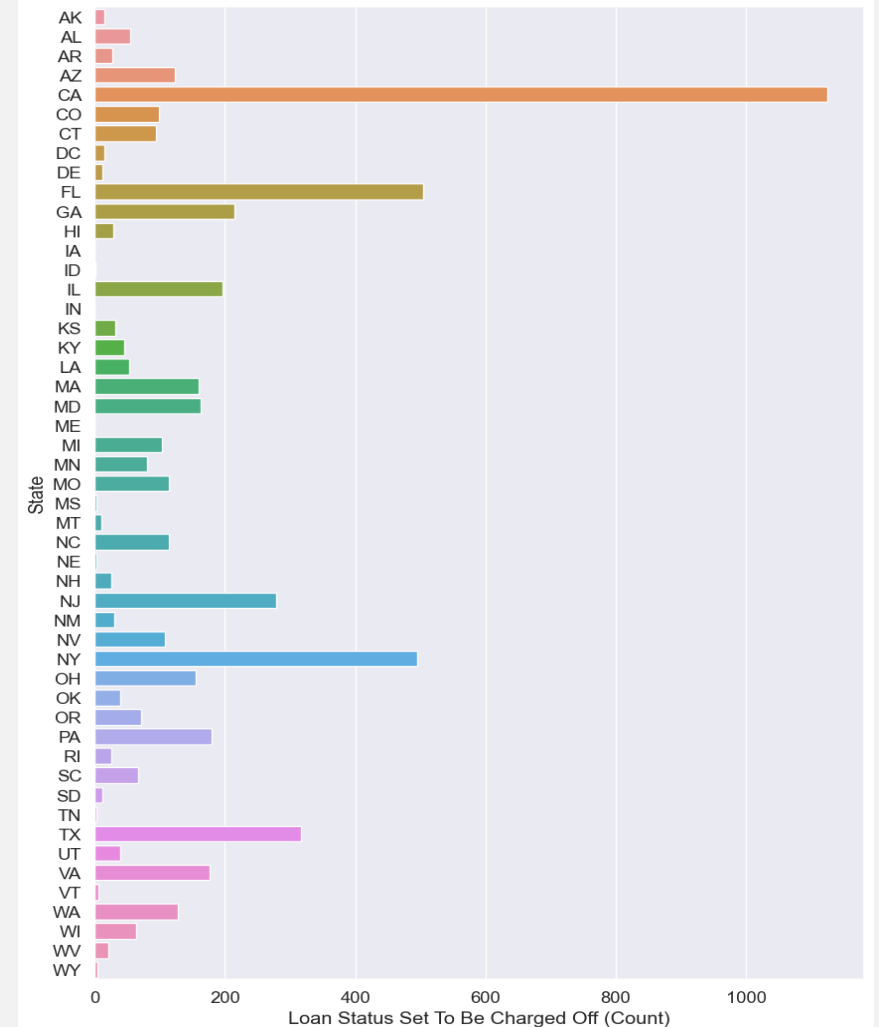
# Bivariate Analysis: Loan Default vs. Term, Grade and Bankruptcy record

- Longer tenure cases (60 months) show higher percentage of default as compared to small tenure loans (30 months).
- Percentage of default is higher if applicant has previous record of bankruptcy. Prior history of bankruptcy increases the chance of default.
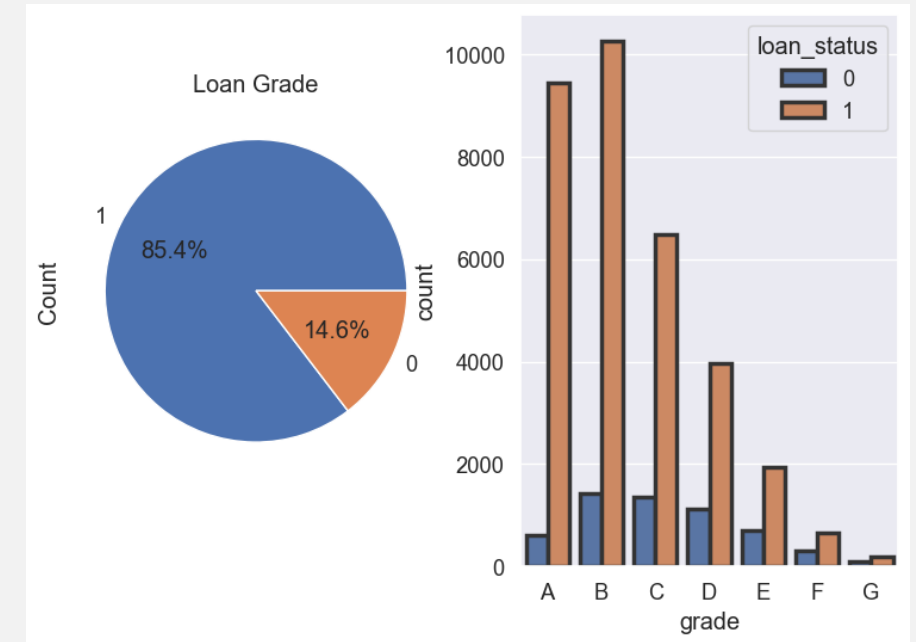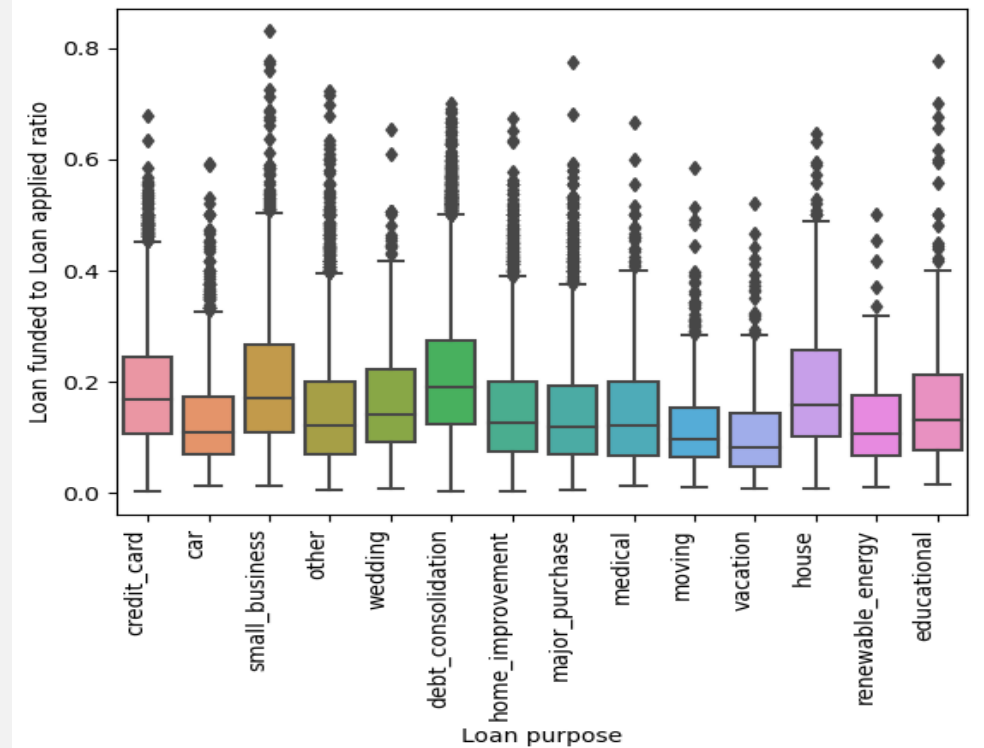
- Proportion of default seems to be lowest in Grade A and generally increases with grades C, D,E,F and G.

# Bivariate Analysis: Distribution of loan to income ratio
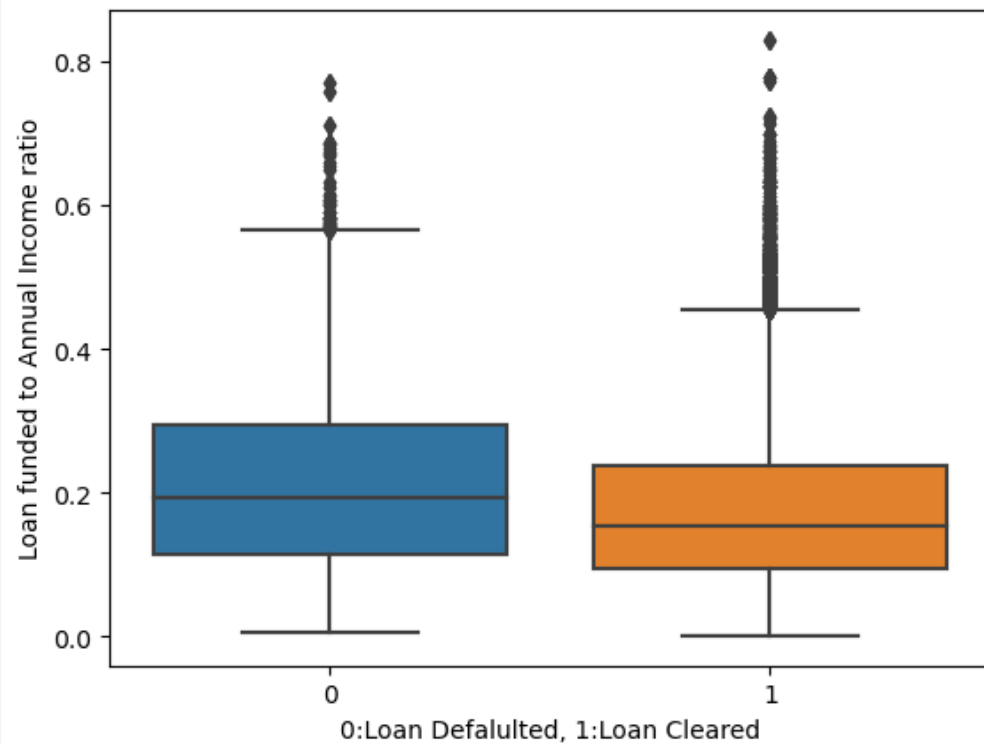
Loan to income ratio has comparatively higher median for the defaulted cases.

Loan to income ratio has comparatively higher median for debt consolidation and credit card categories, which also record highest number of defaults across various loan purposes.

Multivariate Analysis

- Following variables seem to have high positive correlation with one another.
  - loan_amnt
  - funded_amnt
  - funded_amnt_inv
  - Installment

- Following variables seem to have high positive correlation with one another.
  - total_pymnt
  - total_pymnt_inv
  - total_rec_prncp

- total_pymnt is highly positively correlated with funded_amnt.

- total_pymnt_inv is highly positively correlated with funded_amnt_inv.
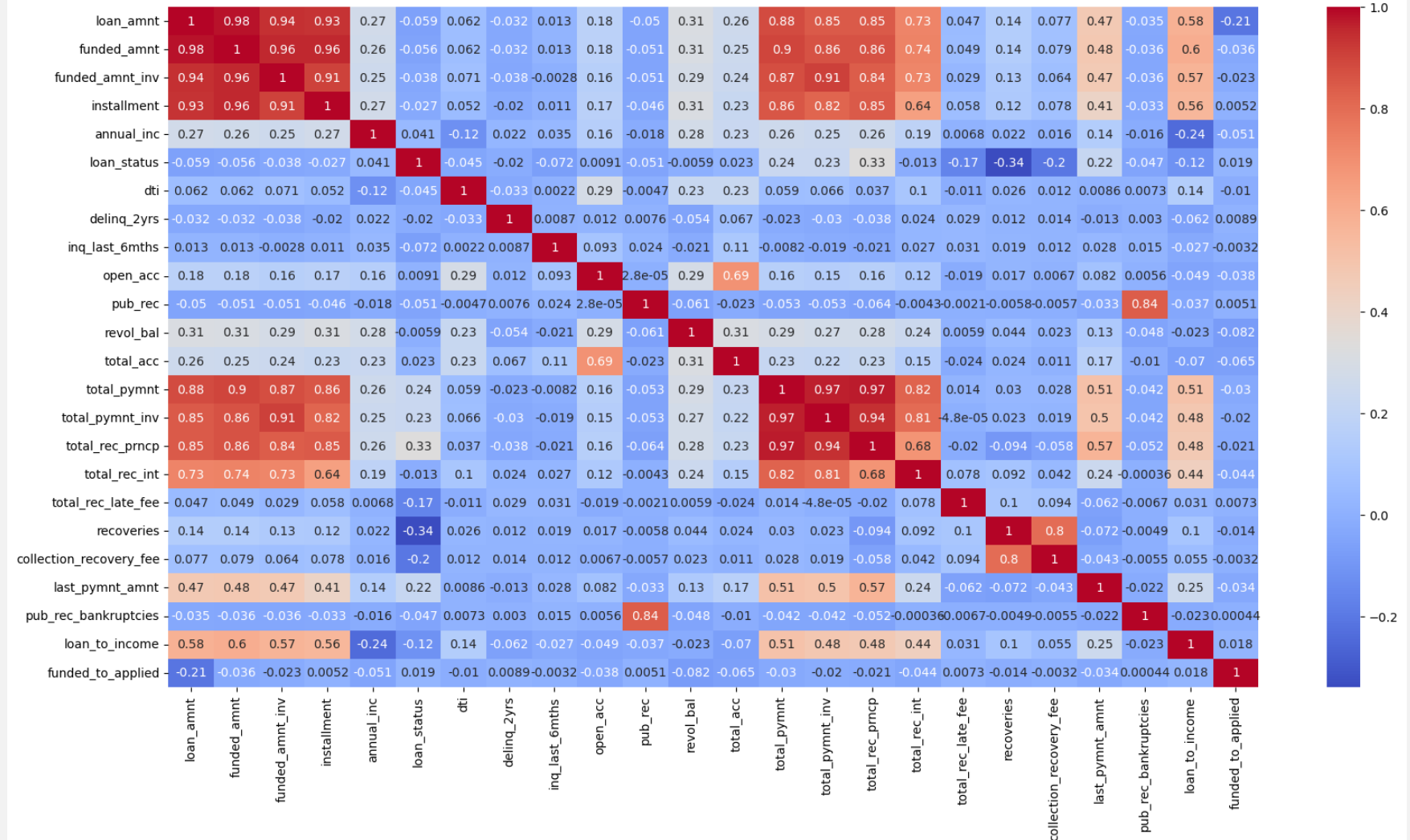
Recommendations

- Chance of loan default are higher if:
    - Loans belong to C,D,E or F grade
    - Borrower has prior history of bankruptcy
    - Loan is being applied for debt consolidation or credit card.
    - Borrower has rented Mortgage & Rent as accommodation compared to own House.

    Such an applications would require extra scrutiny and verification.
- Additional verification would be required for long term loans as they show higher percentage of default as compared to short term loans.
- Detailed verification should be done for borrowers from CA, FL & NY. It is because of large cases of default in such cities.
- For borrowers wherein Debt To Income ration is greater than 15%, loan approval can be deffered.