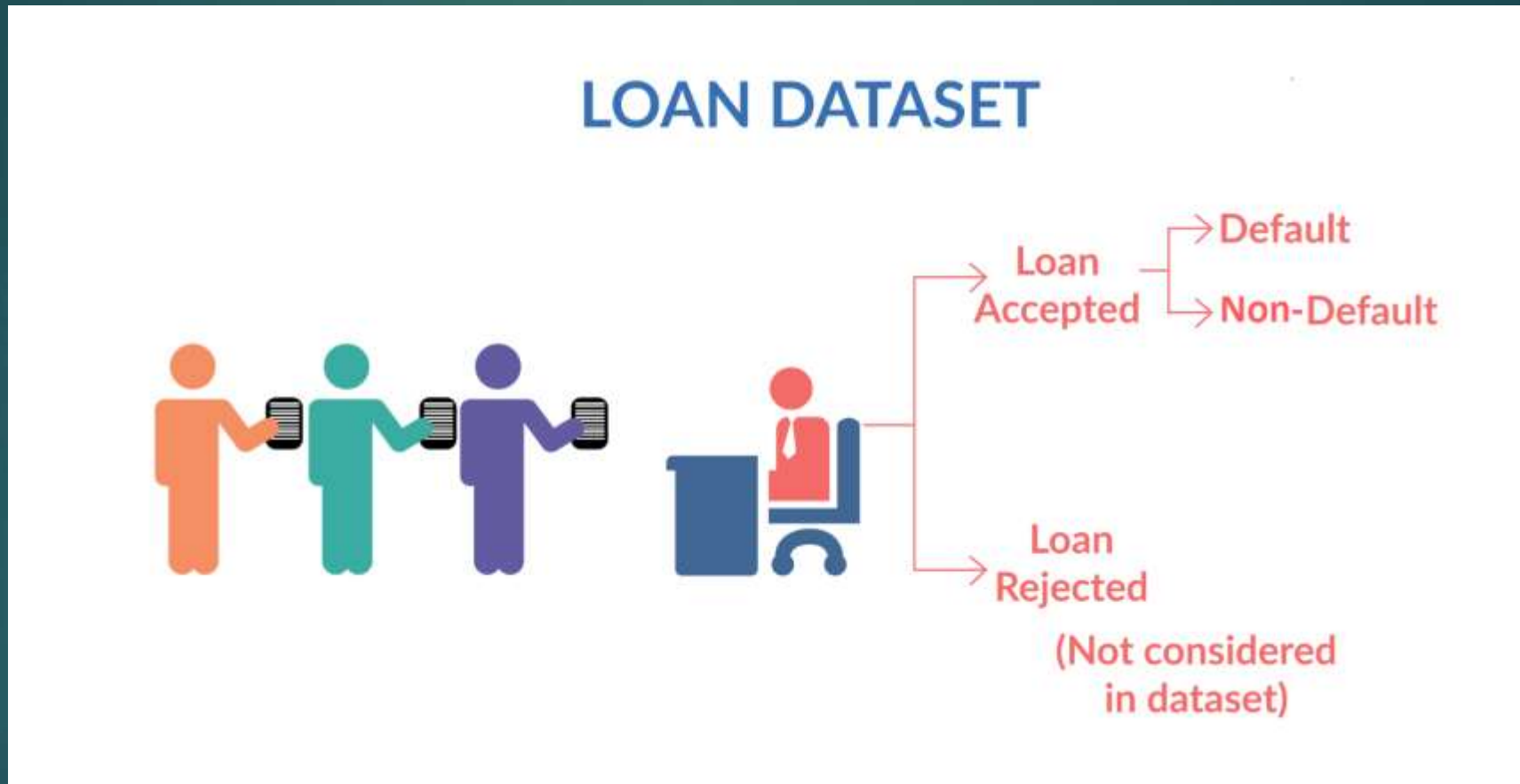# LENDING CLUB CASE STUDY

HARSH RAJ PODDAR

Business Understanding

When the company receives a loan application, 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

The Objective is to understand the real world problem and implement EDA technique to solve those problems.

Benefits of case study:

▶ Gives a idea about how to solve real world problem using EDA.

▶ Develop a basic understanding of risk analytics in banking and financial services

▶ Understand how data is used to minimise the risk of losing money while lending to customers.
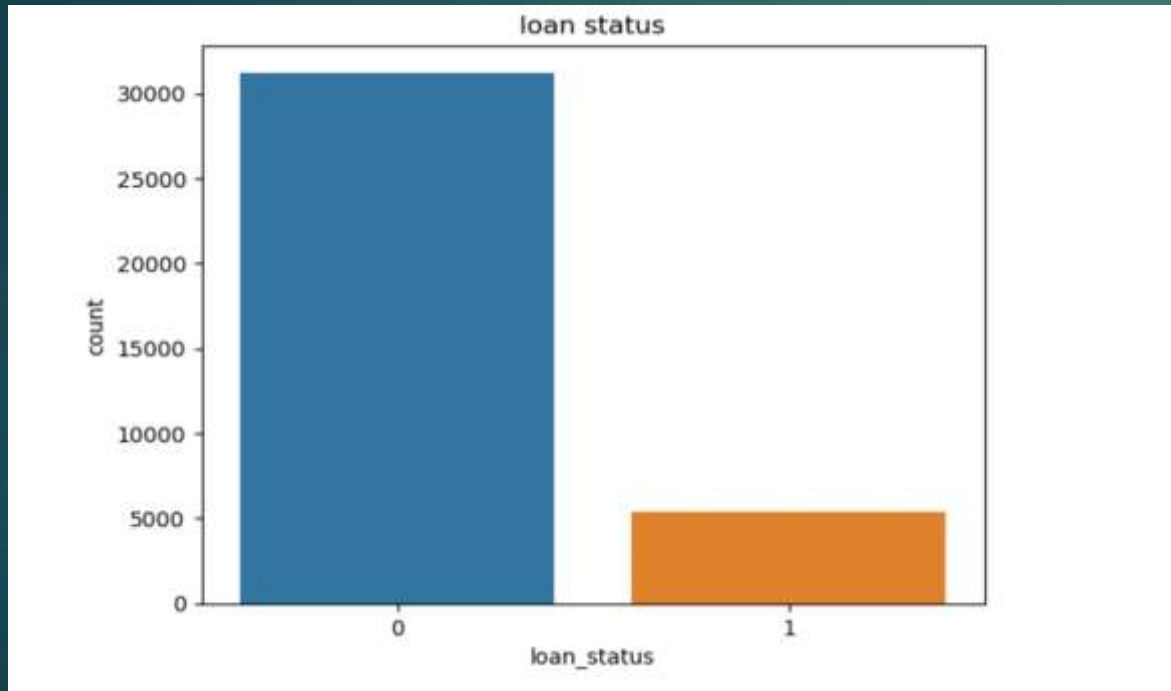
Process

There are four major parts that are needed to be done for this case study:

▶ 1. Data understanding

▶ 2. Data cleaning (cleaning missing values, removing redundant columns etc.)

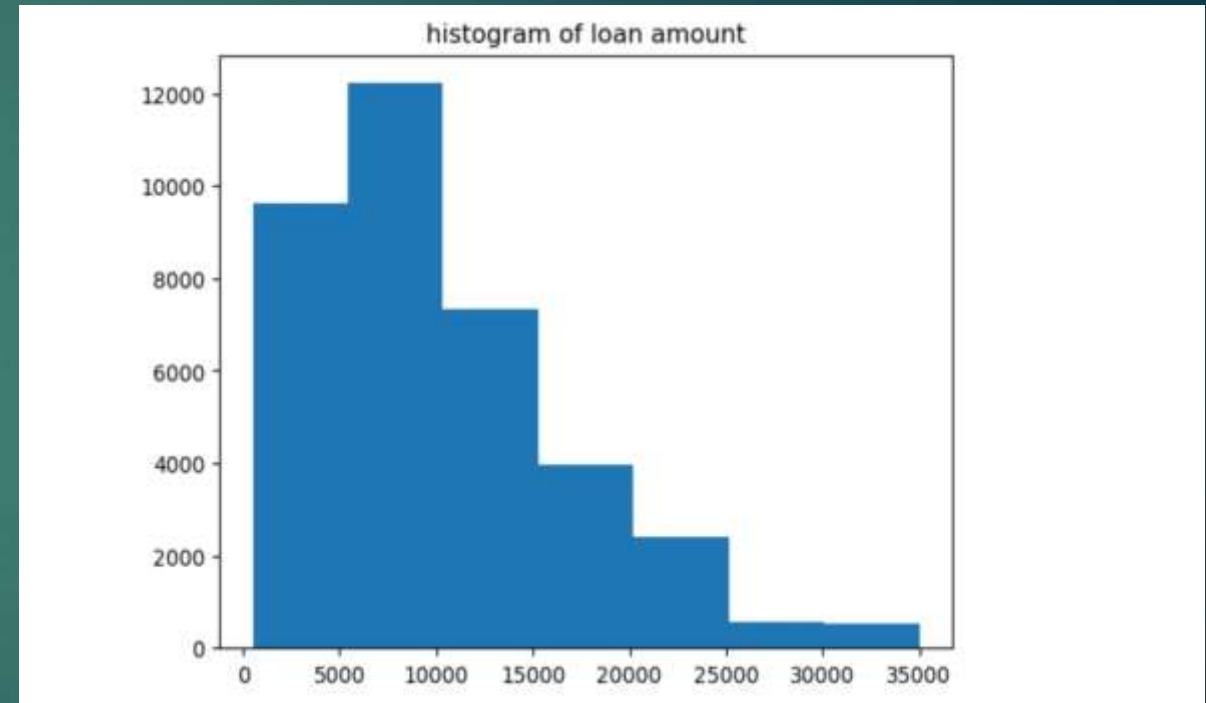▶ 3. Data Analysis

▶ 4. Recommendations
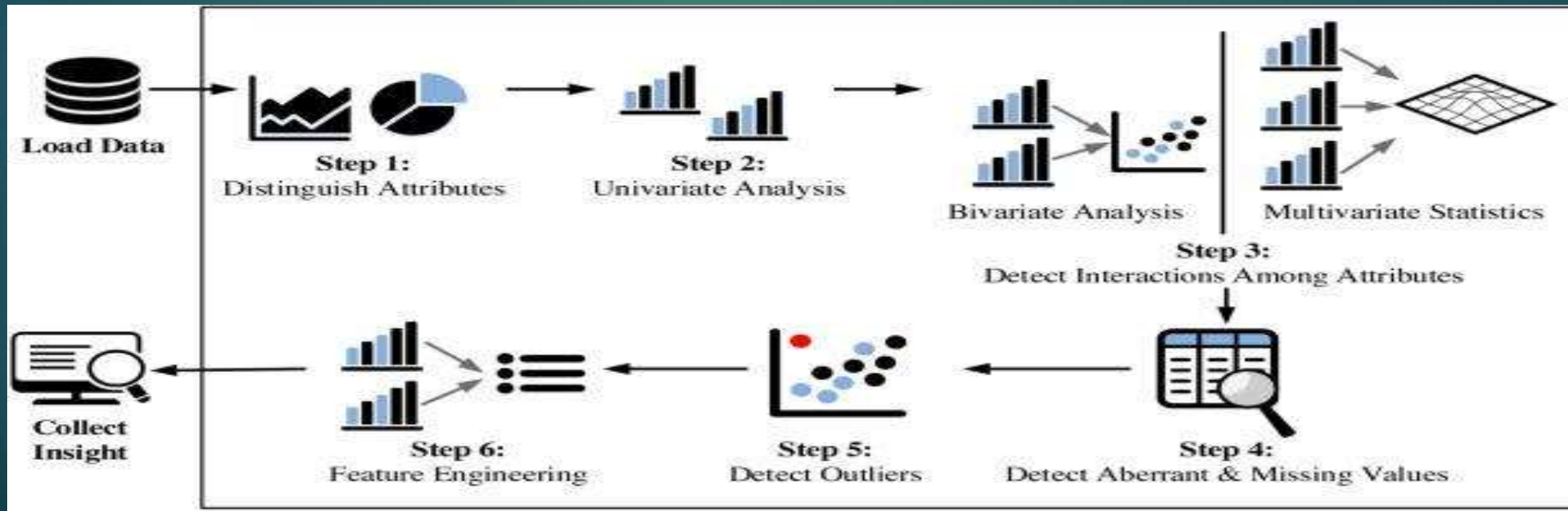
Univariate Analysis
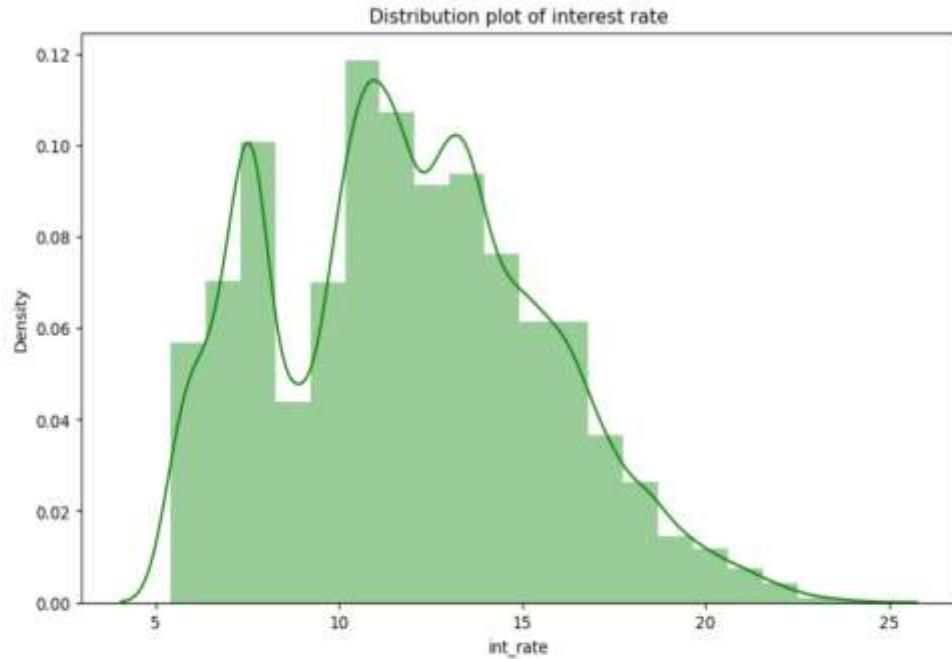
Loan status and Loan amount

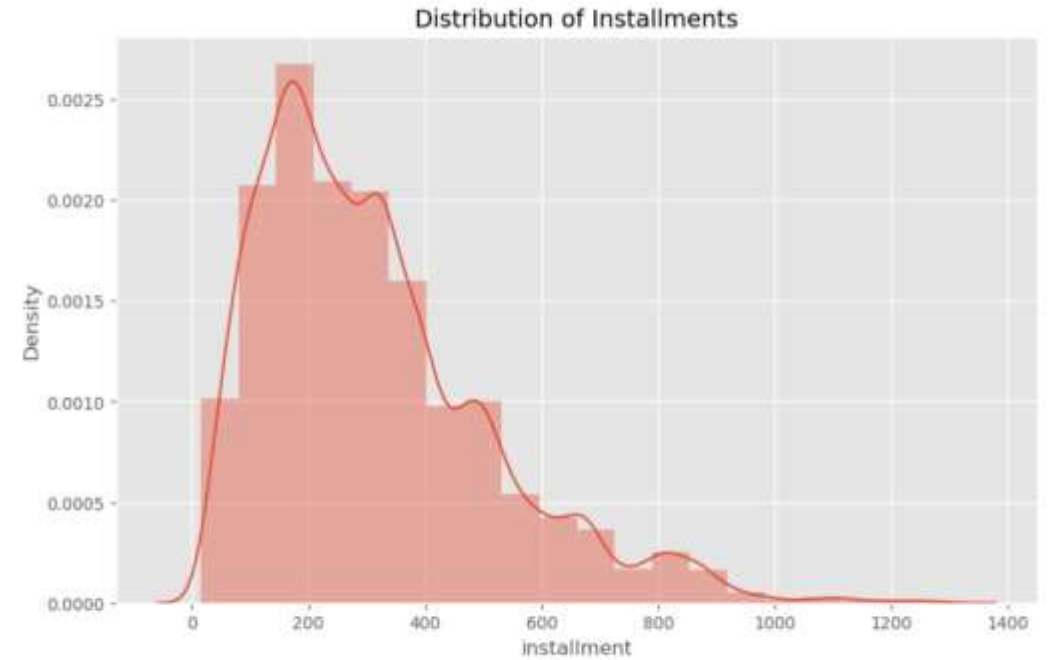The default count is around 14.8%. i.e. 6 times lesser than fully paid count.

75% of loan amount is less than 15000.

Interest Rate and Installments

Distribution plot of interest rate

inference: the interest rate is more crowded around 6-8 and 10-15.

Distribution of Installments

inference: the maximum of installments are between 50 to 400

Interest rate is more crowded around 6-8% and 10-15%.

The maximum of instalments is between 50 to 400.

Grades and Sub Grades

Count of each grade
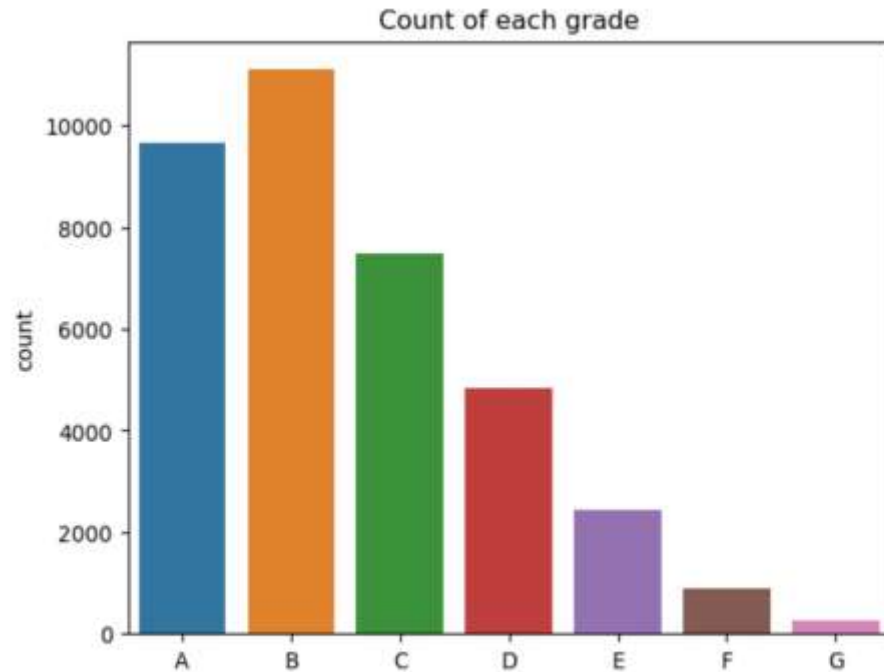
inference: A large amount of grades are A and B compared to rest.

Count of sub grade

inference: mostly the sub grades are A to C.

A and B grades have larger count meaning good rated loans are more.
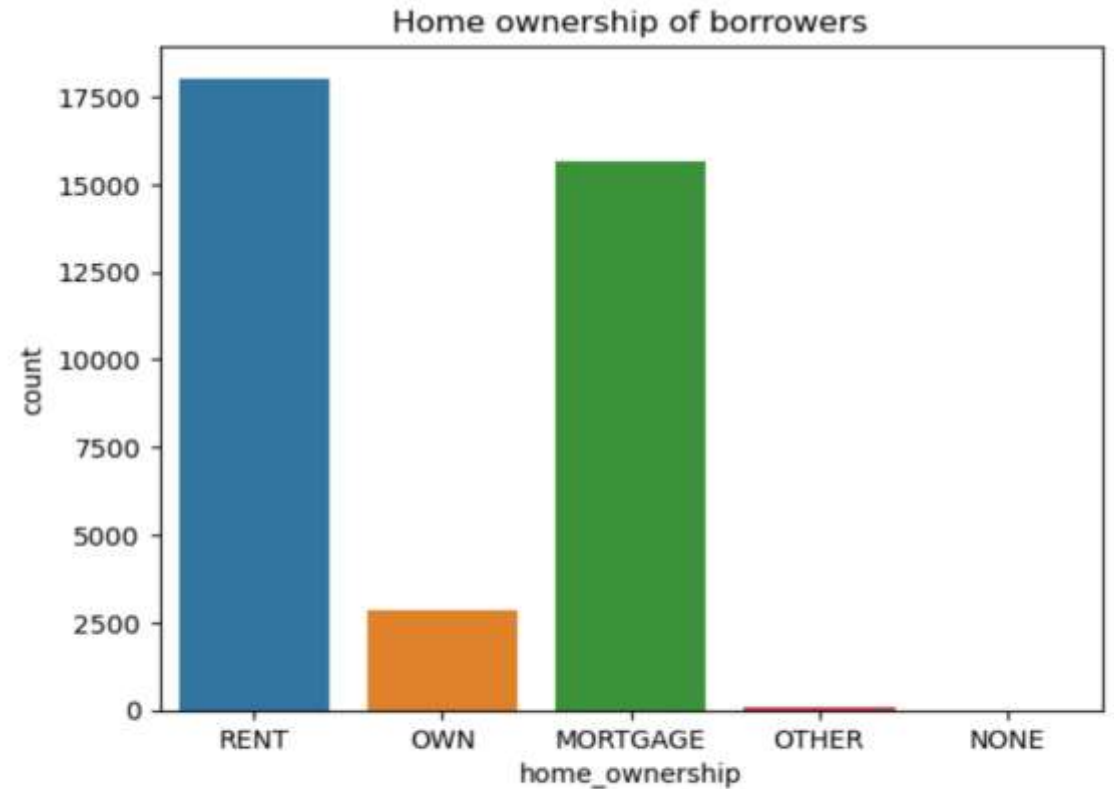
Mostly the sub grades are A and C.
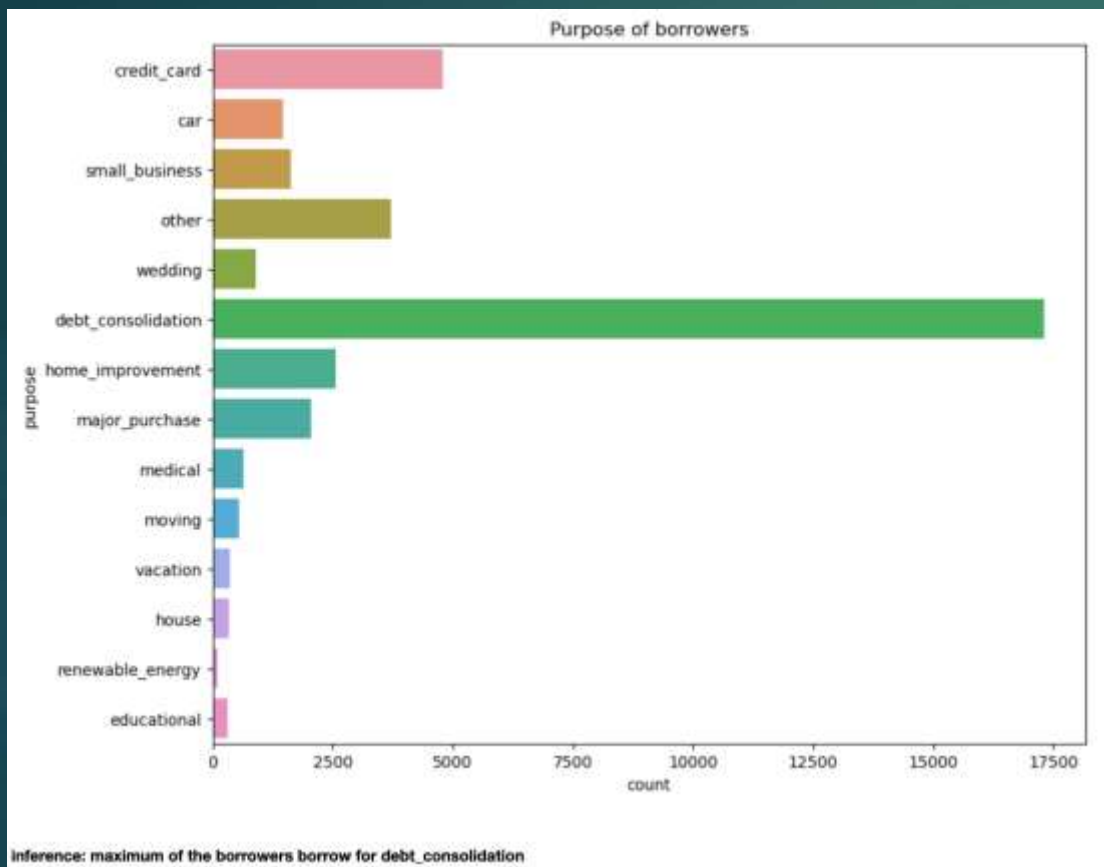
Employment length and Home ownership
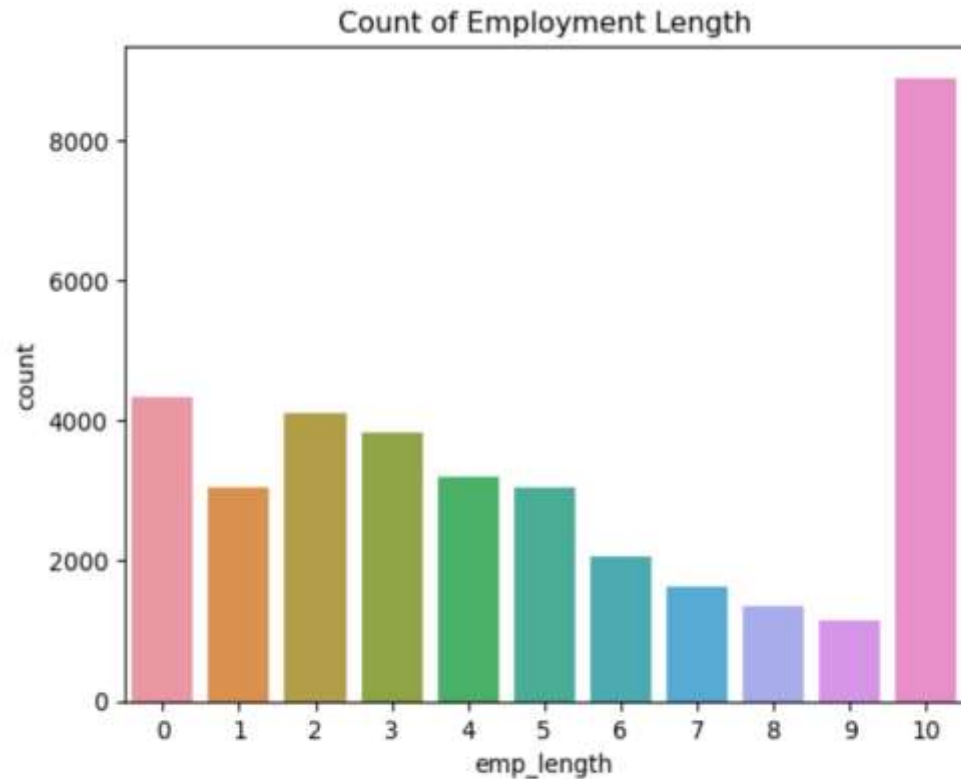
inference: most of the loan borrowers have working experience of more or equal to 10 years

inference: most of the loan borrowers are either paying rent or have mortgage

Purpose of borrowers and DTI

Purpose of borrowers

inference: maximum of the borrowers borrow for debt_consolidation

distributiton of dti

inferene: dti indicated the debt to income ratio thus, it indicates that majority of borrowers have more debt to income. 15 is the highest ratio

Debt consolidation is the major purpose of applying loan
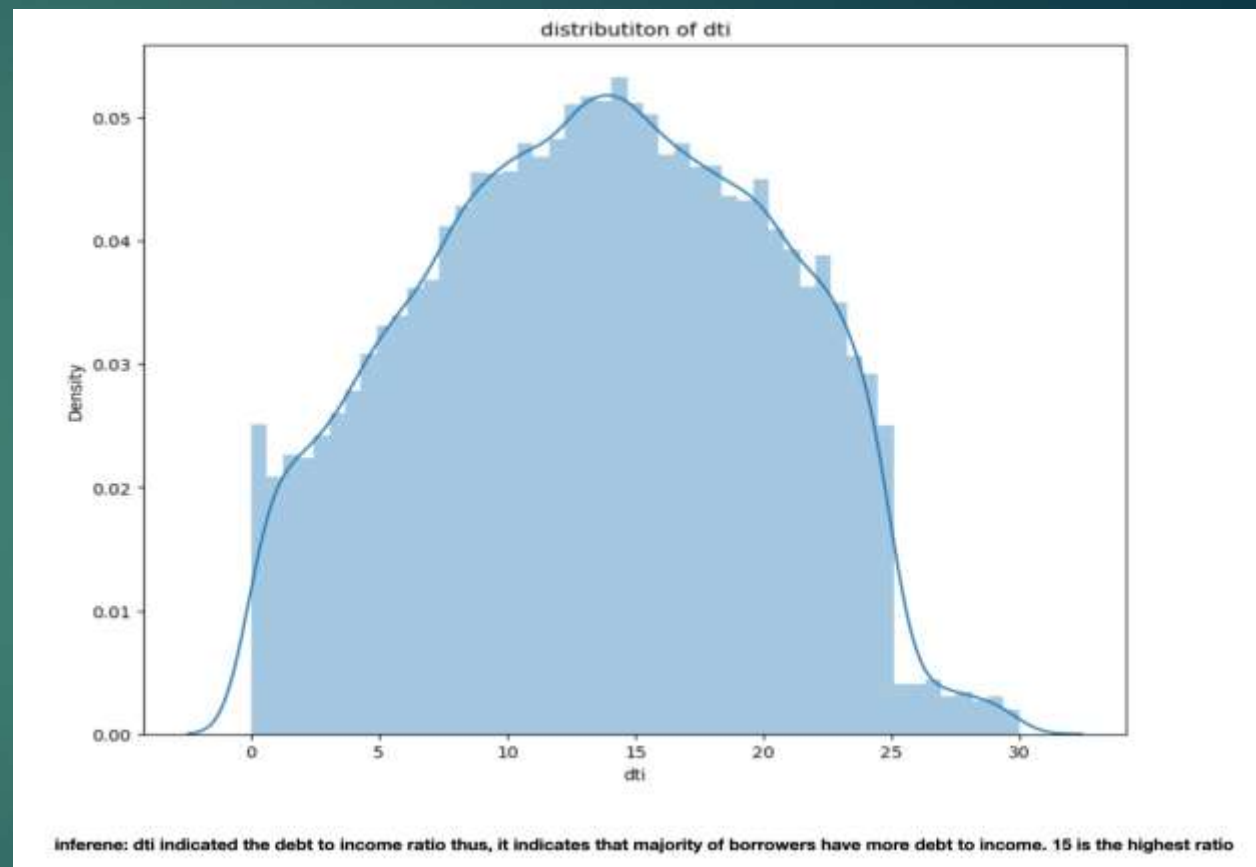
dti indicates the debt to income ratio thus, it indicated that majority of borrowers have more debt to income. 15 is the highest ratio

Issue Year and Issue Month

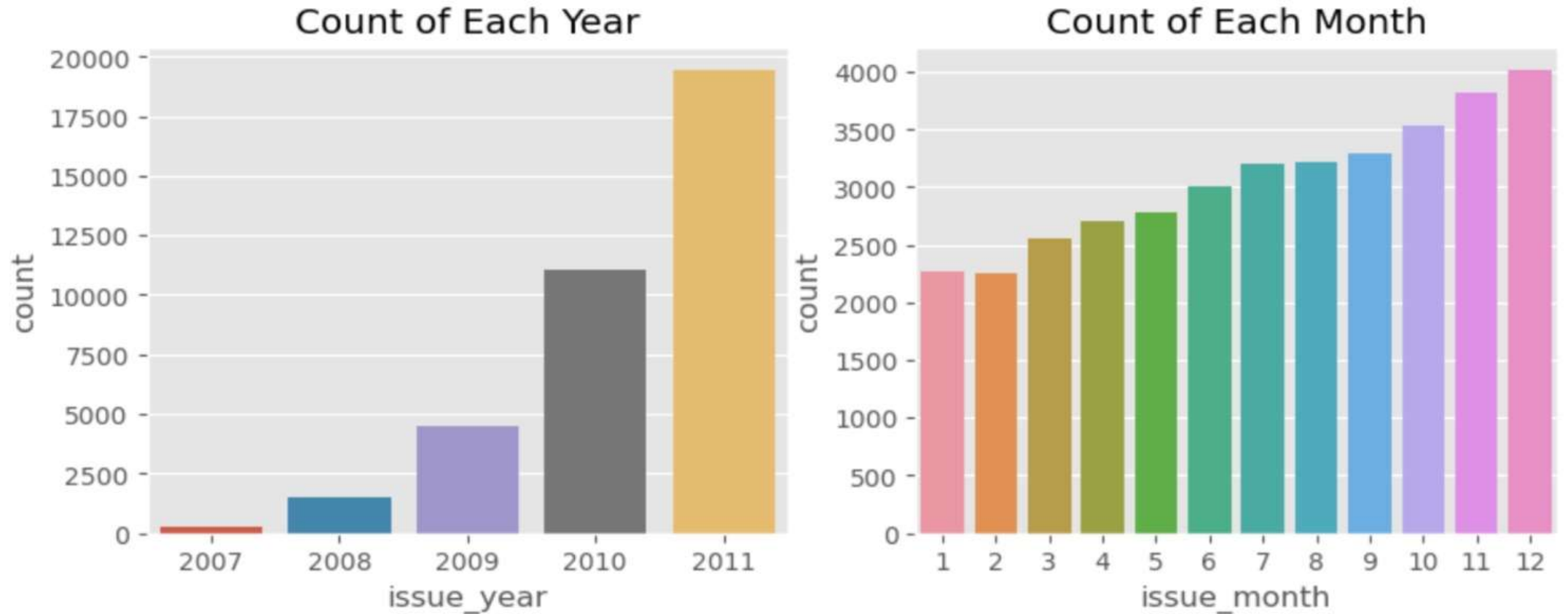inference: with each year , the lending of loans increases and also the lending of loan is more in last quarter of the year

Understanding from univariate analysis

▶ The number of defaulted loan is 6 times less than the number of fully paid loan.

▶ The 36 month term is 3 times more than 60 months. meaning people generally take loans for shorter duration

▶ The interest rate is more crowded around 6-8 and 10-15.

▶ A large amount of loans are with grade 'A' and 'B' compared to rest showing most loans are high grade loans.

▶ Majority of borrowers have working experience greater than 10 years

▶ .Majority of borrowers don't posses property and are on mortgage or rent.

▶ About 50% of the borrowers are verified by the company or have source verified.

▶ A large percentage of loans are taken for debt consolidation followed by credit card

▶ DTI indicated the debt to income ratio thus, it indicates that majority of borrowers have more debt to income. 15 is the highest ratio

▶ Majority of the loans are given in last quarter of the year. The number of loans approved increases with the time at exponential rate, thus we can say that the loan approval rate is increasing with the time.

Segmented Univariate Analysis.

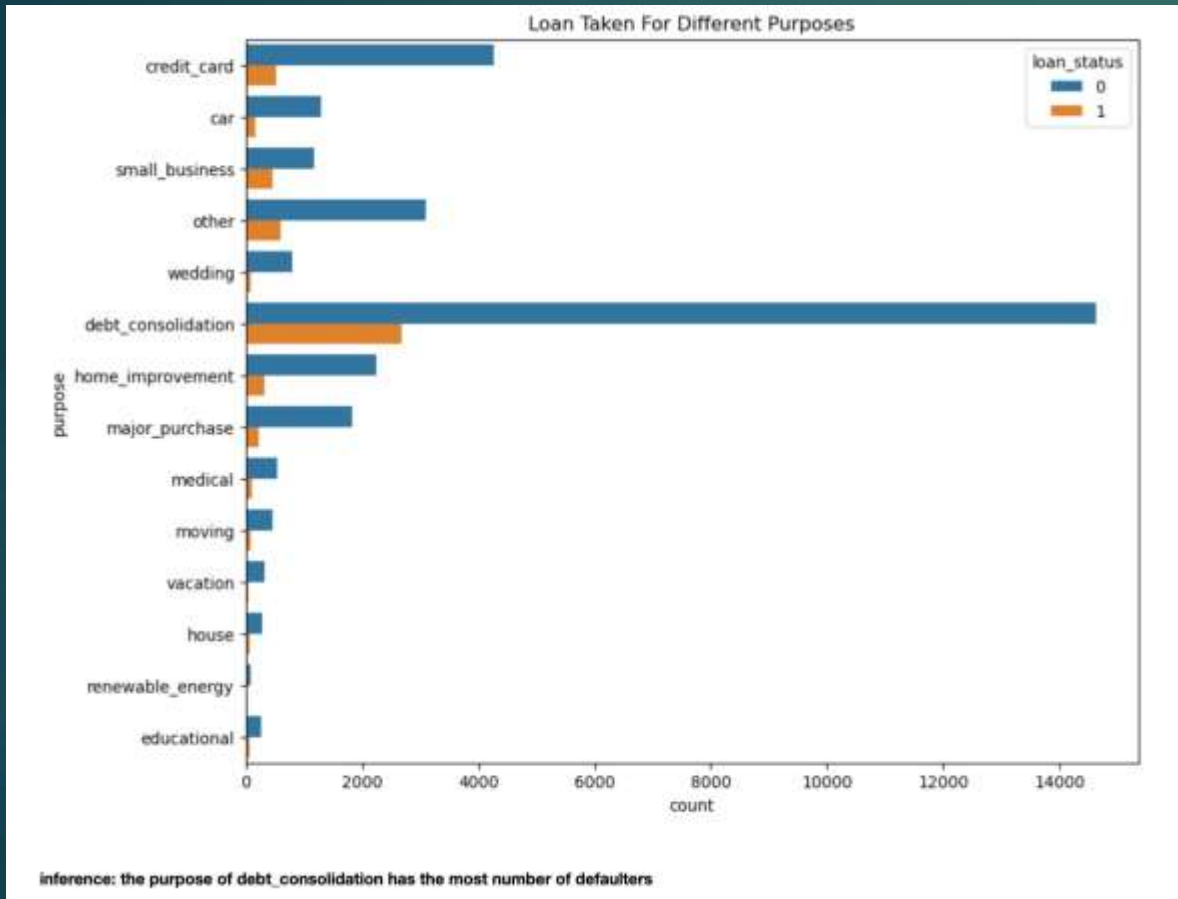Loan status for different grades and home ownership
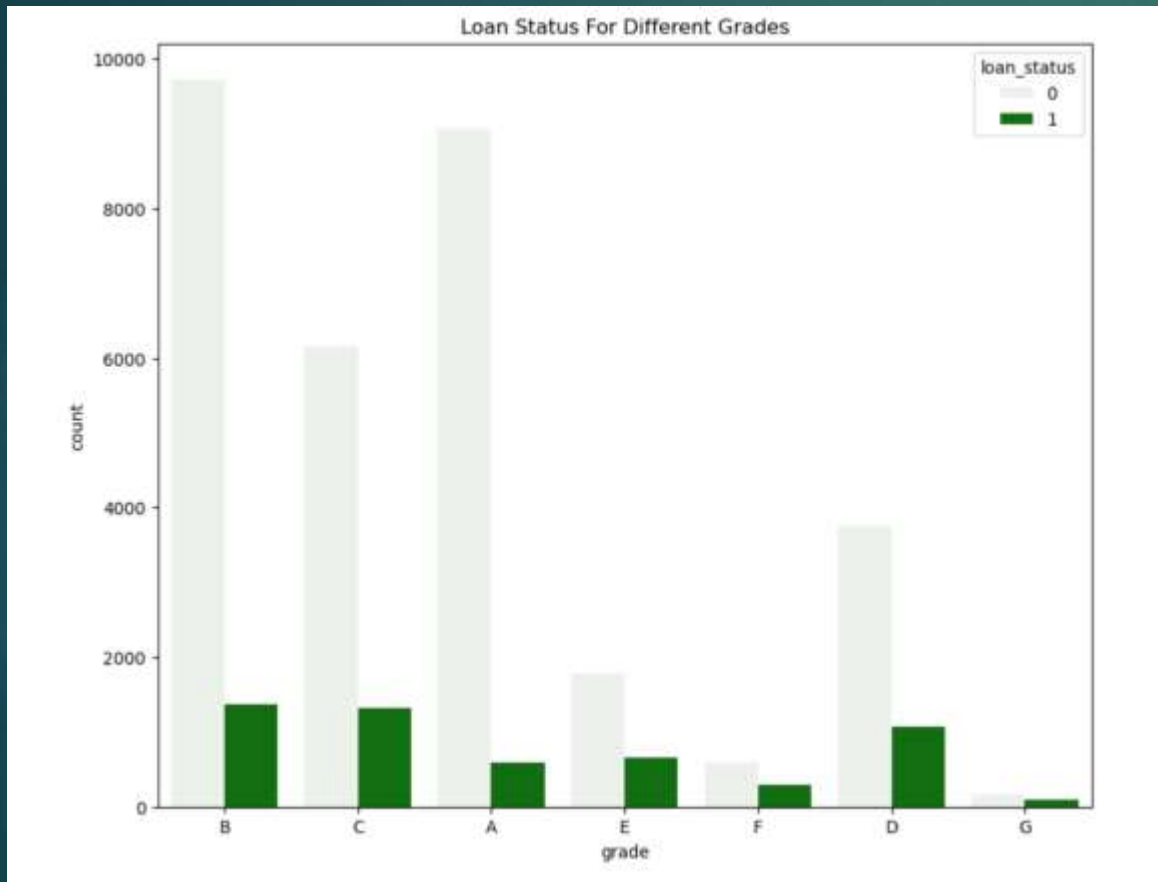
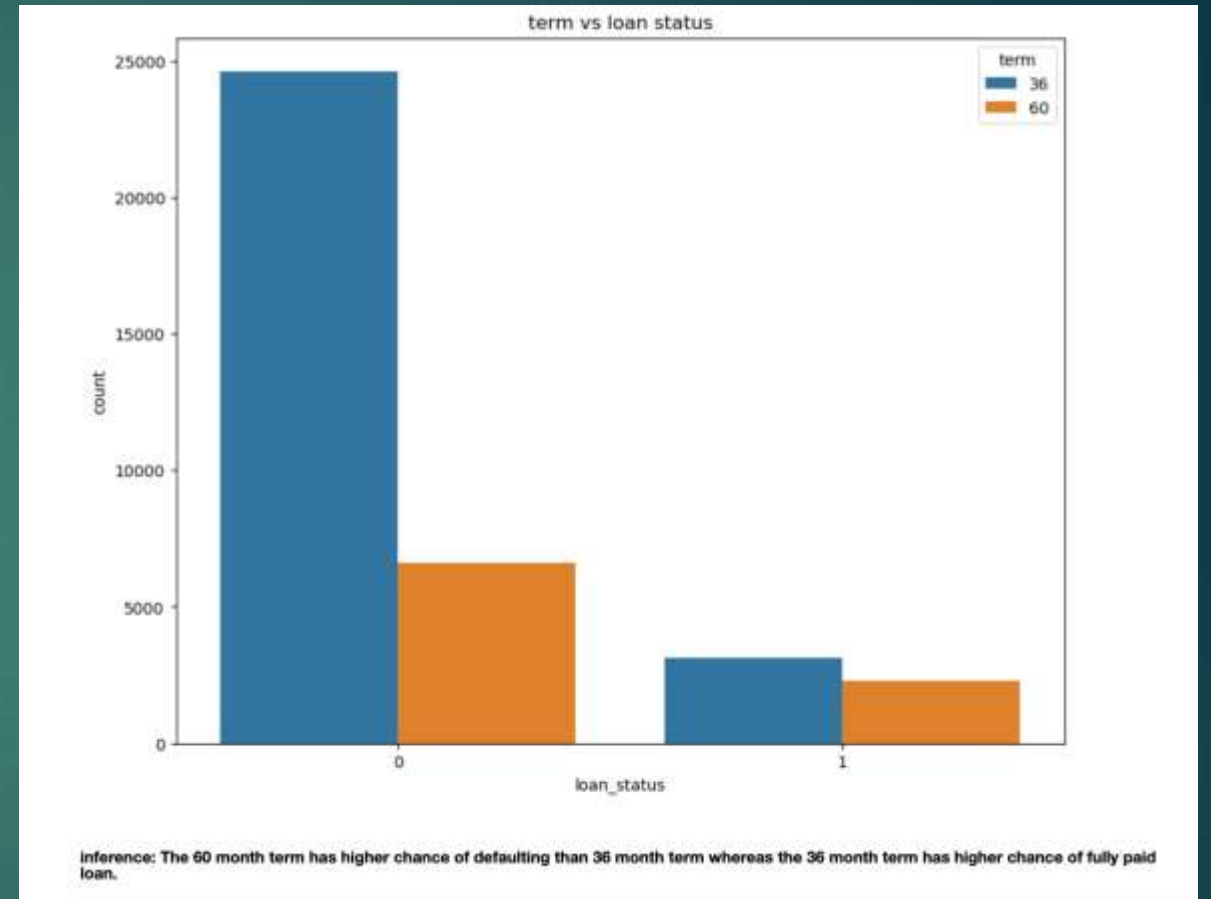Grades B, C , D have the most defaulters.
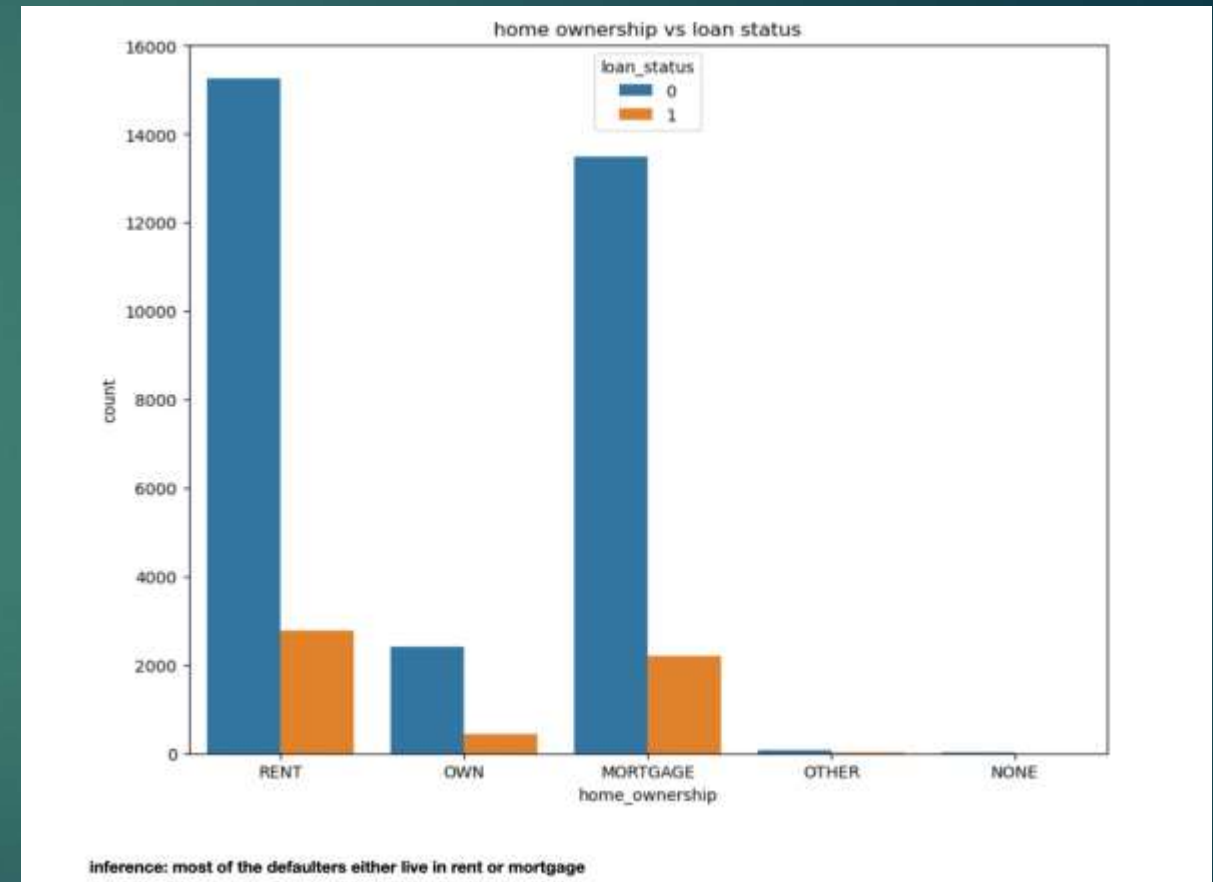
Most of the defaulters rather live in rent or mortgage

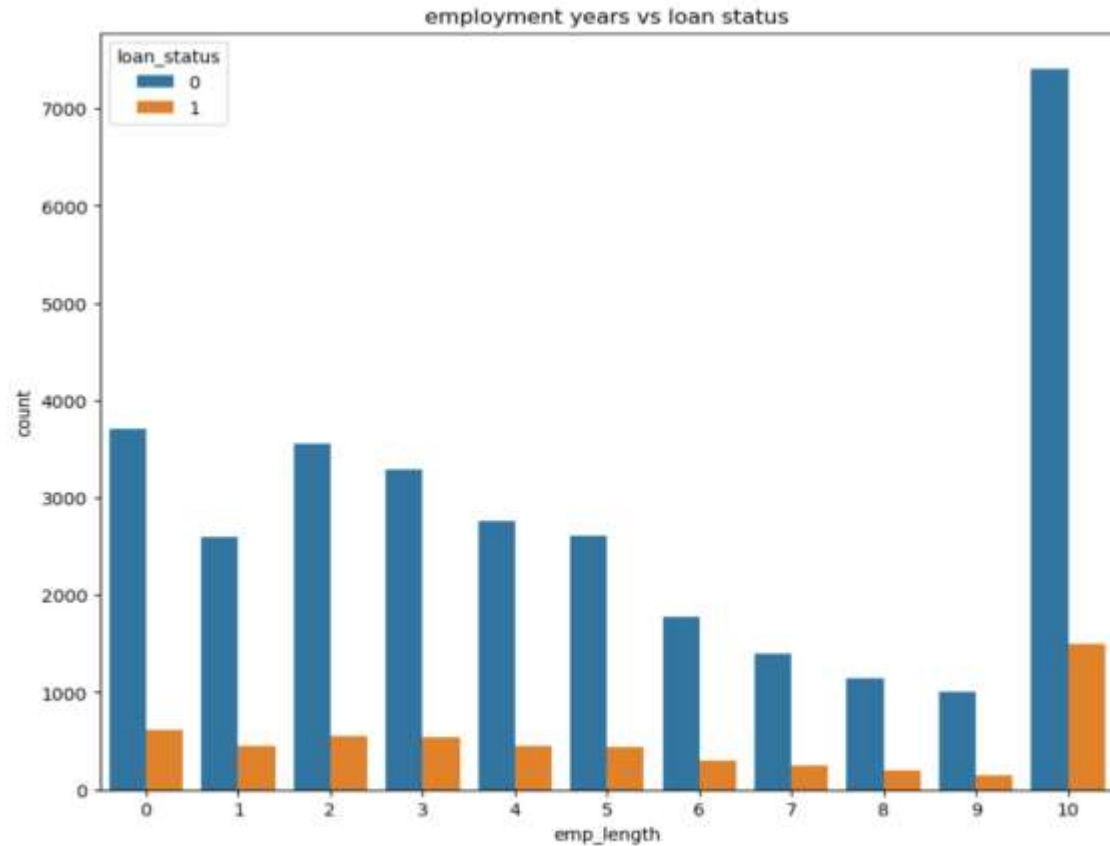# Loan taken for different purpose and term vs loan status.

The purpose of borrowing loan for debt consolidation has the most number of defaulters.

The 60 month term has higher chance of defaulting than 36 month term whereas the 36 month term has higher chance of fully paid loan.

Employment year and loan amount vs loan status

employment years vs loan status

inference: The Employees with 10 or 10+ years of experience are likely to default and have higher chance of fully paying the loan.

Loan Amount vs Loan Status
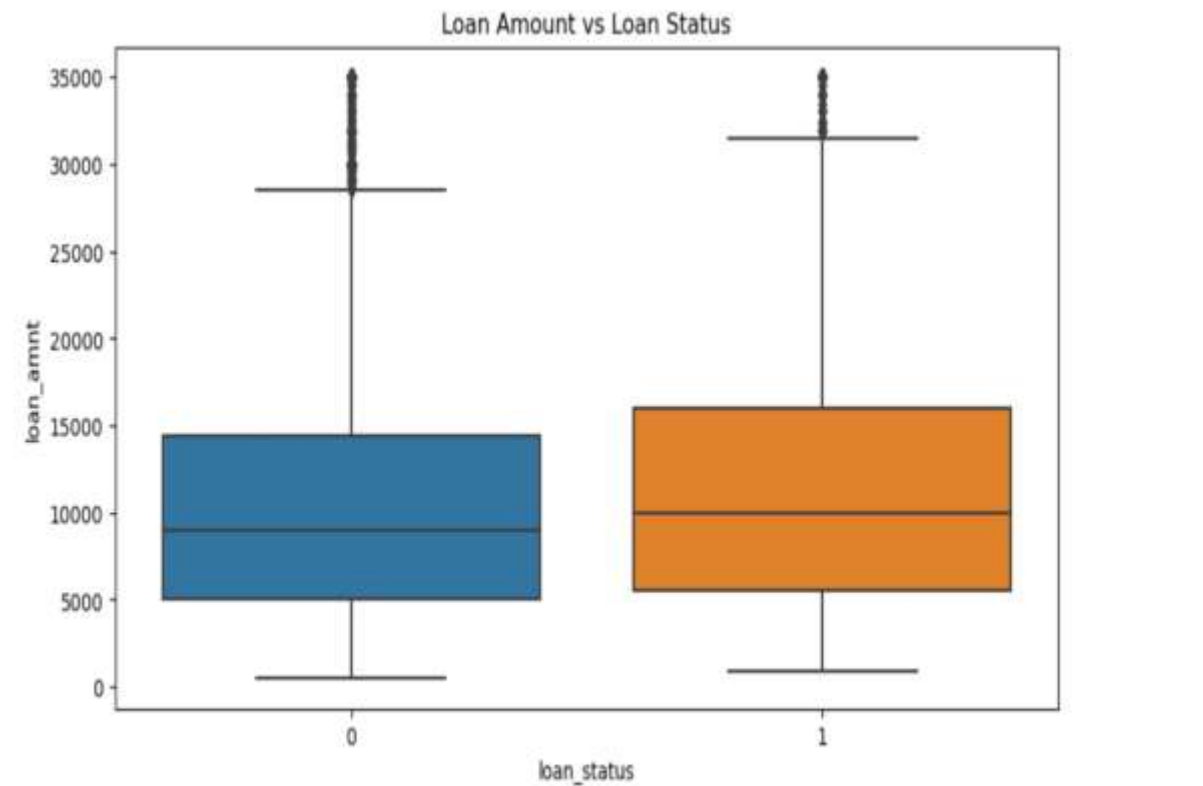
inference: the the median is almost same for defaulters and fully paid but the 75th percentile is larger for defaulters meaning more loan increases the chances of defaulting ¶

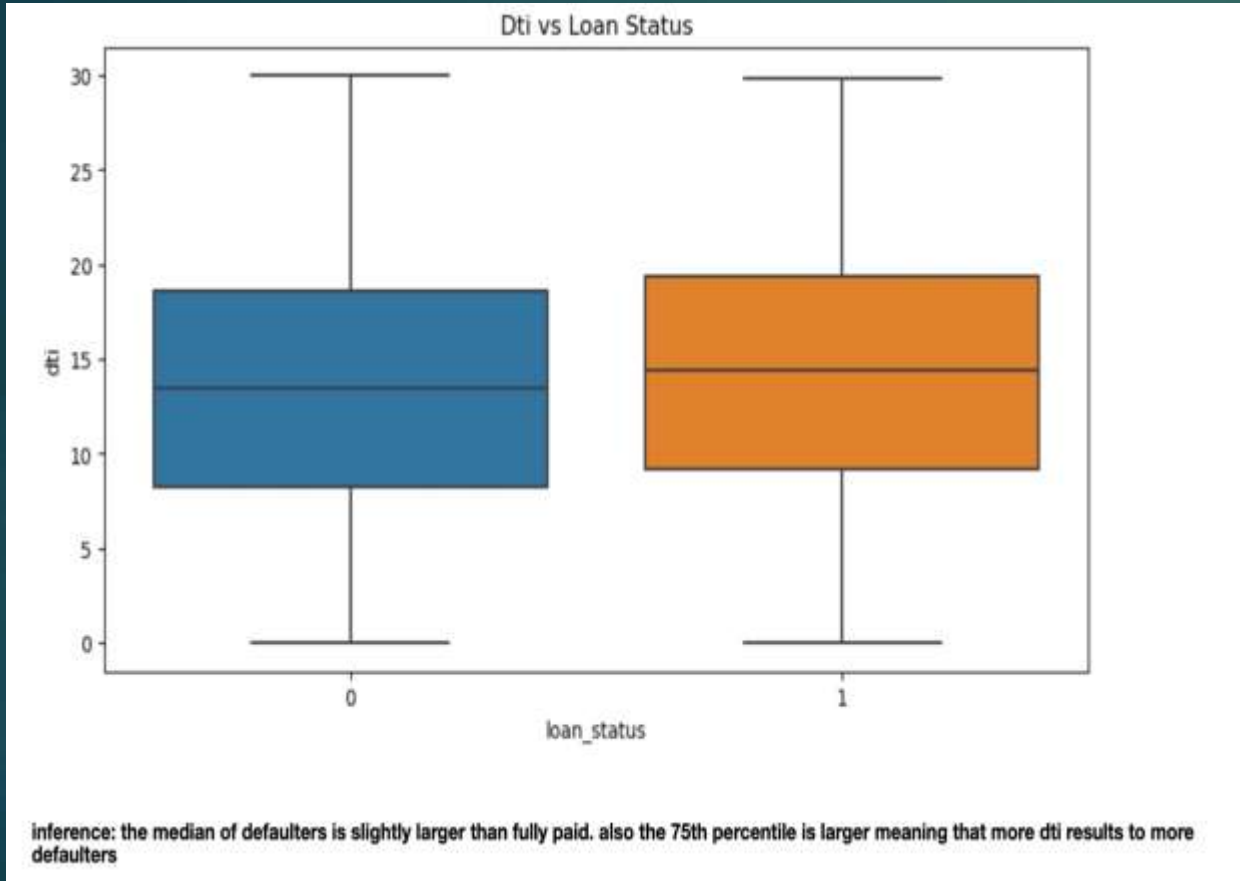The Employees with 10 or 10+ years of experience are likely to default and have higher chance of fully paying the loan

the median is almost same for defaulters and fully paid but the 75th percentile is larger for defaulters meaning more loan increases the chances of defaulting

DTI and Interest rate vs loan status
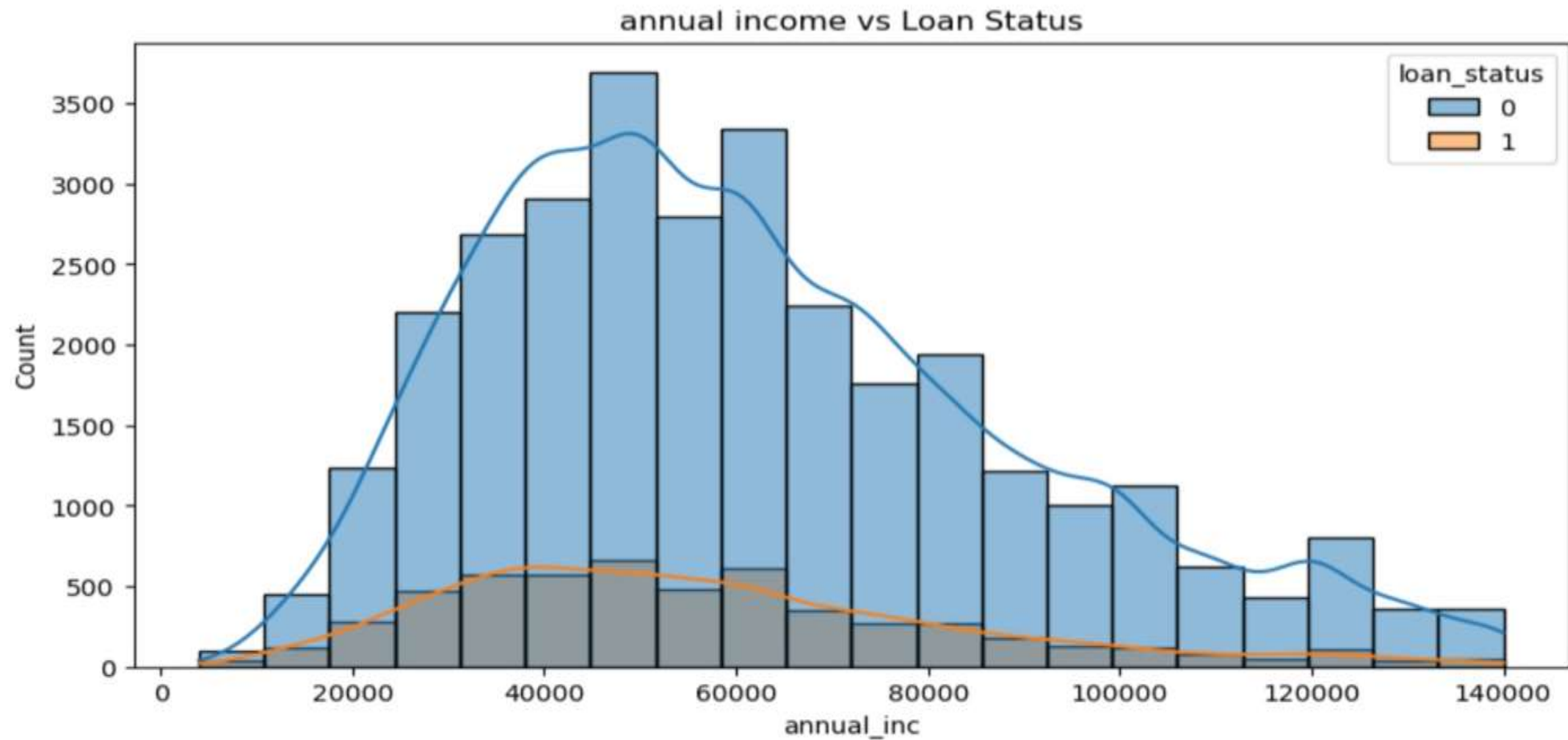

The median of defaulters is slightly larger than fully paid. also the 75th percentile is larger meaning that more DTI results to more defaulters
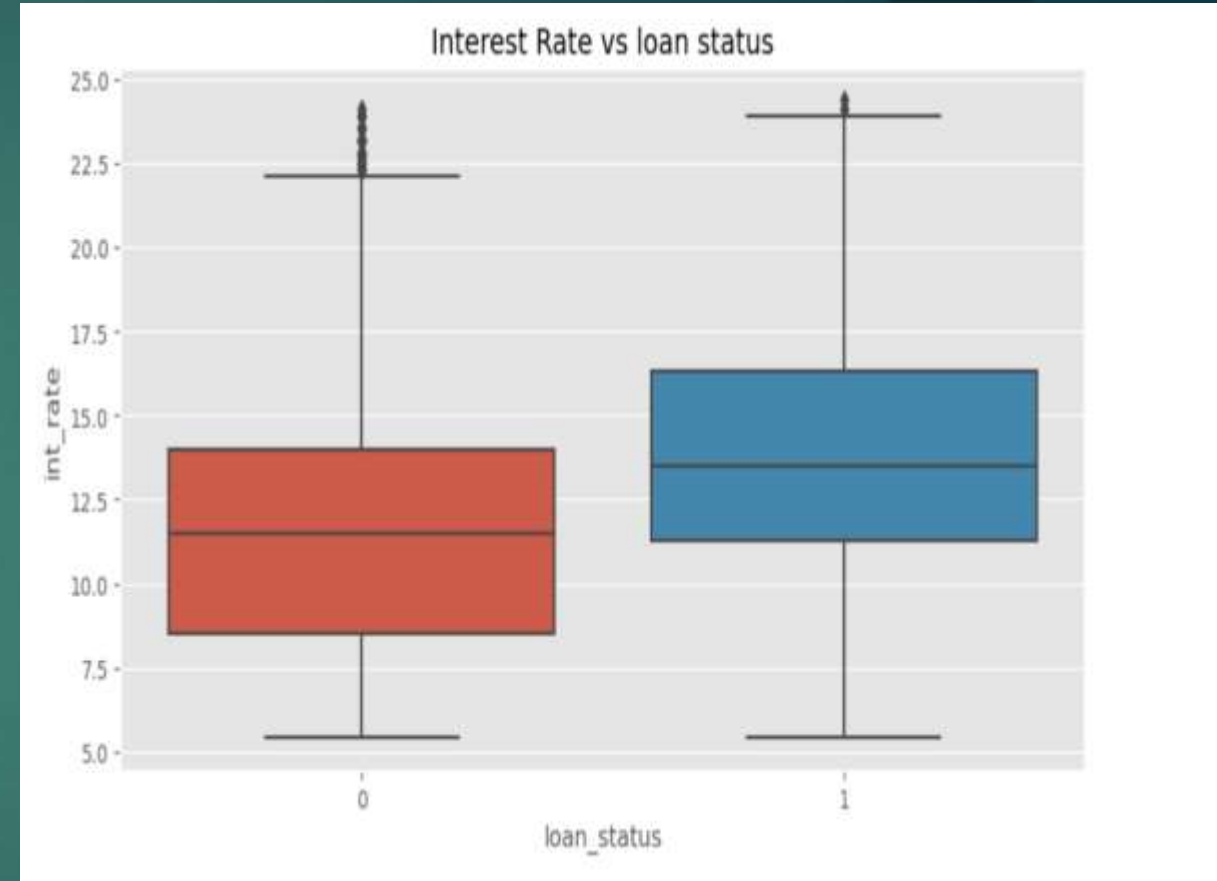
The Interest rate of defaulters is quite high meaning high interest rate tends to have more defaulters

# Annual income vs loan status

annual income vs Loan Status

inference: the annual income of defulters are more between 20000 to 60000

Understanding from Segmented Univariate Analysis.

▶ Grade B, C ,D have the most defaulters

▶ Debt Consolidation is the most popular loan purpose and has highest number of fully paid loan and defaulted loan.

▶ The 60 month term has higher chance of defaulting than 36 month term whereas the 36 month term has higher chance of fully paid loan.

▶ most of the defaulters either live in rent or mortgage.

▶ more loan amount increases the chances of defaulting.

▶ The 75th percentile of defaulters is slightly larger than fully paid meaning that more DTI results to more defaulters.

▶ The Fully paid loan are increasing exponentially with the time compared to defaulted loan.

▶ The default loan amount increases with interest rate.

▶ The Employees with 10+ years of experience are likely to default and have higher chance of fully paying the loan.

Bivariate Analysis

DTI vs grade and Interest rate groups vs grade For loan status
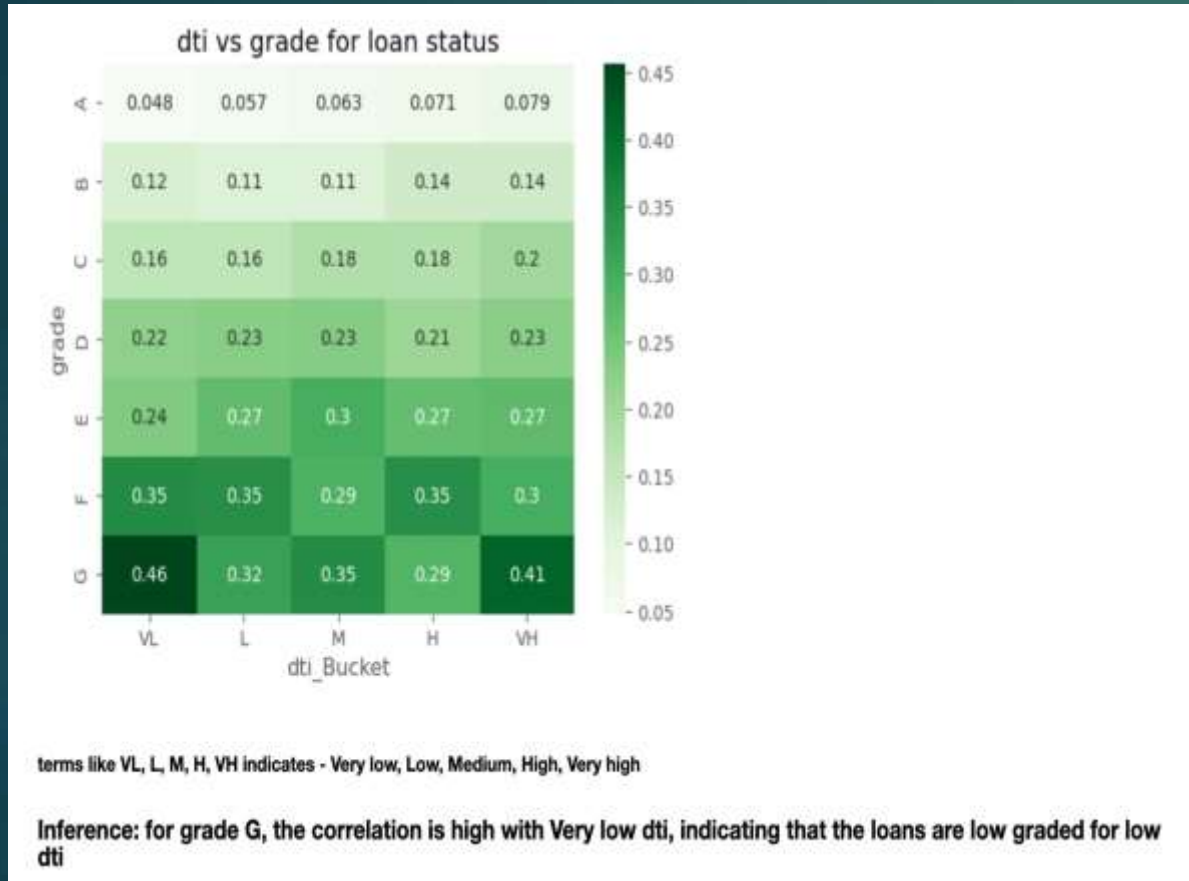
dti vs grade for loan status

terms like VL, L, M, H, VH indicates - Very low, Low, Medium, High, Very high

Inference: for grade G, the correlation is high with Very low dti, indicating that the loans are low graded for low dti

Interest rate groups vs grades for loan status

inference: The lower grades like F,G are higher for high,very high interest rate indicating that for bad graded loans the interest rates are higher

For grade G, the correlation is high with Very low DTI, indicating that the loans are low graded for low DTI
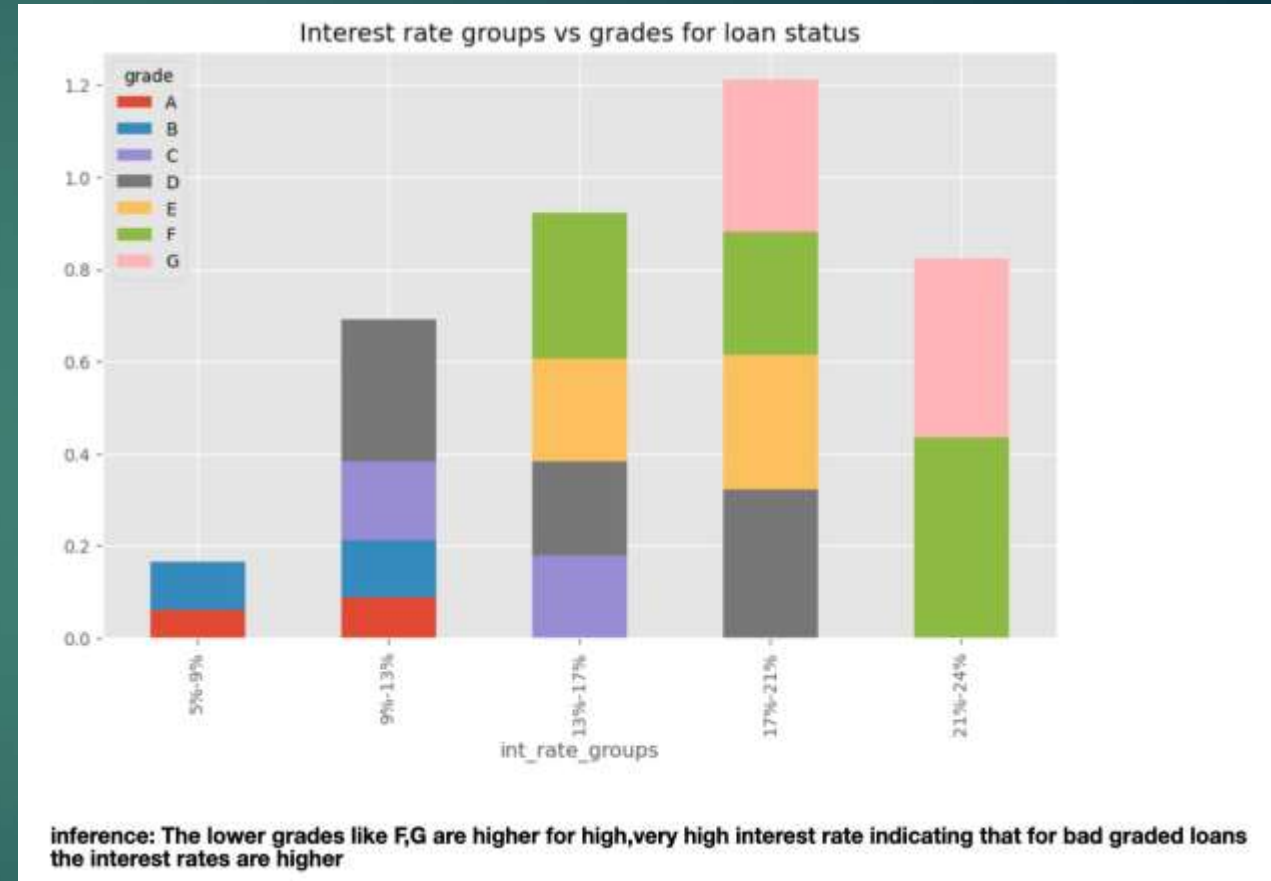
The lower grades like F,G are higher for high , very high interest rate indicating that for bad graded loans the interest rates are higher

Annual income vs home ownership and annual income groups vs loan amount for loan status.

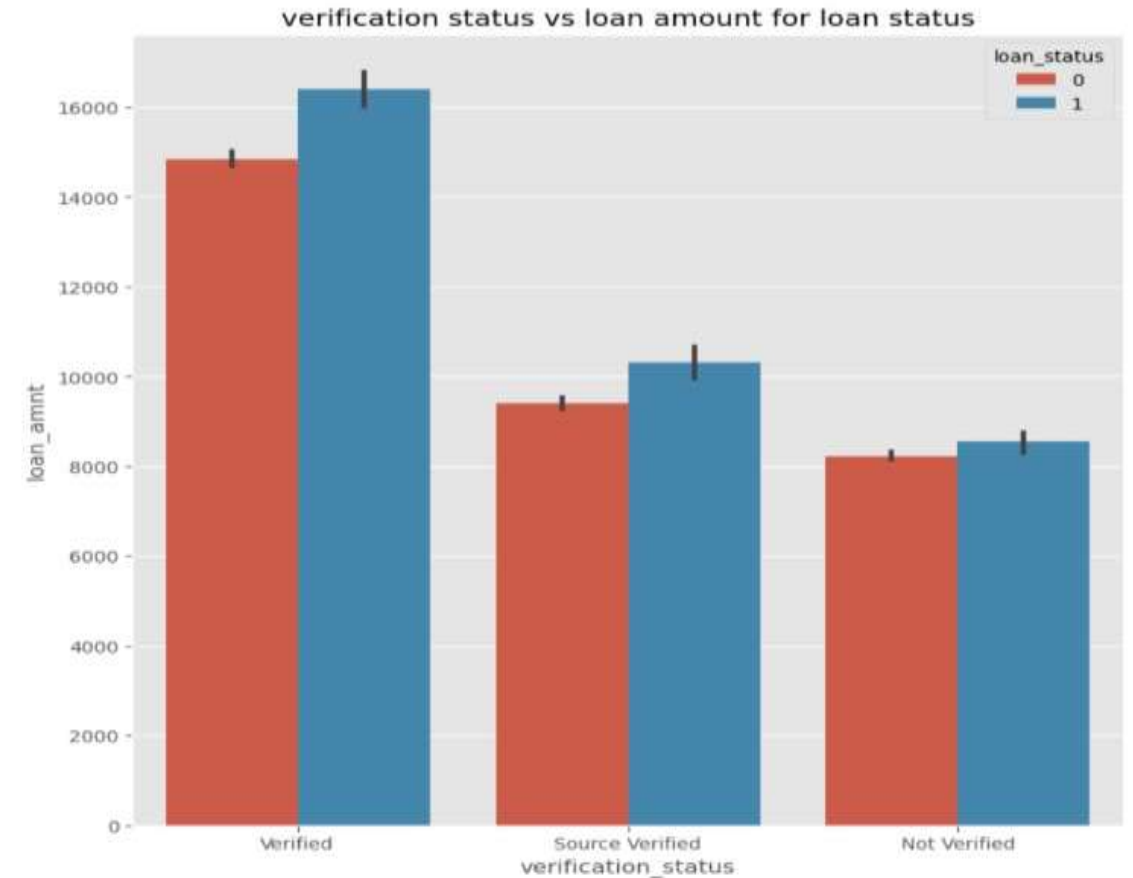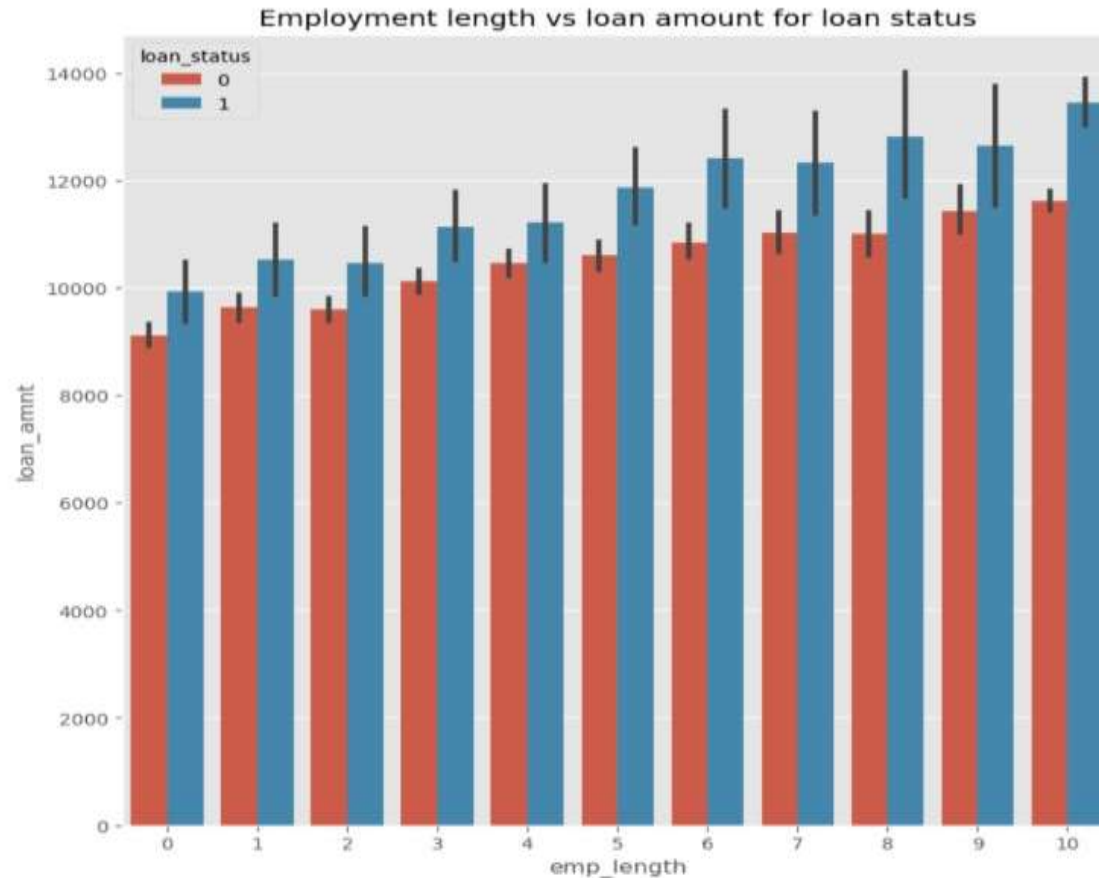annual income vs home ownership for loan status
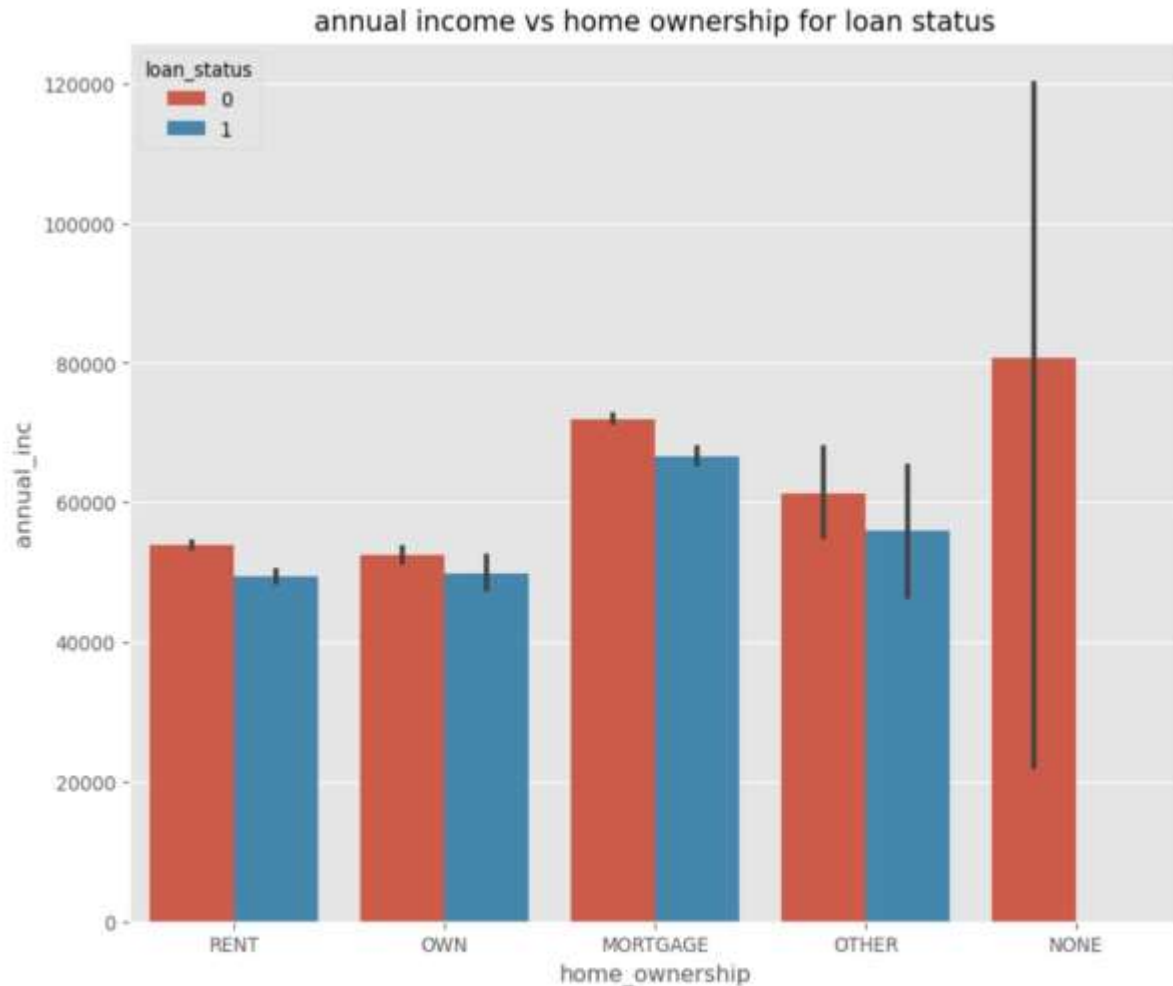
Inference: defaulters are more in case of mortgage

anuual income groups vs loan amount for loan status
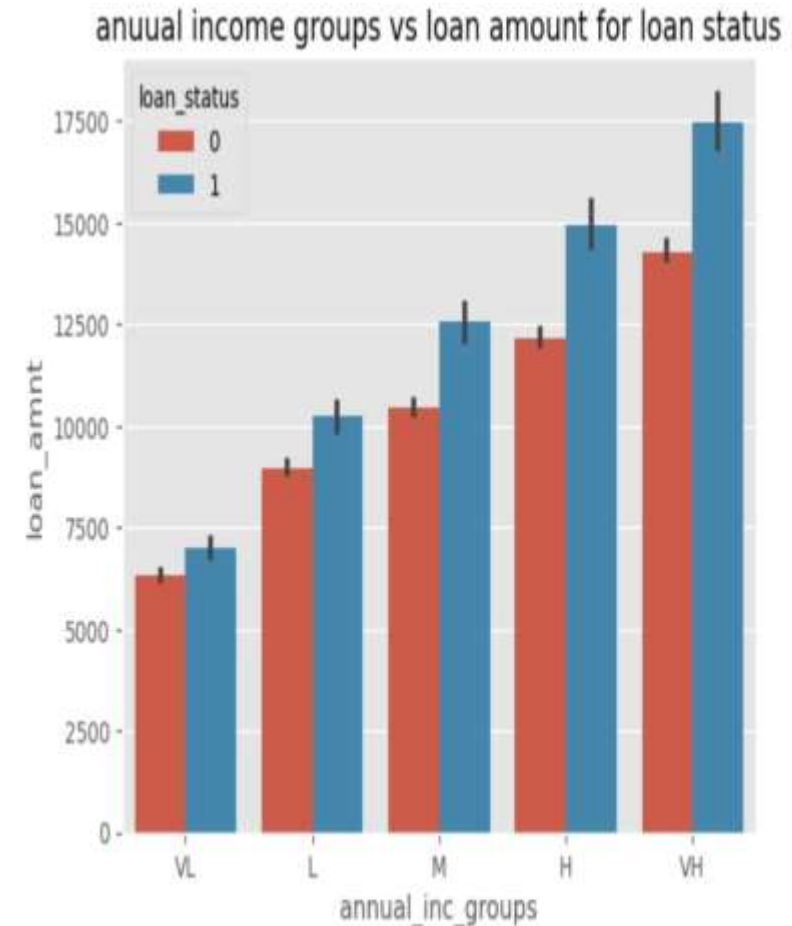
Inference: Across all the income groups, the loan_amount is higher for people who defaulted.

Employment length vs loan amount and verification status vs loan amount for loan status
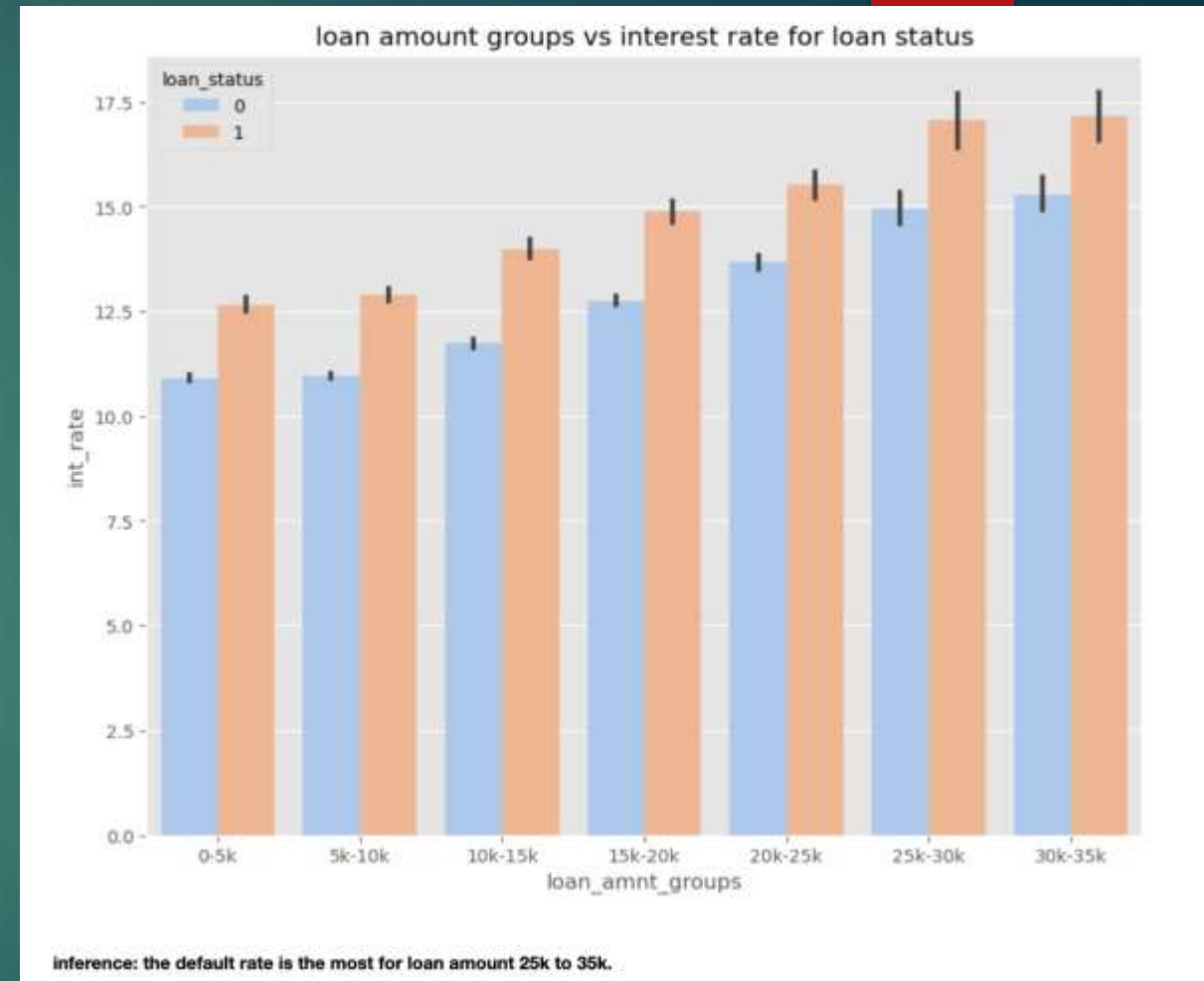

inference: **Employees with longer working history got the loan approved for a higher amount.**

Looking at the verification status data, verified loan applications tend to have higher loan amount.

Loan amount vs purpose and loan amount groups vs interest rate for loan status

loan amount vs purpose for loan status

inference: defaulters are high for small businesses.

loan amount groups vs interest rate for loan status

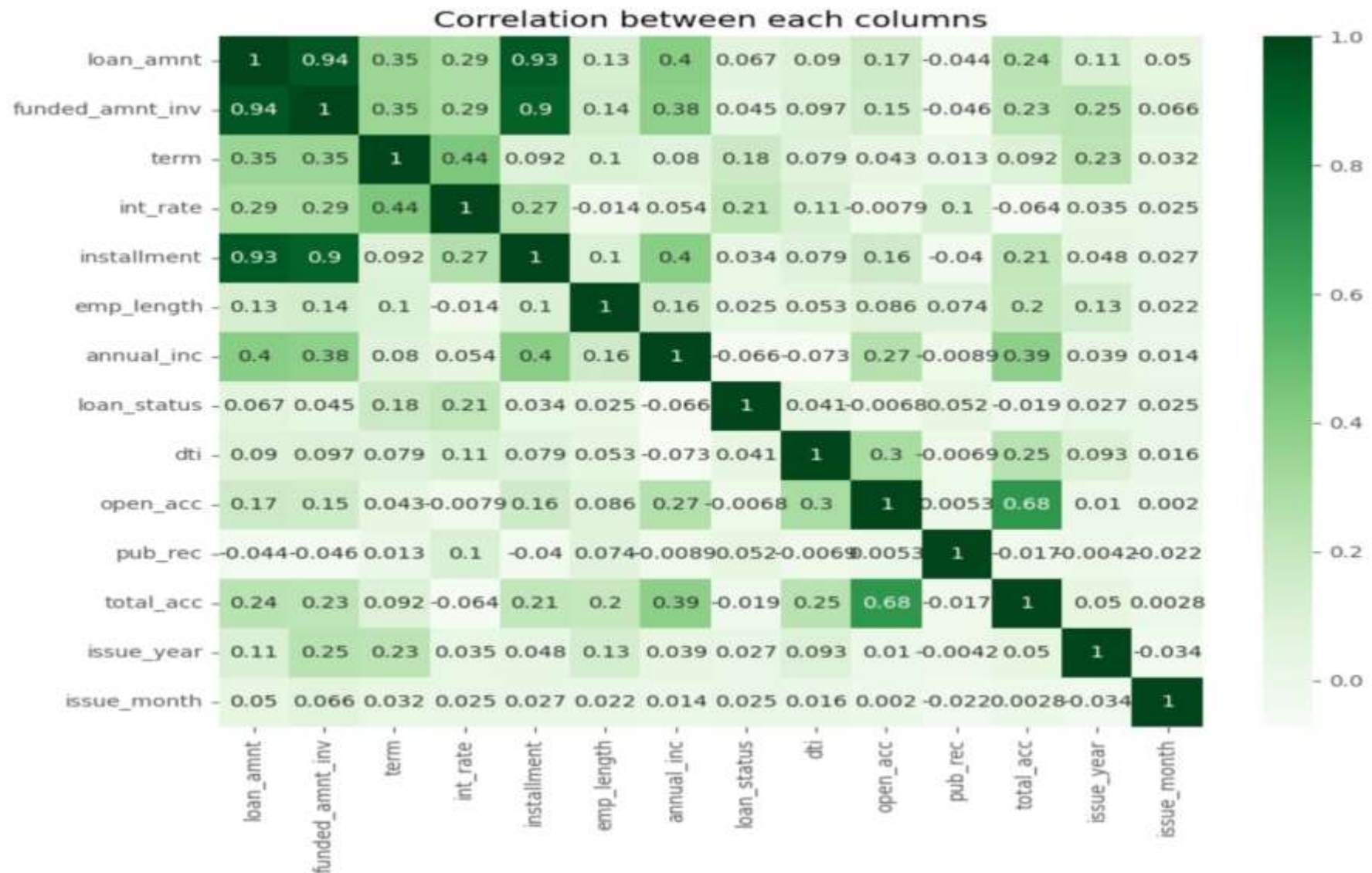inference: the default rate is the most for loan amount 25k to 35k.
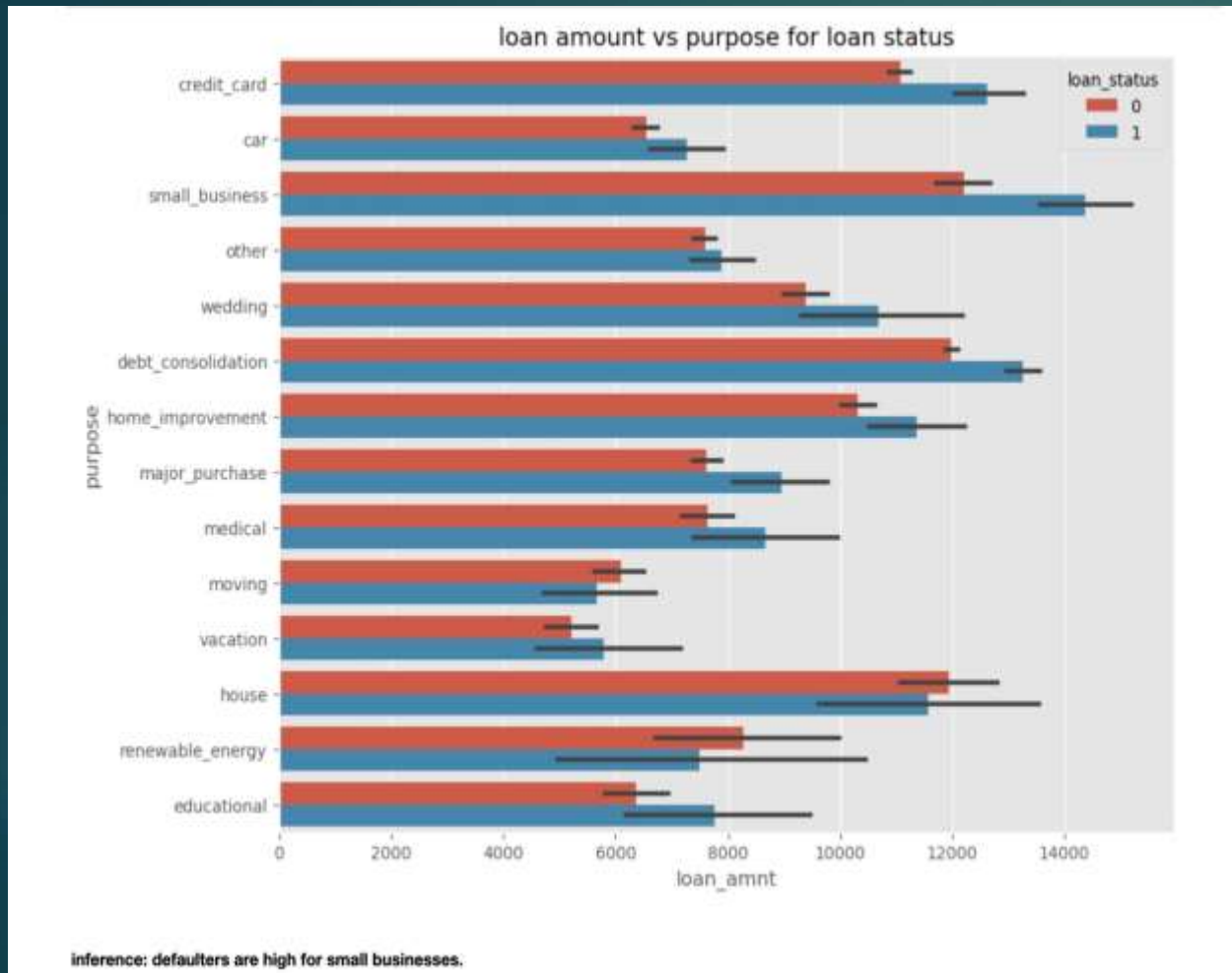
Defaulters are high for small business with loan amount greater than 14k

The default rate is the most for loan amount 25k to 35k with interest rate 15% or above

Correlation between each columns

Correlation between each columns

Understanding of bivariate analysis.

▶ The loans are low graded for low DTI.

▶ The Grade A which is lowest risk also has lowest DTI ratio which we can say that higher grade has lower rate of default.

▶ Interest rate vs grade for loan status indicates that for bad graded loans the interest rates are higher.

▶ Across all the income groups, the loan amount is higher for people who defaulted.

▶ Employees with longer working history got the loan approved for a higher amount also they defaulted the most.

▶ Looking at the verification status data, verified loan applications tend to have higher loan amount.

Conclusion

There is a more probability of defaulting when :

- Applicants taking loan for 'Debt Consolidation'.

- Applicants whose home ownership is 'MORTGAGE'.

- Applicants who receive interest at the rate of 21-24%

- Applicants who have taken a loan in the range 30k - 35k and are charged interest rate of 15-17.5 %.

- Applicants who have taken a loan for small business.

- When grade is F and loan amount is between 15k-20k.

- When employment length is 10yrs or more.

- For grade G and interest rate above 20%.

THANK YOU