

# Lending Club Case Study

By- Ashu Tosh Bharadwaj

Kushagra Sengar

# Index

- Problem Statement
- Goal
- Data Description
- Steps Involved
- Conclusion on various variables
- Driving Variables

# Problem Statement

Being the loan lending company, If one is able to identify risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

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

In this case study, you will use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default.

# Goal

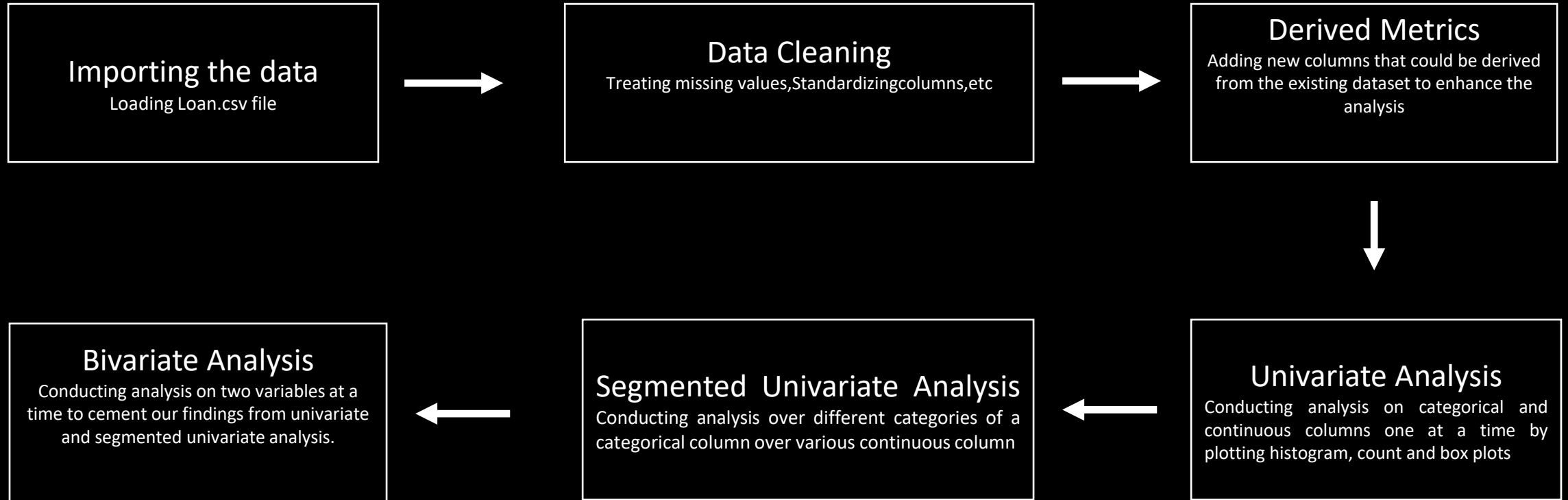
The aim of this case study is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

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. The company can utilise this knowledge for its portfolio and risk assessment

# Data Description

- Dataset used in the case study is Loan.csv which has information about past loan applicants and whether they 'defaulted' or not.
- The dataset has 111 distinct columns and 39717 rows with mixed of continuous and categorical values.

# Steps Involved

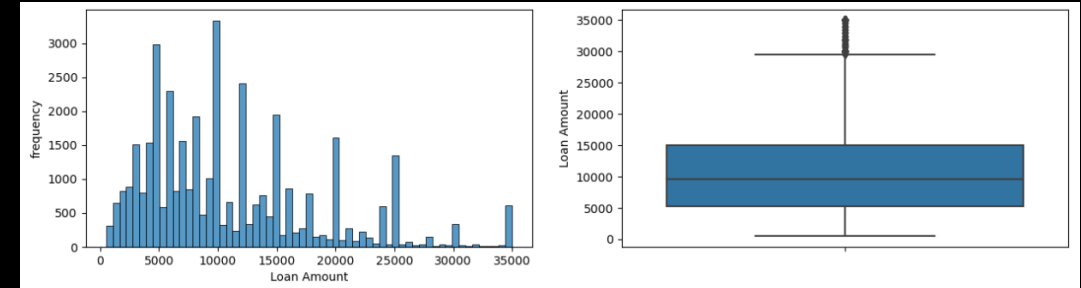


Conclusion on various variables

# Loan Amount

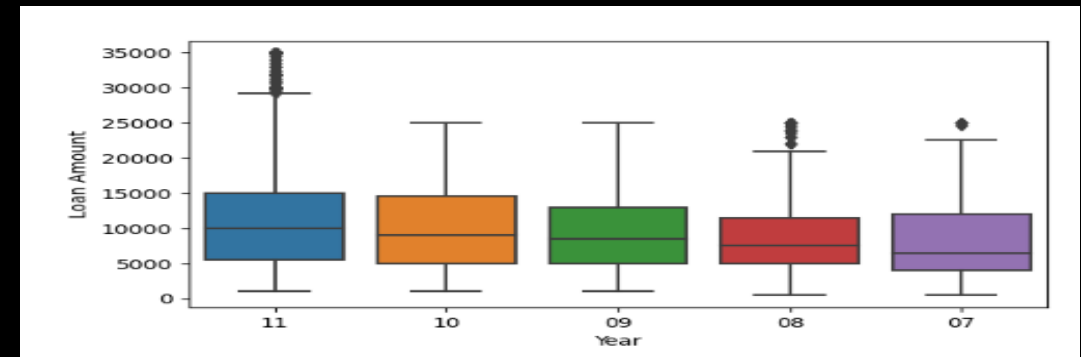
## Univariate Analysis :

The loan amount between 5000 - 10000 is requested the most by the loan applicants



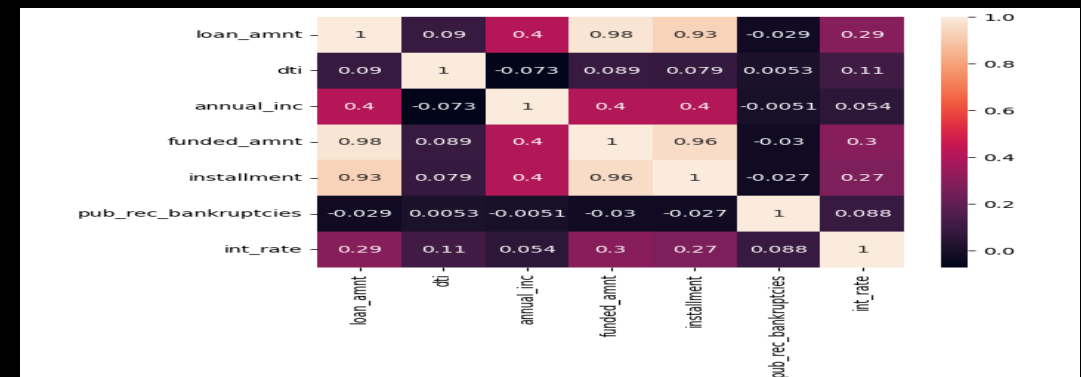
## Segmented Univariate Analysis :

The median loan amount has significantly increased over the years with 2011 leading the race



## Bivariate Analysis :

Loan amount is strongly correlated with Funded amount and installments.

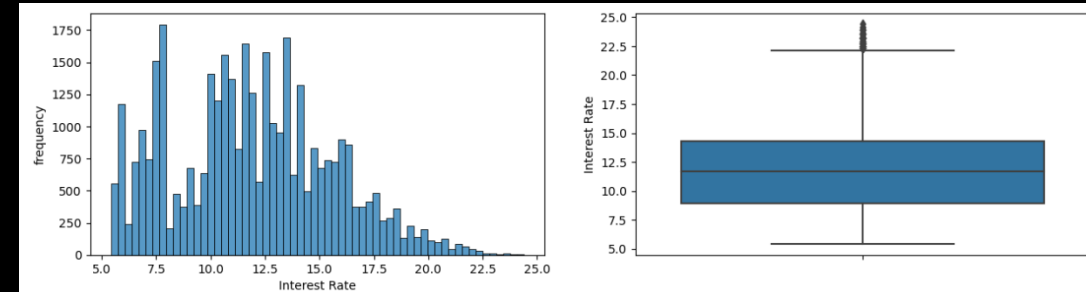




# Rate of Interest

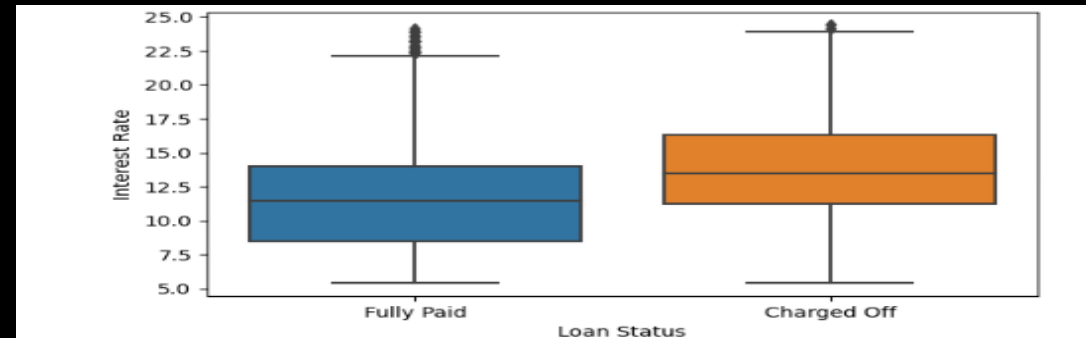
## Univariate Analysis :

Most of the applicants interest rate lies between 10 - 14 %



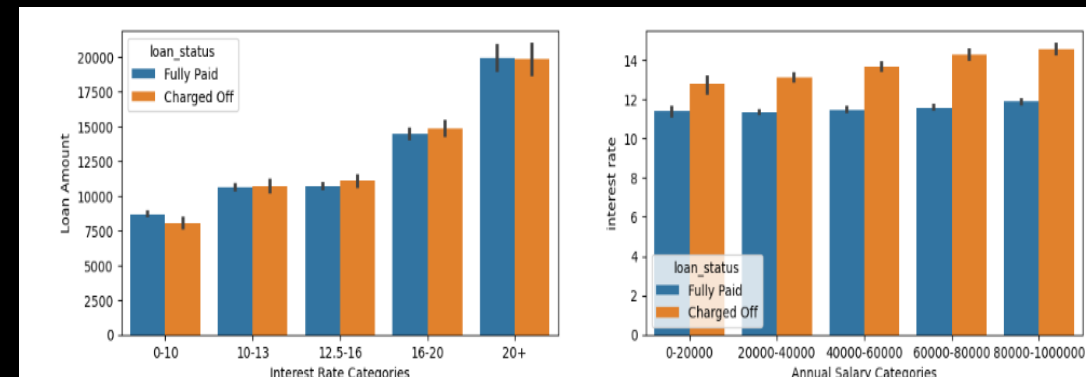
## Segmented Univariate Analysis :

People with high rate of interest are likely to default

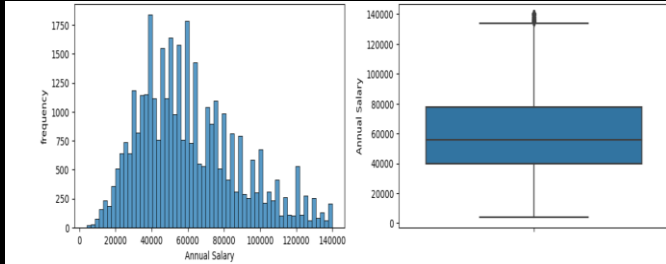


## Bivariate Analysis :

As the loan amount increase, the interest rate increases too people are more likely to be at default.

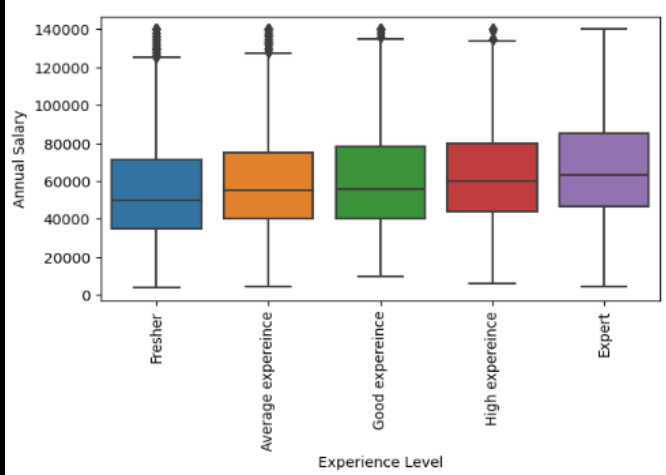
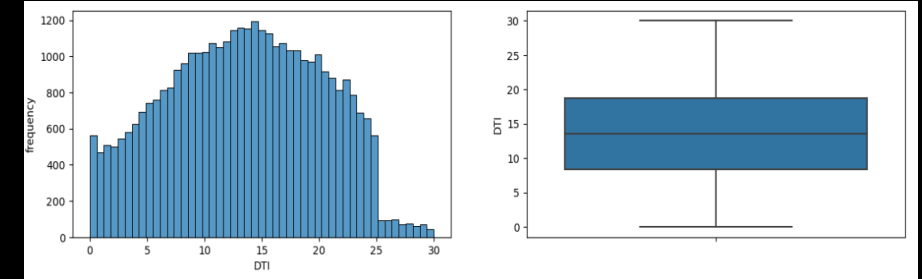


# Annual Salary and DTI



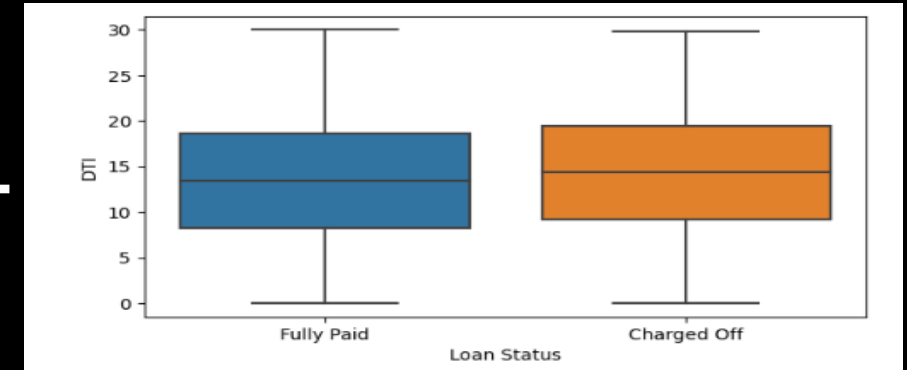
## Univariate Analysis :

Most of the applicants interest rate lies between 10 - 14 %  
People with the annual salary between 40000-60000 have applied for the loans the most than the people falling in other salary brackets



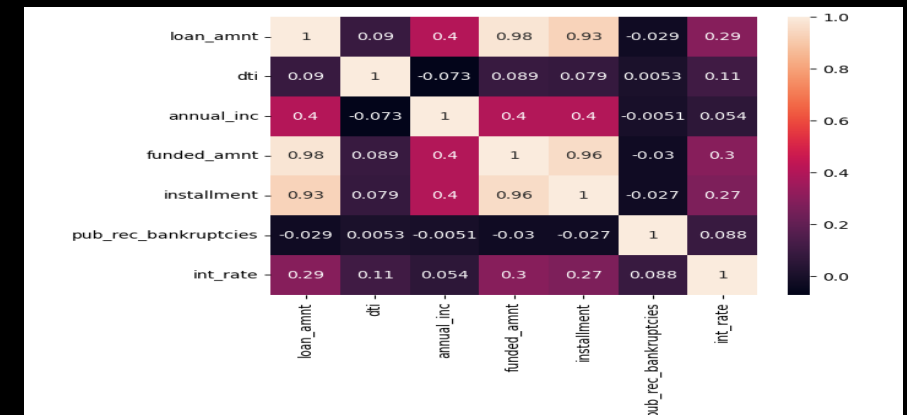
## Segmented Univariate Analysis :

People with high rate of interest are likely to default.  
The experience level plot shows that the average annual salary increase with the year of experience



## Bivariate Analysis :

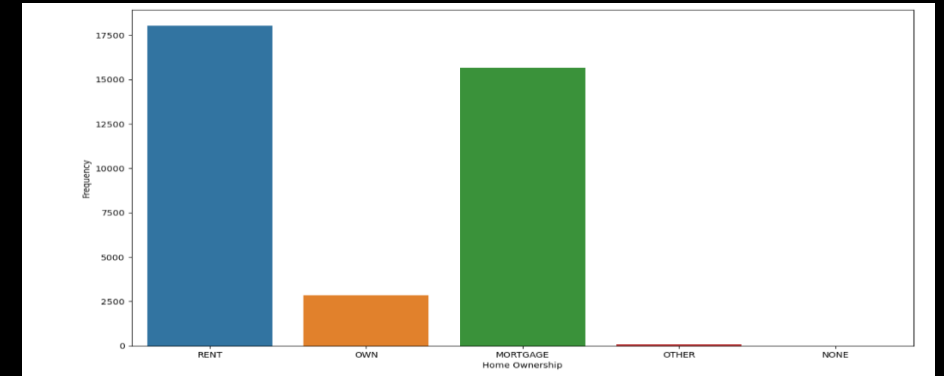
DTI is negatively correlated to the annual salary i.e if the annual salary is less then the DTI is more (monthly debts are more)



# Conclusion - Ownership

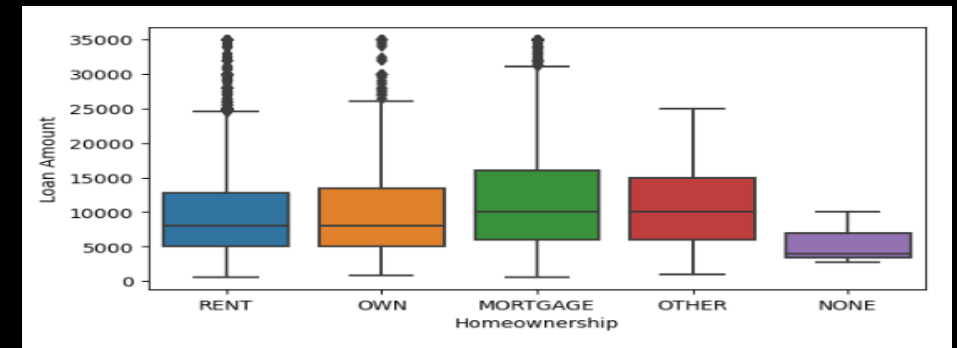
## Univariate Analysis :

Most of the loan applicants are living on a rented property followed by mortgage.



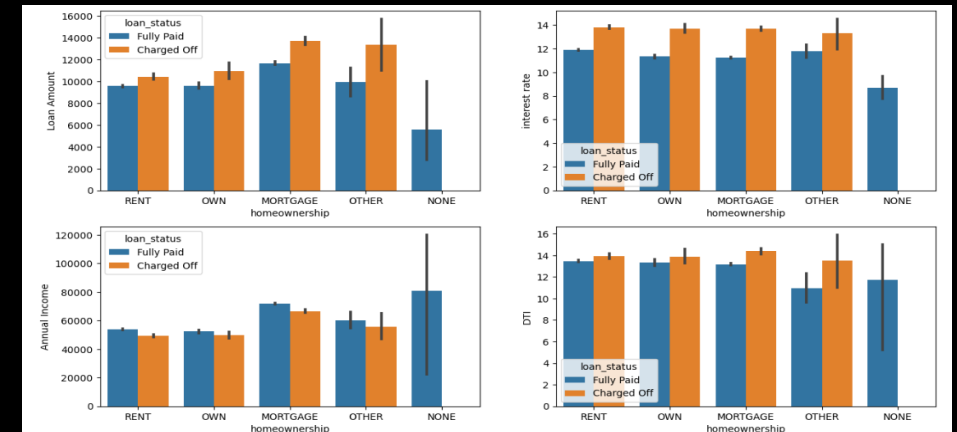
## Segmented Univariate Analysis :

The median loan amount for people who have homeownership of mortgage is the most



## Bivariate Analysis :

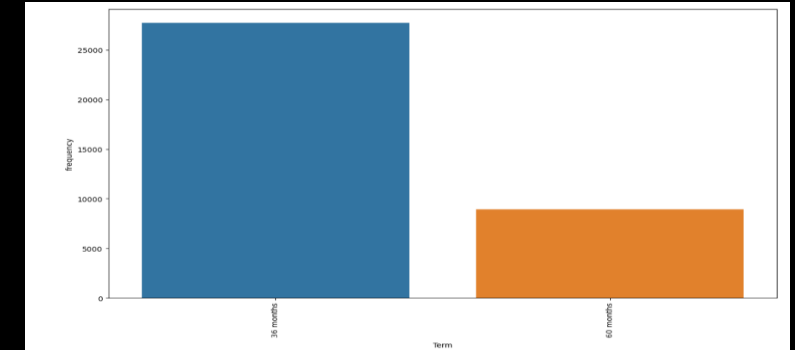
People who have ownership as 'Mortgage' are likely to be at default for higher loan amount, higher rate of interest and higher DTI



# Term

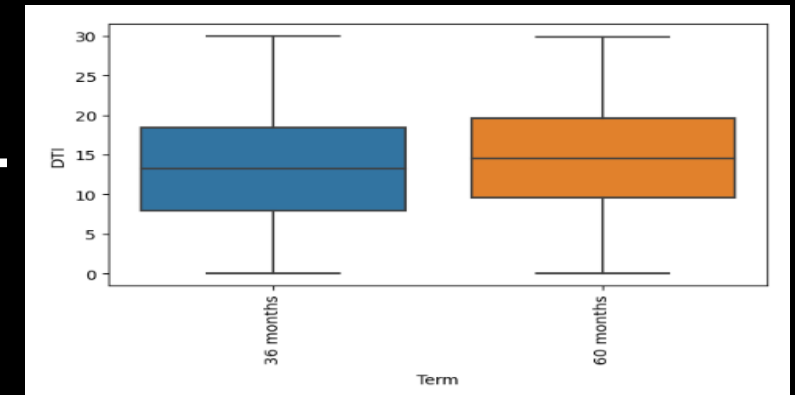
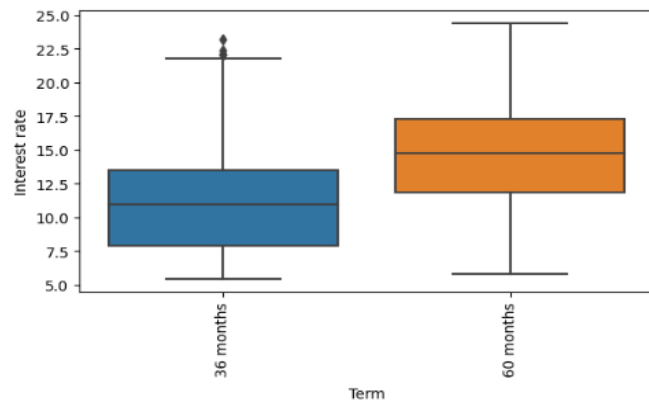
## Univariate Analysis :

36 months EMI is common amongst the loan applicants



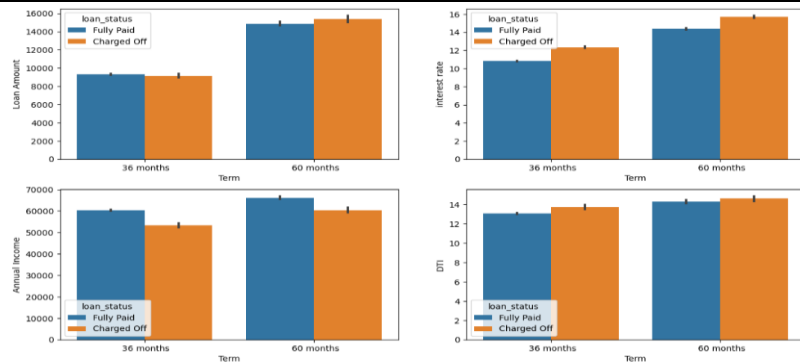
## Segmented Univariate Analysis :

The rate of interest on 60 months term is more than that of 36 months term. There is a slight difference in the DTI of a 36 months tenure applicants as compared to 60 month term.



## Bivariate Analysis :

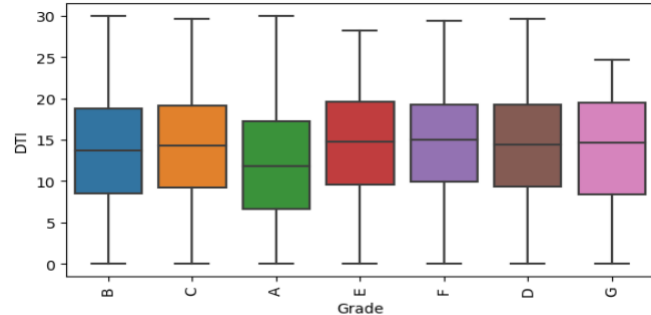
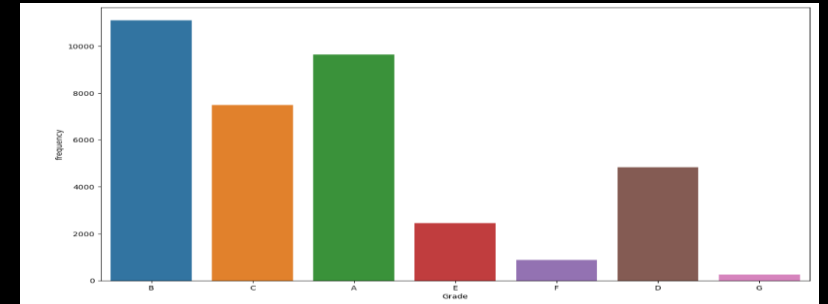
People who took loan for tenure of 60 months are likely to be default for higher amount, rate of interest and DTI.



# Grade

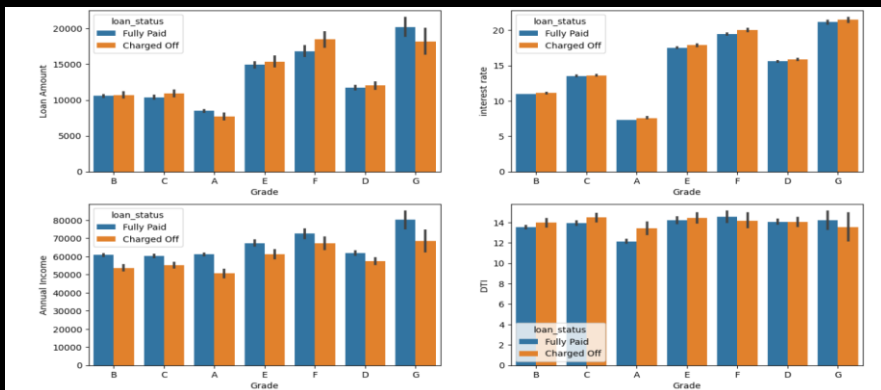
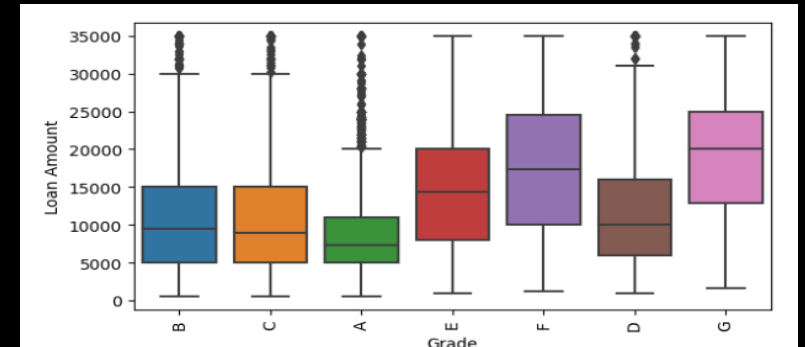
## Univariate Analysis :

Lending Club has classified most of its applicants as Grade B and Grade A



## Segmented Univariate Analysis :

DTI of grade A customer is the lowest. Grade 'G' people are having the average loan amount to be greater than the other.



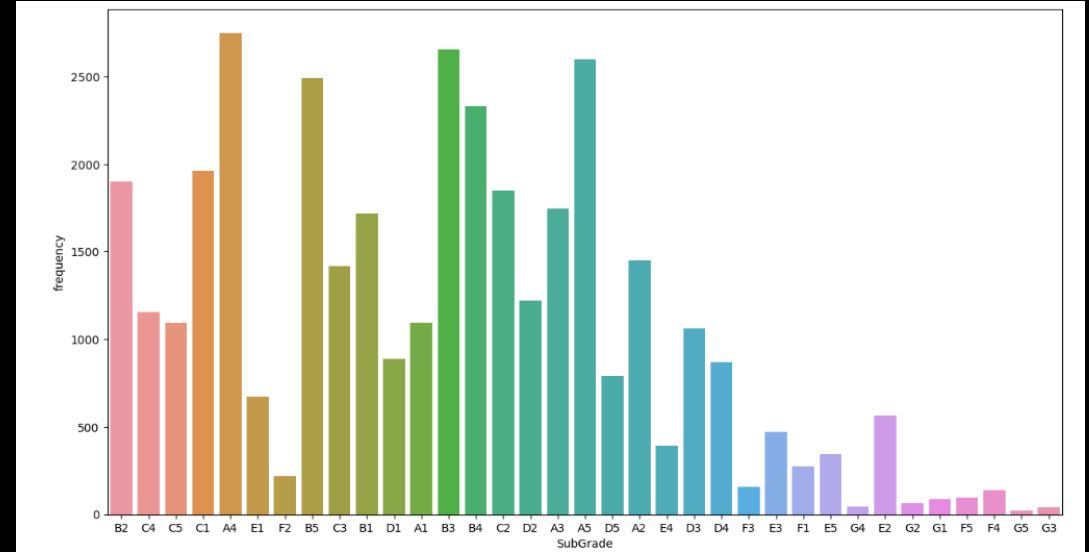
## Bivariate Analysis :

People belonging to Grade 'G' and 'F' tend to pay high rate of interest and hence are likely to be at default. Maximum people who are at default for higher loan amount belong to grade 'F' and 'G'.

# Subgrade

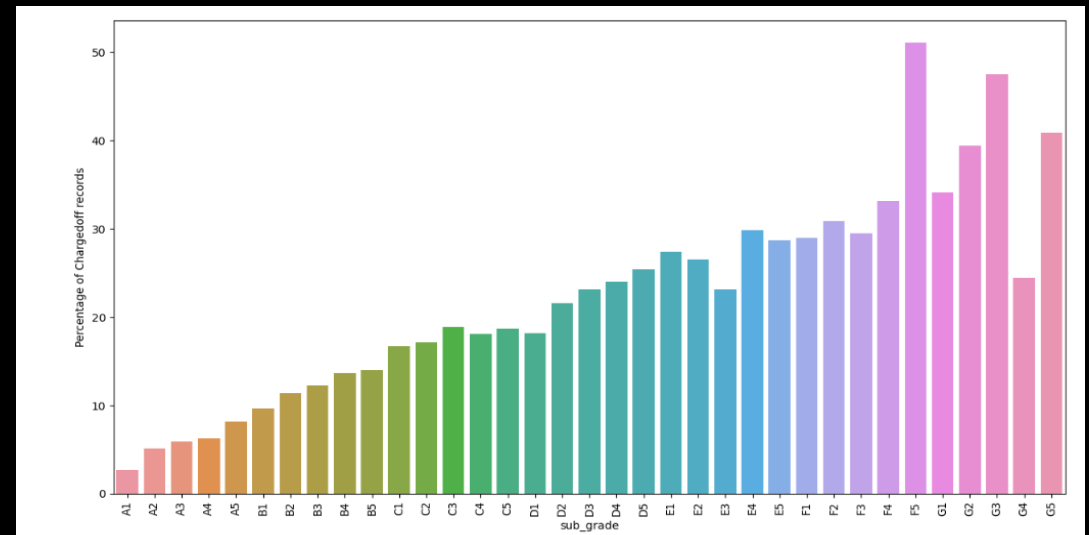
## Univariate Analysis :

There is no such pattern for sub grade that we can assume based on the plot above



## Bivariate Analysis :

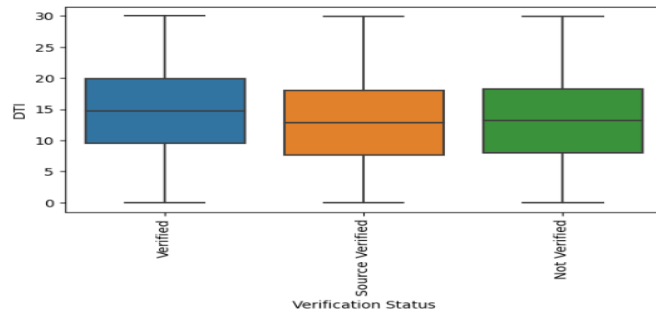
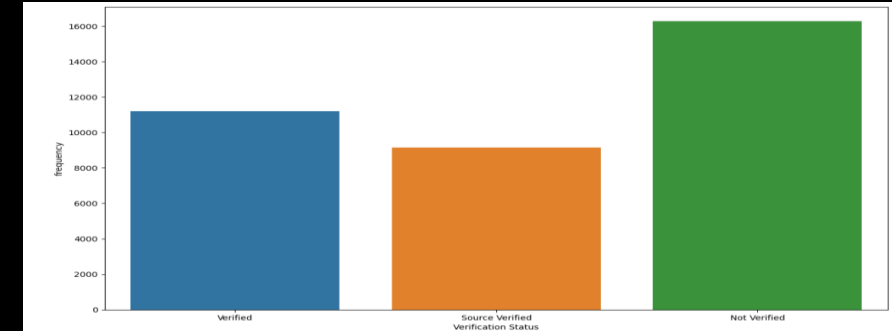
People belonging to subgrade 'F5' have highest charged off percentage.



# Verification Status

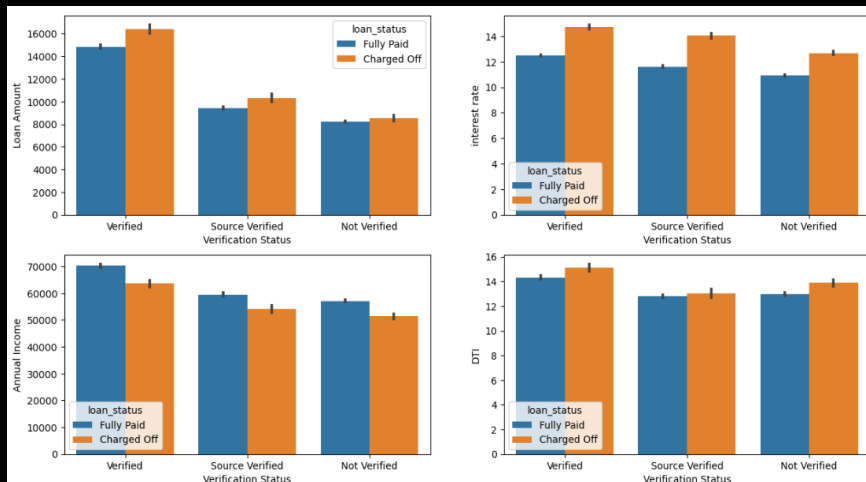
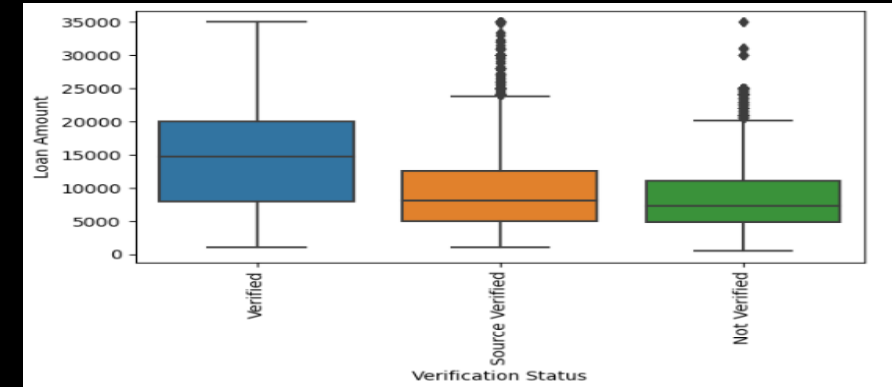
## Univariate Analysis :

Most of the applicants are not verified



## Segmented Univariate Analysis :

DTI of verified people is slightly more than the source verified and not verified. The average loan amount for the people who are verified is more than the people who are either source verified or not verified.



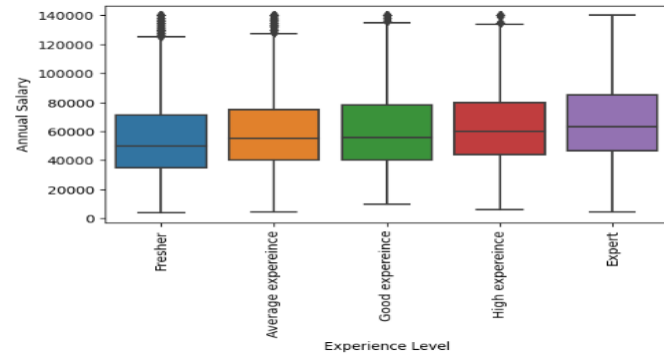
## Bivariate Analysis :

People who are verified usually tends to be charged off when the loan amount is higher and same happens when the DTI is higher too.

# Employee Length

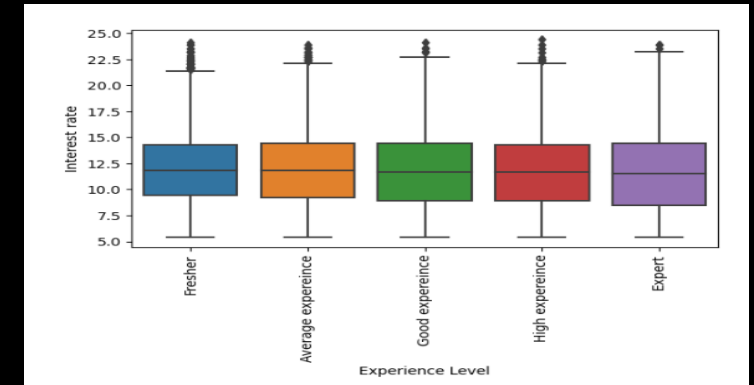
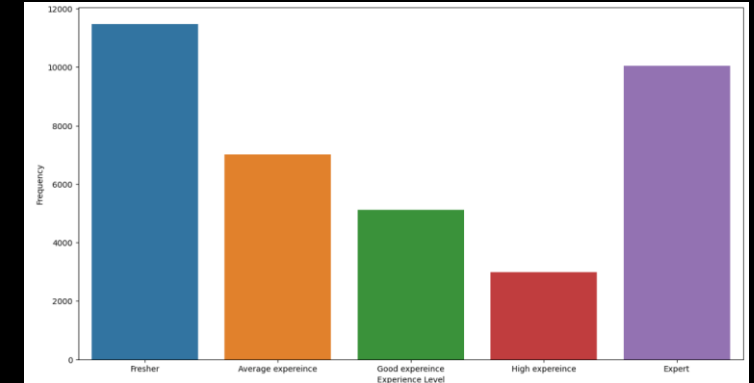
## Univariate Analysis :

Most of the applicants are freshers i.e having experience of less than or equal to 2 years followed by experts having more than 10 years of experience.



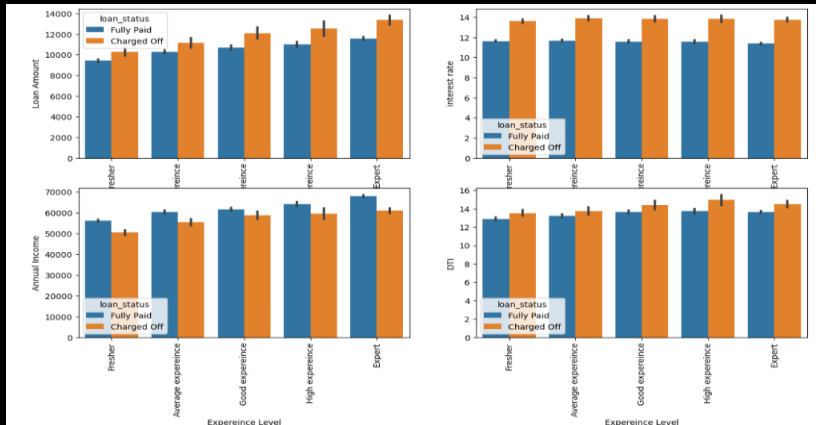
## Segmented Univariate Analysis :

The experience level plot shows that the average annual salary increase with the year of experience. There is no such difference in ROI based on the level of experience.



## Bivariate Analysis :

People who have experience of more than 10 years are likely to be at default for higher amount and rate of interest.

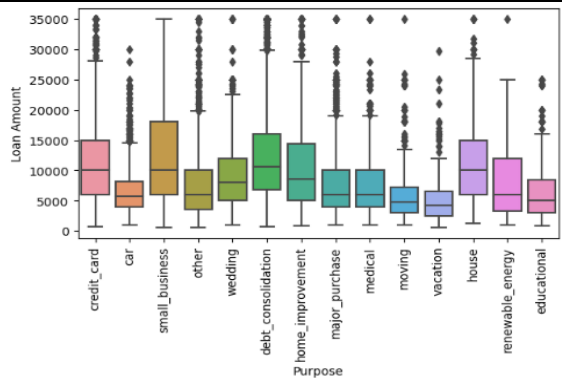
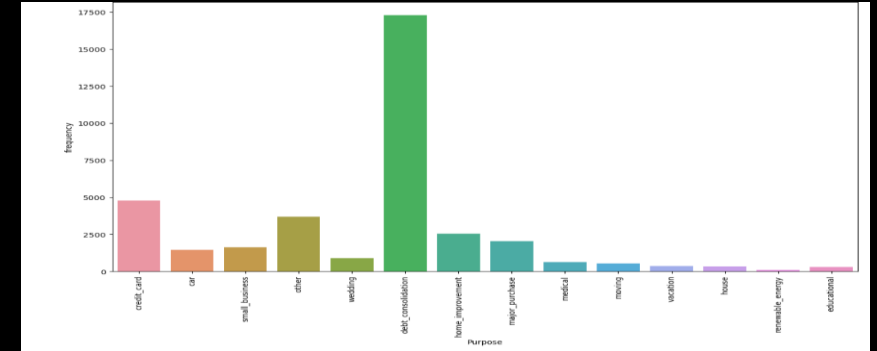




# Purpose

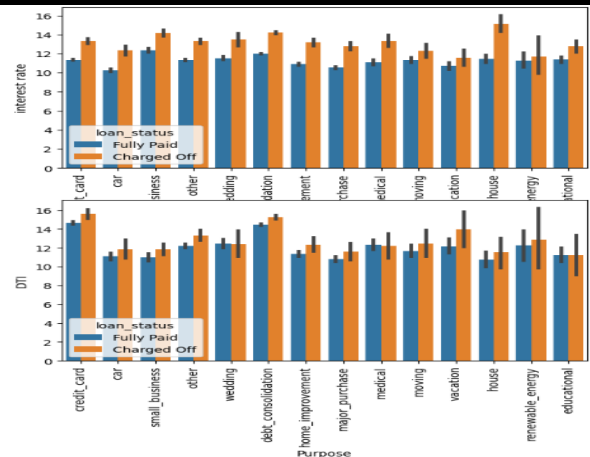
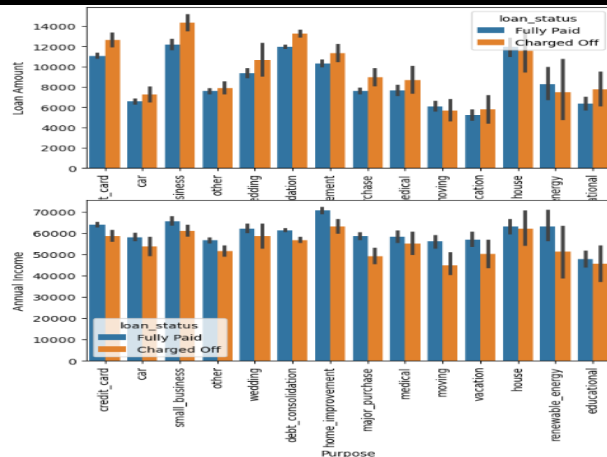
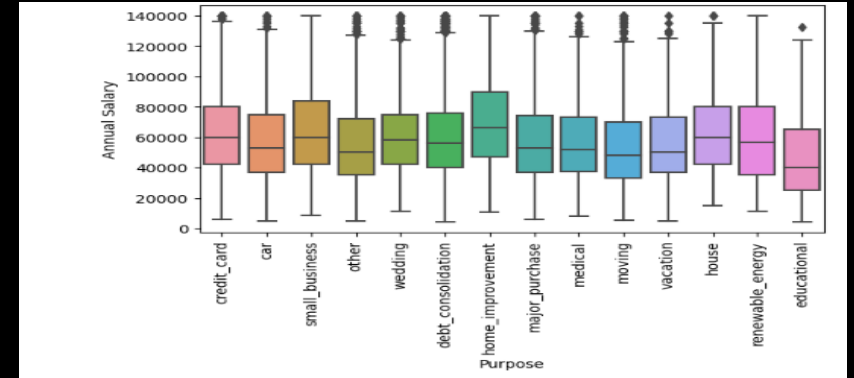
## Univariate Analysis :

Most of the applicants interest rate lies between 10 - 14 %



## Segmented Univariate Analysis :

People who have higher annual salary tend to take loan for home improvement the most and have mortgage as homeownership. The median loan amount for small business is the most, then the second is debt consolidation.



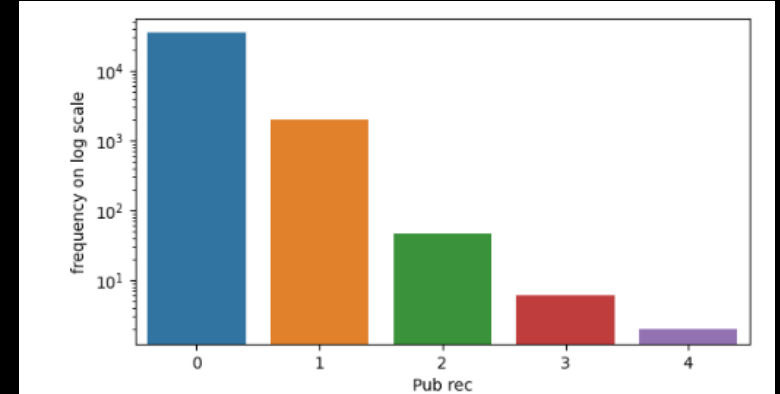
## Bivariate Analysis :

Most of the people who are at default took loan for the purpose of small business when the loan amount is higher while when the rate of interest is high then people who took loan for house and home\_improvement are likely to be at default. Similarly in case of high DTI ,credit card is purpose behind people being at default.

# Derogatory Public Record

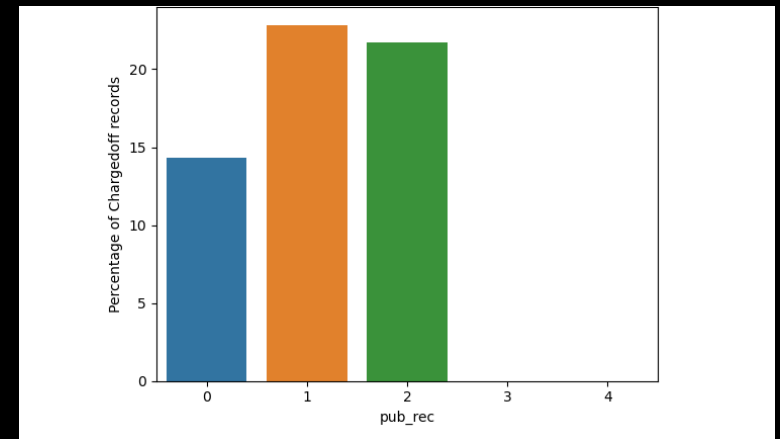
## Univariate Analysis :

Almost 85 % have no bad public record



## Bivariate Analysis :

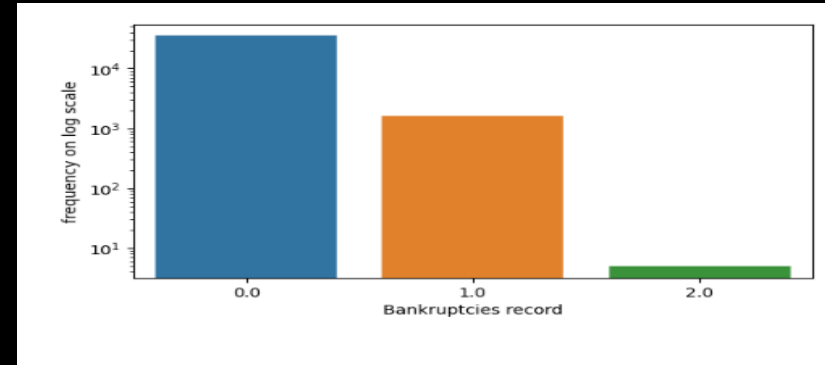
Even though the number of records for derogatory records are low , we cannot ignore that the person is likely to be at default if he/she has bad public record in the past.



# Public Record for bankruptcies

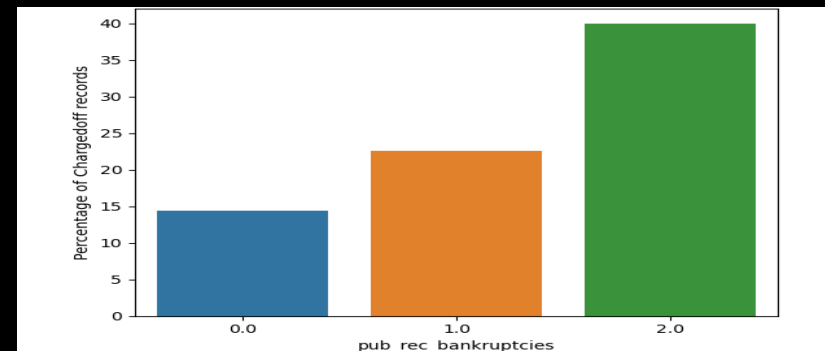
## Univariate Analysis :

Almost 85 % have case of bankruptcies



## Bivariate Analysis :

Even though the number of records for bankruptcies are low , we cannot ignore that the fact that person is likely to be at default if he/she has been involved in bankruptcies in the past.



# Driving Variables

- Interest rate - When the interest rate is high , a person is highly likely to be at default.
- Term - People opting for 60 months tenure are highly likely to be at default.
- Grade - People belonging to grade 'G' and 'F' are likely to be at default.
- Subgrade - People belonging to Subgrade 'F5' are likely to be at default.
- Purpose - People who took loan for credit\_card, House and small business are likely to be at default at higher loan amount.
- Ownership - People have ownership of 'Mortgage' are likely to be at default.
- Employee Length - People having experience of more than 10 years are likely to be at default as they seek higher loan amount which has higher ROI.
- Verification Status - People having verification status as Verified are likely to be at default at higher loan amount.
- Pub record and bankruptcies - People with Derogatory Public Record and any record for bankruptcies are likely to be at default.