

# Lending Club

Misbah Mohammed  
Sandeep Bhatia



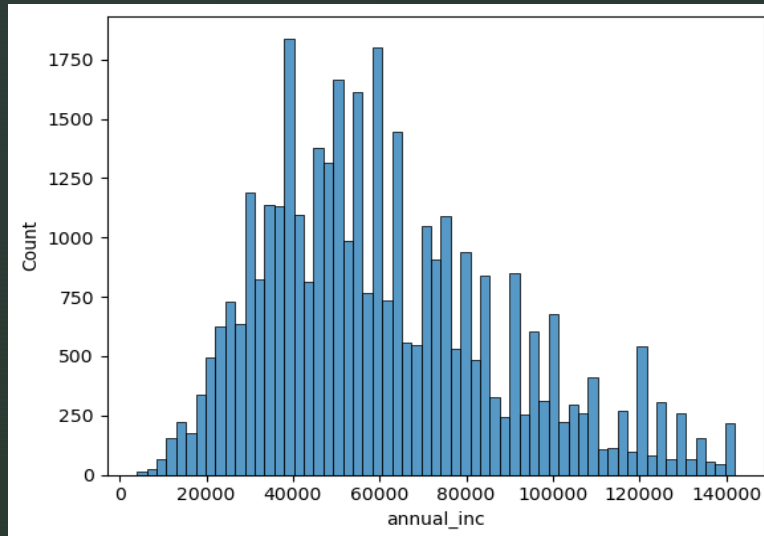
# Problem Statement & Business Objective

- ✓ When the company receives a loan application, the company needs to decide for loan approval based on the applicant's profile.
- ✓ Using Exploratory Data Analysis on given historical loan data, identify how consumer attributes and Loan Attributes influence the tendency of default.
- ✓ identify the key driving factors (or driving variables) behind loan default
- ✓ The company can utilize this knowledge for its portfolio and risk assessment.

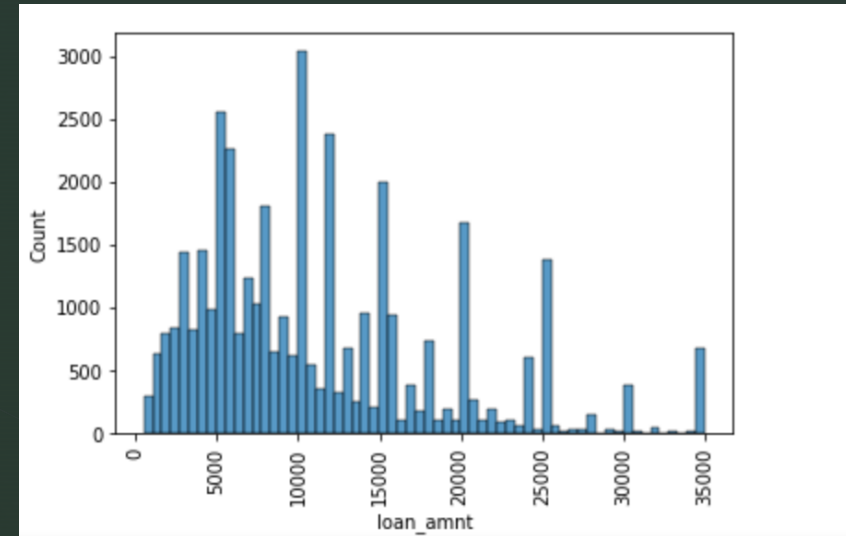
# Data Understanding & Cleansing Approach

- Load Data provided for all the loans issued through the time period 2007 to 2011
- Number of Records – 39717
- Number of fields - 111
- Drop the columns which have all Null values. e.g. - annual\_inc\_joint , dti\_joint, etc. 54 such columns removed which have all Null values.
- Check the unique records in the data set. Identified that there are few columns which only have 1 unique record. These columns will not be of any significance, hence removing these columns. – e.g. - policy\_code, application\_type, etc. Removed 8 such columns.
- Check the unique records in the data set. Identified that there are few columns which have 39717 unique records. e.g. id, member\_id, url. Removed 3 such columns
- Drop columns which don't have significance in analysis .e.g. zip\_code, emp\_title, desc, title. Removed 4 such columns.
- Check the missing values in the columns and removing columns which have more than 50% data missing. e.g - mths\_since\_last\_delinq , mths\_since\_last\_record, next\_pymnt\_d
- Checking the missing values in the columns and deleting the row if those are very low in number as it will not have significant impact on the analysis.
- Categorize the column between Numerical and Categorical

# Univariate Data Analysis – 1/4

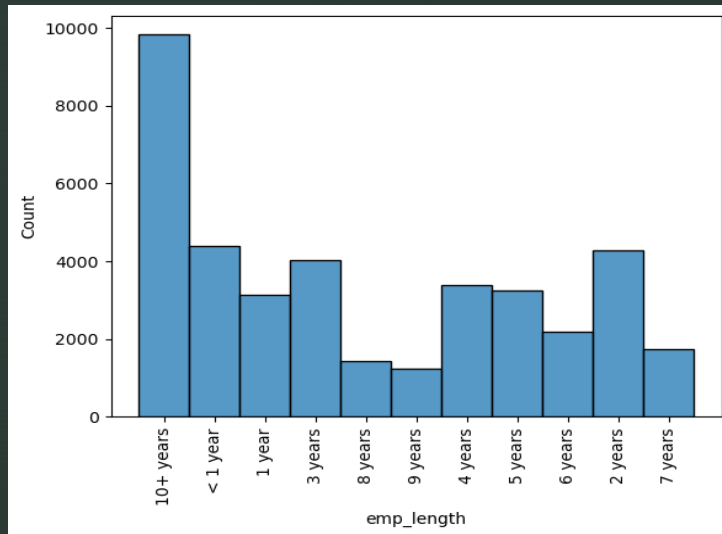


- The annual Income is one of the important factor lesser the income more chances of the person defaults

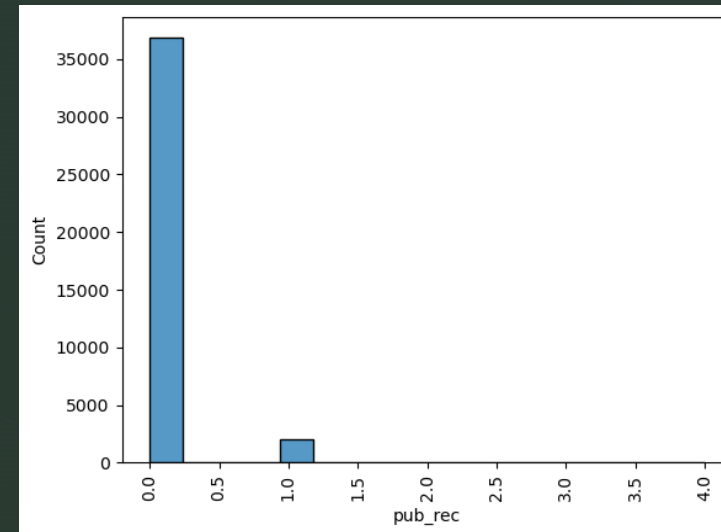


- The higher the loan amount more chances of person defaults.

## Univariate Data Analysis – 2/4

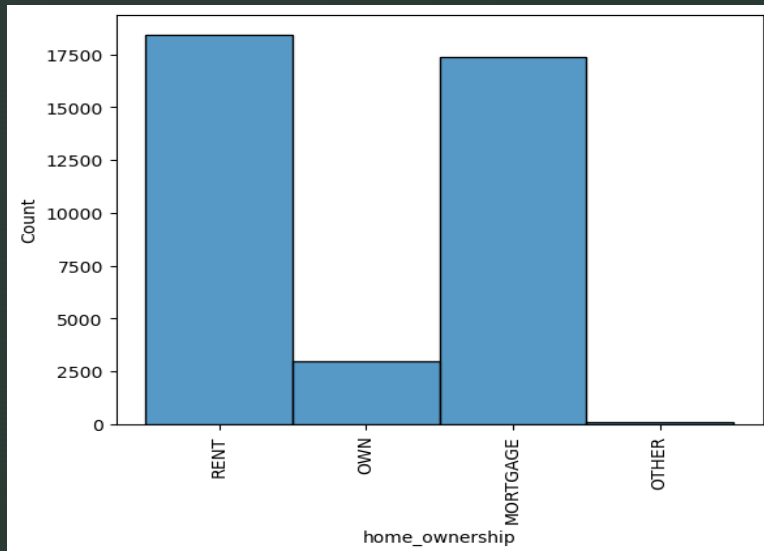


- The more the employment Length , less are the chances to default.

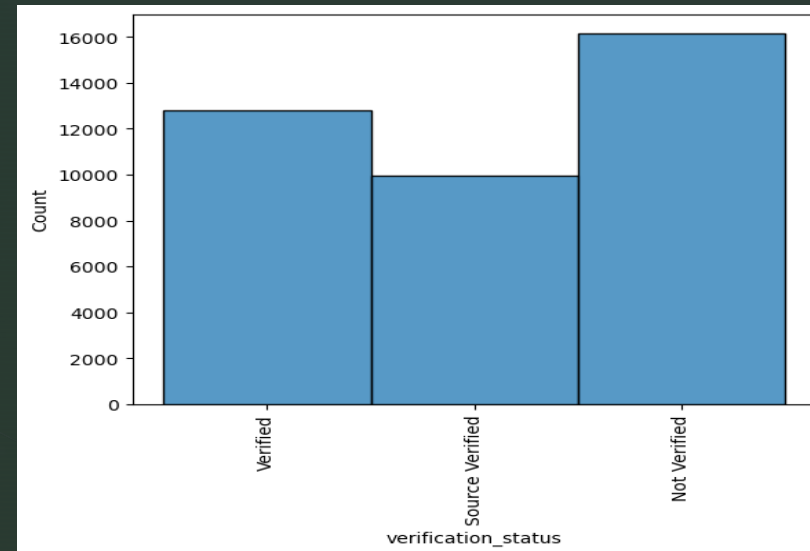


- The people with zero derogatory public records have less chances to default.

## Univariate Data Analysis – 3/4



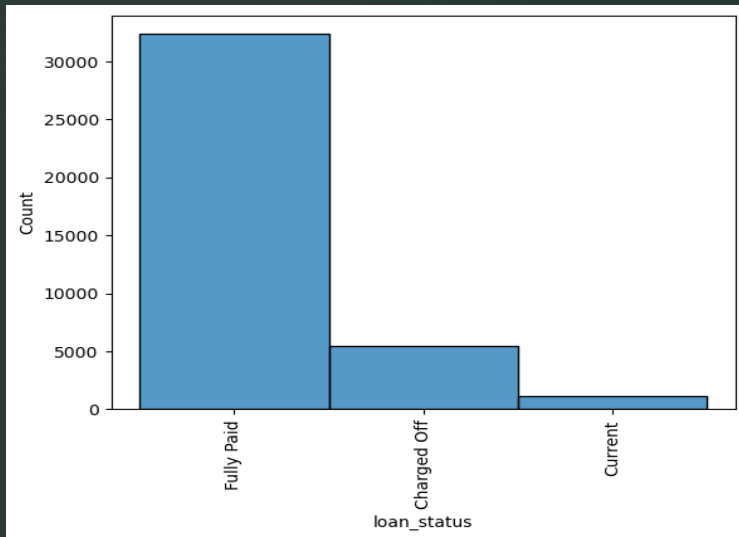
- The people own home have less chances to default



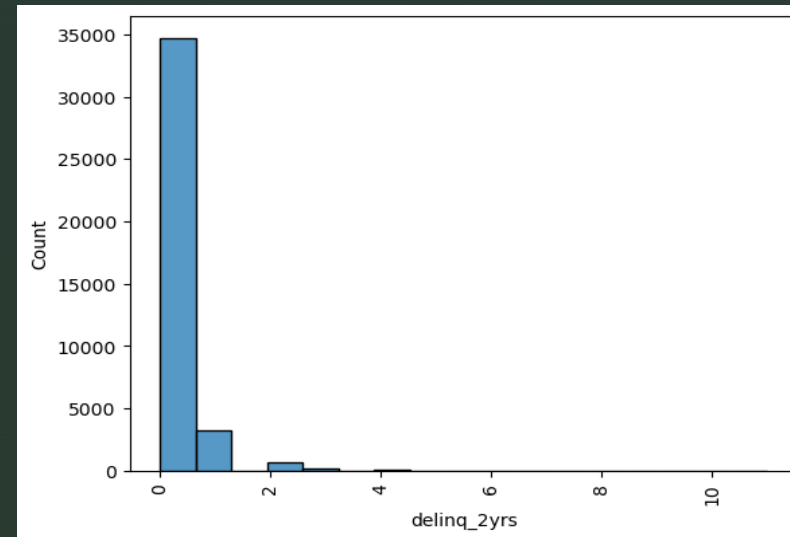
- The income verified by LC less prone to default



## Univariate Data Analysis – 4/4

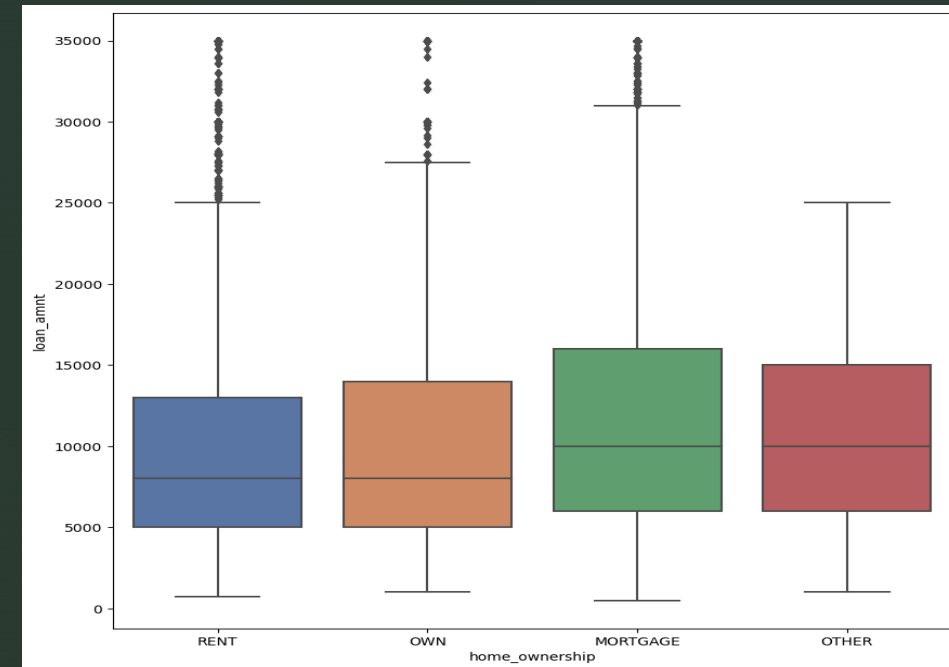
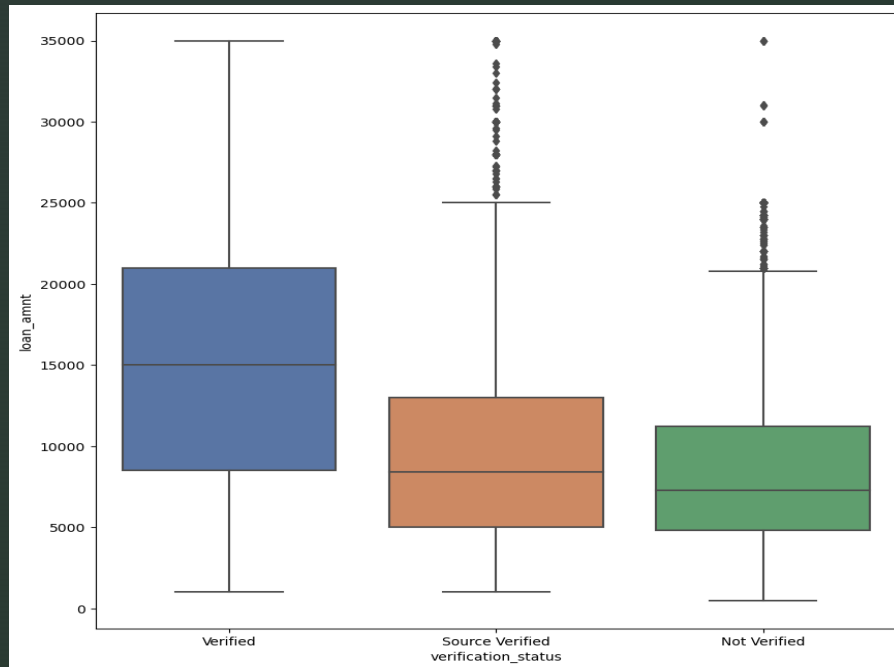


- The people who have fully paid loan have less chances to default.



- People with zero delinquency in last 2 year have less chances to default

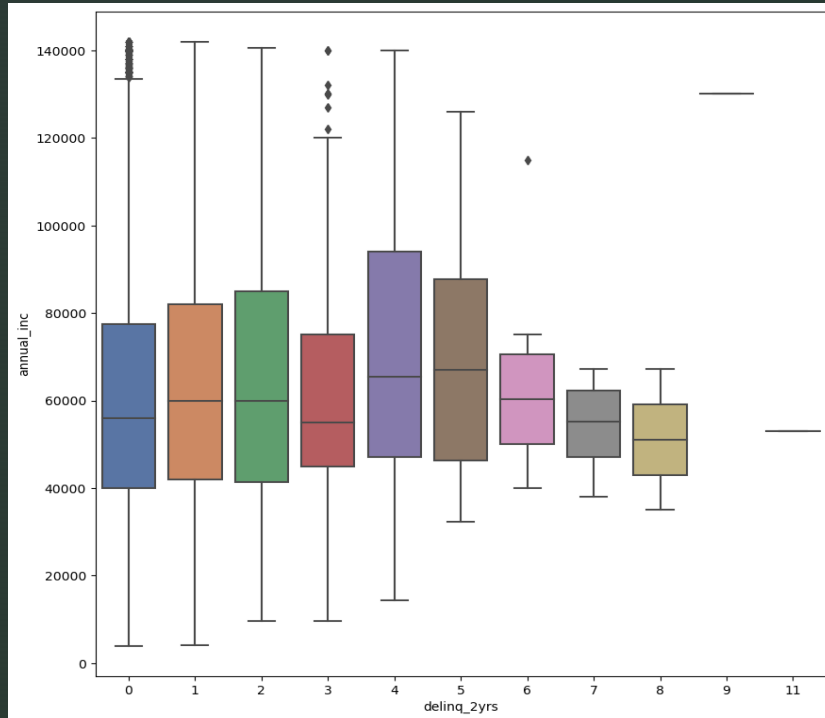
## Bivariate Data Analysis – 1/4



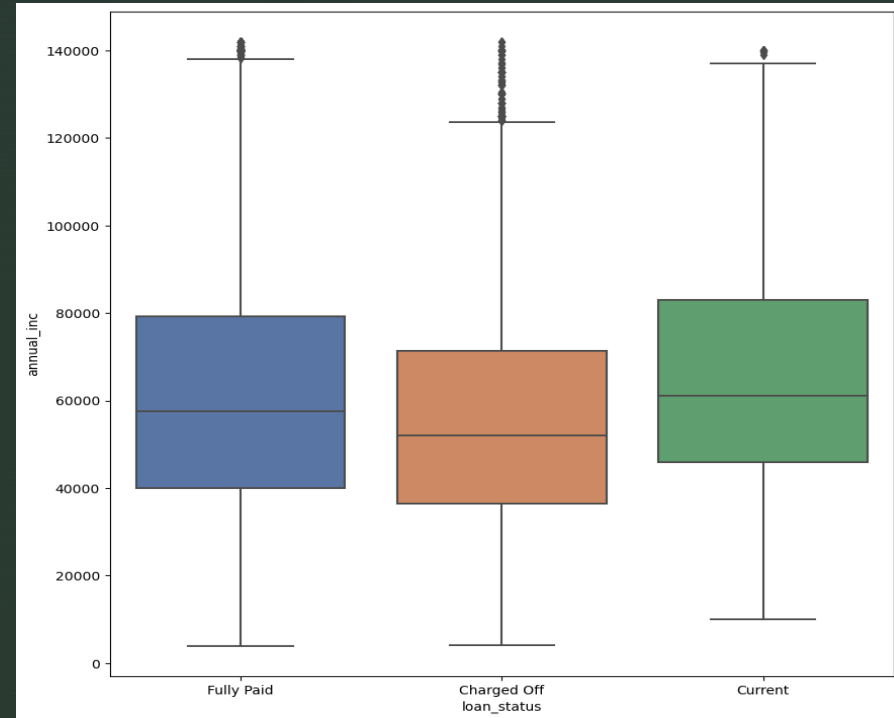
- The people with LC verified can be given higher loan amount.
- The people with own house have less chances to default.



## Bivariate Data Analysis – 2/4

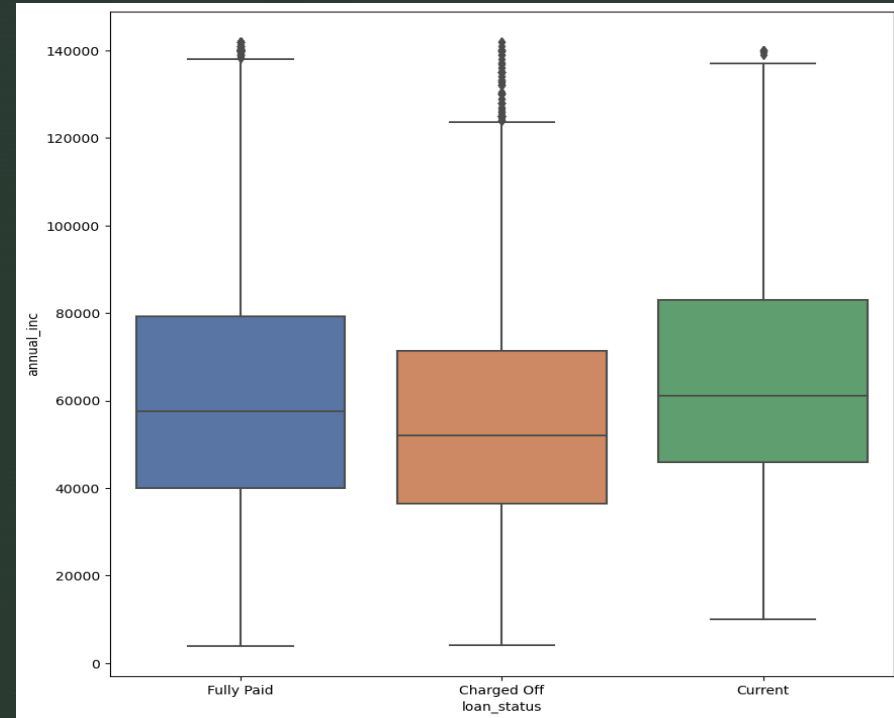
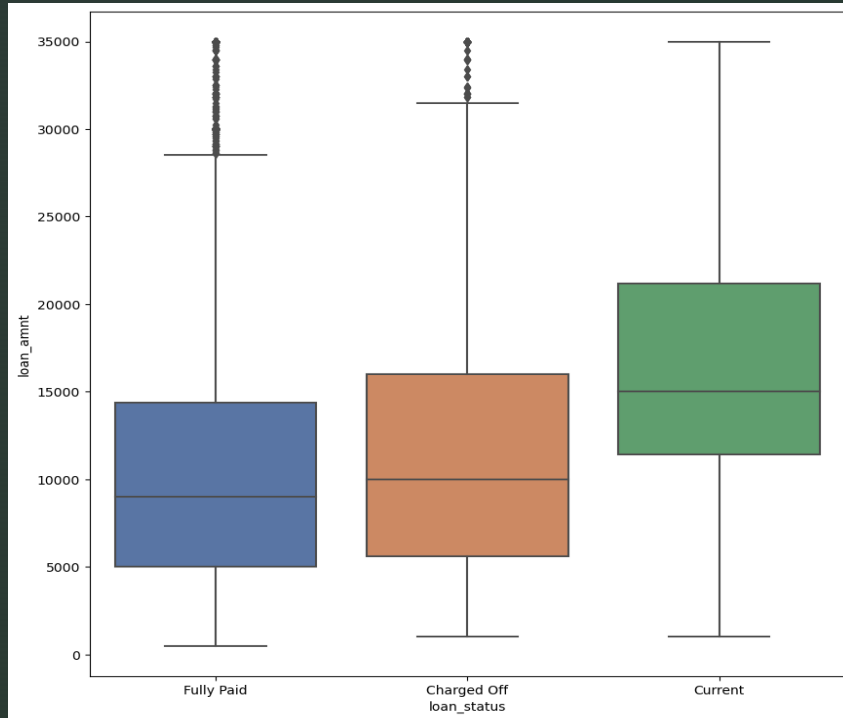


- The people with higher income have less chances of delinquency .



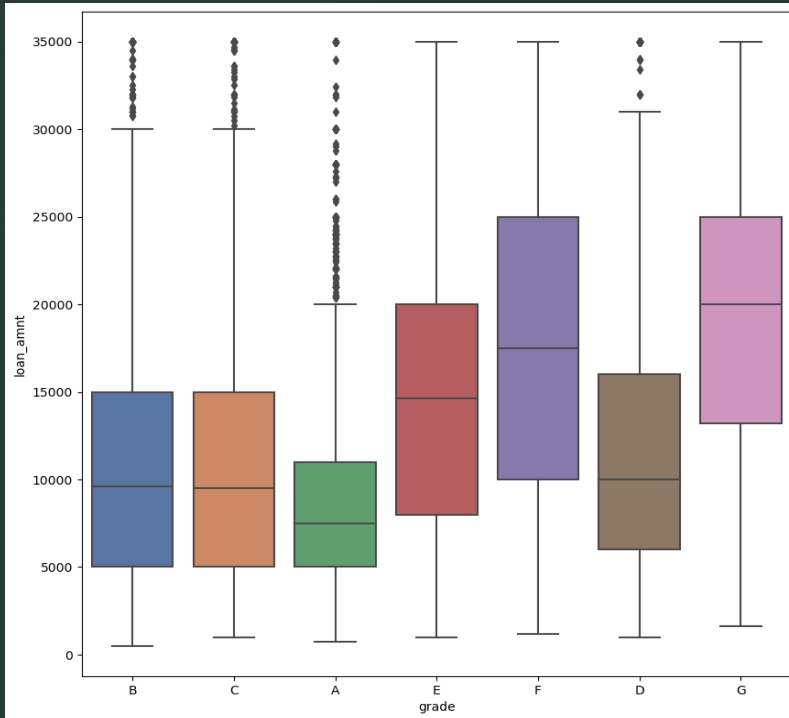
- The people with higher income have fully paid the loan in the past.

## Bivariate Data Analysis – 3/4

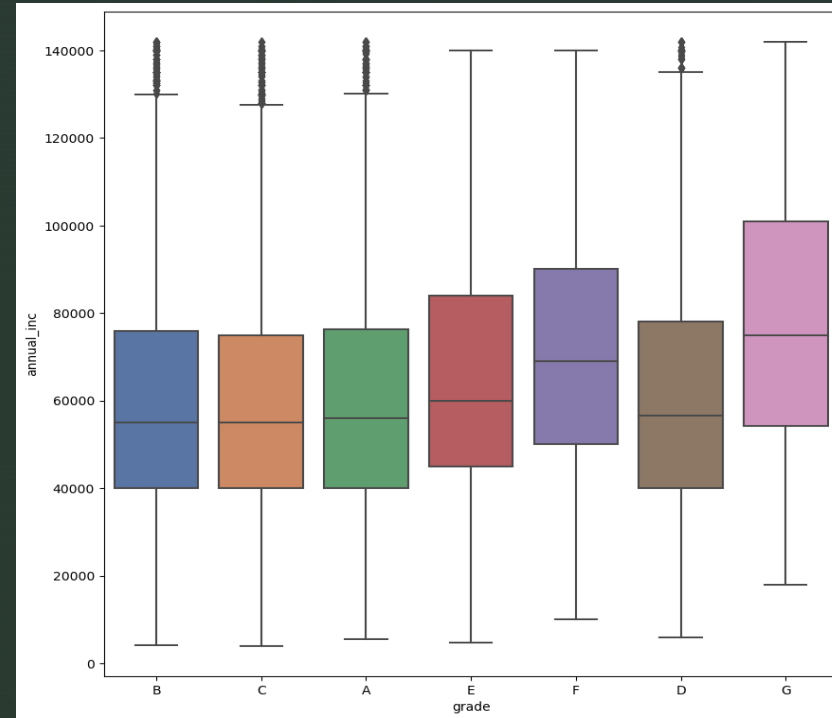


- The lesser loan amount were full paid , Higher the loan amount there are more chances of default.
- The people with higher income have fully paid the loan in the past.

## Bivariate Data Analysis – 4/4



- The higher loan amount were given the higher-grade G & F



- The people with higher income have higher LC assigned loan grades

# Recommendations

- ❑ As per the data analysis, it is clear that customers who have a higher annual income are less prone of being default •
- ❑ Shorter loan term leads to higher chances of being fully paid .
- ❑ The people with LC verified can be given higher loan amount.
- ❑ The people with own house have less chances to default.