

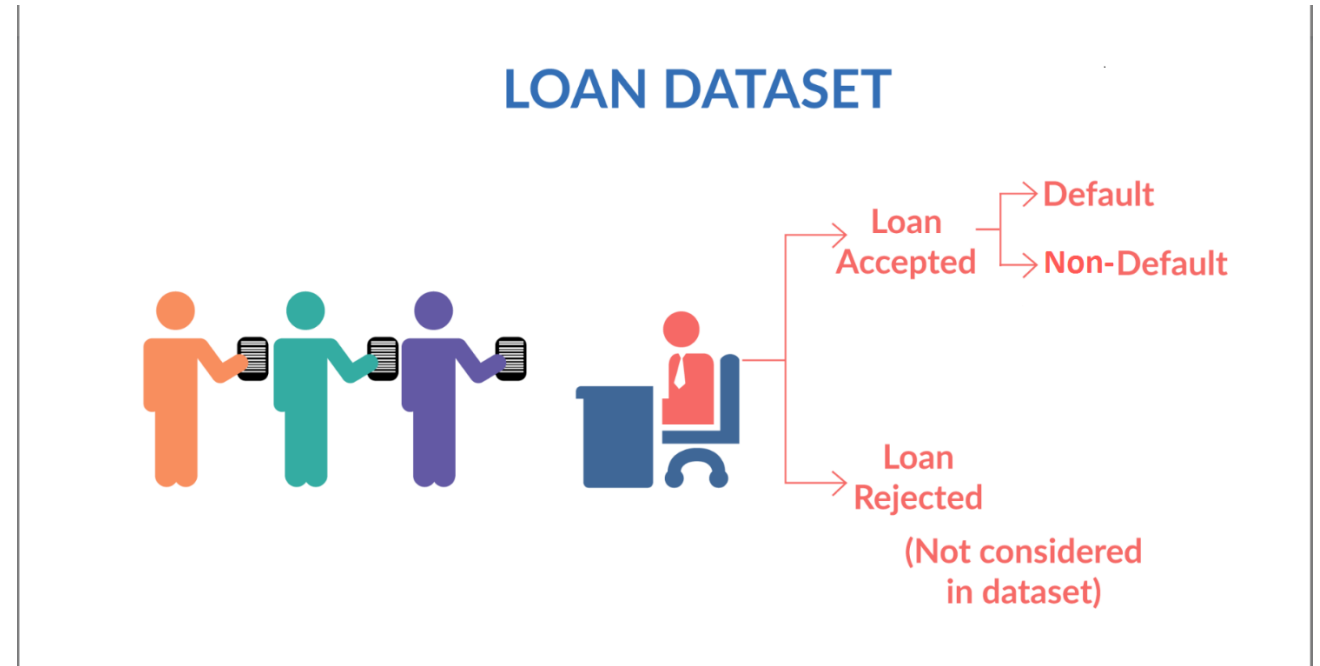
# Lending Club Case Study



- Swarna Polamarasetti

## BUSINESS PURPOSE

By understanding the driving factors (or driver variables) behind loan default, the company can utilize this knowledge for its portfolio and risk assessment.



## PROBLEM STATEMENT

Complete EDA on how consumer attributes and loan attributes influence the tendency of default with the given data set that contains the complete loan data for all loans



# Steps involved in EDA

## Data Analysis

Understanding the columns of dataset  
Figuring out the driving variables.

## Data Cleaning

Removing columns with null and partial null values  
Removing duplicate rows

## Handling Data

Creating derived columns wherever needed.  
Convert columns to desired data type.  
Remove columns with similar values

## Univariate Analysis

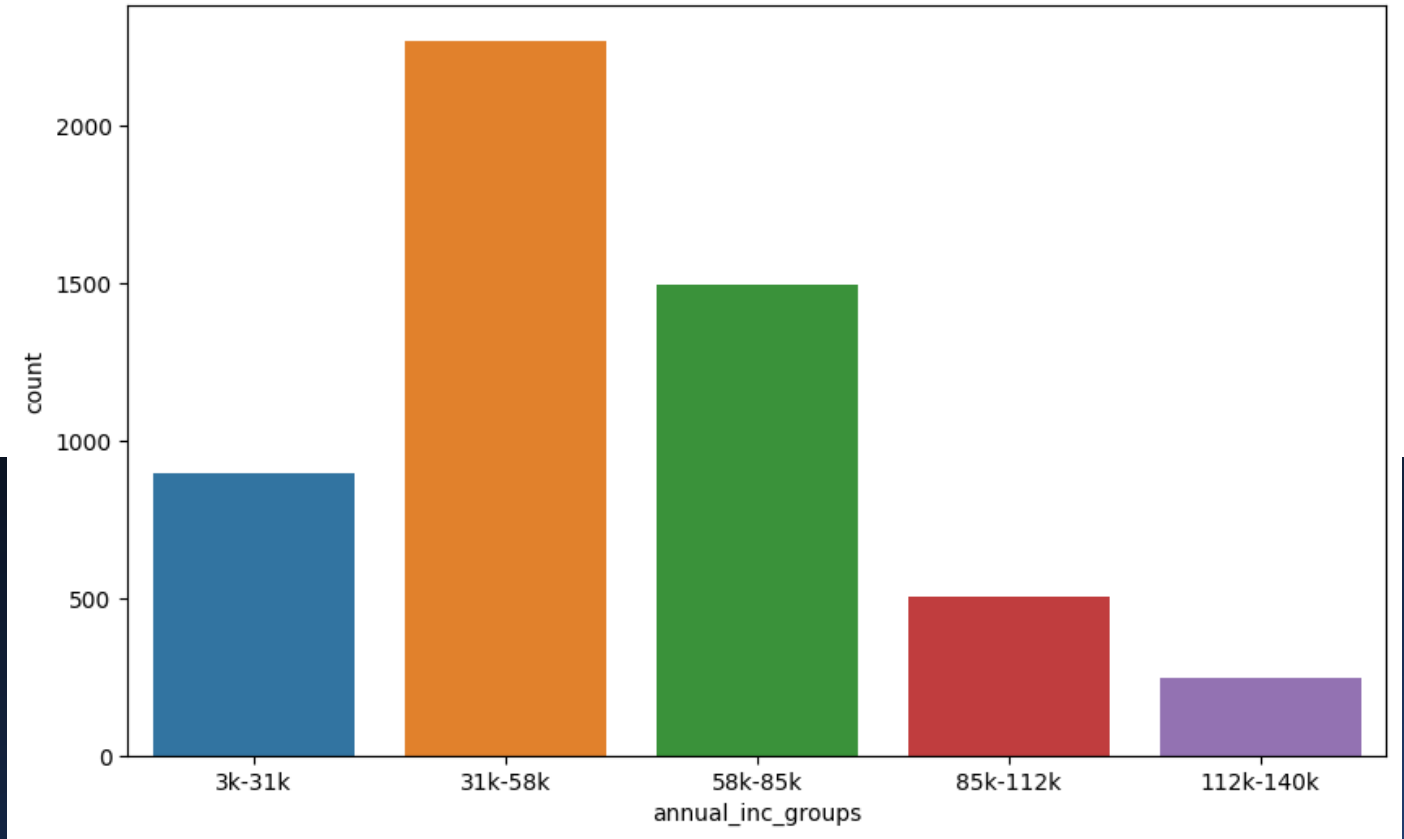
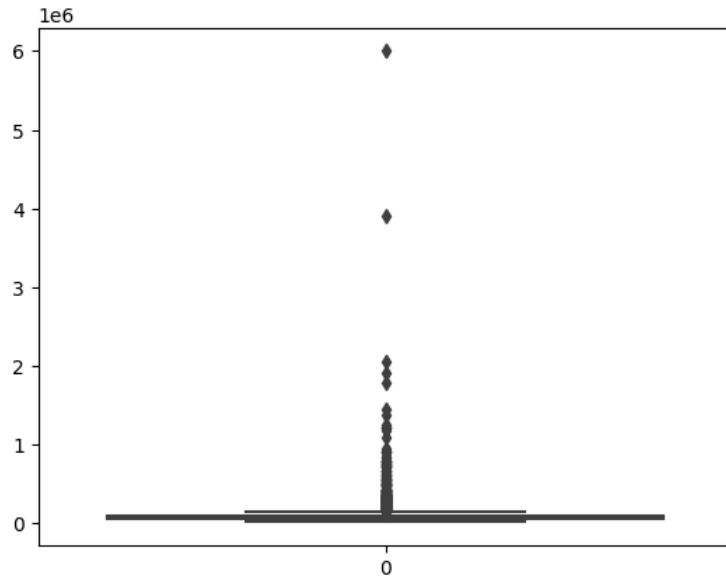
Distribution and variance of data  
Finding outliers and removing them

## Bivariate Analysis

Analyze the impact of continuous and categorical variables on driver variable loan\_status

# Univariate Analysis - 1

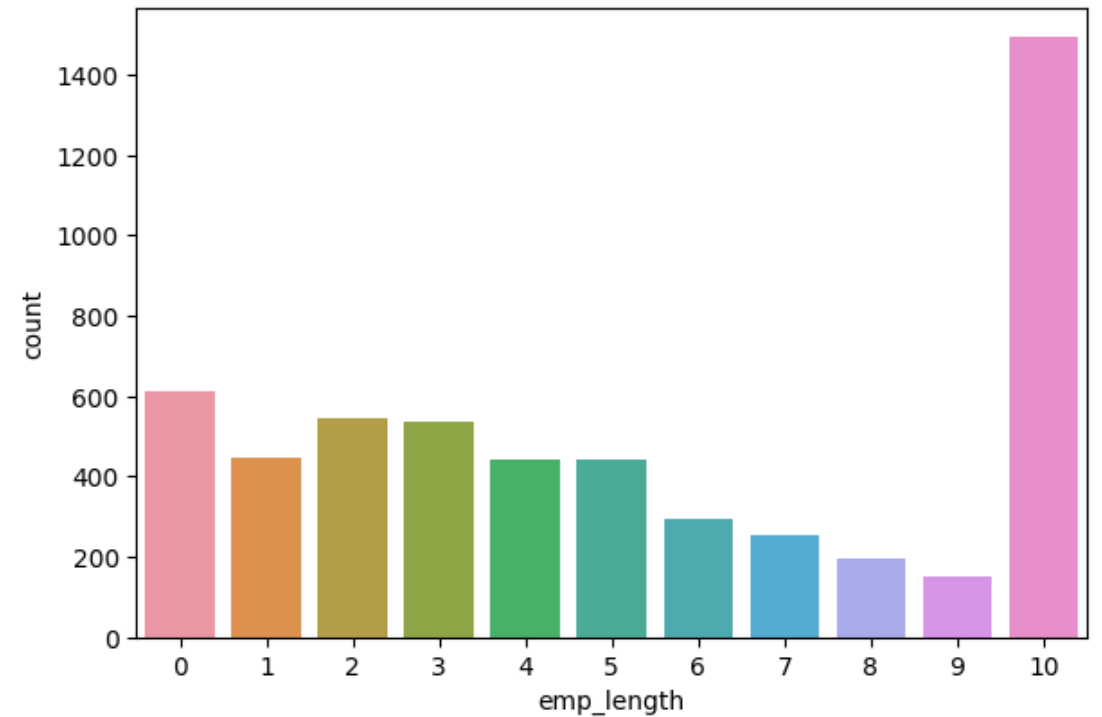
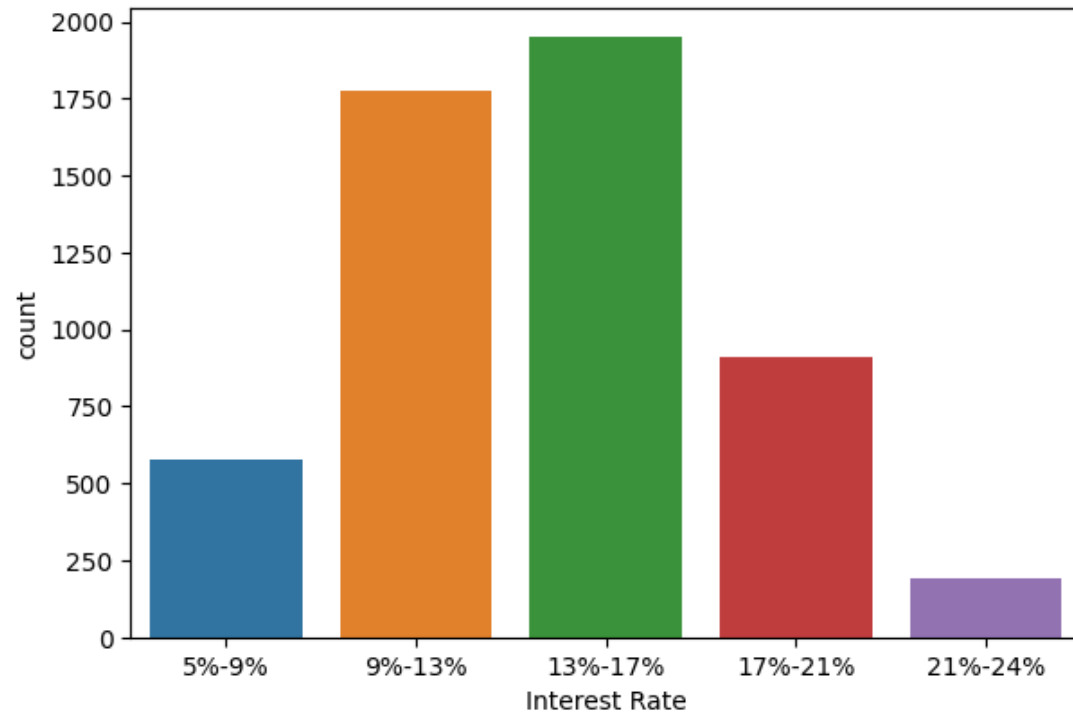
Annual Income - Significant outliers for the field 'annual\_inc'



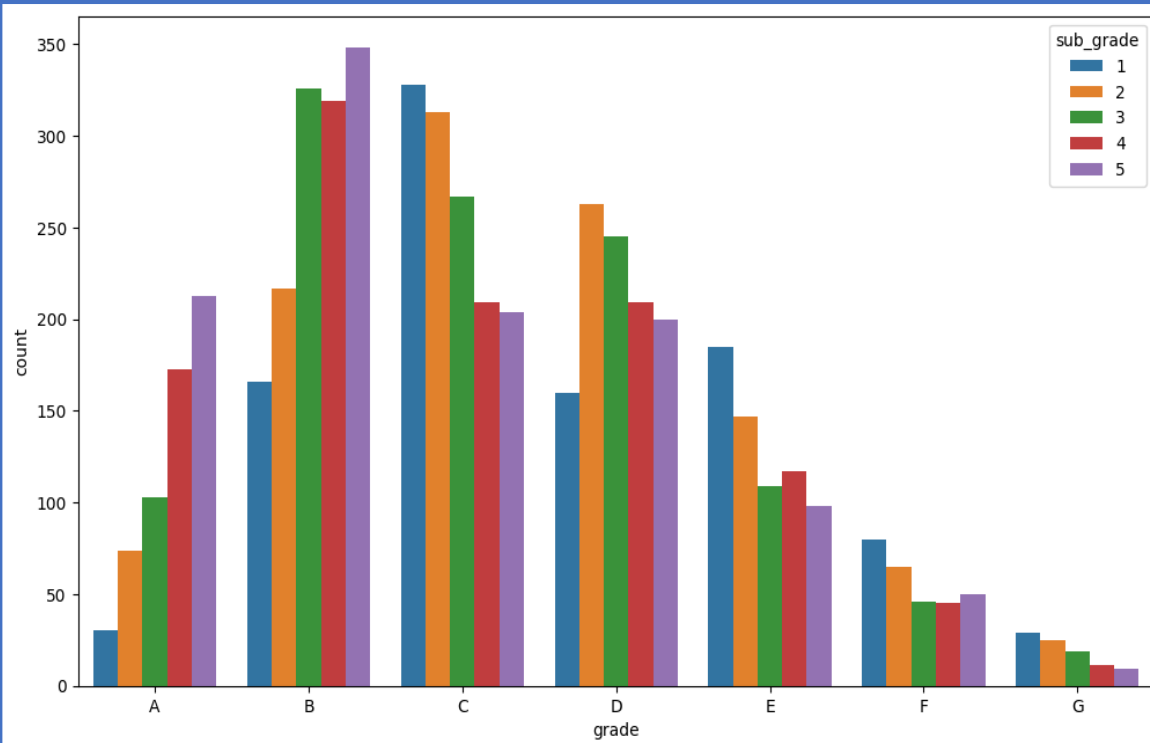
annual\_inc\_groups - A new variable created to cut the data into bins and visualize the data

## Univariate Analysis - 2

Analyzing interest rate with the interest rate bins created and emp\_length



## Univariate Analysis - 3

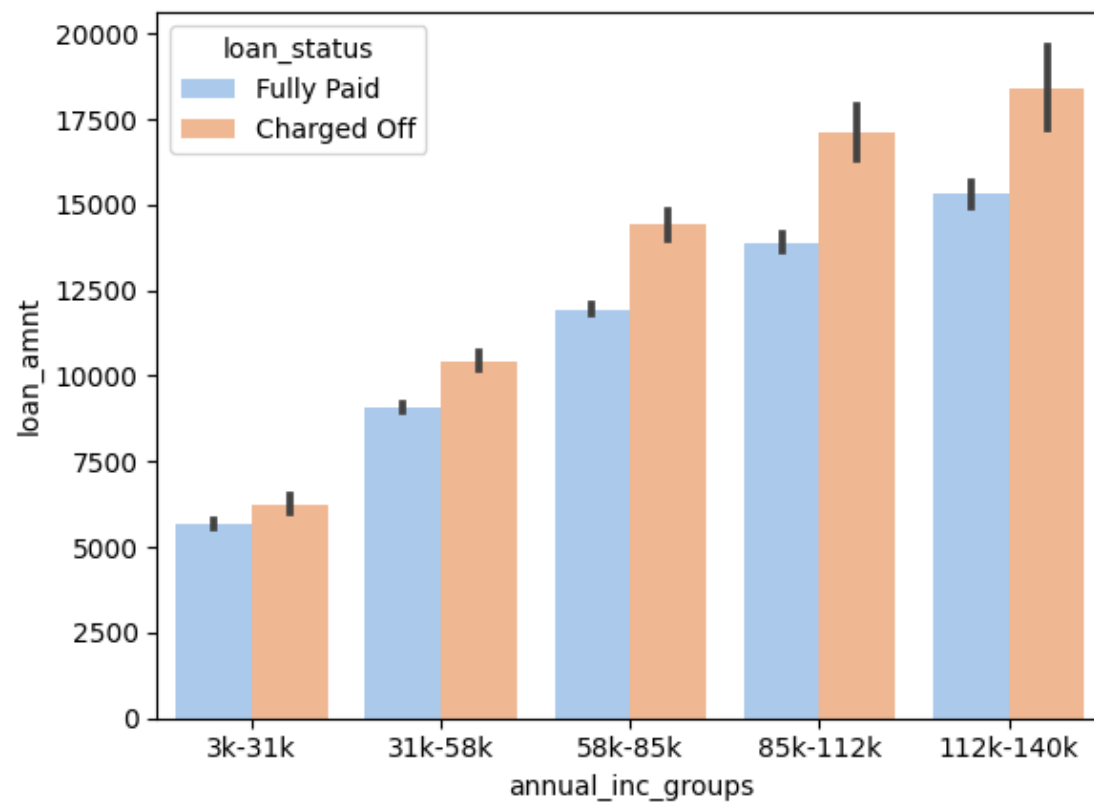


Grade - A,B,C,D are quite dominant amongst all. Grade B counts is almost 350 , whereas grade G is less than 50.



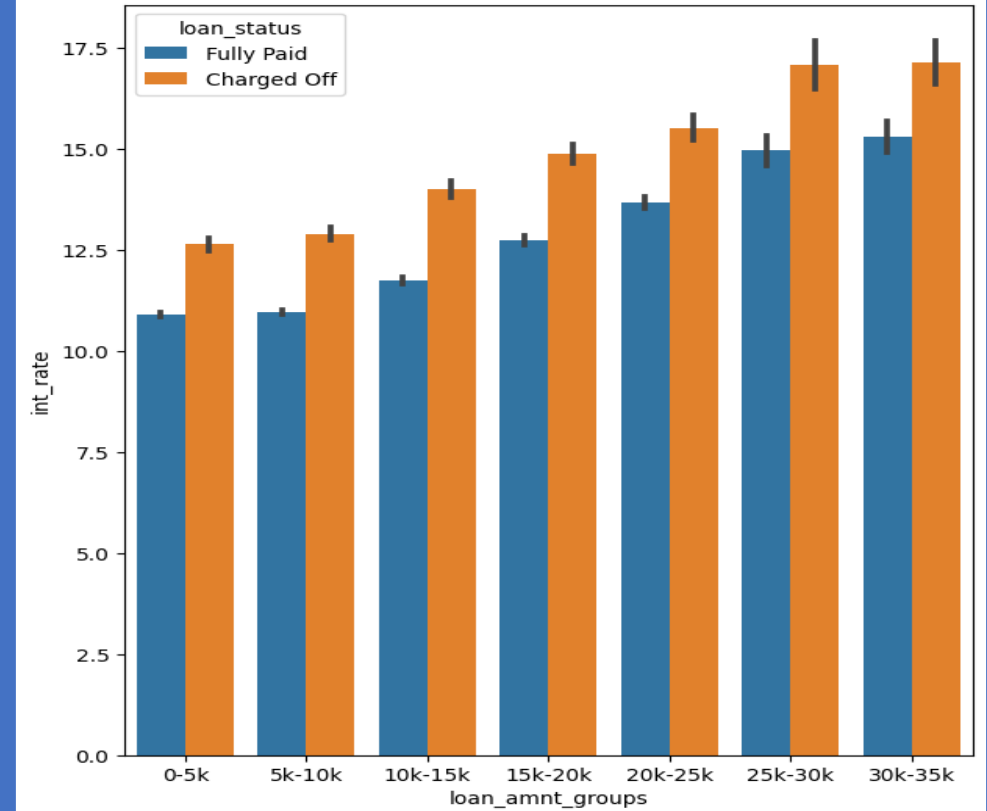
Home Ownership - Most of the loan applicants house type is either Rent or Mortgage. A very small portion of the applicants have their own houses

# Bivariate Analysis - 1



Annual income and Loan amount –

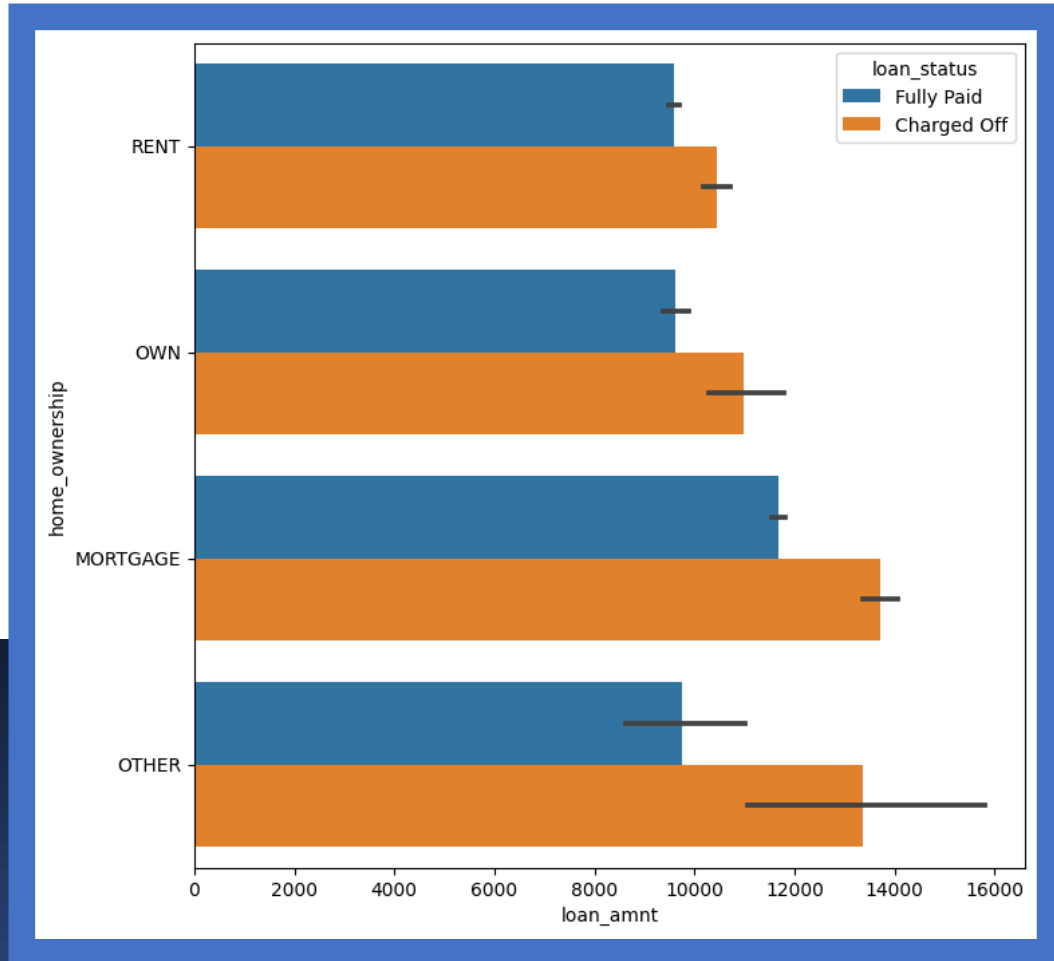
Visualizing the annual income groups wrt loans which are Fully paid or Charged Off



Interest Rate and Loan amount :

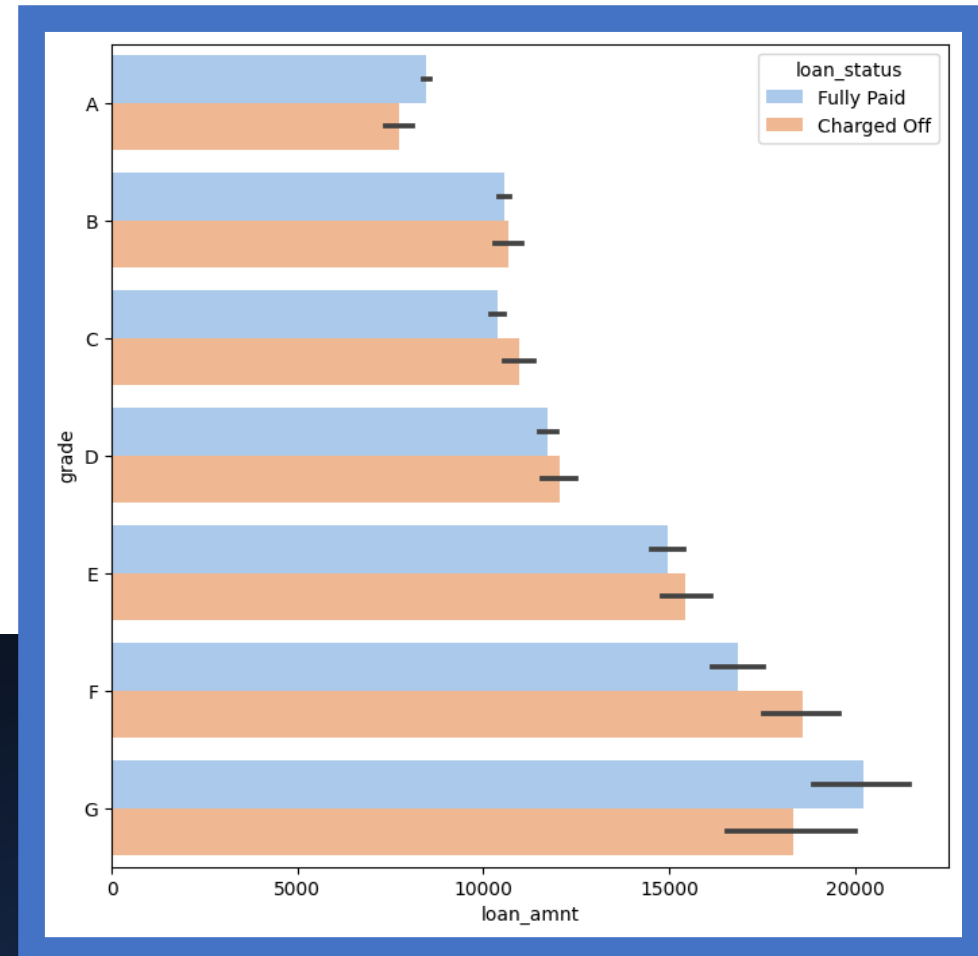
We get a clear indication that chances of loans getting charged off is very low if the interest rate is low.

## Bivariate Analysis - 2



### Loan amount and House Ownership:

Loan amount is higher for people who already has mortgage for home



### Grade and loan amount:

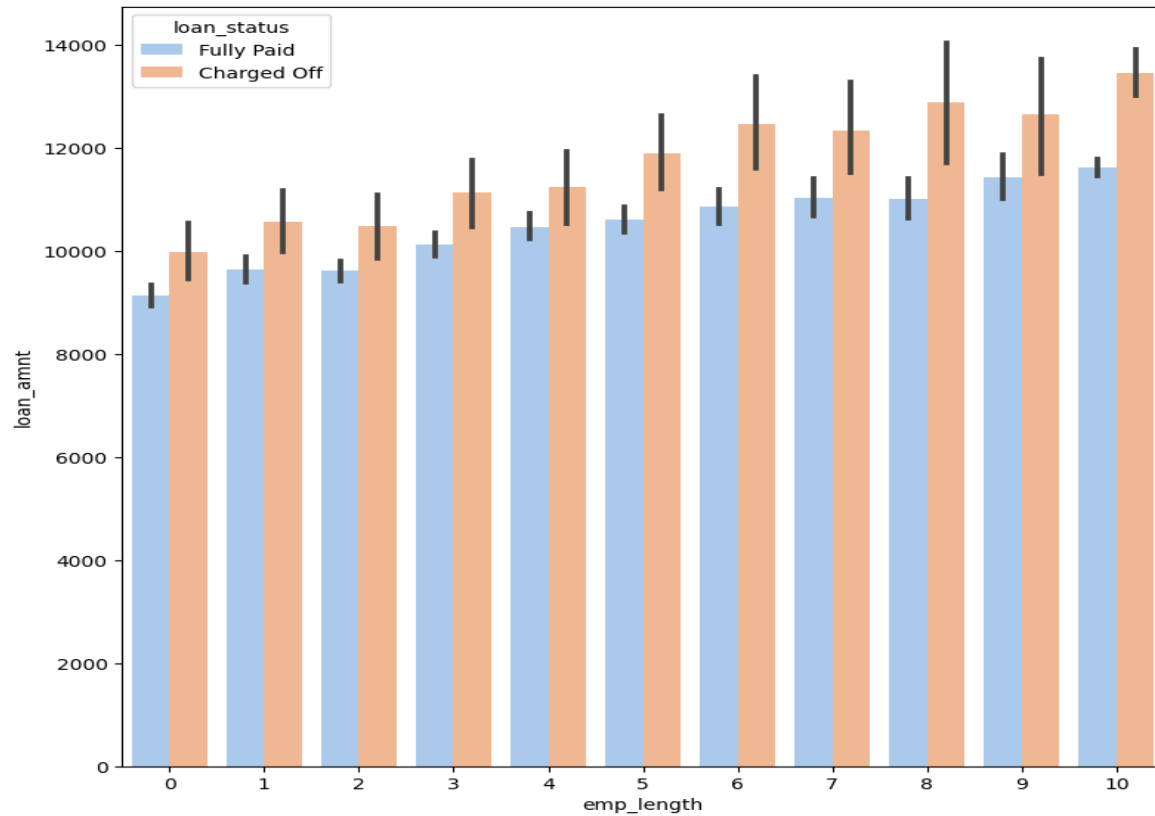
We find an almost linear line between loan grade and the charged off loans. As the loan grade increases the chances of a loan to be charged off increases. It also proves that higher grade loan bears much more credit risk for the lenders.



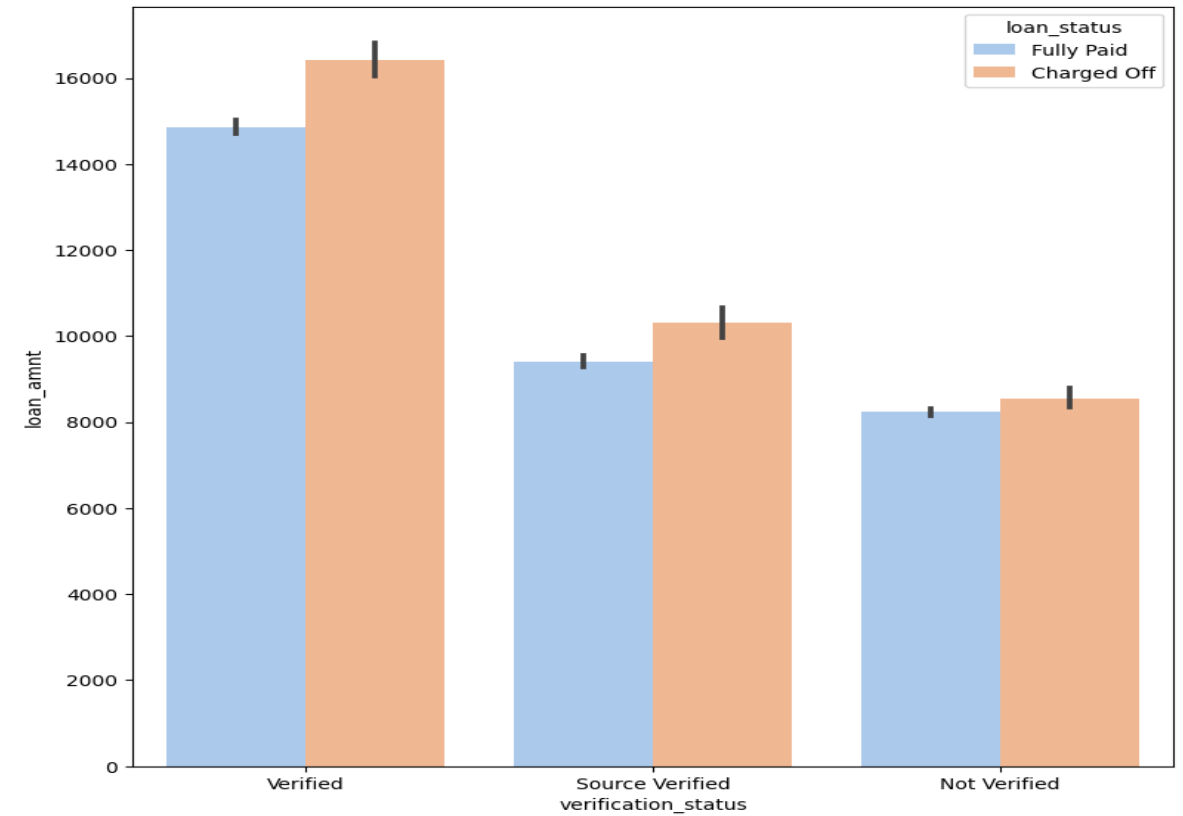
# Bivariate Analysis - 3

## Employment length and Charged Off loan and verification status

We see most loans are charged off when the employment years is 10 years or above



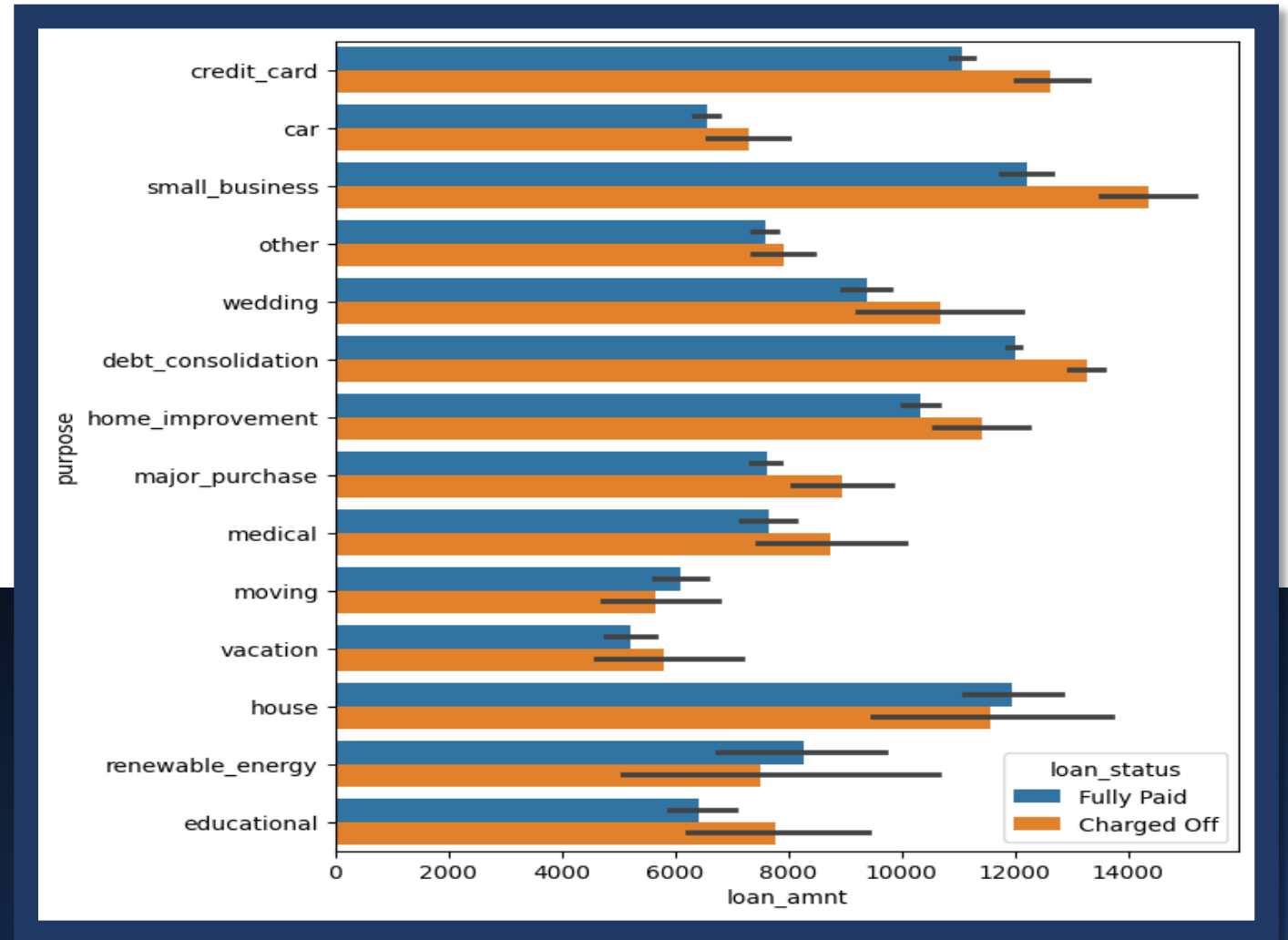
Verification status does not have much influence on Charged off loans.



## Bivariate Analysis - 4

### Purpose and Charged Off loan ratio

We see that for small business the chance is highest that the loan may get charged off. And it is the lowest for loan taken for marriage.



# Inferences

**Based on the EDA with respect to the charged off loans, Probability of defaulting is high when :**

- Applicants who receive interest at the rate of 21-24% and have an income of 70k-80k
- 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 and the loan amount is greater than 14k
- Applicants taking loan for 'home improvement' and have income of 60k -70k
- Applicants whose home ownership is 'MORTGAGE and have income of 60-70k
- Applicants whose home ownership is 'MORTGAGE and have loan of 14-16k
- When grade is F and loan amount is between 15k-20k
- When employment length is 10yrs and loan amount is 12k-14k
- When the loan is verified, and loan amount is above 16k
- For grade G and interest rate above 20%



Thank You!!