



# **Lending Club Case Study**

Creator: Priyanka Jadhav

Case Study: IIIT – Bangalore and UpGrad PG Program

Date: 8<sup>th</sup> March 2023



# Table of Contents

1: General Information (Problem Statement)

2: Data Understanding

3: Data Cleaning

4: Data Analysis

- Univariate Analysis
- Segmented Univariate Analysis
- Bivariate/Multivariate Analysis

5: Observations and Recommendations





# Problem Statement

## **Key Question**

Decipher which types of customers default on a loan.

Lending Club is a consumer finance company that specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily lower-interest-rate loans through a fast online interface.

Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss).

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

The Analysis is divided into four main parts

- 1.Data Understanding
- 2.Data cleaning and Manipulation
- 3.Data Analysis
- 4.Recommendations



# Technology

We have used the following Libraries for this project:

1: Pandas

2: Numpy

3: Matplotlib

4: Seaborn





# Data Understanding

The data given contains information about past loan applicants and whether they 'defaulted' or not.

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

Here we are provided with 39717 records and 111 columns.

Some important columns list:

- loan\_status: shows the status of the borrower's loan - whether its Fully Paid, 'Charged Off' or 'Current'. This is our **target Variable**.  
Here we want to see how other variables/ factors affects on the status of this column. So that in future Lending Club can be prepared in advance to judge a borrower.
- loan\_amnt: Amount applied by the potential borrowers
- funded\_amnt: Amount recommended/approved by Lending Club
- funded\_amnt\_inv: Amount funded by investors
- term: The number of payments on the loan (values are mentioned in months)



# Data Cleaning

1. There are many columns having missing values. Some are having 100% missing values and few are having more than 30%. We have dropped all these values because they are not going to help us in our analysis.
2. There are few columns which datatypes are mentioned as object but it should be numeric. We have converted all variables in the recommended format.




# Data Analysis

The objective is to identify predictors of default so that at the time of loan application, we can use those variables for approval/rejection of the loan.

There are broadly 3 types of variables:

1. Those which are related to the applicant(demographic variables such as age, occupation, employment details etc.)
2. Loan characteristics (amount of loan, interest rate, purpose of loan etc.)
3. Customer behavior variables (those which are generated after the loan is approved such as delinquent 2 years, revolving balance, next payment date etc.)





Customer behavior variables are not available at the time of loan application, and thus they cannot be used as predictors for credit approval.

There are 3 unique values in our target column

1. Fully Paid
2. Current
3. Charged Off

Here Current status means its neither fully paid or charged off. So we got rid of it.





# Outlier Treatment

We have analyzed following few numerical columns where we found the possibly have outliers using box plots:

1. annual\_inc
2. dti
3. loan\_amnt
4. funded\_amnt\_inv

After viewing boxplot visualization and quantile information:

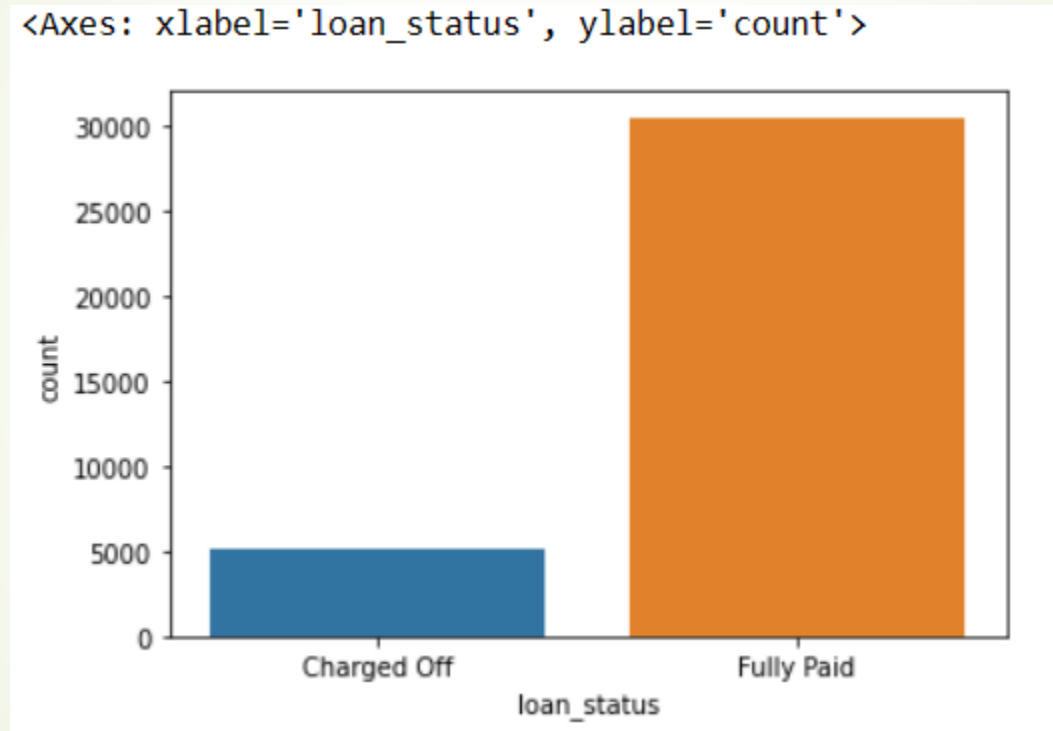
We had removed outliers from annual income and for rest other columns values distribution is far but distribution is continuous that is why no need to remove these values.



# **Visualization**

# Visualizing categorical data

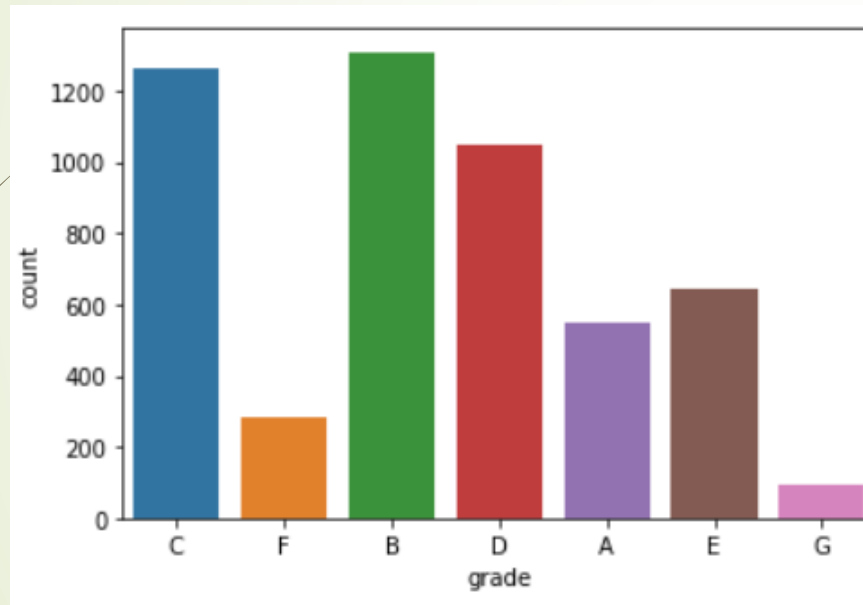
## 1. loan\_status



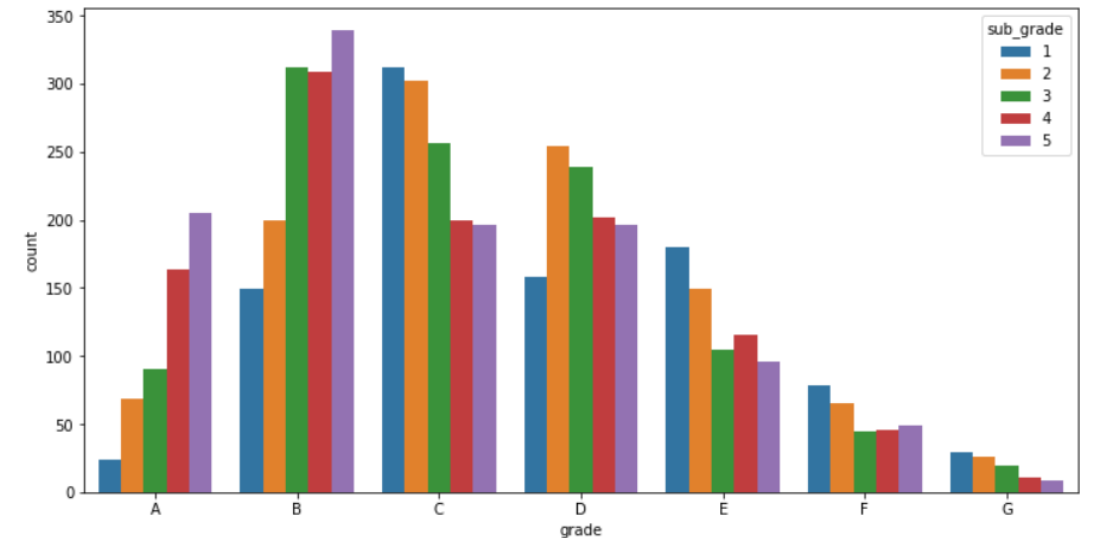
Defaulter rate is 15% appx.

We are analyzing and visualizing only the defaulter data. So sub setting the data while plotting only for '**Charged Off**' loan\_status for below plots.

## 1. Grade & along with Sub\_grade



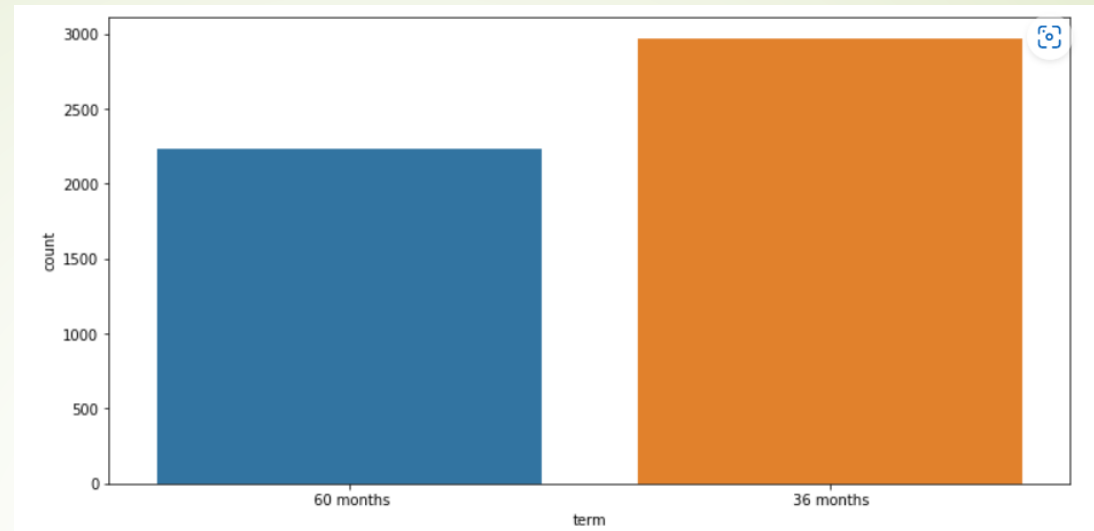
<Axes: xlabel='grade', ylabel='count'>



There are many defaulters with Grade B & C.

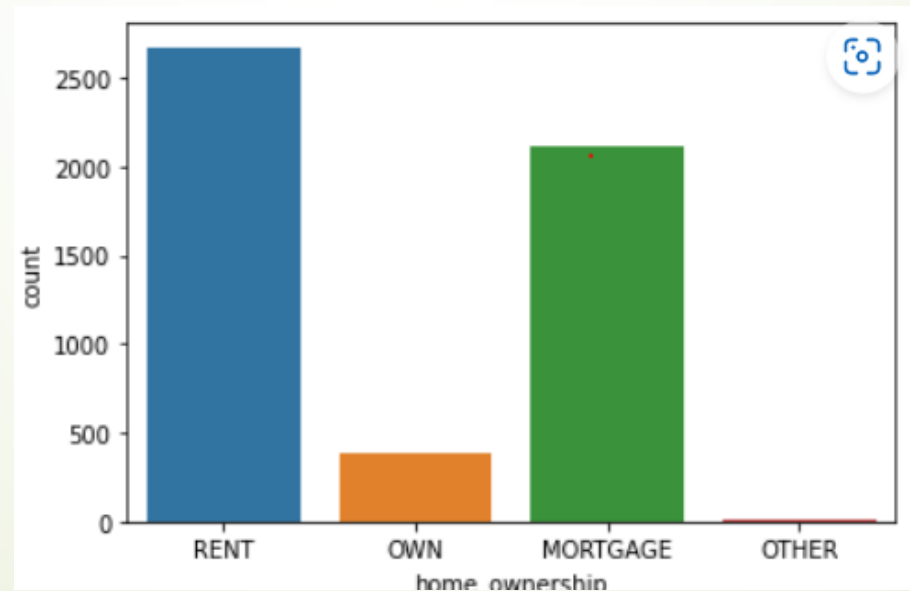
Here if we look at Grade A and its sub\_grade then A1 is better than A2 and so on

## 2. Term



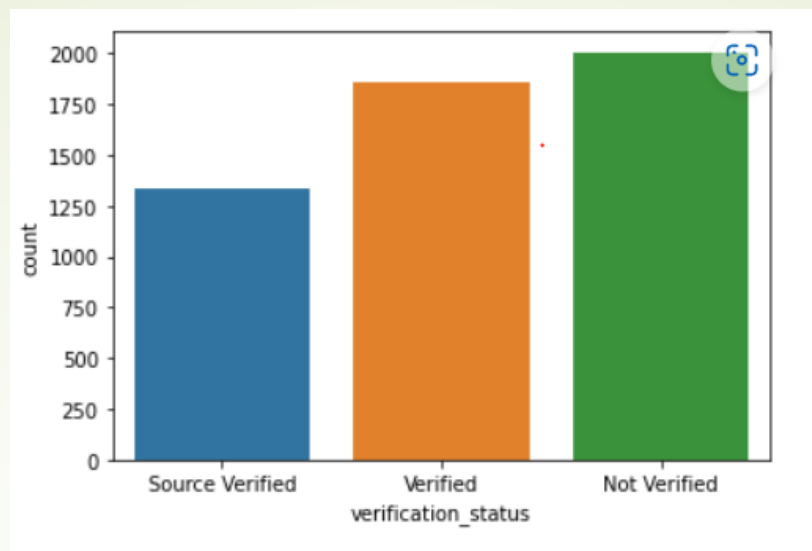
There are more defaulters for 60 months loan than 36 months.

## 3. Home\_ownership



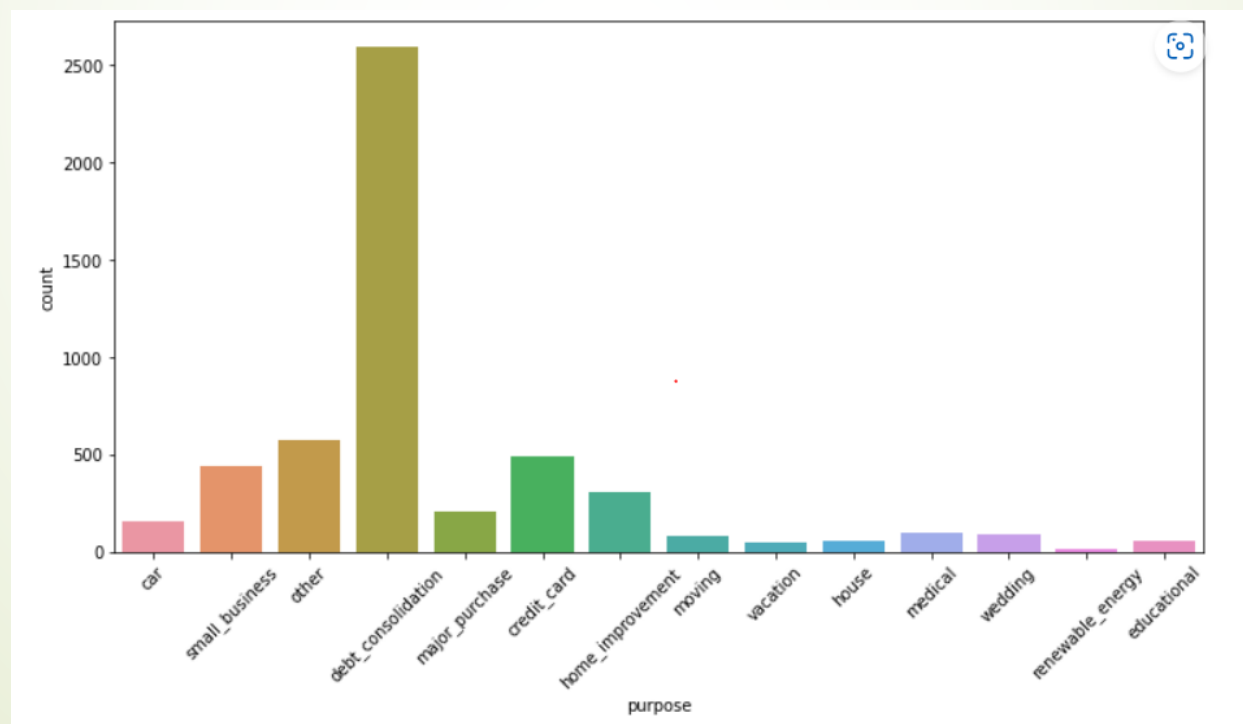
Defaulters are more for whom home ownership is 'RENT'

#### 4. Verification\_status



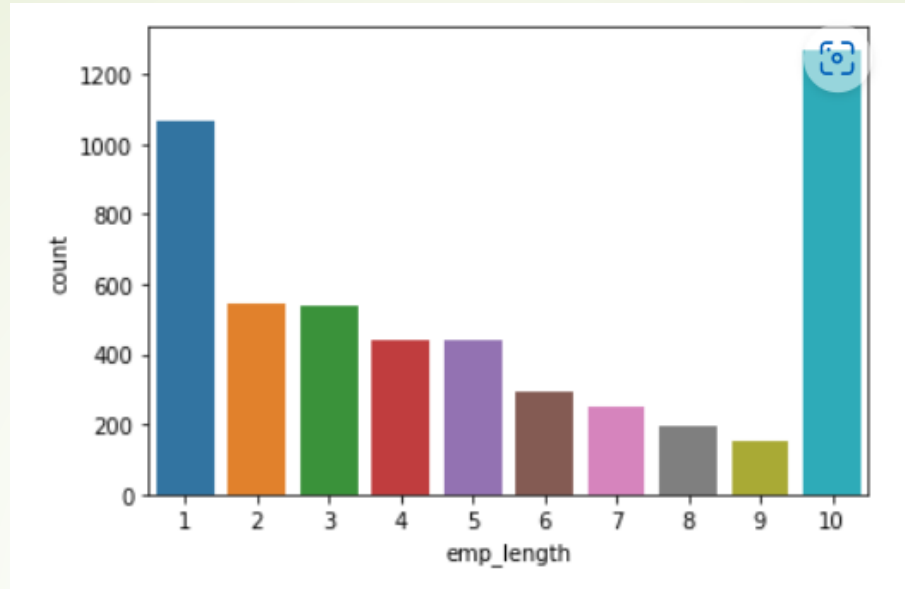
There are maximum defaulters for whom verification has not been done. Source verification has become the fruitful to reduce this number compare to Not verified.

#### 5. Purpose



There are more defaulters who has mentioned there purpose for loan as debt\_consolidation

## 6. emp\_length



More defaulters are the people who are employed 10 or more than 10 years. There are slight less number of defaulters are from the group where people are employed for 1 year.

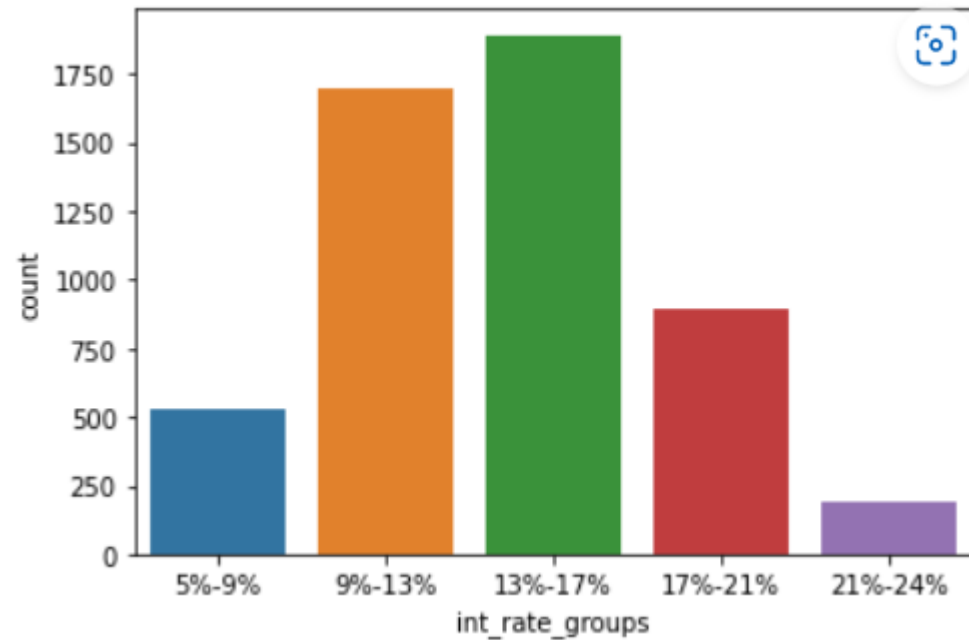
### Analysis using numerical variables as categorical variables

We can convert numerical variables into categorical using bins for the purpose of analysis. We will use here following columns:

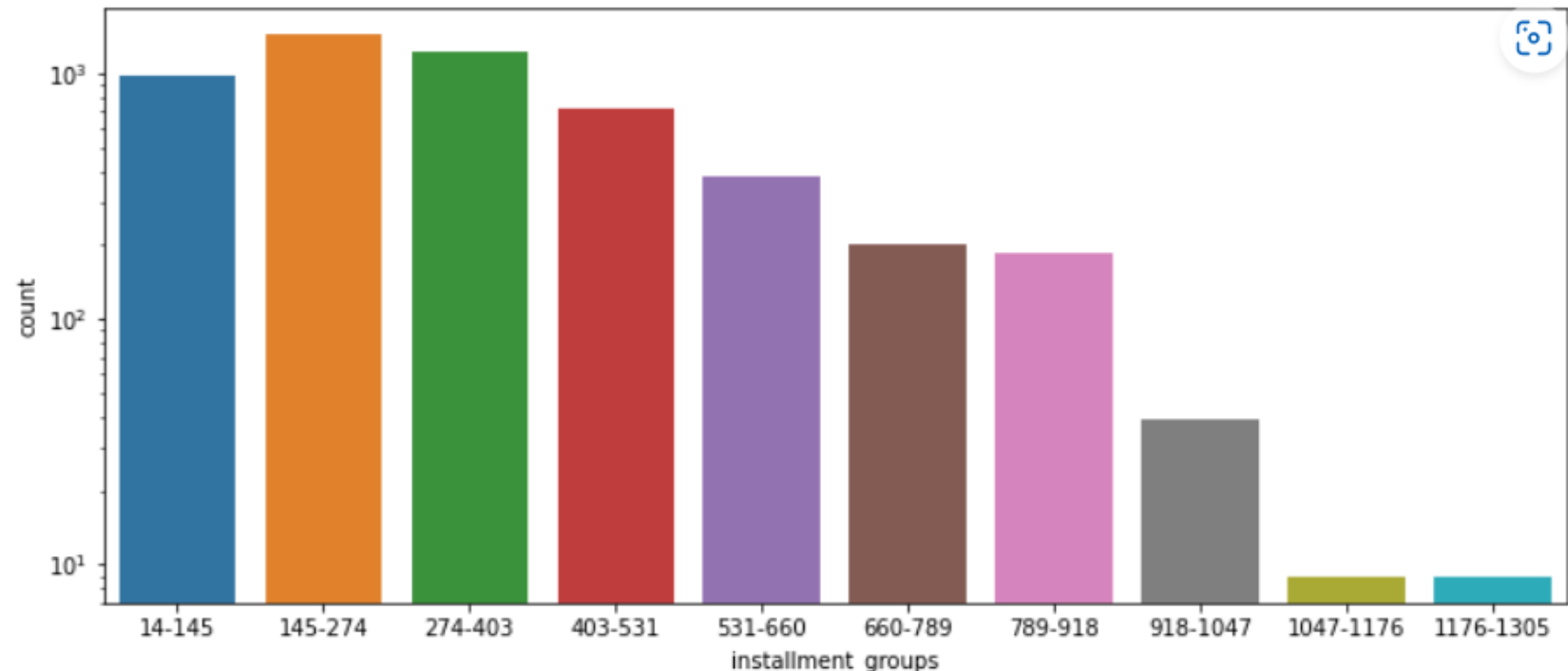
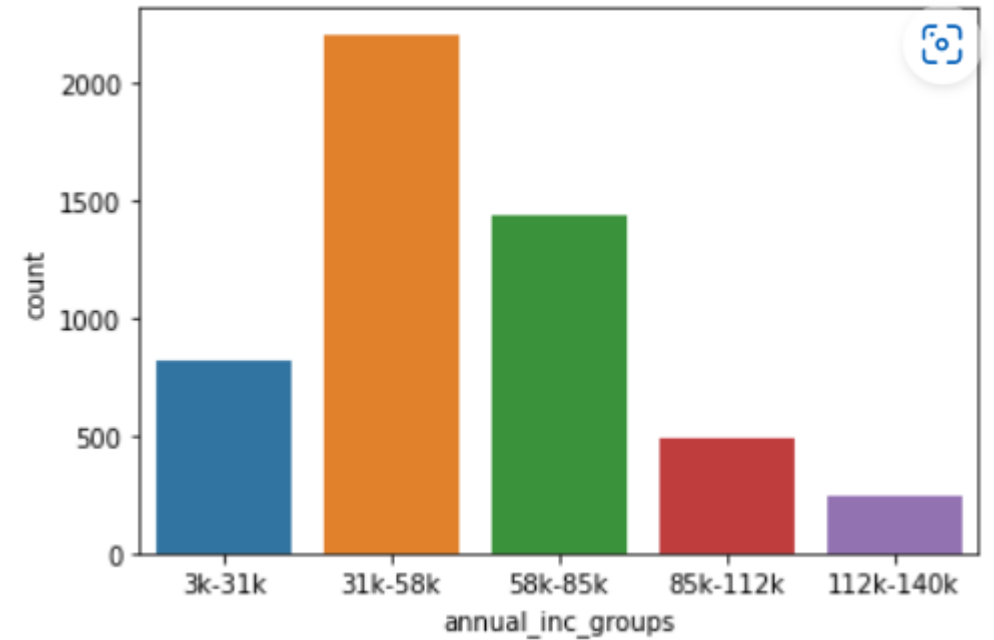
1. int\_rate
2. annual\_inc
3. installment
4. funded\_amnt\_inv
5. loan\_amnt
6. dti



## 1. Interest rate groups

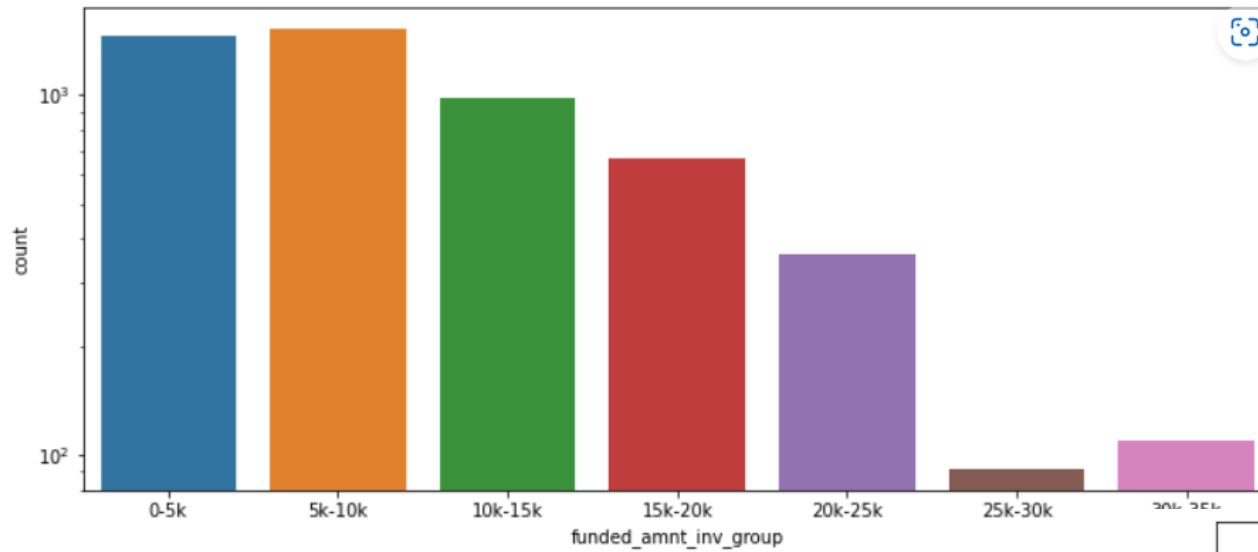


## 2. Annual Income groups

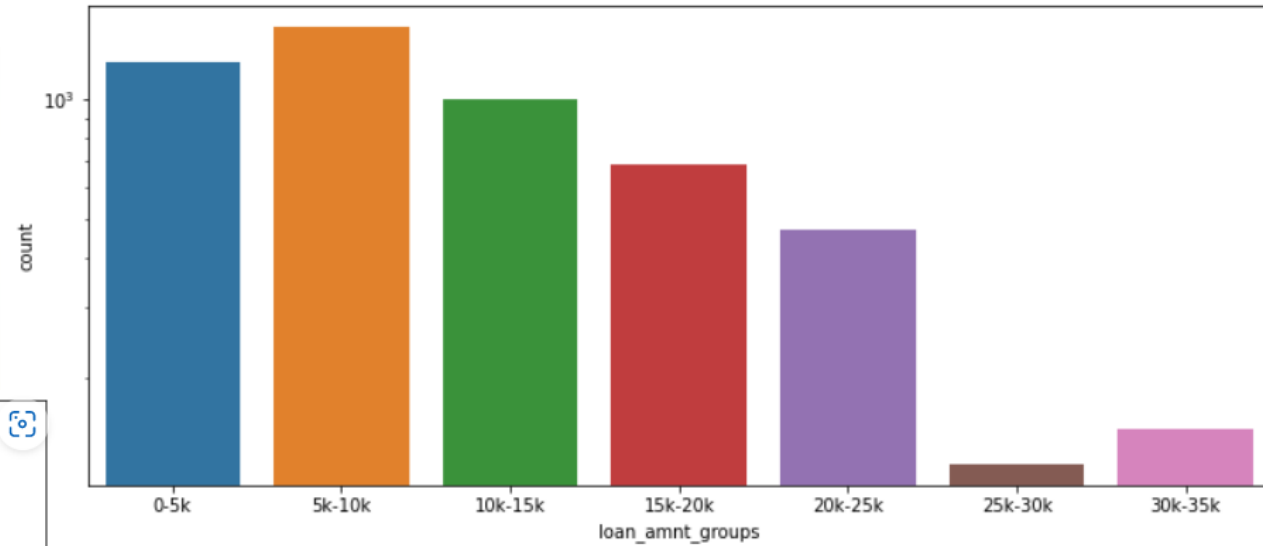


## 3. Installment groups

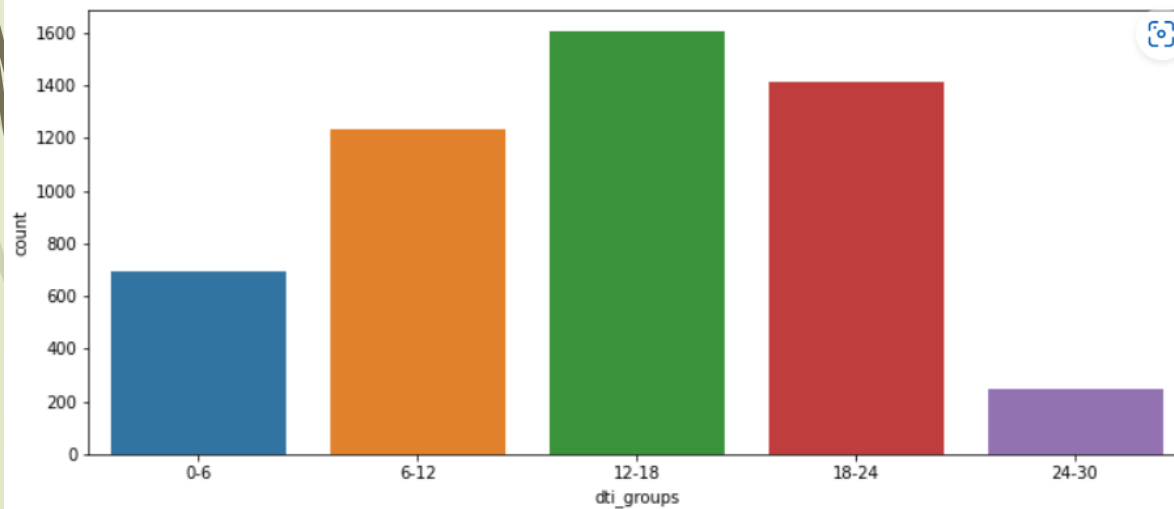
#### 4. Funded Amount Investment Group



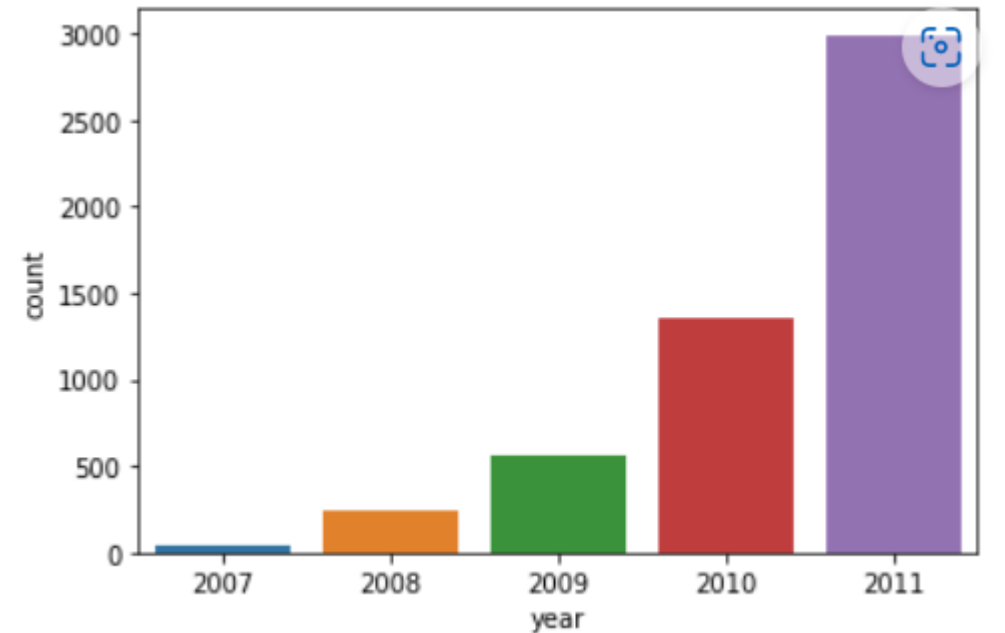
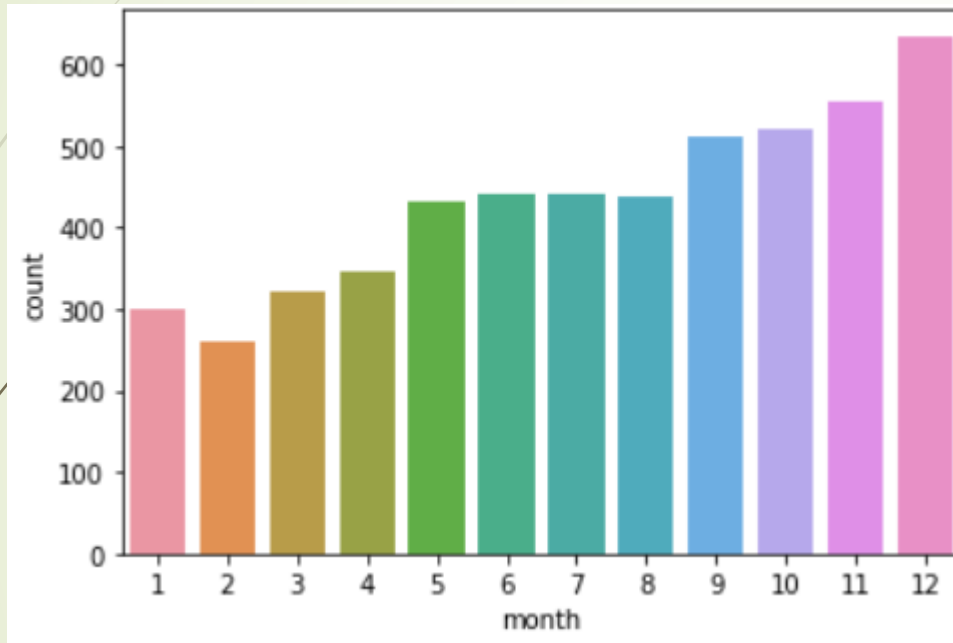
#### 5. Loan Amount groups



#### 6. Dti groups



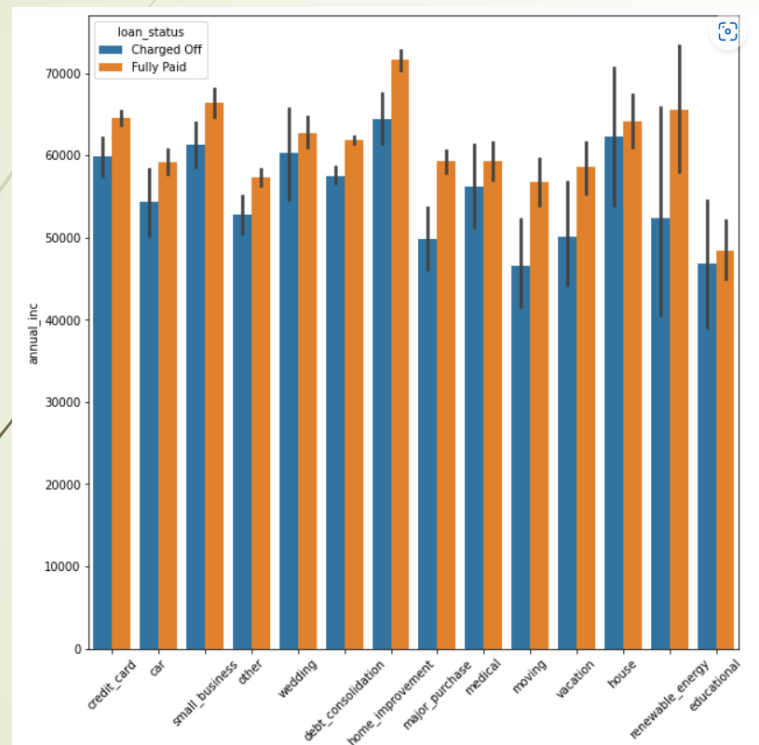
## Distribution of loans across years



We can see that default rate is increasing year by year but a sudden increase is there for the year 2011. and there are more defaulters whose loan is sanctioned in the month of December

# Analysing annual income with other columns for more insights

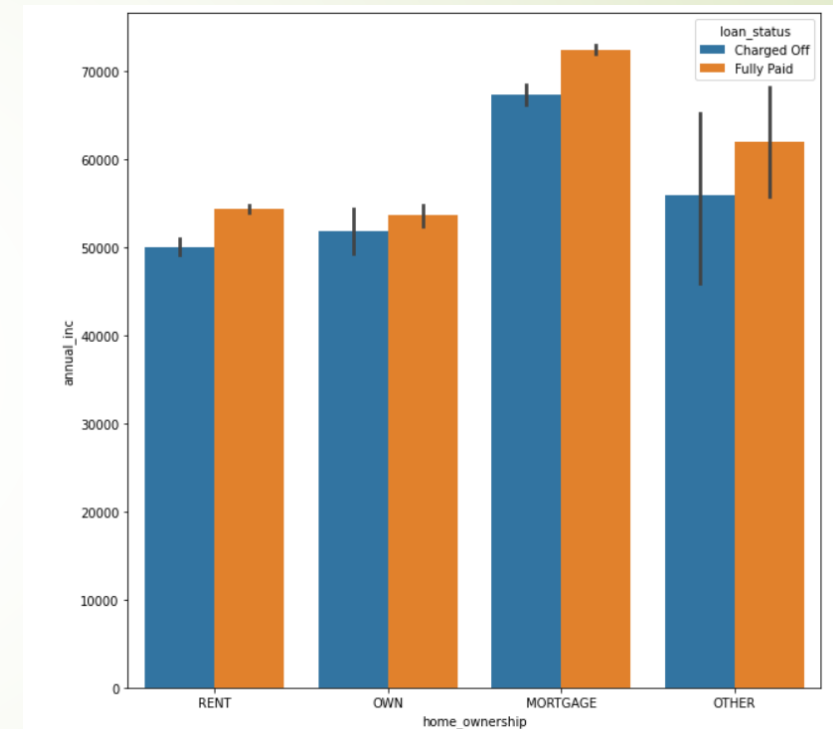
## 1. Purpose



Applicants with higher salary mostly applied loans for "home\_improvement", "house", "renewable\_energy" and "small\_businesses"

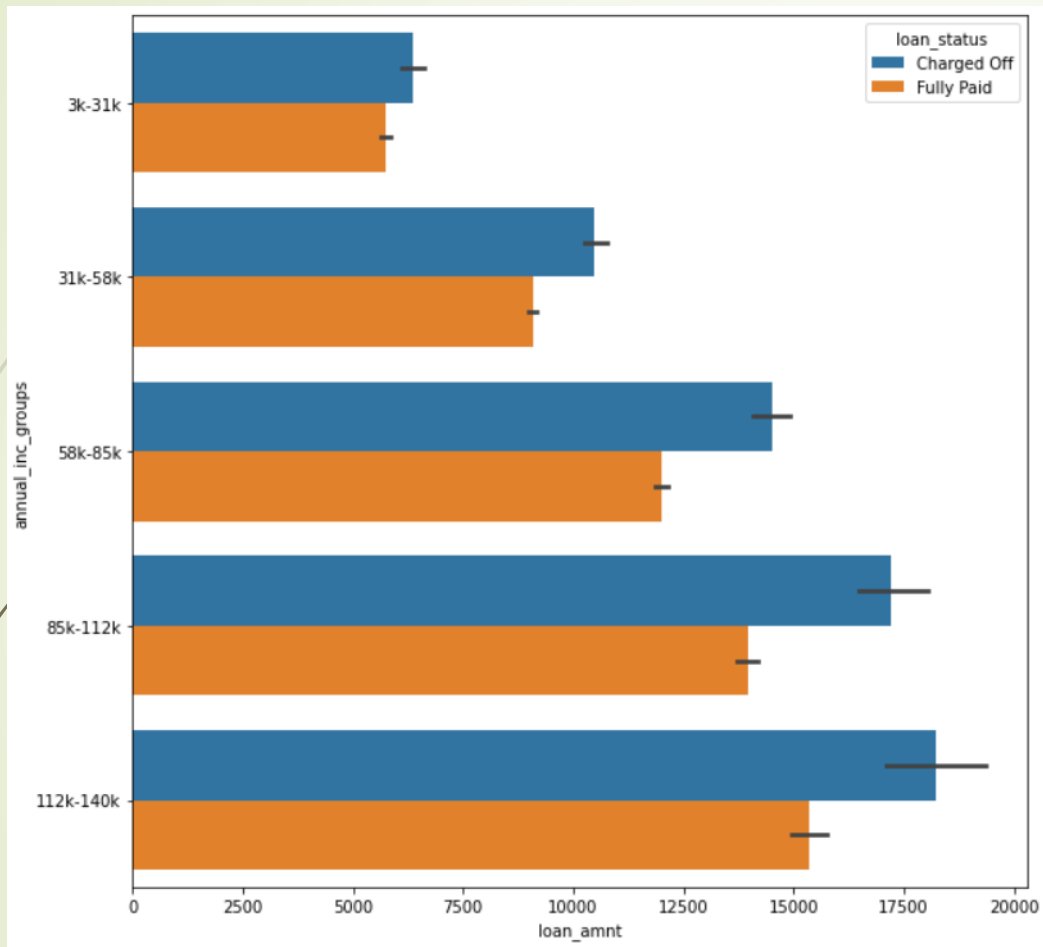
and though the number of defaulters are highest whose purpose for loan was debt\_consolidation, there annual income is not highest.

## 2. Home Ownership



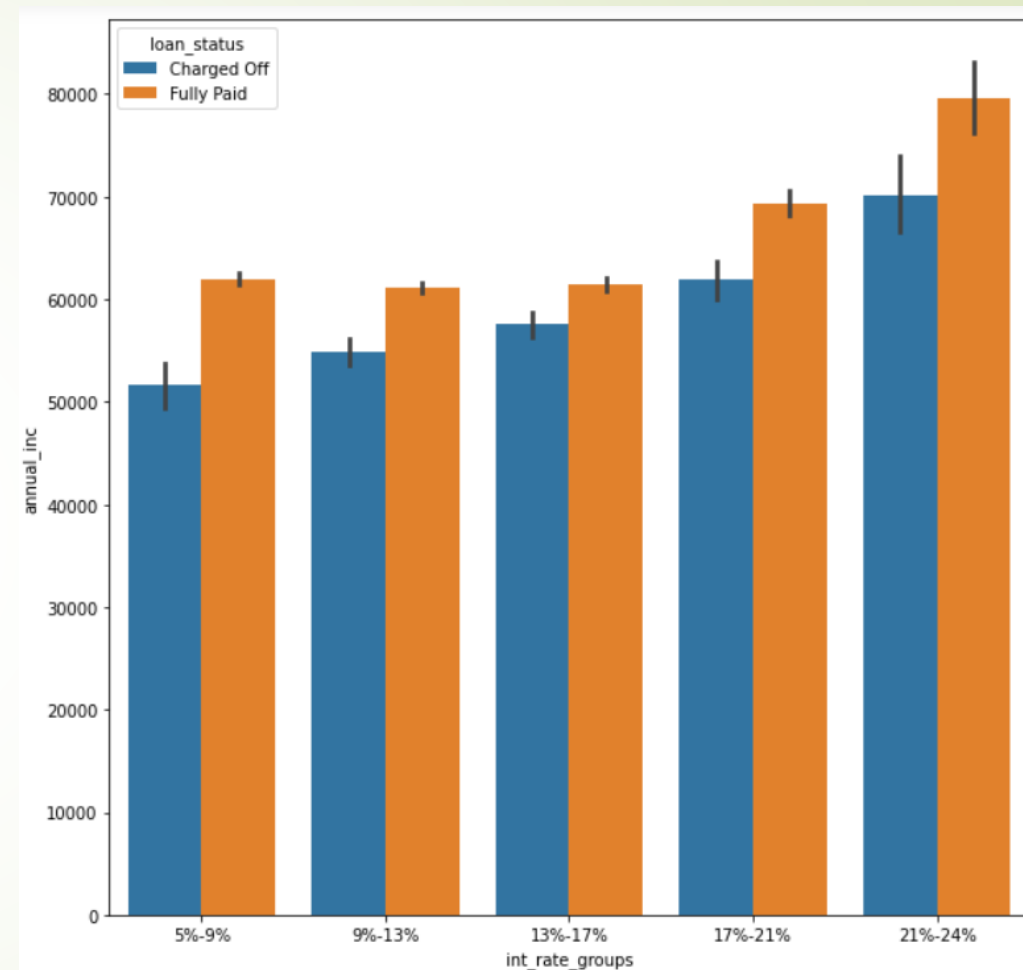
Annual income of defaulters is highest whose home\_ownership is MORTGAGE.

### 3. Loan Amount



Here we can see for all annual income groups defaulters are having higher loan amount. As the loan amount is increasing defaulter rate is also increasing.

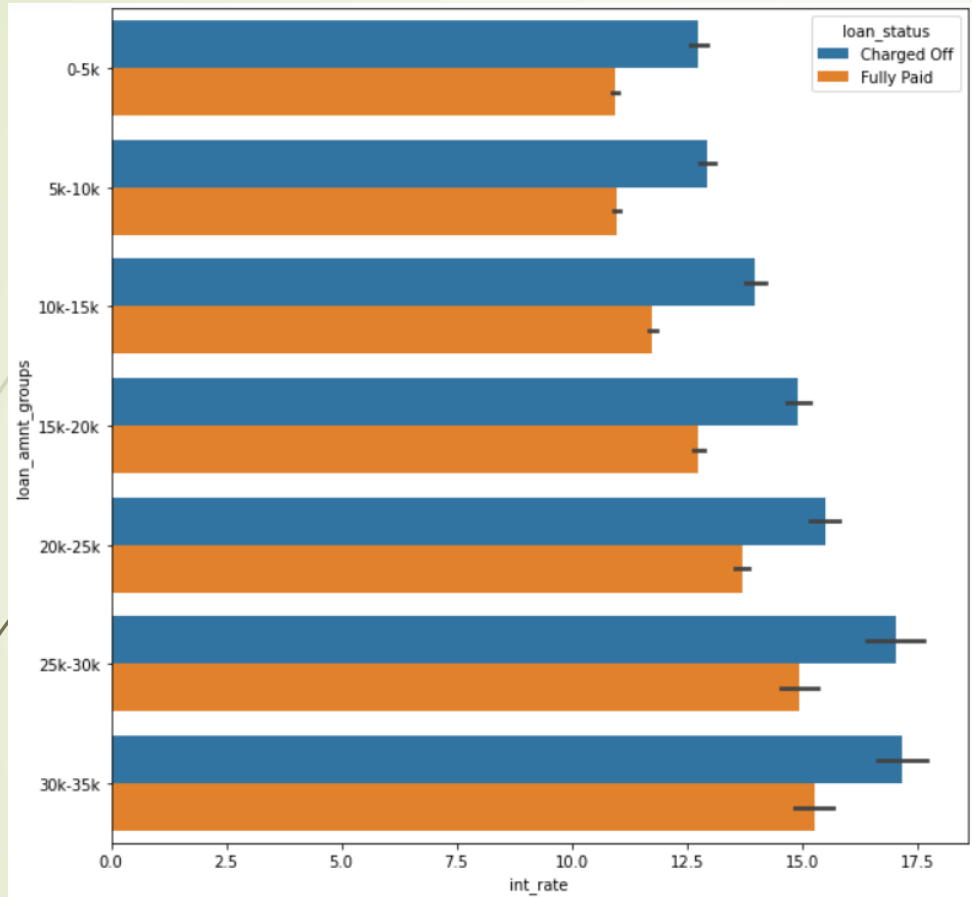
### 4. Interest Rate



Highest defaulters annual\_income is also high and they have got the loans with higher interest as well.

# Analysing loan\_amount with other columns for more insights

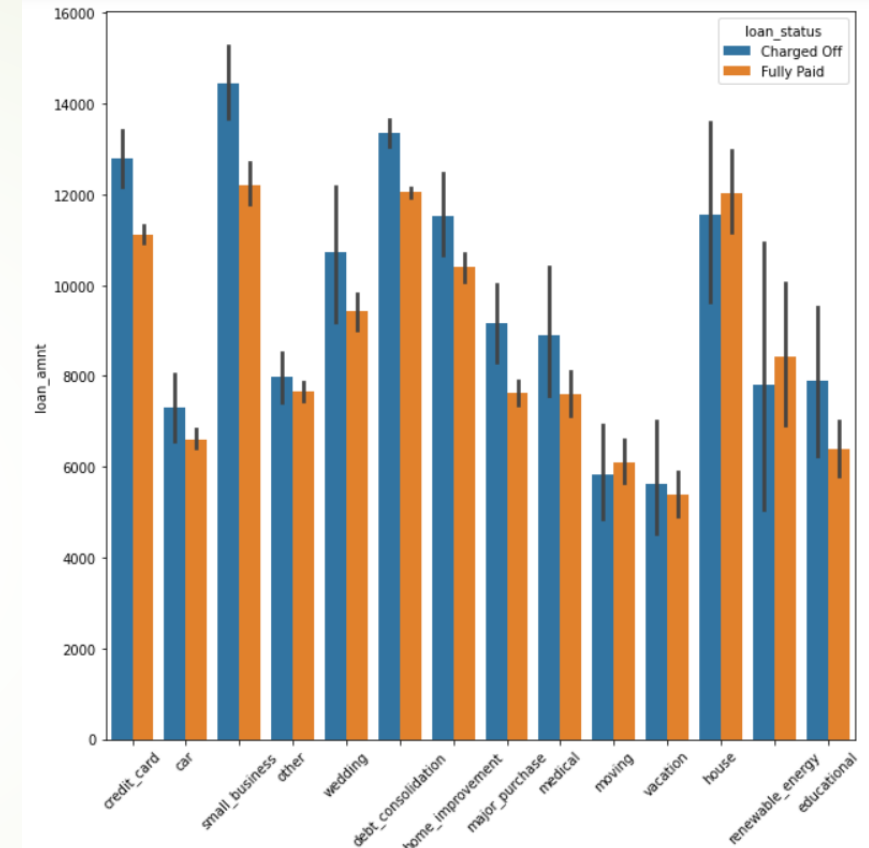
## 1. Interest Rate



As loan amount is increasing, interest rate is also increasing and defaulter rate as well along with this.

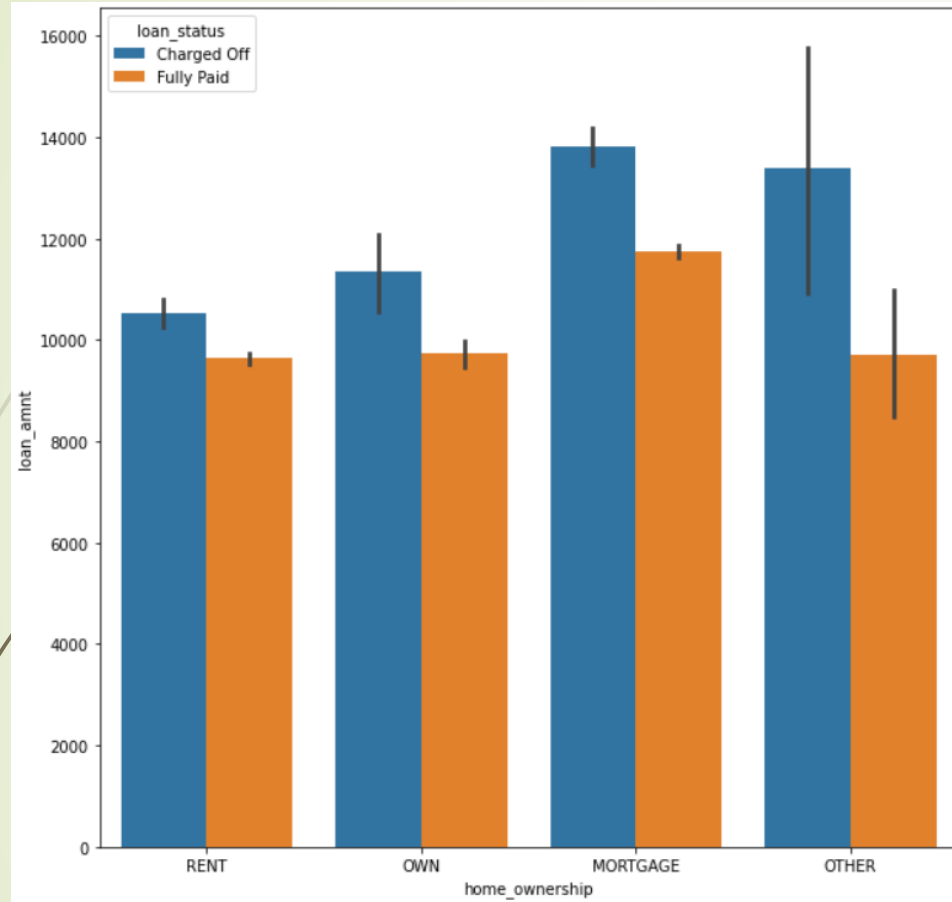
The interest rate for charged off loans is pretty high than that of fully paid loans in all the loan\_amount groups. This can be a pretty strong driving factor for loan defaulting.

## 2. Loan Purpose



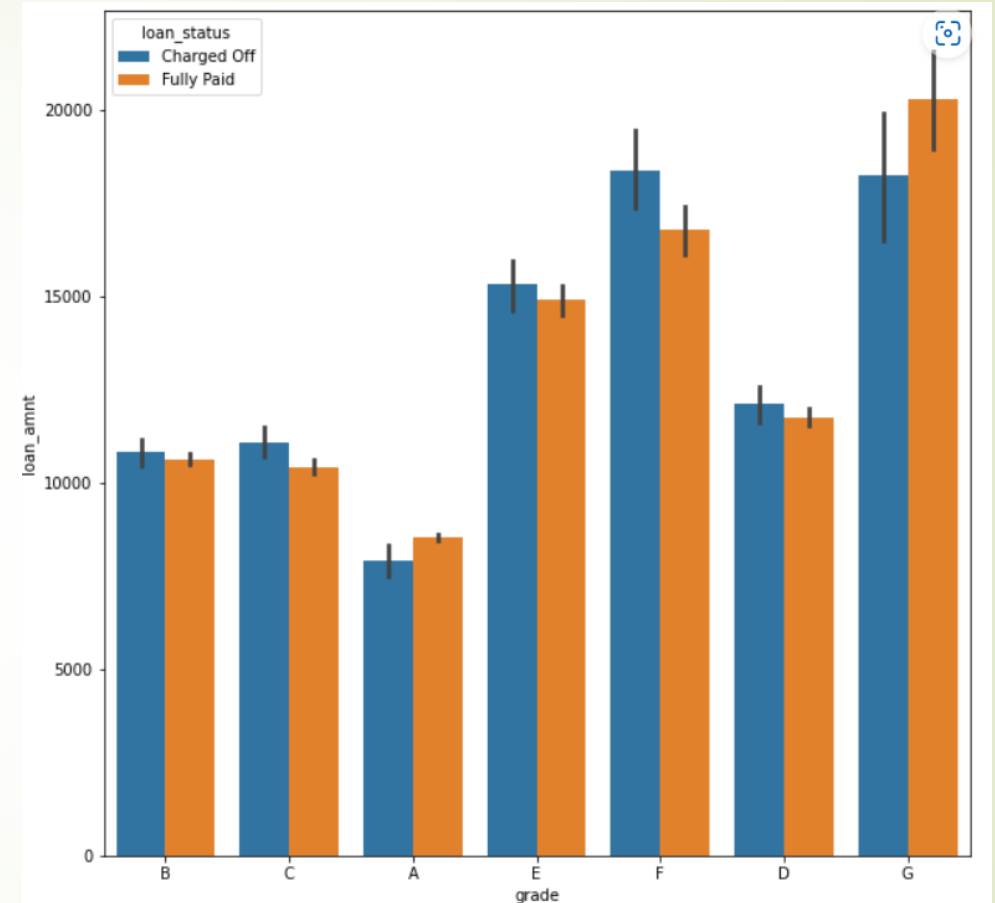
Defaulters have mentioned the purpose for loan is small\_business, received the highest loan amount.

### 3.Home Ownership



We have seen that highest number of defaulters has home\_ownership status as RENT but the one whose status is MORTGAGE has received or avail the highest loan amount.

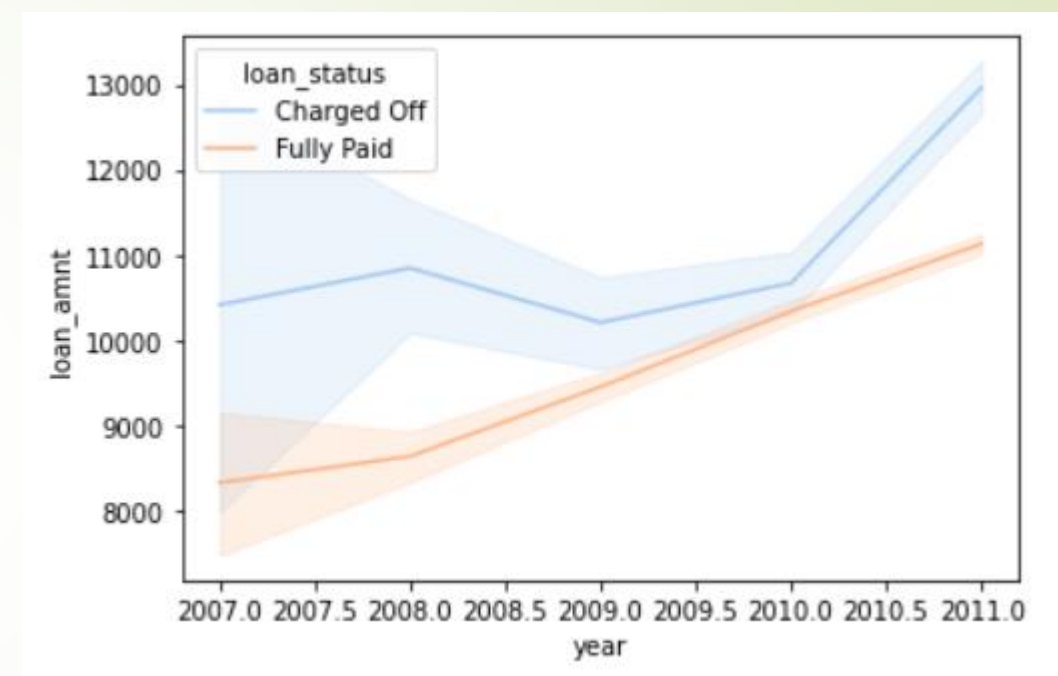
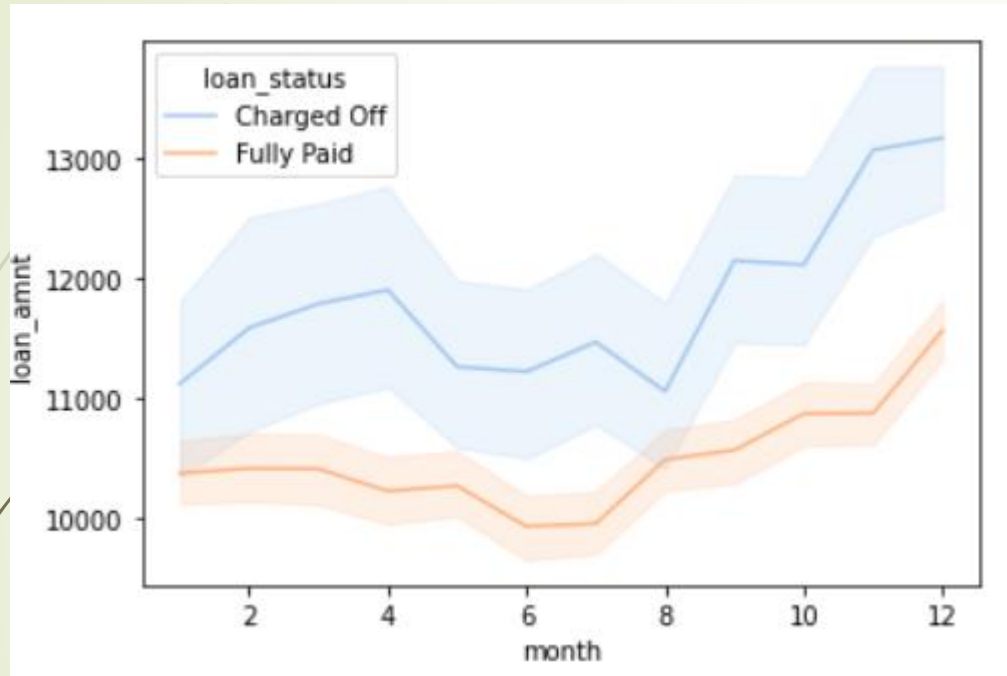
### 4. Grade



Defaulters with grade F & G have received highest loan amount.

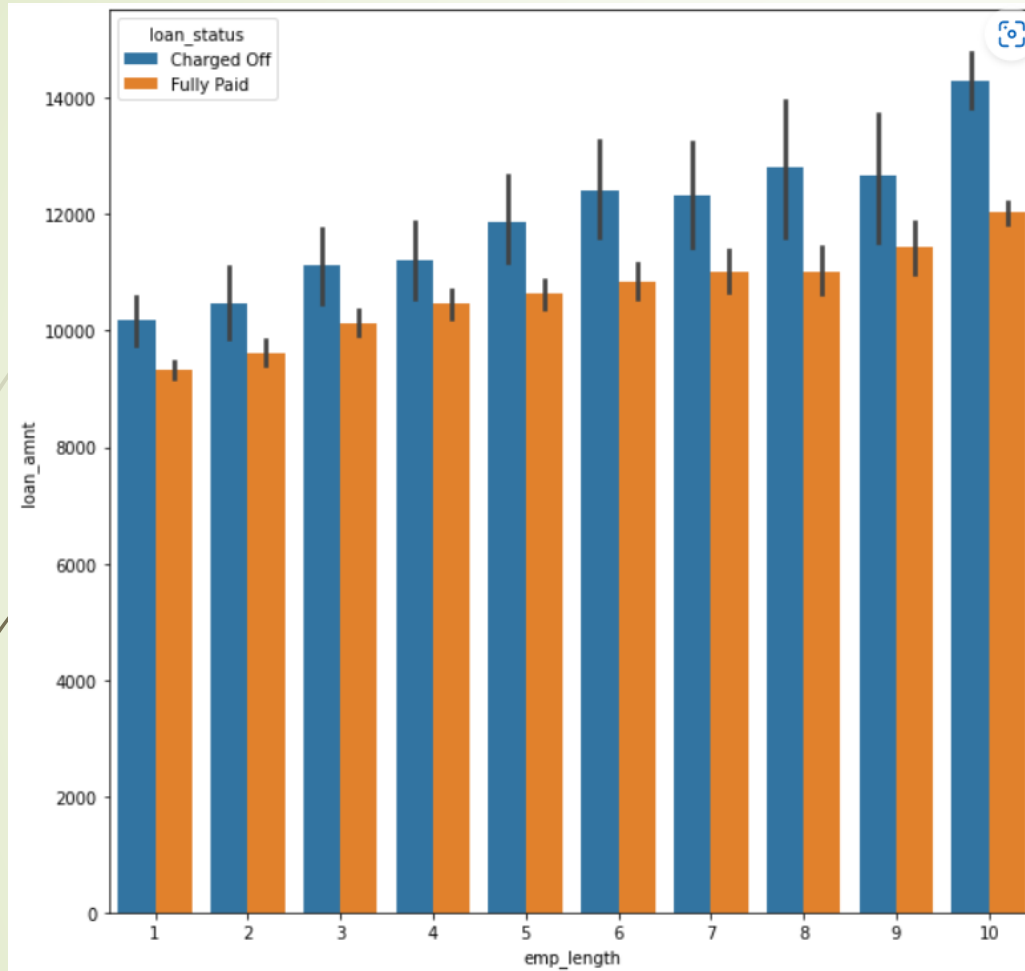


## 5. Loan amount vs month issued and year issued



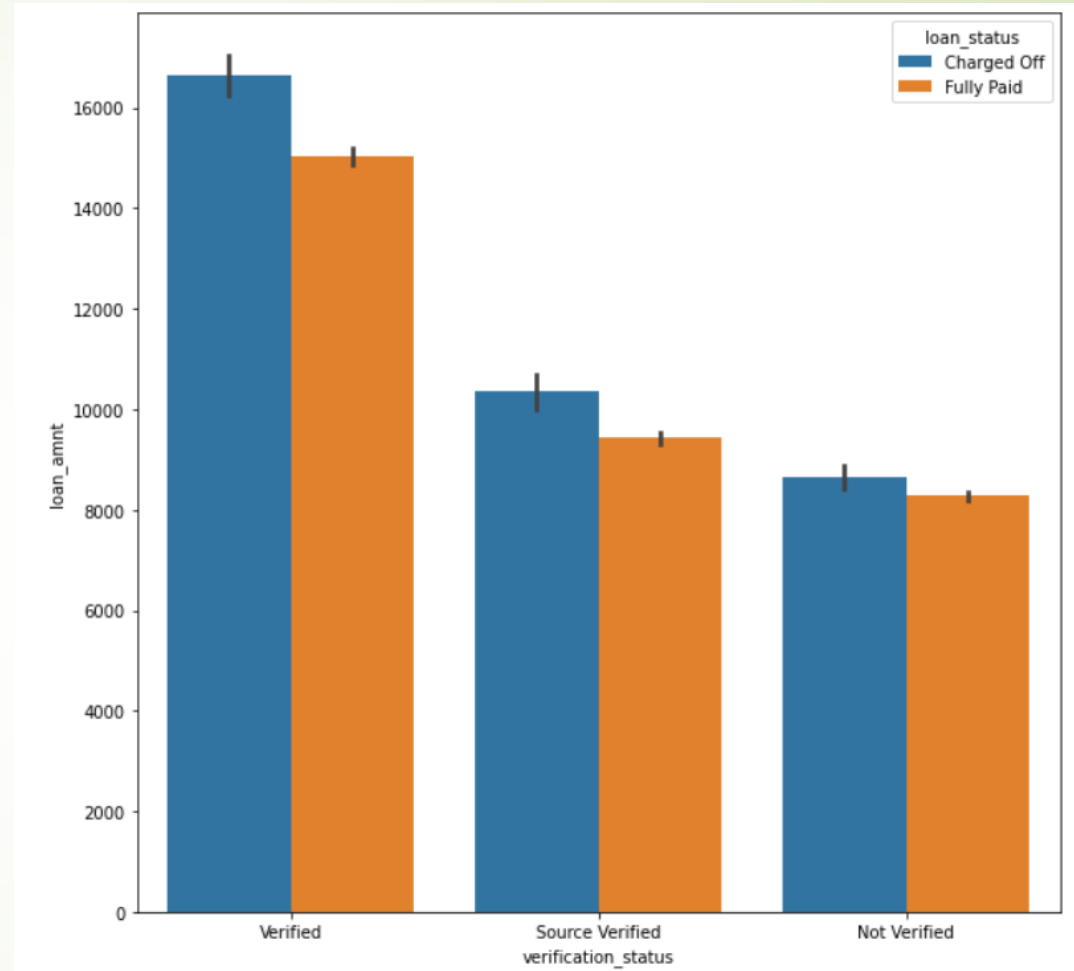
We have seen we were having highest number of defaulters in the year 2011 and here we can also see that loan amount issued in year 2011 is also highest.

## 6. emp\_length



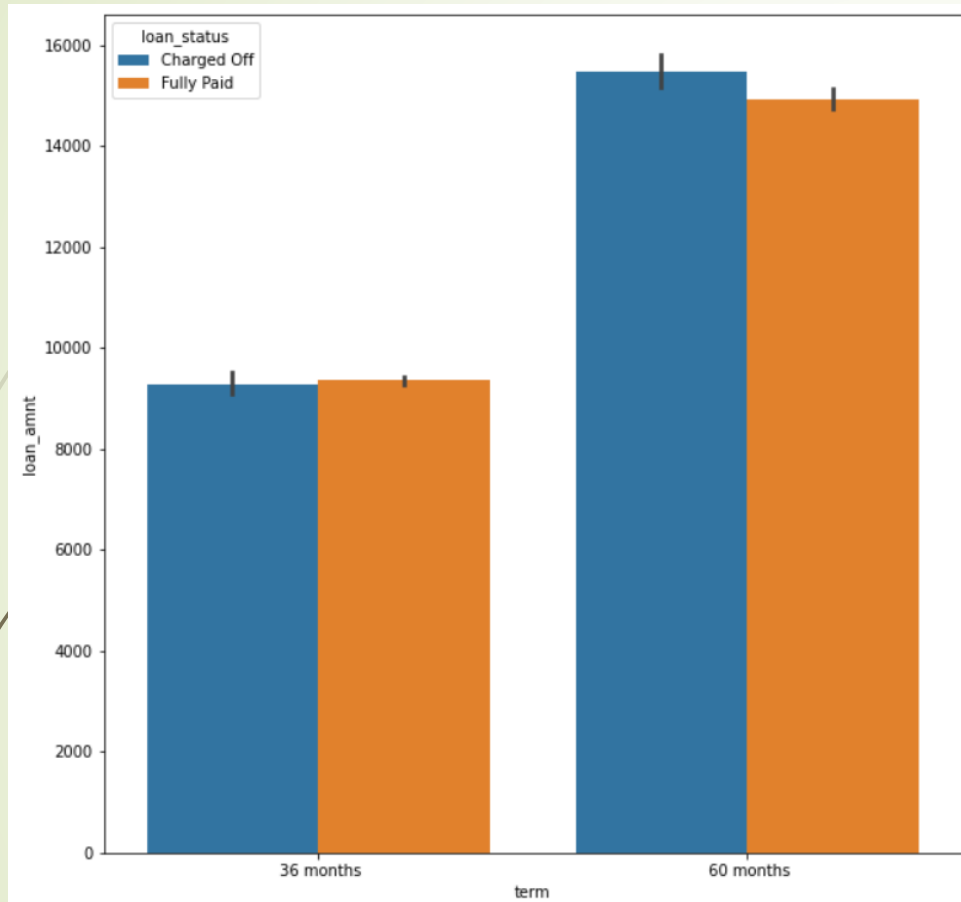
Employees with longer work experience are receiving higher loan\_amnt.

## 7. Verification\_status



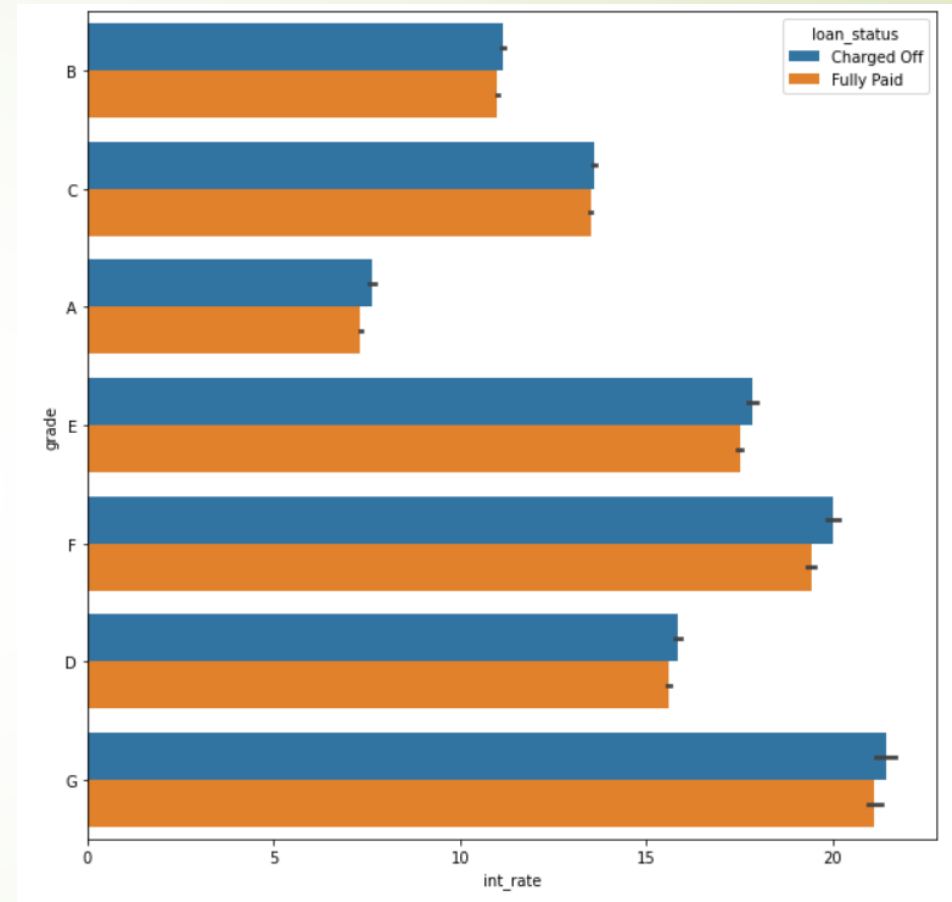
Verified loan applications tend to have higher loan amount, which might indicate that the firms are first verifying the loans with higher values.

## 6. loan\_amount vs term



Applicants who has fully paid the loan\_amount and defaulted have no significant difference in loan\_amounts.  
Applicants applying for long term has applied for more loan.

## Grade vs Interest Rate




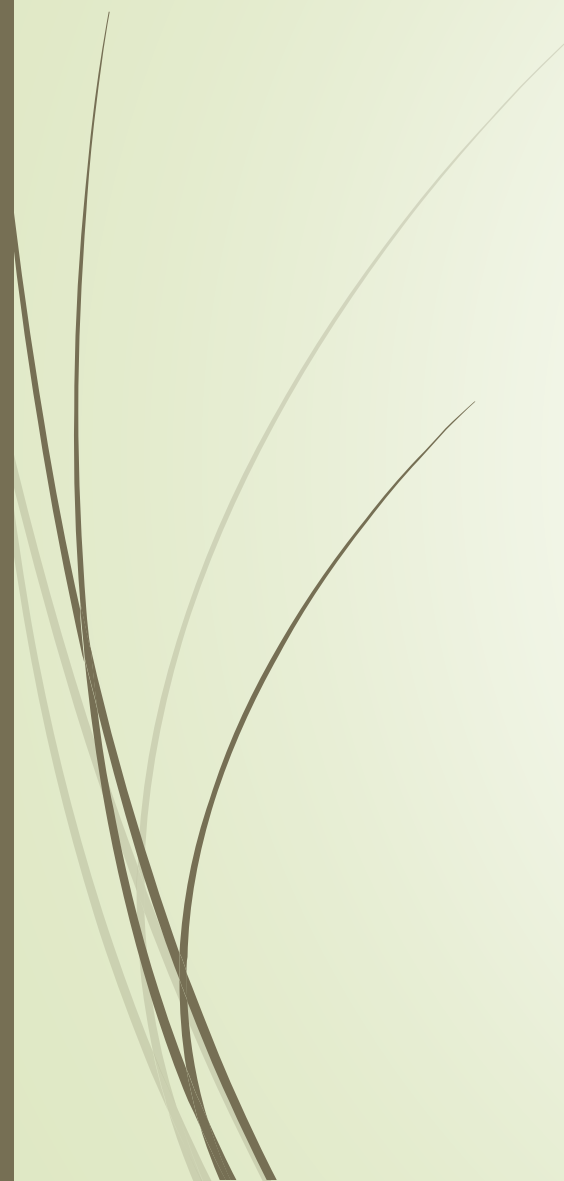
when we go from grade A to F, we can see there is increase in the number of interest rate.


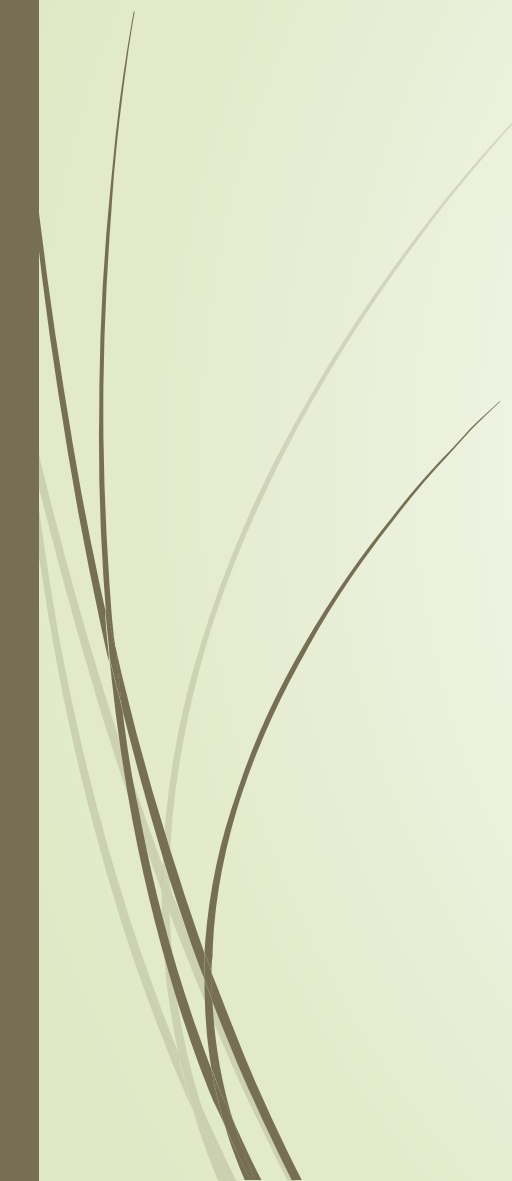


# Conclusions

The analysis with respect to the charged-off loans suggests the following. so that Lending Club can consider these factors before approving any loans.

- Applicants having house\_ownership as 'RENT'
- Applicants who use the loan to clear other debts
- Applicants who receive interest at the rate of 13-17%
- Applicants who have an income of range 31201 - 58402
- Applicants with employment length of 10
- When the funded amount by the investor is between 5000-10000
- Loan amount is between 5429 - 10357
- Dti is between 12-18

- 
- 
- When monthly installments are between 145-274
  - Term of 36 months
  - When the loan status is Not verified
  - When the purpose is 'debt\_consolidation'
  - Grade is 'B'
  - Total grade of 'B5' level.
  - Applicants taking loans for 'home improvement' and have an income of 60k -70k
  - Applicants whose home ownership is 'MORTGAGE and have an income of 60-70k
  - 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 a small business and the loan amount is greater than 14k
  - Applicants whose home ownership is 'MORTGAGE and have a loan of 14-16k
  - When a grade is F and the loan amount is between 15k-20k
  - When employment length is 10yrs and loan amount is 12k-14k
  - When the loan is verified and the loan amount is above 16k
  - For grade G and interest rate above 20%