

# Lending Club Case Study

Ali Abdullrahaman

Krishna Chaitanya

# Introduction

- ▶ This case study will give us an idea of how real business problems are solved using EDA. we will develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

# Business Understanding

- The objective is to help with taken a decision whenever they receive a loan application whether to reject or approve based on certain variables depending on the analysis.

## **Dataset Details:**

The data 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.

# Cleaning the Data and preparation process:

1-Importing the Data.

2-Removing columns which are more than 90% empty.

3-Removing single-valued columns that can't contribute to our analysis.

4-Removing some more columns based on the type of information they provide:

- A- **id**, and **member\_id** are just conventional columns, **zip\_code**, **addr\_state**, **emp\_title**, **url**, **desc**, and **purpose** can also be removed as they don't contribute to finding loan defaults.
- B- **mths\_since\_last\_delinq** is less useful than **delinq\_2yrs**, but **delinq\_2yrs** is highly skewed and more than 38k values are between 0-3 times, by which we can't judge the default tendency of the borrower.
- C- **recoveries** and **collection\_recovery\_fee** are duplicates, so one can be removed.
- D- **out\_prncp**, **chargeoff\_within\_12\_mths**, **total\_rec\_late\_fee**, **last\_pymnt\_amnt**, **revol\_bal** are uncorrelated to the case statement and can be removed.

5-Removing the rows with more the 5 empty values among 31 columns.

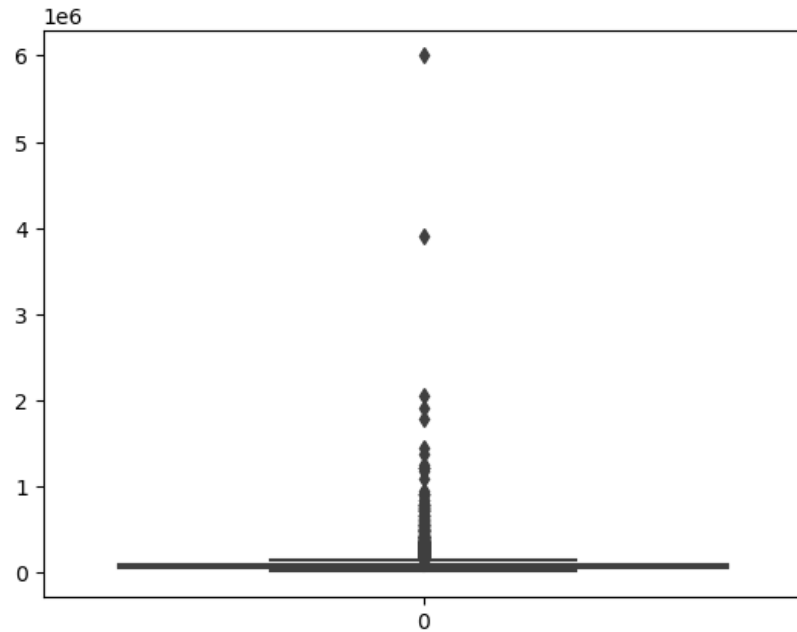
6-Removing/Fixing null values

7-Correcting data types and deriving new columns

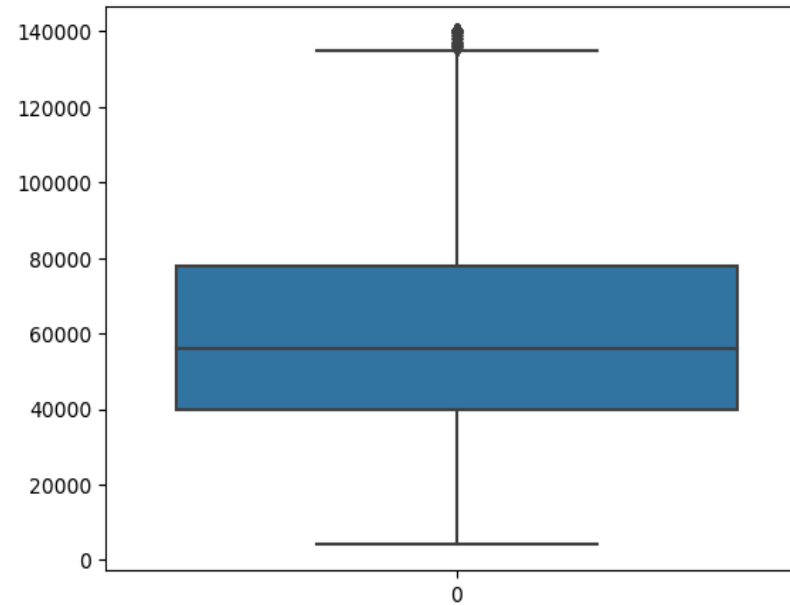
8-Removing outliers.

# Cleaning the Data and preparation process:

Example for Removing outliers for the **annual\_inc** column



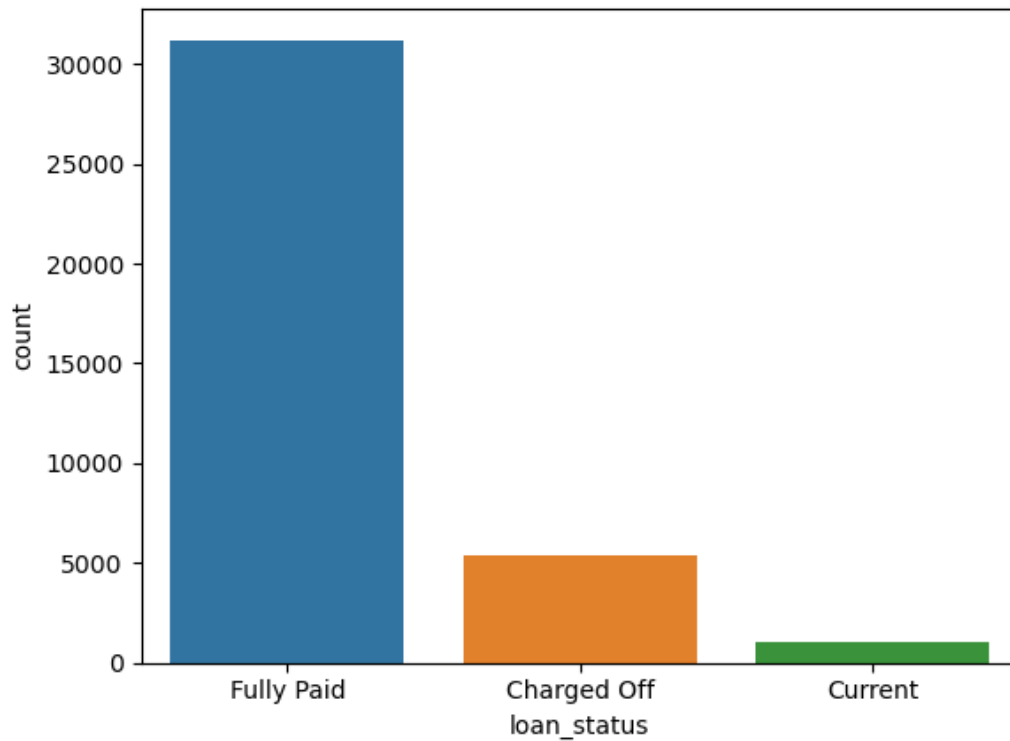
- Annual Inc.: Before the outliers removal.



- Annual Inc.: After the outliers removal.

# Analysis and Visualisation with Univariate:

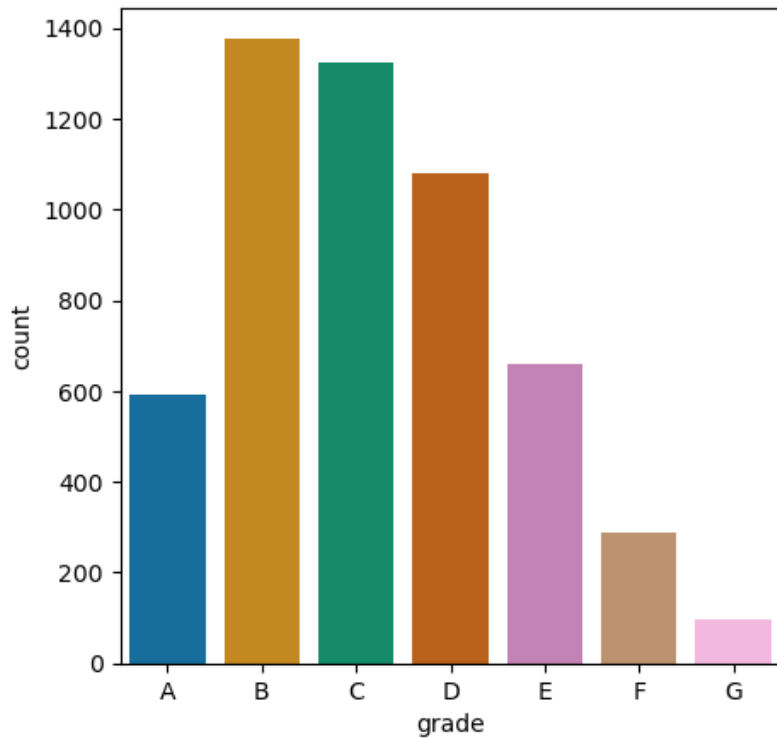
- After the cleaning we start the analysis, we are aiming to find the potential defaulter let's consider only **Charged\_off** for now.



- **Loan Status:** The charged Off loans are much smaller compared to the total count.

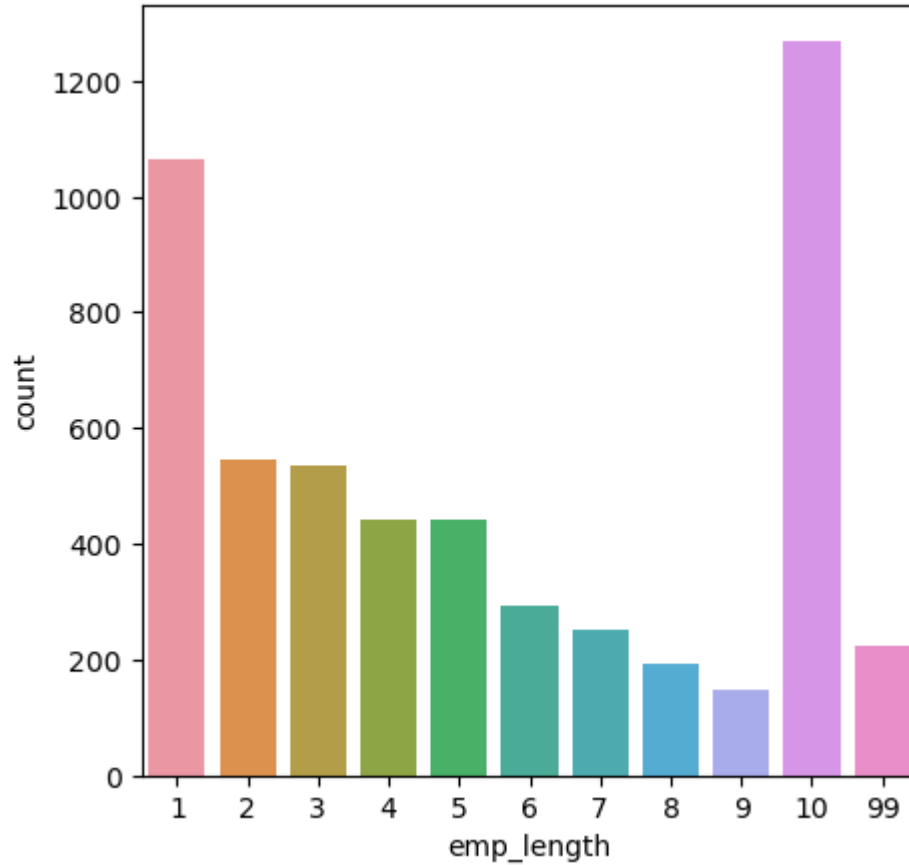
# Analysis and Visualisation with Univariate:

- Analysis of **Charged\_off** wrt **grade**.



# Analysis and Visualisation with Univariate:

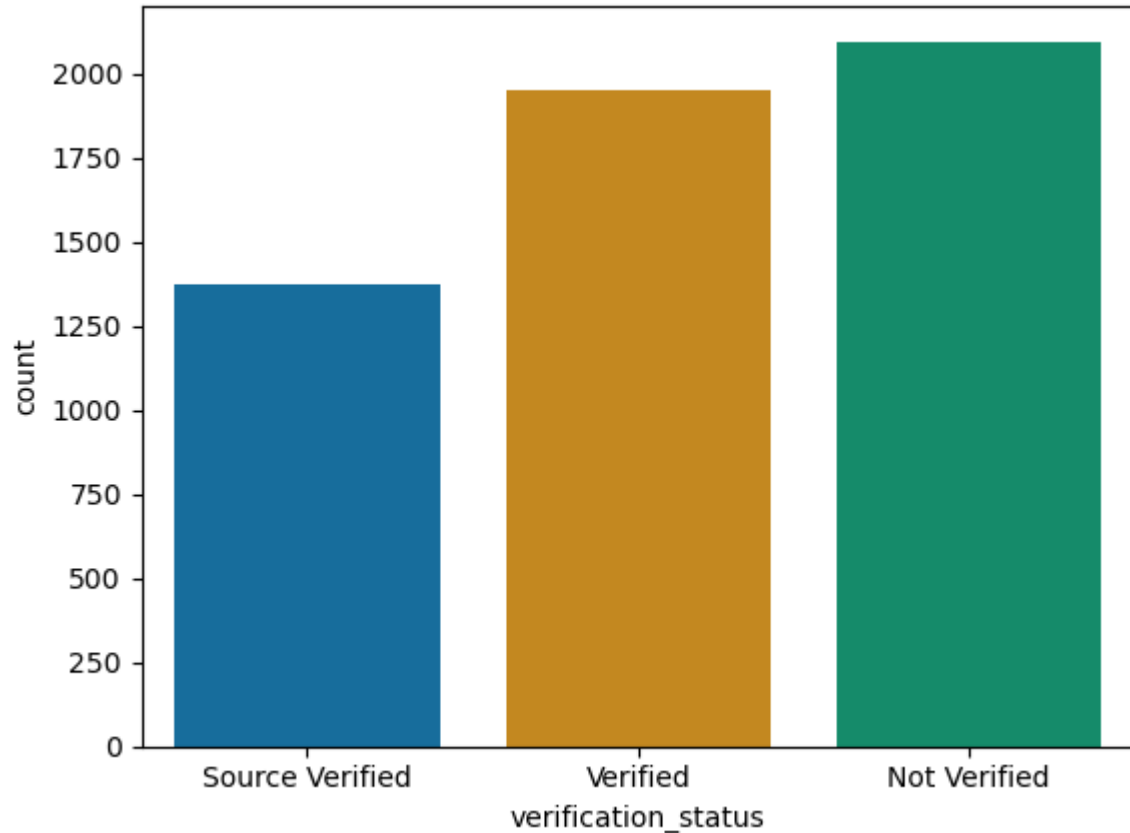
- Analysis of **Charged\_off** wrt **emp\_length**.





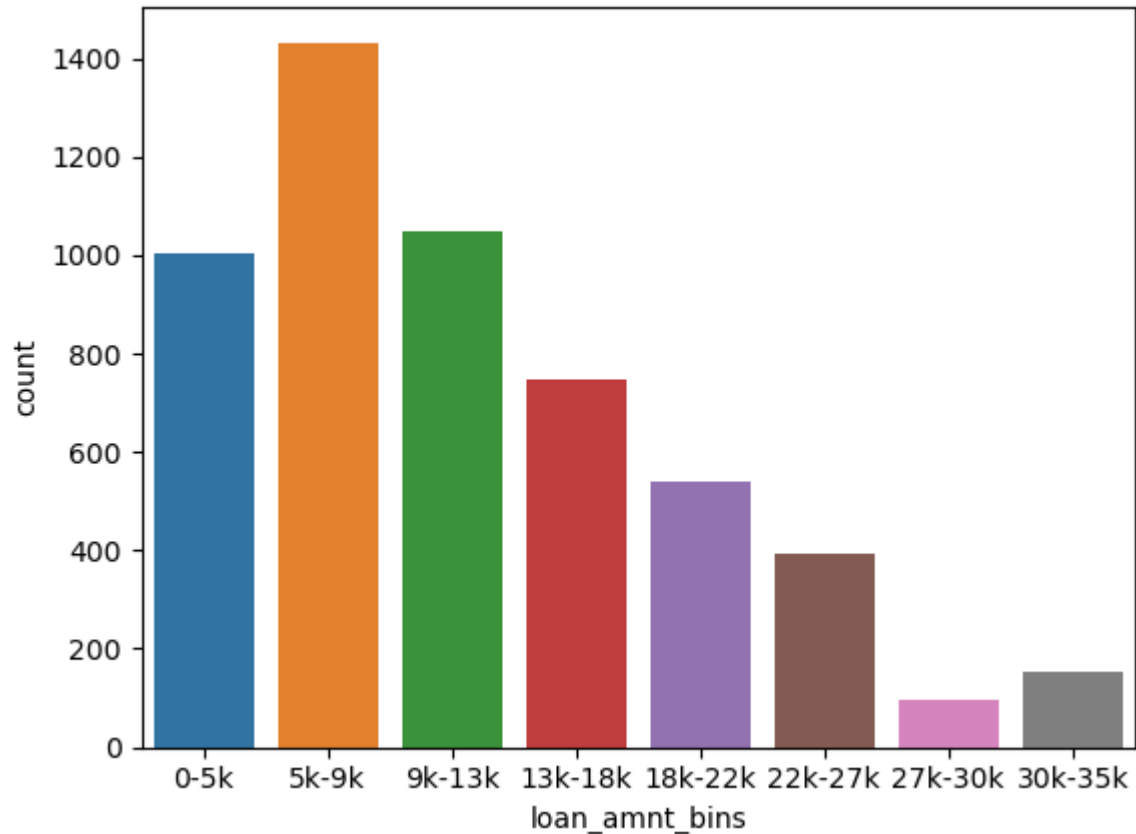
# Analysis and Visualisation with Univariate:

- Analysis of **Charged\_off** wrt **verification\_status**.



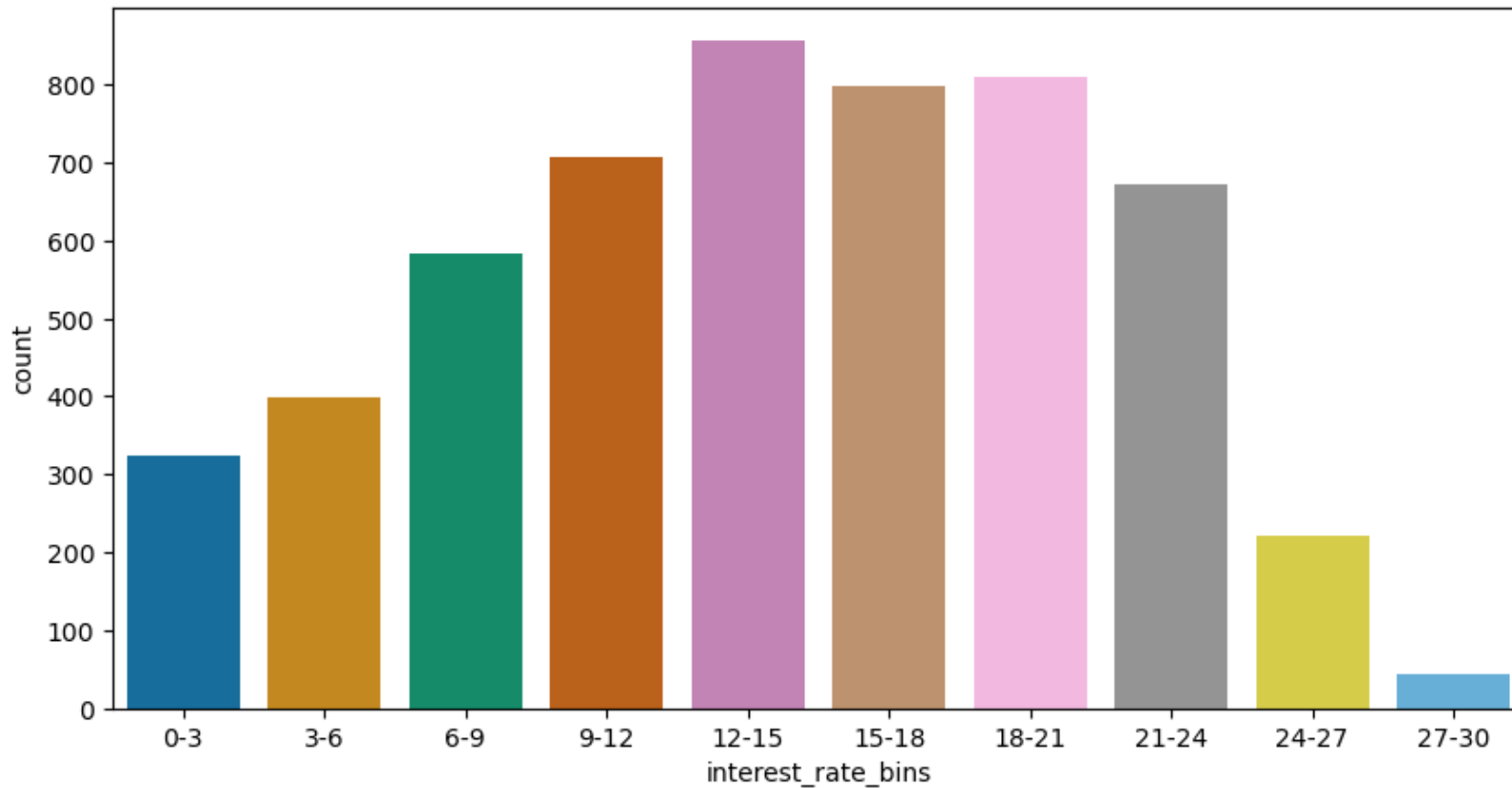
# Analysis and Visualisation with Univariate:

- Analysis of **Charged\_off** wrt **loan\_amnt**.



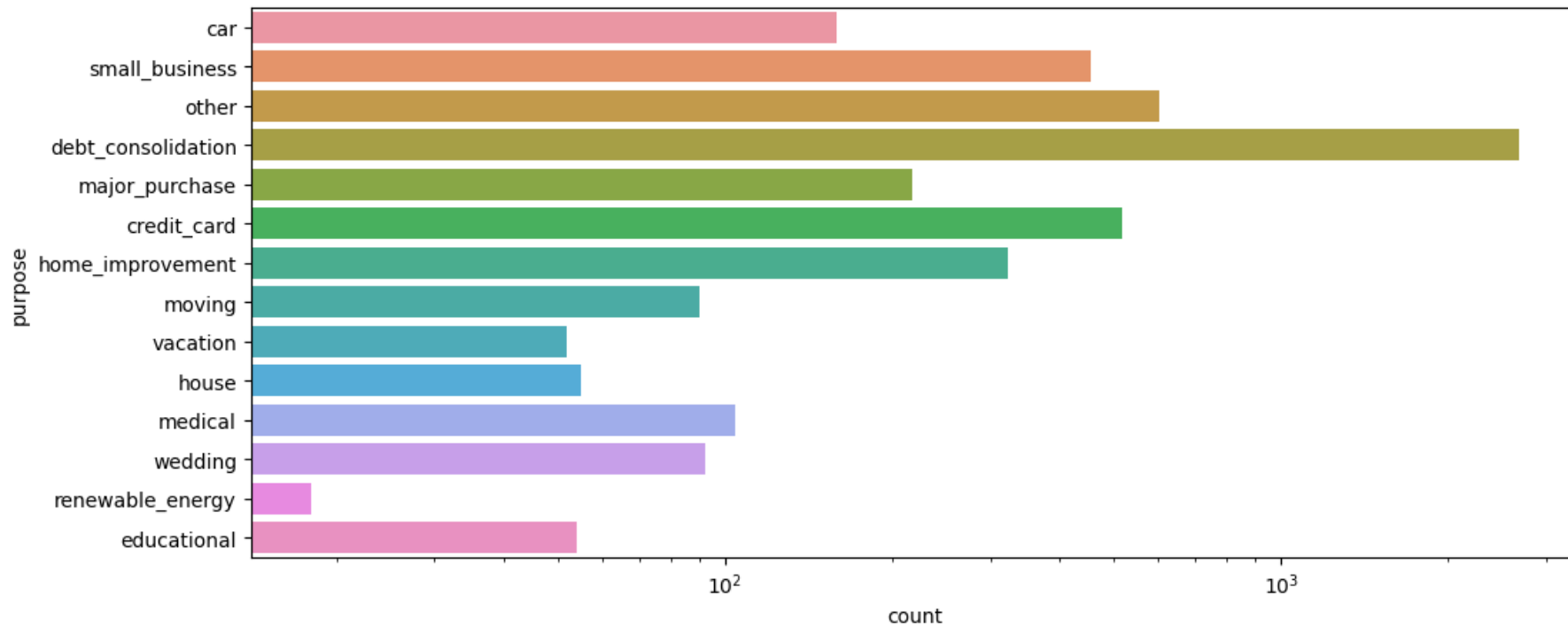
# Analysis and Visualisation with Univariate:

- Analysis of **Charged\_off** wrt **interest\_rate**.



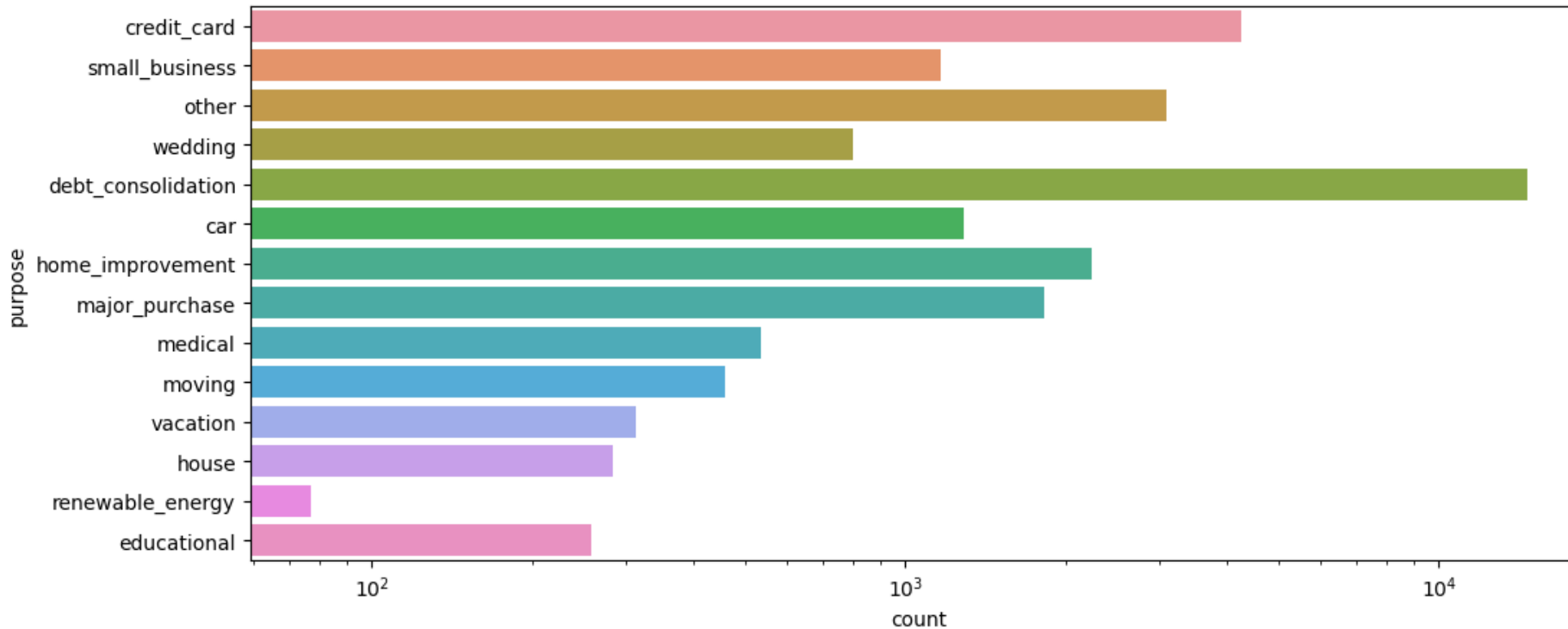
# Analysis and Visualisation with Univariate:

- Analysis of **Charged\_off** wrt **purpose**.



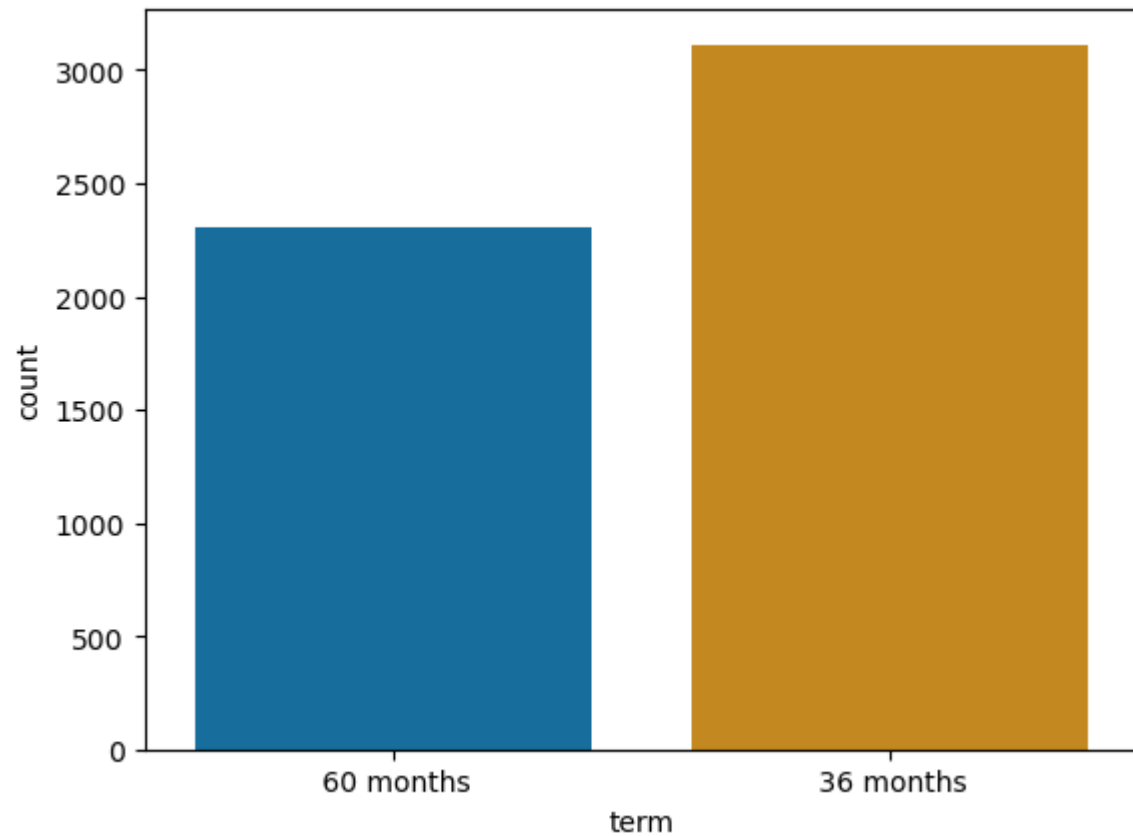
# Analysis and Visualisation with Univariate:

- Analysis of Fully Paid wrt purpose.



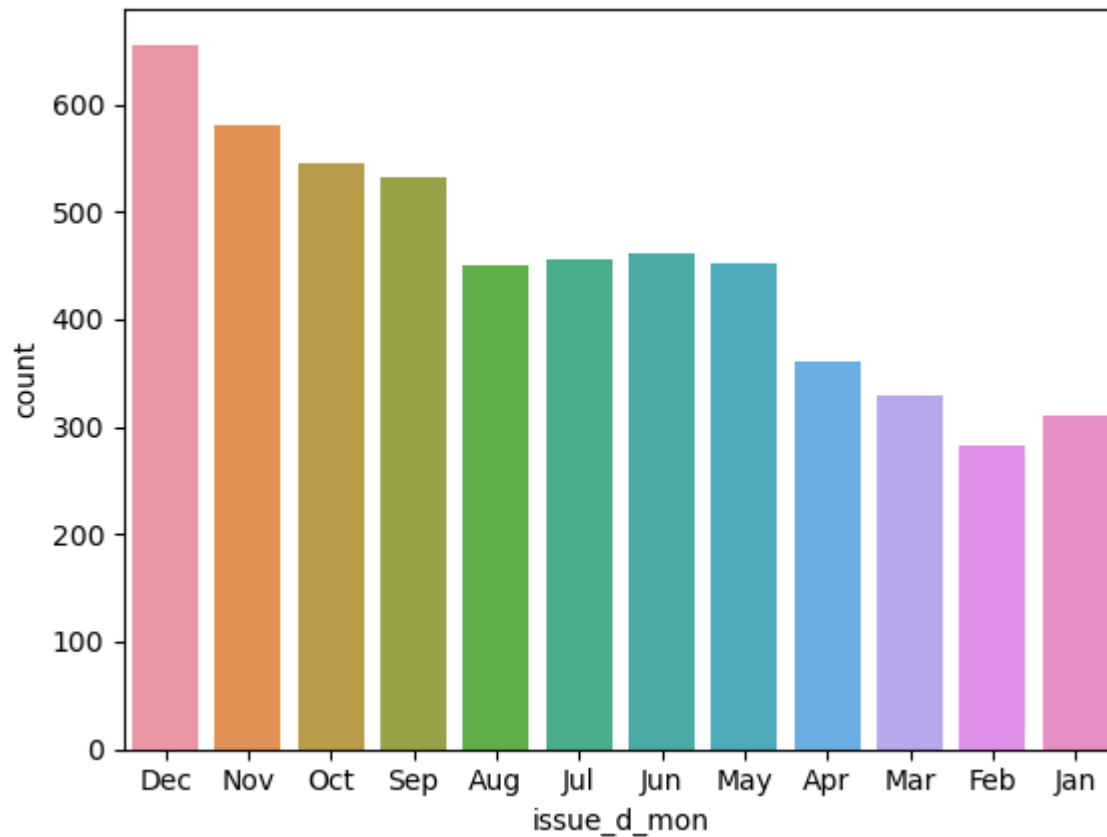
# Analysis and Visualisation with Univariate:

- Analysis of **Charged\_off** wrt **term**.



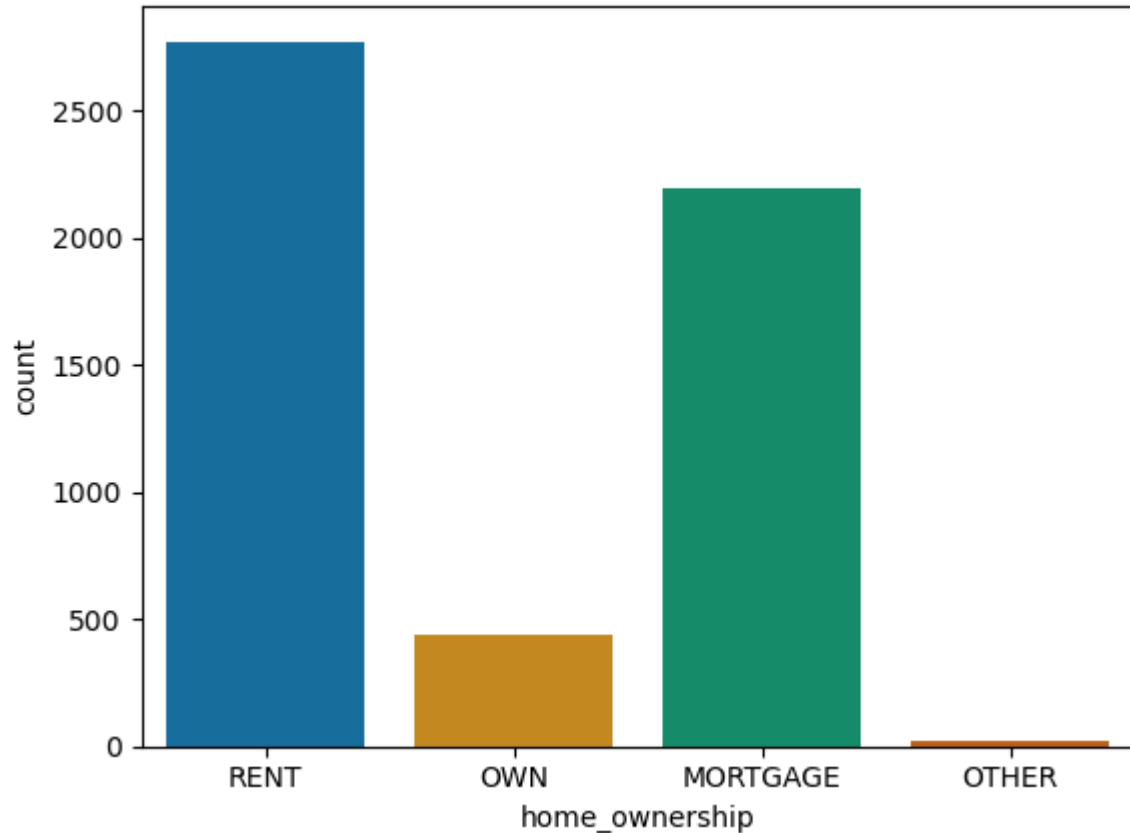
# Analysis and Visualisation with Univariate:

- Analysis of **Charged Off** wrt `issue_d_mon`.



# Analysis and Visualisation with Univariate:

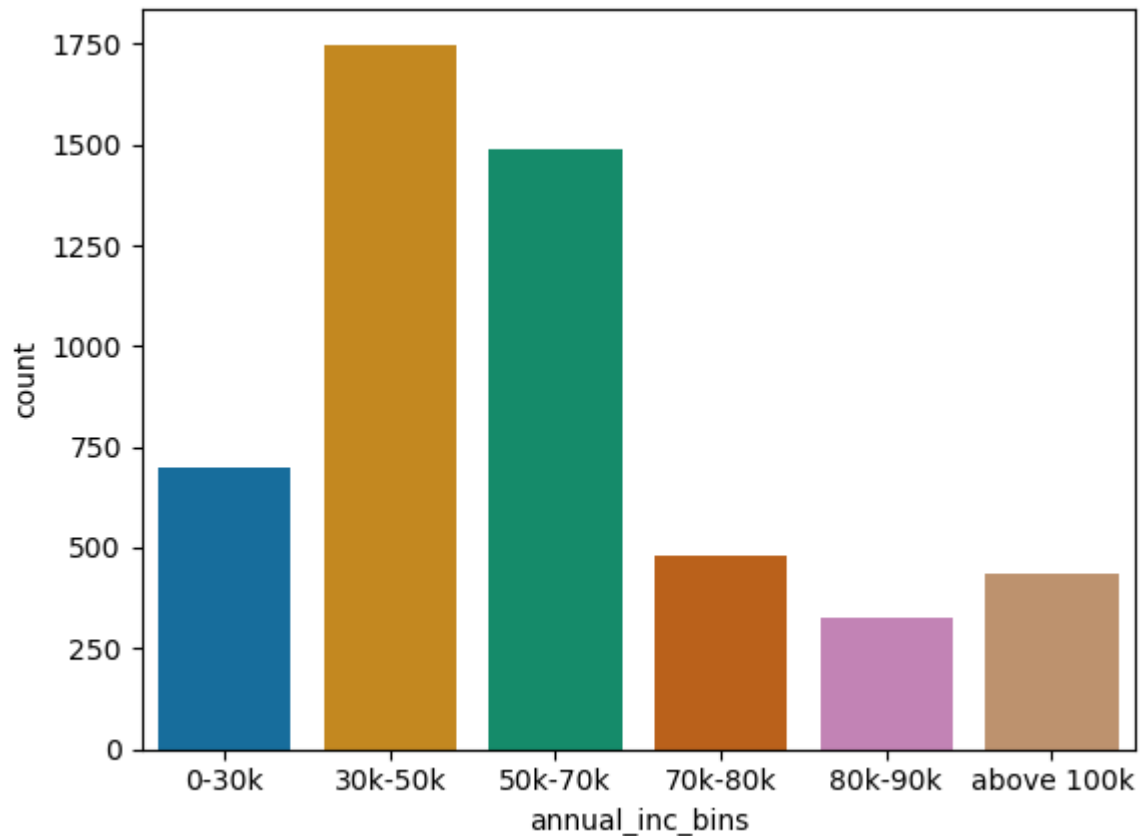
- Analysis of **Charged Off** wrt **home\_ownership**.





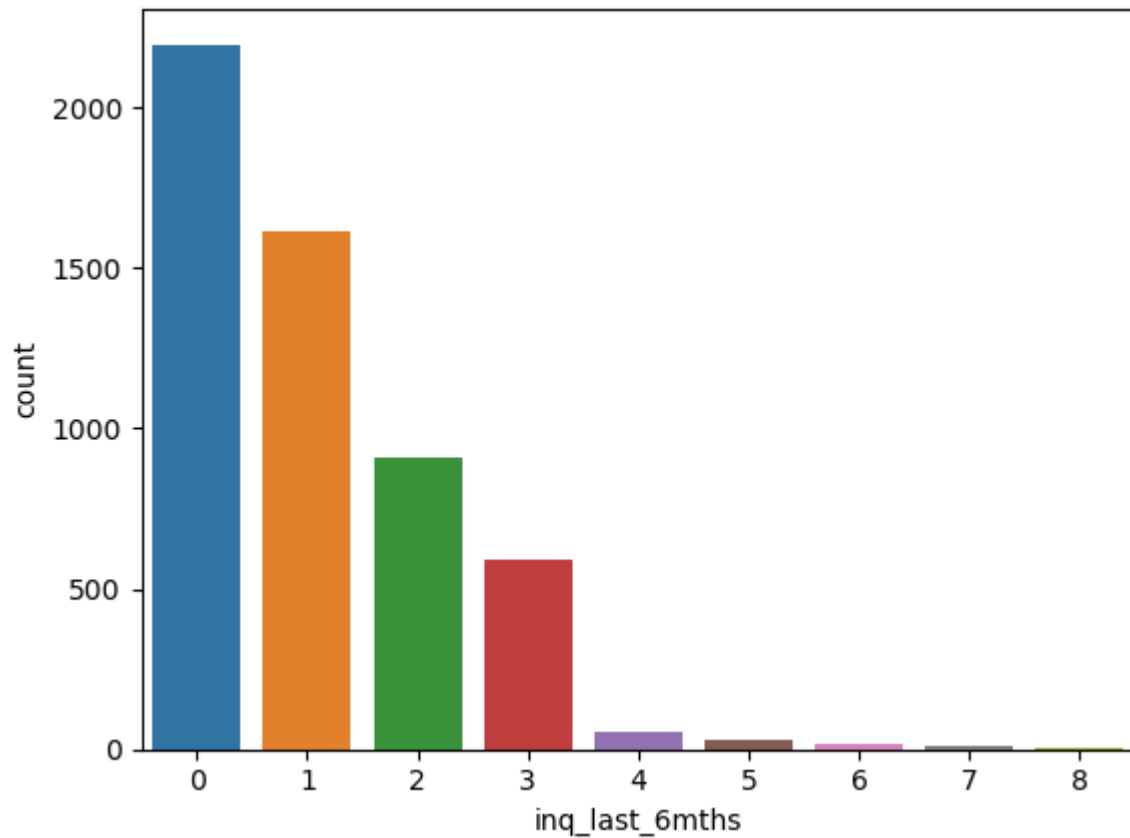
# Analysis and Visualisation with Univariate:

- Analysis of **Charged Off** wrt **annual\_inc**.



# Analysis and Visualisation with Univariate:

- Analysis of **Charged Off** wrt **inq\_last\_6mths**.



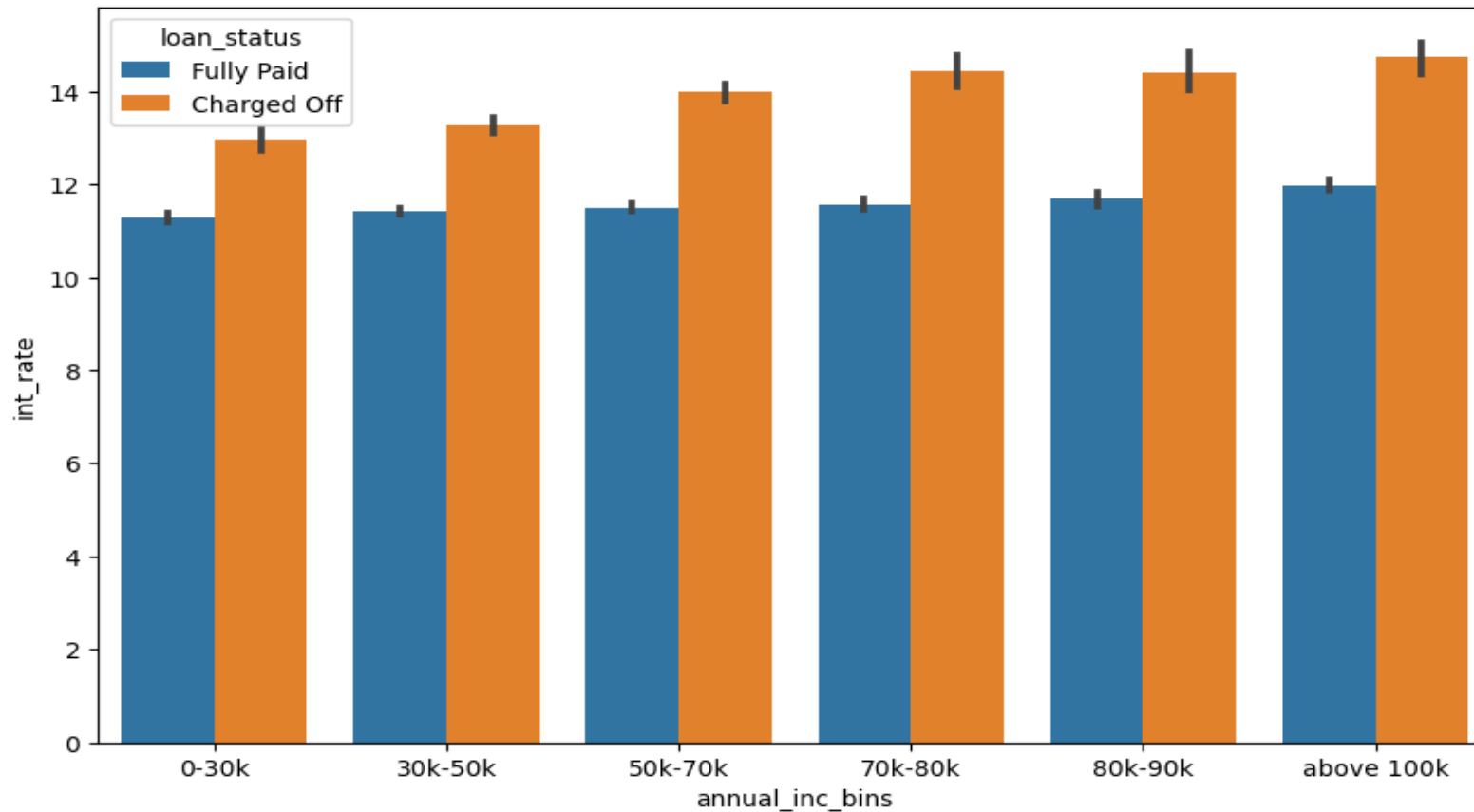
# Conclusions of the Analysis with Univariate:

- ▶ **The defaulting potential is more under the following cases:**
  - When the annual income of the borrower is in the range of 30k to 70k.
  - When the borrower is on either RENT or having a MORTGAGE. Also, people who own a house have a very low chance of defaults.
  - The term of 36 months has more chance of delinquency.
  - When the number of inquiries in the last 6 months is 0.
  - When the interest rates are from 12% to 20% the chances are more.
  - Grade is B or C.
  - When the loan amount is 5k-9k, delinquency is high.
  - When the loan status is Not verified.
  - When the reason is to clear off other debts/loans.
  - When the borrower's employment length is 10+.

For some reason, there is a steady increase in default possibility wrt the months across all 4 years (Jan<Feb<....<Dec).

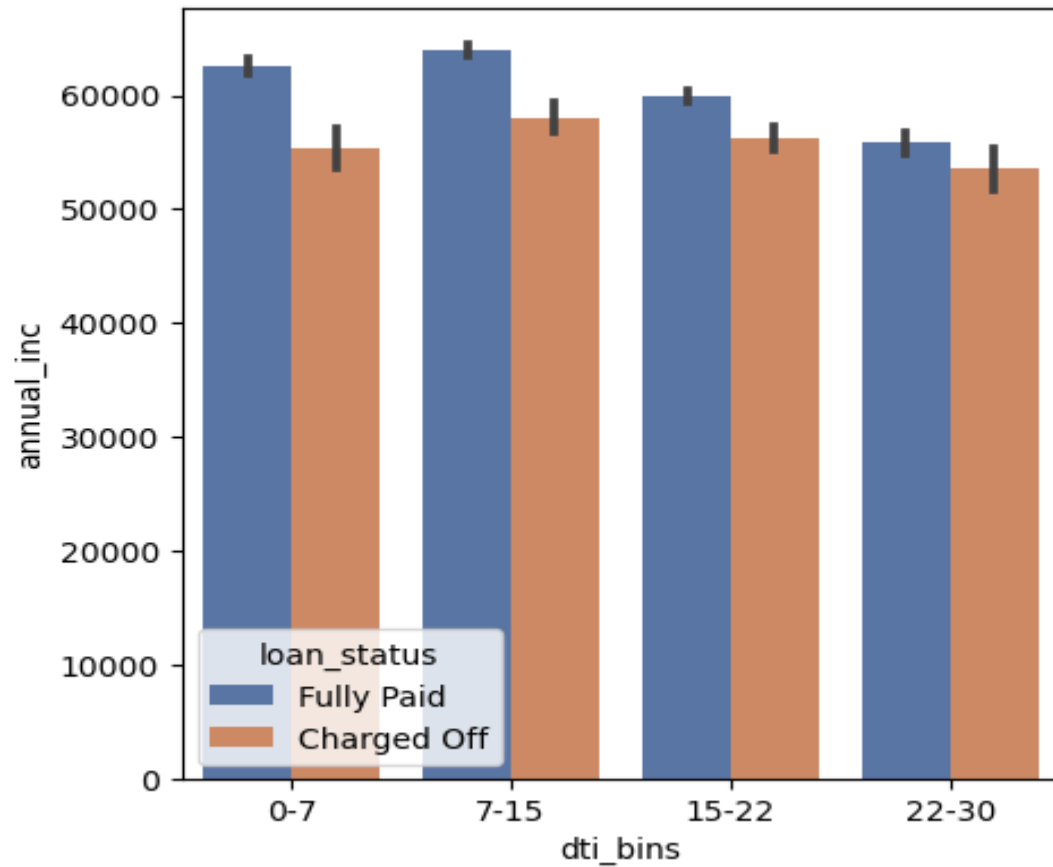
# Analysis and Visualisation with Bivariate:

- Analysis of `loan_status` with Fully Paid and Charged Off with discard the Current.



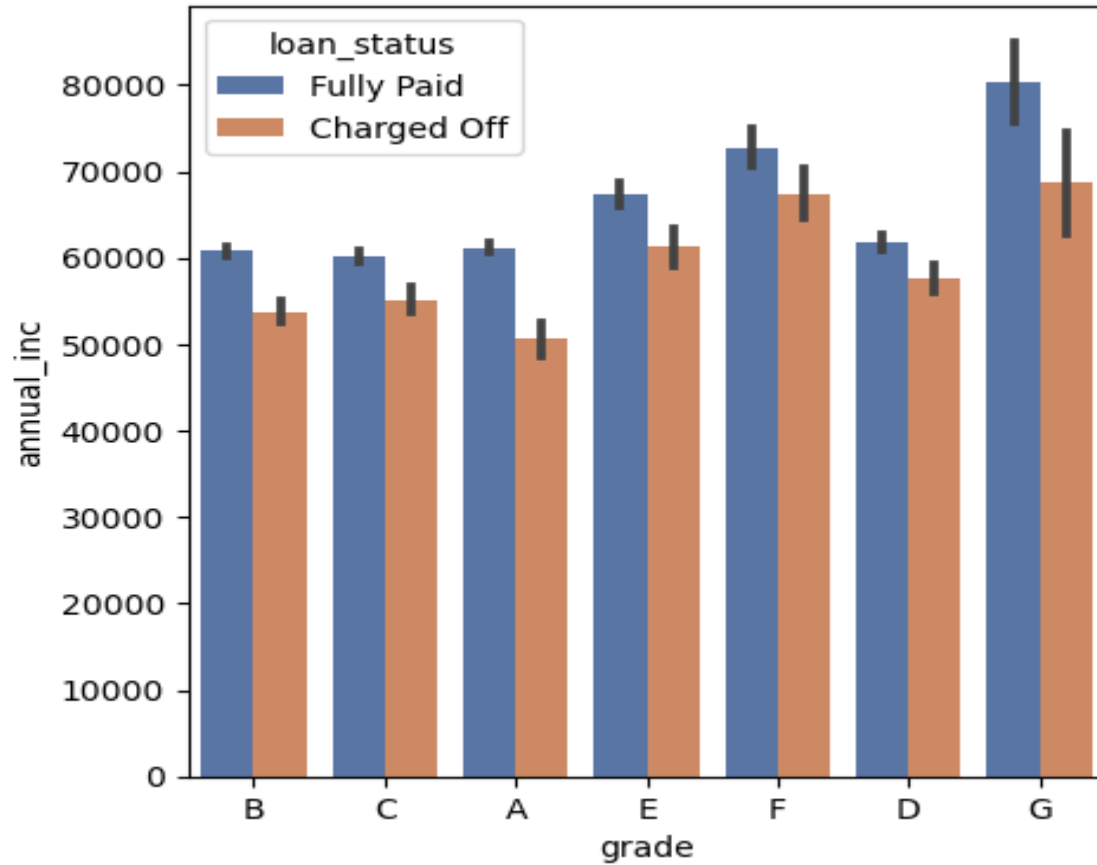
# Analysis and Visualisation with Bivariate:

- Analysis of `loan_status` with `annual_inc` and `DTI`.



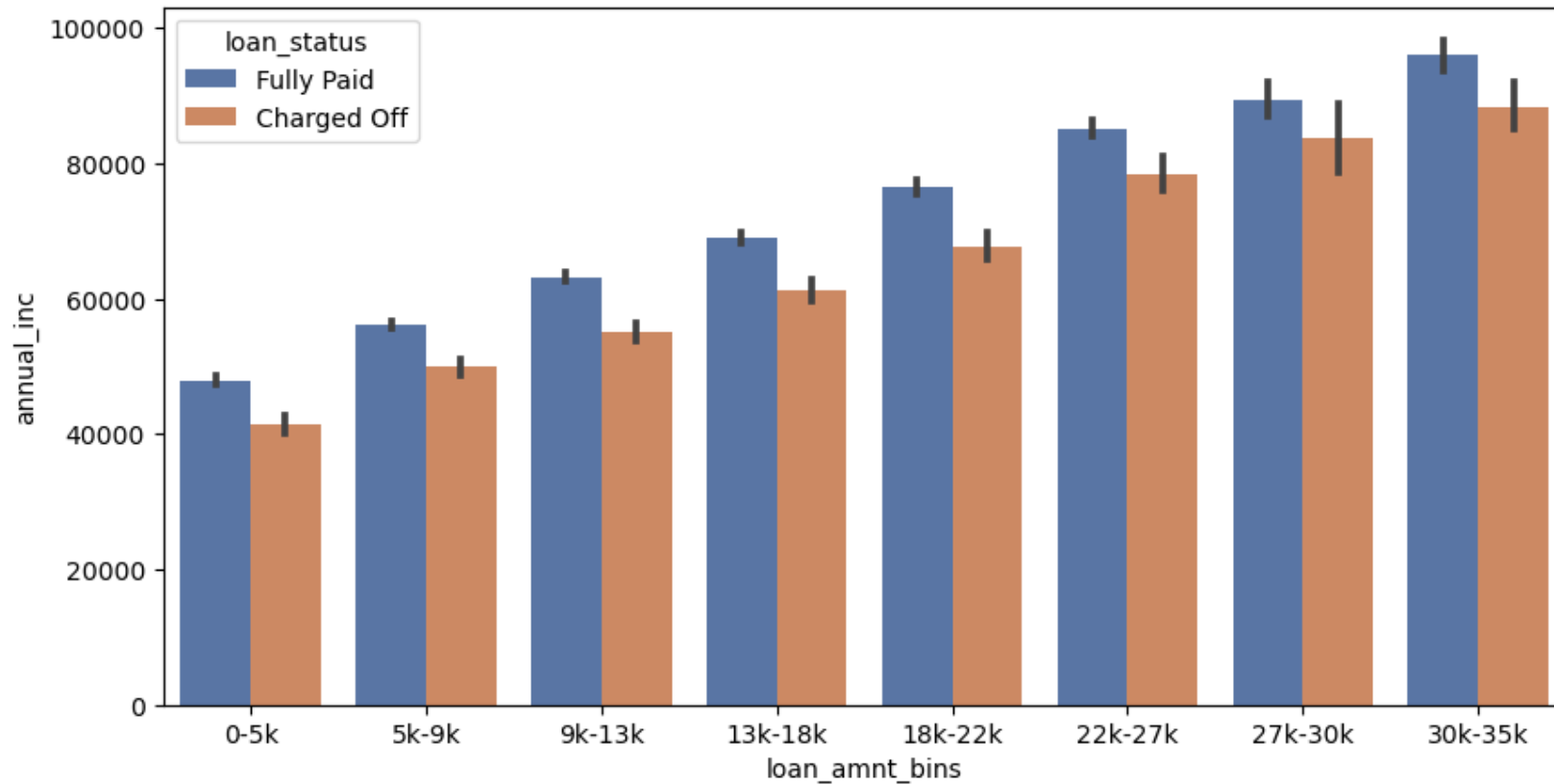
# Analysis and Visualisation with Bivariate:

- Analysis of `loan_status` with `annual_inc` and `grade`.



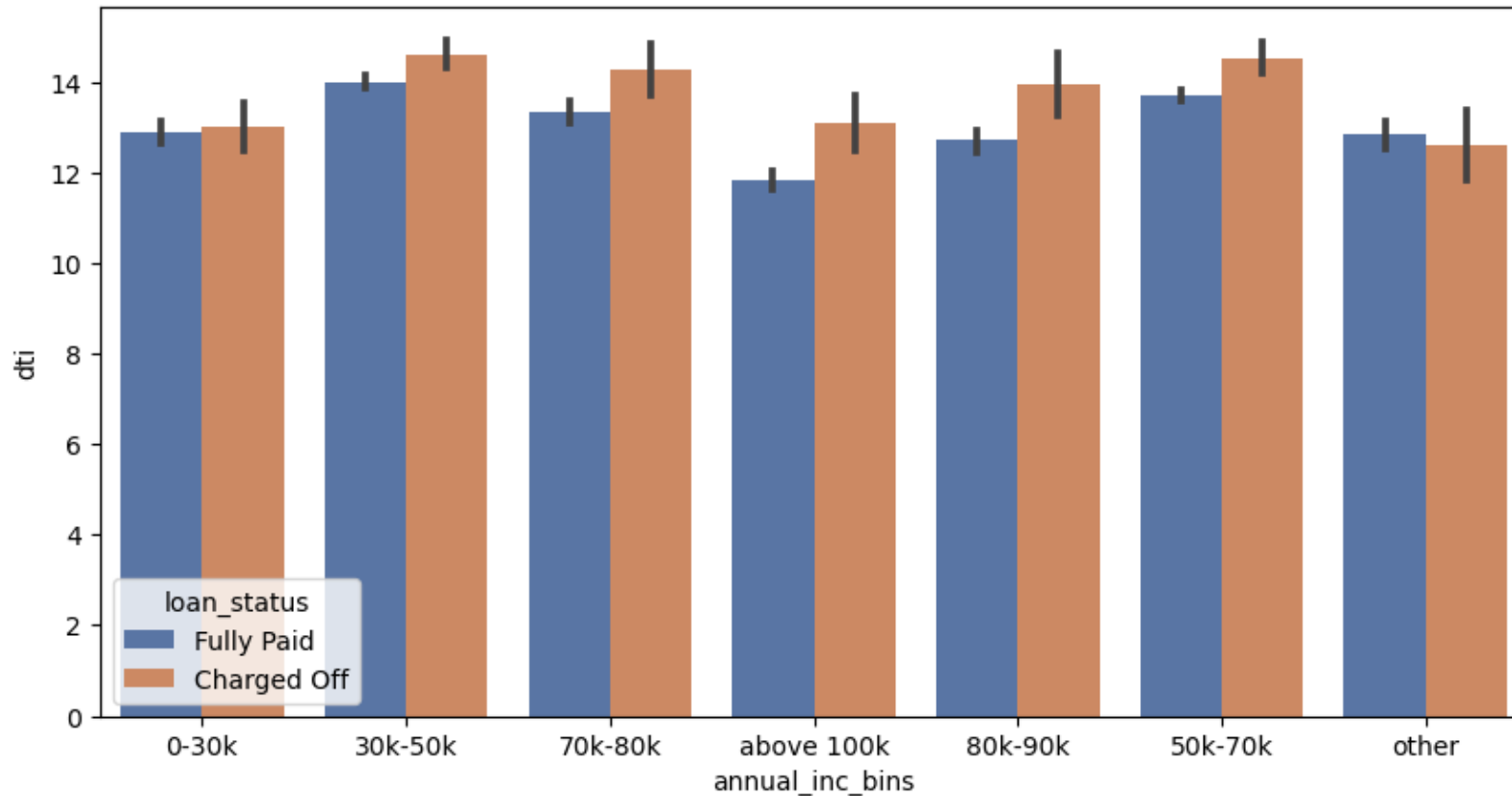
# Analysis and Visualisation with Bivariate:

- Analysis of `loan_status` with `annual_inc` and `loan_amnt_bins`.



# Analysis and Visualisation with Bivariate:

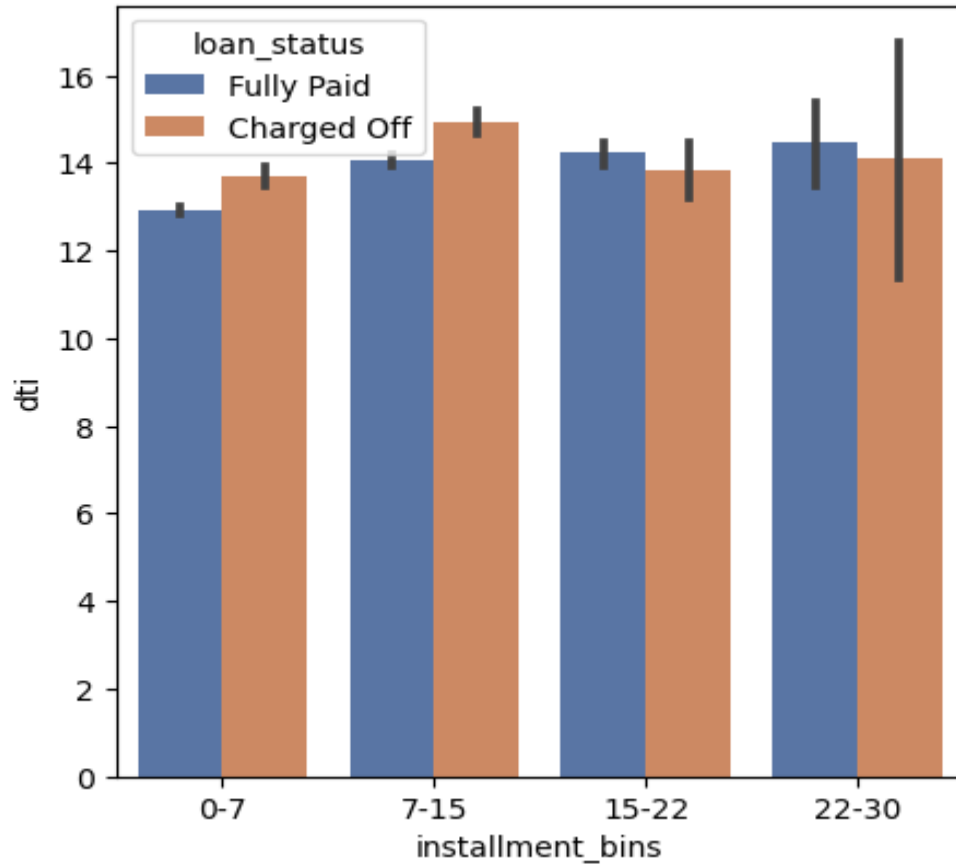
- Analysis of `loan_status` with `DTI` and `annual_inc`.





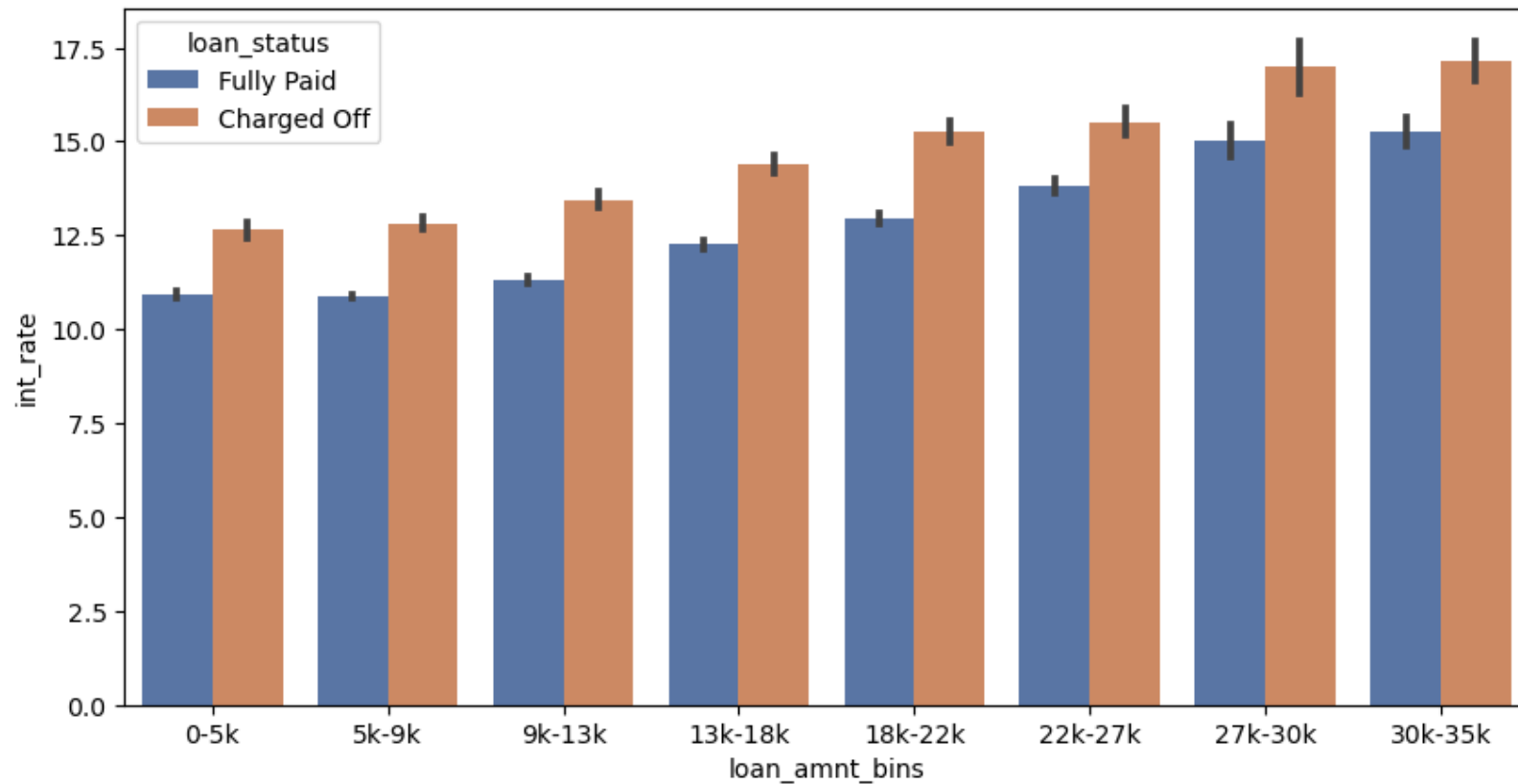
# Analysis and Visualisation with Bivariate:

- Analysis of loan\_status with DTI and installment\_bins.



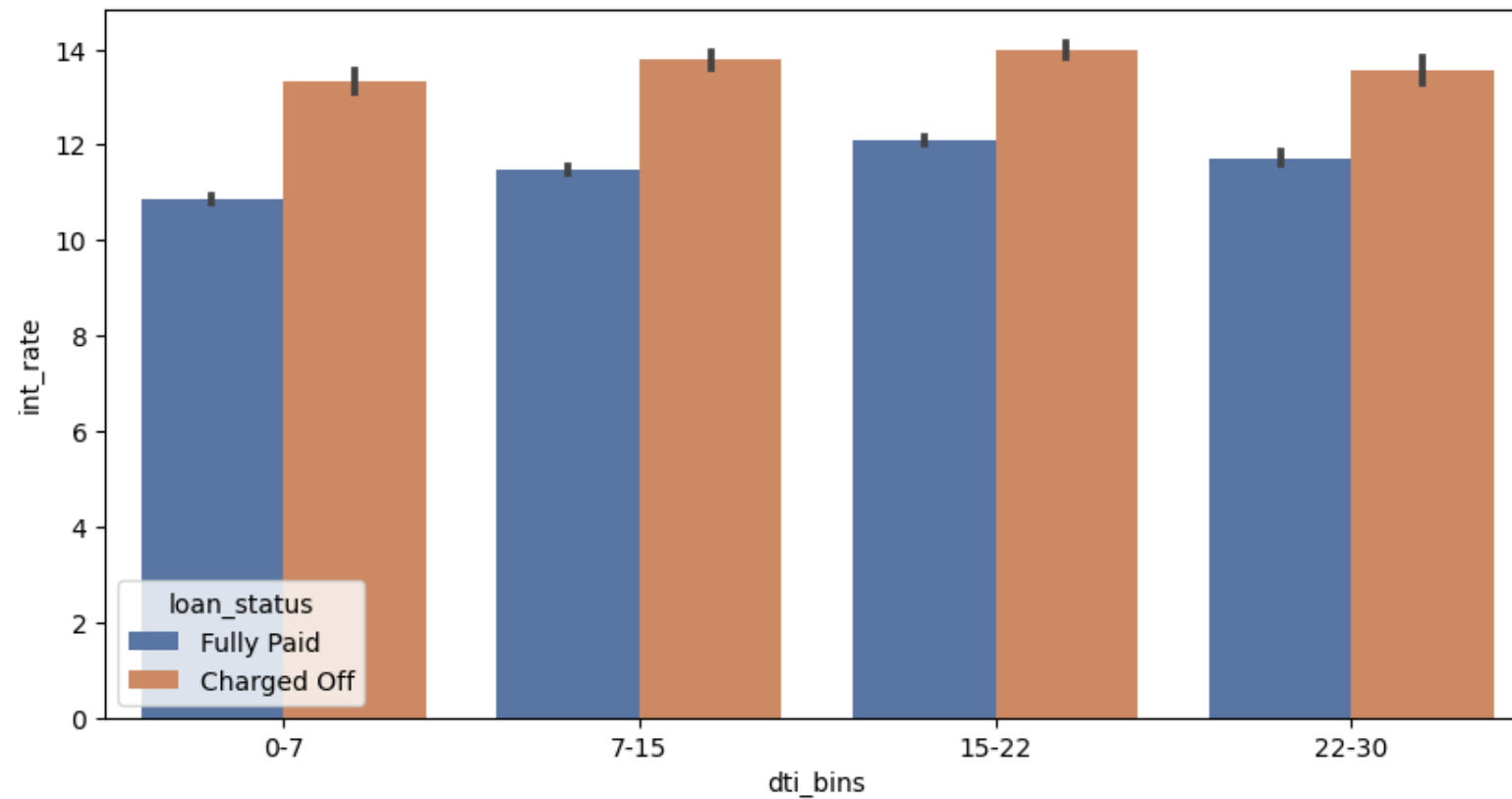
# Analysis and Visualisation with Bivariate:

- Analysis of loan\_status with int\_rate and loan\_amnt.



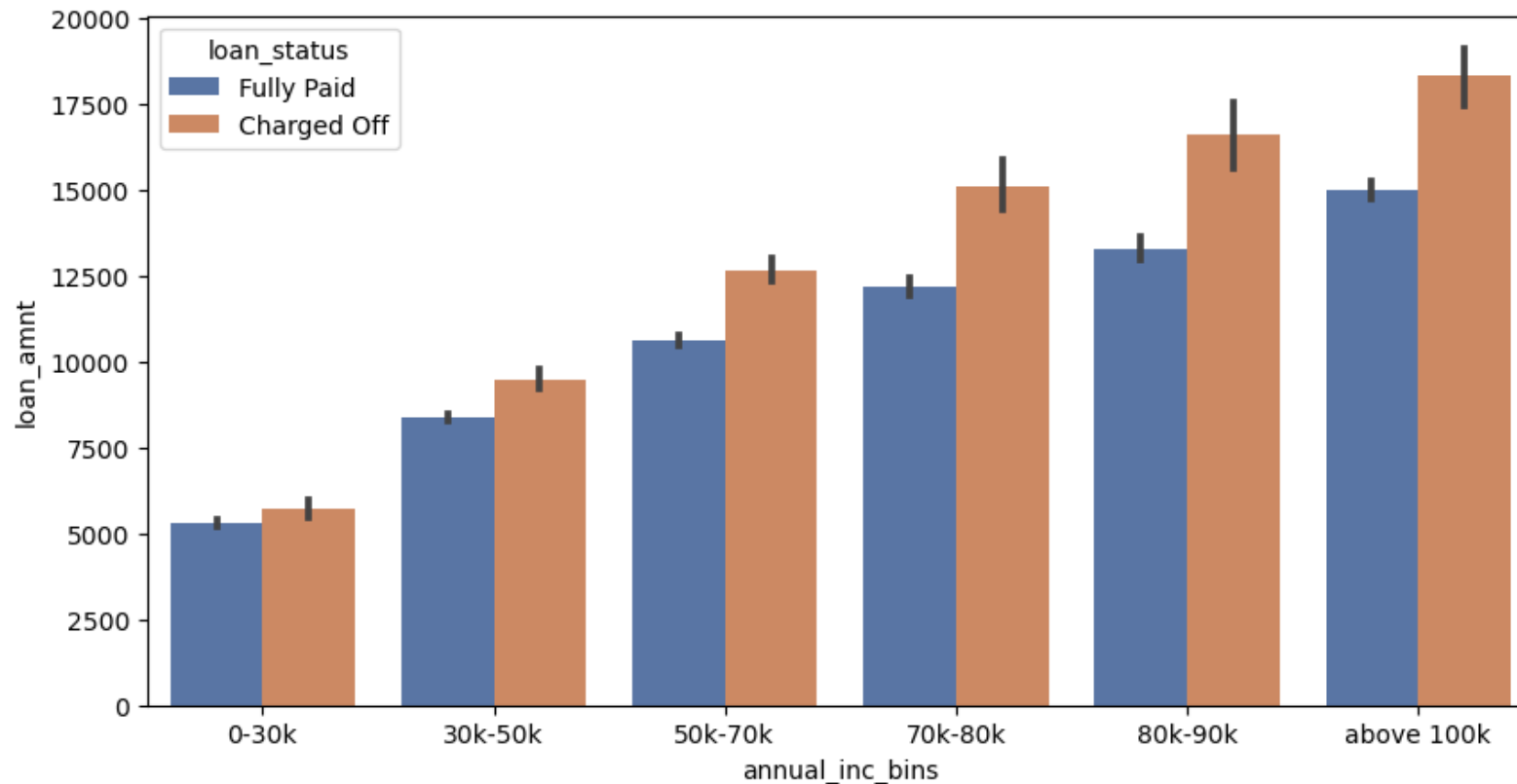
# Analysis and Visualisation with Bivariate:

- Analysis of `loan_status` with `int_rate` and `DTI`.



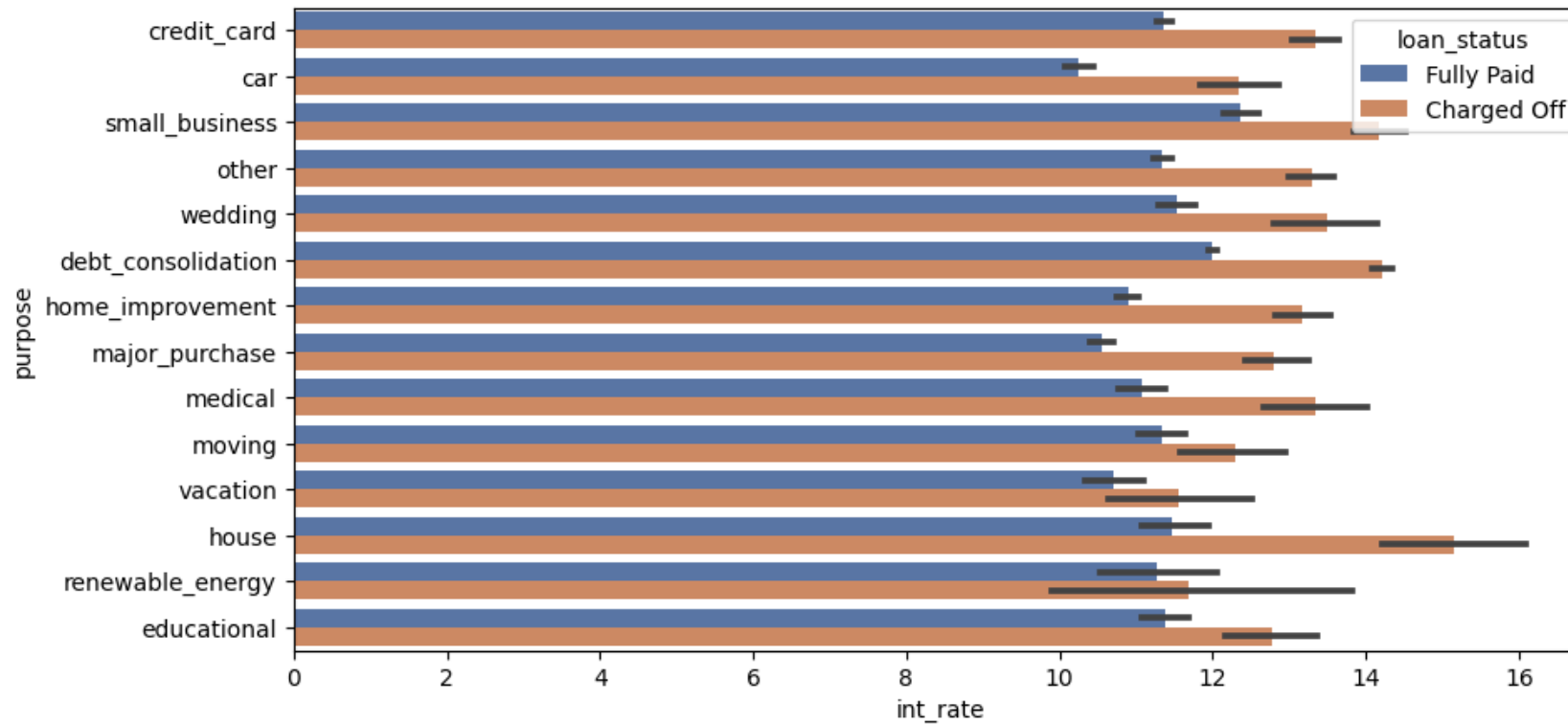
# Analysis and Visualisation with Bivariate:

- Analysis of `loan_status` with `loan_amnt` and `annual_inc`.



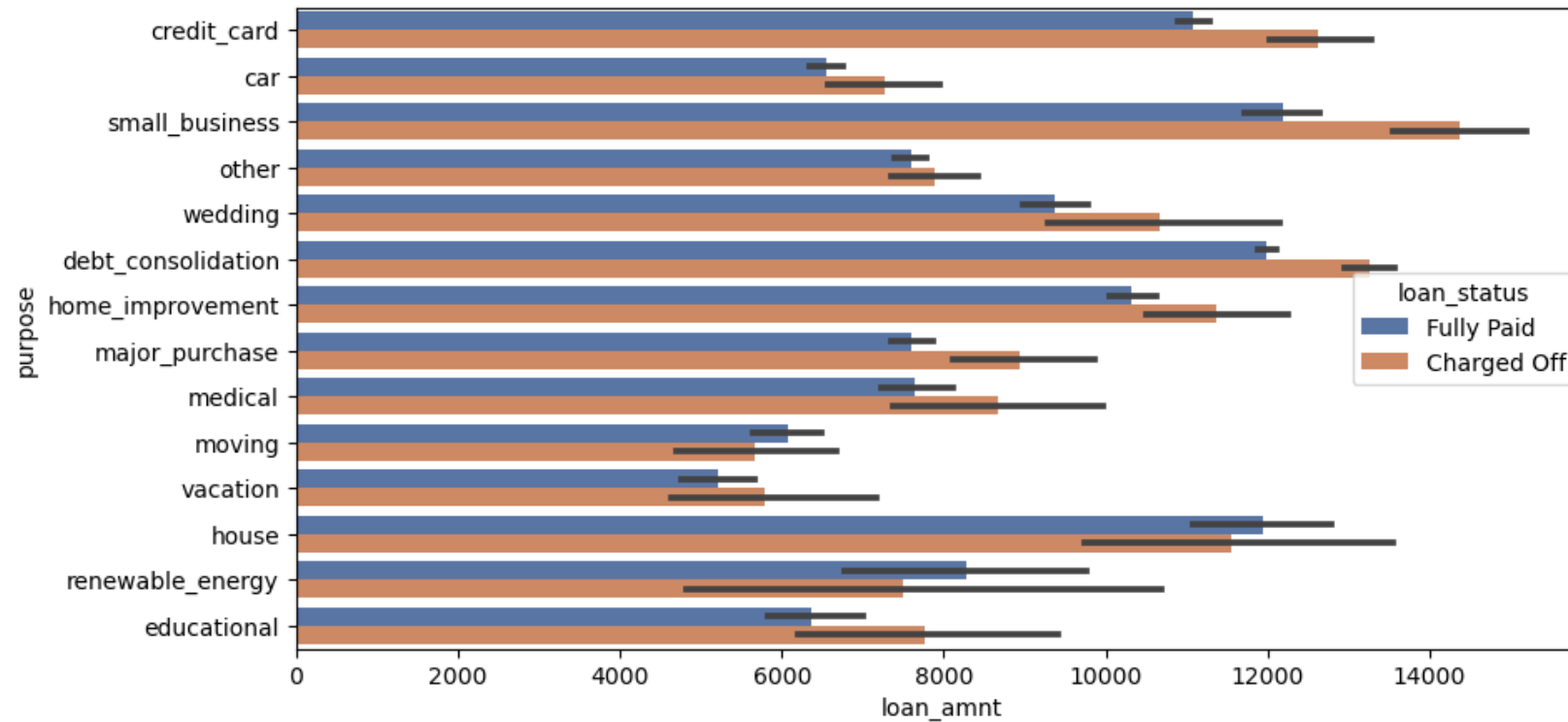
# Analysis and Visualisation with Bivariate:

- Analysis of loan\_status with purpose and int\_rate.



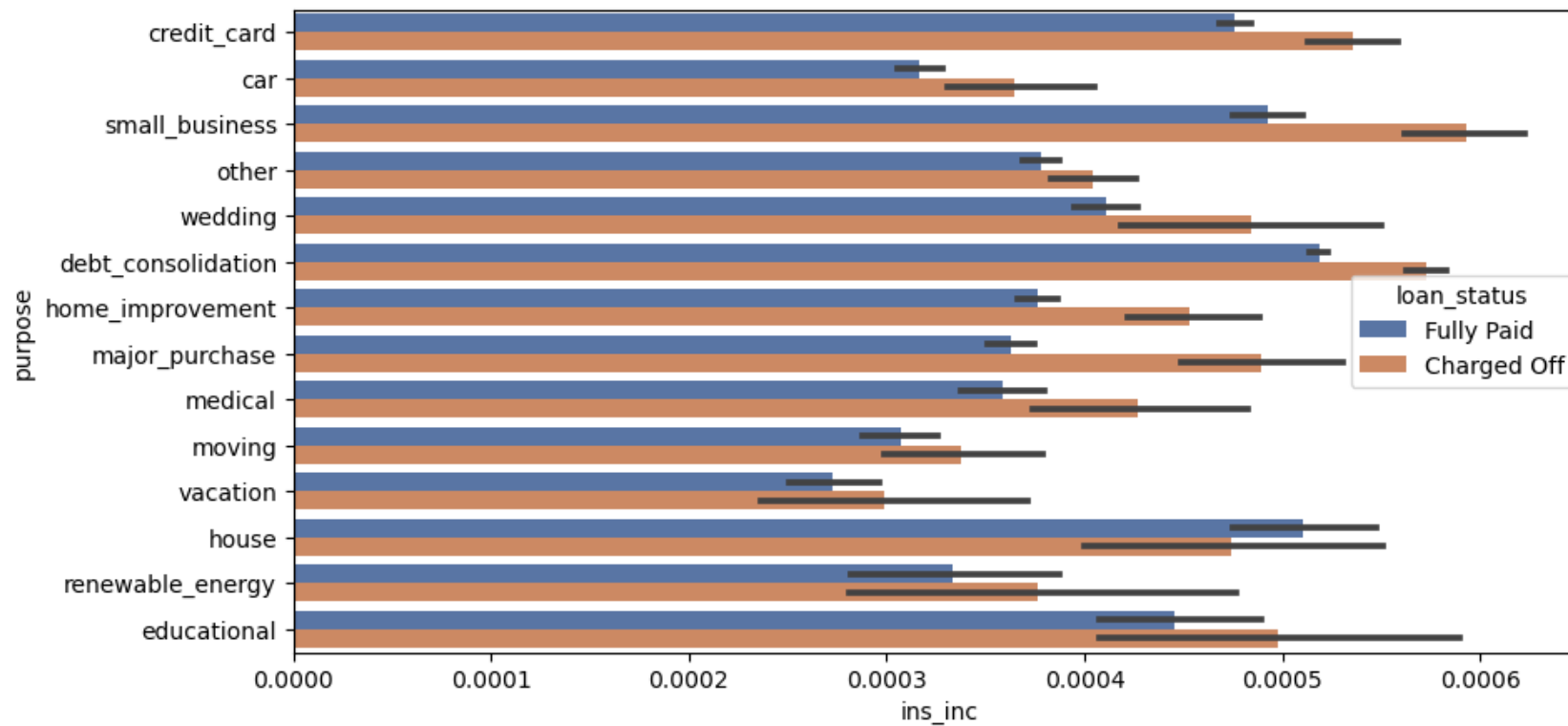
# Analysis and Visualisation with Bivariate:

- Analysis of `loan_status` with `loan_amnt` and `purpose`.



# Analysis and Visualisation with Bivariate:

- Analysis of `loan_status` with `ins_inc` and `purpose`.



# Observations

- In all groups whenever the interest rate is high the possibility of defaulting is also high.
- Whenever the DTI is high, the possibility of defaulting is high. Also true for all brackets/groups.
- Borrowers of Grade G and higher income are less likely to default.
- Across all the variables, the higher the loan granted more is the possibility to default. So finding out the max amount which can be granted for each category would be helpful.
- Borrowers who are of small business and taking a loan of more than 12k are more likely to default.
- Borrowers with education loans of more than 6k are having a hard time closing it. Mostly student loan defaulters.
- Borrowers whose interest rate is more than 11% and who is taking a home loan are more likely to default.
- Borrowers with purpose debt repayment and interest rate of more than 12% are more likely to default.
- Only for home loans, if the ratio of installments to monthly income are high, then more likely the loan will be paid. It's False for all other loan purposes.