International Institute of Information Technology Bangalore

upGrad
#LifeKoKaroLift

LENDING CLUB CASE STUDY

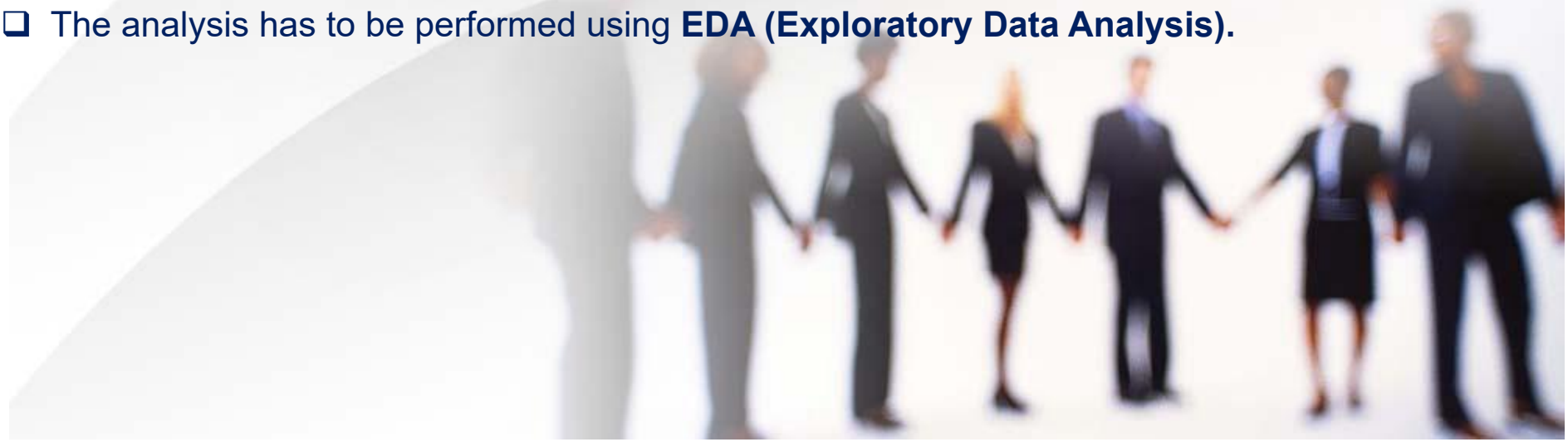
Group Members:
VIpin Jaguri
Deepak Kalarikkal

Problem Statement

- Lending Club is the largest online matketplace facilitating personal as well as business loans.
- It is a **consumer finance company** which specialises in lending various types of loans to urban customers.
- There are two **types of risks** associated with the bank's decision to approve or decline loan request:

  1. If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
  2. 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

- The purpose of our analysis is to understand the driving factors behind loan default. We have to look for indicators(variables) which are correlated with defaulting.
- The analysis has to be performed using **EDA (Exploratory Data Analysis).**



PROBLEM SOLVING METHODOLOGY

- Data understanding and domain knowledge
- Data cleaning
- Datatype treatment
- Derived metrics
- Univariate analysis
- Bivariate analysis
- Data Visualisation and correlation

Data Understanding and Cleaning

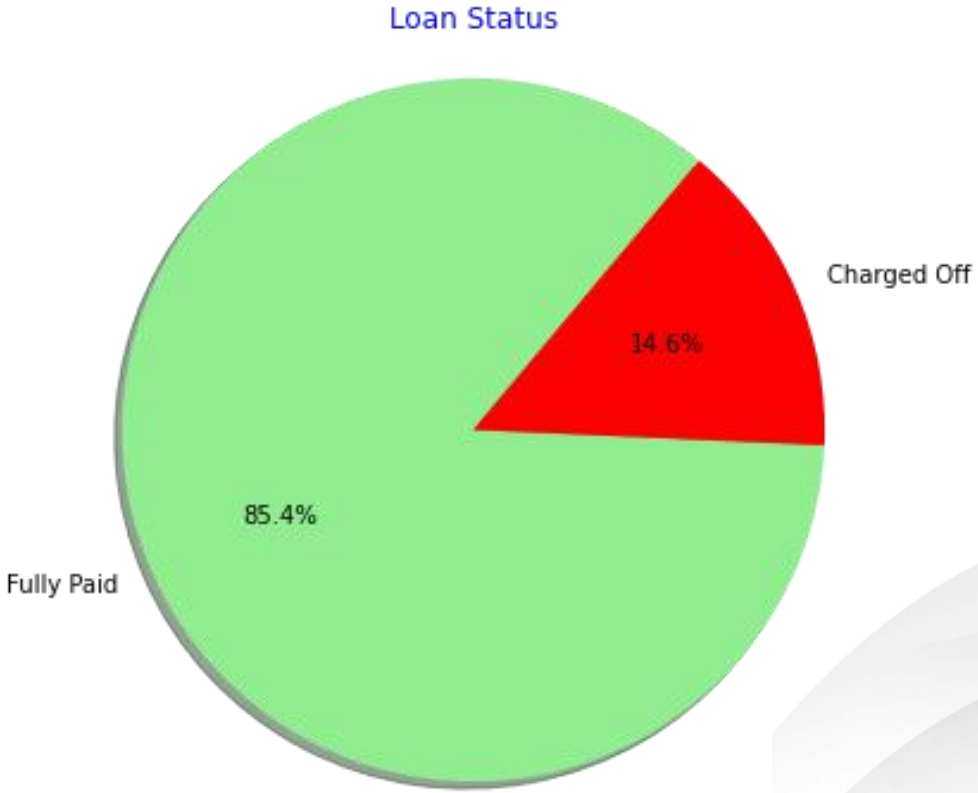
- Loading our dataset and understanding its columns.
- Going through the dataset to understand relevant columns and their interpretations. Domain knowledge is applied to extract some of the insights.

- Fixing Columns:
  - Delete unnecessary columns.
  - Remove columns that have high percentage of missing values.
  - Drop unnecessary columns with string names.
  - Drop columns having only 1 unique values.
  - Drop customer behavior variables as these are not available at time of loan application.

- FIXING ROWS:
  - Removing Irrelevant rows.
  - Imputing missing values with suitable representative values

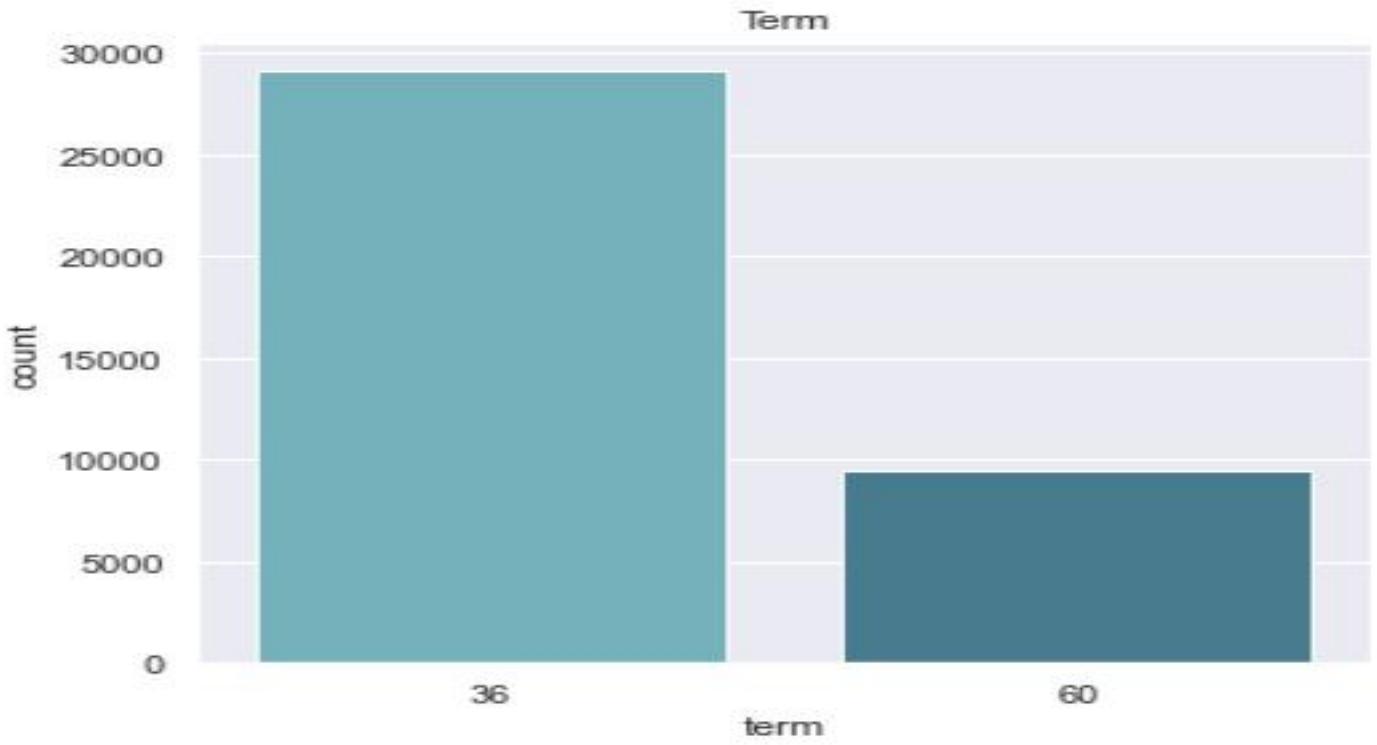
Univariate Analysis

- Plotting Loan_status to visualize distribution of Fully paid and charged off loans.
- We can see that around 15% of the loans resulted in defaults.



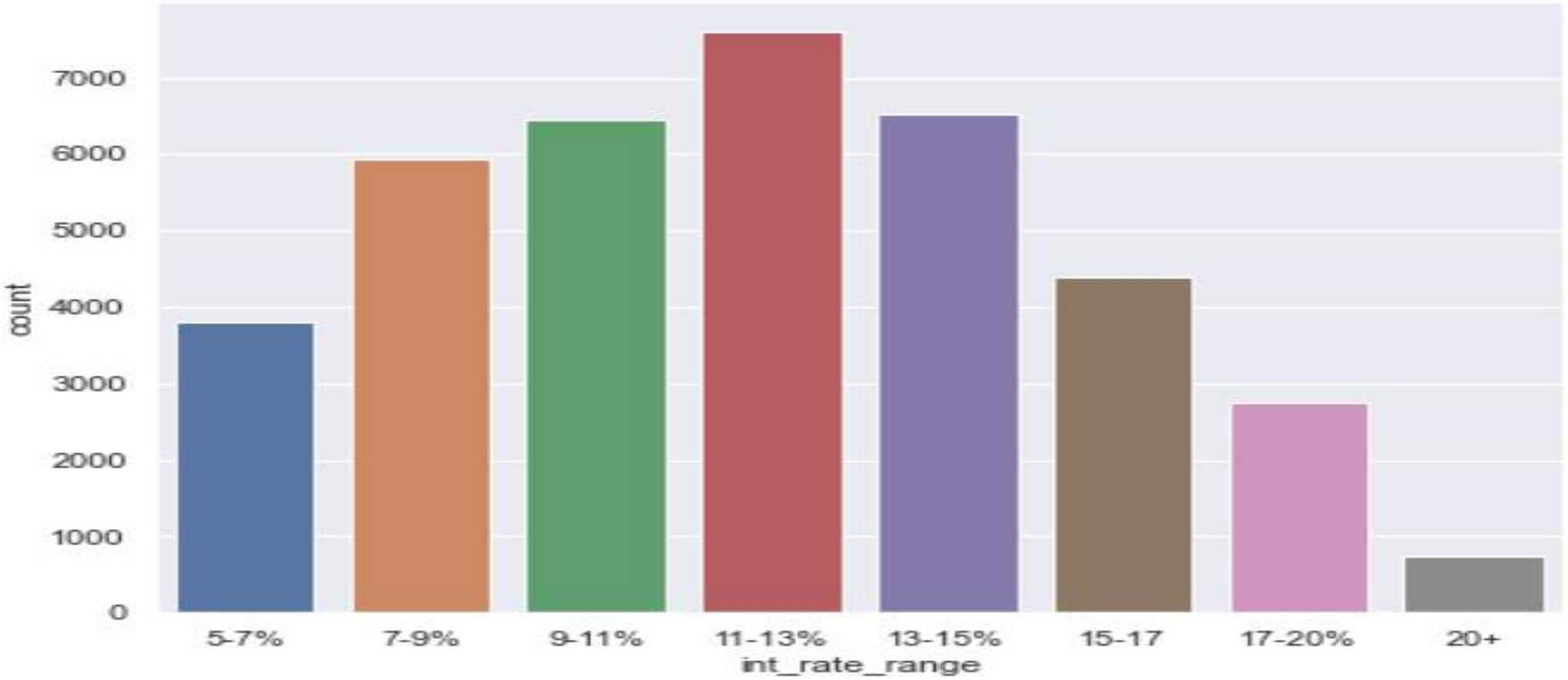
Univariate Analysis

- Plotting terms to visualize distribution of terms across dataset.
- We can see that loans with term 36 months are around 3 times loans with 60 months.



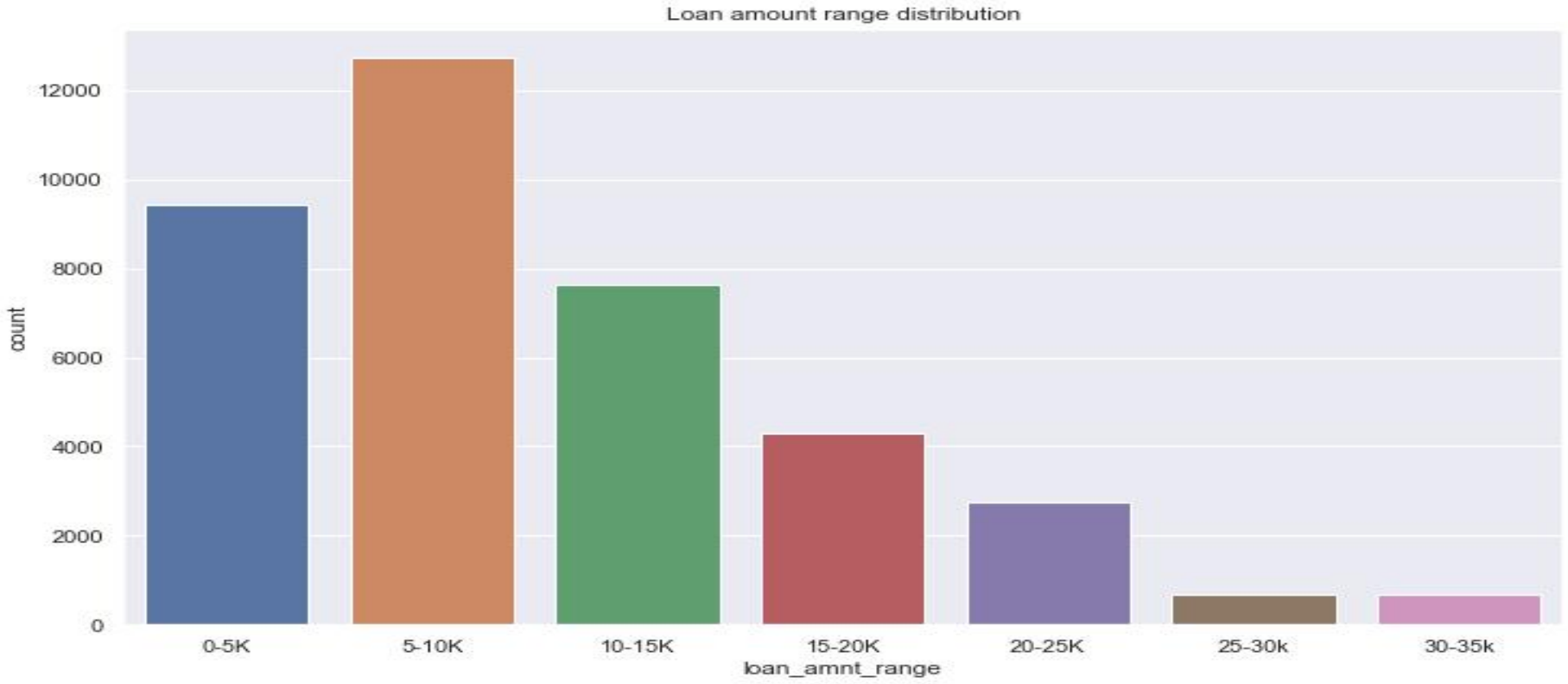
Univariate Analysis

- Plotting interest rates ranges to visualize distribution of interest rates across dataset.
- We can observe that loans are maximum in the interest range 11-13% and loans with higher interest rates are lesser.



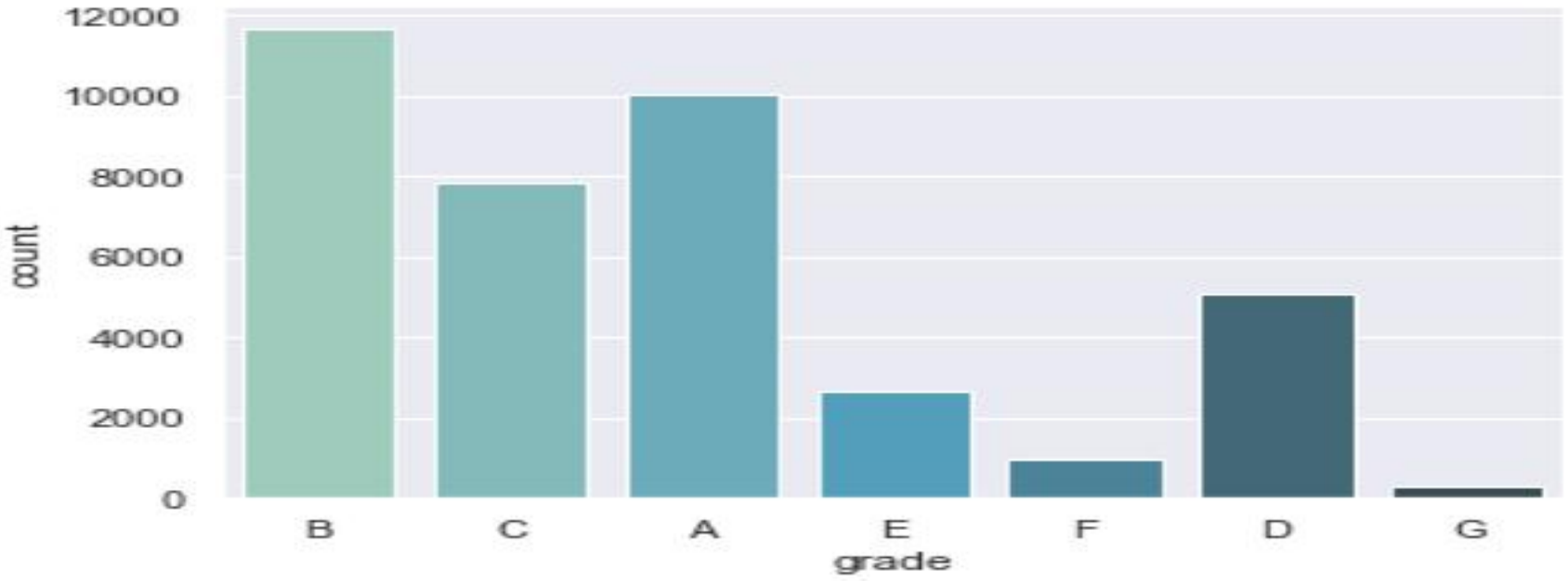
Univariate Analysis

- Plotting loan ammount ranges to visualize distribution of loan ammount range across dataset.
- We can observe that loans are maximum in the loan ammount are maximum in the range 5-10K.



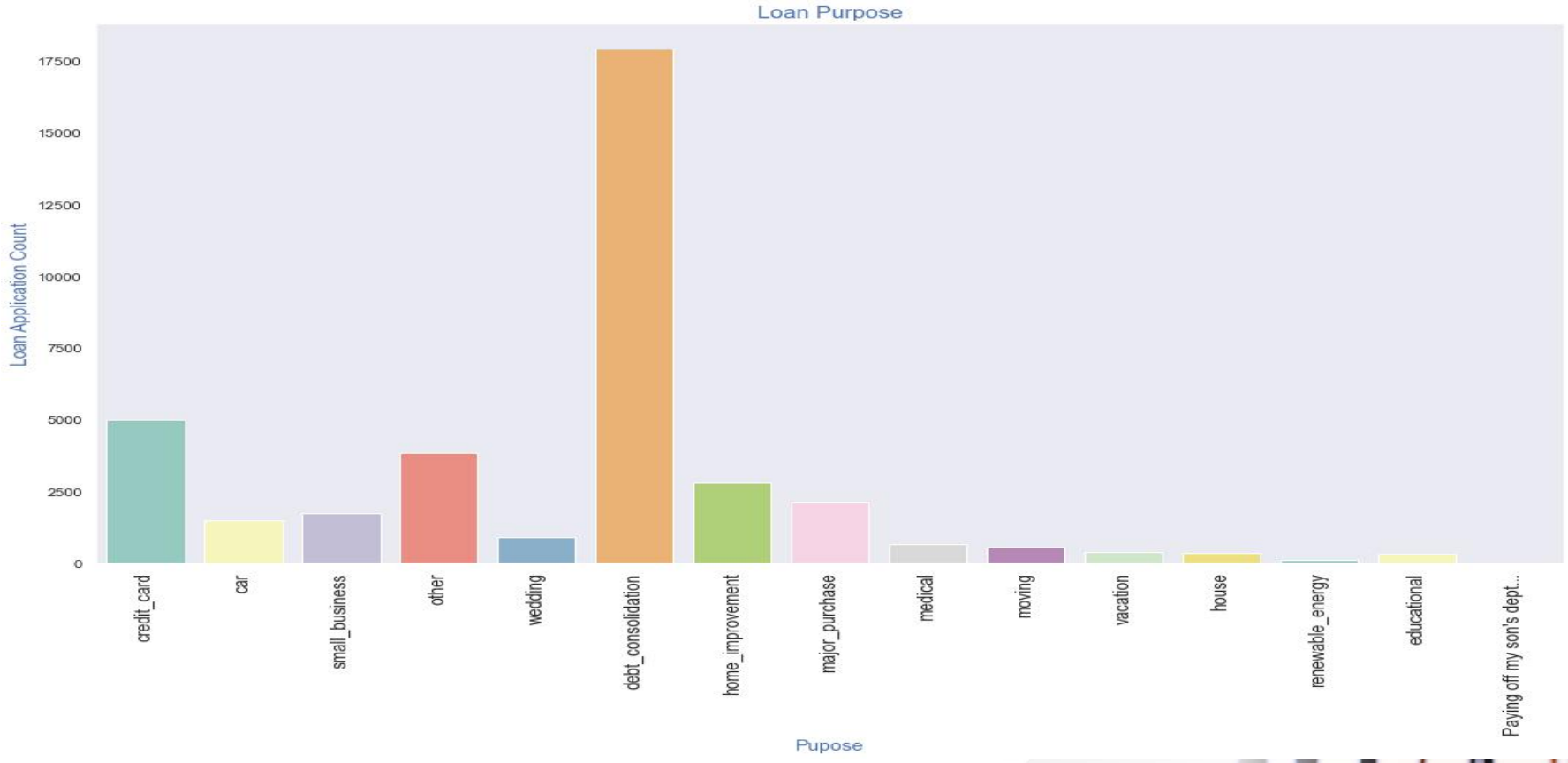
Univariate Analysis

- Plotting grades to visualize distribution of grades across dataset.
- We can observe that maximum loans are of B grades and G grades has minimum loan counts.



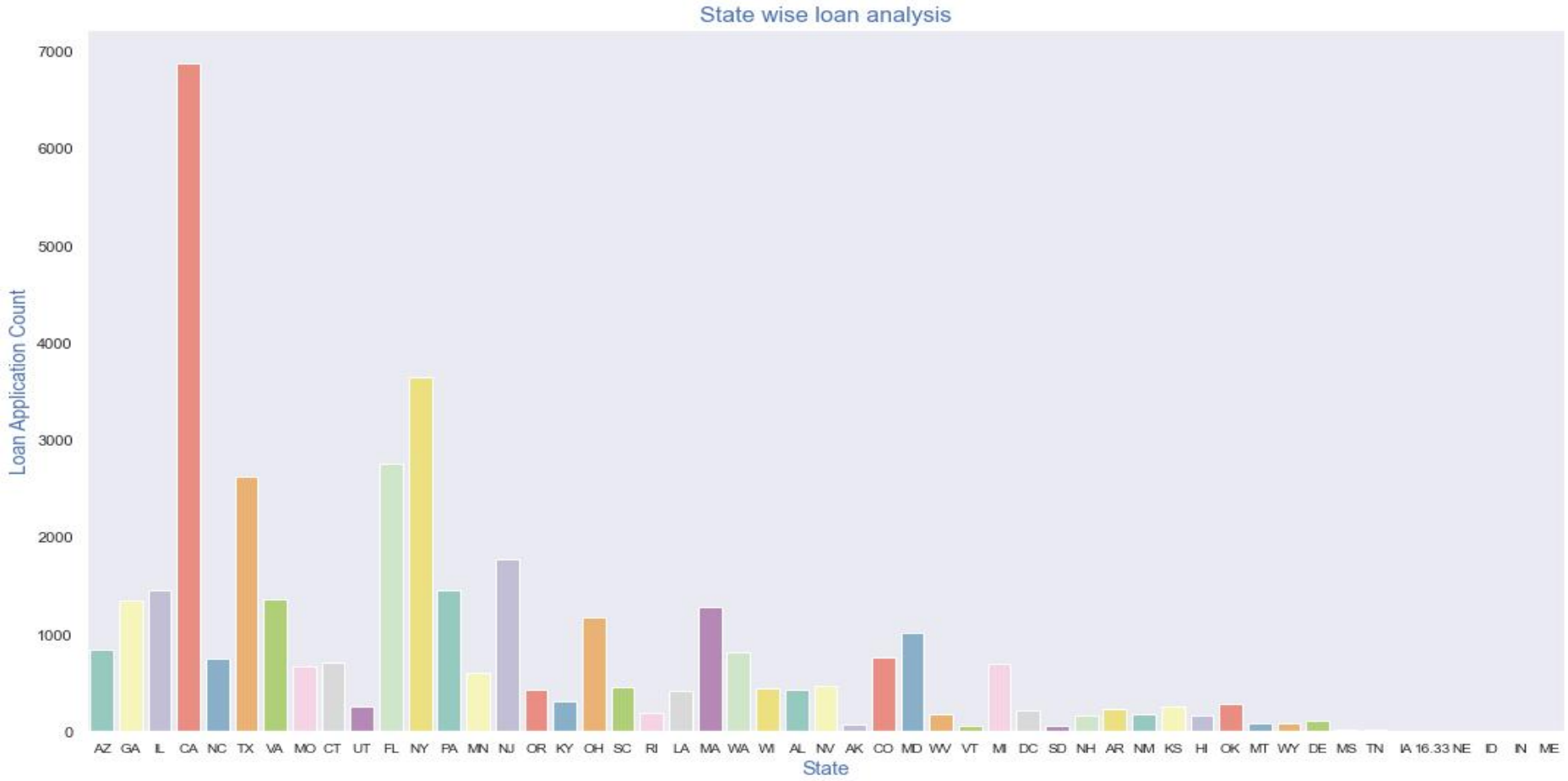
Univariate Analysis

- Plotting purpose of loans to visualize distribution of purpose across dataset.
- We can observe that maximum loans are taken for debt consolidation purpose.



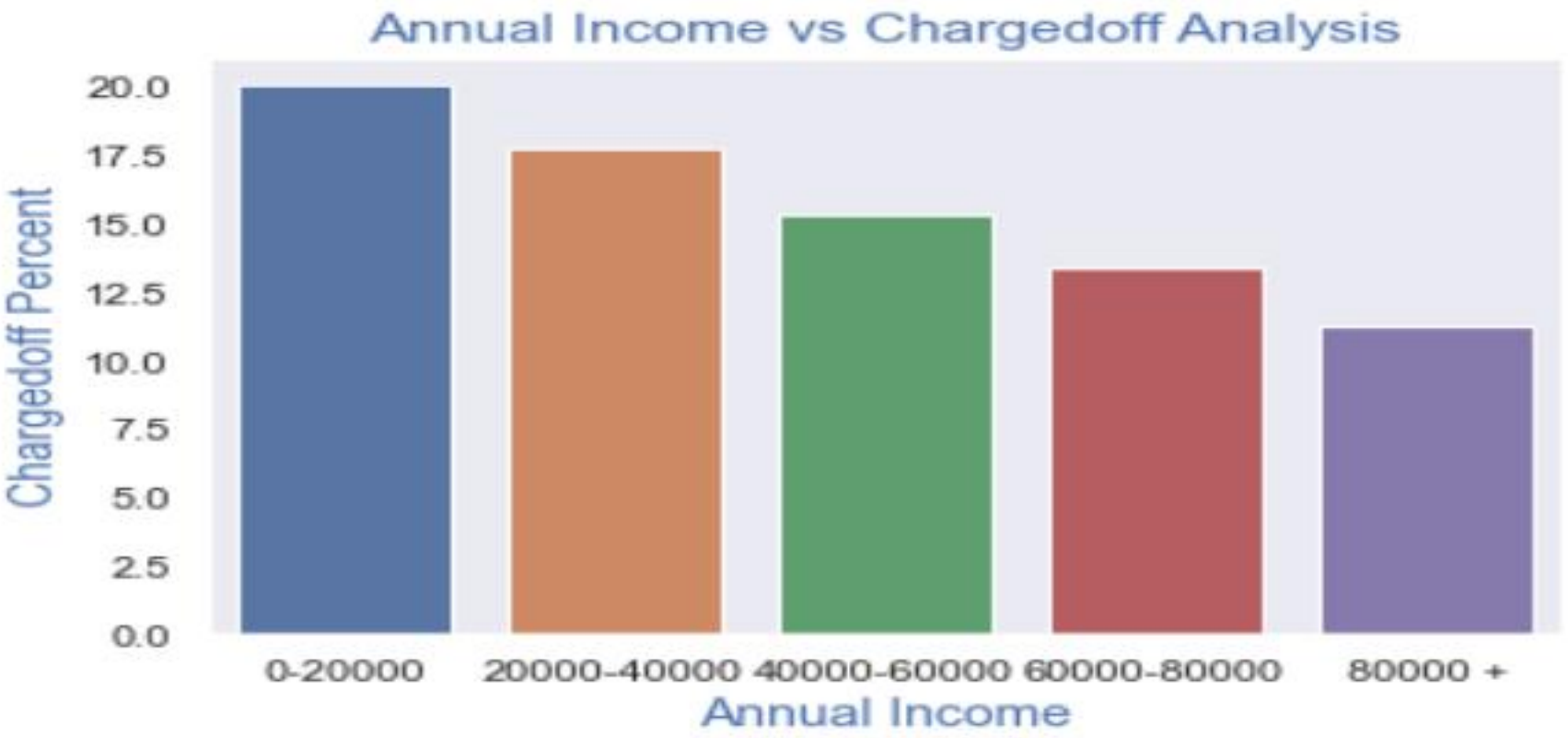
Univariate Analysis

- Plotting purpose of State to visualize distribution of State across dataset.
- We can observe that maximum loans have been from the state of California.



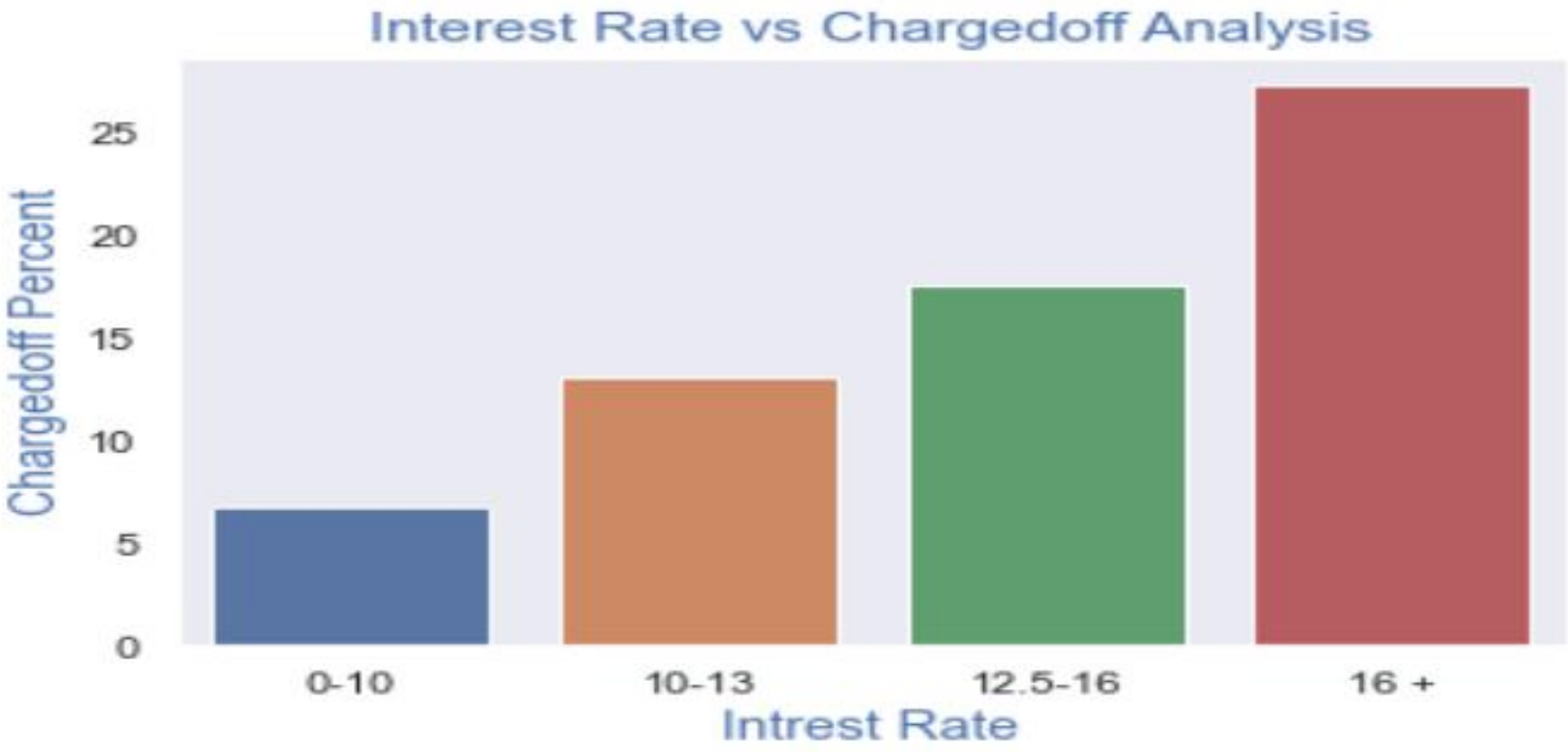
Bivariate Analysis

- Plotting Annual income range against Chargedoff Percentage of loan_status shows that with increase in annaul income, percentage of charged off decreases.



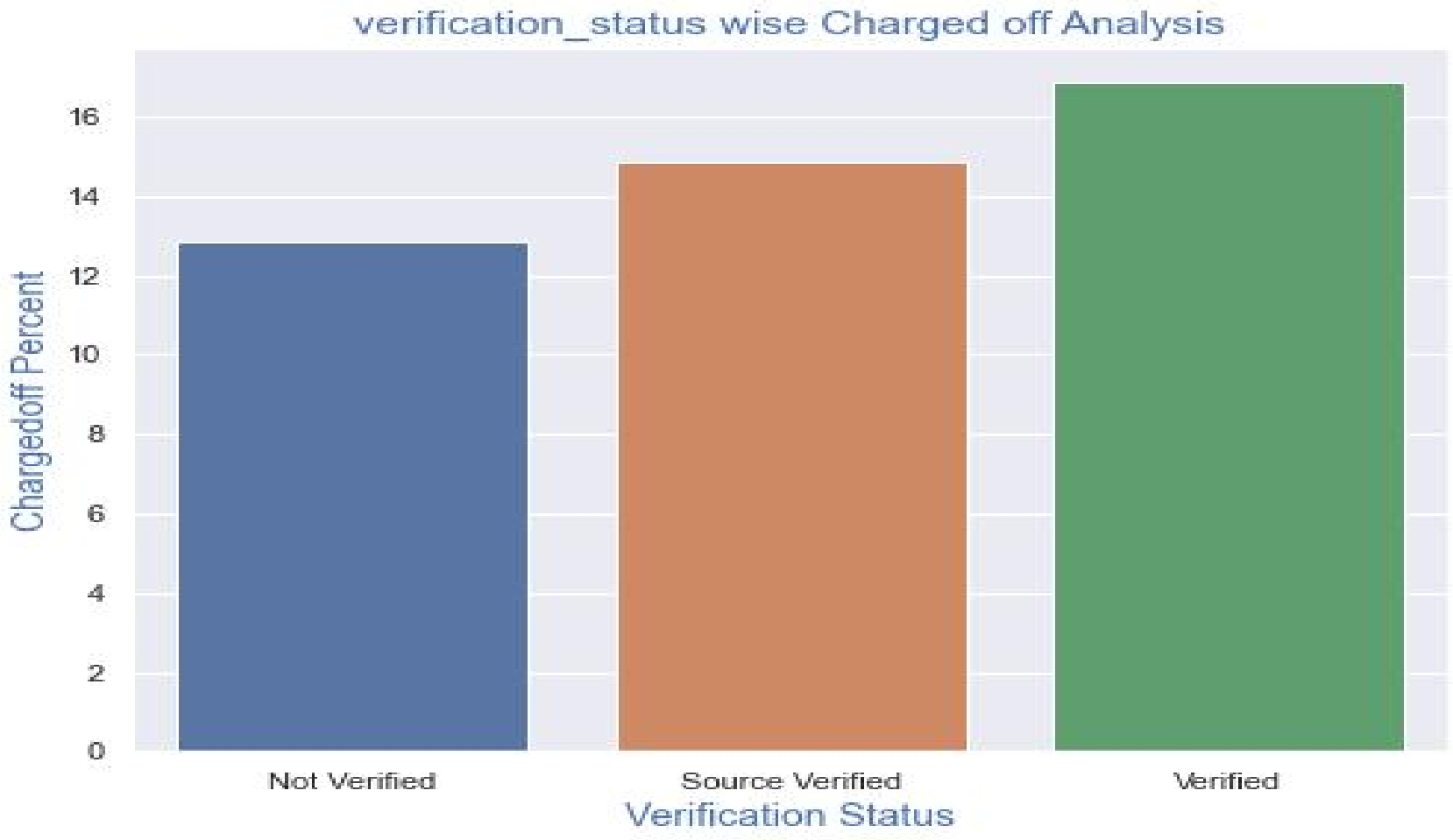
Bivariate Analysis

- Plotting interest range against Chargedoff Percentage of loan_status shows that with increase in interest rates, percentage of charged off increases.



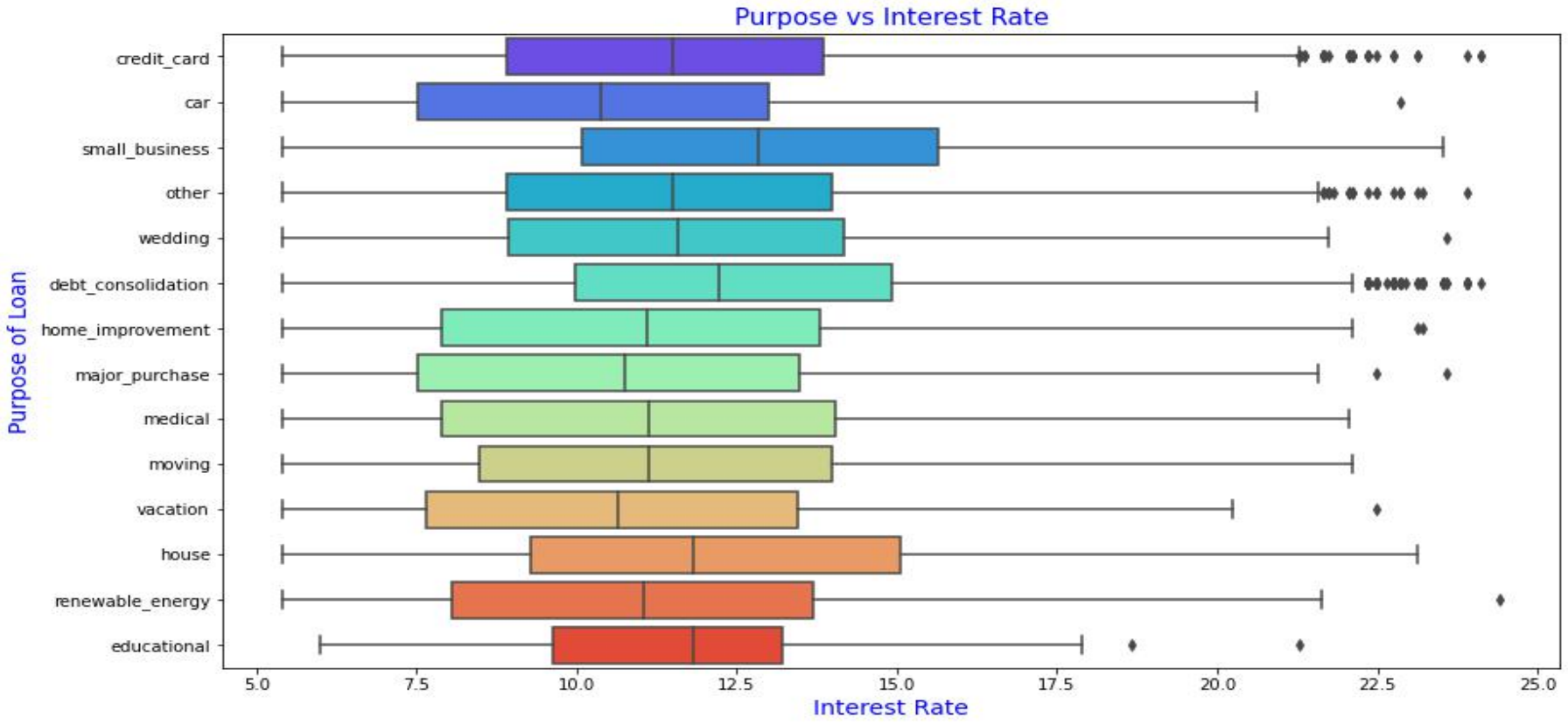
Bivariate Analysis

- Plotting verification status against Chargedoff Percentage of loan_status shows surprising results.
- Loans where verification is done actually have more charged off percentage than those where verification were done at source verification or which were not verified.



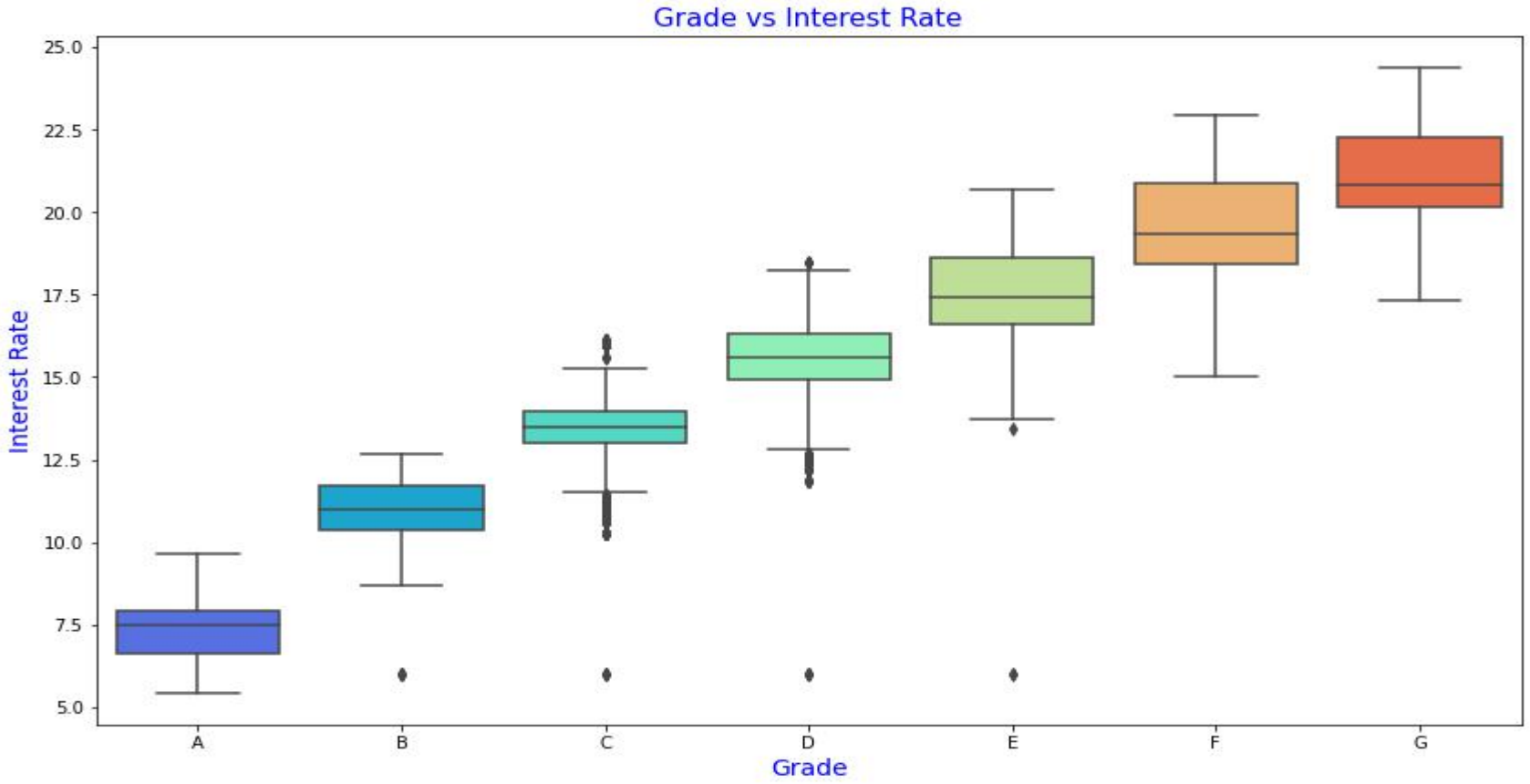
Bivariate Analysis

- We can observe that higher interest rates are charged for small business loan and house loan.



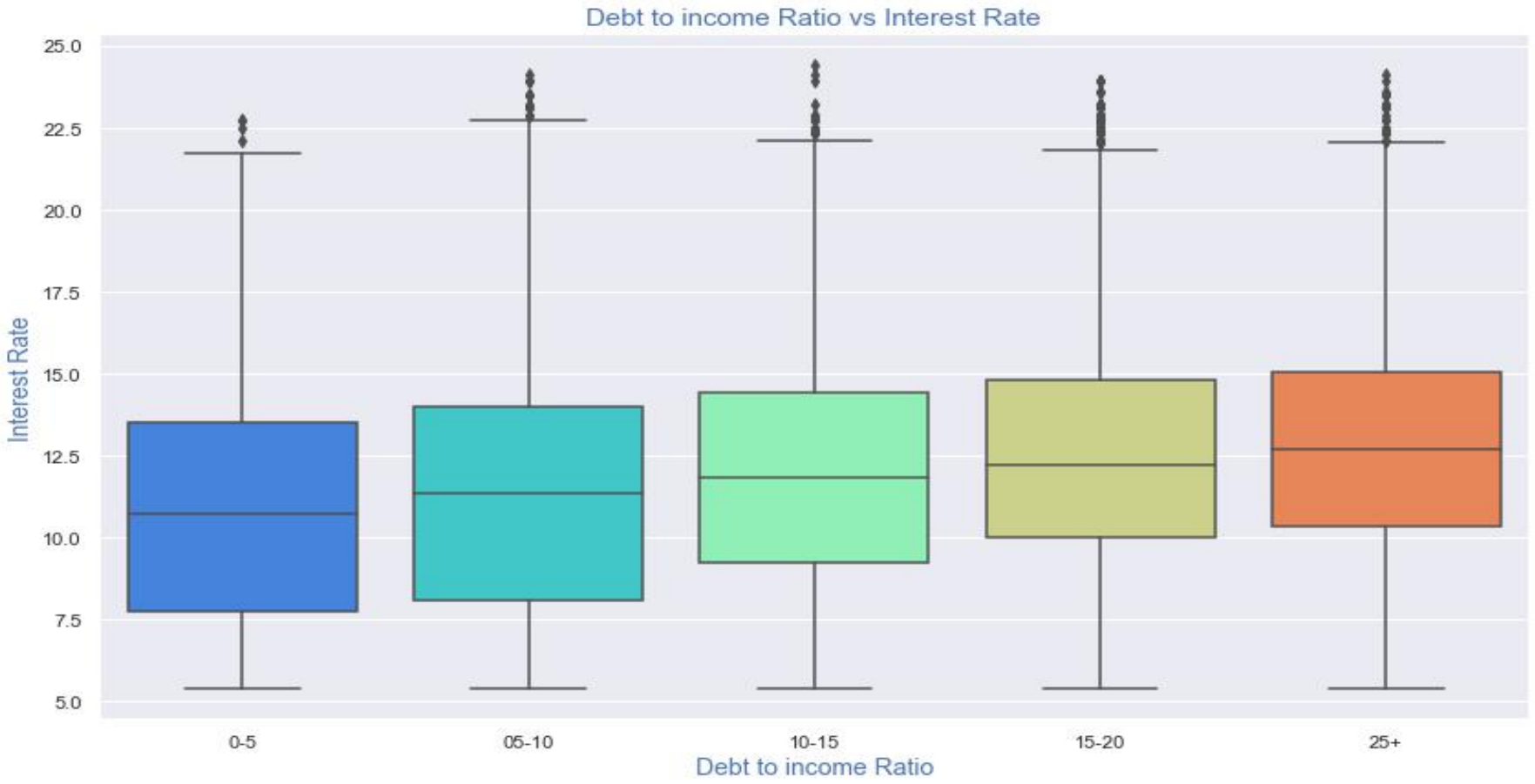
Bivariate Analysis

- Interest Rate increases as we move from Grade A to Grade G

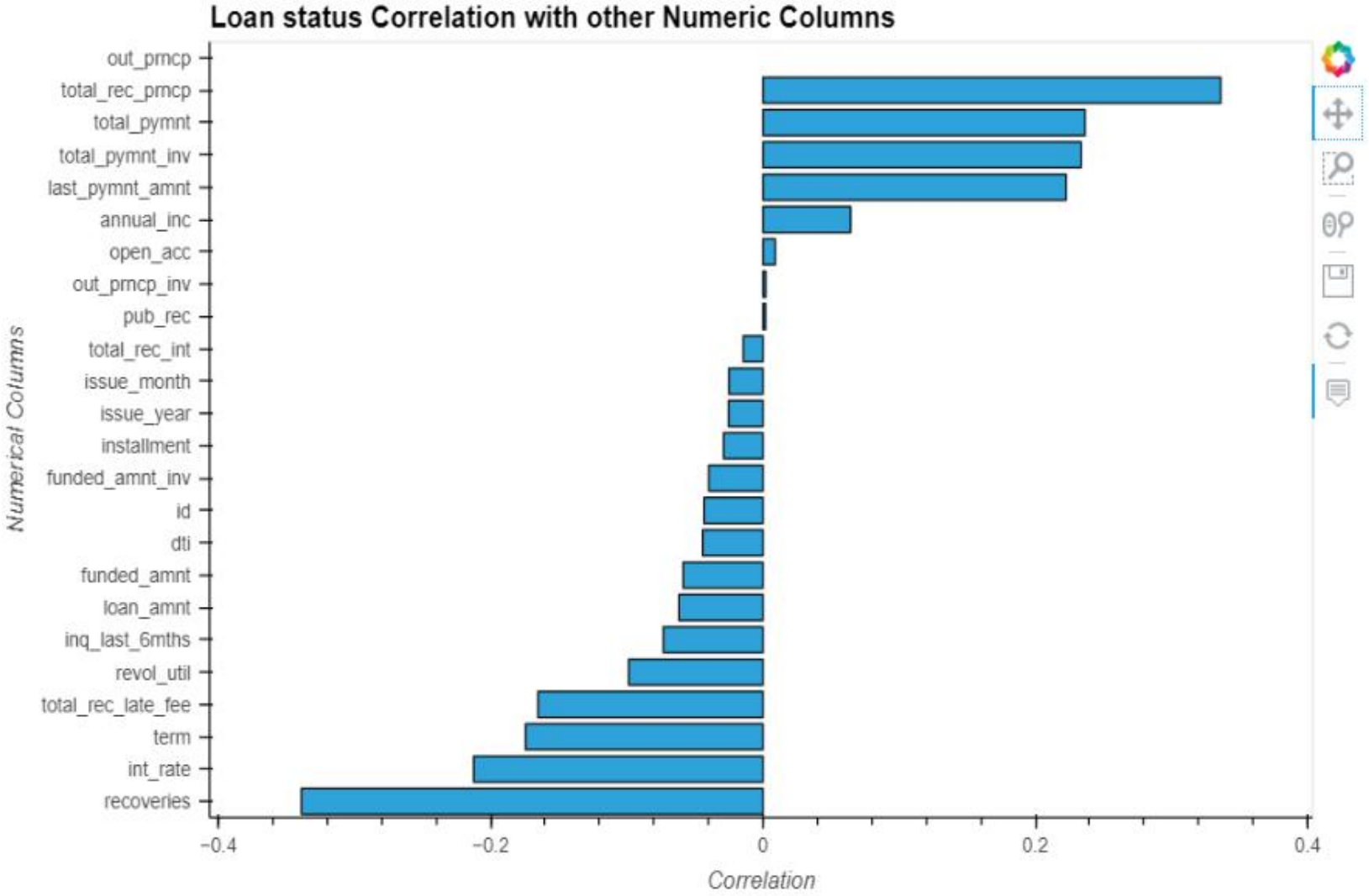


Bivariate Analysis

- Interest rates increases with increase in dti which seems a fair observation.



Correlations



Conclusions from correlations

- Loans with more annual income range have lesser percentage of deafualts than those with lower annual income.
- Loans with 60 moths term have more percentage of loan defaults.
- As the revolving credit utilization percentages of borrowers increases the percentage of loan defaults increases.
- With increase in Term of loan, percentage of loan defaults also increases.

Final Conclusions

- Out of all the borrowers, around 15% of the borrowers defaulted and 85% of the loans were fully paid.
- As we move from Grade A to Grade G, interest rate increases and so does the percentage of loan defaults.
- Another thing to observe is that with increase in annual income, the percentage of loan default decreases.
- With increase in Term of loan, percentage of loan default also increases.

Final Conclusions

- As the revolving credit utilization of borrowers increases the percentage of loan defaults also increases
- Loans taken for the purpose of Debt consolidation and credit card have more percentage of loan defaults.
- In the state of CA ,NY and FL, the percentage of loan defaults is more.
- dti is of great relevance and percentage of loan default also increases with increse in dti.